Act Report

Introduction

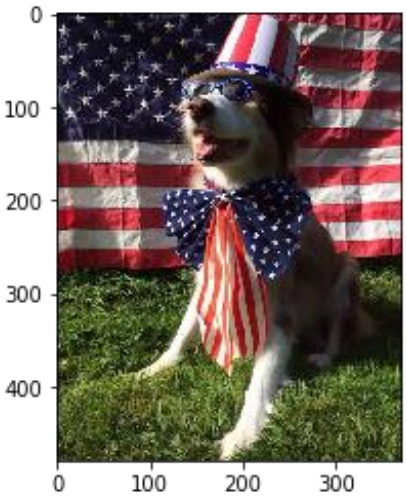
After gathering, assessing, and cleaning the dataset it was time to act on it! I started looking at some interesting facts that may be interesting. Since I have no clue what this dataset should be used for, I used my imagination.

Most liked and retweeted photo.

This photo of a cute doggo with the text “Here's a doggo realizing you can stand in a pool. 13/10 enlightened af” got the most likes and retweets in the history of the twitter account. At a staggering 150,911 Likes and 74,530 retweets this video blew the internet!

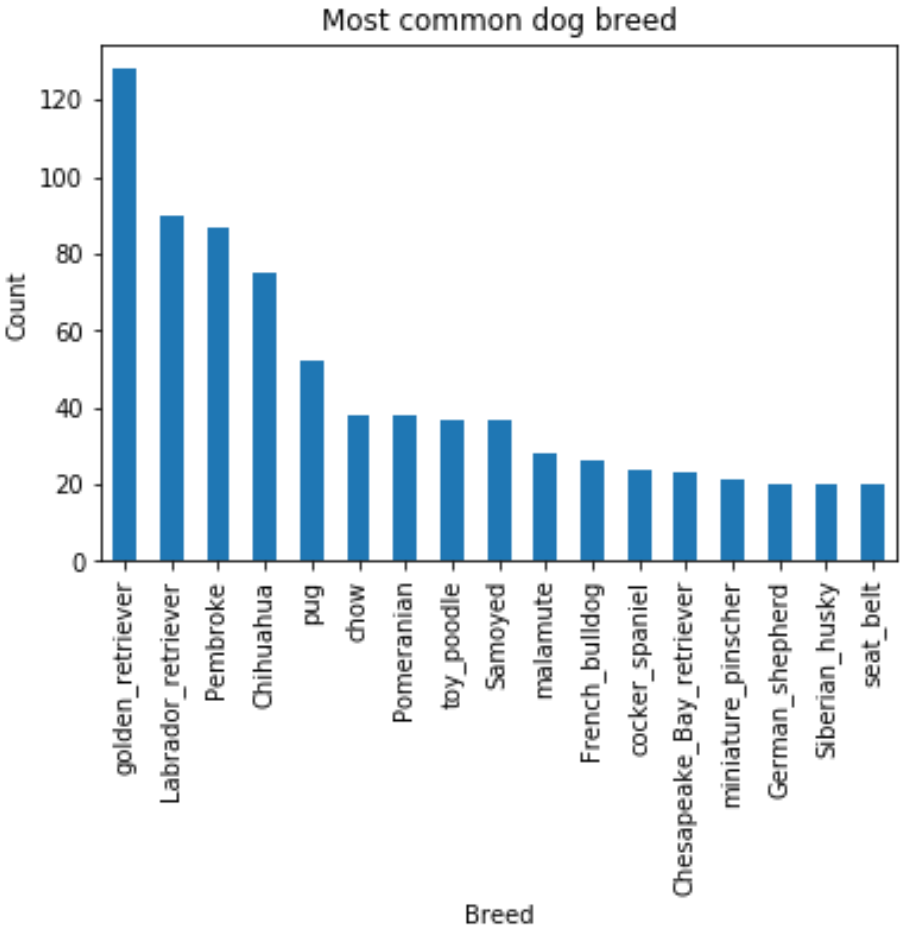
@dog_rates’ favorite dog

This photo with the text “This is Atticus. He's quite simply America af. 1776/10“ got the twitter account’s highest rating ever of 1776



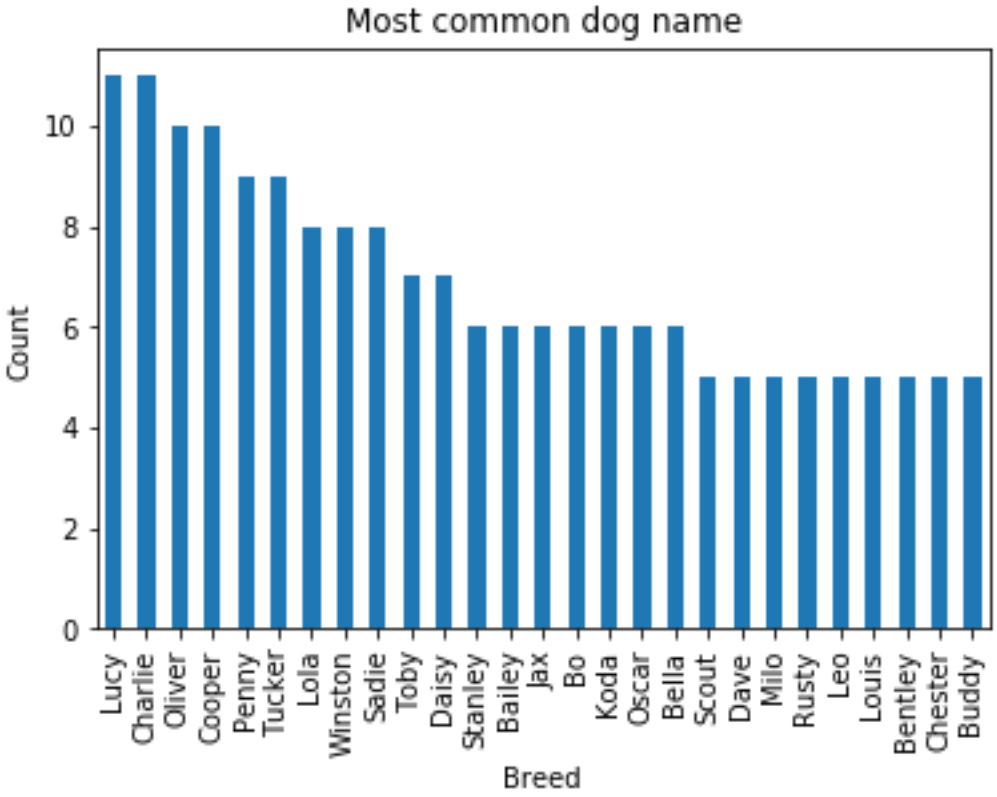
The golden retriever wins the most common doggo award.

This shows the different dog breeds the model predicted from all the @dog_rates photos. The golden retriever is the 3rd most common breed globally but it's the first here. The Labrador Retriever came in second even though it's the most popular breed globally. Is @dog_rates biased to golden retrievers?



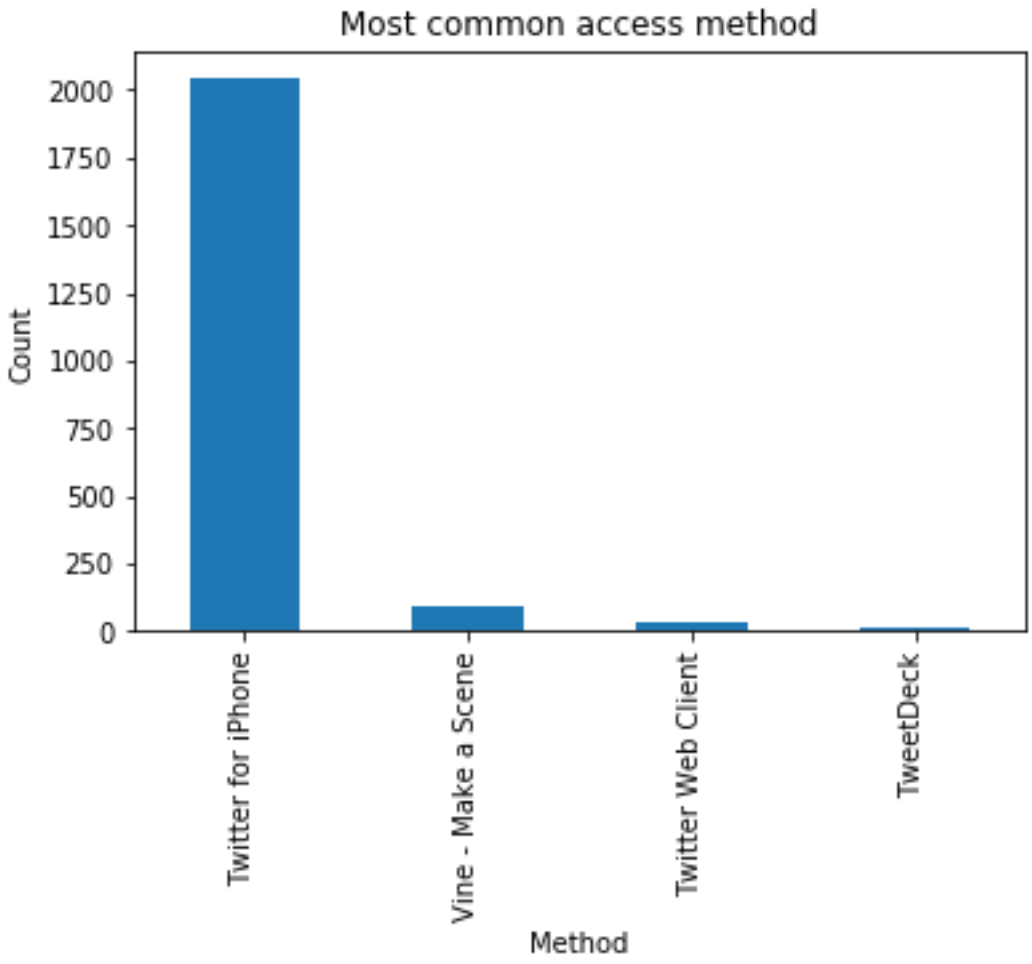
The most common dog name is Lucy!

It didn't really win by much as most of the dog names data is missing. Maybe some NLP can extract the dog names from the tweet text. Lucy is the world's 2nd most common female dog name and Charlie is the 3rd male name. So, the data isn't that far off.



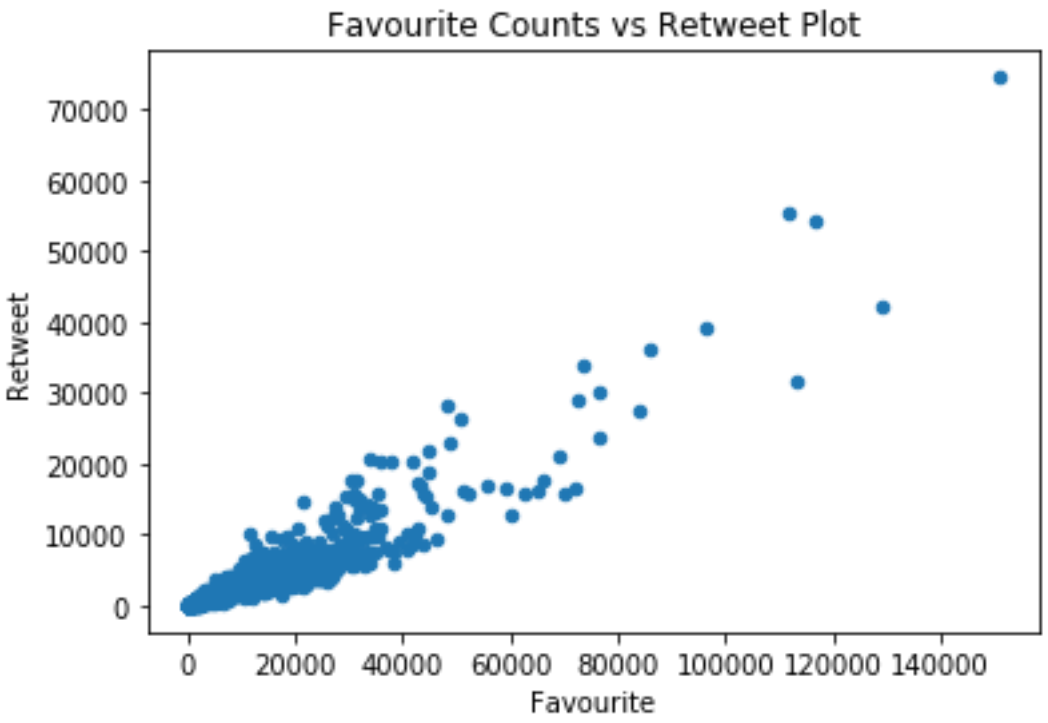
This shows how people access twitter

Apparently, iPhone wins this and there's no competition. I call the dataset rigged.



Favorites vs Retweets

The following figure shows the difference between Favorites and retweets. We can see that for each retweet there are around 3 * as much Favorites.



We're out of visualizations ☹

- The average confidence of the most confident photos is 59% which shows that the model needs some work.
- 26% of the predicted photos predicted something other than the dog in the photo. Maybe a less general model would solve this problem