ACT REPORT

WeRateDog Project

Introduction:

Data visualization is the communication of findings and results in a graphical representation. Communicating by using visual elements such as charts, graphs, or maps. Visualization interprets data. Upon creation of the master dataframe from Data Wrangling stage in WeRaeDog project, data have become ready for analysis and further visualization. Steps are illustrated in Figure 1. and will be detailed further.

Figure 1: Visualization model

1. **Basic Linear Plot:** from the master dataframe, found Rating score which is the result of rating_numerator divided by rating_denominator can be plotted along the time, showing how people used to rate dogs. Figure 2

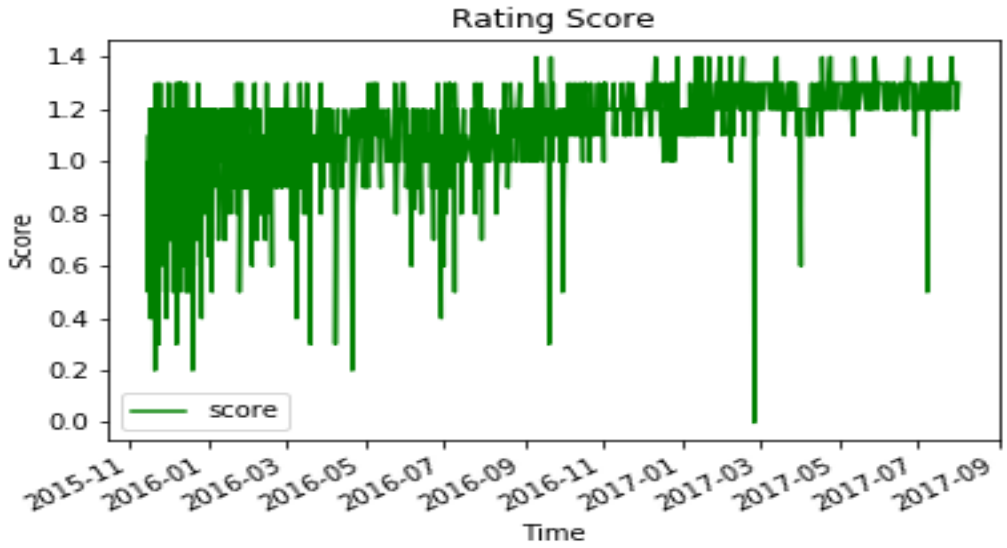

Figure 2: Rating Score

2. **Percentages:** looking for the number of retweets per year from the dataframe indicated that most of the retweets happened to be in year 2016. Retweet would be better to be plotted as a pie chart to indicate the volume per year for instance.

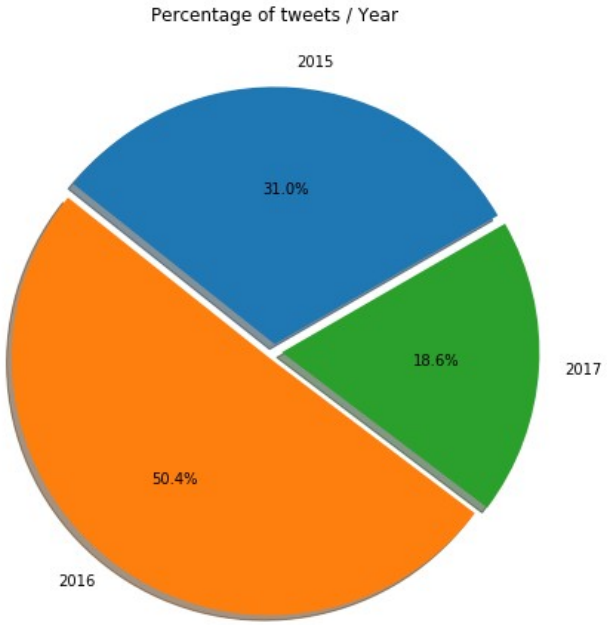

Figure 3: tweet % by year

Similarly, Plotting by month also indicated that most tweets happened to be in November and December consecutively

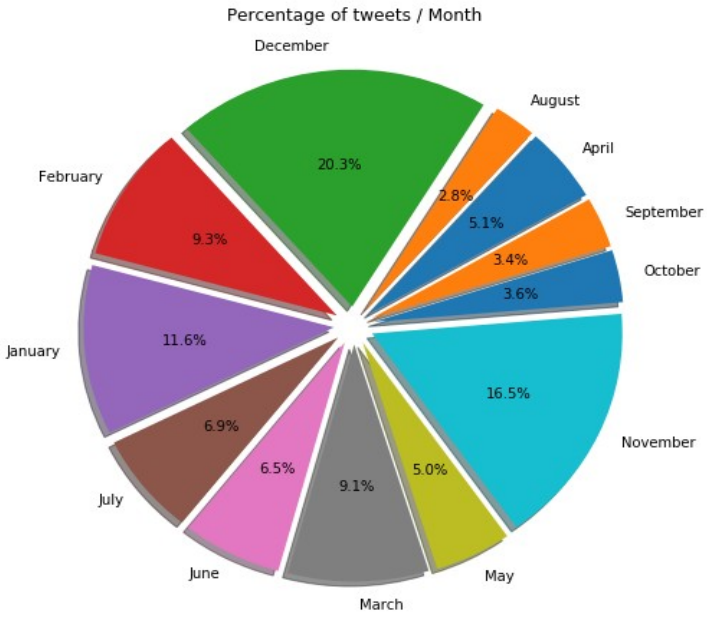

Figure 4: tweet % per month

3. **Correlation:** according to the correlation test on the master dataframe, tweets and favorites tend to have a relationship, a person tweeting a rating of a dog most likely to receive a favorite in his account. This relationship can be visualized using seaborn library heatmap. Figure 5.

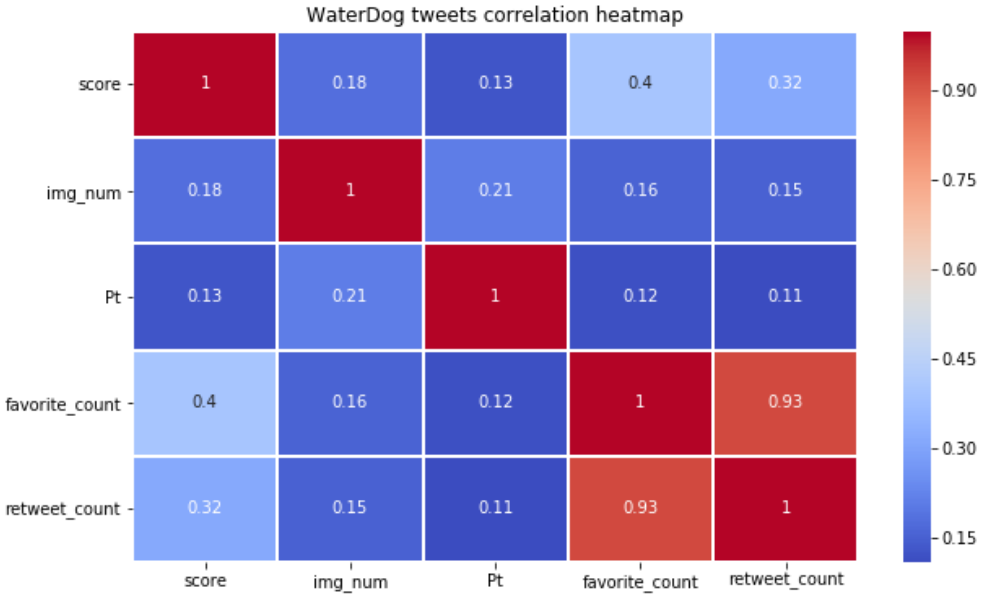

Figure 5: Correlation heatmap

Retweet vs favorites can be further visualized specifically using the scatter plot, showing the strength in relationship. Figure 6 indicates a strong positive relationship between the two parameters.

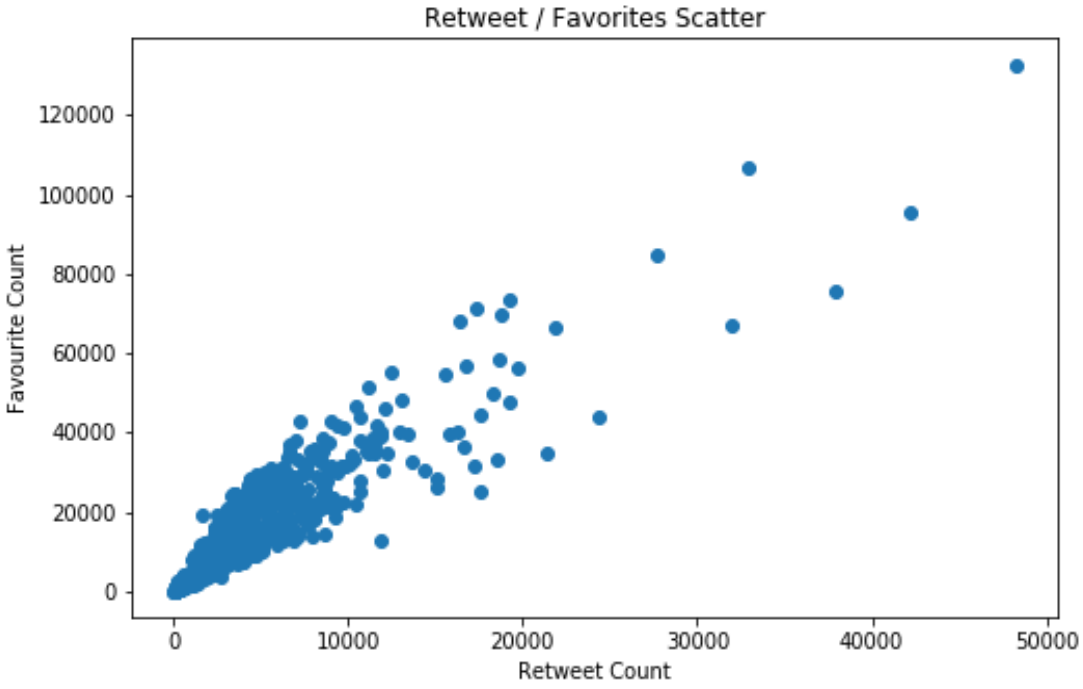

Figure 6: scatter plot

4. **Normal distribution:** the relationship between retweets and favorites also can be visualized on a histogram as in Figure 7.

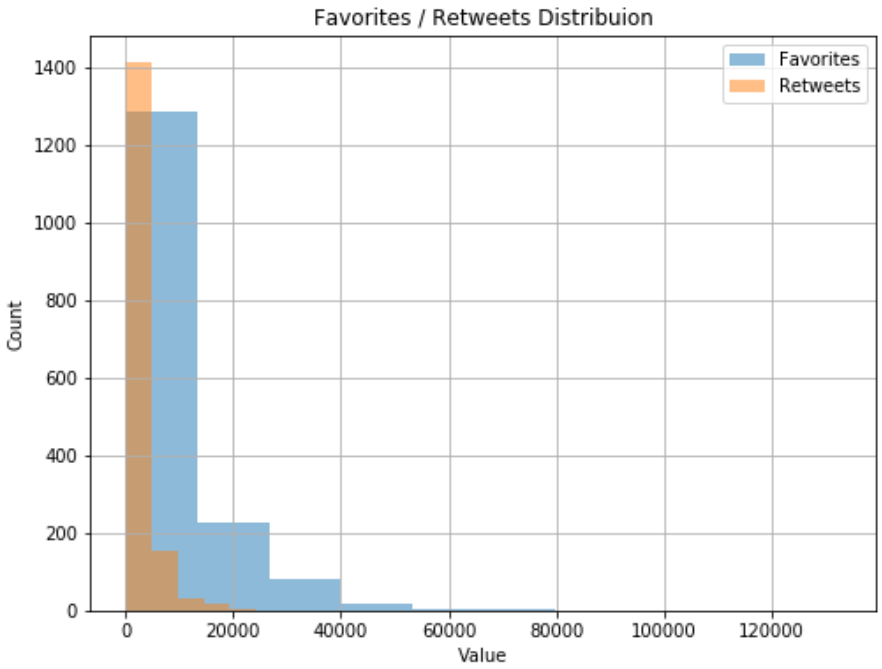

Figure 7: Histogram

Conclusion:

There found to be a strong relationship between tweets and favorites indicating the most likeliness to receive a favorite when tweet a rating of a dog. Dog ratings tend to be high most of the time as indicated by the score value. It also found that most tweets happened to be in November and December, year 2016 have recorded the most count of tweets.