# Universidade Positivo

Francisco Bueno Ghizelini  
 Guilherme Candida

Relatório Geral de Atividades – Projeto "Preço Justo PR

Curitiba, Paraná

2025

Francisco Bueno Ghizelini  
 Guilherme Candida

Análise de Dados do Projeto *Preço Justo PR*

Relatório do trabalho A2 da disciplina de

Data science, apresentada pelo professor Rubem Matimoto Koide pela Universidade Positivo

Curitiba, Paraná

2025

# 

**Data:** 20 de outubro de 2025

**Objetivo do Projeto:** Realizar uma análise de regressão sobre dados de preços de combustíveis no estado do Paraná, com foco no comportamento do preço da gasolina ao longo do tempo. As atividades incluíram a implementação de modelos de regressão linear e não linear, a resolução de desafios técnicos e a criação de uma ferramenta de visualização interativa (dashboard).

#### **1. Análise de Regressão e Geração de Código**

A primeira etapa do projeto consistiu na criação de um script em Python para ser executado no ambiente Google Colab. O objetivo era analisar a tendência dos preços da gasolina com base nos dados fornecidos.

**Atividades Realizadas:**

* **Carregamento e Preparação dos Dados:** O script foi programado para carregar o conjunto de dados, converter a coluna de data para um formato temporal (datetime) e a coluna de valor de venda para um formato numérico. Para a análise de tendência, os preços diários foram agrupados e calculou-se a média de preço para cada dia.
* **Implementação dos Modelos de Regressão:**
  + **Regressão Linear Simples:** Um modelo de linha reta (y = ax + b) foi implementado para identificar a tendência geral dos preços (aumento ou queda constante).
  + **Regressão Polinomial (Parábola):** Um modelo de 2º grau (y = ax² + bx + c) foi criado para capturar tendências de aceleração ou desaceleração nos preços.
  + **Regressões Não Lineares Adicionais:** Foram implementados modelos mais complexos para verificar outros padrões de comportamento:
    - **Exponencial:** y = a \* e^(bx)
    - **Potência:** y = a \* x^b
    - **Logística:** y = L / (1 + e^(-k(x-x0)))
* **Análise de Desempenho:** Para cada modelo, o **Coeficiente de Determinação (R²)** foi calculado para avaliar a qualidade do ajuste aos dados.
* **Visualização:** Gráficos foram gerados para cada modelo, plotando os dados reais (pontos de dispersão) e a curva de regressão ajustada (linha).

#### **2. Diagnóstico e Resolução de Erros Técnicos**

Durante a fase de implementação, surgiram alguns desafios técnicos que exigiram depuração e correção do código.

**Problemas Encontrados e Soluções Aplicadas:**

1. **FileNotFoundError (Arquivo Não Encontrado):**
   * **Diagnóstico:** O código inicial não encontrava o arquivo de dados porque o caminho especificado não correspondia à localização do arquivo no ambiente do Google Colab.
   * **Solução:** O problema foi resolvido de forma definitiva ao se adotar um método de upload interativo (google.colab.files.upload()), que permite ao usuário selecionar o arquivo diretamente, eliminando a necessidade de especificar o caminho manualmente.
2. **UnicodeDecodeError (Erro de Codificação):**
   * **Diagnóstico:** Após corrigir o problema do caminho, um novo erro surgiu, indicando que o arquivo continha caracteres especiais (como "ç") que não eram reconhecidos pela codificação padrão (UTF-8).
   * **Solução:** O código foi ajustado para ler o arquivo especificando a codificação latin-1, que é compatível com arquivos gerados por softwares como o Excel em sistemas Windows.
3. **Error tokenizing data (Erro de Formato de Arquivo):**
   * **Diagnóstico:** O erro final de leitura ocorreu porque o código tentava usar a função pd.read\_csv(), destinada a arquivos de texto, para ler um arquivo nativo do Excel (.xlsx).
   * **Solução:** O código foi corrigido para utilizar a função pd.read\_excel(), que é a ferramenta apropriada para ler planilhas do Excel, resolvendo permanentemente o problema de carregamento de dados.

#### **3. Desenvolvimento do Dashboard Interativo**

Com os modelos de análise funcionando corretamente, a etapa final foi criar uma ferramenta de visualização que permitisse uma exploração dinâmica dos resultados.

**Funcionalidades Implementadas:**

* **Interface Interativa:** Utilizando a biblioteca ipywidgets, foi criado um dashboard diretamente no notebook do Google Colab.
* **Seleção de Modelos:** Um menu suspenso foi adicionado, permitindo ao usuário escolher qual dos modelos de regressão (Linear, Polinomial, Exponencial, etc.) deseja visualizar.
* **Atualização Dinâmica:** Ao selecionar um modelo no menu, o dashboard atualiza automaticamente o gráfico, exibindo a curva de regressão correspondente sobre os dados. Além disso, as informações-chave do modelo, como sua **equação matemática** e o **valor de R²**, são exibidas diretamente no gráfico para fácil interpretação e comparação.

**Conclusão Geral:**

As atividades foram concluídas com sucesso. O projeto evoluiu de um script de análise estática para um dashboard interativo e robusto, capaz de carregar os dados de forma confiável e apresentar os resultados de múltiplos modelos de regressão de maneira clara e dinâmica. A jornada incluiu a resolução de problemas comuns de manipulação de dados em ambientes de nuvem, resultando em uma ferramenta final funcional e informativa para a análise dos preços de combustíveis.