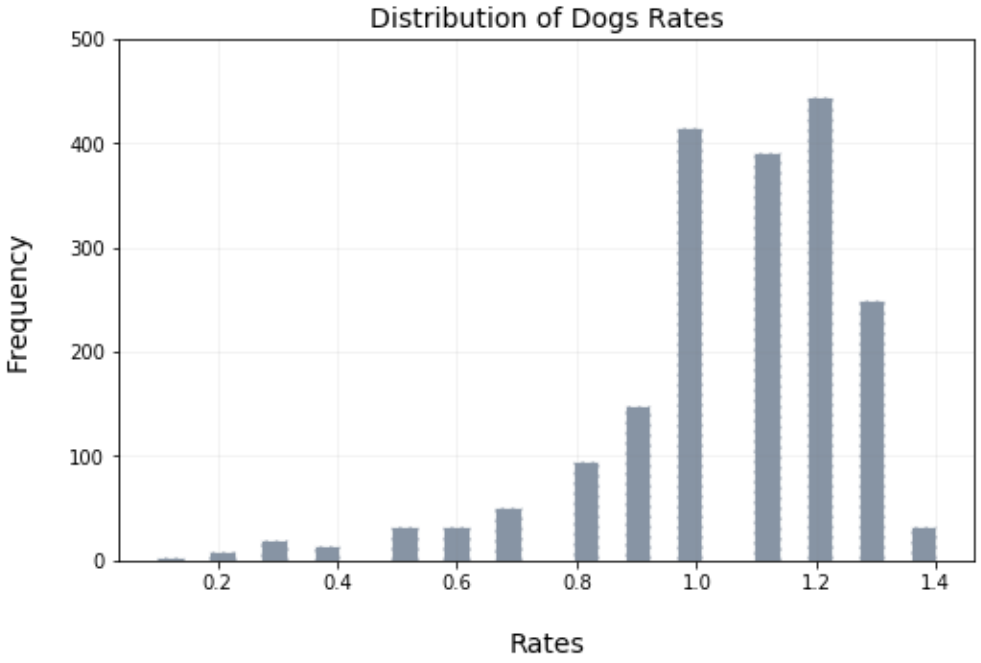
Project#4 - Wrangle and Analyze Data

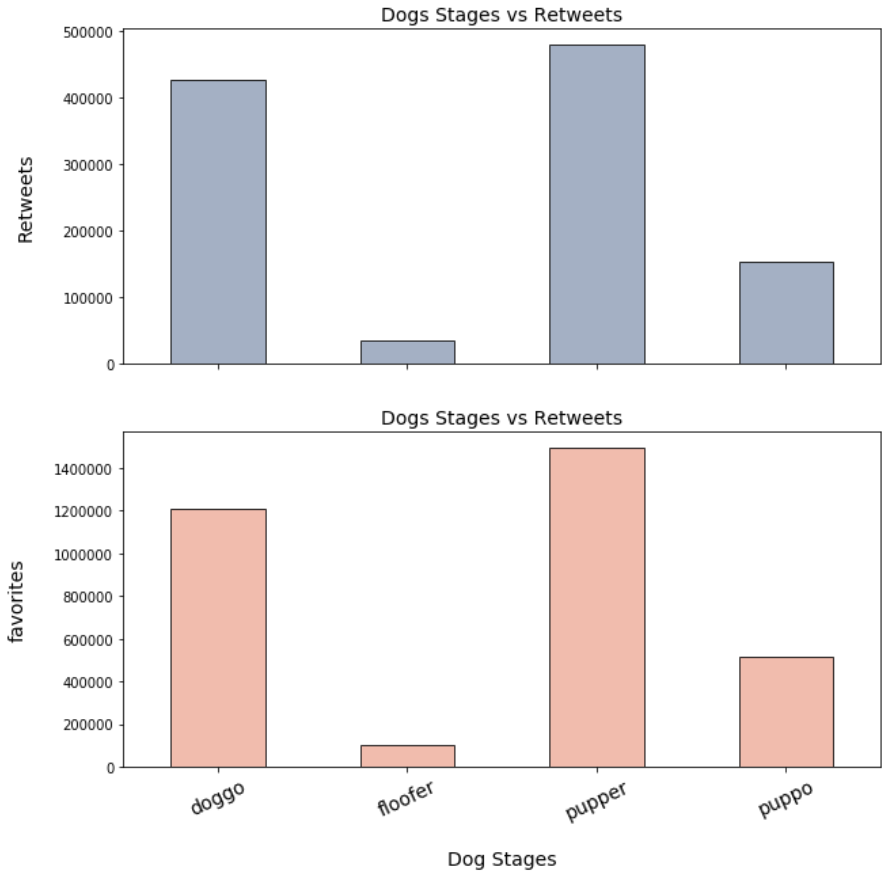
Developed By: Rakan210 on June 22, 2019 - as part of Udacity Data Analyst Nanodegree program (DAND)

Project's Analyzing and Visualizing:

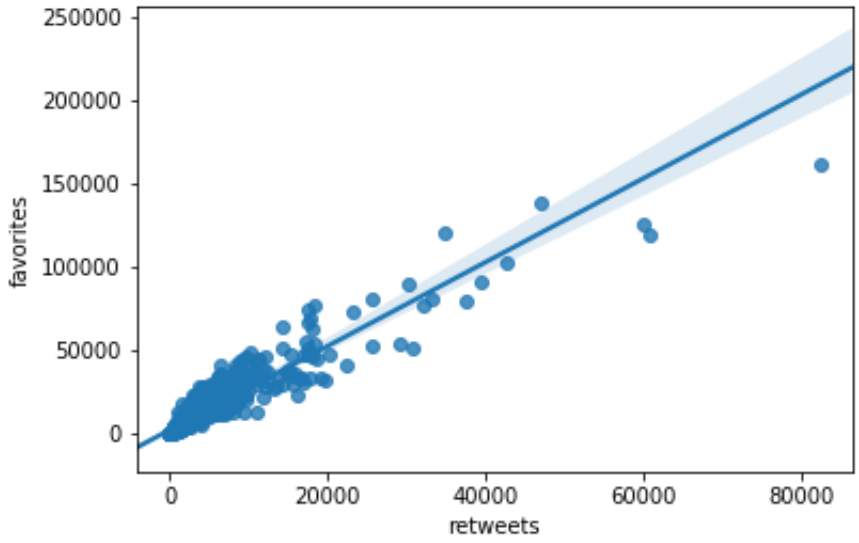
Visualizing the distribution of dogs' rates clearly shows that, the graph is skewed to the right with a majority of rates that are greater than 10:



Considering other point of views such as the number of retweets and favorites for each stages on the dataset (i.e Pupper, Doggo, Puppo and finally Floofer) shows an interesting findings. The number of total retweets for the highest stage, which is Pupper, crosses the boundary of half a million retweets. The same also applied for the number of favorites which crosses the 1.4 million favorites:



This also led us to make a reasonable kind of comparisons that shows the direct relationship between the number of retweets and the number of favorites for all the records in the cleaned dataset. As illustrated below, the graph shows that strong relationship between the two counters. The majority of the tweets got retweeted around 10K times and has been favorite around 25K times.



Project's Insights

1. Funny fact: Most of the dogs in this cleared dataset have rates that are more than 10 out of 10 :). This is coupled with the following facts:
    - Rates Mean: 10.05/10
    - Rates Median: 11/10
    - Rates Max: 14/10
    - Rates Min: 1/10
2. The most favorite Dog Stage is (Pupper) stage, which has both the heights retweet and favorite's rates. This stage is followed by Doggo, Puppo and finally Floofer stages.
3. As the number of retweets increases, the favorite rate is positively increases.