

# **PROJETO APLICADO I**

## **ANÁLISE EXPLORATÓRIA DOS PREÇOS DO LEITE NO MÉXICO (2024–2025)**

**FELIPE YUJI NAKANISHI**

**RAPHAEL CAMARGO EUGENIO DA SILVA**

**RENAN SAMUEL DA SILVA REIS**

**VANESSA BIESEK BARTNICKI**

**Guarulhos – São Paulo**

**2025**

## Sumário

1. Introdução .....	3
2. Contexto do Estudo .....	3
3. Apresentação da Empresa: SNIIM.....	4
4. Problema do Estudo.....	4
5. Estrutura dos Metadados .....	5
6. Referência de Aquisição do Dataset .....	6
7. Descrição da Origem .....	6
8. Descrição do Dataset .....	6
9. Análise Exploratória de Dados .....	8
10. Cronograma.....	19
11. Repositório GitHub .....	20
Referências .....	20

## 1. Introdução

O presente trabalho tem como objetivo a realização de uma análise exploratória de dados (EDA) a partir de informações sobre os preços do leite no México, referentes ao período de 2024 a 2025.

A escolha do tema justifica-se pela relevância do leite como um bem essencial para o consumo humano e pela importância de compreender variações de preços que podem impactar o custo de vida da população.

Neste estudo, pretende-se responder a duas questões principais:

1. Como os preços do leite variaram ao longo do tempo no período analisado?
2. Existem diferenças significativas nos preços entre os estados mexicanos?

## 2. Contexto do Estudo

O presente estudo utiliza como fonte primária de dados o Sistema Nacional de Información e Integración de Mercados (SNIIM), um serviço público da Secretaria de Economia do México. O SNIIM se dedica a coletar e divulgar informações comerciais estratégicas sobre produtos agropecuários, visando promover a transparência e a competitividade no setor. Sua base de dados, que inclui informações diárias sobre preços, volumes e fluxos de produtos, é uma ferramenta essencial para análises de mercado.

Dentre os indicadores disponibilizados, o preço do leite se destaca como um objeto de estudo de grande relevância socioeconômica. Por ser um item fundamental na cesta de consumo da população, suas variações impactam diretamente o custo de vida e refletem dinâmicas complexas de mercado, incluindo fatores sazonais, logísticos e regionais.

Nesse contexto, este trabalho tem como objetivo analisar as variações de preço do leite nos diferentes estados do México, com base nos dados históricos fornecidos pelo SNIIM. A análise buscará identificar padrões, tendências e disparidades regionais, a fim de contribuir para um melhor entendimento da estrutura do mercado de laticínios no país e fornecer subsídios para discussões sobre políticas públicas voltadas à segurança alimentar e estabilidade econômica.

### 3. Apresentação da Empresa: SNIIM

O Sistema Nacional de Información de Mercados (SNIIM) é uma plataforma oficial vinculada ao Ministério da Economia do México. Seu principal objetivo é fornecer informações detalhadas, atualizadas e confiáveis sobre os preços de diversos produtos agrícolas e alimentares comercializados no país. Por meio dessa iniciativa, o governo busca promover a transparência no mercado, facilitar a tomada de decisões e fortalecer as relações comerciais no setor agroalimentar. O SNIIM possui cobertura nacional para muitos produtos perecíveis, abrangendo praticamente todo o mercado mexicano desses produtos no que se refere à oferta de informações. Contudo, dados sobre número de usuários e receita não estão disponíveis para análise.

Como órgão público, o SNIIM não possui declarações formais de missão, visão ou valores. Seus objetivos são definidos pela função pública de serviço à sociedade, com foco no setor agroalimentar. Informações sobre o número de colaboradores também não estão disponíveis.

Na área de ciência de dados, o SNIIM disponibiliza diversos conjuntos de dados públicos, incluindo preços por produto e região, anuários estatísticos, análises comparativas e diretórios, que facilitam a realização de análises avançadas, a criação de dashboards e a elaboração de previsões. Além disso, o módulo “enlaces comerciales” conecta produtores e compradores, principalmente do setor agroalimentar (agrícola, pecuário e pesqueiro), permitindo que usuários publiquem online suas ofertas e demandas, detalhando as características dos produtos.

### 4. Problema do Estudo

Embora o Sistema Nacional de Información e Integración de Mercados (SNIIM) disponibilize dados detalhados sobre os preços do leite no México, existem lacunas importantes na análise dessas informações. Neste estudo, pretende-se responder a duas questões principais: como os preços do leite variaram ao longo do tempo no período analisado (2024-2025); e se existem diferenças significativas nos preços entre os estados mexicanos.

Por exemplo, para a primeira questão, é importante verificar se há variações sazonais nos preços, como aumentos durante períodos de alta demanda (festividades ou férias) ou quedas durante períodos de maior produção. Além disso, mudanças abruptas nos preços podem indicar eventos externos, como problemas logísticos ou políticas públicas que impactam o setor.

Para a segunda questão, espera-se identificar se estados com maior produção leiteira apresentam preços mais baixos devido à maior oferta, enquanto estados mais distantes dos centros produtores podem registrar preços mais elevados por conta dos custos de transporte. Também pode haver diferenças relacionadas à renda média regional e ao poder de compra dos consumidores.

O desafio está em organizar e interpretar dados diários que podem apresentar ruídos, inconsistências e variações regionais complexas. A análise exploratória de dados buscará identificar padrões temporais e espaciais, decompor o problema em aspectos como sazonalidade, oferta e demanda regionais, e construir hipóteses para futuras modelagens preditivas. O objetivo é transformar dados brutos em insights úteis, auxiliando produtores, consumidores e formuladores de políticas públicas na tomada de decisões mais informadas.

## 5. Estrutura dos Metadados

O conjunto de dados utilizado neste estudo é intitulado *Milk Prices in Mexico (2024–2025)*, publicado por José Luis Martínez na plataforma Kaggle em 2024. Os dados foram originalmente extraídos do Sistema Nacional de Información e Integración de Mercados (SNIIM), órgão vinculado à Secretaría de Economía do México, e abrangem o período de janeiro de 2024 a julho de 2025.

O arquivo está disponível em formato CSV e contém informações públicas, não sensíveis, sobre preços do leite por estado, incluindo variáveis como data, localidade, tipo de leite e valor em pesos mexicanos. O dataset é de acesso aberto, sem restrições legais explícitas de uso, sendo apropriado para fins acadêmicos desde que seja devidamente referenciado.

Os dados são de natureza agregada e não contêm informações pessoais, o que os isenta de implicações diretas com legislações como a LGPD. As principais variáveis incluem tipos categóricos (estado, tipo de leite), temporais (data) e numéricos (preço), permitindo análises exploratórias de padrões sazonais e regionais. Parte inferior do formulário

## 6. Referência de Aquisição do Dataset

O dataset foi obtido a partir da plataforma Kaggle, disponível no link:

MARTÍNEZ OL, José Luis. Milk Prices in Mexico (2024–2025). Kaggle, 2024. Disponível em: <https://www.kaggle.com/datasets/jlmartinezol/milk-prices-in-mexico-2024-2025>. Acesso em: 11 set. 2025.

Limitações de uso: O dataset é disponibilizado para fins de pesquisa e aprendizado, podendo não refletir integralmente dados oficiais de mercado.

Período da coleta: janeiro de 2024 a julho de 2025.

## 7. Descrição da Origem

O dataset foi organizado e disponibilizado por José Luis Martínez, membro da comunidade Kaggle, com base em dados fornecidos pela Sistema Nacional de Información e Integración de Mercados (SNIIM).

Os dados foram estruturados de modo a permitir análises comparativas e temporais. A plataforma Kaggle é reconhecida mundialmente como repositório e comunidade voltada à ciência de dados, onde pesquisadores e entusiastas compartilham informações para fins de estudo e experimentação.

## 8. Descrição do Dataset

### 8.1 Estrutura:

As variáveis principais incluem:

- Data de referência (mês/ano);
- Estado mexicano;
- Preço médio do leite (em pesos mexicanos).

#### 8.2 Propósito:

O dataset foi elaborado para permitir análises exploratórias que possibilitem compreender:

- Tendências temporais de variação dos preços;
- Comparações regionais entre estados;
- Reflexões sobre custo de vida e segurança alimentar no México.

#### 8.3 Fenômeno Registrado:

A base registra a variação dos preços de um alimento essencial no mercado de consumo. Esse registro permite estudar possíveis fatores econômicos, sazonais e geográficos que influenciam seu comportamento.

### 9. Análise Exploratória de Dados

A análise exploratória foi iniciada com a inspeção da estrutura do dataset após a limpeza, confirmando o período coberto (jan/2024 a jul/2025) e a presença das variáveis date, state, price e milk\_type. Em seguida, foram calculadas estatísticas descritivas para cada coluna numérica e categórica, incluindo contagem de valores válidos, número de NAs, mínimo, máximo, média, mediana, desvio padrão e amplitude. Essa etapa permitiu identificar a distribuição dos preços e avaliar a consistência dos registros ao longo do tempo.

A análise de valores faltantes revelou a existência de lacunas pontuais em determinadas datas ou estados, as quais foram tratadas mediante remoção de registros insuficientes ou imputação simples quando apropriado. Além disso, outliers foram identificados pelo método do IQR ( $1.5 \times \text{IQR}$ ), permitindo a visualização de possíveis oscilações abruptas causadas por fatores sazonais, logísticos ou erros de registro. Foram comparadas tanto a distribuição geral quanto as distribuições específicas por estado, utilizando histogramas e boxplots.

A análise temporal, organizada a partir da agregação mensal dos preços, possibilitou a identificação de tendências e sazonalidade. A série mostrou padrões de aumento e queda que podem estar associados a ciclos produtivos ou ao comportamento da demanda. Complementando a análise temporal, a comparação entre estados revelou diferenças significativas nos preços médios, destacadas por testes estatísticos como ANOVA e Kruskal–Wallis, demonstrando que a variação entre regiões não é aleatória, mas associada a características econômicas e geográficas específicas.

Por fim, foram gerados gráficos que articulam visualmente os comportamentos observados, como a evolução dos preços ao longo dos meses, a distribuição dos preços entre os estados e a densidade dos valores registrados. Esses elementos visuais, aliados às estatísticas descritivas, permitiram construir uma visão abrangente sobre o mercado de leite no México, fornecendo subsídios importantes para interpretações, discussões e tomadas de decisão baseadas em dados.

```
import os
import warnings
warnings.filterwarnings('ignore')
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split

sns.set(style='whitegrid')

pd.options.display.max_columns = 200
pd.options.display.width = 120

def read_data(path):
    return pd.read_csv(path)
```

```

def basic_report(df):
    print('Dimensão:', df.shape)
    display(df.head())
    print('\nTipos e não-nulos por coluna:')
    display(pd.DataFrame(df.dtypes, columns=['dtype']))
    nulos = df.isnull().sum().to_frame(name='Contagem_nulos')
    nulos['Percentual_nulos (%)'] = (df.isnull().mean() * 100).round(2)
    print('\nNulos por coluna:')
    display(nulos)

def resumo_preco_por_categoria(df, var_cat, var_num='Precio'):

    resumo = (
        df.groupby(var_cat)[var_num]
        .agg(['count', 'mean', 'median', 'std', 'min', 'max'])
        .sort_values(by='mean', ascending=False)
        .round(2)
    )
    return resumo

def plot_media_por_categoria(df, var_cat, var_num='Precio'):
    resumo = df.groupby(var_cat)[var_num].mean().sort_values()
    plt.figure(figsize=(8,5))
    sns.barplot(x=resumo.values, y=resumo.index, palette='viridis')
    plt.title(f'{var_num} médio por {var_cat}')
    plt.xlabel(f'Média de {var_num}')
    plt.ylabel(var_cat)
    plt.show()

def boxplot_preco_por_categoria(df, var_cat, var_num='Precio'):

    plt.figure(figsize=(10,5))
    sns.boxplot(data=df, x=var_cat, y=var_num, palette='Set3')
    plt.title(f'Distribuição de {var_num} por {var_cat}')
    plt.xticks(rotation=45, ha='right')
    plt.show()

def resumo_estatistico(df):
    num_df = df.select_dtypes(include=[np.number])

    resumo = pd.DataFrame({
        'Contagem': num_df.count(),
        'Média': num_df.mean(),
        'Mediana': num_df.median(),
        'Moda': [num_df[col].mode().iloc[0] if not num_df[col].mode().empty else np.nan for col in num_df.columns],
        'Mínimo': num_df.min(),
        'Máximo': num_df.max(),
        'Amplitude': num_df.max() - num_df.min(),
        'Variância': num_df.var(),
        'Desvio_Padrão': num_df.std(),
        'Coef_Variação (%)': (num_df.std() / num_df.mean()) * 100,
        'Assimetria': num_df.skew(),
        'Curtose': num_df.kurt()
    }).round(3)

    return resumo

```

```

def medidas_posicao(df):
    num_df = df.select_dtypes(include=[np.number])

    q = num_df.quantile([0, 0.25, 0.5, 0.75, 1]).T
    q.columns = ['Q0 (mín)', 'Q1 (25%)', 'Q2 (50%)', 'Q3 (75%)', 'Q4 (máx)']
    q['IQR (Q3-Q1)'] = q['Q3 (75%)'] - q['Q1 (25%)']
    return q.round(3)

def plot_distribuicoes(df):
    num_cols = df.select_dtypes(include=[np.number]).columns

    for col in num_cols:
        fig, ax = plt.subplots(1, 2, figsize=(10, 4))

        # Histograma
        sns.histplot(df[col].dropna(), kde=True, ax=ax[0], color='teal')
        ax[0].set_title(f'Distribuição de {col}')

        # Boxplot
        sns.boxplot(x=df[col], ax=ax[1], color='orange')
        ax[1].set_title(f'Boxplot de {col}')

        plt.tight_layout()
        plt.show()

```

```

df = read_data(DATA_PATH)
print('Dados carregados com sucesso')
basic_report(df)

```

... Dados carregados com sucesso  
Dimensão: (52056, 6)

	Fecha	Estado	Ciudad	Tipo	Canal	Precio
0	2025-08-15	Aguascalientes	Aguascalientes	Pasteurizada	Tiendas	27.4
1	2025-08-15	Baja California	Mexicali	Pasteurizada	Tiendas	29.3
2	2025-08-15	Baja California	Tijuana	Pasteurizada	Tiendas	NaN
3	2025-08-15	Baja California Sur	La Paz	Pasteurizada	Tiendas	26.0
4	2025-08-15	Campeche	Campeche	Pasteurizada	Tiendas	NaN

...

Tipos e não-nulos por coluna:

	dtype
Fecha	object
Estado	object
Ciudad	object
Tipo	object
Canal	object
Precio	float64

Nulos por coluna:

	Contagem_nulos	Percentual_nulos (%)
Fecha	0	0.00
Estado	0	0.00
Ciudad	0	0.00
Tipo	0	0.00
Canal	0	0.00
Precio	6511	12.51

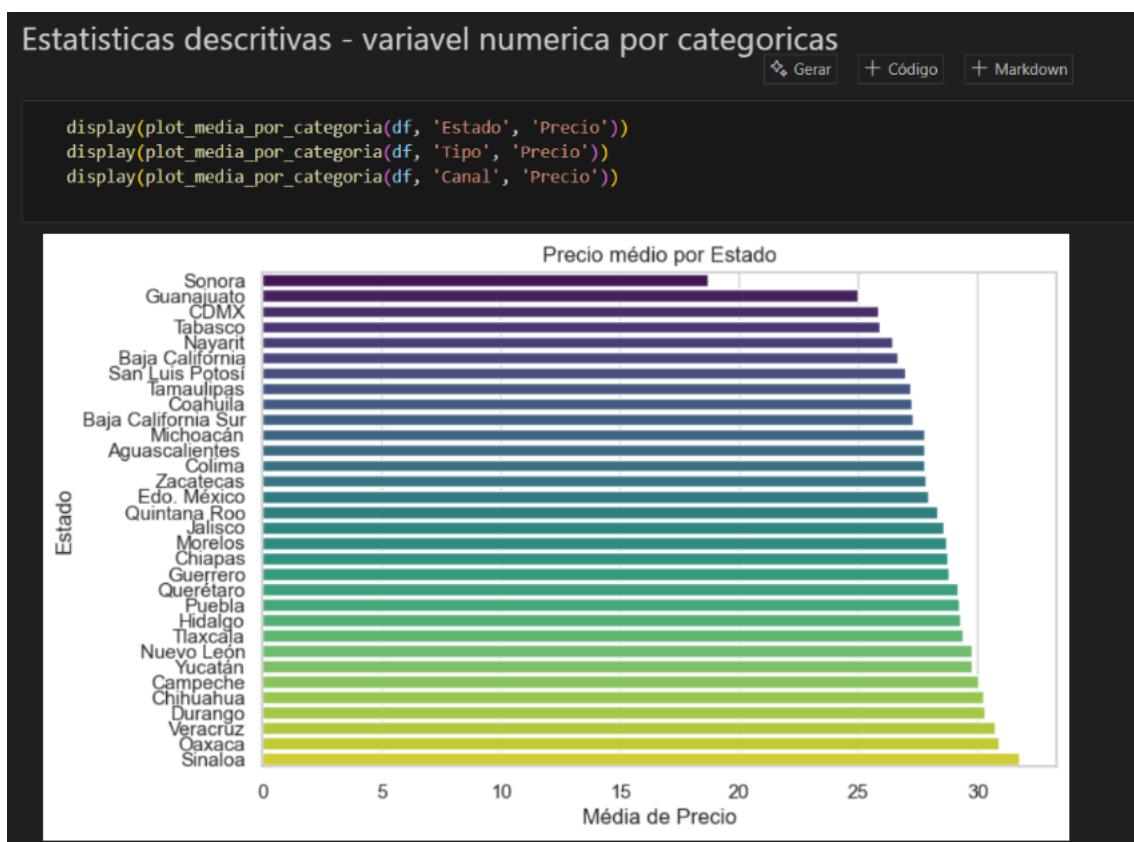
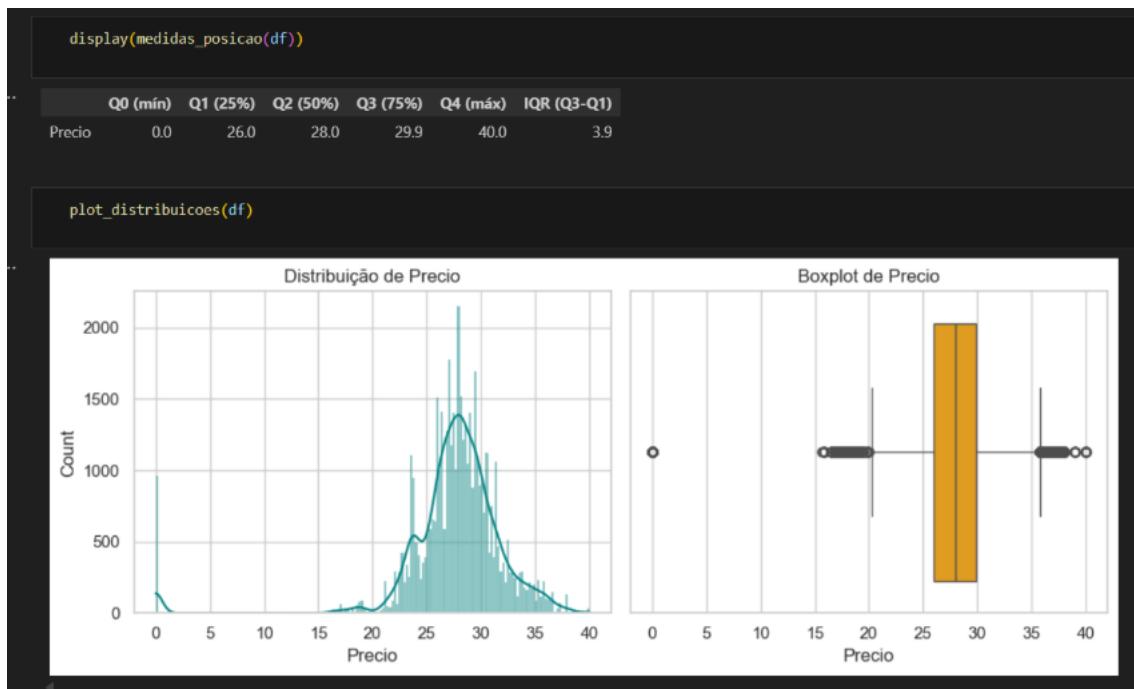
Estatísticas descritivas - variável numérica

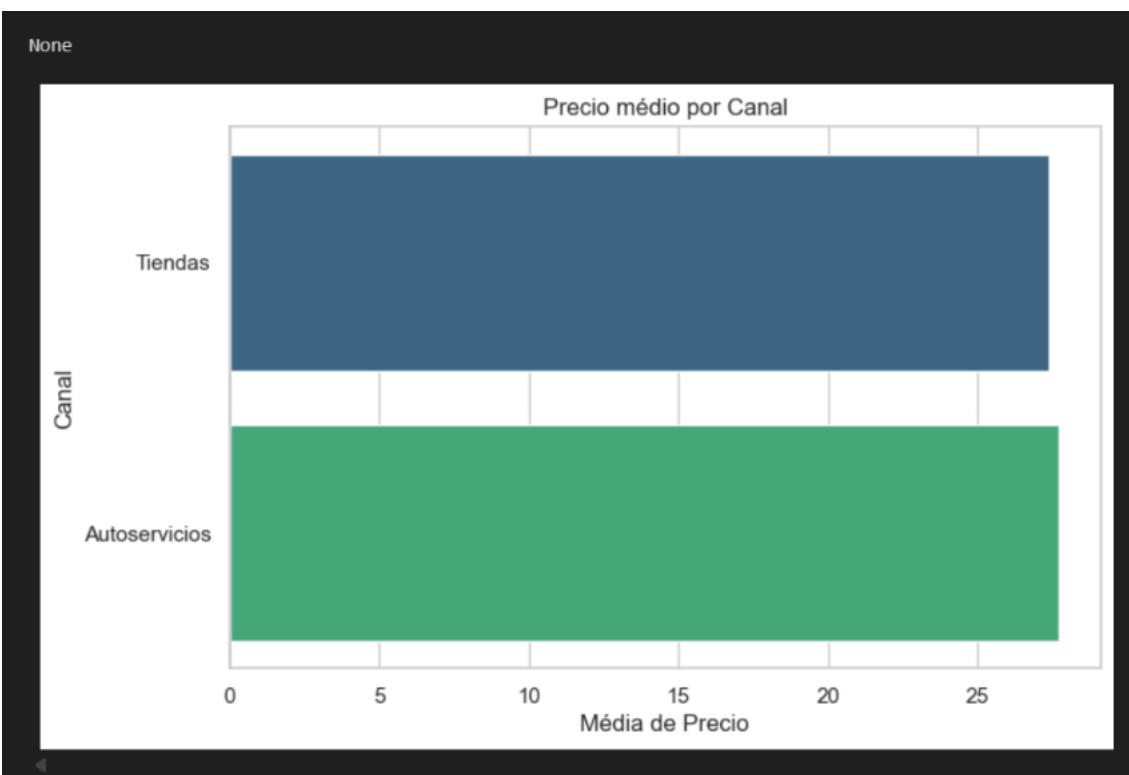
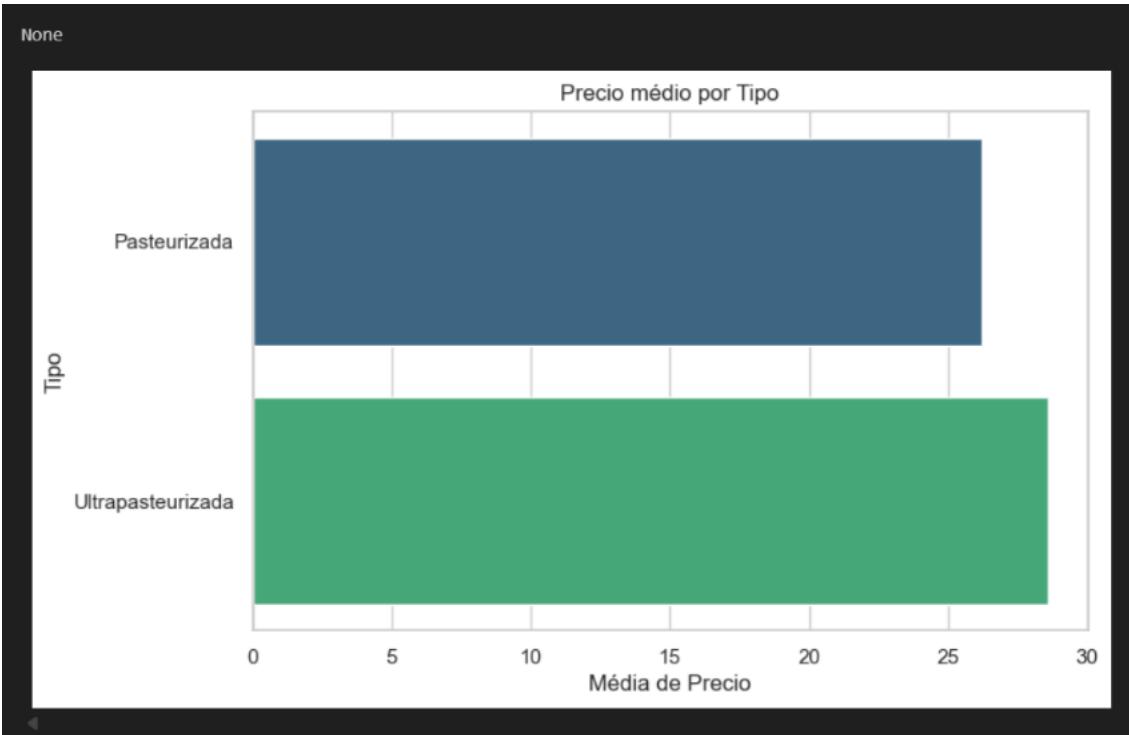
```
resumo_estatistico(df)
```

	Contagem	Média	Mediana	Moda	Mínimo	Máximo	Amplitude	Variância	Desvio_Padrão	Coef_Variação (%)	Assimetria	Curtose
Precio	45545	27.542	28.0	28.0	0.0	40.0	40.0	26.891	5.186	18.828	-2.997	14.422

```
display(medidas_posicao(df))
```

	Q0 (mín)	Q1 (25%)	Q2 (50%)	Q3 (75%)	Q4 (máx)	IQR (Q3-Q1)
Precio	0.0	26.0	28.0	29.9	40.0	3.9





```

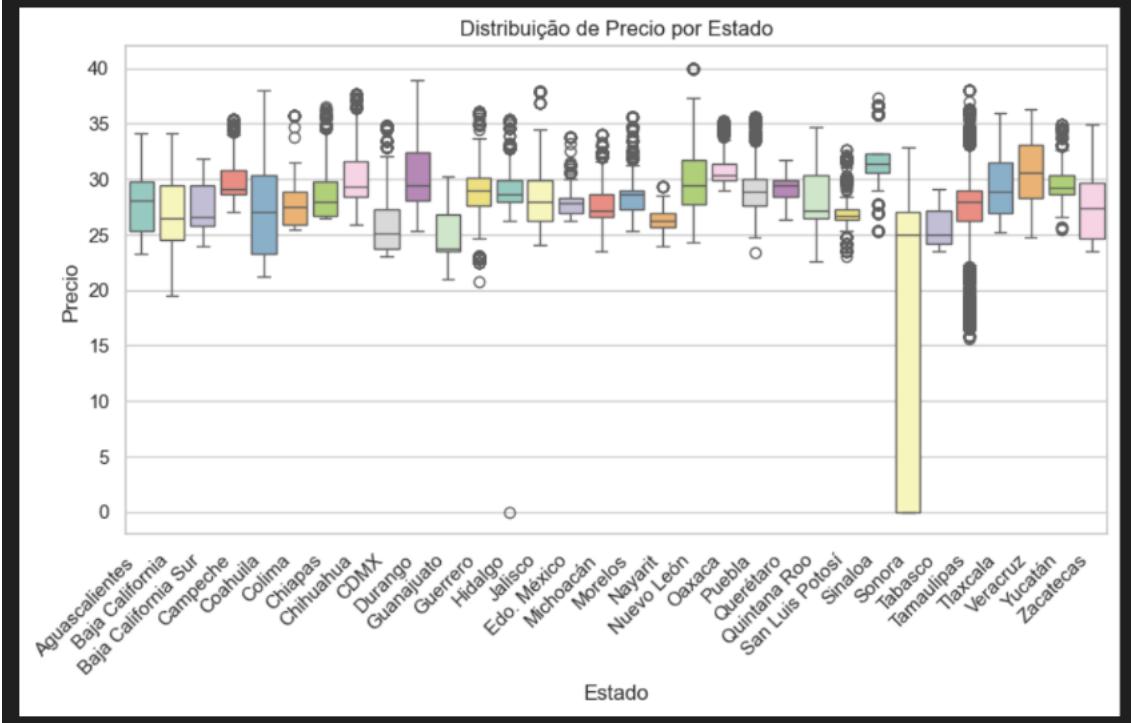
display(resumo_preco_por_categoria(df, 'Estado'))
display(resumo_preco_por_categoria(df, 'Tipo'))
display(resumo_preco_por_categoria(df, 'Canal'))

```

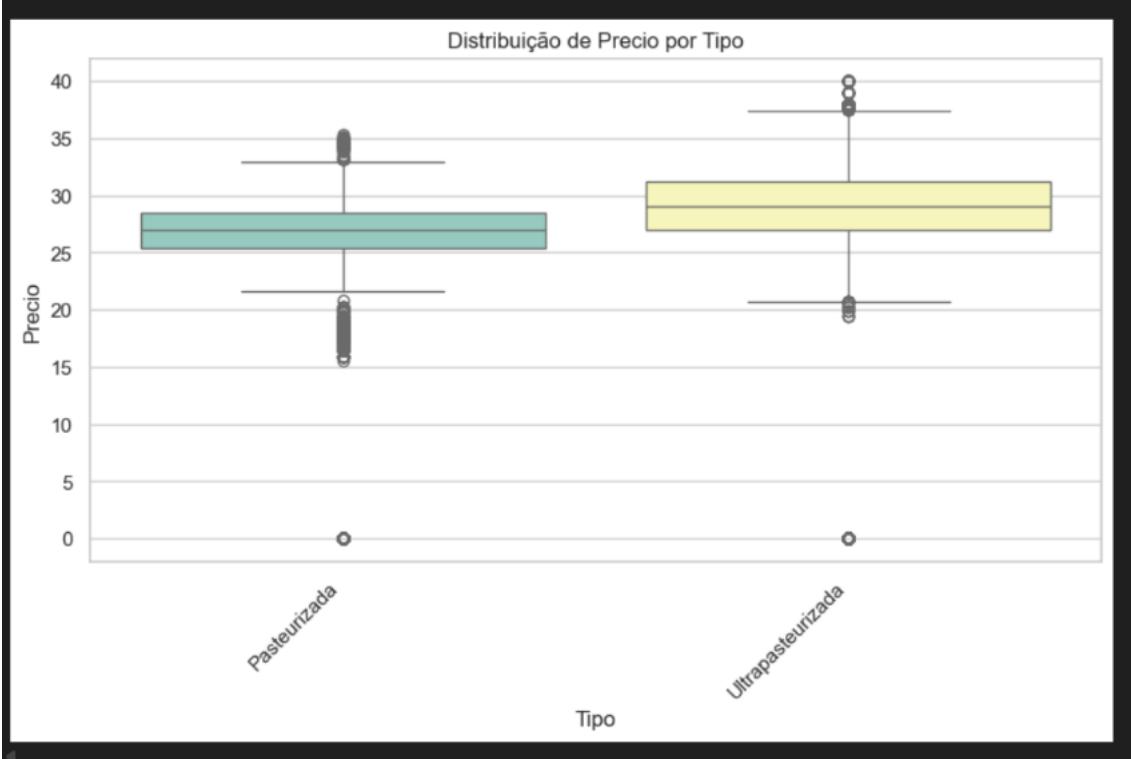
	count	mean	median	std	min	max
Estado						
Sinaloa	964	31.73	31.39	2.80	25.30	37.33
Oaxaca	964	30.91	30.33	1.51	28.97	35.32
Veracruz	1446	30.70	30.63	2.95	24.75	36.35
Durango	1687	30.28	29.45	2.85	25.30	38.95
Chihuahua	1687	30.23	29.29	2.78	25.90	37.72
Campeche	482	30.05	29.07	2.14	27.00	35.40
Yucatán	482	29.77	29.25	1.97	25.45	34.97
Nuevo León	1928	29.75	29.48	3.28	24.33	40.00
Tlaxcala	964	29.39	28.88	3.02	25.25	36.00
Hidalgo	964	29.27	28.67	2.32	0.00	35.42
Puebla	1928	29.23	28.90	2.34	23.35	35.58
Querétaro	964	29.19	29.40	0.97	26.38	31.72
Guerrero	1928	28.79	29.00	2.40	20.80	36.05
Chiapas	964	28.72	28.00	2.39	26.50	36.57
Morelos	964	28.66	28.67	2.15	25.33	35.58
Jalisco	1928	28.57	28.00	2.55	24.10	37.90
Quintana Roo	964	28.32	27.15	2.91	22.57	34.67
Edo. México	964	27.95	27.90	1.35	26.28	33.83

		count	mean	median	std	min	max
<b>Tipo</b>							
Ultrapasteurizada	25303	28.59	29.03	5.23	0.0	40.0	
Pasteurizada	20242	26.23	27.00	4.82	0.0	35.3	
<b>Canal</b>							
Autoservicios	23134	27.72	28.10	5.21	0.0	37.9	
Tiendas	22411	27.36	27.98	5.16	0.0	40.0	

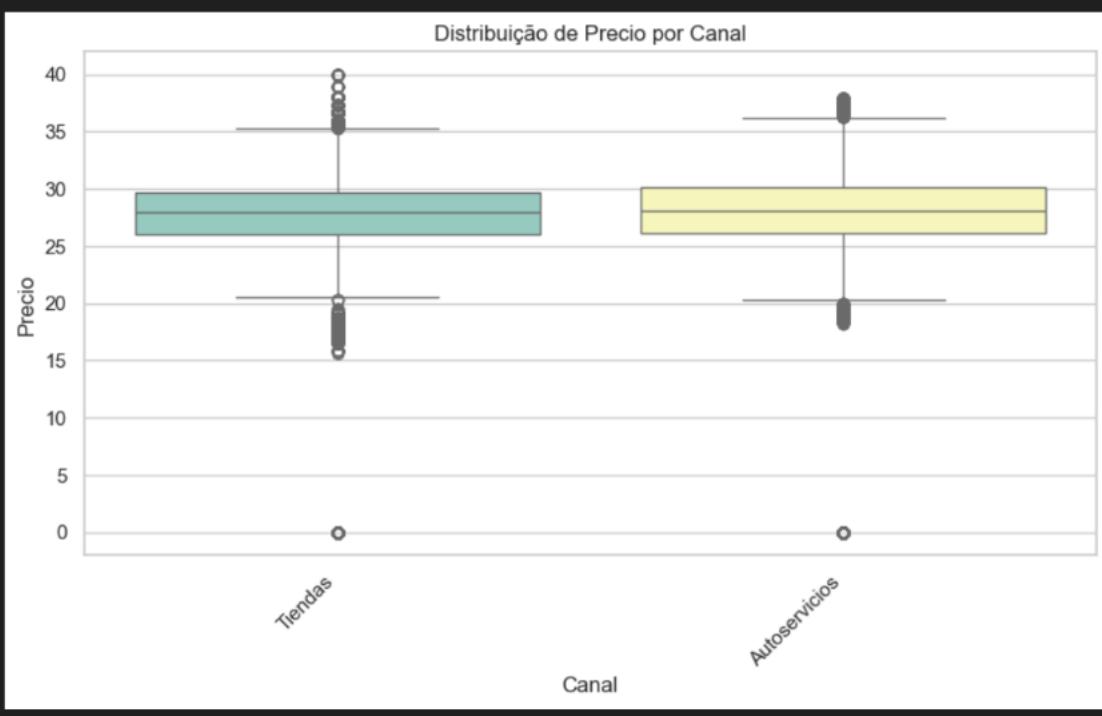
- `display(boxplot_precio_por_categoria(df, 'Estado', 'Precio'))`
- `display(boxplot_precio_por_categoria(df, 'Tipo', 'Precio'))`
- `display(boxplot_precio_por_categoria(df, 'Canal', 'Precio'))`



None



None



```
import matplotlib.dates as mdates

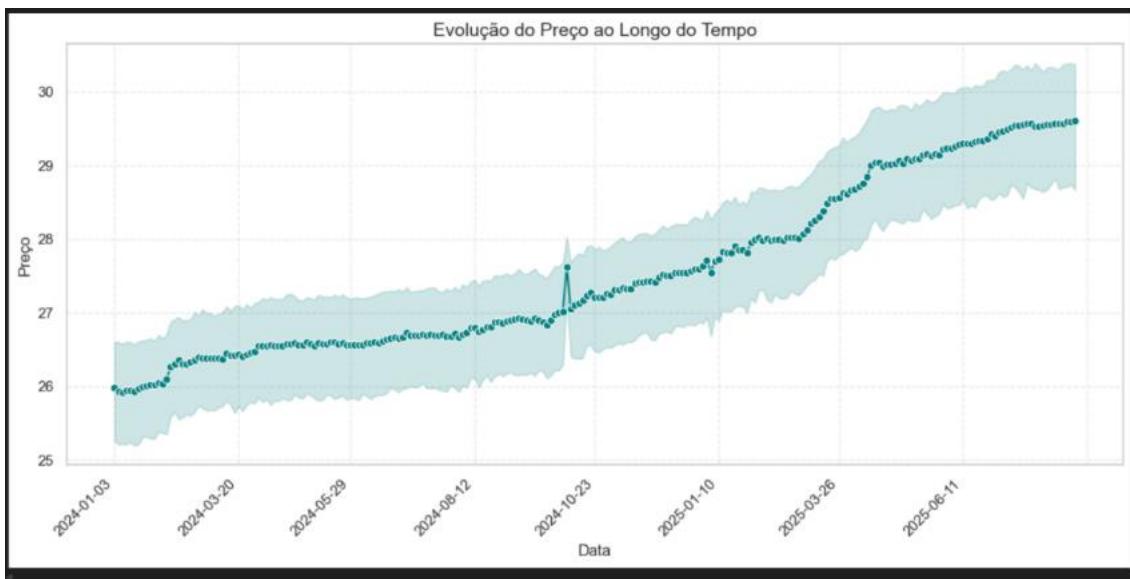
df = df.sort_values('Fecha')

plt.figure(figsize=(12,6))

sns.lineplot(data=df, x='Fecha', y='Precio', marker='o', color='teal')

plt.gca().xaxis.set_major_locator(mdates.AutoDateLocator())
plt.xticks(rotation=45, ha='right')

plt.title('Evolução do Preço ao Longo do Tempo', fontsize=14)
plt.xlabel('Data')
plt.ylabel('Preço')
plt.grid(True, linestyle='--', alpha=0.4)
plt.tight_layout()
plt.show()
```



## 10. Cronograma:

Etapa	Entrega até	Responsável	Atividades Principais
Etapa 1 Metas e milestones	11/09	Grupo	<ul style="list-style-type: none"> <li>- Definição das premissas do projeto</li> <li>- Estabelecimento de objetivos e metas.</li> <li>- Elaboração do cronograma de atividades.</li> <li>- Organização e definição dos grupos.</li> <li>- Introdução ao pensamento computacional em contextos organizacionais.</li> </ul>
Etapa 2 Definição do produto	16/10	Grupo	<ul style="list-style-type: none"> <li>- Elaboração da proposta de solução analítica.</li> <li>- Desenvolvimento da análise exploratória dos dados.</li> </ul>
Etapa 3 Storytelling	13/11	Grupo	<ul style="list-style-type: none"> <li>- Planejamento de como apresentar resultados analíticos.</li> <li>- Aplicação de técnicas de Data Storytelling.</li> </ul>
Etapa 4 Encerramento	27/11	Grupo	<ul style="list-style-type: none"> <li>- Orientação para ajustes finais.</li> <li>- Apresentação dos resultados do projeto.</li> </ul>

## 11. Repositório GitHub

O projeto completo, incluindo os códigos utilizados para análise exploratória e as visualizações gráficas, está disponível em:

[Vanessa186578/Projeto-Aplicado-I: Universidade Presbiteriana Mackenzie - Banco de Dados - 2025 segundo semestre.](https://github.com/Vanessa186578/Projeto-Aplicado-I-Universidade-Presbiteriana-Mackenzie---Banco-de-Dados---2025-segundo-semestre)

## Referências

MARTÍNEZ OL, José Luis. Milk prices in Mexico (2024–2025). Kaggle, 2024. Disponível em: <https://www.kaggle.com/datasets/jlmartinezol/milk-prices-in-mexico-2024-2025>

. Acesso em: 11 set. 2025.

MÉXICO. Secretaría de Economía. SNIIM: Sistema Nacional de Información e Integración de Mercados. Cidade do México: Secretaría de Economía, 2024. Disponível em: <http://www.economia-sniim.gob.mx>

. Acesso em: 11 set. 2025.

SECRETARÍA DE ECONOMÍA (México). [Misión, Visión, Valores]. [S.I.]: SNIIM, [s.d.]. Disponível em: <http://www.economia-sniim.gob.mx/iso9000/Manual%20de%20Calidad/MANUAL%20DE%20CALIDAD%20SNIIM.pdf>

. Acesso em: 16 out. 2025.

SECRETARÍA DE ECONOMÍA (México). Qué es el Sistema Nacional de Información e Integración de Mercados (SNIIM). [S.I.]: Gobierno de México, 2025. Disponível em: <https://www.gob.mx/se/articulos/que-es-el-sistema-nacional-de-informacion-e-integracion-de-mercados-sniim>

. Acesso em: 16 out. 2025.

SECRETARÍA DE ECONOMÍA (México). Sistema Nacional de Información e Integración de Mercados (SNIIM). [S.I.]: [s.n.], [s.d.]. Disponível em: <https://www.economia-sniim.gob.mx/nuevo/Home.aspx?opcion=que.asp>. Acesso em: 16 out. 2025.

SECRETARÍA DE ECONOMÍA (México). Estadísticas – Anuarios del SNIIM. [S.I.]: [s.n.], [s.d.]. Disponível em: [https://www.economia-sniim.gob.mx/SNIIM-AN/estadisticas/i\\_anuariosinsu.asp](https://www.economia-sniim.gob.mx/SNIIM-AN/estadisticas/i_anuariosinsu.asp)

. Acesso em: 16 out. 2025.

SECRETARÍA DE ECONOMÍA (México). Enlaces Comerciales – SNIIM. [S.I.]: [s.n.], [s.d.]. Disponível em: <https://www.economia-sniim.gob.mx/EnlacesComerciales/ecComprar.asp>

. Acesso em: 16 out. 2025.