



TDC  
ONLINE



# Inovação com dados em nuvem

TRILHA

#TheDevConf  
Oracle

## Gerando valor com dados abertos

Daniel Panizzo  
03.12.20 16h10



# Inovação com dados em nuvem

TRILHA

#TheDevConf  
Oracle



Este trabalho está licenciado sob uma Licença Creative Commons Atribuição-Compartilhamento 4.0 Internacional.

Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by-sa/4.0/>.



# Inovação com dados em nuvem

GERANDO VALOR  
COM DADOS  
ABERTOS





# Gerando valor com dados abertos no Oracle Analytics

---

Daniel Panizzo  
Engenheiro de Soluções

December, 2020, Version 2.0  
Copyright © 2020, Oracle and/or its affiliates  
Public

# ÍNDICE

---

<b>INTRODUÇÃO</b>	<b>6</b>
<b>LAB.1: COLETAR OS DADOS ABERTOS</b>	<b>7</b>
Obtendo os códigos do Github	7
Coletando os dados do TSE	9
Coletando os dados do IBGE	14
<b>LAB.2: ARMAZENAR OS DADOS ABERTOS</b>	<b>16</b>
Carregando os dados no Object Storage	16
Armazenando os dados no banco de dados	23
<b>LAB.3: PREPARAR OS DADOS ABERTOS</b>	<b>28</b>
Preparando os municípios	28
Criando a tabela de principal espectro político	30
<b>LAB.4: ANALISAR OS DADOS ABERTOS</b>	<b>31</b>
Conectando o Analytics com o ADW	31
Importando os metadados	35
Ajustando os metadados	38
Visualizando os dados	40

## INTRODUÇÃO

Neste passo-a-passo, iremos gerar valor com dados abertos e, para isso, passaremos pelo fluxo geral de um projetos de análise de dados. Coletaremos os dados em sua forma primária nos repositórios disponibilizados por organizações públicas. Em seguida, armazenaremos estes dados em um banco de dados. Uma vez armazenados, iremos prepará-los e padronizá-los para, ao final, analisá-los no Oracle Analytics.

Para executar este passo-a-passo, será necessário que você esteja familiarizado com as seguintes linguagens:

- Python
- SQL

E que você tenha as seguintes ferramentas:

- [Jupyter Notebook ou JupyterLab](#) com Python 3.6 ou maior, instalado na sua máquina.
  - Necessários para executar os notebooks com código Python que automatizam a coleta dos dados. Sugerimos instalação da distribuição [Anaconda](#), que já inclui o Jupyter Notebook, JupyterLab, Python 3.8 e diversas bibliotecas padrões para Cientistas de Dados.
- Uma conta no [OCI - Oracle Cloud Infrastructure](#) com:
  - Um Autonomous Data Warehouse, necessário para o armazenamento e preparação dos dados. Veja aqui como provisionar um banco de dados autônomo gratuito no OCI.
- [Oracle Analytics Desktop](#) instalado na sua máquina.
  - Necessário para analisar e visualizar os dados.

Durante este passo-a-passo, acessaremos as seguintes origens de dados abertos:

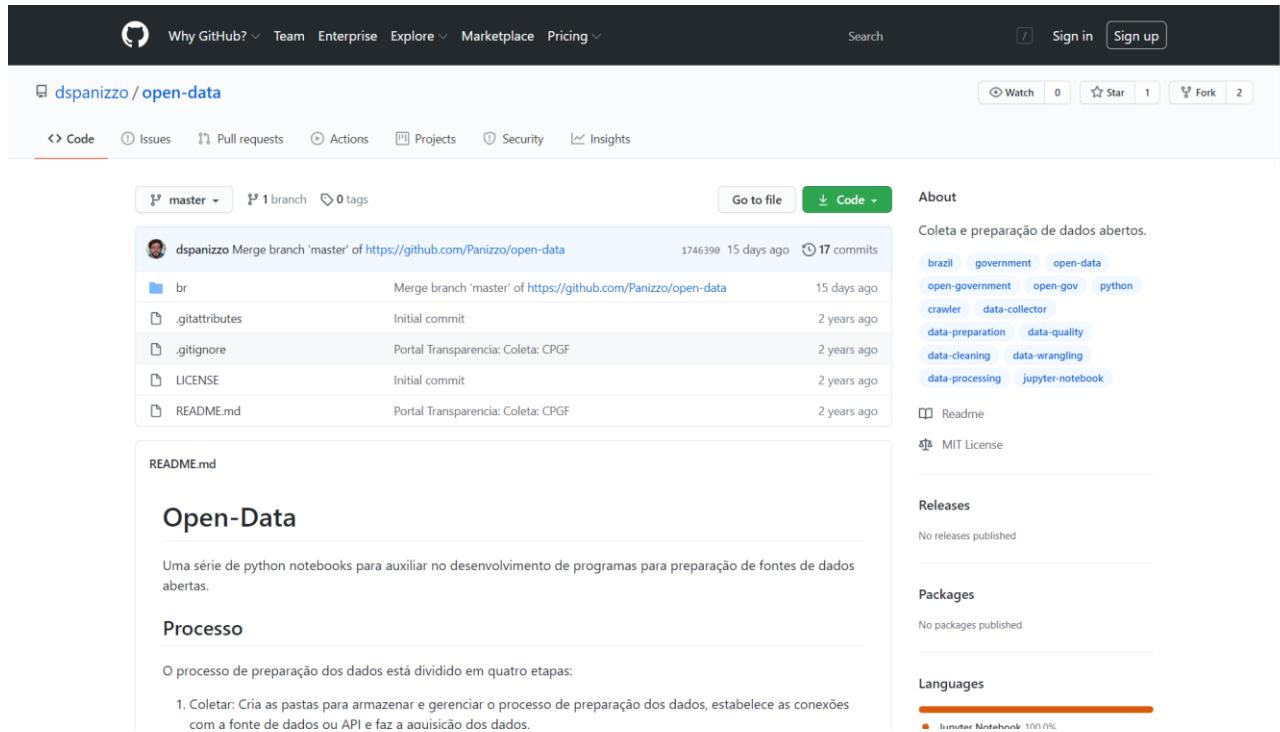
- [Repositório de dados eleitorais do Tribunal Superior Eleitoral](#)
- [API de Serviço de Dados do Instituto Brasileiro de Geografia e Estatística](#)

Outros scripts necessários para a execução deste passo-a-passo estão contidos dentro da pasta “arquivos-de-apoio” dentro do arquivo ZIP.

## LAB.1: COLETAR OS DADOS ABERTOS

### Obtendo os códigos do Github

1. Acesse o repositório [dspanizzo/opendata](https://github.com/dspanizzo/opendata) no Github.



The screenshot shows the GitHub repository page for `dspanizzo/opendata`. The repository has 17 commits, 1 branch, and 0 tags. The README.md file contains the following content:

```
Open-Data

Uma série de python notebooks para auxiliar no desenvolvimento de programas para preparação de fontes de dados abertas.

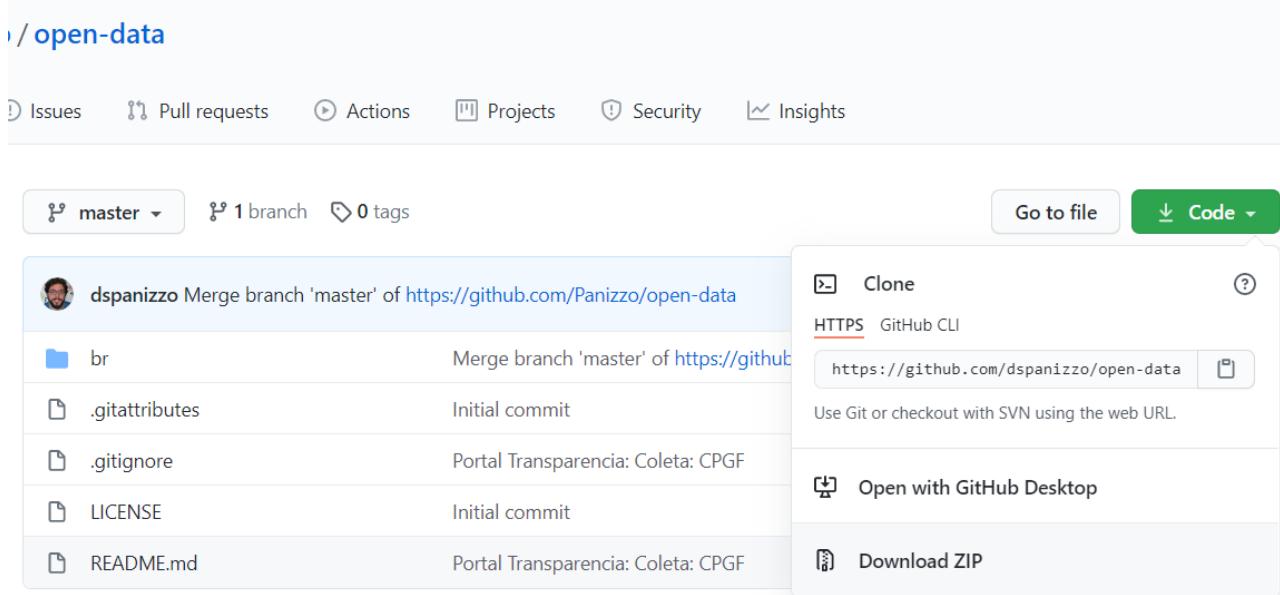
Processo

O processo de preparação dos dados está dividido em quatro etapas:

1. Coletar: Cria as pastas para armazenar e gerenciar o processo de preparação dos dados, estabelece as conexões com a fonte de dados ou API e faz a adquisição dos dados.
```

The 'Code' button is highlighted in green, indicating it was clicked to download the repository as a ZIP file.

2. Clique no botão verde “Code” e selecione a opção “Download ZIP”.



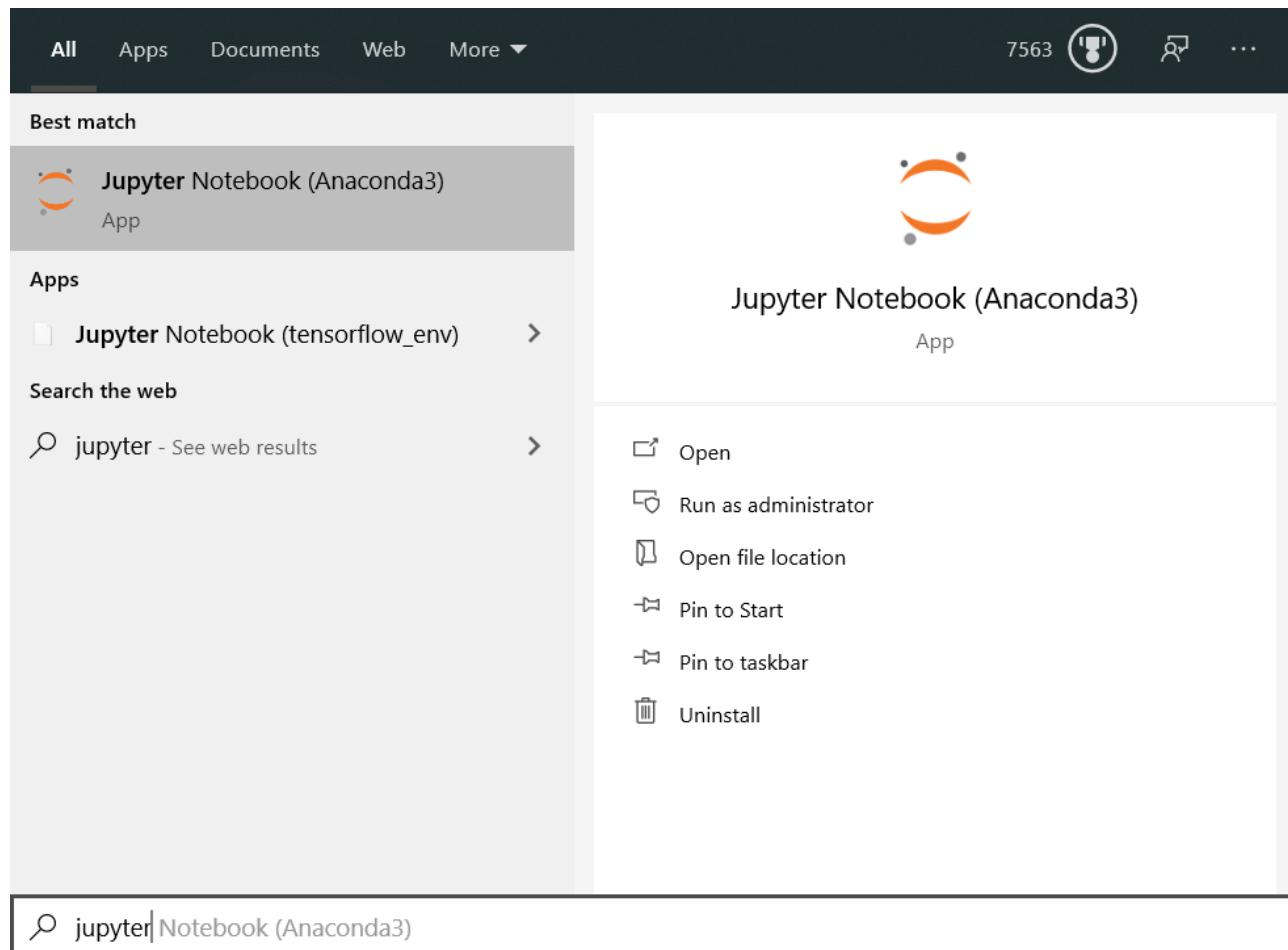
The screenshot shows the GitHub repository page for `dspanizzo/opendata` with the 'Code' button open. The 'Download ZIP' option is selected, and a modal window shows the download progress at 100.0%.

3. Localize no seu computador o local de download do arquivo “open-data-master.zip” e copie-o para outra pasta de sua preferência ( Exemplo: C:\Users\dspanizzo\Documents\TDC ). Em seguida, descompacte-o. Você deverá ver os arquivos listados abaixo.

Documents > TDC > open-data-master >

Name	Date modified	Type	Size
br	12/11/2020 18:15	File folder	
.gitattributes	12/11/2020 18:15	GITATTRIBUTES File	1 KB
.gitignore	12/11/2020 18:15	GITIGNORE File	2 KB
LICENSE	12/11/2020 18:15	File	2 KB
README.md	12/11/2020 18:15	MD File	2 KB

4. No seu computador, abra o Jupyter Notebook ou JupyterLab.



## Coletando os dados do TSE

5. Com o Jupyter aberto no seu browser, navegue pelas pastas até o local da pasta “open-data-master” e, a partir daí, navegue para “br -> tse.jus.br -> votacao-candidato-munzona -> scripts”. Os notebooks dentro desta pasta são responsáveis por coletar e preparar os dados oriundos do Repositório de Dados Eleitorais do TSE.

The screenshot shows the Jupyter Notebook interface with the title bar "jupyter". Below it is a navigation bar with tabs: "Files", "Running", and "Clusters". A message "Select items to perform actions on them." is displayed above a file list. The file list shows the following structure and details:

	Name	Last Modified	File size
0	01-data-collector.ipynb	15 dias atrás	12 kB
1	02-data-cleaner.ipynb	15 dias atrás	19.8 kB
2	03-data-organizer.ipynb	15 dias atrás	11.2 kB
3	04-data-standardizer.ipynb	15 dias atrás	29.5 kB

A message "poucos segundos atrás" is visible above the file list.

6. Clique no notebook “01-data-collector.ipynb” para abrí-lo em uma nova guia do browser.

The screenshot shows the Jupyter Notebook interface with the title bar "jupyter 01-data-collector (autosaved)". Below it is a toolbar with buttons for File, Edit, View, Insert, Cell, Kernel, Widgets, Help, and a Trusted status indicator. The main area contains two code cells:

```
In [1]: import os as os
import requests as req
import zipfile as zipfile
from datetime import datetime
from dateutil.relativedelta import relativedelta

In [2]: def data_directories():
    """
    Funcao de criacao das pastas para armazenamento dos dados, caso nao existam.

    # Directories para armazenamento dos dados coletados
    directories = ['..\data\01-collected',
                  '..\data\02-cleaned',
                  '..\data\03-organized',
                  '..\data\04-standardized']

    # Log: Mensagem de inicio da criacao dos diretorios
    print(str(datetime.now()) + ': Criacao dos diretorios para armazenamento dos dados iniciada.')

    # Para cada diretório na lista de diretórios
    for directory in directories:

        # Verifica se o diretório existe
        if not os.path.exists(directory):

            # Cria o diretório
            os.makedirs(directory)
```

7. Este é o notebook responsável por coletar os arquivos do repositório do TSE. Execute todas as células de código na sequência em que foram criadas (de cima para baixo) clicando no botão “> Run”. A última célula (com a função “main”) será responsável por chamar as demais funções e executar todo o processo. Caso queira verificar os detalhes, todo o código está comentado para que possa compreender cada etapa executada.

Ao finalizar o processo, a última mensagem exibida pelo log da função “main” deverá ser “Coleta dos arquivos finalizada”.

The screenshot shows a Jupyter Notebook interface with the title "jupyter 01-data-collector Last Checkpoint: 11 minutos atrás (autosaved)". The menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Trusted, and Python 3. The toolbar has icons for file operations like Open, Save, and Run. The code cell (In [5]) contains the following Python code:

```

def main():
    """
    Coleta os dados de resultados das eleicoes por candidato, municipio e zona.
    Origem: Tribunal Superior Eleitoral (TSE)
    """

    # Prepara o dicionario de variaveis (kwargs = keyworded arguments)
    kwargs = {}

    # Datas de Inicio e Fim (YYYY-MM-DD)
    # Obs.: Dados disponiveis a partir de 1994-01-01
    kwargs['dt_ini_ref'] = datetime.strptime('2016-01-01', "%Y-%m-%d")
    kwargs['dt_fim_ref'] = datetime.strptime('2016-01-01', "%Y-%m-%d")

    # Coleta os arquivos
    collect_data(**kwargs)

if __name__ == '__main__':
    main()

```

The output cell shows the execution log:

```

2020-11-27 17:52:19.531539: Criacao dos diretorios para armazenamento dos dados iniciada.
2020-11-27 17:52:19.535527: Diretorio ..\data\01-collected criado.
2020-11-27 17:52:19.537524: Diretorio ..\data\02-cleaned criado.
2020-11-27 17:52:19.539518: Diretorio ..\data\03-organized criado.
2020-11-27 17:52:19.541511: Diretorio ..\data\04-standardized criado.
2020-11-27 17:52:19.541511: Criacao dos diretorios para armazenamento dos dados finalizada.
2020-11-27 17:52:19.541511: Coleta dos arquivos iniciada.
2020-11-27 17:52:19.541511: Inicio do download do arquivo ..\data\01-collected\votacao_candidato_munzona_2016.zip
2020-11-27 17:52:19.861948: URL encontrada, fazendo o download do arquivo: ..\data\01-collected\votacao_candidato_munzona_2016.zip
2020-11-27 17:52:49.554205: Descompactando o arquivo ..\data\01-collected\votacao_candidato_munzona_2016.zip
2020-11-27 17:52:53.330453: Fim do download do arquivo ..\data\01-collected\votacao_candidato_munzona_2016.zip
2020-11-27 17:52:53.331451: Verificando a existencia de arquivos nao relacionados com os dados.
2020-11-27 17:52:53.333445: Arquivo ..\data\01-collected\leiaime.pdf removido.
2020-11-27 17:52:53.333445: Remocao dos arquivos nao relacionados com os dados concluida.
2020-11-27 17:52:53.334443: Coleta dos arquivos finalizada.

```

8. Retorne para a guia de navegação de diretórios do Jupyter e clique no notebook “02-data-cleaner.ipynb” para abri-lo em uma nova guia do browser.

The screenshot shows the Jupyter Notebook interface with the title "jupyter". The top navigation bar includes Quit and Logout. Below it is a file list with tabs for Files, Running, and Clusters. The Files tab is selected, showing a list of files in the directory structure:

- 01-data-collector.ipynb (Running, 7 minutos atrás, 12 kB)
- 02-data-cleaner.ipynb (15 dias atrás, 19.8 kB)
- 03-data-organizer.ipynb (15 dias atrás, 11.2 kB)
- 04-data-standardizer.ipynb (15 dias atrás, 29.5 kB)

At the bottom right of the file list are buttons for Upload, New, and Refresh.

9. Este é o notebook responsável por realizar a limpeza dos dados, eliminando caracteres desconhecidos e codificando o arquivo no padrão UTF-8.

```

In [1]: import os
import codecs
from datetime import datetime

In [2]: def encode_to_utf8(**kwargs):
    """
    Funcao para padronizar a codificacao dos arquivos originais em UTF-8.
    Facilita a futura leitura dos arquivos e elimina caracteres desconhecidos.

    Args:
        original_encode(str): Codificacao original do arquivo.
            Ex.: 'cp1252'
        in_data_path(str): Caminho completo para acesso ao arquivo original.
            Ex.: '..\data\01-collected\2012_Funcoes.csv'
        out_data_path(str): Caminho completo para escrita do arquivo codificado
            em UTF-8.
            Ex.: '..\data\02-cleaned\2012_Funcoes.csv'
    """

    # Log: Mensagem de inicio da codificacao
    print(str(datetime.now()) + ': Inicio da codificacao do arquivo ' + kwargs['in_data_path'])

    # Le arquivo na codificacao original
    with codecs.open(kwargs['in_data_path'], 'rb', encoding = kwargs['original_encode'], errors = 'replace') as file:
        # Substitui os caracteres desconhecidos por interrogacao
        lines = file.read().replace('\x1a', '?').replace('\u0085', '?')

    # Escreve o arquivo codificado em UTF-8
    with codecs.open(kwargs['out_data_path'], 'w', encoding = 'utf-8') as file:
        file.write(lines)

    # Log: Mensagem de termino da codificacao
    print(str(datetime.now()) + ': Fim da codificacao do arquivo ' + kwargs['out_data_path'])

```

10. Mais uma vez, execute todas as células de código na sequência em que foram criadas (de cima para baixo) clicando no botão “> Run”. Ao finalizar o processo, a última mensagem exibida pelo log da função “main” deverá ser “Limpeza dos arquivos finalizada”.

```

2020-11-27 18:04:19.468296: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_MS.csv
2020-11-27 18:04:19.489241: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_MS.csv
2020-11-27 18:04:19.490237: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_MT.csv
2020-11-27 18:04:19.511184: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_MT.csv
2020-11-27 18:04:19.511184: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_PA.csv
2020-11-27 18:04:19.550078: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_PA.csv
2020-11-27 18:04:19.551075: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_PB.csv
2020-11-27 18:04:19.571021: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_PB.csv
2020-11-27 18:04:19.572019: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_PE.csv
2020-11-27 18:04:19.617896: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_PE.csv
2020-11-27 18:04:19.619892: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_PI.csv
2020-11-27 18:04:19.637843: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_PI.csv
2020-11-27 18:04:19.638841: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_PR.csv
2020-11-27 18:04:19.706659: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_PR.csv
2020-11-27 18:04:19.709652: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_RJ.csv
2020-11-27 18:04:20.057906: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_RJ.csv
2020-11-27 18:04:20.065885: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_RN.csv
2020-11-27 18:04:20.087829: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_RN.csv
2020-11-27 18:04:20.088826: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_RO.csv
2020-11-27 18:04:20.103784: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_RO.csv
2020-11-27 18:04:20.103784: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_RR.csv
2020-11-27 18:04:20.110765: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_RR.csv
2020-11-27 18:04:20.110765: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_RS.csv
2020-11-27 18:04:20.159178: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_RS.csv
2020-11-27 18:04:20.162171: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_SC.csv
2020-11-27 18:04:20.193088: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_SC.csv
2020-11-27 18:04:20.194084: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_SE.csv
2020-11-27 18:04:20.204057: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_SE.csv
2020-11-27 18:04:20.205054: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_SP.csv
2020-11-27 18:04:20.556770: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_SP.csv
2020-11-27 18:04:20.565748: Inicio da codificacao do arquivo ..\data\01-collected\votacao_candidato_munzona_2016_TO.csv
2020-11-27 18:04:20.577715: Fim da codificacao do arquivo ..\data\02-cleaned\votacao_candidato_munzona_2016_TO.csv
2020-11-27 18:04:20.578713: Limpeza dos arquivos finalizada.

```

11. Retorne para a guia de navegação de diretórios do Jupyter e clique no notebook “03-data-organizer.ipynb” para abri-lo em uma nova guia do browser. Repita as mesmas etapas dos itens anteriores. Este notebook é responsável por estruturar o arquivo em formato tabular (caso seja necessário) e confirmar que todos os registros do conjunto de dados possuem a mesma quantidade de atributos. Ao finalizar o processo, a última mensagem exibida pelo log da função “main” deverá ser “Organizacao dos arquivos finalizada”.

```

2020-11-30 07:14:03.645501: Verificando registros do arquivo votacao_candidato_munzona_2016_PB.csv
2020-11-30 07:14:03.921763: Resultado: Normais = 14765 - Corrigidas = 0
2020-11-30 07:14:03.921763: Verificando registros do arquivo votacao_candidato_munzona_2016_PE.csv
2020-11-30 07:14:04.549316: Resultado: Normais = 37543 - Corrigidas = 0
2020-11-30 07:14:04.549656: Verificando registros do arquivo votacao_candidato_munzona_2016_PI.csv
2020-11-30 07:14:04.747638: Resultado: Normais = 12470 - Corrigidas = 0
2020-11-30 07:14:04.747638: Verificando registros do arquivo votacao_candidato_munzona_2016_PR.csv
2020-11-30 07:14:05.588962: Resultado: Normais = 49842 - Corrigidas = 0
2020-11-30 07:14:05.588962: Verificando registros do arquivo votacao_candidato_munzona_2016_RJ.csv
2020-11-30 07:14:09.332163: Resultado: Normais = 214813 - Corrigidas = 0
2020-11-30 07:14:09.333134: Verificando registros do arquivo votacao_candidato_munzona_2016_RN.csv
2020-11-30 07:14:09.584102: Resultado: Normais = 11214 - Corrigidas = 0
2020-11-30 07:14:09.584856: Verificando registros do arquivo votacao_candidato_munzona_2016_RO.csv
2020-11-30 07:14:09.728716: Resultado: Normais = 7703 - Corrigidas = 0
2020-11-30 07:14:09.728716: Verificando registros do arquivo votacao_candidato_munzona_2016_RR.csv
2020-11-30 07:14:09.783569: Resultado: Normais = 1952 - Corrigidas = 0
2020-11-30 07:14:09.783569: Verificando registros do arquivo votacao_candidato_munzona_2016_RS.csv
2020-11-30 07:14:10.437845: Resultado: Normais = 38030 - Corrigidas = 0
2020-11-30 07:14:10.437845: Verificando registros do arquivo votacao_candidato_munzona_2016_SC.csv
2020-11-30 07:14:10.814874: Resultado: Normais = 21792 - Corrigidas = 0
2020-11-30 07:14:10.815806: Verificando registros do arquivo votacao_candidato_munzona_2016_SE.csv
2020-11-30 07:14:10.927507: Resultado: Normais = 6595 - Corrigidas = 0
2020-11-30 07:14:10.927507: Verificando registros do arquivo votacao_candidato_munzona_2016_SP.csv
2020-11-30 07:14:13.896065: Resultado: Normais = 200913 - Corrigidas = 0
2020-11-30 07:14:13.896065: Verificando registros do arquivo votacao_candidato_munzona_2016_TO.csv
2020-11-30 07:14:14.003630: Resultado: Normais = 7003 - Corrigidas = 0
2020-11-30 07:14:14.003743: Organizacao dos arquivos finalizada.

```

12. Retorne para a guia de navegação de diretórios do Jupyter e clique no notebook “04-data-standardizer.ipynb” para abri-lo em uma nova guia do browser. Repita as mesmas etapas dos itens anteriores. Este notebook é responsável por padronizar os dados e metadados do conjunto de dados. Ao finalizar o processo, a última mensagem exibida pelo log da função “main” deverá ser “Padronizacao dos arquivos finalizada”.

```

2020-11-30 07:19:19.805642: Arquivo votacao_candidato_munzona_2016_AL.csv padronizado.
2020-11-30 07:19:29.802851: Arquivo votacao_candidato_munzona_2016_AM.csv padronizado.
2020-11-30 07:19:30.640604: Arquivo votacao_candidato_munzona_2016_AP.csv padronizado.
2020-11-30 07:19:53.897996: Arquivo votacao_candidato_munzona_2016_BA.csv padronizado.
2020-11-30 07:26:33.347305: Arquivo votacao_candidato_munzona_2016_BRASIL.csv padronizado.
2020-11-30 07:26:45.451411: Arquivo votacao_candidato_munzona_2016_CE.csv padronizado.
2020-11-30 07:26:51.125597: Arquivo votacao_candidato_munzona_2016_ES.csv padronizado.
2020-11-30 07:27:03.355103: Arquivo votacao_candidato_munzona_2016_GO.csv padronizado.
2020-11-30 07:27:13.851176: Arquivo votacao_candidato_munzona_2016_MA.csv padronizado.
2020-11-30 07:28:06.173311: Arquivo votacao_candidato_munzona_2016_MG.csv padronizado.
2020-11-30 07:28:12.873878: Arquivo votacao_candidato_munzona_2016_MS.csv padronizado.
2020-11-30 07:28:19.0005349: Arquivo votacao_candidato_munzona_2016_MT.csv padronizado.
2020-11-30 07:28:32.001166: Arquivo votacao_candidato_munzona_2016_PA.csv padronizado.
2020-11-30 07:28:38.436957: Arquivo votacao_candidato_munzona_2016_PB.csv padronizado.
2020-11-30 07:28:54.373115: Arquivo votacao_candidato_munzona_2016_PE.csv padronizado.
2020-11-30 07:28:59.663889: Arquivo votacao_candidato_munzona_2016_PI.csv padronizado.
2020-11-30 07:29:22.740089: Arquivo votacao_candidato_munzona_2016_PR.csv padronizado.
2020-11-30 07:30:54.897173: Arquivo votacao_candidato_munzona_2016_RJ.csv padronizado.
2020-11-30 07:30:59.607727: Arquivo votacao_candidato_munzona_2016_RN.csv padronizado.
2020-11-30 07:31:03.646350: Arquivo votacao_candidato_munzona_2016_RO.csv padronizado.
2020-11-30 07:31:04.759806: Arquivo votacao_candidato_munzona_2016_RR.csv padronizado.
2020-11-30 07:31:20.914782: Arquivo votacao_candidato_munzona_2016_RS.csv padronizado.
2020-11-30 07:31:30.794182: Arquivo votacao_candidato_munzona_2016_SC.csv padronizado.
2020-11-30 07:31:34.283731: Arquivo votacao_candidato_munzona_2016_SE.csv padronizado.
2020-11-30 07:33:04.735202: Arquivo votacao_candidato_munzona_2016_SP.csv padronizado.
2020-11-30 07:33:07.912923: Arquivo votacao_candidato_munzona_2016_TO.csv padronizado.
2020-11-30 07:33:07.915916: Padronizacao dos arquivos finalizada.

```

13. Ao finalizar a execução dos quatro notebooks, você irá notar que, dentro da pasta “br/tse.jus.br/votacao-candidato-munzona”, temos uma nova pasta chamada “data” com pastas referentes a cada script que executamos.

uments > TDC > open-data-master > br > tse.jus.br > votacao-candidato-munzona > data >

Name	Date modified	Type	Size
01-collected	27/11/2020 17:52	File folder	
02-cleaned	27/11/2020 18:04	File folder	
03-organized	30/11/2020 07:14	File folder	
04-standardized	30/11/2020 07:26	File folder	

14. Dentro das pastas estão todos os conjuntos de dados que foram preparados durante as execuções dos notebooks. Caso queira, você pode acessar cada pasta e validar como cada notebook preparou os dados em cada etapa.

uments > TDC > open-data-master > br > tse.jus.br > votacao-candidato-munzona > data > 04-standardized

Name	Date modified	Type	Size
votacao_candidato_munzona_2016_AC.csv	30/11/2020 07:19	Microsoft Excel Co...	1.159 KB
votacao_candidato_munzona_2016_AL.csv	30/11/2020 07:19	Microsoft Excel Co...	3.373 KB
votacao_candidato_munzona_2016_AM.csv	30/11/2020 07:19	Microsoft Excel Co...	10.065 KB
votacao_candidato_munzona_2016_AP.csv	30/11/2020 07:19	Microsoft Excel Co...	860 KB
votacao_candidato_munzona_2016_BA.csv	30/11/2020 07:19	Microsoft Excel Co...	24.475 KB
votacao_candidato_munzona_2016_BRAS...	30/11/2020 07:26	Microsoft Excel Co...	413.144 KB
votacao_candidato_munzona_2016_CE.csv	30/11/2020 07:26	Microsoft Excel Co...	11.763 KB
votacao_candidato_munzona_2016_ES.csv	30/11/2020 07:26	Microsoft Excel Co...	5.674 KB
votacao_candidato_munzona_2016_GO.csv	30/11/2020 07:27	Microsoft Excel Co...	12.482 KB
votacao_candidato_munzona_2016_MA.csv	30/11/2020 07:27	Microsoft Excel Co...	10.627 KB
votacao_candidato_munzona_2016_MG.csv	30/11/2020 07:28	Microsoft Excel Co...	46.736 KB
votacao_candidato_munzona_2016_MS.csv	30/11/2020 07:28	Microsoft Excel Co...	4.726 KB
votacao_candidato_munzona_2016_MT.csv	30/11/2020 07:28	Microsoft Excel Co...	5.853 KB
votacao_candidato_munzona_2016_PA.csv	30/11/2020 07:28	Microsoft Excel Co...	12.235 KB
votacao_candidato_munzona_2016_PB.csv	30/11/2020 07:28	Microsoft Excel Co...	6.350 KB
votacao_candidato_munzona_2016_PE.csv	30/11/2020 07:28	Microsoft Excel Co...	15.694 KB
votacao_candidato_munzona_2016_PI.csv	30/11/2020 07:28	Microsoft Excel Co...	5.384 KB
votacao_candidato_munzona_2016_PR.csv	30/11/2020 07:29	Microsoft Excel Co...	20.776 KB
votacao_candidato_munzona_2016_RJ.csv	30/11/2020 07:30	Microsoft Excel Co...	89.160 KB
votacao_candidato_munzona_2016_RN.csv	30/11/2020 07:30	Microsoft Excel Co...	4.744 KB
votacao_candidato_munzona_2016_RO.csv	30/11/2020 07:31	Microsoft Excel Co...	3.256 KB
votacao_candidato_munzona_2016_RR.csv	30/11/2020 07:31	Microsoft Excel Co...	817 KB
votacao_candidato_munzona_2016_RS.csv	30/11/2020 07:31	Microsoft Excel Co...	15.898 KB
votacao_candidato_munzona_2016_SC.csv	30/11/2020 07:31	Microsoft Excel Co...	9.074 KB
votacao_candidato_munzona_2016_SE.csv	30/11/2020 07:31	Microsoft Excel Co...	2.834 KB
votacao_candidato_munzona_2016_SP.csv	30/11/2020 07:33	Microsoft Excel Co...	82.261 KB
votacao_candidato_munzona_2016_TO.csv	30/11/2020 07:33	Microsoft Excel Co...	3.055 KB

15. Assim, finalizamos a coleta de dados do TSE.

## Coletando os dados do IBGE

1. Com o Jupyter aberto no seu browser, navegue pelas pastas até o local da pasta “open-data-master” e, a partir daí, navegue para “br -> servicodados.ibge.gov.br -> localidades -> scripts”. Os notebooks dentro desta pasta são responsáveis por coletar e preparar os dados oriundos da [API de Serviço de Dados do IBGE](#).

The screenshot shows a Jupyter Notebook interface with a sidebar on the left and a main content area. The sidebar has tabs for 'Files', 'Running', and 'Clusters'. The main content area shows a file list with the following details:

Name	Last Modified	File size
01-data-collector.ipynb	18 dias atrás	8.67 kB
02-data-cleaner.ipynb	18 dias atrás	5.2 kB
03-data-organizer.ipynb	18 dias atrás	9.21 kB
04-data-standardizer.ipynb	18 dias atrás	17.6 kB

2. Assim como na etapa anterior “Coletando dados do TSE”, repita a mesma sequência de execução dos quatro notebooks dentro desta pasta. O resultado final deverá ser um arquivo de municípios do IBGE (Códigos e nomes de municípios, Estados e Regiões) dentro da pasta “br -> servicodados.ibge.gov.br -> localidades -> data -> 04-standardizer”.

The screenshot shows a file browser interface with a navigation bar at the top. Below the navigation bar, there is a table with the following data:

Name	Date modified	Type	Size
20201130_municipios.csv	30/11/2020 07:47	Microsoft Excel Co...	735 KB

3. Outro arquivo necessário para a futura análise de dados é o de [Malha Geográfica do IBGE](#). É possível visualizar o arquivo que queremos buscar ao acessar a seguinte URL da API:

<https://servicodados.ibge.gov.br/api/v2/malhas/BR?resolucao=5&qualidade=1>

The screenshot shows a web browser with three tabs open. The active tab displays a map of Brazil's administrative divisions, specifically showing the state of São Paulo with its various municipalities outlined in black. The browser's address bar shows the URL: https://servicodados.ibge.gov.br/api/v2/malhas/BR?resolucao=5&qualidade=1

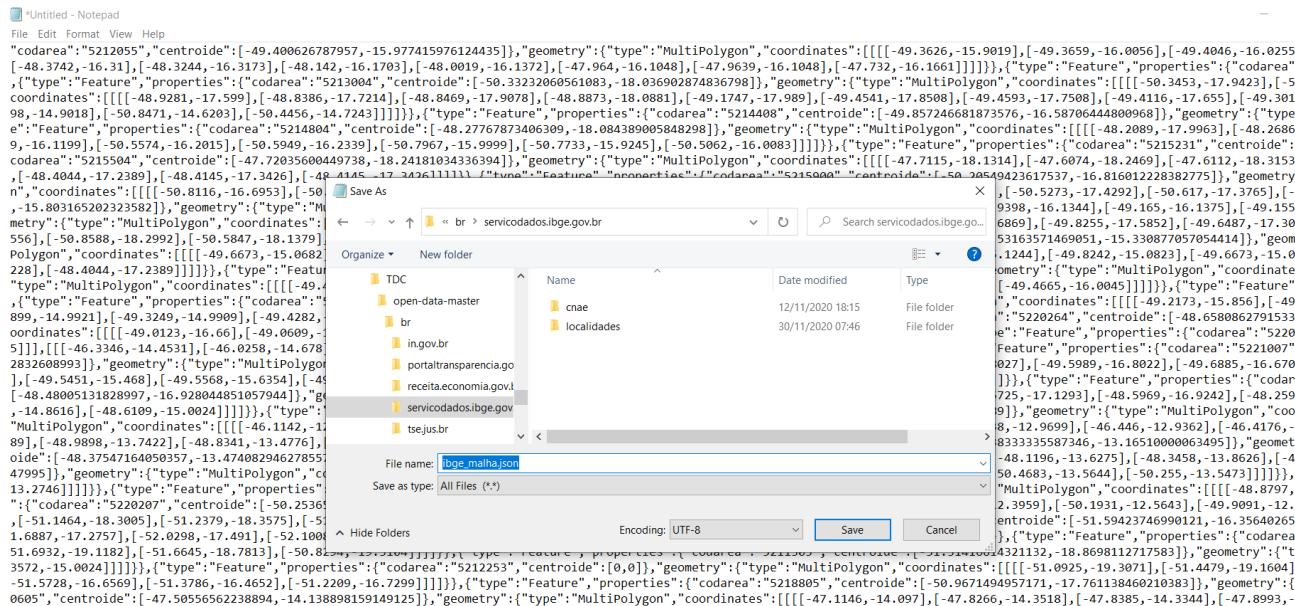


4. Para realizar o acesso a este arquivo em formato GeoJSON, utilize a URL abaixo:

<https://servicodados.ibge.gov.br/api/v2/malhas/BR?resolucao=5&qualidade=1&formato=application/vnd.geo+json>

```
{ "type": "FeatureCollection", "crs": { "type": "name", "properties": { "name": "urn:ogc:def:crs:EPSG:4674" } }, "features": [ { "type": "Feature", "properties": { "codarea": "1100015", "centroide": "[ -62.28182737333638, -12.459978797273866 ]", "geometry": { "type": "Polygon", "coordinates": [ [ [ -62.0584, -11.8673 ], [ -61.8905, -11.9068 ], [ -61.8884, -11.9657 ], [ -62.0006, -12.7921 ], [ -62.4177, -12.2802, -12.5298 ], [ -62.6425, -12.2371 ], [ -62.453, -11.9685 ], [ -62.2067, -11.8319 ], [ -62.0504, -11.8673 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1100023", "centroide": "[ -62.922709055488, -9.86705808864132 ]", "geometry": { "type": "Polygon", "coordinates": [ [ [ -62.5547, -9.6767 ], [ -62.4226, -9.8084 ], [ -62.6699, -10.2882 ], [ -62.6707, -10.4596, -10.0003 ], [ -63.1015, -9.7525 ], [ -62.5547, -9.6707 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1100031", "centroide": "[ -60.66608883210166, -13.4751877693604 ]", "geometry": { "type": "Polygon", "coordinates": [ [ [ -60.9189, -13.5483 ], [ -60.7051, -13.3247 ], [ -60.3716, -13.3185 ], [ -60.7093, -13.6931 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1100049", "centroide": "[ -61.3887907239486, -11.289319572705616 ]", "geometry": { "type": "Polygon", "coordinates": [ [ [ -61.0005, -10.9922 ], [ -61.2117, -11.5291 ], [ -61.5489, -11.6191 ], [ -61.6142, -11.5838 ], [ -61.7957, -10.9875 ], [ -61.0005, -10.9922 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1100056", "centroide": "[ -61.5009, -13.0843 ], [ -60.7429, -13.1223 ]", "geometry": { "type": "Polygon", "coordinates": [ [ [ -60.3311, -13.2579 ], [ -60.3176, -13.3185 ], [ -60.7051, -13.3247 ], [ -60.7415, -13.3249 ], [ -61.9044, -13.2558 ], [ -61.5009, -13.0843 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1100064", "centroide": "[ -60.57367948243636, -13.197209995154414 ]", "geometry": { "type": "Polygon", "coordinates": [ [ [ -60.3311, -13.2579 ], [ -60.3176, -13.3185 ], [ -60.7051, -13.3247 ], [ -60.7415, -13.3249 ], [ -61.9044, -13.2558 ], [ -61.5009, -13.0843 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1100087", "centroide": "[ -61.109951187605205, -12.915424597300866 ]", "geometry": { "type": "Polygon", "coordinates": [ [ [ -60.652, -19.8183 ], [ -60.7429, -13.1223 ], [ -61.5009, -13.0043 ], [ -61.3388, -12.6265 ], [ -60.652, -12.9813 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1100098", "centroide": "[ -60.792094402427536, -11.257143568814529 ]", "geometry": { "type": "Polygon", "coordinates": [ [ [ -63.644, -11.839 ], [ -64.0515, -12.2186 ], [ -63.6847, -12.453 ], [ -64.5945, -12.6121 ], [ -61.5291 ], [ -61.0005, -10.9922 ], [ -60.46, -10.9898 ], [ -60.4585, -10.9893 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1101006", "centroide": "[ -64.5293646232689, -11.3257916 ]", "geometry": { "type": "Polygon", "coordinates": [ [ [ -63.7067, -10.8399 ], [ -63.6627, -11.005 ], [ -63.627, -11.055 ], [ -63.6048, -11.2156 ], [ -65.0289, -11.9976 ], [ -65.364, -10.7615 ], [ -63.7067, -10.8399 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1101014", "centroide": "[ -62.62248513565204, -10.58002309623475 ]", "geometry": { "type": "Polygon", "coordinates": [ [ [ -62.1346, -10.2247 ], [ -62.2596, -10.4768 ], [ -62.5252, -10.7094 ], [ -62.7171, -10.8782 ], [ -63.1426, -11.0713 ], [ -62.788, -10.4924 ], [ -62.6707, -10.2877 ], [ -62.6699, -10.2882 ], [ -62.1346, -10.2247 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1101022", "centroide": "[ -61.78137504691506, -10.46259646073085 ]", "geometry": { "type": "Polygon", "coordinates": [ [ [ -61.632, -10.9862 ], [ -62.1444, -11.0904 ], [ -62.156, -11.0259 ], [ -62.1225, -10.8993 ], [ -61.9105, -10.117 ], [ -61.9726, -10.0675 ], [ -61.9544, -9.9496 ], [ -61.5746, -9.7177 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1101030", "centroide": "[ -62.061512416619784, -9.18082807082085 ]", "geometry": { "type": "Polygon", "coordinates": [ [ [ -61.903, -8.8628 ], [ -61.7132, -8.8628 ], [ -61.5831, -8.7986 ], [ -61.5831, -8.7986 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1101048", "centroide": "[ -62.2426, -9.8046 ], [ -62.5547, -9.6707 ], [ -62.4224, -9.525 ], [ -62.3691, -8.7435 ], [ -62.2688, -8.5797 ], [ -61.903, -8.8628 ] ], "geometry": { "type": "Polygon", "coordinates": [ [ [ -61.9524, -11.2904 ], [ -61.9524, -11.2904 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1101055", "centroide": "[ -62.20032917056549, -11.0861483099515 ], "geometry": { "type": "Polygon", "coordinates": [ [ [ -62.3524, -10.8939 ], [ -62.5252, -10.7094 ], [ -62.2596, -10.4768 ], [ -61.9105, -10.117 ], [ -62.1225, -10.8993 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1101081", "centroide": "[ -60.414828644273, -11.853621987930085 ], "geometry": { "type": "Polygon", "coordinates": [ [ [ -60.46433, -11.4648 ], [ -60.3022, -12.2584 ], [ -60.6499, -12.2613 ], [ -61.1806, -12.1671 ], [ -61.1901, -12.1671 ], [ -61.5345, -11.7431 ], [ -61.5489, -11.6191 ], [ -61.2117, -11.5291 ], [ -60.46433, -11.4648 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1102025", "centroide": "[ -64.28441204528865, -9.158011 ], "geometry": { "type": "Polygon", "coordinates": [ [ [ -62.7439, -8.0452 ], [ -62.738, -8.05 ], [ -62.7276, -8.0624 ], [ -62.7246, -8.0691 ], [ -62.2688, -8.5757 ], [ -62.3691, -8.7435 ], [ -62.7641, -8.798 ], [ -62.8223, -8.798 ], [ -63.7568, -9.3951 ], [ -63.6852, -9.8329 ], [ -64.3411, -10.0003 ], [ -65.333, -9.9442 ], [ -65.4434, -9.6693 ], [ -66.6267, -9.8981 ], [ -66.8102, -9.8181 ], [ -66.4088, -9.4069 ], [ -65.7299, -9.563 ], [ -63.6698, -8.1217 ], [ -63.6624, -8.1011 ], [ -63.6217, -8.0296 ], [ -63.6264, -8.0161 ], [ -63.6214, -7.9765 ], [ -62.7439, -8.0452 ] ] ] } }, "type": "Feature", "properties": { "codarea": "1102054", "centroide": "[ -61.7957, -11.3334 ], [ -61.7957, -11.3334 ], [ -61.9524, -11.2904 ], [ -62.0932, -11.2958 ], [ -62.1592 ] }
```

5. Selecione todo o conteúdo retornado do acesso a esta URL, copie para um editor de texto (Ex.: Notepad), e salve como um arquivo com extensão “.json”.

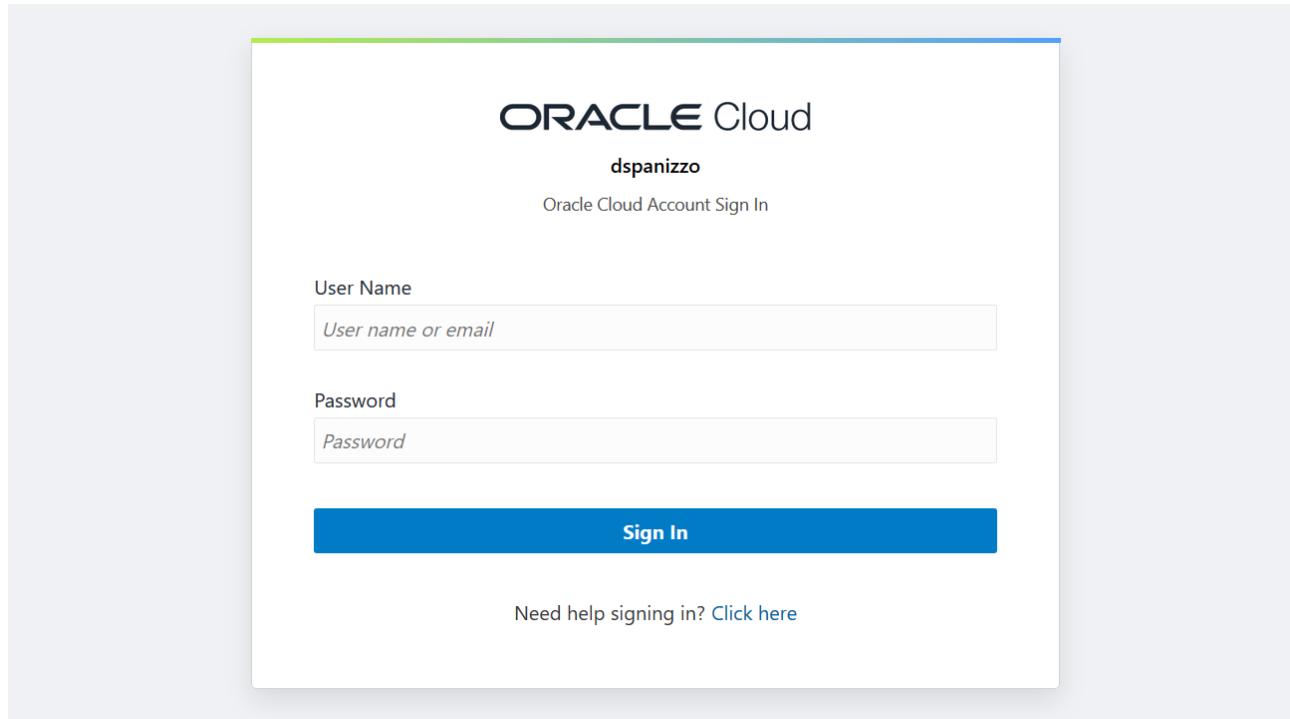


6. Assim, finalizamos a colteia de dados do IBGE.

## LAB.2: ARMAZENAR OS DADOS ABERTOS

### Carregando os dados no Object Storage

1. Com os dados coletados, a próxima etapa será armazenar os dados no banco de dados. Acesse sua conta do Oracle Cloud Infrastructure (OCI) em <https://console.sa-saopaulo-1.oraclecloud.com/>. Caso ainda não tenha uma conta gratuita no OCI, acesse <https://www.oracle.com/br/cloud/free/> para criar uma.



2. No Menu de navegação do OCI, selecione “Armazenamento de Objetos -> Armazenamento do Objetos” entre as opções de “Infraestrutura Básica”.

A screenshot of the Oracle Cloud Infrastructure (OCI) navigation menu. The left sidebar lists various service categories under 'Infraestrutura Básica', including 'Computação', 'Armazenamento em Blocos', 'Armazenamento de Objetos', 'Armazenamento de Arquivos', 'Rede', 'Oracle Database', 'Visão Geral', 'Autonomous Data Warehouse', 'Autonomous JSON Database', 'Autonomous Transaction Processing', 'Bare Metal, VM e Exadata', 'Exadata Cloud@Customer', 'Banco de Dados', 'MySQL', 'Banco de Dados NoSQL', 'Serviços Relacionados ao Banco de Dados', 'Data Safe', 'Dados e IA', 'Big Data', and 'Assistente Digital'. The 'Armazenamento de Objetos' option is highlighted with a teal bar. The main content area shows a 'Controle' section with a card for 'Armazenamento de Objetos' featuring options like 'Criar um banco de dados ATP', 'Criar uma pilha', and 'Pré-instalar Kit de Desenvolvimento do Oracle Cloud'. To the right, there are sections for 'Todos os sistemas operacionais', 'Centro de Contas', 'Upgrade', and 'Novidades'.

3. Clique no botão “Criar Bucket”.

Buckets em analytics *Compartimento*

O Object Storage fornece armazenamento de dados ilimitado, de alto desempenho, durável e seguro. É feito o upload dos dados como objetos armazenados em buckets. [Saiba mais](#)

**Criar Bucket**

Nome	Camada de Armazenamento	Visibilidade	Criado
Nenhum item encontrado.			

Filtros de Tag [adicionar](#) | [limpar](#)

nenhum filtro de tag aplicado

Copyright © 2020. Oracle e/ou suas empresas afiliadas. Todos os direitos reservados.

4. Nas opções de criação do bucket, altere o nome do seu bucket para “opendata” e clique no botão “Criar Bucket”.

Armazenamento de Objetos

**Buckets**

NOME DO BUCKET

CAMADA DE ARMAZENAMENTO

A camada de armazenamento de um bucket só pode ser especificada durante a criação. Depois de definida, não é possível alterar a camada de armazenamento onde reside um bucket.

PADRÃO  
 ARQUIVO COMPACTADO

EVENTOS DE OBJETO [?](#)  
 EMITIR EVENTOS DE OBJETO

CONTROLE DE VERSÃO DO OBJETO [?](#)  
 ATIVAR CONTROLE DE VERSÃO DO OBJETO

CRPTOGRAFIA

CRPTOGRAFAR USANDO CHAVES GERENCIADAS PELA ORACLE  
Deixa todos os assuntos relacionados a criptografia com a Oracle.  
 CRPTOGRAFAR USANDO CHAVES GERENCIADAS PELO CLIENTE  
Exige uma chave válida de um vault ao qual você tem acesso. ([Saiba Mais](#))

TAGS

Tag é um sistema de metadados que permite organizar e rastrear recursos na sua tenancy. As tags são compostas de chaves e valores que podem ser anexados a recursos.  
[Saiba mais sobre tags](#)

NAMESPACE DE TAG   
CHAVE DE TAG   
VALOR

+ Tag Adicional

Copyright © 2020. Oracle e/ou suas empresas afiliadas. Todos os direitos reservados.

5. Clique no bucket “opendata” que acabou de cirar.

The screenshot shows the Oracle Cloud Object Storage Buckets page. On the left, there's a sidebar with 'Armazenamento de Objetos' (Storage) selected. The main area is titled 'Buckets em analytics Compartimento'. It displays a table with one row for the 'opendata' bucket. The table columns are 'Nome' (Name), 'Camada de Armazenamento' (Storage Class), 'Visibilidade' (Visibility), and 'Criado' (Created). The 'opendata' bucket is listed with 'Padrão' as the storage class, 'Privado' (Private) as visibility, and created on 'seg., 30 de nov. de 2020 13:09:01 UTC'. A message at the top says 'Você está usando aproximadamente 0 bytes do limite de 20 GiB de Armazenamento de Objetos e Armazenamento de Arquivos Compactados combinados grátis.' (You are using approximately 0 bytes of the 20 GiB limit for Object Storage and Compressed Files combined free storage.)

6. No seu novo bucket, clique no botão “Fazer Upload”.

The screenshot shows the 'Detalhes do Bucket' (Bucket Details) page for the 'opendata' bucket. On the left, there's a sidebar with 'Armazenamento de Objetos' selected. The main area shows the bucket details. At the top, there's a message about using 0 bytes of the 20 GiB limit. Below it, the bucket name 'opendata' is displayed with a large green button containing a white 'B'. Action buttons include 'Editar Visibilidade', 'Mover Recurso', 'Recipientar', 'Adicionar Tags', and 'Excluir'. Under 'Informações do Bucket', there are two tabs: 'Informações do Bucket' (selected) and 'Tags'. Bucket details shown include: 'Visibilidade: Privado', 'Namespace: grtz7pvp01t', 'Camada de Armazenamento: Padrão', 'Contagem aproximada: 0 objetos', 'ETag: eaf8d3c1-2bf2-49cf-a16d-0d8687acb0db', 'OCID: ...26ow5jrq', 'Chave de Criptografia: Chave gerenciada pela Oracle', 'Criado: seg., 30 de nov. de 2020 13:09:01 UTC', 'Compartimento: analytics', 'Tamanho Aproximado: 0 bytes', 'Emitir Eventos de Objeto: Desativado', and 'Controle de Versão do Objeto: Desativado'. At the bottom, there are sections for 'Recursos' (Resources) and 'Objetos' (Objects), both currently empty.

7. Na tela de Upload de Objetos, clique no link “selecionar arquivos”.

The screenshot shows the Oracle Cloud Storage interface. On the left, there's a sidebar titled 'Armacenamento de Objetos » Detalhes do Bucket' for the 'opendata' bucket. It displays various bucket metadata like Visibility (Private), Namespace, and ETag. Below this are sections for 'Recursos' (Resources) and 'Objetos' (Objects). The 'Objetos' section includes a 'Fazer Upload' button. The main right panel is titled 'Fazer Upload de Objetos' and contains fields for 'PREFIXO DO NOME DO OBJETO' (OPTIONAL) and 'ESCOLHER ARQUIVOS EM SEU COMPUTADOR' (choose files from your computer). A 'Fazer Upload' button is at the bottom. The bottom of the page includes links for 'Termos de Uso e Privacidade' and 'Preferências de Cookies'.

8. e navegue até a pasta onde estao armazenados os arquivos padronizados do TSE (Ex.: C:\Users\developer\Documents\TDC\open-data-master\br\tse.jus.br\votacao-candidato-munzona\data\04-standardized) e selecione todos os arquivos com exceção do arquivo do BRASIL, para não inserir dados redundantes.

The screenshot shows a file selection dialog box. The left sidebar shows a folder structure under 'TDC'. The main area lists files in the '04-standardized' folder, including 'votacao\_candidato\_munzona\_2016\_AC.csv', 'votacao\_candidato\_munzona\_2016\_AL.csv', 'votacao\_candidato\_munzona\_2016\_AM.csv', 'votacao\_candidato\_munzona\_2016\_AP.csv', 'votacao\_candidato\_munzona\_2016\_BA.csv', 'votacao\_candidato\_munzona\_2016\_BRASIL.csv', 'votacao\_candidato\_munzona\_2016\_CE.csv', 'votacao\_candidato\_munzona\_2016\_ES.csv', 'votacao\_candidato\_munzona\_2016\_GO.csv', 'votacao\_candidato\_munzona\_2016\_MA.csv', 'votacao\_candidato\_munzona\_2016\_MG.csv', 'votacao\_candidato\_munzona\_2016\_MS.csv', 'votacao\_candidato\_munzona\_2016\_MT.csv', 'votacao\_candidato\_munzona\_2016\_PA.csv', and 'votacao\_candidato\_munzona\_2016\_PB.csv'. At the bottom, there are buttons for 'File name:' (containing 'votacao\_candidato\_munzona\_2016\_AC.csv'), 'Todos os arquivos (\*.\*)', 'Open', and 'Cancel'.

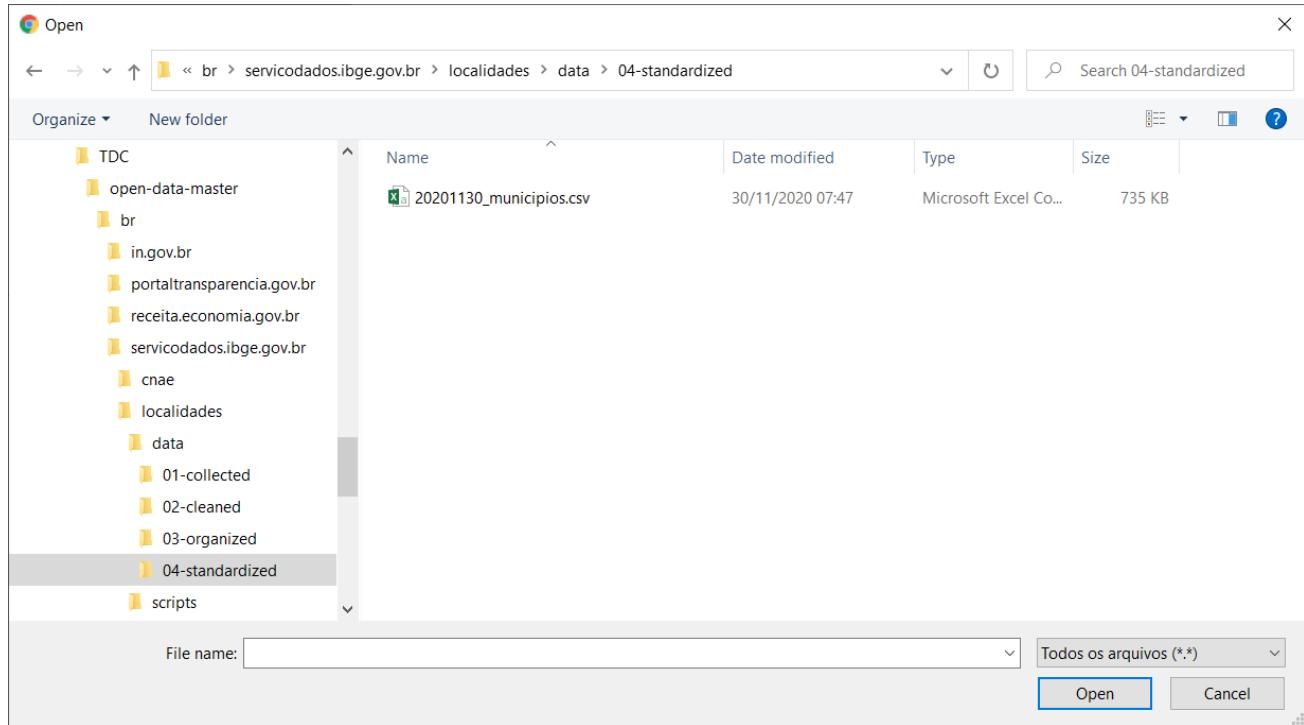
9. Clique no botão “Fazer Upload”.

The screenshot shows the Oracle Cloud Storage interface. On the left, there's a sidebar with 'Recursos' (Resources) and 'Objetos' (Objects) selected. The main area displays a bucket named 'opendata'. A large green button labeled 'B' is visible. The 'Fazer Upload de Objetos' (Upload Objects) dialog box is open, showing a list of files to upload. The 'Fazer Upload' button at the bottom left of the dialog is highlighted in blue.

10. Aguarde a finalização do upload de todos os conjuntos de dados.

The screenshot shows the Oracle Cloud Storage interface after the upload process has completed. The 'Fazer Upload de Objetos' dialog box is still open, but the list of files now shows a 'Status' column where every file is marked as 'Finalizado' (Completed). The 'Abortar' (Abort) button at the bottom left of the dialog is highlighted in red.

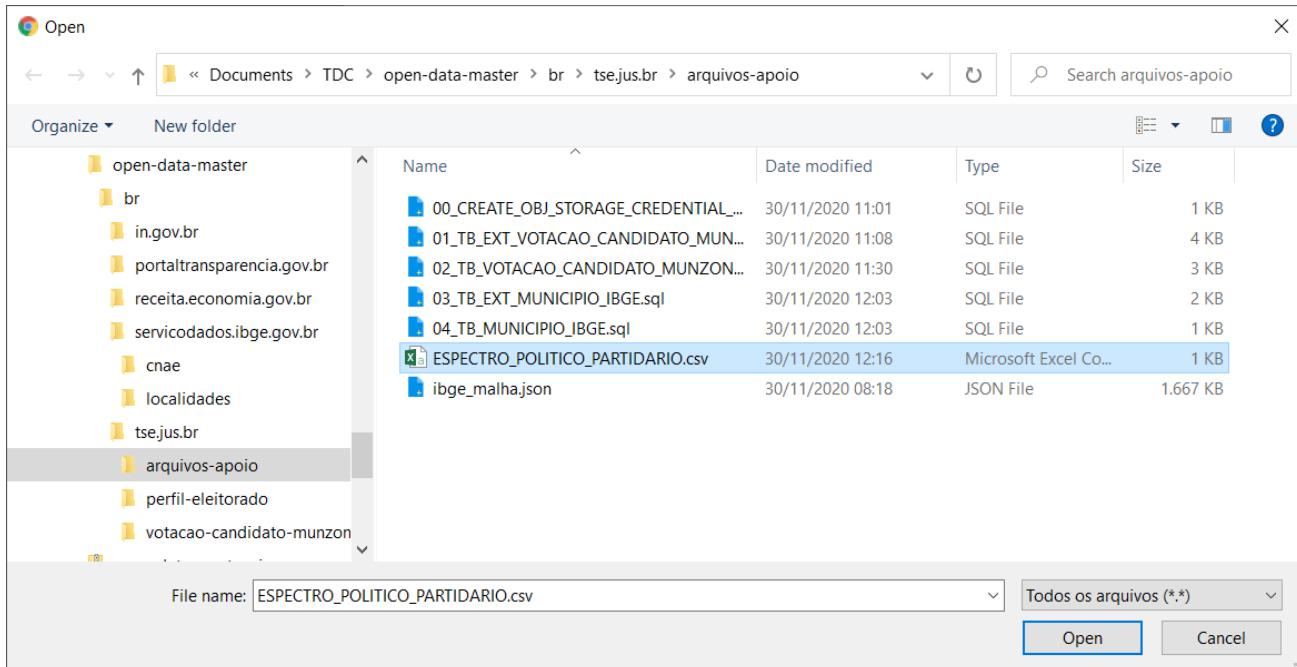
11. Repita os mesmos passos e faça, também, o upload do arquivo padronizado de municípios do IBGE “YYYYMMDD\_municipios.csv”.



12. Aguarde a finalização do upload do conjunto de dados.

The screenshot shows the Oracle Cloud Object Storage interface. The top navigation bar includes "ORACLE Cloud", a search bar, and a location indicator "Brazil East (Sao Paulo)". The main area is titled "Fazer Upload de Objetos" (Make Object Upload). On the left, a sidebar lists "Objetos", "Métricas", "Solicitações Pré-autenticadas", "Solicitações de Serviço", "Regras de Política de Ciclo de Vida", "Política de Replicação", "Regras de Retenção", and "Logs". The central area shows a list of objects with one item uploaded: "20201130\_municipios.csv 734,5 KIB". Below this, there are sections for "PREFÍXO DO NOME DO OBJETO" (OPTIONAL) and "ESCOLHER ARQUIVOS EM SEU COMPUTADOR" (CHOOSE FILES FROM YOUR COMPUTER), which contains a placeholder "Soltar arquivos aqui ou selecionar arquivos". At the bottom right, there is a "Finalizado" (Completed) status indicator. A "Fechar" (Close) button is located at the bottom center of the upload dialog.

13. Novamente, repita os mesmos passos e faça, também, o upload do arquivo de **ESPECTRO\_POLITICO\_PARTIDARIO.csv** disponível na pasta “arquivos-apoio”.



14. Aguarde a finalização do upload do arquivo.

The screenshot shows the Oracle Cloud Object Storage interface with the following details:

- Bucket information:
  - Visibilidade: Privado
  - Namespace: grtzj7pvp01
  - Camada de Armazenamento: F
  - Contagem aproximada: 28 objetos
  - ETag: eaf8d3c1-2bf2-49cf-a16d-0c1d... (with 'Mostrar' and 'Copiar' options)
- Upload process:
  - PREFÍXO DO NOME DO OBJETO: (OPTIONAL)
  - ESCOLHER ARQUIVOS EM SEU COMPUTADOR: (with 'Soltar arquivos aqui ou selecionar arquivos' placeholder)
  - File listed: ESPECTRO\_POLITICO\_PARTIDARIO.csv 978 bytes
  - Status: Finalizado (Completed)
- Resources sidebar:
  - Objetos
  - Métricas
  - Solicitações Pré-autenticadas
  - Solicitações de Serviço
  - Regras de Política de Ciclo de Vida
  - Política de Replicação
  - Regras de Retenção
  - Logs
- Bottom footer: Termos de Uso e Privacidade | Preferências de Cookies | Copyright © 2020, Oracle e/ou suas empresas afiliadas. Todos os direitos reservados.

15. Assim finalizamos o upload dos conjuntos de dados para o Object Storage.

## Armazenando os dados no banco de dados

16. No menu de navegação do OCI, selecione “Autonomous Data Warehouse” entre as opções de “Oracle Database”.

The screenshot shows the OCI console interface. On the left, the navigation sidebar is open, showing various cloud services under "Core Infrastructure". The "Autonomous Data Warehouse" option is selected. The main content area displays several quick-start guides and links:

- AUTONOMOUS TRANSACTION PROCESSING**: Create an ATP database (3-5 mins, Always Free Eligible)
- AUTONOMOUS DATA WAREHOUSE**: Create an ADW database (3-5 mins, Always Free Eligible)
- RESOURCE MANAGER**: Create a stack (2-6 mins, Always Free Eligible)
- OBJECT STORAGE**: Store data (2-6 mins, Always Free Eligible)
- RESOURCE MANAGER**: Preinstall Oracle Cloud Development Kit (10-15 mins, Always Free Eligible)
- SEARCH**: View all my resources

Below these, there are sections for "Key Concepts and Terminology" and "Get Started with FREE training from Oracle University". The right sidebar contains the "Account Center" with user management and upgrade options, and a "What's New" section listing recent updates like Bulk tag editing and ADB-D on Exadata Cloud@Customer.

17. Selecione o seu Autonomous Data Warehouse (ADW). Obs.: Caso ainda não tenha criado um ADW, siga [estes passos](#) para criar um. Também será necessário ter uma credencial de acesso entre o Object Storage e o ADW configurado, veja [aqui](#) como criar esta credencial.

The screenshot shows the OCI console with the "Autonomous Database" section selected in the navigation sidebar. The main content area displays a table titled "Autonomous Databases in analytics Compartment" with one row of data:

Create Autonomous Database							
Display Name	State	Dedicated	OCPUs	Storage (TB)	Workload Type	Autonomous Data Guard	Created
Open Data	Always Free	Available	No	1	0.02	Data Warehouse	Disabled

At the bottom of the table, it says "Displaying 1 Autonomous Database < 1 of 1 >".

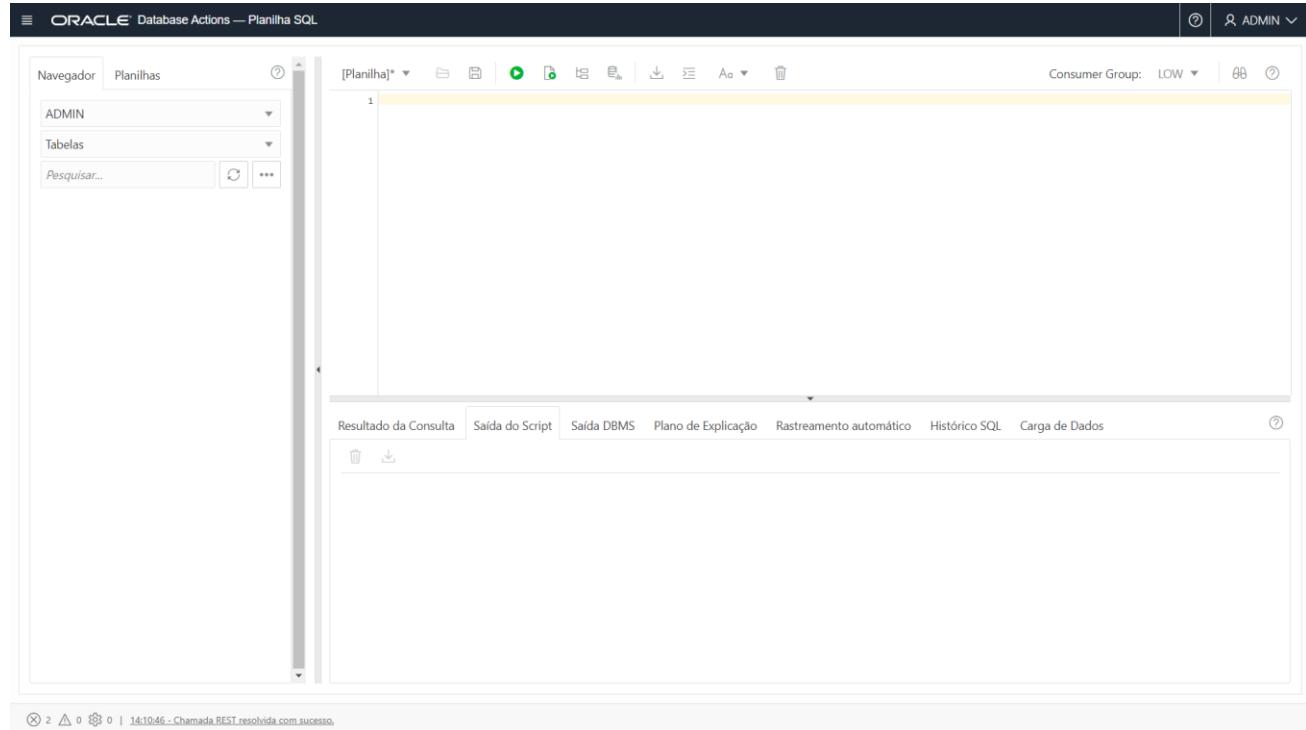
18. No seu ADW, selecione a aba “Tools” (“Ferramentas”) e clique no botão “Open SQL Developer Web” (“Abrir SQL Developer Web”).

The screenshot shows the Oracle Cloud interface for an Autonomous Database. At the top, there's a navigation bar with the Oracle Cloud logo, a search bar, and account information for "Brazil East (Sao Paulo)". Below the navigation, the page title is "Autonomous Database Details". A large green banner on the left says "ADW" and "AVAILABLE". The main content area has tabs: "Autonomous Database Information" (selected), "Tools" (highlighted in blue), and "Tags". Under "Tools", there are four sections: "SQL Developer Web", "Oracle Application Express", "Oracle ML User Administration", and "SODA Drivers". Each section contains a brief description and a "View" or "Open" button. The "SQL Developer Web" section is currently active, showing its description and an "Open SQL Developer Web" button.

19. Acesse o SQL Developer Web com o usuário e senha do ADMIN. Definida durante a criação do seu ADW.

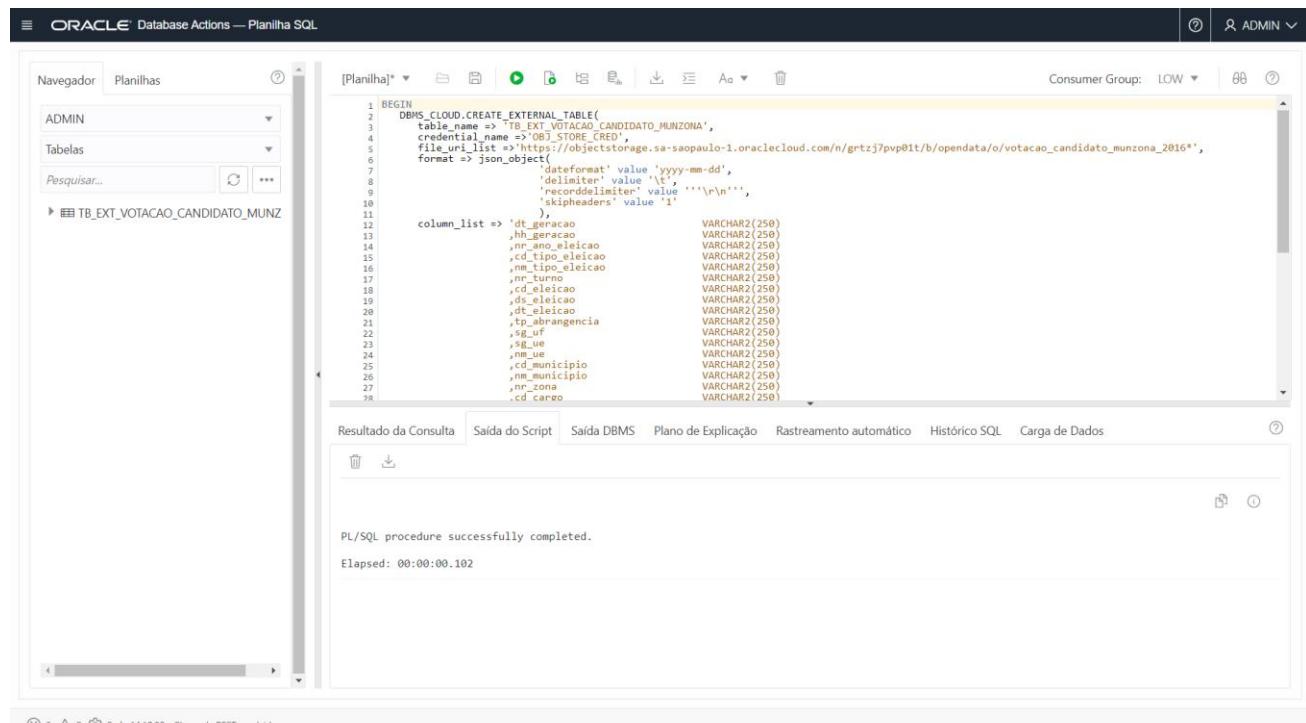
The screenshot shows the Oracle Database Actions sign-in page. It features a logo for "ORACLE Database Actions" and a form with two fields: "Username" containing "ADMIN" and "Password" containing a masked password. Below the form is a blue "Sign in" button.

20. No ambiente do SQL Developer Web, é possível executar algumas instruções SQL para importar os dados armazenados no Object Storage para o ADW. Faremos este processo utilizando o recurso de tabelas externas.



The screenshot shows the Oracle Database Actions interface. The top navigation bar includes tabs for 'Navegador' (Navigator), 'Planilhas' (Sheets), and 'ADMIN'. The main area is titled '[Planilha]\*' and contains a single cell with the number '1'. Below the sheet are tabs for 'Resultado da Consulta' (Query Result), 'Saída do Script' (Script Output), 'Saída DBMS' (DBMS Output), 'Plano de Explicação' (Explanation Plan), 'Rastreamento automático' (Automatic Tracing), 'Histórico SQL' (SQL History), and 'Carga de Dados' (Data Load). A message at the bottom left indicates a successful REST call: '14:10:46 - Chamada REST resolvida com sucesso.'

21. Copie e cole o script “01\_TB\_EXT\_VOTACAO\_CANDIDATO\_MUNZONA.sql” na área de trabalho do SQL Developer Web. Substitua os devidos parâmetros (como o nome da sua credencial de acesso ao Object Storage e o namespace do seu bucket) e execute o script SQL. Este script irá criar uma tabela externa que acessa os arquivos CSV do TSE armazenados no Object Storage.



The screenshot shows the Oracle Database Actions interface with the same layout as the previous one. The main area displays the following PL/SQL code:

```
1 BEGIN
2   DBMS_CLOUD.CREATE_EXTERNAL_TABLE(
3     table_name => 'TB_EXT_VOTACAO_CANDIDATO_MUNZONA',
4     credential_name => '087_STORE_CRED',
5     file_url_list => 'https://objectstorage.sa-saopaulo-1.oraclecloud.com/n/grtzj7pvp01t/b/opendata/o/votacao_candidato_munzona_2016*',
6     format => json_object(
7       'dateformat' value 'yyyy-mm-dd',
8       'encoding' value 'utf-8',
9       'recorddelimited' value '***\r\n***',
10      'skipheaders' value '1'
11    ),
12    column_list => 'dt_geracao          VARCHAR2(250)
13                      ,nr_eleicao           VARCHAR2(250)
14                      ,cd_tipo_eleicao      VARCHAR2(250)
15                      ,nm_tipo_eleicao      VARCHAR2(250)
16                      ,nr_municipio         VARCHAR2(250)
17                      ,cd_municipio         VARCHAR2(250)
18                      ,cd_eleicao           VARCHAR2(250)
19                      ,ds_eleicao           VARCHAR2(250)
20                      ,dt_eleicao           VARCHAR2(250)
21                      ,tp_abrangencia      VARCHAR2(250)
22                      ,s8_ue               VARCHAR2(250)
23                      ,s8_ue               VARCHAR2(250)
24                      ,nm_ue               VARCHAR2(250)
25                      ,cd_municipio         VARCHAR2(250)
26                      ,nm_municipio         VARCHAR2(250)
27                      ,nr_zona             VARCHAR2(250)
28                      ,cd_cargo             VARCHAR2(250))
```

Below the code, the status message 'PL/SQL procedure successfully completed.' is displayed, along with the elapsed time 'Elapsed: 00:00:00.102'.

22. Copie e cole o script “02\_TB\_VOTACAO\_CANDIDATO\_MUNZONA.sql” na área de trabalho do SQL Developer Web. Este arquivo irá copiar os dados da tabela externa que aponta para o Object Storage para uma nova tabela no ADW.

The screenshot shows the Oracle Database Actions interface with the 'Planilha' tab selected. The left sidebar shows 'Navegador' and 'Tabelas' with entries like 'TB\_EXT\_VOTACAO\_CANDIDATO\_MUNZ' and 'TB\_VOTACAO\_CANDIDATO\_MUNZONA'. The main area contains the following SQL code:

```

1 CREATE TABLE TB_VOTACAO_CANDIDATO_MUNZONA AS
2 SELECT
3   TO_DATE(VCH_DT_GERACAO
4         ||VCM_NM_GERACAO
5         , 'YYYY-MM-DD HH24:MI:SS') AS DT_GERACAO
6   ,TO_DATE(VCH_NR_ANO_ELEICAO
7           ||'00' || VCH_NR_DD_ELEICAO
8           , 'YYYY-MM-DD') AS NR_ANO_ELEICAO
9   ,VCM_CD_ELEICAO AS NM_TIPO_ELEICAO
10  ,VCM_NR_TURNO AS NR_TURNO
11  ,VCM_CD_ELEICAO AS CD_ELEICAO
12  ,VCM_CD_ELEICAO AS NM_CARGO
13  ,TO_DATE(VCM_DT_ELEICAO, 'YYYY-MM-DD') AS DT_ELEICAO
14  ,VCM_TP_ABRANGENCIA AS TP_ABRANGENCIA
15  ,VCM_SG_UF AS SG_UF
16  ,VCM_SG_UF AS SG_UF
17  ,VCM_SQ_CUE AS NM_CUE
18  ,VCM_CD_MUNICIPIO AS CD_MUNICIPIO
19  ,VCM_NM_MUNICIPIO AS NM_MUNICIPIO
20  ,VCM_NR_ZONA AS NR_ZONA
21  ,VCM_DS_CARGO AS DS_CARGO
22  ,VCM_DS_CARGO AS DS_CARGO
23  ,VCM_SQ_CANDIDATO AS SQ_CANDIDATO
24  ,VCM_NR_CANDIDATO AS NR_CANDIDATO
25  ,VCM_NR_CANDIDATO AS NM_CANDIDATO
26  ,VCM_NM_CANDIDATO_URNAS AS NM_CANDIDATO_URNAS
27  ,VCM_NM_CANDIDATO_SOCIAL AS NM_CANDIDATO_SOCIAL
28

```

Below the code, the 'Resultado da Consulta' tab shows the output:

```

PL/SQL procedure successfully completed.
Elapsed: 00:00:00.102

```

Table TB\_VOTACAO\_CANDIDATO\_MUNZONA created.

Elapsed: 00:00:30.401

23. Repita o mesmo processo com os scripts “03\_TB\_EXT\_MUNICIPIO\_IBGE.sql” e “04\_TB\_MUNICIPIOS\_IBGE.sql” para termos disponível no ADW os dados de municípios do IBGE.

The screenshot shows the Oracle Database Actions interface with the 'Planilha' tab selected. The left sidebar shows 'Navegador' and 'Tabelas' with entries like 'TB\_EXT\_MUNICIPIO\_IBGE' and 'TB\_MUNICIPIO\_IBGE'. The main area contains the following SQL code:

```

1 CREATE TABLE TB_MUNICIPIO_IBGE AS
2 SELECT
3   CAST(ID_MUNICIPIO AS NUMBER) AS ID_MUNICIPIO
4   ,NM_MUNICIPIO AS NM_MUNICIPIO
5   ,CAST(ID_MICROREGIAO AS NUMBER) AS ID_MICROREGIAO
6   ,NM_MICROREGIAO AS NM_MICROREGIAO
7   ,CAST(ID_MESORREGIAO AS NUMBER) AS ID_MESORREGIAO
8   ,NM_MESORREGIAO AS NM_MESORREGIAO
9   ,CAST(ID_UF AS NUMBER) AS ID_UF
10  ,NM_UF AS NM_UF
11  ,NM_UF AS NM_UF
12  ,CAST(ID_REGIAO AS NUMBER) AS ID_REGIAO
13  ,SG_REGIAO AS SG_REGIAO
14  ,NM_REGIAO AS NM_REGIAO
15  ,TO_DATE(DT_REFERENCIA, 'YYYY-MM-DD') AS DT_REFERENCIA
16  ,NM_FONTE_DADOS AS NM_FONTE_DADOS
17  ,NM_ARQUIVO_DADOS AS NM_ARQUIVO_DADOS
18
19 FROM
20   TB_EXT_MUNICIPIO_IBGE
21 ;

```

Below the code, the 'Resultado da Consulta' tab shows the output:

```

PL/SQL procedure successfully completed.
Elapsed: 00:00:00.021

```

Table TB\_MUNICIPIO\_IBGE created.

24. Repita, também, o processo com o script “05\_TB\_EXT\_ESPECTRO\_POLITICO\_PARTIDARIO.sql” e “06\_TB\_ESPECTRO\_POLITICO\_PARTIDARIO.sql” para termos disponível no ADW a relação de partidos políticos e seu respectivo espectro político.

The screenshot shows the Oracle Database Actions interface with the title "ORACLE Database Actions — Planilha SQL". The left sidebar shows a navigation tree with "ADMIN" selected under "Tabelas", and a list of tables including TB\_ESPECTRO\_POLITICO\_PARTIDARIO, TB\_EXT\_ESPECTRO\_POLITICO\_PAIS, TB\_EXT\_MUNICIPIO\_IBGE, TB\_EXT\_VOTACAO\_CANDIDATO\_I, TB\_MUNICIPIO\_IBGE, and TB\_VOTACAO\_CANDIDATO\_MUN. The main area contains a SQL editor window with the following code:

```
1 CREATE TABLE TB_ESPECTRO_POLITICO_PARTIDARIO AS
2 SELECT
3   SG_PARTIDO AS SG_PARTIDO
4   ,DS_ESPECTRO_POLITICO AS DS_ESPECTRO_POLITICO
5
6 FROM
7   TB_EXT_ESPECTRO_POLITICO_PARTIDARIO
8 ;
```

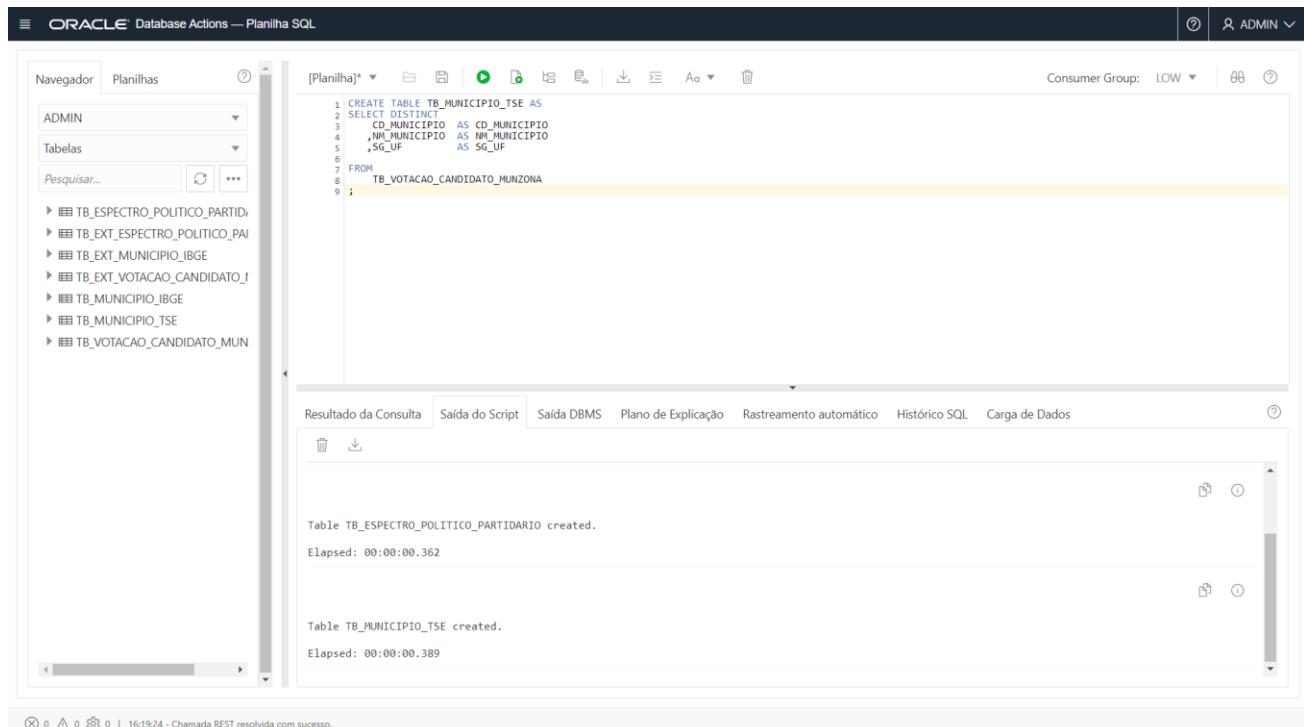
Below the editor, there are tabs for "Resultado da Consulta", "Saída do Script", "Saída DBMS", "Plano de Explicação", "Rastreamento automático", "Histórico SQL", and "Carga de Dados". The "Resultado da Consulta" tab is active, showing the message "PL/SQL procedure successfully completed." and "Elapsed: 00:00:00.036". The "Saída do Script" tab shows the message "Table TB\_ESPECTRO\_POLITICO\_PARTIDARIO created." and "Elapsed: 00:00:00.362". At the bottom of the interface, there is a status bar with icons and the text "16/03/35 - Chamada REST resolvida com sucesso."

## LAB.3: PREPARAR OS DADOS ABERTOS

Com os dados disponíveis no ADW, é hora de iniciar algumas preparações de dados. Como realizar o relacionamento dos municípios do TSE com os do IBGE e criar a tabela que classifica qual o principal espectro político de um município.

### Preparando os municípios

1. Ainda no SQL Developer Web, copie, cole e execute o script “07\_TB\_MUNICIPIO\_TSE.sql” para criar uma tabela com os distintos municípios utilizados pelo TSE.



The screenshot shows the Oracle Database Actions interface with the title "ORACLE Database Actions — Planilha SQL". The left sidebar shows a navigation tree with "ADMIN" selected under "Tabelas", and a search bar. The main area contains a SQL editor with the following code:

```
1 CREATE TABLE TB_MUNICIPIO_TSE AS
2 SELECT DISTINCT
3   CD_MUNICIPIO AS CD_MUNICIPIO
4   ,NM_MUNICIPIO AS NM_MUNICIPIO
5   ,SG_UF
6
7 FROM
8   TB_VOTACAO_CANDIDATO_MUNZONA
9 ;
```

Below the editor, there are tabs for "Resultado da Consulta", "Saída do Script", "Saída DBMS", "Plano de Explicação", "Rastreamento automático", "Histórico SQL", and "Carga de Dados". The "Resultado da Consulta" tab shows the output of the executed query:

Table TB\_ESPECTRO\_POLITICO\_PARTIDARIO created.  
Elapsed: 00:00:00.362

Table TB\_MUNICIPIO\_TSE created.  
Elapsed: 00:00:00.389

At the bottom, a status bar indicates: "X 0 △ 0 ⚙ 0 | 16:19:24 - Chamada REST resolvida com sucesso."

2. Copie, cole e execute o script “08\_TB\_MUNICIPIO\_TSE\_UPDATE.sql” para alterar os nomes de alguns municípios da base do TSE para que fiquem correspondentes com a base do IBGE.

```

1 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'SINIMIPE' WHERE CD_MUNICIPIO = '87847';
2 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'TAQUARIU DO SUL' WHERE CD_MUNICIPIO = '85669';
3 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'BROCHIER' WHERE CD_MUNICIPIO = '85065';
4 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'CANGUÇU' WHERE CD_MUNICIPIO = '85871';
5 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'HERVAL DO OESTE' WHERE CD_MUNICIPIO = '81310';
6 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'PARAI' WHERE CD_MUNICIPIO = '81191';
7 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'PRESIDENTE CASTELLO BRANCO' WHERE CD_MUNICIPIO = '82732';
8 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'LUIZ ALVES' WHERE CD_MUNICIPIO = '81973';
9 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'BALNEÁRIO PICARAS' WHERE CD_MUNICIPIO = '82511';
10 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'URUGUAII' WHERE CD_MUNICIPIO = '82520';
11 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'AMPARO DE SÃO FRANCISCO' WHERE CD_MUNICIPIO = '81232';
12 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'GRACIO CARDOSO' WHERE CD_MUNICIPIO = '81518';
13 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'FLORIANA' WHERE CD_MUNICIPIO = '64238';
14 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'MUNICIPIO DE SÃO GONÇALO DO PARAITINGA' WHERE CD_MUNICIPIO = '67115';
15 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'SUD MENNUCCI' WHERE CD_MUNICIPIO = '71471';
16 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'MOG DAS CRUZES' WHERE CD_MUNICIPIO = '67130';
17 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'TARABAÍ' WHERE CD_MUNICIPIO = '71790';
18 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'VILA VELHA' WHERE CD_MUNICIPIO = '67131';
19 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'ITACOA' WHERE CD_MUNICIPIO = '61683';
20 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'MOG MIRIM' WHERE CD_MUNICIPIO = '67172';
21 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'SÃO VALERIO' WHERE CD_MUNICIPIO = '73466';
22 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'OLIVEIRA DE FATIMA' WHERE CD_MUNICIPIO = '73466';
23 UPDATE TB_MUNICIPIO_TSE SET NM_MUNICIPIO = 'DARCINÓPOLIS' WHERE CD_MUNICIPIO = '73032';
24 COMMIT;

```

3. Copie, cole e execute o script “09\_TB\_LOCALIDADE.sql” para criar uma tabela que unifica os padrões de municípios do IBGE com os códigos utilizados pelo TSE.

```

1 CREATE TABLE TB_LOCALIDADE AS
2 SELECT
3   LOC_ID_MUNICIPIO
4   ,CAST(LOC_ID_MUNICIPIO AS VARCHAR2(10)) AS ID_MUNICIPIO
5   ,NM_MUNICIPIO AS NM_MUNICIPIO
6   ,LOC_ID_MICROREGIAO
7   ,CAST(LOC_ID_MICROREGIAO AS VARCHAR2(10)) AS NM_MICROREGIAO
8   ,LOC_NM_MICROREGIAO AS NM_MICROREGIAO
9   ,LOC_ID_MESOREGIAO
10  ,CAST(LOC_ID_MESOREGIAO AS VARCHAR2(10)) AS ID_MESOREGIAO
11  ,LOC_NM_MESOREGIAO AS NM_MESOREGIAO
12  ,LOC_ID_UF
13  ,CAST(LOC_ID_UF AS VARCHAR2(10)) AS ID_UF
14  ,LOC_SG_UF AS SG_UF
15  ,LOC_NM_UF AS NM_UF
16  ,LOC_ID_REGIAO
17  ,CAST(LOC_ID_REGIAO AS VARCHAR2(10)) AS ID_REGIAO
18  ,LOC_NM_REGIAO AS NM_REGIAO
19  ,LOC_ID_TSE
20  ,CAST(TMT_CD_MUNICIPIO AS NUMBER(38,0)) AS ID_MUNICIPIO_TSE
21  ,CAST(TMT_CD_MUNICIPIO AS VARCHAR2(10)) AS CD_MUNICIPIO_TSE
22 FROM
23   TB_MUNICIPIO_IBGE LOC
24
25 INNER JOIN TB_MUNICIPIO_TSE TMT
26   ON UPPER(TMT.NM_MUNICIPIO) = UPPER(LOC.NM_MUNICIPIO)
27   AND UPPER(TMT.SG_UF) = UPPER(LOC.SG_UF)
28 ;
29
30

```

## Criando a tabela de principal espectro político

4. Copie, cole e execute o script “10\_TB\_ESPECTRO\_POLITICO\_BR.sql” para criar uma tabela que classifica o principal espectro político de cada município durante o primeiro turno das eleições.

The screenshot shows the Oracle Database Actions interface with a SQL worksheet titled "[Planilha]". The code executed is:

```
1 KCREATE TABLE TB_ESPECTRO_POLITICO_BR AS
2 SELECT
3   AUX.NR_ANO_ELEICAO      AS NR_ANO_ELEICAO
4   ,AUX.CD_MUNICIPIO        AS CD_MUNICIPIO
5   ,AUX.DS_ESPECTRO_POLITICO AS DS_ESPECTRO_POLITICO
6   ,AUX.QT_VOTOS_NOMINAIS   AS QT_VOTOS_NOMINAIS
7   ,RANK() OVER(
8     PARTITION BY
9       AUX.NR_ANO_ELEICAO
10      ,AUX.CD_MUNICIPIO
11    ORDER BY
12      AUX.QT_VOTOS_NOMINAIS DESC) AS VL_RANK_ESPECTRO
13
14  FROM
15  (SELECT
16    VCM.NR_ANO_ELEICAO      AS NR_ANO_ELEICAO
17    ,LOC.CD_MUNICIPIO        AS CD_MUNICIPIO
18    ,EPP.DS_ESPECTRO_POLITICO AS DS_ESPECTRO_POLITICO
19    ,SUM(QT_VOTOS_NOMINAIS)   AS QT_VOTOS_NOMINAIS
20  FROM
21    TB_VOTACAO_CANDIDATO_MUNZONA VCM
22
23  INNER JOIN TB_LOCALIDADE LOC
24    ON LOC.CD_MUNICIPIO_TSE = VCM.CD_MUNICIPIO
25
26  INNER JOIN TB_ESPECTRO_POLITICO_PARTIDARIO EPP
27    ON EPP.SG_PARTIDO = VCM.SG_PARTIDO
28
29 WHERE
```

The results section shows the creation of the tables:

Table TB\_LOCALIDADE created.  
Elapsed: 00:00:01.030

Table TB\_ESPECTRO\_POLITICO\_BR created.  
Elapsed: 00:00:00.785

The URL at the bottom of the browser window is: [https://uiws7tupdyp2ze3-opendata.adb.sa-saopaulo-1.oraclecloudapps.com/ords/admin/\\_sdw/?nav=worksheet#](https://uiws7tupdyp2ze3-opendata.adb.sa-saopaulo-1.oraclecloudapps.com/ords/admin/_sdw/?nav=worksheet#)

5. Assim finalizamos a preparação dos dados no ADW.

## LAB.4: ANALISAR OS DADOS ABERTOS

### Conectando o Analytics com o ADW

- No seu ambiente do seu ADW, clique no botão “Conexão do DB”.

The screenshot shows the Oracle Cloud interface for an Autonomous Database. The main title is "Visão Geral » Autonomous Database » Detalhes do Autonomous Database". The "Open Data" section is highlighted, showing the "Always Free" status. Below it, there's a large green "ADW" logo with "DISPONÍVEL" below it. A navigation bar at the top includes "Procurar recursos, serviços e documentação", "Brazíl East (Sao Paulo)", and various icons. The main content area has tabs for "Informações do Autonomous Database", "Ferramentas", and "Tags". The "Informações Gerais" section contains detailed information about the database, such as Name, Type, OCID, Creation Date, and License Type. Other sections include "Infraestrutura", "Autonomous Data Guard", "Fazer Backup", "Rede", and "Manutenção".

- Nas opções de conexão, clique no botão “Download da Wallet”.

This screenshot shows the "Conexão de Banco de Dados" (Connection to Database) modal dialog. It prompts the user to provide client credentials and connection information. The "Fazer Download das Credenciais do Cliente (Wallet)" section allows selecting the wallet type (Instance Wallet) and performing a download or rotation. The background shows the same Autonomous Database details as the previous screenshot, including the "Open Data" section and various configuration tabs.

3. Informe a senha do ADMIN e clique em “Fazer Download” para baixar o arquivo ZIP.

The screenshot shows the Oracle Cloud interface for an Autonomous Database. A modal window titled "Download da Wallet" is open, prompting for a password. The background shows details of the database instance, including its status as "DISPONIVEL" and various configuration parameters like storage and licensing. At the bottom left, a download link for "Wallet\_opendata.zip" is visible.

4. Em posse das credenciais. Abra o seu Oracle Analytics Desktop

The screenshot shows the Oracle Analytics Desktop search interface. The search bar at the bottom contains the query "Analytics". The results are categorized into "Best match" and "Search the web". In the "Best match" section, the "Oracle Analytics Desktop 12c" app is highlighted. In the "Search the web" section, several search results for "Analytics" are listed, including links to Google, Twitter, YouTube, Vidhya, Uber Eats, and Brasil. To the right, a detailed view of the "Oracle Analytics Desktop 12c" app is shown, including its icon, name, type (App), an "Open" button, and a list of recent files: IBGE\_Localidades.xlsx, 200730\_base\_receitas\_municipais.xlsx, Transparencia\_ComprasNET\_Contratos\_Oracle\_v02.xlsx, and ComprasNET\_UASGs.xlsx.

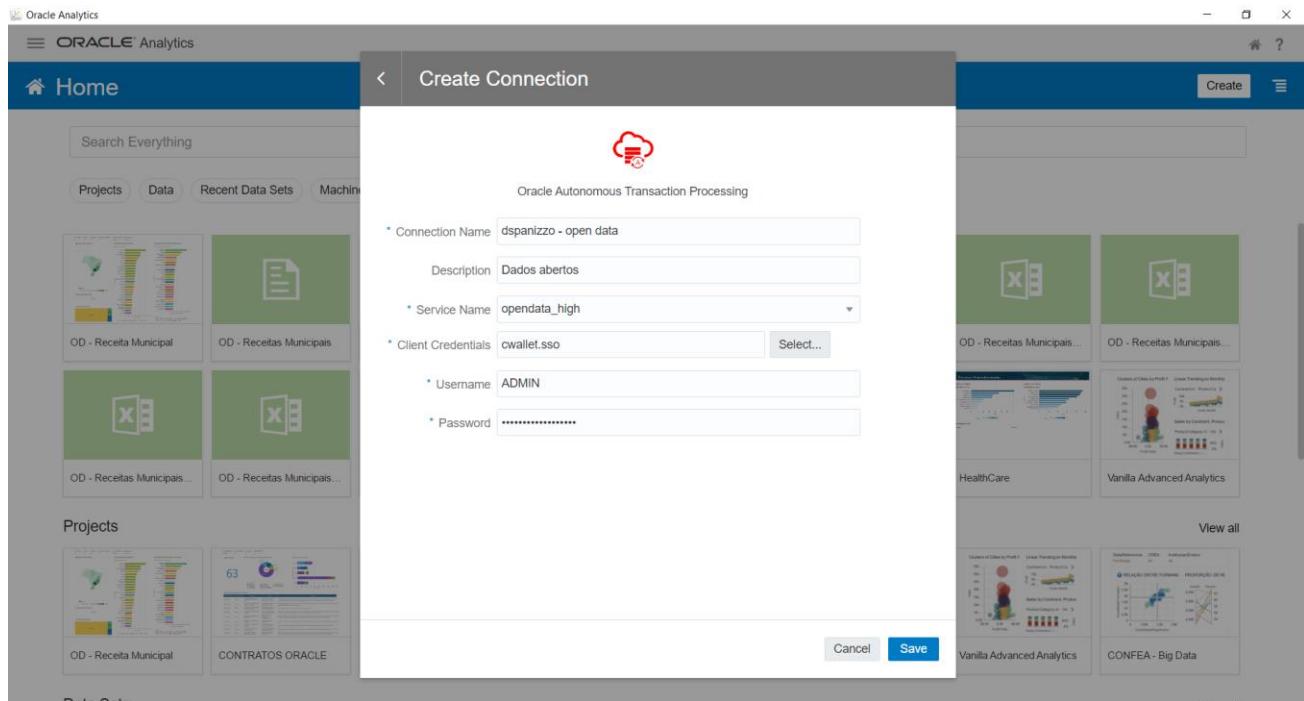
5. Na Home do Oracle Analytics, clique no botão “Create” e, logo em seguida, na opção “Connection”.

The screenshot shows the Oracle Analytics Home interface. At the top right, there is a "Create" button with a dropdown menu. The menu is open, showing options: "Project", "Data Set", "Data Flow", "Sequence", and "Connection". Below the menu, there is a section labeled "</> Script". The main area displays a grid of project cards. Some visible project names include "OD - Receita Municipal", "IBGE - Localidades", "OD - Receitas Municipais...", "CONTRATOS ORACLE", "COMPRASNET - UASG", "HealthInsurance\_PT-BR...", "HealthCare", and "Vanilla Advanced Analytics". Below the grid, there are sections for "Projects" and "Data Sets", each containing several cards.

6. Na tela de criar conexão, selecione a opção “Autonomous Data Warehouse”.

The screenshot shows the "Create Connection" dialog box overlying the Oracle Analytics Home page. The dialog has a title "Select Connection Type" and a search bar. It displays a grid of connection icons with their corresponding names: Oracle Applications, Oracle Autonomous Data Warehouse, Oracle Autonomous Transaction Processing, Oracle Database, Oracle Essbase, Oracle NetSuite, Oracle Service Cloud, Oracle Talent Acquisition, Action Ingres, Action Matrix, Action Vector, Amazon Aurora, Amazon EMR, Amazon Redshift, and Apache Drill. The Oracle Autonomous Data Warehouse icon is highlighted. The background of the dialog shows a blurred view of the Oracle Analytics interface, including the "Projects" and "Data Sets" sections.

7. Na opção “Client Credentials”, informe o local onde armazenou o arquivo ZIP da Wallet do ADW. Em seguida, informe um nome para a conexão, usuário e senha do ADMIN. Clique em “Save”.



8. Com a conexão criada, selecione no menu a opção “Dados”.

9. Na área de “Dados”, selecione a aba “Conexões”.

Dados					
Conjuntos de Dados				Conexões	Fluxos de Dados
Nome		Tipo de Conexão	Descrição	Conexão de Replicação	Proprietário
	dspanizzo - opendata	Oracle Autonomous Transa...	Dados abertos		daniel.panizzo@oracle.com 2 minutos atrás
	DATALAKE	Oracle Autonomous Data ...	Data Lake de dados abertos		daniel.panizzo@oracle.com 25 de Ago de 2020

## Importando os metadados

10. A conexão que acabamos de criar deve aparecer como uma das opções. Clique na nossa conexão “opendata” e a tela importação de metadados deve aparecer. Selecione o esquema “ADMIN”.

The screenshot shows the 'Importar Metadados' (Import Metadata) dialog for the 'dspanizzo - opendata' connection. On the left, there's a list of schemas: ADMIN, APEX\_190200, APEX\_200100, APEX\_200200, APPQOSSYS, AUDSYS, C##ADPSERVICE, C##CLOUDSERVICE, C##OMLIDM, CTXSYS, DBSPWUSER, DBSNMP, DV\$SYS, FLOWS\_FILES, GSMADMIN\_INTERNAL, LBACSYS, MDSYS, OML\$MODELS. On the right, a detailed view of the 'ADMIN' schema is shown:

- Nome: Sem título
- Descrição: Conjunto de Dados Externos
- Conexão: dspanizzo - opendata
- Acesso a D...: Armazenamento Automático
- Novo Indica...: Selecionar Coluna
- Proprietário: daniel.panizzo@oracle.com
- Criado em: Em Andamento
- Modificado ...: Em Andamento
- Atualizado: Nunca

11. As tabelas que criamos no banco de dados devem aparecer como opções. Selecione a tabela “TB\_ESPECTRO\_POLITICO\_BR”.

12. As opções de colunas para serem importadas serão exibidas. Clique no botão “Adicionar Tudo”.

13. Do lado direito, nas propriedades da tabela, altere a opção de “Acesso a Dados” para “Ativo”.

The screenshot shows the Oracle Analytics interface for managing datasets. On the left, there's a sidebar with options like 'Script de Preparação', 'Adicionar Conjunto de D...', 'Resultados', and 'Aplicar Script'. The main area displays a table structure with columns: NR\_ANO\_ELEICAO, CD\_MUNICIPIO, DS\_ESPECTRO\_POLITICO, QT\_VOTOS\_NOMINAIS, and VL\_RANK\_ESPECTRO. To the right, the 'Propriedades' (Properties) panel is open for the dataset 'TB\_ESPECTRO\_POLITICO\_BR'. It includes fields for 'Nome' (Name), 'Descrição' (Description), 'Conexão' (Connection), 'Acesso a Da...' (Access Type), 'Novo Indicad...' (New Indicator), 'Proprietário' (Owner), 'Criado em' (Created on), 'Modificado em' (Modified on), and 'Atualizado' (Last updated). A note in the 'Acesso a Da...' section states: 'Ativo Os dados mais recentes são sempre retornados pela origem ativa. Alguns tipos de preparação e enriquecimento de dados não estão disponíveis.' (Active The most recent data is always returned by the active source. Some types of preparation and enrichment of data are not available.) Below the properties panel, there's a button labeled 'Obter Dados de Visualização' (Get Data for Visualization).

14. Na parte superior, clique no botão “Adicionar”.

This screenshot shows the same Oracle Analytics interface as the previous one, but now the table 'TB\_ESPECTRO\_POLITICO\_BR' has data populated in its preview area. The data consists of five rows with the following values:

NR_ANO_ELEICAO	CD_MUNICIPIO	DS_ESPECTRO_POLITICO	QT_VOTOS_NOMINAIS	VL_RANK_ESPECTRO
01/01/2016 12:00:00.000 AM	1100015	Direita	12.069	1
01/01/2016 12:00:00.000 AM	1100015	Centro	5.554	2
01/01/2016 12:00:00.000 AM	1100015	Centro-direita	3.768	3
01/01/2016 12:00:00.000 AM	1100015	Centro-esquerda	3.501	4

The rest of the interface remains the same, with the properties panel on the right and the 'Script de Preparação' sidebar on the left.

## Ajustando os metadados

15. A ferramenta irá para a etapa de preparação dos dados.

NR_ANO_ELEICAO	CD_MUNICIPIO	DS_ESPECTRO	QT_VOTOS	VL_RANK_ESPECTRO
01/01/2016 12:00:00.000 AM	1703073	Centro-direita	379	4
01/01/2016 12:00:00.000 AM	220850	Centro-direita	1.925	3
01/01/2016 12:00:00.000 AM	2103307	Direita	11.781	4
01/01/2016 12:00:00.000 AM	1505064	Centro	31.468	1
01/01/2016 12:00:00.000 AM	1707702	Direita	459	4
01/01/2016 12:00:00.000 AM	2204105	Centro-esquerda	5.737	1
01/01/2016 12:00:00.000 AM	2210509	Esquerda	16	5
01/01/2016 12:00:00.000 AM	2207959	Centro-direita	122	4
01/01/2016 12:00:00.000 AM	1710706	Centro-esquerda	3.519	1
01/01/2016 12:00:00.000 AM	1600105	Centro	1.634	3
01/01/2016 12:00:00.000 AM	2106904	Esquerda	7.936	2
01/01/2016 12:00:00.000 AM	1714203	Centro-direita	4.717	1
01/01/2016 12:00:00.000 AM	1100031	Centro-direita	339	3
01/01/2016 12:00:00.000 AM	2310951	Esquerda	267	4
01/01/2016 12:00:00.000 AM	2301257	Esquerda	47	4
01/01/2016 12:00:00.000 AM	2206554	Centro-direita	471	4
01/01/2016 12:00:00.000 AM	1701903	Centro	3.646	1
01/01/2016 12:00:00.000 AM	2205704	Centro	739	4
01/01/2016 12:00:00.000 AM	2105706	Centro	2.838	3
01/01/2016 12:00:00.000 AM	1100056	Direita	7.286	1
01/01/2016 12:00:00.000 AM	1715150	Centro	1.701	1
01/01/2016 12:00:00.000 AM	2309805	Direita	905	4
01/01/2016 12:00:00.000 AM	2203869	Centro-esquerda	1.813	1
01/01/2016 12:00:00.000 AM	1718659	Centro	507	3
01/01/2016 12:00:00.000 AM	1100064	Centro	5.950	2
01/01/2016 12:00:00.000 AM	1504752	Centro-direita	9.868	2
01/01/2016 12:00:00.000 AM	2202653	Centro	3.455	2
01/01/2016 12:00:00.000 AM	1100072	Centro-esquerda	6.595	1
01/01/2016 12:00:00.000 AM	1100072	Centro-direita	1.419	2
01/01/2016 12:00:00.000 AM	1703057	Centro-esquerda	544	4
01/01/2016 12:00:00.000 AM	1721257	Centro-esquerda	1.230	1
01/01/2016 12:00:00.000 AM	2109205	Centro	339	5
01/01/2016 12:00:00.000 AM	2102200	Esquerda	9.213	2
01/01/2016 12:00:00.000 AM	1507961	Direita	927	4
01/01/2016 12:00:00.000 AM	2401659	Centro-esquerda	4.869	1
01/01/2016 12:00:00.000 AM	2404507	Centro-esquerda	7.213	2
01/01/2016 12:00:00.000 AM	2302701	Esquerda	993	3
01/01/2016 12:00:00.000 AM	2209609	Direita	3.145	1
01/01/2016 12:00:00.000 AM	2404046	Centro-esquerda	113	4
01/01/2016 12:00:00.000 AM	1100098	Esquerda	501	5
01/01/2016 12:00:00.000 AM	1711506	Centro	144	3
01/01/2016 12:00:00.000 AM	2206209	Direita	7.039	2
01/01/2016 12:00:00.000 AM	2109007	Centro-direita	5.739	2

16. Clique na coluna “VL\_RANK\_ESPECTRO”. As propriedades da coluna irão aparecer na área escura à esquerda na parte inferior. Altere a propriedade “Tratar como” de “Medida” para “Atributo”.

NR_ANO_ELEICAO	CD_MUNICIPIO	DS_ESPECTRO	QT_VOTOS	VL_RANK_ESPECTRO
01/01/2016 12:00:00.000 AM	1703073	Centro-direita	379	4
01/01/2016 12:00:00.000 AM	220850	Centro-direita	1.925	3
01/01/2016 12:00:00.000 AM	2103307	Direita	11.781	4
01/01/2016 12:00:00.000 AM	1505064	Centro	31.468	1
01/01/2016 12:00:00.000 AM	1707702	Direita	459	4
01/01/2016 12:00:00.000 AM	2204105	Centro-esquerda	5.737	1
01/01/2016 12:00:00.000 AM	2210509	Esquerda	16	5
01/01/2016 12:00:00.000 AM	2207959	Centro-direita	122	4
01/01/2016 12:00:00.000 AM	1710706	Centro-esquerda	3.519	1
01/01/2016 12:00:00.000 AM	1600105	Centro	1.634	3
01/01/2016 12:00:00.000 AM	2106904	Esquerda	7.936	2
01/01/2016 12:00:00.000 AM	1714203	Centro-direita	4.717	1
01/01/2016 12:00:00.000 AM	1100031	Centro-direita	339	3
01/01/2016 12:00:00.000 AM	2310951	Esquerda	267	4
01/01/2016 12:00:00.000 AM	2301257	Esquerda	47	4
01/01/2016 12:00:00.000 AM	2206554	Centro-direita	471	4
01/01/2016 12:00:00.000 AM	1701903	Centro	3.646	1
01/01/2016 12:00:00.000 AM	2205704	Centro	739	4
01/01/2016 12:00:00.000 AM	2105706	Centro	2.838	3
01/01/2016 12:00:00.000 AM	1100056	Direita	7.286	1
01/01/2016 12:00:00.000 AM	1715150	Centro	1.701	1
01/01/2016 12:00:00.000 AM	2309805	Direita	905	4
01/01/2016 12:00:00.000 AM	2203869	Centro-esquerda	1.813	1
01/01/2016 12:00:00.000 AM	1718659	Centro	507	3
01/01/2016 12:00:00.000 AM	1100064	Centro	5.950	2
01/01/2016 12:00:00.000 AM	1504752	Centro-direita	9.868	2
01/01/2016 12:00:00.000 AM	2202653	Centro	3.455	2
01/01/2016 12:00:00.000 AM	1100072	Centro-esquerda	6.595	1
01/01/2016 12:00:00.000 AM	1100072	Centro-direita	1.419	2
01/01/2016 12:00:00.000 AM	1703057	Centro-esquerda	544	4
01/01/2016 12:00:00.000 AM	1721257	Centro-esquerda	1.230	1
01/01/2016 12:00:00.000 AM	2109205	Centro	339	5
01/01/2016 12:00:00.000 AM	2102200	Esquerda	9.213	2
01/01/2016 12:00:00.000 AM	1507961	Direita	927	4
01/01/2016 12:00:00.000 AM	2401659	Centro-esquerda	4.869	1
01/01/2016 12:00:00.000 AM	2404507	Centro-esquerda	7.213	2
01/01/2016 12:00:00.000 AM	2302701	Esquerda	993	3
01/01/2016 12:00:00.000 AM	2209609	Direita	3.145	1
01/01/2016 12:00:00.000 AM	2404046	Centro-esquerda	113	4
01/01/2016 12:00:00.000 AM	1100098	Esquerda	501	5
01/01/2016 12:00:00.000 AM	1711506	Centro	144	3
01/01/2016 12:00:00.000 AM	2206209	Direita	7.039	2
01/01/2016 12:00:00.000 AM	2109007	Centro-direita	5.739	2

17. Na parte superior da área escura, clique no botão “Aplicar Script”.

18. No navegador da ferramenta, clique no botão “Criar Projeto”. A etapa de visualização de dados irá iniciar.

## Visualizando os dados

19. Clique com o botão direito na coluna “VL\_RANK\_ESPECTRO” e selecione a opção “Criar Filtro” .

The screenshot shows the Oracle Analytics interface with a context menu open over the 'VL\_RANK\_ESPECTRO' column in the 'Dados' (Data) pane. The menu options are: Criar Melhor Visualização, Selecionar Visualização..., Criar Filtro (which is highlighted), and Explicar VL\_RANK\_ESPECTRO. The main workspace shows a visualization area with a placeholder message 'Clique aqui ou arraste os dados para adicionar um filtro' (Click here or drag data to add a filter). The left sidebar displays the data source 'TB\_ESPECTRO\_POLITICO' and a table named 'NR\_ANO\_ELEICAO' with one row selected.

20. Selecione o valor “1” do filtro “VL\_RANK\_ESPECTRO” para que nossa análise conte comente apenas o principal espectro político do município.

The screenshot shows the Oracle Analytics interface with the 'VL\_RANK\_ESPECTRO' filter set to the value '1'. A modal dialog titled 'VL\_RANK\_ESPECTRO' lists the available values: 1, 2, 3, 4, and 5. The value '1' is selected. The main workspace shows a visualization area with a placeholder message 'Selecionar Visualização para Ver os Detalhes' (Select a visualization to view details). The left sidebar displays the data source 'TB\_ESPECTRO\_POLITICO' and a table named 'NR\_ANO\_ELEICAO' with one row selected.

21. Expanda as opções de data da coluna “NR\_ANO\_ELEICAO”, clique com o botão direito na opção “Ano” e selecione a opção “Criar Filtro”.

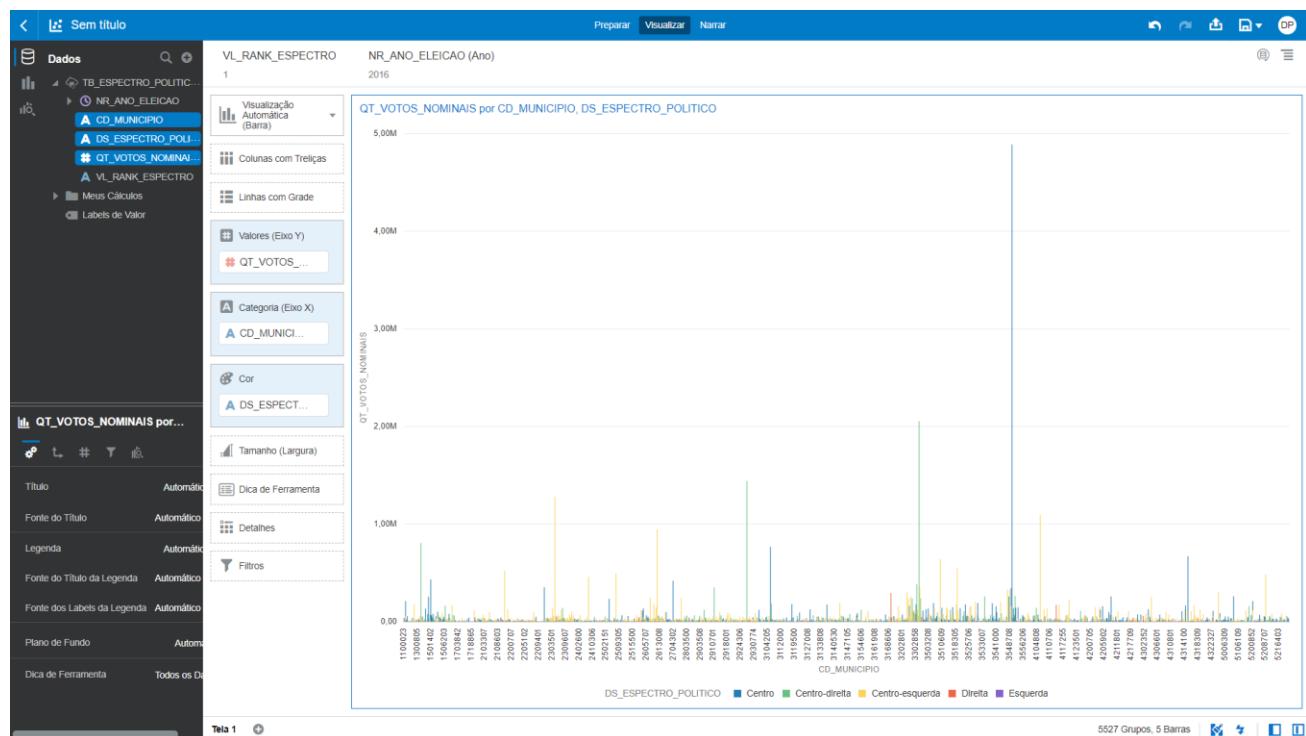
The screenshot shows the Oracle Analytics interface with a blue header bar. In the top right, there are buttons for 'Preparar' (Prepare), 'Visualizar' (View), and 'Narrar' (Narrate). Below the header, on the left, is a sidebar titled 'Dados' (Data) containing a tree view of dimensions and measures. One dimension node, 'NR\_ANO\_ELEICAO', has its 'Ano' (Year) option selected. A context menu is open over this selection, listing options like 'Criar Melhor Visualização' (Create Best Visualization), 'Selecionar Visualização...' (Select Visualization...), and 'Criar Filtro' (Create Filter). The main workspace area contains a visualization titled 'VL\_RANK\_ESPECTRO' with a value of '1'. A tooltip 'Selecionar Visualização para Ver os Detalhes' (Select Visualization to View Details) is visible near the visualization. At the bottom right of the workspace, there are icons for 'Soltar Visualizações ou Dados Aqui' (Drop Visualizations or Data Here) and other interface controls.

22. Selecione o valor “2016” do filtro “Ano” para que nossa análise conte com apenas uma eleição por vez.

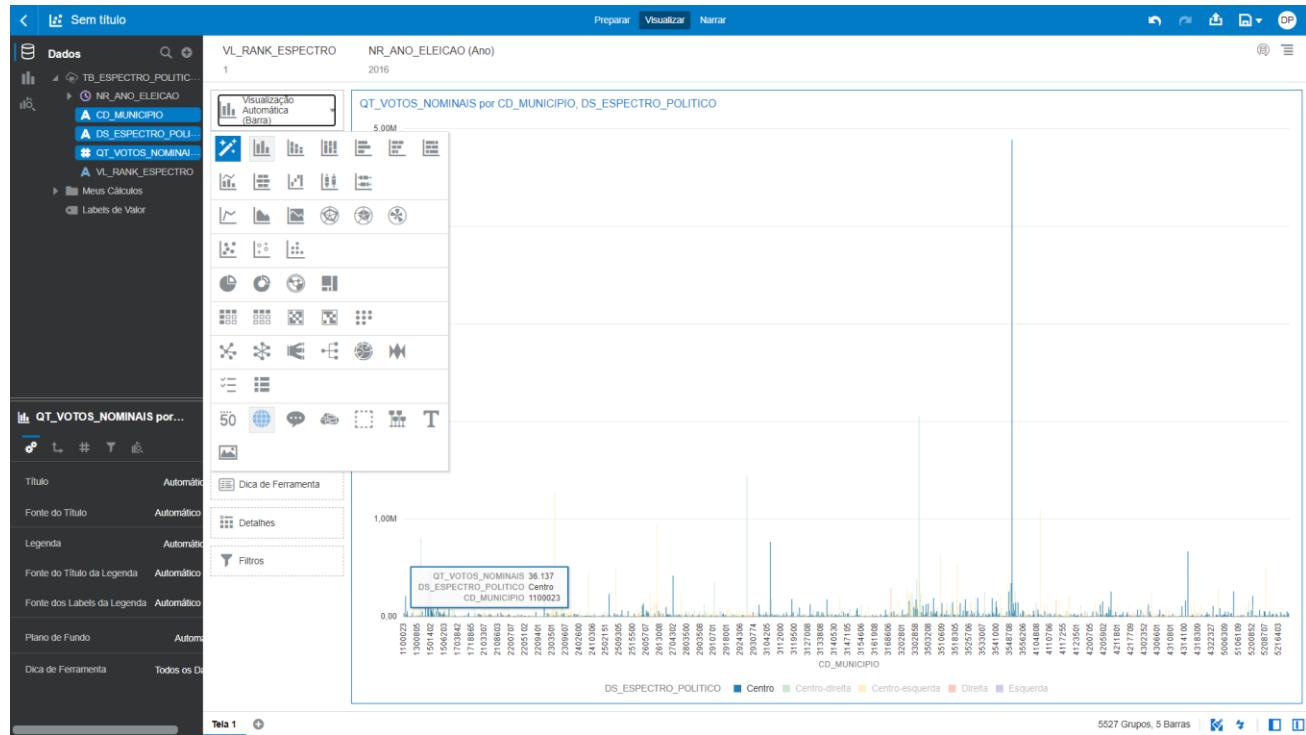
This screenshot shows the same Oracle Analytics interface as the previous one, but with a 'Selecionar' (Select) dialog box overlaid on the workspace. The dialog is centered over the 'Ano' (Year) filter node in the dimension tree. It contains a search bar 'Pesquisar' and a list titled 'Seleções (1/1)' with a single item '2016'. At the bottom of the dialog are buttons for 'Selecionar Tudo' (Select All) and 'Desmarcar Tudo' (Deselect All). The rest of the interface, including the sidebar and workspace, remains consistent with the first screenshot.

23. Segurando o botão no seu teclado CTRL, selecione as colunas “CD\_MUNICIPIO”, “DS\_ESPECTRO\_POLITICO” e “QT\_VOTOS\_NOMINAIS”, arraste e solte na área de visualização de dados

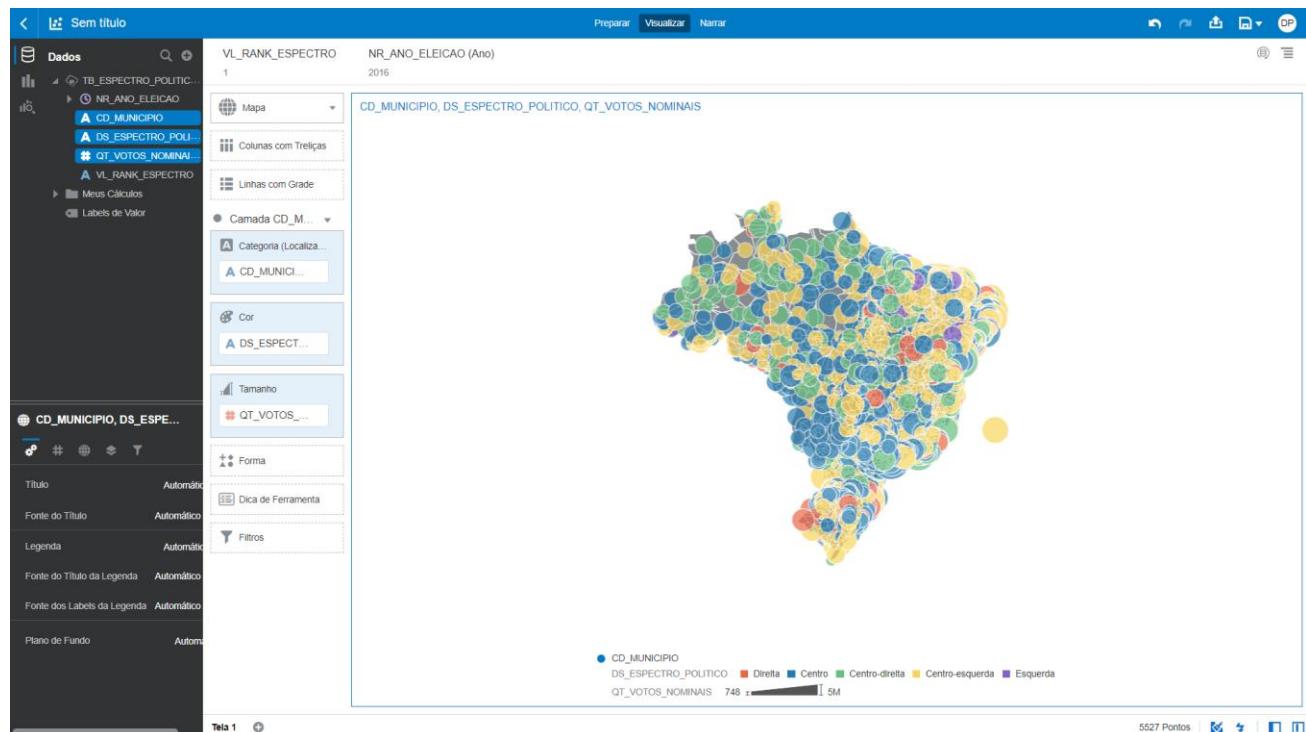
24. Um gráfico de barra será criado.



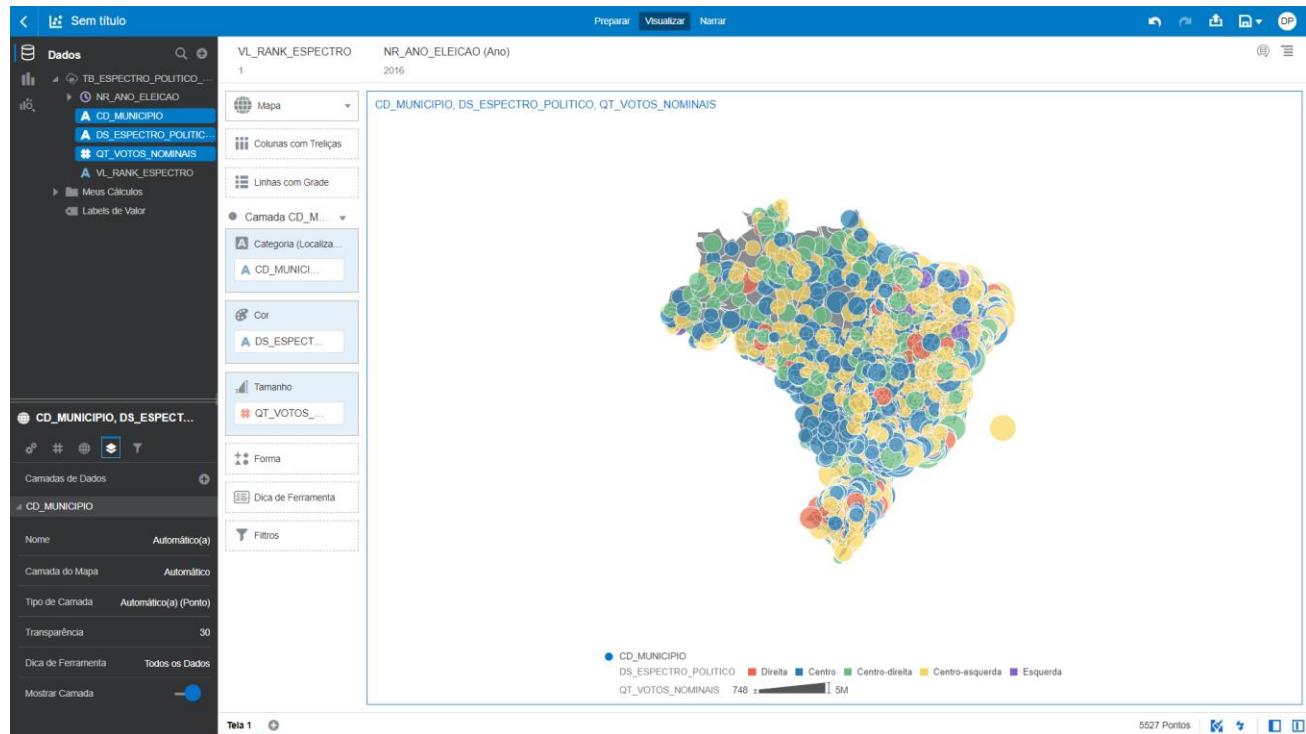
25. Na área da gramática do gráfico (à esquerda do gráfico), clique no botão de tipo Visualização e selecione a opção “Mapa”.



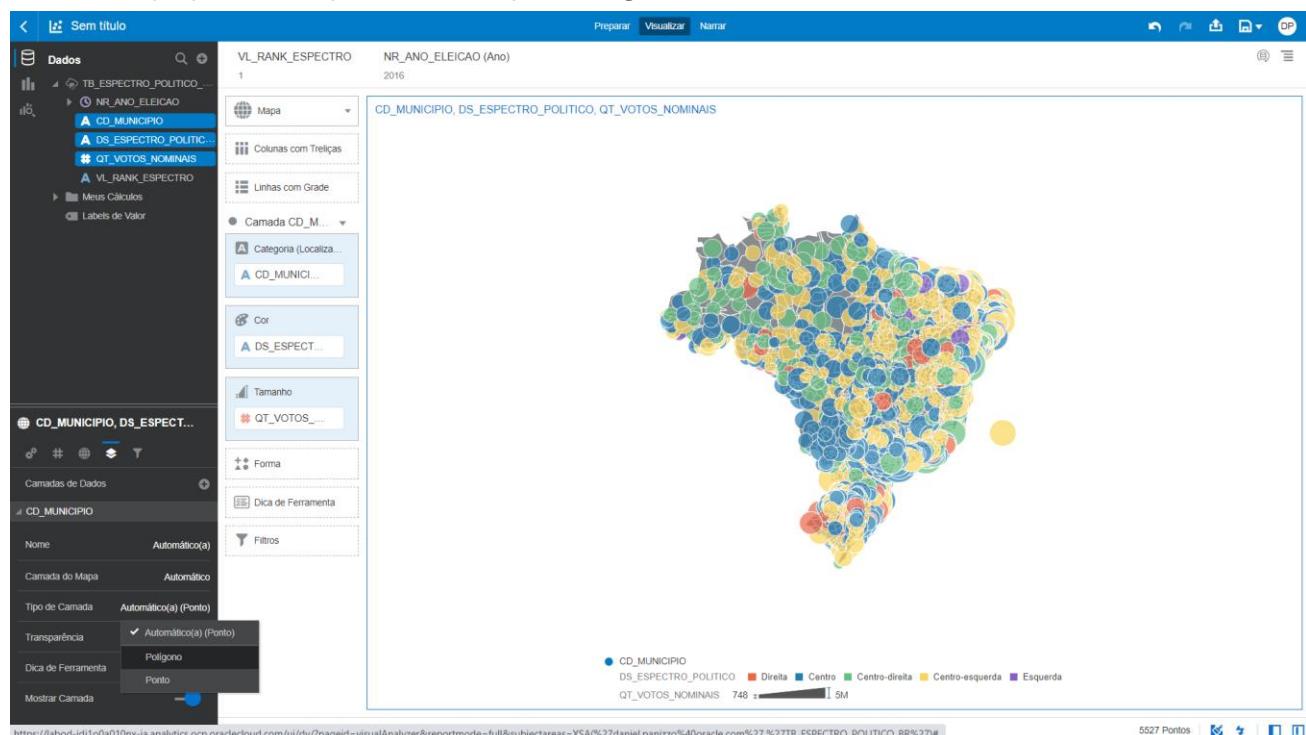
26. Um mapa do Brasil com círculos representando os municípios, deverá aparecer.



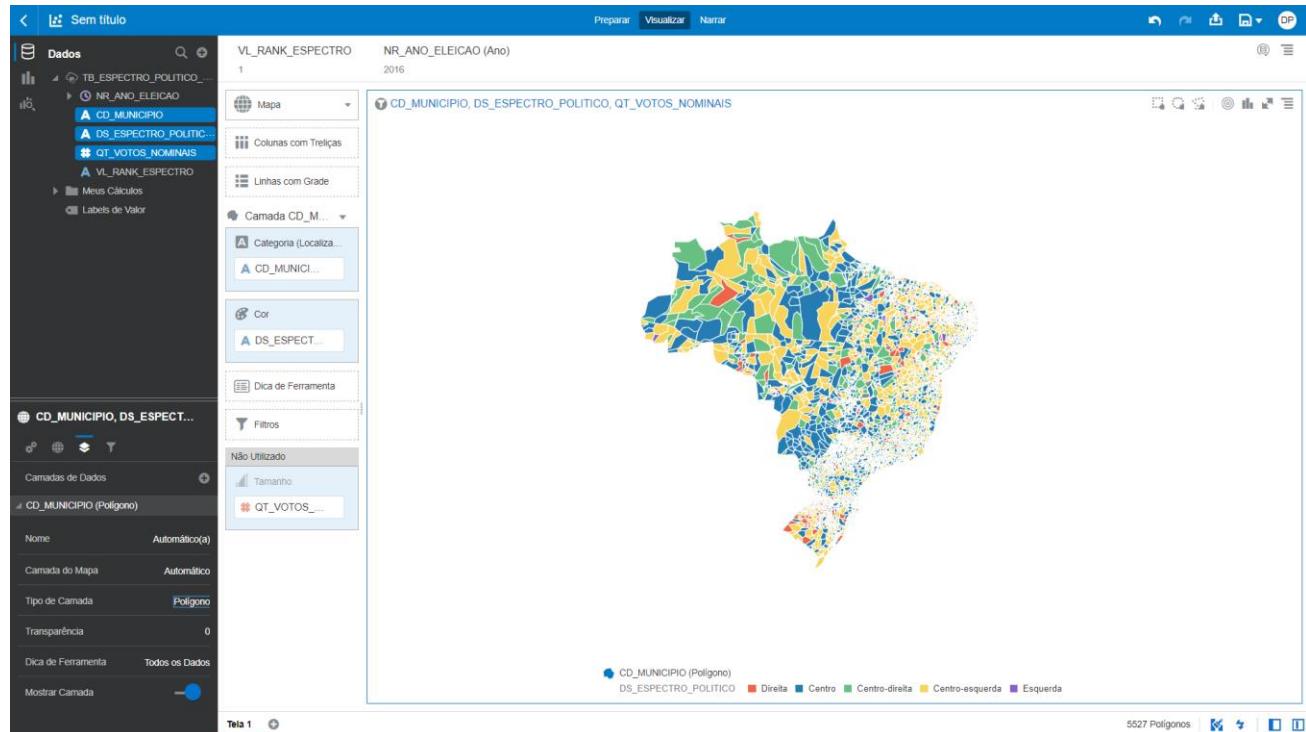
27. Na área escura, nas propriedades do mapa, selecione a aba “Camada de Dados”.



28. Altere a propriedade “Tipo de Camada” para “Polígono”.



29. Agora temos um mapa do Brasil do Principal Espectro Político por Município do primeiro turno das eleições de 2016.



30. Assim finalizamos nossa etapa de análise de dados.