WeRateDogs Twitter Archive - Act Report

Akshai Ramesh
August 2019

WeRateDogs Data

The WeRateDogs Enhanced Twitter archive contains data of about 5000 tweets from @dog_rates twitter account, posted between November 15, 2015 and August 1, 2017. This data is comprised of dog ratings that were taken from the text of the tweet along with dog name and dog stage if present.
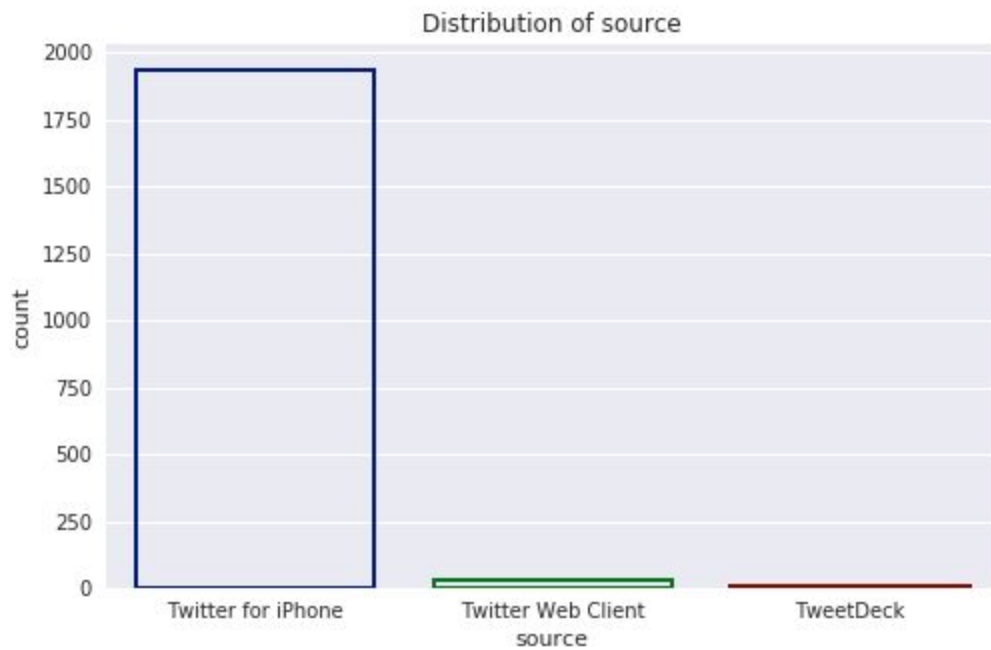
The retweet count and favorite count for each tweet are not included in the enhanced archive and so I had to download this additional data from the twitter account using the tweet ID from the archive.

Along with the Twitter data, I also downloaded an image prediction files from Udacity servers containing the top 3 predictions for dog breeds based on the images from the tweets.
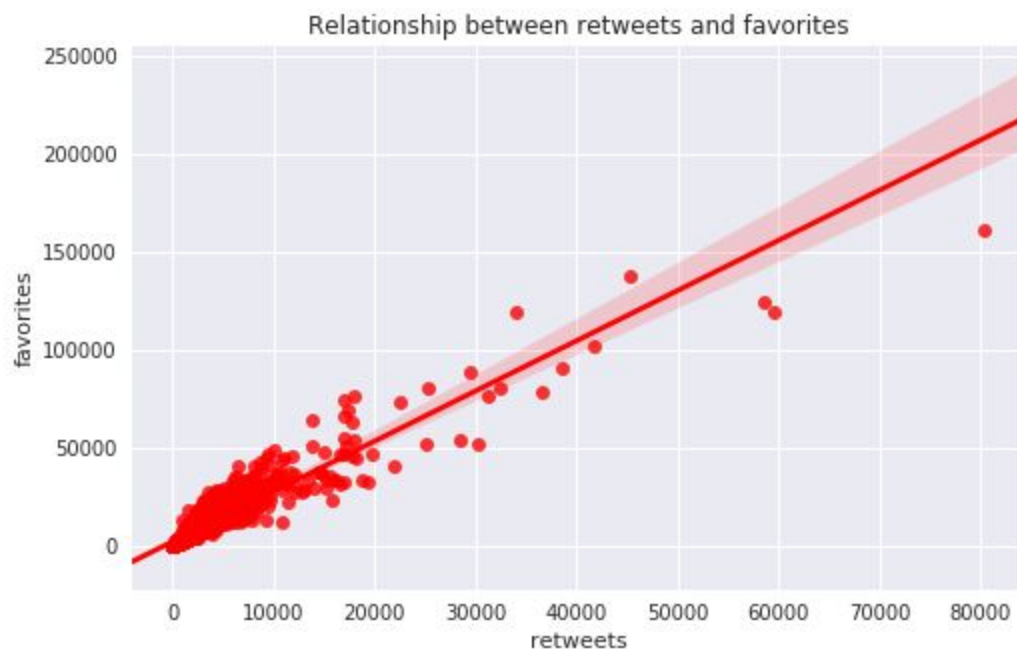
Insights

1. Distribution of source

We can see that the dominant source of tweets is from iPhone twitter app, which is 94% in total. That means the twitter app is the main channel for people using to tweet, retweet, post, and others, while the TweetDesk is pretty rare (less than 1%).
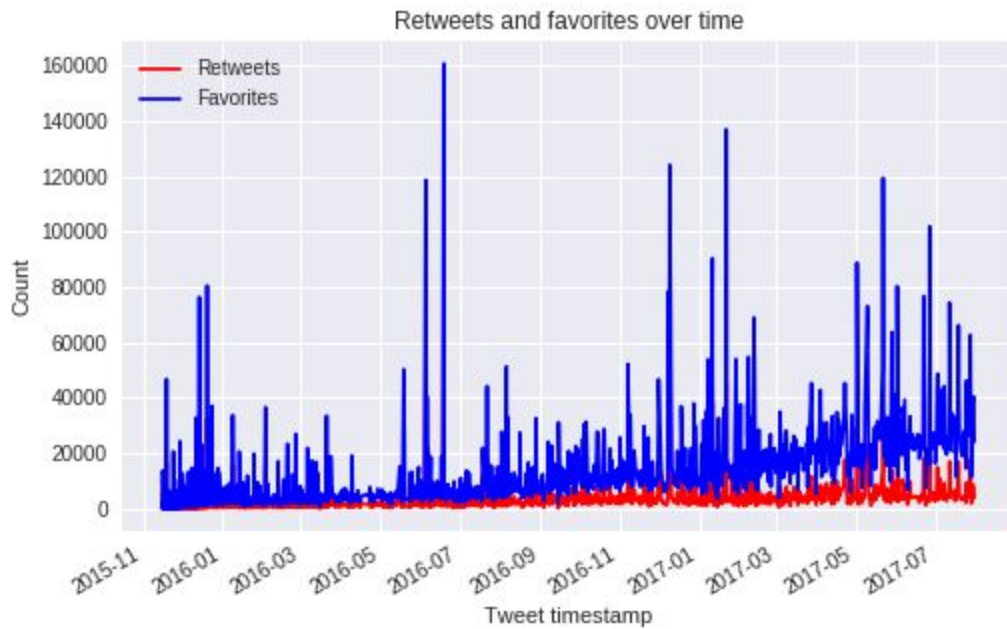
2. **Relationship between retweets and favorites**

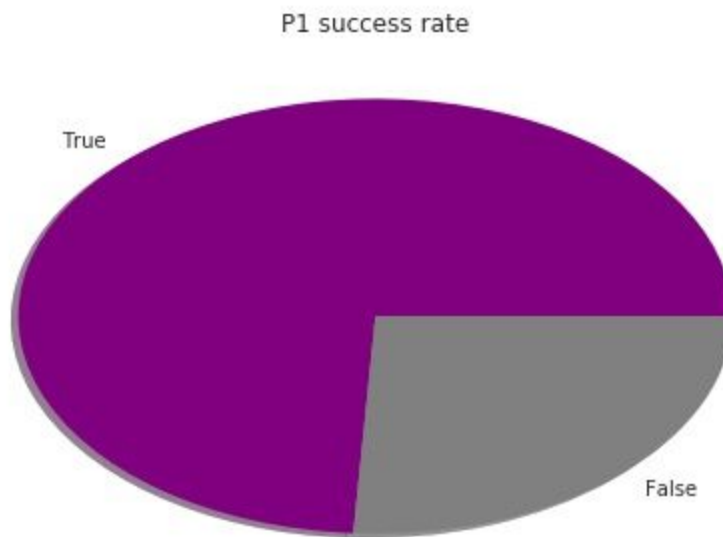The scatter plot shows a strong positive relationship between 'retweet_count' and 'favorite_count'.

3. Retweets and favorites over time

The graph shows that the highest number of retweets are posted mostly during July,2016 and most tweets are liked during July, 2017.
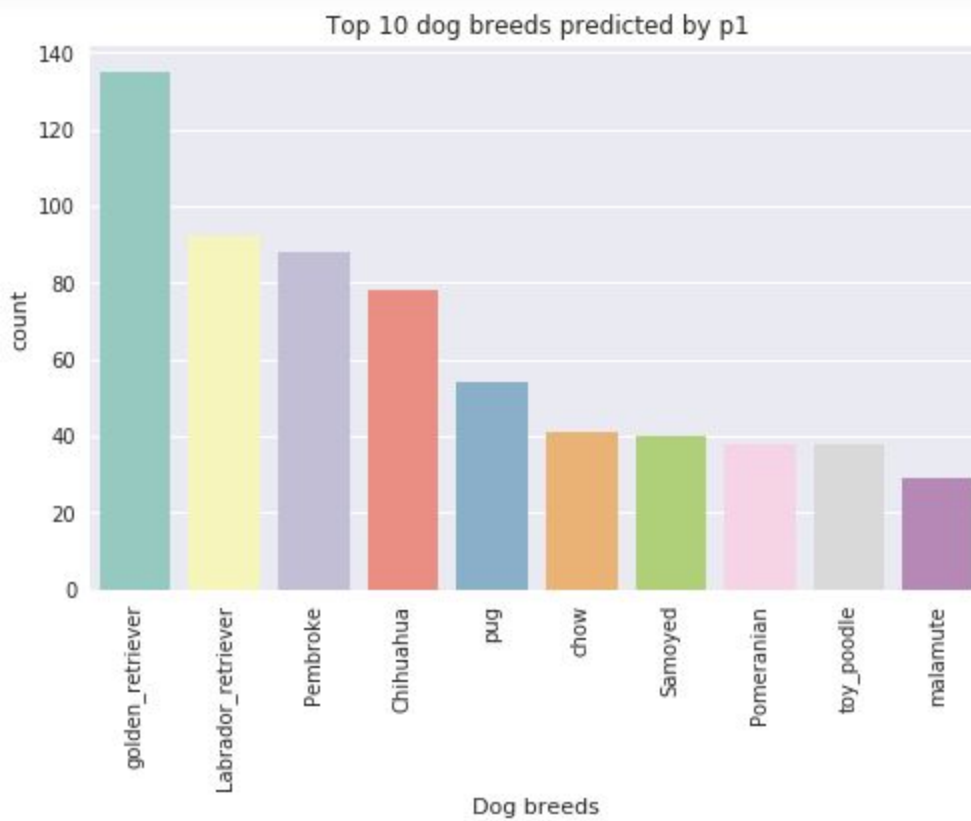

4. Analysis of dog results generated by first algorithm

- First algorithm has predicted around 75% of the dog breeds accurately.

P1 success rate

- 'Golden retriever', 'Labrador retriever' and 'Pembroke' are the top 3 dog breeds analyzed by first algorithm.

- 'Chihuahua', 'French buildog' and 'labrador retriever' are the top 3 breeds that received highest number of likes.

Top 10 breeds that received highest number of likes