CSC1143 DATA MANAGEMENT & VISUALISATION

STOCK MARKET VOLATILITY AND SECTOR DYNAMICS

Abhishek Malaviya and Kishore Magendiran

ABSTRACT

The stock market has undergone profound changes from 1962 to the present. This project focuses on stock market volatility and sector dynamics by analyzing key measures, such as tickers, trading volumes, sectors, and annual data. Time series analysis identifies volatile periods that have influenced market behavior over time.

The study also examines sector-specific dynamics impacting general market performance, highlighting trends over decades, including resilience, emerging opportunities, and key drivers of stability and change.

This analysis of historical stock market data benefits financial analysts, investors, and researchers by offering insights into the factors shaping market trends and sectoral shifts. These findings provide a deeper understanding of the dynamics guiding market performance and inform better decision-making.

DATA COLLECTION

The Dataset:

The dataset that we have used is "*9000+ Tickers of Stock Market Data"*, which is an open dataset that is available on **Kaggle**. It is publicly available without any restrictions. The data is regularly updated. We utilized Python libraries such as *numpy, pandas, and yfinance* to preprocess and integrate additional information into the dataset. The *time* library was used to manage delays between requests while fetching sector information, and *concurrent.futures* was implemented to parallelize this process, ensuring efficient data retrieval.

The 3 'V's:

Our dataset incorporates all three of the big data features.

Volume:

The size of the dataset is 4.44 GB. It has 34,646,258 rows and 11 variables that contain the values of Date, OHLCV(Open, High, Low, Close, Volume), Tickers, Sectors, etc. along with stock tickers, sector information, and other relevant financial data.

Velocity:

The dataset we used is updated frequently and while preprocessing it, we fetched additional data(sector information) from yFinance and integrated it with the existing data.

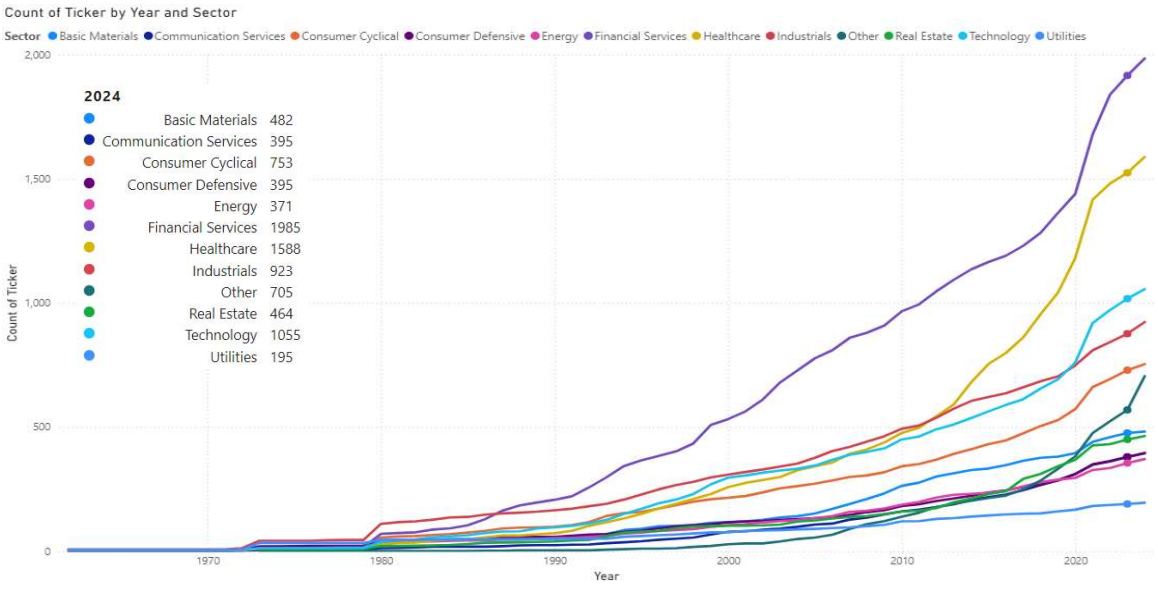
Variety:

The dataset includes a wide range of data, from numerical values, such as OHLCV (open, high, low, close, volume and even percentage change) to categorical data, including stock tickers. Further, the dataset was enhanced by adding sector information from Yahoo Finance for a comprehensive overview of the stock market in various industries.

DATA EXPLORATION, PROCESSING, CLEANING / INTEGRATION

Data Exploration:

The dataset spans over six decades, from 1962 to 2024, with more than 34 million rows and 11 columns. The columns include Date, Ticker, Open, High, Low, Close, Volume, Dividends, Stock Splits, Sector, and Pct Change. During exploration, the distribution of stock prices (Close), trading volume, and other key variables was analyzed across different time periods and sectors. Null values were detected in numeric columns (Open, High, Low, Close, and Volume), and missing sector information for tickers was noted. This exploration revealed patterns in data trends, such as an increase in stock activity in recent decades, aligning with global market growth. Exploratory visualization highlighted outliers, such as unusual spikes in volume or missing data clusters for certain periods. During exploration, we found out that the dataset contained more than 9000 unique tickers (Companies listed on the Stock Market).

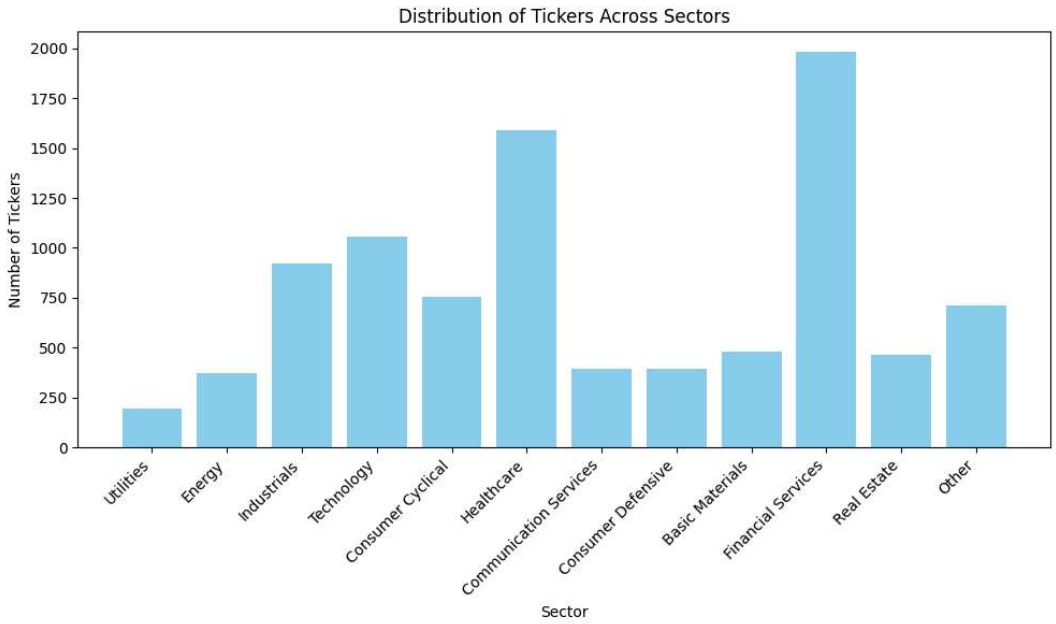


Data Cleaning:

The Date column was converted to a datetime64 format, ensuring seamless time-based operations like filtering and grouping. To handle missing data in numeric fields (Open, High, Low, Close), a time-based interpolation method was applied, leveraging the temporal nature of stock data to fill gaps and N/A values accurately. We couldn't use forward filling, backward filling, mean, median or dropping rows consisting N/A values as it would introduce more noise and would make the data inaccurate. Calculations were performed to generate new metrics, such as the Pct Change column, which measures the daily percentage change in stock prices, offering insights into market volatility. The processed dataset was indexed by Date to enable efficient time-series analysis and grouping by year for aggregate insights.

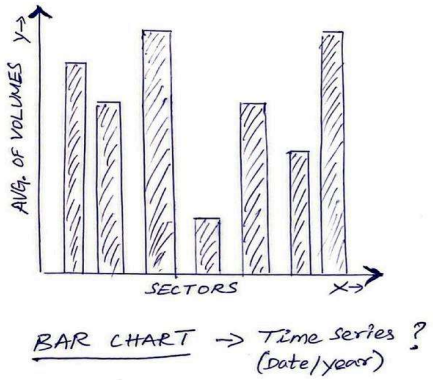
Data Integration:

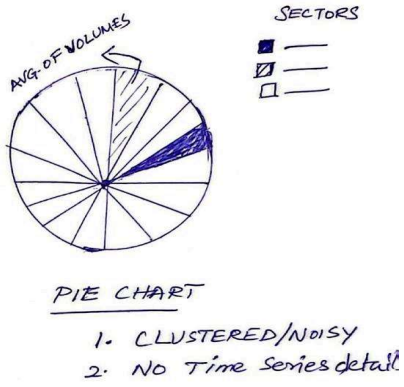
Sector information for each stock ticker was integrated using Python's yfinance library, with multithreading implemented via *concurrent.futures.ThreadPoolExecutor* to fetch data for all tickers in parallel using up to 5 threads. This significantly reduced runtime compared to sequential processing. A 2-second delay per request, managed with the time library, ensured compliance with API usage guidelines. Missing or invalid sector information was handled with exception handling, categorizing it as "Other" to ensure robustness. Tickers were mapped to sectors and added as a new column for sector-wise analysis. Dividends and Stock Splits data were also included to account for corporate actions impacting stock prices, providing a comprehensive view of market behavior. This enriched dataset enabled more detailed visualizations and insights into sectoral performance and market trends.



VISUALISATION

Initially, we tried to use a bar chart to represent the Average volume of trades by sector. However, it was not efficient in showing the trend across time. Then we tried a pie chart, but again it was not ideal because the visualization became too noisy and did not clearly show the insights we wanted to point out.

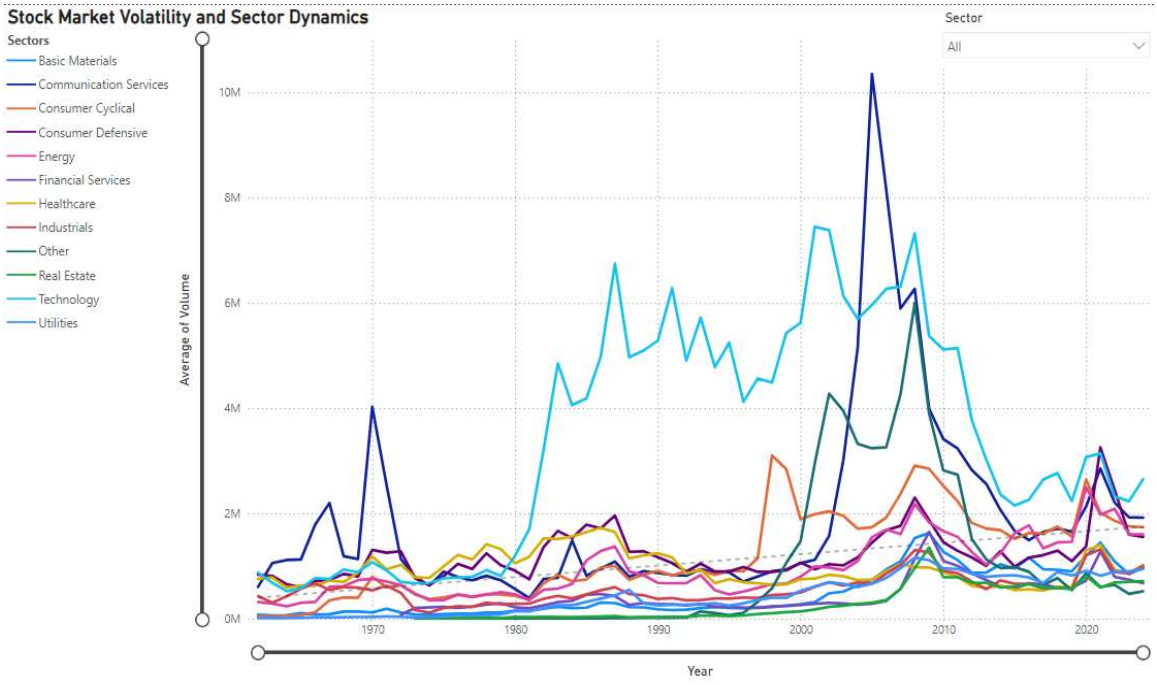




As we reviewed our first approach, we realized that adding the time frame data to both chart types was not useful. In particular, the pie chart became too crowded and failed to show the changes across time.

For this reason, a line graph is used. By applying this type of visualization, time can be mapped out explicitly, and the years will present a proper analysis of the Average Volume of Trades across sectors. This line graph best describes the fluctuations and patterns well, hence being very appropriate for assessing sector performance and trends over an extensive period.

To provide a clear overview of stock market activity, we created a **Line Chart** in Power BI, focusing on the **Average Volume of Trades** across various sectors over time. The X-axis represents the **Year (Date)**, while the Y-axis displays the **Average Volume**, and the **Sector** is used as the legend to differentiate market segments.



This Explanatory visualization is complemented with a slicer for sectors, enabling dynamic exploration of specific sectors' performance. The graph effectively captures the fluctuations in trading volumes, showcasing patterns such as:

- Sector Dominance: Certain sectors, like Financial Services or Technology, exhibit significant spikes in trading volumes during specific periods, reflecting heightened market activity.
- Historical Trends: Observing volume trends over time reveals notable events in the market, such as financial crises or technological booms, that influenced sector-specific trading activity.
- Comparative Analysis: The multi-sector comparison provides valuable insights into which sectors have consistently high or low trading volumes, aiding in identifying stable versus volatile market areas.

Design choice

The design of the main graph was crafted with clarity and interactivity as key priorities:

1. **Color Palette**: A distinct color was used for each sector, ensuring clear differentiation and easy interpretation. This consistent color-coding enhances visual storytelling and supports the quick identification of trends across sectors.

2. **Line Chart Selection**: A line chart was chosen for the main visualization to effectively

depict the temporal evolution of average trade volumes. Its ability to highlight fluctuations and trends over time makes it ideal for analyzing stock market dynamics.

3. **Slicer Integration**: Incorporating a slicer enables us to focus on specific sectors or compare subsets of data, enhancing the interactivity of the chart and allowing for deeper exploration.

4. **Scalable X-Axis**: The time axis is designed to scale seamlessly, allowing users to zoom in on specific time periods or view long-term trends across decades, providing both detailed and high-level perspectives.

5. **Scalable Y-Axis**: The Y-axis, representing average volumes, uses a logarithmic scale to accommodate the vast range of trading data, preventing smaller values from being overshadowed by large spikes.

6. **Minimalistic Layout**: The chart avoids unnecessary embellishments, adhering to a clean and professional design that prioritizes data readability and accessibility.

Animation Choice

While Power BI lacks advanced animation tools, interactive elements were strategically implemented to enhance engagement:

1. **Dynamic Filtering via Slicers**: Slicers allow for real-time updates of the chart when we select different sectors, creating an intuitive exploration experience akin to animation.

2. **Smooth Transitions**: Power BI's built-in smooth transitions between filtered states add a layer of animation, making data updates visually seamless and more engaging.

3. **Highlighting Trends**: Interactive hover effects provide instant insights into specific data points, such as sector-specific peaks or changes in volume.

4. **Drill-Down Features**: Although not a formal animation, drill-down capabilities allow users to interactively explore data at different levels (e.g., monthly or quarterly trends), creating a dynamic storytelling experience.

CONCLUSION

This project explored the dynamics of the stock market from 1962 to the present, with a focus on sector-wise trading volume trends and volatility. Through this analysis, we uncovered key insights into how different sectors respond to historical events and changing market conditions.

Data exploration and visualization were done in combination of Python libraries and Power BI. The Python libraries used were: NumPy and Pandas to explore and preprocess the data, YFinance to pull sector-specific data, and Concurrent.futures to accomplish all of that in

parallel much more quickly than is otherwise possible. Matplotlib for exploratory data analysis was used. In Power BI, an interactive line chart has been prepared, with added slicers for dynamic analysis, facilitated by drill-down functionality and hover and smooth transitions of elements for more user comfort.

Critical Analysis of the Visualization Outcome

The line chart of sector-wise average trading volumes effectively captures long-term trends and key historical patterns. By focusing on trading volumes, the visualization highlights periods of market turbulence, like the dot-com boom and the 2008 financial crisis, as well as the resilience of sectors like Utilities and Consumer Defensive. The interactivity, with features like slicers for sector filtering, adds an exploratory dimension, enabling us to focus on specific sectors or compare performance across time.

Technical Challenges

There were certain functionalities we were unable to achieve due to technical limitations:

- Power BI's limited animation tools constrained the storytelling aspect. A dynamic animation feature to narrate the historical progression of sector activity would have enhanced the storytelling.
- Integrating more advanced chart types, such as heat maps or area charts, was considered but required additional customization tools not readily available in Power BI.
- The complex and huge nature of the dataset made it difficult to visualise every aspect as the computer would run out of memory even when we assigned 60 gbs to Virtual Memory.
- Data retrieval through *yfinance* was time-intensive, even with threading it took more than 60 minutes to fetch sector data. A more robust API or prebuilt dataset with detailed sector information could have streamlined this step.

Collaboration and Effort Sharing

This project was a collaborative effort between Abhishek Malaviya and Kishore Magendiran:

- **Kishore** led the effort in collecting and integrating the dataset, sourcing the raw data from Kaggle, and retrieving sector information through *yfinance*.
- **Abhishek** focused on data cleaning, preprocessing, and handling missing data, ensuring the dataset was ready for analysis. Kishore provided valuable assistance during these steps.
- For the visualization, we contributed equally, brainstorming and building the final line chart and slicer features in Power BI. Finally, the report was co-authored with both members reviewing and refining the content to ensure clarity and accuracy.

Final Reflection

This project demonstrated the power of big data principles in understanding complex market behaviors and trends. While the outcomes aligned with our objectives, addressing the aforementioned improvements and challenges could further enhance the depth and interactivity of the analysis. Ultimately, the collaboration enabled a seamless workflow, leveraging each team member's strengths to produce a robust and engaging visualization of stock market trends.

REFERENCE:

The dataset is Taken from the following websites.

1. Kaggle, "9000 Tickers of Stock Market Data - Full History," https://www.kaggle.com/datasets/jakewright/9000-tickers-of-stock-market-data-full-history/data.
2. YFinance, Used to map each Unique Ticker with the Sector.