

Analizando evolução do preço da gasolina no Brasil

De 2013 até 2023

Se há um produto que causa comoção a qualquer indício de variação, esse produto é a gasolina. Protagonista de muitas notícias e polêmicas, desde assuntos como privatização até estratégia governamental, ela sempre está no meio.

Tendo isso em mente, não seria interessante investigar como esse produto se comportou ao longo dos anos no nosso país? Será que de fato ficou mais cara? Se sim, o quanto mais cara? Que efeitos a pandemia teve sobre seu valor? Essas são algumas dúvidas iniciais que tinha no início deste trabalho.

Todos os dados coletados estão disponíveis pela ANP no link



[Série Histórica de Preços de Combustíveis e de GLP](#)

E todos os códigos utilizados estão disponíveis no meu GitHub pelo link



https://github.com/jorgecarvalhopi/Python_Analise-Precos-Gasolina-Comum.git

Dito isso, vamos iniciar os trabalhos.

O primeiro passo se resumiu em carregar os dados. Estamos falando de cerca de 1.6GB de arquivos .csv, ferramentas mais tradicionais como Excel não seriam eficientes para essa tarefa, por isso o escolhido da vez foi a linguagem Python juntamente com o JupyterLab.

Após isso adicionamos algumas colunas que ajudarão na nossa análise, como a coluna de margem em percentual, esse é um valor que define quantos porcentos (%) do seu lucro bruto com a venda de um item representou do seu faturamento com ele.

Com isso temos uma base com a seguinte forma.

```
<class 'pandas.core.frame.DataFrame'>
Index: 2703118 entries, 0 to 431575
Data columns (total 16 columns):
#   Column                Dtype
---  -
0   i»¿Regiao - Sigla      object
1   EstadoSigla           object
2   Municipio             object
3   Revenda               object
4   CNPJRevenda           object
5   Bairro                object
6   Produto               object
7   DataColeta            datetime64[ns]
8   ValorVenda            float64
9   ValorCompra           float64
10  Bandeira              object
11  Regiao - Sigla        object
12  mes                   int64
13  ano                   int64
14  dia                   int64
15  mrg                   float64
dtypes: datetime64[ns](1), float64(3), int64(3), object(9)
memory usage: 350.6+ MB
```

A coluna chamada “Região Sigla” que está duplicada será descartada para as análises. Também vale destacar que temos na base carregada valores faltantes.

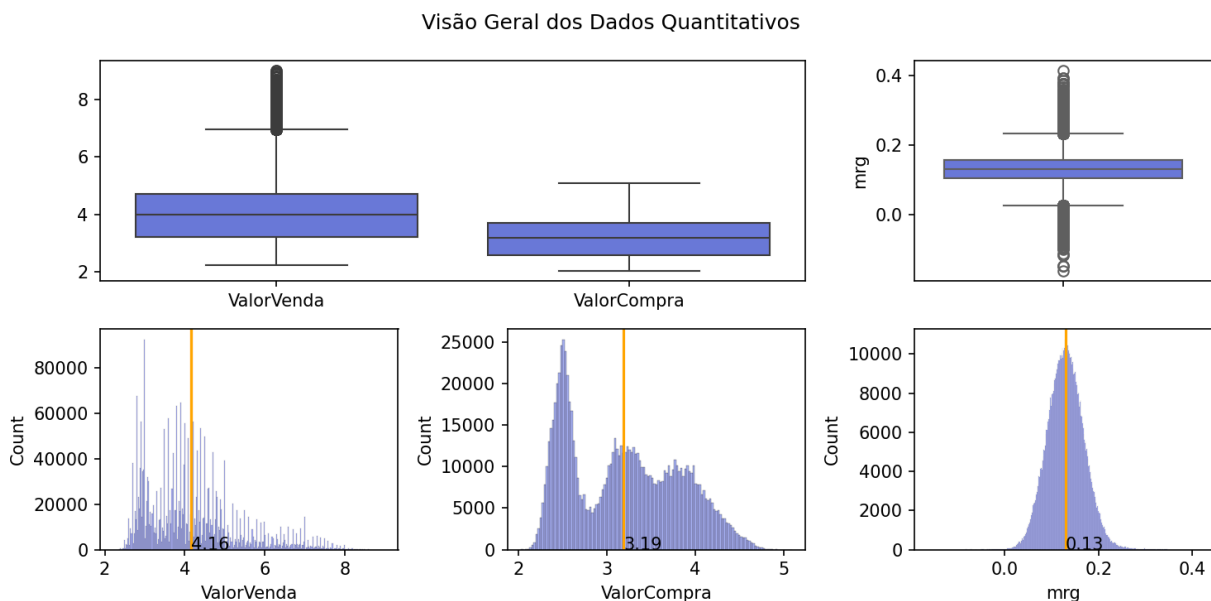
```

i»Regiao - Sigla      119166
EstadoSigla          0
Municipio            0
Revenda              0
CNPJRevenda          0
Bairro               8729
Produto              0
DataColeta           0
ValorVenda           0
ValorCompra          1777820
Bandeira             0
Regiao - Sigla       2583952
mes                  0
ano                  0
dia                  0
mrg                  1777820
dtype: int64

```

O ponto mais relevante nesse caso se encontra na variável Valor Compra, que conta com 1.777.820 valores faltantes, mas será possível lidar com eles depois.

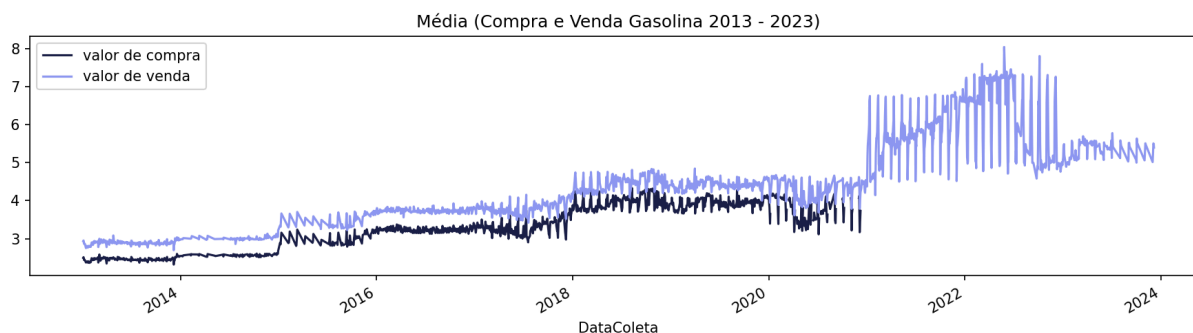
Realizada essa fase preparação, vamos iniciar com uma análise descritiva das nossas variáveis.



Olha só que interessante, perceba que tanto o Histograma como o Boxplot da variável **mrg (margem)** possuem valores extremamente centrados na média, sim possuímos outliers, mas a maior parte das informações está centrada. Contudo, o mesmo não pode ser dito do **Valor de Compra e Venda**.

Ambos possuem grandes variações, o que gera os picos distintos, no caso do valor de compra temos um pico centrado entre 2 e 3, outro entre 3 e 3.5 e um se forma levemente entre 3.7 e 4, o que pode indicar que durante a evolução dessa série existe uma variação significativa.

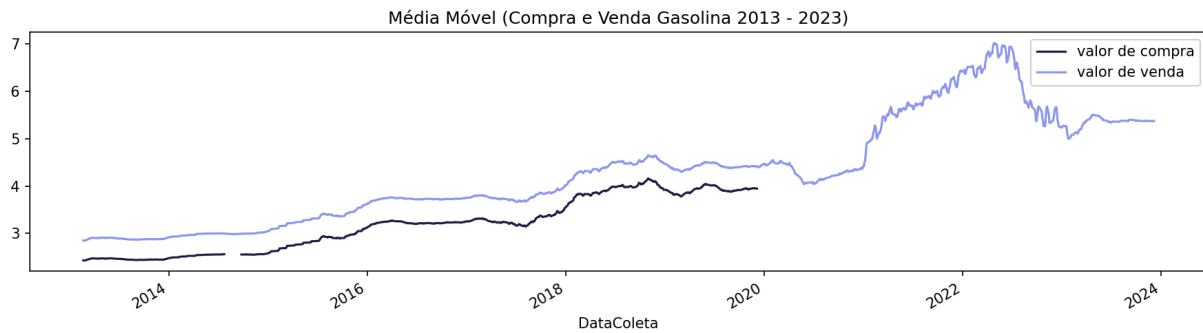
Para visualizar isso melhor, foram criados os gráficos abaixo.



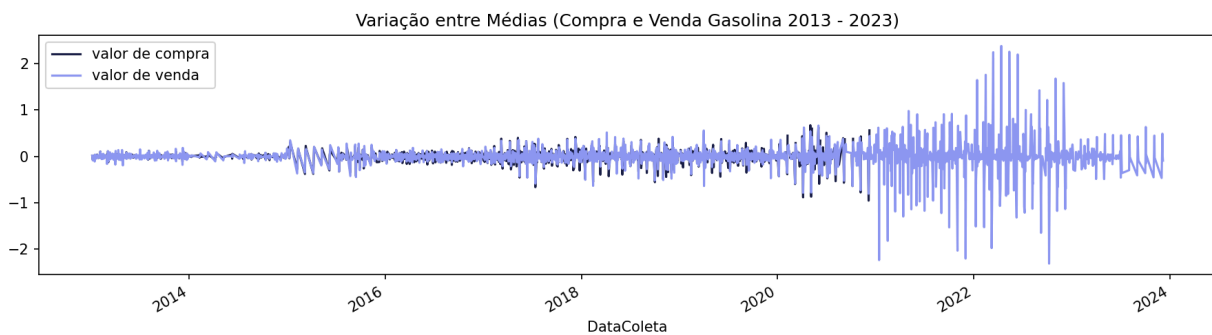
Bingo! Percebemos um claro aumento nos valores que começa a surgir a partir de 2016 e explode após 2020. Reparem que temos aqui a falta visual dos valores de compra, mas vamos revisitar essas análises após simular esses números. Mas voltando ao aumento que observamos, temos alguns possíveis motivos que afetaram isso, medidas governamentais por exemplo.

Vamos relembrar, no Brasil, a Petrobras é a principal responsável pela distribuição de combustíveis, sim, temos outras companhias atuando em alguns pontos da cadeia, mas ela ainda é a dominante. Com isso em mente, medidas que afetem sua operação, afetam esse produto.

Em 2016 tivemos a introdução do PPI (preço de paridade de importação), o que fez com que o preço dos produtos vendidos pela empresa buscassem estar pareados com os preços internacionais, não entrarei em detalhes de como o processo funciona, se é correto ou não, mas fato é que após 2016 conseguimos ver, principalmente pelo gráfico das Média Móvel, o aumento dos valores.

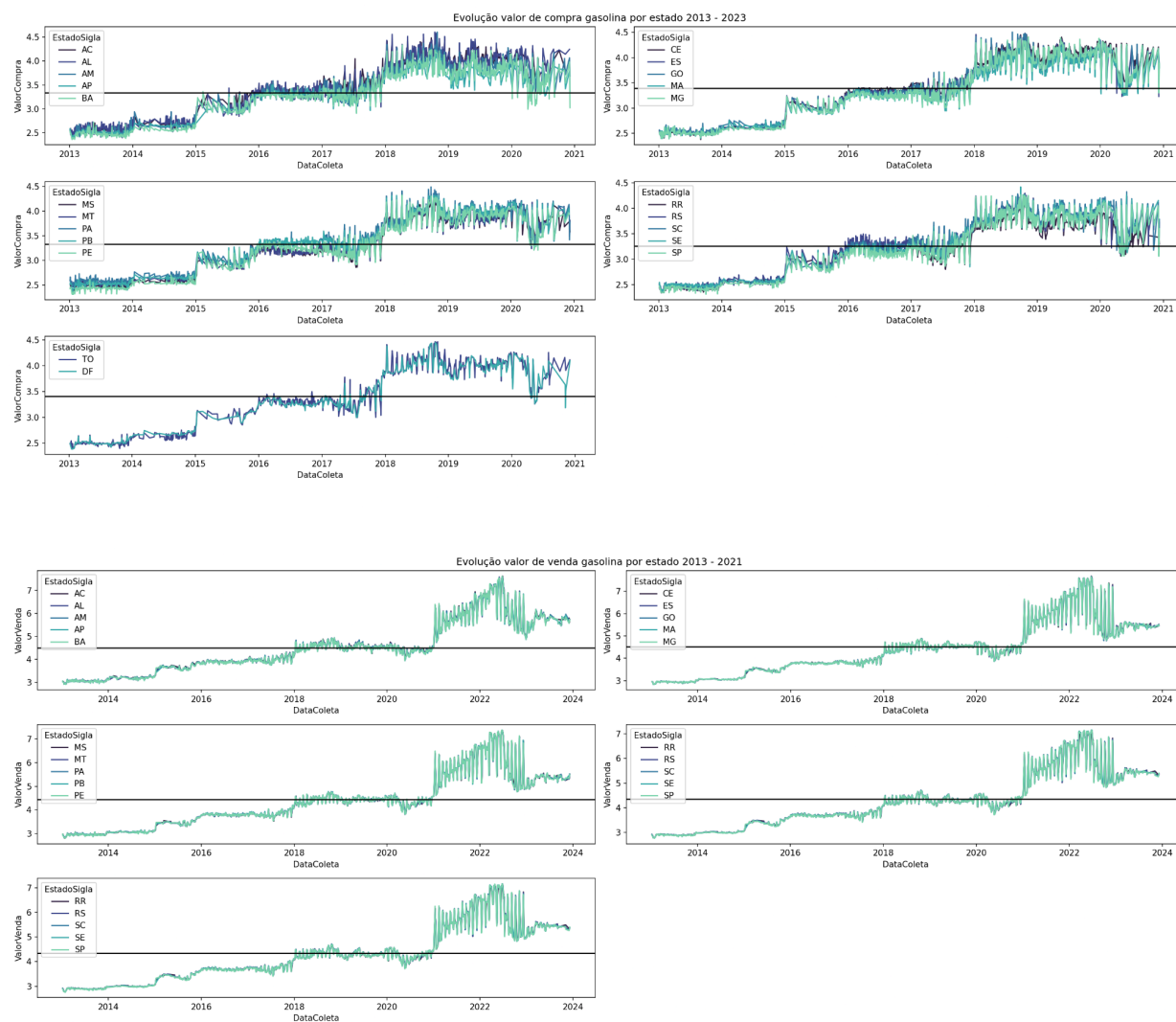


E em 2020, não preciso nem comentar, a pandemia afetou todo o mercado mundial, praticamente todo produto foi afetado, seja em disponibilidade ou preço. Olhemos para a variação entre os valores médios diários ao longo dos anos.

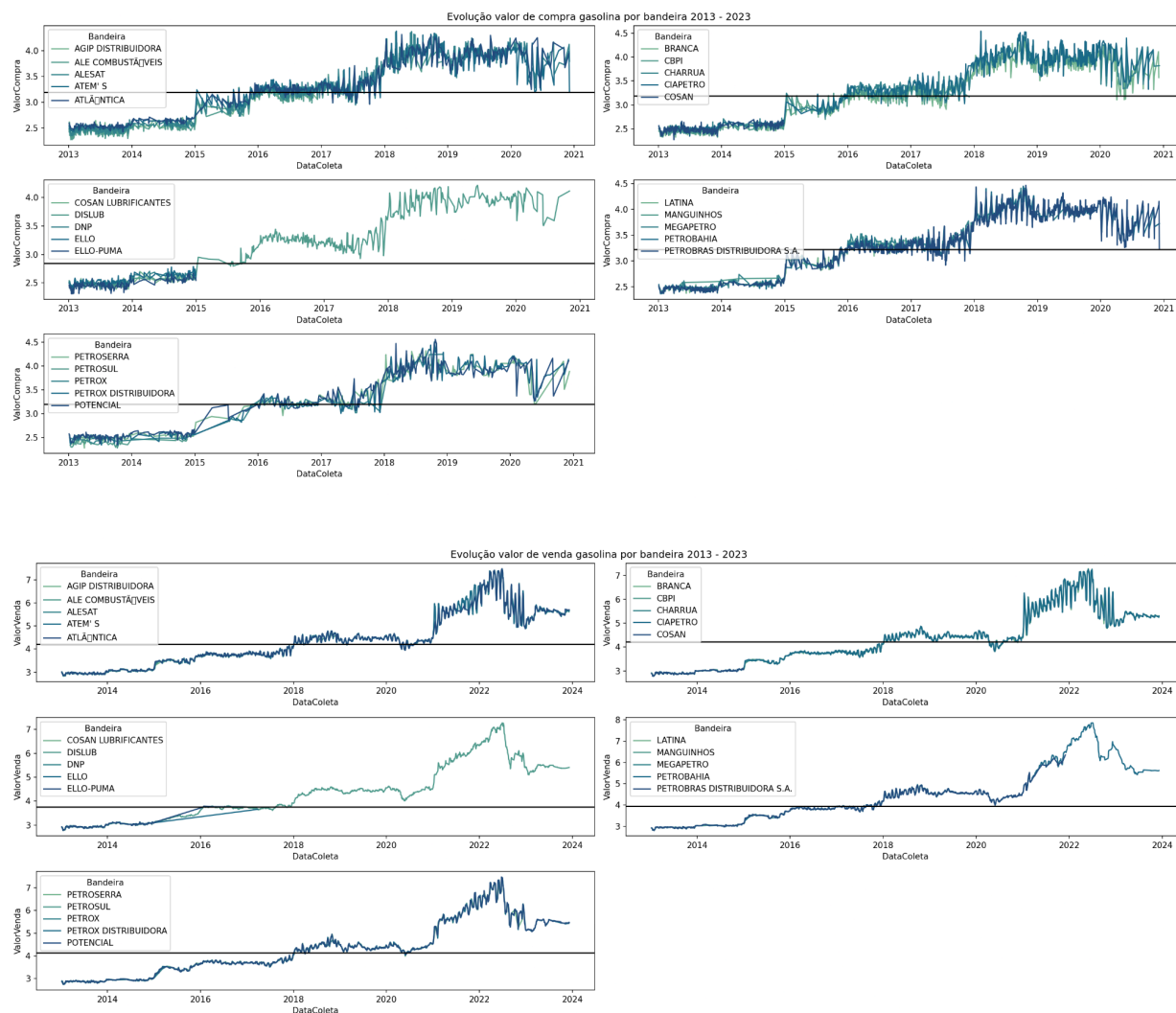


E novamente, além da grande variação após a pandemia, também verificamos que a variação do valor de venda e compra estão sobrepostas, o que corrobora a observação inicial que esse produto segue uma margem de lucro extremamente centrada na média.

Podemos também realizar essa evolução temporal por Unidade Federativa (estados).



Como imaginamos, a evolução é praticamente a mesma, mesmo estratificando por estado, e por fornecedor? Aqui chamado de Bandeira.

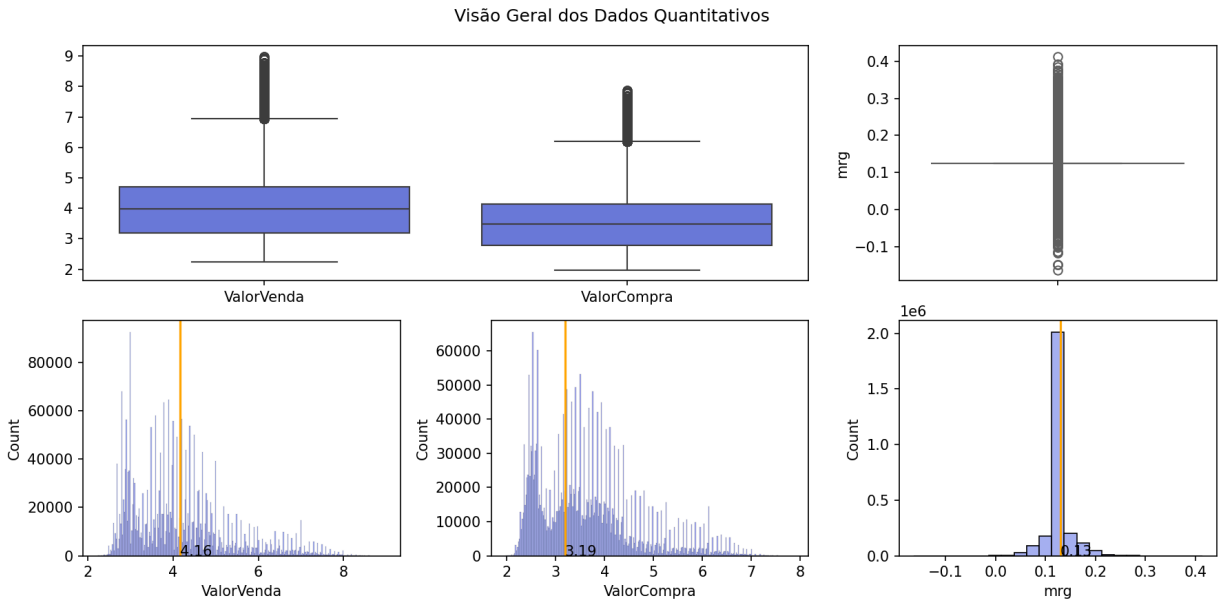


Novamente, enxergamos um padrão muito semelhante aos padrões iniciais encontrados. O que fortalece a ideia de que não há grande variação entre os estados e fornecedores.

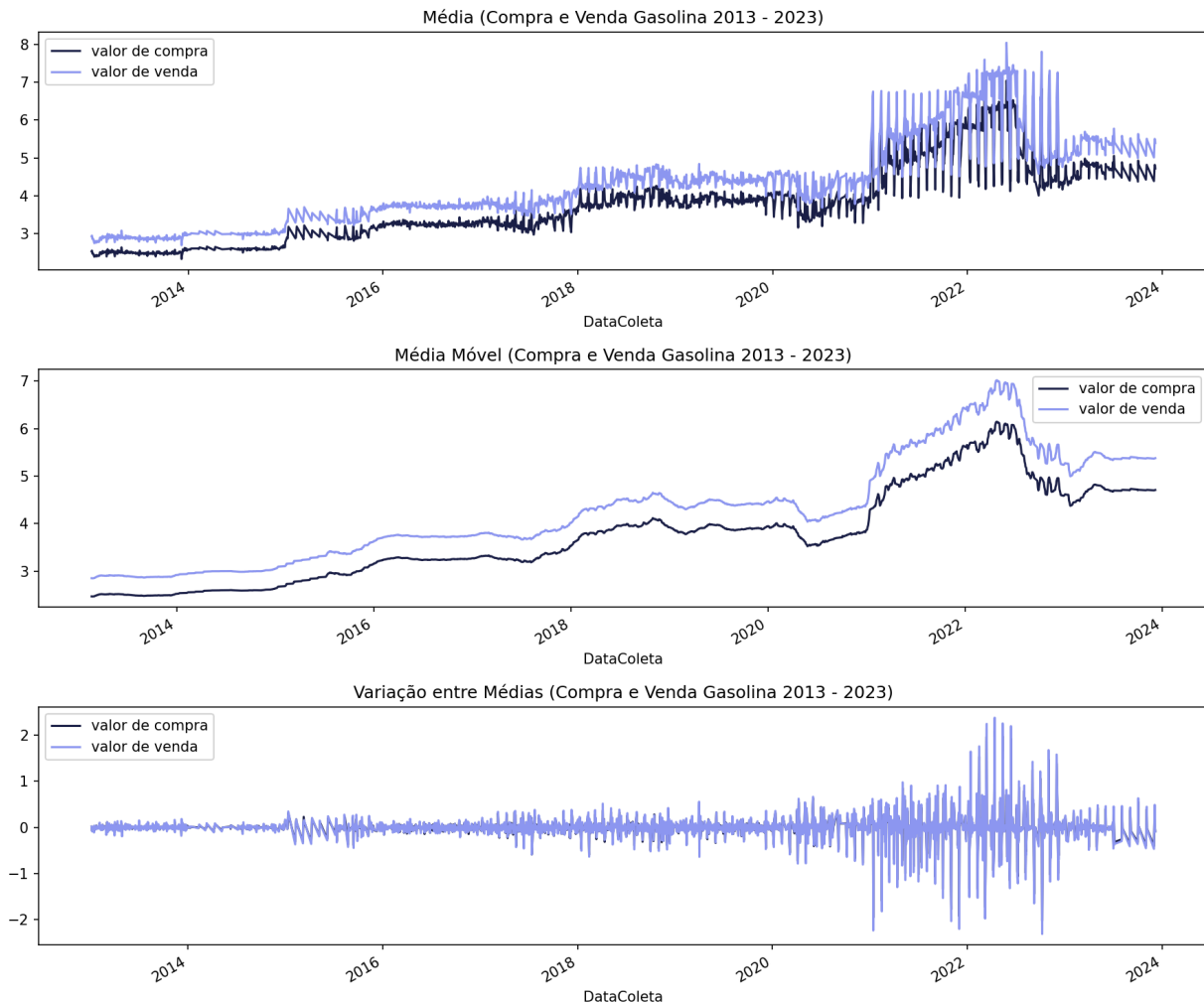
Por fim, lembra que temos uma margem bem centrada em torno de uma mesma média? Podemos então simular como o valor de compra se comportaria com base no valor de venda, através da equação

$$\text{Valor de Compra} = \text{Valor de Venda} - (\text{Valor de Venda} \times \text{Margem})$$

A margem está sendo dada por uma distribuição de números aleatórios entre 0,05 (5%) e 0,16 (16%). Vamos analisar a distribuição dos dados agora.



Olha só, não vemos a presença de mais outliers, olhem no gráfico de caixa (boxplot) da margem. Porém, a distribuição continua muito concentrada em torno da mesma média. E os valores de venda e compra continuam seguindo uma distribuição semelhante, onde os de venda possuem valores um pouco maiores (devido a margem baixa desse produto, analisada em torno de 13%). E o mesmo se mantém na evolução diária, conforme vemos abaixo.



Com isso meus amigos, o que concluimos até aqui?

Concluimos que sim, a gasolina ficou mais cara no Brasil ao longo dos anos. Algo que não foi avaliado aqui foi o possível efeito da inflação nesse cenário (no Brasil medido pelo IPCA).

Porém, mesmo com o produto ficando mais caro, e sua variação ter aumentado assustadoramente pós-pandemia, aparentemente voltando a estar nos padrões pré-covid, a margem de lucro se manteve muito centrada em torno de uma mesma média.

O que podemos esperar para o futuro é continuar observando aumentos, mas aos revendedores de combustível cabe o desafio de trabalhar com uma margem de lucro praticamente fixa. O que pode explicar a tendência desse tipo de negócio buscar agregar valor de outras formas como através de serviços adicionais, lojas de

conveniência integradas aos postos, venda de lubrificantes entre outros, pois depender somente do lucro oriundo dos combustíveis não se mostra uma estratégia sustentável, a não ser que se obtenha um volume que compense o baixo lucro.