

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO
PUC-RJ



Autor: Alexandre José Sá Barreto de Andrade

MVP – Ciência de dados e Analytics – Sprint Engenharia de Dados

PÓS GRADUAÇÃO - CIÊNCIA DE DADOS E ANALYTICS

Sumário

Objetivo do trabalho:	3
Questões serem respondidas:	3
Fontes de dados:	4
Formato dos dados:	4
Arquitetura do projeto:	5
Criação do banco de dados e tabelas no Data Briks:	5
Ativação do ambiente de processamento (SQL Warehouse Serverless)	5
Criação do banco de dados do projeto	5
Upload dos arquivos CSV para o Catalog	5
Criação das tabelas no Databricks a partir dos arquivos CSV	6
ETL (Extrair, Transformar e Carregar)	7
- Padronização dos nomes das colunas	7
- Padronização dos tipos de dados	9
- Criação das tabelas Dimensão:	10
- Criação das tabelas Fato:	15
Registro dos metadados das tabelas	17
Conexão e carga das tabelas no Power Bi Fabric	22
Modelagem de dados - Criação do modelo estrela na plataforma de conexões nativo do Power bi	25
- Tabelas Dimensão:	25
- Tabelas Fato:	26
- Modelo de dados estrela das tabelas relacionadas a PRODUÇÃO:	26
- Modelo de dados estrela das tabelas relacionadas a VENDAS:	27
- Modelo de dados estrela das tabelas relacionadas a COLETA:	28
- Modelo de dados estrela das tabelas relacionadas a RERREFINADOR:	29
Criação dos relatórios de respostas implementados no Power bi.	31
- Relatório de respostas de PRODUÇÃO	31
- Relatório de respostas de VENDAS	32
- Relatório de respostas de COLETAS	33
- Relatório de respostas de RERREFINO	34
AUTO AVALIAÇÃO	35
ENTREGA DO PROJETO – Abaixo os links para acesso aos arquivos que foram gerados para atender ao projeto	35

Objetivo do trabalho:

Esse trabalho construiu um pipeline de dados utilizando tecnologias na nuvem. O pipeline envolveu a busca, coleta, modelagem, carga e análise dos dados e as respostas para as questões levantadas foram entregues através de relatórios criados em PowerBI.

O objetivo principal foi obter um panorama de dados amplo cobrindo todo Brasil com relação ao mercado de óleos lubrificantes. Iremos observar e apurar o volume de óleos lubrificantes acabados (OLAC) vendidos no Brasil e também os volumes coletados e rerrefinados e também a reutilização dos óleos lubrificantes já utilizados OLUC.

A ANP - agência nacional do petróleo possui uma legislação que exige que as empresas produtoras de óleo lubrificante acabado (OLAC) tenham responsabilidade sobre o óleo já utilizado, mantendo uma economia circular.

Desta forma as empresas fabricantes de óleo são responsáveis pela coleta do óleo já utilizado ou contaminado (OLUC) levando em consideração o volume de vendas de cada empresa nas diferentes regiões do Brasil. Esse volume é entregue para empresas de rerrefino com o objetivo de que sejam filtrados, tratados e limpos, para serem reutilizados como insumo de fabricação do óleo novo. Neste processo de coleta existe um certificado que a empresa coletora emite para as empresas originadoras para comprovar a aderência a legislação.

Assim é garantido o processo de reutilização reduzindo impactos na natureza caso esse produto fosse descartado de maneira indevida.

Questões serem respondidas:

1. Perguntas sobre a **produção** de Óleo lubrificante (OLAC) no Brasil em m³.
 - a. Qual é o volume total de produção no Brasil?
 - b. Qual é o volume de produção segmentado por região e estado?
 - c. Quais são os 10 estados que mais produzem no Brasil por região?
 - d. Quais são os produtos óleo produzido no Brasil e o volume de produção?
 - e. Visualização de volume de produção por agrupamentos
2. Perguntas sobre a **vendas** de Óleo lubrificante (OLAC) no Brasil
 - a. Qual é o volume total de vendas no Brasil?
 - b. Qual é o volume de vendas segmentado por região e estado?
 - c. Quais são os 10 estados que mais vendem no Brasil?
 - d. Quais são os produtos óleo mais vendidos no Brasil e o volume de vendas?
 - e. Visualização de volume de vendas por agrupamentos
3. Perguntas sobre **coleta** de Óleo lubrificante (OLUC) no Brasil
 - a. Qual é o volume total de coleta no Brasil?
 - b. Qual é o volume de coleta segmentado por região e estado?
 - c. Quais são os 10 estados que mais coletam no Brasil?
 - d. Quais são as empresas certificadas para fazer a coleta e o seu volume de coleta?

- e. Visualização geográfica das regiões e cidades que possuem coleta
4. Perguntas sobre a **rerrefino** de Oleo lubrificante (OLUC no Brasil)
- a. Qual é o volume total de rerrefino no Brasil?
 - b. Qual é o volume de rerrefino segmentado por região e estado?
 - c. Quais são os 10 estados que mais fazem rerrefino no Brasil?
 - d. Quais são os produtos que utilizam o oleo de refino e volume utilizado
 - e. Visualização geográfica das regiões e cidades que possuem rerrefino

Fontes de dados:

A **Agência Nacional do Petróleo, Gás Natural e Biocombustíveis** disponibiliza por meio do portal **Gov.br – Ministério de Minas e Energia dados** dados numéricos e geográficos sobre o Panorama do Mercado Brasileiro de Lubrificantes.

Disponibiliza informações atualizadas sobre este mercado, com dados retirados do sistema SIMP, sistema para monitorar dados de produção e movimentação de produtos do setor de petróleo, gás e biocombustíveis, alimentado pelas empresas reguladas pela ANP, permitindo ao usuário baixar e interagir com os dados.

Os dados são uma ferramenta de análise destinada a empresas, órgãos de governo, universidades, imprensa e à sociedade como um todo, além de estar em sintonia com a política de transparência amplamente adotada pela ANP e com o papel desempenhado pela Agência na logística reversa de óleo lubrificante usado, nos termos da Resolução Conama 362/2005.

Formato dos dados:

É disponibilizado pelo portal **Agência Nacional do Petróleo, Gás Natural e Biocombustíveis** arquivos no formato .CVS (Arquivos de valores separados por virgulas do Microsoft Excel). Contendo as informações necessárias para as análises.

Sendo cada arquivo com nome original e com seus conteúdos:

Produção nacional de OLAC: (Ano, Mês, Código do Produto, Descrição do Produto, Região do Produtor, UF do Produtor, Código da Operação, Descrição da Operação, Volume(L))

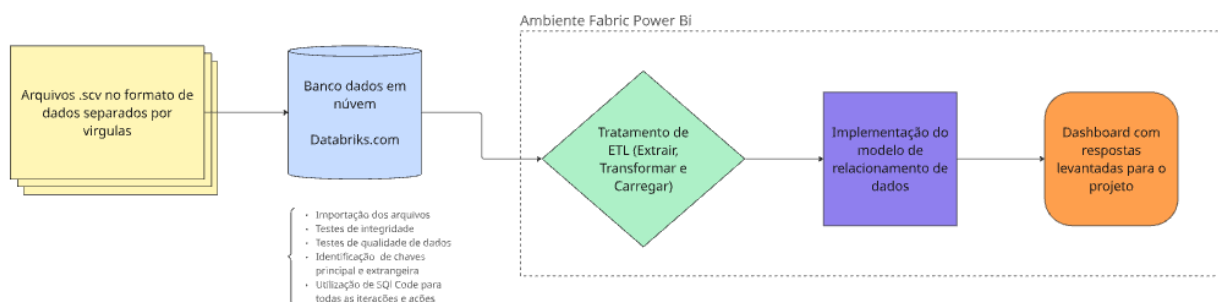
Vendas nacionais de OLAC pelos Produtores e Importadores (Ano, Mês, Código do Regulado – ANP, Agente Regulado, Código do Produto, Descrição do Produto, Região do Destinatário, UF do Destinatário, Volume(L))

Vendas nacionais de OLAC dispensadas de coleta: (Ano, Mês, Código do Regulado – ANP, Agente Regulado, Código do Produto, Descrição do Produto, Região do Destinatário, UF do Destinatário, Volume(L))

Volume de OLUC coletado por Coletor: (Ano, Mês, Código do Regulado – ANP, Agente Regulado, Região da Coleta, UF da Coleta, Volume Coletado(L))

Operações por Rerrefinador: (Ano, Mês, Código do Regulado – ANP, Agente Regulado, Código da Operação, Descrição da Operação, Código do Produto, Descrição do Produto, Região do Destinatário, UF do Destinatário, Volume(L))

Arquitetura do projeto:



Criação do banco de dados e tabelas no Data Briks:

Ativação do ambiente de processamento (SQL Warehouse Serverless)

O warehouse padrão chamado **Serverless Starter Warehouse** foi ativado e configurado para iniciarmos a execução dos dados do projeto.

Criação do banco de dados do projeto

Com o SQL Warehouse ativo, utilizei o editor SQL do Databricks para criar o banco de dados que armazenará todas as tabelas importadas dos anexos.

Código SQL utilizado para realizar a tarefa:




```
[CREATE DATABASE IF NOT EXISTS projeto_MVP_Engenharia_de_Dados;]
```

Upload dos arquivos CSV para o Catalog

Todos os arquivos disponibilizados pela ANP foram enviados para o volume criado dentro do catálogo workspace e do **projeto_mvp_engenharia_de_dados**.


O repositório central de todos os dados brutos do projeto. Esse procedimento garante organização, rastreabilidade e padronização, preparando o ambiente para a criação das tabelas e para as próximas etapas do projeto.


Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados >






 **arquivos_dados_cvs**   Share Upload to this volume

Overview Files Details Permissions


Description

 AI generate Add

/Volumes/workspace/projeto_mvp_engenharia_de_dados/arquivos_dados_cvs  Create directory

Name	Size	Last modified
 anexo_c_producao_nacional.csv	591.03 KB	8 days ago
 anexo_d_vendas_nacionais_prod.csv	17.52 MB	8 days ago
 anexo_e_vendas_nacionais_dispenza.csv	6.06 MB	7 days ago
 anexo_g_volume_coletado_coletor.csv	1.43 MB	8 days ago
 anexo_i_operacoes_rerrefinador.csv	6.16 MB	7 days ago

About this volume

Owner alexandre.andrade@gmail.com 

Criação das tabelas no Databricks a partir dos arquivos CSV

Cada arquivo foi convertido em uma tabela dentro do ambiente Databricks, diretamente através do assistente *Create table from volumes*.

O Databricks detectou automaticamente os tipos de dados e criou as tabelas, tornando-as disponíveis para consulta SQL e para as próximas etapas de tratamento e modelagem.








As tabelas criadas foram:

- [anexo_c_producao_nacional]
- [anexo_d_vendas_nacionais_prod]
- [anexo_e_vendas_nacionais_dispenza]
- [anexo_g_volume_coletado_coletor]
- [anexo_i_operacoes_rerrefinador]


Essa etapa consolidou no Lakehouse a camada de dados brutos, servindo como base para as fases seguintes do projeto.


Tela dentro do Databricks como exemplo da criação de umas das tabelas:


Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados >

 **anexo_c_producao_nacional**    Open in a dashboard  Share  Create 

Overview Sample Data Details Permissions Policies History Lineage Insights Quality

Description 
Created by the file upload UI

 AI generate

Column	Type	Comment	Tags	Column masking 
ANO	int			
MES	int			
CODIGO_PRODUTO	string			
DESCRICAO_PRODUTO	string			
REGIAO_PRODUTOR	string			
UF_PRODUTOR	string			
CODIGO_OPERACAO	string			
DESCRICAO_OPERACAO	string			
VOLUME_L	double			

ETL (Extrair, Transformar e Carregar)

- Padronização dos nomes das colunas

- Foi realizada a padronização dos nomes das colunas para o formato UPPER CASE em todas as tabelas, com uso de underline para separar palavras. Essa padronização mantém um padrão único em todo o projeto.

Os códigos SQL utilizados para todas as tabelas para fazer os ajustes seguem abaixo:

- Alteração nomes colunas [ANEXO_C_PRODUCAO_NACIONAL]

Código SQL utilizado para realizar a tarefa:

```
CREATE OR REPLACE TABLE
projeto_mvp_engenharia_de_dados.ANEXO_C_PRODUCAO_NACIONAL AS
SELECT
    `Ano` AS ANO,
    `Mês` AS MES,
    `Código do Produto` AS CODIGO_PRODUTO,
    `Descrição do Produto` AS DESCRICAO_PRODUTO,
    `Região do Produtor` AS REGIAO_PRODUTOR,
    `UF do Produtor` AS UF_PRODUTOR,
    `Código da Operação` AS CODIGO_OPERACAO,
    `Descrição da Operação` AS DESCRICAO_OPERACAO,
    `Volume(L)` AS VOLUME_L
FROM projeto_mvp_engenharia_de_dados.anexo_c_producao_nacional;
```

- Alteração nomes colunas – [ANEXO_D_VENDAS_NACIONAIS_PROD]

Código SQL utilizado para realizar a tarefa:

```
CREATE OR REPLACE TABLE
projeto_mvp_engenharia_de_dados.ANEXO_D_VENDAS_NACIONAIS_PROD AS
SELECT
    `Ano` AS ANO,
    `Mês` AS MES,
    `Código do Regulado - ANP` AS CODIGO_REGULADO_ANP,
    `Agente Regulado` AS AGENTE_REGULADO,
    `Código do Produto` AS CODIGO_PRODUTO,
    `Descrição do Produto` AS DESCRICAO_PRODUTO,
    `Região do Destinatário` AS REGIAO_DESTINATARIO,
    `UF do Destinatário` AS UF_DESTINATARIO,
    `Volume(L)` AS VOLUME_L
FROM projeto_mvp_engenharia_de_dados.anexo_d_vendas_nacionais_prod;
```

- Alteração nomes colunas – [ANEXO_E_VENDAS_NACIONAIS_DISPENSA]

Código SQL utilizado para realizar a tarefa:

```
CREATE OR REPLACE TABLE
projeto_mvp_engenharia_de_dados.ANEXO_E_VENDAS_NACIONAIS_DISPENSA AS
SELECT
    `Ano` AS ANO,
    `Mês` AS MES,
    `Código do Regulado - ANP` AS CODIGO_REGULADO_ANP,
    `Agente Regulado` AS AGENTE_REGULADO,
    `Código do Produto` AS CODIGO_PRODUTO,
    `Descrição do Produto` AS DESCRICAO_PRODUTO,
    `Região do Destinatário` AS REGIAO_DESTINATARIO,
    `UF do Destinatário` AS UF_DESTINATARIO,
    `Volume(L)` AS VOLUME_L
FROM projeto_mvp_engenharia_de_dados.anexo_e_vendas_nacionais_dispensa;
```

- Alteração nomes colunas – [ANEXO_G_VOLUME_COLETADO_COLETOR]

Código SQL utilizado para realizar a tarefa:

```
CREATE OR REPLACE TABLE
projeto_mvp_engenharia_de_dados.ANEXO_G_VOLUME_COLETADO_COLETOR AS
SELECT
    `Ano` AS ANO,
    `Mês` AS MES,
    `Código do Regulador - ANP` AS CODIGO_REGULADO_ANP,
    `Agente Regulador` AS AGENTE_REGULADOR,
    `Região da Coleta` AS REGIAO_COLETA,
    `UF da Coleta` AS UF_COLETA,
    `Volume Coletado(L)` AS VOLUME_COLETADO_L
FROM projeto_mvp_engenharia_de_dados.anexo_g_volume_coletado_coletor;
```

- Alteração nomes colunas [ANEXO_I_OPERACOES_RERREFINADOR]

Código SQL utilizado para realizar a tarefa:

```
CREATE OR REPLACE
projeto_mvp_engenharia_de_dados.ANEXO_I_OPERACOES_RERREFINADOR AS
SELECT
```



```

`Ano` AS ANO,
`Mês` AS MES,
`Código do Regulado - ANP` AS CODIGO_REGULADO_ANP,
`Agente Regulado` AS AGENTE_REGULADO,
`Código da Operação` AS CODIGO_OPERACAO,
`Descrição da Operação` AS DESCRICAO_OPERACAO,
`Código do Produto` AS CODIGO_PRODUTO,
`Descrição do Produto` AS DESCRICAO_PRODUTO,
`Região do Destinatário` AS REGIAO_DESTINATARIO,
`UF do Destinatário` AS UF_DESTINATARIO,
`Volume(L)` AS VOLUME_L
FROM projeto_mvp_engenharia_de_dados.anexo_i_operacoes_rerrefinador;

```

- Padronização dos tipos de dados

Nesta etapa foram revisados todas as tabelas e **ajustados todos os tipos de dados** para garantir consistência e compatibilidade e qualidade dos dados.

Foram corrigidos campos que deveriam ser numéricos, mas foram importados como texto, e campos que devem permanecer como STRING.

- Ajuste dos tipos de dados da tabela [ANEXO_C_PRODUCAO_NACIONAL]

Código SQL utilizado para realizar a tarefa:

```

CREATE OR REPLACE TABLE anexo_c_producao_nacional AS
SELECT
    CAST(ANO AS INT) AS ANO,
    CAST(MES AS INT) AS MES,
    CAST(CODIGO_PRODUTO AS STRING) AS CODIGO_PRODUTO,
    CAST(DESCRICAO_PRODUTO AS STRING) AS DESCRICAO_PRODUTO,
    CAST(REGIAO_PRODUTOR AS STRING) AS REGIAO_PRODUTOR,
    CAST(UF_PRODUTOR AS STRING) AS UF_PRODUTOR,
    CAST(CODIGO_OPERACAO AS STRING) AS CODIGO_OPERACAO,
    CAST(DESCRICAO_OPERACAO AS STRING) AS DESCRICAO_OPERACAO,
    CAST(VOLUME_L AS DOUBLE) AS VOLUME_L
FROM anexo_c_producao_nacional;

```

- Ajuste dos tipos de dados da tabela [ANEXO_D_VENDAS_NACIONAIS_PROD]

Código SQL utilizado para realizar a tarefa:

```

CREATE OR REPLACE TABLE anexo_d_vendas_nacionais_prod AS
SELECT
    CAST(ANO AS INT) AS ANO,
    CAST(MES AS INT) AS MES,
    CAST(CODIGO_REGULADO_ANP AS STRING) AS CODIGO_REGULADO_ANP,
    CAST(AGENTE_REGULADO AS STRING) AS AGENTE_REGULADO,
    CAST(CODIGO_PRODUTO AS STRING) AS CODIGO_PRODUTO,
    CAST(DESCRICAO_PRODUTO AS STRING) AS DESCRICAO_PRODUTO,
    CAST(REGIAO_DESTINATARIO AS STRING) AS REGIAO_DESTINATARIO,
    CAST(UF_DESTINATARIO AS STRING) AS UF_DESTINATARIO,
    CAST(VOLUME_L AS DOUBLE) AS VOLUME_L
FROM anexo_d_vendas_nacionais_prod;

```

- Ajuste dos tipos de dados da tabela [ANEXO_E_VENDAS_NACIONAIS_DISPENSA]

Código SQL utilizado para realizar a tarefa:

```

CREATE OR REPLACE TABLE anexo_e_vendas_nacionais_dispensa AS
SELECT
    CAST(ANO AS INT) AS ANO,
    CAST(MES AS INT) AS MES,

```

```

CAST(CODIGO_REGULADO_ANP AS STRING) AS CODIGO_REGULADO_ANP,
CAST(AGENTE_REGULADO AS STRING) AS AGENTE_REGULADO,
CAST(CODIGO_PRODUTO AS STRING) AS CODIGO_PRODUTO,
CAST(DESCRICAO_PRODUTO AS STRING) AS DESCRICAO_PRODUTO,
CAST(REGIAO_DESTINATARIO AS STRING) AS REGIAO_DESTINATARIO,
CAST(UF_DESTINATARIO AS STRING) AS UF_DESTINATARIO,
CAST(VOLUME_L AS DOUBLE) AS VOLUME_L
FROM anexo_e_vendas_nacionais_dispensa;

```

- Ajuste dos tipos de dados da tabela [ANEXO_G_VOLUME_COLETADO_COLETOR]

Código SQL utilizado para realizar a tarefa:

```

CREATE OR REPLACE TABLE anexo_g_volume_coletado_coletor AS
SELECT
    CAST(`Ano` AS INT) AS ANO,
    CAST(`Mês` AS INT) AS MES,
    CAST(`Código do Regulado - ANP` AS STRING) AS CODIGO_REGULADO_ANP,
    CAST(`Agente Regulado` AS STRING) AS AGENTE_REGULADO,
    CAST(`Região da Coleta` AS STRING) AS REGIAO_COLETA,
    CAST(`UF da Coleta` AS STRING) AS UF_COLETA,
    CAST(`Volume Coletado(L)` AS DOUBLE) AS VOLUME_COLETADO_L
FROM anexo_g_volume_coletado_coletor;

```

- Ajuste dos tipos de dados da tabela [ANEXO_I_OPERACOES_RERREFINO]

Código SQL utilizado para realizar a tarefa:

```

CREATE OR REPLACE TABLE anexo_i_operacoes_rerrefinador AS
SELECT
    CAST(ANO AS INT) AS ANO,
    CAST(MES AS INT) AS MES,
    CAST(CODIGO_REGULADO_ANP AS BIGINT) AS CODIGO_REGULADO_ANP,
    CAST(AGENTE_REGULADO AS STRING) AS AGENTE_REGULADO,
    CAST(CODIGO_OPERACAO AS BIGINT) AS CODIGO_OPERACAO,
    CAST(DESCRICAO_OPERACAO AS STRING) AS DESCRICAO_OPERACAO,
    CAST(CODIGO_PRODUTO AS BIGINT) AS CODIGO_PRODUTO,
    CAST(DESCRICAO_PRODUTO AS STRING) AS DESCRICAO_PRODUTO,
    CAST(REGIAO_DESTINATARIO AS STRING) AS REGIAO_DESTINATARIO,
    CAST(UF_DESTINATARIO AS STRING) AS UF_DESTINATARIO,
    CAST(VOLUME_L AS DOUBLE) AS VOLUME_L
FROM anexo_i_operacoes_rerrefinador;

```

- Criação das tabelas Dimensão:

Nesta etapa foram criadas as tabelas dimensão:

- Calendário [dim_calendario]
- Estado e Região [dim_uf_estado]
- Produto [dim_produto]
- Regulado ANP [dim_regulado_ano].

Usando como fonte de dados as informações constantes nas tabelas que irão usar essas tabelas dimensões.

- Criação da tabela [dim_uf_regiao]

Código SQL utilizado para realizar a tarefa:

```
CREATE OR REPLACE TABLE dim_uf_regiao AS
SELECT DISTINCT
    UF_PRODUTOR      AS UF,
    REGIAO_PRODUTOR  AS REGIAO
FROM anexo_c_producao_nacional
UNION
SELECT DISTINCT
    UF_DESTINATARIO AS UF,
    REGIAO_DESTINATARIO AS REGIAO
FROM anexo_d_vendas_nacionais_prod
UNION
SELECT DISTINCT
    UF_DESTINATARIO AS UF,
    REGIAO_DESTINATARIO AS REGIAO
FROM anexo_e_vendas_nacionais_dispensa
UNION
SELECT DISTINCT
    UF_COLETA        AS UF,
    REGIAO_COLETA     AS REGIAO
FROM anexo_g_volume_coletado_coletor
UNION
SELECT DISTINCT
    UF_DESTINATARIO AS UF,
    REGIAO_DESTINATARIO AS REGIAO
FROM anexo_i_operacoes_rerrefinador;
```

Catalog

Serverless Starter Warehouse Serverless 2XS

Type to search...

For you All

- My organization
- workspace
 - default
 - information_schema
 - projeto_mvp_engenharia_de_dados
 - Tables (11)
 - dim_calendario
 - dim_operacoes
 - dim_produto
 - dim_regulado_anp
 - dim_uf_regiao**
 - fato_coleta_coletor
 - fato_operacoes_rerrefinador
 - fato_producao_nacional
 - fato_vendas_nacionais
 - fato_vendas_uf
 - fato_vendas_uf_dispensa
 - Volumes (1)

Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados

dim_uf_regiao

Overview **Sample Data** Details Permissions

Ask your question about the sample data...

How many unique states are there in each region?

Sample

	UF	REGIAO
1	AM	NORTE
2	CE	NORDESTE
3	PB	NORDESTE
4	RJ	SUDESTE
5	MT	CENTRO OESTE
6	GO	CENTRO OESTE
7	RN	NORDESTE
8	BA	NORDESTE
9	PE	NORDESTE
10	TO	NORTE
11	PR	SUL
12	RO	NORTE
13	MS	CENTRO OESTE
14	SP	SUDESTE
15	SC	SUL

- Criação da tabela [dim_calendario]

Código SQL utilizado para realizar a tarefa:

```
CREATE OR REPLACE TABLE dim_calendario AS
SELECT
  d AS DATA_INICIO_MES,
  year(d) AS ANO,
  month(d) AS MES,
  date_format(d, 'MMMM') AS NOME_MES,
  date_format(d, 'MMM') AS NOME_MES_ABREV,
  quarter(d) AS TRIMESTRE
FROM (
  SELECT explode(
    sequence(
      make_date(2015, 1, 1),
      make_date(2025, 12, 1),
      interval 1 month
    )
  ) AS d
);
```

The screenshot displays the Amazon Redshift Catalog Explorer interface. On the left, the 'Catalog' pane shows the hierarchy: 'Serverless Starter Warehouse' > 'Serverless' > '2XS' > 'workspace' > 'projeto_mvp_engenharia_de_dados' > 'Tables (11)'. The 'dim_calendario' table is selected. The main pane shows the 'dim_calendario' table details. The 'Sample Data' tab is active, displaying a table with the following columns: DATA_INICIO_MES, ANO, MES, NOME_MES, NOME_MES_ABREV, and TRIMESTRE. The sample data shows dates from 2015-01-01 to 2016-02-01.

	DATA_INICIO_MES	ANO	MES	NOME_MES	NOME_MES_ABREV	TRIMESTRE
1	2015-01-01	2015	1	January	Jan	1
2	2015-02-01	2015	2	February	Feb	1
3	2015-03-01	2015	3	March	Mar	1
4	2015-04-01	2015	4	April	Apr	2
5	2015-05-01	2015	5	May	May	2
6	2015-06-01	2015	6	June	Jun	2
7	2015-07-01	2015	7	July	Jul	3
8	2015-08-01	2015	8	August	Aug	3
9	2015-09-01	2015	9	September	Sep	3
10	2015-10-01	2015	10	October	Oct	4
11	2015-11-01	2015	11	November	Nov	4
12	2015-12-01	2015	12	December	Dec	4
13	2016-01-01	2016	1	January	Jan	1
14	2016-02-01	2016	2	February	Feb	1

- Criação da tabela [dim_produto]

Código SQL utilizado para realizar a tarefa:

```
CREATE OR REPLACE TABLE dim_produto AS
SELECT DISTINCT
  CODIGO_PRODUTO,
  DESCRICAO_PRODUTO
FROM anexo_c_producao_nacional
UNION
SELECT DISTINCT
  CODIGO_PRODUTO,
  DESCRICAO_PRODUTO
FROM anexo_d_vendas_nacionais_prod
UNION
SELECT DISTINCT
  CODIGO_PRODUTO,
  DESCRICAO_PRODUTO
FROM anexo_e_vendas_nacionais_dispensa
UNION
SELECT DISTINCT
```

```

CODIGO_PRODUTO,
DESCRICAO_PRODUTO
FROM anexo_i_operacoes_rerrefinador;

```

The screenshot shows the Databricks Catalog Explorer interface. On the left, the 'Catalog' pane displays the hierarchy: Serverless Starter Warehouse > Serverless > 2XS > workspace > projeto_mvp_engenharia_de_dados > Tables (11) > dim_produto. The 'dim_produto' table is selected. On the right, the 'dim_produto' table is displayed in the 'Sample Data' tab. The table has 14 rows and 2 columns: CODIGO_PRODUTO and DESCRICAO_PRODUTO. The data includes various product codes and descriptions such as 'ÓLEOS BÁSICOS - GRUPO II', 'ÓLEOS LUBRIFICANTES USADOS OU CONTAMINADOS', and 'ÓLEOS BÁSICOS - GRUPO III'.

	CODIGO_PRODUTO	DESCRICAO_PRODUTO
1	610601001	ÓLEOS BÁSICOS - GRUPO II
2	611301001	ÓLEOS BÁSICOS - GRUPO I
3	630101001	ÓLEOS LUBRIFICANTES USADOS OU CONTAMINADOS
4	610501001	OUTROS ÓLEOS LUBRIFICANTES BÁSICOS
5	610701001	ÓLEOS BÁSICOS - GRUPO III
6	620201001	ÓLEOS LUBRIFICANTES PARA AVIAÇÃO
7	620501002	CICLO DIESEL
8	620503001	MULTIFUNCIONAIS (TRATORES, MÁQUINAS AGRÍCOLAS, CONSTRUÇÃO, ETC)
9	620502001	MOTORES 2 TEMPOS
10	620501001	CICLO OTTO
11	620301001	ÓLEOS LUBRIFICANTES MARÍTIMOS
12	620601004	OUTROS ÓLEOS LUBRIFICANTES ACABADOS
13	620101002	ENGRENAGENS E SISTEMAS CIRCULATÓRIOS
14	620101005	ISOLANTE TIPO B

- Criação da tabela [dim_regulado_anp]]

Código SQL utilizado para realizar a tarefa:

```

CREATE OR REPLACE TABLE dim_regulado_anp AS
SELECT DISTINCT
    CODIGO_REGULADO_ANP,
    AGENTE_REGULADO
FROM anexo_d_vendas_nacionais_prod
UNION
SELECT DISTINCT
    CODIGO_REGULADO_ANP,
    AGENTE_REGULADO
FROM anexo_e_vendas_nacionais_dispensa
UNION
SELECT DISTINCT
    CODIGO_REGULADO_ANP,
    AGENTE_REGULADO
FROM anexo_g_volume_coletado_coletor
UNION
SELECT DISTINCT
    CODIGO_REGULADO_ANP,
    AGENTE_REGULADO
FROM anexo_i_operacoes_rerrefinador;

```

Catalog Serverless Starter Warehouse Serverless 2XS

Type to search...

For you **All**

- My organization
 - workspace
 - default
 - information_schema
 - projeto_mvp_engenharia_de_dados
 - Tables (11)
 - dim_calendario
 - dim_operacoes
 - dim_produto
 - dim_regulado_anp**
 - dim_uf_regiao
 - fato_coleta_coletor
 - fato_operacoes_rerrefinador
 - fato_producao_nacional
 - fato_vendas_nacionais
 - fato_vendas_uf
 - fato_vendas_uf_dispenda
 - Volumes (1)

Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados >

dim_regulado_anp ☆

Overview **Sample Data** Details Permissions Policies History Lineage Insights Quality

Ask your question about the sample data...

How many unique AGENTE_REGULADO entries exist? What is the distribution of CODIGO_REGULADO_ANP values? Are there any NULL values in AGE

Sample

	CODIGO_REGULADO_ANP	AGENTE_REGULADO
1	2017243269	RESOLUX DO BRASIL INDUSTRIA E COMERCIO ESPECIALIZADO EM ENERGIA EOLICA LTDA.
2	2004377382	BLASER SWISSLUBE DO BRASIL LTDA
3	2002260769	WD-40 COMPANY BRASIL LTDA
4	19739612	CBDL INDUSTRIA E COMERCIO LTDA
5	2068870997	BITZER COMPRESSORES LTDA
6	43085349	HOMY INDÚSTRIA E COMÉRCIO DE PRODUTOS QUÍMICOS LTDA
7	2002322453	PARTS IMPORT COMÉRCIO DE IMPORTAÇÃO E EXPORTAÇÃO LTDA.
8	2005083080	RACING LUB DO BRASIL IMPORTAÇÃO, EXPORTAÇÃO E COMÉRCIO LTDA
9	7685256	EXTRON INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA - EPP
10	2050606573	EASTMAN CHEMICAL DO BRASIL LTDA.
11	3006294505	NORTLUB RECICLAGEM DE ÓLEOS MINERAIS LTDA
12	2061577904	SOCIEDADE COMERCIAL TOYOTA TSUSHO DO BRASIL LTDA.
13	30523274	LLM LUBRIFICANTES LTDA
14	3613421	PETRONAS LUBRIFICANTES S.A
15	2005524572	ICONIC LUBRIFICANTES S.A.

- Criação da tabela [dim_operacoes]

Código SQL utilizado para realizar a tarefa:

```
CREATE OR REPLACE TABLE dim_operacoes AS
SELECT DISTINCT
    CODIGO_OPERACAO,
    DESCRICAO_OPERACAO
FROM anexo_c_producao_nacional
UNION
SELECT DISTINCT
    CODIGO_OPERACAO,
    DESCRICAO_OPERACAO
FROM anexo_i_operacoes_rerrefinador;
```

Catalog Serverless Starter Warehouse Serverless 2XS

Type to search...

For you **All**

- My organization
 - workspace
 - default
 - information_schema
 - projeto_mvp_engenharia_de_dados
 - Tables (11)
 - dim_calendario
 - dim_operacoes**
 - dim_produto
 - dim_regulado_anp
 - dim_uf_regiao
 - fato_coleta_coletor
 - fato_operacoes_rerrefinador
 - fato_producao_nacional
 - fato_vendas_nacionais
 - fato_vendas_uf
 - fato_vendas_uf_dispenda
 - Volumes (1)

Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados >

dim_operacoes ☆

Overview **Sample Data** Details Permissions Policies History Lineage Insight

Ask your question about the sample data...

How many distinct DESCRICAO_OPERACAO exist? Are there any null values in DESCRICAO

Sample

	CODIGO_OPERACAO	DESCRICAO_OPERACAO
1	1021004	PRODUÇÃO POR MISTURA
2	1022002	CONSUMO PRÓPRIO DE PRODUTO DE ORIGEM INTERNA
3	1012004	DEVOLUÇÃO DE PRODUTO PARA AGENTE REGULADO
4	1011004	RECEBIMENTO DE DEVOLUÇÕES DE AGENTE REGULADO
5	2012001	EXPORTAÇÃO DE PRODUTOS
6	1011001	COMPRA DE AGENTE REGULADO
7	1022004	PERDAS DE PROCESSO
8	1011013	COLETA DE ÓLEO LUBRIFICANTE
9	1012001	VENDA PARA AGENTE REGULADO
10	4030002	ÓLEO LUBRIFICANTE COLETADO CONTRATADO
11	1011002	COMPRA DE AGENTE NÃO REGULADO
12	2011001	IMPORTAÇÃO DE PRODUTOS
13	1022015	SAÍDA PARA PRODUÇÃO POR MISTURA
14	1012999	OUTRAS SAÍDAS NÃO ESPECIFICADAS
15	1012002	VENDA PARA AGENTE NÃO REGULADO

- Criação das tabelas Fato:

Nesta etapa foram criadas as tabelas fato seguindo a estrutura abaixo: A maioria das tabelas foram apenas renomeadas e apenas duas tabelas foram concatenadas.

Tabela Fato	Origem (tabela completa)	Ação realizada
fato_producao_nacional	anexo_c_producao_nacional	Tabela renomeada
fato_vendas_nacionais	anexo_d_vendas_nacionais_prod + anexo_e_vendas_nacionais_dispensa	Join entre duas tabelas
fato_coleta_coletor	anexo_g_volume_coletado_coletor	Tabela renomeada
fato_operacoes_rerrefinador	anexo_i_operacoes_rerrefinador	Tabela renomeada

Criação da tabela: [fato_vendas_nacionais], concatenando as tabelas originais [anexo_d_vendas_nacionais_prod] e [anexo_e_vendas_nacionais_dispensa]

Código SQL utilizado para realizar a tarefa:

```
CREATE OR REPLACE TABLE fato_vendas_nacionais AS
SELECT
    ANO,
    MES,
    CODIGO_REGULADO_ANP,
    AGENTE_REGULADO,
    CODIGO_PRODUTO,
    DESCRICAO_PRODUTO,
    REGIAO_DESTINATARIO,
    UF_DESTINATARIO,
    VOLUME_L,
    'anexo_d_vendas_nacionais_prod' AS FONTE_ANEXO
FROM anexo_d_vendas_nacionais_prod
UNION ALL
SELECT
    ANO,
    MES,
    CODIGO_REGULADO_ANP,
    AGENTE_REGULADO,
    CODIGO_PRODUTO,
    DESCRICAO_PRODUTO,
    REGIAO_DESTINATARIO,
    UF_DESTINATARIO,
    VOLUME_L,
    'anexo_e_vendas_nacionais_dispensa' AS FONTE_ANEXO
FROM anexo_e_vendas_nacionais_dispensa;
```

Tabela: [fato_vendas_nacionais]

Catalog

Serverless Starter Ware... Serverless 2X5

Type to search...

For you All

My organization

workspace

default

information_schema

projeto_mvp_engenharia_de_d...

Tables (11)

dim_calendario

dim_operacoes

dim_produto

dim_regulador_ano

dim_regiao

fato_vendas_nacionais

fato_vendas_uf_dispensa

Volumes (1)

Catalog Explorerworkspaceprojeto_mvp_engenharia_de_dados

fato_vendas_nacionais

Open in a dashboardShare

OverviewSample DataDetailsPermissionsPoliciesHistoryLineageInsightsQuality

Ask your question about the sample data...

What is the total sales volume per year?How does sales volume vary by month and region?Which products have the highest sales volume?

Preview

Sample

ANO	MES	CODIGO_REGULADOR_ANP	AGENTE_REGULADO	CODIGO...	DESCRICAO_PRODUTO	REGIAO...	UF_DES...	1.2 VOLUME_L	FONTE_ANEXO
1	2015	10	200296	TG Indústria e Comércio de Lubrificantes e Graxas Ltda. - EPP	620101002	ENGENHAGENS E SISTEMAS CIRCULATORIOS	SUL	PR	1600.0
2	2015	10	200296	TG Indústria e Comércio de Lubrificantes e Graxas Ltda. - EPP	620501001	CICLO OTTO	SUL	PR	96.0
3	2015	10	200296	TG Indústria e Comércio de Lubrificantes e Graxas Ltda. - EPP	620503001	TRANSMISSOES E SISTEMAS HIDRAULICOS	SUL	PR	80420.0
4	2015	10	200296	TG Indústria e Comércio de Lubrificantes e Graxas Ltda. - EPP	620601004	OUTROS OLÉOS LUBRIFICANTES ACABADOS	SUL	PR	12.0
5	2015	10	245992	CADIM COMÉRCIO IMPORTAÇÃO E EXPORTAÇÃO LTDA	620601004	OUTROS OLÉOS LUBRIFICANTES ACABADOS	SUDESTE	SP	24300.0
6	2015	10	384068	FALUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA	620101002	ENGENHAGENS E SISTEMAS CIRCULATORIOS	SUDESTE	SP	2000.0
7	2015	10	384068	FALUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA	620503001	TRANSMISSOES E SISTEMAS HIDRAULICOS	SUDESTE	SP	1000.0
8	2015	10	384068	FALUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA	620601004	OUTROS OLÉOS LUBRIFICANTES ACABADOS	SUDESTE	SP	32480.0
9	2015	10	616970	TECLUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA	620503001	MOTORES E TRAMPIS	SUL	PR	660.0
10	2015	10	616970	TECLUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA	620601004	OUTROS OLÉOS LUBRIFICANTES ACABADOS	SUL	PR	7526.0
11	2015	10	999042	QUAKER CHEMICAL INDÚSTRIA E COMÉRCIO LA	620601004	OUTROS OLÉOS LUBRIFICANTES ACABADOS	SUDESTE	RJ	194712.0
12	2015	10	1084176	REGELUB LUBRIFICANTES EIRELI	620601004	OUTROS OLÉOS LUBRIFICANTES ACABADOS	SUDESTE	SP	21304.0
13	2015	10	3099204	SPARK LUBRIFICANTES LTDA	620503001	TRANSMISSOES E SISTEMAS HIDRAULICOS	SUDESTE	SP	18253.0
14	2015	10	9788054	EDSON LUBRIFICANTES LTDA	620503001	TRANSMISSOES E SISTEMAS HIDRAULICOS	SUDESTE	ES	15001.0

Tabela: [fato_coleta_coletador]

Catalog

Serverless Starter Ware... Serverless 2X5

Type to search...

For you (All)

My organization

workspace

default

information_schema

projeto_mvp_engenharia_de_d...

Tables (11)

dim_calendario

dim_operacoes

dim_produto

dim_regulador_ano

dim_regiao

fato_coleta_coletador

fato_operacoes_rerrefinador

fato_producao_nacional

fato_vendas_nacionais

fato_vendas_uf

fato_vendas_uf_dispensa

Volumes (1)

Catalog Explorerworkspaceprojeto_mvp_engenharia_de_dados

fato_coleta_coletador

Open in a dashboardShare

OverviewSample DataDetailsPermissionsPoliciesHistoryLineageInsightsQuality

Ask your question about the sample data...

Average volume collected by region?Monthly volume variation by agent?Total volume collected each year?

Preview

Sample

	ANO	MES	CODIGO_REGULADOR_ANP	AGENTE_REGULADO	REGIAO_COLETA	UF_COLETA	1.2 VOLUME_COLETADO_L
1	2015	10	384068	FALUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA.	CO	GO	10669.0
2	2015	10	384068	FALUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA.	CO	MT	22000.0
3	2015	10	384068	FALUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA.	NI	AM	45000.0
4	2015	10	384068	FALUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA.	NE	BA	5680.0
5	2015	10	384068	FALUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA.	S	PR	95546.0
6	2015	10	384068	FALUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA.	S	SC	168795.0
7	2015	10	384068	FALUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA.	SE	MG	104440.0
8	2015	10	384068	FALUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA.	SE	RJ	45000.0
9	2015	10	384068	FALUB INDÚSTRIA E COMÉRCIO DE LUBRIFICANTES LTDA.	SE	SP	548327.0
10	2015	10	3002581865	WJ LUBRIFICANTES LTDA. - EPP	CO	MS	39759.0
11	2015	10	3002581865	WJ LUBRIFICANTES LTDA. - EPP	S	PR	118839.0
12	2015	10	3002581865	WJ LUBRIFICANTES LTDA. - EPP	SE	MG	4156.0
13	2015	10	3002581865	WJ LUBRIFICANTES LTDA. - EPP	SE	SP	384051.0
14	2015	10	3005276147	RS LUBRIFICANTES EIRELI - ME	S	RS	80660.0

Tabela: [fato_operacoes_rerefinador]

Catalog Explorerworkspaceprojeto_mvp_engenharia_de_dados

fato_operacoes_rerrefinador

Open in a dashboardShareCreate

OverviewSample DataDetailsPermissionsPoliciesHistoryLineageInsightsQuality

Ask your question about the sample data...

What is the total volume by year?Which agent handled the highest volume?What are the top 5 operations by volume?

Preview

Sample

ANO	MES	CODIGO...	AGENTE_REGULADO	CODIGO_OPERACAO	DESCRICAO_OPERACAO	CODIGO_PRODUTO	DESCRICAO_PRODUTO	REGIAO_DESTINATARIO	UF_DES...	1.2 VOLUME_L	
1	2007	7	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	54602.0
2	2007	8	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	56811.0
3	2007	8	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	85700.0
4	2007	10	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	54900.0
5	2007	11	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	144716.0
6	2007	12	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	189000.0
7	2008	1	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	161440.0
8	2008	2	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	156101.0
9	2008	3	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	317200.0
10	2008	5	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	140425.0
11	2008	10	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	275200.0
12	2009	2	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	129700.0
13	2009	3	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	427569.0
14	2009	5	3092678432	AVA LUBRIFICANTES E SOLUÇÕES AMBIENTAIS LTDA	1021002	PRODUÇÃO PRÓPRIA	610501001	OUTROS OLÉOS LUBRIF...	S	RS	399000.0

Tabela: [fato_producao_nacional]

Catalog

Serverless Starter Ware... Serverless 2X5

Type to search...

For you All

My organization

workspace

default

information_schema

projeto_mvp_engenharia_de_d...

Tables (11)

dim_calendario

dim_operacoes

dim_produto

dim_regulador_ano

dim_regiao

fato_coleta_coletador

fato_operacoes_rerefinador

fato_producao_nacional

fato_vendas_nacionais

fato_vendas_uf

fato_vendas_uf_dispensa

Volumes (1)

Catalog Explorerworkspaceprojeto_mvp_engenharia_de_dados

fato_producao_nacional

Open in a dashboard

OverviewSample DataDetailsPermissionsPoliciesHistoryLineageInsightsQuality

Ask your question about the sample data...

What is the total volume produced by each region?Which product had the highest production volume in 2023?Are there any months with anomalously low production volumes?

Sample

ANO	MES	CODIGO_PRODUTO	DESCRICAO_PRODUTO	REGIAO_PRODUTOR	UF_PRODUTOR	CODIGO_OPERACAO	DESCRICAO_OPERACAO	1.2 VOLUME_L	
1	2015	1	620101002	ENGENHAGENS E SISTEMAS CIRCULATORIOS	CENTRO OESTE	MS	1021004	PRODUÇÃO POR MISTURA	20.0
2	2015	1	620101002	ENGENHAGENS E SISTEMAS CIRCULATORIOS	SUDESTE	MG	1021004	PRODUÇÃO POR MISTURA	221270.0
3	2015	1	620101002	ENGENHAGENS E SISTEMAS CIRCULATORIOS	SUDESTE	RJ	1021004	PRODUÇÃO POR MISTURA	3166336.0
4	2015	1	620101002	ENGENHAGENS E SISTEMAS CIRCULATORIOS	SUDESTE	SP	1021004	PRODUÇÃO POR MISTURA	2472775.0
5	2015	1	620101002	ENGENHAGENS E SISTEMAS CIRCULATORIOS	SUL	PR	1021004	PRODUÇÃO POR MISTURA	1916.0
6	2015	1	620101002	ENGENHAGENS E SISTEMAS CIRCULATORIOS	SUL	RS	1021004	PRODUÇÃO POR MISTURA	429627.0
7	2015	1	620101004	ISOLANTE TIPO A	SUDESTE	SP	1021004	PRODUÇÃO POR MISTURA	34570.0
8	2015	1	620101005	ISOLANTE TIPO B	SUDESTE	SP	1021004	PRODUÇÃO POR MISTURA	70547.0
9	2015	1	620301001	OLÉOS LUBRIFICANTES MARÍTIMOS	CENTRO OESTE	MS	1021004	PRODUÇÃO POR MISTURA	104.0
10	2015	1	620301001	OLÉOS LUBRIFICANTES MARÍTIMOS	SUDESTE	RJ	1021004	PRODUÇÃO POR MISTURA	2307086.0
11	2015	1	620401001	OLÉOS LUBRIFICANTES FERROVIÁRIOS	SUDESTE	RJ	1021004	PRODUÇÃO POR MISTURA	800667.0
12	2015	1	620501001	CICLO OTTO	CENTRO OESTE	MS	1021004	PRODUÇÃO POR MISTURA	96188.0
13	2015	1	620501001	CICLO OTTO	CENTRO OESTE	MT	1021004	PRODUÇÃO POR MISTURA	14166.0
14	2015	1	620501001	CICLO OTTO	NORTE	AM	1021004	PRODUÇÃO POR MISTURA	89984.0

Registro dos metadados das tabelas

- Os metadados das tabelas foram registrados no banco de dados no Databricks. Descrevendo o dado e especificando as regras, limites, formatos de todos os campos.

Metadado da tabela: [dim_calendario]

Catalog

Serverless Starter Ware... Serverless 2XS

Type to search...

For you All

My organization

workspace

default

information_schema

projeto_mvp_engenharia_de_d...

Tables (11)

dim_calendario

dim_operacoes

dim_produto

dim_regulado_anp

dim_uf_regiao

fato_coleta_coletor

fato_operacoes_rerrefinador

fato_producao_nacional

fato_vendas_nacionais

fato_vendas_uf

fato_vendas_uf_dispensa

Volumes (1)

system

Delta Shares Received

samples

Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados >

dim_calendario

Overview Sample Data Details Permissions Policies History Lineage Insights Quality

Description

AI generate Add

Filter columns...

Column	Type	Comment	Tags
DATA_INICIO_MES	date	Junção entre o número do mês e numero do ano separados por underline, sendo 2 digitos para o mês e quatro digitos para o ano.	
ANO	int	Número de 4 digitos com dados de anos	
MES	int	Número de 2 digitos com dados de meses limitado de 01 a 12	
NOME_MES	string	Nome completo do mês limitado aos 12 meses do ano	
NOME_MES_ABBREV	string	Nome abreviado com três sílabas do mês limitado aos 12 meses do ano	
TRIMESTRE	int	Informação do trimestre, apenas um numeral limitado aos números de 1 ao 4 semestre	

Metadado da tabela: [dim_operacoes]

Catalog

Serverless Starter Ware... Serverless 2XS

Type to search...

For you All

My organization

workspace

default

information_schema

projeto_mvp_engenharia_de_d...

Tables (11)

dim_calendario

dim_operacoes

dim_produto

dim_regulado_anp

dim_uf_regiao

fato_coleta_coletor

fato_operacoes_rerrefinador

fato_producao_nacional

fato_vendas_nacionais

fato_vendas_uf

fato_vendas_uf_dispensa

Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados >

dim_operacoes

Overview Sample Data Details Permissions Policies History Lineage Insights Quality

Description

AI generate Add

Filter columns...

Column	Type	Comment	Tags
CODIGO_OPERACAO	bigint	Código único de identificação da operação com apenas números obrigatoriamente com 7 digitos	
DESCRICAO_OPERACAO	string	Texto descrição do tipo de operação com limite de 90 caracteres	

Metadado da tabela: [dim_produto]

Catalog

Serverless Starter Ware... Serverless 2XS

Type to search...

For you All

My organization

workspace

default

information_schema

projeto_mvp_engenharia_de_d...

Tables (11)

dim_calendario

dim_operacoes

dim_produto

dim_regulado_anp

dim_uf_regiao

fato_coleta_coletor

fato_operacoes_rerrefinador

fato_producao_nacional

fato_vendas_nacionais

fato_vendas_uf

fato_vendas_uf_dispensa

Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados >

dim_produto

Overview Sample Data Details Permissions Policies History Lineage Insights Quality

Description

AI generate Add

Filter columns...

Column	Type	Comment	Tags
CODIGO_PRODUTO	bigint	Código único do produto com apenas números obrigatoriamente com 9 dígitos	
DESCRICAO_PRODUTO	string	Texto descrição do produto com limite de 90 caracteres	

Metadado da tabela: [dim_regulado_anp]

Catalog

Serverless Starter Ware... Serverless 2XS

Type to search...

For you All

My organization

workspace

default

information_schema

projeto_mvp_engenharia_de_d...

Tables (11)

dim_calendario

dim_operacoes

dim_produto

dim_regulado_anp

dim_uf_regiao

fato_coleta_coletor

fato_operacoes_rerrefinador

fato_producao_nacional

fato_vendas_nacionais

fato_vendas_uf

fato_vendas_uf_dispensa

Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados >

dim_regulado_anp

Overview Sample Data Details Permissions Policies History Lineage Insights Quality

Description

AI generate Add

Filter columns...

Column	Type	Comment	Tags
CODIGO_REGULADO_ANP	bigint	Código único do agente regulado com limite de até 10 números	
AGENTE_REGULADO	string	Texto descrição do nome do agente regulado com limite de 90 caracteres	

Metadado da tabela: [dim_uf_regiao]

Catalog

Serverless Starter Ware... Serverless 2XS

Type to search...

For you All

My organization

workspace

default

information_schema

projeto_mvp_engenharia_de_d...

Tables (11)

dim_calendario

dim_operacoes

dim_produto

dim_regulado_anp

dim_uf_regiao

fato_coleta_coletor

fato_operacoes_rerefinador

fato_producao_nacional

fato_vendas_nacionais

fato_vendas_uf

fato_vendas_uf_dispena

Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados >

dim_uf_regiao

Overview Sample Data Details Permissions Policies History Lineage Insights Quality

Description

AI generate Add

Filter columns...

Column	Type	Comment	Tags
UF	string	Texto com dois caracteres representando os estados do Brasil	
REGIAO	string	Texto descritos as regiões do Brasil	

Metadado da tabela: [fato_coleta_coletor]

Catalog

Serverless Starter Ware... Serverless 2XS

Type to search...

For you All

My organization

workspace

default

information_schema

projeto_mvp_engenharia_de_d...

Tables (11)

dim_calendario

dim_operacoes

dim_produto

dim_regulado_anp

dim_uf_regiao

fato_coleta_coletor

fato_operacoes_rerefinador

fato_producao_nacional

fato_vendas_nacionais

fato_vendas_uf

fato_vendas_uf_dispena

Volumes (1)

system

Delta Shares Received

samples

Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados >

fato_coleta_coletor

Overview Sample Data Details Permissions Policies History Lineage Insights Quality

Description

Created by the file upload UI

Filter columns...

Column	Type	Comment	Tags
ANO	int	Informação do ano, com 4 dígitos obrigatoriamente	
MES	int	Informação do mês, com 2 dígitos obrigatoriamente	
CODIGO_REGULADO_ANP	string	Código único do agente regulado com limite de até 10 números	
AGENTE_REGULADO	string	Texto descrição do nome do agente regulado com limite de 90 caracteres	
REGIAO_COLETA	string	Texto descritos com as regiões do Brasil	
UF_COLETA	string	Texto com dois caracteres representando os estados do Brasil	
VOLUME_COLETADO_L	double	Valor numérico, formatado com duas casas decimais separados por ponto "."	

Metadado da tabela: [fato_operacoes_rerrefinador]

Catalog

Serverless Starter Ware... Serverless 2XS

Type to search...

For you All

My organization

- workspace
 - default
 - information_schema
 - projeto_mvp_engenharia_de...
 - Tables (11)
 - dim_calendario
 - dim_operacoes
 - dim_produto
 - dim_regulado_anp
 - dim_uf_regiao
 - fato_coleta_coletor
 - fato_operacoes_rerrefinador
 - fato_producao_nacional
 - fato_vendas_nacionais
 - fato_vendas_uf
 - fato_vendas_uf_dispensa
- Volumes (1)
- system

Delta Shares Received

- samples

Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados >

fato_operacoes_rerrefinador

Overview Sample Data Details Permissions Policies History Lineage Insights Quality

DescriptionCreated by the file upload UI

Filter columns...

Column	Type	Comment	Tags
ANO	bigint	Informação do ano, com 4 dígitos obrigatoriamente	
MES	bigint	Informação do mês, com 2 dígitos obrigatoriamente	
CODIGO_REGULADO_ANP	bigint	Código único do agente regulado com limite de até 10 números	
AGENTE_REGULADO	string	Texto descrição do nome do agente regulado com limite de 90 caracteres	
CODIGO_OPERACAO	bigint	Código único de identificação da operação com apenas números obrigatoriamente com 7 dígitos	
DESCRICAO_OPERACAO	string	Texto descrição do tipo de operação com limite de 90 caracteres	
CODIGO_PRODUTO	bigint	Código único do produto com apenas números obrigatoriamente com 9 dígitos	
DESCRICAO_PRODUTO	string	Texto descrição do produto com limite de 90 caracteres	
REGIAO_DESTINATARIO	string	Texto descritos as regiões do Brasil	
UF_DESTINATARIO	string	Texto com dois caracteres representando os estados do Brasil	
VOLUME_L	double	Valor número, formatado com duas casas decimais separados por ponto "."	

Metadado da tabela: [fato_producao_nacional]

Catalog

Serverless Starter Ware... Serverless 2XS

Type to search...

For you All

My organization

workspace

default

information_schema

projeto_mvp_engenharia_de_d...

Tables (11)

dim_calendario

dim_operacoes

dim_produto

dim_regulado_anp

dim_uf_regiao

fato_coleta_coletor

fato_operacoes_rerrefinador

fato_producao_nacional

fato_vendas_nacionais

fato_vendas_uf

fato_vendas_uf_dispena

Volumes (1)

system

Delta Shares Received

samples

Catalog Explorer > workspace > projeto_mvp_engenharia_de_dados >

fato_producao_nacional

Overview Sample Data Details Permissions Policies History Lineage Insights Quality

Description

Created by the file upload UI

Filter columns...

Column	Type	Comment	Tags
ANO	int	Número de 4 dígitos com dados de anos	
MES	int	Número de 2 dígitos com dados de meses limitado de 01 a 12	
CODIGO_PRODUTO	string	Código único do produto com apenas números obrigatoriamente com 9 dígitos	
DESCRICAO_PRODUTO	string	Texto descrição do produto com limite de 90 caracteres	
REGIAO_PRODUTOR	string	Texto descritos as regiões do Brasil	
UF_PRODUTOR	string	Texto com dois caracteres representando os estados do Brasil	
CODIGO_OPERACAO	string	Código único de identificação da operação com apenas números obrigatoriamente com 7 dígitos	
DESCRICAO_OPERACAO	string	Texto descrição do tipo de operação com limite de 90 caracteres	
VOLUME_L	double	Valor número, formatado com duas casas decimais separados por ponto "."	

Metadado da tabela: [fato_vendas_nacional]

fato_vendas_nacionais

Overview | Sample Data | Details | Permissions | Policies | History | Lineage | Insights | Quality

Description

AI generate | Add

Filter columns...

Column	Type	Comment	Tags
ANO	int	Número de 4 dígitos com dados de anos	
MES	int	Número de 2 dígitos com dados de meses limitado de 01 a 12	
CODIGO_REGULADO_ANP	string	Código único do agente regulado com limite de até 10 números	
AGENTE_REGULADO	string	Texto descrição do nome do agente regulado com limite de 90 caracteres	
CODIGO_PRODUTO	string	Código único do produto com apenas números obrigatoriamente com 9 dígitos	
DESCRICAO_PRODUTO	string	Texto descrição do produto com limite de 90 caracteres	
REGIAO_DESTINATARIO	string	Texto descritos as regiões do Brasil	
UF_DESTINATARIO	string	Texto com dois caracteres representando os estados do Brasil	
VOLUME_L	double	Valor numérico, formatado com duas casas decimais separados por ponto "."	
FONTE_ANEXO	string	Preenchimento automático, Identificador da tabela fonte de dados dessa informação.	

Conexão e carga das tabelas no Power Bi Fabric

- Para construção dos relatórios no Power BI utilizando as tabelas criadas no Databricks, foram configurado uma conexão direta entre as duas ferramentas.

Utilizando a funcionalidade nativa da ferramenta na aba de opção **“Connection Details”** do **“SQL Warehouse”**, foram obtidos os dois parâmetros necessários para a conexão:

- **Server Hostname**
- **HTTP Path**

Essas informações foram inseridas no Power BI na opção nativa da ferramenta **“Conexão > Azure Databricks”**, possibilitando o acesso às tabelas criadas no ambiente e garantindo que os dados possam ser consumidos diretamente para modelagem e criação das visualizações necessárias para proporcionar as respostas previstas nesse projeto.

Serverless Starter Warehouse ●

Overview **Connection details** Monitoring

Use these details to connect to this warehouse



Tableau



Power BI



dbt



Python



Java



Node.js



Go



More tools

Server hostname

dbc-31e69a92-a625.cloud.databricks.com



HTTP path

/sql/1.0/warehouses/ab165e58030437e5



JDBC URL **2.6.25 or later** ▼

jdbc:databricks://dbc-31e69a92-a625.cloud.databricks.com:443/default;transportMode=http;ssl=1;AuthMech=3;httpPath=/sql/1.0/warehouses/ab165e58030437e5;



Databricks supports drivers released within the last two years. [Download drivers here](#)

OAuth URL

https://dbc-31e69a92-a625.cloud.databricks.com/oidc



Azure Databricks

Nome do Host do Servidor ⓘ

dbc-31e69a92-a625.cloud.databricks.com

Caminho HTTP ⓘ

/sql/1.0/warehouses/ab165e58030437e5

⌵ Opções Avançadas (opcional)

Catálogo padrão (opcional) ⓘ

Exemplo: abc

Banco de Dados (opcional) ⓘ

projeto_mvp_engenharia_de_dados

Tags de consulta (opcional) ⓘ

Exemplo: tag1:value1,tag2:value2

Descoberta Automática de Proxy (opcional) ⓘ

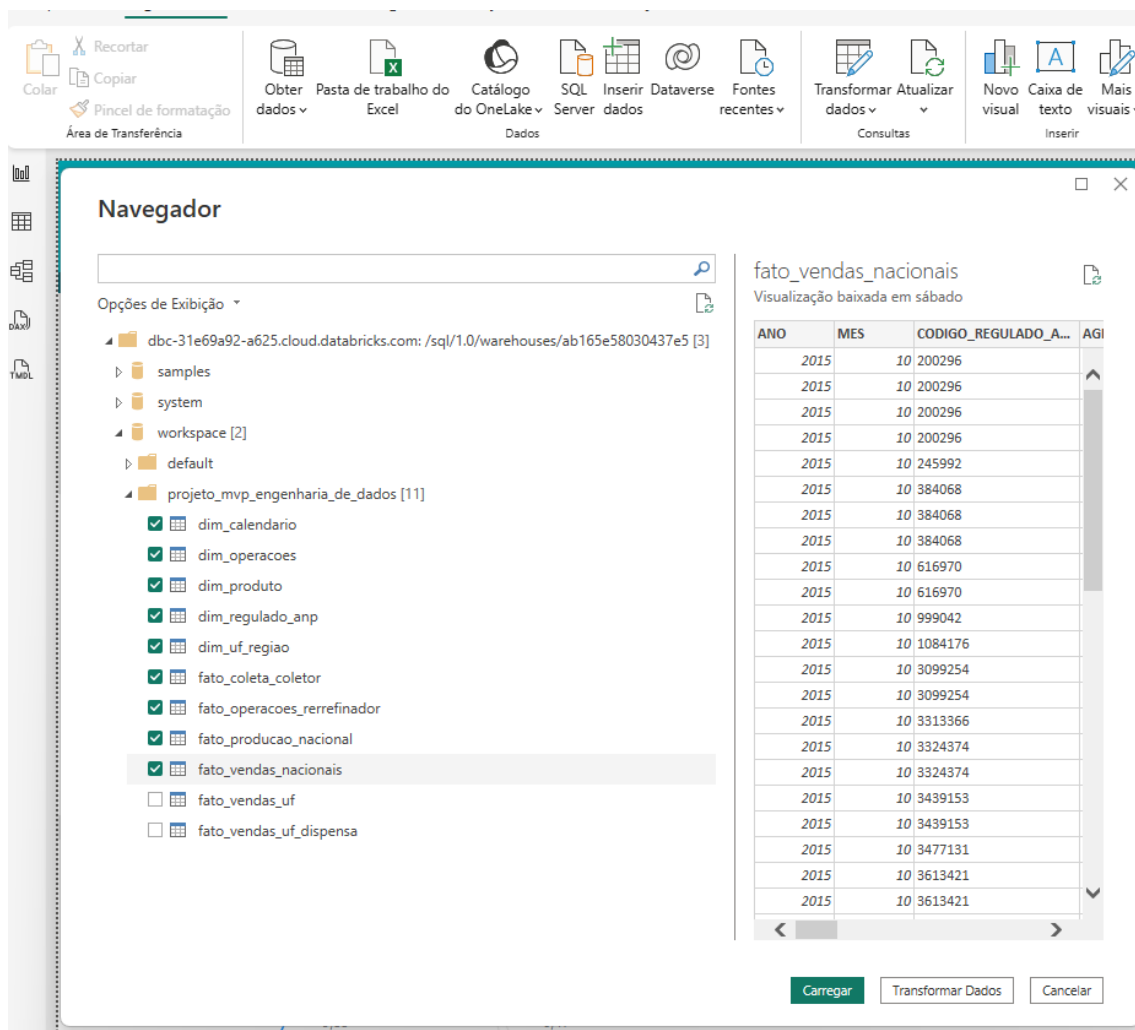
Implementação (opcional) ⓘ

Consulta nativa (Requer: Catálogo padrão) (opcional) ⓘ

Exemplo: selecione * a partir de db.schemaname.tablename

OK

Cancelar



As tabelas fato_vendas_uf_dispenda e fato_vendas_uf não foram carregadas para o modelo pois foram unificadas na tabela nova fato_vendas_nacional

Modelagem de dados - Criação do modelo estrela na plataforma de conexões nativo do Power bi.

Utilizei um modelo estrela clássico, onde as tabelas fato centralizam medidas e chaves estrangeiras, e as dimensões têm chaves primárias, que serão utilizadas para para filtragem de dados e análise.

Foram utilizadas todas as tabelas previstas e criadas para o projeto. Alguns procedimentos adicionais foram realizados para criação da modelagem e conexões e que seguem descritos abaixo.

- Tabelas Dimensão:

- São as tabelas que que serão utilizadas para criação de filtros de acesso e outras funcionalidades.
- Serão conectadas a todas as tabelas fato que possuem seus atributos.

dim_calendario

dim_uf_regiao
dim_produto
dim_regulado_anp
dim_operacoes

- Tabelas Fato:

- São as bases de dados transacionais: que possuem os dados de quantidades, volumes, etc

fato_producao_nacional
fato_vendas_nacionais
fato_coleta_coletor
fato_operacoes_rerefinaria

- Foram ajustados os nomes dos campos para identificação de **Chave primária (PK_)** e **chaves estrangeiras (FK_)**

- Foi criado para todas as tabelas, o campo customizado de data MES_ANO, concatenando os atributos MÊS + ANO = MES_ANO. Ajustando o formato para “data”. Os campos MES e ANO da tabela foram excluídos para não ocuparem espaço desnecessário e evitando redundâncias.

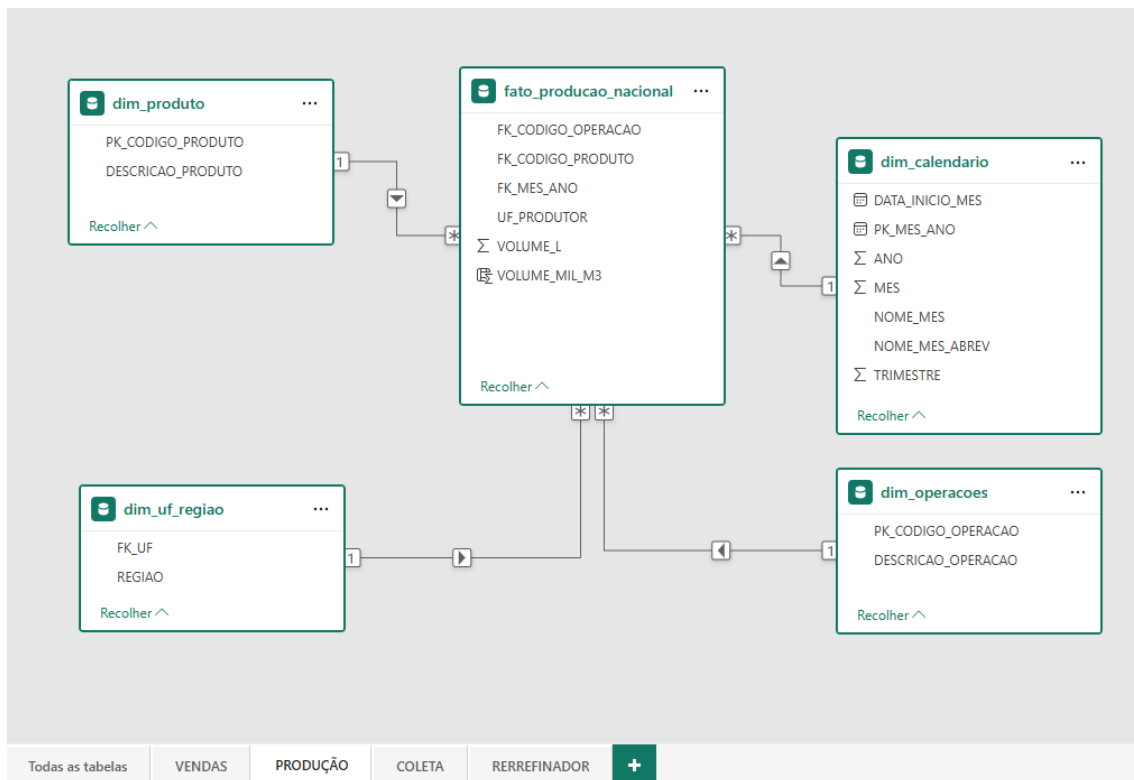
- Modelo de dados estrela das tabelas relacionadas a PRODUÇÃO:

- Foi criado campo customizado de data MES_ANO, concatenando os atributos MÊS + ANO = MÊS/ANO. Ajustando o formato para “data”. Os campos MÊS e ANO da tabela foram excluídos para não ocuparem espaço desnecessário e evitando redundâncias.

- Com o objetivo de reduzir o volume e garantir a qualidade dos dados. Foi excluído o campo DESCRIÇÃO_PRODUTO pois esse campo já consta na tabela dim_produto.

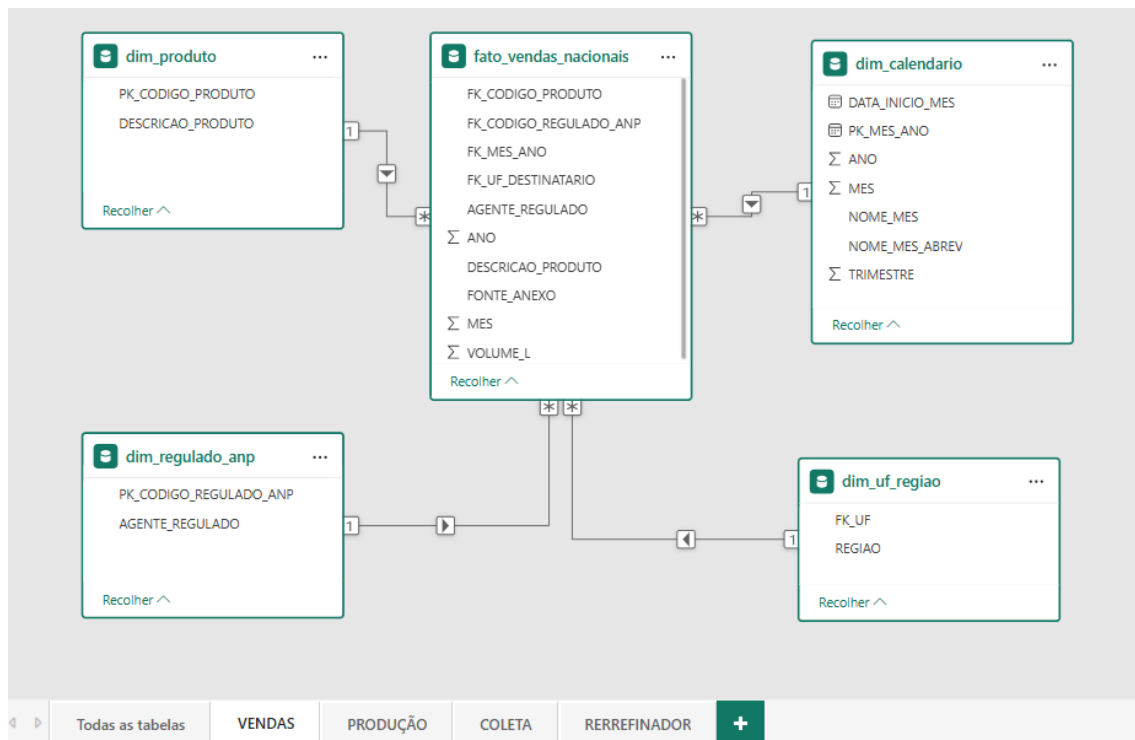
- Com o objetivo de reduzir o volume e garantir a qualidade dos dados. Foi excluído o campo DESCRIÇÃO_OPERACAO pois esse campo já consta na tabela dim_operacao.

- Com o objetivo de reduzir o volume e garantir a qualidade dos dados. Foi excluído o campo REGIAO_PRODUTOR pois esse campo já consta na tabela dim_uf_regiao.



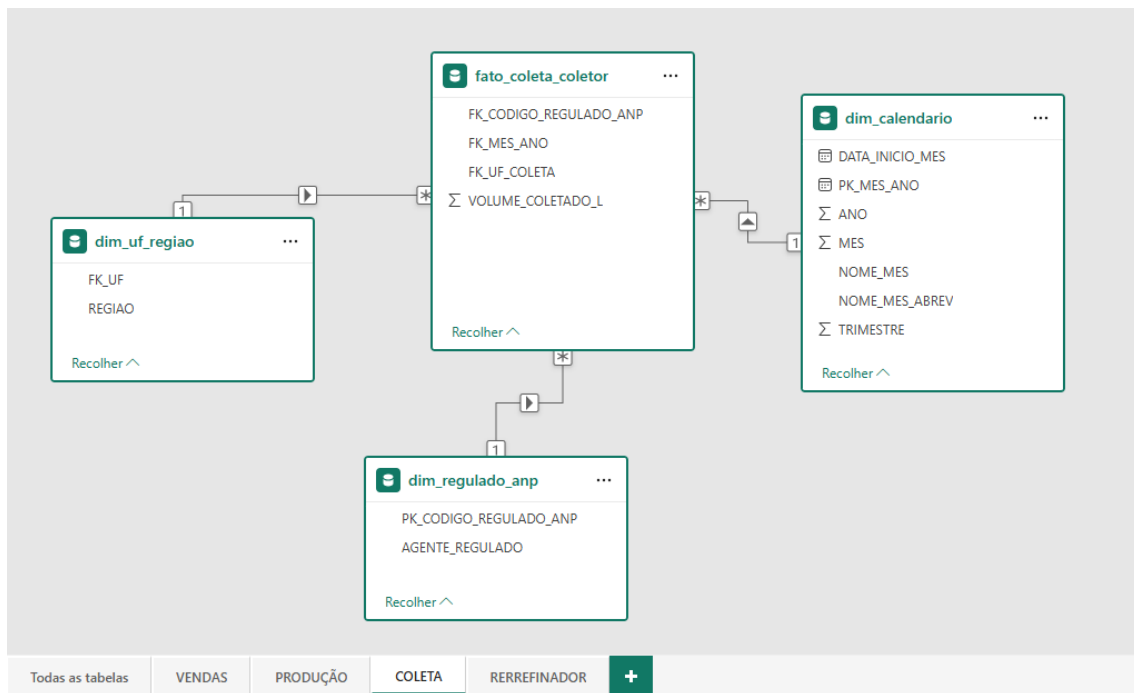
- Modelo de dados estrela das tabelas relacionadas a VENDAS:

- Foi criado campo customizado de data MES_ANO, concatenando os atributos MÊS + ANO = MÊS/ANO. Ajustando o formato para “data”. Os campos MÊS e ANO da tabela foram excluídos para não ocuparem espaço desnecessário e evitando redundâncias.
- Com o objetivo de reduzir o volume e garantir a qualidade dos dados. Foi excluído o campo DESCRIÇÃO_PRODUTO pois esse campo já consta na tabela dim_produto.
- Com o objetivo de reduzir o volume e garantir a qualidade dos dados. Foi excluído o campo REGIAO_DESTINATARIO pois esse campo já consta na tabela dim_uf_região.



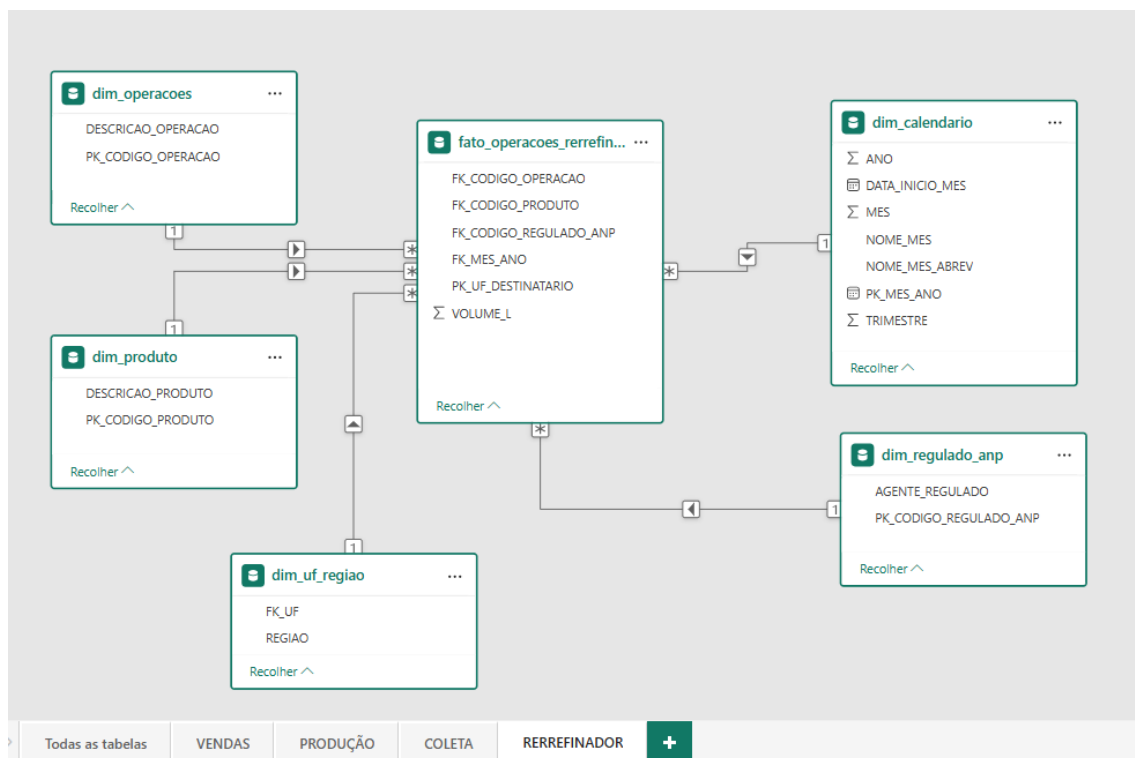
- Modelo de dados estrela das tabelas relacionadas a COLETA:

- Foi criado campo customizado de data MES_ANO, concatenando os atributos MÊS + ANO = MÊS/ANO. Ajustando o formato para "data". Os campos MÊS e ANO da tabela foram excluídos para não ocuparem espaço desnecessário e evitando redundâncias.
- Com o objetivo de reduzir o volume e garantir a qualidade dos dados. Foi excluído o campo AGENTE_REGULADO pois esse campo já consta na tabela dim_regulado_anp
- Com o objetivo de reduzir o volume e garantir a qualidade dos dados. Foi excluído o campo FK_REGIAO_COLETA pois esse campo já consta na tabela dim_uf_região.



- Modelo de dados estrela das tabelas relacionadas a RERREFINADOR:

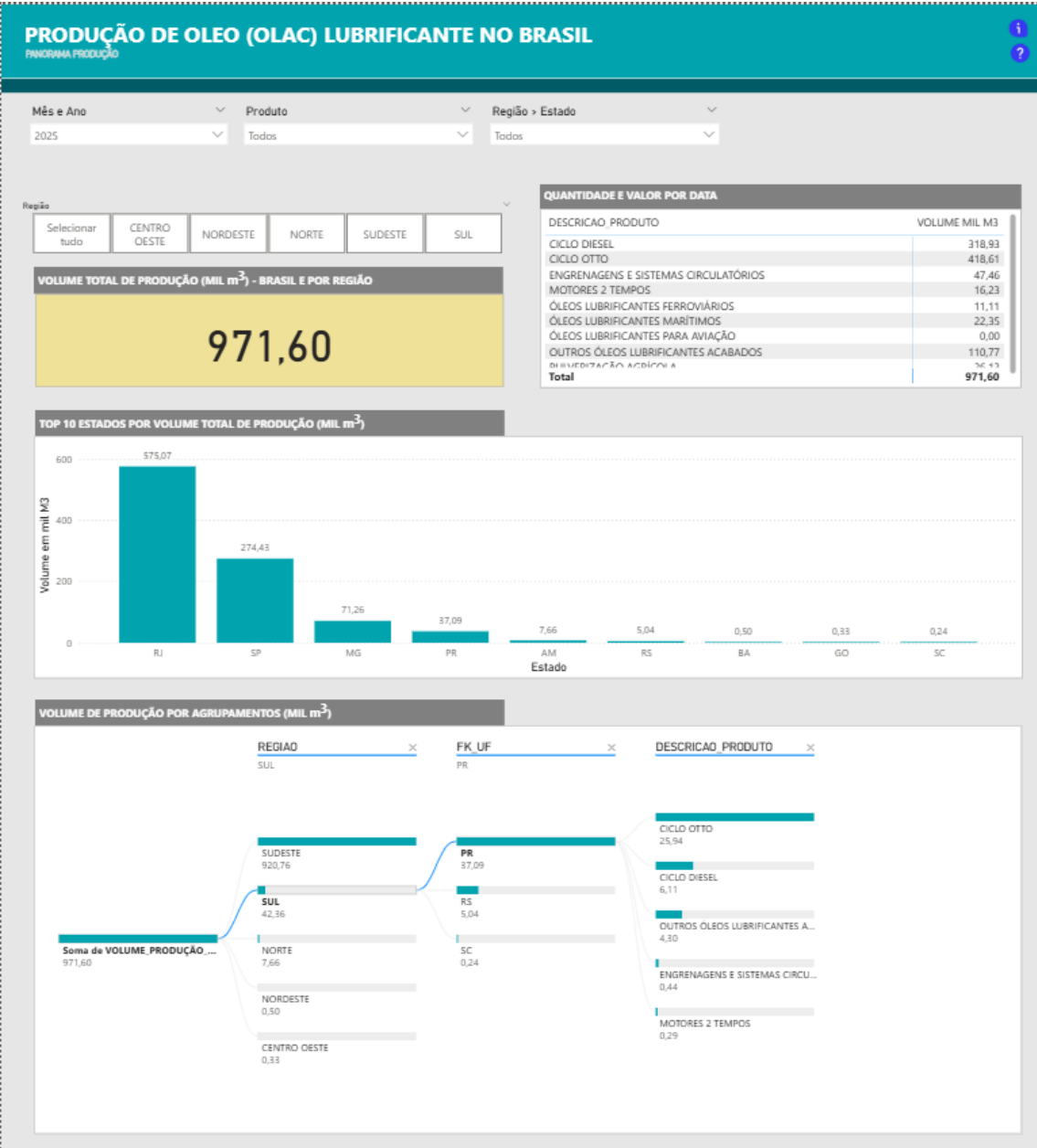
- Foi criado campo customizado de data MES_ANO, concatenando os atributos MÊS + ANO = MÊS/ANO. Ajustando o formato para "data". Os campos MÊS e ANO da tabela foram excluídos para não ocuparem espaço desnecessário e evitando redundâncias.
- Com o objetivo de reduzir o volume e garantir a qualidade dos dados. Foi excluído o campo AGENTE_REGULADO pois esse campo já consta na tabela dim_regulado_anp.
- Com o objetivo de reduzir o volume e garantir a qualidade dos dados. Foi excluído o campo DESCRIÇÃO_OPERACAO pois esse campo já consta na tabela dim_operacao.
- Com o objetivo de reduzir o volume e garantir a qualidade dos dados. Foi excluído o campo DESCRIÇÃO_PRODUTO pois esse campo já consta na tabela dim_produto.
- Com o objetivo de reduzir o volume e garantir a qualidade dos dados. Foi excluído o campo FK_REGIAO_DESTINATARIO pois esse campo já consta na tabela dim_uf_região.



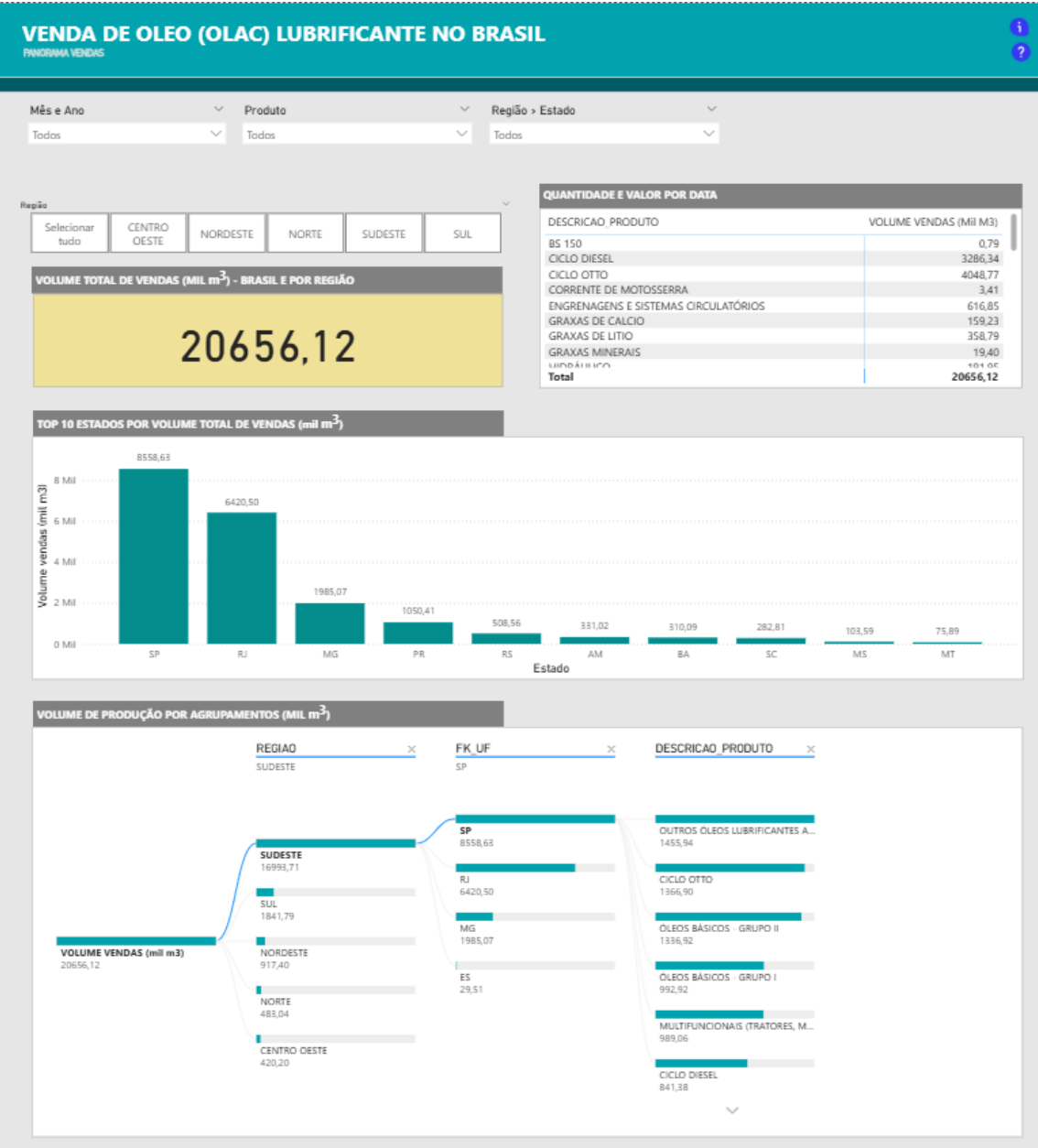
Relatórios de respostas implementados no Power bi.

Arquivo PDF publicado no Github possui também os prints de todo relatório

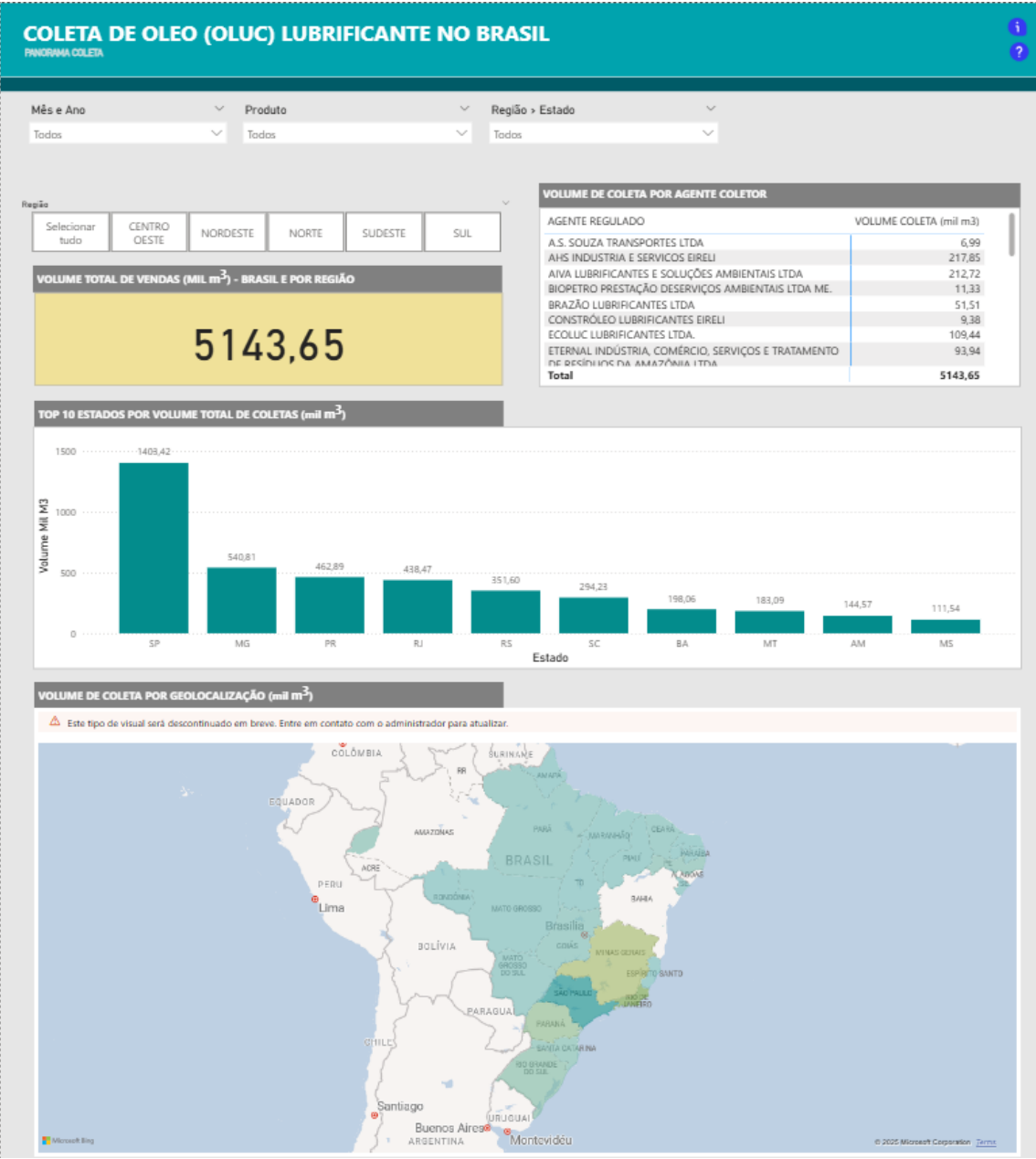
- Relatório de respostas de PRODUÇÃO



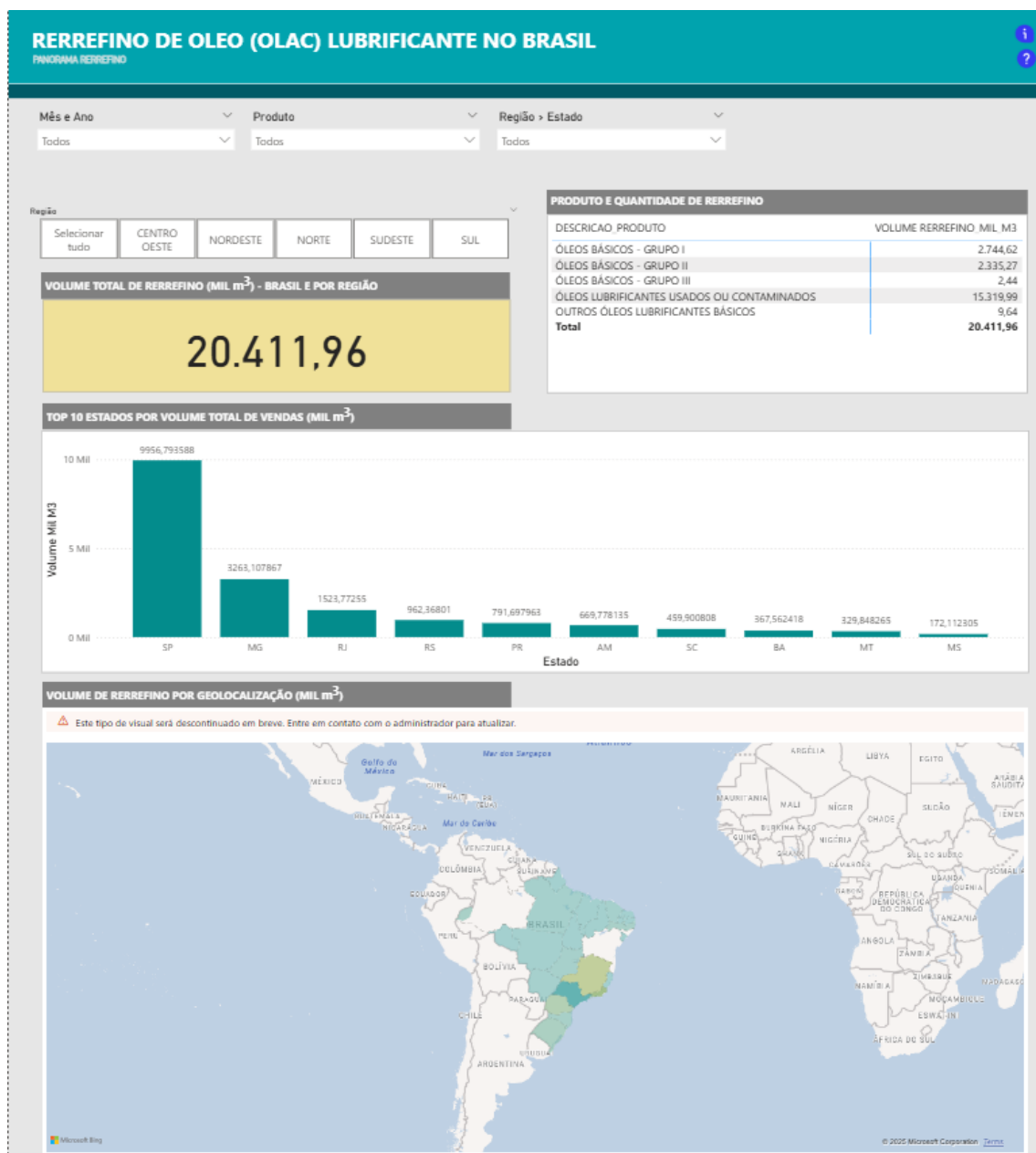
- Relatório de respostas de VENDAS



- Relatório de respostas de COLETAS



- Relatório de respostas de RERREFINO



AUTO AVALIAÇÃO

- O objetivo do trabalho foi criar um panorama amplo do cenário de fabricação e venda de óleo lubrificante no Brasil. Atendendo a respostas interessantes sobre como esse mercado de fabricação e venda funcionam, mostrando o grande volume que circula no país todos os meses e as regiões mais representativas para esse mercado. Constatando que a região sudeste está a frente em volumes de fabricação e venda em relação a todo país.

Durante o levantamento de dados foi observado que a ANP – agência nacional do petróleo exige que as empresas produtoras de óleo garantam um economia circular, sendo também responsável sobre o óleo já utilizado, garantindo a coleta de óleo já utilizado e envio para as empresas de rerrefino.

A ANP gera um certificado para as empresas que fazem o Rerrefino, nesse trabalho a entrega das certificações não foram considerados. O projeto não aprofundou na área de certificação que a ANP gera para as empresas, considerando as empresas já homologadas pela ANP.

Esse projeto também não considerou as metas de coleta geridas pela ANP, que são medidas sobre o óleo já utilizado (OLUC) coletado por região.

As oportunidades de evolução desse projeto podem considerar o aprofundamento das informações sobre Certificados e também Metas. Também tem oportunidade de que sejam criados relacionamentos entre informações como fabricação e venda, coleta e fabricação nos estados do Brasil, trazendo um cenário onde será possível ter visualização de movimentação desse volume.

ENTREGA DO PROJETO – Abaixo os links para acesso aos arquivos que foram gerados para atender ao projeto .

- Github (https://github.com/alexandreandrade/projeto_mvp_engenharia_de_dados)

- Link para o vídeo de resposta do projeto:

<https://youtu.be/08n7WwgzJ-U>