Wrangle and Analyze Data

A report on data analysis of @weratedogs twitter data

As a part of this project we will be wrangling data from WeRateDogs which is a Twitter account that rates people's dogs with a humorous comment about the dog. We have 2 different type of data available:

1. Data to be downloaded from twitter via APIs which contains retweet and favorites count
2. Image predictions data where the dog pictures have been worked on via neural networks to figure out the dog breed.
3. Data with information about tweets such was what source was used, when was the tweet made etc

After cleaning the data using various wrangling techniques and analyzing it, I created following visualizations to answer some of the questions. Questions that I worked on are:

1. Which dogs are more popular? Retweet count and favorite count should give some information about which dog breeds are more popular. We will assume that the P1 prediction about dog breed is correct.

2. Which stage of dog is more popular? We can use the stage and retweet/favorites count.

3. How does the trending of retweet and favorite counts over time look like?

4. Which platform is used by twitter users more for dog related tweets (and possibly for other tweets too)?

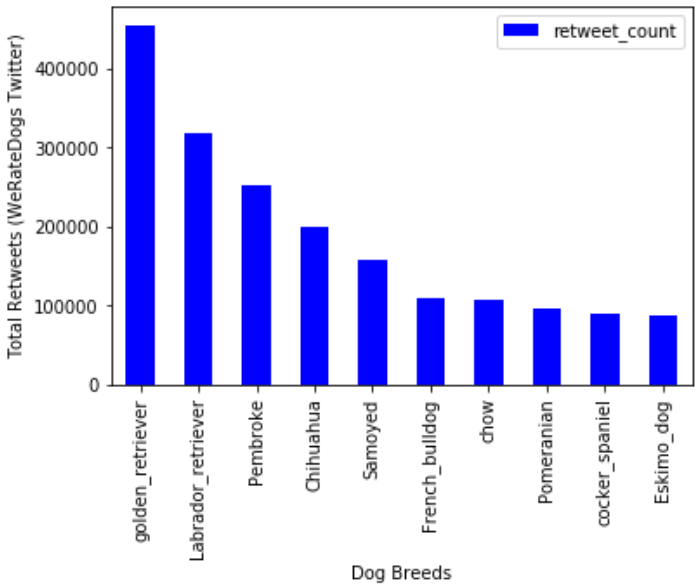
5. Which dog stage and breed receive highest ratings?

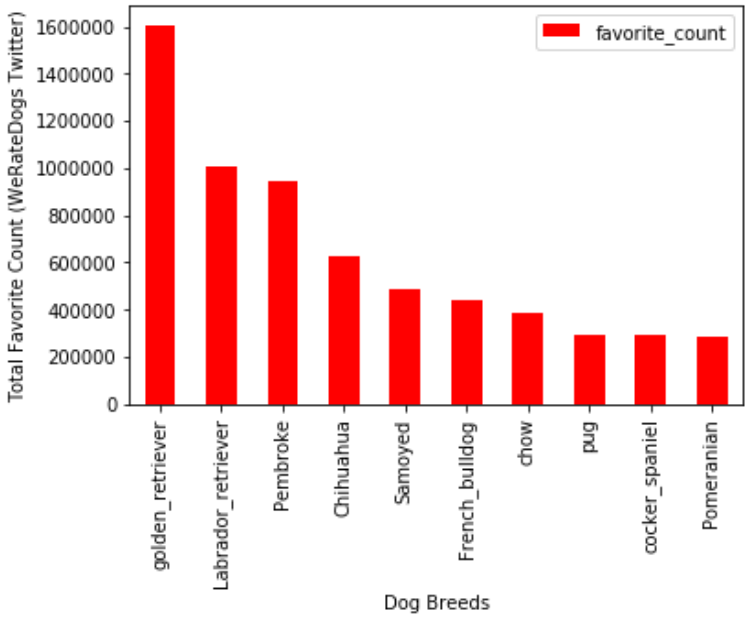
6. Which are the most popular names for the dogs?

Using data visualization using simple matplotlib plots I was able to arrive at answers to all the above questions. In the following few pages you will be able to see the visualization and conclusions.

Popular Dog Breed

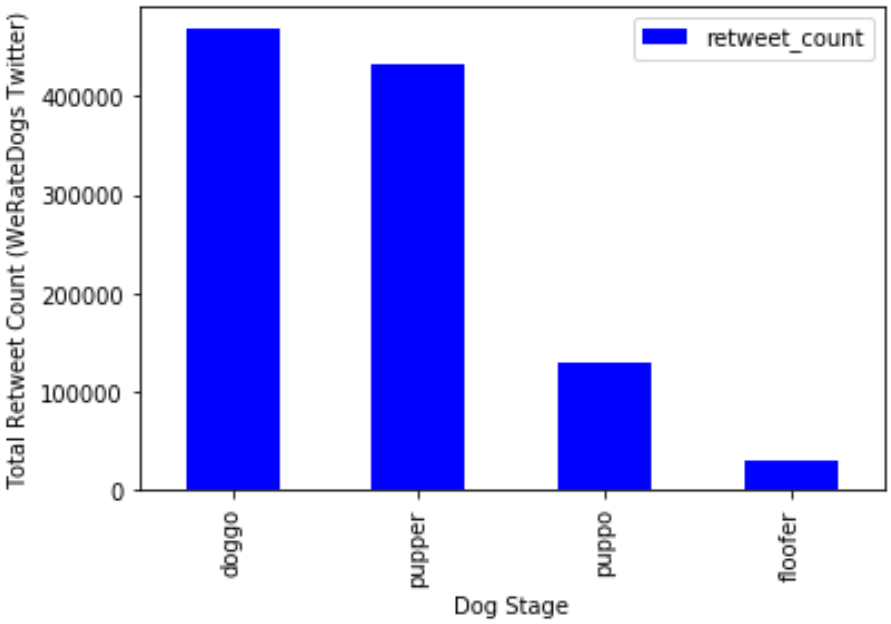
Following chart shows us what are the most retweeted about dog. This show that most popular dog breeds are **Golden Retriever, Labrador Retriever and Pembroke**. There are many dog breeds in the data, however, I used the top 10 breeds to plot this chart. Similar analysis was done using favorites count data too, and conclusion about popular dog breed didn't change.

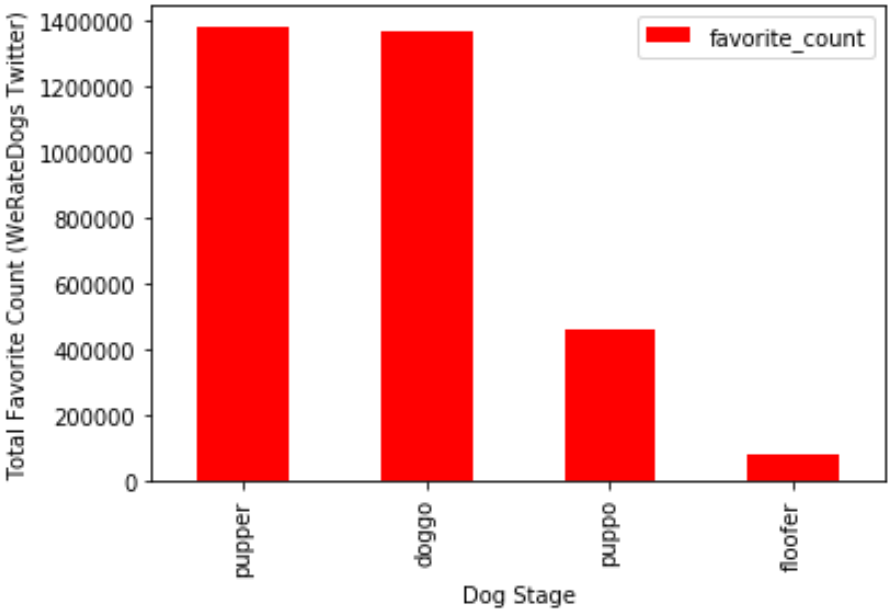




Popular Dog Stage

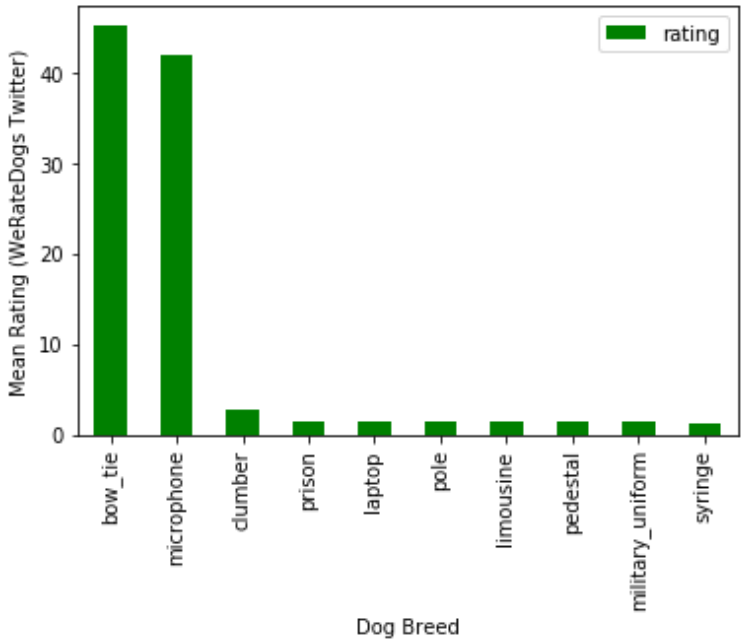
The next is about the popularity of dog stage. I used the retweet count as a measure of popularity again and this time the analysis showed that Doggo and Pupper stages are most popular





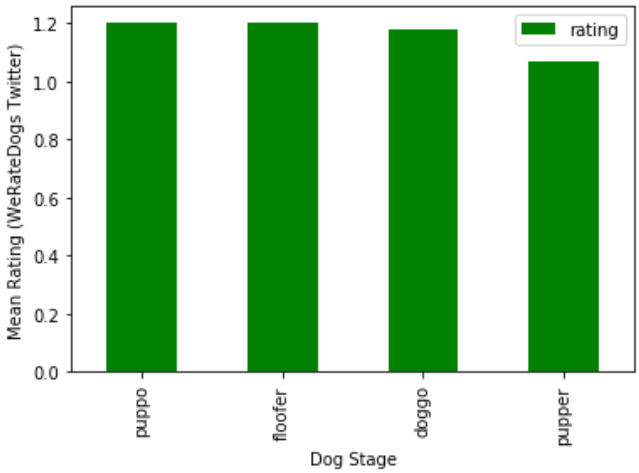
Dog Breed with highest ratings

Some dog breeds get high ratings than others. The rating system is unique here when people rate their dogs out of 10, but often times their rating is more than 10. Ratings were normalized to decimal values before plotting for comparative analysis. The chart below shows that bow tie and microphone as the highest rated dog breeds. However, the different with other breeds is very huge hear and it could be due to anomaly in the collected data.



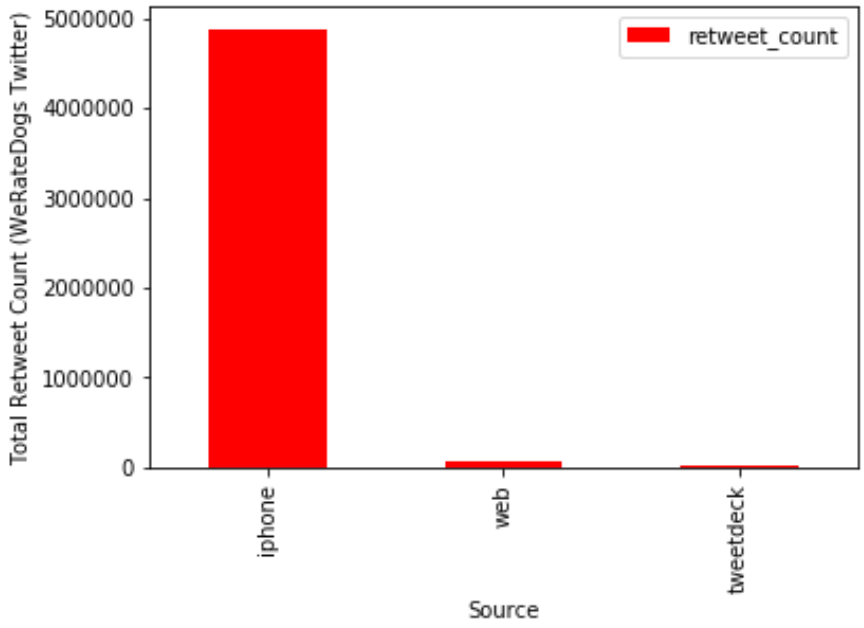
Dog Stage with highest ratings

Similar analysis was done for dog stage to understand which stages get highest rating. The chart obtained shows there is not significant difference between dog stages when it comes to ratings. Hence, all stages have pretty much similar rating. However, there were many records where the dog stages were available which can make the conclusion differ from reality.



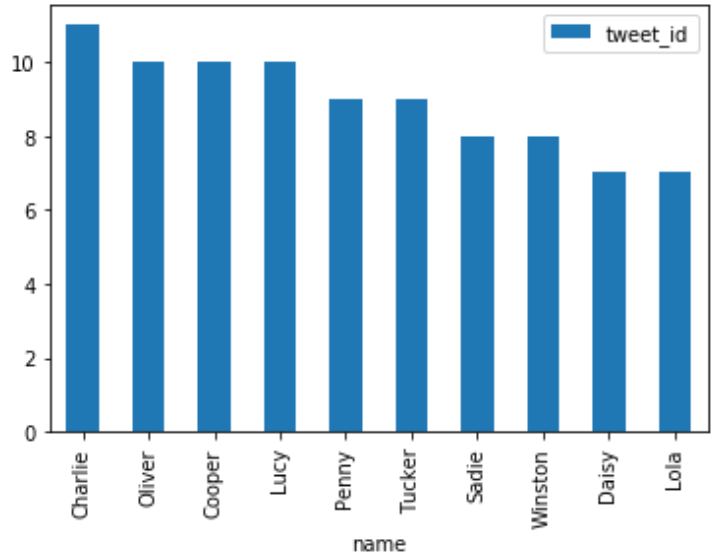
Highest used source for Tweeting

This shouldn't come as a surprise. I found that the highest proportion of tweets originated from Iphone. Compared to Iphone other sources were very rare. Iphone is undoubtedly used as the biggest source of tweets on @weratedogs twitter account (and probably on twitter in general).



Most popular dog names

Analysis on tweets with dog names on it tell us what the top 10 dog name are. However, many tweets have missing dog names. So, the reality can be different from what we could concluded from this data.



Trending of retweet and favorite counts over time

Lastly, I wanted to check the trend over time for twitter use using retweet and favorites count. As is very evident from the chart below and our general experience of the world, online platforms are becoming more and more popular day by day. Compared to 2015, in 2017 there are more favorite counts and more retweets. The trend seems to do upwards and will continue to increase in future.