

# Name: Balakrishna Mupparaju

## Assignment: Week 12 & Project Milestone 5

```
In [6]: import pandas as pd
import numpy as np
from fuzzywuzzy import process
import requests
from bs4 import BeautifulSoup
import time
import warnings
# Suppress all warnings
warnings.filterwarnings("ignore")

# Load Kaggle dataset
kaggle_data = pd.read_csv("/Users/balakrishnamupparaju/Downloads/financials.
```

```
/opt/anaconda3/lib/python3.12/site-packages/fuzzywuzzy/fuzz.py:11: UserWarning: Using slow pure-python SequenceMatcher. Install python-Levenshtein to remove this warning
```

```
warnings.warn('Using slow pure-python SequenceMatcher. Install python-Levenshtein to remove this warning')
```

```
In [8]: kaggle_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 505 entries, 0 to 504
```

```
Data columns (total 14 columns):
```

#	Column	Non-Null Count	Dtype
0	Symbol	505 non-null	object
1	Name	505 non-null	object
2	Sector	505 non-null	object
3	Price	505 non-null	float64
4	Price/Earnings	503 non-null	float64
5	Dividend Yield	505 non-null	float64
6	Earnings/Share	505 non-null	float64
7	52 Week Low	505 non-null	float64
8	52 Week High	505 non-null	float64
9	Market Cap	505 non-null	float64
10	EBITDA	505 non-null	float64
11	Price/Sales	505 non-null	float64
12	Price/Book	497 non-null	float64
13	SEC Filings	505 non-null	object

```
dtypes: float64(10), object(4)
```

```
memory usage: 55.4+ KB
```

```
In [10]: # Replace headers in Kaggle dataset for clarity and consistency
#Converted original headers into more descriptive and consistent names.
```

```

kaggle_data.rename(columns={
    'Symbol': 'Ticker',
    'Name': 'Company_Name',
    'Sector': 'Industry_Sector',
    'Price': 'Stock_Price',
    'Price/Earnings': 'PE_Ratio',
    'Dividend Yield': 'Dividend_Yield',
    'Earnings/Share': 'Earnings_Per_Share',
    '52 Week Low': '52_Week_Low',
    '52 Week High': '52_Week_High',
    'Market Cap': 'Market_Cap',
    'EBITDA': 'EBITDA_Value',
    'Price/Sales': 'Price_to_Sales_Ratio',
    'Price/Book': 'Price_to_Book_Ratio',
    'SEC Filings': 'SEC_Filings_Link'
}, inplace=True)

# Verify header replacements
print("Step: Replaced Headers")
print(kaggle_data.head())

```

Step: Replaced Headers

	Ticker	Company_Name	Industry_Sector	Stock_Price	PE_Ratio
0	MMM	3M Company	Industrials	222.89	24.31
1	AOS	A.O. Smith Corp	Industrials	60.24	27.76
2	ABT	Abbott Laboratories	Health Care	56.27	22.51
3	ABBV	AbbVie Inc.	Health Care	108.48	19.41
4	ACN	Accenture plc	Information Technology	150.51	25.47

	Dividend_Yield	Earnings_Per_Share	52_Week_Low	52_Week_High
0	2.332862	7.92	259.77	175.490
1	1.147959	1.70	68.39	48.925
2	1.908982	0.26	64.60	42.280
3	2.499560	3.29	125.86	60.050
4	1.714470	5.44	162.60	114.820

	Market_Cap	EBITDA_Value	Price_to_Sales_Ratio	Price_to_Book_Ratio
0	1.387211e+11	9.048000e+09	4.390271	11.34
1	1.078342e+10	6.010000e+08	3.575483	6.35
2	1.021210e+11	5.744000e+09	3.740480	3.19
3	1.813863e+11	1.031000e+10	6.291571	26.14
4	9.876586e+10	5.643228e+09	2.604117	10.62

	SEC_Filings_Link
0	http://www.sec.gov/cgi-bin/browse-edgar?action...
1	http://www.sec.gov/cgi-bin/browse-edgar?action...
2	http://www.sec.gov/cgi-bin/browse-edgar?action...
3	http://www.sec.gov/cgi-bin/browse-edgar?action...
4	http://www.sec.gov/cgi-bin/browse-edgar?action...

```

In [12]: # Scrape ticker data from Wikipedia
wiki_url = "https://en.wikipedia.org/wiki/List_of_S%26P_500_companies"
response = requests.get(wiki_url)
soup = BeautifulSoup(response.content, 'html.parser')

```

```

# Extract Ticker, Company, Sector, and Industry
rows = soup.find('table', {'id': 'constituents'}).find_all('tr')
tickers = []
companies = []
sectors = []
industries = []
date_added=[]
#print(rows)
for row in rows[1:]:
    cols = row.find_all('td')
    tickers.append(cols[0].text.strip())
    companies.append(cols[1].text.strip())
    sectors.append(cols[3].text.strip())
    industries.append(cols[4].text.strip())
    date_added.append(cols[5].text.strip())

wiki_data = pd.DataFrame({
    "Ticker": tickers,
    "Company": companies,
    "Sector": sectors,
    "Industry": industries,
    "Date_Added":date_added
})
print("Step 2: Extracted Wikipedia dataset.")
print(wiki_data.head())

```

Step 2: Extracted Wikipedia dataset.

	Ticker	Company	Sector \
0	MMM	3M	Industrial Conglomerates
1	AOS	A. O. Smith	Building Products
2	ABT	Abbott Laboratories	Health Care Equipment
3	ABBV	AbbVie	Biotechnology
4	ACN	Accenture	IT Consulting & Other Services

	Industry	Date_Added
0	Saint Paul, Minnesota	1957-03-04
1	Milwaukee, Wisconsin	2017-07-26
2	North Chicago, Illinois	1957-03-04
3	North Chicago, Illinois	2012-12-31
4	Dublin, Ireland	2011-07-06

```

In [14]: # Replace headers in Wikipedia dataset for consistency with Kaggle dataset
wiki_data.rename(columns={
    'Ticker': 'Ticker', # Already aligned
    'Company': 'Company_Name', # Aligns with Kaggle's 'Company_Name'
    'Sector': 'Industry_Sector', # Matches Kaggle's 'Industry_Sector'
    'Industry': 'Headquarters' # Assuming this column reflects headquarters
}, inplace=True)

# Verify header replacements in Wikipedia dataset
print("Step: Replaced Headers in Wikipedia Dataset")
print(wiki_data.head())

```

Step: Replaced Headers in Wikipedia Dataset

	Ticker	Company_Name	Industry_Sector \
0	MMM	3M	Industrial Conglomerates
1	AOS	A. O. Smith	Building Products
2	ABT	Abbott Laboratories	Health Care Equipment
3	ABBV	AbbVie	Biotechnology
4	ACN	Accenture	IT Consulting & Other Services

	Headquarters	Date_Added
0	Saint Paul, Minnesota	1957-03-04
1	Milwaukee, Wisconsin	2017-07-26
2	North Chicago, Illinois	1957-03-04
3	North Chicago, Illinois	2012-12-31
4	Dublin, Ireland	2011-07-06

```
In [16]: import requests
import time
import pandas as pd

# Placeholder for storing data
api_data = []

# Fetch the full list of tickers from the Wikipedia dataset
tickers = wiki_data['Ticker'].tolist() # Use all available tickers
print(f"Total tickers to process: {len(tickers)}")

# Initialize ticker counter
ticker_count = 0

# Loop through each ticker and fetch data
for ticker in tickers:
    # Define API parameters
    params = {
        "function": "TIME_SERIES_DAILY",
        "symbol": ticker,
        "apikey": "" # Replace with my API key
    }

    # API call
    response = requests.get("https://www.alphavantage.co/query", params=params)
    data = response.json()

    # Process time series data if available
    if "Time Series (Daily)" in data:
        time_series = data["Time Series (Daily)"]
        for date, values in time_series.items():
            api_data.append({
                "Ticker": ticker,
                "Date": date,
                "Open": float(values.get("1. open", 0)),
                "High": float(values.get("2. high", 0)),
                "Low": float(values.get("3. low", 0)),
                "Close": float(values.get("4. close", 0)),
                "Volume": int(values.get("5. volume", 0))
            })
    else:
```

```

        print(f"No data for {ticker}: {data.get('Note', 'Unknown error')}")

    # Respect API rate limits
    time.sleep(12)

    # Increment ticker counter
    ticker_count += 1

    # Break the loop after processing 24 tickers
    if ticker_count >= 24:
        print(f"Processed data for {ticker_count} tickers. Exiting loop.")
        break

# Convert API data to DataFrame
api_df = pd.DataFrame(api_data)
print(f"Total rows fetched from API: {len(api_df)}")

# Save to CSV (optional)
api_df.to_csv("partial_api_data.csv", index=False)
print("Partial API data saved.")

```

Total tickers to process: 503  
 No data for AMCR: Unknown error  
 Processed data for 24 tickers. Exiting loop.  
 Total rows fetched from API: 2300  
 Partial API data saved.

```

In [18]: # Standardize Ticker and Industry_Sector casing
wiki_data['Ticker'] = wiki_data['Ticker'].str.upper()
wiki_data['Industry_Sector'] = wiki_data['Industry_Sector'].str.title() # F
wiki_data['Headquarters'] = wiki_data['Headquarters'].str.title() # Assumir
kaggle_data['Industry_Sector'] = kaggle_data['Industry_Sector'].str.title()

print("Step 4 Part 1: Fixed casing inconsistencies in Wiki and Kaggle dataset")

```

Step 4 Part 1: Fixed casing inconsistencies in Wiki and Kaggle datasets.

```

In [20]: # Fill missing Kaggle data with median/mean values
kaggle_data['Market_Cap'] = kaggle_data['Market_Cap'].fillna(kaggle_data['Ma
kaggle_data['PE_Ratio'] = kaggle_data['PE_Ratio'].fillna(kaggle_data['PE_Rat
print("Step 4.2: Handled missing values in Kaggle data.")

```

Step 4.2: Handled missing values in Kaggle data.

```

In [22]: # Remove duplicate rows from Wikipedia data
wiki_data.drop_duplicates(subset=['Ticker'], inplace=True)
print("Step 4.3: Removed duplicate entries in Wikipedia data.")

```

Step 4.3: Removed duplicate entries in Wikipedia data.

```

In [24]: # Match company names between Kaggle and Wikipedia
"""Corrected the column reference from Company to Company_Name to match the
The Matched_Company column will now store the closest company name match bet

wiki_data['Matched_Company'] = wiki_data['Company_Name'].apply(
    lambda x: process.extractOne(x, kaggle_data['Company_Name'].tolist())[0]
)

```

```
print("Step 4.4: Performed Fuzzy Matching for company names.")
print(wiki_data[['Company_Name', 'Matched_Company']].head())
```

Step 4.4: Performed Fuzzy Matching for company names.

	Company_Name	Matched_Company
0	3M	3M Company
1	A. O. Smith	A.O. Smith Corp
2	Abbott Laboratories	Abbott Laboratories
3	AbbVie	AbbVie Inc.
4	Accenture	Accenture plc

In [26]: **""First Merge:**

Kaggle and Wikipedia datasets are merged using the Ticker column to align fi

**Second Merge:**

The resulting merged dataset is further combined with the API dataset using

**Output:**

The cleaned and unified dataset is saved as final\_dataset.csv for further ar

```
# Merge Kaggle and Wikipedia datasets on Ticker
```

```
merged_data = pd.merge(kaggle_data, wiki_data, on='Ticker', how='inner')
```

```
# Merge the result with API data on Ticker
```

```
final_data = pd.merge(merged_data, api_df, on='Ticker', how='inner')
```

```
# Save the final cleaned dataset to a CSV file
```

```
final_data.to_csv("/Users/balakrishnamupparaju/Downloads/final_dataset.csv",
```

```
print("\nStep 5: Final cleaned dataset saved as 'cleaned_final_dataset.csv'.
```

```
print(final_data.head())
```

Step 5: Final cleaned dataset saved as 'cleaned\_final\_dataset.csv'.

	Ticker	Company_Name_x	Industry_Sector_x	Stock_Price	PE_Ratio	\
0	MMM	3M Company	Industrials	222.89	24.31	
1	MMM	3M Company	Industrials	222.89	24.31	
2	MMM	3M Company	Industrials	222.89	24.31	
3	MMM	3M Company	Industrials	222.89	24.31	
4	MMM	3M Company	Industrials	222.89	24.31	

	Dividend_Yield	Earnings_Per_Share	52_Week_Low	52_Week_High	\
0	2.332862	7.92	259.77	175.49	
1	2.332862	7.92	259.77	175.49	
2	2.332862	7.92	259.77	175.49	
3	2.332862	7.92	259.77	175.49	
4	2.332862	7.92	259.77	175.49	

	Market_Cap	...	Industry_Sector_y	Headquarters	\
0	1.387211e+11	...	Industrial Conglomerates	Saint Paul, Minnesota	
1	1.387211e+11	...	Industrial Conglomerates	Saint Paul, Minnesota	
2	1.387211e+11	...	Industrial Conglomerates	Saint Paul, Minnesota	
3	1.387211e+11	...	Industrial Conglomerates	Saint Paul, Minnesota	
4	1.387211e+11	...	Industrial Conglomerates	Saint Paul, Minnesota	

	Date_Added	Matched_Company	Date	Open	High	Low	Close	\
0	1957-03-04	3M Company	2025-05-30	148.73	149.45	147.60	148.35	
1	1957-03-04	3M Company	2025-05-29	149.87	150.36	148.18	149.63	
2	1957-03-04	3M Company	2025-05-28	149.65	150.00	148.19	148.66	
3	1957-03-04	3M Company	2025-05-27	150.00	150.00	146.66	149.49	
4	1957-03-04	3M Company	2025-05-23	147.45	148.69	147.01	147.62	

	Volume
0	5339916
1	2621435
2	2297543
3	3155330
4	2369839

[5 rows x 25 columns]

```
In [28]: # Ensure the necessary columns exist before deriving new columns
if 'High_Price' in final_data.columns and 'Low_Price' in final_data.columns:
    # Daily Price Range: Difference between High and Low prices
    final_data['Daily_Price_Range'] = final_data['High_Price'] - final_data['Low_Price']

if 'Close_Price' in final_data.columns and 'Open_Price' in final_data.columns:
    # Price Performance Index: Ratio of Close Price to Open Price
    final_data['Price_Performance_Index'] = final_data['Close_Price'] / final_data['Open_Price']

if 'Dividend_Yield' in merged_data.columns and 'Earnings_Per_Share' in merged_data.columns:
    # Dividend to Earnings Ratio: Ratio of Dividend Yield to Earnings per Share
    merged_data['Dividend_to_Earnings_Ratio'] = merged_data['Dividend_Yield'] / merged_data['Earnings_Per_Share']

print("Step: Derived new columns successfully.")
print(final_data.head())
```

Step: Derived new columns successfully.

	Ticker	Company_Name_x	Industry_Sector_x	Stock_Price	PE_Ratio	\
0	MMM	3M Company	Industrials	222.89	24.31	
1	MMM	3M Company	Industrials	222.89	24.31	
2	MMM	3M Company	Industrials	222.89	24.31	
3	MMM	3M Company	Industrials	222.89	24.31	
4	MMM	3M Company	Industrials	222.89	24.31	

	Dividend_Yield	Earnings_Per_Share	52_Week_Low	52_Week_High	\
0	2.332862	7.92	259.77	175.49	
1	2.332862	7.92	259.77	175.49	
2	2.332862	7.92	259.77	175.49	
3	2.332862	7.92	259.77	175.49	
4	2.332862	7.92	259.77	175.49	

	Market_Cap	...	Industry_Sector_y	Headquarters	\
0	1.387211e+11	...	Industrial Conglomerates	Saint Paul, Minnesota	
1	1.387211e+11	...	Industrial Conglomerates	Saint Paul, Minnesota	
2	1.387211e+11	...	Industrial Conglomerates	Saint Paul, Minnesota	
3	1.387211e+11	...	Industrial Conglomerates	Saint Paul, Minnesota	
4	1.387211e+11	...	Industrial Conglomerates	Saint Paul, Minnesota	

	Date_Added	Matched_Company	Date	Open	High	Low	Close	\
0	1957-03-04	3M Company	2025-05-30	148.73	149.45	147.60	148.35	
1	1957-03-04	3M Company	2025-05-29	149.87	150.36	148.18	149.63	
2	1957-03-04	3M Company	2025-05-28	149.65	150.00	148.19	148.66	
3	1957-03-04	3M Company	2025-05-27	150.00	150.00	146.66	149.49	
4	1957-03-04	3M Company	2025-05-23	147.45	148.69	147.01	147.62	

	Volume
0	5339916
1	2621435
2	2297543
3	3155330
4	2369839

[5 rows x 25 columns]

```
In [30]: #Earnings_to_MarketCap_Ratio: Assess the company's earnings relative to its
final_data['Earnings_to_MarketCap_Ratio'] = final_data['Earnings_Per_Share']
#Debt_Equity_Calculation: Use Industry_Sector and financial metrics to deriv
sector_avg_de_ratio = final_data.groupby('Industry_Sector_x')['PE_Ratio'].me
final_data['Sector_Avg_PE_Ratio'] = final_data['Industry_Sector_x'].map(sect
#Volume_to_MarketCap: Ratio of traded volume to market capitalization. This
final_data['Volume_to_MarketCap'] = final_data['Volume'] / final_data['Marke

#Daily_Market_Movement: Aggregate daily price movement (High - Low), providi
final_data['Daily_Market_Movement'] = final_data['High'] - final_data['Low']

#Region_Sector_Combo: Combine Headquarters and Industry_Sector_y for geograp
final_data['Region_Sector_Combo'] = final_data['Headquarters'] + ' - ' + fir

#Time_Trend: Create a rolling average of daily Close prices for a 30-day per
#final_data['30_Day_Rolling_Avg'] = final_data['Close'].rolling(window=30).n
final_data['30_Day_Rolling_Avg'] = final_data['Close'].rolling(window=30).me
```



```
final_data['30_Day_Rolling_Avg'].fillna(method='ffill', inplace=True)
print(final_data[['Date', 'Close', '30_Day_Rolling_Avg']].head(40))
```

	Date	Close	30_Day_Rolling_Avg
0	2025-05-30	148.35	NaN
1	2025-05-29	149.63	NaN
2	2025-05-28	148.66	NaN
3	2025-05-27	149.49	NaN
4	2025-05-23	147.62	NaN
5	2025-05-22	149.78	NaN
6	2025-05-21	149.40	NaN
7	2025-05-20	154.15	NaN
8	2025-05-19	154.14	NaN
9	2025-05-16	153.11	NaN
10	2025-05-15	148.74	NaN
11	2025-05-14	147.88	NaN
12	2025-05-13	150.74	NaN
13	2025-05-12	150.33	NaN
14	2025-05-09	142.60	NaN
15	2025-05-08	141.12	NaN
16	2025-05-07	138.64	NaN
17	2025-05-06	138.02	NaN
18	2025-05-05	140.80	NaN
19	2025-05-02	142.08	NaN
20	2025-05-01	137.90	NaN
21	2025-04-30	138.91	NaN
22	2025-04-29	138.38	NaN
23	2025-04-28	137.93	NaN
24	2025-04-25	137.32	NaN
25	2025-04-24	138.88	NaN
26	2025-04-23	136.06	NaN
27	2025-04-22	136.33	NaN
28	2025-04-21	126.09	NaN
29	2025-04-17	130.21	143.443000
30	2025-04-16	130.46	142.846667
31	2025-04-15	135.26	142.367667
32	2025-04-14	136.01	141.946000
33	2025-04-11	135.95	141.494667
34	2025-04-10	132.97	141.006333
35	2025-04-09	138.32	140.624333
36	2025-04-08	127.16	139.883000
37	2025-04-07	128.55	139.029667
38	2025-04-04	126.91	138.122000
39	2025-04-03	139.74	137.676333

```
In [32]: print(f"Total columns after adding new features: {len(final_data.columns)}")
```

Total columns after adding new features: 31

```
In [56]: final_data.head()
```

Out [56]:

	Ticker	Company_Name_x	Industry_Sector_x	Stock_Price	PE_Ratio	Dividend_Yiel
0	MMM	3M Company	Industrials	222.89	24.31	2.3
1	MMM	3M Company	Industrials	222.89	24.31	2.3
2	MMM	3M Company	Industrials	222.89	24.31	2.3
3	MMM	3M Company	Industrials	222.89	24.31	2.3
4	MMM	3M Company	Industrials	222.89	24.31	2.3

5 rows × 31 columns

```
In [34]: # Get the number of rows
total_rows = final_data.shape[0]
print(f"Total number of rows: {total_rows}")
```

Total number of rows: 2200

```
In [36]: # Convert Market_Cap to a readable format
final_data['Market_Cap'] = final_data['Market_Cap'].apply(lambda x: f"{x:,}")

print("Formatted Market_Cap values:")
print(final_data[['Ticker', 'Market_Cap']].head())

# Disable scientific notation globally
pd.options.display.float_format = '{:,.2f}'.format

print("Disabled scientific notation for all float columns.")
print(final_data.head())
```

Formatted Market\_Cap values:

	Ticker	Market_Cap
0	MMM	138,721,055,226
1	MMM	138,721,055,226
2	MMM	138,721,055,226
3	MMM	138,721,055,226
4	MMM	138,721,055,226

Disabled scientific notation for all float columns.

	Ticker	Company_Name_x	Industry_Sector_x	Stock_Price	PE_Ratio	\
0	MMM	3M Company	Industrials	222.89	24.31	
1	MMM	3M Company	Industrials	222.89	24.31	
2	MMM	3M Company	Industrials	222.89	24.31	
3	MMM	3M Company	Industrials	222.89	24.31	
4	MMM	3M Company	Industrials	222.89	24.31	

	Dividend_Yield	Earnings_Per_Share	52_Week_Low	52_Week_High	\
0	2.33	7.92	259.77	175.49	
1	2.33	7.92	259.77	175.49	
2	2.33	7.92	259.77	175.49	
3	2.33	7.92	259.77	175.49	
4	2.33	7.92	259.77	175.49	

	Market_Cap	...	High	Low	Close	Volume	\
0	138,721,055,226	...	149.45	147.60	148.35	5339916	
1	138,721,055,226	...	150.36	148.18	149.63	2621435	
2	138,721,055,226	...	150.00	148.19	148.66	2297543	
3	138,721,055,226	...	150.00	146.66	149.49	3155330	
4	138,721,055,226	...	148.69	147.01	147.62	2369839	

	Earnings_to_MarketCap_Ratio	Sector_Avg_PE_Ratio	Volume_to_MarketCap	\
0	0.00		24.38	0.00
1	0.00		24.38	0.00
2	0.00		24.38	0.00
3	0.00		24.38	0.00
4	0.00		24.38	0.00

	Daily_Market_Movement	Region_Sector_Combo	\
0	1.85	Saint Paul, Minnesota - Industrial Conglomerates	
1	2.18	Saint Paul, Minnesota - Industrial Conglomerates	
2	1.81	Saint Paul, Minnesota - Industrial Conglomerates	
3	3.34	Saint Paul, Minnesota - Industrial Conglomerates	
4	1.68	Saint Paul, Minnesota - Industrial Conglomerates	

	30_Day_Rolling_Avg
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN

[5 rows x 31 columns]

In [38]: *#Project Milestone 3*

```
import pandas as pd
import requests
```

```

from bs4 import BeautifulSoup
from fuzzywuzzy import process
import warnings
warnings.filterwarnings("ignore")
# =====
# Scrape the Wikipedia S&P 500 data directly from the HTML source.
wiki_url = "https://en.wikipedia.org/wiki/List_of_S%26P_500_companies"
response = requests.get(wiki_url)
soup = BeautifulSoup(response.content, 'html.parser')
table = soup.find('table', {'id': 'constituents'})
rows = table.find_all('tr')

# Gather data from the table rows (skip the header row)
tickers = []
companies = []
sectors = []
industries = []
date_added = [] # Date on which the company was added to the index

for row in rows[1:]: # Skipping header row
    cols = row.find_all('td')
    tickers.append(cols[0].text.strip())
    companies.append(cols[1].text.strip())
    sectors.append(cols[3].text.strip())
    industries.append(cols[4].text.strip())
    date_added.append(cols[5].text.strip())

# Create DataFrame from the scraped data
wiki_data = pd.DataFrame({
    "Ticker": tickers,
    "Company_Name": companies,
    "Sector": sectors,
    "Industry": industries,
    "Date_Added": date_added
})

# =====
# Step #1: Replace Headers
# Description: Rename headers for clarity and consistency with other data sets
wiki_data.rename(columns={
    "Sector": "Industry_Sector" # Aligns with Kaggle and other datasets
}, inplace=True)
print("Step #1: Replaced Headers")
print(wiki_data.head(), "\n")

# =====
# Step #2: Fix Casing/Inconsistent Values
# Description: Convert text columns to Title Case to ensure consistency.
wiki_data['Company_Name'] = wiki_data['Company_Name'].str.title()
wiki_data['Industry_Sector'] = wiki_data['Industry_Sector'].str.title()
wiki_data['Industry'] = wiki_data['Industry'].str.title()
print("Step #2: Fixed casing in text columns")

```

```

print(wiki_data.head(), "\n")

# =====
# Step #3: Identify and Remove Duplicates
# Description: Remove duplicate entries by Ticker & Company_Name to ensure u
num_duplicates = wiki_data.duplicated(subset=['Ticker', 'Company_Name']).sum
print(f"Step #3: Found {num_duplicates} duplicate(s) before removal.")
wiki_data = wiki_data.drop_duplicates(subset=['Ticker', 'Company_Name'])
print("Duplicates removed.")
print(wiki_data.head(), "\n")

# =====
# Step #4: Format Date Fields
# Description: Convert the 'Date_Added' column to datetime format for proper u
wiki_data['Date_Added'] = pd.to_datetime(wiki_data['Date_Added'], errors='co
print("Step #4: Converted 'Date_Added' to datetime format")
print(wiki_data[['Ticker', 'Date_Added']].head(), "\n")

# =====
# Step #5: Clean and Trim Text Fields / Remove Missing Critical Data
# Description: Trim extra whitespace in the 'Ticker' column, standardize it,
wiki_data['Ticker'] = wiki_data['Ticker'].str.strip().str.upper()
wiki_data = wiki_data.dropna(subset=['Ticker'])
print("Step #5: Trimmed Ticker text and dropped rows with missing Ticker")
print(wiki_data.head(), "\n")

# =====
# Final Cleaned Dataset
print("Cleaned Website Dataset (Wikipedia) after all transformations:")
print(wiki_data.to_string(index=False))

```

Step #1: Replaced Headers

	Ticker	Company_Name	Industry_Sector \
0	MMM	3M	Industrial Conglomerates
1	AOS	A. O. Smith	Building Products
2	ABT	Abbott Laboratories	Health Care Equipment
3	ABBV	AbbVie	Biotechnology
4	ACN	Accenture	IT Consulting & Other Services

	Industry	Date_Added
0	Saint Paul, Minnesota	1957-03-04
1	Milwaukee, Wisconsin	2017-07-26
2	North Chicago, Illinois	1957-03-04
3	North Chicago, Illinois	2012-12-31
4	Dublin, Ireland	2011-07-06

Step #2: Fixed casing in text columns

	Ticker	Company_Name	Industry_Sector \
0	MMM	3M	Industrial Conglomerates
1	AOS	A. O. Smith	Building Products
2	ABT	Abbott Laboratories	Health Care Equipment
3	ABBV	AbbVie	Biotechnology
4	ACN	Accenture	It Consulting & Other Services

	Industry	Date_Added
0	Saint Paul, Minnesota	1957-03-04
1	Milwaukee, Wisconsin	2017-07-26
2	North Chicago, Illinois	1957-03-04
3	North Chicago, Illinois	2012-12-31
4	Dublin, Ireland	2011-07-06

Step #3: Found 0 duplicate(s) before removal.

Duplicates removed.

	Ticker	Company_Name	Industry_Sector \
0	MMM	3M	Industrial Conglomerates
1	AOS	A. O. Smith	Building Products
2	ABT	Abbott Laboratories	Health Care Equipment
3	ABBV	AbbVie	Biotechnology
4	ACN	Accenture	It Consulting & Other Services

	Industry	Date_Added
0	Saint Paul, Minnesota	1957-03-04
1	Milwaukee, Wisconsin	2017-07-26
2	North Chicago, Illinois	1957-03-04
3	North Chicago, Illinois	2012-12-31
4	Dublin, Ireland	2011-07-06

Step #4: Converted 'Date\_Added' to datetime format

	Ticker	Date_Added
0	MMM	1957-03-04
1	AOS	2017-07-26
2	ABT	1957-03-04
3	ABBV	2012-12-31
4	ACN	2011-07-06

Step #5: Trimmed Ticker text and dropped rows with missing Ticker

	Ticker	Company_Name	Industry_Sector \
--	--------	--------------	-------------------

0	MMM	3M	Industrial Conglomerates
1	AOS	A. O. Smith	Building Products
2	ABT	Abbott Laboratories	Health Care Equipment
3	ABBV	Abbvie	Biotechnology
4	ACN	Accenture	It Consulting & Other Services

	Industry	Date_Added
0	Saint Paul, Minnesota	1957-03-04
1	Milwaukee, Wisconsin	2017-07-26
2	North Chicago, Illinois	1957-03-04
3	North Chicago, Illinois	2012-12-31
4	Dublin, Ireland	2011-07-06

Cleaned Website Dataset (Wikipedia) after all transformations:

Ticker	Company_Name	Industry	Date_Added
MMM	3M	Industrial Conglomerates	Saint Paul, Minnesota 1957-03-04
AOS	A. O. Smith	Building Products	Milwaukee, Wisconsin 2017-07-26
ABT	Abbott Laboratories	Health Care Equipment	North Chicago, Illinois 1957-03-04
ABBV	Abbvie	Biotechnology	North Chicago, Illinois 2012-12-31
ACN	Accenture	nsulting & Other Services	Dublin, Ireland 2011-07-06
ADBE	Adobe Inc.	Application Software	San Jose, California 1997-05-05
AMD	Advanced Micro Devices	Semiconductors	Santa Clara, California 2017-03-20
AES	Aes Corporation	roducers & Energy Traders	Independent Power P Arlington, Virginia 1998-10-02
AFL	Aflac	Life & Health Insurance	Columbus, Georgia 1999-05-28
A	Agilent Technologies	Sciences Tools & Services	Life Santa Clara, California 2000-06-05
APD	Air Products	Industrial Gases	Upper Macungie Township, Pennsylvania 1985-04-30
ABNB	Airbnb	s, Resorts & Cruise Lines	Hotel San Francisco, California 2023-09-18
AKAM	Akamai Technologies	Services & Infrastructure	Internet Cambridge, Massachusetts 2007-07-12
ALB	Albemarle Corporation	Specialty Chemicals	Charlotte, North Carolina 2016-07-01
ARE	Alexandria Real Estate Equities	Office Reits	Pasadena, California 2017-03-20
ALGN	Align Technology	Health Care Supplies	Tempe, Arizona 2017-06-19
ALLE	Allegion		

Building Products		Dublin, Ireland 2013-12-02
LNT	Alliant Energy	
Electric Utilities		Madison, Wisconsin 2016-07-01
ALL	Allstate	Prop
erty & Casualty Insurance		Northbrook, Illinois 1995-0
7-13		
G00GL	Alphabet Inc. (Class A)	Int
eractive Media & Services		Mountain View, California 2014-0
4-03		
G00G	Alphabet Inc. (Class C)	Int
eractive Media & Services		Mountain View, California 2006-0
4-03		
MO	Altria	
Tobacco		Richmond, Virginia 1957-03-04
AMZN	Amazon	
Broadline Retail		Seattle, Washington 2005-11-18
AMCR	Amcor	Paper & Plastic Packa
ging Products & Materials		Warmley, Bristol, United Kingdom 2019-0
6-07		
AEE	Ameren	
Multi-Utilities		St. Louis, Missouri 1991-09-19
AEP	American Electric Power	
Electric Utilities		Columbus, Ohio 1957-03-04
AXP	American Express	
Consumer Finance		New York City, New York 1976-06-30
AIG	American International Group	
Multi-Line Insurance		New York City, New York 1980-03-31
AMT	American Tower	
Telecom Tower Reits		Boston, Massachusetts 2007-11-19
AWK	American Water Works	
Water Utilities		Camden, New Jersey 2016-03-04
AMP	Ameriprise Financial	Asset M
anagement & Custody Banks		Minneapolis, Minnesota 2005-1
0-03		
AME	Ametek	Electric
al Components & Equipment		Berwyn, Pennsylvania 2013-0
9-23		
AMGN	Amgen	
Biotechnology		Thousand Oaks, California 1992-01-02
APH	Amphenol	
Electronic Components		Wallingford, Connecticut 2008-09-30
ADI	Analog Devices	
Semiconductors		Wilmington, Massachusetts 1999-10-12
ANSS	Ansys	
Application Software		Canonsburg, Pennsylvania 2017-06-19
AON	Aon Plc	
Insurance Brokers		London, United Kingdom 1996-04-23
APA	Apa Corporation	Oil & Gas
Exploration & Production		Houston, Texas 1997-07
-28		
APO	Apollo Global Management	Asset M
anagement & Custody Banks		New York City, New York 2024-1
2-23		
AAPL	Apple Inc.	Technology Hardwa
re, Storage & Peripherals		Cupertino, California 1982-1
1-30		



AMAT	Applied Materials	Semiconduc
tor Materials & Equipment		Santa Clara, California 1995-0
3-16		
APTV	Aptiv	Aut
omotive Parts & Equipment		Dublin, Ireland 2012-1
2-24		
ACGL	Arch Capital Group	Prop
erty & Casualty Insurance		Hamilton, Bermuda 2022-1
1-01		
ADM	Archer Daniels Midland	Agricul
tural Products & Services		Chicago, Illinois 1957-0
3-04		
ANET	Arista Networks	
Communications Equipment		Santa Clara, California 2018-08
-28		
AJG	Arthur J. Gallagher & Co.	
Insurance Brokers		Rolling Meadows, Illinois 2016-05-31
AIZ	Assurant	
Multi-Line Insurance		Atlanta, Georgia 2007-04-10
T	At&T	Integrated T
elecommunication Services		Dallas, Texas 1983-1
1-30		
ATO	Atmos Energy	
Gas Utilities		Dallas, Texas 2019-02-15
ADSK	Autodesk	
Application Software		San Francisco, California 1989-12-01
ADP	Automatic Data Processing	Human Resou
rces & Employment Services		Roseland, New Jersey 1981-0
3-31		
AZO	Autozone	
Automotive Retail		Memphis, Tennessee 1997-01-02
AVB	Avalonbay Communities	Multi
-Family Residential Reits		Arlington, Virginia 2007-0
1-10		
AVY	Avery Dennison	Paper & Plastic Packa
ging Products & Materials		Mentor, Ohio 1987-1
2-31		
AXON	Axon Enterprise	
Aerospace & Defense		Scottsdale, Arizona 2023-05-04
BKR	Baker Hughes	Oil &
Gas Equipment & Services		Houston, Texas 2017-07
-07		
BALL	Ball Corporation	Metal, G
lass & Plastic Containers		Broomfield, Colorado 1984-1
0-31		
BAC	Bank Of America	
Diversified Banks		Charlotte, North Carolina 1976-06-30
BAX	Baxter International	
Health Care Equipment		Deerfield, Illinois 1972-09-30
BDX	Becton Dickinson	
Health Care Equipment		Franklin Lakes, New Jersey 1972-09-30
BRK.B	Berkshire Hathaway	
Multi-Sector Holdings		Omaha, Nebraska 2010-02-16
BBY	Best Buy	Comp
uter & Electronics Retail		Richfield, Minnesota 1999-0
6-29		

TECH	Bio-Techne	Life
Sciences Tools & Services		Minneapolis, Minnesota 2021-0
8-30		
BIIB	Biogen	
Biotechnology	Cambridge, Massachusetts	2003-11-13
BLK	Blackrock	Asset M
anagement & Custody Banks		New York City, New York 2011-0
4-04		
BX	Blackstone Inc.	Asset M
anagement & Custody Banks		New York City, New York 2023-0
9-18		
BK	Bny Mellon	Asset M
anagement & Custody Banks		New York City, New York 1995-0
3-31		
BA	Boeing	
Aerospace & Defense		Arlington, Virginia 1957-03-04
BKNG	Booking Holdings	Hotel
s, Resorts & Cruise Lines		Norwalk, Connecticut 2009-1
1-06		
BSX	Boston Scientific	
Health Care Equipment		Marlborough, Massachusetts 1995-02-24
BMJ	Bristol Myers Squibb	
Pharmaceuticals		New York City, New York 1957-03-04
AVGO	Broadcom	
Semiconductors		Palo Alto, California 2014-05-08
BR	Broadridge Financial Solutions	Data Process
ing & Outsourced Services		Lake Success, New York 2018-0
6-18		
BR0	Brown & Brown	
Insurance Brokers		Daytona Beach, Florida 2021-09-20
BF.B	Brown-Forman	
Distillers & Vintners		Louisville, Kentucky 1982-10-31
BLDR	Builders Firstsource	
Building Products		Irving, Texas 2023-12-18
BG	Bunge Global	Agricul
tural Products & Services		Chesterfield, Missouri 2023-0
3-15		
BXP	Bxp, Inc.	
Office Reits		Boston, Massachusetts 2006-04-03
CHRW	C.H. Robinson	
Air Freight & Logistics		Eden Prairie, Minnesota 2007-03-02
CDNS	Cadence Design Systems	
Application Software		San Jose, California 2017-09-18
CZR	Caesars Entertainment	
Casinos & Gaming		Reno, Nevada 2021-03-22
CPT	Camden Property Trust	Multi
-Family Residential Reits		Houston, Texas 2022-0
4-04		
CPB	Campbell'S Company (The)	
Packaged Foods & Meats		Camden, New Jersey 1957-03-04
COF	Capital One	
Consumer Finance		Tysons Corner, Virginia 1998-07-01
CAH	Cardinal Health	
Health Care Distributors		Dublin, Ohio 1997-05

-27	KMX	Carmax	
Automotive Retail		Richmond, Virginia	2010-06-28
CCL	Carnival	Hotel	
s, Resorts & Cruise Lines		Miami, Florida	1998-1
2-22	CARR	Carrier Global	
Building Products		Palm Beach Gardens, Florida	2020-04-03
CAT	Caterpillar Inc.	Construction Machinery & Heavy	
Transportation Equipment		Irving, Texas	1957-03
-04	CBOE	Cboe Global Markets	F
inancial Exchanges & Data		Chicago, Illinois	2017-0
3-01	CBRE	Cbre Group	
Real Estate Services		Dallas, Texas	2006-11-10
CDW	Cdw Corporation		
Technology Distributors		Vernon Hills, Illinois	2019-09-
23	COR	Cencora	
Health Care Distributors		Conshohocken, Pennsylvania	2001-08
-30	CNC	Centene Corporation	
Managed Health Care		St. Louis, Missouri	2016-03-30
CNP	Centerpoint Energy		
Multi-Utilities		Houston, Texas	1985-07-31
CF	Cf Industries	Fertilizers	
& Agricultural Chemicals		Deerfield, Illinois	2008-08
-27	CRL	Charles River Laboratories	Life
Sciences Tools & Services		Wilmington, Massachusetts	2021-0
5-14	SCHW	Charles Schwab Corporation	Inves
tment Banking & Brokerage		Westlake, Texas	1997-0
6-02	CHTR	Charter Communications	
Cable & Satellite		Stamford, Connecticut	2016-09-08
CVX	Chevron Corporation		
Integrated Oil & Gas		San Ramon, California	1957-03-04
CMG	Chipotle Mexican Grill		
Restaurants		Newport Beach, California	2011-04-28
CB	Chubb Limited	Prop	
erty & Casualty Insurance		Zurich, Switzerland	2010-0
7-15	CHD	Church & Dwight	
Household Products		Ewing, New Jersey	2015-12-29
CI	Cigna		
Health Care Services		Bloomfield, Connecticut	1976-06-30
CINF	Cincinnati Financial	Prop	
erty & Casualty Insurance		Fairfield, Ohio	1997-1
2-18	CTAS	Cintas	Div
ersified Support Services		Mason, Ohio	2001-0
3-01	CSCO	Cisco	
Communications Equipment		San Jose, California	1993-12

-01			
C	Citigroup		
Diversified Banks	New York City, New York	1988-05-31	
CFG	Citizens Financial Group		
Regional Banks	Providence, Rhode Island	2016-01-29	
CLX	Clorox		
Household Products	Oakland, California	1969-03-31	
CME	Cme Group		F
inancial Exchanges & Data	Chicago, Illinois	2006-0	
8-11			
CMS	Cms Energy		
Multi-Utilities	Jackson, Michigan	1957-03-04	
KO	Coca-Cola Company (The)		Soft Drinks
& Non-Alcoholic Beverages	Atlanta, Georgia	1957-0	
3-04			
CTSH	Cognizant		It Co
nsulting & Other Services	Teaneck, New Jersey	2006-1	
1-17			
COIN	Coinbase		F
inancial Exchanges & Data	New York City, New York	2025-0	
5-19			
CL	Colgate-Palmolive		
Household Products	New York City, New York	1957-03-04	
CMCSA	Comcast		
Cable & Satellite	Philadelphia, Pennsylvania	2002-11-19	
CAG	Conagra Brands		
Packaged Foods & Meats	Chicago, Illinois	1983-08-3	
1			
COP	Conocophillips		Oil & Gas
Exploration & Production	Houston, Texas	1957-03	
-04			
ED	Consolidated Edison		
Multi-Utilities	New York City, New York	1957-03-04	
STZ	Constellation Brands		
Distillers & Vintners	Rochester, New York	2005-07-01	
CEG	Constellation Energy		
Electric Utilities	Baltimore, Maryland	2022-02-02	
COO	Cooper Companies (The)		
Health Care Supplies	San Ramon, California	2016-09-23	
CPRT	Copart		Div
ersified Support Services	Dallas, Texas	2018-0	
7-02			
GLW	Corning Inc.		
Electronic Components	Corning, New York	1995-02-27	
CPAY	Corpay		Transaction & Pa
ymment Processing Services	Atlanta, Georgia	2018-0	
6-20			
CTVA	Corteva		Fertilizers
& Agricultural Chemicals	Indianapolis, Indiana	2019-06	
-03			
CSGP	Costar Group		
Real Estate Services	Washington, D.C.	2022-09-19	
COST	Costco		Consumer S
taples Merchandise Retail	Issaquah, Washington	1993-1	
0-01			
CTRA	Coterra		Oil & Gas

Exploration & Production -23		Houston, Texas 2008-06
CRWD	CrowdStrike	
Systems Software		Austin, Texas 2024-06-24
CCI	Crown Castle	
Telecom Tower Reits		Houston, Texas 2012-03-14
CSX	Csx Corporation	
Rail Transportation		Jacksonville, Florida 1957-03-04
CMI	Cummins Construction Machinery & Heavy	
Transportation Equipment -31		Columbus, Indiana 1965-03
CVS	Cvs Health	
Health Care Services		Woonsocket, Rhode Island 1957-03-04
DHR	Danaher Corporation	Life
Sciences Tools & Services		Washington, D.C. 1998-1
1-18		
DRI	Darden Restaurants	
Restaurants		Orlando, Florida 1995-05-31
DVA	Davita	
Health Care Services		Denver, Colorado 2008-07-31
DAY	Dayforce	Human Resou
rcce & Employment Services		Minneapolis, Minnesota 2021-0
9-20		
DECK	Deckers Brands	
Footwear		Goleta, California 2024-03-18
DE	Deere & Company	Agri
cultural & Farm Machinery		Moline, Illinois 1957-0
3-04		
DELL	Dell Technologies	Technology Hardwa
re, Storage & Peripherals		Round Rock, Texas 2024-0
9-23		
DAL	Delta Air Lines	
Passenger Airlines		Atlanta, Georgia 2013-09-11
DVN	Devon Energy	Oil & Gas
Exploration & Production -30		Oklahoma City, Oklahoma 2000-08
DXCM	Dexcom	
Health Care Equipment		San Diego, California 2020-05-12
FANG	Diamondback Energy	Oil & Gas
Exploration & Production -03		Midland, Texas 2018-12
DLR	Digital Realty	
Data Center Reits		Austin, Texas 2016-05-18
DG	Dollar General	Consumer S
taples Merchandise Retail		Goodlettsville, Tennessee 2012-1
2-03		
DLTR	Dollar Tree	Consumer S
taples Merchandise Retail		Chesapeake, Virginia 2011-1
2-19		
D	Dominion Energy	
Multi-Utilities		Richmond, Virginia 2016-11-30
DPZ	Domino'S	
Restaurants		Ann Arbor, Michigan 2020-05-12
DASH	Doordash	Spec
ialized Consumer Services		San Francisco, California 2025-0
3-24		

DOV	Dover Corporation	Industrial Machinery & Supplies & Components
Downers Grove, Illinois 1985-10-31		
DOW	Dow Inc.	
Commodity Chemicals	Midland, Michigan 2019-04-01	
DHI	D. R. Horton	
Homebuilding	Arlington, Texas 2005-06-22	
DTE	Dte Energy	
Multi-Utilities	Detroit, Michigan 1957-03-04	
DUK	Duke Energy	
Electric Utilities	Charlotte, North Carolina 1976-06-30	
DD	Dupont	
Specialty Chemicals	Wilmington, Delaware 2019-06-03	
EMN	Eastman Chemical Company	
Specialty Chemicals	Kingsport, Tennessee 1994-01-01	
ETN	Eaton Corporation	Electric
al Components & Equipment	Dublin, Ireland 1957-03-04	
EBAY	Ebay Inc.	
Broadline Retail	San Jose, California 2002-07-22	
ECL	Ecolab	
Specialty Chemicals	Saint Paul, Minnesota 1989-01-31	
EIX	Edison International	
Electric Utilities	Rosemead, California 1957-03-04	
EW	Edwards Lifesciences	
Health Care Equipment	Irvine, California 2011-04-01	
EA	Electronic Arts	Inter
active Home Entertainment	Redwood City, California 2002-07-22	
ELV	Elevance Health	
Managed Health Care	Indianapolis, Indiana 2002-07-25	
EMR	Emerson Electric	Electric
al Components & Equipment	Ferguson, Missouri 1965-03-31	
ENPH	Enphase Energy	Semiconduc
tor Materials & Equipment	Fremont, California 2021-01-07	
ETR	Entergy	
Electric Utilities	New Orleans, Louisiana 1957-03-04	
EOG	Eog Resources	Oil & Gas
Exploration & Production	Houston, Texas 2000-11-02	
EPAM	Epam Systems	It Co
nsulting & Other Services	Newtown, Pennsylvania 2021-12-14	
EQT	Eqt Corporation	Oil & Gas
Exploration & Production	Pittsburgh, Pennsylvania 2022-11-03	
EFX	Equifax	Resea
rch & Consulting Services	Atlanta, Georgia 1997-06-19	
EQIX	Equinix	
Data Center Reits	Redwood City, California 2015-03-20	
EQR	Equity Residential	Multi
-Family Residential Reits	Chicago, Illinois 2001-12-03	

ERIE	Erie Indemnity	
Insurance Brokers	Erie, Pennsylvania	2024-09-23
ESS	Essex Property Trust	Multi
-Family Residential Reits	San Mateo, California	2014-0
4-02		
EL	Estée Lauder Companies (The)	
Personal Care Products	New York City, New York	2006-01-0
5		
EG	Everest Group	
Reinsurance	Hamilton, Bermuda	2017-06-19
EVRG	Evergy	
Electric Utilities	Kansas City, Missouri	2018-06-05
ES	Eversource Energy	
Electric Utilities	Hartford, Connecticut	2009-07-24
EXC	Exelon	
Electric Utilities	Chicago, Illinois	1957-03-04
EXE	Expand Energy	Oil & Gas
Exploration & Production	Oklahoma City, Oklahoma	2025-03
-24		
EXPE	Expedia Group	Hotel
s, Resorts & Cruise Lines	Seattle, Washington	2007-1
0-02		
EXPD	Expeditors International	
Air Freight & Logistics	Seattle, Washington	2007-10-
10		
EXR	Extra Space Storage	
Self-Storage Reits	Salt Lake City, Utah	2016-01-19
XOM	Exxonmobil	
Integrated Oil & Gas	Irving, Texas	1957-03-04
FFIV	F5, Inc.	
Communications Equipment	Seattle, Washington	2010-12
-20		
FDS	Factset	F
inancial Exchanges & Data	Norwalk, Connecticut	2021-1
2-20		
FICO	Fair Isaac	
Application Software	Bozeman, Montana	2023-03-20
FAST	Fastenal	Trading
Companies & Distributors	Winona, Minnesota	2008-09
-15		
FRT	Federal Realty Investment Trust	
Retail Reits	Rockville, Maryland	2016-02-01
FDX	Fedex	
Air Freight & Logistics	Memphis, Tennessee	1980-12-
31		
FIS Fidelity National Information Services	Transaction & Pa	
yment Processing Services	Jacksonville, Florida	2006-1
1-10		
FITB	Fifth Third Bancorp	
Regional Banks	Cincinnati, Ohio	1996-03-29
FSLR	First Solar	
Semiconductors	Tempe, Arizona	2022-12-19
FE	Firstenergy	
Electric Utilities	Akron, Ohio	1997-11-28
FI	Fiserv	Transaction & Pa
yment Processing Services	Brookfield, Wisconsin	2001-0

4-02		
F	Ford Motor Company	
Automobile Manufacturers		Dearborn, Michigan 1957-03
-04		
FTNT	Fortinet	
Systems Software	Sunnyvale, California	2018-10-11
FTV	Fortive	Industrial Machiner
y & Supplies & Components		Everett, Washington 2016-0
7-01		
FOXA	Fox Corporation (Class A)	
Broadcasting	New York City, New York	2019-03-19
FOX	Fox Corporation (Class B)	
Broadcasting	New York City, New York	2019-03-19
BEN	Franklin Resources	Asset M
anagement & Custody Banks		San Mateo, California 1998-0
4-30		
FCX	Freeport-Mcmoran	
Copper	Phoenix, Arizona	2011-07-01
GRMN	Garmin	
Consumer Electronics	Schaffhausen, Switzerland	2012-12-12
IT	Gartner	It Co
nsulting & Other Services		Stamford, Connecticut 2017-0
4-05		
GE	Ge Aerospace	
Aerospace & Defense		Evendale, Ohio 1957-03-04
GEHC	Ge Healthcare	
Health Care Equipment		Chicago, Illinois 2023-01-04
GEV	Ge Vernova	H
eavy Electrical Equipment		Cambridge, Massachusetts 2024-0
4-02		
GEN	Gen Digital	
Systems Software		Tempe, Arizona 2003-03-25
GNRC	Generac	Electric
al Components & Equipment		Waukesha, Wisconsin 2021-0
3-22		
GD	General Dynamics	
Aerospace & Defense	Falls Church, Virginia	1957-03-04
GIS	General Mills	
Packaged Foods & Meats		Golden Valley, Minnesota 1957-03-0
4		
GM	General Motors	
Automobile Manufacturers		Detroit, Michigan 2013-06
-06		
GPC	Genuine Parts Company	
Distributors		Atlanta, Georgia 1973-12-31
GILD	Gilead Sciences	
Biotechnology	Foster City, California	2004-07-01
GPN	Global Payments	Transaction & Pa
yment Processing Services		Atlanta, Georgia 2016-0
4-25		
GL	Globe Life	
Life & Health Insurance		Mckinney, Texas 1989-04-
30		
GDDY	Godaddy	Internet
Services & Infrastructure		Tempe, Arizona 2024-0
6-24		



GS Investment Banking & Brokerage 7-22	Goldman Sachs New York City, New York	Investment 2002-0
HAL Gas Equipment & Services -04	Halliburton Houston, Texas	Oil & 1957-03
HIG Property & Casualty Insurance 3-04	Hartford (The) Hartford, Connecticut	Property 1957-0
HAS Leisure Products	Hasbro Pawtucket, Rhode Island	1984-09-30
HCA Health Care Facilities 7	Hca Healthcare Nashville, Tennessee	2015-01-2
DOC Health Care Reits	Healthpeak Properties Denver, Colorado	2008-03-31
HSIC Health Care Distributors -17	Henry Schein Melville, New York	2015-03
HSY Packaged Foods & Meats 4	Hershey Company (The) Hershey, Pennsylvania	1957-03-0
HES Integrated Oil & Gas	Hess Corporation New York City, New York	1984-05-31
HPE Printer, Storage & Peripherals 1-02	Hewlett Packard Enterprise Houston, Texas	Technology Hardware 2015-1
HLT Hotels, Resorts & Cruise Lines 6-19	Hilton Worldwide Tysons Corner, Virginia	Hotel 2017-0
HOLX Health Care Equipment	Hologic Marlborough, Massachusetts	2016-03-30
HD Home Improvement Retail 31	Home Depot (The) Atlanta, Georgia	1988-03-
HON Industrial Conglomerates -04	Honeywell Charlotte, North Carolina	1957-03
HRL Packaged Foods & Meats 4	Hormel Foods Austin, Minnesota	2009-03-0
HST Hotel & Resort Reits	Host Hotels & Resorts Bethesda, Maryland	2007-03-20
HWM Aerospace & Defense	Howmet Aerospace Pittsburgh, Pennsylvania	2016-10-21
HPQ Printer, Storage & Peripherals 2-31	Hp Inc. Palo Alto, California	Technology Hardware 1974-1
HUBB Printer & Supplies & Components 0-18	Hubbell Incorporated Shelton, Connecticut	Industrial Machinery 2023-1
HUM Managed Health Care	Humana Louisville, Kentucky	2012-12-10
HBAN Regional Banks	Huntington Bancshares Columbus, Ohio; Detroit, Michigan	1997-08-28
HII Huntington Ingalls Industries		

Aerospace & Defense	Newport News, Virginia	2018-01-03
IBM	Ibm	It Co
Consulting & Other Services	Armonk, New York	1957-0
3-04		
IEX	IDEX Corporation	Industrial Machinery & Supplies & Components
8-09		Lake Forest, Illinois 2019-0
IDXX	IDEXX Laboratories	
Health Care Equipment		Westbrook, Maine 2017-01-05
ITW	Illinois Tool Works	Industrial Machinery & Supplies & Components
2-28		Glenview, Illinois 1986-0
INCY	Incyte	
Biotechnology	Wilmington, Delaware	2017-02-28
IR	Ingersoll Rand	Industrial Machinery & Supplies & Components
3-03		Davidson, North Carolina 2020-0
PODD	Insulet Corporation	
Health Care Equipment		Acton, Massachusetts 2023-03-15
INTC	Intel	
Semiconductors	Santa Clara, California	1976-12-31
ICE	Intercontinental Exchange	F
Financial Exchanges & Data		Atlanta, Georgia 2007-0
9-26		
IFF	International Flavors & Fragrances	
Specialty Chemicals	New York City, New York	1976-03-31
IP	International Paper	Paper & Plastic Packaging Products & Materials
3-04		Memphis, Tennessee 1957-0
IPG	Interpublic Group Of Companies (The)	
Advertising	New York City, New York	1992-10-01
INTU	Intuit	
Application Software	Mountain View, California	2000-12-05
ISRG	Intuitive Surgical	
Health Care Equipment	Sunnyvale, California	2008-06-02
IVZ	Invesco	Asset Management & Custody Banks
8-21		Atlanta, Georgia 2008-0
INVH	Invitation Homes	Single-Family Residential Reits
9-19		Dallas, Texas 2022-0
IQV	Iqvia	Life Sciences Tools & Services
8-29		Durham, North Carolina 2017-0
IRM	Iron Mountain	
Other Specialized Reits		Boston, Massachusetts 2009-01-06
JBHT	J.B. Hunt	Logistics Ground Transportation
7-01		Lowell, Arkansas 2015-0
JBL	Jabil	Electronic Manufacturing Services
2-18		St. Petersburg, Florida 2023-1
JKHY	Jack Henry & Associates	Transaction & Payment Processing Services
1-13		Monett, Missouri 2018-1

J	Jacobs Solutions	C
onstruction & Engineering		Dallas, Texas 2007-1
0-26		
JNJ	Johnson & Johnson	
Pharmaceuticals	New Brunswick, New Jersey	1973-06-30
JCI	Johnson Controls	
Building Products		Cork, Ireland 2010-08-27
JPM	Jpmorgan Chase	
Diversified Banks	New York City, New York	1975-06-30
JNPR	Juniper Networks	
Communications Equipment		Sunnyvale, California 2006-06
-02		
K	Kellanova	
Packaged Foods & Meats		Chicago, Illinois 1989-09-1
1		
KVUE	Kenvue	
Personal Care Products		Skillman, New Jersey 2023-08-2
5		
KDP	Keurig Dr Pepper	Soft Drinks
& Non-Alcoholic Beverages		Burlington, Massachusetts 2022-0
6-21		
KEY	Keycorp	
Regional Banks		Cleveland, Ohio 1994-03-01
KEYS	Keysight Technologies	Electroni
c Equipment & Instruments		Santa Rosa, California 2018-1
1-06		
KMB	Kimberly-Clark	
Household Products		Irving, Texas 1957-03-04
KIM	Kimco Realty	
Retail Reits		Jericho, New York 2006-04-04
KMI	Kinder Morgan	Oil & Gas
Storage & Transportation		Houston, Texas 2012-05
-25		
KKR	Kkr & Co.	Asset M
anagement & Custody Banks		New York City, New York 2024-0
6-24		
KLAC	Kla Corporation	Semiconduc
tor Materials & Equipment		Milpitas, California 1997-0
9-30		
KHC	Kraft Heinz	
Packaged Foods & Meats	Chicago, Illinois; Pittsburgh, Pennsylvania	2015-07-0
6		
KR	Kroger	
Food Retail		Cincinnati, Ohio 1957-03-04
LHX	L3Harris	
Aerospace & Defense		Melbourne, Florida 2008-09-22
LH	Labcorp	
Health Care Services		Burlington, North Carolina 2004-11-01
LRCX	Lam Research	Semiconduc
tor Materials & Equipment		Fremont, California 2012-0
6-29		
LW	Lamb Weston	
Packaged Foods & Meats		Eagle, Idaho 2018-12-0
3		
LVS	Las Vegas Sands	
Casinos & Gaming		Las Vegas, Nevada 2019-10-03

LDOS	Leidos	Div
ersified Support Services		Reston, Virginia 2019-0
8-09		
LEN	Lennar	
Homebuilding		Miami, Florida 2005-10-04
LII	Lennox International	
Building Products		Richardson, Texas 2024-12-23
LLY	Lilly (Eli)	
Pharmaceuticals		Indianapolis, Indiana 1970-12-31
LIN	Linde Plc	
Industrial Gases		Guildford, United Kingdom 1992-07-01
LYV	Live Nation Entertainment	
Movies & Entertainment		Beverly Hills, California 2019-12-2
3		
LKQ	Lkq Corporation	
Distributors		Chicago, Illinois 2016-05-23
LMT	Lockheed Martin	
Aerospace & Defense		Bethesda, Maryland 1957-03-04
L	Loews Corporation	
Multi-Line Insurance		New York City, New York 1995-05-31
LOW	Lowe'S	
Home Improvement Retail		Mooresville, North Carolina 1984-02-
29		
LULU	Lululemon Athletica	Apparel, A
ccessories & Luxury Goods		Vancouver, Canada 2023-1
0-18		
LYB	Lyondellbasell	
Specialty Chemicals		Rotterdam, Netherlands 2012-09-05
MTB	M&T Bank	
Regional Banks		Buffalo, New York 2004-02-23
MPC	Marathon Petroleum	Oil &
Gas Refining & Marketing		Findlay, Ohio 2011-07
-01		
MKTX	Marketaxess	F
inancial Exchanges & Data		New York City, New York 2019-0
7-01		
MAR	Marriott International	Hotel
s, Resorts & Cruise Lines		Bethesda, Maryland 1998-0
5-29		
MMC	Marsh McLennan	
Insurance Brokers		New York City, New York 1987-08-31
MLM	Martin Marietta Materials	
Construction Materials		Raleigh, North Carolina 2014-07-0
2		
MAS	Masco	
Building Products		Livonia, Michigan 1981-06-30
MA	Mastercard	Transaction & Pa
yment Processing Services		Harrison, New York 2008-0
7-18		
MTCH	Match Group	Int
eractive Media & Services		Dallas, Texas 2021-0
9-20		
MKC	Mccormick & Company	
Packaged Foods & Meats		Hunt Valley, Maryland 2003-03-2
0		
MCD	Mcdonald'S	

Restaurants	Chicago, Illinois 1970-06-30	
MCK	Mckesson Corporation	
Health Care Distributors		Irving, Texas 1999-01
-13		
MDT	Medtronic	
Health Care Equipment		Dublin, Ireland 1986-10-31
MRK	Merck & Co.	
Pharmaceuticals	Kenilworth, New Jersey 1957-03-04	
META	Meta Platforms	Int
eractive Media & Services		Menlo Park, California 2013-1
2-23		
MET	Metlife	
Life & Health Insurance		New York City, New York 2000-12-
11		
MTD	Mettler Toledo	Life
Sciences Tools & Services		Columbus, Ohio 2016-0
9-06		
MGM	Mgm Resorts	
Casinos & Gaming		Paradise, Nevada 2017-07-26
MCHP	Microchip Technology	
Semiconductors	Chandler, Arizona 2007-09-07	
MU	Micron Technology	
Semiconductors		Boise, Idaho 1994-09-27
MSFT	Microsoft	
Systems Software	Redmond, Washington 1994-06-01	
MAA	Mid-America Apartment Communities	Multi
-Family Residential Reits		Memphis, Tennessee 2016-1
2-02		
MRNA	Moderna	
Biotechnology	Cambridge, Massachusetts 2021-07-21	
MHK	Mohawk Industries	
Home Furnishings		Calhoun, Georgia 2013-12-23
MOH	Molina Healthcare	
Managed Health Care		Long Beach, California 2022-03-02
TAP	Molson Coors Beverage Company	
Brewers	Chicago, Illinois 1976-06-30	
MDLZ	Mondelez International	
Packaged Foods & Meats		Chicago, Illinois 2012-10-0
2		
MPWR	Monolithic Power Systems	
Semiconductors		Kirkland, Washington 2021-02-12
MNST	Monster Beverage	Soft Drinks
& Non-Alcoholic Beverages		Corona, California 2012-0
6-28		
MCO	Moody'S Corporation	F
inancial Exchanges & Data		New York City, New York 1998-0
7-01		
MS	Morgan Stanley	Inves
tment Banking & Brokerage		New York City, New York 1993-0
7-29		
MOS	Mosaic Company (The)	Fertilizers
& Agricultural Chemicals		Tampa, Florida 2011-09
-26		
MSI	Motorola Solutions	
Communications Equipment		Chicago, Illinois 1957-03
-04		

MSCI	Msci Inc.	F
Financial Exchanges & Data	New York City, New York	2018-04-04
NDAQ	Nasdaq, Inc.	F
Financial Exchanges & Data	New York City, New York	2008-10-22
NTAP	Netapp	Technology Hardwa
re, Storage & Peripherals	San Jose, California	1999-06-25
NFLX	Netflix	
Movies & Entertainment	Los Gatos, California	2010-12-20
NEM	Newmont	
Gold	Denver, Colorado	1969-06-30
NWSA	News Corp (Class A)	
Publishing	New York City, New York	2013-08-01
NWS	News Corp (Class B)	
Publishing	New York City, New York	2015-09-18
NEE	Nextera Energy	
Multi-Utilities	Juno Beach, Florida	1976-06-30
NKE	Nike, Inc.	Apparel, A
ccessories & Luxury Goods	Washington County, Oregon	1988-11-30
NI	Nisource	
Multi-Utilities	Merrillville, Indiana	2000-11-02
NDSN	Nordson Corporation	Industrial Machiner
y & Supplies & Components	Westlake, Ohio	2022-02-15
NSC	Norfolk Southern	
Rail Transportation	Atlanta, Georgia	1957-03-04
NTRS	Northern Trust	Asset M
anagement & Custody Banks	Chicago, Illinois	1998-01-30
NOC	Northrop Grumman	
Aerospace & Defense	West Falls Church, Virginia	1957-03-04
NCLH	Norwegian Cruise Line Holdings	Hotel
s, Resorts & Cruise Lines	Miami, Florida	2017-10-13
NRG	Nrg Energy	Independent Power P
roducers & Energy Traders	Houston, Texas	2010-01-29
NUE	Nucor	
Steel	Charlotte, North Carolina	1985-04-30
NVDA	Nvidia	
Semiconductors	Santa Clara, California	2001-11-30
NVR	Nvr, Inc.	
Homebuilding	Reston, Virginia	2019-09-26
NXPI	Nxp Semiconductors	
Semiconductors	Eindhoven, Netherlands	2021-03-22
ORLY	O'Reilly Automotive	
Automotive Retail	Springfield, Missouri	2009-03-27
OXY	Occidental Petroleum	Oil & Gas
Exploration & Production	Houston, Texas	1957-03-04
ODFL	Old Dominion	Ca
rgo Ground Transportation	Thomasville, North Carolina	2019-1

2-09		
OMC	Omnicom Group	
Advertising	New York City, New York	1997-12-31
ON	On Semiconductor	
Semiconductors	Phoenix, Arizona	2022-06-21
OKE	Oneok	Oil & Gas
Storage & Transportation		Tulsa, Oklahoma 2010-03
-15		
ORCL	Oracle Corporation	
Application Software		Austin, Texas 1989-08-31
OTIS	Otis Worldwide	Industrial Machinery
y & Supplies & Components		Farmington, Connecticut 2020-0
4-03		
PCAR	Paccar	Construction Machinery & Heavy
Transportation Equipment		Bellevue, Washington 1980-12
-31		
PKG	Packaging Corporation Of America	Paper & Plastic Packa
ging Products & Materials		Lake Forest, Illinois 2017-0
7-26		
PLTR	Palantir Technologies	
Application Software		Denver, Colorado 2024-09-23
PANW	Palo Alto Networks	
Systems Software		Santa Clara, California 2023-06-20
PARA	Paramount Global	
Movies & Entertainment		New York City, New York 1994-09-3
0		
PH	Parker Hannifin	Industrial Machinery
y & Supplies & Components		Cleveland, Ohio 1985-1
1-30		
PAYX	Paychex	Human Resou
rce & Employment Services		Penfield, New York 1998-1
0-01		
PAYC	Paycom	Human Resou
rce & Employment Services		Oklahoma City, Oklahoma 2020-0
1-28		
PYPL	Paypal	Transaction & Pa
yment Processing Services		San Jose, California 2015-0
7-20		
PNR	Pentair	Industrial Machinery
y & Supplies & Components		Worsley, United Kingdom 2012-1
0-01		
PEP	Pepsico	Soft Drinks
& Non-Alcoholic Beverages		Purchase, New York 1957-0
3-04		
PFE	Pfizer	
Pharmaceuticals		New York City, New York 1957-03-04
PCG	Pg&E Corporation	
Multi-Utilities		Oakland, California 2022-10-03
PM	Philip Morris International	
Tobacco		New York City, New York 2008-03-31
PSX	Phillips 66	Oil &
Gas Refining & Marketing		Houston, Texas 2012-05
-01		
PNW	Pinnacle West Capital	
Multi-Utilities		Phoenix, Arizona 1999-10-04
PNC	Pnc Financial Services	

Diversified Banks	Pittsburgh, Pennsylvania 1988-04-30
P00L	Pool Corporation
Distributors	Covington, Louisiana 2020-10-07
PPG	Ppg Industries
Specialty Chemicals	Pittsburgh, Pennsylvania 1957-03-04
PPL	Ppl Corporation
Electric Utilities	Allentown, Pennsylvania 2001-10-01
PFG	Principal Financial Group
Life & Health Insurance	Des Moines, Iowa 2002-07-22
PG	Procter & Gamble
Personal Care Products	Cincinnati, Ohio 1957-03-04
PGR	Progressive Corporation
erty & Casualty Insurance	Mayfield Village, Ohio 1997-08-04
PLD	Prologis
Industrial Reits	San Francisco, California 2003-07-17
PRU	Prudential Financial
Life & Health Insurance	Newark, New Jersey 2002-07-22
PEG	Public Service Enterprise Group
Electric Utilities	Newark, New Jersey 1957-03-04
PTC	Ptc Inc.
Application Software	Boston, Massachusetts 2021-04-20
PSA	Public Storage
Self-Storage Reits	Glendale, California 2005-08-19
PHM	Pultegroup
Homebuilding	Atlanta, Georgia 1984-04-30
PWR	Quanta Services
onstruction & Engineering	Houston, Texas 2009-07-01
QCOM	Qualcomm
Semiconductors	San Diego, California 1999-07-22
DGX	Quest Diagnostics
Health Care Services	Secaucus, New Jersey 2002-12-12
RL	Ralph Lauren Corporation
ccessories & Luxury Goods	New York City, New York 2007-02-02
RJF	Raymond James Financial
tment Banking & Brokerage	St. Petersburg, Florida 2017-03-20
RTX	Rtx Corporation
Aerospace & Defense	Waltham, Massachusetts 1957-03-04
0	Realty Income
Retail Reits	San Diego, California 2015-04-07
REG	Regency Centers
Retail Reits	Jacksonville, Florida 2017-03-02
REGN	Regeneron Pharmaceuticals
Biotechnology	Tarrytown, New York 2013-05-01
RF	Regions Financial Corporation
Regional Banks	Birmingham, Alabama 1998-08-28
RSG	Republic Services
tal & Facilities Services	Phoenix, Arizona 2008-12-05
RMD	Resmed



Health Care Equipment	San Diego, California	2017-07-26
RVTY	Revvity	
Health Care Equipment	Waltham, Massachusetts	1985-05-31
ROK	Rockwell Automation	Electric
al Components & Equipment	Milwaukee, Wisconsin	2000-0
3-12		
ROL	Rollins, Inc.	Environmen
tal & Facilities Services	Atlanta, Georgia	2018-1
0-01		
ROP	Roper Technologies	Electroni
c Equipment & Instruments	Sarasota, Florida	2009-1
2-23		
ROST	Ross Stores	
Apparel Retail	Dublin, California	2009-12-21
RCL	Royal Caribbean Group	Hotel
s, Resorts & Cruise Lines	Miami, Florida	2014-1
2-05		
SPGI	S&P Global	F
inancial Exchanges & Data	New York City, New York	1957-0
3-04		
CRM	Salesforce	
Application Software	San Francisco, California	2008-09-15
SBAC	Sba Communications	
Telecom Tower Reits	Boca Raton, Florida	2017-09-01
SLB	Schlumberger	Oil &
Gas Equipment & Services	Houston, Texas	1957-03
-04		
STX	Seagate Technology	Technology Hardwa
re, Storage & Peripherals	Dublin, Ireland	2012-0
7-02		
SRE	Sempra	
Multi-Utilities	San Diego, California	2017-03-17
NOW	Servicenow	
Systems Software	Santa Clara, California	2019-11-21
SHW	Sherwin-Williams	
Specialty Chemicals	Cleveland, Ohio	1964-06-30
SPG	Simon Property Group	
Retail Reits	Indianapolis, Indiana	2002-06-26
SWKS	Skyworks Solutions	
Semiconductors	Irvine, California	2015-03-12
SJM	J.M. Smucker Company (The)	
Packaged Foods & Meats	Orrville, Ohio	2008-11-0
6		
SW	Smurfit Westrock	Paper & Plastic Packa
ging Products & Materials	Dublin, Ireland	2024-0
7-08		
SNA	Snap-On	Industrial Machiner
y & Supplies & Components	Kenosha, Wisconsin	1982-0
9-30		
SOLV	Solventum	
Health Care Technology	Saint Paul, Minnesota	2024-04-0
1		
S0	Southern Company	
Electric Utilities	Atlanta, Georgia	1957-03-04
LUV	Southwest Airlines	
Passenger Airlines	Dallas, Texas	1994-07-01

SWK	Stanley Black & Decker	Industrial Machinery & Supplies & Components	New Britain, Connecticut 1982-09-30
SBUX	Starbucks	Restaurants	Seattle, Washington 2000-06-07
STT	State Street Corporation	Asset Management & Custody Banks	Boston, Massachusetts 2003-03-14
STLD	Steel Dynamics	Steel	Fort Wayne, Indiana 2022-12-22
STE	Steris	Health Care Equipment	Dublin, Ireland 2019-12-23
SYK	Stryker Corporation	Health Care Equipment	Kalamazoo, Michigan 2000-12-12
SMCI	Supermicro	Technology Hardware, Storage & Peripherals	San Jose, California 2024-03-18
SYF	Synchrony Financial	Consumer Finance	Stamford, Connecticut 2015-11-18
SNPS	Synopsys	Application Software	Sunnyvale, California 2017-03-16
SYY	Sysco	Food Distributors	Houston, Texas 1986-12-31
TMUS	T-Mobile Us	Wireless Telecommunication Services	Bellevue, Washington 2019-07-15
TROW	T. Rowe Price	Asset Management & Custody Banks	Baltimore, Maryland 2019-07-29
TTWO	Take-Two Interactive	Interactive Home Entertainment	New York City, New York 2018-03-19
TPR	Tapestry, Inc.	Apparel, Accessories & Luxury Goods	New York City, New York 2004-09-01
TRGP	Targa Resources	Oil & Gas Storage & Transportation	Houston, Texas 2022-10-12
TGT	Target Corporation	Consumer Staples Merchandise Retail	Minneapolis, Minnesota 1976-12-31
TEL	Te Connectivity	Electronic Manufacturing Services	Galway, Ireland 2011-10-17
TDY	Teledyne Technologies	Electronic Equipment & Instruments	Thousand Oaks, California 2020-06-22
TER	Teradyne	Semiconductor Materials & Equipment	North Reading, Massachusetts 2020-09-21
TSLA	Tesla, Inc.	Automobile Manufacturers	Austin, Texas 2020-12-21
TXN	Texas Instruments	Semiconductors	Dallas, Texas 2001-03-12
TPL	Texas Pacific Land Corporation	Oil & Gas	

Exploration & Production		Dallas, Texas 2024-11
-26		
TXT	Textron	
Aerospace & Defense	Providence, Rhode Island	1978-12-31
TMO	Thermo Fisher Scientific	Life
Sciences Tools & Services	Waltham, Massachusetts	2004-0
8-03		
TJX	Tjx Companies	
Apparel Retail	Framingham, Massachusetts	1985-09-30
TKO	Tko Group Holdings	
Movies & Entertainment	New York City, New York	2025-03-2
4		
TSCO	Tractor Supply	
Other Specialty Retail	Brentwood, Tennessee	2014-01-2
4		
TT	Trane Technologies	
Building Products	Dublin, Ireland	2010-11-17
TDG	Transdigm Group	
Aerospace & Defense	Cleveland, Ohio	2016-06-03
TRV	Travelers Companies (The)	Prop
erty & Casualty Insurance	New York City, New York	2002-0
8-21		
TRMB	Trimble Inc.	Electroni
c Equipment & Instruments	Westminster, Colorado	2021-0
1-21		
TFC	Truist Financial	
Diversified Banks	Charlotte, North Carolina	1997-12-04
TYL	Tyler Technologies	
Application Software	Plano, Texas	2020-06-22
TSN	Tyson Foods	
Packaged Foods & Meats	Springdale, Arkansas	2005-08-1
0		
USB	U.S. Bancorp	
Diversified Banks	Minneapolis, Minnesota	1999-11-01
UBER	Uber	Passen
ger Ground Transportation	San Francisco, California	2023-1
2-18		
UDR	Udr, Inc.	Multi
-Family Residential Reits	Highlands Ranch, Colorado	2016-0
3-07		
ULTA	Ulta Beauty	
Other Specialty Retail	Bolingbrook, Illinois	2016-04-1
8		
UNP	Union Pacific Corporation	
Rail Transportation	Omaha, Nebraska	1957-03-04
UAL	United Airlines Holdings	
Passenger Airlines	Chicago, Illinois	2015-09-03
UPS	United Parcel Service	
Air Freight & Logistics	Sandy Springs, Georgia	2002-07-
22		
URI	United Rentals	Trading
Companies & Distributors	Stamford, Connecticut	2014-09
-20		
UNH	Unitedhealth Group	
Managed Health Care	Minnetonka, Minnesota	1994-07-01
UHS	Universal Health Services	

Health Care Facilities 0	King Of Prussia, Pennsylvania 2014-09-2	
VLO Gas Refining & Marketing -20	Valero Energy San Antonio, Texas 2002-12	Oil &
VTR Health Care Reits	Ventas Chicago, Illinois 2009-03-04	
VLTO tal & Facilities Services 0-02	Veralto Waltham, Massachusetts 2023-1	Environmen
VRSN Services & Infrastructure 2-01	Verisign Dulles, Virginia 2006-0	Internet
VRSK rch & Consulting Services 0-08	Verisk Analytics Jersey City, New Jersey 2015-1	Resea
VZ elecommunication Services 1-30	Verizon New York City, New York 1983-1	Integrated T
VRTX Biotechnology	Vertex Pharmaceuticals Boston, Massachusetts 2013-09-23	
VTRS Pharmaceuticals	Viatris Pittsburgh, Pennsylvania 2004-04-23	
VICI Hotel & Resort Reits	Vici Properties New York City, New York 2022-06-08	
V yment Processing Services 2-21	Visa Inc. San Francisco, California 2009-1	Transaction & Pa
VST Electric Utilities	Vistra Corp. Irving, Texas 2024-05-08	
VMC Construction Materials 0	Vulcan Materials Company Birmingham, Alabama 1999-06-3	
WRB erty & Casualty Insurance 2-05	W. R. Berkley Corporation Greenwich, Connecticut 2019-1	Prop
GW y & Supplies & Components 6-30	W. W. Grainger Lake Forest, Illinois 1981-0	Industrial Machiner
WAB Transportation Equipment -27	Wabtec Construction Machinery & Heavy Pittsburgh, Pennsylvania 2019-02	
WBA Drug Retail	Walgreens Boots Alliance Deerfield, Illinois 1979-12-31	
WMT taples Merchandise Retail 8-31	Walmart Bentonville, Arkansas 1982-0	Consumer S
DIS Movies & Entertainment 0	Walt Disney Company (The) Burbank, California 1976-06-3	
WBD Broadcasting	Warner Bros. Discovery New York City, New York 2022-04-11	
WM tal & Facilities Services 8-31	Waste Management Houston, Texas 1998-0	Environmen
WAT	Waters Corporation	Life

Sciences Tools & Services	Milford, Massachusetts	2002-01-02
WEC	Wec Energy Group	
Electric Utilities	Milwaukee, Wisconsin	2008-10-31
WFC	Wells Fargo	
Diversified Banks	San Francisco, California	1976-06-30
WELL	Welltower	
Health Care Reits	Toledo, Ohio	2009-01-30
WST	West Pharmaceutical Services	
Health Care Supplies	Exton, Pennsylvania	2020-05-22
WDC	Western Digital	
re, Storage & Peripherals	Technology Hardwa	
7-01	San Jose, California	2009-0
WY	Weyerhaeuser	
Timber Reits	Seattle, Washington	1979-10-01
WSM	Williams-Sonoma, Inc.	
Homefurnishing Retail	San Francisco, California	2025-03-24
WMB	Williams Companies	
Storage & Transportation	Oil & Gas	
-31	Tulsa, Oklahoma	1975-03
WTW	Willis Towers Watson	
Insurance Brokers	London, United Kingdom	2016-01-05
WDAY	Workday, Inc.	
Application Software	Pleasanton, California	2024-12-23
WYNN	Wynn Resorts	
Casinos & Gaming	Paradise, Nevada	2008-11-14
XEL	Xcel Energy	
Multi-Utilities	Minneapolis, Minnesota	1957-03-04
XYL	Xylem Inc.	
y & Supplies & Components	Industrial Machiner	
1-01	White Plains, New York	2011-1
YUM	Yum! Brands	
Restaurants	Louisville, Kentucky	1997-10-06
ZBRA	Zebra Technologies	
c Equipment & Instruments	Electroni	
2-23	Lincolnshire, Illinois	2019-1
ZBH	Zimmer Biomet	
Health Care Equipment	Warsaw, Indiana	2001-08-07
ZTS	Zoetis	
Pharmaceuticals	Parsippany, New Jersey	2013-06-21

In [50]: *#Final API data set after cleaning*

```
print(api_df.head)
api_data = api_df

print(api_df[:5]) # Display a sample of cleaned data
```

```

<bound method NDFrame.head of
Close      Volume      Ticker      Date      Open      High      Low
0      MMM  2025-05-30  148.73  149.45  147.60  148.35  5339916
1      MMM  2025-05-29  149.87  150.36  148.18  149.63  2621435
2      MMM  2025-05-28  149.65  150.00  148.19  148.66  2297543
3      MMM  2025-05-27  150.00  150.00  146.66  149.49  3155330
4      MMM  2025-05-23  147.45  148.69  147.01  147.62  2369839
...
2295   AMZN  2025-01-13  218.06  219.40  216.47  218.46  27262655
2296   AMZN  2025-01-10  221.46  221.71  216.50  218.94  36811525
2297   AMZN  2025-01-08  223.19  223.52  220.20  222.13  25033292
2298   AMZN  2025-01-07  227.90  228.38  221.46  222.11  28084164
2299   AMZN  2025-01-06  226.78  228.84  224.84  227.61  31849831

```

```
[2300 rows x 7 columns]>
```

```

      Ticker      Date      Open      High      Low      Close      Volume
0      MMM  2025-05-30  148.73  149.45  147.60  148.35  5339916
1      MMM  2025-05-29  149.87  150.36  148.18  149.63  2621435
2      MMM  2025-05-28  149.65  150.00  148.19  148.66  2297543
3      MMM  2025-05-27  150.00  150.00  146.66  149.49  3155330
4      MMM  2025-05-23  147.45  148.69  147.01  147.62  2369839

```

```
In [54]: #APT data frame
api_df.head()
```

```
Out[54]:
```

	Ticker	Date	Open	High	Low	Close	Volume
0	MMM	2025-05-30	148.73	149.45	147.60	148.35	5339916
1	MMM	2025-05-29	149.87	150.36	148.18	149.63	2621435
2	MMM	2025-05-28	149.65	150.00	148.19	148.66	2297543
3	MMM	2025-05-27	150.00	150.00	146.66	149.49	3155330
4	MMM	2025-05-23	147.45	148.69	147.01	147.62	2369839

```
In [56]: wiki_data.head()
```

```
Out[56]:
```

	Ticker	Company_Name	Industry_Sector	Industry	Date_Added
0	MMM	3M	Industrial Conglomerates	Saint Paul, Minnesota	1957-03-04
1	AOS	A. O. Smith	Building Products	Milwaukee, Wisconsin	2017-07-26
2	ABT	Abbott Laboratories	Health Care Equipment	North Chicago, Illinois	1957-03-04
3	ABBV	Abbvie	Biotechnology	North Chicago, Illinois	2012-12-31
4	ACN	Accenture	It Consulting & Other Services	Dublin, Ireland	2011-07-06

```
In [58]: final_data.head()
```

Out [58]:

	Ticker	Company_Name_x	Industry_Sector_x	Stock_Price	PE_Ratio	Dividend_Yiel
0	MMM	3M Company	Industrials	222.89	24.31	2.3
1	MMM	3M Company	Industrials	222.89	24.31	2.3
2	MMM	3M Company	Industrials	222.89	24.31	2.3
3	MMM	3M Company	Industrials	222.89	24.31	2.3
4	MMM	3M Company	Industrials	222.89	24.31	2.3

5 rows × 31 columns

## Merging datasets, storing them in SQLite, performing SQL joins, and visualizing data in a single cell.

```
In [60]: import sqlite3
import pandas as pd
```

```
In [62]: # Connect to SQLite database (or create if it doesn't exist)
conn = sqlite3.connect("sp500_data.db")
cursor = conn.cursor()

# Define table schemas
cursor.execute("""
CREATE TABLE IF NOT EXISTS kaggle_data (
    Ticker TEXT PRIMARY KEY,
    Company_Name TEXT,
    Industry_Sector TEXT,
    Stock_Price REAL,
    PE_Ratio REAL,
    Dividend_Yield REAL,
    Earnings_Per_Share REAL,
    Market_Cap TEXT,
    EBITDA_Value REAL
);
""")

cursor.execute("""
CREATE TABLE IF NOT EXISTS wikipedia_data (
    Ticker TEXT PRIMARY KEY,
```

```

        Company_Name TEXT,
        Industry_Sector TEXT,
        Headquarters TEXT,
        Date_Added TEXT
    );
    """

cursor.execute("""
CREATE TABLE IF NOT EXISTS api_data (
    Ticker TEXT,
    Date TEXT,
    Open_Price REAL,
    High_Price REAL,
    Low_Price REAL,
    Close_Price REAL,
    Trading_Volume INTEGER
);
""")
conn.commit()

```

```

In [150]: #Final clean up for safety
import pandas as pd

# Standardize Kaggle dataset column names
final_data.rename(columns={
    "Company_Name_x": "Company_Name",
    "Industry_Sector_x": "Industry_Sector",
    "Stock_Price": "Stock_Price",
    "PE_Ratio": "PE_Ratio",
    "Market_Cap": "Market_Cap",
}, inplace=True)

# Standardize Wikipedia dataset column names
wiki_data.rename(columns={
    "Industry": "Headquarters",
}, inplace=True)

# Standardize API dataset column names
api_df.rename(columns={
    "Open": "Open_Price",
    "High": "High_Price",
    "Low": "Low_Price",
    "Close": "Close_Price",
    "Volume": "Trading_Volume"
}, inplace=True)

# Convert 'Date' column in API data to datetime format for consistency
api_df["Date"] = pd.to_datetime(api_df["Date"])

```

```

In [82]: """final_data.head()
api_df.head()
wiki_data.head() """

# Insert Kaggle data

```



```

final_data.to_sql("kaggle_data", conn, if_exists="replace", index=False)

# Insert Wikipedia data
wiki_data.to_sql("wikipedia_data", conn, if_exists="replace", index=False)

# Insert API stock market data
api_df.to_sql("api_data", conn, if_exists="replace", index=False)

print("All datasets successfully stored in SQLite.")

```

All datasets successfully stored in SQLite.

```

In [86]: import sqlite3

# Connect to the database
conn = sqlite3.connect("sp500_data.db")
cursor = conn.cursor()

# Define the SQL query
query = """
SELECT k.Ticker, k.Company_Name, k.Industry_Sector, k.Stock_Price, k.PE_Ratio,
       w.Headquarters, w.Date_Added, a.Close_Price, a.Trading_Volume
FROM kaggle_data k
JOIN wikipedia_data w ON k.Ticker = w.Ticker
JOIN api_data a ON k.Ticker = a.Ticker;
"""

# Execute the query and fetch the results
cursor.execute(query)
result = cursor.fetchall()

# Convert result into a Pandas DataFrame for easy visualization
import pandas as pd
columns = ["Ticker", "Company_Name", "Industry_Sector", "Stock_Price", "PE_Ratio",
           "Headquarters", "Date_Added", "Close_Price", "Trading_Volume"]
df = pd.DataFrame(result, columns=columns)

# Display the first few rows of the merged dataset
print(df.head())

# Close the database connection
conn.close()

```

	Ticker	Company_Name	Industry_Sector	Stock_Price	PE_Ratio	\
0	MMM	3M Company	Industrials	222.89	24.31	
1	MMM	3M Company	Industrials	222.89	24.31	
2	MMM	3M Company	Industrials	222.89	24.31	
3	MMM	3M Company	Industrials	222.89	24.31	
4	MMM	3M Company	Industrials	222.89	24.31	

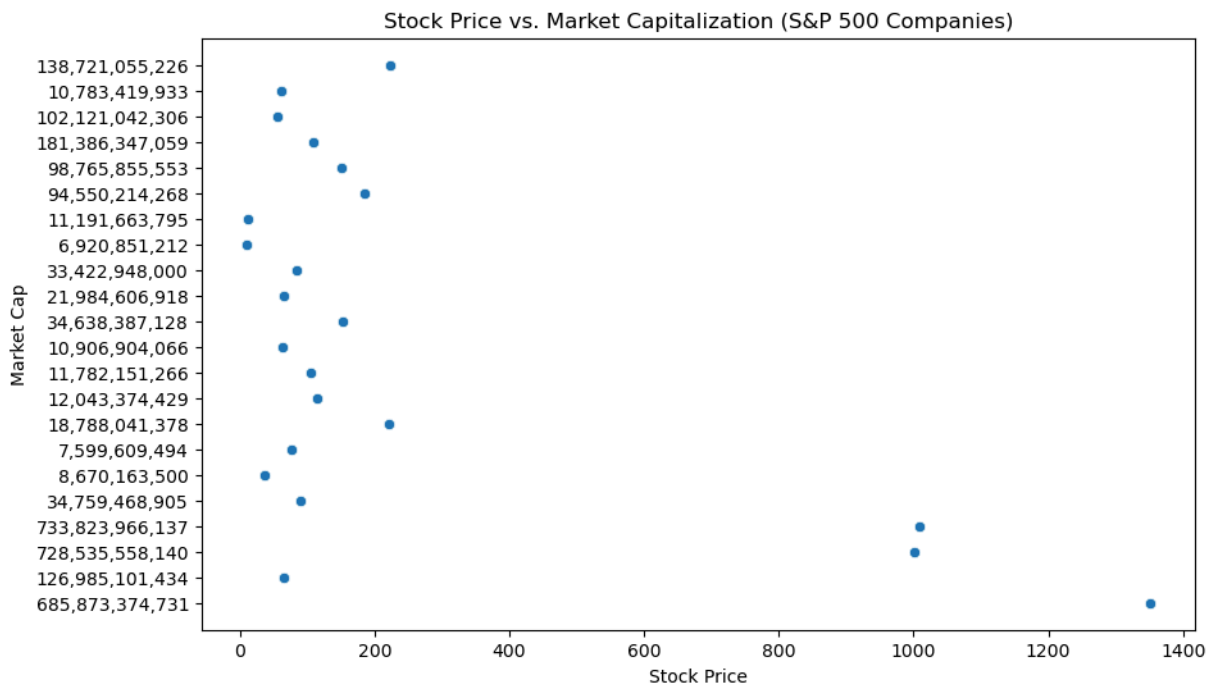
  

		Headquarters	Date_Added	Close_Price	Trading_Volume
0	Saint Paul, Minnesota	1957-03-04 00:00:00	126.09	5971823	
1	Saint Paul, Minnesota	1957-03-04 00:00:00	126.91	9609106	
2	Saint Paul, Minnesota	1957-03-04 00:00:00	127.16	5497234	
3	Saint Paul, Minnesota	1957-03-04 00:00:00	128.55	7174970	
4	Saint Paul, Minnesota	1957-03-04 00:00:00	130.21	4952015	

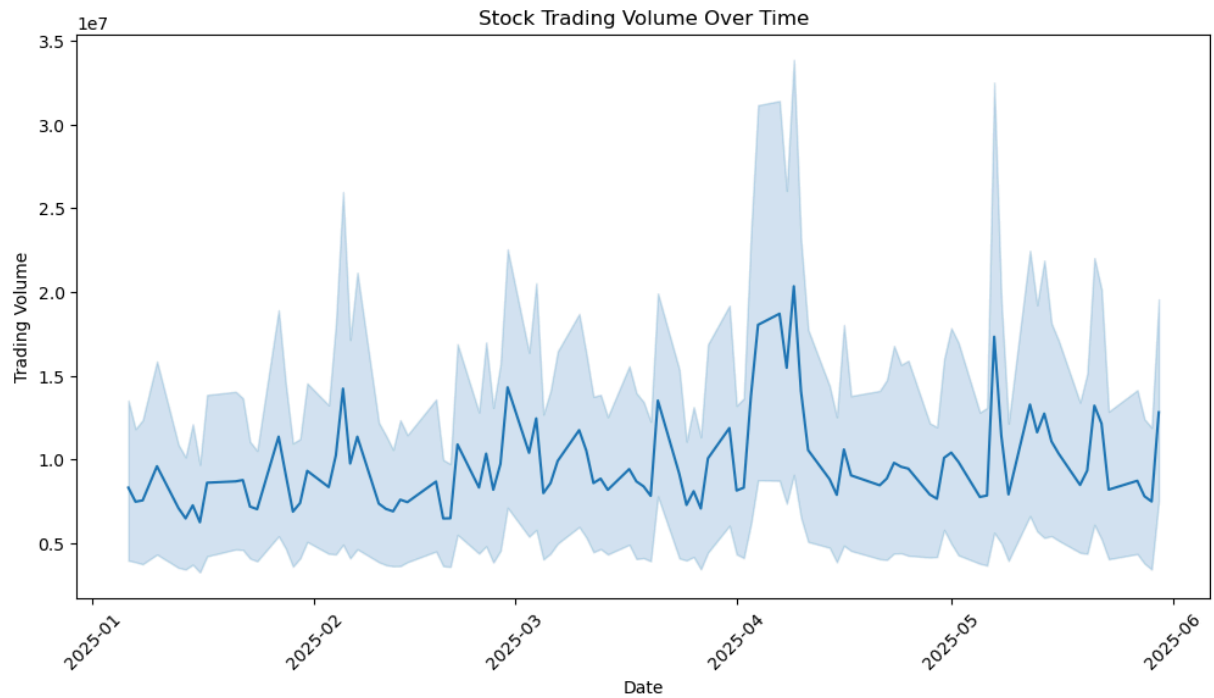
```
In [92]: #Visualizations Using Matplotlib & Seaborn
#we'll create 5 visualizations using different datasets.
#Plot 1: Stock Prices vs. Market Capitalization (Kaggle)

import matplotlib.pyplot as plt
import seaborn as sns

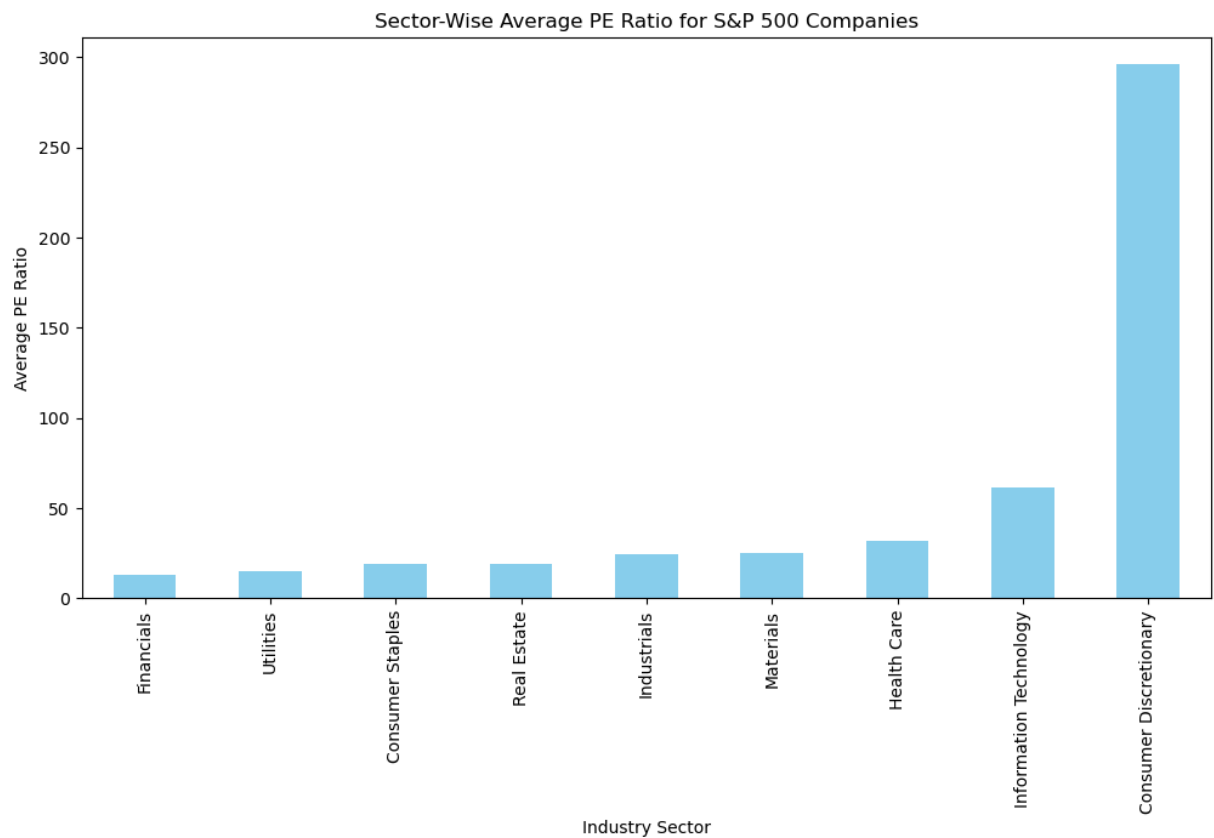
plt.figure(figsize=(10,6))
sns.scatterplot(x=final_data["Stock_Price"], y=final_data["Market_Cap"])
plt.xlabel("Stock Price")
plt.ylabel("Market Cap")
plt.title("Stock Price vs. Market Capitalization (S&P 500 Companies)")
plt.show()
```



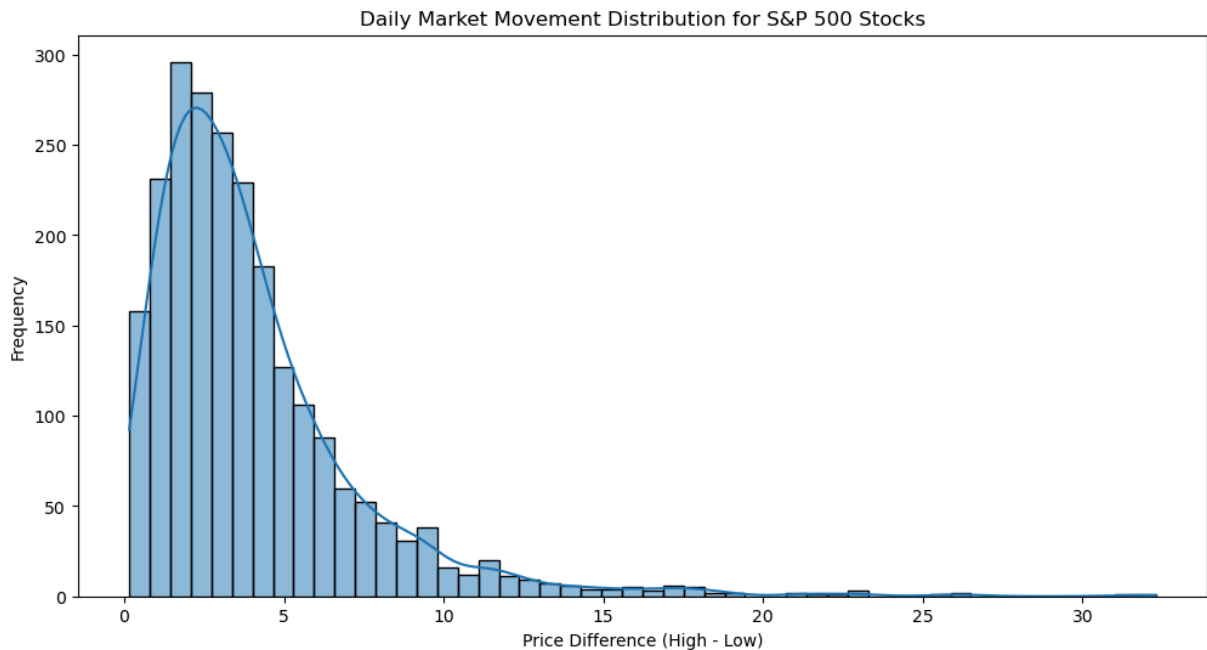
```
In [94]: #Plot 2: Stock Volume Trends Over Time (API)
plt.figure(figsize=(12,6))
sns.lineplot(x=api_df["Date"], y=api_df["Trading_Volume"])
plt.xlabel("Date")
plt.ylabel("Trading Volume")
plt.title("Stock Trading Volume Over Time")
plt.xticks(rotation=45)
plt.show()
```



```
In [96]: #Plot 3: Sector-Wise Average PE Ratio (Merged Data)
sector_avg_pe = final_data.groupby("Industry_Sector")["PE_Ratio"].mean().sort_values()
plt.figure(figsize=(12,6))
sector_avg_pe.plot(kind="bar", color="skyblue")
plt.xlabel("Industry Sector")
plt.ylabel("Average PE Ratio")
plt.title("Sector-Wise Average PE Ratio for S&P 500 Companies")
plt.show()
```



```
In [98]: #Plot 4: Daily Market Movement (API + Kaggle)
api_df["Daily_Market_Movement"] = api_df["High_Price"] - api_df["Low_Price"]
plt.figure(figsize=(12,6))
sns.histplot(api_df["Daily_Market_Movement"], bins=50, kde=True)
plt.xlabel("Price Difference (High - Low)")
plt.ylabel("Frequency")
plt.title("Daily Market Movement Distribution for S&P 500 Stocks")
plt.show()
```



```
In [140... # Increase figure size for better readability
plt.figure(figsize=(20, 10))

# Create scatter plot with adjusted labels and color coding
sns.scatterplot(
    x=wiki_data["Headquarters"],
    y=final_data["Industry_Sector"],
    hue=final_data["Industry_Sector"],
    alpha=0.7,
    palette="Set2"
)

# Improve label readability
plt.xticks(rotation=90, ha="right", fontsize=8) # Rotate x-axis labels for
plt.yticks(fontsize=12) # Increase font size for industry sector labels
plt.xlabel("Headquarters Location", fontsize=14)
plt.ylabel("Industry Sector", fontsize=14)
plt.title("Industry Sector Distribution by Headquarters Location", fontsize=

# Adjust layout to prevent overlapping labels
plt.tight_layout()

# Show plot
plt.show()
```



## Project Summary & Ethical Implications

What I Learned from This Project Key Learnings from the Project This project involved data wrangling, integration, and visualization using three datasets—Kaggle financials, Wikipedia metadata and Alpha Vantage API stock market data—to analyze S&P 500 companies. The process required extensive data cleaning, resolving inconsistencies, ensuring uniform formatting, and conducting SQL joins to merge the datasets into a structured format. The final consolidated dataset enabled insightful visualizations that highlighted stock market trends, sector-wise performance, and financial indicators across different companies.

Through this experience, I learned how to:

Handle multi-source data integration challenges (column misalignment, missing entries, API restrictions).

Perform fuzzy matching techniques to align company names.

Create meaningful financial metrics such as Daily Market Movement, Price Performance Index, and Sector-Average PE Ratio.

Store and manipulate data efficiently in an SQL database for structured queries and analysis.

## Changes Made to the Data

Throughout the project, multiple transformations were performed to enhance data usability:

## **Column Renaming & Standardization**

Adjusted naming conventions across Kaggle, Wikipedia, and API datasets for consistency.

## **Handling Missing Values**

Used median imputation for Market Cap inconsistencies and mean replacement for missing PE Ratios.

## **Data Formatting**

Converted Market Cap to readable numeric formatting (removing scientific notation).

## **Merging Datasets**

Unified financial metrics and company metadata using SQL joins, ensuring proper data alignment.

## **Feature Engineering**

Created new variables like Dividend-to-Earnings Ratio and Volume-to-Market Cap Ratio for deeper insights.

**Legal & Regulatory Considerations** Data used in this project comes from publicly available sources, but it carries ethical and legal implications:

### **Wikipedia & Kaggle Data:**

Public domain, but subject to data accuracy concerns (as they are community-maintained).

### **Alpha Vantage API Usage:**

Must follow terms of service and rate limits to prevent misuse.

SEC & Financial Reporting Standards:

Since the dataset includes financial information, any interpretation should align with investment regulations (avoiding misleading financial advice).

**Potential Risks from Data Transformations** Misalignment of Historical vs. Real-Time Data

The Kaggle dataset contains historical financial metrics, whereas the Alpha Vantage API delivers real-time market data. Merging these sources without careful synchronization may introduce inaccuracies.

## **Data Bias in Company Representation**

The datasets favor large-cap companies with abundant available data, potentially neglecting smaller firms.

#### Fuzzy Matching Errors

Some companies may have multiple ticker symbols due to mergers, leading to incorrect mapping across datasets.

#### Use in Investment Decisions

The analysis should not be interpreted as financial advice, as stock prices are dynamic and influenced by external factors.

**Assumptions Made in Data Cleaning & Transformation** Several assumptions were necessary:

Company names with slight variations were considered matches using fuzzy text processing.

Market Cap median imputation assumes a normal distribution of financial values.

API stock data was considered valid despite potential data gaps due to rate limitations.

#### Credibility of Data Sources To ensure reliability:

Wikipedia data was cross-checked with Kaggle financials to minimize discrepancies.

API financial data was verified by comparing extracted stock prices with external finance sources.

SQL-based integration ensured structured merging without duplicate data conflicts.

**Ethical Acquisition & Handling of Data** All datasets were obtained through publicly available sources, adhering to ethical principles:

No personal or sensitive financial information was extracted.

Data transformations aimed at accuracy and transparency (not manipulation for misleading conclusions).

Open-source data integrity checks were conducted to ensure responsible analysis.

#### Mitigating Ethical Concerns Transparency in Data Cleaning Methods

Clearly documenting assumptions, imputed values, and merge strategies ensures that findings are traceable and reproducible.

#### Cross-validation with External Financial Sources

Comparing API-based stock prices with official records helps confirm accuracy.

### Preventing Misleading Financial Interpretation

Emphasizing that insights are informational rather than predictive analysis avoids unethical financial influence.

### Handling Bias in Industry Representation

Ensuring balanced sector analysis rather than favoring large-cap companies improves dataset fairness.

## Final Thoughts

This project reinforced the importance of structured data integration, careful cleansing, and ethical considerations in financial analysis. While merging multi-source data improves analytical depth, it also requires responsible handling to mitigate biases and inaccuracies. Ensuring transparent methodologies and proper verification guarantees that findings remain valid, ethical, and legally sound.

In [ ]: