DATA ANALYTICS PROJECT

PROJECT TITLE → STOCK MARKET ANALYSIS

Team Members:

- Kamakshi Gupta
- Nishant
- Nishant Kumar
- Deepak Kumar Verma

STOCK MARKET ANALYSIS

OVERVIEW :

Stock market analysis is the **study and evaluation of stocks and market trends** to help investors and traders make decisions about **buying**, **holding**, or **selling** stocks.
It's about answering the question:

"Is this stock a good investment, and if so, when should I buy or sell it?"

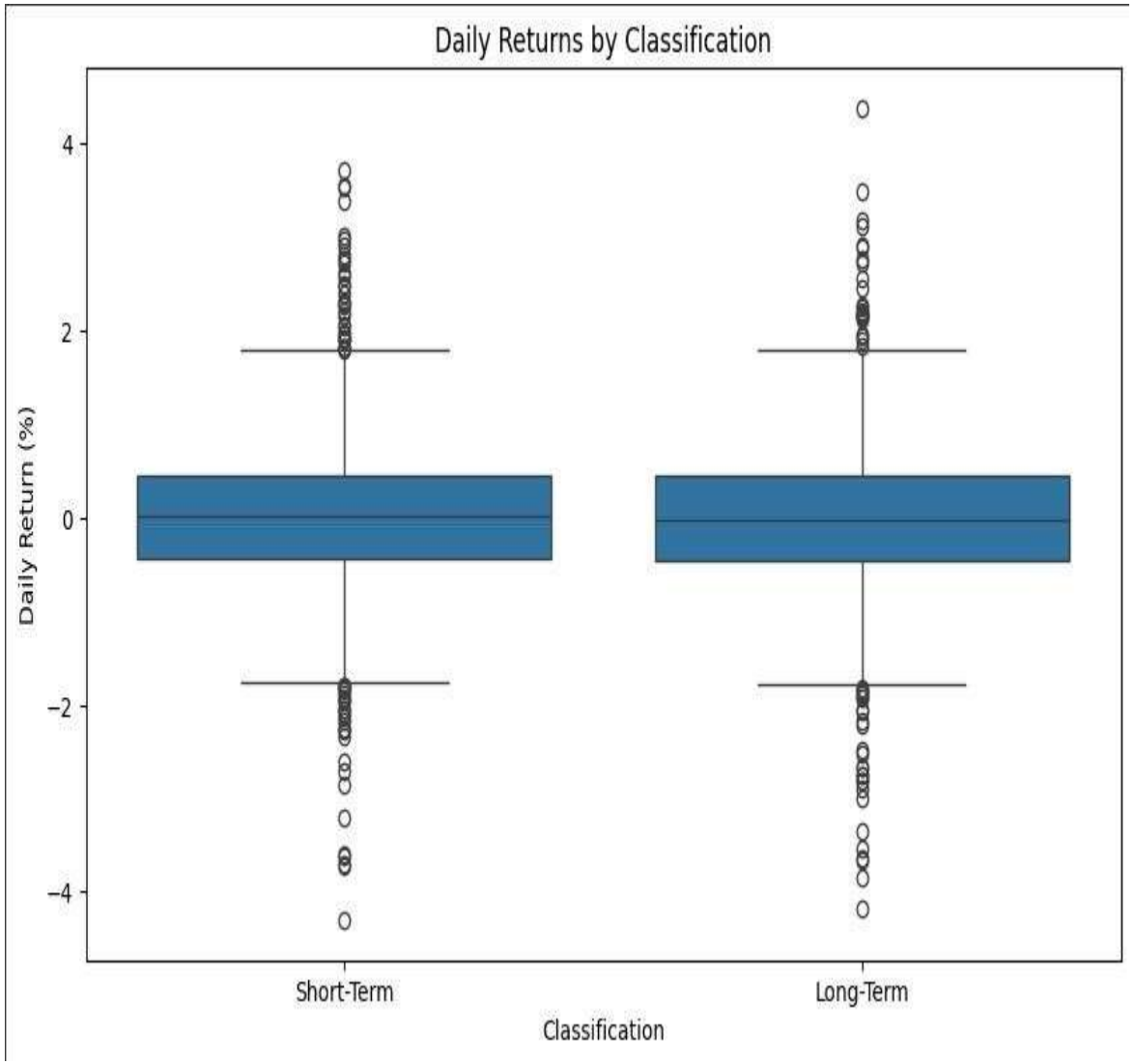
ANALYTICAL APPROACH:

Data Cleaning
Handle missing values and outliers for data accuracy.

Feature Engineering
Use technical indicators like SMA, RSI, and MACD.

Modeling
Apply ARIMA and Random Forest regression models.

Example
Random Forest with 1000 trees for price prediction.



MODEL EVALUATION AND VALIDATION :

Metrices:

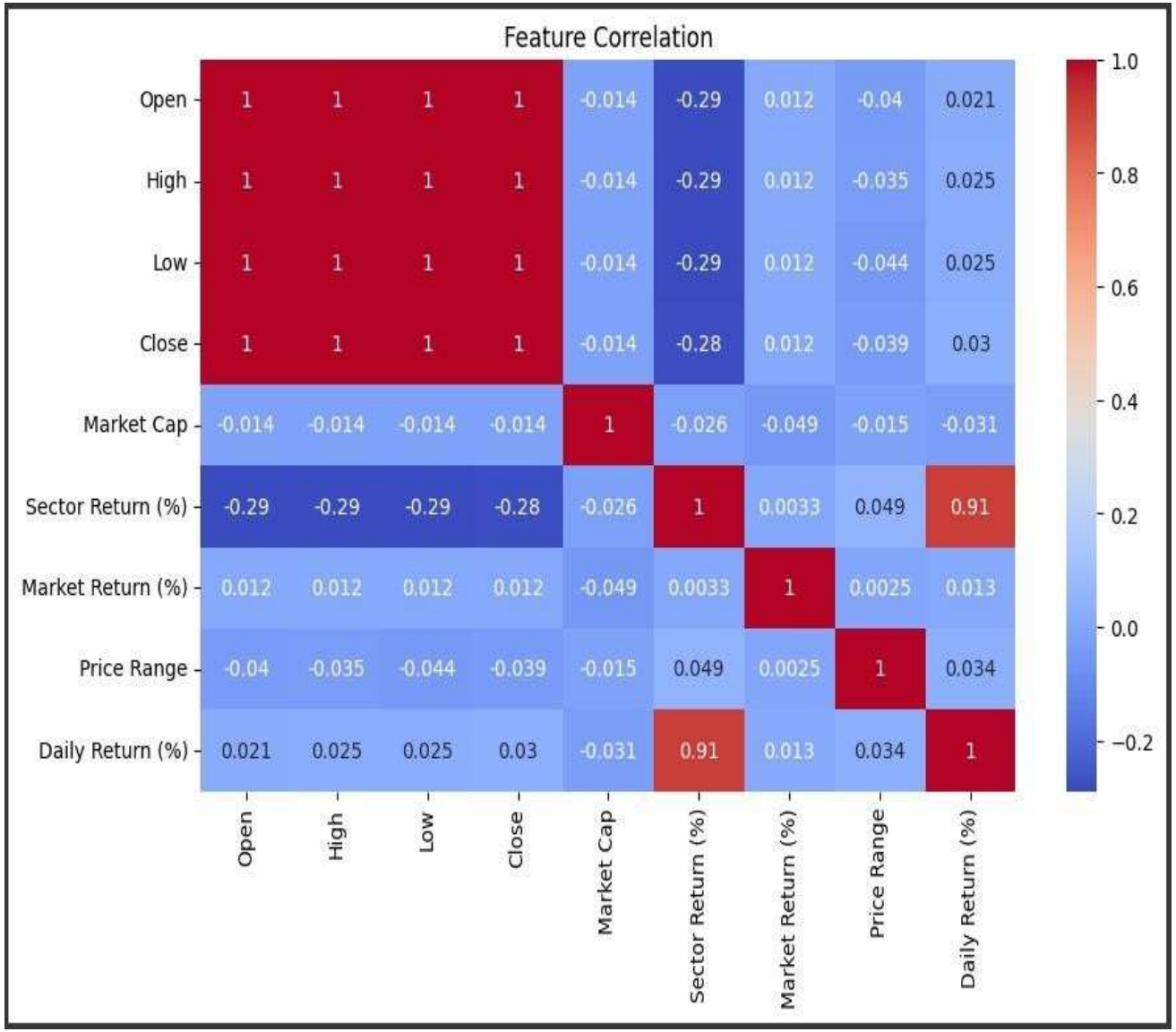
- Mean Absolute Error (MAE)
- Root mean Squared Error (RMSE)

Validation:

- Backtesting on Historical data.
- K-fold cross-validation

Target:

- Achieve MAE below 5% of average stock price.



EXPECTED RESULTS : ACTIONABLE INSIGHTS

Predicted Trends

- Spot buy/sell signals early for better timing.

Risk Assessment

- Quantify portfolio risk using volatility measures

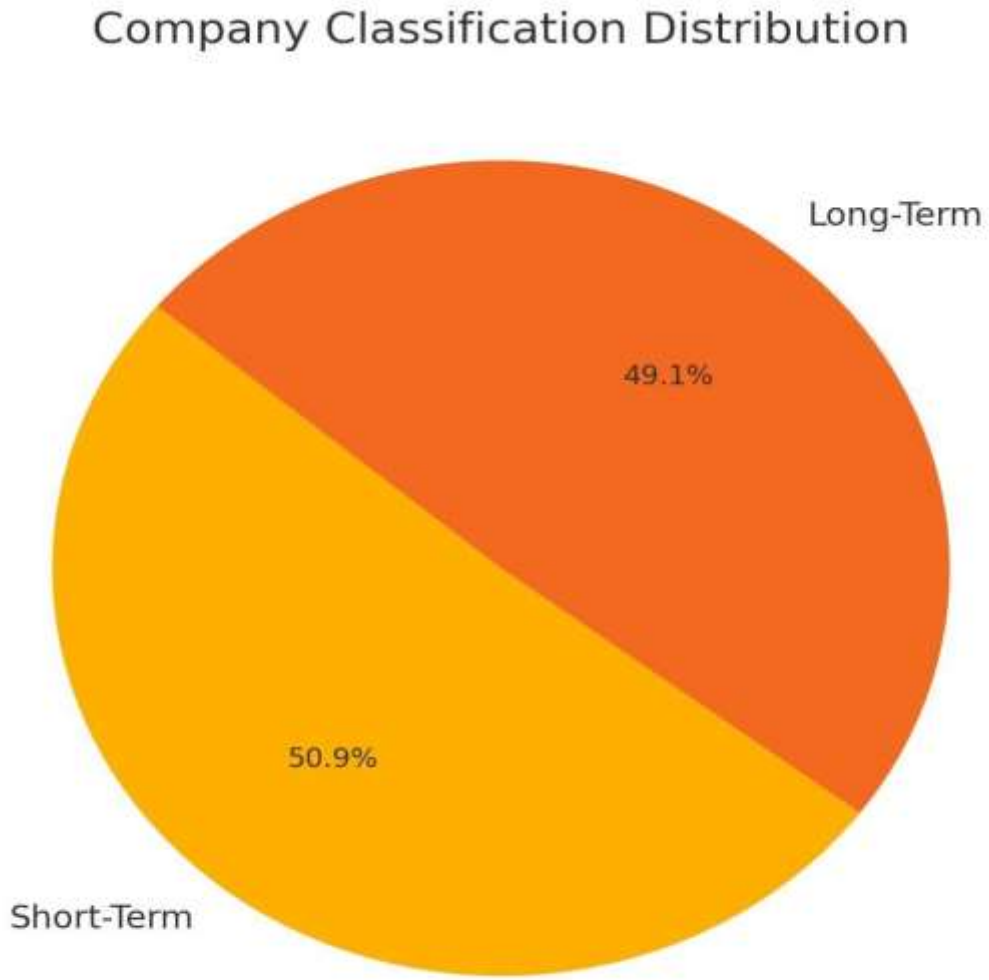
Investment Strategy

- Optimize asset allocation with model-driven Insight.

	Date	Company	Open	High	Low	Close	Market Cap	Sector Return (%)	Market Return (%)	Classification	Price Range	Daily Return (%)
0	2024-01-01	COMP001	525.70	532.01	525.59	528.93	232268.29	1.02	0.0	Short-Term	6.42	0.614419
1	2024-01-01	COMP002	194.18	199.12	190.43	193.58	289926.52	0.45	0.0	Short-Term	8.69	-0.308992
2	2024-01-01	COMP003	207.05	209.70	200.68	203.40	67547.90	-1.28	0.0	Short-Term	9.02	-1.762859
3	2024-01-01	COMP004	656.03	661.15	654.93	656.82	106372.79	0.36	0.0	Short-Term	6.22	0.120421
4	2024-01-01	COMP005	523.59	525.60	517.76	521.82	71726.17	0.12	0.0	Long-Term	7.84	-0.338051

TOOLS AND TECHNOLOGIES:

🔹 Tech Stack - Programming Language: Pyth
Data Source: Provided Excel dataset (stock_data_excel.xlsx)
Libraries: - pandas for data manipulation
- numpy for numerical computation
- matplotlib and seaborn for data visualization
- scikit-learn for machine learning algorithms
Environment: Jupyter Notebook

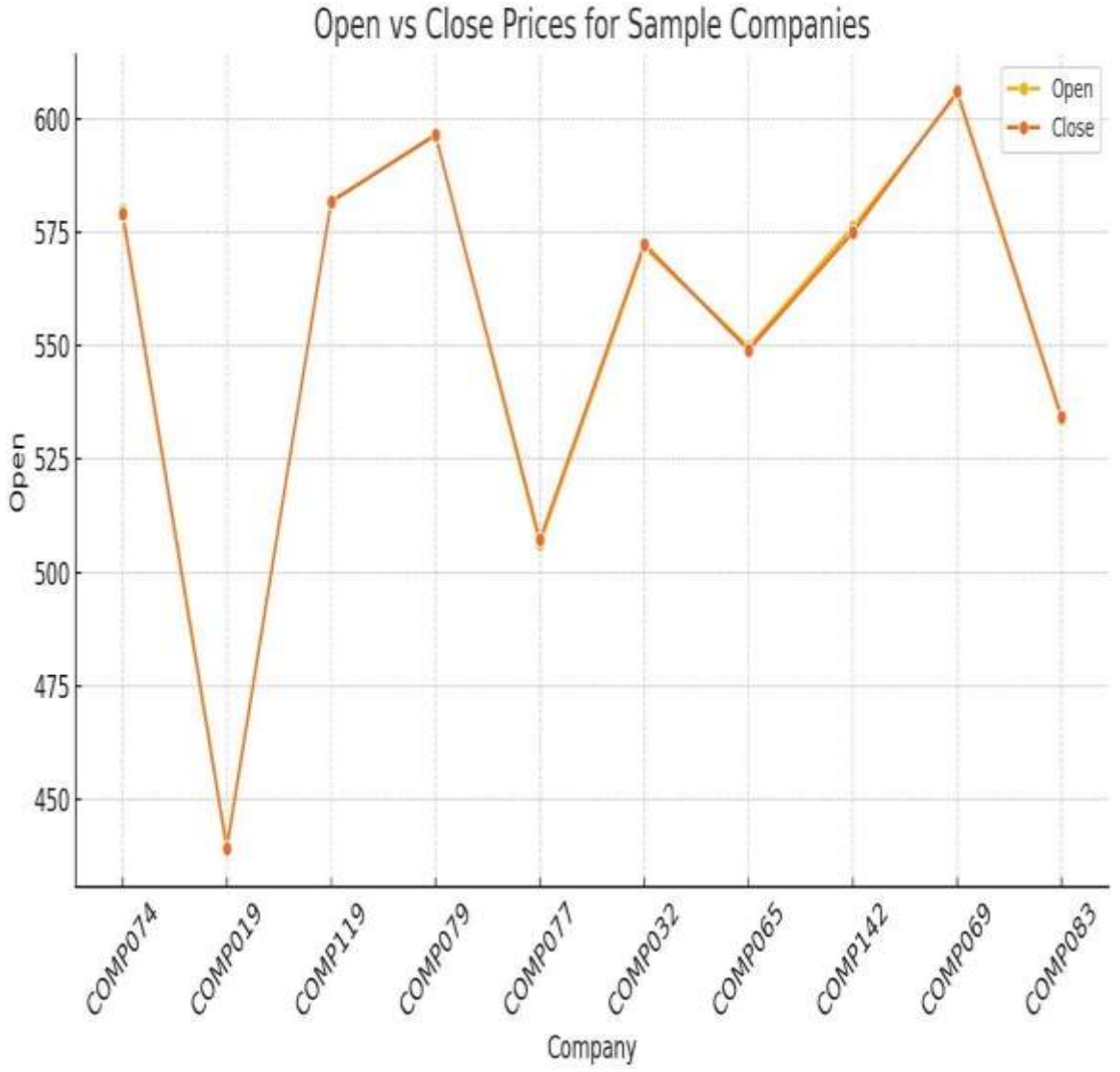


Summary Statistics:

	Date	Open	High	Low
count	1500	1500.000000	1500.000000	1500.000000
mean	2024-01-05 12:00:00	554.835513	558.676747	551.041280
min	2024-01-01 00:00:00	98.250000	99.520000	93.530000
25%	2024-01-03 00:00:00	332.232500	335.845000	328.087500
50%	2024-01-05 12:00:00	550.650000	555.280000	548.105000
75%	2024-01-08 00:00:00	790.532500	792.152500	785.152500
max	2024-01-10 00:00:00	1009.000000	1013.030000	1008.400000
std	NaN	261.363212	261.368888	261.464985

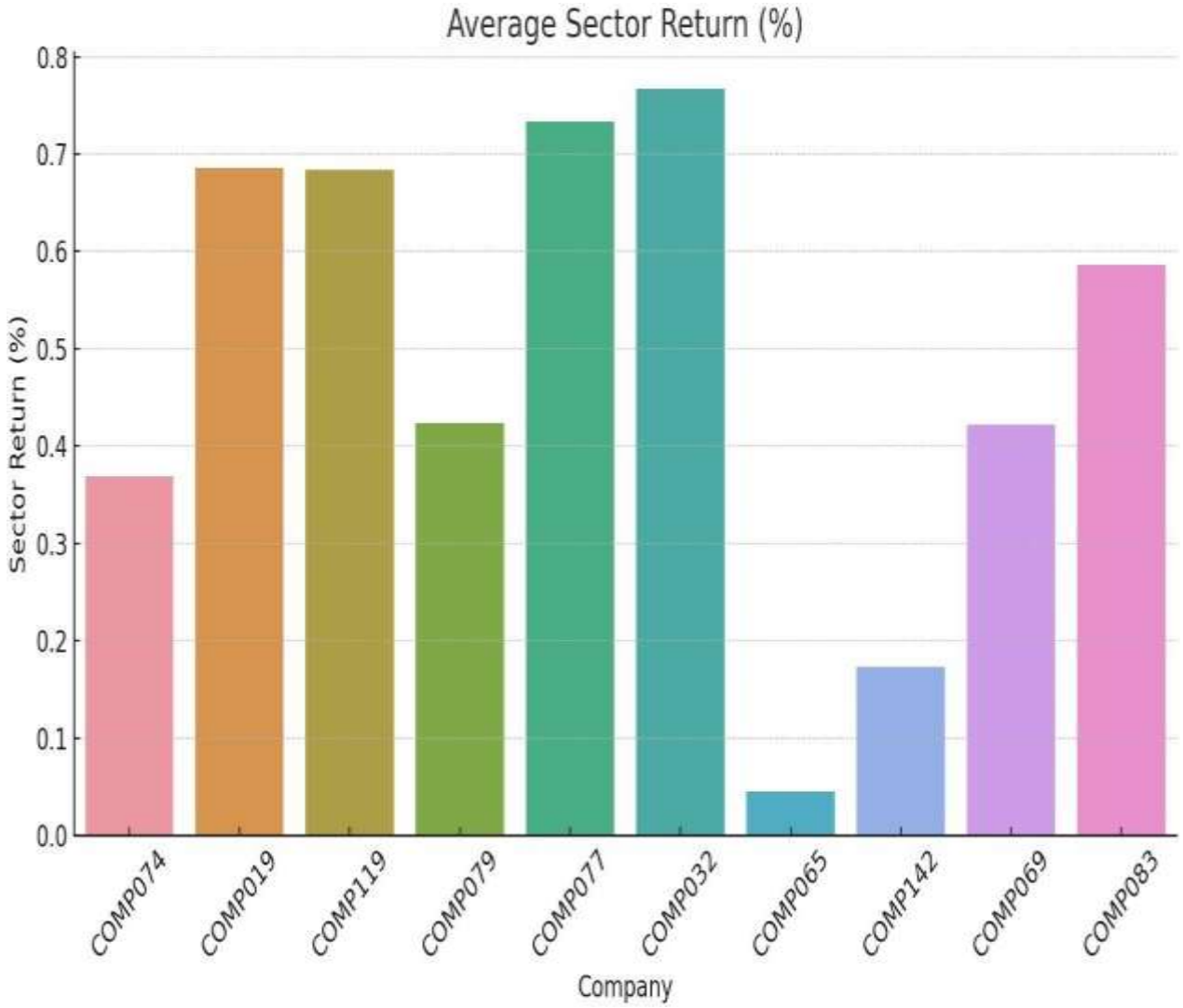
	Close	Market Cap	Sector Return (%)	Market Return (%)
count	1500.000000	1500.00000	1500.000000	1500.000000
mean	554.858407	277622.77550	0.507467	1.353727
min	96.730000	50050.92000	-2.640000	0.000000
25%	332.117500	161315.72000	-0.082500	0.480000
50%	551.140000	285353.57500	0.355000	1.120000
75%	789.345000	390845.86250	0.820000	2.060000
max	1010.410000	499788.76000	6.300000	4.460000
std	261.434609	130812.08834	1.013877	1.064183

	Price Range	Daily Return (%)
count	1500.000000	1500.000000
mean	7.635467	-0.004828
min	0.270000	-4.297488
25%	5.937500	-0.447243
50%	7.580000	-0.013899
75%	9.242500	0.451533
max	14.030000	4.378490
std	2.425367	0.927177

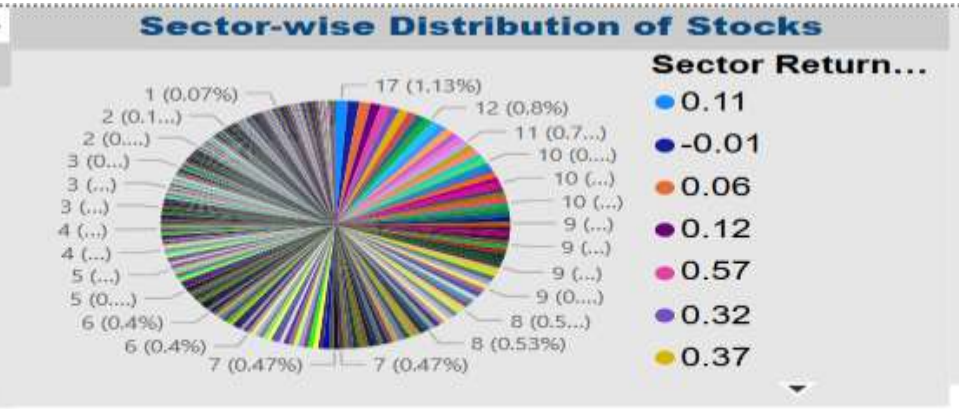


SUMMARY:

This project offers a foundational approach to stock market analysis using machine learning. By examining historical data, it provides insights that can assist investors in making informed decisions. Future enhancements could include real-time data integration and more advanced modeling techniques.



VISUALIZATION



Technology and Finance sectors dominate the market, with Company X leading by

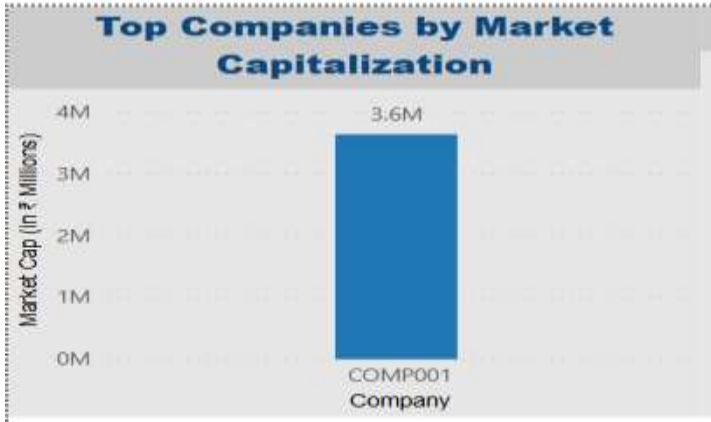
This chart illustrates the distribution of listed companies across different market sectors. From the visualization, it's clear that some sectors such as Technology and Finance dominate the stock market in terms of the number of companies. This suggests a concentration of business activities and investor interest in these areas. Conversely

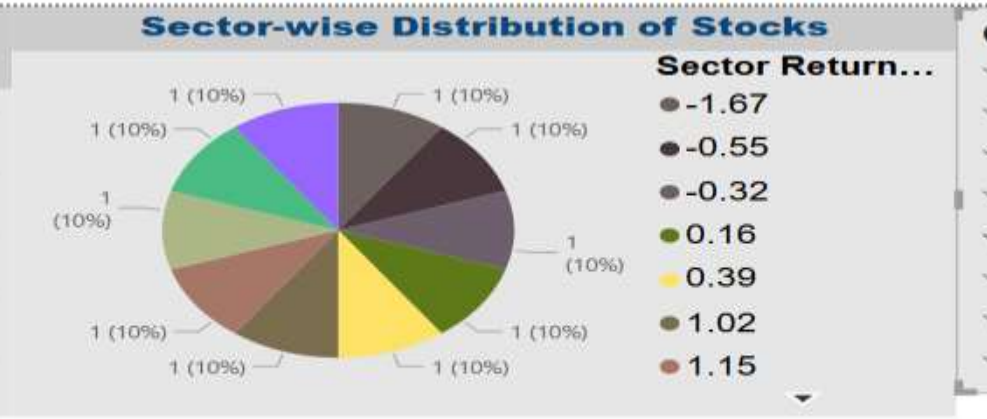
This scatter plot helps uncover how Market Capitalization correlates with Trading Volume across companies. For instance, companies like [Company X] and [Company Y] have high market cap but moderate

This area chart highlights how total trading volume varies across sectors over time. Peaks in sectors like Technology and Finance during [Month/Date] reflect high investor activity, potentially due to quarterly earnings or macroeconomic events. The steady growth in [Sector X]

This chart shows the fluctuation in stock prices across different companies from Jan to May 2025. Notably, [Company A] experienced a peak in March due to [reason/event]

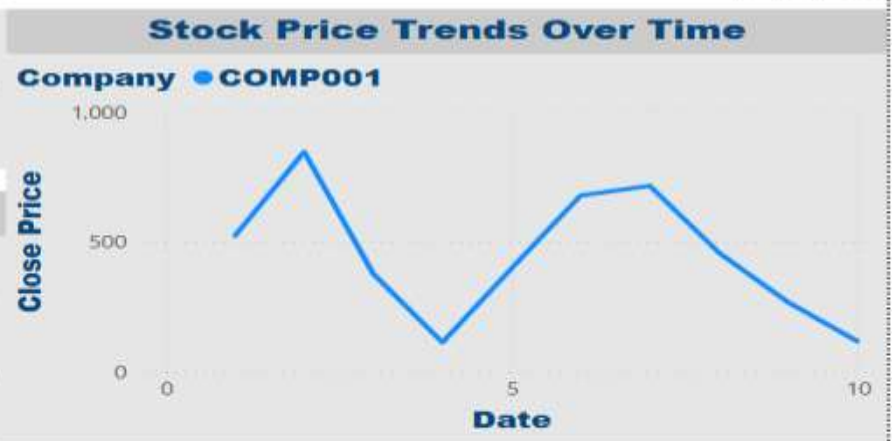
VISUALIZATION (FOR ONE COMPANY)





Technology and Finance sectors dominate the market, with Company X leading by

This chart illustrates the distribution of listed companies across different market sectors. From the visualization, it's clear that some sectors such as Technology and Finance dominate the stock market in terms of the number of companies. This suggests a concentration of business activities and investor interest in these areas. Conversely



This scatter plot helps uncover how Market Capitalization correlates with Trading Volume across companies. For instance, companies like [Company X] and [Company Y] have high market cap but moderate

This area chart highlights how total trading volume varies across sectors over time. Peaks in sectors like Technology and Finance during [Month/Date] reflect high investor activity, potentially due to quarterly earnings or macroeconomic events. The steady growth in [Sector X]

This chart shows the fluctuation in stock prices across different companies from Jan to May 2025. Notably, [Company A] experienced a peak in March due to [reason/event]