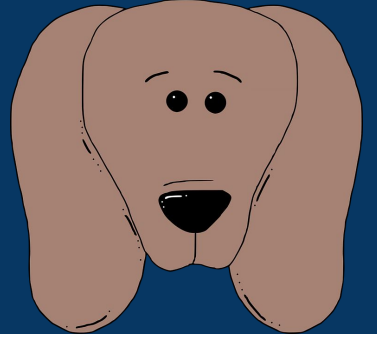
Wrangle and Analyze Data Project
#DAND



Act Report

-Reshu Singh

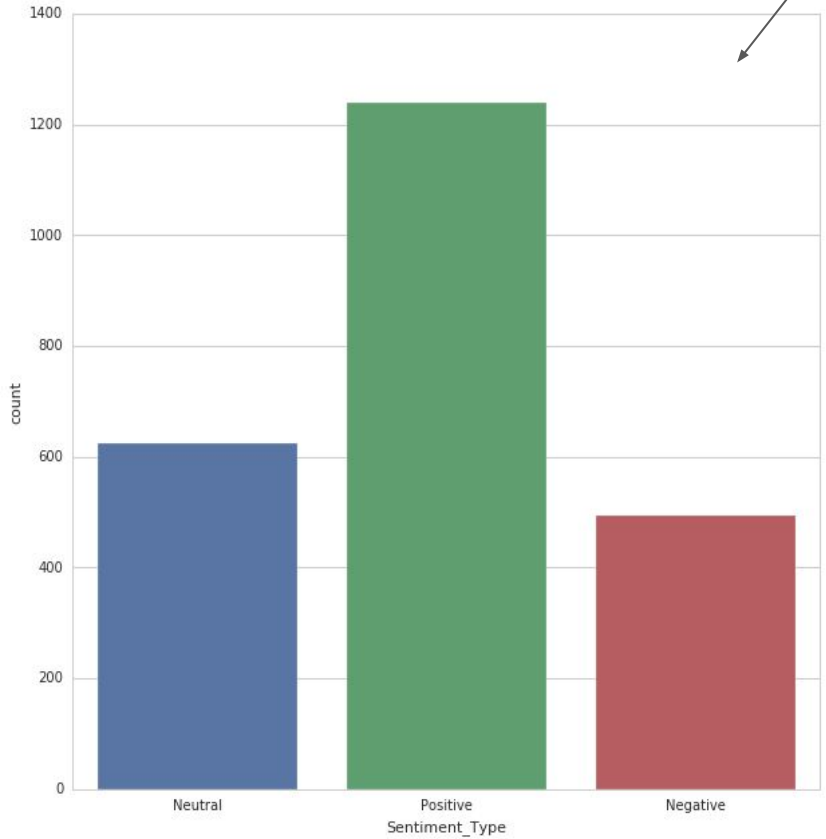
#Project Overview

Real-world data rarely comes clean. Using Python and its libraries, we can gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling. The task is to document our wrangling efforts in a Jupyter Notebook, plus showcase them through analyses and visualizations using Python (and its libraries) .

The dataset that we are wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs.

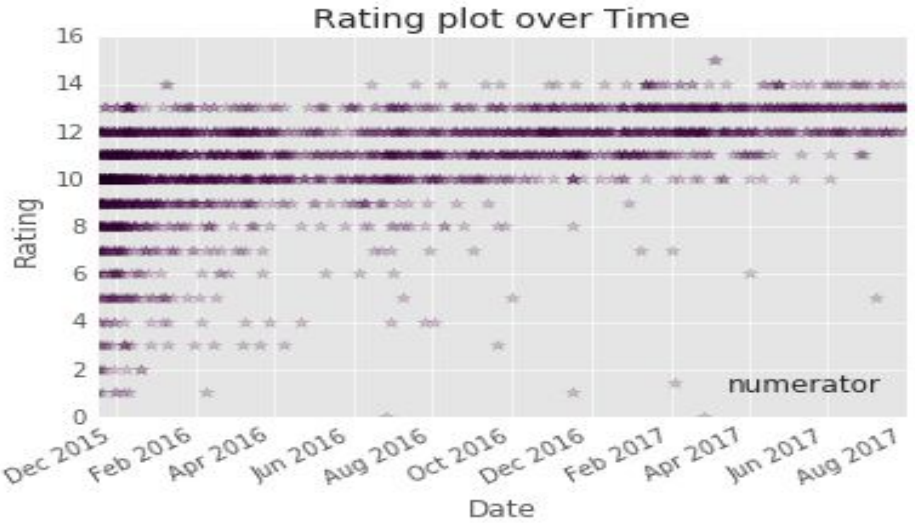
WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog.

#Visuals

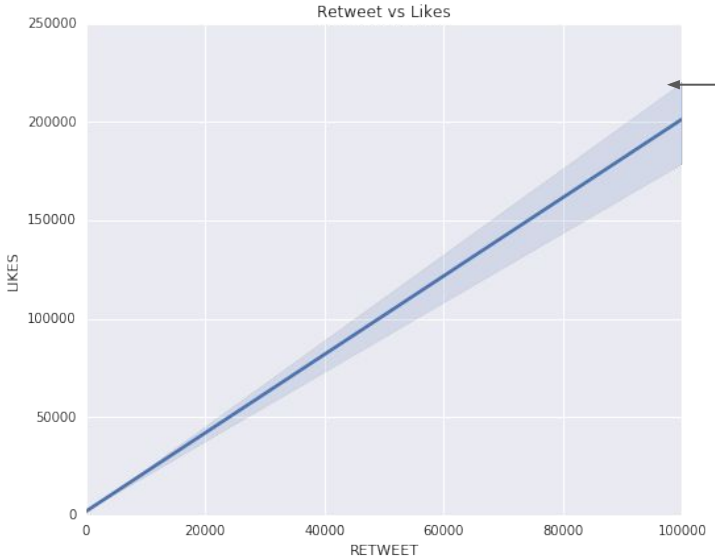


Performed text analysis on ‘text’ column of data frame. Used NLTK and textblob API. Negative,Positive and Neutral polarized texts diiferentiated and found out that most of the texts are positive and then neutral (neither positive nor negative) and at last very less are negative.

This way we can infer that WeRateDogs mostly rate dogs good points and comments positively.The same can be inferred from rating plot below aswell.



#ANALYSIS AND VISUALS



As the retweets increases,so is the likes and vice versa. There is linear correlation here.

At the time the data has been collected, the twitter account seems to be very much active as we can see no. of likes and retweets are very high in number maximum as can be around 100000 to 200000 ranges.

This shows that most dogs are in Puppo category but highest retweets are about doggo category.

There are many outliers as can be seen as trailing tails of boxplots. Plus it can be visualized that the count of retweets are in very large no. Means WeRateDogs is damn popular among its followers.