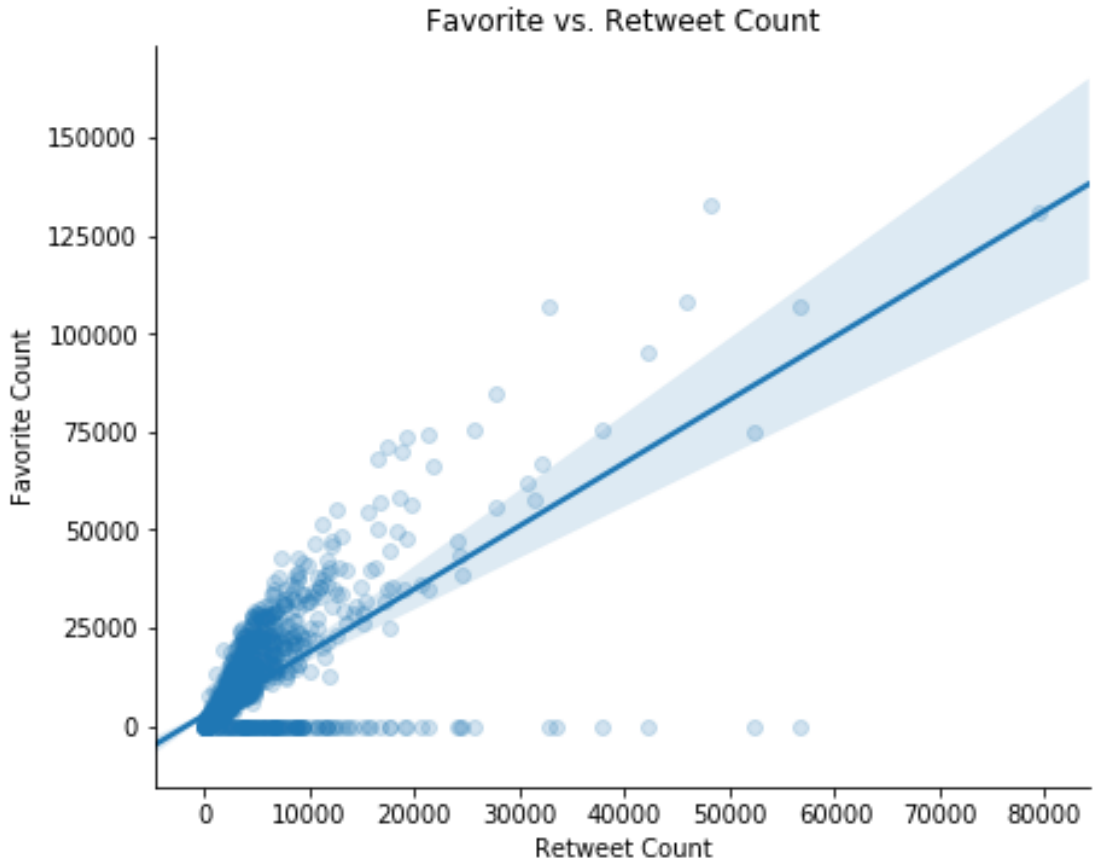
Introduction:

This Wangle and Analyze Data Project is part of the Udacity Data Analyst Nanodegree program, the data set is from the twitter account WeRateDogs, as know as: @dog_rates, which provides information about dogs. The data set contains 2300+ pieces of data, which were stored in 3 different sources. After done gathering, assessing and cleaning the data, from the visualization in the end, we can conclude some insights from the wrangling and analyzing activity.



As the Favorite vs. Retweet count picture shown above, the horizontal axis indicates Retweet count while the vertical axis shows Favorite count. However, even with some outliers existing, from the picture we can see obviously that these two variables have a positive correlation. Majority of the data are seen at the left bottom corner, namely, for the retweet count, are around from 3000 to 10000; while for the favorite count are set in the range of 15000 to 25000.

WeRateDogs asks people to post photos of their dogs, and asks the

followers to rate the dogs, ranging from 1- 10.

As the figure below shows, the horizontal axis indicates the time range, while the vertical axis does the rating score, over the given time range, despite the existence of some outliers, the ratings below 1.25 are majority, these ratings are crowded at the early stage of the time range. However, as time going by, the number of these ratings are decreasing, on the other hand, the fluctuation of the rating scores are narrowed down to a smaller range by time. and the overall ratings are leveled up around the score range of 1 – 1.25.