



PROJETO FINAL MÓDULO 4

Análise exploratória relacionada à série histórica de preço de venda
da gasolina e do etanol



NOSSA EQUIPE



GABRIEL

Dev Team



JEAN

Dev Team



ODILON

Dev Team



PAULO H.

Scrum
Master



ROBERTO

Product
Owner

OBJETO DE ESTUDO

As fontes de dados utilizadas na análise a seguir estão disponíveis no portal dados.gov. Foram utilizados os arquivos dos meses de junho e julho de 2021 contendo a série histórica dos preços da gasolina e do etanol em todo o Brasil.

MATERIAL E MÉTODOS:

Como ferramenta de apoio à análise foi utilizado o Colab apoiado pela biblioteca Pandas;

ETAPAS:

Limpeza dos dados não importantes;
Análise exploratória;



EXECUÇÃO DO TRABALHO - ETAPAS

01

DEFINIÇÃO DAS MACRO TAREFAS

Análise;
GitHub;
Apresentação;
Trello.

02

DIVISÃO DAS TAREFAS POR INTEGRANTE

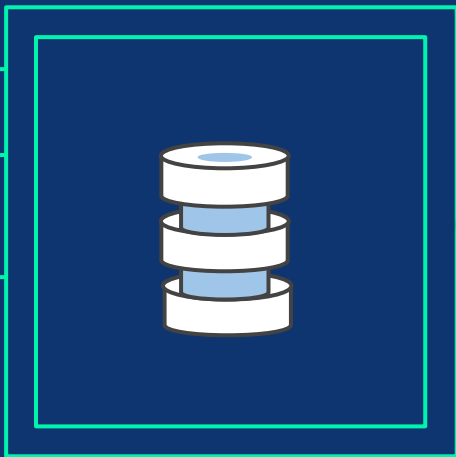
2 perguntas por
integrante;
Papel de cada um
no Grupo.

03

ANÁLISE DOS DADOS

Análise
exploratória dos
dados e
informações
obtidas

RESPOSTAS ÀS PERGUNTAS



RESULTADOS

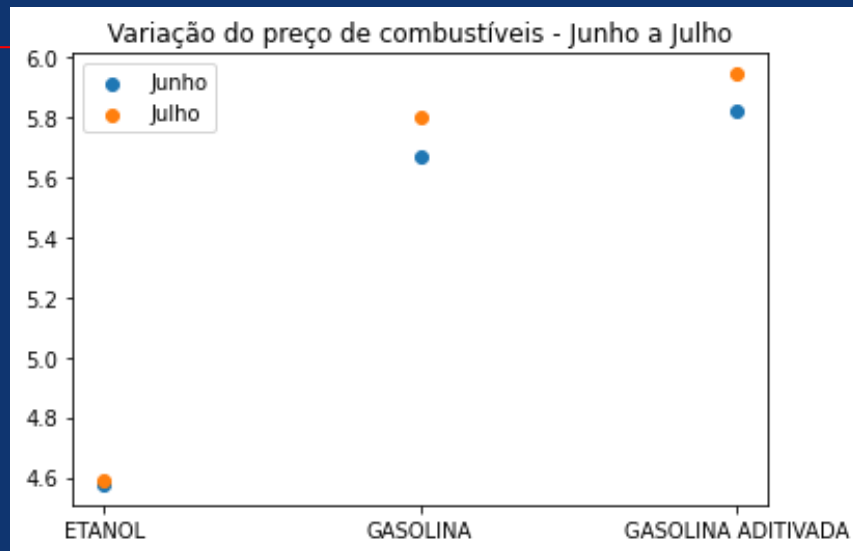


1) Como se comportaram o **preço dos combustíveis** durante os dois meses citados? Os valores do etanol e da gasolina **tiveram uma tendência de queda ou diminuição?**

```
preco_medio_junho = junho.groupby(["Produto"])[\"Valor de Venda\"].mean() #agrupando combustíveis e valor  
preco_medio_junho #calculando a média
```

```
preco_medio_julho = julho.groupby(["Produto"])[\"Valor de Venda\"].mean() #agrupando combustíveis e valor  
preco_medio_julho#calculando a média
```

Resposta: Tendência de alta.





RESULTADOS



2) Qual o preço médio da gasolina e do etanol nesses dois meses?

Filtrando o campo Gasolina - Junho e Julho

```
junho[junho['Produto'] == 'GASOLINA'].shape[0] #criando uma série e retornando o total de linhas de gasolina
```

```
julho[julho['Produto'] == 'GASOLINA'].shape[0] #criando uma série e retornando o total de linhas de gasolina
```

```
gasolina_junho = junho[junho['Produto'] == 'GASOLINA']  
gasolina_junho.mean() #visualizando só o campo gasolina
```



Valor de Venda 5.671506

```
gasolina_julho = julho[julho['Produto'] == 'GASOLINA']  
gasolina_julho.mean() #visualizando só o campo gasolina
```



Valor de Venda 5.800254

Resposta: Variação de R\$ 0,13 centavos para a Gasolina



RESULTADOS

3) Quais os 5 estados com o preço médio da gasolina e do etanol mais caros?

```
# dataframe_JyJ['Valor de Venda']=dataframe_JyJ['Valor de Venda'].str.replace(',','.')
dataframe_JyJ['Valor de Venda']=dataframe_JyJ['Valor de Venda'].astype(float)
etanol = dataframe_JyJ[dataframe_JyJ['Produto']=='ETANOL'].groupby(['Estado - Sigla', 'Produto']).agg( {"Valor de Venda":"mean"}).round(2).sort_values('Valor de Venda',ascending=False).head(5)
gasolina= dataframe_JyJ[dataframe_JyJ['Produto'].str.contains('GASOLINA$')].groupby(['Estado - Sigla', 'Produto']).agg( {"Valor de Venda":"mean"}).round(2).sort_values('Valor de Venda',ascending=False).head(5)
# gasolina_adt= dataframe_JyJ[dataframe_JyJ['Produto'].str.contains('GASOLINA AD')].groupby(['Estado - Sigla', 'Produto']).agg( {"Valor de Venda":"mean"}).round(2).sort_values('Valor de Venda',ascending=False).head(5)
```

```
cinco_estados = pd.concat([etanol, gasolina], join="outer")
cinco_estados
```

Resposta:



Estado – Sigla	Produto	R\$
RS	ETANOL	5.81
AC	ETANOL	5.54
RN	ETANOL	5.52
PA	ETANOL	5.33
AP	ETANOL	5.31
AC	GASOLINA	6.44
RJ	GASOLINA	6.30
RN	GASOLINA	6.15
PI	GASOLINA	6.11
GO	GASOLINA	6.04



RESULTADOS

4) Qual o preço médio da gasolina e do etanol por estado?

```
mediaestado=dataframe_JyJ.groupby(['Produto', 'Estado - Sigla'])['Valor de Venda'].mean().round(2)  
mediaestado.unstack()[2]
```

	AC	AL	AM	AP	BA	CE	DF	ES	GO	MA	MG	MS	MT	PA	PB	PE	PI	PR	RJ	RN	RO	RR	RS	SC	SE	SP	TO
ETANOL	5.54	5.19	4.45	5.31	4.94	5.20	4.88	5.10	4.46	4.96	4.44	4.56	4.12	5.33	4.90	5.08	5.06	4.35	5.31	5.52	5.08	4.86	5.81	5.02	5.05	4.14	5.13
GASOLINA	6.44	5.96	5.56	4.99	5.96	5.81	5.85	5.86	6.04	5.59	6.02	5.82	5.81	5.92	5.51	5.79	6.11	5.52	6.30	6.15	5.85	5.42	6.02	5.49	5.69	5.46	5.93

Resposta: Ao analisar, notamos que o **Acre** possui a maior média de preço para **Gasolina**, no valor de **6,44** e o estado do **Rio Grande do Sul**, possui a maior média para **Etanol** no valor de **5,81**



RESULTADOS

5) Qual o município que possui o menor preço para a gasolina e para o etanol?

Criando Tabela Gasolina:

```
gasolina = ['GASOLINA', 'GASOLINA ADITIVADA']  
selecao = nova_tab['Produto'].isin(gasolina)  
tb_gasolina = nova_tab[selecao]  
tb_gasolina
```

Criando Tabela Etanol:

```
etanol = ['ETANOL']  
selecao = nova_tab['Produto'].isin(etanol)  
tb_etanol = nova_tab[selecao]  
tb_etanol
```

Buscando o menor preço para Gasolina e ETANOL:

```
tb_gasolina.min()
```

```
tb_etanol.min()
```

Junho:

Resposta:

Município ADAMANTINA
Produto **GASOLINA**
Valor de Venda 4,199
dtype: object

Resposta:

Município ADAMANTINA
Produto **ETANOL**
Valor de Venda 3,097
dtype: object

Julho:

Resposta:

Município ABAETETUBA
Produto **GASOLINA**
Valor de Venda 4,549
dtype: object

Resposta:

Município ABAETETUBA
Produto **ETANOL**
Valor de Venda 3,189
dtype: object



RESULTADOS

6) Qual o município que possui o maior preço para a gasolina e para o etanol?

Junho:

Resposta:

Município VOTUPORANGA
Produto **GASOLINA**
Valor de Venda 6,999
dtype: object

Resposta:

Município VOTUPORANGA
Produto **ETANOL**
Valor de Venda 6,999
dtype: object

Julho:

Resposta:

Município XANXERE
Produto **GASOLINA**
Valor de Venda 7,097
dtype: object

Resposta:

Município XANXERE
Produto **ETANOL**
Valor de Venda 6,999
dtype: object

Questão 05 - Entre os meses analisados o município de Adamantina - SP apresentou os menores valores de combustíveis: R\$ 4,19 e R\$ 3,09 respectivamente Gasolina e Etanol.

Questão 06 - O município de Xanxere - SC apresentou os maiores valores de combustíveis: R\$ 6,99 e R\$ 7,09, respectivamente Etanol e Gasolina.

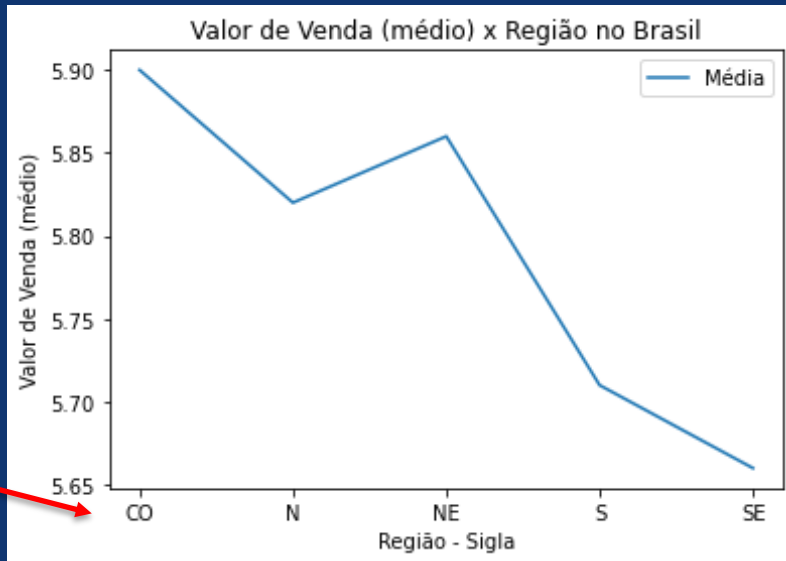


RESULTADOS

7) Qual a região que possui o maior valor médio da gasolina?

```
mediagasolina = df.loc[df['Produto'] == 'GASOLINA'].groupby('Região - Sigla').mean().round(2)
mediagasolina.head()
```

```
plt.plot(mediagasolina, label = 'Média')
plt.title('Valor de Venda (médio) x Região no Brasil')
plt.xlabel('Região - Sigla')
plt.ylabel('Valor de Venda (médio)')
plt.legend()
plt.show()
```



Região - Sigla	Valor de Venda
CO	5.90
N	5.82
NE	5.86
S	5.71
SE	5.66

Resposta: Centro Oeste



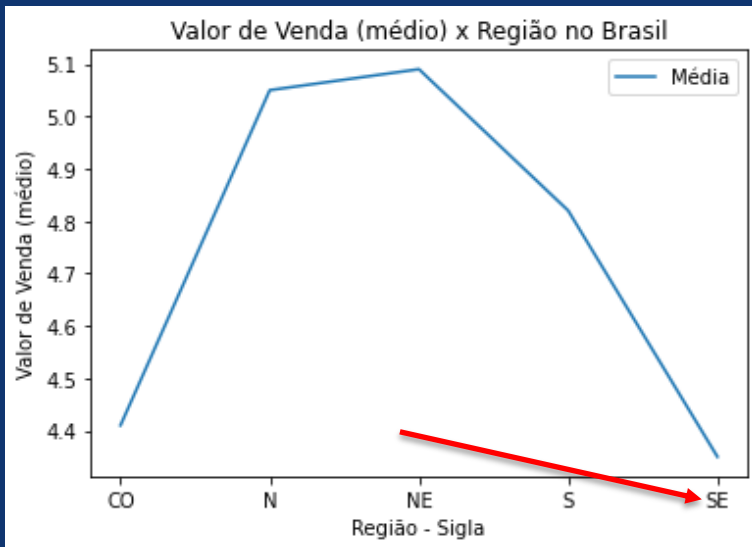
RESULTADOS

8) Qual a região que possui o menor valor médio do etanol?

```
mediaetanol = df.loc[df['Produto'] == 'ETANOL'].groupby('Região - Sigla').mean().round(2)
mediaetanol.head()
```

```
plt.plot(mediaetanol, label = 'Média')
plt.title('Valor de Venda (médio) x Região no Brasil')
plt.xlabel('Região - Sigla')
plt.ylabel('Valor de Venda (médio)')
plt.legend()
plt.show()
```

Resposta: Centro Oeste



Região - Sigla	Valor de Venda
CO	4.41
N	5.05
NE	5.09
S	4.82
SE	4.35



RESULTADOS

9) Há alguma correlação entre o valor do combustível (gasolina e etanol) e a região onde ele é vendido?

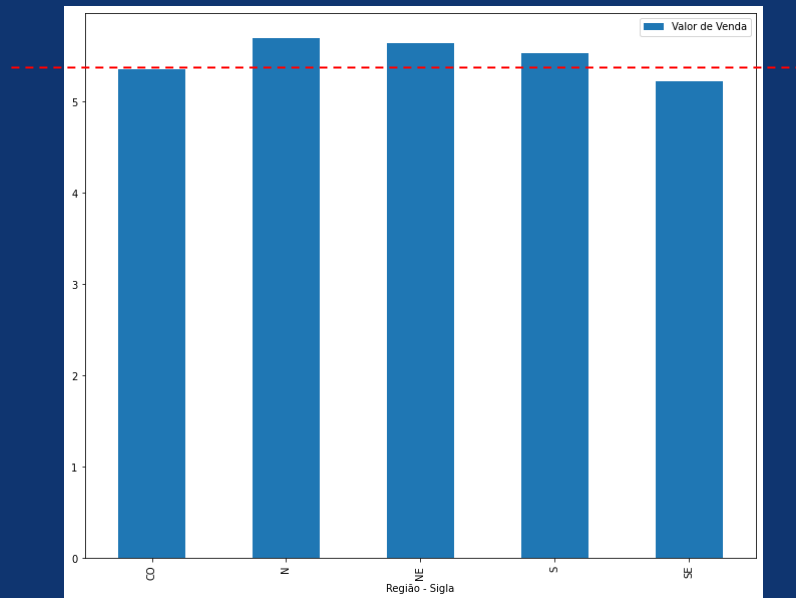
Cálculo da Média por região



```
df.groupby('Região - Sigla').mean()
```

```
df.groupby('Região - Sigla').mean().plot.bar(figsize=(12, 10))
```

Região - Sigla	Valor de Venda
CO	5.349408
N	5.686657
NE	5.633876
S	5.525997
SE	5.223829





RESULTADOS

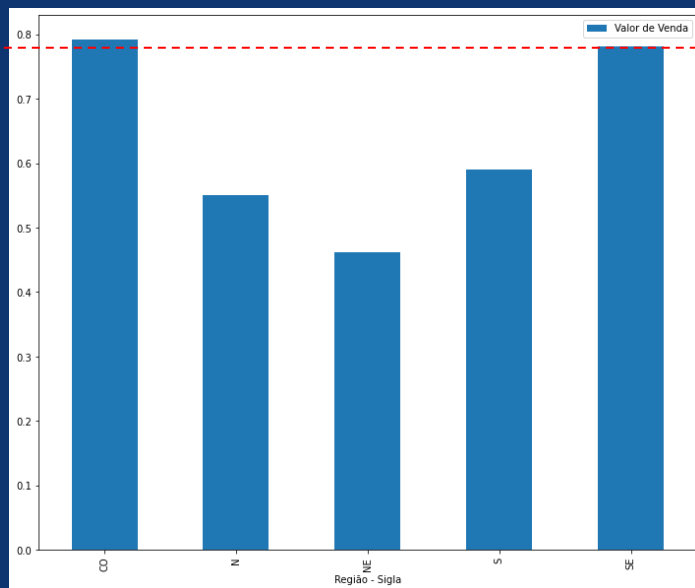
Analisando o **desvio Padrão do valor de Venda** dos combustíveis, é possível observar que a **região Centro Oeste é a que apresenta o maior valor, seguida da região Sudeste, Sul, Norte e Nordeste, respectivamente**. Por essa análise, onde se considera o preço de todos os combustíveis analisados (Gasolina Aditivada, Gasolina e Etanol), podemos observar que **a busca por preços mais baixos por parte do consumidor compensa mais na região Centro Oeste (maior variação) do que na região Nordeste (menor variação)**.

Cálculo do Desvio Padrão por região



```
df.groupby('Região - Sigla').std()
```

Região - Sigla	Valor de Venda
CO	0.791345
N	0.550136
NE	0.462555
S	0.590658
SE	0.781918





RESULTADOS

Região Sudeste com menor preço médio de venda, seguido do Centro Oeste

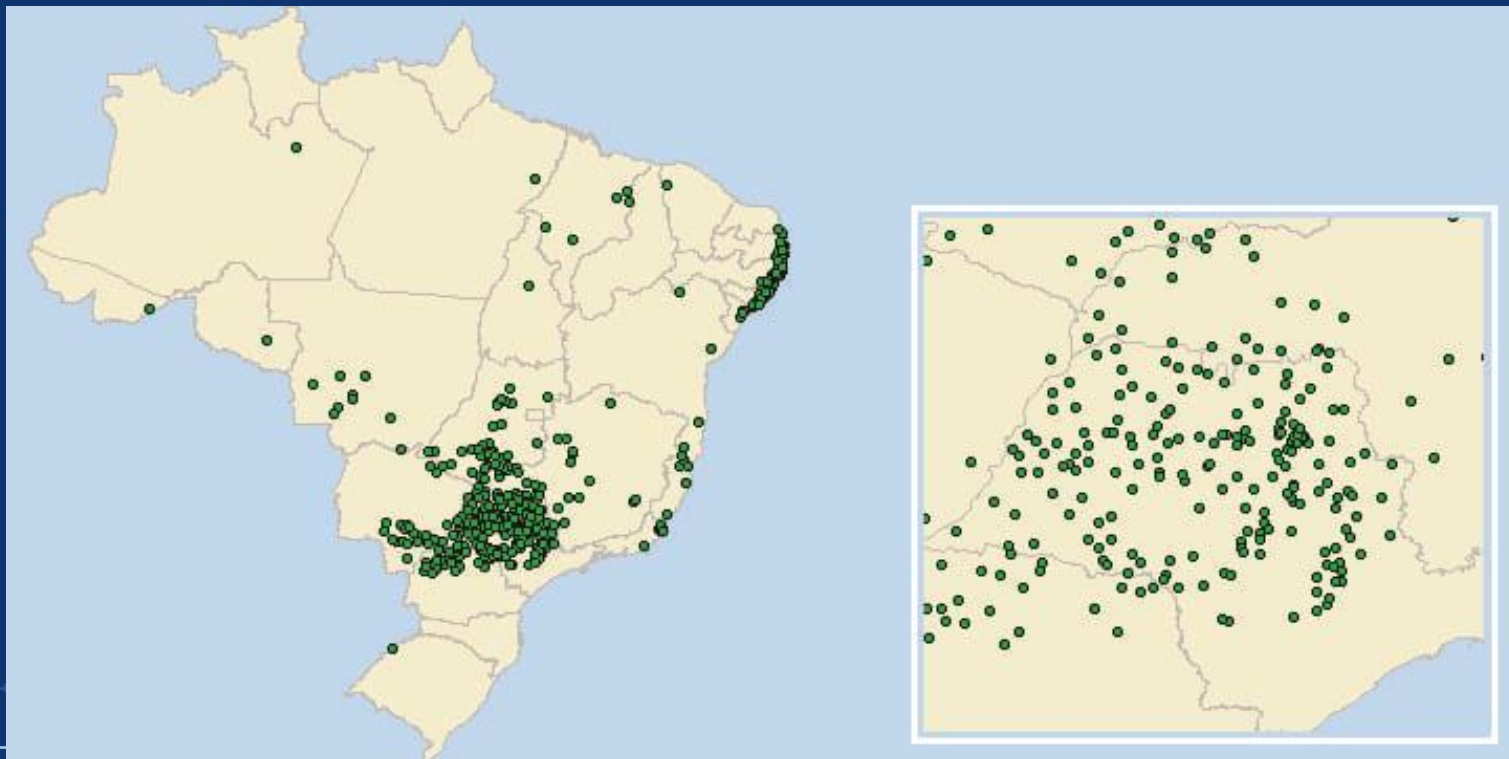
- Concentração do complexo sucroalcooleiro brasileiro na região Centro-Sul (1° no Sudeste, sobretudo em São Paulo, e 2° no Centro Oeste que conta com Goiás e Mato Grosso).
- A proximidade reflete nos custos de frete (maior parte do frete de etanol é feito via rodoviária https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-615/NT-EPE-DPG-SDB-2021-03_Analise_de_Conjuntura_dos_Bicombustiveis_ano_2020.pdf);
- ICMS (com as menores alíquotas em parte do Sudeste e parte do Centro Oeste, onde podem ser metade do percentual de outras regiões com mínimo de 12% e Máximo de 30%).
- Apesar do Nordeste também possuir uma boa quantidade de usinas de Etanol, o ICMS do NE pressiona os valores para cima.
- Com relação à Gasolina (comum e aditivada) o padrão se "repete", com Sudeste com preços mais baixos provavelmente devido à localização de refinarias, menores custo de frete entre as refinarias e os postos de abastecimento e a maior presença da infraestrutura de distribuição por dutos, além da questão do ICMS já descrita acima.



RESULTADOS

Distribuição do setor Sucroalcooleiro no Brasil

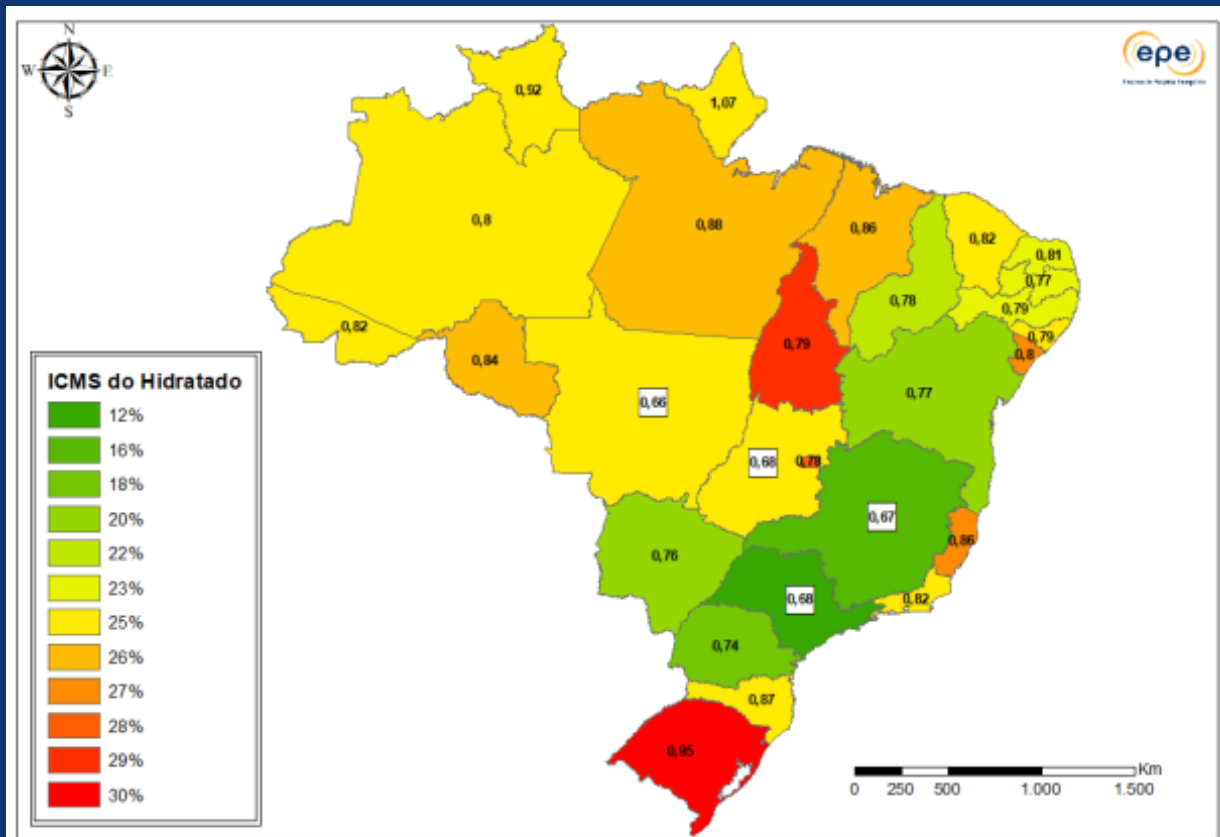
Fonte: <https://gismaps.com.br/en/downloads/usinas-de-acucar-e-alcool-no-brasil-shp/>





RESULTADOS

Alíquotas de ICMS no Etanol Combustível



Fonte: EPE a partir de (ANP, 2021b), (CONFAZ/MF, 2020) e (FECOMBUSTIVEIS, 2021)



RESULTADOS

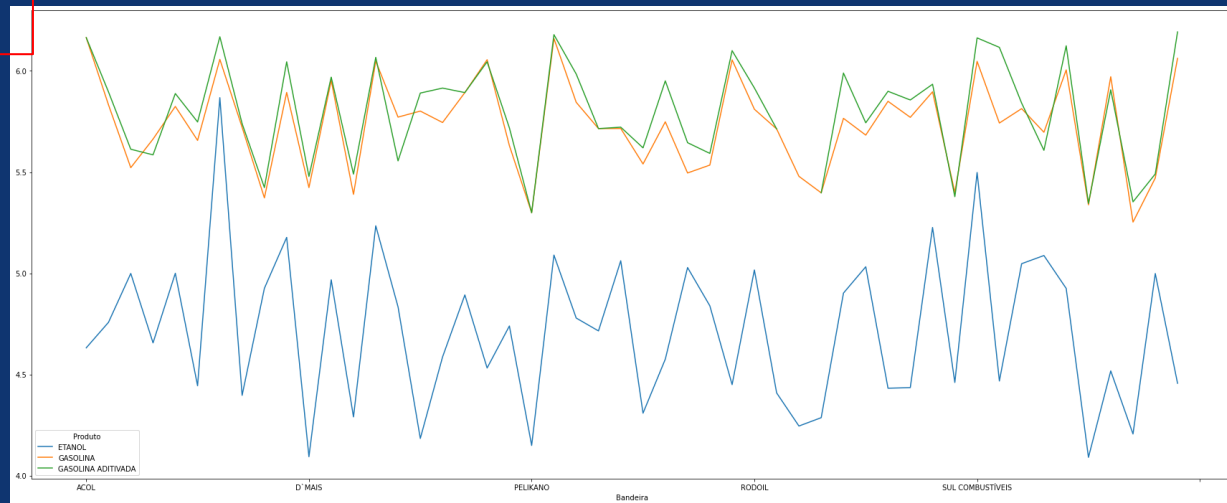
10) Há alguma correlação entre o valor do combustível (gasolina e etanol) e a bandeira que vende ele?

Exibindo os valores de todos os combustíveis por bandeira

```
df.groupby(['Bandeira', 'Produto'])['Valor de Venda'].mean()  
valores_por_bandeira = df.groupby(['Bandeira', 'Produto'])['Valor de Venda'].mean().unstack()
```

```
valores_por_bandeira.plot(figsize=(30,12))
```

Resposta: Sim. Basta observar os picos e vales do gráfico



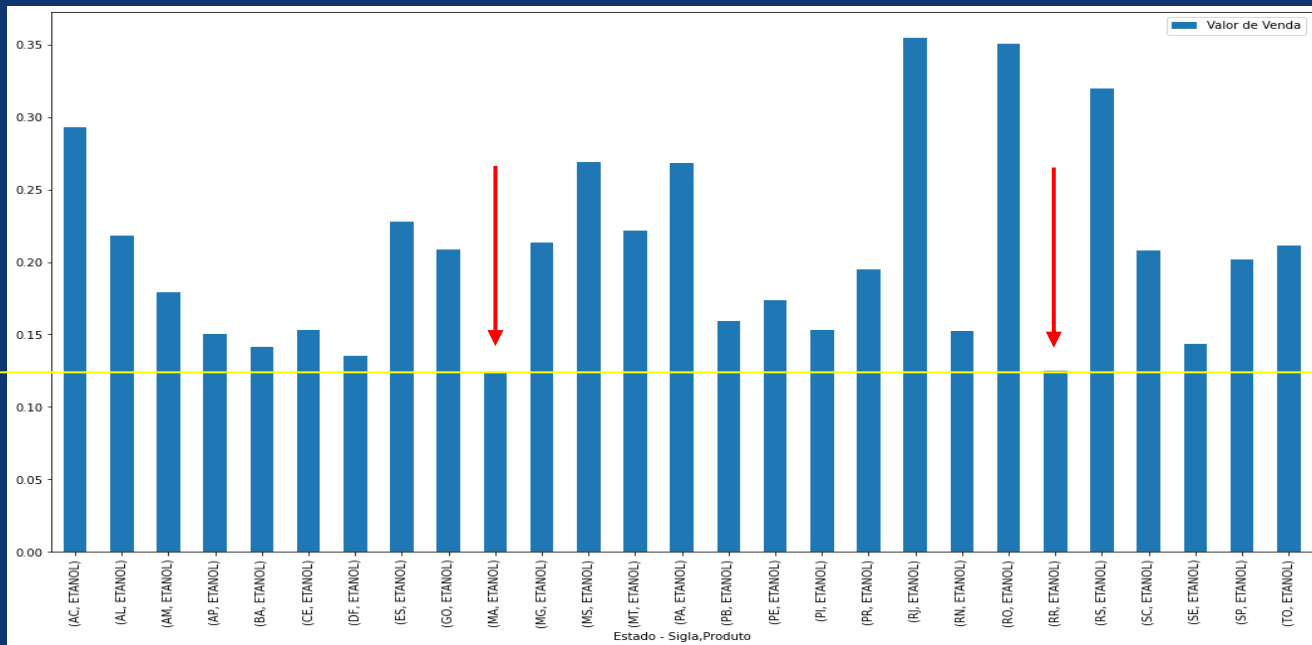


RESULTADOS

11) PERGUNTA EXTRA 1: QUAL O DESVIO PADRÃO DO ETANOL POR ESTADO?

```
df[df['Produto'].str.contains('ETANOL')].groupby(['Estado - Sigla', 'Produto']).std()
```

Resposta: O Desvio Padrão para o Preço do ETANOL por Estado revela que o MA seguido de RR possuem a menor variação de preço, enquanto o Rio de Janeiro possui a maior variação.





RESULTADOS

12) PERGUNTA EXTRA 2: QUAL O DESVIO PADRÃO DA GASOLINA E DO ETANOL POR REGIÃO?

Desvio padrão da Gasolina (Comum e Aditivada)

```
df[df['Produto'].str.contains('GASOLINA')].groupby(['Região - Sigla', 'Produto']).std()
```

Desvio padrão do Etanol

```
df[df['Produto'].str.contains('ETANOL')].groupby(['Região - Sigla', 'Produto']).std()
```

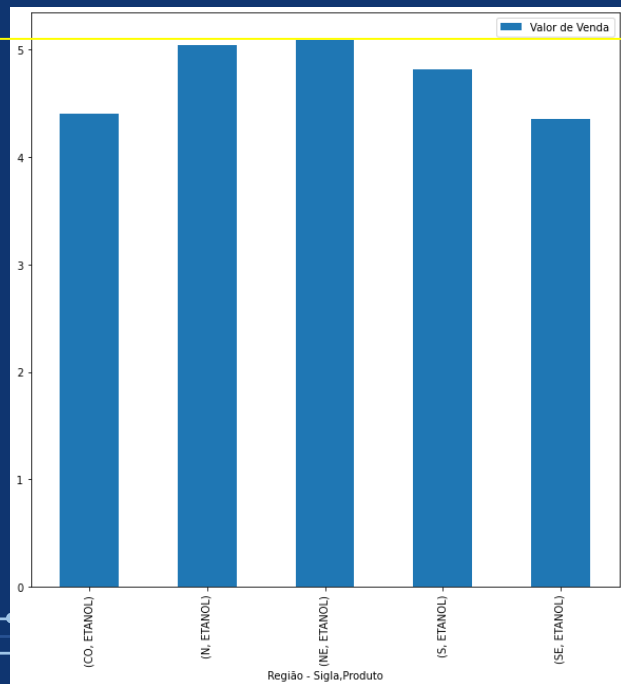
Resposta: O Cálculo do **Desvio Padrão da GASOLINA por região** revela que a região que apresenta o maior valor de Desvio Padrão para a **GASOLINA COMUM** é a **região Norte**, seguida das regiões SE, S, NE e CO, respectivamente. Para a **GASOLINA ADITIVADA**, o maior valor de Desvio Padrão fica para o **SE**, seguido do N, S, NE e CO, respectivamente.

O **Desvio Padrão para o Preço do ETANOL por região** revela que a **região Sul Possui a maior variação**. A **região Nordeste**, que possui a maior Média, possui o menor Desvio padrão.

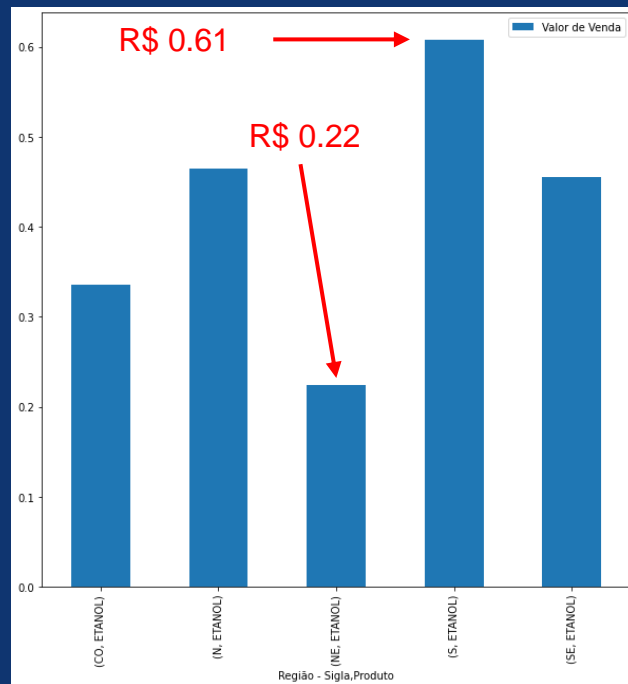


RESULTADOS

Esses resultados apresentam uma situação mais complicada para a região Nordeste. Além de ter preço médio maior, existe uma menor variação nos preços o que revela que o consumidores possuem menos chances de encontrar preços significativamente abaixo da média.



R\$5,09



THANKS!



Perguntas?

@PH, @Roberto,
@Gabriel, @Jean,
@Odilon