

**UNIVERSIDADE PRESBITERIANA MACKENZIE**

BRENNO MONTEIRO DE OLIVEIRA  
BRUNO FERNANDES MASCARINI  
LUCAS CAMARGO SPINELI  
MARINA CAMARGO SPINELI

PROJETO APLICADO I

ANÁLISE DOS DADOS DOS PREÇOS E VOLUME DE VENDA DOS ABACATES NOS EUA

São Paulo  
2024

BRENNO MONTEIRO DE OLIVEIRA  
BRUNO FERNANDES MASCARINI  
LUCAS CAMARGO SPINELI  
MARINA CAMARGO SPINELI

PROJETO APLICADO I

ANÁLISE DOS DADOS DOS PREÇOS E VOLUME DE VENDA DOS ABACATES NOS EUA

Projeto Aplicado I, do segundo período do curso  
de Tecnologia em Ciência de Dados da  
Universidade Presbiteriana Mackenzie.

Prof. Thiago Graziani Traue

São Paulo  
2024

## SUMÁRIO

1. INTRODUÇÃO.....	4
2. CRONOGRAMA.....	5
3. OBJETIVOS DO ESTUDO.....	6
4. CONTEXTO DE ESTUDO.....	7
5. APRESENTAÇÃO DA EMPRESA.....	8
6. DESCRIÇÃO DO DATASET.....	9
7. ANÁLISE EXPLORATÓRIA DE DADOS.....	11
7.1. Teoria por trás do código de análise exploratória de dados.....	11
7.2. Código de análise exploratória.....	14
7.3. Análise dos resultados.....	16
7.4. Solução dos problemas com base nos resultados.....	23
8. CONSIDERAÇÕES FINAIS.....	24
REFERÊNCIAS.....	25
GLOSSÁRIO.....	26

## 1. INTRODUÇÃO

O dataset dá informações detalhadas que partiram de um meme com o preço dos abacates, especialmente entre os millennials, que muitas vezes enfrentam dilemas sobre gastar dinheiro em itens cotidianos versus economizar. Os mais velhos comentavam que os mais novos não conseguiam comprar casas ou coisas/objetos por estar gastando com abacates. Neste contexto, a análise do dataset "Avocado Prices" oferece insights valiosos sobre o comportamento do mercado, influências culturais e a relevância econômica do abacate, considerando a sua popularidade na cultura moderna.

Os dados desse dataset serão analisados minuciosamente para entender a evolução dos preços ao longo do tempo, identificar padrões de consumo e explorar possíveis fatores que influenciam o mercado do abacate. Será também incluído um cronograma de análise que detalhará as etapas do processo, desde a coleta e preparação dos dados até a análise estatística e interpretação dos resultados. O cronograma ajudará a organizar o fluxo de trabalho, garantindo que todas as fases da análise sejam concluídas de forma eficiente e dentro dos prazos estipulados, proporcionando uma compreensão aprofundada das dinâmicas por trás do mercado do abacate.

Toda a documentação, incluindo o vídeo de apresentação deste projeto, ficará disponível no github, no link a seguir:

<https://github.com/lcsspi/Proj-Aplic-I-E-de-casa>



## 2. CRONOGRAMA

O cronograma a seguir ilustra as etapas a serem realizadas neste projeto. Cada integrante do grupo está representado por uma cor, indicando as respectivas responsabilidades de cada um; o “X” marca a expectativa da semana do mês de realização de cada parte do trabalho.

LEGENDA				
	BRENNO MONTEIRO DE OLIVEIRA		LUCAS CAMARGO SPINELI	TODOS
	BRUNO FERNANDES MASCARINI		MARINA CAMARGO SPINELI	

CRONOGRAMA - PROJETO APLICADO I																
Lista de Atividades	RESP.	AGOSTO				SETEMBRO				OUTUBRO				NOVEMBRO		
Organização do Documento				X	X											
Criação do Repositório				X												
Cronograma					X											
Objetivos do projeto					X											
Dataset e Metadados				X												
Organização do Documento								X								
Seção de Proposta Analítica inserida no Documento						X										
Scripts da Análise Exploratória em Python							X									
Seção de Análise Exploratória inserida no Documento								X								
Esboço do Storytelling										X	X					
Scripts da Análise Exploratória em Python									X							
Seção de Análise Exploratória inserida no Documento										X						
Apresentação gravada													X	X	X	
Documento Final																X

### **3. OBJETIVOS DO ESTUDO**

O estudo com o dataset "Avocado Prices" tem como objetivo principal analisar os padrões de variação dos preços de abacates nos Estados Unidos entre os anos de 2015 e 2018. A partir desses dados, busca-se identificar os principais fatores que influenciam as flutuações de preços, levando em consideração elementos como as regiões geográficas, os tipos de abacate (convencional ou orgânico) e os períodos de tempo específicos.

Além disso, o estudo explora as dinâmicas de mercado, incluindo eventos como a crise do abacate de 2017, quando uma combinação de alta demanda e baixa oferta devido a condições climáticas adversas elevou os preços de forma significativa. Com base nesses dados históricos, pretende-se criar modelos que auxiliem na previsão de futuras tendências de preços, oferecendo informações valiosas para o setor.

Outro ponto importante é a análise das preferências dos consumidores. O estudo visa entender melhor o comportamento do mercado, como a crescente demanda por abacates orgânicos e o impacto de diferentes embalagens nas vendas. Esses insights ajudarão a delinear estratégias de comercialização mais eficazes, ajustando a oferta de produtos conforme as preferências dos consumidores.

Em última análise, o estudo fornecerá subsídios para decisões mais informadas por parte da indústria, facilitando o planejamento de ações que promovam o ajuste de preços ou estratégias de marketing direcionadas. Dessa forma, o trabalho contribuirá para uma compreensão mais profunda das dinâmicas de oferta e demanda no mercado de abacates, gerando benefícios para toda a cadeia produtiva.

#### **4. CONTEXTO DE ESTUDO**

O problema que será estudado no dataset "Avocado Prices" envolve a análise e identificação de padrões de variação nos preços de abacates nos EUA entre 2015 e 2018. O estudo vai decompor o problema, examinando como diferentes fatores como regiões geográficas, tipos de abacate (convencional ou orgânico) e períodos de tempo influenciam os preços. A abstração dessas informações permitirá criar modelos para prever tendências futuras e explorar as dinâmicas de mercado que afetam os preços do abacate.



## 5. APRESENTAÇÃO DA EMPRESA

A Hass Avocado Board (HAB) é uma organização sem fins lucrativos criada em 2002 para promover o consumo de abacates Hass nos Estados Unidos. Financiada por produtores e importadores, a HAB reúne dados da indústria e realiza pesquisas nutricionais, ajudando a fortalecer a cadeia de valor do abacate.

**Missão:** A missão da HAB é apoiar os stakeholders globais da indústria de abacates em seus esforços coletivos para expandir o mercado nos EUA, com foco na saúde e bem-estar dos consumidores.

**Visão:** Tornar o abacate Hass a fruta mais popular e desejada nos EUA até 2025, com um aumento anual médio de 6% no valor do consumo per capita.

**Valores:** A organização se baseia na transparência, no fornecimento de dados de qualidade e no fomento de pesquisas científicas sobre nutrição e sustentabilidade para beneficiar toda a indústria.

**Segmento de atuação:** A HAB atua no segmento agrícola, promovendo a comercialização de abacates, principalmente nos EUA. Eles fornecem suporte para produtores e importadores por meio de dados de mercado, projeções de demanda e iniciativas de pesquisa.

**Market share:** Embora não haja uma métrica exata, o abacate Hass domina o mercado de abacates nos Estados Unidos. A HAB desempenha um papel importante ao promover o crescimento contínuo da demanda.

**Número de colaboradores:** A organização é dirigida por um conselho de 12 membros, mas não há informações específicas sobre o número total de colaboradores.

**Iniciativas na área de Data Science:** A HAB tem uma equipe dedicada à coleta e análise de dados de oferta e demanda, ajudando a indústria a tomar decisões informadas. Eles também realizam webinars e disponibilizam ferramentas para auxiliar os stakeholders na análise de dados de mercado.

## 6. DESCRIÇÃO DO DATASET

O dataset contém dados sobre preços médios de abacates, volumes de vendas, tipos de abacate (convencional ou orgânico) e regiões dos EUA. Além de analisar o comportamento do mercado de abacates, incluindo flutuações de preços e padrões de consumo em diferentes regiões e períodos, também é relevante considerar o impacto da crise de abacate nos EUA em 2017. Essa crise resultou de uma combinação de alta demanda e baixa oferta, devido a condições climáticas desfavoráveis e outros fatores, elevando significativamente os preços naquele ano.

Os dados que serão utilizados foram baixados do Kaggle, mas originalmente vêm do Hass Avocado Board. Eles foram agrupados em maio de 2018, reunindo informações semanais sobre vendas e preços de abacates da variedade Hass. Esse dataset inclui dados coletados em pontos de venda, obtidos diretamente das caixas registradoras de diversos tipos de estabelecimentos, como supermercados, clube de compras, farmácias e até mesmo lojas militares.

O arquivo do conjunto de dados está em formato CSV, o que significa que as informações estão organizadas em tabelas, ou seja, cada linha do arquivo representa um registro único e cada coluna (separada por vírgula) corresponde a um atributo específico.

Os dados registrados abrangem as vendas de abacates de 2015 até 2018 e incluem vários atributos que ajudam a descrever as características das transações. Como a data da observação (Date), o preço médio de um abacate (Average Price) e o volume total de abacates vendidos (Total Volume) são os principais e fornecem uma visão geral do mercado.

Além disso, as quantidades de abacates identificadas pelos códigos PLU 4046, 4225 e 4770 são apresentadas em conjunto, o que permite uma análise aprofundada das diferentes variedades. As informações sobre o tipo de embalagem também são fornecidas, elas são organizadas em um total de embalagens (Total Bags) e depois divididos em pequenas (Small Bags), grandes (Large Bags) e extragrandes (XLarge Bags), esses dados são extremamente importantes para auxiliar a compreender o volume em que os clientes preferem realizar suas compras.

A coluna tipo (type) distingue os abacates orgânicos (organic), que são cultivados sem pesticidas e fertilizantes sintéticos, dos convencionais (conventional), que são cultivados com esses produtos. Essa diferença é significativa porque mostra as tendências de consumo em direção a produtos mais sustentáveis. Além disso, a coluna ano (year) mostra o ano da observação, o que permite analisar tendências ao longo do tempo, enquanto a coluna região (region) mostra a localização das vendas, o que permite identificar variações nos padrões de compra regionais.

## **7. ANÁLISE EXPLORATÓRIA DE DADOS**

A análise exploratória de dados (EDA) desempenha um papel crucial na compreensão e interpretação de conjuntos de dados, permitindo a identificação de padrões, tendências e anomalias. Através de visualizações e estatísticas descritivas, buscamos não apenas revelar insights significativos, mas também justificar o uso de diferentes ferramentas e técnicas de visualização.

Ao longo do próximo subtópico, exploraremos cada etapa da análise, enfatizando a importância de cada gráfico e estatística no contexto do mercado de abacates.

### **7.1. Teoria por trás do código de análise exploratória de dados**

A parte teórica da análise exploratória de dados (EDA) é fundamental para interpretar corretamente os resultados e justificar o uso das ferramentas e gráficos que visualizamos no script. Vamos abordar os conceitos e motivações para cada etapa da análise, baseando-nos no script proposto:

#### **1. Estatísticas Descritivas**

Objetivo: Fornecer um resumo das características principais do dataset.

As estatísticas descritivas, como a média, mediana, desvio padrão, e quartis, ajudam a ter uma visão geral dos dados. Elas mostram como os preços e volumes de abacates estão distribuídos, bem como a variação desses valores. Isso é essencial para identificar possíveis anomalias (outliers) ou tendências iniciais.

- Média: Indica o valor central de uma variável, como o preço médio do abacate.
- Desvio padrão: Mostra a variabilidade dos dados. Um desvio padrão alto para o volume de vendas, por exemplo, indica uma grande dispersão entre as regiões ou datas.
- Quartis: Auxiliam a entender a distribuição dos dados e localizar os outliers, caso haja valores muito distantes da média.

#### **2. Preço Médio ao Longo do Tempo**

Objetivo: Detectar tendências temporais nos preços.

O gráfico de linha para o preço médio ao longo do tempo revela variações sazonais, picos ou quedas no preço. Esse tipo de análise é fundamental quando se trabalha com séries temporais, como é o caso de preços e volumes de venda ao longo de meses ou anos.

- Sazonalidade: Pode haver períodos do ano em que o preço do abacate aumenta ou diminui, devido a fatores de demanda e oferta.
- Eventos externos: Flutuações súbitas podem estar relacionadas a fatores externos, como mudanças climáticas, políticas ou logísticas, afetando a produção ou o transporte de abacates.

### 3. Distribuição do Preço Médio

Objetivo: Compreender como os preços estão distribuídos entre as observações.

Um histograma com curva de densidade estimada (KDE) fornece uma visão mais clara da distribuição do preço médio. Ele permite observar se os preços seguem uma distribuição normal, se estão enviesados (skewed) ou se existem múltiplos picos (distribuição multimodal).

- Distribuição simétrica: Indica que a maioria dos preços está concentrada ao redor da média.
- Distribuição enviesada: Se o gráfico for assimétrico, isso pode indicar que há mais observações em uma faixa de preços (mais barato ou mais caro).
- Multimodalidade: Múltiplos picos sugerem a existência de diferentes grupos de preços, possivelmente relacionados a fatores como regiões ou tipo de abacate (orgânico vs convencional).

### 4. Comparação de Preços por Tipo de Abacate

Objetivo: Verificar a diferença de preço entre abacates orgânicos e convencionais.

O boxplot permite comparar visualmente a distribuição dos preços para os dois tipos de abacate (orgânico e convencional). É útil para observar a mediana (linha no meio da caixa), a amplitude interquartil (tamanho da caixa) e os outliers (pontos fora da caixa).

- Diferenças no preço: Se os preços dos abacates orgânicos forem consistentemente mais altos, isso pode ser esperado, dada a natureza do produto, mas é importante verificar a magnitude dessa diferença.

- Outliers: Valores anormalmente altos ou baixos podem representar promoções, sazonalidades ou erros nos dados.

#### 5. Volume Total Vendido por Região

Objetivo: Identificar quais regiões têm maior volume de vendas.

Um gráfico de barras classificado por volume de vendas por região permite comparar a participação de mercado das diferentes regiões. Isso é importante para entender o impacto geográfico nas vendas e identificar onde a demanda é maior.

- Regiões com maior demanda: Regiões que vendem mais abacates podem estar associadas a grandes centros urbanos ou áreas onde o consumo de abacate é culturalmente mais popular.
- Análise geográfica: Entender a distribuição do volume pode ajudar na alocação de recursos, estratégias de marketing e otimização logística.

#### 6. Tendência do Preço Médio por Ano

Objetivo: Avaliar como os preços mudaram ao longo dos anos.

- Agrupar o preço médio por ano e tipo de abacate (orgânico vs convencional) permite visualizar a tendência de preços ao longo do tempo. Isso ajuda a verificar se houve uma tendência de alta ou baixa nos preços e se o comportamento é semelhante para ambos os tipos.
- Tendência de alta ou baixa: Uma tendência de alta nos preços ao longo dos anos pode refletir uma maior demanda, custos de produção ou inflação.
- Diferença entre orgânico e convencional: Isso ajuda a entender se o aumento ou a diminuição dos preços é mais pronunciado em abacates orgânicos ou convencionais.

#### 7. Volume Total por Tipo de Abacate

Objetivo: Comparar o volume de vendas entre abacates convencionais e orgânicos.

Este gráfico de barras permite visualizar se o volume de vendas de abacates orgânicos está crescendo em relação aos convencionais. Com a crescente popularidade de alimentos orgânicos, essa comparação pode fornecer insights sobre as preferências dos consumidores.

- Preferência por tipo: Se o volume de vendas de orgânicos estiver crescendo, pode indicar uma mudança no comportamento do consumidor em direção a escolhas mais sustentáveis e saudáveis.

- Oportunidades de mercado: Vendas mais altas de um tipo específico podem sinalizar áreas para expandir a oferta ou ajustar o marketing.

## 8. Correlação entre Variáveis Numéricas

Objetivo: Entender como diferentes variáveis numéricas estão relacionadas.

A matriz de correlação ajuda a identificar as relações entre variáveis como volume de vendas, preço médio e as diferentes categorias de abacate. O heatmap facilita a visualização de correlações fortes (positivas ou negativas).

- Correlação positiva: Se uma variável, como o volume de vendas de um tipo de abacate, tiver uma alta correlação com o preço médio, isso pode sugerir que a demanda está diretamente relacionada ao preço.
- Correlação negativa: Uma correlação negativa pode sugerir que, conforme o preço aumenta, o volume de vendas diminui, o que é esperado em muitos mercados.

## 7.2. Código de análise exploratória

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np

# Carregar o dataset
file_path =
"https://raw.githubusercontent.com/lcsspi/Proj-Aplic-I-E-de-casa/refs/heads/main/Dataset/avocado.csv"
avocado_data = pd.read_csv(file_path)

# Converter a coluna 'Date' para o formato datetime
avocado_data['Date'] = pd.to_datetime(avocado_data['Date'])

# 1. Estatísticas descritivas
print("Estatísticas descritivas:")
print(avocado_data.describe())

# 2. Preço médio ao longo do tempo
```

```
plt.figure(figsize=(10, 6))
sns.lineplot(x='Date', y='AveragePrice', data=avocado_data, errorbar=None)
plt.title('Preço Médio dos Abacates ao Longo do Tempo', fontsize=16)
plt.xlabel('Data', fontsize=12)
plt.ylabel('Preço Médio (USD)', fontsize=12)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

# 3. Distribuição do preço médio

```
plt.figure(figsize=(10,6))
sns.histplot(avocado_data['AveragePrice'], bins=30, kde=True)
plt.title('Distribuição do Preço Médio', fontsize=16)
plt.xlabel('Preço Médio (USD)', fontsize=12)
plt.ylabel('Frequência', fontsize=12)
plt.tight_layout()
plt.show()
```

# 4. Preço médio por tipo de abacate (convencional vs orgânico)

```
plt.figure(figsize=(10,6))
sns.boxplot(x='type', y='AveragePrice', data=avocado_data)
plt.title('Comparação de Preço Médio por Tipo de Abacate', fontsize=16)
plt.xlabel('Tipo', fontsize=12)
plt.ylabel('Preço Médio (USD)', fontsize=12)
plt.tight_layout()
plt.show()
```

# 5. Volume total vendido por região

```
plt.figure(figsize=(12, 8))
region_volume = avocado_data.groupby('region')['Total Volume'].sum().sort_values(ascending=False)
sns.barplot(x=region_volume.values, y=region_volume.index, hue=region_volume.index,
palette="viridis", legend=False)
plt.title('Volume Total Vendido por Região', fontsize=16)
plt.xlabel('Volume Total', fontsize=12)
plt.ylabel('Região', fontsize=12)
plt.tight_layout()
plt.show()
```

# 6. Tendência do preço médio por ano

```
plt.figure(figsize=(10, 6))
```



```

avocado_data['year'] = avocado_data['Date'].dt.year
sns.lineplot(x='year', y='AveragePrice', hue='type', data=avocado_data, errorbar=None)
plt.title('Tendência do Preço Médio por Ano', fontsize=16)
plt.xlabel('Ano', fontsize=12)
plt.ylabel('Preço Médio (USD)', fontsize=12)
plt.tight_layout()
plt.show()

# 7. Volume total vendido por tipo de abacate
plt.figure(figsize=(10, 6))
type_volume = avocado_data.groupby('type')['Total Volume'].sum()
sns.barplot(x=type_volume.index, y=type_volume.values) # Removido palette
plt.title('Volume Total Vendido por Tipo de Abacate', fontsize=16)
plt.xlabel('Tipo', fontsize=12)
plt.ylabel('Volume Total', fontsize=12)
plt.tight_layout()
plt.show()

# 8. Correlação entre variáveis numéricas
numeric_data = avocado_data.select_dtypes(include=[np.number]) # Seleciona apenas
colunas numéricas
correlation_matrix = numeric_data.corr()

plt.figure(figsize=(10,8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f', linewidths=0.5)
plt.title('Heatmap de Correlações entre Variáveis Numéricas', fontsize=16)
plt.tight_layout()
plt.show()

```

### 7.3. Análise dos resultados

Gráfico 1 - Preço Médio ao Longo do Tempo (Lineplot):

Mostra a evolução do preço médio dos abacates ao longo do tempo.

Esse gráfico ajuda a identificar tendências sazonais ou mudanças significativas nos preços.

Se houver picos ou quedas em certas datas, isso pode indicar sazonalidade (por exemplo, aumentos no preço próximo a festividades) ou eventos econômicos específicos.

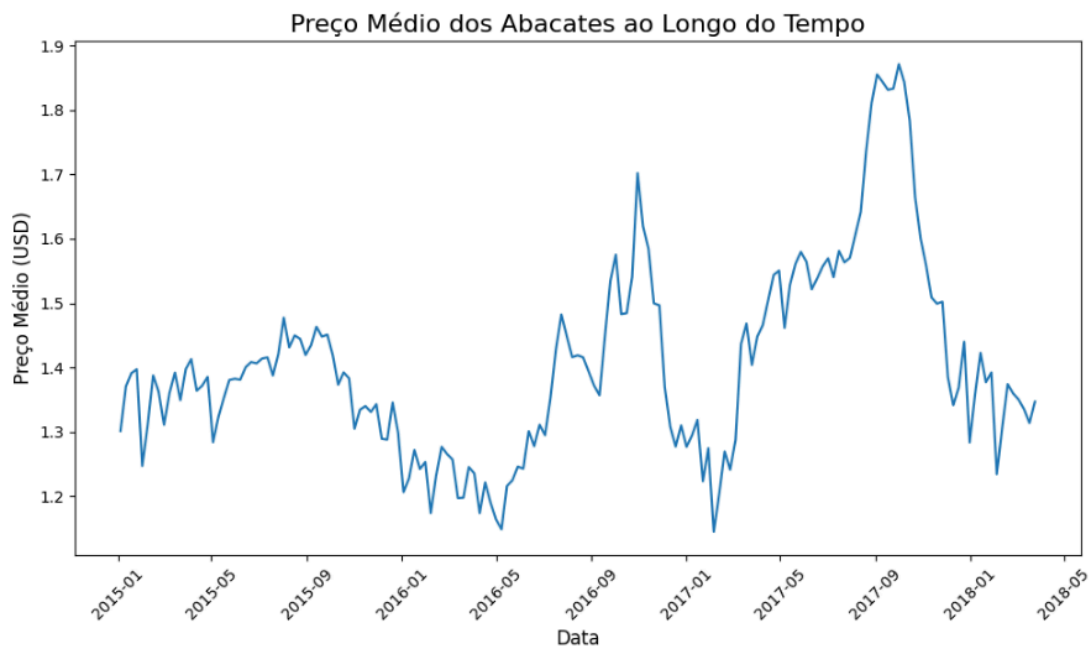


Gráfico 2 - Distribuição do Preço Médio (Boxplot):

O histograma mostra a frequência dos preços em diferentes faixas.

A curva KDE (Kernel Density Estimate) suaviza a distribuição, ajudando a visualizar se há uma média central (preço típico) e como os dados se espalham ao redor dessa média.

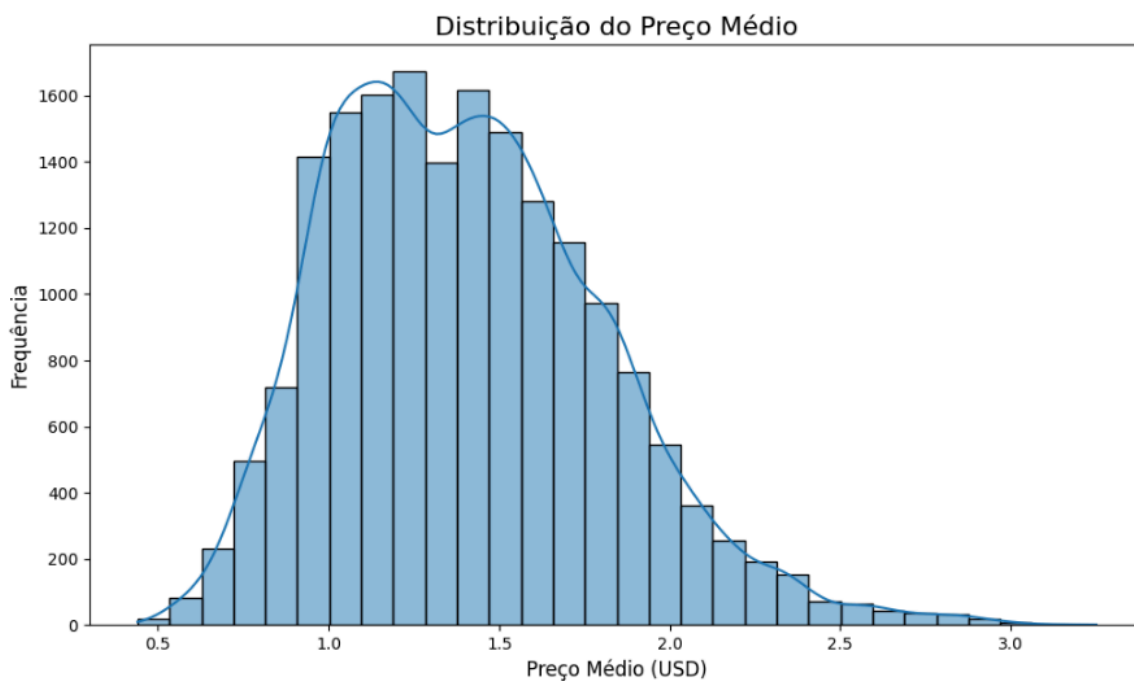


Gráfico 3 - Comparação de Preço Médio por Tipo de Abacate (Boxplot):

Compara os preços médios entre abacates convencionais e orgânicos.

Cada caixa representa a distribuição dos preços para cada tipo, com os limites superiores e inferiores indicando o intervalo interquartil.

Outliers (valores fora do padrão) são marcados como pontos fora da caixa.

Se o preço médio dos abacates orgânicos for maior, isso pode refletir custos adicionais associados à produção orgânica.

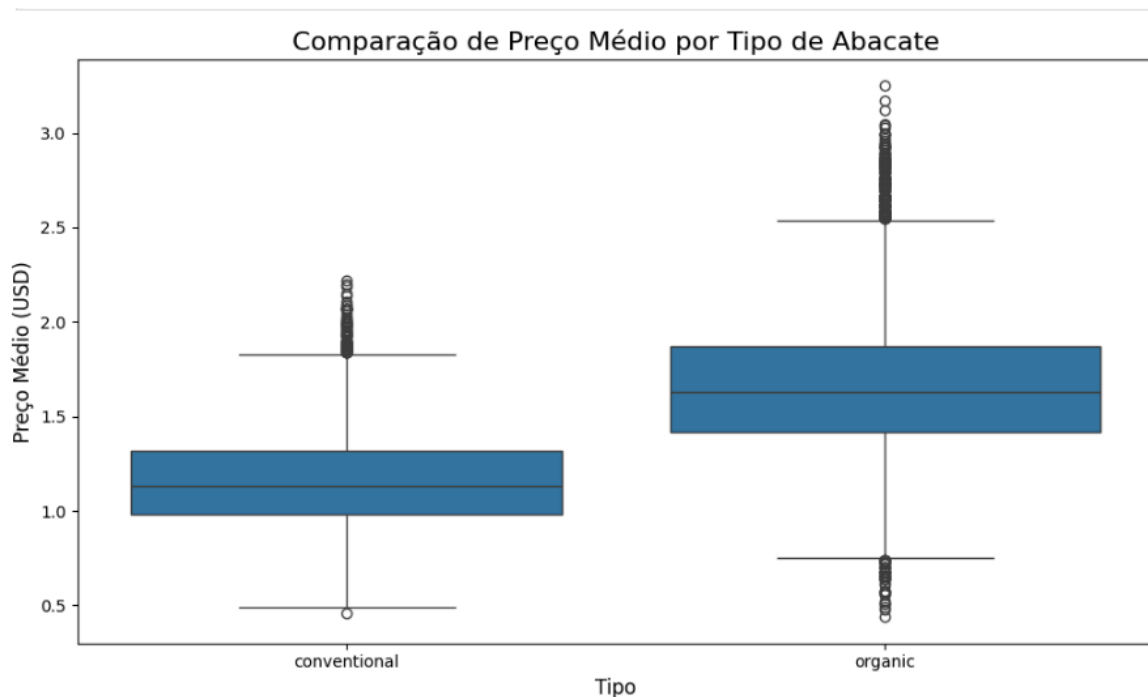


Gráfico 4 - Volume Total Vendido por Região (Barplot Horizontal):

Mostra quais regiões tiveram o maior volume de vendas.

As regiões mais à esquerda possuem maior volume total vendido.

Regiões com vendas expressivas podem ter uma forte demanda ou contar com cadeias de distribuição eficientes.

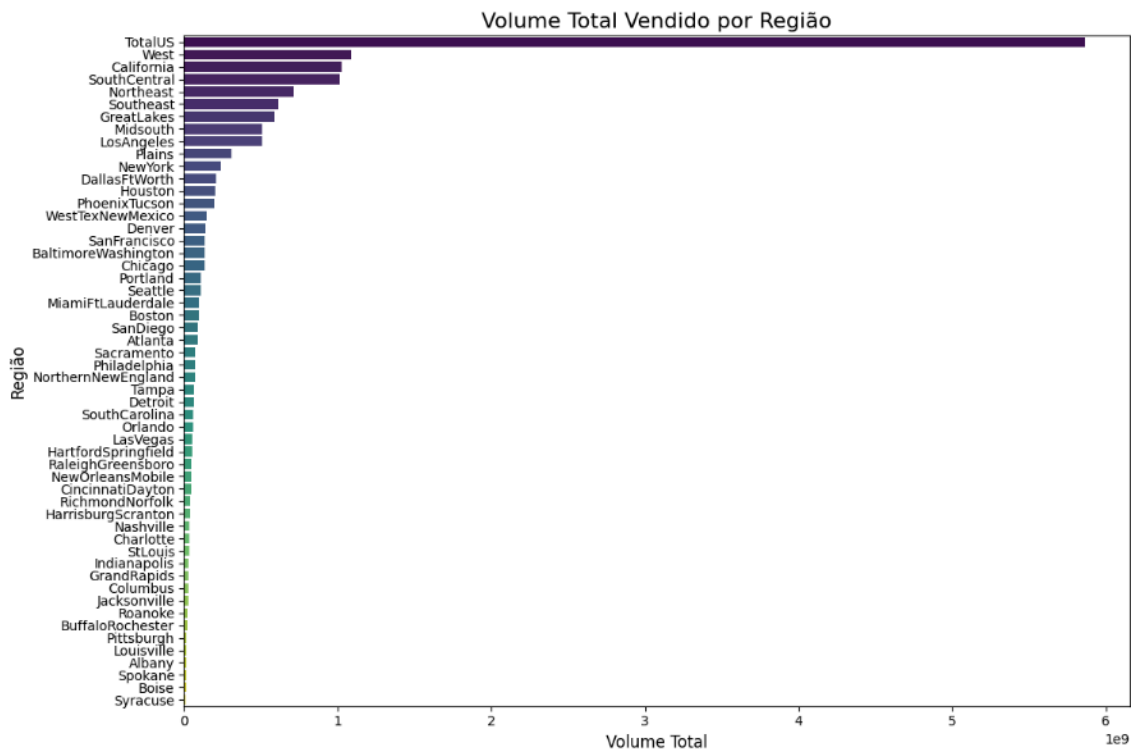


Gráfico 5 - Tendência do Preço Médio por Ano e Tipo (Lineplot):

Analisa como o preço médio variou ao longo dos anos e compara as tendências entre abacates orgânicos e convencionais.

As linhas ascendentes indicam que o preço está aumentando com o tempo.

Comparando as linhas para cada tipo, você pode verificar se a diferença de preço entre abacates orgânicos e convencionais se manteve constante ou mudou ao longo dos anos.

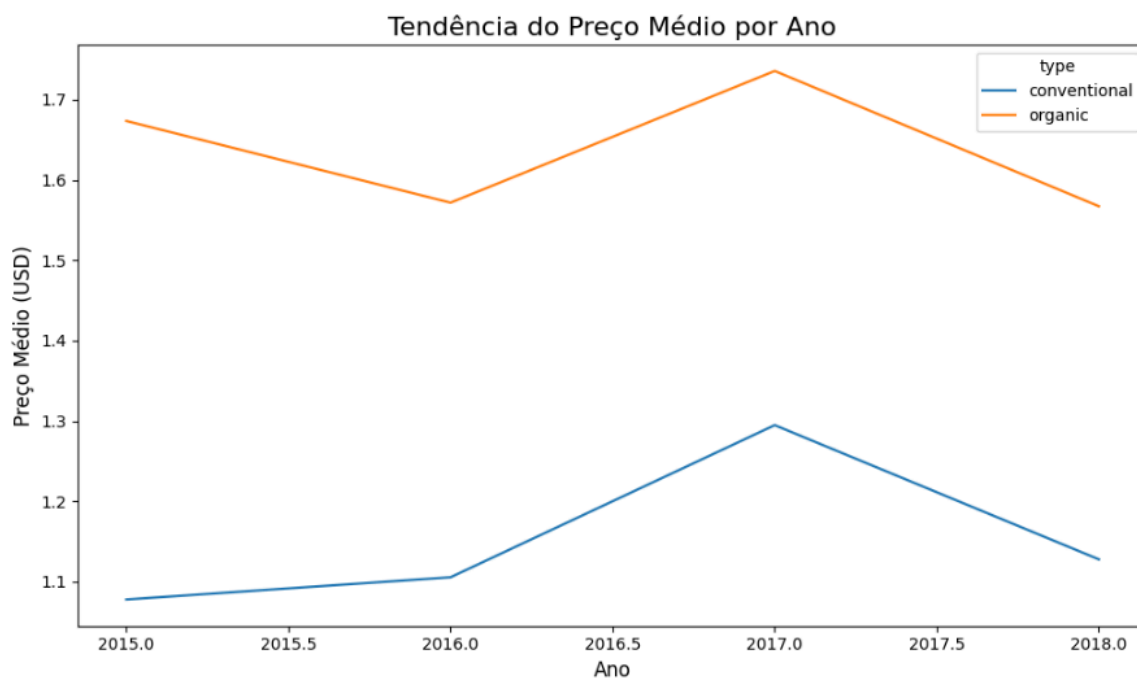
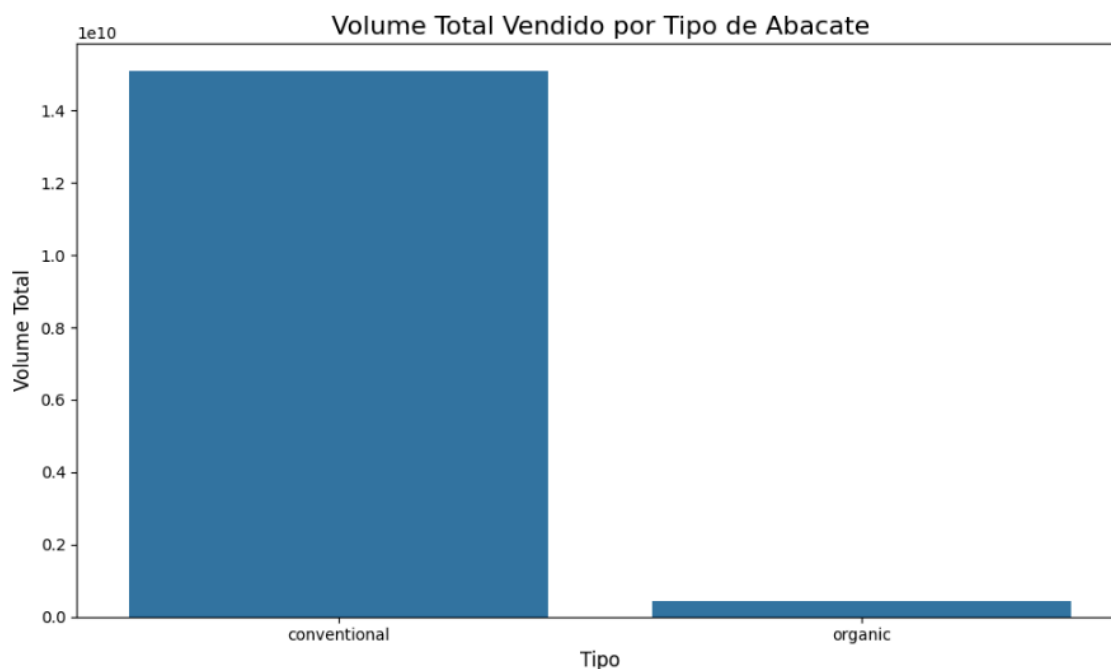


Gráfico 6 - Volume Total Vendido por Tipo de Abacate (Barplot Vertical):

Compara o volume total vendido entre abacates convencionais e orgânicos.

Esse gráfico ajuda a identificar qual tipo de abacate teve maior participação nas vendas.

Se abacates convencionais tiverem maior volume, pode indicar que eles são mais populares ou mais acessíveis.



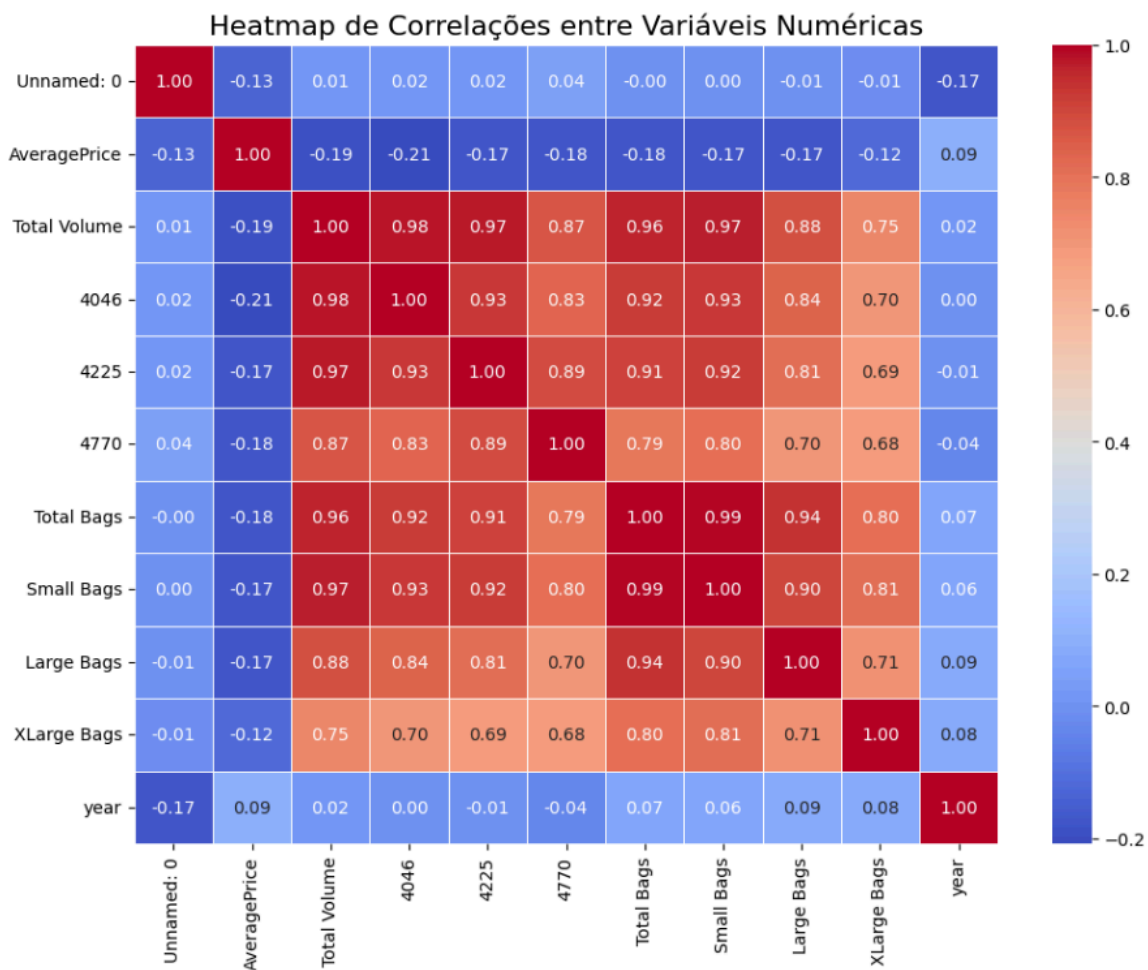
## Gráfico 7 - Heatmap de Correlações entre Variáveis Numéricas

Mostra a correlação entre as variáveis numéricas do dataset.

Valores próximos de +1 indicam forte correlação positiva (quando uma variável aumenta, a outra também aumenta).

Valores próximos de -1 indicam correlação negativa (quando uma variável aumenta, a outra diminui).

Se o preço médio estiver negativamente relacionado com o volume vendido, isso pode sugerir que, quando o preço sobe, a demanda cai, o que faz sentido economicamente.



#### **7.4. Solução dos problemas com base nos resultados**

De acordo com os resultados apresentados, é possível determinar algumas estratégias para impulsionar ainda mais o mercado de abacates.

A primeira estratégia consiste em intensificar a promoção de abacates orgânicos, destacando seus benefícios à saúde e à sustentabilidade. Desenvolver campanhas de marketing direcionadas que enfatizem a qualidade e os atributos dos abacates orgânicos, utilizando dados que mostrem o crescimento da demanda.

A segunda estratégia é monitorar as condições climáticas e seu impacto na oferta de abacates. É preciso desenvolver um sistema de alerta que informe sobre eventos climáticos que possam afetar a produção, permitindo ajustes proativos nas estratégias de marketing e distribuição.

A terceira e última sugestão de solução é o modelo de sugestão de preços. Criar modelos preditivos utilizando os dados históricos para prever tendências de preços, implementar um sistema de análise contínua que ajude os stakeholders a tomar decisões informadas sobre produção e marketing, ajustando-se às flutuações do mercado.

## **8. CONSIDERAÇÕES FINAIS**

A pesquisa sobre as flutuações de preços de abacates nos Estados Unidos entre 2015 e 2018 revelou um cenário complexo, onde o tipo de abacate, a região geográfica e as condições climáticas influenciam significativamente os preços. Observou-se um aumento na demanda por abacates orgânicos, o que representa uma oportunidade de mercado para produtos mais saudáveis e sustentáveis. Além disso, a análise preditiva se mostrou crucial para prever preços e orientar decisões estratégicas, enquanto o monitoramento constante das condições climáticas é essencial para equilibrar oferta e demanda.

Com base nessas descobertas, recomenda-se investir em campanhas de marketing que destaquem os benefícios dos abacates orgânicos, implementar sistemas de alerta para monitorar condições climáticas e utilizar dados da pesquisa para otimizar estratégias de comercialização. Assim, a pesquisa oferece uma base sólida para o desenvolvimento de estratégias mais eficazes, visando um crescimento sustentável da indústria e fortalecimento da cadeia produtiva, consolidando o abacate Hass como um produto valorizado nos Estados Unidos. Em suma, destaca-se a importância da análise de dados e da adaptação às demandas do mercado para garantir o sucesso no setor de abacates.



## REFERÊNCIAS

**Dicio, Dicionário Online de Português.** Disponível em: <<https://www.dicio.com.br/>>. Acesso em: 28 set. 2024.

KIGGINS, J. **Avocado Prices.** Disponível em: <<https://www.kaggle.com/datasets/neuromusic/avocado-prices>>. Acesso em: 26 ago. 2024.

OXFORD DICTIONARY. **Oxford English Dictionary.** Disponível em: <<https://www.oed.com/?tl=true>>. Acesso em: 28 set. 2024.

**Volume and Category Board.** Disponível em: <<https://hassavocadoboard.com/>>. Acesso em: 30 ago. 2024.

## GLOSSÁRIO

**Análise Exploratória de Dados (EDA):** Processo de examinar um conjunto de dados para resumir suas principais características, frequentemente com métodos visuais.

**Cadeia Produtiva:** Conjunto de etapas que um produto percorre, desde a produção até o consumo final.

**Códigos PLU:** Códigos de identificação usados para identificar tipos específicos de produtos, como frutas e vegetais.

**Correlação:** Medida estatística que expressa a extensão em que duas variáveis estão relacionadas.

**Curva de Densidade Estimada (KDE):** Método de suavização que mostra a distribuição de probabilidade de uma variável contínua.

**Desvio Padrão:** Medida de variabilidade que indica o quanto os valores de um conjunto de dados se desviam da média.

**Estatísticas Descritivas:** Resumos numéricos que descrevem as características de um conjunto de dados, como média, mediana, desvio padrão e quartis.

**Market Share:** Participação de mercado, medida pelas vendas de um produto específico; proporção do mercado que ela controla.

**Quartil:** Valores que dividem um conjunto de dados em quatro partes iguais, ajudando a entender a distribuição.

**Sazonalidade:** Variação previsível em um fenômeno ao longo do tempo, frequentemente associada a ciclos sazonais.

**Stakeholder:** Partes interessadas ou envolvidas em um projeto ou setor, como produtores, consumidores e autoridades.

## LISTA DE ILUSTRAÇÕES

Gráfico 1 - Apresenta a evolução do preço médio dos abacates ao longo do tempo.....	17
Gráfico 2 - Apresenta a distribuição dos preços médios dos abacates.....	18
Gráfico 3 - Apresenta a comparação de preços médios entre abacates convencionais e orgânicos.....	19
Gráfico 4 - Apresenta o volume de vendas de abacate por região.....	20
Gráfico 5 - Apresenta a variação de tendências e preços dos abacates ao longo dos anos.....	21
Gráfico 6 - Apresenta uma comparação de volume total vendido entre abacates convencionais e orgânicos.....	21
Gráfico 7 - Apresenta a correlação entre as variáveis numéricas do dataset.....	22