Data Analysis Report
We Rate Dogs
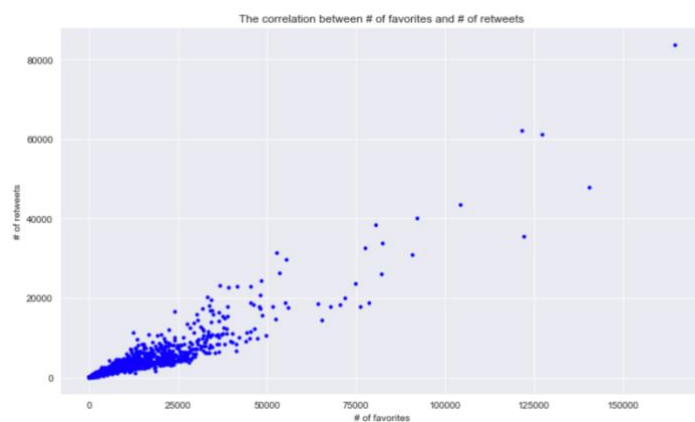
Prepared by: Maria Latysheva, September 2020

This report briefly describes my data findings based on the analysis and visualisation of the data contained in the cleaned dataset We Rate Dogs. This dataset contains information about entries in the We Rate Dogs Twitter account, including favorite count, retweet count, rating of a dog, breed, dog status, etc.

I analysed the dataset to answer the following questions:

1) Is there a strong correlation between retweets and favourite counts? Is there a strong correlation between the favourite counts and the rating?

2) What are the top ten favourite breeds?

3) Which dog status has the highest average rating?

4) What is the distribution of ratings? Which rating is most common?

1. The first question revealed the following insights:

There is a strong correlation between the # of retweets and the # of favorite counts with Pearson's correlation coefficient of 0.93. This may be seen on the following plot:
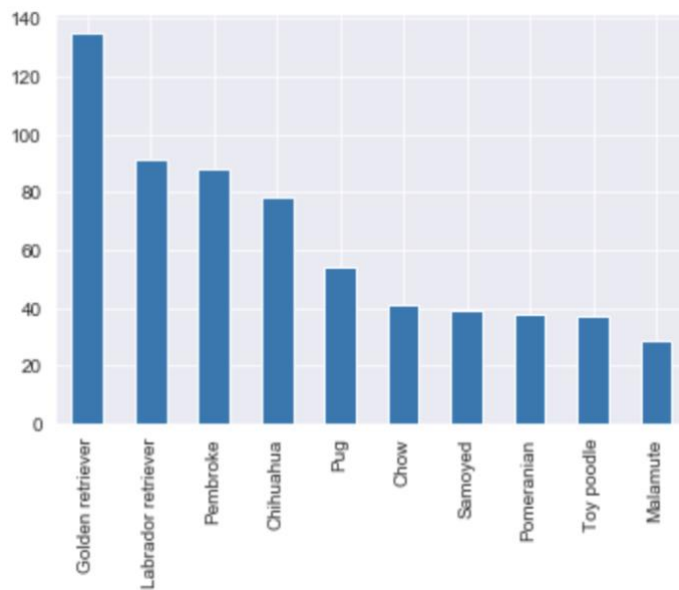

Surprisingly, but the correlation between the # of favorite counts and the rating of a dog is weak with Pearson's correlation coefficient of only 0.39. This is shown on the following graph:

2. The most favorite breed is Golden Retriever in terms of both frequency of occurrence and most favorited tweet.
   The distribution of the top ten breeds in terms of frequency of occurrence looks as follows:

3) The dog status (doggo, pupper, puppo and floofer) did not seem to be interesting in terms of analysis. They were indicated rarely and the frequency of all four statuses was about the same.

4) The most common ratings were 12, 10 and 11. Ratings below 10 were not frequent as the author tried to be nice to the dogs and their masters. The ratings by frequency may be seen on the following chart: