

```
import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
%matplotlib inline
from google.colab import files

uploaded = files.upload()
df = pd.read_csv('data_coffeeshop.csv', header='infer')
df.head(5)
```



Choose Files data\_coffeeshop.csv

- **data\_coffeeshop.csv**(text/csv) - 633903 bytes, last modified: 11/25/2024 - 100% done  
Saving data\_coffeeshop.csv to data\_coffeeshop.csv

	Year of Start	Current State	Size of Site	
0	2008.0	In	20.80	
1	2010.0	In	212.72	
2	2013.0	In	20.04	
3	2012.0	In	64.17	
4	2002.0	Out	11.40	

Next steps:

[Generate code with df](#)

[View recommended plots](#)

[New interactive sheet](#)

1) Genera una tabla de frecuencias para la variable 'yearOfStart' y visualízala por año

## Ordena por año

```
year_f = df['Year of Start'].value_counts().sort_index()
print("Tabla de frecuencias ordenada por año:")
print(year_f)
```



Tabla de frecuencias ordenada por año:

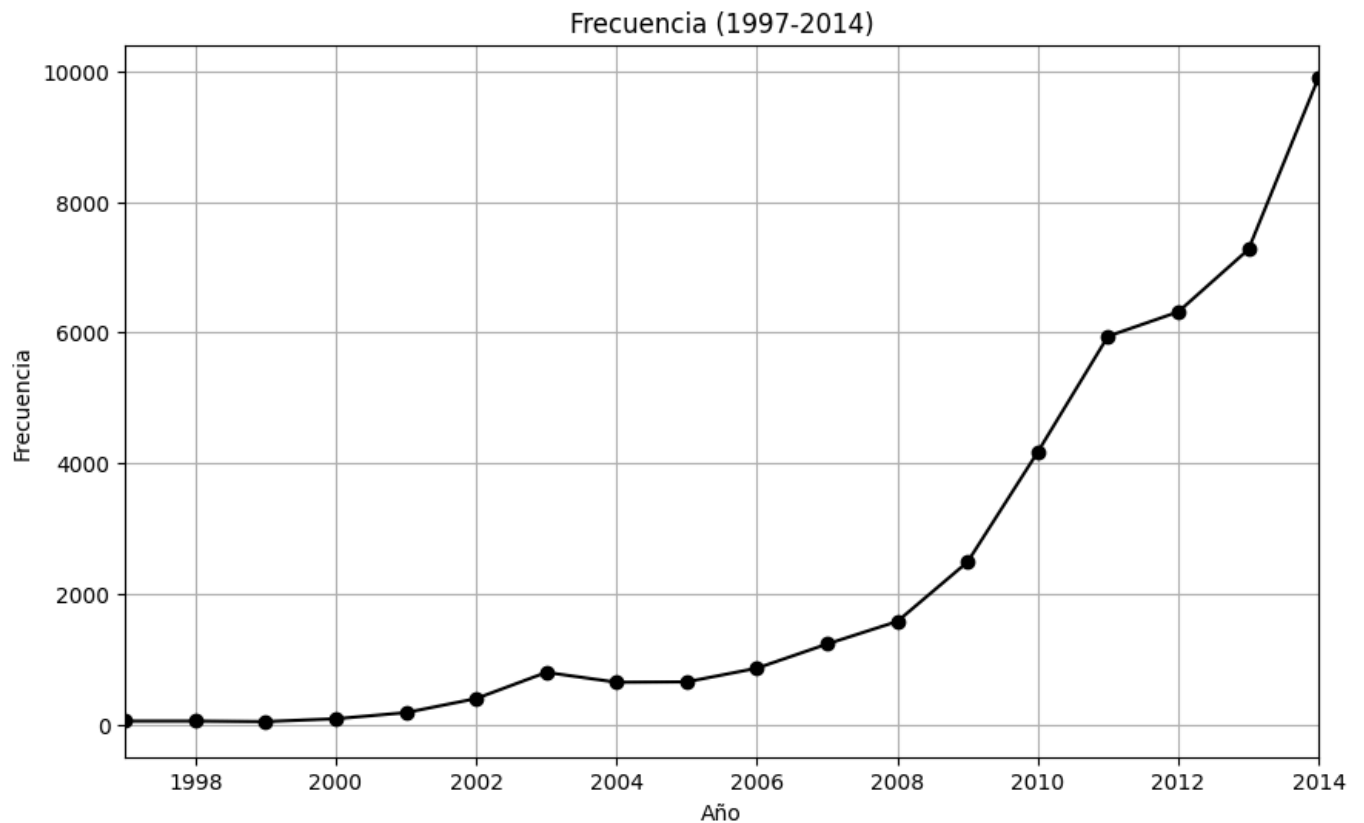
```
Year of Start
1964.0      2
1966.0      2
1967.0      3
1968.0      1
1969.0      2
1970.0      4
1971.0      6
1972.0      3
1974.0      1
1975.0      2
1976.0      5
1979.0      4
1980.0      9
1981.0      8
```

1982.0	12
1983.0	9
1984.0	11
1985.0	18
1986.0	21
1987.0	21
1988.0	26
1989.0	23
1990.0	25
1991.0	28
1992.0	37
1993.0	50
1994.0	48
1995.0	48
1996.0	41
1997.0	54
1998.0	54
1999.0	46
2000.0	89
2001.0	183
2002.0	398
2003.0	799
2004.0	648
2005.0	654
2006.0	863
2007.0	1233
2008.0	1579
2009.0	2489
2010.0	4172
2011.0	5942
2012.0	6315
2013.0	7270
2014.0	9905

Name: count, dtype: int64

## ✓ Dibuja una gráfica de línea de 1997 a 2014

```
plt.figure(figsize=(10, 6))
plt.plot(year_frequency.index, year_frequency.values, marker='o', color='black')
plt.xlim(1997, 2014)
plt.xlabel('Año')
plt.ylabel('Frecuencia')
plt.title('Frecuencia (1997-2014)')
plt.grid(True)
plt.show()
```



- 2) Divide los datos por el estado actual del negocio ('In' o 'Out') y visualiza la tendencia anual de las frecuencias

## Ordena por año

```
_in = df[df['Current State'] == 'In']
_out = df[df['Current State'] == 'Out']

ar_frequency_in = df_in['Year of Start'].value_counts().sort_index()
ar_frequency_out = df_out['Year of Start'].value_counts().sort_index()

t.figure(figsize=(10, 6))
t.plot(year_frequency_in.index, year_frequency_in.values, label='In', marker='o', color='blue')
t.plot(year_frequency_out.index, year_frequency_out.values, label='Out', marker='o', color='red')
t.xlim(1997, 2014)
t.xlabel('Año')
t.ylabel('Frecuencia')
t.title('Tendencia anual por estado (1997-2014)')
t.legend()
t.grid(True)
t.show()
```

