

ESTUDIO DE MERCADO MUNDIAL

Sobre la tabla de datos

El conjunto de datos con el que empezamos es un registro histórico completo de los precios de las acciones de las marcas más reconocidas a nivel mundial, con actualizaciones diarias. La información abarca desde el 1 de enero del año 2000 hasta el 29 de noviembre de 2024, ofreciendo una línea de tiempo extensa sobre el comportamiento del mercado bursátil de diversas marcas globales.

Hipotesis

Hipótesis generales sobre tendencias del mercado:

1. Relación entre sectores y volatilidad:

Los sectores como tecnología y e-commerce presentan una mayor volatilidad en comparación con sectores tradicionales como finanzas o bienes de consumo.

1.1- Sectores de alta volatilidad

Sectores de alta volatilidad comparten características como dependencia de tendencias, factores externos impredecibles, y cambios rápidos en el comportamiento del consumidor. Aunque pueden generar rendimientos elevados en el corto plazo, también representan riesgos significativos para empresas e inversores que no se adapten rápidamente a las condiciones cambiantes. Estos sectores son atractivos para quienes buscan altos rendimientos asumiendo mayores riesgos. Requieren constante innovación y adaptabilidad para permanecer relevantes.

1.2- Sectores de baja volatilidad

Sectores de baja volatilidad se caracterizan por una demanda más estable y predecible, lo que los hace menos vulnerables a las fluctuaciones del mercado. Esto los convierte en opciones más seguras para inversores o empresas que buscan ingresos constantes y sostenidos. Ideales para estrategias de largo plazo y para mantener estabilidad financiera. La competencia puede ser intensa debido a la menor diferenciación y márgenes de ganancia más ajustados.

2. Distribución geográfica y desempeño:

Empresas basadas en países desarrollados (EE. UU., Europa) tienden a tener precios de acciones más altos en comparación con empresas de países en desarrollo.

El poder adquisitivo de las empresas y el poder económico de Estados Unidos se refuerzan mutuamente. El entorno favorable creado por el país impulsa el éxito empresarial, mientras que las empresas fortalecen la economía nacional. Este círculo virtuoso explica el liderazgo económico global de Estados Unidos. Al final, la ventaja económica de Estados Unidos es debida a una fuerte y solida red de beneficios mutuos entre estado y empresas.

3. Ofrecer dividendos significa mayor rendimiento?

A primera vista parecería ser que sí, pues son las empresas mas famosas y con mayor reputación las que los ofrecen. Sin embargo, solo el hecho de ofrecer dividendos parece resultar en una perdida de rendimiento a corto plazo. Mirándolo más de lejos, a largo plazo ofrecer dividendos sí que sale rentable si la empresa es lo suficientemente grande como para permitírselo.

Las empresas que no ofrecen dividendos pueden estar reinvertiendo sus beneficios en crecimiento, lo que podría reflejarse en un mayor rendimiento en términos de cambio porcentual en los precios.

Por otro lado, las empresas que ofrecen dividendos suelen ser más maduras, con un enfoque más conservador hacia el reparto de beneficios, lo cual podría limitar su crecimiento en el corto plazo.

Hipótesis macroeconómicas y externas:

1. Impacto de crisis financieras:

Los precios de las acciones de todos los sectores muestran caídas significativas durante períodos de recesión global (por ejemplo, 2008) pero con diferentes velocidades de recuperación.

Conclusiones

Tecnológicas estadounidenses:

- Estados Unidos ofrece un entorno único donde la relación estado-empresa fomenta la innovación tecnológica con incentivos fiscales y regulación favorable.
- La cultura emprendedora estadounidense facilita la creación de startups tecnológicas con acceso a capital de riesgo abundante y recursos especializados.
- La alta volatilidad del sector tecnológico genera oportunidades significativas de ingresos sin alcanzar niveles de especulación excesiva.
- Las universidades e instituciones de investigación en Estados Unidos atraen talento global y promueven avances tecnológicos de vanguardia.
- La integración de mercados globales permite a las empresas tecnológicas estadounidenses escalar rápidamente y dominar sectores clave.

- Escalabilidad rápida: Los productos digitales pueden expandirse globalmente sin los costos físicos asociados a otros sectores.
 - Altos márgenes de beneficio: Una vez desarrollados, los productos tecnológicos tienen bajos costos marginales de producción.
 - Acceso a datos masivos: La tecnología permite recopilar y analizar datos en tiempo real, mejorando la toma de decisiones y la personalización.
 - Innovación continua: La naturaleza del sector impulsa la creación constante de nuevos productos y servicios que generan ventajas competitivas.
 - Ecosistemas digitales: Las empresas tecnológicas crean plataformas que atraen usuarios, desarrolladores y empresas, fortaleciendo su posición de mercado.
-

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from pandas_profiling import ProfileReport
import warnings

warnings.filterwarnings("ignore")

C:\Users\Usuario\AppData\Local\Temp\ipykernel_2988\3019107045.py:5:
DeprecationWarning: `import pandas_profiling` is going to be
deprecated by April 1st. Please use `import ydata_profiling` instead.
  from pandas_profiling import ProfileReport

df = pd.read_csv(r"C:\Users\Usuario\Desktop\DataScience\THE_BRIDGE\
EDA\data\World-Stock-Prices-Dataset.csv")

columns_translation = {
    "Date": "Fecha",
    "Open": "Apertura",
    "High": "Máximo",
    "Low": "Mínimo",
    "Close": "Cierre",
    "Volume": "Volumen",
    "Brand_Name": "Nombre de Marca",
    "Ticker": "Ticker",
    "Industry_Tag": "Sector/Industria",
    "Country": "País",
    "Dividends": "Dividendos",
    "Stock Splits": "Divisiones de Acciones",
    "Capital Gains": "Ganancias de Capital",
}
df.rename(columns=columns_translation, inplace=True)
```

```

df["Fecha"] = pd.to_datetime(df["Fecha"], utc=True)
df.columns

Index(['Fecha', 'Apertura', 'Máximo', 'Mínimo', 'Cierre', 'Volumen',
      'Nombre de Marca', 'Ticker', 'Sector/Industria', 'País',
      'Dividendos',
      'Divisiones de Acciones', 'Ganancias de Capital'],
      dtype='object')

columns_to_round = ["Apertura", "Máximo", "Mínimo", "Cierre",
                    "Volumen"]
df[columns_to_round] = df[columns_to_round].apply(pd.to_numeric,
errors='coerce')

# Redondear estas columnas a 2 decimales
df[columns_to_round] = df[columns_to_round].round(2)

df.dtypes

Fecha                datetime64[ns, UTC]
Apertura             float64
Máximo              float64
Mínimo              float64
Cierre              float64
Volumen             float64
Nombre de Marca      object
Ticker              object
Sector/Industria     object
País                object
Dividendos          float64
Divisiones de Acciones float64
Ganancias de Capital float64
dtype: object

df.isna().sum()

Fecha                0
Apertura             0
Máximo              0
Mínimo              0
Cierre              0
Volumen             0
Nombre de Marca      0
Ticker              0
Sector/Industria     0
País                0
Dividendos          0
Divisiones de Acciones 0
Ganancias de Capital 299625
dtype: int64

```

```
df['Ganancias de Capital'].isna().sum()/len(df['Ganancias de Capital']) * 100
```

99.99933250341257

```
df.drop('Ganancias de Capital', axis = 1,inplace = True)
```

```
df.isna().sum()
```

Fecha	0
Apertura	0
Máximo	0
Mínimo	0
Cierre	0
Volumen	0
Nombre de Marca	0
Ticker	0
Sector/Industria	0
País	0
Dividendos	0
Divisiones de Acciones	0
dtype: int64	

df

	Fecha	Apertura	Máximo	Mínimo	Cierre	\
0	2024-11-29 05:00:00+00:00	9.86	10.39	9.76	10.34	
1	2024-11-29 05:00:00+00:00	968.09	973.27	962.47	971.88	
2	2024-11-29 05:00:00+00:00	205.83	208.17	204.59	207.89	
3	2024-11-29 05:00:00+00:00	234.84	237.81	233.97	237.33	
4	2024-11-29 05:00:00+00:00	78.20	79.18	78.04	78.77	
...	
299622	2000-01-03 05:00:00+00:00	16.69	16.76	15.95	16.27	
299623	2000-01-03 05:00:00+00:00	7.53	7.53	7.36	7.47	
299624	2000-01-03 05:00:00+00:00	35.79	36.84	34.63	36.26	
299625	2000-01-03 05:00:00+00:00	34.73	34.73	32.93	33.20	
299626	2000-01-03 05:00:00+00:00	5.73	5.79	5.43	5.48	

	Volumen	Nombre de Marca	Ticker	Sector/Industria
\				
0	11650855.0	peloton	PTON	fitness
1	1093161.0	costco	COST	retail
2	24760097.0	amazon	AMZN	e-commerce
3	28220021.0	apple	AAPL	technology
4	6629390.0	nike	NKE	apparel
...

299622	7384400.0	adobe	ADBE	technology
299623	490860.0	unilever	UL	consumer goods
299624	1734000.0	fedex	FDX	logistics
299625	6471267.0	american express	AXP	finance
299626	5825700.0	american eagle outfitters	AE0	apparel

	País	Dividendos	Divisiones de Acciones
0	usa	0.0	0.0
1	usa	0.0	0.0
2	usa	0.0	0.0
3	usa	0.0	0.0
4	usa	0.0	0.0
...
299622	usa	0.0	0.0
299623	netherlands	0.0	0.0
299624	usa	0.0	0.0
299625	usa	0.0	0.0
299626	usa	0.0	0.0

[299627 rows x 12 columns]

Vamos a intentar crear un dataframe con los valores medios de cada empresa para tener un dataframe sin la variable tiempo.

```
new_df = df[['Apertura', 'Máximo', 'Mínimo', 'Cierre', 'Volumen', 'Nombre de Marca', 'Dividendos', 'Divisiones de Acciones']]

df_grouped = new_df.groupby('Nombre de Marca',
as_index=False).median()
df_grouped.sort_values('Volumen', ascending=False)
```

	Nombre de Marca	Apertura	Máximo	Mínimo	Cierre	Volumen
Dividendos \						
8	apple	13.980	14.090	13.840	13.970	267192800.0
0.0						
4	amazon	11.390	11.620	11.280	11.420	96440000.0
0.0						
50	tesla	17.985	18.260	17.695	17.950	82067550.0
0.0						
20	google	28.040	28.250	27.800	28.030	57094848.0
0.0						
36	nvidia	4.480	4.550	4.420	4.490	54451400.0
0.0						
..

```

...
40      porsche      5.260    5.300    5.210    5.265      63100.0
0.0
1      adidas      49.080    49.460    48.940    49.220      30900.0
0.0
56      ubisoft      5.520    5.550    5.460    5.500       9500.0
0.0
10      bmw group    65.060    65.195    64.760    64.910        700.0
0.0
42      puma        45.225    45.390    44.960    45.000         0.0
0.0

```

```

Divisiones de Acciones
8      0.0
4      0.0
50     0.0
20     0.0
36     0.0
..     ..
40     0.0
1      0.0
56     0.0
10     0.0
42     0.0

```

```
[61 rows x 8 columns]
```

Este DataFrame se utilizará **únicamente** para obtener una visión **general** e introductoria de cada empresa. Al eliminar la dimensión temporal y calcular los valores promedio, se pierde información detallada y dinámica sobre cómo cambian las métricas de una empresa a lo largo del tiempo. Sin embargo, este enfoque permite obtener una representación general y **prototípica** de cada empresa, facilitando comparaciones rápidas y el análisis inicial sin distracciones por variaciones temporales. Nuestros análisis uni y multivariante los haremos sobre este último dataframe.

1. Analisis univariante

-categoricas

```

-frecuencias
-Moda
-Absoultos
-Relativos

```

-Numericas

```

-Tendencia Central (mediana, median)
-rangos
-posiciones

```

- dispersion
- distribucion

2. Analisis bivalente

-Categorica-categorica

- chi -2

-Categorica-numerica

- u mann whitney/t-student
- ANOVA

-Numerica-numerica

- Pearson

3. Analisis multivariante

A ojo

Restrospectiva, hipotesis verdaderas?

Descripción de las Variables

Columna/Variable	Descripción
Fecha	Fecha y horario correspondiente al turno de operación.
Apertura	Precio de la acción al inicio del turno.
Máximo	Precio máximo alcanzado durante el turno.
Mínimo	Precio mínimo alcanzado durante el turno.
Cierre	Último precio registrado al final del turno.
Volumen	Valor total de las acciones transaccionadas en ese turno.
Nombre de Marca	Nombre de la empresa o marca correspondiente.
Ticker	Identificador único o nombre de la empresa en la bolsa de valores.
Sector/Industria	Sector económico al que pertenece la empresa.
País	País de origen de la empresa o marca.
Dividendos	Rentas que reciben los inversores por poseer las acciones de la empresa.
Divisiones de Acciones	Número de divisiones de las acciones realizadas por la empresa.

Notas adicionales:

- Los valores como **Apertura**, **Máximo**, **Mínimo**, y **Cierre** permiten estudiar la volatilidad de cada empresa.
- **Volumen** es una métrica clave para entender el nivel de actividad y liquidez en el mercado.
- **Dividendos y Divisiones de Acciones** ofrecen insights sobre las estrategias de retorno y crecimiento de las empresas.

Análisis univariante.

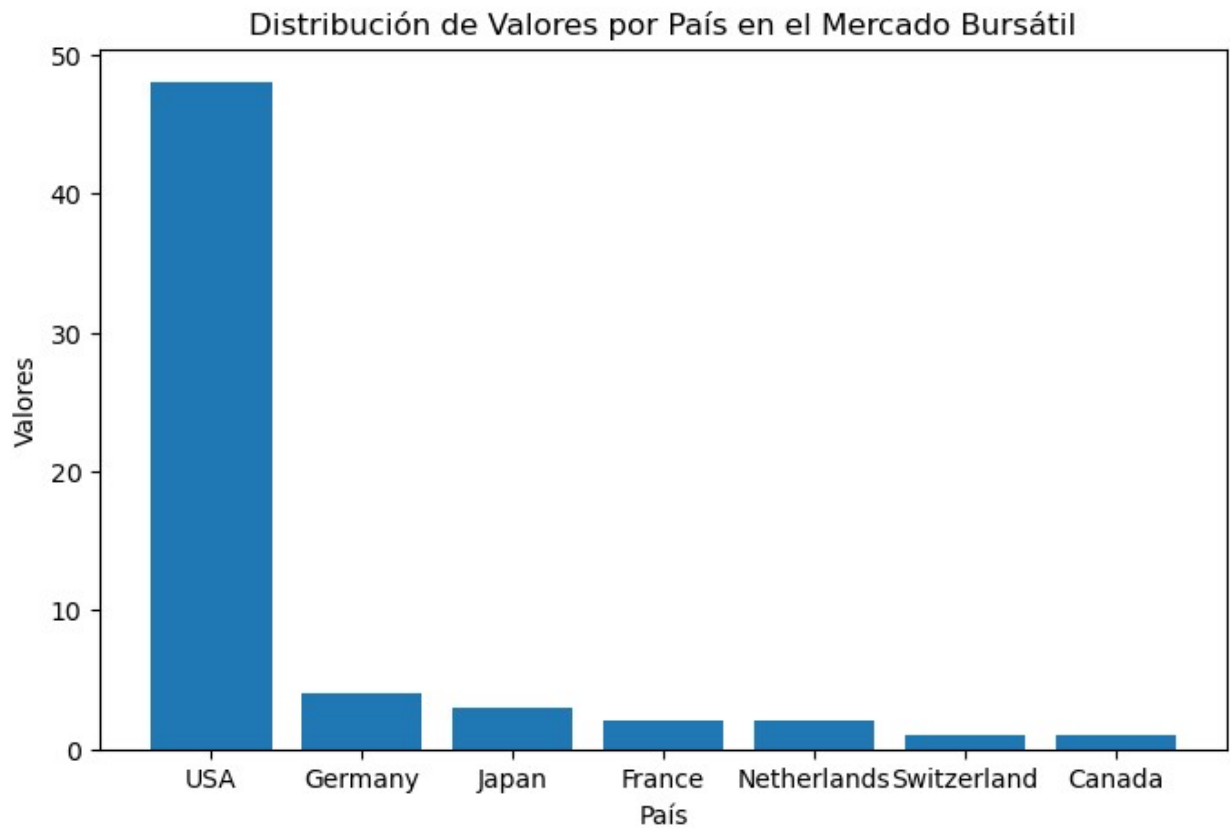
-Categoricas

Nombre de Marca, país y sector/Industria son las variables en las que nos centraremos

```
unique_companies = df.drop_duplicates(subset='Nombre de Marca')
country_counts = unique_companies['País'].value_counts()
country_counts
countries = ['USA', 'Germany', 'Japan', 'France', 'Netherlands',
'Switzerland', 'Canada']
values = [48, 4, 3, 2, 2, 1, 1]

plt.figure(figsize=(8, 5))
plt.bar(countries, values)
plt.title('Distribución de Valores por País en el Mercado Bursátil')
plt.xlabel('País')
plt.ylabel('Valores')

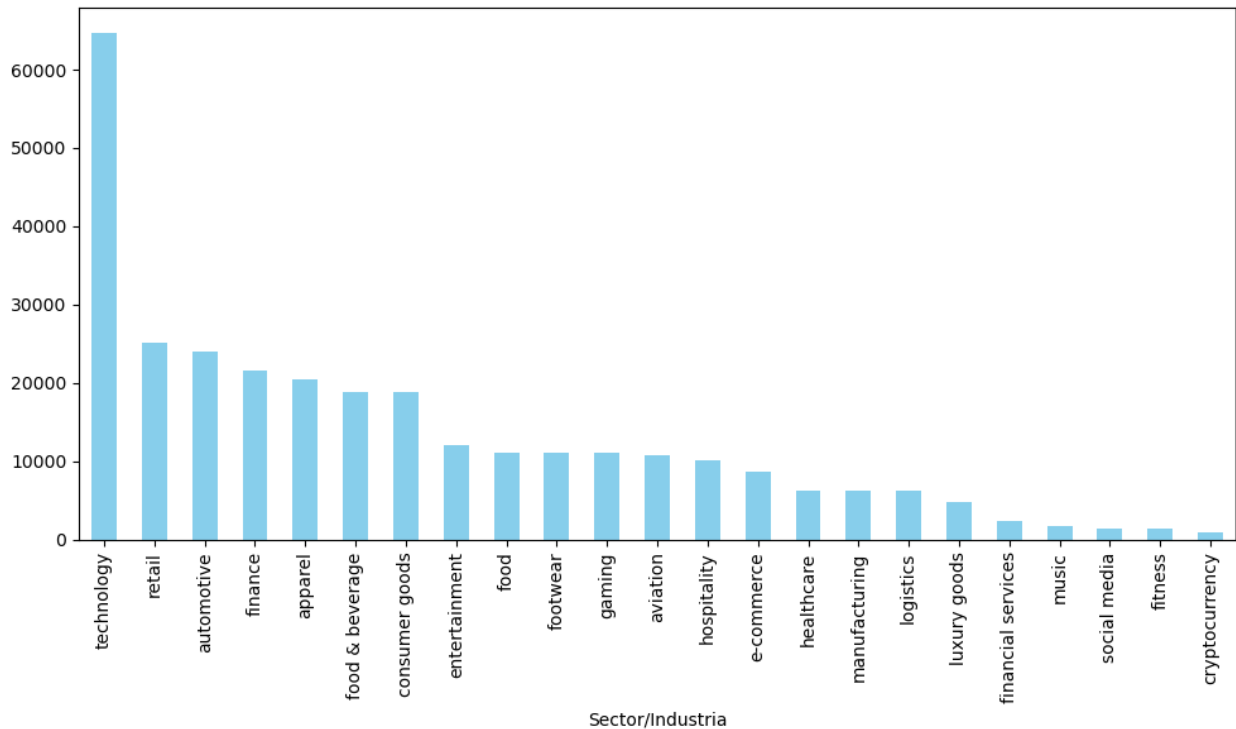
plt.show()
```



```
sector_counts = df['Sector/Industria'].value_counts()

plt.figure(figsize=(10, 6))
sector_counts.plot(kind='bar', color='skyblue')

plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```



```
pd.set_option('display.max_rows', None)
print(df_grouped['Nombre de Marca'])
```

```
0          3m
1      adidas
2      adobe
3      airbnb
4      amazon
5      amd
6  american eagle outfitters
7      american express
8      apple
9      block
10     bmw group
11     chipotle
12     cisco
13     coinbase
14  colgate palmolive
15     costco
16     crocs
17     delta air lines
18     fedex
19     foot locker
20     google
21  hershey company
22     hilton
23     honda
```

```
24      johnson & johnson
25      jpmorgan chase & co
26      logitech
27      lvmh
28      marriott
29      mastercard
30      mcdonald's
31      microsoft
32      netflix
33      nike
34      nintendo
35      nordstrom
36      nvidia
37      peloton
38      philips
39      pinterest
40      porsche
41      procter & gamble
42      puma
43      roblox
44      salesforce / slack
45      shopify
46      southwest airlines
47      spotify
48      starbucks
49      target
50      tesla
51      the coca-cola company
52      the home depot
53      the walt disney company
54      toyota
55      uber
56      ubisoft
57      unlever
58      visa
59      zoom video communications
60      zoominfo
Name: Nombre de Marca, dtype: object
```

En nuestra data table, la mayoría de empresas son de estados unidos, con abrumadora diferencia, y dedicadas al sector tecnologico, aunque la dispersion en cuanto a sector sí es mas variada.

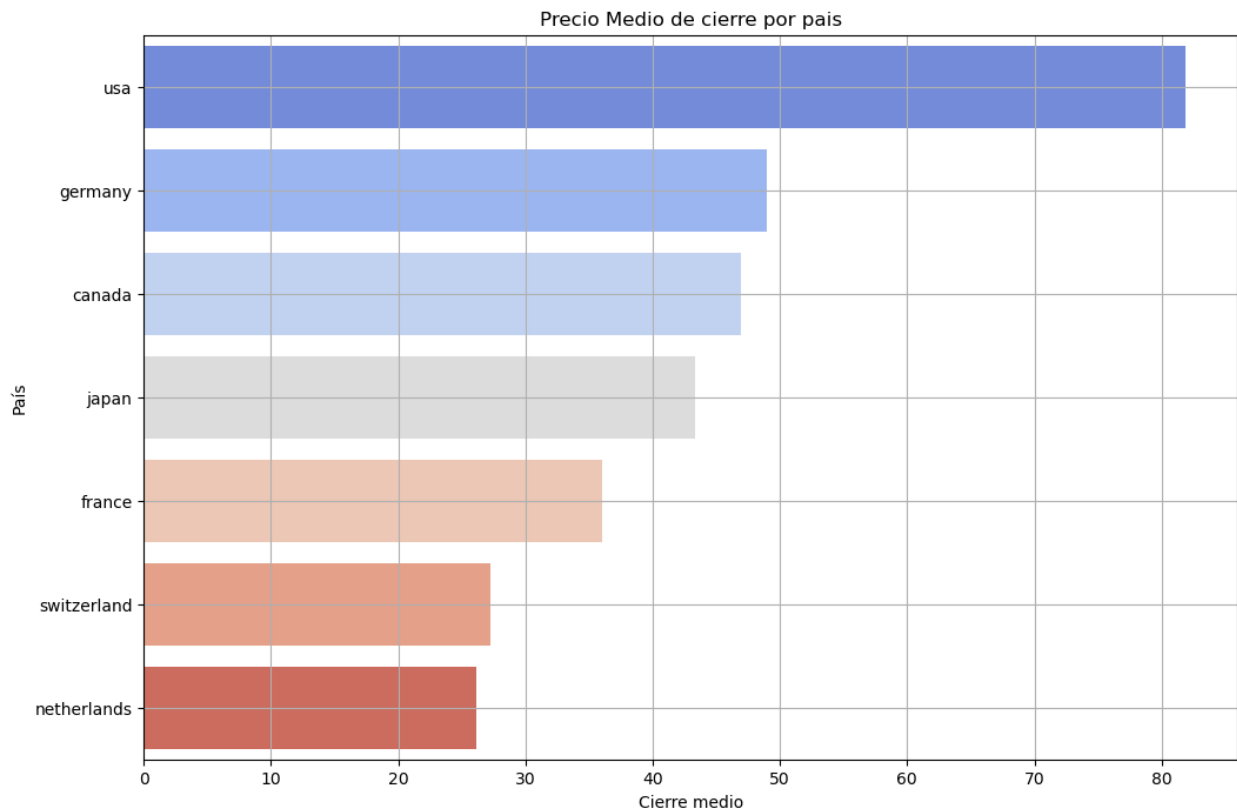
```
country_avg = df.groupby('País')
['Cierre'].mean().sort_values(ascending=False)

plt.figure(figsize=(12, 8))
```

```

sns.barplot(x=country_avg.values, y=country_avg.index, palette =
"coolwarm")
plt.title('Precio Medio de cierre por pais')
plt.xlabel('Cierre medio')
plt.ylabel('País')
plt.grid(True)
plt.show()

```

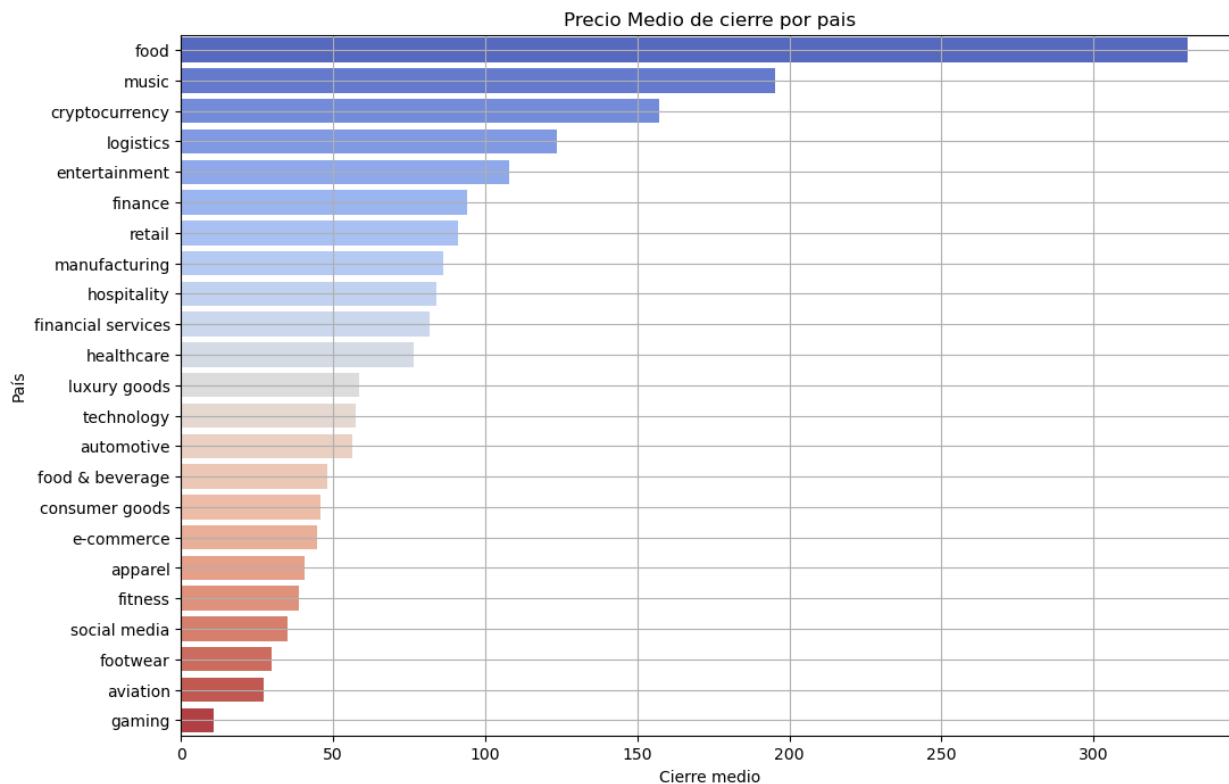


```

country_avg = df.groupby('Sector/Industria')
['Cierre'].mean().sort_values(ascending=False)

plt.figure(figsize=(12, 8))
sns.barplot(x=country_avg.values, y=country_avg.index, palette =
"coolwarm")
plt.title('Precio Medio de cierre por pais')
plt.xlabel('Cierre medio')
plt.ylabel('País')
plt.grid(True)
plt.show()

```



Hipotesis n.2

Distribución geográfica y desempeño:

Empresas basadas en países desarrollados (EE. UU., Europa) tienden a tener precios de acciones más altos en comparación con empresas de países en desarrollo.

Al calcular el precio medio, el hecho de que la gran mayoría de empresas de nuestra tabla sean de estados unidos solo logra una vision mas precisa del precio medio de los cierres de ese país. Con los demas países, sin embargo, tenemos menos información. En cualquier caso, parece que las empresas estadounidenses tienen un mayor poder adquisitivo.

Datos Económicos de los Países Más Ricos del Mundo (2024)

País	PIB Nominal (USD trillones)	Crecimiento Anual (%)	PIB per cápita (USD)	Población (millones)
Estados	28.78	2.8	86,601	332

País	PIB Nominal (USD trillones)	Crecimiento Anual (%)	PIB per cápita (USD)	Población (millones)
Unidos				
China	19.05	4.8	26,310	1,400
Japón	6.57	0.3	53,059	124
Alemania	5.87	0.0	69,845	83
India	4.07	7.0	3,278	1,400
Reino Unido	3.68	1.1	55,888	68
Francia	3.37	1.1	51,410	65
Italia	2.13	0.7	35,243	60
Canadá	2.11	1.2	56,844	38
Brasil	2.05	3.0	9,813	213

Fuentes

- [World Population Review](#)
- [Forbes India](#)

Según world population review y forbes india, Estados Unidos es el país con mayor PIB, si eso es consecuencia de sus poderosas empresas o viceversa (es decir, que las empresas hayan podido crecer gracias al poder adquisitivo del país), lo discutimos a continuación.

¿El poder adquisitivo de Estados Unidos es debido al poder de sus empresas, o el poder de sus empresas se ha conseguido gracias al poder del país?

Las empresas estadounidenses lideran sectores como tecnología, farmacéutica, finanzas y manufactura. Compañías como Apple, Microsoft y Amazon han impulsado la economía al dominar mercados internacionales. Una clara ventaja sobre los demás países. La fortaleza de los mercados financieros, como Wall Street, permite que las empresas estadounidenses accedan a inversiones significativas para su crecimiento global. Lo cual significa una gran ventaja o atajo para las empresas afianzadas o pequeñas con potencial.

2. El poder adquisitivo de las empresas gracias al poder del país

El dólar, como moneda de reserva global, facilita el comercio y la expansión de las empresas estadounidenses. Las relaciones internacionales establecidas por el gobierno abren mercados

para las empresas del país. Universidades y centros de investigación líderes forman talento y generan innovación tecnológica, lo que beneficia directamente al sector empresarial.

Conclusión

El poder adquisitivo de las empresas y el poder económico de Estados Unidos se refuerzan mutuamente. El entorno favorable creado por el país impulsa el éxito empresarial, mientras que las empresas fortalecen la economía nacional. Este círculo virtuoso explica el liderazgo económico global de Estados Unidos. Al final, la ventaja económica de Estados Unidos es debida a una fuerte y solida red de beneficios mutuos entre estado y empresas.

Fuentes:

- [World Population Review](#)
- [Forbes India - World GDP Rankings 2024](#)

Análisis univariante.

-Numericas Tendencia Central (mediana, median) -rangos -posiciones -dispersion -distribucion -----

```
median_values = df_grouped[['Apertura', 'Máximo', 'Mínimo', 'Cierre',  
'Volumen']].median()  
print("Mediana de cada variable:\n", median_values)
```

```
Mediana de cada variable:  
Apertura      35.795  
Máximo        36.815  
Mínimo        35.170  
Cierre        35.525  
Volumen      4663700.000  
dtype: float64
```

Mediana de cada variable:

Apertura: 35.795

Máximo: 36.815

Mínimo: 35.170

Cierre: 35.525

Volumen: 4,663,700

Los precios de Apertura y Cierre están bastante cerca, lo que sugiere que la mayoría de los movimientos de estas empresas no son extremadamente volátiles. La mediana de Volumen es 4,663,700, lo que indica un volumen de negociación medio por acción en este conjunto de datos. Este valor es importante porque refleja el nivel de actividad en el mercado y puede influir en la liquidez de las acciones.

```
range_values = df_grouped[['Apertura', 'Máximo', 'Mínimo', 'Cierre',  
'Volumen']].max() - df[['Apertura', 'Máximo', 'Mínimo', 'Cierre',  
'Volumen']].min()  
print("Rango de cada variable:\n", range_values)
```

```
Rango de cada variable:  
Apertura      4.148850e+02  
Máximo        4.192850e+02  
Mínimo        4.117950e+02  
Cierre        4.150600e+02  
Volumen       2.671928e+08  
dtype: float64
```

Rango de cada variable:

Apertura: 414.885

Máximo: 419.285

Mínimo: 411.795

Cierre: 415.060

Volumen: 267,192,800

Los rangos de los precios son bastante amplios, lo que sugiere que las fluctuaciones de precio entre el mínimo y el máximo de las acciones pueden ser bastante significativas en este conjunto de datos. La diferencia entre los valores de Apertura y Cierre es notablemente pequeña, lo que sugiere que el mercado no tuvo grandes cambios entre el inicio y el cierre en promedio, aunque aún puede haber fluctuaciones intradía.

El rango del Volumen es 267,192,800, que es considerablemente grande. Este valor refleja la diferencia entre el volumen de economía entre las diferentes empresas de la data table.

```
percentiles = df_grouped[['Apertura', 'Máximo', 'Mínimo', 'Cierre',  
'Volumen']].quantile([0.25, 0.5, 0.75])  
print("Percentiles (25%, 50%, 75%) de cada variable:\n", percentiles)
```

```
Percentiles (25%, 50%, 75%) de cada variable:  
      Apertura Máximo Mínimo Cierre  Volumen  
0.25    21.250  21.740  20.89  21.230  1464200.0  
0.50    35.795  36.815  35.17  35.525  4663700.0  
0.75    58.720  59.480  57.84  58.800  9825400.0
```

```
std_devs = df_grouped[['Apertura', 'Máximo', 'Mínimo', 'Cierre',  
'Volumen']].std()  
variances = df_grouped[['Apertura', 'Máximo', 'Mínimo', 'Cierre',
```

```

'Volumen']].var()

print("Desviación estándar de cada variable:\n", std_devs)
print("Varianza de cada variable:\n", variances)

Desviación estándar de cada variable:
  Apertura      5.861505e+01
  Máximo        5.937016e+01
  Mínimo         5.803462e+01
  Cierre         5.859784e+01
  Volumen        3.782019e+07
dtype: float64
Varianza de cada variable:
  Apertura      3.435724e+03
  Máximo        3.524816e+03
  Mínimo         3.368017e+03
  Cierre         3.433707e+03
  Volumen        1.430366e+15
dtype: float64

skewness = df_grouped[['Apertura', 'Máximo', 'Mínimo', 'Cierre',
'Volumen']].skew()
print("Asimetría (skewness) de cada variable:\n", skewness)

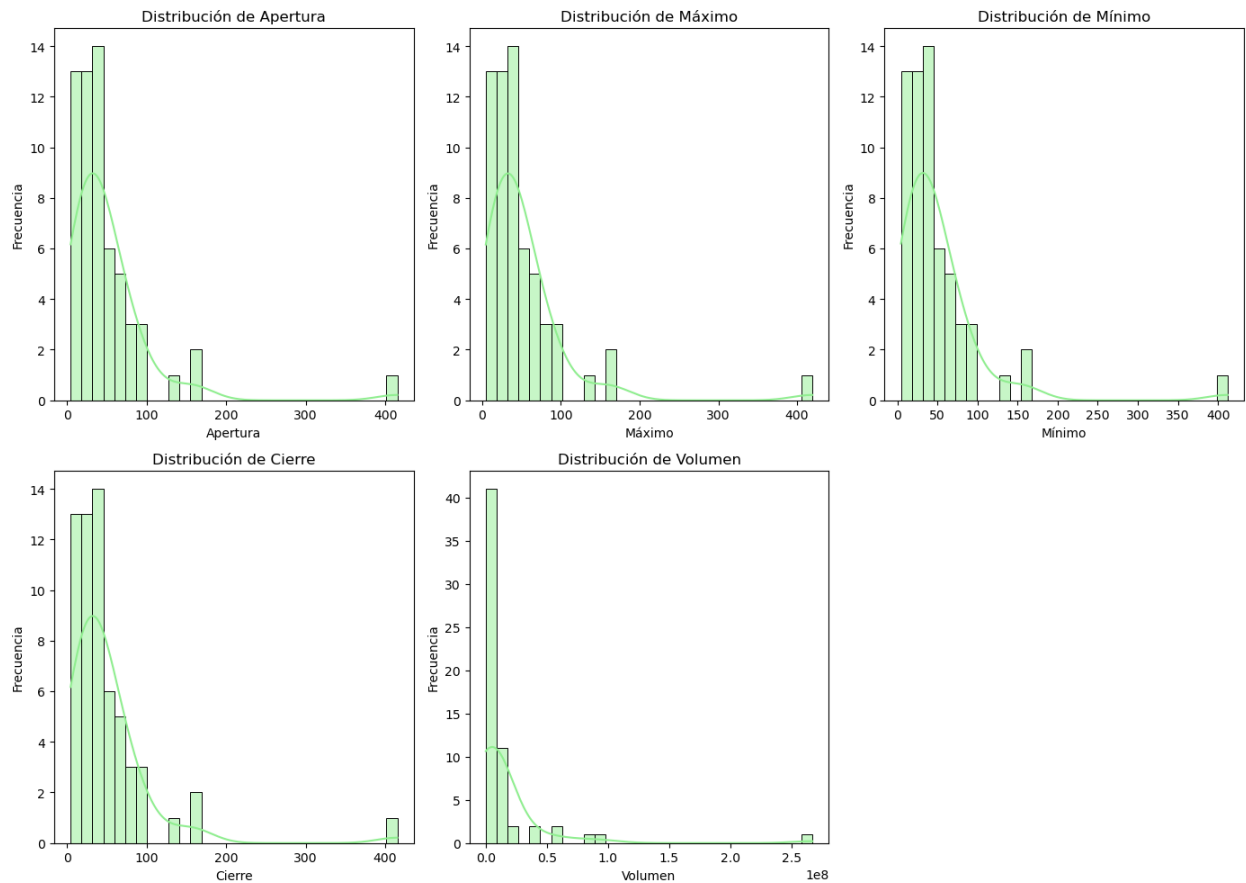
Asimetría (skewness) de cada variable:
  Apertura      4.397439
  Máximo        4.379216
  Mínimo         4.425990
  Cierre         4.404925
  Volumen        5.393190
dtype: float64

plt.figure(figsize=(14, 10))

for i, column in enumerate(['Apertura', 'Máximo', 'Mínimo', 'Cierre',
'Volumen'], 1):
    plt.subplot(2, 3, i)
    sns.histplot(df_grouped[column], kde=True, color='lightgreen',
bins=30)
    plt.title(f'Distribución de {column}')
    plt.xlabel(column)
    plt.ylabel('Frecuencia')

plt.tight_layout()
plt.show()

```



Evidentemente, a cuanto mas ganancia, menos frecuencia. Lo que indica que lo mas común es ganar menos. Desde esta información, podemos tirar del hilo hasta descubrir que hace grandes a las empresas grandes. Vemos que todos los valores tienen casi exactamente la misma frecuencia menos el volumen, el qual es aún más extremo.

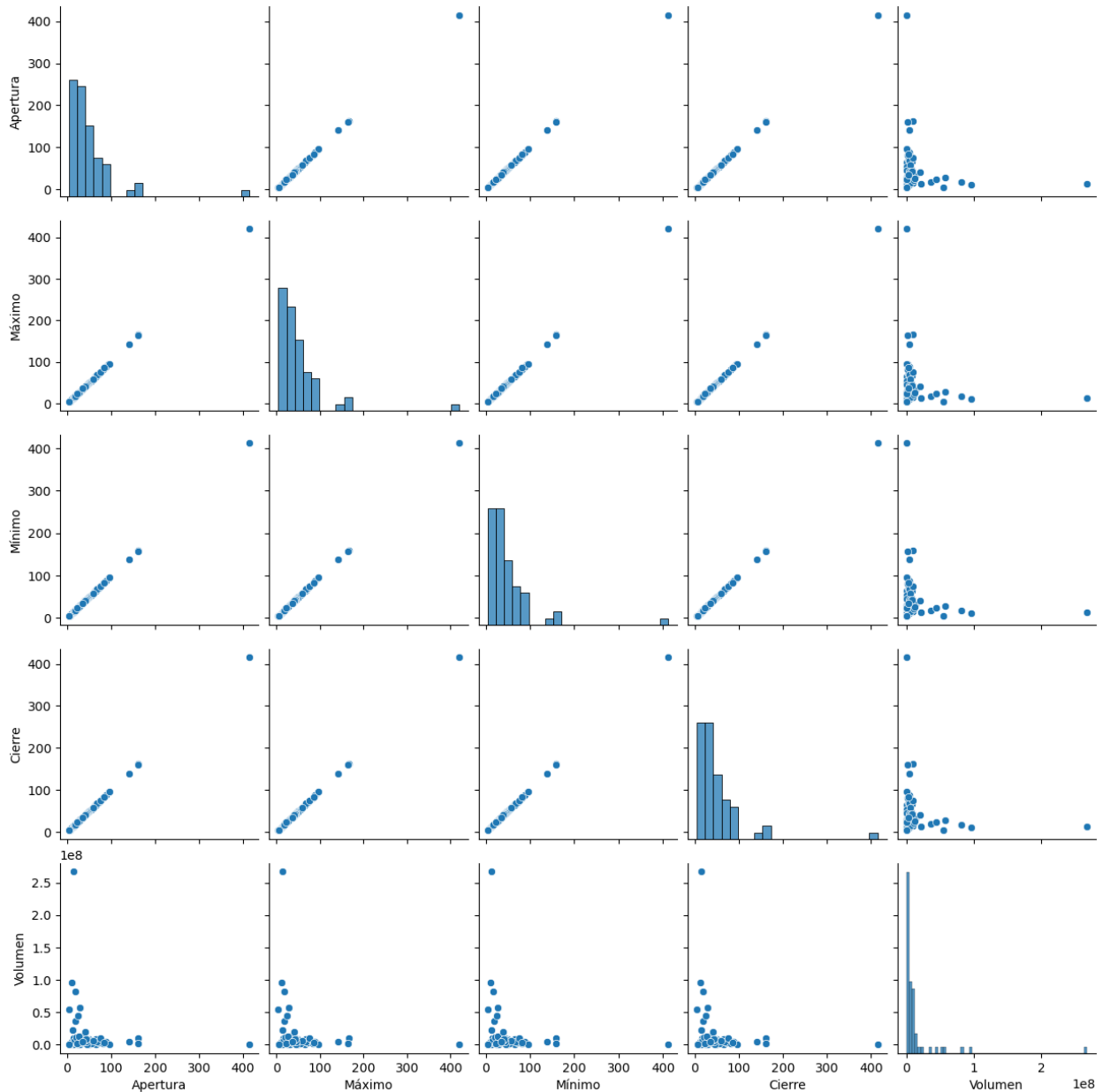
2. Analisis bivalente

-Numerica, numerica

-chi-2

```
columns = ['Apertura', 'Máximo', 'Mínimo', 'Cierre', 'Volumen']
df_num= df_grouped[columns]
sns.pairplot(df_num)
```

```
<seaborn.axisgrid.PairGrid at 0x11d6a58a510>
```



Todas las relaciones que no interactúan con volumen parecen tener una relación proporcionalmente directa positiva (por ejemplo, a cuanto mayor cierre, mayor apertura). En cuanto a cuando interactuamos con el volumen, en todas las ocasiones nos encontramos con una clara tendencia a 0, sin ninguna correlación evidente.

-Categorica, numerica

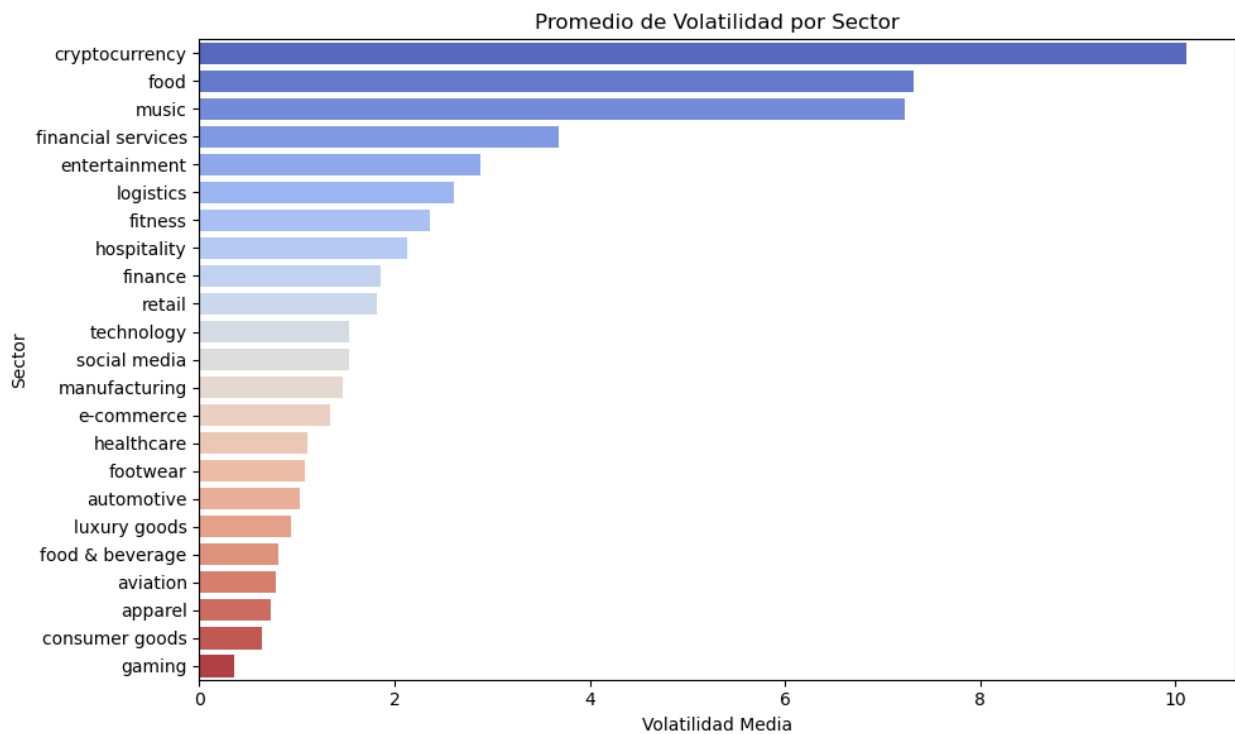
-Sector, volatilidad

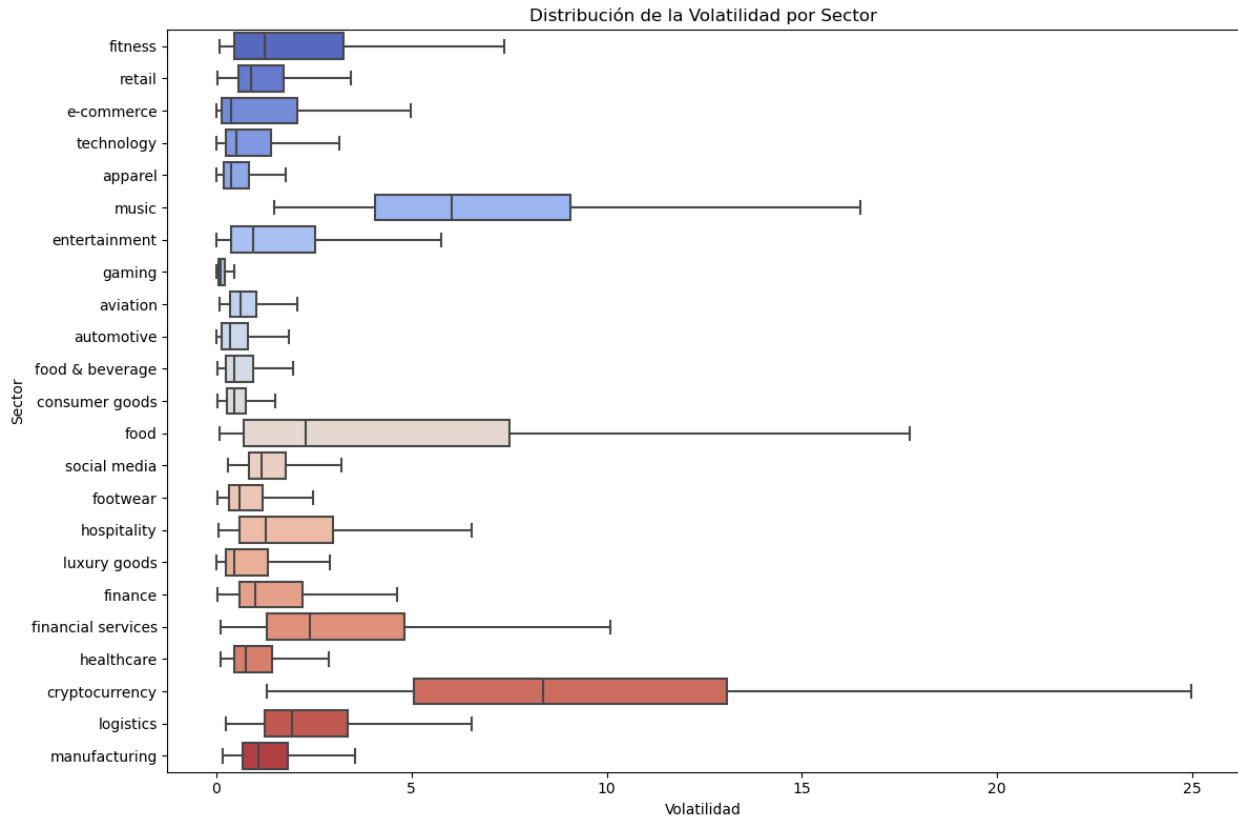
```
df['Volatility'] = df['Máximo'] - df['Mínimo']
```

```
sector_volatility = df.groupby('Sector/Industria')  
['Volatility'].mean().sort_values(ascending=False)
```

```
plt.figure(figsize=(10, 6))
sns.barplot(x=sector_volatility.values, y=sector_volatility.index,
palette="coolwarm")
plt.title('Promedio de Volatilidad por Sector')
plt.xlabel('Volatilidad Media')
plt.ylabel('Sector')
plt.tight_layout()
plt.show()
```

```
plt.figure(figsize=(12, 8))
sns.boxplot(data=df, x='Volatility', y='Sector/Industria',
palette="coolwarm", showfliers=False)
plt.title('Distribución de la Volatilidad por Sector')
plt.xlabel('Volatilidad')
plt.ylabel('Sector')
plt.tight_layout()
plt.show()
```





Mayor volatilidad:

- Criptomonedas
- Musica
- Comida

Menos volatilidad:

- Gaming
- Bienes de consumo
- Sector textil

Sectores de alta volatilidad

Sectores de alta volatilidad comparten características como dependencia de tendencias, factores externos impredecibles, y cambios rápidos en el comportamiento del consumidor. Aunque pueden generar rendimientos elevados en el corto plazo, también representan riesgos significativos para empresas e inversores que no se adapten rápidamente a las condiciones cambiantes. Estos sectores son atractivos para quienes buscan altos rendimientos asumiendo mayores riesgos. Requieren constante innovación y adaptabilidad para permanecer relevantes.

Sectores de baja volatilidad

Sectores de baja volatilidad se caracterizan por una demanda más estable y predecible, lo que los hace menos vulnerables a las fluctuaciones del mercado. Esto los convierte en opciones más

seguras para inversores o empresas que buscan ingresos constantes y sostenidos. Ideales para estrategias de largo plazo y para mantener estabilidad financiera. La competencia puede ser intensa debido a la menor diferenciación y márgenes de ganancia más ajustados.

Tener dividendos aumenta el rendimiento?

1. Definir métricas de rendimiento:

Utilizar el cambio porcentual entre el precio de apertura y el precio de cierre. Cambio_Porcentual

$$\text{Cambio_Porcentual} = \frac{\text{Cierre} - \text{Apertura}}{\text{Apertura}} \times 100$$

Otra opción es analizar el volumen de transacciones como indicador de interés del mercado.

1. Clasificar empresas:

Dividir las empresas en dos grupos: Con dividendos (donde la columna Dividendos > 0). Sin dividendos (donde la columna Dividendos = 0).

1. Comparar el rendimiento promedio: Calcular el promedio del cambio porcentual para ambos grupos. Analizar si hay diferencias significativas.

```
df['Cambio_Porcentual'] = ((df['Cierre'] - df['Apertura']) /
df['Apertura']) * 100

dividendos = df[df['Dividendos'] > 0]
sin_dividendos = df[df['Dividendos'] == 0]

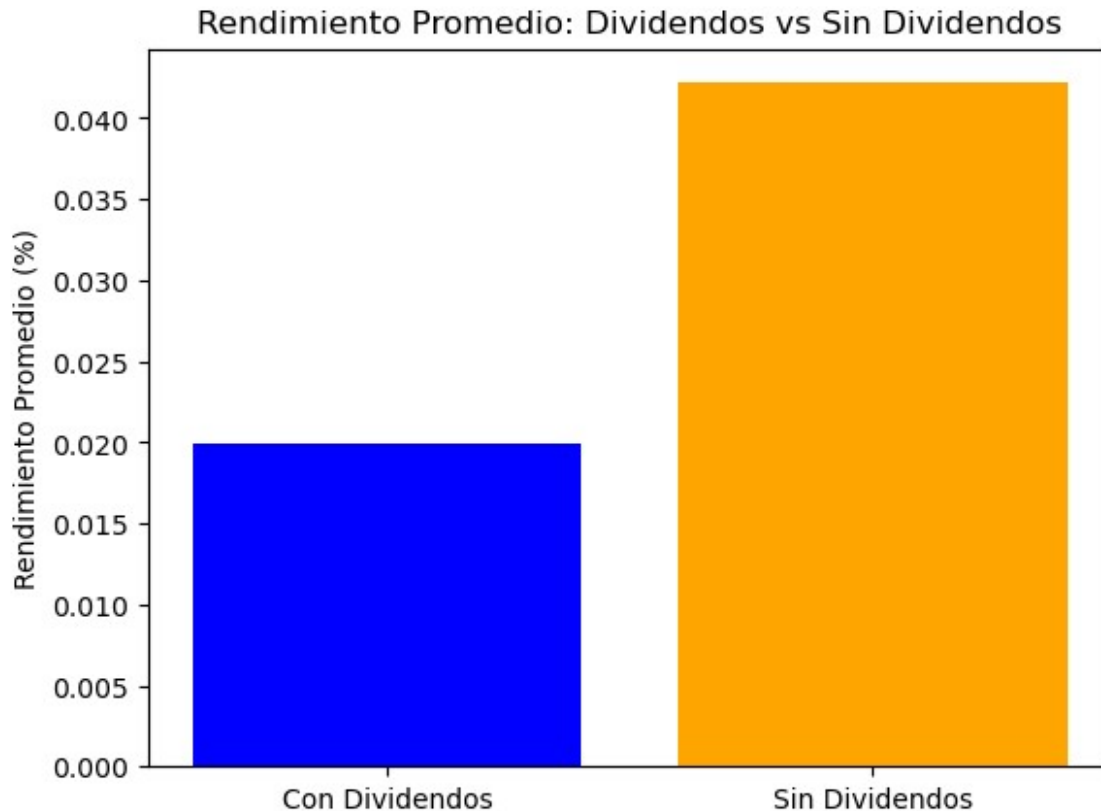
rendimiento_dividendos = dividendos['Cambio_Porcentual'].mean() if not
dividendos.empty else 'Sin datos'
rendimiento_sin_dividendos =
sin_dividendos['Cambio_Porcentual'].mean()

print("Rendimiento promedio (con dividendos):",
rendimiento_dividendos)
print("Rendimiento promedio (sin dividendos):",
rendimiento_sin_dividendos)

categorias = ['Con Dividendos', 'Sin Dividendos']
rendimientos = [rendimiento_dividendos, rendimiento_sin_dividendos]

plt.bar(categorias, rendimientos, color=['blue', 'orange'])
plt.ylabel('Rendimiento Promedio (%)')
plt.title('Rendimiento Promedio: Dividendos vs Sin Dividendos')
plt.show()

Rendimiento promedio (con dividendos): 0.019871880896079823
Rendimiento promedio (sin dividendos): 0.04218118119859028
```



Las empresas que no ofrecen dividendos pueden estar reinvertiendo sus beneficios en crecimiento, lo que podría reflejarse en un mayor rendimiento en términos de cambio porcentual en los precios.

Por otro lado, las empresas que ofrecen dividendos suelen ser más maduras, con un enfoque más conservador hacia el reparto de beneficios, lo cual podría limitar su crecimiento en el corto plazo.

1. Impacto de crisis financieras:

Los precios de las acciones de todos los sectores muestran caídas significativas durante períodos de recesión global (por ejemplo, 2008) pero con diferentes velocidades de recuperación.

```
crisis_data = df[(df['Fecha'] >= '2007-01-01') & (df['Fecha'] <= '2009-12-31')]  
  
sector_stats = crisis_data.groupby('Sector/Industria').agg({'Cierre':  
['min', 'max', 'mean'], 'Volumen': 'sum'})  
sector_stats.columns = ['Min Price', 'Max Price', 'Mean Price', 'Total  
Volume']
```



```

crisis_data['Price Drop (%)'] = (crisis_data['Máximo'] -
crisis_data['Mínimo']) / crisis_data['Máximo'] * 100
max_drops = crisis_data.groupby('Sector/Industria')['Price Drop
(%)'].max()

sector_analysis = pd.merge(sector_stats, max_drops,
on='Sector/Industria')
sector_analysis.rename(columns={'Price Drop (%)': 'Max Drop (%)'},
inplace=True)

plt.figure(figsize=(10, 6))
sector_analysis['Max Drop (%)'].sort_values().plot(kind='barh',
color='skyblue')
plt.title('Caída Máxima por Sector Durante la Crisis Financiera (2007-
2009)')
plt.xlabel('Caída Máxima (%)')
plt.ylabel('Sector')
plt.grid(axis='x', linestyle='--', alpha=0.7)
plt.show()

sectores_a_comparar = ['technology', 'retail', 'finance', 'apparel',
'e-commerce', 'food', 'manufacturing']
filtered_data =
crisis_data[crisis_data['Sector/Industria'].isin(sectores_a_comparar)]

avg_prices = filtered_data.groupby(['Fecha', 'Sector/Industria'])
['Cierre'].mean().unstack()
avg_prices.plot(figsize=(12, 8))
plt.title('Evolución del Precio Promedio por Sector Durante la Crisis
(2007-2009)')
plt.xlabel('Fecha')
plt.ylabel('Precio Promedio (Cierre)')
plt.legend(title='Sector')
plt.grid(alpha=0.5)
plt.show()

df['Año'] = df['Fecha'].dt.year

pre_crisis = df[df['Año'] == 2006]
post_crisis = df[df['Año'] == 2010]

pre_crisis_avg = pre_crisis.groupby('Sector/Industria')
['Cierre'].mean().rename("Cierre Promedio 2006")
post_crisis_avg = post_crisis.groupby('Sector/Industria')

```

```

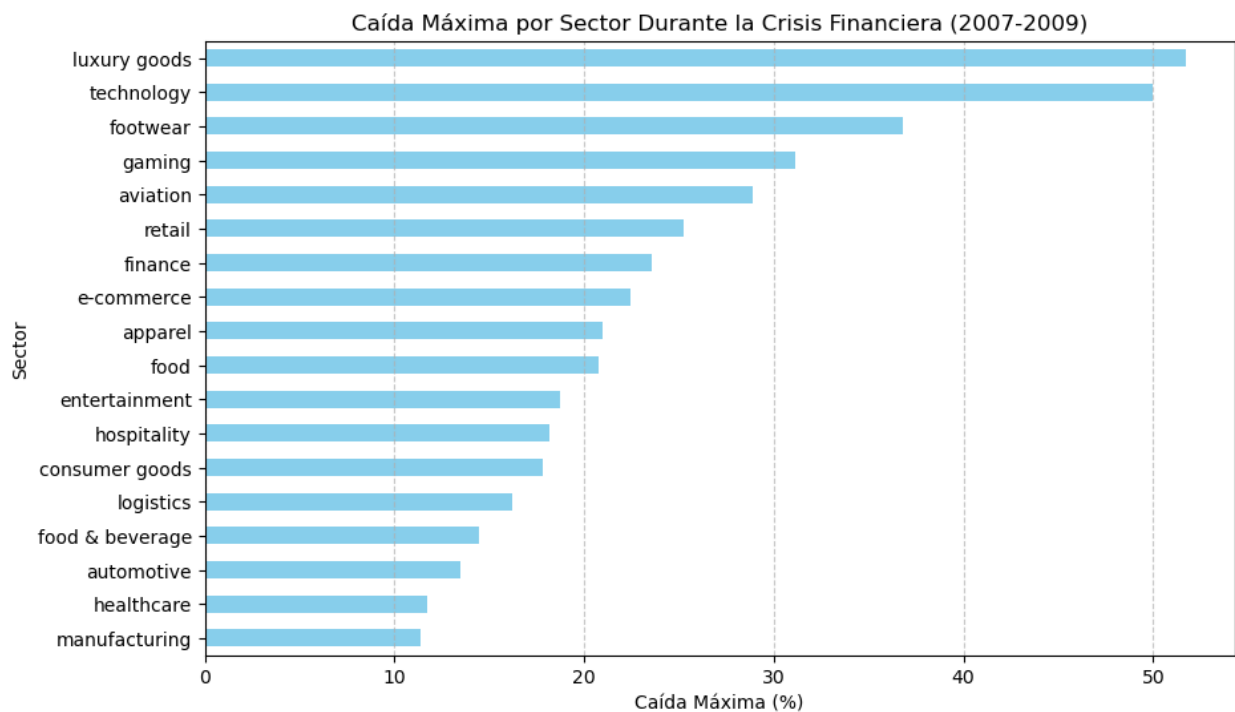
['Cierre'].mean().rename("Cierre Promedio 2010")

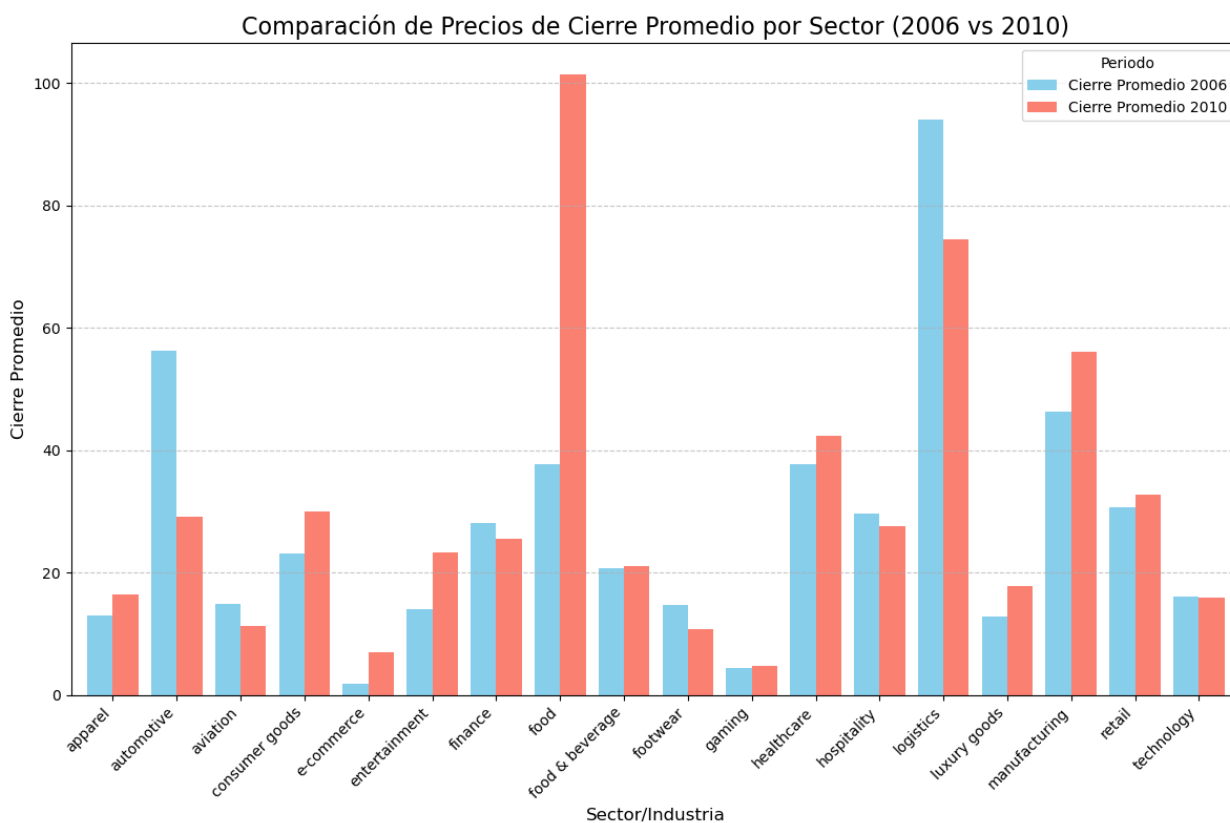
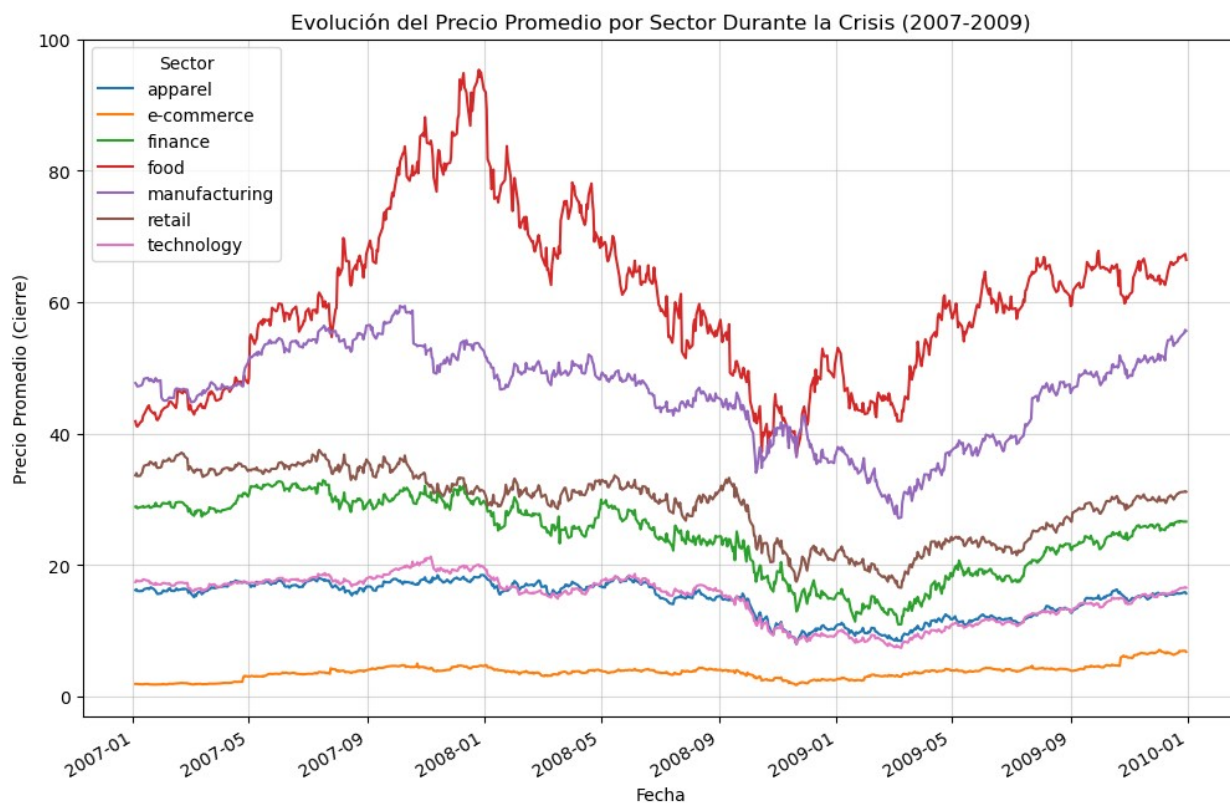
comparison = pd.concat([pre_crisis_avg, post_crisis_avg], axis=1)

comparison.plot(kind='bar', figsize=(12, 8), width=0.8,
color=['skyblue', 'salmon'])
plt.title('Comparación de Precios de Cierre Promedio por Sector (2006
vs 2010)', fontsize=16)
plt.xlabel('Sector/Industria', fontsize=12)
plt.ylabel('Cierre Promedio', fontsize=12)
plt.xticks(rotation=45, ha='right', fontsize=10)
plt.legend(title='Periodo', fontsize=10)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()

plt.show()

```





```

food_2006 = pre_crisis[pre_crisis['Sector/Industria'] == 'food']
food_2010 = post_crisis[post_crisis['Sector/Industria'] == 'food']

# Estadísticas descriptivas para el sector
print("Estadísticas del sector de comida (2006):")
print(food_2006['Cierre'].describe())

print("\nEstadísticas del sector de comida (2010):")
print(food_2010['Cierre'].describe())

# Comparación gráfica para el sector de comida
plt.figure(figsize=(8, 5))
plt.boxplot([food_2006['Cierre'], food_2010['Cierre']],
labels=['2006', '2010'])
plt.title('Comparación de Precios de Cierre - Sector de Comida (2006 vs 2010)')
plt.ylabel('Cierre')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()

```

Estadísticas del sector de comida (2006):

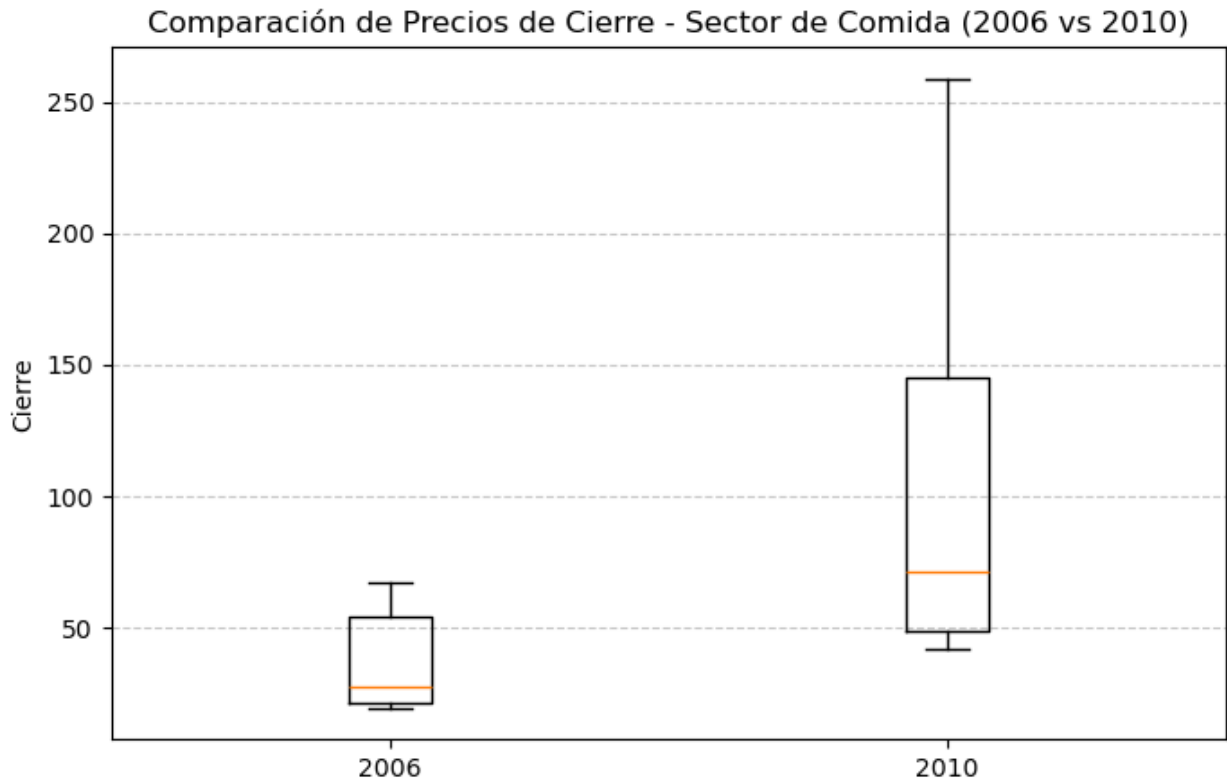
count	486.000000
mean	37.672058
std	16.240692
min	19.430000
25%	21.432500
50%	27.250000
75%	54.065000
max	67.050000

Name: Cierre, dtype: float64

Estadísticas del sector de comida (2010):

count	504.000000
mean	101.432738
std	61.106633
min	41.860000
25%	48.520000
50%	71.470000
75%	144.812500
max	258.820000

Name: Cierre, dtype: float64



Conclusión General

Impacto de la crisis financiera:

- La crisis económica global afectó a todos los sectores, pero el sector de comida demostró resiliencia debido a su naturaleza esencial como un bien de consumo. Sin embargo, las empresas dentro de este sector tuvieron que adaptarse rápidamente a las nuevas condiciones del mercado, incluyendo un enfoque en la diversificación de productos y estrategias de marketing para aumentar el valor percibido de sus productos. El cambio en los precios de cierre puede ser un reflejo de estos ajustes estratégicos y una respuesta a una mayor demanda de productos premium y a una recuperación económica generalizada.

Observaciones adicionales:

- Los sectores de alta volatilidad, como las criptomonedas y la música, suelen verse afectados por cambios rápidos en las tendencias y factores externos impredecibles, lo que contrasta con la estabilidad observada en sectores como el de bienes de consumo y el textil. Esto resalta la importancia de diversificar las inversiones y comprender las dinámicas de cada mercado para minimizar riesgos y aprovechar oportunidades en tiempos de cambio.

Primeras conclusiones conforme a valores generales

```
describe = df.describe()
describe
```

	Apertura	Máximo	Mínimo	Cierre \
count	299627.000000	299627.000000	299627.000000	299627.000000
mean	72.648175	73.470486	71.807613	72.657866
std	137.555866	139.025705	135.995219	137.536848
min	0.200000	0.200000	0.190000	0.200000
25%	15.620000	15.830000	15.400000	15.620000
50%	34.350000	34.760000	33.940000	34.380000
75%	79.360000	80.090000	78.600000	79.360000
max	3445.580000	3463.070000	3370.000000	3427.610000

	Volumen	Dividendos	Divisiones de Acciones
Volatility \			
count	2.996270e+05	299627.000000	299627.000000
mean	2.299714e+07	0.003600	0.000905
std	8.684302e+07	0.069757	0.117439
min	0.000000e+00	0.000000	0.000000
25%	1.361300e+06	0.000000	0.000000
50%	4.585400e+06	0.000000	0.000000
75%	1.216635e+07	0.000000	0.000000
max	7.421641e+09	15.000000	50.000000

	Cambio_Porcentual	Año
count	299627.000000	299627.000000
mean	0.041985	2013.349978
std	1.985232	7.072062
min	-84.414398	2000.000000
25%	-0.763359	2008.000000
50%	0.000000	2014.000000
75%	0.823954	2020.000000
max	39.446930	2024.000000

Volumen:

Alta desviación estándar: La desviación estándar del volumen (86.84 millones) es extremadamente alta, lo que sugiere que algunas empresas tienen un volumen de transacciones

mucho más alto que otras. Este comportamiento podría estar relacionado con la popularidad de las empresas, su tamaño o la volatilidad del mercado. Valores extremos: El volumen mínimo es 0, lo que podría ser un caso raro o un error, mientras que el volumen máximo es superior a 7.4 mil millones ($7.42e+09$), lo que podría estar asociado con empresas muy grandes y líquidas.

Dividendos

Promedio bajo: El promedio de dividendos es 0.0036, lo que sugiere que la mayoría de las empresas en este conjunto de datos no están pagando dividendos, o lo hacen en cantidades muy pequeñas. Desviación estándar baja: La baja desviación estándar (0.0698) en dividendos indica que las empresas que pagan dividendos lo hacen de forma bastante consistente. Sin embargo, dado que la mayoría de las empresas tienen dividendos cercanos a 0, la variabilidad no es tan relevante.

En general

Alta variabilidad en precios y volumen:

La considerable dispersión en los precios de las acciones y el volumen de transacciones sugiere que este dataset contiene tanto empresas de gran capitalización como empresas más pequeñas y menos líquidas. Las empresas con precios extremadamente altos probablemente sean más grandes o más estables.

La mayoría de las empresas no pagan dividendos:

La mayoría de las empresas no están pagando dividendos o lo hacen en cantidades muy pequeñas. Esto podría ser una señal de que muchas de las empresas en el dataset están en fases de crecimiento o reinversión, lo que es común en empresas tecnológicas o startups.

Divisiones de acciones limitadas:

Las divisiones de acciones parecen ser poco comunes, pero las que ocurren podrían indicar estrategias de accesibilidad o crecimiento de empresas que buscan atraer a más inversores.

```
profile = ProfileReport(df, title='World Stocks analysis')
profile.to_notebook_iframe()

{"model_id": "8f6783b1fa0e40f78876b3451db35905", "version_major": 2, "version_minor": 0}

{"model_id": "6236ad05b5514a2f99b2fd766c9a8cfc", "version_major": 2, "version_minor": 0}

{"model_id": "79e470787ae145ca9632eda21f7f4b6d", "version_major": 2, "version_minor": 0}

<IPython.core.display.HTML object>
```

