UDACITY DATA ANALYST DEGREE

Data Wrangling Project II

Submitted by:

Janely Padillo

In this data analytics project, we analyze the tweets from renowned Twitter account We Rate Dogs. Our datasets include information about dog names, categories, tweet image urls, location of tweets, tweet lengths, date and time of tweet creation, tweet image predictions, ratings,favorite counts and retweet counts, among others.

Insights:

1. Out of 17,738,077 total number of favorites as of August 1, 2017, the most favorited tweet only makes up 0.74% or 132,810 favorites. This tweet was created on 01-21-2017 and has a standardized rating of 1.3. The tweet contains 87 characters and the neural network's top 3 predictions for the image are as follows: Lakeland Terrier, Labrador Retriever, Irish Terrier

2. Out of a 5,516,906 total number of retweets as of August 1, 2017, the most retweeted tweet only makes up 0.87% or 79,515 retweets. This tweet was created on 06-18-2016, has a standardized rating 1.3 and there are 91 characters in the tweet. The neural network's top 3 prediction for the image accompanying the tweet are as follows: Labrador Retriever, Ice Bear and Whippet

3. Under dog categories, 66.34% have been assigned to the category pupper, followed 20.6% of doggo assignments, 7.19% puppo assignments, 3.6% for dogs that were assigned multiple categories and fluffo assignments following closely behind at 2.2%.

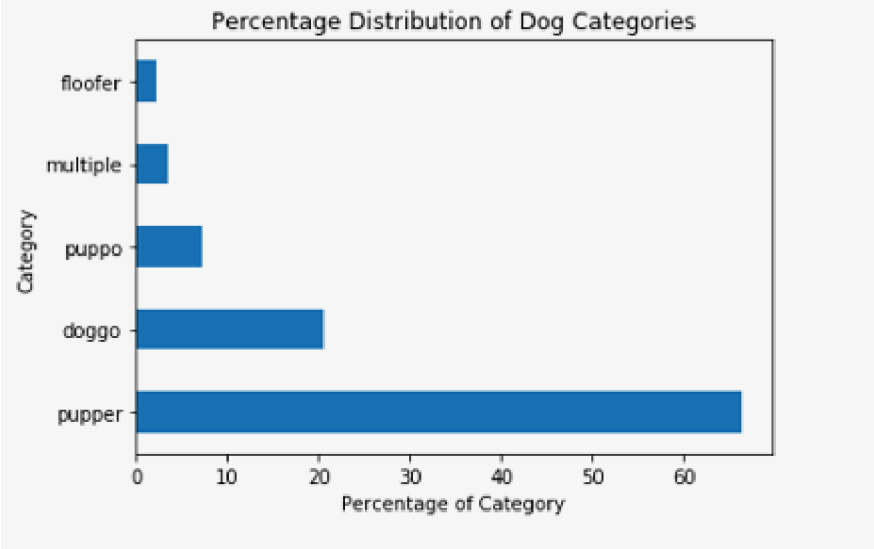

Image 1.1 *A bar graph showing percentage distribution of dog categories pupper, doggo, puppo, multiple dog categories & floofer.*

4. We Rate Dogs standardized ratings, which is the quotient of the rating numerator divided by the rating denominator, has a mean of 1.22 and a standardized deviation of 4.06. The minimum value for rating is 1.0 whereas the maximum value for rating is 177.60, which is regarded as an outlier in this project.

5. The most common predictions by the neural network, which encompassed all of the first, second and third predictions, are as follows: Golden Retriver, Labrador Retriever, Pembroke, Chihuaha and Pug

6. According to the scatter plot created below containing values for favorite counts and retweet counts up until August 1, 2017, it appears that there is a positive correlation between the tue. With that said, since no inferential statistics was applied, we cannot say for certain whether this is matter of fact.

It appears that, according to the scatterplot containing the data of favorite counts and retweet counts, there may be a positive correlation between the two. However, since we did not perform any statistical operations, we are unable to say for certain.

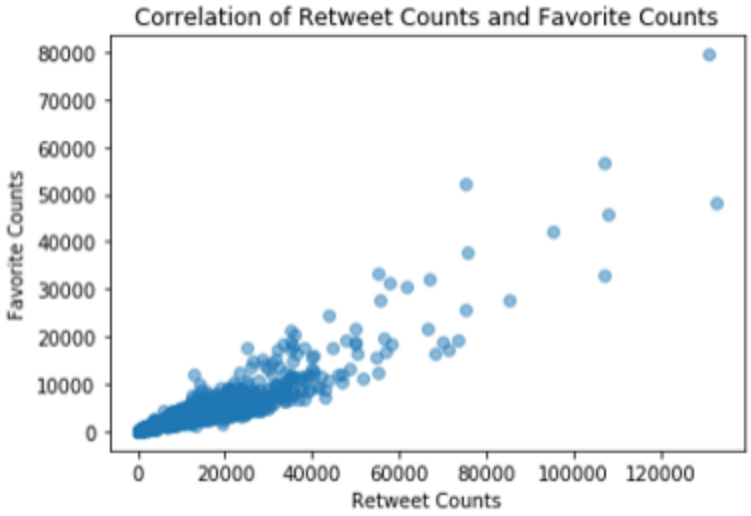

Image 1.2 A scatterplot showing the correlation of retweet and favorite counts

All in all, this dataset shows how popular We Rate Dogs is amassing over 17 million favorites and 5 million in just two years after it was first created. It