**Stock Market Price Prediction:**

There are 2 datasets. One dataset containing the Top 25 headlines for each day for the past 8 years and the other dataset contains stock price data for Dow Jones. This is also the dataset used in my Capstone Project.

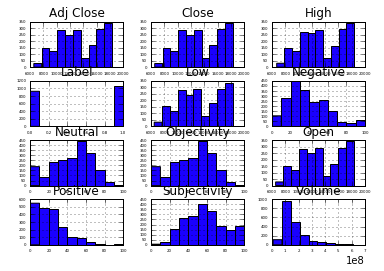
After examining the dataset, I found out there are no outliers. Each row in the headlines dataset was iterated over an algorithm which generated the Subjectivity, Objectivity, Positive, Negative, Neutral sentiments of the respective headlines of each row.

The algorithm was accepting only a single sentence and was providing the respective sentiments in percentage. I modified the algorithm iterate over all of the individuals rows and simultaneously create the Subjectivity, Objectivity, Negative, Positive, Neutral values and assign itself to the columns in the dataframe.

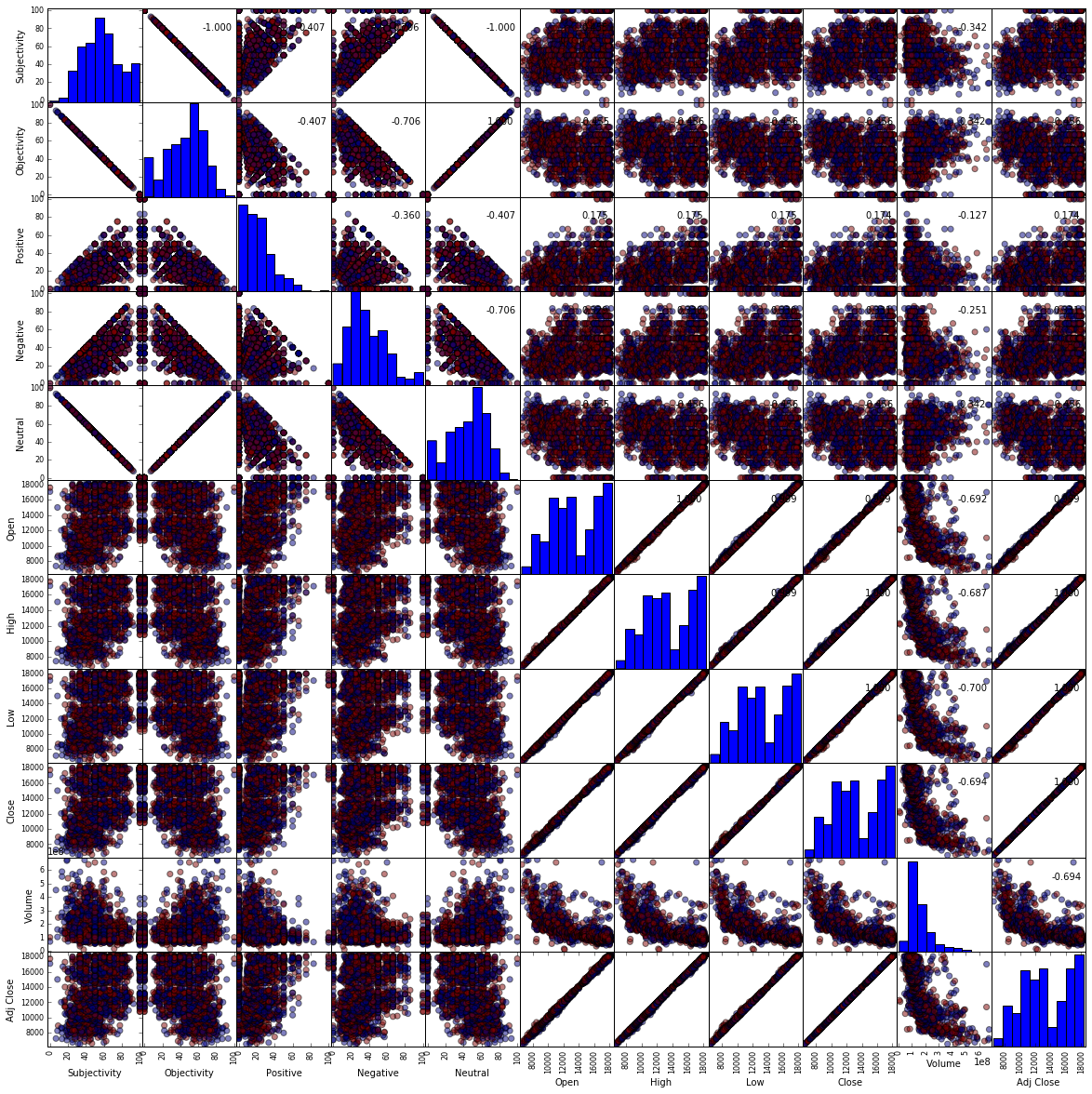
The headlines Top1 through Top25 were concatenated and then passed on to the algorithm.

Merged the newly generated columns with the stock prices dataset. Generated a histogram of each column to see how the data is distributed.

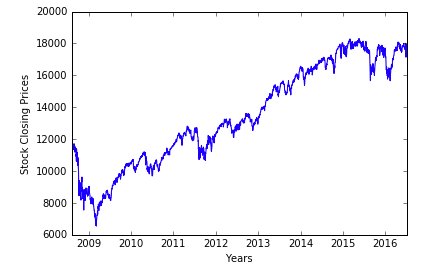
Most of the columns seem to have a normal distribution.



Plotted a scatter matrix to see the correlation of the columns with each other. Below is the scatter matrix.



As this is a time series, I planned to train the data on an ARIMA model. Before training the dataset, it was important to see if we see a trend in our data. So I plotted a line plot of Closing Prices vs the years to see if there is a trend in closing prices over the years. Below is the plot.



There seems to be a positive trend as we can observe the stock prices going upwards over the years.

After creating an ARIMA model, I plotted a graph of the predicted values vs the expected values. The model seems to fit very well. Red line is the predicted value and blue is the expected value. Below is the plot. **Note:** Prediction is on the testing data.

