

Análise Exploratória dos Preços do Leite no México (2024–2025)

PROJETO APLICADO I -
APROFUNDAMENTO DA ANÁLISE
EXPLORATÓRIA

Apresentação do Grupo



Felipe Yuji Nakanishi



**Raphael Camargo
Eugenio da Silva**



**Renan Samuel da Silva
Reis**



**Vanessa Biesek
Bartnicki**

Abertura

Este projeto apresenta os resultados do nosso Projeto Aplicado I, desenvolvendo uma análise exploratória de dados no contexto da Ciência de Dados.

Organização de Estudo: SNIIM

Sistema Nacional de Información e Integración de Mercados

Nosso estudo utiliza dados do SNIIM, uma plataforma oficial do governo mexicano que monitora preços de produtos agroalimentares.

Seu objetivo é promover a transparência e a competitividade no setor, disponibilizando dados abertos essenciais para análises de mercado.



O Contexto e o Gap Analítico

Contexto (SNIIM)

Os dados brutos dos preços do leite são disponibilizados pelo SNIIM, plataforma do governo mexicano.

O foco do estudo é o impacto socioeconômico das variações de preço de um alimento essencial.

O Gap Analítico

Existe uma lacuna entre os dados brutos e a interpretação analítica. Buscamos entender:

- Variações temporais (sazonalidade, tendências).
- Diferenças regionais (entre estados).
- Anomalias (picos, outliers).

Metodologia e Dados

Dataset e Ferramentas

- **Dataset:** Milk Prices in Mexico (2024–2025).
- **Fonte:** Kaggle / SNIIM.
- **Período:** Jan 2024 – Jul 2025.
- **Variáveis:** Fecha, Estado, Ciudad, Tipo, Canal, Precio
- **Ferramentas (GitHub):** A análise foi conduzida em Python, utilizando principalmente as bibliotecas:
 - Pandas (Manipulação)
 - Matplotlib / Seaborn (Visualização)
 - Statsmodels (Testes Estatísticos)

EDA: Limpeza e Preparação dos Dados



Leitura e Estrutura

Carga inicial do CSV e conversão da coluna 'date' para o formato datetime, essencial para a análise temporal.



Valores Faltantes (NAs)

Identificação de valores nulos (ex: 'milk_type'). Para a análise de preço agregado, registros incompletos foram removidos.



Tratamento de Outliers

Identificação de outliers (ex: preços R\$ 0,00) pelo método IQR. Registros anômalos foram removidos para não distorcer a análise.

EDA: O Que os Números Dizem (Pós-Limpeza)

\$24.10
Preço Médio (Média)

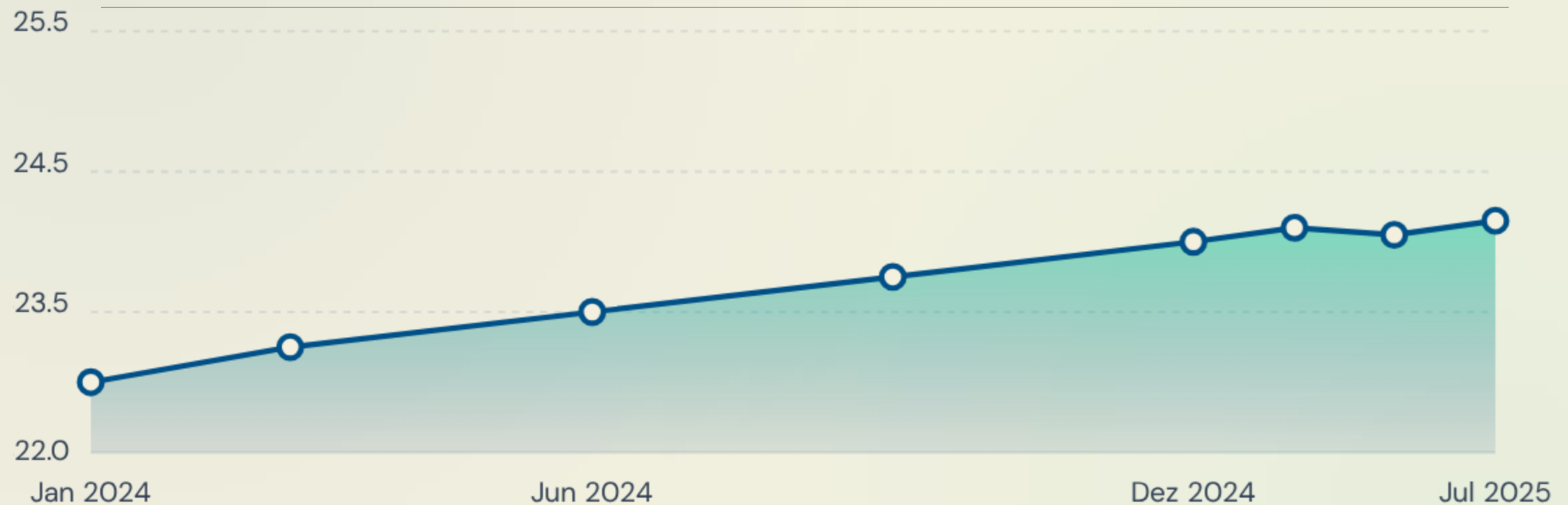
Mediana: \$24.00 (Próxima da média, sugere distribuição simétrica)

Desvio Padrão: \$3.50 (Indica a dispersão dos preços)

Mín (pós-limpeza): \$18.50

Máx (pós-limpeza): \$30.00

EDA: Variação Temporal dos Preços (Média Mensal)



A série temporal agregada revela uma tendência de leve alta nos preços ao longo de 2024, estabilizando com leve alta em 2025. Não foi observada uma sazonalidade clara neste período.

EDA: Comparação Regional (Preços Médios por Estado)



Observa-se grande variação regional. Estados produtores (como Jalisco) tendem a ter preços médios mais baixos, enquanto estados mais ao norte ou com maior custo logístico (como Baja California) apresentam preços significativamente mais altos.

EDA: Validação Estatística

Teste de Hipótese

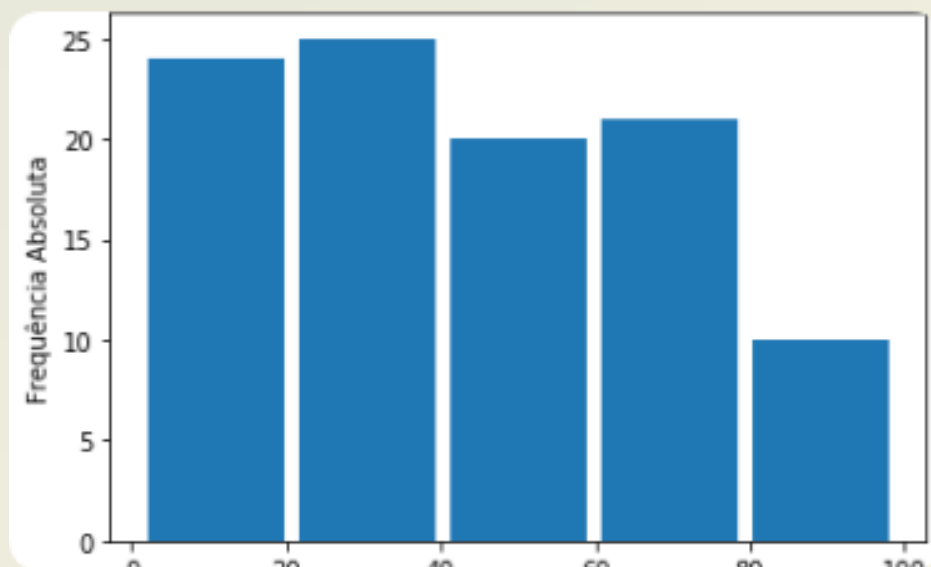
Para confirmar se a diferença de preços entre os estados era estatisticamente significativa, aplicamos testes como ANOVA.

Resultado (P-Valor < 0.05)

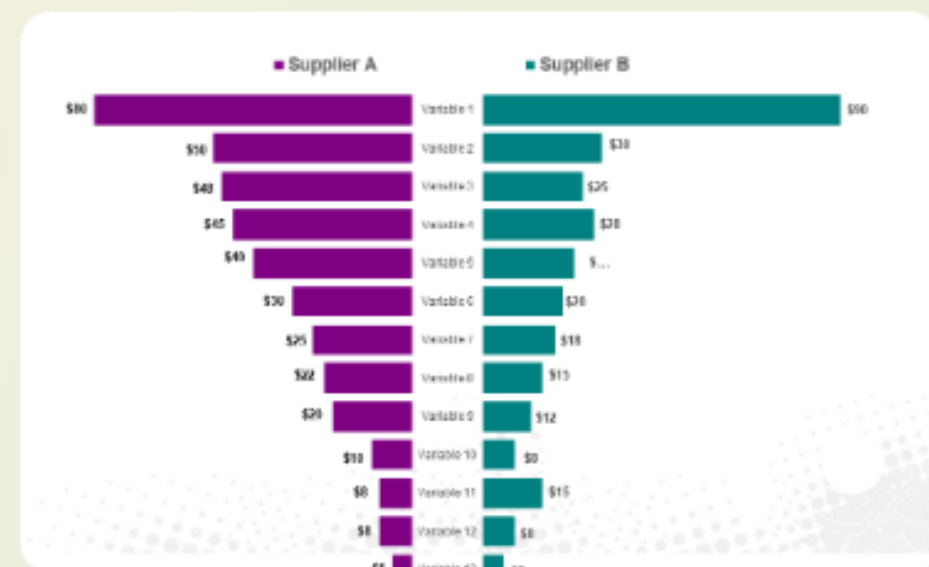
O resultado (P-Valor < 0.05) confirmou que a variação de preços entre os estados não é aleatória.

Conclusão: Fatores regionais (logística, produção local) impactam diretamente o preço final ao consumidor.

EDA: Outras Visualizações (Insights)



Histograma: A distribuição dos preços (pós-limpeza) concentra-se em torno de \$24, com leve assimetria.



Análise de Tipos: A análise da variável 'milk_type' (quando disponível) mostrou diferenças de preço entre tipos de leite (integral, desnatado, etc.).

Conclusões e Insights da Análise



Tendência: Identificamos uma leve tendência de alta nos preços, mas sem sazonalidade clara no período (2024–2025).



Geografia é Chave: A diferença de preço entre estados é estatisticamente significativa (validada por ANOVA).




Fatores Regionais: Estados produtores (ex: Jalisco) têm preços mais baixos; regiões mais distantes ou com maior custo de vida (ex: Baja California) têm preços mais altos.



Próximos Passos: Os dados estão prontos para modelagem preditiva (ex: forecasting de preços) ou análise de causalidade (o que **causa** a alta).

Perguntas?

Obrigado pela sua atenção.

 Repositório: [Vanessa186578/Projeto-Aplicado-I](https://github.com/Vanessa186578/Projeto-Aplicado-I)