



Data Analysis and Visualization of tweets for user @dog_rates

Report Created by

Avinash Korwarkar

as a course requirement for [Udacity's Data Analyst Nanodegree](#) program

Disclaimer:

The dog breeds assumed in this analysis are image predictions from a neural network and we are going to assume these predictions are correct for the purpose of this report.

Analysis 1: Distribution of dog breeds by tweet count

We are able to see that there are 369 different breeds of dogs predicted by the neural network. (Assuming all of the predictions are for dogs and not other things.)

- Let us now see which of these dog breeds have the most number of original tweets.

golden_retriever	137
Labrador_retriever	89
Pembroke	85
Chihuahua	78
pug	51
toy_poodle	38
Pomeranian	38
chow	37
Samoyed	36
malamute	29
cocker_spaniel	26
French_bulldog	23
Chesapeake_Bay_retriever	22
miniature_pinscher	21
seat_belt	21
German_shepherd	20
Siberian_husky	20
Staffordshire_bullterrier	19
Eskimo_dog	18

Fig: Tweet Count of tweets for various breeds of dogs

- We can also look at the above data visually to get a better idea of the tweet distribution by dog breed.

Visualization 1: Distribution of dog breeds by tweet count

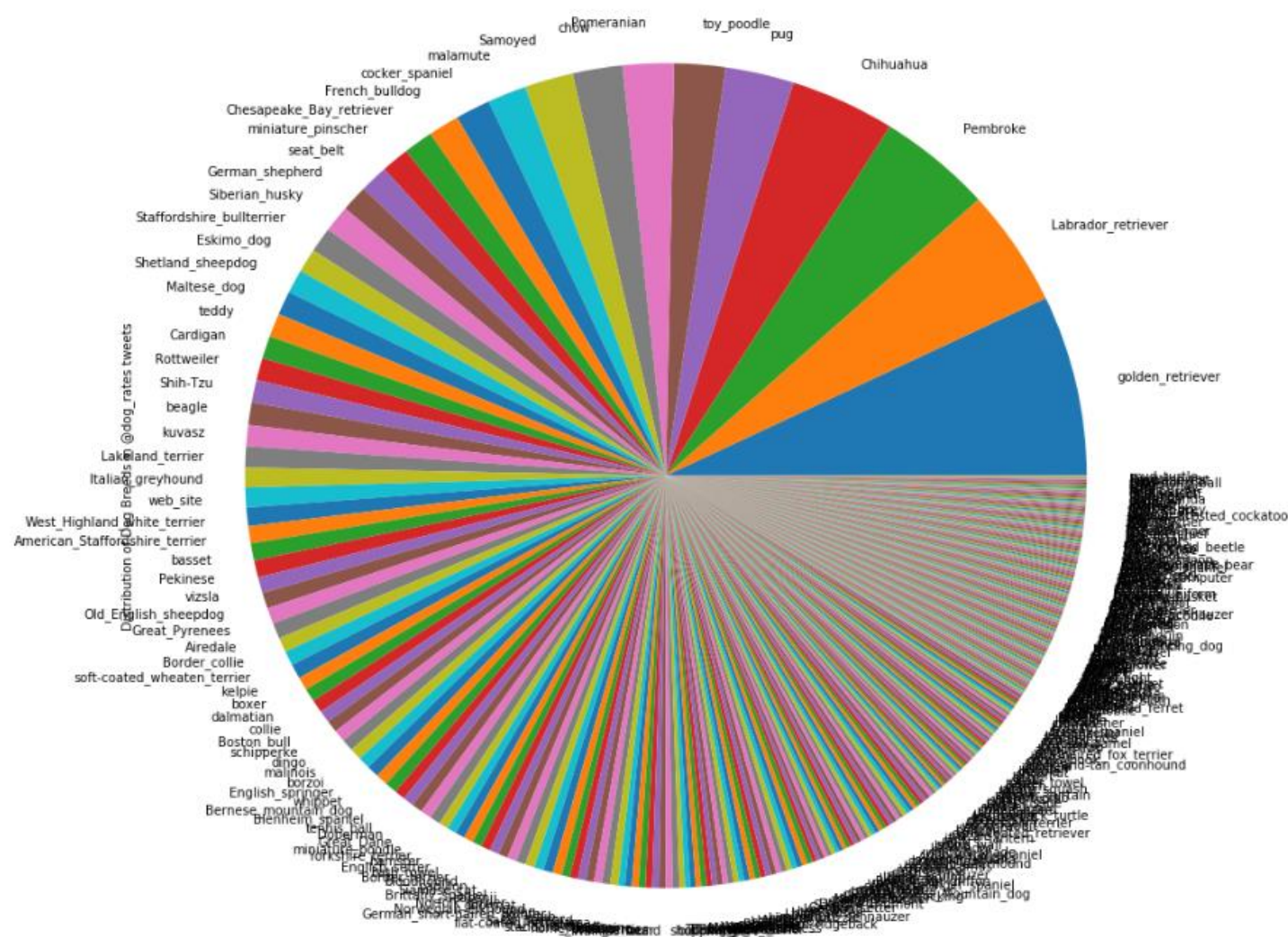


Fig: Pie-Chart showing distribution of tweet count by breed of dog.

Observations:

It looks like Golden Retrievers have the most original tweets (137), followed by Labrador Retrievers (89), Pembrokes (85) and then Chihuahuas (78).

It could be because a lot of people do own Golden Retrievers or Labrador Retrievers as compared to other dogs.

Analysis 2: Highly Rated dog breeds

There were several tweets with multiple dogs and the ratings were multiplied by the number of dogs. So, we had to average out the ratings for each dog. Below are our findings on the average rating ratio (rating numerator / rating denominator) for each dog breed.

predicted_breed	
clumber	2.700000
military_uniform	1.400000
limousine	1.400000
laptop	1.400000
pedestal	1.400000
prison	1.400000
marmot	1.300000
restaurant	1.300000
syringe	1.300000
teapot	1.300000
loupe	1.300000
cliff	1.300000
tiger_shark	1.300000
canoe	1.300000
orange	1.300000
upright	1.300000
quilt	1.300000
barbell	1.300000

Fig: Average Rating Ratio for various breeds of dogs

Observations:

And the award goes to.... Drum roll.....

Clumber..... 27/10

Clumbers have the highest rating followed by military uniforms, limousines, laptops, pedestals and prisons – all at **14/10**.

Analysis 3: Distribution of dog breeds by retweet count

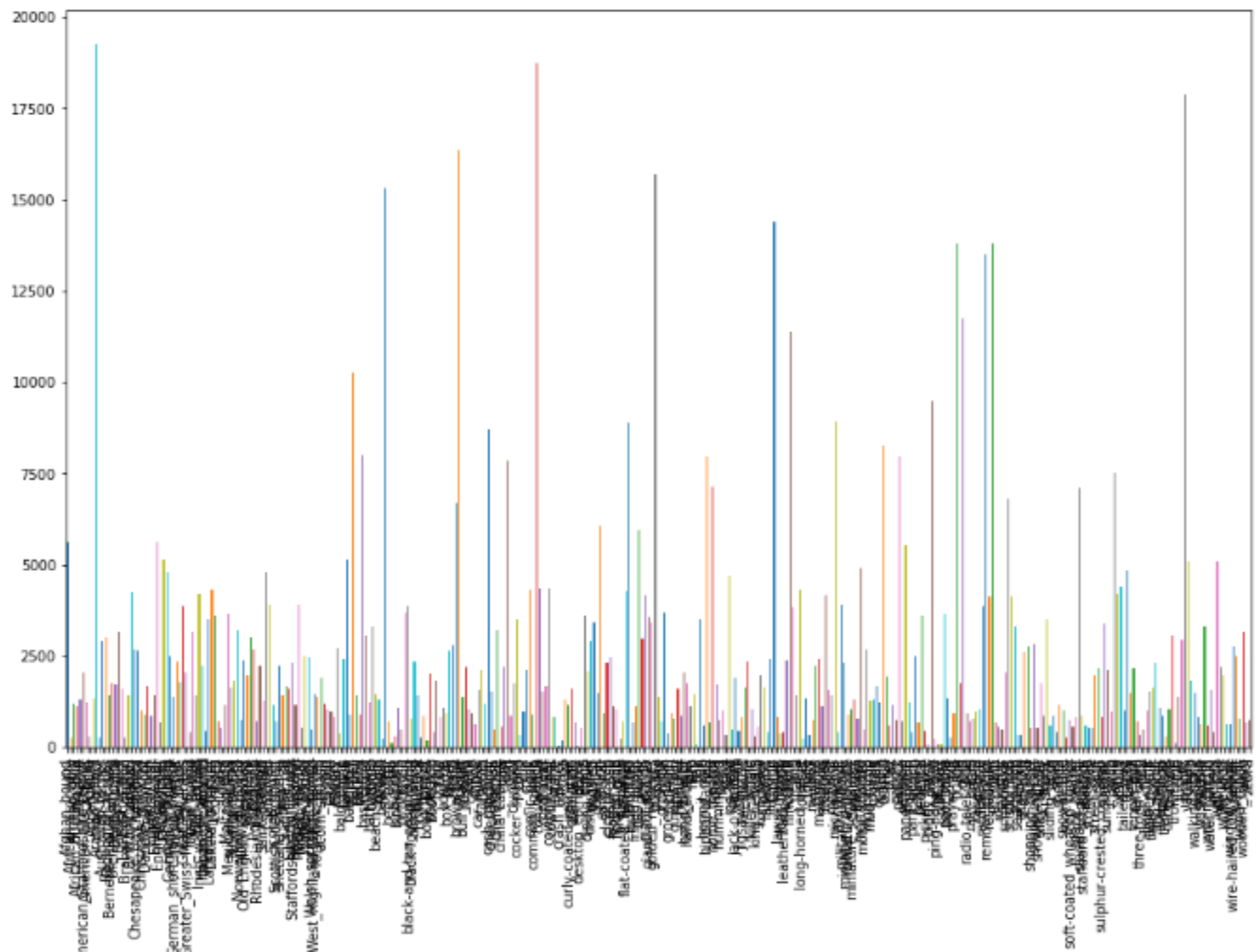
- Retweet Count is a big measure of how much a resonated or piqued people's interest or curiosity. We can try to see which of the breeds of dogs had the most average retweet count.

predicted_breed	
Arabian_camel	19246.000000
conch	18744.000000
upright	17898.000000
bubble	16354.000000
gondola	15695.500000
beaver	15316.000000
laptop	14377.000000
prison	13798.000000
revolver	13796.000000
remote_control	13514.000000
quilt	11758.000000
limousine	11373.000000
barbell	10254.000000
pillow	9489.000000
microphone	8909.000000
fountain	8903.000000
cash_machine	8697.000000
orange	8259.000000

Fig: Average Retweet Count for various breeds of dogs

- We can also look at the above data visually to get a better idea of the average retweet distribution by dog breed. Word of caution – There are a lot of breeds and the visualization is going to be crowded.

Visualization 2



Observations:

American Camels had the highest average retweet count, followed by Conchs, Uprights and Bubbles. There must be definitely something interesting about these breeds in those tweets.

While these are some genuine tweets about American Camels, Conch, Uprights and other breeds, the results of a few other breeds are very interesting. People also tweeted back about prisons, revolvers, remote_controls, microphones, fountains and cash_machines.

Suppose we assume the neural network is NOT predicting correctly, we can very well see why people were wondering why someone would tweet about these things on a dog ratings twitter page.

Analysis 4: Distribution of dog breeds by favorites count

- Favorites Count is a measure of how much people “like”d a certain tweet. Let us try to see which of the breeds of dogs had the most average favorite count.

predicted_breed	
laptop	64878.000000
limousine	45239.000000
conch	45038.000000
fountain	43677.000000
Arabicamel	40436.000000
bubble	39248.000000
orange	37852.000000
beaver	37271.000000
prison	34141.500000
revolver	34017.000000
basketball	33833.000000
barbell	33823.000000
upright	33575.000000
quilt	31927.000000
remote_control	29211.000000
gondola	26495.500000
microphone	24837.000000
paper_towel	23810.500000
cat	22100.000000

Fig: Average Favorites Count for various breeds of dogs

Observations:

Laptops take away the cake with 64878 average favorite count and are the most “liked” dog breeds on the @dog_rates twitter page. They are almost 20,000 favorites higher than the next highest Limousines. There are also Conchs, Fountains, American Camels, bubbles and oranges which are not too far in the popularity list.