Analyzing and Visualizing WeRateDogs

Introduction

This project aims to get insightful information about dog ratings from the twitter page WeRateDogs while demonstrating advanced data wrangling and visualization techniques using various Python libraries. WeRateDogs is a community page on twitter designated to rating dogs on their appearances and stories, and which was formed by the user @dog_rates in 2015. The page has since grown extremely in popularity, with many users sharing its content and requesting their dogs being rated as well. The site is a driving force for the development of the 'dog culture', with it's famous terms like "pupper", "mlem", "floof" etc. They also developed their own unusual rating system over time, in which almost every dog is rated above 10/10, because "they're good dogs".

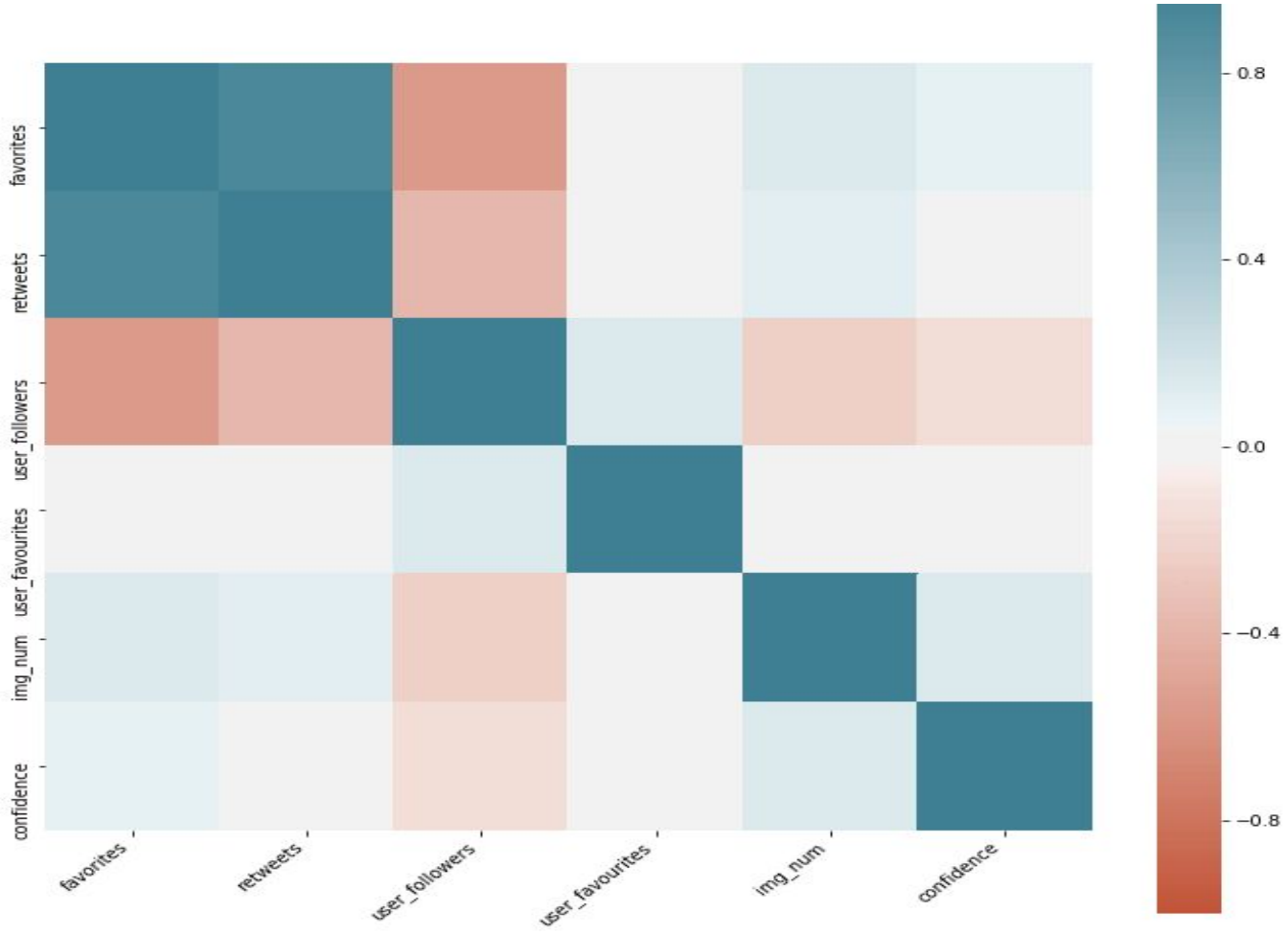
WeRateDogs has over 6 million followers and has received international media coverage. One of those coverage was about the quote “They’re good dogs Brent”, this was an exchange in which WeRateDogs shut down a person having an issue with its rating system in humorous ways.
So what do we see in twitter documents? How are they distributed in their stages? Most accounted Breed of dogs? Looking to breed with favorite and retweet? And looking at dogs name. These are the types of questions we find answers off.

So what do we see in twitter documents?

With the newly created dataset we try to plot correlation matrix with its features and here is what we find off.

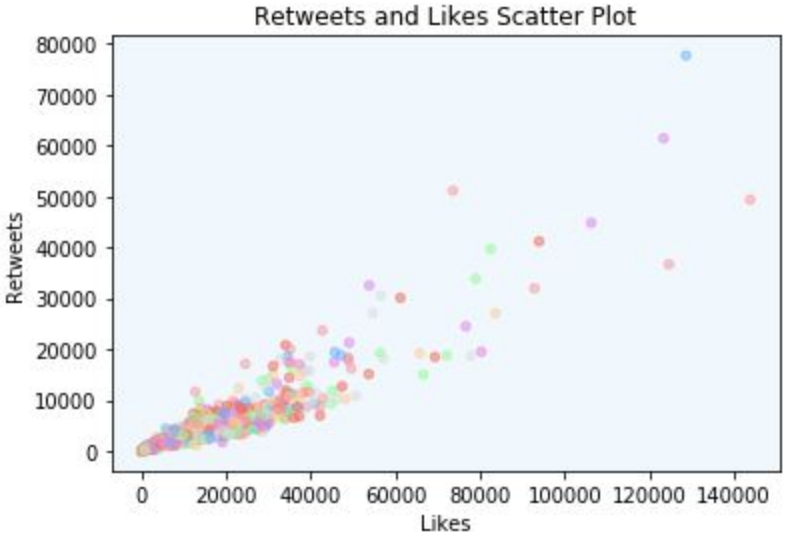
- There is strong correlation between favorites and retweet(more favorites mean more retweets)
- Followers and retweet have negative correlation(-.4) which means they opposite in normal prediction
- Rating_numerator doesn't relate with any features

Look at the diagram below.



Tweet Retweet Vs Favorites

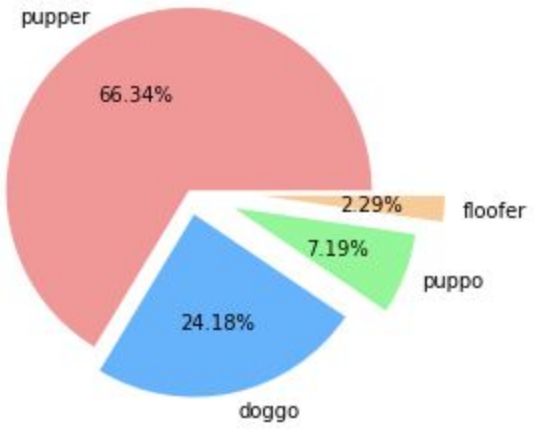
From the above graph we saw that there is a strong positive correlation between Tweet Retweet Vs Favorites and we will try to plot it.



high and positive correlation b/w Favorites and Retweets

Dog stage count:

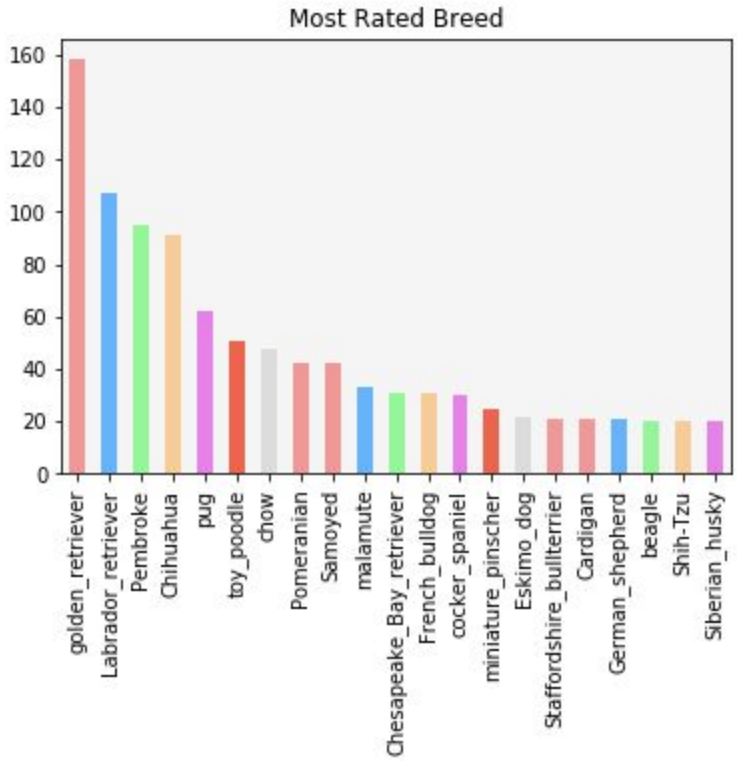
Here we will see the Dog stage distribution of dogs in a newly created dataset.



Here Pupper is leading with a huge gap to second highest one.

Most accounted Breed

We see all breeds of dogs in a dataset where the dog count of all breeds is greater than 20.



We must know any breed is equally important. However, through the data golden_retriever most accounts breed. Followed by Labrador.

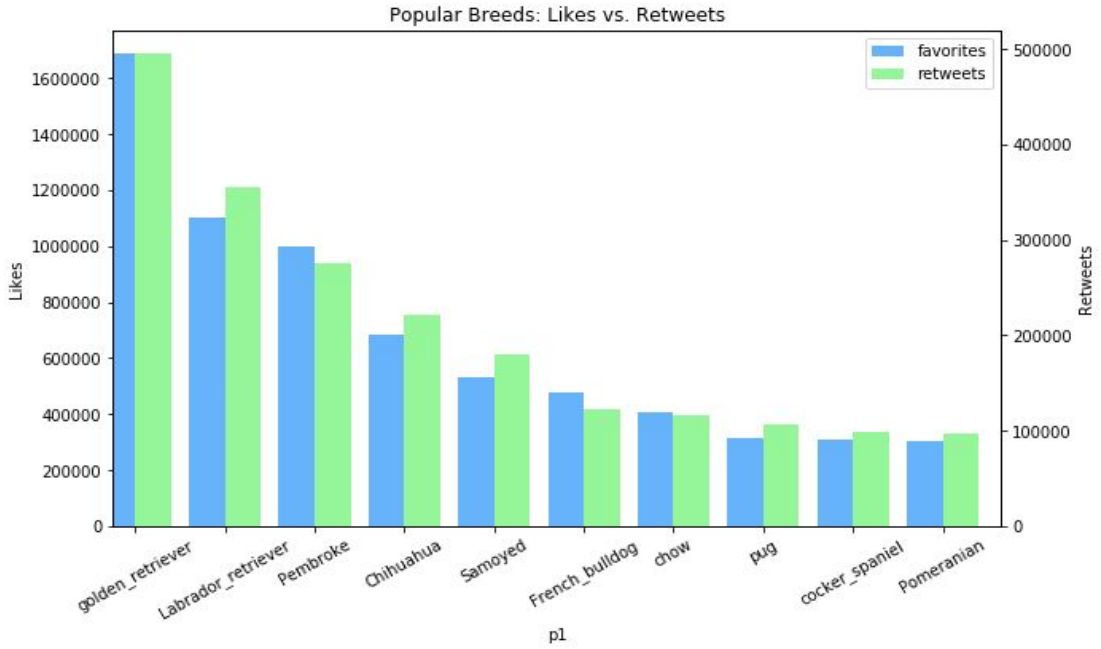
Looking to breed with favorite and retweets using prediction

```
p1
golden_retriever       1687262
Labrador_retriever     1099245
Pembroke                996938
Chihuahua               682941
Samoyed                 533688
French_bulldog          478019
chow                    408406
pug                     314450
cocker_spaniel          310251
Pomeranian              304159
Name: favorites, dtype: int64
```

```
golden_retriever       495259
Labrador_retriever     355794
Pembroke               276009
Chihuahua              221340
Samoyed                179938
French_bulldog         123170
chow                   116442
Pomeranian             106507
cocker_spaniel          98001
pug                     96722
Name: retweets, dtype: int64
```

We know that most accounted breeds are golden retrievers and favorites too.

Same is with cases with retweet of most retweet dog breed so match with most favorite and most account.



And Well in terms of name :
Oliver, Cooper, Charlie, and Lucy is most popular name

Conclusion

- WeRateDogs (@dog_rates) shares beautiful pictures of dogs with adorable poses and expressions. These are rated on a scale of one to ten, however in most cases the ratings were much much higher with max '99/90' and min '0/10'. Most rated were '12/10' and '10/10'.
- Most popular Dog names are Oliver, Cooper, Charlie, and Lucy.
- Most popular breed of dog is golden_retriever and second Labrador_retriever.
- Most of these dogs were in pupper stage.
- And its highly likely that more follower means more retweets.