PLANO DO PROJETO

Comércio Brasileiro

**Cliente**: Maria de Lourdes Rodrigues – EPOC

**Gerente do projeto**: Bismark William. – COEXP

**Término previsto**: 17 de Janeiro de 2022

Cópias distribuídas para:

coexp@coexp.com.br

mariadelurdes@epoc.pt

[professores.bcw4@soulcodeacademy.org](mailto:professores.bcw4@soulcodeacademy.org)

**Sumário**

1 – Termo de abertura de projeto

1.1- Job Description

1.2- Limites e entregas-chave

1.3- Risco do projeto

1.4- Recursos financeiros pré-aprovados

1.5- Lista das partes interessadas

1.6- Requisitos para aprovação do projeto

1.7- Critérios para término do projeto

1.8- Gerente de projeto, responsabilidade e seu nível de autoridade

1.9- Nome e autoridade do patrocinador ou outra(s) pessoa(s) que autoriza(m) o termo de abertura do projeto

2 – Declaração do escopo do projeto

2.1- Descrição do escopo do projeto

2.2- Critérios de aceitação

2.3- Exclusões do projeto

2.4- Premissas

2.5- Escopo – PGE

2.6- Descrição do escopo do projeto

3 – Cronograma

4 – Produtos do Projeto

4.1 - Códigos Python

4.2 - Relatório em Dashboard (Looker Studio)

4.3 - Apresentação do projeto

5 – Tecnologias utilizadas no projeto

5.1- Linguagem da programação

5.2- Programação em nuvem

5.3- MongoDB Atlas

6 - Funcionalidades e demonstração da aplicação

6.1 BigQuery

7 – Insights

8 – Valor total do projeto

8.1 – Consultoria

8.2 – BigQuery

8.3 – VM GCE no Colab por meio do GCP Marketplace

8.4 – Cloud Storage

8.5 – Instância SQL

8.6 – DataProc

8.7 – Dataflow

9 – Serviços Optativos – Adicionais

9.1 – Instância SQL – Para datasets normalizados

9.2 – Instância do MongoDB Atlas – Para datasets originais

10 – Pessoas Desenvolvedoras do Projeto

10.1 – Ana Flavia Dias

10.2 – Ariana Caetano

10.3 – Érika Cunha

10.4 – José Jardim

10.5 – Marcos Valente

## Secção 1

**Termo de abertura de projeto**

**Secção 1**

Subseção 1.1

**Job Description**

A Empresa Portuguesa de Óleos e Combustíveis (EPOC, empresa fictícia), tendo interesse em investir no mercado brasileiro de combustíveis derivados de petróleo, contratou os serviços de Engenharia e Análise de Dados da COEXP. Dentre os objetivos da contratante estão investir na extração de óleo na Bacia de Santos e no refino de combustíveis nos estados de São Paulo e Rio de Janeiro. A fim de embasar seus estudos a respeito da viabilidade do negócio, foram contratados os seguintes serviços:

* A disponibilização de dados brutos e tratados a respeito do volume de vendas combustíveis automotivos (incluindo etanol hidratado) no Brasil, bem como o histórico de preços pagos pelo consumidor final; dados esses que serão posteriormente trabalhados pelo time de analistas de dados da EPOC;
* A disponibilização de dados brutos e tratados a respeito do comércio exterior (importação e exportação) de petróleo e combustíveis automotivos;
* Uma análise preliminar – para coleta de insights – a respeito dos dados disponibilizados, enfatizando-se a distribuição geográfica do consumo de combustíveis automotivos e o impacto que tiveram as crises do coronavírus e da guerra na Ucrânia sobre esse mercado.

A COEXP tradicionalmente oferta – de maneira gratuita – a seus clientes dados a respeito do comércio exterior brasileiro de maneira ampla. Nesse sentido, também se deve disponibilizar dados a respeito das importações e exportações nacionais nos últimos dez anos (2013-2022), bem como uma análise preliminar para coleta de insights.

**Secção 1**

### Subseção 1.2

**Limites e entregas-chave**

O projeto se baseia em dados oficiais disponibilizados pelo Governo Federal (gov.br), como a plataforma COMEX STAT (importações e exportações brasileiras), vinculada ao Ministério da Economia; bem como pesquisas realizadas pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, ANP.

O resultado do serviço será:

• Extração dos dados brutos

• Preparação dos dados (normalização)

• Processamento dos dados originais

• Armazenamento e disponibilização dos dados originais

• Análise dos dados normalizados

• Coleta de insights

• Armazenamento e disponibilização dos dados normalizados

**Secção 1**

### Subseção 1.3

**Risco do projeto**

• Atraso na aquisição e análise dos dados, ocasionando uma prorrogação do prazo de término do projeto;

• Falta de comprometimento por parte cliente ao suporte necessário para que a equipe da COEXP realize a atividade com sucesso.

**Secção 1**

### Subseção 1.4

**Recursos financeiros pré-aprovados**

|  |  |  |
| --- | --- | --- |
| **Orçamento/ Estimativa** | | |
| **Item** | **Valor(R$)** | **%** |
| Serviço COPEX | 18.000,00 | 84,72 |
| Produto: BIG QUERY | Gratuito | 0,00 |
| Produto: VM GCE COLAB | 2.521,15 | 11,86 |
| Produto: Cloud Storage | 1,85 | 0,02 |
| Produto: Instância SQL | 721,06 | 3,39 |
| Total | 21.244,06 | 100,00 |

\*Cotação do dólar em 10-01-2023: $5,30

\*VM GCE COLAB: 5 unid.

**Secção 1**

### Subseção 1.5

**Lista das partes interessadas**

|  |  |  |
| --- | --- | --- |
| **Empresa** | **Participante** | **Função** |
| COPEX | Bismark William | Gerente de projetos |
| EPOC | Maria de Lourdes | Gerente |
| COPEX | Igor Gondim | Diretor de planejamento e projetos especiais |

**Secção 1**

### Subseção 1.6

**Requisitos para aprovação do projeto**

Este projeto será considerado bem-sucedido se o orçamento previsto não for ultrapassado.

A decisão quanto o sucesso do projeto ficará ao encargo do gerente do projeto, o Eng. Bismark William, a cliente, a senhora Maria de Lourdes e o patrocinador, o diretor de planejamento e projetos especiais, Igor Gondim.

**Secção 1**

Subseção 1.7

**Critérios para término do projeto**

Para encerramento deste projeto deve-se cumprir a auditoria necessária, e após a entrega do relatório o projeto será considerado encerrado

**Secção 1**

### Subseção 1.8

**Gerente de projeto, responsabilidade e seu nível de autoridade**

O gerente do projeto será o Eng. Bismark William, Gerente do Depto. COPEX, tem a responsabilidade de:

• Conduzir a identificação dos stakeholders, assim como, suas necessidades e expectativas;

• Coletar requisitos;

• Conduzir todo o processo de planejamento, orientar a execução do projeto, realizar os controles necessários mantendo sempre a linha de base do projeto ajustada;

• Avaliar o desempenho da equipe e mantê-la motivada ao longo de todo o projeto;

* E, como responsabilidade precípua, empenhar-se pessoalmente para que os objetivos do projeto sejam atingidos, respeitando as restrições de escopo, tempo, custo, qualidade, risco e satisfação do cliente.

Também possui autoridade para:

• Alterar o escopo do projeto a qualquer momento;

• Delegar funções para os membros de sua equipe de trabalho citados acima no “Stakeholder Membros da Equipe”;

• Utilizar os recursos financeiros da empresa até o limite do orçamento aprovado para este projeto;

• Contratar empresa especializada para realizar atividades técnicas para o projeto;

• Comprar materiais e equipamentos que estejam dentro do orçamento previsto e com a qualidade técnica aceitável.

**Secção 1**

### Subseção 1.9

**Nome e autoridade do patrocinador ou outra(s) pessoa(s) que autoriza(m) o termo de abertura do projeto**.

Além do cliente do projeto possui autoridade para autorizar as modificações necessárias para este projeto, o patrocinador do mesmo exercido pela senhor Igor Gondim, Diretora de Planejamento e Projetos Especiais.

|  |  |  |
| --- | --- | --- |
| **Aprovações** | | |
| **Participante** | **Assinatura** | **Data** |
| Cliente do Projeto |  |  |
| Patrocinador do Projeto |  |  |
| Gerente do Projeto |  |  |

## Secção 2

**Declaração de escopo do projeto**

**Secção 2**

### Subseção 2.1

**Descrição do escopo do projeto**

O projeto será o serviço de Engenharia e Análise de dados, com vistas a:

1- Disponibilizar dados brutos pertinentes à tomada de decisão do cliente;

2- Disponibilizar os mesmos dados normalizados;

3- Diagnosticar a distribuição da balança comercial do Brasil nos últimos 10 anos, enfatizando-se os produtos de interesse do cliente.

4- Comparar a importação e exportação nos últimos 10 anos, analisando se as crises do Covid-19 e da Guerra da Ucrânia as afetaram.

5- Analisar a evolução do mercado interno de combustíveis, com ênfase para preços repassados ao cliente final, bem como volume de vendas;

6- Para encerrar este projeto serão realizadas análises e avaliação dos resultados para composição do relatório final.

**Secção 2**

### Subseção 2.2

**Critérios de aceitação**

• Todos os datasets devem ser confiáveis, armazenados em locais seguros e categorizados hierarquicamente;

• Devem-se priorizar o modelo de negócio e buscar fontes que atendam a demanda;

• Devem-se verificar métricas de desempenho para analisar o melhor resultado

• Devem-se fazer uma auditoria sobre os dados;

• Devem-se planejar a implantação da política de armazenamentos e gestão de dados na empresa;

• Devem-se realizar um tratamento em todos os dados brutos;

• Devem-se extrair informações relevantes para o projeto;

**Secção 2**

### Subseção 2.3

**Exclusões do projeto**

• Não serão utilizados datasets que não estejam dentro do padrão determinado;

• Não serão feitos acompanhamentos para mensurar outros dados após a finalização do projeto.

**Secção 2**

### Subseção 2.4

**Premissas**

• Os membros da equipe do projeto estarão motivados durante toda a execução do

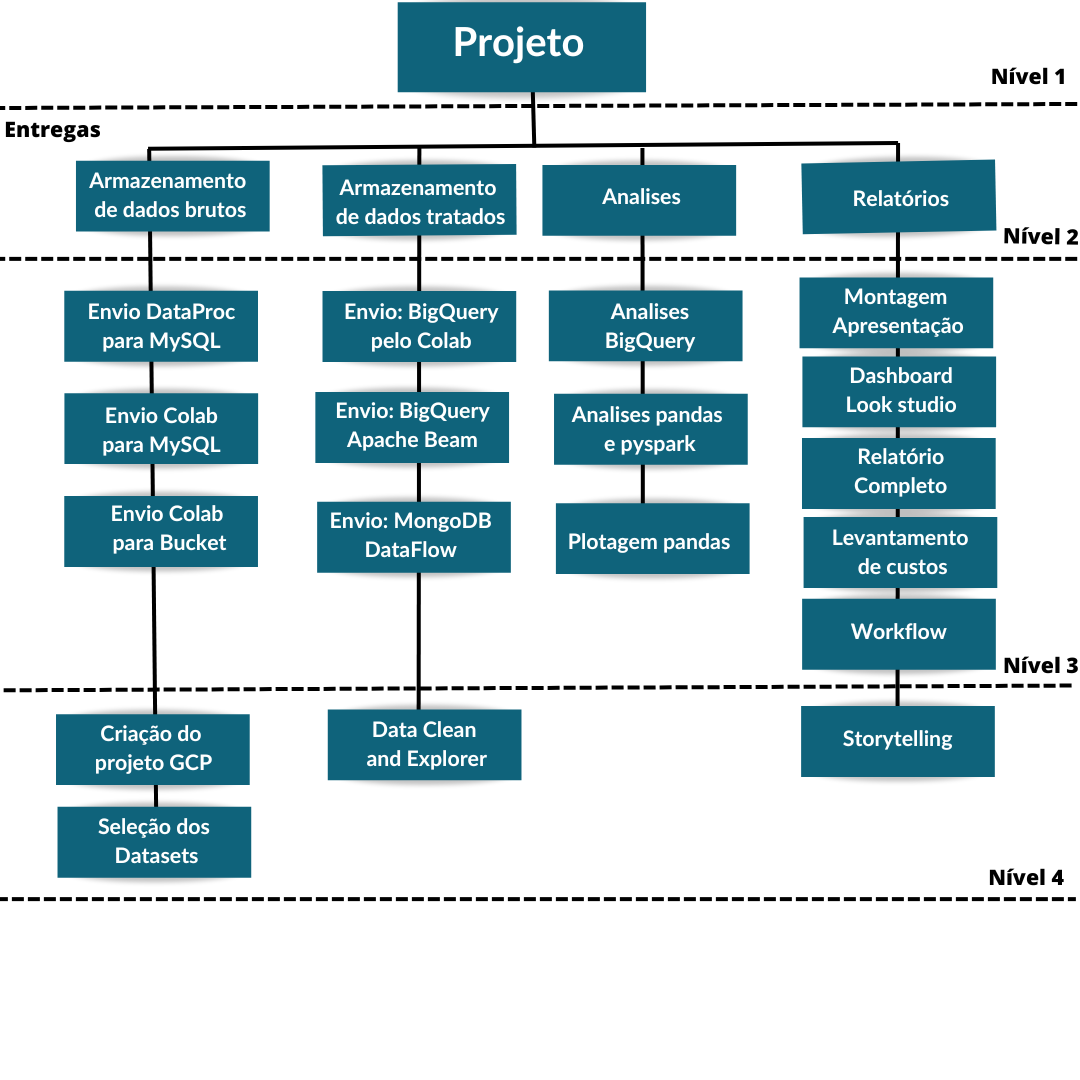
projeto para finalizá-lo dentro prazo e orçamento previsto;

• O valor total estimado do projeto estará disponível para a execução do mesmo;

**Secção 2**

### Subseção 2.5

**Escopo – PGE**



**Secção 2**

*Subseção 2.6*

**Descrição do escopo do projeto**

O projeto será o serviço de análise de dados do mercado brasileiro de combustíveis derivados de petróleo, que terá as seguintes fases:

1ª Fase: Seleção e criação

1.1 - Seleção dos datasets.

1.2 - Criação do projeto na GCP e configurações da infraestrutura (DataLake no Google Cloud Storage (GCS, MySQL, DataProc)

2º Fase: Armazenamento dos datasets no MySQL e GCS

2.1: Envio: pelo Colab para GCS

2.2: Envio: pelo DataProc para MySQL

2.3: Envio: pelo Colab para MySQL

3ª Fase: Data Clean and Explorer com Pandas e Pyspark

4ª Fase - Armazenamento dos dados normalizados em um Data Lake (GCS) e BigQuery

4.1 Envio pelo Colab para GCS

4.2 Envio para o BigQuery dos dados normalizados através de pipeline automatizada (Apache Beam + Dataflow)

4.3 Envio para o BigQuery pelo Colab através de biblioteca (pandas\_gbq)

5ª Fase: Armazenamento dados normalizados no MongoDB Atlas

5.1 Pipeline automatizada MongoDB – Template BigQuery para o MongoDB

5.2 Envio pelo Colab para MongoDB (não concluída por falta de armazenamento)

6ª Fase: Análises Pandas e Pyspark

7ª Fase: Plotagem Pandas

8ª Fase: Análises no BigQuery

9 ª Fase: Elaboração dos relatórios

9.1 - Storytelling

9.2 - Workflow

9.3 - Dashboard Looker Studio

9.4 - Levantamento de custos totais

9.5 - Relatório completo com os insights que justificam todo o processo de ETL.

9.6 - Montagem apresentação

## Secção 3

**Cronograma**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Cronograma do Projeto | | | | | | | | | | | |
| | Dia do mês de Janeiro de 2022 | | | | | | | | | | | |
| Nº | Atividade | 03 | 04 | 05 | 06 | 09 | 10 | 11 | 12 | 13 | 16 |
| 1.1 | Seleção dos datasets |  |  |  |  |  |  |  |  |  |  |
| 1.2 | Criação do projeto na GCP e configurações da infraestrutura (GCS, MySQL, DataProc) |  |  |  |  |  |  |  |  |  |  |
| 2.1 | Envio: pelo Colab (bibliotecas Python) para GCS (DataLake) |  |  |  |  |  |  |  |  |  |  |
| 2.2 | Envio: pelo DataProc para MySQL |  |  |  |  |  |  |  |  |  |  |
| 2.3 | Envio: pelo Colab (bibliotecas Python) para MySQL |  |  |  |  |  |  |  |  |  |  |
| 3.0 | Data Clean and Explorer com Pandas e Pyspark |  |  |  |  |  |  |  |  |  |  |
| 4.1 | Envio pelo Colab (bibliotecas Python) para Google Cloud Storage |  |  |  |  |  |  |  |  |  |  |
| 4.2 | Envio para o BigQuery dados tratados através de pipeline (Apache Beam + Dataflow) |  |  |  |  |  |  |  |  |  |  |
| 4.3 | Envio para o BigQuery dados normalizados pelo Colab através de biblioteca (pandas\_gbq) |  |  |  |  |  |  |  |  |  |  |
| 5.1 | Pipeline automatizada MongoDB – Template BigQuery para o MongoDB |  |  |  |  |  |  |  |  |  |  |
| 5.2 | Envio pelo Colab para MongoDB Atlas (não concluído por falta de armazenamento) |  |  |  |  |  |  |  |  |  |  |
| 6.0 | Analises Pandas e Pyspark |  |  |  |  |  |  |  |  |  |  |
| 7.0 | Plotagem Pandas |  |  |  |  |  |  |  |  |  |  |
| 8.0 | Análises no BigQuery |  |  |  |  |  |  |  |  |  |  |
| 9.1 | Storytelling |  |  |  |  |  |  |  |  |  |  |
| 9.2 | Workflow |  |  |  |  |  |  |  |  |  |  |
| 9.3 | Dashboard Looker Studio |  |  |  |  |  |  |  |  |  |  |
| 9.4 | Levantamento de custos totais |  |  |  |  |  |  |  |  |  |  |
| 9.5 | Relatório completo com os insights que justificam todo o processo de ETL |  |  |  |  |  |  |  |  |  |  |
| 9.6 | Montagem apresentação |  |  |  |  |  |  |  |  |  |  |

## Secção 4

**Produtos do projeto**

Os datasets sobre importação e exportação dos produtos do comércio exterior do Brasil, foram disponibilizados pelo Mistério da Economia no website do Governo Federal. O dataset de importação e o dataset de exportação são detalhadas pelo município e pelo produto, em formato CSV. Atualizados pela última vez com dados até novembro de 2022. Esses dados formam 4 datasets que contém as seguintes informações:

* Ano
* Mês
* Código do produto
* Código da unidade de medida estática
* Código do país de origem ou destino
* Código do estado
* Código de município
* Código de via de transporte
* Código da unidade da receita federal, URF, de embarque ou desembarque
* Quantidade estática
* Volume em quilograma líquido
* Valores em dólares

Como grandes partes dos dados apresentam códigos, as tabelas de correlações de códigos e classificações também são disponibilizadas.

Fonte: <https://www.gov.br/produtividade-e-comercio-exterior/pt-br/assuntos/comercio-exterior/estatisticas/base-de-dados-bruta>

Os datasets sobre o mercado interno de combustíveis são disponibilizados pela Agência Nacional de Petróleo, Gás Natural e Biocombustíveis, ANP, também no website do Governo Federal, no formato de planilhas eletrônicas com a série histórica da pesquisa de Levantamento de Preço. Um dataset contém os preços dos combustíveis pesquisados agregada por municípios, abrangendo os seguintes combustíveis: gasolina, etanol hidratado, óleo diesel, gás natural veicular (GNV) e gás liquefeito de petróleo (GLP). O outro dataset contém as vendas dos derivados de petróleo e etanol, as vendas dos combustíveis por segmento e por tipo, as vendas anuais de etanol hidratado e derivados de petróleo por município e estado.

Fonte: <https://www.gov.br/anp/pt-br/assuntos/precos-e-defesa-da-concorrencia/precos/precos-revenda-e-de-distribuicao-combustiveis/serie-historica-do-levantamento-de-precos>

Fonte: <https://www.gov.br/anp/pt-br/centrais-de-conteudo/dados-abertos/vendas-de-derivados-de-petroleo-e-biocombustiveis>

**Secção 4**

Subseção 4.1

**Códigos Python**

**Colab Síntese: Disponível em:**

[**https://colab.research.google.com/drive/1Q1q3luKpiVYLFPHHXTVjnJICNRrbyprs?usp=sharing**](https://colab.research.google.com/drive/1Q1q3luKpiVYLFPHHXTVjnJICNRrbyprs?usp=sharing)

* Compilado, único, de todos os cadernos de código descritos a seguir;

**Colab 01: Disponível em:**

[**https://colab.research.google.com/drive/14Iuj8jRWyEtyntcf3MUkQyocq\_GKykhO?usp=sharing**](https://colab.research.google.com/drive/14Iuj8jRWyEtyntcf3MUkQyocq_GKykhO?usp=sharing)

* Conexão com GCP;
* Download de datasets originais e upload para GCS;
* Análises com Pandas;
* Normalização de tabelas auxiliares;
* Envio de dados normalizados para Bucket;
* Envio de dados normalizados para BigQuery;

**Colab 02: Disponível em:**

[**https://colab.research.google.com/drive/1lqZ8jtAtEbTtEfDtkLD4EBGSjWkU9mM-?usp=sharing**](https://colab.research.google.com/drive/1lqZ8jtAtEbTtEfDtkLD4EBGSjWkU9mM-?usp=sharing)

* Conexão com GCP;
* Análises de dados normalizados com PySpark;

**Colab 03: Disponível em:**

[**https://colab.research.google.com/drive/1Uz7MRgW-JGnmbZs9RyHgsBVIj6sbkktR?usp=sharing**](https://colab.research.google.com/drive/1Uz7MRgW-JGnmbZs9RyHgsBVIj6sbkktR?usp=sharing)

* Conexão com GCP;
* Normalização de dados com PySpark;
* Análises com Pandas;
* Envio de dados normalizados para BigQuery;
* Envio de dados normalizados para GCS;

**Colab 04: Disponível em:**

[**https://colab.research.google.com/drive/1C4cLmlvRa\_cj7rJm7ZkRNcJezNNqeikL?usp=sharing**](https://colab.research.google.com/drive/1C4cLmlvRa_cj7rJm7ZkRNcJezNNqeikL?usp=sharing)

* Conexão com GCP;
* Normalização de dados com Pandas;
* Envio de dados normalizados para BigQuery;
* Envio de dados normalizados para GCS;

**Colab 05: Disponível em:**

[**https://colab.research.google.com/drive/14dJOfz9ZOs5yktszb7FZCTp78TAUtr48?usp=sharing#scrollTo=MPdIL8rZ-Ocr**](https://colab.research.google.com/drive/14dJOfz9ZOs5yktszb7FZCTp78TAUtr48?usp=sharing#scrollTo=MPdIL8rZ-Ocr)

* Conexão com GCP;
* Normalização de dados com Pandas;
* Análises com pandas;
* Envio de dados normalizados para GCS;

**Colab 06: Disponível em:**

[**https://colab.research.google.com/drive/13BfC99pdar-zl7AqZ1OOrjlS-JBkouwF?usp=sharing**](https://colab.research.google.com/drive/13BfC99pdar-zl7AqZ1OOrjlS-JBkouwF?usp=sharing)

* Envio de datasets de importação e exportação originais para MySQL (DataProc)

**Colab 07: Disponível em:**

[**https://colab.research.google.com/drive/1viCY8ru3IuEHKG5tZKHUHQlm040F3DT-?usp=sharing**](https://colab.research.google.com/drive/1viCY8ru3IuEHKG5tZKHUHQlm040F3DT-?usp=sharing)

* Conexão com GCP;
* Conexão e envio de datasets originais para MySQL;
* Normalização de dados com Pandas;
* Normalização de dados com PySpark;
* Envio de dados normalizados para GCS;
* Envio de dados normalizados para BigQuery;
* Envio de dados normalizados para MongoDB;
* Análise de dados com Pandas.

**Colab 08: Disponível em:**

[**https://colab.research.google.com/drive/15dKDbt0sqRoft0255Q5-mhJgUi-HMmpr?usp=sharing**](https://colab.research.google.com/drive/15dKDbt0sqRoft0255Q5-mhJgUi-HMmpr?usp=sharing)

* Pipeline Apache-Beam – envio de dados Bucket para BigQuery

Subseção 4.2

**Relatório em Dashboard (Looker Studio)**

**Looker Studio 01: Disponível em:**

[**https://datastudio.google.com/reporting/11fe3610-6344-4cfc-9b58-37b575baeee9**](https://datastudio.google.com/reporting/11fe3610-6344-4cfc-9b58-37b575baeee9)

* Dashboard elaborado no Looker Studio;
* Inclui apresentação do projeto.

Subseção 4.3

**Apresentação do projeto**

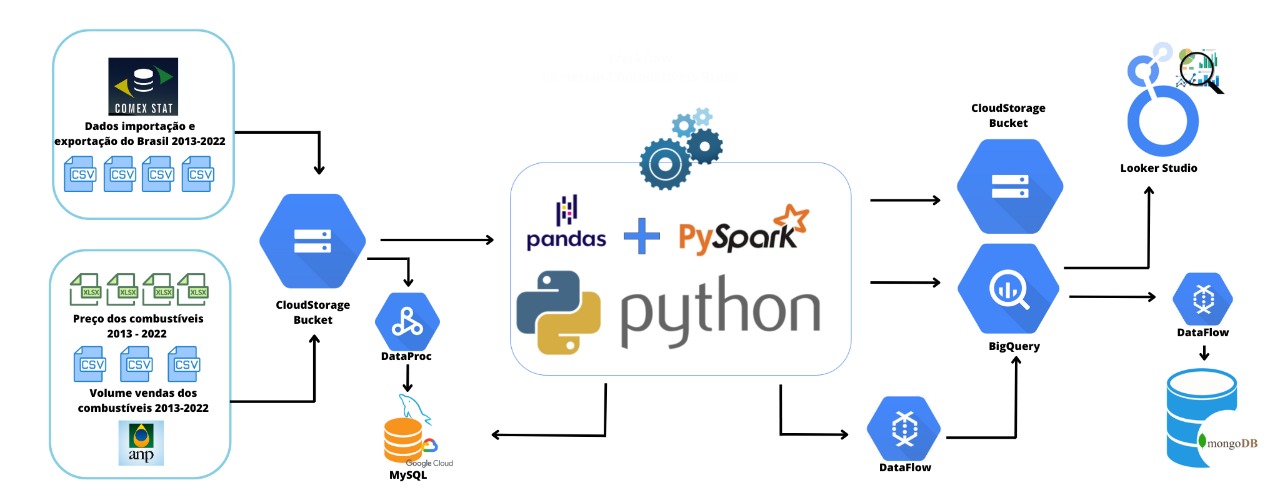
**Apresentação do projeto: Disponível em:**

[**https://datastudio.google.com/reporting/11fe3610-6344-4cfc-9b58-37b575baeee9**](https://datastudio.google.com/reporting/11fe3610-6344-4cfc-9b58-37b575baeee9)

* Slides da apresentação do projeto, acompanhado do Dashboard.

## Secção 5

**Tecnologias utilizadas no projeto**



**Secção 5**

### Subseção 5.1

**Linguagens da programação e bibliotecas:**

* SQL
* Python
* Bibliotecas Python:
* pandas
* pyspark
* numpy
* pandera
* gcsfs
* os
* request
* pymysql
* sqlalchemy
* apache\_beam
* google-cloud-storage
* google.oauth2
* pandas\_gbq
* pymongo

**Secção 5**

### Subseção 5.2

**Programação em nuvem GCP**

* Produtos GCP:
  + BigQuery
  + VM GCE no Colab Por meio do GCP Marketplace
  + Google Cloud Storage (GCS)
  + Instância SQL
  + DataProc
  + Dataflow
  + Looker Studio

**Secção 5**

Subseção 5.3

**MongoDB Atlas**

Criação de um Banco de Dados NO-SQL, utilizando a plataforma do MongoDB, para armazenar os dados tratados.

* Por possuir uma conta gratuita no serviço, não foi possível inserir todos os datasets normalizados. Sob a supervisão do gerente de projetos, senhor Bismarck William, optou-se por manter armazenada somente a tabela de preço de combustíveis, fornecida pela ANP e normalizada pelo time.

Secção 6

**Funcionalidades e Demonstração da Aplicação**

**Secção 6**

Subseção 6.1

**BigQuery**

|  |
| --- |
| #combinar importação por produto com classificação  CREATE OR REPLACE TABLE looker.imp\_prod AS  (SELECT i.ano, i.mes, i.produto, p.pais, i.estado, i.meio\_transporte, i.data,  i.valor\_dolar AS valor\_importado, i.peso\_liquido AS volume\_importado, c.sh6\_capitulo, c.sh6\_secao, c.cuci\_divisao, c.cuci\_secao, c.cgce\_n2, c.isic\_secao  FROM `bc26-ed7-projeto-final.comercio\_brasil.imp\_2013-2022\_tratada\_ncm` AS i  INNER JOIN `bc26-ed7-projeto-final.comercio\_brasil.imp-exp-classificacoes` AS c ON i.mercadoria = c.ncm  INNER JOIN `bc26-ed7-projeto-final.comercio\_brasil.nome\_pais` AS p ON i.cod\_pais = p.cod\_pais); |

|  |
| --- |
| #criar a tabela com o filtro para combustiveis para exportação  CREATE OR REPLACE TABLE looker.exp\_combustivel AS  (SELECT e.ano, e.mes, e.mercadoria AS produto, p.pais, e.estado, e.meio\_transporte, e.data, e.valor\_dolar AS valor\_exportado, e.peso\_liquido AS volume\_exportado  FROM `bc26-ed7-projeto-final.comercio\_brasil.exp\_2013-2022\_tratada\_ncm` AS e  INNER JOIN `bc26-ed7-projeto-final.comercio\_brasil.nome\_pais` AS p ON e.cod\_pais = p.cod\_pais  WHERE e.cod\_mercadoria = 27090010 or e.cod\_mercadoria =27101922 or e.cod\_mercadoria =99980101 or e.cod\_mercadoria =99981002 or e.cod\_mercadoria =27101259 or e.cod\_mercadoria =27101921 or e.cod\_mercadoria =27131100 or e.cod\_mercadoria =27131200 or e.cod\_mercadoria =27111100 or e.cod\_mercadoria = 27101929 or e.cod\_mercadoria =39111029 or e.cod\_mercadoria =27111910); |

|  |
| --- |
| #relação entre o meio de transporte o estado que exporta ordenado pelo volume total  SELECT meio\_transporte, estado,  SUM(valor\_dolar) AS valor\_total, AVG(valor\_dolar) AS valor\_medio,  SUM(peso\_liquido) AS volume\_total, AVG(peso\_liquido) AS volume\_medio  FROM `bc26-ed7-projeto-final.comercio\_brasil.exp\_2013-2022\_tratada\_ncm`  GROUP BY 1, 2 ORDER BY meio\_transporte ,volume\_total DESC; |

|  |
| --- |
| # 10 MENORES PREÇOS, SELECIONAR PRODUTO E ANO: gasolina comun no ano de 2022  SELECT data\_final as Data\_levantamento, estado, municipio,  preco\_medio\_revenda as Preco\_medio  FROM bc26-ed7-projeto-final.comercio\_brasil.preco\_combustiveis\_2013\_2022  WHERE produto = 'GASOLINA COMUM' AND  data\_final >= '2022-01-01 00:00:00' and data\_final < '2023-01-01 00:00:00'  ORDER BY preco\_medio\_revenda ASC LIMIT 10; |

|  |
| --- |
| #Preço mínimo, médio e máximo por ano e regiao dos protudos  SELECT extract(year FROM data\_final) AS ano, produto, regiao,  MIN(preco\_medio\_revenda) AS preco\_minimo,  AVG(preco\_medio\_revenda) AS preco\_medio, MAX(preco\_medio\_revenda) AS preco\_maximo  FROM bc26-ed7-projeto-final.comercio\_brasil.preco\_combustiveis\_2013\_2022  GROUP BY ano, produto, regiao ORDER BY ano DESC, preco\_medio; |

|  |
| --- |
| #os 50 maiores produtos valor por volume (dolar/volume) dos produtos nos 10 ano  SELECT produto, sum(valor\_dolar) as valor\_total,  sum(peso\_liquido) as volume\_total,  (sum(valor\_dolar)/sum(peso\_liquido)) AS dolar\_por\_volume  FROM `bc26-ed7-projeto-final.comercio\_brasil.imp\_2013-2022\_tratada\_ncm`  WHERE peso\_liquido !=0 GROUP BY produto ORDER BY dolar\_por\_volume DESC LIMIT 50; |

|  |
| --- |
| #as via de transporte com de acordo com o volume exportado em cada estado  SELECT meio\_transporte, estado,  SUM(valor\_dolar) AS valor\_total, AVG(valor\_dolar) AS valor\_medio,  SUM(peso\_liquido) AS volume\_total, AVG(peso\_liquido) AS volume\_medio  FROM `bc26-ed7-projeto-final.comercio\_brasil.exp\_2013-2022\_tratada\_ncm`  GROUP BY 1, 2 ORDER BY meio\_transporte, volume\_total DESC; |

|  |
| --- |
| #meses e anos ordenados pelo volume total importado - insight para gráfico  SELECT ano, mes, sum(valor\_dolar) as valor\_total, AVG(valor\_dolar) AS valor\_medio  FROM `bc26-ed7-projeto-final.comercio\_brasil.imp\_2013-2022\_tratada\_ncm` AS i  GROUP BY ano, mes ORDER BY valor\_total DESC; |

|  |
| --- |
| #valor total gasto em importação e o volume em relação a cada seção da Classificação Uniforme para Comercio Internacional em 2022  SELECT c.cuci\_secao, SUM(im.valor\_dolar) AS valor\_total, SUM(im.peso\_liquido) AS volume\_total  FROM `bc26-ed7-projeto-final.comercio\_brasil.imp\_2012-2022\_tratada\_ncm` AS im  INNER JOIN `bc26-ed7-projeto-final.comercio\_brasil.imp-exp-classificacoes` AS c  ON im.mercadoria=c.ncm  WHERE ano=2022 GROUP BY 1 ORDER BY valor\_total; |

Secção 7

Subseção 7.1

**Insights**

O Brasil entre os anos de 2013 e 2022 obteve o valor de:

* US$ 1,9 tri em total de importações;
* US$2,28 tri em total de exportações.

O número de exportações nos últimos 10 anos teve um aumento de 73.869.877.041 aproximadamente 31,7%. Enquanto o número de importação no mesmo período 453.462.225 aproximadamente 0,187%.

A partir de 2015 o saldo da balança comercial tornou-se positivo.

A China, os Estados Unidos e a Argentina representam quase de 50% do valor totas das importações e exportações realizadas nesses anos. Em relação ao volume, o Brasil exportou mais de 300 bilhões em quilograma-litro da China e importou acima de 35 bilhões de quilograma-litro dos Estados Unidos.

O principal meio de transporte usado no comércio exterior é o transporte de carga marítimo.

Combustíveis minerais e máquinas para a indústria são os produtos mais importados, enquanto as sementes e frutos oleaginosos juntos com combustíveis minerais são os produtos mais exportado. Observa-se que os

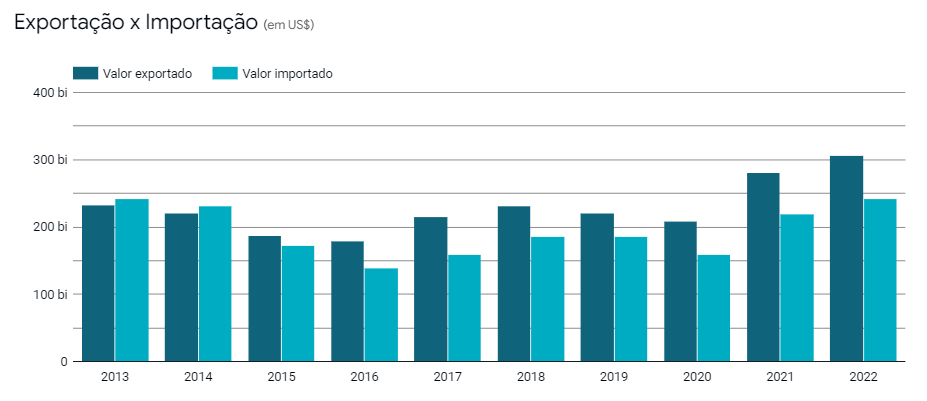
combustíveis minerais são tanto importados quando exportados, pois existem diferem tipos de combustível

com utilizações variadas.

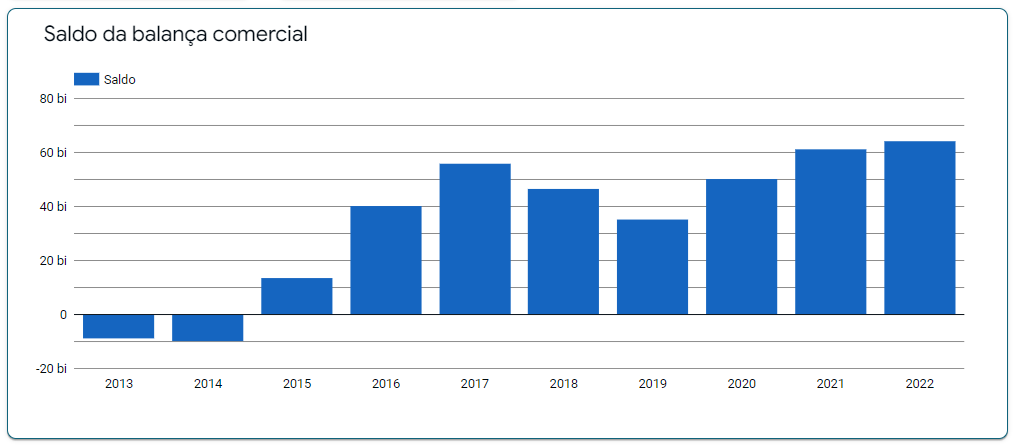
O estado do Rio de Janeiro é o estado que mais exporta combustíveis, mais de 100 trilhões de dólares em 10 anos, sendo $21,8 tri em 2022. Enquanto o estado de São Paulo é o estado que mais importa combustíveis, cerca de $60 trilhões nos 10 anos, e $12,3 tri em 2022.

Destacando, os óleos brutos, que a partir de 2016 obteve um saldo comercial positivo. O valor do óleo bruto na exportação, em dólar por quilograma-litro antes de até 2014 variava acima de 0,6, diminuindo nos anos seguintes para uma variação em 0,2 e 0,3 e voltando a subir nos anos de 2018 e 2019 e nos anos de 2021 e 2020 para acima de 0,4.

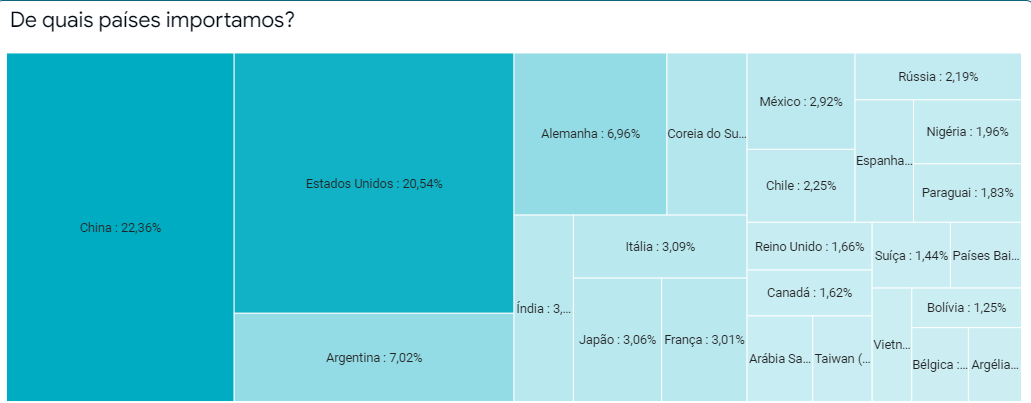
Pode-se ver que o preço da gasolina teve um aumento estável até 2020, passando então por diminuição e crescimento acelerado.

****

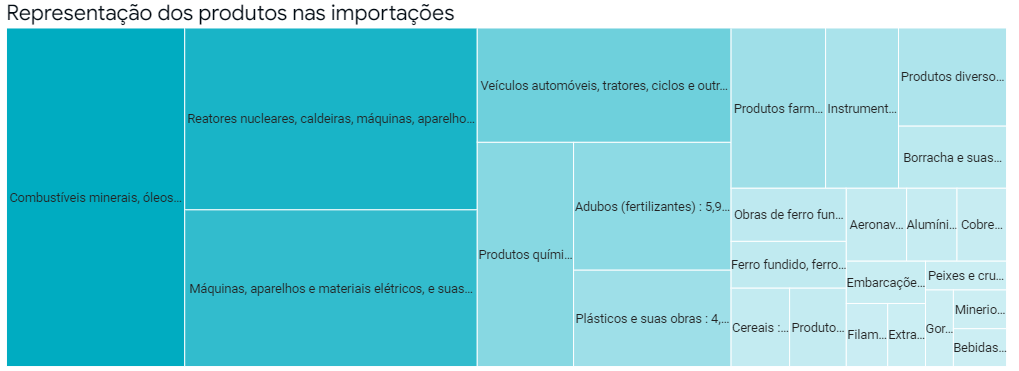
Saldo da balança comercial

****

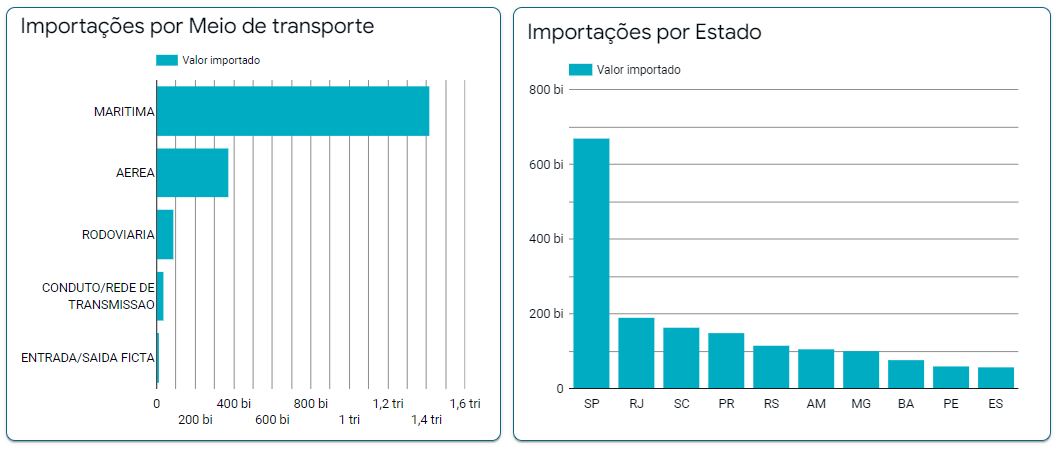
**Países de quem importamos**

****

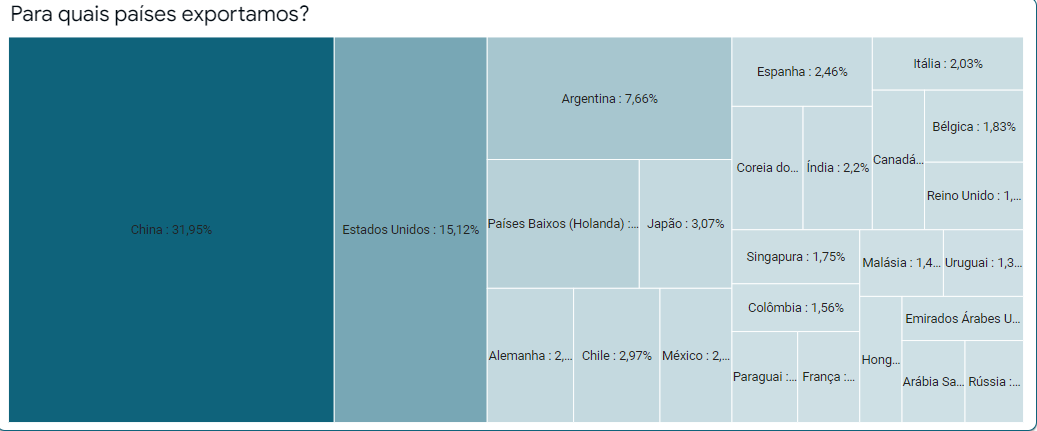
**Produtos importados**

****

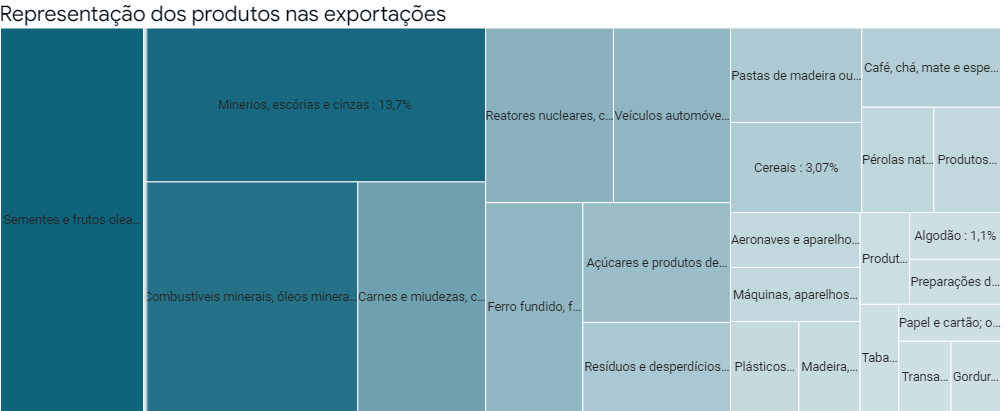
**Meio de transporte e estados importações**

****

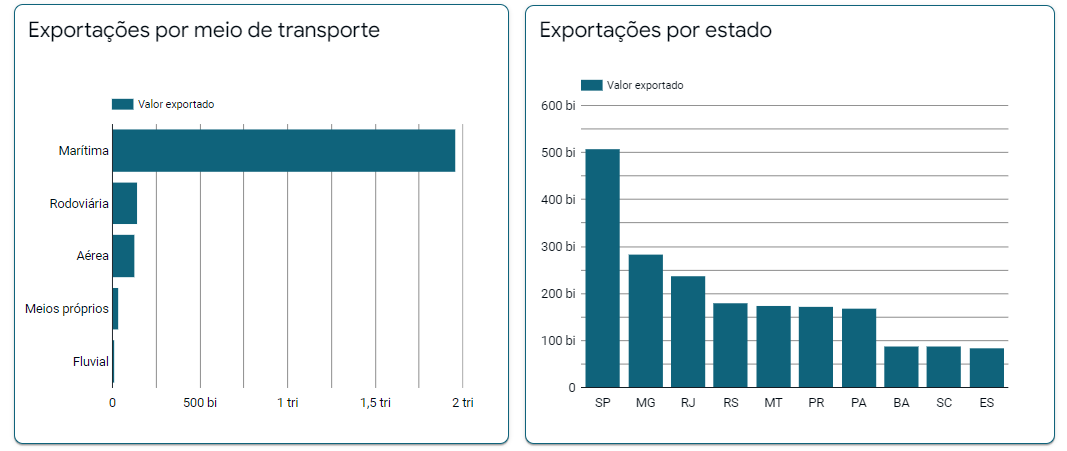
**Exportação países**

****

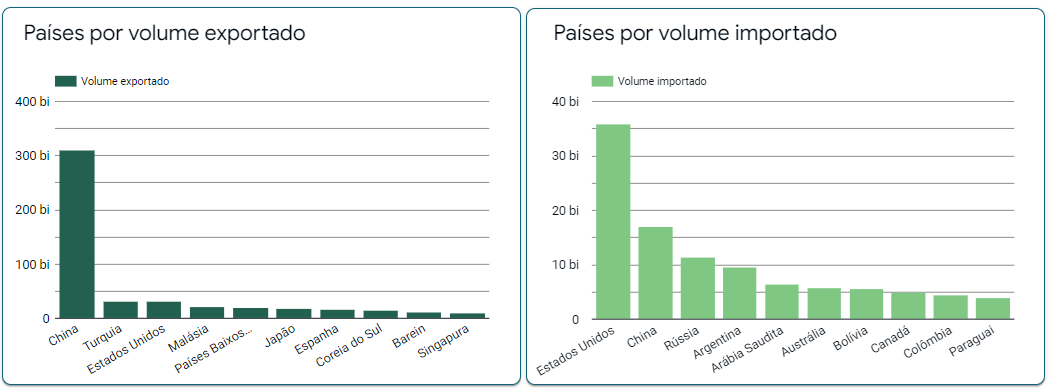
**Produtos nas exportações**

****

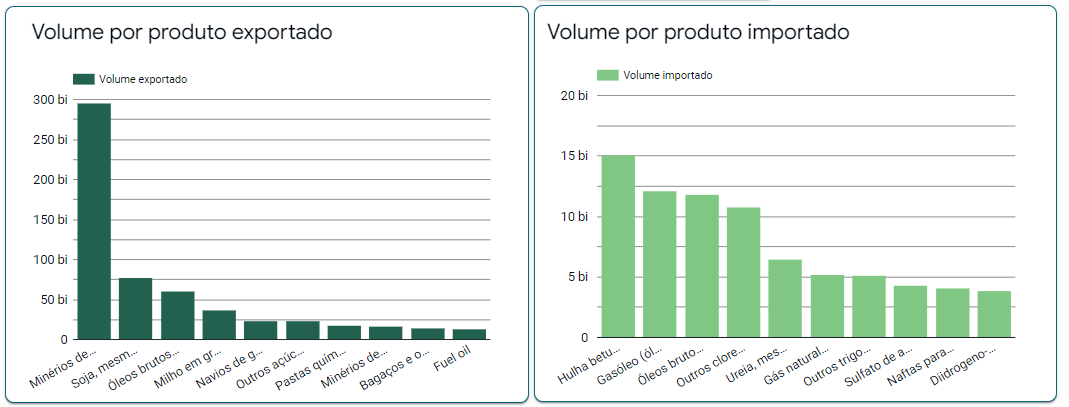
**Exportações transporte e estados**

****

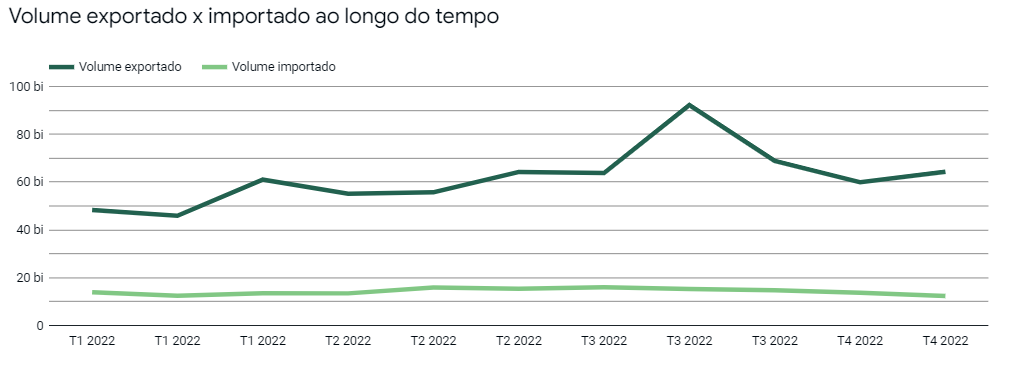
**Países por volume importado x exportado**

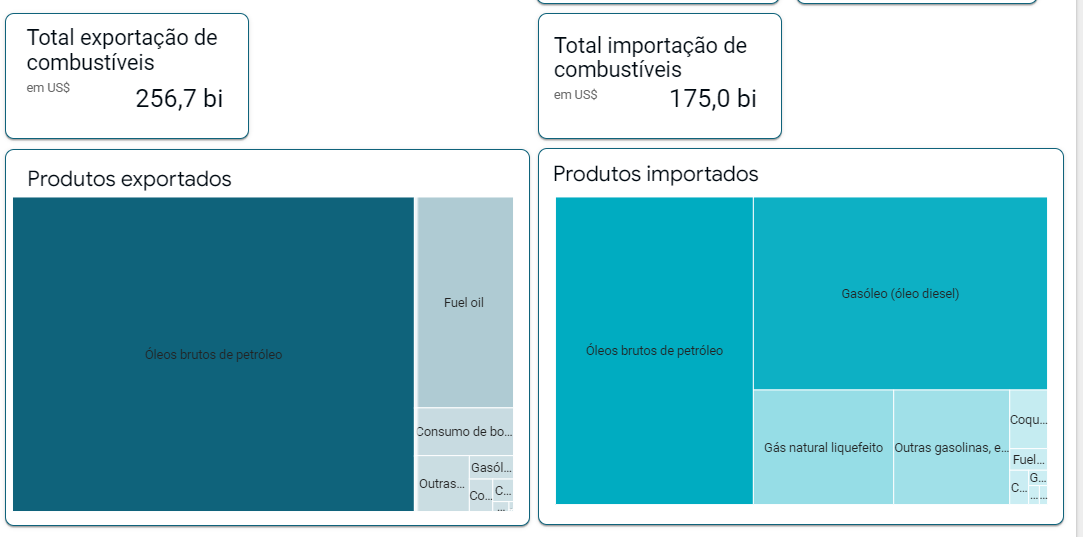
****

**Volume por produto exportado e importado**

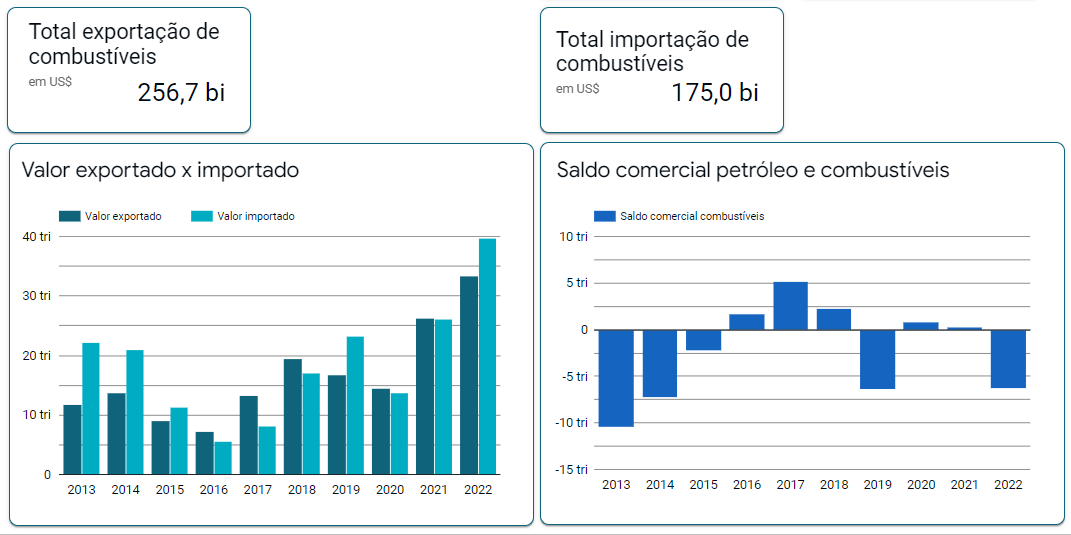
****

**Volume exportado e importado no ano 2022**

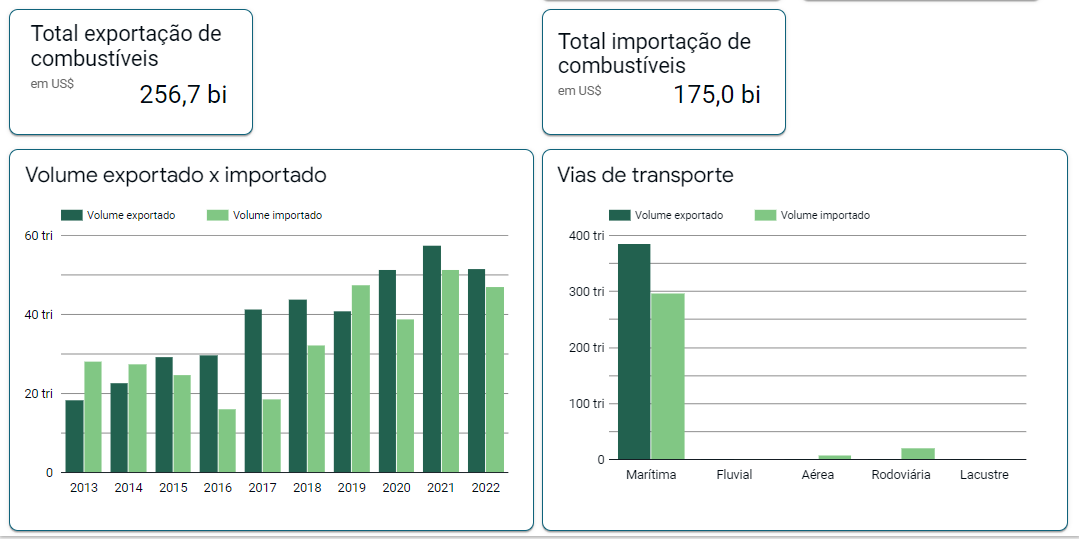




**Petróleo e combustíveis: Balança Comercial**



**Petróleo e combustíveis: Volume e meios de transporte**



## Secção 8

**Valor total do projeto (descritivo produtos e serviços)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Orçamento** | | |
| **Nº** | **Descritivo** | | **Período** | **Valor(R$)** |
|  | **Serviços próprio** | |  |  |
| **8.1** | **Serviço de consultoria** | |  | **18.000,00** |
|  | **Serviço terceirizado** | |  |  |
| **8.2** | **BigQuery** | |  | **0,00** |
| **8.3** | **VM GCE no Colab por meio do GCP Marketplace** | | **Único** | **756,30** |
| **8.4** | **Cloud Storage** | | **Único** | **1,85** |
| **8.5** | **Instância SQL** | | **Mensal** | **721,06** |
| **8.6** | **DataProc** | | **Único** | **1,69** |
| **8.7** | **DataFlow** | | **Único** | **1,11** |
|  |  | |  |  |
|  |  | |  |  |
|  | **Subtotal Serviços Próprios** | |  | **18.000,00** |
|  | **Subtotal Serviços Terceirizados** | |  | **1.527,38** |
|  | **Total** | |  | **19.527,38** |

**\*Cotação do dólar em 10-01-2023 R$5,30**

**\*Valor total de serviços terceirizado em dólar: $288,17**

**Secção 8**

### Subseção 8.1

**Consultoria**

* Integração de sistemas
* Extração e análise de dados
* Dashboard e insights
* Automação
* Entrega dos dados brutos e tratados

**Secção 8**

### Subseção 8.2

**BigQuery**

* Valor Total: Gratuito - Período: Mensal
* 10 GB de armazenamento
* 1 TB consultas/Análise
* Ingestão de dados (carregamento de lote)
* Extração de dados

**Secção 8**

### Subseção 8.3

**VM GCE no Colab por meio do GCP Marketplace**

* Sub-Total:R$151,26 - Período: 3 dias - Quantidade: 5 Total: R$756,30
* Instância de VM, 2 vCPUs + 13 GB (n1-hightmem-2)
* Disco de estado sólido 200 GB
* GPU NVIDIA T4

**Secção 8**

### Subseção 8.4

**Cloud Storage**

* Valor Total:$0,35 - Previsão: Mensal
* 10 GB – Armazenamento padrão
* América do Sul (southamerica-west1)

**Secção 8**

### Subseção 8.5

**Instância SQL**

* Valor Total: $136,05 Previsão: Mensal
* América do Sul (southamerica-west1)
* vCPUs - $45,23
* Memória -$7,67 (GB) = $76,67(+ - 10 GB)
* Armazenamento -$0,25 (GB) = $2,50 (+- 10 GB)
* Rede: entrada: grátis
  + Saída: $11,50
* Exportação sem servidor: $0,015 (GB) = $0,15 (+-10 GB)

**Secção 8**

Subseção 8.6

**DataProc**

* Valor Total: $0,32 Tempo: 2h
* N1-highmem-16
* 16 CPUs
* 104 GB memoria
* Dataproc charge = # of vCPUs \* hours \* Dataproc price
* Dtaproc = 16 \* 2 \* $0.01 = $0.32
* VPC exclusiva: gratuito

**Secção 8**

### Subseção 8.7

**DataFlow**

* Valor Total: $0,21 Tempo: 3h
* US-CENTRAL-1
* Job-lote : $0,071(h) = $0,213
* Armazenamento SSD: $0,000298 (GB/h) =0,006258
* Recursos complementares (BigQuery, Cloud Storage) já inclusos

## Secção 9

**Serviços Optativos – Adicionais**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Serviços Optativos | | | | |
| Nº | | Descritivo | Período | Valor(R$) |
| 9.1 | | Instância SQL | Mensal | 721,06 |
| 9.2 | | MongoDB Atlas | Mensal | 725,04 |
|  | | Total |  | 1.446,10 |

**Secção 9**

### Subseção 9.1

**Instância SQL – Para datasets normalizados**

* Valor Total: $136,05 Previsão: Mensal
* America do Sul(southamerica-west1)
* vCPUs - $45,23
* Memoria -$7,67 (GB) = $76,67(+ - 10 GB)
* Armazenamento -$0,25 (GB) = $2,50 (+- 10 GB)
* Rede: entrada: grátis
  + Saída: $11,50
* Exportação sem servidor: $0,015 (GB) = $0,15 (+-10 GB)

**Secção 9**

### Subseção 9.2

**MongoDB Atlas – Para datasets originais**

* Cluster dedicated – Google Cloud Platform
* Valor Total: $136,80 Previsão: Mensal
* M20M20 (cluster)
* 20GB (armazenamento)
* 3.75 GB 4 GB
* 1vCPUs 2vCPUs ($0,19/h)

## Secção 10

**Pessoas Desenvolvedoras do Projeto**

**Secção 10**

### Subseção 10.1

**Ana Flavia Dias**

Linkedin: www.linkedin.com/in/ana-dias-902559155/

e-mail: amfd88@gmail.com

Github: github.com/anafla7

**Secção 10**

### Subseção 10.2

**Ariana Caetano**

Linkedin: www.linkedin.com/in/ariana-caetano-66141a22/

e-mail: [ariana.caetano@gmail.com](mailto:ariana.caetano@gmail.com)

Github: github.com/ariana-caetano

**Secção 10**

### Subseção 10.3

**Érika Cunha**

Linkedin: www.linkedin.com/in/erikacunha024/

e-mail: [erikancunha024@gmail.com](mailto:erikancunha024@gmail.com)

Github: https://github.com/Erikancunha024

**Secção 10**

### Subseção 10.4

**José Jardim**

Linkedin: www.linkedin.com/in/jos%C3%A9-jardim-pardillos-vieira/

e-mail: [jo.jjpv@gmail.com](mailto:jo.jjpv@gmail.com)

Github: https://github.com/JoseJardimPV

**Secção 10**

### Subseção 10.5

**Marcos Valente Carvalho da Silva**

Linkedin: www.linkedin.com/in/marcosvalentec/

e-mail: [marcos.valente.c@gmail.com](mailto:marcos.valente.c@gmail.com)

Github: https://github.com/MarcosVCS