

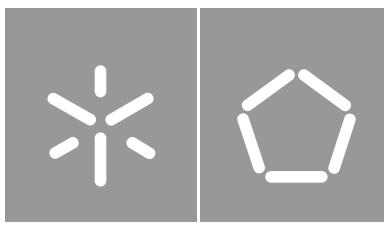


**Universidade do Minho**

Escola de Engenharia

Carlos Amado Lopes Fernandes

**Smart Cities - Otimização inteligente de  
parques de estacionamento**



**Universidade do Minho**

Escola de Engenharia

Carlos Amado Lopes Fernandes

**Smart Cities - Otimização inteligente de  
parques de estacionamento**

Dissertação de Mestrado

Mestrado Integrado em Engenharia e Gestão de Sistemas de  
Informação

Inteligência Artificial

Trabalho efetuado sob a orientação do(a)

**Professor Doutor Carlos Filipe Portela**

**Professor Doutor Manuel Filipe Vieira Torres dos  
Santos**

## Agradecimentos

Com esta dissertação chega ao fim uma etapa que ficará para sempre na minha memória, e quero aproveitar para agradecer a todas as pessoas que tiveram impacto neste meu percurso.

Aos meus pais por todos os esforços e sacrifícios a que se submeteram para que não me faltasse nada durante o curso. Só vocês sabem pelo que passaram para que eu pudesse vir para Guimarães estudar, em busca de um futuro melhor. Quero que saibam que sou-vos eternamente grato e jamais esquecerei o que fizeram por mim.

Ao Prof. Filipe Portela por aceitar orientar-me neste projeto, e por acreditar no meu trabalho. Sem a confiança demonstrada e o conhecimento passado, não seria possível concretizar com sucesso esta dissertação.

Á minha namorada Lola pelo amor e apoio dado, e por me ter incentivado nas fases de maior desmotivação.

Aos pais da Lola por me receberem sempre de braços abertos em sua casa.

Ao meu irmão por estar sempre pronto para me ajudar.

A todos os amigos que fiz durante o curso especialmente ao João Ferreira e ao Hugo Esteves, que são amigos que, certamente, levarei para a vida.

Aos meus amigos de Rio Caldo por toda a amizade e camaradagem.

Aos meus colegas de trabalho e amigos da IOtech, pelas trocas de conhecimento e pelos bons momentos passados na empresa.

## Abstract

The increase in the number of people living in an urban environment has caused some problems in urban mobility, because the investment in infrastructure has not kept pace with the increase. The crowding of cities is a reality to which no one can remain indifferent, and daily the inhabitants and people who frequent urban environments face situations that cause discomfort.

Currently, one of the main problems of urban mobility is traffic congestion. This is largely due to the fact that citizens prefer the use of their own vehicles instead of using public services, which ends up raising some environmental issues in most European cities.

One of the main factors that contributes to traffic congestion is the time spent looking for a parking lot resulting from the difficulty in finding available parking lot.

Many of these problems can be mitigated by optimizing the daily lives in cities. This is precisely the objective of Smart Cities, to ensure more efficiency and effectiveness in city activities. The optimization of city processes is currently dependent on the evolution of information technologies and more precisely on areas such as artificial intelligence, which can provide cities tools and mechanisms that gives them the necessary versatility and intelligence.

The main purpose of this master's dissertation is the optimization of the parking lot searching process. Through Artificial Intelligence and exploration of Machine Learning algorithms, it is intended to provide citizens with information that will enable them to reduce the time spent finding a free parking lot.

To achieve the purpose, data were collected from many platforms, treating them in order to support the Machine Learning models developed in the modeling phase. After this phase was completed, an API Rest was built, which allows the use of intelligent models as a service. Finally, a mobile application was developed that represents the project's functional prototype.

**Keywords:** Urban Mobility, Parking Lot, Smart Cities, Artificial Intelligence, Machine Learning.

## Resumo

O crescente aumento do número de pessoas que vivem em ambiente urbano tem acarretado alguns problemas na mobilidade urbana, isto porque o investimento em infraestruturas não tem acompanhado o aumento. A lotação das cidades é uma realidade à qual não se pode ficar indiferente, e diariamente os habitantes e pessoas que frequentam os ambientes urbanos deparam-se com situações que causam incómodo.

Na atualidade, um dos principais problemas da mobilidade urbana é o congestionamento do trânsito. Isto deve-se muito ao facto dos cidadãos preferirem o uso de transportes motorizados próprios face aos coletivos, o que acaba por levantar algumas questões ambientais em grande parte das cidades europeias.

Um dos principais fatores que contribui para o congestionamento do trânsito é o tempo despendido na procura de um lugar de estacionamento resultante da dificuldade em encontrar lugares disponíveis.

Muitos destes problemas podem ser atenuados otimizando o dia-a-dia das cidades. É precisamente este o objetivo das *Smart Cities*, garantir mais eficiência e eficácia nas atividades citadinas. A otimização dos processos da cidade está, atualmente, dependente da evolução das tecnologias de informação e mais precisamente de áreas como a inteligência artificial, que possam dotar a cidade de ferramentas e mecanismos que lhes confira a versatilidade e inteligência necessária.

A principal finalidade desta dissertação de mestrado inside na otimização do processo de busca de um lugar de estacionamento. Através da Inteligência Artificial e exploração de algoritmos de *Machine Learning* pretende-se fornecer ao cidadão informações que possibilitem a redução do tempo de procura de um lugar para estacionar.

Para atingir a finalidade, foram recolhidos dados de diversas plataformas, tratando-os por forma a darem suporte aos modelos de *Machine Learning* desenvolvidos na fase de modelação. Depois de concluída esta fase, construiu-se uma API Rest que possibilita a utilização dos modelos inteligentes como um serviço. Por fim, foi desenvolvida uma aplicação móvel que representa o protótipo funcional do projeto.

**Palavras-Chave:** Mobilidade Urbana, Parques de Estacionamento, *Smart Cities*, Inteligência Artificial, *Machine Learning*.

# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Enquadramento e Motivação . . . . .	1
1.2	Objetivos . . . . .	3
1.3	Estrutura do Documento . . . . .	4
<b>2</b>	<b>Revisão de Literatura</b>	<b>6</b>
2.1	Mobilidade Urbana . . . . .	7
2.1.1	Urbanização Global e a Mobilidade Urbana . . . . .	8
2.1.2	Mobilidade Urbana nas Áreas Metropolitanas do Porto e Lisboa . . . . .	8
2.1.3	Problemas da Mobilidade Urbana . . . . .	10
2.2	Estacionamento . . . . .	12
2.3	Smart Cities . . . . .	13
2.3.1	Apresentação do conceito . . . . .	14
2.3.2	Âmbito . . . . .	15
2.3.3	Exemplos . . . . .	16
2.4	Inteligência Artificial . . . . .	17
2.4.1	Contextualização e Definições . . . . .	18
2.4.2	Teste de Turing . . . . .	20
2.4.3	A base, os modelos, as técnicas gerais e de aplicação de IA . . . . .	21
2.5	Machine Learning . . . . .	23
2.5.1	Contextualização e Definições . . . . .	23
2.5.2	Tipos de Aprendizagem . . . . .	24
2.5.3	Tipos de Modelos . . . . .	27
2.5.4	Modelos e Algoritmos . . . . .	29
2.6	Deep Learning . . . . .	30
2.7	Trabalhos Relacionados . . . . .	30
<b>3</b>	<b>Abordagem Metodológica</b>	<b>38</b>
3.1	Design Science Research (DSR) . . . . .	38
3.2	CRISP-DM . . . . .	40
3.3	Scrum . . . . .	44

3.4	Tecnologias e Plataformas . . . . .	48
<b>4</b>	<b>Trabalho Realizado</b>	<b>50</b>
4.1	Compreensão dos dados . . . . .	50
4.1.1	Recolha de dados inicial . . . . .	51
4.1.2	Descrição dos dados . . . . .	56
4.1.3	Exploração dos dados . . . . .	67
4.1.4	Qualidade dos dados . . . . .	80
4.2	Preparação dos dados . . . . .	87
4.2.1	Seleção dos dados . . . . .	87
4.2.2	Limpeza e construção dos dados e carregamento para base de dados de estágio	94
4.2.3	Integração e Formatação de dados . . . . .	100
4.2.4	Síntese da recolha, tratamento e carregamento dos dados . . . . .	104
4.3	Modelação . . . . .	105
4.3.1	Seleção das técnicas de modelação . . . . .	105
4.3.2	Criação de cenários de teste . . . . .	106
4.3.3	Construção do modelo . . . . .	109
4.3.4	Avaliação do modelo . . . . .	111
4.3.5	Esquema síntese da modelação . . . . .	115
4.4	Avaliação . . . . .	116
4.4.1	Avaliar resultados . . . . .	116
4.4.2	Rever o processo . . . . .	116
4.4.3	Determinar os próximos passos . . . . .	117
4.5	Implementação . . . . .	118
4.5.1	API . . . . .	118
4.5.2	Aplicação Mobile . . . . .	121
4.5.3	Planear a implementação . . . . .	127
4.5.4	Planear a monitorização e manutenção . . . . .	127
4.5.5	Produzir o relatório final . . . . .	127
4.5.6	Rever o projeto . . . . .	128
<b>5</b>	<b>Conclusão</b>	<b>129</b>
5.1	Reflexão Crítica . . . . .	129

5.2	Análise de Riscos . . . . .	133
5.3	Trabalho Futuro . . . . .	135

**Bibliografia** **136**

**Anexos** **141**

A	Recolha de dados dos Locais . . . . .	141
A.1	Código fonte de recolha dos Locais - Parte 1 . . . . .	141
A.2	Código fonte de recolha dos Locais - Parte 2 . . . . .	142
B	Recolha de dados da Meteorologia . . . . .	143
C	Recolha de dados dos Feriados . . . . .	144
D	Recolha de dados das Ocupações . . . . .	148
E	Exploração dos dados dos Parques de estacionamento . . . . .	149
F	Exploração dos dados dos Locais . . . . .	156
G	Exploração dos dados da Meteorologia . . . . .	164
H	Exploração dos dados dos Feriados . . . . .	179
I	Exploração dos dados das Ocupações . . . . .	185



## Índice de Figuras

Figura 1: Âmbito das <i>Smart Cities</i> . . . . .	15
Figura 2: O teste de Turing . . . . .	21
Figura 3: Inteligência Artificial - Estrutura de 4 níveis . . . . .	22
Figura 4: Tipos de Aprendizagem e Modelos do ML . . . . .	24
Figura 5: Metodologia Design Science Research . . . . .	39
Figura 6: Metodologia CRISP-DM . . . . .	41
Figura 7: Metodologia Scrum . . . . .	46
Figura 8: Parques de estacionamento - Estatísticas Simples do atributo 'parking_group_name' . . . . .	69
Figura 9: Parques de estacionamento - Frequências do atributo 'parking_group_name' . . . . .	69
Figura 10: Parques de estacionamento - Estatísticas Simples do atributo 'park_name' . . . . .	69
Figura 11: Parques de estacionamento - Estatísticas Simples do atributo 'ParkingNumberOfSpaces' . . . . .	70
Figura 12: Parques de estacionamento - Frequências do atributo 'ParkingNumberOfSpaces' . . . . .	70
Figura 13: Parques de estacionamento - Estatísticas do atributo 'ParkingNumberOfSpaces' . . . . .	71
Figura 14: Locais - Estatísticas Simples do atributo 'place_name' . . . . .	72
Figura 15: Locais - Frequências do atributo 'place_name' . . . . .	72
Figura 16: Locais - Estatísticas Simples do atributo 'place_rating' . . . . .	72
Figura 17: Locais - Frequências do atributo 'place_rating' . . . . .	73
Figura 18: Locais - Estatísticas do atributo 'place_rating' . . . . .	73
Figura 19: Locais - Estatísticas Simples do atributo 'place_user_ratings_total' . . . . .	73
Figura 20: Locais - Frequências do atributo 'place_user_ratings_total' . . . . .	74
Figura 21: Locais - Estatísticas do atributo 'place_user_ratings_total' . . . . .	74
Figura 22: Meteorologia - Estatísticas Simples do atributo 'hourly_precipIntensity' . . . . .	75
Figura 23: Meteorologia - Frequências do atributo 'hourly_precipIntensity' . . . . .	75
Figura 24: Meteorologia - Estatísticas do atributo 'hourly_precipIntensity' . . . . .	75
Figura 25: Meteorologia - Estatísticas Simples do atributo 'hourly_temperature' . . . . .	76
Figura 26: Meteorologia - Frequências do atributo 'hourly_temperature' . . . . .	76
Figura 27: Meteorologia - Estatísticas do atributo 'hourly_temperature' . . . . .	76
Figura 28: Feriados - Estatísticas Simples do atributo 'name' . . . . .	77
Figura 29: Feriados - Frequências do atributo 'name' . . . . .	77
Figura 30: Feriados - Estatísticas Simples do atributo 'type' . . . . .	78
Figura 31: Feriados - Frequências do atributo 'type' . . . . .	78

Figura 32: Feriados - Estatísticas Simples do atributo 'public' . . . . .	78
Figura 33: Feriados - Frequências do atributo 'public' . . . . .	79
Figura 34: Ocupação dos parques - Estatísticas Simples do atributo 'occupation' . . . . .	79
Figura 35: Ocupações dos parques - Frequências do atributo 'occupation' . . . . .	80
Figura 36: Ocupações dos parques - Estatísticas do atributo 'occupation' . . . . .	80
Figura 37: Exemplo 1 do atributo 'tariffsAndPayment' . . . . .	95
Figura 38: Exemplo 2 do atributo 'tariffsAndPayment' . . . . .	96
Figura 39: Exemplo 3 do atributo 'tariffsAndPayment' . . . . .	97
Figura 40: Diagrama de entidades e relacionamentos da base de dados de estágio . . . . .	100
Figura 41: Esquema da etapa de integração de dados . . . . .	104
Figura 42: Esquema da recolha, tratamento e carregamento dos dados . . . . .	104
Figura 43: Esquema síntese da modelação . . . . .	116
Figura 44: Documentação API . . . . .	119
Figura 45: Exemplo de resposta da API . . . . .	120
Figura 46: Componentes que interagem com a API . . . . .	121
Figura 47: Aplicação - Splash Screen . . . . .	122
Figura 48: Aplicação - Primeiro ecrã . . . . .	122
Figura 49: Aplicação - Preenchimento do local . . . . .	122
Figura 50: Aplicação - Lista de locais candidatos . . . . .	123
Figura 51: Aplicação - Seleção do Local . . . . .	123
Figura 52: Aplicação - Indicação do dia e hora de chegada . . . . .	124
Figura 53: Aplicação - Ecrã de seleção do parque de estacionamento . . . . .	125
Figura 54: Aplicação - Rota mostrada ao utilizador . . . . .	126
Figura 55: Aplicação - Google Maps para iniciar navegação . . . . .	126
Figura 56: Componentes que interagem com a aplicação . . . . .	126
Figura E.1: Parques de estacionamento - Estatísticas Simples do atributo id . . . . .	150
Figura E.2: Parques de estacionamento - Estatísticas Simples do atributo latitude . . . . .	150
Figura E.3: Parques de estacionamento - Estatísticas Simples do atributo longitude . . . . .	150
Figura E.4: Parques de estacionamento - Estatísticas Simples do atributo 'parking_group_name' . . . . .	151
Figura E.5: Parques de estacionamento - Frequências do atributo 'parking_group_name' . . . . .	151
Figura E.6: Parques de estacionamento - Estatísticas Simples do atributo 'park_name' . . . . .	151
Figura E.7: Parques de estacionamento - Estatísticas Simples do atributo 'parkingHouseNumber' . . . . .	152

Figura E.8: Parques de estacionamento - Frequências do atributo 'parkingHouseNumber' . . . . .	152
Figura E.9: Parques de estacionamento - Estatísticas Simples do atributo 'ParkingNumberOfSpaces' . . . . .	152
Figura E.10: Parques de estacionamento - Frequências do atributo 'ParkingNumberOfSpaces' . . . . .	153
Figura E.11: Parques de estacionamento - Estatísticas do atributo 'ParkingNumberOfSpaces' . . . . .	153
Figura E.12: Parques de estacionamento - Estatísticas Simples do atributo 'ParkingPostalCode' . . . . .	154
Figura E.13: Parques de estacionamento - Frequências do atributo 'ParkingPostalCode' . . . . .	154
Figura E.14: Parques de estacionamento - Estatísticas Simples do atributo 'ParkingStreet' . . . . .	154
Figura E.15: Parques de estacionamento - Frequências do atributo 'ParkingStreet' . . . . .	155
Figura E.16: Parques de estacionamento - Estatísticas Simples do atributo 'UrbanParkingSiteType' . . . . .	155
Figura E.17: Parques de estacionamento - Frequências do atributo 'UrbanParkingSiteType' . . . . .	156
Figura F.1: Locais - Estatísticas Simples do atributo 'park_latitude' . . . . .	156
Figura F.2: Locais - Frequências do atributo 'park_latitude' . . . . .	157
Figura F.3: Locais - Estatísticas Simples do atributo 'park_longitude' . . . . .	157
Figura F.4: Locais - Frequências do atributo 'park_longitude' . . . . .	158
Figura F.5: Locais - Estatísticas Simples do atributo 'park_name' . . . . .	158
Figura F.6: Locais - Estatísticas Simples do atributo 'place_latitude' . . . . .	158
Figura F.7: Locais - Frequências do atributo 'place_latitude' . . . . .	159
Figura F.8: Locais - Estatísticas Simples do atributo 'place_longitude' . . . . .	159
Figura F.9: Locais - Frequências do atributo 'place_longitude' . . . . .	160
Figura F.10: Locais - Estatísticas Simples do atributo 'place_name' . . . . .	160
Figura F.11: Locais - Frequências do atributo 'place_name' . . . . .	161
Figura F.12: Locais - Estatísticas Simples do atributo 'place_vicinity' . . . . .	161
Figura F.13: Locais - Frequências do atributo 'place_vicinity' . . . . .	161
Figura F.14: Locais - Estatísticas Simples do atributo 'place_rating' . . . . .	162
Figura F.15: Locais - Frequências do atributo 'place_rating' . . . . .	162
Figura F.16: Locais - Estatísticas do atributo 'place_rating' . . . . .	162
Figura F.17: Locais - Estatísticas Simples do atributo 'place_user_ratings_total' . . . . .	163
Figura F.18: Locais - Frequências do atributo 'place_user_ratings_total' . . . . .	163
Figura F.19: Locais - Estatísticas do atributo 'place_user_ratings_total' . . . . .	163
Figura G.1: Meteorologia - Estatísticas Simples do atributo 'latitude' . . . . .	164
Figura G.2: Meteorologia - Frequências do atributo 'latitude' . . . . .	164
Figura G.3: Meteorologia - Estatísticas Simples do atributo 'longitude' . . . . .	165

Figura G.4: Meteorologia - Frequências do atributo 'longitude'	165
Figura G.5: Meteorologia - Estatísticas Simples do atributo 'hourly_time'	165
Figura G.6: Meteorologia - Estatísticas Simples do atributo 'hourly_summary'	166
Figura G.7: Meteorologia - Frequências do atributo 'hourly_summary'	166
Figura G.8: Meteorologia - Estatísticas Simples do atributo 'hourly_precipIntensity'	166
Figura G.9: Meteorologia - Frequências do atributo 'hourly_precipIntensity'	167
Figura G.10: Meteorologia - Estatísticas do atributo 'hourly_precipIntensity'	167
Figura G.11: Meteorologia - Estatísticas Simples do atributo 'hourly_precipProbability'	167
Figura G.12: Meteorologia - Frequências do atributo 'hourly_precipProbability'	168
Figura G.13: Meteorologia - Estatísticas do atributo 'hourly_precipProbability'	168
Figura G.14: Meteorologia - Estatísticas Simples do atributo 'hourly_temperature'	168
Figura G.15: Meteorologia - Frequências do atributo 'hourly_temperature'	169
Figura G.16: Meteorologia - Estatísticas do atributo 'hourly_temperature'	169
Figura G.17: Meteorologia - Estatísticas Simples do atributo 'hourly_apparentTemperature'	169
Figura G.18: Meteorologia - Frequências do atributo 'hourly_apparentTemperature'	170
Figura G.19: Meteorologia - Estatísticas do atributo 'hourly_apparentTemperature'	170
Figura G.20: Meteorologia - Estatísticas Simples do atributo 'hourly_dewPoint'	170
Figura G.21: Meteorologia - Frequências do atributo 'hourly_dewPoint'	171
Figura G.22: Meteorologia - Estatísticas do atributo 'hourly_dewPoint'	171
Figura G.23: Meteorologia - Estatísticas Simples do atributo 'hourly_humidity'	171
Figura G.24: Meteorologia - Frequências do atributo 'hourly_humidity'	172
Figura G.25: Meteorologia - Estatísticas do atributo 'hourly_humidity'	172
Figura G.26: Meteorologia - Estatísticas Simples do atributo 'hourly_pressure'	172
Figura G.27: Meteorologia - Frequências do atributo 'hourly_pressure'	173
Figura G.28: Meteorologia - Estatísticas do atributo 'hourly_pressure'	173
Figura G.29: Meteorologia - Estatísticas Simples do atributo 'hourly_windSpeed'	173
Figura G.30: Meteorologia - Frequências do atributo 'hourly_windSpeed'	174
Figura G.31: Meteorologia - Estatísticas do atributo 'hourly_windSpeed'	174
Figura G.32: Meteorologia - Estatísticas Simples do atributo 'hourly_windGust'	174
Figura G.33: Meteorologia - Frequências do atributo 'hourly_windGust'	175
Figura G.34: Meteorologia - Estatísticas do atributo 'hourly_windGust'	175
Figura G.35: Meteorologia - Estatísticas Simples do atributo 'hourly_windBearing'	175

Figura G.36: Meteorologia - Frequências do atributo 'hourly_windBearing' . . . . .	176
Figura G.37: Meteorologia - Estatísticas do atributo 'hourly_windBearing' . . . . .	176
Figura G.38: Meteorologia - Estatísticas Simples do atributo 'hourly_cloudCover' . . . . .	176
Figura G.39: Meteorologia - Frequências do atributo 'hourly_cloudCover' . . . . .	177
Figura G.40: Meteorologia - Estatísticas do atributo 'hourly_cloudCover' . . . . .	177
Figura G.41: Meteorologia - Estatísticas Simples do atributo 'hourly_uvIndex' . . . . .	177
Figura G.42: Meteorologia - Frequências do atributo 'hourly_uvIndex' . . . . .	178
Figura G.43: Meteorologia - Estatísticas do atributo 'hourly_uvIndex' . . . . .	178
Figura G.44: Meteorologia - Estatísticas Simples do atributo 'hourly_visibility' . . . . .	178
Figura G.45: Meteorologia - Frequências do atributo 'hourly_visibility' . . . . .	179
Figura G.46: Meteorologia - Estatísticas do atributo 'hourly_visibility' . . . . .	179
Figura H.1: Feriados - Estatísticas Simples do atributo 'year' . . . . .	179
Figura H.2: Feriados - Frequências do atributo 'year' . . . . .	180
Figura H.3: Feriados - Estatísticas Simples do atributo 'date' . . . . .	180
Figura H.4: Feriados - Frequências do atributo 'date' . . . . .	181
Figura H.5: Feriados - Estatísticas Simples do atributo 'start' . . . . .	181
Figura H.6: Feriados - Frequências do atributo 'start' . . . . .	182
Figura H.7: Feriados - Estatísticas Simples do atributo 'end' . . . . .	182
Figura H.8: Feriados - Frequências do atributo 'end' . . . . .	182
Figura H.9: Feriados - Estatísticas Simples do atributo 'name' . . . . .	183
Figura H.10: Feriados - Frequências do atributo 'name' . . . . .	183
Figura H.11: Feriados - Estatísticas Simples do atributo 'type' . . . . .	183
Figura H.12: Feriados - Frequências do atributo 'type' . . . . .	184
Figura H.13: Feriados - Estatísticas Simples do atributo 'public' . . . . .	184
Figura H.14: Feriados - Frequências do atributo 'public' . . . . .	185
Figura H.15: Feriados - Estatísticas Simples do atributo 'country' . . . . .	185
Figura I.1: Ocupação dos parques - Estatísticas Simples do atributo 'park_name' . . . . .	186
Figura I.2: Ocupações dos parques - Frequências do atributo 'park_name' . . . . .	186
Figura I.3: Ocupação dos parques - Estatísticas Simples do atributo 'park_id' . . . . .	187
Figura I.4: Ocupações dos parques - Frequências do atributo 'park_id' . . . . .	187
Figura I.5: Ocupação dos parques - Estatísticas Simples do atributo 'date' . . . . .	187
Figura I.6: Ocupações dos parques - Frequências do atributo 'date' . . . . .	188

Figura I.7: Ocupações dos parques - Estatísticas do atributo 'date' . . . . .	188
Figura I.8: Ocupação dos parques - Estatísticas Simples do atributo 'timestamp' . . . . .	189
Figura I.9: Ocupações dos parques - Frequências do atributo 'timestamp' . . . . .	189
Figura I.10: Ocupações dos parques - Estatísticas do atributo 'timestamp' . . . . .	189
Figura I.11: Ocupação dos parques - Estatísticas Simples do atributo 'dayOfWeek' . . . . .	190
Figura I.12: Ocupações dos parques - Frequências do atributo 'dayOfWeek' . . . . .	190
Figura I.13: Ocupações dos parques - Estatísticas do atributo 'dayOfWeek' . . . . .	190
Figura I.14: Ocupação dos parques - Estatísticas Simples do atributo 'occupation' . . . . .	191
Figura I.15: Ocupações dos parques - Frequências do atributo 'occupation' . . . . .	191
Figura I.16: Ocupações dos parques - Estatísticas do atributo 'occupation' . . . . .	191

## Índice de Tabelas

Tabela 1: Síntese "Inquérito à Mobilidade nas Áreas Metropolitanas do Porto e de Lisboa"	10
Tabela 2: Definições de IA agrupadas em 4 categorias	19
Tabela 3: Exemplos de Aprendizagem Supervisionada	25
Tabela 4: Exemplos de Aprendizagem Não Supervisionada	26
Tabela 5: Algoritmos de <i>Machine Learning</i> mais populares	29
Tabela 6: Trabalhos Relacionados	33
Tabela 7: Comparação dos trabalhos relacionados com o trabalho a solução a ser desenvolvido	35
Tabela 8: Síntese da análise dos vários trabalhos relacionados	36
Tabela 9: Definição dos <i>Sprints</i>	47
Tabela 10: Materiais e Tecnologias	48
Tabela 11: Resultados da pesquisa por datasets	52
Tabela 12: Descrição do dataset dos parques de estacionamento	57
Tabela 13: Descrição do dataset dos locais	60
Tabela 14: Descrição dos dados meteorológicos	63
Tabela 15: Descrição do dataset dos feriados	66
Tabela 16: Descrição do dataset das ocupações dos parques	67
Tabela 17: Análise de qualidade do dataset dos parques de estacionamento	81
Tabela 18: Análise de qualidade do dataset dos locais	83
Tabela 19: Análise de qualidade do dataset da Meteorologia	85
Tabela 20: Análise de qualidade do dataset dos Feriados	86
Tabela 21: Análise de qualidade do dataset das Ocupações	87
Tabela 22: Seleção dos dados no dataset do dataset dos parques de estacionamento	88
Tabela 23: Seleção dos dados do dataset dos locais	90
Tabela 24: Seleção dos dados do dataset da Meteorologia	91
Tabela 25: Seleção dos dados do dataset dos Feriados	93
Tabela 26: Seleção dos dados do dataset das Ocupações	94
Tabela 27: Descrição da tabela de base de dados 'final_data'	102
Tabela 28: Descrição da tabela de base de dados 'final_data'	107
Tabela 29: Resultados obtidos com o algoritmo K-Nearest Neighbors	113
Tabela 30: Resultados obtidos com o algoritmo Naive Bayes	113
Tabela 31: Resultados obtidos com o algoritmo Decision Trees	114

Tabela 32: Resultados obtidos com o algoritmo Random Forest . . . . .	114
Tabela 33: Resultados obtidos com o algoritmo Gradient Boosting . . . . .	115
Tabela 34: Matriz de concretização de objetivos e resultados esperados . . . . .	131
Tabela 35: Análise do risco . . . . .	134



## Acrónimos

**AML** Área Metropolitana de Lisboa.

**AMP** Área Metropolitana do Porto.

**API** Application Programming Interface.

**BYOD** Bring Your Own Device.

**CEO** Chief Executive Officer.

**CRISP-DM** Cross Industry Standard Process for Data Mining.

**DL** Deep Learning.

**DSR** Design Science Researche.

**IA** Inteligência Artificial.

**INE** Instituto Nacional de Estatística.

**IoT** Internet of Things.

**JSON** JavaScript Object Notation.

**ML** Machine Learning.

**PWA** Progressive Web App.

**REST** Representational State Transfer.

**SI** Sistemas de Informação.

**TI** Tecnologias de Informação.

**TIC** Tecnologias de Informação e Comunicação.

**XML** Extensible Markup Language.

## Glossário

**Deep Learning** DL consiste numa técnica, associada a redes neuronais, na qual os computadores aprendem com a sua experiência estabelecendo uma hierarquia de conceitos, no qual cada conceito é definido relativamente à sua relação com conceitos mais simples, o que permite então que a máquina aprenda conceitos complicados através de conceitos menos complexos. Ao aprender pela própria experiência, o ser humano não tem que especificar formalmente todo o conhecimento necessário.

Adaptado de: <https://www.deeplearningbook.org/>.

**Inteligência Artificial** A inteligência artificial, também conhecida como inteligência de máquinas, é um ramo das ciências da computação que visa produzir software com a capacidade de analisar o ambiente usando regras predeterminadas e algoritmos de pesquisa ou modelos de Machine Learning, de Data Mining e tomar decisões com base nessas análises.

Adaptado de: <https://www.techopedia.com/definition/190/artificial-intelligence-ai>.

**Machine Learning** Machine Learning é uma ramo de inteligência artificial (IA) que visa o desenvolvimento tecnológico do conhecimento humano. O Machine Learning permite que os computadores lidem com novas situações através da análise, aprendizagem, observação e da experiência.

Adaptado de: <https://www.techopedia.com/definition/8181/machine-learning>.

**Smart City** Como qualquer sistema inteligente, uma cidade inteligente é uma cidade em que a recolha de dados através de sensores e as análises dos mesmos são usadas para automatizar e adaptar uma ampla gama de serviços com intuito de atingir um melhor desempenho, a redução custos e do impacto ambiental.

Adaptado de: <https://www.techopedia.com/definition/31494/smart-city>.

# 1 Introdução

Neste capítulo pretende-se apresentar o enquadramento e a motivação para a realização do projeto de dissertação, os objetivos e resultados esperados e também clarificar qual a estrutura do documento.

## 1.1 Enquadramento e Motivação

As cidades europeias enfrentam cada vez mais problemas causados pelo transporte e pelo trânsito, uma vez que a grande maioria dos cidadãos europeus vivem em ambiente urbano, onde partilham as mesmas infraestruturas de mobilidade quer se desloquem em transportes públicos ou em transporte próprio. A questão de como melhorar a mobilidade e, ao mesmo tempo, reduzir o congestionamento, os acidentes e a poluição é um desafio comum a todas as grandes cidades europeias (Costa, Neto, & Bertolde, 2017).

A preferência pela mobilidade através de transportes coletivos face aos veículos próprios é uma medida bem conhecida e benéfica a nível ambiental, financeiro, temporal e até mesmo da segurança (Rodríguez, Comtois, & Slack, 2017). Contudo, e segundo o estudo do Instituto Nacional de Estatística (INE) (Instituto Nacional de Estatística, 2018), muitas vezes os utilizadores não têm acesso a informações acerca dos transportes coletivos, que seriam fundamentais para a sua tomada de decisão, o que os leva a procurar alternativas e por comodidade ( $\approx 50\%$ ), o uso de transporte próprio é o meio de deslocação elegido por grande parte da população que se desloca aos/nos centros urbanos ( $\approx 65\%$ ), sendo o estacionamento dos veículos um dos fatores mais inconvenientes (apenas 7,8% dos residentes da Área Metropolitana do Porto (AMP) justificam a utilização dos automóveis por facilidade de estacionamento). Vários são os problemas que advêm da procura por lugares de estacionamento, como o congestionamento do fluxo de trânsito em ruas com lugares de estacionamento, a poluição causada pelo deslocamento na procura por lugares e, a nível da comodidade, o tempo gasto nesta tarefa (Ibrahim, 2018). O acesso a certas informações, no caso dos utilizadores de veículos próprios, poderá contribuir para a resolução de alguns destes problemas. A solução de muitos destes problemas poderá ser encontrada com investigações em contextos de *Smart Cities* (cidades inteligentes).

Segundo a Autoridade Nacional de Segurança Rodoviária (Júdice, 2014) “uma cidade inteligente é aquela em que é feito investimento em capital humano e social, incentivando a utilização de tecnologias avançadas, as denominadas “TIC” como elemento viabilizador para um crescimento económico susten-

tável, proporcionando uma melhoria na qualidade de vida dos cidadãos residentes e flutuantes, e consequentemente, aquelas que permitem também uma melhor gestão dos recursos naturais e energéticos”.

Podemos definir Inteligência Artificial, (IA), recorrendo à sua origem, como a capacidade das máquinas aprenderem e pensarem de forma semelhante aos humanos (T. M. Mitchell, 1997). De acordo com o estudo “Inteligência Artificial na Europa” de 2018 da consultora Ernst & Young, na última década verificou-se um notório aumento dos investimentos em Inteligência Artificial, tendo sido na sua maioria realizado por fundos de capital privado (Guevara, 2019).

O tema deste projeto de dissertação enquadra-se no ioCity<sup>1</sup>, um projeto da *Start Up* Tecnológica IOTech, que definiu a finalidade deste como o desenvolvimento de uma solução inovadora que combina o conceito de *Smart Cities*, IA, *Progressive Web Apps* (PWA) e *Bring Your Own Device* (BYOD), com a utilização de Sensores e exploração da Internet das Coisas e pretende dar resposta a uma grande problemática na área da mobilidade urbana: as intermináveis filas de trânsito que normalmente se devem, entre outros fatores, à dificuldade em encontrar um parque de estacionamento. Desta forma, O ioCity visa o desenvolvimento de uma PWA inteligente capaz de sugerir a melhor rota/meio de transporte para um determinado local/-parque tendo em conta um conjunto de fatores que podem influenciar a decisão. O sistema permite recolher e analisar dados em tempo-real: tráfego, transportes (localização, taxa de ocupação e afins), e estacionamentos (ocupação e localização).

O foco deste projeto de dissertação será na otimização de parques de estacionamento de forma inteligente, ou seja através da exploração de dados gerados pelos sensores e outros dados disponíveis, combinando com a aplicação de de um conjunto de algoritmos de *Machine Learning* (ML), definir um conjunto de informações úteis para o cidadão e com isso diminuir os desperdícios de tempo à procura de um lugar de estacionamento.

A IOTech<sup>2</sup> é uma empresa que atua na área das Tecnologias da Informação (TI) e apresenta uma visão muito focada que passa por criar um sociedade inclusiva e digital através da desburocratização de processos e do fácil acesso das novas tecnologias aos cidadãos e organizações. A missão da IOTech passa por simplificar os processos de negócio, aproximando as organizações dos seus colaboradores/ clientes / utentes. Para isso, a IOTech procura fornecer valor acrescentado aos seus clientes através do desenvolvimento de aplicações e serviços inteligentes, fiáveis, sustentáveis e *user-friendly*, alicerçadas

---

<sup>1</sup>Projeto submetido ao abrigo do Portugal2020, Aviso N.e 34/SI/2018

<sup>2</sup><https://iotech.pt>

numa forte componente criativa e de inovação, que satisfazem as expectativas dos seus utilizadores, e com isso, fazer a diferença num mercado cada vez mais digital e global.

A motivação para a realização desta dissertação deve-se ao interesse na área de IA e também devido a uma problemática crescente das grandes cidades que a maior parte dos cidadãos acaba por presenciar, a dificuldade na mobilidade urbana. Uma vez que o ioCity é uma solução desenhada a pensar na sociedade e com consequências diretas na vida quotidiana também contribui como fator motivacional.

## 1.2 Objetivos

Este projeto de dissertação pretende dar resposta à seguinte questão de investigação: “De que forma podemos melhorar a mobilidade urbana através da Inteligência Artificial?”

Para dar resposta a esta questão de investigação e assumindo que a dificuldade em encontrar um local para estacionar prejudica a mobilidade urbana, foram definidas duas linhas de Investigação e Desenvolvimento:

1. Monitorização ativa de estacionamento de veículos dentro das cidades;
2. Modelos inteligentes para sugestão da melhor rota para o estacionamento;

Neste projeto foram utilizados modelos matemáticos e IA para, através da análise de um conjunto de dados (taxa de ocupação, tráfego, estacionamento, entre outros), propor de uma forma rápida e simples a melhor rota para um determinado parque de estacionamento de acordo com as necessidades do utilizador (rapidez, tempo de espera, disponibilidade parque, custos, outros). Juntamente com a IOtech, foi estabelecido qual o objetivo principal bem como quais os objetivos secundários. O objetivo principal passava por:

- Perceber até que ponto é possível gerar a rota ideal para um cidadão tendo por base a disponibilidade e a taxa de ocupação/localização dos parques;
- Prever a taxa de ocupação dos parques de estacionamento;

Os objetivos principais dividem-se em objetivos secundários:

- Ter uma gestão mais efetiva dos parques através da otimização da taxa de ocupação;
- Redução de desperdícios de recursos;

Uma vez que todos os objetivos foram cumpridos, obtiveram-se os seguintes resultados:

- Conhecimento acerca da realidade atual sobre a temática;
- Conhecimento das características e tipologias dos parques de estacionamento;
- Exploração de técnicas de IA/ ML;
- Desenvolvimento um conjunto de dados capaz de simular o dia-a-dia dos parques;
- Elaboração de um Protótipo funcional;
- Geração de novo conhecimento que combine a área dos sistemas de informação, interoperabilidade, programação web e inteligência artificial;

Desta forma, para este projeto de dissertação, estabeleceu-se a definição de uma arquitetura de IA com ML capaz de suportar a solução desenvolvida, o desenvolvimento de um algoritmo inteligente que melhor se adegue à realidade e o concebimento de uma aplicação *mobile* que possa servir como prova de conceito.

Por forma possibilitar a realização dos objetivos, numa primeira fase foi realizada uma revisão de literatura que incidiu em conceitos como *Smart Cities*, Parques de estacionamento, IA, ML e *Deep Learning* (DL). Com esta, o objetivo passou por identificar um conjunto de ferramentas, metodologias e técnicas que viabilizassem a realização deste projeto.

### **1.3 Estrutura do Documento**

O documento de dissertação contempla seis capítulos. O primeiro capítulo, o qual se insere este tópico, consiste numa introdução ao projeto no qual é feita uma breve contextualização e explicação em que o projeto se insere bem como quais os seus objetivos, resultado esperados e a motivação para a realização.

No segundo capítulo são abordados todos os temas considerados a base deste projeto, ou seja, Mobilidade Urbana, Parques de Estacionamento, *Smart Cities*, IA, ML e DL. Este capítulo contém também o estado de arte, ou seja, trabalhos da área realizados até ao momento. Por fim, é apresentada uma crítica do estado de arte comparativamente à solução desenvolvida.

O terceiro capítulo contempla as três metodologias aplicadas no desenvolvimento do projeto que são a *Design Science Research* (DSR), aplicada à vertente de investigação, e Scrum utilizada para o desenvolvimento da parte prática, e o *Cross Industry Standard Process for Data Mining* (CRISP-DM) para a parte prática do projeto.

O quarto capítulo apresenta as tarefas de cariz mais prático do projeto, de acordo com as fases do CRISP-DM

O quinto capítulo debruça-se sobre as conclusões retiradas após o desenvolvimento deste projeto de dissertação.

Por último, no sexto capítulo, serão apresentadas todas as referências bibliográficas utilizadas no desenvolvimento do projeto, seguido de alguns anexos considerados relevantes para uma melhor clarificação.

## 2 Revisão de Literatura

Segundo Cardoso et al (Cardoso, Alarcão, & Celorico, 2010) “A actividade de rever a literatura existente implica um processo de transformação, uma passagem da informação ao conhecimento. A informação está nos livros, e artigos (...). O conhecimento está nas pessoas.”

A revisão de literatura pode ser definida como uma pesquisa e avaliação da literatura disponível sobre uma determinada área ou assunto que visa documentar e apresentar alguns conceitos considerados pertinentes para o projeto, resumindo toda a pesquisa realizada de forma a integrar todo o conhecimento retirado sobre o assunto do projeto. Mais ainda, a revisão de literatura possibilita que os estudos de outros sejam um ponto de partida para o desenvolvimento do projeto (Royal Literary Fund, 2019).

O desenvolvimento da revisão de literatura seguiu uma estratégia que consistiu em procurar em plataformas *online*, artefactos que pudessem suportar a mesma. Para a procura foram utilizados os seguintes serviços de indexação: Web of Science Google Scholar, Scopus, RepositóriUM, B-On, Academia.edu e Google. Nestes serviços deu-se prevalência a artigos, livros e dissertações, contudo foram utilizadas também algumas páginas web.

Para a escolha de conteúdo teve que se adotar uma estratégia de seleção. Para definições de conceitos, recorreu-se a investigadores, cientistas e autores referência na área, ou seja os nomes com mais credibilidade na respetiva área. Para o restante conteúdo foram considerados fatores como a relevância científica, a data de publicação e o conteúdo como factor de seleção. A relevância científica foi analisada considerando o número de citações dos artigos, livros ou dissertações. Para a data de publicação deu-se prevalência a conteúdo com data de publicação posterior a 2015. A análise do conteúdo em artigos, livros ou dissertações foi feita através da leitura do abstract. Para páginas web, a data de publicação e a entidade publicadora foram tidos em conta.

A revisão de literatura iniciou-se com a pesquisa sobre a temática onde o ioCity pretende atuar, a mobilidade urbana. Aqui foram procuradas referências que permitiram caracterizar a mobilidade urbana e perceber quais os desafios e problemáticas da área. De seguida, o objetivo foi reunir informações sobre estacionamento, tema desta dissertação, e a procura foi efectuada em repositórios de dissertações de engenharia civil e analisar as referências bibliográficas das mesmas. O passo seguinte foi a revisão de literatura sobre *smart cities*, e aqui a quantidade de resultados encontrada foi muito diversificada, e foram seleccionados artigos de entidades referência na área (Accenture, IBM etc) como ponto de partida.



Finalmente, foi revisto conteúdo acerca das áreas mais técnicas como Inteligência Artificial, *Machine Learning* e *Deep Learning*.

## 2.1 Mobilidade Urbana

A mobilidade é a capacidade de movimentação de um local para outro (Costa et al., 2017). A necessidade de deslocação em ambiente urbano resulta, geralmente, da necessidade de satisfazer uma finalidade como por exemplo emprego, lazer ou o acesso a bens ou serviços (Magagnin, 2008).

Segundo Rodrigue et al (Rodrigue et al., 2017) as áreas urbanas representam os cenários mais complexos de mobilidade de passageiros e mercadorias. Analisando os contextos atuais das cidades, podemos concluir que existem três tipos de transporte que merecem ser destacados:

- **Transporte Coletivo:** Tem o propósito de disponibilizar mobilidade acessível ao público em locais específicos da cidade. Exemplos: Autocarros, metros, comboios etc;
- **Transporte individual:** Inclui qualquer tipo de mobilidade que resulta de uma escolha pessoal e temos como exemplo automóvel, a pé, bicicleta e motociclos;
- **Transporte de carga/mercadorias:** Tem o propósito de transportar mercadorias entre dois pontos. Geralmente está associado a indústrias, centros de distribuição, armazéns ou atividades de retalho.

Os autores (Rodrigue et al., 2017) referem ainda que muitas vezes as deslocações de passageiros e de mercadorias são complementares, contudo também é frequente utilizarem a mesma infraestrutura terrestre e de transporte disponível, incluindo estacionamento, o que nas áreas urbanas está sujeito a sérias restrições de capacidade.

Citando ainda os mesmos autores (Rodrigue et al., 2017) as rotas utilizadas nas cidades tendem ser estáveis (viagens casa - trabalho, distribuição de mercadorias geralmente com rotas fixas). Vários são os fatores que influenciam a rota a realizar, sendo os mais importantes os custos de transporte, tempo e o nível de congestionamento.

### **2.1.1 Urbanização Global e a Mobilidade Urbana**

Segundo Rodrigue et al. (Rodrigue et al., 2017), a urbanização consiste na transição de uma sociedade rural para uma sociedade urbana e estatisticamente reflete-se num aumento da população que vivem em ambiente urbano resultante duma migração entre o meio rural e urbano. Segundo os autores (Rodrigue et al., 2017), urbanização tem sido uma das tendências dominantes, que se verifica desde o século XX e desde 1950 a população mundial urbana aumentou mais do dobro, tendo atingido aproximadamente 4.2 biliões em 2018, o que representa 55% do total de habitantes no planeta Terra e prevê-se que esta tendência seja continua até à segunda metade do século XXI. Em 2050, estima-se que 6.4 biliões de pessoas poderão residir em ambientes urbanos o que se traduzirá em 70% da população mundial.

A falta de planeamento adequado dos sistemas de transporte e o facto da taxa de crescimento da população residente em ambiente urbano ser superior às melhorias das infraestruturas de mobilidade tem levado a um aumento proporcional e muitas vezes exponencial, dos problemas de mobilidade face ao número de deslocações em ambiente urbano (Costa et al., 2017).

Desta forma, a procura por mobilidade urbana e, em particular, a crescente dependência do automóvel particular para viagens intra-metropolitanas, é atualmente um dos temas debatidos nas mais diversas conferências acerca do desenvolvimento urbano sustentável, uma vez que levam à crescente deterioração da mobilidade urbana e acarretam problemas ambientais (Camagni, Gibelli, & Rigamonti, 2002). Embora esta referência já esteja algo distante no tempo, o inquérito realizado nas áreas metropolitanas do Porto e Lisboa (INE, 2018) demonstra que a mesma ainda está atualizada.

### **2.1.2 Mobilidade Urbana nas Áreas Metropolitanas do Porto e Lisboa**

Do “Inquérito à Mobilidade nas Áreas Metropolitanas do Porto e de Lisboa”, realizado no 4º trimestre do ano de 2017 (INE, 2018), é possível obter dados acerca da mobilidade nestas duas áreas. Aqui foram inquiridos aproximadamente 100000 (cem mil) indivíduos, 40393 na área metropolitana do Porto (AMP) e 58751 na área metropolitana de Lisboa (AML), entre os 6 e 84 anos, residentes nas duas áreas metropolitanas. No que diz respeito à população móvel, ou seja, quem realizou pelo menos uma viagem à data do inquérito, a taxa face ao total de população foi de 78.9% na AMP e 80.4% na AML. Nos dias úteis este indicador ascendeu a 82.9% na AMP e 85.1% na AML.

Os resultados divulgados demonstram a centralidade de deslocações dos municípios de Porto e Lisboa nas respetivas áreas metropolitanas e que, em média, cada residente móvel realizou 2.72 deslocações por dia na AMP e 2.60 na AML. De destacar ainda que mais de 90% das movimentações foram intrametropolitanas, ou seja com início e fim em municípios da mesma área metropolitana, mais precisamente 94% e 97% na AMP e AML respetivamente. Das deslocações intrametropolitanas, 71,0% na AMP e 65,4% na AML foram deslocações com origem e destino no mesmo município.

No inquérito realizado (INE, 2018) foi ainda possível apurar os motivos das deslocações. Na AMP 30.3% das deslocações foram motivadas por trabalho, 33.2% por compras e assuntos particulares e 15.7% devido ao acompanhamento de familiares e aqui inclui-se o transporte de crianças de/para a escola. Por sua vez na AML o motivo de trabalho correspondeu a 30.8% das deslocações, os assuntos particulares e compras tiveram um peso de 31.7% e o acompanhamento de familiares ficou-se pelos 15.2%. É importante referir se apenas considerarmos as deslocações aos fins-de-semana estes valores diferiram, uma vez que emergiu um motivo com especial significância, o lazer, que representou 21,7% na AMP na e 22.5% na AML e as deslocações motivadas por compras foram mais expressivas que nos dias úteis (34,6% e 36,1% ao fim de semana face aos 12,3% e 14,0% nos dias úteis, pela mesma ordem).

Este inquérito comprovou ainda que o transporte individual motorizado, por exemplo automóvel individual ou motociclo, é o principal meio de transporte nas duas áreas metropolitanas, com mais expressão na do Porto (69,0%) do que na de Lisboa (59,8%). Nos dias úteis os números ficaram-se pelos 66,5% na AMP e 57,3% na AML. O automóvel foi a escolha em 67,6% do total dos casos na AMP e 58,9% na AML. Se apenas se considerar os dias úteis registaram-se 65,2% e 56,3%, pela mesma ordem. As denominadas deslocações suaves, (pedonal ou bicicleta) atingiram 18,9% na AMP e 23,5% na AML. Analisando apenas as deslocações por transporte individual motorizado (ligeiros de passageiros e motociclos) e por transporte coletivo/público (autocarros e transportes ferroviários), o transporte público/coletivo foi utilizado em 13,9% e 20,9% das deslocações na AMP e AML, respetivamente. Nos dias úteis ficou pelos 16,7% e 24,3% pela mesma sequência.

No que diz respeito a tempos e distâncias, as deslocações realizadas por residentes na AMP tiveram uma duração média de 21.8 minutos e uma distância de 10.1 km. Por sua vez na AML este números ficam-se pelos 24.3 minutos para uma distância média de 10.3 km. Registou-se ainda que, em média, os residentes na AMP demoram 23.6 minutos a chegar ao local de trabalho, percorrendo um total de 12.6 km, enquanto que na AML estes números ficaram-se pelos 28,9 minutos e 13,1 km.

No inquérito realizado, foi questionado qual o motivo para a seleção do método de transporte utilizado. A "rapidez" foi o principal motivo em 58.4% das deslocações na AMP e em 62,5% na AML e o "conforto" em 49,8% e 50,4%, pela mesma ordem. Outros motivos apresentados foram a "rede de transportes públicos sem ligação direta ao destino", "ausência de alternativa" e "serviços de transporte público sem a frequência ou fiabilidade necessárias". No que concerne aos motivos para a escolha dos transportes públicos 52.5% dos residentes na AMP e 45,3% dos residentes na AML apontaram o facto de "não conduzir/não ter transporte individual". A "ausência de alternativa" e o "preço/custo do transporte público" surgem imediatamente a seguir (49,5% e 37,9% na AMP e 43,1% e 35,7% na AML).

Em ambas as áreas metropolitanas, os inquiridos avaliaram positivamente os transportes públicos no que diz respeito à "proximidade à rede/paragens" e à "facilidade de transbordo". Na AMP foi ainda destacado positivamente a "segurança" enquanto que na AML foi a "rapidez". Em contrapartida, foram destacados, na AMP, como aspetos negativos o "preço/custo do transporte público" e o "acesso por pessoas portadoras de deficiência". Na AML este último também foi referido juntamente com a "lotação" nos serviços de transporte.

A Tabela 1 permite sintetizar a informação obtida com o inquérito por forma a tornar mais fácil a análise.

Tabela 1: Síntese "Inquérito à Mobilidade nas Áreas Metropolitanas do Porto e de Lisboa

<b>Indicador</b>	<b>AMP</b>	<b>AML</b>
Taxa de população móvel	78.9%	80.4%
Número de deslocações por dia	2.72	2.60
Duração média das deslocações	21.8 min	24.3 min
Distância média das deslocações	10.1 km	10.3 km
% de deslocações intrametropolitanas	94%	97%
% de deslocações motivadas por trabalho	30.3%	30.8%
% de deslocações realizadas com automóvel	67,6%	58,9%

### 2.1.3 Problemas da Mobilidade Urbana

"Os desafios de transporte mais importantes ocorrem quando os sistemas de transporte urbano, por várias razões, não conseguem satisfazer os requisitos de mobilidade urbana" (Rodrigue et al., 2017). Esta

foi a definição utilizada para a definição dos desafios da Mobilidade Urbana e o mesmo autor enumera também aqueles que são alguns dos desafios impostos:

1. **Dependência do automóvel:** O uso de automóveis está obviamente relacionado a uma variedade de vantagens, como mobilidade de acordo com a necessidade, conforto, velocidade e conveniência. Quando é dada a escolha e a oportunidade, a maioria das pessoas prefere usar um automóvel.
2. **Congestionamento do trânsito e dificuldades de estacionamento:** Com a difusão do automóvel, a procura por infraestruturas de transporte aumentou e o investimento realizado no melhoramento e criação de infraestruturas não acompanhou o crescimento da mobilidade. Como os veículos passam a maior parte do tempo estacionados (95% do tempo, em média), a difusão do automóvel acentuou a problemática na procura por lugares de estacionamento. Desta forma o congestionamento e o estacionamento acabam por estar relacionados, uma vez que os lugares de estacionamento reduzem a capacidade de deslocação pois ocupam demasiado espaço e a procura por lugares cria atrasos adicionais prejudicando a circulação.
3. **Impacto Ambiental e consumo energético:** A poluição, incluindo a sonora, gerado pela circulação é prejudicial à qualidade de vida e até mesmo à saúde das populações urbanas. Além disto, com o aumento da dependência de transportes motorizados crescem as necessidades de petróleo.
4. **Transportes públicos inadequados:** Muitos dos sistemas de transporte público não satisfazem as necessidades de mobilidade devido à sobrelotação que causa desconforto para quem usufrui, ou porque a abrangência não é suficiente e um torna-se entrave à utilização. Mais ainda, como a procura por transportes públicos está sujeita a períodos de altos e baixos, o período de lotação provoca desconforto e os períodos com reduzidos número de passageiros tornam os serviços financeiramente insustentáveis.
5. **Viagens mais longas:** A dificuldade em arranjar habitação nas zonas centrais das cidades, onde geralmente há mais emprego, traduz-se num aumento da distância realizada entre casa e trabalho.
6. **Dificuldades para o transporte não motorizado:** Durante muito tempo os piões e as bicicletas foram desconsideradas nos projetos de infraestruturas e instalações.
7. **Acidentes e Segurança:** O crescimento do tráfego nas áreas urbanas está associado a um número crescente de acidentes e fatalidades. Os acidentes contribuem também de forma significativa para o congestionamento do trânsito.

8. **Altos custos de manutenção de infraestrutura:** Quanto mais extensa a rede ferroviária e rodoviária, maior o custo de manutenção e encargos financeiros.
9. **Distribuição de mercadorias:** O crescimento das entregas ao domicílio tem contribuído para o aumento dos transportes de mercadorias, o que é problemático uma vez que os veículos de transporte de mercadorias compartilham infraestruturas com os restantes veículos.

Desta lista de problemas destaca-se um tema que dá origem a esta dissertação que são as dificuldades de estacionamento. Neste sentido, torna-se pretinente abordar a temática do estacionamento.

## 2.2 Estacionamento

Segundo Litman (Litman, 2016), “o Estacionamento é uma componente essencial do sistema de transportes. Os veículos têm que ser estacionados em cada destino. A oferta de estacionamento afeta a facilidade de chegar aos destinos e conseqüentemente afeta a mobilidade urbana em geral”.

Parar o veículo e deixá-lo desocupado é chamado de estacionamento (DriversEd, 2019). Segundo (Bates & Leibling, 2012), em média, um carro passa 80% do tempo estacionado na casa dos proprietários, e apenas 3 a 4% é tempo de deslocamento. Nos restantes 16% o carro está estacionado noutra local. Embora estes sejam dados de 2012, ainda são atuais, uma vez que segundo David Z. Morris (Morris, 2016) os carros passam 95% do tempo estacionados. A melhoria do tráfego urbano depende muito das condições de estacionamento o que leva obrigatoriamente a cuidados redobrados no que a este aspecto diz respeito, e conseqüentemente dá especial ênfase à gestão do estacionamento (Ison, 2014).

Litman (Litman, 2016) define gestão do estacionamento como um conjunto de políticas e programas que resultam no uso mais eficiente dos recursos alocados aos estacionamento. O mesmo refere ainda que, quando bem aplicada, a gestão do estacionamento pode reduzir de forma significativa as necessidades de estacionamento em situações específicas, o que se traduz em benefícios económicos, sociais e ambientais.

De acordo com Instituto Superior Técnico (IST, 2009) os locais de estacionamento podem ser classificados quanto aos **utentes** e quanto ao **local**. Quanto aos utentes, os estacionamentos podem ser classificados como:

- **Estacionamento público:** é feito sem qualquer discriminação de utilização salvo de tipo de veículo e destina-se ao público em geral, gratuitamente ou mediante o pagamento de taxas.
- **Estacionamento privado:** só pode ser utilizado pelos respectivos proprietários ou seus autorizados, em propriedade particular, área coberta ou em logradouro.
- **Estacionamento de admissão reservada:** a utilização só é permitida a determinados condutores e é normalmente gratuita. É o caso de alguns estabelecimentos comerciais, casas de espectáculo e clubes que proporcionam estacionamento aos seus frequentadores.

Quanto ao local podem ser classificados de:

- **Estacionamento na via pública:** São em geral públicos, à parte os casos em que regulamentação especial estabeleça o contrário.
- **Estacionamento fora da via pública:** Podem dividir-se em dois grupos: os **parques de estacionamento** que têm um só piso, normalmente ao ar livre, embora possam ter cobertura aligeirada e os **auto-silos** que compreendem vários pisos, quer seja em elevação ou subterrâneos.

O Dicionário de Cambridge (Cambridge University Press, 2019) define parque de estacionamento como sendo uma estrutura/construção onde se pode deixar o carro por um período de tempo. Segundo (Brooke, Ison, & Quddus, 2014) quando um condutor chega ao destino escolhe um lugar de estacionamento que preencha os seus requisitos, entre as várias opções possíveis. A escolha depende do tráfego, do tempo que o condutor tem, do preço a pagar para estacionar, das características das infraestruturas de estacionamento e das características individuais do condutor.

Segundo Alsafery et al (Alsafery, Alturki, Reiff-Marganec, & Jambi, 2018), muitos dos problemas relativos ao estacionamento podem ser atenuados através da difusão dos estacionamentos inteligentes, que se enquadra na temática das *Smart Cities*. Desta forma, na próxima secção é abordada esta temática.

## 2.3 Smart Cities

Nesta secção é apresentado o conceito de *Smart Cities*, seguindo-se a apresentação do âmbito e finalizando com a apresentação de alguns exemplos.

### 2.3.1 Apresentação do conceito

Com a acentuação da urbanização, as cidades ganham maior destaque e mais controlo económico e político sobre o seu desenvolvimento. Mais ainda, estas estão cada vez mais dotadas tecnologicamente e a interconexão de todos os seus sistemas é uma tendência, o que permite atingir novos níveis de inteligência. Em contrapartida estas enfrentam também desafios e ameaças à sustentabilidade nos sistemas empresariais, nas infraestruturas como o transporte, a água, a energia e a comunicação. Para aproveitar as oportunidades e criar valor de forma sustentável, as cidades precisam de se tornar “mais inteligentes” (Dirks & Keeling, 2009).

A inteligência de uma cidade pode ser medida segundo a sua capacidade para reunir todos os seus recursos, de maneira eficaz e contínua por forma a atingir os objetivos a que se propôs (International Standards Organization (ISO), 2015). Encontrar uma definição de cidade inteligente é algo difícil, uma vez que existem várias e que geram significado ambíguo. Da maioria das definições, há 5 elementos que emergem pelo facto de serem recorrentes (Cunha, Przeybilovicz, Macaya, & Burgos, 2016):

- Visão Holística;
- Um meio para atingir os objetivos;
- A tecnologia como fator disruptivo;
- Um novo modelo de relações e interações;
- Sustentabilidade;

Não havendo uma definição amplamente aceite, é uma boa abordagem recorrer aos mecanismos de padronização internacionais ou organismos reconhecidos internacionalmente. A União Internacional de Telecomunicações considera que uma *Smart City* é uma cidade inovadora que utiliza as Tecnologias da Informação e Comunicação (TIC) e outros meios para melhorar a qualidade de vida, a competitividade e a eficiência das operações e dos serviços urbanos, garantindo ao mesmo tempo que atende às necessidades das gerações presentes e futuras no que diz respeito aos aspetos económicos e socioambientais (International Telecommunications Union (ITU), 2014). A Accenture (Accenture, 2014) define que, atualmente, uma *Smart City* é aquela que utiliza a tecnologia para prestar de forma mais eficiente os serviços urbanos, melhorar a qualidade de vida dos cidadãos e transformar a relação entre entidades locais, empresas e cidadãos, facilitando o modo de viver. Um elemento diferenciador é que a tecnologia é, desta vez,



disruptiva e altera o padrão de relações entre os agentes citadinos: cidadãos, governo, setor produtivo, terceiro setor.

### 2.3.2 Âmbito

A União Europeia propôs um modelo que decompõe o termo *Smart City* em 6 diferentes âmbitos: *Smart governance*, *Smart economy*, *Smart mobility*, *Smart environment*, *Smart people* e *Smart living* (Marville et al., 2014). O Centro de Inovação do Setor Público da PwC e o IE Business School, de Madrid, em Espanha, propôs uma alteração no modelo, desdobrando *Smart living* em dois âmbitos: “Segurança” e “Sanidade e saúde” (Cunha et al., 2016).

A Figura 1 representa graficamente os conceitos anteriormente referidos, e na mesma é possível perceber quais são os âmbitos das *Smart Cities*.

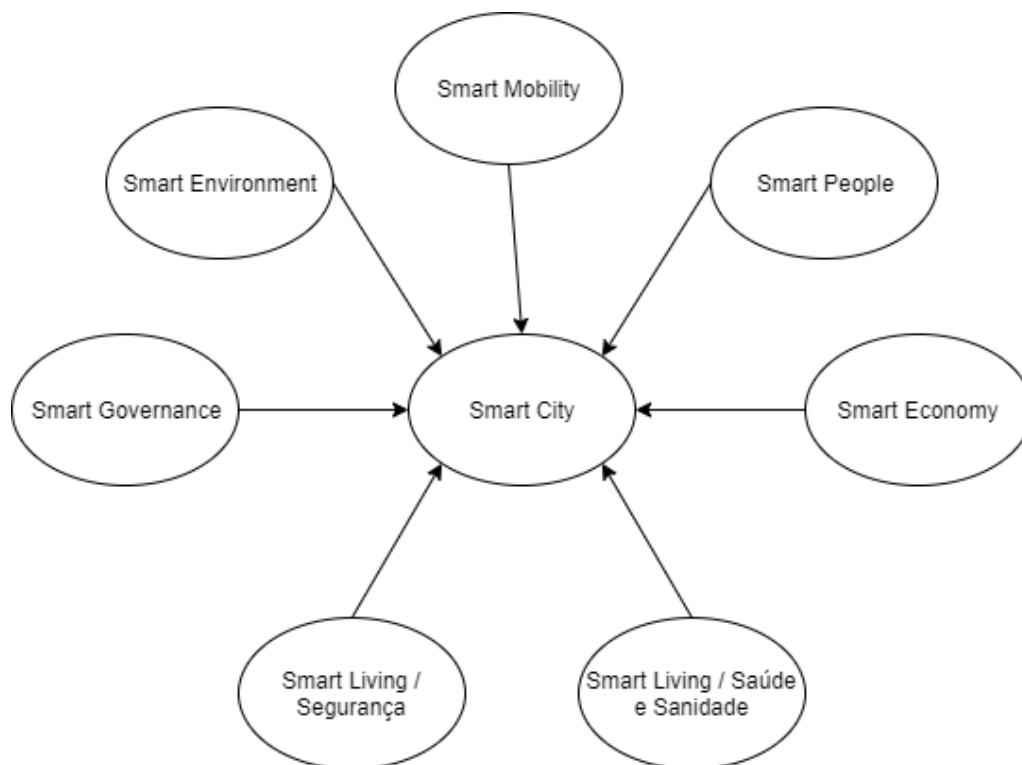


Figura 1: Âmbito das *Smart Cities*, adaptado de (Cunha et al., 2016)

De acordo com (Cunha et al., 2016), estes âmbitos podem ser vistos da seguinte forma:

- **Smart governance:** Por governança inteligentes entende-se a governança dentro e fora da cidade, incluindo serviços e interações que integram organizações públicas, privadas, civis e a comunidade europeia para que a cidade possa funcionar de maneira eficaz como um organismo.

- **Smart economy:** As TIC são utilizadas por forma a aumentar a produtividade, criar novos serviços, produtos e modelos de negócio.
- **Smart mobility:** O foco é a mobilidade e a tecnologia é utilizada por forma a obter um sistema logístico e de transporte integrado e que permita aos fornecer as informações necessárias atempadamente para que os cidadãos mudar facilmente de tipo de transporte.
- **Smart environment:** O setor de atuação é o meio ambiente e o objetivo é conseguir a sustentabilidade ambiental. Como exemplos temos o investimento em energias renováveis, sistemas de medição inteligente de consumos de energia e água, monitorização e controlo da poluição e reutilização e reciclagem de recursos;
- **Smart people:** O foco é a educação e as cidades surgem como impulsionadoras da educação dos seus cidadãos. Como exemplos temos ofertas de cursos de música, dança, línguas estrangeiras, informática etc;
- **Smart living / Segurança:** Neste âmbito as TIC devem contribuir para a segurança dos cidadãos. Exemplos disto temos sistemas de videovigilância, sistemas de localização por satélite, segurança nos transportes;
- **Smart living / Saúde e Sanidade:** O foco é a saúde e a tecnologia é colocada em prol desta área, permitindo uma melhor gestão das emergências, da assistência a idosos, e outros grupos vulneráveis.

### 2.3.3 Exemplos

No estudo "IESE Cities in Motion Index 2019" da IESE Business School (Berrone, Ricart, Duch, & Carrasco, 2019) foi feita uma análise a várias cidades em todo o mundo, considerando um total de 96 indicadores, por forma a obter um classificação das cidades consideradas mais inteligentes. Segue-se então uma lista das 5 cidades mais bem classificadas neste estudo:

1. **Londres:** A cidade de Londres é a cidade que apresenta mais atividade de programadores e *startups* em todo o mundo. Destaque para a sua plataforma de dados abertos "*London Datastore*"<sup>3</sup> que é utilizada por mais de 50.000 pessoas, empresas, desenvolvedores e investigadores, mensalmente. A cidade tem inovado na mobilidade e atualmente tem implementado os denominados

---

<sup>3</sup><https://data.london.gov.uk/>

Ultra (*Urban Light Transit*) que atuam no aeroporto de Heathrow e está a investir também nas melhorias das suas redes ferroviárias.

2. **Nova Iorque:** A cidade conta com mais de 7000 empresas de renome tecnológico e destaca-se pelos serviços integrados de tecnologia como o serviço Wi-Fi gratuito LinkNYC.
3. **Amesterdão:** A cidade provou ser uma importante potência europeia para a FinTech (empresas que trabalham para otimizar o sistema financeiro), eficiência energética e cultura. Aproximadamente, 90% das famílias da cidade usam bicicletas como meio transporte e a cidade apresenta serviços automatizados que permitem a partilha das bicicletas entre os cidadãos.
4. **Paris:** a cidade tem incentivado ao uso de bicicletas e veículos elétricos e é reconhecida pela inovação. Os dados relativos à cidade são públicos e têm investido fortemente numa das maiores redes de transporte da Europa (projeto Grand Paris Express) que permitirá adicionar mais 4 linhas de metro, 200 km de novas linhas ferroviárias e 68 novas estações e todas elas conectadas a um sistema de metro automático.
5. **Reykjavik:** A cidade destaca-se pelas questões ambientais. Mais de 90% da eletricidade e cerca de 80% do total de energia são de origem hidroelétrica e geotérmica o que demonstra a sua preocupação com a dependência de combustíveis fósseis.

Embora esteja fora do top 5, a **Singapura** é a cidade mais bem classificada na dimensão da tecnologia. A cidade fornece a Internet com melhor desempenho aos seus cidadãos e quase 100% da população usa telemóveis. Possui uma cultura tecnológica muito forte, com hospitais que contêm pessoal humano e robôs, táxis autónomos, espaços verdes verticais e jardins com temperatura controlado.

Segundo Pramathi e Vindhya (Pramathi & Vindhya, 2018) as *Smart Cities* do futuro deverão ser suportadas pela Inteligência Artificial (IA) através da incorporação desta tecnologia no dia a dia destas pessoas. De seguida é apresentada a revisão de literatura de IA.

## 2.4 Inteligência Artificial

Esta secção visa apresentar o conceito de IA começando pela sua contextualização e definições, seguindo-se uma breve abordagem ao teste de Turing e terminando com uma visão geral sobre a base, os modelos, as técnicas gerais e de aplicação de IA.

## 2.4.1 Contextualização e Definições

A IA já é uma realidade com a qual vivemos há muito tempo nos mais diversos contextos, contudo muitas vezes sem termos noção disso. Atualmente este é um tema ao qual se dá grande destaque uma vez que é reconhecida sua potencialidade e o quanto pode ajudar na melhoria da qualidade de vida da população (Machado, 2018). Segundo Winston, a IA pode ajudar-nos a resolver problemas difíceis do mundo real, criando novas oportunidades nos negócios, na engenharia e em outras áreas de aplicação. O mesmo autor refere que o objetivo de engenharia da IA é resolver problemas do mundo real, enquanto que o objetivo científico é explicar os vários tipos de inteligência. (Winston, 1992).

O termo IA foi proposto por Jonh McCarthy, em 1956 na conferência de Dartmouth, tendo sido inspirado no teste de Turing, referindo-se a um ramo das ciências da computação que procura que os computadores se comportem como humanos adquirindo as suas capacidades perçetivas e cognitivas. A definição de IA permanece ainda um pouco vaga consequência das diferentes visões e abordagens de quem se dedica a esta área (Li & Jiang, 2017). Na sequência disto, Stuart Russel e Peter Norvig (Russel & Norvig Peter, 2010) reuniram algumas das definições dos vários estudiosos desta área e dividiram-nas em 4 categorias, na qual cada categoria caracteriza a forma como a máquina se comporta:

- **Age Humanamente:** A máquina age como um humano. Na sua excelência, uma máquina que se comporte como um humano será capaz de passar no teste de Turing (abordado no próximo sub capítulo);
- **Pensa Humanamente:** É uma teoria associada à ciência cognitiva na qual se procura que a máquina pense como um humano, ou seja não interessa só a forma como a máquina se comporta, a forma como raciocina tem que ser semelhante à do ser humano. Para isto, é necessário perceber o funcionamento da mente humana e uma vez que exista uma teoria suficientemente precisa, torna-se possível expressá-la num programa de computador;
- **Pensa Racionalmente:** Nesta abordagem a IA relaciona-se com a lógica, uma vez que existe uma notação lógica criada no século XIX para a resolução de problemas e passando essas regras para uma máquina torna-se possível a resolução, isto se o problema for solucionável. Um exemplo de um problema de lógica é o seguinte "Todos os Homens são mortais. Sócrates é um Homem, logo Sócrates é mortal";
- **Age Racionalmente:** Espera-se que a máquina opere de forma autónoma, perceba o seu ambi-

ente, persista por um período prolongado, se adapte às mudanças e que crie e persiga objetivos. Ou seja, desta forma espera-se que a máquina alcance o melhor resultado ou, em situações de incerteza, o melhor resultado possível.

Na Tabela 2, a seguir apresentada, encontram-se as definições reunidas por Stuart Russel e Peter Norvig (Russel & Norvig Peter, 2010), na qual cada definição foi associada a uma das categorias acima apresentada.

Tabela 2: Definições de IA agrupadas em 4 categorias. Adaptado de (Russel & Norvig Peter, 2010).

<b>Categoria</b>	<b>Definições</b>
<b>Age Humanamente</b>	<p>“A arte de criar máquinas que executam funções que requerem inteligência quando executadas por pessoas” (Kurzweil, 1990)</p> <p>“O estudo de como fazer os computadores executarem coisas, que atualmente, o humano executa melhor” (Rich &amp; Knight, 1991)</p>
<b>Pensa Humanamente</b>	<p>“O novo e empolgante esforço para fazer os computadores pensarem... máquinas com mentes, no sentido pleno e literal.” (Haugeland, 1985)</p> <p>“A automação de atividades que são associadas ao pensamento humano, como tomada de decisão, resolução de problemas, aprendizagem...” (Bellman, 1978)</p>
<b>Pensa Racionalmente</b>	<p>“O estudo das faculdades mentais através do uso de modelos computacionais” (Charniak &amp; McDermott, 1985)</p> <p>“O estudo da computação que permite às máquinas perceber, raciocinar e agir” (Winston, 1992)</p>
<b>Age Humanamente</b>	<p>“Inteligência computacional é o estudo do design de agentes inteligentes.” (Poole, Mackworth, &amp; Goebel, 1998)</p> <p>“IA... está preocupado com o comportamento inteligente em artefatos” (Nilsson, 1998)</p>

Mais recentemente, Miguel Guevara, Coordenador de desenvolvimento do Centro de Computação Gráfica<sup>4</sup> referiu que hoje em dia existe uma definição mais simples de IA, “sistemas que realizam ações que, se efetuadas por humanos, seriam consideradas inteligentes.” (Guevara, 2019)

Existem 4 ingredientes básicos para se fazer IA: dados, recursos computacionais, algoritmos e talento para conseguir agrupar tudo. O facto de existir um volume de dados substancialmente maior do que os restantes fatores demonstra que muito há ainda a fazer na IA. (Media, 2017) Contudo, com a evolução de conceitos como *Big Data*, *Internet of Things*, a dimensão e disponibilidade dos sistemas tem contribuído para o desenvolvimento da IA (Machado, 2018).

## 2.4.2 Teste de Turing

Sendo a primeira definição de IA baseada no teste de Turing, torna-se obrigatório abordar este tema. O teste surgiu num dos artigos mais antigos relacionados com inteligência de máquinas “*Computer machinery and Intelligence*” e foi escrito por Alan Turing em 1950. O teste mede o desempenho de uma máquina em relação ao do ser humano, e ainda hoje se mantém atual pois permite padronizar a forma como podemos identificar se uma máquina é inteligente. Turing chamou ao teste “Jogo da Imitação”, e para que o teste seja possível tem que existir uma pessoa responsável por colocar questões, o interrogador, uma máquina e uma pessoa que irão responder às questões colocadas. O interrogador é colocado num local em que não lhe permite ver nem falar diretamente com nenhum dos outros, e pode apenas comunicar com eles utilizando um dispositivo que permita inserir texto. À medida que as questões vão sendo respondidas, o interrogador é desafiado a distinguir a máquina do humano. Caso o interrogador não consiga fazer a distinção, então, segundo Turing, a máquina é inteligente (Luger & Stubblefield, 1998).

---

<sup>4</sup><http://www.ccg.pt/>

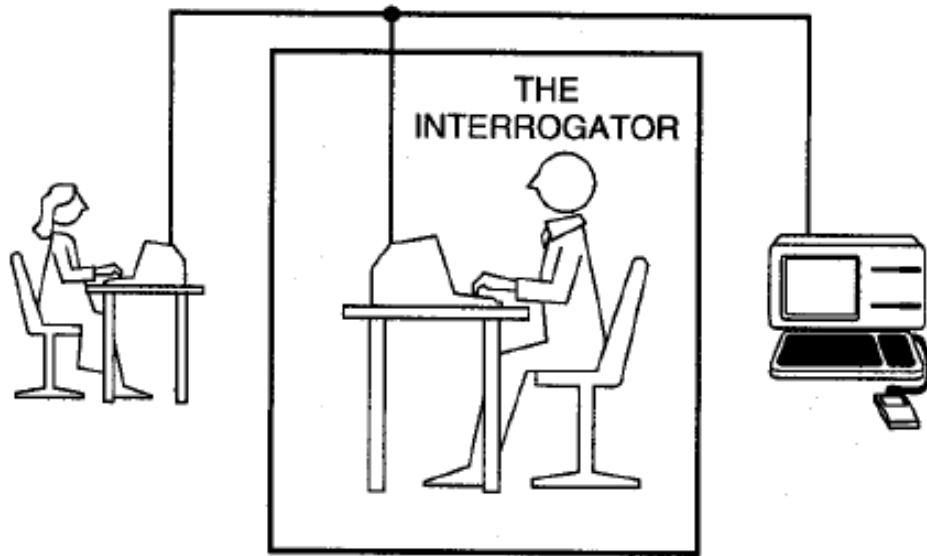


Figura 2: O teste de Turing, retirado de (Luger & Stubblefield, 1998)

Para passar no teste a máquina deve ter as seguintes competências (Russel & Norvig Peter, 2010):

- Processamento Natural de Linguagem para permitir que esta comunique fluentemente;
- Representação de conhecimento para armazenar o que a mesma sabe ou escuta.
- Raciocínio automatizado para utilizar a informação armazenada por forma a responder às questões e tirar novas conclusões;
- *Machine Learning* para adaptar-se a novas circunstâncias e extrapolar padrões;
- Visão computacional para reconhecer objetos;
- Robótica para manipular objetos e mover-se sobre os mesmos;

### 2.4.3 A base, os modelos, as técnicas gerais e de aplicação de IA

Em 2017, Li e Jiang (Li & Jiang, 2017) efectuaram uma revisão das técnicas de IA e as suas aplicações, como resultado obtiveram uma estrutura de 4 níveis, no qual cada nível suporta o que lhe sucede, e permite mais facilmente perceber a abrangência da IA.

No primeiro nível da *framework* desenvolvida está a base da IA. Ou seja, são as áreas ou ciências que constituem os fundamentos teóricos e no qual a IA foi inspirada e entre as mais populares temos a Estatística, a Física, a Teoria das Probabilidades, a Teoria dos Jogos, a Biologia, a Teoria dos Grafos, as Ciências Neurais, as Ciências Cognitivas, a Psicologia e a Sociologia.

No segundo nível encontram-se os modelos e os algoritmos, que são as técnicas fundamentais de IA e geralmente são baseados nas descobertas das áreas do primeiro nível. São vários os algoritmos e modelos até então desenvolvidos, contudo merecem particular destaque os seguintes: *support vector machine*, o algoritmo de otimização da colónia de formigas, o algoritmo imune, o algoritmo *Fuzzy*, árvore de decisão, algoritmo genético, algoritmo de enxame de partículas, redes neurais e *deep learning*.

No nível que se segue, ou seja, no terceiro, estão as técnicas gerais, ou seja são técnicas que não tem um propósito específico e aqui temos: extração de características, *clustering*, reconhecimento de padrões, *machine learning*, controlo inteligente, representação de conhecimento, mineração de conhecimento, entre outros.

No quarto e último nível temos as técnicas de aplicação de IA, ou seja onde a IA pode ser utilizada para satisfazer uma necessidade específica. Aqui consideram-se, entre outras, as seguintes técnicas: Reconhecimento de voz, Visão Computacional, Percepção do ambiente, Identificação Biométrica, Processamento de Linguagem Natural, Sistema especialista, previsão de tendências, deteção de anomalias, interação homem-máquina, múltiplos agentes.

Na figura 3, apresentada de seguida, é possível ver esquematicamente a estrutura de 4 níveis anteriormente referida.

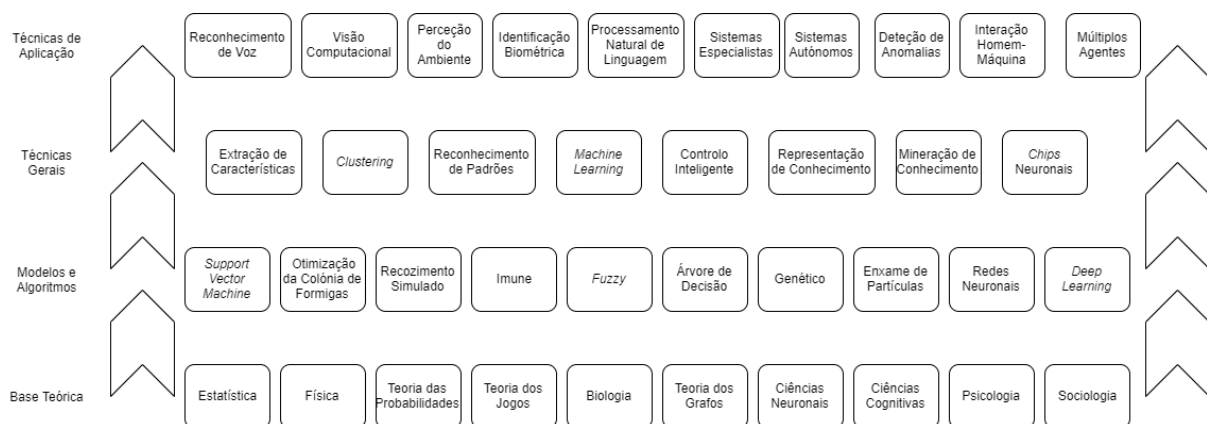


Figura 3: Inteligência Artificial - Estrutura de 4 níveis, adaptado de (Li & Jiang, 2017)



Há ainda duas técnicas de aplicação que merecem ser enunciadas, e que não estão mencionadas na estrutura anteriormente apresentada, que são os jogos e a robótica (Pannu & Student, 2015).

## 2.5 Machine Learning

Esta secção visa apresentar a área tecnológica que será o foco deste projeto, uma vez que é através desta que se pretende a otimização inteligente dos parques de estacionamento. Neste sentido, primeiramente é feita uma contextualização acerca da área e uma apresentação de algumas definições, seguindo-se os vários tipos de aprendizagem e modelos, terminando com a apresentação de alguns modelos e algoritmos de *Machine Learning* (ML).

### 2.5.1 Contextualização e Definições

Neste projeto de dissertação o ML assume particular destaque. Ainda há muita confusão entre a população em geral acerca do que é ML, uma vez que este termo é muitas vezes confundido com IA ou então são vistos como sinónimos. Na verdade o ML pode ser visto como um ramo da IA (Iriondo, 2018) no qual é possível criar modelos de dados que permitam entender esses mesmos dados e estando os modelos ajustados e aptos podem ser usados para prever e entender aspetos acerca de novos dados (VanderPlas, 2017). Tom M. Mitchell, uma referência e um dos pioneiros nesta área diz ainda que “*Machine Learning* é o estudo de algoritmos computacionais que permite que os programas de computador melhorem automaticamente através da experiência” (T. M. Mitchell, 1997).

Segundo Shai Shalev-Shwartz et al (Shalev-Shwartz & Ben-David, 2014), o ML é necessário e muito útil nas seguintes situações:

- **Tarefas executadas por Humanos/Animais:** Tarefas rotineiras difíceis de traduzir programaticamente. Exemplos deste tipo de tarefas é a condução, reconhecimento de voz, perceção de imagem etc;
- **Tarefas que vão além da capacidade Humana:** Tarefas que implicam a análise de grandes volumes de dados e dados de áreas complexas como dados da astrologia, previsões meteorológicas, transformação de arquivos médicos em conhecimento médico, análise do genoma humano. Ou seja, encontrar padrões e significado em situações que vão para além da capacidade Humana.

A Figura 4 contém um esquema adaptado de um livro que permite identificar no segundo nível quais os tipos de aprendizagem é possível adotar no ML e no terceiro nível os tipos de modelo mais comuns para os respetivos tipos de aprendizagem.

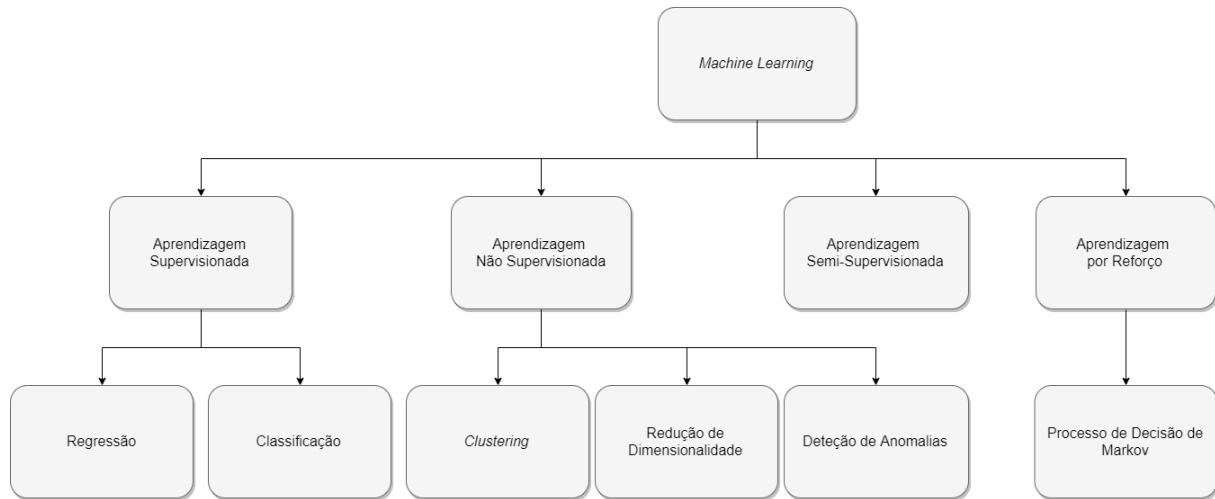


Figura 4: Tipos de Aprendizagem e Modelos do ML, adaptado de (Swamynathan, 2017)

## 2.5.2 Tipos de Aprendizagem

Como pode ser visto na Figura 4, atualmente existem 4 principais abordagens de ML ou tipos de aprendizagem, a aprendizagem supervisionada, aprendizagem semi-supervisionada, aprendizagem não supervisionada e aprendizagem por reforço. De seguida são apresentados, individualmente, os tipos de aprendizagem.

### a) Aprendizagem supervisionada

O objetivo da aprendizagem supervisionada é inferir uma função ou mapeamento de dados de treino. (Mohammed, Khan, & Bashier, 2016). A máquina é ensinada com recurso a exemplos. São fornecidos exemplos das entradas e saídas desejadas utilizando assim dados rotulados (com a resposta definida). A “máquina” (também conhecida como algoritmo) usa essa entrada para determinar correlações e lógica que podem ser usadas para prever a resposta. (Nevala, 2017). Ou seja, geralmente é quando se tenta prever uma variável dependente a partir de uma lista de variáveis independentes (Honda, Facure, & Yaohao, 2017).

A Tabela 3 contém alguns exemplos de problemas no qual se poderia utilizar aprendizagem supervisionada, em que na primeira coluna temos as variáveis independentes e na segunda a variável dependente, a

variável que se pretende prever.

Tabela 3: Exemplos de Aprendizagem Supervisionada, retirado de (Honda et al., 2017)

<b>Var. Independentes</b>	<b>Var. Dependentes</b>
Anos de Carreira, Formação, Idade	Salário
Idade do Carro, Idade do Motorista	Risco de Acidente Automóvel
Texto de um livro	Escola Literária
Temperatura	Receita de vendas de um gelado
Imagem da Rodovia	Ângulo da direção de um carro autónomo

A aprendizagem supervisionada contém dois grupos ou categorias de algoritmo, sendo eles a regressão e a classificação e segundo Honda et al (Honda et al., 2017) as técnicas mais comuns são regressão linear, regressão logística, redes neurais artificiais, *support vector machine*, árvores de decisão, *k-nearest neighbors*, Estatística Bayesiana, etc. A aplicabilidade deste tipo de aprendizagem é muito vasta, seguem-se alguns exemplos (Nevala, 2017):

- Determinação de risco: Determinar se um doente está em risco de ter algum problema de saúde; Determinar se um determinado equipamento está em risco de falhar tendo por base o histórico; Determinar a probabilidade de um indivíduo deixar de pagar uma prestação de empréstimo;
- Interação Personalizada: Determinar qual o conteúdo a mostrar a um indivíduo tendo por base o seu comportamento, intenções e gostos;
- Detecção de Fraudes: Identificação transações potencialmente fraudulentas;
- Reconhecimento de imagem, texto e voz: Identificação de objetos e pessoas em imagens; Sistemas que agem mediante indicações verbais;
- Segmentação de clientes: Segmentação de clientes com base no seu histórico, preferências e intenções;

## **b) Aprendizagem não supervisionada**

Na aprendizagem não supervisionada a figura do supervisor é eliminada, ou seja, não são fornecidas respostas nem dados de treino. (Mohammed et al., 2016). O objetivo passa por encontrar estruturas ocultas nos dados não rotulados analisando os mesmos e identificando padrões. Por conseguinte a máquina

determina correlações e relacionamentos analisando os dados disponíveis. Neste tipo de aprendizagem modela-se de acordo com a forma como os humanos vêem o mundo, definindo inferências e agrupando coisas de acordo com a observação e intuição (Nevala, 2017).

De uma forma geral, com aprendizagem não supervisionada o que se pretende é obter conhecimento mais representativo acerca dos dados que se tem. Um exemplo muito em voga nos tempos atuais são os sistemas de recomendação de filmes ou músicas.

A Tabela 4 contém mais alguns exemplos, em que na primeira coluna estão os dados que poderão ser vir de input para os modelos e na segunda coluna a informação que se pode extrair (Honda et al., 2017):

Tabela 4: Exemplos de Aprendizagem Não Supervisionada, retirado de (Honda et al., 2017)

<b>Dados</b>	<b>Forma Representativa</b>
Transações bancárias	Normalidade da transação
Registros de Compras	Associação entre produtos
Dados Multidimensionais	Dados com dimensão reduzida
Registos de Compras	Perfil dos consumidores
Palavras num texto	Representação matemática das palavras

Segundo Honda et al (Honda et al., 2017) as técnicas de aprendizagem não supervisionada são redes neuronais artificiais, *Clustering k-means*, *support vector machine*, *Clustering Hierárquico*, mapas auto-organizados etc. Segue-se uma lista de alguns exemplos práticas de aplicações deste tipo de aprendizagem (Nevala, 2017):

- Análise de carrinhos de compras: Análise de produtos que geralmente são comprados em conjunto;
- Detecção de anomalias ou intrusões: Detecção de comportamentos inesperados; Detecção de anomalias em sistemas ou anomalias em operações que possam indicar situações de fraude; Identificação de zonas com elevado risco de surto de doenças;
- Identificação de semelhanças: Identificação de novos grupos ou classificações de clientes tendo por base o seu comportamento, influências, rede de contactos etc;

### c) **Aprendizagem semi-supervisionada**

A aprendizagem semi-supervisionada é utilizada para resolver problemas semelhantes aos da aprendizagem supervisionada (Nevala, 2017). Neste tipo de aprendizagem são utilizados dados rotulados e não rotulados, ou seja para alguns dados são fornecidas as respostas e para outros não. Esta combinação permite gerar um modelo apropriado de classificação de dados. O objetivo da aprendizagem semi-supervisionada passa por otimizar a forma como se prevê classes de dados futuros comparativamente aos modelos em que apenas se usam dados rotulados (Mohammed et al., 2016).

### d) **Aprendizagem por reforço**

Segundo Nevala (Nevala, 2017) a aprendizagem por reforço assemelha-se a ato de ensinar alguém a jogar um jogo. As regras e objetivos estão definidos, no entanto o resultado depende do jogador que deve ajustar a sua abordagem ao ambiente, à sua habilidade e às ações de determinado oponente. Por outras palavras, é fornecida à máquina um conjunto de ações, regras e estados finais possíveis. Tendo tudo isto, a máquina aprende a explorar as regras para criar o resultado desejado, ou seja determinar uma série de ações que levarão a um resultado ótimo ou desejável.

Nevala (Nevala, 2017) refere ainda que as técnicas de aprendizagem por reforço mais comuns são as redes neuronais artificiais, autómatos de aprendizagem, processo de decisão de markov, *Q-Learning*. A mesma autora anteriormente referenciada enuncia uma série de aplicações práticas para este tipo de aprendizagem entre os quais:

- Jogos: Executar a melhor estratégia para ganhar um jogo;
- Robótica: Demonstrar como executar uma série de tarefas;
- Navegação: Identificar a melhor rota para ir de um ponto A para um ponto B tendo por base o trânsito, as condições climáticas e outras condições;

## 2.5.3 **Tipos de Modelos**

Como é possível ver na Figura 4, presente no secção 2.5.1, os tipos de modelos dependem do tipo de aprendizagem. Para a aprendizagem supervisionada existem dois tipos de modelos mais comuns que são os modelos de regressão e de classificação. Para aprendizagem não supervisionada as técnicas mais

comuns são o *clustering*, a redução de dimensionalidade, e a deteção de anomalias. Para a aprendizagem por reforço o mais comum é o processo de desição de Markov (Swamynathan, 2017).

## a) Regressão

Nos modelos deste tipo, o resultado esperado é uma variável contínua, ou seja um valor real (Brownlee, 2016). Exemplos de casos de uso deste tipo de modelos são a previsão das vendas de retalho, previsão do número de funcionários para cada turno, número de lugares de estacionamento necessários para uma loja, etc (Swamynathan, 2017).

## b) Classificação

Nos modelos deste tipo, o resultado esperado é uma variável discreta ou categórica (Brownlee, 2016). Para isto, o número de classes possíveis tem que ser igual ou superior a 2, e o algoritmo deve aprender as características dos dados de entrada de cada classe com base em dados históricos e prever a classe nos novos dados de entrada (futuros). Um exemplo esclarecedor é a identificação de um email se é *spam* ou não *spam* (Swamynathan, 2017).

## c) Clustering

Um problema de *clustering*, é quando se deseja definir algum tipo de agrupamento dos dados considerando alguma similaridade (Brownlee, 2016), ou seja o objetivo passa por dividir os dados de entrada e, grupos lógicos de itens relacionados. Alguns exemplo deste tipo de exemplos são artigos e notícias semelhantes, agrupamento de clientes tendo por base o seu perfil, etc (Swamynathan, 2017)

## d) Redução de Dimensionalidade

Neste tipo de técnica, o objetivo passa por simplificar o conjunto de dados de entrada, fazendo o mapeamento para um espaço dimensional inferior. Um dos exemplos é a análise de um conjunto de dados de grande dimensão requer esforço em termos computacionais, desta forma torna-se útil identificar as variáveis preponderantes ou seja aquelas que contribuem para a variabilidade dos dados (Swamynathan, 2017).

## e) Detecção de Anomalias

Neste tipo de técnica, o objetivo consiste em identificar itens que não estão em conformidade com um padrão ou comportamento esperado comparativamente a outros itens de um determinado conjunto de dados. A aplicabilidade deste tipo de técnicas é útil por exemplo para monitorizar a integridade de sistemas ou para detetar fraudes (Swamynathan, 2017).

### 2.5.4 Modelos e Algoritmos

Sendo o *Machine Learning*, segundo Mitchell (T. M. Mitchell, 1997), baseado no estudo de algoritmos torna-se mandatório fazer uma abordagem acerca dos mesmos. Um algoritmo é um método que requer sempre uma série de operações finitas (Filho, 2017). De entre os vários algoritmos de ML, existem alguns que merecem particular destaque por serem os mais populares. Neste sentido, a Tabela 5 apresenta alguns dos algoritmos que merecem destaque, indicando qual o algoritmo, o tipo de aprendizagem e o tipo de modelo (Labs, 2019):

Tabela 5: Algoritmos de *Machine Learning* mais populares

<b>Algoritmo</b>	<b>Tipo de Aprendizagem</b>	<b>Tipo de Modelo</b>
Naive Bayes	Supervisionada	Classificação
K-Means Clustering	Não Supervisionada	Clustering
Support Vector Machine	Supervisionada	Classificação
Linear Regression	Supervisionada	Regressão
Logistic Regression	Supervisionada	Classificação
Decision Tree	Supervisionada	Regressão / Classificação
Random Forest	Supervisionada	Regressão / Classificação
K-Nearest Neighbors	Supervisionada	Classificação
Principal Component Analysis	Não Supervisionada	Redução de dimensionalidade
Redes Neurais Artificiais	Supervisionada	Regressão / Classificação

## 2.6 Deep Learning

Desde 2010 que o termo *Deep Learning* (DL) tem estado em particular destaque na área de IA e muitas vezes atribui-se o significado de IA a DL. Na verdade, o DL é apenas uma técnica ou sub-área do ML que por sua vez é sub-área da IA. (M. Mitchell, 2019).

De uma forma sucinta, o DL consiste numa técnica, associada a redes neuronais, na qual os computadores aprendem com a sua experiência estabelecendo uma hierarquia de conceitos, no qual cada conceito é definido relativamente à sua relação com conceitos mais simples, o que permite então que a máquina aprenda conceitos complicados através de conceitos menos complexos. Ao aprender pela própria experiência, o ser humano não tem que especificar formalmente todo o conhecimento necessário. (Goodfellow, Bengio, & Courville, 2016)

Este método de treino é denominado "*Deep*" uma vez que a rede neuronal é constituída por mais do que uma camada de processamento e o número de camadas define a profundidade da rede. Isto faz com que as redes neuronais seja mais complexas do que as convencionais. É importante realçar ainda que, as pesquisas mais recentes mostram que as redes neuronais profundas mais bem sucedidas são aquelas cuja estrutura se assemelha à do cérebro humano (M. Mitchell, 2019).

Atualmente, o DL é a base de muitos sistemas avançados de ML e permitiu melhorar a forma com se lida com dados não estruturados como imagens, sons e vídeos contribuindo de forma significativa para avanços no reconhecimento de imagem, som e vídeo por parte das máquinas (Nevala, 2017).

## 2.7 Trabalhos Relacionados

Ao nível da monitorização ativa de estacionamento de veículos dentro das cidades é possível constatar a existência de várias soluções, quer comerciais, quer de investigação, relativas à monitorização de locais de parqueamento. Foram identificadas três estratégias de gestão de lugares de estacionamento nos centros urbanos: sistemas de sensorização; algoritmos de análise de tendências e previsão ou; por pórtricos ou parquímetros. Dos sistemas identificados, é possível concluir que a maioria das soluções existentes recorrem à terceira abordagem para efetuar a contagem de lugares ocupados. Desses sistemas destacam-



se os da **Via Verde**<sup>5</sup>, da **Telpark**<sup>6</sup>, da startup **ParkMe**<sup>7</sup> e o **ePark** da EMEL<sup>8</sup> que são, geralmente, pagos.

Em Portugal, por exemplo, existe a solução da Via Verde, onde o utente pode estacionar em determinados parques e lugares de estacionamento de rua usando o identificador RFID adquirido para circular nas autoestradas. No caso dos lugares de rua, o identificador é complementado pela aplicação, onde pode ser feita a gestão de pagamentos e prolongar o tempo de estacionamento. Na solução da Telpark, idêntica à solução da Via Verde, o identificador RFID é substituído pela matrícula do carro. No entanto, atualmente, ambos os serviços apenas se encontram disponíveis em certas localidades. Um sistema semelhante à aplicação para smartphone da Via Verde para gestão de estacionamento na via pública é o ePark, da EMEL, apenas disponível na zona de Lisboa. Além de apenas funcionarem em parques com parquímetro ou pórtico, as soluções não disponibilizam a informação sobre os lugares livres aos utentes.

Dos sistemas comerciais estudados, os da **eNWAVE**<sup>9</sup>, da **Smart Parking**<sup>10</sup> e o **ParkingCloud da All Traffic Solutions**<sup>11</sup> são os que mais se assemelham à solução desenvolvida. Estes recorrem a sistemas de sensorização baseados em câmaras, magnetómetros, lasers, ultrassons, sensores indutivos e leitores de placas instalados no solo, à superfície ou suspensos para a deteção de lugares de estacionamento livres. A informação da taxa de ocupação e lugares livres é apresentada numa aplicação, disponível para os utilizadores. No entanto, são tecnologias que carecem de funcionalidades como reserva de lugares, previsão de ocupação (com base em dados previamente registados), sugestão de melhores parques e rotas tendo em conta o endereço de destino e conexão com outras aplicações de suporte (como o Google Maps, por exemplo).

Em 2013 é apresentado por Xu et al (Xu & Xu, 2016) um trabalho que descreve uma solução sem necessidade de instalação de infraestruturas de apoio como sensores ou câmaras. A solução deteta, através da informação do GPS, acelerómetro e *Bluetooth* do utilizador onde e quando o veículo é estacionado e quando o lugar fica livre. Contudo, nesta solução, a informação recolhida é processada para estimar em tempo-real a disponibilidade de estacionamento numa certa rua, não estando desenvolvida nenhuma

---

<sup>5</sup><https://www.viaverde.pt/via-verde/onde-e-como-usar/parques>

<sup>6</sup><https://www.telpark.com/en/how-it-works/>

<sup>7</sup><https://www.parkme.com/pt/>

<sup>8</sup><https://www.emel.pt/pt/>

<sup>9</sup><https://www.nwave.io/>

<sup>10</sup><https://www.smartparking.com/smartpark-system>

<sup>11</sup><https://www.alltrafficsolutions.com/our-solutions/>

plataforma que disponibilize essa informação ao utente.

Na revisão de literatura existem alguns estudos na área dos parques inteligentes tendo como principal foco a previsão da taxa de ocupação, sendo um dos mais relevantes o Smart Santander (MITOS) (Vlahogianni, Kepaptsoglou, Tsetsos, & Karlaftis, 2016). Esta aplicação (MITOS) tem por base a localização de uma pessoa, o tráfego, entre outras variáveis, e utilizando Redes Neurais Artificiais tenta prever a probabilidade de um lugar estar disponível quando o condutor chegar ao parque. No entanto, a aplicação não tem a capacidade de sugerir várias possibilidades, ou seja, o condutor tem de procurar vários parques e ver qual o melhor pois não consegue colocar opções para afinar o modelo. Caso a escolha do utilizador não seja a melhor o condutor fica sem saber se haveria ou não melhor opção. Destaque ainda para o trabalho de Fedchenkov et. al (Fedchenkov et al., 2018) que criou um modelo de regressão que integra o GPS do carro e tendo em conta a disponibilidade e o valor a pagar prevê o número de lugares vagos e sugere o melhor parque para estacionar. Relativamente ao projeto anterior este já permite indicar qual o melhor parque, no entanto, não utilizada dados em tempo-real para treinar o modelo o que limita a aprendizagem automática e a fiabilidade da resposta. Este projeto de dissertação permite ter o melhor dos dois projetos juntamente com características próprias e diferenciadoras, ou seja, tem as principais funcionalidades do MITOS, a disponibilização de várias opções para que o cidadão possa decidir, e a possibilidade de integrar tudo numa só plataforma e de acordo com as necessidades do cidadão.

Com a existência de várias soluções que se enquadram na mesma temática deste projeto de dissertação, torna-se importante realizar uma comparação entre as mesmas e a solução a ser desenvolvida. O primeiro passo foi selecionar algumas características consideradas pertinentes em soluções que visam otimizar parques de estacionamento para que se possa analisar a existência das mesmas nos trabalhos relacionados e no produto a realizar neste projeto de dissertação. De seguida são apresentadas 3 tabelas que permitem comparar soluções. Na Tabela 6 estão listadas as várias soluções encontradas, na qual se pode ver quais os nomes e links para as soluções. A Tabela 7 contém a análise de uma série de características consideradas pertinentes das várias soluções e na última linha da tabela é feita a mesma análise para o projeto de dissertação. Por fim, na Tabela 8 é feita uma síntese da análise realizada na Tabela 7, excluindo da síntese o trabalho a desenvolver neste projeto de dissertação.

Tabela 6: Trabalhos Relacionados

#	Nome	Link
1	Via Verde	<a href="https://www.viaverde.pt/particulares/Via-Verde/Onde-e-como-usar/Parques">https://www.viaverde.pt/particulares/Via-Verde/Onde-e-como-usar/Parques</a>
2	Telpark	<a href="https://www.telpark.com/en/">https://www.telpark.com/en/</a>
3	Real Time	<a href="http://realtimelparking.ca/">http://realtimelparking.ca/</a>
4	Real-Time Parking Startup ParkMe Launches An Android App	<a href="https://techcrunch.com/2013/05/20/parkme-android/">https://techcrunch.com/2013/05/20/parkme-android/</a>
5	nwave	<a href="https://www.nwave.io/?gclid=EAIaIQobChMI5LkT0-04QIVUkTTCh3LwgbDEAAYASAAEgJQuFD_BwE">https://www.nwave.io/?gclid=EAIaIQobChMI5LkT0-04QIVUkTTCh3LwgbDEAAYASAAEgJQuFD_BwE</a>
6	SmartParking	<a href="https://www.smartparking.com/smartpark-system/smart-app">https://www.smartparking.com/smartpark-system/smart-app</a>
7	Real Time Parking	<a href="https://www.researchgate.net/publication/262275607_Real-Time_Street_Parking_Availability_Estimation">https://www.researchgate.net/publication/262275607_Real-Time_Street_Parking_Availability_Estimation</a>
8	ParkingCloud	<a href="https://www.alltrafficsolutions.com/solutions/parkingcloud/">https://www.alltrafficsolutions.com/solutions/parkingcloud/</a>
9	eparkomat	<a href="https://www.eparkomat.com/">https://www.eparkomat.com/</a>
10	Real-Time Smart Parking Systems Integration in Distributed ITS for Smart Cities	<a href="https://new.hindawi.com/journals/jat/2018/1485652/">https://new.hindawi.com/journals/jat/2018/1485652/</a>
11	iParking – a real-time parking space monitoring and guiding system	<a href="https://www.sciencedirect.com/science/article/abs/pii/S2214209616301930">https://www.sciencedirect.com/science/article/abs/pii/S2214209616301930</a>
12	A Real-Time Parking Prediction System for Smart Cities	<a href="https://www.tandfonline.com/doi/abs/10.1080/15472450.2015.1037955">https://www.tandfonline.com/doi/abs/10.1080/15472450.2015.1037955</a>

Continua na próxima página

#	Nome	Link
13	ePark	<a href="https://www.emel.pt/pt/onde-estacionar/via-publica/meios-de-pagamento/epark/">https://www.emel.pt/pt/onde-estacionar/via-publica/meios-de-pagamento/epark/</a>
14	A Reservation-based Smart Parking System	<a href="https://digitalcommons.unl.edu/computerscidiss/29/">https://digitalcommons.unl.edu/computerscidiss/29/</a>
15	SmartParking: A Secure and Intelligent Parking System	<a href="https://ieeexplore.ieee.org/abstract/document/5746481">https://ieeexplore.ieee.org/abstract/document/5746481</a>
16	Inrix ParkMe	<a href="https://play.google.com/store/apps/details?id=com.parkme.consumer&amp;hl=en-ca">https://play.google.com/store/apps/details?id=com.parkme.consumer&amp;hl=en-ca</a>
17	JustPark	<a href="https://play.google.com/store/apps/details?id=com.justpark.jp">https://play.google.com/store/apps/details?id=com.justpark.jp</a>
18	Parker, Find available parking	<a href="https://play.google.com/store/apps/details?id=com.streetline.parker&amp;hl=en-ca">https://play.google.com/store/apps/details?id=com.streetline.parker&amp;hl=en-ca</a>
19	ParkMe	<a href="https://www.parkme.com/pt/">https://www.parkme.com/pt/</a>

Tabela 7: Comparação dos trabalhos relacionados com o trabalho a solução a ser desenvolvido

Sistema	País	Tipo de Sensores	Tipo de Ambiente	Tipo de Solução	Tipo de Aplicação	Tipo de Software	Características									
							Parquímetros	Gestão do Parque	% Ocupação	Reserva de Parque	Ocupação em tempo real	Previsão da ocupação	Usa IA	Dados de ocupação disponível para os utilizadores	Sugestão de melhor parque para estacionar	Conectividade Externa / API
1	Portugal	Outros	Ambos	Comercial	Ambos	Proprietário	Sim	Não	Não	Não	Não	Não	Não	Não	Não	Não
2	Portugal e Espanha	GPS	Ambos	Comercial	Mobile	Proprietário	Sim	Não	Não	Não	Não	Não	Não	Não	Não	Não
3	Canadá	Outros	Ambos	Comercial	Ambos	Proprietário	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
4	15000Cidades	Outros	Ambos	Comercial	Ambos	Proprietário	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
5	Estados Unidos	Outros	Ambos	Comercial	Nenhuma	Proprietário	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
6	Vários países	GPS	Ambos	Comercial	Mobile	Proprietário	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
7	Estados Unidos	Outros	Ambos	Investigação	Nenhuma	Investigação	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
8	Estados Unidos	Outros	Ambos	Comercial	Nenhuma	Proprietário	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
9	República Checa	GPS	Ambos	Comercial	Mobile	Proprietário	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
10	Portugal	Outros	Ambos	Investigação	Nenhuma	Investigação	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
11	Taiwan	Outros	Ambos	Investigação	Nenhuma	Investigação	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
12	Espanha	Outros	Ambos	Investigação	Nenhuma	Investigação	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
13	Portugal	Outros	Ambos	Comercial	Mobile	Proprietário	Sim	Não	Não	Não	Não	Não	Não	Não	Não	Não
14	Estados Unidos	Outros	Ambos	Investigação	Nenhuma	Investigação	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
15	Estados Unidos	Outros	Ambos	Investigação	Nenhuma	Investigação	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
16	Estados Unidos	Outros	Ambos	Comercial	Mobile	Proprietário	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
17	Estados Unidos	GPS	Ambos	Comercial	Mobile	Proprietário	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
18	Estados Unidos	GPS	Ambos	Comercial	Mobile	Proprietário	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não
19	Vários países	GPS	Ambos	Comercial	Mobile	Proprietário	Sim	Não	Não	Não	Não	Não	Não	Não	Não	Não
#	Portugal	Vários	Ambos	Investigação e Comercial	PWA	Investigação e Proprietário	Não	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim

Tabela 8: Síntese da análise dos vários trabalhos relacionados

<b>Opção</b>	<b>Quantidade</b>	<b>Percentagem</b>
<b>Tipo de solução</b>		
Comercial	13	68.42%
Investigação	6	31.58%
<b>Tipo Aplicação</b>		
Ambos	3	15.79%
Mobile	8	42.11%
Web	0	0.00%
Nenhuma	8	42.11%
<b>Tipo de Ambiente</b>		
Indoor	0	0.00%
Outdoor	0	0.00%
Ambos	19	100%
<b>Tipo de Sensor</b>		
GPS	6	31.58%
<b>Características</b>		
Parquímetro	7	36.84%
Gestão do Parque	14	73.68%
% de Ocupação	9	47.37%
Reserva de Parque	7	36.84%
Ocupação em tempo real	8	42.11%
Previsão da ocupação	3	15.79%
Usa IA	2	10.53%
Dados de ocupação utilizadores	4	21.05%
Sugestão melhor parque	0	0.00%
Conetividade Externa / API	1	5.26%

Da análise da Tabela 8, podemos concluir que a solução que se pretende desenvolver contém algumas características que ainda não são comuns nas soluções semelhantes, uma vez que 9 das 10 características tem uma presença inferior a 50% nas 19 soluções analisadas. A solução que se pretende desenvolver

destaca-se pela utilização de IA, coisa que ainda é pouco recorrente nos sistemas semelhantes, e pela sugestão do melhor parque para estacionar, algo que ainda não é feito hoje em dia.

### 3 Abordagem Metodológica

Considerando a complexidade associada ao desenvolvimento de um projeto de dissertação, é recomendada a adoção de metodologias que forneçam um conjunto de princípios, técnicas e boas práticas para um âmbito específico. Para a concretização desta dissertação, foram seguidas 3 abordagens metodológicas, o *Design Science Research* (DSR) enquanto metodologia científica, o *Cross Industry Standard Process for Data Mining* (CRISP-DM) para a componente mais prática do projeto e o Scrum como metodologia para gerir o projeto de forma ágil. Nesta secção serão então abordadas as três metodologias selecionadas.

#### 3.1 Design Science Research (DSR)

Segundo Peffers et al. (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007), o DSR incorpora os princípios, as práticas e os procedimentos necessários para realizar projetos de investigação. Os autores (Peffers et al., 2007) dividem a metodologia em seis fases, em que nas duas primeiras fases é identificado o problema e são definidos os objetivos da solução. Após a conceção da solução é efetuada a respetiva demonstração e avaliação. Sendo a solução validada, os resultados são comunicados através de publicações. É importante realçar que apesar de ser um processo sequencial, não é obrigatório começar o processo pela primeira atividade.

A Figura 5 visa representar esquematicamente a metodologia DSR. Posteriormente serão abordadas as fases da metodologia.



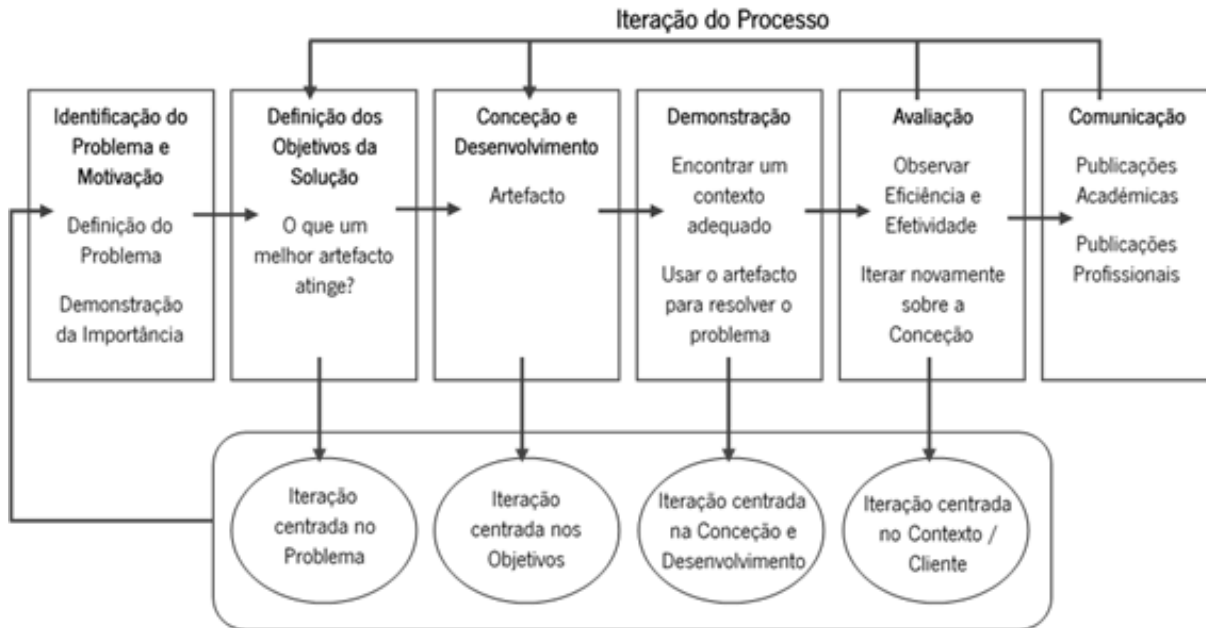


Figura 5: Metodologia Design Science Research, adaptado de (Peffer et al., 2007)

A adoção deste metodologia deve-se, primeiramente, à flexibilidade inerente ao seu processo de iteração, que permite que se regresse a fases anteriores, e depois porque o problema e os objetivos são definidos de uma forma clara. Segue-se uma explicação detalhada das várias fases desta metodologia segundo Peffer et al. (Peffer et al., 2007):

1. **Identificação do Problema e Motivação:** definição da questão de investigação e justificação do valor da solução. Como a definição do problema consiste em fornecer uma solução, é útil enquadrar o problema conceptualmente de forma a demonstrar a sua complexidade. A justificação do valor da solução permite facilitar a compreensão do problema, sobre o ponto de vista do autor. Assim, relativamente à presente dissertação, esta fase identifica a questão de investigação do projeto e respetiva motivação. A motivação, de acordo com o que está na [secção 1.1](#), para a realização desta dissertação deve-se ao interesse na área de Inteligência Artificial (IA) e contribuir para a melhoria da mobilidade urbana.
2. **Definição dos Objetivos da Solução:** identificação dos objetivos de uma solução, a partir da definição do problema e inferir o que é exequível. Os objetivos podem ser quantitativos ou qualitativos, sendo identificados através da especificação do problema. Desta forma, relativamente à presente dissertação, esta fase define os objetivos e resultados esperados do projeto, sendo necessário efetuar um enquadramento conceptual do tema. Os objetivos da solução, que estão presentes na [secção 1.2](#), são perceber até que ponto é possível gerar a rota ideal para um cidadão

tendo por base a disponibilidade e a taxa de ocupação/localização dos parques e prever a taxa de ocupação dos parques de estacionamento.

3. **Conceção e Desenvolvimento:** criação de um artefacto. O artefacto pode assumir a forma de uma construção, modelação, metodologia ou instanciação. Esta atividade inclui a definição do objetivo do artefacto, a sua arquitetura e por fim, a sua criação. Este ponto pressupõe também o estudo da literatura selecionada. Relativamente à dissertação, esta fase define a conceção e desenvolvimento do artefacto do protótipo de Otimização Inteligente de Parques de Estacionamento, presente na secção 4.5.2.
4. **Demonstração:** utilizar o artefacto para resolver uma ou mais instâncias do problema, através de simulações, casos de estudo, testes, entre outros. Assim, nesta dissertação, esta fase pressupõe o teste e validação do funcionamento total do sistema, desde a recolha dos dados à apresentação da informação ao utilizador final.
5. **Avaliação:** analisar o impacto do artefacto na solução para o problema. Esta atividade envolve a comparação dos objetivos de uma solução com os resultados observados no uso do artefacto na demonstração. Dependendo da natureza do local do problema e do artefacto, a avaliação pode assumir muitas formas. Assim, para esta dissertação, o protótipo será avaliado mediante o cumprimento dos objetivos e resultados esperados, como apresentado na secção 5.1.
6. **Comunicação:** envolve a comunicação do problema e a importância do artefacto, a sua utilidade e eficácia para investigadores e outros públicos-alvo relevantes. É considerado uma estrutura comum para trabalhos empíricos de pesquisa, onde a comunicação requer conhecimento da cultura disciplinar. Neste caso o protótipo é apresentado neste relatório de dissertação e na apresentação a da mesma.

## 3.2 CRISP-DM

Segundo Chapman et Al. (Chapman et al., 2000) o CRISP-DM é uma metodologia de *Data Mining* que fornece um conjunto de indicações, tarefas e técnicas para a realização de projetos nesta área. Na metodologia está pressuposta a existência de um ciclo de vida num projeto de *Data Mining* que contém fases, os relacionamentos entre as mesmas e as tarefas associadas a cada fase. Embora a metodologia forneça uma sequência de fases considerada mais adequada ou frequente, esta não é rígida, isto porque

o resultado de cada fase é que determina qual a próxima fase ou tarefa específica a ser realizada, sendo muito frequente voltar a fases anteriores para efetuar ajustes.

O CRISP-DM foi selecionado uma vez que fornece todas as práticas e tarefas para a realização dos artefactos do DSR e como a componente prática desta dissertação contempla a compreensão do negócio e dos dados e a geração de modelos, o CRISP-DM adequa-se.

De seguida, na Figura 6, está representado um esquema que ilustra o ciclo de vida de um projeto segundo o CRISP-DM e a sua versatilidade. Posteriormente serão detalhadas as fases e quais as tarefas associadas a cada fase. (Chapman et al., 2000).

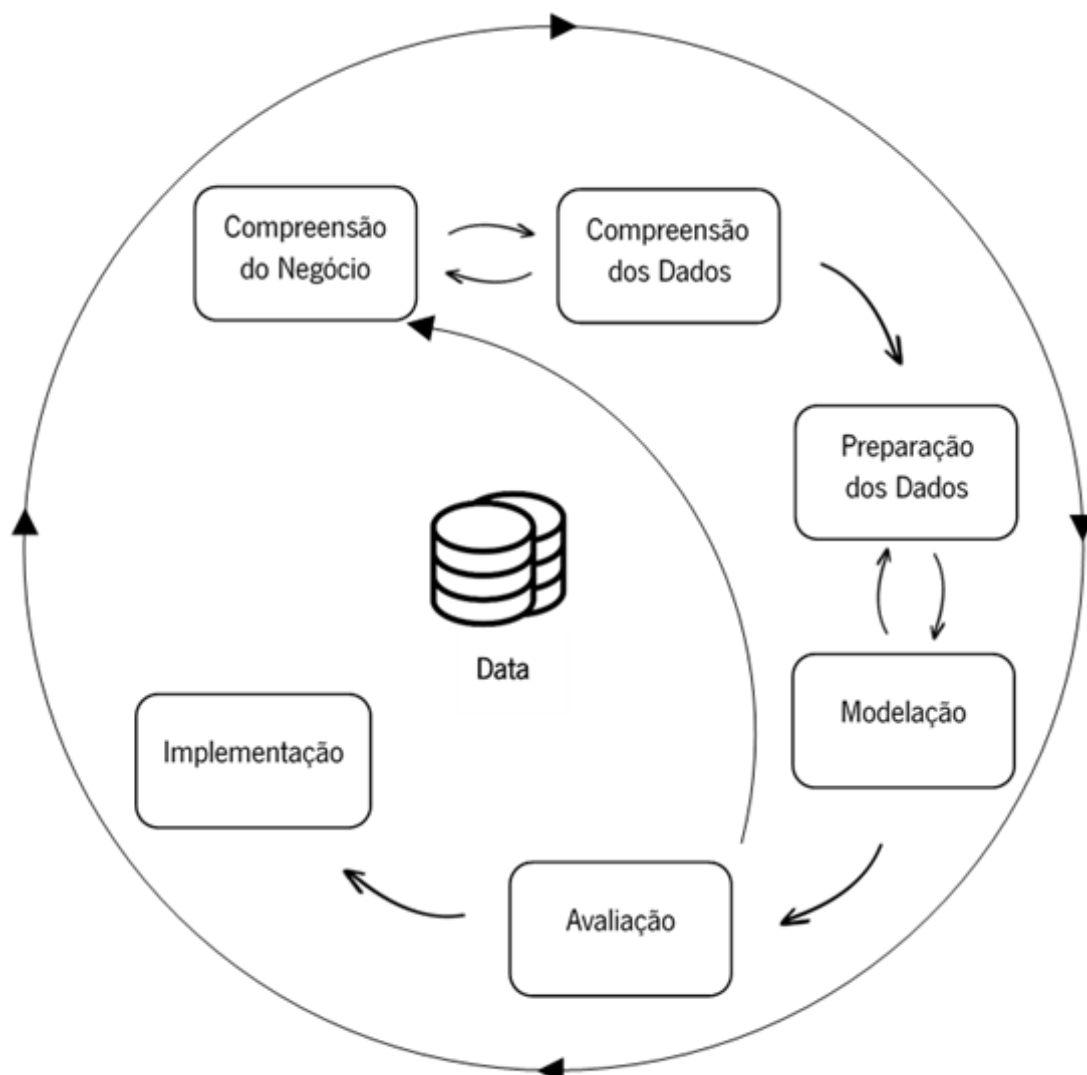


Figura 6: Metodologia CRISP-DM, adaptado de (Chapman et al., 2000)

- **Compreensão do Negócio:** Esta fase visa a compreensão dos objetivos e requisitos do projeto

numa perspetiva de negócio, convertendo estes num problema de *Data Mining* e num plano preliminar projetado para que se atinjam os objetivos. Nesta dissertação foram realizadas as seguintes tarefas:

- Determinação os objetivos de negócio, presente na secção secção 1.2
  - Avaliação da situação, que pode ser observada anteriormente na revisão de literatura efectuada, nomeadamente no levantamento do estado da mobilidade urbana (secção 2.1)
  - Determinação dos objetivos do projeto, que podem ser consultados também na secção 1.2
  - Produção de um plano do projeto
- **Compreensão dos Dados:** Esta fase inicia-se com a recolha e obtenção dos dados e algumas atividades ou procedimentos que permitam compreender os mesmos, ou seja, que permita que se fique mais familiarizado com os dados, identificando problemas de qualidade dos dados, ter as primeiras intuições e / ou detetar subconjuntos interessantes para formular hipóteses. Nesta dissertação, associada a esta fase foram realizadas as seguintes tarefas:
    - Recolha de dados inicial, na qual foram extraídos dados de diversas plataformas por forma a dar suporte aos modelos; (secção 4.1.1)
    - Descrição dos dados, em que os diversos atributos do dataset foram descritos e o dataset como um todo (secção 4.1.2)
    - Exploração dos dados, na qual se analisaram os diversos atributos dos datasets com recurso a ferramentas de visualização e análise estatística (secção 4.1.3)
    - Verificação a qualidade dos dados, no qual foram verificados quais os problemas detetados nos dados (secção 4.1.4)
  - **Preparação dos Dados:** A fase de preparação dos dados contempla todas as atividades necessárias para obter o *dataset* final a partir de dados crus ou não processados. Nesta fase deste projeto de dissertação foram realizadas as seguintes tarefas:
    - Seleção dos dados, na qual foram selecionados quais os dados a considerar na fase de modelação (secção 4.2.1)
    - Limpeza dos dados, no qual foram filtradas as informações relevantes do dataset e melhorada a qualidade dos dados (secção 4.2.2)

- Construção dos dados, presente na mesma secção da limpeza, no qual foram adicionados novos atributos consideradas relevantes (secção 4.2.2)
  - Integração dos dados, em que os dados foram integrados numa única tabela de base de dados (secção 4.2.3)
  - Formatação dos dados, que foi incluído na mesma secção do que a Integração, e nesta tarefa alguns dados foram formatados de acordo com uma base de dados criada (secção 4.2.3)
- **Modelação:** Nesta fase, várias técnicas de modelação (ou modelos) são selecionados, aplicados e otimizados para a obtenção de melhores resultados. Esta fase contempla algumas tarefas realizadas neste projeto de dissertação:
    - Seleção das técnicas de modelação, na qual foram selecionados quais os algoritmos de *Machine Learning* a utilizar (secção 4.3.1)
    - Geração de cenários de testes, tarefa esta na qual foram definidos cenários para a construção dos modelos (secção 4.3.2)
    - Construção dos modelos, na qual foram construídos os modelos de ML (secção 4.3.3)
    - Avaliação dos modelos, tarefa na qual foram avaliados os resultados dos vários modelos, para os diferentes cenários (secção 4.3.4)
  - **Avaliação:** Fase em que se avalia o modelo para garantir que o mesmo permite atingir os objetivos de negócio. No final desta fase deve ser tomada uma decisão acerca do utilização e aproveitamento dos resultados atingidos. Esta fase pressupõe a realização das seguintes tarefas:
    - Avaliação dos resultados, tarefa na qual foi avaliado o resultado global da fase de modelação (secção 4.6.1)
    - Revisão do processo, no qual é analisado o processo de modelação e fazendo a correspondência com a avaliação previamente realizada (secção 4.6.2)
    - Determinação dos próximos passos, em que foram definidas quais as tarefas a fazer futuramente por forma a tornar os modelos mais eficientes (secção 4.6.3)
  - **Implementação:** Nesta fase, o conhecimento adquirido tem que ser organizado e apresentado por forma a que o cliente o possa utilizar. Ou seja tornar útil o conhecimento adquirido para a tomada de decisão. Nesta fase foram realizadas as seguintes tarefas:

- Planeamento da implementação, secção 4.7.1
- Planeamento da monitorização e manutenção, secção 4.7.2
- Produção do relatório final, secção 4.7.3
- Revisão do projeto, secção 4.7.4

### 3.3 Scrum

Segundo os criadores (Schwaber & Sutherland, 2017), o Scrum é uma *framework* ágil que pode ser utilizada para resolver problemas complexos, entregando produtos com maior valor possível. O Scrum utiliza uma abordagem iterativa e incremental por forma a haver mais previsibilidade e controlo do risco e que assenta em três pilares de implementação a transparência, inspeção e adaptação. A transparência implica que os aspetos mais importantes do processo sejam visíveis por aqueles que são responsáveis pelo resultado. A inspeção requer que os utilizadores do Scrum inspecionem com regularidade os artefactos Scrum e o respetivos progressos aos objetivos. A adaptação implica que caso seja detetado algo que não corresponde ao pretendido, o processo ou o material em produção deva ser ajustado. De uma forma geral, o Scrum contém funções, eventos, artefactos e regras.

O Scrum foi selecionada por forma a gerir o desenvolvimento da parte prática da dissertação de forma ágil. Sendo esta componente bastante complexa, o Scrum permite de forma iterativa lidar com a complexidade.

**Equipa Scrum** Segundo o Guia de Scrum (Schwaber & Sutherland, 2017) as equipas Scrum são auto-organizadas e multifuncionais. Ao serem auto-organizadas a equipa escolhe a melhor abordagem para realizar o seu trabalho. Com multi-funcionalidade, a equipa é dotada de todas as competências necessárias para realizar o seu trabalho sem depender de outros. A equipa Scrum é constituída pelos seguintes elementos:

- **Product Owner:** é o dono do produto, o responsável por maximizar o valor do produto e do trabalho da Equipa de desenvolvimento, sendo então responsável por fazer toda a gestão do *Product Backlog*. O *Product Owner* deste projeto de dissertação é Filipe Portela, *Chief Executive Officer* (CEO) da IOTech.
- **Equipa de Desenvolvimento:** é formada por profissionais que trabalham para entregar um

incremento potencialmente utilizável no produto. Neste projeto de dissertação a equipa de desenvolvimento é composta por uma única pessoa, Carlos Fernandes.

- **Scrum Master:** é o responsável por promover e suportar o Scrum, ou seja ajuda toda a equipa Scrum a entender a teoria, as práticas, as regras e os valores do Scrum.

**Eventos Scrum** Os eventos Scrum estão estabelecidos por forma a criar rotinas e minimizar a necessidade de reuniões não definidas no guião e todos os eventos têm uma janela temporal bem definida, bem como os assuntos a serem tratados e quem deve estar presente. Para se falar dos eventos do Scrum, primeiramente tem que se introduzir o termo *Sprint* que é a base do Scrum. O *Sprint* é uma janela temporal limitada de um mês ou menos no qual se trabalha com vista ao atingimento de um objetivo ou seja, acaba por ser por si só um projeto. Isto permite que o Scrum seja um processo iterativo, reduzindo assim o risco no desenvolvimento. O Scrum contempla os seguintes eventos:

- **Planeamento do Sprint:** é uma reunião realizada no início de cada *Sprints*, que tem uma duração máxima de 8h para *Sprints* de um mês, na qual participa toda a equipa Scrum e aqui é planeado o que pode ser entregue no final da *Sprints*, ou seja, o que pode ser feito e como pode ser feito.
- **Reunião Diária:** é um evento diário, no qual apenas a equipa de desenvolvimento participa e está limitado a um máximo de 15 minutos. O objetivo é inspecionar o trabalho realizado desde a última reunião diária e prever o trabalho a realizar de seguida.
- **Revisão do Sprint:** é um evento que decorre no final de cada *Sprint*, no qual participa toda a equipa Scrum e tem duração máxima de 4h para *Sprints* de um mês. O objetivo é inspecionar o incremento (trabalho realizado) e adaptar o *Product Backlog* se necessário. Assim sendo é uma reunião de ponto de situação e destina-se a obter *feedback* e a fomentar a colaboração.
- **Retrospectiva do Sprint:** é um evento que ocorre após a revisão do *Sprint* e antes do planeamento do próximo. Aqui toda a equipa Scrum participa e tem duração máxima de 3h para *Sprints* de um mês. É uma oportunidade para que a equipa Scrum analise o que correu mal e crie um plano de melhorias para o futuro.

**Artefactos do Scrum** Os artefactos do Scrum representam o trabalho ou valor e visam fomentar a transparência. São especificamente desenhados de forma a que sejam entendidos de igual forma por todos. Os artefactos do Scrum são os seguintes:

- **Product Backlog:** O *Product Backlog* é uma lista ordenada de tudo o que é conhecido ser necessário no produto. É a única fonte de requisitos para quaisquer alterações a serem efetuadas nesse produto. O *Product Owner* é responsável pelo *Product Backlog* incluindo o seu conteúdo, disponibilidade e ordenação.
- **Sprint Backlog:** O *Sprint Backlog* é o conjunto de itens do *Product Backlog* selecionados para o *Sprint*, em conjunto com um plano para os entregar e assim concretizar o Objetivo do *Sprint*
- **Incremento:** O incremento é a soma de todos os itens pertencentes ao *Product Backlog* que foram concluídos durante o *Sprint*, juntamente com o valor de todos os incrementos concluídos em *Sprints* anteriores.

A Figura 7, pretende demonstrar esquematicamente o funcionamento da metodologia Scrum, contento as funções, eventos e artefactos.

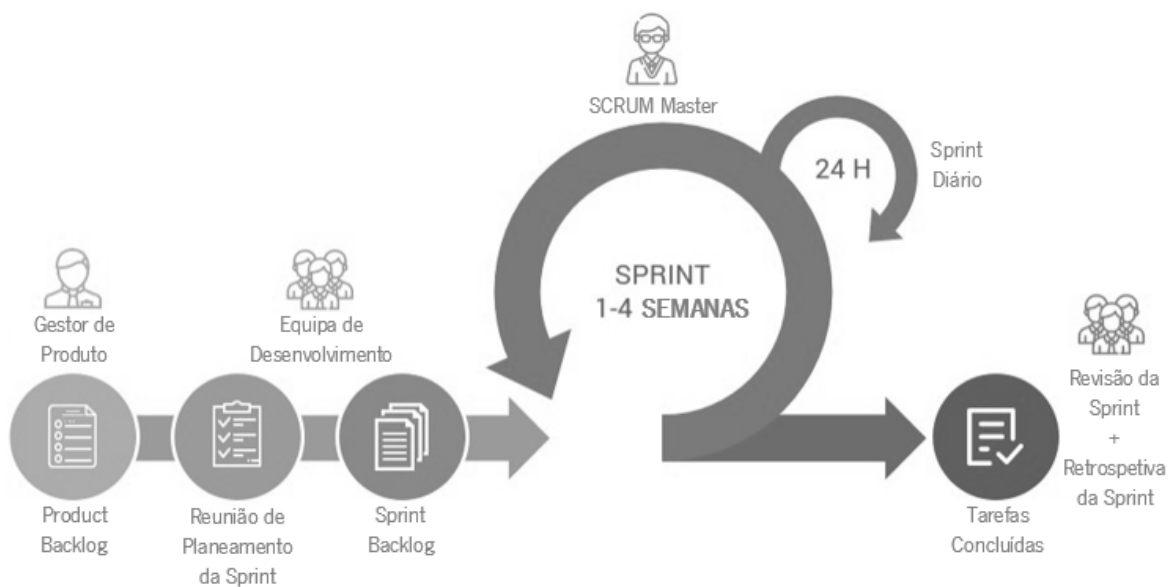


Figura 7: Metodologia Scrum

Na Tabela 9, apresentada de seguida, são apresentadas as tarefas realizadas nas *Sprint* do projeto, indicando para cada *Sprint* qual a data de início, a data de fim, e as tarefas realizadas.



Tabela 9: Definição dos *Sprints*

<b>Sprint</b>	<b>Início</b>	<b>Fim</b>	<b>Tarefas Realizadas</b>
1	03-02-2020	16-02-2020	Recolha inicial de dados
2	17-02-2020	01-03-2020	
3	02-03-2020	15-03-2020	
4	16-03-2020	29-03-2020	Descrição dos dados;
5	30-03-2020	12-04-2020	
6	13-04-2020	26-04-2020	Exploração dos dados;
7	27-04-2020	10-05-2020	
8	11-05-2020	24-05-2020	Qualidade dos dados;
9	25-04-2020	07-06-2020	Seleção dos dados;
10	08-06-2020	21-06-2020	Limpeza e construção dos dados e carregamento para base de dados de estágio;
11	22-06-2020	05-07-2020	
12	06-07-2020	19-07-2020	Integração e Formatação dos dados; Síntese da recolha, tratamento e carregamento dos dados;
13	20-07-2020	02-08-2020	Seleção das técnicas de modelação; Criação dos cenários de teste;
14	03-08-2020	16-08-2020	Construção do modelo;
15	17-08-2020	30-08-2020	
16	31-08-2020	13-09-2020	
17	14-09-2020	27-09-2020	
18	28-09-2020	11-10-2020	
19	12-10-2020	25-10-2020	Avaliação do modelo; Esquema síntese da modelação;
20	26-10-2020	08-11-2020	Avaliar resultados; Rever o processo; Determinar próximos passos;
21	09-11-2020	22-11-2020	API;
22	23-11-2020	06-12-2020	Aplicação mobile;
23	07-12-2020	20-12-2020	

Continua na próxima página

<b>Sprint</b>	<b>Início</b>	<b>Fim</b>	<b>Tarefas Realizadas</b>
24	21-12-2020	03-01-2021	Planear a implementação; Planear a monitorização e manutenção;
25	04-01-2021	17-01-2021	Produzir relatório final; Reflexão crítica; Análise de riscos; Trabalho Futuro;

### 3.4 Tecnologias e Plataformas

Por forma a desenvolver este projeto científico foi necessário escolher uma série de tecnologias. Na Tabela 10 são enumeradas as várias tecnologias, plataformas e softwares utilizados bem como o propósito que servem neste projeto.

Tabela 10: Materiais e Tecnologias

<b>Nome</b>	<b>Tipo</b>	<b>Propósito</b>
Draw.IO	Plataforma como serviço	Plataforma utilizada para desenhar muitos dos diagramas apresentados no documento
Github	Plataforma como serviço	Plataforma utilizada para gerir os repositórios Git
Google Cloud Platform	Plataforma como serviço	Cloud utilizada para alojar a máquina utilizada para o desenvolvimento das <i>scripts</i> e módulos de extração, tratamento e carregamento de dados e desenvolvimento dos modelos de <i>Machine Learning</i> (ML)
DarkSky API	Plataforma como serviço	Plataforma utilizada para extrair os dados meteorológicos necessários
Festivo API	Plataforma como serviço	Plataforma utilizada para extrair os dados dos feriados necessários
Continua na próxima página		

<b>Nome</b>	<b>Tipo</b>	<b>Propósito</b>
Google Places	Plataforma como serviço	Plataforma utilizada para extrair os locais mais próximos dos parques, extrair fotografias de locais e procurar locais
Google Directions	Plataforma como serviço	Plataforma utilizada para traçar rotas
Datastore Brussels API	Plataforma como serviço	Plataforma utilizada para extrair os dados dos parques de estacionamento de Bruxelas
Python	Tecnologia	Linguagem utilizada para o desenvolvimento dos módulos e <i>scripts</i> para a recolha, tratamento e carregamento de dados e para o desenvolvimento dos modelos de ML
IONIC	Tecnologia	Tecnologia utilizada para desenvolver a aplicação mobile
Angular	Tecnologia	<i>Framework</i> utilizada para desenvolver a aplicação mobile
MySQL	Tecnologia	Sistema de gestão de base de dados utilizado para armazenar todos os dados do projeto
Git	Tecnologia	Sistema de controlo de versões utilizado para gerir todo o código do projeto
PyCharm	Software	Software utilizado para desenvolver todas as <i>scripts</i> e módulos de Python
Visual Studio Code	Software	Editor utilizado para escrever o código da aplicação mobile
Android Studio	Software	Software utilizado para executar a aplicação no telemóvel
MySQL Workbench	Software	Ferramenta utilizada para a concepção da base de dados
Chrome Remote Desktop	Software	Software utilizado para aceder à máquina remotamente
Talend Open Studio for Data Quality	Software	Software utilizado para efectuar análises aos dados

## 4 Trabalho Realizado

Neste secção é apresentado todo o trabalho realizado neste projeto de dissertação. A divisão dos capítulos segue a estrutura da metodologia prática adotada, o *Cross Industry Standard Process for Data Mining* (CRISP-DM). Importa realçar ainda que a primeira fase do CRISP-DM, a compreensão do negócio, encontra-se na introdução do projeto, quando são definidos os objetivos do projeto (1.2) e na revisão de literatura na secção em que se aborda a temática da mobilidade urbana (2.1), sendo que neste capítulo apenas se aborda a componente prática do projeto.

### 4.1 Compreensão dos dados

Depois de feita a compreensão do negócio, conclui-se que para a realização deste projeto de dissertação serão necessárias diversos dados que possam suportar os modelos de Machine Learning (ML) que farão a previsão das ocupações dos parques de estacionamento. Os dados necessários que serão o foco desta fase são os seguintes:

1. Parques de estacionamento
  - Localização
  - Capacidade do parque
  - Preço
  - Acessibilidade
2. Histórico de ocupações dos parques
  - Data
  - Dia da semana
  - Hora
  - Taxa de ocupação
3. Locais nas proximidades dos parques
4. Meteorologia
  - Temperatura
  - Precipitação

5. Histórico de eventos ocorridos

6. Histórico de feriados e férias

### **4.1.1 Recolha de dados inicial**

O objetivo nesta etapa consiste em efetuar uma recolha de dados inicial que sirva de base para o projeto, começando naturalmente pela obtenção de um *dataset* de parques de estacionamento, com as respetivas características, seguido dos locais das proximidades de cada parque, o histórico de meteorologia, os eventos nas proximidades, o histórico de feriados, finalizando com o histórico de ocupação de cada parque de estacionamento

#### **a) Parques de estacionamento**

Inicialmente, esta tarefa passava por contactar algumas empresas responsáveis pela gestão de parques de estacionamento, tais como SABA, Braga Parque ou EMEL, com vista à realização de uma parceria para o projeto na qual as empresas disponibilizariam todos os dados necessários. Contudo, devido ao surto pandémico da COVID-19, muitas destas empresas suspenderam as suas ações e a possibilidade da realização da parceria ficou sem efeito, originando aqui um problema de dados.

Para colmatar a falha de dados foi necessário recorrer a uma solução de recurso. Desta forma iniciou-se um processo de procura de diversos datasets em plataformas de Open Data. O objetivo foi encontrar datasets com o maior número de informações sobre o parque (horário de funcionamento, preço, lugares reservados, localização etc) e também, para cada parque, o histórico de ocupações. A Tabela 11 apresenta 9 resultados que se destacam, indicando para cada qual a abrangência geográfica, o link para o dataset e a descrição.

Tabela 11: Resultados da pesquisa por datasets

<b>Abrangência</b>	<b>Link</b>	<b>Descrição</b>
Parque em Itália	<a href="http://cnrpark.it/">http://cnrpark.it/</a>	Dados da ocupação de um parque em Itália com recurso a reconhecimento de imagem. Apenas abrange um máximo de 23 dias.
Bristol	<a href="https://opendata.bristol.gov.uk/explore/dataset/car-parks/information/">https://opendata.bristol.gov.uk/explore/dataset/car-parks/information/</a>	Dados dos parques de estacionamento de bristol. Tem um limite de 10000 registos e destes apenas 5 tem informações sobre ocupação.
Maple Garage em Evanstone	<a href="https://data.cityofevanston.org/Parking/Maple-Garage-2015-2017-Occupancy-Rates/rddj-qr8r">https://data.cityofevanston.org/Parking/Maple-Garage-2015-2017-Occupancy-Rates/rddj-qr8r</a>	Contagem dos carros presentes no parque de estacionamento em questão, tendo dois registos por cada dia, num total de 2 anos.
Birmingham	<a href="https://data.birmingham.gov.uk/dataset/birmingham-parking">https://data.birmingham.gov.uk/dataset/birmingham-parking</a>	Ocupações em diferentes horas do dia de cada um dos parques de Birmingham operados pela entidade que gere o parque de estacionamento do aeroporto de Birmingham.
Vila de Luxemburgo	<a href="https://data.public.lu/en/reuses/car-parking-get-current-situation/">https://data.public.lu/en/reuses/car-parking-get-current-situation/</a>	Dados do estado atual de todos os parques de estacionamento integrados na Vila de Luxemburgo.
Los Angeles	<a href="https://data.lacity.org/A-Livable-and-Sustainable-City/LADOT-Parking-Meter-Occupancy/e7h6-4a3e">https://data.lacity.org/A-Livable-and-Sustainable-City/LADOT-Parking-Meter-Occupancy/e7h6-4a3e</a>	Dados acerca dos estacionamentos na via pública, obtidos através da análise dos tickets retirados dos parquímetros.
Continua na próxima página		

<b>Abrangência</b>	<b>Link</b>	<b>Descrição</b>
Melbourne	<a href="https://data.melbourne.vic.gov.au/Transport/On-street-Parking-Bay-Sensors/vh2v-4nfs">https://data.melbourne.vic.gov.au/Transport/On-street-Parking-Bay-Sensors/vh2v-4nfs</a>	Estado atual de cada lugar dos diversos parques de estacionamento.
Seattle	<a href="https://data.seattle.gov/Transportation/2019-Paid-Parking-Occupancy-Year-to-date-/qktt-2bsy">https://data.seattle.gov/Transportation/2019-Paid-Parking-Occupancy-Year-to-date-/qktt-2bsy</a>	Dados da ocupação dos parques de estacionamento pagos da via pública de Seattle.
Bruxelas	<a href="http://opendatastore.brussels/dataset/parking-occupancy">http://opendatastore.brussels/dataset/parking-occupancy</a>	Dados acerca dos parques de estacionamento de Bruxelas, contendo também alguns dados de ocupações.

Após a pesquisa e análise global de cada um dos datasets acima referidos, pode-se concluir que nenhum tem informação suficientemente completa para serem aplicados modelos de previsão. Muito dos datasets não contemplam dados relativos às características do parque e nenhum possui histórico de ocupação razoável para este projeto.

O dataset selecionado como base para este projeto será o de Bruxelas, uma vez que é o que contém mais informações das características dos parques, sendo assim o mais completo.

## **b) Locais**

Segundo Ionita et al. (Ionita, Pomp, Cochez, Meisen, & Decker, 2018) as informações dos locais nas proximidades do parque influenciam a taxa de ocupação. Os autores referem que, por exemplo, se um parque estiver situado perto de uma zona de restaurantes, a ocupação será tendenciosamente mais elevada nas horas de almoço e jantar.

Assim sendo, para cada parque de estacionamento de Bruxelas obtido anteriormente, foi feita recolha dos locais nas proximidades, utilizando o Google Places Application Programming Interface (API)<sup>12</sup>. Um

<sup>12</sup><https://developers.google.com/places/web-service/search>

dos métodos desta API permite, dada uma latitude e uma longitude, obter os locais nas proximidades ordenados por proeminência (*Nearby Search requests*).

Primeiramente, o dataset dos parques de estacionamento foi convertido de *Extensible Markup Language* (XML) para *JavaScript Object Notation* (JSON), e guardado num ficheiro denominado “original-parking-dataset.json”. Depois, com recurso a uma *script* Python, foram extraídos as latitudes e longitudes de cada parque e armazenadas num outro ficheiro JSON “parking-data-pre-places.json”. Esta extração foi feita, uma vez que a API da Google apenas necessita da latitude e longitude e assim a iteração torna-se mais célere. No Anexo A.1 é possível aceder ao código que permite a implementação do que foi dito anteriormente.

Posteriormente foi desenvolvido um módulo em Python que faz a iteração no ficheiro “parking-data-pre-places.json” e realiza um pedido GET à API Google Places para o URL `https://maps.googleapis.com/maps/api/place/nearbysearch/output?parameters` passando os parâmetros requeridos, que neste caso foi a latitude a longitude e a distância a considerar. Os resultados foram armazenados num ficheiro JSON denominado “places.json”. Este ficheiro contém então para cada parque de estacionamento os locais com proeminência na área. No Anexo A.2, é possível visualizar a implementação deste módulo.

### **c) Meteorologia**

As condições meteorológicas são também um fator que influencia a taxa de ocupação dos parques de estacionamento (Lijbers, 2016). Neste sentido, foi feita uma pesquisa por plataformas que permitissem obter o histórico para a cidade de Bruxelas. Foram analisadas algumas soluções e foi selecionada a DarkSky, uma vez que é a que melhor serve os efeitos de previsão, devido ao elevado volume de dados históricos (RapidAPI Team, 2020).

A DarkSky disponibiliza uma API e com recurso a uma *script* Python foi efectuada a extração do histórico das condições meteorológicas para Bruxelas desde 01-01-2015 até 31-12-2019, inclusive. Importante referir que a API possibilita que sejam fornecidas uma latitude, longitude e uma data e retorna as condições meteorológicas verificadas nas várias horas para a data indicada. Ou seja, obteve-se o histórico por hora, dum período de 5 anos, sendo que o resultado da extração foi armazenado num ficheiro denominado “meteo.json”.



É importante referir que foi assumido que as condições meteorológicas são homogêneas em Bruxelas, ou seja, que não variam consideravelmente de local para local. Desta forma, foram consideradas a latitude e longitude central de Bruxelas para evitar a recolha do histórico para as várias latitudes e longitudes dos parques, uma vez que isto acarretaria custos de performance e económicos consideráveis.

O código fonte de recolha dos dados meteorológicos encontra-se disponível no Anexo B.

## **d) Eventos**

A existência de eventos nas proximidades é uma variável que impacta a ocupação dos parques (Lijbers, 2016). Contudo não foram encontrados datasets públicos que permitam a utilização deste fator preditivo pelo que nesta fase não será considerado.

## **e) Feriados**

Segundo a literatura, os feriados são também um fator que influencia a taxa de ocupação dos parques de estacionamento (Lijbers, 2016). Neste sentido foi feita uma pesquisa com o objetivo de encontrar uma fonte de dados. Das várias possibilidades existentes, foi selecionada a plataforma Festivo<sup>13</sup>. Esta seleção justifica-se pelo facto da plataforma disponibilizar uma API REST e os dados fornecidos suprem as necessidades do projeto, no que a este fator diz respeito.

Para a recolha dos dados foi desenvolvido um módulo em Python que efetua a recolha do histórico de feriados para os anos 2015 até 2020, inclusive. O resultado foi armazenado num ficheiro JSON denominado “holidays.json”.

Importante referir que a API retorna os feriados da Bélgica, não fornecendo assim os feriados Regionais. Neste sentido foi adicionado para cada ano o Iris Day, que é um feriado de Bruxelas que ocorre todos os anos no dia 8 de Maio.

Em anexo foi colocado o código fonte que permitiu a recolha dos feriados (Anexo C).

---

<sup>13</sup><https://getfestivo.com/>

## f) Ocupação

O histórico de ocupações para cada parque de estacionamento ao longo do tempo é um dos dados mais importantes para efeitos de previsão (Lijbers, 2016). Contudo, durante a realização de diversas pesquisas não foram encontrados dados. Desta forma, para ultrapassar este problema optou-se por gerar ocupações.

Para gerar dados recorreu-se a uma *script* Python e utilizando uma distribuição normal, uma vez que é uma distribuição geralmente utilizada para modelar fenómenos naturais (Sousa, 2019).

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

A função da distribuição normal recebe dois parâmetros, a média e o desvio padrão elevado ao quadrado (variância), que foram definidos como 65 e 20, respetivamente.

$$X \sim \mathcal{N}(65, 20^2).$$

Para cada parque de estacionamento foram geradas ocupações a cada 60 minutos, desde 01-01-2015 até 01-01-2020. Devido à limitação de recursos de hardware, foram somente geradas ocupações para 10 parques, uma vez que para efectuar esta geração de dados para todos os parques, seriam necessários muitos recursos computacionais para conseguir fazer a integração de todos os dados à hora.

A *script* de Python que permitiu gerar as ocupações foi colocada no Anexo D.

### 4.1.2 Descrição dos dados

#### a) Parques de estacionamento

O formato inicial do dataset dos parques de estacionamento de Bruxelas era XML, contudo foi convertido para JSON para permitir uma melhor visualização e manipulação.

O dataset contempla dados de 63 parques de estacionamento, denominados de `parkingSite`, divididos por 18 grupos de parques, denominados de `parkingRecords`. Os `parkingRecords` são um conjunto de parques de estacionamento que têm em comum a mesma entidade responsável pela gestão.

O dataset contém alguns metadados que foram omitidos por não se revelarem úteis para este projeto. Segue-se uma apresentação da estrutura do dataset, com a explicação dos diversos atributos dos parkingRecord e ParkingSite.

Na Tabela 12 são apresentados os vários atributos do dataset em que para cada indica qual o objeto JSON, o tipo de dado e a descrição.

Tabela 12: Descrição do dataset dos parques de estacionamento

	<b>Objeto</b>	<b>Tipo</b>	<b>Descrição</b>
parkingRecord	@type **	Texto	Significado de parkingRecord
	@id **	Texto	Identificador único do parkingRecord
	@version **	Data	Data à qual os dados do parkingRecord foram recolhidos
	parkingName	Objeto	Objeto que contém o nome do parkingRecord
	parkingRecordVersionTime **	Data	Data à qual os dados do parkingRecord foram recolhidos
	operator *	Objeto	Objeto que contém as informações de quem operacionaliza o parque (Moradas, contactos, website, coordenadas etc)
	parkingLocation **	Objeto	Objeto que contém a localização do parkingRecord
Continua na próxima página			

<b>Objeto</b>		<b>Tipo</b>	<b>Descrição</b>
parkingSite	@type **	Texto	Significado de parkingSite
	@id	Texto	Identificador único do parque de estacionamento
	@version **	Data	Data da recolha das informações do parque
	parkingName	Objeto	Objeto que contém o nome do parque de estacionamento
	parkingRecordVersionTime **	Data	Data da recolha dos dados do parque de estacionamento
	parkingNumberOfSpaces	Inteiro	Número de lugares do parque de estacionamento
	responsibleAuthority *	Objeto	Objeto que contém as informações da pessoa responsável pelo parque
	operator *	Objeto	Objeto que contém as informações da entidade que operacionaliza o parque
	parkingLocation	Objeto	Objeto que contém as coordenadas do parque
Continua na próxima página			

<b>Objeto</b>		<b>Tipo</b>	<b>Descrição</b>
	onlyAssignedParking *	Objeto	Objeto que contém as restrições de acesso ao parque (tamanho dos veículos por exemplo)
	tariffsAndPayment *	Objeto	Objeto que contém as informações de pagamentos
	parkingEquipmentOrServiceFacility *	Array	Array que contém os serviços e extras disponibilizados
	groupOfParkingSpaces *	Objeto	Objeto que contém as informações dos lugares disponibilizados para deficientes
	parkingSiteAddress	Objeto	Objeto que contém a morada do parque
	openingTimes *	Objeto	Objeto que contém os horários para os vários dias da semana do parque
	parkingAccess *	Array	Array que contém as informações dos acessos ao parque. Entradas, saídas, acessos de cadeiras de rodas etc
	urbanParkingSiteType	Texto	Tipo de parque de estacionamento

## b) Locais

O dataset dos locais contém o nome, latitude e longitude dos vários parques de estacionamento anteriormente recolhidos e para cada parque existe um array que contém os 60 locais mais próximos. Ou seja, este dataset contempla os dados identificativos dos 63 parques de estacionamento, e os 60 locais mais próximos de cada um.

Segue-se, na Tabela 13, a apresentação da estrutura do dataset, indicando para cada atributo qual o tipo de dado e a descrição.

Tabela 13: Descrição do dataset dos locais

Objeto				Tipo	Descrição
parkingSite	name			Texto	Nome do parque de estacionamento
	parkingLocation	latitude		Decimal	Latitude do parque de estacionamento
		longitude		Decimal	Longitude do parque de estacionamento
	geometry	location	latitude	Decimal	Latitude do ponto de interesse
			longitude	Decimal	Longitude do Ponto de interesse
	Continua na próxima página				

PointOfInterests

Objeto					Tipo	Descrição			
			viewport	northeast	lat **	Decimal	Latitude da janela de exibição no mapa a nordeste		
					lng **	Decimal	Longitude da janela de exibição no mapa a nordeste		
				southwest	lat **	Decimal	Latitude da janela de exibição no mapa a sudoeste		
					lng **	Decimal	Longitude da janela de exibição no mapa a sudoeste		
			icon **					Texto	URL do icon do elemento no mapa
			id **					Texto	Identificador Google
			name					Texto	Nome do local
			place_id **					Texto	Identificador único do local
Continua na próxima página									

<b>Objeto</b>		<b>Tipo</b>	<b>Descrição</b>
	plus_code **	Objeto	Referência do local codificada em "plus_code".
	reference **	Texto	Referência google
	scope **	Texto	Abrangência - Sempre "GOOGLE"
	types *	Array	Array que contém as várias categorias/tipologias que o local se insere
	rating	Decimal	Média das classificações dos utilizadores no Google
	user_ratings_total	Inteiro	Número total de classificações dos utilizadores da Google
	vacinity	Texto	Nome de um local próximo



## c) Meteorologia

O dataset da meteorologia contém o histórico das condições meteorológicas de Bruxelas entre 01-01-2015 e 31-12-2019, ou seja um total de 1826 dias, que se traduz em 1826 registos. Cada registo contém várias informações entre as quais o rastreamento por hora das condições meteorológicas.

Na Tabela 14 é apresentada a estrutura do ficheiro JSON no qual é indicado qual para cada atributo qual o tipo de dado e a descrição.

Tabela 14: Descrição dos dados meteorológicos

Objeto		Tipo	Descrição	
latitude		Decimal	Latitude da qual foram extraídos os dados meteorológicos	
longitude		Decimal	Longitude da qual foram extraídos os dados meteorológicos	
timezone		Texto	Fuso horário considerado	
hourly	summary **	Texto	Resumo da informação horária	
	icon **	Texto	Texto identificador para utilização de icons	
	data	time	Inteiro	Dia e hora do registo no formato Unix
		summary	Texto	Resumo horário da meteorologia
		icon **	Texto	Texto identificador para utilização de icons
		precipIntensity	Decimal	Intensidade da precipitação
		precipProbability	Decimal	Probalilidade de ocorrer precipitação
		temperature	Decimal	Temperatura do ar
		apparentTemperature	Decimal	Temperatura aparente (temperatura equivalente à sentida pelo homem).
		dewPoint	Decimal	Ponto de condensação da água
		humidity	Decimal	Humidade relativa

Continua na próxima página

		<b>Objeto</b>	<b>Tipo</b>	<b>Descrição</b>
		pressure	Decimal	Pressão do ar
		windSpeed	Decimal	Velocidade do vento
		windGust	Decimal	Rajada de vento
		windBearing	Decimal	Direção do vento
		cloudCover	Decimal	Percentagem de céu nublado
		uvIndex	Decimal	Nível de raios UV
		visibility	Decimal	Visibilidade média
daily	data	time **	Inteiro	Dia e hora do registo no formato Unix
		summary **	Texto	Resumo diário da meteorologia
		icon **	Texto	Texto identificador para utilização de ícons
		sunriseTime **	Inteiro	Tempo exato do nascer do sol
		sunsetTime **	Inteiro	Tempo exato do pôr do sol
		moonPhase **	Decimal	Identificador da fase da lua
		precipIntensity **	Decimal	Intensidade da precipitação
		precipIntensityMax **	Decimal	Precipitação máxima registada no dia
		precipIntensityMaxTime **	Inteiro	Momento exato da precipitação máxima
		precipProbability **	Decimal	Probabilidade de ocorrer precipitação
		precipType **	Texto	Tipo de precipitação
		temperatureHigh **	Decimal	Temperatura mais alta do dia
		temperatureHighTime **	Inteiro	Momento exato em que se registou a temperatura mais alta do dia
		temperatureLow **	Decimal	Temperatura mais baixa do dia
temperatureLowTime **	Inteiro	Momento exato em que se registou a temperatura mais baixa do dia		
apparentTemperatureHigh **	Decimal	Máximo de temperatura aparente registada		
Continua na próxima página				

	<b>Objeto</b>	<b>Tipo</b>	<b>Descrição</b>
	apparentTemperatureHighTime **	Inteiro	Momento exato que se registou o valor mais elevado de temperatura aparente
	apparentTemperatureLow **	Decimal	Mínimo de temperatura aparente registada
	apparentTemperatureLowTime **	Inteiro	Momento exato em que se registou valor mais baixo de temperatura aparente
	dewPoint **	Decimal	Ponto de condensação da água
	humidity **	Decimal	Humidade relativa
	pressure **	Decimal	Pressão do ar
	windSpeed **	Decimal	Velocidade do vento
	windGust **	Decimal	Rajada de vento mais forte
	windGustTime **	Inteiro	Momento do dia em que se registou a rajada de vento mais forte
	windBearing **	Decimal	Direção do vento
	cloudCover **	Decimal	Percentagem de céu nublado
	uvIndex **	Inteiro	Nível de raios UV
	uvIndexTime **	Inteiro	Momento do dia em que se registou o maior índice de emissão de raios UV
	visibility **	Decimal	Visibilidade Média
	temperatureMin **	Decimal	Temperatura mínima do dia
	temperatureMinTime **	Inteiro	Momento exato em que se registou a temperatura mínima do dia
	temperatureMax **	Decimal	Temperatura máxima do dia
	temperatureMaxTime **	Inteiro	Momento exato em que se registou a temperatura máxima do dia
	apparentTemperatureMin **	Decimal	Temperatura mínima aparente

Continua na próxima página

Objeto		Tipo	Descrição
	apparentTemperatureMinTime **	Inteiro	Momento do dia em que se registou a temperatura mínima aparente
	apparentTemperatureMax **	Decimal	Temperatura máxima aparente
	apparentTemperatureMaxTime **	Inteiro	Momento do dia em que se registou a temperatura máxima aparente
offset		Inteiro	Fuso horário em horas

## d) Feriados

No que a feriados diz respeito, foram extraídos 19 por ano. Como foram considerados 6 anos, conclui-se que foram recolhidos um total de 114 feriados.

De seguida, na Tabela 15, são apresentados os diversos atributos associados aos feriados, indicado para cada, qual o tipo de dado e a descrição.

Tabela 15: Descrição do dataset dos feriados

Objeto		Tipo	Descrição
Year		Texto	Atributo que permite identificar o ano dos feriados aglomerados no array de holidays
holidays	date	Data	Data do feriado (YYYY-MM-DD)
	start	Data	Data de início do feriado com hora associada (YYYY-MM-DDTHH:mm:ss.sssZ)
	end	Data	Data de término do feriado com a hora associada (YYYY-MM-DDTHH:mm:ss.sssZ)
	name	Texto	Nome do feriado
Continua na próxima página			

	<b>Objeto</b>	<b>Tipo</b>	<b>Descrição</b>
	type	Texto	Tipologia do feriado (public, bank, school, observance)
	public	Booleano	Atributo que indica se um feriado é público ou não
	country	Texto	País onde o feriado se regista

## e) Ocupação dos parques

Para cada um dos 63 parques de estacionamento foram geradas ocupações de 60 em 60 minutos considerando o período de 01-01-2015 até 01-01-2020.

A Tabela 16 apresenta para cada atributo qual o tipo de dado e a descrição associados a um registo de ocupação.

Tabela 16: Descrição do dataset das ocupações dos parques

<b>Atributo</b>	<b>Tipo</b>	<b>Descrição</b>
park_name	Texto	Nome do parque
park_id	Texto	Identificador único do parque
date	Date	Data do registo no formato (yyyy-mm-dd hh:mm:ss)
dayOfWeek	Inteiro	Dia da semana do registo
timestamp	Inteiro	Unix timestamp do registo
occupation	Decimal	Percentagem de ocupação do parque

### 4.1.3 Exploração dos dados

Para a exploração dos dados recorreu-se ao Talend Open Studio for Data Quality. Importante denotar que para tirar proveito das funcionalidades que esta ferramenta disponibiliza foi necessário converter todos os dados até então recolhidos para o formato csv. Para efetuar esta conversão foram desenvolvidas

*scripts* de Python.

Para as várias colunas dos diferentes datasets foram efetuadas análises estatísticas que permitem obter maior conhecimento acerca dos dados. As análises efectuadas contemplam a análise de estatísticas simples, frequências de valores e estatísticas mais avançadas.

As estatísticas simples contemplam indicadores como número de linhas, o número de nulos, o número de valores únicos e o número de valores duplicados. Na análise da frequência dos valores é possível verificar quais os valores mais comuns e os menos. As estatísticas avançadas contemplam indicadores como a média, a mediana, os quartis, entre outros.

Desta forma, para atributos textuais, fora efectuadas análises das estatísticas simples e da frequência de valores. Para os atributos numéricos, além dos mencionados anteriormente, foi acrescentada a análise das estatísticas avançadas.

## **a) Parques de estacionamento**

Na exploração do dataset dos parques de estacionamento alguns atributos foram descartados ou por serem irrelevantes para o projeto ou devido à necessidade de processamento para se conseguir extrair informação, processamento este que será feito numa fase posterior. Assim sendo, os atributos que na etapa da descrição de dados a) estão marcados com \*\* são atributos previamente descartados. Os campos destacados com \* são os atributos que necessitam de algum processamento para que se possa obter informação utilizável.

Uma vez que o número de atributos é considerável, seguem-se os resultados para uma série de atributos seleccionados. Em anexo podem ser acedidos os resultados completos (Anexo E).

- **parking\_group\_name** - parkingRecord.parkingName.values.value.#text

▼ Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	15	23.81%
Unique Count	8	12.70%
Duplicate Count	7	11.11%
Blank Count	0	0.00%

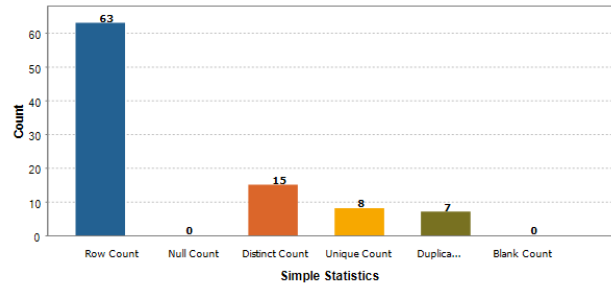
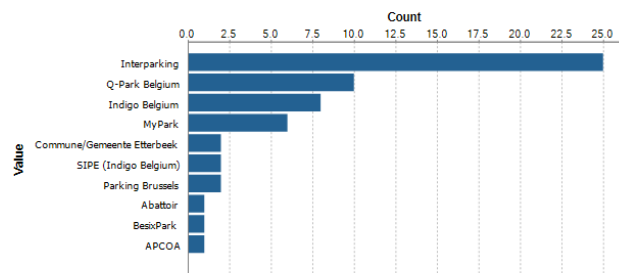


Figura 8: Parques de estacionamento - Estatísticas Simples do atributo 'parking\_group\_name'

▼ Value Frequency

Value	Count	%
Interparking	25	39.68%
Q-Park Belgium	10	15.87%
Indigo Belgium	8	12.70%
MyPark	6	9.52%
Commune/Gemeente Etter...	2	3.17%
SIPE (Indigo Belgium)	2	3.17%
Parking Brussels	2	3.17%
Abattoir	1	1.59%
BesixPark	1	1.59%
APCOA	1	1.59%



▼ Value Low Frequency

Value	Count	%
Rauwers Parkingshop	1	1.59%
Abattoir	1	1.59%
Commune/Gemeente Mole...	1	1.59%
Hobinvest	1	1.59%
Hotel Warwick	1	1.59%
Commune/Gemeente Ixelle...	1	1.59%
BesixPark	1	1.59%
APCOA	1	1.59%
Commune/Gemeente Etter...	2	3.17%
Parking Brussels	2	3.17%

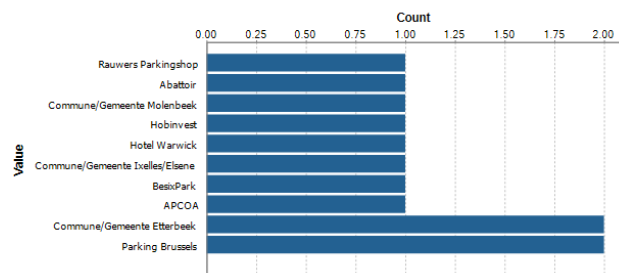


Figura 9: Parques de estacionamento - Frequências do atributo 'parking\_group\_name'

- **parking\_name** - parkingSite.parkingName.values.value[0].#text

▼ Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	63	100.00%
Unique Count	63	100.00%
Duplicate Count	0	0.00%
Blank Count	0	0.00%

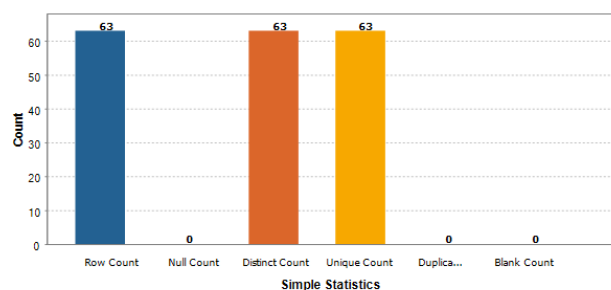


Figura 10: Parques de estacionamento - Estatísticas Simples do atributo 'park\_name'

- **parkingNumberOfSpaces** - parkingSite.parkingNumberOfSpaces

Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	59	93.65%
Unique Count	55	87.30%
Duplicate Count	4	6.35%

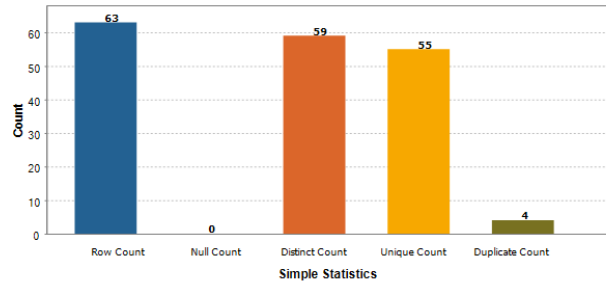
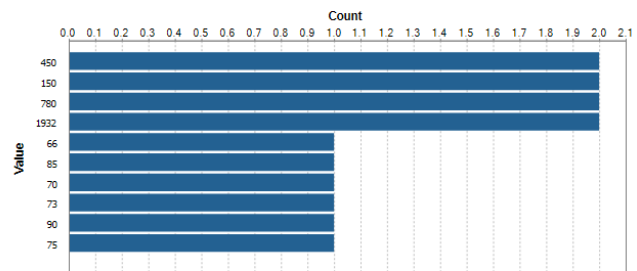


Figura 11: Parques de estacionamento - Estatísticas Simples do atributo 'ParkingNumberOfSpaces'

Value Frequency

Value	Count	%
450	2	3.17%
150	2	3.17%
780	2	3.17%
1932	2	3.17%
66	1	1.59%
85	1	1.59%
70	1	1.59%
73	1	1.59%
90	1	1.59%
75	1	1.59%



Value Low Frequency

Value	Count	%
96	1	1.59%
66	1	1.59%
115	1	1.59%
85	1	1.59%
70	1	1.59%
73	1	1.59%
105	1	1.59%
90	1	1.59%
75	1	1.59%
92	1	1.59%

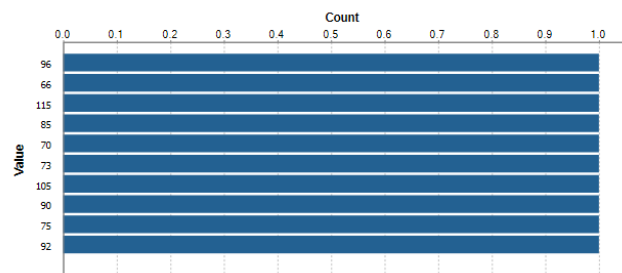
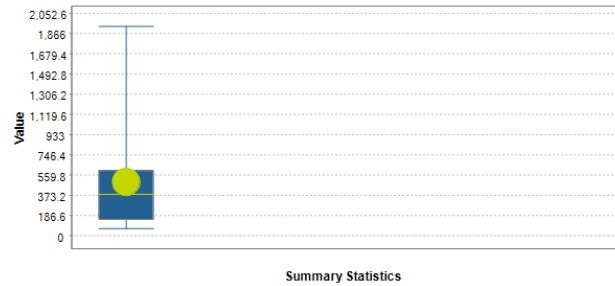


Figura 12: Parques de estacionamento - Frequências do atributo 'ParkingNumberOfSpaces'



## ▼ Summary Statistics

Label	Value
Mean	495.85714285714283
Median	381.0
Inter Quartile Range	452.0
Lower Quartile	150.0
Upper Quartile	602.0
Range	1866.0
Minimum	66
Maximum	1932



## ▼ Mode

Mode
150

Figura 13: Parques de estacionamento - Estatísticas do atributo 'ParkingNumberOfSpaces'

## b) Locais

Na exploração do dataset dos locais foram descartados alguns atributos que não considerados importantes para o propósito do projeto. Também os atributos que implicam algum tipo de processamento para se conseguir extrair informação relevante não foram introduzidos na ferramenta utilizada para a exploração dos dados. Mais uma vez, os atributos destacados com \*\* na fase de descrição dos dados dos locais b), são atributos previamente descartados e os quais não serão feitas análises exploratórias. Os atributos marcados com \* necessitam de algum processamento para se extrair informação, pelo que serão apenas abordados na etapa da "Qualidade dos dados". De seguida segue-se o resultado da análise para cada atributo selecionado, sendo que a análise completa pode ser acedida no Anexo F.

- **place\_name** - parkingSite.PointsOfInterests.name

▼ Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	869	22.99%
Unique Count	357	9.44%
Duplicate Count	512	13.54%
Blank Count	0	0.00%

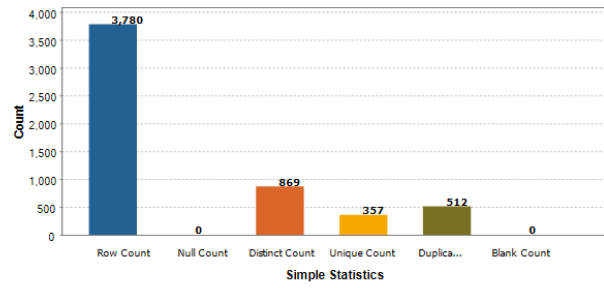
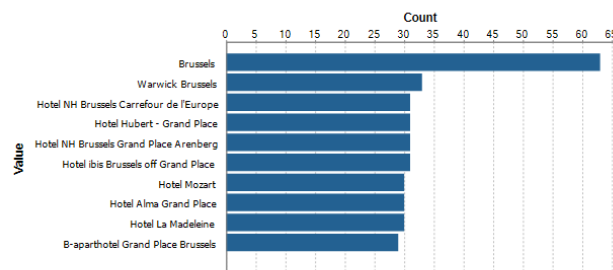


Figura 14: Locais - Estatísticas Simples do atributo 'place\_name'

▼ Value Frequency

Value	Count	%
Brussels	63	1.67%
Warwick Brussels	33	0.87%
Hotel NH Brussels Carrefour...	31	0.82%
Hotel Hubert - Grand Place	31	0.82%
Hotel NH Brussels Grand Pl...	31	0.82%
Hotel ibis Brussels off Gran...	31	0.82%
Hotel Mozart	30	0.79%
Hotel Alma Grand Place	30	0.79%
Hotel La Madeleine	30	0.79%
B-aparthotel Grand Place B...	29	0.77%



▼ Value Low Frequency

Value	Count	%
ACTIVA	1	0.03%
A.C.E.	1	0.03%
AXA Assistance Broker Solu...	1	0.03%
A.D.I.R. Asbl	1	0.03%
A l'Entreprise Franky sprl	1	0.03%
ACV Nationaal	1	0.03%
AWEX - Agence wallonne à...	1	0.03%
AccessAndGo	1	0.03%
Action Service Company	1	0.03%
Administration communale...	1	0.03%

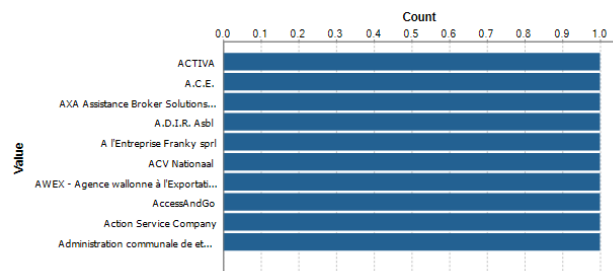


Figura 15: Locais - Frequências do atributo 'place\_name'

- **rating** - parkingSite.PointsOfInterests.rating

▼ Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	37	0.98%
Unique Count	3	0.08%
Duplicate Count	34	0.90%
Blank Count	359	9.50%

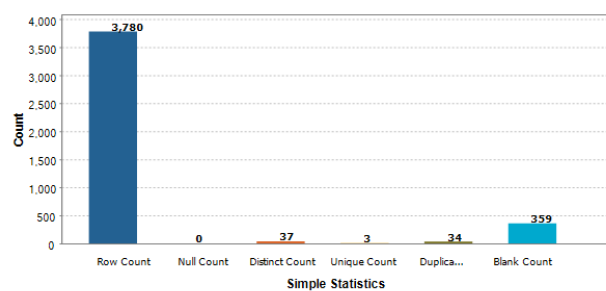
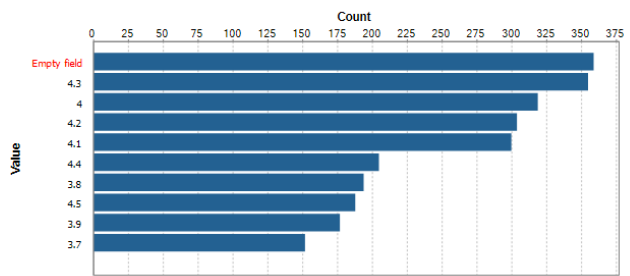


Figura 16: Locais - Estatísticas Simples do atributo 'place\_rating'

Value Frequency

Value	Count	%
Empty field	359	9.50%
4.3	355	9.39%
4	319	8.44%
4.2	304	8.04%
4.1	300	7.94%
4.4	205	5.42%
3.8	194	5.13%
4.5	188	4.97%
3.9	177	4.68%
3.7	152	4.02%



Value Low Frequency

Value	Count	%
2.1	1	0.03%
1.7	1	0.03%
1.9	1	0.03%
1.5	3	0.08%
1.8	3	0.08%
2.2	5	0.13%
1	10	0.26%
2.8	11	0.29%
2.3	16	0.42%
2.6	20	0.53%

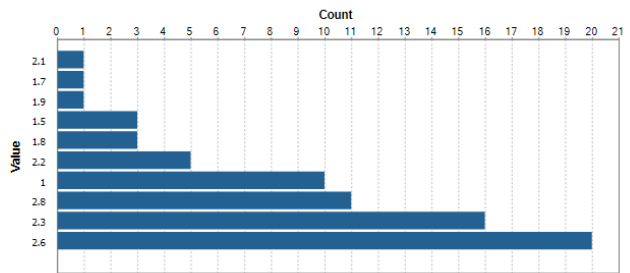


Figura 17: Locais - Frequências do atributo 'place\_rating'

Summary Statistics

Label	Value
Mean	3.9358082432037413
Median	4.1
Inter Quartile Range	0.5999999999999996
Lower Quartile	3.7
Upper Quartile	4.3
Range	4.0
Minimum	1.0
Maximum	5.0

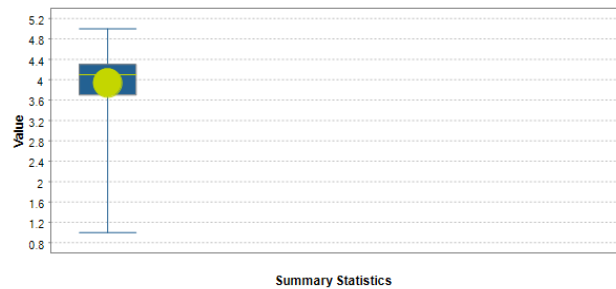


Figura 18: Locais - Estatísticas do atributo 'place\_rating'

• user\_ratings\_total - parkingSite.PointsOfInterests.user\_ratings\_total

Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	300	7.94%
Unique Count	49	1.30%
Duplicate Count	251	6.64%
Blank Count	359	9.50%

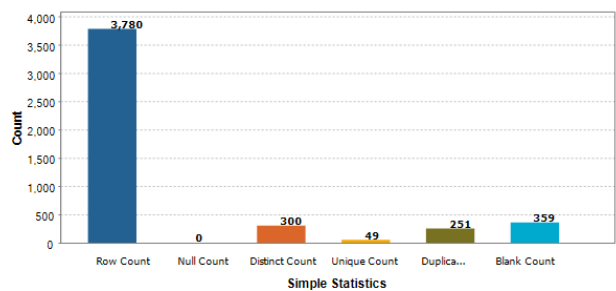
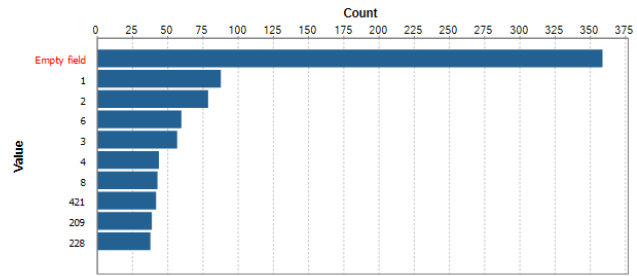


Figura 19: Locais - Estatísticas Simples do atributo 'place\_user\_ratings\_total'

▼ Value Frequency

Value	Count	%
Empty field	359	9.50%
1	88	2.33%
2	79	2.09%
6	60	1.59%
3	57	1.51%
4	44	1.16%
8	43	1.14%
421	42	1.11%
209	39	1.03%
228	38	1.01%



▼ Value Low Frequency

Value	Count	%
143	1	0.03%
1128	1	0.03%
103	1	0.03%
125	1	0.03%
136	1	0.03%
104	1	0.03%
107	1	0.03%
118	1	0.03%
150	1	0.03%
142	1	0.03%

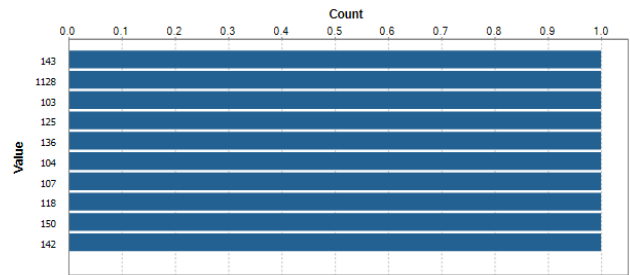


Figura 20: Locais - Frequências do atributo 'place\_user\_ratings\_total'

▼ Summary Statistics

Label	Value
Mean	3.9358082432037413
Median	4.1
Inter Quartile Range	0.5999999999999996
Lower Quartile	3.7
Upper Quartile	4.3
Range	4.0
Minimum	1.0
Maximum	5.0

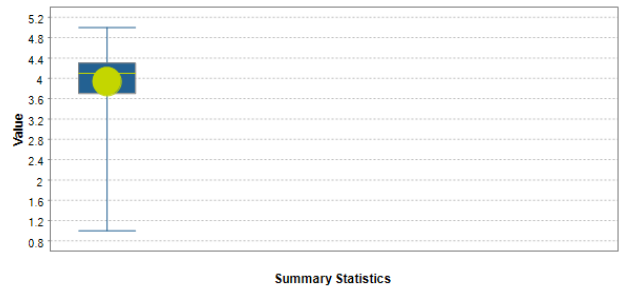


Figura 21: Locais - Estatísticas do atributo 'place\_user\_ratings\_total'

### c) Meteorologia

Do dataset da meteorologia, e tal como nas explorações anteriormente efectuadas, também foram descartados da análise exploratória alguns atributos considerados pouco relevantes para o propósito do projeto. Estes atributos encontram-se precedidos de um \* na etapa de descrição dos dados c). Seguem-se então os resultados extraídos da ferramenta utilizada para a exploração dos dados, para uma série de atributos selecionados. No Anexo G é possível consultar os resultados completos.

- **hourly\_precipIntensity** - hourly.data[n].precipIntensity

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	4916	11.82%
Unique Count	2806	6.74%
Duplicate Count	2110	5.07%

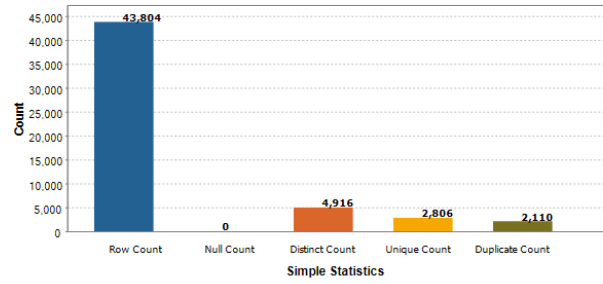
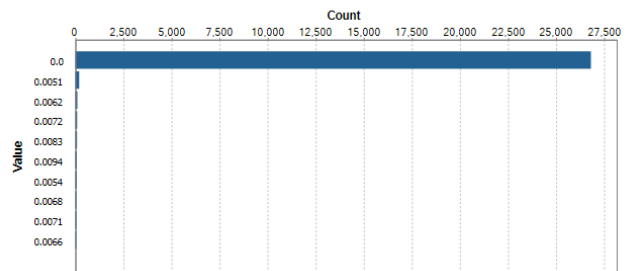


Figura 22: Meteorologia - Estatísticas Simples do atributo 'hourly\_precipIntensity'

Value Frequency

Value	Count	%
0.0	26817	64.46%
0.0051	197	0.47%
0.0062	106	0.25%
0.0072	95	0.23%
0.0083	86	0.21%
0.0094	79	0.19%
0.0054	71	0.17%
0.0068	62	0.15%
0.0071	60	0.14%
0.0066	55	0.13%



Value Low Frequency

Value	Count	%
0.0338	1	2.404E-3%
0.0481	1	2.404E-3%
0.0437	1	2.404E-3%
0.0645	1	2.404E-3%
0.074	1	2.404E-3%
0.0357	1	2.404E-3%
0.0635	1	2.404E-3%
0.0694	1	2.404E-3%
0.0697	1	2.404E-3%
0.0665	1	2.404E-3%

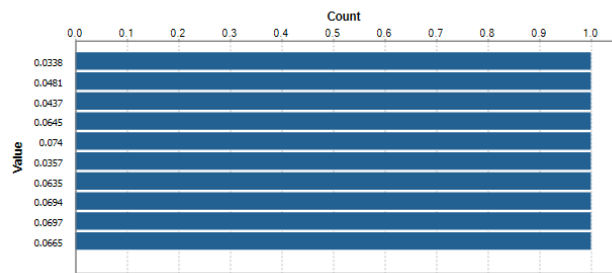


Figura 23: Meteorologia - Frequências do atributo 'hourly\_precipIntensity'

Summary Statistics

Label	Value
Mean	0.07832104946877554
Median	0.0
Inter Quartile Range	0.0201
Lower Quartile	0.0
Upper Quartile	0.0201
Range	27.3784
Minimum	0.0
Maximum	27.3784

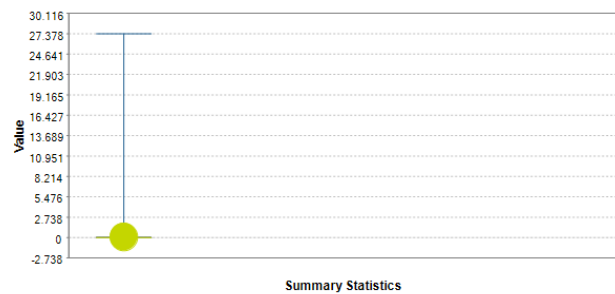


Figura 24: Meteorologia - Estatísticas do atributo 'hourly\_precipIntensity'

- **hourly\_temperature** - hourly.data[n].temperature

▼ Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	3548	8.10%
Unique Count	386	0.88%
Duplicate Count	3162	7.22%

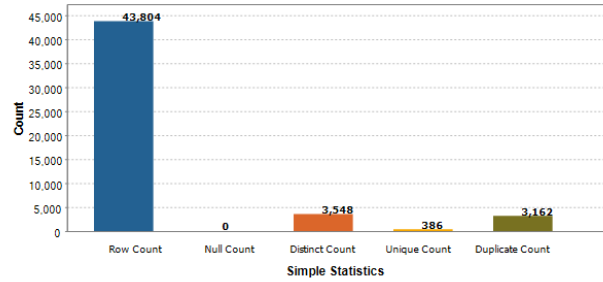
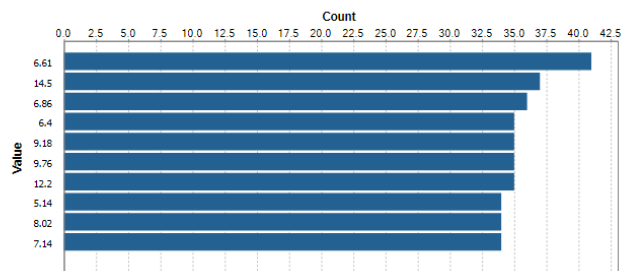


Figura 25: Meteorologia - Estatísticas Simples do atributo 'hourly\_temperature'

▼ Value Frequency

Value	Count	%
6.61	41	0.09%
14.5	37	0.08%
6.86	36	0.08%
6.4	35	0.08%
9.18	35	0.08%
9.76	35	0.08%
12.2	35	0.08%
5.14	34	0.08%
8.02	34	0.08%
7.14	34	0.08%



▼ Value Low Frequency

Value	Count	%
-6.49	1	2.283E-3%
-8.19	1	2.283E-3%
-7.67	1	2.283E-3%
-6.55	1	2.283E-3%
-7.62	1	2.283E-3%
-8.15	1	2.283E-3%
-8.07	1	2.283E-3%
-6.62	1	2.283E-3%
-7.04	1	2.283E-3%
-6.94	1	2.283E-3%

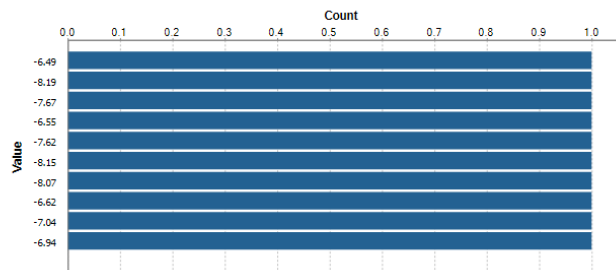


Figura 26: Meteorologia - Frequências do atributo 'hourly\_temperature'

▼ Summary Statistics

Label	Value
Mean	11.653023377927948
Median	11.29
Inter Quartile Range	10.0
Lower Quartile	6.46
Upper Quartile	16.46
Range	46.61
Minimum	-8.19
Maximum	38.42

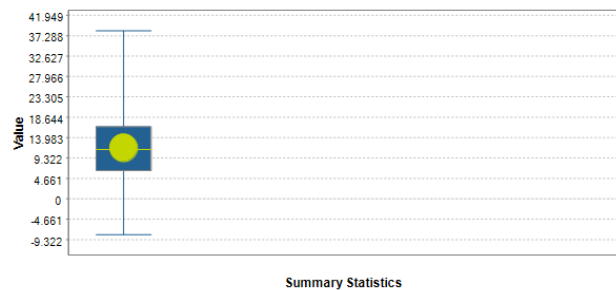


Figura 27: Meteorologia - Estatísticas do atributo 'hourly\_temperature'

## d) Feriados

O dataset que contém o histórico de feriados em Bruxelas também foi explorado com recurso à ferramenta Talend Open Studio for Data Quality. Segue-se os resultados extraídos da ferramenta, para uma série de atributos selecionados. No Anexo H é possível aceder aos resultados completos.

- **name** - holidays[n].name

### Simple Statistics

Label	Count	%
Row Count	114	100.00%
Null Count	0	0.00%
Distinct Count	19	16.67%
Unique Count	0	0.00%
Duplicate Count	19	16.67%
Blank Count	0	0.00%

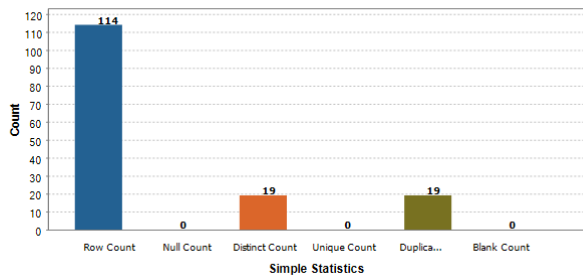
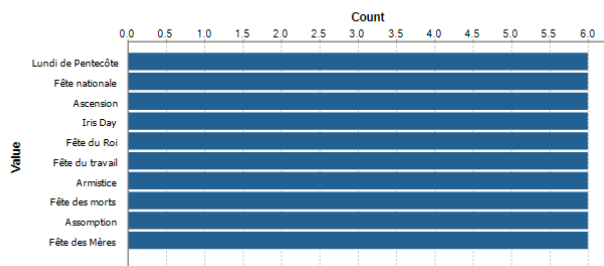


Figura 28: Feriados - Estatísticas Simples do atributo 'name'

### Value Frequency

Value	Count	%
Lundi de Pentecôte	6	5.26%
Fête nationale	6	5.26%
Ascension	6	5.26%
Iris Day	6	5.26%
Fête du Roi	6	5.26%
Fête du travail	6	5.26%
Armistice	6	5.26%
Fête des morts	6	5.26%
Assomption	6	5.26%
Fête des Mères	6	5.26%



### Value Low Frequency

Value	Count	%
Lundi de Pentecôte	6	5.26%
Fête nationale	6	5.26%
Ascension	6	5.26%
Iris Day	6	5.26%
Fête du Roi	6	5.26%
Fête du travail	6	5.26%
Armistice	6	5.26%
Fête des morts	6	5.26%
Assomption	6	5.26%
Fête des Mères	6	5.26%

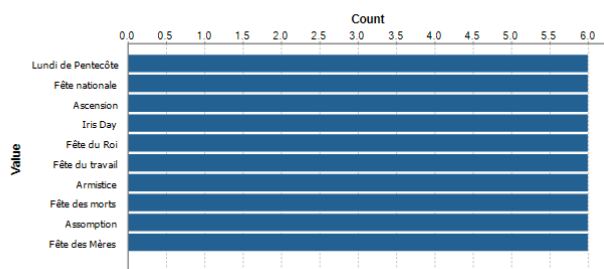


Figura 29: Feriados - Frequências do atributo 'name'

- **type** - holidays[n].type

▼ Simple Statistics

Label	Count	%
Row Count	114	100.00%
Null Count	0	0.00%
Distinct Count	3	2.63%
Unique Count	0	0.00%
Duplicate Count	3	2.63%
Blank Count	0	0.00%

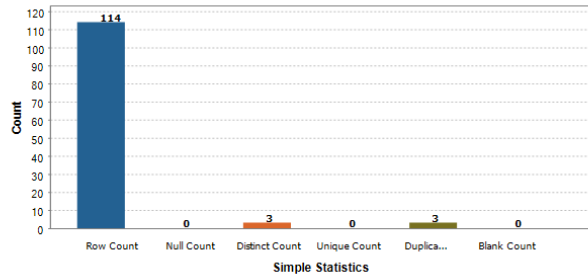
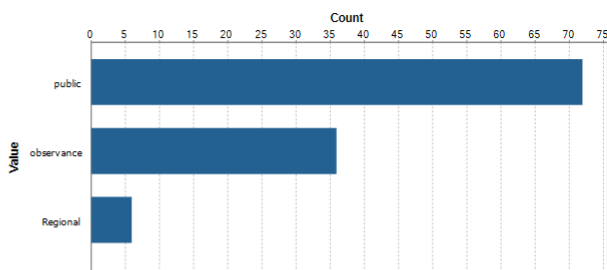


Figura 30: Feriados - Estatísticas Simples do atributo 'type'

▼ Value Frequency

Value	Count	%
public	72	63.16%
observance	36	31.58%
Regional	6	5.26%



▼ Value Low Frequency

Value	Count	%
Regional	6	5.26%
observance	36	31.58%
public	72	63.16%

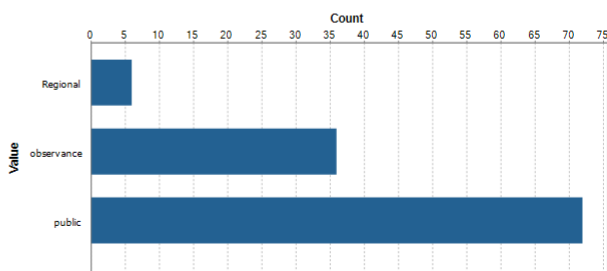


Figura 31: Feriados - Frequências do atributo 'type'

- **public** - holidays[n].public

▼ Simple Statistics

Label	Count	%
Row Count	114	100.00%
Null Count	0	0.00%
Distinct Count	2	1.75%
Unique Count	0	0.00%
Duplicate Count	2	1.75%
Blank Count	0	0.00%

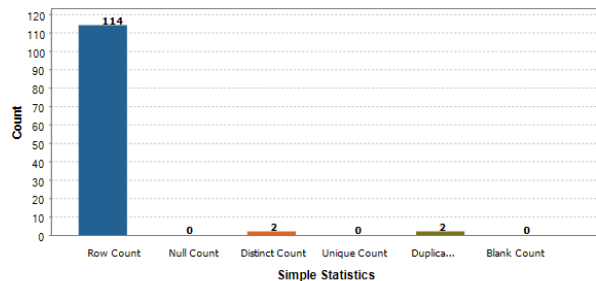
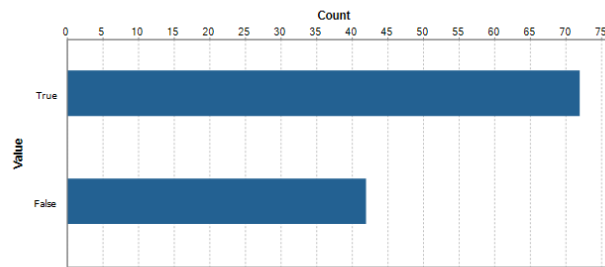


Figura 32: Feriados - Estatísticas Simples do atributo 'public'



▼ Value Frequency

Value	Count	%
True	72	63.16%
False	42	36.84%



▼ Value Low Frequency

Value	Count	%
False	42	36.84%
True	72	63.16%

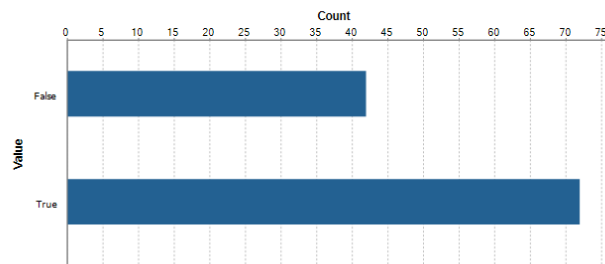


Figura 33: Feriados - Frequências do atributo 'public'

## e) Ocupação dos parques

O dataset que contém o histórico de ocupações dos vários parques de Bruxelas também foi explorado com recurso à ferramenta Talend Open Studio for Data Quality. Segue-se os resultados extraídos da ferramenta, apenas para o atributo 'occupation', sendo que os restantes podem ser acedidos no Anexo I.

- **occupation**

▼ Simple Statistics

Label	Count	%
Row Count	5521824	100.00%
Null Count	0	0.00%
Distinct Count	1034	0.02%
Unique Count	33	5.976E-4%
Duplicate Count	1001	0.02%

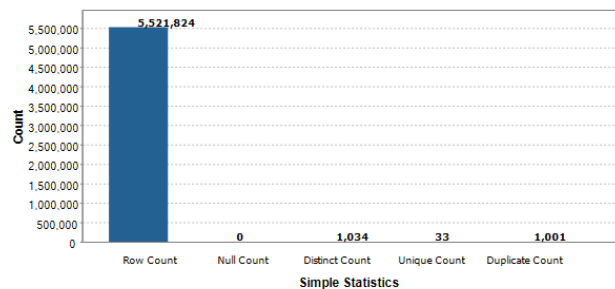
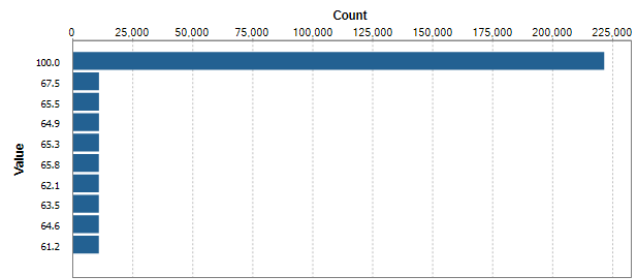


Figura 34: Ocupação dos parques - Estatísticas Simples do atributo 'occupation'

Value Frequency

Value	Count	%
100.0	221693	4.01%
67.5	11113	0.20%
65.5	11108	0.20%
64.9	11105	0.20%
65.3	11103	0.20%
65.8	11097	0.20%
62.1	11084	0.20%
63.5	11069	0.20%
64.6	11069	0.20%
61.2	11051	0.20%



Value Low Frequency

Value	Count	%
0.3	50	9.055E-4%
0.4	51	9.236E-4%
1.1	51	9.236E-4%
0.5	52	9.417E-4%
0.1	52	9.417E-4%
0.2	53	9.598E-4%
0.6	54	9.779E-4%
1.4	64	1.159E-3%
0.8	64	1.159E-3%
1.6	64	1.159E-3%

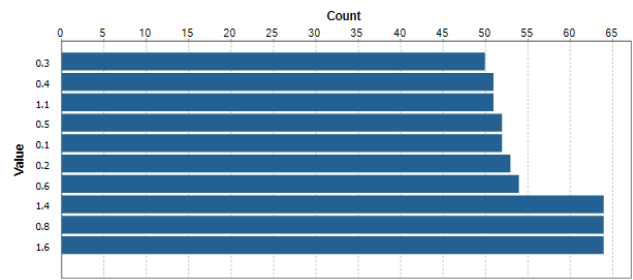


Figura 35: Ocupação dos parques - Frequências do atributo 'occupation'

Summary Statistics

Label	Value
Mean	64.66735513482502
Median	65.0
Inter Quartile Range	27.0
Lower Quartile	51.5
Upper Quartile	78.5
Range	100.0
Minimum	0.0
Maximum	100.0

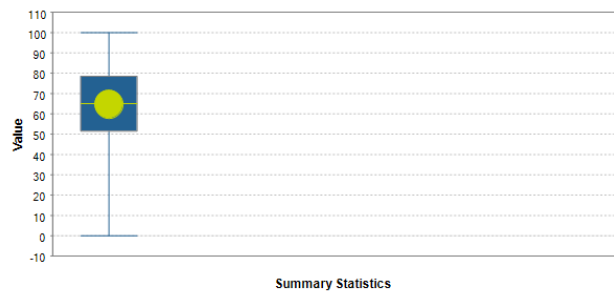


Figura 36: Ocupação dos parques - Estatísticas do atributo 'occupation'

### 4.1.4 Qualidade dos dados

O resultado da secção anterior 4.1.3 serve de base para esta no qual se pretende fazer um relatório de qualidade dos dados. Neste sentido, segue-se para cada dataset uma análise de qualidade, no qual são descritos erros e/ou problemas identificados e a respetiva proposta de correção.

#### a) Parques de estacionamento

O dataset dos parques de estacionamento de Bruxelas contempla diversas informações acerca dos parques de estacionamento, o que possibilita efetuar diversas correlações entre as características do parque e as taxas de ocupação.

**Análise geral** Dataset muito completo, mas ao mesmo tempo complexo de ser utilizado e de se extrair toda a informação necessária.

Na Tabela 17 é possível verificar para cada atributo do dataset quais os problemas detetados e apresentando, caso seja necessário, uma proposta de resolução.

Tabela 17: Análise de qualidade do dataset dos parques de estacionamento

<b>Atributo</b>	<b>Problemas detetados</b>	<b>Proposta de resolução</b>
parkingRecord[n].parkingName	Não foram detetados problemas	-
parkingRecord[n].operator	Objeto complexo no qual é necessário fazer seleção da informação	Analisar cada um dos atributos do objeto e caso se decida utilizar uma determinada coluna o nome da mesma será operador_<nome_da_coluna>
parkingRecord[n].parkingSite[x].@id	Não foram detetados problemas	-
parkingRecord[n].parkingSite[x].parkingName	Contém o nome em Francês e Holandês	Considera-se apenas o nome em Francês
parkingRecord[n].parkingSite[x].parkingNumberOfSpaces	Não foram detetados problemas	-
parkingRecord[n].parkingSite[x].responsibleAuthority	Objeto complexo no qual é necessário fazer seleção da informação	Analisar cada um dos atributos do objeto e caso se decida utilizar uma determinada coluna o nome da mesma será responsibleAuthority_<nome_da_coluna>
Continua na próxima página		

<b>Atributo</b>	<b>Problemas detetados</b>	<b>Proposta de resolução</b>
parkingRecord[n].parkingSite[x].operator	Objeto complexo no qual é necessário fazer seleção da informação	Analisar cada um dos atributos do objeto e caso se decida utilizar uma determinada coluna o nome da mesma será operador_<nome_da_coluna>
parkingRecord[n].parkingSite[x].parkingLocation	Não foram detetados problemas	-
parkingRecord[n].parkingSite[x].onlyAssignedParking	Objeto que necessita de formatação para se extrair informação acerca das dimensões máximas dos veículos permitidas	Criar três atributos, altura máxima, largura máxima e comprimento máximo
parkingRecord[n].parkingSite[x].tariffsAndPayment	Não tem qualquer tipo de padrão o preçário *	Descartar os parques que não tem preço hora
parkingRecord[n].parkingSite[x].parkingEquipmentOrServiceFacility	Atributo que necessita de processamento para obter informação útil	Serão consideradas todas os serviços e equipamentos presentes no dataset e para cada parque haverá um booleano a indicar se tem ou não o serviço/equipamento
parkingRecord[n].parkingSite[x].groupOfParkingSpaces	Não foram detetados problemas	-
parkingRecord[n].parkingSite[x].parkingSiteAddress	Objeto complexo no qual é necessário fazer seleção da informação	Analisar cada um dos atributos do objeto e caso se decida utilizar uma determinada coluna o nome da mesma será parkingSiteAddress_<nome_da_coluna>
Continua na próxima página		

<b>Atributo</b>	<b>Problemas detetados</b>	<b>Proposta de resolução</b>
parkingRecord[n].parkingSite[x].openingTimes	Em alguns casos não é indicado a hora de abertura e fecho	Utilizar hora de abertura e fecho caso exista senão assume-se que está sempre aberto
parkingRecord[n].parkingSite[x].parkingAccess	Objeto complexo no qual é preciso processamento para obter informação	Registar número de entradas e saídas
parkingRecord[n].parkingSite[x].urbanParkingSiteType	Não foram detetados problemas	-

## b) Locais

**Análise geral** O dataset contém informações necessárias para que seja possível medir o impacto da proximidade de determinados locais aos parques de estacionamento. O dataset contempla 60 lugares relevantes para cada parque de estacionamento. Sendo um número elevado será necessário definir qual o número de locais a considerar.

De seguida é apresentada uma tabela, Tabela 18, na qual é possível verificar quais os problemas detetados e a proposta de resolução destes problemas para os diversos atributos do dataset.

Tabela 18: Análise de qualidade do dataset dos locais

<b>Atributo</b>	<b>Problemas detetados</b>	<b>Proposta de resolução</b>
parkingSite[n].name	Não foram detetados problemas	-
parkingSite[n].parkingLocation.latitude	Não foram detetados problemas	-
parkingSite[n].parkingLocation.longitude	Não foram detetados problemas	-
parkingSite[n].pointOfInterests[x].geometry.location.latitude	Não foram detetados problemas	-
Continua na próxima página		

<b>Atributo</b>	<b>Problemas detetados</b>	<b>Proposta de resolução</b>
parkingSite[n].pointOfInterests[x]. geometry.location.longitude	Não foram detetados problemas	-
parkingSite[n].pointOfInterests[x]. name	Não foram detetados problemas	-
parkingSite[n].pointOfInterests[x]. types	Array que pode ter um tamanho variável	Definir um máximo de tipos e cada tipo origina uma coluna
parkingSite[n].pointOfInterests[x]. rating	Valores a Branco e/ou nulo	Descarta-se o atributo
parkingSite[n].pointOfInterests[x]. user_ratings_total	Valores a Branco e/ou nulo	Descarta-se o atributo
parkingSite[n].pointOfInterests[x]. vacinity	Valores a Branco e/ou nulo	Colocar "não definido" nos valores a branco e/ou nulos

### c) Meteorologia

**Análise geral** O dataset contém o histórico das condições meteorológicas em Bruxelas. Desta forma assume-se que para toda a cidade as condições meteorológicas são uniformes, pois num contexto ideal seria benéfico ter as condições meteorológicas para cada uma das latitudes e longitudes dos diferentes parques, contudo isto implicaria aumentar o número de pedidos efectuados à API da Darksky o que traduziria num aumento dos custos do projeto.

A Tabela 19 apresenta quais os problemas detetados e a respetiva resposta de resolução para os diversos atributos do dataset.

Tabela 19: Análise de qualidade do dataset da Meteorologia

<b>Atributo</b>	<b>Problemas detetados</b>	<b>Proposta de resolução</b>
latitude	Não foram detetados problemas	-
longitude	Não foram detetados problemas	-
hourly.data[x].time	Não foram detetados problemas	-
hourly.data[x].summary	Valores a Branco e/ou nulo	Descartar a coluna
hourly.data[x].precipIntensity	De 18 de Maio de 2018 a 20 de Agosto de 2018 este campo é nulo	Considerar o registo anterior
hourly.data[x].precipProbability	De 18 de Maio de 2018 a 20 de Agosto de 2018 este campo é nulo	Considerar o registo anterior
hourly.data[x].temperature	Valores a Branco e/ou nulo	Considerar o registo anterior
hourly.data[x].apparentTemperature	Valores a Branco e/ou nulo	Considerar o registo anterior
hourly.data[x].dewPoint	Não foram detetados problemas	-
hourly.data[x].humidity	Valores a Branco e/ou nulo	Considerar o registo anterior
hourly.data[x].pressure	Valores a Branco e/ou nulo	Considerar o registo anterior
hourly.data[x].windSpeed	Valores a Branco e/ou nulo	Considerar o registo anterior
hourly.data[x].windGust	Valores a Branco e/ou nulo	Considerar o registo anterior
hourly.data[x].windBearing	Valores a Branco e/ou nulo	Considerar o registo anterior
hourly.data[x].cloudCover	Valores a Branco e/ou nulo	Considerar o registo anterior
Continua na próxima página		

<b>Atributo</b>	<b>Problemas detetados</b>	<b>Proposta de resolução</b>
hourly.data[x].uvIndex	Valores a Branco e/ou nulo	Considerar o registo anterior
hourly.data[x].visibility	Valores a Branco e/ou nulo	Considerar o registo anterior

## d) Feriados

**Análise geral** O dataset contém o histórico de feriados da Bélgica e foi adicionado o Iris Day que é um feriado exclusivo de Bruxelas. Desta forma é possível que existam feriados de Bruxelas que não estão a ser considerados uma vez que não são retornados pela API os feriados locais.

A Tabela 20 sintetiza a análise de qualidade efectuada no dataset dos feriados, indicando para cada atributo quais os problemas detetados e as respetivas propostas de resolução.

Tabela 20: Análise de qualidade do dataset dos Feriados

<b>Atributo</b>	<b>Problemas detetados</b>	<b>Proposta de resolução</b>
Year	Não foram detetados problemas	-
holidays[x].date	Não foram detetados problemas	-
holidays[x].start	Não foram detetados problemas	-
holidays[x].end	Não foram detetados problemas	-
holidays[x].name	Não foram detetados problemas	-
holidays[x].type	Não foram detetados problemas	-
holidays[x].public	Não foram detetados problemas	-
holidays[x].country	Não foram detetados problemas	-



## e) Ocupação dos parques

**Análise geral** Tal como explicado na etapa de recolha de dados f), os dados de ocupação foram gerados, o que não se traduz numa solução ideal, porém é a solução possível face às adversidades causadas pela pandemia da Covid-19 na obtenção dos dados.

Na Tabela 21 são apresentados os problemas detetados e as propostas de resolução para os atributos do dataset do histórico de ocupações dos parques de estacionamento.

Tabela 21: Análise de qualidade do dataset das Ocupações

<b>Atributo</b>	<b>Problemas detetados</b>	<b>Proposta de resolução</b>
park_name	Não foram detetados problemas	-
park_id	Não foram detetados problemas	-
date	Não foram detetados problemas	-
dayOfWeek	Não foram detetados problemas	-
timestamp	Não foram detetados problemas	-
occupation	Não foram detetados problemas	-

## 4.2 Preparação dos dados

Esta secção debruçar-se-á sobre todas as decisões e alterações efectuadas aos dados por forma a estes darem suporte ao modelos inteligentes, começando pela seleção dos dados, passando depois para a limpeza, construção e carregamento para a base de dados de estágio, passando ainda pela integração e formatação dos dados e finalizando com uma síntese da recolha, tratamento e carregamento dos dados.

### 4.2.1 Seleção dos dados

Depois de feita a análise dos dados, urge a seleção de quais os dados mais relevantes para a obtenção de modelos preditivos mais eficientes. Duas das formas de seleção de dados consiste na eliminação de linhas e eliminação de colunas. Na etapa anterior, na fase de verificação de qualidade dos dados, para alguns dos problemas detetados foi definido que a proposta de resolução do problema passaria pela

eliminação das linhas. Desta forma, nesta fase o foco será determinar quais os atributos a considerar e a descartar.

## a) Parques de estacionamento

Na Tabela 22, apresentada de seguida, é possível verificar para cada atributo se o mesmo será ou não utilizado e qual o motivo para ser utilizado (ponto verde) ou descartado (ponto vermelho).

Tabela 22: Seleção dos dados no dataset do dataset dos parques de estacionamento

<b>Atributo</b>	<b>Utilização</b>	<b>Motivo</b>
parkingRecord[n].parkingName	●	Este atributo será utilizado como forma de perceber o impacto do grupo de parques a que um parque pertence na taxa de ocupação
parkingRecord[n].operator	●	Este objeto será descartado uma vez que contempla diversos dados considerados pouco relevantes e com baixo impacto na variável target
parkingRecord[n].parkingSite[x].@id	●	Esta coluna será descartada pelo facto de ser um identificador do sistema
parkingRecord[n].parkingSite[x].parkingName	●	Este atributo será utilizado pelo facto do nome ser um atributo único e ser possível assim correlacionar o parque com a taxa de ocupação
parkingRecord[n].parkingSite[x].parkingNumberOfSpaces	●	Este atributo será utilizado pelo facto de ser um fator decisivo para a determinação da taxa de ocupação
parkingRecord[n].parkingSite[x].responsibleAuthority	●	Este objeto será descartado uma vez que contempla diversos dados considerados pouco relevantes e com baixo impacto na variável target
parkingRecord[n].parkingSite[x].operator	●	Este objeto será descartado uma vez que contempla diversos dados considerados pouco relevantes e com baixo impacto na variável target
Continua na próxima página		

<b>Atributo</b>	<b>Utilização</b>	<b>Motivo</b>
parkingRecord[n].parkingSite[x].parkingLocation	●	Este atributo será utilizado por forma a garantir a consistência relacional entre os datasets
parkingRecord[n].parkingSite[x].onlyAssignedParking	●	Este atributo será utilizado com o intuito de perceber qual o impacto das restrições de acesso na taxa de ocupação do parque
parkingRecord[n].parkingSite[x].tariffsAndPayment	●	Este atributo será utilizado para perceber qual o impacto do preço do parque na taxa de ocupação
parkingRecord[n].parkingSite[x].parkingEquipmentOrServiceFacility	●	Objeto utilizado por forma a perceber qual a correlação entre as condições que o parque oferece e a taxa de ocupação
parkingRecord[n].parkingSite[x].groupOfParkingSpaces	●	Atributo utilizado por forma a perceber qual o impacto da percentagem de números de lugares disponíveis para pessoas com deficiência na taxa de ocupação dos parques
parkingRecord[n].parkingSite[x].parkingSiteAddress	●	Objeto descartado por ser meramente informacional
parkingRecord[n].parkingSite[x].openingTimes	●	Atributo utilizado uma vez que o horário de funcionamento afeta diretamente a taxa de ocupação
parkingRecord[n].parkingSite[x].parkingAccess	●	Atributo descartado por ser meramente informacional e com pouca relevância na variável target
parkingRecord[n].parkingSite[x].urbanParkingSiteType	●	Este atributo será descartado pelo facto de todos os parques serem do mesmo tipo

## b) Locais

A Tabela que se segue, Tabela 23, apresenta para os vários atributos dos locais, quais os que serão utilizados (ponto verde) e quais os que serão descartados (ponto vermelho) indicando para ambos os casos um motivo.

Tabela 23: Seleção dos dados do dataset dos locais

<b>Atributo</b>	<b>Utilização</b>	<b>Motivo</b>
parkingSite[n].name	●	Atributo meramente informacional pelo que será descartado, uma vez que o nome do parque já está presente no dataset dos parques
parkingSite[n].parkingLocation. latitude	●	Esta coluna será apenas utilizada para fazer o correto mapeamento entre o parque de estacionamento e os locais. Desta forma para os modelos será descartada
parkingSite[n].parkingLocation. longitude	●	Esta coluna será apenas utilizada para fazer o correto mapeamento entre o parque de estacionamento e os locais. Desta forma para os modelos será descartada
parkingSite[n].pointOfInterests[x]. geometry.location.latitude	●	Atributo utilizado uma vez que conjuntamente com a longitude permite identificar o local
parkingSite[n].pointOfInterests[x]. geometry.location.longitude	●	Atributo utilizado uma vez que conjuntamente com a latitude permite identificar o local
parkingSite[n].pointOfInterests[x]. name	●	Atributo que permitirá identificar o local
parkingSite[n].pointOfInterests[x]. types	●	Atributo utilizado com o intuito de perceber qual a correlação entre os tipos de espaços circundantes a um parque e a sua taxa de ocupação
parkingSite[n].pointOfInterests[x]. rating	●	Atributo descartado uma vez que a maioria dos locais não tem rating
parkingSite[n].pointOfInterests[x]. user_ratings_total	●	Atributo descartado uma vez que a maioria dos locais não tem rating
parkingSite[n].pointOfInterests[x]. vacinity	●	Atributo descartado pelo facto de ser meramente informacional, e sem interessa para o projeto em questão

### c) Meteorologia

Na Tabela 24 é possível verificar a seleção dos atributos efectuada no dataset da Meteorologia. Para cada atributo indica se o mesmo é ou não utilizado e qual o motivo para a decisão. O ponto verde significa que o atributo é considerado e em contra partida o ponto vermelho significa que o atributo é descartado.

Tabela 24: Seleção dos dados do dataset da Meteorologia

Atributo	Utilização	Motivo
latitude	●	Atributo descartado pelo facto de ser pouco relevante e ser igual para todos os registos meteorológicos
longitude	●	Atributo descartado pelo facto de ser pouco relevante e ser igual para todos os registos meteorológicos
hourly.data[x].time	●	Atributo utilizado por forma a manter a consistência entre os datasets
hourly.data[x].summary	●	Atributo descartado pelo facto de ser um resumo textual dos vários indicadores meteorológicos
hourly.data[x].precipIntensity	●	Indicador utilizado por forma a perceber o impacto da precipitação nas taxas de ocupação dos parques de estacionamento
hourly.data[x].precipProbability	●	Indicador utilizado por forma a perceber o impacto da precipitação nas taxas de ocupação dos parques de estacionamento
hourly.data[x].temperature	●	Atributo utilizado por forma a perceber qual o impacto da temperatura nas opções dos condutores, no que a parques de estacionamento diz respeito
hourly.data[x].apparentTemperature	●	Atributo utilizado por forma a perceber qual o impacto da temperatura nas opções dos condutores, no que a parques de estacionamento diz respeito
Continua na próxima página		

<b>Atributo</b>	<b>Utilização</b>	<b>Motivo</b>
hourly.data[x].dewPoint	●	Atributo utilizado por forma a perceber qual o impacto do ponto de condensação nas opções dos condutores, no que a parques de estacionamento diz respeito
hourly.data[x].humidity	●	Atributo utilizado por forma a perceber qual o impacto da humidade nas opções dos condutores, no que a parques de estacionamento diz respeito
hourly.data[x].pressure	●	Atributo utilizado por forma a perceber qual o impacto da pressão nas opções dos condutores, no que a parques de estacionamento diz respeito
hourly.data[x].windSpeed	●	Este atributo será utilizado por forma a perceber o impacto do vento nas decisões dos automobilistas na hora de estacionar
hourly.data[x].windGust	●	Este atributo será utilizado por forma a perceber o impacto do vento nas decisões dos automobilistas na hora de estacionar
hourly.data[x].windBearing	●	Este atributo será utilizado por forma a perceber o impacto do vento nas decisões dos automobilistas na hora de estacionar
hourly.data[x].cloudCover	●	Este atributo será utilizado por forma a perceber o impacto da percentagem de nebulosidade nas decisões dos automobilistas na hora de estacionar
hourly.data[x].uvIndex	●	Este atributo será utilizado por forma a perceber o impacto da emissão de raios ultravioleta nas decisões dos automobilistas na hora de estacionar
hourly.data[x].visibility	●	Este atributo será utilizado por forma a perceber o impacto das condições de visibilidade nas decisões dos automobilistas na hora de estacionar

## d) Feriados

A Tabela 25 apresenta, para cada atributo, a decisão de utilização (ponto verde) ou descarte (ponto vermelho) indicando para ambos os casos qual o motivo.

Tabela 25: Seleção dos dados do dataset dos Feriados

<b>Atributo</b>	<b>Utilização</b>	<b>Motivo</b>
Year	●	Atributo descartado uma vez que o ano encontra-se também no atributo date
holidays[x].date	●	Atributo utilizado para perceber se a afluência ao parques de estacionamento sofre alterações nos feriados
holidays[x].start	●	Atributo descartado uma vez que para o que é pretendido e tendo em conta que os feriados tem todos duração de um dia, o campo date é suficiente
holidays[x].end	●	Atributo descartado uma vez que para o que é pretendido e tendo em conta que os feriados tem todos duração de um dia, o campo date é suficiente
holidays[x].name	●	Atributo utilizado por forma a perceber o impacto de um feriado específico nas opções de estacionamento
holidays[x].type	●	Atributo descartado pelo facto de não se considerar relevante para o negócio
holidays[x].public	●	Atributo descartado pelo facto de ser pouco relevante para o negócio
holidays[x].country	●	Atributo descartado pelo facto de ser sempre igual e meramente informacional

## e) Ocupação dos parques

De seguida, na Tabela 26, é possível verificar entre os vários atributos os que foram utilizados (ponto verde) e descartados (ponto vermelho), sendo que em qualquer um dos casos é indicado o motivo.

Tabela 26: Análise de qualidade do dataset das Ocupações

Atributo	Utilização	Motivo
park_name	●	Atributo descartado por ser meramente informacional e estar já considerado no dataset dos parques de estacionamento
park_id	●	Atributo utilizado por forma a garantir a consistência entre os datasets
date	●	Atributo considerado por forma a indicar a que momento temporal se refere a taxa de ocupação
dayOfWeek	●	Atributo utilizado por forma a perceber se o dia da semana afeta a taxa de ocupação dos parques de estacionamento
timestamp	●	Atributo considerado por forma a indicar a que momento temporal se refere a taxa de ocupação
occupation	●	Variável target que que servirá para treinar os modelos

### 4.2.2 Limpeza e construção dos dados e carregamento para base de dados de estágio

Nesta fase do desenvolvimento do projeto de dissertação, o objetivo passou por tentar aumentar a qualidade dos dados, conseguir extrair e filtrar as informações relevantes de cada dataset e efetuar o carregamento dos dados para tabelas de base de dados de estágio. Tal como em grande parte das etapas anteriores, toda esta etapa foi realizada com recurso a scripts de Python e a base de dados selecionada foi MySQL.



## a) Parques de estacionamento

Sendo o dataset dos parques de estacionamento algo complexo, foi necessário algum tratamento da informação para se conseguir extrair informação útil para dar suporte ao projeto. De seguida são apresentadas todas as intervenções feitas nos vários atributos do dataset.

- **Atributo 'tariffsAndPayment'**: Da etapa da qualidade dos dados 4.1.4 concluiu-se que não existia nenhum padrão para o preçário, no que diz respeito à unidade temporal considerada, o que dificulta a extração direta da informação. De seguida são apresentadas três figuras que demonstram a inexistência de um padrão para o preçário do parque.

```
"tariffsAndPayment": {
  "acceptedMeansOfPayment": [
    "cash",
    "paymentCard"
  ],
  "paymentMode": "payPriorToExit",
  "chargeBand": [
    {
      "@id": "F6678352-88FF-4177-9BA5-4F7B945B31FA",
      "@version": "2017-05-04T11:48:17.3330000+02:00",
      "chargeCurrency": "eur",
      "chargeBandName": {
        "values": {
          "value": [
            {
              "@lang": "fr",
              "#text": "Par heure"
            },
            {
              "@lang": "nl",
              "#text": "Per uur"
            }
          ]
        }
      }
    }
  ],
  "charge": {
    "charge": "1.5"
  }
},
```

Figura 37: Exemplo 1 do atributo 'tariffsAndPayment'

```
"tariffsAndPayment": {
  "acceptedMeansOfPayment": [
    "cash",
    "paymentCard"
  ],
  "paymentMode": "payPriorToExit",
  "chargeBand": [
    {
      "@id": "466FAAE6-3740-47FF-B1CC-F37274D97B9F",
      "@version": "2017-05-04T12:06:27.5170000+02:00",
      "chargeCurrency": "eur",
      "chargeBandName": {
        "values": {
          "value": [
            {
              "@lang": "fr",
              "#text": "Par 50 min."
            },
            {
              "@lang": "nl",
              "#text": "Per 50 min."
            }
          ]
        }
      },
      "charge": {
        "charge": "2"
      }
    }
  ],
}
```

Figura 38: Exemplo 2 do atributo 'tariffsAndPayment'

```

"tariffsAndPayment": {
  "acceptedMeansOfPayment": "paymentCard",
  "chargeBand": [
    {
      "@id": "37F172E5-C4E0-4414-A8E3-F275FBE379F6",
      "@version": "2018-04-17T22:03:21.0130000+02:00",
      "chargeCurrency": "eur",
      "chargeBandName": {
        "values": {
          "value": [
            {
              "@lang": "fr",
              "#text": "Premières 15 minutes sont gratuites"
            },
            {
              "@lang": "nl",
              "#text": "Eerste 15 min zijn gratis"
            }
          ]
        }
      },
      "charge": {
        "charge": "0"
      }
    },
    {
      "@id": "DFDB7498-9A39-413B-AAC1-F90B06E5C3F8",
      "@version": "2018-04-17T22:03:21.0130000+02:00",
      "chargeCurrency": "eur",
      "chargeBandName": {
        "values": {
          "value": [
            {
              "@lang": "fr",
              "#text": "Lundi / Samedi (04:00 - 09:00) : par heure"
            },
            {
              "@lang": "nl",
              "#text": "Maandag / Zaterdag (04:00 - 09:00) : per uur"
            }
          ]
        }
      },
      "charge": {
        "charge": "1"
      }
    }
  ],
}

```

Figura 39: Exemplo 3 do atributo 'tariffsAndPayment'

Desta forma o que se fez foi considerar apenas os parques que tem o preço por hora e com a unidade temporal e pelo menos 1 destas formatações 'Par heure', '1 heure' e '1h' e gerou-se um novo campo denominado **'hourly\_price'** e apenas este foi carregado, descartando-se agora o atributo 'tariffsAndPayment' (que é um objeto). A destacar que daqui resultam 49 parques que contém o preço no formato indicado anteriormente.

- **Atributo 'openingTimes'** Deste atributo foram gerados dois novos denominados **'openTime'** e **'closeTime'** que são respetivamente a hora de abertura e de fecho do parque. Com isto o atributo 'openingTimes' (que é um objeto) é substituídos por estes.
- **Atributo 'onlyAssignedParking'**: Deste objeto foram criados três novos atributos denominados

'**max\_vehicle\_height**', '**max\_vehicle\_width**' e '**max\_vehicle\_length**' que representam os limites máximos de dimensões dos veículos autorizados a entrar no respetivo parque. Com isto, também o objeto 'onlyAssignedParking' é descartado.

- **Atributo 'groupOfParkingSpaces'**: Deste objeto foi criado um novo campo denominado '**handicapped\_places**' que representa o número de lugares disponíveis para deficientes e desta forma o atributo 'groupOfParkingSpaces' deixa de ter relevância.
- **Atributo 'parkingEquipmentOrServiceFacility'**: Deste atributo foram criadas 17 novas colunas que servirão como variáveis *dummy* e que indicam se o parque possui ou não determinado equipamento ou serviço. Com isto, também o objeto 'parkingEquipmentOrServiceFacility' é descartado.

## b) Locais

Este dataset, embora tenha uma estrutura mais simples comparativamente ao dos parques de estacionamento, também foi necessário aplicar algum tratamento e filtro nos dados, os quais são apresentados de seguida.

- **Atributo 'types'**: Sendo este atributo um array que pode ter uma dimensão variável, em que cada elemento do array indica uma tipologia para o local em questão, foram criadas 4 colunas 'type1', 'type2', 'type3' e 'type4', que corresponde aos 4 primeiros tipos do array. Caso o array não tenha 4 ou mais elementos as colunas criadas por excesso assumem valor nulo.

## c) Meteorologia

No dataset da meteorologia não foi necessário nenhum tratamento específico nesta fase, a forma de lidar com os valores nulos é explicada posteriormente.

## d) Feriados

Neste dataset apenas foi necessário garantir que as datas estivessem no formato correto por forma a garantir uma correta inserção das mesmas na base de dados.

## e) Ocupação dos parques

Tal como referido anteriormente, foram gerados dados para 10 parques de estacionamento. Nesta fase de limpeza e construção dos dados antes do carregamento da base de dados, foi necessário filtrar estas ocupações, considerando apenas os registos que estivessem no intervalo do horário de funcionamento do parque.

## f) Esquema da base de dados de estágio

Esta etapa de limpeza e construção terminou com o carregamento dos dados para tabelas de estágio, que permitem uma melhor integração numa fase posterior. Esta base de dados de estágio contempla 5 tabelas sendo elas:

- parks - tabela que contém os dados dos parques;
- places - tabela que armazena os 60 espaços circundantes a cada parque;
- meteorologies - tabela que contém o histórico das condições meteorológicas para Bruxelas;
- holidays - tabela que contém o histórico de feriados de Bruxelas;
- occupations - tabela que armazena o histórico de ocupações;

As tabelas 'parks', 'places' e 'occupations' relacionam-se com recurso a chaves estrangeiras. As restantes tabelas não se encontram relacionadas. As tabelas 'holidays' e 'meteorologies' não estão relacionadas com a tabela 'occupations' uma vez que para efectuar a ligação seria necessário efectuar uma iteração por cada ocupação nas tabelas, e uma vez que esta é uma base de dados de estágio optou-se por não fazer, e efectuar as junções com *scripts* de Python.

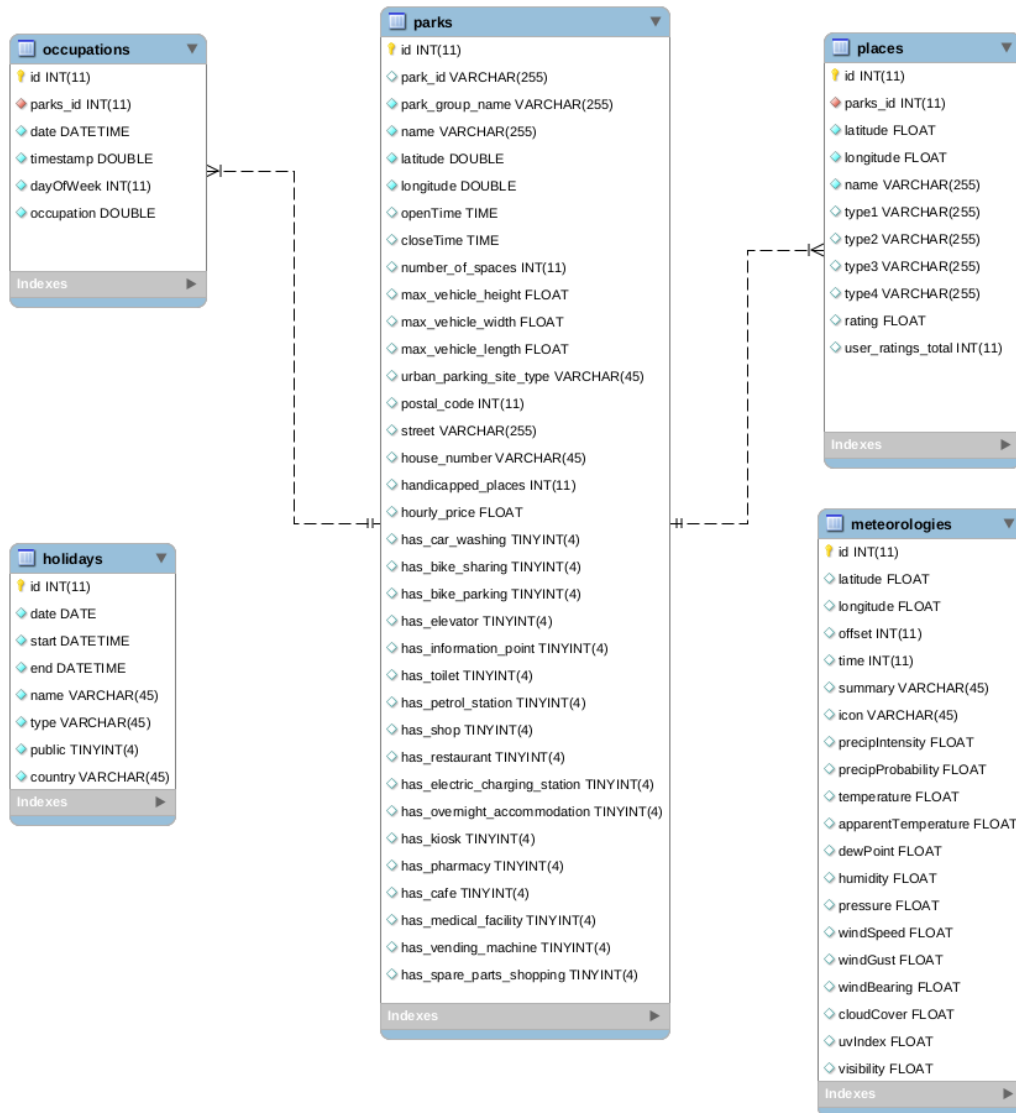


Figura 40: Diagrama de entidades e relacionamentos da base de dados de estágio

É de denotar ainda que, embora se tenha gerado apenas ocupações para 10 parques por motivos já anteriormente explicados, todos os 49 parques resultantes desta etapa fora carregados. Assim sendo, caso se pretenda gerar dados para mais parques, apenas tem que se adicionar os parques à *script* de geração, ficando tudo o resto integro.

### 4.2.3 Integração e Formatação de dados

Nesta fase do projeto, o objetivo passou por integrar todas as informações por forma a ter a informação centralizada. As tabelas de estágio anteriormente descritas serviram de suporte para esta integração e com recurso a uma *script* de Python fez-se a junção da informação todas as tabelas.

A tabela das ocupações serviu de base para a integração, uma vez que é esta que contém a informação que necessita de ser complementada por forma a dar suporte aos algoritmos de inteligência e é a tabela que tem menor granularidade. Iterou-se nesta percorrendo todas as linhas e fazendo as respetivas junções com as diferentes tabelas. De realçar que para cada ocupação foi criada uma nova coluna 'isWeekend' do tipo booleano, que indica se determinada ocupação se registou num feriado ou não, com recurso ao atributo 'dayOfWeek'. Também a data deu origem a 3 novas colunas 'day', 'month' e 'year' que contém respetivamente o dia, o mês e o ano em que se registou determinada taxa de ocupação.

A junção da tabela dos parques de estacionamento, tabela 'parks', foi possível com recurso à chave estrangeira que a tabela das ocupações possui, 'park\_id'.

Uma vez que a tabela dos locais, tabela 'places', possui um elevado número de locais circundantes para cada parque e existem algumas colunas que são meramente informativas (latitude e longitude por exemplo), o que se fez foi integrar apenas qual o tipo1 dos primeiros 5 locais (recorde-se que estes estão ordenados por relevância). Daqui resultaram 5 novas colunas 'place1Type', 'place2Type', 'place3Type', 'place4Type', 'place5Type'. Desta forma, é torna-se possível perceber qual o impacto dos tipos de locais circundantes na taxa de ocupação de um determinado parque.

Para integrar a tabela dos feriados, 'holidays', gerou-se uma nova coluna denominada 'isHoliday' do tipo booleano, que determina se uma ocupação ocorreu num dia de feriado ou não. Para popular esta coluna, em cada ocupação foi verificada se a data coincide com alguma data presente na tabela dos feriados.

A tabela da meteorologia foi integrada fazendo a correspondência da coluna 'time' da tabela da meteorologia com a coluna 'date' da tabela das ocupações.

Por fim, tendo todos os dados integrados, foi criada uma tabela na base de dados denominada 'final\_data' que contém todos os dados integrados, sendo então esta a tabela que armazena os dados que darão suporte aos algoritmos de Machine Learning. A descrição da mesma é apresentada de seguida, na Tabela 27, na qual são enumerados os vários atributos e o tipo de dados.

Tabela 27: Descrição da tabela de base de dados 'final\_data'

<b>Atributo</b>	<b>Tipo</b>
id	INT(11)
parkName	VARCHAR(255)
data	DATETIME
day	INT(11)
month	INT(11)
year	INT(11)
hour	INT(11)
weekDay	INT(11)
occupancy	Double
isWeekend	Boolean
isHoliday	Boolean
precipIntensity	Double
precipProbability	Double
temperature	Double
apparentTemperature	Double
dewPoint	Double
humidity	Double
pressure	Double
windSpeed	Double
windGust	Double
windBearing	Double
cloudCover	Double
uvIndex	Double
visibility	Double
place1Type	VARCHAR(45)
place2Type	VARCHAR(45)
place3Type	VARCHAR(45)
place4Type	VARCHAR(45)
Continua na próxima página	



<b>Atributo</b>	<b>Tipo</b>
place5Type	VARCHAR(45)
parkingGroupName	VARCHAR(45)
numberOfSpaces	INT(11)
handicappedPlaces	INT(11)
hourlyPrice	Double
has_car_washing	Boolean
has_bike_sharing	Boolean
has_bike_parking	Boolean
has_elevator	Boolean
has_information_point	Boolean
has_toilet	Boolean
has_petrol_station	Boolean
has_shop	Boolean
has_restaurant	Boolean
has_electric_charging_station	Boolean
has_kiosk	Boolean
has_pharmacy	Boolean
has_cafe	Boolean
has_medical_facility	Boolean
has_vending_machine	Boolean
has_spare_parts_shopping	Boolean

Na figura que se segue é apresentado um esquema que permite perceber a ordem de concretização das integrações de dados efectuada.

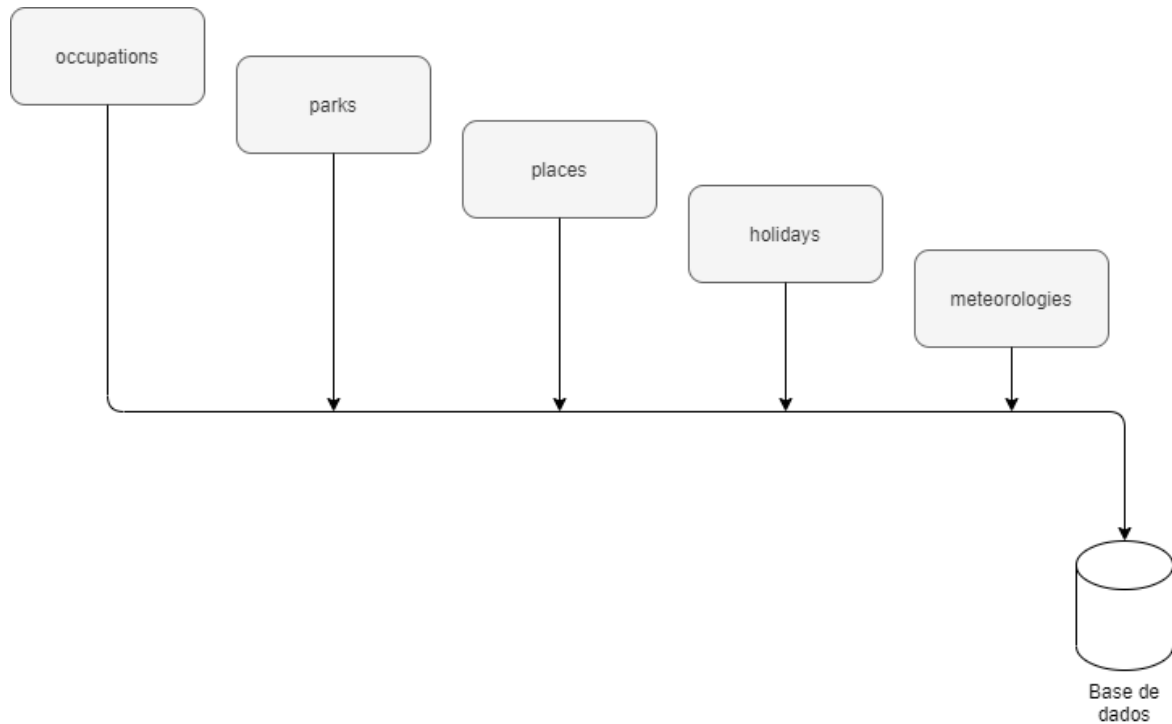


Figura 41: Esquema da etapa de integração de dados

#### 4.2.4 Síntese da recolha, tratamento e carregamento dos dados

A etapa de integração e formatação dos dados finaliza o capítulo de preparação dos dados. De seguida é apresentado um esquema que permite sintetizar a recolha, o tratamento e o carregamento de dados efectuado.

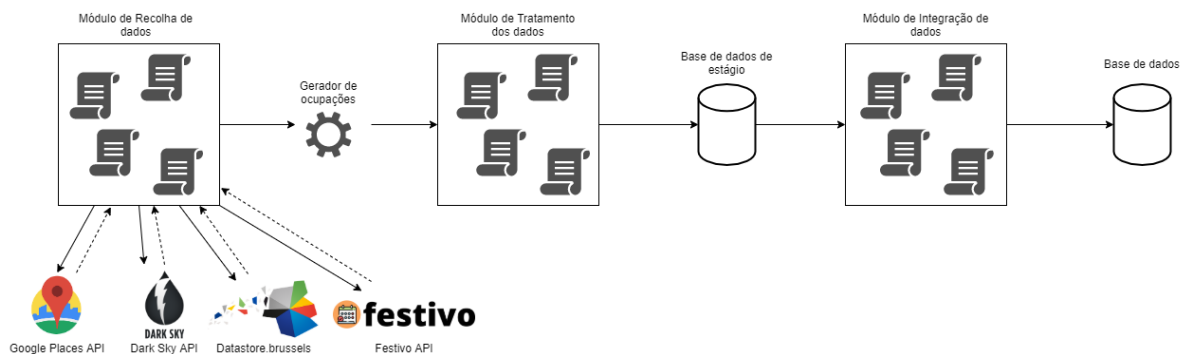


Figura 42: Esquema da recolha, tratamento e carregamento dos dados

## 4.3 Modelação

Nesta fase, o objetivo passa por selecionar, aplicar e otimizar vários modelos por forma a obter melhores preditores que possibilitarão a previsão da variável target de uma forma mais eficiente possível. Nesta secção serão abordadas então as tarefas realizadas e decisões tomadas para melhor garantir a concretização do objetivo anteriormente enunciado.

### 4.3.1 Seleção das técnicas de modelação

O primeiro passo na modelação é selecionar as técnicas a serem utilizadas. Para que esta seleção seja adequada definiu-se claramente qual a variável target ou variável dependente, e aqui está claro que a taxa de ocupação é o que se pretende prever, e estando a taxa de ocupação associada a um tempo, então este é um problema de previsão de uma série temporal.

Depois de definida a target foi necessário decidir qual a abordagem a aplicar, se regressão ou classificação, uma vez que esta decisão tem impacto nos modelos possíveis de serem utilizados. Embora a variável target seja uma variável contínua, optou-se por utilizar classificação e assim em vez de prever uma taxa de ocupação, o objetivo passará por prever uma classe que corresponderá a um intervalo de ocupações, e decidiu-se que existiriam 10 classes:

- classe 0: ]-0.1, 10]
- classe 1: ]10, 20]
- classe 2: ]20, 30]
- classe 3: ]30, 40]
- classe 4: ]40, 50]
- classe 5: ]50, 60]
- classe 6: ]60, 70]
- classe 7: ]70, 80]
- classe 8: ]80, 90]
- classe 9: ]90, 100]

Sendo este um problema de Classificação multiclasse, os algoritmos mais populares são k-Nearest Neighbors, Decision Trees, Naive Bayes, Random Forest e Gradient Boosting. O algoritmo Support Vector Machine também pode ser adaptado através da criação de múltiplas classificações binárias (Brownlee, 2020). Contudo como este algoritmo não é o mais eficiente quando existe um elevado número de variáveis explicativas ou independentes, foi descartado. Desta forma os algoritmos selecionados foram os seguintes:

- k-Nearest Neighbors;
- Naive Bayes;
- Decision Trees;
- Random Forest;
- Gradient Boosting;

### 4.3.2 Criação de cenários de teste

Antes de proceder à construção dos modelos é importante definir quais os cenários a serem testados. Desta forma foram desenvolvidos sete cenários sendo eles:

- **cenário 1:** Cenário personalizado, foram adicionadas as variáveis que segundo a literatura tem mais impacto na taxa de ocupação dos parques;
- **cenário 2:** Cenário no qual foram consideradas todas as variáveis sem exceção;
- **cenário 3:** Cenário no qual não foram consideradas as variáveis independentes associadas à meteorologia;
- **cenário 4:** Cenário no qual não foram consideradas as características do parque como o preço, o número de lugares e os restantes;
- **cenário 5:** Cenário no qual foram excluídos os dados relativos à tipologia dos espaços circundantes ao parque de estacionamento;
- **cenário 6:** Cenário no qual não foram considerados os feriados;
- **cenário 7:** Cenário no qual foram consideradas todas as variáveis exceto os equipamentos e serviços disponibilizados;

Por forma a facilitar a verificação de quais as variáveis utilizadas para os vários cenários, segue-se uma Tabela na qual os cenários foram colocados nas colunas e as diversas variáveis nas linhas. No cruzamento entre cada variável e cenário, se tiver preenchido com um marcador, significa que a variável é utilizada no respetivo cenário.

Tabela 28: Descrição da tabela de base de dados 'final\_data'

Cenários Variáveis	1	2	3	4	5	6	7
parkName	X	X	X	X	X	X	X
day	X	X	X	X	X	X	X
month	X	X	X	X	X	X	X
year	X	X	X	X	X	X	X
hour	X	X	X	X	X	X	X
weekDay	X	X	X	X	X	X	X
occupancy	X	X	X	X	X	X	X
isWeekend	X	X	X	X	X	X	X
isHoliday	X	X	X	X	X		X
precipIntensity	X	X		X	X	X	X
precipProbability	X	X		X	X	X	X
temperature	X	X		X	X	X	X
apparentTemperature	X	X		X	X	X	X
dewPoint		X		X	X	X	X
humidity		X		X	X	X	X
pressure		X		X	X	X	X
windSpeed	X	X		X	X	X	X
windGust		X		X	X	X	X
windBearing		X		X	X	X	X
cloudCover		X		X	X	X	X
uvIndex		X		X	X	X	X
visibility		X		X	X	X	X

Continua na próxima página

Cenários Variáveis	1	2	3	4	5	6	7
place1Type	X	X	X	X		X	X
place2Type	X	X	X	X		X	X
place3Type	X	X	X	X		X	X
place4Type	X	X	X	X		X	X
place5Type	X	X	X	X		X	X
parkingGroupName	X	X	X		X	X	X
numberOfSpaces	X	X	X		X	X	X
handicappedPlaces		X	X		X	X	X
hourlyPrice	X	X	X		X	X	X
has_car_washing		X	X		X	X	
has_bike_sharing		X	X		X	X	
has_bike_parking		X	X		X	X	
has_elevator		X	X		X	X	
has_information_point		X	X		X	X	
has_toilet		X	X		X	X	
has_petrol_station		X	X		X	X	
has_shop		X	X		X	X	
has_restaurant		X	X		X	X	
has_electric_charging_station		X	X		X	X	
has_overnight_accommodation		X	X		X	X	
has_kiosk		X	X		X	X	
has_pharmacy		X	X		X	X	
has_cafe		X	X		X	X	
has_medical_facility		X	X		X	X	
has_vending_machine		X	X		X	X	
has_spare_parts_shopping		X	X		X	X	

Uma vez que foram definidos 7 cenários e seleccionados 5 modelos, no total foram induzidos 35 modelos.

### 4.3.3 Construção do modelo

Na fase da construção do modelo foram tomadas decisões e definidos um conjunto de configurações que importa apresentar.

Os modelos foram construídos em Python com recurso à biblioteca Sklearn, importando sempre a classe respetiva ao modelo que se pretendia implementar. A fonte de dados para treinar os modelos foi a tabela 'final\_data' (4.2.3), e para efectuar a conexão à base de dados MySQL foi utilizada a biblioteca SQLAlchemy que permitiu a criação de uma conexão e efectuar queries. Como foi referido na etapa anterior, os vários cenários utilizam diferentes conjuntos de variáveis, e para os vários cenários foram construídos os modelos seleccionados. Foi criada uma lista que contém as variáveis a utilizar em cada cenário, que permite restringir as colunas da base de dados a considerar.

Por forma a manipular mais facilmente os dados, foi importada a biblioteca Pandas. A função cut desta biblioteca permitiu fazer a associação das ocupações para as respetivas classes anteriormente definidas (4.3.1). Foram passados dois parâmetros, qual a coluna do dataframe a utilizar e em quantos elementos se pretende dividir. Esta função permitiu criar uma nova coluna denominada 'class' que, para cada registo de ocupação, definiu a classe a que este pertence, mais tarde fez-se a categorização desta coluna com recurso à função LabelEncoder do Sklearn, resultando daqui a coluna class\_code que passou a ser a variável target. Posto isto, a coluna 'occupancy' e 'class' foram removidas.

Para efectuar a divisão do dataset em treino e teste foi utilizado cross\_validation (classe KFold do Sklearn). Foram efectuados dois testes para decidir qual o fold a utilizar, 10 e 2. Como se verificou que a variação nos resultados para este conjunto de dados foi insignificante, utilizou-se 2 folds por uma questão de desempenho.

Antes de entrar nas configurações de cada modelo, importa destacar que foram criadas variáveis dummy, por forma a refletir de forma numérica os valores das variáveis categóricas. Para isto, utilizou-se a função get\_dummies da biblioteca Pandas.

Por forma a ter mais capacidade computacional, recorreu-se serviço da Google que permite a computação na cloud. Desta forma foi utilizada uma máquina da Google com 64GB de RAM e 16 vCPUs.

De seguida, são apresentadas as várias configurações definidas e utilizadas para cada modelo.

## a) K-Nearest Neighbors

Para implementar o algoritmo K-Nearest Neighbors recorreu-se à utilização da classe `KNeighborsClassifier` do Sklearn. Por forma a considerar a distância euclidiana entre pontos, definiu-se o parâmetro 'metric' com o valor 'minkowski' e com parâmetro  $p=2$ .

O parâmetro que se tentou afinar foi o número de neighbors (parâmetro 'n\_neighbors'). Testou-se com 5, 10 e 20. Verificou-se diferenças nos resultados eram residuais, ou seja os resultados, para este conjunto de dados, foram idênticos nos 3 casos testados.

## b) Naive Bayes

Recorreu-se à classe `GaussianNB` do módulo `naive_bayes` da biblioteca Sklearn para implementar o algoritmo de Naive Bayes. Embora geralmente este algoritmo apresente boa precisão, é limitado no que diz respeito às afinações. Isto levou a que se utilizassem os parâmetros padrão.

## c) Decision Trees

Para efectuar a classificação utilizando árvores de decisão, recorreu-se à classe `DecisionTreeClassifier` do módulo `Tree` da biblioteca Sklearn. No que diz respeito a afinações, uma que pode ter impacto é a fórmula ou critério que define as ramificações da árvore. O parâmetro que tem impacto nisto é o 'criterion' e existem duas possibilidades de definição para este parâmetro 'gini' ou 'entropy'. Verificou-se que para o conjunto de dados utilizado o 'entropy' é o que apresenta melhores resultados.

Embora existam outros parâmetros que geralmente tem impacto nos resultados, para este conjunto de dados o impacto foi residual. Isto certamente estará relacionado com o facto dos dados terem sido gerados. Posto isto foram utilizados os valores por defeito.



## d) Random Forest

Para implementar o algoritmo Random Forest, e uma vez que este é um problema de classificação, importou-se a classe `RandomForestClassifier` do módulo `ensemble` do `Sklearn`. Também para este algoritmo, e uma vez que acaba por estar relacionado com o anterior, foi definido o critério 'entropy' para a definição das ramificações das árvores, isto porque foi o que apresentou melhores resultados. Também o número de árvores a utilizar foi afinada. Começou-se com 10, foi-se aumentando e analisando os resultados. Concluiu-se que 100 seria o número mais acertado, uma vez que a partir daqui os ganhos não foram substanciais.

## e) Gradient Boosting

Para a utilização do algoritmo Gradient Boosting, importou-se a classe `GradientBoostingClassifier` do módulo `ensemble` do `Sklearn`. À medida que se foi tentando afinar os parâmetros do algoritmo, verificou-se que os ganhos eram pouco significativos. Desta forma optou-se por utilizar os valores por defeito.

### 4.3.4 Avaliação do modelo

Após a construção dos modelos, estes foram executados para cada cenário. No final da execução foram registadas e anotadas as métricas que permitem avaliar a qualidade do modelo. Como este se trata de um problema de classificação as métricas registadas foram a acuidade, a precisão, o recall e f1-score.

Segundo Purva Huilgol (Huilgol, 2019) o f1-score é uma boa métrica a considerar para casos onde temos classes não balanceadas. Como podemos ver na lista que se segue, as classes não estão balanceadas uma vez que por exemplo a classe 0 tem apenas 1034 registos ( $\approx 0.28\%$ ), e a classe 6 tem 72348 registos ( $\approx 19.63\%$ ).

- classe 0: 1034 registos ( $\approx 0.28\%$ ).
- classe 1: 3384 registos ( $\approx 0.92\%$ ).
- classe 2: 10233 registos ( $\approx 2.78\%$ ).
- classe 3: 24281 registos ( $\approx 6.59\%$ ).

- classe 4: 45181 registos ( $\approx 12.26\%$ ).
- classe 5: 64602 registos ( $\approx 17.52\%$ ).
- classe 6: 72348 registos ( $\approx 19.63\%$ ).
- classe 7: 64364 registos ( $\approx 17.46\%$ ).
- classe 8: 44587 registos ( $\approx 12.09\%$ ).
- classe 9: 38629 registos ( $\approx 10.48\%$ ).

Por forma a dar mais peso ao f1-score face à acuidade e uma vez que os dados foram gerados tendo em conta uma média e um desvio padrão, foi desenvolvida uma métrica dinâmica,  $m$ , e que tem em atenção a distribuição dos dados.

$$m = \text{acuidade} \times (50 - p)\% + \text{f1-score} \times (50 + p)\%$$

em que  $p$  é obtido em função do valor da média ( $\mu$ ) e do desvio padrão ( $\sigma$ ) do gerador dos dados.

$$\mu \text{ — } 50$$

$$\sigma \text{ — } p$$

Para o conjunto de dados em questão, com  $\mu = 65$  e  $\sigma = 20$ ,

$$p = \frac{20 \times 50}{65} \iff p = 15.38$$

De seguida são apresentados os resultados obtidos, no formato de tabela em que é possível verificar quais os resultados das métricas para determinado algoritmo nos vários cenários. Importa realçar que a última coluna das tabelas consiste no valor calculado para a métrica desenvolvida  $m$ .

Tabela 29: Resultados obtidos com o algoritmo K-Nearest Neighbors

	<b>K-Nearest Neighbors</b>				
	<b>Acuidade</b>	<b>Precisão</b>	<b>Recall</b>	<b>F1 Score</b>	<b>m</b>
<b>Cenário 1</b>	0,1442	0,1	0,1	0,1	0,1153
<b>Cenário 2</b>	0,1444	0,1	0,1	0,1	0,1154
<b>Cenário 3</b>	0,1447	0,1	0,1	0,1	0,1155
<b>Cenário 4</b>	0,1458	0,1	0,1	0,1	0,1159
<b>Cenário 5</b>	0,1444	0,1	0,1	0,1	0,1154
<b>Cenário 6</b>	0,1444	0,1	0,1	0,1	0,1154
<b>Cenário 7</b>	0,1444	0,1	0,1	0,1	0,1154

Tabela 30: Resultados obtidos com o algoritmo Naive Bayes

	<b>Naive Bayes</b>				
	<b>Acuidade</b>	<b>Precisão</b>	<b>Recall</b>	<b>F1 Score</b>	<b>m</b>
<b>Cenário 1</b>	0,1781	0,09	0,1	0,07	0,1074
<b>Cenário 2</b>	0,1707	0,08	0,1	0,07	0,1049
<b>Cenário 3</b>	0,1826	0,08	0,1	0,06	0,1024
<b>Cenário 4</b>	0,1779	0,08	0,1	0,06	0,1008
<b>Cenário 5</b>	0,1725	0,08	0,1	0,07	0,1055
<b>Cenário 6</b>	0,1701	0,08	0,1	0,07	0,1047
<b>Cenário 7</b>	0,178	0,09	0,1	0,07	0,1074

Tabela 31: Resultados obtidos com o algoritmo Decision Trees

	<b>Decision Trees</b>				
	<b>Acuidade</b>	<b>Precisão</b>	<b>Recall</b>	<b>F1 Score</b>	<b>m</b>
<b>Cenário 1</b>	0,1481	0,1	0,1	0,1	0,1167
<b>Cenário 2</b>	0,1468	0,1	0,1	0,1	0,1162
<b>Cenário 3</b>	0,1459	0,1	0,1	0,1	0,1159
<b>Cenário 4</b>	0,1463	0,1	0,1	0,1	0,1160
<b>Cenário 5</b>	0,1463	0,1	0,1	0,1	0,1160
<b>Cenário 6</b>	0,1454	0,1	0,1	0,1	0,1157
<b>Cenário 7</b>	0,1476	0,1	0,1	0,1	0,1165

Tabela 32: Resultados obtidos com o algoritmo Random Forest

	<b>Random Forest</b>				
	<b>Acuidade</b>	<b>Precisão</b>	<b>Recall</b>	<b>F1 Score</b>	<b>m</b>
<b>Cenário 1</b>	0,1552	0,1	0,1	0,1	0,1191
<b>Cenário 2</b>	0,1542	0,1	0,1	0,1	0,1188
<b>Cenário 3</b>	0,1535	0,1	0,1	0,1	0,1185
<b>Cenário 4</b>	0,1477	0,1	0,1	0,1	0,1165
<b>Cenário 5</b>	0,1538	0,1	0,1	0,1	0,1186
<b>Cenário 6</b>	0,1538	0,1	0,1	0,1	0,1186
<b>Cenário 7</b>	0,1542	0,1	0,1	0,1	0,1188

Tabela 33: Resultados obtidos com o algoritmo Gradient Boosting

	<b>Gradient Boosting</b>				
	<b>Acuidade</b>	<b>Precisão</b>	<b>Recall</b>	<b>F1 Score</b>	<b>m</b>
<b>Cenário 1</b>	0,196	0,09	0,1	0,03	0,0875
<b>Cenário 2</b>	0,196	0,1	0,1	0,03	0,0875
<b>Cenário 3</b>	0,1961	0,07	0,1	0,03	0,0875
<b>Cenário 4</b>	0,1961	0,09	0,1	0,03	0,0875
<b>Cenário 5</b>	0,196	0,1	0,1	0,03	0,08745
<b>Cenário 6</b>	0,196	0,1	0,1	0,03	0,0875
<b>Cenário 7</b>	0,1961	0,09	0,1	0,03	0,0875

Analisando os resultados obtidos, podemos concluir que a assertividade dos modelos desenvolvidos é baixa, contudo não se traduz em algo inesperado uma vez que as variáveis explicativas não tem correlação direta com a target devido à forma como as ocupações foram geradas ([secção 4.1.1 f](#))

Considerando a métrica desenvolvida ( $m$ ) por forma a dar mais relevância ao f1-score face à acuidade, conclui-se que o algoritmo Random Forest é o que apresenta melhores resultados, e é com o cenário 1 que a métrica desenvolvida assume os valores mais elevados.

Depois de analisados os resultados, foi necessário exportar o melhor modelo e as colunas utilizadas nesse modelo por forma a conseguir reutilizar posteriormente sem ter que estar sempre a treinar o mesmo. Para efectuar esta exportação recorreu-se ao módulo pickle, que efectua a exportação para um ficheiro que pode ser posteriormente utilizado.

### 4.3.5 Esquema síntese da modelação

De seguida é apresentado um esquema que permite perceber quais os componentes que interagem e resultam do módulo de modelação desenvolvido.

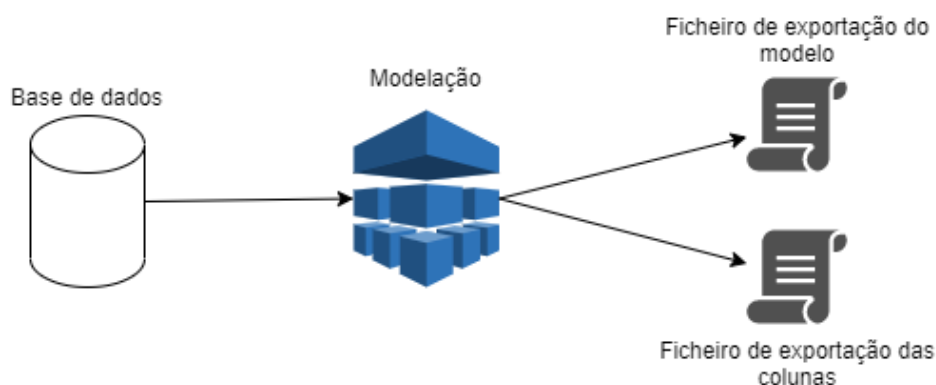


Figura 43: Esquema síntese da modelação

## 4.4 Avaliação

Depois de concluída a modelação é importante efectuar uma avaliação dos modelos desenvolvidos. Desta forma, as subsecções inseridas nesta secção visam atingir melhores modelos no futuro. Posto isto seguem-se a avaliação dos resultados, a revisão do processo e a determinação dos próximos passos.

### 4.4.1 Avaliar resultados

Como pode ser verificado nas métricas apresentadas na [secção 4.3.4](#), a baixa eficiência dos modelos de previsão é um fator que se destaca. Nenhum dos 35 modelos desenvolvidos conseguiu ter, por exemplo, uma acuidade superior a 20%. Isto demonstra que a qualidade dos dados é baixa, e isso pode ser observado pela distribuição dos registos do dataset pelas respectivas classes ([secção 4.3.4](#)), uma vez que, por exemplo, a classe 0 tem apenas aproximadamente 0.28% dos registos, enquanto que a classe 6 tem 19.63%, aproximadamente. Isto faz com que os modelos sejam mais eficientes a prever a classe 6 do que a classe 0.

De destacar que não foi possível definir nenhum treshold uma vez que os dados não davam garantias de sucesso.

### 4.4.2 Rever o processo

À avaliação menos positiva efectuada anteriormente estão associadas razões para tal. As métricas baixas dos modelos de previsão justificam-se pelo facto das ocupações terem sido geradas e desta forma

as variáveis explicativas não tem correlação com a target. Claro está que é também importante complementar os dados com informações de eventos nas proximidades dos parques de estacionamento, coisa que não se afigura fácil uma vez que implica analisar o impacto dos eventos, e não existe uma plataforma que agregue os eventos para todos os países/cidades.

### 4.4.3 Determinar os próximos passos

Após a análise, avaliação dos resultados e revisão do processo, é importante definir quais os próximos passos a seguir. No seguimento do surgimento de alguns entraves ao normal desenrolar do projeto, criam-se oportunidades de melhoria sendo elas:

- **Obter dados de ocupações reais e aumentar os parques considerados em Bruxelas:** Devido aos problemas de obtenção de dados, surge a necessidade de obter o histórico de ocupações reais dos parques considerados no projeto. Muito provavelmente com isto, os modelos de previsão serão mais eficientes. Importante também será aumentar o leque de parques em Bruxelas uma vez que foram apenas considerados 10 parques nos treinos dos modelos devido a limitações de hardware que impacta nos tempos de execução.
- **Criar uma estratégia de implementação e difusão da solução:** Sendo esta uma solução que poderá ser global, é importante a mesma ir crescendo geograficamente de forma gradual, ou seja neste momento a solução encontra-se estável para os parques de Bruxelas considerados, mas é importante definir quais os países e cidades que se pretende abranger a curto/médio prazo por forma a preparar a solução e planear a implementação, isto porque para cada parque é necessário obter dados que possam suportar os modelos.
- **Definição de parceiros estratégicos:** Depois de definido o item anterior, é necessário criar uma rede de parceiros que possibilitem o fornecimento de dados e facilitem a difusão da solução. Estes fornecedores de dados podem ser entidades gestoras de parques, por exemplo, que podem neste projeto ser um stakeholder importante.
- **Obter e considerar dados de eventos nas próximas iterações:** É necessário que para cada região considerada se obtenha dados acerca dos eventos existentes. Esta tarefa é algo complexa uma vez que há a necessidade de efectuar uma análise acerca do impacto do evento, uma vez que existem eventos que não tem impacto relevante nas taxas de ocupação dos parques de estacionamento nas proximidades. O primeiro passo é tentar perceber quais os eventos impactantes

em Bruxelas, e a obtenção de um parceiro com conhecimento de causa poderá ser um auxílio importante no filtro e na obtenção dos dados.

## **4.5 Implementação**

Depois de concluídos e avaliados os modelos desenvolvidos segue-se a implementação que contempla a descrição da API Rest e da aplicação mobile desenvolvidas, as fases de planeamento da implementação, planeamento da monitorização e manutenção, produção do relatório final e revisão do projeto. Estas fases são descritas de seguida.

### **4.5.1 API**

Uma vez terminada a modelação procedeu-se à criação de um mecanismo que permitisse a utilização do melhor modelo preditivo. Para isto recorreu-se à construção de uma API REST por forma a disponibilizar uma forma de utilizar o modelo como um serviço. A API foi construída em Python com recurso à framework Flask.

Na API foi desenvolvida uma rota que tem um método POST que recebe no body um JSON com 3 parâmetros, a latitude e longitude do local para onde se pretende ir e a data de chegada ao mesmo, como pode ser visto na Figura 44.



## POST scenario1\_10parks

```
35.197.237.135:60001/10parks/scenario1/random_forest
```

### BODY raw

```
{
  "latitude": 50.8503,
  "longitude": 4.3517,
  "date": "2020-12-10 05:00:00"
}
```

Figura 44: Documentação API

Através dos ficheiros anteriormente exportados do melhor modelo e das colunas utilizadas no mesmo criou-se um serviço inteligente. Através da iteração nas colunas utilizadas é possível saber quais os dados de input que o modelo necessita.

Importa realçar que uma vez que só existem dados históricos de meteorologia, foi necessário a integração da API Darksy com a finalidade de obter os dados meteorológicos no dia e hora indicados. Com queries à base de dados consegue-se obter os restantes dados e passar para o modelo. O modelo é executado para todos os parques da base de dados por forma a retornar para cada um qual a classe prevista, que depois é mapeado para o respetivo intervalo de ocupação.

Uma vez que a distância do parque de estacionamento ao local de destino é um fator que impacta na decisão de escolha do parque, foi efectuada uma integração com o Google Distances que permite então obter não só a distância como o tempo a pé de cada parque até ao local de destino.

Assim sendo, a rota retorna todos os parques indicando alguns dados do mesmo, qual a taxa de ocupação prevista para o dia e hora indicados e qual a distância e o tempo a pé do parque até ao local

de destino. O critério de seleção dos dados que a API deve retornar baseia-se nos factores que podem impactar a decisão de estacionamento. Segue-se um excerto de um exemplo de resposta retornada pelo serviço.

```
{
  "address": "Avenue Hippocrate",
  "distance": {
    "distance": {
      "text": "5.2 mi",
      "value": 8448
    },
    "duration": {
      "text": "1 hour 50 mins",
      "value": 6616
    }
  },
  "house_number": "11",
  "latitude": 50.852821,
  "longitude": 4.455578,
  "park_group_name": "Q-Park Belgium",
  "park_id": "AE979507-05C2-47C2-91BB-61D2F38E13A5",
  "park_name": "Hippocrate",
  "postal_code": 1200,
  "prediction": "40% - 50%",
  "price": 1.5
},
```

Figura 45: Exemplo resposta da API

Depois de terminado o desenvolvimento da API, disponibilizou-se a mesma numa máquina da Google por forma a permitir o acesso externo através da utilização do IP da máquina.

Na figura que se segue é possível perceber de forma gráfica os componentes que interagem com a API desenvolvida.

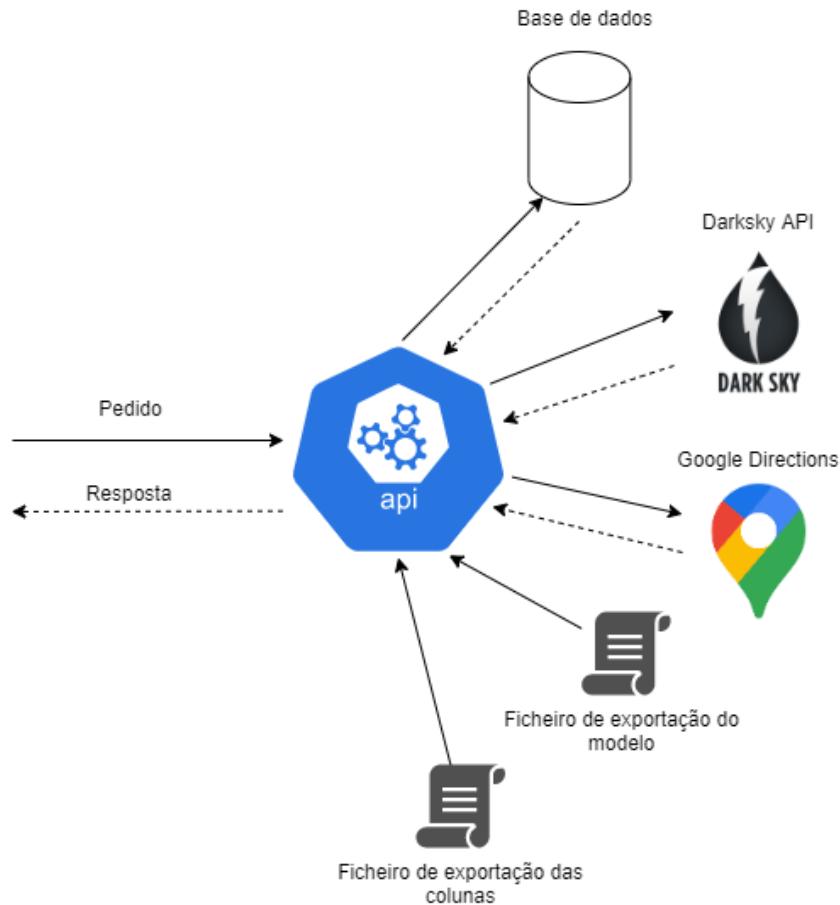


Figura 46: Componentes que interagem com a API

## 4.5.2 Aplicação Mobile

Um dos resultados esperados deste projeto de dissertação passa pelo desenvolvimento de um protótipo funcional. Optou-se pelo desenvolvimento de uma aplicação mobile que permita auxiliar o cidadão na tomada de decisão de estacionamento.

### a) Descrição da aplicação

A aplicação mobile foi desenvolvida em IONIC com Angular. A escolha da tecnologia prende-se com a rapidez de desenvolvimento e com a facilidade em “deploy” tanto para Android como para iOS. Decidiu-se que a aplicação seria denominada de “IOParking” com o *slogan* “Smart Parking for Everyone”, como pode ser visto na Figura 47.



Figura 47: Aplicação - Splash Screen

Depois de entrar na aplicação, é mostrado ao utilizador um campo de texto onde este deve colocar o local para onde se pretende dirigir, como pode ser comprovado nas duas figuras que se seguem (Figura 48 e Figura 49).

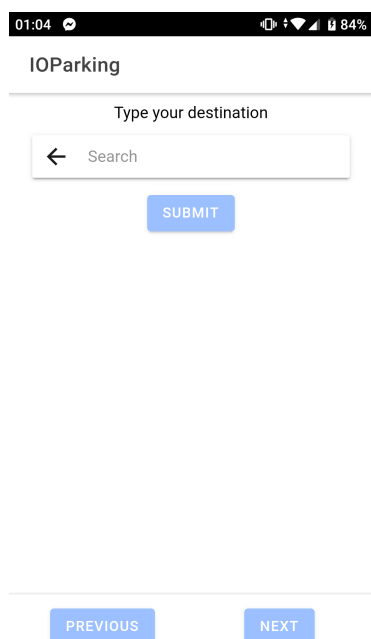


Figura 48: Aplicação - Primeiro ecrã

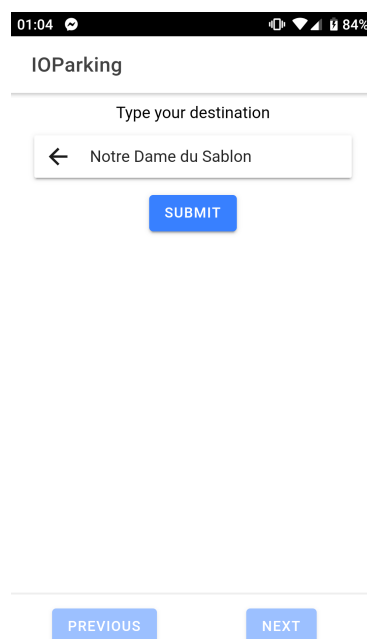


Figura 49: Aplicação - Preenchimento do local

Preenchendo então o local e clicando no botão de submissão, é mostrada uma lista com possíveis locais, considerando o texto inserido. Estes locais são obtidos com recurso à API Google Places, a função “findplacefromtext” que permite obter o locais candidatos a partir de um texto. É através desta API que se consegue obter a latitude e longitude associadas ao destino do utilizador. As duas figuras que se seguem, Figura 50 e Figura 51, são, respetivamente, um exemplo de uma pesquisa efectuada à API e da seleção por parte do utilizador.

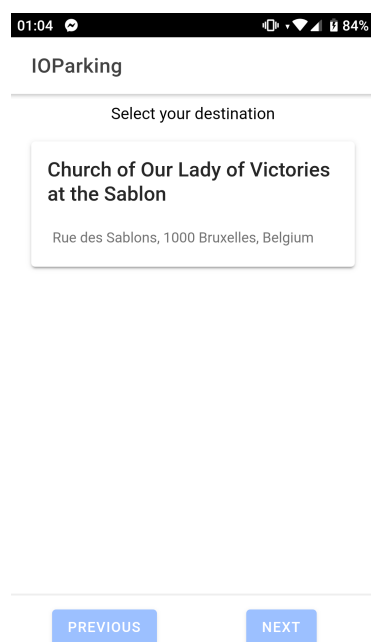


Figura 50: Aplicação - Lista de locais candidatos

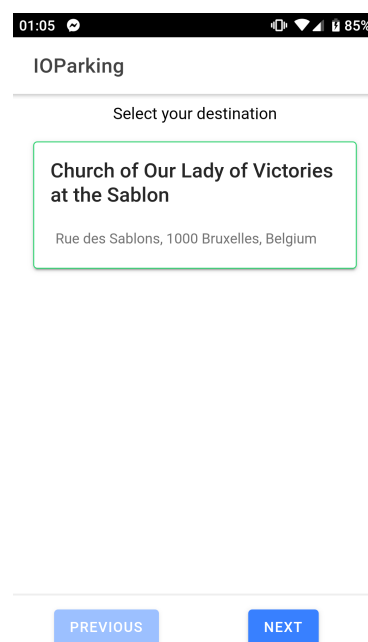


Figura 51: Aplicação - Seleção do Local

Depois de seleccionar o local correto, o utilizador deve indicar qual o dia e a hora de chegada ao mesmo. Importante realçar que se adicionou uma fotografia do local previamente seleccionado para que o utilizador comprove que é o correto, sendo que a fotografia foi extraída do Google Places com a função “photo”. A Figura 52 apresenta um exemplo disto mesmo.

01:05

IOParking

Place Name: Church of Our Lady of Victor...

Address: Rue des Sablons, 1000 Bruxelles...

2021-01-05 11:00

SUBMIT

PREVIOUS GO

Figura 52: Aplicação - Indicação do dia e hora de chegada

Depois de indicado qual o local pretendido e o dia e hora de chegada, é efectuado um pedido à API anteriormente desenvolvida (secção 5.5.1), fornecendo estes mesmos dados, para que esta devolva as previsões de ocupações e os restantes dados (preço, tempo a pé, distância dia parque - local etc) para cada parque aberto no dia e hora indicados. Estes dados retornados pela API são os que têm impacto na decisão de estacionamento.

O informação retornada pela API é mostrada ao utilizador num formato de lista, e com esta o utilizador deve decidir qual o parque que mais satisfaz as suas pretensões (Figura 53).

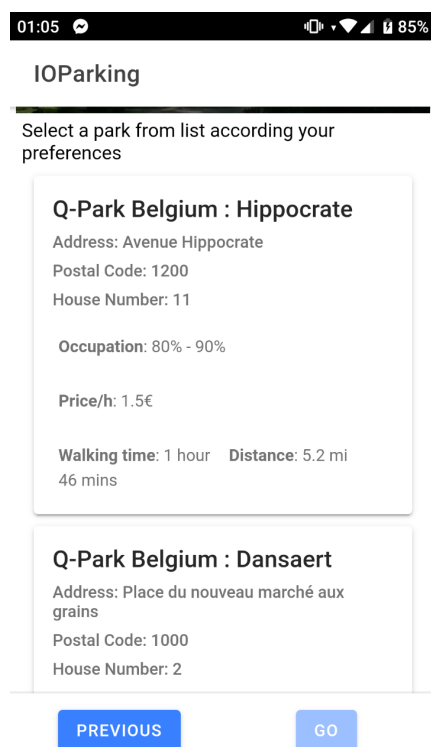


Figura 53: Aplicação - Ecrã de seleção do parque de estacionamento

Escolhendo um parque, é extraída a localização do utilizador e é mostrada a rota desta até ao parque selecionado, indicando o tempo de viagem e custos associados com portagens, caso seja o caso. Tudo isto é obtido com integração serviço Google Directions (Figura 54).

Caso o utilizador pretenda iniciar a navegação, é lançada a aplicação GOOGLE MAPS indicando qual a origem e o destino da rota (Figura 55).

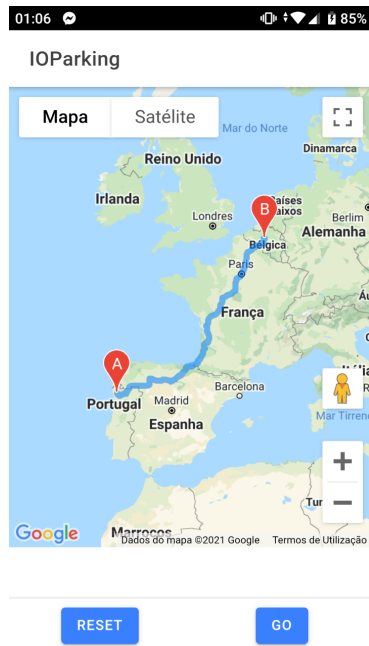


Figura 54: Aplicação - Rota mostrada ao utilizador

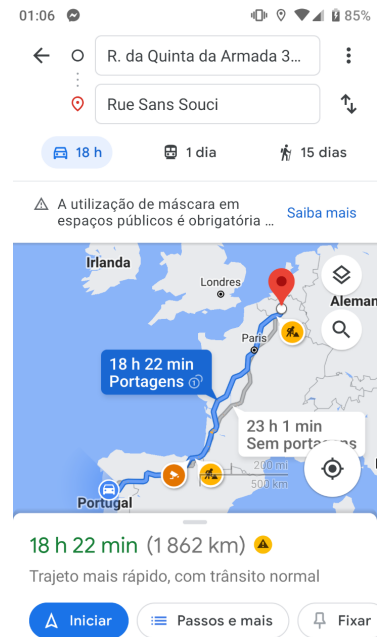


Figura 55: Aplicação - Google Maps para iniciar navegação

## b) Esquema de utilização da aplicação

Na figura que se segue é apresentado um esquema que permite perceber os componentes que interagem com a aplicação e que contribuem para a melhoria da usabilidade da mesma.

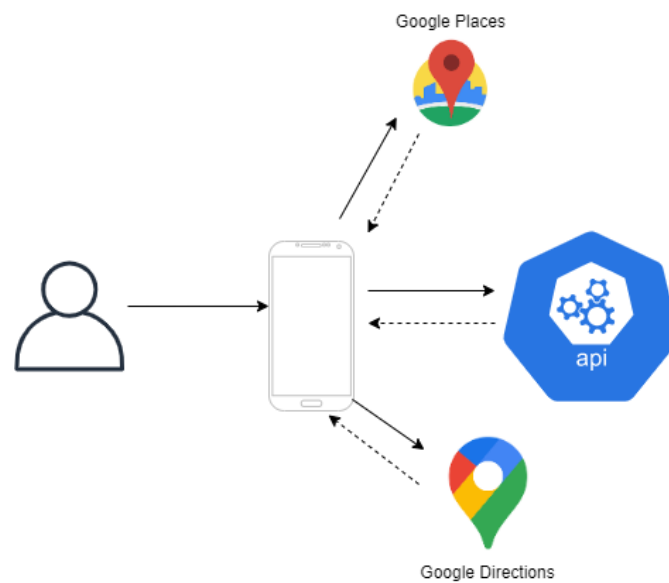


Figura 56: Componentes que interagem com a aplicação



### **4.5.3 Planear a implementação**

Após a criação da solução, esta carece de implementação em ambiente real. Estando este projeto associado a um co-financiamento pendente de resposta, a implementação apenas será possível aquando a vinda da mesma. Contudo espera-se a obtenção de parceiros estratégicos, nomeadamente, entidades gestoras de parques de estacionamento que possam, por um lado facultar dados alusivos à ocupação dos parque que detém, e por outro queiram testar e disponibilizar a solução aos seus clientes. As Câmaras Municipais poderão ser fortes parceiros uma vez que estas controlam os parques públicos municipais, podendo ainda auxiliar na difusão da solução.

A disponibilização da aplicação mobile na Play Store e na APP Store constituem também tarefas da implementação.

Importante também será registar um domínio por forma a associar a API a esse domínio, ao invés de aceder por IP da máquina onde esta se encontra alojada.

### **4.5.4 Planear a monitorização e manutenção**

A monitorização do projeto passará por controlar o histórico de pedidos que são efectuados à API analisando os códigos HTTP de resposta. Caso existam erros nos registos da API, estes servirão de base para proceder à correção e manutenção da solução.

Na manutenção serão ainda incluídas o desenvolvimento de novas funcionalidades e possíveis otimizações dos modelos de ML.

Ao nível da monitorização, a utilização dos registos referidos anteriormente poderá servir de base para realizar uma análise estatística de quais os locais mais requisitados, quais os parques mais escolhidos e a distribuição horária da utilização dos mesmos.

### **4.5.5 Produzir o relatório final**

O presente relatório surge como resultado da fase de produção do relatório final, presente na sexta fase da metodologia CRISP-DM, na qual foram cruzados os resultados das execuções das demais atividades requisitadas na supracitada metodologia, bem como das experiências efetivamente realizadas, desde o

conjunto de tarefas de análise e controlo e tratamento de dados à utilização dos diferentes algoritmos de ML. Nele são ainda descritos, organizados e analisados os resultados dessas mesmas experiências, bem como um conjunto de explicações, contextualizações e descrições de tudo que foi feito até ao término do projeto.

#### **4.5.6 Rever o projeto**

Este projeto contém pontos positivos e negativos. Como factores negativos é de destacar a baixa eficiência dos modelos, que se traduzem em previsões que se afastam da realidade numa parte considerável dos casos. Outro fator que se pode considerar como negativo é a pouca abrangência geográfica da solução, isto é, a solução, no ponto em que se encontra, apenas contém parques de Bruxelas.

Positivamente é de realçar a disponibilização de um serviço que fornece vários dados que auxiliam a tomada de decisão de estacionamento, podendo este facilmente ser acedido. A aplicação mobile desenvolvida, que tanto funciona em Android como iOS, é também um fator positivo na medida em que se pode utilizar realmente a solução. Nesta também, importa realçar as integrações das APIS da Google que permitem melhorar a experiência e usabilidade da aplicação.

## 5 Conclusão

Nesta secção do documento, são apresentadas as principais conclusões obtidas com a realização deste trabalho científico, a análise dos riscos subjacentes ao projeto e ainda uma abordagem ao trabalho futuro.

### 5.1 Reflexão Crítica

Na introdução deste documento foram propostos alguns objetivos e resultados esperados para este projeto de dissertação, tendo sido definidos dois objetivos principais que se subdividiram em dois objetivos secundários.

O primeiro objetivo principal passava por tentar perceber até que ponto é possível gerar uma rota ideal para um cidadão tendo por base a disponibilidade, a taxa de ocupação e a localização dos parques de estacionamento. Este objetivo foi atingido, e comprovável com a utilização da aplicação mobile desenvolvida, uma vez que é possível traçar uma rota desde da localização atual até ao parque de estacionamento que melhor satisfaz as necessidades do cidadão.

O segundo objetivo principal consistia em prever a taxa de ocupação dos parques de estacionamento. Este objetivo foi parcialmente atingido, uma vez que, devido a fatores alheios a esta dissertação, as métricas de avaliação dos modelos são pouco satisfatórios. Por exemplo, o melhor modelo conseguiu acuidade de aproximadamente 20%, embora claro está que, era algo previsível uma vez que a geração de ocupações dos parques foi uma solução de recurso para o problema de obtenção de dados. Por consequente, podemos dizer que a aplicação está preparada para utilizar os modelos inteligentes, no entanto os modelos carecem de otimização com a utilização de dados de ocupações reais. De destacar que a solução é escalável e que apenas carece de uma atualização dos dados, ou seja, à partida, não será necessário trabalho extra aquando da inserção de dados reais das ocupações.

No que diz respeito a objetivos secundários, embora se possa prever que venham a concretizar-se, os mesmos só poderão ser verificados aquando da implementação e difusão da solução. O primeiro objetivo secundário passa por ter uma gestão mais efetiva dos parques de estacionamento através da otimização da taxa de ocupação. Uma vez que são fornecidos dados acerca das taxas de ocupação dos parques aos utilizadores da solução, e aquando da concretização do trabalho futuro que contemplará a obtenção de ocupações reais que terão impacto na eficiência dos modelos preditivos, os cidadãos estarão em condições de efectuar uma melhor escolha acerca de qual o parque de estacionamento devem optar.

O segundo objetivo secundário passava por reduzir o desperdício de recursos. Este só será possível no momento em que se definirem parcerias estratégicas com entidades que gerem parques de estacionamento, por forma a fornecer-lhes previsões das taxas de ocupação dos seus parques nas diferentes horas do dia, nos vários dias do ano. Isto faz com que esta solução tenha também impacto na tomada de decisão das entidades gestoras de parques. Os próprios cidadãos, ao poderem utilizar uma aplicação que lhes permite seleccionar o parque mais apropriado, previsivelmente reduzirá do tempo de procura de parques de estacionamento.

Depois de efectuado este projeto de dissertação, era esperada a obtenção de determinados resultados. Tal como indicado na introdução deste documento foram definidos 6 resultados esperados.

Era espectável a obtenção de conhecimento acerca da realidade atual da temática, e com a revisão de literatura efectuada podemos concluir que foi atingido.

Outro resultado esperado passava por conhecer as características e tipologias de parques de estacionamento, e tanto com a análise dos datasets obtidos como com a secção da revisão de literatura dedicada ao estacionamento (2.2) este resultado foi conseguido.

A exploração de técnicas de Inteligência Artificial (IA)/Machine Learning (ML), que corresponde ao terceiro resultado esperado, foi atingido não só na revisão de literatura sobre a temática (2.4) como no trabalho realizado, na fase de modelação (4.3).

O quarto resultado esperado passava por desenvolver um conjunto de dados real capaz de simular o dia-a-dia dos parques, e podemos dizer que este resultado foi alcançado, embora com uma parte dos dados gerada. Este resultado pode ser verificado na fase de recolha de dados inicial (4.1.1).

A elaboração de um protótipo funcional constituía o quinto resultado esperado. Este resultado foi atingido com o desenvolvimento da aplicação mobile (4.5.2) que possibilitou a utilização dos modelos inteligentes previamente desenvolvidos e disponibilizados via API.

Por último e não menos importante, a geração de novo conhecimento que combine a área dos sistemas de informação, interoperabilidade, programação web e inteligência artificial foi apontado como o sexto resultado esperado e o mesmo foi atingido. Os artefactos que correspondem à geração de novo

conhecimento são os modelos inteligentes, a API Rest e a aplicação mobile. Todo o conhecimento desenvolvido e a área de atuação deste projeto de dissertação enquadram-se nos sistemas de informação. A interoperabilidade foi conseguida com recurso à integração das diversas APIs que comunicam com o módulo de recolha de dados, com a API desenvolvida e com a aplicação mobile. Os conhecimentos de programação web foram aplicados na integração das várias APIs e no desenvolvimento da API Rest. Por fim, a inteligência artificial que por si só, foi a temática em foco neste projeto de dissertação e na qual foi despendido mais tempo.

Na Tabela 34, apresentada de seguida, é possível verificar quais as componentes do trabalho realizado onde se pode comprovar a concretização dos vários objetivos e resultados esperados. Desta forma nas linhas encontram-se os objetivos principais, secundários e resultados esperados, e nas colunas apenas as componentes do trabalho necessárias.

Tabela 34: Matriz de concretização de objetivos e resultados esperados

<b>Objetivos</b>	<b>Revisão de Literatura</b>	<b>Recolha de dados</b>	<b>Modelos Inteligentes</b>	<b>API</b>	<b>Aplicação Mobile</b>
<b>Gerar rota ideal tendo por base a taxa de ocupação e localização dos parques</b>					X
<b>Prever a taxa de ocupação dos parques de estacionamento</b>			X		
<b>Otimizar a taxa de ocupação dos parques</b>					X
<b>Reduzir desperdícios de recursos</b>				X	

Continua na próxima página

<b>Objetivos</b>	<b>Revisão de Literatura</b>	<b>Recolha de dados</b>	<b>Modelos Inteligentes</b>	<b>API</b>	<b>Aplicação Mobile</b>
<b>Conhecer a realidade atual sobre a temática</b>	X				
<b>Conhecer as características e tipologias dos parques de estacionamento</b>	X				
<b>Explorar técnicas de IA/ML</b>			X		
<b>Desenvolver um conjunto de dados real capaz de simular o dia-a-dia dos parques</b>		X			
<b>Elaborar um Protótipo funcional</b>					X
<b>Gerar novo conhecimento que combine a área dos sistemas de informação, interoperabilidade, programação web e inteligência artificial</b>			X		

A questão de investigação deste projeto de dissertação consistia em perceber “De que forma podemos melhorar a mobilidade urbana através da Inteligência Artificial?”. Estando o projeto concluído e tendo sido cumpridos os objetivos, estão reunidas as condições para dar uma resposta a esta questão. E a resposta prende-se com o que foi desenvolvido neste projeto de dissertação, ou seja o desenvolvimento de soluções que utilizem modelos de Inteligência Artificial e que sejam disponibilizados como um serviço para que o cidadão possa utilizar e facilitar a sua tomada de decisão no dia-a-dia em ambientes urbanos

Globalmente, pode dizer-se que o projeto de dissertação representa um importante avanço para o ioCity, uma vez que em caso de aprovação, todo o trabalho desenvolvido será considerado e utilizado. Mais se acrescenta que, os resultados atingidos permitiram o desenvolvimento de uma nova ioSolution para a IOTECH. Esta nova solução, denominada IOParking, com os devidos parceiros, poderá revelar-se impactante na medida em que é uma solução que se enquadra na missão da empresa.

Em suma, deste projeto de dissertação resultaram 35 modelos de Machine Learning, 1 API Restful e 1 aplicação mobile.

## 5.2 Análise de Riscos

Na fase inicial deste projeto de dissertação foi efectuado um levantamento dos vários factores de risco que podia comprometer o sucesso do mesmo.

A Tabela 35 apresenta a análise efectuada, estando os itens ordenados de forma descendente pelo nível de gravidade. Para cada um dos riscos identificados, foram estimados a sua probabilidade e impacto. A gravidade ou severidade foi calculada multiplicando a probabilidade pelo impacto. As consequências e estratégias de mitigação para cada risco também foram detalhadas. Depois de finalizado o projeto foi adicionada uma nova coluna que indica se determinado risco ocorreu ou não.

A escala usada foi a seguinte:

1. P (Probabilidade): de **1** a **5**, em que 1 indica um risco com baixa probabilidade de ocorrer e 5 um risco com alta probabilidade
2. I (Impacto): de **1** a **5**, em que 1 indica um risco com baixo impacto e 5 um risco de elevado impacto

3. G (Gravidade): Resultado da multiplicação da probabilidade pelo impacto. Desta forma varia de **1** a **25**, sendo que quanto mais elevado for o valor, mais gravidade está associada.

Tabela 35: Análise do risco

<b>Risco</b>	<b>P</b>	<b>I</b>	<b>G</b>	<b>Consequências</b>	<b>Ocorreu</b>	<b>Mitigação</b>
Não obtenção de datasets de qualidade	4	5	20	Pode comprometer a realização do projeto ou levar a resultados pouco satisfatórios	Sim	Pesquisa de dados em plataformas de referência e geração de dados
Elevado grau de complexidade desta área de investigação	4	4	16	Pode levar a atrasos na entrega dos artefactos devido ao processo de investigação ser mais demorado	Sim	Aplicar com rigor a metodologia Scrum por forma a lidar com a complexidade do projeto
Falta de experiência em projetos da área	4	4	16	Pode comprometer o cumprimento dos objetivos do projeto e levar a atrasos nas entregas dos artefactos	Sim	Ler documentação, tutoriais, livros e material relacionados. Solicitar o fornecimento de diretrizes ao orientador
Poder computacional reduzido	3	3	9	Isto pode comprometer a execução dos modelos de ML e IA e consequentemente levar a resultados de menor qualidade	Sim	Recorrer a serviços <i>cloud</i> ajustando o poder computacional às necessidades
Os resultados obtidos não irem ao encontro do que é expectável	2	4	8	Isto pode comprometer a qualidade dos artefactos do projeto e a normal execução do IOCity	Não	Adiar a data prevista para a entrega dos artefactos de pesquisa científica. Validar constantemente com o <i>Product Owner</i> os resultados obtidos

Continua na próxima página



Risco	P	I	G	Consequências	Ocorreu	Mitigação
Dificuldade de comunicação com o orientador	1	4	4	Isto pode levar a interpretações incorretas do projeto, reduzindo assim a probabilidade de sucesso	Não	Tentar incentivar a ocorrência de reuniões semanais com o orientador e utilizar plataformas de comunicação adequadas
Planeamento inadequado	1	4	4	Pode levar a atrasos na entrega dos artefactos e na redução da qualidade dos mesmos	Não	Aplicação de agilidade no processo através da priorização das atividades e a execução de ciclos iterativos de investigação e desenvolvimento
Perda de arquivos	1	4	4	Pode colocar em causa todo o planeamento realizado	Sim	Dar prevalência à utilização de armazenamento em <i>cloud</i>
Avarias nas infraestruturas utilizadas	1	3	3	Pode colocar em causa a entrega atempada dos artefactos.	Sim	Utilização de serviços <i>cloud</i> e se possível com redundância por forma a ter <i>backups</i>

### 5.3 Trabalho Futuro

Depois de concluídos este projeto de dissertação importa deixar algumas anotações acerca de qual o trabalho a realizar no futuro.

Uma vez que esta dissertação está associada ao ioCity, projeto sujeito a uma candidatura PT2020 a qual ainda não se obteve resposta, o futuro dependerá muito do resultado da mesma. Desta forma a primeira tarefa será aguardar pela resposta e em caso de ser afirmativa deve-se concretizar os pontos anteriormente mencionados na secção 4.4.3. Além disto será importante converter os protótipos em soluções comerciais, adicionando robustez à API e à aplicação mobile desenvolvidas. Por exemplo, a adição de autenticação OAuth 2.0 é algo que necessariamente terá que ser feito.

## Bibliografia

- Accenture. (2014). *Technology Vision 2014 Building Cities for the Digital Citizen* (Tech. Rep.). Retrieved from [https://www.accenture.com/t00010101t000000\\_\\_w\\_\\_/\\_au-en/\\_acnmedia/pdf-11/accenture-technology-vision-2014-building-cities.pdf](https://www.accenture.com/t00010101t000000__w__/_au-en/_acnmedia/pdf-11/accenture-technology-vision-2014-building-cities.pdf)
- Alsafery, W., Alturki, B., Reiff-Marganiec, S., & Jambi, K. (2018). Smart Car Parking System Solution for the Internet of Things in Smart Cities.
- Bates, J., & Leibling, D. (2012). *Spaced Out Perspectives on parking policy* (Tech. Rep.). RAC Foundation. Retrieved from [www.racfoundation.org](http://www.racfoundation.org)
- Bellman, R. E. (1978). *An Introduction to Artificial Intelligence: Can Computers Think?* (illustrated edition ed.). Boyd & Fraser Pub. Co.
- Berrone, P., Ricart, J. E., Duch, A., & Carrasco, C. (2019, 5). *IESE Cities in Motion Index 2019* (Tech. Rep.). Retrieved from <https://www.ieseinsight.com/fichaMaterial.aspx?pk=154263&idi=2&origen=3> doi: 10.15581/018.ST-509
- Brooke, S., Ison, S., & Quddus, M. (2014). *"Parking Choice", Parking Issues and Policies*. Emerald Group Publishing Limited.
- Brownlee, J. (2016). *Supervised and Unsupervised Machine Learning Algorithms*. Retrieved from <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>
- Brownlee, J. (2020, 4). *4 Types of Classification Tasks in Machine Learning*. Retrieved from <https://machinelearningmastery.com/types-of-classification-in-machine-learning/>
- Camagni, R., Gibelli, M. C., & Rigamonti, P. (2002). Urban mobility and urban form: the social and environmental costs of different patterns of urban expansion. *Ecological Economics*, 40, 199–216.
- Cambridge University Press. (2019). *Parking Garage - meaning in the Cambridge English Dictionary*. Retrieved from <https://dictionary.cambridge.org/dictionary/english/parking-garage>
- Cardoso, T., Alarcão, I., & Celorico, J. A. (2010). *Revisão da Literatura e Sistematização do Conhecimento* (2-2010 ed.). Porto: Porto Editora.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 - Step-by-step data mining guide* (Tech. Rep.).
- Charniak, E., & McDermott, D. (1985). *Introduction to artificial intelligence*. Boston: Addison-Wesley

Longman Publishing Co.

- Costa, P. B., Neto, G. C., & Bertolde, A. I. (2017). Urban Mobility Indexes: A Brief Review of the Literature. In *Transportation research procedia* (Vol. 25, pp. 3645–3655). Elsevier B.V. doi: 10.1016/j.trpro.2017.05.330
- Cunha, M. A., Przeybilowicz, E., Macaya, J. F. M., & Burgos, F. (2016). *Smart Cities: Transformação digital de cidades* (1st ed.). São Paulo.
- Dirks, S., & Keeling, M. (2009). *A vision of smarter cities - How cities can lead the way into a prosperous and sustainable future IBM Global Business Services* (Tech. Rep.). IBM. Retrieved from [https://www-03.ibm.com/press/attachments/IBV\\_Smarter\\_Cities\\_-\\_Final.pdf](https://www-03.ibm.com/press/attachments/IBV_Smarter_Cities_-_Final.pdf)
- DriversEd. (2019). *Parking - Driving Information*. Retrieved from <https://driversed.com/driving-information/driving-techniques/parking.aspx>
- Fedchenkov, P., Anagnostopoulos, T., Zaslavsky, A., Ntalianis, K., Sosunova, I., & Sadov, O. (2018). An Artificial Intelligence Based Forecasting in Smart Parking with IoT. In O. Galinina, S. Andreev, S. Balandin, & Y. Koucheryav (Eds.), *Internet of things, smart spaces, and next generation networks and systems* (pp. 33–40). Cham: Springer International Publishing.
- Filho, W. F. (2017). *Computer Science Distilled: Learn the Art of Solving Computational Problems*. Code Energy LLC.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. Retrieved from <http://www.deeplearningbook.org>
- Guevara, M. (2019). *Como trazer a inteligência artificial para a sua empresa*. Retrieved from <http://www.ccg.pt/como-trazer-a-inteligencia-artificial-para-a-sua-empresa/>
- Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. MIT Press.
- Honda, H., Facure, M., & Yaohao, P. (2017). *Os Três Tipos de Aprendizado de Máquina*. Retrieved from <https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/>
- Huilgol, P. (2019, 8). *Accuracy vs. F1-Score*. Retrieved from <https://medium.com/analytics-vidhya/accuracy-vs-f1-score-6258237beca2>
- Ibrahim, H. E.-D. (2018). Car Parking Problem in Urban Areas, Causes and Solutions. In *1st international conference on towards a better quality of life, 2017* (p. 13). Retrieved from <https://ssrn.com/abstract=3163473>
- INE. (2018). *Inquérito à Mobilidade nas Áreas Metropolitanas do Porto e de Lisboa 2017 Predomínio do automóvel nas deslocações dos residentes das Áreas Metropolitanas de Porto e Lisboa* (Tech. Rep.). INE.

- Instituto Nacional de Estatística. (2018). *Mobilidade e funcionalidade do território nas Áreas Metropolitanas do Porto e de Lisboa 2017*. Lisboa: Instituto Nacional de Estatística, I. P. Retrieved from [www.ine.pt](http://www.ine.pt)
- International Standards Organization (ISO). (2015). *Smart cities Preliminary Report 2014* (Tech. Rep.). Retrieved from [http://www.iso.org/iso/smart\\_cities\\_report-jtc1.pdf](http://www.iso.org/iso/smart_cities_report-jtc1.pdf)
- International Telecommunications Union (ITU). (2014). *Smart sustainable cities: An analysis of definitions* (Tech. Rep.). ITU. Retrieved from [www.itu.int/en/ITU-T/focusgroups/ssc/Documents/Approved\\_Deliverables/TR-Definitions.docx](http://www.itu.int/en/ITU-T/focusgroups/ssc/Documents/Approved_Deliverables/TR-Definitions.docx)
- Ionita, A., Pomp, A., Cochez, M., Meisen, T., & Decker, S. (2018, 6). Where to park? predicting free parking spots in unmonitored city areas. In *Acm international conference proceeding series*. Association for Computing Machinery. doi: 10.1145/3227609.3227648
- Iriondo, R. (2018). *Machine Learning vs. AI, Important Differences Between Them*. Retrieved from <https://medium.com/datadriveninvestor/differences-between-ai-and-machine-learning-and-why-it-matters-1255b182fc6>
- Ison, S. (2014). *Parking Management Policy: its potential in improving urban traffic flows* (Tech. Rep.). Loughborough University. Retrieved from [https://www.acea.be/uploads/publications/21th\\_SAG.pdf](https://www.acea.be/uploads/publications/21th_SAG.pdf)
- IST. (2009). *Estacionamento. Apontamentos da cadeira de Transportes do Mestrado Integrado em Engenharia Civil. J.M. Viegas, Instituto Superior Técnico (IST)*. Retrieved from <https://fenix.tecnico.ulisboa.pt/downloadFile/3779572236300/Estacionamento.pdf>
- Júdice, D. (2014). *Smart cities*. Retrieved from <http://www.ansr.pt/SegurancaRodoviaria/ArtigosTecnicos/Pages/default.aspx>
- Kurzweil, R. (1990). *The Age of Intelligent Machines*. MIT Press.
- Labs, I. (2019). *Top 10 Machine Learning Algorithms And Its Use Cases*. Retrieved from <https://medium.com/@imaginorlabs/top-10-machine-learning-algorithms-and-its-use-cases-fc303daa2003>
- Li, X., & Jiang, H. (2017). Artificial Intelligence Technology and Engineering Applications. *ACES Journal*, 32, 8.
- Lijbers, J. (2016). *Predicting parking lot occupancy using Prediction Instrument Development for Complex Domains* (Unpublished doctoral dissertation). University of Twente.
- Litman, T. (2016). *Parking Management Strategies, Evaluation and Planning* (Tech. Rep.). Victoria Transport Policy Institute. Retrieved from [https://www.vtpi.org/park\\_man.pdf](https://www.vtpi.org/park_man.pdf)

- Luger, G. F., & Stubblefield, W. A. (1998). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (Third Edition ed.). Addison Wesley Longman.
- Machado, J. (2018). *Inteligência artificial já é o presente*. Retrieved from <http://algoritmi.uminho.pt/inteligencia-artificial-ja-e-o-presente/>
- Magagnin, R. C. (2008). The perception of the expert on urban mobility theme. *Magazine Transport*, vol.16(1).
- Manville, C., Europe, R., Millard, J., Technological Institute, D., Liebe, A., & Massink, R. (2014). *Directorate General for Internal Policies. Policy Department A: Economic and scientific policy. Mapping Smart cities in the EU* (Tech. Rep.). União Europeia. Retrieved from [http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/507480/IPOL-ITRE\\_ET\(2014\)507480\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/507480/IPOL-ITRE_ET(2014)507480_EN.pdf)
- Media, O. (2017). *Artificial Intelligence Now*.
- Mitchell, M. (2019). *Artificial Intelligence* (First Edition ed.).
- Mitchell, T. M. (1997). *Machine Learning* (1edition ed.). McGraw-Hill Education.
- Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). *Machine Learning: Algorithms and Applications* (1st Edition ed.). CRC Press.
- Morris, D. (2016). *Today's Cars Are Parked 95% of the Time*. Retrieved from <https://fortune.com/2016/03/13/cars-parked-95-percent-of-time/>
- Nevala, K. (2017). *The Machine Learning Primer* (Tech. Rep.). SAS.
- Nilsson, N. J. (1998). *Artificial Intelligence: A New Synthesis* (1st edition ed.). Morgan Kaufmann.
- Pannu, A., & Student, M. T. (2015). *Artificial Intelligence and its Application in Different Areas* (Vol. 4; Tech. Rep. No. 10).
- Peppers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007, 12). A Design Science Research Methodology for Information Systems Research. *J. Manage. Inf. Syst.*, 24(3), 45–77. Retrieved from <https://doi.org/10.2753/MIS0742-1222240302> doi: 10.2753/MIS0742-1222240302
- Poole, D., Mackworth, A., & Goebel, R. (1998). *Computational Intelligence: A Logical Approach* (1st Edition ed.). Oxford University Press.
- Pramathi, J. N., & Vindhya, P. M. (2018). Artificial Intelligence in Smart City Analysis. *International Conference on Smart Systems and Inventive Technology*.
- RapidAPI Team. (2020). *Top 7 Best Free Weather APIs to Access Global Weather Data (Updated for 2020)*. Retrieved from <https://rapidapi.com/blog/access-global-weather-data>

-with-these-weather-apis/

- Rich, E., & Knight, K. (1991). *Artificial Intelligence* (2nd Edition ed.). McGraw-Hill Education.
- Rodrigue, J.-P., Comtois, C., & Slack, B. (2017). *The Geography of Transport Systems* (FOURTH EDI ed.). New York: Routledge.
- Royal Literary Fund. (2019). *What is a literature review?* Retrieved from <https://www.rlf.org.uk/resources/what-is-a-literature-review/>
- Russel, S., & Norvig Peter. (2010). *Artificial Intelligence - A Modern Approach* (Third Edition ed.). New Jersey.
- Schwaber, K., & Sutherland, J. (2017). *The Scrum Guide* (Tech. Rep.). Retrieved from [https://www.scrum.org/resources/scrum-guide?gclid=Cj0KCQiA04XxBRD5ARIsAGFygj9UayJTjVDwxevvQN0175E-w6IvQaZ6X2Ju\\_9PRIumDMpv6jAkU0kAaAtvbEALw\\_wcB](https://www.scrum.org/resources/scrum-guide?gclid=Cj0KCQiA04XxBRD5ARIsAGFygj9UayJTjVDwxevvQN0175E-w6IvQaZ6X2Ju_9PRIumDMpv6jAkU0kAaAtvbEALw_wcB)
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press. Retrieved from <http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning>
- Sousa, J. (2019, 10). O papel da distribuição normal na Estatística. , 1–1. Retrieved from <http://hdl.handle.net/10400.3/5363>
- Swamynathan, M. (2017). *Mastering Machine Learning with Python in Six Steps*. Apress. doi: 10.1007/978-1-4842-2866-1
- VanderPlas, J. (2017). *Python Data Science Handbook*. Retrieved from [www.allitebooks.com](http://www.allitebooks.com)
- Vlahogianni, E. I., Kepaptsoglou, K., Tsetsos, V., & Karlaftis, M. G. (2016, 3). A Real-Time Parking Prediction System for Smart Cities. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 20(2), 192–204. doi: 10.1080/15472450.2015.1037955
- Winston, P. H. (1992). *Artificial Intelligence* (Third Edition ed.). Addison-Wesley.
- Xu, G., & Xu, Y. (2016). *GPS - Theory, Algorithms and Applications*. Springer, Berlin, Heidelberg. doi: <https://doi.org/10.1007/978-3-662-50367-6>

## Anexos

### A Recolha de dados dos Locais

#### A.1 Código fonte de recolha dos Locais - Parte 1

```
1 import json
2
3 ## Function to arrange data and extract info of original-paking-data
4
5 with open('../original-parking-data.json') as json_file:
6     data = json.load(json_file)
7
8     data_holder = []
9
10    for record in data['payloadPublication']['genericPublicationExtension']
11    ['parkingTablePublication']['parkingTable']['parkingRecord']:
12        x = {
13            'parkingRecordName': record['parkingName']['values']['value']['
14            #text'],
15            'parkingRecordLocation': record['parkingLocation']['
16            pointByCoordinates']['pointCoordinates'],
17            'parkingSite': [],
18        }
19        try:
20            if isinstance(record['parkingSite'], list):
21                for site in record['parkingSite']:
22                    k = {
23                        'name': site['parkingName']['values']['value'][0]['
24                        #text'],
25                        'parkingLocation': site['parkingLocation']['
26                        pointByCoordinates']['pointCoordinates'],
27                        'PointOfInterests': []
28                    }
29                    x['parkingSite'].append(k)
30            else:
31                k = {
```

```

27         'name': record['parkingSite']['parkingName']['values']['
'value'][0]['#text'],
28         'parkingLocation': record['parkingSite']['
parkingLocation']['pointByCoordinates']['pointCoordinates'],
29         'PointOfInterests': []
30     }
31     x['parkingSite'].append(k)
32
33     data_holder.append(x)
34 except:
35     print('Parking record without parking sites: ' + record['
parkingName']['values']['value']['#text'])
36
37     with open('../parking-data-pre-places.json', 'w', encoding='utf-8') as
f:
38         json.dump(data_holder, f, ensure_ascii=False, indent=4)
39

```

## A.2 Código fonte de recolha dos Locais - Parte 2

```

1     import requests
2     import json
3     import time
4
5     def get_nearby_places_by_distance(latitude, longitude, api_key,
data_holder, nextPageToken):
6         url = 'https://maps.googleapis.com/maps/api/place/nearbysearch/
json?location=' + str(latitude) + ',' + str(
7             longitude) + '&rankby=distance&key=' + api_key + '&
pagetoken=' + nextPageToken
8         response = requests.get(url).json()
9         if response['status'] == 'OK':
10            print("SUCCESS: request successfully done with url " + url)
11            data_holder.extend(response['results'])
12            if 'next_page_token' in response:
13                print("Next Page: waiting 3 seconds")
14                time.sleep(3)

```



```

15         get_nearby_places_by_distance(latitude, longitude,
api_key, data_holder, response['next_page_token'])
16     else:
17         print("Error code: " + response['status'] + ". Sleeping 3
seconds and re-doing request")
18         time.sleep(3)
19         get_nearby_places_by_distance(latitude, longitude, api_key,
data_holder, nextPageToken)
20
21
22     with open('../parking-data-pre-places.json') as json_file:
23         parking_data = json.load(json_file)
24
25
26     for parking_record in parking_data:
27         for parking_site in parking_record['parkingSite']:
28             print("#####PARKING SITE INIT: " + parking_site['name'
])
29             get_nearby_places_by_distance(parking_site['
parkingLocation']['latitude'], parking_site['parkingLocation']['
longitude'], google_api_key, parking_site['PointOfInterests'], '')
30             print("#####PARKING SITE DONE: " + parking_site['name'
])
31
32
33     with open('../places.json', 'w', encoding='utf-8') as f:
34         json.dump(parking_data, f, ensure_ascii=False, indent=4)
35

```

## B Recolha de dados da Meteorologia

```

1     from datetime import timedelta, date
2
3
4     def daterange2(start, finish, hours):
5         for n in range(int((finish - start).days)):
6             for h in hours:
7                 yield str(start + timedelta(n)) + 'T' + h

```

```
8
9
10     def get_historical_weather(region, data_holder, secret, latitude,
longitude, timestamp):
11         url = "https://api.darksky.net/forecast/" + secret + "/" +
latitude + "," + longitude + "," + timestamp + "?units=si&exclude=
currently,flags"
12         response = requests.get(url).json()
13         if response:
14             print("SUCCESS: " + ": " + region + " with timestamp " +
timestamp)
15             data_holder.append(response)
16         else:
17             print("ERROR :" + "sleeping 3seconds and re-doing request
with date: " + timestamp)
18             time.sleep(3)
19             get_historical_weather(region, data_holder, secret,
latitude, longitude, timestamp)
20
21
22     secret = "<secret key>"
23
24     start_date = date(2015, 1, 1) # One day before the desired
iteration
25     end_date = date(2020, 1, 1) # One day after the desired iteration
26
27
28     latitude = 50.85045
29     longitude = 4.34878
30     data = []
31
32     for d in helper.daterange2(start_date, end_date, ["00:00:00"]):
33         get_historical_weather('brussels', data, secret, str(latitude),
str(longitude), d)
34
```

## C Recolha de dados dos Feriados

```
1  import requests
2  import json
3  import time
4
5  api_key = "<secret key>"
6
7  holidays = [
8      {
9          "year": 2015,
10         "holidays": [
11             {
12                 "date": "2015-05-08",
13                 "start": "2015-05-08T00:00:00.000Z",
14                 "end": "2015-05-09T00:00:00.000Z",
15                 "name": "Iris Day",
16                 "type": "Regional",
17                 "public": False,
18                 "country": "BE"
19             }
20         ]
21     },
22     {
23         "year": 2016,
24         "holidays": [
25             {
26                 "date": "2016-05-08",
27                 "start": "2016-05-08T00:00:00.000Z",
28                 "end": "2016-05-09T00:00:00.000Z",
29                 "name": "Iris Day",
30                 "type": "Regional",
31                 "public": False,
32                 "country": "BE"
33             }
34         ]
35     },
36     {
37         "year": 2017,
38         "holidays": [
39             {
```

```
40         "date": "2017-05-08",
41         "start": "2017-05-08T00:00:00.000Z",
42         "end": "2017-05-09T00:00:00.000Z",
43         "name": "Iris Day",
44         "type": "Regional",
45         "public": False,
46         "country": "BE"
47     }
48 ]
49 },
50 {
51     "year": 2018,
52     "holidays": [
53         {
54             "date": "2018-05-08",
55             "start": "2018-05-08T00:00:00.000Z",
56             "end": "2018-05-09T00:00:00.000Z",
57             "name": "Iris Day",
58             "type": "Regional",
59             "public": False,
60             "country": "BE"
61         }
62     ]
63 },
64 {
65     "year": 2019,
66     "holidays": [
67         {
68             "date": "2019-05-08",
69             "start": "2019-05-08T00:00:00.000Z",
70             "end": "2019-05-09T00:00:00.000Z",
71             "name": "Iris Day",
72             "type": "Regional",
73             "public": False,
74             "country": "BE"
75         }
76     ]
77 },
78 {
```

```
79     "year": 2020,
80     "holidays": [
81         {
82             "date": "2020-05-08",
83             "start": "2020-05-08T00:00:00.000Z",
84             "end": "2020-05-09T00:00:00.000Z",
85             "name": "Iris Day",
86             "type": "Regional",
87             "public": False,
88             "country": "BE"
89         }
90     ]
91 }
92 ]
93
94 def get_holidays_in_year(year, country, data_holder, api_key):
95     url = "https://getfestivo.com/v1/holidays?api_key=" + api_key + "&
country=" + country + "&year=" + str(year)
96     response = requests.get(url).json()
97     if response:
98         print("SUCCESS: Data retrieved to " + str(year))
99         data_holder.extend(response['holidays']['holidays'])
100     else:
101         print("ERROR :" + "sleeping 3seconds and re-doing request with
year: " + str(year))
102         time.sleep(3)
103         get_holidays_in_year(year, country, data_holder, api_key)
104
105
106 for record in holidays:
107     get_holidays_in_year(record['year'], 'BE', record['holidays'],
api_key)
108
109
110 with open('../holidays.json', 'w', encoding='utf-8') as f:
111     json.dump(holidays, f, ensure_ascii=False, indent=4)
112
113
```

## D Recolha de dados das Ocupações

```

1  import json
2  import utils.helper as helper
3  from datetime import date, datetime
4  import numpy as np
5  import random
6
7  with open('../original-parking-data.json') as json_file:
8      data = json.load(json_file)
9
10
11     start_date = datetime(2015, 1, 1) # One day before the desired
iteration
12     end_date = datetime(2020, 1, 1)
13
14     holder = []
15     park_counter = 1
16
17     for record in data['payloadPublication']['genericPublicationExtension']
['parkingTablePublication']['parkingTable']['parkingRecord']:
18         try:
19             if isinstance(record['parkingSite'], list):
20                 for site in record['parkingSite']:
21
22                     for date in helper.dateRangeMinutes(start_date,
end_date, 60):
23                         print("PARKCOUNTER: " + str(park_counter))
24                         print("##### SITE: " + site['parkingName']['values
']['value'][0]['#text'] + " && DATE: " + str(date))
25
26                         occupation_generated = random.gauss(65, 20)
27
28                         x = {
29                             "park_name": site['parkingName']['values']['
value'][0]['#text'],
30                             "park_id": site['@id'],
31                             "date": str(date),

```

```

32         "dayOfWeek": date.weekday(),
33         "timestamp": str(date.timestamp()),
34         "occupation": 100 if occupation_generated > 100
else occupation_generated
35     }
36     holder.append(x)
37     park_counter += 1
38     else:
39         for date in helper.dateRangeMinutes(start_date, end_date):
40             print("PARKCOUNTER: " + str(park_counter))
41             print("##### SITE: " + record['parkingSite']['
parkingName']['values']['value'][0]['#text'] + " && DATE: " + str(date))
42
43             occupation_generated = random.gauss(65, 20)
44             x = {
45                 "park_name": record['parkingSite']['parkingName']['
values']['value'][0]['#text'],
46                 "park_id": record['parkingSite']['@id'],
47                 "date": str(date),
48                 "dayOfWeek": date.weekday(),
49                 "timestamp": str(date.timestamp()),
50                 "occupation": 100 if occupation_generated > 100
else occupation_generated
51             }
52             holder.append(x)
53             park_counter += 1
54         except:
55             print('Parking record without parking sites: ' + record['
parkingName']['values']['value']['#text'])
56
57     with open('../occupation.json', 'w', encoding='utf-8') as f:
58         json.dump(holder, f, ensure_ascii=False, indent=4)
59
60

```

## E Exploração dos dados dos Parques de estacionamento

- **id** - parkingSite.id

▼ Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	63	100.00%
Unique Count	63	100.00%
Duplicate Count	0	0.00%
Blank Count	0	0.00%

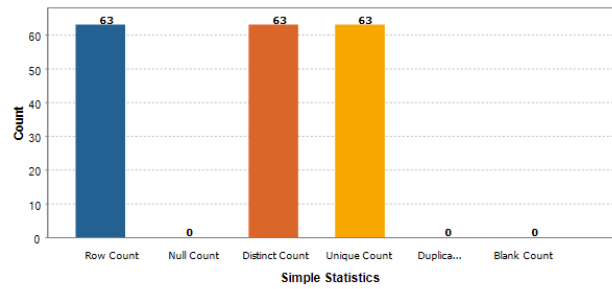


Figura E.1: Parques de estacionamento - Estatísticas Simples do atributo 'id'

- **latitude** - parkingSite.parkingLocation.pointByCoordinates.pointCoordinates.latitude

▼ Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	63	100.00%
Unique Count	63	100.00%
Duplicate Count	0	0.00%

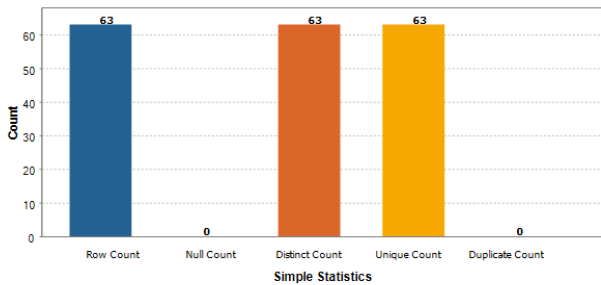


Figura E.2: Parques de estacionamento - Estatísticas Simples do atributo 'latitude'

- **longitude** - parkingSite.parkingLocation.pointByCoordinates.pointCoordinates.longitude

▼ Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	63	100.00%
Unique Count	63	100.00%
Duplicate Count	0	0.00%

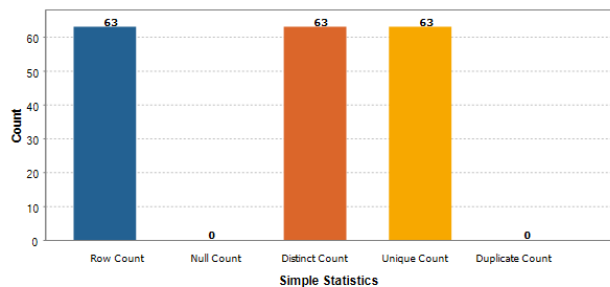


Figura E.3: Parques de estacionamento - Estatísticas Simples do atributo 'longitude'

- **parking\_group\_name** - parkingRecord.parkingName.values.value.#text



▼ Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	15	23.81%
Unique Count	8	12.70%
Duplicate Count	7	11.11%
Blank Count	0	0.00%

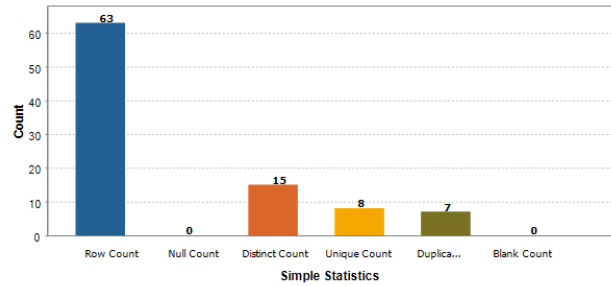
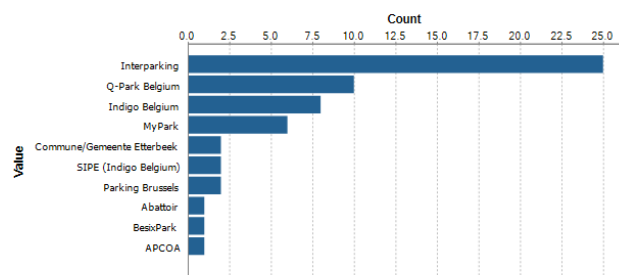


Figura E.4: Parques de estacionamento - Estatísticas Simples do atributo 'parking\_group\_name'

▼ Value Frequency

Value	Count	%
Interparking	25	39.68%
Q-Park Belgium	10	15.87%
Indigo Belgium	8	12.70%
MyPark	6	9.52%
Commune/Gemeente Etter...	2	3.17%
SIPE (Indigo Belgium)	2	3.17%
Parking Brussels	2	3.17%
Abattoir	1	1.59%
BesixPark	1	1.59%
APCOA	1	1.59%



▼ Value Low Frequency

Value	Count	%
Rauwers Parkingshop	1	1.59%
Abattoir	1	1.59%
Commune/Gemeente Mole...	1	1.59%
Hobinvest	1	1.59%
Hotel Warwick	1	1.59%
Commune/Gemeente Ixelle...	1	1.59%
BesixPark	1	1.59%
APCOA	1	1.59%
Commune/Gemeente Etter...	2	3.17%
Parking Brussels	2	3.17%

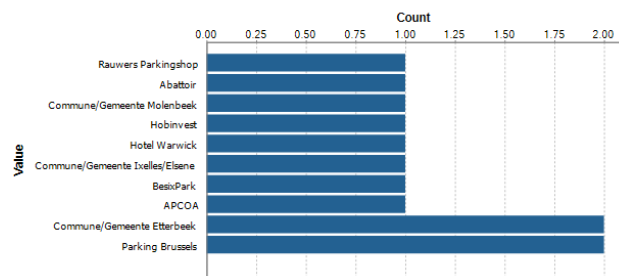


Figura E.5: Parques de estacionamento - Frequências do atributo 'parking\_group\_name'

- **parking\_name** - parkingSite.parkingName.values.value[0].#text

▼ Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	63	100.00%
Unique Count	63	100.00%
Duplicate Count	0	0.00%
Blank Count	0	0.00%

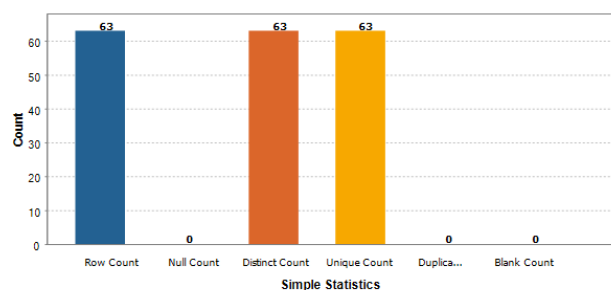


Figura E.6: Parques de estacionamento - Estatísticas Simples do atributo 'park\_name'

- **parkingHouseNumber** - parkingSite.parkingSiteAddress.contactDetailsHouseNumber

Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	45	71.43%
Unique Count	35	55.56%
Duplicate Count	10	15.87%
Blank Count	5	7.94%

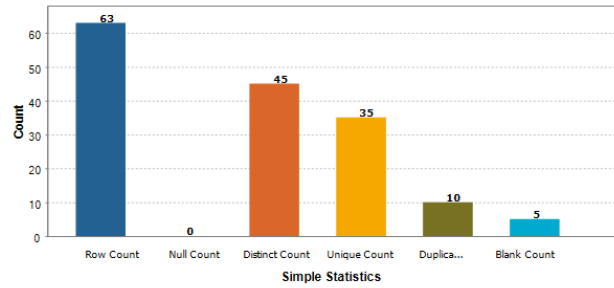
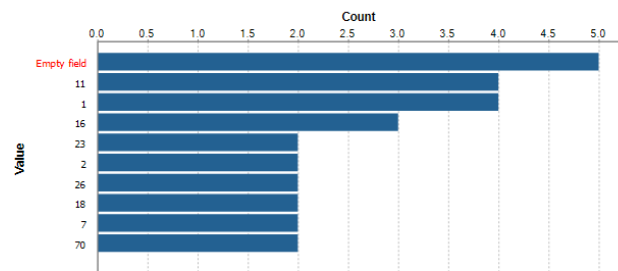


Figura E.7: Parques de estacionamento - Estatísticas Simples do atributo 'parkingHouseNumber'

Value Frequency

Value	Count	%
Empty field	5	7.94%
11	4	6.35%
1	4	6.35%
16	3	4.76%
23	2	3.17%
2	2	3.17%
26	2	3.17%
18	2	3.17%
7	2	3.17%
70	2	3.17%



Value Low Frequency

Value	Count	%
110	1	1.59%
101	1	1.59%
14	1	1.59%
103	1	1.59%
104	1	1.59%
128	1	1.59%
107	1	1.59%
164-166	1	1.59%
10	1	1.59%
120	1	1.59%

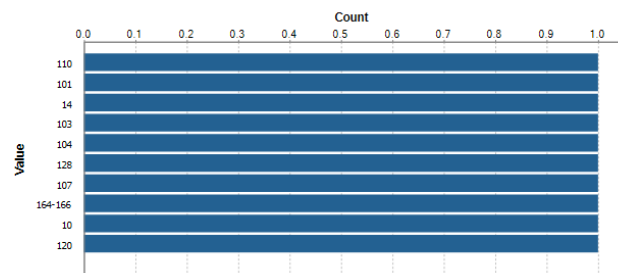


Figura E.8: Parques de estacionamento - Frequências do atributo 'parkingHouseNumber'

- **parkingNumberOfSpaces** - parkingSite.parkingNumberOfSpaces

Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	59	93.65%
Unique Count	55	87.30%
Duplicate Count	4	6.35%

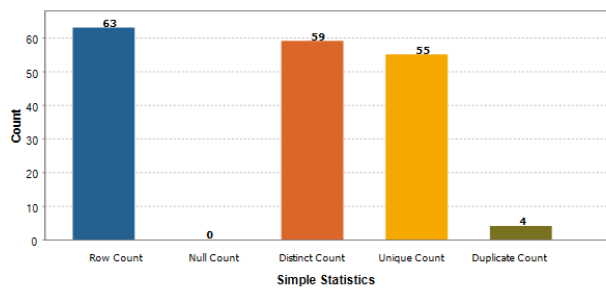
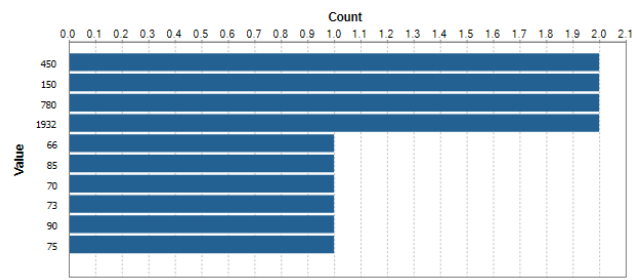


Figura E.9: Parques de estacionamento - Estatísticas Simples do atributo 'ParkingNumberOfSpaces'

▼ Value Frequency

Value	Count	%
450	2	3.17%
150	2	3.17%
780	2	3.17%
1932	2	3.17%
66	1	1.59%
85	1	1.59%
70	1	1.59%
73	1	1.59%
90	1	1.59%
75	1	1.59%



▼ Value Low Frequency

Value	Count	%
96	1	1.59%
66	1	1.59%
115	1	1.59%
85	1	1.59%
70	1	1.59%
73	1	1.59%
105	1	1.59%
90	1	1.59%
75	1	1.59%
92	1	1.59%

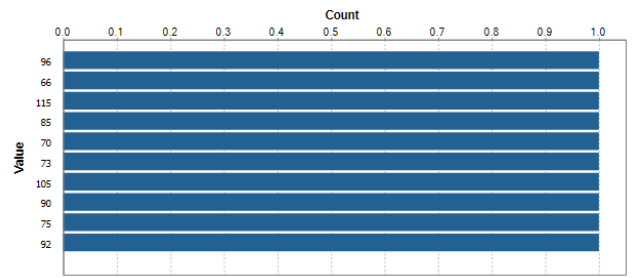
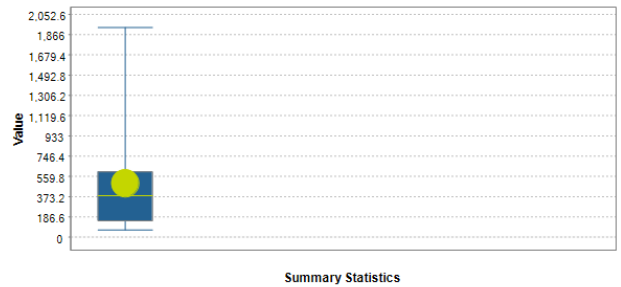


Figura E.10: Parques de estacionamento - Frequências do atributo 'ParkingNumberOfSpaces'

▼ Summary Statistics

Label	Value
Mean	495.85714285714283
Median	381.0
Inter Quartile Range	452.0
Lower Quartile	150.0
Upper Quartile	602.0
Range	1866.0
Minimum	66
Maximum	1932



▼ Mode

Mode	150
------	-----

Figura E.11: Parques de estacionamento - Estatísticas do atributo 'ParkingNumberOfSpaces'

- **parkingPostalCode** - parkingSite.parkingSiteAddress.contactDetailsPostCode

▼ Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	12	19.05%
Unique Count	5	7.94%
Duplicate Count	7	11.11%

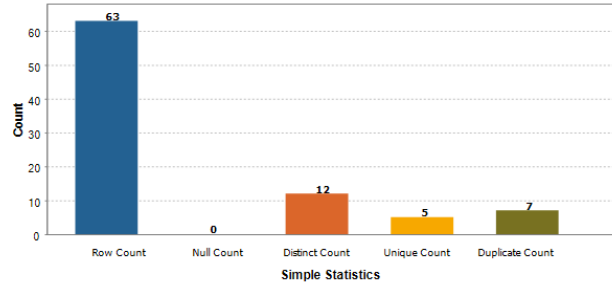
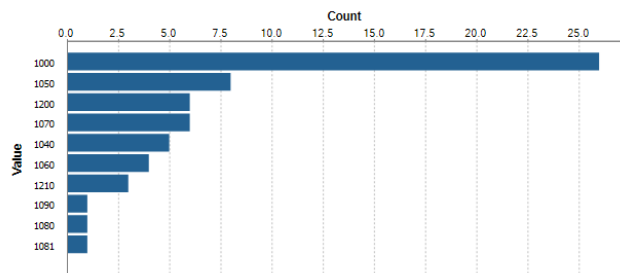


Figura E.12: Parques de estacionamento - Estatísticas Simples do atributo 'ParkingPostalCode'

▼ Value Frequency

Value	Count	%
1000	26	41.27%
1050	8	12.70%
1200	6	9.52%
1070	6	9.52%
1040	5	7.94%
1060	4	6.35%
1210	3	4.76%
1090	1	1.59%
1080	1	1.59%
1081	1	1.59%



▼ Value Low Frequency

Value	Count	%
1090	1	1.59%
1080	1	1.59%
1081	1	1.59%
1180	1	1.59%
1150	1	1.59%
1210	3	4.76%
1060	4	6.35%
1040	5	7.94%
1200	6	9.52%
1070	6	9.52%

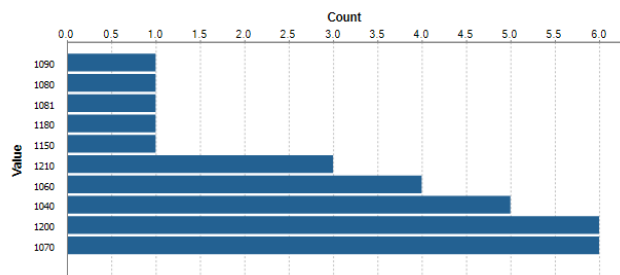


Figura E.13: Parques de estacionamento - Frequências do atributo 'ParkingPostalCode'

- **parkingStreet** - parkingSite.parkingSiteAddress.contactDetailsStreet

▼ Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	57	90.48%
Unique Count	51	80.95%
Duplicate Count	6	9.52%
Blank Count	0	0.00%

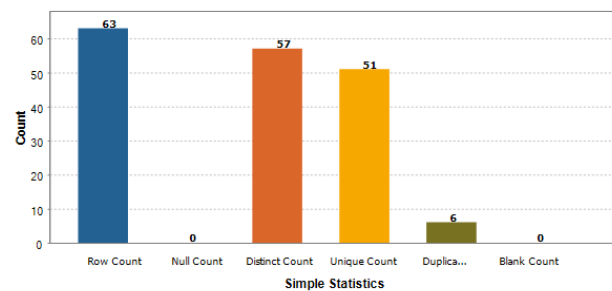
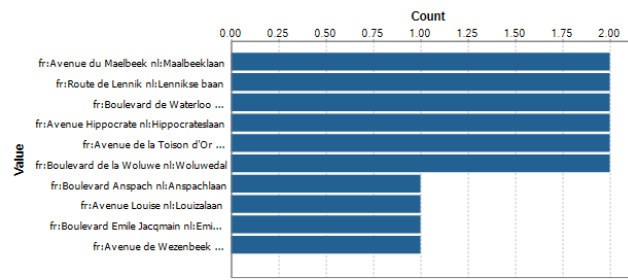


Figura E.14: Parques de estacionamento - Estatísticas Simples do atributo 'ParkingStreet'

▼ Value Frequency

Value	Count	%
fr:Avenue du Maelbeek nl:...	2	3.17%
fr:Route de Lennik nl:Lennik...	2	3.17%
fr:Boulevard de Waterloo n...	2	3.17%
fr:Avenue Hippocrate nl:Hi...	2	3.17%
fr:Avenue de la Toison d'Or...	2	3.17%
fr:Boulevard de la Woluwe ...	2	3.17%
fr:Boulevard Anspach nl:An...	1	1.59%
fr:Avenue Louise nl:Louizal...	1	1.59%
fr:Boulevard Emile Jacqmai...	1	1.59%
fr:Avenue de Wezenbeek nl...	1	1.59%



▼ Value Low Frequency

Value	Count	%
fr:Boulevard Josse Leemans...	1	1.59%
fr:Boulevard Pachéco nl:Pac...	1	1.59%
fr:Boulevard du Jardin Bota...	1	1.59%
fr:Boulevard Anspach nl:An...	1	1.59%
fr:Avenue Louise nl:Louizal...	1	1.59%
fr:Boulevard Emile Jacqmai...	1	1.59%
fr:Boulevard Lambertmont n...	1	1.59%
fr:Boulevard Sylvain Dupuis...	1	1.59%
fr:Avenue de Wezenbeek nl...	1	1.59%
fr:Boulevard Maurice Lem...	1	1.59%

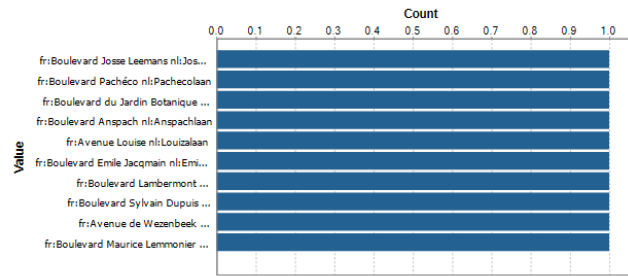


Figura E.15: Parques de estacionamento - Frequências do atributo 'ParkingStreet'

- **urbanParkingSiteType** - parkingSite.urbanParkingSiteType

▼ Simple Statistics

Label	Count	%
Row Count	63	100.00%
Null Count	0	0.00%
Distinct Count	2	3.17%
Unique Count	0	0.00%
Duplicate Count	2	3.17%
Blank Count	2	3.17%

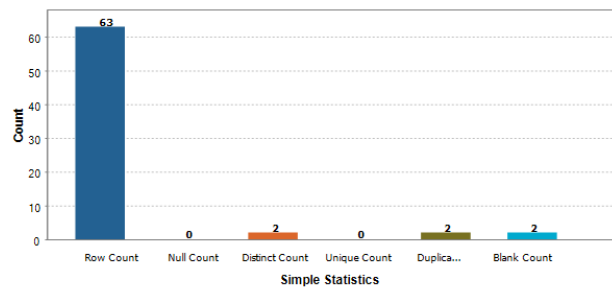
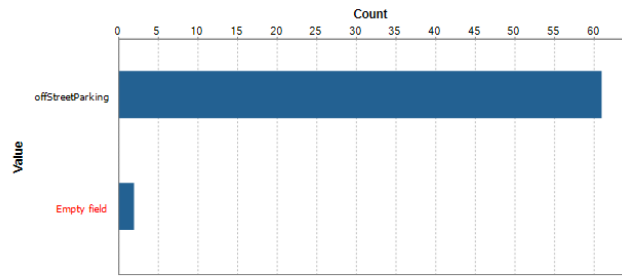


Figura E.16: Parques de estacionamento - Estatísticas Simples do atributo 'UrbanParkingSiteType'

Value Frequency

Value	Count	%
offStreetParking	61	96.83%
Empty field	2	3.17%



Value Low Frequency

Value	Count	%
Empty field	2	3.17%
offStreetParking	61	96.83%

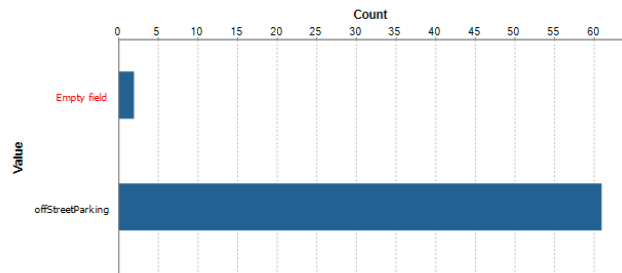


Figura E.17: Parques de estacionamento - Frequências do atributo 'UrbanParkingSiteType'

## F Exploração dos dados dos Locais

- **park\_latitude** - parkingSite.parkingLocation.latitude

Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	63	1.67%
Unique Count	0	0.00%
Duplicate Count	63	1.67%

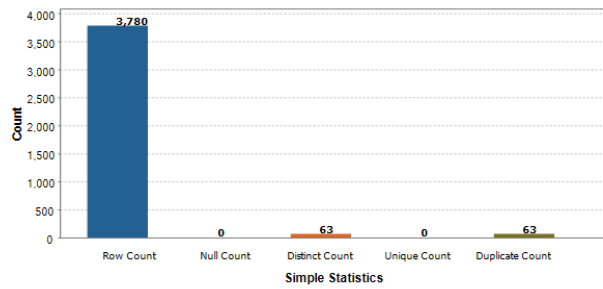
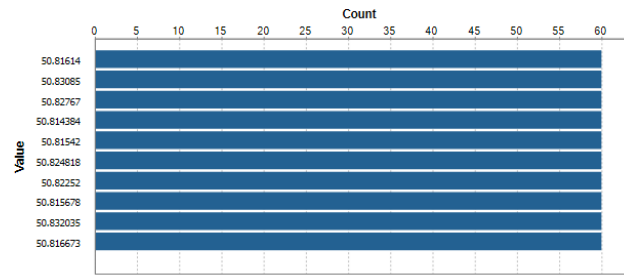


Figura F.1: Locais - Estatísticas Simples do atributo 'park\_latitude'

▼ Value Frequency

Value	Count	%
50.81614	60	1.59%
50.83085	60	1.59%
50.82767	60	1.59%
50.814384	60	1.59%
50.81542	60	1.59%
50.824818	60	1.59%
50.82252	60	1.59%
50.815678	60	1.59%
50.832035	60	1.59%
50.816673	60	1.59%



▼ Value Low Frequency

Value	Count	%
50.81614	60	1.59%
50.83085	60	1.59%
50.82767	60	1.59%
50.814384	60	1.59%
50.81542	60	1.59%
50.824818	60	1.59%
50.82252	60	1.59%
50.815678	60	1.59%
50.832035	60	1.59%
50.816673	60	1.59%

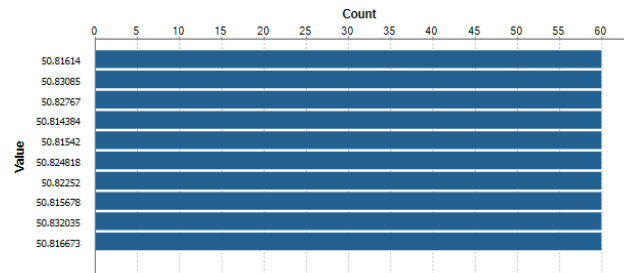


Figura F.2: Locais - Frequências do atributo 'park\_latitude'

- **park\_longitude** - parkingSite.parkingLocation.longitude

▼ Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	63	1.67%
Unique Count	0	0.00%
Duplicate Count	63	1.67%

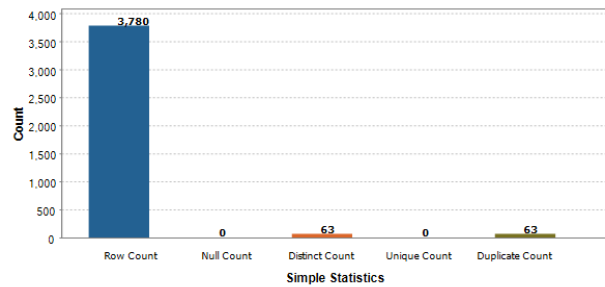
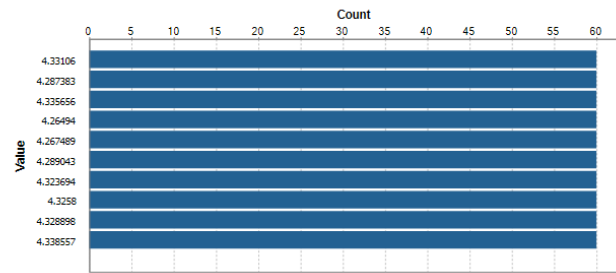


Figura F.3: Locais - Estatísticas Simples do atributo 'park\_longitude'

Value Frequency

Value	Count	%
4.33106	60	1.59%
4.287383	60	1.59%
4.335656	60	1.59%
4.26494	60	1.59%
4.267489	60	1.59%
4.289043	60	1.59%
4.323694	60	1.59%
4.3258	60	1.59%
4.328898	60	1.59%
4.338557	60	1.59%



Value Low Frequency

Value	Count	%
4.33106	60	1.59%
4.287383	60	1.59%
4.335656	60	1.59%
4.26494	60	1.59%
4.267489	60	1.59%
4.289043	60	1.59%
4.323694	60	1.59%
4.3258	60	1.59%
4.328898	60	1.59%
4.338557	60	1.59%

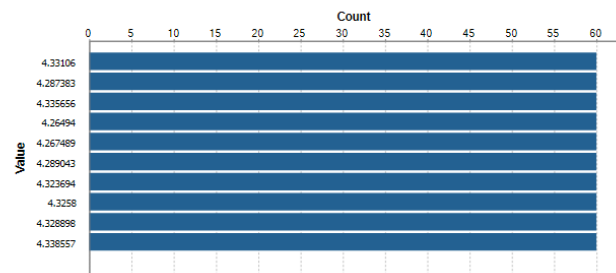


Figura F.4: Locais - Frequências do atributo 'park\_longitude'

- **park\_name** - parkingSite.name

Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	63	1.67%
Unique Count	0	0.00%
Duplicate Count	63	1.67%
Blank Count	0	0.00%

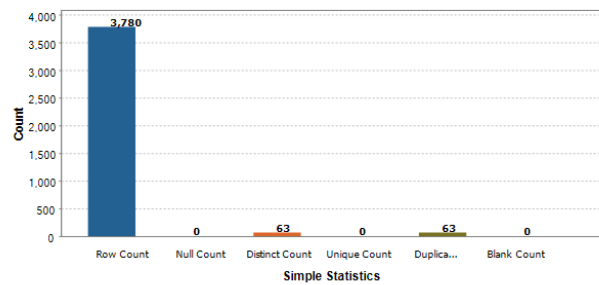


Figura F.5: Locais - Estatísticas Simples do atributo 'park\_name'

- **place\_latitude** - parkingSite.PointsOfInterests.geometry.location.latitude

Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	824	21.80%
Unique Count	321	8.49%
Duplicate Count	503	13.31%

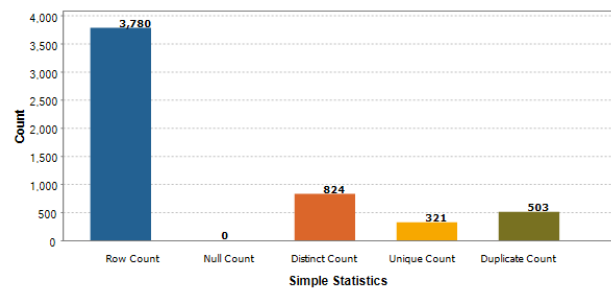
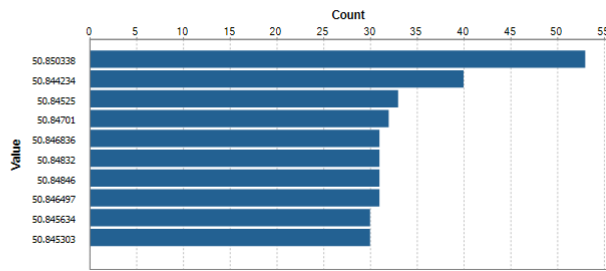


Figura F.6: Locais - Estatísticas Simples do atributo 'place\_latitude'



Value Frequency

Value	Count	%
50.850338	53	1.40%
50.844234	40	1.06%
50.84525	33	0.87%
50.84701	32	0.85%
50.846836	31	0.82%
50.84832	31	0.82%
50.84846	31	0.82%
50.846497	31	0.82%
50.845634	30	0.79%
50.845303	30	0.79%



Value Low Frequency

Value	Count	%
50.81242	1	0.03%
50.812786	1	0.03%
50.80741	1	0.03%
50.812984	1	0.03%
50.810474	1	0.03%
50.812508	1	0.03%
50.812378	1	0.03%
50.808315	1	0.03%
50.8091	1	0.03%
50.810745	1	0.03%

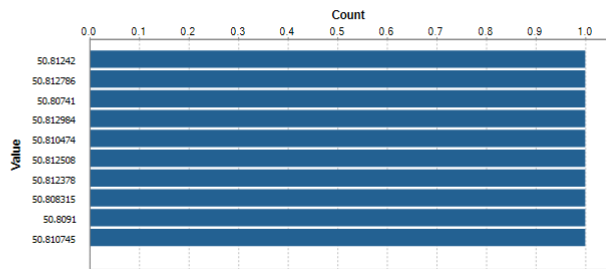


Figura F.7: Locais - Frequências do atributo 'place\_latitude'

- **place\_longitude** - parkingSite.PointsOfInterests.geometry.location.longitude

Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	847	22.41%
Unique Count	338	8.94%
Duplicate Count	509	13.47%

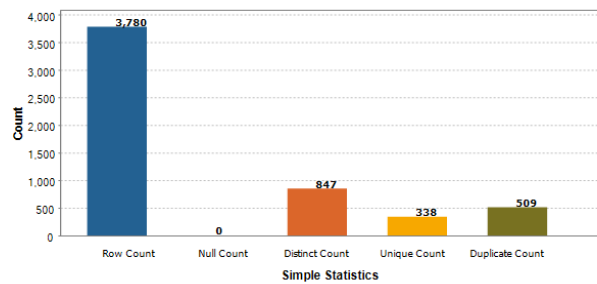
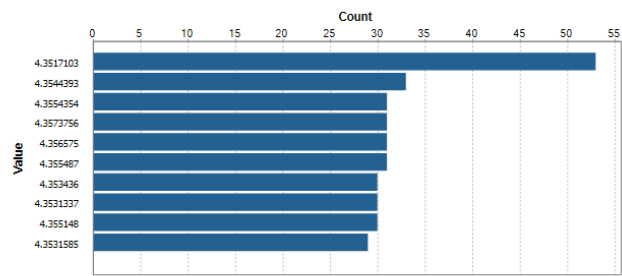


Figura F.8: Locais - Estatísticas Simples do atributo 'place\_longitude'

Value Frequency

Value	Count	%
4.3517103	53	1.40%
4.3544393	33	0.87%
4.3554354	31	0.82%
4.3573756	31	0.82%
4.356575	31	0.82%
4.355487	31	0.82%
4.353436	30	0.79%
4.3531337	30	0.79%
4.355148	30	0.79%
4.3531585	29	0.77%



Value Low Frequency

Value	Count	%
4.266324	1	0.03%
4.266295	1	0.03%
4.264495	1	0.03%
4.2640986	1	0.03%
4.2625127	1	0.03%
4.2650914	1	0.03%
4.265985	1	0.03%
4.2534275	1	0.03%
4.252749	1	0.03%
4.2660775	1	0.03%

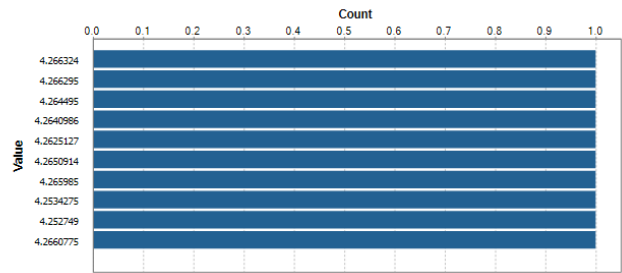


Figura F.9: Locais - Frequências do atributo 'place\_longitude'

- **place\_name** - parkingSite.PointsOfInterests.name

Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	869	22.99%
Unique Count	357	9.44%
Duplicate Count	512	13.54%
Blank Count	0	0.00%

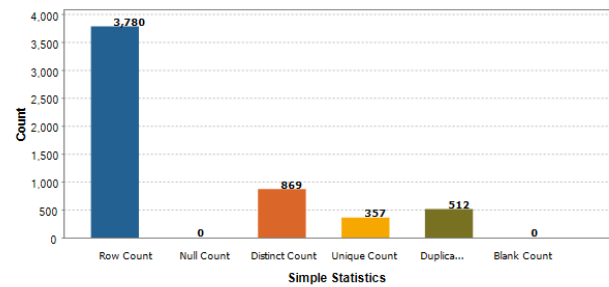
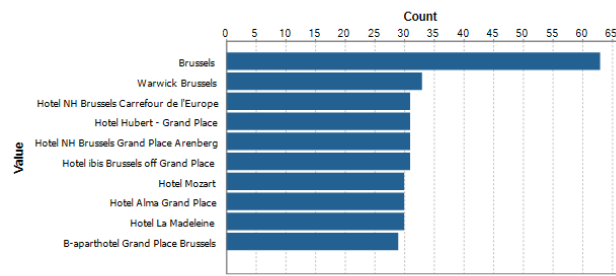


Figura F.10: Locais - Estatísticas Simples do atributo 'place\_name'

Value Frequency

Value	Count	%
Brussels	63	1.67%
Warwick Brussels	33	0.87%
Hotel NH Brussels Carrefou...	31	0.82%
Hotel Hubert - Grand Place	31	0.82%
Hotel NH Brussels Grand Pl...	31	0.82%
Hotel ibis Brussels off Gran...	31	0.82%
Hotel Mozart	30	0.79%
Hotel Alma Grand Place	30	0.79%
Hotel La Madeleine	30	0.79%
B-aparthotel Grand Place B...	29	0.77%



Value Low Frequency

Value	Count	%
ACTIVA	1	0.03%
A.C.E.	1	0.03%
AXA Assistance Broker Solu...	1	0.03%
A.D.I.R. Asbl	1	0.03%
A l'Entreprise Franky sprl	1	0.03%
ACV Nationaal	1	0.03%
AWEX - Agence wallonne à...	1	0.03%
AccessAndGo	1	0.03%
Action Service Company	1	0.03%
Administration communale...	1	0.03%

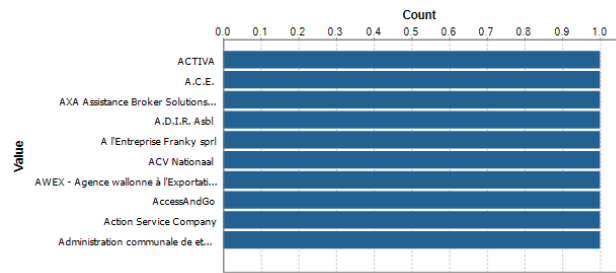


Figura F.11: Locais - Frequências do atributo 'place\_name'

- **place\_vicinity** - parkingSite.PointsOfInterests.vicinity

Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	779	20.61%
Unique Count	299	7.91%
Duplicate Count	480	12.70%
Blank Count	10	0.26%

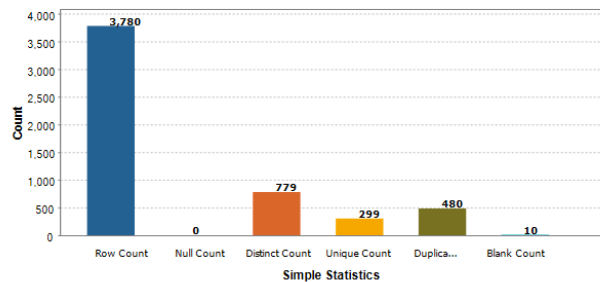


Figura F.12: Locais - Estatísticas Simples do atributo 'place\_vicinity'

Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	779	20.61%
Unique Count	299	7.91%
Duplicate Count	480	12.70%
Blank Count	10	0.26%

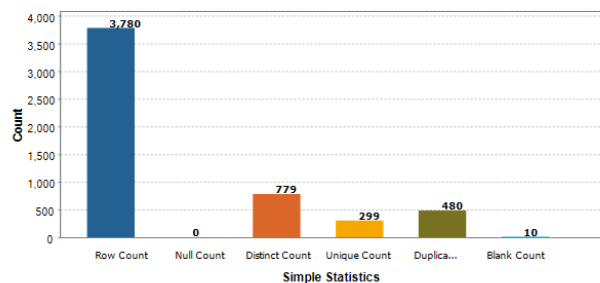


Figura F.13: Locais - Frequências do atributo 'place\_vicinity'

- **rating** - parkingSite.PointsOfInterests.rating

Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	37	0.98%
Unique Count	3	0.08%
Duplicate Count	34	0.90%
Blank Count	359	9.50%

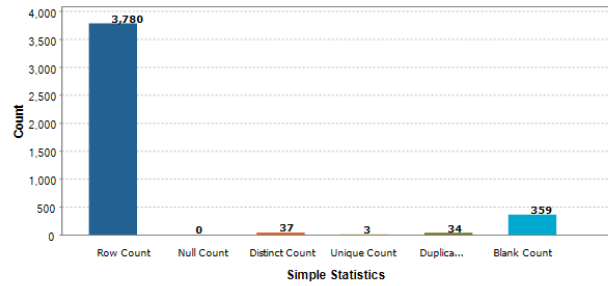
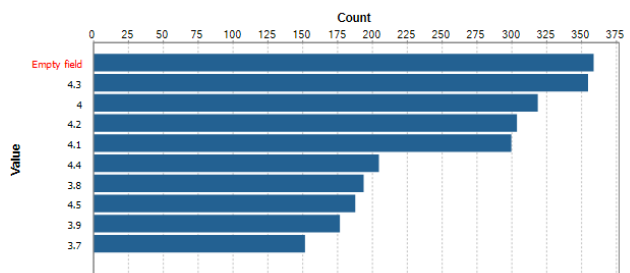


Figura F.14: Locais - Estatísticas Simples do atributo 'place\_rating'

Value Frequency

Value	Count	%
Empty field	359	9.50%
4.3	355	9.39%
4	319	8.44%
4.2	304	8.04%
4.1	300	7.94%
4.4	205	5.42%
3.8	194	5.13%
4.5	188	4.97%
3.9	177	4.68%
3.7	152	4.02%



Value Low Frequency

Value	Count	%
2.1	1	0.03%
1.7	1	0.03%
1.9	1	0.03%
1.5	3	0.08%
1.8	3	0.08%
2.2	5	0.13%
1	10	0.26%
2.8	11	0.29%
2.3	16	0.42%
2.6	20	0.53%

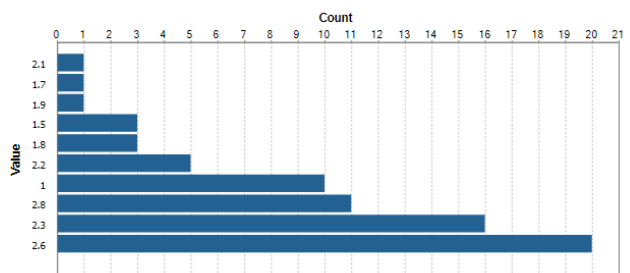


Figura F.15: Locais - Frequências do atributo 'place\_rating'

Summary Statistics

Label	Value
Mean	3.9358082432037413
Median	4.1
Inter Quartile Range	0.5999999999999996
Lower Quartile	3.7
Upper Quartile	4.3
Range	4.0
Minimum	1.0
Maximum	5.0

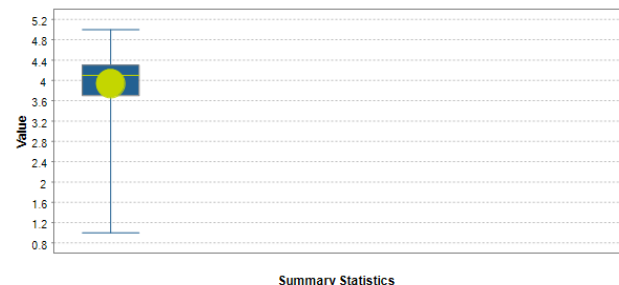


Figura F.16: Locais - Estatísticas do atributo 'place\_rating'

- **user\_ratings\_total** - parkingSite.PointsOfInterests.user\_ratings\_total

▼ Simple Statistics

Label	Count	%
Row Count	3780	100.00%
Null Count	0	0.00%
Distinct Count	300	7.94%
Unique Count	49	1.30%
Duplicate Count	251	6.64%
Blank Count	359	9.50%

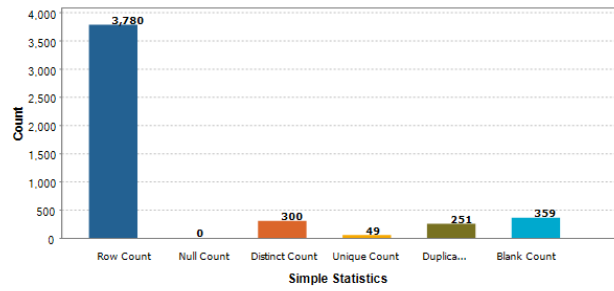
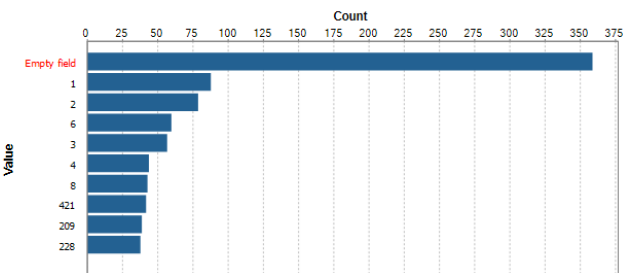


Figura F.17: Locais - Estatísticas Simples do atributo 'place\_user\_ratings\_total'

▼ Value Frequency

Value	Count	%
Empty field	359	9.50%
1	88	2.33%
2	79	2.09%
6	60	1.59%
3	57	1.51%
4	44	1.16%
8	43	1.14%
421	42	1.11%
209	39	1.03%
228	38	1.01%



▼ Value Low Frequency

Value	Count	%
143	1	0.03%
1128	1	0.03%
103	1	0.03%
125	1	0.03%
136	1	0.03%
104	1	0.03%
107	1	0.03%
118	1	0.03%
150	1	0.03%
142	1	0.03%

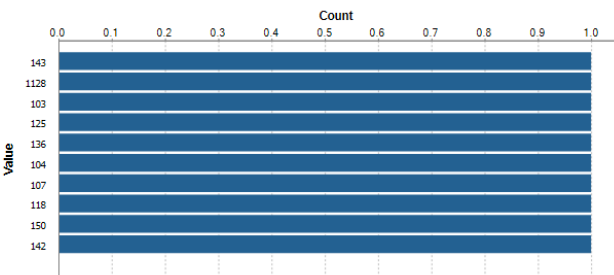


Figura F.18: Locais - Frequências do atributo 'place\_user\_ratings\_total'

▼ Summary Statistics

Label	Value
Mean	3.9358082432037413
Median	4.1
Inter Quartile Range	0.5999999999999996
Lower Quartile	3.7
Upper Quartile	4.3
Range	4.0
Minimum	1.0
Maximum	5.0

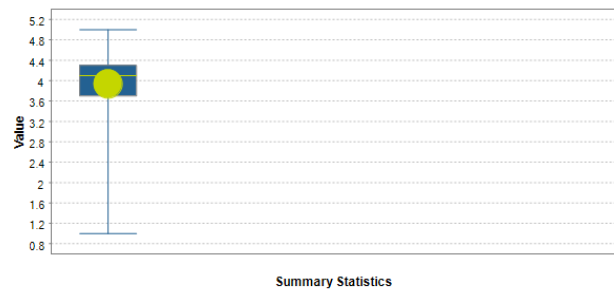


Figura F.19: Locais - Estatísticas do atributo 'place\_user\_ratings\_total'

## G Exploração dos dados da Meteorologia

- latitude

▼ Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	1	2.283E-3%
Unique Count	0	0.00%
Duplicate Count	1	2.283E-3%

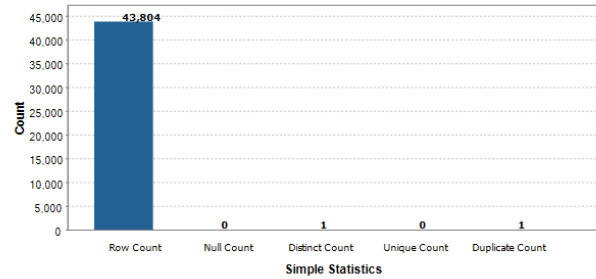
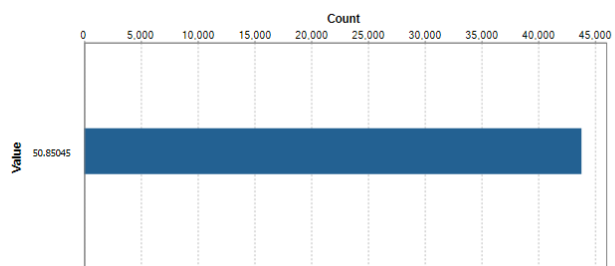


Figura G.1: Meteorologia - Estatísticas Simples do atributo 'latitude'

▼ Value Frequency

Value	Count	%
50.85045	43804	100.00%



▼ Value Low Frequency

Value	Count	%
50.85045	43804	100.00%

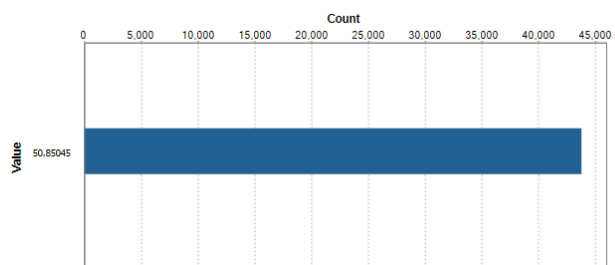


Figura G.2: Meteorologia - Frequências do atributo 'latitude'

- longitude

▼ Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	1	2.283E-3%
Unique Count	0	0.00%
Duplicate Count	1	2.283E-3%

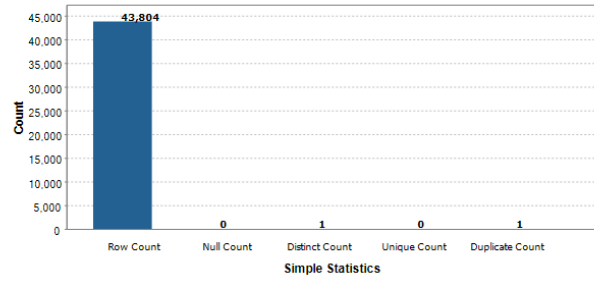
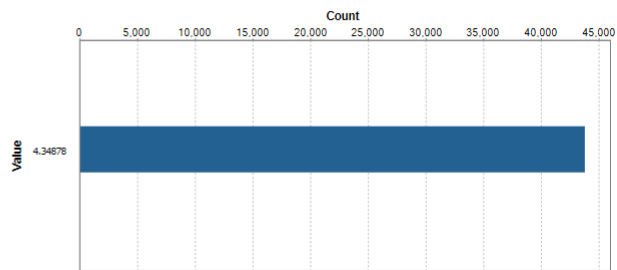


Figura G.3: Meteorologia - Estatísticas Simples do atributo 'latitude'

▼ Value Frequency

Value	Count	%
4.34878	43804	100.00%



▼ Value Low Frequency

Value	Count	%
4.34878	43804	100.00%

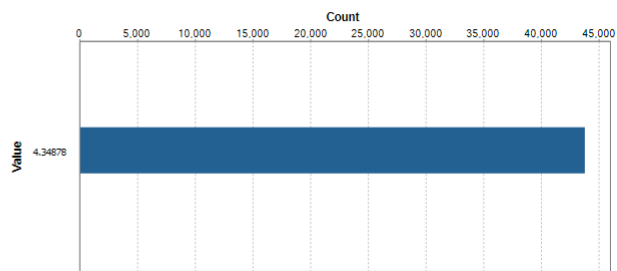


Figura G.4: Meteorologia - Frequências do atributo 'longitude'

- **hourly\_time** - hourly.data[n].time

▼ Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	43804	100.00%
Unique Count	43804	100.00%
Duplicate Count	0	0.00%

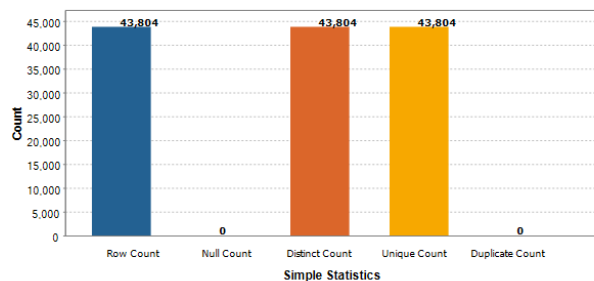


Figura G.5: Meteorologia - Estatísticas Simples do atributo 'hourly\_time'

- **hourly\_summary** - hourly.data[n].summary

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	30	0.07%
Unique Count	2	4.566E-3%
Duplicate Count	28	0.06%
Blank Count	2219	5.07%

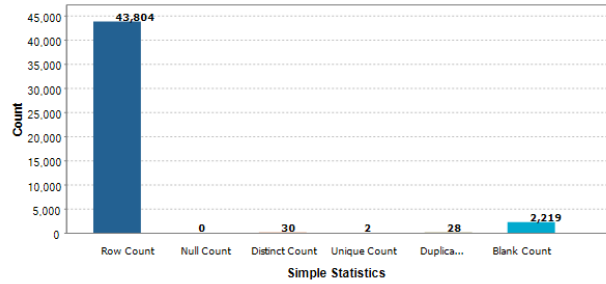
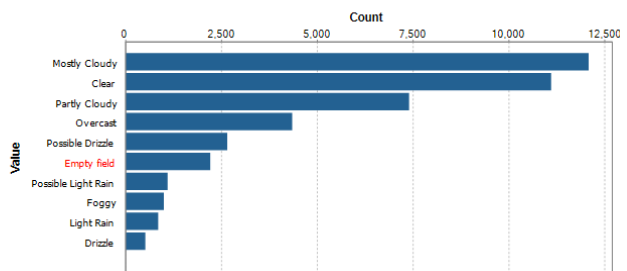


Figura G.6: Meteorologia - Estatísticas Simples do atributo 'hourly\_summary'

Value Frequency

Value	Count	%
Mostly Cloudy	12094	27.61%
Clear	11115	25.37%
Partly Cloudy	7408	16.91%
Overcast	4357	9.95%
Possible Drizzle	2662	6.08%
Empty field	2219	5.07%
Possible Light Rain	1106	2.52%
Foggy	1007	2.30%
Light Rain	861	1.97%
Drizzle	523	1.19%



Value Low Frequency

Value	Count	%
Humid and Partly Cloudy	1	2.283E-3%
Windy and Foggy	1	2.283E-3%
Possible Light Rain and Da...	2	4.566E-3%
Dangerously Windy and Ov...	3	6.849E-3%
Heavy Rain	3	6.849E-3%
Humid	4	9.132E-3%
Flurries	4	9.132E-3%
Possible Drizzle and Dange...	5	0.01%
Windy	5	0.01%
Windy and Partly Cloudy	5	0.01%

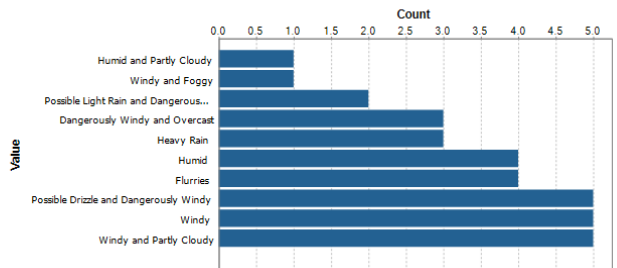


Figura G.7: Meteorologia - Frequências do atributo 'hourly\_summary'

hourly\_precipIntensity - hourly.data[n].precipIntensity

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	4916	11.82%
Unique Count	2806	6.74%
Duplicate Count	2110	5.07%

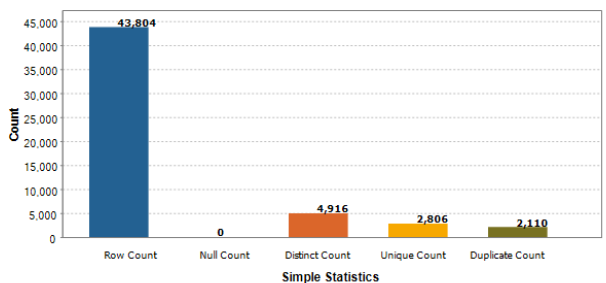
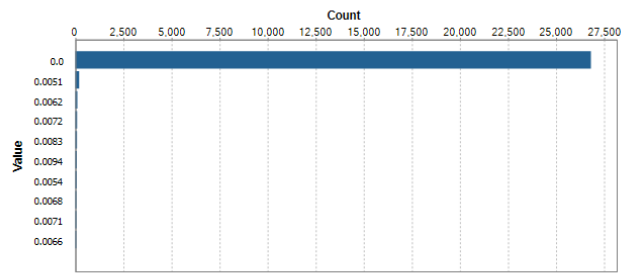


Figura G.8: Meteorologia - Estatísticas Simples do atributo 'hourly\_precipIntensity'



Value Frequency

Value	Count	%
0.0	26817	64.46%
0.0051	197	0.47%
0.0062	106	0.25%
0.0072	95	0.23%
0.0083	86	0.21%
0.0094	79	0.19%
0.0054	71	0.17%
0.0068	62	0.15%
0.0071	60	0.14%
0.0066	55	0.13%



Value Low Frequency

Value	Count	%
0.0338	1	2.404E-3%
0.0481	1	2.404E-3%
0.0437	1	2.404E-3%
0.0645	1	2.404E-3%
0.074	1	2.404E-3%
0.0357	1	2.404E-3%
0.0635	1	2.404E-3%
0.0694	1	2.404E-3%
0.0697	1	2.404E-3%
0.0665	1	2.404E-3%

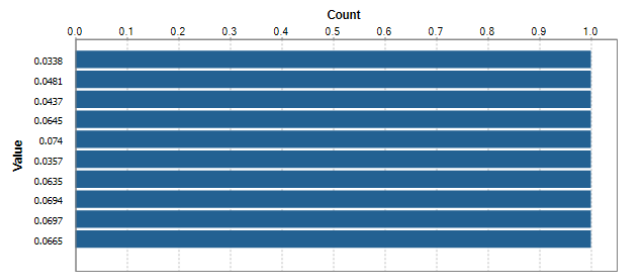


Figura G.9: Meteorologia - Frequências do atributo 'hourly\_precipIntensity'

Summary Statistics

Label	Value
Mean	0.07832104946877554
Median	0.0
Inter Quartile Range	0.0201
Lower Quartile	0.0
Upper Quartile	0.0201
Range	27.3784
Minimum	0.0
Maximum	27.3784

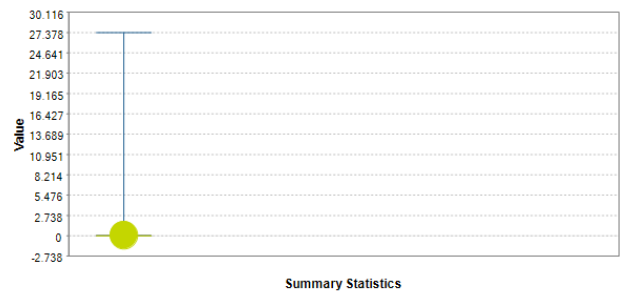


Figura G.10: Meteorologia - Estatísticas do atributo 'hourly\_precipIntensity'

- **hourly\_precipProbability** - hourly.data[n].precipProbability

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	95	0.23%
Unique Count	5	0.01%
Duplicate Count	90	0.22%

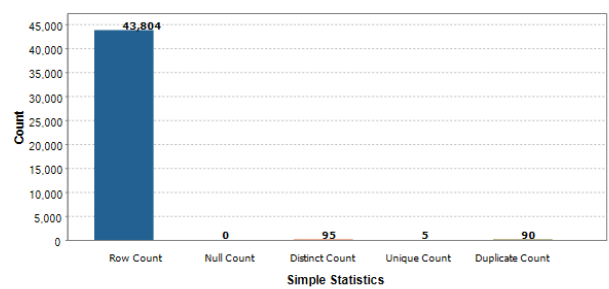
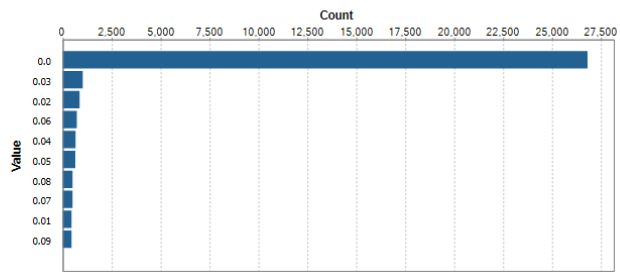


Figura G.11: Meteorologia - Estatísticas Simples do atributo 'hourly\_precipProbability'

Value Frequency

Value	Count	%
0.0	26817	64.46%
0.03	1024	2.46%
0.02	857	2.06%
0.06	716	1.72%
0.04	657	1.58%
0.05	644	1.55%
0.08	501	1.20%
0.07	498	1.20%
0.01	447	1.07%
0.09	447	1.07%



Value Low Frequency

Value	Count	%
0.9	1	2.404E-3%
0.95	1	2.404E-3%
0.92	1	2.404E-3%
0.93	1	2.404E-3%
0.94	1	2.404E-3%
0.89	2	4.807E-3%
0.88	3	7.211E-3%
0.86	4	9.615E-3%
0.87	4	9.615E-3%
0.85	9	0.02%

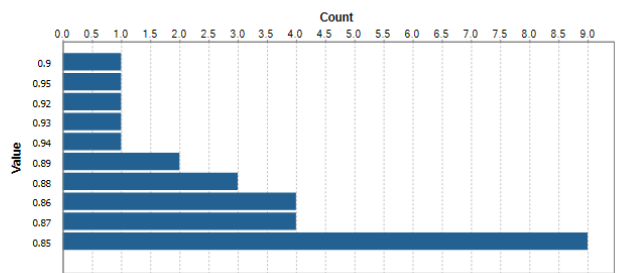


Figura G.12: Meteorologia - Frequências do atributo 'hourly\_precipProbability'

Summary Statistics

Label	Value
Mean	0.08944786308350561
Median	0.0
Inter Quartile Range	0.07
Lower Quartile	0.0
Upper Quartile	0.07
Range	0.95
Minimum	0.0
Maximum	0.95

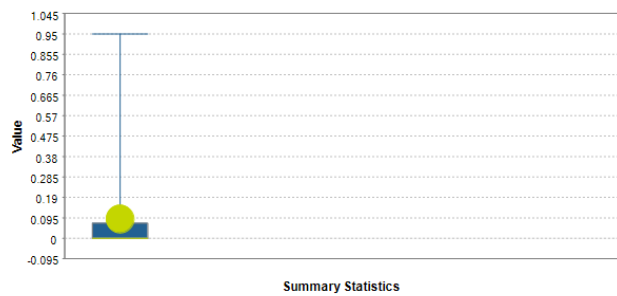


Figura G.13: Meteorologia - Estatísticas do atributo 'hourly\_precipProbability'

- **hourly\_temperature** - hourly.data[n].temperature

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	3548	8.10%
Unique Count	386	0.88%
Duplicate Count	3162	7.22%

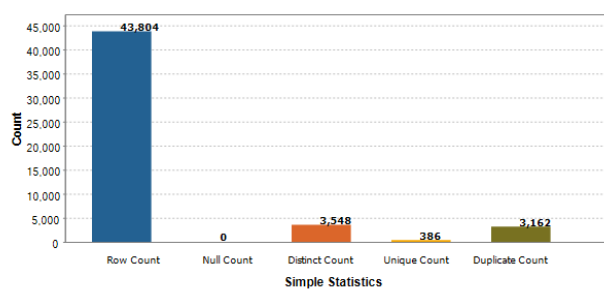
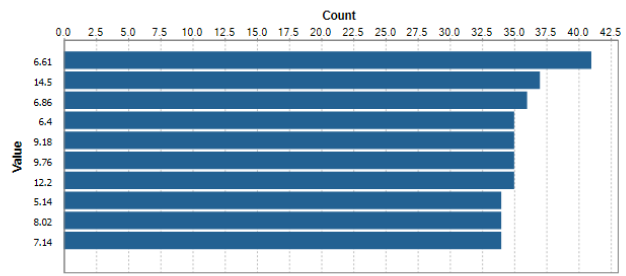


Figura G.14: Meteorologia - Estatísticas Simples do atributo 'hourly\_temperature'

Value Frequency

Value	Count	%
6.61	41	0.09%
14.5	37	0.08%
6.86	36	0.08%
6.4	35	0.08%
9.18	35	0.08%
9.76	35	0.08%
12.2	35	0.08%
5.14	34	0.08%
8.02	34	0.08%
7.14	34	0.08%



Value Low Frequency

Value	Count	%
-6.49	1	2.283E-3%
-8.19	1	2.283E-3%
-7.67	1	2.283E-3%
-6.55	1	2.283E-3%
-7.62	1	2.283E-3%
-8.15	1	2.283E-3%
-8.07	1	2.283E-3%
-6.62	1	2.283E-3%
-7.04	1	2.283E-3%
-6.94	1	2.283E-3%

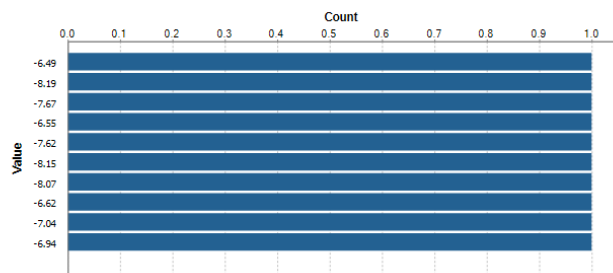


Figura G.15: Meteorologia - Frequências do atributo 'hourly\_temperature'

Summary Statistics

Label	Value
Mean	11.653023377927948
Median	11.29
Inter Quartile Range	10.0
Lower Quartile	6.46
Upper Quartile	16.46
Range	46.61
Minimum	-8.19
Maximum	38.42

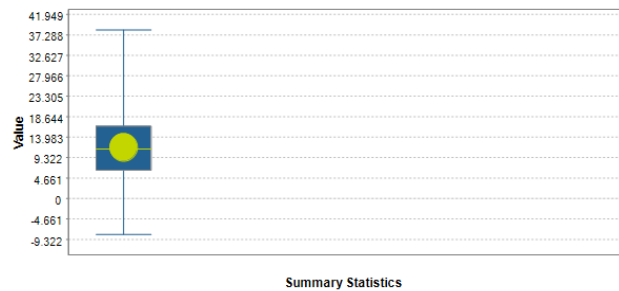


Figura G.16: Meteorologia - Estatísticas do atributo 'hourly\_temperature'

hourly\_apparentTemperature - hourly.data[n].apparentTemperature

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	3948	9.01%
Unique Count	473	1.08%
Duplicate Count	3475	7.93%

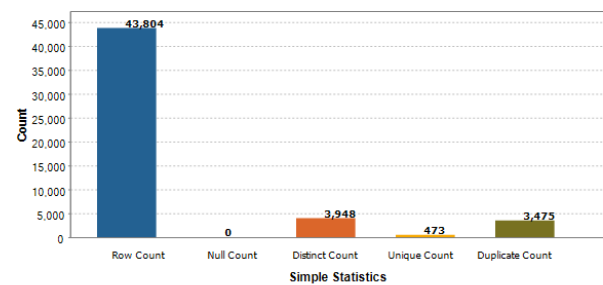
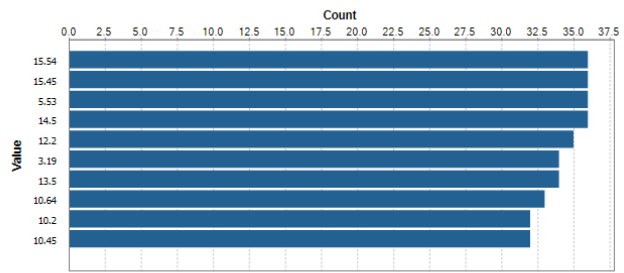


Figura G.17: Meteorologia - Estatísticas Simples do atributo 'hourly\_apparentTemperature'

Value Frequency

Value	Count	%
15.54	36	0.08%
15.45	36	0.08%
5.53	36	0.08%
14.5	36	0.08%
12.2	35	0.08%
3.19	34	0.08%
13.5	34	0.08%
10.64	33	0.08%
10.2	32	0.07%
10.45	32	0.07%



Value Low Frequency

Value	Count	%
-12.79	1	2.283E-3%
-12.48	1	2.283E-3%
-13.35	1	2.283E-3%
-13.0	1	2.283E-3%
-12.6	1	2.283E-3%
-13.5	1	2.283E-3%
-12.49	1	2.283E-3%
-12.63	1	2.283E-3%
-12.52	1	2.283E-3%
-12.68	1	2.283E-3%

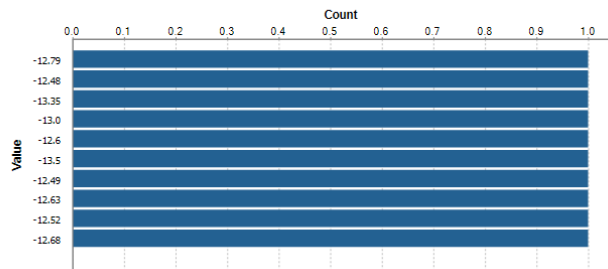


Figura G.18: Meteorologia - Frequências do atributo 'hourly\_apparentTemperature'

Summary Statistics

Label	Value
Mean	10.718571526414319
Median	11.29
Inter Quartile Range	12.170000000000002
Lower Quartile	4.31
Upper Quartile	16.48
Range	52.33
Minimum	-13.5
Maximum	38.83

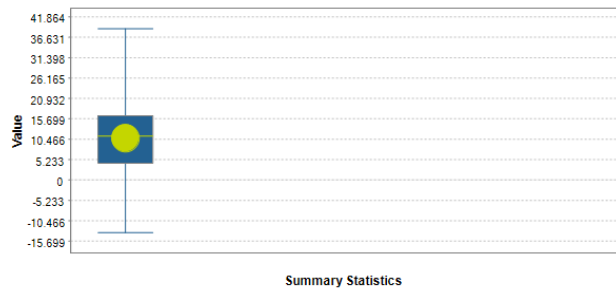


Figura G.19: Meteorologia - Estatísticas do atributo 'hourly\_apparentTemperature'

hourly\_dewPoint - hourly.data[n].dewPoint

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	2919	6.66%
Unique Count	385	0.88%
Duplicate Count	2534	5.78%

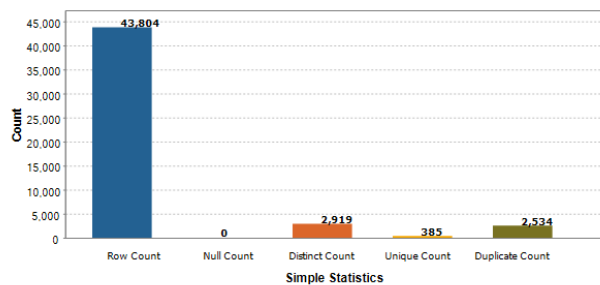
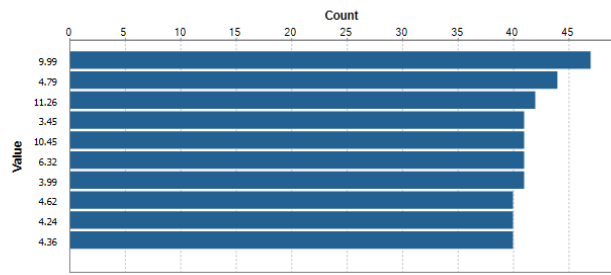


Figura G.20: Meteorologia - Estatísticas Simples do atributo 'hourly\_dewPoint'

Value Frequency

Value	Count	%
9.99	47	0.11%
4.79	44	0.10%
11.26	42	0.10%
3.45	41	0.09%
10.45	41	0.09%
6.32	41	0.09%
3.99	41	0.09%
4.62	40	0.09%
4.24	40	0.09%
4.36	40	0.09%



Value Low Frequency

Value	Count	%
-15.07	1	2.283E-3%
-14.81	1	2.283E-3%
-15.17	1	2.283E-3%
-15.59	1	2.283E-3%
-15.48	1	2.283E-3%
-15.43	1	2.283E-3%
-15.16	1	2.283E-3%
-14.91	1	2.283E-3%
-15.19	1	2.283E-3%
-14.5	1	2.283E-3%

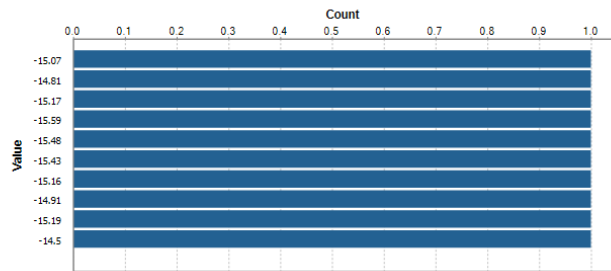


Figura G.21: Meteorologia - Frequências do atributo 'hourly\_dewPoint'

Summary Statistics

Label	Value
Mean	7.0800404072687435
Median	7.21
Inter Quartile Range	8.3
Lower Quartile	3.03
Upper Quartile	11.33
Range	38.6
Minimum	-15.59
Maximum	23.01

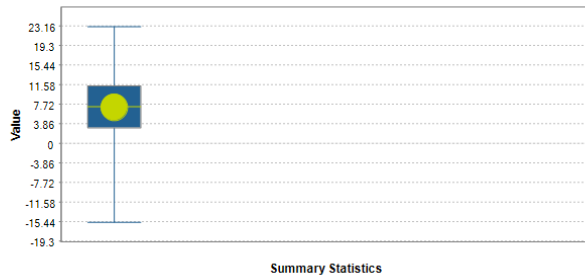


Figura G.22: Meteorologia - Estatísticas do atributo 'hourly\_dewPoint'

hourly\_humidity - hourly.data[n].humidity

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	83	0.19%
Unique Count	2	4.566E-3%
Duplicate Count	81	0.18%

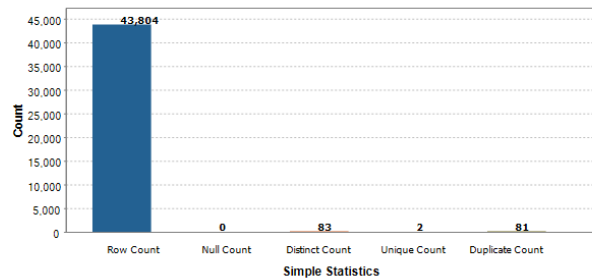
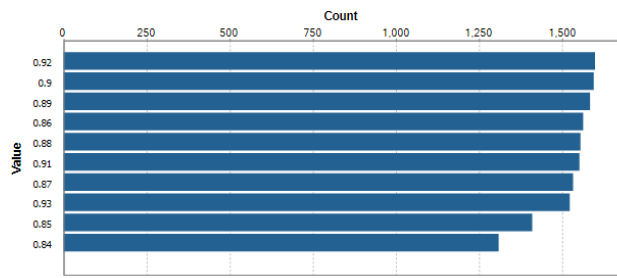


Figura G.23: Meteorologia - Estatísticas Simples do atributo 'hourly\_humidity'

Value Frequency

Value	Count	%
0.92	1599	3.65%
0.9	1595	3.64%
0.89	1584	3.62%
0.86	1563	3.57%
0.88	1555	3.55%
0.91	1552	3.54%
0.87	1533	3.50%
0.93	1523	3.48%
0.85	1410	3.22%
0.84	1309	2.99%



Value Low Frequency

Value	Count	%
0.15	1	2.283E-3%
0.22	1	2.283E-3%
0.2	2	4.566E-3%
0.18	2	4.566E-3%
0.21	2	4.566E-3%
0.23	5	0.01%
0.24	7	0.02%
0.25	17	0.04%
0.27	22	0.05%
0.26	23	0.05%

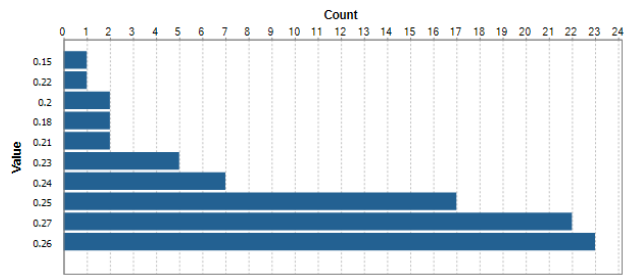


Figura G.24: Meteorologia - Frequências do atributo 'hourly\_humidity'

Summary Statistics

Label	Value
Mean	0.7567065430802247
Median	0.8
Inter Quartile Range	0.21999999999999997
Lower Quartile	0.66
Upper Quartile	0.88
Range	0.85
Minimum	0.15
Maximum	1.0

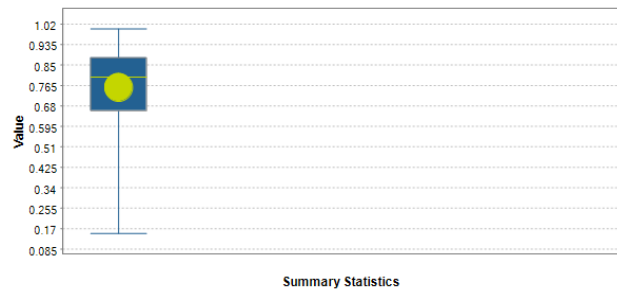


Figura G.25: Meteorologia - Estatísticas do atributo 'hourly\_humidity'

hourly\_pressure - hourly.data[n].pressure

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	684	1.65%
Unique Count	52	0.13%
Duplicate Count	632	1.52%

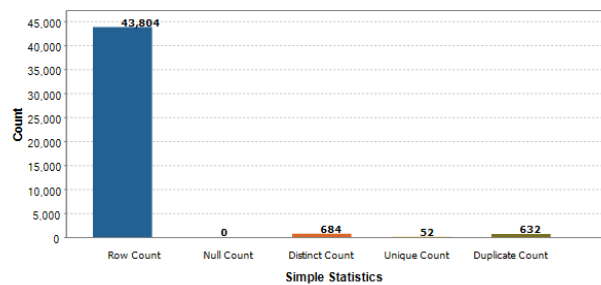
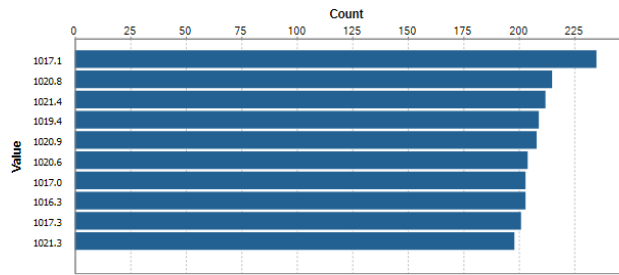


Figura G.26: Meteorologia - Estatísticas Simples do atributo 'hourly\_pressure'

Value Frequency

Value	Count	%
1017.1	235	0.57%
1020.8	215	0.52%
1021.4	212	0.51%
1019.4	209	0.50%
1020.9	208	0.50%
1020.6	204	0.49%
1017.0	203	0.49%
1016.3	203	0.49%
1017.3	201	0.48%
1021.3	198	0.48%



Value Low Frequency

Value	Count	%
972.8	1	2.409E-3%
974.3	1	2.409E-3%
975.0	1	2.409E-3%
973.1	1	2.409E-3%
974.6	1	2.409E-3%
973.9	1	2.409E-3%
974.9	1	2.409E-3%
974.2	1	2.409E-3%
974.7	1	2.409E-3%
971.7	1	2.409E-3%

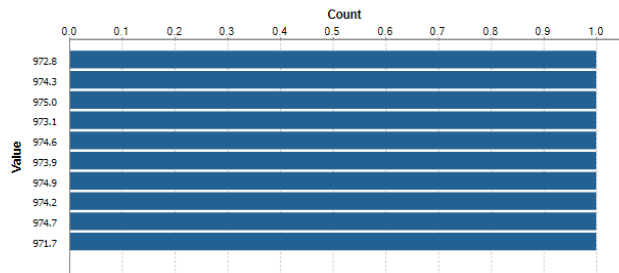


Figura G.27: Meteorologia - Frequências do atributo 'hourly\_pressure'

Summary Statistics

Label	Value
Mean	1016.5891289685408
Median	1017.4
Inter Quartile Range	12.300000000000068
Lower Quartile	1010.9
Upper Quartile	1023.2
Range	74.299999999999995
Minimum	971.7
Maximum	1046.0



Summary Statistics

Figura G.28: Meteorologia - Estatísticas do atributo 'hourly\_pressure'

- hourly\_windSpeed - hourly.data[n].windSpeed

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	1123	2.56%
Unique Count	112	0.26%
Duplicate Count	1011	2.31%

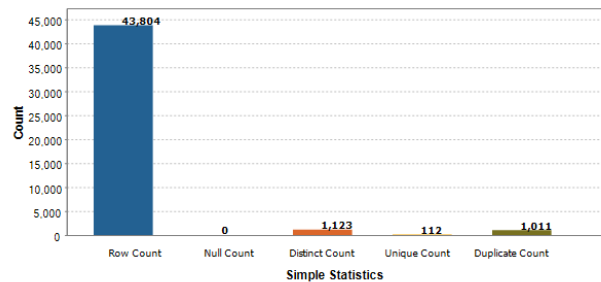
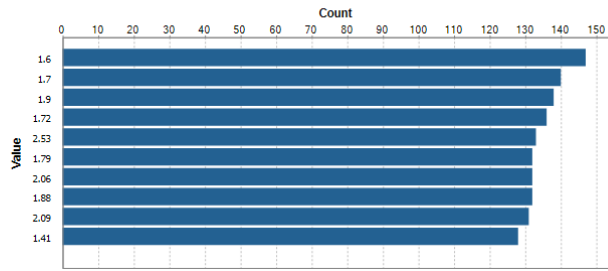


Figura G.29: Meteorologia - Estatísticas Simples do atributo 'hourly\_windSpeed'

Value Frequency

Value	Count	%
1.6	147	0.34%
1.7	140	0.32%
1.9	138	0.32%
1.72	136	0.31%
2.53	133	0.30%
1.79	132	0.30%
2.06	132	0.30%
1.88	132	0.30%
2.09	131	0.30%
1.41	128	0.29%



Value Low Frequency

Value	Count	%
9.06	1	2.283E-3%
8.42	1	2.283E-3%
9.42	1	2.283E-3%
8.23	1	2.283E-3%
9.23	1	2.283E-3%
8.47	1	2.283E-3%
9.24	1	2.283E-3%
0.03	1	2.283E-3%
9.44	1	2.283E-3%
8.08	1	2.283E-3%

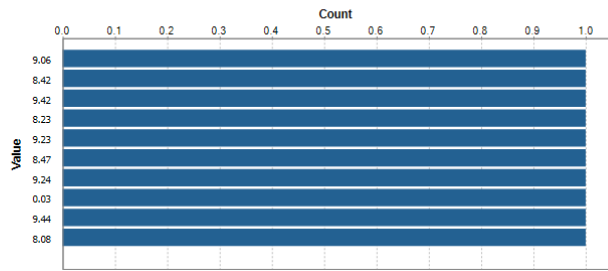


Figura G.30: Meteorologia - Frequências do atributo 'hourly\_windSpeed'

Summary Statistics

Label	Value
Mean	3.1259550176960844
Median	2.76
Inter Quartile Range	2.37
Lower Quartile	1.74
Upper Quartile	4.11
Range	14.520000000000001
Minimum	0.03
Maximum	14.55

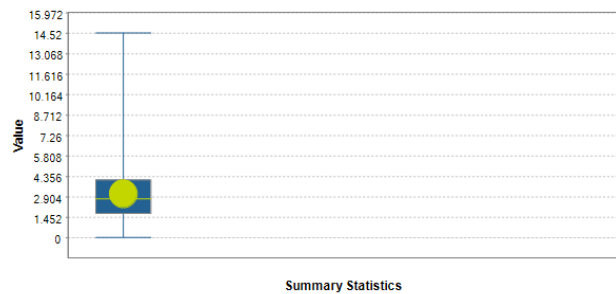


Figura G.31: Meteorologia - Estatísticas do atributo 'hourly\_windSpeed'

hourly\_windGust - hourly.data[n].windGust

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	1870	4.50%
Unique Count	171	0.41%
Duplicate Count	1699	4.09%

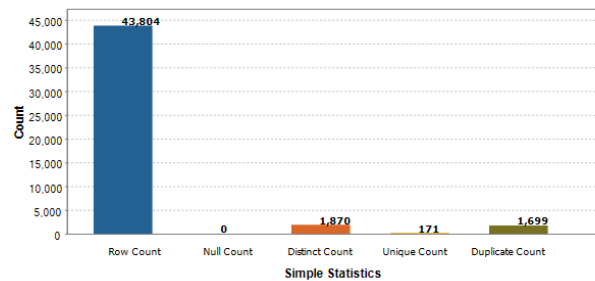
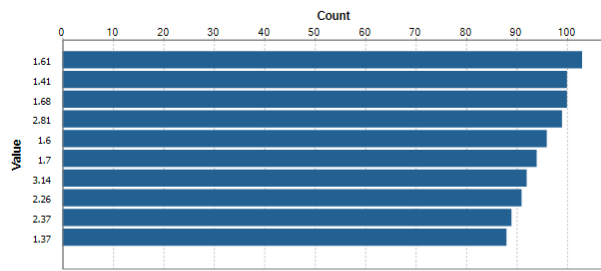


Figura G.32: Meteorologia - Estatísticas Simples do atributo 'hourly\_windGust'



Value Frequency

Value	Count	%
1.61	103	0.25%
1.41	100	0.24%
1.68	100	0.24%
2.81	99	0.24%
1.6	96	0.23%
1.7	94	0.23%
3.14	92	0.22%
2.26	91	0.22%
2.37	89	0.21%
1.37	88	0.21%



Value Low Frequency

Value	Count	%
11.03	1	2.406E-3%
13.48	1	2.406E-3%
14.25	1	2.406E-3%
12.82	1	2.406E-3%
12.66	1	2.406E-3%
13.96	1	2.406E-3%
13.85	1	2.406E-3%
0.03	1	2.406E-3%
13.02	1	2.406E-3%
13.54	1	2.406E-3%

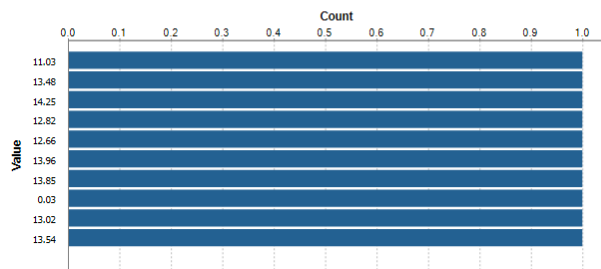


Figura G.33: Meteorologia - Frequências do atributo 'hourly\_windGust'

Summary Statistics

Label	Value
Mean	4.511364172139232
Median	3.48
Inter Quartile Range	3.65
Lower Quartile	2.03
Upper Quartile	5.68
Range	23.66
Minimum	0.03
Maximum	23.69

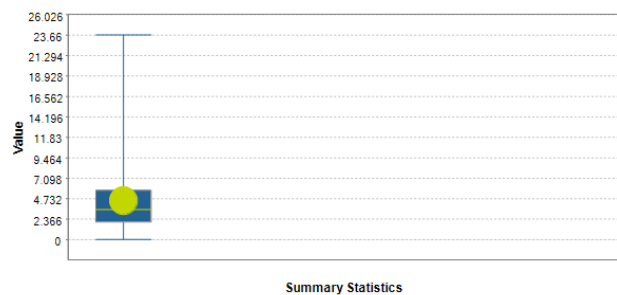


Figura G.34: Meteorologia - Estatísticas do atributo 'hourly\_windGust'

hourly\_windBearing - hourly.data[n].windBearing

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	360	0.82%
Unique Count	0	0.00%
Duplicate Count	360	0.82%

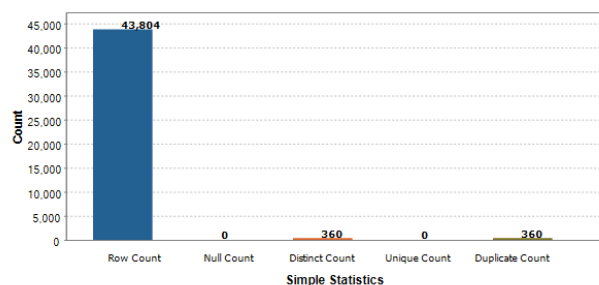
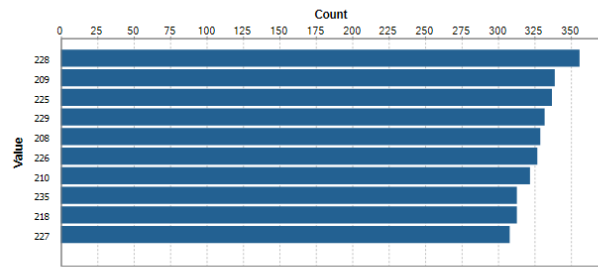


Figura G.35: Meteorologia - Estatísticas Simples do atributo 'hourly\_windBearing'

Value Frequency

Value	Count	%
228	356	0.81%
209	339	0.77%
225	337	0.77%
229	332	0.76%
208	329	0.75%
226	327	0.75%
210	322	0.74%
235	313	0.71%
218	313	0.71%
227	308	0.70%



Value Low Frequency

Value	Count	%
93	44	0.10%
111	46	0.11%
98	49	0.11%
131	49	0.11%
140	49	0.11%
96	50	0.11%
103	50	0.11%
121	50	0.11%
137	50	0.11%
89	51	0.12%

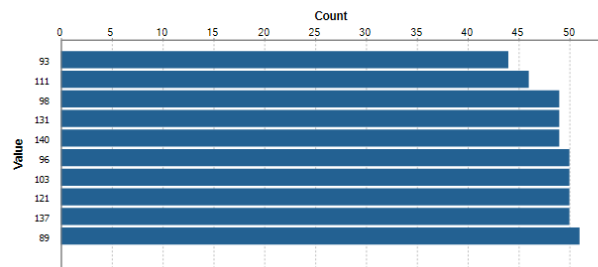


Figura G.36: Meteorologia - Frequências do atributo 'hourly\_windBearing'

Summary Statistics

Label	Value
Mean	192.2646420824295
Median	209.0
Inter Quartile Range	110.0
Lower Quartile	137.0
Upper Quartile	247.0
Range	359.0
Minimum	0
Maximum	359

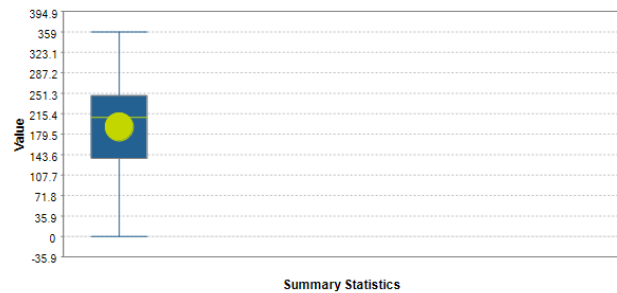


Figura G.37: Meteorologia - Estatísticas do atributo 'hourly\_windBearing'

hourly\_cloudCover - hourly.data[n].cloudCover

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	101	0.23%
Unique Count	0	0.00%
Duplicate Count	101	0.23%

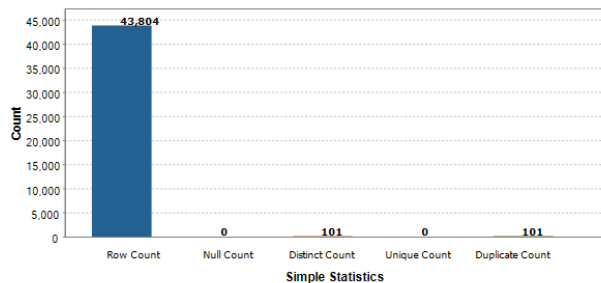
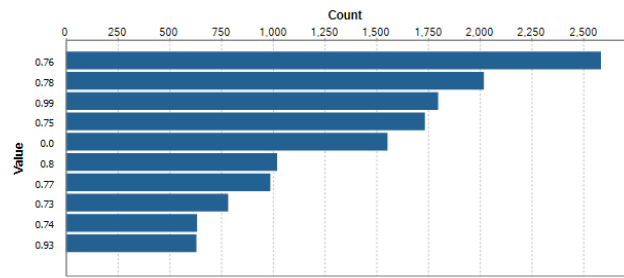


Figura G.38: Meteorologia - Estatísticas Simples do atributo 'hourly\_cloudCover'

▼ Value Frequency

Value	Count	%
0.76	2586	5.95%
0.78	2020	4.65%
0.99	1799	4.14%
0.75	1735	3.99%
0.0	1554	3.58%
0.8	1021	2.35%
0.77	988	2.27%
0.73	784	1.80%
0.74	634	1.46%
0.93	631	1.45%



▼ Value Low Frequency

Value	Count	%
0.09	101	0.23%
0.08	104	0.24%
0.07	118	0.27%
0.06	119	0.27%
0.04	160	0.37%
0.91	162	0.37%
0.9	167	0.38%
0.59	171	0.39%
0.57	188	0.43%
0.58	189	0.44%

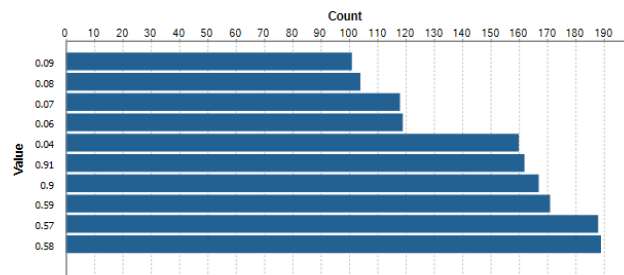


Figura G.39: Meteorologia - Frequências do atributo 'hourly\_cloudCover'

▼ Summary Statistics

Label	Value
Mean	0.5580657934114506
Median	0.64
Inter Quartile Range	0.5
Lower Quartile	0.28
Upper Quartile	0.78
Range	1.0
Minimum	0.0
Maximum	1.0

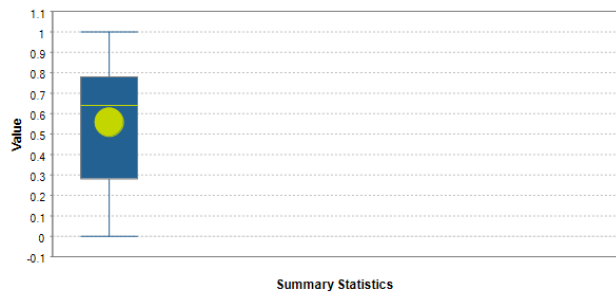


Figura G.40: Meteorologia - Estatísticas do atributo 'hourly\_cloudCover'

• **hourly\_uvIndex** - hourly.data[n].uvIndex

▼ Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	10	0.02%
Unique Count	0	0.00%
Duplicate Count	10	0.02%

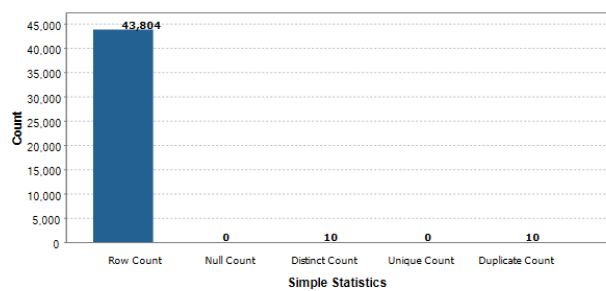
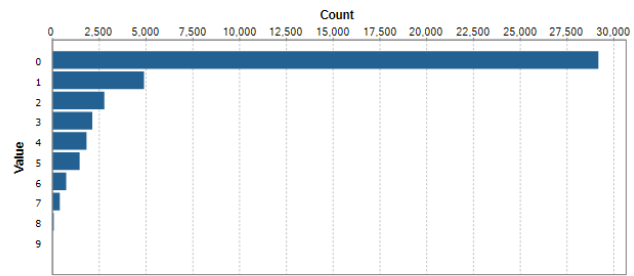


Figura G.41: Meteorologia - Estatísticas Simples do atributo 'hourly\_uvIndex'

Value Frequency

Value	Count	%
0	29202	66.86%
1	4918	11.26%
2	2800	6.41%
3	2157	4.94%
4	1849	4.23%
5	1479	3.39%
6	758	1.74%
7	416	0.95%
8	96	0.22%
9	4	9.158E-3%



Value Low Frequency

Value	Count	%
9	4	9.158E-3%
8	96	0.22%
7	416	0.95%
6	758	1.74%
5	1479	3.39%
4	1849	4.23%
3	2157	4.94%
2	2800	6.41%
1	4918	11.26%
0	29202	66.86%

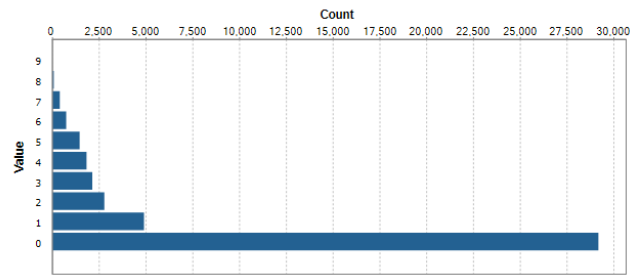


Figura G.42: Meteorologia - Frequências do atributo 'hourly\_uvIndex'

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	10	0.02%
Unique Count	0	0.00%
Duplicate Count	10	0.02%

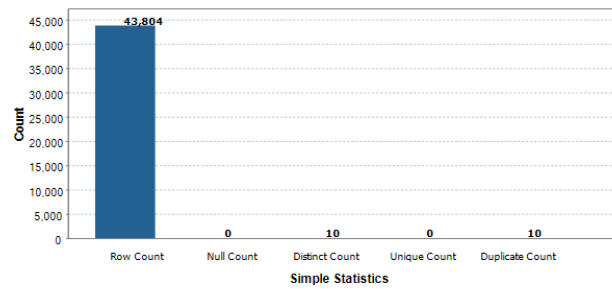


Figura G.43: Meteorologia - Estatísticas do atributo 'hourly\_uvIndex'

hourly\_visibility - hourly.data[n].visibility

Simple Statistics

Label	Count	%
Row Count	43804	100.00%
Null Count	0	0.00%
Distinct Count	5900	13.47%
Unique Count	3529	8.06%
Duplicate Count	2371	5.41%

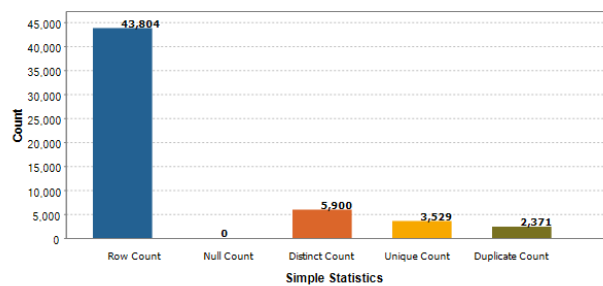
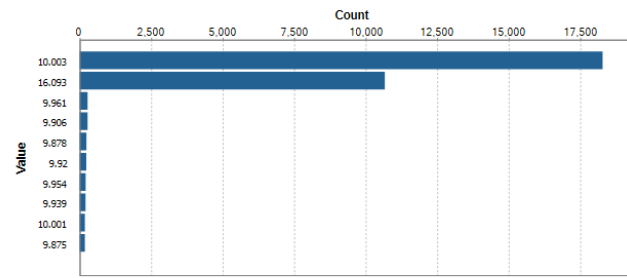


Figura G.44: Meteorologia - Estatísticas Simples do atributo 'hourly\_visibility'

Value Frequency

Value	Count	%
10.003	18286	41.75%
16.093	10672	24.37%
9.961	279	0.64%
9.906	278	0.63%
9.878	238	0.54%
9.92	232	0.53%
9.954	211	0.48%
9.939	206	0.47%
10.001	184	0.42%
9.875	183	0.42%



Value Low Frequency

Value	Count	%
0.0	1	2.283E-3%
0.089	1	2.283E-3%
0.126	1	2.283E-3%
0.116	1	2.283E-3%
0.054	1	2.283E-3%
0.07	1	2.283E-3%
0.127	1	2.283E-3%
0.111	1	2.283E-3%
0.113	1	2.283E-3%
0.065	1	2.283E-3%

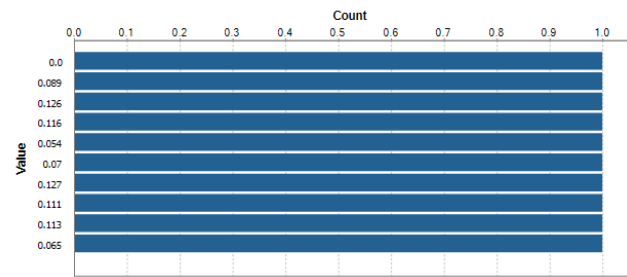


Figura G.45: Meteorologia - Frequências do atributo 'hourly\_visibility'

Summary Statistics

Label	Value
Mean	10.736197469973057
Median	10.003
Inter Quartile Range	3.5439999999999987
Lower Quartile	9.784
Upper Quartile	13.328
Range	16.093
Minimum	0.0
Maximum	16.093

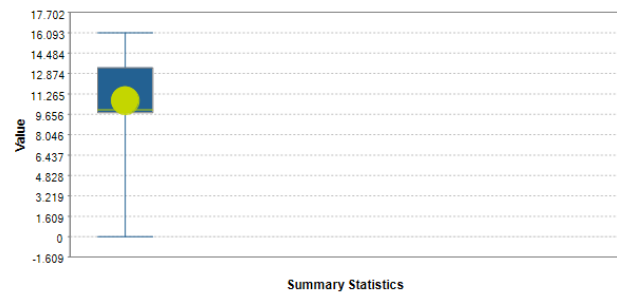


Figura G.46: Meteorologia - Estatísticas do atributo 'hourly\_visibility'

## H Exploração dos dados dos Feriados

- **year** - year

Simple Statistics

Label	Count	%
Row Count	114	100.00%
Null Count	0	0.00%
Distinct Count	6	5.26%
Unique Count	0	0.00%
Duplicate Count	6	5.26%

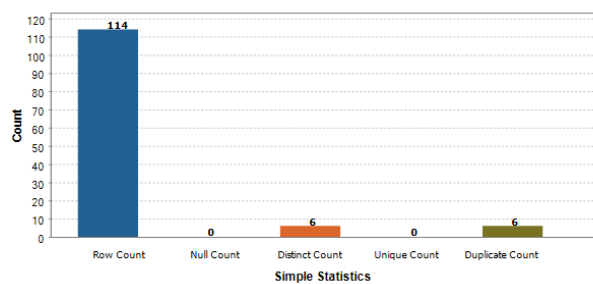
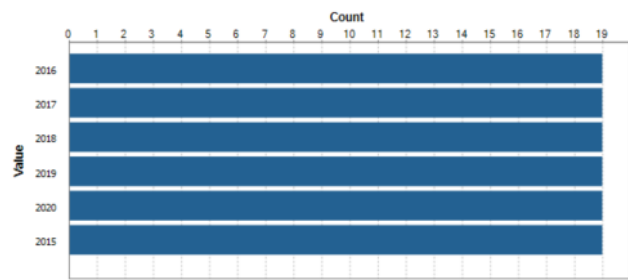


Figura H.1: Feriados - Estatísticas Simples do atributo 'year'

▼ Value Frequency

Value	Count	%
2016	19	16.67%
2017	19	16.67%
2018	19	16.67%
2019	19	16.67%
2020	19	16.67%
2015	19	16.67%



▼ Value Low Frequency

Value	Count	%
2016	19	16.67%
2017	19	16.67%
2018	19	16.67%
2019	19	16.67%
2020	19	16.67%
2015	19	16.67%

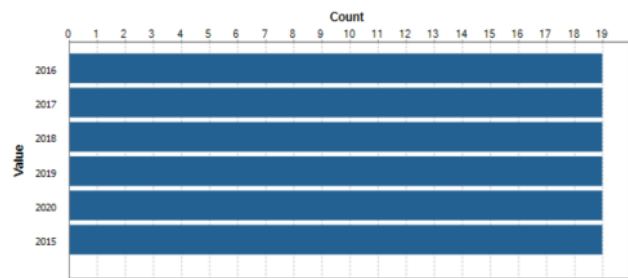


Figura H.2: Feriados - Frequências do atributo 'year'

- **date** - holidays[n].date

▼ Simple Statistics

Label	Count	%
Row Count	114	100.00%
Null Count	0	0.00%
Distinct Count	113	99.12%
Unique Count	112	98.25%
Duplicate Count	1	0.88%

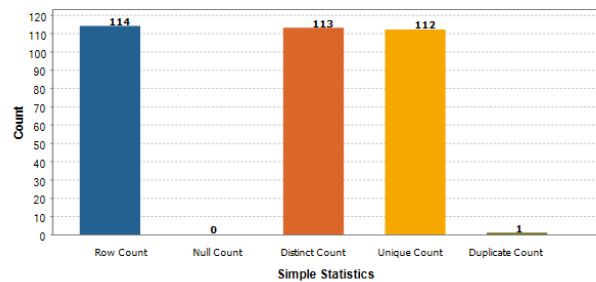
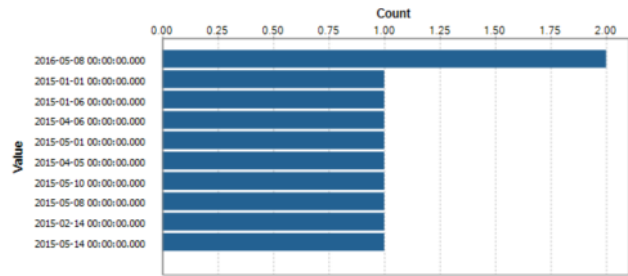


Figura H.3: Feriados - Estatísticas Simples do atributo 'date'

Value Frequency

Value	Count	%
2016-05-08 00:00:00.000	2	1.75%
2015-01-01 00:00:00.000	1	0.88%
2015-01-06 00:00:00.000	1	0.88%
2015-04-06 00:00:00.000	1	0.88%
2015-05-01 00:00:00.000	1	0.88%
2015-04-05 00:00:00.000	1	0.88%
2015-05-10 00:00:00.000	1	0.88%
2015-05-08 00:00:00.000	1	0.88%
2015-02-14 00:00:00.000	1	0.88%
2015-05-14 00:00:00.000	1	0.88%



Value Low Frequency

Value	Count	%
2015-01-01 00:00:00.000	1	0.88%
2015-01-06 00:00:00.000	1	0.88%
2015-04-06 00:00:00.000	1	0.88%
2015-05-01 00:00:00.000	1	0.88%
2015-04-05 00:00:00.000	1	0.88%
2015-05-10 00:00:00.000	1	0.88%
2015-05-08 00:00:00.000	1	0.88%
2015-02-14 00:00:00.000	1	0.88%
2015-05-14 00:00:00.000	1	0.88%
2015-05-24 00:00:00.000	1	0.88%

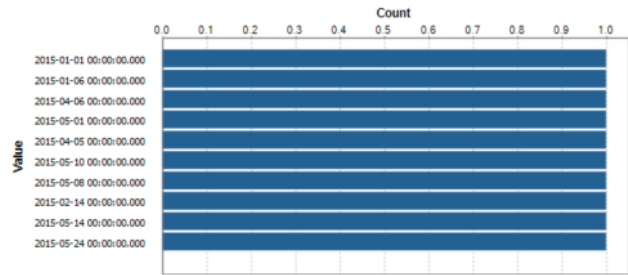


Figura H.4: Feriados - Frequências do atributo 'date'

- **start** - holidays[n].start

Simple Statistics

Label	Count	%
Row Count	114	100.00%
Null Count	0	0.00%
Distinct Count	113	99.12%
Unique Count	112	98.25%
Duplicate Count	1	0.88%

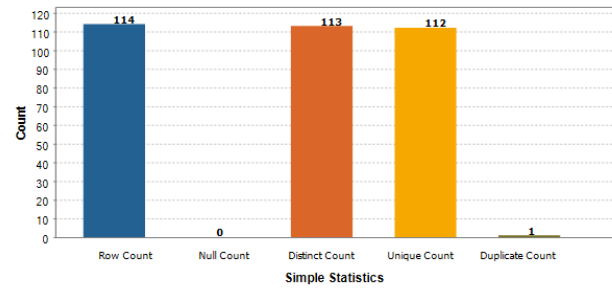
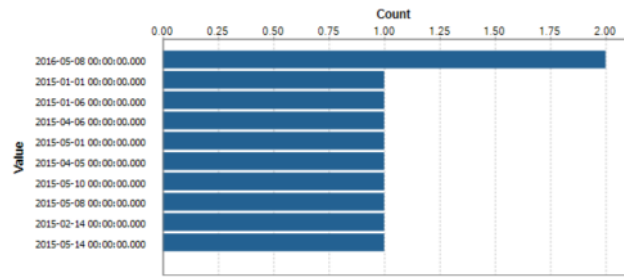


Figura H.5: Feriados - Estatísticas Simples do atributo 'start'

Value Frequency

Value	Count	%
2016-05-08 00:00:00.000	2	1.75%
2015-01-01 00:00:00.000	1	0.88%
2015-01-06 00:00:00.000	1	0.88%
2015-04-06 00:00:00.000	1	0.88%
2015-05-01 00:00:00.000	1	0.88%
2015-04-05 00:00:00.000	1	0.88%
2015-05-10 00:00:00.000	1	0.88%
2015-05-08 00:00:00.000	1	0.88%
2015-02-14 00:00:00.000	1	0.88%
2015-05-14 00:00:00.000	1	0.88%



Value Low Frequency

Value	Count	%
2015-01-01 00:00:00.000	1	0.88%
2015-01-06 00:00:00.000	1	0.88%
2015-04-06 00:00:00.000	1	0.88%
2015-05-01 00:00:00.000	1	0.88%
2015-04-05 00:00:00.000	1	0.88%
2015-05-10 00:00:00.000	1	0.88%
2015-05-08 00:00:00.000	1	0.88%
2015-02-14 00:00:00.000	1	0.88%
2015-05-14 00:00:00.000	1	0.88%
2015-05-24 00:00:00.000	1	0.88%

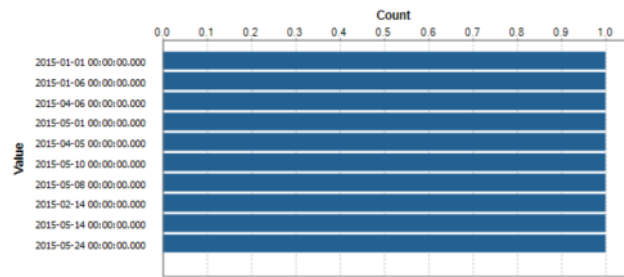


Figura H.6: Feriados - Frequências do atributo 'start'

- **end** - holidays[n].end

Simple Statistics

Label	Count	%
Row Count	114	100.00%
Null Count	0	0.00%
Distinct Count	113	99.12%
Unique Count	112	98.25%
Duplicate Count	1	0.88%

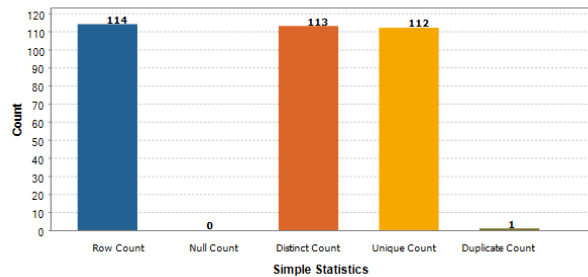


Figura H.7: Feriados - Estatísticas Simples do atributo 'end'

Simple Statistics

Label	Count	%
Row Count	114	100.00%
Null Count	0	0.00%
Distinct Count	113	99.12%
Unique Count	112	98.25%
Duplicate Count	1	0.88%

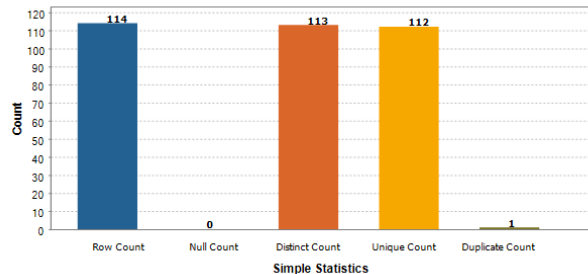


Figura H.8: Feriados - Frequências do atributo 'end'

- **name** - holidays[n].name



Simple Statistics

Label	Count	%
Row Count	114	100.00%
Null Count	0	0.00%
Distinct Count	19	16.67%
Unique Count	0	0.00%
Duplicate Count	19	16.67%
Blank Count	0	0.00%

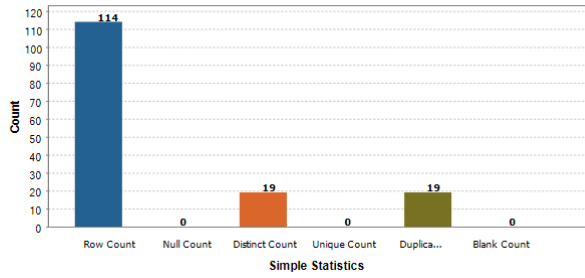
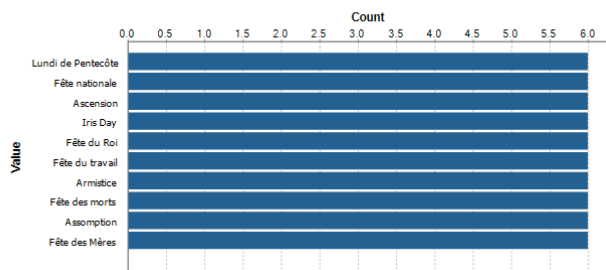


Figura H.9: Feriados - Estatísticas Simples do atributo 'name'

Value Frequency

Value	Count	%
Lundi de Pentecôte	6	5.26%
Fête nationale	6	5.26%
Ascension	6	5.26%
Iris Day	6	5.26%
Fête du Roi	6	5.26%
Fête du travail	6	5.26%
Armistice	6	5.26%
Fête des morts	6	5.26%
Assomption	6	5.26%
Fête des Mères	6	5.26%



Value Low Frequency

Value	Count	%
Lundi de Pentecôte	6	5.26%
Fête nationale	6	5.26%
Ascension	6	5.26%
Iris Day	6	5.26%
Fête du Roi	6	5.26%
Fête du travail	6	5.26%
Armistice	6	5.26%
Fête des morts	6	5.26%
Assomption	6	5.26%
Fête des Mères	6	5.26%

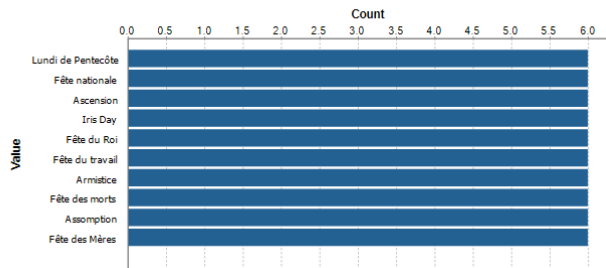


Figura H.10: Feriados - Frequências do atributo 'name'

- **type** - holidays[n].type

Simple Statistics

Label	Count	%
Row Count	114	100.00%
Null Count	0	0.00%
Distinct Count	3	2.63%
Unique Count	0	0.00%
Duplicate Count	3	2.63%
Blank Count	0	0.00%

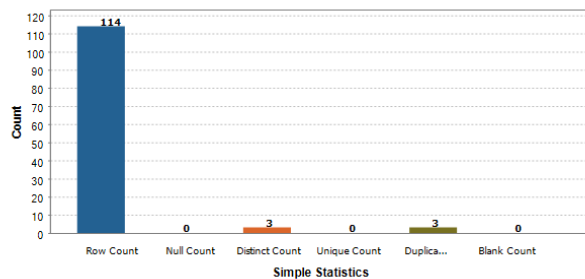
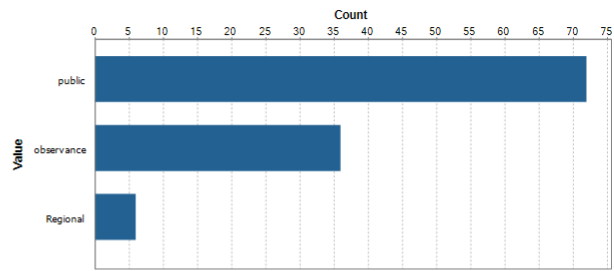


Figura H.11: Feriados - Estatísticas Simples do atributo 'type'

Value Frequency

Value	Count	%
public	72	63.16%
observance	36	31.58%
Regional	6	5.26%



Value Low Frequency

Value	Count	%
Regional	6	5.26%
observance	36	31.58%
public	72	63.16%

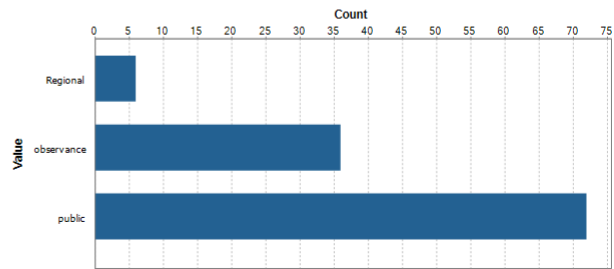


Figura H.12: Feriados - Frequências do atributo 'type'

- **public** - holidays[n].public

Simple Statistics

Label	Count	%
Row Count	114	100.00%
Null Count	0	0.00%
Distinct Count	2	1.75%
Unique Count	0	0.00%
Duplicate Count	2	1.75%
Blank Count	0	0.00%

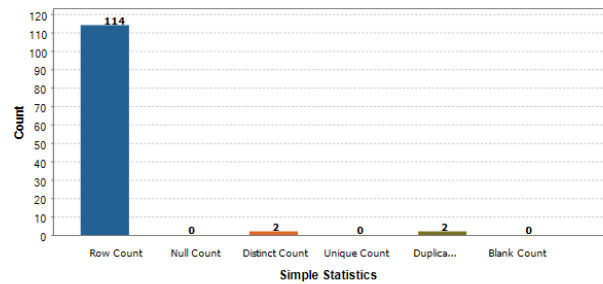
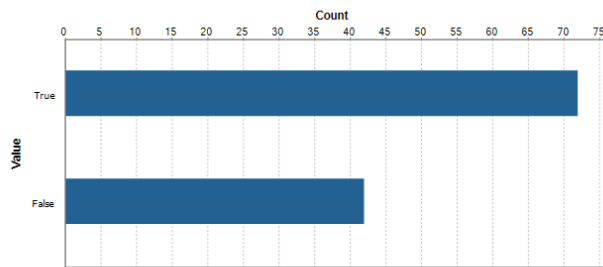


Figura H.13: Feriados - Estatísticas Simples do atributo 'public'

▼ Value Frequency

Value	Count	%
True	72	63.16%
False	42	36.84%



▼ Value Low Frequency

Value	Count	%
False	42	36.84%
True	72	63.16%

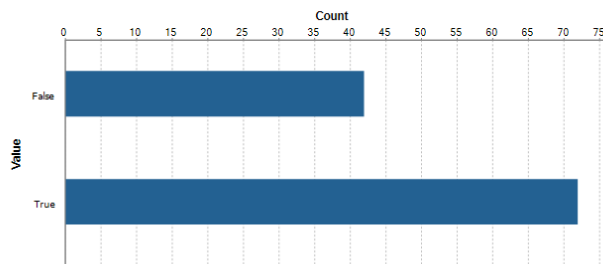


Figura H.14: Feriados - Frequências do atributo 'public'

- **country** - holidays[n].country

▼ Simple Statistics

Label	Count	%
Row Count	114	100.00%
Null Count	0	0.00%
Distinct Count	1	0.88%
Unique Count	0	0.00%
Duplicate Count	1	0.88%
Blank Count	0	0.00%

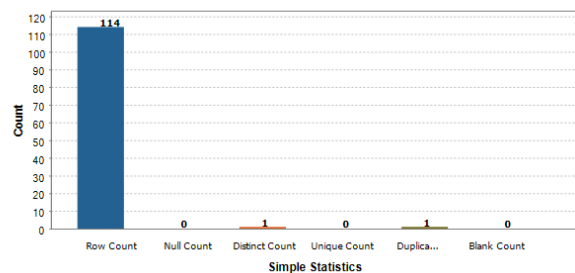


Figura H.15: Feriados - Estatísticas Simples do atributo 'country'

## I Exploração dos dados das Ocupações

- **park\_name**

▼ Simple Statistics

Label	Count	%
Row Count	5521824	100.00%
Null Count	0	0.00%
Distinct Count	63	1.141E-3%
Unique Count	0	0.00%
Duplicate Count	63	1.141E-3%
Blank Count	0	0.00%

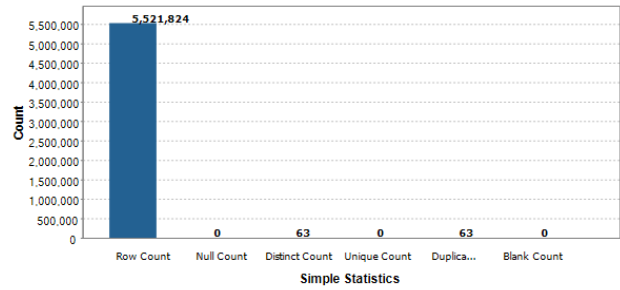
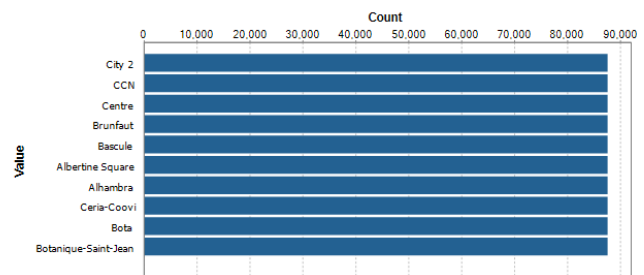


Figura I.1: Ocupação dos parques - Ocupação dos parques - Estatísticas Simples do atributo 'park\_name'

▼ Value Frequency

Value	Count	%
City 2	87648	1.59%
CCN	87648	1.59%
Centre	87648	1.59%
Brunfaut	87648	1.59%
Bascule	87648	1.59%
Albertine Square	87648	1.59%
Alhambra	87648	1.59%
Ceria-Coovi	87648	1.59%
Bota	87648	1.59%
Botanique-Saint-Jean	87648	1.59%



▼ Value Low Frequency

Value	Count	%
City 2	87648	1.59%
CCN	87648	1.59%
Centre	87648	1.59%
Brunfaut	87648	1.59%
Bascule	87648	1.59%
Albertine Square	87648	1.59%
Alhambra	87648	1.59%
Ceria-Coovi	87648	1.59%
Bota	87648	1.59%
Botanique-Saint-Jean	87648	1.59%

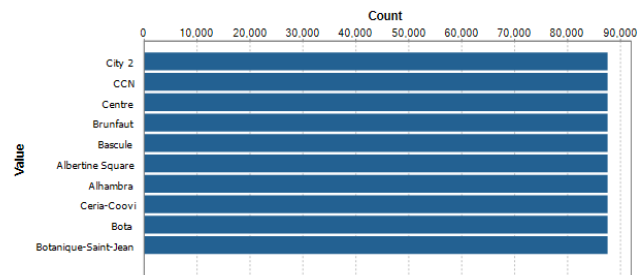
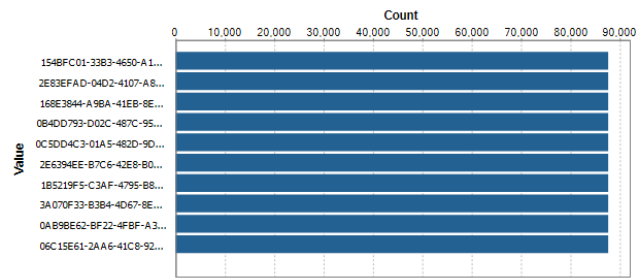


Figura I.2: Ocupação dos parques - Frequências do atributo 'park\_name'

- park\_id

Value Frequency

Value	Count	%
154BFC01-33B3-4650-A1B...	87648	1.59%
2E83EFAD-04D2-4107-A83...	87648	1.59%
168E3844-A9BA-41EB-8EC...	87648	1.59%
0B4DD793-D02C-487C-95...	87648	1.59%
0C5DD4C3-01A5-482D-9D...	87648	1.59%
2E6394EE-B7C6-42E8-B093...	87648	1.59%
1B5219F5-C3AF-4795-B829...	87648	1.59%
3A070F33-B3B4-4D67-8E6...	87648	1.59%
0AB9BE62-BF22-4FBF-A30...	87648	1.59%
06C15E61-2AA6-41C8-921...	87648	1.59%



Value Low Frequency

Value	Count	%
154BFC01-33B3-4650-A1B...	87648	1.59%
2E83EFAD-04D2-4107-A83...	87648	1.59%
168E3844-A9BA-41EB-8EC...	87648	1.59%
0B4DD793-D02C-487C-95...	87648	1.59%
0C5DD4C3-01A5-482D-9D...	87648	1.59%
2E6394EE-B7C6-42E8-B093...	87648	1.59%
1B5219F5-C3AF-4795-B829...	87648	1.59%
3A070F33-B3B4-4D67-8E6...	87648	1.59%
0AB9BE62-BF22-4FBF-A30...	87648	1.59%
06C15E61-2AA6-41C8-921...	87648	1.59%

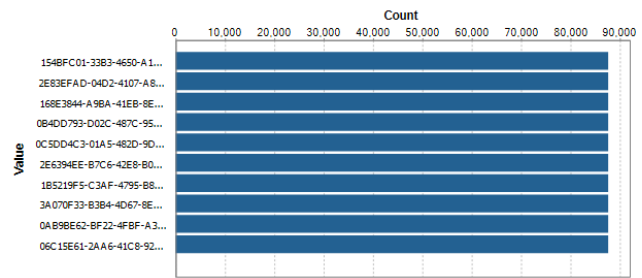


Figura I.3: Ocupação dos parques - Estatísticas Simples do atributo 'park\_id'

Simple Statistics

Label	Count	%
Row Count	5521824	100.00%
Null Count	0	0.00%
Distinct Count	63	1.141E-3%
Unique Count	0	0.00%
Duplicate Count	63	1.141E-3%
Blank Count	0	0.00%

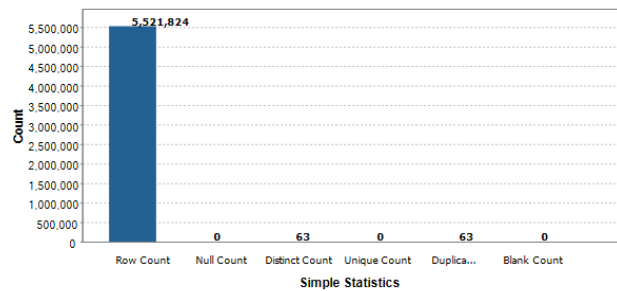


Figura I.4: Ocupação dos parques - Frequências do atributo 'park\_id'

date

Simple Statistics

Label	Count	%
Row Count	5521824	100.00%
Null Count	0	0.00%
Distinct Count	87638	1.59%
Unique Count	0	0.00%
Duplicate Count	87638	1.59%

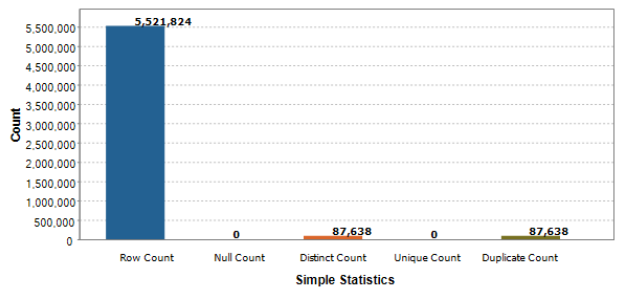


Figura I.5: Ocupação dos parques - Estatísticas Simples do atributo 'date'



Simple Statistics

Label	Count	%
Row Count	5521824	100.00%
Null Count	0	0.00%
Distinct Count	87648	1.59%
Unique Count	0	0.00%
Duplicate Count	87648	1.59%

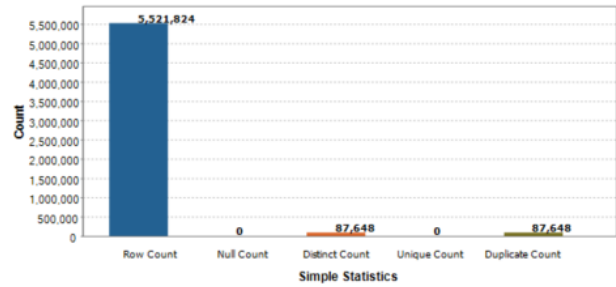
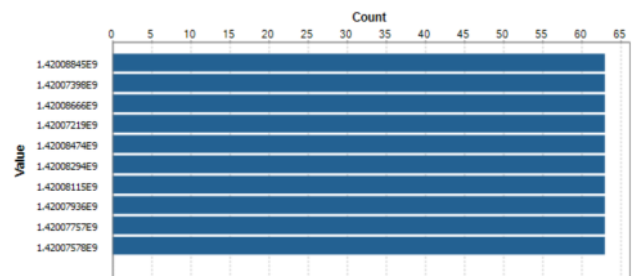


Figura I.8: Ocupação dos parques - Estatísticas Simples do atributo 'timestamp'

Value Frequency

Value	Count	%
1.42008845E9	63	1.141E-3%
1.42007398E9	63	1.141E-3%
1.42008666E9	63	1.141E-3%
1.42007219E9	63	1.141E-3%
1.42008474E9	63	1.141E-3%
1.42008294E9	63	1.141E-3%
1.42008115E9	63	1.141E-3%
1.42007936E9	63	1.141E-3%
1.42007757E9	63	1.141E-3%
1.42007578E9	63	1.141E-3%



Value Low Frequency

Value	Count	%
1.42008845E9	63	1.141E-3%
1.42007398E9	63	1.141E-3%
1.42008666E9	63	1.141E-3%
1.42007219E9	63	1.141E-3%
1.42008474E9	63	1.141E-3%
1.42008294E9	63	1.141E-3%
1.42008115E9	63	1.141E-3%
1.42007936E9	63	1.141E-3%
1.42007757E9	63	1.141E-3%
1.42007578E9	63	1.141E-3%

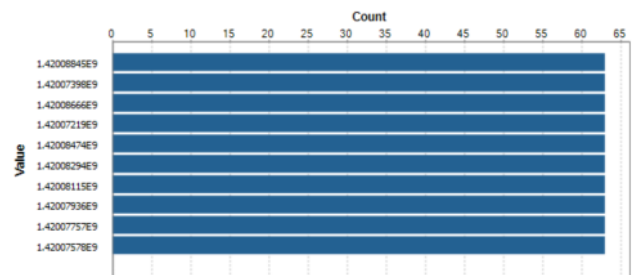


Figura I.9: Ocupação dos parques - Frequências do atributo 'timestamp'

Summary Statistics

Label	Value
Mean	1.4989545E9
Median	1.498954496E9
Inter Quartile Range	7.888314E7
Lower Quartile	1.459512958E9
Upper Quartile	1.538396098E9
Range	1.5776461E8
Minimum	1.42007219E9
Maximum	1.5778368E9

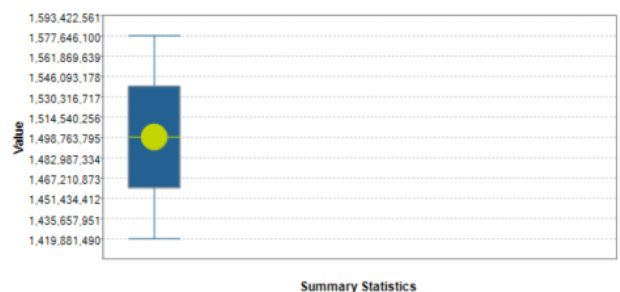


Figura I.10: Ocupação dos parques - Estatísticas do atributo 'timestamp'

- dayOfWeek

Simple Statistics

Label	Count	%
Row Count	5521824	100.00%
Null Count	0	0.00%
Distinct Count	7	1.268E-4%
Unique Count	0	0.00%
Duplicate Count	7	1.268E-4%

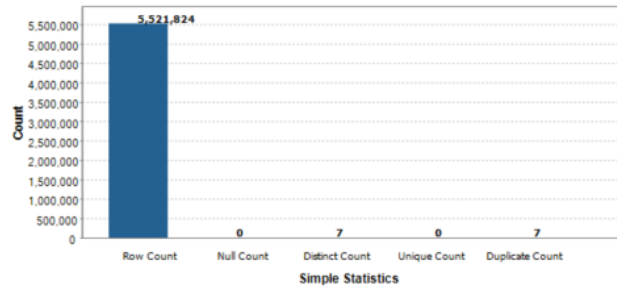
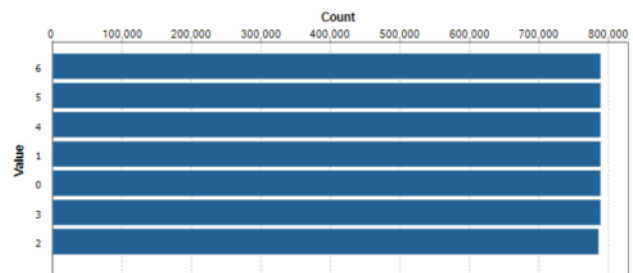


Figura I.11: Ocupação dos parques - Estatísticas Simples do atributo 'dayOfWeek'

Value Frequency

Value	Count	%
6	789264	14.29%
5	789264	14.29%
4	789264	14.29%
1	789264	14.29%
0	789264	14.29%
3	789201	14.29%
2	786303	14.24%



Value Low Frequency

Value	Count	%
2	786303	14.24%
3	789201	14.29%
6	789264	14.29%
5	789264	14.29%
4	789264	14.29%
1	789264	14.29%
0	789264	14.29%

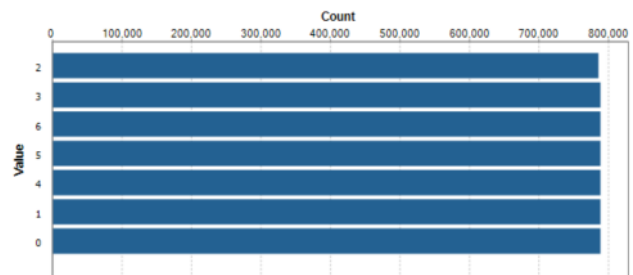


Figura I.12: Ocupação dos parques - Frequências do atributo 'dayOfWeek'

Summary Statistics

Label	Value
Mean	3.000536235852501
Median	3.0
Inter Quartile Range	4.0
Lower Quartile	1.0
Upper Quartile	5.0
Range	6.0
Minimum	0
Maximum	6

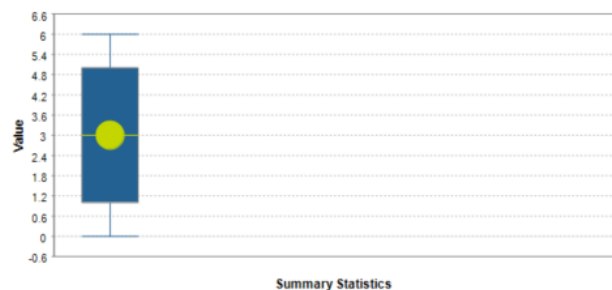


Figura I.13: Ocupação dos parques - Estatísticas do atributo 'dayOfWeek'

- **occupation**



Simple Statistics

Label	Count	%
Row Count	5521824	100.00%
Null Count	0	0.00%
Distinct Count	1034	0.02%
Unique Count	33	5.976E-4%
Duplicate Count	1001	0.02%

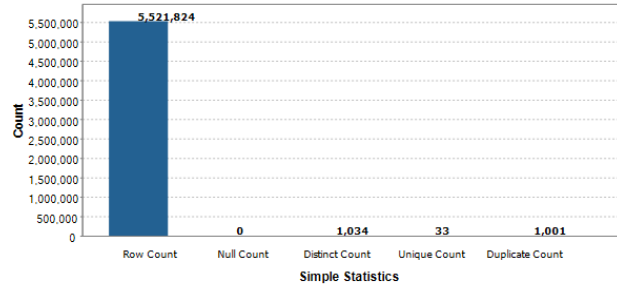
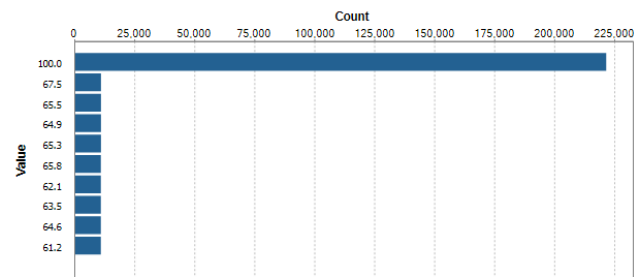


Figura I.14: Ocupação dos parques - Estatísticas Simples do atributo 'occupation'

Value Frequency

Value	Count	%
100.0	221693	4.01%
67.5	11113	0.20%
65.5	11108	0.20%
64.9	11105	0.20%
65.3	11103	0.20%
65.8	11097	0.20%
62.1	11084	0.20%
63.5	11069	0.20%
64.6	11069	0.20%
61.2	11051	0.20%



Value Low Frequency

Value	Count	%
0.3	50	9.055E-4%
0.4	51	9.236E-4%
1.1	51	9.236E-4%
0.5	52	9.417E-4%
0.1	52	9.417E-4%
0.2	53	9.598E-4%
0.6	54	9.779E-4%
1.4	64	1.159E-3%
0.8	64	1.159E-3%
1.6	64	1.159E-3%

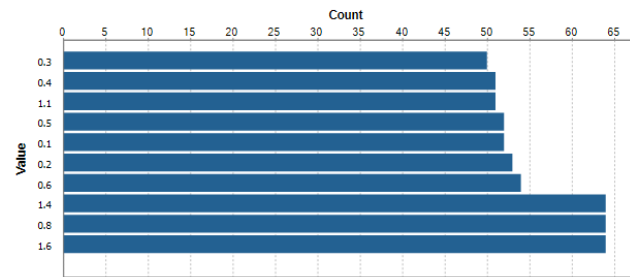


Figura I.15: Ocupação dos parques - Frequências do atributo 'occupation'

Summary Statistics

Label	Value
Mean	64.66735513482502
Median	65.0
Inter Quartile Range	27.0
Lower Quartile	51.5
Upper Quartile	78.5
Range	100.0
Minimum	0.0
Maximum	100.0

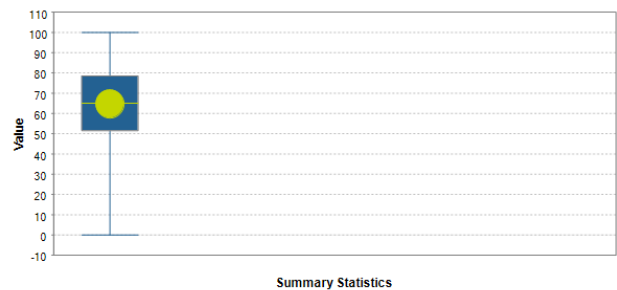


Figura I.16: Ocupação dos parques - Estatísticas do atributo 'occupation'