2020

Dogs

We Rate Dogs

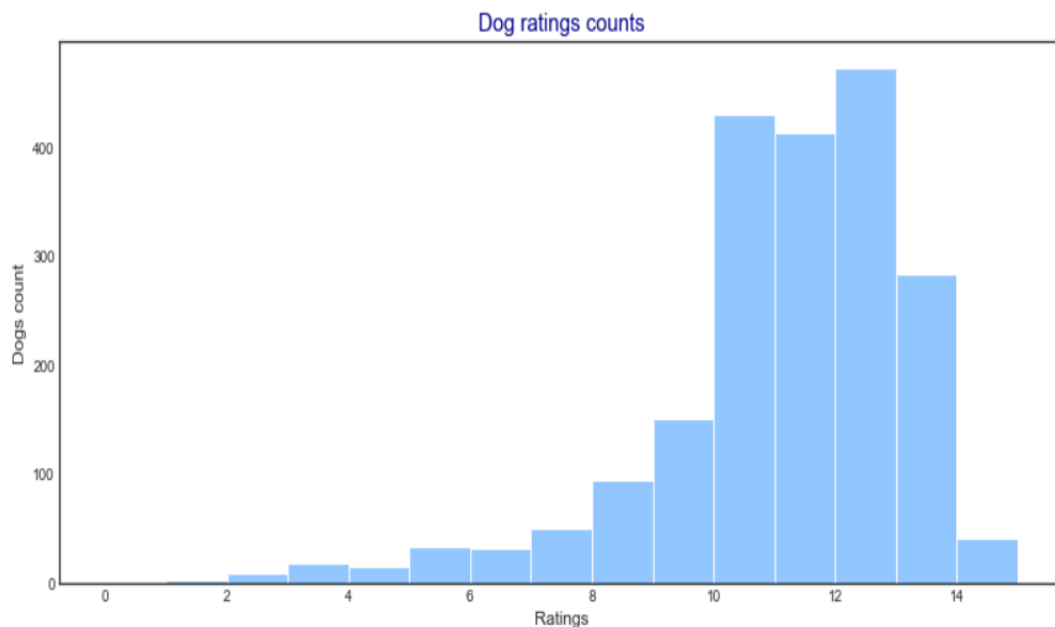Data Wrangling Act Report

Fagr Ahmed

12/13/2020

WerateDogs Data

The WerateDogs Enhanced Twitter archive contains data extracted from 2356 of the 5000+ tweets from the @dog_rates twitter account, posted between November 15, 2015 and August 1, 217. This data is comprised of dog ratings that were taken from the text of the tweet along with the dog name and dog stage if present. The retweet count and favorite count for each tweet were 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 predictions file from Udacity servers containing the top 3 predictions for dog breeds based on the images from the tweets (there can be up to 4).

Wrangling Data

Before I could begin the analysis, the data had to be wrangled into shape to make it easier. I assessed the data both visually and programmatically for quality and tidiness; the quality of data is determined mainly by looking at several aspects or dimensions to ensure that it is complete, valid, accurate and consistent. After cleaning many of the issues found during the assessment, there were about 1950 tweets with good quality data.
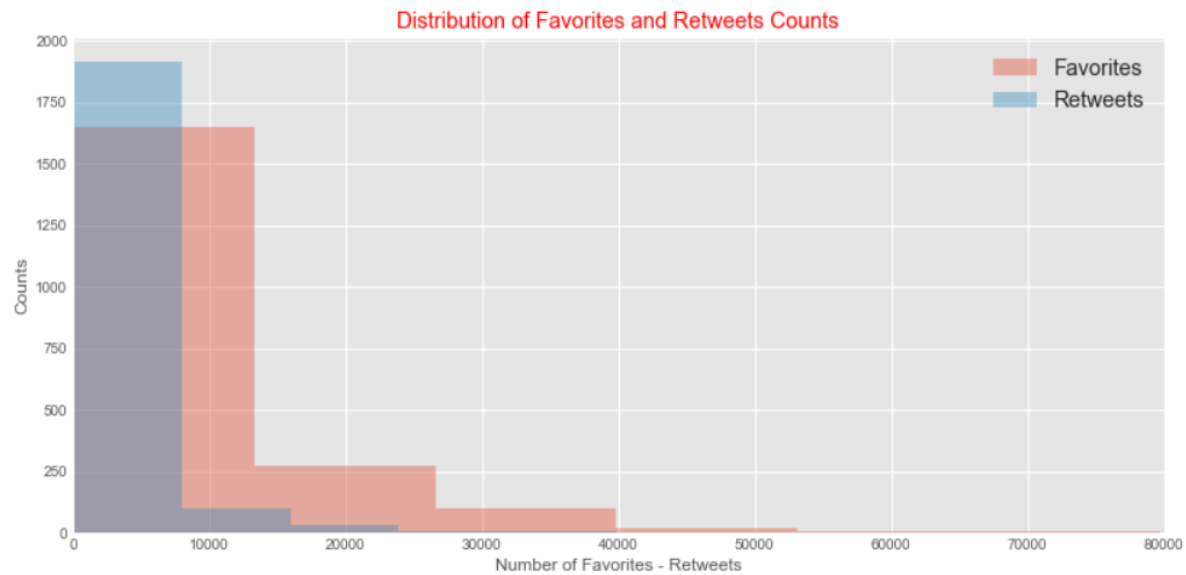
Ratings analysis

After cleaning rating values, these values are used to develop a histogram between rating scores and dogs counts that received that score. And it looks like most of dogs are good that their ratings are above 10.
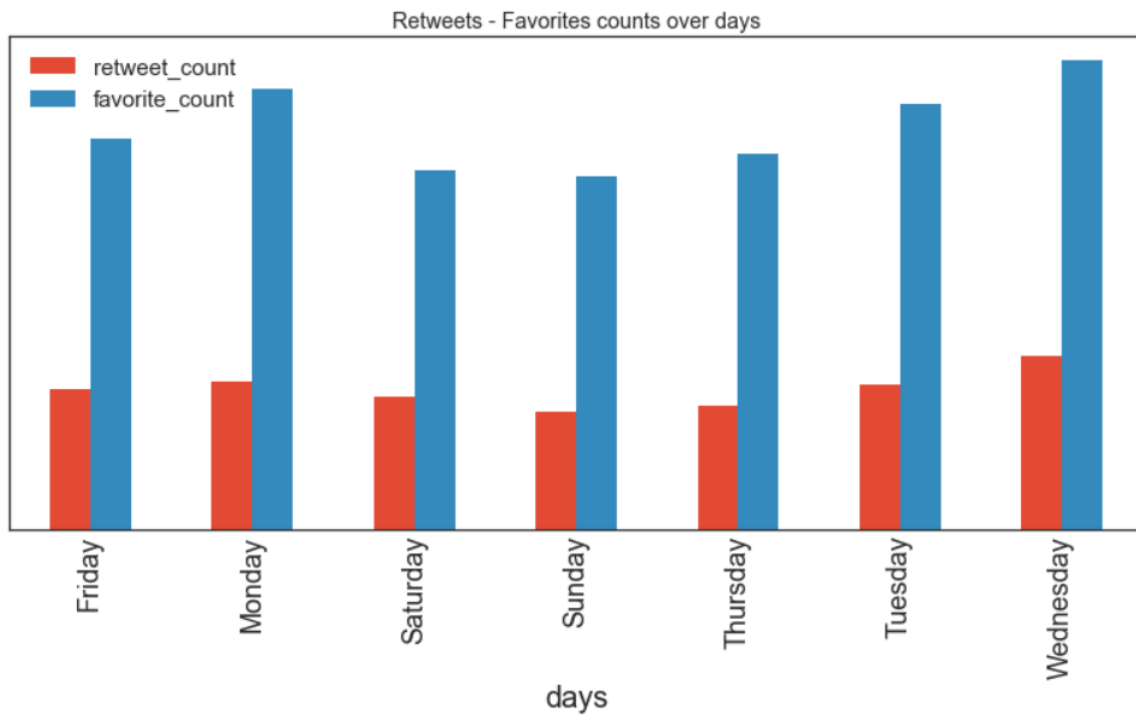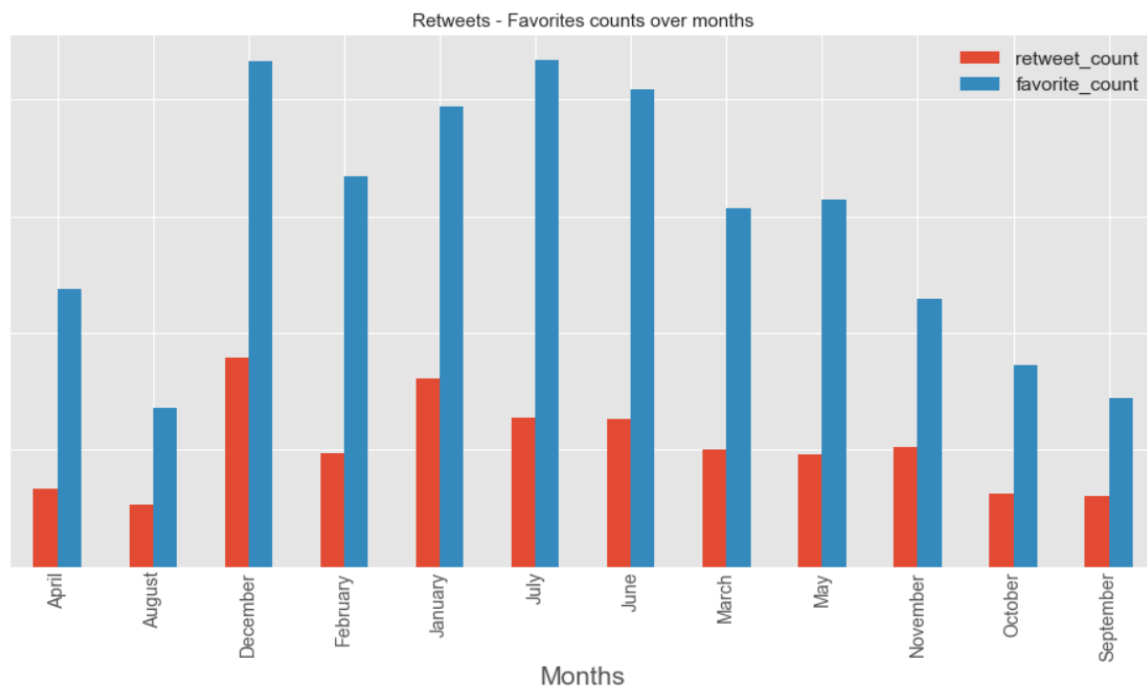
Favorites and retweets analysis

We want to understand the relationship between favorites and retweet. Do people like tweets only? Retweet without liking the tweet? Using the favorites and retweets records to answer these questions

Distribution of Favorites and Retweets Counts

Scatter plot of Favorites and Retweets counts

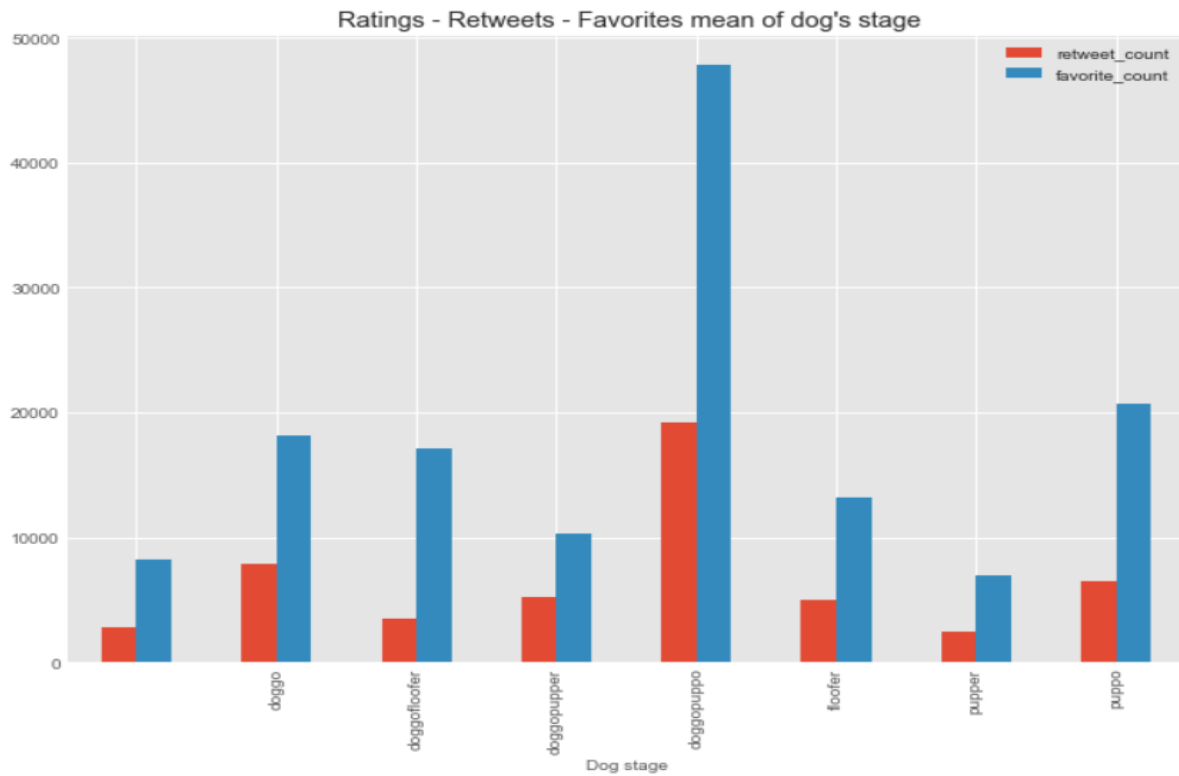# Favorites and retweets analysis by months and days:

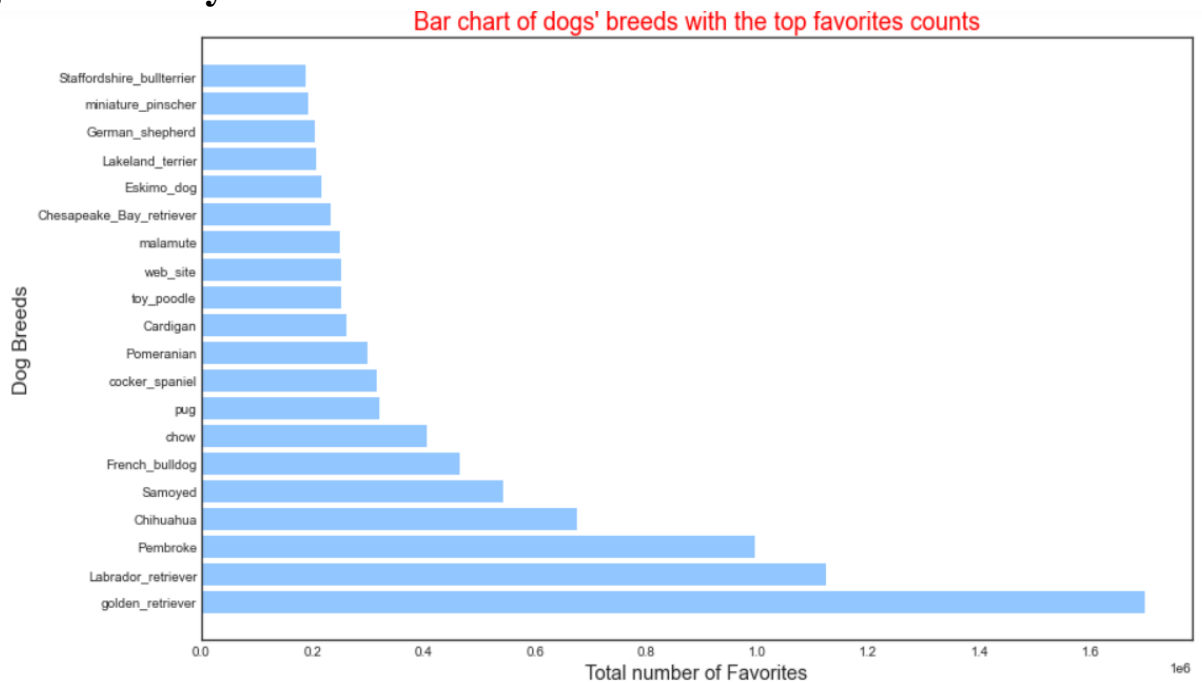The Highest number of retweets and favorites were on Wednesday

The Highest number of retweets and favorites were in January and December.

Dog stage analysis

Dog breed analysis

We have a winner! Golden Retriever is the most popular dog.