



A COEXP é uma consultoria de inteligência voltada a atender empresas estrangeiras que vislumbram fazer negócios no Brasil. Focada em disponibilizar e analisar dados confiáveis, seus produtos procuram facilitar a tomada de decisões de grandes stakeholders internacionais.

- **Stakeholder**
 - EPOC - Empresa Portuguesa de Óleos e Combustíveis
- **Projeto**
 - Engenharia e Análise de dados do Comércio Brasileiro
- **Objetivo**
 - Dados confiáveis para embasar estudos de viabilidade



Ana Flávia Dias



Ariana Caetano



Erika Cunha



José Jardim



Marcos Valente

- **Dados**
- Comércio Exterior (Exportação e Importação) de petróleo e combustíveis automotivos;
 - Volume de combustíveis automotivos (incluindo etanol hidratado) vendidos no Brasil;
 - Histórico de preços pagos pelo consumidor final.



➤ Utilização de infraestrutura e ferramentas Google Cloud Platform

- Disponibilização dos dados originais em Data Lake (Cloud Storage) e MySQL;
- Disponibilização dos dados tratados em Cloud Storage, BigQuery e MongoDB;
- Análise dos dados em BigQuery e Looker Studio.



➤ Análises

- Visualização das importações e exportações brasileiras;
- Visualizar o impacto das crises da COVID-19 e da Guerra da Ucrânia sobre a balança comercial;
- Verificar quais os países são maiores parceiros comerciais do Brasil;
- Visualizar os maiores exportadores e importadores de petróleo e combustíveis;
- Visualizar como as regiões e sub-regiões do Brasil importam e exportam petróleo e combustíveis;
- Verificar qual meio de transporte é mais utilizado nas importações e exportações desses bens.





Google Cloud

- Criação de Projeto na organização COEXP para o cliente EPOC e configurar os papéis de acesso IAM.



IAM



Cloud
Storage



Cloud SQL



BigQuery



mongoDB®

- Criação de conta para o cliente EPOC.

Linguagens de programação

- SQL, principalmente DQL
- Python e as bibliotecas:
 - pandas ○ apache_beam
 - pyspark ○ google.oauth2
 - numpy ○ os
 - pandera ○ request
 - sqlalchemy ○ gcsfs
 - pymysql ○ google-cloud-storage
 - pymongo ○ pandas_gbq



BigQuery



DataProc



Colab



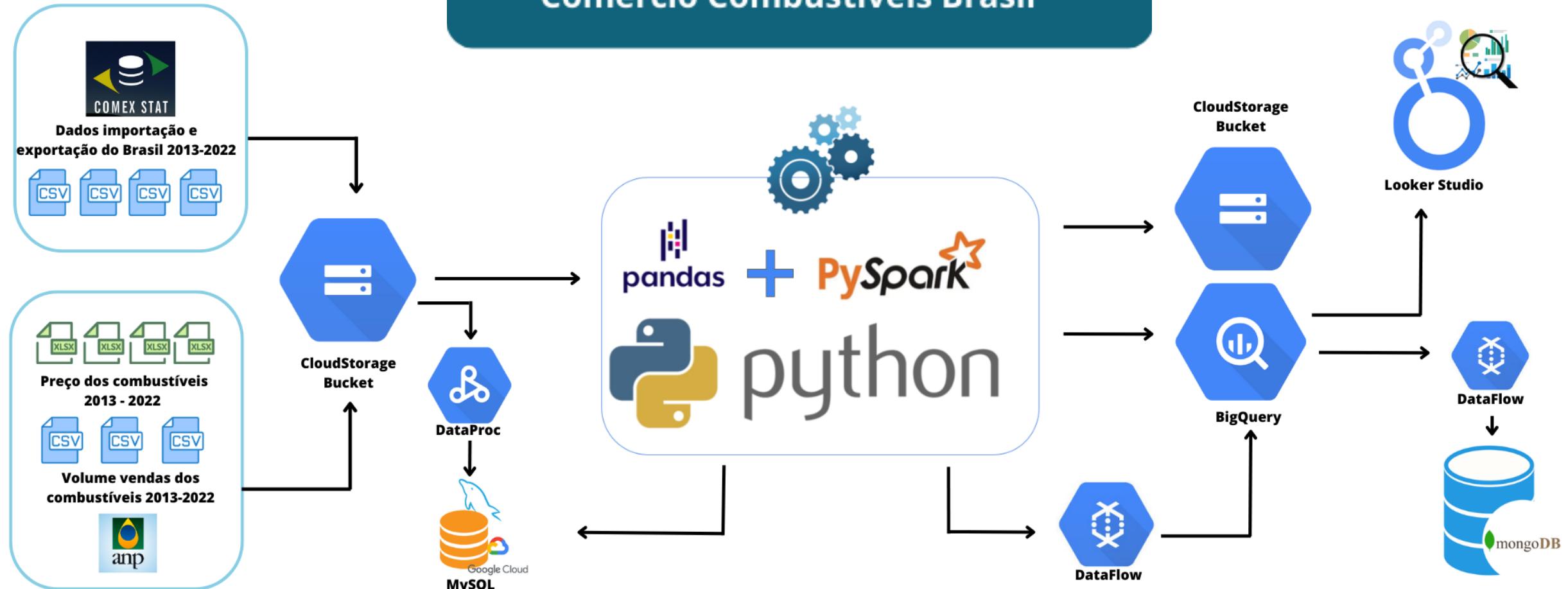
Google Cloud
Marketplace

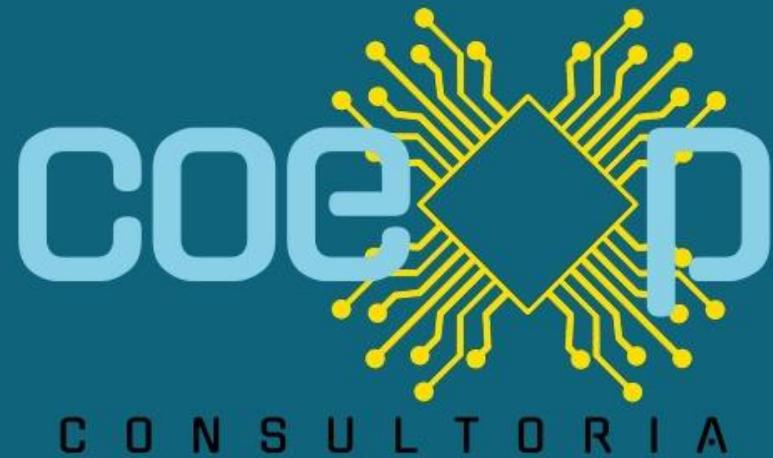


Dataflow

Workflow

Comércio Combustíveis Brasil





Fase 1: Seleção dos datasets e disponibilização



Ministério da Economia
Sec. Esp. de Comércio Exterior



Disponibilizadas pelos meios eletrônicos oficiais do Governo Federal.

Estatísticas de Comércio Exterior em Dados Abertos



Ministério da Economia

Sec. Esp. de Comércio Exterior

Quatro datasets principais:

- Exportação por NCM
- Importação por NCM
- Exportação por Município
- Importação por Município

>> Corte temporal: 2012 - 2022
>> Totalizando 56,929,067 linhas
>> Com referências a 16 tabelas de Correlações de Códigos e Classificações

NMC: Nomenclatura Comum do Mercosul



Lançamentos
Produtos
Data
Origem no Brasil
Destino
Valor
Volume
Meio de transporte



Preços de Revenda e de Distribuição de Combustíveis

Levantamento por municípios (semanal)

Dividido em três dataframes (2013 a 2017; 2018 a 2021; 2022)

>> Totalizando 1.190.920 linhas



Semana levantamento
Município
Produto
Postos pesquisados
Preço de revenda (maior,
menor e médio)
Dados estatísticos

Vendas de derivados de petróleo e biocombustíveis

- Vendas de combustíveis por segmento (metros cúbicos) 2012-2022
- Vendas de GLP por tipo de vasilhame (metros cúbicos) 2007-2022
- Vendas de óleo diesel por tipo (metros cúbicos) 2013-2022



>> Disponibilizadas por mês
>> Totalizando 9.440 linhas



Data (mês e ano)
Unidade Federativa
Produto
Tipo de venda
Quantidade de vendas

Armazenamento dos datasets

➤ **Datasets originais**

- Google Cloud Storage
- MySQL



Google Cloud Storage



Cloud SQL



Google BigQuery



➤ **Datasets normalizados:**

- Google Cloud Storage
- MongoDB
- BigQuery

➤ **Configuração das ferramentas de armazenamento e das chaves de acesso**



Google BigQuery



Cloud SQL



Google Cloud Storage



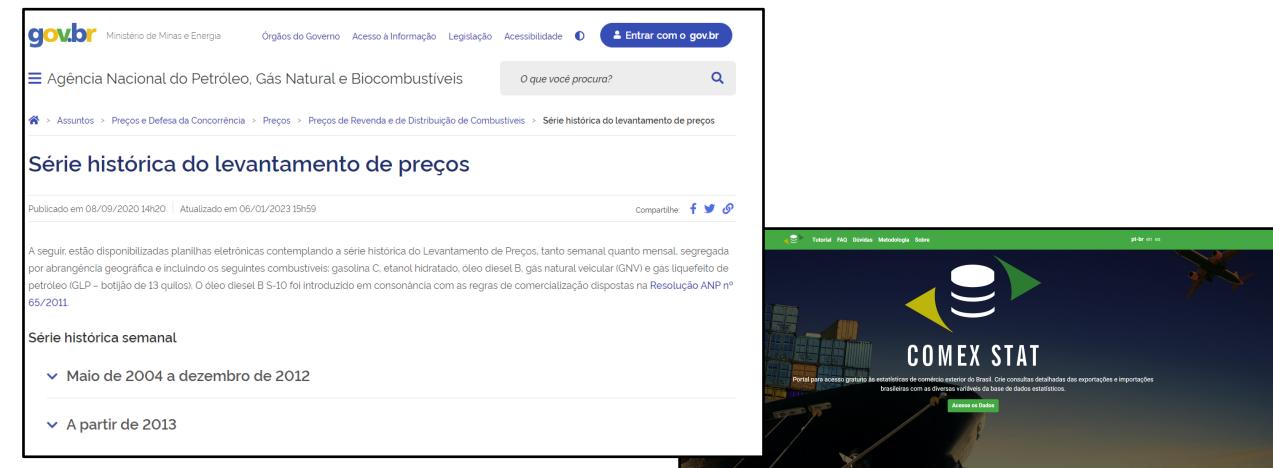
Extração dos datasets

```
1 import requests
```

+ Código + Texto

```
1 link_comb = 'https://www.gov.br/anp/pt-br/centrais-de-conteudo/dados-abertos/arquivos/vdpb/vcs/vendas-combustiveis-segmento-m3-2012-2022.csv'
2
3 link_glp = 'https://www.gov.br/anp/pt-br/centrais-de-conteudo/dados-abertos/arquivos/vdpb/vct/vendas-glp-tipo-vasilhame-m3-2007-2022.csv'
4
5 link_oleo = 'https://www.gov.br/anp/pt-br/centrais-de-conteudo/dados-abertos/arquivos/vdpb/vct/vendas-oleo-diesel-tipo-m3-2013-2022.csv'
6
```

```
1 #pegar o conteúdo do link
2 request_text = requests.get(link_comb).content
3 #enviar o conteúdo para o bucket
4 bucket.blob('comex-original-link/vendas-combustiveis-segmento-m3-2012-2022').upload_from_string(request_text, 'csv')
```



The screenshot shows a web page from the Brazilian National Agency for Petroleum, Natural Gas and Biofuels (ANP). The URL is <https://www.gov.br/anp/pt-br/centrais-de-conteudo/dados-abertos/arquivos/vdpb/vct/vendas-combustiveis-segmento-m3-2012-2022.csv>. The page title is "Série histórica do levantamento de preços". It displays a table with historical price data for different fuel types. The table includes columns for 'Preço' (Price), 'Unidade' (Unit), 'Mês' (Month), and 'Ano' (Year). The data spans from May 2004 to December 2012, with a note indicating that data from January 2013 onwards is available. The page also features a sidebar with links to other ANP services like 'Tutorial', 'FAQ', 'Divulgação', 'Monitórgia', and 'Sobre'.

Armazenamento dos datasets originais

➤ Envio para o Google Cloud Storage (bucket)

- Seleção dos 10 últimos anos (2013 - 2022)
- Otimização do processo de tratamento

```
[ ] 1 # criação do caminho dentro da bucket (que já está acessada) para salvar o parquet
[ ] 2 path = 'gs://bc26-datasets-comex/IMP_2012-2022 (MUN).parquet'

[ ] 1 # SALVAMENTO DO ARQUIVO NO FORMATO PARQUET DENTRO DA BUCKET
[ ] 2 df1.to_parquet(path, index=False, partition_cols=['CO_ANO'])

[ ] 1 # COMPROVAÇÃO DA EFICIÊNCIA DO FORMATO PARQUET:
[ ] 2 # O DF QUE LEVAVA APROXIMADAMENTE 3 MINUTOS PARA SER MONTADO A PARTIR DO CSV
[ ] 3 # AGORA É MONTADO PELO PANDAS EM APROXIMADAMENTE 10 SEGUNDOS
[ ] 4 # (ganho de eficiência nada negligenciável)
[ ] 5 df1 = pd.read_parquet(path)
```



Carregamento dos datasets originais

➤ Envio para o servidor MySQL

- Seleção dos 10 últimos anos (2013 - 2022)
- Carregamento pelo colab para o servidor
- Carregamento com uso do DataProc



Cloud Dataproc

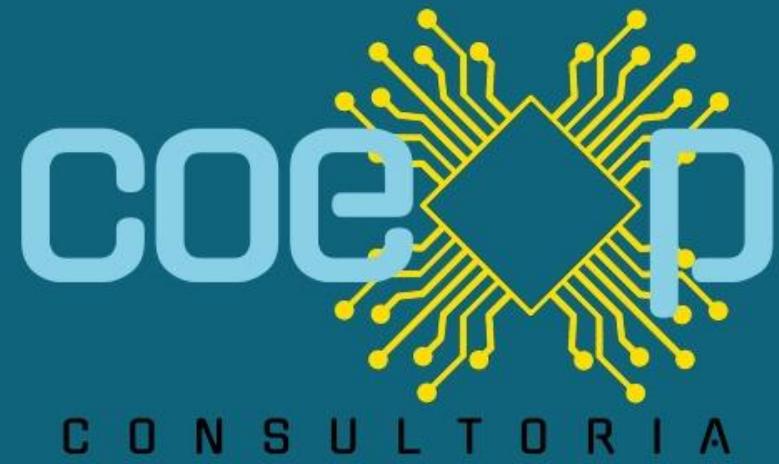


Cloud SQL

Envio do dataframe com os dados de combustíveis do ano de 2018 a 2021

```
1 #envio do df de combustíveis 2018 a 2021 para a tabela combustivel_2018_2021
2 try:
3     dfsql = df_comb_2018_2021.to_sql('combustivel_2018_2021', connection_comb , if_exists='replace', index = False)
4 except:
5     print('Erro!')
6 else:
7     print('Tabela criada com sucesso');
```

Tabela criada com sucesso



Fase 2: Tratamento dos dados

Normalização dos dados



- ✓ Verificação das duplicatas
- ✓ Renomear as colunas
- ✓ Remoção de espaços
- ✓ Exclusão das colunas frete e valor
- ✓ Substituição valores 0 por NAN

- ✓ Substituição dos códigos(valores) através de join e merge com a tabela referência
- ✓ Correção do tipos/formatos de dados por coluna
- ✓ Validação por Pandera
- ✓ Salvar em Parquet

➤ Substituição de códigos pelo seu real significado (nome)

- Manualmente (método replace);
- “Merge” com as tabelas referências

```
df1['ncm_unidade'].value_counts(ascending=True).head(50) #13 valores
```

14	21	Metro	21
18	36	Mil quilowatt hora	36
22	893	Grama líquido	893
12	4624	Milheiro	4624
20	12903	Dúzia	12903
19	14052	Quilate	14052
17	92185	Litro	92185
15	147096	Metro quadrado	147096
21	150677	Tonelada métrica líquida	150677
16	150771	Metro cúbico	150771
13	229814	Pares	229814
11	4163069	Número unidade	4163069
10	8181704	Quilograma líquido	8181704
		Name: ncm_unidade, dtype: int64	

antes

depois

➤ Remoção de dados duplicados

#Verificando df_ncm_sh4		
df_ncm_sh4.head()		
	CO_SH4	NO_SH4_POR
0	101	Cavalos, asininos e muares, vivos
1	101	Cavalos, asininos e muares, vivos
2	101	Cavalos, asininos e muares, vivos
3	101	Cavalos, asininos e muares, vivos
4	101	Cavalos, asininos e muares, vivos

```
#Eliminando códigos repetidos do df_pais #5324 repetidos
df_ncm_sh4= df_ncm_sh4.drop_duplicates(subset='CO_SH4', keep='first')
```

➤ “Merge” de dados

```
#Unindo com a descrição coluna mercadoria
df2= pd.merge(df2,df_ncm_sh4, left_on='ncm_mercadoria',right_on='CO_SH4')
```

```
df2.head(2)
```

	ano	mes	ncm_mercadoria	estado	peso_liquido	valor_dolar	país	municipio	CO_SH4	NO_SH4_POR
0	2012	Setembro	8803	SP	10195	3960721	Europa	São José dos Campos	8803	Partes dos veículos e aparelhos das posições 8...
1	2012	Julho	8803	SP	10145	3931274	Europa	São José dos Campos	8803	Partes dos veículos e aparelhos das posições 8...

Normalização dos dados

Antes

df1.head()											
	CO_ANO	CO_MES	CO_NCM	CO_UNID	CO_PAIS	SG_UF_NCM	CO_VIA	CO_URF	QT_ESTAT	KG_LIQUIDO	VL_FOB
12602064	2012	6	87089100	11	97	PR	7	145200	90	1194	21925
12602065	2012	6	33051000	10	249	SP	1	817800	3763	3763	17164
12602066	2012	6	94035000	11	628	RS	1	1017700	19576	561623	1490086
12602067	2012	6	85371090	10	589	PR	4	817700	2063	2063	84614
12602068	2012	6	83012000	10	158	SP	7	1010900	352	352	14720

Depois

#validador do DF schema1.validate(df1)												
	ano	mes	ncm_mercadoria	ncm_unidade	estado	meio_transporte	qnt_estatistica	peso_liquido	valor_dolar	pais	urf_localizacao	mercadoria
0	2012	Junho	87089100	Número unidade	PR	Rodoviária	9	1194	21925.0	América do Sul	0145200 - CORUMBA	Radiadores para tratores e veículos automóveis
1	2012	Novembro	87089100	Número unidade	PR	Rodoviária	12	314	39258.0	América do Sul	0145200 - CORUMBA	Radiadores para tratores e veículos automóveis
2	2012	Junho	87089100	Número unidade	não declarada	Rodoviária	2	15	262.0	América do Sul	0145200 - CORUMBA	Radiadores para tratores e veículos automóveis
3	2012	Fevereiro	87089100	Número unidade	SP	Rodoviária	18	211	4405.0	América do Sul	0145200 - CORUMBA	Radiadores para tratores e veículos automóveis
4	2012	Novembro	87089100	Número unidade	SP	Rodoviária	81	82	1389.0	América do Sul	0145200 - CORUMBA	Radiadores para tratores e veículos automóveis

Normalização dos dados



```
# Análise das tabelas contendo números tipo float

df_numeros = dfs.select('preco_medio_revenda', 'desvio_padrao_revenda', 'preco_min_revenda', 'preco_max_revenda', 'margem_media_revenda', 'coef_variacao_revenda')

df_numeros.show()

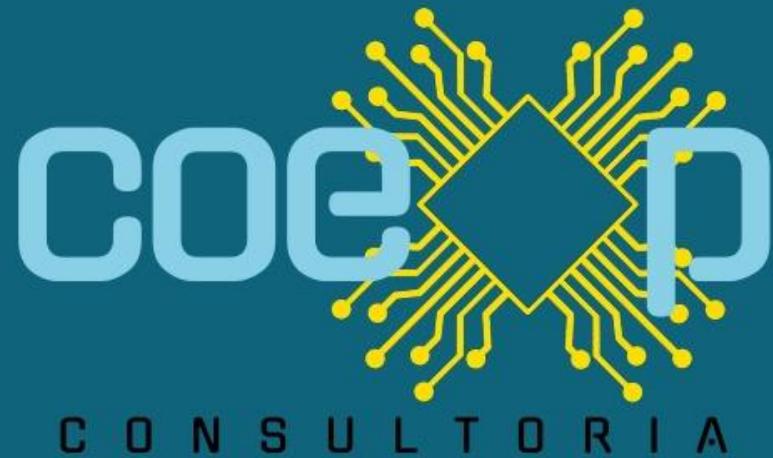
+-----+-----+-----+-----+-----+-----+
|preco_medio_revenda|desvio_padrao_revenda|preco_min_revenda|preco_max_revenda|margem_media_revenda|coef_variacao_revenda|
+-----+-----+-----+-----+-----+-----+
|      2.273|     0.022864401|        2.19|       2.299|        0.405|        0.01|
|      2.134|     0.043358967|        2.09|       2.19|        0.268|        0.02|
|      1.865|     0.054574985|        1.69|       1.899|        0.36|        0.029|
|      1.771|     0.02253203|        1.73|       1.79|        0.206|        0.013|
|      1.918|     0.10539576|        1.65|       1.999|        0.376|        0.055|
|      1.955|     0.044013128|        1.89|       1.999|        0.401|        0.023|
|      1.939|     0.08711017|        1.85|       2.089|        0.411|        0.045|
|      1.847|     0.070510305|        1.67|       1.99|        0.201|        0.038|
|      1.81|     0.069522895|        1.75|       1.99|        NaN|        0.038|
+-----+-----+-----+-----+-----+-----+
```



```
# Alterar dado 'R$/13Kg' para 'R$/13kg'  
dfs = dfs.withColumn('unidade_medida', F regexp_replace(F.col('unidade_medida'), 'R$/13Kg', 'R$/13kg'))  
  
# Alterar dado 'R$/m³' para 'R$/m3'  
dfs = dfs.withColumn('unidade_medida', F regexp_replace(F.col('unidade_medida'), 'R$/m³', 'R$/m3'))
```

```
# Arredondar valores apresentados na coluna 'desvio_padrao_revenda' (três casas decimais)  
dfs = dfs.withColumn('desvio_padrao_revenda', F.round(F.col('desvio_padrao_revenda'), 3))
```





Fase 3: Disponibilização dos datasets normalizados

Armazenamento dos datasets normalizados

- **Envio para o Data Lake**
(Google Cloud Storage)



Google Cloud Storage

- **Envio para o BigQuery**
 - Envio usando Google Colaboratory
 - Envio usando Pipeline: modelo criado no Apache Beam e executado no Dataflow



Google Dataflow



Google BigQuery

```
1 from pymongo import MongoClient
2
3 uri = "mongodb+srv://comerciobr.nslqhu7.mongodb.net/?authSource=%24external&authMechanism=MONGODB-X509&retryWrites=true&w=majority"
4 client = MongoClient(uri,
5                      tls=True,
6                      tlsCertificateKeyFile='/content/X509-cert-800051816302881169.pem')
7
8 #criação do database
9 db = client['comercio-br']
10 #criação da collection
11 collection = db['volume-preco-combustiveis-mensal-2013-2022']
12 doc_count = collection.count_documents({})
13 print(doc_count)
```

```
1 #envio dos dados
2 df_dict = df_combustivel.to_dict("records")
3
4 collection.insert_many(df_dict)
```

```
1 doc_count = collection.count_documents({})
2 print(doc_count)
```

1190920

-  **Dataflow**
-  Overview
-  **Jobs**
-  Pipelines
-  Workbench
-  Snapshots
-  SQL Workspace

[←](#) Create job from template

Job name * preco-combustiveis-bigquery-mongodb
 Must be unique among running jobs

Regional endpoint * us-west1 (Oregon)

Choose a Dataflow regional endpoint to deploy worker instances and store job metadata. You can optionally deploy worker instances to any available Google Cloud region or zone by using the worker region or worker zone parameters. Job metadata is always stored in the Dataflow regional endpoint. [Learn more](#)

Dataflow template * BigQuery to MongoDB

A batch pipeline which reads data rows from BigQuery and writes them to MongoDB as documents.

Required parameters

MongoDB Connection URI * mongodb+srv://comercio_br_sc:pLhq71cvjdeHIYUT@comerciobr.nslqhu7.mongodb.net

URI to connect to MongoDB Atlas.

Mongo database * comercio-combustiveis

Database in MongoDB to store the collection from. Example: my-db.

Mongo collection * preco-combustiveis-2013-2022

Name of the collection inside MongoDB database. Example: my-collection.

BigQuery source table * bc26-ed7-projeto-final.comercio_brasil.preco_combustiveis_2013_2022

BigQuery source table spec. Example: bigquery-project.dataset.input_table.

Encryption

Google-managed encryption key
 No configuration required

Customer-managed encryption key (CMEK)
 Manage via Google Cloud Key Management Service



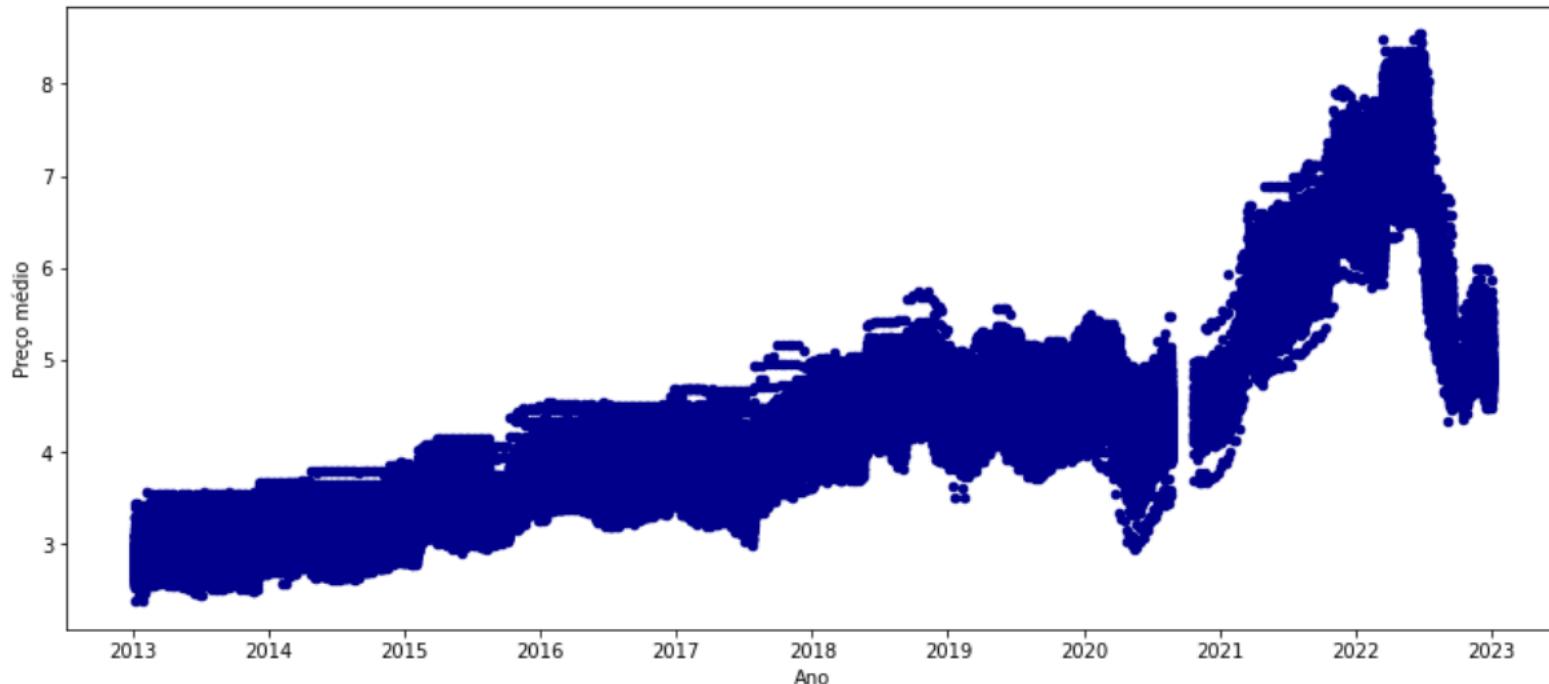
Fase 4: Análise dos datasets normalizados

Análise dos datasets normalizados

```
#Tendência na variação do preço médio da gasolina comum no Brasil  
#Dados serão melhor explorados no Looker Studio  
  
filtro_gasolina = df_plot.produto == 'GASOLINA COMUM'  
df_gasol = df_plot.loc[filtro_gasolina]  
  
df_gasol.plot.scatter(x = 'data_final', y = 'preco_medio_revenda', c = 'DarkBlue', xlabel = 'Ano', ylabel = 'Preço médio', figsize = (14,6))
```



Dataset: Preço Combustíveis



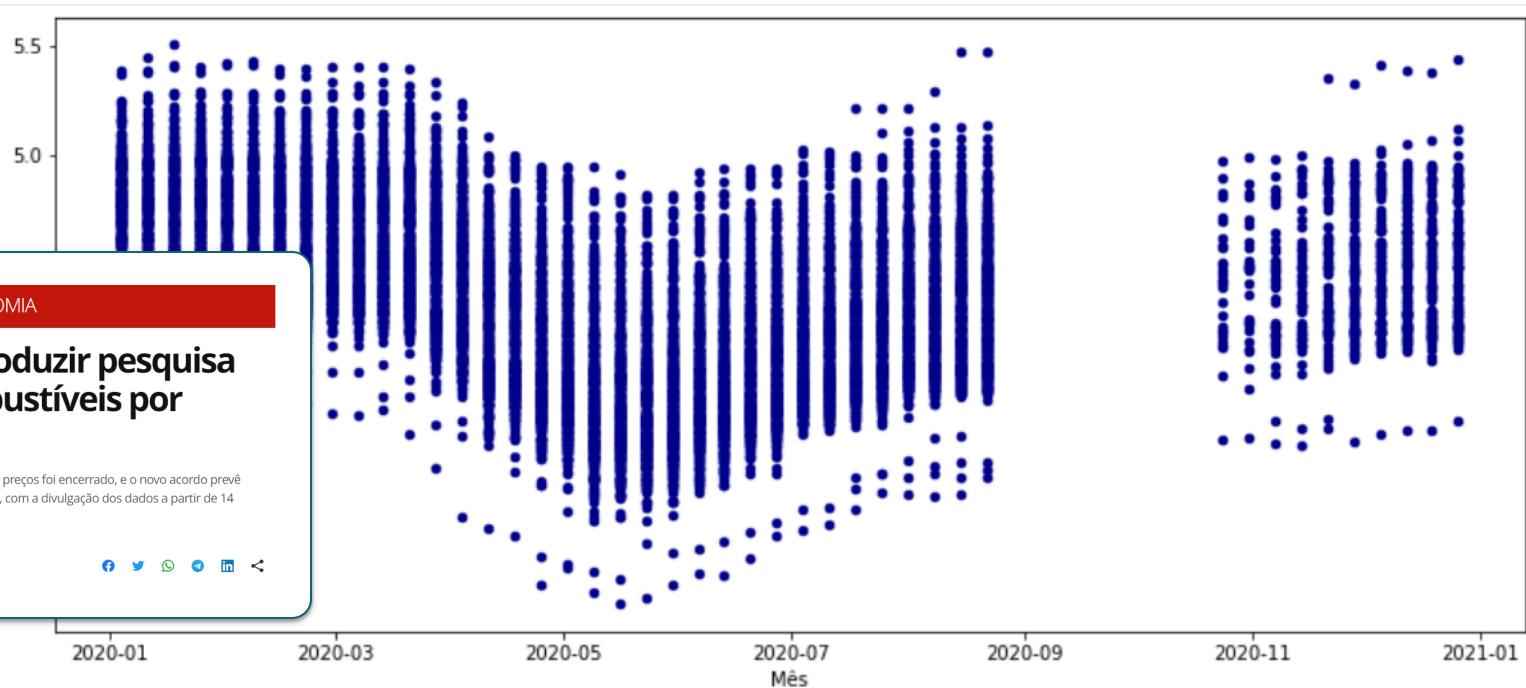
Análise dos datasets normalizados

```
# Percebeu-se um gap de dados no dataframe no terceiro semestre de 2020

filtro_ano = df_plot.data_final.dt.year == 2020

df_gasol_ano = df_plot.loc[filtro_ano & filtro_gasolina]

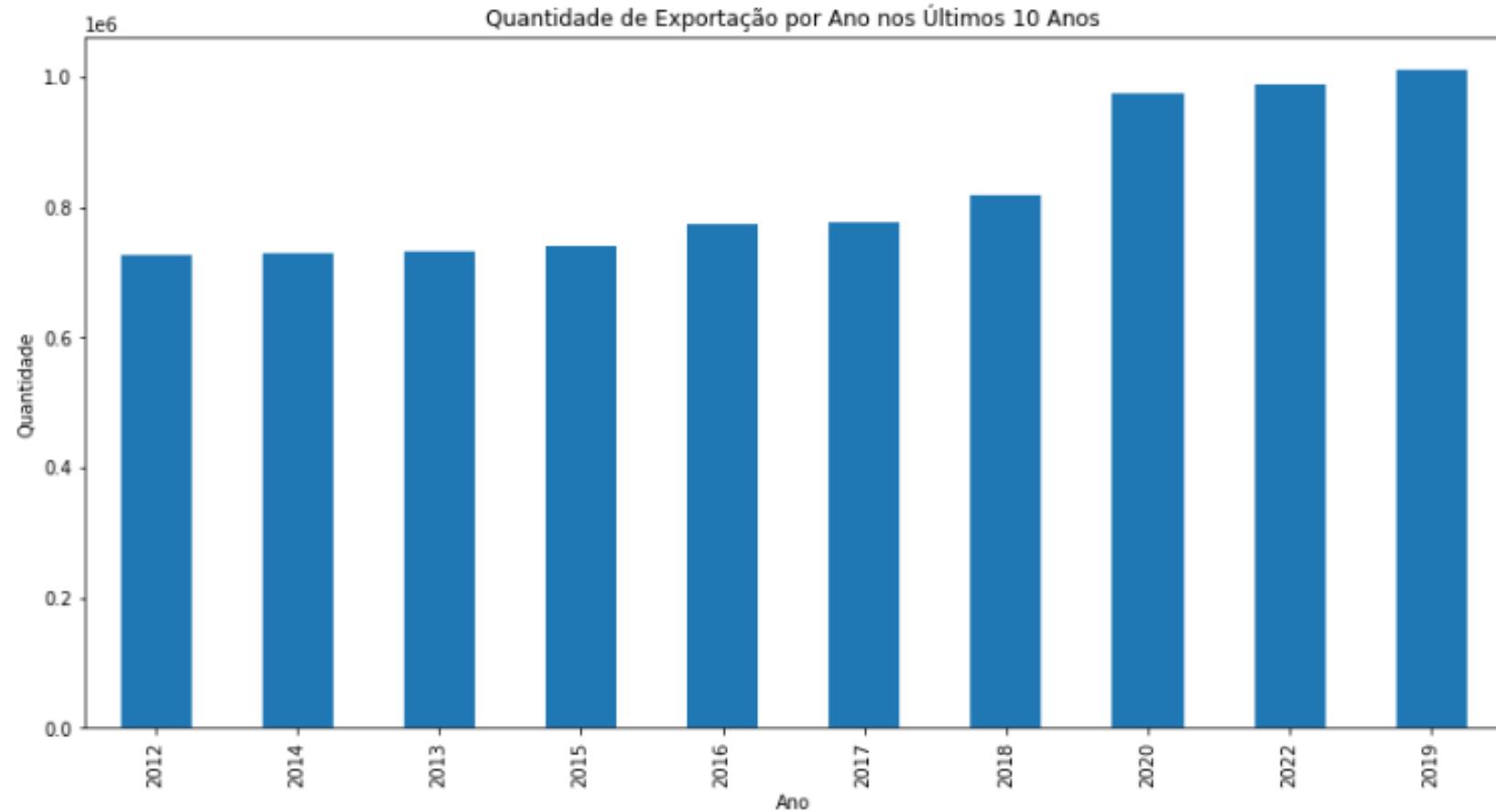
df_gasol_ano.plot.scatter(x ='data_final', y = 'preco_medio_revenda', c = 'DarkBlue', xlabel = 'Mês', ylabel = 'Preço médio', figsize = (14,6))
```



Análise dos datasets normalizados



**Dataset
Exportação**



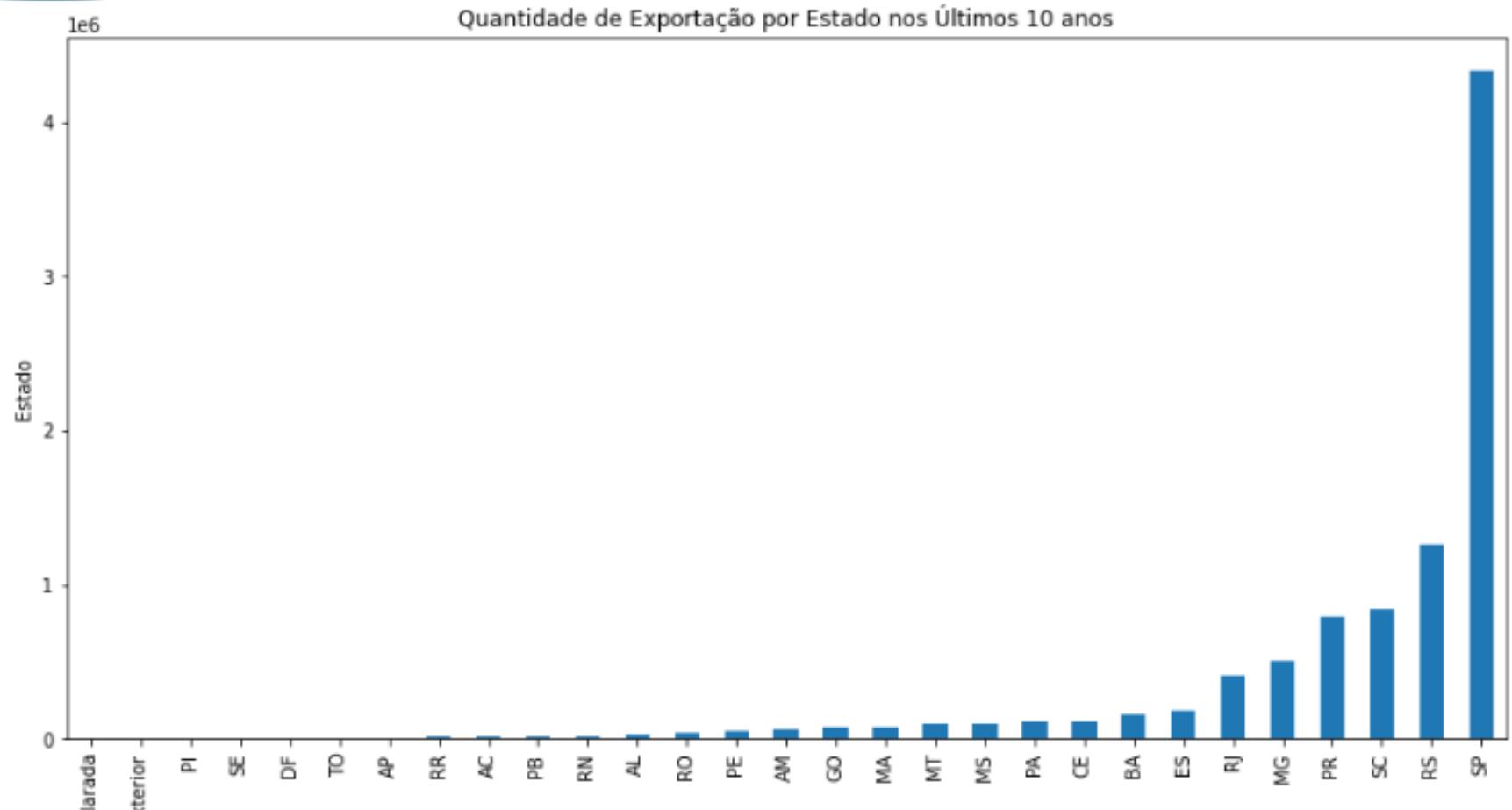
```
1 exp_ncm_tratado.groupby(['ano']).size().sort_values(ascending=True).head(10)
```

```
2 .plot.bar(figsize=(14,7), xlabel='Ano', ylabel='Quantidade', title='Quantidade de Exportação por Ano nos Últimos 10 Anos')
```

Análise dos datasets normalizados



Dataset Exportação



```
1 exp_ncm_tratado.groupby(['estado']).size().sort_values(ascending=True).head(35)
2 .plot.bar(figsize=(14,7), xlabel='Ano', ylabel='Estado', title='Quantidade de Exportação por Estado nos Últimos 10 anos')
```

Análise dos datasets normalizados

```
1 #os 10 produtos importados com maior valor em dolares no 10 anos.  
2 dfip.select(F.col('produto')).orderBy(F.col('valor_dolar').desc()).show(10,truncate=False)
```

```
+-----+  
| produto |  
+-----+  
| Plataformas de perfuração ou de exploração, flutuantes ou submersíveis |  
| Plataformas de perfuração ou de exploração, flutuantes ou submersíveis |  
| Plataformas de perfuração ou de exploração, flutuantes ou submersíveis |  
| Barcos-faróis/guindastes/docas/diques flutuantes, etc. |  
| Óleos brutos de petróleo |  
| Óleos brutos de petróleo |  
| Outras máquinas de sondagem/perfuração |  
| Barcos-faróis/guindastes/docas/diques flutuantes, etc. |  
| Óleos brutos de petróleo |  
| Óleos brutos de petróleo |  
+-----+
```



**Dataset:
Importação**

Análise dos datasets normalizados



Dataset: Exportação

valor_dolar	urf_localizacao	mercadoria	dolar_por_vol
25000000	AEROPORTO INTER...	Quadros, pinturas...	8333333.33
6355500	AEROPORTO INTER...	Pedras preciosas ...	6355500.0
15680906	SAO JOSE DOS CA...	Aviões e outros v...	5226968.67
4177010	IRF/BELO HORIZO...	Rubis, safiras e ...	4177010.0
4168620	AEROPORTO INTER...	Outros compostos ...	4168620.0
3874555	IRF/BELO HORIZO...	Rubis, safiras e ...	3874555.0
3250000	AEROPORTO INTER...	Outras pedras pre...	3250000.0
3243515	AEROPORTO INTER...	Outras pedras pre...	3243515.0
3094742	AEROPORTO INTER...	Rubis, safiras e ...	3094742.0
5616051	PORTO DE PARANAGUA	Biodiesel e suas ...	2808025.5

```
1 #adicionar um filtro, excluido, a mercadoria energia elétrico
2 dfexp_prod.filter(F.col('mercadoria') != 'Energia elétrica').orderBy(F.col('dolar_por_vol').desc()).show(10)
```

Análise dos datasets normalizados



Dataset Exportação

```
1 #rankear os produtos dentro de cada estado pelo volume
2 dfexp_prod.withColumn('rank', F.rank().over(w2)).show()
```

ano	mes	cod_mercadoria	unidade	cod_pais	estado	meio_transporte	cod_urf	qnt_estatistica	peso_liquido	valor_dolar	urf_localizacao	mercadoria	rank
2016	Fevereiro	12019000	Tonelada métrica ...	160	AM	Marítima	227600	29714	29714231	10257947	PORTO DE MANAUS Soja, mesmo tritu...	Soja, mesmo tritu...	1
2019	Janeiro	12019000	Tonelada métrica ...	160	AM	Marítima	227600	21435	21434816	7677952	PORTO DE MANAUS Soja, mesmo tritu...	Soja, mesmo tritu...	2
2019	Fevereiro	12019000	Tonelada métrica ...	160	AM	Marítima	227600	15004	15004225	5112690	PORTO DE MANAUS Soja, mesmo tritu...	Soja, mesmo tritu...	3
2016	Outubro	25171000	Quilograma líquido	169	AM	Fluvial	227600	12705700	12705700	283457	PORTO DE MANAUS Calhaus, cascalho...	Calhaus, cascalho...	4
2015	Novembro	25171000	Quilograma líquido	169	AM	Fluvial	227600	11772800	11772800	306875	PORTO DE MANAUS Calhaus, cascalho...	Calhaus, cascalho...	5
2016	Maio	25171000	Quilograma líquido	169	AM	Fluvial	227600	10383800	10383800	257337	PORTO DE MANAUS Calhaus, cascalho...	Calhaus, cascalho...	6
2021	Dezembro	19012000	Quilograma líquido	850	AM	Rodoviária	260151	9558190	9558190	5603152	PACARAIMA Misturas e pastas...	Misturas e pastas...	7
2021	Agosto	19012000	Quilograma líquido	850	AM	Rodoviária	260151	9538500	9538500	5917227	PACARAIMA Misturas e pastas...	Misturas e pastas...	8
2021	Maio	19012000	Quilograma líquido	850	AM	Rodoviária	260151	9234830	9234830	5518708	PACARAIMA Misturas e pastas...	Misturas e pastas...	9
2021	Novembro	19012000	Quilograma líquido	850	AM	Rodoviária	260151	8636000	8636000	5291031	PACARAIMA Misturas e pastas...	Misturas e pastas...	10

Análise dos datasets normalizados



**Seleção
Funções
Agrupamento
Ordenação**

RUN SAVE SHARE SCHEDULE MORE ✓

```

1 #valor total e volume total importado em cada seção do ISIC por ano e
2 #ordenado pelo valor por volume
3 SELECT i.ano, c.isic_secao,
4 SUM(i.valor_dolar) AS valor_total,
5 SUM(i.peso_liquido) AS volume_total,
6 AVG(i.valor_dolar)/AVG(i.peso_liquido) AS valor_por_volume
7 FROM `bc26-ed7-projeto-final.comercio_brasil.imp_2013-2022_tratada_ncm` AS i
8 INNER JOIN `bc26-ed7-projeto-final.comercio_brasil.imp-exp-classificacoes` AS c
9 ON i.mercadoria=c.ncm
10 GROUP BY 1, 2
11 ORDER BY i.ano DESC, valor_por_volume DESC;
12

```

Press Alt+F1 for /

SAVE RESULTS EXPLORE

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS		EXECUTION GRAPH
Row	ano	isic_secao	valor_total	volume_total	valor_por_volume	PR
1	2022	Outros Produtos	2138388108	420267655	5.0881577074...	P
2	2022	Indústria de Transformação	214500428...	103231873...	2.0778507730...	R
3	2022	Agropecuária	5215272351	10042570244	0.5193164921...	I
4	2022	Indústria Extrativa	20100259540	43169303944	0.4656146313...	C

Análise dos datasets normalizados



**Produto de
maior volume
exportado**

RUN SAVE SHARE SCHEDULE MORE

```
1 #produto exportado com maior volume
2 SELECT e.mercadoria,
3     e.data,
4     e.estado,
5     p.pais
6 FROM `bc26-ed7-projeto-final.comercio_brasil.exp_2013-2022_tratada_ncm` AS e
7 INNER JOIN `bc26-ed7-projeto-final.comercio_brasil.nome_pais` AS p
8 ON e.cod_pais = p.cod_pais
9 ORDER BY peso_liquido DESC
10 LIMIT 1;
```

Query results

SAVE RESULTS

Row	mercadoria	data	estado	pais
1	Navios de guerra	2022-08-01...	RJ	Turquia

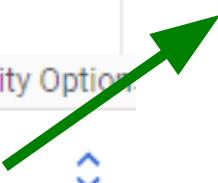
Selecionar dados dos dataset como fonte de dados para um relatório visual interativo.

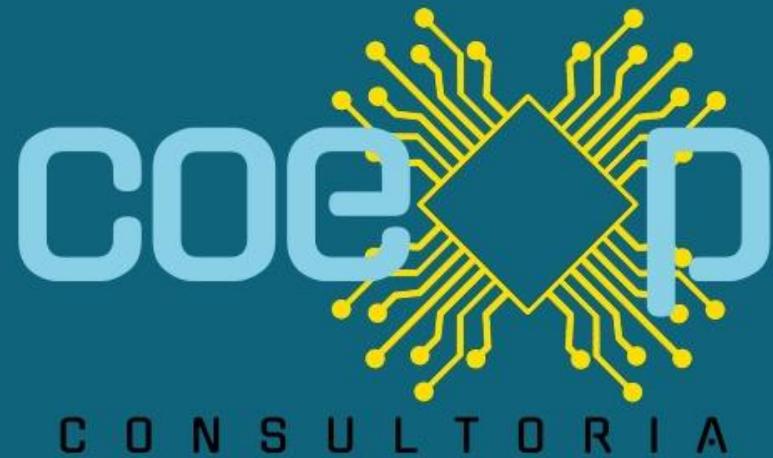
RUN SAVE ▾ SHARE ▾ SCHEDULE ▾ MORE ▾ Query completed.

```
1 #criar a tabela com o filtro para combustiveis para exportação
2 CREATE OR REPLACE TABLE comercio_brasil.exp_combustivel AS (
3     SELECT e.ano, e.mes, e.mercadoria AS produto, p.pais, e.estado, e.meio_transporte, e.data,
4     e.valor_dolar AS valor_exportado, e.peso_liquido AS volume_exportado
5     FROM `bc26-ed7-projeto-final.comercio_brasil.exp_2013-2022_tratada_ncm` AS e
6     INNER JOIN `bc26-ed7-projeto-final.comercio_brasil.nome_pais` AS p ON e.cod_pais = p.cod_pais
7     WHERE e.cod_mercadoria = 27090010 OR e.cod_mercadoria =27101922 OR e.cod_mercadoria =99980101 OR
8     e.cod_mercadoria =99981002 OR e.cod_mercadoria =27101259 OR e.cod_mercadoria =27101921 OR e.
9     cod_mercadoria =27131100 OR e.cod_mercadoria =27131200 OR
10    e.cod_mercadoria =27111100 OR e.cod_mercadoria = 27101929 OR e.cod_mercadoria =39111029 OR e.
11    cod_mercadoria =27111910);
```

Press Alt+F1 for Accessibility Options

Query results SAVE RESULTS ▾ EXPLORE DATA ▾





Fase 5: Análise de dados com o Looker Studio

Brasil: Exportação x Importação

Total importação

em US\$

1,9 tri

Total exportação

em US\$

2,28 tri

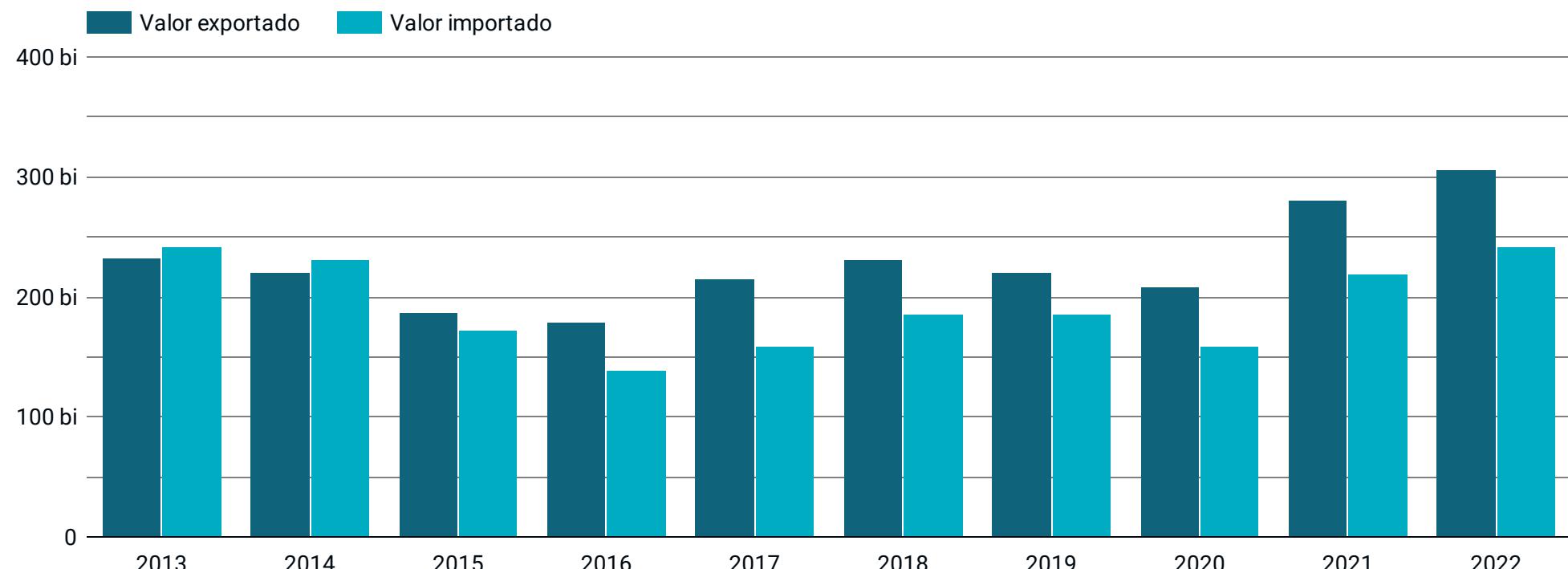
Selecionar CUCI

Código CUCI

Selecionar país

País

Exportação x Importação (em US\$)



Brasil: Balança Comercial

Total importação

em US\$

1,9 tri

Total exportação

em US\$

2,28 tri

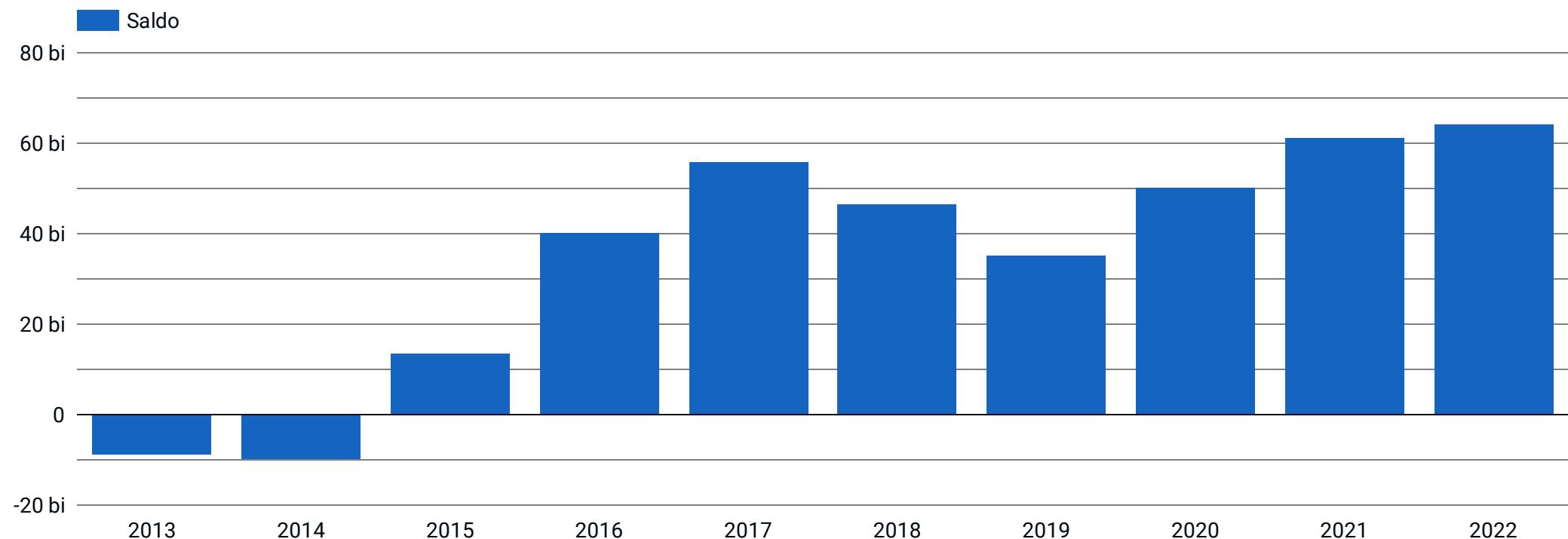
Selecionar CUCI

Código CUCI

Selecionar país

País

Saldo da balança comercial



Importação: Países

Total importação

em US\$

1,9 tri

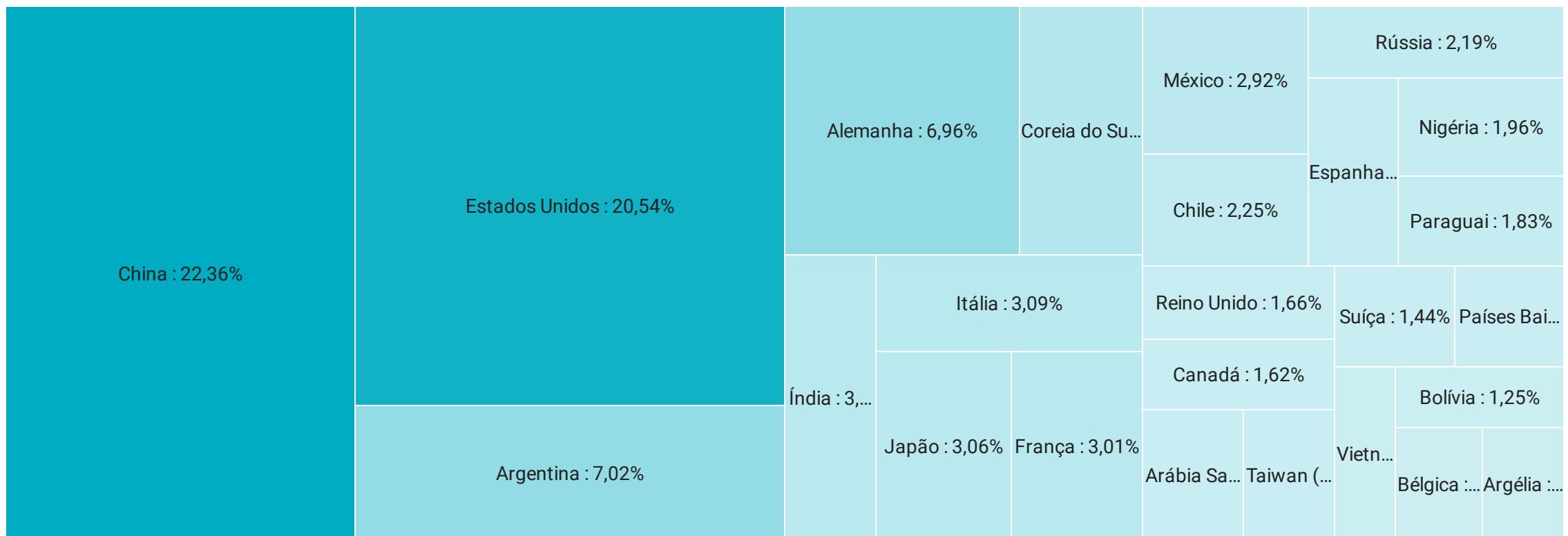
Selecionar setor

Selecionar ano

Setor produtivo

Ano

De quais países importamos?



Importação: Produtos

Total importação

em US\$

1,9 tri

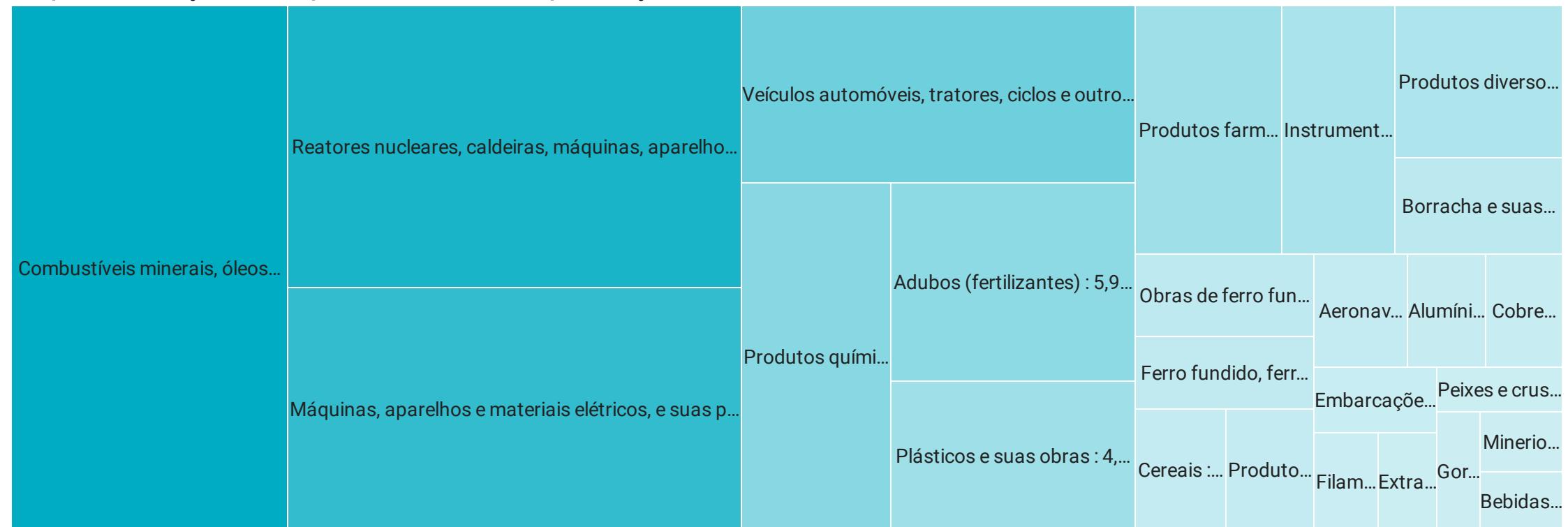
Selecionar setor

Selecionar ano

Setor produtivo

Ano

Representação dos produtos nas importações



Importação: Estados e meios de transporte

Total importação

em US\$

1,9 tri

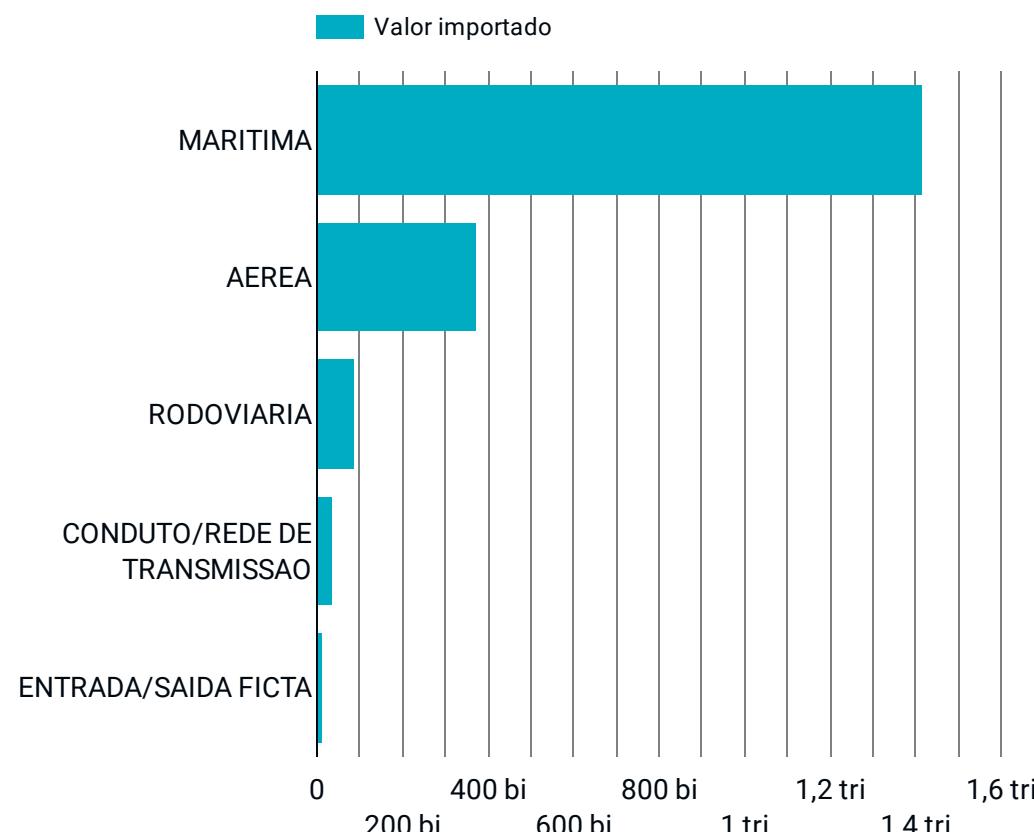
Selecionar setor

Selecionar ano

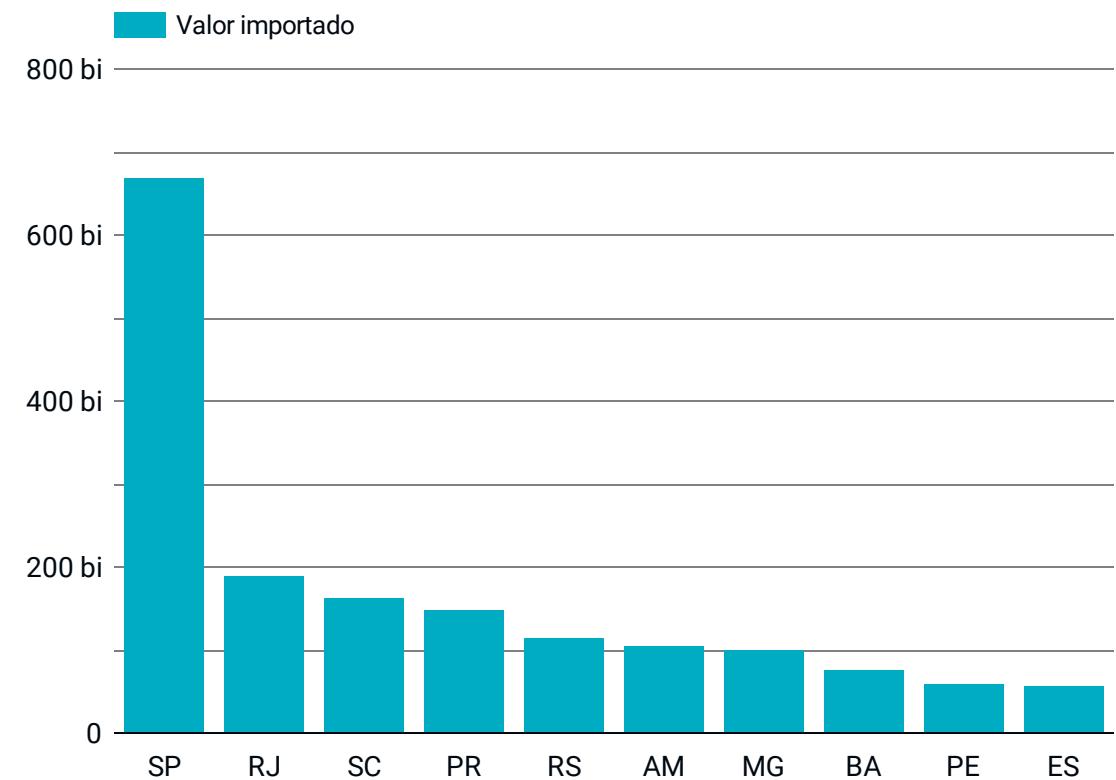
Setor produtivo

Ano

Importações por Meio de transporte



Importações por Estado



Exportação: Países

Total exportação

em US\$

2,3 tri

Selecionar setor

Selecionar ano

Setor produtivo

Ano

Para quais países exportamos?

China : 31,95%

Estados Unidos : 15,12%

Argentina : 7,66%

Espanha : 2,46%

Itália : 2,03%

Bélgica : 1,83%

Coreia do S... Índia : 2,2%

Canadá...

Reino Unido : 1,...

Singapura : 1,75%

Malásia : 1,4... Uruguai : 1,3...

Colômbia : 1,56%

Emirados Árabes Un...

Alemanha : 2,... Chile : 2,97% México : 2,...

Paraguai : ... França : ...

Hong...

Arábia Sa... Rússia : ...

Exportação: Produtos

Total exportação

em US\$

2,3 tri

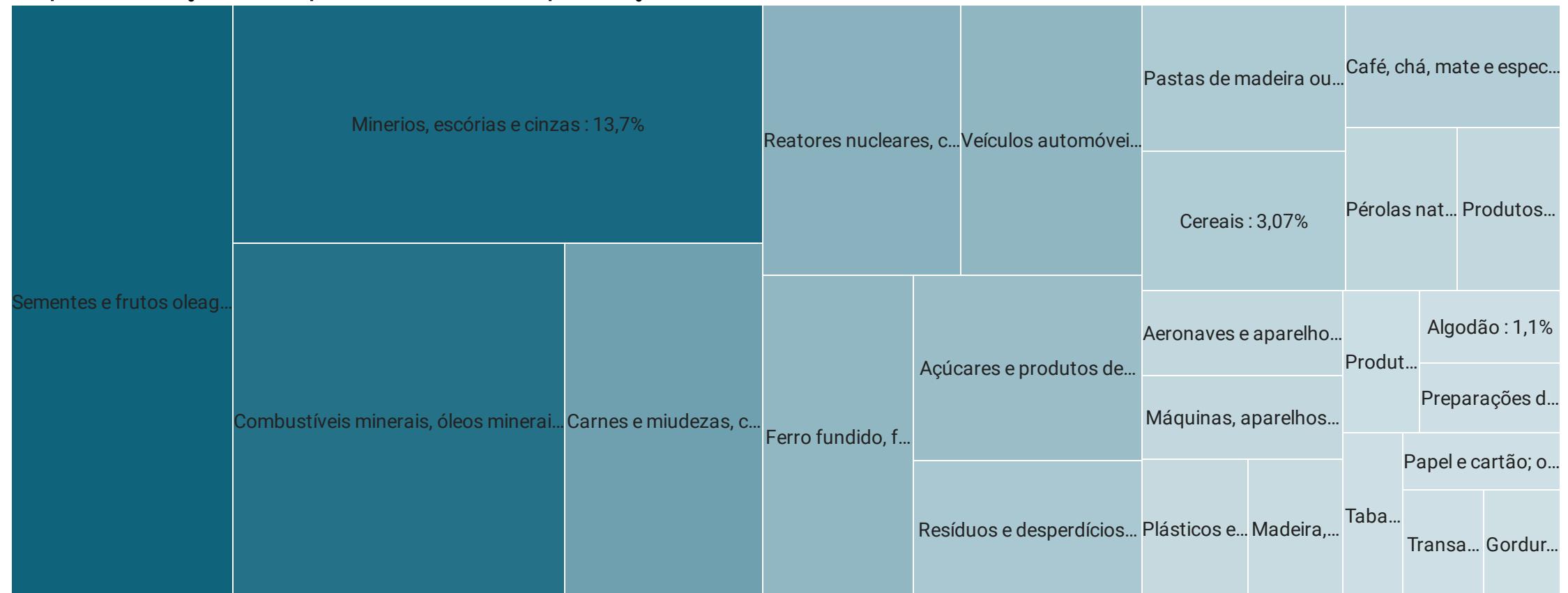
Selecionar setor

Selecionar ano

Setor produtivo

Ano

Representação dos produtos nas exportações



Exportação: Estados e meios de transporte

Total exportação

em US\$

2,3 tri

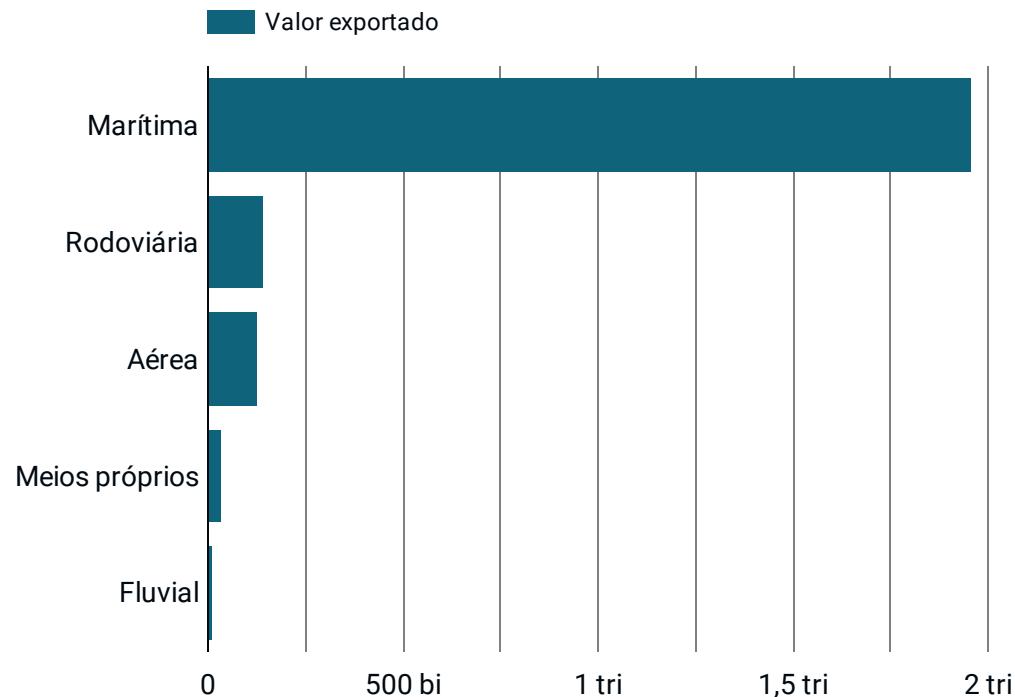
Selecionar setor

Selecionar ano

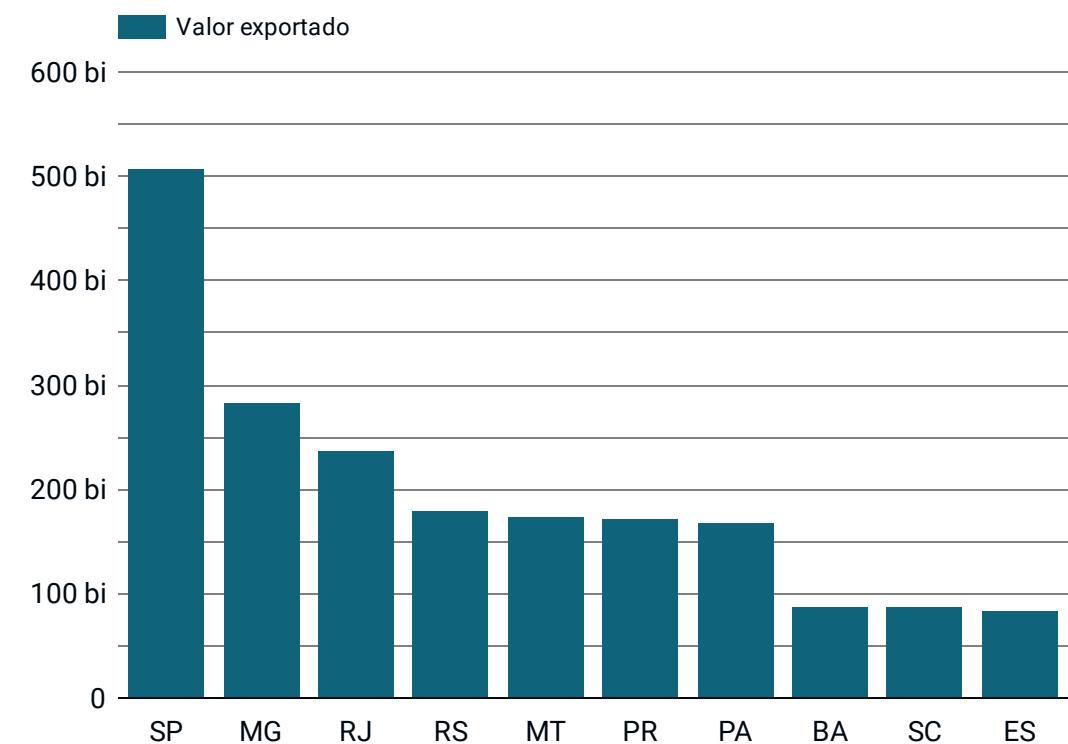
Setor produtivo

Ano

Exportações por meio de transporte



Exportações por estado



Volume: Exportação e Importação por países

Total exportação

em US\$

679,3 bi

Selecionar CUCI

Selecionar ano

Seção CUCI

Ano: 2022

(1) ▾

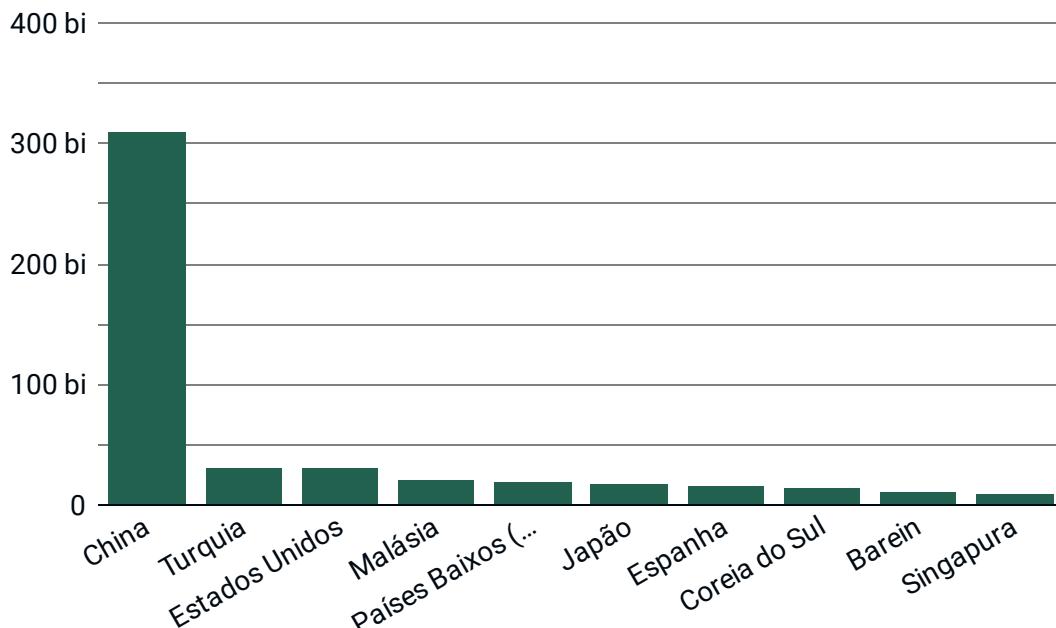
Total importação

em US\$

242,0 bi

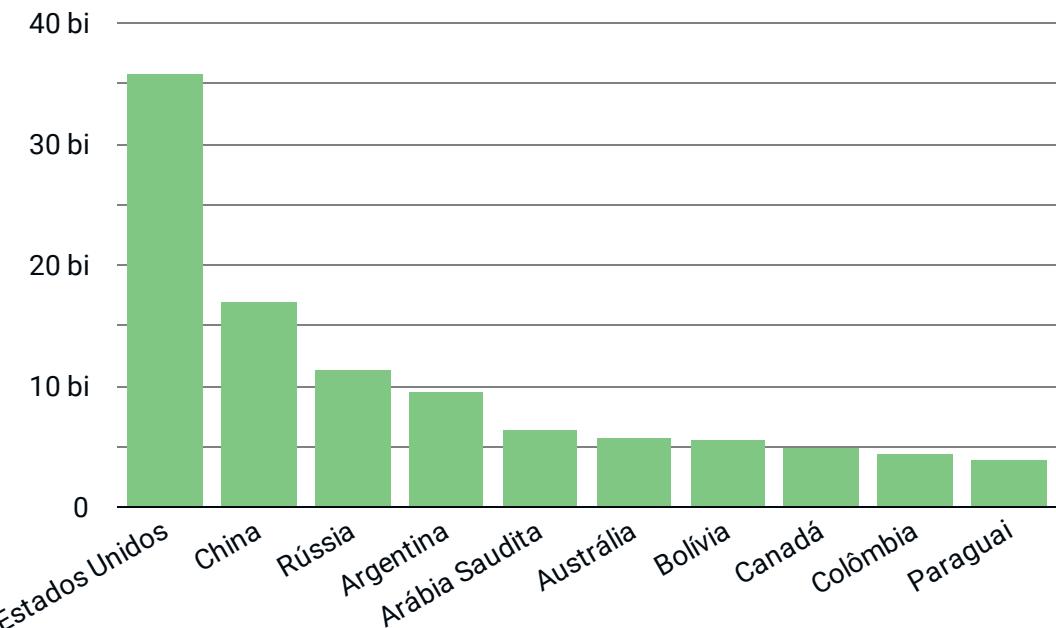
Países por volume exportado

Volume exportado



Países por volume importado

Volume importado



Volume: Exportação e Importação por produto

Total exportação

em US\$

679,3 bi

Selecionar CUCI

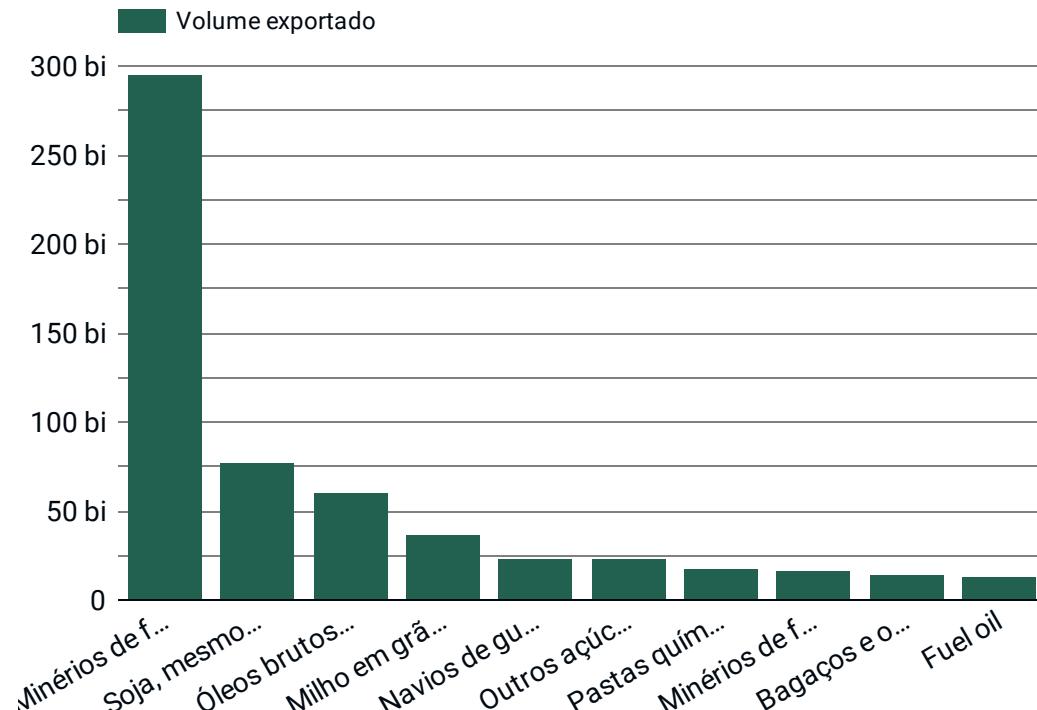
Selecionar ano

Seção CUCI

Ano: 2022

(1) ▾

Volume por produto exportado

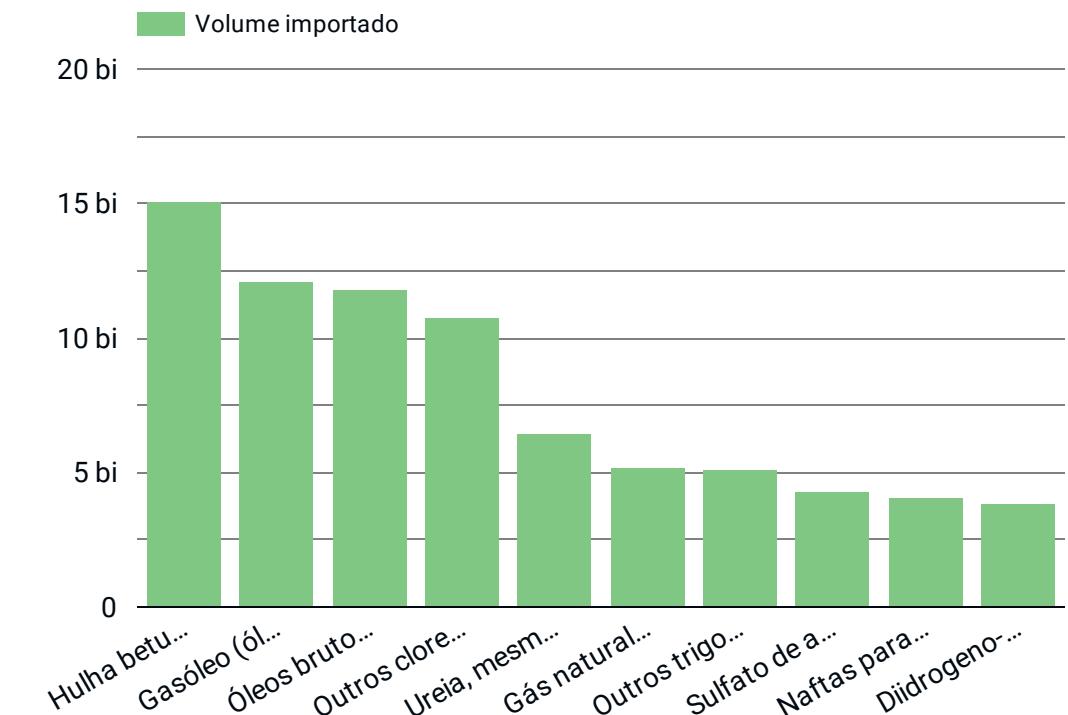


Total importação

em US\$

242,0 bi

Volume por produto importado



Volume: Exportação x Importação

Total exportação
em US\$

679,3 bi

Selecionar CUCI

Selecionar ano

Seção CUCI

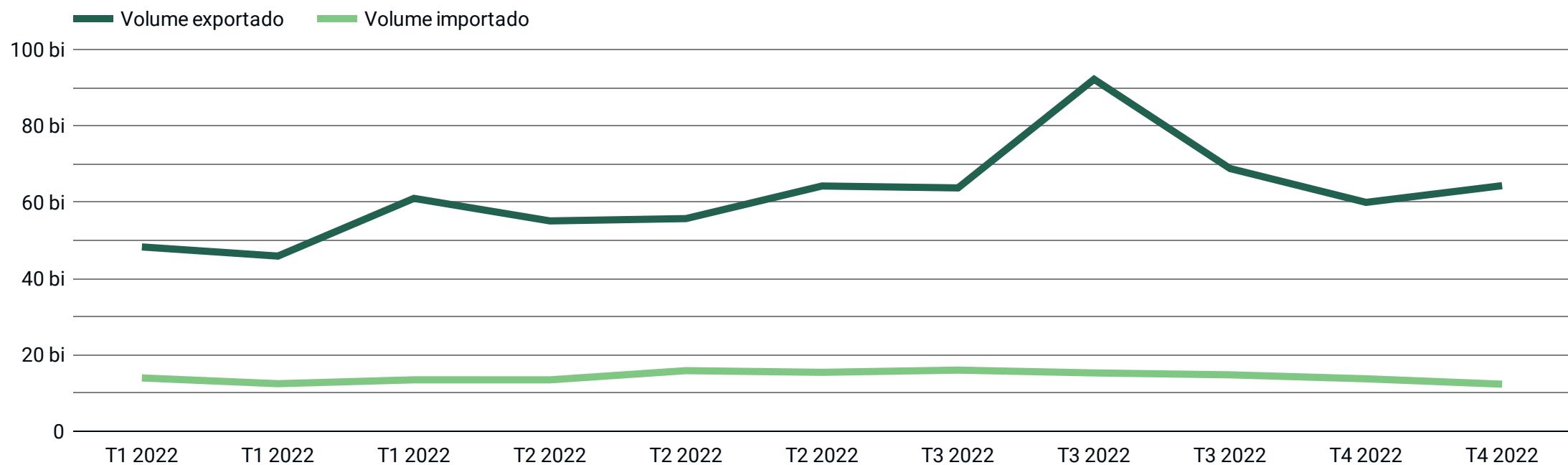
Ano: 2022

(1) ▾

Total importação
em US\$

242,0 bi

Volume exportado x importado ao longo do tempo



Petróleo e combustíveis: Comércio Externo

Total exportação de
combustíveis
em US\$
256,7 bi

Produtos exportados

Óleos brutos de petróleo

Fuel oil

Consumo de bor...

Outras... Gasól...
Co... Co...
...

Selecionar produto

Produto

Selecionar data

Selecionar período

Total importação de
combustíveis
em US\$
175,0 bi

Produtos importados

Gasóleo (óleo diesel)

Óleos brutos de petróleo

Gás natural liquefeito Outras gasolinhas, e...

Coqu...

Fuel...

G...

C...

...

Petróleo e combustíveis: Balança Comercial

Total exportação de
combustíveis
em US\$
205,5 bi

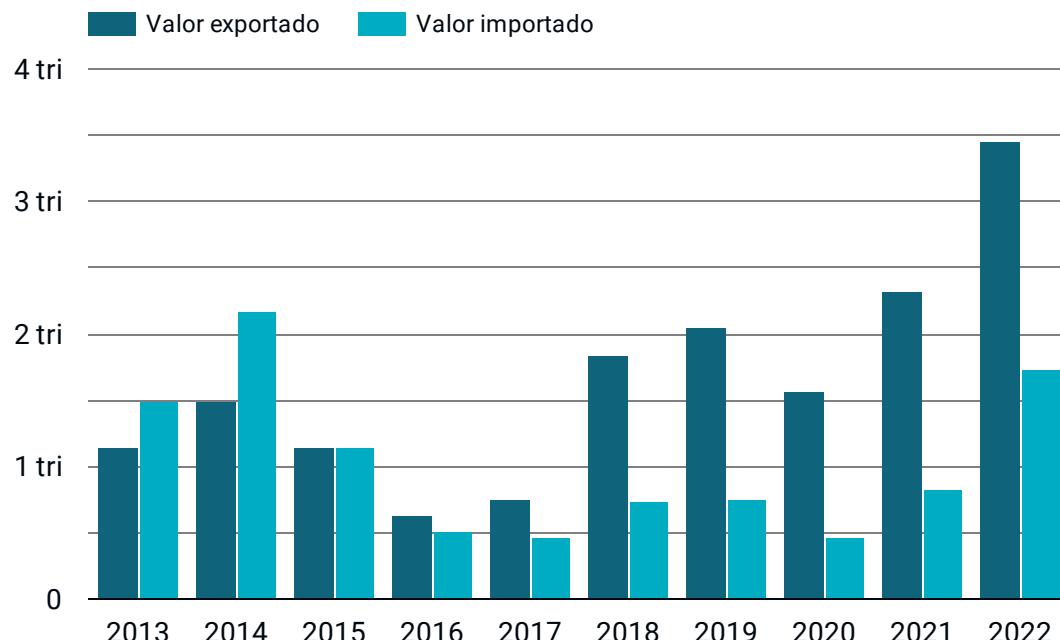
Selecionar produto

Selecionar data

Produto: Óleos brut... (1) ▾

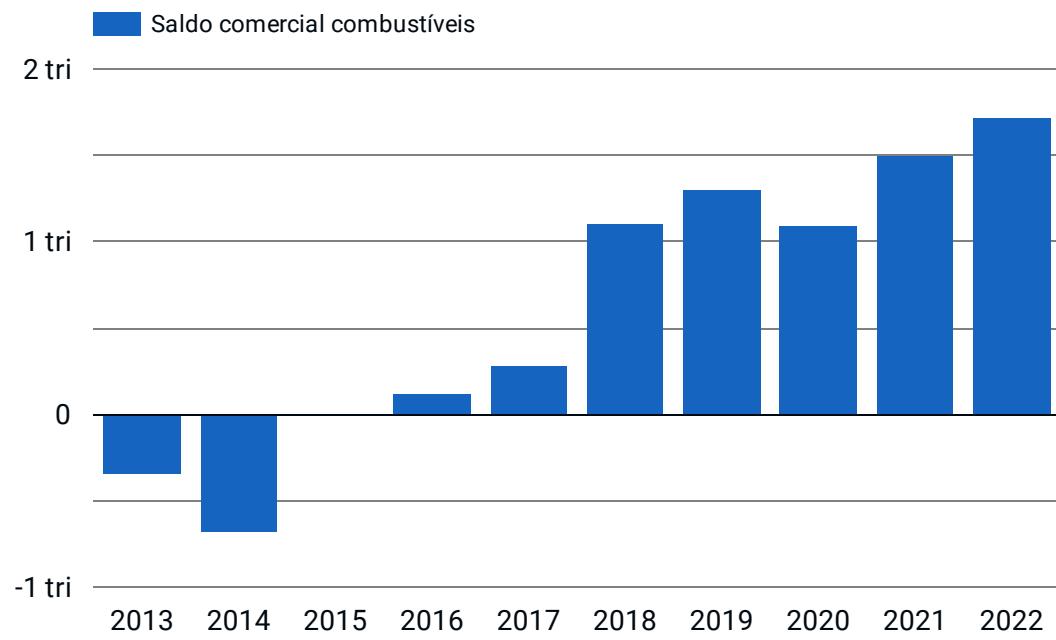
Selecionar período ▾

Valor exportado x importado



Total importação de
combustíveis
em US\$
70,1 bi

Saldo comercial petróleo e combustíveis



Petróleo e combustíveis: Volume e meios de transporte

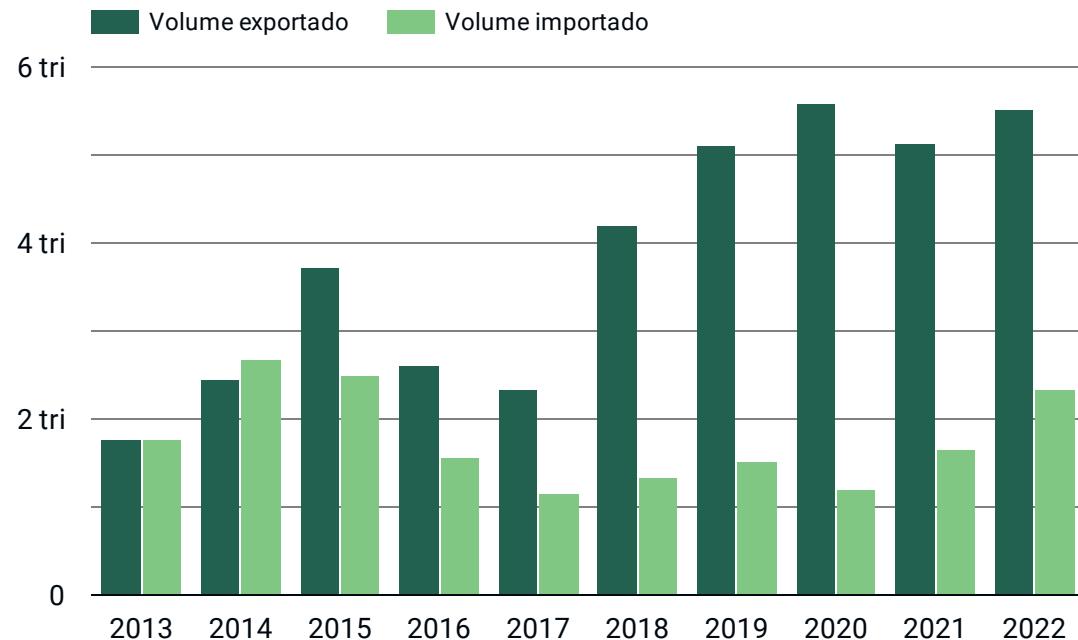
Total exportação de combustíveis
em US\$
205,5 bi

[Selecionar produto](#)
[Selecionar data](#)

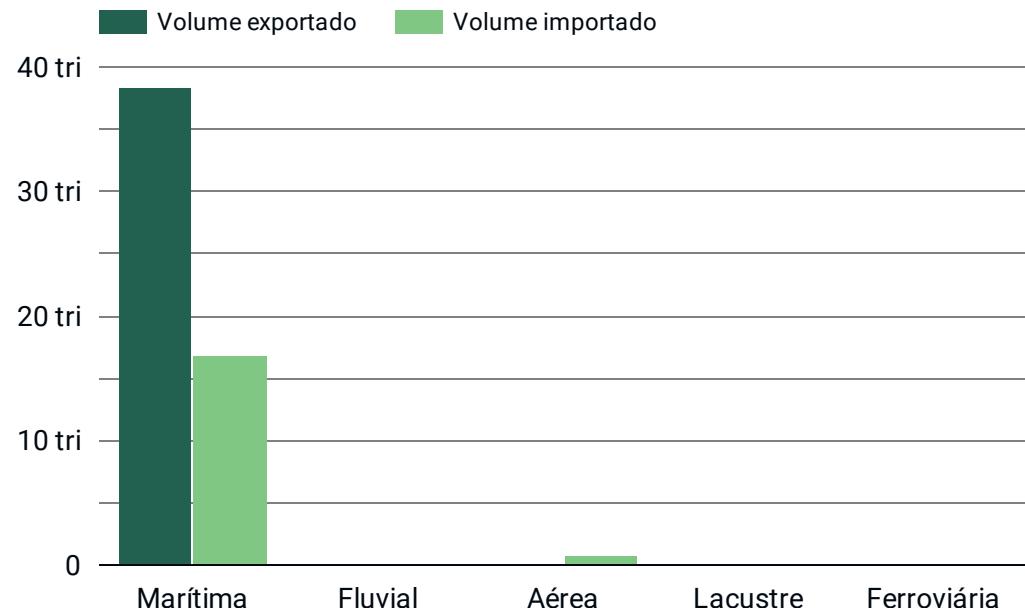
Produto: Óleos brut... (1) ▾

[Selecionar período](#) ▾

Volume exportado x importado



Vias de transporte



Petróleo e combustíveis: Países

Total exportação de petróleo e combustíveis
em US\$
205,5 bi

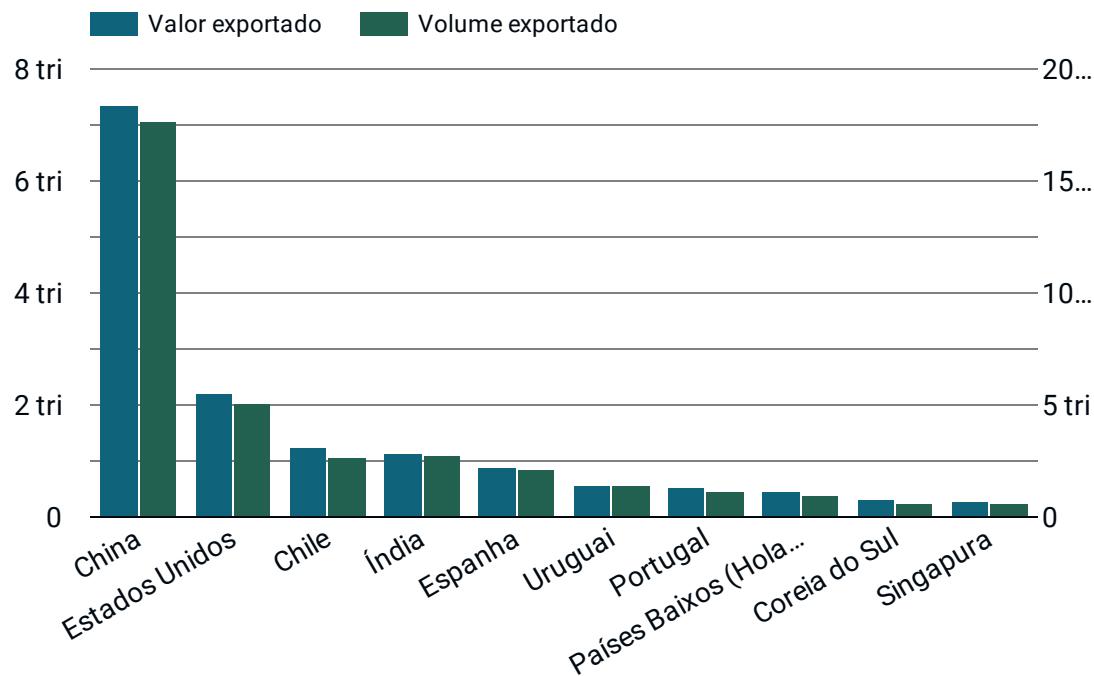
Selecionar produto

Produto: Óleos brut... (1) ▾

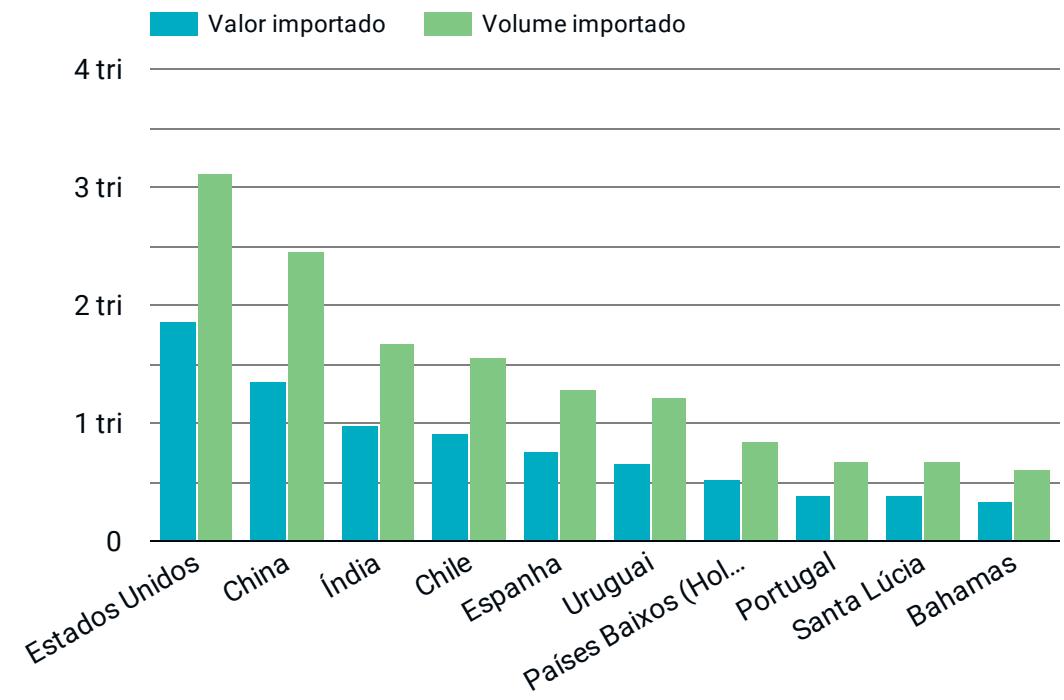
Selecionar ano

Ano ▾

Exportação por país



Importação por país



Petróleo e combustíveis: Estados

Total exportação de petróleo e combustíveis
em US\$
256,7 bi

Selecionar produto

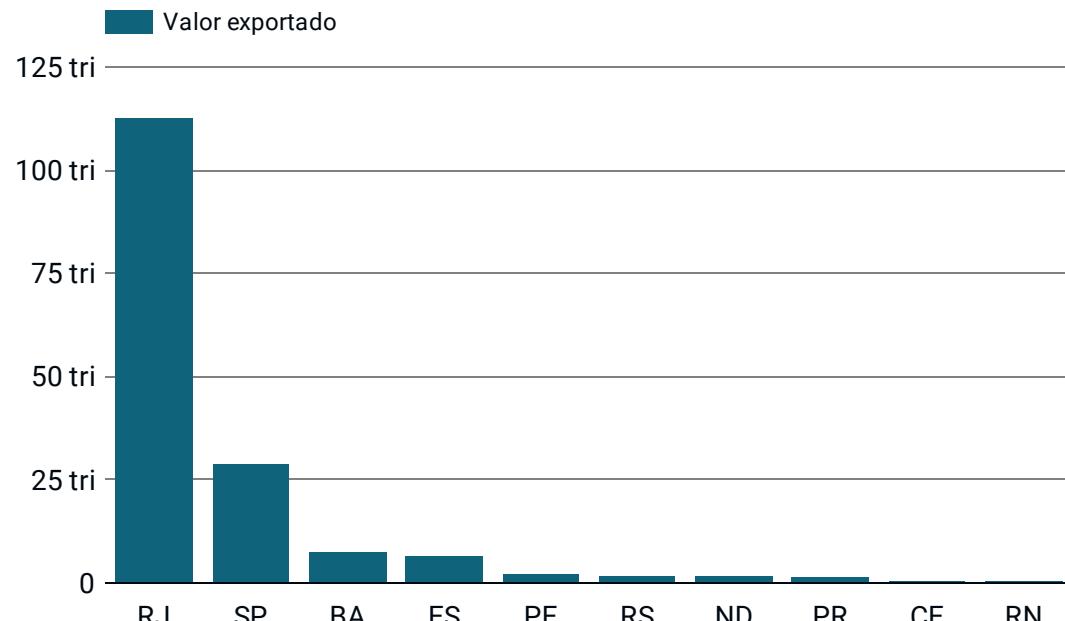
Selecionar ano

Produto

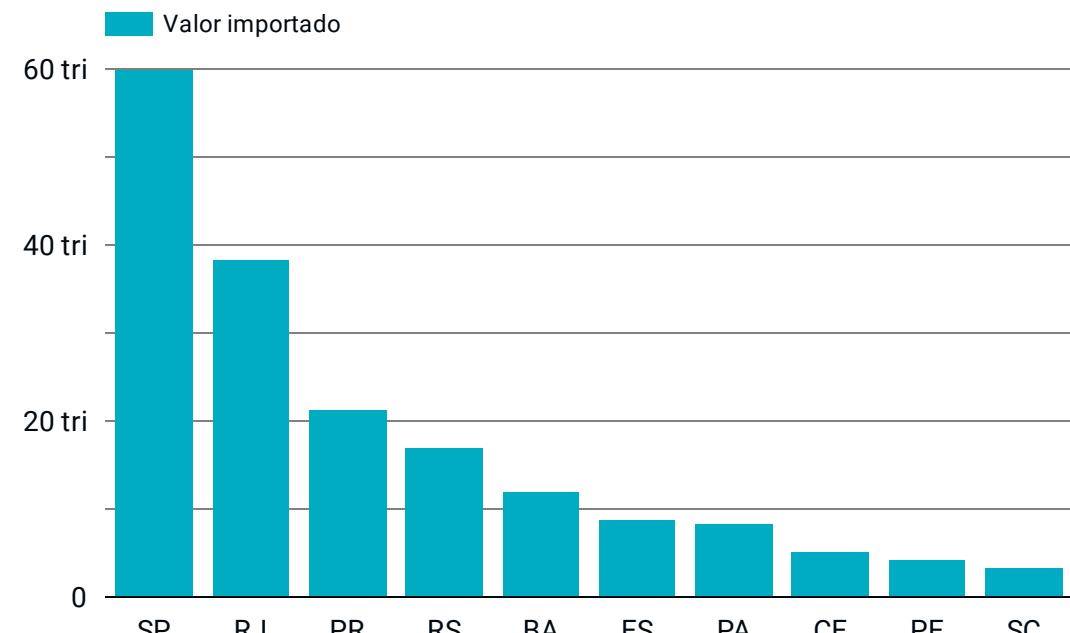
Ano

Total importação de petróleo e combustíveis
em US\$
175,0 bi

Exportação por estado



Importação por estado



Petróleo e combustíveis: Volume

Total exportação de petróleo e combustíveis em US\$ 205,5 bi

Selecionar produto

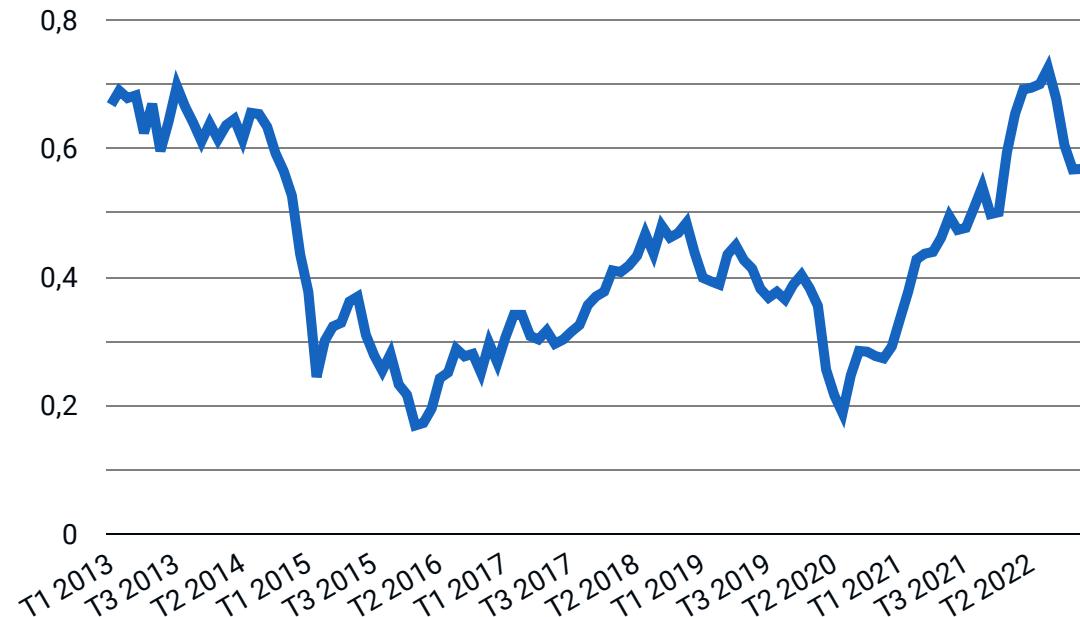
Selecionar ano

Produto: Óleos brut... (1) ▾

Ano ▾

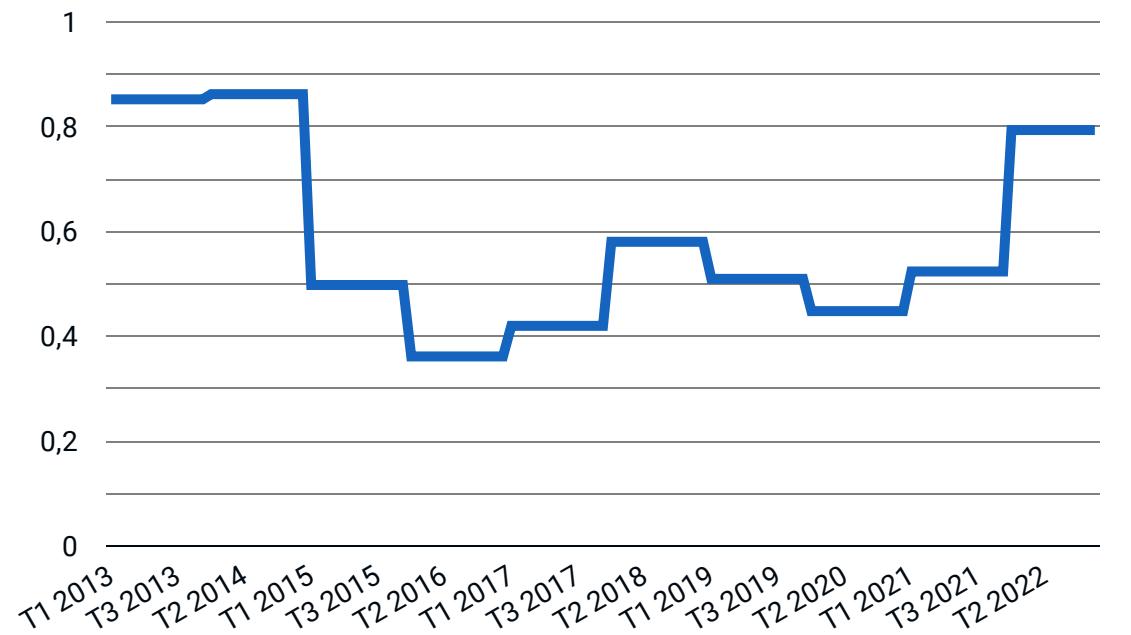
Valor por volume na exportação

— Mediana valor por volume exportado



Valor por volume na importação

— Mediana valor por volume importado



Combustíveis: Comércio Interno

Selecionar estado

Estado

Selecionar produto

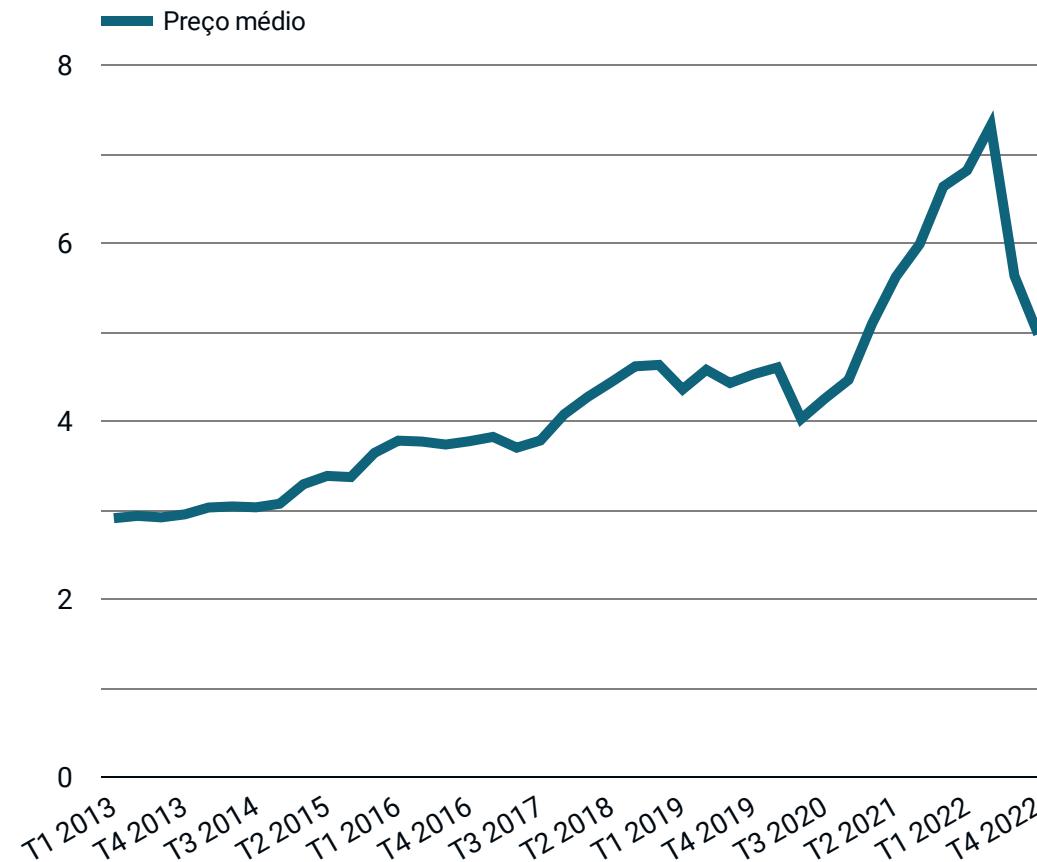
produto: GASOLINA COMUM (1)

Unidade de medida ▾

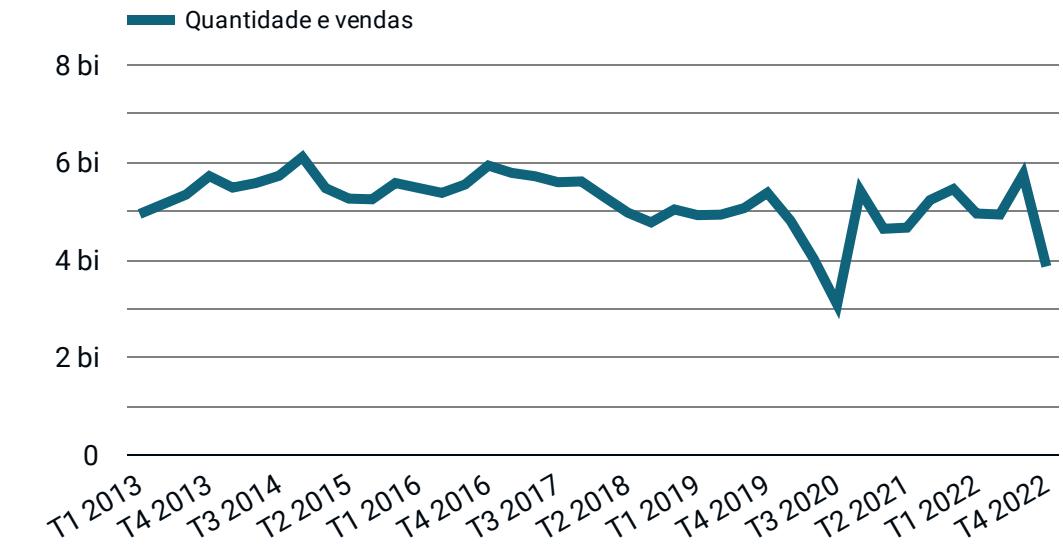
1. R\$/l

1 - 1 / 1 < >

Preços médios por produto (2013-2022)



Volume de vendas por produto (2013-2022)



CORREIO BRAZILIENSE Acervo

BRASIL

Imagens de satélite mostram como a quarentena reduz a poluição

Imagens de satélite revelam diminuição dos níveis de dióxido de nitrogênio em São Paulo, Rio, Brasília e outras capitais. Especialistas avaliam o impacto da mudança na saúde das pessoas e propõem uma reflexão sobre modelos econômicos e sistemas de transporte.

Combustíveis: Pandemia e Guerra

Selecionar produto

Selecionar estado

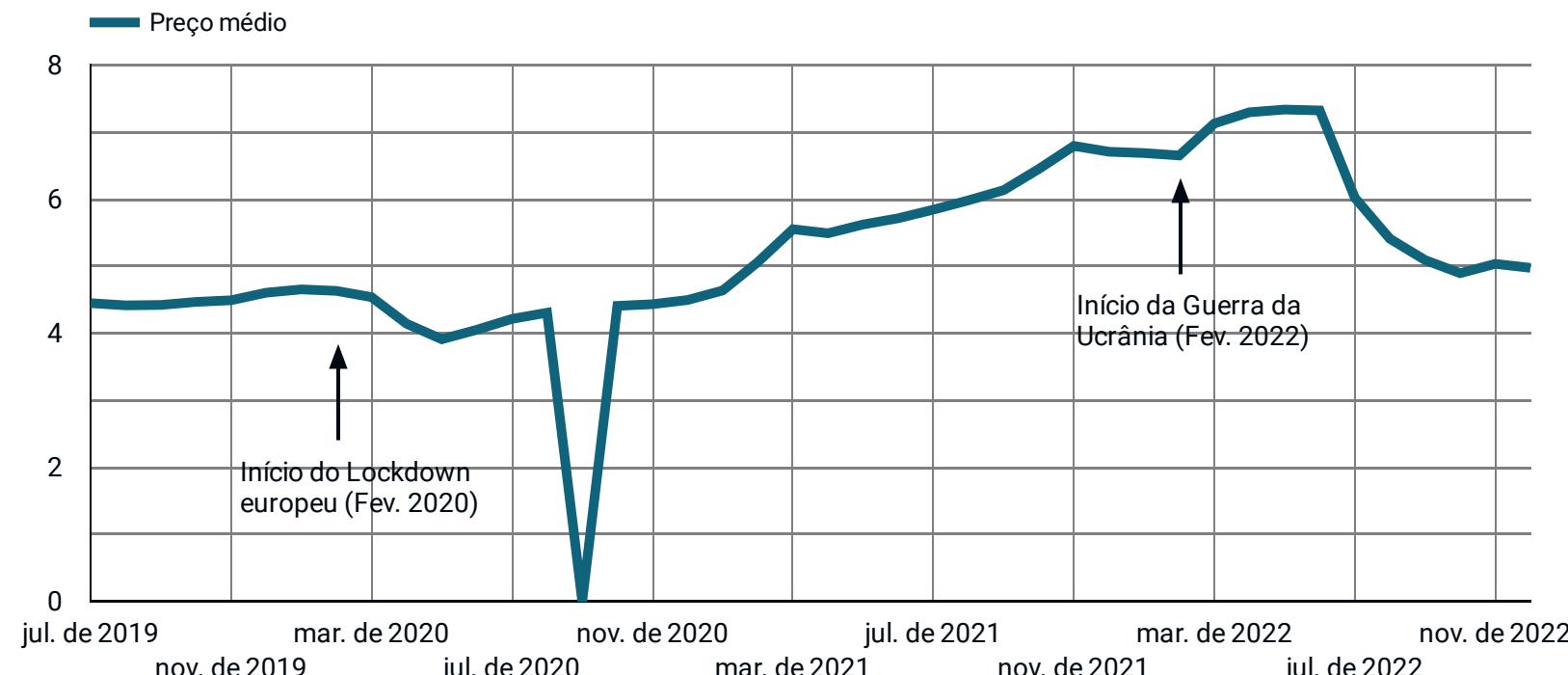
produto: GASOLINA COMUM

(1) ▾

Estado

▼

Impacto da pandemia e da Guerra na Ucrânia sobre o preço médio dos combustíveis (Agosto 2019 - Dezembro 2022)



g1

ECONOMIA

ANP deixará de produzir pesquisa de preços de combustíveis por duas semanas

Contrato com a empresa que realizava a pesquisa de preços foi encerrado, e o novo acordo prevê a próxima coleta de preços a partir de 8 de setembro, com a divulgação dos dados a partir de 14 de setembro.

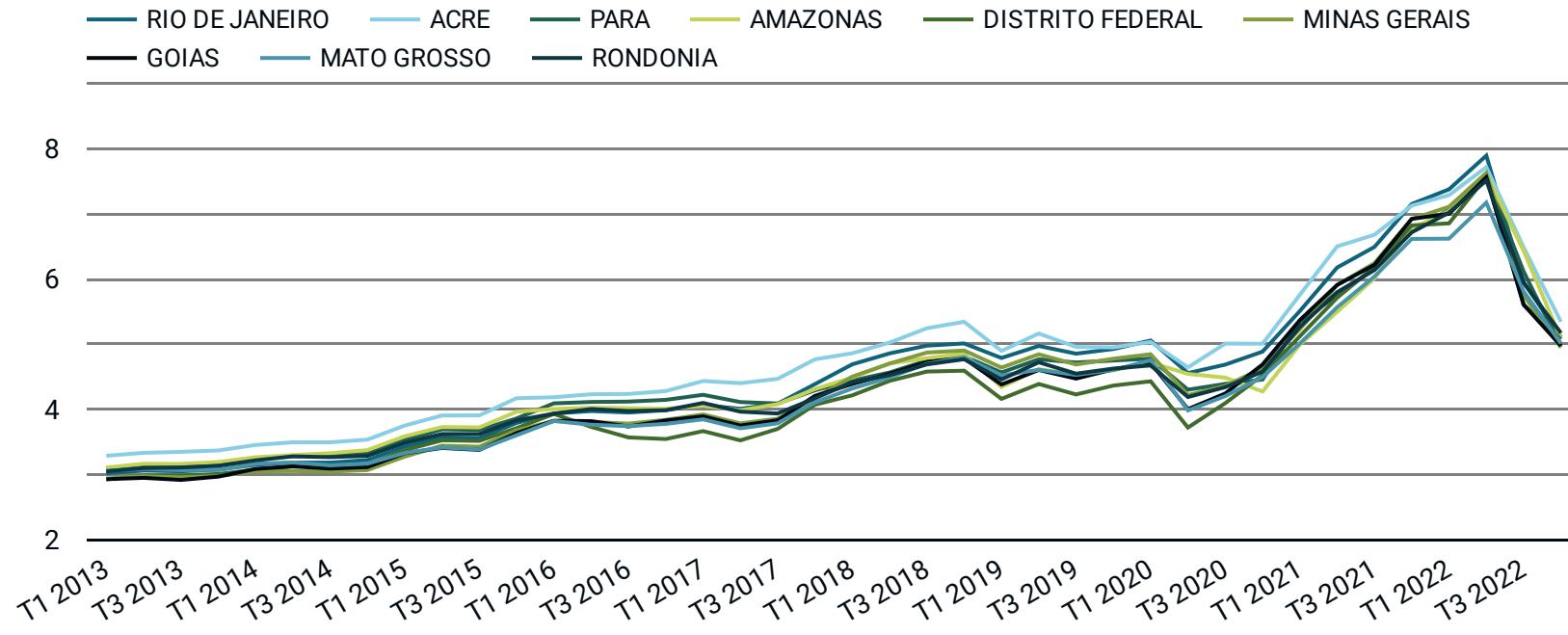
Por G1

25/08/2020 11h41 · Atualizado há 2 anos



Combustíveis Brasil: Preço

Maiores preços por estado



Selecionar região

Região

Selecionar produto

Produto: GASOLINA... (1)

Unidade de medida ▾

1. R\$/l

1 - 1 / 1 < >

Preço médio por produto por região

Região	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Ano / Preço médio
CENTRO OESTE	3,01	3,14	3,44	3,77	3,85	4,55	4,48	4,36	6,16	6,47	
NORDESTE	2,9	3,01	3,39	3,79	3,78	4,46	4,52	4,36	6	6,62	
NORTE	3,13	3,26	3,64	4,02	4,07	4,63	4,6	4,44	6,09	6,63	
SUDESTE	2,88	3	3,33	3,7	3,77	4,5	4,46	4,31	5,9	6,33	
SUL	2,9	3,01	3,34	3,77	3,81	4,43	4,36	4,22	5,93	6,26	

Combustíveis Brasil: Preços máximos e mínimos

[Selecionar produto](#)
[Selecionar estado](#)
[Selecionar ano](#)

produto: GASOLINA ... (1) ▾

Estado ▾

Ano ▾

Maiores preços

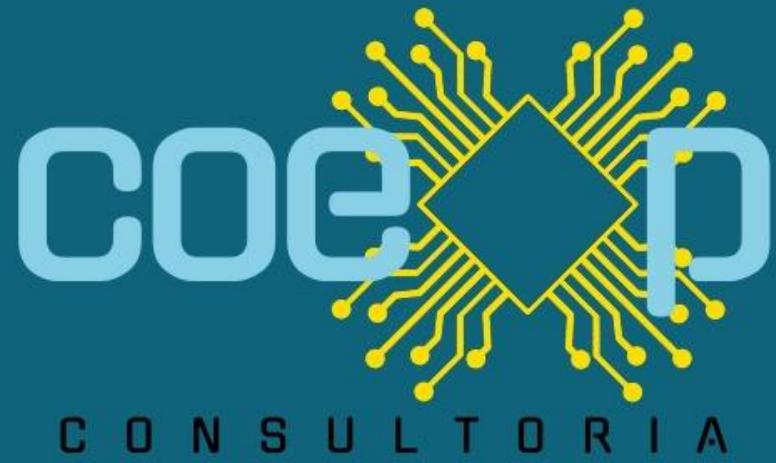
	Data	Estado	Município	Preço médio
1.	25 de jun. de 2022 00...	BAHIA	CAETITE	8,54
2.	18 de jun. de 2022 00...	BAHIA	CAETITE	8,54
3.	18 de jun. de 2022 00...	BAHIA	TEIXEIRA DE ...	8,5
4.	12 de mar. de 2022 0...	BAHIA	EUNAPOLIS	8,5
5.	4 de jun. de 2022 00...	RIO DE JA...	ANGRA DOS ...	8,49
6.	25 de jun. de 2022 00...	RIO DE JA...	ANGRA DOS ...	8,47
7.	25 de jun. de 2022 00...	RIO DE JA...	TRES RIOS	8,45
8.	25 de jun. de 2022 00...	BAHIA	TEIXEIRA DE ...	8,45
9.	21 de mai. de 2022 0...	BAHIA	PORTO SEGU...	8,37
10.	18 de jun. de 2022 00...	BAHIA	VITORIA DA C...	8,37

1 - 10 / 224422 < >

Menores preços

	Data	Estado	Município	Preço médio
1.	5 de jan. de 2013 00...	SAO PAU...	OURINHOS	2,38
2.	26 de jan. de 2013 00...	SAO PAU...	OURINHOS	2,38
3.	12 de jan. de 2013 00...	SAO PAU...	OURINHOS	2,39
4.	19 de jan. de 2013 00...	SAO PAU...	OURINHOS	2,39
5.	29 de jun. de 2013 00...	SAO PAU...	OURINHOS	2,44
6.	6 de jul. de 2013 00:0...	SAO PAU...	OURINHOS	2,44
7.	22 de jun. de 2013 00...	SAO PAU...	OURINHOS	2,45
8.	15 de jun. de 2013 00...	SAO PAU...	OURINHOS	2,45
9.	2 de fev. de 2013 00:0...	SAO PAU...	OURINHOS	2,46
10.	23 de nov. de 2013 0...	BAHIA	LAURO DE FR...	2,47

1 - 10 / 224422 < >



Fase 6: Análise de custos

Ferramentas utilizadas na GCP



Google Cloud Storage

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

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



Google Cloud Platform Marketplace

Ferramentas utilizadas na GCP



Google 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

Ferramentas utilizadas na GCP



Cloud Dataproc

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

Ferramentas utilizadas na GCP



Google BigQuery

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

Ferramentas utilizadas na GCP



Cloud SQL

- **Instância SQL – Para datasets originais 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)

Valor total do projeto(descritivo produtos e serviços)

Orçamento			
Nº	Descriutivo	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

Sugestão de novos produtos e serviços opcionais



➤ **MongoDB Atlas**
Armazenar datasets originais
Cluster dedicated – Google Cloud Platform

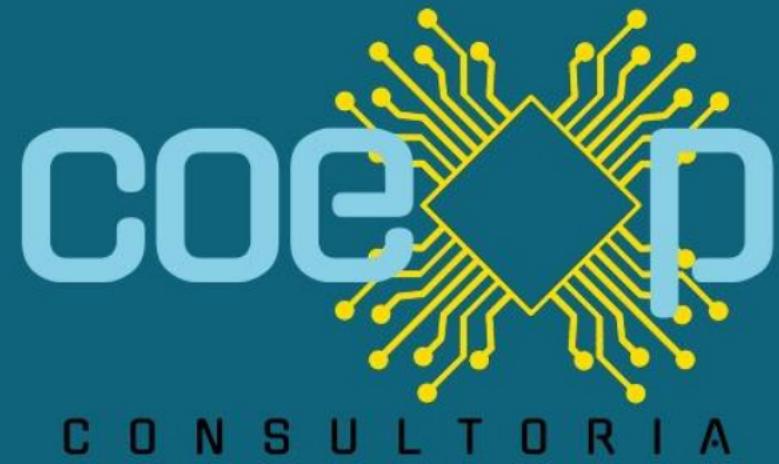
Valor Total: \$136,80 Previsão: Mensal

- M20M20 (cluster)
- 20GB (armazenamento)
- 3.75GB 4GB
- 1 vCPUs 2 vCPUs (\$0,19/h)

Sugestão de novos produtos e serviços opcionais



- **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 (+- 10GB)
 - Rede: entrada : grátis - Saída: \$11,50
 - Exportação sem servidor: \$0,015(GB) = \$0,15 (+10 GB)



Obrigado!

Contatos

Ana Flavia Dias

 amfd88@gmail.com
 encurtador.com.br/xBTUZ
 github.com/anafla7

Ariana Caetano

 ariana.caetano@gmail.com
 encurtador.com.br/bmsEK
 github.com/ariana-caetano

Erika Cunha

 erikancunha024@gmail.com
 encurtador.com.br/suxJ8
 github.com/Erikancunha024

José Jardim

 jo.jpv@gmail.com
 encurtador.com.br/alyB2
 github.com/JoseJardimPV



Marcos Valente

 marcos.valente.c@gmail.com
 encurtador.com.br/kvAU7
 github.com/MarcosVCS