



UNIVERSIDADE FEDERAL DE MINAS GERAIS

Departamento de Ciência da Computação

Belo Horizonte, Minas Gerais

Introdução a Banco de Dados

Trabalho Prático - PARTE 1

Aynoa Souza Jorgino, Daniel Lopes Cardoso, Filipe Davi Gomes, Kamily
Victoria Araújo Ferreira, Mateus Souza Barreto

Introdução

A proposta deste trabalho consiste em promover o acesso, coleta, gerenciamento, integração e análise de conjuntos de dados públicos. A revelação é que, a partir de dados brutos, seja possível criar um novo produto de dados com valor agregado, resultado da integração e da análise. Este documento corresponde ao relatório da Parte 1 do trabalho prático. As seções a seguir descreverão as fontes de dados públicas selecionadas, o processo técnico de aquisição e carga dos dados no Dbeaver e, por fim, o esquema conceitual e o dicionário de dados desenvolvidos para a integração dos conjuntos de dados.

1	Fontes de Dados	2
2	Processo de Aquisição e Carga	2
3	Esquema Conceitual	3
4	Dicionário de Dados	5
4.1	Metadados	5
4.2	Principais colunas (Dicionários das três tabelas)	5
5	Bibliografia	7

1 Fontes de Dados

Foram selecionados três conjuntos de dados públicos referentes ao ciclo político e financeiro de 2022. A seleção partiu da exigência de se obter dados que mostrariam o aspecto do mesmo problema, neste caso, o ciclo de gastos políticos no Brasil, desde a campanha eleitoral até o exercício do mandato parlamentar. Os conjuntos de dados foram obtidos de duas fontes principais: o Tribunal Superior Eleitoral (TSE) e a Câmara dos Deputados.

Table 1: Conjuntos de dados selecionados para o trabalho.

Conjunto de Dados	Órgão Produtor	Fonte (Link)	Arquivo Utilizado
Candidatura 2022	Tribunal Superior Eleitoral (TSE)	Portal de Dados Abertos (TSE)	<code>consulta_cand_2022_BRASIL.csv</code>
Despesas de Campanha 2022	Tribunal Superior Eleitoral (TSE)	Portal de Dados Abertos (TSE)	<code>despesas_contratadas_candidatos_2022_BRASIL.csv</code>
Cota Parlamentar 2022	Câmara dos Deputados	Portal de Dados Abertos (Câmara)	<code>Ano-2022.csv</code>

A correlação entre as fontes é o que viabiliza a análise integrada. Os dados do TSE permitem vincular um candidato (`candidatos_unicos`) aos seus gastos de campanha (`despesas_campanha`). A expectativa de realização de análise é, a partir desta primeira integração, identificar os candidatos que foram eleitos e cruzá-los com os dados da Câmara dos Deputados. Isso permitirá uma análise de comparação entre os gastos de campanha de um indivíduo com seus gastos públicos durante o exercício do mandato (`gastos_parlamentares`).

2 Processo de Aquisição e Carga

Primeiramente, os três arquivos CSV obtidos das fontes foram importados para o banco de dados. Para garantir o sucesso da importação inicial e evitar erros de "parse" de dados "sujos" ou mal formatados, todas as colunas foram carregadas com o tipo `text`. Esta abordagem permitiu que os dados crus ficassem disponíveis dentro do SGBD (Dbeaver) para a etapa de preparação.

Após a carga inicial, iniciou-se o processo de preparação e estruturação dos dados. Colunas que representavam dados numéricos, datas ou valores monetários foram inspecionadas e convertidas para seus tipos semânticos corretos. Por exemplo, colunas como `ano_eleicao` foram alteradas de `text` para `integer`, `dt_nascimento` para `date`, e valores monetários como `vlrliquido` para `numeric(10, 2)`.

Uma parte crucial deste processo foi a definição das chaves primárias para permitir a integração e a criação do esquema conceitual. Foi realizada uma análise exploratória preliminar para verificar a viabilidade de chaves primárias naturais (compostas). Esta análise revelou inconsistências nos dados de origem, como a presença de linhas duplicadas na combinação de colunas que deveriam ser únicas (ex: em `despesas_campanha`). Por esta razão, optou-se por adotar chaves primárias artificiais, como `id_despesa` (tipo `BIGSERIAL`), para garantir a integridade referencial do banco de dados e viabilizar a carga completa dos dados.

3 Esquema Conceitual

O esquema conceitual, representado na Figura 1 como um esquema lógico relacional, detalha a estrutura das três tabelas (`candidatos_unicos`, `despesas_campanha`) e

(`gastos_parlamentares`) e as ligações entre elas. A integração dos dados, conforme solicitado pelo trabalho, é viabilizada pelas colunas que correlacionam os conjuntos de dados, como `nr_cpf_candidato` (ligando `candidatos` a `despesas`) e `idecadastro` (identificador do parlamentar). Esse esquema foi criado pelo próprio Dbeaver.

Um dos principais desafios durante a fase de carga e modelagem foi a definição das chaves primárias (PKs). A primeira abordagem foi tentar identificar uma chave primária natural (composta) para cada tabela, baseada nos atributos dos dados de origem.

No entanto, a análise exploratória preliminar (uma etapa da Parte 2 realizada para viabilizar a carga da Parte 1) revelou inconsistências e problemas de qualidade nos dados. Na tabela `despesas_campanha`, por exemplo, a combinação de colunas que naturalmente identificaria uma despesa (como `nr_cpf_candidato`, `nr_cpf_cnpj_fornecedor`, `dt_despesa` e `nr_documento`) apresentou duplicatas, tornando-a inviável como PK.

Por esta razão, conforme descrito na seção anterior sobre o processo de carga, optou-se por adotar chaves primárias artificiais. Foram criadas as colunas `id_despesa` (em `despesas_campanha`) e `id_gasto` (em `gastos_parlamentares`), ambas do tipo `BIGSERIAL`, para servir como chaves primárias únicas e não nulas. Esta abordagem garante a integridade referencial do banco de dados e simplifica as operações de junção (*JOIN*) necessárias para a análise integrada.

A imagem do diagrama é observada na próxima página.

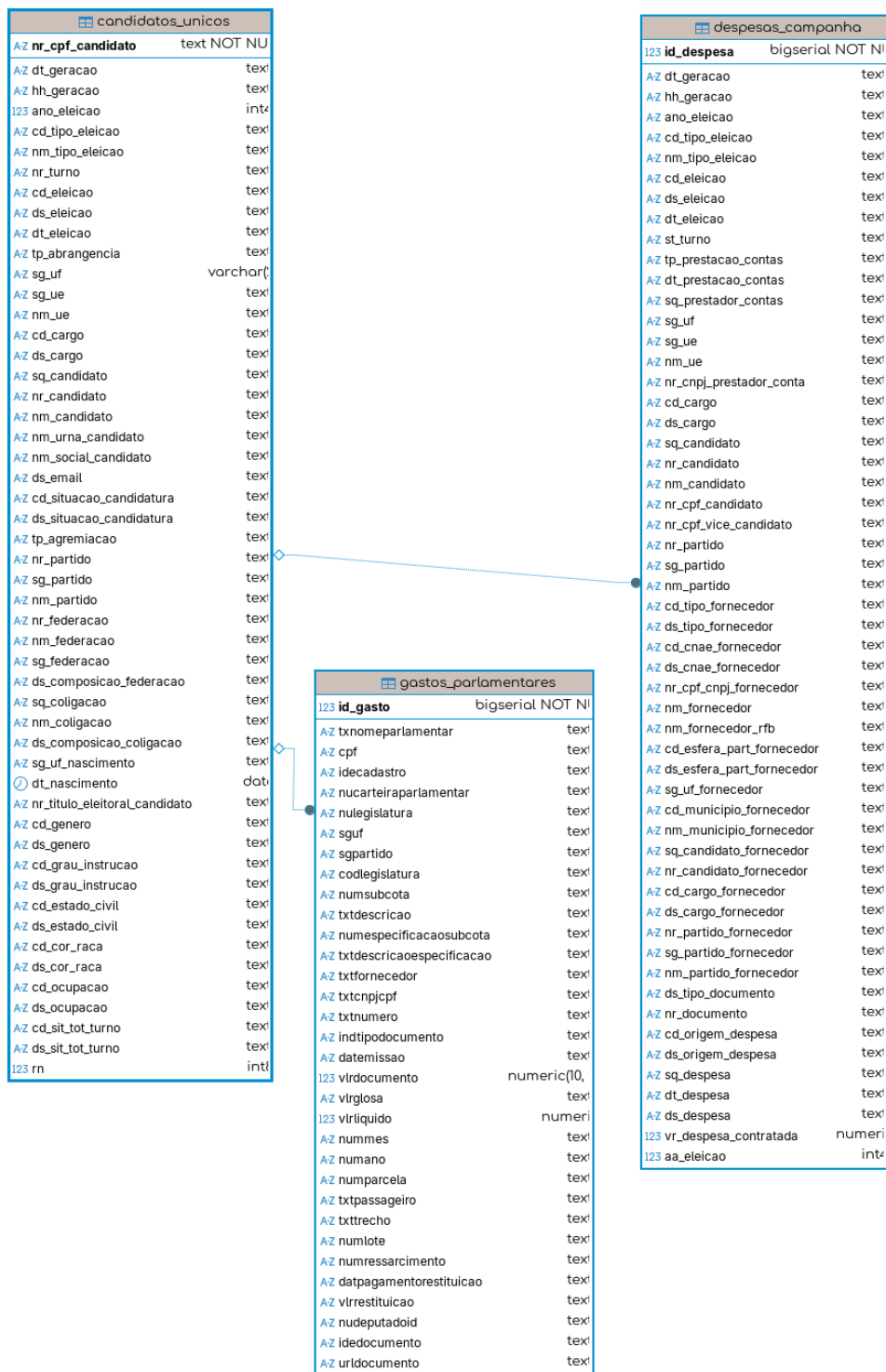


Figure 1: Diagrama

4 Dicionário de Dados

Este quadro resume as informações de proveniência e contexto dos dados utilizados.

4.1 Metadados

Table 2: Metadados dos conjuntos de dados.

Atributo	Descrição
Fontes	Portal de Dados Abertos do TSE; Portal de Dados Abertos da Câmara
Órgão Produtor	Tribunal Superior Eleitoral (TSE); Câmara dos Deputados
Data de Obtenção	10/10/2025
Data de Referência	2022 (Ano eleitoral e ano da cota parlamentar)
Cobertura Geográfica	Nacional (Brasil)
Limitações Registradas	Como será detalhado na Parte 2, foram identificadas "inconsistências" nos dados brutos, como duplicatas em chaves naturais candidatas, exigindo a criação de chaves artificiais.

4.2 Principais colunas (Dicionários das três tabelas)

A tabela a seguir descreve a estrutura final da tabela *candidatos_unicos*, após o processo de limpeza e conversão de tipos de dados. Os principais tipos.

Table 3: Dicionário da tabela *candidatos_unicos*.

Nome	Descrição	Tipo	Tamanho	Restrições
nr_cpf_candidato	CPF	varchar	11	PK
nm_candidato	Nome completo.	text		NOT NULL
ano_eleicao	Ano da eleição.	integer		NOT NULL
sg_uf	Sigla da UF.	varchar	2	
ds_cargo	Descrição do cargo.	text		
dt_nascimento	Data de nascimento.	date		
ds_genero	Gênero.	text		

Descreve os gastos da Cota para Exercício da Atividade Parlamentar (CEAP).

Table 4: Dicionário da tabela *gastos_parlamentares*.

Nome	Descrição	Tipo	Tamanho	Restrições
id_gasto	Chave primária artificial.	bigserial		PK
idecadastro	ID do parlamentar na Câmara.	text		(Chave de integração)
txnomeparlamentar	Nome do parlamentar.	text		
numano	Ano do gasto.	integer		
nummes	Mês do gasto.	integer		
txtdescricao	Descrição da cota/-gasto.	text		
datemissao	Data de emissão do fiscal.	date		
vlrliquido	Valor líquido ressarcido.	numeric	(10, 2)	

Descreve as despesas de campanha. Utiliza uma chave primária criada (*id_despesa*).

Table 5: Dicionário da tabela *despesas_campanha*.

Nome	Descrição	Tipo	Tamanho	Restrições
id_despesa	Chave primária artificial.	bigserial		PK
nr_cpf_candidato	CPF do candidato.	varchar	11	FK (candidatos_unicos)
dt_despesa	Data da despesa.	date		
nr_cpf_cnpj_fornecedor	CPF/CNPJ do fornecedor.	text		
nm_fornecedor	Nome do fornecedor.	text		
vr_despesa_contratada	Valor da despesa.	numeric	(10, 2)	
ds_despesa	Descrição da despesa.	text		

5 Bibliografia

- [1] TRIBUNAL SUPERIOR ELEITORAL (TSE). *Candidatos - 2022*. Portal de Dados Abertos. Disponível em: <https://dadosabertos.tse.jus.br/dataset/candidatos-2022>. Acesso em: 11 out. 2025.
- [2] TRIBUNAL SUPERIOR ELEITORAL (TSE). *Prestação de Contas Eleitorais - 2022*. Portal de Dados Abertos. Disponível em: <https://dadosabertos.tse.jus.br/dataset/prestacao-de-contas-eleitorais-2022>. Acesso em: 11 out. 2025.
- [3] CÂMARA DOS DEPUTADOS. *Cota para Exercício da Atividade Parlamentar (Arquivos Estáticos)*. Portal de Dados Abertos. Disponível em: <https://dadosabertos.camara.leg.br/swagger/api.html?tab=staticfile>. Acesso em: 11 out. 2025.