

# Analysing the US Stock Market and Commodities Data

Srihari Kamath

Ashtavinayak Pande

Chinmay Shanbhag

## 1) Abstract:

This project presents a comprehensive analysis of the US stock market and commodities data from 2020 to 2024. The study focuses on trend analysis, correlation analysis, event impact assessment, and forecasting. Using data analytics techniques and machine learning models, the project aims to provide insights into market behaviour, particularly in response to significant events like the COVID-19 pandemic. We employed various methodologies to extract meaningful patterns and predictions from the complex financial data.

## 2) Introduction:

The financial markets, particularly the US stock market and commodities sector, have experienced significant volatility and transformation in recent years. This project seeks to unravel the complexities of these markets by analysing historical data, identifying trends, and forecasting future movements. The study period, spanning from 2020 to 2024, encompasses the tumultuous era of the COVID-19 pandemic, providing a unique opportunity to examine market resilience and adaptation to unprecedented global events.

## 3) Objectives:

The primary objectives of this research project are:

- To conduct a comprehensive trend analysis, identifying patterns and cyclical movements in stock prices and commodity markets.
- To perform correlation analysis, exploring interdependencies between various stocks and commodities to understand their relationships.
- To assess the impact of significant economic and geopolitical events, particularly the COVID-19 pandemic, on market behaviour.
- To develop and evaluate forecasting models for projecting future trends based on historical data.

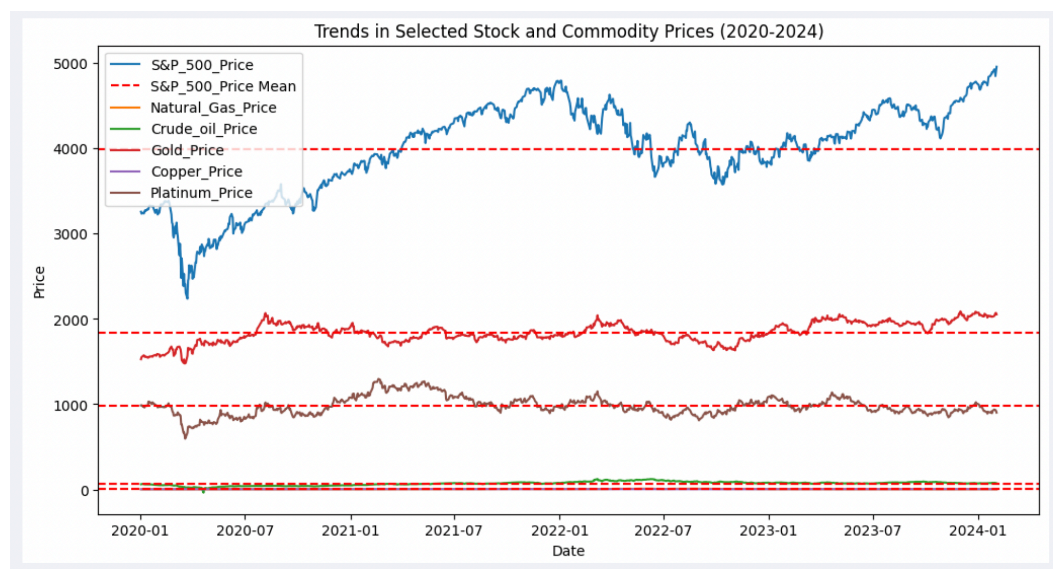
## 4) Methodology:

### 4.1) Data Collection and Preprocessing:

- **Data Source:** The dataset on the US Stock market was obtained from Kaggle (Link in Reference section).
- **Data Cleaning:**
  1. Null value check was performed across all columns.
  2. Irrelevant columns (Platinum\_Vol and Unnamed:0) were removed.
  3. Missing values were imputed using the mean of respective columns.
- **Data Transformation:** Date formats were standardized to ensure accurate time-series analysis.

### 4.2) Exploratory Data Analysis (EDA):

- **Dataset Overview:** The dataset contained 39 columns and 1013 rows, including features such as dates, stock prices, commodity prices (natural gas, crude oil, copper, gold), and trading volumes.
- **Visualization:** Trends of selected stocks and commodities were plotted to observe key patterns and relationships over time.



**Fig1**

The fig above shows the time scale of stocks and commodities over the period of 4 years.

#### 4.3) Trend Analysis:

- **Linear Regression:** Applied to examine trends in Apple's stock price and trading volume from 2020 to 2024.

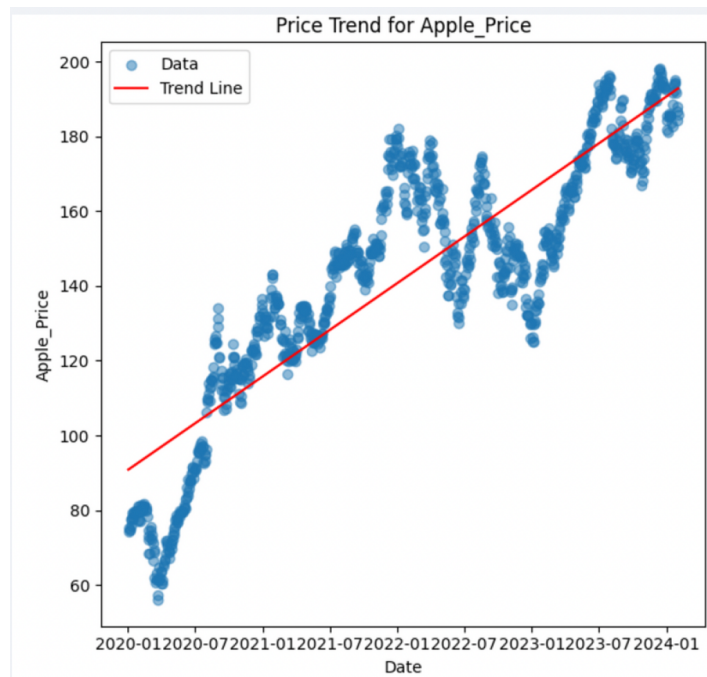


Fig 2

Above figure shows the upward trend for apple stock prices over the period of four years.

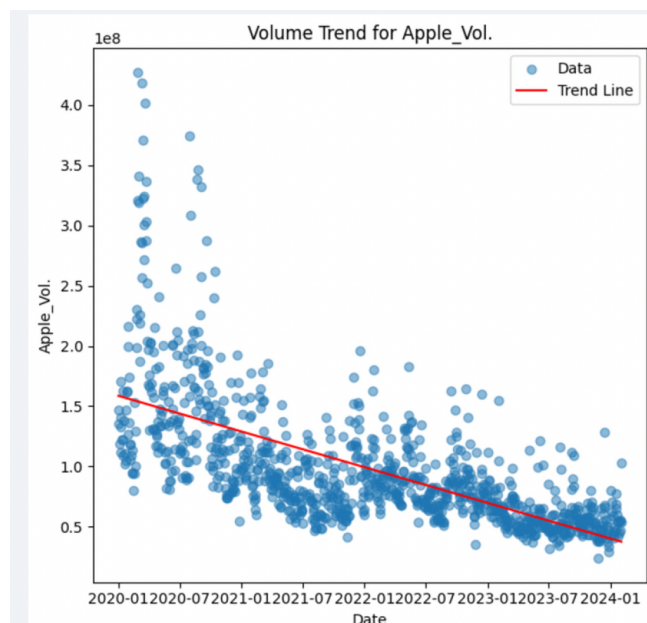


Fig 3

Fig 3 shows decreasing trading volume may indicate reduced speculative activity as prices stabilize.

#### 4.4) Event Impact Analysis:

- **Focus Event:** The impact of COVID-19 on stock prices was analysed.
- **Time Periods:** The analysis covered pre-COVID, during COVID waves, and post-COVID phases.

	Apple_Mean	Apple_Volatility	Apple_Percentage_Change \
Pre-COVID	77.902750	3.299683	9.877085
First Wave	70.067903	6.546386	3.862994
Second Wave	125.765763	8.049987	-13.710514
Omicron Wave	171.322712	5.952488	-13.922481

	S&P500_Mean	S&P500_Volatility	S&P500_Percentage_Change
Pre-COVID	3277.780750	95.037565	10.277840
First Wave	2770.236613	207.575514	11.786228
Second Wave	3676.620847	115.977982	0.909171
Omicron Wave	4571.575424	149.490284	4.092908

Table 1

This table shows the stock changes over the Covid-19 period with differentiating as per pandemic waves.

- **Selected Stocks:** Apple Inc. and S&P 500 were chosen as key representatives of the stock market.

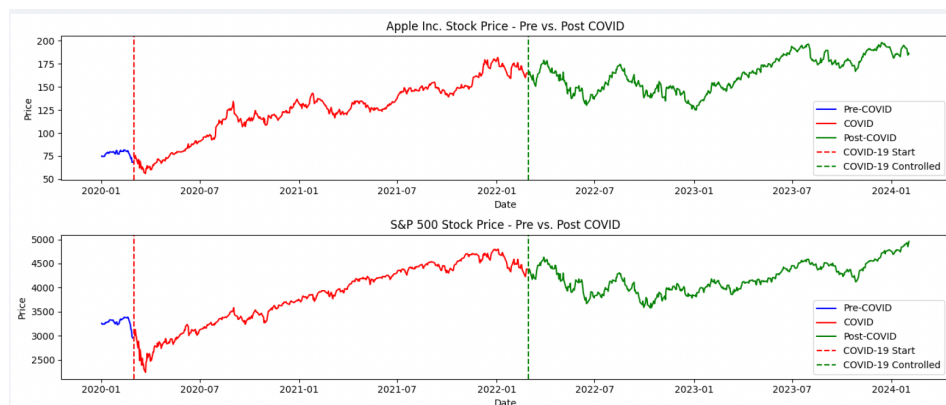


Fig 4

Fig 4 shows the stock prices fell after covid then started to rise during Covid and then started to stabilize post Covid.

#### 4.5) Forecasting:

- **Models Employed:**

1. LSTM (Long Short-Term Memory)
2. ARIMA (Autoregressive Integrated Moving Average)
3. SARIMA (Seasonal ARIMA)

- **Performance Metrics:** MAPE (Mean Absolute Percentage Error) and SMAPE (Symmetric Mean Absolute Percentage Error) were used to evaluate model performance.

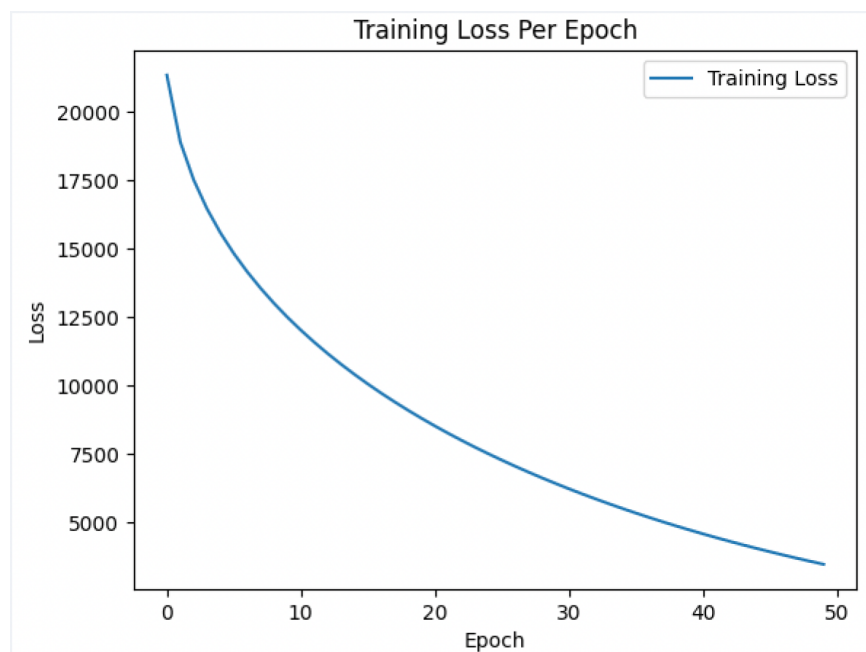


Fig 5

Fig 5 shows the training loss per epoch for LSTM. The following fig shows the Performance metrics for LSTM.

**Performance Metrics:**

- Training Error: MAPE: 2.61%, SMAPE: 2.65%
- Test Error: MAPE: 5.02%, SMAPE: 4.87%

ARIMA & SARIMA failed to give reliable results thus are not included in the report. We achieved relatively accurate predictions from LSTM but results varied due to market volatility.

## 5) Results and Discussion

### 5.1) Trend Analysis

- **Apple Stock Price:**

1. A positive trend was observed, indicating steady growth over time.
2. The linear regression trend line highlighted an overall upward trajectory.

- **Apple Trading Volume:**

1. A negative trend was noted, suggesting a gradual decline as the stock price increased.
2. This decline in volume may indicate reduced speculative activity as prices stabilized.

### 5.2) Event Impact Analysis (COVID-19)

- **Volatility Patterns:**

1. S&P 500 experienced peak volatility during the First Wave.
2. Apple's volatility peaked during the Second Wave.

- **Price Recovery:**

1. Both Apple and S&P 500 showed significant price recovery post-COVID.
2. The highest mean prices for both were observed during the Omicron Wave.

- **Percentage Changes:**

1. Apple experienced sharp percentage declines during the Second and Omicron Waves.
2. S&P 500 showed more stability, with a notable positive change during the First Wave.

- **Market Resilience:**

1. The trends highlighted the market's ability to adapt to the pandemic's economic impacts.

### 5.3) Forecasting Results

- **LSTM Model Performance:**

1. Training Error: MAPE: 2.61%, SMAPE: 2.65%
2. Test Error: MAPE: 5.02%, SMAPE: 4.87%
3. The model achieved relatively accurate predictions, though performance varied due to market volatility.

- **Other Models:**

1. ARIMA and SARIMA models failed to produce reliable results and were not included in the final analysis.

### 6) Conclusion:

This comprehensive analysis of the US stock market and commodities data from 2020 to 2024 has revealed several key insights:

- The stock market, particularly exemplified by Apple Inc., showed resilience and growth despite the challenges posed by the COVID-19 pandemic.
- Trading volumes for Apple displayed a declining trend, possibly indicating a shift in investor behaviour or market dynamics.
- The impact of COVID-19 on the stock market was significant but varied across different phases of the pandemic, with markets showing remarkable recovery and adaptation.
- The LSTM model demonstrated promising results in forecasting stock prices, outperforming traditional time series models like ARIMA and SARIMA.
- The study underscores the importance of considering external events and their complex interactions with market behaviour in financial analysis and forecasting.

These findings contribute to a deeper understanding of market dynamics during periods of global crisis and provide valuable insights for investors, policymakers, and financial analysts. Future research could explore more sophisticated modelling techniques and incorporate a broader range of economic indicators to enhance predictive accuracy.

## 7) Future Aspects:

- **Percentage Change Analysis:** Instead of using direct stock values, future iterations of the project could focus on percentage changes. This approach would provide a more normalized view of price movements across different stocks and commodities, allowing for better comparisons regardless of their absolute price levels.
- **Random Walk Model:** Implementing a random walk model could offer insights into the efficiency of the market. This model assumes that stock price changes are random and unpredictable, which could be compared against the current predictive models to assess market predictability.

## 8) References:

- [Kaggle Dataset of US stock market.](#)
- [Regression references.](#)
- [How to analyze financial datasets.](#)
- [Autoregressive integrated moving average](#)