Stock Price Analysis

In this python notebook, I have analyzed the stock data which was retrieved from AWS Athena.

In this project, streaming data was collected through AWS Lamdda function and then that was stored in a AWS S3 bucket through Kinesis. Finally, from that S3 bucket, I queried few data in the AWS Athena by creating a Crawler in AWS Glue.

Data Analysis with Visualization

I have downloaded the result file from AWS Athena. I will be using the Google Colab to do some analysis.

At first, I am importing the necessary libraries for the analysis.

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

After importing, I am importing the csv file from my drive and converting that into a dataframe.

In [21]:
```
from google.colab import drive
drive.mount('/content/gdrive')
```

```
Mounted at /content/gdrive
```

In [23]:
```
import pandas as pd
df=pd.read_csv('results.csv')
```

In [25]:
```
df.head(10)
```

Out[25]:

	name	max_high	ts	hour
0	BYND	74.543999	2021-11-30 09:35:00-05:00	9
1	BYND	73.279999	2021-11-30 10:00:00-05:00	10
2	BYND	71.040001	2021-11-30 11:20:00-05:00	11
3	BYND	71.019997	2021-11-30 12:30:00-05:00	12
4	BYND	71.239998	2021-11-30 13:55:00-05:00	13
5	BYND	71.400002	2021-11-30 14:00:00-05:00	14
6	BYND	71.279999	2021-11-30 15:45:00-05:00	15
7	DDOG	186.289993	2021-11-30 09:35:00-05:00	9
8	DDOG	184.100006	2021-11-30 10:00:00-05:00	10
9	DDOG	178.539001	2021-11-30 11:10:00-05:00	11

Data Visualization

In this section, I will plot few basic graphs to visualize the data.

Barplot

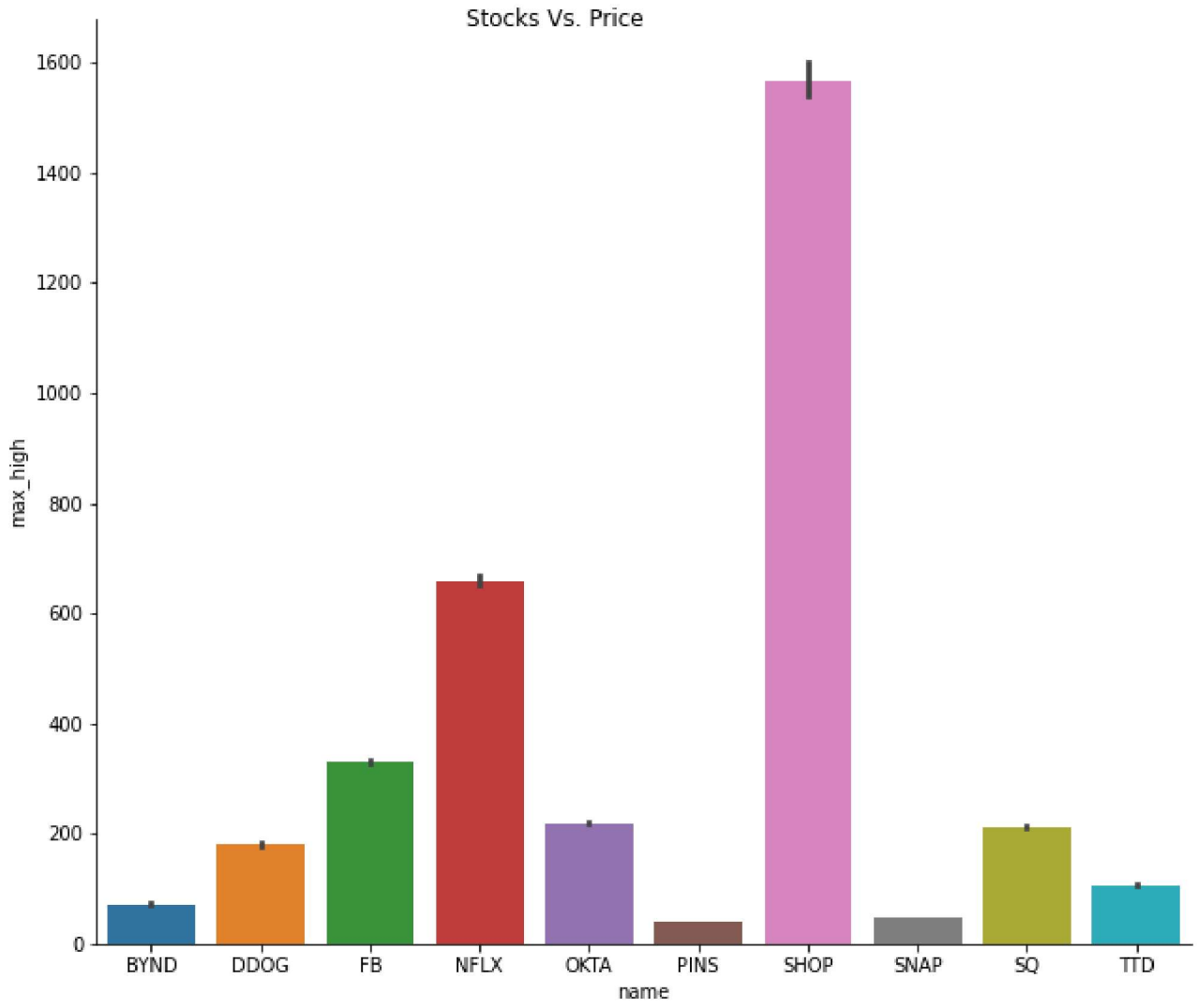
In the first plot, I am creating a bar chart of stocks versus its mean price.

In [39]:
```
ax = sns.catplot(x='name',y='max_high',data=df, kind='bar')

ax.fig.set_size_inches(10,8)

ax.fig.suptitle('Stocks Vs. Price', fontsize = 12)
```

Out[39]: Text(0.5, 0.98, 'Stocks Vs. Price')



So, according to the above barchart, shopify has the highest price while Pininterest has lowest.

The following two plots are line graphs for hourly time series plots for **Facebook** and **Netflix**. I have chosen these two because these two are very popularly used medias now a days.

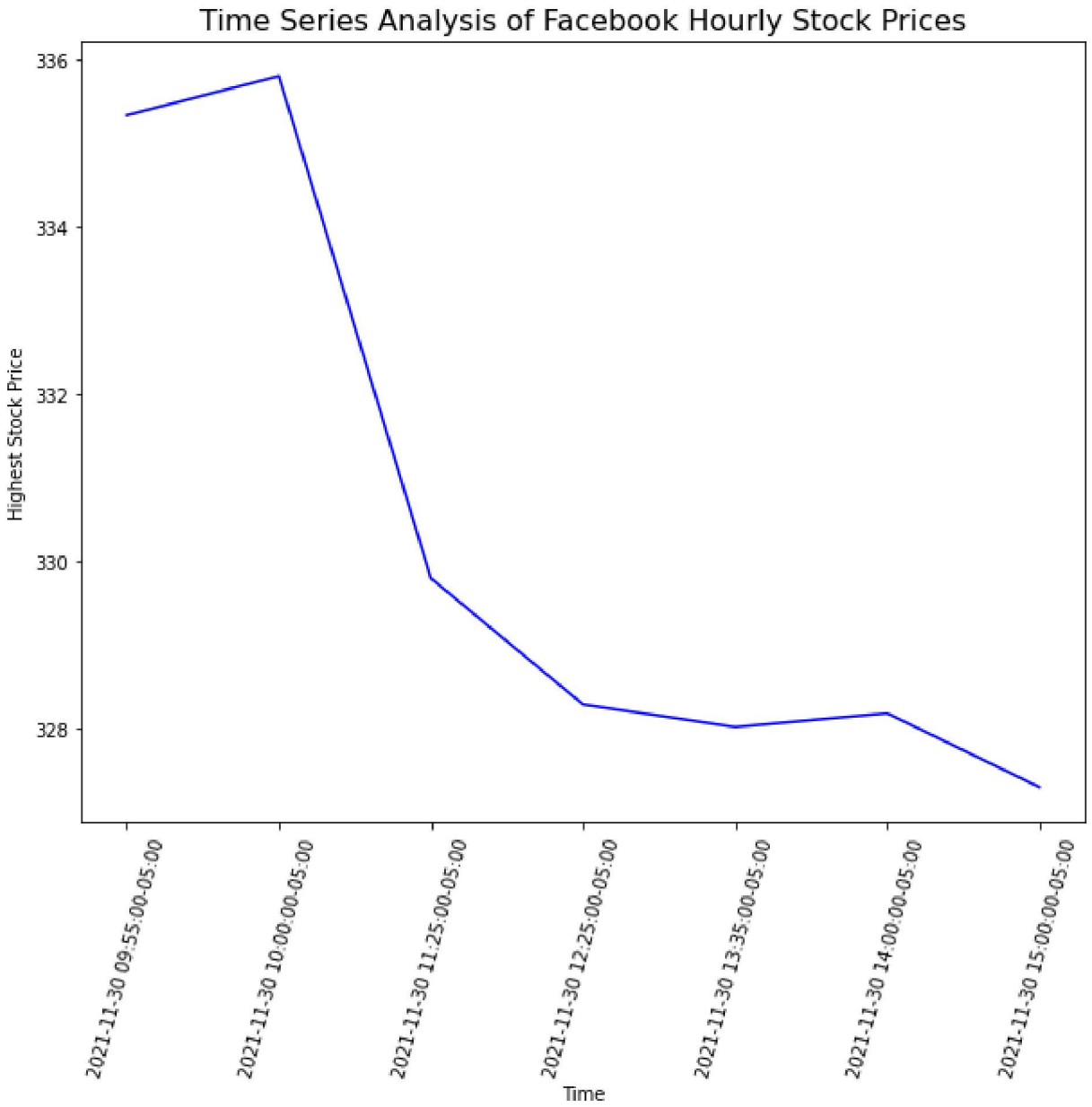
Time Series Analysis (Line Graphs)

Facebook

In [46]:

```
Facebook_data = df[df['name'] == 'FB']
fig, ax1 = plt.subplots(figsize = (10, 8))
sns.lineplot(x = 'ts', y = 'max_high', color = 'blue', data = Facebook_data, ax = ax1)
plt.xticks(rotation=75)
ax1.set_title('Time Series Analysis of Facebook Hourly Stock Prices', fontsize = 16)
ax1.set_xlabel('Time')
ax1.set_ylabel('Highest Stock Price')

plt.show()
```

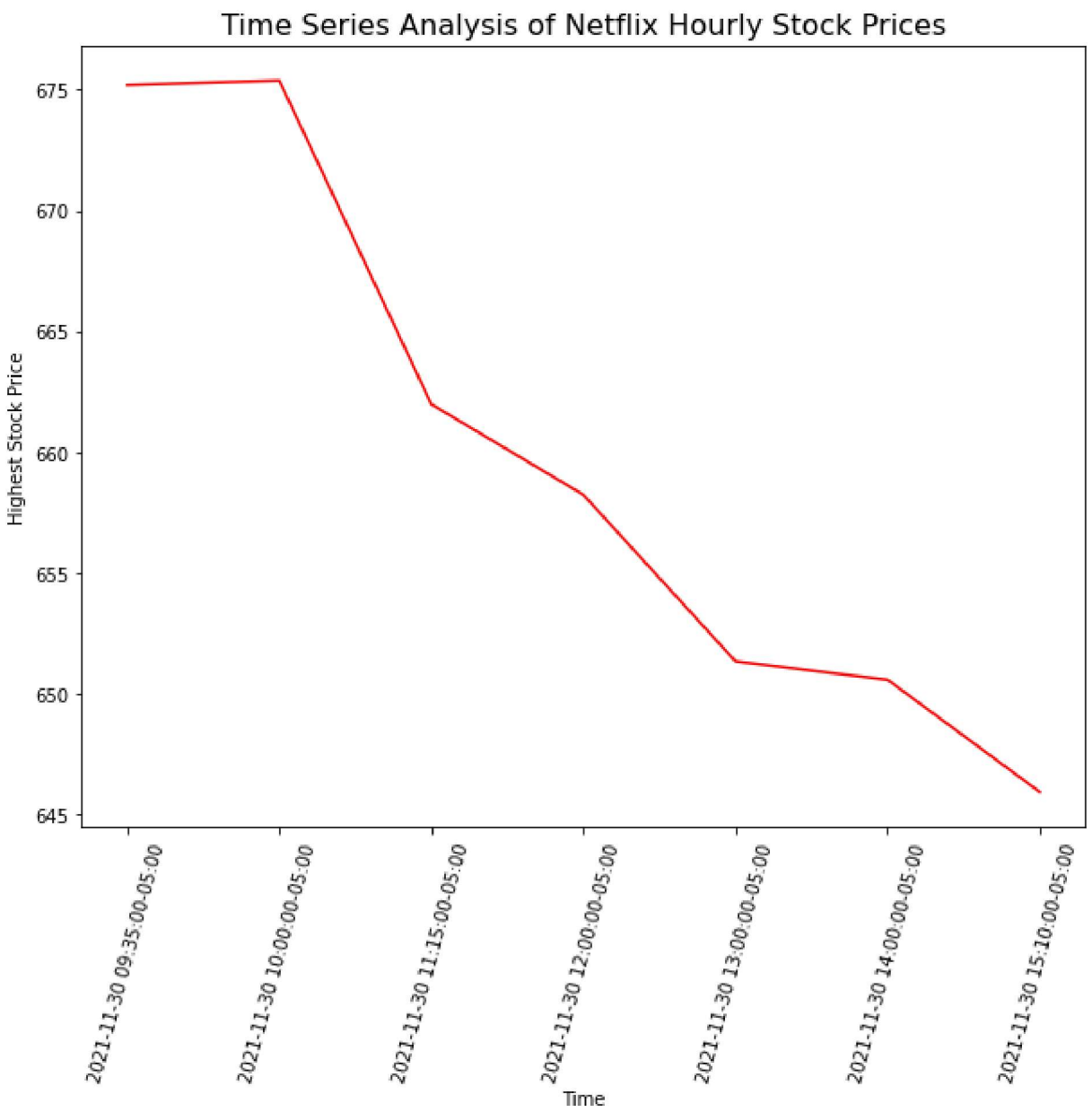

Time Series Analysis of Facebook Hourly Stock Prices
Highest Stock Price
336
334
332
330
328
2021-11-30 09:55:00-05:00
2021-11-30 10:00:00-05:00
2021-11-30 11:25:00-05:00
2021-11-30 12:25:00-05:00
2021-11-30 13:35:00-05:00
2021-11-30 14:00:00-05:00
2021-11-30 15:00:00-05:00
Time

Netflix

In [49]:

```
Netflix_data = df[df['name'] == 'NFLX']
fig, ax2 = plt.subplots(figsize = (10, 8))
sns.lineplot(x = 'ts', y = 'max_high', color = 'red', data = Netflix_data, ax = ax2)
plt.xticks(rotation=75)
ax2.set_title('Time Series Analysis of Netflix Hourly Stock Prices', fontsize = 16)
ax2.set_xlabel('Time')
ax2.set_ylabel('Highest Stock Price')

plt.show()
```



So, both of the time series plot have a similar trend. Since these are very prominent companies, the stock prices are very high for these companies. However, interestingly, both of these companies stock price had fallen drastically at the same time (between **9.30 am to 10.30 am**), and then it started to went up.

This gives us an indication that there might have been an industry wide issue at that specific time so both prominent companies stock had fallen.

This was a very minimal graphical analysis of the stock prices which I extracted from AWS S3 bucket with by quering in AWS Athena.