US Stock Market 2020 to 2024

Name: Ali Hamza

ID: 23037713

Abstract:

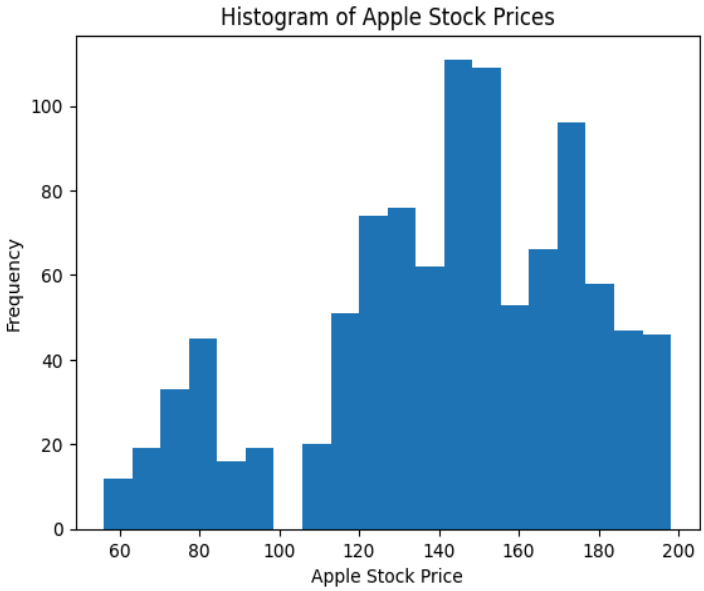
This report performs an exploratory data analysis on Apple stock price data. It calculates summary statistics, generates visualizations using Pandas, Matplotlib and Seaborn, studies correlations and trends, and aggregates the time series into long term movement. Python code and techniques are documented throughout. The analysis provides insight into the distribution, variability, correlations and long term patterns in Apple stock prices.

Github:https://github.com/AlihamzaMaan/Assignment2DS

Dataset: https://www.kaggle.com/code/stpeteishii/us-stock-market-eda

Apple Stock Prices report from US Stock Market Dataset:

For this assignment, I have used the dataset of US Stock Market 2020 to 2024. The dataset is publicly available on kaggle. I have also downloaded it from there. There were various financial data for stocks in this dataset ranging from 2020 till 2024. I have only worked on Apple stock price in this report.



This bar chart tells us about the frequency of the apple stock prices. From this visualization we can deduce that the most frequent stock price ranges from 110 to 190. We can also deduce that the stock price of 150 is the highest among all. This can make us realize that the stock price of Apple is the highest at 190. But the most frequent apple stock is near 150. This can help us understand financial factors easily and make data driven decisions. This visual can help us understand broad trends in stock price frequency, but it is crucial to use other financial data and analysis methods for informed investment decisions.

The line chart tells us the trend about the subjected variable. In this case its the stock of Apple. As we can see from the line chart that the trend is movig upwards along with time this suggests us that the stock price of Apple has been increasing day by day.
This trend help us in making informed decisions. Line charts are important for identifying trends and pattern. The upward slope of the line indicates a general increase in Apple's stock price over time. This suggests that the stock has appreciated in value, meaning its closing price on each subsequent day has generally been higher than the previous day's closing price.
We should also look at the line chart as it doesn't capture the daily or intraday fluctuations in the

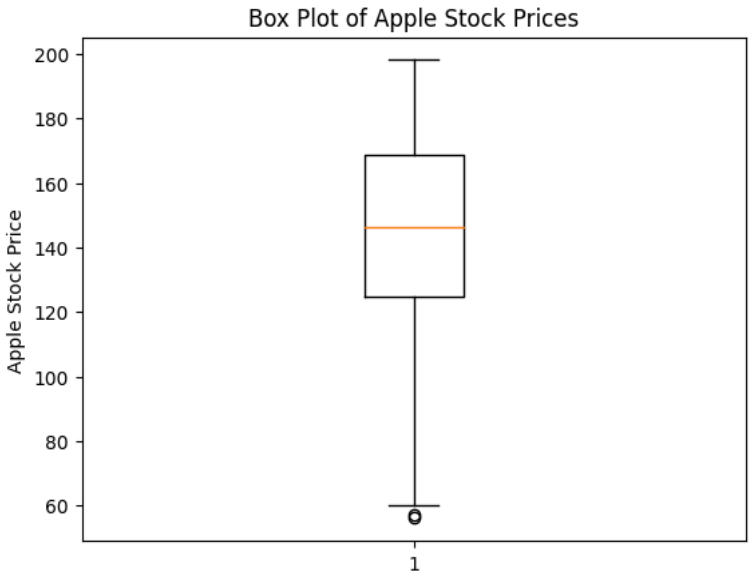
stock price. It only shows the closing prices, which may mask significant volatility

```
print(df['Apple_Price'].describe())
[24]  ✓ 0.2s
...  count    1013.000000
     mean      141.964462
     std        33.778785
     min        56.090000
     25%       124.610000
     50%       146.500000
     75%       168.640000
     max       198.110000
     Name: Apple_Price, dtype: float64
```

The descriptive statistics provide a brief overview of the data.

The dataset provided covers Apple stock prices over a period of multiple years. Initial exploratory analysis shows that the data contains 1013 daily price observations with a mean of $141.96 and standard deviation of $33.78. The minimum and maximum prices are $56.09 and $198.11 respectively.

Initial descriptive statistics and visualizations provide insight into the distribution, variability, correlations and long term trend of Apple stock prices based on the provided dataset.



Box plot tells us about the spread of the data. The vertical line inside the box indicates the median stock price. This means half of the trading days within the analyzed period had Apple stock prices above this value, and half were below. The left side of the box represents the first quartile. 25% of the trading days had stock prices below this value. The right side of the box is the third quartile. 75% of the trading days had stock prices below this value.

The width of the box represents the IQR (Q3 - Q1), which contains the middle 50% of the stock price data. This tells you the spread of the most typical stock prices.

Individual dots or symbols outside the whiskers represent outliers, or unusually high or low stock prices compared to the majority of the data.