

# Analysis of WeRateDogs Twitter data

*By Geo Niju Shanth G*

In this data wrangling project I have worked with three different source of data. It was very challenging but finally able to achieve what I wanted. We have one file which is provided by udacity as csv file other source one is from a http link and the other through twitter API. Twitter archive data is provided in the form of csv files. The image predictions data is programmatically downloaded using requests python library file as a tsv file. Each tweet's JSON data is queried from Twitter API using Python's Tweepy library. Tweepy was really interesting where in we can analyse the data on a real time basis. In the case of this project we have pulled the data for corresponding tweet ids from which we can get the retweet and favourite counts.

The first step in gathering the insight was to do statistical analysis on the resultant dataset using describe function. Then I started asking questions like which twitter image was the least rated one. The answer was found by fetching the tweet id corresponding to the lease numerator rating. Below is the image of the dog that was rated very low.

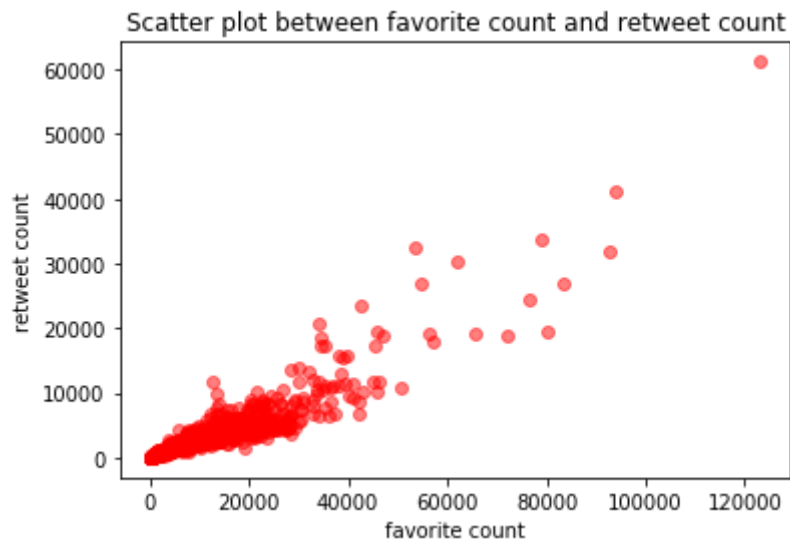


The next question I was trying to answer was which dog was rated very high. Below is the screenshot of the image which I have got while answering the question.

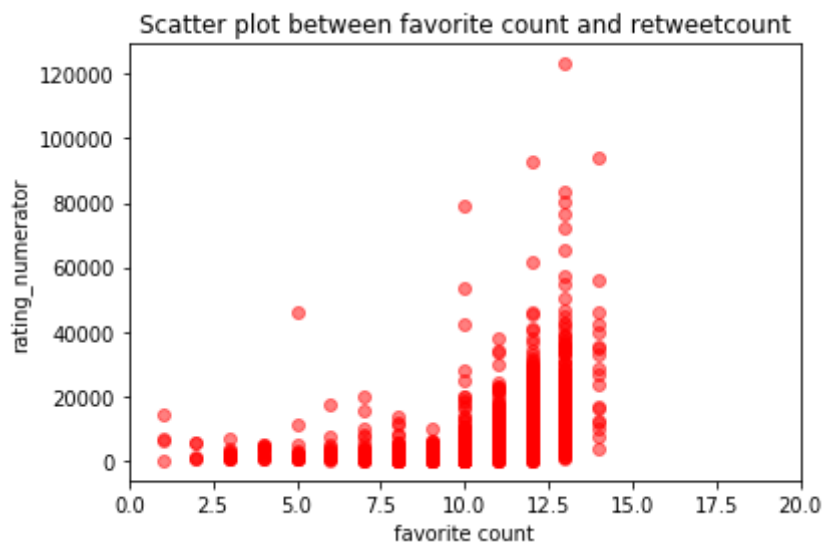


Then my question was which were the top 5 dogs which were having more retweet counts. It turns out to be Conch, Upright, Bubble, Gondola and Angora had the most retweeets. The next question was which were the top 5 dogs which were having the more favourite counts. It was found to be Angora, Limousine, Conch, Bubble and Beaver.

For visualization I have tried to understand the relationship between favourite count and retweet count, favourite count and the rating numerator using scatter plot. Below are the visualizations.



From the above plot we can see that the favorite count and retweet count has a positive correlation between them.



From the above plot we can see that the tweets with higher favorite count are associated with higher rating.

From the above analysis we can see that the retweet and favorite count are strongly correlated.