We Rate Dogs

Data wrangle and analysis of dogs from Twitter

With more than 300 million monthly active users, Twitter is one of the most popular social networks worldwide.

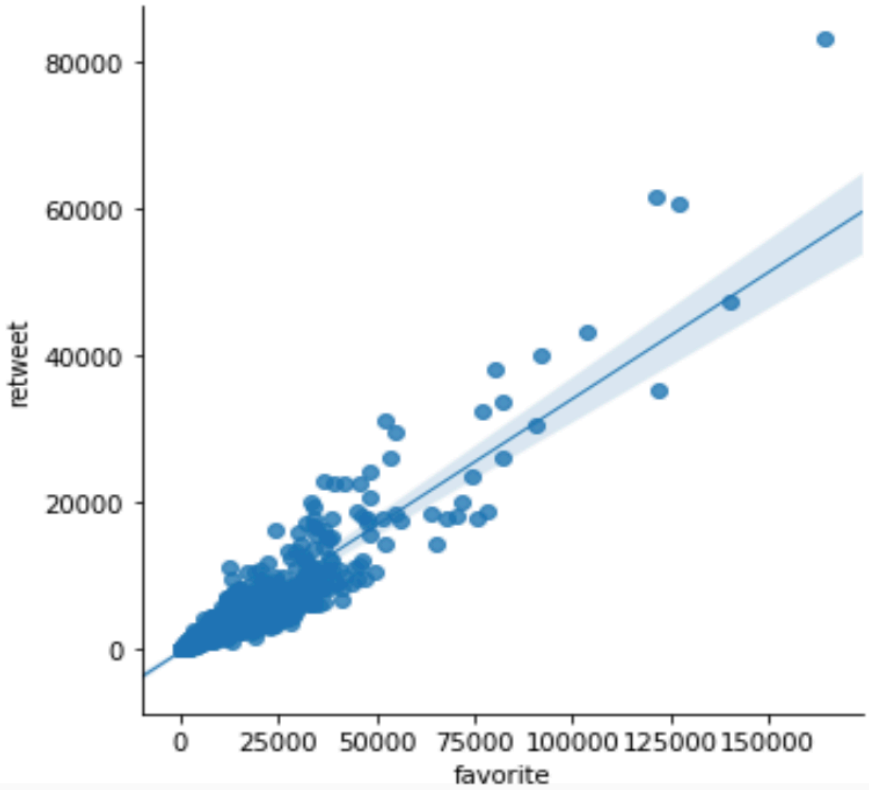
For this project, we want to gather the data from @dog_rates, which is a very popular Twitter account with over 7 million followers. The account gain its popularity by rating dogs worldwide and rating is based on faction (x out of 10).

About the data

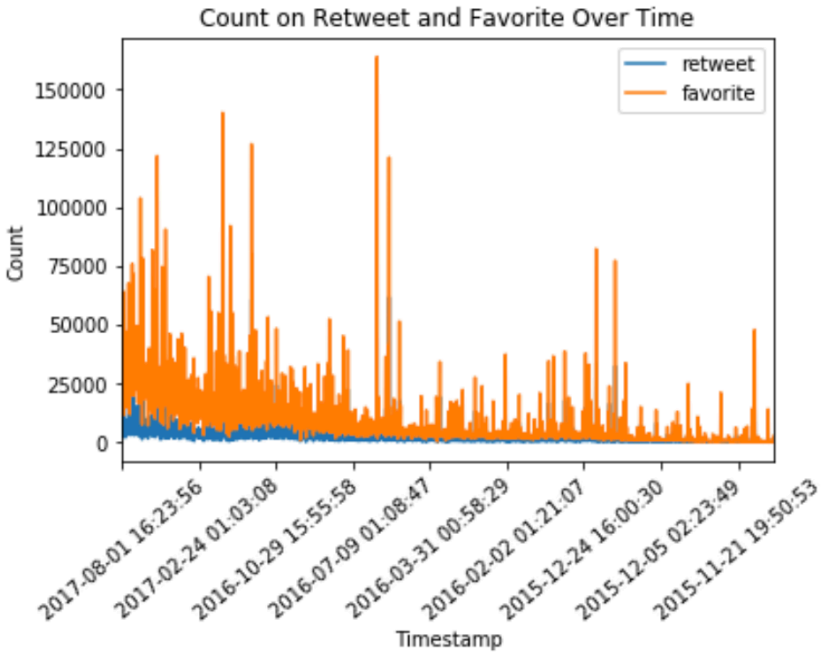
We got three sources for the project. First was provided by Udacity, an archive of past tweets in above account, containing id, sources, text description and so on. The second was about images predictions about the photos upload by users. The third one was retrieved from Twitter API, about the number of retweet and favorite, which were key metrics to define if this dog was popular! By combining and cleaning these sources, we get 2117 tweets as our data to finish analyzation.

Rating Over Time

First I observe the relationship between counts of favorite and retweets. Normally they are positive linear relationship. Here' the graph.

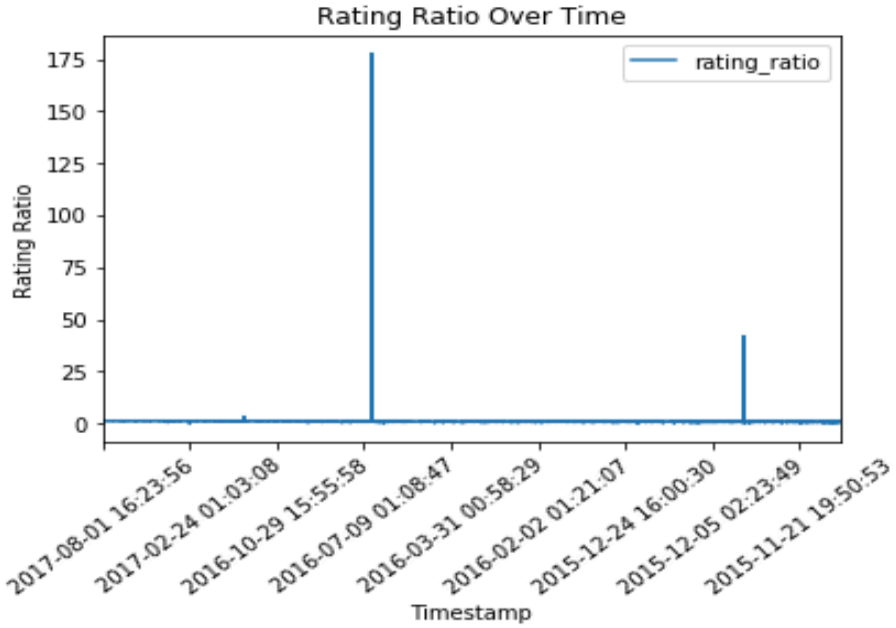


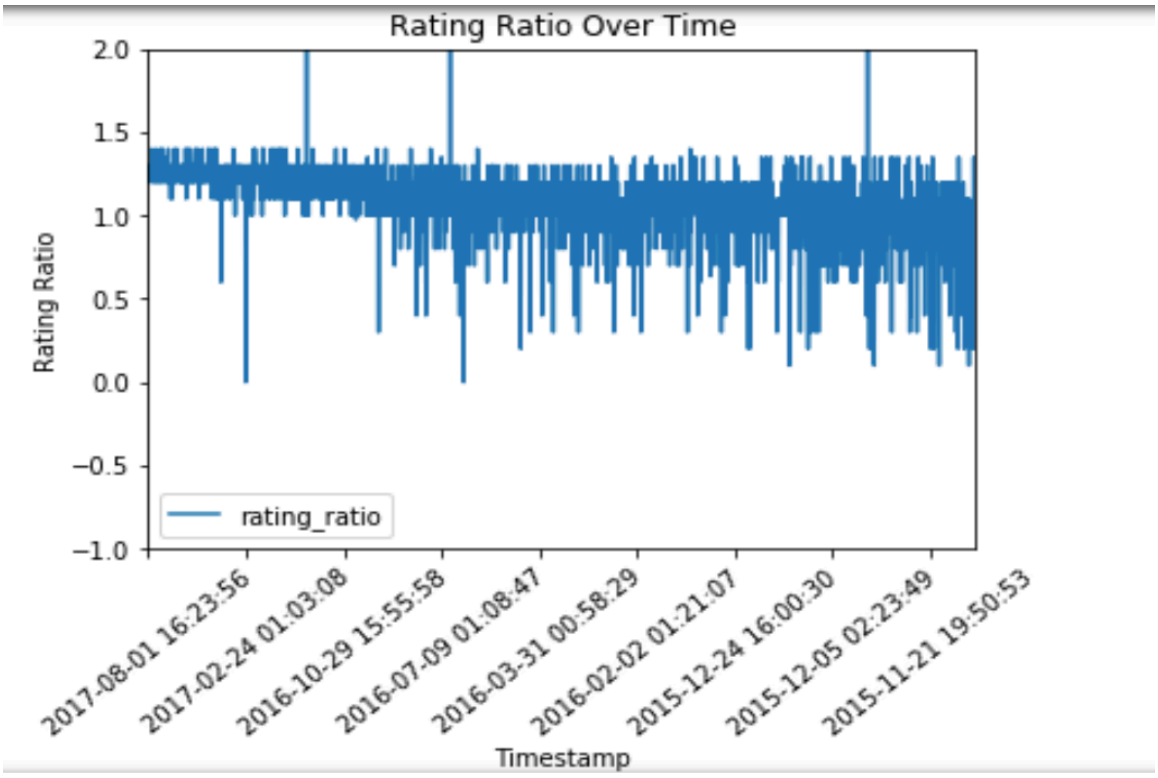
Since there is a column about time, which is also an important feature to determine the favor by users. I decide to take a look at the count on favorite and retweet over time.



The plot shows, the number of retweets is increasing by time, while not for favorite. It makes sense, as time goes, the counts on retweet increase. However, for counts on favorite, it depends on how many people like this dog. But in general, the counts on favorite is increasing, it means more and more people follow @dog_rates.

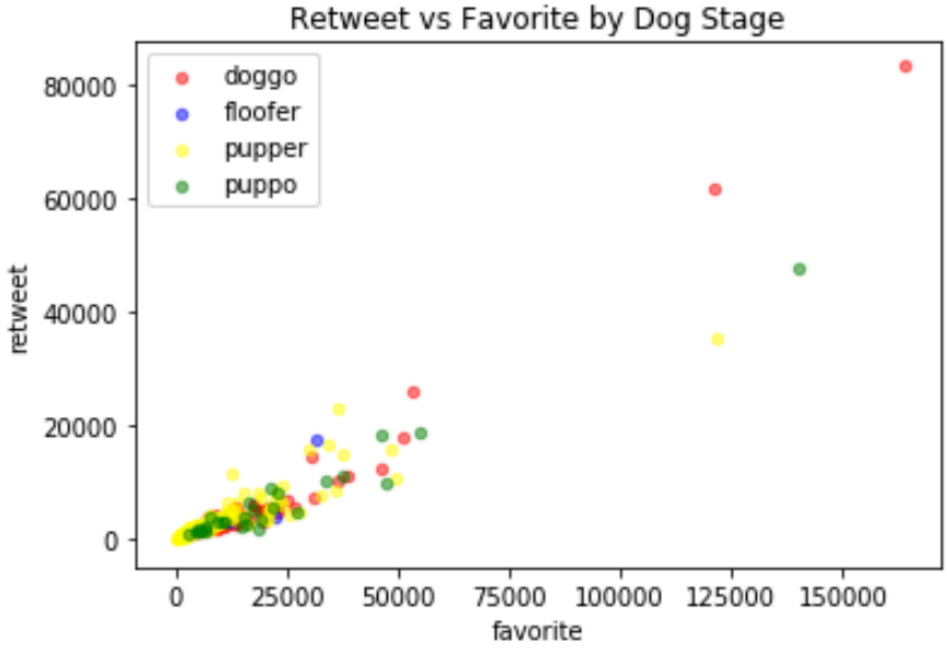
Also, we check the rating ration over time.





By Dog Stage

Now let' s plot the relationship between retweet and favorite again, but identify each stage (doggo, floofer, puppper, puppo).

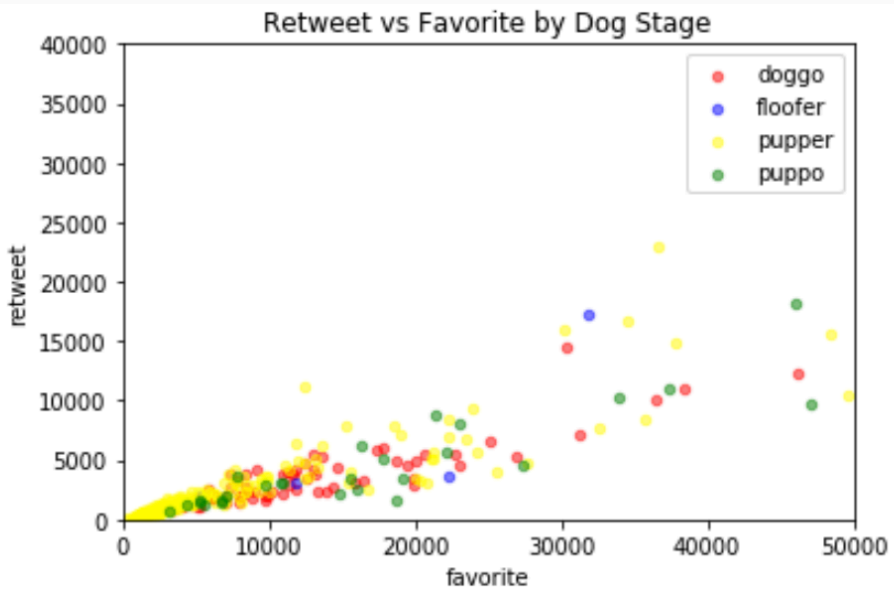



```
twitter_archive_clean.dog_stage.value_counts()
```

```
pupper      242
doggo        81
puppo        29
floofer       4
Name: dog_stage, dtype: int64
```

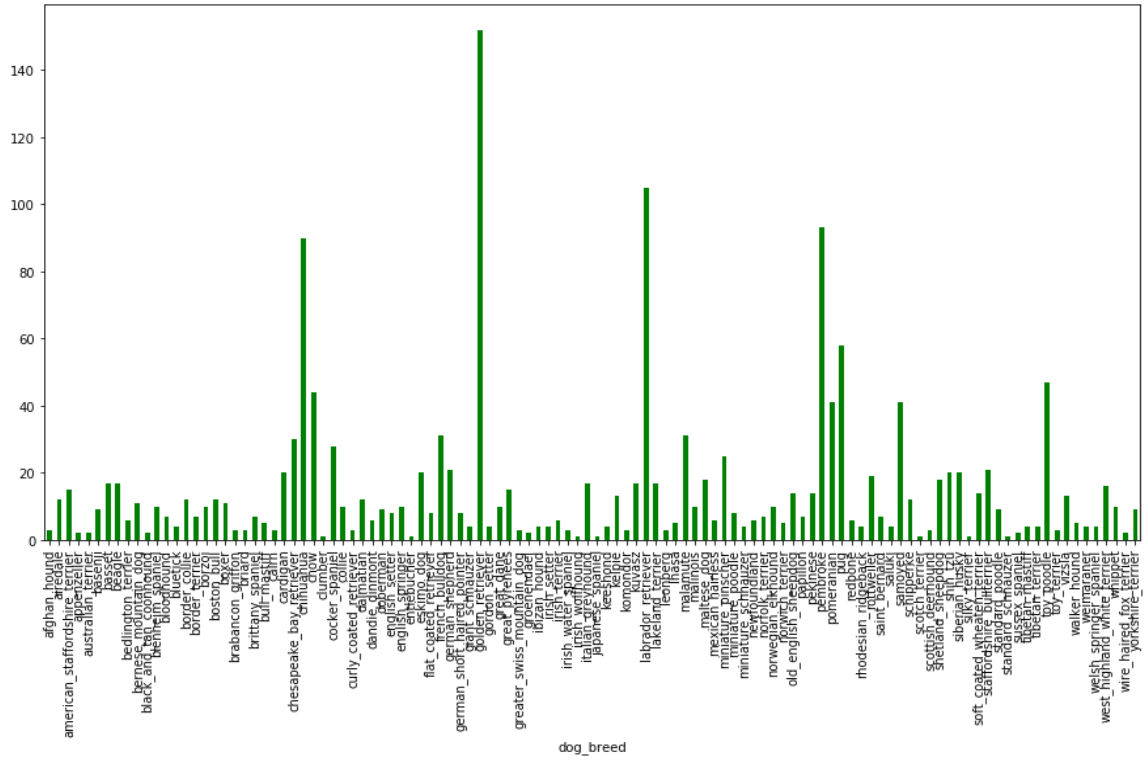
Pupper is most popluar, tho there are only 356 records contain the stage info. And still, the relationship between retweet and favorite is positive liner.

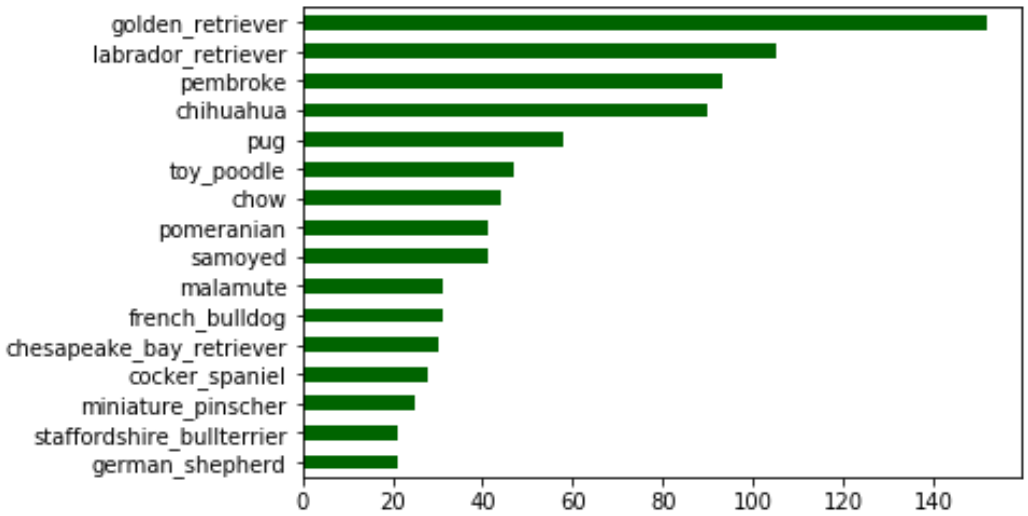
Let' s take a closer look at the graph.



Most Popular Breed

We are lacking information about the breed, so based on three prediction, I use 1st prediction as primary, if it' s not defined as dog, use the 2nd prediction, otherwise, use the third one. In the end, we get 1624 records out of 2117. That makes the analysis more convinced. From the graph, gold retriever, labrador retriever, Pembroke (also known as corgi) and chihuahua are the top 4 popular breed among all 112 breeds in the dataset.


dog_breed


golden_retriever
labrador_retriever
pembroke
chihuahua
pug
toy_poodle
chow
pomeranian
samoyed
malamute
french_bulldog
chesapeake_bay_retriever
cocker_spaniel
miniature_pinscher
staffordshire_bullterrier
german_shepherd
0
20
40
60
80
100
120
140