Stock Market Analysis

Group : No Name
By - Kathan Shah, Nishit Salot, Shubhangkar Girish Jain

Overview

Stock Markets have been one of the most lucrative ways to make money since the beginning of time. Predicting and Analysing Stock Prices has been one of the biggest problems in the stock markets. To minimize speculations and increase the accuracy of data this project will focus on Analyzing and Predicting Stock Market Market movements and volatility.

The Data

Stock Market Price Data - India

Source : Kaggle

2017-2021

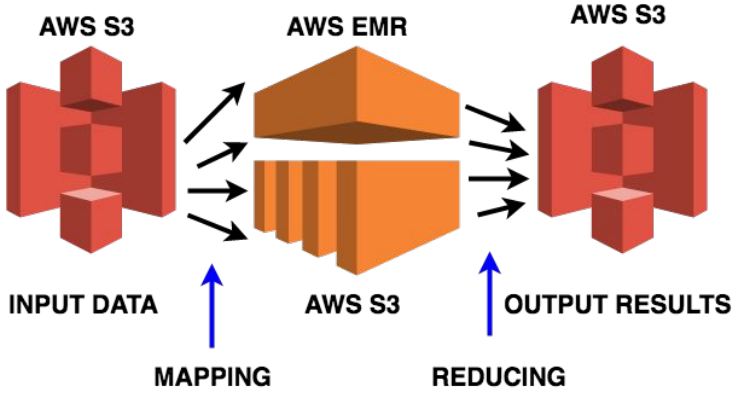
There are a total of 160 stocks with their characteristics for every minute on any trading day within the given time period

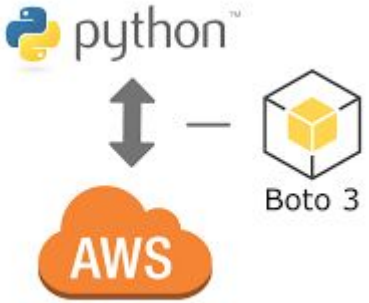
Business Questions

1. What is the price of the stock at the end of the fiscal year ?
2. What trend is the stock following ?
3. Which Stock has the highest returns ?
4. List the days in which the stock experiences maximum volatility.

Our Approach

1. Insert data in AWS S3 Bucket
2. Use EMR a to extract data from S3 and query data using SparkSQL to find few insights into our business questions.
3. Use BOTO3 to use S3 data locally in python jupyter notebook.
4. Running ARIMA Model (Predictive Analysis) in jupyter notebook to predict Stock trends and answer other business questions.






Descriptive Analysis

Q1. How will you resample the data on Daily Basis for a readable view ?

Date	OPEN	LOW	HIGH	CLOSE	VOLUME
2017-01-02	340.0	337.3	348.95	346.0	8372.0
2017-01-03	349.5	341.8	355.0	350.0	20826.0
2017-01-04	353.3	341.5	355.65	350.0	131883.0
2017-01-05	352.35	349.55	354.95	351.5	8818.0
2017-01-06	352.0	349.0	354.9	349.1	8128.0
2017-01-09	351.75	348.0	352.0	348.5	6310.0
2017-01-10	350.0	348.0	352.25	350.5	10186.0
2017-01-11	350.5	346.75	351.35	351.25	17155.0
2017-01-12	350.5	349.05	351.95	350.0	8469.0
2017-01-13	348.1	348.1	354.0	351.95	12967.0
2017-01-16	351.9	350.0	355.0	353.95	8673.0
2017-01-17	353.55	352.1	354.5	354.0	7386.0
2017-01-18	355.0	352.5	360.75	356.5	27663.0
2017-01-19	355.2	350.05	359.7	352.5	16641.0
2017-01-20	355.0	342.5	355.0	343.25	28588.0
2017-01-23	340.6	336.55	353.35	353.0	15594.0
2017-01-24	350.0	346.5	353.85	346.5	20011.0
2017-01-25	347.1	345.6	354.25	351.5	62783.0
2017-01-27	352.5	352.45	362.2	360.5	45212.0
2017-01-30	361.5	358.6	364.9	359.5	43072.0
2017-01-31	357.55	356.5	361.85	360.0	26472.0

In [131]:
```
#Resampling on Daily Basis
df.createOrReplaceTempView('data');
query2 = spark.sql("SELECT DATE(timestamp) as Date,FIRST(open) as OPEN,MIN(low) as LOW,MAX(high) as HIGH,LAST(close) as
query2.show(50)
```

Q2. What is the percentage return on a particular security?

In [160]:
```
#Returns on a Stock
df.createOrReplaceTempView('data');
query3 = spark.sql("SELECT LAST(close)/FIRST(close)* 100 as Percentage_Return from data")
query3.show()
```

```
VBox()

FloatProgress(value=0.0, bar_style='info', description='Progress:', layout=Layout(height='25px', width='50%'),…

+-----------------+
|Percentage_Return|
+-----------------+
|370.5882352941177|
+-----------------+
```

Q3. What is the maximum volatility of a particular stock over the entire time frame ?

In [174]:
```
#Amount Traded
df.createOrReplaceTempView('data');
query4 = spark.sql("SELECT DATE(timestamp) ,ROUND(LAST(close)-FIRST(open),2) as volatility from data group by DATE(time
query4.show()
```

timestamp	volatility
2020-04-27	95.0
2020-03-13	88.8
2020-01-29	88.0
2018-01-05	85.0
2020-04-28	68.0
2020-08-20	57.1
2017-03-02	55.35
2020-11-05	54.0
2020-11-27	51.35
2019-10-18	48.5
2019-05-30	45.0
2020-03-31	44.0
2018-10-03	43.0
2020-02-20	42.15
2019-09-20	41.2
2020-12-07	41.15
2020-08-19	40.95
2018-08-30	39.75
2020-01-28	39.15
2019-02-04	38.5

only showing top 20 rows

Q4. What is the price of this stock at the end of each fiscal year ?

Year	Value
2017	155.96093333333346
2018	94.97439999999999
2019	98.12440000000004

Q5. What is the most recent trend that the stock is following ?

```
ITC = df[80]
if ITC[-7:-6]['open'].values < ITC[-1:]['open'].values:
    print("This stock has an Upward Trend")
else:
    print("This stock has an Downward trend")
```

```
This stock has an Upward Trend
```

Q6. Which stock has the highest returns ?

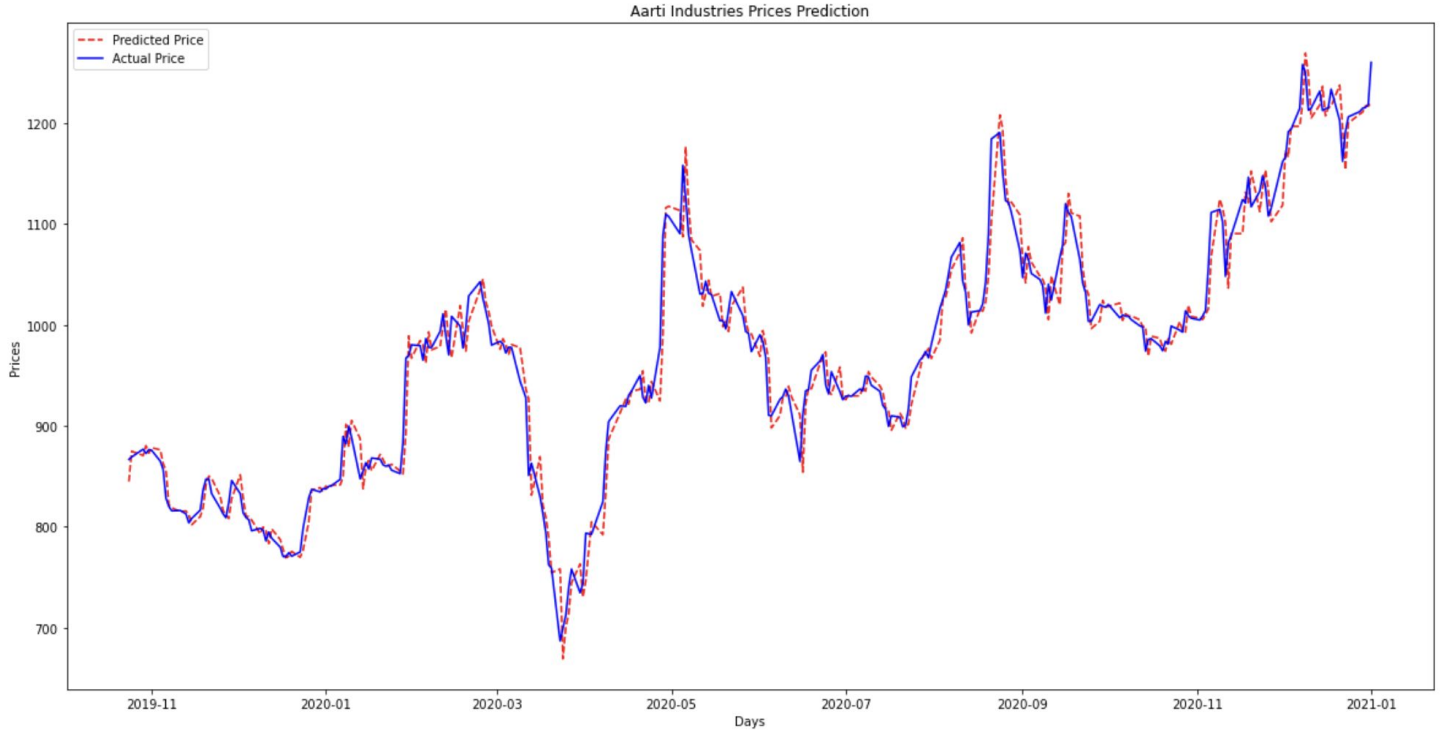
```
maximum = 0
index = 0
for i in range(0,159):
    temp=0
    temp = df[i][-1:]['cum_return'].values
    if temp>maximum:
        maximum = temp
        index=i
```

```
The stock Dhani gives the highest return of 1565.9289754748181 %
```


Predictive Analysis

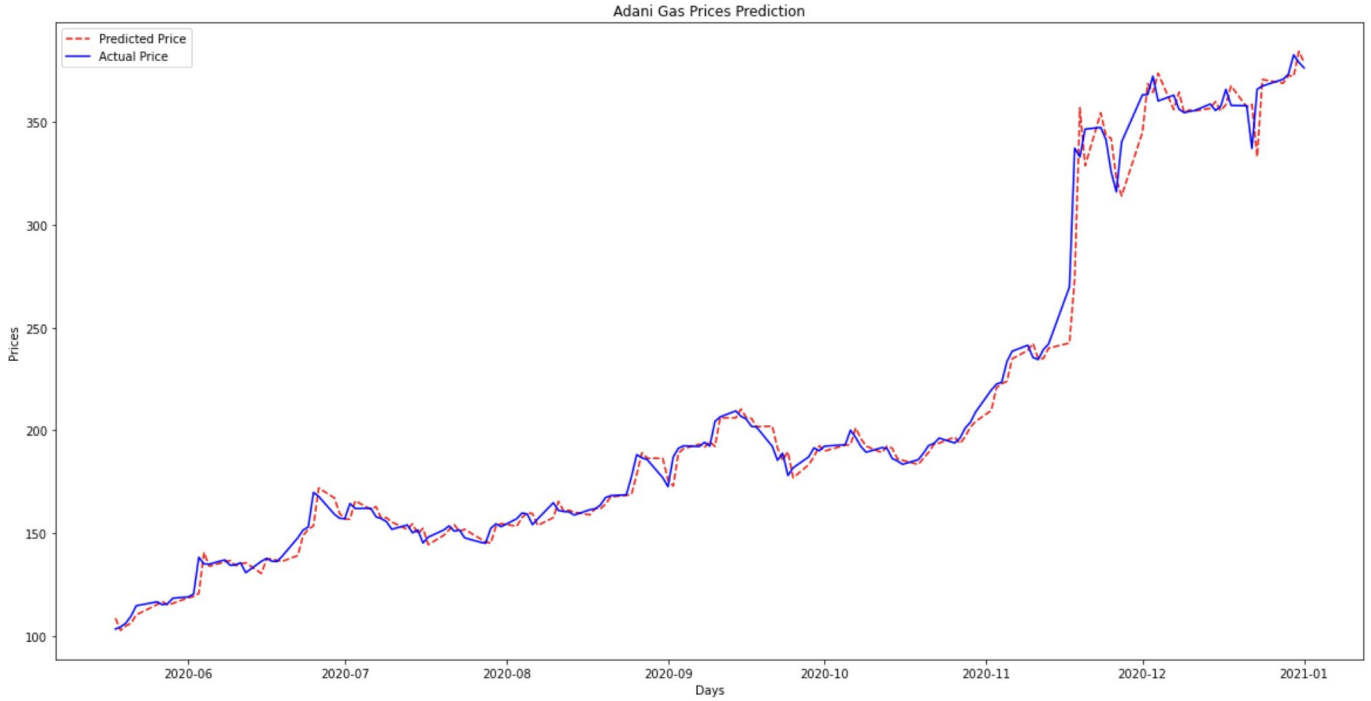
Trend analysis

```
Testing Mean Squared Error is 506.4753180312519
```

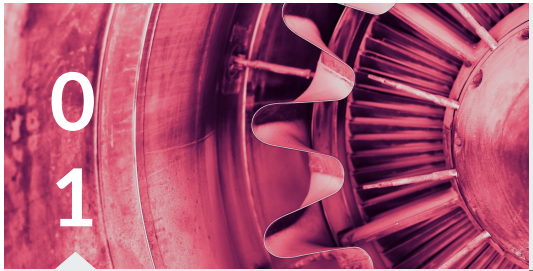


Trend analysis

```
Testing Mean Squared Error is 76.29030694398338
```



Further Analysis Ideas



Suggesting tactical moves to maximize gains

Considering External Factors into research and prediction.

Implement SVM and outlier detection to improve model performance.

Conclusions

1.

Prediction models as a sole measure should not be used to trade the stock markets with real money.

2.

Analysing previous prices is a good way to get rough idea of future price movements.

3.

Stock market price movements are controlled a lot more by external factors than they are by previous numbers and patterns.

Thank you.