REPORT: ACT_REPORT

Introduction

The project was about the analysis of data in a Twitter account by the user name @dog_rate also called WeRateDogs. On the account, Twitter user rate dogs and provide comments on them. The rating has a denominator of 10 with a denominator of a number more than 10.

In this project, we were provided with two data sets, *twitter_archive_enhanced.csv* and *image predictions.tsv* data. The first data contained tweet details except for the favorite counts and the retweet counts. This data was extracted manually. The second data contained the details of dog images. A link was provided, hence, the data was programmatically extracted. Since the data from twitter_archive had retweets as well, we used the tweet_ids from the image-prediction data to filter our data having original tweets only. This was based on an assumption that the image is only attached to the original tweets.

The information contained in the two data sets was not enough to make informed insights on dog ratings, as such, we gathered a third data set from the Twitter account using the Twitter API.

The data was cleaned and the following questions were answered;

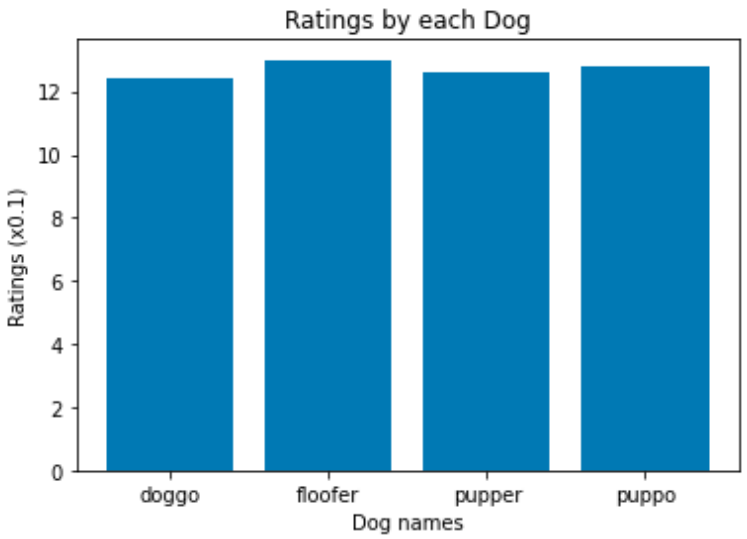
- Which dogs have the highest and the lowest average rating?
- What is the dog ranking based on the average favorite counts and the average number of retweets?
- Do the dogs with higher have higher favorite counts and number of retweets
- What is the effective percentage ranking of the dogs based on the favorite counts, number of retweets, and ratings? Does the ranking based on the effective percentage agree with the other rankings used earlier?

Results

In the project, found that Puppo is the most highly ranked dog name followed by popper and doggo is the least ranked one. We were not able to get a better rank for floofer because we had limited data to understand help us rank it. It only have one value.

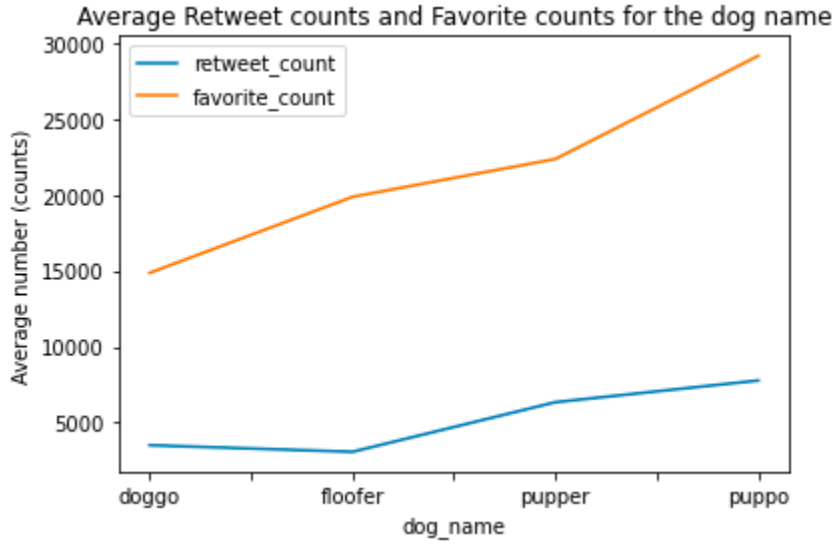
The bar chart and the table below shows the findings described above.

Dog_name	Rating over 10
Doggo	12.448276
Floofer	13.000000
Pupper	12.625000
Puppo	12.769231



We also found that the dog names with higher ranking had more average favorite counts and retweet counts compared to others. On the other hand, the dog name with the lowest ranking had fewer average favorite counts and retweet counts compared to others. The table and the graph below show the details.

dog_name	retweet_count	favorite_count
doggo	3497.793103	14864.551724
floofer	3061.000000	19881.000000
pupper	6334.708333	22379.000000
puppo	7765.230769	29184.230769



Finally, we came up with a metric that combined the ranking, the favorite count and the retweet counts. The new metric gave us results that were in agreement with the rankings of the dogs as done with the earlier metrics. The table below shows the results.

dog_name	retweet_count	favorite_count	rating_x0.1	Effective_percentage
doggo	16.931306	17.222525	24.483993	19.545941
floofer	14.816979	23.034736	25.569156	21.140290
pupper	30.663587	25.928995	24.831585	27.141389
puppo	37.588128	33.813744	25.115266	32.172379

The main limitation is that the scope of the data was limited to 2017 or later and most rows had no dog ratings. As such, we had very few data rows at the end of the cleaning. For instance, the floofer dog had one rating only in the final data. This does did not give us a sufficient result that can be replicated when the same report is done on a bigger data set.