

eleicao_22

May 26, 2023

1 Eleição 2022

O trabalho envolve o processamento dos dados da Eleição Nacional no Brasil de 2022.

1.1 Dados

Os dados são os valores dos votos registrados nos 2 turnos (1º turno ocorrido em 2 de outubro, 2º turno ocorrido em 30 de outubro) para as eleições de deputados estaduais, deputados federais, senadores e presidente da república.

Os arquivos contendo os valores das votações, os marcos temporais de abertura e encerramento das urnas, quantidades de votos, nomes de candidatos e partidos estão disponibilizados cada Unidade da Federação (UF). Os dados dentro de cada arquivo são codificados em ISO 8859-1 (Latin1), separados por “;” e os registros ausentes ou conceitualmente sem informação (nome de candidato para votos na legenda, por exemplo) são registrados como “#NULO#” para campos alfabéticos e -1 para campos numéricos.

1.1.1 Obtenção

Os dados foram obtidos do portal oficial do Tribunal Superior Eleitoral (TSE) em <https://dadosabertos.tse.jus.br/>.

Os arquivos, separados por Unidade Federativa (UF) por motivo de tamanho, tem os caracteres codificados em ISO 8859-1 e os dados (variáveis e seus valores) são disponibilizadas em arquivos CSV (*comma separated values* - valores separados por vírgulas).

1.1.2 Seleção de variáveis e conversões

A seguir são listadas as variáveis, segundo a descrição do documento que acompanha os dados, os comentários sobre seus valores e se foram ou não utilizadas.

Variável	Descrição	Comentário
DT_GERACAO (não utilizada)	Data da extração dos dados para geração do arquivo.	Os valores são 05/10/2022 e 31/10/2022.
HH_GERACAO (não utilizada)	Hora da extração dos dados para geração do arquivo com base no horário de Brasília.	Os valores são 13:21:57 para 05/10/2022 e 15:35:43 para 31/10/2022.

Variável	Descrição	Comentário
ANO_ELEICAO (não utilizada)	Ano de referência da eleição para geração do arquivo. Observação: Para eleições suplementares o ano de referência da eleição é o da eleição ordinária correspondente. Por exemplo: Em 2016 houve eleições ordinárias. Após a data desta eleição ordinária e antes da próxima, houve eleições suplementares em 2017, 2018 e 2019. As informações destas eleições suplementares estarão divulgadas no arquivo gerado para as Eleições 2016.	O valor é sempre 2022.
CD_TIPO_ELEICAO (não utilizada)	Código do tipo de eleição. Pode assumir os valores: 0 - Eleição Ordinária, 1 - Eleição Extraordinária, 2 - Consulta Popular.	O valor é sempre 0.

Variável	Descrição	Comentário
NM_TIPO_ELEICAO (não utilizada)	<p>Nome do tipo de eleição.</p> <p>Observação: As eleições ordinárias são previstas em Lei, possuem data certa para serem realizadas, ocorrem em anos pares e possuem a periodicidade de 04 em 04 anos. Nas eleições ordinárias nacionais são eleitos os cargos de Presidente, Governadores, Deputados (Federais e Estaduais) e Senadores. Nas eleições ordinárias municipais são eleitos os cargos de Prefeito e Vereadores. As eleições suplementares são aquelas que não têm periodicidade pré-determinada ou definida e ocorrem quando, eventualmente, se fizerem necessárias. As consultas populares ocorrem sempre que a população é convocada a opinar diretamente sobre um assunto específico e importante. Ela pode ser realizada de duas formas: plebiscito (quando o cidadão opina previamente sobre a possível criação de uma lei) e referendo (quando uma lei aprovada por um órgão legislativo é submetida à aceitação ou não dos eleitores).</p>	O valor é sempre Eleição Ordinária .

Variável	Descrição	Comentário
CD_PLEITO	Código do pleito é uma informação de eleições em uma determinada data. Um Pleito pode ter uma ou mais eleições. Pleito eleitoral: Assim se diz em alusão à luta ou disputa, que se fere nas eleições, para designar o desenrolar destas. E, desse modo, extensivamente, é a expressão usada para designar as próprias eleições, no período em que se registrarem as votações.	Os valores são 406 para o 1º turno e 407 para o 2º turno.
DT_PLEITO (não utilizada)	Data do pleito da eleição.	Os valores são 02/10/2022 para o 1º turno e 30/10/2022 para o 2º turno.

Variável	Descrição	Comentário
NR_TURNO	<p>Número do turno da eleição.</p> <p>Observação: No Brasil, as eleições realizam-se por meio de dois sistemas: o sistema majoritário (aplicado aos cargos de Presidente, Vice-Presidente, Governador, Vice-Governador, Prefeito, Vice-Prefeito e Senador) e o sistema proporcional (aplicado aos cargos de Deputado Federal, Deputado Estadual, Deputado Distrital e Vereador). O sistema majoritário consiste em declarar eleito o candidato que tenha recebido a maioria dos votos válidos (excluídos os votos em brancos e os votos nulos). Caso o candidato ao cargo indicado no sistema majoritário, com exceção do cargo de Senador, não alcance maioria absoluta destes votos válidos no primeiro turno (mínimo de 50% + 1), haverá segundo turno em que concorrerão apenas os dois candidatos mais votados. O segundo turno das eleições no Brasil ocorre para os cargos de Presidente, Vice-Presidente da República, Governadores e Vice-Governadores dos Estados e do Distrito Federal e para Prefeitos e Vice-Prefeitos de Municípios com mais de 200 mil eleitores. Nos municípios cujo eleitorado é igual ou menor que 200 mil e para o cargo de Senador elege-se o candidato que tenha alcançado a maioria simples dos votos.</p>	Os valores são 1 para o 1º turno e 2 para o 2º turno.

Variável	Descrição	Comentário
CD_ELEICAO	Código único da eleição no âmbito da Justiça Eleitoral. Observação: Este código é único por eleição e por turno, ou seja, cada turno possui seu código de eleição.	Os valores são: 544 (Eleição geral federal 2022), 546 (Eleições gerais estaduais 2022), 548 (Eleição conselho distrital 2022) para o 1º turno (CD_PLEITO 406) e 545 (Eleição geral federal 2022), 547 (Eleições gerais estaduais 2022) para o 2º turno.
DS_ELEICAO	Descrição da eleição.	Os valores são: Eleição geral federal 2022 (544 no 1º turno, 545 no 2º turno), Eleições gerais estaduais (546 no 1º turno, 547 no 2º turno), Eleição conselho distrital (548) para o 1º turno.
SG_UF	Sigla da Unidade da Federação em que ocorreu a eleição.	Os valores são AC, AL, AM, AP, BA, CE, DF, ES, GO, MA, MG, MS, MT, PA, PB, PE, PI, PR, RJ, RN, RO, RR, RS, SC, SE, SP, TO para os estados brasileiros e ZZ para urnas no exterior.
CD_MUNICIPIO	Código TSE do município onde ocorreu a eleição.	O código é único para cada cidade. Cada estado pouco populoso parece ter reservado 1000 códigos para suas cidades e os demais estados 2000.
NM_MUNICIPIO	Nome do município onde ocorreu a eleição.	Os valores são texto com nome da cidade (exemplo ITABAIANA, CAMPINAS etc.).
NR_ZONA	Número da Zona eleitoral onde ocorreu a eleição.	Número único por estado, incremental iniciando em 1. Uma zona pode envolver mais que um município. Uma zona possui várias seções.
NR_SECAO	Número da Seção eleitoral onde ocorreu a eleição.	Cada seção é de uma zona. Zona e seção só são únicas dentro do estado.

Variável	Descrição	Comentário
NR_LOCAL_VOTACAO	Número do local de votação referente ao boletim de urna.	O local de votação é único por zona, mas repete-se a cada cidade. Cada zona ocorre em um único local de votação, mas o número do local de votação só é único dentro da cidade; aparentemente toda cidade inicia a numeração em 1015 dos seus locais de votação.
CD_CARGO_PERGUNTA	Código do cargo do candidato ou pergunta, no caso de plebiscito.	Os valores são 1 (Presidente), 3 (Governador), 5 (Senador), 6 (Deputado Federal), 7 (Deputado Estadual), 8 (Deputado Distrital), 25 (Conselheiro Distrital).
DS_CARGO_PERGUNTA	Descrição do cargo do candidato ou pergunta, no caso de plebiscito.	Os valores são Presidente, Governador, Senador, Deputado Federal, Deputado Estadual, Deputado Distrital, Conselheiro Distrital.

Variável	Descrição	Comentário
NR_PARTIDO	Número do partido de origem do candidato.	Os valores são os códigos dos partidos: 10 (REPUBLICANOS), 11 (PROGRESSISTAS), 12 (Partido Democrático Trabalhista), 13 (Partido dos Trabalhadores), 14 (Partido Trabalhista Brasileiro), 15 (Movimento Democrático Brasileiro), 16 (Partido Socialista dos Trabalhadores Unificado), 18 (Rede Sustentabilidade), 19 (Podemos), 20 (Partido Social Cristão), 21 (Partido Comunista Brasileiro), 22 (Partido Liberal), 23 (Cidadania), 27 (Democracia Cristã), 28 (Partido Renovador Trabalhista Brasileiro), 29 (Partido da Causa Operária), 30 (Partido Novo), 33 (Partido da Mobilização Nacional), 35 (Partido da Mulher Brasileira), 36 (AGIR), 40 (Partido Socialista Brasileiro), 43 (Partido Verde), 44 (UNIÃO BRASIL), 45 (Partido da Social Democracia Brasileira), 50 (Partido Socialismo e Liberdade), 51 (Patriota), 55 (Partido Social Democrático), 60 (Nome Partido 60), 65 (Partido Comunista do Brasil), 70 (AVANTE), 77 (Solidariedade), 80 (Unidade Popular), 85 (Nome Partido 85), 90 (Partido Republicano da Ordem Social). O valor -1 registra voto em branco ou nulo.

Variável	Descrição	Comentário
SG_PARTIDO	Sigla do partido de origem do candidato.	Os valores são os códigos dos partidos: REPUBLICANOS (REPUBLICANOS), PP (PROGRESSISTAS), PDT (Partido Democrático Trabalhista), PT (Partido dos Trabalhadores), PTB (Partido Trabalhista Brasileiro), MDB (Movimento Democrático Brasileiro), PSTU (Partido Socialista dos Trabalhadores Unificado), REDE (Rede Sustentabilidade), PODE (Podemos), PSC (Partido Social Cristão), PCB (Partido Comunista Brasileiro), PL (Partido Liberal), CIDADANIA (Cidadania), DC (Democracia Cristã), PRTB (Partido Renovador Trabalhista Brasileiro), PCO Partido da Causa Operária, NOVO (Partido Novo), PMN (Partido da Mobilização Nacional), PMB (Partido da Mulher Brasileira), AGIR (AGIR), PSB (Partido Socialista Brasileiro), PV (Partido Verde), UNIÃO (UNIÃO BRASIL), PSDB (Partido da Social Democracia Brasileira), PSOL (Partido Socialismo e Liberdade), PATRIOTA (Patriota), PSD (Partido Social Democrático), P60 (Nome Partido 60), PC do B (Partido Comunista do Brasil), AVANTE (AVANTE), SOLIDARIEDADE (Solidariedade), UP (Unidade Popular), P85 (Nome Partido 85), PROS (Partido Republicano da Ordem Social). O valor #NULO# registra voto em branco ou nulo.

Variável	Descrição	Comentário
NM_PARTIDO	Nome do Partido de origem do candidato.	Ver comentários de SG_PARTIDO e NR_Partido. O valor #NULO# registra voto em branco ou nulo.
DT_BU_RECEBIDO (não utilizada)	Data do boletim de urna recebido.	Um marco temporal com data e horário no formato DD/MM/AAAA HH:MM:SS (exemplo: 02/10/2022 19:39:56, 02/10/2022 20:40:41 etc.).
QT_APTOS	Quantidade de eleitores aptos a votar naquele município, zona e seção.	Valor inteiro não negativo.
QT_COMPARECIMENTO (não utilizada)	Quantidade de eleitores que compareceram às eleições naquele município, zona e seção.	Valor inteiro não negativo. Não foi utilizada por ser redundante: QT_COMPARECIMENTO = QT_APTOS - QT_ABSTENCOES.
QT_ABSTENCOES	Quantidade de eleitores que não compareceram às eleições naquele município, zona e seção.	Valor inteiro não negativo.
CD_TIPO_URNA (não utilizada)	Código de identificação do tipo de urna.	Sempre 1 (Apurada).
DS_TIPO_URNA (não utilizada)	Descrição de identificação do tipo de urna.	Sempre Apurada.
CD_TIPO_VOTAVEL	Código do tipo de votável.	Os valores são 1 (Nominal), 2 (Branco), 3 (Nulo) e 4 (Legenda).
DS_TIPO_VOTAVEL	Descrição do tipo votavel.	Os valores são Nominal, Branco, Nulo e Legenda.

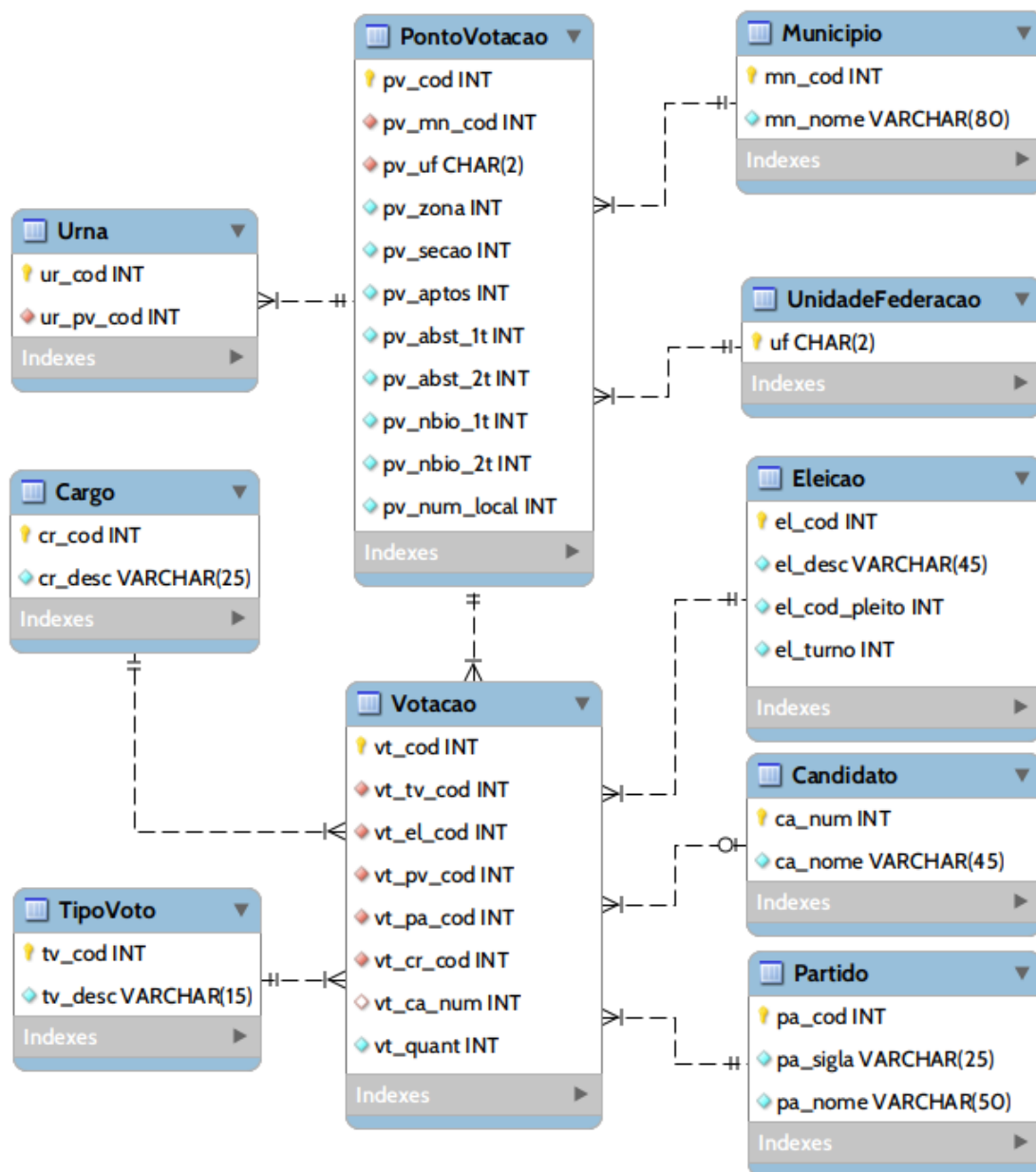
Variável	Descrição	Comentário
NR_VOTAVEL	Pode assumir os valores: Número do candidato (quando voto nominal); Número do partido (quando voto em legenda); Número 95 (quando voto em branco); Número 96 (quando voto nulo); Número 97 (quando voto anulado e apurado em separado) e Número 98 (quando voto anulado).	Os valores são os números dos partidos ou os códigos de branco, nulo ou anulado (exemplo 13, 95 etc.).
NM_VOTAVEL	Pode assumir os valores: Nome do candidato (quando voto nominal ou voto anulado); Nome do partido (quando voto em legenda); Branco (quando voto em branco); Nulo (quando voto nulo) e “Voto anulado e apurado em separado” (quando voto anulado e apurado em separado).	Os valores são os nomes dos candidatos (ou do partido, quando voto na legenda) ou Branco, Nulo ou Anulado (exemplo: REQUIÃO, PTB etc.).
QT_VOTOS	Quantidade de votos recebidos pelo votável.	Inteiro não negativo.
NR_URNA_EFETIVADA	Número da urna efetivada.	Um valor inteiro de sete dígitos (exemplo: 1797374, 1269487 etc.). Em geral é um valor para cada município/zona/seção, mas pode ter mais que 1. Um NR_LOCAL_VOTACAO tem várias NR_URNA_EFETIVADA
CD_CARGA_1_URNA_EFETIVADA (não utilizada)	Código da Carga 1 da urna efetivada.	Um valor com seis grupos de três dígitos (exemplo: 711.637.365.686.035.465., 509.516.924.804.368.201. etc.).
CD_CARGA_2_URNA_EFETIVADA (não utilizada)	Código da Carga 2 da urna efetivada.	Um valor de 5 ou 6 dígitos (exemplo: 44612, 381354 etc.).

Variável	Descrição	Comentário
CD_FLASHCARD_URNA_EFETIVADA (não utilizada)	Código do Flashcard de urna efetivada.	Um valor hexadecimal de 8 dígitos (exemplo: 83F0903D, 00F22BC0 etc.).
DT_CARGA_URNA_EFETIVADA (não utilizada)	Data da carga da urna efetivada.	Um marco temporal com data e horário no formato DD/MM/AAAA HH:MM:SS (exemplo: 21/09/2022 16:58:00, 21/09/2022 09:50:00 etc.).
DS_CARGO_PERGUNTA_SECAO (não utilizada)	Informação de código cargo pergunta e seção. Apresentando os valores agrupados de CD_CARGO_PERGUNTA CD_SECAO, caso um dos valores seja nulo, deverá retornar “-” no lugar do código.	O valor é um texto com CD_CARGO_PERGUNTA e CD_SECAO separados por um traço “-” (exemplo 6 - 151, 1 - 96 etc.).
DS_AGREGADAS (não utilizada)	Lista das seções agregadas.	Normalmente contém #NULO#. Para alguns registros contém um valor (exemplo 426 , 79 etc.) ou uma sequência de valores separados por barra / (exemplo 184 / 185, 76 / 152 / 155 / 157 / 161 etc.).
DT_ABERTURA (não utilizada)	Data/hora de abertura da urna eletrônica para votação.	Um marco temporal com data e horário no formato DD/MM/AAAA HH:MM:SS (exemplo: 02/10/2022 08:02:54, 02/10/2022 08:00:01 etc.).
DT_ENCERRAMENTO (não utilizada)	Data/hora de encerramento da urna eletrônica para votação.	Um marco temporal com data e horário no formato DD/MM/AAAA HH:MM:SS (exemplo: 02/10/2022 17:00:40, 30/10/2022 17:05:13 etc.).
QT_ELEITORES_BIOMETRIA_NH	Quantitativo de eleitores com biometria, mas que não foram habilitados por meio dela.	Valor inteiro não negativo.

Variável	Descrição	Comentário
DT_EMISSAO_BU (não utilizada)	Data/hora de emissão do boletim de urna (BU).	Um marco temporal com data e horário no formato DD/MM/AAAA HH:MM:SS (exemplo: 30/10/2022 17:07:22, 02/10/2022 17:09:17 etc.).
NR_JUNTA_APURADORA (não utilizada)	Número da junta apuradora.	Os valores que apareceram na amostragem são -1 (indicando não disponível), 1 e 6.
NR_TURMA_APURADORA (não utilizada)	Número da turma apuradora.	Os valores que apareceram na amostragem são -1 (indicando dado não disponível) e 1.

1.2 Modelo de dados

Após estudo das variáveis e seus valores, foi construído um modelo de dados para armazenar em um Sistema Gerenciador de Banco de Dados as informações normalizadas.



1.3 Importação dos dados

1.3.1 Processamento inicial

Os dados originais foram processados de modo a serem adequados ao modelo e inseridos em um Sistema Gerenciador de Bancos de Dados. Um programa em Java foi utilizado alguns ajustes nos dados a fim de facilitar o manuseio e a importação em outros sistemas:

- Conversão de ISO 8859-1 para UTF-8;
- Apenas variáveis selecionadas foram mantidas;
- Foi mantida aspas nos valores apenas para variáveis alfanuméricas;
- Separador passou a ser vírgula (,) uma vez que não há variáveis com casas decimais que pudessem causar confusão quanto à utilização da vírgula.