Act Report
Kyle Santana
3/27/2018

Visualization and Insights of We Rate Dogs Twitter Data

I combined and analyzed WeRateDogs twitter data from the three sources below:

1. The WeRateDogs Twitter archive (twitter_archive_enhanced.csv) provided as a csv file from Udacity.

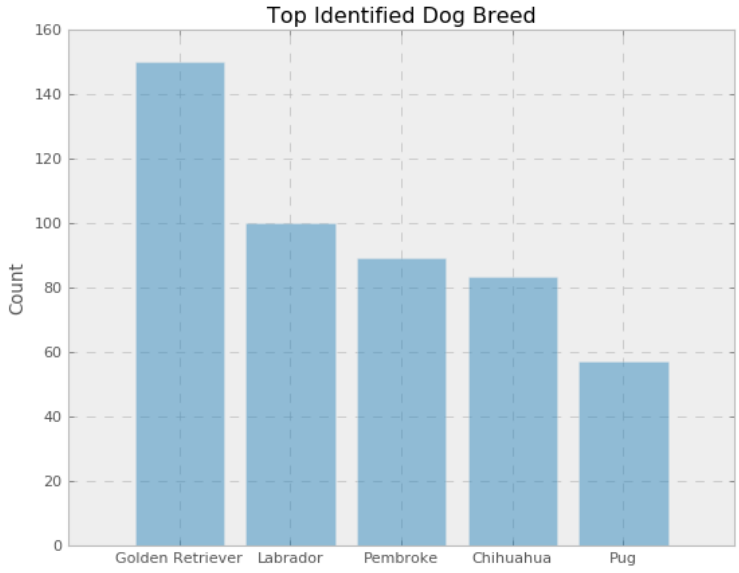
2. The tweet image predictions, i.e., what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. The file (image_predictions.tsv) was downloaded programmatically from Udacity's servers.

3. Using the tweet IDs in the WeRateDogs Twitter archive, I queried the Twitter API for each tweet's JSON data using Python's Tweepy library and stored each tweet's entire set of JSON data in a file called tweet_json.txt file.

After cleaning the data I found the following insights below:

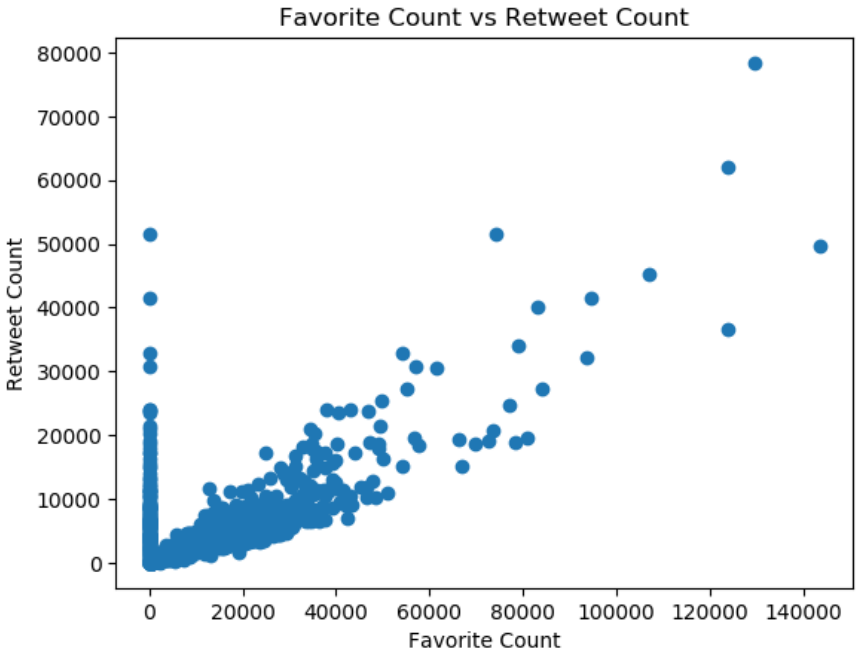
Insight 1

According to the breed classification by the Neural network the top five identified breeds were Golden Retriever, Labrador, Pembroke, Chihuahua, and Pug. What was also interesting among the five top breeds is the average of the favorite and retweet count were near the top of the list as well. While this does not conclude that if you have a top identified breed you will receive a high number of retweets and favorites, it does warrant further study to look at the relationship between breed and favorite/rewet count.



Insight 2

Looking further into favorite count vs retweet count to determine if there is any type of relationship that can be deduced. I decided to use a scatter plot to plot all the data points using the x axis as the favorite count and the y axis as the retweet count. As we can see from the graph below there appears to be a positive linear relationship between Favorite and Retweet count.



Insight 3

Which dog had the most favorites? To come up with the answer I sorted the entire dataframe by favorite count by largest and only kept the first three rows. I then found the dog below with over 143,000 favorites.

Insight 4

Which dog had the most retweets? To come up with the answer I sorted the entire dataframe by retweet count by largest and only kept the first three rows. I then found the dog below with over 78,000 retweets.