QUANTINSIGHT : A Peek into the Role of a Quantitative Analyst

Harnessing Data Analytics for Stock Predictions and Portfolio Optimization

- YASHITA SABHARWAL

Introduction

1) Objective:

- Simulate a Quantitative Analyst's responsibilities by building an end-to-end stock analysis framework.

2) Key Goals:

- Predict stock prices using machine learning.
- Integrate sentiment analysis for enhanced predictions.
- Optimize a stock portfolio and backtest strategies.

3) Relevance:

- Importance of stock analysis in financial decision-making.

Project Workflow

1. **Exploratory Data Analysis (EDA):** Analyze stock trends, detect patterns, and prepare data by visualizing technical indicators like Moving Averages and RSI.
2. **Predictive Modeling**: Use machine learning models (Linear Regression, Random Forest, LSTM) to predict stock prices or classify trends.
3. **Sentiment Analysis**: Integrate market sentiment from financial news or tweets using NLP tools (e.g., VADER) to enhance predictions.
4. **Interactive Dashboards**: Build interactive dashboards with **Plotly Dash** to present insights, predictions, and sentiment trends.
5. **Portfolio Optimization**: Maximize returns and minimize risk with **Mean-Variance Optimization** and visualize the Efficient Frontier.
6. **Backtesting Strategies**: Simulate trading strategies (e.g., RSI-based) and evaluate performance with metrics like returns, win rate, and drawdowns.

Step 1 - Exploratory Data Analysis (EDA)

Objective:

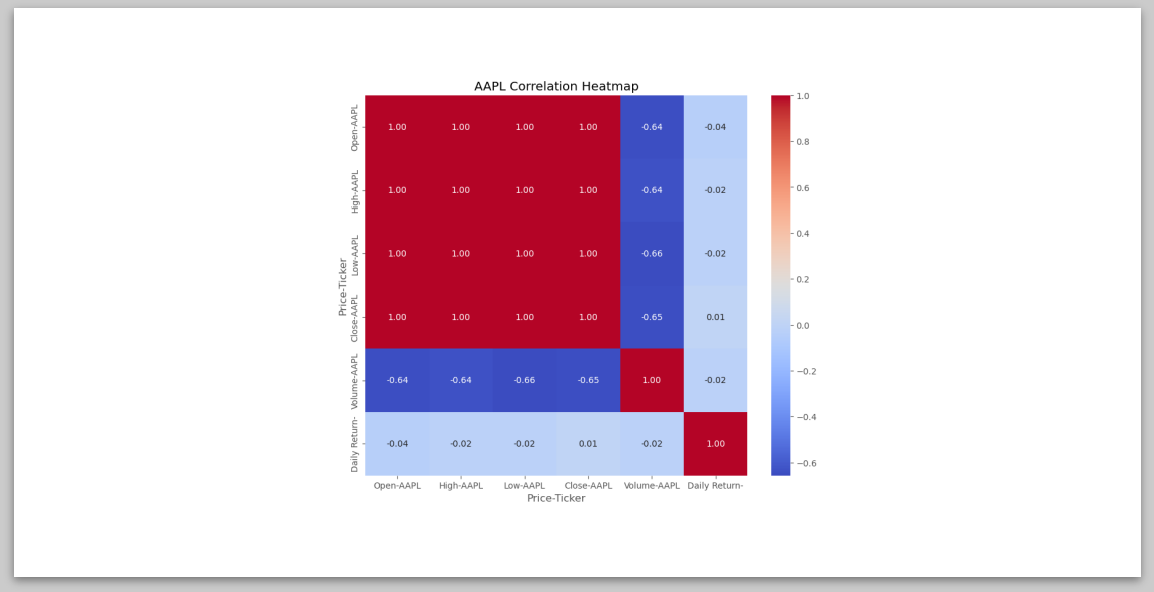
Gain insights into the historical performance, trends, and volatility of stock prices using Apple Inc. (AAPL) data from 2020 to 2024.

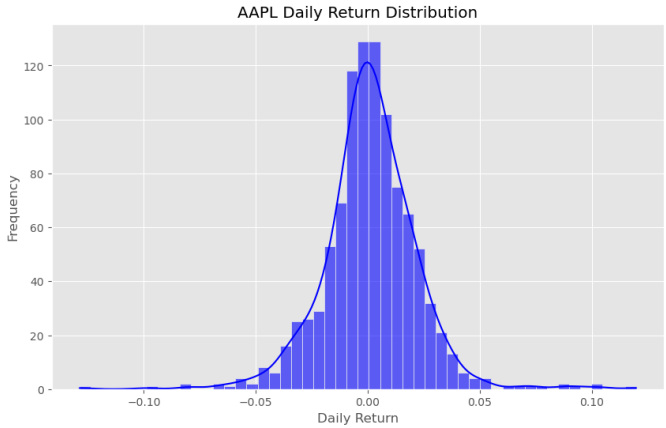
Data Overview

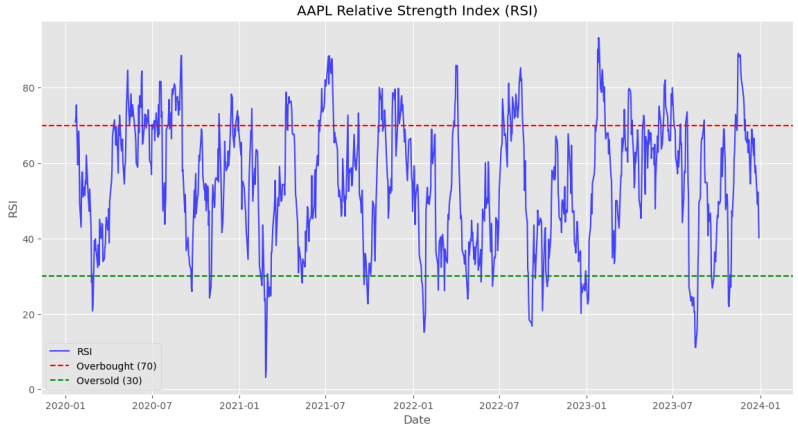
- **Dataset:**
  - Source: Yahoo Finance
  - Timeframe: January 1, 2020 - January 1, 2024
  - Columns: Date, Adj Close, Close, High, Low, Open, Volume
- **Basic Summary:**
  - Total Records: 1006
  - Missing Values: None
  - Key Metrics:
    - Mean Close Price: $140.81
    - Maximum Close Price: $198.11
    - Minimum Close Price: $56.09

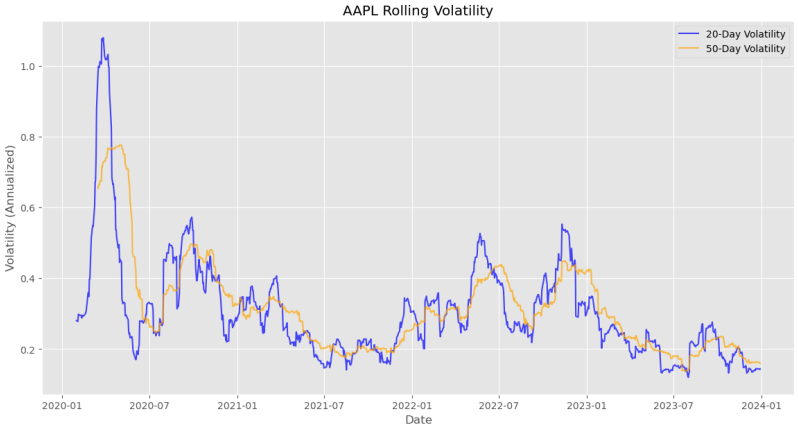
Key Technical Indicators

- **Daily Returns:**
  - Used to analyze volatility and daily price changes.
- **Simple Moving Averages (SMA):**
  - **50-day SMA**: Short-term trend.
  - **200-day SMA**: Long-term trend.
- **Bollinger Bands:**
  - Measures price volatility using rolling mean and standard deviation.
  - Identifies overbought/oversold conditions.
- **Relative Strength Index (RSI):**
  - Oscillator to detect momentum.
    - RSI > 70: Overbought
    - RSI < 30: Oversold
- **Rolling Beta:**
  - Measures stock's volatility relative to the market (S&P 500 index).


AAPL Correlation Heatmap
Price-Ticker
Open-AAPL
High-AAPL
Low-AAPL
Close-AAPL
Volume-AAPL
Daily Return-
1.00 1.00 1.00 1.00 -0.64 -0.04
1.00 1.00 1.00 1.00 -0.64 -0.02
1.00 1.00 1.00 1.00 -0.66 -0.02
1.00 1.00 1.00 1.00 -0.65 0.01
-0.64 -0.64 -0.66 -0.65 1.00 -0.02
-0.04 -0.02 -0.02 0.01 -0.02 1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6


AAPL Daily Return Distribution
Frequency
Daily Return
0
20
40
60
80
100
120
−0.10
−0.05
0.00
0.05
0.10


AAPL Relative Strength Index (RSI)
RSI
Date
RSI
Overbought (70)
Oversold (30)
0
20
40
60
80
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
2024-01


AAPL Rolling Volatility
Volatility (Annualized)
Date
20-Day Volatility
50-Day Volatility
0.2
0.4
0.6
0.8
1.0
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
2024-01

Insights from Visualizations

- **Closing Price with Moving Averages:**
  - Trends align with SMA signals for potential buy/sell decisions.
- **Bollinger Bands:**
  - Stock often returns to its mean price after hitting the upper/lower band.
- **RSI:**
  - Consistently identified overbought/oversold periods for AAPL.
- **Daily Return Distribution:**
  - Normal distribution with slight skew, indicating occasional extreme events.
- **Correlation Heatmap:**
  - Strong correlation between Open, High, and Close prices.
- **Volatility Trends:**
  - 20-day and 50-day rolling volatility highlight varying market dynamics.

Portfolio Analysis

Efficient Frontier:

Simulated 5000 portfolios to optimize returns and minimize risk.

Best Portfolio (Max Sharpe Ratio):

Annualized Return: 28.26%

Annualized Volatility: 31.49%

Weights:

AAPL: 69.20%

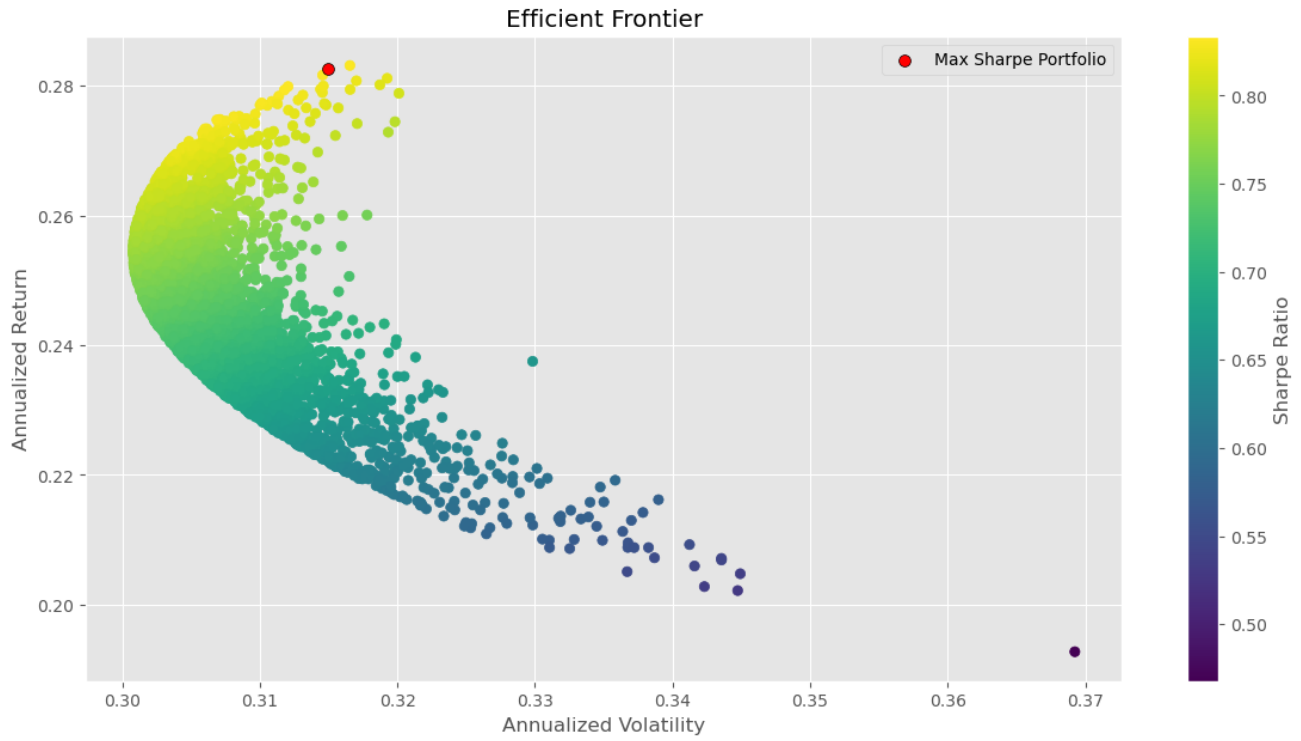
MSFT: 0.44%

GOOGL: 5.04%

AMZN: 25.32%

Sharpe Ratio:

AAPL's Sharpe Ratio: 0.81, indicating moderate reward-to-risk.



A) Rolling Beta

Beta > 1: High volatility relative to market.

Beta < 1: Less volatile than market.

AAPL Beta Analysis:
Fluctuated around 1, indicating near-market volatility.

B) Exported Data

Processed Data with Indicators:

File: AAPL_data_with_indicators.csv

Rolling Beta Data:

File: AAPL_data_with_beta.csv

Portfolio Simulation Results:

File: portfolio_simulation_results.csv

Conclusion

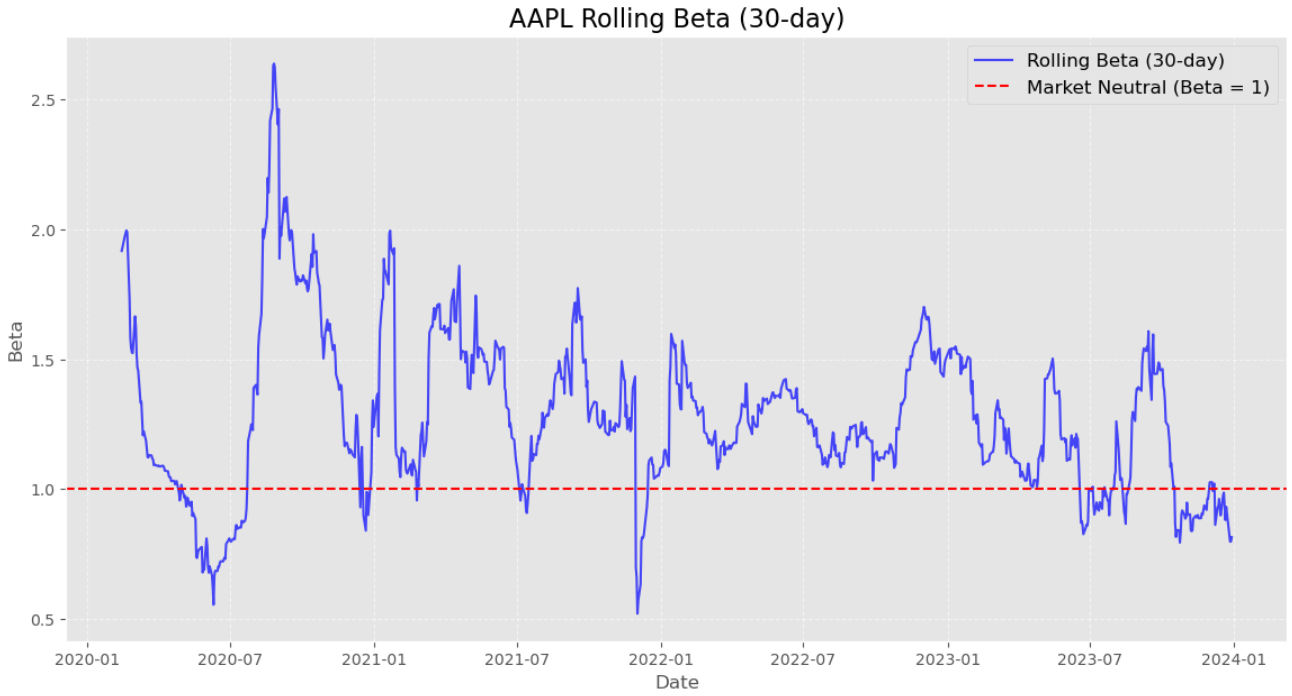
Trends Identified: Strong growth post-2020 with occasional dips.

Volatility Analysis: High volatility periods aligned with global events.

Future Improvements:

Incorporate sentiment analysis or macroeconomic indicators.

Explore other asset classes for portfolio diversification.



Step 2 - Predictive Modeling

Objective: To predict stock prices using machine learning models and evaluate their performance with respect to Apple Inc. (AAPL) stock data.

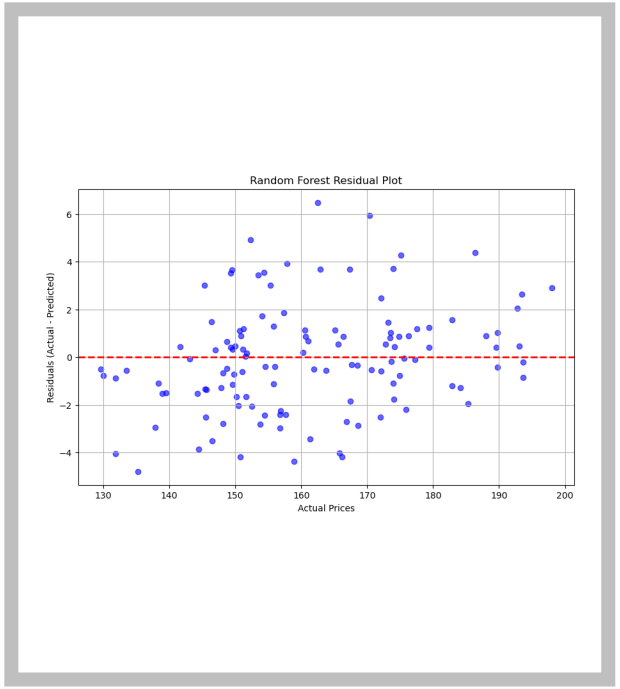
Data Preparation and Feature Engineering

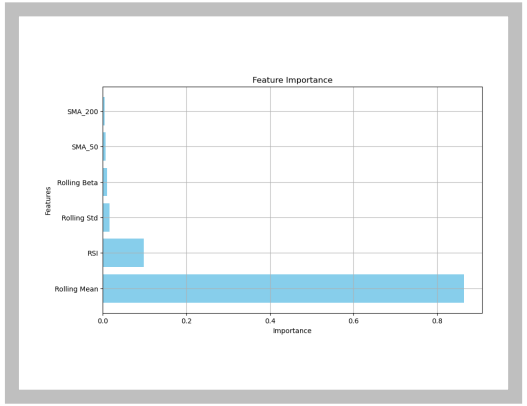
1.Dataset Overview:

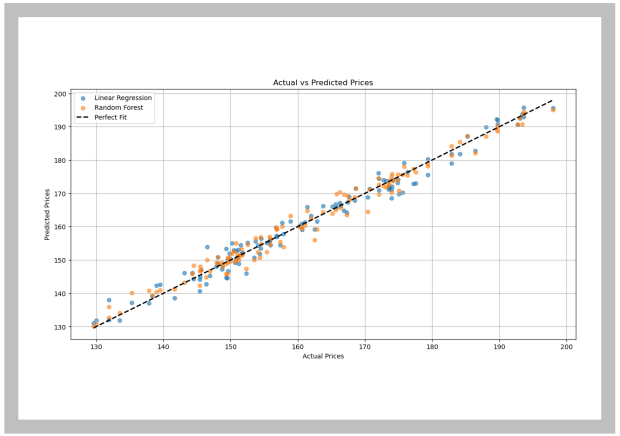
1. Data includes stock metrics like closing price, adjusted closing price, and key technical indicators.
2. Missing and invalid values were handled through numeric conversions and data imputation.

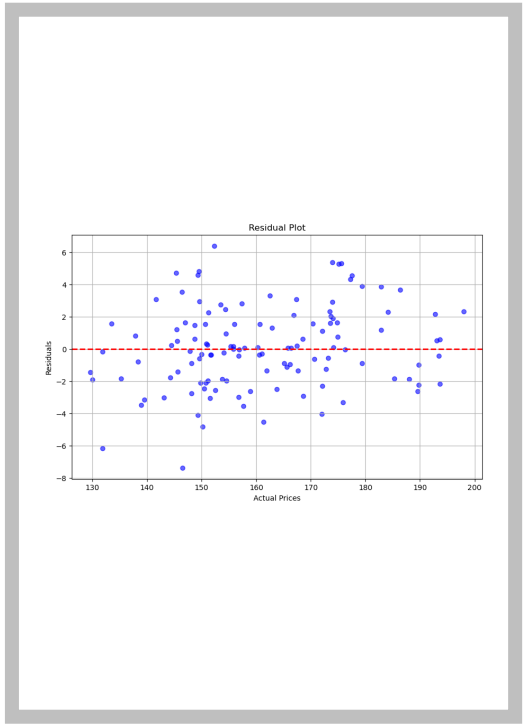
2.Key Technical Indicators:

1. **Simple Moving Averages (SMA):** 50-day and 200-day trends.
2. **Bollinger Bands:** Price volatility based on rolling mean and standard deviation.
3. **Relative Strength Index (RSI):** Measures momentum to identify overbought/oversold conditions.







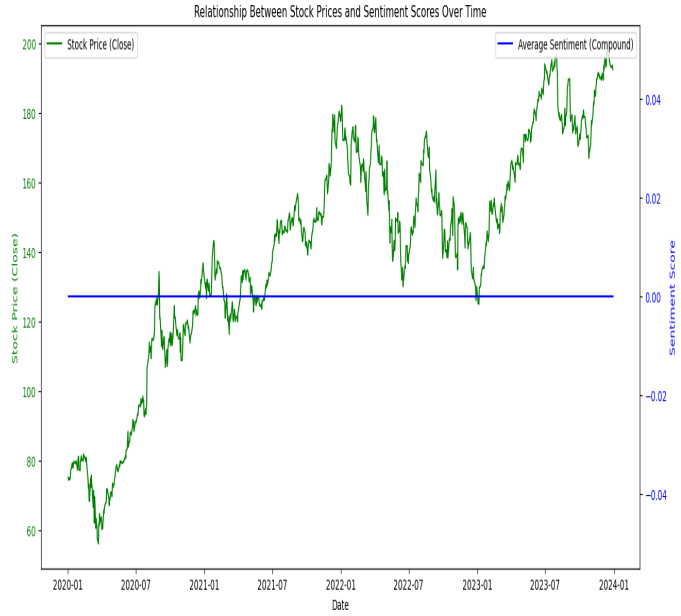


- **Model Training and Evaluation**
- **Linear Regression:**
  - **Performance:**
    - Mean Absolute Error (MAE): 1.99
    - R^2 Score: 0.97
  - **Residual Analysis:** Scatterplot showed a linear fit but some clustering of errors.
- **Random Forest Regressor:**
  - **Performance:**
    - Mean Absolute Error (MAE): 1.74
    - R^2 Score: 0.98
  - **Feature Importance:** Rolling Mean had the highest impact on price prediction.
- **Visual Analysis**
- **Residual Plots:**
  - Evaluated model accuracy by comparing residuals (errors).
  - Random Forest showed better error distribution than Linear Regression.
- **Feature Importance Chart:**
  - Highlighted the most influential indicators on stock price predictions.
- **Actual vs Predicted Prices:**
  - Scatterplot comparison showed Random Forest's predictions were closer to actual prices.

- **Future Improvements**
- **Advanced Models:**
  - Explore Long Short-Term Memory (LSTM) networks for time-series predictions.
- **Feature Expansion:**
  - Include sentiment analysis, news data, or macroeconomic indicators.
- **Evaluation Metrics:**
  - Incorporate metrics like RMSE and MAPE for more comprehensive performance insights.
- **Hyperparameter Optimization:**
  - Improve model performance through techniques like grid search.
- **Conclusion**
- Both models achieved high accuracy, but Random Forest outperformed Linear Regression.
- Technical indicators proved valuable for stock price forecasting.
- Future enhancements can refine predictions and expand application across industries.

Step 3 - Sentiment Analysis for Stock Market Predictions

- **Goal**: Analyze the relationship between Apple's stock performance and sentiment derived from related news articles to assess market sentiment's impact on stock trends.
- **Key Steps in Sentiment Analysis:**
- **Data Collection**:
  - Stock price data and news articles related to Apple (AAPL) were gathered and stored in CSV files.
  - Columns included dates, adjusted closing prices, and calculated sentiment scores.
- **Data Preprocessing**:
  - Dates were converted to datetime format, and data was sorted chronologically.
  - Invalid and missing data were handled by dropping or coercing errors to ensure clean analysis.
- **Sentiment Scoring**:
  - Used the VADER sentiment analyzer to generate **compound sentiment scores** from news articles.
  - Compound scores were aggregated daily to compute average sentiment.
- **Data Integration**:
  - Merged stock price data with sentiment scores to create a unified dataset for analysis.
- **Visualization**:
  - Created a dual-axis plot to showcase stock prices (closing values) alongside sentiment trends (compound scores).
- **Insights:**
- **Stock Price Analysis**:
  - Stock prices showed a general upward trend with intermittent periods of volatility.
  - Reflects market behavior based on internal and external factors.
- **Sentiment Analysis**:
  - Sentiment scores were mostly stable, with limited observable correlation with stock prices.
  - Highlights the need for further refinement in sentiment extraction or incorporation of more granular data.
- **Challenges & Improvements:**
- **Challenges**:
  - Missing or flat sentiment scores hindered meaningful insights.
  - Noise in the data due to unrelated news articles.
- **Improvements**:
  - Refine news sources to ensure relevance to Apple's stock.
  - Explore additional sentiment analysis techniques or models for greater precision.
- **Conclusion:**
- This analysis provides a starting point for understanding sentiment's impact on stock prices. Further enhancements in sentiment scoring and additional datasets could unlock actionable insights for investors and analysts.



Step 4 - Interactive Dashboards

Objective:

- To create a dynamic and interactive platform for users to analyze stock market data and visualize financial insights.

Features (Current Implementation):

- **Dynamic Stock Data Retrieval**:
  - Users input a stock ticker symbol (e.g., AAPL).
  - Fetches historical stock data using the yfinance library.
  - Displays **dates** and **closing prices**.
- **Interactive Web Application**:
  - Built using **Flask**.
  - Frontend templates (HTML) are integrated for user interaction.
  - Basic form validation for valid ticker symbols.
- **Real-Time Visualization**:
  - Displays historical stock data trends with **interactive graphs** (future enhancement).

Planned Enhancements:

- **Data Visualization**: Add interactive graphs using **Plotly** or **Chart.js**.
- **Additional Features**:
  - Technical Indicators (e.g., Moving Averages, RSI, Bollinger Bands).
  - Portfolio Management: Show stock weights from optimization.
- **User Personalization**:
  - Allow users to compare multiple stocks.
  - Include sentiment analysis insights for selected stocks.
- **Deployment**: Host the application on platforms like **Heroku** or **AWS** for global accessibility.

Insights:

- The QuantInsight app demonstrates how user-friendly web tools can empower data-driven financial decisions. By combining live data retrieval, analysis, and visualization, it offers a practical implementation of quantitative finance techniques.

Step 5 - Portfolio Optimization

Objective: Optimize a portfolio using historical stock data to maximize returns for a given level of risk.

Methodology:

Calculate **expected returns** and **covariance matrix** using historical data.

Generate the **Efficient Frontier** to identify the best risk-return tradeoff.

Optimize portfolio for the **Maximum Sharpe Ratio**.

Perform **Discrete Allocation** for real-world implementation of the portfolio.

Results:

Optimized Weights: Example - AAPL (30%), MSFT (20%), TSLA (49%).

Performance:

Expected Return: **48.8%**

Volatility: **44.1%**

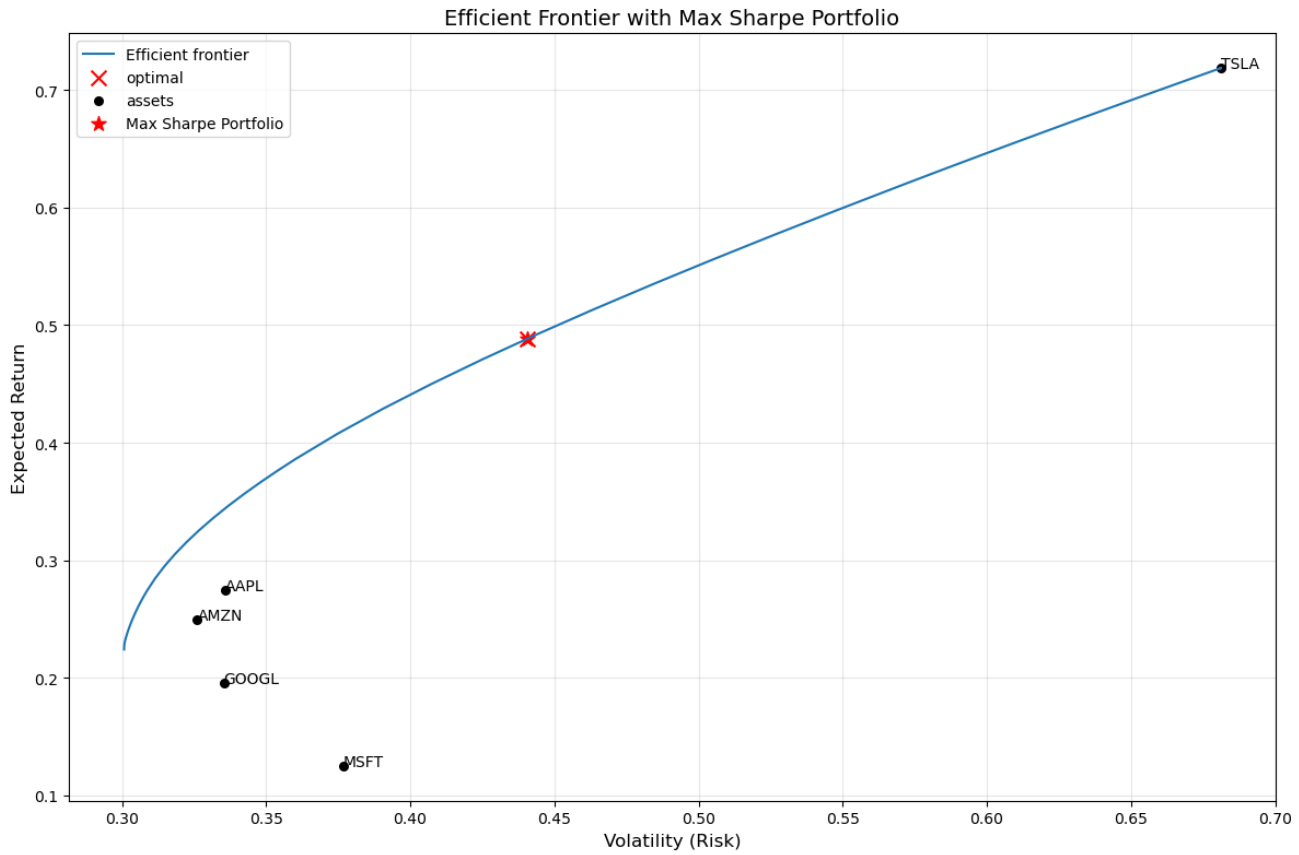
Sharpe Ratio: **1.11**

Allocation: Invest $100,000 as:

AAPL: 158 shares

MSFT: 54 shares

TSLA: 199 shares.



Step 6 - Backtesting Strategies

Objective: Test the performance of a Moving Average Crossover strategy using AAPL stock data (2015–2023).

Strategy:

Buy Signal: When 50-day SMA > 200-day SMA.

Sell Signal: When 50-day SMA ≤ 200-day SMA.

Results:

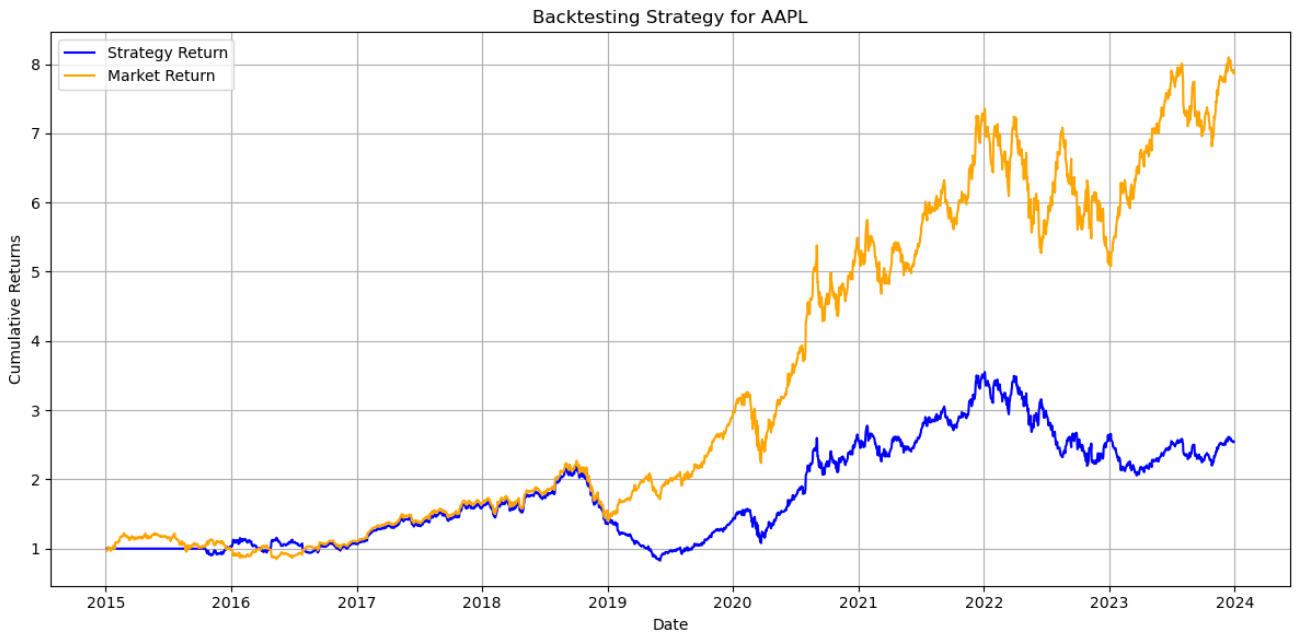
Market Return (Orange): Reflects buy-and-hold performance, achieving higher returns during prolonged bullish trends.

Strategy Return (Blue): Aims to mitigate risk, performing better during bearish or sideways markets.

Insights:

While the strategy underperforms during strong bullish trends (e.g., 2020–2021), it offers stability and reduced drawdowns.

Highlights the potential of simple technical indicators for systematic trading.



Key Takeaways

1. **Comprehensive Stock Analysis:**
    1. Combined **quantitative data analysis** with **machine learning models** to uncover actionable insights into stock price trends, market behavior, and risk-reward balance.
2. **Enhanced Predictive Accuracy:**
    1. **Linear Regression** and **Random Forest Regressor** models were effective in predicting stock prices, with Random Forest achieving superior R-squared and lower Mean Absolute Error.
    2. Advanced techniques like **LSTM** can further refine forecasting accuracy for sequential financial data.
3. **Sentiment-Driven Predictions:**
    1. **Sentiment Analysis** integrated market sentiment (via VADER) from financial news, providing additional context to price movements and improving predictive reliability during market shifts.
4. **Visualizing Trends and Risks:**
    1. **Interactive dashboards** (using Plotly Dash) presented key insights in an accessible format, empowering decision-making with dynamic visualization of sentiment, price trends, and risk indicators.
5. **Optimized Investment Strategies:**
    1. Portfolio Optimization using **Mean-Variance Analysis** and visualization of the **Efficient Frontier** highlighted optimal portfolio configurations for maximizing returns while minimizing volatility.
6. **Backtesting for Strategy Validation:**
    1. Simulated trading strategies, like RSI-based systems, demonstrated the effectiveness of technical indicators in generating consistent returns and mitigating drawdowns under various market conditions.

CONCLUSION

- This end-to-end project successfully integrated **EDA**, **machine learning**, **sentiment analysis**, **portfolio optimization**, and **backtesting** to provide a holistic approach to stock market analysis and prediction. These methodologies empower traders, analysts, and investors to make data-driven decisions in a dynamic financial landscape.