



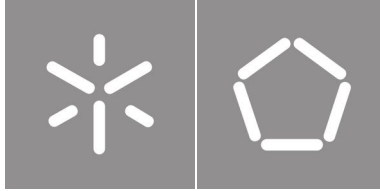
**Universidade do Minho**

Escola de Engenharia

Maria Inês Vieira Pinto

**Plataforma online para visualização  
de tendências e opiniões de eventos  
desportivos com base em redes sociais**





**Universidade do Minho**

Escola de Engenharia

Maria Inês Vieira Pinto

**Plataforma online para visualização  
de tendências e opiniões de eventos  
desportivos com base em redes sociais**

Dissertação de Mestrado  
Mestrado em Engenharia Informática

Trabalho efetuado sob a orientação de:  
**Professor Doutor José Manuel Ferreira Machado**  
**Doutor Hugo Daniel Abreu Peixoto**

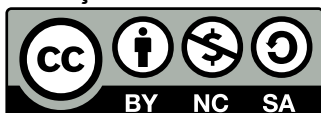
## **DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS**

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositoriUM da Universidade do Minho.

### ***Licença concedida aos utilizadores deste trabalho***



**Creative Commons Atribuição-NãoComercial-Compartilhalgal 4.0 Internacional**  
**CC BY-NC-SA 4.0**

<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.pt>

# Agradecimentos

A realização deste projeto de dissertação de mestrado não teria sido possível sem o apoio e estímulo de diversas pessoas. Gostaria, de expressar toda minha gratidão a todos aqueles que, de forma direta ou indireta, contribuíram para que este objetivo fosse concluído.

Ao orientador desta dissertação, o Professor Doutor José Machado mas especialmente ao coorientador Doutor Hugo Peixoto pela orientação prestada, pelo seu incentivo, disponibilidade e apoio que sempre demonstrou.

Agradeço à MOG Technologies pela oportunidade concedida para a realização deste projeto e por me acolherem tão bem. Em particular, ao Pedro Santos e Alexandre Ulisses que amavelmente me receberam, apoiaram, ajudaram e me forneceram toda a informação disponível necessária.

De modo particular, quero agradecer a toda a minha família e namorado, que foram preponderantes ao longo deste percurso, pela compreensão, apoio, motivação demonstrada e por serem o meu suporte diário.

Aos meus colegas de trabalho e amigos, Alexandra Branco e Luís Cerqueira por todo o incentivo nos dias bons mas também pelo apoio nos dias maus.

Aos meus amigos e colegas do mestrado agradeço a ajuda, o convívio, as boas discussões, a partilha de ideias e sobretudo, a camaradagem, vividos durante, não só em todo o meu percurso académico, mas até aos dias de hoje.

A todos, os meus sinceros e profundos agradecimentos!

### **DECLARAÇÃO DE INTEGRIDADE**

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

Vila Verde, 26 de Setembro de 2021  
(Local) (Data)

Maria Inês Vieira Pinto  
(Maria Inês Vieira Pinto)

# Resumo

## **Plataforma online para visualização de tendências e opiniões de eventos desportivos com base em redes sociais**

A extracção e o processamento de dados nas redes sociais são complementares e, por vezes, fundamentais no apoio à tomada de decisões em diversos negócios. Os eventos desportivos não são excepção, podendo dessa forma beneficiar dessas técnicas. Nesse contexto, os clubes desportivos podem aproveitar as informações obtidas nas redes sociais durante determinados eventos para dar suporte nas decisões, podendo dessa forma melhorar a experiência do espectador e, assim, atrair mais audiência.

Foi necessário realizar uma análise ao tipo de dados a serem recolhidos das redes sociais, assim como a melhor forma para os exibir ao utilizador, dando ao mesmo uma boa experiência. Foi elaborado um protótipo como solução para o desafio, apresentando assim uma plataforma *web*, alinhada às melhores práticas de *User Experience* e *User Interface*, que auxilia clubes desportivos e produtores de conteúdo televisivo a visualizar facilmente dados sociais, durante ou após um evento desportivo em específico.

Foram apresentados os resultados obtidos com esta proposta, demonstrando que é possível obter informações valiosas e relevantes para os clubes desportivos, conseguindo captar e reunir certos conhecimentos em tempo real.

**Palavras-chave:** Dashboard, Estatísticas desportivas, Análise semântica, Análise de dados, Eventos Desportivos.

# Abstract

## **Online platform for view trends and opinions of sports events, collected from social networks**

Data extraction and processing of social media are complementary and sometimes fundamental to support decision making in several businesses. Sports events are no exception and can benefit from these techniques. In this context, sports clubs can take advantage of the information obtained from social media during events to support decisions that can improve the spectator experience and thus attract more audiences.

It was necessary to carry out an analysis of the type of data to be collected from social networks, as well as the best way to display them to the user, giving them a good experience. A prototype was developed as a solution to the challenge, thus presenting a web platform, aligned with the best practices of User Experience and User Interface, which helps sports clubs and TV content producers to easily visualize social data, during or after a specific sporting event.

The results obtained with this proposal were presented, demonstrating that it is possible to obtain valuable and relevant information for sports clubs, managing to capture and gather certain knowledge in real time.

**Keywords:** Dashboard, Sports statistics, Semantic analysis, Data analysis, Sports events.



# Índice

<b>Índice de Figuras</b>	<b>xi</b>
<b>Índice de Tabelas</b>	<b>xiv</b>
<b>Siglas</b>	<b>xvi</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Contexto e Problemas . . . . .	2
1.2 Motivação . . . . .	3
1.3 Questão de Investigação e Objetivos . . . . .	4
1.4 Estrutura do documento . . . . .	5
<b>2 Estado da arte</b>	<b>7</b>
2.1 Visualização de dados nas Redes Sociais . . . . .	7
2.1.1 Twitter . . . . .	10
2.1.2 Instagram . . . . .	16
2.2 Visualização de dados no Desporto . . . . .	23
2.3 Trabalho Relacionado . . . . .	33
2.4 Comparação do trabalho relacionado . . . . .	38
2.5 Representação de dados . . . . .	39
<b>3 Ferramentas e Metodologias</b>	<b>44</b>
3.1 Ferramentas . . . . .	44
3.1.1 Tecnologias de Front-end . . . . .	44
3.1.2 Ferramentas de Visualização de Dados . . . . .	46
3.1.3 Tecnologias de back-end . . . . .	49
3.1.4 Base de dados . . . . .	51
3.2 Metodologias . . . . .	53

3.2.1	Data Driven Design . . . . .	53
3.2.2	Análise de dados . . . . .	65
<b>4</b>	<b>PLAYOFF - Prova de Conceito, Requisitos e Arquitetura do sistema</b>	<b>67</b>
4.1	Prova de Conceito . . . . .	67
4.1.1	Testes de fluxo - Texto simples . . . . .	67
4.1.2	Testes de fluxo - Texto simples com utilizador específico . . . . .	70
4.2	Requisitos . . . . .	74
4.2.1	Requisitos Funcionais . . . . .	74
4.2.2	Requisitos Não Funcionais . . . . .	75
4.3	Arquitetura do Sistema . . . . .	77
4.3.1	Data Source - API Twitter . . . . .	77
4.3.2	Core . . . . .	79
4.3.3	Data Visualization . . . . .	84
4.3.4	Data Storage . . . . .	88
<b>5</b>	<b>Design e Desenvolvimento</b>	<b>90</b>
5.1	Mockups . . . . .	90
5.2	Desenvolvimento . . . . .	92
5.2.1	Página Principal . . . . .	92
5.2.2	Página Dashboard . . . . .	96
<b>6</b>	<b>Resultados</b>	<b>100</b>
6.1	Análise de Resultados . . . . .	100
6.2	Análise de SWOT . . . . .	107
<b>7</b>	<b>Conclusão e Trabalho Futuro</b>	<b>109</b>
	<b>Bibliografia</b>	<b>111</b>
	<b>Apêndices</b>	
<b>A</b>	<b>Anexo A</b>	<b>118</b>
<b>B</b>	<b>Anexo B</b>	<b>119</b>
<b>C</b>	<b>Anexo C</b>	<b>120</b>

## Índice de Figuras

1	Pirâmide do Conhecimento . . . . .	1
2	Conta do <i>Tottenham Hotspur</i> no <i>Twitter</i> e no <i>Instagram</i> . . . . .	9
3	Evolução das redes sociais até ao ano 2021 . . . . .	10
4	<i>Tweets</i> de Portugal sobre <i>Covid-19</i> [42] . . . . .	12
5	Tipo de sentimentos pelas regiões do país [42] . . . . .	12
6	Mapa Geográfico que representa o tipo de sentimento [47] . . . . .	14
7	<i>Tweet</i> no canal do <i>Manchester United</i> . . . . .	15
8	Relação volume de tendências/tempo no <i>Twitter</i> [49] . . . . .	15
9	<i>Twitter</i> Vs TV - Tipo de evento mais memorável [83] . . . . .	16
10	Evolução do <i>Instagram</i> ao longo do tempo [59] . . . . .	17
11	Publicação do jogador Cristiano Ronaldo . . . . .	18
12	Gráfico que analisa os filtros nas histórias do <i>Instagram</i> [14] . . . . .	19
13	Mapa geográfico interativo dos Estados Unidos [14] . . . . .	19
14	Nuvem de Palavras das <i>hashtags</i> em <i>Nova York</i> [14] . . . . .	20
15	Nuvem de Palavras das <i>hashtags</i> em <i>Los Angeles</i> [14] . . . . .	20
16	Relação tipo de imagem vs número de gostos [60] . . . . .	21
17	Relação tipo de imagem vs número de comentários [60] . . . . .	21
18	Imagem narrativa vs Imagem comunicativa [60] . . . . .	22
19	Porcentagem do tipo de evento no <i>Instagram</i> da CBG [46] . . . . .	23
20	Protagonistas das publicações no <i>Instagram</i> da CBG [46] . . . . .	23
21	Artigos sobre dados desportivos ao longo dos anos [62] . . . . .	24
22	<i>Website</i> da plataforma <i>SofaScore</i> [68] . . . . .	25
23	<i>Software</i> da <i>Prozone Sports</i> [11] . . . . .	26
24	Plataforma <i>NBAVis</i> [51] . . . . .	27
25	Resultados dos jogos da primeira liga inglesa [62] . . . . .	29
26	Relação na distância do lance com a percentagem de golos [77] . . . . .	30

27	Visualização interativa no <i>Website</i> da empresa <i>Smartbase</i> [66]	31
28	Exemplo de um gráfico comunicativo [43]	32
29	Exemplo de um gráfico interpretativo[43]	33
30	Exemplo da plataforma <i>Modern Research Lite</i> [69]	34
31	Gráfico das tendências de saúde pública [24]	34
32	Análise de sentimento dos <i>tweets</i> agrupados [12]	35
33	Exemplo do investigador <i>John Naujoks</i> [40]	36
34	Sumarização de eventos para jogos desportivos [35]	37
35	Estudo do Campeonato do Mundo FIFA 2014 [90]	37
36	Tipos de visualização	41
37	<i>Github</i> : Número de <i>downloads</i>	44
38	<i>StackOverflow</i> : Percentagem de questões feitas [28]	45
39	Ranking das linguagens de programação [79]	51
40	Três fases da visão do utilizador	53
41	Conhecimento do utilizador [43]	54
42	Etapas do processo de design de visualização de dados	55
43	Dois tipos de gráficos que representam a mesma informação [43]	57
44	Previsão do mercado de <i>Big Data</i> [1]	58
45	Gráfico do nº de passes de cada jogador por cada ano [43]	60
46	Gráfico do nº de passes de cada jogador pela idade [43]	60
47	Comprimento versus Área [43]	61
48	Modelos de cores [54]	63
49	Exemplo da composição de dados num gráfico [20]	64
50	Ciclo de desenvolvimento do design	65
51	Passos para a análise de dados	66
52	Criação do conjunto de dados de <i>streaming</i> no <i>Power Bi</i>	68
53	Criação do fluxo no <i>Power Automate</i>	69
54	Operação de deteção do sentimento no <i>Power Automate</i>	69
55	Gráfico que representa o sentimento dos <i>tweets</i> durante o jogo	70
56	Esquema representativo do <i>Power Automate</i> com o <i>Power Bi</i>	71
57	Esquema das sub condições criadas no <i>Power Automate</i>	72
58	Fluxo criado no <i>Power Automate</i> para o canal do <i>Chelsea</i>	72
59	Sub condição do sentimento muito negativo	73
60	Tipo de sentimento predominante no canal do <i>Chelsea</i>	73
61	Tipo de sentimento predominante no canal do <i>Tottenham</i>	74
62	Componentes da arquitetura do sistema	77

---

63	Arquitetura de alto nível . . . . .	78
64	Sistema <i>HTTP Streaming</i> [34] . . . . .	78
65	Paleta de cores . . . . .	90
66	<i>Mockup</i> da Página Principal . . . . .	91
67	<i>Mockup</i> da Página da <i>Dashboard</i> . . . . .	92
68	Gráficos utilizados da biblioteca <i>ApexCharts</i> . . . . .	94
69	1º Parte-Página Principal . . . . .	95
70	2º Parte-Página Principal . . . . .	96
71	Página da <i>dashboard</i> . . . . .	97
72	<i>Grids</i> na Página da <i>dashboard</i> . . . . .	97
73	Campos para o utilizador preencher . . . . .	98
74	Gráfico na página da <i>dashboard</i> . . . . .	99
75	1º Parte- Antes do jogo . . . . .	101
76	2º Parte- Antes do jogo . . . . .	101
77	<i>Tweets</i> do jogador <i>N'Golo Kanté</i> . . . . .	102
78	Resultados ao minuto 15' do jogo . . . . .	102
79	<i>Tweets</i> do jogador <i>Werner</i> . . . . .	103
80	<i>Tweets</i> do jogador <i>Kai Havertz</i> . . . . .	104
81	Resultados no início da segunda parte do jogo . . . . .	104
82	Resultados ao minuto 56' do jogo . . . . .	105
83	Resultados ao minuto 56' do jogo . . . . .	105
84	<i>Tweets</i> sobre o jogo . . . . .	106
85	Resultado no final do jogo . . . . .	106
86	<i>Tweets</i> sobre o jogador <i>Kanté</i> . . . . .	107
87	Análise de <i>SWOT</i> . . . . .	108
88	Comparação de características de diferentes gráficos [80, 82, 81, 2, 43] . . . . .	118
89	Comparação de características de diferentes gráficos [80, 82, 81, 2, 43] . . . . .	119
90	Comparação de características de diferentes gráficos [80, 82, 81, 2, 43] . . . . .	120

## Índice de Tabelas

1	Matriz de comparação dos trabalhos relacionados . . . . .	38
2	Tabela de comparação das ferramentas <i>front-end</i> . . . . .	45
3	Tabela de estudo das ferramentas de visualização [15, 6, 19, 16, 4, 33, 32] . . . . .	47
4	Vantagens e Desvantagens das bibliotecas selecionadas . . . . .	49
5	Estudo das ferramentas de back-end [53, 52, 23, 61, 58, 44] . . . . .	50
6	Estudo e comparação das bases de dados [72, 18, 50, 17] . . . . .	52
7	Tipos de interatividade na visualização de dados . . . . .	62
8	Requisitos Funcionais - Página Principal . . . . .	75
9	Requisitos Funcionais - Página <i>Dashboard</i> . . . . .	76
10	Requisitos Não Funcionais . . . . .	76

## Índice de Excertos de Código

1	Receção de dados via verbo <i>POST</i> . . . . .	79
2	Função <i>openStreamOfTweets</i> . . . . .	80
3	Rotas de resposta <i>GET</i> criadas . . . . .	81
4	Exemplo dos dados dos <i>tweets</i> recolhidos . . . . .	82
5	Definição do modelo de um <i>tweet</i> . . . . .	83
6	Exemplo de <i>query</i> usando o <i>Sequelize</i> . . . . .	84
7	Exemplo da utilização do <i>dialog</i> . . . . .	85
8	Exemplo de evento utilizado no gráfico . . . . .	86
9	Exemplo do <i>slideshow</i> criado . . . . .	86
10	Exemplo do <i>SingleDatePicker</i> criado . . . . .	87

# Siglas

<b>API</b>	Application Programming Interface <a href="#">x</a> , <a href="#">10</a> , <a href="#">18</a> , <a href="#">46</a> , <a href="#">77</a> , <a href="#">78</a> , <a href="#">79</a> , <a href="#">80</a> , <a href="#">81</a>
<b>CBG</b>	Confederação Brasileira de Ginástica <a href="#">xi</a> , <a href="#">22</a> , <a href="#">23</a>
<b>CBS</b>	Columbia Broadcasting System <a href="#">20</a>
<b>CSR</b>	Corporate Social Responsibility <a href="#">35</a>
<b>DIKW</b>	Data-Information-Knowledge-Wisdom <a href="#">1</a>
<b>ESPN</b>	Entertainment and Sports Programming Network <a href="#">20</a>
<b>HTTP</b>	Hypertext Transfer Protocol <a href="#">xiii</a> , <a href="#">78</a> , <a href="#">79</a> , <a href="#">81</a> , <a href="#">87</a>
<b>IT</b>	Information Technology <a href="#">35</a>
<b>JSON</b>	JavaScript Object Notation <a href="#">46</a> , <a href="#">51</a> , <a href="#">79</a>
<b>MIT</b>	Massachusetts Institute of Technology <a href="#">48</a> , <a href="#">49</a>
<b>NBA</b>	National Basketball Association <a href="#">26</a> , <a href="#">27</a>
<b>noSQL</b>	Not Only SQL <a href="#">51</a> , <a href="#">52</a>
<b>NPM</b>	Node Package Manager <a href="#">46</a> , <a href="#">50</a>
<b>ORM</b>	Object Relational Mapping <a href="#">83</a>
<b>SQL</b>	Structured Query Language <a href="#">51</a> , <a href="#">52</a>
<b>SWOT</b>	Strengths, Weaknesses, Opportunities e Threats <a href="#">xiii</a> , <a href="#">100</a> , <a href="#">107</a> , <a href="#">108</a>



- UEFA** Union of European Football Associations 100
- UI** User Interface 5, 46, 90, 110
- UX** User Experience 5, 46, 90, 110
- 
- WWE** World Wrestling Entertainment 8



## Introdução

Os **dados** ajudam-nos a entender e a interagir com o mundo. Atualmente, muitas empresas estão a colecionar dados com grande velocidade, seja para fins de conformidade, elaboração de relatórios ou visualização dos mesmos e a ideia é que essas informações possam vir a ser valiosas.

A **informação** [36] é um dos bens mais valiosos para as empresas que pretendem melhorar os seus negócios e se destacarem no mercado. A informação e o conhecimento assumem-se atualmente como factores críticos de sucesso para a produtividade, competitividade e criação de valor.

A Hierarquia *DIKW* (*Data-Information-Knowledge-Wisdom*) ou Pirâmide do Conhecimento [31] é estruturada em quatro partes, sendo os dados colocados como base da informação, conhecimento e sabedoria (visível na Figura 1).

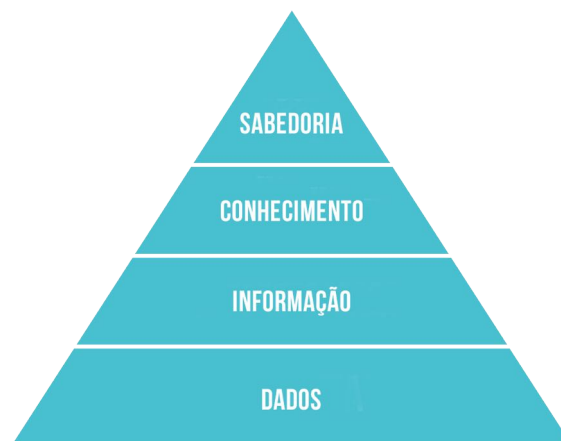


Figura 1: Pirâmide do Conhecimento

Neste capítulo, é apresentada uma introdução ao trabalho que irá ser desenvolvido no contexto desta dissertação, começando por explicar em que contexto e problemas é que o projeto se enquadra. O envolvimento e opinião do público é bastante importante nos dias de hoje para vários setores, sendo o desporto um deles. Contudo, ainda surgem obstáculos em aproveitar e fazer bom uso dos variados pontos de vista de quem assiste. Seguidamente, é apresentada a motivação deste tema de projeto e também são

descritos os objetivos a serem atingidos com a execução deste trabalho, assim como é também delineada a estrutura do documento.

## 1.1 Contexto e Problemas

A transmissão de eventos desportivos [64] em direto é, sem dúvida, uma das tipologias de conteúdos audiovisuais que atinge maior volume de espetadores finais. Motivo pelo qual, no desporto internacional moderno, a dinâmica espetador-desportista(s) é, hoje, vista como uma relação bilateral crucial para o sucesso de televisões e clubes. Estas dinâmicas reconhecem que o envolvimento do público, a liberdade de expressão do mesmo e a **interação** com a performance a que assiste são de grande importância, pois moldam a atmosfera, amplificam a excitação e o comprometimento com o espetáculo e impactam diretamente no resultado final da experiência [71]. Cada vez mais, o pensamento e opinião do público tem vindo a ser exibido nas redes sociais, sendo estas abertas a todo o tipo de pessoas de maneira simples e importante no nosso dia a dia, permitindo também a troca de opiniões entre os mais variados utilizadores.

No mundo atual, o número de utilizadores das **redes sociais** é de cerca de 3,6 mil milhões [74], um número grande o suficiente para se tirar proveito e realizar estudos sobre preferências gerais, ou até específicas de um determinado grupo de utilizadores. Algo que não deve faltar nos departamentos de *marketing* e vendas de determinadas empresas, é o contínuo estudo de hábitos de pesquisa do utilizador para desenvolvimento de estratégias para seu próprio usufruto. A nível do desporto, a afluência nas redes sociais tem vindo a **aumentar** significativamente, seja no *Facebook*, *Twitter*, *Instagram* ou outras redes sociais. **A recolha de dados das redes sociais** durante um determinado evento desportivo permite obter informação sobre o impacto do mesmo na audiência e, assim, clubes e estações de televisão podem adotar medidas que permitam enriquecer a experiência dos espetadores, atraindo audiências e consequentemente aumentando as suas receitas [41].

Por outro lado, a forma como esta **informação é apresentada** aos organizadores do evento influencia de forma muito significativa a perceção que estes têm da mesma e é também importante para entender, aproximar e fortalecer a relação do público alvo com as marcas de um certo clube e até mesmo com os próprios clubes. Desta forma, são necessárias ferramentas que permitam aos clubes e estações de televisão tomar decisões estratégicas mais conscientes e fundamentadas [45], com base nas **opiniões** dos utilizadores e é neste contexto que surge este trabalho, criando assim novas dinâmicas inclusivas entre os adeptos e os clubes desportivos.

Os diferentes pontos onde se pretende intervir passam por fornecer a diferentes utilizadores (como clubes desportivos, departamento de *marketing* de um clube e estações de televisão) uma boa visualização de dados que permita aos mesmos uma absorção rápida da informação, permitindo assim perceber as opiniões e tendências dos adeptos. Uma rápida filtragem do que os adeptos estão a sentir relativamente a um evento desportivo em tempo real permite obter soluções e ações rápidas sobre um determinado

contexto.

Embora os avanços tecnológicos tenham enriquecido a experiência do evento tanto para os espectadores no local como para os que assistem remotamente, o desafio de realmente gerir e atender às preferências e necessidades individuais dos diversos públicos ainda não foi totalmente atingido, nem pelas estações de televisão, nem pelos clubes desportivos. Apesar de se apresentarem como entidades cada vez mais relevantes na produção de conteúdos media, ainda recorrem a soluções tecnológicas com uma linha de orientação diferente das que são apresentadas às estações de televisão, fazendo com que não beneficiem dos anos acumulados de experiência destas últimas.

Persistem, deste modo, desafios, tanto para as estações de televisão, que necessitam de fornecer aos utilizadores um modelo único de serviços sincronizados, em camadas e personalizados, por meio de comunicação multiplataforma e convergência na distribuição, como para os clubes desportivos. Estes devem tirar partido da customização da experiência, através da combinação única de várias fontes de informação no mesmo local, no sentido de fazer com que os adeptos adiram ainda mais à identidade do clube, proporcionando-lhe novas oportunidades de negócio.

Esta dissertação foi realizada em contexto empresarial, sendo a mesma relevante para um projeto desta empresa chamado *PlayOff*. Este projeto tem como principal objetivo a criação de um ambiente de transmissão de media reativa multimodal, dirigido a estações de televisão e clubes, através do desenvolvimento de:

- Um ecossistema *Cloud* para personalização da experiência do utilizador, dirigido às estações de televisão e aos clubes e baseado na fusão, num ambiente de convergência, de diversas fontes de conteúdos desportivos multimédia em direto;
- Um ambiente integrado capaz de fornecer, às estações de televisão e aos clubes, serviços de *data analytics* em tempo real, baseados na análise da audiência e padrões de consumo de conteúdo;
- Um conjunto de ferramentas multicamada para uma gestão eficaz e enriquecimento de conteúdos e uma sumarização multimodal do evento, centradas na geração de estatísticas que sumarizem o evento e que o comparem com outros eventos semelhantes.

No contexto deste projeto, surge esta dissertação que se foca, principalmente, no segundo e terceiro ponto apresentados acima. O propósito final passa por construir uma aplicação que receba dados em tempo real e os apresente de forma dinâmica e intuitiva aos utilizadores para os mesmos tirarem conclusões relevantes acerca de um determinado tema ou tópico.

## 1.2 Motivação

A **visualização de dados**, nos dias de hoje, tem bastante importância e permite compreender variados conceitos ou identificar novos padrões, tendo acesso a quantidades massivas de dados (**Big Data**) que

podem facilitar tomadas de decisão mais bem fundamentadas [86]. Esta prática tornou-se fundamental para quaisquer empresas, assim como por exemplo, o surgimento de novas tecnologias que permitem visualizações interativas que possibilitou um grande avanço nesta área. De nada adianta uma análise perfeita sem uma boa apresentação do resultado. Pode parecer uma afirmação exagerada, mas a visualização criada a partir da **análise** tem o potencial de enaltecer cada aspecto importante de ser observado, assim como, identificar as informações rapidamente e certas relações que possam existir entre os dados, e também identificar tendências. Explorar e analisar os vastos volumes de dados tornou-se cada vez mais difícil e complexo. Muitas vezes as mudanças acontecem a uma velocidade que supera a capacidade de aprendizagem e compreensão. Da mesma forma, há uma crescente necessidade de encontrar mecanismos para comunicar a informação às pessoas de forma eficiente e eficaz.

Fruto das alterações dos padrões de consumo de media, que mudaram para um paradigma de ecrãs múltiplos, promovendo o aumento da procura por uma maior flexibilidade e fluidez de visualização e dando mais ênfase à programação em direto e à utilização de tecnologias híbridas de radiodifusão, os espetadores encontram-se, cada vez mais, num modo *multitasking*, em busca de informações adicionais e da partilha de experiências, seja através da interação na *web* ou em plataformas de *social media*, cruzando a transmissão principal com informação de outros canais.

Torna-se, assim, claro que, para angariar e influenciar audiências da transmissão de desporto em direto, é necessário aproveitar, com sucesso, o conteúdo do espaço físico do evento, vinculá-lo, alinhá-lo com os conteúdos das plataformas de *social media* e de *open data* e, finalmente, disponibilizá-lo aos espetadores, através de ecossistemas atraentes e personalizados. Por outras palavras, é necessário fornecer aos espetadores ferramentas para que os que estão no estádio possam interagir com conteúdo complementar empolgante e para que os que estão “em casa” possam ter controlo total sobre conteúdo personalizado, ao mesmo tempo que são abertos canais diretos de participação e interação.

Para além disso, ofereceu a possibilidade de trabalhar pela primeira vez no mundo empresarial e integrar um projeto de investigação nacional. Esta dissertação destaca-se por ser parte de um projeto nacional que junta universidades e empresas com uma forte componente em inovação e investigação.

### 1.3 Questão de Investigação e Objetivos

Para esta dissertação, surge o objetivo de criar uma **plataforma online** que permita a clubes desportivos e estações de televisão visualizar os dados recolhidos de redes sociais sobre um determinado evento desportivo, tendo em consideração as boas práticas de **User Experience** e **User Interface**, nomeadamente **Data Driven Design**. Idealiza-se então, que seja desenvolvida uma plataforma online capaz de representar informação proveniente de redes sociais, em tempo real, de uma forma dinâmica, com vários níveis de informação, customizável e que permita aos clubes tomar decisões mais informadas no sentido de melhorar a **experiência dos espetadores**. Posto isto, é também importante que a informação exibida esteja representada por tipos de sentimentos, sendo importante para o utilizador captar uma

visão geral das emoções dos adeptos em um determinado acontecimento/evento.

Com isto, levanta-se uma questão de investigação que se pretende que seja respondida com esta dissertação - Recorrendo à boa prática definida pelo *Data Driven Design*, será possível construir uma plataforma online que permita a clubes e estações de televisão visualizar os dados recolhidos de redes sociais próximo do tempo real, sobre um determinado evento desportivo de forma a acompanhar a interação com os seus seguidores?

Tendo isso em mente, os principais objetivos são listados a seguir:

- Compreender os conceitos de **Data Driven Design, Big Data e Streaming Data**;
- Implementar uma arquitetura de sistema, que permite receber dados contínuos relevantes das redes sociais para posterior visualização gráfica;
- Fazer uma análise sobre boas práticas de exposição de dados, onde os mesmos são melhor assimilados pelo público alvo;
- Fazer um levantamento e um estudo comparativo de ferramentas já existentes para visualização de dados;
- Explorar e definir uma metodologia de análise de dados;
- Compreender a importância da seleção de dados e o seu impacto no desempenho e demonstração visual;
- Encontrar mecanismos de análise simples do texto que avaliem uma mensagem em três tipos de sentimento - positiva, negativa ou neutra - permitindo assim obter dados do tipo de emoções que os adeptos sentem num determinado momento;
- Desenhar uma plataforma online que permita a clubes desportivos e estações de televisão visualizar dados recolhidos de redes sociais, já processados, sobre um determinado evento desportivo tendo em consideração as boas práticas de **UX e UI design**;
- Implementar um protótipo da plataforma;
- Testar e validar o protótipo desenvolvido.

## 1.4 Estrutura do documento

Esta dissertação vai ser contemplada com sete capítulos diferentes:

- **Introdução:** Neste primeiro capítulo foi introduzido o contexto do projeto, problemas que originaram a sua existência e a motivação que levou à ideia desta dissertação. De seguida, são definidos os objetivos que se pretendem alcançar e descrita a estrutura do documento.

- **Estado de Arte:** Neste segundo capítulo, são referenciados termos e definições sobre determinados tópicos que estão a ser analisados atualmente. Pretende estabelecer um ponto de partida entre trabalhos relacionados com o contexto do projeto.
- **Ferramentas e Metodologias:** Este terceiro capítulo passa pelo estudo e seleção de ferramentas existentes para a elaboração do projeto assim como a descrição das metodologias a serem utilizadas.
- **Prova de Conceito, Requisitos e Arquitetura do sistema:** Neste quarto capítulo é introduzida e recomendada uma descrição da arquitetura de sistema proposta, explicando o funcionamento de cada componente. São listados todos os requisitos funcionais e não funcionais deste projeto assim como é mostrada a prova de conceito feita com o objetivo de tentar provar o conceito teórico desta dissertação.
- **Design e Desenvolvimento:** Aqui é feita uma descrição e explicação da composição do projeto, nomeadamente os métodos de *design* utilizados, funcionando como um manual de utilizador.
- **Resultados:** Este capítulo serve para os resultados obtidos nesta dissertação serem apresentados e explicados, percebendo se os mesmos vão de encontro aos objetivos iniciais.
- **Conclusão e Trabalho Futuro:** Neste último capítulo é necessário explicar as conclusões tiradas, o estado atual do plano de trabalho proposto e o trabalho futuro a realizar.



## Estado da arte

Basta olhar para o mundo ao nosso redor para entendermos que em qualquer lugar, os dados são essenciais e serem exibidos de uma forma simplista e explicativa é fundamental para a percepção de quem os interpreta. Nos jornais televisivos, jornais tradicionais, *websites*, plataformas de comunicação social, redes sociais, estão presentes em todo o lado.

Neste capítulo, aborda-se inicialmente a importância da visualização dos dados nas redes sociais e o seu impacto. Segue-se a visualização de dados no desporto, onde são abordados temas como tipos de plataformas onde é possível visualizar este tipo de dados assim como, os tipos de dados desportivos mais utilizados e os vários tipos de experiência que os utilizadores podem desfrutar. De seguida, são dados exemplos de trabalhos relacionados que englobam as redes sociais e o desporto, comparando-os com este projeto. Por último, são estudados diferentes tipos de visualização para diferentes tipos de dados.

### 2.1 Visualização de dados nas Redes Sociais

A visualização de dados tem vindo a ser cada vez mais explorada e melhorada para satisfazer as necessidades dos dias de hoje. Com ela surge o contexto de busca e avaliação de dados na procura de uma inovação de formas de comunicação. Esta forma de comunicação estimula e ultrapassa limites de muitas pessoas, sendo conhecida como um multi-talento ou seja como um conjunto de múltiplas competências com uma evolução cada vez maior [43].

Os jogos tornam o desporto tão interessante para os adeptos demonstrarem as suas opiniões em tempo real e envolverem-se com as equipas, sendo isso uma vantagem para as mesmas. As redes sociais são um fator crítico e de extrema importância nos tempos que correm sendo elas uma fonte de informação para grandes ou pequenas empresas. Em 2020, mais de 3,6 mil milhões de pessoas usam as redes sociais em todo o mundo, um número projetado para aumentar para quase 4,41 mil milhões em 2025 [74]. Um dos grandes objetivos na busca de dados das redes sociais passa por, verificar as reações despertadas no público e rapidamente as equipas de *marketing* entenderem o que se passa e tirarem vantagens e partido das opiniões do público, percebendo assim a tendência e quanto mais próximo do tempo real melhor.

A análise de redes sociais é uma abordagem metodológica na área de ciências sociais (que estuda os aspetos sociais de indivíduos ou grupos) para descrever, compreender e explicar uma determinada estrutura social [25]. Tornou-se cada vez mais popular na última década, trazendo como consequência a consecutiva mudança na maneira como as pessoas reagem aos acontecimentos e interagem umas com as outras [10]. O conteúdo partilhado nas redes sociais gera milhares de dados diariamente - quanto maior o número de pessoas que utilizam as redes sociais, mais conteúdo existe. Dessa forma, a probabilidade de envolvimento com determinado conteúdo aumenta assim como, a percepção do valor do mesmo.

O *marketing* e publicidade estão bem presentes em todas as plataformas de *social media*, o que se traduz num *marketing* de relacionamento. É definido como o processo de identificação e visa estabelecer, manter, aprimorar e, quando necessário, encerrar relacionamentos com clientes e outras partes interessadas, de forma a que os objetivos de todas as partes sejam atendidas [39]. Desta forma, o papel das redes sociais nos desportos concentra-se em:

- Manter os adeptos/fãs atualizados sobre as notícias do clube em questão como o mercado de jogadores, contratos e anúncios importantes da liga;
- Oferecer aos adeptos a oportunidade de se envolver com as equipas e expressar a sua opinião;
- Destacar eventos e notícias da comunidade desportiva relacionados à equipa.

Atualmente, os atletas de diferentes tipos de desportos e clubes fazem parte das contas mais seguidas nas redes sociais. Tem-se o exemplo do jogador Cristiano Ronaldo, sendo este atualmente a figura pública com mais seguidores no *Instagram* (252 milhões), um dos maiores ícones a nível mundial, trazendo consigo os seus maiores fãs para qualquer clube que o mesmo seja transferido.

As redes sociais e os desportos estão a evoluir e com elas a constante adaptação das equipas e dos atletas. Segundo o *website* da *Sprout Social*, o evento da *Wrestlemania* da *WWE* marcou 13,8 milhões de interações nas redes sociais de todas as plataformas em Abril de 2020 [8].

O *Twitter* e o *Instagram* são sem dúvida uma das redes sociais mais utilizadas mundialmente [84]. Dessa forma pode-se comunicar com os adeptos através de: por exemplo, eventos online no *Twitter* como jogos ou conferências de imprensa, estabelecer e fornecer aos fãs uma *hashtag* específica sobre uma equipa ou um jogo em particular, elaborar sessões de perguntas e respostas no *Instagram* de forma a reunir informações sobre o que os adeptos pensam e até mesmo partilhar histórias no *Instagram* regularmente sobre informações de um jogo ou acontecimentos ocorridos antes e após o mesmo.

Os clubes desportivos cada vez mais pretendem destacar-se nas suas contas nas redes sociais de formas criativas, sendo que estas servem também para campanhas publicitárias e *marketing*. Cada vez mais é importante explorar novos caminhos e oportunidades para se diferenciarem e envolverem cada vez mais os seus adeptos. Temos o exemplo do clube inglês *Tottenham Hotspur*, que utiliza o *Twitter* e o *Instagram* para criar entusiasmo durante uma partida de um jogo. Faz uso de uma *hashtag* #COYS (“*Come on you Spurs!*”) no *Twitter* para fazer referência a este clube em todos os *tweets* que são postados como o exemplo representado no lado esquerdo da figura 2, onde para além de variadas informações sobre jogos ou notícias do clube, existe também a partilha vídeos, em primeira mão, pela equipa digital responsável das redes sociais do clube. No *Instagram*, são tipicamente fornecidas atualizações sobre um jogo que esteja a decorrer, anúncios de golos, algumas imagens do estádio, publicidade a marcas e também questionários para recolher informações sobre a opiniões dos adeptos, como por exemplo o melhor jogador em campo representado no lado direito da figura 2.



Figura 2: Conta do *Tottenham Hotspur* no *Twitter* e no *Instagram*

Como é possível observar na Figura 3, com base na empresa *Statista* (empresa especializada em dados de mercado e consumidores) [70] e na empresa *Social Blade* (serviço de rastreamento de estatísticas e análises de mídias sociais) [67], foi elaborado o seguinte gráfico que mostra o crescimento das redes sociais até ao ano de 2021. Posto isto, foram selecionadas e estudadas duas redes sociais que servem como base para extrair informações - o *Twitter* e *Instagram* - apesar de não se encontrarem em primeiro lugar no que toca à popularidade deste ano, destacam-se por primariamente serem feitas para partilhar informação reduzida e sucinta. Nas subsecções abaixo referentes a cada uma são listados todos os motivos chave para esta escolha.

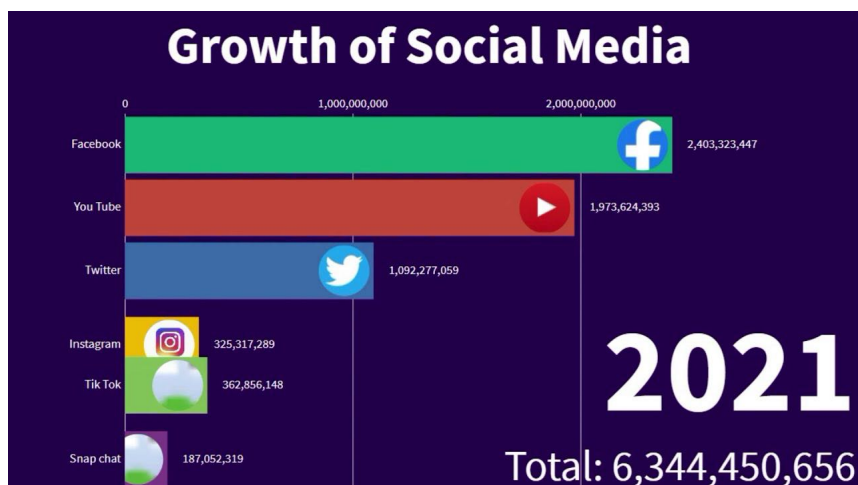


Figura 3: Evolução das redes sociais até ao ano 2021

### 2.1.1 Twitter

As redes sociais em geral e particularmente o *Twitter*, oferecem potencial para explorar um trabalho empírico que começa a desvendar novas relações sociais que se orientam em torno de sujeitos e objetos digitais [88]. São utilizadas para estudo de ciências sociais como referido anteriormente e servem como ferramentas de análise, visualização e previsão de dados, onde a originalidade se sobrepõe [3].

Criado em 2006, o *Twitter* é uma rede social e ferramenta de microblog que faz parte de um grupo de tecnologias geralmente classificadas como "*social media*", fornecendo conteúdo rico e interativo [39]. O *Twitter* registou um crescimento recorde de utilizadores no segundo trimestre de 2020, alcançando assim a marca de 186 milhões de utilizadores, um aumento de 34% em relação ao ano anterior de 2019 [39]. Nesta plataforma as pessoas são capazes de expressar e partilhar as suas opiniões instantaneamente.

Para a maioria, o *Facebook* é usado num contexto de privacidade relativa, pois permite partilhar informações entre uma rede estabelecida de contactos pessoais, ou seja, entre amigos na própria plataforma.

Em contraste, o *Twitter* não se limita a uma rede pré-aprovada sendo a maioria das contas registadas públicas. Os *tweets* são constituídos por mensagens curtas limitadas a 140 caracteres onde links, fotos ou imagens podem ser adicionadas para transmitir mais informações. Qualquer *tweets* pode ser re-emitido (o vólgo *retwittar*), citado ou pesquisado, com o potencial de ser visto por qualquer pessoa e permitir a interação uns com os outros [7].

Esta plataforma é uma das redes sociais mais utilizadas e influentes, sendo que retrata uma disponibilidade parcial e gratuita de dados, recebendo atualmente cerca de 190 milhões de *tweets* por dia sobre diversos tópicos onde são partilhadas opiniões sobre produtos ou serviços [47]. Os próprios termos e serviços declaram que as publicações dos utilizadores que são públicas, podem ser disponibilizadas a terceiros, tornando assim esta uma boa plataforma a ser utilizada para extração de dados. Esta plataforma é de longe a mais estudada por se diferenciar ao colocar os dados mais acessíveis aos investigadores. Dispõe de uma *API* de *streaming* de dados com cerca de 5 milhões de *tweets* recolhidos diariamente, onde

os utilizadores expressam publicamente as suas reações a determinados eventos e geram uma enorme fonte de dados para pesquisas sociais, funcionando assim como uma fonte de avaliação do sentimento do utilizador [29]. Essa análise de sentimento nos *tweets* pretende avaliar as opiniões das pessoas em sentimentos negativos ou positivos ou até mesmo em estados de emoção (por exemplo, feliz, stressado, triste, revoltado) através de métodos de análise de texto [26].

Com o enorme fluxo de dados gerados no *Twitter*, a diversidade de comentários e a distribuição de opiniões ao longo do tempo torna a análise dos dados do *Twitter* muito desafiante onde se destaca o desafio - trabalhar com enormes volumes de dados (*Big Data*) e com eles são apresentadas variadas técnicas para a análise de sentimentos e visualização dos mesmos:

- Baseado na extração, mapeamento e medição dos tópicos de análise de sentimento das opiniões do utilizador;
- Análise de fluxo que identifica *tweets* interessantes com base nas suas características de densidade, negatividade e influência;
- Mapas geográficos baseados em pixels que representam tipos de sentimentos que visualizam grandes volumes de dados numa única visualização.

Através da plataforma *Twitter*, podemos expôr dados de diversas maneiras e com diferentes intuitos. É possível uma exploração mais poderosa da informação contida nos *tweets* de diferentes utilizadores, combinando com a análise automatizada de sentimento, surgindo assim variados tipos de visualização.

É exemplo disso, o caso mostrado na Figura 4 e 5 sobre um tópico em específico - *Covid-19* - onde os objetos de interesse são padrões emergentes de relacionamento e é possível visualizar a sua interação com os atributos de identificação. Foi determinado um certo tópico - neste caso um dos grandes *trendings* do *Twitter*, selecionou-se e extrairam-se atributos deste tópico, detetou-se opiniões sobre os atributos e mediu-se o valor do sentimento.

Os dados são da AP Exata [5], uma empresa de análise de dados que pretendia estudar os tipos de sentimentos predominantes no *Twitter*, recolhendo informações dos *tweets* em Portugal durante um período de tempo sobre o covid-19, uma pandemia que veio abalar o mundo no ano de 2020. Esta empresa aplica um algoritmo para medir os diferentes tipos de sentimentos existentes e basearam-se na teoria das emoções do psicólogo norte-americano *Robert Plutchik* que distinguiu oito emoções primárias que considera relevantes e influenciam a escolha e decisões dos seres humanos - raiva, medo, tristeza, aversão, surpresa, antecipação, confiança e alegria [42].

Na figura 4 é possível observar o gráfico obtido sobre o sentimento dos utilizadores do *Twitter* ao longo do mês de Março onde *Robert* enquadra os sentimentos em acontecimentos, dizendo “Nota-se tristeza face à economia, confiança face à união de todos, mas o que predomina é o medo. Medo de se ser contaminado, ou de contaminar alguém.” Feita uma análise ao gráfico é possível observar que o medo e a tristeza são os sentimentos que predominam nas publicações sobre o novo coronavírus

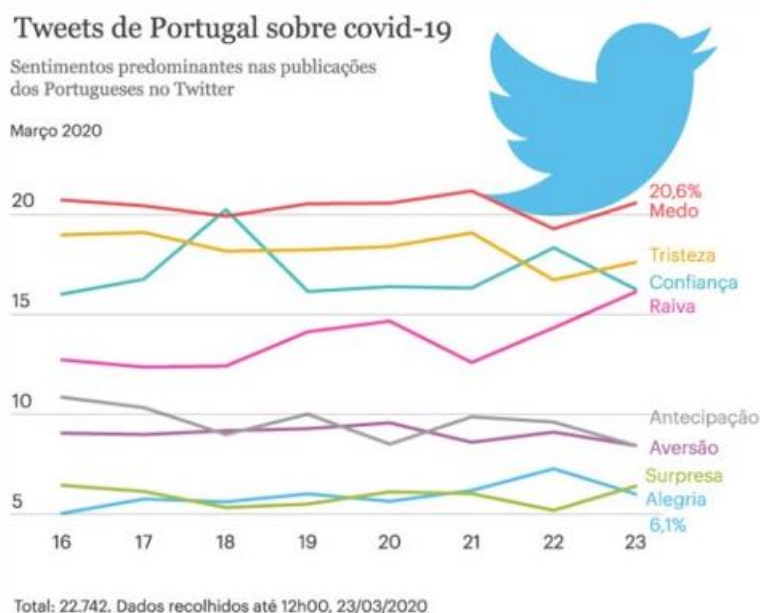


Figura 4: Tweets de Portugal sobre Covid-19 [42]

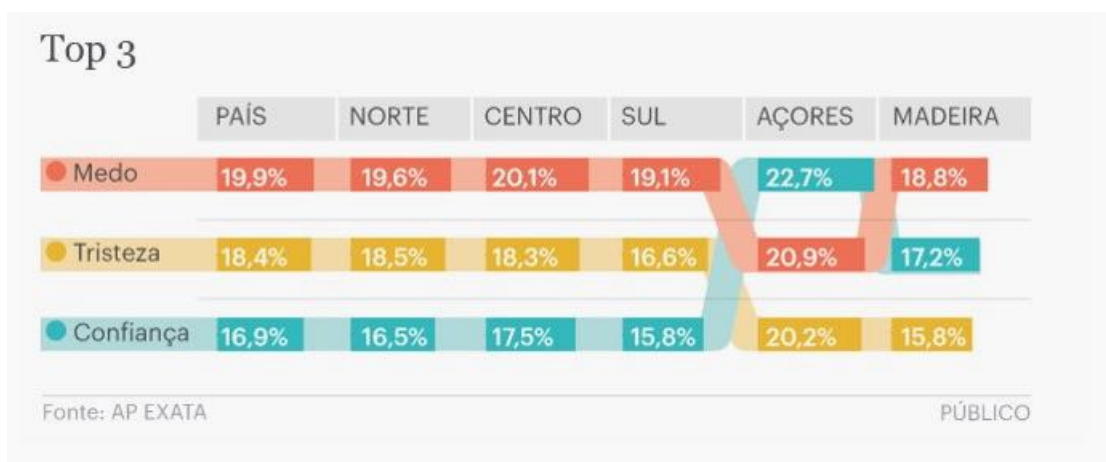


Figura 5: Tipo de sentimentos pelas regiões do país [42]

feitas pelos utilizadores nesta plataforma em Portugal. A confiança vem em terceiro lugar mas é possível observar uma enorme quebra no dia 18 de Março que se deve, supostamente, à data onde foi declarado o estado de emergência do país. Seguidamente e em quarto lugar, o sentimento raiva começou a aumentar significativamente a partir desse dia devido à entrada em vigor das novas medidas aplicadas pelo governo. O jornal Público [42] relata que "Foram identificadas no *Twitter* certas expressões (por exemplo, "guerra ao vírus", "situação grave" e "pânico total") que dominavam em cerca de 20% (4500) das publicações partilhadas desde dia 16 de Março, data em que as aulas começaram a ser lecionadas através da Internet. Entre dia 18 de Março e dia 19 de Março, a confiança presente nos *tweets* desceu de 20% para 16%". Como é possível analisar, o sentimento da raiva a partir do dia 21 de Março começou novamente a

ascender paralelamente em conjunto com o sentimento de medo e tristeza, enquanto o sentimento de confiança decrescia.

No gráfico da Figura 5 é possível visualizar e analisar o tipo de sentimentos mais predominantes nas principais regiões do país através das publicações do *Twitter* em Portugal. Esta tabela mostra que nos arquipélagos, a confiança é mais forte do que a tristeza. Segundo o jornal Público [42] deve-se ao facto de que as pessoas idealizam que numa ilha é mais fácil controlar a pandemia e por essa mesma razão, as pessoas sentem-se menos receosas. Afirmam também que "no Norte, Centro e Sul o tipo de sentimento predominante é comum a todos eles e é sem dúvida o medo (a rondar os 19% e 20%), seguido da tristeza (18%) e em último a confiança (a rondar entre os 16% e os 17%)".

Este exemplo é baseado na extração, mapeamento e medição dos tópicos de análise de sentimento das opiniões do utilizador, sendo que agora vai ser dado outro exemplo. Este passa pela representação dados de sentimentos através de mapas geográficos que representam em pixeis o sentimento e permitem a visualização de grandes quantidades de dados através da interatividade. A ideia principal é permitir aos utilizadores otimizar o grau de sobreposição e distorção para uma melhor visualização possível. Cada píxel é formado por um único *tweet* relativo aos sentimentos que se dividem em três categorias: negativo, neutro e positivo - cada um destes representado por uma cor atribuída. Para além de ser possível visualizar o tipo de sentimentos predominantes em cada lugar, revela também o volume de comentários que foi publicado no *Twitter* de cada país.

Analisando o mapa geográfico que representa o sentimento presente nos *tweets* sobre o filme *Panda Kung-Fu* é obviamente notável que o sentimento mais predominante é o positivo face a este tópico como é possível visualizar na Figura 6. Por outro lado, entende-se que os comentários com as mesmas posições são colocadas em torno da posição original com base em cada valor de sentimento. Exemplo disto é, a borda vermelha em torno da área circular presente no país Chile que representa *tweets* negativos (vermelhos) nas mesmas posições que os *tweets* positivos (verdes) centrados na área circular [47].

O *Twitter* abrange todo o tipo de conteúdo que é possível imaginar mas o tema de desporto tornou-se um veículo de *social media* favorito pois permite fornecer variadas perspectivas sobre os jogos a diversas organizações, funcionando como um veículo de publicidade e *marketing*.

Um estudo [39] observou que 15% dos utilizadores que são os profissionais de *marketing* seguem mais de 2.000 utilizadores, em oposição a 0,29% dos utilizadores gerais do *Twitter*. Temos o exemplo do clube *New England Patriots* que posta *tweets* frequentemente para aumentar o entusiasmo sobre os próximos jogos, fornece comentários jogada a jogada durante o jogo e, terminado cada jogo é possível acompanhar repetições do jogo e as entrevistas com jogadores e ainda compartilham fotos e *tweets* de fãs, o que cria uma conexão e envolvimento ainda maior [39]. O objetivo pretendido é o aumento de envolvimento com os adeptos, construir relacionamentos, divulgar informações sobre determinados eventos e gerar receitas.

Esta plataforma dispõe de recursos de *trending* (tópicos mais falados em determinados momentos) utilizando *hashtags* de tendência ou apenas palavras ilustrativas de um assunto relevante o que se torna interessante pois esta plataforma fornece as suas próprias ferramentas de análise.



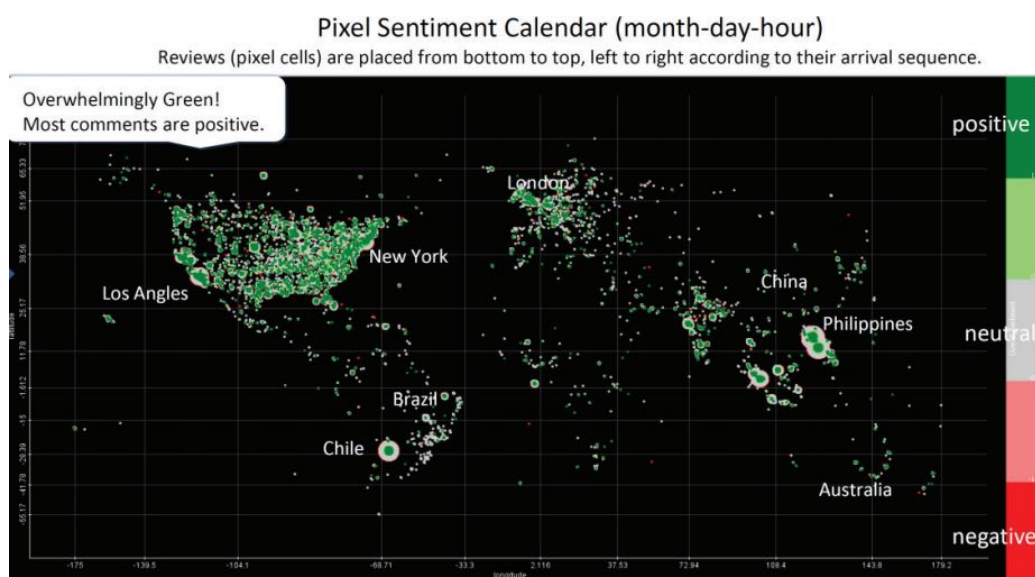


Figura 6: Mapa Geográfico que representa o tipo de sentimento [47]

Uma análise do conteúdo dos *tweets* feitos pelos atletas mostra que os mesmos continuam interações com os adeptos, informações privilegiadas sobre a equipa ou o desporto, discussões sobre outros desportos, promoções de publicidade, links para outro conteúdo e entretenimento de uma forma geral [39].

### 2.1.1.1 Twitter no Desporto

Os níveis de atividade da ação de ver jogos na televisão são enormes e vieram a ser complementados pelas redes sociais para um aumento do envolvimento do público, onde são discutidas e partilhadas diversas opiniões entre os utilizadores. O *Twitter* pode ser uma ferramenta extremamente valiosa para redes de TV, agências e patrocinadores que procuram compreender melhor o público. Na Figura 7 está representado um exemplo de um *tweet* no canal do *Manchester United* sobre o homem do jogo, sendo este tipo de informações importantes para captar as opiniões dos utilizadores. Dessa forma, foram realizados vários estudos para entender a relação existente entre as transmissões ao vivo de jogos com o envolvimento dos adeptos nas redes sociais. A empresa *Nielsen Social* [49] comparou os resultados de quase 300 espectadores a assistir a transmissões ao vivo de oito programas no horário nobre com a atividade minuto a minuto do *Twitter* para os mesmos programas. Com este estudo foi encontrada uma forte correlação de cerca de 79,5% com o envolvimento neurológico dos utilizadores (uma combinação de emoções, memórias e atenções), o que representa os grandes níveis de envolvimento entre o público em geral.

Na figura 8 retirada do blog do *Twitter* é representado o volume de tendências para um determinado episódio no *Empire* (website de conteúdo televisivo e jogos), medido pela *Nielsen* (empresa global de informação, dados e medição), onde é visível que os principais *tweets* por minuto se alinham com certos



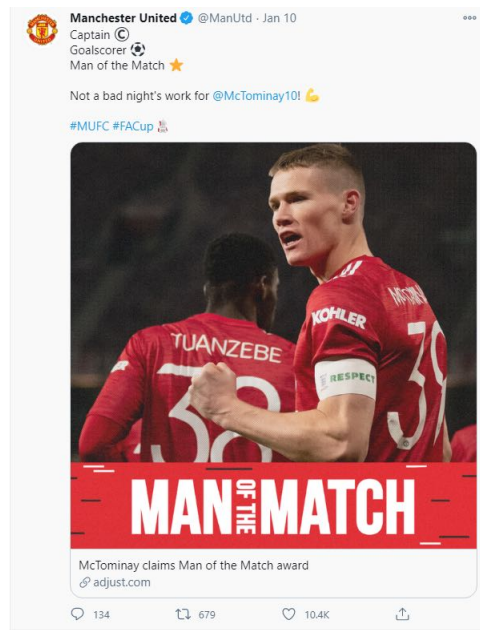


Figura 7: Tweet no canal do Manchester United

picos dramáticos na programação, representando assim a correlação com a atividade presente no *Twitter* versus TV [49].

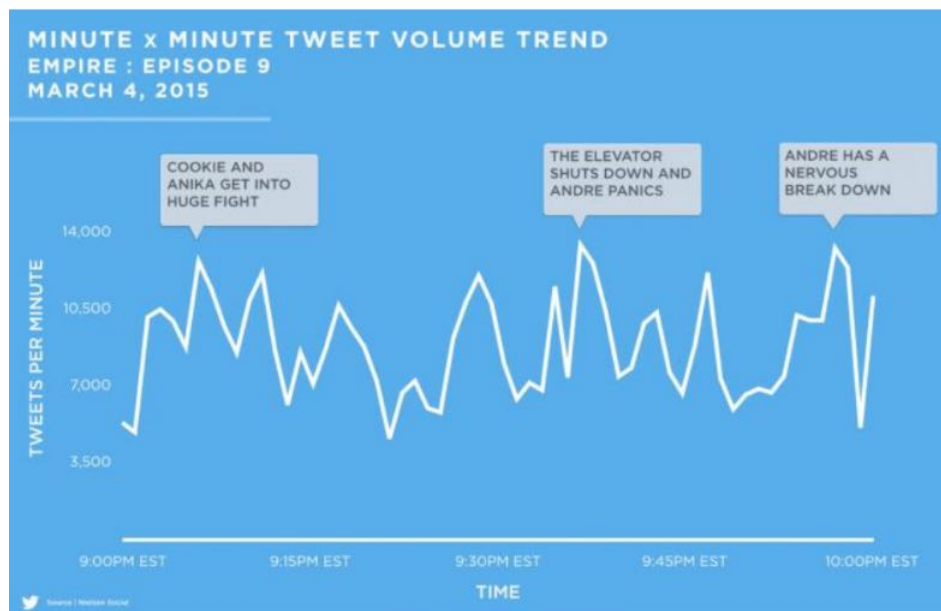


Figura 8: Relação volume de tendências/tempo no *Twitter* [49]

Desta forma, é possível concluir que o *Twitter* pode e influencia o comportamento dos adeptos paralelamente com a típica TV tradicional. Esta rede social representa uma ferramenta perspicaz para a tomada de decisão baseada em dados, permitindo construir audiências e impulsionar ações. Contempla benefícios para gerar audiências que se estendem também para as marcas e profissionais de *marketing*.

Um estudo [83] feito pela equipa de *marketing* representado na Figura 9 da empresa *Twitter* garante que "os utilizadores mostram níveis mais elevados de relacionamento (+ 31%) e memorização (+ 35%) quando estão a acompanhar na TV e no *Twitter* simultaneamente do que apenas a acompanhar o evento na TV. Para além de que, mesmo quando esta rede social é a única fonte para acompanhar um evento desportivo, os utilizadores continuam a exibir um maior envolvimento (+ 60%) e memorização (+ 59%) em comparação com a exibição apenas na TV, demonstrando a experiência envolvente que o *Twitter* proporciona".

Para além destas estatísticas que foram feitas, afirmam também que a publicidade e anúncios têm mais impacto nos adeptos que acompanham os desportos no *Twitter* e na TV paralelamente. O *Twitter* realmente muda a experiência do utilizador ao assistir a desportos ao vivo na televisão, sendo usado como um complemento.

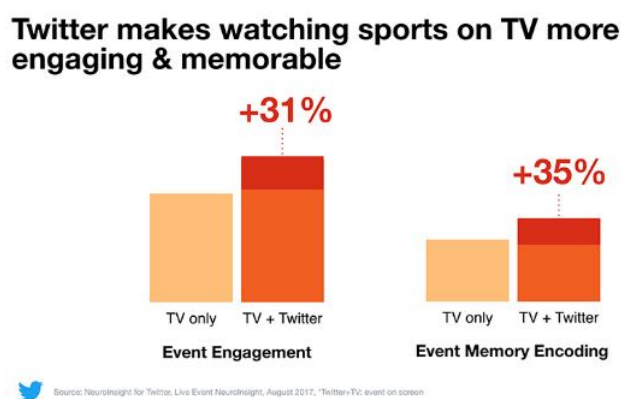


Figura 9: *Twitter* Vs TV - Tipo de evento mais memorável [83]

Este tipo de exemplos são bastantes relevantes de maneira a ser perceptível as opiniões utilizadas nas redes sociais que são fundamentais nos dias de hoje. A interação dos utilizadores deve ser importante para qualquer empresa, que serve para tirar conclusões e de certa forma, engendrar maneiras de incentivar ainda mais os adeptos para com os seus clubes e sem dúvida que o *Twitter* é uma das plataformas que contempla inúmeras características importantes para a análise de dados.

### 2.1.2 Instagram

O *Instagram* é uma plataforma social de partilha de fotografias que se encontra activa desde 2010. Tornou-se numa ferramenta cada vez mais popular para as organizações desportivas compartilharem conteúdos de recursos visuais. Como é possível verificar na Figura 10 o *Instagram* apresenta no mesmo período uma forte tendência de subida, duplicando o volume de interações.

Desde 2013 esta plataforma já multiplicou 5 vezes o número de apreciações entre os inquiridos neste estudo da *Marktest* [57], "aproximando-se já dos 85%, contra os 98% do *Facebook* e tem crescido de forma evidente entre os utilizadores de redes sociais em Portugal, duplicando de 2018 para 2019". É

utilizado por mais de um mil milhões de utilizadores no mundo. Permite capturar uma fotografia/vídeos com o nosso dispositivo móvel e compartilhá-los na hora com os seguidores, através de publicações no perfil ou até mesmo através das famosas histórias que duram 24 horas. Os utilizadores podem seguir ou ser seguidos por outros membros, gostar das fotos, realizar comentários às fotos publicadas ou reagir com *emojis* às histórias.

Esta plataforma é um espaço de criatividade que oferece grandes oportunidades de *marketing* e publicidade gratuita para profissionais, negócios e marcas. Os utilizadores produzem dados diariamente sendo esta plataforma uma fonte de dados incomparável para investigadores de redes sociais, servindo para visualizar padrões e divisões sociais.

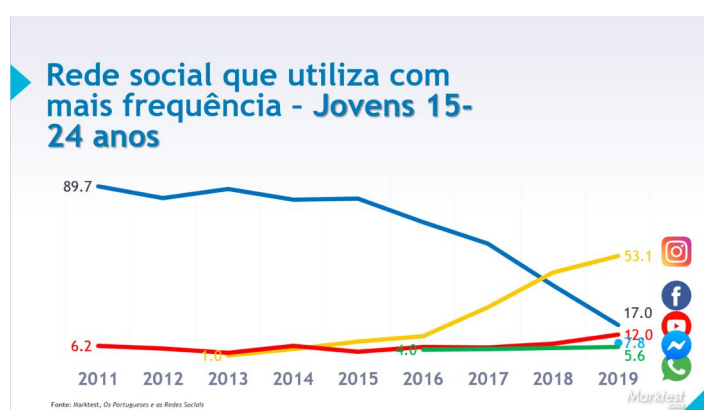


Figura 10: Evolução do *Instagram* ao longo do tempo [59]

O *Instagram* oferece um espaço essencial para a partilha de conteúdo e permite que seja possível criar uma conexão com a força da cultura através das vozes dos criadores de conteúdo que a definem e propagam. Desta forma, esta plataforma proporciona [37] a capacidade de incentivar os movimentos sociais e culturais servindo como um atalho para gerar sentimentos.

Embora o sucesso em campo continue a ser importante, as equipas desportivas profissionais precisam de se adaptar à indústria de entretenimento em constante expansão, aproveitando o valor da sua marca. Existe um aumento da popularidade do *Instagram* entre as equipas desportivas, atletas e fãs, combinados com um potencial de *marketing* à medida que amadurecem. Esta plataforma tem vindo a ser utilizada para estes fins e principalmente para investigações do comportamento do consumidor relativamente à publicidade de determinadas marcas.

As histórias do *marketing* surgiram e vieram tornar-se numa das grandes tendências para os consumidores de redes sociais. A mesma lógica se aplica nos desportos, principalmente com atletas. Através das histórias desta rede social, os jogadores podem mostrar o seu dia-a-dia.

O *Instagram* permite aproximar as pessoas a produtos ou marcas, através da publicidade de criadores de conteúdo. Cada vez mais, as empresas querem compartilhar a história da marca em parceria com criadores de conteúdo, produzindo um conteúdo autêntico, popular e atrativo, capaz de gerar resultados comerciais. Esta plataforma disponibiliza certos mecanismos de publicidade que permite que

os criadores de conteúdo marquem os seus produtos nas publicações deles. Dessa forma, as pessoas podem descobrir a sua empresa e comprar na própria, sem ser necessário sair do *Instagram*, como é possível observar na Figura 11 na conta oficial do Cristiano Ronaldo, uma publicação em parceria com uma marca muito conhecida, a *Nike*.

É a terceira celebridade mundial mais bem paga por publicação, sendo o futebolista o único português no top 10 com maior número de seguidores no mundo. Sendo este uma das pessoas mais influentes no mundo, a publicidade nas suas publicações gera imensas receitas para as empresas de determinadas marcas [65].



Figura 11: Publicação do jogador Cristiano Ronaldo

Para além disso, a análise de dados do *Instagram* serve para aprender sobre as diferenças culturais existentes. Temos o exemplo de um projeto [14] representado na Figura 12 que analisa o uso dos filtros nas histórias do *Instagram* pela quantidade de número de fotos. Uma das grandes vantagens deste tipo de análise passa por perceber qual o filtro mais utilizado e que melhora a qualidade de uma fotografia, alcançando assim um maior número de gostos.

A própria plataforma disponibiliza o "*Instagram-analytics*" [76] que contém uma grande quantidade de dados em bruto do *Instagram* para uso dos utilizadores/investigadores e respetiva análise de dados. Outro exemplo como o *Instastats* (sistema de estatísticas) [38] que pode ser acedido através de *scripts* na linguagem de programação *Python* para extrair dados da *API* do *Instagram*. Cada vez mais são utilizadas este tipo de ferramentas para reunir informações de dados disponíveis nesta rede social para posterior análise e estudo.

Para além disso, o projeto falado anteriormente [14] também analisa e representa as *hashtags* mais populares por 50 cidades dos Estados Unidos representando assim eventos populares ou termos específicos que existam para diferentes cidades. Um dos objetivos deste projeto é aprender sobre os diferentes tipos de culturas e diferenças de conteúdo ajudando assim a captar as tendências artísticas e tendências

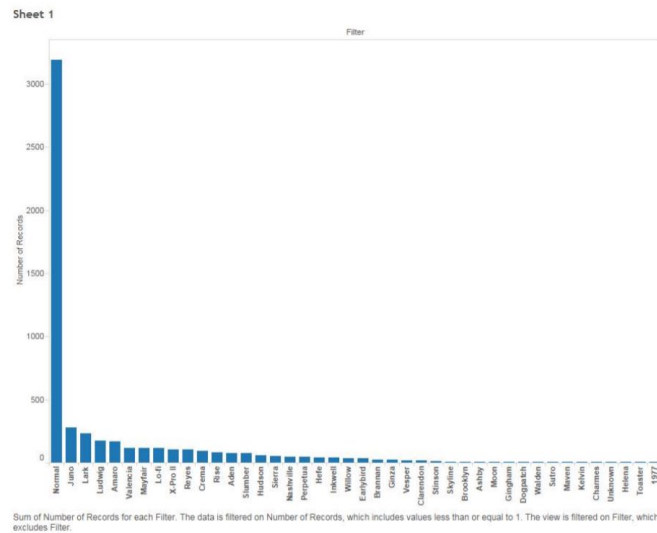


Figura 12: Gráfico que analisa os filtros nas histórias do *Instagram* [14]

de eventos. O projeto dispõe de um mapa geográfico interativo representado por cada cidade mostrado na Figura 13 e após clicar no ponto da cidade o mesmo fornece um nuvem de palavras com uma exibição das *tags* mais utilizadas nessa mesma cidade. Na nuvem de palavras o tamanho maior do texto significa mais publicações com essa *tag* e menor tamanho de texto significa menos publicações com essa *tag*.

Analisa amostras de um conjunto de dados de cerca de 550.000 fotos do *Instagram* da cidade de *Nova York* e *Los Angeles* [14]. Como é possível verificar na nuvem de palavras de *Los Angeles* na Figura 15, visualiza-se as palavras "graduação" e "moda", o que pode implicar que esta cidade é uma cidade energética. Por outro lado na Figura 14, para a cidade de *Nova York*, observa-se palavras referentes a muitas animações japonesas. Uma possível explicação é que *Nova York* podia estar a realizar uma exposição de anime naquele momento [14].

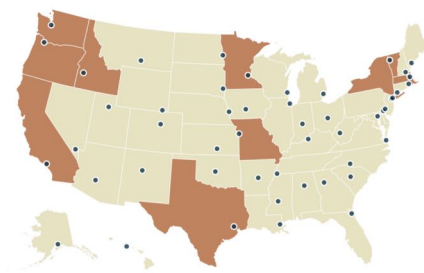


Figura 13: Mapa geográfico interativo dos Estados Unidos [14]

### 2.1.2.1 Instagram no Desporto

A *social media* é um espaço crescente para a comunicação e as imagens e informações compartilhadas nas redes sociais frequentemente são um fator de extrema importância para aumentar o envolvimento



Figura 14: Nuvem de Palavras das *hashtags* em Nova York [14]

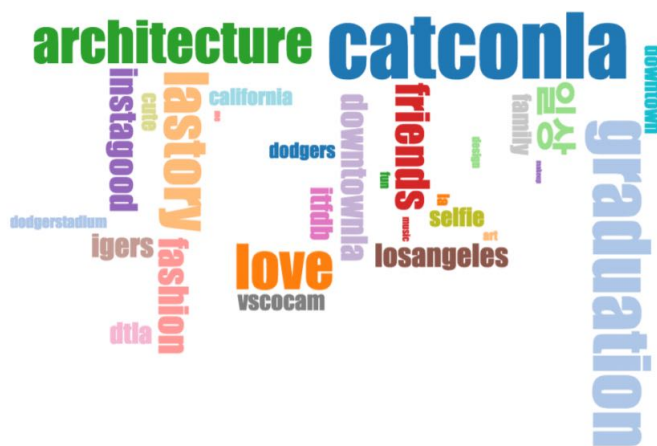


Figura 15: Nuvem de Palavras das *hashtags* em Los Angeles [14]

do público. A comunicação visual tornou-se numa forma dominante de nos comunicarmos e os criadores de conteúdo costumam usar uma mistura de texto e imagens para transmitir determinadas mensagens.

A notícia tornou-se cada vez mais um bem-estar social e as redes sociais são um canal de comunicação bastante importante e que atinge de uma forma significativa toda a população. São utilizadas para debates e discussões que envolvem e influenciam ideologias. [46].

É importante entender quais os tipos de imagens e formas de comunicar que geram um maior relacionamento com o público nas redes sociais. Posto isto, o Romney e Johnson realizaram um estudo [60] para perceber a relação entre uma imagem narrativa e meta-comunicativa com os quatro principais canais de desporto no Instagram : *ESPN*, *FOX Sports Media Group*, *CBS Sports* e *National Broadcasting Corporation Sports* - de modo a tentar perceber a relação entre o tipo de imagem mais cativante para o público, através do número de gostos representado na Figura 16 e dos números de comentários na Figura 17.

Uma imagem narrativa é constituída por um autor, uma finalidade e uma transacção, normalmente em movimento e uma imagem comunicativa passa por imagens transformadas/manipuladas, com adição de textos por exemplo. Na Figura 18 encontra-se um exemplo deste tipo de imagem. Romney e Johnson dizem que "estudos semelhantes sobre a estrutura de notícias de televisão concordam que

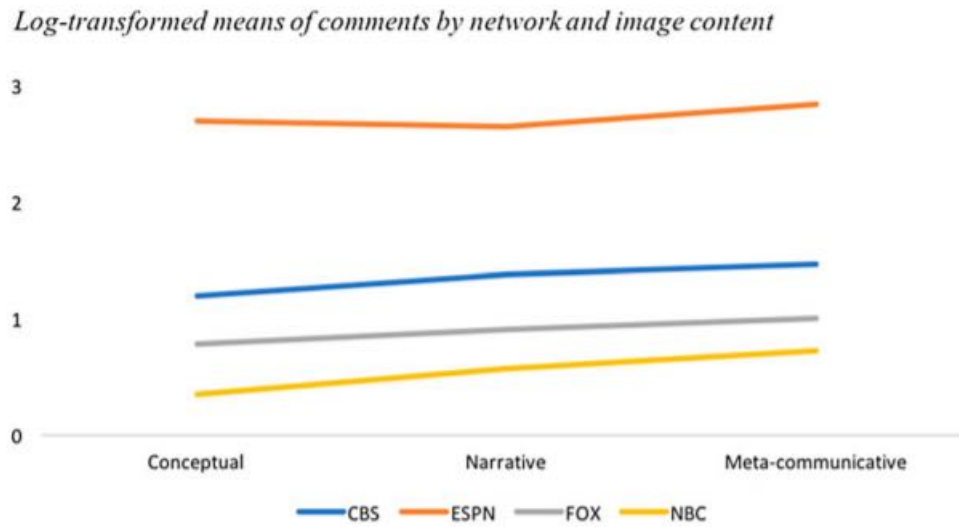


Figura 16: Relação tipo de imagem vs número de gostos [60]

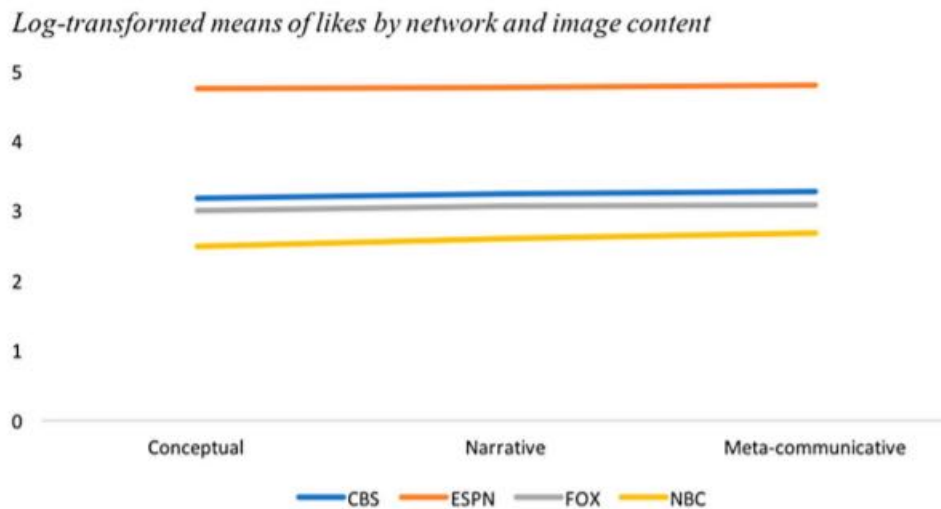


Figura 17: Relação tipo de imagem vs número de comentários [60]

uma imagem narrativa é um método eficaz para se conectar com o público sendo que este tipo de imagem age narrativamente da mesma maneira que os textos tradicionais”. Para além disso, compartilhar imagens narrativas através de plataformas digitais pode servir como um importante contribuidor para o desenvolvimento da comunidade [60]. Os resultados deste estudo indicaram que as imagens que contêm qualidades narrativas ou comunicativas geram um maior relacionamento entre o público. O tipo de imagem comunicativa, no qual as imagens são intensificadas através de manipulação, inclusão de texto ou elementos gráficos, mostrou sinais de relacionamento mais fortes, tanto no número de gostos quanto



nos comentários.



Figura 18: Imagem narrativa vs Imagem comunicativa [60]

Por outro lado, estudou-se um projeto [46] que realizou uma pesquisa quantitativa com o objetivo de avaliar o conteúdo (imagem e texto) das publicações publicadas na rede social *Instagram* da Confederação Brasileira de Ginástica (@cbginastica). Sendo o Brasil um dos maiores consumidores destes ambientes virtuais, em 2020 considerou-se que 85% dos 66,3 milhões de brasileiros utilizam redes sociais. Dessa forma, foram reunidas todas as publicações feitas de Maio de 2013 a Dezembro de 2016 com um total de 2679 publicações [46].

Conforme as publicações, foram estudados vários campos como as modalidades de ginástica retratadas nas publicações, o tipo de evento e a pessoa a quem se refere cada publicação.

Com base nos dados recolhidos, obtiveram-se vários resultados visualmente no tipo de evento citado nas publicações como as competições nacionais, internacionais, treinos, cursos, entrevistas, visitas, eventos e outros. O objetivo era verificar qual destes tipos de eventos era mais frequente no *Instagram* da CBG e como é visível na Figura 19 a maioria das publicações foca-se na participação nos campeonatos internacionais com 59,8% e seguidamente nos campeonatos nacionais com 14,7%, enquanto que os eventos como festivais e cursos foram raramente mencionados.

Para além da representação visual referida anteriormente, foi também criado um outro gráfico mas desta vez referente à pessoa a quem a publicação se refere ou seja, se a publicação era sobre o atleta, técnico, árbitro, público, múltiplos (por exemplo sobre o atleta e o técnico) ou sobre outros. Esta representação está demonstrada na Figura 20. Referem que durante a análise foi perceptível que quando dois sujeitos apareciam na foto, o número de gostos era significativo quando as publicações eram sobre o atleta e o técnico.

Por outro lado, quando eram outros sujeitos, como por exemplo o atleta e o dirigente os números representavam menos de 1%. Resumidamente o atleta revela uma grande importância nas publicações nas redes sociais com 66% de interação a partir dos utilizadores.



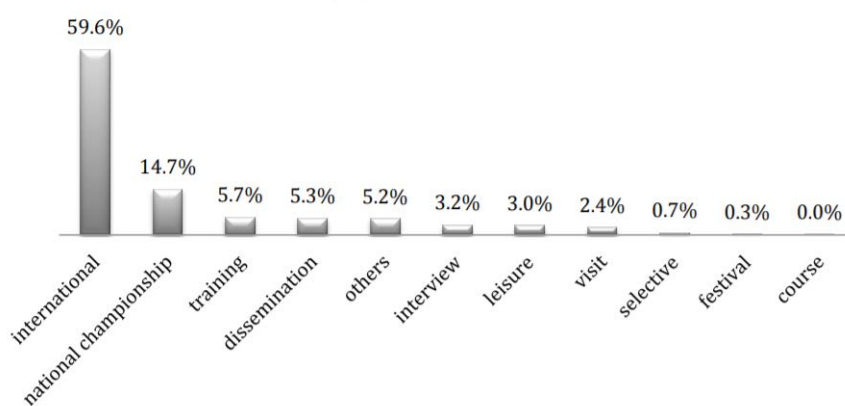


Figura 19: Percentagem do tipo de evento no *Instagram* da CBG [46]

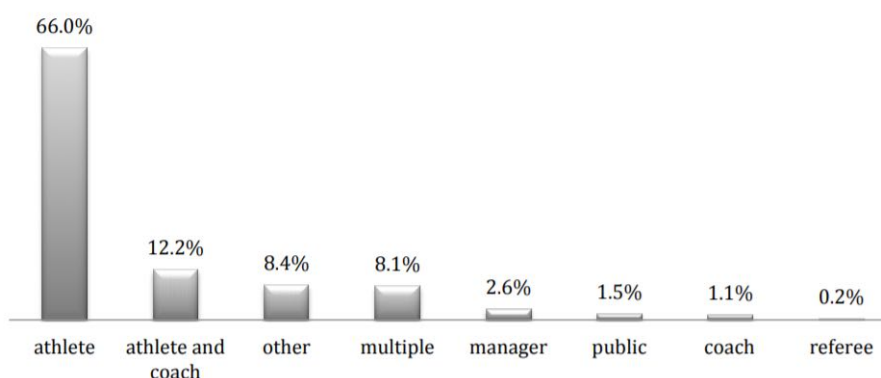


Figura 20: Protagonistas das publicações no *Instagram* da CBG [46]

## 2.2 Visualização de dados no Desporto

As organizações desportivas perceberam e têm vindo a aumentar a importância do uso das redes sociais para uma maior conexão com os adeptos. Hoje em dia temos disponível *feeds* de notícias, *blogs*, sites interativos e fóruns de discussão sobre todo o tipo de temas relacionado com o desporto.

Quase todos os desportos, sejam eles individuais ou coletivos, produzem grandes quantidades de dados como resultado dos jogos, dados que podem ser estudados e usados para melhorar o futuro assim como aumentar a diversão dos adeptos, auxiliar nas informações das redes sociais e nas tomadas de decisão. Esta área da visualização de dados no desporto tem vindo a ganhar cada vez mais popularidade focando-se mais especificamente no basebol, basquetebol, futebol, futebol americano, hóquei no gelo e no golfe. Contudo, ainda existe escassez de ferramentas de representação de dados desportivos que permitam a exploração e análise de dados interativamente [85].

A análise de desportos tem sido uma tendência extremamente popular e que continua a crescer significativamente nos últimos 20 anos [62]. Existe uma enorme afluência na visualização de determinados desportos na televisão, maioritariamente por indivíduos do sexo masculino e é necessário os adeptos/espectadores perceberem e inculcirem algumas percepções significativas sobre um determinado evento.

Como é possível visualizar na Figura 21, o número de artigos científicos sobre a visualização de dados no desporto, tanto a nível académico como a nível de investigação, tem sofrido um gradual aumento ao longo dos anos, sendo que este é um tema cada vez mais investigado e por essa razão, existe uma pesquisa cada vez maior de informações relevantes sobre este tema.

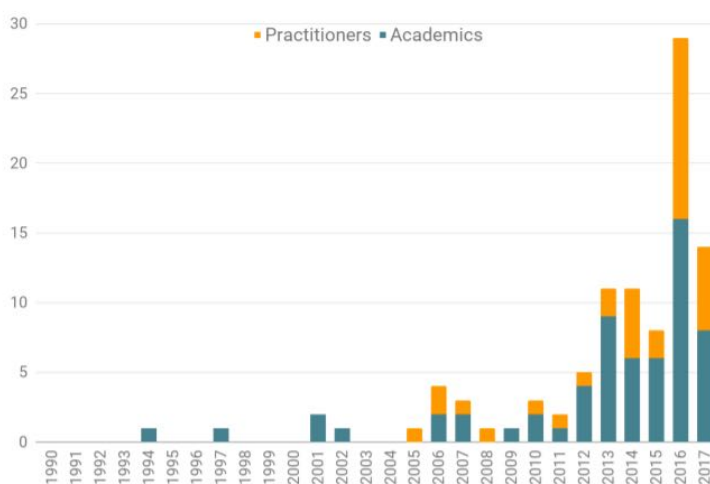


Figura 21: Artigos sobre dados desportivos ao longo dos anos [62]

Atualmente, existem imensos dados sobre clubes, jogadores e eventos desportivos e a contínua massificação de conteúdo multimédia existente tem promovido o consumo desses conteúdos e como resultado a análise de dados no desporto está-se a tornar cada vez mais extensa e diversificada. O acesso instantâneo de informações nas plataformas móveis através da internet e a capacidade de interagir com essas informações está a mudar a forma como os adeptos consomem certos tipos de conteúdo. Existem grandes evidências de que a análise bem sucedida de dados desportivos pode resultar num melhor desempenho de uma equipa [62].

Hoje em dia, basta aceder a diversos *websites*, fóruns, aplicações para telemóveis e até mesmo jornais em papel para ter acesso a dados estatísticos e informativos sobre desporto mas as visualizações interativas tornaram-se mais acessíveis e significativas do que as estatísticas tradicionais. Os sites de transmissão de televisão e apostas usam cada vez mais este tipo de estatísticas e gráficos para avaliar e prever eventos desportivos [62].

No desporto a recolha de estatísticas dos jogos não é nada de novo mas surgiu uma nova era de análise de dados, onde os valores dos dados são cada vez mais importantes e transformou a maneira como um clube e os fãs interagem com os eventos no campo. Para os envolvidos no jogo, a análise permite um entendimento mais objetivo e profundo do que os utilizadores pensam e para os fãs, pode revelar pormenores e informações sobre um determinado clube ou jogador e, em última análise, facilitar uma compreensão mais profunda de determinados temas que os clubes queiram ver resolvidos.

### 2.2.0.1 Plataformas de visualização de dados desportivos

A visualização de dados é importante em todos os setores da indústria mas no que toca ao desporto o efeito talvez seja o mais óbvio. Hoje em dia existe um leque enorme de websites ou aplicações que exibem estatísticas de clubes e jogadores durante e após os jogos. Seguidamente, vão ser retratadas algumas delas.

Tem-se o exemplo do *website SofaScore* [68] representado na Figura 22, um *software* de desporto muito poderoso onde se pode consultar jogos ao vivo, resultados e estatísticas. Contempla mais de 20 tipos de desportos, mais de 5.000 ligas e torneios e milhões de eventos.

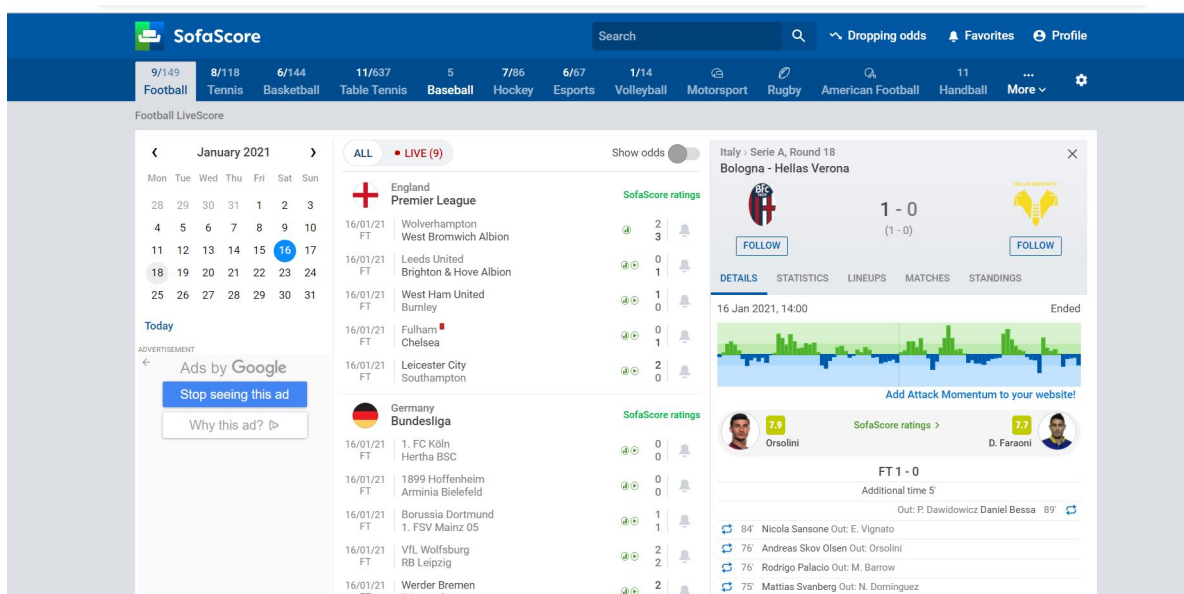


Figura 22: Website da plataforma SofaScore [68]

Este *website* permite aos utilizadores acompanharem todas as ligas e competições de uma ampla gama de desportos: futebol, basquetebol, hóquei no gelo, ténis, críquete, basebol, rugby, futebol americano, andebol, voleibol, ciclismo, pólo aquático, futsal, snooker, dardos, badminton, floorball, bandy, ténis de mesa, voleibol de praia e diversos desportos de automobilismo, nomeadamente cronometragem do Formula Live, MotoGP, Superbike, Nascar, rally e *Deutsche Tourenwagen Masters*. Segundo o *website* do *Google Play*, esta plataforma permite [55] visualizar todos os placares e jogos ao vivo da *Premier League* e outras ligas europeias, ter acesso a quem marcou um golo e obter estatísticas detalhadas sobre jogadores e outros eventos desportivos, visualizar todas as transferências, pontos fortes e fracos do jogador para uma visão geral completa do desempenho de um jogador numa temporada assim como todos os movimentos de um jogador, nomeadamente as posições, as jogadas feitas e perdidas, entre outros. Para além disso, contempla mais de 300 estatísticas utilizadas por jogadores profissionais, sites de notícias e analistas e é possível representar graficamente cinco atributos principais dos jogadores para descobrir rapidamente o quão completo um determinado jogador é.

Como outro exemplo, tem-se a *Prozone Sports* [11], uma empresa que oferece *software* de análise de desempenho de alto nível para as equipas desportivas, incluindo a visualização de dados a ser visível na Figura 23. Os dados não são nada sem uma visualização adequada. Esta empresa pega em dados complexos e reúne-os de forma a serem compreendidos por qualquer pessoa, sendo um profissional ou não, e a aplicação de uma boa visualização de dados contribui para a sua boa credibilidade. Temos o exemplo de um treinador, onde o seu trabalho é melhorado quando os dados são de fácil assimilação e compreensão, permitindo ao mesmo dedicar mais e melhor o seu tempo com os atletas.

Empresas como a *Prozone* e também a *Opta Sports* [89] fornecem diferentes tipos de visualização sobre determinados jogos e jogadores tendo um grande impacto sobre os distribuidores e fabricantes de videojogos. O *software* de visualização da *Prozone* foi adicionado a uma séria bastante popular de videojogos chamada *Football Manager* [27], dando aos jogadores a capacidade de examinar dados sobre a sua equipa de maneira semelhante às equipas profissionais.

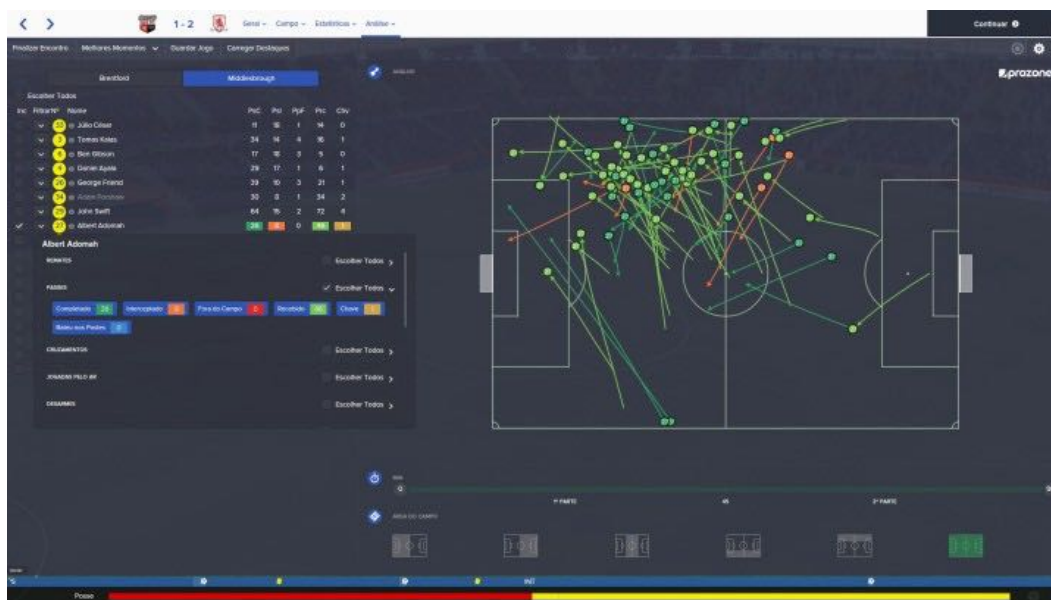


Figura 23: Software da Prozone Sports [11]

Surgiu também uma plataforma chamada *NBAVi* [51] que permite a visualização e exploração de dados dos jogadores e das equipas das últimas temporadas da *NBA*. Esta aplicação utiliza um gráfico de dispersão onde é possível selecionar a equipa a exibir assim como os tipos de dados a serem exibidos no eixo x e y e as cores que representam os dados. Como é possível visualizar na Figura 24 foi selecionada a equipa da *NBA San Antonio Spurs*, e no campo do eixo do x fez-se uso dos pontos da equipa e para o eixo do y do número total de *rebounds* da equipa. *Rebounds* significa uma estatística concedida a um jogador que recupera a bola após uma cesta de campo perdida ou de um lance livre. Abaixo da exibição do gráfico de dispersão, há um resumo dos jogos da equipa durante o ano, mostrando quantos jogos ganharam ou perderam e a pontuação correspondente. Para além disso, é um sistema interativo onde existe a possibilidade de mover o cursor do rato sobre um determinado ponto no gráfico de forma a obter

mais detalhe sobre o mesmo, como por exemplo, a qual jogador e jogo o ponto representa. É possível também clicar numa marca dentro do gráfico de dispersão ou numa das imagens dos jogadores que estão posicionados à direita para saber mais sobre cada um deles [51].

Esta plataforma possui um menu superior no canto esquerdo onde pode ser selecionada uma equipa em específico atualizando todos os gráficos de dispersão com as performances dos jogadores da equipa selecionada, em cada jogo da temporada. É possível selecionar todas as equipas do *NBA* simultaneamente onde marcas no gráfico de dispersão representam uma equipa de um determinado jogo.

Mostra um enorme conjunto de dados como os minutos que foram jogados, tentativas, golos e percentagens de *field goal* (forma de pontuar no futebol americano), os lances livres feitos, tentativas e percentagens, ressaltos ofensivos e defensivos ( é uma estatística concedida a um jogador que recupera a bola após uma cesta de campo perdida ou lance livre) , o número de assistências, roubos de bola, bloqueios, reviravoltas no jogo, faltas pessoais e pontos [51].

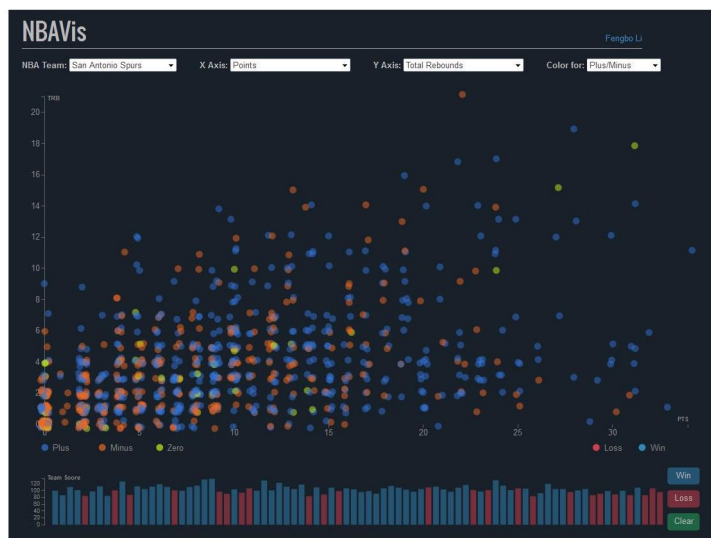


Figura 24: Plataforma NBAVis [51]

### 2.2.0.2 Tipos de dados desportivos

Os eventos desportivos geram grandes quantidades de dados, em tempo real, que podem e devem ser aproveitados pela *social media* para enriquecer o entretenimento dos adeptos. Por outro lado, os atletas vêm o retorno da sua exposição nas redes sociais através da rentabilização da sua imagem individual e da mesma forma, os clubes em que jogam. Quanto às organizações e empresas podem produzir uma compreensão mais profunda do próprio desporto. Estes dados quando são organizados e transformados, fornecem dados valiosos e potencializam uma nova visão no campo do desporto.

Existem diferentes tipos de dados desportivos que podem ser usados para visualização de dados, como por exemplo dados que contêm estatísticas de resumos de um evento desportivo, dados sobre

determinadas ações e trajetórias no jogo ou até sobre tipos de desportos, clubes, jogadores, participantes, entre outros.

A combinação de rede social com dados de treino e competição mostram incentivos e motivação valiosos para o envolvimento em desportos e atividades físicas.

A grande quantidade de dados vindos do desporto até podem ser utilizados na medicina desportiva, sendo que podem ser tomadas decisões baseadas nos dados como preparar planos de treino, prever riscos de lesões nos atletas e preparar estratégias de recuperação [9].

Segundo *Basole* [9], os dados desportivos tendem a ser hiper variáveis, temporais, relacionais, hierárquicos ou uma combinação dos mesmos, o que leva a alguns desafios de visualização.

O sistema de dados da "*Major League Baseball*" [73], que é a principal liga de beisebol profissional nos Estados Unidos da América, permite rastrear as localizações dos jogadores e das bolas, bem como eventos de jogo significativos. Segundo *Lage, et al.* este sistema contempla um conjunto de ferramentas de visualização interativas de consulta bastante simples sobre uma coleção de jogos e análise de tendências, permitindo aos jogadores e treinadores obter uma visão espaço-temporal de um jogo de beisebol, quantificando e avaliando as habilidades por cada jogador. O utilizador pode consultar as jogadas, visualizar espacialmente os jogos individuais ou em grupos, filtrar interativamente as jogadas, desenhar no campo as regiões pretendidas para consultar as jogadas e exportar os dados.

Os adeptos consomem informações relacionados ao desporto todos os dias, principalmente através das redes sociais que resume um todo de informações, tendo em conta o desempenho das equipas ou de jogadores em específico, as táticas das equipas, previsões da equipa vencedora que passa pela intuição humana ou preferências individuais, etc.

O que torna os desportos um domínio de interesse único é a variedade e a especificidade dos dados que gera. Segundo *Perin, Vuillemot, Stolper, Stasko, Wood e Carpendale* identificam três tipos de dados desportivos que mais são utilizados nas visualizações desportivas [62]:

- **Dados de pontuação:** dados que contêm resumos estatísticos de um determinado evento desportivo;
- **Dados de rastreamento:** dados sobre ações e trajetórias durante um jogo;
- **Metadados:** dados sobre o desporto ou os seus participantes.

Os dados de pontuação resumem estatisticamente um determinado jogo, podendo incluir estatísticas de um resumo assim como dados de eventos mais detalhados. Este tipo de dados normalmente são mais abstratos e seguem estritamente as regras que definem diferentes tipos de eventos e ocorrências durante um evento desportivos sendo que, envolvem uma ampla variedade de abordagens de visualização. São exemplos de visualização as tabelas e gráficos que classificam as características dos campeonatos (pontuações, metas, avaliações, golos, resultados, entre outros) que evoluem ao longo do tempo têm sido a forma padrão de representar e analisar campeonatos.



Os desportos coletivos normalmente são disputados durante dias ou meses, sendo os resultados desses jogos inseridos em tabelas e tomadas decisões e previsões (como por exemplo, quem vence um campeonato ou quem desce para uma liga abaixo) [62].

Tem-se como exemplo o gráfico representado na Figura 25 onde é possível visualizar os resultados das equipas da primeira liga da Inglaterra ao longo do tempo, jogo a jogo. É possível observar no gráfico detalhes sobre cada jogo que a equipa jogou, nomeadamente o resultado representado por uma cor (vermelho para derrota, amarelo para empate e azul para vitória) e são representadas umas linhas verticais conectam a equipa em foco ao seu adversário em cada dia de jogo onde vai ser possível visualizar as “lacunas” entre a equipa em foco e cada um dos seus adversários. Este tipo de gráficos possuem interação para selecionar uma equipa e explorar a mesma num contexto mais amplo de uma temporada.

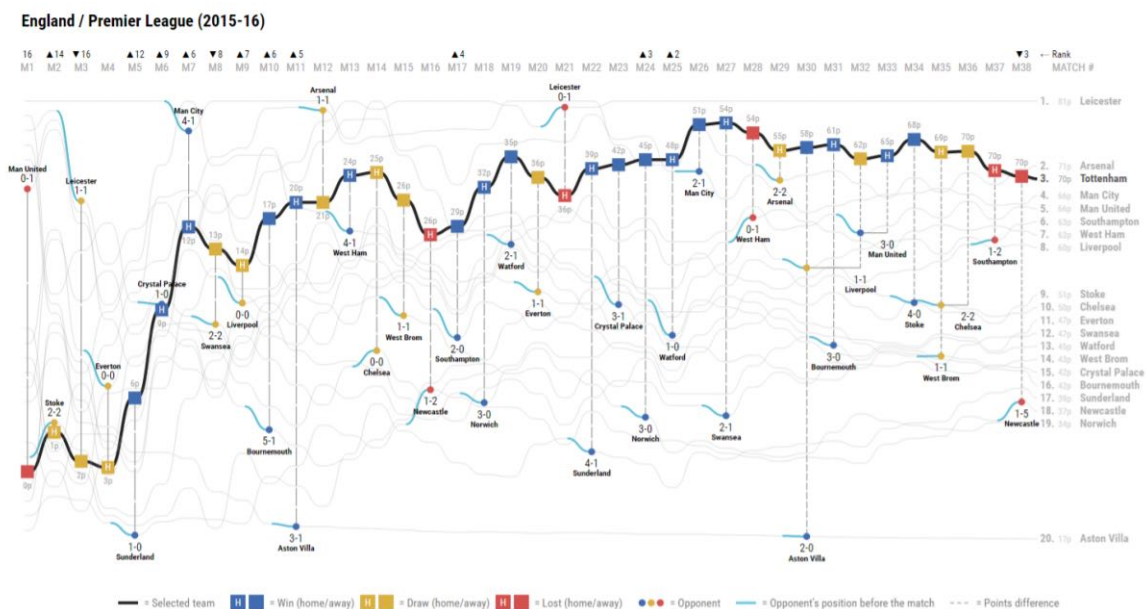


Figura 25: Resultados dos jogos da primeira liga inglesa [62]

Relativamente aos dados de rastreamento [62], estes contrastam com a abordagem dos dados de pontuação sendo que o volume, a variedade e a precisão crescem exponencialmente. A maioria dos dados de rastreamento recolhem, através de *Machine Learning*, informações espaço-temporais precisas sobre os jogadores e as equipas em tempo real durante o decorrer de um jogo. Para além disso existem inúmeros dados adicionais relacionados aos desportos que são utilizados no contexto de determinados jogos ou eventos, como a capacidade de um estádio, as características físicas de determinados jogadores, a velocidade de uma bola, as posições de um jogador em campo, entre outros.

Este tipo de dados podem trazer desafios na mineração e visualização de dados pois normalmente envolvem combinação de algoritmos e também podem ser recolhidos através das redes sociais.

*Morris* [77] pretendeu estudar e exemplificar se a distância do lance tem sido confiável e o quão tem sido decisiva para a percentagem de *field goals* (representa uma forma de pontuar no futebol americano).

Utilizou um gráfico por data que permitiu que o utilizador detectar facilmente a relação na distância do lance com a percentagem de golos no futebol americano de forma a revelar padrões e semelhanças que possam existir. É possível observar na Figura 26 que a percentagem de *field goals* da equipa *Kickers* no ano de 2015 foi boa a qualquer distância, sendo a percentagem de golos sempre superior a 70%.

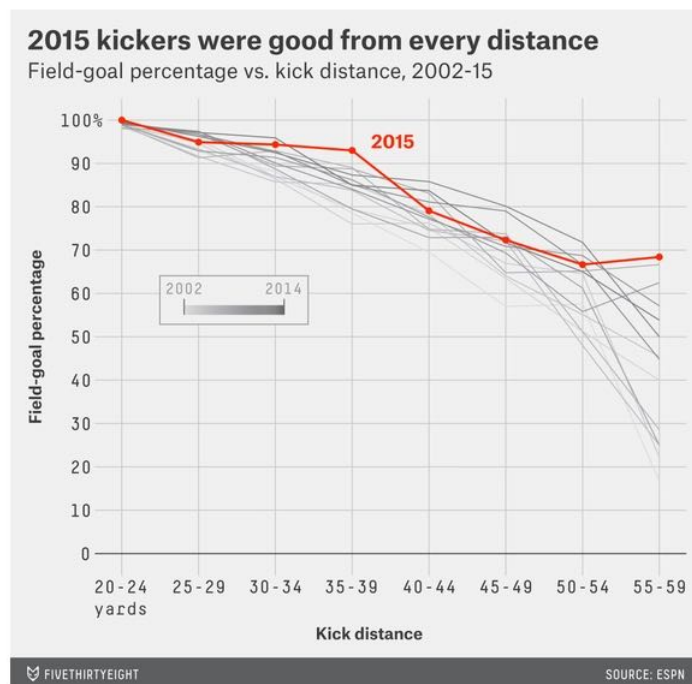


Figura 26: Relação na distância do lance com a percentagem de golos [77]

Os metadados são muito utilizados nos dados desportivos sendo que permitem fornecer detalhes de um determinado conteúdo e contexto (como as características físicas de um jogador ou as dimensões do campo). São dados sobre outros dados e facilitam o entendimento dos relacionamentos e a utilidade das informações dos dados. Podem utilizar redes sociais como o *Twitter* (recolha de informação dos *tweets*), programas de TV ou até mesmo para apostas desportivas que hoje em dia adquirem milhões de clientes. No entanto, como os metadados são muito diversificados, as visualizações dos mesmos variam muito.

Segundo *Perin et al.* [62], os metadados desportivos mais comuns a serem utilizadas referem-se às visualizações de desportos, jogadores e características dos jogos como pontos de referência, o desempenho dos jogadores e das equipas, as táticas das equipas, assim como a própria estrutura da competição e a visualização detalhada de pontos chave num determinado torneio. Este tipo de dados também pode ser utilizado para mostrar a nacionalidade dos atletas de determinadas ligas desportivas de forma a representar a evolução ao longo da história no que toca à diversidade de nacionalidades.

A utilização da idade dos jogadores, por exemplo, é um indicador poderoso no que toca às suas exibições e expectativas face ao jogo e à equipa, assim como o salário que normalmente se espera que o valor de mercado de um jogador ou de uma equipa seja proporcional ao desempenho ou potencial.



### 2.2.0.3 Tipos de visualizações

Apesar da visualização de dados no desporto estar constantemente a evoluir, passa por muitos desafios como por exemplo no que toca à qualidade dos dados disponíveis. Os diferentes tipos de visualização reconhecidos dividem-se em analítico (ou exploratório), narrativo (ou comunicativo) e interpretativo (expositor) [62].

Uma **visualização analítica ou exploratória** como exemplo na Figura 27, procura obter uma melhor percepção dos dados e das suas características, focando-se em ajudar os espectadores a formar as suas próprias interpretações e pretendo obter indicadores sobre determinados eventos. O conhecimento obtido a partir dos dados pode ser usado para futuras tomadas de decisão. É possível oferecer ao utilizador a capacidade de filtrar uma exibição para mostrar apenas certas categorias de interesse, alternar a exibição para diferentes parâmetros de dados e ter a possibilidade de o utilizador contribuir com os próprios dados para uma visualização mais válida.

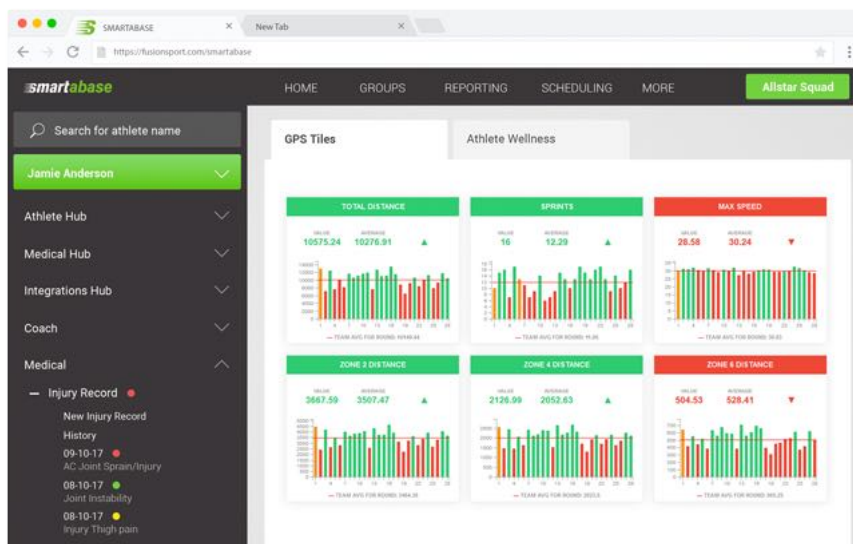


Figura 27: Visualização interativa no Website da empresa Smartbase [66]

Outro tipo de visualização é o **tipo narrativo ou comunicativo** presente na Figura 28 que gira em torno da apresentação de dados de uma maneira que seja mais facilmente compreendida e comunicada pelos utilizadores. Pretende enfatizar visualmente os valores de maior interesse recorrendo a propriedades de cores contrastantes pode oferecer dicas que estabelecem a hierarquia de importância. Propriedades de anotação, como legendas, resumos, animações e narrações podem acelerar a interpretação. O visualizador precisa de ser capaz de identificar os *insights* mais relevantes e interessantes. No passado, as estatísticas desempenharam um papel importante para a visualização de dados no desporto, mas o consumidor de informações de hoje é mais exigente e pretende visualizações eficazes que não podem ser transmitidos facilmente por estatísticas.

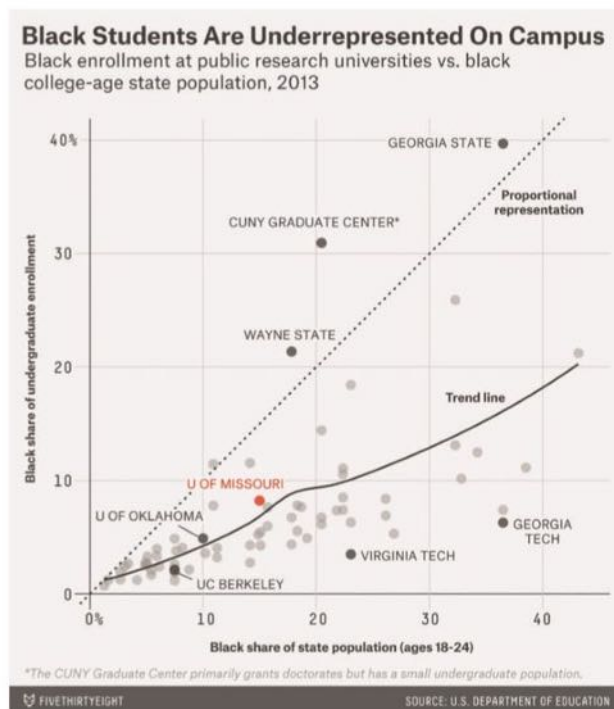


Figura 28: Exemplo de um gráfico comunicativo [43]

Nos tipos de **visualização de interpretação (expositor)** os espectadores têm que fazer o trabalho de interpretar e traduzir diversos significados do que estão a ver, contando com sua própria capacidade de perceber os recursos de uma visualização. Se existe um público-alvo muito específico, que sabe que tem conhecimento suficiente sobre o domínio e a análise que forneceu, pode não ser necessário enfatizar nenhum dos principais indicadores como faria com uma visualização narrativa ou comunicativa.

O utilizador vai formar a sua própria interpretação sobre o que vê apresentado visualmente. Na figura 29 é apresentada uma visualização de dados sobre os 100 atletas mais bem pagos do mundo. O utilizador pode retirar informações com o seu ponto de vista como por exemplo que a mulher continua a ser discriminada no que toca ao desporto ou por outro lado, existem muitas mais pessoas do sexo masculino a praticar desporto do que pessoas do sexo feminino.

O uso da visualização no desporto e em equipas pode ser dividido em três categorias: centrado no atleta (que se concentra em apoiar os atletas, bem como os treinadores), centrado no espectador (apoiar os espectadores ou o público) e centrado na apreciação coletiva. Os desportos coletivos são uma componente fundamental na sociedade de hoje, disponível através de quase todos os meios de comunicação. Exigem a colaboração e envolvimento de vários participantes em simultâneo. mas também se tornam naturalmente mais complexos ao longo do tempo.

O desporto é uma indústria multibilionária. O futuro está repleto de oportunidades no que toca à visualização de dados desportivos sendo que cada vez mais se torna um hábito o acesso imediato a informações. Existe uma quantidade crescente de pesquisas em torno de dados desportivos e tornar as



Figura 29: Exemplo de um gráfico interpretativo[43]

visualizações desses mesmos dados simples, acessíveis e identificáveis em formas digitais favorece não só os utilizadores, como os jogadores, treinadores, equipas e toda a parte envolvido em tópicos desportivos. Combinar esses tópicos de pesquisa e informação oferece muitas oportunidades para trabalhos futuros.

## 2.3 Trabalho Relacionado

Hoje em dia, como referido anteriormente, existem diversas ferramentas no mercado que retiram e estudam informações das redes sociais ou sobre. É exemplo disto o produto *Modern Research Lite*, criado pela *Sprinklr*, uma empresa de *software*. Este projeto pretende ser uma solução de auto-atendimento para grandes empresas descobrirem rapidamente *insights* essenciais da marca através de dados sociais [69]. Permite agrupar as publicações da marca por temas, tons e categorias sendo assim possível explorar percepções competitivas mais ricas, mostrando ao utilizador dados estatísticos sobre as opiniões da sua marca (ver Figura 30).

Neste conjunto de aplicações, existem ainda aplicações que extraem dados e possuem grande foco na rede social *Twitter*. Como é o caso do artigo dos autores *Oduwa Edo-Osagie*, *Beatriz De La Iglesia*, *Iain Lake* e *Obaghe Edeghere* [24] que pretendem estudar os dados da rede social *Twitter* para fins de saúde pública. Teve como objetivo principal examinar as tendências nos domínios de aplicação da saúde pública estudados ao longo dos anos. O gráfico representado na Figura 31 representa o número de artigos para um determinado ano e domínio de saúde pública. Isso mostra que parece realmente haver uma tendência na atividade para diferentes domínios da saúde pública. É possível perceber que alguns domínios decrescem e outros crescem constantemente. A detecção de eventos, a vigilância e a farmacovigilância tiveram aumentos constantes na atividade, liderando os outros domínios. No entanto, desde 2016, a pesquisa nesses três domínios diminuiu ligeiramente, mudando o foco para os outros

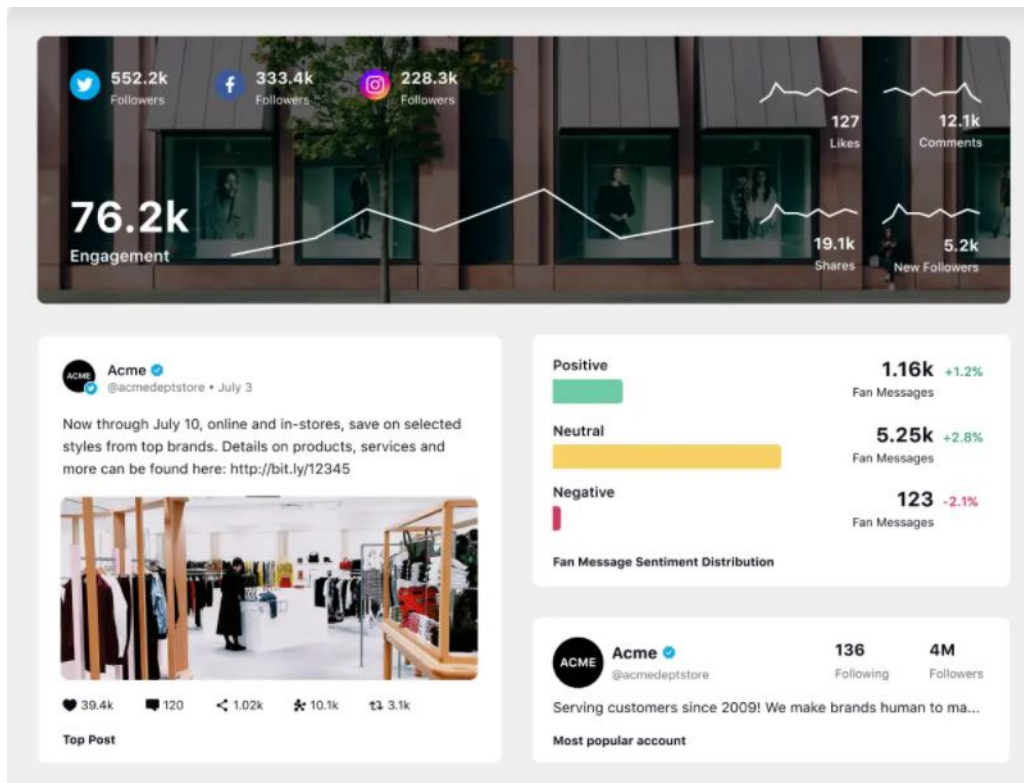


Figura 30: Exemplo da plataforma *Modern Research Lite* [69]

domínios.

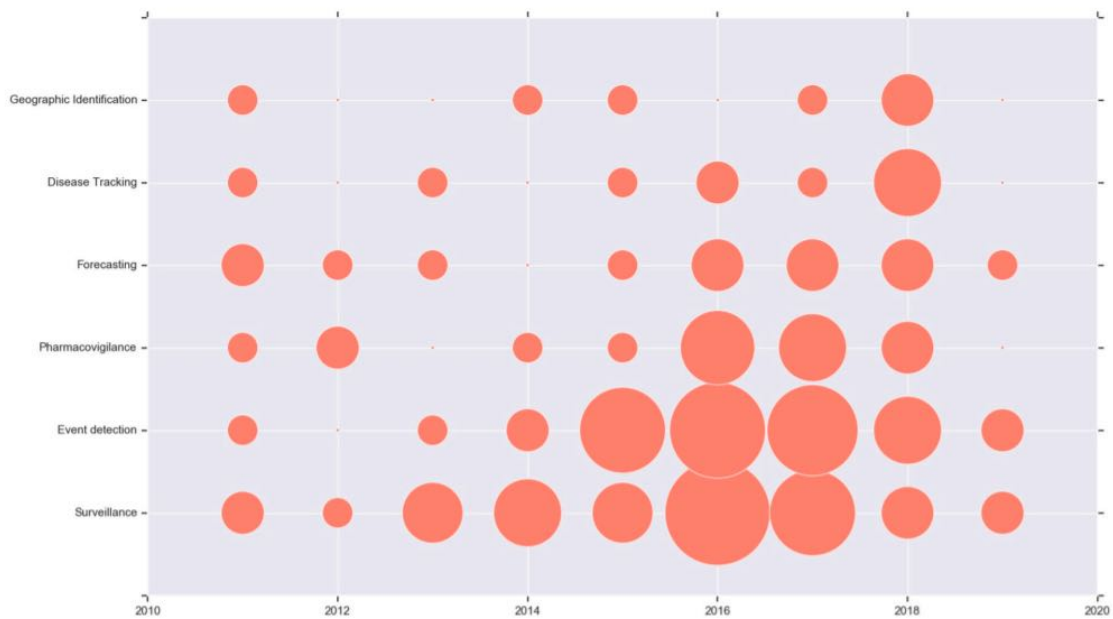


Figura 31: Gráfico das tendências de saúde pública [24]

Outro exemplo é o estudo do *Bongsug* [12] que pretendeu visualizar os *insights* da *hashtag #supply-chain* de forma a obter informações sobre a gestão da cadeia de suprimentos. É uma gestão de uma rede interligada de negócios envolvidos na provisão final de pacotes, produtos e serviços requeridos por clientes finais. O autor separou em cinco grupos, usando temas como *CSR*, Risco, Logística, Fabricação e *IT* e foi feita análise de sentimento dos *tweets* relacionados com estes temas de forma a entender a opinião dos utilizadores. Foi construído um gráfico como é possível observar na Figura 32 que representa a percentagem de quantidade de *tweets* por cada grupo de produtos/serviços. Através do gráfico, foi possível perceber que os *tweets* relacionados a riscos parecem ser mais negativos do que os outros grupos de *tweets*.

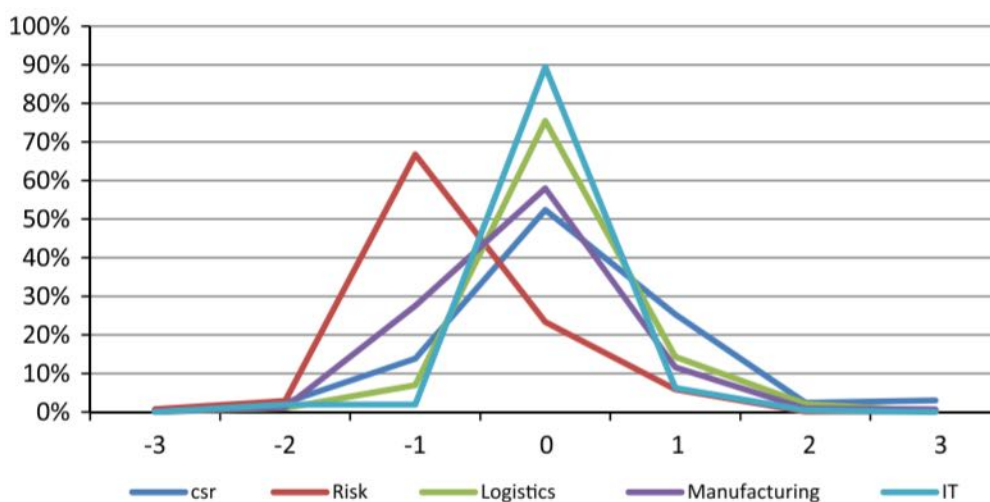


Figura 32: Análise de sentimento dos *tweets* agrupados [12]

Apesar destes exemplos, ainda existe um conjunto mais pequeno de extração de dados do *Twitter* que se foca na área do desporto. A análise de sentimento no *Twitter* tem sido um tema cada vez mais abordado e do qual, no tema desportivo já existem alguns exemplos e casos de uso relacionados com o mesmo. É aqui o principal foco desta dissertação, fazendo uma seleção dos trabalhos que:

- Usam a rede social *Twitter* para extrair informação sobre desporto;
- Utilizam análise de tópicos ou sentimentos;
- Permitem ou não interactividade dos utilizadores com os resultados.

É o caso do investigador *John Naujoks* [40] que utilizou *packages* de análise de sentimento e recolha de *tweets* para desta forma, obter o sentimento correspondente à equipa *Cleveland Browns*, durante cada dia do mês de setembro do ano 2019, como é possível observar na Figura 33.

Foi realizada a extração dos dados e posterior visualização num gráfico de barras. A partir dos dados recolhidos, o mesmo afirma que o sentimento dos fãs/adeptos estava muito positivo no início da

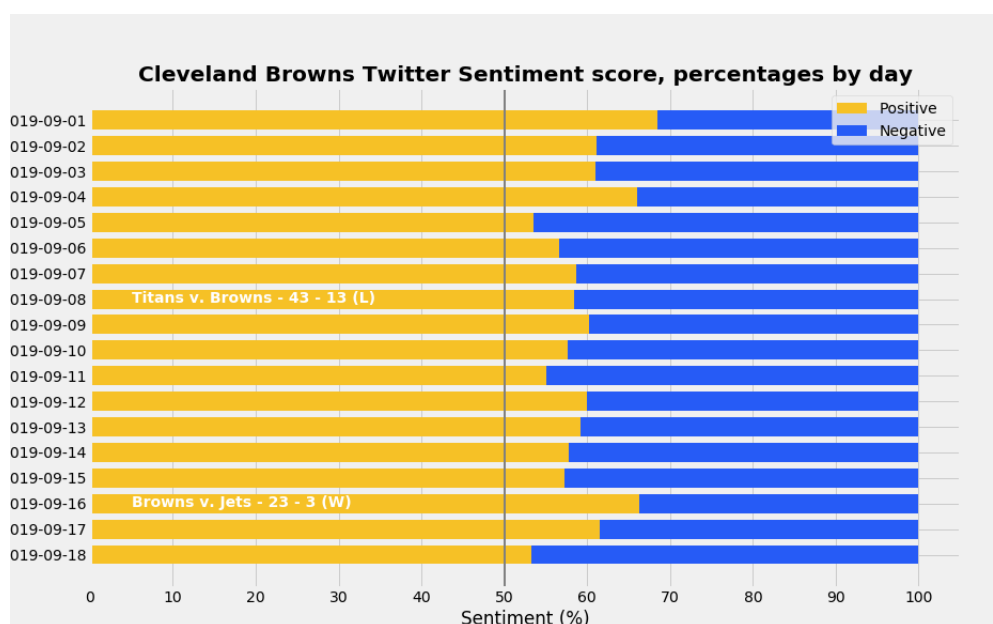


Figura 33: Exemplo do investigador *John Naujoks* [40]

temporada (com 68,3% de *tweets* positivos), tendo vindo a baixar devido à derrota no primeiro jogo da equipa. Este é um dos vários exemplos construídos que prevalece a importância do uso das redes sociais, paralelamente com o tema desporto.

Outra aplicação baseada numa sumarização de eventos para jogos desportivos utilizando os dados em *streaming* do *Twitter* (ver na Figura 34) com dados desportivos é a proposta realizada por *Yue Huang*, *Chao Shen* e *Tao Li* [35]. É necessário os utilizadores inserirem o horário de início e término do evento e um conjunto de palavras-chave, incluindo nomes de equipas, jogadores e treinadores. Esta aplicação permite:

- Detectar os momentos importantes durante o evento e mostrar os *tweets* relativos a esses momentos;
- Os utilizadores podem escolher reproduzir um evento anterior ou seguir um evento em tempo real;
- O sistema indica a quantidade de *tweets* de cada palavra chave adicionada inicialmente. O utilizador pode clicar na palavra-chave, e é aberta uma página detalhada com todos os *tweets* sobre essa mesma palavra, sendo representado o número de positivos e negativos (análise sentimental).

É certo que a análise das redes sociais nos dias de hoje é bastante importante para extrair informações relevantes. A análise de sentimento dessas informações também permite obter uma análise ao que os utilizadores estão a sentir, face a um determinado acontecimento. É exemplo disso, um sistema que durante o Campeonato do Mundo FIFA 2014, fornece acesso programático para ler e gravar dados do *Twitter* [90]. Para além disso, o presente projeto analisou as emoções ao nível da palavra, ou seja, foram



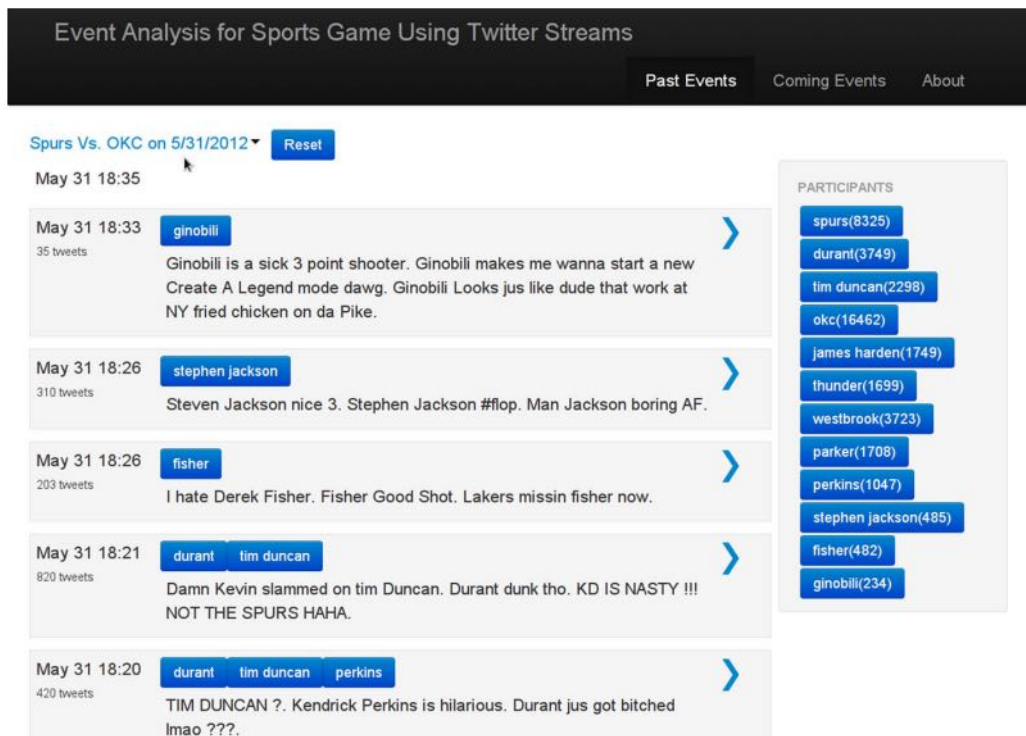


Figura 34: Sumarização de eventos para jogos desportivos [35]

analisadas as emoções e sentimentos de cada uma delas. Com isso, foram construídos vários gráficos com as emoções ao longo do tempo, como é visível um exemplo na Figura 35.

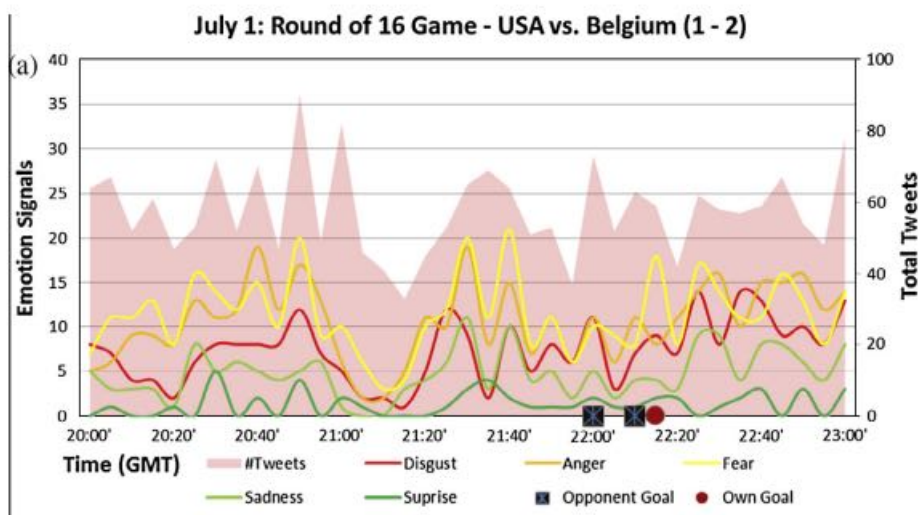


Figura 35: Estudo do Campeonato do Mundo FIFA 2014 [90]

Yang Yu e Xiao Yang [90], relacionaram as emoções com os acontecimentos do jogo, dizendo "Os tweets mostraram uma mistura de raiva e medo durante o jogo. Nenhuma das equipas marcou durante

o tempo regulamentar. No início do tempo de compensação, a antecipação atingiu o pico. No entanto, a expectativa e a alegria diminuíram, o medo e a raiva aumentaram pouco depois, quando a Bélgica marcou seu primeiro golo aos 92 minutos e subiu novamente quando a Bélgica marcou pela segunda vez. Aí os fãs americanos mostraram pouca expectativa e alegria e muita raiva e medo. No final do jogo, a alegria, voltou a subir, indicando que um jogo disputado, apesar de uma derrota, ainda pode ser uma fonte de diversão.”

## 2.4 Comparação do trabalho relacionado

Como vimos anteriormente, existem alguns casos hoje em dia, que se relacionam com o trabalho que se pretende realizar. Este estudo é feito de forma a entender que tipos de aplicações/trabalhos já existem acerca do tema dados em *streaming* de redes sociais, nomeadamente o *Twitter* com dados desportivos. Dessa forma, é bom compará-los e perceber que melhorias ou novas funcionalidades se pode implementar e conseqüentemente, trazer um novo produto mais enriquecido.

Depois de várias pesquisas sobre trabalhos relacionados, como mostrado acima, e apesar de não existirem muitos trabalhos nesta área do desporto, foi elaborada uma tabela comparativa que é possível observar na Tabela 1. Esta permite comparar as funcionalidades/objetivos de outros projetos com o presente nesta dissertação. Na tabela estão representadas as funcionalidades que se pretendem com este projeto e perceber se outros projetos correspondem ou se assemelham às mesmas.

	Os dados são representados em gráficos	Suporta streaming de dados	Possui um bom design	Permite uma análise do sentimento dos tweets	Permite ao utilizador construir os seus próprios gráficos	Permite ao utilizador a interatividade com os dados
Cleveland Browns Twitter Sentiment score, percentages by day	Verde	Vermelho	Vermelho	Verde	Vermelho	Vermelho
Event Analysis for Sports Game Using Twitter Streams	Vermelho	Verde	Vermelho	Verde	Vermelho	Verde
Study of Campeonato do Mundo FIFA 2014	Verde	Vermelho	Verde	Verde	Vermelho	Vermelho
Plataforma Online para visualização de tendências e opiniões de eventos desportivos com base em redes sociais	Verde	Verde	Verde	Verde	Verde	Verde

Tabela 1: Matriz de comparação dos trabalhos relacionados

Através da Tabela 1, é possível concluir que a proposta desta dissertação possui uma maior cobertura nas vertentes que se pretende alcançar comparado a outros, tais como:



- Esta plataforma fornece uma diversidade de gráficos de forma a mostrar diferentes tipos de dados;
- Os dados são recebidos em *streaming*, ou seja, o utilizador consegue visualizar dados em tempo real.
- Esta plataforma tem em conta as boas práticas de *User Interface* e *User Experience*, de forma a facultar ao utilizador um bom design e uma boa interatividade com os dados. Pretende dar ao utilizador uma experiência agradável e intuitiva.
- É exibida ao utilizador uma análise do sentimentos dos *tweets* sobre um determinado tema, permitindo assim ao mesmo perceber o que os utilizadores estão a sentir;
- O utilizador pode construir os seus próprios gráficos com informações do *Twitter* e exportá-los.

Com isto, e ainda que, existam algumas funcionalidades noutros projetos já existentes, esta dissertação tem vários pontos que permite dar novas soluções/respostas aos utilizadores. Em suma, esta subsecção serve para mostrar que as soluções existentes não se enquadram da melhor maneira nos objetivos pretendidos para este projeto e que a solução proposta pretende resolver essas limitações.

## 2.5 Representação de dados

A realização de uma análise através de representações visuais aumenta a capacidade de entendimento e absorção de informações. O avanço das tecnologias e ferramentas gráficas, impulsionaram a diversidade de modelos gráficos que o utilizador pode adotar como recurso.

A visualização de dados afeta positivamente o processo de tomada de decisão de uma organização com representações visuais interativas dos dados e segundo o *website Analytiks* fornece benefícios como [22]:

- **Descobrir relacionamentos:** Com a visualização de dados é possível identificar as correlações entre o relacionamento das variáveis, podendo tomar melhores decisões de negócio;
- **Visualizar tendências ao longo do tempo:** É uma das mais valias sendo que, as tendências ao longo do tempo dizem-nos o que se deve fazer e os pontos que se devem seguir;
- **Visualizar frequências:** A frequência está intimamente relacionada às tendências ao longo do tempo. Ao analisar valores e frequências nos dados expostos obtém-se uma melhor ideia de como os potenciais clientes podem agir e reagir a diferentes estratégias de marketing, por exemplo;
- **Examinar o mercado:** Na visualização de dados reúnem-se informações de diferentes mercados de modo fornecer indicadores sobre os públicos em que se deve concentrar uma maior atenção e também de quais se deve ficar longe. Obtém-se assim uma imagem mais clara das oportunidades dos mercados;

- **Reagir ao mercado:** A capacidade de obter informações de forma rápida e fácil com dados exibidos de forma clara permite que as empresas ajam e respondam às descobertas rapidamente e que evitem certos erros.

Consumir dados na forma de gráfico, ao contrário de ler uma tabela, permite que um visualizador processe certas características de vários pontos de dados simultaneamente para identificar as inclinações, planos, picos e depressões, bem como lacunas e cruzamentos entre as linhas. Os diferentes tipos de dados que lidamos podem ter uma grande influência nos tipos de gráficos que podemos ou não utilizar. A representação de dados envolve a codificação visual, usando várias propriedades visuais (marcas e atributos) com o propósito de explorar, analisar e comunicar dados.

As marcas representam dados como registros distintos ou agrupamentos discretos. São alguns exemplos delas os pontos, linhas, formas e formatos. Os atributos representam os valores associados a cada conjunto de dados, definindo variações na aparência visual dos gráficos, como a variação do comprimento, área, ângulo, volume, cores, padrões e símbolos para representar/distinguir certos tipos de valores.

Os diferentes tipos de gráficos oferecem diferentes formas de representar dados, cada um com diferentes combinações de marcas e atributos. Resumidamente, as marcas e os atributos são os ingredientes e os tipos de gráficos as receitas. Sendo assim, definiu-se as cinco principais famílias de gráficos:

- **Catagóricos** - Compara as categorias e distribuições de valores quantitativos;
- **Hierárquicos** - Revela os relacionamentos e hierarquias de uma certa parte para um todo;
- **Relacionais** - Explora as correlações e conexões;
- **Temporais** - Traça tendências e intervalos ao longo do tempo;
- **Espaciais** - Mapea padrões espaciais.

Todos os tipos de gráficos oferecem determinados valores para diferentes situações, apenas é necessário usar um critério para os selecionar em certas circunstâncias específicas. Ter bem presente a diversidade de tipos de gráficos que existem e que tipos de dados os mesmos conseguem representar, é uma tarefa necessária para uma prévia análise do tipo de visualizações que conseguiremos visualizar conforme o tipo de dados que temos em nossa posse, como por exemplo, dados catagóricos, booleanos, temporais, qualilativos e quantitativos.

Desta forma e visto que existe um leque enorme de tipos de gráficos, selecionou-se e estudou-se os 12 gráficos mais relevantes presentes na Figura 36, na ótica deste projeto, que contemplam os tipos de dados mais comuns e utilizados e que visam satisfazer todas as necessidades visuais que se pretende.

Foram elaboradas tabelas com informações sobre cada tipo de visualização representadas nos Anexos A, B e C, nomeadamente: uma descrição que serve como explicação para que tipos de dados são utilizados num determinado gráfico; dicas de utilizações (informações sobre a interatividade, anotações,



Figura 36: Tipos de visualização

cores e composição que beneficiam e satisfazem as condições de certos tipos de visualização) e também alternativas que visam atingir o mesmo objetivo e necessidades de um certo tipo de gráfico.

Começando pela primeira tabela construída visível no Anexo A selecionou-se gráficos majoritariamente utilizados na representação de informações:

- **Gráfico de barras:** Representa valores quantitativos de cada item, como itens de categorias e itens temporais. As barras que o mesmo constitui (horizontais ou verticais) exibem comparações de forma a preservar uma distinção clara entre cada categoria;
- **Histograma:** Exibe a frequência e distribuição de dados quantitativos em valores agrupados num período de tempo ou para certos itens de dados. A variação do tamanho da linha ou da área da forma representa a frequência dos dados;
- **Nuvem de Palavras:** É possível visualizar a frequência de palavras de dados textuais. O tamanho de cada palavra é dimensionado de acordo com a frequência do seu uso. A utilização de cores permite realçar ainda mais as frequências existentes;
- **Gráfico circular:** Representa as proporções de quantidades para diferentes categorias. Forma um todo, ou seja, o total de todos os valores do setor deve ser de 100 por cento. Divide-se em setores para cada categoria com um determinado ângulo, cada um representado por diferentes cores que serve para classificar cada categoria e maximizar a diferença visual.

Em seguida está representada no Anexo B, uma nova tabela elaborada com outros quatro tipos de gráficos que pretendem satisfazer as necessidades deste projeto e que visam representar o tipo de dados facultados:

- **Gráfico de Dispersão:** Exibem a relação entre duas variáveis quantitativas para diferentes itens de categoria. A cor nos pontos é usada para distinguir as dimensões categóricas. O objetivo final passa por avaliar o impacto que uma variável tem sobre a outra, observando assim as relações existentes. Podem surgir alguns típicos padrões, como a correlação negativa (certos valores aumentam enquanto outros diminuem), correlação positiva (ambos os valores aumentam) e nulos onde não existe correlação.
- **Gráfico de Linha:** É possível visualizar como os valores quantitativos mudam ao longo do tempo para diferentes itens de categoria. Várias categorias podem ser exibidas na mesma visualização cada uma representada por uma linha, normalmente com atributos de cores associados para revelar tendências ou visualizar relações que existam;
- **Gráfico de Fluxo:** Mostra como os valores quantitativos mudam ao longo de um período de tempo para vários itens de categoria. A área ocupada pelas várias camadas é preenchida com cores de modo a representar uma escala de valor quantitativa ou para associar a certas categorias;
- **Gráfico de Área:** Exibe os valores quantitativos ao longo do tempo para itens de categoria. São equivalentes aos gráficos de linhas mas a área é totalmente preenchida por uma cor ou textura. São bastante úteis para revelar tendências e relações, sendo que é possível visualizar várias camadas de áreas de diferentes categorias.

Finalmente criou-se uma última tabela que se encontra no Anexo C, com outro tipos de gráficos que se pretende utilizar para este projeto. Para além de gráficos que representem dados numéricos ou categóricos, decidiu-se adicionar um outro que faça uso de imagens, marcas ou símbolos :

- **Gráfico de Radar:** Transpõe valores de variáveis quantitativas para formar padrões gerais, servindo assim para comparar dados quantitativos com múltiplas variáveis. Este tipo de gráfico funciona melhor quando os pares vizinhos possuem algum algum valor comparável significativo;
- **Pictograma:** Exibem valores quantitativos para diferentes itens de categoria. A base do pictograma é representada pela repetição de marcas, símbolos ou imagens para representar uma contagem quantitativa associada. Cada marca pode corresponder a uma ou mais unidades quantitativas.
- **Caixa de Bigodes:** Representa a distribuição de de valores quantitativos para diferentes categorias, através dos quartis. As linhas do gráfico indica os quartis inferior e superior e os *outliers* são apresentados por pontos individuais fora das caixas.
- **Gráfico de Bolhas:** Exibe a relação entre três variáveis quantitativas para diferentes itens de categoria, representando assim múltiplas variáveis. Usa normalmente círculos como marca de forma para cada categoria ao longo de cada eixo quantitativo com uma variação no tamanho dos círculos que irá representar a terceira medida quantitativa.

Em resumo, este subcapítulo passou por um estudo de alguns tipos de visualização que visam satisfazer as necessidades deste projeto. Para além de perceber que tipos de dados conseguimos representar em cada um deles, achou-se também necessário recolher informação sobre alguns aspetos essenciais para o design e interatividade dos mesmos.

Decidir o tipo de visualização que pretendemos utilizar é um passo bastante importante e que não deve ser escolhido aleatoriamente. Devemos considerar o tipo específico de dados com o qual estamos a trabalhar, bem como definir o público alvo que se pretende atingir e que mensagem desejamos transmitir. Os dados e o visual precisam de trabalhar em conjunto para contar uma história.

A visualização de dados é realmente importante - desde professores a entender os resultados dos testes dos alunos até às empresas que utilizam dados para obter informações importantes. Hoje em dia, mais e mais empresas recolhem grandes quantidades de dados de forma rápida e eficaz, o que também exige uma maneira de classificar, compreender e explicar esses dados de uma forma que faça sentido para as partes interessadas [87].

Os gráficos revelam informações como as distribuições incomuns de dados, certos padrões locais, lacunas que possam existir, valores ausentes e discrepantes. Os gráficos levantam questões que estimulam a pesquisa e sugerem ideias. Tanto os gráficos estáticos, dinâmicos mas especialmente os interativos têm muito a acrescentar.

Não importa qual o negócio ou a empresa, a visualização de dados pode ajudar, fornecendo dados da maneira mais eficiente possível. A visualização de dados pega em dados brutos, modela-os e entrega os dados para que as conclusões possam ser alcançadas. Em análises avançadas, os cientistas de dados estão a criar algoritmos de aprendizagem para compilar melhor os dados essenciais em visualizações que são mais fáceis de entender e interpretar [22].

## Ferramentas e Metodologias

### 3.1 Ferramentas

Para este projeto em concreto, e orientando aos resultados que se pretendem obter, esta secção é dividida em: tecnologias de *front-end*, ferramentas de visualização de dados, tecnologias de *back-end* e base de dados. O principal objetivo desta abordagem é analisar as tecnologias existentes em cada uma das áreas mencionadas acima, para posteriormente fazer uma escolha das melhores tecnologias para desenvolver este projeto.

#### 3.1.1 Tecnologias de Front-end

Nos dias de hoje existem muitas tecnologias de *front-end*. Não sendo possível estudar todas, decidiu-se selecionar as 3 mais utilizadas hoje em dia - *Reactjs*, *Vuejs* e *Angular*. Como é visível na Figura 37, a biblioteca *Reactjs* é sem dúvida a que possui um maior número de *downloads* nos últimos 5 anos no *GitHub*, assim como, uma maior percentagem de questões feitas sobre a mesma na plataforma *StackOverflow*, na figura 38.

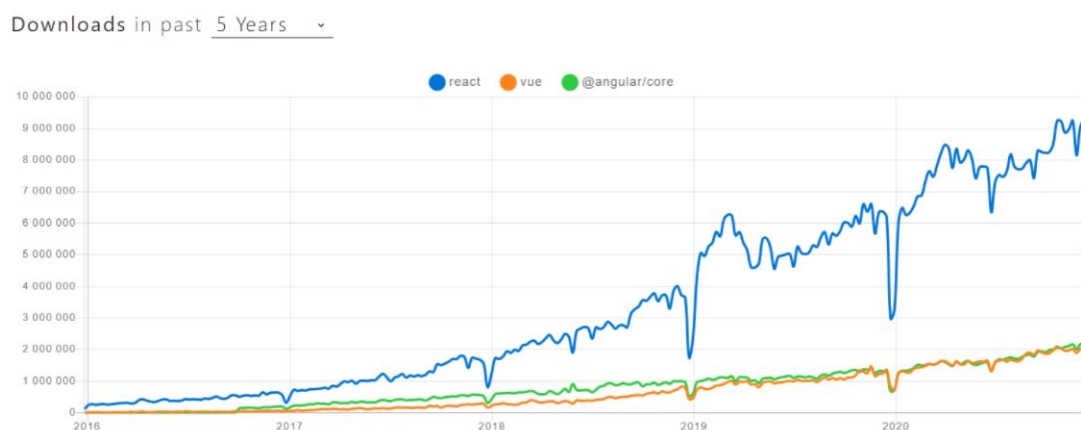


Figura 37: *GitHub*: Número de *downloads*

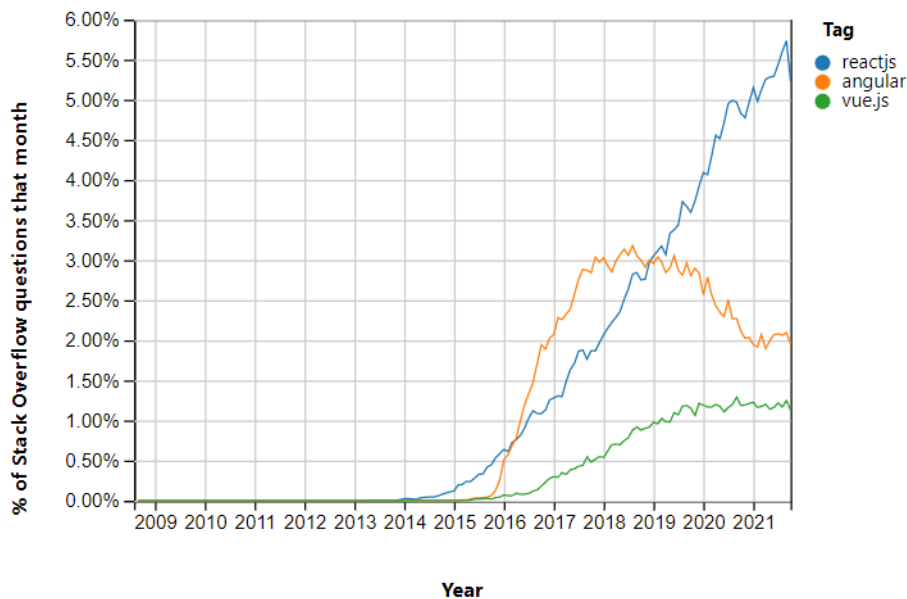


Figura 38: *StackOverflow*: Percentagem de questões feitas [28]

Sendo estas as tecnologias mais utilizadas para *front-end*, foi necessário criar uma tabela comparativa de forma a auxiliar na decisão de escolha da tecnologia mais adequada para o projeto (ver na Tabela 2).

Esta comparação foi feita tendo em conta os seguintes campos: popularidade, curva de aprendizagem, compatibilidade com versões anteriores, documentação e a performance da tecnologia.

	Reactjs	Vuejs	Angular
Popularidade	Boa	Média	Média
Curva de aprendizagem	Boa	Boa	Má
Compatibilidade com versões anteriores	Boa	Boa	Má
Documentação	Boa	Boa	Boa
Performance (Criar uma tabela com 10.000 linhas)	2866 milissegundos	1447 milissegundos	6361 milissegundos

Tabela 2: Tabela de comparação das ferramentas *front-end*

Os dados presentes na tabela foram escritos tendo em conta *Elar Saks* [63] e *Yoni Flenner* [78]. Como é possível observar através da tabela, a ferramenta *Reactjs* foi a escolhida. Apesar de a performance estar abaixo do *Vuejs*, esta ferramenta apresenta grandes qualidades nos outros campos. Para além

disso, sendo este trabalho realizado em contexto empresarial, a mesma tem uma preferência por esta tecnologia sendo esta a que é usada internamente, sendo assim um requisito a utilização desta tecnologia.

### 3.1.2 Ferramentas de Visualização de Dados

Atualmente, a visualização de dados, como dito anteriormente, tem sido uma prática cada vez mais utilizada nas tomadas de decisão. Foram criadas **ferramentas de visualização** que acompanham esse crescimento com atualizações constantes, de modo a obter um alto nível de boas práticas de visualização de dados.

O **objetivo** chave deste projeto está centrado na qualidade da visualização de dados e para isso, o estudo de ferramentas nesta área, é algo bastante valioso. É imprescindível uma pesquisa sobre as ferramentas que existem hoje em dia, de modo a perceber quais as mais relevantes para este projeto, tendo em conta a evolução das mesmas ao longo dos anos e certas características necessárias para o tipo de dados que pretendemos visualizar. Foi elaborada a **Tabela 3** para resumir todo o estudo feito de cada uma das ferramentas usadas para visualização de dados, que inclui as principais características/-funcionalidades que uma ferramenta de visualização deve ter/suportar.

Para este efeito, reuniram-se 10 ferramentas/bibliotecas mais comuns e utilizadas para a representação visual de dados: **Cubejs, Chartjs, AmCharts, High Charts, Chartist, FusionCharts, ApexCharts, Nivo, Recharts e Grafana**.

Para facilitar o processo de estudo e escolha das ferramentas, foram escolhidos 14 campos todos eles necessários na perspectiva do projeto, não só a nível gráfico, como também a nível de implementação, nomeadamente: não serem necessários conhecimentos informáticos para a implementação da ferramenta, apresentar uma boa **UI/UX** e uma documentação bastante completa, permitir interatividade nos gráficos, possuir a capacidade de receber dados em **streaming**, apresentar facilidade de integração, ter uma boa capacidade de armazenamento, disponibilizar uma **API**, suportar uma variedade de gráficos e conectores, permitir a customização de gráficos. Para além destes, complementou-se este estudo comparativo com os valores dos tamanhos das respetivas bibliotecas de integração para **Nodejs**, tipo de licença e por último, o número de **downloads** do **NPM (Nodejs)** de cada ferramenta, sendo que as ferramentas/bibliotecas que têm vindo a crescer a nível de número de (**downloads**) estão representadas por um sinal mais (+) e aquelas que se têm mantido iguais/constantes ao longo do tempo estão representadas por um sinal de igual (=) - de salientar que algumas informações que não constavam na documentação de certas bibliotecas, foram assumidas como limitações para as mesmas.

Ainda no contexto dos campos estudados, foram impostos alguns limites no suporte de gráficos e conectores, sendo que, se definiu um número mínimo de gráficos e conectores que a ferramenta/biblioteca deverá suportar. Foram escolhidos 7 tipos de gráficos (barras, espacial, linhas, circular, dispersão, fluxo e histograma) e 6 conectores de dados (**MySQL, Oracle, PostgreSQL, JSON, MariaDB e MongoDB**).

A pesquisa realizada para cada uma das ferramentas e a posterior análise dos dados da tabela permitiram perceber quais as ferramentas/bibliotecas que preenchem o maior número de requisitos



	Cubejs	Chartjs	AmCharts	High Charts	Chartist	FusionCharts	ApexCharts	Nivo	Recharts	Grafiáa
Boa UI/UX sem customização	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓
Documentação bastante completa	✓	✓	✓	✓	✗	✓	✓	✓	✗	✓
Permite interatividade	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓
Capacidade de lidar com streaming	✓	✓	✓	✓	✓	✓	✓	✗	✓ <sup>(2)</sup>	✓
Facilidade de integração em soluções de terceiros	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Data Querying ilimitado	✓ <sup>(1)</sup>	✓ <sup>(1)</sup>	✗	✓ <sup>(1)</sup>	✗	✓ <sup>(1)</sup>	✓ <sup>(1)</sup>	✓ <sup>(1)</sup>	✓ <sup>(1)</sup>	✓
Disponibilização de uma API	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Suporta o total de gráficos necessários <sup>(3)</sup>	✗	✗	✓	✓	✗	✓	✓	✓	✓	✓
É possível alterar a customização dos gráficos	✗ <sup>(5)</sup>	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tamanho dos pacotes de integração no Nodejs	100 kB	1.41 MB	22.6 MB	42.7 MB	535 kB	4.53 MB	3.96 MB	✗	643 MB	38.3 MB
Licença	MIT/Apache 2.0	MIT	MIT	Commercial	MIT	Commercial	MIT	MIT	MIT	Apache 2.0
Número de downloads <sup>(4)</sup>	96 (±)	1,161,259 (+)	1,992 (+)	1,747,871 (±)	279,570 (±)	11,168 (+)	128,243 (+)	3,376 (+)	185,115 (+)	22 (±)

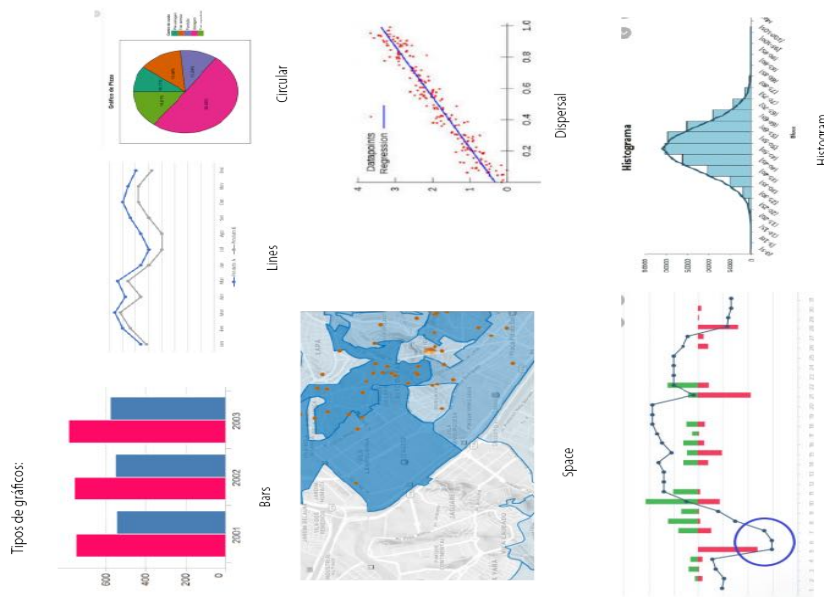


Tabela 3: Tabela de estudo das ferramentas de visualização [15, 6, 19, 16, 4, 33, 32]

(1) - Não foi encontrada nenhuma documentação sobre limites de armazenamento. Assume-se que a biblioteca não tem nenhuma limitação.  
 (2) - Não foi encontrada nenhuma documentação oficial sobre suporte de streaming de dados. Presumiu-se que a biblioteca não lida com dados em streaming.  
 (3) - Selecionou-se 7 tipos de gráficos que as bibliotecas devem suportar. Um exemplo de cada tipo de gráfico selecionado encontra-se à direita da tabela.  
 (4) - A métrica utilizada para esta linha foram os downloads do npm, das bibliotecas de integração para o Nodejs no dia 8 de Outubro de 2020.  
 (5) - Não foi encontrada nenhuma documentação oficial sobre a customização de gráficos. Presumiu-se que a biblioteca tem algumas limitações.

(campos). No seguimento desta análise, selecionaram-se três ferramentas que se encontram coloridas a verde na **Tabela 3: Chartjs, High Charts e ApexCharts**.

Foram analisadas e comparadas as características das várias bibliotecas para construção de gráficos, de modo a perceber-se quais seriam as mais completas para este projeto. O foco principal passou por perceber que tipo de licença apresentavam e, paralelamente, o número de **downloads**, verificando-se assim se os valores das mesmas têm aumentado ou mantido constante ao longo do tempo. Assim, selecionaram-se três bibliotecas que preenchem o maior número de campos: **Chartjs, High Charts e ApexCharts**.

Entre estas bibliotecas, surgem algumas diferenças bastante relevantes: A biblioteca **Chartjs** é a que apresenta um maior número de desvantagens, uma vez que não suporta o total de gráficos pretendidos. Em contrapartida, o seu tipo de licença é o **MIT** e como se pode concluir pelos dados apresentados na tabela, é uma biblioteca que tem vindo a crescer gradualmente e apresenta um dos maiores números de **downloads**. A biblioteca **High Charts**, de todas as bibliotecas, é a que apresenta um maior número de **downloads** e que se tem mantido igual ao longo do tempo. Apresenta, no entanto, uma grande desvantagem, que se prende com o facto de ter uma licença do tipo **Commercial**. Por outro lado, temos a biblioteca **ApexCharts** com o tipo de licença **MIT**, que é o mais procurado, e para além disso - apesar de os números de **downloads** não serem tão elevados como os das outras bibliotecas, tem vindo a crescer significativamente ao longo do tempo. Por estas razões, a maior indecisão foca-se principalmente entre as bibliotecas **High Charts e Apex Charts**, sendo que o único contra é o tipo de licença.

O objetivo pretendido com a **Tabela 3** passou por fazer a seleção das 5 bibliotecas e o próximo passo passará por analisar com mais detalhe as mesmas, tentando-se perceber qual se adequa melhor ao propósito final. Elaborou-se a Tabela 4, mas desta vez, com as vantagens e desvantagens de cada uma das bibliotecas selecionadas da tabela anterior para ser tomada uma decisão mais consistente, percebendo assim quais faria mais sentido testar com alguns protótipos de **streaming** de dados e, consequentemente, com a visualização gráfica dos mesmos.

Relativamente às bibliotecas, conclui-se que a biblioteca **Chartjs** é a menos completa. Para além de contemplar gráficos muito básicos, os mesmos não são interativos o que seria um requisito essencial na realização do projeto. Apesar de ter algumas vantagens, não apresenta alguns dos requisitos necessários para a concretização dos objetivos a que se propõe este projeto, sendo que existem outras que o fazem e de uma forma mais eficaz. Fica-se assim, dividido entre as bibliotecas **High Charts e Apex Charts**. A biblioteca **High Charts** é claramente a que apresenta um maior número de vantagens, nomeadamente ao nível de funcionalidades nos gráficos - sendo que por outro lado, esta apresenta uma customização mais restrita e algumas limitações no desenvolvimento de detalhe nos gráficos. O facto de ter uma licença do tipo **commercial** torna-a uma grande desvantagem para este projeto. Em relação à biblioteca **Apex Charts**, é também uma biblioteca com grandes qualidades de visualização de dados. Apresenta bons tipos de gráficos, com recurso a interatividade, uma boa documentação e a possibilidade de alterar a customização de gráficos. Tem, no entanto, algumas desvantagens expectáveis, nomeadamente o facto de uma grande quantidade de dados, tornar a sua renderização mais lenta e o facto de possuir

	Chartjs	High Charts	Apex Charts
Vantagens	<ul style="list-style-type: none"> <li>- Os gráficos são totalmente responsivos.</li> <li>- A documentação do chart.js é bem organizada e fornece informações detalhadas sobre o uso de cada recurso.</li> <li>- Muitos plugins disponíveis via NPM e podemos facilmente integrar com React, AngularJS, etc...</li> <li>- Fácil de implementar</li> <li>- Altamente escalonável e responsivo durante a renderização de dados.</li> <li>- Licença MIT</li> </ul>	<ul style="list-style-type: none"> <li>- Funciona em todos os navegadores de telemóvel e desktop</li> <li>- Sobre qualquer uma das licenças, gratuitas ou não, existe permissão para fazer download do código-fonte e fazer as próprias edições</li> <li>- Gráficos responsivos</li> <li>- Variedade de gráficos</li> <li>- Curva de aprendizagem baixa e poderosa</li> <li>- Documentação fácil de analisar</li> <li>- Gráfico de exportação integrado como está, para o arquivo de imagem</li> <li>- Esquema de cores fáceis de personalizar</li> <li>- Sintaxe de configuração simples</li> <li>- Gráficos invertidos</li> <li>- Gratuitos para usos não comerciais</li> <li>- Não é necessária a utilização de plugins</li> <li>- Os utilizadores podem configurar o sistema facilmente</li> <li>- Fácil integração de APIS e JSON.</li> <li>- Pequenas dependências</li> <li>- Lida com grandes quantidades de dados</li> <li>- Boa documentação</li> <li>- Gráficos com boa aparência</li> <li>- Zoom de dados</li> </ul>	<ul style="list-style-type: none"> <li>- Código aberto com licença MIT</li> <li>- Renderização de gráficos em SVG</li> <li>- Oferece recursos de zoom</li> <li>- Gráficos interativos</li> <li>- Boa documentação</li> <li>- Gráficos responsivos</li> <li>- É dinâmico - permite que ao carregar nos dados selecionados, crie outros gráficos com base nessas seleções.</li> <li>- Alta performance</li> <li>- Permite adicionar anotações/legendas aos gráficos</li> <li>- É possível exportar painéis completos</li> <li>- É possível criar o próprio tema de cores</li> <li>- Implementação simples</li> <li>- Zoom nos gráficos</li> </ul>
Desvantagens	<ul style="list-style-type: none"> <li>- Recursos limitados</li> <li>- Baseado no ecrã Canvas com base em bitmap e compartilha os mesmos problemas dos formatos não vetoriais</li> <li>- Gráficos não interativos</li> <li>- Gráficos muito básicos (apenas 6 tipos)</li> </ul>	<ul style="list-style-type: none"> <li>- Relativamente complexo</li> <li>- A customização é mais restrita (quando precisamos de algo mais do que tamanho e cor)</li> <li>- Necessários mais ajustes para torná-lo responsivo</li> <li>- As funções de detalhe são mais complicadas de desenvolver</li> <li>- Necessidade de melhorar os dados e a configuração do display</li> <li>- Licença Commercial.</li> </ul>	<ul style="list-style-type: none"> <li>- Renderização lenta quando suporta muitas animações</li> <li>- Se usamos um conjunto de dados com mais do que mil valores, também apresentará algum atraso.</li> <li>- Possui sete dependências, o que pode se tornar um problema no futuro, se algumas delas falhar (todas essas dependências dizem respeito ao svg).</li> </ul>

Tabela 4: Vantagens e Desvantagens das bibliotecas selecionadas

dependências. Para além disso, é uma biblioteca com o tipo de licença **MIT**, apresentando assim código aberto. Após esta análise comparativa ao nível dos campos especificados na tabela e tendo em conta o tipo de licença (**MIT**) sendo este, um requisito com um peso elevado na tomada de decisão, conclui-se que a biblioteca com mais vantagens - e a que preenche todos os requisitos necessários no âmbito deste projecto - é a biblioteca **Apex Charts**, recaindo sobre esta a escolha para este projeto.

### 3.1.3 Tecnologias de back-end

O *back-end* serve para fazer a ponte entre os dados que recebe e as bibliotecas que mostram os dados em visualizações gráficas. Por essa mesma razão decidiu-se fazer um estudo sobre 4 tipos de tecnologias mais utilizadas para esta função - *Nodejs*, *Django*, *Laravel* e *Ruby*, elaborando assim a Tabela 5. Foram estudados vários campos para cada uma delas como o tipo de licença, a velocidade de desenvolvimento, a escalabilidade, se possui um bom suporte de dados em *streaming*, a curva de aprendizagem, a sintaxe, a sua popularidade atualmente, a documentação, a performance, a utilização para aplicações *web* assim

como o número de estrelas que possui no *GitHub*.

### Comparação das ferramentas

	Nodejs	Django	Laravel	Ruby
Tipo	Software	Web Framework	PHP Framework	Linguagem de programação
Linguagem	JavaScript	Python	PHP	Ruby
Licença	Open Source	Open Source	Open Source	Open Source
Velocidade de desenvolvimento	Medíocre	Bom	Medíocre	Bom
Escalabilidade	Bom	Bom	Medíocre	Medíocre
Dados em Streaming	Bom	Medíocre	Medíocre	Medíocre
Curva de aprendizagem	Mais complexo	Menos complexo	Mais complexo	Menos complexo
Sintaxe	Mais apelativa	Menos apelativa	Mais apelativa	Mais apelativa
Popularidade	Bom	Medíocre	Bom	Medíocre
Documentação	Bom	Bom	Bom	Mau
Performance	Bom	Medíocre	Bom	Medíocre
Utilizado para aplicações web	Bom	Medíocre	Bom	Bom
Github (estrelas)	55 330	37 544	63 600	47 500

Tabela 5: Estudo das ferramentas de back-end [53, 52, 23, 61, 58, 44]

Analisando o estudo feito na tabela anterior, existem três ferramentas que se destacam - *Nodejs*, *Django* e *Laravel*, sendo que a linguagem *Ruby* apresenta variadas características não tão relevantes para este projeto. A *framework Django* realmente revela bastantes vantagens no que toca a conter uma curva de aprendizagem menos complexa, assim como possuir uma boa documentação, escalabilidade e velocidade de desenvolvimento. Apesar disto, não apresenta uma sintaxe muito apelativa e uma grande popularidade atualmente. Seguem-se as ferramentas *Nodejs* e *Laravel*. Comparando as mesmas percebe-se que a ferramenta *Nodejs* se destaca pela escalabilidade e por possuir um bom suporte de dados em *streaming* enquanto que a *framework Laravel* se destaca pela popularidade na plataforma *GitHub* apesar de o *Nodejs* se posicionar em segundo lugar neste campo.

A linguagem de programação *JavaScript* é a mais popular no mundo, sendo uma das mais utilizadas para construir páginas *web* [13]. Segundo a *RedMonk*, uma empresa de analistas, a linguagem *JavaScript* continua a ser a número um, com a linguagem *Python* e *Java* em segundo lugar. É possível observar na Figura 39 o *ranking* das linguagens de programação medido através da popularidade no *GitHub*, assim como na plataforma *Stack Overflow* no ano de 2020 [79]. Através das trends do *NPM* durante o ano de 2020 até ao mês inicial de 2021, consegue-se claramente perceber o destaque que o *Nodejs* teve no número de *downloads* nos últimos 6 meses.

Depois de todo o estudo feito a ferramenta escolhida para *back-end* foi o *software Nodejs*. Para além de utilizar a linguagem *JavaScript* e da importância da mesma nos dias de hoje como referido anteriormente, é uma ferramenta que apresenta todas as características necessárias para este projeto. É uma ferramenta de código aberto, possui uma sintaxe apelativa e tem variadas vantagens quanto à escalabilidade, à popularidade, à performance e no que toca à receção de dados em *streaming*.

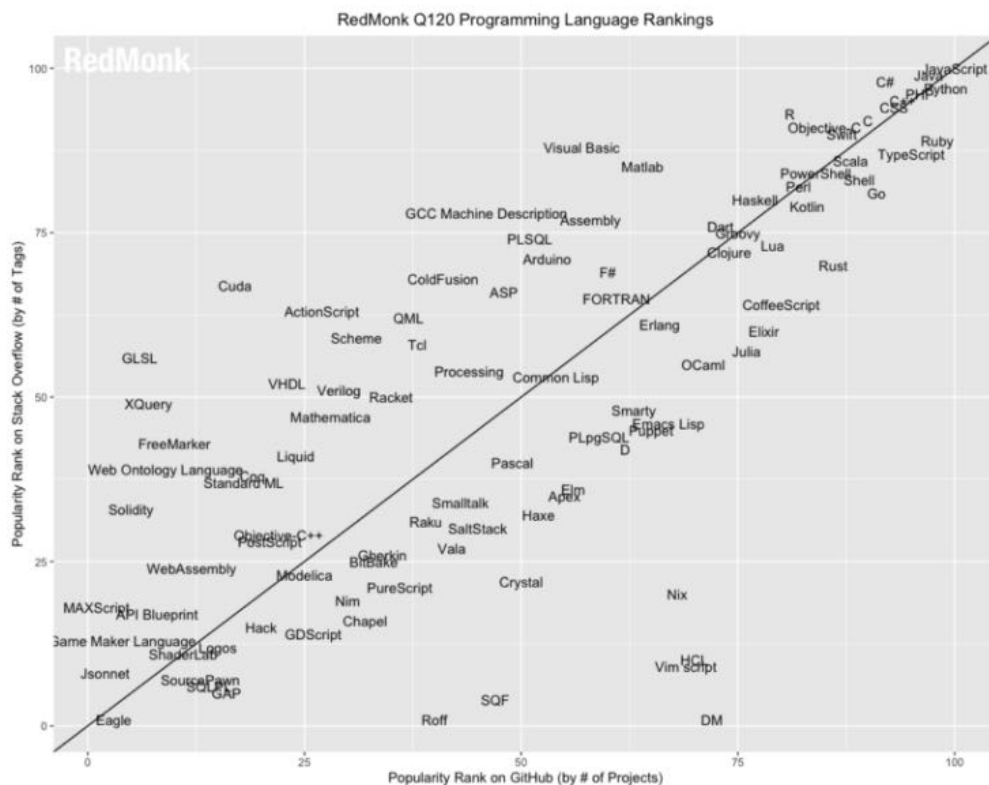


Figura 39: Ranking das linguagens de programação [79]

### 3.1.4 Base de dados

Uma base de dados tem como função gerir variados conjuntos de informações para facilitar a organização, manutenção e pesquisa de dados. Uma base de dados é imprescindível para este projeto e por essa razão decidiu-se seleccionar três delas para estudo e posterior decisão. O estudo das bases de dados *MySQL*, *MongoDB* e *PostgreSQL* encontra-se na Tabela 6.

O *MongoDB* é uma base de dados *noSQL* orientada e excelente para lidar com grandes volumes de dados. O facto de ser documental e não estruturada, permite que um documento represente toda a informação necessária em formato *JSON*, tornando-se por isso uma base de dados não relacional.

A base de dados *PostgreSQL* é muito utilizada para sistemas *Web* com recursos de consultas complexas, facilidade de acesso, chaves estrangeiras e suporta *SQL* (Linguagem de Consulta Estruturada). Tira vantagens como o suporte de grandes volumes de dados, segurança, flexibilidade e alto desempenho.

## Bases de dados

	MySQL	MongoDB	PostgreSQL
Licença	GPLv2	SSPL	BSD
API's e métodos de acesso	API nativa	Protocolo que utiliza JSON	API de streaming nativa
Suporte SQL	Sim	Leitura apenas	Sim
Suporte para múltiplas linguagens	Sim	Sim	Sim
Possui chaves estrangeiras	Sim	Não	Sim
Manipulação de dados em simultâneo	Sim	Sim	Sim
Possibilidade de manter recursos em memória	Sim	Sim	Não
Suporta extensibilidade	Não	Sim	Sim
Base de dados relacional	Sim	Não	Sim
Migração de dados	Sim	Sim	Sim
Documentação	Bom	Bom	Mediocre
Segurança	Bom	Bom	Bom
Flexibilidade	Bom	Bom	Bom
Lida com grandes volumes de dados	Bom	Bom	Bom
Popularidade	Bom	Mediocre	Bom
Curva de aprendizagem	Fácil	Fácil	Difícil
Estrutura de dados	SQL	NoSQL	SQL
Processamento de queries	Queries processadas simultaneamente	Queries processadas uma a uma	Queries processadas simultaneamente

Tabela 6: Estudo e comparação das bases de dados [72, 18, 50, 17]

Em relação à base de dados *MySQL* utiliza também a linguagem *SQL* e é vantajoso em variados aspetos tais como no desempenho, confiança, migração de dados, flexibilidade assim como uma boa documentação. É uma base de dados bastante popular.

Tanto a base de dados *MySQL* e *PostgreSQL* são bases de dados relacionais.

Realizado o estudo, optou-se pela escolha de bases de dados relacionais sendo que estas utilizam linguagem *SQL* permitindo vincular várias informações diferentes através do uso de chaves estrangeiras, enquanto que nas bases de dados *noSQL* é necessário realizar múltiplas consultas e unir os dados manualmente dentro do código.

Ficou-se assim com dúvidas entre a base de dados *MySQL* e *PostgreSQL* sendo que são ambas bastante populares. A base de dados *PostgreSQL* tem desvantagens no que toca à possibilidade de manter recursos em memória enquanto que na base de dados *MySQL* o nível de extensibilidade é reduzido. Por todas as razões apresentadas, a base de dados escolhida foi a *PostgreSQL*.

## 3.2 Metodologias

Nesta secção vão ser explicadas as metodologias escolhidas e utilizadas para este projeto. A primeira - chamada *Data Driven Design* - corresponde à forma como se apresenta os dados ao utilizador. A outra metodologia explicada diz respeito à forma como é feita a análise e seleção dos dados.

### 3.2.1 Data Driven Design

A representação visual permite mostrar os recursos mais relevantes de determinados dados. Consumir esses dados na forma de gráfico, ao contrário de ler uma tabela, permite que um visualizador possa identificar as inclinações, planos, picos e depressões, bem como lacunas e cruzamentos entre as linhas.

Para um observador, que tenta conscientemente compreender uma visualização e extrair entendimento da mesma, passa por três fases diferentes representadas na Figura 40: **perceber, interpretar e compreender**.



Figura 40: Três fases da visão do utilizador

A primeira fase passa por fazer a leitura do gráfico, percebendo que tipo de dados existem, quais as características observáveis assim como identificar padrões ou diferenças relevantes. A segunda fase é interpretar, traduzindo a compreensão em significados quantitativos ou qualitativos de como entender as características mais importantes para cada um de nós.

A última e terceira fase passa por compreender, levando os espectadores a considerarem o que os dados significam para eles mesmos. Os tipos de conhecimento do utilizador estão visíveis na Figura 41. Nesta fase é importante entender o que foi aprendido com aquele tipo de visualização e se reforçou o conhecimento do utilizador ou se aprendeu coisas novas, percebendo se a experiência teve um impacto negativo ou positivo. Por essa razão, o pretendido com uma boa visualização de dados é que o espectador desconheça e que passe a conhecer mais acerca de um determinado assunto ou que conheça mas que fique ainda a conhecer ainda mais. Se o público sabe e não sabe sobre um assunto isso terá uma influência significativa no que os mesmos consideram uma visualização acessível. Por essa razão é necessário simplificar quando o público não tiver conhecimento do assunto e não precisar de adquirir um conhecimento profundo do mesmo ou então esclarecer quando o público não tem conhecimento mas tem capacidade de lidar com um determinado assunto.



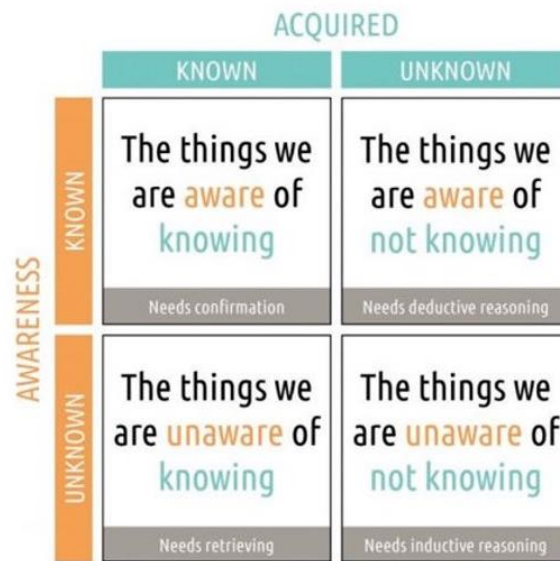


Figura 41: Conhecimento do utilizador [43]

O *Data-Driven Design* [43] é um método de desenvolvimento de produtos e serviços a partir de dados. Todas as etapas do processo pretendem corresponder a informações concretas, que são analisadas tanto de fontes quantitativas, quanto qualitativas.

Uma boa utilização dos dados leva diretamente a melhores resultados nos negócios. Uma pesquisa do *Center for Digital Business* do MIT [21] constatou que “As empresas no uso de tomada de decisão orientada a dados eram, em média, 5% mais produtivas e 6% mais rentáveis que os seus concorrentes”.

Este método traz consigo bastantes benefícios para o negócio, tais como [21]:

- **Agilizar processos:** A utilização de dados antigos, ou o levantamento de dados através de *feedbacks*, pesquisas e entrevistas, garante a agilidade aos processos, melhorando a qualidade dos mesmos;
- **Uma contínua melhoria por um custo menor:** Se antigamente era necessário esperar longos períodos de tempo para levantar métricas e, a partir delas, planejar determinadas alterações no projeto, com esta metodologia o processo de melhoria é ágil e constante;
- **Sistema intuitivo:** Com este método um dos grandes benefícios é ter conhecimento suficiente sobre o utilizador final para apresentar as informações mais relevantes e que os mesmos, entendam os conteúdos de forma simples e intuitiva, com autonomia e conforto. Um dos maiores erros passa pelas informações confusas, pouco intuitivas ou que não levam em conta os desejos do utilizador;
- **Uma melhor experiência do cliente significa mais vendas:** O design serve para criar experiências e quando o mesmo é feito através de informações pré-concebidas sobre o cliente, o



resultado é uma experiência melhor. E experiências melhores vendem mais.

Segundo o *UX Collective*, o *website* de comércio electrónico chamado "*Music & Arts*" utilizou testes de usabilidade e avaliação para modificar o *design* do *website* e desta forma entender o impacto que traria nas vendas. Após a conclusão do projeto, as vendas online aumentaram cerca de 30% no ano seguinte o que demonstra que um bom *design* e experiência do utilizador é um ponto realmente importante a ter-se em atenção. Por essa razão é necessário seguir alguns passos no que toca ao *design* de visualização de dados, definidos na Figura 42:

- **1º Passo** - Planear, definir e iniciar o projeto: como por exemplo definir as características do perfil do público alvo, definir requisitos necessários, informações sobre as fontes de dados, *et al*;
- **2º Passo** - Trabalhar com os dados: recolher, tratar e preparar os dados;
- **3º Passo** - Estabelecer o pensamento editorial: definir e preparar o que se pretende mostrar ao público;
- **4º Passo** - Realizar a solução de *design*: como por exemplo a realização de *mockups*.

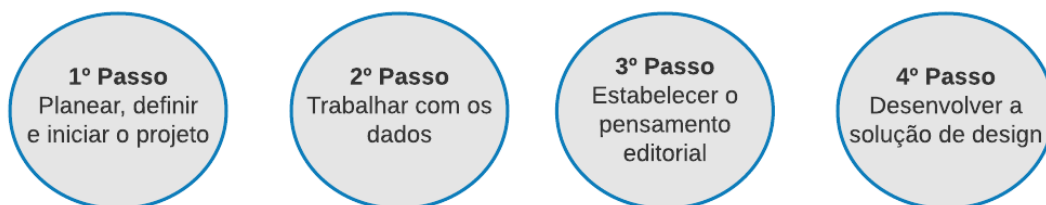


Figura 42: Etapas do processo de design de visualização de dados

### 3.2.1.1 1º Passo

O primeiro passo do processo de *design* de visualização passa pelo planeamento e definição do projeto. Trata-se de identificar os pontos relevantes que impulsionam o projeto e perceber quais as perspectivas oferecem descobertas mais relevantes ou interessantes. Identificar as circunstâncias e o leque de pessoas que teriam interesse no trabalho. Nesta fase, é necessário entender que existem bons princípios que devem ser seguidos, como possuir um *design* confiável, elegante e acessível. No *design* de visualização é necessário dar um entendimento útil ao público, ou seja, deve ser relevante para o assunto e para as suas necessidades e não deve exigir esforço para as perceber ou interpretar.

Para uma boa visualização de dados é necessário representar um assunto relevante para um determinado leque de pessoas, se não podemos não ter motivação do público. Deve-se possuir uma interface

simples e intuitiva mas por vezes pôr tudo simples de mais, pode-se correr o risco de o assunto ser simples de mais ao ponto da obscuridade e conseqüentemente, remover subtilezas e técnicas importantes. Por essa mesma razão é importante estudar o público que pretendemos atingir e a forma como são representados os dados, tendo assim duas opções - simplificar quando o público não tiver conhecimento do assunto e não precisar de adquirir um conhecimento profundo do mesmo ou então esclarecer quando o público não tem conhecimento mas tem capacidade de lidar com o assunto. Já Kirk [43] dizia "Se tudo o que se conseguir é reforçar a compreensão existente ou apenas adicionar um grão extra de compreensão, isso possivelmente representará sucesso".

É preferível a utilização de gráficos familiares e que estamos acostumados a visualizar mas quando existir bons motivos para se utilizarem gráficos diferentes do habitual, devem ser utilizados visto que pode, ser a única maneira de mostrar certo tipo de informação. Cada caso é um caso e em cada um deles as escolhas de *design* devem oferecer uma solução de *design* bastante atraente que não prejudique a mensagem. Para isso é importante entender certos tipos de erros de *design* que não devem/podem ser cometidos como a inexistência de legendas, as proporções exageradas das imagens, a má utilização de cores, o tamanho de escalas ou sequência do *layout*, *et al.*

Para além disso o chamado "espectro do tom" é muito importante para as escolhas de *design* pois ao selecionarmos o tom mais adequado para um determinado projeto, estamos a decidir se damos mais ênfase à capacidade do visualizador de ler ou sentir os dados. As cores mais escuras servem para leitura enquanto as cores mais vivas são utilizadas para "sentir". É também importante perceber que tipos de experiência se pretende que os utilizadores usufruam, estando estas divididas em três tipos como falado no capítulo 2:

- **As visualizações explicativas**, que oferecem uma experiência caracterizada pelo visualizador assumir a responsabilidade de apresentar observações e interpretações importantes, de forma a ajudar o utilizador a assimilar mais rapidamente o significado do que é apresentado;
- **As visualizações exploratórias ou interativas**, que diferem das explicativas por serem mais focadas em ajudar os utilizador, dando-lhes a capacidade de filtrar uma exibição para mostrar apenas certas categorias de interesse ou alternar a exibição para diferentes parâmetros de dados;
- **As visualizações que necessitam de interpretação** são caracterizadas por não serem explicitamente explicativas nem funcionalmente exploratórias. Com visualizações interpretativas os espectadores têm que fazer o trabalho de interpretar o significado contando com sua própria capacidade de perceber e traduzir os recursos de uma visualização.

Temos o exemplo da Figura 43 onde estão representados dois gráficos de barras sobre o mesmo assunto - a percentagem de cada faixa etária (dos 16 anos aos mais de 75 anos) que são proprietários de habitação por cada ano (de 1981 a 2011/12). Analisando bem os dois gráficos que contêm a mesma informação, o gráfico que melhor se adequa para este tipo de informação é sem dúvida o do lado esquerdo,

por variadas razões. Primeiramente as cores, como dito anteriormente são algo fulcral e que têm de ser muito bem seleccionadas o que não acontece com as do gráfico do lado direito sendo estas muito fortes e inadequadas para este tipo de informação assim como o tipo de letra e o título. Para além disto não faz uso de legendas e no eixo y, onde são representadas as percentagens, faz uso do símbolo de percentagem (%) em todos os números, outra situação que não deve acontecer e deve estar representado como no gráfico do lado esquerdo, apenas uma vez. As imagens de fundo também desviam um pouco a atenção do utilizador, tornando a visualização não tão simples. Por último as referências nos gráficos tornam a informação mais credível e aceite juntos dos telespectadores, o que acontece com o gráfico do lado esquerdo.

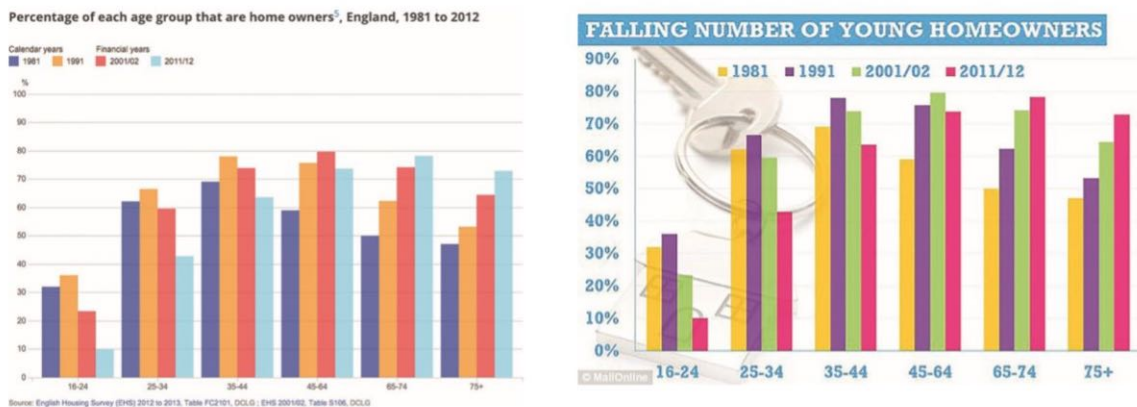


Figura 43: Dois tipos de gráficos que representam a mesma informação [43]

### 3.2.1.2 2º Passo

Nesta segunda etapa do processo de visualização de dados, existem 4 atividades principais:

- **Adquirir e recolher os dados;**
- **Examinar os dados:** familiarizar-se com o tipo de dados e as condições que os mesmos possam trazer;
- **Transformar os dados:** refinar os dados de modo a torná-los mais relevantes para um determinado projeto;
- **Explorar os dados:** usando análise de dados e técnicas de pesquisa para descobrir determinados *insights* que possam existir.

Trabalhar com dados normalizados é sempre o caminho a seguir mas neste capítulo apenas se irá falar da exploração de dados, visto que esta dissertação não passa pela recolha e transformação de dados.

As características dos dados terão uma forte influência sobre o que pode ou não ser qualificado como uma solução de *design* adequada. Temos de deixar claro a base sobre qual a amostra é formada e se pode representar e comunicar os dados de uma forma fiel ao público. Por essa razão é importante realizar um estudo sobre o tipo de dados recolhidos pois estes têm uma grande influência em: determinar o tipo de métodos de análise estatística que se pode utilizar; moldar as perspectivas editoriais que se pode seguir; filtrar os tipos de gráficos específicos que podemos ou não usar com determinados tipos de dados; escolher as cores adequadas e orientar as decisões de composição sobre tamanho, posicionamento e *layout* e refletir sobre as experiências de trabalhar com uma variedade maior de dados qualitativos ou quantitativos, enfatizando os mais relevantes e esquematizando a forma como se pode representar.

Na figura 44 temos um exemplo de uma representação visual de diversos tipos de dados a acontecer ao mesmo tempo. O gráfico representa a previsão do mercado de *big data* em todo o mundo de 2011 a 2026 por diferentes segmentos (em bilhões de dólares americanos). Os dados estão representados pelos diferentes anos comparativamente ao volume do mercado em bilhões de dólares americanos e distribuídos por diferentes categorias (serviços profissionais, gestão de dados, *networking*, et al). Em suma, neste gráfico estão representados tanto dados qualitativos como dados quantitativos sendo necessária uma esquematização de como os representar, que neste caso optou-se por atribuir diferentes cores a cada tipo de categoria e posicioná-las de forma crescente (da que possui maior volume de mercado para a menos).

Big Data Market Worldwide Segment Revenue Forecast 2011-2026  
**Big Data Market Forecast Worldwide from 2011 to 2026, by segment**  
 (in billion U.S. dollars)

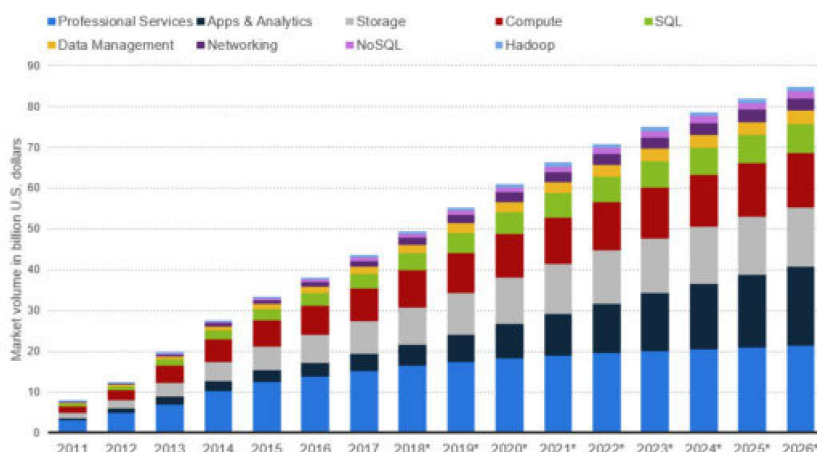


Figura 44: Previsão do mercado de *Big Data* [1]

Desta forma é importante descobrir os *insights* e as qualidades potenciais de compreensão que os dados podem fornecer. A parte da exploração de dados depende muito da informação que se pretende

transmitir ao público alvo mas ainda mais da metodologia que se utiliza para a análise de dados (a ser falada no subcapítulo 3.2.2) e do instinto do analista e do seu raciocínio. Este tipo de raciocínio pode seguir duas vertentes: uma curiosidade ou uma hipótese sobre o conhecimento de um determinado assunto a fim de determinar se há alguma evidência ou interesse na conclusão final ou o tipo de raciocínio mais "aberto" com base num determinado instinto sobre o que pode ser de interesse.

### 3.2.1.3 3º Passo

Para a terceira etapa do processo de *design* de visualização tem-se o pensamento editorial que pode desencadear a necessidade de recolher ou adicionar mais dados sobre um assunto específico para apoiar os ângulos de análise ou as dimensões de enquadramento. Pode também influenciar a necessidade de cálculos adicionais, agrupamentos ou modificações gerais para refinar a preparação dos dados para uma melhor exibição.

Este passo serve para retratar o que se vai mostrar visualmente ao público a fim de satisfazer as suas necessidades. A relevância do foco editorial está principalmente associada a visualizações explicativas, em que elevar os principais *insights* à superfície de um ecrã é um atributo-chave da experiência que os mesmos fornecem. As três perspectivas a serem consideradas para se moldar o pensamento editorial são [43]:

- **Ângulo** - Refere-se à escolha do ângulo de análise - da visão dos dados - que melhor apoiará o entendimento necessário para um determinado assunto. Muitas vezes, tem-se a ideia de querer incluir vários ângulos para servir os interesses do maior número de pessoas mas é necessário ver tudo como um só para abranger o máximo de público possível, tentando ter vários ângulos diferentes com diferentes gráficos.
- **Enquadramento** - Depois de se definir qual o ângulo ou ângulos de análise que é necessário incluir, o enquadramento é a segunda perspectiva editorial preocupada em refinar o conteúdo a ser incluído na sua análise. Esta perspectiva serve para atribuir um limite de uma exibição e apenas o que o público será capaz de processar.
- **Foco** - Preocupa-se com a seleção dos dados para a qual se pretende dar destaque e foco na atenção do utilizador. O foco recai muito sobre a escolha de contrastar visualmente características que consideramos mais importantes do que outras.

Na figura 45 está representado os dados referentes a uma equipa onde é possível visualizar o recorde do número de passes de cada jogador respetivo a cada ano. Definiu-se para este exemplo as seguintes perspectivas do ângulo editorial: o ângulo selecionado reflete os valores quantitativos a mudarem ao longo do tempo (ano) pelas várias categorias (nomes dos jogadores); no enquadramento limita-se um período entre 1930 e 2014 e um limite quantitativo para o número de passes estipulando um mínimo de 30; relativamente ao foco distingue-se o jogador com mais passes (*Peyton Manning*), para ajudar na

visualização de perceber qual dos outros jogadores pode ter uma hipótese de obter esse recorde. Na figura 46 está representado na mesma os dados referentes à mesma equipa de anteriormente mas desta vez em vez de se visualizar o número de passes de cada jogador pelo respetivo ano, visualiza-se pela respetiva idade. Ou seja, apresenta o mesmo enquadramento e foco que o gráfico anterior mas o ângulo é diferente. O ângulo neste gráfico passa a ser os valores quantitativos a mudarem ao longo do tempo (idade) pelas várias categorias (nomes dos jogadores).

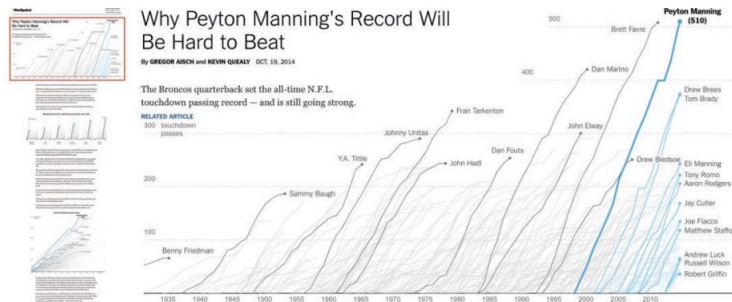


Figura 45: Gráfico do nº de passes de cada jogador por cada ano [43]

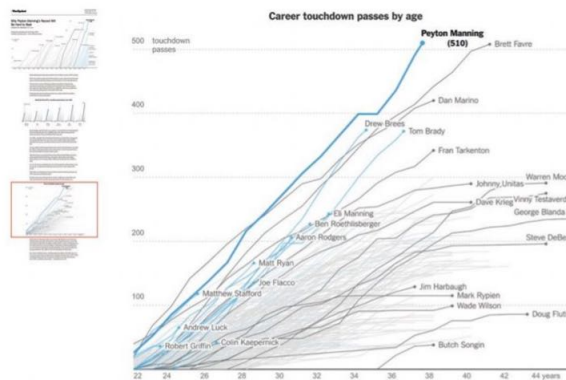


Figura 46: Gráfico do nº de passes de cada jogador pela idade [43]

O pensamento editorial é um fator de extrema importância e que influencia o *design* de visualização de dados, mais especificamente em cinco camadas:

- Representação de dados;
- Interatividade;
- Legendas;
- Cor;
- Composição (sequência de gráficos).

### 3.2.1.4 Representação de dados

A representação de dados diz respeito às diferentes propriedades visuais que existem assim como os diferentes gráficos que oferecem diferentes formas estabelecidas de representar dados (as diferentes formas de representar dados estão detalhadas no capítulo estado de arte, mais concretamente no subcapítulo 2.3), tendo em atenção que por vezes o que se está a tentar representar pode não ser possível utilizando um gráfico tradicional.

Representar diferentes tipos de dados é um desafio e para isso é necessário entender como encaixar diferentes tipos de dados (qualitativos e quantitativos) de diferentes formas, seja através de tamanhos, áreas, volumes, cores, texturas, símbolos, *et al.* No exemplo da Figura 47 em ambos os casos a representação B toma o valor de 5 embora que na representação B do gráfico de barras o tamanho 5 parece correto mas o mesmo não acontece com o círculo B sendo que parece menos correto. O nosso sistema é precioso no que toca a julgamentos relativos a uma linha, em comparação com uma forma. Isso é explicado pelo facto de se julgar que a variação no tamanho das linhas envolve uma mudança no comprimento enquanto que a a variação no tamanho de uma forma geométrica como um círculo ocorre numa dimensão quadrática (área).

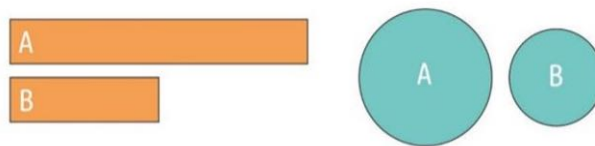


Figura 47: Comprimento versus Área [43]

### 3.2.1.5 Interatividade

Hoje em dias os gráficos interativos são bastante relevantes e importantes para os utilizadores. Na interatividade existem determinados eventos que são definidos pela ação do utilizador, o controlador que se trata do recurso ao qual a ação do evento é aplicada como um menu lateral e a função que representa a operação executada, como os cliques no rato. É necessário entender quais os tipos de interatividade que existem, em que situações os podemos utilizar e também os recursos das tecnologias que temos acesso. A Tabela 7 mostra que tipos de eventos de interatividade são adequados para permitir a interatividade entre o utilizador e o sistema em situações de visualização de dados (como filtrar, realçar, animar, navegar, *et al.*).

A interatividade não é só utilizada nas experiências interpretativas mas também pode ser utilizadas nas experiências explicativas, podendo através da interatividade fazer a transição dos utilizadores para uma interface mais exploratória que os mesmos procurem as suas particulares curiosidades e ajudá-los a visualizar uma outra perspectiva.



	INTERATIVIDADE					
	Filtragem	Realce	Participação	Anotação	Animação	Navegação
Selecionar um botão ou link	✓	✓	✓	✓	✗	✓
Selecionar um item de uma lista de menu	✓	✓	✓	✗	✗	✗
Selecionar vários itens de uma caixa de seleção ou lista de menu	✓	✓	✓	✗	✗	✗
Botão de opção para alterar o estado de um gráfico	✓	✓	✓	✗	✗	✗
Alterar a posição de um objeto ao longo de um controlador deslizante com escala	✓	✓	✓	✗	✓	✓
Alterar a posição de dois objetos ao longo de um controlador deslizante com escala (para criar um intervalo)	✓	✓	✗	✗	✗	✗
Inserir um valor numa caixa de entrada	✓	✓	✓	✗	✗	✗
Selecionar uma marca num gráfico	✗	✓	✗	✓	✗	✗
Passar o rato sobre uma marca num gráfico	✗	✓	✗	✓	✗	✗
Carregar uma página web	✗	✗	✗	✗	✓	✗
Selecionar um botão (reproduzir, pausar, parar, botões de velocidade, etc)	✗	✗	✗	✗	✓	✗
Navegar detalhadamente em torno de um ecrã	✗	✗	✗	✗	✗	✓
Navegar por uma sequência de páginas	✗	✗	✗	✗	✗	✓
Selecionar e desenhar uma região de interesse	✗	✗	✗	✗	✗	✓

Tabela 7: Tipos de interatividade na visualização de dados

Esta deve ser fácil para os utilizadores interagirem mas acima de tudo visualmente simples para não criar determinados obstáculos para os que visualizam. Com isto é necessário possuir um *design* confiável e com um desempenho favorável para a utilização dos utilizadores.

### 3.2.1.6 Legendas

As legendas ou anotações servem como uma assistência para a compreensão de um determinado contexto, função ou o propósito de um projeto de forma a ajudar os utilizadores a perceberem e a interpretar as representações de dados. São exemplos disso:

- **Títulos e apresentações:** Formas de títulos curtos com observações importantes que apoiam uma determinada visualização assim como detalhes do projeto, histórias, descrições, comentários, *et al.*
- **Guia de utilizador:** Num gráfico interativo pode ser necessário fornecer instruções mais detalhadas para auxiliar o utilizador;



- **Rótulos e legendas dos gráficos:** Como títulos de eixo e referências ao longo de cada eixo podem ser uma preciosa ajuda no que toca à visualização gráfica;
- **Apoios de legendas:** Como marcas de escala, referências e marcadores que servem para fornecer contexto e esclarecer determinados assuntos;
- **Notas de rodapé:** Como créditos a autores e origem da fonte de dados são importantes para manter a confiança com os utilizadores;
- **Legendas para os leitores:** A principal diferença entre uma legenda e um guia de utilizador é que uma legenda oferece muito menos explicação escrita, enquanto um guia do leitor orienta mais ativamente o visualizador durante a tarefa de leitura.

Estes tipos de anotações e legendas são extremamente importantes para dar ênfase à informação contida nos gráficos mas a assistência em excesso tornam-se uma sobrecarga. O público é um dos grandes fatores que influenciam nas decisões deste tipo de apoios sendo que é necessário ter em atenção a quem se está a dirigir e as principais características do público-alvo, percebendo por exemplo se estão familiarizados com o assunto do projeto. As decisões sobre o tipo de letra também são necessárias ter em conta.

### 3.2.1.7 Cor

A cor é o estímulo visual mais potente. As escolhas que fazemos terão um impacto imediato no olho do espectador, oferecendo pistas sensoriais sobre o significado dos dados. É importante ter em conta a visão geral dos modelos de cores que existem - cor, saturação e brilho como é visível na Figura 48.



Figura 48: Modelos de cores [54]

As cores servem para tornar legível determinados dados, garantindo as diferenças e associações através de cores significativas. Elas podem ser utilizadas para colorir escalas quantitativas, classificações categóricas, legendas e até recursos interativos. A seleção de cores é influenciada pelo contexto e formato de cada projeto, por isso é necessário ter em atenção de cores potencialmente restritas para um determinado contexto. Muitas das vezes é necessário errar na seleção das cores e voltar a tentar até se obter o resultado pretendido.

### 3.2.1.8 Composição

A variação de posição e tamanho transmite significado. Um gráfico maior do que outro implicará um maior nível de importância. Gráficos de tamanhos iguais, mas localizados em lugares diferentes, farão com que a atenção seja comandada pelo posicionamento ou por ser apresentado primeiramente numa sequência como o exemplo representado na Figura 49. Assim as visualizações de dados podem estar ordenadas por ordem alfabética, cronologicamente, por ordem ordinal e até espacialmente. Para a distribuição da composição hierárquica é importante ter em conta o espaço existente para trabalhar, intervalos de valores assim como a quantidade de ângulos possíveis. A hierarquia traz consigo alguns desafios no que toca à elegância e disposição de dados. É necessário organizar o conteúdo dos dados e pensar em maneiras de dimensionar diferentes tipos de gráficos.

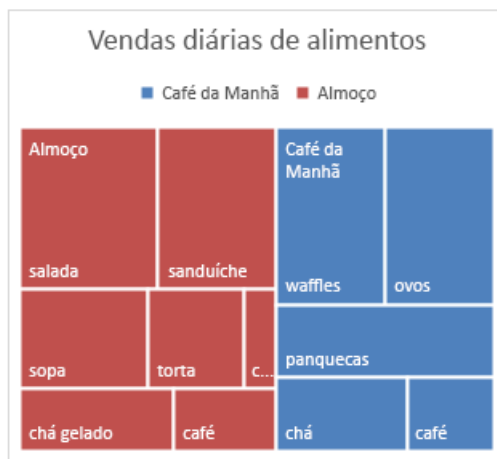


Figura 49: Exemplo da composição de dados num gráfico [20]

### 3.2.1.9 4º Passo

Medir a eficácia de uma visualização de dados continua a ser uma tarefa difícil. Na figura 50 está demonstrado o quarto e último passo do ciclo de desenvolvimento do *design* do projeto que passa pelas seguintes etapas:

- Criar *mockups* ou protótipos;
- Realizar testes: Os testes são indispensáveis para qualquer projeto e passa por solucionar quaisquer problemas que existam para ser um sistema confiável, acessível e elegante. Podem realizados por nós próprios, pelo público ou por qualquer pessoa que tenha o mínimo de conhecimento sobre o projeto;
- Refinar e completar: Esta etapa serve para corrigir erros ou problemas identificados, melhorar conteúdo que esteja em falta, rever os requisitos definidos inicialmente, *et al.*

- Lançar e avaliar o projeto.



Figura 50: Ciclo de desenvolvimento do design

### 3.2.2 Análise de dados

A análise de dados [30] é definida como um processo de limpeza, transformação e modelação de dados de forma a descobrir informações úteis para a tomada de decisões de negócios. O objetivo da análise de dados é extrair informações úteis dos dados e tomar decisões.

Sendo que é necessário analisar os dados e perceber quais os mais relevantes para o projeto assim como as suas relações, é necessário definir uma metodologia. Foram assim definidos quatro passos baseados na metodologia segundo Doyle [48] a ver na Figura 51:

- **1ª Etapa** – Definir a fonte de dados necessária - Determinar a origem de onde se irá retirar os dados fundamentais para o objetivo final;
- **2ª Etapa** - Selecionar e analisar os dados mais importantes e relevantes para o projeto - É possível recorrer a gráficos, tabelas e outros recursos visuais que permitam uma visualização mais clara dos dados recolhidos;
- **3ª Etapa** – Definir perguntas - diz respeito à definição das perguntas que se pretende responder ao analisar os dados que se tem em mãos, arranjando assim possíveis soluções para um determinado problema ou oportunidade. As perguntas devem o mais claras e concisas possíveis. Com a realização destas perguntas pode ser necessário rever os dados ou recolher mais;
- **4ª Etapa** – Interpretar os resultados - Depois de analisar os dados e ver os mesmos respondidos é hora de interpretar os resultados obtidos. Nesta fase final é necessário compreender se os dados

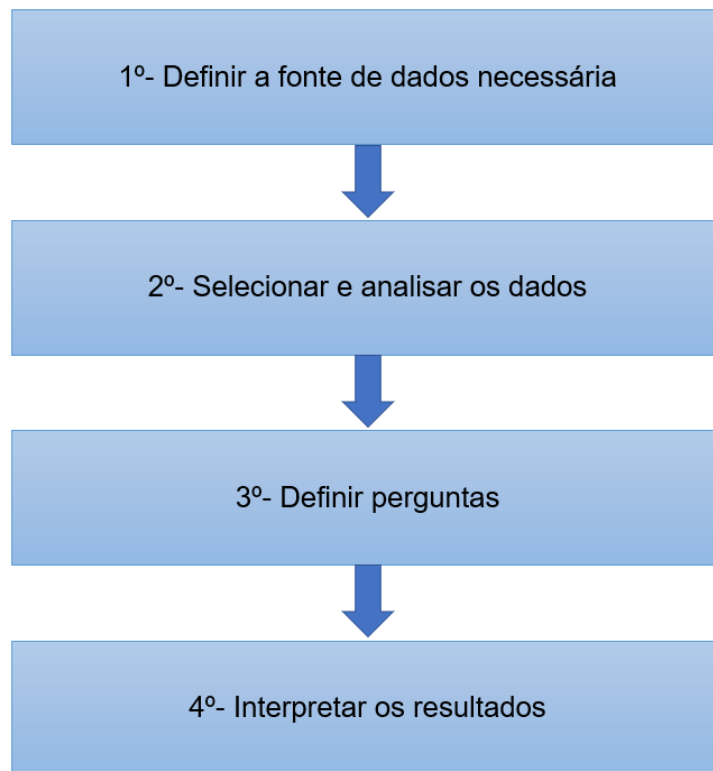


Figura 51: Passos para a análise de dados

respondem a todas as perguntas que foram feitas e se existem limitações de representação de dados.

# PLAYOFF - Prova de Conceito, Requisitos e Arquitetura do sistema

O projeto *PlayOff* passa pelo desenho e desenvolvimento de uma plataforma online que permita a clubes desportivos e estações de televisão visualizar os dados recolhidos de redes sociais, sobre um determinado evento desportivo. Idealizou-se então, que seja desenvolvida uma plataforma online capaz de representar a informação em tempo real proveniente de redes sociais, neste caso do *Twitter*, de uma forma dinâmica, com vários níveis de informação, customizável e que permita aos clubes tomar decisões mais informadas no sentido de melhorar a experiência dos espetadores. Esta dissertação serve então como um protótipo para o projeto *PlayOff* elaborado em contexto empresarial.

## 4.1 Prova de Conceito

O *Power Bi* é um serviço de análise de negócios da *Microsoft* que fornece visualizações interativas e recursos de *business intelligence* [56]. Este permite aos utilizadores criarem os seus próprios relatórios e *dashboards*. O objetivo principal a ser testado no *Power Bi* foi a sua capacidade de resposta a dados recebidos em *streaming* através de vários fluxos, atualização dos mesmos assim como, a demonstração gráfica e a boa visualização de dados. Sendo esta uma ferramenta genérica, pretendeu-se usá-la como prova de conceito para a ideia desta dissertação ser validada.

### 4.1.1 Testes de fluxo - Texto simples

Para a criação de fluxos, utilizou-se o *Power Automate* sendo este também um serviço da *Microsoft* baseado na *cloud*, que permite aos utilizadores na linha de negócio criar fluxos de trabalho de forma prática e simples que automatizam os processos em diferentes aplicações e serviços. Num primeiro exemplo, decidiu-se obter dados sobre o jogo entre o Atlético de Madrid e o Barcelona, sendo o objetivo pretendido visualizar qual o sentimento dos utilizadores do *Twitter* em relação a este jogo ao longo do tempo. Inicialmente, começou-se pela criação dos campos do conjunto de dados em *streaming* que se

pretendiam receber do fluxo criado como está representado na Figura 52. Sendo assim, optou-se por criar um campo de texto (que recebe o texto dos *tweets*), um campo com valor numérico (que continha o valor do sentimento do utilizador respetivo de cada *tweet*) e também um campo *DateTime* (contém a data em que cada *tweet* foi criado).

Novo conjunto de dados de streaming

Crie um conjunto de dados de streaming e integre a sua API no seu dispositivo ou aplicativo para enviar dados. [Saiba mais sobre a API.](#)

\* Obrigatório

Nome do conjunto de dados \*

JogoAtleticoBarca

Valores do fluxo \*

TextoTweet	Texto	🗑️
SentimentoTweet	Número	🗑️
DataTweet	DateTime	🗑️
Inserir um novo nome de valor	Texto	🗑️

```
[
  {
    "TextoTweet" : "AAAAA555555",
    "SentimentoTweet" : 98.6,
    "DataTweet" : "2021-01-21T23:18:55.194Z"
  }
]
```

Análise de dados histórica

Ativado

Figura 52: Criação do conjunto de dados de *streaming* no *Power Bi*

Para a realização deste pequeno protótipo, começou-se por selecionar no conector *Twitter*, já disponibilizado pelo *Power Automate*, o fluxo pretendido, sendo neste caso, um fluxo sobre a análise de sentimentos de *tweets*, um tema muito idêntico ao que se vai realizar nesta dissertação.

Criado o fluxo, foi hora de construir e definir parâmetros para o mesmo. Começou-se por definir o texto de procura que se pretendia quando um *tweet* era postado, onde se optou pela *hashtag* "*#AtletiBarca*" sendo que, foi a *hashtag* criada e assumida para aquele jogo.

Seguidamente, criou-se uma ação de detetar o sentimento do texto do *tweet* (o *Power Automate* disponibiliza várias ações já pré-definidas) para desta forma, nos retornar o valor do sentimento de cada *tweet* acerca de cada texto de procura. Segundo os documentos da *Microsoft*, esta operação retorna uma pontuação numérica entre 0 e 1 [75]. As pontuações próximas de 1 como é possível verificar na Figura 54 indicam sentimento positivo, enquanto as pontuações próximas de 0 indicam sentimento negativo. Por último é adicionada a ação do *Power Bi* (também disponibilizada pelo *Power Automate*), onde se seleciona o *workspace* e os campos do conjunto de dados do *Power Bi* onde queremos inserir os valores que este fluxo gera, visível na Figura 53.

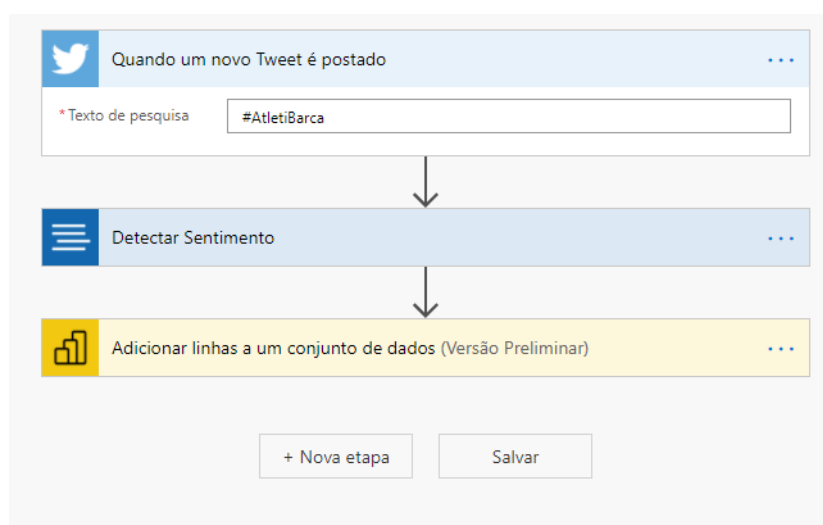


Figura 53: Criação do fluxo no *Power Automate*

### Detect Sentiment

Operation ID: DetectSentimentV2

Returns a numeric score between 0 and 1. Scores close to 1 indicate positive sentiment, while scores close to 0 indicate negative sentiment.

#### Parameters

Name	Key	Required	Type	Description
Text	text		string	The text to analyze.
Language	language		string	This is the 2 letter ISO 639-1 representation of a language. For example, use "en" for English; "es" for Spanish etc.

Figura 54: Operação de detecção do sentimento no *Power Automate*

Criado o fluxo no *Power Automate* com todas as ações necessárias, seguidamente criou-se o relatório para a visualização de dados. Para isso foi necessário ir ao *dataset* criado inicialmente no *Power BI*. Clicando nas variadas opções do mesmo, bastou selecionar a opção "Criar relatório" de forma a criar todas as visualizações interativas pretendidas. Seguidamente, visualizou-se do lado direito todos os campos criados e para criar gráficos com estes dados, bastou escolher o tipo de gráfico e arrastar para os valores do eixo do x e y os dados pretendidos. Foram selecionados os valores do sentimento para o eixo do y, representados com valores entre 0 e 2 que mais perto de 0 representa valores negativos e mais próximos do 2, valores positivos. No eixo do x foi adicionada a data, sendo que o pretendido era obter um gráfico que representasse o sentimento dos adeptos durante o tempo do jogo. O gráfico obtido é possível consultar na figura 55.

Analisando o mesmo foi possível entender que existiram grandes aumentos e decadências a nível de sentimento e por essa razão, tentou-se enquadrar com acontecimentos que ocorreram durante o jogo

naqueles minutos. Ao minuto 47 da partida a equipa do Atlético de Madrid marca um golo o que levou ao sentimento dos adeptos no *Twitter* começar a subir visivelmente.

De seguida, viu-se uma descida acentuada que chega mesmo a atingir valores muito baixos de sentimento que se deve ao facto do jogador Piqué se ter lesionado gravemente, podendo levar esse mesmo acontecimento à finalização do resto da sua carreira. O jogo continua a decorrer e ao minuto 78 saiu um cartão amarelo para a equipa do Atlético de Madrid com uma falta provocada pelo jogador Giménez, levando novamente a uma descida no nível de sentimento dos adeptos. O jogo continua e o Atlético de Madrid vence o Barcelona pela primeira vez na LaLiga, voltando a equilibrar os níveis de sentimentos.



Figura 55: Gráfico que representa o sentimento dos *tweets* durante o jogo

#### 4.1.2 Testes de fluxo - Texto simples com utilizador específico

Depois de testado o protótipo no *Power Bi*, através de um fluxo criado no *Power Automate*, realizou-se novamente outro fluxo mas desta vez a retirar *tweets* apenas de um determinado canal, sendo neste caso referente ao clube Chelsea, um dos maiores do futebol inglês.

O pretendido era visualizar o sentimentos dos adeptos relativamente ao jogo entre o *Chelsea* e o *Tottenham*, mas desta vez sendo um pouco diferente do exemplo anterior. O processo foi igual ao mencionado anteriormente mas em vez de se adicionar de imediato os dados no *Power Bi*, criou-se uma condição. Para explicar melhor este processo, encontra-se na figura 56 um esquema representativo de como se realiza a receção de dados em *streaming* no *Power Bi* através do *Power Automate* e a respetiva visualização de dados. O texto de procura foi a *hashtag* "*CHETOT*", novamente uma *hashtag* criada para o jogo que iria acontecer. O segundo passo passou pela ação de detetar o sentimento do texto do *tweet* e, visto que o pretendido era obter dados apenas de um canal do *Twitter*, foi criada a condição "*TweetedBy ChelseaFC*", o canal oficial do clube *Chelsea*.



A condição passou por verificar se o *tweet* foi *tweetado* pelo canal *ChelseaFC* que caso fosse verdadeira, existiu uma ação de adicionar linhas a um conjunto de dados no *Power Bi*. Se a condição fosse falsa, não adicionava os dados.

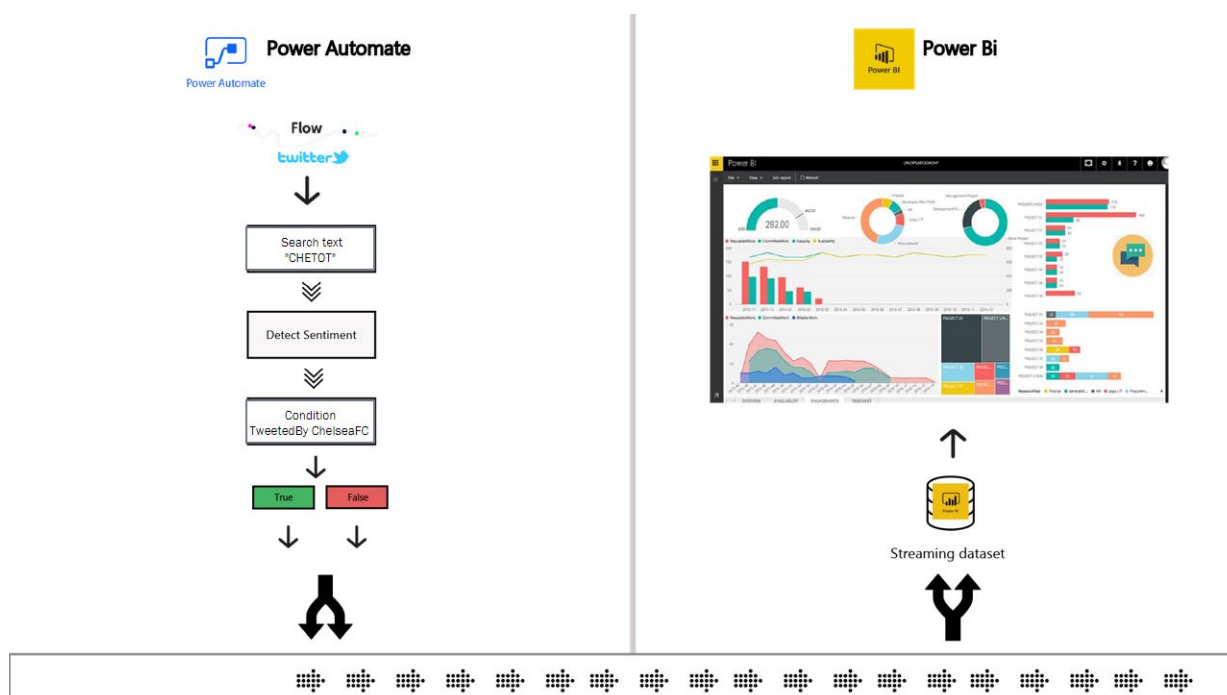


Figura 56: Esquema representativo do *Power Automate* com o *Power Bi*

Decidiu-se ir mais além e em vez de se obter o valor do sentimento num intervalo de 0 a 1, dividiu-se e agrupou-se em 5 categorias de sentimento onde na Figura 56 é possível observar o fluxo criado no *Power Automate* - muito negativo, negativo, neutro, positivo e muito positivo tomando assim distintos valores para cada um:

- Muito negativo - [0 a 0.2];
- Negativo - ]0.2 a 0.4];
- Neutro - ]0.4 a 0.6];
- Positivo - ]0.6 a 0.8];
- Muito Positivo - ]0.8 a 1].

Para a divisão do campo sentimento em categorias, foi necessário criar várias sub condições dentro da condição criada anteriormente para verificar a que intervalos de número correspondiam o valor do sentimento do *tweet* recolhido. Na Figura 57 é possível observar as sub condições criadas no *Power Automate*. Para isso foram utilizadas as expressões que o *Power Automate* também providencia.

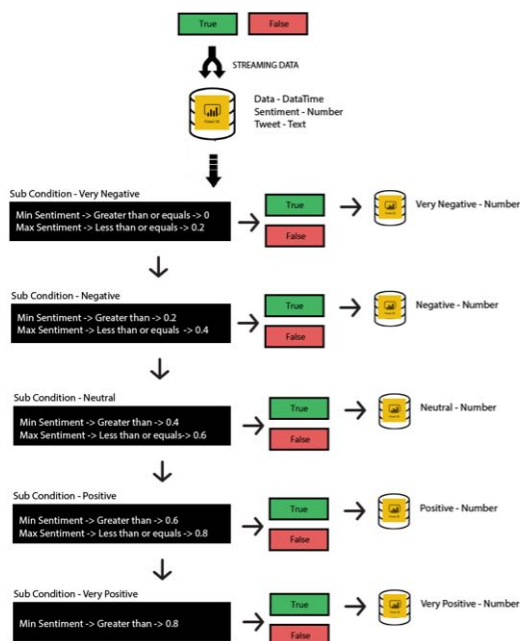


Figura 57: Esquema das sub condições criadas no *Power Automate*

Como é possível observar na Figura 58, tem-se o exemplo para a condição do sentimento "muito negativo" onde se utilizaram as funções de máximo e mínimo para definir o valor desse intervalo, que sendo tipo de sentimento muito negativo tomava o intervalo com valores compreendidos de [0 a 0.2].

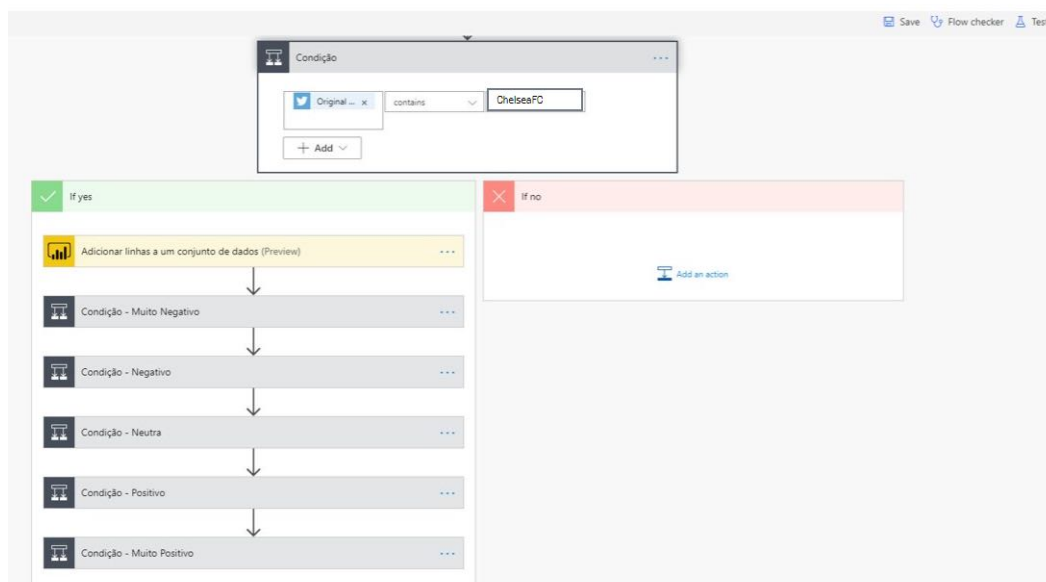


Figura 58: Fluxo criado no *Power Automate* para o canal do *Chelsea*

Se o valor do sentimento recolhido de um *tweet* pertencesse a esse intervalo, os valores do sentimento seriam guardados na variável "Very Negative". O mesmo acontece para as restantes categorias - Negativo, Neutro, Positivo e Muito Positivo.

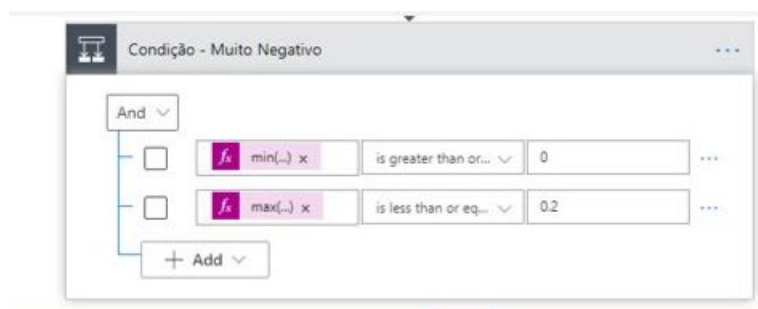


Figura 59: Sub condição do sentimento muito negativo

O resultado obtido está visível na Figura 60 e 61. O objetivo pretendido era entender quais os tipos de sentimentos que predominavam nos *tweets* no canal oficial do *Chelsea* no *Twitter*, relativamente ao famoso jogo referido anteriormente com a *hashtag* definida para o mesmo. Foram recolhidos no total 25 *tweets* publicados ao longo do jogo. Analisando o gráfico construído (na Figura 60) é notório que o sentimento predominante foi o muito negativo, tendo esse os valores mais baixos de sentimento o que se deve ao facto de o *Chelsea* ter perdido o jogo. O sentimento negativo foi o segundo mais influente deste estudo, com um total de 6 *tweets*. Realizado o mesmo estudo mas para o canal do clube *Tottenham (SpursOfficial)* visível na Figura 61, o rival do *Chelsea*, tendo alterado as *hashtags* de pesquisa pois cada canal cria e atribui as suas próprias *hashtags*. Foi possível observar várias diferenças relativamente aos sentimentos dos adeptos, o que já era de esperar visto que esta equipa saiu vitoriosa do jogo. O sentimento dominante nos *tweets* foi sem dúvida o Muito positivo, atingindo valores altíssimos de sentimento.



Figura 60: Tipo de sentimento predominante no canal do *Chelsea*

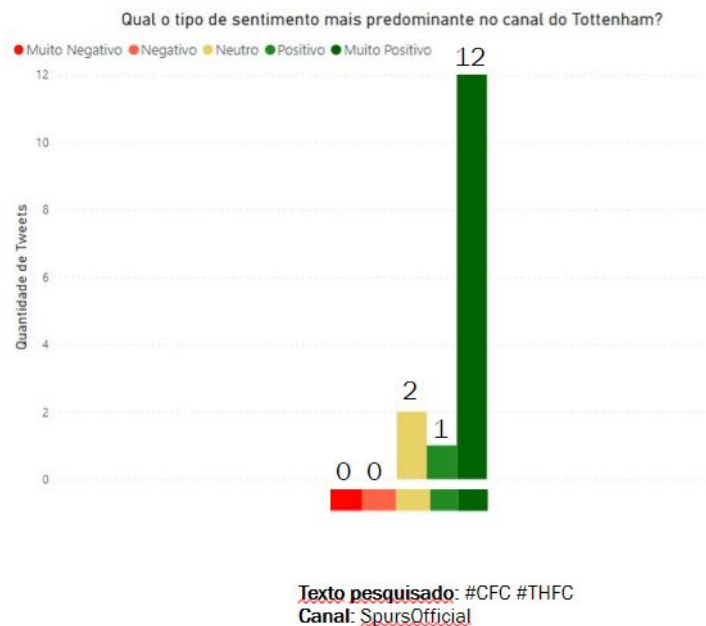


Figura 61: Tipo de sentimento predominante no canal do *Tottenham*

## 4.2 Requisitos

Os requisitos do sistema são declarações articuladas de forma clara sobre o que um sistema deve ser capaz de fazer para satisfazer as necessidades e requisitos dos intervenientes. Devem ser definidos em duas categorias claras, funcionais e não funcionais:

- Os requisitos funcionais descrevem o comportamento exigido e as funções do sistema.
- Os requisitos não funcionais descrevem os critérios específicos que podem ser usados para avaliar o funcionamento de um sistema, como por exemplo o desempenho, a segurança e a disponibilidade.

### 4.2.1 Requisitos Funcionais

Dados os objetivos deste projeto, os principais requisitos do sistema foram decididos através de reuniões periódicas com os *stakeholders* da empresa e são identificados e listados nesta secção. Por uma questão de organização, esta secção é dividida entre a página principal do projeto e a página da *dashboard*, sendo que para cada uma os requisitos necessários são diferentes mas todos classificados com prioridade **alta**.

#### 4.2.1.1 Página Principal

O principal objetivo da página principal é exibir informações sobre os *tweets* de um determinado assunto/evento. Para além de um painel geral, o utilizador tem acesso a um leque de diferentes tipos de

gráficos com diferentes informações detalhadas interativamente. Na Tabela 8 estão representados os requisitos definidos para esta página, sendo estes funcionalidades para os utilizadores recolher dados de uma maneira rápida e eficaz. Em cada requisito é representado também o seu nível de prioridade, podendo ter três tipos: alta, média e baixa.

Requisitos Funcionais do Utilizador	
RFP1	O utilizador deve ser capaz de inserir o texto de pesquisa e/ou o canal do Twitter.
RFP2	O utilizador deve ser capaz de visualizar um conjunto de informações gerais em <i>streaming</i> .
RFP3	O utilizador deve ser capaz de visualizar informações em gráficos em <i>streaming</i> .
RFP4	O utilizador deve ser capaz de parar e reiniciar a coleta e tratamento dos dados na página.
RFP5	O utilizador deve ser capaz de aceder à página da <i>dashboard</i> .
RFP6	O utilizador deve ser capaz de visualizar detalhadamente as informações nos gráficos.
RFP7	O utilizador deve ser capaz de interagir com os gráficos.
RFP8	O utilizador deve ser capaz de fazer download de cada gráfico.

Tabela 8: Requisitos Funcionais - Página Principal

#### 4.2.1.2 Página Dashboard

A página da *dashboard* permite ao utilizador construir os seus próprios gráficos com o assunto que o mesmo pesquisou. Possibilita ao utilizador a oportunidade de selecionar diferentes tipos de dados, tipos de gráficos e até mesmo customizá-lo. Para isso foram definidos os seguintes requisitos presentes na Tabela 9. De notar que em todos os requisitos desta página como na página principal, foram definidos com prioridade alta por serem extremamente necessários para atingir o objetivo final deste projeto.

#### 4.2.2 Requisitos Não Funcionais

Sendo que requisitos não funcionais não interferem diretamente no desenvolvimento do sistema mas sim, pretendem estabelecer como o sistema se comportará em determinadas situações, foram definidos na Tabela 10 os mais importantes a seguir neste projeto.

Requisitos Funcionais do Utilizador	
RFD1	O utilizador deve ser capaz de voltar à página anterior sem perder os dados.
RFD2	O utilizador deve ser capaz de customizar as entradas de dados dos gráficos.
RFD3	O utilizador deve ser capaz de alterar as entradas de dados dos gráficos.
RFD4	O utilizador deve ser capaz de aceder a um calendário para inserir a data.
RFD5	O utilizador deve ser capaz de visualizar um gráfico.
RFD6	O utilizador deve ser capaz de mudar a cor do gráfico.
RFD7	O utilizador deve ser capaz de fazer download do gráfico.
RFD8	O utilizador deve ser capaz de reorganizar a dashboard.

Tabela 9: Requisitos Funcionais - Página *Dashboard*

Requisitos Não Funcionais do Utilizador		Descrição
RNF1	<b>Usabilidade</b>	Toda a interface visual deve ser de fácil interpretação e utilização assim como, ambas as páginas devem ser intuitivas para o utilizador. Um utilizador deve poder terminar uma determinada tarefa num prazo especificado (ou numa quantidade de cliques do rato). Com isto, os utilizadores devem expressar a satisfação de usar o sistema.
RNF2	<b>Confiabilidade</b>	Este requisito é importante no que toca à consistência na transmissão de dados entre o <i>back-end</i> e o <i>front-end</i> , e vice-versa, sendo importante evitar a existência de falhas em ambos. A capacidade de recuperação de falhas, o tempo que o sistema deverá estar disponível para uso, assim como a precisão de confiabilidade são fatores importantes. Estes possuem coeficientes entre 0 (grande quantidade de erros) e 1 (sem erros absolutos), que devem obter valores que comprovem a habilidade do sistema, tomando valores entre 0,55 e 1.
RNF3	<b>Desempenho</b>	O desempenho especifica a quantidade de tempo disponível para o sistema para terminar tarefas e transações especificadas. Por essa razão deve haver um tempo de resposta entre um a três segundos, sendo o tempo máximo de resposta de cinco segundos.
RNF4	<b>Suportabilidade</b>	A capacidade de suporte consiste na adaptabilidade, compatibilidade, configurabilidade, instalação, nível de suporte, escalabilidade e estabilidade. Este sistema deve estar acessível em vários navegadores ( <i>Chrome, Firefox, Microsoft Edge</i> , entre outros menos relevantes), bem como em diversos dispositivos móveis (nomeadamente computador e <i>tablet</i> ). Deve possuir um <i>design</i> responsivo de forma a se adaptar a qualquer um deles.

Tabela 10: Requisitos Não Funcionais

## 4.3 Arquitetura do Sistema

Esta arquitetura divide-se em quatro componentes principais, representados na Figura 62, que vão ser explicados detalhadamente nos subcapítulos abaixo:

- **Data Source;**
- **Core;**
- **Data Visualization;**
- **Data Storage.**

Sendo assim, foi construída uma arquitetura de alto nível que se encontra na Figura 63 onde são ilustrados os componentes do *software* utilizados assim como os relacionamentos existentes entre os mesmos. Dado o estudo das ferramentas a visualizar no capítulo 3.1, decidiu-se optar por utilizar as seguintes tecnologias: o *software* *Nodejs*, a biblioteca *ApexCharts* incorporada na biblioteca *JavaScript* *Reactjs* e a base de dados *PostgreSQL*.

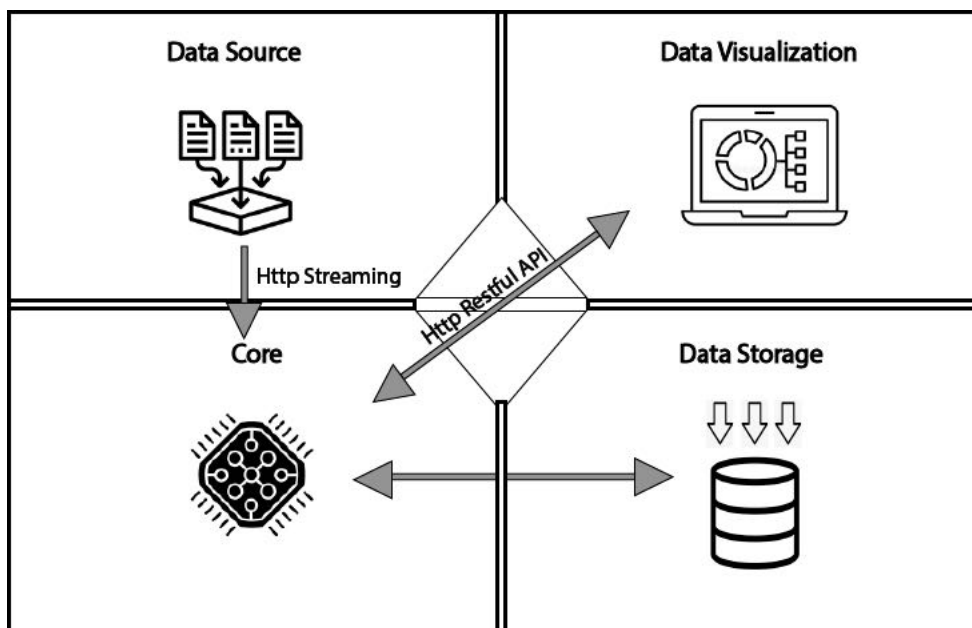


Figura 62: Componentes da arquitetura do sistema

### 4.3.1 Data Source - API Twitter

A visualização de dados em tempo real é um dos pontos mais relevantes deste projeto e por essa razão, foi necessário encontrar um sistema que extraísse e recolhesse dados para este efeito. Feita uma pesquisa, optou-se por utilizar a *API do Twitter*, nomeadamente a parte de consumir dados em *streaming*. Foram

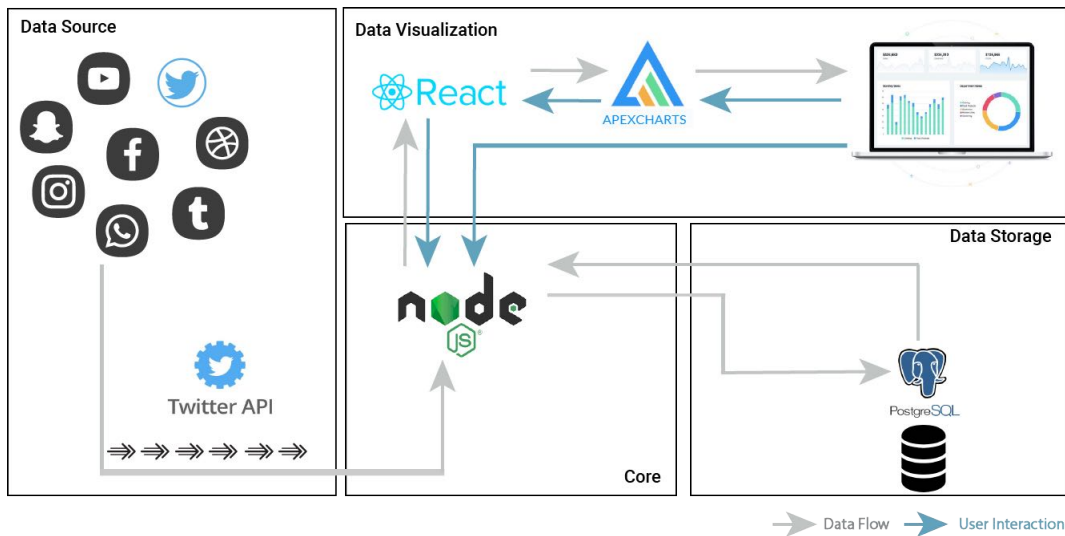


Figura 63: Arquitetura de alto nível

usados protocolos *HTTP Streaming* que proporcionam uma conexão "long-lived" para envio de dados contínuos (ver o sistema deste protocolo na Figura 65). O cliente envia uma solicitação ao servidor e o servidor mantém a resposta aberta por um período indefinido. Essa conexão permanecerá aberta até que um cliente a feche ou ocorra um evento do lado do servidor. Se não houver novos dados para enviar, a aplicação envia uma série de *ticks keep-alive* para que a conexão não feche. Este componente é estruturado em *JavaScript (Nodejs)*, que devolve fluxos com informações dos *tweets* do *Twitter*. Esses dados foram recolhidos e tratados no componente *Core*.

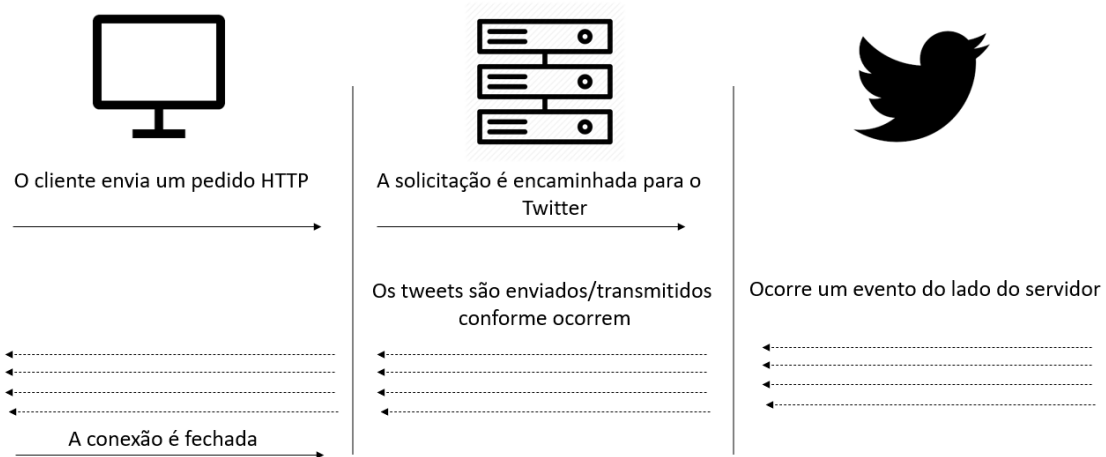


Figura 64: Sistema *HTTP Streaming* [34]

O *Twitter* disponibiliza uma plataforma para programadores onde cada utilizador para utilizar a sua *API* necessita de responder a uma série de questões para esta ser disponibilizada ou não. Sendo aceite, o utilizador pode desfrutar das informações utilizando as suas *keys* e *tokens* de acesso fornecidas. Esta



*API* (Interface de programação de aplicações) utiliza protocolos *HTTP* de forma a fornecer dados através de uma conexão de *streaming* e dessa forma, dados individuais são transmitidos e codificados em *JSON*.

O pretendido era reunir um conjunto de *tweets* sobre um determinado tema, sendo inserido pelo utilizador o texto de pesquisa ou um determinado perfil do *Twitter*, na interface do projeto. Para esse efeito, foram utilizadas duas rotas disponibilizadas por esta *API*:

- ***statuses/filter***: recebe o texto de pesquisa, o perfil de um utilizador ou ambos e retorna em *streaming* todos os *tweets* de acordo com os campos definidos. Esta rota retorna um *JSON* com o campo "track" que filtra pela palavra/termo de pesquisa pretendida(o) e o campo "follow" que serve para retornar todos os *tweets* de um determinado utilizador utilizando o *id* desse mesmo canal do *Twitter*.

**Exemplo: *T.stream('statuses/filter', track: 'mango', follow: '123456789');***

Neste caso, a *API* apenas vai devolver *tweets* que contiverem a palavra "mango" e que forem *tweetados* ou *retweetados* através do utilizador "Reuters" (utilizador respetivo ao *id* 123456789) que é um canal de notícias inglês. Este canal é uma fonte de informação credível e muito ativo no *Twitter*, daí ter sido utilizado para este exemplo.

- ***users/show***: nesta rota o parâmetro passado é o nome do perfil de um utilizador (exemplo: @Reuters) e esta rota do tipo *GET* retorna o *ID* respetivo, sendo que, na rota acima descrita o campo do perfil de um utilizador apenas é válido com o *ID*.

**Exemplo: *T.get('users/show', screen\_name: @Reuters);***

Para além destas rotas disponibilizadas pela *API* do *Twitter*, foi também utilizada a rota ***trends/place*** que retorna os 50 principais tópicos de tendência de um determinado local, ou seja, as tão famosas *trends* do *Twitter* de um determinado país. Para isso é necessário transmitir o *id* do país que neste caso foi o de Portugal.

**Exemplo: *T.get('trends/place', id: '23424925');***

Em suma, todas estas rotas foram necessárias e utilizadas para extrair dados necessários para este projeto.

### 4.3.2 Core

Construído em *Nodejs*, é responsável por orquestrar toda a informação e servir de encaminhador da mesma. Neste componente, alguns tipos de dados foram tratados de forma a obter informações necessárias para os objetivos deste trabalho e comunicados com o componente *Data Visualization*, através de uma *RESTful API*, servindo assim como uma ferramenta de *backend*. Este componente permite a recolha dos dados em tempo real do *Twitter*, selecionando os campos pretendidos de cada *tweet* e guardando-os e/ou exibindo-os no *frontend*. Para iniciar este *streaming*, o *backend* é responsável por captar um texto de pesquisa ou o nome de um utilizador do *frontend* como é possível verificar no código na lista 1.

Excerto de Código 1: Receção de dados via verbo *POST*

```
router.post('/dataForm', function(req, res) {
```

```

let { hashtag, follow } = req.body;
// Se contém só o username
if(hashtag === '' && follow !== '') {
  twitterSentiment.getUserIdFromUsername(follow, undefined);
  res.status(201);
// Se contém só a hashtag
} else if(follow === "" && hashtag !== "") {
  twitterSentiment.changeHashtag(hashtag);
  res.status(201)
// Se contém os dois (hashtag e follow)
} else if(follow !== "" && hashtag !== "") {
  twitterSentiment.getUserIdFromUsername(follow, hashtag);
  res.status(201);
} else {
  // lida quando não contém nem hashtag nem username
  res.status(400);
  console.log("User & hashtag empty");
}
});

```

Este componente contém uma rota *POST* onde verifica se existe um desses campos inseridos pelo utilizador ou se existem ambos. Caso existam, o sistema executa determinadas funções para fazer a recolha dos *tweets* com esses parâmetros nomeadamente, os pedidos *GET* e *STREAM* explicados no sub capítulo acima da *API* do *Twitter*. São feitas várias configurações para definir os parâmetros que vão ser usados na rota final. Sempre que a função *openStreamOfTweets* é chamada (código presente na lista 2), o *streaming* de dados é aberto e a rota de *stream* da *API* do *Twitter* é chamada, começando assim, a serem recebidos *tweets* em tempo real onde são definidos e tratados.

#### Excerto de Código 2: Função *openStreamOfTweets*

```

_openStreamOfTweets () {
  this.streamIsOpen = true;

  var self = this;
  this.stream = self.twitterClient.stream('statuses/filter',
    self._getConfig());

  self.stream.on('tweet', function (tweet) {
    .....
  });
}

```

```
}

```

Para a resposta com os *tweets* em tempo real foi criada uma rota do tipo *GET* e outra do mesmo tipo para as *trends* do momento (como é possível observar no código da lista 3), de forma a comunicar com o *front-end*. Foram construídas variadas rotas para responderem às solicitações do cliente através de métodos *HTTP*.

Excerto de Código 3: Rotas de resposta *GET* criadas

```
router.get('/apiTwitter', function(req, res) {
  .....
  res.json( twitterSentiment.tweetsWithSentiments );
});

router.get('/trends', function(req, res) {
  T.get('trends/place', params, (err, data, response) => {
    .....
    res.json( data );
  })
})
```

Todas as outras rotas foram do tipo *POST*, com o objetivo de receber e enviar informações processadas sendo estas rotas utilizadas para o formulário da página principal (como referido acima) mas também para a página da *dashboard*. Nessa página existem vários formulários e campos que o utilizador necessita de preencher para criar a sua *dashboard*. E por essa razão, é necessário este componente para realizar as *queries* à base de dados e posteriormente comunicá-los com o componente de *front-end*.

Para a recolha e tratamento de dados foi necessário definir os campos dos dados, dos quais se pretendia para posterior utilização. Para além disso, foram também utilizados alguns *packages*, fazendo uso dos *tweets* de forma a retirar outras informações relevantes para este projeto. Nos sub capítulos é explicado com mais detalhe os dados recolhidos e tratados, assim como os pacotes utilizados.

#### 4.3.2.1 Dados recolhidos e tratados

Sendo que este componente serviu para recolher e tratar dados a serem posteriormente utilizados no *frontend* e inseridos na base de dados, foi elaborada uma análise aos dados provenientes da *API* do *Twitter*. Com isso, foram selecionados os seguintes dados para exibir em tempo real na página principal:

- *Id* do *tweet*;
- Texto do *tweet*;
- Data de criação do *tweet*;
- Idioma do *tweet*;

- Número de *retweets* do *tweet*;
- Utilizador que publicou o *tweet*;
- Imagem do utilizador que publicou o *tweet*;
- Género do utilizador que publicou o *tweet*;
- Sentimento do *tweet*;

Tendo assim como exemplo no código da lista 4:

Excerto de Código 4: Exemplo dos dados dos *tweets* recolhidos

```
self.tweetsWithSentiments.push({
  id: tweet.id,
  text: tweet.text,
  .....
});
```

Por outro lado, foram também selecionados alguns tipos de dados para serem guardados na base de dados para posterior utilização na página da *dashboard*, página essa onde o utilizador constrói os seus próprios gráficos. É possível observar no subcapítulo *Data Storage* informação mais detalhada sobre os dados utilizados.

#### 4.3.2.2 Package Sentiment

Neste componente foram utilizados vários *packages*, como o caso do *package sentiment*. Foi utilizado para, através do texto de cada *tweet*, retornar o respetivo valor do sentimento com valores entre 0 e 1, sendo 0 um sentimento muito negativo e 1 um sentimento muito positivo. Optou-se por dividir os sentimentos em três tipos:

- **Negativos** : valores entre  $\geq 0$  e  $\leq 0.3$ ;
- **Neutros**: valores entre  $> 0.3$  e  $< 0.7$ ;
- **Positivos**: valores entre  $\geq 0.7$  e  $\leq 1$ ;

Esta análise de sentimento nos *tweets* pretende avaliar as opiniões das pessoas em sentimentos negativos, neutros e positivos ou até mesmo em estados de emoção, por exemplo, feliz, com stress, triste e/ou revoltado, através de métodos de análise de texto.

```
let result = sentiment.analyze(tweet.text);
let sentimento = (result.score + 5)/10;
```

### 4.3.2.3 Package Gender-Detection

Outro *package* utilizado designa-se por *gender-detection*, que teve como base recolher o género do utilizador de cada *tweet* através do *username*. Este pacote, através do *username* de um utilizador, encarrega-se de identificar o possível género, sendo masculino, feminino ou sem género, sendo estas as identificações possíveis.

```
genderUser = gender.detect(tweet.user.name);
```

### 4.3.2.4 Sequelize

Sendo que este componente ficou também encarregue de fazer a ligação entre o componente *Data Storage*, servindo como um comunicador da base de dados, foi utilizada uma "promise-based" *Nodejs ORM* chamada *Sequelize*, que oferece suporte a transações e a relações, funcionando como uma ponte entre o *backend* e a base de dados *PostgreSQL*. Foi feita a conexão com a base de dados, e definido o respetivo modelo de cada *tweet* como é possível observar na lista 5. Foi utilizado também para a realização de *queries* à base de dados, de forma a obter os dados para os gráficos criados pelo utilizador, na página *Dashboard*.

Excerto de Código 5: Definição do modelo de um *tweet*

```
const Tweet = sequelize.define('tweets', {
  id: {
    type: DataTypes.STRING,
    primaryKey: true,
    autoIncrement: true
  },
  data: {
    type: DataTypes.DATE
  },
  sentiment: {
    type: DataTypes.DOUBLE,
  },
  hashtag: {
    type: DataTypes.STRING,
    allowNull: true
  },
  follow: {
    type: DataTypes.INTEGER,
    allowNull: true
  },
},
```

```

    hour: {
      type: DataTypes.STRING
    },
    gender: {
      type: DataTypes.STRING
    },
    retweets: {
      type: DataTypes.INTEGER
    }
  });

```

Os controladores deste componente, que servem como roteadores para lidar com solicitações feitas à aplicação, são eles também orientados pelo *Sequelize*. Temos um exemplo abaixo na lista 6, onde mostra uma *query* feita à base de dados utilizando esta promessa.

Excerto de Código 6: Exemplo de *query* usando o *Sequelize*

```

module.exports.getByDatesEixoHashtagUser = (dataInicial,
  dataFinal, eixoX, eixoY, hashtag, utilizador) => {
  return Tweet.sequelize.query('SELECT \${eixoX}, \${eixoY} FROM
    tweets WHERE data BETWEEN '\${dataInicial}'
    AND '\${dataFinal}' AND hashtag='\${hashtag}'
    AND follow='\${utilizador}');
};

```

### 4.3.3 Data Visualization

No componente *Data Visualization* foi utilizada a biblioteca *JavaScript* chamada *Reactjs*, utilizada para criar a interface nas páginas *web*. Todos os dados foram obtidos através de rotas definidas no componente *Core* e comunicadas com este componente. Para o *frontend* deste projeto foram construídas duas páginas diferentes: a principal e a página em que o utilizador consegue construir os seus próprios gráficos.

#### 4.3.3.1 Página Principal

Na primeira página, o o utilizador insere um termo de pesquisa ou um utilizador do qual pretende que sejam exibidos os dados através de um formulário. Esses filtros são tratados no componente *Core* e os dados, já tratados, são enviados para este componente em tempo real através de métodos *POST*. Neste componente são definidos e implementados os gráficos a serem utilizados, feita a renderização dos mesmos com as respetivas *series* e *labels* para cada um exibir tal como as *trends* numa tabela e os dados do *slideshow*. Para além disso, visto que o utilizador possui um botão para começar ou parar o

*streaming* de dados, é feito neste componente ações que comunicam com o *back-end* de modo a "abrir ou fechar" os dados em tempo real. São utilizadas também janelas pop-up (dialogs) para quando não existem dados serem exibidas ao utilizador, como é mostrado no exemplo da lista 7.

Excerto de Código 7: Exemplo da utilização do *dialog*

```
{ this . state . dialog === true ? <AlertDialog
      closeDialog = { this . closeDialog . bind ( this ) }
cancelSearch = { this . cancelSearch . bind ( this ) } /> : null }
```

### 4.3.3.2 Biblioteca ApexCharts

A biblioteca *ApexCharts* foi introduzida neste componente de forma a serem criados todos os gráficos interativos e atualizáveis em tempo real. Esta fornece uma ampla variedade de gráficos flexíveis e responsivos, sendo de código aberto. Todos os dados recebidos do componente *Core* são injetados nos gráficos e os mesmos, são posteriormente customizados com *design* já definido. Para isto, foram decididos os gráficos a serem utilizados e de seguida importados/construídos, conforme os dados disponíveis:

- **Gráfico circular** - Percentagem de linguagem de cada utilizador;
- **Gráfico de Barras** - Sentimento ao longo do tempo e Número de *retweets* ao longo do tempo;
- **Gráfico de Área** - Sentimento ao longo do tempo;
- **Gráfico Radial** - Percentagem de sentimento de cada utilizador e Género de cada utilizador;
- **Gráfico de Dispersão** - Tipo de sentimento (positivo, negativo ou neutro) por cada *tweet*;
- **Tabela** - Top 10 de *trends* no momento.

Todos os gráficos são criados em ficheiros numa pasta "*Charts*" individualmente e importados para o ficheiro da página principal. Sendo que esta biblioteca permite interactividade, configuração e personalização dos gráficos, foi isso mesmo que foi feito. É possível definir a família fonte da letra, cores, tamanhos, animações, eventos de interatividade, responsividade, gradientes, funções personalizadas, títulos, subtítulos, temas, entre outros. A própria biblioteca permite definir nos gráficos eventos onde construímos a lógica para uma determinada ação. É exemplo disso o código abaixo presente na lista 8, onde foi criado um evento que permite ao utilizador clicar num *tweet* do gráfico e este ser reencaminhado para o respetivo *tweet* completo na página oficial do *Twitter*, de modo a podê-lo visualizar mais detalhadamente.

No capítulo 5 é explicado o manual do utilizador em que são explicadas todos os tipos de interatividade que o utilizador tem com o *front-end*.

Excerto de Código 8: Exemplo de evento utilizado no gráfico

```
chart: {  
  fontFamily: 'Roboto',  
  .....  
  events: {  
    markerClick: function (event, chartContext,  
      {seriesIndex, dataPointIndex, config}) {  
      var valoresId = _this.getValoresApiTwitter()[19];  
      window.location.href = 'https://twitter.com/anyuser/  
        /status/' + valoresId[dataPointIndex];  
    }  
  }  
}
```

#### 4.3.3.3 React-twitter-widgets

Permite exibir *widgets* do *Twitter*, ou seja, uma interface gráfica que mostra o *tweet* como ele é visualmente no site oficial, com o número de gostos e de comentários. Foi utilizado para o *slideshow*, como mostra o exemplo de código da lista 9, e apenas é necessário fornecer o *id* respectivo de cada *tweet* para a exibição do mesmo. Para isso são filtrados no componente *back-end* o top 3 de sentimentos mais negativos e positivos, sendo estes exibidos no *slideshow* através deste componente.

Excerto de Código 9: Exemplo do *slideshow* criado

```
<Slider {... settings} idsPositivos={this.props.idsPositivos}  
idsNegativos={this.props.idsNegativos}>  
  {this.state.lista.map((id) => {  
    return (  
      <div className="tweetSlideshow">  
        <Tweet tweetId={id} />  
      </div>  
    )  
  })}  
</Slider>
```

#### 4.3.3.4 Página Dashboard

O ponto principal desta página é permitir ao utilizador a criação dos seus próprios gráficos customizados. Para isso foi criado um sistema de *grids* (utilizando o *package gridlayout*), onde estão inseridos vários



campos que o utilizador consegue seleccionar, definindo assim os dados para construir os seus próprios gráficos. Estes dados são enviados para o componente *Core* utilizando o package *axios*, onde são realizadas as respetivas *queries* à base de dados para dessa forma, serem devolvidos todos os dados que o utilizador pretende que sejam visíveis nos gráficos. Todos estes os gráficos disponíveis foram utilizados novamente pela biblioteca *ApexCharts*. Esta biblioteca permite também o *download* dos gráficos, servindo assim de apoio para o utilizador.

#### 4.3.3.5 Axios

Foi utilizado este *package* que permite realizar solicitações *HTTP* para *Node.js* utilizando neste caso, métodos *POST*. É utilizado neste componente para comunicar com o componente *Core* efetuando *queries* à base de dados conforme os campos selecionados pelo utilizador nos formulários desta página, recebendo assim os dados filtrados para serem injetados e finalmente exibidos no(s) gráfico(s).

#### 4.3.3.6 GridLayout

O sistema *GridLayout*, um *package* do *Reactjs*, foi utilizado para a construção de um sistema de *layout* com três *grids*. Este foi selecionado por permitir responsividade e arrastar e largar (o famoso *drag and drop*). Tem como vantagem dar ao utilizador a oportunidade de posicionar e aumentar/diminuir a sua *grid*, permitindo assim uma melhor visualização e comparação dos dados obtidos nos gráficos.

#### 4.3.3.7 React-select

Em cada *grid*, existe um formulário com vários campos que o utilizador tem de seleccionar para ser devolvido um gráfico customizado. Foi utilizado o *react-select* para criar os formulários, tendo este um bom *design* para componentes *React.js* e extremamente personalizáveis.

#### 4.3.3.8 React-dates

Visto que nos formulários do utilizador, o mesmo pode seleccionar a data ou intervalo de data do qual pretende ver *tweets*, foi utilizado o pacote *react-select*. Este permite a utilização de calendários interativos (*Date Picker*). Para isso foram utilizados dois tipos: *SingleDatePicker* e *DateRangePicker* - o primeiro permite apenas seleccionar uma data e o segundo permite seleccionar um intervalo de datas. No código da lista 10 é possível visualizar um exemplo de utilização de um *single datepicker* onde se define um máximo de intervalo de 8 dias.

Excerto de Código 10: Exemplo do *SingleDatePicker* criado

```
<SingleDatePicker
  appendToBody
  date={this.state.date}
  onChange={date => this.setState({date})}
```

```
focused={this.state.focused}
onFocusChange={({focused}) => this.setState({focused})}
id="your_unique_id"
isOutsideRange={(day) => day.isAfter(moment()) ||
day.isBefore(moment().subtract(9, 'days'))}
/>
```

#### 4.3.3.9 React Color

Este *package* foi utilizado para a customização dos gráficos criados pelo utilizador, ou seja, o mesmo permite ao utilizador escolher a cor pretendida para cada tipo de gráfico. Disponibiliza um *Color Picker*, onde é selecionada a cor e essa é injetada no campo "color" de cada gráfico.

#### 4.3.4 Data Storage

O componente *Data Storage*, formado por uma base de dados *PostgreSQL*, permite guardar informação durante um período de tempo, funcionando assim como uma *cache* temporal para analisar os dados a posteriori. Decidiu-se optar por armazenar os dados durante 8 dias, sendo necessário definir um intervalo de tempo máximo pois a quantidade de dados é enorme e ocupa muito espaço. Sendo assim, o utilizador consegue fazer proveito de tempo suficiente para comparar dados que pretenda em diferentes dias, tendo uma semana para o fazer.

De lembrar, que apenas foram feitos a persistência de algum tipo de dados de cada *tweet* que serão utilizados na página da criação de uma *dashboard*:

- **Id** do *tweet*;
- **Data** do *tweet*;
- **Hora** do *tweet*;
- Número de **Retweets** do *tweet*;
- **Género** do utilizador que publicou o *tweet*;
- Valor do **Sentimento** do *tweet*;
- **Hashtag** - Termo de pesquisa respetivo que o utilizador pesquisou;
- **Follow** - Perfil do utilizador respetivo que o utilizador pesquisou.

Todos estes campos foram guardados para o utilizador ter várias opções de dados para os eixos dos gráficos, como o valor do sentimento, a data, entre outros. Os campos *hashtag* e *follow* foram necessários

para filtrar pelos dados respetivos e retorná-los sobre cada tema/assunto que o utilizador teria pesquisado para a construção dos seus próprios gráficos.

## Design e Desenvolvimento

Neste capítulo é apresentado todo o trabalho desenvolvido em termos gráficos e de usabilidade. O capítulo cobre desde as *mockups* realizadas numa fase prévia do desenvolvimento até à explicação de como usar as funcionalidades do protótipo desenvolvido. Este capítulo também pretende ser entendido como um guia de apoio para a utilização da solução desenvolvida.

### 5.1 Mockups

Como escrito no capítulo 3, nas metodologias deste projeto um dos passos do desenvolvimento do *design* passa pela criação de *mockups*. Dessa forma foi selecionada uma paleta de cores (ver na Figura 65) a ser utilizada para a criação dos esboços, e posteriormente nas páginas *web*. Sendo o projeto uma visualização de *tweets* negativos, neutros e positivos, foram definidas estas cores para representar cada uma delas.



Figura 65: Paleta de cores

Desta forma e tendo em conta as boas práticas de *UX* e *UI*, foram elaboradas *mockups* de modo a fazer uma demonstração e avaliação do *design*, percebendo como se enquadravam melhor para este projeto.

Era importante definir um conjunto de ideias:

- Logótipo do projeto;
- Como se disponibilizariam os gráficos na página de modo a ser intuitivo e perceptível para o utilizador;
- Como seriam facultados os dados ao utilizador sendo absorvidos de forma rápida e eficaz;
- Como estaria construída a página para criação dos gráficos.

Sendo assim, foram realizadas duas *mockups*, uma para a página principal e outra para a página da *dashboard* como é possível observar nas Figuras 66 e 67 .

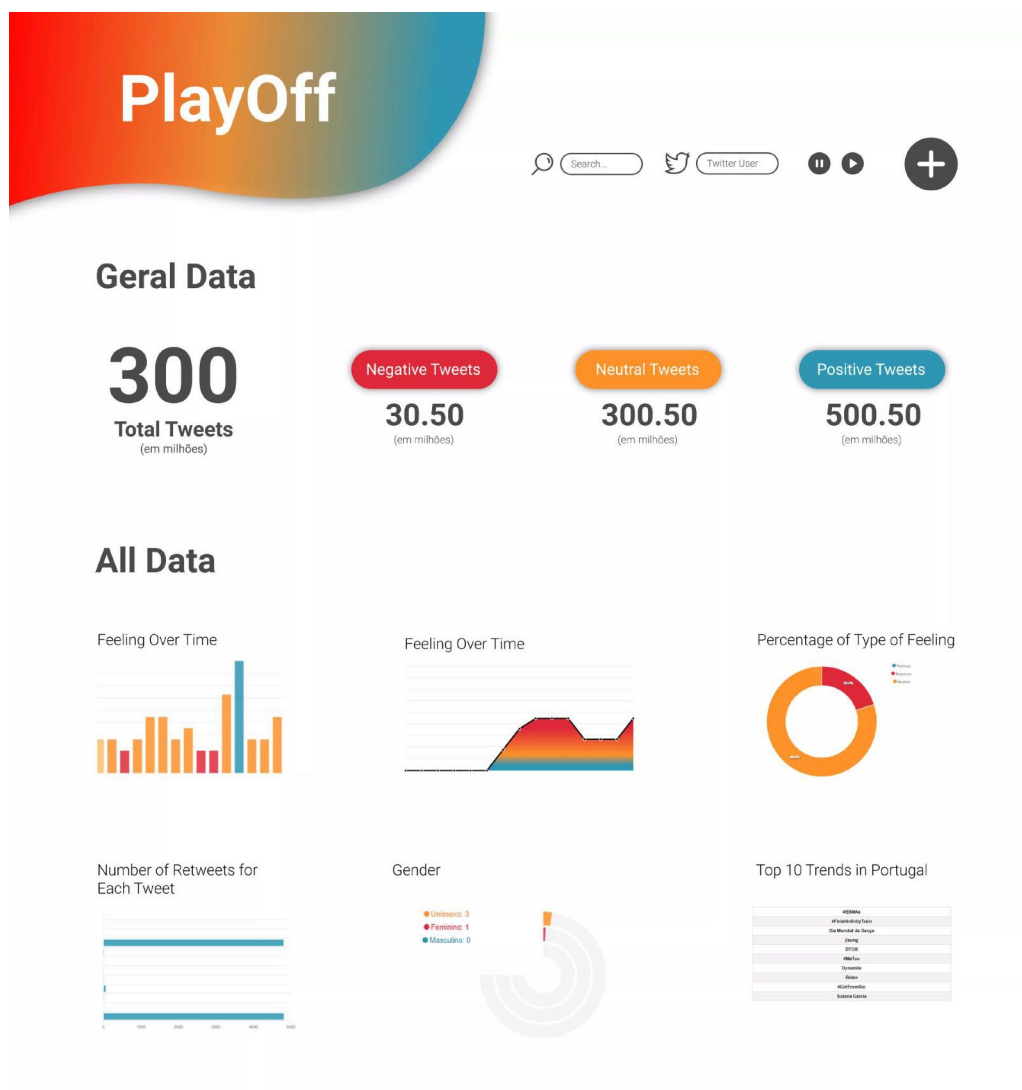


Figura 66: *Mockup* da Página Principal

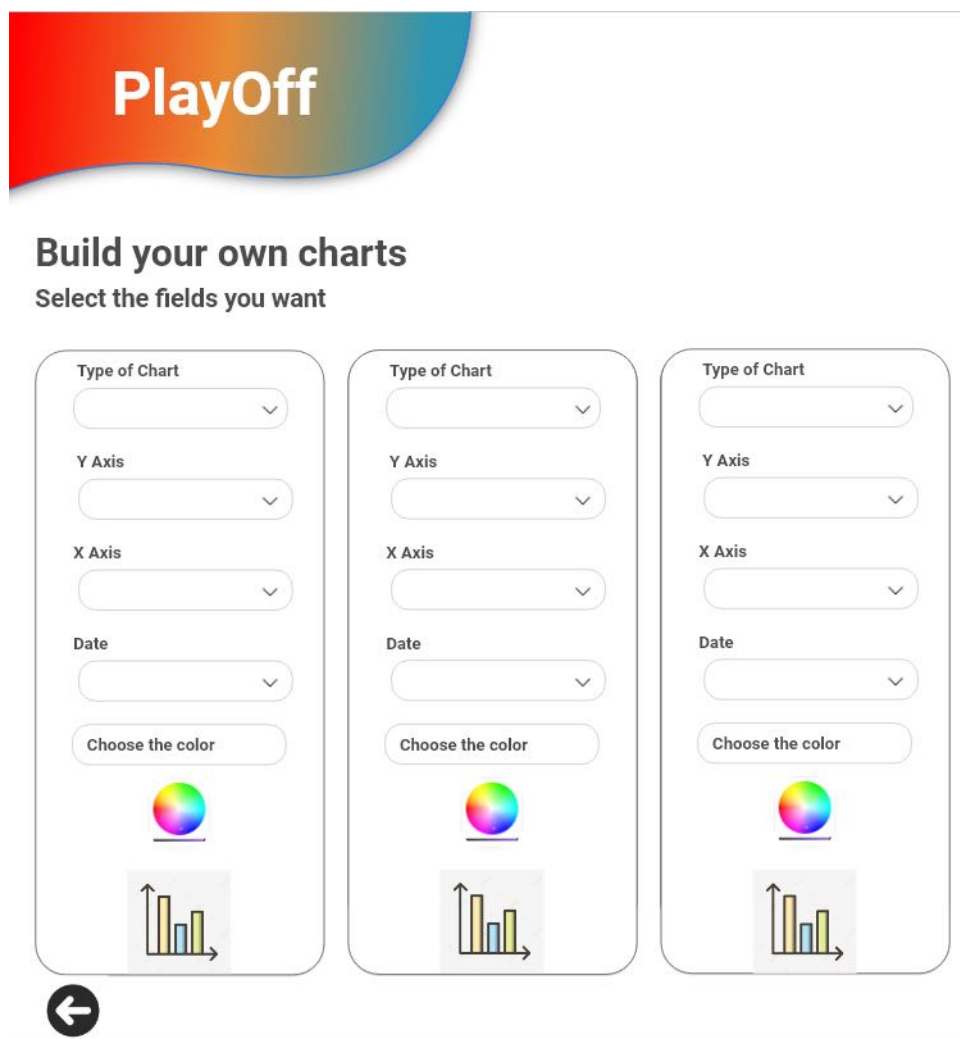


Figura 67: Mockup da Página da Dashboard

## 5.2 Desenvolvimento

Como descrito anteriormente, foi construída uma página web que recebe tweets em streaming diretamente da API disponibilizada pelo Twitter. O propósito deste trabalho passa por facultar ao utilizador dados que ocorram durante um determinado evento desportivo em tempo real para que seja possível extrair informações relevante sobre o mesmo que, muitas das vezes, não se tem a noção dentro do estádio.

### 5.2.1 Página Principal

Nesta página o utilizador é capaz de inserir o texto de pesquisa numa caixa de texto (por exemplo, uma hashtag para um determinado tópico ou acontecimento), onde vão ser devolvidos todos os tweets que contêm essa mesma palavra. Para além disso, também permite inserir o perfil do Twitter que recolhe

todos os *tweets* publicados pelo mesmo. Decorridos 5 minutos e se, não existirem dados conforme o que o utilizador pesquisou, é aberta uma janela *pop up* que questiona o utilizador se ele pretende cancelar a sua pesquisa ou continuar a aguardar que surjam novos dados.

Começando a gerar dados, na página é possível observar o total de *tweets* recolhidos assim como, o total de *tweets* negativos, neutros e positivos. Em cada *tweet* é utilizada uma metodologia que através do texto, classifica com um destes três tipos de sentimentos. Para além disso, é fornecido também ao utilizador botões de "*start and stop*" que permite que utilizador comece a sua pesquisa, assim como pausar a mesma. O utilizador também tem a opção de clicar no ícon *plus* para ser redirecionado para a página da *dashboard*.

Para uma boa exposição de dados, optou-se por dividir a página numa matriz 3/2 (3 colunas por 2 linhas) em que cada uma delas representa o tipo de gráfico utilizado (visível na Figura 68). Este tipo de *design* facilita uma boa absorção de informação, e ao mesmo tempo dota a plataforma de um *layout* prático e responsivo. Relativamente ao tipo de gráficos, foram seleccionados aqueles que, geralmente facilitam a transmissão da informação contida e consoante o tipo de dados disponíveis, transpondo isto para temas desportivos. Como referido no capítulo anterior, na secção *Data Visualization*, foram definidos os tipos de variáveis a utilizar sendo estes valores quantitativos ou qualitativos. Para além disso, foi também elaborada uma tabela disponível nos Anexos A, B e C onde é feito um estudo detalhado de cada tipo de gráfico e quais as variáveis que melhor se adequam para cada um. Posto isto e tendo em conta todos os estudos feitos, foram seleccionados um total de sete gráficos:

- **Gráfico Circular e Radial** - Valores quantitativos (percentagem) por cada *tweet* - Proporções de quantidades por categorias;
- **Gráfico de Barras** - Valores quantitativos por cada *tweet*, ou seja, por diferentes items de categorias;
- **Gráfico de Área** - Valores quantitativos por cada *tweet* que mudam ao longo do tempo;
- **Gráfico de Dispersão** - Coleção de pontos quantitativos por cada *tweet* - Mostra a relação entre variáveis quantitativas para diferentes items de categoria.



Figura 68: Gráficos utilizados da biblioteca ApexCharts

Na primeira linha e primeira célula da matriz utilizou-se um gráfico circular onde é possível observar a percentagem do idioma de cada *tweet*, conseguindo assim tirar conclusões sobre que língua debate mais sobre um determinado tema. Nos segundos gráficos, representados na segunda célula, o utilizador tem a opção de escolher entre um gráfico de barras ou de área através de um botão *switch* e é capaz de observar o valor do sentimento ao longo do tempo, dividido numa escala de cores para dessa forma ser mais perceptível ver momentos em que o sentimento é mais negativo ou positivo. Como complemento à informação disponibilizada, é permitido ao utilizador ver a informação detalhada sobre esse mesmo *tweet* (texto, imagem do utilizador que o publicou e valor do sentimento), assim como, ser redirecionado para a página do *tweet*. Por último, o gráfico da terceira célula é também circular e apresenta de forma geral e compreensível a percentagem de *tweets* negativos, positivos e neutros. Seguidamente na segunda linha da matriz e visível na primeira célula, é apresentado outro gráfico onde se visualiza o género dos utilizadores, estando este dividido em três – sem género, masculino e feminino - permitindo tirar conclusões sobre que tipo de género fala mais sobre um determinado tema. Do lado direito junta-se um gráfico de barras que representa o número de *retweets* de cada *tweet*, estando estes organizados por ordem crescente, conseguindo perceber facilmente qual deles tem um maior número de *retweets* naquele momento. Este gráfico apresenta a mesma funcionalidade que o de cima, sendo possível ver mais detalhadamente cada *tweet*. Por último a nível de gráficos e demonstrado na terceira célula, inseriu-se um gráfico de dispersão que representa os dados como uma coleção, conseguindo assim atingir mais facilmente em que momento existe um aumento/crescimento de cada tipo de sentimento de cada *tweet* ao longo do tempo. De notar que em todos os gráficos, é permitido ao utilizador ter acesso:



- Texto do *tweet*;
- Imagem do utilizador que publicou o *tweet*;
- Valor que estiver a ser representado naquele gráfico, como número de *retweets*, valor do sentimento, entre outros;
- É permitido ao utilizador clicar em cima do ponto do gráfico referente a um *tweet* e será encaminhado para a página do *Twitter* do respetivo.

Para além dos gráficos, nesta mesma página é também possível configurar um painel que mostra as tendências de um dado país. Neste exemplo em específico da dissertação, o mesmo está configurado para Portugal e mostra os 10 temas mais ativos do momento.

Por fim, é apresentado ao utilizador um *widget* (elemento de interação) com um *slideshow* onde o utilizador é capaz de contemplar o top 3 de *tweets* negativos ou positivos naquele momento. Aqui também é permitido ao utilizador clicar e o mesmo vai ser redirecionado para o *tweet* correspondente na página oficial do *Twitter*. Nas figuras 69 e 70 é possível observar o resultado e aspeto final da página.

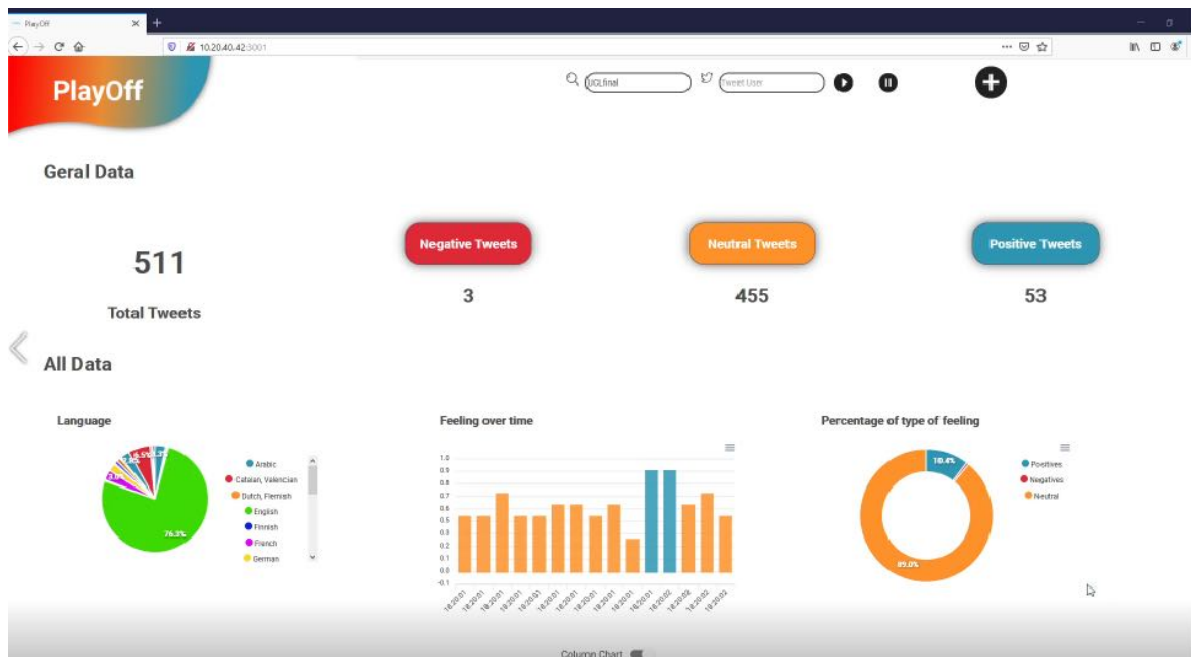


Figura 69: 1º Parte-Página Principal

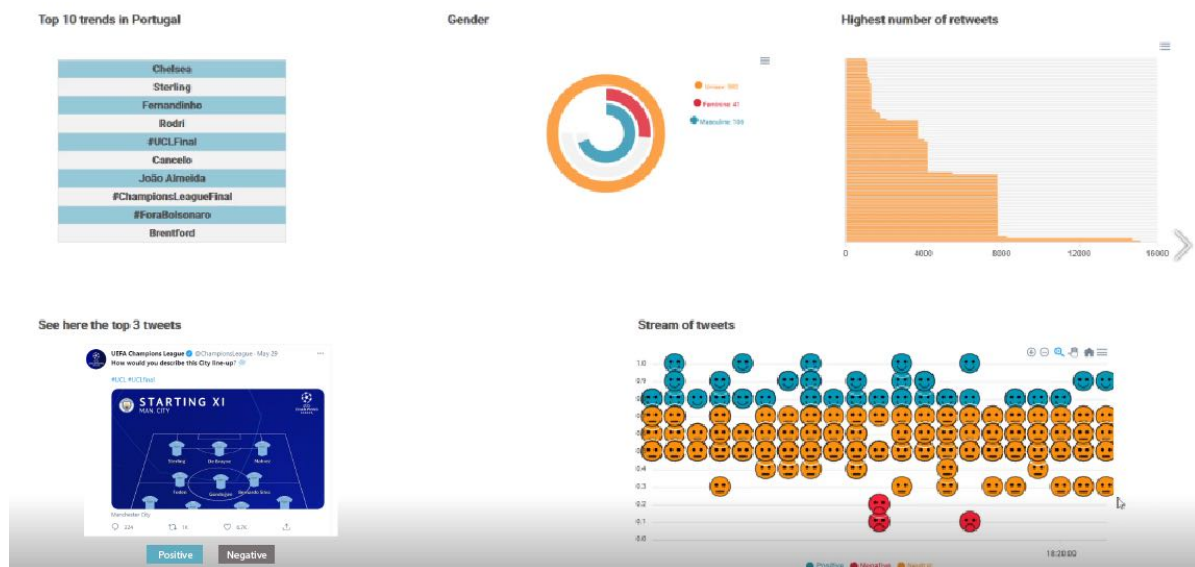


Figura 70: 2º Parte-Página Principal

## 5.2.2 Página Dashboard

Esta página foi construída de forma a fornecer ao utilizador oportunidade de construir os seus próprios gráficos e relatórios com os dados que o mesmo pesquisou na página principal. Para isso e para uma boa prática de *User Interface* e *User Experience* foram adicionadas 3 *grids* às quais o utilizador pode dar uso (ver na Figura 71). Como é visível na Figura 72 estas *grids* são responsivas e permite que o utilizador as arraste no ecrã dinamicamente, assim como aumentar e diminuir o seu tamanho. Isto permite ao utilizador "manipular" a sua própria visualização dos dados assim como, organizá-los de diferentes formas e perspectivas e até comparar os gráficos de diferentes posições.

Para a construção do seu gráfico, o utilizador tem vários campos que necessita de escolher:

- **Tipo de gráfico** - Existem 3 tipos de gráficos que podem ser escolhidos: barras, linhas e área;
- **Data** - O utilizador tem a possibilidade de num calendário, seleccionar a data do qual pretende que sejam exibidos os dados, até no máximo de 8 dias. É permitido ao utilizador inserir apenas uma data ou um intervalo de datas no calendário;
- **Eixo do x** - É disponibilizada uma lista de dados que podem ser utilizados para o eixo do X do gráfico, como por exemplo: valor do sentimento e nº de *retweets*;
- **Eixo do y** - É disponibilizada uma lista de dados que podem ser utilizados para o eixo do Y do gráfico, como por exemplo: data (hora, minuto e segundo) e o género do utilizador;
- **Cor do gráfico** - Por último, o utilizador tem a opção de seleccionar a cor pretendida para o gráfico.

**PlayOff**

**Build your own charts**  
Select the fields you want

🏠

**Type of Chart**  
Select... | ▾

**X Axis**  
Select... | ▾

**Y Axis**  
Select... | ▾

**Date**  
Select... | ▾

Choose the color of the chart

Figura 71: Página da *dashboard*

**Type of Chart**  
Area | ▾

**X Axis**  
Gender | ▾

**Y Axis**  
Retweets | ▾

**Date**  
Interval between two dates | ▾

Start Date → End Date

**Type of Chart**  
Select... | ▾

**X Axis**  
Select... | ▾

**Y Axis**  
Select... | ▾

**Date**  
Only one date | ▾

Date

Done

Figura 72: *Grids* na Página da *dashboard*

Como mostrado na Figura 73 e dito anteriormente, é facultado ao utilizador um formulário com vários campos para o mesmo selecionar. No campo da data, é apresentado ao utilizador um *DatePicker* em

que o mesmo seleciona a data pretendida ou então um intervalo de datas.

Depois de todos os campos preenchidos, basta apenas clicar no botão "Done" e o gráfico revela-se. De notar, que todos os campos necessitam de ser preenchidos pois caso isso não aconteça, surgirá uma mensagem nos campos em falta a lembrar que é necessário preencher.

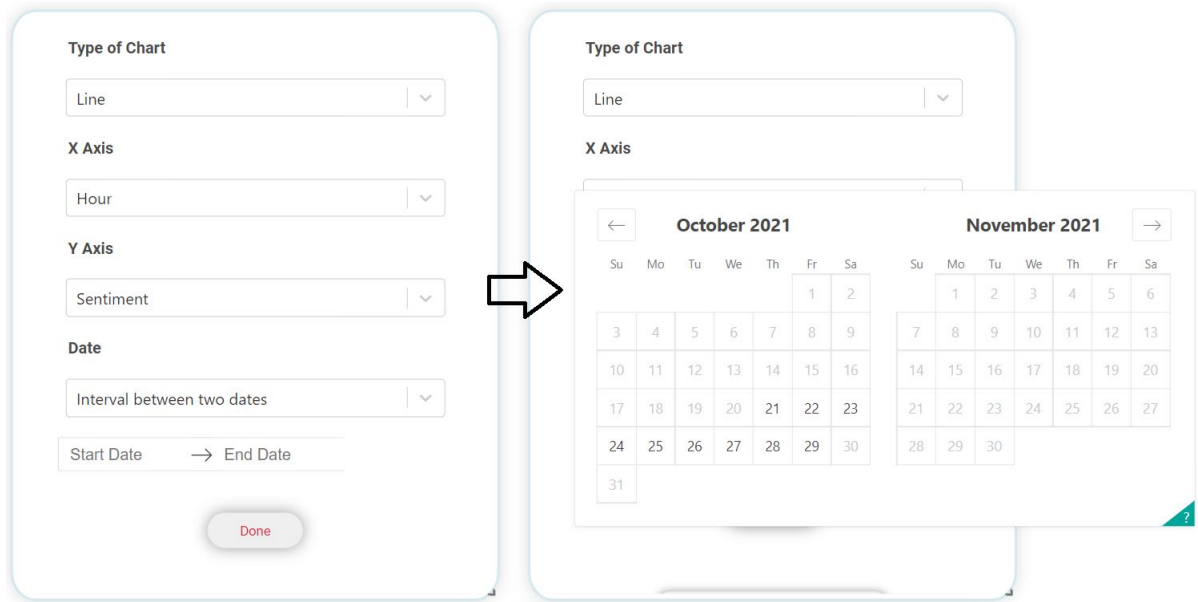


Figura 73: Campos para o utilizador preencher

Assim que o gráfico se revela, depois de selecionados todos os campos pretendidos, o utilizador pode alterar a cor do gráfico através de um *ColorPicker* e fazer *download* do mesmo para uma imagem (visível na Figura 74).

Como referido no Capítulo 4, foram utilizadas metodologias de *design* interativo para dar uma boa experiência ao utilizador. Optou-se por tanto na página principal como na da *dashboard* o utilizador usufruir da interatividade com os gráficos e com a aplicação, como o caso das *grids* em que o utilizador consegue manipulá-las e ajustá-las ao seu gosto. Isto pretende melhorar a experiência do utilizador, fornecendo funcionalidades de uma maneira acessível e esteticamente agradável. A paleta de cores também foi escolhida e utilizada propositadamente para enfatizar neste caso, o tipo de sentimento dos *tweets*, proporcionando assim uma visão geral mais apelativa. Com isto, a biblioteca *JavaScript Reactjs* também foi a escolhida por ser possível criar *UI's* interativas e complexas, permitindo encapsular e criar componentes com estado próprio, fácil de desenvolver, manter e reaproveitar. Com a composição dos componentes, é possível criar vistas, que de forma eficiente, são atualizadas, sempre que o seu estado altera. Possui várias bibliotecas, como é o caso da *Material-UI* que fornece componentes e animações modernas e paralelamente intuitivas. Estas são *componente-based*, escritas em cima do *React*, o que permite criar vistas interactivas e apelativas na aplicação.

Em resumo, foi elaborada uma interface de modo a que os utilizadores usufruam da sua interatividade e que a informação seja transmitida através de boas práticas visuais. Com isto, está assente uma página

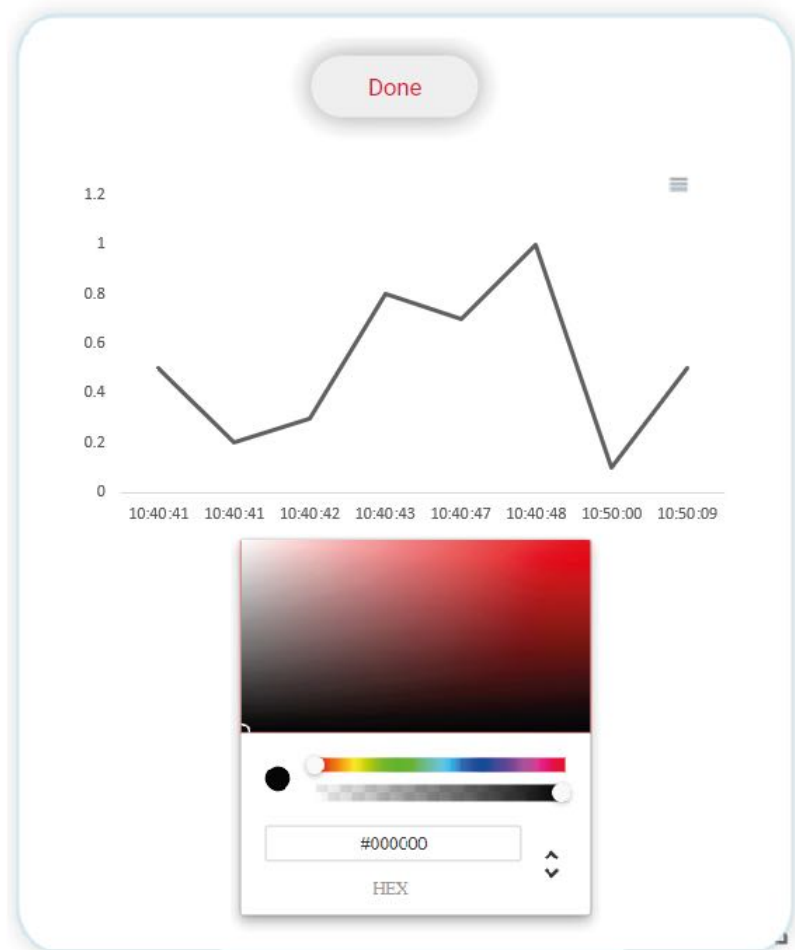


Figura 74: Gráfico na página da *dashboard*

que disponibiliza um total de sete gráficos com informações sobre os *tweets* que representam as opiniões dos adeptos. Na outra página, é apresentado também um sistema de *grids* em que os utilizadores podem criar/construir os seus próprios relatórios com as informações dos *tweets*, dando a possibilidade de os mesmos selecionarem os tipos de dados que pretendem. Estes tipos de funcionalidades tornam este projeto mais intuitivo e enriquecedor.

## Resultados

Nesta secção vão ser verificados e discutidos os resultados que se obtiveram com este projeto, percebendo se se enquadram nos objetivos pretendidos inicialmente para o mesmo. Vai ser também exibida uma análise de *SWOT*, funcionando como um sistema simples de análise.

### 6.1 Análise de Resultados

Esta dissertação nasce para ajudar a validar e materializar algumas ideias presentes no projeto *PlayOff*. O trabalho desenvolvido serve assim como um protótipo, passando por avaliar a sua real aplicabilidade no mundo do desporto.

Realizada a parte técnica do projeto, foi altura da realização de testes e validação da solução. Sendo que o objetivo principal passava por abordar opiniões desportivas, optou-se por fazer um teste com a final da Liga dos Campeões 2021, que teve lugar no dia 29 de maio de 2021, entre o clube *Chelsea* e o *Manchester City*, com início às 20:00 horas em Portugal Continental. A recolha de dados iniciou-se um pouco antes do início da partida e terminou-se um pouco depois do apito final do jogo. Durante 2 horas e 7 minutos, foram recolhidos em média 600 *tweets* por minuto. Para obter apenas informação relacionada com o evento desportivo, aplicou-se a *hashtag* *#UCLFinal* como termo de pesquisa. Esta *hashtag* foi escolhida e promovida pelos promotores oficiais do evento desportivo, neste caso a *UEFA*. Minutos antes do jogo e iniciada a partida, tentou-se perceber o que estavam a sentir os adeptos face a esta final e como resultado obtido entendeu-se, através das opiniões recolhidas, que o sentimento predominante era o neutro, que significa não existia grande entusiasmo, mas por outro lado, era visível vários *tweets* de apoio dos utilizadores para com os seus clubes. Pelo que se vê na Figura 75, os sentimentos negativos eram pouquíssimos, contrariamente aos positivos. Para além disso, muitos adeptos também faziam referência ao jogador *N'Golo Kanté*, dizendo que o mesmo seria um homem para fazer a diferença neste jogo, apostando várias expectativas no mesmo como é possível observar no *slideshow* da Figura 76 e nos *tweets* detalhados na Figura 77.



Figura 75: 1º Parte- Antes do jogo

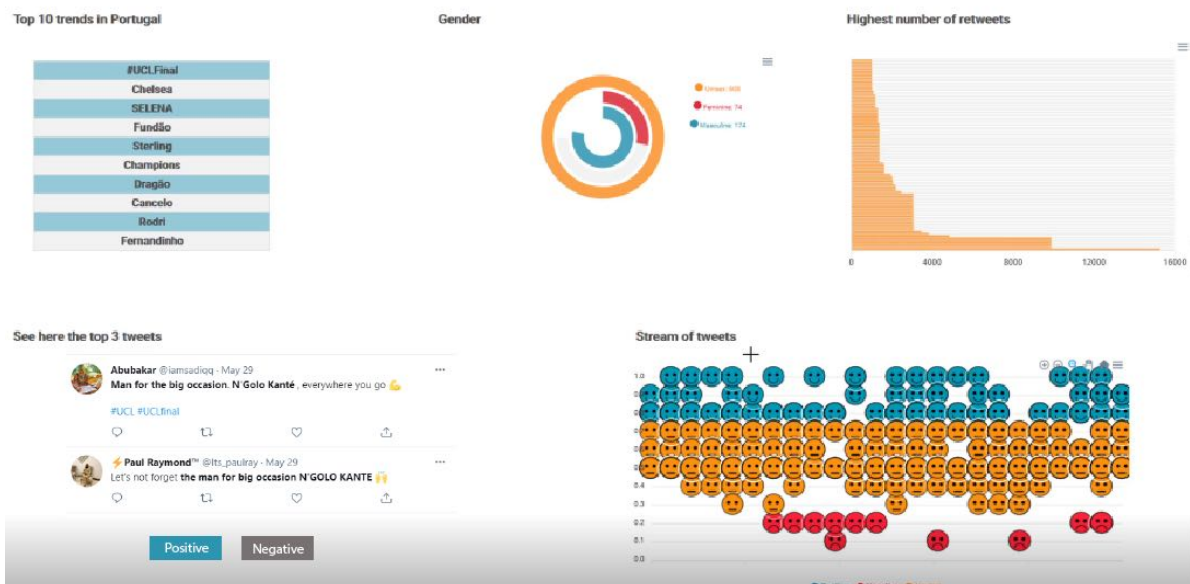


Figura 76: 2º Parte- Antes do jogo



Figura 77: Tweets do jogador N'Golo Kanté

Ao minuto 15, é possível observar que o número de tweets negativos aumentou para 6.2%. Comparado com os acontecimentos do jogo, percebeu-se pelos textos dos tweets que Werner, jogador do Chelsea, falhou duas oportunidades de golo o que se transformou numa grande revolta por parte dos adeptos. No Twitter foram vários os comentários que mostravam decepção com o jogador. Este conjunto de situações, que criou um aumento de tweets negativos, levou aos adeptos usarem comentários como “Werner é tão mau”, demonstrando assim críticas à exibição do jogador (ver nas Figuras 78 e 79).



Figura 78: Resultados ao minuto 15' do jogo





Figura 79: Tweets do jogador Werner

Depois de um conjunto de comentários a reclamar sobre as falhas do clube *Chelsea*, a quantidade de *tweets* positivos começou a aumentar significativamente. Isto relaciona-se com uma fase mais positiva do *Chelsea* e ao golo obtido. O momento do golo levou a um conjunto de *tweets* (na Figura 80) que refletiam a alegria por parte dos adeptos que comentavam a “incrível tática” do jogador *Kai Havertz*, marcador do golo, e elogiando o treinador do clube, *Tuchel*. Às 21 horas e 37 minutos, como se pode observar na Figura 81, iniciou-se a segunda parte do jogo e os *tweets* com sentimento positivo continuavam a aumentar passando para os 39,7%, com os *tweets* a refletirem felicidade pelo *Chelsea* estar em vantagem.

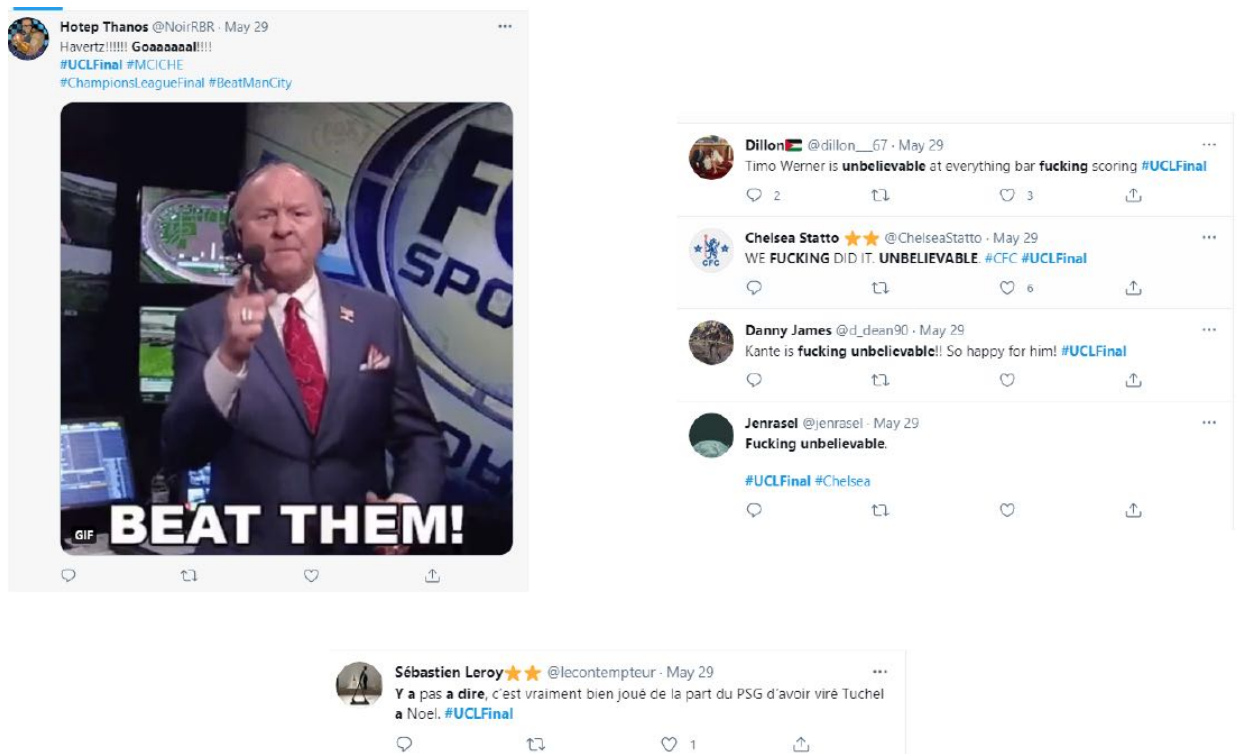


Figura 80: Tweets do jogador Kai Havertz

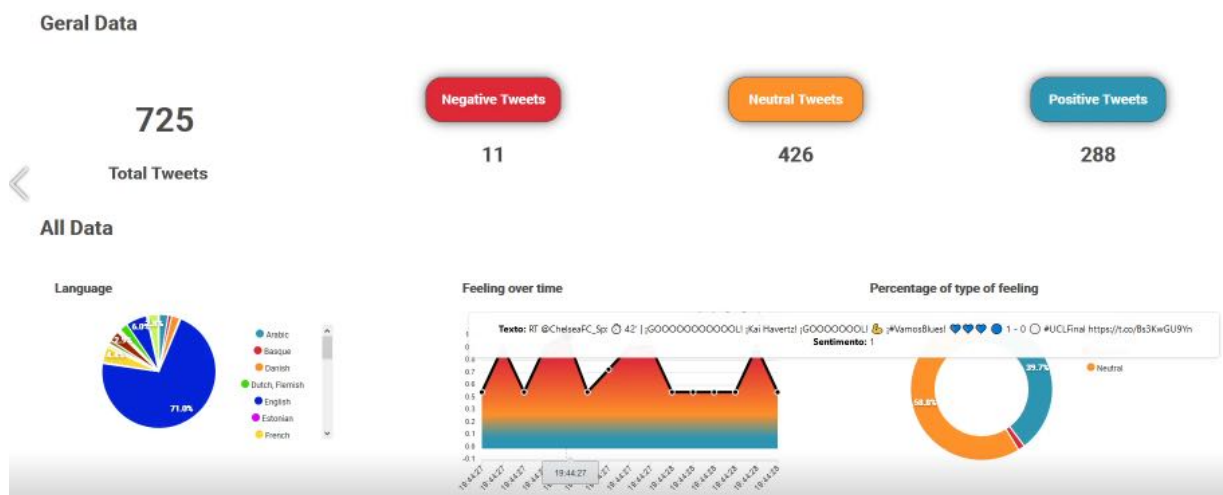


Figura 81: Resultados no início da segunda parte do jogo

Ao minuto 56', volta a acontecer um novo episódio que levou a uma subida no número de tweets negativos que coincidiu com um choque entre dois jogadores, terminando com uma lesão grave para o jogador Kevin de Bruyne que é substituído. Nas Figuras 82 e 83 consegue-se compreender que este foi um momento de tristeza para os adeptos que reagem no Twitter dizendo que “é uma tristeza acabar

a época desta forma” e “o futebol pode ser cruel”. Para além disso, o nome “*Bruyne*” passou para o número quatro do top das *trends* em Portugal.



Figura 82: Resultados ao minuto 56' do jogo



Figura 83: Resultados ao minuto 56' do jogo

A análise deste jogo continuou e eram vários os *tweets* dos utilizadores que continuavam a demonstrar que acreditavam na vitória do *Chelsea* e a elogiar a equipa defensivamente, mas por outro lado, diziam que o jogo estaria tenso. Para muitos, o jogo está demasiado aborrecido, outros criticam o treinador do *Manchester City* dizendo que ele não é digno de estar numa final na Liga dos Campeões. Os comentários dividem-se e alguns adeptos chegam mesmo a dizer que o treinador do *Chelsea* não tem capacidades para ensinar um bom estilo de futebol, como é possível observar na Figura 84.

No final da partida, vê-se na Figura 85 que o número de *tweets* positivos é evidentemente superior ao de negativos. Muitos dos comentários positivos publicados referiam-se ao jogador *Kanté* do *Chelsea*,

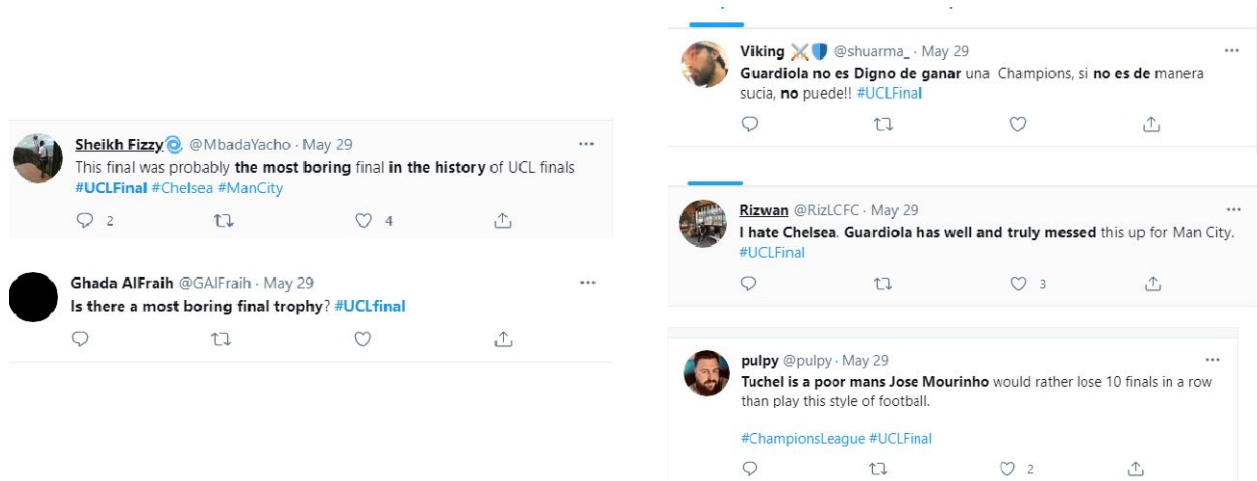


Figura 84: Tweets sobre o jogo

indicando como o mesmo quebra padrões no futebol e que coincide com a atribuição do galardão recebido “Homem do jogo “da partida. Estes tweets podem ser visualizados na Figura 86.

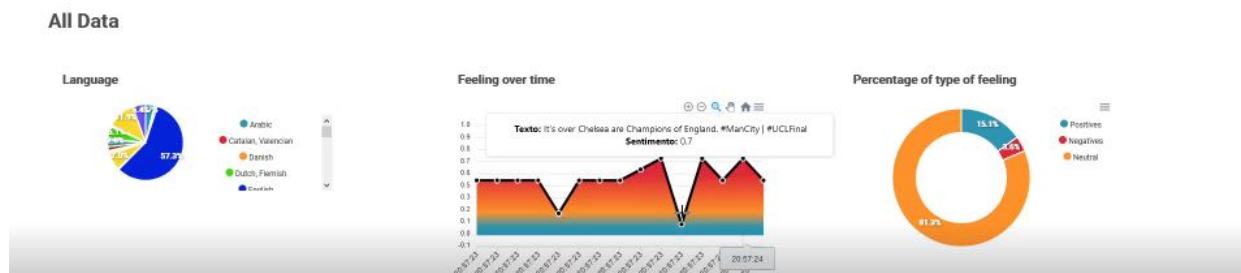


Figura 85: Resultado no final do jogo



Figura 86: Tweets sobre o jogador Kanté

O grande propósito dos testes realizados é perceber se este projeto é fácil de digerir e interpretar pelas equipas, direções de clubes, equipas de *marketing*, entre outras. Dependendo dos recursos que o jogo e a equipa têm disponíveis, tomar decisões baseadas nas informações das redes sociais.

Com isto, as opiniões dos utilizadores nas redes sociais são bastante importantes de forma aos clubes perceberem o que podem fazer para melhorar as relações com os seus adeptos e até mesmo de maneira a captar informações que os mesmos desconhecem. Neste jogo, principalmente na segunda parte, existiram *tweets* a indicar que o jogo estava aborrecido. Isto pode ser útil para uma equipa de *marketing* criar ações que melhorem a experiência de quem assiste o jogo pela televisão. Outros casos, como ser possível extrair quais os jogadores que recebem mais elogios durante uma partida, pode potenciar campanhas publicitárias em conjunto com outras marcas. Tendo isso em conta, é também possível, por exemplo, evidenciar algum plano no campo onde o jogador apareça, de forma a potenciar o impacto da campanha.

Por outro lado, assimilar quais os jogadores ou treinadores que sofrem mais críticas também são pontos indispensáveis para os clubes melhorarem esse tipo de opiniões, e desta forma aumentar o valor dos seus ativos.

## 6.2 Análise de SWOT

A análise de *SWOT* que significa as forças, fraquezas, oportunidades e ameaças. Esta análise é amplamente utilizada para análise de projetos. Como dito anteriormente, divide-se em 4 etapas:

- **Forças:** Aqui verifica-se os fatores positivos do projeto em relação aos existentes no mercado;
- **Fraquezas:** São delineados os pontos do projeto que precisam de ser melhorados;

- **Oportunidades:** As oportunidades são o resultado das forças e fraquezas existentes somadas a quaisquer iniciativas externas que colocam o projeto numa melhor posição competitiva;
- **Ameaças:** Deve-se prever e analisar onde os principais concorrentes se conseguem sobressair.

Com isto, foi elaborada a seguinte análise de *SWOT* de forma a realizar um diagnóstico a este projeto, como é possível observar na Figura 87.

<p style="text-align: center;"><b>Forças</b></p> <ul style="list-style-type: none"> <li>- Plataforma interativa e com um bom <i>design</i>;</li> <li>- O utilizador tem disponível várias informações em diferentes gráficos;</li> <li>- O utilizador pode customizar os seus próprios gráficos e fazer <i>download</i> dos mesmos;</li> <li>- Sistema <i>web</i>, agnóstico ao depósito do utilizador;</li> <li>- Pouca concorrência ou inexistente neste mercado.</li> </ul>	<p style="text-align: center;"><b>Fraquezas</b></p> <ul style="list-style-type: none"> <li>- Dependente de fontes de dados externas;</li> <li>- Limitado aos dados existentes nas redes sociais;</li> <li>- Qualidade e veracidade dos dados;</li> <li>- Assente em sistemas abertos e públicos;</li> <li>- Muito difícil de proteger a propriedade intelectual da solução;</li> <li>- Mercado muito pequeno e limitado a médios e grandes clubes e associações desportivas.</li> </ul>
<p style="text-align: center;"><b>Oportunidades</b></p> <ul style="list-style-type: none"> <li>- Valorizar <i>assets</i> estratégicos dos clubes desportivos nas transmissões televisivas;</li> <li>- Campanhas de <i>marketing</i> mais assertivas e orientadas ao que o público procura;</li> <li>- Valorizar/medir parcerias e colaborações durante a transmissão de um jogo;</li> <li>- Observar e interpretar o ambiente nas redes sociais da mesma forma que se faz nos recintos desportivos.</li> </ul>	<p style="text-align: center;"><b>Ameaças</b></p> <ul style="list-style-type: none"> <li>- Cópia da solução por outras empresas;</li> <li>- Concorrência de um futuro projeto aberto e suportado pela comunidade sem custos para os clubes;</li> <li>- Incoerência nos dados recolhidos;</li> <li>- Manipulação dos dados via sistemas automatizados (<i>bots</i>);</li> <li>- Atualização/remoção das <i>API's</i> das redes sociais.</li> </ul>

Figura 87: Análise de *SWOT*

## Conclusão e Trabalho Futuro

Dada a questão de investigação inicial, ficou respondido e provado que é possível construir uma plataforma online que permita a clubes e estações de televisão visualizar os dados recolhidos de redes sociais sobre um determinado evento desportivo melhorando assim a interação com os seguidores e tendo em conta boas práticas de *design*. Todos os objetivos foram delineados, estudados e trabalhados, como é possível observar durante a escrita desta dissertação. Este trabalho permite criar uma ponte no sentido inverso, isto é, dos telespectadores para os produtores do evento, de forma a estes obterem *feedback* sobre as ações que escolhem para a transmissão televisiva, assim como, para os representantes e responsáveis de decisão dos clubes desportivos.

A arquitetura proposta na dissertação foi aplicada e os resultados obtidos do protótipo foram ao encontro do objetivo. O sucesso da arquitetura deve-se à sua modularidade e robustez do processamento dos dados (desde o angariamento dos dados do *Twitter* até à exibição dos dados tratados na interface gráfica). A interface gráfica, por sua vez demonstrou a sua versatilidade em exibir os resultados de forma contínua mesmo em momentos de grande fluxo de dados. Também a sua capacidade de customizar gráficos e tratar dados após o evento desportivo permite criar relatórios personalizados para posterior discussão/análise dos mesmos.

O grande objetivo deste projeto era demonstrar que é possível extrair informação relevante, em tempo real das redes sociais e expô-los de formas adequadas e esse objetivo foi cumprido. Possibilita uma exploração mais poderosa da informação contida nos *tweets* de diferentes utilizadores, combinando esta com uma análise automatizada de sentimento, potenciando assim variados tipos de visualização.

O *Twitter* pode ser uma ferramenta extremamente valiosa para estações de televisão, agências e patrocinadores que procuram compreender melhor o público e este trabalho mostrou claramente como o contacto com os adeptos deve ser aproveitado e estudado. Assim, a interação dos utilizadores deve ser importante para qualquer empresa, uma vez que serve para assistir tomadas de decisão e de certa forma, engendrar maneiras de aproximar ainda mais os adeptos dos seus clubes. Para além desta dissertação fornecer variada informação disponível de diferente forma aos utilizadores, permite também que o mesmo construa os seus próprios gráficos. Isto é bastante importante para que os utilizadores conseguiram estudar e comparar diferentes acontecimentos em diferentes datas. Em resumo os objetivos pretendidos



foram concluídos e aprovados pela empresa tendo sido criada uma plataforma que permite a clubes desportivos e estações de televisão visualizar os dados recolhidos de redes sociais sobre um determinado evento desportivo, tendo em consideração boas práticas de *User Experience* e *User Interface*.

Um dos grandes pontos compreendidos e observados com este trabalho, é a reação do público que se expressa nas redes sociais, num curto espaço de tempo, sobre os acontecimentos dentro de campo. Este facto é evidenciado quando acontecem momentos chave, como o exemplo do golo ou da lesão de um jogador, as pessoas reagem naturalmente a esses tipos de estímulos obtidos na transmissão televisiva. Outro facto interessante é que momentos antes do jogo, não estando a transmissão televisiva ativa, o público já demonstra atividade nas suas opiniões.

Como trabalho futuro, pretende-se integrar outras redes sociais, como o *Instagram* e *Facebook*, que seriam também muito relevantes para reforçar a extração de opiniões e tendências dos utilizadores. Para além disso, uma funcionalidade que interessaria bastante realizar seria a análise de sentimentos em emoções. Neste momento os sentimentos apenas estão divididos em três tipos: negativo, positivo e neutro; e um objetivo seria passar essas emoções para estados de espírito, como por exemplo, nervoso, ansioso, com dúvidas, entre outros.

Para além disso, apesar de esta dissertação ser um protótipo para validar ideias, seria também interessante perceber se o mesmo está alinhado com o público alvo. Desta forma, iria ser entendido se as informações disponíveis assim como as práticas de *UX* e *UI*, estão elaboradas de maneira perceptível e enriquecedora para clubes desportivos.

Em suma, o trabalho obtido nesta dissertação é relevante e será adotado pelo projeto nacional *PlayOff* que terminará em 2024. O projeto *PlayOff* aborda mais temas que não foram contemplados nesta dissertação mas está no plano atual endereçar alguns dos pontos apresentados no trabalho futuro, nomeadamente, integração de pelo menos mais uma rede social e avaliação da solução por profissionais da área desportiva e televisiva.



## Bibliografia

- [1] 5 gráficos que explican el boom del Big Data. <https://www.businessinsider.es/5-graficos-explican-boom-big-data-266741>. Acedido em: 2021-01-11 (ver p. 58).
- [2] 9 Types of Graphic Design Your Team Needs to Know About. <https://learn.g2.com/types-of-graphic-design>. Acedido em: 2021-01-16 (ver pp. 118–120).
- [3] L. J. W. D. Z. K. A. “[Lecture Notes in Computer Science] Algorithmics of Large and Complex Networks Volume 5515 | | Group-Level Analysis and Visualization of Social Networks”. Em: 10.1007/978-3-642-02094-0 (Chapter 16 2009). doi: [10.1007/978-3-642-02094-0\\_16](https://doi.org/10.1007/978-3-642-02094-0_16) (ver p. 10).
- [4] AmCharts Review. <https://stackshare.io/amcharts> (ver p. 47).
- [5] AP Exata - Data Driven Strategy. <https://www.agenciaexata.com/>. Acedido em: 2021-02-10 (ver p. 11).
- [6] APEXCHARTS.JS - Modern Interactive Open-source Charts. <https://apexcharts.com/> (ver p. 47).
- [7] L. H. J. B. M. A. A. S. B. “Decoding Twitter”. Em: *Annals of Surgery* 264 (6 dez. de 2016). doi: [10.1097/sla.0000000000001824](https://doi.org/10.1097/sla.0000000000001824) (ver p. 10).
- [8] B. Barnhart. *How to use social media in sports to keep fans engaged*. <https://sproutsocial.com/insights/social-media-in-sports/>. Acedido em: 2021-01-08. Ago. de 2020 (ver p. 8).
- [9] D. Basole Rahul C.; Saupe. “Sports Data Visualization [Guest editors’ introduction]”. Em: *IEEE Computer Graphics and Applications* 36 (5 set. de 2016). doi: [10.1109/mcg.2016.85](https://doi.org/10.1109/mcg.2016.85) (ver p. 28).
- [10] A. Y. C. K. H. F. K. I. O. Benoît. “Community extraction and visualization in social networks applied to Twitter”. Em: *Information Sciences* (set. de 2017). doi: [10.1016/j.ins.2017.09.022](https://doi.org/10.1016/j.ins.2017.09.022) (ver p. 8).
- [11] P. Bradley et al. “The reliability of ProZone MatchViewer: a video-based technical performance analysis system”. Em: (2017). url: <https://doi.org/10.1080/24748668.2007.11868415> (ver p. 26).

- [12] B. ( Chae. “Insights from hashtag #supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research”. Em: (2015). Acedido em: 2021-01-18. doi: [10.1016/j.ijpe.2014.12.037](https://doi.org/10.1016/j.ijpe.2014.12.037). url: <https://www.sciencedirect.com/science/article/abs/pii/S0925527314004319?via%5C%3Dihub> (ver p. 35).
- [13] R. Chan. *The 10 most popular programming languages, according to the Microsoft-owned GitHub*. <https://www.businessinsider.com/most-popular-programming-languages-github-2019-11>. Acedido em: 2021-01-15. Nov. de 2019 (ver p. 50).
- [14] S. Chang. “Instagram Post Data Analysis”. Em: (2016), p. 5. doi: [arXiv:1610.02445v1](https://arxiv.org/abs/1610.02445v1). url: <https://arxiv.org/pdf/1610.02445.pdf> (ver pp. 18–20).
- [15] *Chart.js Review*. <https://www.slant.co/options/10578/chart-js-review> (ver p. 47).
- [16] *Chartist.js Review*. <https://www.slant.co/options/10579/chartist-js-review> (ver p. 47).
- [17] *Comparing Database Management Systems: MySQL, PostgreSQL, MSSQL Server, MongoDB, Elasticsearch, and others*. <https://www.altexsoft.com/blog/business/comparing-database-management-systems-mysql-postgresql-mssql-server-mongodb-elasticsearch-and-others/>. Acedido em: 2020-12-14 (ver p. 52).
- [18] *Comparing MongoDB vs PostgreSQL*. <https://www.mongodb.com/compare/mongodb-postgresql>. Acedido em: 2020-12-14 (ver p. 52).
- [19] *Comparison table*. <https://github.com/naver/billboard.js/wiki/Comparison-table> (ver p. 47).
- [20] *Criar um gráfico de mapa de árvore em Office*. <https://support.microsoft.com/pt-br/office/criar-um-gr%C3%A1fico-de-mapa-de-%C3%A1rvore-em-office-dfe86d28-a610-4ef5-9b30-362d5c624b68> (ver p. 64).
- [21] Z. S. Data. *Data-Driven Design: o que é, como funciona e benefícios para negócios*. <https://zooxsmart.com/pt/data-driven-design-o-que-e-beneficios/>. Acedido em: 2020-12-15. 2020 (ver p. 54).
- [22] *Data Visualization: What It Is, Why It's Important & How to Use It for SEO*. <https://www.searchenginejournal.com/what-is-data-visualization-why-important-seo/288127/>. Acedido em: 2021-01-20 (ver pp. 39, 43).
- [23] *Django vs Laravel vs Node.js*. <https://medium.com/@manant.letsnurture/django-vs-laravel-vs-node-js-278c84c2cd47>. Acedido em: 2020-12-18 (ver p. 50).
- [24] O. Edo-Osagie et al. “A scoping review of the use of Twitter for public health research”. Em: (2020). Acedido em: 2021-01-18. doi: [10.1016/j.combiomed.2020.103770](https://doi.org/10.1016/j.combiomed.2020.103770). url: <https://www.sciencedirect.com/science/article/pii/S0010482520301426> (ver pp. 33, 34).

- [25] F. C. Filho e L. A. Santos. “Potencialidades e limitações da metodologia de análise de rede: um modelo teórico voltado para as Ciências Sociais”. pt. Em: *Comunicação e Sociedade* 33 (jun. de 2018), pp. 183–198. issn: 2183-3575. url: [http://www.scielo.mec.pt/scielo.php?script=sci\\_arttext&pid=S2183-35752018000100006&nrm=iso](http://www.scielo.mec.pt/scielo.php?script=sci_arttext&pid=S2183-35752018000100006&nrm=iso) (ver p. 8).
- [26] Florence Ying Wang et al. “SentiCompass: Interactive visualization for exploring and comparing the sentiments of time-varying twitter data”. Em: (2015), pp. 129–133. doi: [10.1109/PACIFICVIS.2015.7156368](https://doi.org/10.1109/PACIFICVIS.2015.7156368) (ver p. 11).
- [27] *Football Manager Video Games*. <https://www.footballmanager.com/>. Acedido em: 2021-01-20 (ver p. 26).
- [28] *Front-end frameworks popularity (React, Vue, Angular and Svelte)*. <https://gist.github.com/tkrotoff/b1caa4c3a185629299ec234d2314e190> (ver p. 45).
- [29] J. P. C. Gaspar Bogueira Fernando Batista. “Sistema Inteligente de Recolha, Armazenamento e Visualização de Informação proveniente do Twitter”. Em: (2016), p. 22 (ver p. 11).
- [30] Guru99. *What is Data Analysis? Research | Types | Methods | Techniques*. <https://www.guru99.com/what-is-data-analysis.html>. Acedido em: 2021-01-05. Dez. de 2020 (ver p. 65).
- [31] *Hierarquia DIKW*. [https://pt.wikipedia.org/wiki/Hierarquia\\_DIKW](https://pt.wikipedia.org/wiki/Hierarquia_DIKW). Acedido em: 2020-12-05 (ver p. 1).
- [32] *Highcharts Ratings and Reviews*. <https://www.trustradius.com/products/highcharts/reviews?qs=pros-and-cons> (ver p. 47).
- [33] *Highcharts Review*. <https://comparecamp.com/highcharts-review-pricing-pros-cons-features/> (ver p. 47).
- [34] *HTTP Streaming*. <https://www.pubnub.com/learn/glossary/what-is-http-streaming/> (ver p. 78).
- [35] Y. Huang, C. Shen e T. Li. “Event summarization for sports games using twitter streams”. Em: (2017). Acedido em: 2021-01-10. doi: [10.1007/s11280-017-0477-6](https://doi.org/10.1007/s11280-017-0477-6). url: <https://link.springer.com/article/10.1007/s11280-017-0477-6> (ver pp. 36, 37).
- [36] *Informação: a importância dos dados para as empresas*. <https://administradores.com.br/noticias/informacao-a-importancia-dos-dados-para-as-empresas>. Acedido em: 2020-12-10 (ver p. 1).
- [37] Instagram. *BRANDED CONTENT*. <https://business.instagram.com/a/branded-content-ads>. Acedido em: 2021-01-07. Jun. de 2019 (ver p. 17).
- [38] *InstaStats*. <http://instastatsapp.herokuapp.com/>. Acedido em: 2021-01-05 (ver p. 18).

- [39] W. J. C. S. J. S. James. "The value of Twitter for sports fans". Em: *Journal of Direct Data and Digital Marketing Practice* 16 (1 jul. de 2014). doi: [10.1057/ddmp.2014.36](https://doi.org/10.1057/ddmp.2014.36) (ver pp. 8, 10, 13, 14).
- [40] N. John. "Are you ready for some football sentiment analysis?" Em: (2019). Acedido em: 2021-01-09. url: <https://medium.com/swlh/are-you-ready-for-some-football-sentiment-analysis-22655017bf93> (ver pp. 35, 36).
- [41] C. F. Jon Rokne. "ASONAM '13: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining". Em: (2013). url: <https://dl.acm.org/doi/10.1145/2492517.2500235> (ver p. 2).
- [42] Karla Pequenino. *O que revelam as publicações do Twitter em Portugal? Medo e tristeza dominam, confiança oscila*. <https://www.publico.pt/2020/03/24/tecnologia/noticia/revelam-publicacoes-twitter-portugal-medo-tristeza-dominam-confianca-oscila-1908930>. Acedido em: 2020-12-23. Março 24 de 2020 (ver pp. 11-13).
- [43] A. Kirk. *Data Visualisation: A Handbook for Data Driven Design*. Sage Publications Ltd., 2016. isbn: 1473912148 (ver pp. 7, 32, 33, 54, 56, 57, 59-61, 118-120).
- [44] *Laravel Vs NodeJS: Which One Is Better?* <https://codersera.com/blog/laravel-vs-node-js-which-one-is-better/>. Acedido em: 2020-12-18 (ver p. 50).
- [45] X. Q. Y. L. N. T. G. Lis. "Making data visualization more efficient and effective: a survey". Em: (2020). url: <https://doi.org/10.1007/s00778-019-00588-3> (ver p. 2).
- [46] M. S. B. Lorena Nabanete dos Reis-Furtado Tamiris Lima Patrício e M. V. Carbinatto. "SPORT AND SOCIAL MEDIA: ANALYSIS OF THE BRAZILIAN GYMNASTICS FEDERATION'S INSTAGRAM". Em: (2021), p. 11. doi: [10.4025/jphyseduc.v32i1.3213](https://doi.org/10.4025/jphyseduc.v32i1.3213). url: <http://www.periodicos.uem.br/ojs/index.php/RevEducFis/article/view/48127/751375150992> (ver pp. 20, 22, 23).
- [47] H. M. R. C. J. H. D. U. K. D. A. H. L.-E. H. Mei-Chun. "[IEEE 2011 IEEE Conference on Visual Analytics Science and Technology (VAST) - Providence, RI, USA (2011.10.23-2011.10.28)] 2011 IEEE Conference on Visual Analytics Science and Technology (VAST) - Visual sentiment analysis on twitter data streams". Em: (2011). doi: [10.1109/VAST.2011.6102472](https://doi.org/10.1109/VAST.2011.6102472) (ver pp. 10, 13, 14).
- [48] *Metodologia de análise de dados: 6 etapas para colocar em prática*. <https://www.siteware.com.br/processos/metodologia-analise-de-dados/>. Acedido em: 2021-01-25 (ver p. 65).
- [49] A. Midha. *Nielsen: Twitter x TV activity levels indicate total audience engagement*. [https://blog.twitter.com/en\\_us/a/2015/nielsen-twitter-x-tv-activity-levels-indicate-total-audience-engagement.html](https://blog.twitter.com/en_us/a/2015/nielsen-twitter-x-tv-activity-levels-indicate-total-audience-engagement.html). Acedido em: 2021-01-10. Mar. de 2015 (ver pp. 14, 15).

- [50] *Mongodb VS Mysql VS Postgresql*. <https://stackshare.io/stackups/mongodb-vs-mysql-vs-postgresql>. Acedido em: 2020-12-14 (ver p. 52).
- [51] *NBAVis: Explorando os dados da temporada da NBA*. <https://www.cc.gatech.edu/gvu/i/sportvis/nbaVis/>. Acedido em: 2021-01-15 (ver pp. 26, 27).
- [52] *Node.js vs Django: Key differences, popularity, use cases and more*. <https://www.simform.com/blog/nodejs-vs-django/>. Acedido em: 2020-12-18 (ver p. 50).
- [53] *Node.js vs. Django: Is JavaScript Better Than Python?* <https://dzone.com/articles/nodejs-vs-django-is-javascript-better-than-python>. Acedido em: 2020-12-18 (ver p. 50).
- [54] *O PROJETO CARTOGRÁFICO*. <https://icaci.org/files/documents/wom/04IMYWoMpt.pdf> (ver p. 63).
- [55] G. Play. *Soccer Scores and Sports Livescore - SofaScore*. [https://play.google.com/store/apps/details?id=com.sofascore.results&hl=pt\\_PT&gl=US](https://play.google.com/store/apps/details?id=com.sofascore.results&hl=pt_PT&gl=US). Acedido em: 2021-01-15. Abr. de 2015 (ver p. 25).
- [56] *Power BI*. [https://pt.wikipedia.org/wiki/Power\\_BI](https://pt.wikipedia.org/wiki/Power_BI). Acedido em: 2020-01-20 (ver p. 67).
- [57] P. M. bibinitperiod Publicidade. *Dia mundial das redes sociais*. <https://www.meiosepublicidade.pt/2020/06/dia-mundial-das-redes-sociais/>. Acedido em: 2021-01-05. Jun. de 2020 (ver p. 16).
- [58] *Python vs. Ruby vs. Node.js – Which Platform Is a Fit for Your Project?* <https://railsware.com/blog/python-vs-ruby-vs-node-js-which-platform-is-a-fit-for-your-project/>. Acedido em: 2020-12-18 (ver p. 50).
- [59] *Quem usa e o que faz nas redes sociais?* <https://www.marktest.com/wap/a/n/id~26c7.aspx>. Acedido em: 2021-01-10 (ver p. 17).
- [60] R. G. Romney Miles; Johnson. “Show me a story: narrative, image, and audience engagement on sports network Instagram accounts”. Em: *Information Communication & Society* (jun. de 2018). doi: 10.1080/1369118X.2018.1486868 (ver pp. 20–22).
- [61] *Ruby on Rails vs. Node.js: Which One Is the Best Solution?* <https://jelvix.com/blog/ruby-on-rails-vs-node-js>. Acedido em: 2020-12-18 (ver p. 50).
- [62] P. C. V. R. S. C. S. J. W. J. C. S. “State of the Art of Sports Data Visualization”. Em: *Computer Graphics Forum* 37 (3 jun. de 2018). doi: 10.1111/cgf.13447 (ver pp. 23, 24, 28–31).
- [63] E. Saks. “JavaScript frameworks: Angular vs React vs Vue”. Em: (2019). Acedido em: 2021-01-15. url: <https://www.theseus.fi/bitstream/handle/10024/261970/Thesis-Elar-Saks.pdf?sequence=2> (ver p. 45).

- [64] S. J.-A. A. M. N. M. A. Sekkaki. "Soccer Video Summarization Using Video Content Analysis and Social Media Streams". Em: (2014). url: <https://dl.acm.org/doi/10.1109/BDC.2014.20> (ver p. 2).
- [65] A. Silva. *Cristiano Ronaldo é dos mais bem pagos no Instagram*. <https://espalhafactos.com/2020/07/17/cristiano-ronaldo-e-dos-mais-bem-pagos-no-instagram/>. Acedido em: 2021-01-08. Jul. de 2020 (ver p. 18).
- [66] *SmartBase Solutions*. <https://www.smartbasesolutions.com/>. Acedido em: 2021-02-11 (ver p. 31).
- [67] *Social Blade, Analytics made easy*. <https://socialblade.com/> (ver p. 9).
- [68] *SofaScore*. <https://www.sofascore.com/>. Acedido em: 2021-01-15 (ver p. 25).
- [69] *Sprinklr Introduces Modern Research Lite*. <https://www.sprinklr.com/newsroom/sprinklr-introduces-modern-research-lite/>. Acedido em: 2021-01-17 (ver pp. 33, 34).
- [70] *Statista, Global No.1 Business Data Platform*. <https://www.statista.com/> (ver p. 9).
- [71] A. O. Suin Kim Jinyeong Bak. "Do You Feel What I Feel? Social Aspects of Emotions in Twitter Conversations". Em: (2012). url: <https://www.researchgate.net/publication/265804086> (ver p. 2).
- [72] *System Properties Comparison MongoDB vs. MySQL vs. PostgreSQL*. <https://db-engines.com/en/system/MongoDB%3BMySQL%3BPostgreSQL> (ver p. 52).
- [73] L. M. O. J. P. C. D. C. J. D. C. S. C. T. "StatCast Dashboard: Exploration of Spatiotemporal Baseball Data". Em: *IEEE Computer Graphics and Applications* 36 (5 set. de 2016). doi: [10.1109/MCG.2016.101](https://doi.org/10.1109/MCG.2016.101) (ver p. 28).
- [74] H. Tankovska. *Number of social network users worldwide from 2017 to 2025*. <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>. Acedido em: 2021-01-12. Jan. de 2021 (ver pp. 2, 7).
- [75] *Text Analytics*. <https://docs.microsoft.com/pt-br/connectors/cognitiveservicestextanalytics/detect-sentiment> (ver p. 68).
- [76] *The Best Instagram Analytics Tools*. <https://www.brandwatch.com/blog/instagram-analytics-tools/>. Acedido em: 2021-01-15 (ver p. 18).
- [77] *The Haters Are Losing The War On Kickers*. <https://fivethirtyeight.com/features/the-haters-are-losing-the-war-on-nfl-kickers/>. Acedido em: 2021-01-13 (ver pp. 29, 30).
- [78] *The Mindset of Developers: Angular vs. React vs. Vue [Part 2]*. <https://blog.testproject.io/2020/11/23/the-mindset-of-developers-angular-vs-react-vs-vue-part-2/>. Acedido em: 2021-01-15 (ver p. 45).

- [79] *The RedMonk Programming Language Rankings: June 2020*. <https://redmonk.com/sogrady/2020/07/27/language-rankings-6-20/>. Acedido em: 2021-01-10 (ver pp. 50, 51).
- [80] *Tipos de Gráficos Estatísticos*. <https://www.matematica.pt/util/resumos/tipos-graficos-estatisticos.php>. Acedido em: 2021-01-16 (ver pp. 118–120).
- [81] *Top 15 Types Of Graphic Designs*. <https://www.designhill.com/design-blog/top-15-types-of-graphic-designs/>. Acedido em: 2021-01-16 (ver pp. 118–120).
- [82] *Tudo sobre Gráficos*. <https://www.todamateria.com.br/tipos-de-graficos/>. Acedido em: 2021-01-16 (ver pp. 118–120).
- [83] *Twitter changes the live TV sports viewing experience*. <https://marketing.twitter.com/en/insights/twitter-changes-the-live-tv-sports-viewing-experience>. Acedido em: 2021-01-10 (ver p. 16).
- [84] *Twitter, Facebook, or Instagram? Which Platform(s) You Should Be On*. <https://blog.hubspot.com/marketing/twitter-vs-facebook>. Acedido em: 2020-12-12 (ver p. 8).
- [85] *Visualização de dados desportivos*. <https://www.cc.gatech.edu/gvu/ii/sportvis/>. Acedido em: 2021-01-15 (ver p. 23).
- [86] *What is Big Data & why is Big Data important in today's era*. <https://medium.com/@syedjuaid.h47/what-is-big-data-why-is-big-data-important-in-todays-era-8dbc9314fb0a/>. Acedido em: 2021-01-12 (ver p. 4).
- [87] *Why Data Visualization Is Important*. <https://analytiks.co/importance-of-data-visualization/>. Acedido em: 2021-01-20 (ver p. 43).
- [88] M. L. Williams, P. Burnap e L. Sloan. “Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users’ Views, Online Context and Algorithmic Estimation”. Em: *Sociology* 51.6 (2017). PMID: 29276313, pp. 1149–1168. doi: 10.1177/0038038517708140. eprint: <https://doi.org/10.1177/0038038517708140>. url: <https://doi.org/10.1177/0038038517708140> (ver p. 10).
- [89] *WORLD LEADERS IN SPORTS DATA*. <https://www.optasports.com/>. Acedido em: 2021-01-12 (ver p. 26).
- [90] Y. Yang e X. Wang. “World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans’ tweets”. Em: (2015). Acedido em: 2021-01-10. doi: 10.1016/j.chb.2015.01.0750747-5632. url: <https://www.sciencedirect.com/science/article/abs/pii/S074756321500103X> (ver pp. 36, 37).



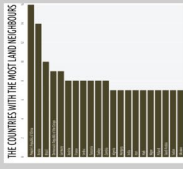
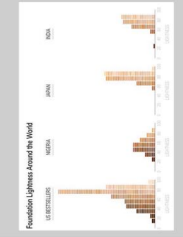
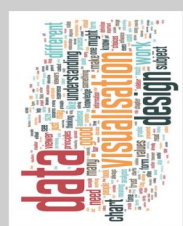
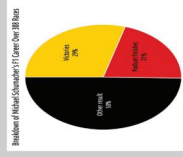
Tipos de Visualização	Gráfico de Barras	Histograma	Nuvem de Palavras	Gráfico Circular
Exemplo				
Descrição	Um gráfico de barras exibe valores quantitativos para diferentes itens de categoria. O gráfico compreende marcas de linha (barras) com o atributo de tamanho (comprimento ou altura) usado para representar o valor quantitativo de cada item.	Exibe a frequência e distribuição de medidas quantificadas em vários agrupados para itens de dados. A representação é formada pela variação do tamanho da linha (se os agrupamentos de valores têm intervalos iguais) ou da área da forma (se os agrupamentos de valores têm intervalos de valores diferentes) para representar a frequência das medições.	Mostra a frequência de itens de palavras individuais de dados textuais. Cada item é representado por palavras e em seguida, o tamanho da fonte de cada um é proporcional ao acordo com a frequência do seu uso. É importante lembrar que é efetivamente a área da palavra que codifica a sua medida quantitativa.	É possível visualizar as proporções de quantidades para diferentes categorias de constituintes formam um todo. Divide-se em setores para cada categoria com um certo ângulo que representa as proporções percentuais, separados por cores. O total de todos os valores do setor deve ser de 100%, caso contrário, o gráfico será corrompido.
Interatividade / Anotações	<ul style="list-style-type: none"> <li>Incluir dispositivos gráficos (marcas de escala, linhas de grade, etc) ajuda numa melhor precisão.</li> <li>Se incluímos rótulos de escala de eixo, não será necessário rotular diretamente cada valor da barra.</li> </ul>	<ul style="list-style-type: none"> <li>Incluir dispositivos gráficos (marcas de escala, linhas de grade, etc) ajuda numa melhor precisão.</li> <li>Se incluímos rótulos de escala de eixo, não será necessário rotular diretamente cada valor da barra.</li> </ul>	<ul style="list-style-type: none"> <li>Permitir a interatividade para os utilizadores interirem, filtrarem e examinarem as palavras com uma maior profundidade.</li> <li>Podem ser auxiliada com uma palavra-chave para explicar como as escalas de tamanho de fonte se equiparam à frequência de palavras.</li> </ul>	<ul style="list-style-type: none"> <li>Aplicar rótulos diretamente a cada categoria e um valor associado para aumentar a legibilidade.</li> </ul>
Composição / Cores	<ul style="list-style-type: none"> <li>As barras devem ser dimensionadas proporcionalmente de acordo com o valor quantitativo associado.</li> <li>Não há diferença significativa na percepção entre gráficos de barras organizados vertical ou horizontalmente.</li> <li>Incluir uma pequena lacuna entre cada barra ajudará a apresentar uma distinção clara entre cada item da categoria.</li> <li>É necessário tomar a classificação dos valores no gráfico o mais significativa possível.</li> </ul>	<ul style="list-style-type: none"> <li> Ao contrário do gráfico de barras, não deve haver nenhuma (ou, no máximo, uma muito fina) lacuna entre as barras para ajudar a emergir a forma coletiva das frequências.</li> <li>A classificação das caixas de valores quantitativos deve ser apresentada em ordem crescente.</li> </ul>	<ul style="list-style-type: none"> <li>Podemos utilizar variadas cores para acentuar ainda mais as frequências maiores que existem e também para organizar agrupamentos por categoria.</li> <li>As cores utilizadas devem ser explicadas através da inclusão de uma legenda.</li> <li>O arranjo das palavras é baseado num processo de layout que calcula o melhor posicionamento de cada palavra para ocupar o espaço ideal.</li> </ul>	<ul style="list-style-type: none"> <li>A cor é usada para classificar as associações categóricas de cada setor, portanto, é importante variar a propriedade de cada cor para maximizar a diferença visual.</li> <li>O posicionamento da primeira linha na posição vertical, forme e uma linha de base (lin) para ajudar a avaliar o valor do ângulo do primeiro setor.</li> <li>A ordenação dos setores usando valores descrescentes ou características ordinais ajuda com a legibilidade geral e a clareza de estorço.</li> </ul>
Alternativas	<ul style="list-style-type: none"> <li>Mostrar valores quantitativos ao longo do tempo.</li> <li>Gráficos de mini-barras</li> <li>Gráfico de barras agrupadas para comparar dois ou mais valores subjacentes.</li> <li>Gráfico de barras empilhadas que estabelece uma relação parte-para-todo.</li> <li>Gráficos de pontos onde existem grandes valores quantitativos com uma faixa estreita de diferença.</li> <li>Gráfico de símbolo proporcional para diversos tamanhos de valor e muitos itens categóricos.</li> </ul>	<ul style="list-style-type: none"> <li>Pirâmide populacional;</li> <li>Gráficos de caixa</li> <li>Gráficos de pontos para revelar detalhes mais granulares numa escala quantitativa.</li> </ul>	<ul style="list-style-type: none"> <li>Gráfico de barras para exibir conteúdos de texto;</li> <li>Gráfico proporcional em que o rótulo da palavra fica dentro de uma marca de forma dimensionada.</li> </ul>	<ul style="list-style-type: none"> <li>Gráfico de rosca, que geralmente, serve para acomodar uma propriedade de rotulagem.</li> <li>Gráfico de barras para exibir e comparar vários setores.</li> <li>Mapa de árvore, especialmente se os agrupamentos hierárquicos estiverem organizados.</li> <li>Gráfico de barras empilhadas que é bastante útil quando existe um relacionamento ordinal entre os valores categóricos.</li> </ul>

Figura 88: Comparação de características de diferentes gráficos [80, 82, 81, 2, 43]



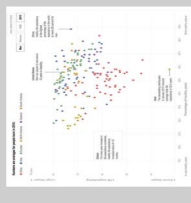
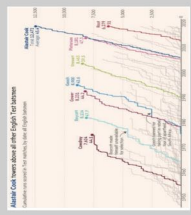
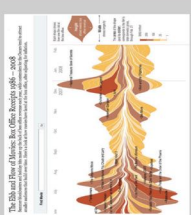
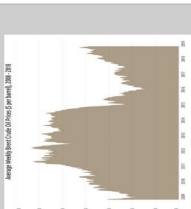
Tipos de Visualização	Gráfico de Dispersão	Gráfico de Linha	Gráfico de Fluxo	Gráfico de Área
Exemplo				
Descrição	Um gráfico de dispersão exibe a relação entre duas variáveis quantitativas para diferentes itens de categoria. O display é formado por marcas de pontos para cada item, plotados posicionalmente em relação a cada eixo quantitativo. Os atributos de marca de cor são usados, normalmente, para distinguir dimensões categoricas nos itens.	Um gráfico de linha mostra como os valores quantitativos mudam ao longo do tempo para diferentes itens de categoria. Várias categorias podem ser exibidas na mesma visualização, cada uma representada por uma linha discreta, geralmente com atributos de cores categoricas ou editoriais. Utilizam-se linhas de conexão para as linhas curvas para servir para enfatizar uma tendência.	Mostra como os valores quantitativos mudam ao longo do tempo para vários itens categoricos. São geralmente usados quando temos muitas categorias concorrentes e constituímos em qualquer ponto no tempo e essas categorias podem começar e parar em pontos diferentes no tempo, em vez de continuar ao longo do período apresentado.	Mostra como os valores quantitativos mudam ao longo do tempo para um único item categorico. As linhas de conexão unem itens adjacentes e relacionados para formar declives que são entendidos ao longo de toda a escala de tempo para exibir uma mudança completa ao longo do tempo. Para acentuar a forma das tendências, a área abaixo da linha é preenchida com cores, o que significa que a altura da área em qualquer ponto também revela sua quantidade.
Interatividade / Anotações	<ul style="list-style-type: none"> <li>Incluir dispositivos gráficos (marcas de escala, linhas de grade, etc.) ajuda numa melhor precisão.</li> <li>Utilizar linhas de referência para melhorar a interpretação.</li> </ul>	<ul style="list-style-type: none"> <li>A interatividade torna-se útil quando temos muitas linhas categoricas e desejamos permitir que o utilizador isole apenas uma determinada categoria de interesse.</li> <li>As etiquetas de classificação podem ser demarcadas da posição vertical ao longo da escala.</li> <li>Incluir dispositivos gráficos (marcas de escala, linhas de grade, etc.) ajuda numa melhor precisão.</li> </ul>	<ul style="list-style-type: none"> <li>A interatividade torna-se útil quando temos muitas linhas categoricas e desejamos permitir que o utilizador isole apenas uma determinada categoria de interesse.</li> <li>Revelar o valor quantitativo, o tempo e o rótulo da categoria em qualquer ponto do gráfico através de seleção de dados é uma dica bastante útil.</li> <li>As legendas diretas nas camadas do gráfico podem ser possíveis com a utilização de uma cor clara.</li> </ul>	<ul style="list-style-type: none"> <li>Oferecer breves descrições, anotações, entre outros.</li> <li>Incluir dispositivos gráficos (marcas de escala, linhas de grade, etc.) ajuda numa melhor precisão.</li> </ul>
Composição / Cores	<ul style="list-style-type: none"> <li>Utilizar círculos preenchidos, não preenchidos ou semitransparentes para ajudar a transmitir a frequência do valor.</li> <li>Geralmente, têm uma proporção de aspeto quadrada para ajudar os padrões a aparecerem de maneira mais evidente.</li> <li>Se existir uma variável quantitativa (emprego) a ser afetada por outra variável (emprego cultural), é prática geral colocar a primeira no eixo y e a última no eixo x.</li> </ul>	<ul style="list-style-type: none"> <li>Quando existem muitas categorias, em vez de colorir cada linha, pode ser viável enfatizar as linhas de interesse usando matiz de cor ou saturação.</li> <li>As dimensões do gráfico têm de ser cuidadosamente consideradas (altura, largura, etc.).</li> <li>O sequenciamento de valores tende a seguir uma direção cronológica de esquerda para a direita para o eixo x, baseado no tempo e valores baixos subindo para valores altos no eixo y.</li> <li>Os gráficos de linha nem sempre exigem que a origem do eixo quantitativo comece em zero, pois o tamanho de um valor é representado pela posição ao longo de uma escala, e não pelo tamanho de uma linha ou forma.</li> </ul>	<ul style="list-style-type: none"> <li>A área ocupada por esta camada é preenchida com um atributo de cor para representar uma escala de valor quantitativa adicional ou para associar classificações categoricas.</li> <li>O arranjo de empilhamento das múltiplas camadas categoricas pode deslocar para cima e para baixo na dimensão implícita do eixo y, a fim de otimizar o layout, mas não para indicar qualquer noção de valores positivos ou negativos.</li> </ul>	<ul style="list-style-type: none"> <li>As dimensões do gráfico têm de ser cuidadosamente consideradas (altura, largura, etc.).</li> <li>As inclinações para cima e para baixo podem parecer mais significativas se a largura do gráfico for estreita e menos significativa se for mais alargada.</li> <li>O sequenciamento de valores tende a seguir uma direção cronológica de esquerda para a direita para o eixo x, baseado no tempo e valores baixos subindo para valores altos no eixo y.</li> <li>No contrário do gráfico de linha, o eixo quantitativo para gráficos de área deve ter origem em zero.</li> </ul>
Alternativas	<ul style="list-style-type: none"> <li>Gráfico de bolha</li> <li>Matriz de gráfico de dispersão que envolve uma visão única de múltiplos gráficos de dispersão com diferentes combinações de variáveis quantitativas usada para explorar possíveis relações entre conjuntos de dados multivariados.</li> </ul>	<ul style="list-style-type: none"> <li>Gráfico de relevo para mostrar classificações ao longo do tempo.</li> <li>Gráfico de inclinação para comparar tendências em dois pontos ao mesmo tempo.</li> <li>Gráfico de barras quando temos quantidades para certos períodos discretos em vez de uma série contínua de medições pontuais.</li> </ul>	<ul style="list-style-type: none"> <li>Gráfico de área empilhada para poucos itens categoricos distintos.</li> <li>Gráfico de barras empilhadas, se existir poucas categorias a serem incluídas.</li> </ul>	<ul style="list-style-type: none"> <li>Gráfico de área empilhada, usado para mostrar como várias categorias formam um todo e como essa composição muda ao longo do tempo.</li> <li>Gráfico de densidade, usado para mostrar a distribuição de valores num eixo quantitativo, em vez de um eixo temporal.</li> <li>Gráfico de linha, se quisermos comparar com vários itens categoricos distintos.</li> </ul>

Figura 89: Comparação de características de diferentes gráficos [80, 82, 81, 2, 43]

# Anexo C

Tipos de Visualização	Gráfico de Radar	Pictograma	Gráfico de Caixa	Gráfico de Bolhas
Exemplo				
Descrição	<p>Plota valores em várias variáveis quantitativas para a formação de padrões gerais. Ele usa um layout radial (circular) que compreende vários eixos que emergem dos raios centrais de uma roda, um para cada variável. As linhas ou o preenchimento da forma são coloridos para ênfase ou para diferenciação de categoria quando mais de um item é plotado.</p> <ul style="list-style-type: none"> <li>Incluir dispositivos gráficos (marcas de escala, linhas de grade, etc) ajuda numa melhor precisão.</li> <li>Se os valores quantitativos estiverem em escalas diferentes, é necessário exibir os intervalos de valores de cada uma.</li> </ul>	<p>Escolha valores quantitativos para diferentes itens da categoria primária com a opção de usar a categoria secundária. A base do pictograma é a repetição do uso de marcas de pontos, na forma de símbolos ou imagens, para representar uma contagem quantitativa associada. Cada marca de ponto pode ser representativa de uma ou mais unidades quantitativas (por exemplo, um único símbolo pode representar 100 pessoas).</p> <ul style="list-style-type: none"> <li>A escolha dos símbolos deve ser o mais intuitiva possível.</li> <li>Caso contrário, quaisquer legendas devem ser apresentadas perto do display para permitir uma referência rápida para determinar a associação, categoria e quantitativa de cada variação de símbolo usada.</li> </ul>	<p>Exibe a distribuição e a forma de uma série de valores quantitativos para diferentes categorias. O display é formado por uma combinação de linhas e marcadores de ponto para indicar (por meio de posição e comprimento), normalmente, cinco medidas estatísticas diferentes. Um único gráfico será produzido para cada agrupamento de categorias relevantes.</p> <ul style="list-style-type: none"> <li>Incluir dispositivos gráficos (marcas de escala, linhas de grade, etc) ajuda numa melhor precisão.</li> <li>As anotações diretas normalmente são restritas apenas a certos pontos.</li> </ul>	<p>Mostra a relação entre três variáveis quantitativas para diferentes itens de categoria. Em contraste com o gráfico de dispersão, o gráfico de bolha usa marcas de forma (geralmente círculos) para cada item da categoria, plotados posicionalmente ao longo de cada eixo quantitativo com a variação no tamanho de cada marca representando uma terceira medida quantitativa.</p> <ul style="list-style-type: none"> <li>Um útil recurso interativo seria permitir a filtragem ou destaque de certos itens categóricos.</li> <li>As opções de seleção para revelar dicas de ferramentas podem ser úteis.</li> <li>Incluir dispositivos gráficos (marcas de escala, linhas de grade, etc) ajuda numa melhor precisão.</li> <li>Se incluímos rótulos de escala de eixo, não é necessário rotular cada valor de reticulado, pois causará sobrecarga de rótulos.</li> </ul>
Interatividade / Anotações	<ul style="list-style-type: none"> <li>Se as quantidades de marcadores excederem uma única linha, tente fazer o número de unidades por linha logicamente "contáveis".</li> <li>Para ajudar na legibilidade, certifique-se de que há uma lacuna suficientemente perceptível entre os clusters de unidades agrupadas.</li> <li>Procure tomar a classificação de valores no gráfico o mais significativa possível.</li> </ul>	<ul style="list-style-type: none"> <li>O eixo de valor quantitativo não precisa de começar em zero, a menos que esse signifique algo importante para a interpretação.</li> <li>Não há diferenças significativas na percepção entre gráficos de caixa dispostos no vertical ou no horizontal.</li> <li>Quando temos vários gráficos dentro do mesmo gráfico, devemos sempre que possível, tornar a classificação de categorias significativa, talvez organizando os valores em ordem crescente ou decrescente com base nos valores médianos.</li> </ul>	<ul style="list-style-type: none"> <li>Quaisquer cores utilizadas devem ser explicadas através da inclusão de uma legenda. Se as cores estiverem a ser usadas para distinguir diferentes categorias, é conveniente que sejam o mais visivelmente diferentes possível.</li> <li>As formas maiores podem se sobrepor, em termos espaciais, a outros valores ou até mesmo ocultá-los completamente. O uso de bordas de contorno e cores semitransparentes pode ajudar a evitar o efeito de oclusão total. Trabalhamos com a área da forma, não com o raio/diâmetro.</li> <li>O eixo quantitativo não precisa de ter origem em zero, a menos que seja significativo.</li> <li>Idealmente, um gráfico de bolha terá uma proporção de aspecto quadrada (lido alta quanto larga) para ajudar os padrões a aparecerem de forma mais evidente.</li> </ul>	
Composição / Cores	<ul style="list-style-type: none"> <li>Gráfico de radar: caso importante preservar o layout radial</li> <li>Gráfico de barras</li> <li>Gráfico de dispersão</li> <li>Se tivermos vários itens de categoria, podemos considerar o uso de pequenos múltiplos formatados por eixos distintos para cada item individual.</li> </ul>	<ul style="list-style-type: none"> <li>Gráfico de velas usa um método semelhante e é frequentemente usado para mostrar a distribuição e os valores quantitativos para eventos que mudam constantemente, como por exemplo, a análise do mercado de ações ao longo de um determinado período de tempo.</li> </ul>	<ul style="list-style-type: none"> <li>Gráfico de dispersão, se a terceira variável quantitativa for removida. Nas variações no gráfico de bolhas podemos usar diferentes formas geométricas como marcadores.</li> </ul>	
Alternativas				

Figura 90: Comparação de características de diferentes gráficos [80, 82, 81, 2, 43]



