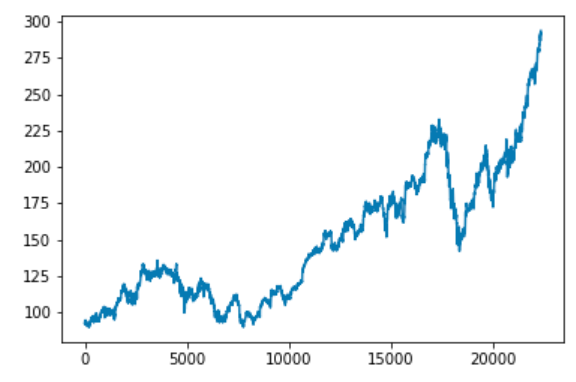
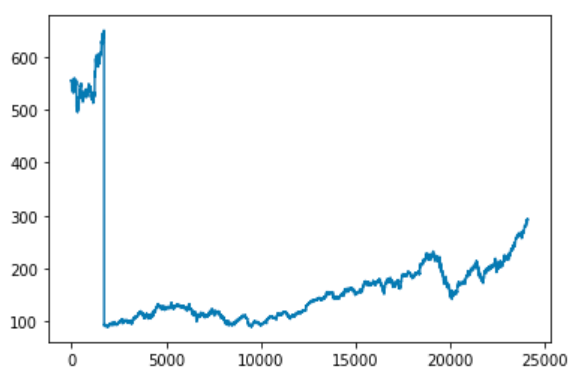
**LINEAR REGRESSION**

**Data and the Processes**

The data consists of an hourly Apple Inc. (AAPL) data for stock price, volume, market depth, number of trades and spread. Relying on the data taken from Reuters, I have performed a linear regression and decision tree regression. Observing the results from linear regression, I have seen that the R2 ratio was slightly high. Stocks have various factors that affect their prices accordingly, thus it was not very logical to have only 4 of these variables explaining the 0.144 percent of the price.

At this point, I have decided to eliminate outliers, which was a sharp price decrease on AAPL stock price on 2014. From 25,000 data points, I have eliminated approximately 1,700 of them. The elimination of these data given me a better line to observe and examine for relations.

**\**The Following Graphs has y-axis as price, and x-axis as data points:*

*\*Before the elimination \*After the elimination*

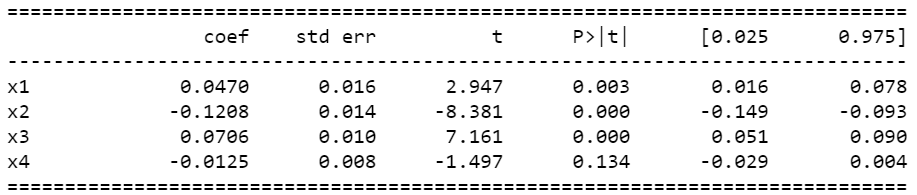
In addition to data elimination, I have standardized all of the values for the linear regression analysis to understand the relation better and reduce the errors due to the data range difference between spread and volume. Although I have also performed regression for the old data, I will only discuss the regression after the elimination.

**Hypothesis**

The hypothesis of this work was to understand if there was any relation between the coefficients of the X variables:

y = β1x1 + β2x2 + ... + βkxk + ε

y-hat = b0 + b1x1 + b2x2 + ...+ bkxk

The regression result is as follows:

Where:

x1 is Number of trades  
 x2 is Depth  
 x3 is Volume  
 x4 is Spread

Choosing alpha value as 0.05, the results show us all of the X variables contribute to the stock price explanation except the spread.

**Understanding the Coefficients**

Number of trades has a positive effect on price, where depth is negative; which shows us the people who are trading AAPL at the time period of the data are more likely to sell. In other words, depth-number of trades relation explains the part of people that wants to sell this stock. Also, the volume is positive, thus it contributes positively to the price increase.

The coefficients tell us the effect of volume and number of trades combined are close to depth’s negative effect. Thus, the depth is a more important factor while explaining the market price rather than the volume and the number of trades.

**DECISION TREE REGRESSION**

**Background and Process**

I have performed the decision tree analysis for both the data I have gathered from Reuters and to the data I have calculated through python myself. The data calculated consists of the RSI, Simple moving average, exponential moving average and MACD-Signal.

RSI is relative strength index which is used to see if the market is oversold or overbought. MACD is used by traders to see if there is any convergence/divergence in different spanned moving averages of the same stock. MACD usually used with Exponential moving averages. I have also added a signal line to understand the convergence even beforehand.

Then, I have took the difference of MACD and Signal values to observe when the subtraction changes its sign. In other words, when there is a convergence in the MACD-Signal lines.

**Results**

I have observed that both mean absolute error and root mean squared error were lower on the data taken from Reuters. Sample observations of predicted y and actual y are also inside the code.