

Predição e recomendação de derivativo de commodities

Renata Nunes Pereira, Rodrigo Vidal Cabral, Marina Brayner dos Santos, Luiz Philipe Rangel, Isnard

CONTEXTO

O que são Commodities?

Commodities são todas as matérias-primas essenciais que possuem baixo nível de industrialização. Elas são produzidas em grande quantidade sem diferenças entre marcas e podem ser estocadas sem perda de qualidade, como petróleo, trigo e outras. As commodities servem de base para a indústria no mundo inteiro e são comercializadas globalmente em grande volume.

Os 4 tipos de commodities mais populares no mercado são: agrícola (ex.: trigo), mineral (ex.: ouro), financeira (ex.: índices como o Ibovespa) e de recursos energéticos (ex.: Etanol).

Algumas commodities podem ser negociadas na B3 (Bolsa de Valores brasileira, BM&FBovespa) através do mercado futuro. No entanto, a maior parte das negociações de commodities são realizadas no balcão da B3.

O Mercado de Balcão da B3 é o ambiente que permite a realização de operações que não estão registradas na Bolsa de Valores. Isso não quer dizer, no entanto, que não haja controle ou fiscalização sobre essas transações. O segmento depende da organização e administração de instituições participantes desse mercado.

As principais características do Mercado de Balcão são:

- Inexistência de um espaço físico de negociação.
- Flexibilidade quanto aos registros das transações.

- Possibilidade de negociação de ativos que não estão habilitados para negociação em Bolsa.

No Mercado de Balcão, as corretoras e distribuidoras de valores e bancos de investimento conectam pessoas que desejam comprar e vender ativos.

O Brasil é um grande produtor e exportador de commodities. Por isso, o mercado de negociação é tão ativo e oferece muitas oportunidades tanto para empresas como para pessoas físicas.

Como mencionado anteriormente, são insumos “in natura”, matérias-primas que podem passar por um pequeno processo industrial para aumentar sua duração. No Brasil, as principais commodities negociadas são agrícolas ou minerais, tais como: soja, trigo, laranja, petróleo, minério de ferro, ouro e boi.

Como são fundamentais para o mundo todo, as commodities têm um grande peso na economia, sendo indispensáveis para nossa sobrevivência e o desenvolvimento dos países. Por isso, a negociação dos preços é realizada em escala global.

E para que isso aconteça, é necessária uma estrutura organizada capaz de tornar possível a negociação em escala global. Tais preços sofrem oscilações de acordo com a oferta e procura, assim como o preço de qualquer mercadoria.

A negociação no mercado de commodities do Brasil é feita sob a forma de contratos no chamado mercado futuro, que no Brasil faz parte da B3.

Para você ter ideia do peso das commodities na economia, selecionamos alguns destaques:

- As commodities são responsáveis por 65% do valor das exportações brasileiras.
- Apenas em março de 2017, o volume exportado da soja em grãos subiu para 9 milhões de toneladas. A receita, para US\$ 3,5 bilhões.
- O valor diário das exportações de carne “in natura” é de mais de R\$50 milhões.
- A soja, minério de ferro e petróleo juntos representam 72% das receitas de exportação e fazem parte da balança comercial brasileira.

O que são derivativos?

Derivativos são contratos que derivam a maior parte de seu valor de um ativo subjacente, taxa de referência ou índice. O ativo subjacente pode ser físico (café, ouro, etc.) ou financeiro (ações, taxas de juros, etc.), negociado no mercado à vista ou não (é possível construir um derivativo sobre outro derivativo). Os derivativos podem ser classificados em contratos a termo, contratos futuros, opções de compra e venda, operações de swaps, entre outros, cada qual com suas características.

Contratos a termo se refere a um contrato entre duas partes. Estas, por sua vez, desejam comprar ou vender um ativo em uma data no futuro, porém com preço acordado hoje. A principal diferença do contrato a vista é que este é um acordo para realizar a operação do ativo hoje. Contratos futuros é uma reguladora desse tipo de contrato. Com um contrato futuro, vendedor e comprador se comprometem com uma negociação em uma data futura.

Em relação ao Mercado de opções, nesse mercado se negocia o direito de comprar ou vender um ativo por um preço fixo em uma data futura. Quem adquire o ativo em questão paga um valor ao vendedor. Entretanto, o valor mencionado não é o preço do ativo, mas sim uma quantia paga pela possibilidade de comprar o bem em uma data futura.

Por último, Swap representa um acordo entre duas partes: investidores, empresas, ou ambos. Eles trocarão entre si fluxos de caixa baseados em um valor de referência. Prazo e outras condições devem ser preestabelecidos. Esse processo também é conhecido como “troca de riscos”. Afinal, os participantes negociam a rentabilidade de dois ativos ou mercadorias diferentes.

Os derivativos, em geral, são negociados sob a forma de contratos padronizados. Isto é, previamente especificados (quantidade, qualidade, prazo de liquidação e forma de cotação do ativo-objeto sobre os quais se efetuam as negociações), em mercados organizados, com o fim de proporcionar, aos agentes econômicos, oportunidades para a realização de operações que viabilizem a transferência de risco das flutuações de preços de ativos e de variáveis macroeconômicas.

Como o nome sugere, derivativos utilizam outros ativos como referência, o dólar por exemplo, e são contratos que terão valores econômicos de acordo com a variação do valor, dentro de um determinado tempo. É o que acontece com contratos futuros, por exemplo. Você negocia o contrato de um ativo

baseado no preço atual, com vencimento em uma data futura, apostando na possibilidade de oscilação de valor, fato que gera rentabilidade em seu investimento.

O mercado futuro de derivativos também pode ser uma opção para quem deseja se proteger contra a alta taxa de volatilidade do mercado, isto é, estão em constante variação de preço, essa operação é denominada hedge.

Essa proteção é feita a partir de uma estratégia de definição do preço futuro de uma ação ou outro ativo, tomando como base os preços e as cotações atuais do mercado.

Podemos dizer que os produtores de commodities, como milho, café e boi gordo, por exemplo, realizam a venda da produção por meio de contratos futuros. Isso quer dizer que eles fixam o preço agora, mas repassam o dinheiro, ou os próprios produtos em alguns casos, só no futuro. Essa operação ajuda a controlar o mercado, evitando que a imprevisibilidade prejudique as transações e afete todo o resultado da produção desses produtos.

O que é o Mercado Futuro?

No mercado futuro você compra ou vende contratos com um prazo determinado no futuro. Existe um lote mínimo de contratos que você deve negociar, podendo ser sacas de café, arrobas de boi gordo, ou até mesmo pontos de índices, por exemplo.

Quando uma pessoa compra ou vende um contrato futuro ela assume uma posição nesse mercado. A posição será "comprado" ou "vendido" dependendo da estratégia adotada. Assim sendo, quem está comprando lucra com a alta na cotação do ativo e perde quando há queda no preço. Bem como, quem está vendendo lucra com a queda na cotação do ativo e perde quando há alta no preço.

Essa compreensão é fundamental para entender o que é Mercado Futuro e seus pontos positivos estão listados a seguir:

- Facilidade
- Flexibilidade
- Alavancagem

- Controle de perdas
- Liquidez

Quem compra e vende nesse mercado garante resultados diários com base na diferença entre o valor negociado e a cotação alcançada. Essa é uma espécie de apuração de perdas e ganhos feita pela Bolsa de Valores, conhecida como ajuste diário.

Podemos resumir essa dinâmica em 3 etapas:

1. Início: tudo começa quando você compra ou vende um contrato futuro, escolhendo o ativo do seu interesse estabelecendo o preço e a quantidade.
2. Ajustes diários: entre o início e o término da operação, seus investimentos são ajustados diariamente pela própria Bolsa de Valores. É por isso que podemos entender o Mercado Futuro como um mercado de "soma zero".
3. Encerramento: ocorre quando o investimento chega ao vencimento ou você encerrar antes do prazo, é o momento de obter o lucro ou o prejuízo por meio da soma dos resultados diários.

Ressalta-se que os ativos negociados no Mercado Futuro são comercializados e consumidos praticamente no mundo todo. Dessa forma, os horários de negociação, para iniciar ou encerrar um investimento, podem mudar ao longo do ano. Assim, a Bolsa brasileira pode funcionar em horários mais próximos das negociações nos Estados Unidos e Europa.

PROBLEMA DO NEGÓCIO

Baixa operação de derivativos de commodities para médias e pequenas empresas devido ao pouco conhecimento de operação no mercado.

IMPACTO

Alavancar o mercado de derivativos, principalmente para proteção de pequenos e médios produtores que não possuem áreas financeiras específicas para isso ou com baixo conhecimento do mercado (ex.: cooperativas).

DESENHO DA SOLUÇÃO

Recomendar para os operadores de commodities, tipos de derivativos que são melhores para sua proteção de acordo com a característica de cada derivativo. Por exemplo, identificar quais commodities são recomendados para alta ou baixa volatilidade da moeda indexada.

O presente trabalho terá como foco inicial a avaliação de apenas um nicho, como o petróleo, por exemplo. A seguir, demais áreas serão abordadas, focando sempre nas commodities mais operadas.

REFERÊNCIAS

HULL, J. C. Opções, futuros e outros derivativos. 9 ed. Porto Alegre: Bookman, 2016.

https://www.investidor.gov.br/menu/Menu_Investidor/derivativos/Derivativos_introducao.html

<https://www.topinvest.com.br/o-que-sao-derivativos/>

<https://blog.toroinvestimentos.com.br/hedge-o-que-e>

DADOS

BUSINESS QUESTIONS	DATOS	BASE	CÁLCULO	FONTE DO DADO
Recomendar para os operadores de commodities, tipos de derivativos que são melhores para sua proteção de acordo com a característica de cada derivativo.	Arquivos .xlsx	Arquivos do excel da base “cotacoes” do período 26/08/2019 a 18/05/2022 & Arquivos do excel da base “termo_mercadoria” do mesmo período	A partir da inserção do Dollar calcular o IPCA acumulado do dia	B3 e Banco Central

FONTE

Os dados utilizados serão da B3 e do Banco Central

Os dados da B3 foram conseguidos com sigilo, pois não podemos expor os clientes e suas transações.

Os dados da B3 foram entregues em duas principais tabelas, uma “cotações e a outra “termo_mercadoria, porém, subdividido em 671 arquivos para a primeira tabela e 713 arquivos para a segunda, todos no formato .xlsx, respectivamente. O período vai de 26/08/2019 até 18/05/2022.

Os dados do Banco Central foram extraídos via Scrapping.

Datasets

B3

“cotacao”

data_cotacao	datetime64[ns]
cod_ric	object
cod_ticker	object
cotacao_tipo	object
fator_conversao	float64
bolsa_negociacao	object
tipo_mercadoria	object
mercadoria	object
moeda	object
unidade_mercadoria	object
valor_cotacao	float64
dtype:	object

“termo_mercadoria”

data_referencia	datetime64[ns]
data_inicio	datetime64[ns]
data_vencimento	datetime64[ns]
numero_contrato	object
mercadoria	object
tipo	object
bolsa	object
unidade_negociacao	object
moeda	object
quantidade_negociada	int64
vendedor	object
comprador	object
dtype:	object

Banco central

USD	float64
dtype:	object


```
IPCA          float64
ipca_acumulado float64
dtype: object
```

OBS: tivemos que criar uma função para ler todos os arquivos do excel, vide:

```
#criando dataframe termo_mercadoria
path = r'c:\xxxxxxxxxx'
files = os.listdir(path)
df_tm = pd.DataFrame()
#print(files)
files_xlsx = [path+'\\' + f for f in files if f[-4:] == 'xlsx']
#print(files_xlsx)
for f in files_xlsx:
    data = pd.read_excel(f)
    df_tm = df_tm.append(data)
```

```
criando dataframe cotacao
path = r'C:\xxxxxxxxxx'
files = os.listdir(path)
df_ct = pd.DataFrame()

#print(files)
files_xlsx = [path+'\\' + f for f in files if f[-4:] == 'xlsx']
#print(files_xlsx)
for f in files_xlsx:
    data = pd.read_excel(f, sheet_name='Commodity')
    df_ct = df_ct.append(data)
```

EDA (Análise Exploratória de dados)

Juntamos os DataFrames “cotacoes e ‘termo-mercadoria’ por meio de um “merge’ e acrescentamos os dados do Banco Central.

Para melhor manuseio do DataFrame final salvamos ele como mastertable.pkl

Vide:

```
<class 'pandas.core.frame.DataFrame'>
```

DatetimeIndex: 242578 entries, 2019-07-24 to 2022-05-18

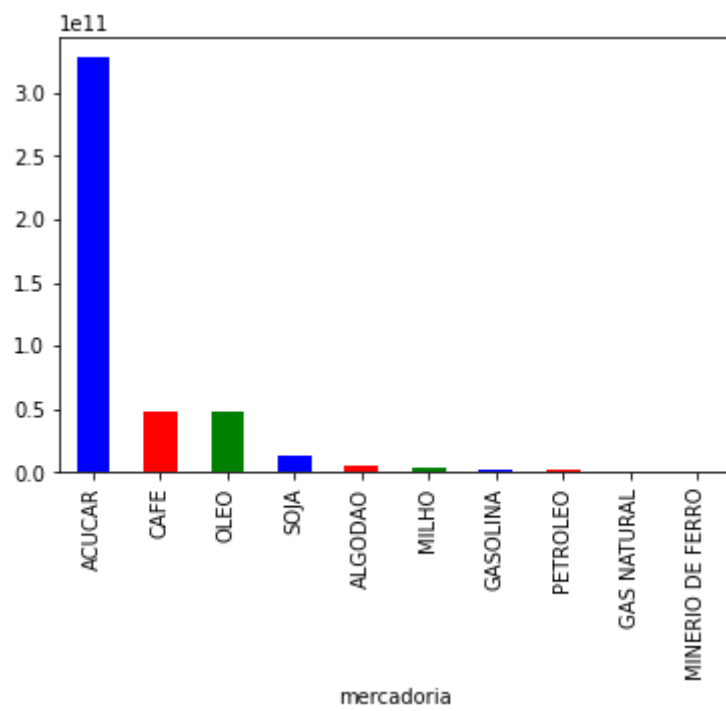
Data columns (total 21 columns):

#	Column	Non-Null Count	Dtype
0	anomes	242578 non-null	object
1	mercadoria	242578 non-null	object
2	tipo	242578 non-null	object
3	bolsa	242578 non-null	object
4	unidade_negociacao	242578 non-null	object
5	moeda	242578 non-null	object
6	quantidade_negociada	242578 non-null	int64
7	comprador_vendedor	242578 non-null	object
8	cod_ric	242523 non-null	object
9	cod_ticker	237308 non-null	object
10	fator_conversao	242578 non-null	float64
11	valor_cotacao	242578 non-null	float64
12	usd	242578 non-null	float64
13	notional_usd	242578 non-null	float64
14	notional_brl	242578 non-null	float64
15	mes	242578 non-null	object
16	ano	242578 non-null	object
17	ipca	242578 non-null	float64
18	ipca_acumulado	242578 non-null	float64
19	notionalusd_ipca	242578 non-null	float64
20	notionalusd_norm	242578 non-null	float64

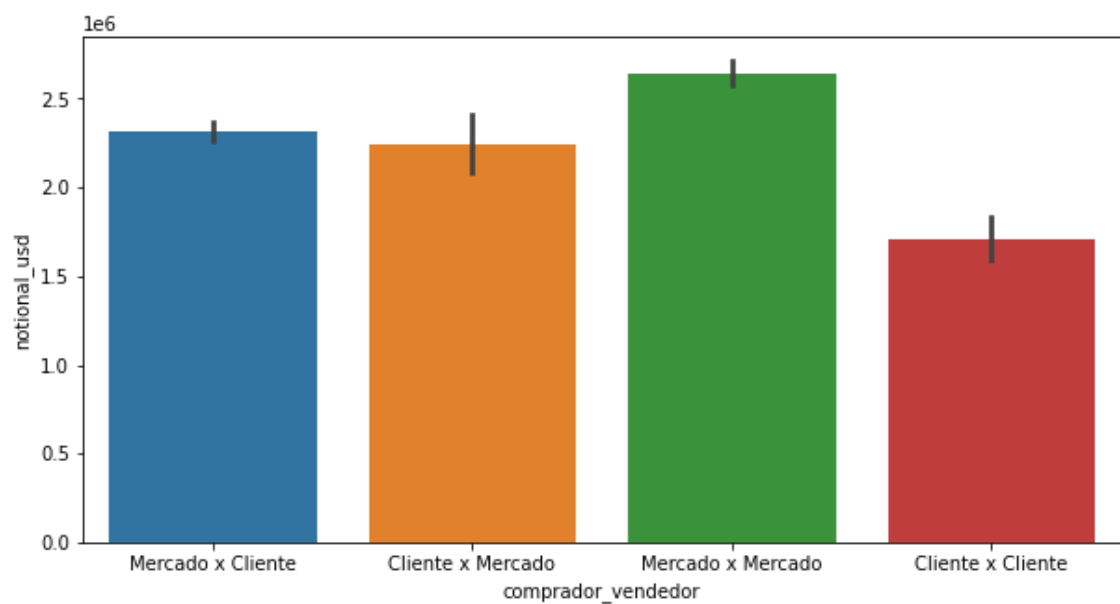
dtypes: float64(9), int64(1), object(11)

memory usage: 40.7+ MB

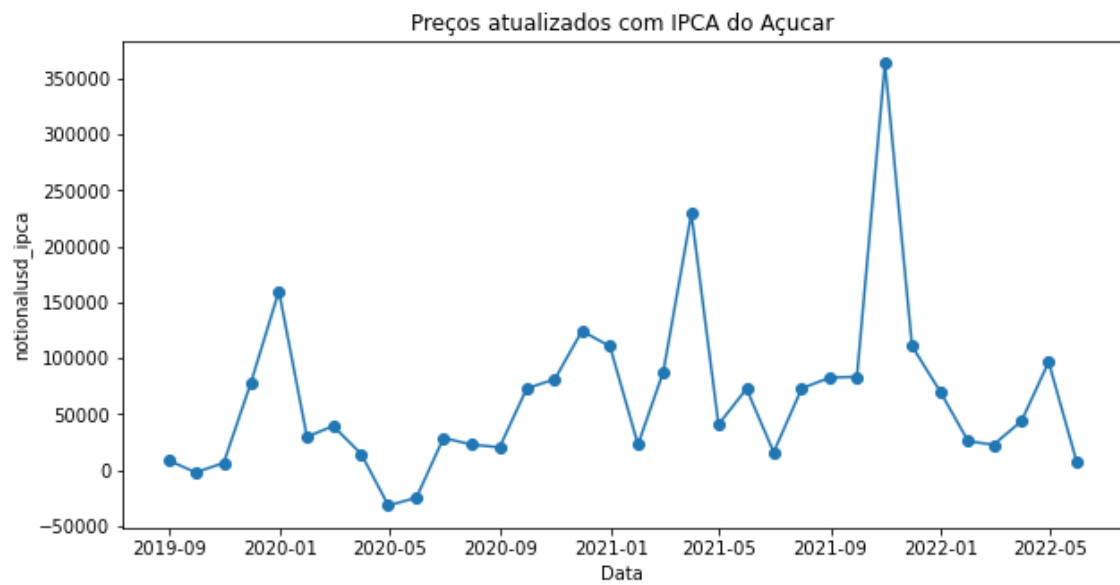
Principais Commodities negociadas:



Negociação:



Preços atualizados com IPCA do açúcar:



Testando a série:

Dickey-Fuller Aumentado

Teste Estatístico: -1.6430

Valor-p: 0.4607167493

Valores Críticos:

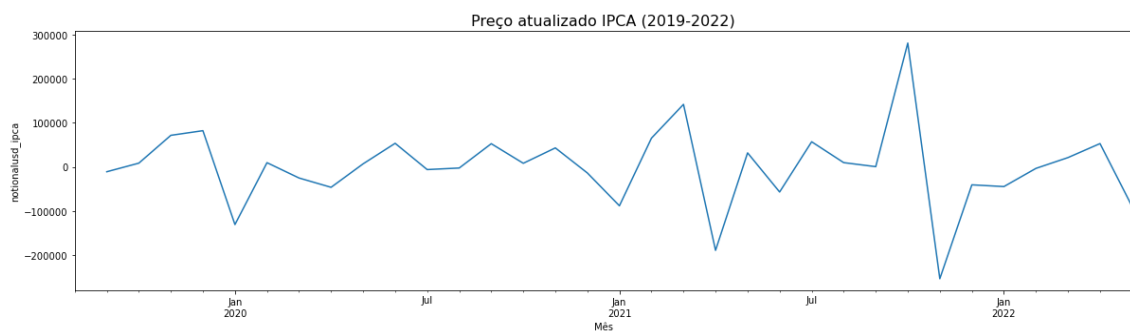
1%: -3.4377

5%: -2.8648

10%: -2.5685

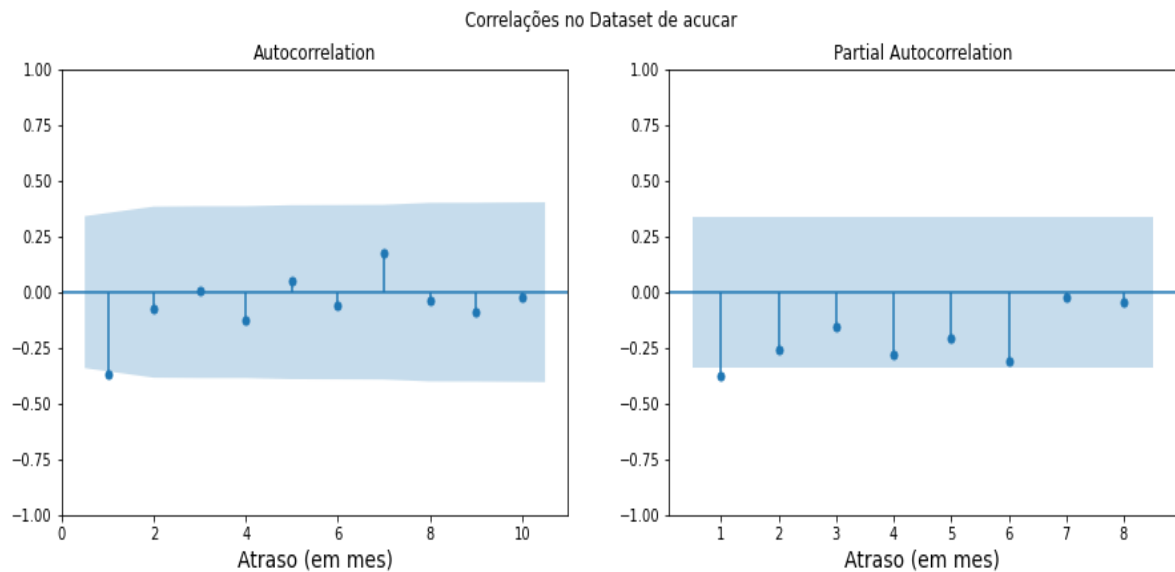
Failed to Reject Ho - Time Series is Non-Stationary

Agora com o comando diff()



Testando:

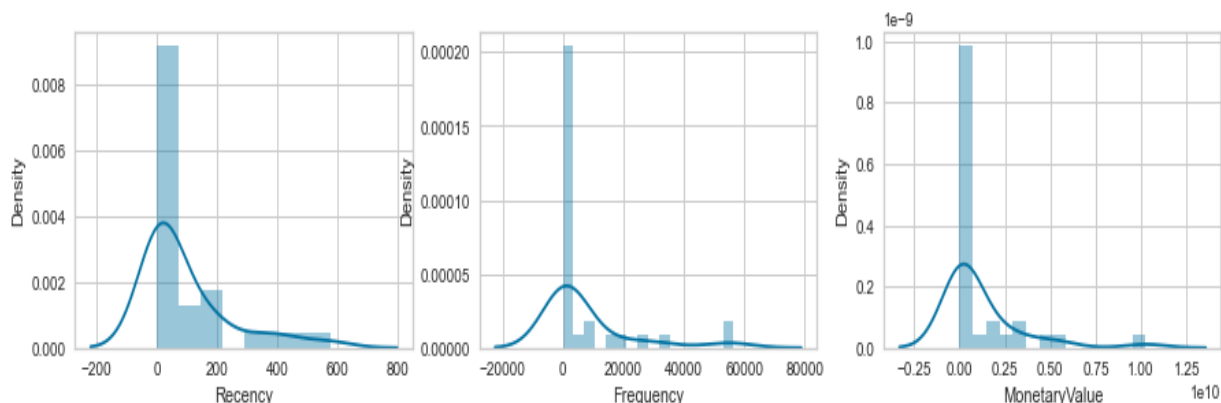
p-valor (<.05 indica estacionariedade) desta série é 2.3228205981705583e-12



ABORDAGENS

Primeira: Aplicamos uma abordagem interessante que foi a segmentação de mercado para as commodities.

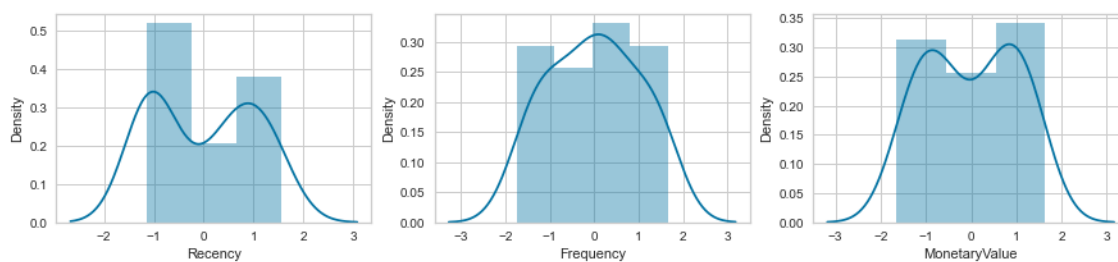
- Segmentação de mercado é o processo de identificar, em um conjunto heterogêneo de consumidores, grupos com características e comportamentos semelhantes. É uma importante ferramenta para entender o mercado e direcionar ações específicas para cada tipo de cliente.
- Existem várias metodologias para efetuar a segmentação. Uma muito conhecida é a RFM, que visa dividir grupos de acordo com os seguintes critérios:
 - Recency: quando a última compra foi efetuada;
 - Frequency: com que frequência compras são efetuadas;
 - Monetary Value: quanto o cliente costuma gastar.
- Vamos adaptar esse tipo de segmentação para as negociações de commodities, onde:
 - Recency: quando a última operação foi efetuada;
 - Frequency = Frequencia que a mercadoria é negociada
 - Monetary value = quanto costumam negociar da mercadoria



Acima da figura estão impressos os valores de skew de cada variável, uma medida de assimetria; quanto mais distantes de zero, mais assimétricas são as distribuições.

Podemos aplicar transformações de potência para aumentar a simetria das distribuições, tornando-as mais "gaussianas". Para isso, utilizaremos a função `power_transform` do módulo `sklearn.preprocessing`:

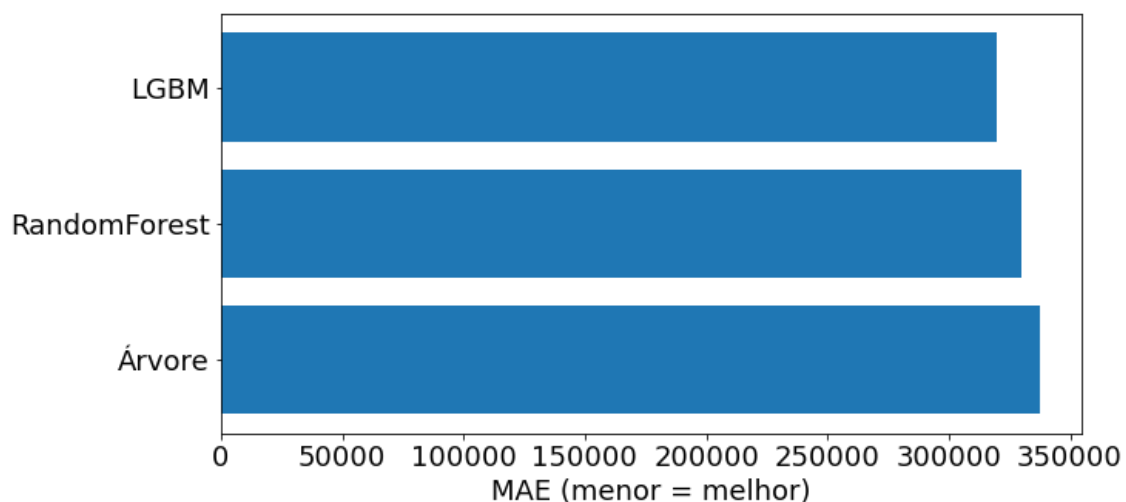
Visualizando as distribuições resultantes:



Estamos preparando os dados para utilizar um modelo de clusterização KMeans. A partir dessa clusterização decidimos fazer um portfólio de commodities usando **oHierarchical Risk Parity**.

Segunda: predição de uma commodities, utilizando os modelos: DecisionTree, RandomForest e LGBM.

1. Construímos as features usando o passado da base também a variável resposta usando o futuro.
2. Definimos que nossa target será a previsão dos próximos 30 dias.
3. Usamos o módulo TimeSeriesSplit para fazer validação cruzada.
4. Obtivemos um objeto split, como em um KFold usado em validações cruzadas de problemas não temporais. Checamos o tamanho de cada fold a cada iteração, porém, ainda precisamos separar dados que o modelo nunca terá acesso, mesmo na validação cruzada. Como temos dados temporais e queremos testar a performance no futuro, faremos uma separação Out of time, “o corte”.
5. Separamos os últimos 6 meses para teste.
6. Escolhendo uma métrica e encontrando o modelo ideal
 - a. mean_squared_error ou MSE: Média das diferenças entre a variável verdadeira e a predita ao quadrado. Quanto menor, melhor!
 - b. mean_absolute_error ou MAE: Média das diferenças entre a variável verdadeira e a predita em módulo. Quanto menor, melhor!
 - c. root_mean_squared_error ou RMSE: Raíz quadrada do mean_squared_error. Quanto menor, melhor!
 - d. r2_score ou apenas r2: Percentual da variância total da variável resposta que pode ser explicada pelo modelo além do que a média já explica. Quanto maior, melhor!
7. Em séries temporais, há uma grande preferência por uso de métricas com escala, como o MAE ou o RMSE. Para este problema, inicialmente, iremos usar o MAE.
8. Usamos a crossvalidação temporal e a métrica MAE para avaliar a performance destes modelos e fazer um gráfico.



Pelo gráfico acima, o modelo que parece ter sido o melhor com os hiperparâmetros usados é o LGBM. Então escolhemos este modelo para otimização

Seleção de variáveis: inferindo a janela ideal de defasagem.

Na grande maioria dos problemas de séries temporais, as bases de dados não são tão grandes assim, exceto em casos em que a captura de dados é muito alta (dados em segundos, milissegundos, etc.) Entretanto, dependendo do cenário, podemos ter bases de dados com muitas features.

Testamos alguns tamanhos de base:

Número de features com 10 defasagens: 52

Número de features com 15 defasagens: 72

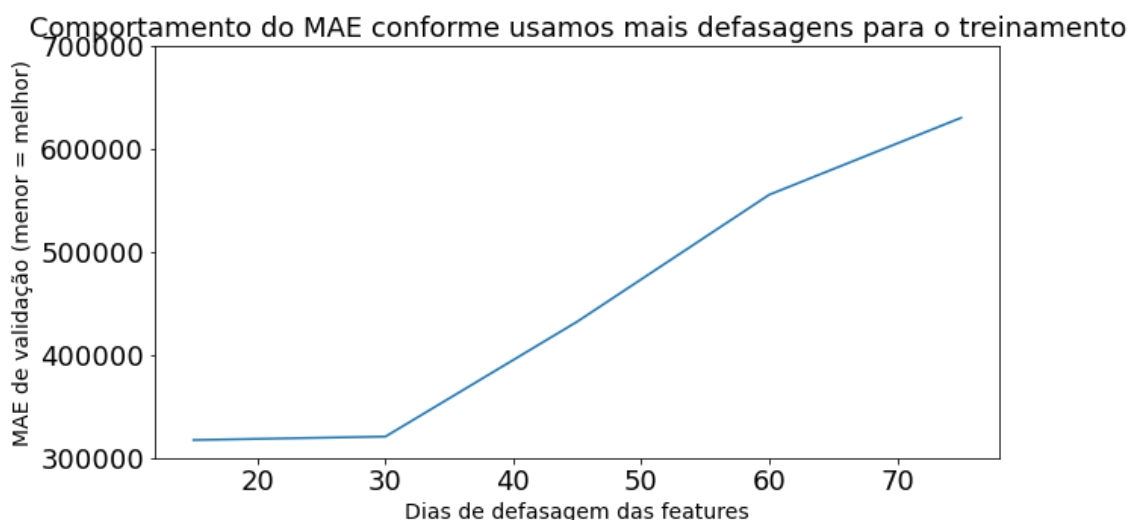
Número de features com 20 defasagens: 92

Número de features com 30 defasagens: 132

Número de features com 45 defasagens: 192

Número de features com 60 defasagens: 252

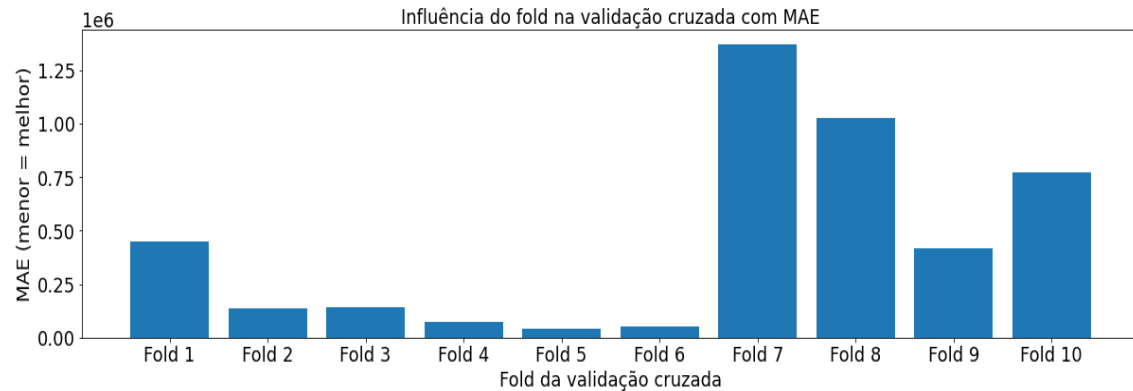
Até mesmo pensando na implantação de modelos, muitas features podem tornar o modelo pesado e complexo, então de alguma maneira tivemos que fazer uma seleção de variáveis baseada nos resultados. Uma forma bacana foi analisar se nossa performance melhora muito na medida em que usamos mais features para o modelo



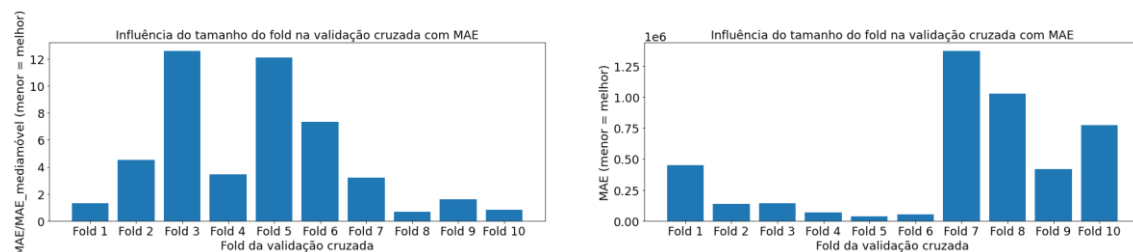
Pelo gráfico acima, vemos que usando mais features defasadas não ganhamos tanta performance. Sendo assim, vamos travar 15 dias e não em 30.

Um dos pontos mais importantes da validação cruzada de séries temporais é a escolha da métrica. Isso importa bastante pois o tamanho do conjunto de treino varia a cada iteração.

Vamos avaliar isso usando o `cross_val_score`.



Quanto menos dados temos para treinar o modelo, mais underfitting teremos, e a nossa métrica precisa ser robusta a isso para evitarmos conclusões incorretas. Uma solução simples para isso é comparar a nossa métrica com a métrica de um modelo de base. No caso do `r2_score`, o modelo de base é uma média simples da variável resposta. Calculamos para todos os dados a média móvel do período de defasagem para o passado, e usamos a métrica deste modelo "simples" para avaliar o nosso.

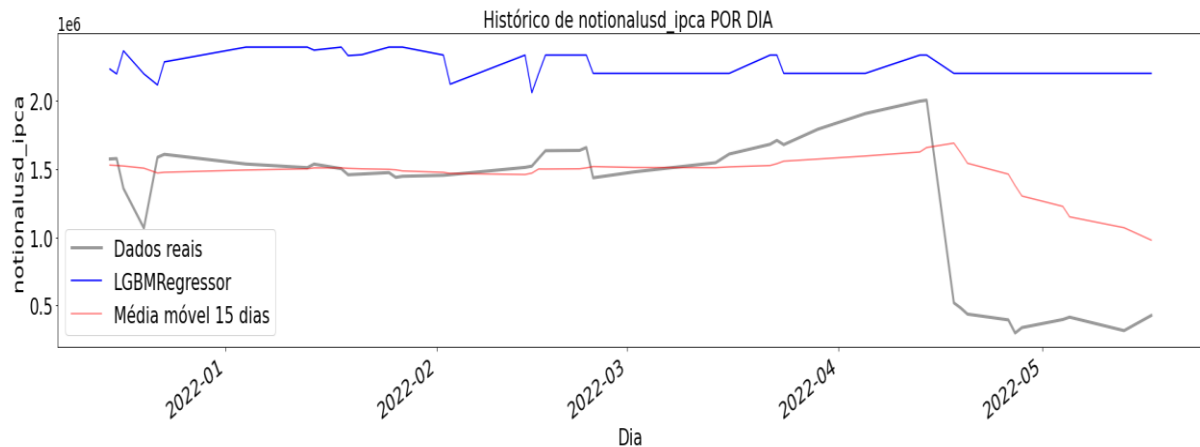


Otimização de Hiperparâmetros e modelo final.

Agora com nosso modelo escolhido, otimizamos os hiperparâmetros e plotamos nosso modelo final.

Para isso, usamos o `GridSearchCV` específico para o `LGBM`.

Plotamos os dados de treino e teste para 6 meses do conjunto de teste



Não tivemos uma performance boa com o LGBMRegressor

Usando o Keras, construímos uma rede neural adicionando camadas uma após a outra, de forma a buscar resolver o nosso problema. Apenas para recapitular:

Construímos nossa variável resposta como sendo o valor médio de notionalusd_ipca 15 dias seguintes.

Consideramos features 15 anteriores.

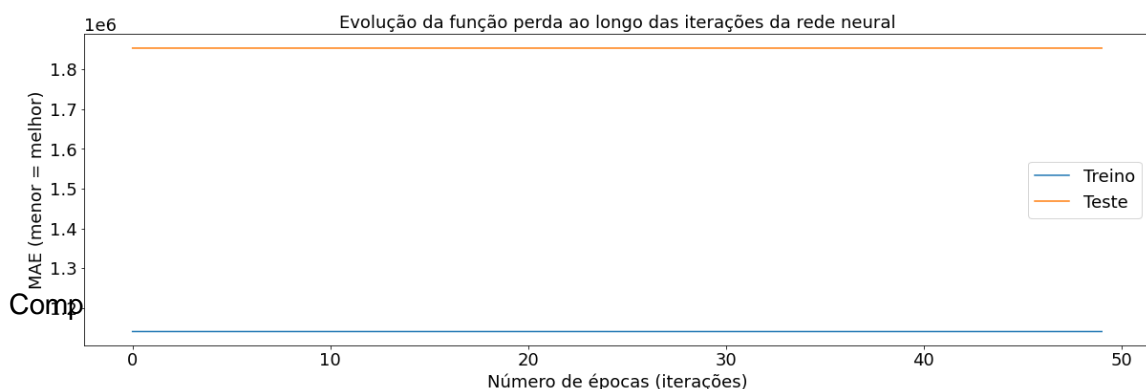
Nenhum modelo entende que essa ordem existe. Quando adicionamos uma camada LSTM(Long Short Term Memory) à rede neural, estamos dizendo que a ordem dos dados deve afetar a resposta, poranto ganhamos uma nova camada de impacto em relação aos outros modelos.

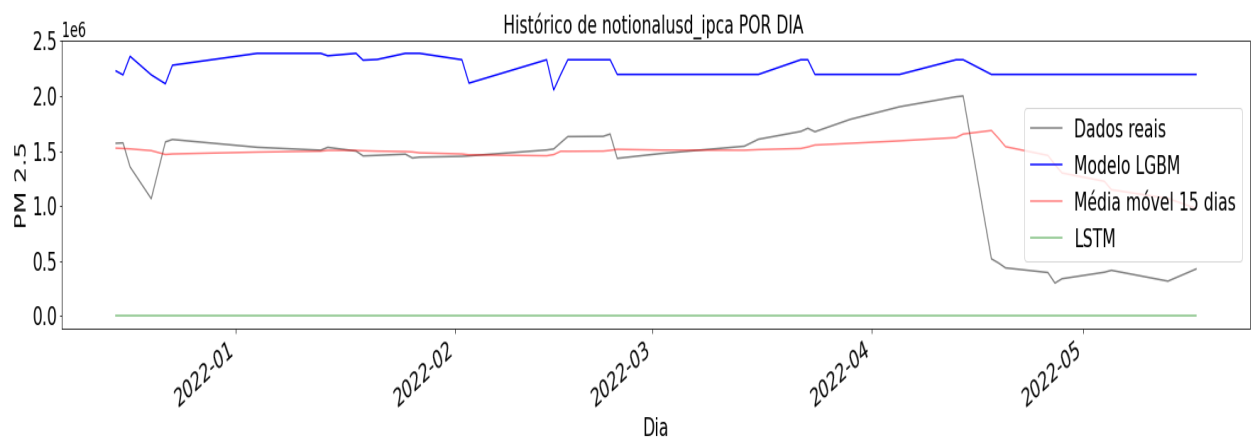
Para treinar um modelo no Keras, por ser uma rede neural, fizemos mais dois tratamentos:

- Escala dos dados, pois não é um modelo robusto a escala.
- Adicionar uma dimensão (para se tornar um tensor), mesmo que tenha dimensão 1.

Usamos um modo em que o modelo nos diz a função custo a cada iteração.

Plotamos a função custo (MAE), no treino e no teste, ao longo das iterações.





Observamos que mesmo com esse modelo a média móvel ainda é melhor.

Decidimos então partir para o modelo ARIMA.