WeRateDogs

Udacity Data-Wrangling Project
Summer Cook

INTRODUCTION

WeRateDogs is a Twitter user that was become an institution with wide appeal and recognition. They post pictures, comments and ratings of dogs. This project is an investigation into tweets from this user. The main questions of interest are:

- What is the relationship between ratings from the user and popularity among followers?
- Which breeds of dogs are most common and which are most popular?
- How close is the relationship between retweets and favorites?
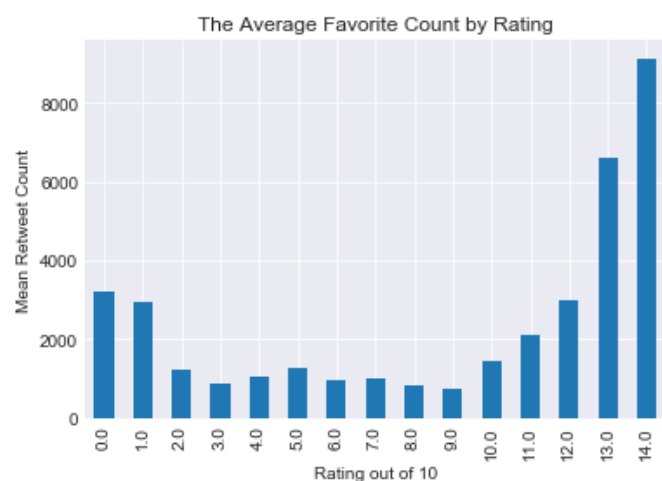- How have the variables changed over time?

SOURCES

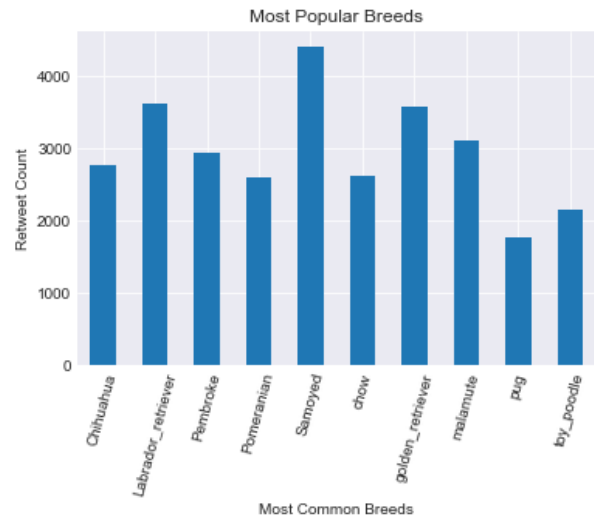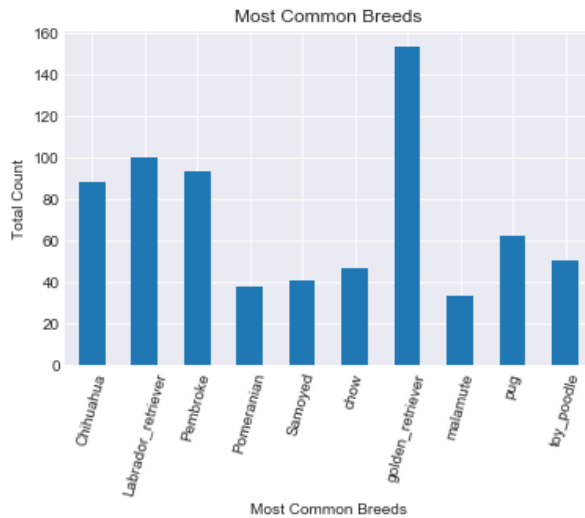Data for this investigation comes from three sources:

1. csv file including a table of Twitter archive data including tweet ID, source, dog's development stage, rating, retweet info, reply info and tweet text. This table set the temporal boundaries of the investigation which were: 12 November 2015 – 1 August 2017.
2. Twitter API which was to be used to extract missing data from the archive data, most importantly, retweet count and favorite count. This was extracted programmatically to a txt. file using python.
3. Tsv file from a Udacity server including data created with an algorithm that predicts the type of dog from the photo connected to the tweet. This was downloaded programmatically with python.
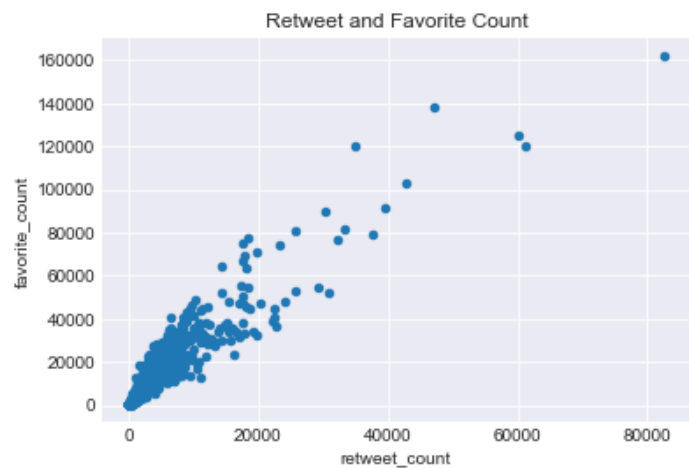
INSIGHTS AND CONCLUSIONS

Ratings are correlated with popularity, but the relationship is not as strong as I expected. R-squared = 0.089.
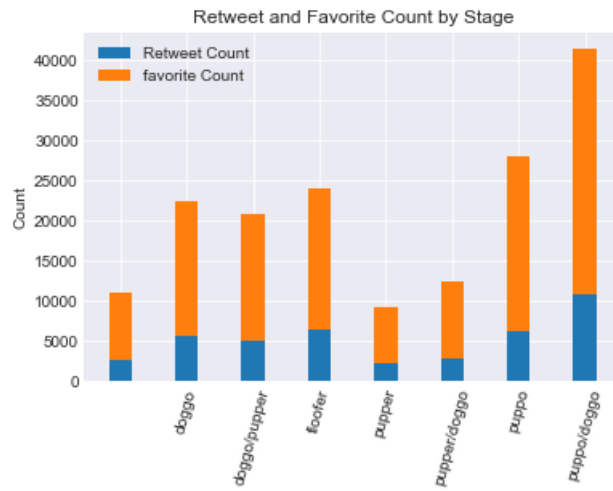
There are more tweets about Golden Retrievers than any other breed. However, among the 10 most common dogs, Samoyed has the highest mean retweet count. The dog with the highest retweet and favorite count is a laborador retriever.

Most Common Breeds

Most Popular Breeds

The correlation between retweet counts and favorite counts is very strong. R-Squared = 0.859. They are consistent measures of popularity.

Retweet and Favorite Count

Ratings Over Time

Ratings seem to have increased over time.

Retweet and Favorite Count by Stage

Dogs in the puppo/doggo stage have the highest mean favorite count.

Wednesday is the day most likely for a tweet to be posted.

Retweet Count by Day of the Week