Title: TESLA STOCK PREDICTION

1. Overview: Brief Description of the project

In our project we have taken the dataset related to Stock market of Tesla. After considering the suitable dataset then we need to do the data cleaning. If the considered dataset is already cleaned without outliers then we can directly do the exploratory analysis, there by building our regression model and predict the stock price. Here we are considering the aspects open, high, low and close. The final expected resultant would be in the linear regression format. The final resulted graph would be sketched using actual and predicted values. During this process we will be encountered with errors: mean absolute error and mean squared error.

1. Tools: You are using tools for this project.

* Programming language: Python.
* Software: Jupyter Notebook.
* Dataset: Tesla Stock Analysis which was considered from kaggle.
* OS: Windows 7 or above.
* Processor: Intel i3 or above.
* RAM: 512GB or above.

1. Dataset Description and Descriptive/Exploratory analysis(dependencies):

**Link of the dataset you are using :** [Tesla Stock Price | Kaggle](https://www.kaggle.com/datasets/rpaguirre/tesla-stock-price)

**More details about the data sets (number and list of variables, numerical, categorical, number of samples:-**

The dataset contains the stock prices of Tesla stock from 2016 to 2021. We have the following attributes:

Date – Date of the stock taken

Open – Opening price of stock on a given date

High – highest price of stock on a given date

Low – lowest price of stock on a given date

Close – closing price of stock on a given date

Adj close – Closing price after adjustments for all splits and dividends

Volume – Number of shares on a given date

Among all the attributes, date is in the form of string.

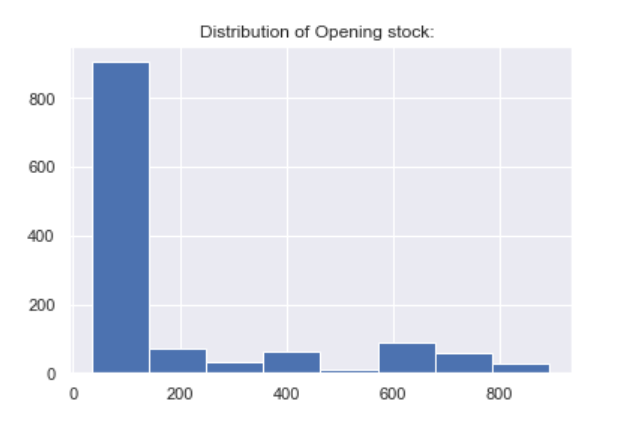
Open, close, high, low, adj close, volume are numerical values.

We have data from 08/16/2016 to 08/13/2021.

**Show data distribution:**

We have shown the distribution of stock with the help of histogram.

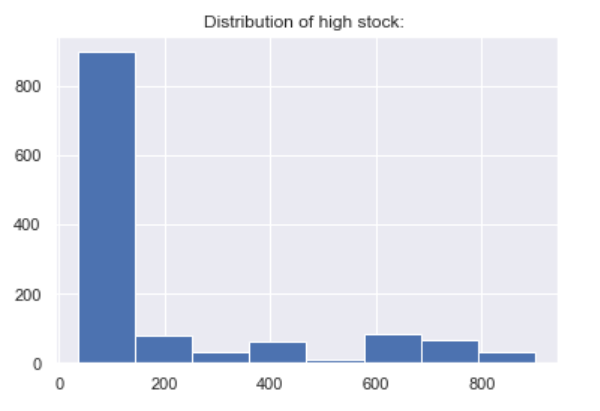
1. Distribution of Opening stock:



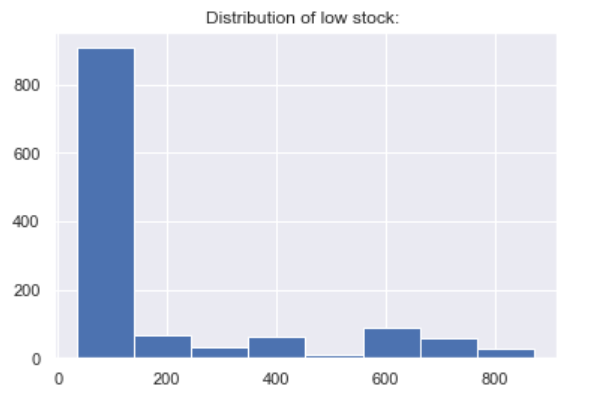
1. Distribution of Closing Stock:



1. Distribution of High Stock:



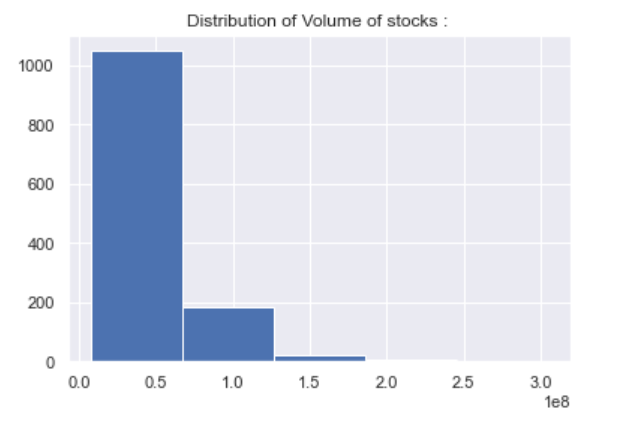
1. Distribution of Low Stock:



1. Distribution of Adjused Closing Stock:



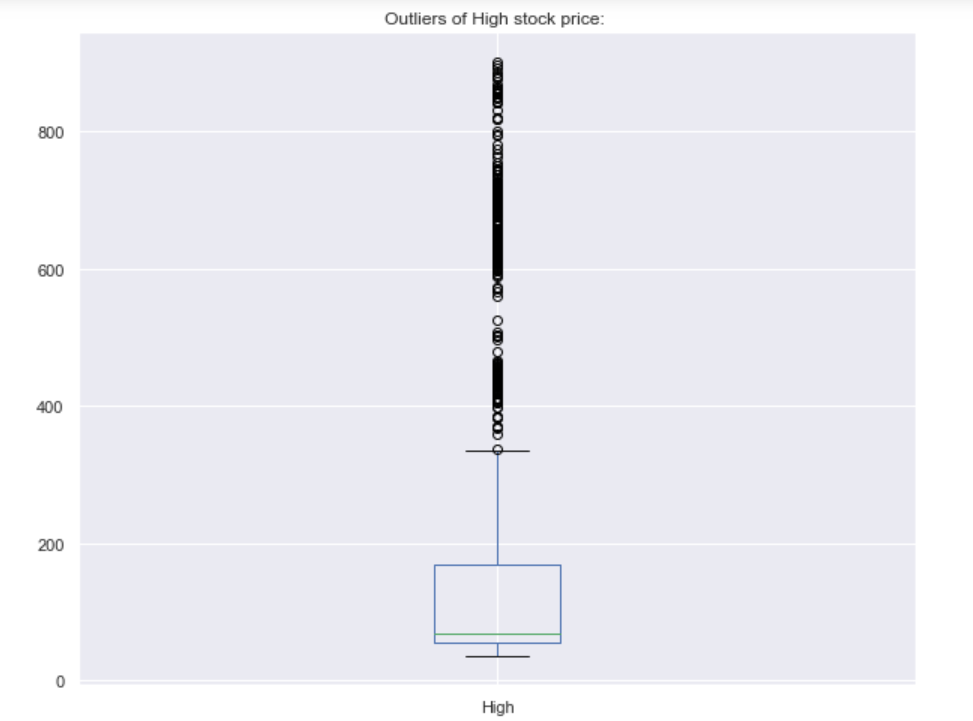
1. Distribution of Volume of stock:



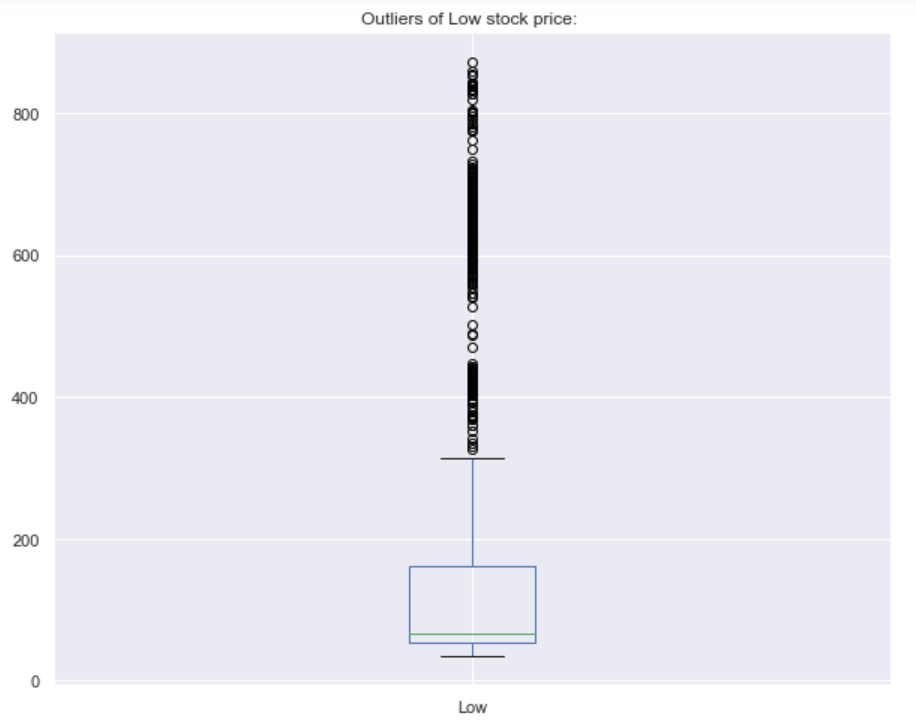
We observe that all the distributions of the attributes are right skewed because they have a tail on the right side of the distribution. This is also known as positively skewed. This means there are very few values in the dataset of each attribute which have high values and larger occurrences of lower values. With the help of this distribution we can understand that there is a possibility of data being small during a particular year and an unexpected hike in price has occurred which is giving a big variation in distribution. We can clearly see them in further process of our analysis.

**Detect outliers**

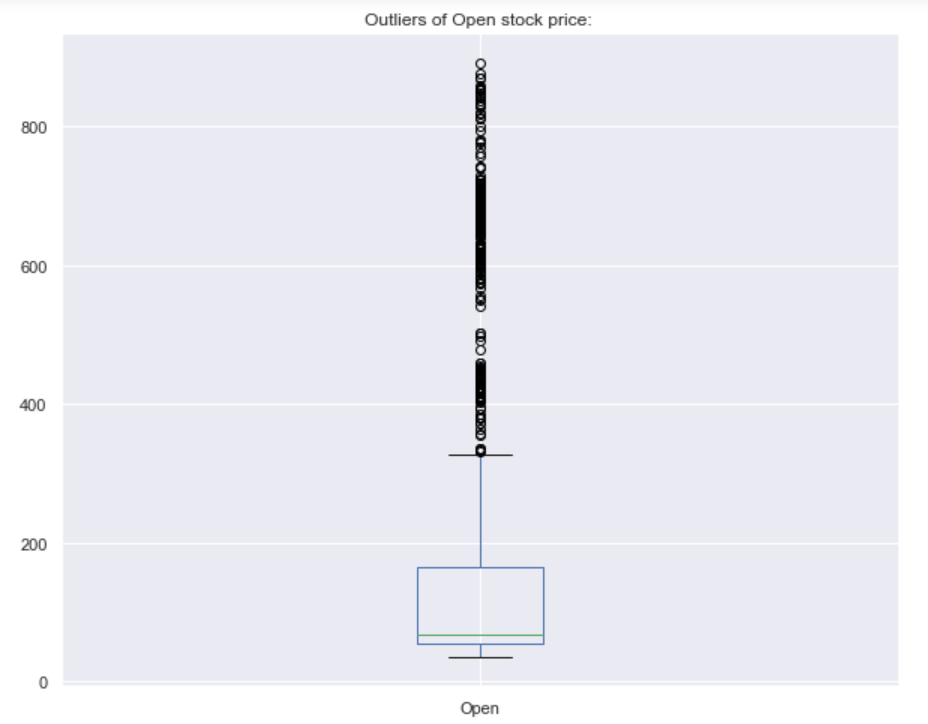
1. Outliers of high stock:

****

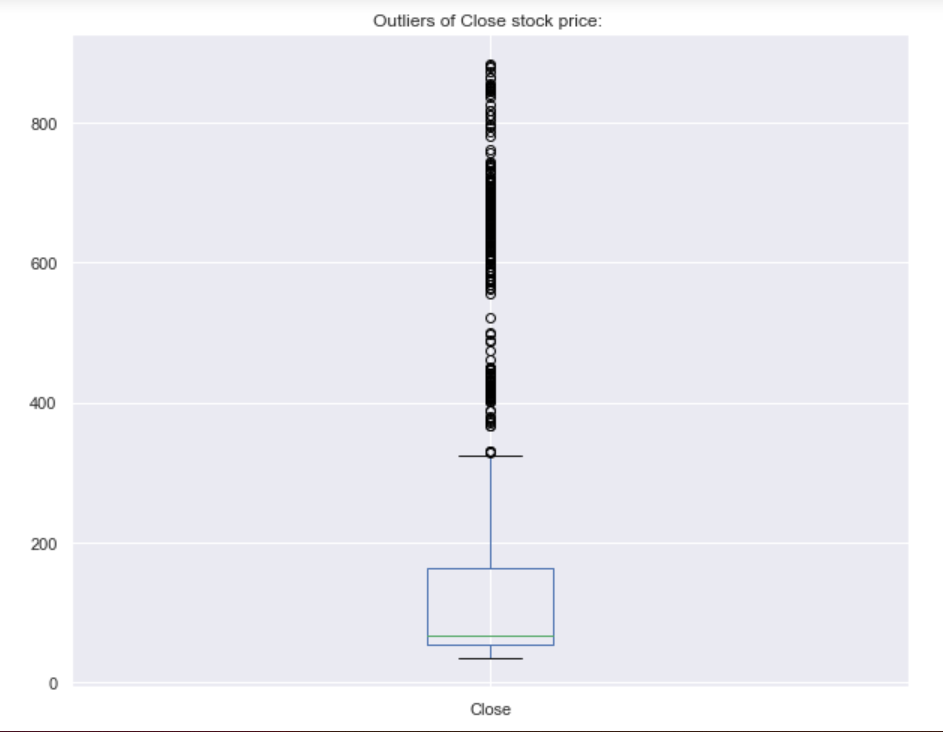
1. Outliers of Low stock:



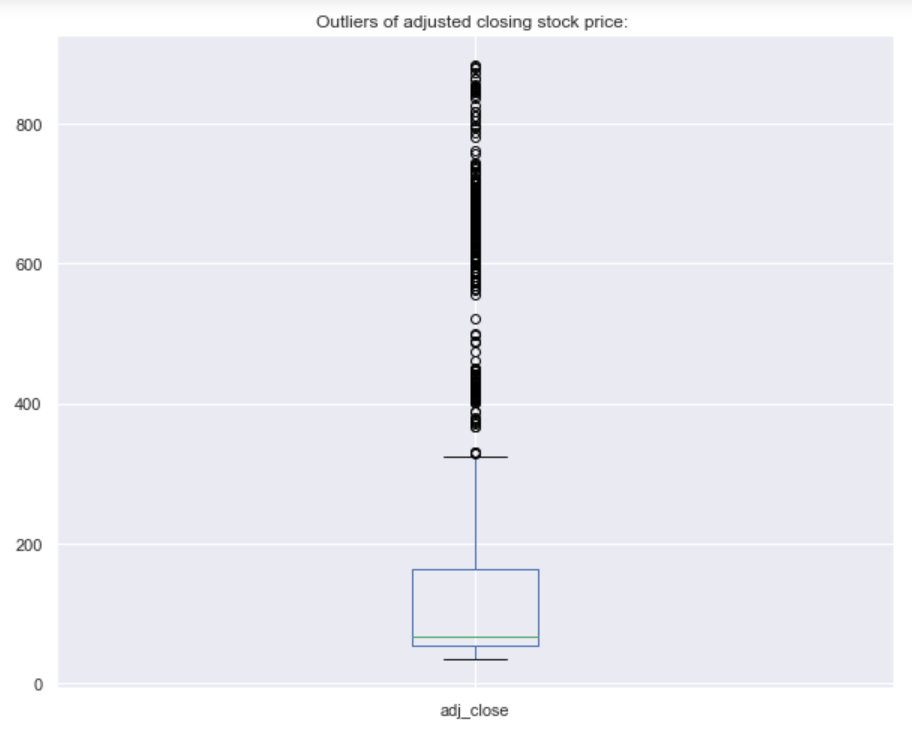
1. Outliers of Open stock:



1. Outliers of Close stock:



1. Outliers of Adjusted closing stock:



Outliers are those values which are different, not normal and peculiar values when compared to the values in the dataset. These outliers affect the entire distribution of the data with their abnormality. Box plot is a graphical way to represent outliers which can be depicted by quartiles.

  7. **Data cleaning/transformation:-**

Here the data cleaning is used to removes data. Which not related to the dataset. This processing is known as data cleaning. The data transformation is used to converting data from one format to another format or one structure to another structure is known as data transformation.

In our dataset, there are no null values of NA values. Therefore, we consider the data is already clean.

We have converted the data type of date to date-time format because we would use this format to extract month and year in our code for further analysis.

After viewing the distribution of data attributes, we have found that there are outliers in our attributes. We have tried to detect the outliers using IQR and quantiles and removed the outliers.

  8. **Statistical Analysis:**

1. What is the change in high and low stock price across months ?
2. What is the change in volume of stocks across 5 year-span?
3. What is the average closing price of the stock across all the years?
4. What is the right time to invest in Tesla to earn maximum profit?
5. How much deviation is there in stock from open and close value?
6. How much deviation in stock price from beginning date to end date from the dataset?
7. Show every year stock increase
8. How much stocks got affected due to covid-19?
9. What is the difference between pre-covid 1 year stock and post-covid 1 year stock?

9. **Plans for final outcomes:**

* We would build a model with the data we have using linear regression.
* We would also try to show pictorial representation of the correlations between various attributes considered.
* We would also show the variation in actual and predicted values with the regression model.

10. **Conclusion and possible progress direction and barriers**

By next phase we would answer the statistical analysis questions we have raised. We would use the variables used in our class for our analysis.

11. **Reference: Link any tutorial, code, blogs you are using**

<https://www.kaggle.com/code/davidolamidebalogun/tesla-stock-prediction-using-linearreg/notebook>

<https://www.kaggle.com/code/ysthehurricane/tesla-stock-price-prediction-using-gru-tutorial/data>

<https://www.kaggle.com/code/ysthehurricane/advanced-stock-pred-using-svr-rfr-knn-lstm-gru>

<https://www.kaggle.com/code/anandhuh/extracting-visualizing-stock-data-2022>