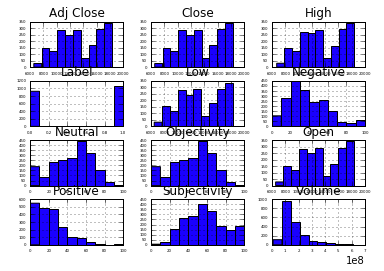
**Inferential Statistics:**

Inferential Statistics in this project mostly involved performing data quality checks at different steps of data wrangling process. I performed checking the statistics of individual columns to check for any anomalies or any outliers in the dataset. I was not able to find any outliers in the dataset. Only thing was that some of the columns had NaN values which further dealt with in the data wrangling step. The NaN values were replaced with the mean values of the set.

This was decided based on checking the distributions of the individual columns in the dataset. As most of the columns in the dataset seems to be normally distributed, the best decision was to replace the NaNs with the mean value of the respective column.

This is how the distribution looks like for individual features in the dataset:

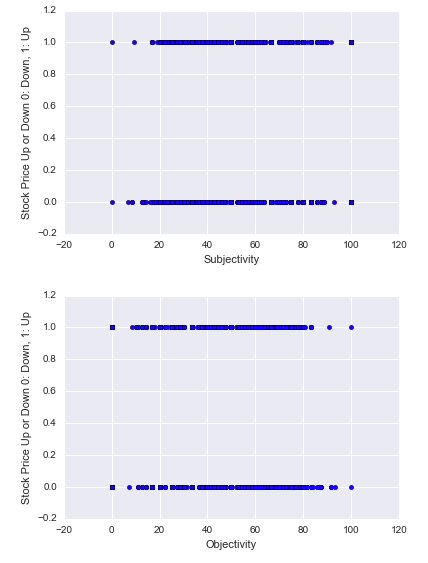


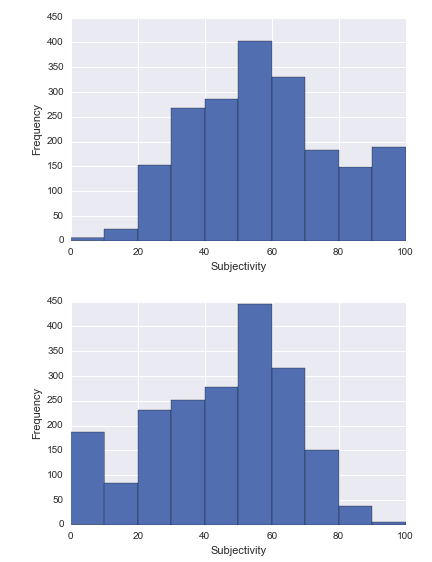
Plot 1: Scatter plot of Stock Prices vs the Subjectivity. Stock Value of 0 means the Stock Value reduced since the previous day. Stock Value of 1 means the Stock Value increased or remained the same since the previous day.

Plot 2: Scatter plot of Stock Prices vs the Objectivity. Stock Value of 0 means the Stock Value reduced since the previous day. Stock Value of 1 means the Stock Value increased or remained the same since the previous day.

Plot 3: Histogram of Subjectivity column. The x axis are the values of Subjectivity and y axis is its respective frequency. The plot seems to be normally distributed.

Plot 4: Histogram of Objectivity column. The x axis are the values of Objectivity and y axis is its respective frequency. The plot seems to be normally distributed.





**Correlation Map for features:**

Now, we will plot a heat map and a scatter matrix to see the correlation of the columns with each other. You can see the heat map with pearson correlation values in the plot below.

This gave me a better understanding to see if there are any dependant variables or if any of the variables are highly correlated.

Some variables Subjectivity, Objectivity are negatively correlated. There are very few variables which seem to have a very high correlation. Thus, at this point we can conclude that we do not need any sort of dimensionality reduction technique to be applied.

