

# TSLA Stock Price Prediction.

Ramachandra Srikanth Naradasu

## Executive Summary

Stock analysis is a method for investors and traders to make buying and selling decisions. By studying and evaluating past and current data, investors and traders attempt to gain an edge in the markets by making informed decisions. There are two basic types of stock analysis: fundamental analysis and technical analysis. In this paper, The main focus is on TSLA Stock Price value. TSLA Stock price has been on the rise recently, with a +100% spike in the last few months. The year wise trend of stock price gives a basic view of stock value over years. Since this is a time series analysis, using ARIMA to generate the model and forecast the stock price value. This Dataset Contains 2518 Observations or (trading days) and it has 7 variables. I'm going to perform Multiple linear regression model to estimate the TSLA stock close price and perform ARIMA model on this time series data to forecast and predict the Future TSLA stock price value. From the interpretation of the model it can be concluded that there are several other factors other than the variables in the dataset for the increase in the price of the stock value. The goal of this project is to estimate the relationship between the close variable and other variables and predict the close price value of TSLA stock by forecasting the generated ARIMA model.

## **A. Introduction**

Tesla Inc. designs, develops, manufactures, and sells electric vehicles and stationary energy storage products. Tesla shares are now up roughly 260% year to date. The Tesla stock price values are collected from Yahoo Finance.

This data can help predict the relationship between variables and generate the time series plot. The trend of the variables across the years are analyzed and informed decisions are made in future based on the historical data.

This data helps in estimating the growth of the company and necessary changes to be made for the better value. Here, The main focus is on the Close price of the TSLA Stocks because the close price is used in predicting the future price.

## **B. Data Description**

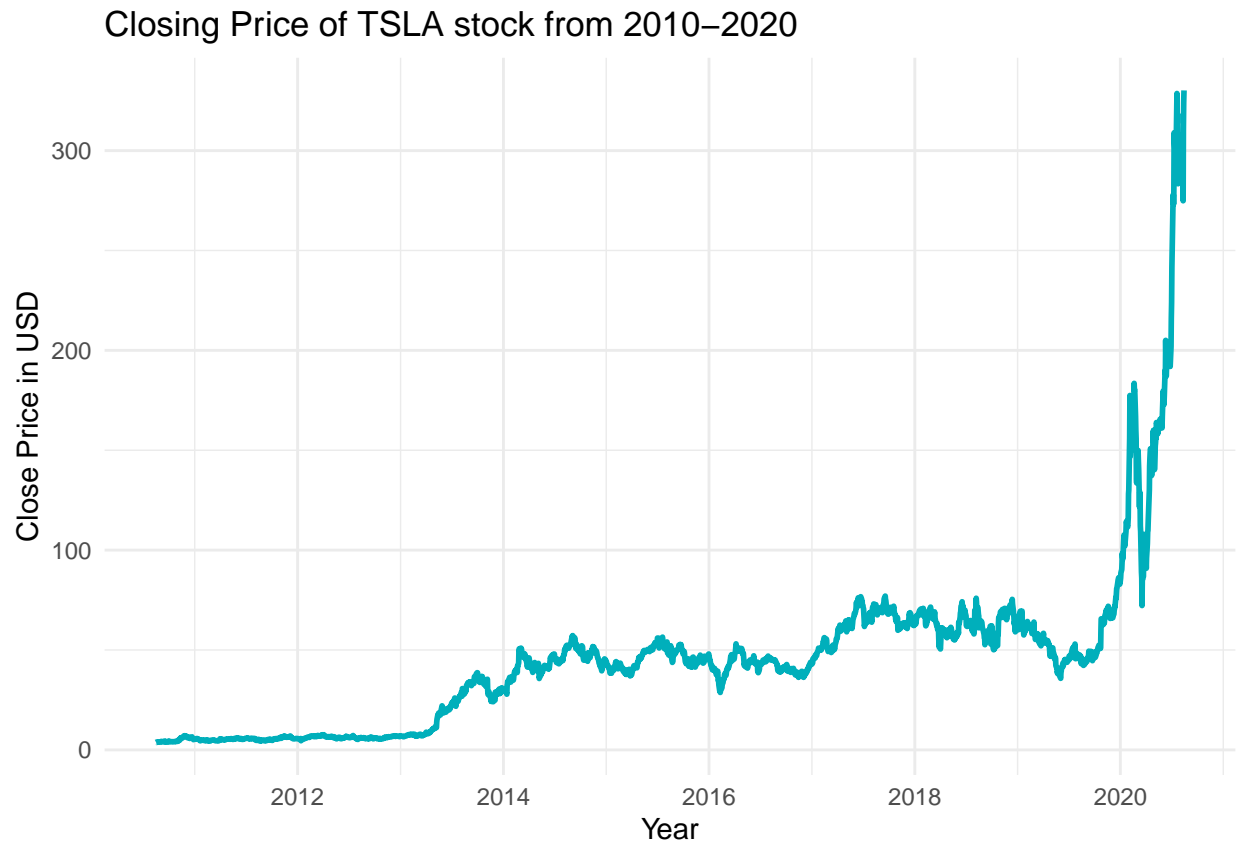
The TSLA dataset contains 2518 observations or (Trading days) between August 16,2010 and August 14,2020. This is a time series data with irregular interval of date. This dataset contains the information about the Open, High, Low, Close, Adj.Close, Volume. All variables except Date are numeric type. The max, min and median values of each variable are calculated.

## **C. PreProcessing Data**

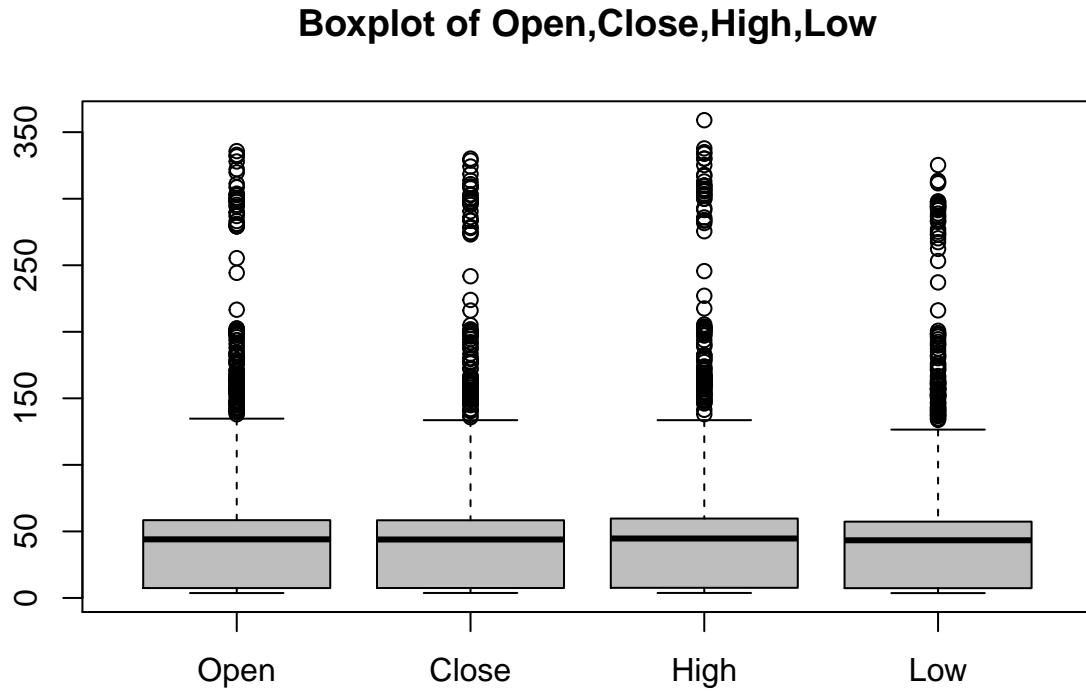
Since this is a timeseries data we converted the date column into date class using the function `as.date` and also checked the number of null values present in the dataset and remove them or impute them accordingly. There are 0 Null values in the dataset. Here the values for Close and Adj.Close are similar and hence the Adj.Close column can be removed from the dataset or it can be retained as it does not effect our calculations. Here, the format indicates the format of date column. There is no much preprocessing necessary for the dataset.

## D. Exploratory Data Analysis

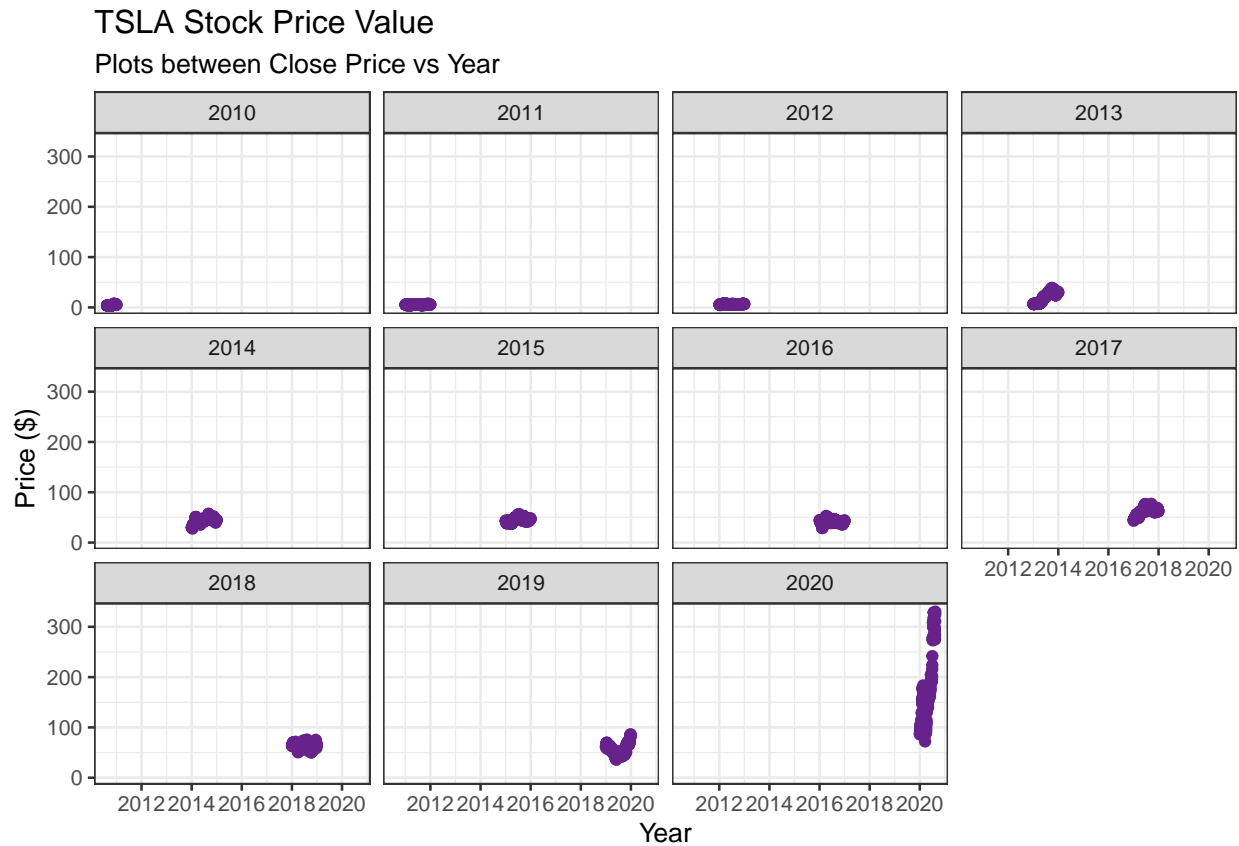
The below plot shows the time series plot of Close price value across the period of years 2010 to 2020. The plot has constant Close price value until 2012 and it spiked in year 2013 then increased gradually till 2019. But, there is a sudden spike of the value in the year 2020.



The below plot shows the boxplot of Variables Open,High,Low,Close. There are outliers present in the plot which can not be neglected in this case as they are useful in the calculations and analysis of the model. These outliers are present due to the sudden spike of values in the year 2020 due to several external factors. Notice that the high value has one outlier which is beyond the outlier interval and that could be the value of Maximum of High price value between years 2010 to 2020. The outlier interval ranges from 110 till 300.



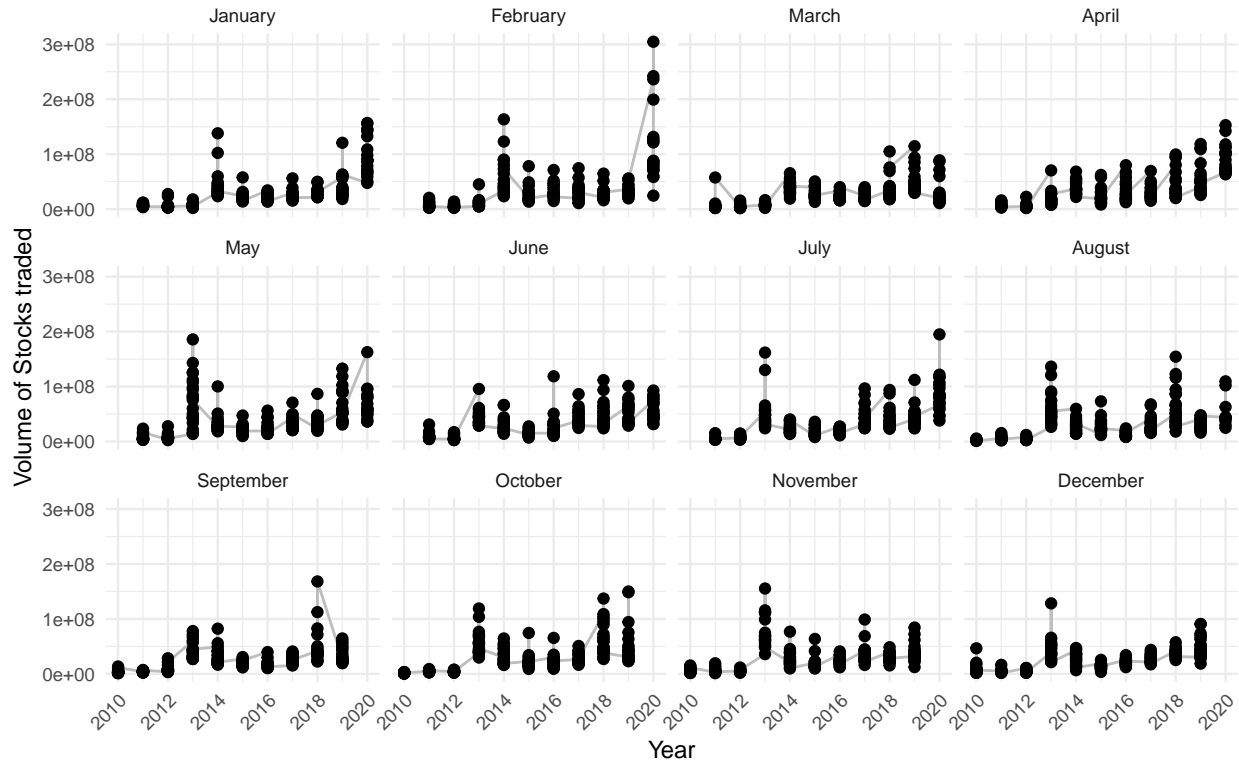
This plot shows the variation of Close price across the years. The plots are created using facet in order to plot the values across each year. From the plot it can be observed that close price value is constant in years 2010, 2011, 2012. The Close price value spiked by a large value in the year 2020. In the year 2020 the value spiked upto \$340.



These are plots that shows the variation of volume of TSLA stocks traded each month of every year. The x-axis ranges from 2010 - 2020. From the plots it is observed that the highest volume of stocks are traded in the month of february in year 2020. Also it is observed there are few spiked values in volume of stocks traded in the year 2013. The volume of stocks traded is not constant across the years. The variation of volume of stocks is due to several factors including the external factors.

### Plots of Volume of TSLA Stocks traded.

Every month from 2010–2020.



## E. Empirical Analysis

In the Multiple Linear regression model, I'm going to estimate the TSLA Close price value based on the Open, High and Close value. It is observed that the p value of f statistic is highly significant which indicates that there is one predictor variable which is strongly correlated to the output variable. The intercepts of High and Low price values are significant whereas, The open Price value is insignificant. Hence, I excluded the open price and generated the model again. Based on the summary it can be interpreted that on a particular day for a fixed high Price value an increase in **\$25 of Low price value leads to increase in Close price value by  $0.58 \times 25 = \$14.5$** . The adjusted  $R^2 = 0.99$ , meaning that "99% of the variance in the measure of close price can be predicted by high price value and low price value TSLA stocks. In the above multiple regression model, the **RSE is 1.248 corresponding to 2% error rate.**

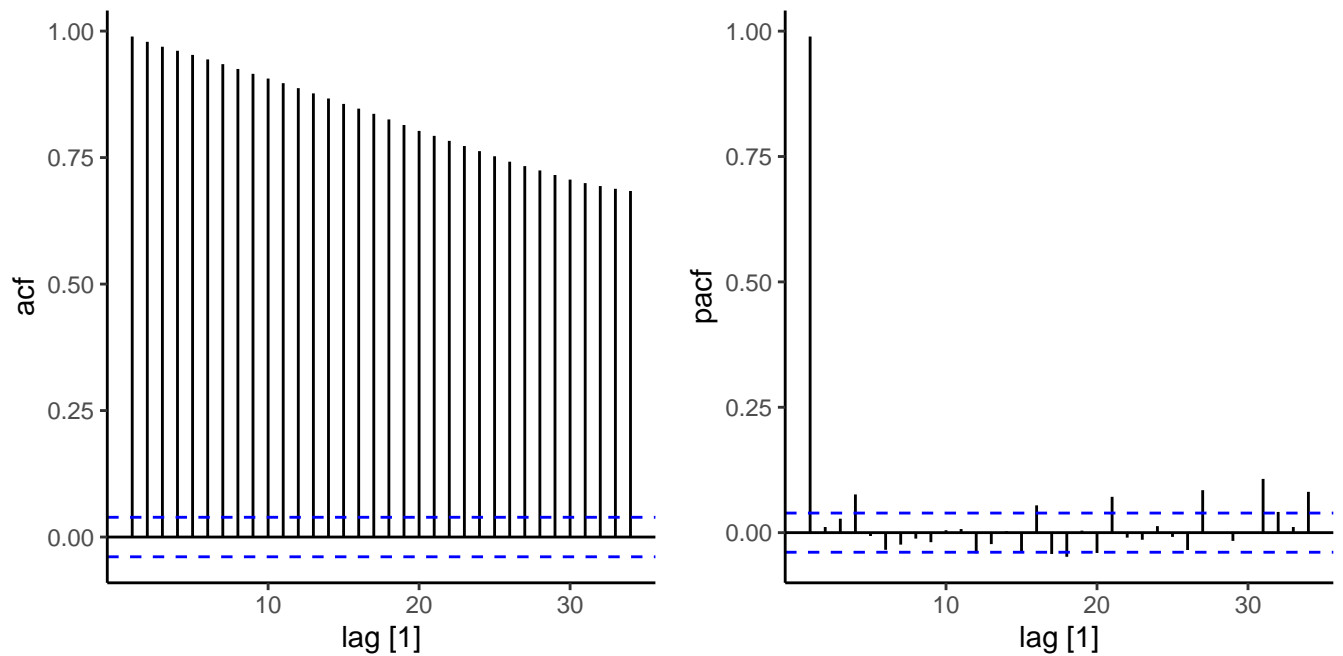
```
##
## Call:
## lm(formula = close ~ open + high + low, data = tsla)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3318  -0.2350   0.0301   0.1734  13.4351
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.05957    0.02895  -2.058  0.0397 *
## open        -0.65529    0.01605 -40.836 <2e-16 ***
## high         0.76868    0.01252  61.390 <2e-16 ***
## low          0.89141    0.01239  71.935 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9681 on 2514 degrees of freedom
## Multiple R-squared:  0.9995, Adjusted R-squared:  0.9995
## F-statistic: 1.684e+06 on 3 and 2514 DF, p-value: < 2.2e-16

##
## Call:
## lm(formula = close ~ high + low, data = tsla)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.6828  -0.3125   0.0051   0.2020  17.7402
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.04934    0.03732  -1.322   0.186
## high         0.42356    0.01191  35.555 <2e-16 ***
## low          0.58149    0.01263  46.036 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.248 on 2515 degrees of freedom
## Multiple R-squared:  0.9992, Adjusted R-squared:  0.9992
## F-statistic: 1.519e+06 on 2 and 2515 DF, p-value: < 2.2e-16

## [1] 0.02735096
```

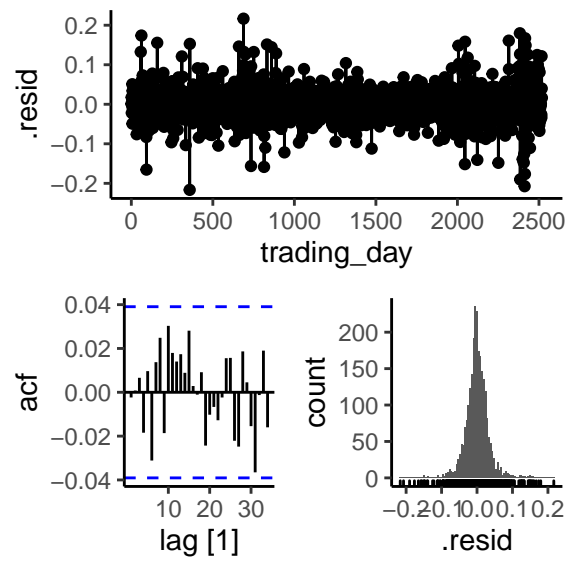
Another model I'm going to perform is ARIMA model. Since this is a time series data, using the ARIMA i'm going to forecast the TSLA stock close price value in future. On analyzing the time series graph it can be clearly seen that there is trend in the series and variance across few intervals. Also, On observing the ACF and PACF plots it can be concluded that the series is not stationary as the ACF plot has exponentially decreasing trend and PACF plot has one spike at one lag value. Hence the variance can be stabilized by taking logarithmic of the data. The Overall mean of the curve is stabilized by taking the first difference. The stationarity can be found out by doing kpss test. The higher P value indicates that the curve is stationary. Here, the p value is 0.1 and hence not rejecting the null hypothesis. Then, The ACF and PACF plots of the differenced data indicate that there is white noise. Hence the model parameters are estimated as (0,1,0). After the running the choosed model and also forcing arima to choose the model the parameter values are similar. The AICC values are same for both the models. The residuals of the model are displayed. The ACF plot of residuals shows all the lags inside the interval which indicate that there is white noise and all the information is captured and hence the forecasting can be done to predict the future TSLA stock close price value. The histogram of the residuals approximately shows the normal distribution. The forecast shows the prediction for 100 days which predicts darker area is 85% confidence interval and Lighter area is 90% confidence interval and the prediction for 100 days show that there will be increase in the TSLA Stock close price value.

### ACF and PACF plots

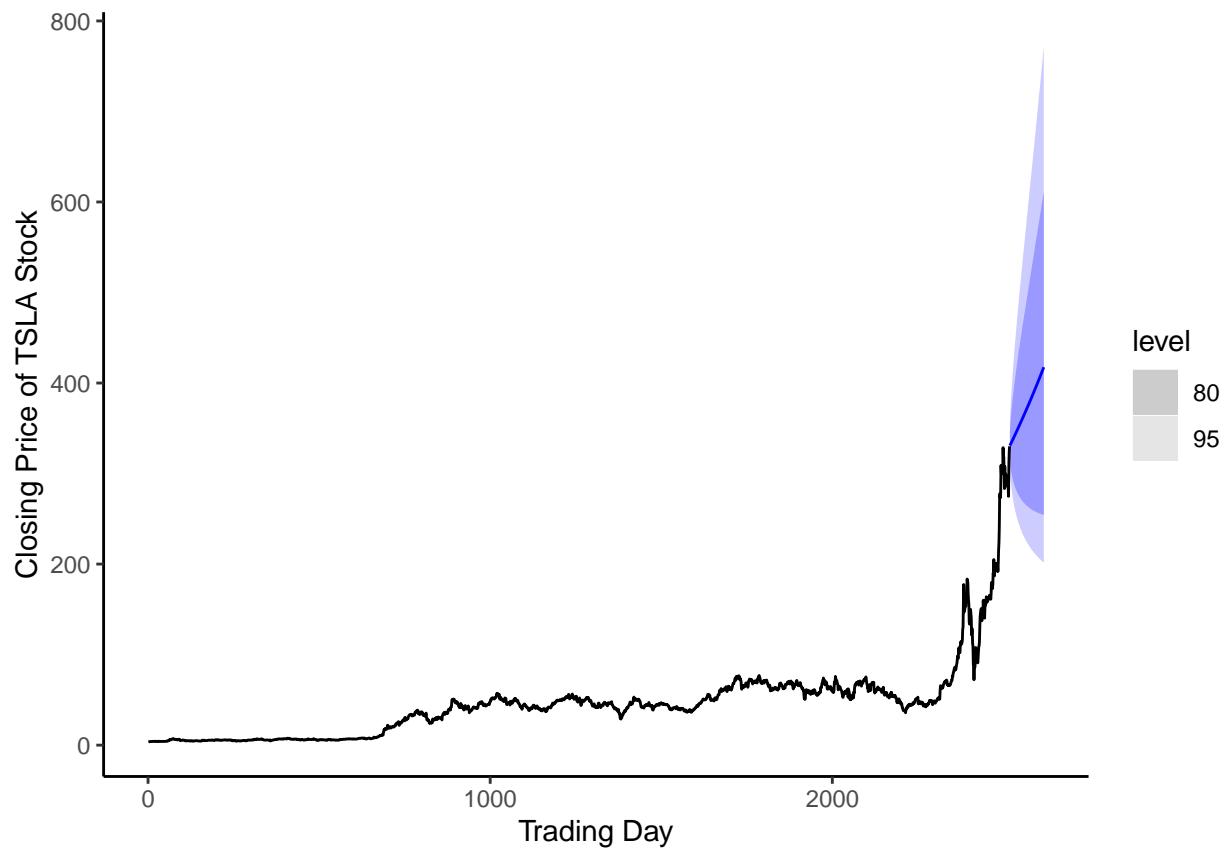




### ACF plot of Residuals



### Forecasting the time series, (h=100) days



## F. Conclusions

Using TSLA dataset from yahoo finance, i analyzed the relationship between the close variable and other variables and how other variables affect the close price value. The multiple linear regression model predicts that “99% of the variance in the measure of close price can be predicted by high price value and low price value TSLA stocks. The forecast of ARIMA model shows the prediction for 100 days and there will be increase in the TSLA Stock close price value. ARIMA model has been used extensively in the field of finance and economics as it is known to be robust, efficient and has a strong potential for short-term share market prediction. All investments and trading in the stock market involve risk. Any decisions to place trades in the financial markets, including trading in stock or options or other financial instruments is a personal decision that should only be made after thorough research, including a personal risk and financial assessment.

## G. Sources

- [<https://finance.yahoo.com/quote/TSLA/history?period1=1281830400&period2=1597449600&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true>]
- [<https://www.earthdatascience.org/courses/earth-analytics/time-series-data/date-class-in-r/>]
- [<https://rpubs.com/kapage/523169>]
- [<http://www.sthda.com/english/articles/40-regression-analysis/168-multiple-linear-regression-in-r/>]