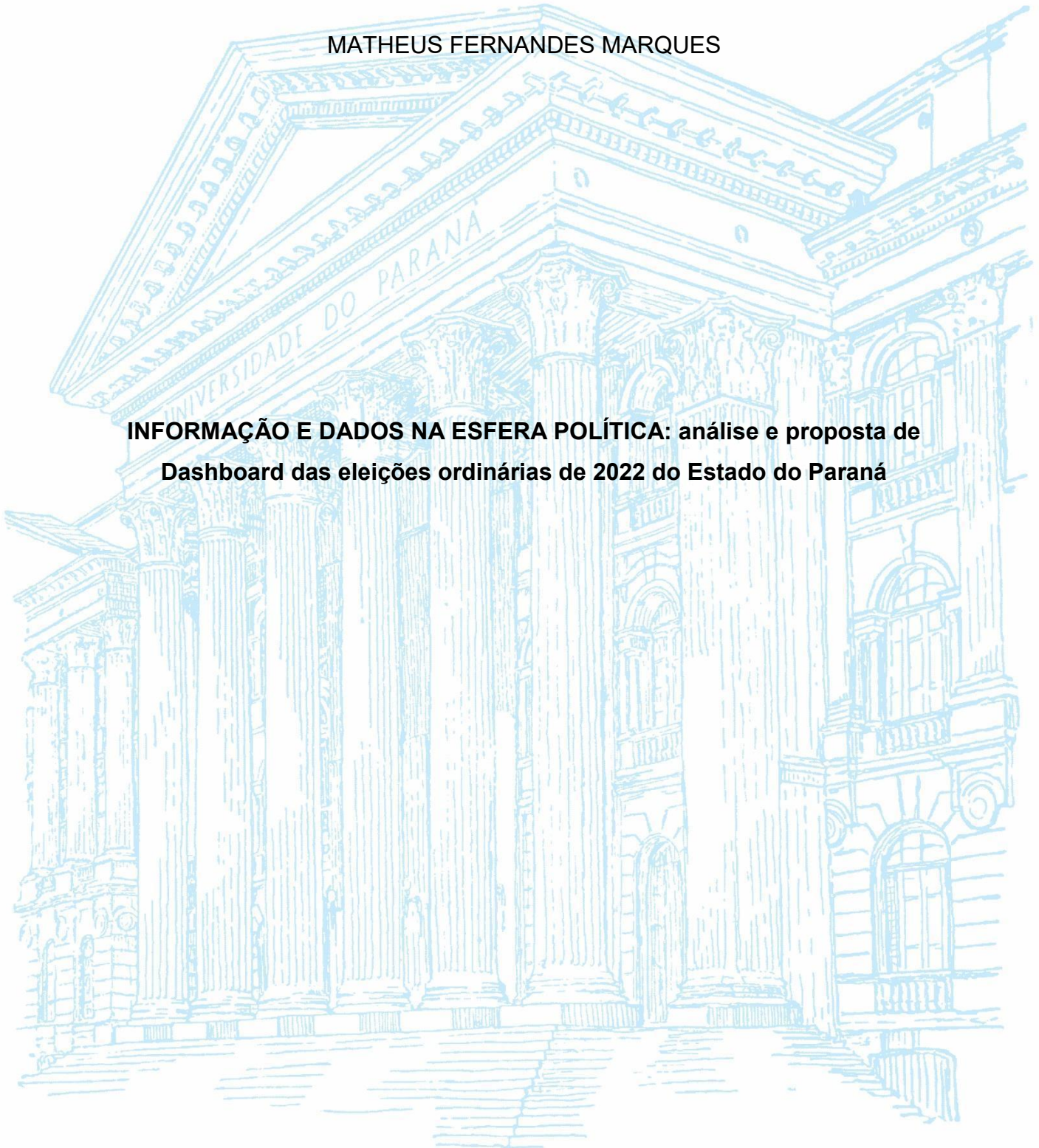


UNIVERSIDADE FEDERAL DO PARANÁ

MATHEUS FERNANDES MARQUES

**INFORMAÇÃO E DADOS NA ESFERA POLÍTICA: análise e proposta de
Dashboard das eleições ordinárias de 2022 do Estado do Paraná**



CURITIBA

2023

Matheus Fernandes Marques

O fantasma dos dados: análise e proposta de Dashboard das eleições ordinárias de 2022 do Estado do Paraná

Trabalho de conclusão de curso apresentado ao curso de Gestão da Informação, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Gestão da Informação.

Professor Orientador: Prof. Dr. Andre Vieira de Freitas Araujo

CURITIBA

2023

RESUMO

Tem como objetivo propor uma releitura do existente Dashboard das Eleições ordinárias do Brasil de 2022 utilizando de dados disponibilizados pelo Tribunal Superior Eleitoral. Discute os diferentes aspectos de um produto/serviço informacional, à luz das humanidades digitais, área do conhecimento que abrange a existência de Dashboards. Com relação a metodologia, é uma pesquisa de abordagem prática, definindo e processos que visam o levantamento de dados do produto/serviço informacional existente, de possíveis melhorias no processo de extração, tratamento, disponibilização de dados, e visual das informações disponibilizadas pelo órgão federal. Mostra-se uma análise do impacto social dos problemas existentes na dashboard que possibilitam a interpretação errada de dados e informações, acarretando na criação de fake news, problema de grande destaque atualmente na sociedade brasileira. Além da discussão sobre problemas existentes, é feito o processo de extração, tratamento e disponibilização das informações, e a criação de uma nova proposta dashboard com a finalidade de sanar os problemas identificados no processo de análise. Mostram-se os resultados obtidos junto de uma análise da nova proposta de *dashboard*, apresentando uma série de melhorias para os problemas observados. A nova proposta de dashboard junto da descrição das alterações de cores, implementação de visuais. Conclui definindo as limitações da pesquisa sendo principalmente o conjunto de cores do órgão, que faz referência a um dos candidatos, junto dos dados sem o complemento de votos em trânsito, enquanto finaliza propondo novas pesquisas utilizando de mineração de dados para observar características dos eleitores e quais propostas os mesmos buscam em um candidato.

Palavras-chave: Produto/serviço informacional; Dashboard; Visualização da informação; Dados eleitorais; Eleição brasileira.

ABSTRACT

Its objective is to propose a rereading of the existing Dashboard of the Ordinary Elections of Brazil in 2022 using data provided by the Superior Electoral Court. It discusses the different aspects of an information product/service, in the light of digital humanities, an area of knowledge that encompasses the existence of Dashboards. Regarding the methodology, it is a research with a practical approach, defining and processes that aim to collect data from the existing informational product/service, possible improvements in the process of extraction, treatment, data availability, and visual of the information made available by the federal agency. It shows an analysis of the social impact of the existing problems in the dashboard that allow the wrong interpretation of data and information, resulting in the creation of fake news, a problem of great prominence in Brazilian society today. In addition to the discussion of existing problems, the process of extracting, processing and making available information is carried out, and the creation of a new dashboard proposal in order to remedy the problems identified in the analysis process. The results obtained are shown together with an analysis of the new dashboard proposal, presenting a series of improvements for the observed problems. The new dashboard proposal next to the description of color changes, implementation of visuals. It concludes by defining the limitations of the research, mainly the set of colors of the body, which refers to one of the candidates, together with data without the complement of votes in transit, while it ends by proposing new research using data mining to observe characteristics of voters and what proposals they look for in a candidate.

Keywords: Informational product/service; Dashboard; Information visualization; Data; Brazilian election.

LISTA DE FIGURAS

FIGURA 1 - Processo de GI 1.....	24
FIGURA 2 - Home page Dashboard das eleições ordinárias de 2022.....	37
FIGURA 3 - Home page Dashboard das eleições ordinárias de 2022 - Botão “Ver lista Completa”	38
FIGURA 4 - Menu do Dashboard.....	39
FIGURA 5 - Página de totalização	40
FIGURA 6 - Página Favoritos	41
FIGURA 7 - Página Dados de Urna	42
FIGURA 8 - Script python - carregamento de dados	47
FIGURA 9 - Script python - criação de engine e integração com o banco de dados	47
FIGURA 10 - Banco de dados - Visão pós integração.....	48
FIGURA 11 - Banco de dados - Visão pós criação de view.....	49
FIGURA 12 - Interface Power BI.....	49
FIGURA 13 - Interface Power BI - conexão com o banco de dados.....	50
FIGURA 14 - Interface Power BI - dados após a conexão – Estado.....	51
FIGURA 15 - Interface Power BI - dados após a conexão – Federação.....	51
FIGURA 16 - Proposta de nova dashboard - interface de filtros.....	52
FIGURA 17 - Proposta de nova dashboard - Exemplo medida	52
FIGURA 18 - Proposta de nova dashboard - interface geral - Federal	53
FIGURA 19 - Proposta de nova dashboard - interface geral - Estadual	54
FIGURA 20 - Proposta de nova dashboard - Análise - Filtros	55
FIGURA 21 - Proposta de nova dashboard - Análise - Ícones.....	56
FIGURA 22 - Proposta de nova dashboard - Análise – Gráficos	56

SUMÁRIO

1. INTRODUÇÃO.....	7
1.1 PROBLEMATIZAÇÃO	9
1.2 OBJETIVOS	9
1.2.1 OBJETIVO GERAL.....	10
1.2.2 OBJETIVOS ESPECÍFICOS	10
1.3 JUSTIFICATIVA	10
1.3.1 ASPECTO SOCIAL, CIENTÍFICO E PESSOAL.....	11
2. REFERENCIAL TEÓRICO	13
2.1 SERVIÇOS/PRODUTOS DE INFORMAÇÃO	13
2.2 HUMANIDADES DIGITAIS.....	15
2.3 BASE DE DADOS	16
2.2.1 DADOS ELEITORAIS	18
2.4 VISUALIZAÇÃO DA INFORMAÇÃO	19
2.5 TÉCNICAS DE VISUALIZAÇÃO DE DADOS.....	21
2.6 PROCESSO DE ETL	23
2.6.1 LINGUAGENS DE PROGRAMAÇÃO	26
2.7 ANÁLISE DE DADOS.....	31
3. METODOLOGIA.....	32
3.1 CAMPO DE PESQUISA	32
3.2 ANÁLISE DO PRODUTO EXISTENTE E ESTABELECIMENTO DO ESCOPO DO NOVO PRODUTO	33
3.3 PROCESSO DE ANÁLISE, TÉCNICAS DE COLETA, TRATAMENTO, E DISPONIBILIZAÇÃO DE DADOS(ELT)	33
4. AVALIAÇÃO DA DASHBOARD EXISTENTE	36
4.1 DESCRIÇÃO E ANÁLISE DA DASHBOARD	36
4.2 IMPACTO SOCIAL DOS DADOS	43
5. RELEITURA DO DASHBOARD DAS ELEIÇÕES ORDINÁRIAS DE 2022 - ESTADO DO PARANÁ	45
5.1 PROCESSO DE ETL	45
5.2 CRIAÇÃO DA DASHBOARD	51
5.3 ANÁLISE DO NOVO DASHBOARD.....	55
6. CONSIDERAÇÕES FINAIS.....	58
7. REFERÊNCIAS	60

1 INTRODUÇÃO

Dissertar sobre dados envolve grande interdisciplinaridade, pois o tema abrange diversas especialidades de diferentes áreas como Tecnologia da Informação, Ciência da Informação e Administração. O tema que inicialmente é simples pode acabar se tornando extremamente complexo envolvendo diferentes conceitos, porém, o dado em si nunca será o produto final, mas sim um meio, uma ferramenta que está circunscrita em contextos, realidades sociais e temporalidades específicas. Gonzalez de Gomez (2017, p.1) enfatiza que

Os estudos da informação formam parte das mudanças das matrizes paradigmáticas das ciências contemporâneas. Mudanças que são consistentes com o deslocamento do conhecimento, da comunicação e da linguagem de seu locus investigativo prévio, no domínio reflexivo dos estudos filosóficos e humanísticos, conforme modos de olhar históricos, genealógicos ou interpretativos, para compor um domínio de construção de objetos ao que se aplicam metodologias explicativas, de concepção e modelização. (GOMEZ , 2017)

Ao mesmo tempo, as tecnologias digitais vão compor o domínio investigativo das ciências humanas e sociais, além de providenciar recursos procedimentais inovadores (como na e-ciência e na ciência aberta e colaborativa). Em sua própria constituição, os estudos da informação surgem num cruzamento de fronteiras, e o entendimento dessa posição, na formação histórica das ciências e das técnicas, seria parte importante daquilo que será objeto de sua crítica e reflexão. Portanto, conforme Gomez (2017, p.1), podemos afirmar que a fronteira tênue de tecnologia e ciências humanas abordada não é somente uma mescla, mas sim um tópico por si só: humanidades digitais.

Segundo Moura (2019), humanidades digitais “são uma nova forma de se abordar as pesquisas em humanidades e ciências sociais, particularmente atravessadas pelo imperativo tecnológico.”, ou seja, uma junção de áreas diferentes, que formam uma específica. O tema principal do trabalho não é a tecnologia da informação pura ou a Ciência da Informação, mas sim a mesclagem de ambos: produtos/serviços de informação.

O serviço/produto informacional que analisamos na pesquisa é uma *Dashboard*, produto informacional que pode ser definido como a junção de ciências exatas e humanas, pois, *Dashboards* permitem a visualização dos dados, utilizando da tecnologia para projetar e destacar informações pertinentes, e, ocultar dados desnecessários que não são úteis para o usuário. Segundo Succi, et al. (2013, tradução livre).

Uma *Dashboard* é o uso da tecnologia com Ciência da Informação, que fornece uma diferente abordagem para alguns produtos do campo da Ciência da Informação, como: relatórios de negócios, FAQ's, memorandos, entre outros. (SUCCI, et al. . 2013, tradução livre).

Devido a tal embasamento, podemos considerar a *Dashboard* um filho das humanidades digitais, pois mescla tecnologia, ciências humanas, ciências sociais aplicadas e ciências exatas, podendo até mesmo abranger ainda mais áreas dependendo de seu contexto.

Uma *Dashboard* é principalmente um produto/serviço informacional, conforme FONSECA(2002) afirma que uma dashboard pode empregar tanto a função de produto como serviço, pois, a mesma se trata de um insumo, fonte para produção de conteúdo intelectual, quanto um serviço de utilidade que traz benefícios para seus usuários. A dashboard é produto se origina do dado e tem como principal objetivo transmitir informação ou conhecimento para aquele que irá consumir deste. A fonte de um produto varia entre

“fontes mais diversificadas e dispersas, sem tratamento analítico, tais como: tabelas estatísticas, jornais, revistas, relatórios, documentos oficiais, cartas, filmes, fotografias, pinturas, tapeçarias, relatórios de empresas, vídeos de programas de televisão, etc.” (FONSECA, 2002).

Da mesma forma, o produto informacional em si varia, porém, ele tende a ser algo mais visual, sendo normalmente uma *Dashboard* ou uma planilha.

O produto informacional *Dashboard* pode ser estudado por diferentes perspectivas como design, tecnologia, ciência de dados e até mesmo psicologia, porém, continua sendo um produto resultante de processos de uma ciência social aplicada, a gestão da informação, onde existem profissionais que atualmente trabalham com uma área específica para isso: visualização da

informação. Dentro da visualização são empregados vários conceitos com o objetivo de aumentar a eficácia da transmissão da informação, conforme cita Freitas et.al(2001) a visualização da informação é uma área que aplica técnicas

visando auxiliar o processo de análise e compreensão de um conjunto de dados, através de representações gráficas manipuláveis. Uma técnica de visualização é baseada numa representação visual e em mecanismos de interação que possibilitam ao usuário manipular essa representação de modo a melhor compreender o conjunto de dados ali representado.(FREITAS et.al, 2001)

Nesse sentido, o presente trabalho é uma releitura produto/serviço informacional *Dashboard* da eleição ordinária de 2022 do Estado do Paraná. Procura-se apontar problemas referentes à visualização dos dados que impactam na maneira pela qual a sociedade obtém e interpreta a informação disponível nesse produto informacional, impactando no melhor uso da informação e seu entendimento, evitando conclusões equivocadas por membros da sociedade.

1.1 PROBLEMATIZAÇÃO

Os problemas que norteiam a pesquisa são: Como os dados estão apresentados e visualizados na *Dashboard* da Eleição Geral Ordinária 2022 no Estado do Paraná? Quais características, natureza do produto informacional e o impacto da visualização dos dados e informações apresentadas? Em que medida um novo *Dashboard*, enquanto um produto/serviço informacional, poderá otimizar a análise, compreensão e interpretação dos dados eleitorais?

1.2 OBJETIVOS

Buscando nortear o estudo proposto, foram definidos os objetivos que visam propor uma releitura com possíveis soluções para os problemas encontrados na *Dashboard* da Eleição Geral Ordinária 2022 no Estado do Paraná adequados ao padrão indicado na literatura da área.

1.2.1 OBJETIVO GERAL

Elaborar um novo produto/serviço informacional no formato de *Dashboard* que utiliza de técnicas de visualização da informação, adotando como base o *Dashboard* do Tribunal Superior Eleitoral e os dados da eleição de 2022.

1.2.2 OBJETIVOS ESPECÍFICOS

Visando alcançar o objetivo geral, propomos os seguintes objetivos específicos:

- Apresentar o processo de ETL;
- Analisar o Dashboard da Eleição Geral Ordinária 2022 com foco no Estado do Paraná.
- Levantar, extrair, tratar e disponibilizar (ETL) os dados referentes à Eleição Geral Ordinária 2022 com foco no Estado do Paraná;
- Apresentar um novo Dashboard que integre as mudanças propostas na pesquisa.

1.3 JUSTIFICATIVA

Quando falamos de dados e informação, temos que ter ciência de que é um assunto delicado que gera impactos em diferentes áreas da sociedade, principalmente porque um dado errado ou mal apresentado pode ter uma alta repercussão, podendo ser uma fake news ou, até mesmo, gerar diversas informações falsas.

Outro fato que justifica o desenvolvimento de uma *Dashboard* é a produção em massa de novos painéis, já que o original não atende. Diversos jornais e meios de comunicação¹ acabaram por criar visões que atendiam melhor os usuários do que a versão do Órgão público, assim, notasse uma falha na Dashboard, pois, se ela atendesse corretamente seu propósito, estaria anexada nos sites do meio jornalístico.

Portanto, um portal de extrema importância como o TSE, pertencente ao governo federal, deve apresentar seus dados referente à eleição de maneira aberta, porém, apenas ser algo público não é suficiente, é necessário que essa informação esteja em um formato claro, adaptado a atender melhor os possíveis usuários com visualizações e apresentação de dados coerentes com sugestões da literatura pertinente, que o Dashboard atual não apresenta no momento. A importância do estudo e apresentação correta do produto informacional impacta na taxa efetiva de consumo e do correto consumo, entendimento completo do produto. Como afirma Porto et al (2012), o estudo dos dados eleitorais no Brasil é fundamental para entender o cenário político do país e informar seus cidadãos sobre as principais questões e candidatos em cada ciclo eleitoral. Ele fornece insights sobre as atitudes e preferências dos eleitores brasileiros e pode ser usado para desenvolver estratégias de campanha e propostas políticas mais eficazes.

1.3.1 ASPECTO SOCIAL, CIENTÍFICO E PESSOAL²

Visando a sociedade, proponho essa repaginação na Dashboard das Eleições Ordinárias do Estado do Paraná de 2022 para que seja possível evitar

¹ Globo: <https://infograficos.oglobo.globo.com/politica/eleicoes-2022/apuracao-2022-em-tempo-real-consulte-resultado.html#/presidente> ; Folha de S. Paulo: <https://www1.folha.uol.com.br/poder/eleicoes/2022/apuracao/2turno/brasil/>;

² Artigos sobre o tema: “Fake news: arma potente na batalha de narrativas das eleições 2018” (ALMEIDA, 2018); “O Enfrentamento Das Fake News No Processo Eleitoral A Partir Das Leis De Proteção De Dados” (SANTOS e FERNANDES, 2021)

casos de desinformação por equívoco na leitura de dados, e, que também acaba por auxiliar na comprovação de informações corretas.

No aspecto científico, a contribuição é principalmente por se tratar da real aplicação de métodos internacionalmente utilizados na realidade brasileira, adequando um produto/serviço informacional público a tal padrão, enquanto percebe-se as singularidades do mesmo e o adequa para estabelecer um parâmetro de referência para entusiastas de visualização da informação e análise de dados, para que os mesmos tenham uma referência nacional no qual podem se inspirar para o desenvolvimento de novos produtos.

Para o ponto de vista pessoal, o tema escolhido da pesquisa se deve a minha afinidade pela área, já que, desde minha adolescência já tinha interesse na área de Tecnologia da Informação, porém, fazer programas simples, como desenvolvedor, nunca me soou interessante, mas, através do uso da linguagem de programação, junto dos conceitos da Ciência da informação, agregar valor tornou-se um tópico qual possuo interesse, e devido a tal gosto por dados, informação, e conhecimento agregando valor, escolhi o tema.

2 REFERENCIAL TEÓRICO

Para o desenvolvimento de um novo produto de informação, é necessário um embasamento tanto da área de Ciência da Informação, quanto de visualização e Tecnologia da Informação. Esse conhecimento teórico da área existe para que boas práticas ocorram, enquanto segue a essência da pesquisa: um produto de informação de formato tecnológico.

Para a escolha do referencial teórico utilizado na pesquisa, foi primeiramente entendido do que se trata um Dashboard e sua definição como produto de informação. Após, foi aprofundado nos insumos utilizados para o desenvolvimento do produto, as bases de dados, onde, além dessa perspectiva do requisito para o produto, temos também o seu vínculo multidisciplinar a ser analisado. Também observamos técnicas e processos necessários para a melhor produção do produto/serviço conforme a literatura pertinente da área, junto de um entendimento das ferramentas a serem utilizadas e da técnica empregada.

2.1 SERVIÇOS/PRODUTOS DE INFORMAÇÃO

Segundo Akkari e Carpi (2019, p. 160), o desenvolvimento de um produto/serviço de informação consiste “no processo a partir do qual informações sobre o mercado são transformadas nas informações e nos bens necessários para a produção de um produto com fins comerciais”, ou seja, entendendo o mercado, as tendências e as necessidades do usuário cria-se um produto provindo dessas informações. E, para sintetizar melhor um produto, alguns aspectos básicos são levados em conta como: funcionalidade, segurança, estética, economia, sustentabilidade, ergonomia, produtividade e operacionalidade. Basicamente, levantamento de informações para problemas mercadológicos.

Produtos de informação são aqueles passíveis de consumo, ou seja, organizações bem sucedidas otimizam o uso da informação nos seus produtos e serviços, e determinam formas de satisfazer as necessidades dos seus clientes associando, desassociando e reassociando informação aos produtos.

De acordo com SILVA (2004, p. 33) produtos de informação se caracterizam como um conjunto de atributos tangíveis constituídos através do processo de gestão da informação, para o atendimento de necessidades. Segundo o autor, acerca dos serviços de informação é possível dizer que a característica essencial para o consumidor é satisfazer as necessidades e demandas de informação de seus “usuários”, criando aspectos memoráveis e surpreendentes, fazendo com que ele experiencie novas e boas sensações, mesmo que alguns destes não compreendam as suas necessidades informacionais. Olhando para o aspecto empresarial, o serviço de informação é parte de um produto necessário para suprir uma necessidade identificada, ou não, e ganhar vantagem competitiva dentro do ambiente de mercado. Tem aspectos importantes como a criação de valor, relacionamentos e capacidade de atender as expectativas de seus stakeholders.

Pode-se definir serviços como sendo uma experiência vivenciada, fortemente associado às relações de consumo e troca de bens, sendo serviço qualquer atividade onde a pessoa que o acessa não obtém a propriedade exclusiva do que foi adquirido. (SANTOS, 2022, p.2).

Abell (1994) explica que para a prestação de serviços de informação é necessário, inicialmente, identificar necessidades, promover meios confiáveis de captação e manipulação da informação, bem como promover acesso a essa informação, tanto para o staff que planeja, quanto para o operacional de uma organização.

Conforme entendemos a composição de um produto/serviço de informação, também faz-se necessário compreender a qual área do conhecimento a pesquisa e desenvolvimento permite devido a inúmeras quantidades de formatos e meios de disponibilização, sendo assim, podemos

dizer que uma área do conhecimento para essa junção acaba por abraçar o campo: Humanidades digitais.

2.2 HUMANIDADES DIGITAIS

O termo humanidades digitais parece um tanto confuso, afinal é difícil assimilar algo humano e digital/tecnológico sendo um só ser. Para melhor entendimento vamos usar da definição de Moura (2019) que afirma que as humanidades digitais são

uma nova forma de se abordar as pesquisas em humanidades e ciências sociais, particularmente atravessadas pelo imperativo tecnológico. Os referidos atravessamentos acontecem sob a égide de intensos fluxos de informação que permitem, em termos técnicos, a circulação de “significantes” livres da custódia dos “significados” (BAUMAN, 2012, p. 34), pela conectividade always on, “o fio invisível que se multiplica entrelaçando consciências, espaços, perguntas, desejos” (SANTAELLA; LEMOS, 2010, p. 55) em face do uso de tecnologias wireless e dispositivos móveis. Atualmente, a constatação da complexidade para se identificar a autoridade epistêmica trouxe inúmeros desafios às diferentes áreas do conhecimento. (MOURA, 2019)

Podemos dizer que essa conexão em real time, vinte e quatro horas por dia durante os sete dias da semana, forçam as fronteiras entre as diferentes áreas do conhecimento a desmoronar.

Segundo Alvarez et. al. (2020) as

Humanidades Digitais são uma área interdisciplinar que combina humanidade com a tecnologia para estudar o impacto das tecnologias digitais na cultura, comunicação e conhecimento humano. Este campo abrange uma ampla gama de assuntos, desde a preservação e acesso a materiais culturais e históricos, até a criação de novos meios de expressão e arte digital. (ALVAREZ et. al. ,2020)

Também é abordado o estudo de questões éticas e políticas relacionadas à tecnologia e sua relação com a sociedade, incluindo questões de privacidade, segurança de dados e acesso equitativo à informação. Além disso, as Humanidades Digitais exploram a forma como a tecnologia está mudando a maneira como as pessoas aprendem, se comunicam e criam conhecimento conforme afirma Moura (2019).

A área de abordagem das humanidades digitais é interdisciplinar, incorporando elementos de diferentes “casas” de conhecimento, como humanas, exatas e biológicas, também mesclam todas as disciplinas específicas, por exemplo, em humanas: história, literatura, arte, filosofia, antropologia e estudos culturais. A tecnologia utilizada nas Humanidades Digitais inclui ferramentas de computação, como análise de dados, visualização de dados, processamento de linguagem natural e inteligência artificial, tópicos extremamente abordados no conjunto do trabalho.

O objetivo das Humanidades Digitais é fornecer uma compreensão mais profunda da cultura, da sociedade e do conhecimento humano, bem como apoiar a criação de novas formas de expressão e arte. A pesquisa nas Humanidades Digitais é importante para compreender como a tecnologia está moldando a sociedade e como podemos usá-la de maneira responsável e ética para melhorar a vida humana.

Uma das formas pela qual ocorre a compreensão e como podemos utilizar a tecnologia de maneira mais ética e responsável é no campo da visualização da informação, área cujo impacto interfere na compreensão do consumidor do produto/serviço da informação. Como tantos outros produtos/serviços informacionais, uma visualização da informação também precisa beber de uma fonte, uma origem, afinal, uma fonte de dados ou informações existentes é necessária para sua construção, sendo uma base de dados em nosso caso.

2.3 BASE DE DADOS ABERTOS

Uma base de dados consiste num arquivo, digital ou físico, que possui símbolos em seu corpo. Um exemplo seria um ficheiro, qual possui folhas com símbolos escritos, que sozinhos não significam muito, porém, dado o contexto do tema deste ficheiro, passam a ter significado, se tornando informação.

Segundo Silva (2022), uma base de dados pode ser definida como um

sistema informacional baseado em registros de acervos geralmente bibliográficos, que formam uma coleção digital de referência, organizada e alimentada por bibliotecas, corporações, associações, universidades ou por voluntários (SILVA, 2022)

Um banco de dados é uma base de dados, mas, não necessariamente toda base de dados é um banco de dados. É necessário ter isso em mente, já que, possuímos base de dados físicas existentes, principalmente no governo, sendo que, boa parte já é digital, mas, não está inserida em formato de banco de dados, muito dos dados está contido em arquivos de texto, planilhas, e/ou comprimidos em formatos de arquivos como JSON.

Quando falamos de base de dados abertos estamos nos referindo a dados disponibilizados tanto pela iniciativa privada quanto pública que podem ser acessados por qualquer usuário. Segundo Silva (2010)

o conceito (...) se relaciona com um entendimento de que a forma como os governos disponibilizam suas informações permite que a inteligência coletiva crie melhores formas de trabalhar com elas do que os próprios governos poderiam fazer (SILVA, 2010)

Silva (2010) também cita a utilidade dos dados abertos para aperfeiçoamento de tecnologias e processos que beneficiam a sociedade como um todo.

Diniz (2010) cita a disponibilização de dados abertos afirmando benefícios, pois

A disponibilização de dados governamentais abertos permite que as informações sejam utilizadas da maneira e conveniência do interessado de tal forma que elas possam ser misturadas e combinadas para agregar mais valor aos dados (DINIZ, 2010)

Podemos afirmar que as bases de dados abertos são essenciais para superar limitações cotidianas que os usuários de informações podem vir a encontrar, já que com eles é possível entender, utilizar e melhorar produtos informações e produtos, conseqüentemente trazendo evoluções em diversas áreas.

A base de dados do TSE, com os dados da Eleição Geral Ordinária 2022 no Estado do Paraná, é uma base de dados aberta e se encontra tanto no formato de banco de dados, quanto no de arquivo digital. Para realizar a junção

do banco com os arquivos, será realizada a unificação, tornando as diversas fontes em uma única base de dados. Portanto, nosso caso é específico, iremos utilizar de uma base de dados, qual a temática dos dados é eleitoral.

Nos últimos anos, os dados eleitorais no Brasil tornaram-se mais acessíveis e fáceis de analisar, graças aos avanços da tecnologia e à disponibilidade de novas fontes de dados, como mídias sociais e pesquisas online. Esses desenvolvimentos proporcionam novas oportunidades para pesquisadores, partidos políticos e jornalistas entenderem as tendências e a dinâmica das eleições brasileiras. No geral, os dados eleitorais no Brasil desempenham um papel crítico na formação do cenário político do país e na informação de seus cidadãos sobre as principais questões e candidatos em cada ciclo eleitoral.

2.2.1 DADOS ELEITORAIS

Um dado é um símbolo, palavra, número fora de contexto, que, não representa algo significativo para alguém, segundo Angeloni (2003) “São símbolos e imagens que não dissipam nossas incertezas. Eles constituem a matéria-prima da informação. Dados sem qualidade levam a informações e decisões da mesma natureza”. Porém, um dado enriquecido, com o contexto correto, vem a se tornar informação.

Quando citamos dados eleitorais, de certa forma, estamos adquirindo símbolos, números, palavras, disponibilizadas pelo TSE, que de certa forma estão inseridas em um contexto: Eleição Geral Ordinária 2022. Mesmo esses dados pertencendo a tal contexto, eles não deixam de ser apenas dados, já que por si só não possuem significado algum para quem os recebe, mas ao mesmo tempo cria-se um paradigma: ele está inserido no contexto da eleição. Mas, por tal dado não estar contextualizado e não ter seu significado explícito por si só, dependendo de indicadores e classificação, ele não é considerado uma informação, sendo assim, apenas um dado eleitoral.

Portanto, a pesquisa trata dos dados eleitorais da Eleição Geral Ordinária 2022 do Estado do Paraná, sendo esses dados eleitorais constituídos por um conjunto de base de dados referente às eleições ordinárias. Porém, o conceito de dados eleitorais é o mesmo para todo o Brasil.

O Brasil realiza eleições presidenciais a cada quatro anos para determinar o chefe do Executivo do país. Os dados eleitorais no Brasil são coletados e analisados por várias organizações, incluindo autoridades eleitorais do país, partidos políticos e organizações de pesquisa independentes.

Segundo Maria Tereza Sadek (2021), os dados eleitorais no Brasil fornecem informações valiosas sobre o cenário político e os padrões de votação dos cidadãos do país. A coleta e análise de dados eleitorais tornou-se cada vez mais importante no era digital, pois é usado para rastrear a popularidade de candidatos e partidos políticos e para entender os fatores que influenciam o comportamento do eleitor. Esses dados públicos ainda se encontram num formato que impossibilite rastrear o voto do eleitor, porém, é possível observar as tendências e fazer levantamentos baseado em dados das seções eleitorais, quais, representam a vontade dessa região eleitoral e, conseqüentemente, a vontade do eleitor.

Tendo a fonte de dados, passamos a poder modelar um produto/serviço informacional, e, a forma abordada para a modelagem foi a construção da visualização da informação.

2.4 VISUALIZAÇÃO DA INFORMAÇÃO

Nesta pesquisa, o objetivo será alcançado através do desenvolvimento de uma nova *Dashboard*, que, se trata de um produto/serviço de informação focado principalmente na transmissão da informação através de sua visualização, seja

por gráficos ou mapas com dados referentes a Eleição Geral Ordinária 2022 no Estado do Paraná.

O objetivo principal de um produto/serviço informacional focado na visualização da informação, segundo Sasieta e Pacheco(2001), é

representar visualmente informações com o objetivo de facilitar a compreensão dos conjuntos de dados que, à primeira vista, são voluminosos ou não são facilmente de compreender em seu estado original, tornando as informações relevantes em objetos mais naturais, mais semelhantes a outros objetos que já são conhecidos (SASIETA e PACHECO, 2001)

A visualização da informação por si só é uma ciência social aplicada, já que abrange grande multidisciplinaridade, como Tecnologia da Informação., Ciência da Informação e Design. Como uma ciência, ela possui seus próprios princípios, porém, para maior entendimento da motivação de seu uso, podemos citar Knaflic (2019) "Temos todas as informações do mundo na ponta de nossos dedos não facilita nossa comunicação: dificulta. Quanto mais informações temos de lidar, mais difícil é filtrar o que é mais importante".

Portanto, a área de visualização da informação, através de seus princípios e boas práticas, leva o consumidor do produto/serviço informacional a interpretar de maneira eficiente e completa as informações apresentadas no produto. Conforme Succi (2013) afirma, *Dashboard* é usado para descrever um sistema que visualiza dados úteis para a tomada de decisão. Os *Dashboards*, como em um carro, têm o objetivo de informar sem distrair os usuários de suas tarefas. Portanto, os dados nos painéis são resumidos usando gráficos, tabelas, medidores e assim por diante. Para permitir que os usuários interpretem um elemento no painel corretamente, os painéis geralmente permitem que eles vejam os dados originais nos quais a compactação foi baseada. Portanto, para tal interpretação ocorrer, deve existir uma boa visualização, que seja coerente com os dados a serem apresentados e respeite o jogo de cores do contexto no qual se baseia, e, no qual o ser humano identifica com maior facilidade os sinais.

Deve-se entender que algumas técnicas de visualização devem ser empregadas para que ocorra a correta e eficiente interpretação do que é mostrado em uma Dashboard.

2.5 TÉCNICAS DE VISUALIZAÇÃO DE DADOS

A visualização de dados é a representação de dados ou informações em um formato gráfico ou pictórico. Considerando a grande quantidade de dados produzidos durante as eleições, torna-se muito difícil analisá-los em sua forma bruta. Portanto, diferentes técnicas de visualização de dados devem ser empregadas com o intuito de auxiliar no processo de análise. Dependendo do tipo de dado (temporal, quantitativo, qualitativo) e do contexto a ele empregado, é determinado um tipo de técnica de visualização específica.

Essa técnica trata-se de uma forma de representação: um gráfico. Essa forma de representação pode ter diferentes formatos, como linhas, pontos, até mesmo regiões quando falamos de gráficos em formato de mapa. Além disso, existem também diferentes propriedades que mudam o sentido da mensagem a ser passada: tamanho, forma, espessura, textura e cor.

A seguir é descrito as principais e mais populares técnicas de visualização da informação, conforme entendimento do livro *Storytelling com dados* (Knaflic, 2019).

Começaremos pelo gráfico de barras, ele é um tipo de visualização de dados que representa dados categóricos usando barras retangulares. Cada barra representa uma categoria e a altura ou comprimento da barra representa a magnitude dos dados dentro dessa categoria. Os gráficos de barras são usados para comparar o tamanho de diferentes categorias, para mostrar mudanças ao longo do tempo ou para mostrar as proporções de diferentes categorias em um todo. As categorias podem ser ordenadas vertical ou horizontalmente, e as barras podem ser exibidas adjacentes ou empilhadas. Os gráficos de barras são uma maneira simples e eficaz de exibir e comparar dados.

Outro tipo de gráfico é o de linhas, o qual representa uma série de pontos de dados conectados por uma linha. Os pontos de dados em um gráfico de linhas geralmente são plotados ao longo de um eixo comum, e a linha conecta os pontos de dados para mostrar tendências e padrões ao longo do tempo. Os gráficos de linhas são frequentemente usados para mostrar alterações nos dados em um intervalo contínuo, como um mês, trimestre ou ano. Eles são comumente usados para mostrar mudanças nas tendências, como preços de ações, números de vendas ou crescimento populacional. Os gráficos de linha também podem ser usados para comparar dados em várias séries, exibindo várias linhas no mesmo gráfico, cada uma representando uma série separada de dados. Eles são uma maneira simples e eficaz de visualizar tendências nos dados ao longo do tempo.

Um dos mais polêmicos gráficos é o de pizza, pois, para muitos autores como Knaflic(2019), o mesmo é muito específico, e deve ser utilizado raramente, pois é ineficiente para transmitir informação.

O gráfico de pizza trata-se de uma representação dos dados como um círculo dividido em setores, onde cada setor representa uma parte do todo. O tamanho de cada setor é proporcional à quantidade que ele representa, e a soma de todos os setores é igual a 100%. Os gráficos de pizza são usados para mostrar a proporção de diferentes categorias em um todo e são frequentemente usados quando os dados consistem em um pequeno número de categorias. Cada setor em um gráfico de pizza é normalmente rotulado com seu nome e valor de categoria correspondente, e uma legenda pode ser usada para identificar as categorias. Ele é uma maneira simples e eficaz de exibir proporções relativas e são amplamente usados em negócios, finanças e outros campos, porém, devem somente ser utilizado quando há poucas categorias, pois, devido ao seu formato, acaba por ser de difícil entendimento caso exista muita informação diferente a ser transmitida.

Temos também o gráfico de dispersão, que representa duas variáveis contínuas, plotando pontos de dados individuais como coordenadas em um gráfico bidimensional. Cada ponto de dados é representado como um único ponto no gráfico, com a posição do ponto representando os valores das duas

variáveis. Os gráficos de dispersão são usados para mostrar a relação entre duas variáveis e podem revelar padrões como correlação, agrupamento ou outliers. Os gráficos de dispersão também podem ser usados para visualizar a distribuição dos dados, mostrando a densidade de pontos em diferentes áreas do gráfico. A orientação dos pontos no gráfico de dispersão pode indicar uma relação positiva, negativa ou neutra entre as variáveis, e a forma dos pontos pode fornecer informações adicionais sobre os dados. Os gráficos de dispersão são uma ferramenta poderosa para explorar e compreender as relações nos dados.

Ainda em gráficos, temos alguns específicos para representações de intensidade, como o heatmap. É uma forma gráfica de realizar a representação dos valores de dados como cores em uma grade ou matriz bidimensional. A cor de cada célula na matriz representa o valor do ponto de dados correspondente, com valores mais altos normalmente representados por cores mais escuras e valores mais baixos representados por cores mais claras. Os mapas de calor são usados para mostrar padrões ou correlações em grandes conjuntos de dados e podem revelar tendências e relacionamentos que podem não ser imediatamente aparentes em outras formas de visualização.

Os mapas de calor são particularmente úteis para comparar várias variáveis, pois permitem a representação simultânea de várias dimensões de dados. Eles podem ser usados em muitos campos, incluindo finanças, biologia e marketing, para explorar relações entre variáveis e identificar áreas de interesse. Os mapas de calor podem ser criados usando vários esquemas de cores e podem ser animados para mostrar alterações ao longo do tempo ou em resposta à interação do usuário.

2.6 PROCESSO DE ETL

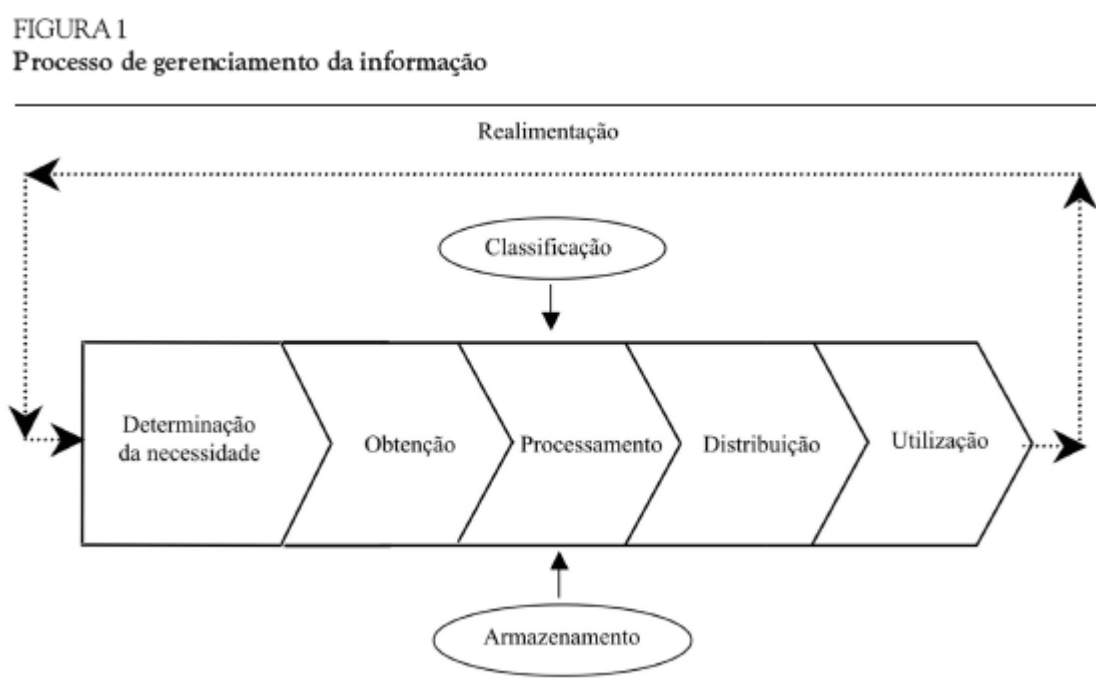
Com essas técnicas para melhor apresentar os dados em mente, deve-se formar um esboço mental de como será o produto informacional final, porém,

antes de começar a parte gráfica, deve-se voltar um passo atrás, e entender a melhor forma de obter os dados, o insumo, para alimentar a Dashboard. Para isso é aplicado a técnica “extract, treat and load”, ou seja, o processo de ETL.

“A ETL é definida como um processo para extrair dados de um sistema de Bases de Dados (BD), sendo esses dados processados, modificados, e posteriormente inseridos numa outra BD”. (ABELHA; MACHADO, 2010).

Uma ETL é um processo que lembra a gestão da informação, devido a sua forma de processo, já que, enquanto na ETL temos a fase de levantar o que é pertinente, extrair, tratar, e disponibilizar, na gestão da informação temos as diferentes etapas com processos semelhantes conforme indicado na FIGURA 1. Segundo Moraes e Escrivão Filho (2006) trata-se da determinação das necessidades, obtenção, processamento (classificação e armazenamento), distribuição e utilização da informação. Ambos os processos também são feitos de forma cíclica, porém, o de ETL pode ocorrer de forma única, ao contrário do processo de gestão da informação que é um ciclo.

FIGURA 1 - Processo de GI 1



Fonte: Moraes e Escrivão Filho (2006, p. 126).

O processo de ETL faz a integração entre diversos dados, de diversas fontes, de diversos formatos, qual, unifica o mesmo em uma única base de dados, de modo com que ela possa ser consumida e atenda a necessidade do usuário, onde, essa necessidade pode ser a própria base, ou, um produto/serviço informacional melhor trabalhado.

O processo de ETL como um todo pode ser feito de diversas maneiras, desde utilizando softwares a linguagens de programação, ou, um mix dos mesmos. Para este trabalho utilizaremos deste misto, utilizando softwares de planilhas e linguagem de programação para realização do processo de ETL.

O processo de ETL envolve extrair dados de várias fontes, transformá-los em um formato consistente e carregá-los em um sistema de destino, como um data warehouse, para análise e geração de relatórios. Suas etapas são:

- **Extração:** durante o estágio de extração, os dados são recuperados de fontes como bancos de dados, arquivos simples e APIs.
- **Transformação:** o estágio de transformação envolve limpeza, filtragem e conversão dos dados em um formato compatível com o sistema de destino. Esta etapa também envolve a aplicação de qualquer enriquecimento e agregação de dados necessários.
- **Disponibilização:** por último os dados transformados são carregados no sistema de destino, seja esse sistema um produto/serviço informacional, ou um sistema terceiro. Esta etapa pode ser realizada por meio de processamento em lote ou integração de dados em tempo real, porém, comumente requer uma estrutura muito cara conforme o mais próximo do tempo real, sendo assim, essa disponibilização tem o tempo de sua integração definida conforme necessidade e poder aquisitivo.

Mesmo sendo um processo estrutural de computadores, as ETL's "devem ser flexíveis e adaptáveis para acomodar requisitos de negócios e fontes de dados em constante mudança. Automatizando o processo ETL usando

ferramentas e tecnologias podem melhorar a qualidade dos dados e reduzir o tempo e os recursos necessários para concluir o processo” (Leonard, Fast, 2004)

Um dos requisitos para a ETL é o uso de linguagens de programação, que tem papel fundamental quando entramos no campo da computação, pois, através delas, conseguimos gerar e obter dados.

2.6.1 LINGUAGENS DE PROGRAMAÇÃO

De modo geral, as linguagens são utilizadas para comunicação, como por exemplo o português é utilizado no Brasil para nos comunicarmos no território brasileiro, assim como o inglês nos Estados Unidos da América. Porém, não podemos utilizar do português para programar computadores, o problema encontra-se na ambiguidade das linguagens naturais, ou seja, uma sentença escrita em português, por exemplo, pode representar diferentes significados, dependendo do contexto ou até mesmo possuir diferentes interpretações. Segundo Macedo et al,

Existem relatos de linguagens de programação muito antes de 1940, que foi quando as primeiras linguagens de programação modernas e os computadores começaram a surgir. As linguagens de programação no começo eram códigos matemáticos. A ideia de uma linguagem de programação era um código especializado para uma aplicação. As linguagens de programação surgiram da evolução da lógica matemática, no qual abstrai conceitos complexos da matemática e podia ser utilizada para resolver problemas específicos. (Macedo et al, 2019)

Podemos afirmar que uma linguagem de programação se baseia numa sequência de ordens, em formato imperativo, onde, o computador, entende essa instrução como algo que deve ser feito ao pé da letra, onde a língua utilizada é própria da máquina. Tal formato é feito para que, através dessa linguagem própria, seja evitado erros semânticos e de significados múltiplos que são encontrados em todas as línguas, principalmente ao comparar uma língua com outra.

Outras figuras extremamente importantes quando falamos de linguagens de programação são API's e bibliotecas. Ambas são ferramentas da própria linguagem de programação selecionada.

Uma API (Application Programming Interface) é um conjunto de protocolos, rotinas e ferramentas para a construção de software e aplicativos. Ele define a maneira como diferentes sistemas de software interagem entre si e acessam os dados uns dos outros.

Em outras palavras, elas representam uma maneira de diferentes aplicativos se comunicarem e trocarem informações, sem estarem fortemente acoplados uns aos outros. Elas permitem que diferentes sistemas se comuniquem e compartilhem dados e funcionalidades. Por exemplo, uma API pode permitir que um site de clima acesse e exiba dados de um serviço de previsão do tempo ou permita que um aplicativo móvel acesse dados de uma plataforma de mídia social.

Existem dois diferentes tipos de API's: abertas e fechadas. As APIs abertas estão disponíveis para desenvolvedores e outras organizações, permitindo que acessem e usem os dados e a funcionalidade fornecidos pela API. As APIs fechadas estão disponíveis apenas para um número limitado de desenvolvedores ou para a empresa que criou a API.

As APIs desempenham um papel crítico no desenvolvimento de software moderno, permitindo comunicação perfeita e troca de dados entre diferentes sistemas e aplicativos.

Já as bibliotecas na linguagem de programação, segundo Macedo et al (2019) é definida como "um código pré-escrito e reutilizável que pode ser facilmente integrado a um aplicativo de software para fornecer funcionalidade adicional. As bibliotecas geralmente contêm um conjunto de funções, classes e módulos que podem ser usados para executar tarefas comuns, como E/S de arquivo, manipulação de strings, cálculos matemáticos e comunicação de rede.". As bibliotecas são uma parte crucial do processo de desenvolvimento de software, pois permitem que os desenvolvedores economizem tempo e esforço, evitando a necessidade de escrever código de baixo nível a partir do zero. Em

vez disso, eles podem simplesmente usar uma biblioteca para executar tarefas específicas e se concentrar em escrever a lógica de alto nível de seu aplicativo, software ou processo de programação (ETL).

Existem diferentes tipos de bibliotecas disponíveis para diferentes linguagens de programação, como bibliotecas padrão incluídas na linguagem, bibliotecas de terceiros que podem ser baixadas da Internet e bibliotecas personalizadas criadas internamente. As bibliotecas podem ser usadas para uma ampla variedade de finalidades, desde tarefas simples, como ler e gravar arquivos, até tarefas complexas, como aprendizado de máquina e visão computacional.

O uso de bibliotecas é uma prática comum no desenvolvimento de software, pois ajuda a aumentar a eficiência, confiabilidade e manutenibilidade dos aplicativos de software. As bibliotecas também ajudam a promover a reutilização de código e a melhorar a qualidade geral do software, permitindo que os desenvolvedores desenvolvam soluções existentes, em vez de ter que reinventar a roda toda vez que precisam realizar uma tarefa específica.

Mesmo tão presentes, as linguagens de programação são uma ferramenta para atividades de diversas áreas, onde, para Hadley Wickham, G. G. (2017) "As linguagens de programação são a ferramenta básica para a manipulação de dados em escala."

As linguagens de programação desempenham um papel crucial no processo de análise de dados. Eles fornecem as ferramentas e estruturas necessárias para coletar, manipular, analisar e visualizar dados.

Os dados podem ser armazenados de várias formas, como bancos de dados, planilhas ou arquivos de texto, e uma linguagem de programação permite que um programador interaja com esses dados. Por exemplo, um programador pode usar uma linguagem como Python, SQL ou R para recuperar dados de um banco de dados e executar limpeza e pré-processamento de dados, como lidar com valores ausentes ou discrepantes.

Linguagens de programação e a análise de dados estão intimamente ligadas, com as linguagens de programação fornecendo as ferramentas necessárias para coletar, manipular, analisar e visualizar dados. A capacidade de analisar dados com eficiência está se tornando cada vez mais importante no mundo orientado a dados de hoje, e as linguagens de programação desempenham um papel crucial nesse processo. Portanto, para execução da pesquisa foi escolhido duas linguagens de programação que suprem todas as necessidades do trabalho: Python e SQL.

White e Conway (2012) define python como

uma linguagem de programação de alto nível, interpretada e de uso geral. Criado no final dos anos 1980 por Guido van Rossum, o Python foi projetado para enfatizar a legibilidade do código e possui uma sintaxe que permite aos programadores expressar conceitos em menos linhas de código do que outras linguagens de programação, como C++ ou Java.(WHITE e CONWAY, 2012)

Alguns dos principais recursos do Python incluem:

- **Digitado dinamicamente:** Python usa tipagem dinâmica, o que significa que você não precisa especificar o tipo de dados de uma variável ao declará-la. Em vez disso, o Python descobrirá automaticamente o tipo de dados em tempo de execução.
- **Interpretado:** Python é uma linguagem interpretada, o que significa que o código é executado linha por linha, em vez de ser compilado em código de máquina antes da execução. Isso facilita a depuração e o teste do código, além de torná-lo mais flexível.
- **Orientado a objetos:** Python é uma linguagem de programação orientada a objetos, o que significa que fornece suporte para encapsulamento, herança e polimorfismo. Isso torna mais fácil organizar e reutilizar o código e também fornece uma maneira mais intuitiva de pensar sobre problemas complexos.
- **Grande biblioteca padrão:** o Python possui uma biblioteca padrão grande e abrangente, que fornece suporte para uma ampla gama de tarefas, desde conectar-se a um banco de dados, trabalhar com HTTP e FTP, até analisar XML e JSON.

Python é amplamente usado para uma variedade de tarefas, incluindo desenvolvimento web, computação científica, análise de dados, inteligência artificial e muito mais. Também é conhecido por sua simplicidade e facilidade de uso, o que o torna uma escolha popular para iniciantes e programadores experientes.

Já o SQL (Structured Query Language) é uma linguagem de programação usada para gerenciar e manipular bancos de dados relacionais. Ele é usado para inserir, atualizar e recuperar dados de um banco de dados. Segundo Winand(2011) SQL e suas versões são uma linguagem de programação poderosa amplamente utilizada para realizar estudos e análise de dados.

Essa linguagem é padrão para diversos bancos de dados relacionais, como MySQL, Oracle e Microsoft SQL Server. Alguns dos principais recursos do SQL incluem:

- *Data Definition Language (DDL)*: SQL inclui um conjunto de comandos para definir a estrutura de um banco de dados, incluindo a criação de tabelas, definição de relacionamentos entre tabelas e definição de restrições.
- *Data Manipulation Language (DML)*: SQL inclui um conjunto de comandos para inserir, atualizar e excluir dados em um banco de dados, bem como recuperar dados usando instruções SELECT.
- *Transações*: o SQL oferece suporte a transações, que permitem que várias operações sejam executadas como uma única unidade de trabalho. Se qualquer parte de uma transação falhar, toda a transação é revertida, garantindo a integridade dos dados.
- *Índices*: o SQL oferece suporte a índices, que fornecem acesso rápido aos dados armazenados em um banco de dados. Isso pode melhorar significativamente o desempenho ao trabalhar com grandes conjuntos de dados.

O SQL é amplamente usado por empresas, organizações governamentais e indivíduos para gerenciamento de dados. É uma ferramenta crítica para análise de dados e tomada de decisão e é usada por analistas de dados, cientistas de dados e administradores de banco de dados.

Mesmo com essas linguagens, ainda assim, temos a utilização de software de planilhas, cuja principal função é dar uma visão mais user friendly de dados, porém, ao mesmo tempo, possibilitando a execução de diversos processos de manipulação de dados.

2.7 ANÁLISE DE DADOS

Com esse conjunto de linguagens, processos, e ferramentas, é necessário a execução do principal processo para que ocorra a criação de um produto/serviço informacional de qualidade, a análise de dados. O processo de análise de dados é dividido em examinar, limpar, transformar e modelar dados, com o objetivo de descobrir informações úteis, tirar conclusões e apoiar a tomada de decisões, conforme afirma Drew Conway e John Myles White em seu livro "Machine Learning for Hackers", "a análise de dados é o processo de coletar dados, limpá-los, transformá-los e modelá-los com o objetivo de descobrir informações úteis que possam informar a tomada de decisões.". Envolve a aplicação de técnicas estatísticas, matemáticas e computacionais para extrair *insights* e conhecimento dos dados.

De acordo com Hadley Wickham e Garrett Grolemund (2016, p. 81) "A análise de dados é um campo interdisciplinar que combina análise estatística, pensamento computacional e conhecimento de domínio para extrair *insights* e conhecimento dos dados". A análise de dados é um componente crítico da ciência de dados e é usada em uma ampla variedade de campos, incluindo negócios, finanças, ciências sociais, engenharia e saúde. O objetivo da análise de dados é transformar dados brutos em informações significativas e acionáveis que possam ser usadas para tomar decisões informadas.

O processo de análise de dados geralmente envolve várias etapas, incluindo limpeza de dados para remover erros e inconsistências, exploração de dados para entender as características dos dados, visualização de dados para representar os dados em um formato gráfico e modelagem estatística para testar hipóteses e fazer previsões. A análise de dados pode ser realizada usando uma ampla variedade de ferramentas, incluindo planilhas, software estatístico e linguagens de programação.

A análise de dados está se tornando cada vez mais importante à medida que as organizações geram e coletam grandes quantidades de dados, e é um componente crítico de muitos processos de tomada de decisão orientados por dados. Ele ajuda as organizações a entender seus dados, identificar padrões e tomar decisões informadas com base nos *insights* obtidos com os dados.

3. METODOLOGIA

Nesta seção será apresentado o campo de pesquisa, o produto/serviço informacional (objeto) da pesquisa, o escopo técnico da análise do produto/serviço informacional existente, técnicas de coleta, tratamento, análise e distribuição dos dados, junto da delimitação da população de dados qual será o foco do trabalho. Com a finalidade de mostrar como o processo de pesquisa foi realizado, como o material de estudo foi escolhido, entre outras informações pertinentes para possível verificação e auditoria do método e linha de trabalho escolhida.

3.1 CAMPO DE PESQUISA

O campo de pesquisa optado foi o instrumental prático, sendo o processo de pesquisa, coleta, tratamento, e distribuição dos dados e informações de

caráter histórico referente à Eleição Geral Ordinária 2022, tendo foco nos dados e informações do estado do Paraná. Foram utilizados livros, artigos científicos, bases de dados e API disponibilizada pelo TSE.

3.2 ANÁLISE DO PRODUTO EXISTENTE E ESTABELECIMENTO DO ESCOPO DO NOVO PRODUTO

Uma análise de todo o produto/serviço informacional *Dashboard* das eleições ordinárias de 2022 será realizada, sendo constatado seus principais pontos negativos e positivos. Os principais pontos a serem analisados, conforme o livro *Storytelling com dados*(Knafllic, 2019) são:

- Cores
- Tipos de gráficos
- Formato dos dados apresentados
- Navegação dentro do produto
- Pertinência das informações apresentadas
- Padrão de formato dos dados

3.3 PROCESSO DE ANÁLISE, TÉCNICAS DE COLETA, TRATAMENTO, E DISPONIBILIZAÇÃO DE DADOS(ELT)

Para exercer essa coleta, tratamento e disponibilização e o novo visual do Dashboard, deve-se antes fazer uma análise do Dashboard já existente, levando em conta o impacto social do mesmo, junto do seu valor como humanidade digital. Essa análise irá levantar os pontos de melhoria do produto/serviço informacional existente, percorrendo os tópicos de análise como produto informacional/serviço de informação, seu impacto como visualização de informação, adequação dos gráficos apresentados aos dados existentes, e, o valor que o mesmo agrega como análise de dados.

Como técnica de coleta de dados, a presente pesquisa obtém dados disponibilizados por parte do Tribunal Superior Eleitoral, e filtra campos necessários e pertinentes. Com as fontes explícitas, é feita a extração dos dados através da linguagem de programação Python, qual se comunica com base de dados disponibilizada pelo TSE. Dentre as principais fontes podemos listar:

- <https://dadosabertos.tse.jus.br/dataset/resultados-2022-boletim-de-urna>
- <https://dadosabertos.tse.jus.br/dataset/resultados-2022>
- <https://resultados.tse.jus.br/oficial/app/index.html#/eleicao/resultados>
- <https://www.portaltransparencia.gov.br/api-de-dados>
- API dados TSE.

Para o tratamento dos dados, todas as bases de dados foram exportadas para “data frames” através da linguagem Python, onde receberam o tratamento para que bugs, erros de digitação humano, e remoção de caracteres especiais ocorressem. Também ocorreu a criação de relações entre as bases de dados e a filtragem dos dados por estado, já que, o foco principal é a base de dados das eleições no estado do Paraná.

Terminando o tratamento dos dados ocorreu a fase de carregamento dos mesmos, eles foram disponibilizados num banco de dados (SQL Server), como forma de ficarem armazenados num ambiente com backup e com a facilidade de evitar reexecução dos códigos de tratamento, além da acessibilidade diretamente através de código sql para análises simples, e, python para complexas.

3.4 ELABORAÇÃO DA PROPOSTA DE PRODUTO INFORMACIONAL

Como ferramenta para o produto/serviço informacional *Dashboard* será escolhido a ferramenta PowerBI³, a mesma permite visualizações dinâmicas de informação, alterações de paleta de cores de forma personalizada, e, diversos tipos de gráficos de mapa. Foi realizada a conexão entre o PowerBI e o ambiente de PostgreSQL, que contém os dados tratados.

Na produção do novo produto/serviço foi utilizado o escopo previamente definido para o desenvolvimento do *Dashboard* e suas diferentes visões.

³ O Power BI é uma coleção de serviços de software, aplicativos e conectores que trabalham juntos para transformar suas fontes de dados não relacionadas em informações coerentes, visualmente envolventes e interativas.(MICROSOFT, 2023)

4 AVALIAÇÃO DA DASHBOARD EXISTENTE

Conforme os objetivos apresentados, faz-se necessário descrever o atual estado do produto informacional, para que possa ser entendido exatamente do que se trata, junto de suas características e finalidade do mesmo.

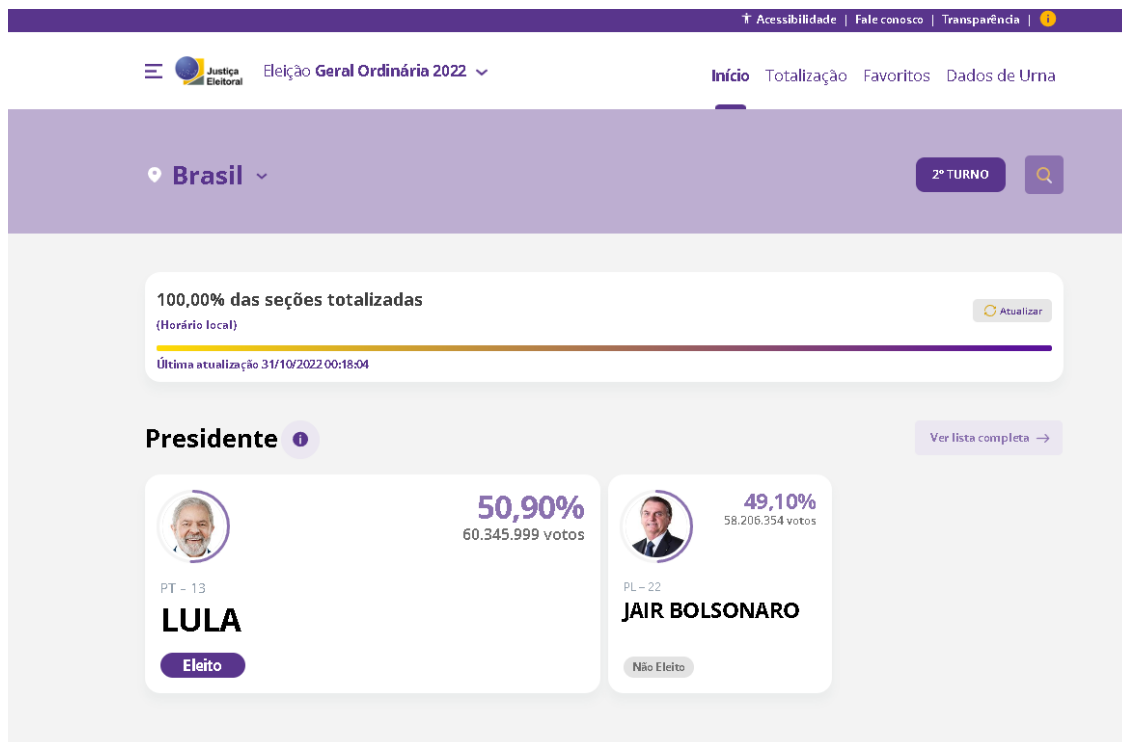
Além disso, deve-se ser analisada, sob o olhar de um gestor da informação, as características, os dados e informações, junto da forma qual é apresentada. Basicamente será uma análise contemplando diversos pontos técnicos, informacionais, e da visualização da informação.

Por fim, será abordado o impacto social desse produto/serviço de informação, na perspectiva da sociedade como um todo, não somente focado no Estado, porém, entende-se que os problemas existentes no país, também são problemas do estado do Paraná, pois o mesmo faz parte do Brasil.

4.1 DESCRIÇÃO E ANÁLISE DA DASHBOARD

O atual *Dashboard* das eleições ordinárias de 2022 pode ser acessado via portal do TSE, ou link direto: <https://resultados.tse.jus.br/>. Ao acessar o site da apuração dos resultados, o usuário é direcionado a uma aba simples, como mostrado na figura 2, com informações sobre os candidatos a presidente competindo no segundo turno e seu respectivo número de votos, % de votos, total de urnas apurados e uma tag indicando o Presidente eleito e o não eleito.

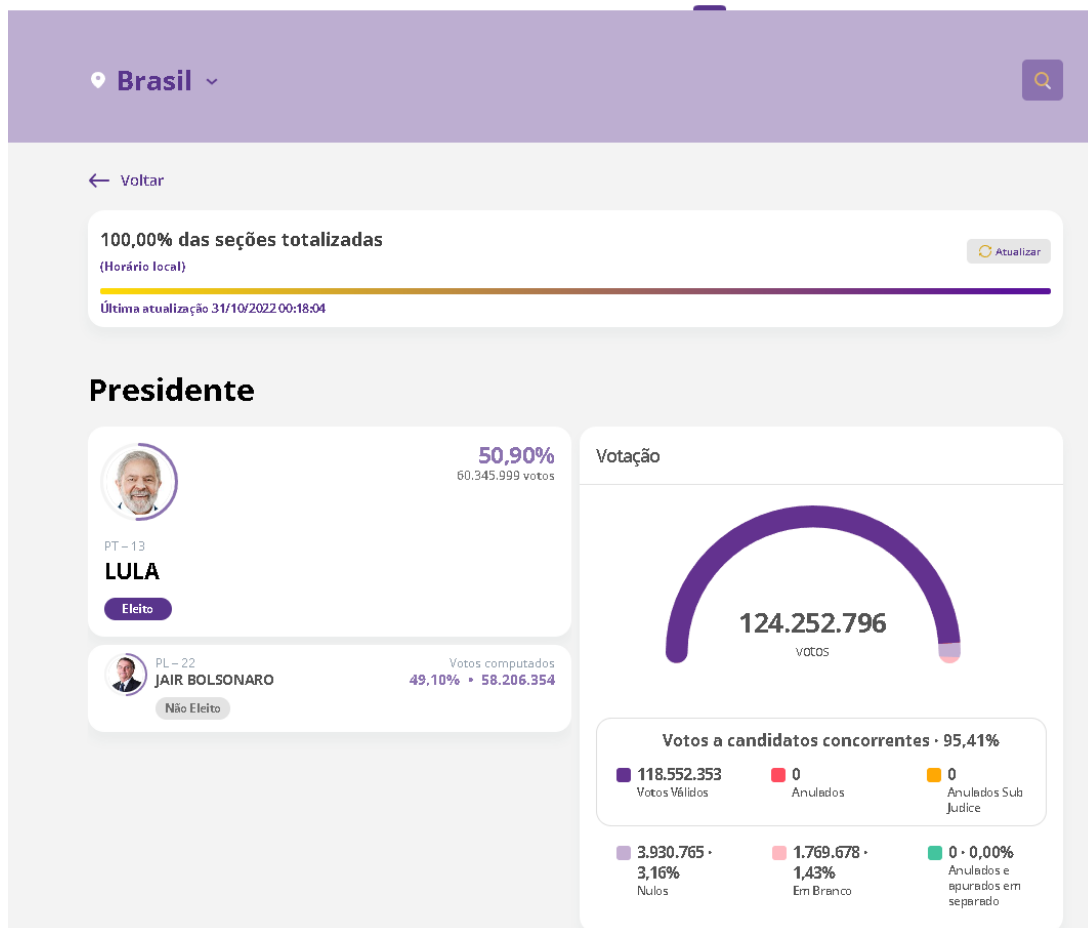
FIGURA 2 - Home page Dashboard das eleições ordinárias de 2022



Fonte: TSE (2023)

Existem outros botões de navegação, como o contido na palavra “Brasil” porém não é deixado explícito que é um botão, uma falha da visualização da informação deste produto. Na figura 3 temos o resultado de utilizar do botão “Ver lista completa”, ele apresenta os mesmos dados, porém, com o detalhamento de votos válidos, anulados, nulos, e em branco.

FIGURA 3 - Home page Dashboard das eleições ordinárias de 2022 - Botão “Ver lista Completa”

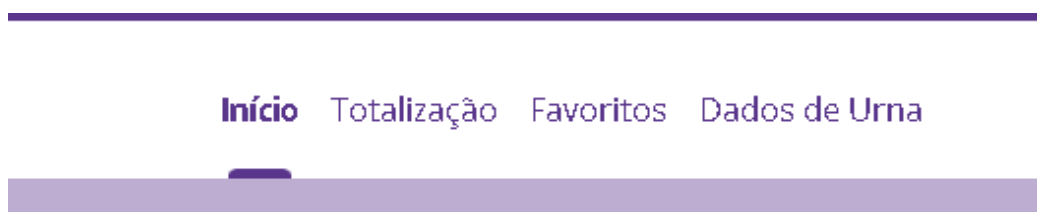


Fonte: TSE (2023)

As informações apresentadas após clicar no botão não diferem tanto da página inicial, porém, um problema é constante, o jogo de cores: conforme o Knaflic (2019) apresenta em seu livro, as cores possuem significado e deve ser usadas para indicar algo positivo, atenção, e, negativo, conforme a situação das informações apresentadas. Além do desrespeito desse conceito, as cores em tonalidades de roxo não condizem com alguma cor utilizada no governo brasileiro, tal escolha faz com que, para pessoas que observam pela primeira vez e não notam o botão Brasil, não seja possível ter a certeza de que se trata das eleições brasileiras.

Também existe outras opções no menu superior do site conforme indicado na figura 4:

FIGURA 4 - Menu do Dashboard



Fonte: TSE (2023)

O menu de Totalização apresenta tela idêntica ao disponível no botão “ver lista completa”.

FIGURA 5 - Página de totalização

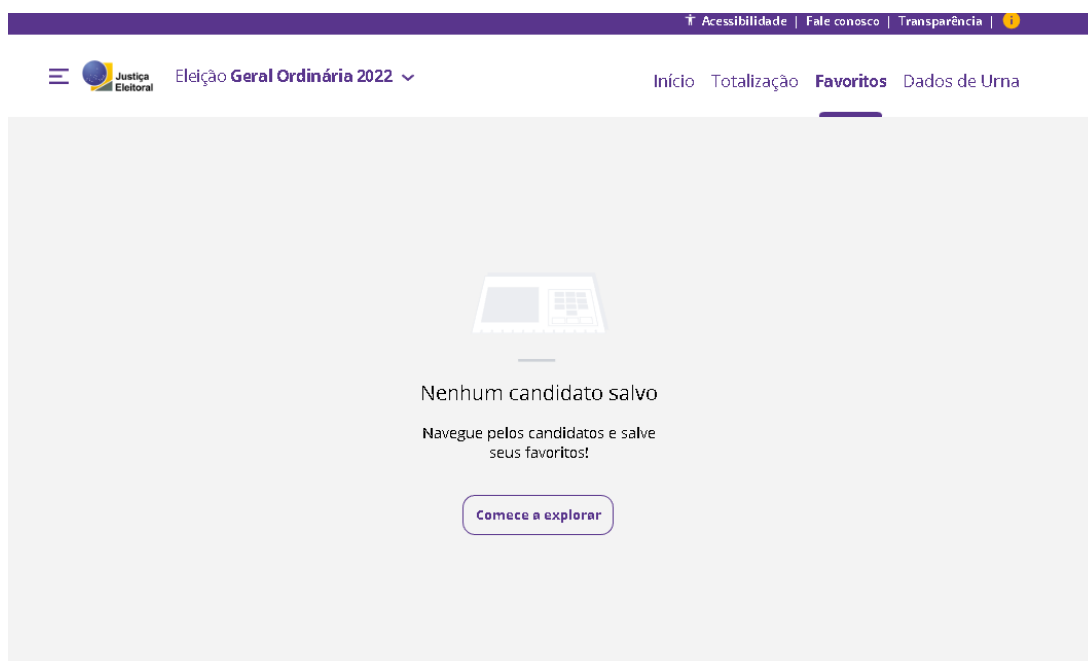


Fonte: TSE (2023)

Um problema com essa página é que, além da grande semelhança com a anterior, ela ainda possui o botão cargo sem funcionar, ele apenas tem a opção de presidente.

Na opção favoritos, figura 6, é possível salvar candidatos para acompanhar se ele foi ou não eleito, porém, tal processo necessita de cadastro do governo.

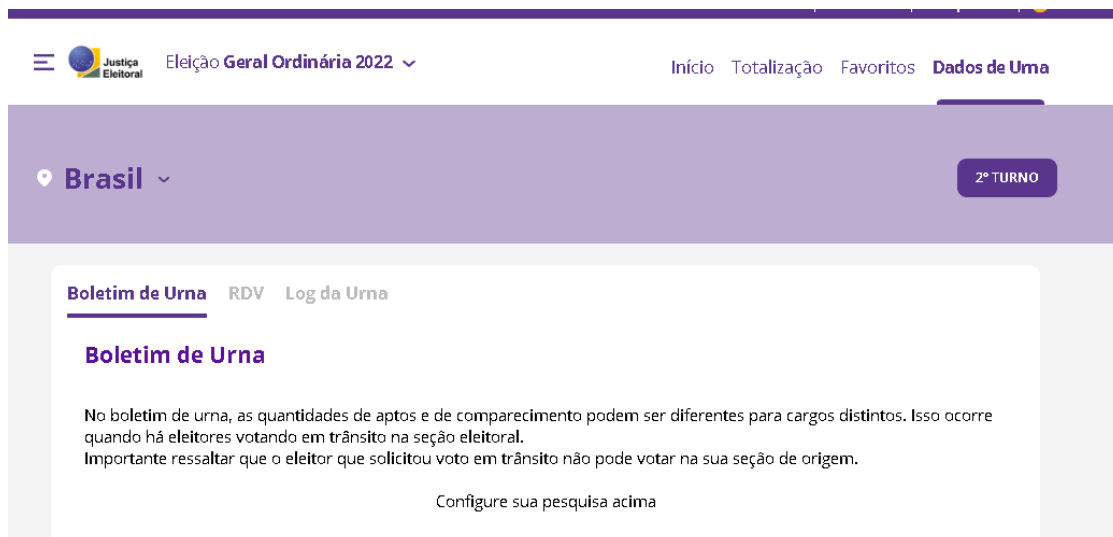
FIGURA 6 - Página Favoritos



Fonte: TSE (2023)

A última aba é “Dados de Urna”, reproduzido na figura , onde são apresentados os relatórios de urna, RDV, e o Log de Urna em formato digital. Além dos problemas indicados em outras páginas como cor, falta de clareza das informações, nesta, os dados dependem de configuração prévia para serem apresentados, porém, não existe indicativo fora a frase “Configure sua pesquisa acima”, sendo que não há espaço claro para fazer esta configuração.

FIGURA 7 - Página Dados de Urna



Fonte: TSE (2023)

Além dos problemas citados, outro é o de detalhamento dos dados, não é apresentado nenhum indicador como “seção de voto em trânsito” nas urnas, isso faz com que exista situações em que os votos numa seção eleitoral sejam muito acima do que o boletim tem como limite, fazendo com que fake news como “fraude nas urnas” sejam espalhadas, dando brecha para que o processo eleitoral seja alvo de diversas críticas e acusações por erro no produto informacional.

A forma que o produto/serviço informacional apresentado deixa a desejar, mesmo o produto tendo grande utilidade para a sociedade como fonte de informação segura e confiável, sendo que, a falta de jogo de cores, melhor organização das informações e a falta de clareza referente a detalhes da dinâmica da eleição acabam por gerar uma espécie de telefone sem fio, onde quem utiliza o produto informacional entende de maneira equivocada e retransmite o erro, virando uma bola de neve de informações incorretas.

4.2 IMPACTO SOCIAL DOS DADOS

Dados e informações tem impacto em nosso dia a dia de diversas maneiras, desde a quantidade de ovos a ser utilizado em uma receita, ou o melhor trajeto para ir a faculdade, quanto a regra matemática para fazer uma prova. Porém, algo que chama a atenção é a interpretação da informação, ou seja, para cada pessoa, o modo com o qual é captado a informação, acaba por gerar diferentes conclusões e pensamentos.

Portanto, podemos afirmar que existe uma guerra entre a verdade e a mentira na própria informação e sua forma de propagação, onde

narrativas fraudulentas exploram divergências políticas em torno de temas, personagens e linguagens, enviesando fatos e realidades de contextos regionais, nacionais e globais. A polarização política tem ajudado a criar um clima favorável para a propagação e validação de informações falsas (DOURADO, 2020).

No contexto das eleições ordinárias no Brasil, é criado por si um fantasma devido à própria forma de disponibilização de dados. Segundo o dicionário Aurélio, um fantasma é uma “imagem fantasiosa, ilusória, que infunde terror; produto da imaginação.” (AURÉLIO, 2023), e, de certa forma, é exatamente o resultado de um produto informacional não claro ao passar sua mensagem. Um fantasma de dados seria o produto da imaginação de uma pessoa que não compreendeu corretamente a mensagem do produto informacional de forma clara, e em alguns casos propagando o uso incorreto dos dados/informações apresentados, fazendo com que mais e mais pessoas acreditem nessa figura medonha. Em outras palavras, o fantasma de dados acaba por se tornar numa fake news, causada muitas vezes pela fértil imaginação de uma pessoa junto da falta de clareza existente num serviço/produto informacional, onde este produto contém falhas que permite tal enviesamento.

Segundo a prefeitura de Curitiba, em 2018, ao analisar 57 fake news de um total de 346, foi identificado mais de 4 milhões de compartilhamentos, somente na reta final das eleições ordinárias de 2018. O problema se agrava ainda mais quando o próprio fantasma é criado pelo próprio TSE, como foi o caso

de uma fake news criada com os dados apresentados do próprio aplicativo do órgão, conforme o Estadão declara

Para fazer o vídeo, o homem utiliza o aplicativo de Resultados da Justiça Eleitoral. Ele mostra os dados referentes à seção 100 da 276ª zona, de Uberaba. Para o cargo de presidente, a urna tinha 429 eleitores aptos e registrou 340 votos (172 para Lula e 127 para Bolsonaro). Nos demais cargos, a urna contava com 348 aptos dos quais 263 compareceram. Essa diferença ocorre porque a seção eleitoral foi utilizada também para o voto em trânsito para as pessoas que assim solicitaram dentro do prazo previsto em lei. (ESTADÃO, 2022)

Esse é um exemplo claro de falha no serviço/produto informacional, que impacta de maneira negativa no processo eleitoral, fazendo com que ocorra falta de credibilidade tanto no processo quanto no órgão público.

O impacto social observado seria o impacto das próprias fake news, onde, nas eleições, essas criações da imaginação humana podem até mesmo mudar o resultado de uma eleição. Porém, existe outro impacto, nos seres humanos, já que

o fenômeno das fake news está ligado ao grande problema filosófico da percepção da realidade. Da mesma forma que os prisioneiros da caverna do Platão confundem as sombras de dois objetos como seus próprios objetos, o homem contemporâneo parece confundir os acontecimentos projetados nas telas com os acontecimentos reais. Esse problema de raízes filosóficas pode representar uma ameaça política porque as democracias contemporâneas se baseiam na existência de uma cidadania livre e bem vivida, capaz de tomar decisões complexas em contextos políticos polarizados. (ALFONSO et al, 2019)

Portanto, corremos o risco de termos diversas pessoas alienadas, e com o senso de razão alterada, por um fantasma de dados, que, apesar de um problema que pode passar despercebido, acaba por impactar a vida de milhões. Tendo a noção desse impacto, erros e acertos da dashboard existente, podemos começar o processo de melhoria, e proposta da releitura do Dashboard das Eleições Ordinárias de 2022 do Estado do Paraná.

5. RELEITURA DO DASHBOARD DAS ELEIÇÕES ORDINÁRIAS DE 2022 - ESTADO DO PARANÁ

Após ter a avaliação do Dashboard, incluindo suas características, pontos fortes e pontos fracos, iniciamos o processo de desenvolvimento de uma releitura, focada no Estado do Paraná, apresentando melhorias nos processos e pontos fracos do produto/serviço informacional. Para melhor organização o tópico será separado em: processo de ETL, montagem da Dashboard.

Temos alguns pré requisitos para a criação da releitura:

- Python instalado no computador (preferencialmente Anaconda pois contém a maioria das bibliotecas necessárias)
- PostgreSQL instalado.
- PowerBI instalado.
- Software de manipulação de arquivos CSV.
- Banco de dados com o nome “eleicao_2022”

Arquivos referentes a parte de programação do processo se encontram disponíveis no GitHub⁴, que tem instruções mais detalhadas referente ao processo, explicando pontos que podem ser alterados para a adaptação do Dashboard para outros estados, por exemplo.

5.1 PROCESSO DE ETL

Para o início da releitura foi preciso realizar a coleta dos dados referentes à eleição ordinária de 2022, esses dados se encontram no site oficial do TSE (dados abertos), no formato de CSV. Outra opção seria utilizar da API para

⁴ Link do repositório do código: https://github.com/FernandesMarques/TCC_Dashboard_Eleicao

realizar essa coleta, mas, com a finalidade de permitir a replicação do trabalho⁵ mesmo após anos, foi optado por utilizar os arquivos disponibilizados.

São necessários dois arquivos do TSE, um referente a votação para presidente, contendo dados do primeiro e segundo turno, e, um arquivo com os dados referente às eleições do estado do Paraná (cargos de governador, deputado e senador). Também foi necessário a utilização de dois arquivos de dados em csv, um referente às unidades de voto em trânsito, e outro referente aos candidatos à presidência e seus respectivos partidos. Para replicação do processo para outros estados, deve-se utilizar da tabela referente a voto em trânsito no mesmo formato, apenas alterando os dados dos locais de votação do voto em trânsito. Após ter todos os arquivos em mãos, iremos utilizar para o processo de extração dos dados a linguagem de programação python e suas bibliotecas: pandas; io. psycopg2; sqlalchemy.

Para a execução do código, foi utilizado o jupyter notebook para melhor documentação, deste modo temos:

1. Coleta de dados e exportação dataframe –representado na figura 8

⁵ Os dados da API mudam conforme os anos, não contendo o histórico completo com a finalidade de otimizar a mesma em épocas de eleição.

FIGURA 8 - Script python - carregamento de dados

```
In [8]: import pandas as pd

df = pd.read_csv('votacao_secao_2022_PR.csv', encoding='latin1', sep=';')
df = df.drop(columns=['DT_GERACAO', 'HH_GERACAO', 'DT_ELEICAO', 'DS_ELEICAO'])
df.rename(columns={'ANO_ELEICAO': 'ano_eleicao', 'CD_TIPO_ELEICAO': 'cd_tipo_eleicao', 'NM_TIPO_ELEICAO': 'nm_tipo_eleicao'})
print(df)
```

	ano_eleicao	cd_tipo_eleicao	nm_tipo_eleicao	nr_turno	\
0	2022	2	ELEIÇÃO ORDINÁRIA	1	
1	2022	2	ELEIÇÃO ORDINÁRIA	1	
2	2022	2	ELEIÇÃO ORDINÁRIA	1	
3	2022	2	ELEIÇÃO ORDINÁRIA	1	
4	2022	2	ELEIÇÃO ORDINÁRIA	1	
...	
3397513	2022	2	ELEIÇÃO ORDINÁRIA	1	
3397514	2022	2	ELEIÇÃO ORDINÁRIA	1	
3397515	2022	2	ELEIÇÃO ORDINÁRIA	1	
3397516	2022	2	ELEIÇÃO ORDINÁRIA	1	
3397517	2022	2	ELEIÇÃO ORDINÁRIA	1	

	cd_eleicao	tp_abrangencia	sg_uf	sg_ue	nm_ue	cd_municipio	...	\
0	546	E	PR	PR	PARANÁ	75833	...	
1	546	E	PR	PR	PARANÁ	74500	...	
2	546	E	PR	PR	PARANÁ	78735	...	
3	546	E	PR	PR	PARANÁ	75353	...	
4	546	E	PR	PR	PARANÁ	76090	...	

Fonte: Elaborado pelo autor (2023)

É feito o carregamento do arquivo CSV através da linguagem python, onde é criado um dataframe com esses dados. Após é feito o tratamento das colunas, descartando colunas que não serão úteis e padronizando a nomenclatura das colunas úteis.

2. Execução do processo de carregamento dos dados - representado na figura 9

FIGURA 9 - Script python - criação de engine e integração com o banco de dados

```
In [9]: from sqlalchemy import create_engine
import psycopg2
import io
engine = create_engine(
    'postgresql+psycopg2://postgres:30265351@localhost:5432/eleicao_2022')

df.head(0).to_sql('eleicao_ordinaria_2022', engine, if_exists='replace', index=False)

conn = engine.raw_connection()
cur = conn.cursor()
output = io.StringIO()
df.to_csv(output, sep='\t', header=False, index=False)
output.seek(0)
contents = output.getvalue()
cur.copy_from(output, 'eleicao_ordinaria_2022', null='')

conn.commit()
cur.close()
conn.close()
```

Fonte: Elaborado pelo autor (2023)

Depois desse tratamento, criamos um “motor” (engine) para pegar os dados do servidor do banco de dados, e, conseguir carregar o banco com os dados do dataframe. Alguns parâmetros específicos são definidos neste processo, onde o principal é o nome da tabela e a propriedade de replace ativa, cuja função é substituir a tabela existente pela nova com os dados do data frame caso essa mesma tabela já exista.

Ao executar o código iremos obter o seguinte resultado, representado pela figura 10, no banco de dados:

FIGURA 10 - Banco de dados - Visão pós integração

ANO_ELEICAO	CO_TIPO_ELEICAO	NM_TIPO_ELEICAO	NRL_TURNO	CO_ELEICAO	TP_ABANGENCIA	SG_LIF	SG_LIE	NM_LIE	CD_MUNICIPIO	NM_MUNICIPIO
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	75953	GUARAPUAVA
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	74500	MARIPÁ
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	78735	SÃO JOSÉ DO CAUÁ
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	75353	CURITIBA
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	76090	IRETAMA
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	78972	SÃO TOMÉ
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	77810	FORTI AMAZONAS
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	79359	UMUARAMA
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	76678	LONDRIA
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	75353	CURITIBA
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	77011	MEDIANERA
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	75310	CRUZEIRO DO SUL
2022	2	ELEIÇÃO ORDINÁRIA	1	546	E	PR	PR	PARANÁ	78859	SÃO JOSÉ DOS PINHAIS

Fonte: Elaborado pelo autor (2023)

Iremos então realizar um tratamento nos dados, pois, temos os dados referente a votação do Paraná e presidente do Brasil, mas, com algumas “sujeiras”, entre elas a falta de indicação do voto em trânsito, e a falta de indicação do partido dos candidatos à presidência. O tratamento será feito através da linguagem SQL, no próprio administrador de banco de dados, utilizando as duas tabelas auxiliares de voto em trânsito e de candidatos à presidência. Iremos fazer a criação de duas views, conforme ilustrado na figura 11, com colunas de dados complementares, com a marcação do voto em trânsito

com a coluna “vt”, junto das colunas “sigla_partido”, “sigla_nome”, “nome_partido”.

FIGURA 11 - Banco de dados - Visão pós criação de view

vt text	nome_partido text	sigla_partido text	sigla_nome text	cidade_estado text	unidade_federativa text
[nul]	Partido dos Trabalhadores	PT	PT-Partido dos Trabalhadores	SIQUEIRA CAMPOS-Paraná	Paraná
[nul]	Partido dos Trabalhadores	PT	PT-Partido dos Trabalhadores	MARINGÁ-Paraná	Paraná
[nul]	Movimnetno Democrático Brasileiro	MDB	MDB-Movimnetno Democrático Brasileiro	LONDRINA-Paraná	Paraná
[nul]	Movimnetno Democrático Brasileiro	MDB	MDB-Movimnetno Democrático Brasileiro	PARANAGUÁ-Paraná	Paraná
[nul]	Movimnetno Democrático Brasileiro	MDB	MDB-Movimnetno Democrático Brasileiro	MEDIANEIRA-Paraná	Paraná
[nul]	Partido Liberal	PL	PL-Partido Liberal	ARAUJÁRIA-Paraná	Paraná
[nul]	Partido Liberal	PL	PL-Partido Liberal	MARINGÁ-Paraná	Paraná
[nul]	Movimnetno Democrático Brasileiro	MDB	MDB-Movimnetno Democrático Brasileiro	PARANAGUÁ-Paraná	Paraná
[nul]	Movimnetno Democrático Brasileiro	MDB	MDB-Movimnetno Democrático Brasileiro	PONTAL DO PARANÁ-Paraná	Paraná
[nul]	Movimnetno Democrático Brasileiro	MDB	MDB-Movimnetno Democrático Brasileiro	GUAPIRAMA-Paraná	Paraná
[nul]	VOTO BRANCO	VOTO BRANCO	VOTO BRANCO	PINHAIAS-Paraná	Paraná

Fonte: Elaborado pelo autor (2023)

Com os dados tratados e disponibilizados nessas views, iremos para a parte de visualização, no Power BI, carregando esses dados através do software, que disponibiliza a opção de obter dados, conforme indica a figura 12.

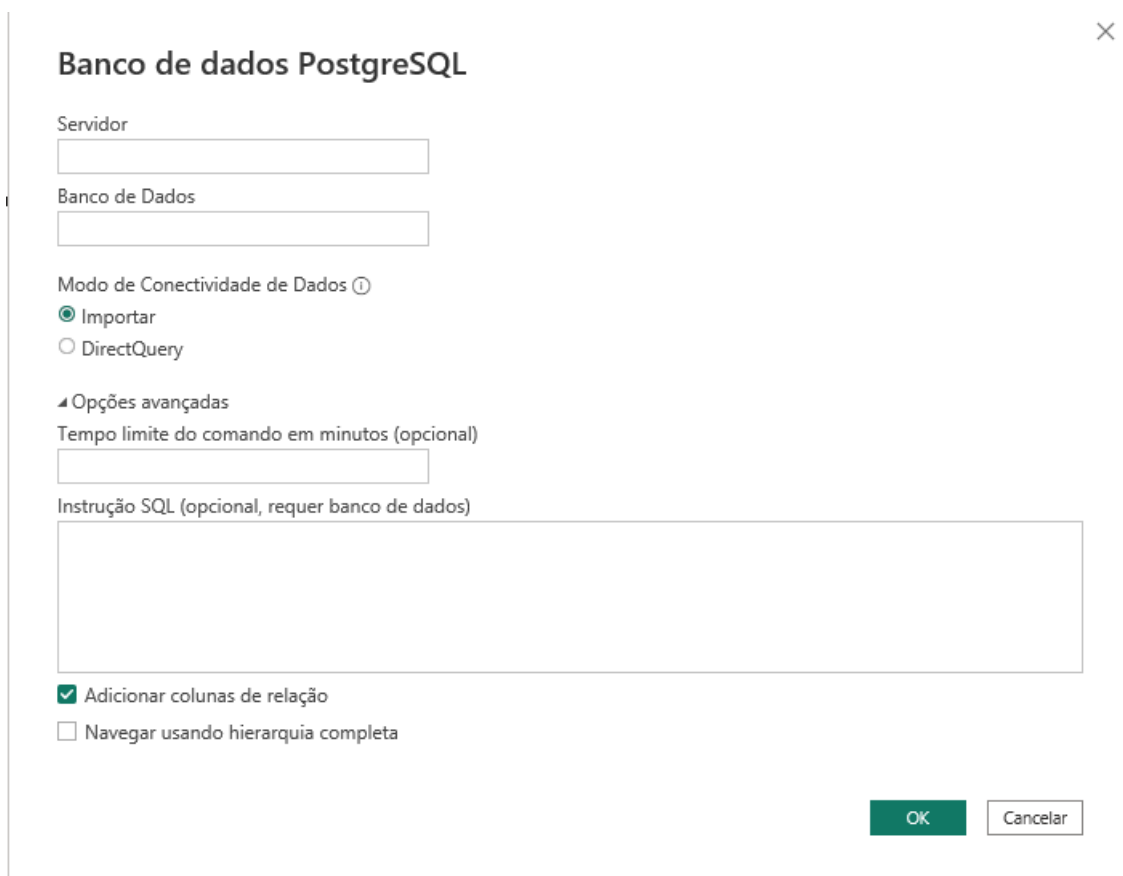
FIGURA 12 - Interface Power BI



Fonte: Elaborado pelo autor (2023)

Iremos obter os dados, procurando pelo banco de dados (PostgreSQL) e preenchendo o servidor, e o banco de dados, junto do código de seleção da tabela que desejamos, as duas views criadas anteriormente, indicado na figura 13 abaixo.

FIGURA 13 - Interface Power BI - conexão com o banco de dados



The image shows a dialog box titled "Banco de dados PostgreSQL" with a close button (X) in the top right corner. The dialog contains the following fields and options:

- Servidor:** A text input field.
- Banco de Dados:** A text input field.
- Modo de Conectividade de Dados:** Two radio button options: "Importar" (selected) and "DirectQuery".
- Opções avançadas:** A section header with a downward arrow.
- Tempo limite do comando em minutos (opcional):** A text input field.
- Instrução SQL (opcional, requer banco de dados):** A large text area for entering SQL queries.
- Adicionar colunas de relação:** A checked checkbox.
- Navegar usando hierarquia completa:** An unchecked checkbox.
- Buttons:** "OK" and "Cancelar" buttons at the bottom right.

Fonte: Elaborado pelo autor (2023)

Então, teremos todos os dados necessários para a criação do Dashboard das Eleições Ordinárias de 2022 do Estado do Paraná, conforme figura 14 e 15.

FIGURA 14 - Interface Power BI - dados após a conexão – Estado

D_MUNICIPIO	NM_MUNICIPIO	NR_ZONA	NR_SECAO	CD_CARGO	DS_CARGO	NR_VC
75353	CURITIBA	177	460	7	DEPUTADO ESTADUAL	
75353	CURITIBA	177	99	7	DEPUTADO ESTADUAL	Linha 6.169 de 3.397.518
75353	CURITIBA	177	99	7	DEPUTADO ESTADUAL	
75353	CURITIBA	177	548	7	DEPUTADO ESTADUAL	
75353	CURITIBA	177	780	7	DEPUTADO ESTADUAL	
75353	CURITIBA	177	558	7	DEPUTADO ESTADUAL	

Fonte: Elaborado pelo autor (2023)

FIGURA 15 - Interface Power BI - dados após a conexão – Federação

ANO_ELEICAO	CD_TIPO_ELEICAO	NM_TIPO_ELEICAO	NR_TURNO	CD_ELEICAO	TP_ABRANGENCIA	SG_UF	SG_UF	NM_UF	CD_MUNICIPIO	NM_MUNICIPIO	NR_ZONA	NR_SECAO	CD_CARGO	DS_CARGO	NR_VOZES
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	252	7	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	253	2	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	257	14	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	253	5	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	347	3	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	257	3	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	260	17	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	405	5	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	347	354	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	405	15	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	6	14	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	280	7	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	417	28	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	260	19	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	249	7	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	253	28	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	252	10	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	348	2	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	389	32	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	326	6	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	254	7	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	405	17	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	348	4	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	389	4	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	251	11	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	349	578	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	376	7	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	328	2	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	248	15	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	405	10	1	PRESIDENTE		
2022	2	ELEIÇÃO ORDINÁRIA	7	544 F	SP	BR	BRASIL	7102	SÃO PAULO	411	6	1	DISPENSANTE		

Fonte: Elaborado pelo autor (2023)

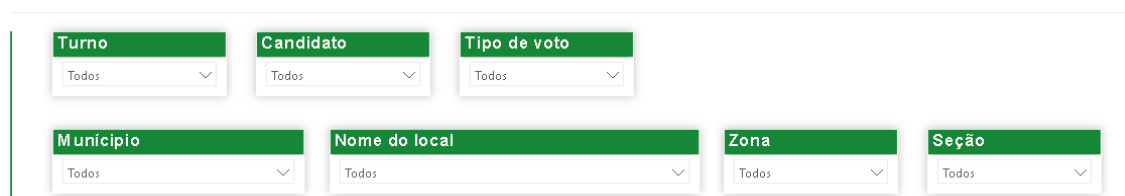
5.2 CRIAÇÃO DA DASHBOARD

Para a criação da Dashboard⁶, foi utilizado principalmente das cores símbolo do TSE, o verde, amarelo e azul. A principal cor escolhida foi o verde, que representa a maior parte da logo símbolo do órgão. A Dashboard foi dividida

⁶O dashboard criado pode ser acessado através do link: <https://app.powerbi.com/groups/me/reports/c4f7ebd2-8158-4876-bb66-ec4d61a5b7ea/ReportSection?experience=power-bi>

em duas, a primeira página referente aos dados da eleição presidencial, e a segunda referente a votação dos representantes estaduais. Na figura 16 são indicados os filtros que foram criados para permitir a interação do usuário com os gráficos.

FIGURA 16 - Proposta de nova dashboard - interface de filtros



Fonte: Elaborado pelo autor (2023)

Os gráficos a serem utilizados são: gráficos de barras, heatmap, funil, e, gráfico de colunas. A escolha deles se deve a simples noção de visualização, não exigindo conhecimento prévio, sendo intuitivos por assim dizer.

Algumas medidas (campos calculados) foram implementadas, eles permitem o funcionamento dinâmico do painel de modo com que o usuário possa navegar, e conforme utiliza dos filtros e outros recursos da ferramenta o painel atualiza os campos referente a porcentagem e total do tipo de voto de forma dinâmica e automática. Uma medida deve ser declarada com formulas matemáticas semelhantes à do software excel, como indicado abaixo na figura 17.

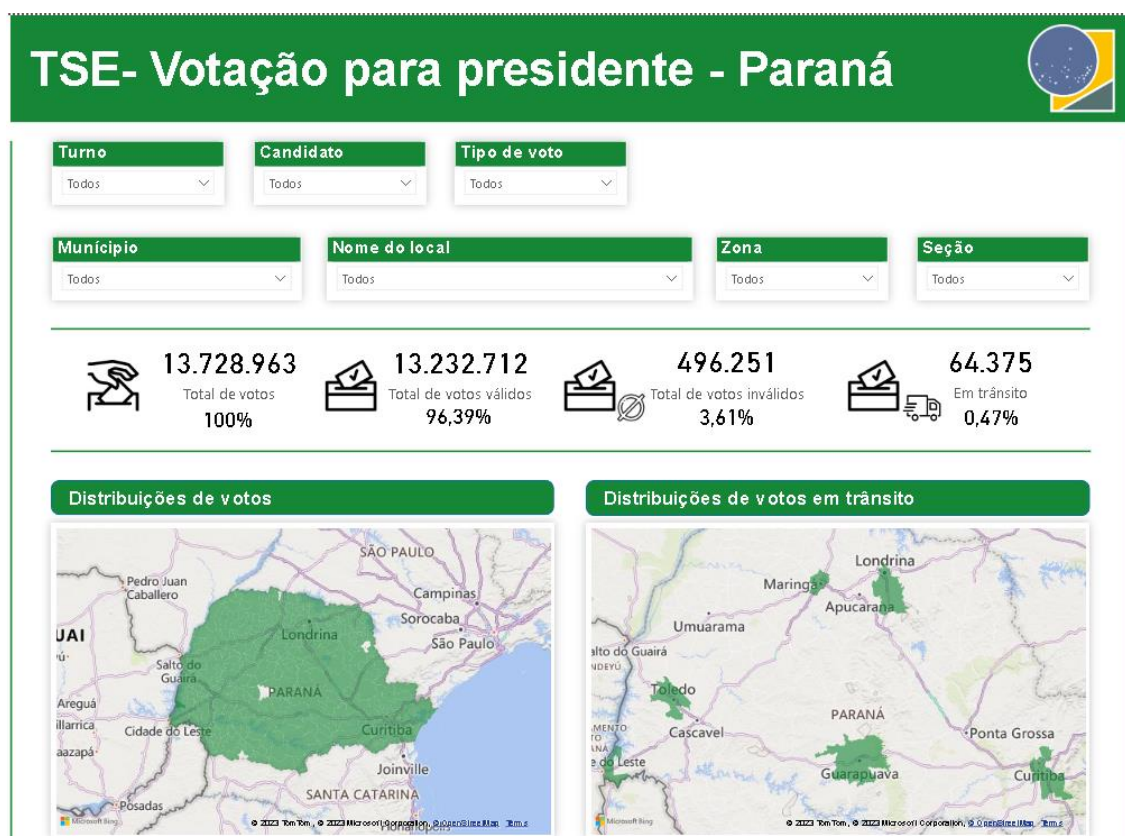
FIGURA 17 - Proposta de nova dashboard - Exemplo medida

```
1 total invalido pr = ((calculate(sum('public eleicao_ordinaria_2022'[qt_votos]),'public eleicao_ordinaria_2022'[nm_votavel] = "VOTO NULO") + calculate(sum('public eleicao_ordinaria_2022'[qt_votos]),'public eleicao_ordinaria_2022'[nm_votavel] = "VOTO BRANCO")))
```

Fonte: Elaborado pelo autor (2023)

Deste modo, temos a formação da página referente a eleição presidencial abaixo, figura 18:

FIGURA 18 - Proposta de nova dashboard - interface geral - Federal



Fonte: Elaborado pelo autor (2023)

A página referente a eleição de governadores segue a mesma linha de pensamento, conforme figura 19:

FIGURA 19 - Proposta de nova dashboard - interface geral - Estadual



Fonte: Elaborado pelo autor (2023)

É necessário lembrar que os dados abordados são específicos ao estado do Paraná devido a delimitação do tema, porém, o mesmo pode ser adaptado para ser replicado em outros estados, apenas alterando a delimitação e fontes feita na parte da ETL.

Com essa releitura, procura-se sanar problemas referente a cores, e falta de indicações de dados, principalmente do voto em trânsito. Foi seguido os princípios de Knaflic para a escolha da visualização, porém, tenta-se manter a identidade visual do Dashboard, através do conjunto de cores e logo do TSE.

5.3 ANÁLISE DO NOVO DASHBOARD

Ao realizar o primeiro contato com a proposta Dashboard, já é possível notar diferenças principalmente no quesito de cores, onde, a palheta foi alterada para ser majoritariamente o verde, como exemplifica a figura 20, com alguns detalhes em amarelo e azul contidos na logo do TSE, habilitando para alguns gráficos a possibilidade de utilizar mais de uma cor.

FIGURA 20 - Proposta de nova dashboard - Análise - Filtros



TSE- Votação para presidente - Paraná

Turno
Todos

Candidato
Todos

Tipo de voto
Todos

Município
Todos

Nome do local
Todos

Zona
Todos

Seção
Todos

Fonte: Elaborado pelo autor (2023)

Também foram inseridos filtros claros sobre os dados que segmentam a votação, dentre eles o turno, candidato, tipo de voto (Voto nominal ou em trânsito), município, nome do local, zona e seção eleitoral para a votação de presidente, enquanto para a parte de votação estadual tiveram esses filtros, com a adição de um para cargo, já que, temos a eleição de senadores, deputados federais e estaduais, governador.

Na figura 21 podemos ver ícones intuitivos foram introduzidos a proposta de Dashboard, com a finalidade de deixar os números mais intuitivos, permitindo a interpretação e associação dos números até por parte de usuários não alfabetizados ou familiarizados com a tecnologia.

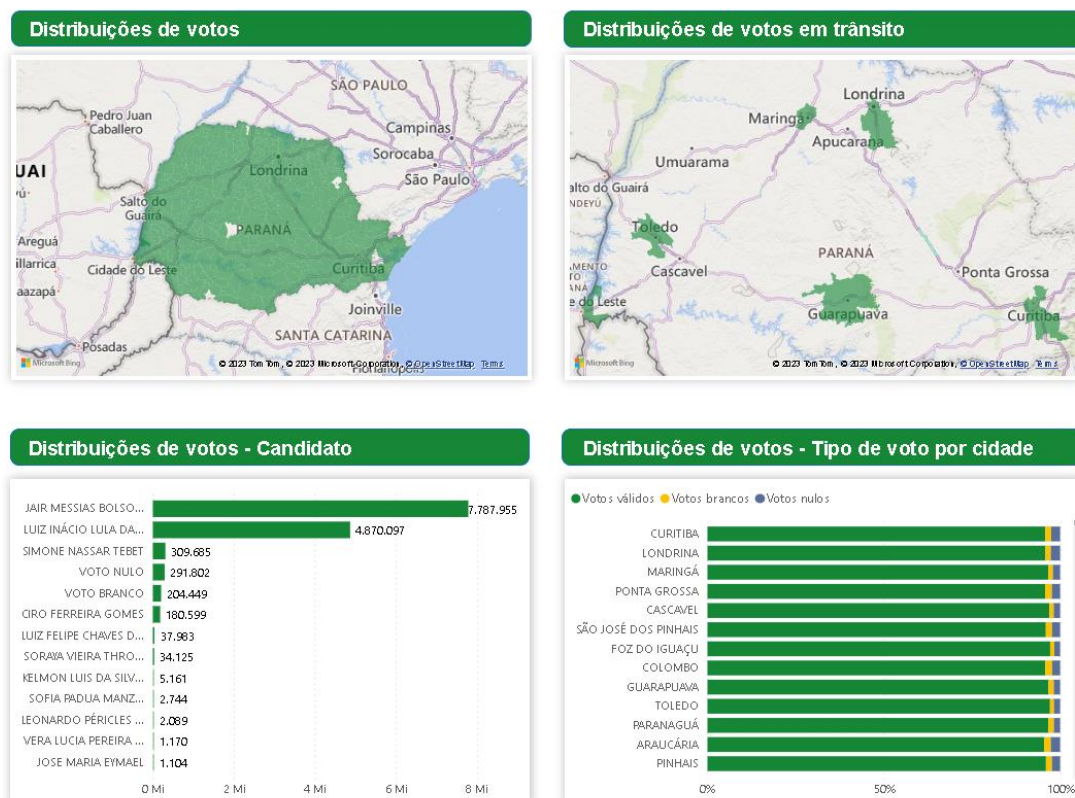
FIGURA 21 - Proposta de nova dashboard - Análise - Ícones



Fonte: Elaborado pelo autor (2023)

No quesito gráficos, os anteriores de rosca, conforme literatura pertinente sugere (Knaflic, 2019), foram removidos, dando espaço aos gráficos da figura 22: heatmap, gráfico de barras empilhadas, e gráfico de barras 100% empilhadas.

FIGURA 22 - Proposta de nova dashboard - Análise – Gráficos



Fonte: Elaborado pelo autor (2023)

Com a utilização dos filtros junto das informações contidas nos gráficos, é possível sanar as dúvidas que o Dashboard antigo esclareceu, junto de alguns aspectos que o mesmo não deixa claro, como a situação de voto em trânsito.

Devido ao visual mais compacto do Dashboard focando somente nos principais números, a parte de exploração de dados fica a critério do usuário, onde o mesmo pode ver com exatidão os números e valores em todos os gráficos ao colocar o cursor do mouse em cima do preenchimento dos gráficos.

Apesar de parecer que os dados estão sendo ocultados, os mesmos são referentes a tópicos específicos, onde, a proposta de produto não consegue deixar tão explícito devido ao foco nos principais números, tornando os números específicos um adicional, que pode ser encontrado de forma mais intuitiva que a versão disponibilizada pelo TSE.

6. CONSIDERAÇÕES FINAIS

O crescimento da produção de dados é evidente, cada vez mais potencializado pelo forte acesso da população a mídias digitais, e se espera que além deste aumento ocorra a expansão do consumo de dados pelas diferentes camadas da sociedade brasileira. Para auxiliar no consumo destes dados, e assegurar a geração de informação e conhecimento de qualidade, a gestão da informação torna-se um campo extremamente necessário, para atuar nos diversos aspectos informacionais, desde a coleta e tratamento, a disponibilização e criação de produtos informacionais.

Neste contexto, o consumo de dados deve ser feito e fontes confiáveis e claras, como o existente Dashboard das Eleições Ordinárias de 2022 falha em atender, portanto, existe a necessidade da entrega do produto/serviço informacional gerado neste trabalho. Neste contexto, a repaginação apresentada sana os principais problemas de fontes, enquanto mantém um visual mais adequado ao governo, com um jogo de cores assertivo com a imagem do órgão público.

O objetivo geral deste trabalho era oferecer uma versão melhorada do produto/serviço informacional existente, melhorando os aspectos tanto na parte de coleta e tratamento dos dados quanto da visualização da informação. Para alcançar este objetivo, foram estabelecidos os objetivos específicos, que de forma geral podem ser definidos em três grupos: Conceituar um serviço/produto informacional e definir visualização de dados e suas técnicas; analisar o dashboard existente, junto do seu impacto social; criar um processo de ETL e uma proposta alternativa de Dashboard das Eleições Ordinárias de 2022 do estado do Paraná.

Perguntas que nortearam o trabalho foram respondidas por completo, sendo elas “Como os dados estão apresentados e visualizados na *Dashboard* da Eleição Geral Ordinária 2022 no Estado do Paraná? Quais características, natureza do produto informacional e o impacto da visualização dos dados e informações apresentadas? Em que medida uma nova *Dashboard*, enquanto um produto/serviço informacional, poderá otimizar a análise, compreensão e

interpretação dos dados eleitorais?”. No decorrer da pesquisa existe uma análise para entender o dashboard e como os dados são apresentados no formato atual, e através do referencial teórico temos um melhor entendimento da natureza do produto e de seu impacto, enquanto é apontado na nova proposta de *Dashboard*, no momento de análise, o impacto da otimização do produto/serviço informacional.

Limitações foram encontradas no decorrer do trabalho: dados com formatação indevida; dificuldade encontrada nas cores do dashboard; capacidade de processamento do computador. Os dados disponibilizados pelo órgão apresentavam diversos erros de preenchimento, não indicando corretamente nomes de locais de votação, e principalmente sua localização. Outro problema foi relacionado a cores, pois, nas eleições de 2022 tivemos o problema de um candidato se apossar das cores do órgão TSE, o azul verde e amarelo. Para contornar a situação das cores foi escolhida uma palheta focada no verde, com tonalidade semelhante ao logo, mas, que se utilizado de forma majoritária não remete ao candidato, mas sim ao órgão público. Referente a limitação do computador, se deve ao grande fluxo de dados, já que a base consta com centenas de milhões de votos, onde, essa situação foi contornada com o estudo em cima dos dados apenas do Paraná, mas, em caso de ampliação do estudo será necessário uso de computadores com maior capacidade de processamento.

Todos os objetivos propostos foram alcançados, porém, durante a execução do trabalho, alguns pontos de melhorias possíveis foram observados, dentre eles o de automatização da coleta de dados diretamente da API do TSE, qual mesmo com suas falhas e indisponibilidades após as eleições, pode ser utilizado para criação de um dashboard que acompanharia em tempo real o processo eleitoral de forma precisa. Além disso, outro aspecto a ser melhorado seria a enriquecimento de dados e a disponibilização destes para estudos mais aprofundados de outras áreas, onde informações com renda média, índice de desenvolvimento humano e outras métricas podem ser utilizadas para traçar perfis e serem usados para entender cada vez melhor a população votante brasileira.

7. REFERÊNCIAS

AKKARI, Alessandra Cristina Santos; CARPI, José Renato. **Desenvolvimento de Produto**. Londrina: Editora e Distribuidora Educacional S.A., 2019. 160 p.

ANGELONI, Maria Terezinha. Elementos intervenientes na tomada de decisão. **Ciência da Informação**, [S.L.], v. 32, n. 1, p. 17-22, abr. 2003. IBICT. <http://dx.doi.org/10.1590/s0100-19652003000100002>. Disponível em: <https://www.scielo.br/j/ci/a/3RVhpdpmmsgkwCxtCC6sXkt/?lang=pt>. Acesso em: 14 mar. 2023.

ASIETA, Héctor Andrés Melgar; PACHECO, Roberto Carlos dos Santos. O Paradigma Qualitativo na Visualização da Informação: Uma Revisão Sistemática. *In*: SASIETA, Héctor Andrés Melgar; PACHECO, Roberto Carlos dos Santos. **O Paradigma Qualitativo na Visualização da Informação: Uma Revisão Sistemática**. [S. l.], 1 out. 2001. Disponível em: Acesso em: 25 nov. 2022.

BORGES, Mônica Erichsen Nassif; CARVALHO, Natália Guiné de Mello. Produtos e serviços de informação para negócios no Brasil: características. **Ciência da Informação**, Brasília, v. 27, n. 1, p. 76-81, jan./abr., 1998. Disponível em: <https://www.scielo.br/j/ci/a/5cfwLhngnxXbkDZKXGJrnWm/?lang=pt>. Acesso em: 27 nov. 2022.

CONWAY, Drew. **Machine Learning for Hackers**. [S. l.]: O'Reilly Media, Inc., 25/02/2012. 303 p. Disponível em: https://www.iiis.org/CDs2011/CD2011CSC/CISCI_2011/PapersPdf/CA294BQ.pdf. Acesso em: 25 nov. 2022.

DAVENPORT, T. H. *Ecologia da informação: porque só a tecnologia não basta para o sucesso na era da informação*. Tradução de Bernadette Siqueira Abrão. São Paulo: Futura, 1998.

DAVENPORT, T.H.; PRUSAK, L. *Conhecimento empresarial: como as organizações gerenciam o seu capital intelectual*. Rio de Janeiro: Campus, 1998 – 4ª edição

DINIZ, V. Como conseguir dados governamentais abertos. In: CONGRESSO CONSAD DE GESTÃO PÚBLICA, III, Brasília, 2010

DIOUF, Papa Senghane; BOLY, Aliou; NDIAYE, Samba. Variety of data in the ETL processes in the cloud: state of the art. **2018 Ieee International Conference On Innovative Research And Development (Icird)**, [S.L.], v. 0, n. 0, p. 0-0, 11 maio 2018. IEEE. <http://dx.doi.org/10.1109/icird.2018.8376308>. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8376308>. Acesso em: 05 dez. 2022.

ESTEVAM, Evandro Cesar. **A Utilização da Visualização da Informação na Gestão de Organizações de Saúde**. 2019. 109 f. TCC (Graduação) - Faculdade de Ciência da Computação, universidade estadual paulista, São José do Rio Preto, 2019. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/181262/estevam_ec_me_sjr_p.pdf?sequence=3&isAllowed=y. Acesso em: 12 dez. 2022.

FERREIRA, João; MIRANDA, Miguel; ABELHA, António; MACHADO, José. O Processo ETL em Sistemas Data Warehouse. **INForum 2010 - II Simpósio de Informática**, [S. I.], ano 2010, p. 757-765, 9 set. 2010. Disponível em: https://www.researchgate.net/profile/Jose-Machado-23/publication/265195317_O_Processo_ETL_em_Sistemas_Data_Warehouse/inks/5580380a08aea3d7096e442e/O-Processo-ETL-em-Sistemas-Data-Warehouse.pdf. Acesso em: 21 nov. 2022.

FREITAS, Carla Maria Dal Sasso; CHUBACHI, Olinda Mioka; LUZZARDI, Paulo Roberto Gomes; CAVA, Ricardo Andrade. Introdução à visualização de informações. **RITA**, [S. I.], ano 2011, v. 8, n. 2, p. 143-158, 1 out. 2001. Disponível em: <https://lume.ufrgs.br/handle/10183/19398>. Acesso em: 25 nov. 2022.

FREITAS, J. A. de C.; BALANIUK, R.; DA SILVA, A. P. B.; DA SILVEIRA, V. S. **O ecossistema de dados abertos do governo federal: composição e desafios**. *Ciência da Informação*, [S. I.], v. 47, n. 2, 2018. DOI: 10.18225/ci.inf.v47i2.3952. Disponível em: <http://revista.ibict.br/ciinf/article/view/3952>.

HOFFMANN, Yohana Taise; ALVAREZ, Edgar Bisset; MARTÍ-LAHERA, Yohannis. Análise textual com IRaMuTeQ de pesquisas recentes em História da educação matemática no Brasil: um exemplo de humanidades digitais. **Investigación Bibliotecológica**: archivonomía, bibliotecología e información, [S.L.], v. 34, n. 84, p. 103, 2 jul. 2020. Universidad Nacional Autónoma de México. <http://dx.doi.org/10.22201/iibi.24488321xe.2020.84.58097>.

JANES, Andrea; SILLITTI, Alberto; SUCCI, Giancarlo. Effective Dashboard Design. **A GQM+STRATEGIES APPROACH**, [S. l.], p. 17-24, 31 ago. 2016. Disponível em: https://www.researchgate.net/publication/286996830_Effective_Dashboard_design. Acesso em: 15 fev. 2023.

KNAFLIC, Cole Nussbaumer. **Storytelling com dados**: Um guia sobre visualização de dados para profissionais de negócios. 2. ed. atual. [S. l.]: Alta Books, 2019. 256 p.

MICROSOFT. O que é Power BI?. In: **Microsoft**. [S. l.], 22 mar. 2023. Disponível em: <https://learn.microsoft.com/pt-br/power-bi/fundamentals/power-bi-overview>. Acesso em: 15 maio 2023.

MOURA, Maria Aparecida. Ciência da Informação e humanidades digitais: mediações, agência e compartilhamento de saberes. **Perspectivas em Ciência da Informação**, [S.L.], v. 24, n. , p. 57-69, mar. 2019. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/1981-5344/3893>. Disponível em: <https://www.scielo.br/j/pci/a/GzKTfV7D4s7RLMM6qgpHkJF/>. Acesso em: 02 mar. 2023.

PORTO, Mauro; THUSSU, Daya. **Media Power and Democratization in Brazil**. New York: Routledge, 2012. 240 p. Disponível em: <https://www.taylorfrancis.com/books/mono/10.4324/9780203119631/media-power-democratization-brazil-mauro-porto-daya-thussu>. Acesso em: 03 fev. 2023.

SADEK, Maria Tereza. Eleições 1988: retórica ou rumo pluripartidário. **Lua Nova**: Revista de Cultura e Política, [S.L.], n. 17, p. 111-132, jun. 1989. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0102-64451989000200007>. Disponível em: <https://www.scielo.br/j/ln/a/8WywhGQb7ytP3kQjnxzJ6pK/?lang=pt>. Acesso em: 25 nov. 2022.

SANTOS, Isabel Lima dos. Elaboração de produtos e serviços de informação: conceitos e etapas chave. **ConCI: Convergências em Ciência da Informação**, v. 5, n. dossiê, p. 1-19, 1 maio 2022. Disponível em: <https://seer.ufs.br/index.php/conci/article/view/16835>. Acesso em: 27 nov. 2022.

SILVA, Francisco Sávio da; COSTA, Marcílio Herculano da; DANTAS, Jefferson Fernandes; LLARENA, Rosilene Agapito da Silva. Produtos e serviços informacionais nos arquivos do Mercosul: análise das páginas web dos arquivos nacionais. **Archeion Online**, João Pessoa, v. 6, n. 1, p. 23-47, jul. 2018. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/archeion/article/view/36825>. Acesso em: 15 27 nov. 2022.

SILVA, D. B. **Transparência na esfera pública interconectada**. 2010. (Dissertação de Mestrado), Faculdade Cásper Líbero, São Paulo.

SILVA, Gustavo Castro. O banco de dados bibliográfico João Guimarães Rosa. **Revista Brasileira de Literatura Comparada**, [S.L.], v. 24, n. 46, p. 215-223, abr. 2022. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/2596-304x20222446gcs>. Disponível em: <https://www.scielo.br/j/rblc/a/Gdrwzgp9r6fZB9sypsxfNCC/?lang=pt>. Acesso em: 03 dez. 2022

TUKEY, John. **Exploratory Data Analysis**. [S. l.]: O'Reilly Media, Inc., 11/01/1977. 688 p. Disponível em: https://www.iiis.org/CDs2011/CD2011CSC/CISCI_2011/PapersPdf/CA294BQ.pdf. Acesso em: 25 nov. 2022.

WICKHAM, Hadley; GROLEMUND, Garrett. **R for Data Science**. [S. l.]: O'Reilly Media, Inc., 01/12/2016. 520 p. Disponível em: https://www.iiis.org/CDs2011/CD2011CSC/CISCI_2011/PapersPdf/CA294BQ.pdf. Acesso em: 25 nov. 2022.

WINAND, Markus. SQL Performance Explained Everything Developers Need to Know about SQL Performance. [S. l.]: Marcus Winand, 01/02/2012. 204 p.