

## 1. Título

Compreendendo fatores que podem influenciar o preço dos combustíveis dos estados brasileiros. Período de 2004 a 2021.

## 2. Nome da equipe e integrantes

### Os reveladores das faces dos dados

- Caique Salvador Noboa, 1904949, *caique2nobia*, [nobia@alunos.utfpr.edu.br](mailto:nobia@alunos.utfpr.edu.br), Eng. Comp, UTFPR;
- Maria Gabriela Rodrigues Policarpo, 2450151, *mariagabrielapolicarpo*, [mariagabrielapolicarpo@alunos.utfpr.edu.br](mailto:mariagabrielapolicarpo@alunos.utfpr.edu.br), BSI, UTFPR;
- Moisés Bryan Carneiro Roja, 2243814, *moisesbryan*, [moisesbryan@alunos.utfpr.edu.br](mailto:moisesbryan@alunos.utfpr.edu.br), Eng. Mecatrônica, UTFPR;
- Bruno Souza Zabot, 2238039, *bzabot*, [zabot@alunos.utfpr.edu.br](mailto:zabot@alunos.utfpr.edu.br), Eng. Mecânica, UTFPR;

## 3. Obtenção e tratamento de dados

### Tabelas utilizadas

- **Tabela principal - Preço dos combustíveis dos estados brasileiros de 2004 a 2021**
  - **Fontes dos dados:** A tabela foi obtida no *Kaggle*.
  - **Procedimentos de limpeza e processamento:** A tabela com os dados de preço médio de revenda e distribuição dos estados entre 2004 a 2021 foi formatada pelo arquivo "*Leitura\_e\_Procedimento de Limpeza*". Nesse arquivo, primeiramente, foram renomeadas as colunas que possuíam acento e "ç", além de adequar o nome dos óleos (que havia sido cadastrado com dois tipos, com e sem acento) para ficar sem acento em todos os casos. Finalizando as formatações textuais, todos os estados, que estavam escritos por extenso, foram substituídos pelas respectivas siglas.

Além disso, foi observado que seis colunas numéricas estavam com tipo de dado 'objeto', isso também foi ajustado para numérico. Além disso, foram checados os tipos de combustíveis analisados e excluídos da tabela todos os gasosos.

Após isso a tabela foi exportada para CSV com o nome '*br\_oil\_prices\_formatado.csv*'.

- **População por estado**

- **Fontes dos dados:** A tabela foi obtida através de uma base de dados no [Big Query](#), e foi utilizada a seguinte query para obter os dados:  

```
SELECT ano, sigla_uf, SUM(populacao) as populacao FROM `basedosdados.br_ibge_populacao.municipio` GROUP BY sigla_uf, ano;
```

Então foi salvo o resultado em CSV, e não foi necessária limpeza nos dados.

- **PIB dos estados**

- **Fontes dos dados:** A tabela foi obtida através deste [link](#), porém as colunas eram: "Estado, 2004, 2005.... 2019", e o valor das células era o valor do PIB por estado e por ano. Então foi necessário utilizar o comando *melt* do Pandas para realizar a conversão para possuir as colunas: "Estado, ano, pib".
- **Procedimentos de limpeza e processamento:** Foi convertido o nome do estado para a sigla, para seguir o padrão da tabela de População por estado.

- **Preço do barril do petróleo mundial**

- **Fontes dos dados:** A tabela foi obtida no *Kaggle*, os dados abrangiam até o mês de janeiro de 2021, então de fevereiro a maio de 2021 os dados foram coletados de forma manual.
- **Procedimentos de limpeza e processamento:** Foi feita a conversão de coluna para tipo data, filtro do período de datas necessário, média mensal da coluna preço do petróleo.

- **Preço do dólar**

- **Fontes dos dados:** A tabela foi obtida através deste [site](#) que possui histórico da cotação do dólar.
- **Procedimentos de limpeza e processamento:** Na parte de limpeza, colunas que tinham acento no nome foram modificadas, a numeração de valores estava em vírgula e foi trocada por ponto, foi feita conversão de colunas de tipo string para float. Como não havia uma coluna para média diária, foi criada a partir da média do valor mínimo e máximo do dólar naquele dia, em seguida foi feita uma coluna com média mensal. As datas foram convertidas para o formato internacional.

- **Salário mínimo**

- **Fontes dos dados:** A tabela foi preenchida manualmente, compreende 3 colunas: a data que o salário se tornou vigente, a data final e o salário mínimo em reais.
- **Procedimentos de limpeza e processamento:** Por ser uma tabela pequena criada por nós mesmos, não foi necessário realizar a limpeza dos dados.

## 4. Cobertura e distribuição dos dados

- Preço dos combustíveis por estado

A tabela lida *'br\_oil\_prices\_formatado'* foi utilizada pelo código *'Análise\_Exploratória\_PreçoPorEstado'*. O código inicialmente, verifica se não há algum estado que difere dos demais no número de medições.

Após isso foi separado e plotados os gráficos dos preços dos combustíveis pelo tempo, isso foi feito para cada uma das cinco regiões.

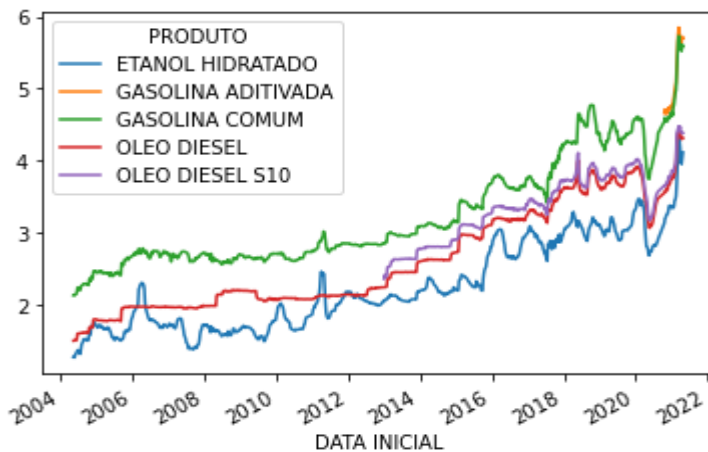


Gráfico 1: Preço dos combustíveis líquidos na região Centro-Oeste.

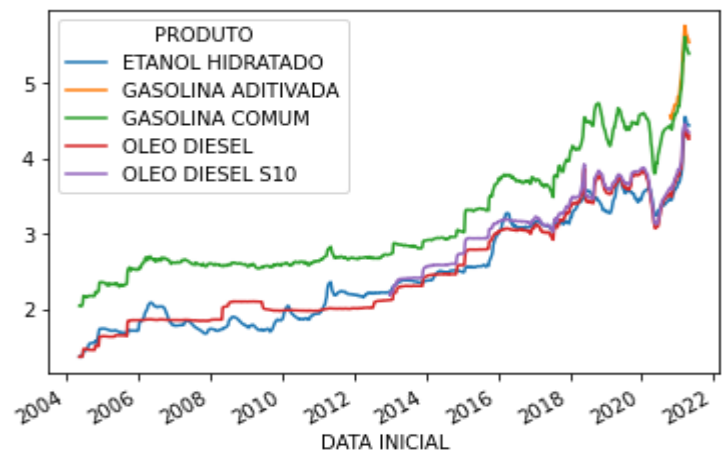


Gráfico 2: Preço dos combustíveis líquidos na região Nordeste.

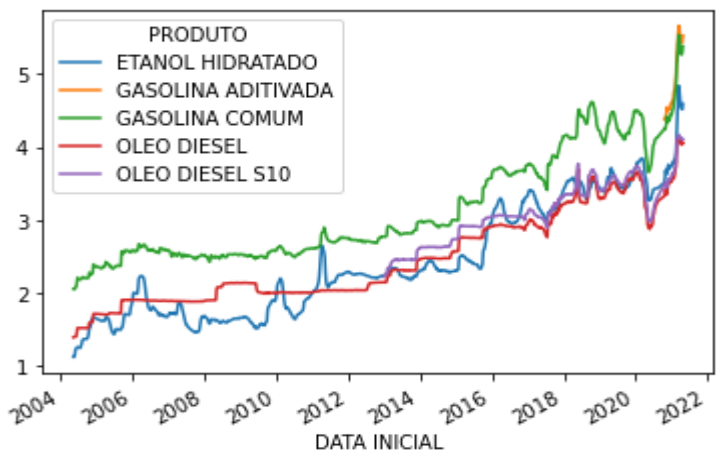


Gráfico 3: Preço dos combustíveis líquidos na região Sul.

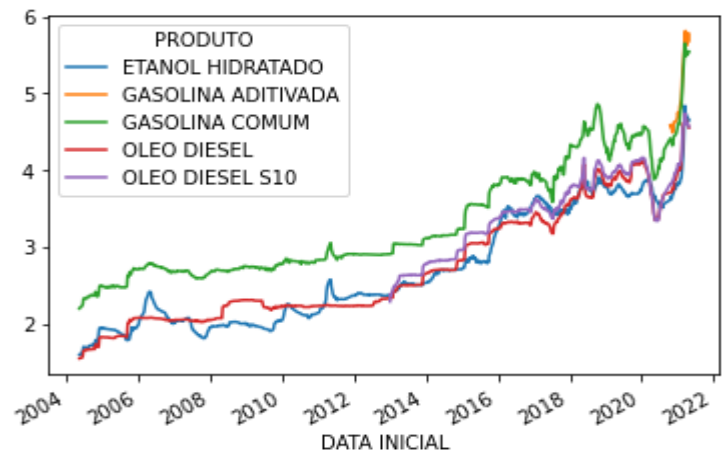


Gráfico 4: Preço dos combustíveis líquidos na região Sudeste.

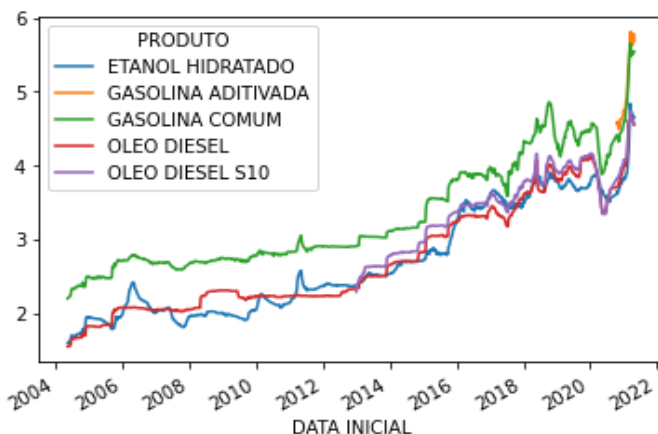


Gráfico 5: Preço dos combustíveis líquidos na região Norte.

Com a análise visual dos gráficos acima é possível perceber alguns comportamentos comuns em todas as regiões, como um aumento repentino em 2011, uma região de instabilidade, com alguns vales e picos entre 2018 e 2020, uma queda repentina entre 2020 e 2021 e um pico bruto em 2022.

Além disso, foram analisadas as diferenças entre o preço médio de revenda e o preço médio de distribuição dos combustíveis, isso é, em termos grosseiros, o lucro bruto dos postos de combustíveis. Para isso, foi criada a correlação '*DELTA*', que é simplesmente a diferença entre as colunas '*PREÇO MÉDIO REVENDA*' e '*PREÇO MÉDIO DISTRIBUIÇÃO*'.

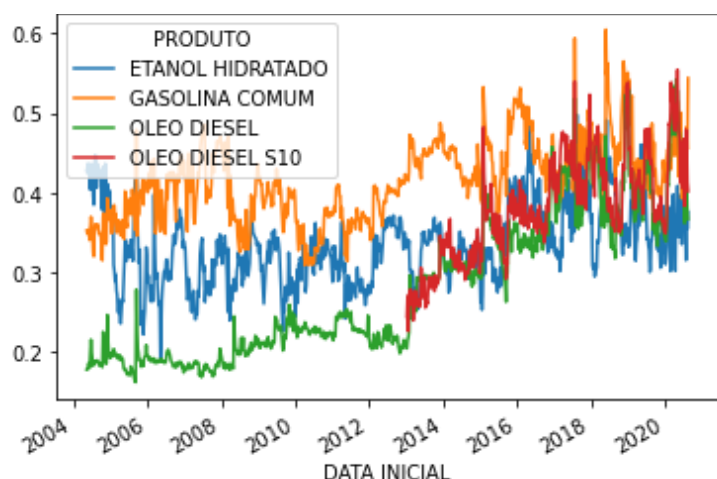


Gráfico 6: Lucro bruto por litro de combustível na região Centro-Oeste.

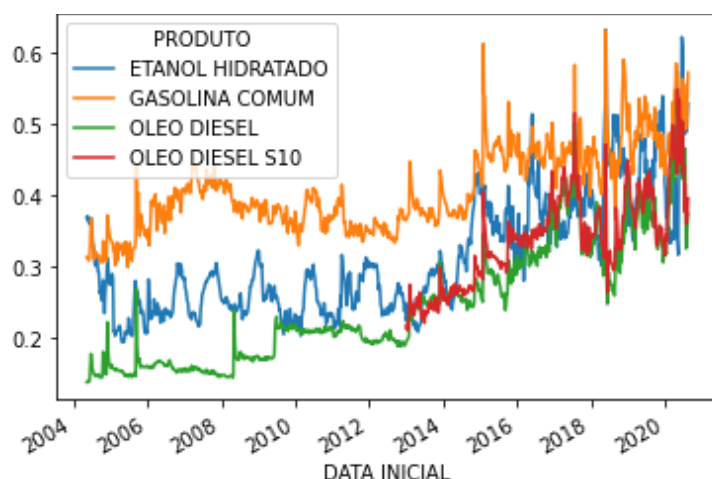


Gráfico 7: Lucro bruto por litro de combustível na região Nordeste.

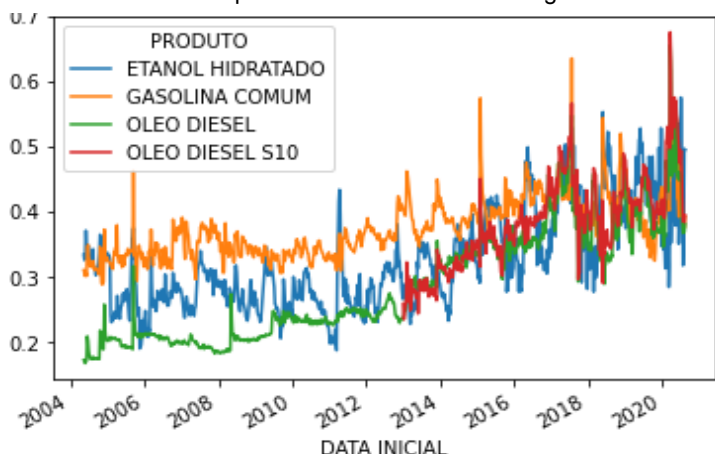


Gráfico 8: Lucro bruto por litro de combustível na região Sul.

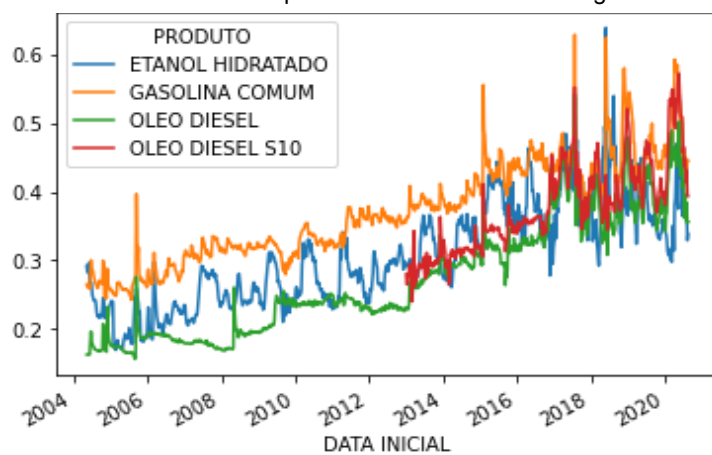


Gráfico 9: Lucro bruto por litro de combustível na região Sudeste.

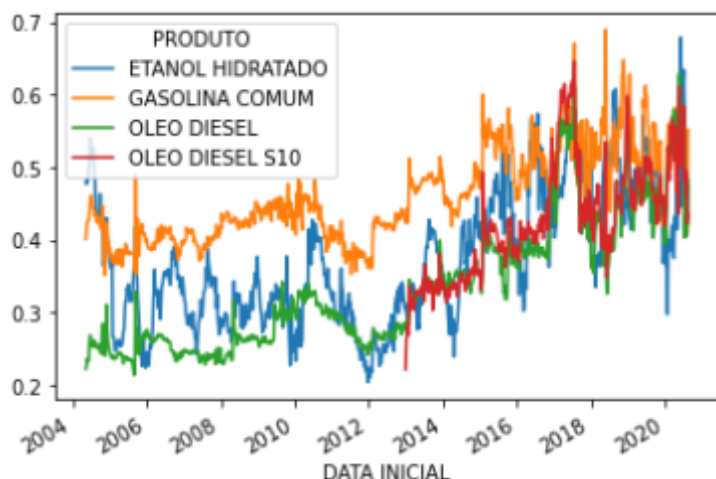


Gráfico 10: Lucro bruto por litro de combustível na região Norte.

A partir desses gráficos, percebe-se que existe uma instabilidade muito alta num período de dois anos, há claramente um pico no ano de 2017 e no início do ano de 2020 há um vale que talvez esteja relacionado com o vale que também ocorreu nos gráficos do preço médio de revenda dos combustíveis.

Ademais, pensando em correlacionar as datas dos combustíveis foi criado, para os combustíveis etanol hidratado, gasolina comum e óleo diesel, uma variável da diferença entre as datas de medição final e inicial do preço dos combustíveis. Essa análise foi feita somente nos três combustíveis citados, pois somente as medições deles foram feitas a partir de 2004.

PRODUTO	PREÇO INICIAL	PREÇO FINAL	DELTA
ETANOL HIDRATADO	1.27525	2.91200	1.63675
GASOLINA COMUM	2.13300	4.38300	2.25000
OLEO DIESEL	1.50250	3.47825	1.97575

Tabela 1: Diferença entre o preço final e inicial da região Centro-Oeste

PRODUTO	PREÇO INICIAL	PREÇO FINAL	DELTA
ETANOL HIDRATADO	1.367333	3.401333	2.034000
GASOLINA COMUM	2.045667	4.368625	2.322958
OLEO DIESEL	1.367778	3.421857	2.054079

Tabela 2: Diferença entre o preço final e inicial da região Nordeste

PRODUTO	PREÇO INICIAL	PREÇO FINAL	DELTA
ETANOL HIDRATADO	1.593286	3.546500	1.953214
GASOLINA COMUM	2.194143	4.387667	2.193524
OLEO DIESEL	1.545429	3.699333	2.153905

Tabela 3: Diferença entre o preço final e inicial da região Norte

PRODUTO	PREÇO INICIAL	PREÇO FINAL	DELTA
ETANOL HIDRATADO	1.05475	3.1370	2.08225
GASOLINA COMUM	1.98575	4.3315	2.34575
OLEO DIESEL	1.38050	3.3310	1.95050

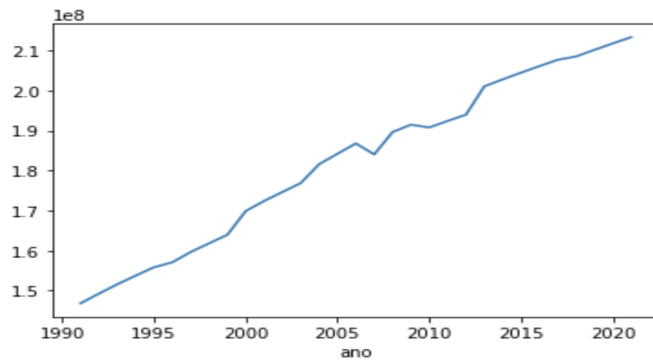
Tabela 4: Diferença entre o preço final e inicial da região Sudeste

PRODUTO	PREÇO INICIAL	PREÇO FINAL	DELTA
ETANOL HIDRATADO	1.125	3.439667	2.314667
GASOLINA COMUM	2.057	4.145333	2.088333
OLEO DIESEL	1.394	3.262667	1.868667

Tabela 5: Diferença entre o preço final e inicial da região Sul

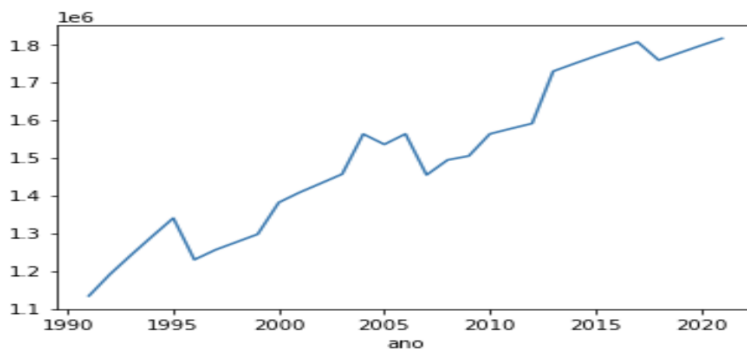
A partir desses dados, podemos perceber que não há uma diferença muito expressiva entre os valores para cada região, com exceção do Etanol na região Centro-Oeste e Sul.

- **População por estado**



A tabela de população por estado possui o intervalo de 1991 até 2021 com valores anuais, e possui os valores para cada estado brasileiro. Foi adicionada uma coluna chamada "*REGIOES*" para associar cada estado à sua respectiva região. O gráfico à esquerda mostra a soma da população de todos os estados por ano.

Gráfico 11: Soma da população de cada estado por ano (em centenas de milhões).



Alguns estados, como Rondônia, possuem um comportamento estranho, como ilustrado no gráfico ao lado.

Gráfico 12: População de Rondônia por ano (em milhões de habitantes).

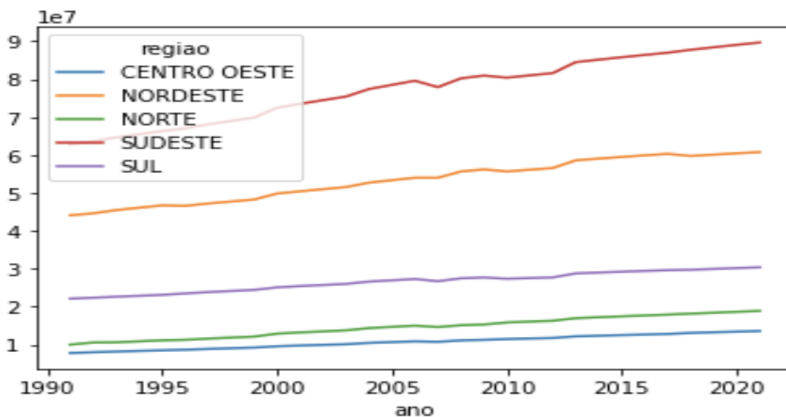


Gráfico 13: População de cada região, por ano (em dezenas de milhões).

- **PIB por estado**

A tabela de PIB por estado brasileiro possui o intervalo de 2002 até 2019 e com valores anuais, e possui os valores para cada estado brasileiro.

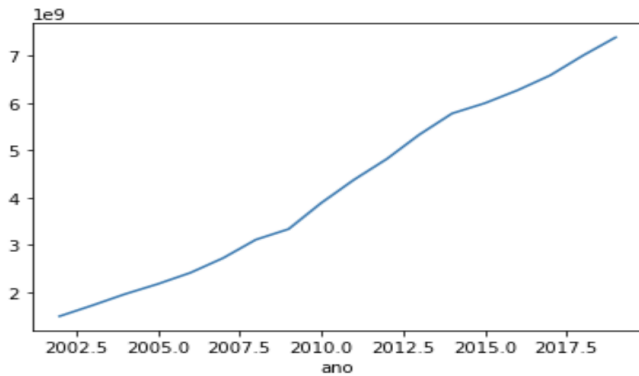


Gráfico 14: Soma do PIB de cada estado, por ano (em trilhões de reais).

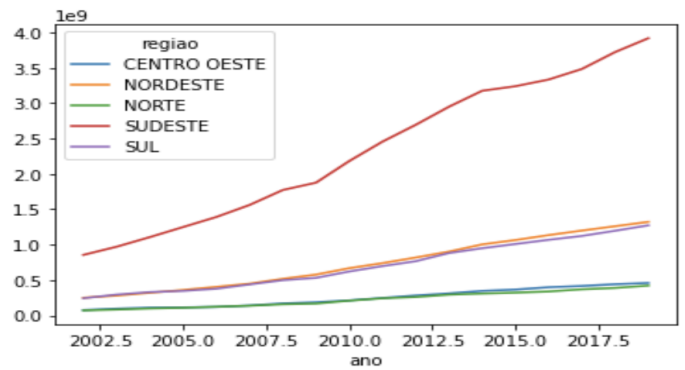


Gráfico 15: PIB de cada região do Brasil, por ano (em trilhões de reais).

- **Salário mínimo**

A tabela de Salário Mínimo brasileiro possui o intervalo de 2004 até 2021.

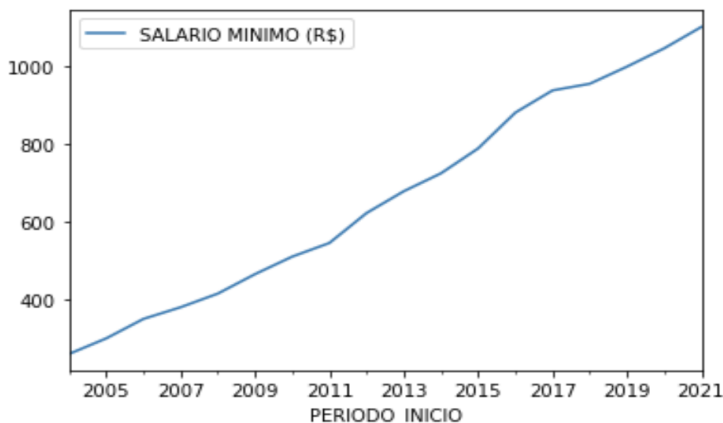


Gráfico 16: Salário mínimo por ano no Brasil (em reais).



Gráfico 17: Salário mínimo em dólar

- **Preço do dólar**

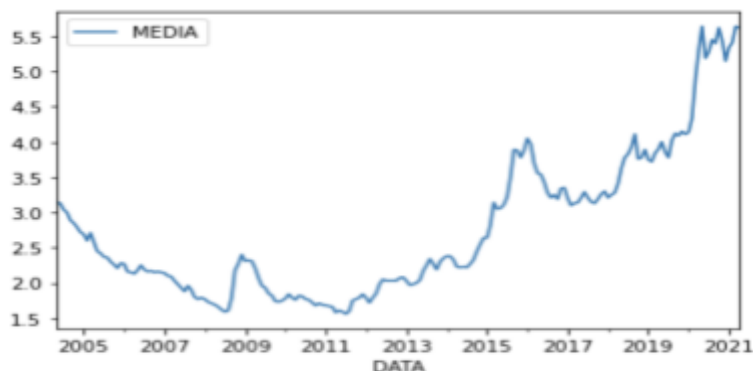


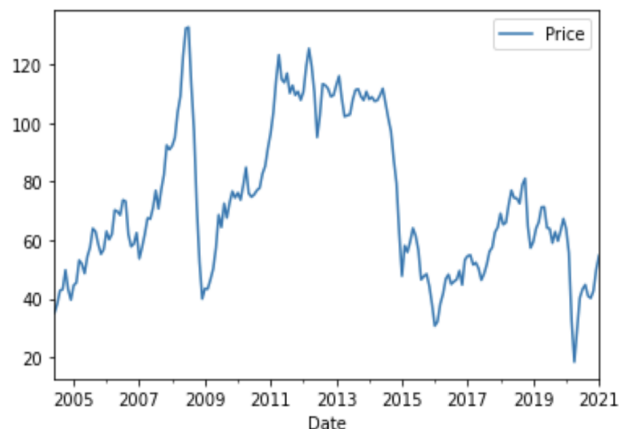
Gráfico 18: Valor do dólar em reais, mensalmente.

A tabela de Preço do Dólar possui o intervalo de 2004 a 2021 com valores mensais.



## 5. Identificação de padrões e perguntas de pesquisa

- **Os combustíveis são mais baratos nos estados que os produzem?**
  - Analisando o gráfico dos combustíveis das cinco regiões, é possível perceber que eles geralmente possuem um comportamento semelhante de variação, mas há alguns casos em que são diferentes. O preço do óleo diesel e do etanol são bastante semelhantes, mas na região Centro-Oeste o preço do etanol foi, na maior parte do tempo, inferior ao do diesel. Talvez isso tenha a ver com o fato de os 3 estados da região estarem entre os 10 maiores produtores nacionais do combustível ([Fonte](#)). Além disso, possuem uma população menor que as dos demais estados, o que torna a demanda menor e, possivelmente, o preço também.
- **Como a pandemia influenciou o preço dos combustíveis?**
  - Os gráficos de 1 a 5 indicam que todos os combustíveis tiveram uma queda brusca logo no início da pandemia, retomaram o valor anterior e atingiram a máxima do período nos últimos meses de dados disponíveis. Então a pandemia influenciou no aumento do preço dos combustíveis.
- **Se considerarmos quantos litros se podia comprar com um salário mínimo em cada ano, o poder de compra do brasileiro aumentou ou diminuiu?**
  - O gráfico à esquerda abaixo demonstra a variação da quantidade de galões de petróleo que se podia comprar com um salário mínimo ao longo dos anos. Se compararmos os extremos é possível perceber que o poder de compra do combustível aumentou, mas ocorreram diversas variações. No período de 2007 a 2009 ocorreu uma queda, o que pode ser explicada pelo momento de crise que o mundo estava passando nesse período. Já a variação mais brutal de todas ocorreu em abril de 2020, um mês após o início da pandemia. O gráfico à direita mostra a variação do preço do barril do petróleo, que caiu muito nesse mês de abril. Isso deve ter acontecido porque, após decreto da pandemia, grande parte das pessoas permaneceram sem sair de casa, consequentemente sem gastar combustível e, por isso, a diminuição da demanda ocasionou na diminuição do preço.





## 6. Discussões

Nesta fase de análise exploratória, foi possível aplicar gráficos para entender o comportamento de algumas variáveis que são essenciais para responder às perguntas do objetivo do trabalho. Foi possível perceber padrões de comportamento entre as regiões, em que os aumentos, picos, estabilidades e instabilidades aconteciam geralmente no mesmo período. Mas em alguns casos houve uma ligeira diferença, como o Etanol que era um pouco mais barato na região Centro-Oeste e um pouco mais caro na região Sul, em comparação às três demais regiões. Dessa forma, com os dados obtidos, há indícios de que as perguntas da pesquisa podem ser respondidas.

No que tange os próximos passos, ainda é necessário:

- 1) Relacionar o preço do barril do petróleo mundial com o preço dos combustíveis nos estados. Quando o preço do barril de petróleo mundial crescia, o preço dos combustíveis nos estados cresciam também?
- 2) Relacionar o preço do barril do petróleo com o preço do dólar. Necessariamente quando o preço do dólar aumentava, o preço do barril do petróleo aumentava também?
- 3) Clusterizar os dados entre o preço médio de distribuição e o preço médio de revenda nos estados e relacionar com o índice de Gini para avaliar se nos estados com maior desigualdade social, o lucro bruto dos postos de combustíveis também é maior.
- 4) Obter os dados de índice de Gini dos estados.