Analyzing, and Visualizing

Author: Gour Bera

Date: 1st-Jul-18

Introduction

The dataset that I am be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

WeRateDogs downloaded their Twitter archive and sent it to Udacity via email exclusively for us to use in this project. This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017.

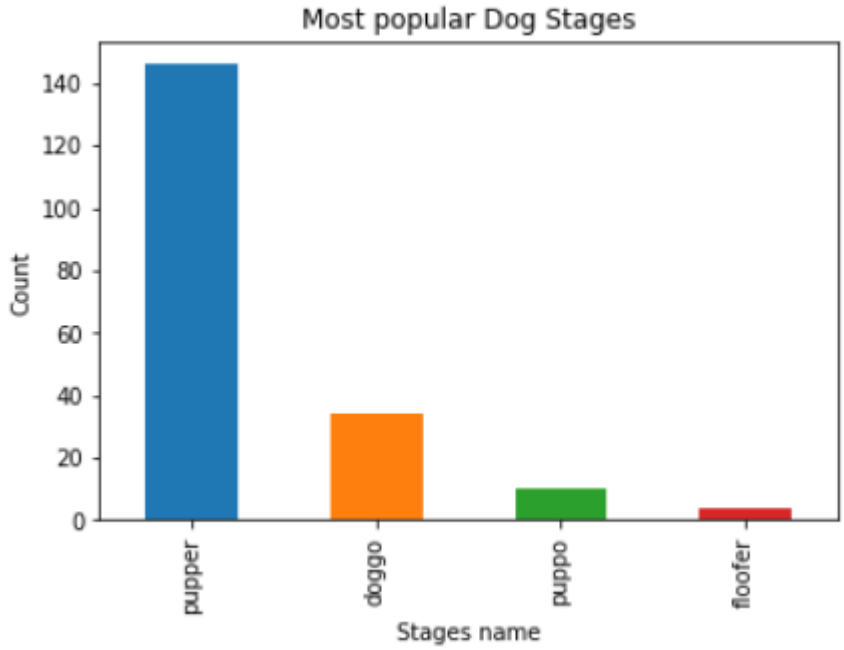
About the Data

To analyze the tweets from WeRateDogs™, we have used three different sources. The first source is an archive of the past tweets from @dog_rates (https://twitter.com/dog_rates) provided via a CSV from Udacity. The second source is from the Twitter API used to retrieve more information about the tweets like number of each was retweeted. The third data source provides us the predicted dog breed in each tweet's image programmatically determined from a neural network. This was also provided by Udacity. This third source is particularly important as it was not determined how accurate the predictions were.

Combining this data from these sources, we ended up analyzing 2174 tweets from WeRateDogs™. These tweets were from before August 2017 and were used to analyze the Most popular dog_stages, Correlation between favorite_count and retweet_count based on time and Top 10 Dog Breed.

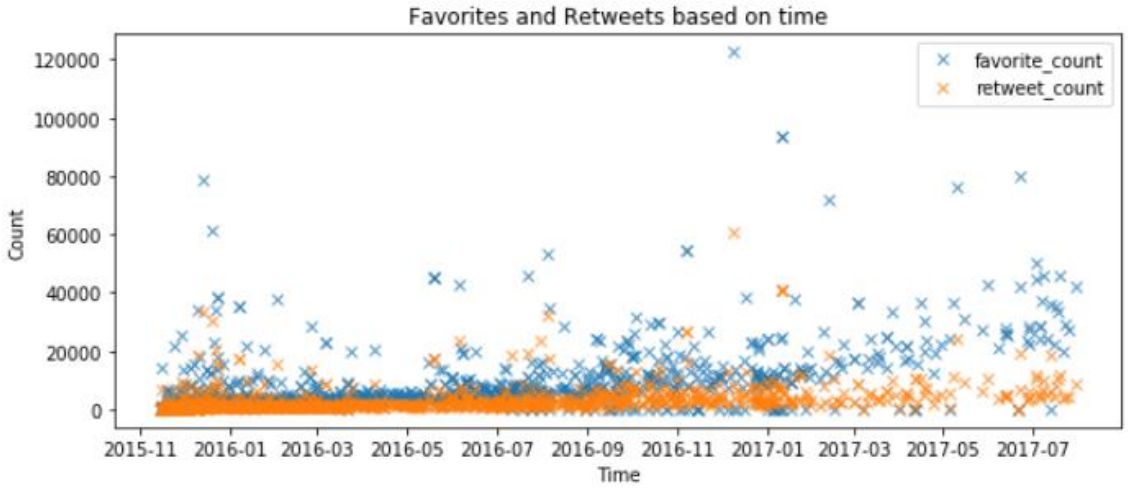
Most popular dog_stages

As per the data and from the visualization, the most popular Dog Stages are as follows:



Correlation between favorite_count and retweet_count based on time

As the correlation map shows if the count of retweet is high for a specific tweet the count of favorites goes high as well.



It will be more clear by looking at below table-

	favorite_count	retweet_count	rating
favorite_count	1.000000	0.791124	0.023180
retweet_count	0.791124	1.000000	0.024069
rating	0.023180	0.024069	1.000000

Most Popular Dog based on favorite_count, retweet_count and rating¶

1. The most popular dog based on favorite_count seems to be Stephan with 122636 favorite_count
2. The most popular dog based on retweet _count seems to be Stephan with 60873 favorite_count
3. The most popular dog based on rating seems to be Atticus with 177.6 rating

Top 10 Dog Breed

Top 10 dog bread are