S&P 500 Portfolio Analysis – Project Report

# 1. Introduction

This project is a comprehensive end-to-end data analysis of S&P 500 stock market data spanning a period of five years (2013–2017). It utilizes Python, Jupyter Notebooks, Pandas, NumPy, Matplotlib, Seaborn, PostgreSQL, and Excel. The primary objective is to build a well-structured portfolio analysis tool with data cleaning, exploratory analysis, risk metrics, forecasting models, SQL-driven insights, and an interactive Excel dashboard.

# 2. Tools and Technologies Used

**Languages & Libraries:** Python, Pandas, NumPy, Matplotlib, Seaborn  
**Environment:** Jupyter Notebook  
**Database:** PostgreSQL (SQL)  
**Reporting & Dashboarding:** Excel

# 3. Project Folder Structure

S&P500-Portfolio-Analysis/  
├── data/  
│ ├── raw/ # Kaggle CSVs  
│ └── processed/ # Cleaned daily returns, metrics  
├── notebooks/  
│ ├── 01\_data\_cleaning.ipynb  
│ ├── 02\_eda.ipynb  
│ ├── 03\_risk\_metrics.ipynb  
│ ├── 04\_sql\_queries.ipynb  
│ ├── 05\_forecasting.ipynb  
│ └── 06\_dashboard\_summary.ipynb  
├── sql/  
│ ├── schema.sql  
│ ├── load\_data.sql  
│ └── queries.sql  
├── scripts/  
│ ├── utils.py  
│ └── load\_to\_postgres.py  
├── reports/  
│ ├── visuals/  
│ ├── portfolio\_insights.pdf  
│ └── dashboard.xlsx  
├── requirements.txt  
├── README.md  
└── .gitignore

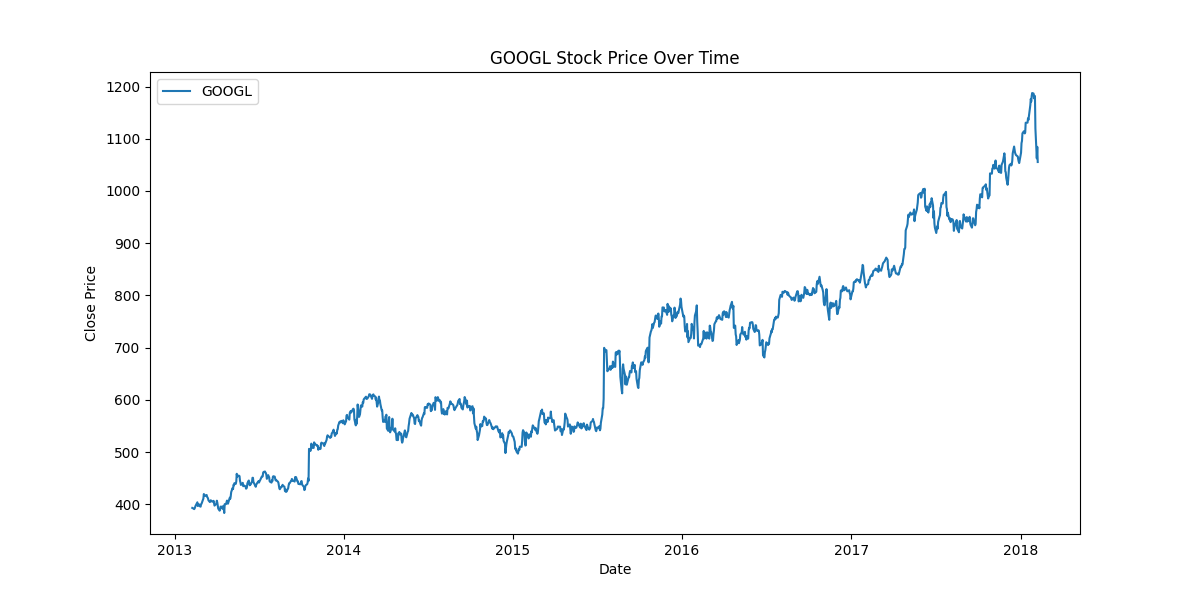
# 4. Project Modules & Description

## Notebook 01: Data Cleaning

Loaded and cleaned the raw stock data (all\_stocks\_5yr.csv), removed nulls, parsed dates, and calculated daily returns.

## Notebook 02: Exploratory Data Analysis

Explored stock distribution, sectoral breakdown, average returns, price trends, and volatility across selected tickers.



## Notebook 03: Risk Metrics

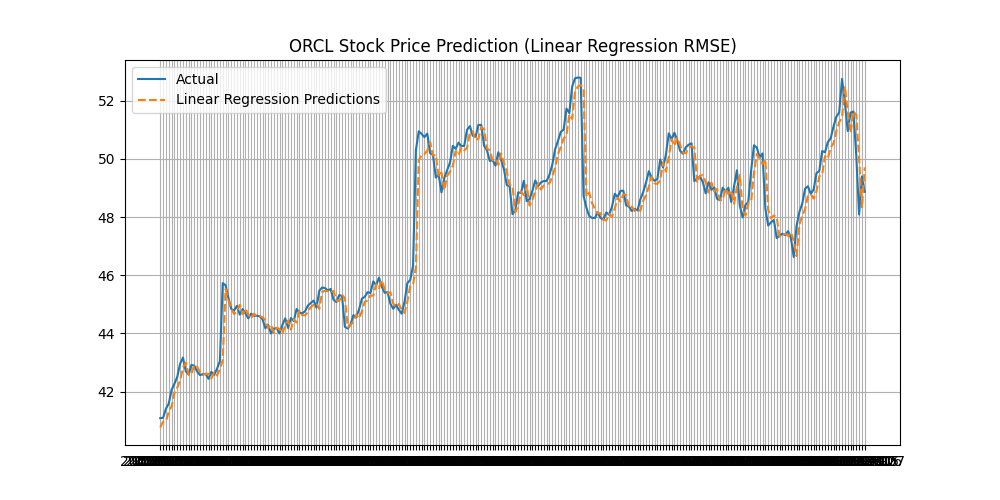
Calculated Sharpe Ratio, Volatility, Beta, and Maximum Drawdown for selected stocks and compared with market.

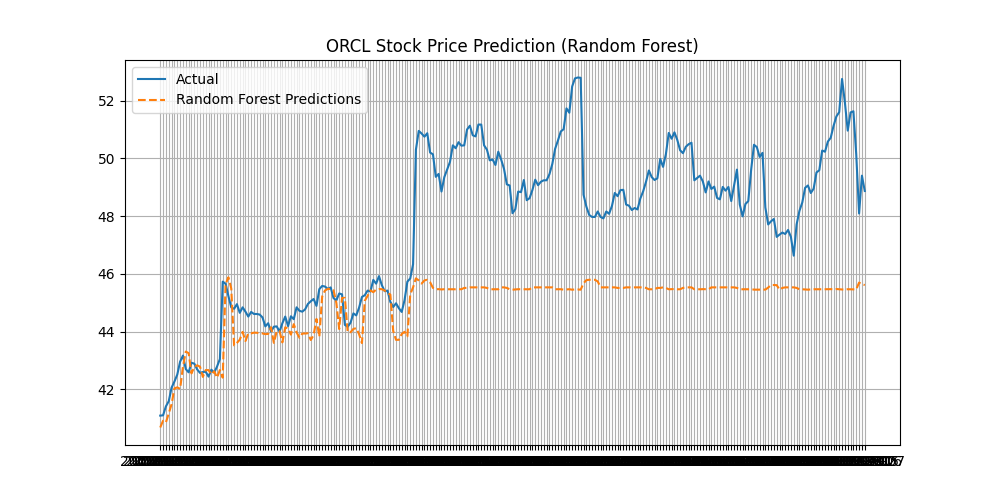
## Notebook 04: SQL Queries

Created SQL schema, loaded cleaned data into PostgreSQL, and executed queries to extract business insights.

## Notebook 05: Forecasting

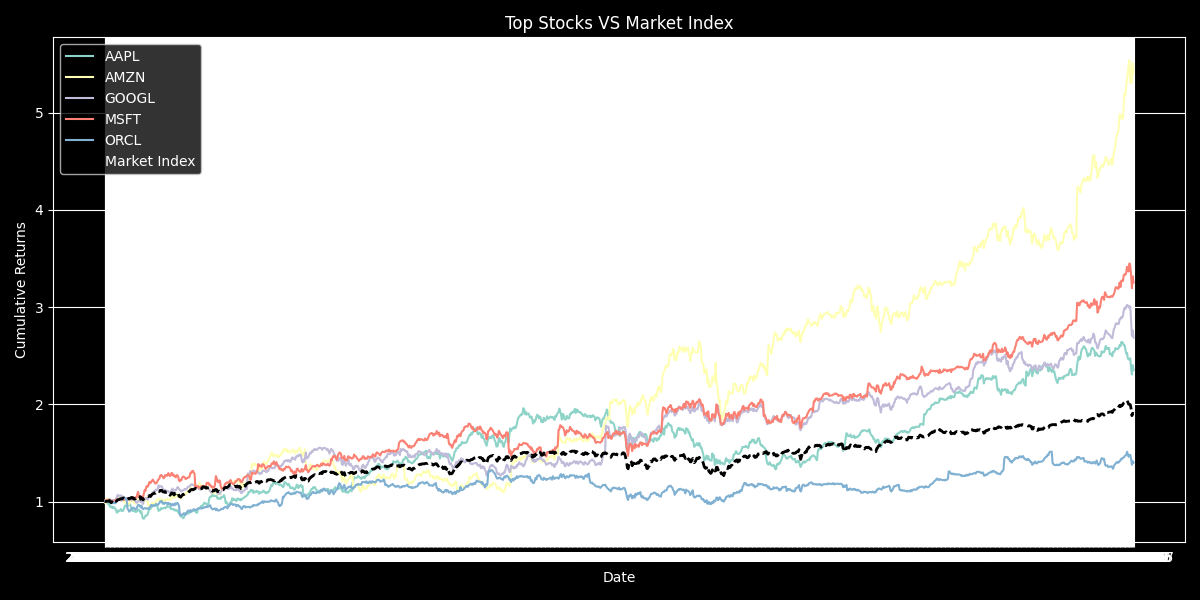
Used ARIMA modeling to predict stock prices (especially ORCL) using historical trends and validated prediction curves.





## Notebook 06: Dashboard Summary

Created summary visuals and KPIs, developed comparison charts between selected stocks and the broader market index.



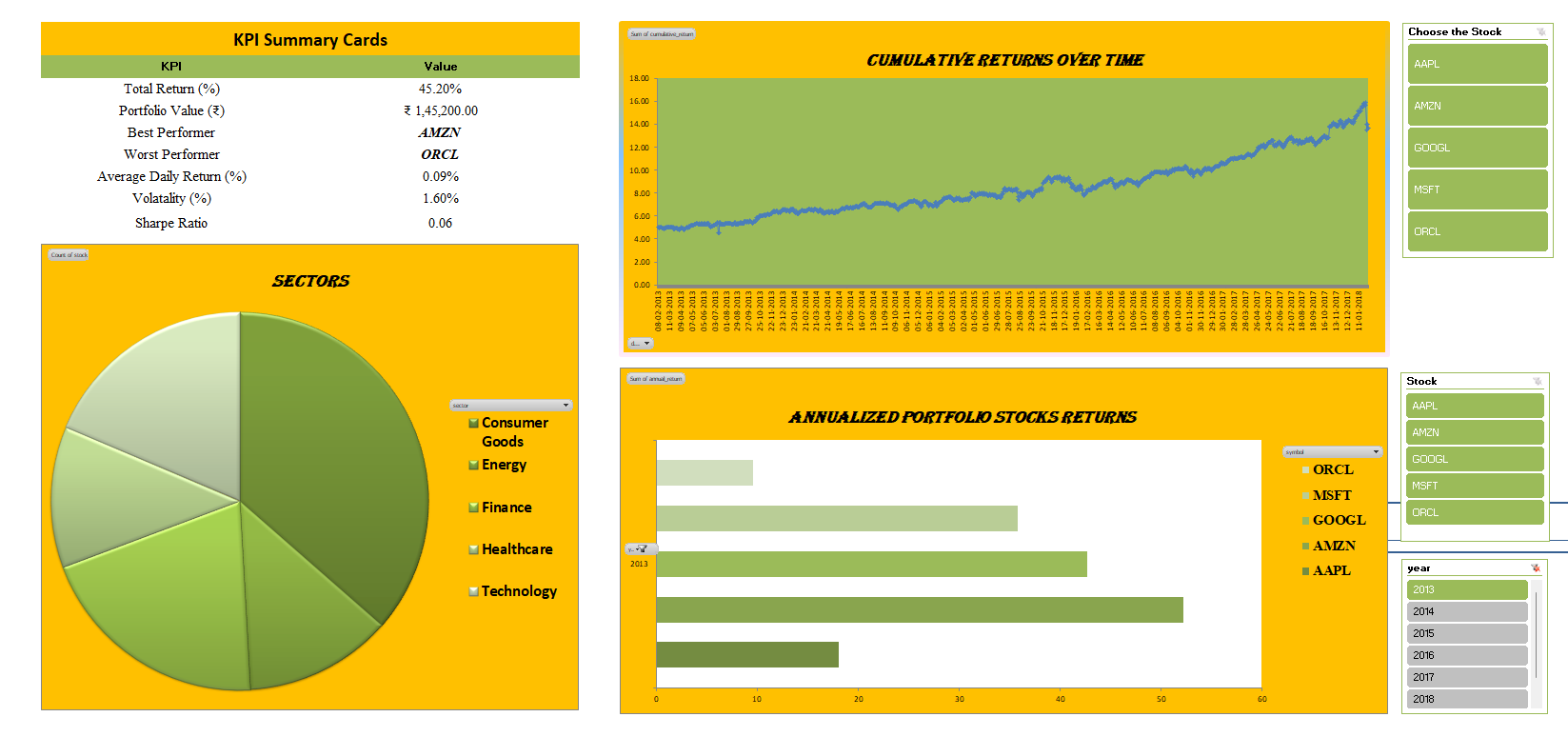
## Additional Notebook: ORCL Predictive Modeling

Focused on Oracle's stock and built a predictive model specifically to forecast its future price based on past behavior.

# 5. Excel Dashboard Highlights

The final deliverable includes an Excel file with the following sheets:

1. Portfolio Overview – Summary of selected portfolio, returns, and sectors.  
   2. KPI Dashboard – Sharpe ratio, volatility, beta, and more.  
   3. Visualizations – Stock trends over time.  
   4. Predictive Modeling – Forecasted values for ORCL.  
   5. SQL Insights – Extracted insights from SQL queries.  
   6. Insights & Recommendations – Actionable points for investors.



# 6. Conclusion & Learnings

This project reflects a deep understanding of financial data analysis, from preprocessing to actionable insights. It demonstrates skill in working with large datasets, applying statistical techniques, visualizing data trends, and generating insights using both Python and SQL. The Excel dashboard complements the Jupyter analysis, bridging technical output with business-friendly presentation.