

Introdução

Este relatório tem como objetivo apresentar as etapas de carga e transformação dos dados do Perfil do Eleitorado dos estados de Rondônia (RO) e Tocantins (TO), com base em informações disponibilizadas pelo Tribunal Superior Eleitoral (TSE). A análise foi conduzida utilizando dados públicos referentes ao eleitorado para as eleições municipais de 2024, com foco na comparação de perfis entre os dois estados.

Carga e Transformação do Dataset de Perfil do Eleitorado - Estado de Rondônia

1. Carga do Dataset

Fonte dos Dados:

O dataset do **Perfil do Eleitorado** foi obtido através do site de dados abertos do TSE (https://dadosabertos.tse.jus.br/pt_PT/dataset/eleitorado-atual), contendo informações sobre o eleitorado do estado de Rondônia.

Passos de Carga:

- **Importação:** O arquivo CSV correspondente ao eleitorado do estado de Rondônia foi importado para o Microsoft Excel.
- **Linhas e Colunas:** O dataset original possui **775644 linhas** e **30 colunas**, com um total de **23269320 células preenchidas**.

2. Transformação do Dataset

2.1 Eliminação de Colunas (Projeção)

Colunas Mantidas:

1. **NM_MUNICIPIO** - Nome do município
2. **CD_GENERO** - Código de gênero
3. **DS_GENERO** - Descrição de gênero
4. **CD_ESTADO_CIVIL** - Código do estado civil
5. **DS_ESTADO_CIVIL** - Descrição do estado civil
6. **CD_FAIXA_ETARIA** - Código da faixa etária
7. **DS_FAIXA_ETARIA** - Descrição da faixa etária
8. **CD_GRAU_ESCOLARIDADE** - Código do grau de escolaridade
9. **DS_GRAU_ESCOLARIDADE** - Descrição do grau de escolaridade
10. **QT_ELEITORES_PERFIL** - Quantidade de eleitores por perfil

Colunas Removidas:

As colunas que foram **removidas** no passo 1 (eliminação de colunas) por serem desnecessárias para a análise do **Perfil do Eleitorado** são:

- DT_GERACAO
- HH_GERACAO
- ANO_ELEICAO
- SG_UF
- CD_MUNICIPIO
- NR_ZONA
- NR_SECAO
- NR_LOCAL_VOTACAO
- NM_LOCAL_VOTACAO
- CD_RACA_COR
- DS_RACA_COR
- CD_IDENTIDADE_GENERO
- DS_IDENTIDADE_GENERO
- CD_QUILOMBOLA
- DS_QUILOMBOLA
- CD_INTERPRETE_LIBRAS
- DS_INTERPRETE_LIBRAS
- QT_ELEITORES_BIOMETRIA
- QT_ELEITORES_DEFICIENCIA
- QT_ELEITORES_INC_NM_SOCIAL

Resultado:

Após a remoção de colunas, o número de colunas foi reduzido de **30 para 10**, e o total de células preenchidas foi recalculado, resultando em **7756440 células preenchidas**.

2.2 Eliminação de Linhas (Seleção ou Filtro)

Foi aplicado o filtro nas colunas **DS_FAIXA_ETARIA** e **DS_ESTADO_CIVIL** para identificar e remover as linhas com valores "Inválido" ou "NÃO INFORMADO".

- Na coluna **DS_FAIXA_ETARIA**, as linhas com o valor "**Inválido**" foram ocultadas.
- Na coluna **DS_ESTADO_CIVIL**, as linhas com o valor "**NÃO INFORMADO**" foram ocultadas.

Resultado:

Após a eliminação das linhas inválidas:

- O número total de linhas foi reduzido de **775644 para 774922**.
- O novo total de células preenchidas é **7749220**, mantendo as colunas de A até J.

3. Comparação de Tamanho do Dataset

- **Antes da Exclusão:**
 - Número de colunas: **30**
 - Número de linhas: **775644**
 - Total de células com dados: **23269320**
- **Após a Exclusão:**
 - Número de colunas: **10**
 - Número de linhas: **774922**
 - Total de células com dados: **7749220**

Carga e Transformação do Dataset de Perfil do Eleitorado - Estado de Pernambuco

1. Eliminação de Colunas (Projeção)

Objetivo:

Reduzir o número de colunas no dataset de eleitores de Tocantins, mantendo apenas as colunas necessárias para a análise do perfil eleitoral, eliminando as colunas irrelevantes.

Colunas a Manter:

As colunas mantidas para a análise do Perfil do Eleitorado foram as seguintes:

- **NM_MUNICIPIO** - Nome do município
- **CD_GENERO** - Código de gênero
- **DS_GENERO** - Descrição de gênero
- **CD_ESTADO_CIVIL** - Código do estado civil
- **DS_ESTADO_CIVIL** - Descrição do estado civil
- **CD_FAIXA_ETARIA** - Código da faixa etária
- **DS_FAIXA_ETARIA** - Descrição da faixa etária
- **CD_GRAU_ESCOLARIDADE** - Código do grau de escolaridade
- **DS_GRAU_ESCOLARIDADE** - Descrição do grau de escolaridade
- **QT_ELEITORES_PERFIL** - Quantidade de eleitores por perfil

Colunas Removidas:

As colunas removidas no processo de projeção por serem consideradas irrelevantes para a análise foram:

- DT_GERACAO
- HH_GERACAO
- ANO_ELEICAO
- SG_UF
- NR_ZONA
- NR_SECAO
- NR_LOCAL_VOTACAO
- NM_LOCAL_VOTACAO
- CD_RACA_COR
- DS_RACA_COR
- CD_IDENTIDADE_GENERO
- DS_IDENTIDADE_GENERO
- CD_QUILOMBOLA
- DS_QUILOMBOLA
- CD_INTERPRETE_LIBRAS
- DS_INTERPRETE_LIBRAS
- QT_ELEITORES_BIOMETRIA
- QT_ELEITORES_DEFICIENCIA
- QT_ELEITORES_INC_NM_SOCIAL

Verificação do Resultado:

Após a eliminação das colunas desnecessárias, o número de colunas foi reduzido para **10**, com as colunas de A até J. O número total de células com dados foi recalculado, resultando em **7046610 células preenchidas**.

Comparação do Tamanho do Dataset:

- **Antes da Exclusão:**
 - **Número de colunas: 30**
 - **Total de células com dados: 21139830**
- **Depois da Exclusão:**
 - **Número de colunas: 10**
 - **Total de células com dados: 7046610**

2. Eliminação de Linhas (Filtro)

Objetivo:

Remover linhas que continham dados inválidos, como "Inválido" ou "NÃO INFORMADO", nas colunas relevantes, garantindo assim que o dataset contenha apenas dados válidos.

Passos de Filtro Aplicado:

- **Aplicar Filtros:**

- Foi aplicado o filtro nas colunas relevantes do Perfil do Eleitorado, como **DS_FAIXA_ETARIA**, para identificar e remover as linhas com dados inválidos.
- **Identificação de Dados Inválidos:**
 - Na coluna **DS_FAIXA_ETARIA**, as linhas com o valor "**Inválido**" foram ocultadas. Não havia valores "**NÃO INFORMADO**" em outras colunas.
- **Exclusão de Linhas com Dados Inválidos:**
 - Após aplicar o filtro, as linhas com dados inválidos foram ocultadas. As linhas com dados válidos foram selecionadas e mantidas na planilha.

Verificação do Resultado:

Após a exclusão das linhas inválidas:

- O número de linhas foi reduzido para **703221**
- O número total de células preenchidas foi recalculado, resultando em **7032210 células preenchidas**.

Comparação do Tamanho do Dataset:

- **Antes da Exclusão:**
 - Número de linhas: **704661**
 - Total de células com dados: **7046610**
- **Depois da Exclusão:**
 - Número de linhas: **703221**
 - Total de células com dados: **7032210**