

Wrangle and Analyze Data

By Fahad Alsaeed

Introduction:

This project focuses on wrangling and visualizing a dataset that has been gathered from different sources. In order to explore a twitter account called @WeRateDogs where the account post a tweet of a dog and letting the audience rate.

Insights:

Most Common Dog Type:

Based on figure 1, we can see the golden retriever has the most of rated dogs with more than 140 ratings.

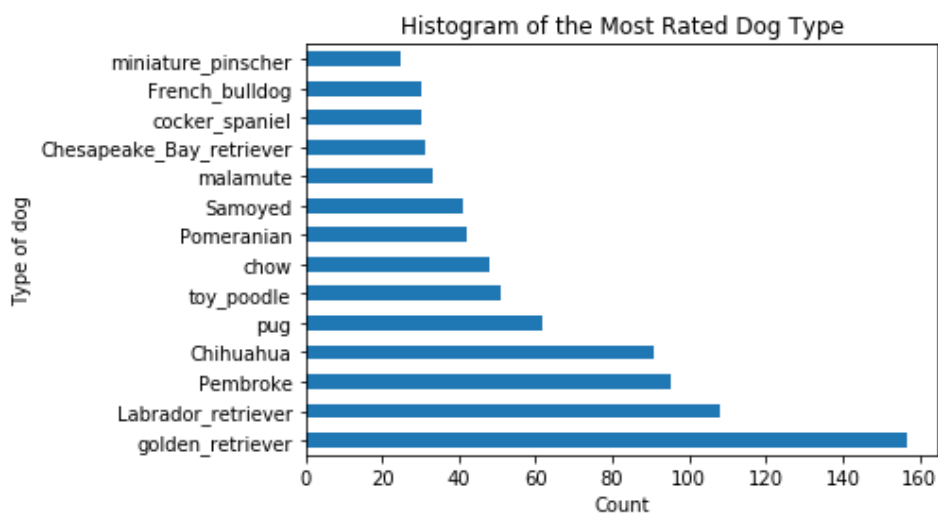


Figure 1

The retriever is very popular based on blue dot on the very right, in the plot below, represents the approximately 140 ratings of golden retriever and average these ratings are high 11.59.

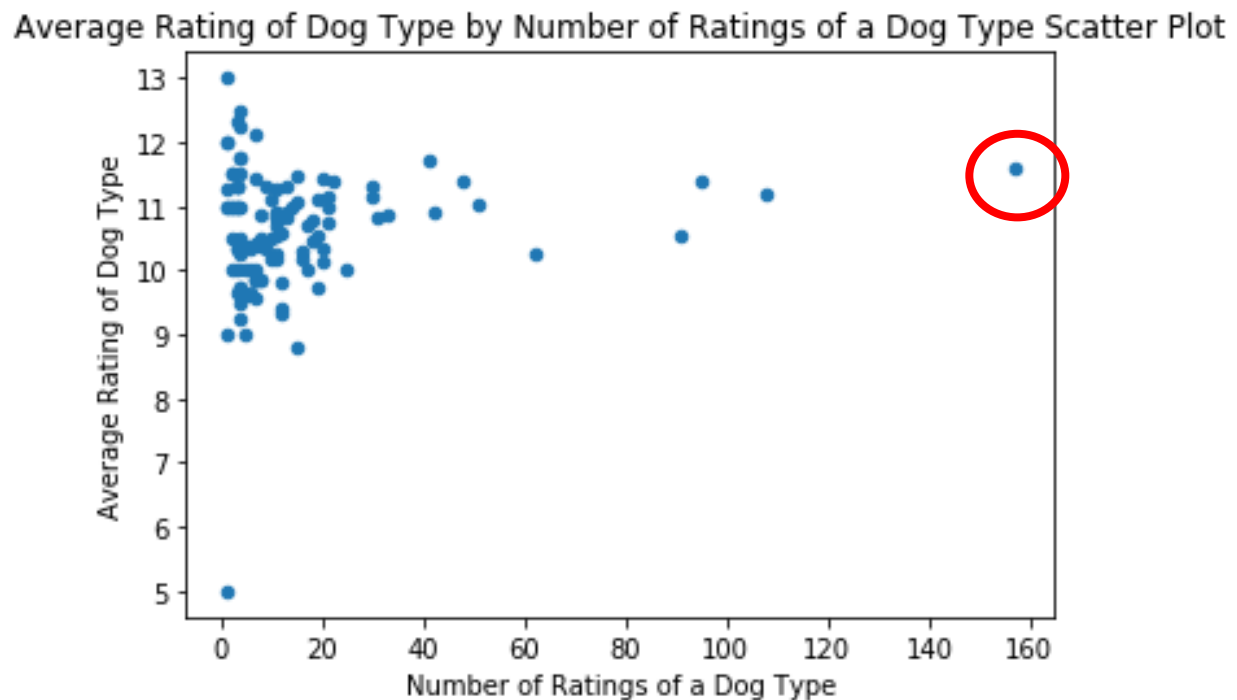


Figure 2

Strong Correlation between Favorites and Retweets:

Based on figure 3 the correlation matrix has a strong positive relationship between retweets and favorite counts (0.93). This suggest that as a tweet gains favorites, one can expect to see retweets to increase.

	tweet_id	rating_numerator	rating_denominator	favorites	retweets	user_followers
tweet_id	1.000000	0.093698	-0.030030	0.600701	0.374413	-0.3179
rating_numerator	0.093698	1.000000	0.972494	0.064033	0.049640	-0.0384
rating_denominator	-0.030030	0.972494	1.000000	-0.029111	-0.020822	-0.0014
favorites	0.600701	0.064033	-0.029111	1.000000	0.931396	-0.2006
retweets	0.374413	0.049640	-0.020822	0.931396	1.000000	-0.1384
user_followers	-0.317977	-0.038463	-0.001443	-0.200685	-0.138442	1.0000
user_favourites	0.023126	-0.001071	-0.008354	-0.005178	-0.010884	0.9285
year	0.879381	0.110515	-0.002658	0.515775	0.317409	-0.3087
month	-0.284839	-0.083318	-0.039875	-0.144325	-0.080050	0.1501
day	-0.009679	-0.006058	0.005061	-0.021099	-0.038263	-0.0052
rating	0.543096	0.236434	0.009506	0.407274	0.308235	-0.1635

Figure 3

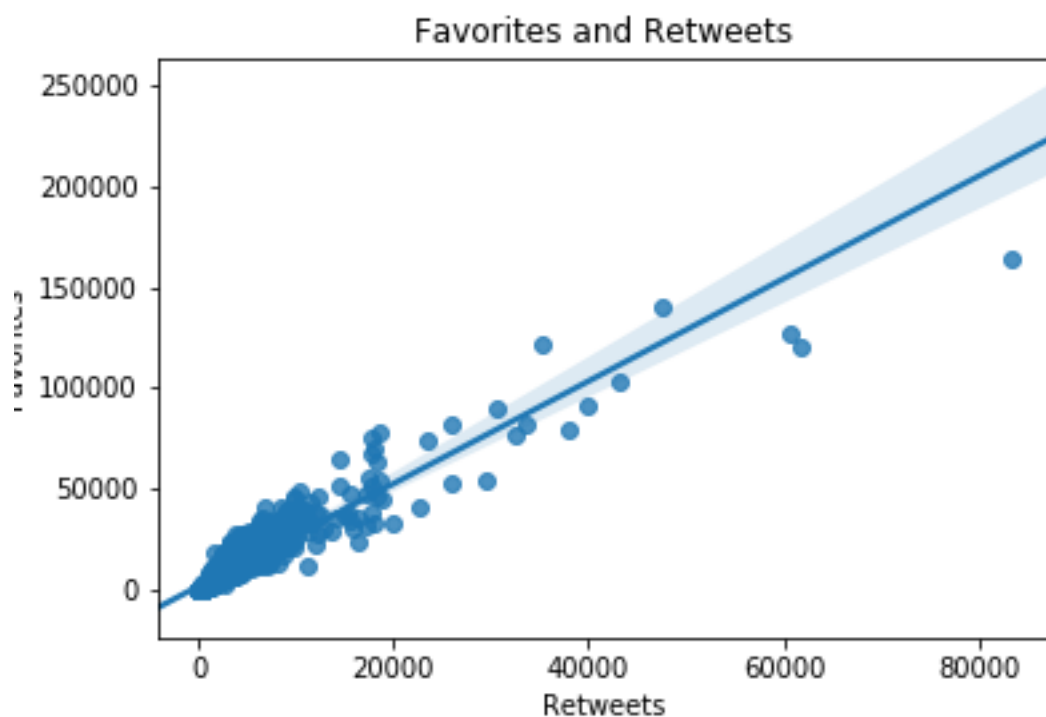


Figure 4

Retweet counts:

Many tweets have been retweeted more than 1000 times some more than even 50000 times. There is not a clear relationship between the ratings and the retweets.

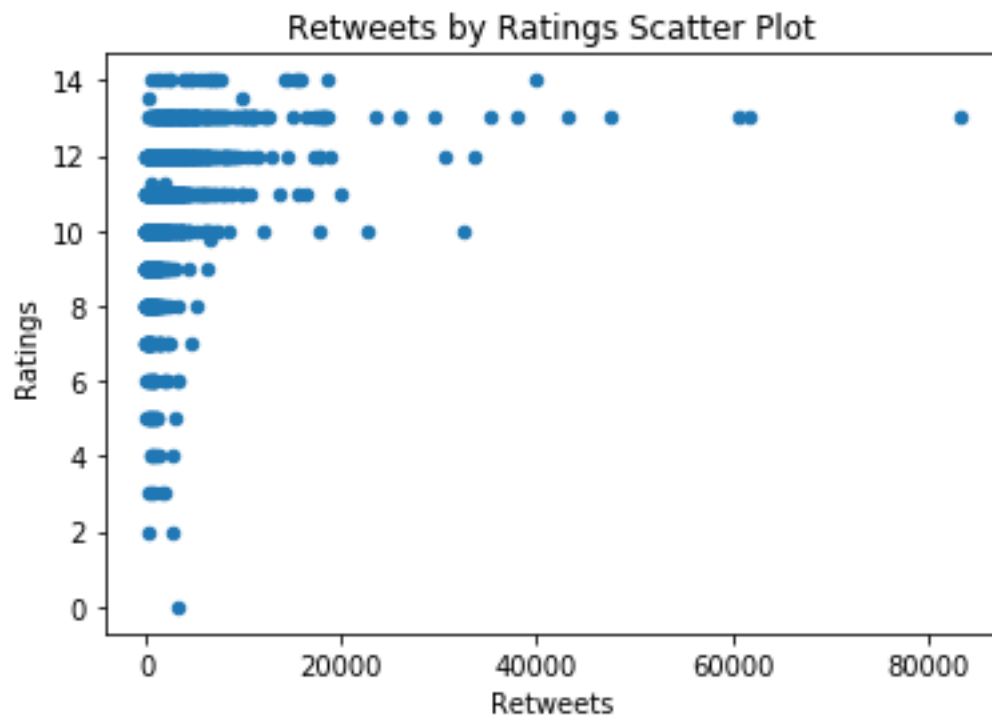


Figure 5