Case Study

Statistical analysis of stock market

Instructor:-Ramesh Athe

Done by :-Devansh Purwar

● Introduction

Stock market known for risk and multibagger returns in current times.Although various investors do every research in a company for which they are going to buy stocks by own & a single share of the stock represents fractional ownership of the corporation in proportion to the total number of shares.Various Fintech startups start working in trading algorithms to analysis each stock wrt to other companies stock and provide consultancy to investors.

The stock market prediction has been quite a desired research topic for years. Although there is the Efficient Market Hypothesis which proposes that stock prices follow a random walk and, thus, cannot be predicted [1-2], there is also the Adaptive Market Hypothesis states that prices of stocks can be predicted [3] and the latter is what this paper would like to focus on. When it comes to making trading decisions and stock price predictions, financial experts mainly make use of Technical Analysis, Fundamental Analysis [4-6], or both. Fundamental Analysis deals more with the business performance of a company while Technical Analysis makes use of the corporation's historical stock data and a few statistical tools to interpret those figures to come up with better predictions. This paper is concentrated on using several statistical methods for stock market analysis and prediction. Specifically, the following methods would be discussed: Exponential Smoothing, Mean Square Error (MSE), and AutoRegressive Integrated Moving Average (ARIMA).

The following libraries were used in conducting the statistical analysis:

- Matplotlib [11] was used for plotting the graphs for the actual time-series as well as the predicted trends.
- Pandas [12] was used for reading values from CSV files as DataFrames.
- Numpy [13] was used to perform matrix operations like flip, reshape, and create random matrices.

● Need of Statistical analysis

There is a lot of data involved in share market trend analysis and in order to start analysing, we must first identify which sector we must pick. The focus can either be on the type of industry like the pharmaceutical sector or in the kind of investments, like the bond market. Only when you select your sector, can you start analysing it. The stock market trend analysis includes both external and internal forces that affect it. Changes in a similar industry or the introduction of a new governmental regulation qualify as forces impacting the market. Analysts then take this data and attempt to predict the direction the market will take, moving forward.
As an investor, we must understand the sense behind the stock market trading trend. As you wouldn't drive your car on the wrong side of a one-way street, similarly, it's advisable not to trade against the trends in a market.

● Objectives

• To study the characteristics of stock prices of selected companies by using descriptive statistics.
• To analyze monthly fluctuations in the stock prices using the simple moving average method.
• To study the price movements in the selected stocks using the money flow index.
• To study the current trend and strength of the three selected stocks using the relative strength index.
• To examine the risk and volatility associated with stock prices using a beta index.

● Explanation of jupyter notebook program:-

- Imported various python libraries
- Fetched datasets of Apple, Blackberry, Microsoft, Dow Jones
- Data visualization of closing prices of these stocks on a monthly, yearly basis using matlplotlib library

- **Resampling and analyzing stock price via monthly and weekly value means** of closing prices on a quarterly basis
- Each stock price change for calculating returns for intraday trading and graphical representation
- Each stock price percentage change wrt previous market closing to calculate risk and graphical representation
- Calculated **the volatility of Apple stock** by using function and graphical representation

```
daily_pct_change_ap.rolling(min_periods).std() *
np.sqrt(min_periods)
```

- **Boxplot** of daily returns of apple stock

- Insights of **Probability Distribution** of daily returns

count	1512.000000
mean	0.001275
std	0.018656
min	-0.128647
25%	-0.006477
50%	0.000899
75%	0.010031
max	0.119808

- **Median** → 0.0008990978620975199
- **Skew** → -0.08089673
- **Kurtosis** → 6.59610886
- **SkewtestResult**(statistic=-1.28841983, pvalue=0.19759986)
- **KurtosistestResult**(statistic=14.56808655, pvalue=4.48271671e-48)
- The returns aren't **normally distributed** or not , a Q-Q plot will give us graphically clear information.
- The straight line is what we expect for **a normal distribution**, while the blue line is what we get from our data. It's clear that the quantiles of our dataset aren't comparable with the quantiles of a normal distribution with the same mean and standard deviation.

- So, returns aren't normally distributed and this makes models like **Geometric Brownian Motion** (which assumes normal returns) only an approximation of the reality.

- We have ensured the non-normality of the returns probability distribution, I also figured out at the raw time series. It's clear that there are time periods with high volatility and other periods with low volatility. This phenomenon is called volatility clustering and it's very common in the stock market. Practically speaking, the standard deviation changes during time, making the time series non-stationary.

- A closer look at the **20-days rolling standard deviation** shows that it's not a constant value, but it has spikes and oscillations. The fat tails of the distribution may be caused by these volatility spikes, which create non-neglectable outliers.

- Plotted the partial autocorrelation function, which gives us some ideas about the autocorrelation of the time series and the possibility that this **autocorrelation can be used for prediction purposes in some ARIMA models.**

- I've shown a simple statistical analysis of Apple stock price. The returns aren't normally distributed, since they are skewed and have fat tails. The time series is not stationary because the standard deviation changes over time. The autocorrelation of returns time series is very low at almost any lag, making ARIMA models useless.
- So, predicting stock prices using statistics and machine learning is a great challenge.

- Represented Apple shares volume traded graph over years
- **Used rolling means /moving average**(Trends and Seasonality) to analyze data points by creating a series of averages of different subsets of the full data set
- Find out the **range of Apple stock** closing price and volume

Probability Distribution of daily Closing price

- **Mean** → 48.104024427908435

- **Standard deviation** →25.08002387240191
- **Covariance** → 52.1370595717793

- Also, find out the **Correlation between the closing price and volume of stock** traded

- **Linear Regression Using Stocks closing price and Volume of stocks** traded
- Estimated coefficients:
  b_0 = 46.688442038431624
  b_1 = 9.699031025972516e-09

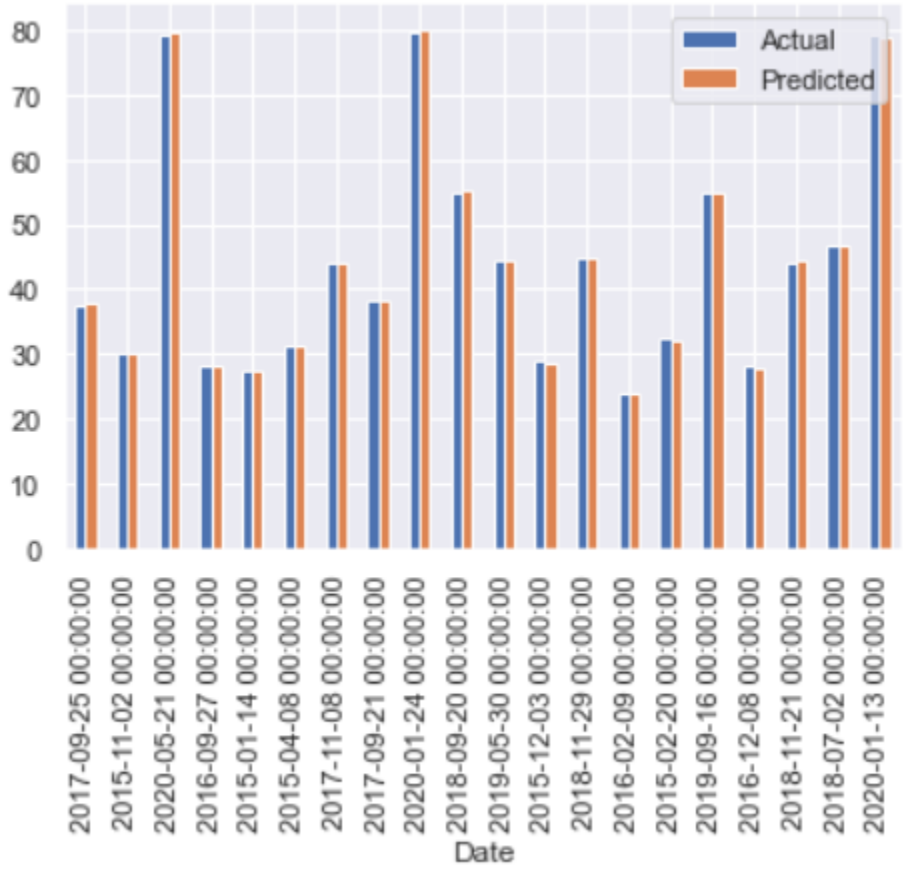
- Also, find out **Central Moments** of the closing price of the stock

```
        0th moment =  1.0
        1st moment =  0.0
 2nd moment =  629.0075974402504
 3rd moment =  26162.73552939183
4th moment =  2053524.4047724027
```

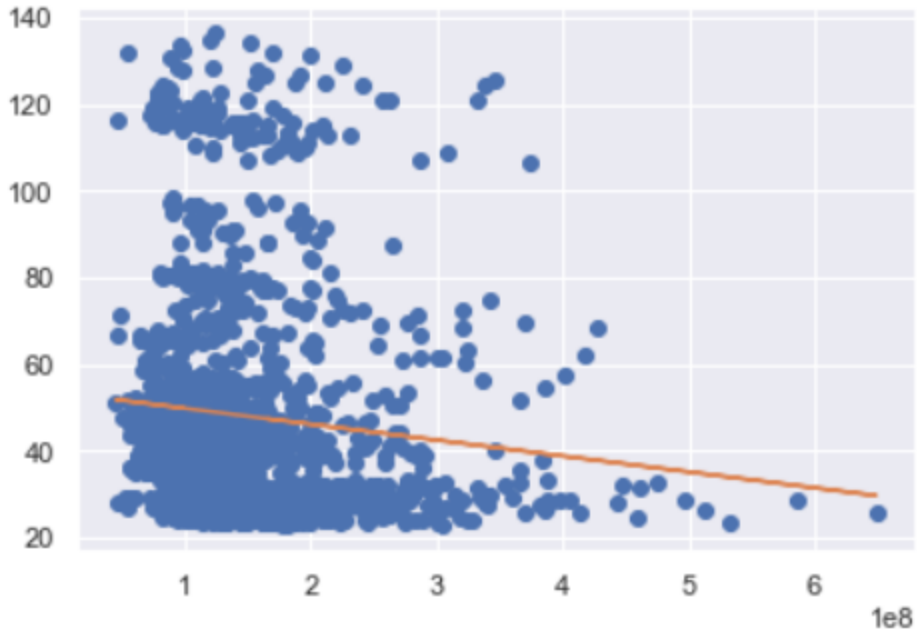
- **Curve fitting of stock price and volume in a straight line**
- **Used linear regression for prediction/estimation** of apple stock price

Results:-

- **Prediction of closing stock price**



- **Curve fitting closing price of stock and volume traded**



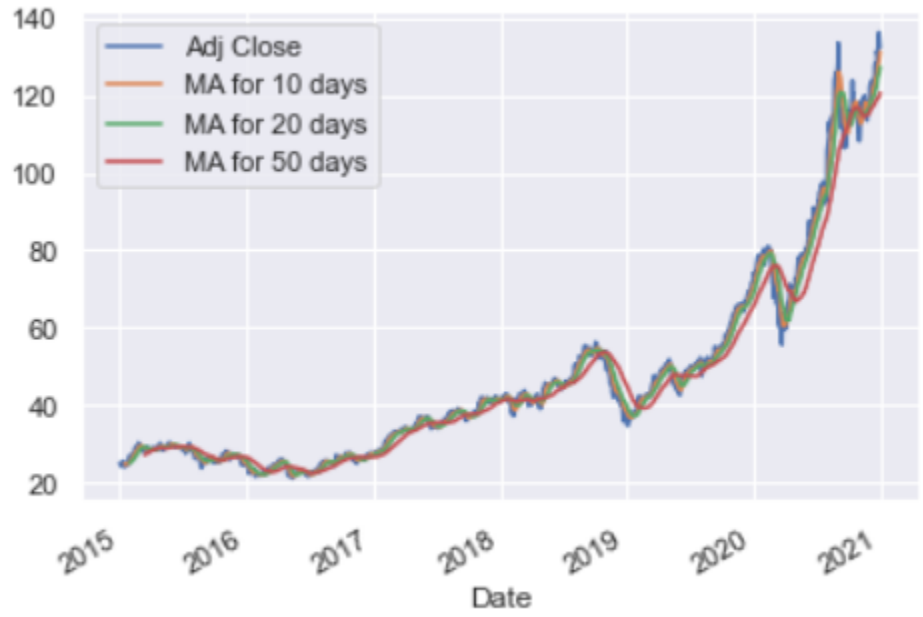
- **Linear regression Using Stocks closing price and Volume of stocks traded**

```
Estimated coefficients:
b_0 = 46.688442038431624
b_1 = 9.699031025972516e-09
```

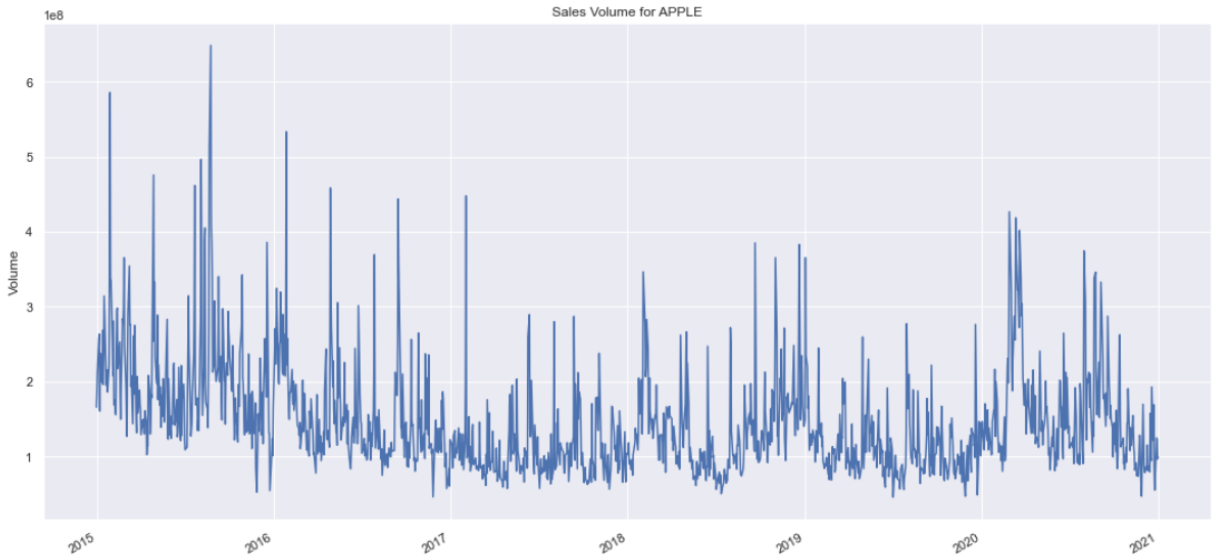
- **Correlation**

	Close	Volume
Close	1.000000	-0.104599
Volume	-0.104599	1.000000

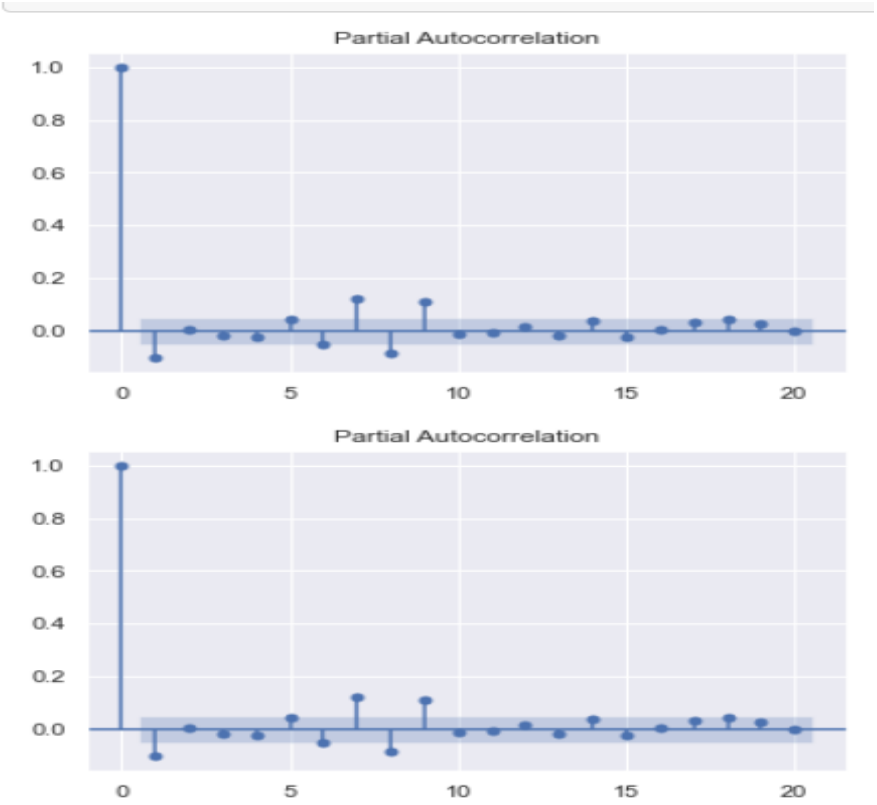
- **Moving Average**



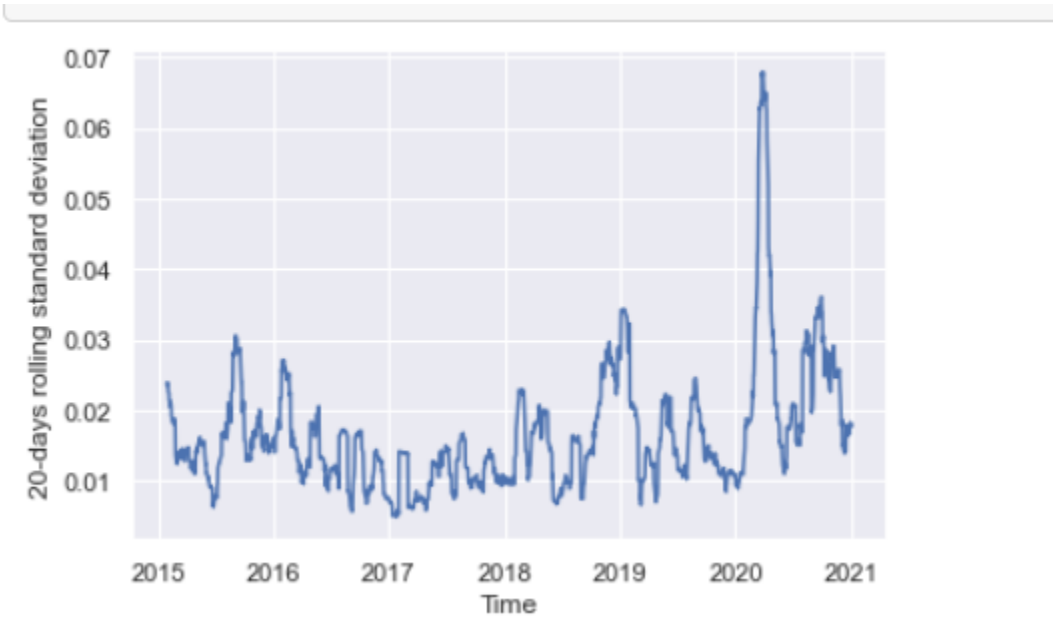
- **Sales Volume of Apple**



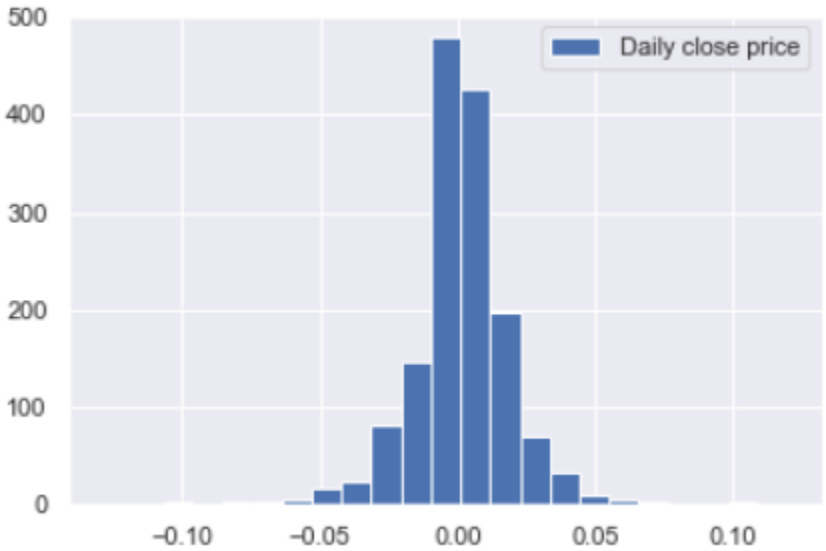
- **Partial Autocorrelation**



- **20 days rolling standard deviation**

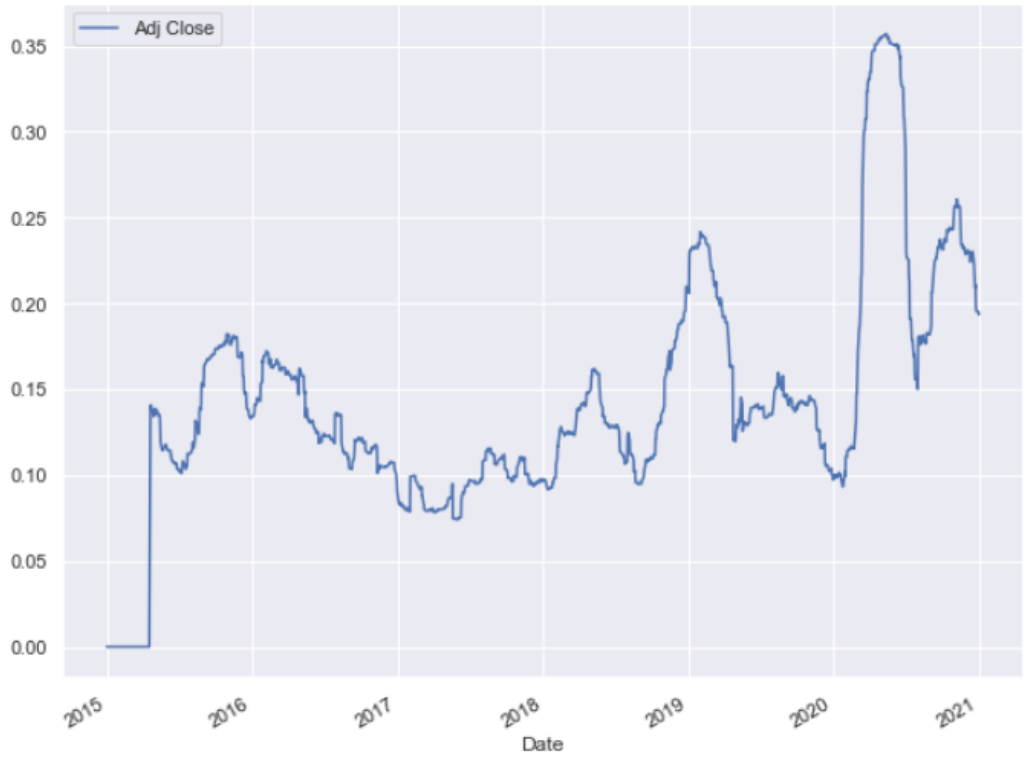


- **histogram for calculating returns with the help of daily close price**

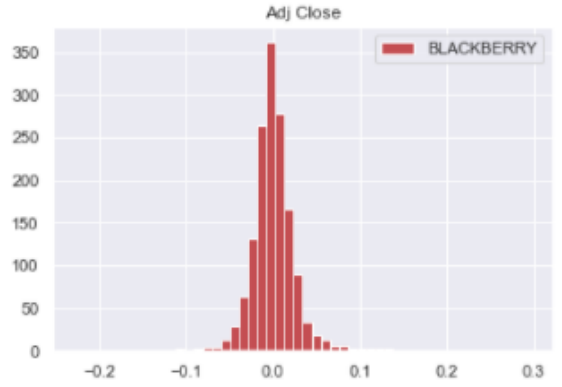


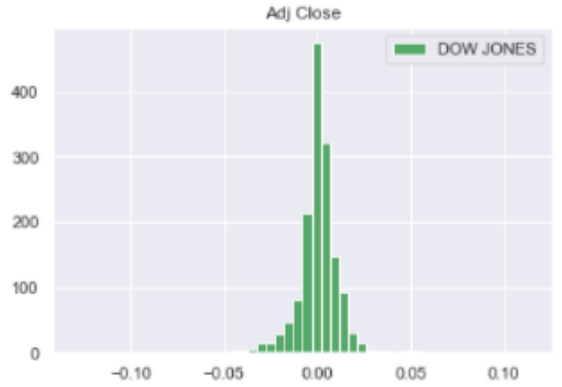
- **Box plot**

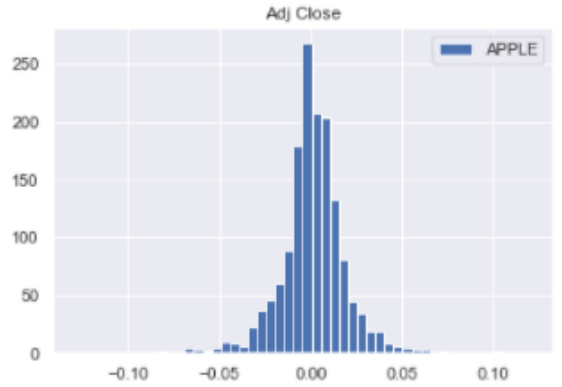
- **Volatility of stock**

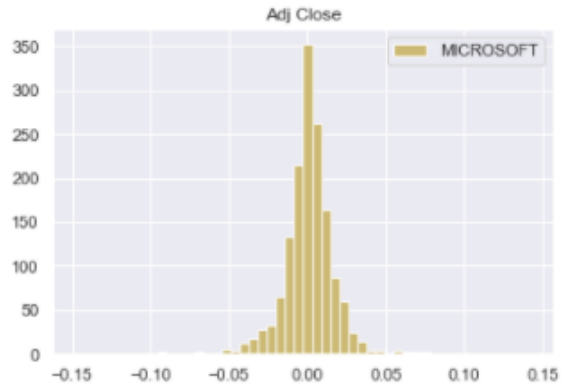


- **Stock price percentage change wrt previous market closing to calculate risk**

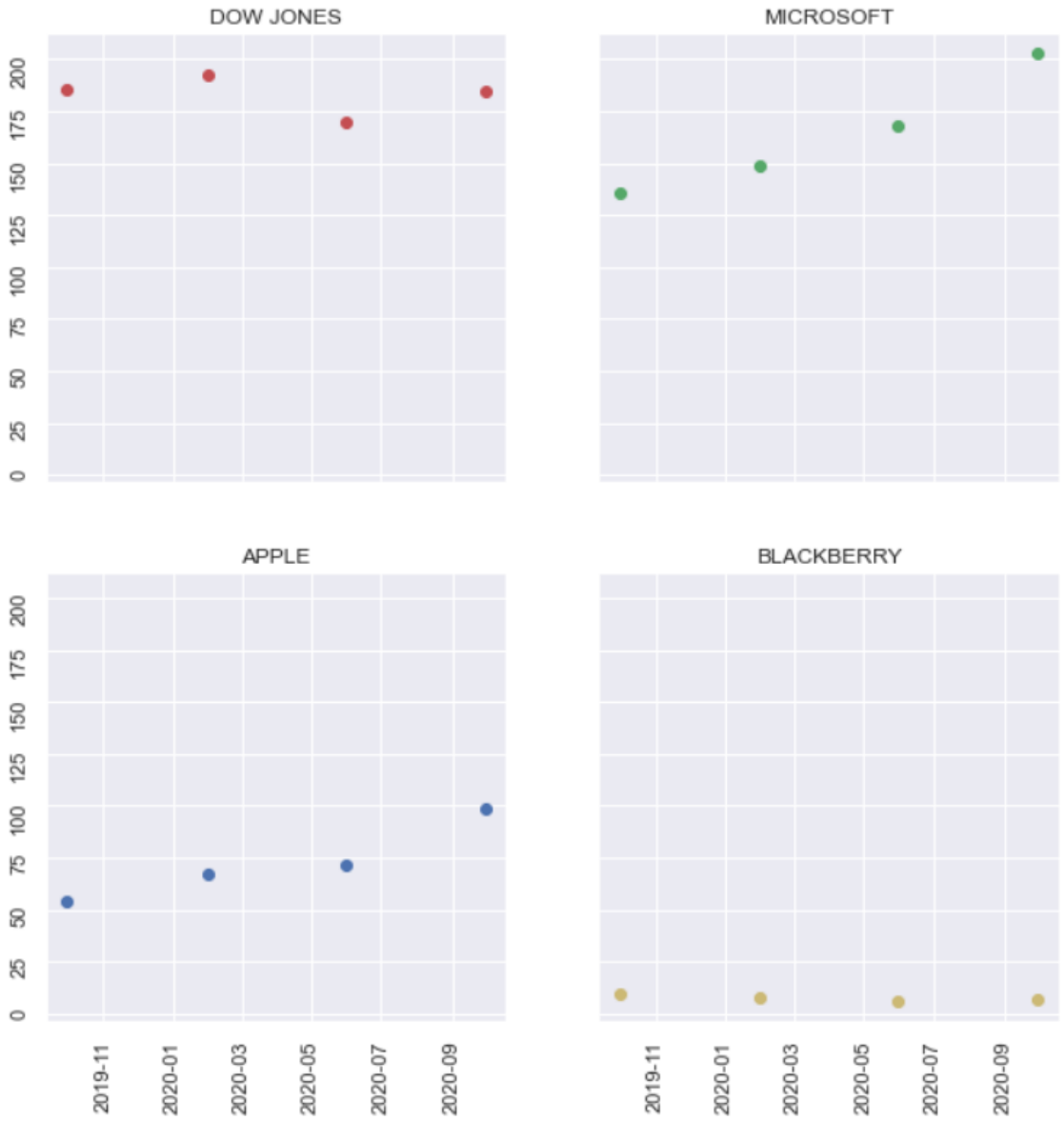








- **Resampling and analyzing stocks price monthly via mean values**



- **Conclusion**

The present study discussed the stock prices of the selected companies using statistical methods. The basic statistics along with box plot, histogram diagram, and normal p-plot helped to decipher the dynamics of the stock prices during the studied period. It is found that stock prices are not normally distributed and the stock price of Gammon is the poorest performer among all four. The simple moving average is applied to the spread of open-close prices and close price of the stocks which also showed a decreasing trend in stock prices in general. However, it also gave certain clues for investors to make necessary decisions with stocks.

- Dow Jones,Apple,Microsoft stocks distribution are left skewed and Blackberry stock distribution is symmetrical from 2015 to 2021 Jan.
- There is gradual increase in stock price during new products launch and significant decrease in March 2020 due to COVID19.
- There has been a huge increase of 485.02% (119.80 USD) in the stock price of Apple since the past five years as compared to other stocks

In a simple statistical analysis of Apple stock price.The returns aren't normally distributed, since they are skewed and have fat tails. The time series is not stationary because the standard deviation changes over time. The autocorrelation of returns time series is very low at almost any lag, making ARIMA models useless.

The stock market is always considered a challenge for statistics. Somebody thinks that knowing the statistics of a market lets us beat it and earn money. The reality can be quite different.So, predicting stock prices using statistics and machine learning is a great challenge. Some results have been achieved using LSTM models, but we are very far from clearly modeling a stock market in a money-making way.

Dataset reference → Yahoo Finance