

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Sistema Web para Visualização e Análise de Dados Abertos Governamentais sobre Eleições no Brasil

Ana Carolina Fidelis Gonçalves

JUIZ DE FORA
MARÇO, 2021

Sistema Web para Visualização e Análise de Dados Abertos Governamentais sobre Eleições no Brasil

ANA CAROLINA FIDELIS GONÇALVES

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Sistemas de Informação
Orientador: Fabrício Martins Mendonça

JUIZ DE FORA
MARÇO, 2021

SISTEMA WEB PARA VISUALIZAÇÃO E ANÁLISE DE DADOS ABERTOS GOVERNAMENTAIS SOBRE ELEIÇÕES NO BRASIL

Ana Carolina Fidelis Gonçalves

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS
EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTE-
GRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE
BACHAREL EM SISTEMAS DE INFORMAÇÃO.

Aprovada por:

Fabício Martins Mendonça
Doutor em Ciências da Informação

Luciana Conceição Dias Campos
Doutora em Engenharia Elétrica

Mário Antônio Ribeiro Dantas
Doutor em Ciência da Computação

JUIZ DE FORA
15 DE MARÇO, 2021

Aos meus amigos e irmãos.

Aos pais, pelo apoio e sustento.

Resumo

A integração de dados abertos governamentais e seu uso como ferramenta de análise é um problema cuja complexidade aumenta muito com banco de dados de grandes volumes. Anualmente os órgãos públicos brasileiros geram um grande conjunto de informações que podem ser utilizadas para análise pelos cidadãos. Apesar disso, os formatos de distribuição são heterogêneos, não padronizados segundo normas de publicação (por exemplo, do consórcio W3C) e a descentralização dos dados dificulta sua utilização. Além disso, a maior parte da população não possui conhecimentos básicos em Informática para acessar os dados publicados.

Neste trabalho propomos uma solução de *Business Intelligence* (BI) com dados abertos governamentais na área eleitoral. O sistema web proposto é composto por um servidor, banco de dados, bibliotecas e ferramentas visuais de dados. A aplicação foi criada a partir da construção de um *Data Warehouse* (DW) extraído de dados governamentais públicos em formato CSV de repositórios de dados eleitorais do Tribunal Superior Eleitoral. A partir da construção do DW, foram aplicadas técnicas de *Data Analytics* com o objetivo de integrar os dados dos repositórios de dados eleitorais do Tribunal Superior Eleitoral (TSE) com o estudo de Espectro Político e, na sequência, gerar *Dashboards* como modelos visuais que facilitam a análise dessas informações. O resultado esperado com essa proposta é permitir aos cidadãos brasileiros terem maior capacidade de utilização das informações públicas na área eleitoral.

Palavras-chave: *Business Intelligence*, *Data Warehouse*, ETL, Dados Abertos Governamentais, Dados Eleitorais.

Abstract

The integration of government open data and its use as an analysis tool is a problem whose complexity increases greatly with large volume databases. Every year Brazilian public agencies generate a large set of information that can be used for analysis by citizens. Nevertheless, the distribution formats are heterogeneous, non-standardized according to publication norms (for example, of the W3C consortium) and the decentralization of the data makes its use difficult. In addition, most of the population does not have basic computer skills to access the published data.

In this work we propose a Business Intelligence (BI) solution with open governmental data in the electoral area. The proposed web system is composed of a server, database, libraries and visual data tools. The application was created from the construction of a Data Warehouse (DW) extracted from public government data in CSV format from electoral data repositories of the Tribunal Superior Eleitoral (TSE). After the construction of the DW, Data Analytics techniques were applied with the objective of integrating the data from the electoral data repositories of the TSE with the Political Spectrum study, and then generate Dashboards as visual models that facilitate the analysis of this information. The expected result with this proposal is to allow Brazilian citizens to have a greater capacity to use public information in the electoral area.

Keywords: Business Intelligence, Data Warehouse, ETL, Open Government Data, Electoral Data.

Agradecimentos

Primeiramente quero agradecer a Deus, pois me deu saúde e força para conseguir concluir o curso.

A minha querida mãe, Ana Maria Fidelis Gonçalves, mulher forte e trabalhadora, pelo seu apoio incondicional e sustento durante toda a trajetória. Dedico este projeto de pesquisa ao meu saudoso pai, Pedro Roseno José Gonçalves, pois enquanto vivo me ensinou que a fé alinhada com estudo e trabalho seria a chave para realização dos meus sonhos.

Ao meu irmão, Pedro Henrique Fidelis Gonçalves, pois juntos na universidade fomos fonte de apoio e amizade um para com o outro.

A todos os amigos e familiares que me apoiaram e entenderam os momentos que não pude estar presente. Em especial a minha querida mãe do coração, Luziana Gracina de Almeida, que foi apoio, força, aconchego desde o início desta caminhada.

Ao professor Fabrício Martins pela orientação, pela paciência e profissionalismo, sem o qual este trabalho não se concluiria.

Aos professores do Departamento de Ciência da Computação pelos seus ensinamentos e aos funcionários do curso, que durante esses anos contribuíram de algum modo para o meu enriquecimento pessoal e profissional.

Por último, agradeço aos profissionais que trabalhei durante esses anos, pois muito agregaram ao conhecimento profissional. Não posso esquecer dos grandes amigos que fiz na faculdade, pois com sua forma alegre ou algumas vezes mal-humorada me ajudaram durante todos esses anos e tornaram essa etapa mais divertida.

“Que todos os nossos esforços estejam sempre focados no desafio à impossibilidade. Todas as grandes conquistas humanas vieram daquilo que parecia impossível”.

Charles Chaplin

Conteúdo

Lista de Figuras	8
Lista de Tabelas	9
Lista de Abreviações	10
1 Introdução	11
1.1 Problema	12
1.2 Objetivos	13
1.3 Metodologia	13
1.4 Organização do Trabalho	15
2 Fundamentação Teórica	16
2.1 <i>Business Intelligence</i>	16
2.1.1 <i>Data Warehouse</i>	18
2.1.2 ETL	24
2.1.3 <i>Dashboards</i> para BI	26
2.2 Dados Abertos Governamentais	26
2.2.1 Dados Abertos no Brasil	28
2.3 Eleições no Brasil	30
2.4 Espectro Político	31
3 Trabalhos Relacionados	34
4 Requisitos de Negócios e Modelagem	37
4.1 Requisitos de Negócios	37
4.1.1 Regras de Negócio	38
4.1.2 Requisitos Funcionais	39
4.1.3 Requisitos Não Funcionais	39
4.2 Modelagem Dimensional	40
4.2.1 Dimensão Abrangência	40
4.2.2 Dimensão Ano	40
4.2.3 Dimensão Cargo	41
4.2.4 Dimensão Tipo de Eleição	41
4.2.5 Dimensão Turno	42
4.2.6 Dimensão Eleição	42
4.2.7 Dimensão Região	43
4.2.8 Dimensão Estado	43
4.2.9 Dimensão Município	44
4.2.10 Dimensão Espectro Político	44
4.2.11 Dimensão Partido	44
4.2.12 Dimensão Cor/Raça	45
4.2.13 Dimensão Gênero	45
4.2.14 Dimensão Situação de Totalização	46
4.2.15 Fato Candidaturas	46

4.2.16	Fato Perfil Candidaturas	50
5	Desenvolvimento	54
5.1	Ferramentas Utilizadas	54
5.1.1	Banco de Dados MariaDB	54
5.1.2	Pentaho Data Integration (PDI)	54
5.1.3	APIs	56
5.1.4	Laravel e Lumen	56
5.1.5	<i>React</i>	57
5.1.6	<i>Highcharts</i>	57
5.2	Experimento	58
5.2.1	Arquivos	58
5.2.2	Processo ETL	59
5.2.3	Visualização dos dados	60
6	Resultados	65
6.1	<i>Dashboards</i> sem o filtro Espectro Político	65
6.1.1	Número de Candidatos	66
6.1.2	Número de Eleitos	66
6.1.3	Candidatos e Eleitos	67
6.1.4	Eleitos - Cenário Político	68
6.1.5	Candidatos por gênero	69
6.1.6	Candidatos por cor/raça	69
6.2	<i>Dashboards</i> com o filtro Espectro Político selecionado	70
6.2.1	Número de candidatos do espectro	71
6.2.2	Número de eleitos do espectro	71
6.2.3	Candidatos por gênero	72
6.2.4	Eleitos por gênero	72
6.2.5	Candidatos por cor/raça	74
6.2.6	Eleitos por cor/raça	75
6.2.7	Candidatos e Eleitos por cargo	76
6.2.8	Candidatos eleitos por partido	76
6.3	Considerações	77
7	Conclusão	79
	Bibliografia	81

Lista de Figuras

2.1	Tecnologias de informação usadas como ferramentas de BI. Fonte: adaptação de (BRAGA, 2005)	18
2.2	Modelo Estrela. Fonte: (MACHADO, 2004)	23
2.3	Modelo Floco de Neve. Fonte: (MACHADO, 2004)	24
2.4	Processo ETL. Fonte: (ELIAS, 2014)	25
2.5	Posição ideológica dos partidos brasileiros Fonte: (BOLOGNESI; RIBEIRO; CODATO, 2020)	33
4.1	Eleições Gerais - Modelo lógico Fato Candidaturas. Fonte: Elaborado pelos autores	47
4.2	Eleições Municipais - Modelo lógico Fato Candidaturas. Fonte: Elaborado pelos autores	49
4.3	Eleições Gerais - Modelo lógico Fato Perfil Candidaturas. Fonte: Elaborado pelos autores	51
4.4	Eleições Municipais - Modelo lógico Fato Perfil Candidaturas. Fonte: Elaborado pelos autores	53
5.1	Pentaho Data Integration. Fonte: Elaborado pelos autores	55
5.2	Conexão do PDI com o MariaDB. Fonte: Elaborado pelos autores	60
5.3	Exemplo da carga de dimensões. Fonte: Elaborado pelos autores	60
5.4	Exemplo da carga dos fatos. Fonte: Elaborado pelos autores	61
5.5	Processo de desenvolvimento. Fonte: Elaborado pelos autores	61
5.6	Tela do sistema versão <i>desktop</i> . Fonte: Elaborado pelos autores	62
5.7	Tela do sistema versão <i>mobile</i> . Fonte: Elaborado pelos autores	63
5.8	Tela do sistema <i>dashboards</i> do espectro político. Fonte: Elaborado pelos autores	64
6.1	Gráfico Número de candidatos. Fonte: Elaborado pelos autores	66
6.2	Gráfico Número de eleitos. Fonte: Elaborado pelos autores	67
6.3	Gráfico Candidatos e Eleitos. Fonte: Elaborado pelos autores	68
6.4	Gráfico Eleitos - Cenário Político. Fonte: Elaborado pelos autores	68
6.5	Gráfico Candidatos por gênero. Fonte: Elaborado pelos autores	69
6.6	Gráfico Candidatos por cor/raça. Fonte: Elaborado pelos autores	70
6.7	Card Número de candidatos. Fonte: Elaborado pelos autores	71
6.8	Card Número de eleitos. Fonte: Elaborado pelos autores	71
6.9	Gráfico Candidatos por gênero. Fonte: Elaborado pelos autores	72
6.10	Gráfico Eleitos por gênero. Fonte: Elaborado pelos autores	73
6.11	Eleitorado por gênero. Fonte: Portal estatísticas eleitorais do TSE	74
6.12	Candidatos por cor/raça. Fonte: Elaborado pelos autores	75
6.13	Eleitos por cor/raça. Fonte: Elaborado pelos autores	75
6.14	Candidatos e Eleitos por cargo. Fonte: Elaborado pelos autores	76
6.15	Candidatos eleitos por partido. Fonte: Elaborado pelos autores	77

Lista de Tabelas

4.1	Regras de Negócios. Fonte: Elaborado pelos autores	39
4.2	Tabela dimensão Abrangência. Fonte: Elaborado pelos autores	40
4.3	Tabela dimensão Ano. Fonte: Elaborado pelos autores	41
4.4	Tabela dimensão Cargo. Fonte: Elaborado pelos autores	41
4.5	Tabela dimensão Tipo de Eleição. Fonte: Elaborado pelos autores	42
4.6	Tabela dimensão Turno. Fonte: Elaborado pelos autores	42
4.7	Tabela dimensão Eleição. Fonte: Elaborado pelos autores	43
4.8	Tabela dimensão Região. Fonte: Elaborado pelos autores	43
4.9	Tabela dimensão Estado. Fonte: Elaborado pelos autores	43
4.10	Tabela dimensão Município. Fonte: Elaborado pelos autores	44
4.11	Tabela dimensão Espectro Político. Fonte: Elaborado pelos autores	44
4.12	Tabela dimensão Partido. Fonte: Elaborado pelos autores	45
4.13	Tabela dimensão Cor/Raça. Fonte: Elaborado pelos autores	45
4.14	Tabela dimensão Gênero. Fonte: Elaborado pelos autores	45
4.15	Tabela dimensão Situação de Totalização. Fonte: Elaborado pelos autores .	46
4.16	Eleições Gerais - Tabela Fato Candidaturas. Fonte: Elaborado pelos autores	46
4.17	Eleições Municipais - Tabela Fato Candidaturas. Fonte: Elaborado pelos autores	48
4.18	Eleições Gerais - Tabela Fato Perfil Candidaturas. Fonte: Elaborado pelos autores	50
4.19	Eleições Municipais - Tabela Fato Perfil Candidaturas. Fonte: Elaborado pelos autores	52
5.1	Etapas do desenvolvimento. Fonte: Elaborada pelos autores	58
5.2	Arquivos utilizados. Fonte: Elaborado pelos autores	59

Lista de Abreviações

API	<i>Application Programming Interface</i>
BI	<i>Business Intelligence</i>
CGU	Controladoria Geral da União
CNJ	Conselho Nacional de Justiça
CSV	<i>Comma Separated Values</i>
DM	<i>Data Mart</i>
DW	<i>Data Warehouse</i>
ETL	<i>Extract, Transform and Load</i>
JSON	<i>JavaScript Object Notation</i>
LAI	Lei de Acesso à Informação
OLAP	<i>Online Analytical Processing</i>
PDI	<i>Pentaho Data Integration Community</i>
TSE	Tribunal Superior Eleitoral

1 Introdução

Seguindo as diretrizes da Lei de Acesso à Informação (BRASIL, 2011), o Brasil disponibiliza alguns dados governamentais em formato aberto para a população. Os dados são fornecidos pelos órgãos públicos e podem ser encontrados em seus sites, seja nas esferas municipal, estadual ou federal e nos três poderes - legislativo, executivo e judiciário e podem conter informações sobre eleições, projetos de lei, gastos públicos, orçamento, verbas distribuídas, gastos de agentes públicos entre outras informações, sendo assim uma grande fonte de pesquisa e informação. Podem ser considerados dados abertos um conjunto de dados que estão submetidos a uma licença aberta que permite seu uso e distribuição livremente,(ALCANTARA et al., 2015).

Tais dados podem ser utilizados para determinar o Índice de Desenvolvimento Humano de um município, o perfil dos políticos de uma determinada região, a quantidade de recursos destinados para projetos sociais, quanto gastam os políticos, entre outros. Esses dados podem ser relevantes quando transformados em informações e disponibilizados para estudos, partidos, políticos e para a população em geral.

O formato como os dados são disponibilizados aos interessados (que podem ser pesquisadores, partidos políticos, jornalistas e até o cidadão curioso) podem acabar os desencorajando em sua busca por informações. O tratamento dos dados muitas vezes requer um conhecimento de *hardware* e *software*. O *hardware* precisa ser dimensionado para suportar um grande volume de dados e na parte de *software*, pode ser necessário desenvolver planilhas, gráficos e banco de dados.

O uso de Sistemas de Informação aliado as técnicas da área de Ciências de Dados permite o tratamento, o armazenamento e a análise desses dados transformando-os em informações úteis para tomada de decisão e conhecimento. Este processo é chamado de Inteligência de Negócios (*Business Intelligence* - BI), segundo (NOVATO, 2017), e é utilizado dentro de organizações públicas e privadas. Porém, dada as características dos dados públicos: grande volume, diversas fontes de dados e necessidade de tratamento dos dados faz com que a utilização das diversas técnicas de BI auxiliem na busca de

informações do setor público.

1.1 Problema

No Brasil o acesso à informação pública é garantido pela Constituição Federal (BRASIL, 1988). A lei da Transparência sancionada em 27 de maio de 2009 (BRASIL, 2009), assegura ao cidadão que o estado, município e a União divulgarão dados na Internet em tempo real "[...] liberação ao pleno conhecimento e acompanhamento da sociedade, em tempo real; de informações pormenorizadas sobre a execução orçamentária e financeira, em meios eletrônicos de acesso público". Em complemento a esta lei temos a Lei nº 12.527, de 18 de novembro de 2011, também conhecida como Lei de Acesso à Informação (LAI), que determina aos órgãos públicos divulgarem, independentemente de solicitações, informações de interesse geral ou coletivo, salvo aquelas cuja confidencialidade esteja prevista no texto legal. Isto deverá ser feito através de todos os meios disponíveis e obrigatoriamente na internet (BRASIL, 2011).

As leis brasileiras garantem o acesso aos dados governamentais, porém cada vez mais se faz necessário a utilização de sistemas de informação para auxiliar na visualização e análise desses dados por cidadãos, uma vez que para a participação deles na gestão pública é fundamental que os mesmos tenham conhecimento e compreensão das informações divulgadas (FIGUEIREDO; SANTOS, 2013).

Os Dados Abertos Governamentais no Brasil são encontrados em diversos formatos, sendo o mais comum: *.csv*, classificado como 3 estrelas, segundo o modelo de classificação de *Open Data* de Tim Berners-Lee, (HAUSENBLAS, 2012). Nesse formato um usuário comum não consegue manipular os dados contudo, é possível a utilização das técnicas de *Business Intelligence* para manipulação dos dados.

Apontado os problemas neste trabalho levantamos a seguinte questão: Como as técnicas de *Business Intelligence* podem ser utilizadas na construção de ferramentas que permitem a análise de dados públicos e a construção de *dashboards* para visualização de informações?

1.2 Objetivos

O objetivo do trabalho é o desenvolvimento de um sistema web que permite a geração de *dashboards* a partir das técnicas de Data Analytics para expor informações das eleições do Brasil ao cidadão. Como objetivo específico temos a criação dos seguintes *dashboards*:

- Construir o *Data Warehouse* (DW) sobre os dados do repositório das eleições brasileiras do site do TSE;
- Construir *dashboards* que exibem informações sobre: Candidaturas e Eleitos por partido, por cargo, por gênero e por raça;
- Construir *dashboards* que exibem informações sobre: Candidatos e Eleitos por Espectro Político¹;
- Fornecer uma visualização de dados mais clara e transparente para o cidadão.

1.3 Metodologia

Na presente seção, é feita a caracterização desta pesquisa quanto à sua tipologia. Para tanto, consultou-se (SILVA; MENEZES, 2005) e (GIL et al., 2002), como referências sobre metodologia de pesquisa científica, na literatura, como forma de classificar o trabalho em questão.

Segundo (SILVA; MENEZES, 2005) as pesquisas podem ser classificadas como **básica ou aplicada**. Este trabalho adota a metodologia de **pesquisa aplicada**, uma vez que o objetivo é trazer conhecimentos para a sociedade, dando acesso e visibilidade aos dados e informações das eleições. Segundo o autor, a pesquisa aplicada tem como objetivo "gerar conhecimentos para aplicação prática e dirigidos à solução de problemas específicos. Envolve verdades e interesses locais".

Ao se tratar da abordagem do problema conforme (GIL et al., 2002) a pesquisa pode ser classificada em: *(i) pesquisa quantitativa*, traduz em números e taxas as opiniões

¹Segundo (SCHEEFFER et al., 2016), Espectro Político é a classificação de diferentes posições políticas dos partidos entre dois polos que se distinguem: a esquerda e a direita.

ou informações coletadas e utiliza-se de técnicas estatísticas para classificação e análises; *(ii) pesquisa qualitativa*, o conjunto de categoria é modificado e refeito várias vezes no processo com o objetivo de obter resultados mais significativos, a pesquisa é descritiva e as informações não são quantificáveis e os dados obtidos são analisados individualmente. Considerando tais definições, o presente trabalho pode ser classificado como uma **pesquisa qualitativa**.

Quando considera-se os objetivos do trabalho (GIL et al., 2002) classifica as pesquisas em três grupos: *(i) exploratória*, cuja finalidade é proporcionar maiores informações sobre determinado assunto, facilitar a delimitação de um tema de trabalho; *(ii) descritiva*, que pretende descrever as características de determinada população, envolvendo técnicas padronizadas de coleta de dados; e *(iii) explicativa*, que registra, analisa e interpreta os fenômenos estudados, procurando identificar suas razões, seus fatores determinantes e suas causas. Baseado nessas definições este trabalho se encaixa em uma **pesquisa exploratória**.

Ao analisar os procedimentos técnicos (GIL et al., 2002) classifica as pesquisas como: pesquisa bibliográfica, pesquisa documental, pesquisa experimental, levantamento, estudo de caso, pesquisa ação e pesquisa participante. Conforme esta classificação o trabalho trata de uma **pesquisa bibliográfica**, pois é realizada a partir de material já publicado, como: artigos, dissertações, livros e trabalhos de conclusão de curso.

Inicialmente, foram realizadas pesquisas bibliográficas a respeito dos diversos assuntos relacionados ao projeto com o objetivo de abordar os pontos fundamentais que serão utilizados no processo de desenvolvimento do sistema web. Na sequência foi feito um planejamento de quais dados eleitorais seriam úteis para a presente pesquisa e quais *dashboards* deveriam ser construídos para trazer informação relevante ao cidadão brasileiro, utilizando-se de ferramentas apropriadas para construção destes modelos gráficos. Em seguida, foram realizados uma série de ciclos incrementais de planejamento, priorização de tarefas, desenvolvimento e testes. E finalmente o desenvolvimento de um sistema web, como prova de conceito, para testes e avaliação da proposta de pesquisa

1.4 Organização do Trabalho

O presente trabalho está estruturado em capítulos da seguinte forma:

- Capítulo 1: Introdução - breve apresentação do tema, objetivos e metodologia deste trabalho.
- Capítulo 2: Fundamentação Teórica - revisão dos conceitos fundamentais para o entendimento deste trabalho como *Business Intelligence*, *Data Warehouse*, Dados Abertos Governamentais, Eleições no Brasil e Espectro Político.
- Capítulo 3: Trabalhos Relacionados - apresenta três trabalhos: o primeiro faz a classificação do sistema web desenvolvido como aplicativo cívico, os demais apresentam sistemas que se assemelham ao aplicativo desenvolvido bem como as técnicas de *Business Intelligence* por eles utilizadas.
- Capítulo 4: Modelagem - apresentação das regras de negócio, requisitos e modelagem dimensional do projeto de DW.
- Capítulo 5: Desenvolvimento - descreve todas etapas e ferramentas utilizadas no desenvolvimento do sistema web.
- Capítulo 6: Resultados - análise das consultas obtidas no sistema e comparação com os *dashboards* do aplicativo de estatísticas eleitorais do Tribunal Superior Eleitoral.
- Capítulo 7: Conclusão - considerações finais relacionadas ao objetivo do trabalho e sua contribuição, além de indicar os trabalhos futuros.

2 Fundamentação Teórica

O presente capítulo apresenta os principais conceitos utilizados no desenvolvimento deste trabalho. A primeira seção traz o conceito de *Business Intelligence* (BI). Na sequência, é discutido sobre os Dados Abertos bem como o acesso a informações providas por órgãos públicos e sua importância para a população. Por fim, discorre-se sobre o Tribunal Superior Eleitoral (TSE), órgão responsável por todo o processo eleitoral do Brasil e a principal fonte de dados utilizada neste trabalho, além de uma breve definição de Espectro Político.

2.1 *Business Intelligence*

As organizações, sejam elas públicas ou privadas, geram através de seus sistemas de informação um grande volume de dados. Os dados de maneira isolada geram pouca informação para tomada de decisão, tornando-se necessário a adoção de ferramentas para descoberta de informações. No setor empresarial, as tecnologias de inteligência de negócio fazem com que a empresa obtenha vantagem competitiva face aos seus concorrentes. Já nos órgãos públicos, a utilização desses recursos podem promover a melhoria da qualidade de vida da população com o aproveitamento correto dos recursos públicos ou ainda na divulgação das informações da gestão pública, permitindo ao cidadão ter informações e conhecimento das aplicações dos recursos públicos pelos governantes.

O termo *Business Intelligence* (BI), em tradução livre: Inteligência de Negócios, surge dentro desse contexto e pode ser definido como um conjunto de ferramentas, métodos e tecnologias que auxiliam o processo de extração de dados com o objetivo de transformá-los em informações que servirão de suporte para a tomada de decisão. Segundo (TURBAN et al., 2009), o termo BI foi criado na década de 1990 pelo Gartner Group, mas o conceito começou a ser difundido na década de 1970 com o início dos sistemas de geração de relatório. Esses sistemas eram estáticos, bidimensionais e não havia recursos de análise. Posteriormente, na década de 1980, surgiram os sistemas de informações exe-

cutivas (EIS) que introduziram os relatórios dinâmicos multidimensionais, prognósticos e previsões, análise de tendências, detalhamento, acesso a status e fatores críticos de sucesso. Novos recursos foram agregados aos EIS com o tempo e a nomenclatura BI passou a ser adotada para nomear esses sistemas.

Conforme (BATISTA, 2017), BI é “um conjunto de ferramentas e aplicativos que oferece aos tomadores de decisão possibilidade de organizar, analisar, distribuir e agir, ajudando a organização a tomar decisões melhores e mais dinâmicas”, podemos concluir que o uso das tecnologias de inteligência de negócio são aplicáveis independente dos motivos que se deseja a informação ou da decisão que se pretende embasar. Este conceito engloba algumas tecnologias, como *Data Warehouse* (DW) e *Data Marts* (DM), ferramentas de ETL (*Extract, Transform e Load*), fundamentais na transformação de dados transacionais em informacional, além de aplicações, como OLAP (*On-Line Analytical Processing*) e *Data Mining*.

Segundo (BRAGA, 2005), o objetivo de utilização dessas tecnologias é garantir um processo de gestão do conhecimento dentro das organizações através da manipulação do grande volumes de dados, a fim de extrair informações e conhecimento dos mesmos, para que se possa tomar decisões a partir de dados históricos. A figura 2.1, ilustra como essas ferramentas são usadas para dar suporte às atividades de tomada de decisão nas organizações.

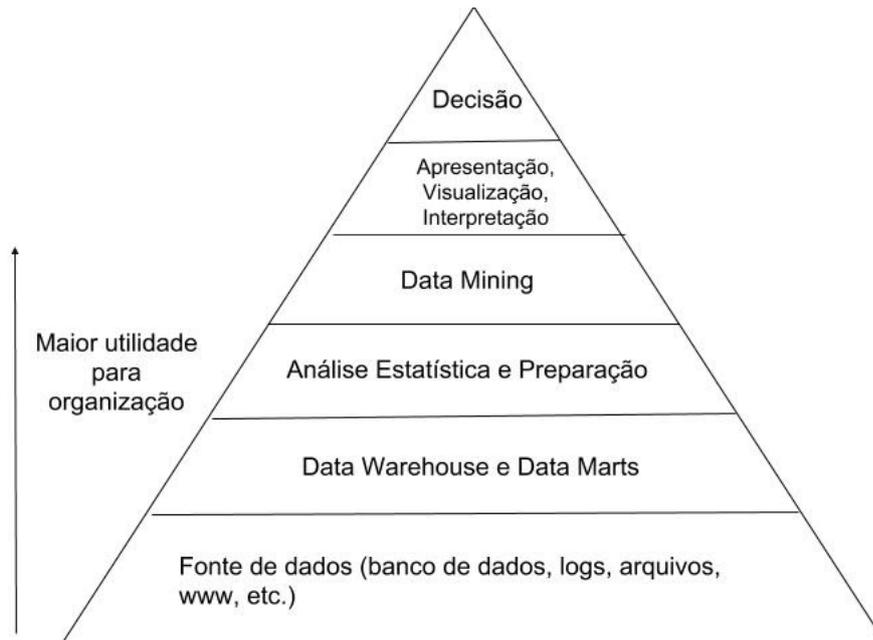


Figura 2.1: Tecnologias de informação usadas como ferramentas de BI. Fonte: adaptação de (BRAGA, 2005)

Neste trabalho serão utilizadas técnicas para construção de *Data Warehouse*, o processo de ETL para tratamento dos dados abertos e os *Dashboards* como forma de visualizar as informações obtidas.

2.1.1 *Data Warehouse*

O grande volume de dados gerados diariamente pelas organizações, públicas ou privadas, são armazenados em banco de dados transacionais, esses dados constantemente são atualizados e envolvidos em diversas transações. As redundâncias neste tipo de banco não ocorrem e informações históricas não ficam armazenadas por muito tempo. Devido a estas características, a consulta de dados históricos para a tomada de decisão não é satisfatória, uma vez que existe um grande número de operações que fazem com que as consultas se tornem lentas e em alguns casos inviável.

Para solucionar este problema temos o *Data Warehouse*(DW) que é um banco de dados específico para a análise de dados históricos. Segundo (INMON, 2005), DW pode ser entendido como: “Uma coleção de dados, orientada a assunto, integrada, variante no tempo e não volátil, para suporte ao gerenciamento dos processos de tomada de decisão”. Para (PERKINS, 1995), o DW é visto como um armazém de dados que provê as necessi-

dades de informações de uma organização. Essas informações podem ser complexas, mas com mecanismos simples elas podem ser obtidas de forma rápida e precisa.

Conforme (TAURION, 1998), ao reunir informações dispersas de diversos sistemas, o *Data Warehouse* permite que sejam feitas consultas e análises bastante eficazes, transformando dados esparsos em informações antes inacessíveis ou sub-aproveitadas, que podem ser convertidas em estratégias para os negócios. Quando analisamos o DW com dados abertos governamentais podemos dizer que as informações podem ser convertidas em estratégias para melhoria da qualidade de vida da população ou ainda como instrumento de fiscalização das ações dos governos, por parte dos cidadãos.

(KIMBALL; ROSS, 2011) afirma, que um DW fornece acesso a dados que são consistentes podendo ser separados e combinados usando qualquer medição. Outra possibilidade é ter acesso a um conjunto de ferramentas para consulta, análise e apresentação das informações. Sendo assim o *Data Warehouse*, é um local onde se publica dados confiáveis e a qualidade desses impulsiona à reengenharia de negócios.

Algumas vezes, BI é confundido com DW. Porém ele é uma ferramenta incorporada em um projeto de *Business Intelligence*. De modo geral, um DW deve possibilitar a análise de grandes volumes de dados, ser adaptável, apresentar dados íntegros e seguros a fim de servir como suporte para tomada de decisão.

Data Mart

De acordo com (KIMBALL; ROSS, 2011) o *Data Mart* (DM) é uma parte do todo que compõe o *Data Warehouse*, ou seja, o *Data Mart* representa os dados de um único processo de negócio. Enquanto o DW abrange todas as grandes áreas da empresa, o DM atua como um subconjunto dos dados de um *Data Warehouse*, fracionado e levemente resumido.

Bill Inmon e Ralph Kimball, dois pioneiros nos estudos de *Data Warehouse*, diferem em suas opiniões sobre como um DW deve ser projetado da perspectiva de organização. Na visão de (INMON, 2005), deve-se modelar um DW único para toda organização e posteriormente partir para construção de DM por assuntos ou departamentos, essa abordagem é conhecida como *top-down*. Já para (KIMBALL; ROSS, 2011) a

modelagem do DW é justamente contrária, primeiramente deve-se construir *Data Marts* separados que, conforme forem evoluindo deverão ser integrados para chegar a um modelo único onde o *Data Warehouse* é um conjunto de *Data Marts*, essa abordagem é conhecida como *bottom-up*.

A abordagem *top-down* requer uma visão geral da solução e possui grande abrangência, porém implica em grande prazo de implementação. É indicada para projetos em que todo o escopo é conhecido e é mais simples de ser implementado. Já a abordagem *bottom-up* é um método que tem maiores resultados de sucesso, pois o processo de desenvolvimento é evolutivo, apresenta um retorno rápido uma vez que a estrutura pode ser utilizada antes da completa construção do DW. É indicado para projetos onde é difícil enxergar os limites e sua complexidade total.

Modelagem Dimensional

Segundo (TAURION, 1998), a modelagem dimensional é uma metodologia que permite modelar logicamente dados para melhorar o desempenho de consultas e prover facilidade de utilização a partir de um conjunto de eventos básicos de medição. Os modelos dimensionais são compreensíveis, previsíveis, ampliáveis e resistentes ao ataque específico de grupos de usuários de negócio, por se manter fiel à simplicidade, ter uma perspectiva voltada para as necessidades analíticas da empresa, e especialmente ao seu formato simétrico, em que todas as dimensões normalmente são iguais pontos de entrada na tabela de fatos. Em outras palavras, a modelagem dimensional é a técnica de projeto lógica utilizada para modelar *Data Warehouse*. Os dados são organizados em dois tipos de tabelas: Fato e Dimensões.

Tabela Fato

Conforme (KIMBALL, 2008), a tabela de fatos contém as medidas numéricas produzidas por um evento de medição operacional no mundo real. Assim, o projeto fundamental de uma tabela de fatos é inteiramente baseado em uma atividade física e não é influenciado pelos relatórios eventuais que podem ser produzidos. Além das medidas numéricas, uma tabela de fatos sempre contém chaves estrangeiras para cada uma de

suas dimensões associadas, bem como chaves de dimensão degeneradas opcionais e selos de data / hora. As tabelas de fatos são o principal alvo de cálculos e agregações dinâmicas decorrentes de consultas.

Em outras palavras a tabela fato armazena o que ocorreu, é o fato propriamente dito, por isso ela tem esse nome. Normalmente, as tabelas de fatos ocupam mais espaço do que as de dimensão. Aproximadamente 90% de um DW é ocupado por elas, uma vez que estas armazenam um registro para cada acontecimento que representa.

(KIMBALL, 2008) ainda agrupa as tabelas fatos em três tipos: granularidade transacional, granularidade *snapshot* e granularidade *snapshot* cumulativa, cada um com suas respectivas características:

- Granularidade transacional: é medida em um único instante, as medidas, valores e a informação são só válidos para aquele instante e contexto específico, normalmente é uma das tabelas fatos que possui o maior número de registros.
- Granularidade *snapshot*: corresponde a um intervalo de tempo pré-definido, diferente da transacional, as informações resumem as informações durante ou ao final do intervalo de tempo. É uma das tabelas mais utilizadas para se criar um relatório. Um dos pontos positivos é ser possível garantir que todas as chaves estrangeiras estejam presentes.
- Granularidade *snapshot* cumulativa: ela é usada quando um evento possui um início e fim bem definidos, por exemplo, em matrículas de alunos, registros de call centers, processamento de pedidos. Essa tabela é revisada e reescrita à medida que o evento progride através das suas etapas de início, meio e fim. Normalmente o *snapshot* cumulativo é muito menor do que as outras tabelas, por utilizar a estratégia de reescrita.

Na concepção de (KIMBALL, 2008) é importante saber que para cada situação a tabela será diferente. Ela deve responder às perguntas de negócio dos usuários chaves e o objetivo do DW deve estar muito claro antes de qualquer coisa. Identificar os assuntos, granularidade e a complexidade é essencial para um funcionamento eficiente, pois por

exemplo, quanto menor for o grão dos dados em uma tabela fato, maior será a quantidade de dados e detalhes.

Tabela Dimensão

Após entender o que é um fato e a sua granularidade é preciso saber com qual informação ela será agrupada e cruzada. A tabela dimensão possui elementos que participam da tabela fato, determinando um contexto do assunto, contendo as características do evento (KIMBALL; ROSS, 2011; CECI, 2012).

As tabelas de dimensão são compostas por atributos que contém a descrição do negócio. Seus atributos são fontes das restrições de consultas, agrupamento dos resultados e cabeçalhos para relatórios. Ela possui aspectos pelos quais se pretende observar as métricas relativas ao processo modelado. Estas tabelas apresentam uma chave primária que lhes dá unicidade e serve de chave estrangeira para as operações de junção com as tabelas de fatos.

Para (CECI, 2012) uma forma de criar dimensões é após identificar os assuntos que o DW irá tratar, ou seja, os fatos, deve-se levantar as dimensões que são utilizadas para agrupar e cruzar as informações contidas no DW, para isso é necessário responder a questões do tipo: ‘quem?’, ‘quando?’, ‘o quê?’ e ‘onde?’. Por exemplo, se a tabela fato for a venda de um produto, pode-se perguntar ‘quem’ realizou a venda, ‘quando’ ocorreu, ‘o que’ foi vendido e ‘onde’ a venda ocorreu, com isso, é possível ver que nesse exemplo simples terá quatro dimensões: de vendedor, tempo, produto e local.

Diferentemente das tabelas fatos as dimensões não possuem muitas linhas, porém diversas colunas e devem ser entendidas como tabelas de filtro para as consultas das análises e relatórios. Ainda de acordo com (KIMBALL; ROSS, 2011) as tabelas dimensões são consideradas a alma do DW, pois somente com elas é possível dar sentido aos dados, criando análises importantes para o negócio, sendo impulsionadores da experiência do BI como um todo para seus usuários.

Existem dois tipos de modelos dimensionais para *Data Warehouse*, são eles Estrela (*Star Schema*) e Floco de Neve (*Snow Flake*).

Modelo Estrela (*Star Schema*)

Segundo (PUJARI, 2001) o esquema estrela é um paradigma de modelagem em que o DW contém uma grande tabela fato central única e um conjunto de tabelas dimensões menores, uma para cada dimensão, cada uma única e altamente desnormalizada, conforme a figura 2.2. Possuindo as seguintes vantagens:

- Fácil de entender.
- Fácil de definir hierarquias.
- Reduz o número de JOINS² físicos.
- Requer baixa manutenção.
- Metadados muito simplificados.

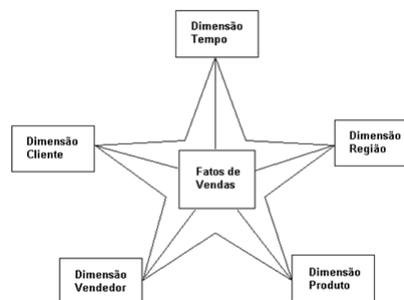


Figura 2.2: Modelo Estrela. Fonte: (MACHADO, 2004)

(BARBIERI, 2011) aponta que o modelo tem como desvantagem a criação de diversos DM sem uma perfeita ligação entre eles e uma provável duplicação de esforços na fase de extração, preparação e carga de dados. O autor afirma que a fácil implementação, a possibilidade de entrega de projetos em pequenas partes e o custo de espaço em disco - que hoje é bem barato - acaba fazendo com que esse modelo seja o mais utilizado.

Modelo Floco de Neve (*Snow Flake*)

O modelo consiste em uma extensão do modelo Estrela onde cada uma das “pontas da estrela” passa a ser o centro de outras estrelas. Isto porque cada tabela de dimensão

²A tradução para o português da palavra JOIN é juntar. O comando JOIN do SQL tem a função básica de agregar tabelas mediante a um campo que faça sentido às mesmas.

seria normalizada, “quebrando-se” a tabela original ao longo de hierarquias existentes em seus atributos.

Para (BARBIERI, 2011) a ênfase do esquema flocos de neve, proposto por Inmon, sempre foi um “grande depósito central de informações empresariais tratadas, limpas e integradas, construído inicialmente, e de onde outros depósitos secundários são originados e construídos”. Segundo o autor o modelo se concentra no estilo mais tradicional de construção de bancos de dados, muito próximo daquele sugerido nos primeiros projetos de bancos de dados, nos quais se buscava uma forte integração entre todos os dados da empresa que habitavam áreas funcionais diferentes. A figura 2.3 mostra uma representação do modelo.



Figura 2.3: Modelo Floco de Neve. Fonte: (MACHADO, 2004)

O Modelo Floco de Neve reduz o espaço de armazenamento dos dados dimensionais mas acrescenta várias tabelas ao modelo, deixando-o mais complexo, tornando mais difícil a navegação pelos *softwares* que utilizarão o banco de dados. Um outro fator é que mais tabelas serão utilizadas para executar uma consulta, então mais JOINS de instrução SQL³ serão feitos, tornando o acesso aos dados mais lento do que no modelo estrela.

2.1.2 ETL

ETL significa *Extract, Transform and Load*, em tradução literal significa Extração, Transformação e Carga. Esta é uma etapa de construção de um DW que é uma das mais críticas. Conforme (KIMBALL; ROSS, 2011) ETL é a fundação do DW. A etapa é essencial e necessita de um bom planejamento. Quando esta fase é bem plane-

³SQL(*Structured Query Language*) é a linguagem usada para executar comando em bancos de dados relacionais, isto é, baseado em tabelas.

jada a qualidade dos dados e a padronização consistente são reforçadas. Na opinião de (KIMBALL; ROSS, 2011), ETL é uma atividade *backend* e não vista pelo usuário final, porém pode consumir facilmente mais de 70% dos recursos destinados à implementação e manutenção do DW. O ETL é muito mais que um meio de pegar os dados do banco de origem para o *data warehouse*:

- Remove erros e corrige dados não padronizados ou não existentes na estrutura.
- Provê documentação das medidas.
- Captura o fluxo de dados.
- Ajusta os dados de múltiplas fontes de origem.
- Estrutura os dados para serem utilizados em ferramentas diversas.

Na fase de extração os dados são obtidos de suas fontes e conduzidos para a *staging area* (área de transição ou área temporária), onde são convertidos para um único formato. Na transformação são criadas medidas, limpeza de dados, seleção de colunas que serão utilizadas, unificação dos dados no contexto desejado, mesclagem e controle dos dados, como por exemplo geração de chaves primárias. A carga consiste em fisicamente estruturar e carregar os dados para dentro da camada de apresentação seguindo o modelo dimensional. A figura 2.4 mostra este processo.

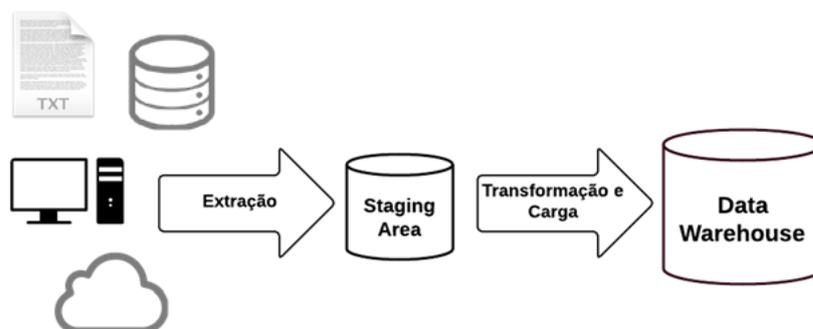


Figura 2.4: Processo ETL. Fonte: (ELIAS, 2014)

O processo de ETL é essencial para a criação das estruturas de dimensões e fatos no ambiente do DW. É ele que faz a ligação entre o operacional e o DW. O processo deve

ser bem planejado para evitar transtornos futuros e até mesmo para que não ocasione, em casos extremos, a interrupção dos sistemas operacionais da empresa. Dessa forma, o DW terá informações tratadas, com qualidade e grande valor para apoiar as decisões organizacionais. Um bom ETL deve ter escalabilidade e ser manutenível. Devemos analisar a janela de operação, pois como trabalha com grandes volumes de dados não é em qualquer momento que ele poderá ser executado. Devemos também analisar a periodicidade de execução.

2.1.3 *Dashboards* para BI

Dashboards ou painéis de controle são ferramentas de visualização que viabilizam a visualização de métricas e indicadores do negócio. Interface personalizável e capacidade de disponibilizar as informações de forma inteligente em tempo real, a partir de diversas fontes, são características fundamentais dos *Dashboards* de BI.

Segundo (FEW, 2006) *dashboards* são parte do gênero Ferramentas de Visualização de BI assim como os Indicadores, Relatórios, Alertas, etc. Mas o que os diferencia das demais ferramentas é a exibição visual das informações relevantes, consolidadas e organizadas em uma única tela, de modo que possa ser monitorada em único olhar.

Os *dashboards* podem ser compostos de textos e gráficos, porém, com ênfase nos últimos, pois quando bem elaborados, sua capacidade de comunicação é bem mais rápida, objetiva e direta. Além disso, podem ser compostos de indicadores, metas, alertas, índices, coeficientes, a depender da criatividade do design de *dashboards*.

2.2 Dados Abertos Governamentais

Segundo a (KNOWLEDGE, 2021), dados abertos são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras. Dentro dos dados abertos existe a categoria de Dados Abertos Governamentais.

Segundo (BRASIL, 2011), Dados Abertos Governamentais são dados produzidos pelo governo e colocados à disposição das pessoas de forma a tornar possível não apenas

a leitura e acompanhamento, mas também a sua reutilização em novos projetos, sítios e aplicativos; seu cruzamento com outros dados de diferentes fontes; e sua disposição em visualizações interessantes e esclarecedoras.

Existem oito princípios que estabelecem o que os dados abertos governamentais devem ser. Segundo (VAZ; RIBEIRO; MATHEUS, 2010), esses princípios foram criados em 2007 pelo grupo de especialistas denominado *OpenGovData*. São eles:

1. *Completos*. Todos os dados públicos estão disponíveis. Dado público é o dado que não está sujeito a limitações válidas de privacidade, segurança ou controle de acesso.
2. *Primários*. Os dados são apresentados tais como os coletados na fonte, com o maior nível possível de granularidade e sem agregação ou modificação.
3. *Atuais*. Os dados são disponibilizados tão rapidamente quanto necessário à preservação do seu valor.
4. *Acessíveis*. Os dados são disponibilizados para o maior alcance possível de usuários e para o maior conjunto possível de finalidades
5. *Compreensíveis por máquinas*. Os dados são razoavelmente estruturados de modo a possibilitar processamento automatizado.
6. *Não discriminatórios*. Os dados estão disponíveis para todos, sem exigência de requerimento ou cadastro.
7. *Não proprietários*. Os dados são disponíveis em formato sobre o qual nenhuma entidade detenha controle exclusivo.
8. *Livres de licenças*. Os dados não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial. Restrições sensatas relacionadas à privacidade, segurança e privilégios de acesso são permitidas.

No artigo (EAVES, 2009), o autor que é especialista em políticas públicas e ativista dos dados abertos, propõe as três leis dos dados abertos governamentais, que são:

1. Se o dado não pode ser encontrado e indexado na web, ele não existe.

2. Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado.
3. Se algum dispositivo legal não permitir sua replicação, ele não é útil.

Segundo (BRASIL, 2011), os dados abertos causam três efeitos sobre as políticas públicas, são eles: Inclusão, Transparência e Responsabilidade. O fornecimento de dados em formatos padronizados e acessíveis permite que qualquer usuário utilize qualquer ferramenta de software para adaptá-los às suas necessidades, por isso temos a inclusão como um dos efeitos. Uma vez que informações do setor público estão acessíveis e podem ser usadas da maneira adequada, permitindo que se tenha uma melhor ideia do governo, temos uma melhora na transparência. Quando o conjunto de dados abertos apropriados são devidamente associados, podem oferecer vários pontos de vista sobre o desempenho do governo no atingimento de suas metas em políticas públicas, temos então a responsabilidade como efeito.

Percebemos que a utilização de dados abertos governamentais associados a tecnologias de sistema de informação tornam possível a construção de serviços que operam em benefícios da população e dos governos. Ao cidadão é possível acompanhar a gestão pública e cobrar por melhorias uma vez que ele percebe as necessidades da comunidade já aos governos é possível melhorar e propor novos serviços e benefícios para a população a partir da análise dos dados.

2.2.1 Dados Abertos no Brasil

O acesso aos dados públicos no Brasil é garantido com base na Constituição de 1988, que determina que, a não ser em casos de exceção, as informações do governo devem ser públicas. Porém movimentos que buscam a abertura de dados governamentais no país surgiram a partir de 2005 e com o passar dos anos houve a intenção do governo federal de tornar a sua administração mais transparente por meio da publicação de informações de interesse da sociedade na internet (SILVA; PINHEIRO, 2017).

A Lei Complementar nº 101, de 4 de maio de 2000 estabeleceu normas públicas voltadas para gestão das finanças públicas e responsabilidade fiscal. Esta lei é um dos

maiores instrumentos de transparência em relação aos gastos públicos no Brasil e define padrões a serem seguidos pela União, Estados, Distrito Federal e Municípios (BRASIL, 2000).

A criação da Controladoria Geral da União (CGU) em 2003 representou um grande avanço do governo federal na intenção de permitir o acesso do cidadão à informação, transparência das ações do Estado e combate à corrupção. A CGU é o órgão de controle interno do Governo Federal responsável por realizar atividades relacionadas à defesa do patrimônio público e ao incremento da transparência da gestão, por meio de ações de auditoria pública, correição, prevenção e combate à corrupção e ouvidoria (BRASIL, 2019).

Em 2004, o Portal da Transparência do Governo Federal foi lançado pelo Ministério da Transparência e a CGU. Este é um site de livre acesso no qual o cidadão pode encontrar informações sobre como o dinheiro público é utilizado, além de se informar sobre assuntos relacionados à gestão pública do Brasil. Desde a criação, a ferramenta ganhou novos recursos, aumentou a oferta de dados ano após ano e consolidou-se como importante instrumento de controle social, com reconhecimento dentro e fora do país (BRASIL, 2020a). Segundo (VICTORINO et al., 2017), em 2006 através da Portaria Interministerial - CGU/MPOG n. 140, de 16 de março de 2006 - foi determinado pela CGU que é responsabilidade dos órgãos e entidades da Administração Pública Federal manter em seus sites informações detalhadas determinados de aspectos, como, por exemplo, execução orçamentária, licitações, contratações, entre outros. Estas devem ser mantidas em páginas específicas, denominadas Páginas de Transparência Pública.

Em 2009, a Lei Complementar nº 131, determinou a União, os Estados e Municípios disponibilizar em meio eletrônico e em tempo real, informações pormenorizadas sobre a execução orçamentária e financeira. Esta lei altera a Lei de Responsabilidade Fiscal (BRASIL, 2009). A Lei nº 12.527 sancionada em 2011, conhecida como Lei de Acesso à Informação (LAI) foi a responsável pelo início da disponibilização de dados abertos governamentais, pois ela regulariza o acesso a esse conteúdo. Segundo a lei é dever dos órgão e entidades promover, independente de requerimentos, a divulgação em seus sítios na Internet de informações de interesse coletivo por eles produzidas ou custodiadas. Hoje,

todos os sites ligados ao governo em qualquer instância, possui uma área de divulgação dos dados, o que facilita a busca pelo cidadão (BRASIL, 2011). No ano de 2016 tem-se a publicação do Decreto N° 8.777 que instituiu a Política de Dados Abertos do Poder Executivo Federal, destacando a criação da Infraestrutura Nacional de Dados Abertos (INDA) e do Plano de Dados Abertos (PDA) - um instrumento que operacionaliza a Política de Dados Abertos do Poder Executivo Federal (BRASIL, 2016).

Podemos observar que existem diversos dispositivos legais que contribuíram para a abertura e consolidação de dados nos portais eletrônicos de órgãos públicos no Brasil, no entanto ainda temos algumas dificuldades no acesso à informação, são elas: o formato dos dados que podem variar, pois cada site adota um formato, e a segunda dificuldade está na distribuição de dados por diversos sites da internet uma vez que cada entidade é responsável pela divulgação dos seus dados.

Neste trabalho a fonte de dados utilizada é do Tribunal Superior Eleitoral que possui um repositório de dados eleitorais publicados desde as eleições de 1945. Estes dados são disponibilizados, em formato *.zip*, para pesquisadores, imprensa e cidadãos em geral, principalmente aqueles que tenham interesse em analisar os dados de candidatura, resultados e prestação de contas.

2.3 Eleições no Brasil

O Tribunal Superior Eleitoral (TSE) é órgão máximo da Justiça Eleitoral e exerce papel fundamental na construção e no exercício da democracia brasileira. Suas principais competências estão fixadas pela Constituição Federal e pelo Código Eleitoral (Lei n° 4.737, de 15.7.1965⁴). O órgão tem ação conjunta com os tribunais regionais eleitorais (TREs), que são os responsáveis diretos pela administração do processo eleitoral nos estados e municípios brasileiros.

Dentre as ações mais usadas na atuação jurisdicional da Justiça Eleitoral destacam-se a ação de impugnação de registro de candidatura (AIRC), o recurso contra expedição de diploma (RCED), a ação de investigação judicial eleitoral (AIJE), com as respectivas representações e reclamações e a ação de impugnação de mandato eletivo (AIME). Cabe

⁴Acessar: http://www.planalto.gov.br/ccivil_03/leis/14737compilado.htm

também ao Tribunal apreciar a prestação de contas eleitorais e partidárias e julgar os recursos advindos dos Tribunais Regionais Eleitorais, tais como o recurso especial eleitoral (REspe), previsto no art. 276, I, do Código Eleitoral, e o recurso ordinário (RO), disposto no art. 276, II, do mesmo diploma (BRASIL, 2020b).

Além das competências descritas acima, o TSE como órgão público preza pela transparência do processo eleitoral e por isso divulga os dados das eleições do Brasil em seu site. As informações das eleições são divulgadas através de uma plataforma de estatísticas eleitorais⁵ desenvolvida pelo próprio órgão ou ainda através do Repositório de dados eleitorais que fornece os dados brutos das eleições.

O Repositório de dados eleitorais⁶ é uma compilação de informações brutas das eleições, desde de 1945, voltada para pesquisadores, imprensa e pessoas interessadas em analisar os dados de eleitorado, candidaturas, resultados e prestação de contas. Todos os arquivos fornecidos estão em formato TXT ou CSV e podem ser importados para qualquer programa estatístico, base de dados ou planilha eletrônica. Com o repositório é possível a criação de novos *softwares* para consultas e filtros, além de ser possível fazer o cruzamento desses dados com dados abertos de outras fontes.

Neste trabalho foram selecionados os dados referentes aos candidatos, partidos e resultados para criar visualizações referentes ao perfil de candidaturas e partidos brasileiros a fim de que com esses dados possam ter informações sobre o cenário político brasileiro.

2.4 Espectro Político

Desde o processo de redemocratização e abertura política após o regime militar, o Brasil dispõe de um sistema multi partidário bastante fragmentado, com trinta e três partidos políticos registrados atualmente no TSE. O que se observa nesse cenário é a dificuldade do conhecimento sobre as legendas e a classificação dos partidos conforme as ideologias por eles propagadas (MACIEL; ALARCON; GIMENES, 2018).

Segundo (SCHEEFFER et al., 2016), o espectro ideológico serve para classifi-

⁵Acessar: <https://www.tse.jus.br/eleicoes/estatisticas/estatisticas-eleitorais>

⁶Acesse: <https://www.tse.jus.br/hotsites/pesquisas-eleitorais/index.html>

car diferentes posições políticas entre dois pólos que se distinguem: a esquerda e a direita. O conceito de “ideologia” é complexo e varia conforme o tempo e a conjuntura histórico-cultural, podendo representar um “conjunto de ideias e valores respeitantes à ordem pública e tendo como função orientar os comportamentos políticos coletivos”.

De acordo com (MACIEL; ALARCON; GIMENES, 2018), o termo “direita” e “esquerda” surgiram no contexto da Revolução Francesa (1789) para indicar posicionamentos políticos. A esquerda priorizava o igualitarismo e a direita a aristocracia. Com o passar do tempo os termos foram associados ao liberalismo e conservadorismo. Sendo que a esquerda prioriza a igualdade e a direita a liberdade.

(HEYWOOD; OLIVEIRA, 2010) classifica as ideologias em “clássicas e “novas”. A ideologia clássica é pautada no liberalismo, conservadorismo, socialismo, nacionalismo e fascismo. Já as as novas ideologias abrangem movimentos como o feminismo, ecologismo, fundamentalismo religioso e multiculturalismo. O autor pondera que são chamadas de “novas” por conferirem notoriedade a certas áreas de debate ideológico anteriormente desconsideradas e por fomentarem o surgimento de perspectivas ideológicas novas e contestadoras. Esses conceitos são aplicados por cientistas políticos para a classificação de um partido político.

(BOLOGNESI; RIBEIRO; CODATO, 2020) apresenta vários métodos para analisar o espectro do partido. Porém segundo os autores é que não há partido que seja absolutamente coerente em suas posições. Logo, os cientistas políticos costumam criar classificações intermediárias como centro-direita ou centro-esquerda. Há posições predominantes (e não exclusivas) vinculadas mais à direita no primeiro caso e mais à esquerda no segundo. Já os partidos de centro seriam aqueles que mesclam posições ora mais à direita, ora mais à esquerda, sem que ocorra o domínio de uma visão sobre a outra.

Buscando solucionar o dilema de onde se posicionam os partidos políticos brasileiros, os autores (BOLOGNESI; RIBEIRO; CODATO, 2020) fizeram um questionário para que 519 cientistas políticos respondessem. A partir da resposta foi feita uma média das posições ideológicas das legendas existentes no Brasil. Os respondentes foram solicitados a classificar os partidos em uma escala de zero a dez, em que zero representa a posição mais à esquerda e dez, mais à direita. A figura 2.5 apresenta as posições de cada partido

segundo a pesquisa.

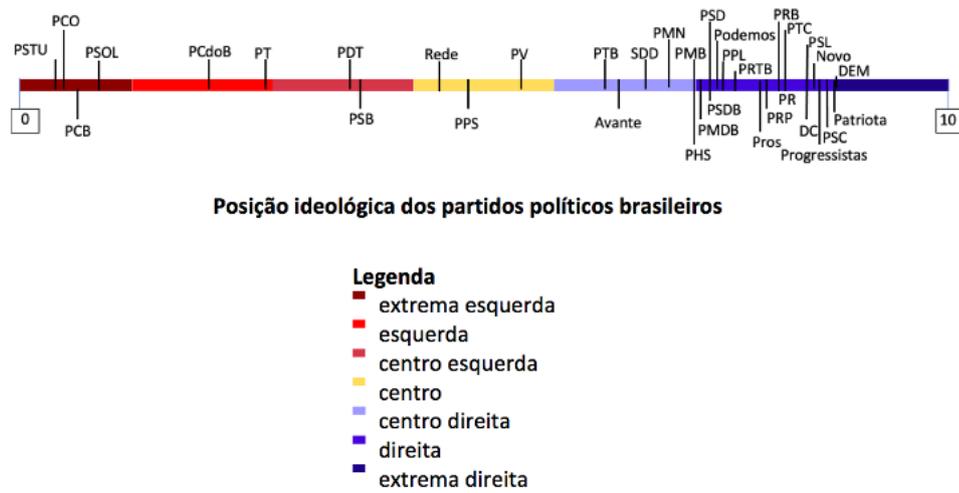


Figura 2.5: Posição ideológica dos partidos brasileiros Fonte: (BOLOGNESI; RIBEIRO; CODATO, 2020)

No presente trabalho utilizamos essa classificação para os partidos. As legendas são classificadas em Extrema-esquerda, Esquerda, Centro-Esquerda, Centro, Centro-direita, Direita, Extrema-direita. Esta definição nos ajudará a entender os *dashboards* relacionados a classificação dos resultados das eleições pelo espectro político.

3 Trabalhos Relacionados

O presente trabalho foi motivado por pesquisas sobre a manipulação de grandes volumes de dados e dados abertos governamentais. O interesse pelo assunto foi despertado ao cursar as disciplinas DCC127 - MINERAÇÃO DE DADOS e DCC147 - DATA WAREHOUSE. A justificativa da pesquisa e a proposta de solução tem origem na observação dos resultados de três estudos. Tais estudos estão relatados em (GOMES; SCHNEIDER; BEZERRA, 2018), (PIZANI; SIQUEIRA, 2018) e (VEIGA; GUIMARÃES; TANAKA, 2015). Nestes estudos foram relatados o planejamento da ação, sua execução, o processamento e tratamento dos dados, apresentação de resultados, métodos e recursos utilizados, adaptações e melhorias necessárias e dificuldades encontradas.

(GOMES; SCHNEIDER; BEZERRA, 2018) identifica e analisa dezoito aplicativos⁷ cívicos desenvolvidos no Brasil, que utilizam dados de financiamento de campanhas eleitorais, atuação política nos poderes executivos, legislativo e judiciário, e sua repercussão midiática e informacional. Com o objetivo de entender de que modo esses aplicativos apropriam-se dos dados abertos governamentais e como sua análise e categorização pode contribuir para a transparência política e a participação cidadã.

Segundo (GOMES; SCHNEIDER; BEZERRA, 2018) o levantamento dos aplicativos cívicos nacionais, foi feito mediante a coleta de notícias de jornais, congressos e sites especializados em dados abertos e em política. Realizaram também uma pesquisa informal com os membros do grupo de pesquisa dos autores, para que fosse indicados sites e outros grupos que estivessem utilizando *dataset* sobre os temas desejados. Outra forma de descoberta foi a utilização de alertas no Google Alertas com o objetivo de receberem notificações de qualquer notícia com os temas "Aplicativos Cívicos" ou "Tecnologias Cívicas". Com a pesquisa foi possível catalogar dezoito aplicativos que foram categorizados em três formas: patrocinador, assunto e o tipo de licença de uso do *software*. Com essa categorização eles observaram que mais da metade desses aplicativos estavam associados a

⁷O termo aplicativo aqui se refere a um software que possa ser utilizado através da infraestrutura da Internet, de navegadores para a World Wide Web (www), ou outros recursos computacionais e não para referenciar aplicativos de smartphones.

jornais ou a reportagens jornalísticas, mesmo estes sendo financiados por ONGs ou pelos próprios desenvolvedores estes aplicativos fazem interface com matérias jornalísticas.

(GOMES; SCHNEIDER; BEZERRA, 2018) conseguiram observar que existem poucos casos de dados abertos gerados por organizações não governamentais, sejam elas com ou sem fins lucrativos. A grande maioria dos dados abertos encontrados são produzidos pelo Estado. Os autores concluíram ainda existe pouca reflexão sobre a utilização de aplicativos cívicos no Brasil. Estudos sobre a construção de cada aplicativo e o acesso a ele podem trazer novas perspectivas sobre a utilização dos dados abertos.

O trabalho (PIZANI; SIQUEIRA, 2018) apresenta e implementa uma solução de BI com os dados abertos disponibilizados pelo Conselho Nacional de Justiça (CNJ). Esses dados divulgam a realidade dos tribunais brasileiros a respeito de sua estrutura, litigiosidade e análises essenciais para a gestão judiciária. Segundo o autor, o CNJ disponibiliza os dados, porém desnormalizados e além disso os nomes das colunas e indicadores disponibilizados no arquivo estão em um padrão que não significa nada para o analista de negócio ou partes interessadas, o que dificulta a sua compressão.

No desenvolvimento do trabalho (PIZANI; SIQUEIRA, 2018) primeiramente levantou-se os requisitos do sistema, baseadas nas questões sobre qual a solução *Business Intelligence* deseja-se responder. Após o levantamento das perguntas chaves são criadas as regras de negócios do projeto. A modelagem do projeto de DW optou pelo esquema estrela, pois encontraria melhor performance no processamento de dados. O modelo lógico possui uma tabela fato e três tabelas dimensões. Após este processo é feita a etapa ETL e armazenamento dos dados no banco de dados. Para visualização dos indicadores da justiça o autor utiliza a ferramenta Power BI⁸. Com a conclusão do desenvolvimento o relatório foi apresentado a um especialista em tribunais de justiça, e a conclusão do mesmo é que a visualização gráfica é válida e útil, pois as análises feitas são solicitadas constantemente dentro dos tribunais e por advogados. A dificuldade apontada pelo autor do trabalho foi a classificação dos indicadores, pois apesar do CNJ disponibilizar a documentação dos indicadores estes não estavam em um formato compatível que permitisse a combinação com os dados disponibilizados. Como solução foi feito um trabalho manual de preenchimento

⁸Acesse: <https://powerbi.microsoft.com/en-us/what-is-power-bi/>

utilizando a ferramenta Excel.

De (PIZANI; SIQUEIRA, 2018) é possível concluir que criar visualizações e análises para um público alvo vasto é uma atividade complexa, pois é necessário generalizar, todavia é preciso conter informações precisas, claras e importantes, por esse motivo, foi criada a seção de perguntas estratégicas para que o projeto não fugisse do escopo.

Em (VEIGA; GUIMARÃES; TANAKA, 2015) é abordado o uso de técnicas de *Business Intelligence* na análise de dados governamentais. O assunto escolhido foi a eleição de 2014, cujo os dados estão publicados no Portal do Tribunal Superior Eleitoral. O trabalho tem como objetivo demonstrar a viabilidade do uso de técnicas de BI, usando a plataforma aberta Pentaho BI Suite⁹, no auxílio ao entendimento dos Dados Abertos Governamentais, que são disponibilizados em formato bruto, de difícil compreensão pelo cidadão comum.

Em (VEIGA; GUIMARÃES; TANAKA, 2015) é abordado o uso de técnicas de *Business Intelligence* na análise de dados governamentais. O assunto escolhido foi a eleição de 2014, cujo os dados estão publicados no Portal do Tribunal Superior Eleitoral. O trabalho tem como objetivo demonstrar a viabilidade do uso de técnicas de BI, usando a plataforma aberta Pentaho BI Suite, no auxílio ao entendimento dos Dados Abertos Governamentais, que são disponibilizados em formato bruto, de difícil compreensão pelo cidadão comum.

⁹Acesse: <https://www.techtem.com.br/pentaho-software-bi/>

4 Requisitos de Negócios e Modelagem

Neste capítulo é apresentado a modelagem dimensional do projeto de DW. Começando pelas regras de negócio, requisitos funcionais e não funcionais e apresentado o modelo proposto para o *Data Warehouse* de dados eleitorais.

4.1 Requisitos de Negócios

Quando se desenvolve uma solução utilizando BI para esta etapa é necessário responder questões sobre o negócio ou questões específicas que o cliente tenha. As perguntas que costumam ser respondidas para iniciar o levantamento de requisitos são:

- “O que aconteceu ?”
- “Quando aconteceu?”
- “Quantas vezes isso aconteceu?”
- “Onde aconteceu?”
- “Quem fez?”

Após isso, é necessário entender quem é o público alvo. Neste trabalho o alvo são os cidadãos brasileiros, esse público é bem variado, porém possuem em comum o interesse de informação sobre as eleições de forma simples, rápida, intuitiva e comparativa.

Com essa consideração as perguntas estratégicas que o sistema deverá responder são:

- Qual espectro político que possui mais candidatos?
- Qual espectro político que possui mais candidatos eleitos?
- Como é a distribuição de eleitos por espectro político no Brasil após as eleições?
- Qual espectro político possui mais mulheres?

- Qual o espectro político possui mais negros?
- Qual partido possui mais candidatos dentro de um espectro político?
- Qual partido mais elegeu candidatos em seu espectro político?

Para responder esses questionamentos foi necessário estudar os arquivos do repositório eleitoral e identificar quais seriam necessários utilizar. O repositório possui vários arquivos que são classificados por assuntos, são eles: candidatos, comparecimento e ausência, eleitorado, partidos, pesquisas eleitorais, prestação de contas eleitorais, prestação de contas partidárias, processual e resultados. Para este trabalho foi identificado que os arquivos necessários seriam dos assuntos candidatos e resultados. Como o TSE não apresenta uma classificação dos partidos pelo espectro político foi necessário buscar a fonte que poderia fornecer essa classificação para ser integrada com os dados eleitorais.

Após os estudos dos dados foi possível planejar as tabelas dimensões e tabelas fatos necessárias para responder às questões levantadas e assim criar as regras de negócio e especificar os requisitos. As subseções a seguir apresenta o que foi levantado.

4.1.1 Regras de Negócio

A tabela 4.1, detalha as regras de negócios levantadas neste projeto:

REGRA DE NEGÓCIO	DESCRIÇÃO
RN01 - Tabelas dimensões	O DW deve conter as seguintes tabelas dimensões: dm_abrangencia, dm_ano, dm_cargo, dm_cor_raca, dm_eleicao, dm_espectro_politico, dm_estado, dm_genero, dm_municipio, dm_partido, dm_regiao, dm_situacao_totalizacao, dm_tipo_eleicao e dm_turno
RN02 - Tabelas fatos	O DW deve conter as seguintes tabelas fatos: ft_eg_candidaturas, ft_em_candidaturas, ft_eg_candidaturas_perfil, ft_em_candidaturas_perfil
RN03 - Transformação de dados	Os dados devem ser transformados para que se adequem a estrutura do modelo previamente definida
RN04 - Carga de dados	Após a etapa de transformação, os dados estruturados devem ser inseridos no banco de dados.

Tabela 4.1: Regras de Negócios. Fonte: Elaborado pelos autores

4.1.2 Requisitos Funcionais

- O sistema não deve permitir inserção de dados pelo usuário.
- O sistema não deve permitir a remoção de dados pelo usuário final.
- O sistema deve permitir realizar funcionalidades de acesso a visualização e análise de dados.

4.1.3 Requisitos Não Funcionais

- O sistema deve ser responsivo, ou seja, ser adaptado para utilização em aparelhos móveis.
- Deve ser utilizado o banco de dados *MariaDB versão 10.4.11*¹⁰ ou superior, para armazenar os dados.
- A etapa de ETL deve utilizar o *Pentaho Data Integration Community 9.0*¹¹.
- As páginas com os *dashboards* devem ser desenvolvidas com a biblioteca *React*¹² do

¹⁰ Acesse: <https://mariadb.org/about/>

¹¹ Acesse: https://help.pentaho.com/Documentation/9.0/Products/Pentaho_Data_Integration

¹² Acesse: <https://reactjs.org/docs/getting-started.html>

javascript.

- Para visualização dos dados deve ser utilizado a biblioteca *Highcharts*¹³.

4.2 Modelagem Dimensional

Para este projeto de DW o esquema floco de neve foi escolhido para evitar as redundâncias e acrescentar mais detalhes ao modelo.

O modelo lógico possui quatro tabelas fatos e treze tabelas dimensões. As tabelas fatos são diferenciadas pela abrangência da eleição, ou seja, para cada assunto que será analisado foi criada uma tabela fato para as eleições municipais e para as eleições gerais.

Nas subseções seguintes, são apresentados as tabelas dimensões e fato que compõe o modelo dimensional.

4.2.1 Dimensão Abrangência

Esta dimensão possui os atributos que descrevem qual é a abrangência da eleição em um ano e os cargos que podem ser disputados. As abrangências são: Estadual, Federal e Municipal.

TABELA: dm_abrangencia	
COLUNA	DESCRIÇÃO
id	Identificador da abrangência
nome	Descrição da abrangência eleitoral

Tabela 4.2: Tabela dimensão Abrangência. Fonte: Elaborado pelos autores

4.2.2 Dimensão Ano

A dimensão Ano contém o atributo ano de referência da eleição e a chave estrangeira abrangência. Esta dimensão é utilizada como filtro de pesquisa pelo usuário final.

¹³Acesse: <https://www.highcharts.com/docs/getting-started/system-requirements>

TABELA: dm_ano	
COLUNA	DESCRIÇÃO
id	Identificador do ano de referência da a eleição
abrangencia_id	Abrangência do ano eleitoral

Tabela 4.3: Tabela dimensão Ano. Fonte: Elaborado pelos autores

4.2.3 Dimensão Cargo

A dimensão Cargo possui os atributos para descrever os cargos para os quais os candidatos concorrem nas eleições.

TABELA: dm_cargo	
COLUNA	DESCRIÇÃO
id	Identificador do cargo
nome	Descrição do cargo a que o candidato concorre
abrangencia_id	Abrangência do cargos que o candidato concorre

Tabela 4.4: Tabela dimensão Cargo. Fonte: Elaborado pelos autores

4.2.4 Dimensão Tipo de Eleição

Esta dimensão possui atributos que descrevem as características da eleição. Esta dimensão pode assumir os seguintes valores: Eleição Ordinária, Eleição Suplementar e Consulta Popular.

Segundo o TSE as eleições ordinárias são previstas em lei, ocorre periodicamente de quatro em quatro anos. As eleições suplementares são aquelas que não têm periodicidade e ocorrem quando, eventualmente, se fizerem necessárias. As consultas populares ocorrem sempre que a população é convocada para opinar sobre um assunto específico, um exemplo foi o referendo de armas realizado em 2005, (ELEITORAIS, 2020).

A dimensão é utilizada como filtro para consulta pelo usuário final.

TABELA: dm_tipo_eleicao	
COLUNA	DESCRIÇÃO
id	Identificador do tipo de eleição
nome	Nome do tipo de eleição

Tabela 4.5: Tabela dimensão Tipo de Eleição. Fonte: Elaborado pelos autores

4.2.5 Dimensão Turno

Esta dimensão contém os atributos que definem o turno eleitoral. No sistema é utilizada como filtro de pesquisa.

Segundo (ELEITORAIS, 2020), no Brasil as eleições ocorrem por meio de dois sistemas: majoritário e sistema proporcional. O sistema majoritário consiste em declarar eleito o candidato que tenha recebido a maioria dos votos válidos e em caso de não se ter a maioria absoluta dos votos válidos no primeiro turno, cinquenta e um por cento dos votos válidos, haverá um segundo turno com os dois candidatos mais votados, este sistema é utilizado para os cargos de governadores, presidente e prefeito. O sistema proporcional é aplicado aos cargos de deputado, vereador e senador.

TABELA: dm_turno	
COLUNA	DESCRIÇÃO
id	Identificador do turno
nome	Nome do turno

Tabela 4.6: Tabela dimensão Turno. Fonte: Elaborado pelos autores

4.2.6 Dimensão Eleição

A dimensão eleição possui os atributos para descrever as eleições. Esta dimensão substitui a dimensão tempo, pois descreve quando aconteceu e como foi. Ela é utilizada pelo usuário final como filtro de pesquisa.

TABELA: dm_eleicao	
COLUNA	DESCRIÇÃO
id	Identificador da eleição
ano_id	Identificador do ano que ocorreu a eleição
tipo_eleicao_id	Identificador do tipo da eleição
turno_id	Identificador do turno
estado_id	Identificador da unidade federativa que ocorreu a eleição. Utilizada quando a eleição é suplementar.
municipio_id	Identificador do município que ocorreu a eleição. Utilizada quando a eleição é suplementar.
codigo_tse	Código único que identifica a eleição no âmbito da Justiça Eleitoral
nome	Descrição da eleição

Tabela 4.7: Tabela dimensão Eleição. Fonte: Elaborado pelos autores

4.2.7 Dimensão Região

A dimensão Região possui atributos que descreve os agrupamentos dos estados brasileiros. Ela é utilizada pelo usuário final como filtro de pesquisa.

TABELA: dm_regiao	
COLUNA	DESCRIÇÃO
id	Identificador da região
nome	Nome do região

Tabela 4.8: Tabela dimensão Região. Fonte: Elaborado pelos autores

4.2.8 Dimensão Estado

A dimensão Estado possui atributos que descrevem as unidades federativas do Brasil. Ela é utilizada pelo usuário final como filtro de pesquisa.

TABELA: dm_estado	
COLUNA	DESCRIÇÃO
id	Identificador do estado (Sigla da unidade da federação)
regiao_id	Identificador da região a qual pertence o estado
nome	Nome do estado

Tabela 4.9: Tabela dimensão Estado. Fonte: Elaborado pelos autores

4.2.9 Dimensão Município

A dimensão Município possui atributos que descrevem as cidades do Brasil. Ela é utilizada pelo usuário final como filtro de pesquisa.

TABELA: dm_municipio	
COLUNA	DESCRIÇÃO
id	Identificador do município
estado_id	Identificador do estado que o município pertence
codigo_tse	Número que identifica o município no sistema do TSE
nome	Nome do município

Tabela 4.10: Tabela dimensão Município. Fonte: Elaborado pelos autores

4.2.10 Dimensão Espectro Político

Esta dimensão possui os atributos que descrevem os espectros políticos que classificam os partidos políticos. Esses espectros são comentados no capítulo dois e pode ser utilizado como filtro de pesquisa pelo usuário final.

TABELA: dm_espectro_politico	
COLUNA	DESCRIÇÃO
id	Identificador do espectro político
nome	Nome do espectro
cor	Número hexadecimal que representa a cor do espectro político

Tabela 4.11: Tabela dimensão Espectro Político. Fonte: Elaborado pelos autores

4.2.11 Dimensão Partido

Esta dimensão possui os atributos que descrevem os partidos e as legendas concorrentes às vagas.

TABELA: dm_partido	
COLUNA	DESCRIÇÃO
id	Identificador do partido
espectro_id	Identificador do espectro político que o partido pertence
numero	Número do partido
sigla	Sigla da legenda
nome	Nome da legenda
cor	Número hexadecimal que representa a cor do partido

Tabela 4.12: Tabela dimensão Partido. Fonte: Elaborado pelos autores

4.2.12 Dimensão Cor/Raça

Esta dimensão possui os atributos que descrevem a cor/raça que o candidato declara ao se cadastrar para concorrer às eleições.

TABELA: dm_cor_raca	
COLUNA	DESCRIÇÃO
id	Identificador da cor/raça
codigo_tse	Código para identificação no sistema do TSE
nome	Nome da cor/raça

Tabela 4.13: Tabela dimensão Cor/Raça. Fonte: Elaborado pelos autores

4.2.13 Dimensão Gênero

Esta dimensão possui os atributos que descrevem os gêneros que o candidato declara ao se inscrever no processo eleitoral.

TABELA: dm_genero	
COLUNA	DESCRIÇÃO
id	Identificador do gênero
codigo_tse	Código para identificação no sistema do TSE
nome	Nome do gênero

Tabela 4.14: Tabela dimensão Gênero. Fonte: Elaborado pelos autores

4.2.14 Dimensão Situação de Totalização

Esta dimensão possui os atributos que descrevem a situação do candidato no turno da eleição após a totalização dos votos.

TABELA: dm_situacao_totalizacao	
COLUNA	DESCRIÇÃO
id	Identificador da situação
codigo_tse	Código para identificação no sistema do TSE
nome	Descrição da situação

Tabela 4.15: Tabela dimensão Situação de Totalização. Fonte: Elaborado pelos autores

4.2.15 Fato Candidaturas

Os fatos possuem as medidas sobre os candidatos e a sua situação após o turno da eleição. Como dito anteriormente existe a tabela fato para as eleições gerais¹⁴ e outra para as eleições municipais¹⁵. A diferença entre ambas se encontra na chave estrangeira que representa a dimensão localidade, onde para primeira tabela é utilizada a dimensão estado e na segunda a dimensão município.

TABELA: ft_eg_candidaturas	
COLUNA	DESCRIÇÃO
eleicao_id	Chave estrangeira para a tabela de dimensão eleição
estado_id	Chave estrangeira para a tabela de dimensão estado
partido_id	Chave estrangeira para a tabela de dimensão partido
cargo_id	Chave estrangeira para a tabela de dimensão cargo
situacao_totalizacao_id	Chave estrangeira para a tabela de dimensão situação de totalização
quantidade	Quantidade de candidatos

Tabela 4.16: Eleições Gerais - Tabela Fato Candidaturas. Fonte: Elaborado pelos autores

A figura 4.1 mostra o modelo lógico para este fato.

¹⁴Nas eleições gerais os seguintes cargos são disputados: Governador, Vice-Governador, Deputado Estadual, Deputado Federal, Deputado Distrital, Senador, Senador - 1º Suplente, Senador - 2º Suplente, Presidente e Vice-Presidente

¹⁵Nas eleições municipais os cargos disputados são: Prefeito, Vice-Prefeito e Vereador

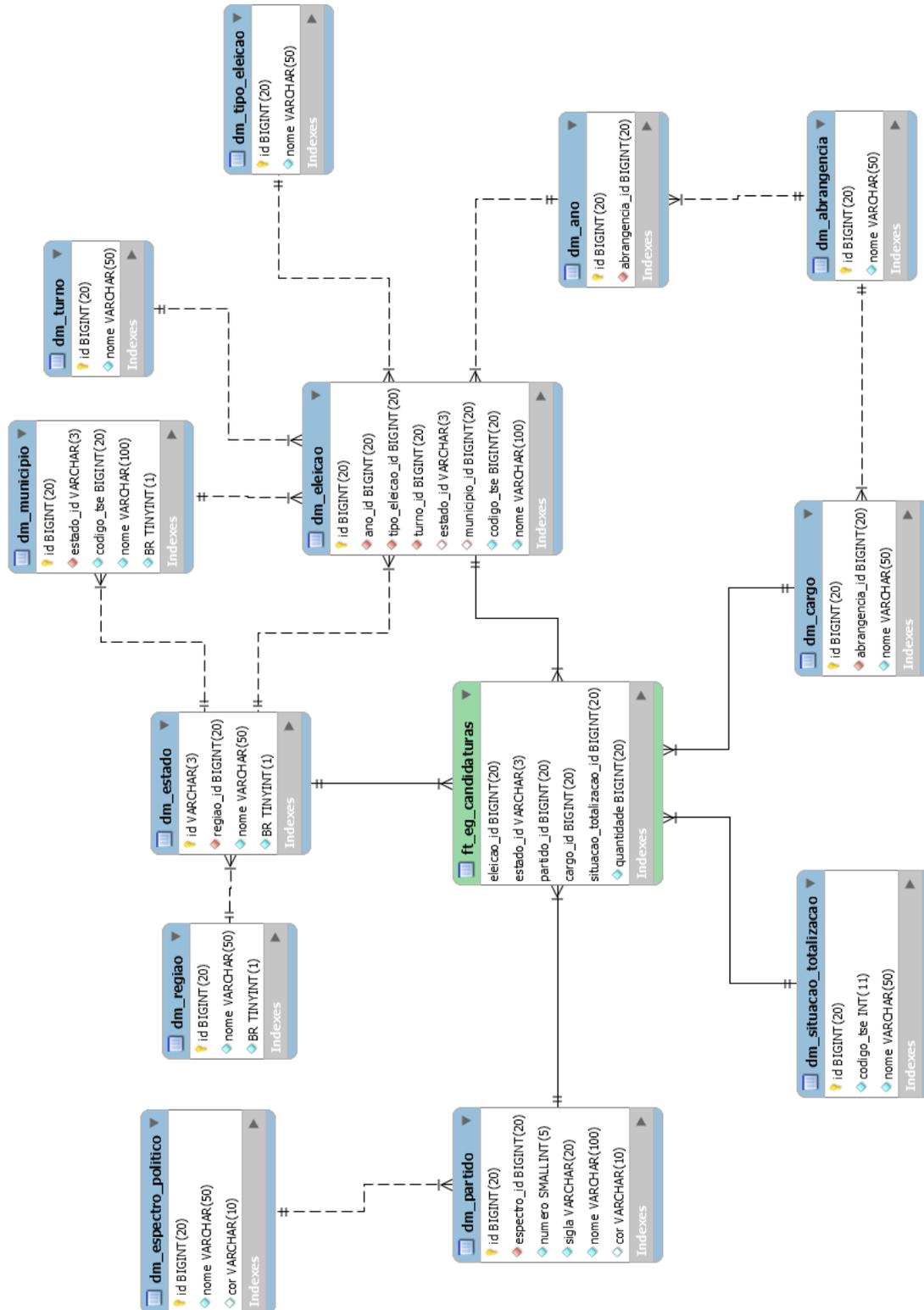


Figura 4.1: Eleições Gerais - Modelo Lógico Fato Candidaturas. Fonte: Elaborado pelos autores

TABELA: ft_em_candidaturas	
COLUNA	DESCRIÇÃO
eleicao_id	Chave estrangeira para a tabela de dimensão eleição
municipio_id	Chave estrangeira para a tabela de dimensão município
partido_id	Chave estrangeira para a tabela de dimensão partido
cargo_id	Chave estrangeira para a tabela de dimensão cargo
situacao_totalizacao_id	Chave estrangeira para a tabela de dimensão situação de totalização
quantidade	Quantidade de candidatos

Tabela 4.17: Eleições Municipais - Tabela Fato Candidaturas. Fonte: Elaborado pelos autores

A figura 4.2 mostra o modelo lógico para este fato.

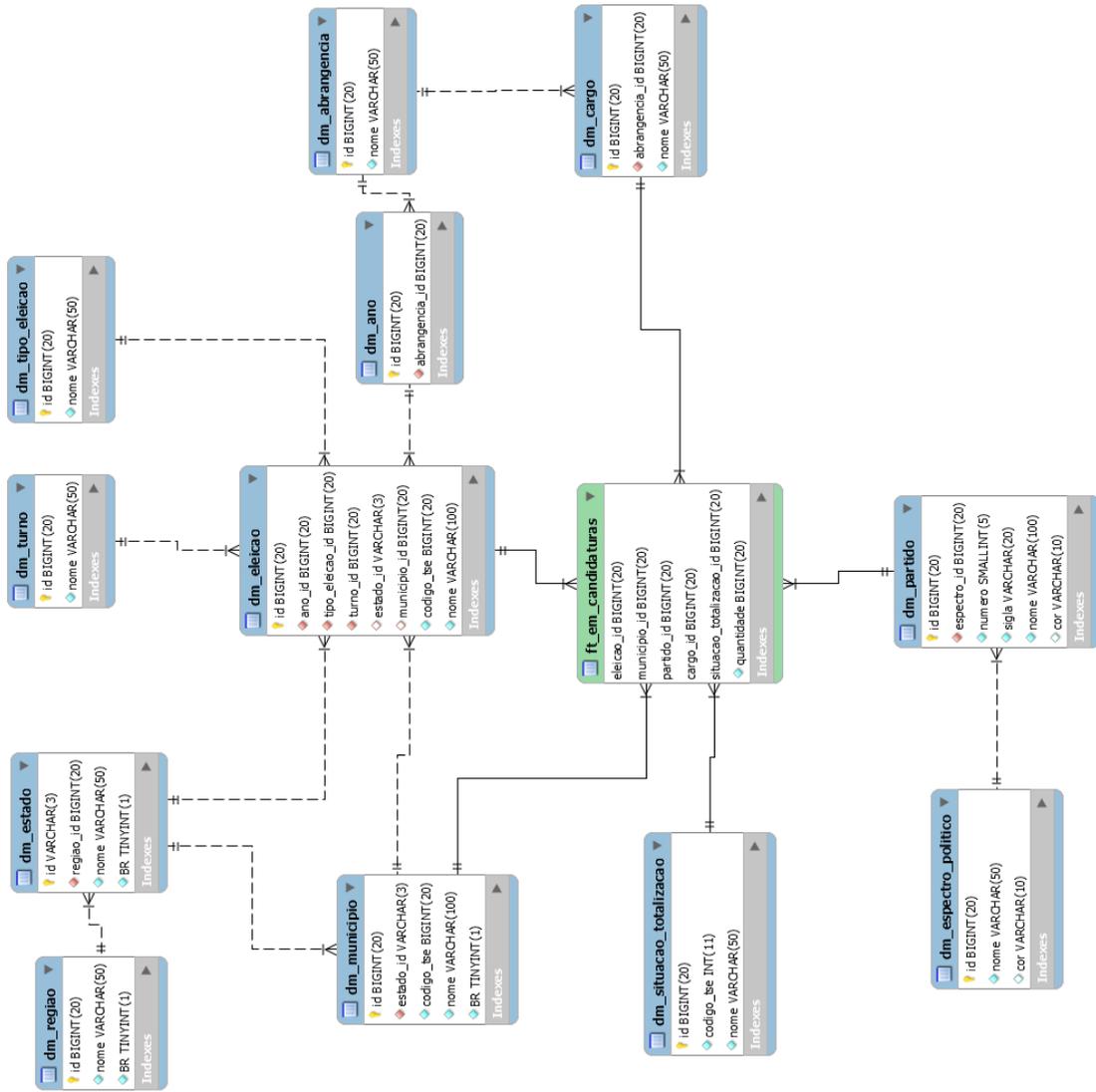


Figura 4.2: Eleições Municipais - Modelo Lógico Fato Candidaturas. Fonte: Elaborado pelos autores

4.2.16 Fato Perfil Candidaturas

Estes fatos possuem as medidas sobre os perfis dos candidatos são eles: gênero e cor/raça. Para a tabela fato que representa as eleições gerais a dimensão estado é a chave estrangeira que representa a localidade. Já na tabela fato para as eleições municipais a localidade é representada pela chave estrangeira dimensão município.

TABELA: ft_eg_candidaturas_perfil	
COLUNA	DESCRIÇÃO
eleicao_id	Chave estrangeira para a tabela de dimensão eleição
estado_id	Chave estrangeira para a tabela de dimensão estado
partido_id	Chave estrangeira para a tabela de dimensão partido
cargo_id	Chave estrangeira para a tabela de dimensão cargo
situacao_totalizacao_id	Chave estrangeira para a tabela de dimensão situação de totalização
cor_raca_id	Chave estrangeira para a tabela de dimensão cor/raça
genero_id	Chave estrangeira para a tabela de dimensão gênero
quantidade	Quantidade de candidatos

Tabela 4.18: Eleições Gerais - Tabela Fato Perfil Candidaturas. Fonte: Elaborado pelos autores

A figura 4.3 mostra o modelo lógico para este fato.

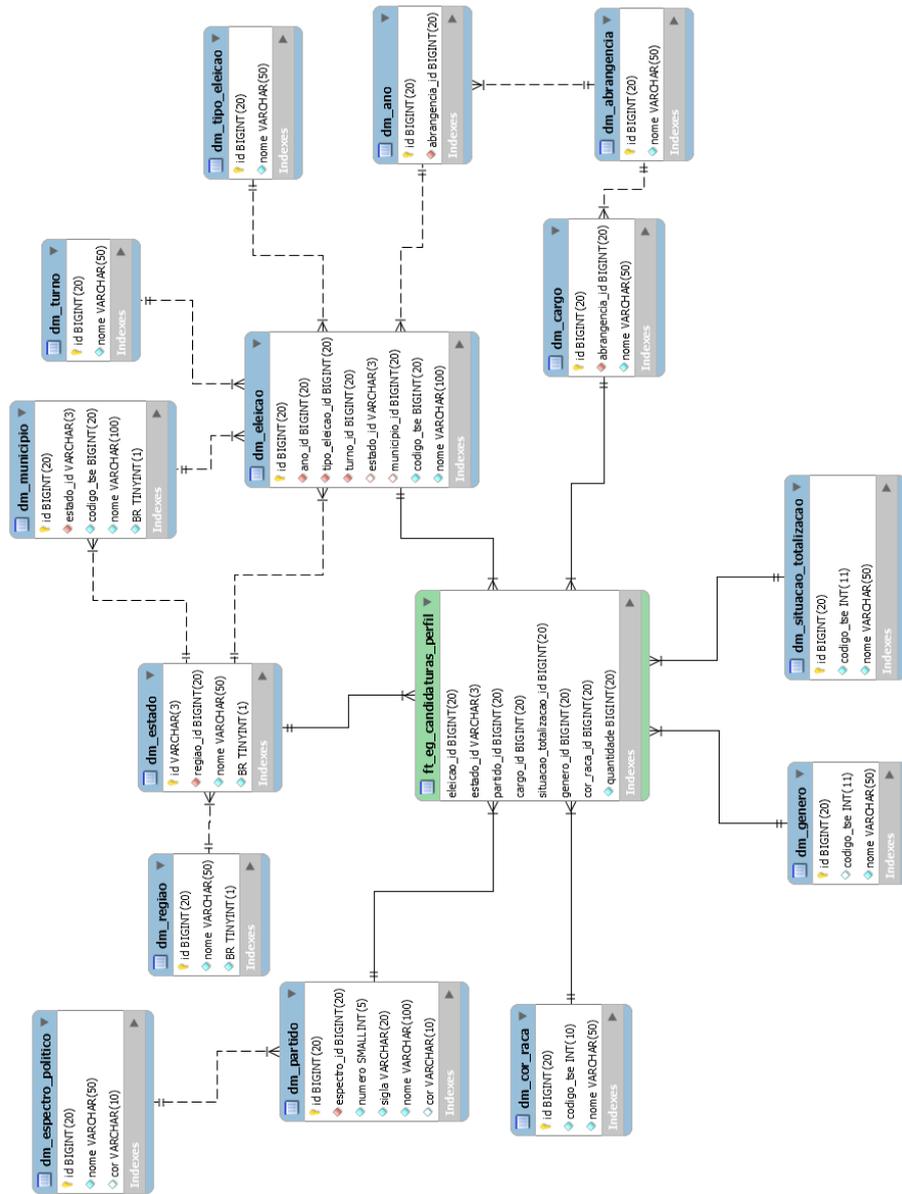


Figura 4.3: Eleições Gerais - Modelo lógico Fato Perfil Candidaturas. Fonte: Elaborado pelos autores

TABELA: ft_em_candidaturas_perfil	
COLUNA	DESCRIÇÃO
eleicao_id	Chave estrangeira para a tabela de dimensão eleição
municipio_id	Chave estrangeira para a tabela de dimensão município
partido_id	Chave estrangeira para a tabela de dimensão partido
cargo_id	Chave estrangeira para a tabela de dimensão cargo
situacao_totalizacao_id	Chave estrangeira para a tabela de dimensão situação de totalização
cor_raca_id	Chave estrangeira para a tabela de dimensão cor/raça
genero_id	Chave estrangeira para a tabela de dimensão gênero
quantidade	Quantidade de candidatos

Tabela 4.19: Eleições Municipais - Tabela Fato Perfil Candidaturas. Fonte: Elaborado pelos autores

A figura 4.4 mostra o modelo lógico para este fato.

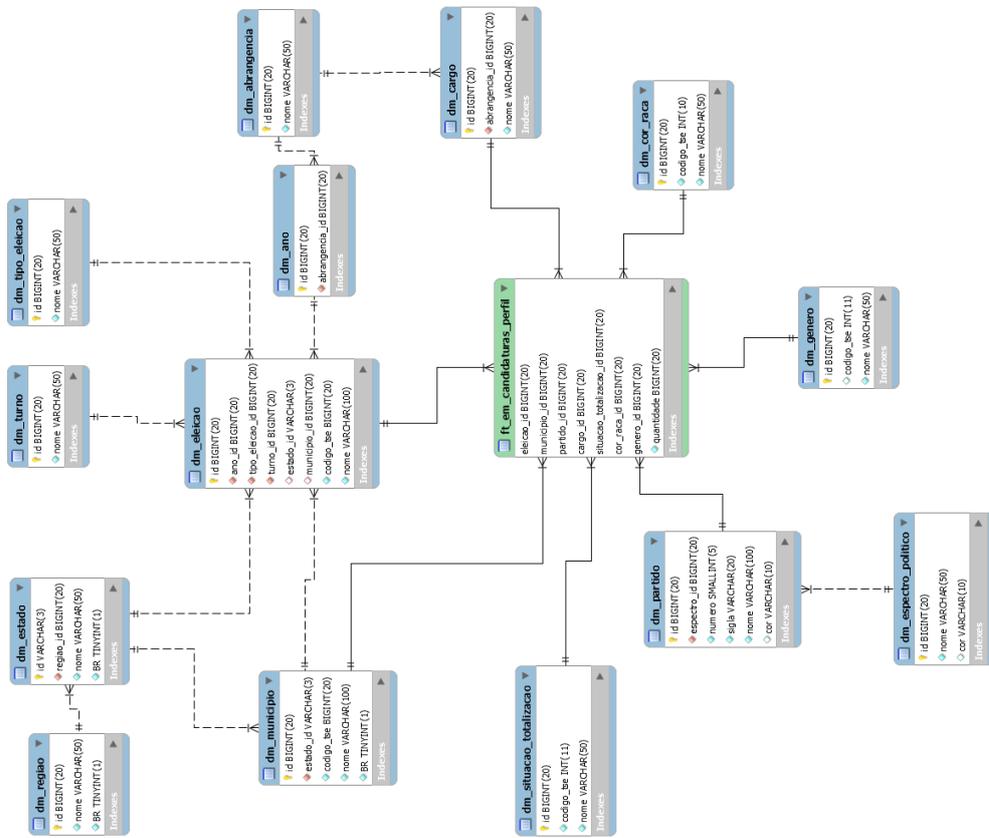


Figura 4.4: Eleições Municipais - Modelo lógico Fato Perfil Candidaturas. Fonte: Elaborado pelos autores

5 Desenvolvimento

Neste capítulo é apresentado as ferramentas utilizadas, o processo de desenvolvimento e resultados obtidos. Este capítulo foi fundamentado a partir dos estudos feitos nos capítulos anteriores. Com base nos estudos de BI foram pensadas as melhores práticas para a criação do sistema.

5.1 Ferramentas Utilizadas

Nesta seção serão expostas as ferramentas utilizadas durante o desenvolvimento, seja para armazenamento, transformação, extração, cargas, modelagem e visualização.

5.1.1 Banco de Dados MariaDB

MariaDB¹⁶ é um gerenciador de banco de dados relacional bastante popular. É mantido pelos desenvolvedores do MySQL e tem garantia de código aberto, além de fornecer uma interface SQL para acesso dos dados. Projetado como substituto do MySQL ele é utilizado, pois é rápido, escalável e robusto. Possui um rico ecossistema de mecanismo de armazenamento, plugins e muitas outras ferramentas que o tornam versátil para uma ampla variedade de casos de uso.

5.1.2 Pentaho Data Integration (PDI)

O PDI¹⁷, também conhecido como Kettle, faz parte da suíte da Pentaho. Esta aplicação provê recursos de extração, transformação e carregamento facilitando a limpeza, padronização e armazenamento dos dados para a utilização por demais tecnologias. Neste trabalho é utilizada a versão *Community*, que é gratuita, porém possui os mesmos recursos da versão paga.

O Kettle é escrito em Java, open source, possui uma documentação atualizada e

¹⁶ Acessar: <https://mariadb.org/about/>

¹⁷ Acessar: https://help.pentaho.com/Documentation/9.1/Products/Pentaho_Data_Integration

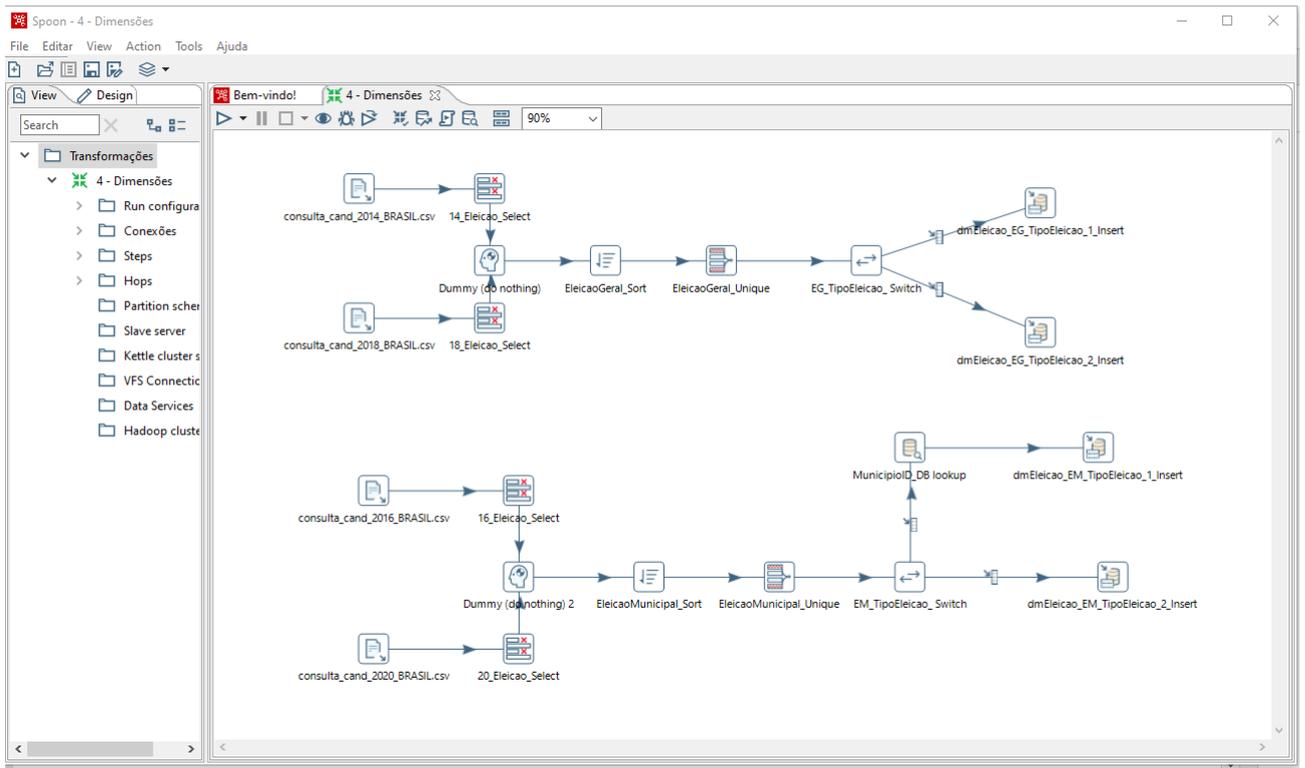


Figura 5.1: Pentaho Data Integration. Fonte: Elaborado pelos autores

comunidade ativa. Existem três aplicações diferentes, são elas:

- *Spoon* - Interface gráfica que permite projetar o fluxo de dados através de *steps* de entrada de dados passando pela transformação e gerando a saída dos dados.
- *Pan* - Esta ferramenta executa as transformações do *Spoon* através de linha de comando, normalmente a sua execução é agendada e executada por batch.
- *Kitchen* - Ferramenta utilizada para execução de *Jobs* criados pelo *Spoon* através de linha de comando.

A transformação é uma sequência lógica de tarefas em um fluxo de dados, cada passo é chamado de *step* e as ligações de *hop* e é executado de forma paralela. O *Job* atua de forma diferente da transformação, ele é executado de forma sequencial sendo possível combinar uma sequência de transformações para serem executadas, além disso é possível criar armazenamento de *logs* ou mandar um email para o responsável comunicando o sucesso ou a falha na transformação.

5.1.3 APIs

O acrônimo API vem do inglês “*Application Programming Interface*” que traduzido significa Interface de Programação de Aplicações. Segundo (PIRES, 2017) é um conjunto de rotinas e padrões de programação para acesso a um aplicativo de software ou plataforma baseado na Web que permite o acesso de outras aplicações para utilização de suas funcionalidades, não precisando conhecer os detalhes da implementação.

O sistema web proposto utiliza uma API para realizar as consultas no DW e as páginas desenvolvidas fazem requisições a esta API para ter acesso aos dados e criar os *dashboards*. Esta separação permite que outras aplicações possam fazer uso do *Data Warehouse*.

5.1.4 Laravel e Lumen

O Laravel¹⁸ é um framework, open source sob a licença MIT, de aplicação web desenvolvido com a linguagem PHP (*Hypertext Preprocessor*) com uma sintaxe expressiva e elegante. Ele busca facilitar o desenvolvimento de tarefas comuns usadas em muitos projetos da web. Utiliza o padrão de projeto Model, View e Controller (MVC) e hoje é amplamente utilizado.

O Lumen¹⁹ é um micro-framework derivado do Laravel que possui uma quantidade de recursos limitada em relação ao Laravel, porém permite a adição de novos recursos de terceiros. Ele foi criado para dar suporte ao desenvolvimento de APIs e microsserviços. Apresenta maior rapidez no desenvolvimento uma vez que o desenvolvedor precisa se preocupar apenas com o necessário para o seu projeto. Tende a ser mais leve devido a sua menor estrutura se comparada com o seu pai. Permite também a criação de aplicações mais robustas.

Neste projeto optamos pelo Lumen para a criação da API, pois como será utilizado outras bibliotecas para construção da camada de visualização prezamos por uma estrutura mais leve.

¹⁸Acessar: <https://laravel.com/>

¹⁹Acessar: <https://lumen.laravel.com/>

5.1.5 *React*

*React*²⁰ é uma biblioteca desenvolvida em JavaScript para construção de interfaces de usuário. A biblioteca permite que a criação de UIs²¹ seja uma tarefa fácil. As *views* são criadas de maneira simples e o *React* atualiza e renderiza de forma eficiente apenas os componentes necessários na medida em que os dados mudam. É baseada em componentes, ou seja, a lógica é escrita em JavaScript e não em templates, dessa forma é possível passar tipos de dados ao longo da sua aplicação e ainda manter o estado fora do DOM²².

Neste trabalho a biblioteca é utilizada para a construção dos *dashboards*, fazendo constantes requisições na API do *data warehouse* para consultas e atualizações dos filtros e gráficos, além da estruturação de todas as páginas do sistema web.

5.1.6 *Highcharts*

O *Highcharts*²³ é uma biblioteca desenvolvida JavaScript que permite a criação de diversos tipos de gráficos. Possui adaptadores para outras tarefas, como a exportação dos dados, que podem ser incluídas conforme a necessidade dos usuários. A biblioteca foi criada em 2009 a partir da dificuldade de um programador em criar gráficos com plugins de Flash e outras soluções proprietárias. Os gráficos são gerados a partir de dados que são convertidos no formato JSON (*JavaScript Object Notation*). Com isso, as mais diversas linguagens para criação de páginas podem reconhecer e executar os códigos dos gráficos.

Os gráficos são gerados com qualidade e rapidez mesmo quando se está acessando um banco de dados. A estrutura do *Highcharts* é simples, mesmo quando se trabalha com consultas OLAP. Pelas considerações citadas a biblioteca se mostrou interessante para a utilização no projeto.

²⁰Acessar: <https://pt-br.Reactjs.org/>

²¹A sigla UI significa *User Interface*, em português, Interface do Usuário. Acessar: <https://searcharchitecture.techtarget.com/definition/user-interface-UI>

²²DOM (*Document Object Model*) é a representação de dados dos objetos que compõem a estrutura e o conteúdo de um documento na Web. Acessar: https://developer.mozilla.org/en-US/docs/Web/API/Document_Object_Model/Introduction

²³Acessar: <https://www.highcharts.com/products/highcharts/>

5.2 Experimento

De acordo com a problemática e conforme o objetivo, ambos descritos no capítulo 1, este trabalho consiste na criação de uma solução de BI com as fontes de dados disponibilizadas no repositório eleitoral pelo TSE. O processo de desenvolvimento pode ser resumido na tabela abaixo.

PASSO	DESCRIÇÃO
1	O TSE trata os dados e prepara os arquivos referentes às eleições no repositório.
2	Os dados são disponibilizados ao público no site do repositório em formato <i>.csv</i> para <i>download</i> .
3	A modelagem lógica do DW é feita.
4	O banco de dados é criado e as tabelas dimensões e fatos são criadas.
5	Os dados passam por um tratamento prévio e são incluídos no banco de dados utilizado na ferramenta <i>Pentaho Data Integration</i> .
6	As consultas aos fatos são criadas com micro framework <i>Lumen</i> e disponibilizadas a partir das URLs.
7	Os <i>dashboards</i> são criados com os dados disponíveis.
8	Usuários finais utilizam os <i>dashboards</i> .

Tabela 5.1: Etapas do desenvolvimento. Fonte: Elaborada pelos autores

5.2.1 Arquivos

Por se tratar de várias bases, é possível derivar diversas análises pertinentes. A escolha dos arquivos foi baseada nas perguntas estratégicas levantadas no capítulo 4. Logo os arquivos escolhidos foram:

NOME DO ARQUIVO	FONTE	DESCRIÇÃO
consulta_cand_ANO_ELEIÇÃO_BRASIL.csv	TSE	Possui os dados de todos os candidatos às eleições em todo o território nacional.
votacao_candidato_munzona_ANO_ELEIÇÃO_BRASIL.csv	TSE	Possui os dados da votação para todos os candidatos do país.
Estados.csv	Autores	Arquivo com os dados das unidades federativas.
Regiões.csv	Autores	Contém os dados das regiões territoriais do Brasil
Região x Estado.csv	Autores	Arquivo com dos estados e suas regiões.
Abrangência.csv	Autores	Contém os dados da abrangência territorial das eleições
Ano.csv	Autores	Contém os anos das eleições associado a abrangência.
Cargo.csv	Autores	Contém os cargos eleitorais associado com a sua abrangência.
Espectro Político.csv	TSE e Autores	Possui os espectros políticos para classificação dos partidos políticos.
Partidos.csv	TSE Autores	Partidos políticos brasileiros associados ao espectro político.
Tipo de Eleição.csv	TSE	Tipos de eleições do Brasil
Turno.csv	TSE	Contém as descrições dos turnos eleitorais.

Tabela 5.2: Arquivos utilizados. Fonte: Elaborado pelos autores

5.2.2 Processo ETL

Os dados são disponibilizados em sua forma primária bruta, a fim de permitir o maior número possível de levantamento e análises. Portanto, podem surgir diversas agregações e derivações a partir desses dados dependendo somente do utilizador. Para cada arquivo CSV disponibilizado pelo TSE existe um PDF com a descrição de cada coluna. Neste trabalho foram utilizados os arquivos das eleições dos anos 2014, 2016, 2018 e 2020.

Foi escolhido o sistema de gerência de banco de dados MariaDB para armazenar as tabelas resultantes da aplicação, como mostra a conexão Brasil Aberto na figura 5.2.

O PDI foi utilizado para tratar os dados e popular o banco de dados. A carga do DW foi dividida em duas fases: na primeira fase foi realizada a carga das tabelas dimensões. Como as tabelas dimensões possuem relacionamentos, foi necessário dividir a carga em quatro etapas para que fosse totalmente concluída.

Após as dimensões foi feita a carga das tabelas fatos. Devido a limitação de memória essa carga foi dividida pelos anos das eleições, ou seja, para cada ano eleitoral foi criado uma transformação. Na figura 2.4 é mostrado o exemplo da carga.

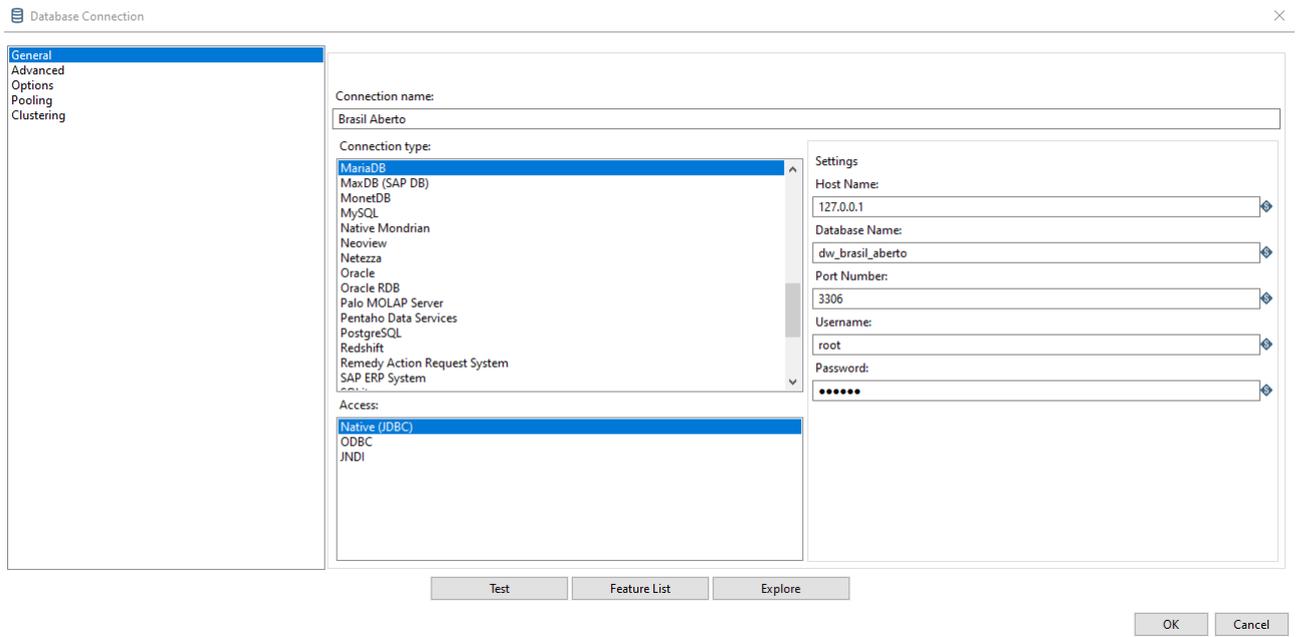


Figura 5.2: Conexão do PDI com o MariaDB. Fonte: Elaborado pelos autores

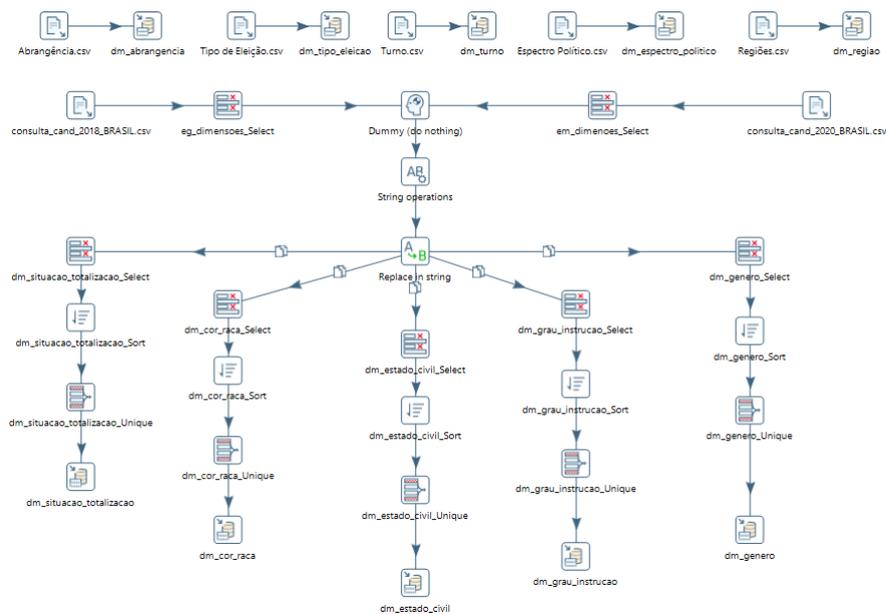


Figura 5.3: Exemplo da carga de dimensões. Fonte: Elaborado pelos autores

5.2.3 Visualização dos dados

Conforme explicado nas seções anteriores para a visualização dos dados foi desenvolvido um sistema web. O processo de desenvolvimento do sistema pode ser melhor explicado conforme a figura 5.5.

A aplicação cliente é a que o usuário final tem acesso. Foi desenvolvida com o *React* e utiliza a biblioteca *Highchartse* para exibir os resultados das consultas analíticas.

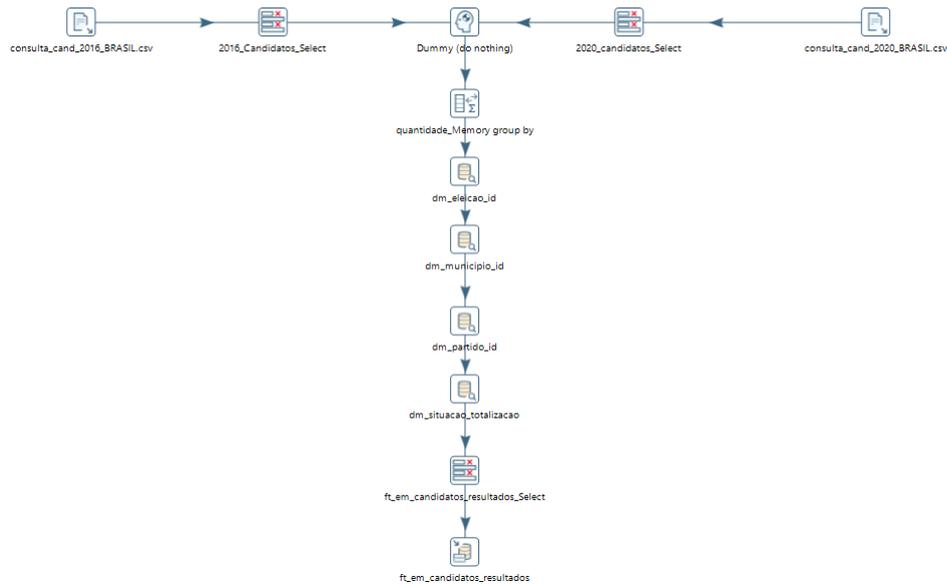


Figura 5.4: Exemplo da carga dos fatos. Fonte: Elaborado pelos autores

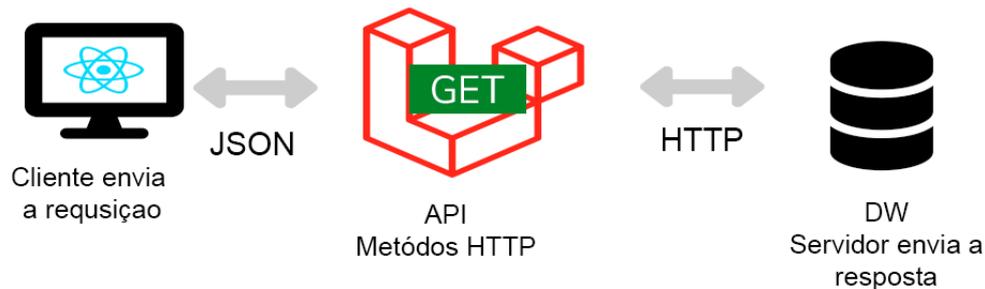


Figura 5.5: Processo de desenvolvimento. Fonte: Elaborado pelos autores

Esta camada comunica com a API através do método GET e recebe como resposta os dados em formato JSON.

A API por sua vez consulta o DW para criar os resultados JSON. Ela recebe os filtros da pesquisa que são podem ser as seguintes dimensões: Ano, Região, Estado, Município, Turno, Tipo de Eleição, Eleição e Espectro Político. Após a realização da consulta, ela prepara uma coleção com os dados obtidos e devolve a resposta ao cliente. Para o desenvolvimento foi utilizado o framework *Lumen*, pois este oferece um conjunto de bibliotecas internas que facilitam o desenvolvimento de APIs.

A figura 5.6 ilustra a tela do sistema na versão *desktop* e a figura 5.7 ilustra a tela na versão *mobile*.

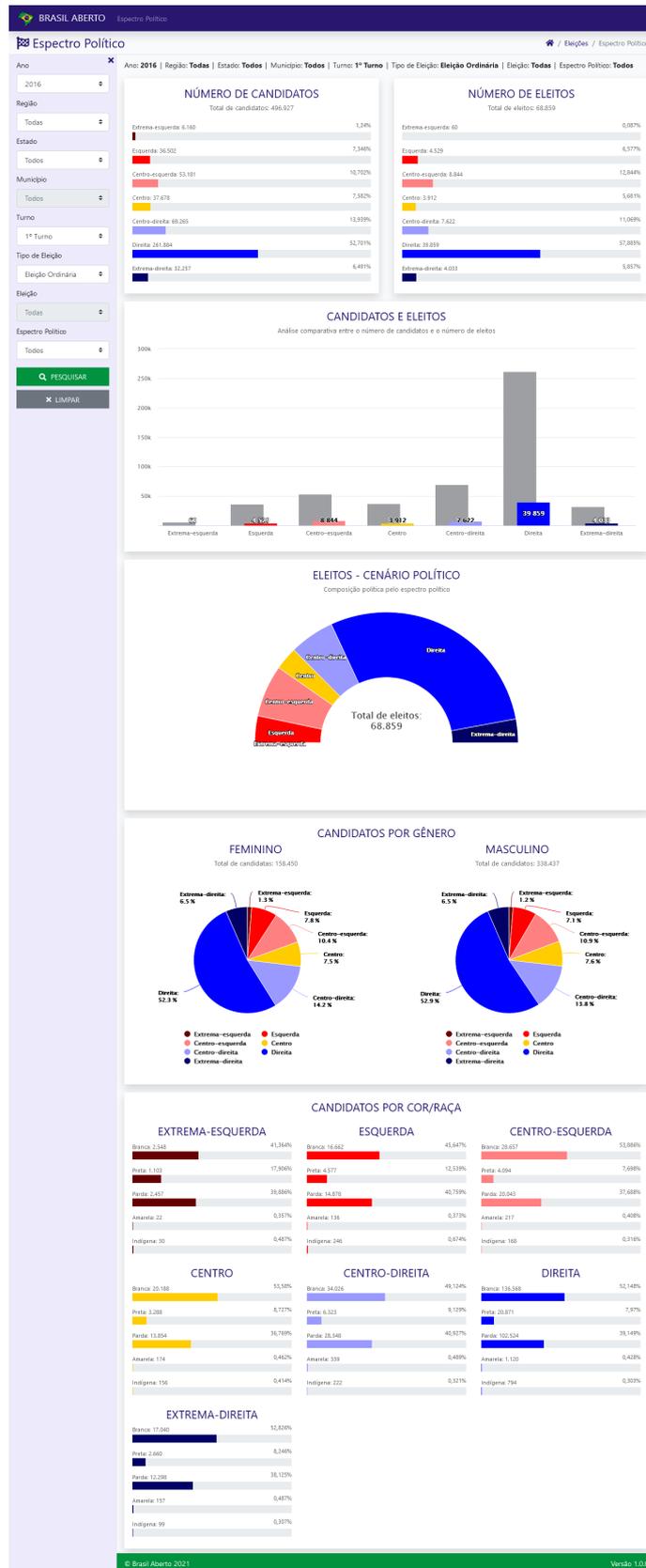


Figura 5.6: Tela do sistema versão *desktop*. Fonte: Elaborado pelos autores



Figura 5.7: Tela do sistema versão *mobile*. Fonte: Elaborado pelos autores

Por padrão, os dados exibidos são sempre do último ano de eleição. Todos os filtros de consulta interagem com os elementos gráficos. Um detalhe que foi criado é com relação a seleção do filtro Espectro Político, pois quando este é selecionado há um detalhamento do espectro selecionado e os gráficos mudam conforme é mostrado na figura 5.8.

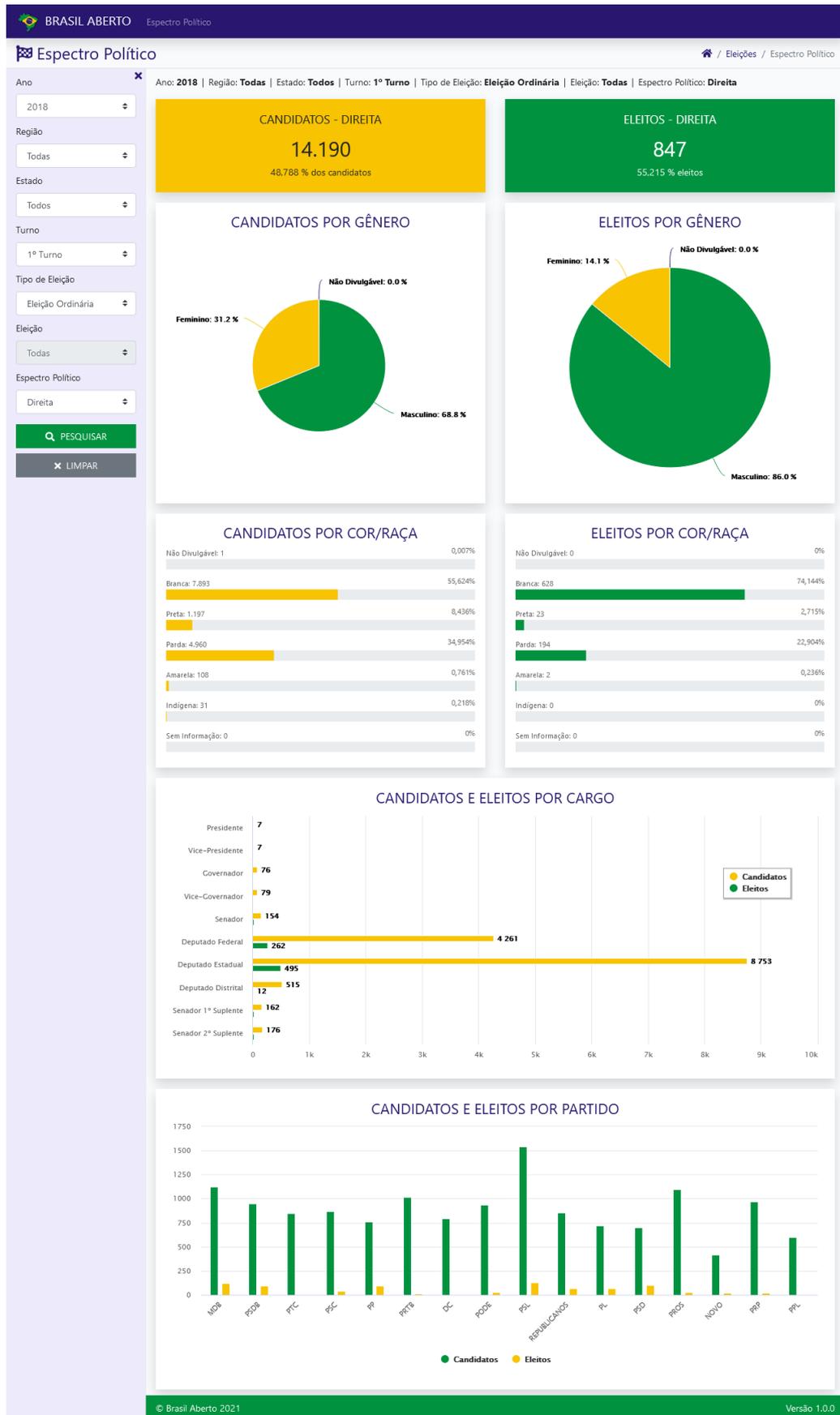


Figura 5.8: Tela do sistema dashboards do espectro político. Fonte: Elaborado pelos autores

6 Resultados

No presente capítulo será apresentado os *dashboards* criados. O objetivo foi elaborar *dashboards* que possam trazer informações novas e úteis, as quais não podem ser encontradas atualmente no portal de estatísticas eleitorais do TSE²⁴. Um ponto importante a ser observado é que os filtros de consultas são os mesmos utilizados pela página “Candidaturas” do portal, porém foi acrescentado o filtro Espectro Político.

Existem dois cenários relacionados a escolha do filtro, o primeiro é sem a seleção do filtro Espectro Político e o segundo cenário é com o este filtro selecionado. As próximas seções apresentam os resultados de cada cenário.

6.1 *Dashboards* sem o filtro Espectro Político

Considerando o cenário sem a escolha do Espectro Político para realização das buscas foram criados seis *dashboards* apresentados nas subseções que seguem.

Os filtros utilizados para obter os resultados apresentados na próximas subseções são:

- Ano: 2020
- Região: Todas
- Estado: Todos,
- Turno: 1º Turno
- Tipo de Eleição: Eleição Ordinária
- Eleição: Todas
- Espectro Político: Todos

²⁴Acessar: <https://www.tse.jus.br/eleicoes/estatisticas/estatisticas-eleitorais>

6.1.1 Número de Candidatos

Este gráfico busca informar a quantidade e percentual de candidatos pelos espectros políticos. Com ele é possível analisar qual espectro possui mais candidatos nas eleições. Este *dashboard* responde a pergunta: “Qual espectro político que possui mais candidatos?”.

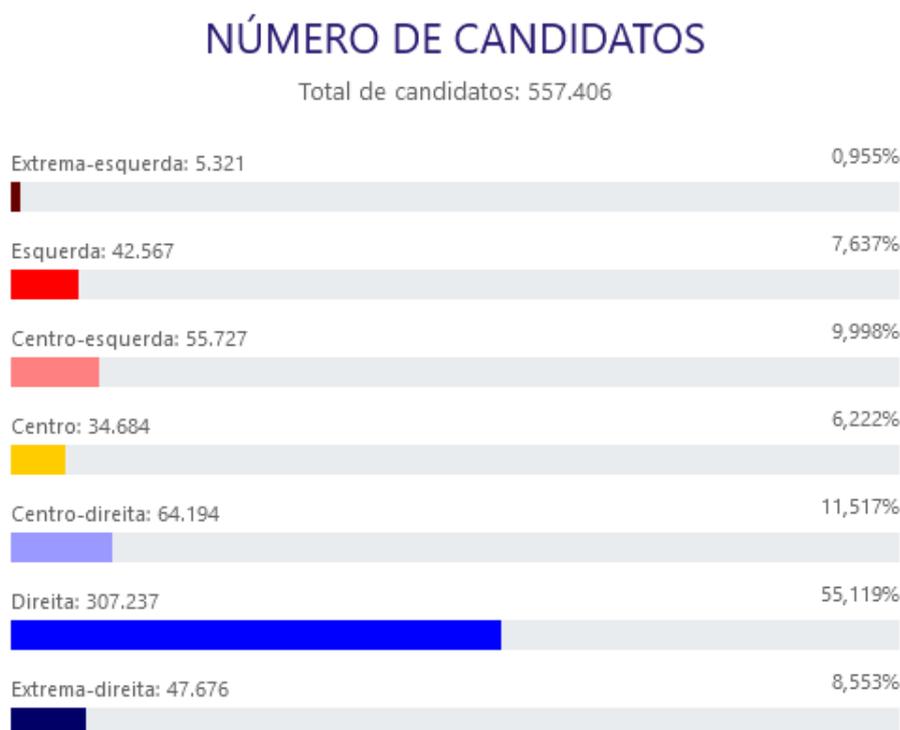


Figura 6.1: Gráfico Número de candidatos. Fonte: Elaborado pelos autores

A utilização dos filtros de pesquisa revela que nas eleições municipais de 2020 o espectro político com mais candidatos é a Direita. Mostra também que a Extrema-Esquerda é o espectro com menor número de candidatos.

6.1.2 Número de Eleitos

O gráfico Número de eleitos apresenta o quantitativo dos candidatos eleitos por espectro político. Esta análise busca responder à pergunta: “Qual espectro político que possui mais candidatos eleitos?”.



Figura 6.2: Gráfico Número de eleitos. Fonte: Elaborado pelos autores

Na pesquisa para as eleições de 2020 foi possível observar que a Direita conseguiu eleger aproximadamente 61% de seus candidatos. É possível observar também que a Extrema-Esquerda elegeu menos de 1% de seus candidatos.

6.1.3 Candidatos e Eleitos

Este gráfico oferece uma visão comparativa entre o número de eleitos e o número de candidatos, revelando também o percentual de candidatos eleitos em relação à quantidade total de candidatos do espectro.

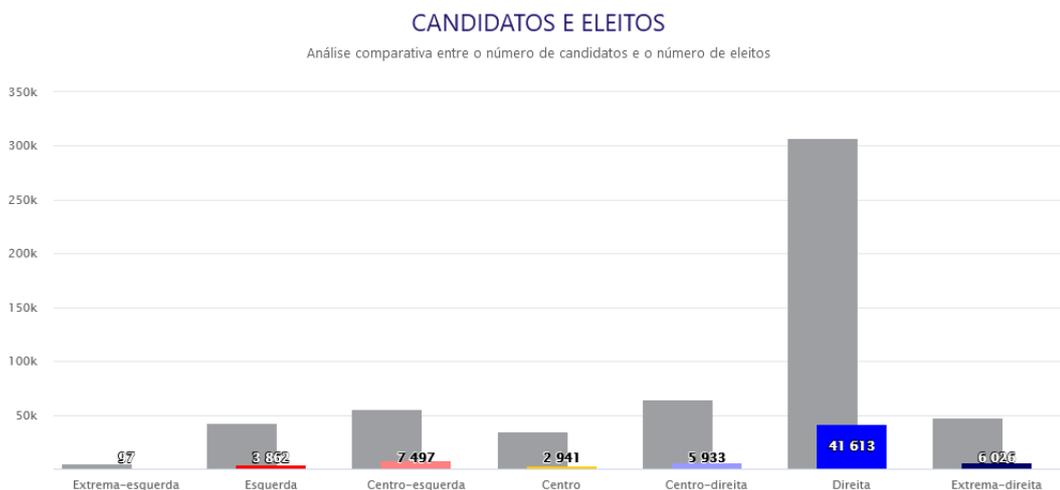


Figura 6.3: Gráfico Candidatos e Eleitos. Fonte: Elaborado pelos autores

É possível observar com este gráfico que nenhum espectro político conseguiu eleger 50% dos seus candidatos nas eleições municipais de 2020.

6.1.4 Eleitos - Cenário Político

O gráfico busca revelar a composição política distribuída pelos espectros políticos com os eleitos. Ele busca responder a pergunta: “Como é a distribuição de eleitos por espectro político no Brasil após as eleições?”.

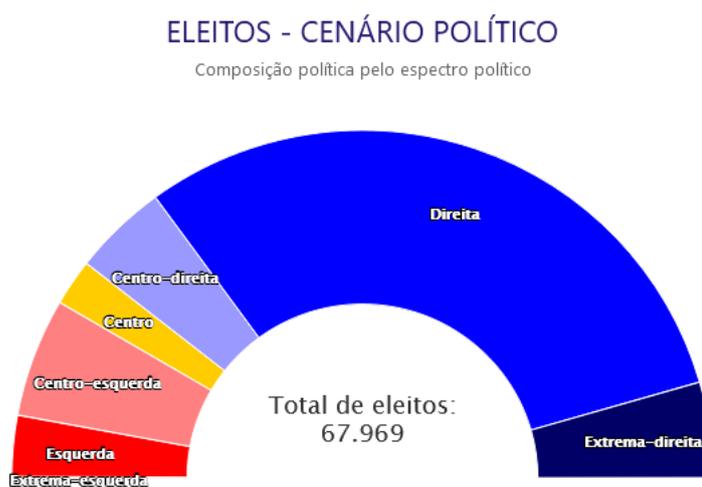


Figura 6.4: Gráfico Eleitos - Cenário Político. Fonte: Elaborado pelos autores

Esse *dashboard* ajuda a ter uma visualização do cenário político pós votação. Ele

mostra que os partidos da Direita nas eleições de 2020 possui maior espaço no cenário político, pois consegue eleger mais candidatos.

6.1.5 Candidatos por gênero

Esses *dashboard* apresenta a composição de homens e mulheres pelos espectros políticos. Ele responde a seguinte questão: “Qual espectro político possui mais mulheres?”

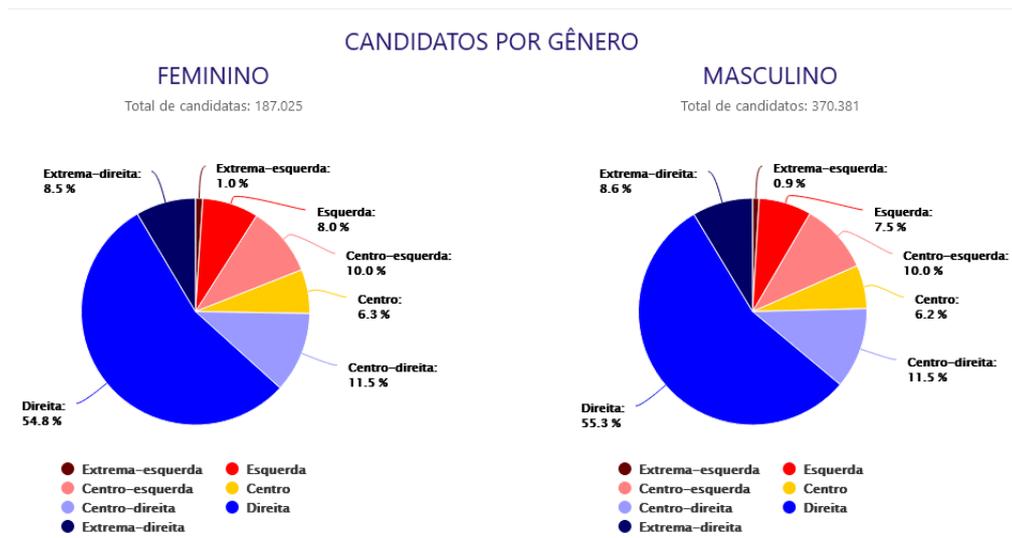


Figura 6.5: Gráfico Candidatos por gênero. Fonte: Elaborado pelos autores

6.1.6 Candidatos por cor/raça

O conjunto de gráficos abaixo apresenta a composição dos espectros políticos por cor/raça. Com eles é possível responder a questão: “Qual espectro político possui mais negros?”

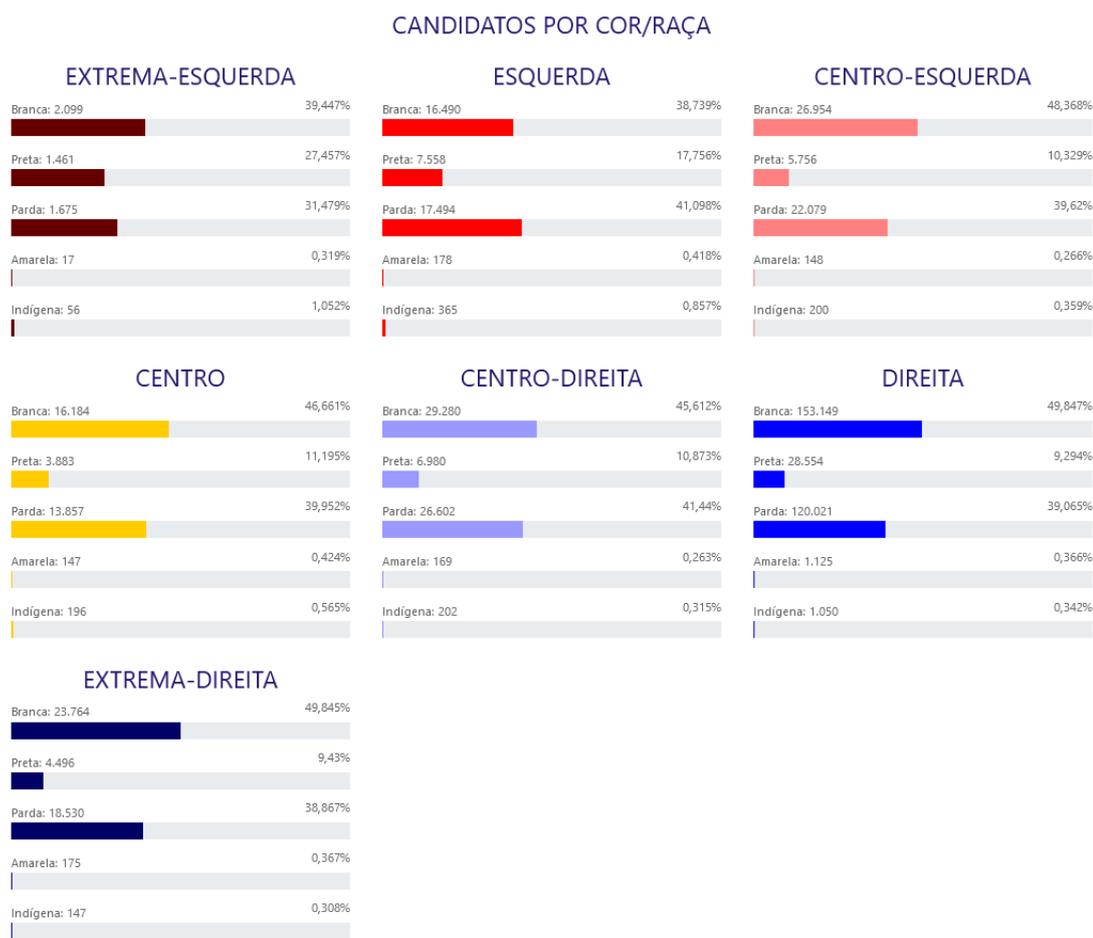


Figura 6.6: Gráfico Candidatos por cor/raça. Fonte: Elaborado pelos autores

Nas eleições de 2020 a Esquerda teve 17,756% dos candidatos Pretos. Na Direita que é o espectro político com maior número de candidatos observou-se que apenas 9,294% dos seus candidatos são pretos. Com esse *dashboards* pode-se observar que em todos os espectros políticos os candidatos são em maioria cor branca, seguido da cor parda.

6.2 Dashboards com o filtro Espectro Político selecionado

Ao selecionar um espectro político os *dashboards* mudam, pois será feita uma análise detalhada do espectro selecionado. Para esse cenário tem-se sete *dashboards*.

Os filtros utilizados para obter os resultados apresentados nessa seção são:

- Ano: 2020

- Região: Todas
- Estado: Todos,
- Turno: 1º Turno
- Tipo de Eleição: Eleição Ordinária
- Eleição: Todas
- Espectro Político: **Direita**

6.2.1 Número de candidatos do espectro

Este *dashboard* é um *card* que apresenta o número de candidatos de um espectro político e qual o seu percentual em relação ao total de candidatos concorrentes na eleição. O *card* é um detalhamento do gráfico apresentado na figura 6.1.



Figura 6.7: Card Número de candidatos. Fonte: Elaborado pelos autores

6.2.2 Número de eleitos do espectro

Este *dashboard* é um *card* que apresenta o número total de candidatos eleitos, o percentual de eleitos dentro do espectro político e o percentual com relação ao número total de candidatos que concorreram às eleições. Ele é um complemento do gráfico apresentado na figura 6.2.



Figura 6.8: Card Número de eleitos. Fonte: Elaborado pelos autores

6.2.3 Candidatos por gênero

Com este gráfico é possível saber a porcentagem de candidatos por gênero.

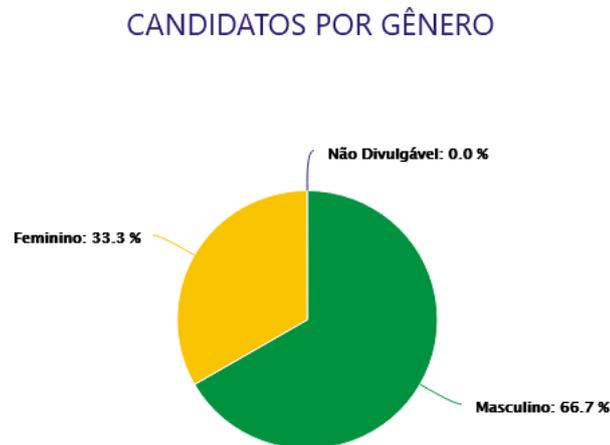


Figura 6.9: Gráfico Candidatos por gênero. Fonte: Elaborado pelos autores

O que se pode observar é que embora a participação feminina nas eleições venha aumentado o número de mulheres candidatas ainda é menos que a metade do número de candidatos.

6.2.4 Eleitos por gênero

O gráfico Eleitos por gênero apresenta o percentual de candidatos eleitos pelo gênero no espectro político.

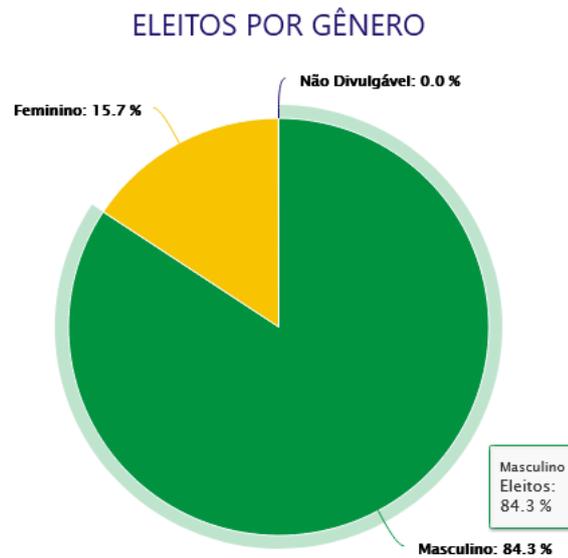


Figura 6.10: Gráfico Eleitos por gênero. Fonte: Elaborado pelos autores

Foi possível observar que o número de mulheres eleitas não ultrapassa os 50% em todos os espectros políticos.

Observamos que o número de participação feminina na política ainda é bem menor se comparado ao número de homens, porém quando analisamos o eleitorado pelo gênero é possível constatar que as mulheres são 52% dos eleitores brasileiros. Em teoria as mulheres podem decidir as eleições, porém ainda são eleitas poucas representantes no Brasil.

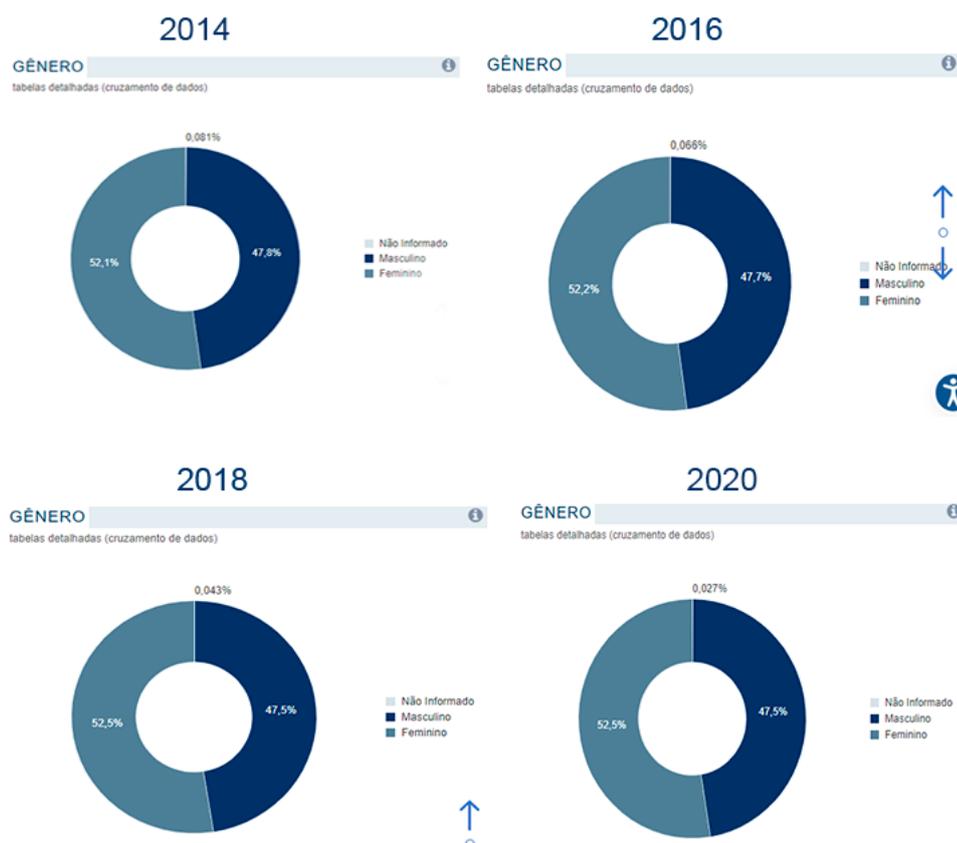


Figura 6.11: Eleitorado por gênero. Fonte: Portal estatísticas eleitorais do TSE

²⁵Acessar: <https://www.tse.jus.br/eleicoes/estatisticas/estatisticas-eleitorais>

6.2.5 Candidatos por cor/raça

O *dashboard* apresenta os candidatos por raça dentro de um espectro político. Esse gráfico é um detalhamento dos gráficos apresentados na figura 6.6.

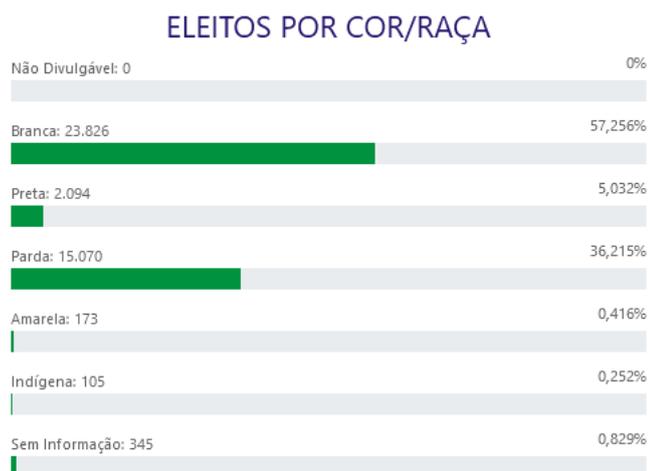


Figura 6.12: Candidatos por cor/raça. Fonte: Elaborado pelos autores

6.2.6 Eleitos por cor/raça

Este gráfico representa os eleitos distribuídos cor/raça. Observamos que o perfil de eleitos predominante é de candidatos brancos. Essa tendência é seguida em todos os espectros políticos.

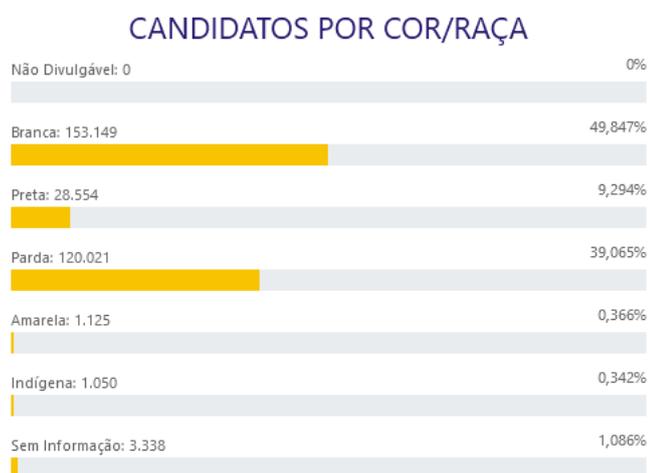


Figura 6.13: Eleitos por cor/raça. Fonte: Elaborado pelos autores

Observamos o número de candidatos eleitos pretos, pardos, amarelos e indígena não ultrapassa os 50%. Esse padrão se repete em todos os espectros políticos.

6.2.7 Candidatos e Eleitos por cargo

O gráfico abaixo representa a distribuição de candidatos e eleitos por cargo pelo espectro político. Pode observar com este *dashboard* quais são os cargos com mais disputa e qual o percentual de eleitos por cargos.

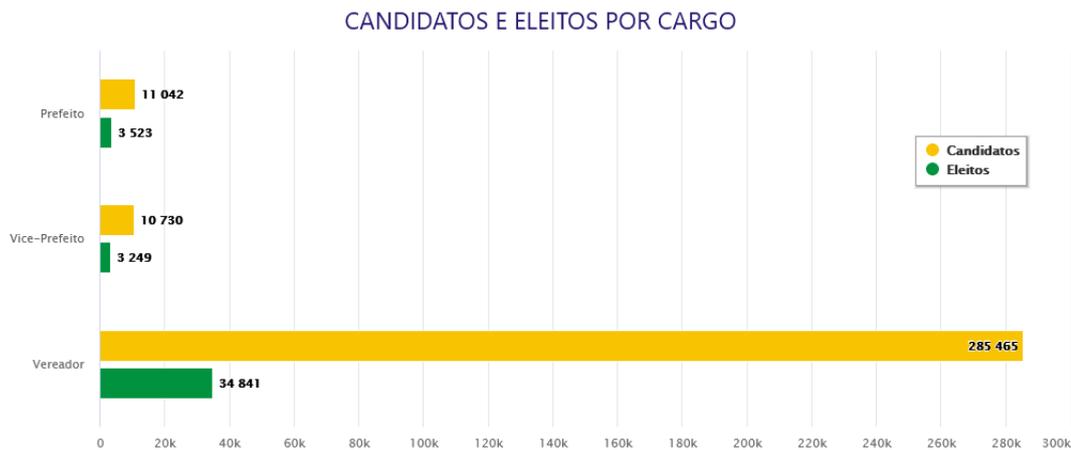


Figura 6.14: Candidatos e Eleitos por cargo. Fonte: Elaborado pelos autores

6.2.8 Candidatos eleitos por partido

O último gráfico apresentado representa uma comparação entre os candidatos e eleitos por partidos do espectro político. Com este *dashboard* é possível descobrir qual partido teve mais candidatos e qual partido teve mais eleitos na eleição. Ele responde a duas perguntas estratégicas: “Qual partido possui mais candidatos dentro de um espectro político?” e “Qual partido mais elegeu candidatos em seu espectro político?”.



Figura 6.15: Candidatos eleitos por partido. Fonte: Elaborado pelos autores

Foi possível observar que nas eleições municipais de 2020 o partido MDB, antigo PMDB, é partido da Direita como maior número de candidatos e também o maior número de candidatos eleitos. Este partido se consolida como o partido de maior representatividade, pois é o partido com mais políticos eleitos.

6.3 Considerações

Durante o desenvolvimento do sistema foi utilizada a teoria descrita no capítulo 2 e foi possível identificar que com o uso de ferramentas como o PDI e a criação dos *dashboards*, assim como o conhecimento técnico, é possível criar análise de dados de forma intuitiva e transparente.

Apesar do desafio de colocar em prática a teoria aprendida durante o desenvolvimento, a principal dificuldade encontrada foi na definição de quais assuntos seriam abordados na análise, pois existe um universo imenso de pesquisas que podem ser abordadas a partir dos dados eleitorais e delimitar o que poderia ser abordado demandou muito tempo.

Uma outra dificuldade identificada foi no momento de criação das consultas, pois era necessário realizar vários JOINS nas consultas, conforme foi explicado no capítulo 2. Em alguns momentos as respostas chegaram a demorar 1 min para ser concluída e nestes momentos era necessário revisar o modelo lógico proposto para buscar melhorar o

desempenho das consultas.

Apesar das dificuldades encontradas durante o desenvolvimento, a estrutura criada atende os requisitos definidos e foi possível atender os objetivos específicos do trabalho.

7 Conclusão

Neste trabalho, foi desenvolvido uma solução de BI a partir dos dados abertos sobre as eleições, esses dados são fornecidos pelo Tribunal Superior Eleitoral. As eleições definem os rumos políticos e econômicos do Brasil, portanto é importante oferecer aos cidadãos brasileiros soluções que o permitam utilizar das informações públicas para poderem tomar as melhores decisões ao votar e também no acompanhamento dos projetos votados no congresso, no senado e nas câmaras municipais.

Criar visualizações e análises para um público alvo vasto é uma atividade complexa, pois é necessário generalizar, todavia é preciso conter informações precisas, claras e importantes, por esse motivo, foram criadas as perguntas estratégicas para que o projeto não fugisse do escopo. Essas perguntas foram utilizadas como base para os requisitos e regras, e como consequência, para a modelagem e para os *dashboards*.

Todo o desenvolvimento do sistema foi feito utilizando metodologias e conceitos descritos no trabalho sendo possível realizar todos os passos planejados e de acordo com os requisitos.

O grande desafio do trabalho foi a classificação dos partidos políticos pelo espectro político, uma vez que essa informação não é obtida no repositório do TSE foi necessário buscar fontes que forneçam essa informação. Nas buscas foi possível descobrir que existem várias técnicas para se fazer esta classificação e após diversas leituras foi possível escolher uma técnica e assim classificar os partidos, a abordagem utilizada neste trabalho foi explicada no capítulo 2.

É possível afirmar que a presente pesquisa cumpriu com os objetivos pretendidos na utilização de conceitos e ferramentas de BI para geração e análise de dados das fontes sobre dados eleitorais no Brasil. Cabe destacar também que um projeto de *Business Intelligence* é um trabalho de constante melhorias e evoluções. A partir deste trabalho é possível desenvolver aperfeiçoamentos na proposta apresentada, criação de novos fatos e novas análises, pois trata-se de uma fonte de dados atualizada e com diversos assuntos a serem explorados.

A primeira sugestão de trabalhos futuros é criar a opção de exportar os dados exibidos nos *dashboards* em formatos *.csv* e a exportação dos gráficos para os formatos *.jpg* ou *.png*. Como foi criada uma API de consulta, outra sugestão é documentar a API e publicar a mesma para que outros desenvolvedores ou pesquisadores possam acessar as informações pelas URIs²⁶, possibilitando o surgimento de novas aplicações.

Concluindo, a última sugestão de trabalho futuro é a criação de novas páginas com algumas novas visualizações, como por exemplo, uma página de coligações partidárias. As coligações representam a união de dois ou mais partidos para alcançar objetivos eleitorais comuns. Podem ser exploradas, por exemplo, as formações das coligações avaliando se partidos de espectros políticos diferentes se unem, avaliar qual partido de uma coligação elege mais candidatos e verificar quais coligações possuem mais recursos financeiros.

²⁶URI (*Uniform Resource Identifier*), em português: Identificador de Recursos Universal. Acesse: https://developer.mozilla.org/pt-BR/docs/Web/HTTP/Basics_of_HTTP/Identifying_resources_on_the_Web

Bibliografia

- ALCANTARA, W. et al. Desafios no uso de dados abertos conectados na educação brasileira. In: SBC. *Anais do IV Workshop de Desafios da Computação Aplicada à Educação*. [S.l.], 2015. p. 11–20.
- BARBIERI, C. *BI2–Business intelligence: Modelagem & Qualidade*. [S.l.]: Elsevier Editora, 2011.
- BATISTA, E. d. O. *Sistemas de informação*. [S.l.]: Saraiva Educação SA, 2017.
- BOLOGNESI, B.; RIBEIRO, E. A.; CODATO, A. *Esquerda, centro ou direita? Como classificar os partidos no Brasil*. 2020. Disponível em: <https://observatoriodaseleicoes.com.br/esquerda-centro-ou-direita-como-classificar-os-partidos-no-brasil/>. Acesso em: 10/02/2021.
- BRAGA, L. P. V. B. *Introdução à Mineração de Dados-2a edição: Edição ampliada e revisada*. [S.l.]: Editora E-papers, 2005.
- BRASIL. *CONSTITUIÇÃO DA REPÚBLICA FEDERATIVA DO BRASIL DE 1988*. 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 10/02/2021.
- BRASIL. *LEI COMPLEMENTAR Nº 101, DE 4 DE MAIO DE 2000*. 2000. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm. Acesso em: 10/02/2021.
- BRASIL. *LEI COMPLEMENTAR Nº 131, DE 27 DE MAIO DE 2009*. 2009. Disponível em: http://www.planalto.gov.br/CCivil_03/leis/LCP/Lcp131.htm. Acesso em: 10/02/2021.
- BRASIL. *LEI Nº 12.527 DE 18 DE NOVEMBRO DE 2011*. 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 10/02/2021.
- BRASIL. *DECRETO Nº 8.777, DE 11 DE MAIO DE 2016*. 2016. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm. Acesso em: 10/02/2021.
- BRASIL. *Controladoria-Geral da União (CGU)*. 2019. Disponível em: <https://www.gov.br/pt-br/orgaos/controladoria-geral-da-uniao>. Acesso em: 10/02/2021.
- BRASIL. *O que é e como funciona*. 2020. Disponível em: <http://www.portaltransparencia.gov.br/sobre/o-que-e-e-como-funciona>. Acesso em: 10/02/2021.
- BRASIL. *O TSE*. 2020. Disponível em: <https://www.tse.jus.br/o-tse/sobre-o-tse/apresentacao>. Acesso em: 10/02/2021.
- BRASIL, W. W. W. W. C. *Manual dos Dados Abertos: Governo*. 2011. Disponível em: https://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf. Acesso em: 03/02/2021.

- CECI, F. Business intelligence. *Palhoça, Unisul*, 2012.
- EAVES, D. *The Three Laws of Open Government Data*. 2009. Disponível em: <https://eaves.ca/2009/09/30/three-law-of-open-government-data>. Acesso em: 03/02/2021.
- ELEITORAIS, R. de dados. *Candidatos*. 2020. Disponível em: https://cdn.tse.jus.br/estatistica/sead/odsele/consulta_cand/consulta_cand_2020/leiam.pdf. Acesso em: 28/02/2021.
- ELIAS, D. *Entenda o processo de ETL*. 2014. Disponível em: <https://canaltech.com.br/business-intelligence/entendendo-o-processo-de-etl-22850/>. Acesso em: 10/02/2021.
- FEW, S. *Information dashboard design: The effective visual communication of data*. [S.l.]: O'Reilly Media, Inc., 2006.
- FIGUEIREDO, V. d. S.; SANTOS, W. J. L. d. Transparência e controle social na administração pública. *Temas de Administração Pública*, v. 8, n. 1, p. 1–20, 2013.
- GIL, A. C. et al. *Como elaborar projetos de pesquisa*. [S.l.]: Atlas São Paulo, 2002. v. 4.
- GOMES, J. C.; SCHNEIDER, M. A. F.; BEZERRA, A. C. Aplicativos cívicos: apropriação de dados abertos governamentais pela sociedade. *XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (XIX ENANCIB); XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (XIX ENANCIB)*, v. 24, n. 2, 2018.
- HAUSENBLAS, M. *As 5 estrelas dos Dados Abertos*. 2012. Disponível em: <https://5stardata.info/pt-BR/>.
- HEYWOOD, A.; OLIVEIRA, I. d. A. R. de. *Ideologias políticas, [v. 2]: do feminismo ao multiculturalismo*. [S.l.]: Editora Ática, 2010.
- INMON, W. H. *Building the data warehouse*. [S.l.]: John wiley & sons, 2005.
- KIMBALL, R. *Fact Tables*. 2008. Disponível em: <https://www.kimballgroup.com/2008/11/fact-tables/>. Acesso em: 10/02/2021.
- KIMBALL, R.; ROSS, M. *The data warehouse toolkit: the complete guide to dimensional modeling*. [S.l.]: John Wiley & Sons, 2011.
- KNOWLEDGE, O. *O que são Dados Abertos*. 2021. Disponível em: http://opendatahandbook.org/guide/pt_BR/what-is-open-data/. Acesso em: 03/02/2021.
- MACHADO, F. N. R. *Tecnologia e projeto de Data Warehouse*. [S.l.]: Saraiva Educação SA, 2004.
- MACIEL, A. P. B.; ALARCON, A. de O.; GIMENES, É. R. Partidos políticos e espectro ideológico: parlamentares, especialistas, esquerda e direita no brasil. *Revista Eletrônica de Ciência Política*, v. 8, n. 3, 2018.
- NOVATO, D. *O que é Business Intelligence?* 2017. Disponível em: <https://www.oficinadanet.com.br/post/13153-o-que-e-business-intelligence>. Acesso em: 10/02/2021.
- PERKINS, A. Developing a data warehouse. *The Enterprise Engineering Approach, Visible Systems Corporation*, v. 96, 1995.

- PIRES, J. *O que é API? REST e RESTful? Conheça as definições e diferenças!* 2017. Disponível em: <https://becode.com.br/o-que-e-api-rest-e-restful/>. Acesso em: 12/02/2021.
- PIZANI, A. L.; SIQUEIRA, H. B. Geração de indicadores da justiça: utilizando uma solução de business intelligence com dados abertos. *Sistemas de Informação-Florianópolis*, 2018.
- PUJARI, A. K. *Data mining techniques*. [S.l.]: Universities press, 2001.
- SCHEEFFER, F. et al. Ideologia e comportamento parlamentar na câmara dos deputados: faz sentido ainda falar em esquerda e direita? 2016.
- SILVA, E. L. d.; MENEZES, E. M. A pesquisa e suas classificações. ----- *Metodologia da pesquisa e elaboração de dissertação*, v. 4, 2005.
- SILVA, P. N.; PINHEIRO, M. M. K. Dados governamentais abertos e a lei de acesso à informação: diagnóstico nas universidades públicas federais brasileiras. 2017.
- TAURION, C. O data warehouse será útil para a sua organização. *Developers' Magazine*, n. 18, p. 26–27, 1998.
- TURBAN, E. et al. *Business intelligence: um enfoque gerencial para a inteligência do negócio*. [S.l.]: Bookman Editora, 2009.
- VAZ, J. C.; RIBEIRO, M. M.; MATHEUS, R. Dados governamentais abertos e seus impactos sobre os conceitos e práticas de transparência no brasil. *Cadernos ppg-au/ufba*, v. 9, n. 1, 2010.
- VEIGA, J. C. de A.; GUIMARÃES, J. C. B.; TANAKA, A. K. Análise de dados abertos governamentais usando técnicas de business intelligence: um estudo de caso das eleições 2014. 2015.
- VICTORINO, M. de C. et al. Uma proposta de ecossistema de big data para a análise de dados abertos governamentais concetados. *Informação & sociedade*, Universidade Federal da Paraíba-Programa de Pós-Graduação em Ciência da Computação, v. 27, n. 1, 2017.