



Universidad Autónoma de Baja California



Facultad de Ciencias

Programación para ciencias de datos

Proyecto final

keymi Aylin Ponce Aguirre
matrícula: 374479

Ensenada, Baja California a 5 de Junio del 2024

¿Cómo varían los precios de las tortillas en México según el tipo de tienda y la ubicación?

Para responder esa pregunta utilizamos el siguiente dataset <https://www.kaggle.com/datasets/richave/tortilla-prices-in-mexico>

1. Comenzamos importando las librerías que se utilizaran para el análisis.

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

2. Importamos el dataset y lo guardamos en price_df.

```
dataset = '/content/tortilla_prices.csv'
price_df = pd.read_csv(dataset)
price_df.head()
```

	State	City	Year	Month	Day	Store type	Price per kilogram
0	Aguascalientes	Aguascalientes	2007	1	10	Mom and Pop Store	9.9
1	Baja California	Mexicali	2007	1	10	Mom and Pop Store	NaN
2	Baja California	Tijuana	2007	1	10	Mom and Pop Store	10.0
3	Baja California Sur	La Paz	2007	1	10	Mom and Pop Store	10.0
4	Campeche	Campeche	2007	1	10	Mom and Pop Store	10.0

3. Lo primero que se realizará será juntar las columnas Year, Month y Day en una sola, para crear la columna fecha.

```
price_df['fecha'] = pd.to_datetime(price_df[['Year', 'Month', 'Day']])
price_df.head()
```

	State	City	Year	Month	Day	Store type	Price per kilogram	fecha
0	Aguascalientes	Aguascalientes	2007	1	10	Mom and Pop Store	9.9	2007-01-10
1	Baja California	Mexicali	2007	1	10	Mom and Pop Store	NaN	2007-01-10
2	Baja California	Tijuana	2007	1	10	Mom and Pop Store	10.0	2007-01-10
3	Baja California Sur	La Paz	2007	1	10	Mom and Pop Store	10.0	2007-01-10
4	Campeche	Campeche	2007	1	10	Mom and Pop Store	10.0	2007-01-10

4. Se eliminarán las columnas Year, Month y Day

```
price_df.drop(['Year', 'Month', 'Day'], axis=1, inplace=True)
price_df.head()
```

	State	City	Store type	Price per kilogram	fecha
0	Aguascalientes	Aguascalientes	Mom and Pop Store	9.9	2007-01-10
1	Baja California	Mexicali	Mom and Pop Store	NaN	2007-01-10
2	Baja California	Tijuana	Mom and Pop Store	10.0	2007-01-10
3	Baja California Sur	La Paz	Mom and Pop Store	10.0	2007-01-10
4	Campeche	Campeche	Mom and Pop Store	10.0	2007-01-10

5. Utilizaremos el `.describe` que nos mostrará un resumen estadístico de los precios de las tortillas. Incluye información como la cantidad de datos, el promedio, la desviación estándar, los cuartiles y los valores mínimo y máximo. Esta información es útil para entender la distribución y la variabilidad de los precios, lo que nos ayuda a comparar y analizar cómo varían los precios de las tortillas según el tipo de tienda y la ubicación en México.

```
descripcion = price_df['Price per kilogram'].describe()
print(descripcion)
```

```
count    273902.000000
mean      12.092700
std        4.376921
min        0.000000
25%        9.570000
50%       11.180000
75%       14.000000
max       31.000000
Name: Price per kilogram, dtype: float64
```

6. Utilizaremos el siguiente código para calcular y mostrar el precio promedio mensual de las tortillas. Primero convertimos la columna de fecha al formato `datetime` y la establecemos como el índice del `DataFrame`. Luego, utiliza el método `resample()` para agrupar los datos por mes y calcula el precio promedio de cada mes. Esta información no muestra como varían los precios.

```
price_df['fecha'] = pd.to_datetime(price_df['fecha'])
price_df.set_index('fecha', inplace=True)

promedio_mensual = price_df['Price per kilogram'].resample('M').mean()
print(promedio_mensual)
```

```
fecha
2007-01-31    7.459833
2007-02-28    7.392718
2007-03-31    7.368155
2007-04-30    7.377820
2007-05-31    7.394428
```

7. Lo mismo, solo que por año.

```
promedio_anual = price_df['Price per kilogram'].resample('Y').mean()
print(promedio_anual)
```

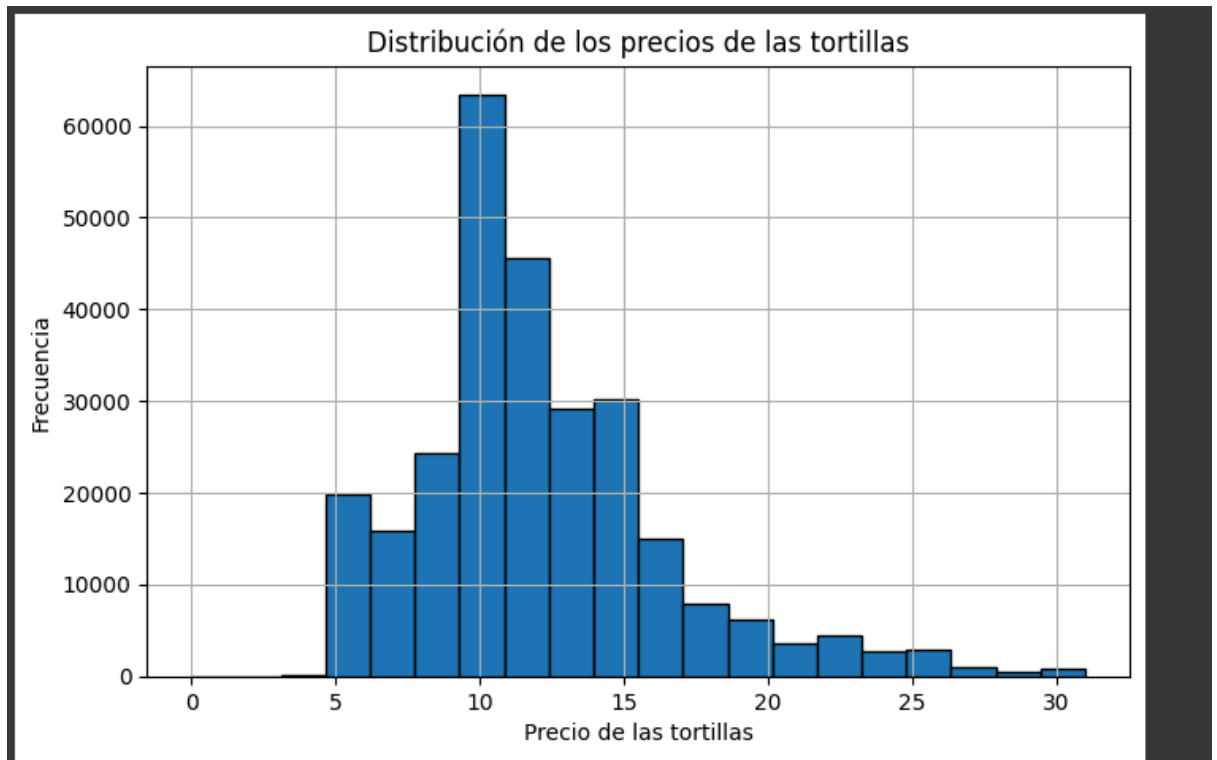
```
fecha
2007-12-31    7.399119
2008-12-31    7.598889
2009-12-31    8.094949
2010-12-31    8.670437
2011-12-31    9.737406
```

8. Se crea un gráfico de barras (histograma) que muestra cómo se distribuyen los precios de las tortillas. Cada barra representa un rango de precios y su altura indica cuántas veces aparece ese rango en los datos. El gráfico nos ayuda a entender la variabilidad y la tendencia central de los precios de las tortillas.

```
plt.figure(figsize=(8, 5))
plt.hist(price_df['Price per kilogram'], bins=20, edgecolor='black')

plt.xlabel('Precio de las tortillas')
plt.ylabel('Frecuencia')
plt.title('Distribución de los precios de las tortillas')

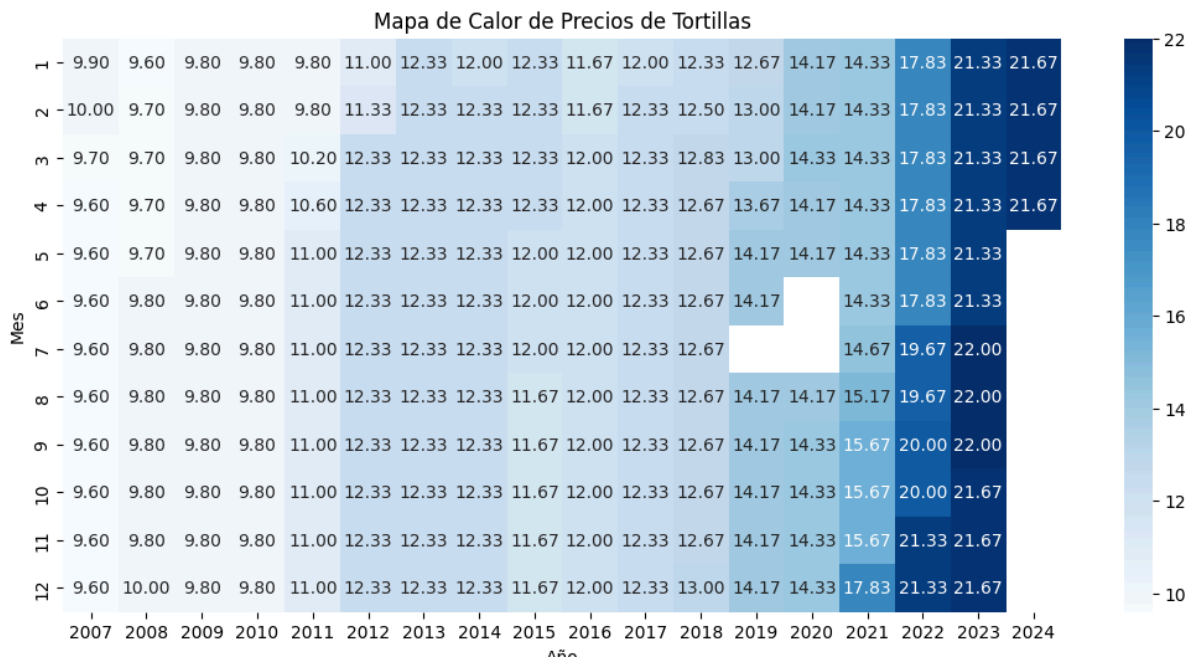
plt.grid(True)
plt.show()
```



9. Utilizamos un mapa de calor que muestra cómo cambian los precios de las tortillas con el tiempo. Elimina datos duplicados, como si tuvieras dos registros para el mismo mes y año. Luego organiza los datos de forma que los meses sean las filas y los años las columnas, y cada celda del mapa contiene el precio promedio de las tortillas para ese mes y año en particular. Las celdas tienen colores más oscuros si los precios son más altos y colores más claros si son más bajos.

```
price_df_cleaned = price_df.drop_duplicates(subset=['Month', 'Year'], keep='first')

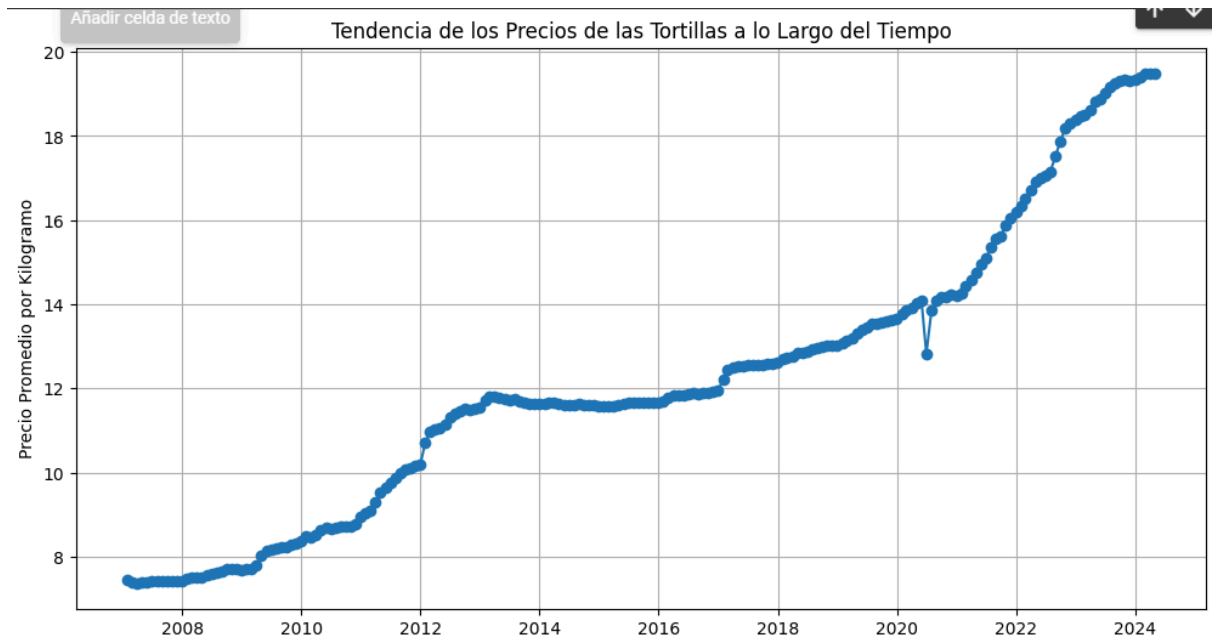
plt.figure(figsize=(13, 6))
price_heatmap = price_df_cleaned.pivot(index='Month', columns='Year', values='Price per kilogram')
sns.heatmap(price_heatmap, fmt=".2f", annot=True, cmap='Blues')
plt.title('Mapa de Calor de Precios de Tortillas')
plt.xlabel('Año')
plt.ylabel('Mes')
plt.show()
```



10. Utilizando un gráfico de líneas podemos analizar cómo cambian los precios promedio mensuales de las tortillas en México a lo largo del tiempo y luego muestra estos cambios en un gráfico. Primero, calcula el precio promedio de las tortillas para cada mes. Luego, representa estos precios en un gráfico donde cada punto muestra el precio promedio de las tortillas en un mes específico.

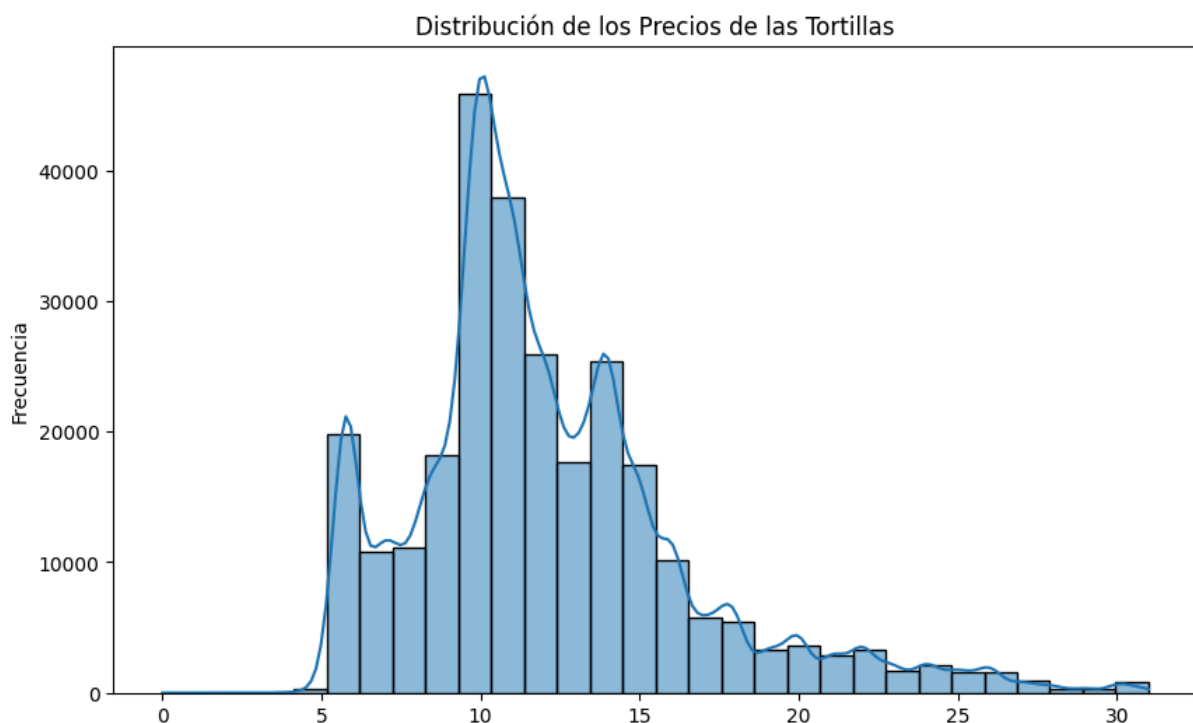
```
price_df['fecha'] = pd.to_datetime(price_df['fecha'])
monthly_avg = price_df.resample('M', on='fecha')['Price per kilogram'].mean()

plt.figure(figsize=(12, 6))
plt.plot(monthly_avg, marker='o', linestyle='--')
plt.xlabel('Fecha')
plt.ylabel('Precio Promedio por Kilogramo')
plt.title('Tendencia de los Precios de las Tortillas a lo Largo del Tiempo')
plt.grid(True)
plt.show()
```

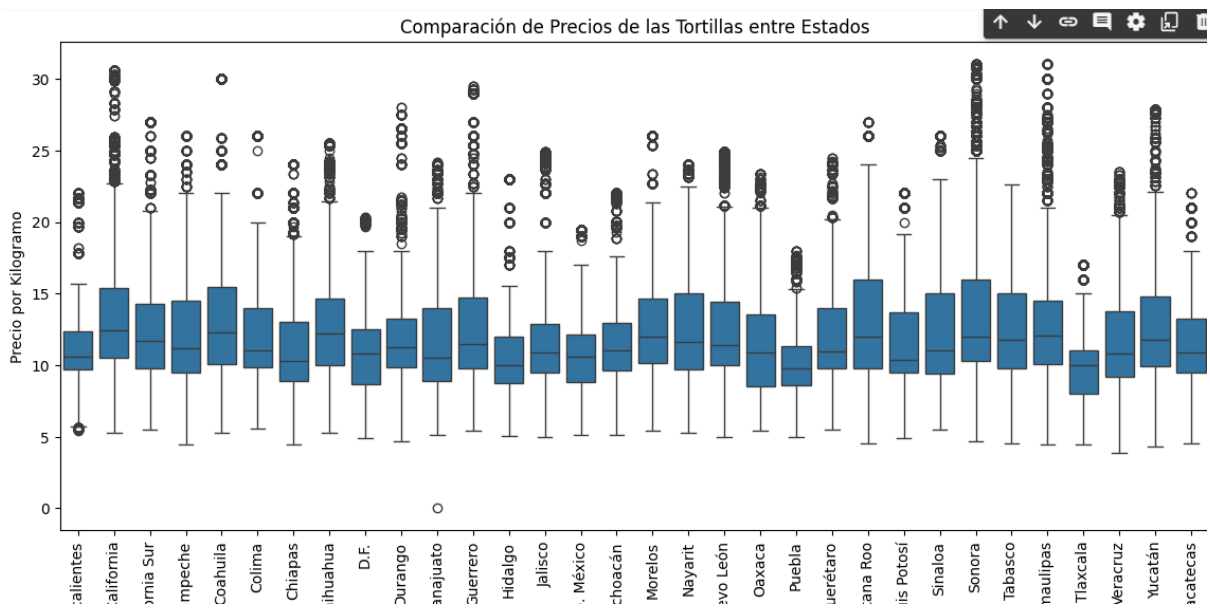


11. Utilizamos un gráfico llamado histograma que muestra cómo se distribuyen los precios por kilo de las tortillas. Divide los precios en grupos llamados bins y crea barras para representar cuántas veces ocurren diferentes rangos de precios. Además, incluye una línea que muestra la forma general de la distribución de los precios. El eje horizontal representa el precio por kilogramo y el eje vertical muestra la frecuencia de cada rango de precio.

```
plt.figure(figsize=(10, 6))
sns.histplot(price_df['Price per kilogram'], bins=30, kde=True)
plt.xlabel('Precio por Kilogramo')
plt.ylabel('Frecuencia')
plt.title('Distribución de los Precios de las Tortillas')
plt.show()
```

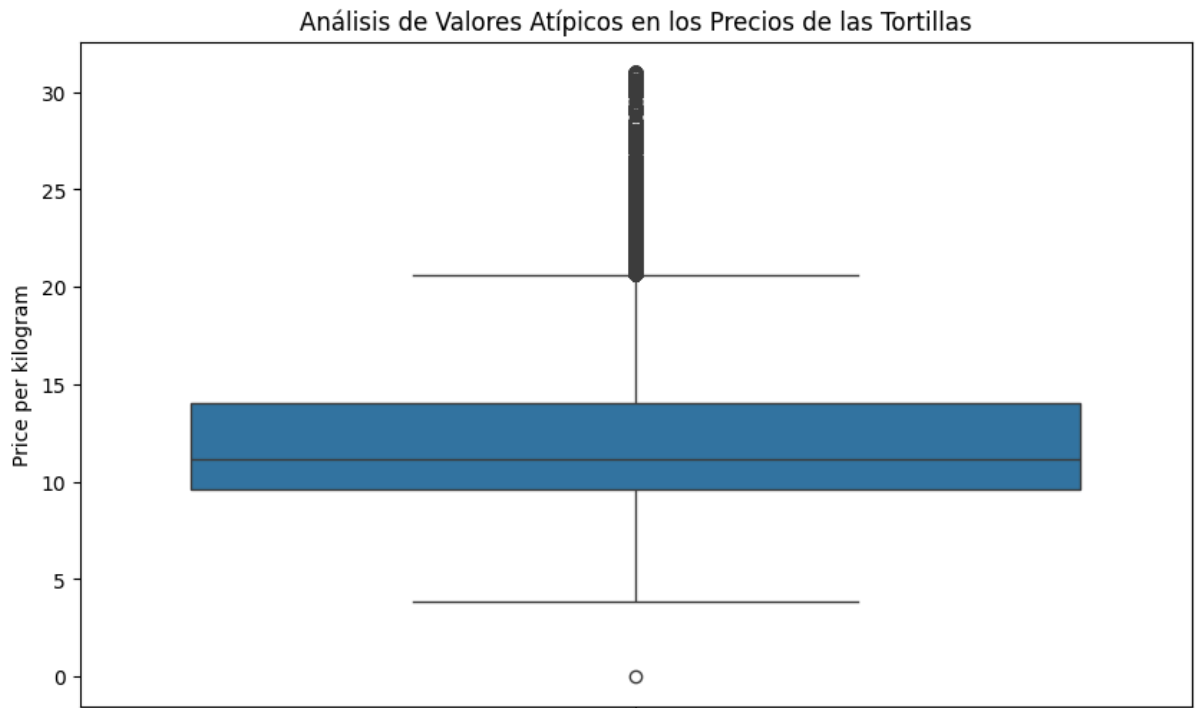


12. Utilizando un boxplot comparamos los precios por kilo de las tortillas entre distintos estados de México. Cada caja en el gráfico muestra la distribución de los precios en un estado específico, ayudando a identificar la mediana, los cuartiles y los valores atípicos.



13. Utilizamos un boxplot vertical que muestra cómo están distribuidos los precios por kilo de las tortillas y resalta cualquier valor atípico que pueda existir. Esto nos ayuda a identificar fácilmente los precios inusuales que pueden ser significativamente diferentes del resto de los precios

```
plt.figure(figsize=(10, 6))
sns.boxplot(y=price_df['Price per kilogram'])
plt.title('Análisis de Valores Atípicos en los Precios de las Tortillas')
plt.show()
```

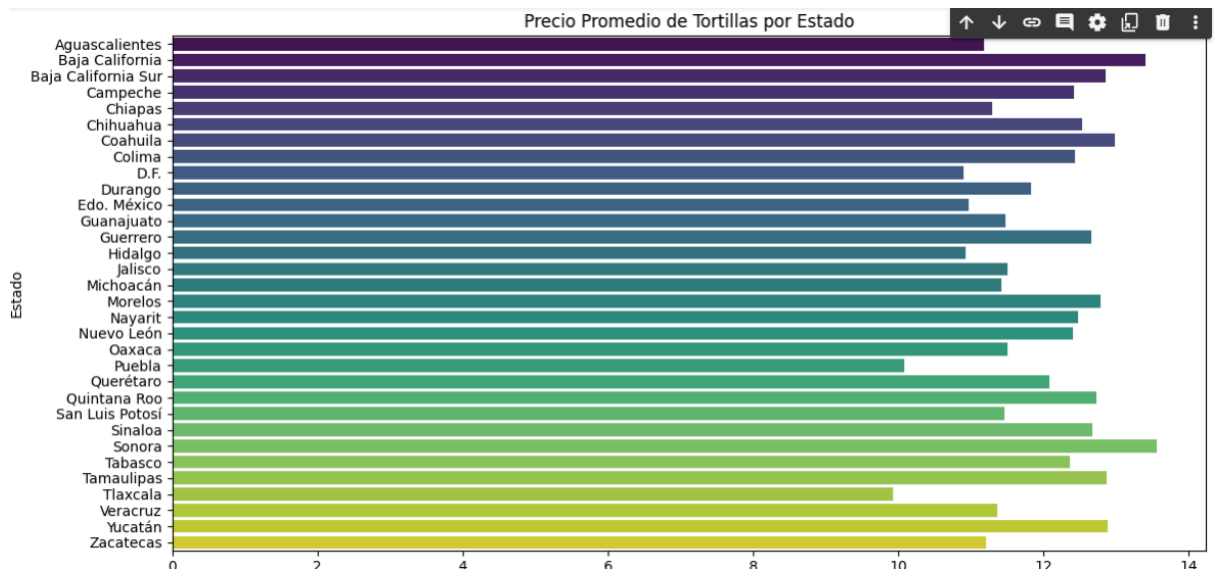
14. Realizamos una gráfica que calcula cuánto cuestan en promedio las tortillas por kilogramo en cada estado de México y muestra esta información en un gráfico de barras horizontal. Cada barra representa un estado y su longitud indica el precio promedio de las tortillas en ese estado.

```
promedio = price_df.groupby('State')['Price per kilogram'].mean().reset_index()

plt.figure(figsize=(14, 8))
sns.barplot(x='Price per kilogram', y='State', data=promedio, palette='viridis')

plt.xlabel('Precio Promedio por Kilogramo')
plt.ylabel('Estado')
plt.title('Precio Promedio de Tortillas por Estado')

plt.tight_layout()
plt.show()
```



15. Creamos un gráfico de pastel que nos muestra cuánto contribuye cada tipo de tienda a los precios de las tortillas en México. Cada porción del pastel representa un tipo de tienda, y su tamaño muestra el porcentaje del total de precios que corresponde a ese tipo de tienda.

```
def grafico_porcentaje_precios_tipo_tienda(df):
    plt.figure(figsize=(10, 8))

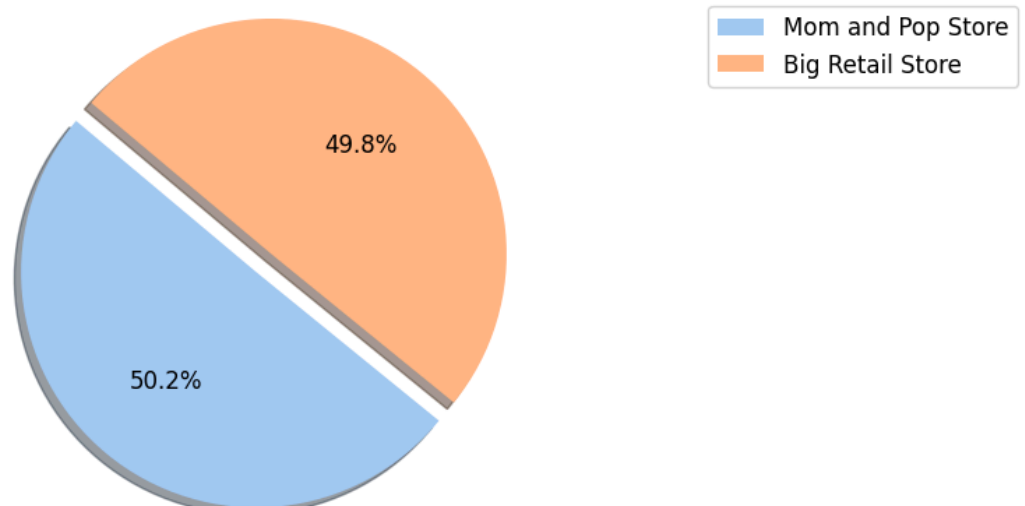
    tienda_tipo = df['Store type'].value_counts()
    explode = [0.1] + [0] * (len(tienda_tipo) - 1)

    tienda_tipo.plot(kind='pie', autopct='%1.1f%%', colors=sns.color_palette('pastel'),
                     labels=None, explode=explode, shadow=True, startangle=140, textprops={'fontsize': 12})
    plt.title('Porcentaje de Precios por Tipo de Tienda', fontsize=18, fontweight='bold')
    plt.axis('equal')
    plt.ylabel('')

    plt.legend(tienda_tipo.index, loc='upper left', bbox_to_anchor=(1, 1), fontsize=12)
    plt.text(0.5, -0.1, "Este gráfico de pastel muestra el porcentaje de precios por tipo de tienda.",
            ha='center', fontsize=12, transform=plt.gca().transAxes)
    plt.show()

grafico_porcentaje_precios_tipo_tienda(price_df)
```

Porcentaje de Precios por Tipo de Tienda



Este gráfico de pastel muestra el porcentaje de precios por tipo de tienda.

16. Se crea un ciclo que calcula el precio promedio de las tortillas para diferentes tipos de tiendas. Utiliza un for para recorrer cada tipo de tienda en los datos y calcular el precio promedio de las tortillas en cada uno. Luego imprime estos precios promedio, indicando cuánto cuesta en promedio un kilo de tortillas en cada tipo de tienda.

```
promedio_prices = {}

for store_type in price_df['Store type'].unique():
    store_prices = price_df[price_df['Store type'] == store_type]['Price per kilogram']

    mean_price = store_prices.mean()

    promedio_prices[store_type] = mean_price

for store_type, mean_price in promedio_prices.items():
    print(f"El precio promedio de las tortillas en tiendas de tipo '{store_type}' es: {mean_price:.2f}")
```

El precio promedio de las tortillas en tiendas de tipo 'Mom and Pop Store' es: 14.34
El precio promedio de las tortillas en tiendas de tipo 'Big Retail Store' es: 9.82

17. Utilizando un ciclo que analiza cómo cambian los precios promedio mensuales de las tortillas para distintos tipos de tiendas a lo largo del tiempo. Primero, crea una nueva columna que combina el año y el mes. Luego, calcula el precio promedio mensual de las tortillas para cada tipo de tienda y lo almacena en un diccionario. Después, muestra estos precios promedio mensuales en un gráfico, donde cada tipo de tienda tiene su propia línea.

```
price_df['Year-Month'] = price_df['fecha'].dt.to_period('M')

average_prices_by_month_and_store = {}

for store_type in price_df['Store type'].unique():
    store_prices = price_df[price_df['Store type'] == store_type]

    monthly_avg_prices = store_prices.groupby('Year-Month')['Price per kilogram'].mean()

    average_prices_by_month_and_store[store_type] = monthly_avg_prices

for store_type, monthly_avg_prices in average_prices_by_month_and_store.items():
    print(f"\nPrecio promedio mensual de las tortillas en tiendas de tipo '{store_type}':")
    print(monthly_avg_prices)
    monthly_avg_prices.plot(label=store_type)

plt.title('Precio promedio mensual de las tortillas por tipo de tienda')
plt.xlabel('Año')
plt.ylabel('Precio por kilogramo')
plt.legend(title='Tipo de tienda')
plt.show()
```

Precio promedio mensual de las tortillas en tiendas de tipo 'Mom and Pop Store':

Year-Month

2007-01	9.305598
2007-02	8.958901
2007-03	8.902242
2007-04	8.904966
2007-05	8.925956

...

2023-12	24.409938
2024-01	24.469344
2024-02	24.538737
2024-03	24.551128
2024-04	24.596748

Freq: M, Name: Price per kilogram, Length: 208, dtype: float64

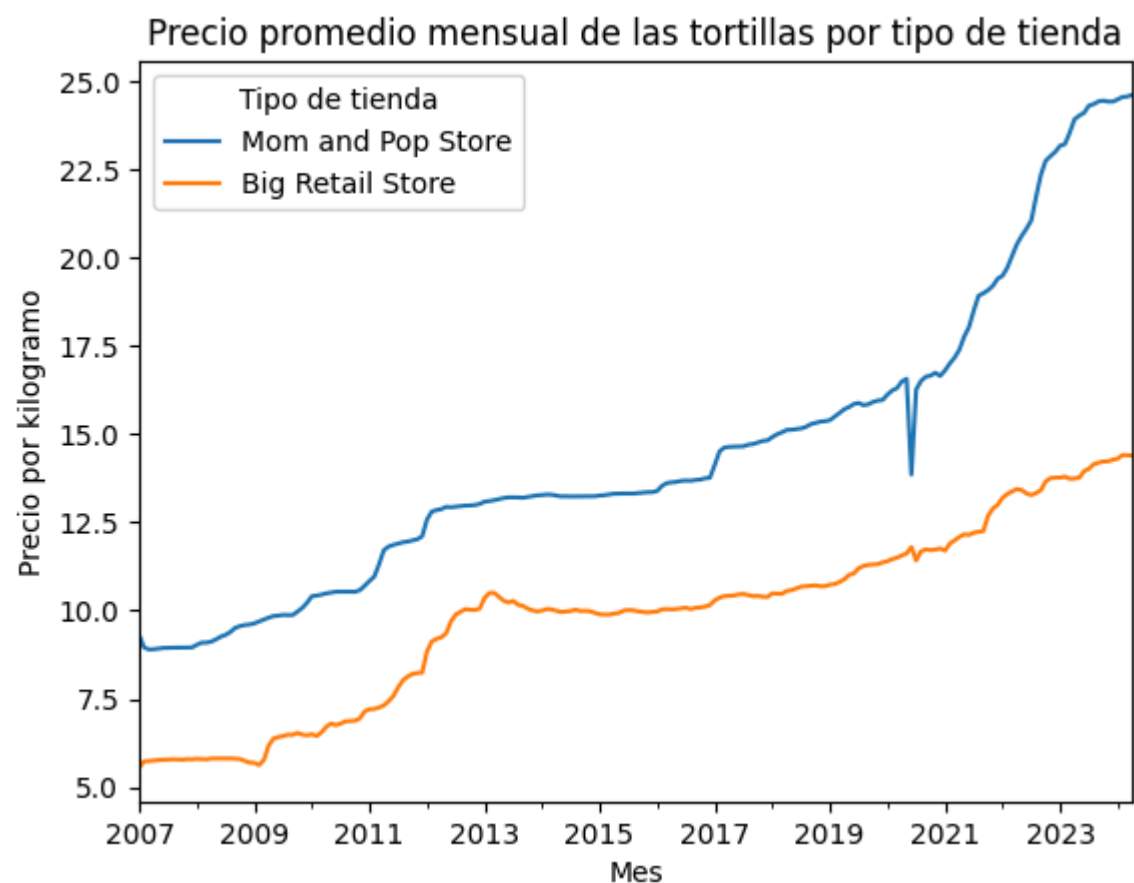
Precio promedio mensual de las tortillas en tiendas de tipo 'Big Retail Store':

Year-Month

2007-01	5.553962
2007-02	5.732565
2007-03	5.742023
2007-04	5.759045
2007-05	5.771009

...

2023-12	14.270046
2024-01	14.301155
2024-02	14.403552



18. Creamos una condición que divide los precios de las tortillas en tres grupos: Bajo, Medio y Alto, basándose en su posición en la distribución de precios. Primero, calcula los cuartiles de los precios. Luego, asigna cada precio a una categoría según si es menor que el primer cuartil (Bajo), está entre el primer y tercer cuartil (Medio), o es mayor que el tercer cuartil (Alto). Finalmente, muestra la cantidad de precios en cada categoría en un gráfico de barras.

```
q25 = price_df['Price per kilogram'].quantile(0.25)
q75 = price_df['Price per kilogram'].quantile(0.75)

def clasificar_precio(precio):
    if precio < q25:
        return 'Bajo'
    elif precio > q75:
        return 'Alto'
    else:
        return 'Medio'

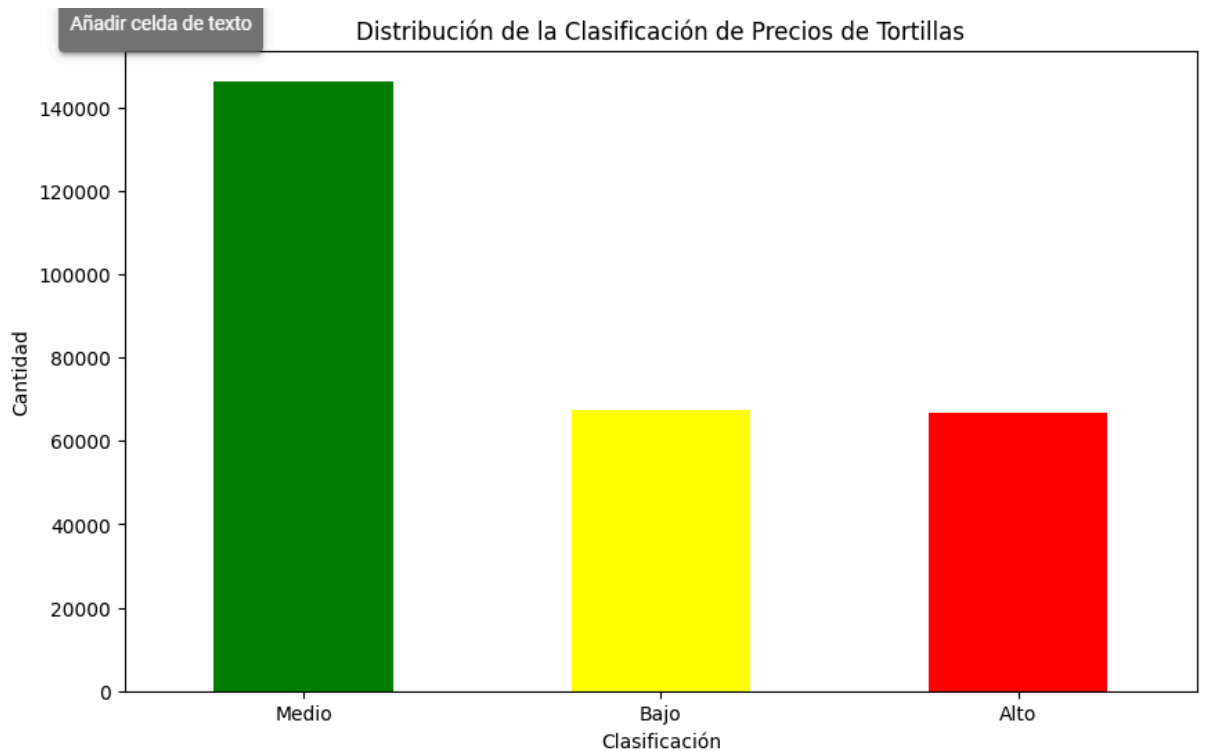
price_df['Clasificación Precio'] = price_df['Price per kilogram'].apply(clasificar_precio)

print(price_df.head())

clasificacion_counts = price_df['Clasificación Precio'].value_counts()
|
plt.figure(figsize=(10, 6))
clasificacion_counts.plot(kind='bar', color=['green', 'yellow', 'red'])
plt.title('Distribución de la Clasificación de Precios de Tortillas')
plt.xlabel('Clasificación')
plt.ylabel('Cantidad')
plt.xticks(rotation=0)
plt.show()
```

	State	City	Year	Month	Day	Store type \
0	Aguascalientes	Aguascalientes	2007	1	10	Mom and Pop Store
1	Baja California	Mexicali	2007	1	10	Mom and Pop Store
2	Baja California	Tijuana	2007	1	10	Mom and Pop Store
3	Baja California Sur	La Paz	2007	1	10	Mom and Pop Store
4	Campeche	Campeche	2007	1	10	Mom and Pop Store

	Price per kilogram	Clasificación Precio
0	9.9	Medio
1	NaN	Medio
2	10.0	Medio
3	10.0	Medio
4	10.0	Medio



Conclusión:

Los precios de las tortillas en México muestran variaciones significativas tanto según el tipo de tienda como por la ubicación geográfica. El análisis revela que distintos tipos de establecimientos tienen precios promedio diferentes por kilogramo de tortillas, y que estos precios varían de manera notable entre diferentes estados y ciudades del país. Esto sugiere que factores como la competencia local, la disponibilidad de materias primas y las políticas de precios regionales influyen en la fijación de precios de las tortillas en México.