In [1]:
```
import pandas as pd
import numpy as np
import yfinance as yf
```

In [2]:
```
start = "2025-01-01"
```

In [3]:
```
overall = pd.read_html("https://en.wikipedia.org/wiki/List_of_S%26P_500_companies")[0]
```

In [4]: overall

Out[4]:

	Symbol	Security	GICS Sector	GICS Sub-Industry	Headquarters Location	Date added	CIK	Founded
0	MMM	3M	Industrials	Industrial Conglomerates	Saint Paul, Minnesota	1957-03-04	66740	1902
1	ACE	A. O. Smith	Industrials	Building Products	Milwaukee, Wisconsin	2017-07-26	91142	1916
2	AET	Abbott	Health Care	Health Care Equipment	North Chicago, Illinois	1957-03-04	1800	1888
3	ABBV	AbbVie	Health Care	Biotechnology	North Chicago, Illinois	2012-12-31	1551152	2013 (1888)
4	ACN	Accenture	Information Technology	IT Consulting & Other Services	Dublin, Ireland	2011-07-06	1467373	1989
...	...	...	...	...	...	...	...	...
498	XYL	Xylem Inc.	Industrials	Industrial Machinery & Supplies & Components	White Plains, New York	2011-11-01	1524472	2011
499	YUM	Yum! Brands	Consumer Discretionary	Restaurants	Louisville, Kentucky	1997-10-06	1041061	1997
500	ZBRA	Zebra Technologies	Information Technology	Electronic Equipment & Instruments	Lincolnshire, Illinois	2019-12-23	877212	1969
501	ZBH	Zimmer Biomet	Health Care	Health Care Equipment	Warsaw, Indiana	2001-08-07	1136869	1927
502	ZTS	Zoetis	Health Care	Pharmaceuticals	Parsippany, New Jersey	2013-06-21	1555280	1952

503 rows × 8 columns

In [5]:
```
stocks = overall.Symbol
```

In [6]:
```
stocks.head(5)
```

Out[6]:
```
0    MMM
1    ACE
2    AET
3   ABBV
4    ACN
Name: Symbol, dtype: object
```

In [7]:
```
stocks = stocks.to_list()
```

In [8]:
```
overall = overall[overall["Date added" >= start]
```

In [9]:
```
removed = pd.read_html(
    "https://en.wikipedia.org/wiki/List_of_S%26P_500_companies")[1]["Date", "Removed"]
```

In [10]: removed

Out[10]:

	Date		Removed
	Date	Ticker	Security
0	June 24, 2024	RHI	Robert Half
1	June 24, 2024	CMA	Comerica
2	June 24, 2024	SAPI	Sturma, Inc.
3	May 8, 2024	PXD	Pioneer Natural Resources
4	April 3, 2024	XRAY	Dentsply Sirona
...	...	...	...
345	June 9, 1999	HPH	Hamilton/Inger Industries
344	December 11, 1998	LDW	Amsco
345	December 11, 1998	GRN	General Re
346	December 11, 1998	SUN	SunAmerica
347	June 17, 1997	UDL	UDLAs

348 rows × 3 columns

In [11]:
```
removed = removed.set_index(removed.Date.Date)
```

In [12]:
```
removed.index = pd.to_datetime(removed.index)
```

In [13]:
```
removed = removed[removed.index >= start]
```

In [14]:
```
removed = removed.Removed.dropna()
```

In [15]: removed

Out[15]:

	Ticker	Security
Date		
2024-06-24	RHI	Robert Half
2024-06-24	CMA	Comerica
2024-06-24	SAPI	Sturma, Inc.
2024-05-08	PXD	Pioneer Natural Resources
2024-04-03	XRAY	Dentsply Sirona
...	...	...
2015-03-23	AVP	Avon Products
2015-03-18	CFN	Carefusion
2015-03-12	PETM	PetSmart
2015-01-27	SWY	Safeway
2015-01-27	COV	Covidien

203 rows × 2 columns

In [16]:
```
stocks.extend(removed.Ticker.to_list())
```

In [17]:
```
df = yf.download(stocks,start=start)["Close"]
```

```
[*********************100%***********************]  24 of 700 completed
ERRORS: possibly delisted; No price data found  (1d 2025-01-01 -> 2024-08-11)
[*********************100%***********************]  24 of 700 completed
ERRORS: possibly delisted; No price data found  (1d 2025-01-01 -> 2024-08-11)
[*********************100%***********************]  32 of 700 completed
ERRORS: possibly delisted; No price data found  (1d 2025-01-01 -> 2024-08-11)
[*********************100%***********************]  32 of 700 completed
ERRORS: possibly delisted; No price data found  (1d 2025-01-01 -> 2024-08-11)
[*********************100%***********************]  125
```