

# Inovação com dados em nuvem

Planejando sua *road trip* com  
dados abertos e *analytics*

Daniel Panizzo  
08.06.2021



Este trabalho está licenciado sob uma Licença Creative Commons Atribuição-Compartilhalgual 4.0 Internacional.  
Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by-sa/4.0/>.



# Planejando sua road trip com dados abertos e analytics

---

Daniel Panizzo  
Engenheiro de Soluções

June, 2021, Version 2.0  
Copyright © 2021, Oracle and/or its affiliates  
Public

# ÍNDICE

---

<b>INTRODUÇÃO</b>	<b>5</b>
<b>LAB.1: COLETAR OS DADOS ABERTOS</b>	<b>6</b>
Obtendo os códigos do Github	6
Coletando os dados da PRF	8
Coletando os dados do INMET	13
<b>LAB.2: ARMAZENAR OS DADOS ABERTOS</b>	<b>14</b>
Carregando os dados no Object Storage	14
Armazenando os dados no banco de dados	19
<b>LAB.4: ANALISAR OS DADOS ABERTOS</b>	<b>23</b>
Conectando o Analytics com o ADW	23
Importando os metadados	27
Visualizando os dados	29

## INTRODUÇÃO

Neste passo-a-passo, iremos analisar grandes volumes de dados abertos e, para isso, passaremos pelo fluxo geral de um projeto 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 superior, 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:

- [Dados abertos do Histórico Meteorológico das estações automáticas do INMET](#)
- [Dados abertos de acidentes agrupados por ocorrência da PRF](#)

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/pendata](https://github.com/dspanizzo/pendata) no Github.

The screenshot shows the GitHub repository page for `dspanizzo/open-data`. The repository has 1746398 commits, 15 days ago, and 17 commits. The files listed are `br`, `.gitattributes`, `.gitignore`, `LICENSE`, and `README.md`. The `README.md` file is open, showing the title "Open-Data" and a description: "Uma série de python notebooks para auxiliar no desenvolvimento de programas para preparação de fontes de dados abertas." The "Processo" section describes the data preparation process in four steps: 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 aquisição dos dados.

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

The screenshot shows the GitHub repository page for `dspanizzo/open-data` with the "Code" dropdown menu open. The menu options are "Clone", "Open with GitHub Desktop", and "Download ZIP". The "Clone" option is selected, showing the "HTTPS" method and the URL `https://github.com/dspanizzo/open-data`. The "Open with GitHub Desktop" option is also visible.

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.

All Apps Documents Web More ▾ 7563

**Best match**

**Jupyter Notebook (Anaconda3)**  
App

**Apps**

**Jupyter Notebook (tensorflow\_env)** >

**Search the web**

jupyter - See web results >

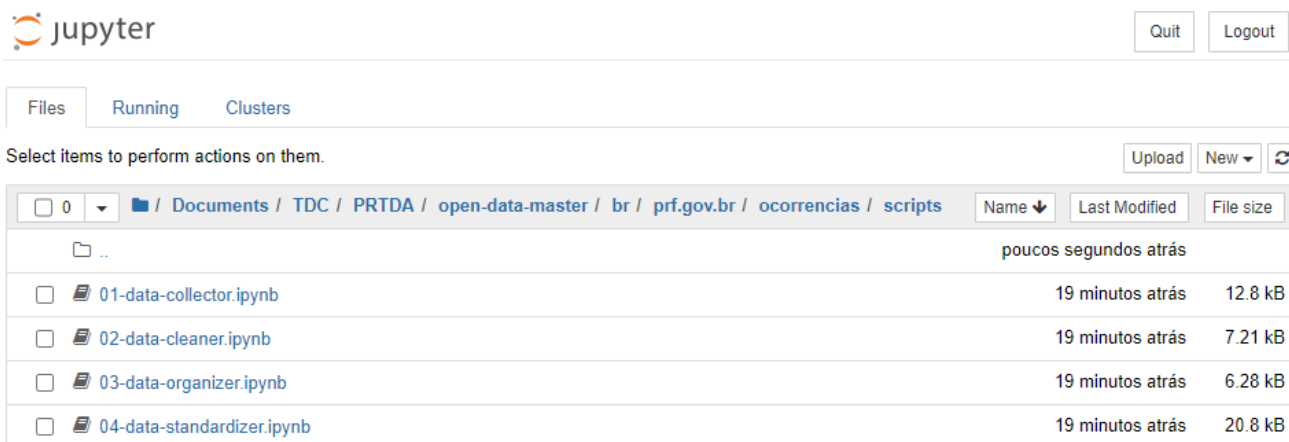
**Jupyter Notebook (Anaconda3)**  
App

- Open
- Run as administrator
- Open file location
- Pin to Start
- Pin to taskbar
- Uninstall

jupyter | Notebook (Anaconda3)

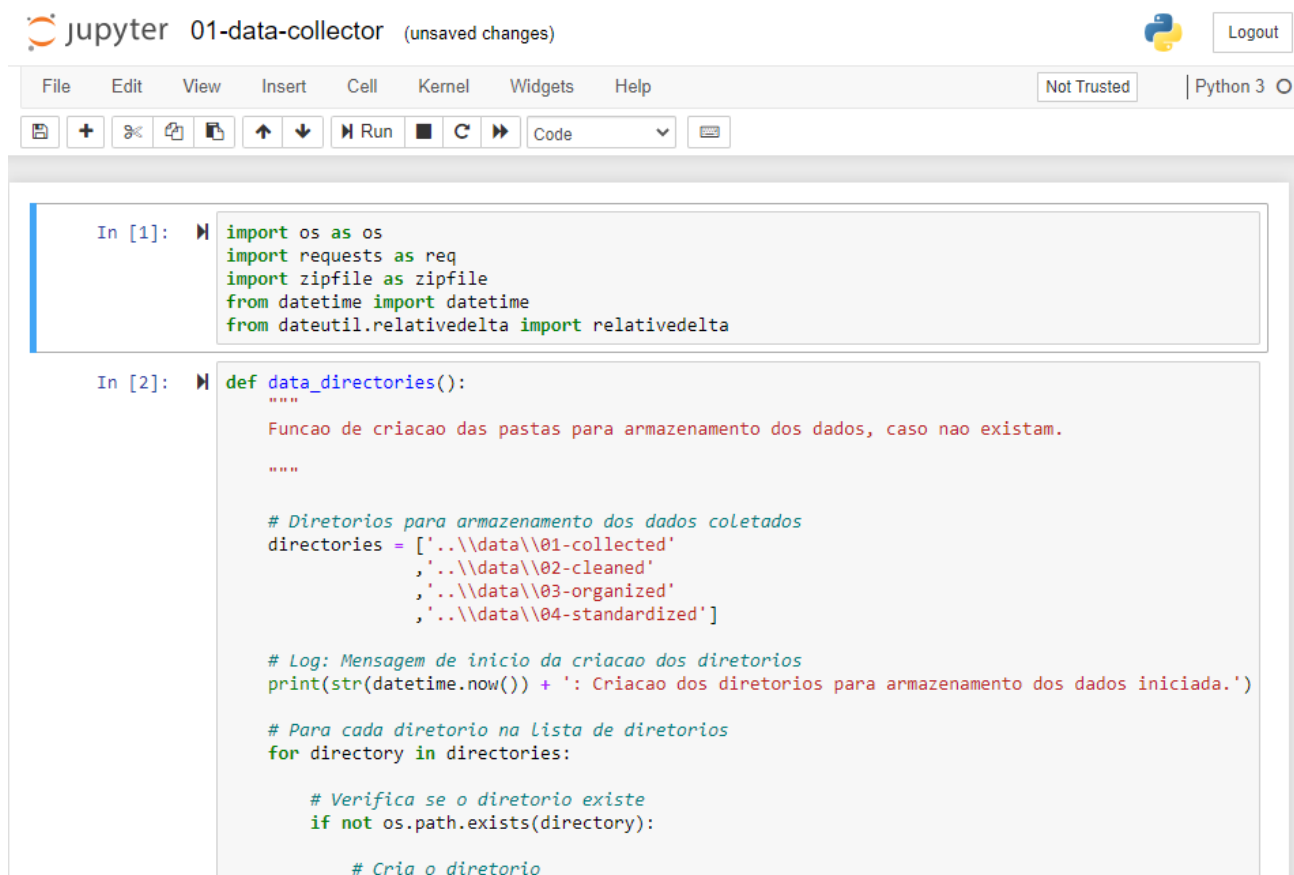
## Coletando os dados da PRF

- 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 -> prf.gov.br -> ocorrencias -> scripts”. Os notebooks dentro desta pasta são responsáveis por coletar e preparar os dados oriundos dos Dados Abertos de Ocorrências da Polícia Rodoviária Federal.



Name	Last Modified	File size
01-data-collector.ipynb	19 minutos atrás	12.8 kB
02-data-cleaner.ipynb	19 minutos atrás	7.21 kB
03-data-organizer.ipynb	19 minutos atrás	6.28 kB
04-data-standardizer.ipynb	19 minutos atrás	20.8 kB

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



```
In [1]: import 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.
    """
    # Diretorios 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 diretorio na lista de diretorios
    for directory in directories:
        # Verifica se o diretorio existe
        if not os.path.exists(directory):
            # Cria o diretorio
```



7. Este é o notebook responsável por coletar os arquivos do repositório da Polícia Rodoviária Federal. 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 “Fim do download do arquivo...”.

```
In [7]: def main():
        """
        Coleta os dados de acidentes por ocorrencia nas rodovias federais.
        Origem: Polícia Rodoviaria Federal (PRF)
        """

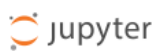
        # Prepara o dicionario de variaveis (kwargs = keyworded arguments)
        kwargs = {}

        # Coleta os arquivos
        collect_data(**kwargs)

    if __name__ == '__main__':
        main()





2021-05-30 21:28:40.015309: Criacao dos diretorios para armazenamento dos dados iniciada.
2021-05-30 21:28:40.016306: Directorio ..\data\01-collected ja existe.
2021-05-30 21:28:40.016306: Directorio ..\data\02-cleaned ja existe.
2021-05-30 21:28:40.016306: Directorio ..\data\03-organized ja existe.
2021-05-30 21:28:40.016306: Directorio ..\data\04-standardized ja existe.
2021-05-30 21:28:40.017304: Criacao dos diretorios para armazenamento dos dados finalizada.
2021-05-30 21:28:40.017304: Coleta dos arquivos iniciada.
2021-05-30 21:28:40.017304: Inicio do download do arquivo ..\data\01-collected\jdDLrQIf33xXSce.zip
2021-05-30 21:28:40.298118: URL encontrada, fazendo o download do arquivo: ..\data\01-collected\jdDLrQIf33xXSce.zip
2021-05-30 21:28:40.633190: Descompactando o arquivo ..\data\01-collected\jdDLrQIf33xXSce.zip
2021-05-30 21:28:40.717693: Fim do download do arquivo ..\data\01-collected\jdDLrQIf33xXSce.zip
2021-05-30 21:28:40.718659: Inicio do download do arquivo ..\data\01-collected\kRBUylqz6DyQznN.zip
2021-05-30 21:28:40.956394: URL encontrada, fazendo o download do arquivo: ..\data\01-collected\kRBUylqz6DyQznN.zip
2021-05-30 21:28:41.360641: Descompactando o arquivo ..\data\01-collected\kRBUylqz6DyQznN.zip
2021-05-30 21:28:41.452467: Fim do download do arquivo ..\data\01-collected\kRBUylqz6DyQznN.zip
2021-05-30 21:28:41.452467: Inicio do download do arquivo ..\data\01-collected\MaC6cieXSfACNWT.zip
2021-05-30 21:28:41.709859: URL encontrada, fazendo o download do arquivo: ..\data\01-collected\MaC6cieXSfACNWT.zip
2021-05-30 21:28:42.494102: Descompactando o arquivo ..\data\01-collected\MaC6cieXSfACNWT.zip
2021-05-30 21:28:42.587850: Fim do download do arquivo ..\data\01-collected\MaC6cieXSfACNWT.zip
2021-05-30 21:28:42.587850: Inicio do download do arquivo ..\data\01-collected\nqvFu7xEf6HhbAq.zip
2021-05-30 21:28:42.842960: URL encontrada, fazendo o download do arquivo: ..\data\01-collected\nqvFu7xEf6HhbAq.zip
2021-05-30 21:28:43.444564: Descompactando o arquivo ..\data\01-collected\nqvFu7xEf6HhbAq.zip
2021-05-30 21:28:43.555218: Fim do download do arquivo ..\data\01-collected\nqvFu7xEf6HhbAq.zip
2021-05-30 21:28:43.555218: Verificando a existencia de arquivos nao relacionados com os dados.
2021-05-30 21:28:43.555218: Remocao dos arquivos nao relacionados com os dados concluida.
2021-05-30 21:28:43.555218: 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.

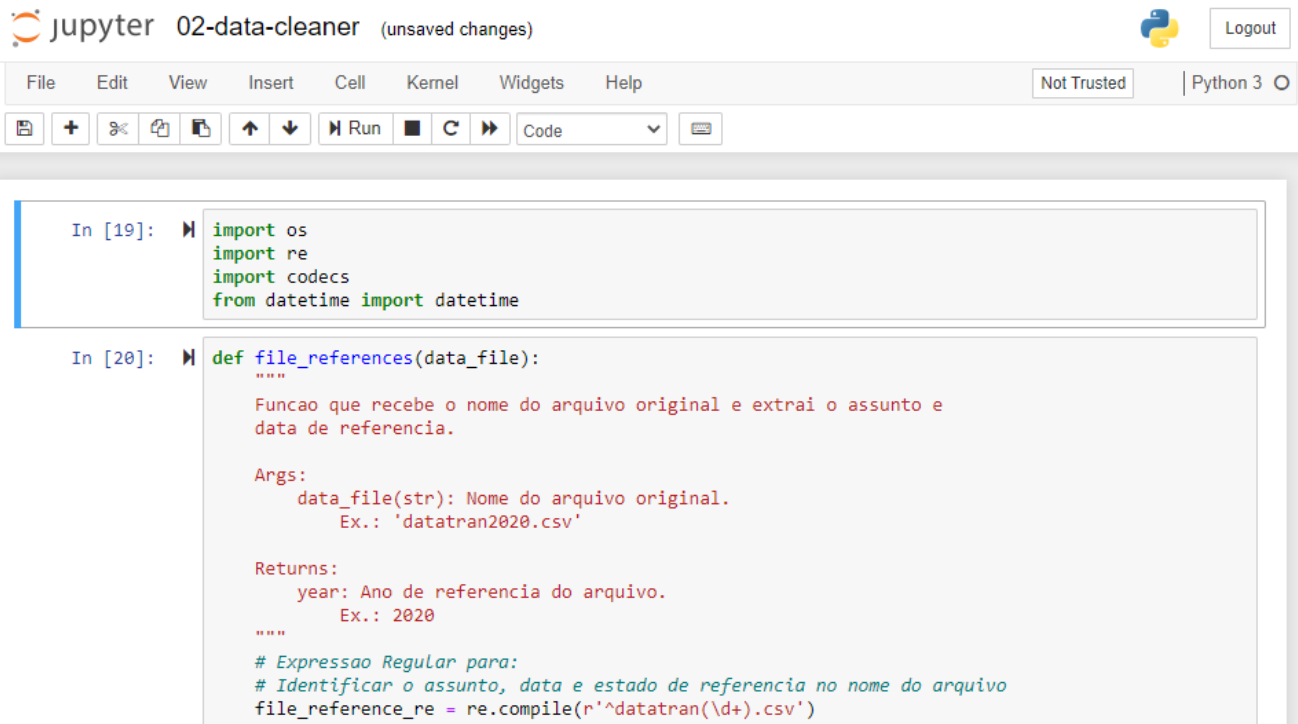
 Quit Logout

Files Running Clusters

Select items to perform actions on them. Upload New Refresh

<input type="checkbox"/>	0	Name	Last Modified	File size
<input type="checkbox"/>	..	poucos segundos atrás		
<input type="checkbox"/>	 01-data-collector.ipynb	Running	poucos segundos atrás	12.8 kB
<input type="checkbox"/>	 02-data-cleaner.ipynb		21 minutos atrás	7.21 kB
<input type="checkbox"/>	 03-data-organizer.ipynb		21 minutos atrás	6.28 kB
<input type="checkbox"/>	 04-data-standardizer.ipynb		21 minutos atrás	20.8 kB

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



```
In [19]: import os
import re
import codecs
from datetime import datetime

In [20]: def file_references(data_file):
        """
        Funcao que recebe o nome do arquivo original e extrai o assunto e
        data de referencia.

        Args:
            data_file(str): Nome do arquivo original.
            Ex.: 'datatran2020.csv'

        Returns:
            year: Ano de referencia do arquivo.
            Ex.: 2020
        """
        # Expressao Regular para:
        # Identificar o assunto, data e estado de referencia no nome do arquivo
        file_reference_re = re.compile(r'^datatran(\d+).csv')
```

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”.

```
In [23]: def main():
        """
        Limpa os dados de acidentes por ocorrencia nas rodovias federais.
        Origem: Policia Rodoviaria Federal (PRF)
        """

        ### Limpa os arquivos
        clean_files()

    if __name__ == '__main__':
        main()

2021-05-30 21:57:58.586982: Limpeza dos arquivos iniciada.
2021-05-30 21:57:58.586982: Inicio da codificacao do arquivo ..\data\01-collected\datatran2017.csv
2021-05-30 21:57:58.712085: Fim da codificacao do arquivo ..\data\02-cleaned\datatran2017.csv
2021-05-30 21:57:58.712085: Inicio da codificacao do arquivo ..\data\01-collected\datatran2018.csv
2021-05-30 21:57:58.789029: Fim da codificacao do arquivo ..\data\02-cleaned\datatran2018.csv
2021-05-30 21:57:58.791015: Inicio da codificacao do arquivo ..\data\01-collected\datatran2019.csv
2021-05-30 21:57:58.868634: Fim da codificacao do arquivo ..\data\02-cleaned\datatran2019.csv
2021-05-30 21:57:58.870629: Inicio da codificacao do arquivo ..\data\01-collected\datatran2020.csv
2021-05-30 21:57:58.928028: Fim da codificacao do arquivo ..\data\02-cleaned\datatran2020.csv
2021-05-30 21:57:58.930022: 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 dividir o arquivo por tipo de registro, separando CNPJs, Sócios e CNAEs. Ao finalizar o processo, a última mensagem exibida pelo log da função “main” deverá ser “Organizacao dos arquivos finalizada”.

```
In [8]: def main():
        """
        Organiza os dados de acidentes por ocorrencia nas rodovias federais.
        Origem: Policia Rodoviaria Federal (PRF)
        """

        ### Organiza os arquivos
        organize_files()

    if __name__ == '__main__':
        main()

2021-05-30 21:59:11.723043: Organizacao dos arquivos iniciada.
2021-05-30 21:59:11.723043: Verificando registros do arquivo datatran2017.csv
2021-05-30 21:59:12.721097: Resultado: Normais = 89563 - Corrigidas = 0
2021-05-30 21:59:12.721097: Verificando registros do arquivo datatran2018.csv
2021-05-30 21:59:13.502251: Resultado: Normais = 69295 - Corrigidas = 0
2021-05-30 21:59:13.502251: Verificando registros do arquivo datatran2019.csv
2021-05-30 21:59:14.224149: Resultado: Normais = 67446 - Corrigidas = 0
2021-05-30 21:59:14.224149: Verificando registros do arquivo datatran2020.csv
2021-05-30 21:59:14.895655: Resultado: Normais = 63530 - Corrigidas = 0
2021-05-30 21:59:14.895655: 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”.

```
In [13]: def main():
        """
        Padroniza os dados de acidentes por ocorrencia nas rodovias federais.
        Origem: Policia Rodoviaria Federal (PRF)
        """

        ### Padroniza os arquivos
        standardize_files()

    if __name__ == '__main__':
        main()

2021-06-06 15:23:39.375784: Padronizacao dos arquivos iniciada.
2021-06-06 15:24:06.518630: Arquivo datatran2017.csv padronizado.
2021-06-06 15:24:28.052108: Arquivo datatran2018.csv padronizado.
2021-06-06 15:24:47.910601: Arquivo datatran2019.csv padronizado.
2021-06-06 15:25:06.204641: Arquivo datatran2020.csv padronizado.
2021-06-06 15:25:06.217605: 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.

TDC > PRTDA > open-data-master > br > prf.gov.br > ocorrencias > data

Name	Date modified	Type	Size
01-collected	06/06/2021 15:22	File folder	
02-cleaned	06/06/2021 15:22	File folder	
03-organized	06/06/2021 15:22	File folder	
04-standardized	06/06/2021 15:25	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.


TDC > PRTDA > open-data-master > br > prf.gov.br > ocorrencias > data > 04-standardized

Name	Date modified	Type	Size
 datatran2017.csv	06/06/2021 15:24	Microsoft Excel Co...	27.262 KB
 datatran2018.csv	06/06/2021 15:24	Microsoft Excel Co...	21.304 KB
 datatran2019.csv	06/06/2021 15:24	Microsoft Excel Co...	20.573 KB
 datatran2020.csv	06/06/2021 15:25	Microsoft Excel Co...	19.487 KB

15. Assim, finalizamos a coleta de dados da Polícia Rodoviária Federal.

## Coletando os dados do INMET





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 -> inmet.gov.br -> historico -> scripts”. Os notebooks dentro desta pasta são responsáveis por coletar e preparar os dados oriundos do INMET.

 jupyter Quit Logout

Files Running Clusters









Select items to perform actions on them. Upload New Refresh

☐ 0 / Documents / TDC / PRTDA / open-data-master / br / inmet.gov.br / historico / scripts Name Last Modified File size

poucos segundos atrás		
<input type="checkbox"/>  01-data-collector.ipynb	41 minutos atrás	10.9 kB
<input type="checkbox"/>  02-data-cleaner.ipynb	41 minutos atrás	2.17 MB
<input type="checkbox"/>  03-data-organizer.ipynb	41 minutos atrás	1.47 MB
<input type="checkbox"/>  04-data-standardizer.ipynb	41 minutos atrás	862 kB

2. Assim como na etapa anterior “Coletando dados da PRF”, repita a mesma sequência de execução dos quatro notebooks dentro desta pasta. O resultado final deverá ser uma série de arquivos de cada estação automática do INMET dentro da pasta “br -> inmet.gov.br -> historico -> data -> 04-standardizer”.

TDC > PRTDA > open-data-master > br > inmet.gov.br > historico > data > 04-standardized

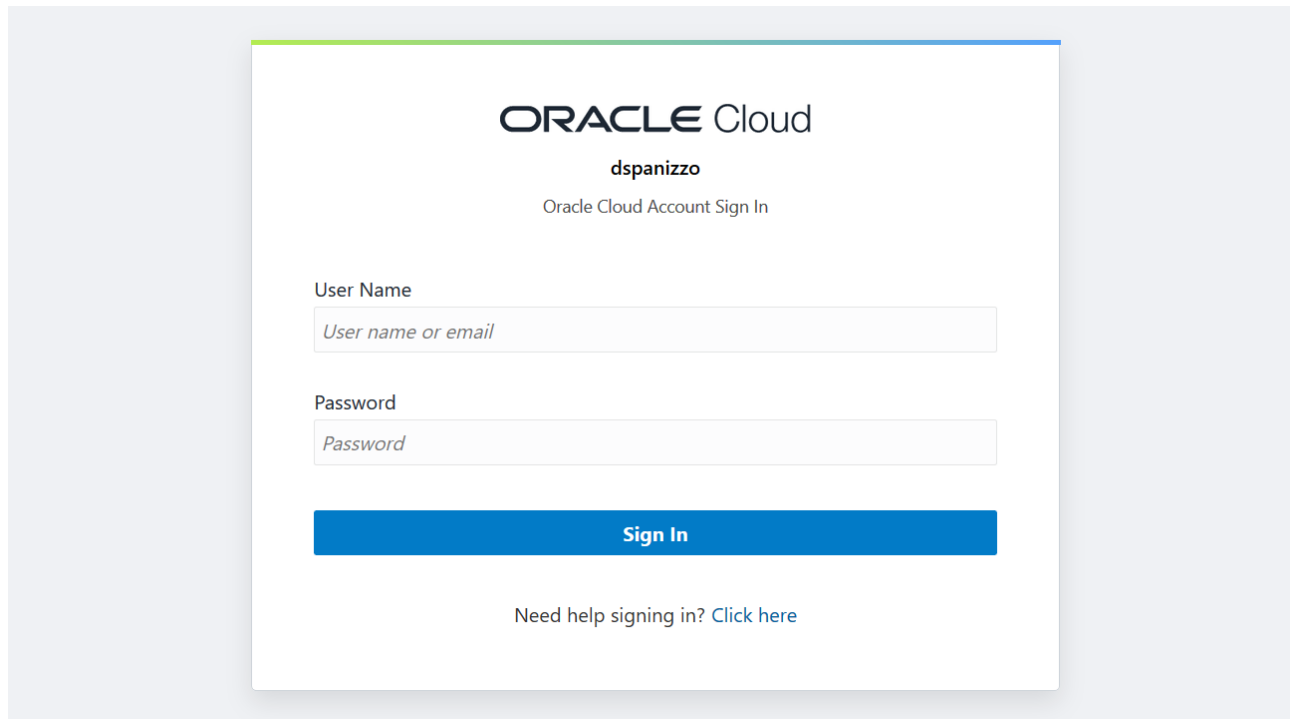
Name	Date modified	Type	Size
 INMET_CO_DF_A001_BRASILIA_01-01-20...	06/06/2021 15:39	Microsoft Excel Co...	2.264 KB
 INMET_CO_DF_A042_BRAZLANDIA_01-0...	06/06/2021 15:39	Microsoft Excel Co...	2.302 KB
 INMET_CO_DF_A045_AGUAS EMENDAD...	06/06/2021 15:39	Microsoft Excel Co...	2.372 KB
 INMET_CO_DF_A046_GAMA (PONTE ALT...	06/06/2021 15:39	Microsoft Excel Co...	2.432 KB
 INMET_CO_DF_A047_PARANOA (COOPA-...	06/06/2021 15:39	Microsoft Excel Co...	2.414 KB
 INMET_CO_GO_A002_GOIANIA_01-01-20...	06/06/2021 15:39	Microsoft Excel Co...	2.228 KB
 INMET_CO_GO_A003_MORRINHOS_01-0...	06/06/2021 15:39	Microsoft Excel Co...	2.273 KB
 INMET_CO_GO_A005_PORANGATU_01-01...	06/06/2021 15:40	Microsoft Excel Co...	2.007 KB

3. Assim, finalizamos a coleta de dados do INMET.

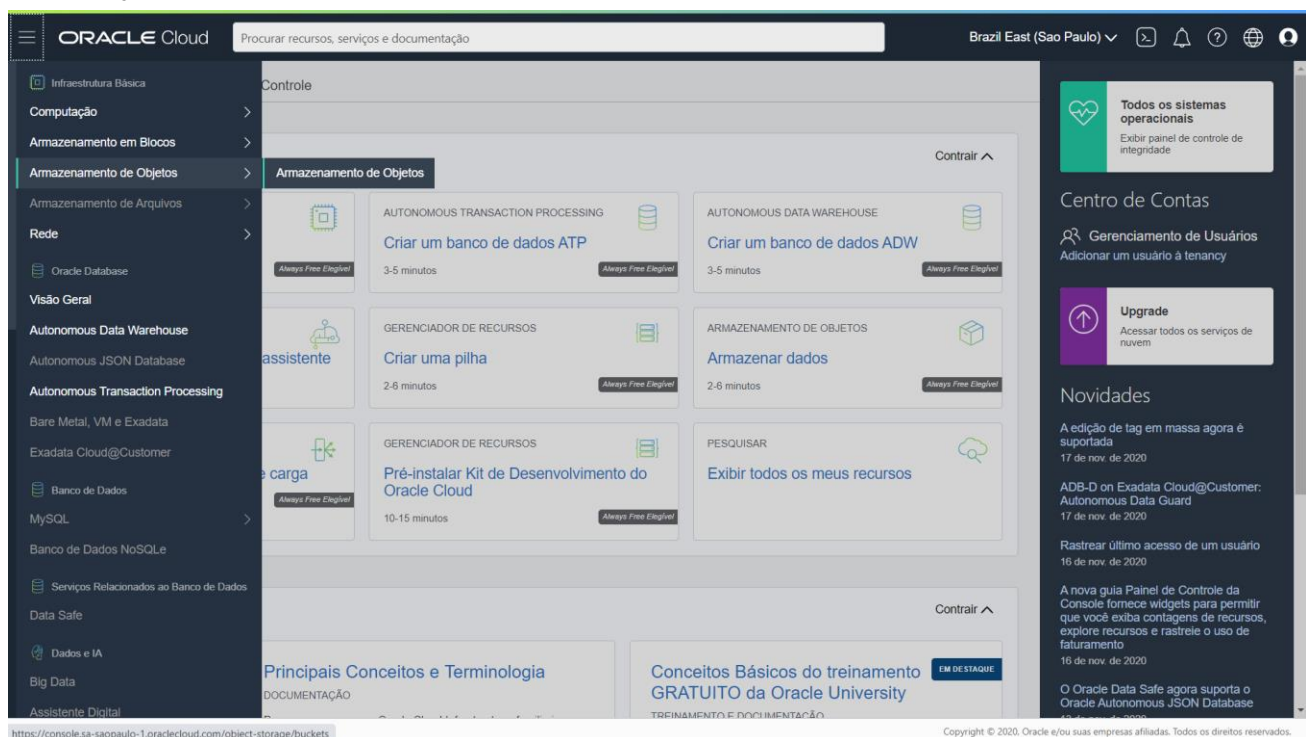
## LAB.2: ARMAZENAR OS DADOS ABERTOS

### 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”.





### 3. Clique no botão “Criar Bucket”.

The screenshot shows the Oracle Cloud console interface. At the top, there's a navigation bar with the Oracle Cloud logo, a search bar, and the region 'Brazil East (Sao Paulo)'. The main content area is titled 'Buckets em analytics Compartimento'. Below the title, there's a message about storage limits: 'Você está usando aproximadamente 0 bytes do limite de 20 GiB de Armazenamento de Objetos e Armazenamento de Arquivos Compactados combinados grátis. Faça upgrade para usar armazenamento ilimitado. Mostrar detalhes.' A blue button labeled 'Criar Bucket' is visible. Below this, there's a table with columns: 'Nome', 'Camada de Armazenamento', 'Visibilidade', and 'Criado'. The table is currently empty, showing 'Nenhum item encontrado.' and 'Mostrando 0 itens < 1 de 1 >'. On the left sidebar, there's a menu with 'Armazenamento de Objetos' and 'Escopo da Lista' set to 'analytics'.

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

The screenshot shows the 'Criar Bucket' dialog box in the Oracle Cloud console. The dialog has a title bar with 'Ajuda' and 'Cancelar' buttons. The main content area is titled 'CRIAR BUCKET' and contains several sections: 'NOME DO BUCKET' with a text input field containing 'opendata'; 'CAMADA DE ARMAZENAMENTO' with a note about specifying the layer during creation and two radio buttons: 'PADRÃO' (selected) and 'ARQUIVO COMPACTADO'; 'EVENTOS DE OBJETO' with a note and a checkbox 'EMITIR EVENTOS DE OBJETO'; 'CONTROLE DE VERSÃO DO OBJETO' with a note and a checkbox 'ATIVAR CONTROLE DE VERSÃO DO OBJETO'; 'CRİPTOGRAFIA' with a note and two radio buttons: 'CRİPTOGRAFAR USANDO CHAVES GERENCIADAS PELA ORACLE' (selected) and 'CRİPTOGRAFAR USANDO CHAVES GERENCIADAS PELO CLIENTE'; and 'TAGS' with a note and a link 'Saiba mais sobre tags'. At the bottom, there's a table for 'TAGS' with columns 'NAMESPACE DE TAG', 'CHAVE DE TAG', and 'VALOR'. The first row has 'Nenhuma (adicionar uma ta...)' in the first column. A '+ Tag Adicional' button is at the bottom right.

5. Clique no bucket “open-data” que acabou de criar.

Armazenamento de Objetos

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](#)

Você está usando aproximadamente 0 bytes do limite de 20 GiB de Armazenamento de Objetos e Armazenamento de Arquivos Compactados combinados grátis. [Faça upgrade](#) para usar armazenamento ilimitado. [Mostrar detalhes](#).

[Criar Bucket](#)

Nome	Camada de Armazenamento	Visibilidade	Criado
<a href="#">opendata</a>	Padrão	Privado	seg., 30 de nov. de 2020 13:09:01 UTC

Mostrando 1 item < 1 de 1 >

Termos de Uso e Privacidade Preferências de Cookies

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

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

Armazenamento de Objetos » Detalhes do Bucket

opendata

[Editar Visibilidade](#) [Mover Recurso](#) [Criptografar](#) [Adicionar Tags](#) [Excluir](#)

Informações do Bucket

Tags

Visibilidade: Privado

Namespace: grtzj7pvp01t

Camada de Armazenamento: Padrão

Contagem aproximada: 0 objetos

ETag: eaf8d3c1-2bf2-49cf-a16d-0d8687acb0db

OCID: ...26ow5jrj [Mostrar](#) [Copiar](#)

Chave de Criptografia: Chave gerenciada pela Oracle [Atribuir](#)

Criado: seg., 30 de nov. de 2020 13:09:01 UTC

Compartimento: analytics

Tamanho Aproximado: 0 bytes

Emitir Eventos de Objeto: Desativado [Editar](#)

Controle de Versão do Objeto: Desativado [Editar](#)

Recursos

Objetos

[Fazer Upload](#) [Mais Ações](#)

Pesquisar por prefixo

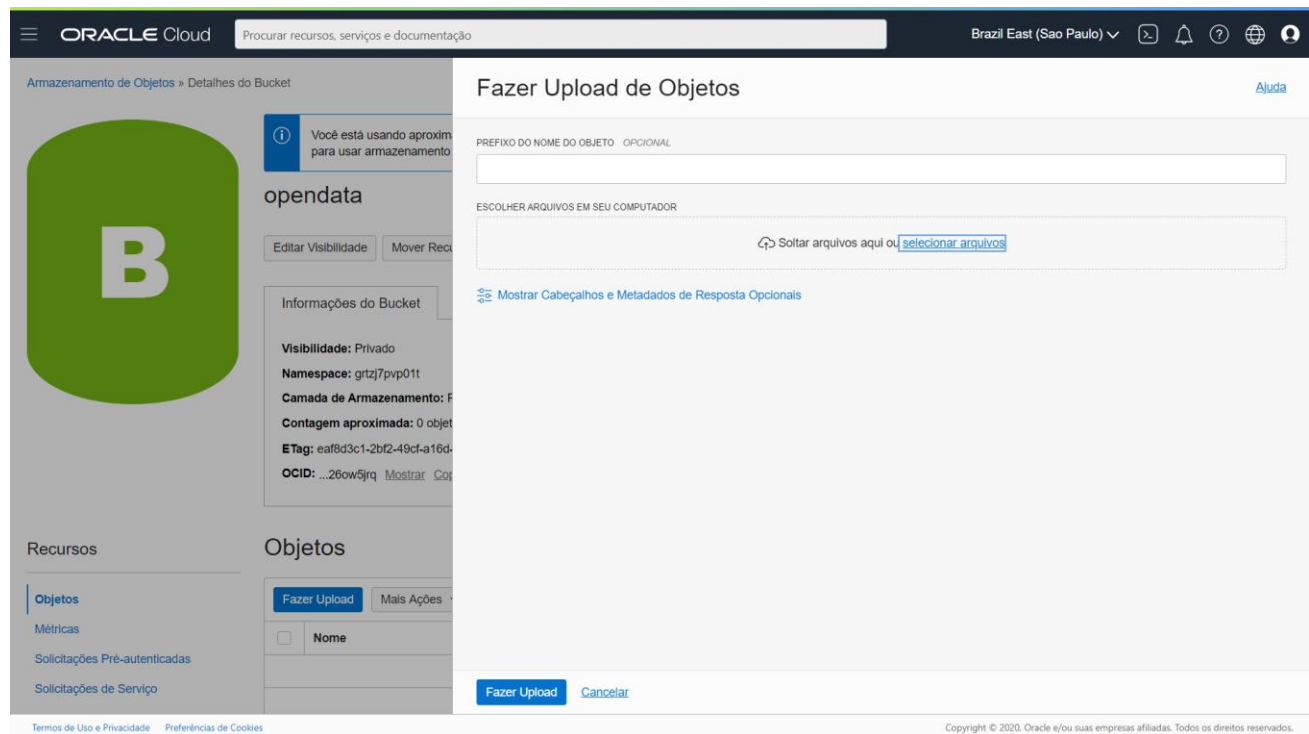
Nome	Última Modificação	Tamanho	Status
Nenhum item encontrado.			

Termos de Uso e Privacidade Preferências de Cookies

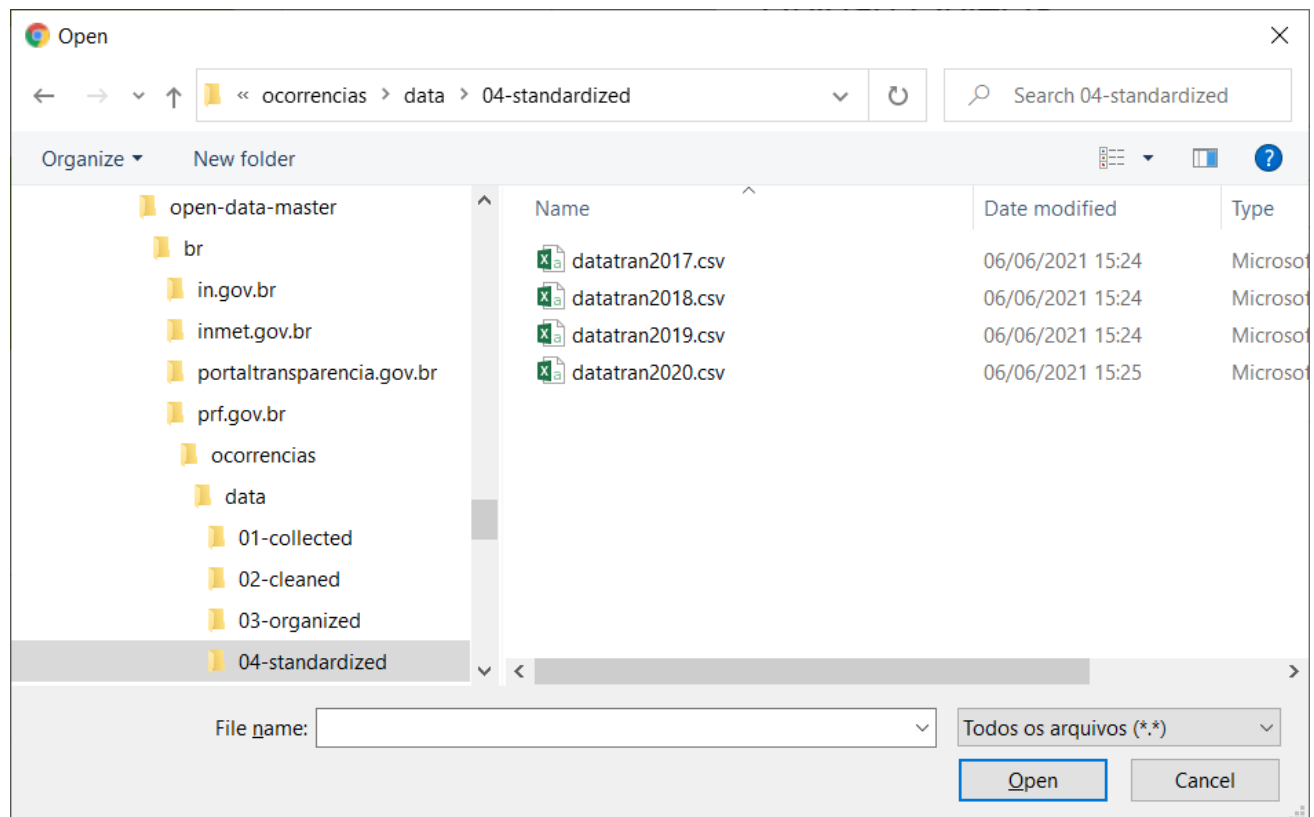
Copyright © 2020, Oracle e/ou suas empresas afiliadas. Todos os direitos reservados.



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



8. Navegue até a pasta onde estão armazenados os arquivos padronizados da Polícia Rodoviária Federal e selecione todos os arquivos.



9. Clique no botão “Fazer Upload”.

The screenshot shows the Oracle Cloud console interface. On the left, there's a sidebar with 'Resources' and 'Objects' sections. The main area displays the 'prf-ocorrencias' bucket details, including visibility (Private), namespace (id100a010nx), and storage tier (Standard). The 'Upload Objects' dialog is open, showing a list of files to be uploaded. A warning message indicates that 4 existing objects will be replaced. The 'Upload' button is highlighted at the bottom of the dialog.

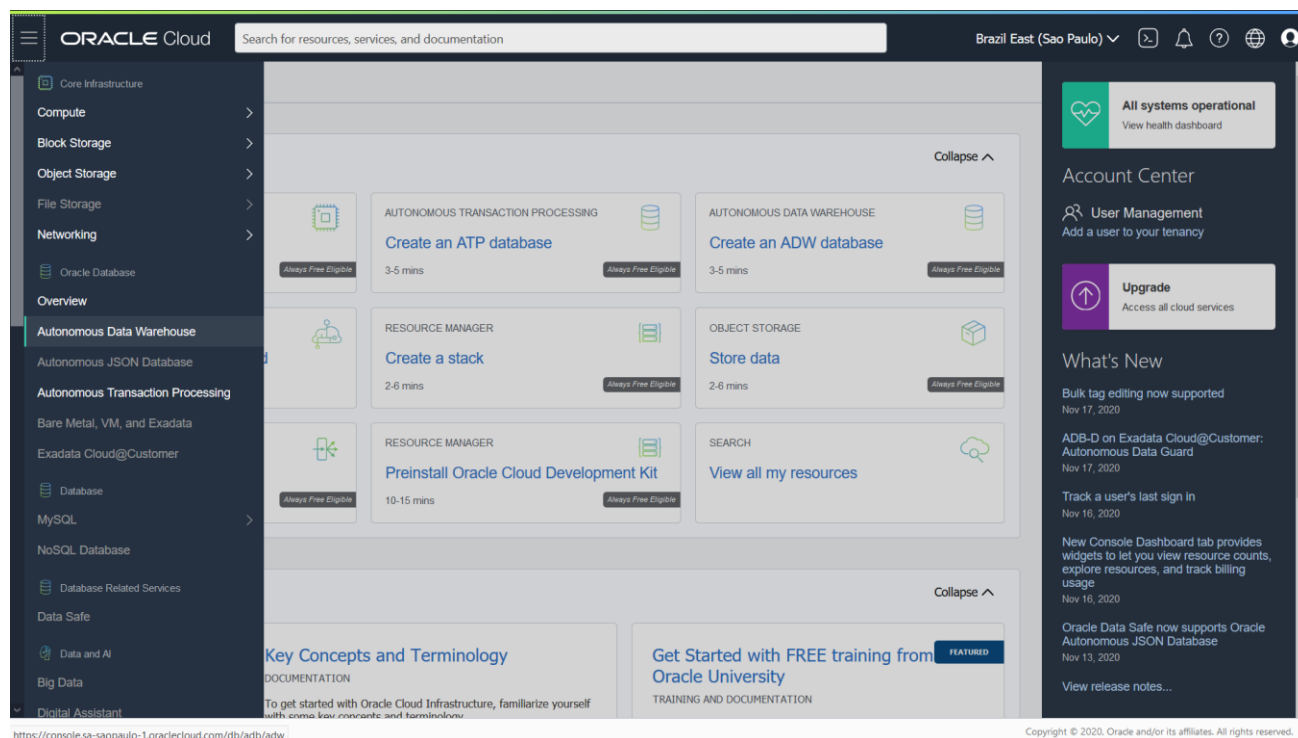
10. Aguarde a finalização do upload de todos os conjuntos de dados.

The screenshot shows the Oracle Cloud console interface. On the left, there's a sidebar with 'Resources' and 'Objects' sections. The main area displays the 'prf-ocorrencias' bucket details, including visibility (Private), namespace (id100a010nx), and storage tier (Standard). The 'Upload Objects' dialog is open, showing a list of files to be uploaded. The status of each file is 'Finished'. The 'Close' button is highlighted at the bottom of the dialog.

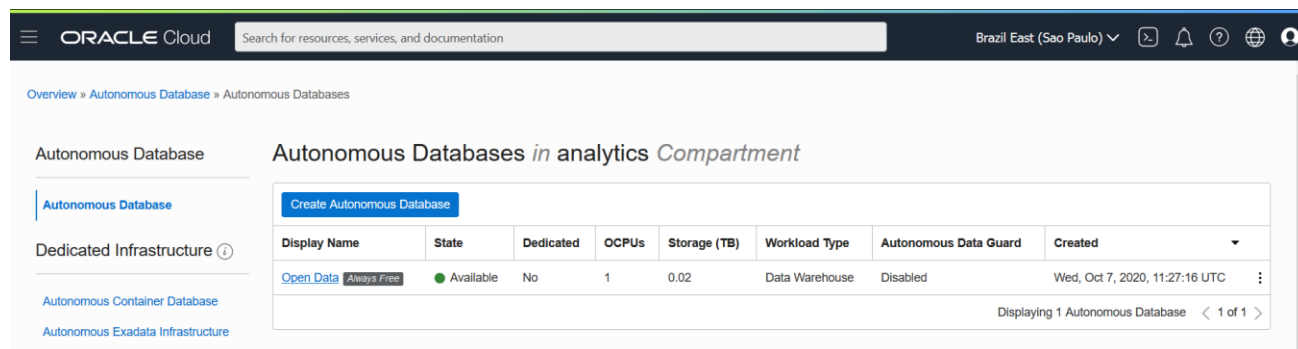
11. Repita os mesmos passos e faça, também, o upload dos arquivos padronizados de histórico das estações automáticas do INMET.
12. Aguarde a finalização do upload do conjunto de dados.
13. Assim finalizamos o upload dos conjuntos de dados para o Object Storage.

## Armazenando os dados no banco de dados

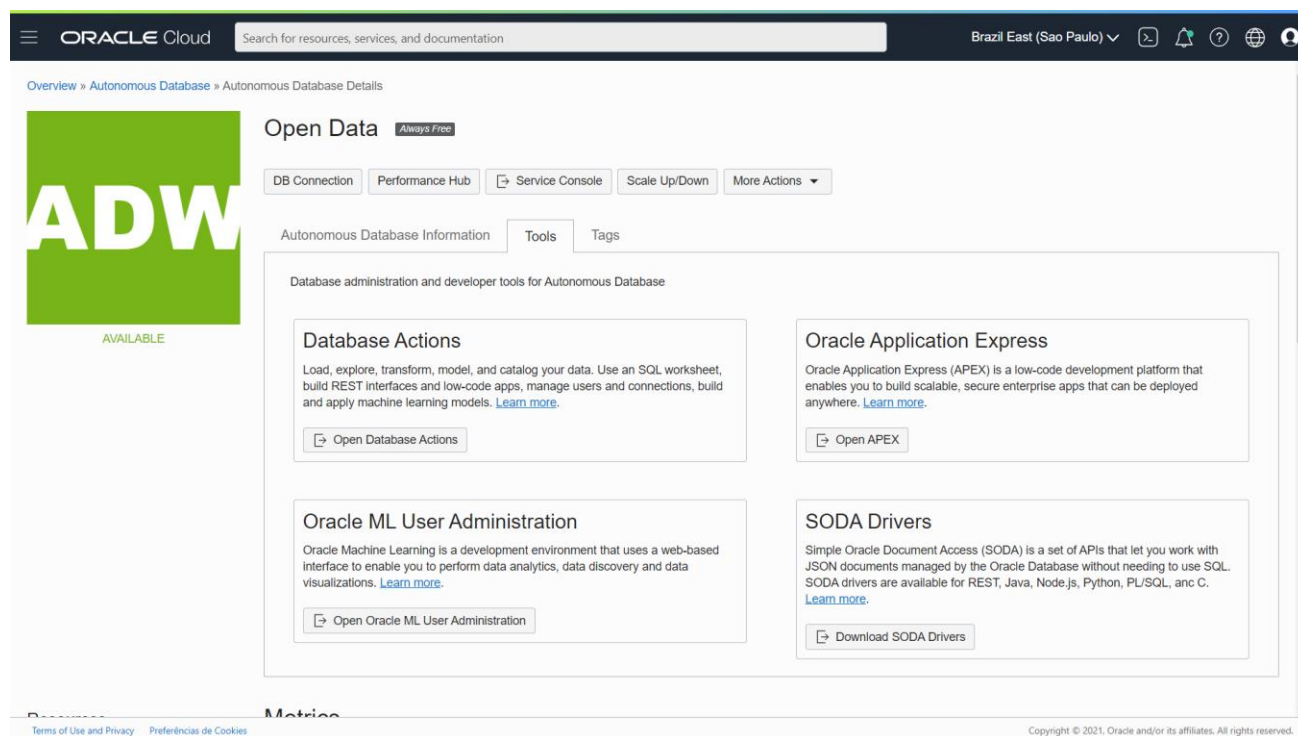
14. No menu de navegação do OCI, selecione “Autonomous Data Warehouse” entre as opções de “Oracle Database”.



15. 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.



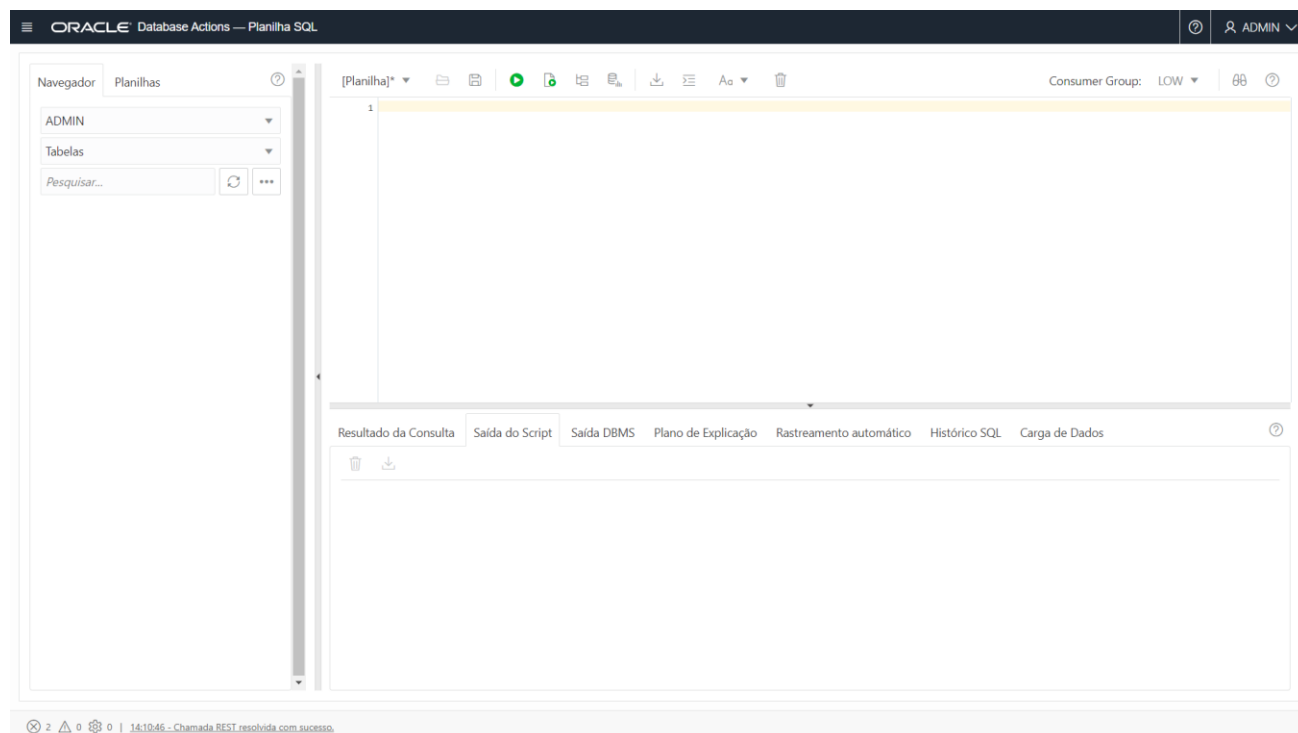
16. No seu ADW, selecione a aba “Tools” (“Ferramentas”) e clique no botão “Database Tools” (“SQL Developer Web”).



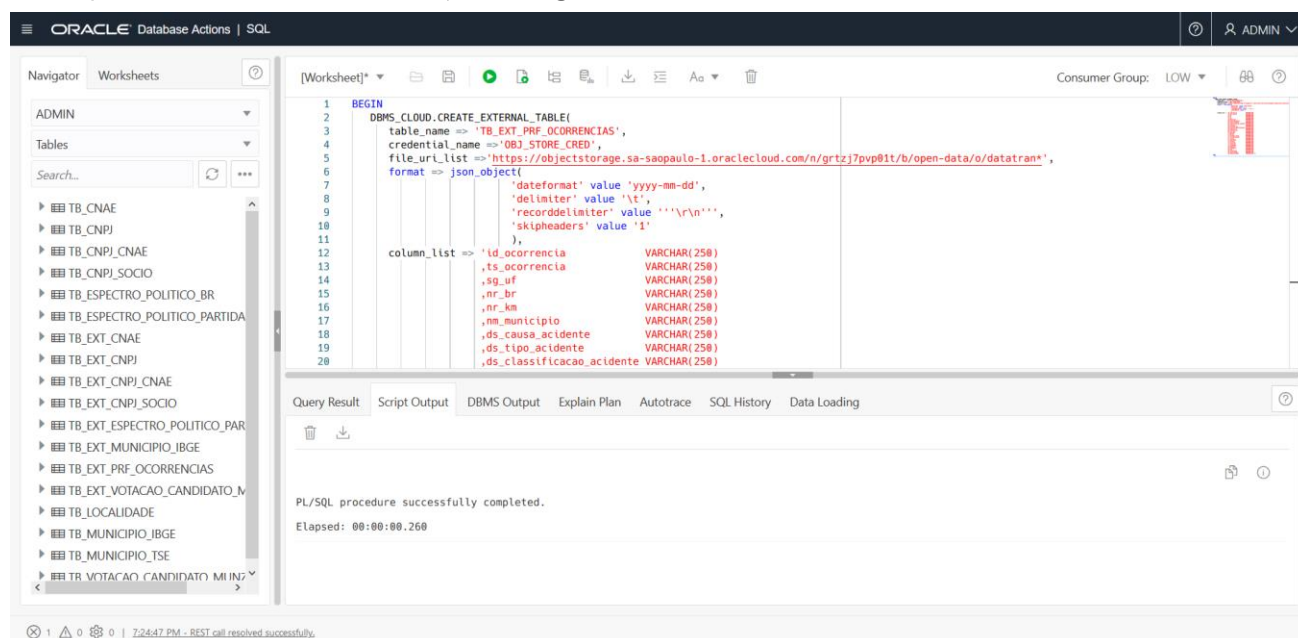
17. Acesse o SQL Developer Web com o usuário e senha do ADMIN. Definida durante a criação do seu ADW.



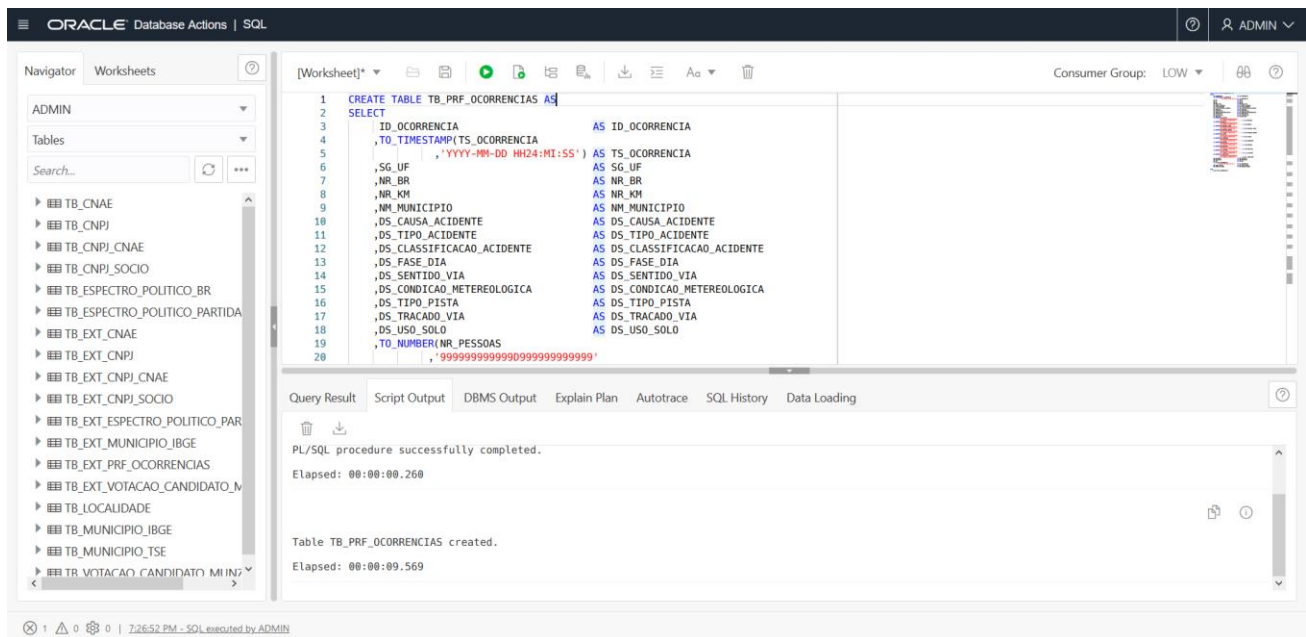
18. No ambiente do SQL Developer Web, é possível executar algumas instruções SQL para importar os dados armazenados no Object Storage para o ADW. Faremos este processo utilizando o recurso de tabelas externas.



19. Copie e cole o script “01\_TB\_EXT\_PRF\_OCORRENCIAS.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 armazenados no Object Storage.



20. Copie e cole o script “02\_TB\_PRF\_OCORRENCIAS.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.



21. Repita o mesmo processo com os scripts

- “03\_TB\_EXT\_INMET\_HISTORICO.sql”;
- “04\_TB\_INMET\_HISTORICO.sql”;

Assim teremos todos os dados armazenados dentro do banco de dados.

## LAB.4: ANALISAR OS DADOS ABERTOS

### Conectando o Analytics com o ADW

1. No sua ambiente do seu ADW, clique no botão “Conxão do DB”.

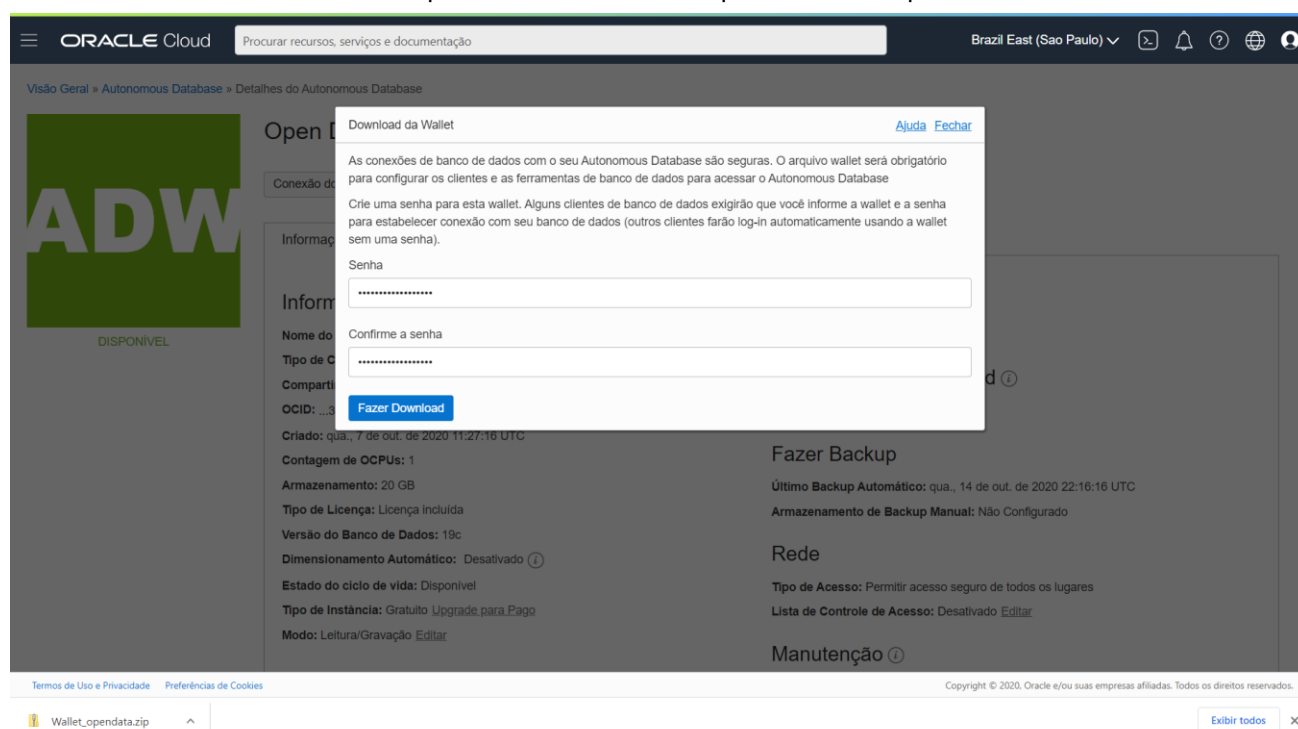
The screenshot shows the Oracle Cloud console interface. At the top, there's a navigation bar with the Oracle Cloud logo and a search bar. Below it, the breadcrumb trail reads 'Visão Geral » Autonomous Database » Detalhes do Autonomous Database'. The main content area is titled 'Open Data' with a sub-label 'Always Free'. On the left, there's a large green 'ADW' logo with the word 'DISPONÍVEL' below it. To the right of the logo, there are several buttons: 'Conexão do BD', 'Hub de Desempenho', 'Console de Serviço', 'Ampliar/Reduzir', and 'Mais Ações'. Below these buttons, there are tabs for 'Informações do Autonomous Database', 'Ferramentas', and 'Tags'. The 'Informações do Autonomous Database' tab is active, showing a list of details under 'Informações Gerais' and 'Infraestrutura'. The 'Informações Gerais' section includes fields like 'Nome do Banco de Dados', 'Tipo de Carga de Trabalho', 'Compartimento', 'OCID', 'Criado', 'Contagem de OCPUs', 'Armazenamento', 'Tipo de Licença', 'Versão do Banco de Dados', 'Dimensionamento Automático', 'Estado do ciclo de vida', 'Tipo de Instância', and 'Modo'. The 'Infraestrutura' section includes 'Infraestrutura Dedicada', 'Autonomous Data Guard', 'Fazer Backup', 'Rede', and 'Manutenção'.

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

The screenshot shows the Oracle Cloud console interface with a modal dialog box open. The dialog box is titled 'Conexão de Banco de Dados' and has a close button 'Fechar'. Inside the dialog, there's a message: 'Você precisará das credenciais do cliente e das informações de conexão para conectar-se ao seu banco de dados. As credenciais do cliente incluem a wallet.' Below this, there's a section titled 'Fazer Download das Credenciais do Cliente (Wallet)'. It contains a message: 'Para fazer download das credencias de cliente selecione o tipo de wallet e clique em Fazer Download da Wallet. Você será solicitado a criar uma senha para a wallet.' There's a dropdown menu for 'Tipo de Wallet' with 'Wallet da Instância' selected. Below the dropdown, there are two buttons: 'Download da Wallet' and 'Girar Wallet'. At the bottom of the dialog, it says 'Wallet girada pela última vez: -' and there's a 'Fechar' button.

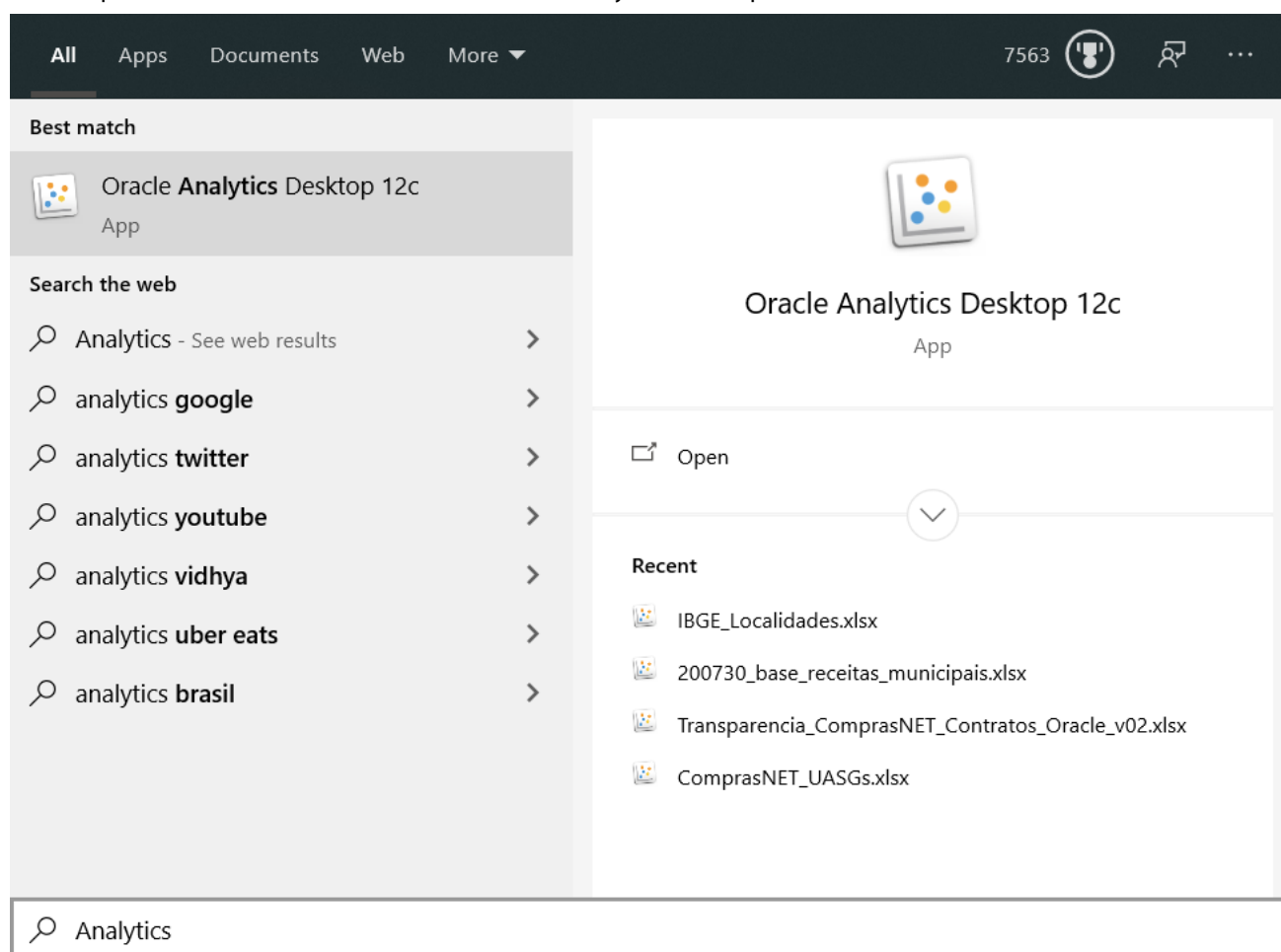


3. Informe a senha do ADMIN e clique em “Fazer Download” para baixar o arquivo ZIP.



The screenshot shows the Oracle Cloud console interface. A modal window titled "Download da Wallet" is open, displaying instructions for downloading the wallet and creating a password. The modal includes fields for "Senha" and "Confirme a senha", and a "Fazer Download" button. The background shows the details of an Autonomous Database, including its name, type, and status.

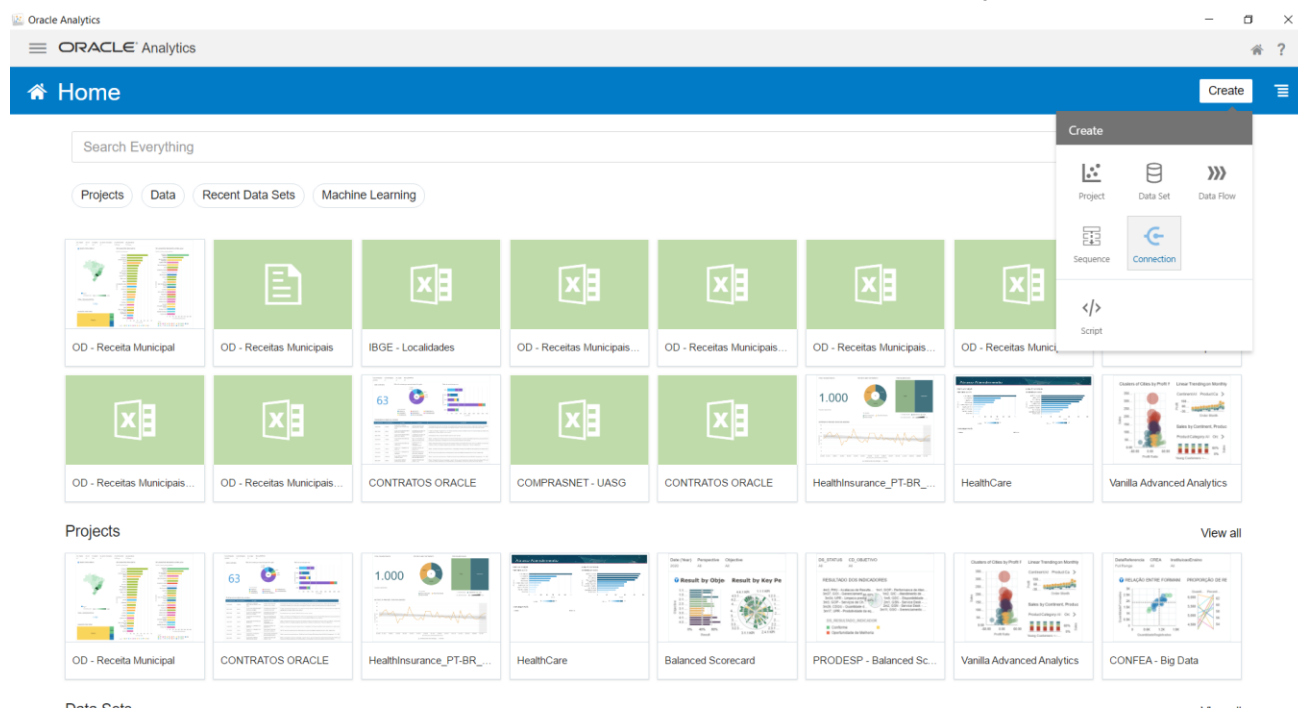
4. Em posse das credenciais. Abra o seu Oracle Analytics Desktop



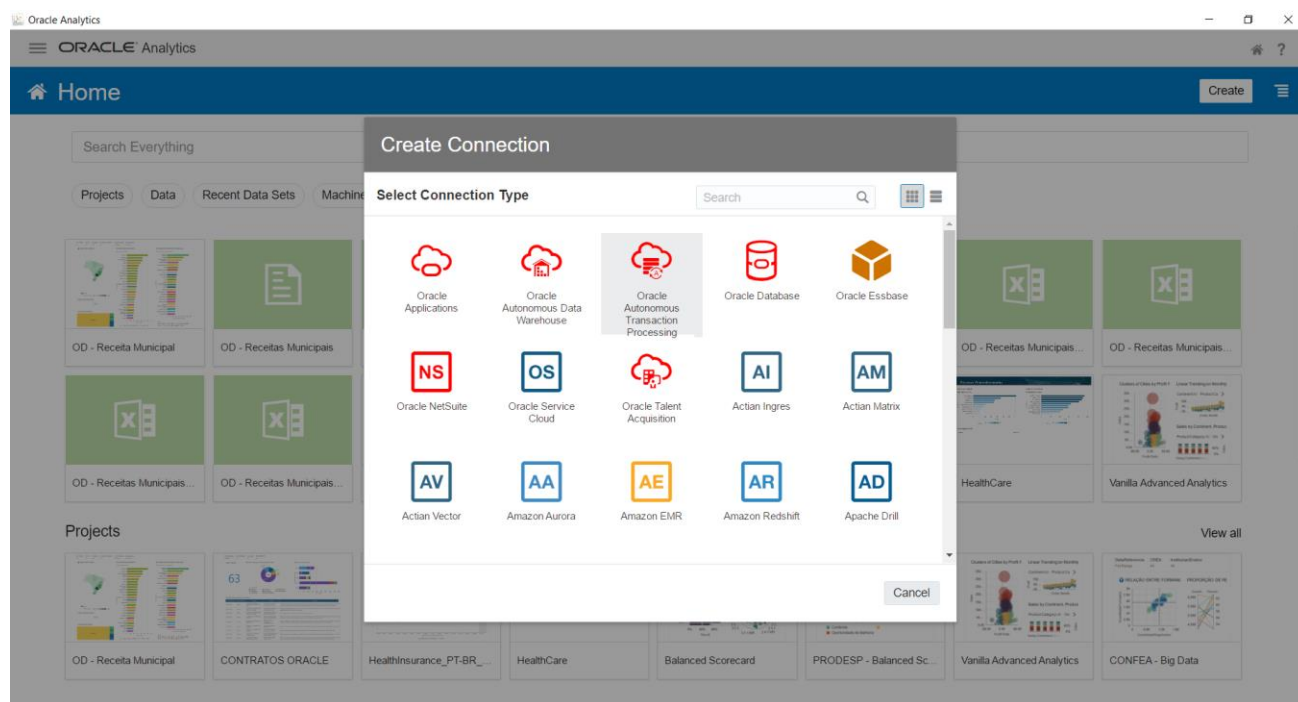
The screenshot shows the Oracle Analytics Desktop application interface. The top navigation bar includes tabs for "All", "Apps", "Documents", "Web", and "More". The main content area displays "Best match" results for "Oracle Analytics Desktop 12c App". A search bar on the left shows "Analytics" with suggestions like "Analytics - See web results", "analytics google", "analytics twitter", "analytics youtube", "analytics vidhya", "analytics uber eats", and "analytics brasil". The right panel shows the "Oracle Analytics Desktop 12c App" with an "Open" button and a list of recent files.



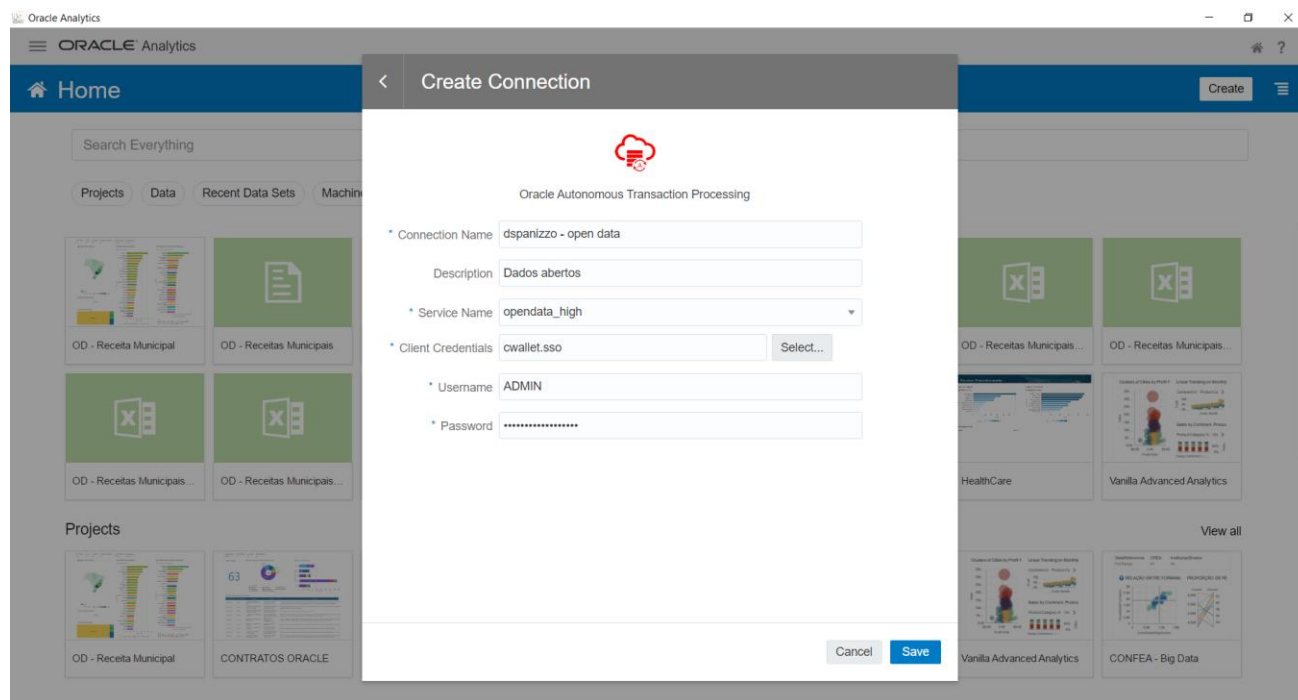
5. Na Home do Oracle Analytics, clique no botão “Create” e, logo em seguida, na opção “Connection”.



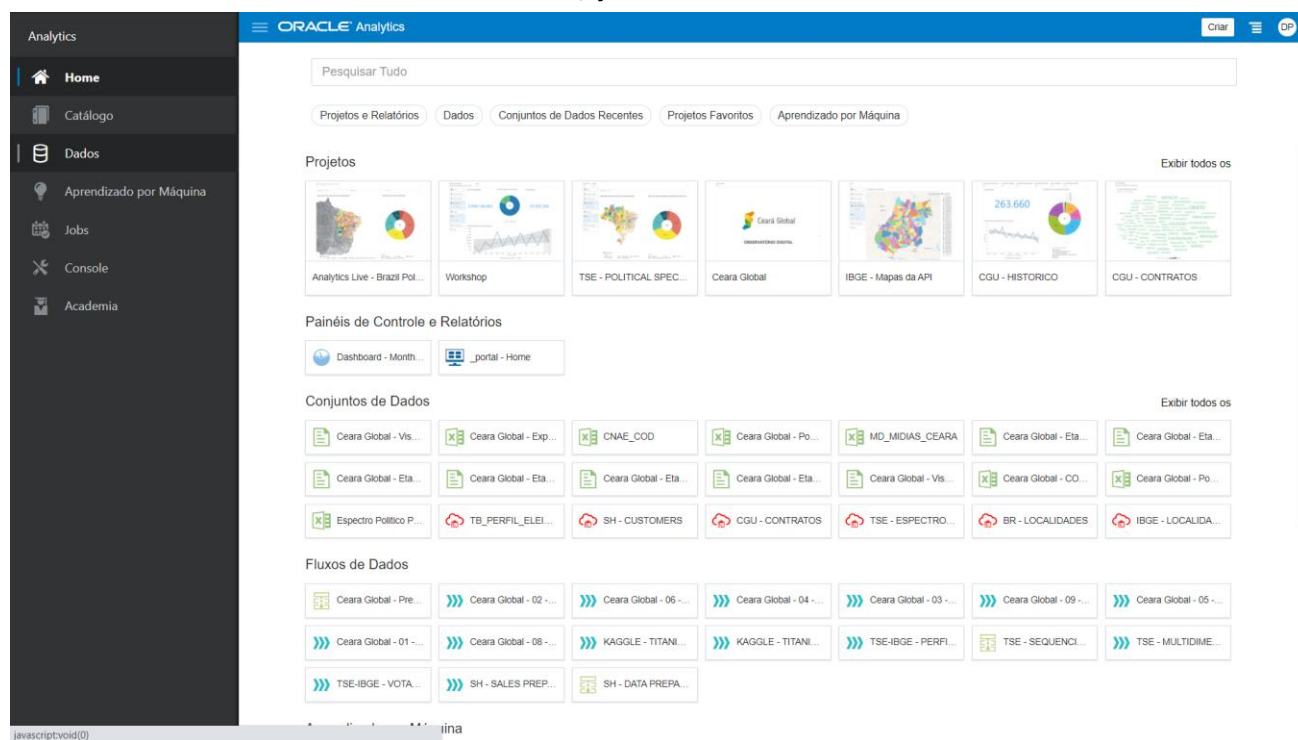
6. Na tela de criar conexão, selecione a opção “Autonomous Data Warehouse”.



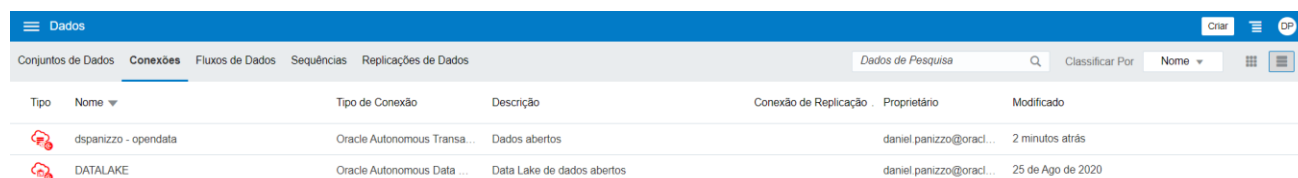
7. Na opção “Client Credentials”, informe o local onde armazenou o arquivo ZIP da Wallet do ADW. Em seguida, informe um nome para a conexão, usuário e senha do ADMIN. Clique em “Save”.



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



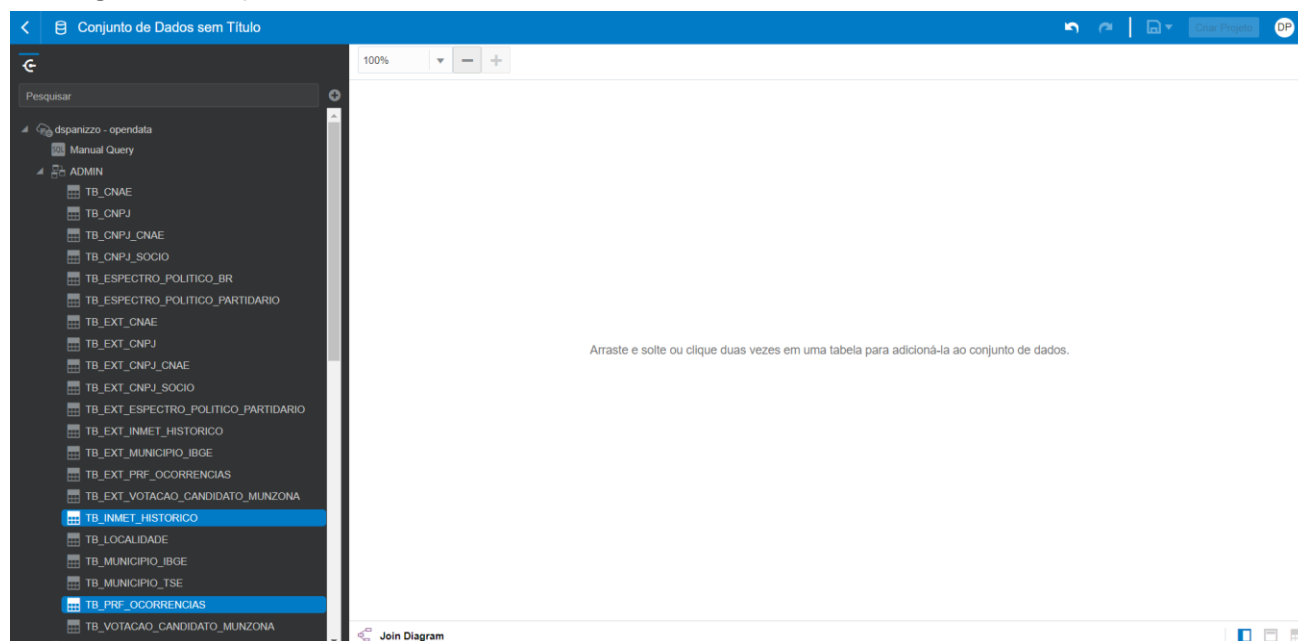
9. Na área de “Dados”, selecione a aba “Conexões”.



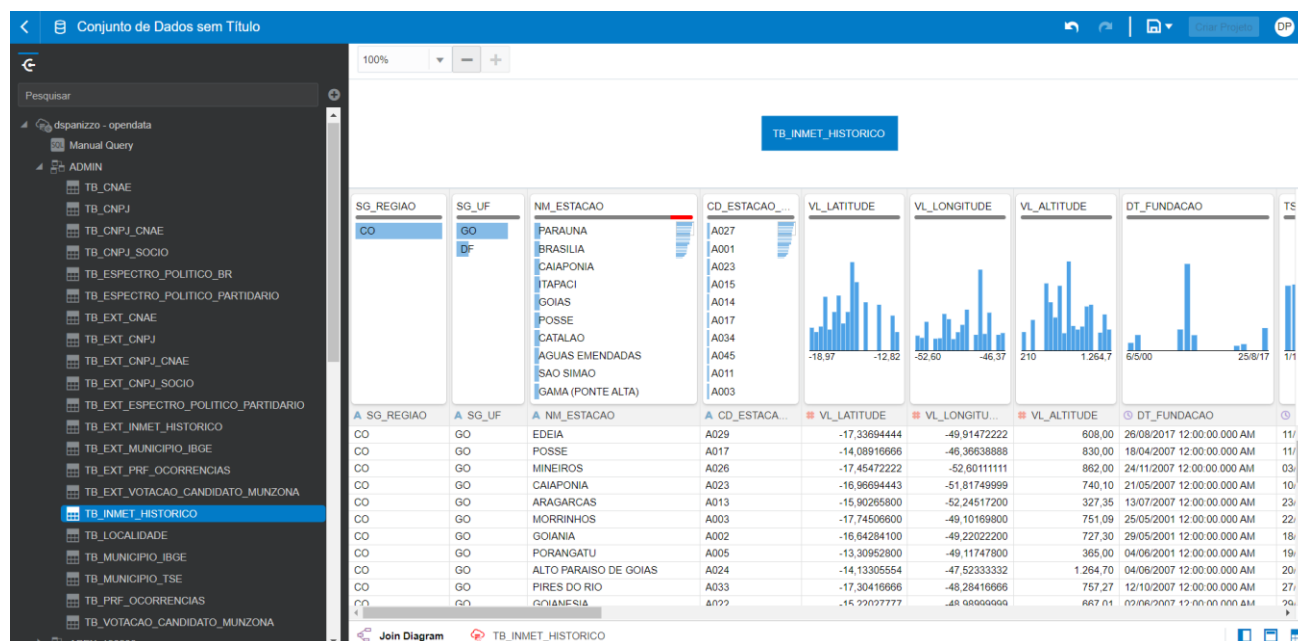
Dados						
Conjuntos de Dados Conexões Fluxos de Dados Sequências Replicações de Dados						
Dados de Pesquisa Q Classificar Por Nome						
Tipo	Nome	Tipo de Conexão	Descrição	Conexão de Replicação	Proprietário	Modificado
	dspanizzo - opendata	Oracle Autonomous Transa...	Dados abertos		daniel.panizzo@oraci...	2 minutos atrás
	DATA LAKE	Oracle Autonomous Data ...	Data Lake de dados abertos		daniel.panizzo@oraci...	25 de Ago de 2020

## Importando os metadados

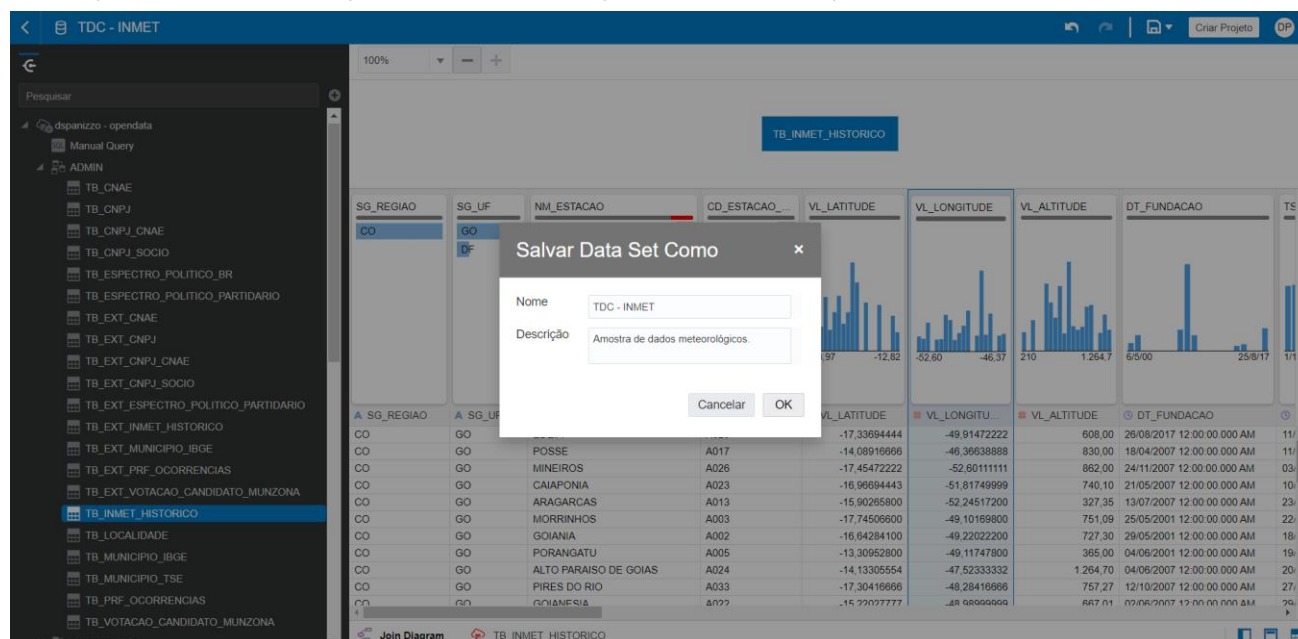
10. Ao selecionar a conexão que acabamos de criar, um painel irá aparecer com os esquemas do seu ADW. Dentro do esquema “ADMIN”, selecione a tabela “TB\_INMET\_HISTORICO” arraste-a para a área de diagrama ou clique duas vezes na tabela.



11. Uma amostra da tabela “TB\_INMET\_HISTORICO” será disponibilizada, informando valores e qualidade de dados em cada coluna.



12. Clique no ícone de salvar para definir um nome para o seu novo conjunto de dados.



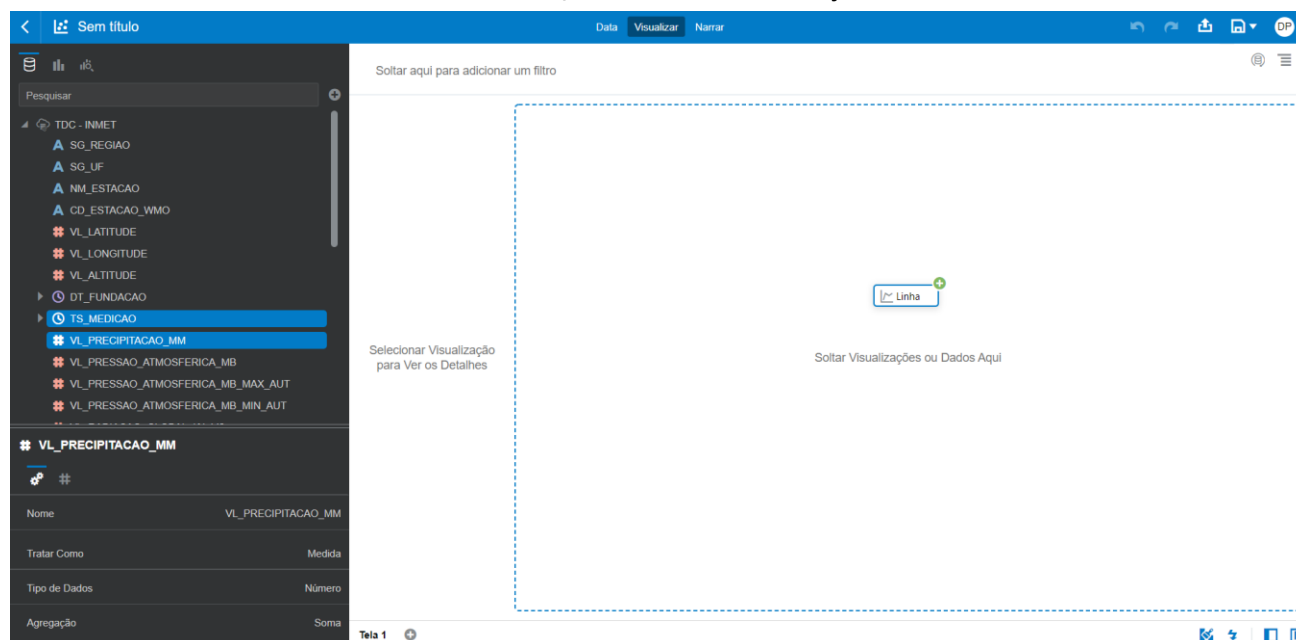
13. Com o conjunto de dados salvo, clique no botão “Criar Projeto” no canto superior direito.



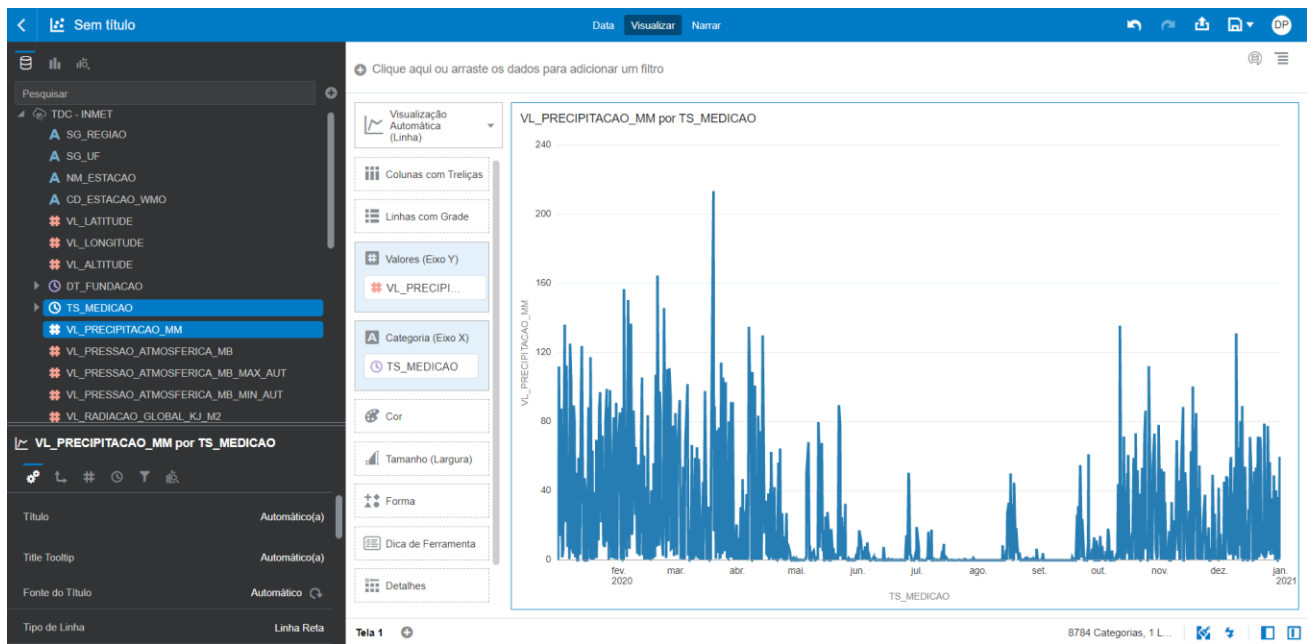
TB\_INMET\_HISTORICO

## Visualizando os dados

14. Na área de “Dados” (menu esquerdo), segurando CTRL, selecione as colunas TS\_MEDICAO e VL\_PRECIPITACAO\_MM. Arraste as colunas para a área de visualização de dados.



15. Agora já podemos explorar os detalhes meteorológicos longo do tempo. Agora a criatividade da exploração é com você!



16. Assim finalizamos nossa etapa de análise de dados.