

ANALYSIS OF WERATEDOGS TWITTER DATASET

Tayebwa Crispus

6th September, 2022

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. The account was started in 2015 by college student Matt Nelson, and has received international media attention both for its popularity and for the attention drawn to social media copyright law when it was suspended by Twitter for breaking these aforementioned laws. – Wikipedia.

Description of the data

The data used in this project was extracted from three sources.

The first source is a dataset of archived tweets from the WeRateDogs twitter account, @dog_rates (https://twitter.com/dog_rates) that was provided as a csv file by Udacity.

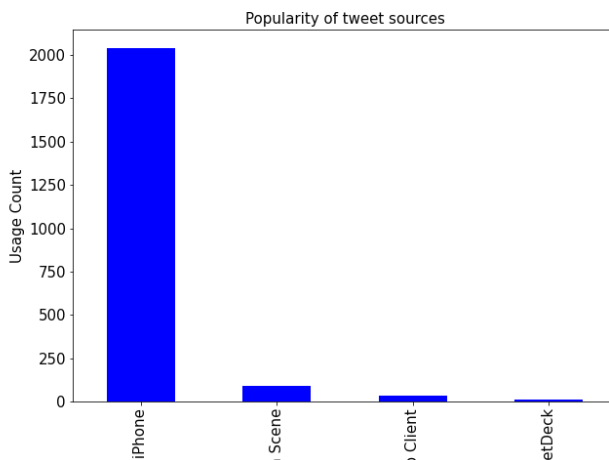
The second data source is a dataset of predicted dog breeds in each tweet's media. This was gathered and provided to me by Udacity as well.

The third source was the Twitter API - *Tweepy*. I used tweepy to wrangle two additional parameters for tweets, that is – tweet retweet count and like count.

Realized insights and visualizations

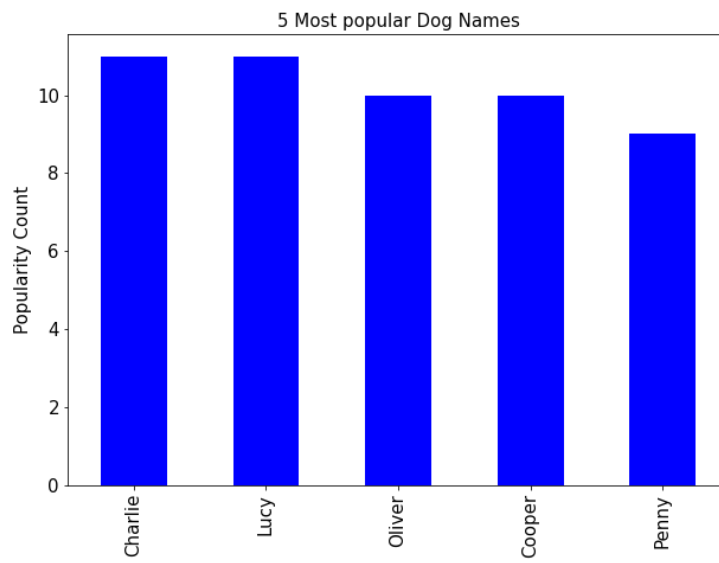
1. Popularity of source of tweets

My analysis gave me the realization that Twitter for iPhone is the most common/popular tweet source. This source numerates to over 2000 times as compared to the other three sources, that do not collectively make a quarter of the Twitter for iPhone numbers.



2. The most popular dog names

I realized that the five most popular dog names are Charlie, Lucy, Cooper, Penny, and Oliver. However, most dogs did not have a name record, and instead had None as their name record. Thus, I chose to disregard this event when carrying out this analysis.



3. Correlation between retweet count and like count

There is an almost perfect positive correlation between favorited tweet count and retweeted tweet count - 0.9262542799659841. This could indicate that users who liked a tweet did most likely also retweet it.

