

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS NÚCLEO DE
EDUCAÇÃO A DISTÂNCIA**

Pós-graduação *Lato Sensu* em Engenharia de Dados

Érika Kelly Domingues Nunes Rodrigues Andrade

**ANÁLISE DOS DADOS DE SEGURANÇA PÚBLICA DO MINISTÉRIO DA
JUSTIÇA NO BRASIL E DO INSTITUTO DE SEGURANÇA PÚBLICA DO RIO DE
JANEIRO**

Belo Horizonte

2024

Érika Kelly Domingues Nunes Rodrigues Andrade

**ANÁLISE DOS DADOS DE SEGURANÇA PÚBLICA DO MINISTÉRIO DA
JUSTIÇA NO BRASIL E COM ÊNFASE NO RIO DE JANEIRO**

Trabalho de Conclusão de Curso, apresentado
ao Curso de Especialização em Engenharia de
Dados como requisito parcial à obtenção do
título de especialista.

Belo Horizonte

2024

SUMÁRIO

1. Introdução	3
1.1. Contextualização.....	3
1.2. O problema proposto.....	5
1.3. Objetivos	5
2. Coleta de Dados	6
3. Ingestão/Processamento/Tratamento de Dados	7
4. Orquestração de Dados.....	11
5. Visualização de dados.....	18
6. Apresentação dos Resultados.....	19
7. Conclusão.....	20
8. Links.....	20

1. Introdução

1.1. Contextualização

A segurança pública é dever do Estado e é de extrema importância para sobrevivência e direito humano, é a garantia da proteção aos direitos individuais de cada cidadão à convivência social isenta de ameaça de violência, fazendo com que possam exercer seu direito de cidadania em segurança, como trabalhar, conviver em sociedade e se divertir.

Assegurar que o direito à segurança pública está sendo aplicado é o que alguns órgãos, tem fiscalizado e coletado dados para analisar e tomar medidas para ajudar a resolver problemas permanentes e evolutivos em determinados locais, como o Rio de Janeiro, Estado nacionalmente conhecido pelo seu índice de periculosidade e falta de segurança pública.

Em 2023, a gestão do MJSP buscou trabalhar junto com as 27 unidades da Federação na construção do Sistema Único de Segurança Pública (SUSP). O MJSP doou mais de R\$ 389 milhões em equipamentos, viaturas, armamentos e drones aos Estados, em um esforço de melhorar e ampliar o trabalho local das polícias militar e civil. Soma-se a esse esforço, a nova política de restrição de armas de fogo, que reduziu a circulação destas em território nacional. ¹

Para o ministro Flávio Dino, ainda há muito a fazer para que os números informados pelos Estados e demais órgãos tenham cada vez mais precisão e qualidade. *“É um processo essencial que nasceu da Lei 13.675. Números cada vez mais confiáveis irão gerar ações mais eficientes”*, afirmou.

¹ Por **Agência Gov**. Disponível em <<https://agenciagov.ebc.com.br/noticias/202401/balanco-2023-brasil-tem-menor-numero-de-assassinatos-dos-ultimos-14-anos#:~:text=O%20Brasil%20fechou%20o%20ano,de%20brasileiros%20e%20brasileiras%20salvas>>. Publicado em: 28 janeiro 2024.

Mas os números ainda são muito altos em relação a outros países, no G20 o Brasil fica para trás quando o assunto é segurança pública.² O estudo Panorama da competitividade dos países do G20 Brasil 2024, elaborado pela Federação das Indústrias do Estado do Rio de Janeiro (Firjan), coloca o país na 25ª posição em uma lista de 27 nações, superando apenas México e África do Sul.

A real aplicação de medidas de segurança pública alinhada às estatísticas é de extrema importância para que os números baixem a cada estudo, para isso além de capacitação e observância, é necessária a análise dos dados coletados e criação de medidas específicas para cada caso e região.

Como vimos, os dados são fontes essenciais para estudos e análises a fim de melhorar esse cenário e justificar investimentos no setor de segurança pública, a simples coleta de dados não gera informação, é necessário que esses dados recebam tratamento e análise exploratória, assim como criação de medidas e indicadores para que eles possam mostrar informações e gerar insights. Neste projeto vou mostrar algumas análises feitas com os dados coletados de segurança pública.

² Por **A Tribuna**. Disponível em <<https://www.tribunarij.com.br/materia/firjan-diz-que-seguranca-publica-e-o-principal-desafio-do-brasil-no-g20>>. Publicado em: 29 janeiro 2024.

1.2. O problema proposto

Neste projeto, serão utilizados dados de segurança pública nacional e do Rio de Janeiro, com a ingestão, orquestração e tratativa para que gere informações úteis para análise e aplicação de melhorias nas medidas de segurança locais. Daremos um foco especial na Cidade do Rio de Janeiro, com dados específicos do Estado e dentro disso também levantaremos estatísticas sobre feminicídios em suas localidades.

A importância dessa análise se dá ao fato de que medidas públicas, e até privadas, de segurança podem ser tomadas a partir das informações geradas.

Serão analisados, os dados de segurança pública do Ministério da Justiça e Segurança Pública (MJSP – Brasil) e Instituto de Segurança Pública (ISP – RJ).

O objetivo dessa análise é realizar algumas análises sobre os dados de segurança pública por Unidades Federativas, a fim de verificar os registros mais recorrentes em cada UF e por sexo das vítimas, esses dados são de 2015 a 2021, sobre os dados de segurança pública do Rio de Janeiro por Municípios, a fim de verificar os registros mais recorrentes em cada Município, esses dados são de 2003 a 2023 e sobre os dados de feminicídio, no Rio de Janeiro por Municípios, a fim de verificar a estatística desse tipo de ocorrência, esses dados são de 2016 a 2023.

1.3. Objetivos

Este estudo visa empregar técnicas de Engenharia de Dados para coletar informações e tratá-las de modo que gere ou agregue valor ao dado para futuras análises.

Com o objetivo de analisar os índices de criminalidade recorrentes em algumas regiões e mostrar o cenário da segurança pública no Rio de Janeiro, levantando enfaticamente também dados de feminicídio no Estado.

É fundamental salientar que essa abordagem também leva em conta a necessidade de um conhecimento específico e que tenha poderes para utilizar essa informação, uma vez que não cabe à população a aplicação de medidas.

2. Coleta de Dados

Os dados que serão analisados foram coletados do site do Ministério da Justiça de Segurança Pública (MJSP) e do Instituto de Segurança Pública, esses dados não tem relação entre si, são duas bases diferentes para o mesmo assunto, sendo que a primeira é uma análise abrangente por UF e a segunda é uma análise somente do Estado do Rio de Janeiro por Município e suas delegacias. Além disso utilizaremos dados do IBGE, para a padronização dos municípios e regiões.

Base de Dados de Segurança Pública do MJSP ¹ – Essa base de dados traz as informações de números de vítimas, femininas ou masculinas, por ocorrências, Estado, mês e ano, no âmbito nacional. 2015 a 2021

Base de Dados de Segurança Pública do ISP-RJ ² – Essa base de dados traz as informações de números de vítimas, por ocorrências, Circunscrições Integradas de Segurança Pública, Município, mês e ano, no Estado do Rio de Janeiro. 2003 a 2023

Base de Dados de Feminicídio do ISP-RJ ² – Essa base de dados traz as informações de números de vítimas por feminicídio, por Circunscrições Integradas de Segurança Pública, mês e ano, no Estado do Rio de Janeiro. 2016 a 2023

Dicionário de Dados do ISP-RJ ³ – Essa base traz o dicionário de dados para a criação da dimensão de tipos de ocorrências para os dados do ISP-RJ.

Base de Dados CISP do ISP-RJ ⁴ – Essa base traz a informação da localidade da CISP de ocorrência dos crimes.

Base de Dados de Municípios do IBGE ⁵ – Essa base traz a informação dos municípios e suas hierarquias para possíveis análises territoriais mais abrangentes.

¹ <https://www.gov.br/mj/pt-br/assuntos/sua-seguranca/seguranca-publica/estatistica/dados-nacionais-1/base-de-dados-e-notas-metodologicas-dos-gestores-estaduais-sinesp-vde-2022-e-2023>

² <https://www.ispdados.rj.gov.br/estatistica.html>

³ <https://www.ispdados.rj.gov.br/Conteudo.html>

⁴ <https://www.ispdados.rj.gov.br/Notas.html>

⁵ <https://www.ibge.gov.br/explica/codigos-dos-municipios.php>

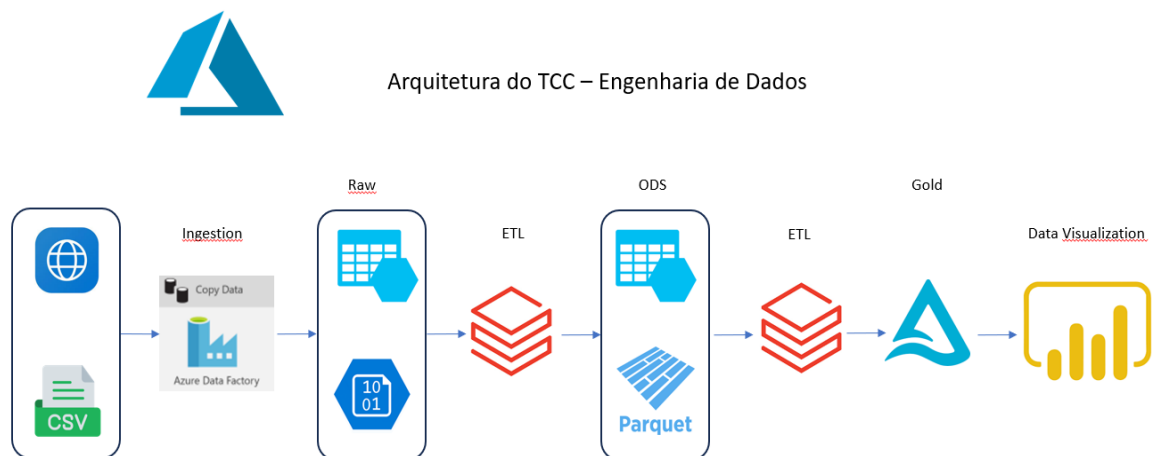
3. Ingestão/Processamento/Tratamento de Dados

Aqui apresento a arquitetura desse projeto, todas as ferramentas que foram utilizadas para o carregamento, transformação e processamento dos dados utilizados.

Optei por utilizar a arquitetura Azure, utilizando a conta de estudante pelo e-mail da minha matrícula na PUC Minas.

Criei toda a arquitetura do zero desde o resource group até os resources, keys, connections, integration runtimes, linked services, datasets, pipelines, notebooks, workspaces, dashboards, etc.

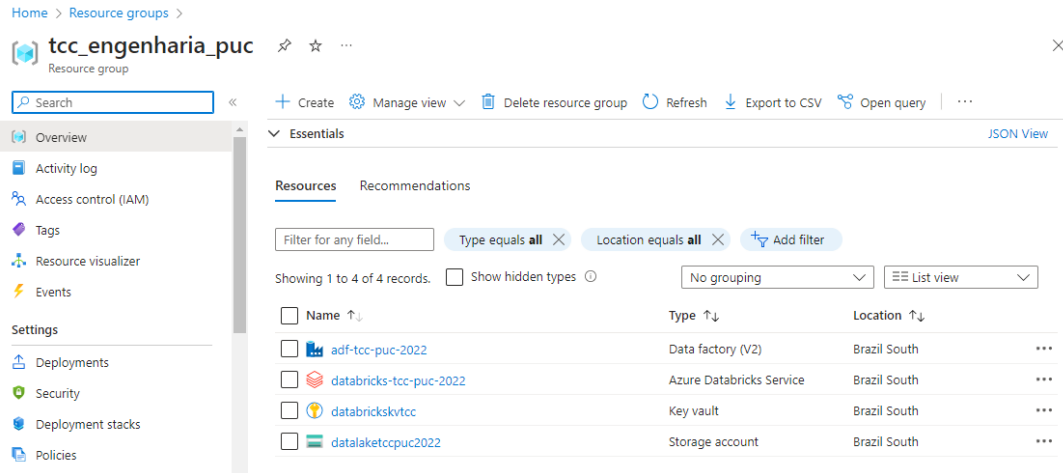
Figura 1: Screenshot da arquitetura do projeto



Fonte: Autor

É importante enfatizar que os dados apresentam uma configuração estruturada no formato "CSV". Adicionalmente, os dados já se encontram decodificados, fator que diminui a quantidade de etapas necessárias para sua manipulação. Finalmente, os conjuntos de dados contém informações numéricas e textuais

Figura 2: Screenshot dos recursos criados: Storage Account, Azure Data Factory, Databricks



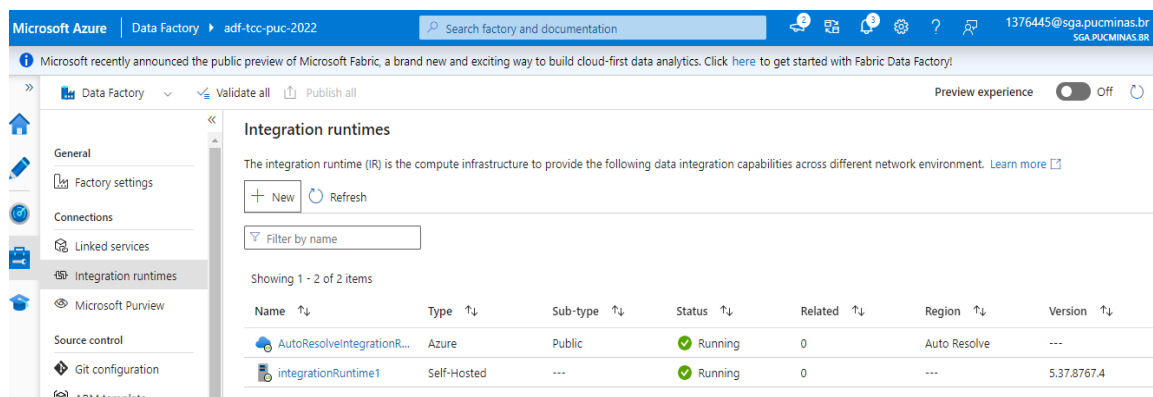
Fonte: Autor

Copiando os arquivos csv para uma Conta de Armazenamento do Azure (Storage Account)

- Instalação do Integration Runtime
- Configuração do Integration Runtime no Data Factory
- Criação dos Containers na Conta de Armazenamento
- Criação dos Linked Services
- Criação dos Datasets

Foi necessária a instalação do Integration Runtime na minha máquina para autorizar a transferência dos arquivos para o Azure

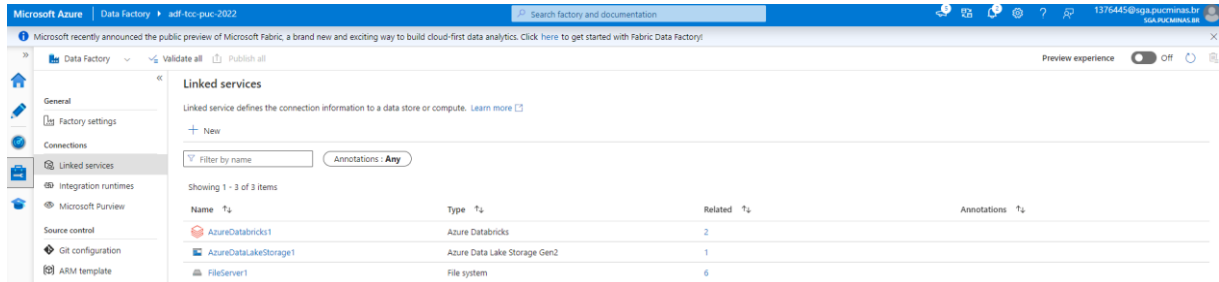
Figura 3: Screenshot dos Integrations Runtimes



Fonte: Autor

Para essa arquitetura também precisei criar os linked services para conectar os recursos necessários ao Data Factory (Databricks, File Server e Datalake)

Figura 4: Screenshot dos Linked Services



Fonte: Autor

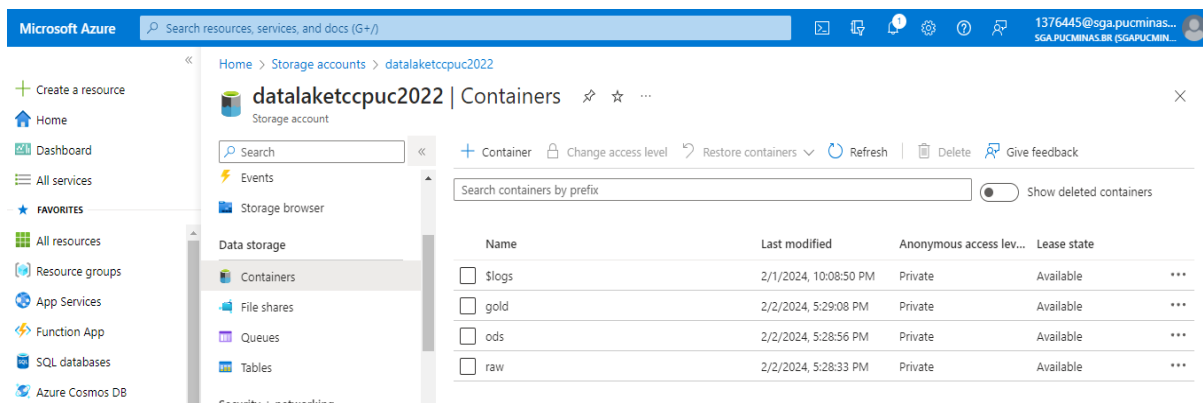
Também fiz a criação da Conta de Armazenamento no Azure, para salvar os arquivos em suas respectivas camadas:

Raw – Nessa camada os arquivos foram transferidos da forma que foram extraídos da fonte.

ODS – Nessa camada os arquivos foram salvos em parquet já com algumas limpezas efetuadas, como exclusão de vazios, substituição de caracteres especiais.

Gold – Nessa camada os arquivos foram salvos em parquet já com os tratamentos efetuados para camada de visualização, como transformação de dados, adição de colunas, criação de tabelas auxiliares ou dimensões, formatações de campos, etc.

Figura 5: Screenshot do Storage Account e seus Containers

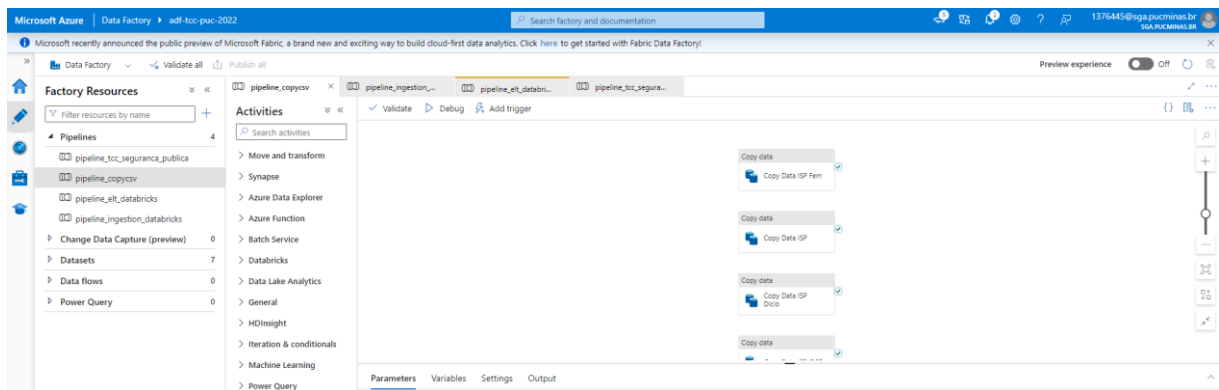


Fonte: Autor

- Criação da pipeline de cópia dos dados

No Data Factory foi criada uma pipeline para a cópia dos arquivos locais no datalake.

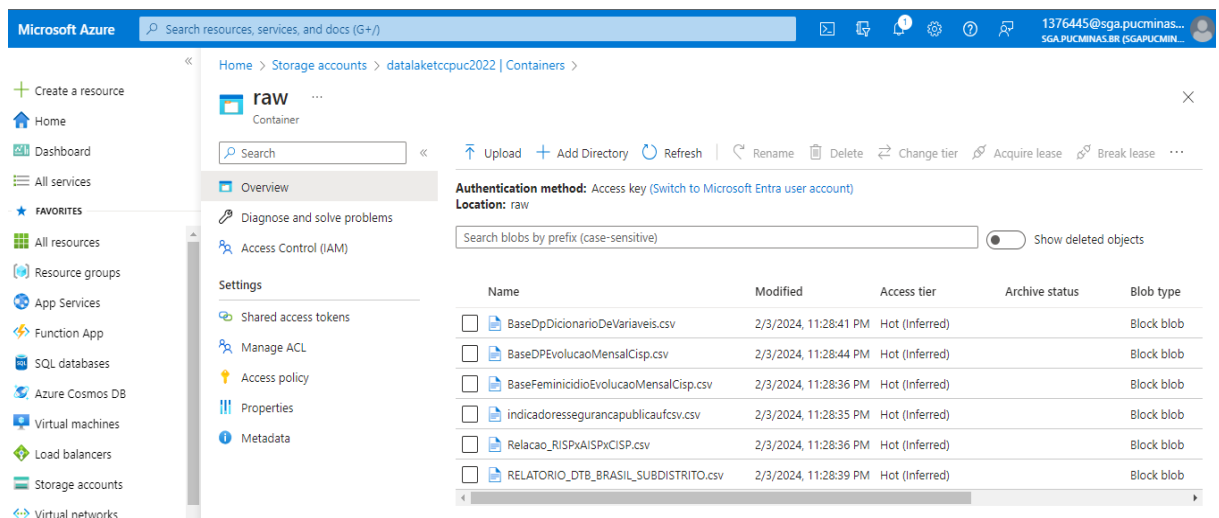
Figura 6: Screenshot da pipeline de copia dos dados



Fonte: Autor

Essa pipeline salva assim os arquivos csv no datalake na camada raw para serem tratados posteriormente:

Figura 7: Screenshot dos arquivos parquet salvos na camada raw



Fonte: Autor

4. Orquestração de Dados

Para a realização da orquestração de dados iremos utilizar o Databricks que irá coletar os dados, limpar, transformar e gravar os dados em delta tables e posteriormente o Data Factory para executar todo o processamento em sequência.

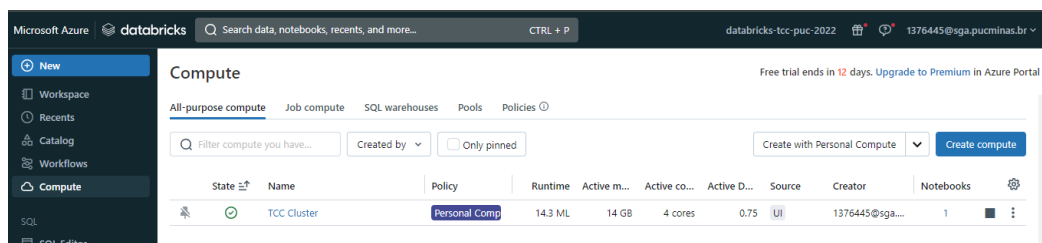
Coleta – Lendo os arquivos csv na camada raw da Conta de Armazenamento do Azure (Storage Account) com o Databricks para limpeza dos dados e criação dos arquivos parquet na camada ODS.

A transformação em arquivos parquet é feita para a otimização dos processamentos e consultas, além de ocupar menos espaço no datalake.

Para a configuração e utilização do Databricks:

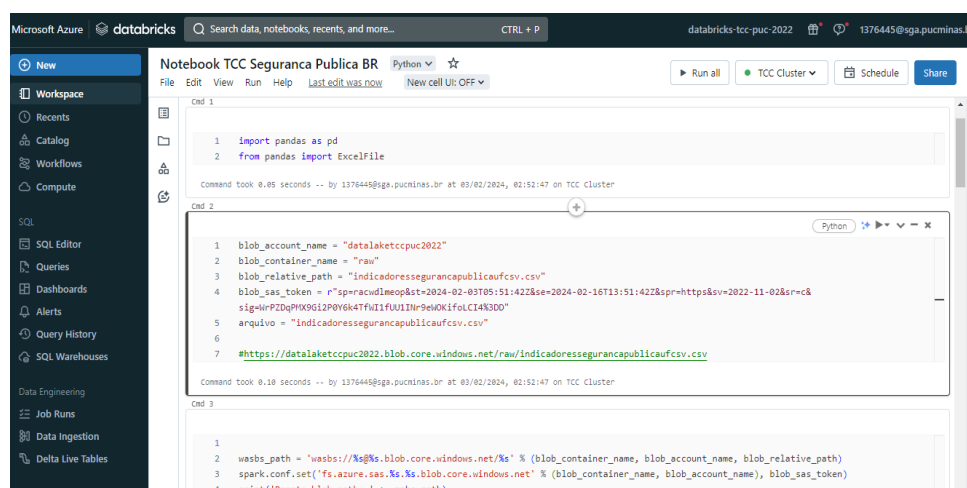
- Criação do Cluster
- Criação do Notebook

Figura 8: Screenshot dos Cluster criado para o processamento do Databricks



Fonte: Autor

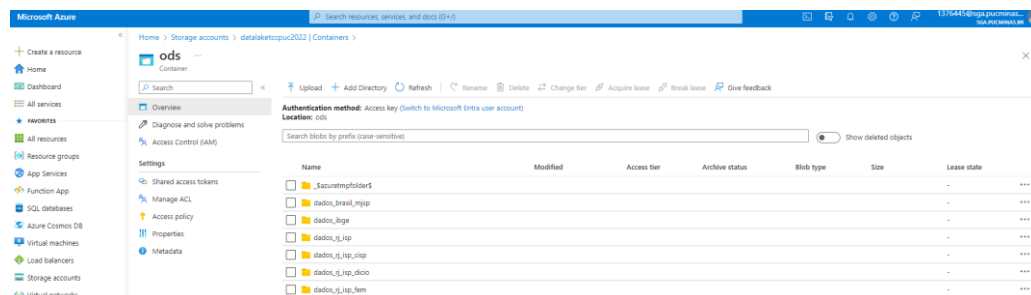
Figura 9: Screenshot do notebook de coleta de dados



Fonte: Autor

O notebook executa a coleta dos arquivos na camada raw, faz uma limpeza inicial dos dados e salva assim os arquivos parquet no datalake na camada ODS, por pastas dos conjuntos de dados para serem utilizados posteriormente na fase da geração das delta tables:

Figura 10: Screenshot das pastas dos arquivos parquet salvos na camada ods



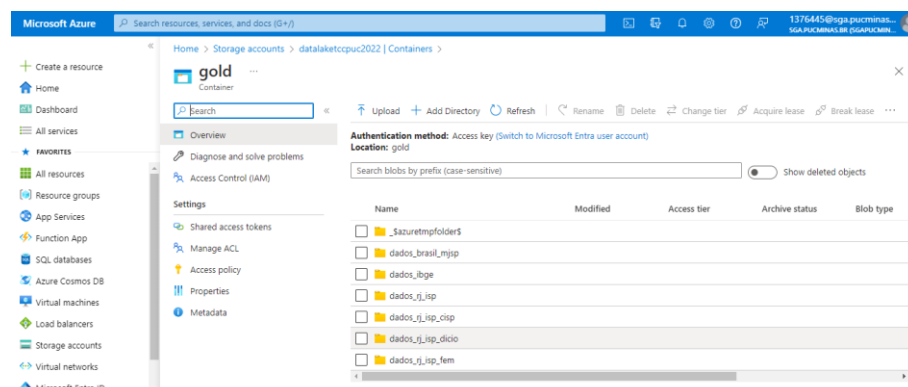
Fonte: Autor

Transformação – Lendo os arquivos parquet na camada ODS da Conta de Armazenamento do Azure (Storage Account) com o Databricks para transformação dos dados e criação dos arquivos parquet na camada gold.

Nessa etapa de transformação e limpeza de dados iremos usar o Python e SQL para fazer a limpeza e tratamento de dados para que possamos remover alguns dados que não são necessário para análise futura. Dentro do Databricks temos a possibilidade de usar diversas linguagem para tratamento de dados.

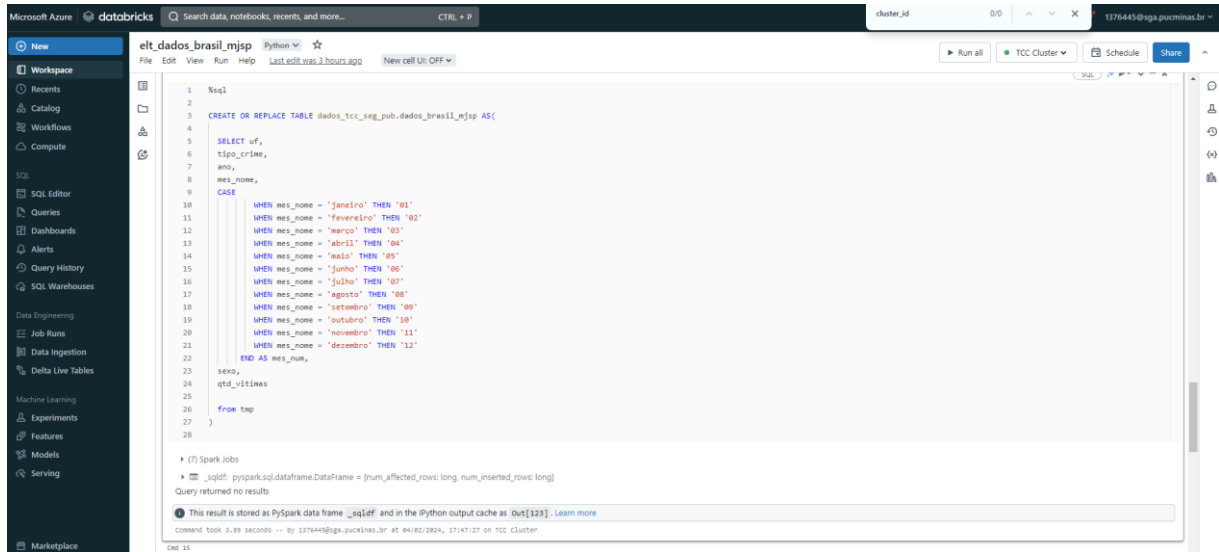
Nesta etapa também é dada a criação das delta tables que serão as tabelas finais a serem consumidas pela camada de visualização, que neste projeto será o Power BI.

Figura 10: Screenshot das pastas dos arquivos parquet salvos na camada ods



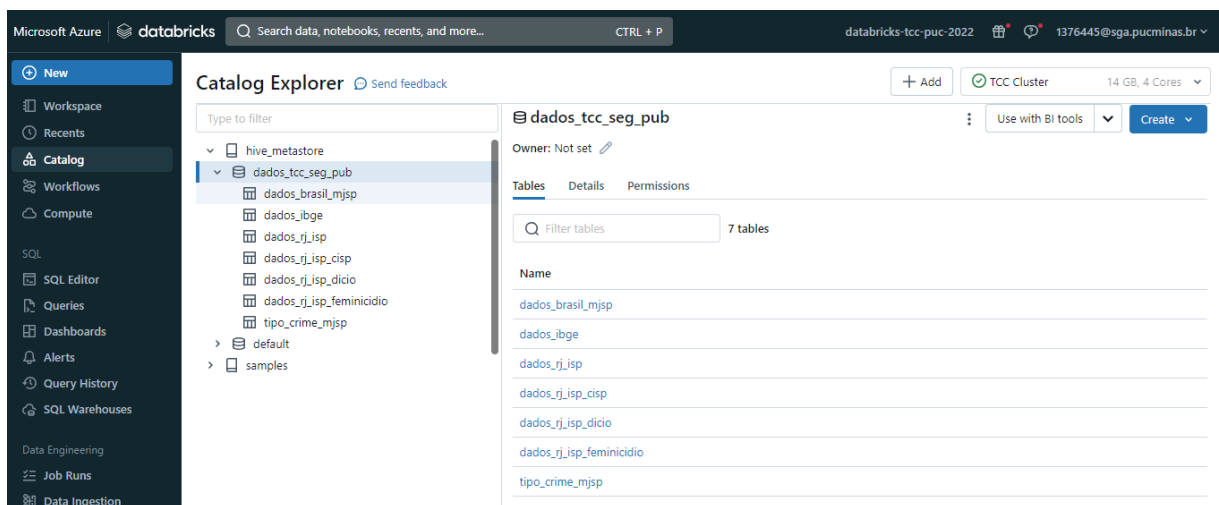
Fonte: Autor

Figura 11: Screenshot do notebook que cria as delta tables no Databricks para consumo



Fonte: Autor

Figura 12: Screenshot das delta tables criadas do Databricks para consumo

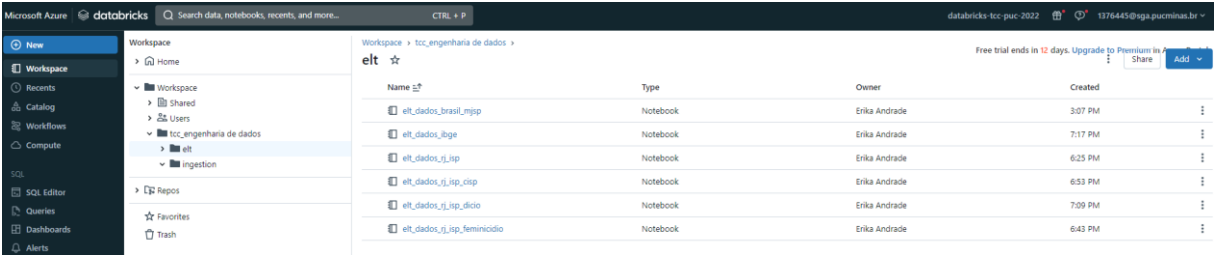


Fonte: Autor

Vale demonstrar a organização feita dentro do Databricks na Workspace.

Criei uma pasta para o projeto e pastas diferentes para os processos de ingestão e transformação. Além disso, criei o database do projeto onde ficam armazenadas as deltas como podem ver na imagem acima.

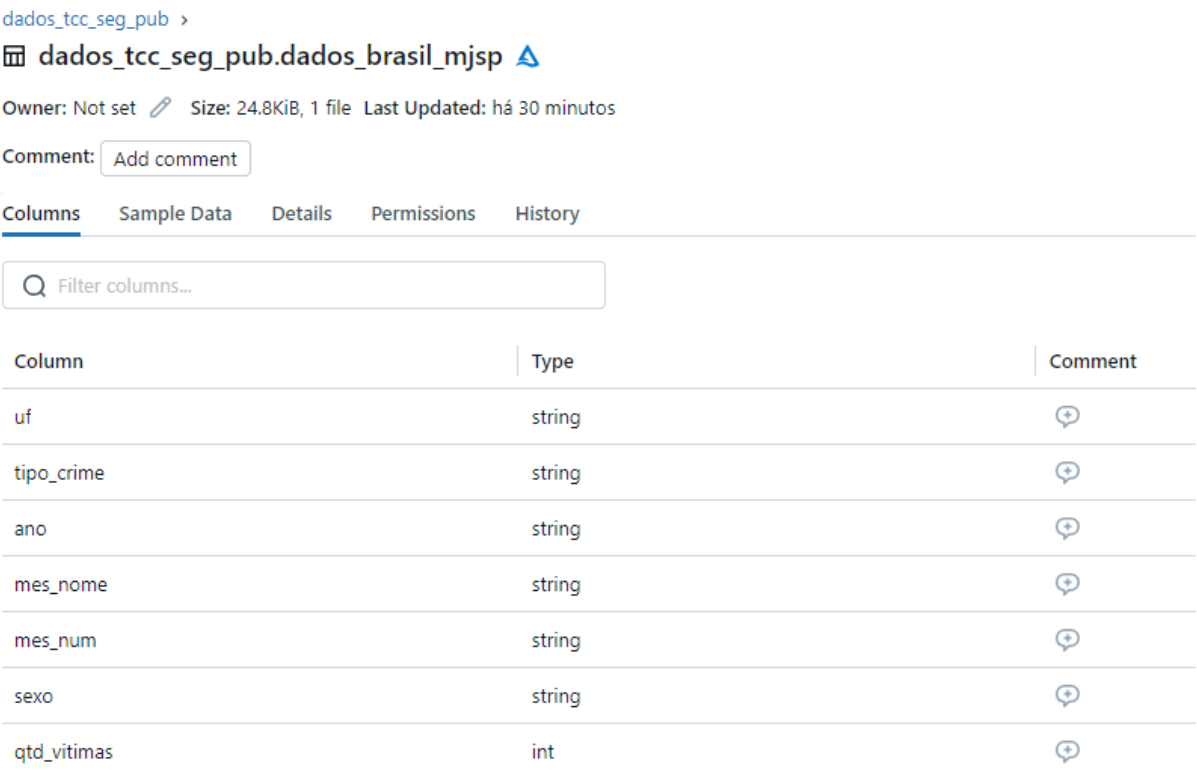
Figura 13: Screenshot do Workspace e sua organização no projeto



Fonte: Autor

Abaixo detalho as tabelas que serão utilizadas na camada de visualização.

Figura 14: Screenshot da estrutura da tabela de dados de crimes ocorridos no Brasil



Fonte: Autor

Figura 15: Screenshot da estrutura da tabela de dados de crimes ocorridos no Rio de Janeiro

dados_tcc_seg_pub >

dados_tcc_seg_pub.dados_rj_isp

Use with BI tools

Create

Owner: Not set

Size: 1.2MiB, 2 files

Last Updated: há 31 minutos

Comment:

Add comment

Columns

Sample Data

Details

Permissions

History

Q Filter columns...

Column	Type	Comment
cisp	string	
mes	string	
ano	string	
mes_ano	string	
aisp	string	
risp	string	
munic	string	
mcirc	string	
regiao	string	
hom_doloso	string	
lesao_corp_morte	string	
latrocinio	string	
cvli	string	
hom_por_interv_policial	string	
letalidade_violenta	string	

Fonte: Autor

Figura 16: Screenshot da estrutura da tabela de dados de feminicídios ocorridos no Rio de Janeiro

dados_tcc_seg_pub >

dados_tcc_seg_pub.dados_rj_isp_feminicidio

Use with BI tools

Create

Owner: Not set

Size: 7.5KiB, 1 file

Last Updated: há 32 minutos

Comment:

Add comment

Columns

Sample Data

Details

Permissions

History

Q Filter columns...

Column	Type	Comment
cisp	string	
ano	string	
mes	string	
feminicidio	string	
feminicidio_tentativa	string	
fase	string	

Fonte: Autor

Figura 17: Screenshot da estrutura da tabela complementar de locais de ocorrencias

dados_tcc_seg_pub >

dados_tcc_seg_pub.dados_rj_isp_cisp

Use with BI tools

Create

Owner: Not set

Size: 9.9KiB, 1 file

Last Updated: há 33 minutos

Comment:

Add comment

Columns

Sample Data

Details

Permissions

History

Filter columns...

Column	Type	Comment
risp	string	
aisp	string	
cisp	string	
unidade_territorial	string	
municipio	string	
regiao_governo	string	

Fonte: Autor

Figura 18: Screenshot da estrutura da tabela dicionario de dados de ocorrencias

dados_tcc_seg_pub >

dados_tcc_seg_pub.dados_rj_isp_dicio

Use with BI tools

Create

Owner: Not set

Size: 5.5KiB, 1 file

Last Updated: há 34 minutos

Comment:

Add comment

Columns

Sample Data

Details

Permissions

History

Filter columns...

Column	Type	Comment
variavel	string	
descricao_variavel	string	
contagem	string	
observacao	string	
grupo	string	

Fonte: Autor

Figura 19: Screenshot da estrutura da tabela de municípios do IBGE

dados_tcc_seg_pub >

dados_tcc_seg_pub.dados_ibge

Use with BI tools

Create

Owner: Not set

Size: 13.2KiB, 1 file

Last Updated: há 35 minutos

Comment:

Add comment

Columns

Sample Data

Details

Permissions

History

Filter columns...

Column	Type	Comment
uf	string	
nome_uf	string	
nome_microrregiao	string	
municipio	string	

Fonte: Autor

PADRÃO ABNT

FAZER ULTIMAS TRANSFORMACOES NAS TABELAS PARA MODELAR

CRIAR POWER BI COM CONEXÃO DATABRICKS

FAZER MODELAGEM DOS DADOS NO POWER BI

CRIAR DASHBOARDS

SALVAR IMAGENS

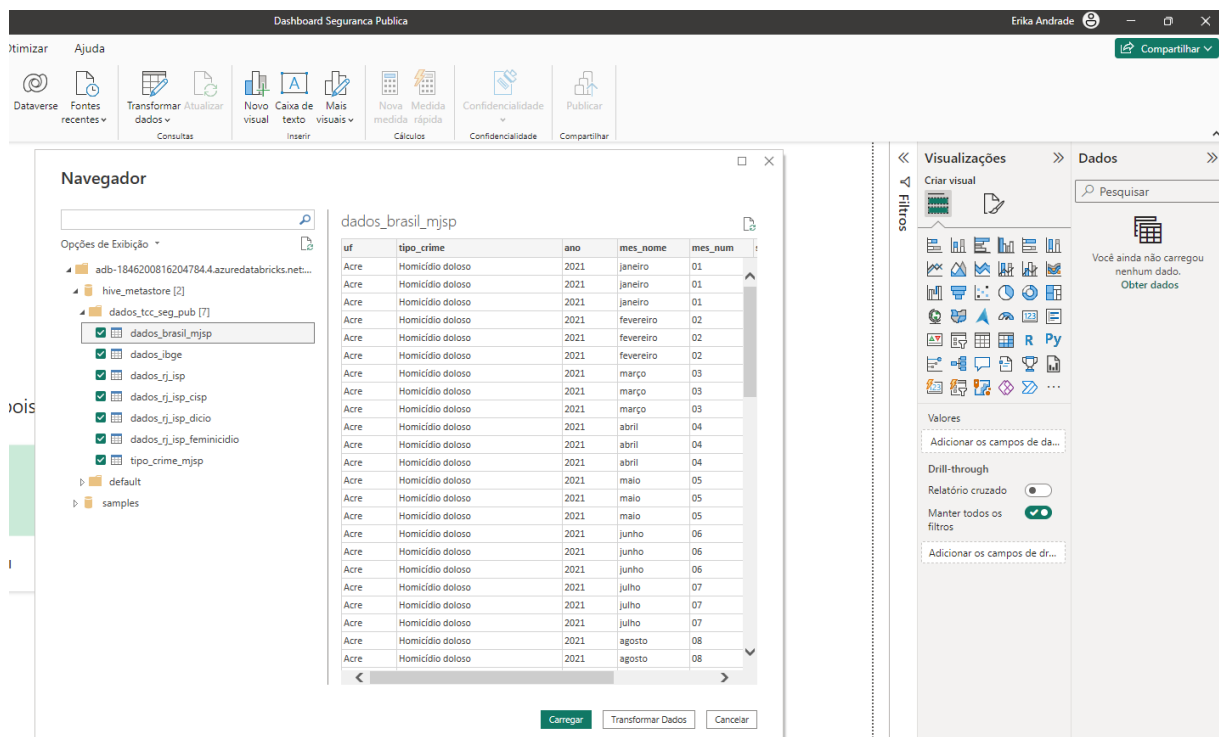
SALVAR ARQUIVOS NO GIT

FAZER VIDEO

5. Visualização de dados

Após a criação das tabelas vamos criar a visualização dos dados, para isso iremos utilizar a ferramenta Power BI Desktop, também com o uso da licença de estudante da PUC, obtendo dados diretamente do Databricks por Direct Query

Figura 20: Screenshot da conexão do Power BI com o Databricks



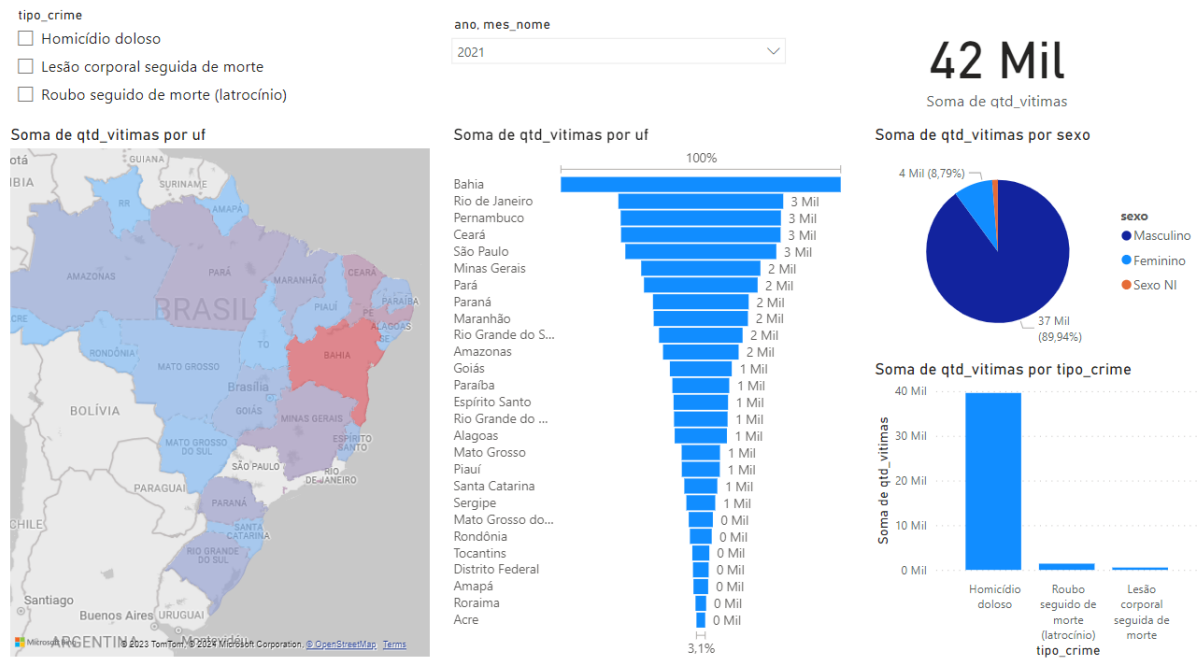
Fonte: Autor

Serão importados os dados que foram gravados no Databricks nas Delta Tables como mostrado anteriormente.

Nesse Dashboard foram criados alguns filtros por crime e ano. Podemos visualizar no mapa o Estado com mais incidência criminal, assim como no funil ao lado.

No visual de pizza podemos segmentar a visualização por sexo e abaixo a relação comparativa por tipo de crime

Figura 21: Screenshot da conexão do Power BI com o Databricks



Fonte: Autor

6. Apresentação dos Resultados

Os resultados evidenciados neste estudo corroboram importantes discussões no que se refere à criação de estratégias e condições apropriadas, para a melhoria da segurança pública baseada em dados históricos.

Esta reflexão deve perpassar pelos profissionais da força do Estado, gestores públicos e sociedade civil organizada. Nesse sentido, medidas de segurança pública aplicadas às localidades analisadas se fazem eficientes quando são efetivas por historicidade.

7. Conclusão

Considerando o desafio de criar um processo de engenharia de dados, com todo seu mecanismo funcional, de acordo com o objetivo proposto e as ferramentas utilizadas, o estudo foi concluído com êxito demonstrando a realidade das informações em relação aos dados coletados de segurança pública.

Os visuais permitiram a visualização dos índices criminais no Brasil e no Rio de Janeiro.

Portanto, concluímos que este estudo atingiu seu objetivo, ilustrando o potencial das ferramentas de engenharia de dados e na assistência à tomada de decisão por parte dos gestores de segurança pública.

8. Links

Link para o vídeo:

<https://www.youtube.com/watch?v=yCv5CajRCLg>

Github

[erikakelly/projeto-edd-seguranca-publica \(github.com\)](https://github.com/erikakelly/projeto-edd-seguranca-publica)

9. Referencias

PYTHON. Disponível em: <<https://www.python.org/>>.

Data Factory. Disponível em: <<https://learn.microsoft.com/en-us/training/modules/intro-to-azure-data-factory/>>.

Databricks Disponível em: <<https://www.databricks.com/learn>>.

Power BI Disponível em: <<https://learn.microsoft.com/pt-br/training/powerplatform/power-bi>>.

Azure Disponível em: <<https://learn.microsoft.com/en-us/training/azure/>>.

10. Apêndice

Toda documentação, assim como os códigos, estão no github.