

Analysis and Visualization of WeRateDogs

Introduction

When it comes to dealing with real-world data rarely comes clean. Using Python and its libraries, we gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling.

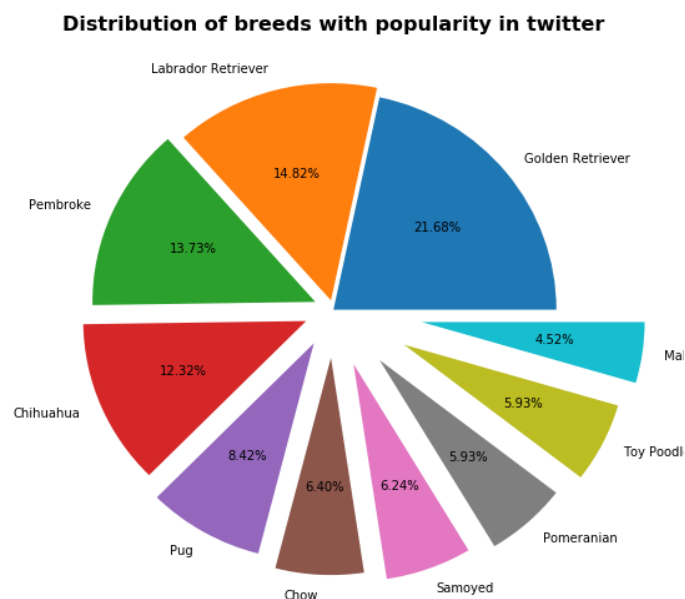
The dataset that I'll be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. WeRateDogs 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 me 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.

Which dog type (breed) is most common dog in the tweet dataset? What dog type has the highest average rating? These questions and more are answered in the following insights:

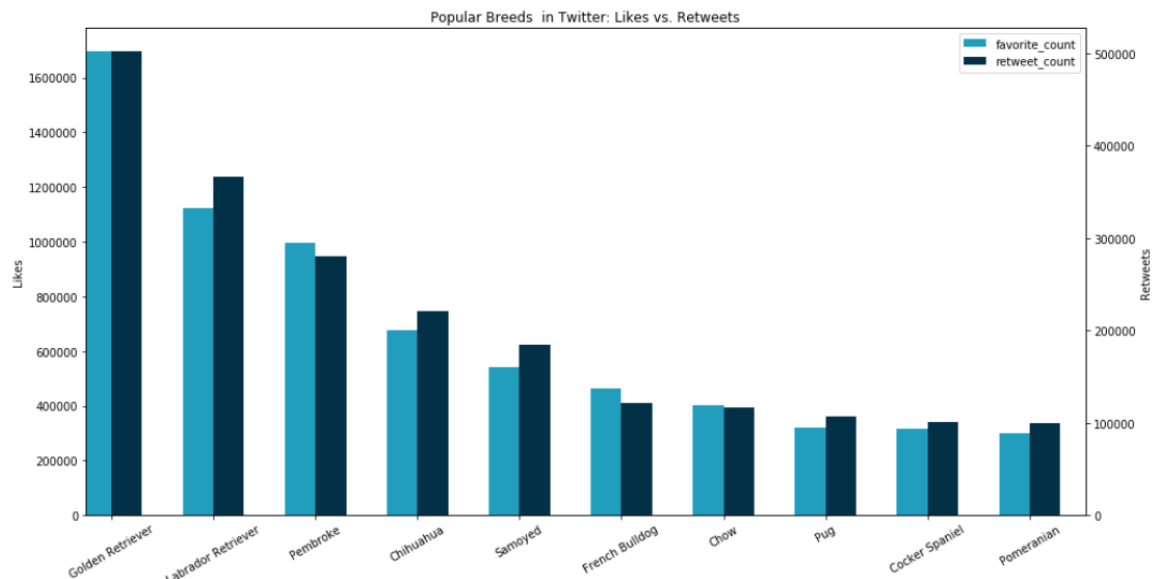
The Most Popular Breed

WeRateDogs has over 6000+ tweets. I was able to analyzed around 1900+ tweets. The most rated dog was golden retriever with more than 140 ratings. We can see that golden retriever definitely stands out, claiming the first rank in popular dogs.



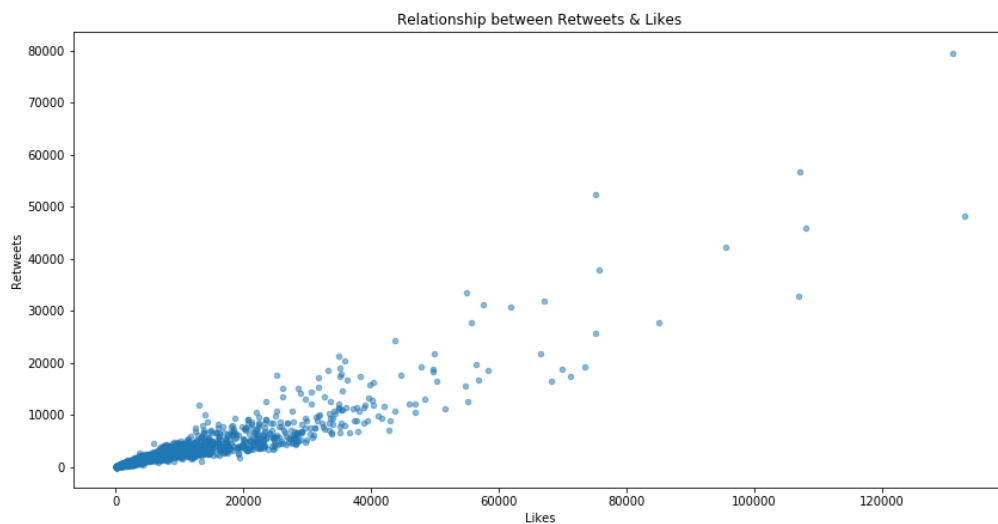
Most Popular breed in terms of retweet and likes

So here, we find that the breed of dog that has the most attention of people. Such that the people tend to like and retweet them. Well it may not surprise you, and the winner is golden retriever!



Correlation between Retweets and Likes

From the figure we can draw out that there is positive correlation between likes and retweets.



Tweeting Trend Over Time

So, how's the number game going on with WeRateDogs. Judging based on the graph it seems that it is gradually decreasing month after month.

