

## Estudo de Caso: Análise do Preço do GLP no Brasil

**Contexto:** Você foi contratado(a) como analista de dados júnior em uma consultoria especializada em mercado de combustíveis. Seu primeiro desafio é analisar o comportamento do preço do Gás Liquefeito de Petróleo (GLP) em diferentes regiões do Brasil nos últimos 6 meses. Para isso, você recebeu 6 arquivos CSV, cada um contendo dados de um mês específico. Sua tarefa é consolidar, tratar e extrair insights relevantes desses dados para apresentar à equipe.

**Objetivo:** Compreender a variação do preço do GLP, identificar momentos de oscilação e analisar o impacto regional, utilizando as habilidades de manipulação e análise de dados em Python.

**Dados:** Serão fornecidos 6 arquivos CSV, um para cada semestre totalizando 3 anos de informações. Cada arquivo CSV deve conter, no mínimo, as seguintes colunas (adapte se necessário):

- `DATA_COLETA`: Data da coleta do preço.
  - `PRODUTO`: Nome do produto (GLP).
  - `REGIÃO`: Região geográfica (Norte, Nordeste, Sudeste, Sul, Centro-Oeste).
  - `ESTADO`: Sigla do estado.
  - `MUNICÍPIO`: Nome do município.
  - `PREÇO_VENDA`: Preço de venda do GLP (R\$).
  - `UNIDADE_MEDIDA`: Unidade de medida (Litro, Kg, etc. - para GLP geralmente é Kg ou Litro).
- 

### Instruções e Tópicos da Avaliação:

Os alunos deverão entregar um script Python (ou um Jupyter Notebook) contendo todo o código e as análises, acompanhado de um breve relatório/resumo dos principais insights encontrados.

#### 1. Concatenar os Dados (20 pontos)

- Carregue os 6 arquivos CSV fornecidos para DataFrames separados.
- Concatene todos os DataFrames em um único DataFrame chamado `df_glp_consolidado`.
- Verifique as dimensões do DataFrame consolidado (número de linhas e colunas).
- Mostre as 5 primeiras linhas do DataFrame consolidado.

#### 2. Tratar os Dados (20 pontos)

- Verifique se há valores ausentes (NaN) no DataFrame. Se houver, discuta a estratégia para tratá-los (ex: preencher com a média, remover as linhas, etc.) e implemente o tratamento escolhido.
- Verifique os tipos de dados de cada coluna. Converta as colunas de data (se necessário) para o tipo `datetime`.

- Converta a coluna `PREÇO_VENDA` para o tipo numérico (float), tratando possíveis erros de conversão (ex: vírgulas como separador decimal).
- Identifique e trate possíveis outliers na coluna `PREÇO_VENDA` (opcional, mas um diferencial para alunos mais avançados).

### 3. Separar por Ano

- Crie uma nova coluna chamada `ANO` extraindo o ano da coluna `DATA_COLETA`.
- Considerando que a base é de 6 meses, os dados podem abranger dois anos distintos (ex: novembro e dezembro de 2024 e janeiro a abril de 2025). Os alunos devem ser capazes de separar e analisar cada ano, mesmo que um deles tenha poucos meses.

### 4. Cálculos Estatísticos

- Calcule as seguintes estatísticas descritivas para a coluna `PREÇO_VENDA` no DataFrame consolidado:
  - o Média
  - o Mediana
  - o Desvio Padrão
  - o Mínimo
  - o Máximo
- Calcule as mesmas estatísticas descritivas, mas agrupadas por `REGIÃO`.
- Apresente os resultados de forma clara.

### 5. Variação do Preço do GLP

- Calcule a variação percentual do preço médio do GLP ao longo dos meses.
- Apresente um gráfico de linha mostrando a evolução do preço médio mensal do GLP.

### 6. Em que Momento Ele Oscilou

- Identifique os meses ou períodos em que o preço médio do GLP apresentou as maiores variações (positivas e negativas).
- Os alunos podem usar diferentes abordagens para identificar a oscilação, como a diferença em relação ao mês anterior, ou a amplitude dos preços em um determinado período.
- Justifique a metodologia utilizada.

### 7. Destacar os Dados por Região

- Apresente as estatísticas descritivas do preço do GLP para cada região (mínimo, máximo, média, desvio padrão).
- Crie um boxplot ou um gráfico de barras comparando o preço médio do GLP entre as diferentes regiões.
- Discuta as diferenças observadas no preço do GLP entre as regiões.

### 8. Momentos em que o Preço foi Afetado

- Com base nas análises anteriores (variação e oscilação), os alunos deverão inferir e descrever os momentos em que o preço do GLP parece ter sido mais afetado.
  - Incentive-os a pensar em possíveis fatores externos que poderiam influenciar o preço (sem a necessidade de dados externos, apenas com base nas oscilações observadas).
  - Exemplos de "momentos afetados" podem ser picos, quedas bruscas ou períodos de alta instabilidade.
- 

### Recursos e Dicas para os Alunos:

- Utilize a biblioteca `pandas` para manipulação de dados.
- Utilize `matplotlib` ou `seaborn` para visualização de dados.
- Documente seu código com comentários.
- Mantenha o código organizado e legível.
- Crie um relatório conciso no final do notebook ou em um arquivo separado, resumindo as principais descobertas e insights.

### Critérios de Avaliação:

- **Correto e Funcional:** O código executa sem erros e produz os resultados esperados.
- **Organização e Legibilidade:** O código é bem estruturado, com comentários e variáveis claras.
- **Precisão das Análises:** As análises estatísticas e as conclusões são precisas e bem fundamentadas.
- **Visualização de Dados:** Os gráficos são apropriados e transmitem a informação de forma clara.
- **Interpretação e Insights:** A capacidade de extrair e comunicar insights relevantes dos dados.
- **Tratamento de Erros/Exceções:** Abordagem para dados ausentes, tipos incorretos, etc.