

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE COMUNICAÇÕES E ARTES
DEPARTAMENTO DE JORNALISMO E EDITORAÇÃO

BIANCA APARECIDA MUNIZ PEREIRA

DEZ ANOS DE COBERTURA RACIAL DO JORNAL FOLHA DE SÃO PAULO POR
MEIO DE TÍTULOS DE REPORTAGENS: UMA ABORDAGEM COM JORNALISMO
DE DADOS

Relatório de pesquisa apresentado para a disciplina
“Projeto Experimental em Jornalismo”, realizado sob
orientação da Profa. Dra. Dayana Karla Melo da Silva.

SÃO PAULO

2023

SUMÁRIO

1	INTRODUÇÃO.....	3
1.1	O início do Jornalismo de Dados	4
1.2	Jornalismo de dados no Brasil.....	6
1.3	Raça e racismo.....	7
1.4	Cor e raça na população brasileira	8
1.5	Raça e imprensa	10
1.6	Título e pauta.....	11
2	JUSTIFICATIVA	12
3	OBJETIVOS.....	13
4	METODOLOGIA	13
4.1	Definição dos termos de pesquisa	13
4.2	Coleta de dados	14
4.3	Limpeza dos dados.....	14
4.4	Categorização dos títulos.....	15
4.5	Criando os dataframes para análise.....	15
4.6	Análise descritiva dos dados	16
4.7	Construção de uma narrativa visual.....	16
5	RESULTADOS	17
6	CONSIDERAÇÕES FINAIS	17
7	REFERÊNCIAS BIBLIOGRÁFICAS.....	19

1 INTRODUÇÃO

O Jornalismo de Dados, também conhecido como jornalismo baseado em/guido por dados, é uma prática que emerge em um contexto de progresso tecnológico que permite uma ampla disseminação de informações no ambiente digital. Não há um consenso entre a definição do jornalismo de dados, de modo geral, ele pode ser definido como um “fluxo de trabalho” em que dados - produzidos constantemente em diversos formatos, como textos, imagens e áudios - são passíveis de investigação jornalística, orientam decisões, fundamentam narrativas, análises e visualizações. Nesse contexto de grande produção de informações e dados, e de maior presença de ferramentas digitais, torna-se essencial inserir nos meios de comunicação o conhecimento de como lidar com essas ferramentas e enfrentar os desafios de uma sociedade cada vez mais “dataficada” (BOUNEGRU; GRAY, 2021; CENTRE, 2010; HERAVI et al., 2021; MEJIAS; COULDRY, 2019).

O Jornalismo de Dados também pode ser considerado um processo de refinamento do dado bruto, no qual habilidades provenientes de áreas como estatística, programação e design são incorporadas aos dados com o objetivo de agregar valor às informações e facilitar a comunicação sobre elas.



Figura 1. O processo do jornalismo orientado por dados (adaptado de: http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf)

Há, ainda, outras definições para Jornalismo de Dados, que se complementam. Mastrella (2014, p. 14) discorre sobre algumas delas em seu Trabalho de Conclusão de Curso, que também diz que essa falta de consenso atrapalha a criação de “um corpo teórico mais estruturado para o segmento”

Segundo Howard (2014), “Jornalismo de Dados é contar histórias com números, ou procurar história neles. É tratar o dado como uma fonte para complementar os testemunhos humanos, as versões oficiais e de especialistas”. Para Trédan (2015), trata-se de uma prática baseada na “coleta, na análise estatística, na visualização de dados na Internet e no seu tratamento jornalístico”. Parasie e Dagiral (2013) falam em produção de conteúdos inovadores com base em “técnicas computacionais usadas para coletar, processar, analisar e visualizar conjuntos de dados”. Alguns teóricos optam pelo termo Jornalismo Computacional, como Hamilton e Turner (2009), que o classificam como “a combinação de algoritmos, dados e conhecimento das ciências sociais para complementar a função de fiscalizar do jornalismo.

Desse modo, as definições de Jornalismo de Dados compreendem desde o JD como produto, bem como processo. Os produtos de Jornalismo de Dados fazem referência aos trabalhos com visualizações de dados (com gráficos, infográficos, mapas interativos ou não) e aplicativos web, enquanto os processos se referem ao fluxo de trabalho do JD, que incluem a coleta dos dados, organização e análise dessas informações, que podem ser feitas de maneira automatizada ou não, com uso de softwares de computador, linguagens de programação, calculadoras e outras técnicas (GEHRKE et al., 2021; ROYAL; BLASINGAME, 2015).

1.1 O início do Jornalismo de Dados

O Jornalismo de Dados tem suas raízes na Computer Assisted Reporting (Reportagem Assistida por Computador) e no Jornalismo de Precisão. O Jornalismo de Precisão foi criado por Philip Meyer. Enquanto trabalhava para o jornal estadunidense Detroit Free Press, Meyer produziu uma reportagem que se utilizou de um computador IBM para ser desenvolvida: por meio de análise de dados de um questionário, a matéria pode estabelecer relações entre os conflitos raciais na cidade de Detroit e fatores sociodemográficos da população. A contribuição de Philip Meyer para o desenvolvimento e popularização da técnica de jornalismo de precisão

foi de grande importância para a evolução do jornalismo investigativo contemporâneo. É autor do livro intitulado “Precision Journalism: A Reporter’s Introduction to Social Science Methods” (Jornalismo de Precisão: Introdução dos Métodos de Ciências Sociais para Jornalistas), que se tornou uma referência na área (TRÄSEL, 2014; VENTURA, 2018; apud SILVA, 2019).

Um dos primeiros usos da reportagem com auxílio de computador foi na eleição presidencial dos EUA de 1952, quando a rede de televisão CBS utilizou o computador UNIVAC na previsão do resultado (COX, 2000). Anos mais tarde, computadores nas redações se popularizaram com a popularização da tecnologia: de acordo com Garrison (2001), reportagens com o auxílio de computador faziam parte de cerca de 90% das redações de jornais com circulação de 20 mil exemplares em 1998 nos EUA.

Apesar desses dois marcos na história do Jornalismo de Dados, estudiosos também atribuem o início do campo com Florence Nightingale (1820-1910), em 1850. Durante a Guerra da Crimeia (1853-1856), a enfermeira britânica ficou conhecida pelo seu trabalho em prol da melhoria das condições sanitárias e de higiene dos soldados feridos. Para tal, ela se valeu de seus conhecimentos em matemática e estatística, sendo que um de seus trabalhos mais notáveis foi a criação do gráfico estatístico conhecido como “Diagrama da Rosa”. Tal gráfico ilustrou a taxa de mortalidade entre soldados feridos durante a guerra em questão, e pode ser considerado um dos primeiros exemplos de jornalismo de dados, por empregar dados no processo de criação e no produto final, com a finalidade de comunicar informações (BRADSHAW, 2020).

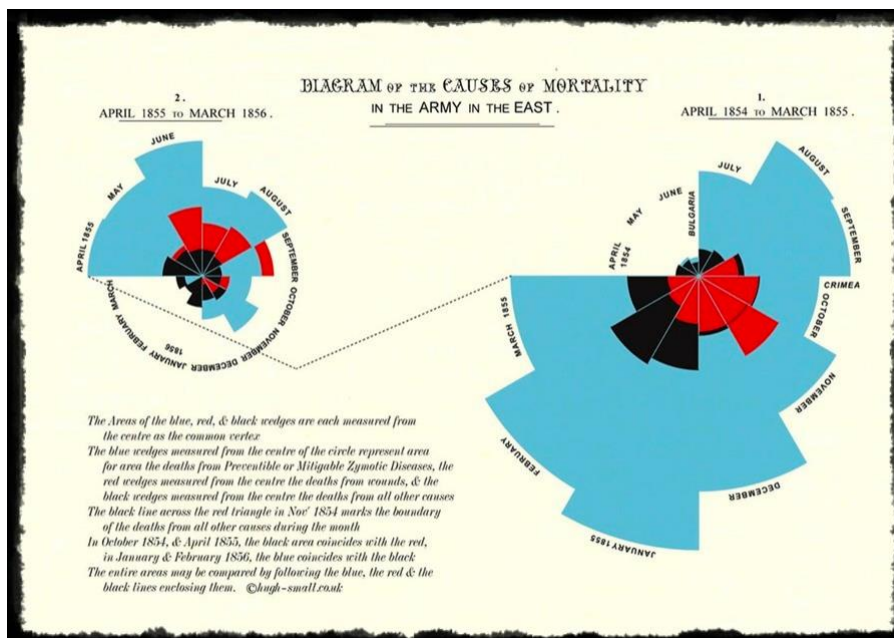


Figura 2 – O Diagrama da Rosa. Fonte: Site Medium. Disponível em <<https://medium.com/geotech/florence-nightingale-o-diagrama-da-rosa-e-o-impacto-visual-6f0e1fe5844d>>

No entanto, foi apenas com o advento da tecnologia digital e a disponibilidade de grandes quantidades de dados em formatos digitais que o Jornalismo de Dados se tornou uma prática comum nas redações, substituindo o uso do termo "jornalismo assistido por computador" no início dos anos 2000 (ROYAL ET AL, 2015).

O Jornalismo de Dados tem sido utilizado para investigações em áreas como política, saúde, finanças e esportes, e tem sido fundamental para a cobertura de temas complexos, como mudanças climáticas e pandemias.

1.2 Jornalismo de dados no Brasil

Muitas organizações de mídia em todo o mundo passaram a investir em equipes de Jornalismo de Dados, e surgiram diversas iniciativas de treinamento e desenvolvimento de habilidades em Jornalismo de Dados. No Brasil, equipes de dados e cursos sobre o tema foram essenciais para o crescimento da área no país.

A introdução de computadores pessoais na rotina das redações se iniciou nos anos 80, no jornal Folha de S. Paulo. A tecnologia possibilitou ao jornal ser o pioneiro na informatização no processo jornalístico, sendo a “primeira redação informatizada na América do Sul” (SILVA, 2019, p. 31). Junto com o Estado de São Paulo, foram os primeiros a implementar técnicas de Jornalismo de Dados na redação, por meio das reportagens assistidas por computador.

Ambos os jornais consolidaram suas equipes de dados em 2012: o Estadão Dados e o DeltaFolha. Cabe destacar o papel de instituições que fomentam e financiam as práticas jornalísticas para o desenvolvimento dessas equipes, como a parceria com o Knight Center for Journalism in the Americas, da Universidade do Texas em Austin, na criação do núcleo de dados do jornal Estadão; e com programa Knight International Journalism Fellowship, do Centro Internacional para Jornalistas (ICFJ), que participou na elaboração do FolhaSPDados, em 2012, que futuramente serviria de base para a formação da equipe de jornalismo de dados da Folha, a DeltaFolha (FOLHA DE S. PAULO, 2012; SOUZA, 2021; FOLHA DE S. PAULO, 2019). Também é importante o papel dos treinamentos oferecidos pela Associação Brasileira de Jornalismo Investigativo e iniciativas como a Escola de Dados no letramento de dados em redações (DE LIMA-SANTOS, 2019, p.7).

Na história do JD no Brasil, um dos nomes que se destacam é o de Claudio Weber Abramo (1952-2018). Abramo foi um jornalista e ativista pela transparência pública, sendo um dos fundadores da organização Transparência Brasil, em 2000. Sua contribuição para o JD se dá em grande parte pelo fortalecimento do acesso à informação pública, essencial nos trabalhos com dados públicos, ou seja, em dados que deveriam ser disponibilizados publicamente para os cidadãos por entes dos governos federal, estaduais e municipais.

1.3 Raça e racismo

O conceito de raça deriva da Idade Média, inicialmente utilizado com uma interpretação das noções de linhagem e ancestral comum. Posteriormente, com as “Grandes Navegações” do

século XV, raça passou a classificar a diversidade humana, justificando relações de dominação e sujeição entre classes sociais, “pois foi destes encontros entre a ”civilização” europeia e os índios americanos, os negros africanos e os asiáticos que surgiu a necessidade de classificar e definir o que era e quem era a humanidade” (SCHUCMAN, 2012, p.31).

Com o Iluminismo, raça passou a incorporar uma interpretação biológica na categorização de povos distintos. Entretanto, a adoção das Ciências Naturais para explicar a diversidade humana acabou legitimando teorias raciais que hierarquizavam povos, e justificando a dominação europeia sobre povos de outros continentes, como a África e as Américas.

Foram estabelecidos critérios para essa classificação ao longo dos séculos. Inicialmente, a cor da pele era fundamental, e posteriormente foram adicionados critérios como formato do nariz, lábios, queixo, formato do crânio, entre outros.

De acordo com Almeida (2019, p. 112):

Conforme este raciocínio, as sociedades contemporâneas, mesmo após o fim oficial dos regimes escravistas, permaneceriam presas a padrões mentais e institucionais escravocratas, ou seja, racistas, autoritários e violentos. Dessa forma, o racismo seria uma espécie de resquício da escravidão, uma contaminação essencial que, especialmente nos países periféricos, impediria a modernização das economias e o aparecimento de regimes democráticos. No caso dos países centrais as marcas da escravidão poderiam ser vistas na discriminação econômica e política a que são submetidas as minorias raciais, como é o caso da população negra e latina nos EUA e dos imigrantes não-brancos na Europa”.

Segundo Abrahão (2010, p. 31):

As construções sobre a “raça negra” antecedem ao racismo na medida em que se valem da intencionalidade de construir a superioridade de “uns” em detrimento de “outros”. Em outros termos, foi a vontade da diferenciação que levou à necessidade de construção da “raça”.

1.4 Cor e raça na população brasileira

No Brasil, algumas teorias tentaram explicar a formação do povo brasileiro. Uma delas foi a da “democracia racial”. Gilberto Freire (1933) foi um dos pensadores que popularizou essa ideia no Brasil: o mito refere-se, especialmente ao longo do século XX, a uma teoria de

que o país era caracterizado pela harmonia e pela igualdade racial, apesar das evidentes diferenças sociais e econômicas entre brancos e não brancos. Esse pensamento sugeriu que as relações raciais no Brasil eram mais amistosas e menos problemáticas do que em outros lugares, como nos Estados Unidos, onde o racismo era mais flagrante e institucionalizado.

Inspirada nos diários e registros de pessoas dos séculos XVIII e XIX e tendo como plano de fundo os canaviais nordestinos da época, Casa Grande & Senzala (1933) retrata a formação do povo brasileiro por meio da miscigenação harmoniosa entre as três etnias que habitavam o país: índios, portugueses e africanos. Tal visão defendida por Freyre (2004) pode ser apresentada sob a ideia de uma “democracia racial”, que fazia com que as relações escravocratas no Brasil fossem amenizadas, em comparação com as dos Estados Unidos, que, de acordo com ele, eram bem mais severas”. (SILVA; OLIVEIRA SILVA; SILVA NASCIMENTO, 2014)

O mito incorporava a miscigenação entre diferentes grupos étnicos, o que tornava a sociedade brasileira, na teoria, menos propensa ao racismo. Segundo Santos (2018, p. 7):

Trata-se de uma visão de mundo que reinventa o país, visto que revela a possibilidade de convivência dos diferentes grupos socioculturais (europeus, africanos e indígenas) então residentes dentro das fronteiras político-geográficas brasileiras. Deve-se lembrar de que até as primeiras décadas do século XX uma questão polarizava o debate político brasileiro, a saber, a constituição de uma nação unitária e progressista, baseada na lógica da mestiçagem.

Como um mito, a democracia racial não se mostrava na realidade. Lélia Gonzalez (1984) foi uma das críticas ao “discurso dominante” da democracia racial. De acordo com Hertz (2022), citando Lélia:

[...] esse “discurso dominante”, no termo de Gonzalez - sintetizado na noção de “democracia racial” - obstaculizou a percepção e o combate em torno das discriminações e exclusões raciais na sociedade brasileira, ainda mais quando consideramos que ele se torna uma “imagem de Brasil” oficializada e promulgada pelo Estado, não só em âmbito nacional, mas também internacional.

De acordo com a Pesquisa Nacional por Amostra de Domicílios (PNAD) Contínua, realizada pelo Instituto Brasileiro de Geografia e Estatística em 2022, a população negra compõe 56% do total de habitantes brasileiros (IBGE, 2022). Mesmo sendo maioria, em

proporção, da população brasileira, negros ainda são oprimidos em diferentes esferas, como herança das teorias raciais que justificavam a sua dominação.

Segundo pesquisa da ONG Oxfam denominada “A distância que nos une – Um retrato das Desigualdades Brasileiras” (BRASIL, 2017), usando dados da PNAD Contínua do IBGE, o rendimento de negros é menor do que o de brancos em todos os níveis educacionais. Cargos gerenciais, normalmente associados à liderança, são ocupados majoritariamente por brancos: cerca de 69%

Por outro lado, a taxa de homicídios por 100 mil habitantes também reforça a desigualdade racial. Enquanto para uma pessoa branca a taxa é de 11,5, para pretos o valor é de 21,9 e pardos, 34,1 (IBGE, 2020). A taxa de analfabetismo de pessoas negras é mais que o dobro da taxa da população branca em diferentes faixas etárias (IBGE, 2022).

Essas estatísticas apenas resumem o “racismo a brasileira”. Enquanto negros sofrem os efeitos oriundos do racismo, a branquitude é universalizada e isso acontece sem o devido debate. O racismo à brasileira tem sua existência, de acordo com Antônio (2015, p. 368)

[...] fundada em sistemas de ideias, teorias, discursos e ações articulados pelo pensamento único, pela cristalização e pela naturalização ativadas historicamente pelas políticas e ideologias de cordialidade no âmbito das relações raciais; pelo mito de democracia racial, pelas políticas e ideologias de branqueamento; pela branquitude universalizada e, para fechar o sistema, pelas perversas políticas de silêncio e de ignorância pré-estabelecida em torno do debate e da superação do racismo à brasileira”.

1.5 Raça e imprensa

A grande imprensa brasileira também reflete a desigualdade social marcada pelo racismo. O estudo do grupo GEMAA “Raça, gênero e imprensa: quem escreve nos principais jornais do Brasil?” (PORTELA et al., 2023) explora os dados de cor e raça da redação de três jornais brasileiros (O Globo, O Estado de São Paulo e Folha de São Paulo), e também se

aprofunda em outros recortes, como idade e gênero da equipe editorial desses veículos. O relatório conclui que

A mídia brasileira continua reproduzindo padrões de forte desigualdade racial. Quando se trata da produção de notícias e circulação da informação, a discussão sobre desigualdades ultrapassa a questão da justa proporcionalidade e alcança o problema da invisibilização de grupos sociais na produção das narrativas e informações que instruem o processo de formação de opinião e os entendimentos socialmente partilhados”. (PORTELA et al., 2023, p. 3)

1.6 Título e pauta

Se os negros estão sub-representados nas redações, como eles chegam e são representados nas notícias?

Pensando nessa questão, é interessante observar o papel dos títulos em uma notícia. Segundo o Dicionário de Oxford, “Título”, é o “nome ou expressão que se coloca no começo de um livro, em seus capítulos, em publicação jornalística, peça teatral, filme, música etc., que pode indicar o assunto ou simplesmente identificar, individualizar a obra ou o trabalho”. Desse modo, o título é o primeiro contato que o leitor, ouvinte ou espectador têm com a produção que vai experienciar.

O título antecede a pauta, que também segundo o adotado pela Folha de São Paulo (2022), é o “primeiro roteiro para a produção de reportagens, não é uma simples ideia, mas um plano de ação” e considera “questões básicas a serem respondidas ao longo da apuração, enfoque, formato, hipótese de trabalho, material de apoio, imagens”.

Moraes (2022, p. 10) conceitua a pauta como “a coluna vertebral da notícia, aquilo que dá forma ao conteúdo jornalístico oferecido ao público” e complementa, chamando atenção para um ponto de vista político da pauta:

A pauta vai enquadrar algo, a partir de um ou poucos olhares, que será compartilhado para um público mais amplo. Ela diz: algo está acontecendo e vocês devem saber.(...) Assim, tanto os elementos presentes na concepção de uma reportagem quanto aqueles que foram descartados são sempre escolhas políticas e, sim, também arbitrárias. Toda pauta organiza e desorganiza visibilidades e invisibilidades, toda pauta hierarquiza e

desierarquiza vozes e representações, toda pauta estrutura e desestrutura discursos” (MORAES, 2022, p. 10).

Considerando esses apontamentos, é possível inferir que, no processo jornalístico, a pauta antecede o título para o repórter, mas o que chega primeiro ao leitor é o título. Assim, podemos reformular a questão feita anteriormente: se os negros estão sub-representados nas redações, como eles viram pauta nas notícias e quais títulos eles estampam nos portais e manchetes de jornais?

2. JUSTIFICATIVA

Diante dos conceitos e histórico apresentados, este trabalho se propõe a analisar notícias com abordagem racial publicadas pelo jornal Folha de São Paulo, de forma online, entre os anos de 2013 e 2022. Isso será realizado por meio de técnicas de Jornalismo de Dados e apresentado em um site com visualizações e análises de dados, mostrando as dimensões do JD como processo e produto.

Com cerca de 41 milhões de visitas no mês de outubro de 2023 (SIMILARWEB, 2023), o jornal Folha de São Paulo é uma das principais mídias informativas. Com tamanha repercussão, é possível acreditar que a maneira como representa pessoas negras pode impactar na forma como são tratadas na realidade. A escolha pela Folha de São Paulo ao invés de outros veículos também se deu pela possibilidade de raspar o site com a certeza da legalidade dessa ação.

Optou-se por analisar o período de 2013 a 2022, pois essa seleção permitiu a coleta de uma extensa quantidade de informações, as quais poderiam ser processadas por ferramentas de Jornalismo de Dados. Essa abordagem visa verificar o potencial dessas técnicas em investigações envolvendo objetos de estudo de grande escala.

Desse modo, este trabalho pretende contribuir para o debate sobre o uso de ferramentas de tecnologia e análise de dados no âmbito jornalístico, como forma de extrair informações e

pautas a partir de um conjunto de dados. Utilizando de títulos de notícias como objeto de estudo, o trabalho também propõe uma maneira de avaliar a existência de padrões de cobertura, bem como fazer um histórico do que foi noticiado em um longo período de tempo, por meio de ferramentas automatizadas. Além disso, pretende-se questionar como padrões nos títulos podem corroborar ou evitar vieses na cobertura racial, o que é possível a partir de uma ampla base de dados.

3. OBJETIVOS

O presente trabalho tem como objetivo geral:

- Analisar dez anos de cobertura racial pelo jornal Folha de São Paulo utilizando técnicas do Jornalismo de Dados.

Como objetivos específicos, pretendeu-se:

- Criar uma base de dados de notícias com temática racial do portal Folha de S. Paulo;
- Analisar a base de dados criada com técnicas de Jornalismo de Dados, explorando o fluxo de trabalho com dados;
- Discutir os achados por meio de uma narrativa com dados

4. METODOLOGIA

4.1 Definição dos termos de pesquisa

Alguns termos foram utilizados nas pesquisas no site da Folha de São Paulo. São termos que são associados às pessoas negras. São eles: “afrodescendente”, “homem negro”, “homens negros”, “homens pardos”, “mulata”, “mulato”, “mulheres negras”, “mulheres pardas”, “mulher negra”, “negritude”, “pessoas negras”, “pessoas pardas”, “raciais”, “racial”.

O trabalho não incluiu a pesquisa com “mulher preta”, “homem preto” e suas variações no plural. Apesar da compreensão de que a população negra é composta por pretos e pardos, a inclusão do adjetivo preto trazia mais notícias fora do contexto racial.

4.2 Coleta de dados

A coleta de notícias foi realizada utilizando a linguagem de programação Python, a partir de Web Scraping (ou "raspagem de dados da web") que consiste em uma técnica de coleta de dados na qual um programa automatizado extrai informações a partir do código HTML de uma página web.

Foram usadas as bibliotecas Selenium e BeautifulSoup para extrair o título da reportagem, sua url, data de publicação, descrição e seção. O período de publicação das matérias foi de janeiro de 2013 a dezembro de 2022 e a execução do código da raspagem das páginas foi feita uma vez por ano e por termo de pesquisa.

Foram coletados 31 mil e 513 títulos do portal da Folha de São Paulo, obtidos por meio do resultado das pesquisas no site com os termos citados na etapa anterior.

4.3 Limpeza dos dados

Os dados coletados da etapa anterior passaram por um pré-processamento. É nessa etapa que os títulos são formatados para serem lidos igualmente pelo computador. A limpeza consistiu na remoção de quebras de linha e formatação das datas, exclusão de colunas em excesso, padronização dos títulos para letra minúscula, remoção de pontuação e remoção de valores duplicados.

Em seguida, as informações foram filtradas por editoria. De interesse para este trabalho, foram escolhidas as editorias Mundo, Esporte, Mercado, Ilustrada, Opinião, Equilíbrio e Saúde, Poder, Ilustríssima, Cotidiano.

Stopwords: Stopwords são palavras comuns que são frequentemente usadas em uma linguagem, mas que geralmente não contribuem muito para o significado do texto em que aparecem. Essas palavras incluem artigos, preposições, conjunções e outras palavras comuns, como “o”, “a”, “um”, “de”, “para”, “em”, “com”, “e”, entre outras. Em muitas tarefas de processamento de linguagem natural, como análise de sentimentos, classificação de texto e recuperação de informações, as stopwords são removidas do texto para reduzir o ruído e melhorar a precisão dos resultados.

4.4 Categorização dos títulos

Os títulos foram distribuídos em duas categorias: VERDADEIRO, quando apresenta características que possibilita entender que se trata de uma notícia com temática racial; ou FALSO, se não der esse entendimento. Essa categorização foi feita automaticamente e manualmente.

Na categoria VERDADEIRO, foram incluídos automaticamente os títulos que:

- Apresentam um ou mais dos termos no singular ou plural;
- Apresentam, no título, uma personalidade negra;

Manualmente, os títulos que não seguiram os critérios acima foram lidos e atribuídos os valores VERDADEIRO ou FALSO quando o contexto da frase passa o entendimento de que se trata de uma notícia com temática racial.

4.5 Criando os dataframes para análise

Os títulos categorizados como VERDADEIRO foram utilizados para criar quatro outros arquivos, que foram utilizados para cada análise:

- títulos_palavras.csv: neste arquivo, cada registro é uma palavra de um título. Com ele é possível ver a quantidade de ocorrências de uma palavra e como elas se distribuem por anos;
- títulos_bigramas.csv: semelhante ao arquivo anterior, neste caso cada linha é um bigrama (dupla de palavras que ocorrem simultaneamente em um texto);
- títulos_entidades.csv: para cada notícia foram extraídas as entidades reconhecidas por aprendizado de máquina. Foi possível reconhecer pessoas, organizações e lugares. Também foram extraídos os verbos, adjetivos e pronomes.

4.6 Análise descritiva dos dados

A análise descritiva foi feita observando termos mais frequentes nos títulos, existência ou ausência de padrões presentes nesses textos. Para isso, os dados gerados na etapa anterior foram explorados para responder às questões:

- a) Quais são as palavras mais frequentes no corpus textual geral?
- b) Quais são as palavras mais frequentes no corpus textual por editoria e por ano?
- c) Quais são os bigramas mais frequentes no corpus textual geral?
- d) Quais são os bigramas mais frequentes no corpus textual por editoria e por ano?
- e) Quais são as personalidades e organizações mais citados nos títulos?

4.7 Construção de uma narrativa visual

Com as análises e entrevistas, foi produzida uma narrativa de dados em um site, de forma que os resultados obtidos se tornassem uma reportagem de dados. As visualizações de dados foram construídas a partir dos templates do Flourish Studio, e a página web foi produzida utilizando linguagens de marcação e estilo (HTML e CSS) e Javascript. Também foram disponibilizados, em uma página web feita com Quarto, os códigos e métodos descritos.

5 RESULTADOS

A partir das informações coletadas, foram produzidos:

- Um site, contendo as anotações metodológicas presentes neste relatório de pesquisa, de modo a tornar acessível e aberto as considerações, comentários e métodos utilizados para análise, disponível em <https://biancamuniz.quarto.pub/tcc/>
- Uma narrativa orientada por dados, resumindo achados deste trabalho, disponível em https://biamuniz.github.io/tcc_storytelling
- Arquivos nos formatos CSV e RDS, contendo os dados utilizados para as análises, disponível em https://github.com/biamuniz/dados_tcc.

6 CONSIDERAÇÕES FINAIS

O presente Trabalho de Conclusão de Curso atingiu os objetivos propostos, além de fomentar o uso de técnicas de Jornalismo de Dados para a análise crítica de vieses, conscientes ou não, da cobertura jornalística sobre temas relacionados à população negra.

No entanto, cabe destacar algumas limitações. O Jornalismo de Dados, por utilizar diferentes conhecimentos (como estatística, programação, design, apuração e escrita), é uma área multidisciplinar. Para um bom Jornalismo de Dados, seria interessante realizar este trabalho com uma equipe maior e diversificada em campos de atuação.

Os desafios relacionados ao Processamento de Linguagem Natural na Língua Portuguesa também devem ser considerados. Diferentemente da extensa quantidade de trabalhos e materiais sobre como realizar reconhecimento de entidades, classificação de texto e análise pós-tagging na língua inglesa, os exemplos em português ainda são escassos. Em alguns momentos, foram adaptadas técnicas de um vocabulário diferente do falado no Brasil, o que pode ter afetado algumas conclusões.

Para além desses fatores, apesar do trabalho não se propor a aprofundar em questões raciais, seria proveitoso complementar os achados da pesquisa com entrevistas com especialistas e jornalistas. Eles poderiam comentar e analisar, a partir da sua prática jornalística e acadêmica sobre a cobertura racial, o que os dados revelam.

Por fim, a pesquisa fortalece o potencial de técnicas de Jornalismo de Dados na rotina jornalística. Com os avanços da Inteligência Artificial e de ferramentas de investigação orientadas por dados, é necessário que o profissional ao menos compreenda as finalidades do DDJ e como ele afeta o fluxo de trabalho na construção de uma reportagem. Além disso, o trabalho sugere uma aplicação do Jornalismo de Dados na avaliação da cobertura midiática.

7 REFERÊNCIAS BIBLIOGRÁFICAS

ABRAHÃO, B. DE L. **O “preconceito de marca” e a ambiguidade do “racismo à brasileira” no futebol**. Rio de Janeiro: Universidade Gama Filho-Programa de Pós-Graduação em Educação Física: Tese de Doutorado, 2010.

ALMEIDA, S. **Racismo estrutural**. São Paulo: Pólen Produção Editorial LTDA, 2019.

ANTONIO, C. F. Racismo à brasileira: a perspectiva totalitária da branquitude. In: **África e Brasil**, 2015. p. 366-398.

BOUNEGRU, L.; GRAY, J. **The Data Journalism Handbook: Towards a Critical Data Practice**. Amsterdam: Amsterdam University Press, 2021.

BRADSHAW, N. A. Florence Nightingale (1820–1910): an unexpected master of data. **Patterns**, v. 1, n. 2, 2020.

BRASIL, O. A distância que nos une: um retrato das desigualdades brasileiras. Disponível em: <<https://www.oxfam.org.br/publicacao/a-distancia-que-nos-une-um-retrato-das-desigualdades-brasileiras/>>.

CENTRE, E. J. **Data-Driven Journalism: What Is There to Learn?** [s.l.] European Journalism Center, 2010. Disponível em: <http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf>.

COX, M. **The Development Of Computer-Assisted Reporting**. Chapel Hill, EUA: Newspaper Division, Association for Education in Journalism; Mass Communication, 2000.

DE-LIMA-SANTOS, M. F. Tem# DDJBR aqui? Mapeando a presença do jornalismo de dados no Brasil. Intercom–Sociedade Brasileira de Estudos Interdisciplinares Da Comunicação. Anais... 2019.

FREYRE, G. **Casa grande e senzala**. Rio de Janeiro: Schmidt Editor, 1933.

GARRISON, B. Computer-assisted reporting near complete adoption. **Newspaper Research Journal**, v. 22, n. 1, p. 65–79, 2001.

GEHRKE, M. et al. **Fluxo de trabalho com dados - Do zero à prática** (Escola de Dados). [s.l: s.n.].

GOMES, T. B.; COSTA, G. S. Caça-cliques no jornalismo: 5 padrões de títulos para atrair leitores no Facebook. **Periódicos UFPB**, Ano XII, n. 7, 2016.

GONZALEZ, L. Racismo e sexismo na cultura brasileira. **Revista ciências sociais hoje**, v. 2, n. 1, p. 223–244, 1984.

HERAVI, B. et al. Preserving data journalism: A systematic literature review. **Journalism Practice**, p. 1–23, 2021.

IBGE. **PNAD Contínua - Pesquisa Nacional por Amostra de Domicílios Contínua: características gerais dos domicílios e dos moradores**, 2022. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2022.

LAGE, N. **Estrutura da notícia**. São Paulo: Ática. 2005.

MASTRELLA, B. **Os discursos de legitimação do jornalismo guiado por dados**. 2019.

MEJIAS, U. A.; COULDRY, N. Datafication. **Internet Policy Review**, v. 8, n. 4, 2019.

MORAES, F. **A pauta é uma arma de combate: subjetividade, prática reflexiva e posicionamento para superar um jornalismo que desumaniza**. [s.l.] Arquipélago Editorial, 2022.

PAULO, F. DE S. **Manual da redação: as normas de escrita e conduta do principal jornal do país**. Folha de S. Paulo, 2022.

PORTELA, P. et al. **Raça, gênero e imprensa: quem escreve nos principais jornais do Brasil?** [s.l.] GEMAA, IESP-UERJ, 2023. p. 1–21

ROYAL, C.; BLASINGAME, D. Data journalism: An explication. International Symposium in Online Journalism. Anais...2015.

S. PAULO, F. DE. FOLHA LANÇA SITE DE DADOS ABERTOS FOCADO NAS ELEIÇÕES DE SÃO PAULO. Disponível em: <<https://www1.folha.uol.com.br/blogs/novo->

em-folha/2022/01/folha-realiza-segunda-edicao-de-treinamento-para-profissionais-negros.shtml>.

SCHUCMAN, L. V. Entre o “encardido”, o “branco” e o “branquíssimo”: raça, hierarquia e poder na construção da branquitude paulistana. tese de doutorado—São Paulo: Instituto de Psicologia, Universidade de São Paulo, 2012.

SILVA, A. H. DA; OLIVEIRA SILVA, L. E. DE; SILVA NASCIMENTO, W. DA. A formação histórico-cultural e a ideia de pessimismo racial no Brasil. **Revista Três Pontos**, 2014.

SILVA, L. D. J. Jornalismo de dados: conceitos, fluxos, ferramentas e formação acadêmica. 2019.

SOUZA, V. D. F. B. JORNALISMO DE DADOS E COBERTURA POLÍTICA: UMA ANÁLISE DA ATUAÇÃO DO DELTAFOLHA NAS ELEIÇÕES MUNICIPAIS DE 2020. **Revista Miquel**, v. 4, n. 4, 2021.

TRÄSEL, M. Entrevistando Planilhas: Estudo das crenças e do ethos de um grupo de profissionais de Jornalismo Guiado por Dados no Brasil. tese de doutorado—Porto Alegre: Pontifícia Universidade Católica do Rio Grande do Sul, 2014.

VENTURA, M. P. Jornalismo de Dados como diferencial: o caso do Nexo. **Revista Pauta Geral: Estudos em Jornalismo**, p. 240–245, 2018.