

# Análisis del Conjunto de Datos de Precios de Aguacate

**Conjunto de Datos de Precios de Aguacate:** El conjunto de datos "Precios de Aguacate", obtenido de Kaggle, es un conjunto de datos ampliamente utilizado para proyectos de análisis de datos y aprendizaje automático. Proporciona datos históricos sobre precios y ventas de aguacates en varias regiones de los Estados Unidos. Este conjunto de datos es valioso para entender las tendencias en los precios de los aguacates, los volúmenes de ventas y su relación con diferentes factores.

## Atributos Clave

- **Columnas:** El conjunto de datos incluye varias columnas de información. Algunas de las columnas clave típicamente encontradas en este conjunto de datos incluyen:
  - **Fecha** (Date): La fecha de observación.
  - **Precio Promedio** (AveragePrice): El precio promedio de los aguacates.
  - **Volumen Total** (Total Volume): El volumen total de aguacates vendidos.
  - **4046**: Volumen de aguacates Hass pequeños vendidos.
  - **4225**: Volumen de aguacates Hass grandes vendidos.
  - **4770**: Volumen de aguacates Hass extra grandes vendidos.
  - **Bolsas Totales** (Total Bags): Total de bolsas de aguacates vendidas.
  - **Bolsas Pequeñas** (Small Bags): Bolsas de aguacates pequeños vendidas.
  - **Bolsas Grandes** (Large Bags): Bolsas de aguacates grandes vendidas.
  - **Bolsas Extra Grandes** (XLarge Bags): Bolsas de aguacates extra grandes vendidas.
  - **Tipo** (Type): El tipo de aguacates, generalmente categorizados como convencionales u orgánicos.
  - **Región** (Region): La región o ciudad dentro de los Estados Unidos donde se registraron los datos.
- **Rango de Fechas:** El conjunto de datos abarca un rango de fechas, lo que permite el análisis de series de tiempo. Puedes examinar cómo cambian los precios y ventas de aguacates a lo largo de diferentes estaciones y años.
- **Regiones:** Se proporciona información para varias regiones o ciudades a través de los Estados Unidos, lo que permite el análisis de variaciones de precios y ventas en diferentes mercados.
- **Tipos:** El conjunto de datos distingue entre diferentes tipos de aguacates, como convencionales y orgánicos, lo que puede ser útil para comparar tendencias de precios entre estas categorías.
- **Volumen:** Están disponibles datos sobre el volumen total de aguacates vendidos. Esta métrica de volumen se utiliza a menudo para analizar la demanda del mercado.
- **Precio Promedio:** El conjunto de datos contiene el precio promedio de los aguacates, una métrica fundamental para entender las tendencias de precios.

# Casos de Uso

- Este conjunto de datos se utiliza comúnmente para aprender y practicar el análisis de datos, visualización de datos y modelado de regresión en proyectos de ciencia de datos y aprendizaje automático.
  - Sirve como un recurso valioso para entender cómo trabajar con datos del mundo real, extraer conocimientos y tomar decisiones basadas en datos.
- 

## Actividades de Análisis

### 1. Análisis de Series Temporales

**Resumen:** El análisis de series temporales permite identificar patrones, tendencias y estacionalidades en los precios y volúmenes de ventas de aguacates a lo largo del tiempo.

#### 1. Descomposición de Series Temporales de Precios:

- **Uso de Datos:** Usa la columna `AveragePrice` y `Date`.
- **Esperado:** Utiliza la función `seasonal_decompose` de la librería `statsmodels` para descomponer la serie temporal de precios en componentes de tendencia, estacionalidad y ruido.
  - Convierte `Date` a tipo `datetime` usando `pd.to_datetime()`.
  - Agrupa los datos por `Date` y calcula el promedio de `AveragePrice` utilizando `groupby()` si es necesario.
  - Visualiza los componentes descompuestos usando `matplotlib` para cada uno de ellos.

#### 2. Análisis de Estacionalidad por Región:

- **Uso de Datos:** Usa las columnas `AveragePrice`, `Date` y `Total Volume`.
- **Esperado:** Utiliza gráficos de líneas para visualizar cómo varían los precios de aguacates por región a lo largo de diferentes estaciones del año.
  - Agrupa los datos por `region` y `Date` utilizando `groupby()`.
  - Calcula el promedio de `AveragePrice` para cada región.
  - Representa gráficamente las tendencias utilizando `plt.plot()` de `matplotlib`.

#### 3. Comparación de Precios Promedio Mensuales:

- **Uso de Datos:** Usa las columnas `AveragePrice` y `Date`.
- **Esperado:** Calcula y compara los precios promedio mensuales.
  - Agrupa los datos por mes usando `pd.Grouper` con `freq='M'`.
  - Calcula el promedio de `AveragePrice` para cada mes con `mean()`.
  - Visualiza los resultados con un gráfico de líneas usando `plt.plot()`.

#### 4. Tendencia de Ventas a lo Largo del Tiempo:

- **Uso de Datos:** Usa las columnas `Total Volume` y `Date`.

- **Esperado:** Analiza cómo varía el volumen total de ventas a lo largo del tiempo.
  - Agrupa los datos por Date y suma el Total Volume usando `groupby()`.
  - Visualiza los resultados usando un gráfico de líneas con `plt.plot()` para mostrar la tendencia.

#### 5. **Análisis de Cambios en Precios Anuales:**

- **Uso de Datos:** Usa las columnas `AveragePrice` y `year`.
- **Esperado:** Observa las diferencias anuales en los precios promedio.
  - Agrupa los datos por year utilizando `groupby()`.
  - Calcula el promedio de `AveragePrice` para cada año.
  - Representa los resultados en un gráfico de barras usando `plt.bar()` que compare los precios de cada año.

## 2. Gráficos para Visualización de Datos

**Resumen:** La visualización de datos es clave para identificar patrones y relaciones entre diferentes variables. Los gráficos apropiados pueden proporcionar información valiosa sobre el comportamiento de los precios y volúmenes de ventas.

#### 1. **Gráfico de Violín de Volumen de Ventas por Región:**

- **Uso de Datos:** Usa las columnas `Total Volume` y `region`.
- **Esperado:** Visualiza la distribución de ventas en diferentes regiones.
  - Utiliza la función `violinplot` de `seaborn` para crear gráficos de violín.
  - Configura los ejes para mostrar la relación entre `Total Volume` y `region`.
  - Añade etiquetas y títulos usando `plt.title()` y `plt.xlabel()` para facilitar la interpretación.

#### 2. **Boxplot Comparativo de Precios entre Años:**

- **Uso de Datos:** Usa las columnas `AveragePrice` y `year`.
- **Esperado:** Genera boxplots para comparar la distribución de precios.
  - Utiliza `boxplot` de `seaborn` para crear boxplots que comparen `AveragePrice` entre diferentes años.
  - Asegúrate de que cada boxplot repre