



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DO SUL DE MINAS GERAIS CAMPUS MACHADO
CURSO BACHARELADO EM SISTEMAS DE INFORMAÇÃO**

**Gabriel Caproni Pegoraro
Guilherme Henrique de Souza Silva
Luiz Henrique Silvério de Souza
Pedro Ferreira Franco
Sander Gustavo Piva**

Proposta de Heurística: Modelagem Macroeconômica de Ativos

MACHADO

2025

1. Qual problema geral será resolvido?

O problema geral a ser resolvido é a Quantificação da Sensibilidade do Retorno de Ações a Fatores Macroeconômicos.

Em essência, a heurística busca explicar e medir o quanto as variáveis de risco e de política monetária (como juros, inflação e risco país) influenciam a volatilidade e o retorno diário/mensal de ações de empresas brasileiras de capital aberto. O objetivo é transformar o complexo ambiente econômico em um modelo matemático interpretável.

2. Qual o domínio/negócio em que o problema está inserido?

- Descrição: O problema está inserido no domínio do Mercado Financeiro e da Econometria Aplicada a Finanças (Asset Pricing). Especificamente, ele se aplica à análise de Renda Variável, onde a precificação dos ativos é crucial.
- Cenário: O cenário é o Mercado de Capitais Brasileiro, que é altamente influenciado pelas decisões do Banco Central (Selic¹, Câmbio²) e pela percepção global de risco (Risco Brasil (CDS)³). O ambiente é caracterizado pela interconexão entre o micro (preço da ação) e o macro (economia do país).
- Atores Impactados: Gestores de Fundos, que precisam alocar capital de forma eficiente; Analistas de Mercado, que criam relatórios de *valuation*

3. Quais os objetivos da heurística?

Objetivo	Descrição
Objetivo Principal (Maximização)	Maximizar a Capacidade Explicativa do modelo sobre a variação do Retorno das Ações.

¹ Selic: Juro básico da economia brasileira que influencia todas as outras taxas de juros.

² Câmbio: Preço de uma moeda estrangeira (ex: Dólar) em relação à moeda nacional (Real).

³ Risco Brasil (CDS): Medida do custo de seguro contra calote da dívida pública do Brasil; indica o risco percebido pelos investidores.

Objetivo Secundário (Viabilidade)	Fornecer uma solução rápida e aproximada (heurística) para a precificação de ativos, utilizando um modelo linear simples (Regressão Múltipla) em vez de modelos financeiros complexos (como APT ou Fama-French).
Critério Sucesso	Obter um modelo onde pelo menos uma das variáveis macroeconômicas (Selic, Cambio, Risco País) seja estatisticamente significativa (p-valor < 0.05) para o Retorno das Ações.

4. Qual a solução (método) proposta?

Tópico	Detalhamento da Heurística
Método Proposto	Regressão Linear Múltipla.
Funcionamento (Heurística Aproximação)	A Regressão Múltipla serve como uma heurística de aproximação ao assumir uma relação linear entre as variáveis macroeconômicas e o retorno da ação. Ela estima coeficientes que quantificam a variação esperada no retorno da ação (Y) para cada unidade de variação nos fatores econômicos (X), simplificando a complexidade do mercado.
Função do Modelo	$\text{Retorno da Ação}(Y) \approx \beta_0 + \beta_1 \cdot \text{Selic} + \beta_2 \cdot \text{Cambio} + \beta_3 \cdot \text{Risco(CDS)}$
Justificativa para a Heurística	A Regressão Linear é a heurística mais apropriada por ser computacionalmente eficiente, extremamente interpretável (os coeficientes explicam o impacto do fator) e fornece uma base sólida e rápida para entender a sensibilidade da carteira aos fatores de risco sistêmicos, atendendo à limitação de tempo e recursos do projeto.

5. Quais as métricas para avaliar se a solução é satisfatória?

Métrica	Descrição	Critério de Satisfação
Poder Explicativo (Acurácia)	R^2 Ajustado (Coeficiente de Determinação). Mede a porcentagem da variação no Retorno da Ação que é explicada pelas variáveis macroeconômicas.	Um R^2 Ajustado positivo (acima de zero) e o mais próximo possível de 1,0.
Relevância Fator (Significância)	p-valor dos Coeficientes. Mede a probabilidade de um coeficiente ser zero (ou seja, a variável não ter impacto).	p-valor < 0.05 para as variáveis macro selecionadas.
Eficiência (Tempo Execução)	Tempo de execução do cálculo do modelo de regressão no ambiente de programação.	O modelo deve ser executado em milissegundos (muito rápido), comprovando a eficiência do método heurístico.

6. Quais as limitações da heurística proposta?

Limitação	Explicação
Suposição Linearidade	A heurística assume que a relação entre economia e preço das ações é linear, o que é uma simplificação. Na realidade, choques de mercado são frequentemente não-lineares.
Limitação Causalidade	O modelo de regressão é explicativo, mas não estabelece causalidade. Ele pode apenas mostrar a correlação, sendo sensível a fenômenos não incluídos no modelo (ex: eventos políticos, notícias).

7. Quais são os dados necessários e como serão obtidos?

Dado Necessário	Tipo Frequência	Fonte de Obtenção	Tratamento Essencial
Retorno da Ação (Y)	Preço Fechamento Ajustado (Diário)	API yfinance para ações: -ITUB4 (Banco Itaú) -PETR4 (Petrobrás) -VALE3 (Vale do Rio Doce). Link: api_preco_acao	Criação da Variável: Calcular o retorno diário a partir do preço. Cotação em reais
Selic	Taxa de Juros (Diário)	Banco Central do Brasil (SGS), Pesquisa textual: Selic, código: 432. Link: bcb_tx_selic	Uniformização da data dos registros, transformação tipos de dados, verificação de valores nulos, remoção eventual de registros conforme critério técnico
Câmbio	Dólar Americano (Diário)	Banco Central do Brasil (SGS), Pesquisa textual: Taxa Câmbio, código: 1 . Link: bcb_tx_cambio	Uniformização da data dos registros, transformação tipos de dados, verificação de valores nulos
Risco Brasil	(CDS)	Investing.com Link: investing.com_risco_cds	Uniformização da data dos registros, transformação tipos de dados, verificação de valores nulos, cálculo retorno logaritmo

8 Etapas para obtenção da base de dados (dataframe)

As etapas para obtenção do dataframe do projeto deverão seguir, em linhas gerais, os seguintes passos:

1. Serão carregados os dataframes referentes à Selic e ao Câmbio. Uma limpeza inicial será realizada
2. Observando as datas do dataframe da Selic e Câmbio, o número de registros da Selic é um pouco maior. Selecionar os registros do dataframe Selic de acordo com as datas do dataframe do Câmbio⁴
3. Realizar pré-processamento e união (merge) dos dataframes
4. Carregar dataframe dos preços das ações (fechamento) do banco Itaú (ITUB4), Petrobrás (PETR4) e Vale do Rio Doce (VALE3), usando API. Realizar uma limpeza inicial, se necessário
5. Selecionar os registros do dataframe dos preços de acordo com as datas do dataframe do Câmbio
6. Realizar pré-processamento e merge dos dataframes
7. Carregar o dataframe correspondente ao Risco Brasil (CDS)
8. Selecionar os registros do dataframe Risco Brasil (CDS) de acordo com as datas do dataframe do Câmbio
9. Realizar pré-processamento e merge dos dataframes. Salvar dataframe parcial na extensão 'csv'
10. Calcular o retorno logaritmo dos preços das ações e salvar como o dataframe "final_merged_dfs_with_log_returns.csv"

Nesse ponto, a princípio, uma exploração inicial com gráficos deverá ser aplicada, seguido do aprendizado de máquina Regressão Múltipla. Uma vez validado o modelo, após eventuais ajustes e testes complementares, este será exportado na forma de um arquivo JSON para ser utilizado em um app Flutter.

⁴ datas do dataframe do Câmbio: Adotou-se essa referência, pois os dataframes precisam ser do mesmo tamanho para uni-los (merge). Para evitar inferência, optou-se pelo menor dataframe ~1250 registros

9. Coleta

Em relação a coleta, o respectivo projeto foi composto pela união de distintos conjuntos de dados (dataframes) conforme mencionado nas seções anteriores. Os dataframes empregados foram: Taxa Selic, Taxa de Câmbio, Risco Brasil - CDS e os preços das ações das empresas (Itaú, Petrobras, Vale do Rio Doce).

Para facilitar o acesso, disponibiliza-se os links a seguir:

- Taxa Selic: [df_selic](#)
- Taxa Câmbio: [df_cambio](#)
- Risco Brasil - CDS: [df_cds](#)

Os preços das ações citadas, embora a fonte de dados original seja oriunda do link: [df_preco](#), na prática, utilizou-se uma API (código auxiliar) do Yahoo Finances para extração. O código em python encontra-se neste link: [api_preco_acao](#)

10. Limpeza

Uma vez que os dados foram carregados, em relação ao procedimento de limpeza, as verificações básicas foram aplicadas para checar a presença de valores missing ou similares. A princípio, não houve problemas nessa etapa. A seguir, alguns links com trechos de código com as verificações:

- Dataframe Selic: [carregar_df_selic_verificar_valores_missing](#)
- Dataframe Cambio: [carregar_df_cambio_verificar_valores_missing](#)
- Dataframe CDS: [carregar_df_cds_verificar_valores_missing](#)
- Dataframe Preco Acoes: [carregar_df_preco_acoes_verificar_valores_missing](#)

11. Pré-processamento e Transformação

O processo de pré - processamento e transformação dos dados foi fundamental para consolidar as diversas fontes de informação em uma base única e consistente, apta a alimentar o modelo de Regressão Linear Múltipla. Este procedimento foi estruturado sequencialmente, iniciando-se pela carga e verificação das bases macroeconômicas primárias: a Taxa Selic e a Taxa de Câmbio.

Inicialmente, procedeu-se à leitura dos arquivos brutos (formato .csv). Para a Taxa de Câmbio, foi necessário um tratamento específico na conversão da coluna de datas para o formato datetime, garantindo a ordenação cronológica correta. Como observado na análise exploratória, a base do Câmbio possuía um número de registros inferior à da Selic. Link: [conversao_coluna_datas_df_cambio](#)

Dessa forma, adotou-se a base de Câmbio como referência temporal, filtrando o *dataframe* da Selic para manter apenas as datas coincidentes, assegurando a integridade do alinhamento temporal entre as variáveis. Link: [filtrar_df_selic](#)

Após a padronização das datas e renomeação das colunas para termos mais intuitivos (ex: Taxa Selic - a.a. e Taxa Cambio u.m.c./US\$), realizou-se a primeira união (*merge*) entre os dados da Selic e do Câmbio, utilizando a data como chave de ligação (índice). Links:

- Renomeação coluna selic: [renomear_coluna_selic](#)
- Renomeação coluna câmbio: [renomear_coluna_cambio](#)
- Merge dos frames selic, câmbio e CDS: [merge_df_selic_com_df_cambio](#)

Na sequência, foram integrados os dados de renda variável. O *dataframe* contendo os preços históricos de fechamento das ações (Itaú, Petrobras e Vale) passou pelo mesmo processo de validação de datas e conversão para *datetime*. Em seguida, executou-se um novo *merge* para unificar esses preços à base macroeconômica já consolidada. Link: [validacao_conversao_datetime_merge_df_selic_cambio_preco](#)

A etapa seguinte envolveu a incorporação do fator de risco, o Risco Brasil (CDS). O *dataframe* correspondente foi carregado e suas datas convertidas. Link: [ajustes_conversao_tipo_datetime_df_cds](#)

Aplicou-se um filtro para selecionar apenas os registros de CDS que correspondiam às datas já existentes no conjunto de dados principal. Link: [filtrar_datas_df_cds](#)

Logo, procedeu-se à união dos frames. Link: [merge_df_selic_cambio_preco_cds](#)

Com todas as variáveis integradas, realizou-se uma reorganização das colunas para uma ordem lógica e a renomeação das variáveis de ações para os nomes das companhias (Itaú, Petrobras, Vale Rio Doce), visando facilitar a interpretação. Links:

- Reorganização das colunas do frame: [reorganizacao_colunas](#)
- Renomeação das colunas do frame: [renomear_colunas](#)

Um tratamento crítico foi aplicado à coluna da Taxa Selic. Originalmente importada como texto contendo símbolos percentuais e vírgulas, a variável foi convertida para tipo numérico (*float*) e sua escala foi ajustada (divisão por 100).

- Conversão para tipo float: [converter_tipo_float_coluna_selic](#)
- Ajuste de escala: [ajustar_escala_coluna_selic](#)

Para manter a consistência matemática com as demais taxas. Adicionalmente, aplicou-se arredondamento em diversas colunas para padronização das casas decimais. Link: [arredondar_valores](#)

Na sequência, o *dataframe* parcial foi salvo. Link: [salvar_df_parcial](#)

Depois de carregar o dataframe parcial em uma variável, executou-se a engenharia de atributos (*feature engineering*) necessária para a modelagem financeira. Foi calculado o Retorno Logarítmico para cada ativo (Itaú, Petrobras e Vale), utilizando a fórmula $\ln(P_t/P_{t-1})$. Essa transformação é essencial para estabilizar a variância e normalizar a série temporal de preços.

- Cálculo retorno logaritmo das ações: [calcula_retorno_log_preco_acoes](#)

O conjunto de dados resultante, contendo as variáveis macroeconômicas originais e os retornos calculados, foi salvo no arquivo 'final_merged_dfs_with_log_returns.csv'.

- Salvar frame: [salvar_frame_com_retornos_log_precos_acoes](#)

12. Modelo de Regressão Linear Múltipla: Preparação

Antes de executar o modelo preditivo, observou-se a necessidade de realizar ajustes finais no dataframe consolidado, especificamente a inclusão das variáveis transformadas 'RETORNO_LOG_CAMBIO' e 'RETORNO_LOG_CDS'.

A opção pela transformação dos preços e taxas em retornos logarítmicos (valores relativos) fundamenta-se na necessidade de modelagem econométrica e financeira. Para análises de risco e regressão, é crucial trabalhar com a variação percentual dos ativos e não com seus valores absolutos, garantindo a estacionariedade das séries temporais.

Houve, contudo, uma exceção importante: a Taxa Selic. O cálculo de um "retorno logarítmico" para a Selic resultaria na variação percentual diária de uma taxa que já é anual, dificultando a interpretação econômica no contexto de retornos diários de ativos. A teoria econômica sugere que o nível da taxa de juros (e não sua variação percentual) é o fator determinante para o custo de oportunidade e precificação. Portanto, a Selic foi mantida em sua forma original (taxa nominal) para medir o efeito do nível da política monetária sobre os retornos.

12.1 Tratamento dos fatores macroeconômicos

Nesta etapa, procedeu-se à criação das colunas de retorno logarítmico para as variáveis independentes de Câmbio e Risco Brasil (CDS). Utilizando a biblioteca NumPy, calculou-se o logaritmo natural da razão entre o valor atual e o valor do dia anterior (lag de 1 dia). Em seguida, os resultados foram arredondados para seis casas decimais para manter a precisão numérica. Link:

- Cálculo retorno log (câmbio + CDS): [calcula_retorno_log_cambio_cds](#)

12.2 Limpeza de dados (NaN) e Definição de Variáveis

A operação de cálculo dos retornos (que envolve lags) e a conversão da Selic geraram, inevitavelmente, valores nulos (NaN) nas primeiras linhas da série temporal. Para garantir a robustez do modelo, removeu-se qualquer registro que contivesse valores nulos nas colunas de interesse ('Taxa Selic - a.a.', 'RETORNO_LOG_CAMBIO', 'RETORNO_LOG_CDS' e demais retornos).

Após a limpeza, definiu-se o conjunto de variáveis independentes (X) composto pela Selic, Retorno do Câmbio e Retorno do CDS. Adicionalmente, foi inserida uma constante ao modelo (intercepto), requisito estatístico para a estimação correta da regressão linear via Ordinary Least Squares (OLS). Link:

- Limpeza dados (NaN) e definição de variáveis independentes: [limpeza_apos_criacao_retornos_log_cambio_cds_definicao_var_x](#)

12.3 Exportação do Dataframe Final

Com todos os ajustes realizados, o dataframe definitivo, contendo tanto os dados originais quanto as variáveis transformadas necessárias para o modelo preditivo, foi salvo em formato CSV sob o nome `df_final_model.csv`. Este arquivo serve como base imutável para a execução das regressões que serão apresentadas na seção de resultados. Link: [df_final_model](#)

13. Gráficos

Nesta etapa, procedeu-se à análise visual dos dados por meio de gráficos de série temporal e de correlação. O objetivo foi identificar tendências, volatilidade e relações preliminares entre as variáveis macroeconômicas e os preços dos ativos antes da modelagem econométrica. A geração dos gráficos foi realizada utilizando as bibliotecas `matplotlib` e `seaborn`, conforme detalhado no código do projeto.

13.1 Análise dos Fatores Macroeconômicos

Inicialmente, foram plotadas as séries temporais dos fatores macroeconômicos: Taxa Selic, Taxa de Câmbio e Risco Brasil (CDS). A análise visual revela insights importantes sobre a evolução econômica no período estudado.

Em relação à Taxa Selic, observa-se uma política monetária cíclica. A série inicia-se em um período de taxas baixas (política expansionista), seguida por uma elevação significativa e rápida, refletindo a atuação do COPOM para conter a inflação ou estabilizar o cenário macroeconômico, e termina com a taxa sustentada em um patamar mais elevado.

O gráfico da Taxa de Câmbio (U.M.C./US\$) caracteriza-se por uma volatilidade considerável. Foram registrados momentos de forte desvalorização da moeda

nacional, marcados por picos que sugerem reações a instabilidades políticas ou incertezas econômicas globais e internas. A série demonstra sensibilidade a choques, com vales e picos bem definidos.

Já a série do CDS (Risco Brasil), que mensura a percepção de risco de crédito do país, demonstra uma tendência geral de melhora na confiança. O gráfico inicia com níveis de risco mais elevados, associados a uma maior fragilidade, e apresenta uma queda gradual ao longo do tempo, indicando um aumento da confiança dos investidores na solvência brasileira.

- Ver código de geração dos gráficos (Seção 16.1): [grafico1](#)

13.2 Retornos Logarítmicos dos Fatores Macroeconômicos

Para adequar as variáveis ao modelo de regressão linear e analisar a volatilidade percentual, foram gerados os gráficos dos retornos logarítmicos do Câmbio (RETORNO_LOG_CAMBIO) e do CDS (RETORNO_LOG_CDS). Essas séries transformadas permitem visualizar a estacionariedade e a magnitude das variações diárias, removendo a tendência de longo prazo observada nos preços brutos.

- Ver código de geração dos gráficos (Seção 16.2): [grafico2](#)

13.3 Análise dos Preços das Ações

A análise das séries temporais dos preços das ações do Itaú (ITUB4), Petrobras (PETR4) e Vale (VALE3) revelou comportamentos distintos, influenciados por seus respectivos setores.

O ativo Itaú (ITUB4) apresentou uma tendência geral de valorização no longo prazo, apesar de flutuações com quedas e recuperações. A ação demonstra resiliência, sendo influenciada principalmente pela taxa Selic e pela saúde geral da economia doméstica.

A Petrobras (PETR4) exibiu alta volatilidade, com picos e vales acentuados, refletindo sua forte sensibilidade ao preço internacional do petróleo, fatores geopolíticos e mudanças na gestão ou políticas governamentais. A tendência mostrou-se mais errática, porém com ganhos expressivos em períodos favoráveis à commodity.

As ações da Vale (VALE3) apresentaram a volatilidade típica do setor de commodities, fortemente correlacionadas com a demanda por minério de ferro e a variação cambial. O gráfico exhibe fortes oscilações, mas mantém uma tendência de valorização no longo prazo, alinhada aos ciclos globais de materiais básicos.

- Ver código de geração dos gráficos (Seção 16.3): [grafico3](#)

13.4 Retornos Logarítmicos das Ações

De maneira análoga aos fatores macroeconômicos, foram plotados os retornos logarítmicos das ações. Esta visualização é crucial para identificar a dispersão dos retornos (risco) e confirmar a adequação dos dados para a Modelagem de Regressão Múltipla, focando na variação diária em vez do preço nominal.

- Ver código de geração dos gráficos (Seção 16.4): [grafico4](#)

13.5 Matriz de Correlação

Por fim, foi gerada uma matriz de correlação (mapa de calor) para quantificar a relação linear entre as variáveis independentes (Selic, Retorno Câmbio, Retorno CDS) e as variáveis dependentes (Retornos de ITUB4, PETR4 e VALE3).

A análise do mapa de calor permite validar as hipóteses econômicas iniciais. Observou-se, por exemplo, o comportamento inverso esperado entre risco e retorno: à medida que os valores do Câmbio ou do risco CDS aumentam (indicando deterioração do cenário), os retornos logarítmicos das ações tendem a diminuir, evidenciando uma correlação negativa. Essa ferramenta visual foi fundamental para confirmar quais variáveis macroeconômicas possuem maior poder explicativo sobre cada ativo antes da execução do modelo preditivo.

- Ver código de geração da matriz (Seção 16.5): [correlacao](#)

14. Modelo de Regressão Linear Múltipla: Execução e Resultados

Nesta etapa, o modelo de Regressão Linear Múltipla foi executado iterativamente para cada um dos ativos (Itaú, Petrobras e Vale), testando diferentes combinações de variáveis independentes (Fatores Macroeconômicos) para identificar a configuração com melhor poder explicativo (R^2 Ajustado) e significância estatística (p -valor <0.05).

14.1 Teste 1: Modelo Completo (Selic + Câmbio + CDS)

Inicialmente, o modelo foi rodado considerando todas as variáveis macroeconômicas disponíveis: Taxa Selic, Retorno do Câmbio e Retorno do CDS.

A análise dos resultados indicou que essa combinação não é válida. Observou-se que a variável Taxa Selic não atendeu ao critério de significância estatística ($p<0.05$) para a maioria dos ativos. No caso da Vale do Rio Doce, por exemplo, o p -valor para a Selic foi de aproximadamente 0.10, o que sugere que, neste modelo multivariado, a Selic não contribui significativamente para explicar a variação dos retornos diários. Link: Combinação completa das variáveis: [mp_selic_cambio_cds_summary](#)

14.2 Teste 2: Modelo sem Selic (Câmbio + CDS)

Diante da insignificância da Selic, procedeu-se à execução de um segundo modelo contendo apenas as variáveis Retorno do Câmbio e Retorno do CDS.

Esta combinação mostrou-se válida para todos os casos analisados. Ambas as variáveis apresentaram significância estatística ($p < 0.05$) para as três ações. Destaca-se o resultado para o Itaú, onde o modelo obteve o maior poder explicativo entre os testes, com um R2 Ajustado de aproximadamente 0.074 (7.4%). Link:

- Combinação Câmbio + CDS: [mp_cambio_cds_summary](#)

14.3 Testes Complementares (Sem Câmbio e Apenas CDS)

Para fins de validação, foram realizados outros dois testes:

- Sem Câmbio (Selic + CDS): A combinação mostrou-se inválida, pois a Taxa Selic continuou apresentando p-valor acima do limite de tolerância (0.05). Link: [mp_selic_cambio_summary](#)
- Apenas CDS: Embora válida estatisticamente ($p < 0.05$), o poder explicativo (R2 Ajustado) foi consideravelmente inferior ao modelo que inclui o Câmbio, reduzindo a utilidade da heurística. Link: [mp_cds_summary](#)

14.4 Considerações Finais e Seleção do Modelo

Com base nos testes realizados, a combinação Câmbio + Risco Brasil (CDS) foi a selecionada para o modelo final da heurística.

A escolha justifica-se pelos seguintes fatores:

- Equilíbrio: Oferece o melhor compromisso entre simplicidade (parsimônia) e capacidade de explicar a variação nos retornos das ações.
- Significância: Ambas as variáveis (RETORNO_LOG_CAMBIO e RETORNO_LOG_CDS) são estatisticamente significativas para todos os ativos.
- Multicolinearidade Aceitável: A presença de multicolinearidade (indicada pelo Condition Number) sugere que Câmbio e CDS se movem juntos, o que é esperado economicamente. Contudo, isso não invalida a capacidade preditiva do modelo conjunto, sendo uma limitação aceitável para a proposta da heurística.

15. Salvar modelo preditivo no formato JSON

Após a obtenção do melhor modelo preditivo, foi implementado um código complementar para se obter o respectivo modelo no formato de um arquivo JSON (chave: valor) de modo que pudesse servir de backend para uma aplicação em Flutter. Links:

- Salvando modelo preditivo (Câmbio+CDS) em formato JSON:
[salvar_mp_cambio_cds_em_formato_json](#)
- Fazendo download arquivo JSON:
[download_mp_cambio_cds_em_formato_json](#)

16. Conclusão:

Esse trabalho de ciência de dados implementou técnicas fundamentais de extração, limpeza e transformação de dados, seguido de uma análise descritiva e exploratória, justamente para entender como a variação de fatores macroeconômicos (Ex: Taxa Selic, Câmbio e Risco Brasil - CDS) poderiam influenciar no retorno logarítimo de ações de empresas importantes como o Itaú, Petrobrás, Vale do Rio Doce.

Pode-se concluir que à medida que o Câmbio e o CDS aumentam, o retorno logarítimo do preço das ações dessas empresas diminui. Essa informação funciona como um termômetro, indicando ao investidor que os preços das ações tendem a diminuir, ajudando-o a tomar decisões quanto aos seus investimentos.