

**FACULDADE DE TECNOLOGIA DA BAIXADA SANTISTA
RUBENS LARA**

**PROJETO DE INICIAÇÃO CIENTÍFICA:
ESTUDO E APLICAÇÃO DE TÉCNICAS PARA MODELAR E PREVER
PADRÕES EM SÉRIES TEMPORAIS EM DADOS ECONÔMICOS DE
MERCADO**

Nome do aluno: Gabriel Andrade do Nascimento - 6º ciclo - Ciências de Dados

Orientadora: Profª. Ms. Márcia Roberta Dos Santos Pires Da Silva

Junho – 2024

SUMÁRIO

<i>Introdução.....</i>	3
Objetivos do Trabalho.....	3
Relevância do Projeto	5
Revisão da Literatura.....	5
<i>Metodologia.....</i>	6
Levantamento Bibliográfico.....	6
Ferramentas e Linguagem de Programação	6
Desenvolvimento e Controle de Versão	7
<i>Análise dos Dados e Discussão.....</i>	8
Definição de variáveis padrão para as análises.....	8
intervalo_de_corte_outliers	8
intervalo_de_agrupamento.....	8
Análise Exploratória dos Dados	9
Discussão a respeito das explorações	15
Análises Preditivas.....	16
Discussão a respeito das previsões.....	20
<i>Considerações finais.....</i>	21
<i>Referências</i>	22

Introdução

A análise de séries temporais é uma ferramenta essencial na Ciência de Dados, permitindo compreender padrões e comportamentos em dados que evoluem ao longo do tempo. Em dados econômicos de mercado, essas análises são cruciais para identificar tendências e eventos que impactam decisões econômicas e estratégias de investimento.

Este campo, no entanto, enfrenta desafios significativos. A instabilidade dos mercados financeiros, influenciada por fatores macroeconômicos, políticos e globais, dificulta a modelagem precisa. Eventos extraordinários, como crises ou choques econômicos, adicionam complexidade e requerem técnicas robustas para identificar e gerenciar anomalias, evitando distorções nas análises. Além disso, dados de alta frequência, comuns no mercado financeiro, demandam métodos avançados para reduzir ruídos e extrair padrões significativos.

A presente pesquisa foca na análise das interações entre o valor do Dólar, o preço do das ações da Petrobras, Aramco, Exxon e Chevron e derivados de petróleo como gasolina, diesel e GLP. Por meio de técnicas estatísticas avançadas e aprendizado de máquina, busca-se explorar a correlação entre essas variáveis e prever preços futuros desses derivados para consumidores finais.

A relevância do estudo reside na complexidade desses fatores e em suas implicações para a economia brasileira e o cotidiano da população. Compreender as dinâmicas do setor de energia é essencial para a formulação de políticas públicas eficazes e estratégias econômicas.

Assim, este trabalho contribui não apenas para o entendimento histórico das flutuações e correlações no mercado energético, mas também fornece insights estratégicos que fortalecem a interseção entre economia, ciência de dados e o setor energético nacional.

Objetivos do Trabalho

Este estudo tem como objetivo aprofundar a compreensão de aspectos fundamentais do contexto econômico brasileiro, inicialmente focando na investigação da complexa correlação entre o preço das ações da Petrobras, o valor do Dólar e os preços dos combustíveis, como gasolina, diesel e Gás Liquefeito de Petróleo (GLP).

Contudo, ao longo da análise, essas variáveis demonstraram uma correlação menos significativa do que se esperava, o que exigiu a inclusão de novos fatores no estudo. Entre as variáveis adicionais inseridas estão a cotação do Riyal (moeda da Arábia Saudita) e os valores de fechamento das ações das petrolíferas ARAMCO, CHEVRON e EXXON, outras petrolíferas que apresentaram correlações mais expressivas no decorrer da pesquisa.

OBJETIVOS GERAIS

1. Compreender as inter-relações Econômicas: Investigar a relação entre o preço das ações da Petrobras, Aramco, Exxon e Chevron, o valor do Dólar e do Riyal os preços dos combustíveis, e como essas variáveis impactam o cenário econômico brasileiro.
2. Desenvolver Modelos de Previsão: Aplicar técnicas avançadas de modelagem estatística e aprendizado de máquina para prever tendências futuras nos preços dos combustíveis.
3. Contribuir para a Formulação de Políticas Públicas: Fornecer *insights* que possam ser aplicados na formulação de estratégias econômicas e políticas públicas mais informadas, com base na análise das séries temporais dos dados econômicos.

OBJETIVOS ESPECÍFICOS

1. Análise de Dados Históricos: Realizar uma exploração detalhada dos dados econômicos históricos, identificando padrões, tendências, sazonalidades e outliers.
2. Avaliar a eficiência dos modelos usando métricas como MAE, RMSE e MAPE, para identificar o mais adequado a cada conjunto de dados.
3. Integração de Variáveis Macroeconômicas: Incorporar fatores como inflação e indicadores de mercado para aumentar a precisão das previsões.
4. Visualização de Resultados: Desenvolver visualizações claras e interativas com ferramentas como Seaborn e Plotly, facilitando a interpretação dos dados pelos stakeholders.
5. Documentação e Disseminação: Registrar o processo de modelagem e os resultados, compartilhando-os por meio de relatórios, artigos acadêmicos e apresentações.

6. Propostas de Melhoria: Sugerir aprimoramentos contínuos nos modelos, considerando os avanços tecnológicos para aumentar a precisão e aplicabilidade em diferentes cenários econômicos.

Relevância do Projeto

Este projeto explora a análise de séries temporais em dados econômicos, uma ferramenta essencial para compreender dinâmicas financeiras. Ao investigar a relação entre preços do petróleo, inflação e combustíveis, fatores cruciais para a economia brasileira, o estudo aplica técnicas avançadas de modelagem para prever tendências e subsidiar políticas públicas eficazes. Sua contribuição está na interseção entre economia, ciência de dados e setor energético, oferecendo insights estratégicos para gestão econômica e tomada de decisões.

Revisão da Literatura

A previsão de séries temporais é amplamente explorada na literatura acadêmica, destacando-se como um tema central em economia, finanças e ciência de dados. Diversos estudos contribuem significativamente para o desenvolvimento de metodologias e ferramentas aplicadas à análise de dados complexos.

O estudo “The Role of Temporal Dependence in Factor Selection and Forecasting Oil Prices”, de Kyle E. Blinder, Mohsen Pourahmadi e James W. Mjelde, ressalta a relevância da dependência temporal na previsão de preços do petróleo. Integrar esse aspecto nos modelos resulta em maior precisão, fundamental tanto para decisões empresariais quanto para políticas econômicas, oferecendo direções importantes para estudos futuros.

Já “Stock Market Analysis: A Review and Taxonomy of Prediction Techniques”, de Dev Shah, Haruna Isah e Farhana Zulkernine, apresenta uma taxonomia detalhada das técnicas de previsão de mercado financeiro. Este guia estruturado auxilia profissionais a lidar com as incertezas do mercado de ações, consolidando-se como uma referência na área.

No campo de séries temporais sazonais, o trabalho “Time Series Forecasting by a Seasonal Support Vector Regression Model”, de Ping-Feng Pai e colaboradores, propõe o modelo SSRV (Seasonal Support Vector Regression). A pesquisa demonstra que esse modelo supera métodos tradicionais, como ARIMA, em termos de precisão, especialmente em dados com forte sazonalidade.

A abordagem híbrida de Yanpeng Zhang e coautores, apresentada em “A Novel Fuzzy Time Series Forecasting Model Based on Multiple Linear Regression and Time Series Clustering”, combina séries temporais fuzzy, regressões lineares múltiplas e clustering. Essa inovação melhora a precisão das previsões e expande o potencial de técnicas avançadas de aprendizado de máquina no contexto de séries temporais.

O modelo CRISP-DM (Cross-Industry Standard Process for Data Mining), proposto por Rüdiger Wirth e Jochen Hipp no estudo “CRISP-DM: Towards a Standard Process Model for Data Mining”, estrutura projetos de mineração de dados em fases interconectadas. Este modelo, amplamente adotado, oferece flexibilidade e sistematicidade, sendo referência em diversos setores industriais.

Por fim, “A Prediction Approach for Stock Market Volatility Based on Time Series Data”, de Sheikh Mohammad Idrees, M. Afshar Alam e Parul Agarwal, aborda a previsão de volatilidade no mercado de ações utilizando técnicas de aprendizado de máquina aplicadas a séries temporais. Os resultados destacam avanços na gestão de riscos e formulação de estratégias financeiras robustas.

Essas contribuições formam a base teórica e metodológica para o desenvolvimento de modelos mais precisos e eficientes na análise de séries temporais econômicas.

Metodologia

Levantamento Bibliográfico

O projeto iniciou-se com um levantamento bibliográfico abrangente, focado em técnicas de ciência de dados aplicadas à economia, especialmente modelos de aprendizado de máquina para séries temporais. A pesquisa incluiu a análise de abordagens e metodologias voltadas à previsão de padrões econômicos, com ênfase no mercado petrolífero.

Foram revisados estudos de caso e pesquisas que exploram diversos modelos de previsão, como redes neurais e métodos estatísticos. A seleção criteriosa desses materiais garantiu uma base teórica sólida para embasar o desenvolvimento e a aplicação das técnicas previstas no projeto.

Ferramentas e Linguagem de Programação

Para a implementação das técnicas e construção dos modelos de previsão, a linguagem de programação Python foi escolhida como a principal ferramenta. Python oferece uma vasta gama de bibliotecas e *frameworks* que facilitam a aplicação de

técnicas de ciência de dados e aprendizado de máquina. Dentre as bibliotecas selecionadas, destacam-se:

- **Scikit-learn:** Para a implementação de algoritmos de aprendizado de máquina e modelos de previsão.
- **Pandas:** Para a manipulação e análise de dados estruturados.
- **Numpy:** Para operações matemáticas e de álgebra linear.
- **Matplotlib e Seaborn:** Para a visualização de dados e a criação de gráficos.
- **Statsmodels:** Para a modelagem estatística e análise de séries temporais.

Essas ferramentas foram escolhidas por sua robustez, flexibilidade e ampla aceitação na comunidade de ciência de dados, garantindo a eficácia e a eficiência no desenvolvimento dos modelos.

Desenvolvimento e Controle de Versão

Para o armazenamento e controle de versão dos scripts de análise, o GitHub foi selecionado. Esta plataforma oferece uma integração eficiente com o ambiente de desenvolvimento escolhido, o Visual Studio Code. O uso do GitHub proporciona diversas vantagens, incluindo:

- **Facilidade de Integração:** Permite a colaboração contínua e sincronização dos códigos desenvolvidos.
- **Branches e Commits:** Facilitam a gestão das alterações e atualizações no código, garantindo a segurança e a rastreabilidade das modificações.
- **Restauração de Versões:** A possibilidade de retornar a versões anteriores do código contribui para a resolução rápida de problemas e a manutenção da integridade dos scripts.

Análise dos Dados e Discussão

Definição de variáveis padrão para as análises

Antes de tudo, vale considerar que, a fim de melhorar a estrutura e a visualização dos gráficos, foram criadas as variáveis “intervalo_de_corte_outliers” e “intervalo_de_agrupamento”.

intervalo_de_corte_outliers

Esta variável é utilizada para definir o intervalo de dias para a detecção de outliers. Ao agrupar os dados em intervalos de x dias, é possível identificar e tratar outliers de maneira mais eficaz. A lógica por trás desta abordagem é que ao aumentar o intervalo de corte, mais discrepâncias podem ser detectadas, o que ajuda a suavizar a curva dos dados.

- **Objetivo:** Suavizar a curva de dados e melhorar a detecção de outliers;
- **Impacto:** Quanto maior o intervalo, mais outliers serão identificados, contribuindo para uma análise mais limpa e precisa dos dados;
- **Uso:** É particularmente útil em séries temporais onde a detecção de outliers pode ser crítica para a precisão dos modelos preditivos.

intervalo_de_agrupamento

Esta variável se refere ao intervalo utilizado para agrupar os dados com o objetivo de visualização. Ao definir um intervalo de agrupamento, o número de pontos de dados exibidos nos gráficos de linha pode ser reduzido, facilitando a interpretação visual dos dados.

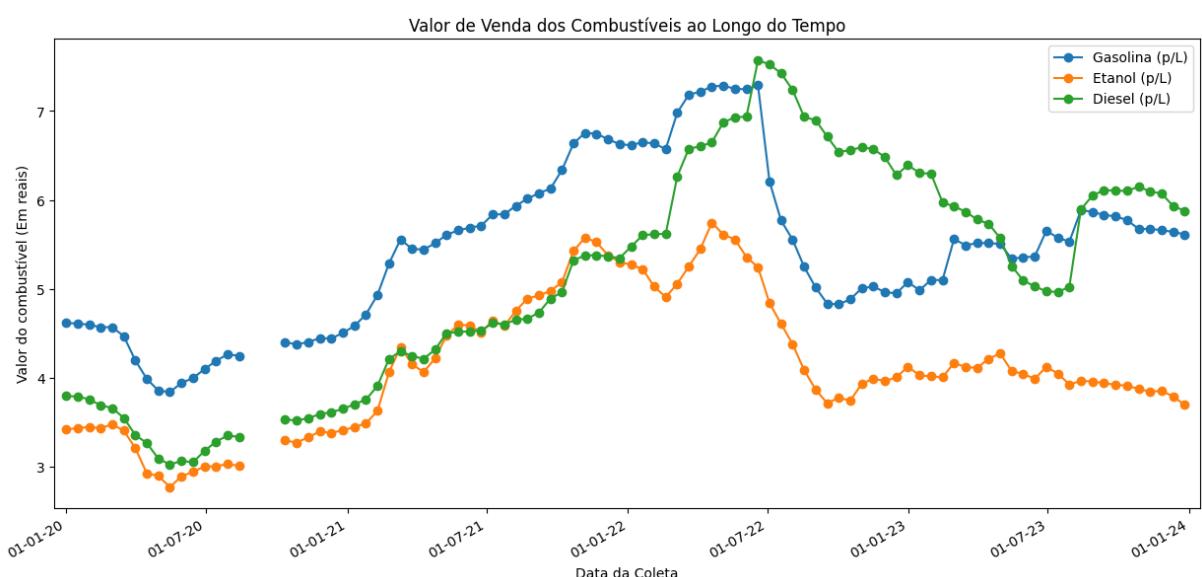
- **Objetivo:** Melhorar a visualização dos dados ao reduzir o número de pontos mostrados em gráficos.
- **Impacto:** Com intervalos maiores, os gráficos se tornam menos congestionados, permitindo uma interpretação mais clara das tendências e padrões ao longo do tempo.
- **Uso:** Essencial para a criação de visualizações de dados que são fáceis de entender e interpretar, especialmente em apresentações e relatórios.

Essas variáveis são fundamentais para o tratamento e visualização eficaz dos

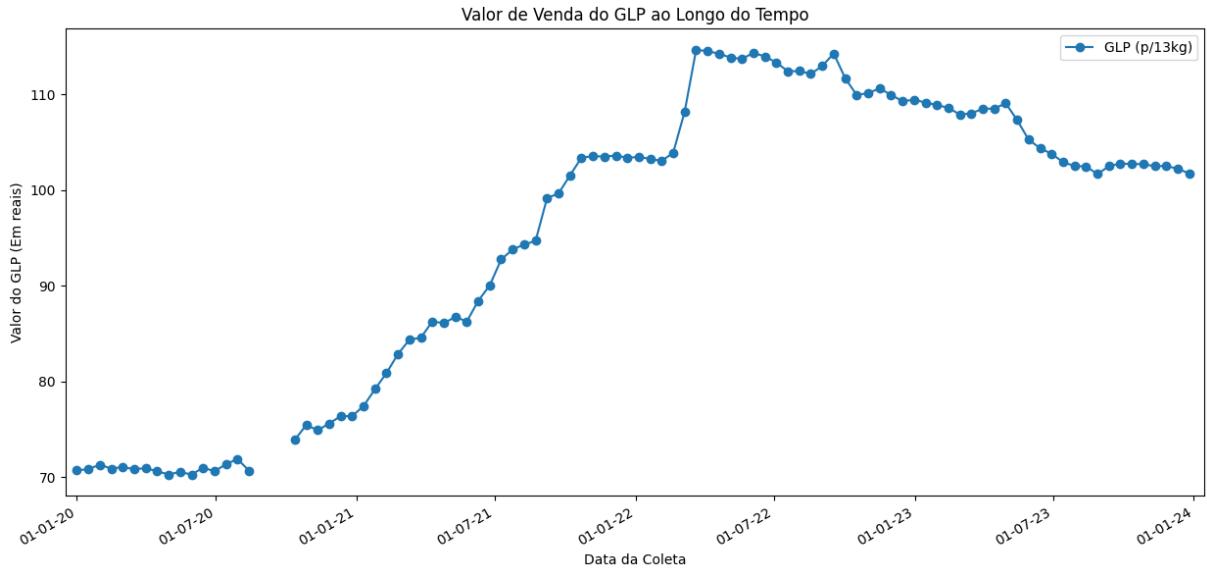
dados. O intervalo_de_corte_outliers assegura que os outliers sejam identificados e tratados adequadamente, melhorando a qualidade dos dados utilizados nos modelos. Já o intervalo_de_agrupamento facilita a criação de gráficos mais limpos e interpretáveis, o que é crucial para a comunicação dos resultados das análises. A definição e o ajuste cuidadoso desses intervalos são essenciais para o sucesso das análises de séries temporais no contexto econômico e de mercado.

Assim, assumindo que definimos ambas as variáveis com o valor “15”, podemos prosseguir as análises.

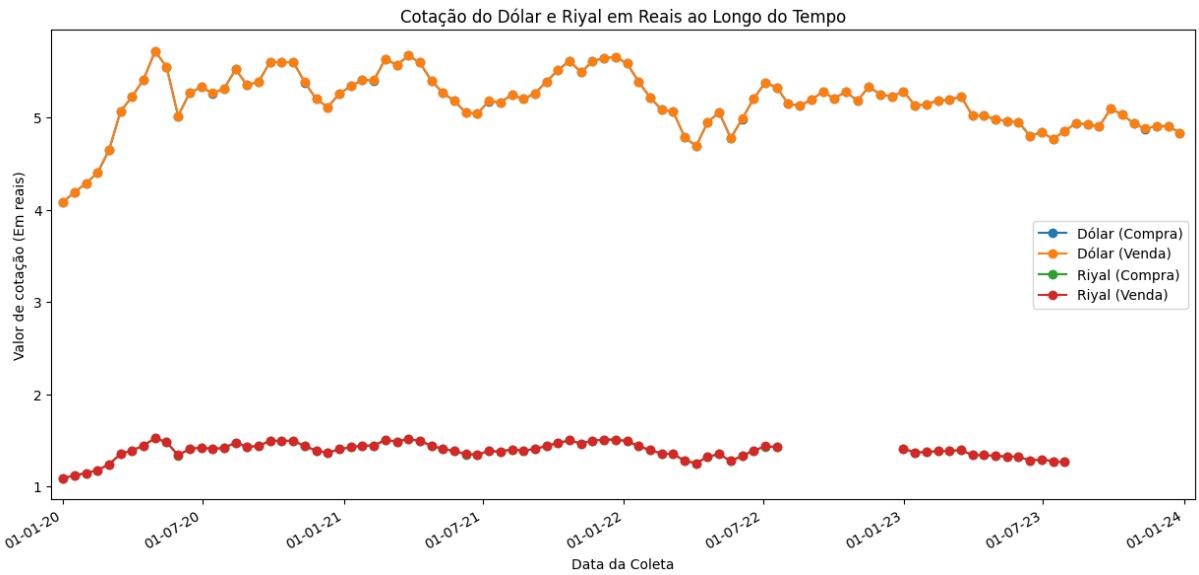
Análise Exploratória dos Dados



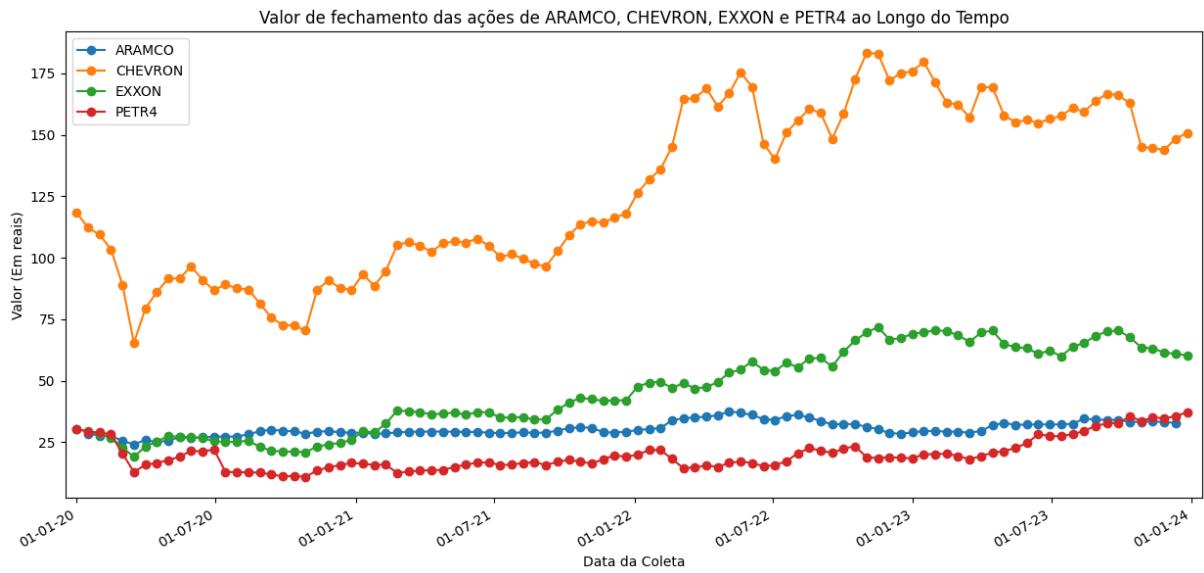
Os preços dos combustíveis exibem um comportamento similar ao longo do tempo, o que indica uma forte correlação entre as variáveis. Observa-se uma tendência de aumento nos preços a partir do início de 2020, coincidente com o início da pandemia de COVID-19. A ausência de dados no segundo semestre de 2020 é provavelmente causada pela pandemia.



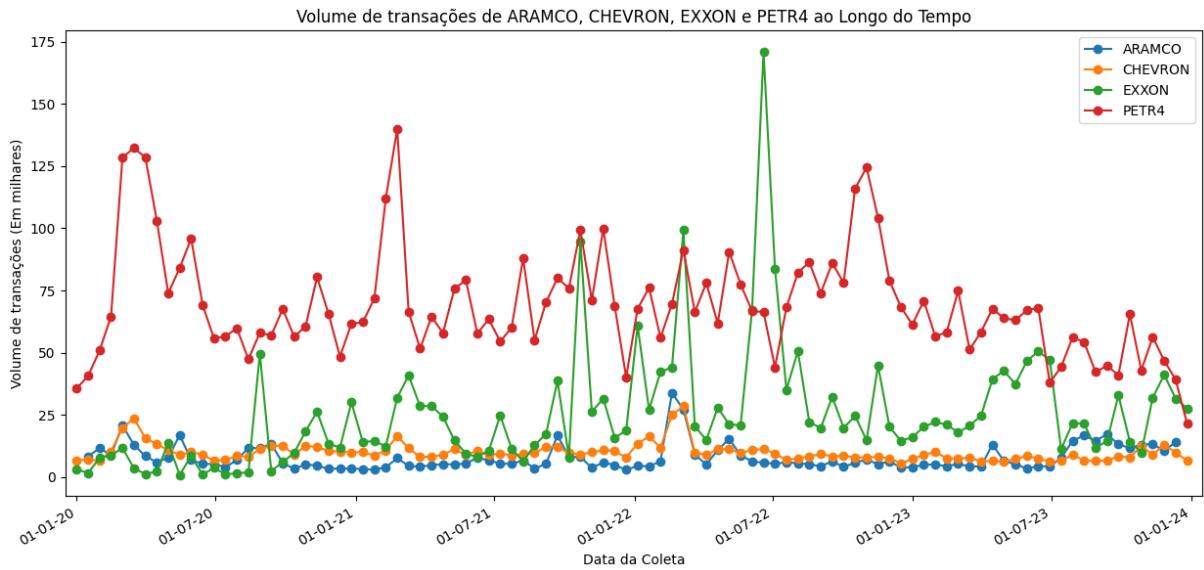
Observa-se um crescimento acentuado no valor do GLP desde o início de 2021 até meados de 2022. A ausência de dados no segundo semestre de 2020 pode ser atribuída, novamente, às implicações da pandemia de COVID-19.



Há uma significativa lacuna de dados sobre o Rival no segundo semestre de 2023, cuja causa não pôde ser identificada. Ainda assim, os dados disponíveis revelam uma alta instabilidade no comportamento do Dólar, enquanto o Rival apresenta relativa estabilidade ao longo dos últimos anos, mesmo durante a pandemia de COVID-19. Notavelmente, as curvas de variação das duas moedas exibem formatos semelhantes.

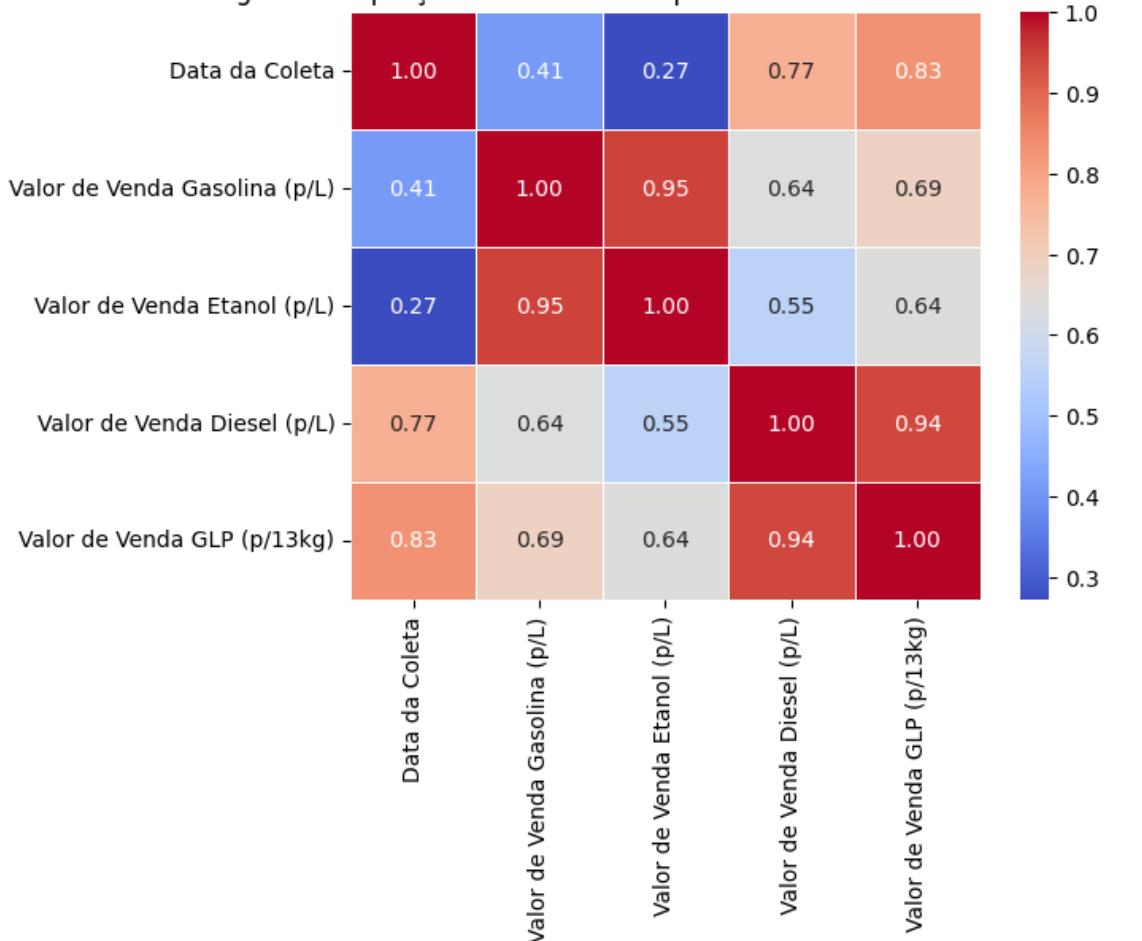


Nesse ponto, podemos observar que a CHEVRON, visivelmente, tem um valor de fechamento maior que as demais petrolíferas. No entanto, também é notável a maior volatilidade do valor dela. Já as demais petrolíferas mantêm um valor de mercado estável com o tempo, com o início de um leve crescimento da EXXON a partir do começo de 2022.



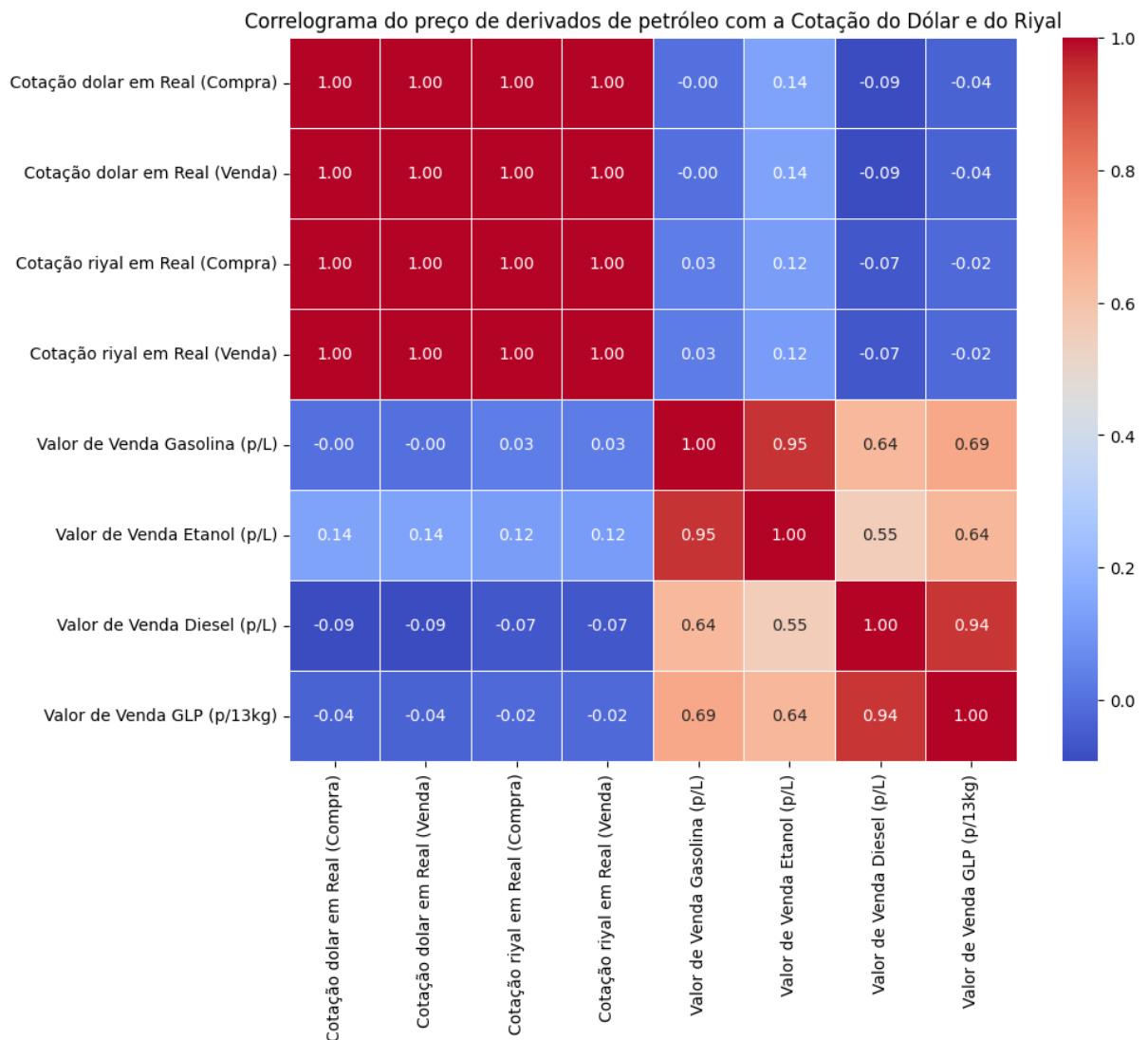
Com esse gráfico, fica perceptível que a Petrobras (PETR4) tem visivelmente um número de transações mais alto que as demais petrolíferas, também variando mais que as demais. Além disso, a EXXON também tem um número de transações maior e com maior variação, mas não se compara a Petrobras. Por fim, as demais petrolíferas seguem uma linha mais estável e com menor número de transações.

Correlograma do preço de derivados de petróleo com a Data da Coleta do dado

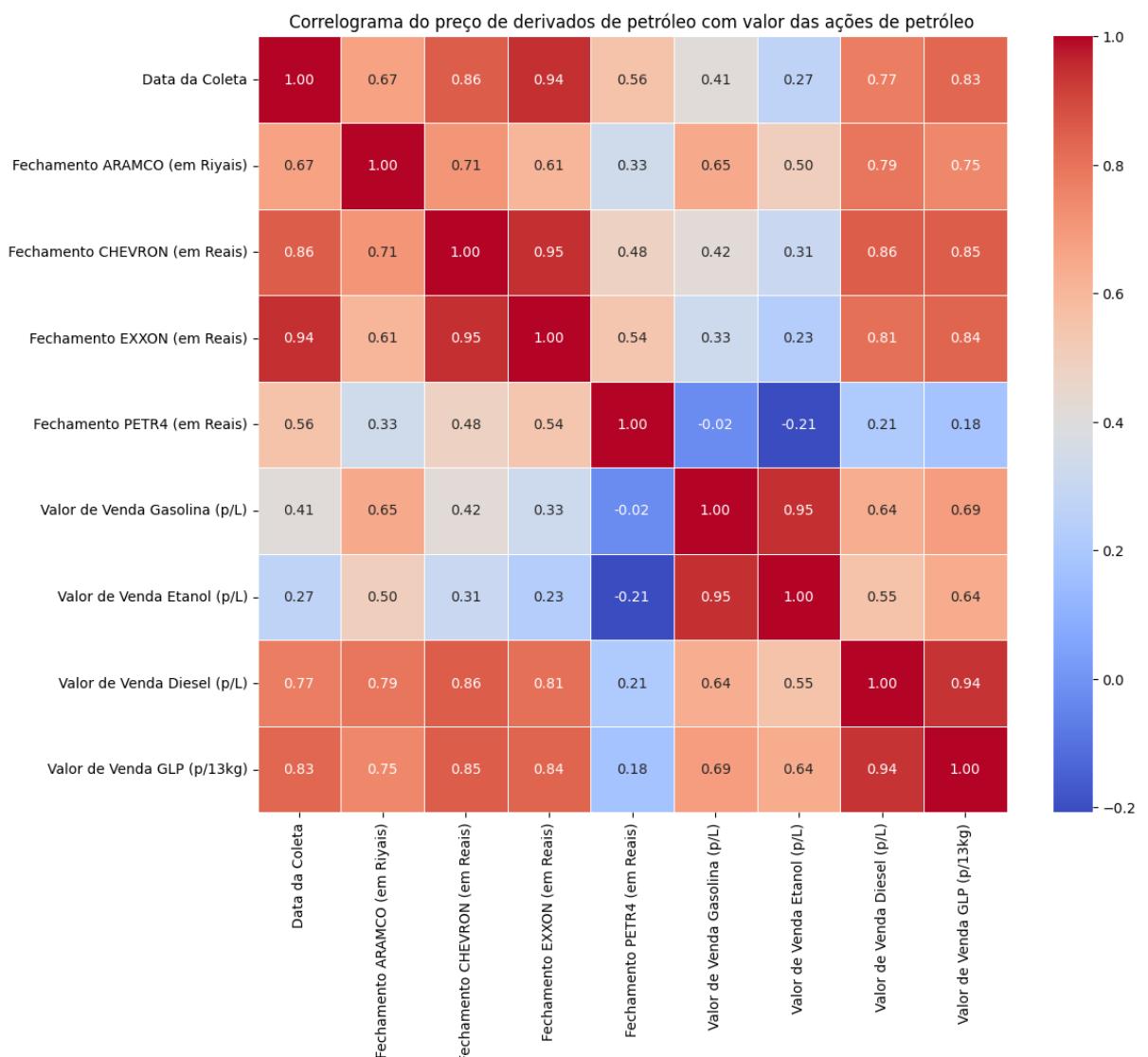


A forte correlação positiva entre a data de coleta e o valor de venda do GLP indica que o preço de venda do gás de cozinha está aumentando ao longo do tempo. O mesmo padrão é observado nos demais derivados de petróleo, embora com uma correlação menos intensa.

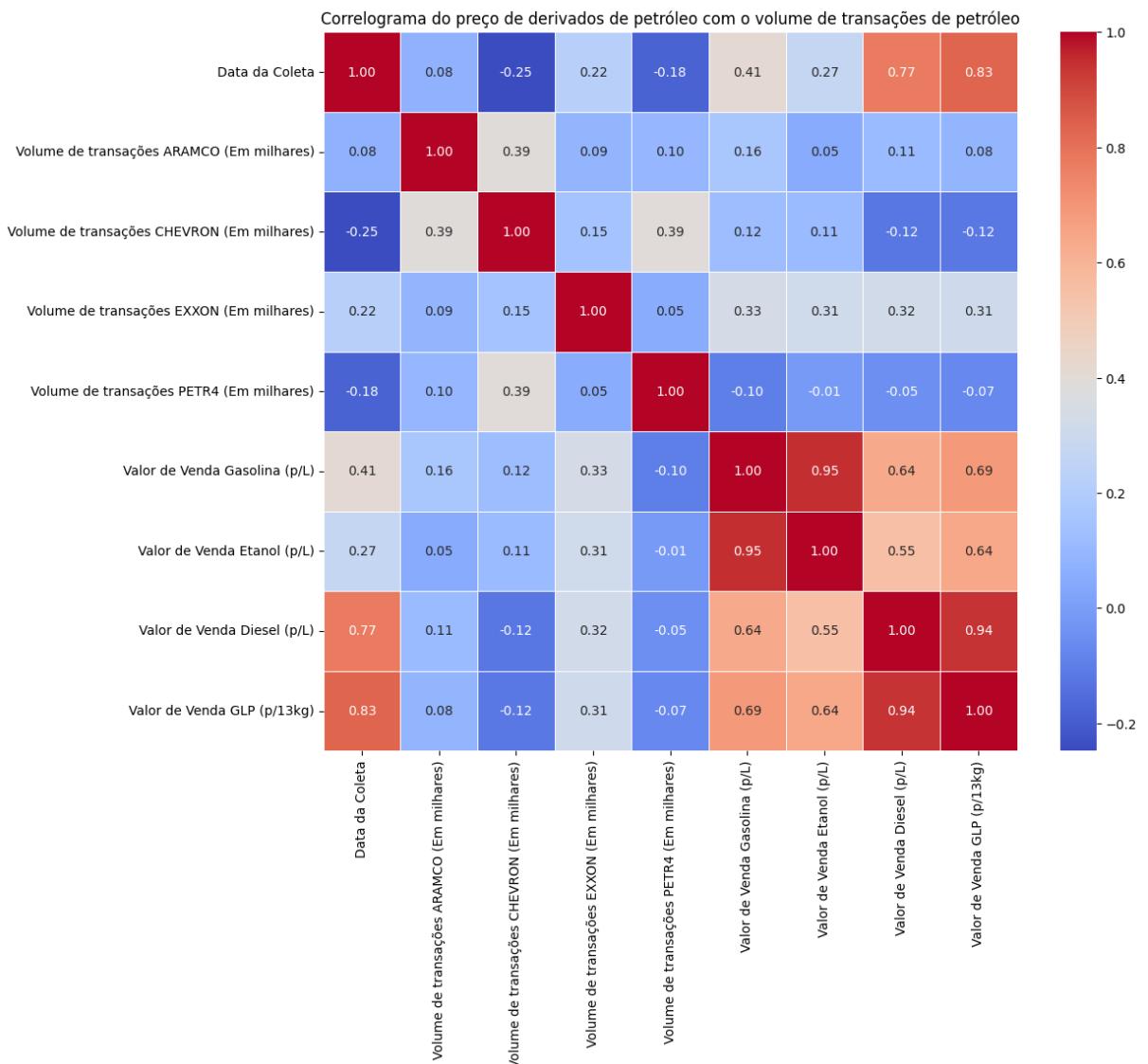
Também é notável a forte correlação positiva entre os preços da gasolina e do etanol, assim como entre os preços do diesel e do GLP. Estes resultados abrem possibilidades para futuras investigações mais aprofundadas.



Contrariando as expectativas, a correlação entre a cotação do Dólar e do Riyal com os preços dos derivados de petróleo no Brasil mostrou-se quase inexistente. No entanto, ambas as moedas apresentam uma relação ligeiramente mais forte com o preço do etanol, embora essa correlação ainda seja considerada baixa.



As correlações entre os valores de fechamento das petrolíferas e o preço do Diesel e do GLP é relativamente forte, enquanto a correlação delas com os valores da Gasolina e do Etanol são mais fracos. A única com relações mais forte com a Gasolina e Etanol é a ARAMCO, mas ainda não é uma relação extremamente forte. Assim, essa observação abre possibilidade de uma expansão de pesquisa futura, investigando o motivo pelo qual o Fechamento das ações da ARAMCO tem uma relação expressivamente maior com o valor da gasolina e do etanol no país.



Por outro lado, o volume de transações mostra uma correlação extremamente fraca com os preços de todos os derivados de petróleo, com uma ligeira elevação na EXXON.

Discussão a respeito das explorações

As análises exploratórias mostraram padrões importantes, como a forte correlação entre os preços de combustíveis derivados de petróleo. Foi observada uma tendência de aumento dos preços durante a pandemia de COVID-19, especialmente em 2020 e 2021. No entanto, a falta de dados em períodos específicos e discrepâncias, como a lacuna de informações sobre o Riyal, apresentaram desafios.

Além disso, notou-se que variáveis como o preço das ações de grandes petrolíferas (ARAMCO, CHEVRON e EXXON) e seus volumes de transações

apresentaram relações mais significativas com os preços do diesel e do GLP, enquanto a Petrobras e o Dólar tiveram menos influência do que inicialmente esperado. Isso indicou a necessidade de expandir o escopo de variáveis para capturar melhor as dinâmicas do mercado.

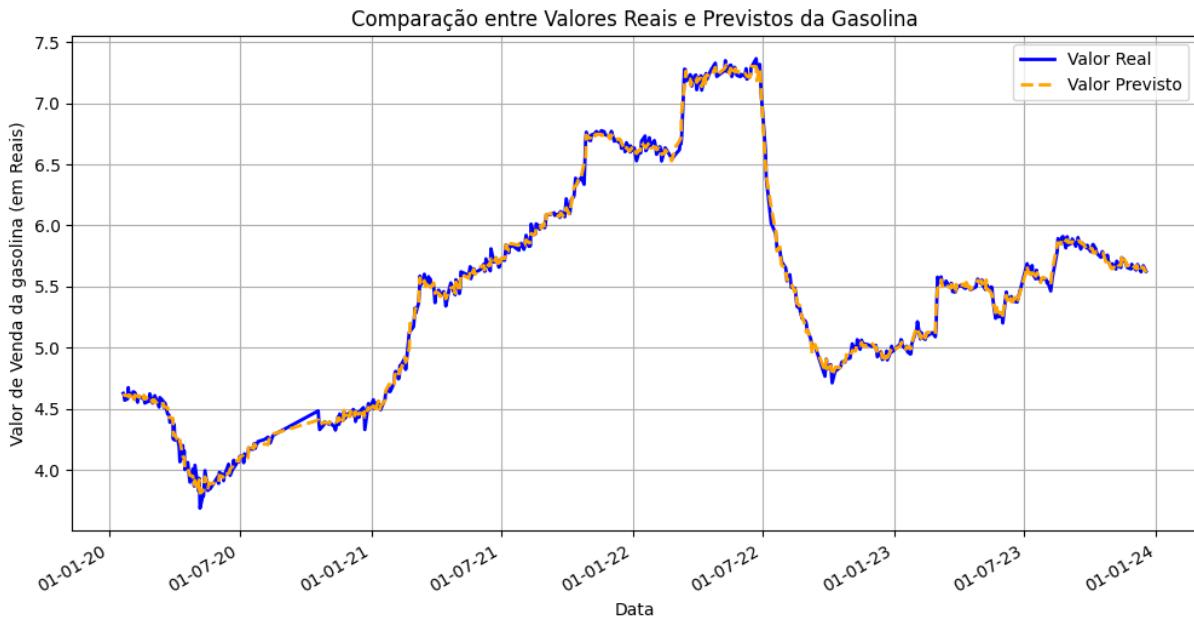
Por outro lado, as correlações positivas entre gasolina e etanol, e entre diesel e GLP, reforçam a interdependência desses derivados, abrindo espaço para estratégias preditivas baseadas nesses pares. A detecção e tratamento de outliers, por meio de intervalos definidos, contribuíram para suavizar os dados e melhorar a visualização, o que foi essencial para garantir a robustez das análises subsequentes.

Análises Preditivas

Ao analisar correlações com os valores *target* apresentados anteriormente, foi decidido que, para cada um dos derivados de petróleo, iremos basear os modelos nas seguintes variáveis:

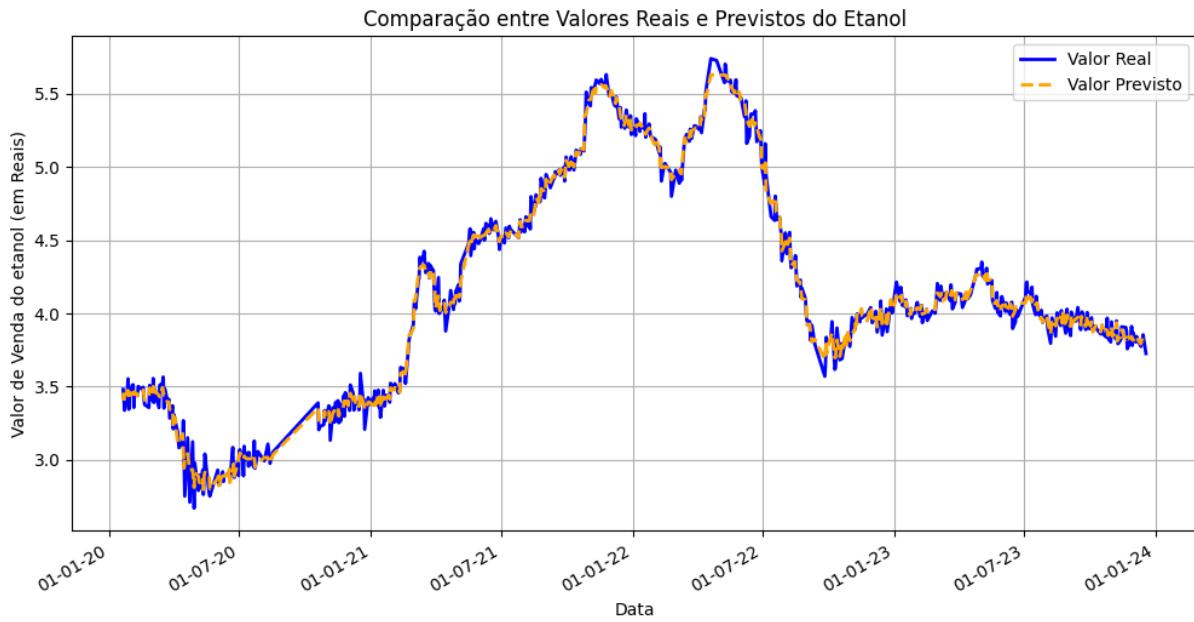
- Gasolina: Data de Coleta, Volume de Transações da EXXON e Fechamento da ARAMCO e CHEVRON.
- Etanol: Data de Coleta, Volume de Transações da EXXON, Fechamento da ARAMCO.
- Diesel: Data de Coleta, Volume de Transações da EXXON, Fechamento da ARAMCO, CHEVRON e EXXON.
- GLP: Data de Coleta, Volume de Transações da EXXON, Fechamento da ARAMCO, CHEVRON e EXXON.

Após definidas as variáveis de predição de cada um dos derivados de petróleo, podemos seguir com os resultados.



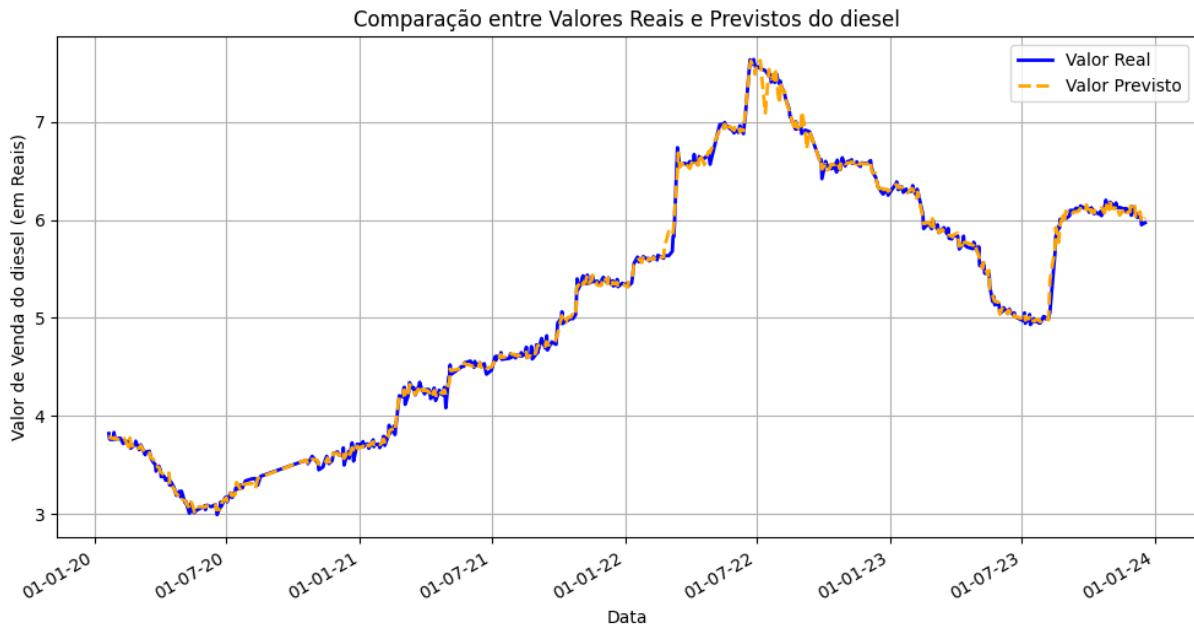
	MAE	RMSE	MAPE
Gasolina	0.03	0.04	0.49%

Para a gasolina, o modelo obteve um erro absoluto médio (MAE) de R\$ 0,03 por litro, o que representa uma diferença mínima entre os valores reais e previstos. A raiz do erro quadrático médio (RMSE), de R\$ 0,04, está próxima do MAE, indicando que os erros são uniformemente distribuídos, sem discrepâncias significativas. Além disso, o erro percentual médio (MAPE) foi de apenas 0,49%, um valor excepcionalmente baixo. Isso demonstra que o modelo consegue captar com alta precisão as variações do preço da gasolina, tanto em termos absolutos quanto relativos.



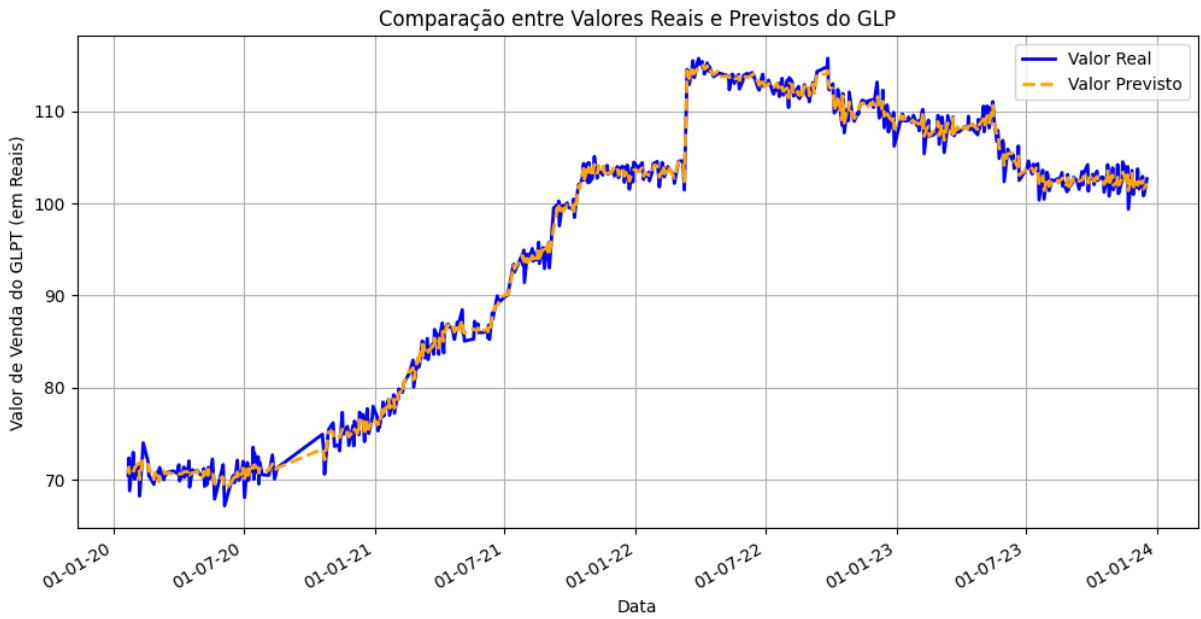
	MAE	RMSE	MAPE
Etanol	0.04	0.05	0.99%

No caso do etanol, o modelo apresentou um MAE de R\$ 0,04 por litro, um pouco superior ao da gasolina, mas ainda dentro de uma faixa muito pequena de erro. O RMSE, de R\$ 0,05, é ligeiramente mais elevado, sugerindo que erros maiores podem ocorrer ocasionalmente, embora não em magnitude preocupante. O MAPE, de 0,99%, indica que o erro percentual médio é inferior a 1%, o que é um excelente resultado. No geral, o modelo conseguiu prever bem os preços do etanol, embora haja espaço para pequenos ajustes visando reduzir discrepâncias ocasionais.



	MAE	RMSE	MAPE
Diesel	0.03	0.05	0.60%

Para o diesel, o MAE foi de R\$ 0,03 por litro, igual ao da gasolina, enquanto o RMSE foi de R\$ 0,05, levemente maior, sugerindo uma maior sensibilidade a erros isolados. Ainda assim, o MAPE de 0,60% é extremamente baixo, reafirmando a robustez do modelo. Assim como na gasolina, o modelo demonstrou alto nível de precisão na previsão dos preços do diesel, captando bem as flutuações ao longo do tempo.



	MAE	RMSE	MAPE
GLP	0.53	0.69	0.58%

O desempenho do modelo para o GLP também foi satisfatório, considerando a faixa de preços observada, que variou entre R\$ 70 e R\$ 120 durante a análise. O MAE foi de R\$ 0,53, e o RMSE de R\$ 0,69, indicando erros absolutos maiores em comparação aos outros derivados, mas que ainda representam menos de 1% do intervalo de variação. O MAPE, de 0,58%, reflete um erro percentual médio muito baixo, demonstrando que, proporcionalmente, o modelo conseguiu prever os preços com alta precisão. Esse resultado sugere que os valores absolutos maiores de erro não comprometem a qualidade geral das previsões, dada a escala mais elevada dos preços do GLP.

Discussão a respeito das previsões

Os modelos de previsão foram implementados utilizando o LGBMRegressor, uma técnica de aprendizado de máquina baseada em gradient boosting, altamente eficaz para tarefas de regressão. Esta escolha foi motivada pela sua capacidade de lidar com dados não estacionários e de capturar interações complexas entre variáveis, superando abordagens tradicionais como ARIMA e Prophet em precisão e flexibilidade e em exigência de processamento.

Os resultados mostraram um desempenho sólido em todas as variáveis target, com erros percentuais médios (MAPE) inferiores a 1%. Destaca-se a previsão para a gasolina, que apresentou o menor erro absoluto médio (MAE de R\$ 0,03 por litro) e um RMSE próximo, sugerindo uma boa uniformidade nos erros. Para o GLP, embora os valores absolutos de erro tenham sido mais elevados devido à escala de preços maior, o MAPE foi de apenas 0,58%, evidenciando a precisão proporcional do modelo.

A estratégia de segmentar as variáveis preditoras específicas para cada derivado de petróleo foi essencial para o sucesso dos modelos. Por exemplo, ao utilizar o fechamento das ações da ARAMCO e da CHEVRON para o diesel e o GLP, o modelo conseguiu capturar as nuances de mercado com maior eficácia. Essa abordagem garantiu que os dados com maior relevância contribuíssem diretamente para as previsões.

Considerações finais

O estudo demonstrou que, embora as relações esperadas inicialmente entre o Dólar e os derivados de petróleo tenham sido fracas, a inclusão de novas variáveis permitiu melhorar significativamente os modelos de previsão. As análises exploratórias confirmaram a interdependência entre os preços de derivados e as influências macroeconômicas, enquanto as previsões mostraram-se robustas, com baixa margem de erro.

Como principais contribuições, o trabalho fornece um framework replicável para a análise de séries temporais em contextos econômicos, destacando o impacto de variáveis globais e locais no mercado de combustíveis no Brasil. Além disso, os resultados obtidos podem servir como base para decisões mais informadas por stakeholders, incluindo formuladores de políticas públicas e agentes do mercado.

Por fim, recomenda-se que estudos futuros expandam a coleta de dados para incluir mais fatores macroeconômicos, como taxas de juros e políticas fiscais, além de investigar técnicas híbridas de aprendizado de máquina para potencializar a precisão dos modelos. A integração de métodos avançados, como redes neurais profundas, também poderá oferecer insights mais profundos e úteis para o setor energético.

Referências

- CHEN, C. W. S.; LIU, F.C.; SO, M.K.P. **A review of threshold times series models in finance.** Statistics and interface, 2011. Disponível em: <https://www.intlpress.com/site/pub/pages/journals/items/sii/content/vols/0004/0002/a012/>. Acesso em: 13 dezembro 2023.
- BLINDER, K.E.; POURAHMADI, M. and MJELDE, J.W. **The role of temporal dependence in factor selection and forecasting oil prices.** Empirical Economics. 2017. Disponível em: <https://doi.org/10.1007/s00181-018-1574-9>. Acesso em: 25 de março de 2024.
- IDREES, S. M.; ALAM, M. A. and AGARWAL, P. **A Prediction Approach for Stock Market Volatility Based on Time Series Data,** in IEEE Access, vol. 7, p. 17287 - 17298, 2019. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8626097> Acesso em: 25 de março 2024.
- PAI, Ping-Feng; LIN, Kuo-Ping; LIN, Chi-Shen; CHANG, Ping-Teng. **Time series forecasting by a seasonal support vector regression model.** Journal Expert Systems with Applications. V. 37, p. 4261- 4265. 2010. Disponível em: www.elsevier.com/locate/eswa. Acesso em: 26 de março de 2024.
- PALMA, W. **Time series analysis.** Wiley & Sons. 2019. ISBN 978-1118634325.
- SHAH, Dev; ISAH, Haruma; ZULKERNINE, Farhana. **Stock Market Analysis: A Review and Taxonomy of Prediction Techniques.** International Journal of Financial Studies. 2019. DOI 103390/ijfs7020026.
- SHUMWAY, R. H.; STOFFER, D. S. **Time Series Analysis and Its Applications: With R Examples.** 1. ed. New York: Springer, 2006. ISBN 978-1441978646.
- WIRTH, Rüdiger; HIPP, Jochen. **CRISP-DM: Towards a Standard Process Model for Data Mining.** Proceedings of the 4th international conference on the practical applications. 2000. Disponível em: <http://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf>. Acesso em: 26 de março de 2024.
- ZHANG, Yanpeng; QU, Hua; WANG, Weipeng; ZHAO, Jihong. **A Novel Fuzzy Time Series Forecasting Model Based on Multiple Linear Regression and Series Clustering.** Mathematical Problems in Engineering. Article ID 9546792. Hindawi. 2020. Disponível em: <https://doi.org/10.1155/2020/9546792>. Acesso em: 26 de março de 2024.
- ZHANG, Y.; YAN, B.; AASMA, M. **A novel deep learning framework: Prediction and analysis of financial time series using CEEMD and LSTM.** Expert Systems with Applications, v. 159, 2020. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0957417420304334>. Acesso em: 11 de dezembro 2023.