AI60003
ASSIGNMENT-1<STAGE 3> REPORT
DEBANJAN SAHA (19CS30014)

>> Brief Problem Statement: Analyzing the time series forecasting data of the stock prices of top software companies i.e FAANG(Facebook+Amazon+Apple+Netflix+Google) companies and analyzing the risk factors involved in investing in these companies and predicting the future stock prices using AI methods.

>> [IMPORTANT] EXTERNAL REFERENCE: THIS ASSIGNMENT WAS IMPLEMENTED BY ME IN THE FOLLOWING NOTEBOOK: GOOGLE COLAB NOTEBOOK (PLEASE WAIT A FEW SECONDS AFTER OPENING THE NOTEBOOK FOR THE IMAGES TO LOAD)

>> Dataset:

- ➔ **Link:** https://www.kaggle.com/specter7/amazon-amzn-historical-stock-price-data

- ➔ **Information:** The above link gives us a dataset containing the stock prices of FAANG companies ranging from 2016 to March 2021 (i.e the past 5 years)
- ➔ **Description of Dataset:** Each of the datasets for each company gives us information about 6 different things for each particular date. These are as follows:
  - ◆ **High =>** Highest Price of the stock for that particular date
  - ◆ **Low =>** Lowest Price of the stock for that particular date
  - ◆ **Open =>** Opening Price of the stock
  - ◆ **Close =>** Closing Price of the stock
  - ◆ **Volume =>** Total amount of Trading Activity
  - ◆ **AdjClose =>** Adjusted values factor in corporate actions such as dividends, stock splits, and new share issuance

>> A Quick Summary of Stage-2 :

- ➔ As a part of the goals of the problem that I had proposed, I had summarized my goals as the following in stage-2 of the assignment:-
  - Generating different data analysis figures to visualize the datasets
  - The change in the price of the stocks over time and analyzing the time series
  - The daily return of the stocks on average
  - The moving average of the various stocks
  - Correlation between different stocks
  - How can we attempt to predict future stock behaviour? (Predicting the stock opening/closing price of one/more companies using algorithms described below)
  - How much value do we put at risk by investing in a particular stock?
  - An attempt to answer the question of whether it is beneficial to invest in these stocks.
- ➔ Hence, in the following part, I will try to elaborate on each of these points using my implementation results and plots etc.

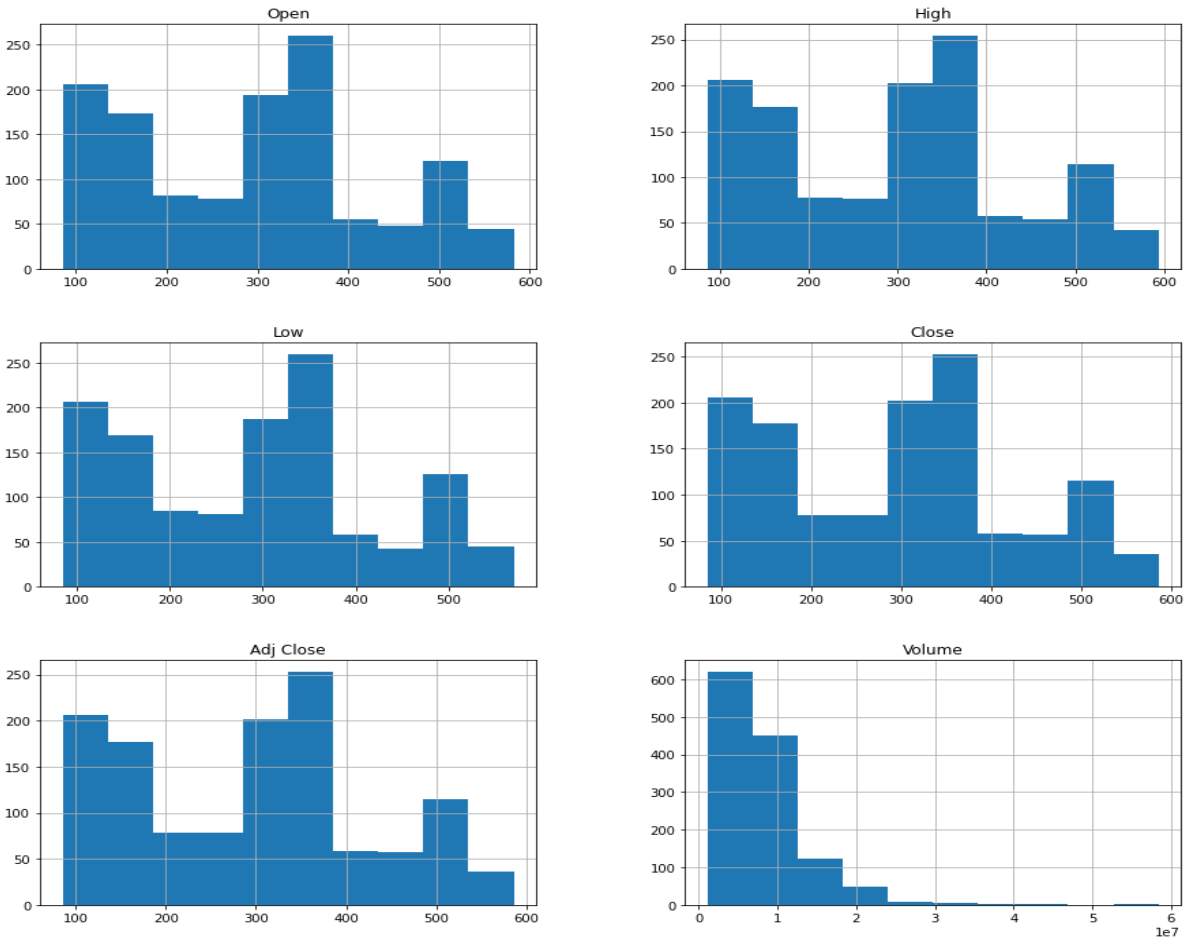
>> Implementation Report and Analysis:

1. Analyzing and Visualizing the Data:

- I used the **NetFlix** dataset to get a basic idea about the dataset and found the following basic information.
- As we can see, there are 1260 rows in this dataset of NetFlix and all of them are non-null. So there is no need for data imputation or filling the missing values.
- I also analyzed the datasets for the same for the other companies and found that there are no missing values at all. Hence it was a good dataset choice for me.

```
<class 'pandas.core.frame.DataFrame'>
Index: 1260 entries, 2016-03-28 to 2021-03-26
Data columns (total 6 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Open       1260 non-null   float64
 1   High       1260 non-null   float64
 2   Low        1260 non-null   float64
 3   Close      1260 non-null   float64
 4   Adj Close  1260 non-null   float64
 5   Volume     1260 non-null   int64
dtypes: float64(5), int64(1)
memory usage: 68.9+ KB
```

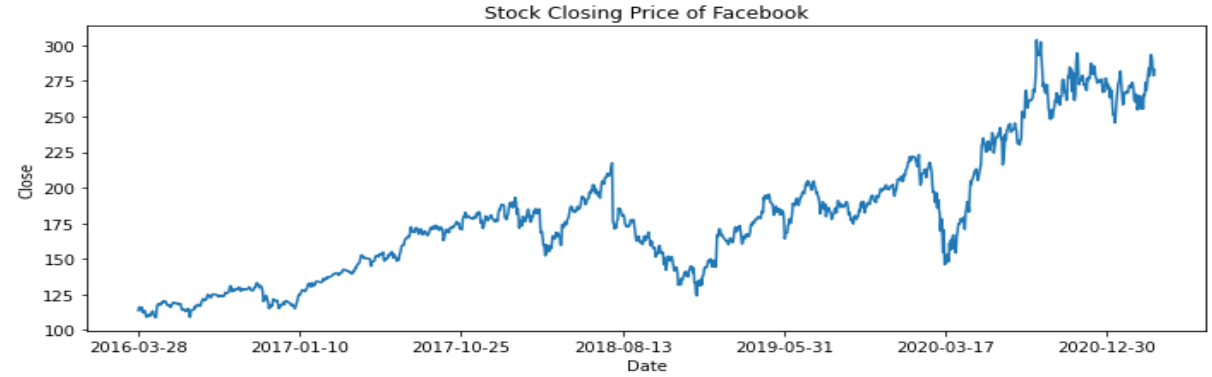
- If we see the distribution of each of the columns of this dataset, we can see the following histograms :

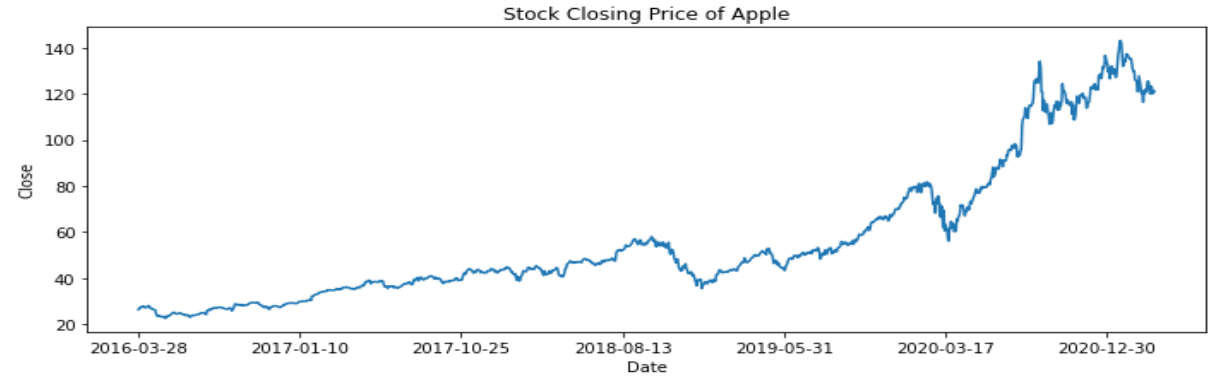


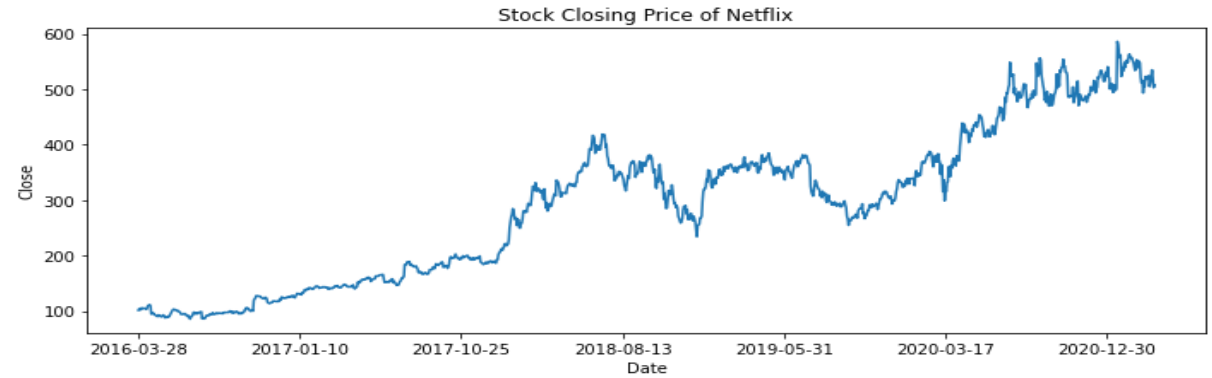
- We can see that volume of stock trading has a high distribution in the range of [0,1] and the other columns are somewhat distributed among all the ranges with significant values.

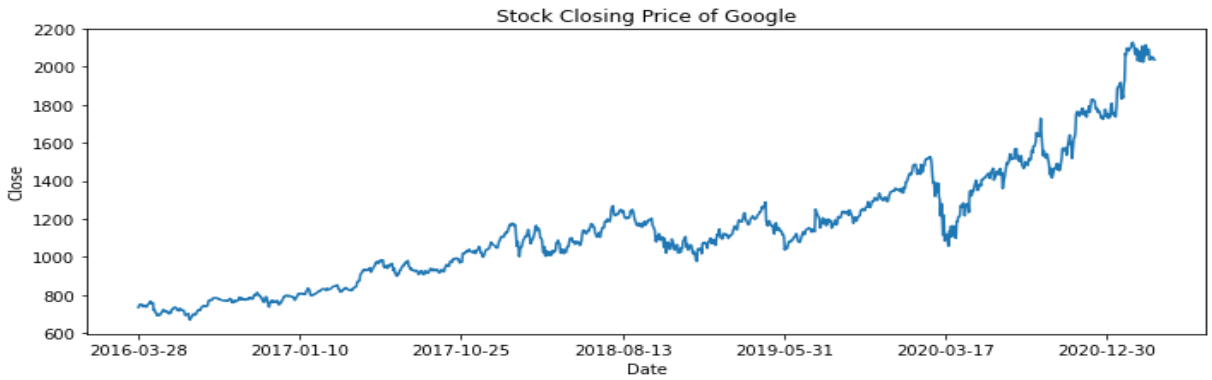
2. What was the change in the price of the stock over time?

- Here I plotted the time series stock closing price of each of the company and the plot is as follows

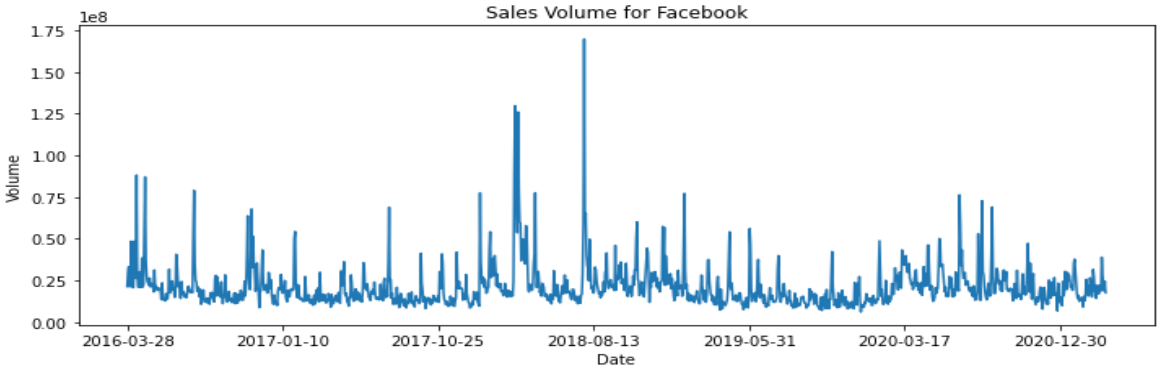


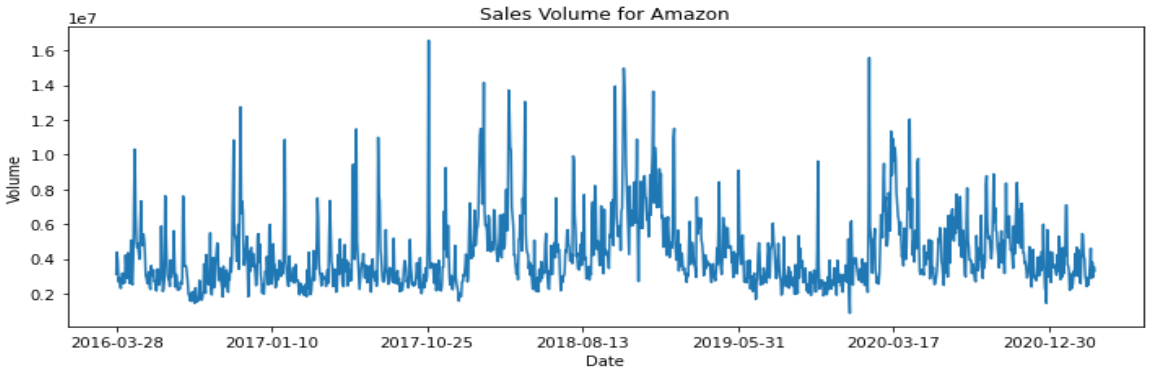


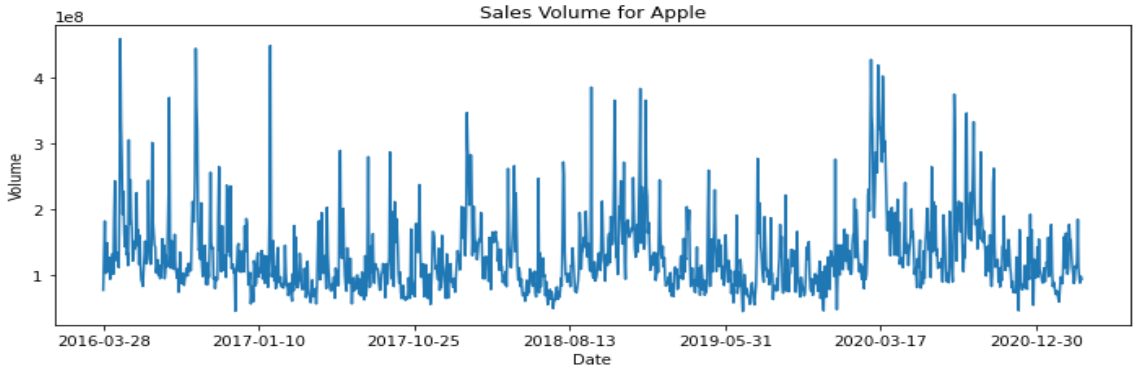


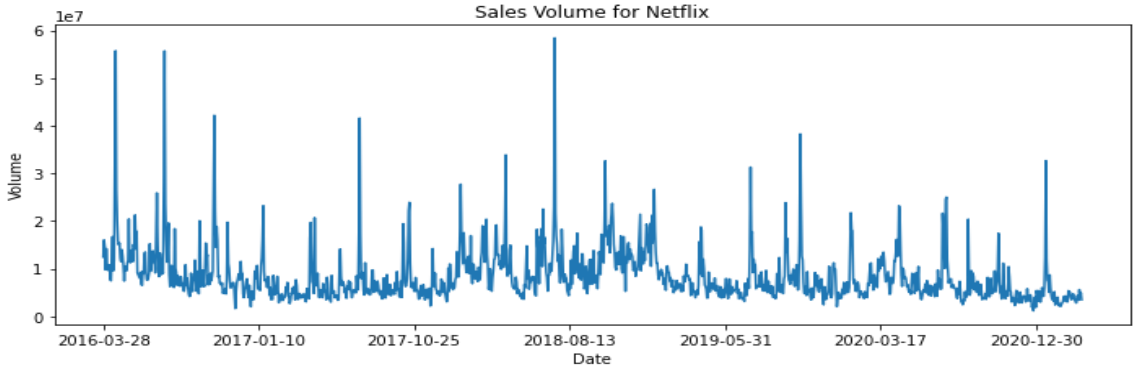


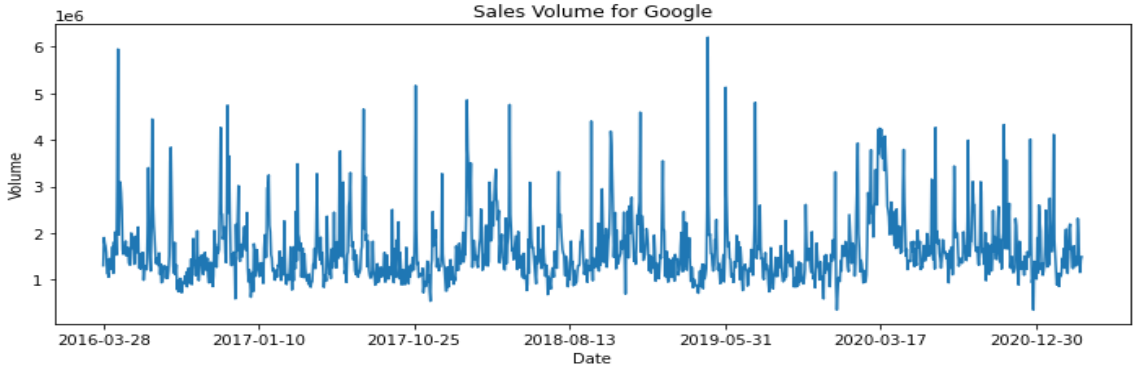
- We can see that all the companies have a very good upward curve in the time series of the stock closing price since 2016. **However, around 2020-March, due to the sudden wave of the pandemic, there is a sharp decline in every company's stock closing price**. It affected **Facebook** the most as can be seen from the graph.
- Now let us see another interesting plot for the volume traded per day for each company







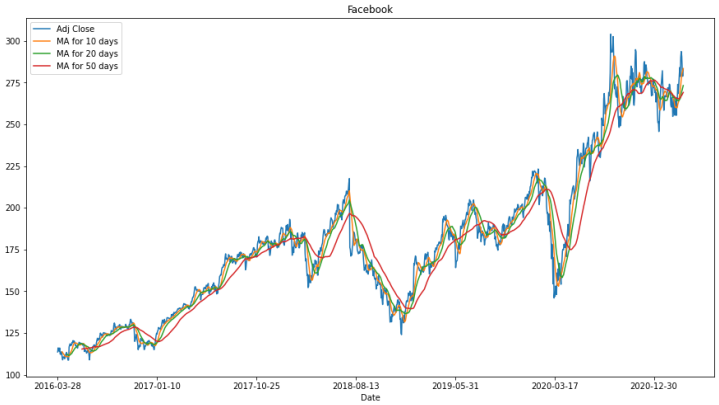


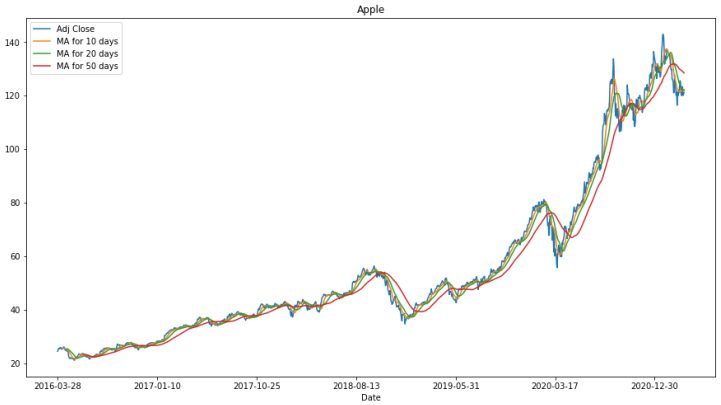


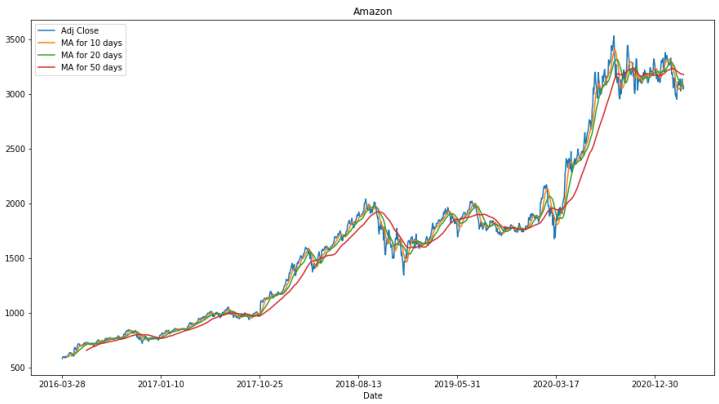
- From the above plot, we can see that all the companies had a sudden increase in the volume of trading during the pandemic. This is most prominent for Google, NetFlix, Apple and Amazon. **This is due to the fact that people sitting at home were more prone to incorporate the technologies and online services such as Amazon product delivery, google meet etc into their lives**.

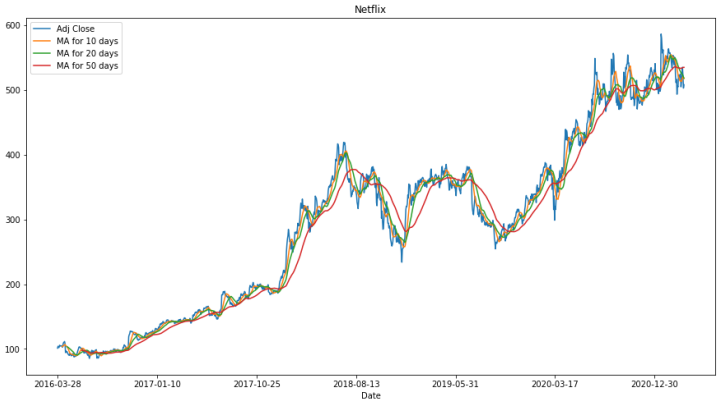
3. What was the moving average (MA) of the various stocks?

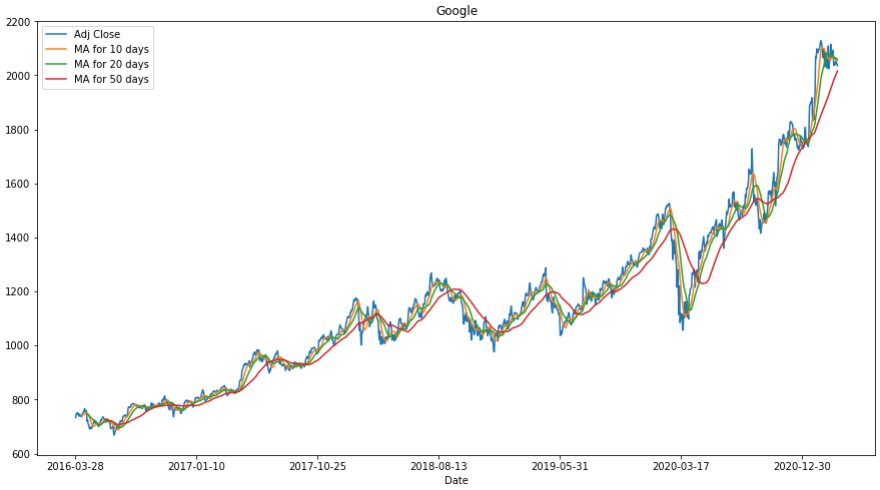
- I plotted the moving average of the adjusted value factors of the stocks for each of the companies and found that the moving average for 10,20 and 50 days were following the normal curve and did not deviate much from the original values. Hence, it proves the consistency of the stocks of these companies and shows why it is the best option to invest in.





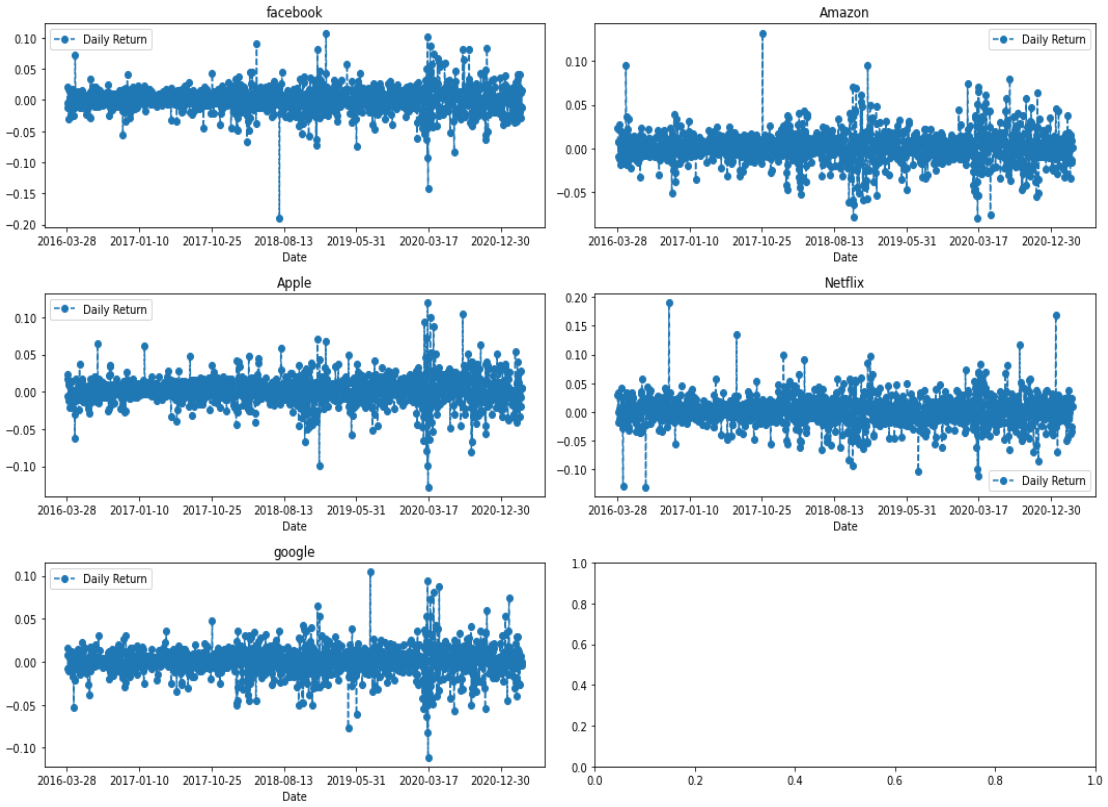




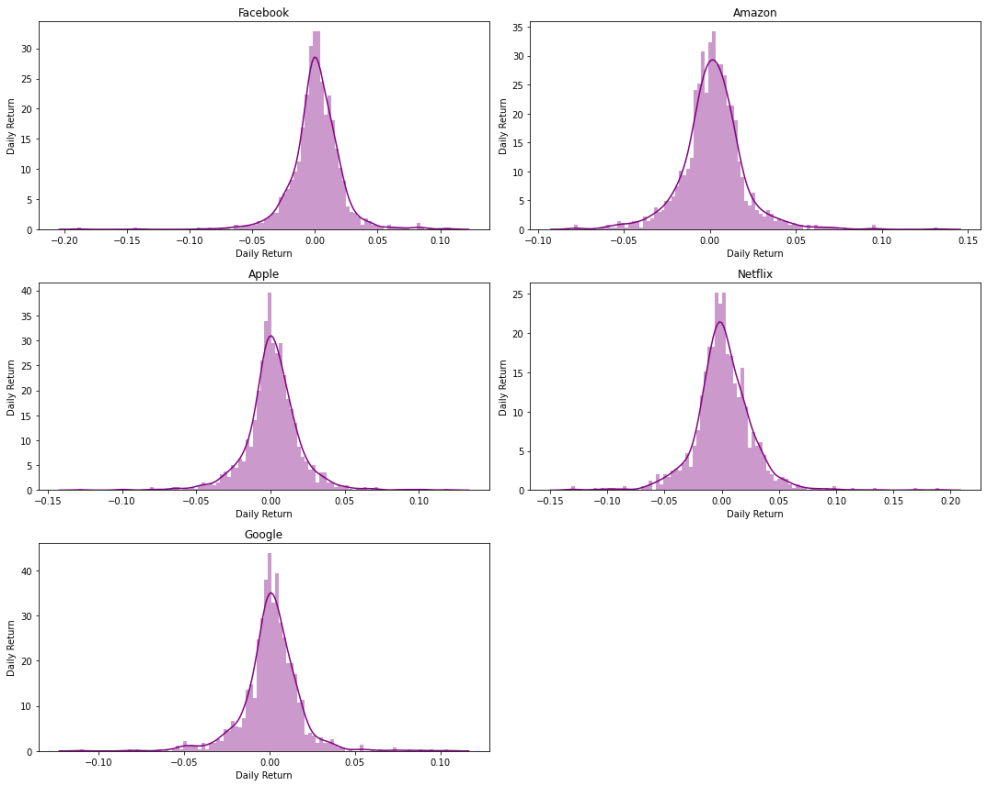


4. **<u>What was the daily return of the stock on average?</u>**

- Now that we've done some baseline analysis, let's go ahead and dive a little deeper. We're now going to analyze the risk of the stock. In order to do so, we'll need to take a closer look at the daily changes of the stock and not just its absolute value. I used pandas to retrieve the daily returns for the FAANG companies' stock. It actually shows the percentage change of the 'adj_close' values of the stocks which gives our daily returns.



- Then, I looked at the overall average daily return using a histogram. I used seaborn to create both a histogram and KDE plot on the same figure. As we can see from the plot below, the plots of the daily return follows a **normal** distribution.

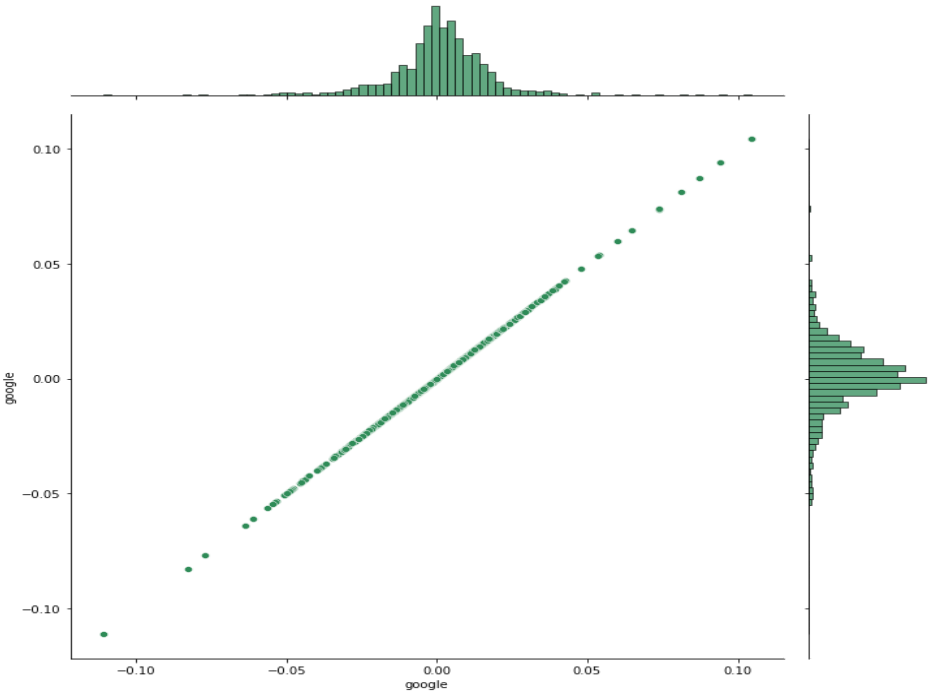


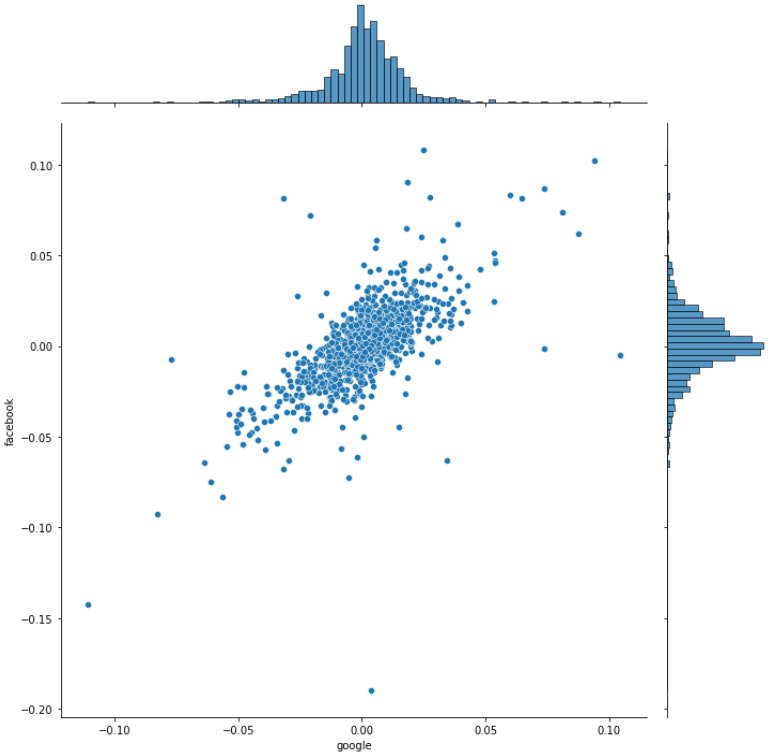
5. **<u>What was the correlation between different stocks closing prices?</u>**

- Now, I tried to compare and check if there is any relation between the stock closing prices of these companies. For example, the dataset for this operation was manipulated as below:

Date	facebook	amazon	apple	netflix	google
2016-03-28	113.690002	579.869995	26.297501	101.209999	733.530029
2016-03-29	116.139999	593.859985	26.920000	104.129997	744.770020
2016-03-30	114.699997	598.690002	27.389999	102.190002	750.530029
2016-03-31	114.099998	593.640015	27.247499	102.230003	744.950012
2016-04-01	116.059998	598.500000	27.497499	105.699997	749.909973

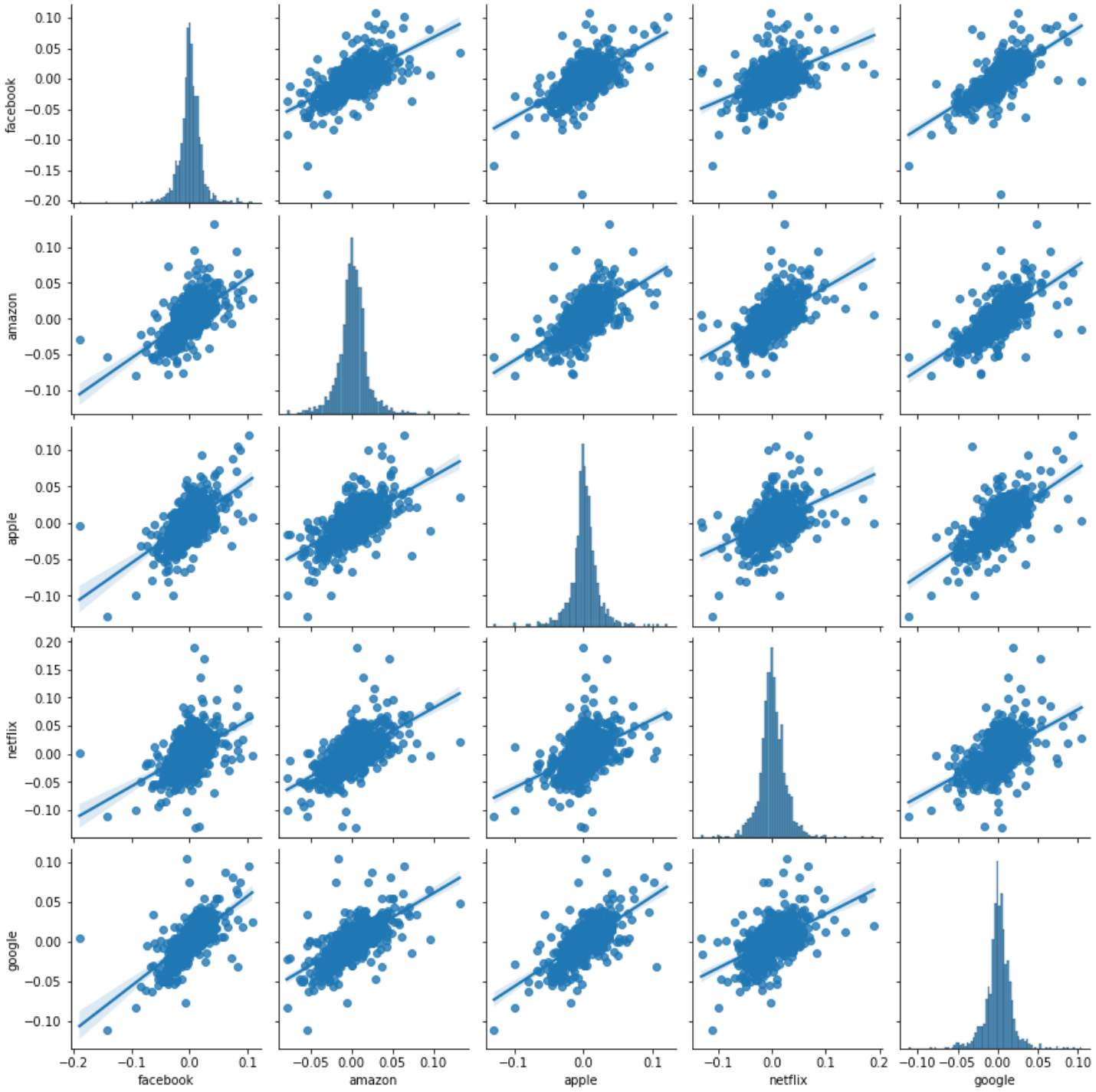
- Now, I plotted the correlation of Google with itself. It is supposed to be a linear relation because comparing with itself is a straight line in terms of correlation.



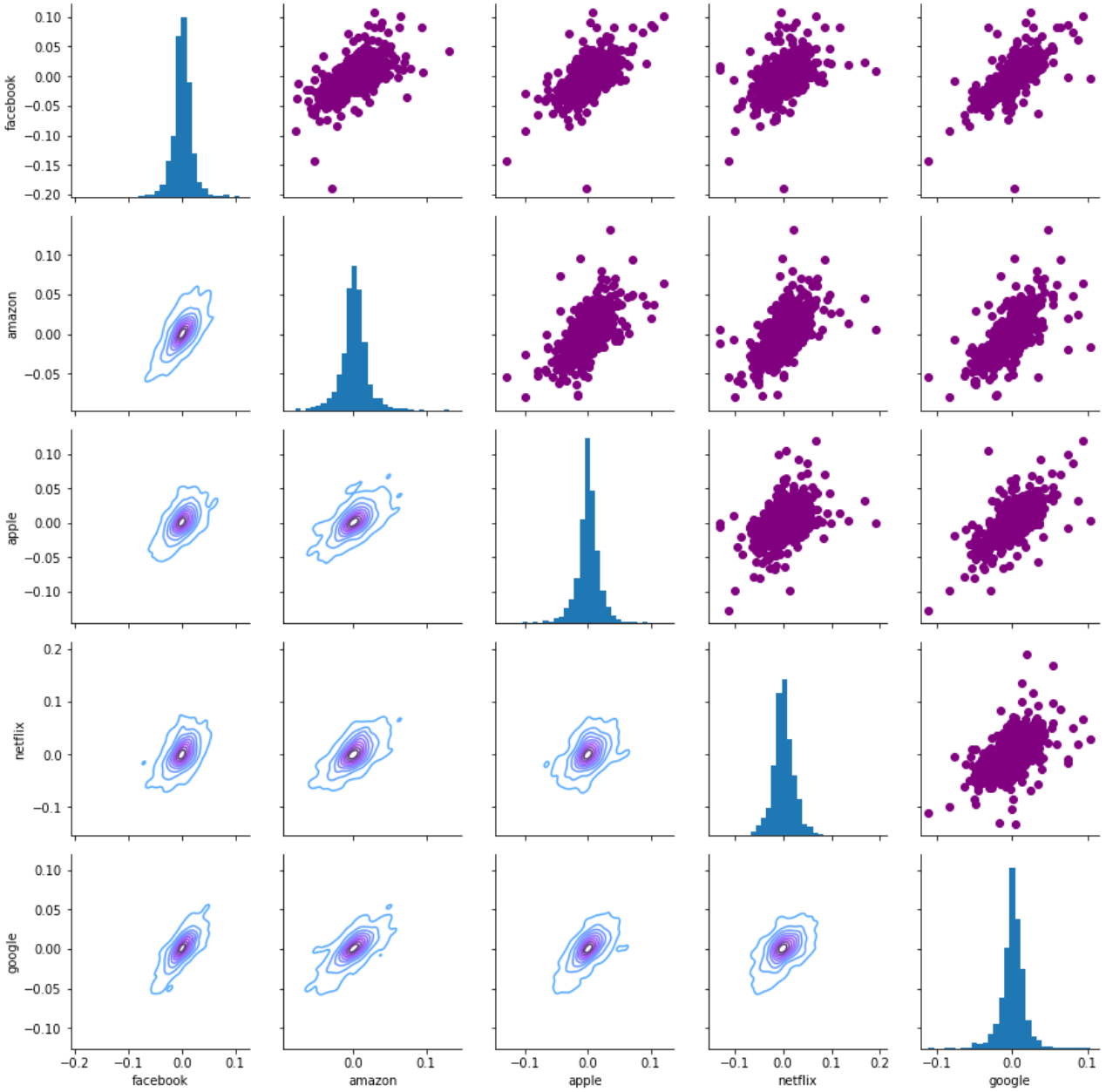


- Similarly, I plotted the correlation plot using seaborn library's jointplot method of Google and Facebook. As we can see, if two stocks are perfectly (and positively) correlated with each other a linear relationship between their daily return values should occur. The clustered data points are distributed as if they are near a straight line, making it more visible that they are highly correlated.

- Now I went through the documentation of seaborn library and found out some quick methods for auto-analysis. Here are the results of **seaborn.PairGrid** method:



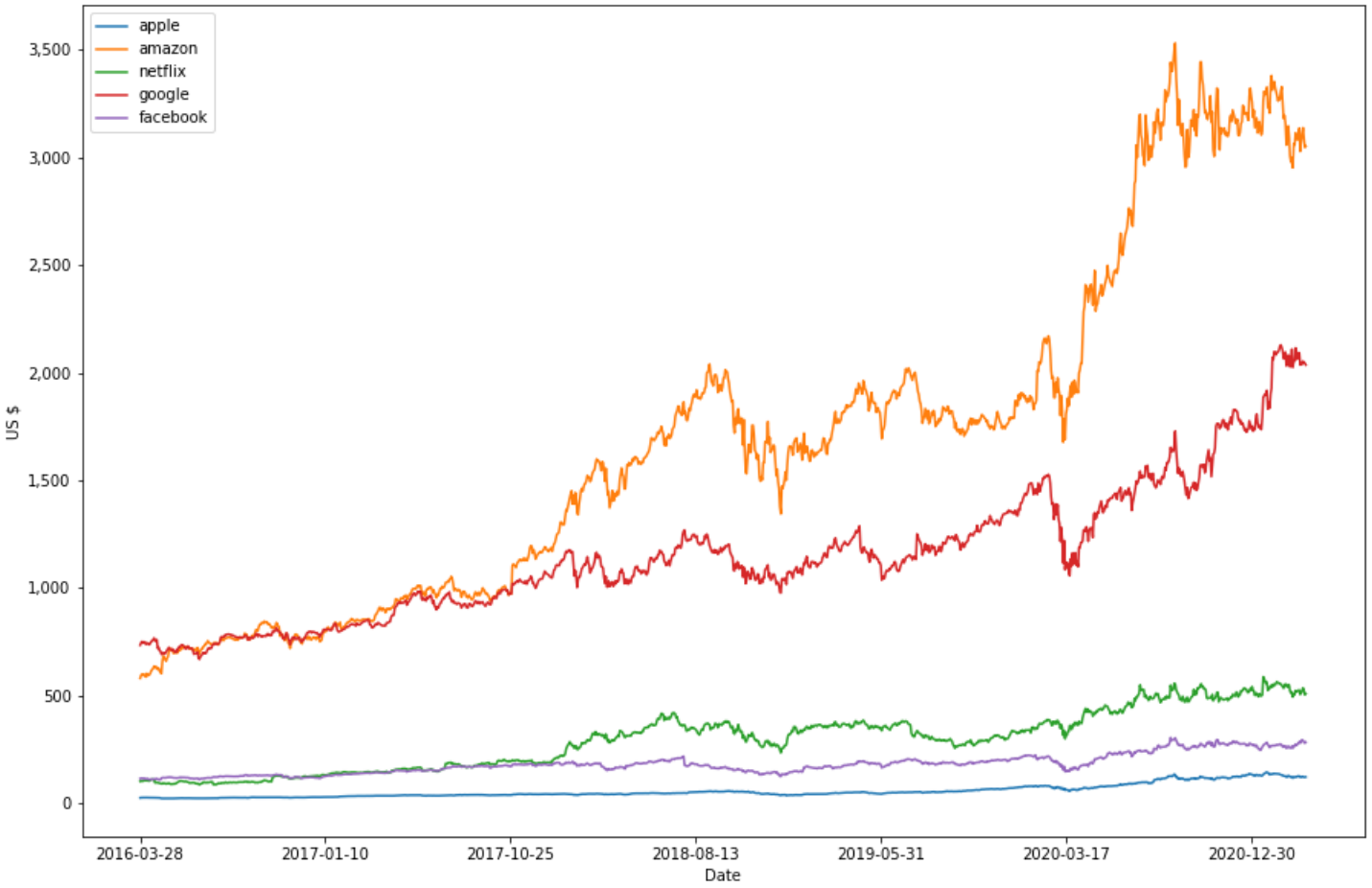
- Above we can see all the relationships on daily returns between all the stocks. A quick glance shows an interesting correlation between Google and Amazon daily returns. It might be interesting to investigate that individual comparison. While the simplicity of just calling seaborn.pairplot() is fantastic we can also use seaborn.PairGrid() for full control of the figure, including what kind of plots go in the diagonal, the upper triangle, and the lower triangle. Below is an example of utilizing the full power of seaborn to achieve this result.
- Below, I did some more analysis using this library method for better visualisation of the data.



- Now, below I tried to plot the heatmap of the daily returns of each company. We can see here numerically and visually that Google and Facebook had the strongest correlation of daily stock return. It is quite interesting to see this kind of correlation.

6. **Return on Investment for FAANG Companies:**

- I tried to estimate the return on investment for each of these FAANG companies and for that I chose the Adj_Close(Adjusted values factor in corporate actions such as dividends, stock splits, and new share issuance) values for each company and plotted them in a single graph. Needless to say, the **return on investment(ROI)** will also follow the trend of the following graph.



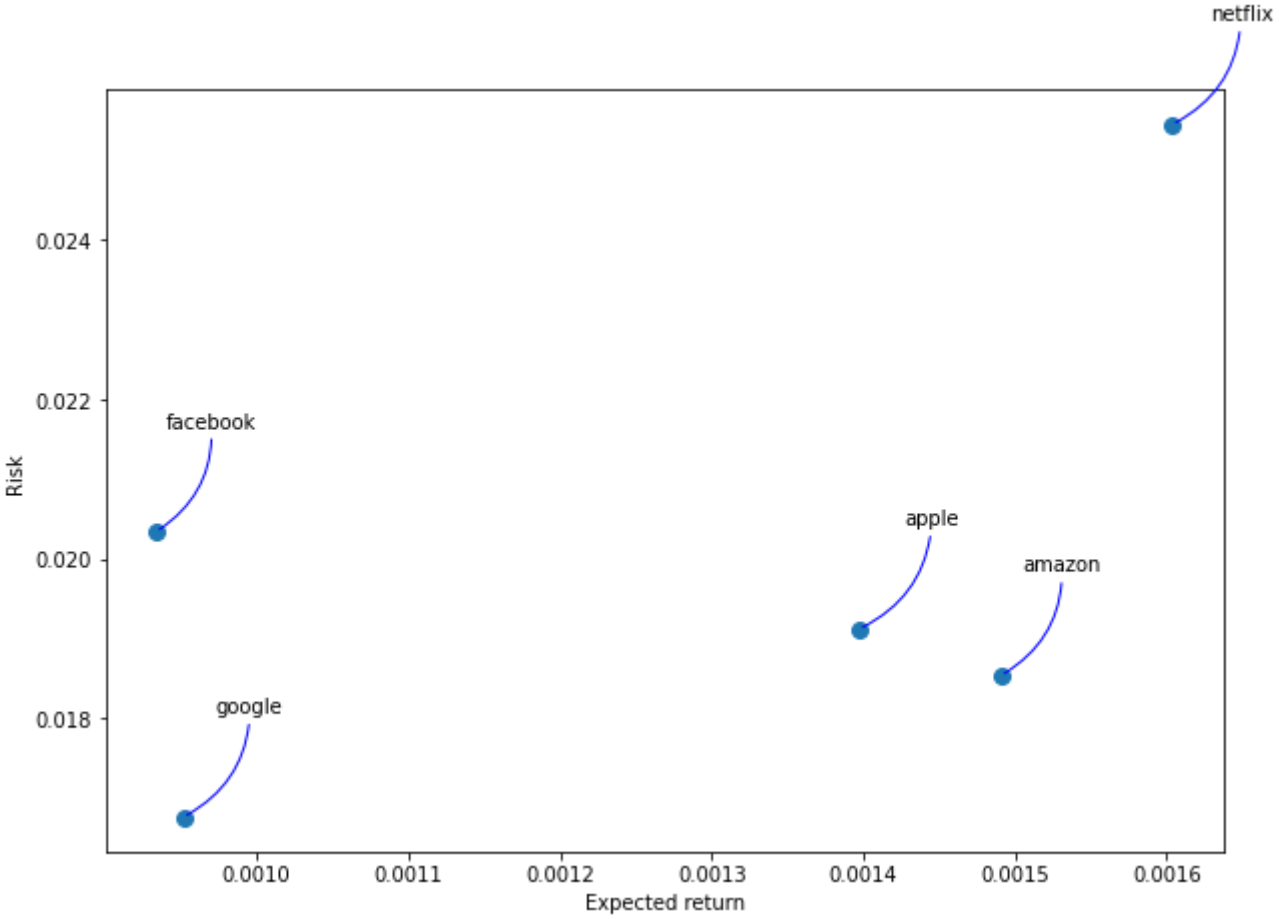
- As I mentioned earlier, the ROI(return on investment) value for each of the companies are as follows.

	Asset	Price point bought at	Stock acquired	Current price	Current value	ROI value	ROI percentage
1	amazon	579.87	0.34	3052.03	1037.69	837.69	418.85
2	netflix	101.21	1.98	508.05	1005.94	805.94	402.97
0	apple	24.46	8.18	121.21	991.50	791.50	395.75
3	google	733.53	0.27	2035.55	549.60	349.60	174.80
4	facebook	113.69	1.76	283.02	498.12	298.12	149.06

- We can say that it is very beneficial to invest in Amazon, Netflix and Apple as their return values are very high and one common thing among all the FAANG companies is that they all give us a very high amount of profit.

7. How much value do we put at risk by investing in a particular stock?

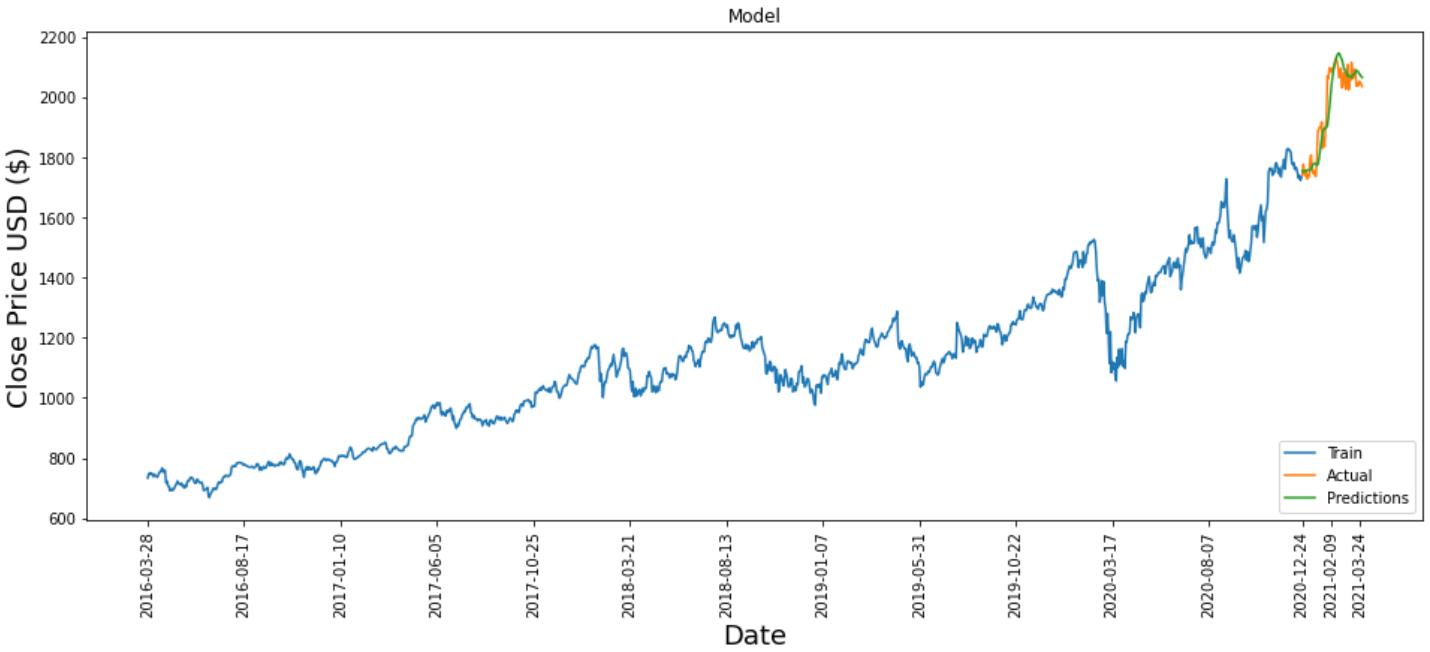
- Now, I analyzed the risk factors of investing in each of these companies by plotting the scatter plot using the mean and standard deviation of the returns of these companies.



- As we can see from the plot above, the risk factors are somewhat close to each other and for NetFlix, the risk factor is the highest. However, NetFlix is giving the highest expected return as well. Similarly, Google has the lowest risk factor but its expected return is also low. Similarly, we can see the other companies as well compared to the above.

8. Predicting the closing price stock price of GOOGLE:

- I used the AI algorithms related to Neural networks to train a model to predict the stock closing prices of Google for the last 3 months of the dataset.
- I used an LSTM model with 128, 64 and 25 hidden layers respectively and trained the model with all the dataset of google except the last 60 days period.
- The last 60 days i.e the data from December 2020 to March 2021 was used to predict the values of the stock closing prices of google.
- Below I have shown the results of the prediction with the actual values in the same graph.



- Here are some values from the predictions. As we can see the values are pretty similar and close to each other. The column 'Close' denotes the actual values of the stock closing prices and the 'Predictions' show the predicted values generated from the model.

- The average error obtained from this prediction was found to be = 2.0858351647561486%. This is pretty decent considering the size of our training set was in the order of thousands. We can conclude that our model or neural network has learned quite well to predict the future values.

Date	Close	Predictions
2020-12-24	1738.849976	1755.337646
2020-12-28	1776.089966	1750.640869
2020-12-29	1758.719971	1751.998169
2020-12-30	1739.520020	1754.799927
2020-12-31	1751.880005	1756.189697
...	...	...
2021-03-22	2038.589966	2083.253418
2021-03-23	2052.959961	2077.733398
2021-03-24	2045.060059	2073.399658
2021-03-25	2044.359985	2069.226318
2021-03-26	2035.550049	2065.537598

➔ **Conclusion:**

- As a conclusion of the above implementations, I would like to highlight that I have tried to touch upon every aspect of the goal that I tried to explore.
- We saw that it is highly beneficial to invest in the stocks of the FAANG companies as they provide a high amount of ROI (Return on Investment) even with a small amount of investment.
- I also analyzed the risk factors involved in investing in each of these companies and we saw that the risk factor coefficients were pretty close to each other where the risk increased with increase in average returns from a company.
- However, from the point of view of future predictions, we can see the graphs are never going downward. Although things went a bit rough during the pandemic, even after that, the graphs of stock prices have an upward inclination which proves it won't be that much of a risk if we are willing to invest in the stocks of these companies.