EAS 509

Group 18: Project Progress Report

Members: Kumar Utkarsh(kumarutk), Abhinav Nehra(anehra), Gaurav Kumar Jha(gjha)

Title: Sector-wise Stock Price Segmentation and Forecasting to uncover Investment Opportunities
Importance:
Understanding and predicting stock market trends within specific industry sectors is crucial for investors and analysts to identify potential investment opportunities, optimize portfolio strategies, and mitigate risks eventually. With the help of this project we aim to leverage different data mining, statistical analysis and machine learning techniques learned in class to analyse historical stock price data, segment stocks within industry sectors using possible clustering algorithms like K-means, analyse feature importance using random forest/boosting and forecast future price trends using Time Series models . By uncovering market segmentation, trend identification, and investment opportunities within each sector, this project will provide valuable insights and actionable recommendations for stakeholders in the financial markets.

Note: As per the feedback received during the Project Proposal, we completely agree that feeding raw data might not be the most effective approach for achieving optimal clustering results. To address this concern, we plan to transform and select relevant features that capture the underlying patterns and dynamics in the stock price data before feeding them to the clustering algo.

What We Want to Do:

- Exploratory Data Analysis
- Descriptive Statistics
- Data Preprocessing

- Clustering
- Time Series Analysis(ARIMA)
- Analysis and Interpretation

What we did so far:

- **Data Preprocessing**: Converted Date column, handled missing values, checked and handled outliers and scaled features.
- **Exploratory Data Analysis:** Visualized stock prices over time and checked summary statistics.
- **Clustering:** Applied K-means clustering to segment stocks within industry sectors and identify the industry sectors which have the highest dominating importance in each cluster
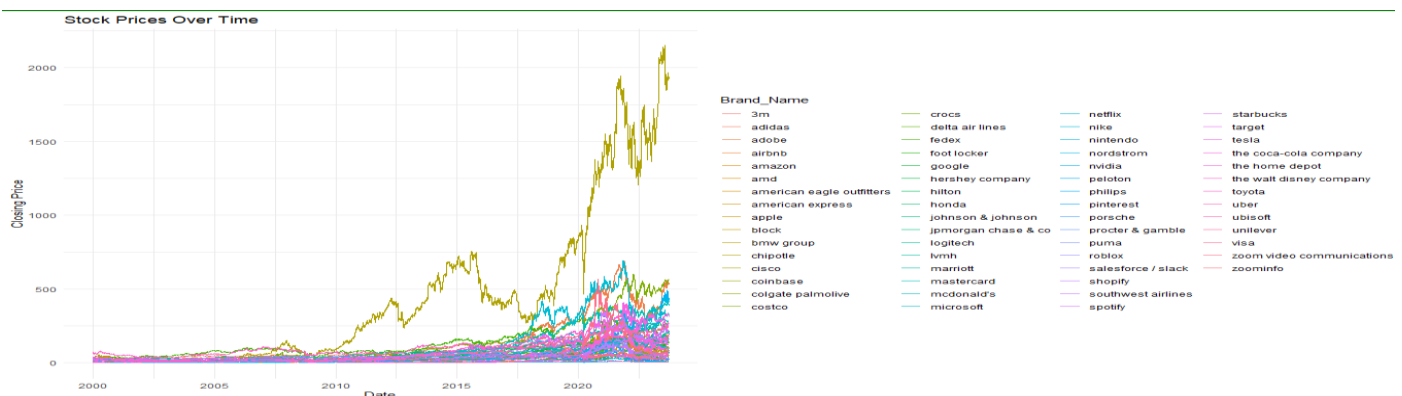
Link to Data Source:
World Stock Prices (Daily Updating) Dataset