

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI
CURSO DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO
CALINY BASÍLIO DE OLIVEIRA

VIZUALIZAÇÃO DE DADOS EXTRAÍDOS DO TWITTER: uma proposta de *dashboard*
utilizando a biblioteca *Chart.js*

Diamantina
2023

CALINY BASÍLIO DE OLIVEIRA

**VIZUALIZAÇÃO DE DADOS EXTRAÍDOS DO *TWITTER*: uma proposta de *dashboard*
utilizando a biblioteca *Chart.js***

Trabalho de Conclusão de Curso apresentado como parte dos requisitos exigidos para obtenção do título de Bacharel em Sistemas de Informação, da Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM).

Orientador: Caroline Queiroz Santos

**Diamantina
2023**



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI

FOLHA DE APROVAÇÃO

Caliny Basílio de Oliveira

VISUALIZAÇÃO DE DADOS EXTRAÍDOS DO TWITTER:
UMA PROPOSTA DE DASHBOARD UTILIZANDO A BIBLIOTECA CHART.JS

Trabalho de Conclusão de Curso apresentado ao Departamento de Computação como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação pela Universidade Federal dos Vales do Jequitinhonha e Mucuri.

Aprovada em 07/02/2023

BANCA EXAMINADORA

Profa. Caroline Queiroz Santos
Faculdade de Ciências Exatas - UFVJM

Caroline Miranda Barroso
Faculdade de Ciências Exatas - UFVJM

Cláudia Beatriz Berti
Faculdade de Ciências Exatas - UFVJM



Documento assinado eletronicamente por **Claudia Beatriz Berti, Servidor (a)**, em 07/02/2023, às 17:09, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Caroline Queiroz Santos, Servidor (a)**, em 07/02/2023, às 17:11, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Caroline Miranda Barroso, Servidor (a)**, em 07/02/2023, às 17:12, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufvjm.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0976262** e o código CRC **F304D6F0**.

Dedico este trabalho à minha família, em especial à minha mãe Gessy e meu irmão Diego, sem eles eu não teria conseguido atravessar tantas barreiras e superar tantas adversidades na vida.

AGRADECIMENTOS

Gostaria de expressar minha gratidão a todos que contribuíram para a realização desse momento tão especial. Agradeço, em particular, à minha família, aos meus grandes amigos Camila e Tiago, ao Yuri Faustino, à minha orientadora Carol, ao querido Xena e ao Professor Leonardo Turchi, cujas colaborações foram imprescindíveis para a conclusão deste trabalho. Além disso, agradeço aos meus colegas de turma pela companhia e troca de conhecimentos ao longo da graduação. Este Trabalho é o resultado de um esforço coletivo, e sou profundamente grata a todos os envolvidos.

”Brechas serão vossos caminhos, pois tudo que é fácil é vão”(SANT, 2015).

RESUMO

A visualização de dados é uma técnica fundamental para a análise e compreensão de informações, pois permite que os dados sejam apresentados de forma clara e concisa. Baseado nessa importância, foi proposto o desenvolvimento de um *dashboard* utilizando a biblioteca Chart.js para auxiliar os integrantes do Projeto “Nos fios das controvérsias: um estudo comparativo sobre mídia, futebol e suas dimensões políticas em Belo Horizonte e Montevideu”, na análise de dados coletados da rede social Twitter, provenientes de jornalistas pré-selecionados. Esses dados incluem *tweets*, *replies* e *likes*. O objetivo do *dashboard* é fornecer uma interface intuitiva e interativa, com representações gráficas para facilitar a compreensão e análise dos dados. Contribuindo na identificação de tendências, padrões e relações entre diferentes variáveis dos dados, o que é complexo de realizar em dados brutos. Através dos resultados foi possível identificar lacunas na coleta dos dados que impedem a elaboração de visualizações consideradas críticas para a análise dos dados.

Palavras-chave: Redes Sociais. Twitter. Política. Futebol. Dashboard. Análise de Dados. Visualização de Dados.

ABSTRACT

Data visualization is a foundational technique earmarked for analyzing and understanding information, since it allows data to be shown in a clear, concise manner. Based on its importance, the development of a dashboard was proposed, using the Chart.js library to help the project members of “On controversies wires: a comparative study on media, soccer and their political dimensions in Belo Horizonte and Montevideo“, in the analysis of collected data from social network Twitter, from pre-selected journalists. This data includes tweets, replies and likes. The purpose of the dashboard is to provide an intuitive and interactive interface with graphical representations which facilitate the understanding and analysis of data. Contributing to the identification of trends, patterns and relationships between different data variables, which is complex to perform in raw data. Through the results, it was possible to identify gaps in data collections that restrict perspectives elaboration and critical opinions for data analysis..

Keywords: Social networks. Twitter. Politics. Dashboard. Data analysis. Data Visualization.

LISTA DE ILUSTRAÇÕES

Figura 1 – Linha do tempo de lançamento das principais redes sociais	24
Figura 2 – Exemplo <i>tweet</i>	26
Figura 3 – Página inicial do <i>Twitter</i>	26
Figura 4 – Possíveis interações no <i>Tweet</i>	27
Figura 5 – 5Vs do Big Data	29
Figura 6 – Painel Coronavírus	32
Figura 7 – Painel Google <i>Analytics</i>	32
Figura 8 – Visualizações geradas	33
Figura 9 – Metodologia	37
Figura 10 – Diagrama Relacional	39
Figura 11 – Página Inicial de <i>Login</i>	41
Figura 12 – Visão geral do <i>dashboard</i>	41
Figura 13 – Gráfico Incidência de Palavras (<i>tweets</i> jornalistas)	42
Figura 14 – Tabela Incidência de Palavras	42
Figura 15 – Gráfico de Linhas (<i>tweets</i> jornalistas)	42
Figura 16 – Gráfico Incidência de Palavras (<i>replies</i>)	43
Figura 17 – Gráfico Incidência de Palavras (<i>likes</i>)	44
Figura 18 – Contador e Período de Coleta	44
Figura 19 – Gráfico de pizza - Quantidade de Jornalistas	45
Figura 20 – Tabela Jornalista e Usuários	45

LISTA DE TABELAS

Tabela 1 – <i>Ranking</i> redes sociais	25
Tabela 2 – Jornalistas	36

LISTA DE ABREVIATURAS E SIGLAS

UFVJM	Universidade Federal dos Vales do Jequitinhonha e Mucuri
FACET	Faculdade de Ciências Exatas
HTML	Linguagem de Marcação de Hipertexto
CSS	Cascading Style Sheets
UNIFAL	Universidade Federal de Alfenas
BI	Business Intelligence

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Motivação	21
1.2	Objetivos	22
1.2.1	<i>Objetivo geral</i>	22
1.2.2	<i>Objetivos específicos</i>	22
2	REFERENCIAL TEÓRICO	23
2.1	Redes Sociais	23
2.2	Visualização de Dados	27
2.3	Análise de Dados	28
2.4	Big Data	29
2.5	Business Intelligence	30
2.6	<i>Dashboard</i>	30
3	TRABALHOS RELACIONADOS	33
4	NOS FIOS DAS CONTROVÉRSIAS: UM ESTUDO COMPARATIVO SOBRE MÍDIA, FUTEBOL E SUAS DIMENSÕES POLÍTICAS EM BELO HORIZONTE E MONTEVIDÉU	35
5	METODOLOGIA	37
6	RESULTADOS	39
6.1	O Banco de Dados	39
6.2	<i>O Dashboard</i>	40
6.3	Representações Gráficas Geradas	41
7	DISCUSSÃO DOS RESULTADOS	47
8	CONSIDERAÇÕES FINAIS	49
	REFERÊNCIAS	51
9	SEMI-ESTRUTURA ENTREVISTA INICIAL	55
10	FORMULÁRIO VALIDAÇÃO DO <i>DASHBOARD</i>	57

1 INTRODUÇÃO

Diariamente são gerados uma quantidade massiva de dados, sendo esses considerados o novo petróleo da atualidade, um recurso natural e valioso para as organizações. Explorando esses dados gerados da maneira correta, é possível elencar, por exemplo, os gostos, hábitos e perfis dos consumidores. Com essas informações, as organizações conseguem criar novos produtos, serviços, marketing direcionado, afim de gerar uma experiência personalizada para os perfis identificados nas análises (BELTRAMINI; GUNTHER, 2022).

Na atualidade, as redes sociais, como *Facebook*, *Twitter* e *Tiktok*, são responsáveis por gerar a maioria dos dados no cenário virtual, pois conta com usuários assíduos. O *Twitter*, por exemplo, conta com milhões de usuários ativos diariamente, postando e interagindo em *posts* sobre temas diversos, gerando uma extensa quantidade de dados a todo instante. Esses dados gerados pelos usuários do *Twitter* podem e são explorados de diversas formas, tanto para fins comerciais quanto acadêmicos.

Entretanto, um montante de dados brutos não gera informação útil para tomadas de decisões, é necessário coletá-los, podendo ocorrer de diversas formas, como manualmente ou com auxílio de alguma ferramenta, após isso, é indispensável filtrá-los e analisá-los (CHEN; CHIANG; STOREY, 2012). Para analisar os dados coletados, uma visualização inclinada para o objetivo da análise torna o processo mais assertivo e rápido.

No presente trabalho, é proposto um *dashboard* para exposição de dados gerados através de coleta por *script* da rede social *Twitter*, a fim de colaborar com o Projeto “Nos fios das controvérsias: um estudo comparativo sobre mídia, futebol e suas dimensões políticas em Belo Horizonte e Montevideú” (PACHECO *et al.*, 2022) na análise dos dados, além de evidenciar a importância da visualização de dados nesses cenários.

1.1 Motivação

A atual monografia visa contribuir com o Projeto de pesquisa “Nos fios das controvérsias: um estudo comparativo sobre mídia, futebol e suas dimensões políticas em Belo Horizonte e Montevideú” (PACHECO *et al.*, 2022), que está em desenvolvimento no Laboratório de Etnografia - LABET da Universidade Federal de Alfenas (UNIFAL), pelo Professor Dr. Leonardo Turchi Pacheco. O Projeto tem o intuito de realizar um estudo comparativo entre jornalistas esportivos que trabalham em rádios do Brasil e do Uruguai, tendo como um dos objetivos primários compreender a relação entre política e futebol através da interação entre jornalistas e torcedores mediada pela rede social *Twitter*.

No Projeto Nos Fios, foram selecionados dezessete jornalistas, sendo dez do Brasil e sete do Uruguai, para coleta e análise de *twittes* e interações de suas contas no *twitter*. A coleta desses dados ocorre de forma manual e através de um *script* desenvolvido por Santos (2022), bolsista do Projeto.

Durante entrevista com o coordenador do Projeto, ele expressou preocupação quanto ao modo como seriam visualizados os dados coletados, uma vez que essa etapa ainda não havia sido planejada. Visando auxiliá-los na visualização objetiva desses dados, propomos o

desenvolvimento de *dashboard* com o intuito de promover uma visualização personalizada e análise assertiva dos dados coletados.

1.2 Objetivos

1.2.1 Objetivo geral

Desenvolver um *dashboard* para apresentação de dados provenientes do *Twitter*, promovendo a visualização dos dados de forma personalizada, a ferramenta tem como objetivo auxiliar o usuário na melhor compreensão dos dados coletados, com um design objetivo e interativo.

1.2.2 Objetivos específicos

Em busca de atingir o objetivo geral, foram definidos os seguintes objetivos específicos:

- I. Revisar a bibliografia de trabalhos relacionados existentes;
- II. Realizar entrevista com o Professor Dr. Leonardo Turchi Pacheco;
- III. Definir quais dados e gráficos a serem expostos no *dashboard* com base na entrevista mencionada no item anterior;
- IV. Plotar gráficos e informações utilizando os arquivos *.json* disponibilizados pelo Projeto Nos Fios;
- V. Apresentar gráficos gerados e analisar os dados com base na bibliografia existente.

2 REFERENCIAL TEÓRICO

Neste capítulo, é apresentada a fundamentação teórica indispensável para o desenvolvimento e domínio do estudo deste trabalho, onde descrevemos uma visão ampla sobre redes sociais, aprofundando no *Twitter*, por ser a rede social fonte de dados para o estudo.

Além disso, descrevemos brevemente as tecnologias utilizadas no desenvolvimento da aplicação fruto do presente trabalho de conclusão de curso.

2.1 Redes Sociais

Marteleto (2001) retrata redes sociais como um agrupamento de participantes independentes, onde compartilham seus interesses e valores, unindo recursos e ideias. De forma semelhante, Silva e Ferreira (2007) qualificam redes sociais como grupo de pessoas, empresas ou qualquer outra entidade criada socialmente, conectadas entre si por meio de um conjunto de relações sociais, como amizade, relação profissional, comércio ou apenas troca de informações.

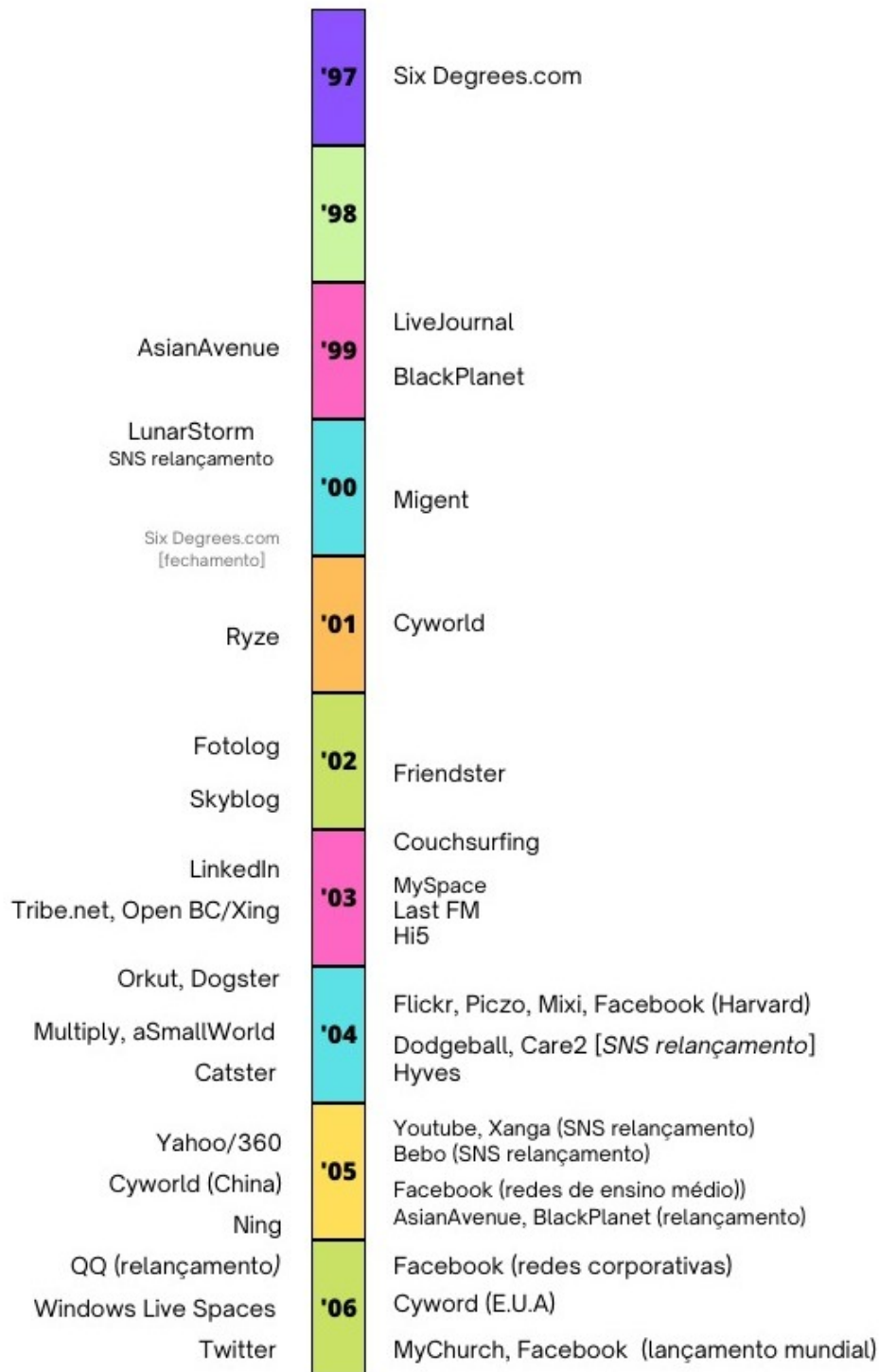
Rede social *on-line* é sintetizada como uma organização social composta por indivíduos, associações, empresas ou outras entidades sociais, onde os usuários estão conectados por um ou diversos elos, seja familiar, de amizade, comercial ou amoroso (FERREIRA, 2011). Com base nessas relações, os atores compartilham informações, popularidade, crenças e posicionamentos políticos.

Surgindo nos anos 90, de acordo com (BOYD; ELLISON, 2008), o *Six Degrees* foi a primeira rede social *on-line* conhecida, com lançamento em 1997 por Andrew Weinreich, um empreendedor americano. O *Six Degrees* era um site de rede social *on-line* que permitia que os usuários criassem perfis e adicionassem amigos. Os usuários podiam adicionar informações pessoais, como sua localização, interesses e ocupação, e podiam procurar outros usuários para adicionar à sua lista de amigos. No entanto, devido a problemas de escalabilidade e falta de financiamento, o site foi encerrado em 2000.

Em seguida, fundado em 1999 (Figura 1), o *LiveJournal* era um serviço de *blog* e rede social popular nos anos 2000, onde usuários podiam criar seus próprios blogs pessoais e se conectar com outros usuários através de comentários e amizades, incluindo também recursos como listas de amigos, comunidades baseadas em interesses e recursos de grupo (BOYD; ELLISON, 2008).

Ainda observando a *timeline* proposta por (BOYD; ELLISON, 2008), é possível verificar que foram lançadas diversas redes sociais ao longo dos anos, sendo que as mais populares nos anos 2000, inclui o *LinkedIn* (2002), *MySpace* (2003), *Facebook* (2004), *YouTube* (2005), *Twitter* (2006), devido as datas de lançamentos de algumas redes serem posteriores a criação da *timeline*, elas não estão presentes na figura, como *Instagram* (2010), *WhatsApp* (2009) e *TikTok* (2016).

Figura 1 – Linha do tempo de lançamento das principais redes sociais



Fonte: Adaptado (BOYD; ELLISON, 2008)

Martins (2022) apresentou um *ranking* das 10 (dez) redes sociais mais populares no Brasil em 2022, conforme consta na Tabela 1.

Tabela 1 – Ranking redes sociais

Posição	Rede Social
1	WhatsApp
2	YouTube
3	Instagram
4	Facebook
5	TikTok
6	Facebook Messenger
7	LinkedIn
8	Pinterest
9	Twitter
10	Snapchat

Fonte: Adaptado (MARTINS, 2022)

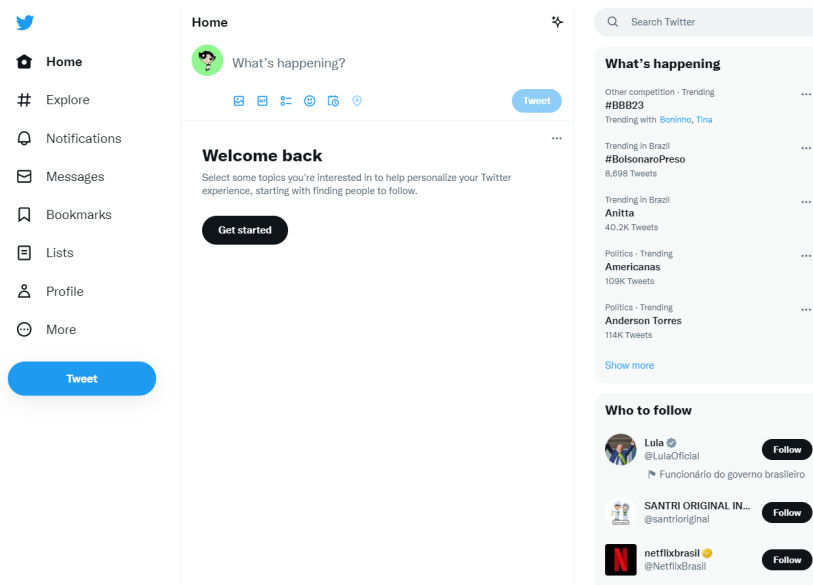
Neste trabalho, foi utilizada a rede social Twitter, fundada em Março de 2006 por Jack Dorsey, Evan Williams, Biz Stone e Noah Glass, nos Estados Unidos. O *Twitter* inicialmente era um projeto paralelo da empresa *Odeo*, uma ferramenta de *podcasting* (MEIO&MENSAGEM, 2022). Os criadores desejavam uma plataforma com compartilhamento de mensagens similar a troca de *SMS* (mensagem de texto), a nomearam de *Twtrr* (sem as vogais), com uso restrito apenas aos colaboradores da *Odeo*. Após seis meses do lançamento, o nome foi alterado para *Twitter* (SMAAL, 2010).

Ainda em 2006, especificamente no mês de julho, a versão completa da rede social foi disponibilizada ao público, inicialmente com a limitação de 140 caracteres por postagem, ação que recebe o nome de *tweet* (Figura 2) (MEIO&MENSAGEM, 2022). Em 2017, o limite de caracteres dobrou, sendo possível postar mensagens curtas de até 280 caracteres, contendo além de texto, imagens, vídeos, *links*, *gifs*, enquetes, *emojis*, cronograma e localização, desde que seja respeitado o referido limite.

A plataforma é destinada aos usuários que desejam saber mais sobre o que está acontecendo e sobre o que as pessoas estão falando no momento Twitter (2022). Com o intuito de instigar o usuário a *Tweetar* com frequência, no perfil dele é apresentada a mensagem “*O que está acontecendo?*”, conforme Figura 3.

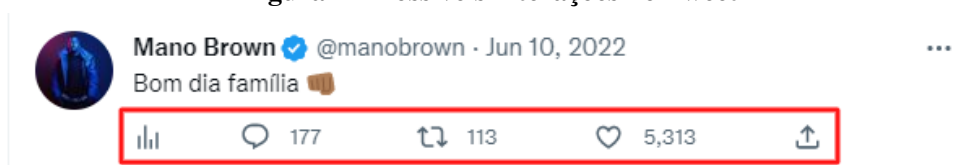
Figura 2 – Exemplo *tweet*

Fonte: Twitter (2023)

Figura 3 – Página inicial do *Twitter*

Fonte: Twitter (2023)

Após publicados, os *tweets* possuem quatro possibilidades de interação (figura 4): *Reply*, *Retweet*, *Like* e *Share*. Além disso, a plataforma fornece espaço para conversação privada, e uma funcionalidade denominada *Trending Topics*, onde é apresentado os assuntos em alta e mais comentados do mundo naquele momento.

Figura 4 – Possíveis interações no Tweet

Fonte: Twitter (2023)

Dang (2021) indica que o *Twitter* registrou 211 milhões de usuários ativos diariamente no terceiro trimestre de 2021, sendo essa extensa quantidade de usuários ativos, responsável por gerar a cada segundo uma diversa e massiva base de dados. Em Gluck (2017) é exposto que dados extraídos de redes sociais, são superiores as demais fontes de informações, sendo o *Twitter* responsável por apontar o que é importante para os usuários naquele momento, tornando-o rico em dados para estudo e aplicação em áreas distintas, como para análise de opinião, tendências de mercado, análise de redes sociais, desenvolvimento de uma pesquisa acadêmica e campanha de *marketing* direcionado.

2.2 Visualização de Dados

Conforme exposto por Card, Mackinlay e Shneiderman (1999), visualização de dados corresponde à representação visual de informações e dados, através de elementos como gráficos, diagramas e mapas, tendo como intuito ampliar a interpretação do usuário. Em suma, a visualização de dados tem como objetivo aumentar a cognição do usuário em relação aos dados expostos.

A história da visualização de dados, segundo Friendly e Denis (2001), remonta ao século XVI, quando os navegadores europeus começaram a usar diagramas geométricos, tabelas de posição das estrelas e mapas para representar e analisar dados. Com a expansão marítima da Europa, novas técnicas e instrumentos foram desenvolvidos, permitindo a criação de formas mais precisas e avançadas de apresentação visual do conhecimento (FRIENDLY; DENIS, 2001).

Desde então, cientistas e matemáticos desenvolveram técnicas e ferramentas para representar e analisar dados, incluindo William Playfair, que introduziu gráficos de barras e gráficos de linhas no século XVIII, e Florence Nightingale, que usou gráficos de setores para mostrar a mortalidade em hospitais durante a Guerra da Crimeia (FRIENDLY; DENIS, 2001).

A evolução tecnológica tem revolucionado a forma como os dados são visualizados, tornando-os mais acessíveis e compreensíveis para o público. De acordo com Chen, Chiang e Storey (2012), representação gráfica dos dados é considerada uma das maneiras mais impactantes de comunicação e é vista como uma das principais ferramentas para o processo decisório.

Antes da popularização da tecnologia, a visualização de dados era geralmente limitada a gráficos e tabelas estáticas, conforme exposto por Friendly e Denis (2001), tornando difícil para as pessoas compreenderem informações complexas. No entanto, com o avanço da tecnologia, a visualização de dados tornou-se mais avançada e interativa, possibilitando que os usuários explorem e compreendam os dados de forma mais eficaz.

Uma das principais contribuições da tecnologia para a visualização de dados é a capacidade de criar gráficos e *dashboards* interativos. De acordo com o *CEO* da Tableau, Christian Chabot, os indivíduos precisam ser capazes de ver, explorar e compreender seus dados. A interatividade é a chave para tornar isso possível. Gráficos interativos permitem que os usuários explorem os dados de diferentes perspectivas, ajudando-os a entender tendências, padrões e relações entre diferentes conjuntos de dados (CHABOT, 2013). A tecnologia está transformando a maneira como as pessoas veem e entendem os dados, com gráficos interativos e ferramentas de inteligência artificial sendo apenas algumas das suas manifestações.

A visualização de dados é uma ferramenta valiosa para extrair informações e conhecimentos dos dados. Entretanto, não é garantido que ela por si só extraia o máximo de informações possíveis. É essencial combinar a visualização de dados com outras técnicas de análise, como estatísticas e modelagem, para obter uma compreensão completa dos dados. Além disso, a escolha adequada dos tipos de gráficos e visualizações para representar os dados é crucial para a eficácia da análise (HEALY, 2018).

A qualidade e eficácia da visualização dos dados está diretamente ligada a coleta, devido a isso se faz necessário o planejamento desta etapa, fundamental para garantir que as informações coletadas sejam relevantes, precisas e completas. Tal planejamento permite identificar os dados necessários para atingir os objetivos do Projeto, além de estabelecer como eles serão coletados, armazenados e analisados. Sem um planejamento adequado, é possível que dados importantes sejam perdidos ou que sejam coletados dados desnecessários, o que pode levar a conclusões equivocadas e a desperdício de tempo e recursos. Além disso, o planejamento da coleta de dados também é importante para garantir que as informações coletadas sejam de qualidade e possam ser utilizadas para tomar decisões estratégicas (EVERGREEN, 2018).

2.3 Análise de Dados

A análise de dados é um processo utilizado para examinar, limpar, transformar e modelar dados com o objetivo de obter informações valiosas e conhecimento. Essa técnica é amplamente utilizada em diversas áreas, como negócios, ciência, saúde e jornalismo, para tomar decisões baseadas em dados e entender tendências e padrões (BOEHMKE, 2016).

De acordo com Chen, Chiang e Storey (2012), a análise de dados envolve quatro etapas principais: coleta de dados, limpeza de dados, análise exploratória e análise estatística. A coleta de dados é o primeiro passo, onde os dados são coletados de fontes internas e externas, como bancos de dados, questionários, sensores e redes sociais. A limpeza de dados é a etapa seguinte, onde os dados são examinados e limpos para remover ruídos, *outliers* e dados faltantes. A análise exploratória é a etapa onde os dados são visualizados e explorados para obter *insights* e conhecimentos Tukey (1977). Por fim, na análise estatística, os dados são utilizados para testar hipóteses e modelos estatísticos (CHEN; CHIANG; STOREY, 2012).

É importante destacar que, como mencionado anteriormente, a análise de dados não se limita somente à visualização de dados, mas também inclui outras técnicas como estatísticas e modelagem. Segundo Boehmke (2016) a representação gráfica dos dados é fundamental para

uma análise eficaz, mas não é suficiente. É crucial selecionar as visualizações e tipos de gráficos adequados para transmitir de forma clara e precisa as informações contidas nos dados.

Em suma, a análise de dados é um processo que permite obter informações valiosas e conhecimento a partir dos dados, para tomar decisões baseadas em dados e entender tendências e padrões. É composta por diversas etapas, além de combinar técnicas de visualização, estatísticas e modelagem para obter uma compreensão completa dos dados.

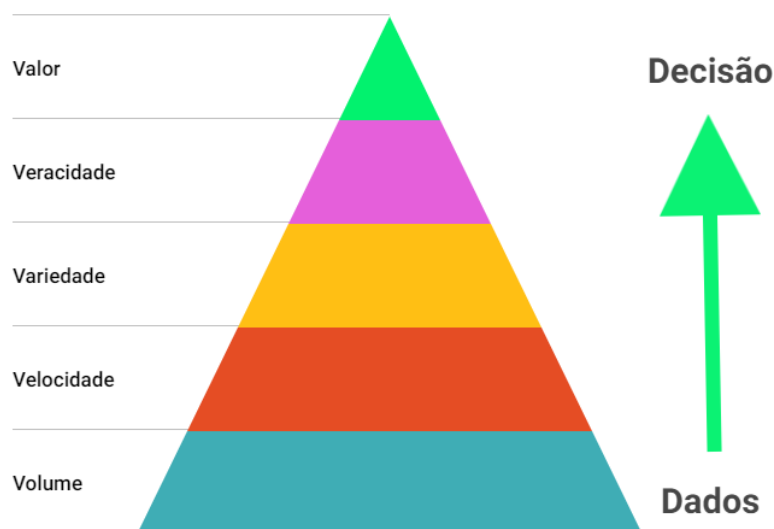
2.4 Big Data

Big data é definido como dados em grande escala e diversidade, onde devido a essas características se torna inviável o processamento e análise por softwares tradicionais, necessitando de outras abordagens para o tratamento (ORACLE, 2022). Em outras palavras, Taurion (2013) expõem que, *big data* tem como foco principal questões de volume de conjunto de dados extremamente grandes, gerados a partir de práticas tecnológicas. O Autor ainda afirma que *big data* não é teoria, pois geramos dados a cada instante, sendo que ao analisar os padrões e correlações nesse imenso volume de dados, pode-se gerar informações valiosas para todos os setores da nossa sociedade.

Em resumo, o *Big Data* consiste em 5Vs (figura 5) comumente conhecidos na bibliografia (RIBEIRO, 2021):

- I. Volume: determinado pela quantidade de dados;
- II. Velocidade: aponta a periodicidade de atualização dos dados;
- III. Variedade: refere-se aos diferentes formatos e tipos dos dados;
- IV. Veracidade: tange à confiabilidade e qualidade dos dados;
- V. Valor: resultante da combinação dos 4Vs anteriores, o valor é as informações geradas que agregam nas tomadas de decisão da organização.

Figura 5 – 5Vs do Big Data



Fonte: Adaptado (FALSARELLA; JANNUZZI; SUGAHARA, 2017)

Conclui-se que, tomadas de decisão baseada em dados, ajuda com diversas atividades de negócios, desde a experiência do cliente até a análise avançada, assim como com desenvolvimento de novos produtos, manutenção preditiva, fraude e conformidade, eficiência operacional e promoção da inovação (ORACLE, 2022).

2.5 Business Intelligence

Primak (2008) expõem que o termo *Business Intelligence* (BI) surgiu na década de 80, mas as técnicas e ferramentas utilizadas na BI existem há décadas. A análise de dados e a criação de relatórios, por exemplo, são utilizadas desde o século XIX para ajudar as organizações nas tomadas de decisões. No entanto, foi somente na década de 1980 que o termo *Business Intelligence* foi definido e começou a ser usado para descrever técnicas e ferramentas utilizadas para coletar, integrar, analisar e fornecer acesso a informações relevantes para auxílio nas tomadas de decisões. O objetivo do BI é transformar dados brutos em informações úteis, capazes de aprimorar o desempenho da empresa e contribuir para o alcance de seus objetivos.

As ferramentas de BI incluem softwares de análise de dados, *dashboards*, relatórios, mineração de dados e visualização de dados. Essas ferramentas permitem aos usuários transformar grandes volumes de dados em informações fáceis de entender e utilizar para tomar decisões estratégicas e táticas. Além disso, as ferramentas de BI permitem integrar dados de fontes diferentes e combinar dados históricos com dados em tempo real, o que permite uma análise mais completa e precisa (PRIMAK, 2008). Neste trabalho focaremos na visualização de dados, mais especificamente no formato de *dashboards*.

2.6 Dashboard

Definindo *dashboards* de forma ampla, o autor Few (2004) enfatiza a eficiência de sua natureza comunicativa como um atributo importante:

Um *dashboard* é uma exibição visual das informações mais importantes necessárias para atingir um ou mais objetivos; [essas informações estão] consolidadas e organizadas em uma única tela para que possam ser monitoradas em um relance (FEW, 2004, p. 3).

De outro modo, Arbex (2013) aponta *dashboards* como painéis com componentes visuais, onde a leitura dos dados ocorre de modo simultâneo e integral, e não uma leitura em momentos distintos, como ocorre em textos tabulares. Complementando, Manoel (2020) afirma que para uma análise assertiva, é preciso assegurar a consistência e confiabilidade dos dados.

Os *dashboards* podem ser categorizados de diversas formas, sendo que cada categoria serve a um propósito específico e é projetado para atender às necessidades de informação de diferentes grupos de usuários. Uma das formas de classificar os *dashboards* é pelo comportamento de seus elementos visuais, o que permite classificá-los em três tipos diferentes: operacional, estratégico e analítico (FEW, 2006).

O *dashboard* operacional é projetado para monitorar as atividades diárias e fornecer informações em tempo real para equipes operacionais. Ele é descrito por Few (2006), como um painel que permite acompanhar o desempenho, identificar rapidamente problemas e tomar medidas imediatas para solucioná-los.

Já o *dashboard* estratégico é projetado para acompanhar o desempenho de métricas-chave e fornecer uma visão geral do negócio para a liderança. Ele fornece uma visão geral e consolidada das métricas e indicadores-chave de desempenho (KPI) mais relevantes para a empresa, permitindo aos líderes tomar decisões informadas e baseadas em dados.

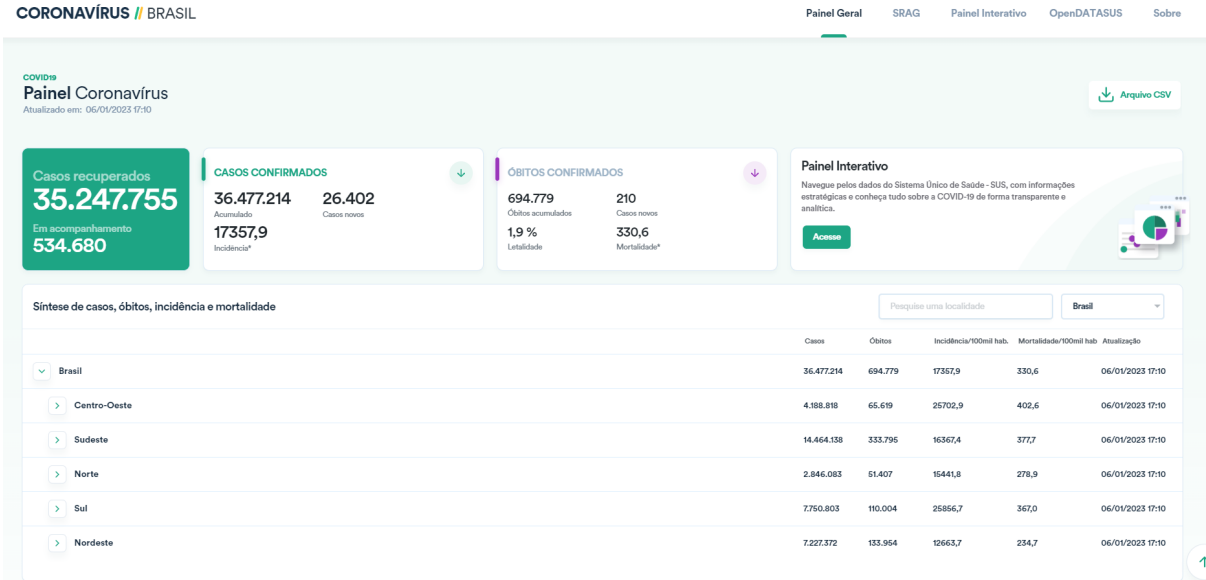
Por fim, temos o analítico, uma ferramenta valiosa que permite aos usuários interagir e avaliar dados de forma intuitiva e eficiente, a fim de retirar *insights* relevantes e informações valiosas para a tomada de decisões baseadas em dados. Ainda segundo Few (2006), este painel é projetado para ajudar os usuários a compreender facilmente tendências e padrões presente nos dados analisados, tornando-se uma ferramenta essencial para a tomada de decisões informadas e fundamentadas.

O *dashboard* estando alinhado com o objetivo dos usuários, há inúmeras vantagens em sua utilização, que incluem agilidade no acesso as informações relevantes de forma rápida e fácil, sem a necessidade de navegar por relatórios extensos; tomadas de decisão baseadas em dados; personalização da ferramenta para atender às necessidades específicas de uma empresa ou usuário, permitindo-lhes acessar apenas as informações relevantes; projetados para serem fáceis de usar, mesmo para usuários sem habilidades técnicas, a facilidade de uso também é uma vantagem; e por fim, o *dashboard* permite compartilhar informações e discutir sobre elas, auxiliando na colaboração entre equipes.

Atualmente existem várias tecnologias disponíveis para criação de dashboards, algumas das quais incluem: *Power BI*, ferramenta de análise de dados da *Microsoft* que permite criar *dashboards* e relatórios avançados; *Tableau*, ferramenta de visualização de dados amplamente utilizada para criar *dashboards* interativos e explorações de dados; e a *Chart.js*, uma biblioteca *JavaScript open-source* que permite criar gráficos e visualizações de dados de forma fácil e rápida. A *Chart.js* suporta uma variedade de tipos de gráficos, incluindo gráficos de barras, linhas, áreas, pizza e radar, sendo altamente personalizável, permitindo alteração de cores, formas e outros aspectos dos gráficos.

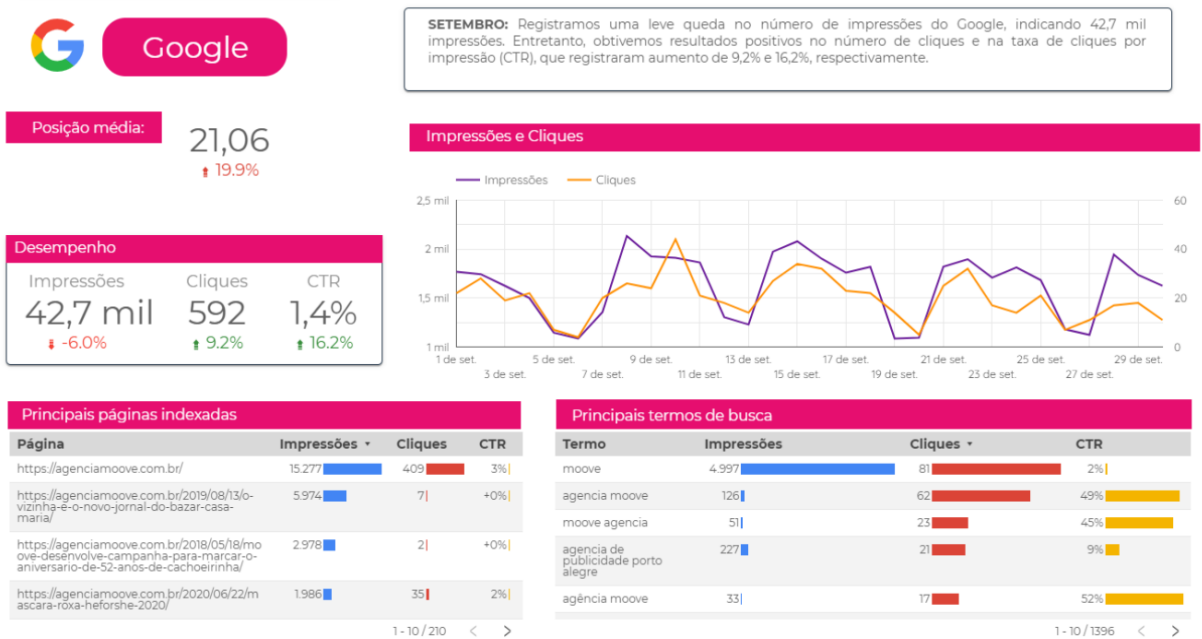
Assim como há uma variedade extensa de tecnologias para concepção da ferramenta, a aplicabilidade dela se estende a diversos segmentos e *layouts*, como exemplos, na exibição de dados referentes ao *Covid-19* (Figura 6), e no monitoramento de dados do *Google Analytics* (Figura 7).

Figura 6 – Painel Coronavírus



Fonte: Ministerio da Saude - Brasil (2023)

Figura 7 – Painel Google Analytics



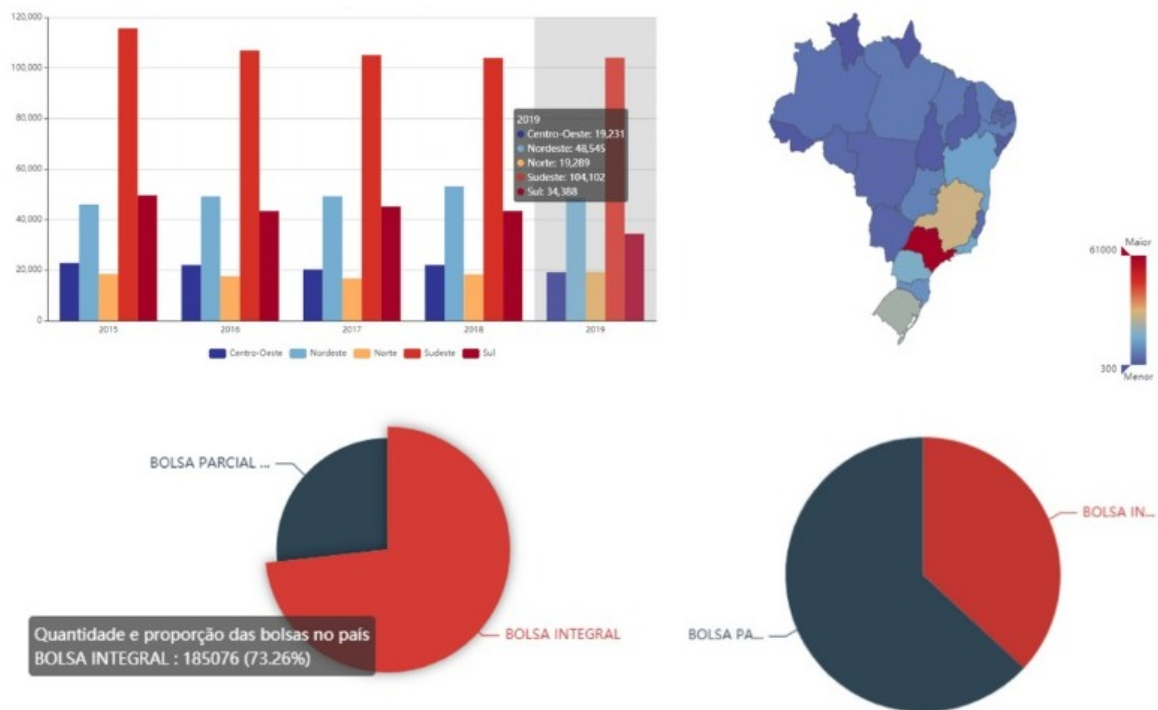
Fonte: Fuscaldo (2022)

3 TRABALHOS RELACIONADOS

Neste capítulo são descritos trabalhos encontrados através de buscas realizadas em bibliotecas digitais, como Google Acadêmico, *SciELO* e Periódicos *Capes*, direcionadas aos trabalhos com ligação direta ao tema proposto na presente monografia.

Trierveiler (2020) propõem uma aplicação web para exibição de dados do PROUNI, retirados do Portal Brasileiro de Dados, utilizando *PHP*, *JavaScript* e a mesma biblioteca utilizada no presente trabalho, a *Chart.js*. Fazendo uso das tecnologias mencionadas, o autor apresenta quatro visualizações distintas (Figura 8), onde é possível visualizar a quantidade de bolsas por estado brasileiro, por região, por instituição de ensino, além de a proporção de bolsas parciais e integrais, permitindo a visualização e análise dos dados de forma clara e direta, algo difícil de se fazer através das planilhas de dados brutos expostas no Portal Brasileiro de Dados Abertos.

Figura 8 – Visualizações geradas



Fonte: Trierveiler (2020)

Santos *et al.* (2016) entregaram um estudo sobre como as técnicas de visualização de dados podem ajudar os jornalistas a compreender e extrair informações dos dados. No trabalho os tweets sobre a Copa do Mundo de 2014 foram coletados e processados, e as informações obtidas foram apresentadas aos jornalistas através de visualizações em grupos focais. A análise dos comentários revelou que as visualizações são consideradas uma ferramenta valiosa para o jornalismo e que as visualizações interativas são preferenciais. Além disso, as visualizações podem ser usadas como um meio de iniciar análises e devem ser acompanhadas de explicações contextuais.

No estudo apresentado Santos *et al.* (2017), é analisado como visualizações interativas de dados do Twitter podem beneficiar jornalistas e profissionais de mídia. Os dados foram coletados durante as eleições presidenciais de 2014 no Brasil. Foi realizada uma pesquisa com 50 participantes para entender se as informações obtidas das visualizações interativas apresentadas podem apoiar atividades de trabalho de jornalistas e profissionais de mídia. A análise revelou que os participantes consideraram as visualizações de dados um recurso valioso em suas atividades de trabalho, mas também apontou que a visualização de dados precisa ser melhor introduzida nas práticas de trabalho desses profissionais.

Obach e Jaques (2019) propuseram um dashboard para um sistema tutor inteligente baseado em passos, objetivando permitir que o professor tenha acesso às informações coletadas pelo sistema tutor, a fim de ajudá-lo a compreender as necessidades de seus alunos. Esse sistema tem como diferencial a utilização de técnicas de mineração de dados para apresentar as informações ao professor de forma clara e intuitiva, por meio de uma interface gráfica. A contribuição desse trabalho é justamente a combinação de técnicas de mineração de dados com ferramentas de visualização para fornecer informações valiosas ao professor, permitindo-o tomar decisões mais informadas sobre seus alunos.

Terra (2015) apresenta um protótipo de extrator de dados do trânsito da Serra Gaúcha, região metropolitana do Rio Grande do Sul, utilizando a rede social Twitter como fonte. A partir desses dados coletados, o autor utilizou um minerador de dados para traçar perfis do trânsito da localidade mencionada, concluindo que é possível enriquecer as informações relacionadas com o trânsito a partir de perfis traçados através de dados provenientes de redes sociais.

Perante os trabalhos relacionados neste capítulo, é possível afirmar a importância de uma coleta de dados direcionada, assim como comprovar a relevância e impacto da visualização na análise dos dados. A proposta desta monografia visa utilizar visualização de dados personalizada, buscando facilitar o acompanhamento das coletas e auxiliar na análise assertiva dos dados coletados.

4 NOS FIOS DAS CONTROVÉRSIAS: UM ESTUDO COMPARATIVO SOBRE MÍDIA, FUTEBOL E SUAS DIMENSÕES POLÍTICAS EM BELO HORIZONTE E MONTEVIDÉU

O projeto de pesquisa “Nos fios das controvérsias: um estudo comparativo sobre mídia, futebol e suas dimensões políticas em Belo Horizonte e Montevideú”¹ é coordenado por um professor da Universidade Federal de Alfenas (curso de Jornalismo) e conta com a colaboração de uma professora da Universidade Federal dos Vales do Jequitinhonha e Mucuri (curso de Sistemas de Informação). Ele possui duração de dois anos (2022 a 2024) e propõe pensar o tema da política no universo do futebol por meio da mídia, numa perspectiva que compara as abordagens de jornalistas esportivos que trabalham em rádios em Belo Horizonte e Montevideú (Tabela 2).

O recorte para selecionar os jornalistas foram: Jornalistas Homens, que trabalham em rádio e possuem mais de 100K seguidores na rede social *Twitter*. A proposta é seguir as controvérsias por meio das postagens de jornalistas de Belo Horizonte e Montevideú e seus desdobramentos, como:

- I. as mensagens publicadas por eles,
- II. os comentários sobre a mensagem,
- III. os *likes* dos jornalistas,
- IV. os *retweets* dos jornalistas na rede social.

As postagens são coletadas de forma manual e de forma automatizada, utilizando a API do *Twitter*. De acordo com Pacheco *et al.* (2022), o Projeto Nos Fios tem como objetivo utilizar a antropologia para compreender como a política e o futebol estão inter-relacionados. O futebol é um esporte globalmente popular, podendo ser utilizado como meio para expressar opiniões políticas, criar identidades culturais e unir ou dividir grupos sociais. A antropologia, como disciplina que se concentra no estudo das sociedades e culturas humanas, é uma ferramenta ideal para analisar como o futebol e a política se relacionam e influenciam a sociedade. Por exemplo, é possível estudar como o futebol é utilizado como ferramenta de propaganda política, como os jornalistas esportivos se posicionam, como os clubes de futebol são utilizados como símbolos de identidade nacional, como as práticas esportivas podem ser usadas para resolver conflitos políticos, e como as políticas públicas podem afetar as práticas esportivas, incluindo problemas como as brigas de torcida.

Inicialmente, em entrevista com o professor coordenador do Projeto, quando questionado sobre as visualizações de dados do Projeto, ele informou que ainda estavam atrasados nesse aspecto, focando apenas na coleta, o que estava deixando-o receoso, dado a quantidade de dados para analisar, destacando ainda que essa era a sua primeira vez trabalhando com dados provenientes de redes sociais.

Utilizando os dados coletados pelo Projeto mencionado, o foco deste trabalho é entregar um *dashboard*, para visualizações dos dados de forma objetiva, onde seja possível obter

¹ Edital 01/2021 – Demanda Universal - FAPEMIG. PROCESSO N°: APQ-01761-21

uma visão geral rápida e fácil compreensão dos dados. Considerando que o Projeto está em fase inicial de coleta e tratamento dos dados, o *dashboard* proposto é apenas um protótipo para inicialização da visualização das informações.

Tabela 2 – Jornalistas

Localidade	Usuário Twitter
BH	@hevertonfutebol
BH	@Igortep
BH	@joaovitorxavier
BH	@thiagoreisbh
BH	@juniorbrasil
BH	@samuelvenancio
BH	@mariocaixa
BH	@maisvibrante
BH	@claudiorez
BH	@alvarodamiao
Montevidéu	@Buysan
Montevidéu	@Roberto_Moar
Montevidéu	@SebasGiovannelli
Montevidéu	@MartinCharquero
Montevidéu	@KesmanAlberto
Montevidéu	@MKesman
Montevidéu	@RodrigoRomano76

Fonte: Adaptado (PACHECO *et al.*, 2022)

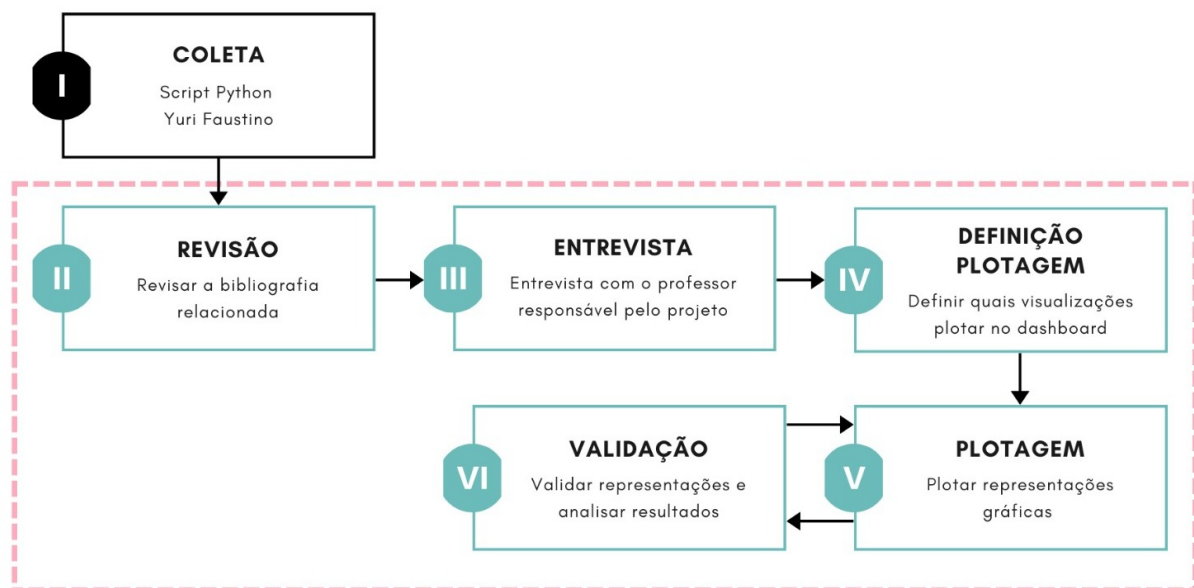
5 METODOLOGIA

A monografia em questão é uma pesquisa qualitativa aplicada, cujo propósito é solucionar um problema atual através da aplicação prática e metodológica. Para atingir esse objetivo, foram empregadas técnicas de Ciências de Dados, Engenharia de Software e Visualização de Dados, utilizando como fonte de dados a plataforma *Twitter*.

O referencial teórico é baseado em dissertações, livros, artigos científicos, artigos de conteúdo técnico (como documentações do *Twitter*) e monografias.

Para atingir o objetivo geral, o trabalho foi dividido em 06 (seis) etapas, conforme consta na Figura 9.

Figura 9 – Metodologia



Fonte: Elaborada pelo autor (2023)

- I. Coletar dados no *Twitter* via *script* (ação feita por integrantes do Projeto Nos Fios);
- II. Revisar a bibliografia de trabalhos relacionados existentes;
- III. Realizar entrevista semi-estruturada com o Professor Dr. Leonardo Turchi Pacheco, com o objetivo de elicitare os dados primordiais para análise dos dados do Projeto Nos Fios;
- IV. Definir quais dados e gráficos a serem expostos no *dashboard* com base na entrevista mencionada no item anterior;
- V. Plotar gráficos e informações utilizando os arquivos *.json* disponibilizados pelo Projeto Nos Fios;
- VI. Apresentar as representações gráficas geradas e analisá-las com base na bibliografia existente.

No desenvolvimento da aplicação web, em formato de *dashboard*, para exibição dos dados extraídos do twitter, utilizamos a *IDE Visual Studio Code*¹, *PostgreSQL*², a versão 1.5.0

¹ <https://visualstudio.microsoft.com/pt-br/>

² <https://www.postgresql.org/>

do *template Argon Dashboard Angular*³, a biblioteca *Chart.JS*⁴, a linguagem de programação *TypeScript*⁵, os *frameworks Angular*⁶ e *Express*⁷, *Node.js*⁸, e as linguagens de marcação *HTML* e *CSS*. Quanto para a hospedagem da aplicação na *web*, foi utilizado as plataformas *Vercel*⁹ e *Railway*¹⁰.

³ (<https://www.creative-tim.com/product/argon-dashboard-angular>)

⁴ (<https://www.chartjs.org/>)

⁵ (<https://www.typescriptlang.org/>)

⁶ (<https://angular.io/>)

⁷ (<https://expressjs.com/pt-br/>)

⁸ (<https://nodejs.org/en/>)

⁹ (<https://vercel.com/>)

¹⁰ (<https://railway.app/>)

6 RESULTADOS

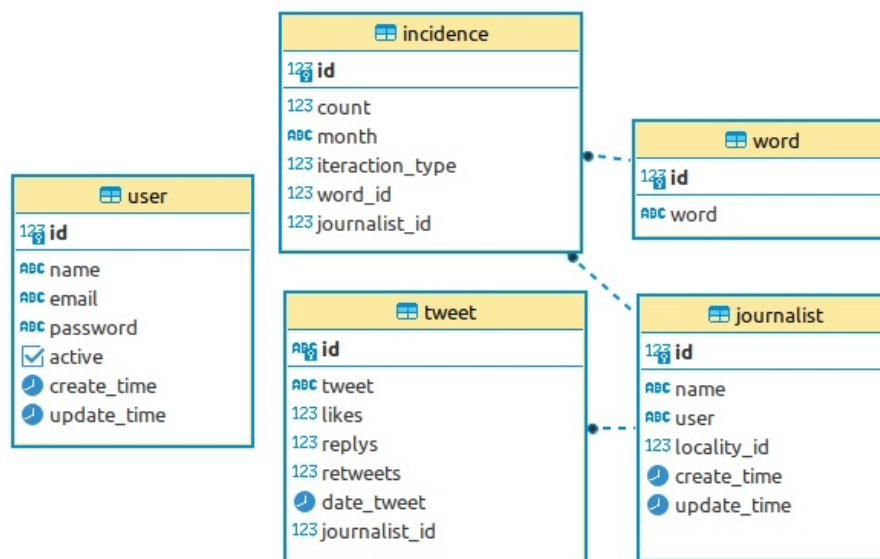
Nesta seção são apresentados os gráficos, filtros e demais visualizações geradas para o *dashboard* proposto, além da estrutura do banco de dados PostgreSQL. Para desenvolvimento da aplicação web fruto do presente trabalho, utilizou-se como fonte de dados o *dataset* gerado através de extração realizada pelo *script* de Santos (2022), disponibilizado pelo Projeto Nos fios das controvérsias: um estudo comparativo sobre mídia, futebol e suas dimensões políticas em Belo Horizonte e Montevideú, já para concepção da parte gráfica, utilizou-se as tecnologias *HTML*, *CSS*, *TypeScript*, *Node.js*, *Angular* e *PostgreSQL*, resultando em uma ferramenta de design intuitivo, prático e funcional.

O código fonte do *dashboard* está disponível no GitHub^{1,2} para consulta e uso. Os usuários interessados podem acessar o repositório para visualizar e utilizar o código para fins educacionais ou comerciais. O código está escrito em linguagem de programação especificada e pode ser facilmente modificado de acordo com as necessidades do usuário. Além disso, é possível contribuir para o desenvolvimento do código, enviando solicitações de *pull* e colaborando com a comunidade.

6.1 O Banco de Dados

Para o desenvolvimento da base de dados consumida pelo painel da aplicação foi utilizado o sistema gerenciador de banco de dados objeto relacional *PostgreSQL*. A base de dados é composta por 5 (cinco) tabelas principais, que auxiliam tanto o processo de autenticação, quanto a geração das estatísticas para o painel. As tabelas criadas foram denominadas como: *user*, *word*, *journalist*, *tweet* e *incidence*, conforme consta no diagrama da Figura 10.

Figura 10 – Diagrama Relacional



Fonte: Elaborada pelo autor (2023)

¹ <<https://github.com/calinybasilio/dashboard-tcc-interface>>

² <<https://github.com/calinybasilio/dashboard-tcc-api>>

A tabela *user* armazena as informações dos usuários que podem acessar a aplicação, com isso sua principal utilização é nesse processo de autenticação e liberação do acesso ao painel.

A tabela *word* armazena todas as palavras chaves destacadas dos tweets no processo de coleta, e posteriormente são relacionadas às incidências das mesmas e com os jornalistas para gerar as estatísticas gerais, como a tabela de palavras.

A tabela *journalist* armazena as informações dos jornalistas, dentre elas a identificação do usuário do twitter, localidade (Belo horizonte ou Montevideú), dentre outras. Ela também é utilizada para auxiliar no processo de geração das estatísticas, como no gráfico de número de jornalistas por localidade e tabela de jornalistas.

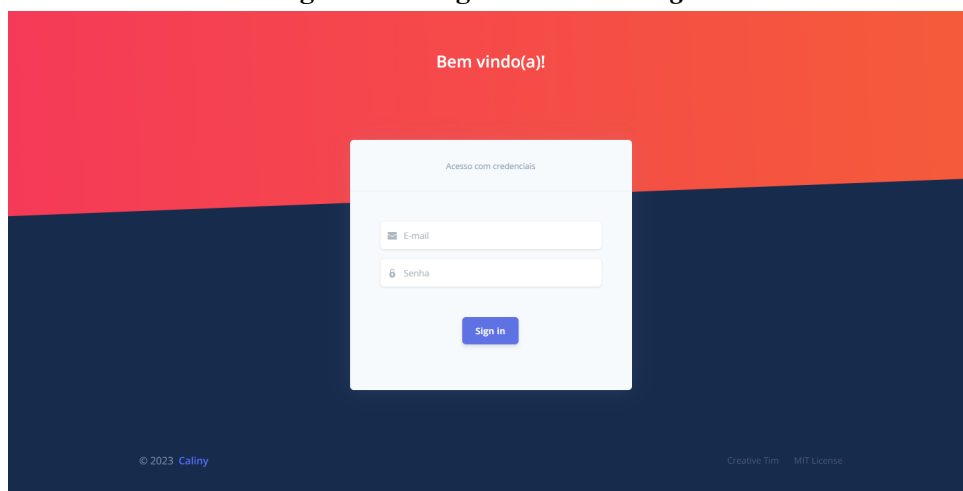
A tabela de *incidence* armazena a quantidade de aparições das palavras nos *tweets* de cada jornalista, separada pelo mês da postagem do *tweet*. Essa tabela, em conjunto das tabelas de *word* e *journalist*, é a principal fonte de auxílio da geração das estatísticas, utilizada tanto na tabela de palavras quanto no gráfico de incidência dos *tweets*.

E por fim, temos a tabela de *tweet*, a qual armazena as informações gerais dos *tweets* coletados, como o próprio texto, relação com o jornalista, a data de publicação/coleta, quantidade de *likes* e *replies* recebidos, dentre outras. Ela é utilizada nas estatísticas de número total de *tweets* coletados, período das coletas e número de *tweets* coletados por mês.

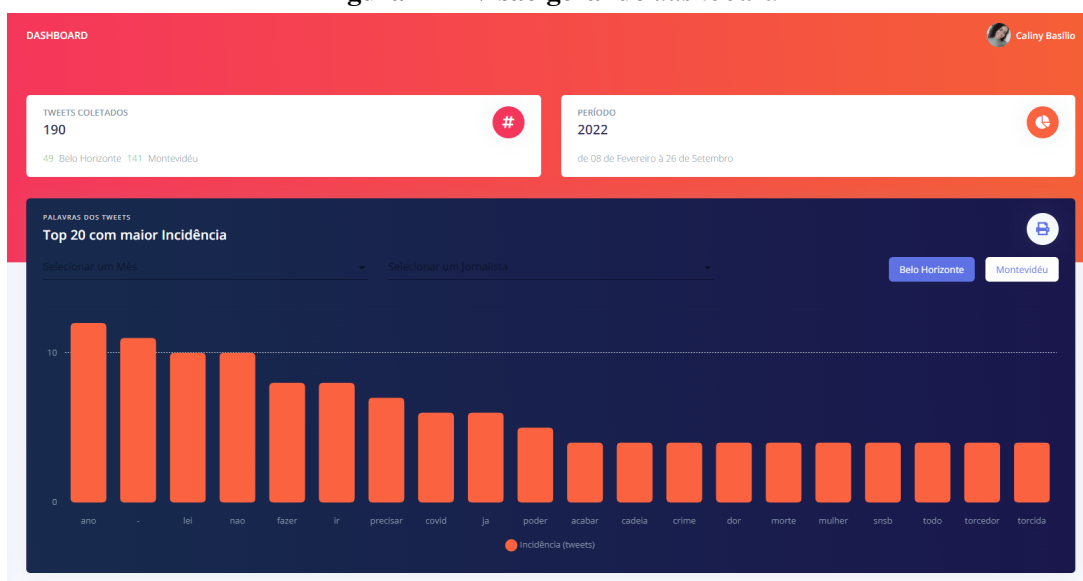
6.2 O Dashboard

Ao acessar o *dashboard* é necessário inserir as credenciais de *login*, conforme consta na Figura 11, previamente cadastradas pelo administrador. Após efetuar o *login*, o usuário é direcionado a página do *dashboard*, onde ele tem a visão geral de das representações gráficas geradas (Figura 12):

- I. Contador e Período de Coleta
- II. Gráfico de Pizza Jornalista por localidade;
- III. Gráfico de Barras Incidência de Palavras (*tweets* jornalistas);
- IV. Gráfico de Barras Incidência de Palavras (*replies*);
- V. Gráfico de Barras Incidência de Palavras (*likes*);
- VI. Gráfico de Linhas Coleta *tweets*
- VII. Tabela Incidência de Palavras;
- VIII. Tabela de Jornalistas com a localidade e respectivos usuários ('@') do *Twitter*.

Figura 11 – Página Inicial de *Login*

Elaborada pelo autor (2023)

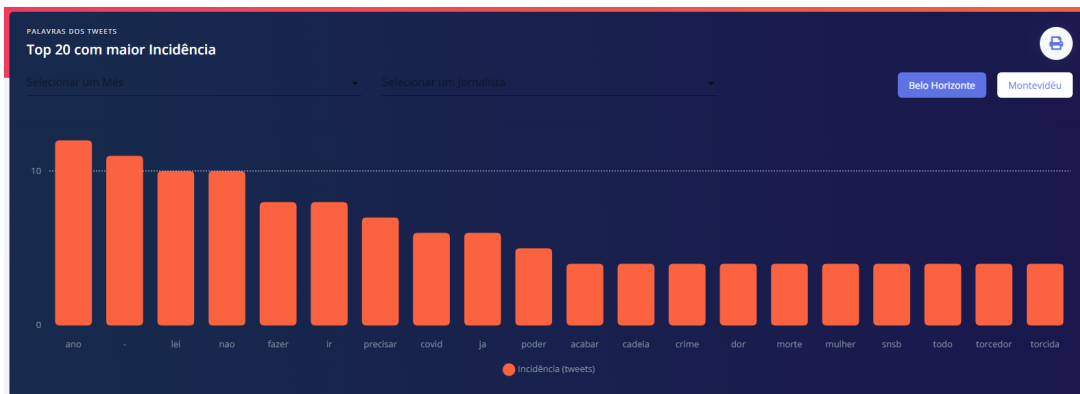
Figura 12 – Visão geral do *dashboard*

Elaborada pelo autor (2023)

6.3 Representações Gráficas Geradas

Com o propósito de exibir as palavras mais tuitadas pelos jornalistas acompanhados na análise do Projeto, criou-se um gráfico de barras onde é possível visualizar as vinte palavras mais incidentes, podendo filtrar entre jornalistas de Belo Horizonte e Montevidéu, conforme consta na Figura 13. Essa representação visual tem como intuito auxiliar o usuário a identificar padrões e reincidência nos tweets, podendo assim correlacionar os assuntos publicados com acontecimentos daquele período. Além do gráfico, gerou-se uma tabela (Figura 14) complementar ao gráfico de barras da Figura 13, onde todas as palavras tuitadas são expostas ordenadas pela quantidade de aparições.

Figura 13 – Gráfico Incidência de Palavras (*tweets* jornalistas)



Elaborada pelo autor (2023)

Figura 14 – Tabela Incidência de Palavras

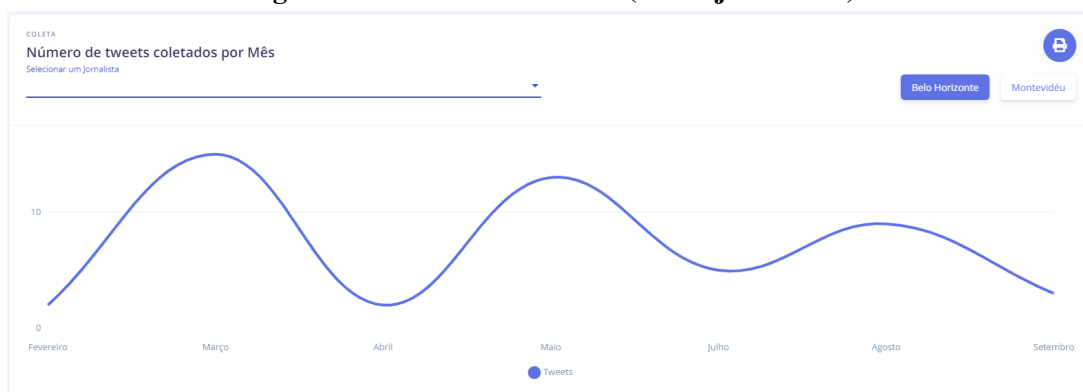
Tabela com o eixo Y rotulado 'Palavras' e o eixo X rotulado 'Incidência (tweets)'. A tabela mostra a incidência de 12 palavras diferentes.

Palavra	Incidência (tweets)
ano	12
-	11
lei	10
nao	10
fazer	8
ir	8
precisar	7
covid	6
ja	6
poder	5
acabar	4

Elaborada pelo autor (2023)

Plotou-se também um gráfico de linhas (Figura 15) onde é possível visualizar a quantidade de *tweets* em todo período da coleta, assim como filtrar por jornalista, localidade e mês das postagens. Tal gráfico é útil para mostrar a tendência dos *tweets*, assim a identificação de padrões, flutuações e mudanças na direção dos dados. Além disso, um gráfico de linhas é e fácil entendimento e interpretação, mesmo para usuários sem muita experiência em análise de dados.

Figura 15 – Gráfico de Linhas (*tweets* jornalistas)



Elaborada pelo autor (2023)

Em entrevista com o Professor Leonardo reconheceu-se a necessidade de acompanhar os “fios” gerados através dos *tweets* realizados pelos jornalistas, com isso, foi plotado o gráfico de barras com as vinte palavras mais incidentes nos *replies* (comentários) das postagens dos jornalistas. Com auxílio desse gráfico, é possível identificar as palavras mais incidentes nos *replies*, podendo assim correlacionar as palavras com acontecimentos no mundo futebolístico ou político.

Figura 16 – Gráfico Incidência de Palavras (*replies*)



Elaborada pelo autor (2023)

Considerando que os jornalistas acompanhados não se posicionam politicamente de forma direta, identificou-se a necessidade de acompanhar os conteúdos curtidos por eles, assim é possível detectar a inclinação política de cada um, em consequência disso, plotou-se o gráfico de barras com a incidência de palavras presentes nos *likes* distribuídos por esses jornalistas (Figura 17).

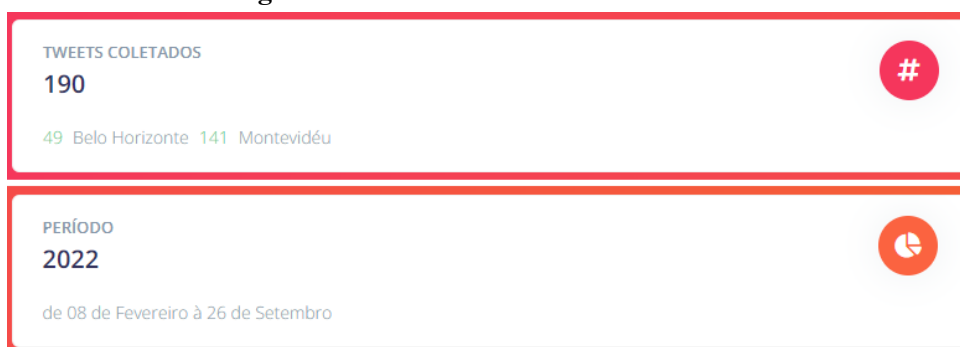
Figura 17 – Gráfico Incidência de Palavras (*likes*)



Elaborada pelo autor (2023)

No *dashboard* também é possível visualizar *cards* com o período e número de coletas realizadas (Figura 18), com isso, o usuário consegue identificar facilmente qual o período referente aos dados exibidos, além da quantidade de *tweets* coletados por localidade.

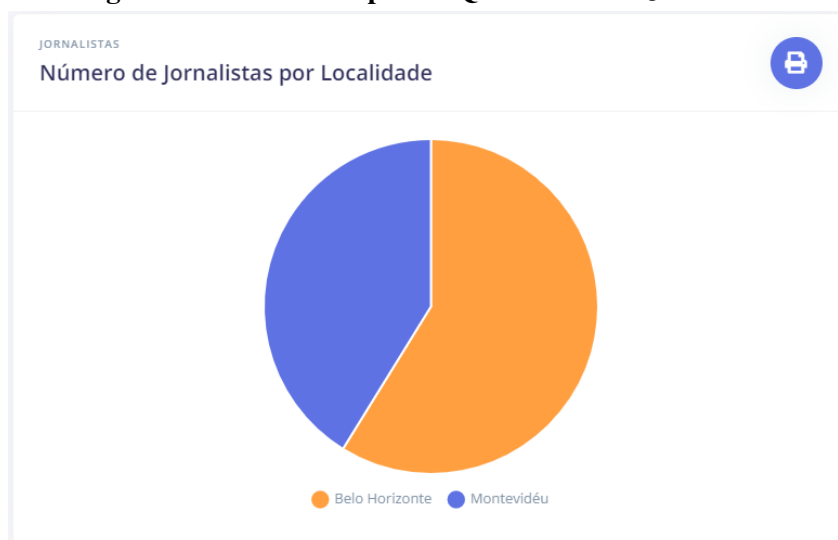
Figura 18 – Contador e Período de Coleta



Elaborada pelo autor (2023)

Plotamos também um gráfico de pizza (Figura 19), possuindo a funcionalidade *hover* (suspender) habilitada, que permite exibir informações sobre o número de jornalistas atribuídos a uma determinada localidade quando o cursor do mouse é movido sobre ela. Em complemento, criamos uma tabela (Figura 20) constando nome, usuário do *Twitter* e a respectiva localidade do jornalista.

Figura 19 – Gráfico de pizza - Quantidade de Jornalistas



Elaborada pelo autor (2023)

Figura 20 – Tabela Jornalista e Usuários

NOME	USUÁRIO (e)	CIDADE
Heverton Guimarães	@Hevertonfutebol	Belo Horizonte
Igor Assunção	@igorstep	Belo Horizonte
João Vitor Xavier	@joaovitorxavier	Belo Horizonte
Thiago Reis	@thiagoreisbh	Belo Horizonte
Junior Brasil	@juniorbrasil	Belo Horizonte
Samuel Venâncio	@samuelvenancio	Belo Horizonte
Mário Henrique Caixa	@mariocaixa	Belo Horizonte
Alberto Rodrigues	@maisvibrante	Belo Horizonte
Claudio Rezende	@claudiorez	Belo Horizonte
Ávaro Damiao	@alvarodamiao	Belo Horizonte
Buysan	@Buysan	Montevideu
Roberto Moar	@Roberto_Moar	Montevideu
Sebas Giovanelli	@SebasGiovanelli	Montevideu
Martin Charquero	@MartinCharquero	Montevideu
Alberto Kesman	@KesmanAlberto	Montevideu
Martin Kesman	@MKesman	Montevideu
Rodrigo Romano	@RodrigoRomano76	Montevideu

20 itens por página

Elaborada pelo autor (2023)

Em todas as representações gráficas geradas, exceto os *cards*, foi implementada a funcionalidade de impressão, onde quando acionada gera arquivo (*pdf*) daquele gráfico ou tabela, utilizando os filtros pré-selecionados, funcionalidade útil para compartilhamento das representações gráficas com terceiros.

7 DISCUSSÃO DOS RESULTADOS

Em geral, as representações gráficas entregues foram consideradas satisfatórias pelo responsável do projeto. Nos fios das controvérsias, sendo destacada a eficácia das visualizações em transmitir informações de forma clara e concisa, permitindo interpretar um grande volume de dados de forma dinâmica. Em entrevista com o professor Leonardo Turchi, foi sinalizado que o Projeto não havia definido como se daria a visualização dos dados coletados, estando focados apenas na coleta, o que gerava preocupação, dado ao volume de dados gerados.

Conforme exposto por Evergreen (2018), a falta de planejamento na coleta de dados pode resultar em diversos problemas, incluindo dados incompletos, inconsistentes ou até mesmo inválidos. Isso pode afetar diretamente a qualidade e a precisão das análises e visualizações, tornando-as inúteis ou pouco confiáveis. Em resumo, planejar a visualização de dados antes da coleta, é essencial para garantir que os dados sejam coletados de forma apropriada e possam ser visualizados de maneira eficaz. Isso permitirá uma análise e interpretação mais precisa dos dados e a identificação de padrões e tendências importantes.

A qualidade e eficácia das representações gráficas geradas foram limitadas aos dados fornecidos pelo Projeto, pois algumas plotagens consideradas importantes não puderam ser implementadas, devido a falta de dados ou correlação entre os parâmetros. Contudo, todos os dados coletados através da *API* do *Twitter* e disponibilizados para desenvolvimento deste trabalho, foram manipulados para gerar representações com a máxima qualidade, a fim de entregar uma ferramenta útil para auxílio nas análises de dados do referido Projeto.

Ao acessar a plataforma é solicitado login e senha para visualizar o *dashboard*, visando a segurança dos dados na *web*. Há apenas um usuário criado na ferramenta atualmente, não havendo distinção de tipos de utilizadores, uma vez que a ferramenta é destinada para uso exclusivo do Projeto referenciado, sendo responsabilidade do professor fornecer (ou não) os dados de acesso a terceiros.

No *dashboard* foi implementado um contador (Figura 18) para mensurar a quantidade total de *tweets* coletados de cada região, e um *card* com o período de coleta dos dados exibidos, visualizações para acompanhamento da base gerada e do período de coleta. Além disso, dado a importância dos gráficos interativos na exploração dos dados, conforme indicado por Healy (2018), propomos o gráfico de barras com a incidência de palavras dos *tweets* (Figura 13), onde é possível filtrar por jornalista (autor), localidade (Montevidéu e Belo Horizonte), podendo filtrar também pelos meses de incidências das palavras, além de classificação por tema que a palavra incidente está relacionada. A classificação das palavras ocorreu de forma manual, considerando as questões abordadas no *tweet* do jornalista, essencial para correlacionar os dados com o acontecimento (político ou não) que gerou aquele *tweet*.

De forma similar ao gráfico anterior (Figura 13), plotou-se um gráfico de incidência de palavras nos *likes* atribuídos pelos jornalistas (Figura 17), onde é possível filtrar por localidade, além da verificação a quantidade de aparições de cada palavra; e outro gráfico de barras com

incidência de palavras em *replies* (Figura 16) onde também é possível filtrar por localidade e visualizar as palavras mais incidentes.

Gerou-se também um gráfico de linhas (Figura 15), onde é possível filtrar por região e jornalista, apresentando as quantidades de *tweets* coletados ao longo dos meses, útil para visualizar tendências, variações e comparações, permitindo ao observador identificar rapidamente a evolução das grandezas ao longo do tempo e detectar mudanças significativas na tendência. Tufté (1983) enfatiza a importância de selecionar a representação gráfica adequada para os dados e a informação que se deseja comunicar, evitando distorções ou enganos intencionais nos dados através de escolhas inadequadas de representação gráfica.

Ainda apresentamos um gráfico de pizza (Figura 19), onde é possível consultar a quantidade de jornalistas de acordo com a localidade, em complemento, criamos a tabela 20 para visualizar nome do jornalista, localidade, e usuário do *Twitter*. Tais representações foram sinalizadas pelo professor como “importante para comparar os jornalistas”.

Em todas as representações gráficas geradas, há a funcionalidade para impressão, onde o usuário pode salvar a visualização em .pdf, caso tenha necessidade de acessar desta forma, ou até mesmo de compartilhar com algum colega. Funcionalidade útil, considerado que não há necessidade de compartilhamento de dados de acesso a plataforma para terceiros.

Ao ser questionado quanto a navegabilidade, interatividade e clareza dos elementos visuais apresentados na ferramenta, o professor avaliou com nota máxima, demonstrando satisfação com as visualizações geradas, sinalizando que elas contribuem fortemente para análise dos dados. O que corresponde ao exposto por Santos *et al.* (2016): visualização de dados permite uma análise mais aprofundada dos conjuntos de dados, tornando possível identificar tendências, padrões e exceções que podem não ser evidentes ao examinar os dados de forma manual.

Os principais questionamentos foram ao gráfico de *likes*, onde sinalizaram a necessidade de um filtro correlacionando o jornalista que curtiu o *tweet* relacionando ao usuário que tuitou, o que se repete no gráfico de *replies*, que também possui a necessidade relacionar jornalista e usuário publicante, o que de acordo com Gelman e Hill (2007) é uma técnica importante, pois com a correlação pode-se identificar relações entre diferentes variáveis, por mais que apenas isso não prove causalidade, é um ponto de partida importante para investigar se há alguma relação entre os dados presentes. As funcionalidades mencionadas não foram implementadas, devido a base de dados disponibilizados pelo Projeto não conterem parâmetros para correlacioná-los desta forma.

Em alguns casos, os jornalistas selecionados não se posicionam politicamente de forma explícita, sendo necessário acompanhar além dos *tweets*, os *replies* e *likes*, para compreender o posicionamento deles conforme conteúdo presente nas interações. Dado a esse fato, faz-se necessário ajustar o *script* de coleta, tendo em vista que com o *script* atual não são coletados dados suficientes para correlacioná-los neste sentido.

8 CONSIDERAÇÕES FINAIS

Ao longo deste trabalho, foi possível compreender a importância da visualização e análise de dados na tomada de decisões e na compreensão de fenômenos complexos. A capacidade de processar e exibir grandes conjuntos de informações de maneira clara, interativa e intuitiva tem permitido aos profissionais de negócios, cientistas, pesquisadores e outros especialistas a tomarem decisões baseadas em dados.

A revisão bibliográfica mostrou que a visualização de dados é primordial para transmitir informações de bases extensas ou complexas, de maneira clara e concisa. Além disso, os avanços tecnológicos, como a computação em nuvem, *big data* e a inteligência artificial, permitiram o acesso e análise de grandes quantidades de dados em tempo real, o que tem possibilitado a criação de visualizações avançadas e interativas.

Um *dashboard* foi apresentado para visualização de dados provenientes do *Twitter*, sendo tais dados coletados utilizando a *API* do *Twitter* e fornecidos pelo Projeto “Nos fios das controvérsias: um estudo comparativo sobre mídia, futebol e suas dimensões políticas em Belo Horizonte e Montevidéu”. Após a finalização e entrega da ferramenta contendo quatro gráficos interativos, duas tabelas, um gráfico estático, dois *cards* informativos, foi aplicado um formulário ao professor responsável pelo Projeto, a fim de entender se a aplicação atendia as necessidades do Projeto na análise dos dados. Verificando as respostas obtidas, constatou-se que as visualizações geradas foram consideradas satisfatórias, entretanto, houve solicitação de inserção de filtros de data, além de filtro correlacionando o autor receptor do *like/reply*, e autor do *tweet*, nos gráficos de *likes* (Figura 17) e *replies* (Figura 16). Além de uma solicitação para plotagem de representação gráfica onde constasse os usuários que mais recebiam *likes* de cada jornalista. Contudo, nos dados coletados e disponibilizados pelo Projeto não havia informações suficientes para correlacionar os dados, e realizar as inserções solicitadas.

Em síntese, as representações gráficas geradas foram consideradas pertinentes para auxílio na análise dos dados coletados pelo Projeto Nos Fios, todavia, se o planejamento da visualização tivesse ocorrido antes da coleta, seria possível selecionar os dados mais relevantes e garantir que fossem coletados e organizados de forma adequada. Isso permitiria uma visualização e análise mais eficiente, entregando melhor comunicação dos resultados. Além disso, ao planejar a visualização, é possível identificar possíveis limitações dos dados e planejar soluções para superá-las.

Com base nos *feedbacks* recebidos e necessidades do Projeto Nos Fios, foram identificados *insights* para trabalhos futuros: ajuste do *script* para coleta de dados inclinada às visualizações, como plotagem de gráficos com correlação entre jornalista e usuário receptor do *like*, assim como em *replies*, além de coletar dados para plotagem de gráfico que indique quais os usuários recebem mais *likes* de cada jornalista.

REFERÊNCIAS

- ARBEX, L. F. S. Visualização dos dados estatísticos da uerj: proposta de dashboards baseados no trabalho de jacques bertin. **Universidade do Estado do Rio de Janeiro**, 2013. Disponível em: <https://www.bddd.uerj.br:8443/handle/1/9108>. Acesso em: 01 jan. 2023.
- BELTRAMINI, F.; GUNTHER, L. E. O "novo petróleo" da sociedade contemporânea. **Revista Relações Internacionais do Mundo Atual Unicuritiba**, v. 1, n. 34, 2022. Disponível em: <http://dx.doi.org/10.21902/Revrma.v1i34.5736>. Acesso em: 11 jan. 2023.
- BOEHMKE, B. C. **Data Wrangling with R**. [S.l.]: Springer, 2016.
- BOYD, D. M.; ELLISON, N. B. Social network sites: Definition, history, and scholarship. **Journal of Computer-Mediated Communication**, v. 13, p. 210–230, 2008. Disponível em: <https://doi.org/10.1111/j.1083-6101.2007.00393.x>. Acesso em: 11 jan. 2023.
- CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. C. **Readings in Information Visualization: Using Vision to Think**. [S.l.]: Academic Press, 1999.
- CHABOT, C. **The key to data visualization: Interactivity**. 2013. Disponível em: <https://www.tableau.com/about/blog/2013/10/key-data-visualization-interactivity-18269>. Acesso em: 10 jan. 2023.
- CHEN, H.; CHIANG, R. H.; STOREY, V. C. Business intelligence and analytics: From big data to big impact. **Journal of Computer-Mediated Communication**, v. 36, n. 4, p. 24, 2012. Disponível em: <https://doi.org/10.2307/41703503>. Acesso em: 11 jan. 2023.
- DANG, S. Twitter frustra previsões de crescimento de usuários. **Folha de São Paulo**, p. 1, 2021. Disponível em: <https://www1.folha.uol.com.br/mercado/2021/10/twitter-frustra-previsoes-de-crescimento-de-usuarios.shtml>. Acesso em: 05 mar. 2022.
- EVERGREEN, S. D. H. **Presenting data effectively**. [S.l.]: Sage Publishing, 2018.
- FALSARELLA, O. M.; JANNUZZI, C. A. S. C.; SUGAHARA, C. R. Gestão da estratégia empresarial: Proposição de um modelo de monitoramento informacional na era big data. **Revista Digital Biblioteconomia e Ciência da Informação**, v. 15, p. 420–441, 2017. Disponível em: <http://eprints.rclis.org/31862/1/8647124-26242-5-PB.pdf>. Acesso em: 05 jan. 2023.
- FERREIRA, G. C. Redes sociais de informação: uma história e um estudo de caso. **Perspectivas em Ciência da Informação**, v. 16, n. 3, p. 208–231, 2011. Disponível em: <https://doi.org/10.1590/S1413-99362011000300013>. Acesso em: 11 dez. 2022.
- FEW, S. Dashboard confusion. **Perceptual Edge**, p. 1–4, 2004. Disponível em: https://www.perceptualedge.com/articles/ie/dashboard_confusion.pdf. Acesso em: 10 jan. 2023.
- FEW, S. **Information Dashboard Design: The Effective Visual Communication of Data**. Sebastopol, CA: O'Reilly, 2006.
- FRIENDLY, M.; DENIS, D. J. Milestones in the history of thematic cartography, statistical graphics, and data visualization. **Journal of Computer-Mediated Communication**, p. 13, 2001. Disponível em: <http://www.datavis.ca/milestones>. Acesso em: 17 jan. 2023.
- FUSCALDO, G. M. **Dashboards de Marketing: como nós fazemos na Moove**. 2022. Disponível em: <https://marketing.twitter.com/pt/insights/how-social-data-informs-business-decisions-and-drives-social-change>. Acesso em: 10 jan. 2023.

GELMAN, A.; HILL, J. **Data analysis using regression and multilevel/hierarchical models**. [S.l.]: Cambridge University Press, 2007.

GLUCK, E. **Como os dados das redes sociais influenciam as decisões das empresas e promovem mudanças sociais**. 2017. Disponível em: <https://marketing.twitter.com/pt/insights/how-social-data-informs-business-decisions-and-drives-social-change>. Acesso em: 13 jan. 2023.

HEALY, K. **Data Visualization: A Practical Introduction**. [S.l.]: Princeton University Press, 2018.

MANOEL, V. R. Desenvolvimento de uma ferramenta de apoio à gestão utilizando business intelligence. **Universidade Estadual Paulista**, 2020. Disponível em: <https://www.sorocaba.unesp.br/Home/Graduacao/EngenhariadeControleeAutomacao/galdenoro1906/revisado-tcc---vinicius-rodrigues-manoel-2.pdf>. Acesso em: 02 jan. 2023.

MARTELETO, R. M. Análise de redes sociais - aplicação nos estudos de transferência da informação. **Instituto Brasileiro de Informação em Ciência e Tecnologia. Ciência da Informação. Brasília**, v. 30, n. 2, p. 71–81, 2001. Disponível em: <https://doi.org/10.1590/S0100-19652001000100009>. Acesso em: 11 jan. 2023.

MARTINS, D. **As 10 Redes Sociais mais usadas no Brasil em 2022**. 2022. Disponível em: <https://rockcontent.com/br/blog/as-maiores-redes-sociais/>. Acesso em: 11 jan. 2023.

MEIO&MENSAGEM. **História do Twitter: da origem da rede social até a compra por Elon Musk**. 2022. Disponível em: <https://www.meioemensagem.com.br/proxima/pxx-noticias/historia-do-twitter>. Acesso em: 13 jan. 2023.

MINISTERIO DA SAUDE - BRASIL. **Dashboard Coronavírus**. 2023. Disponível em: <https://covid.saude.gov.br/>. Acesso em: 08 jan. 2023.

OBACH, R. L.; JAQUES, P. A. Um dashboard educacional para um sistema tutor baseado em passos. **VIII Congresso Brasileiro de Informática na Educação**, 2019. Disponível em: [10.5753/cbie.sbie.2019.1561](https://doi.org/10.5753/cbie.sbie.2019.1561). Acesso em: 08 jan. 2023.

ORACLE. **O que é Big Data?** 2022. Disponível em: www.oracle.com/br/big-data/what-is-big-data/. Acesso em: 05 mar. 2022.

PACHECO, L. T.; SAMONETTO, B. J.; SANTOS, Y. S. dos; SANTOS, C. Q. Gt11: Antropologia das práticas esportivas e de lazer: Destecendo fios e costurando controvérsias: esporte e política nos twitters de jornalistas de belo horizonte e montevidéu. **Reunião Brasileira de Antropologia**, v. 33, 2022. Disponível em: encurtador.com.br/pzHY1. Acesso em: 13 jan. 2023.

PRIMAK, F. V. **Decisões com BI (Business Intelligence)**. [S.l.]: Ciência Moderna, 2008.

RIBEIRO, T. V. Análise e visualização de dados aplicadas à evolução de projetos de software. **Universidade Federal de Pernambuco**, 2021. Disponível em: <https://repositorio.ufpe.br/handle/123456789/40513>. Acesso em: 11 dez. 2022.

SANT. **Coragem**. 2015. Disponível em: <https://open.spotify.com/track/4zbsPUNxO06xPhI0uufgir?si=7fb30effd5f04eb8>. Acesso em: 08 dez. 2022.

SANTOS, C. Q.; CUNHA, H. S. da; TEIXEIRA, C. R. G.; SOUZA, D. R. de; TIETZMANN, R.; MANSSOUR, I. H.; SILVEIRA, M. S.; RUIZ, D. D. A.; TRASEL, M.; BARROS, R. C. Media professionals' opinions about interactive visualizations of political polarization during brazilian presidential campaigns on twitter. **Proceedings of the 50th Hawaii International Conference on System Science**, 2017. Disponível em: <http://hdl.handle.net/10923/13267>. Acesso em: 11 jan. 2023.

SANTOS, C. Q.; TIETZMANN, R.; TRASEL, M.; MORAES, S. M. W.; MANSSOUR, I. H.; SILVEIRA, M. S. Can visualization techniques help journalists to deepen analysis of twitter data? exploring the ? germany7x1 brazil? case. **Proceedings of the 49th Annual Hawaii International Conference on System Sciences**, p. 1530–1605, 2016. Disponível em: <http://hdl.handle.net/10923/13814>. Acesso em: 01 jan. 2023.

SANTOS, Y. S. dos. **Minerador - Nos Fios da Controvérsia**. 2022. Disponível em: <https://github.com/YuriSilv/Minerador---Nos-Fios-da-Controversia>. Acesso em: 28 nov. 2022.

SILVA, A. B. O.; FERREIRA, M. A. T. Gestão do conhecimento e capital social: as redes e sua importância para as empresas. **Informação & Informação**, v. 12, n. 1 - esp., p. 125–156, 2007. Disponível em: <https://doi.org/10.5433/1981-8920.2007v12n1espp125>. Acesso em: 11 jan. 2023.

SMAAL, B. **A história do Twitter**. 2010. Disponível em: <https://www.tecmundo.com.br/rede-social/3667-a-historia-do-twitter.htm>. Acesso em: 13 jan. 2023.

TAURION, C. **Big Data**. Rio de Janeiro, RJ: BRASPORT Livros e Multimídia Ltda. Livros, 2013.

TERRA, E. D. Ferramenta para extração de dados do twitter para mineração de dados. **Universidade de Caixias do Sul**, 2015. Disponível em: <https://repositorio.ucs.br/handle/11338/1124?locale-attribute=es>. Acesso em: 14 dez. 2022.

TRIERVEILER, M. F. A. Visualização de dados abertos educacionais do prouni com chartjs. **Universidade Federal de Santa Catarina**, 2020. Disponível em: <https://repositorio.ufsc.br/handle/123456789/218418t>. Acesso em: 14 dez. 2022.

TUFTE, E. R. **The Visual Display of Quantitative Information**. [S.l.]: Graphics Press, 1983.

TUKEY, J. W. **Exploratory data analysis**. [S.l.]: Reading, MA: Addison-Wesley, 1977.

TWITTER. **Twitter é o lugar certo**. 2022. Disponível em: <https://about.twitter.com/pt>. Acesso em: 14 dez. 2022.

TWITTER. **Twitter**. 2023. Disponível em: <https://twitter.com/>. Acesso em: 13 jan. 2023.

9 SEMI-ESTRUTURA ENTREVISTA INICIAL

Objetivando compreender o projeto “Nos fios das controvérsias: um estudo comparativo sobre mídia, futebol e suas dimensões políticas em Belo Horizonte e Montevideu”, sua motivação, objetivos, a coleta de dados e visualização deles, fizemos uma entrevista semi-estruturada com o Prof. Leonardo Turchi seguindo o seguinte escopo:

1. Qual o objetivo do projeto “Nos fios das controvérsias: um estudo comparativo sobre mídia, futebol e suas dimensões políticas em Belo Horizonte e Montevideu” e a motivação dele?
2. A coleta dos dados está ocorrendo de quais formas?
3. Quais os dados considerados mais importantes na análise?
4. Como feita essa categorização de importância dos dados a serem coletados?
5. Como foram definidos os campos de pesquisa (racismo, feminismo, nazismo)?
6. A visualização dos dados é de suma importância para a análise dos dados. Como ocorre essa visualização atualmente?
7. Quais dados julga importante visualizar? E quais os filtros ou possíveis dados cruzados poderiam facilitar o processo de análise?
8. O que o projeto pretende entregar como contribuição para a sua área de pesquisa?

10 FORMULÁRIO VALIDAÇÃO DO *DASHBOARD*

Ao finalizar o *dashboard*, apresentamos a ferramenta ao Prof. Leonardo, e posteriormente enviamos um formulário para validação, onde constava as seguintes perguntas:

1. Como julga a navegabilidade do *dashboard*?
2. Como classifica a clareza do elementos visuais apresentados na ferramenta?
3. Os botões de navegação e outros elementos interativos estão funcionando corretamente?
4. O quão pertinentes são as representações visuais geradas para atender o objetivo geral do seu projeto?
5. Há alguma consideração que deseja adicionar sobre as questões abordadas na pergunta anterior?
6. Descreva qual a sua percepção sobre o gráfico de barras onde constam as incidências de palavras dos *tweets* de cada jornalista, considere os filtros disponíveis na ferramenta (insira todas as informações e *feedbacks* que julgar necessário).
7. Descreva qual a sua percepção sobre o gráfico de barras onde constam as incidências de palavras dos *replies*, considere os filtros disponíveis na ferramenta (insira todas as informações e *feedbacks* que julgar necessário).
8. Descreva qual a sua percepção sobre o gráfico de barras onde constam as incidências de palavras dos *likes* nos *tweets* (de terceiros), considere os filtros disponíveis na ferramenta (insira todas as informações e *feedbacks* que julgar necessário).
9. Previamente foi solicitado que no *dashboard* constasse uma tabela com os dados dos jornalistas (localidade e usuário). Informe qual a importância dessa visualização na ferramenta e repasse o seu *feedback* quanto layout da tabela apresentada.
10. Há algum elemento visual (como gráfico ou tabela) que julgue necessário **alterar** no *dashboard*?
11. Há algum elemento visual (como gráfico ou tabela) que julgue necessário **adicionar** ao *dashboard*?
12. Insira no campo baixo todas as considerações adicionais, seja relacionada as perguntas apresentadas, sugestão de melhoria ou *feedbacks*.

