Udacity's Data Wrangling Project Report

Prepared by: Suhaila Ehab Alkholy

Project Overview

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. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage. In this report I report on my efforts to gather, assess and clean the twitter data to gain some useful insights. Which do you think is the most loveable dog?

Methodology

Data Gathering

- This is the part where I start to gather the data.
- Firstly, I import the Twitter archive as a DataFrame into my Jupyter Notebook
- Secondly, I decided to download the image prediction file programmatically and then create a DataFrame with its data using Python's Pandas library
- Thirdly, I used my Twitter's Developer account and Python's Tweepy to create an API object. Each Tweet is JSON file and I wanted to import all of its data into a text file where each new tweet forms a new line. I then create a Panda's DataFrame with this data.

Data Assessment

This is the part where I assess the gathered data.

- I assess the data visually and programmatically.
- I find 8 quality issues and 2 Tidiness issues.

Data Cleaning

This is the part where I clean the data, and store the data to CSV files.

- I use the method of **define**, **code** and **test**.
- For each issue I define how I would tackle it, then I write the code and then test if the quality has been resolved.
- I then store the data to CSV files to start my analysis process.

My Findings:

## Total average monthly Interactions
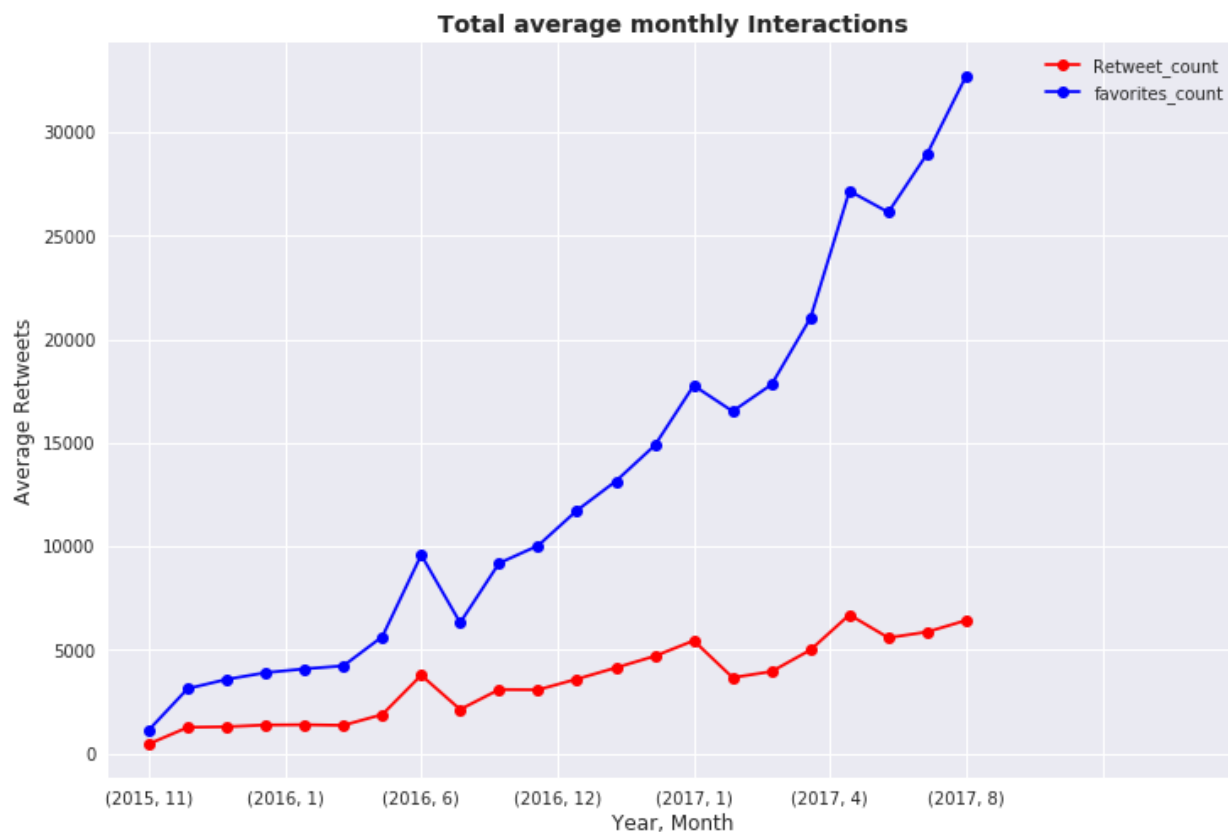
First Insight:

The Twitter account WeRateDogs gets more popular over the years. There is a linear increase in interaction level through favorites and retweets as shown by the figure above. There is a sudden spike in May, 2016 which could be a Tweet that went viral.

Second Insight:

Interactions with Tweets through favoriting is always higher than the tweet being Retweeted. The stage of 'doggo-puppo' shows the highest favorites and retweets count, puppo is the second highest.
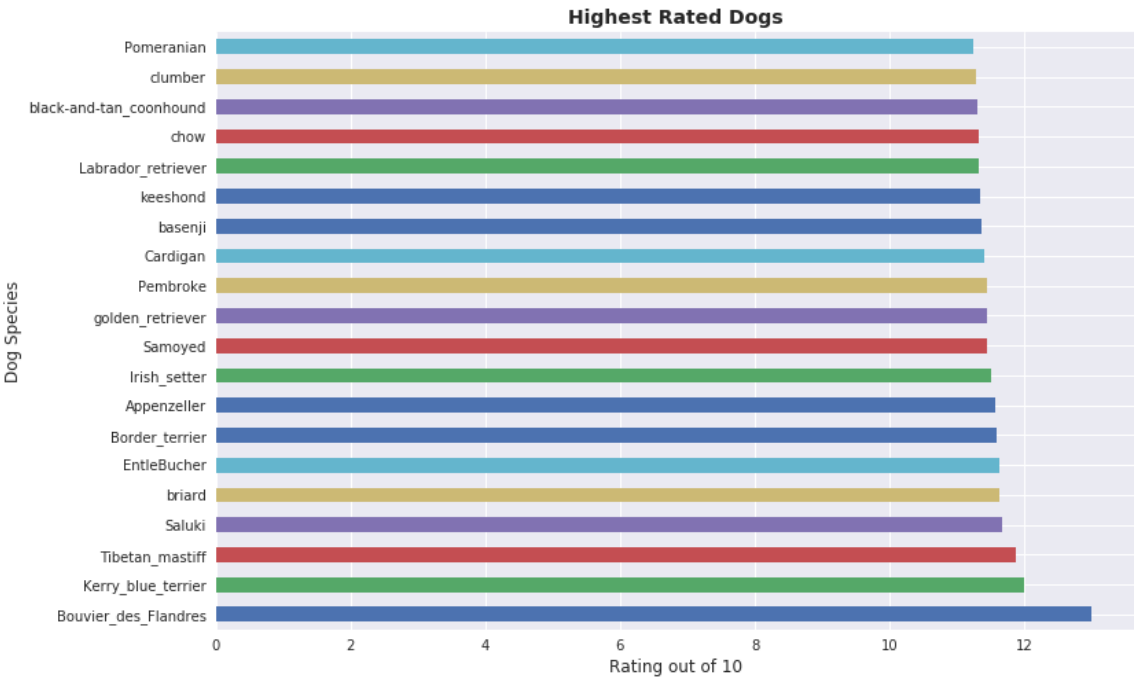
Interactions Based on Dog Stage in Life
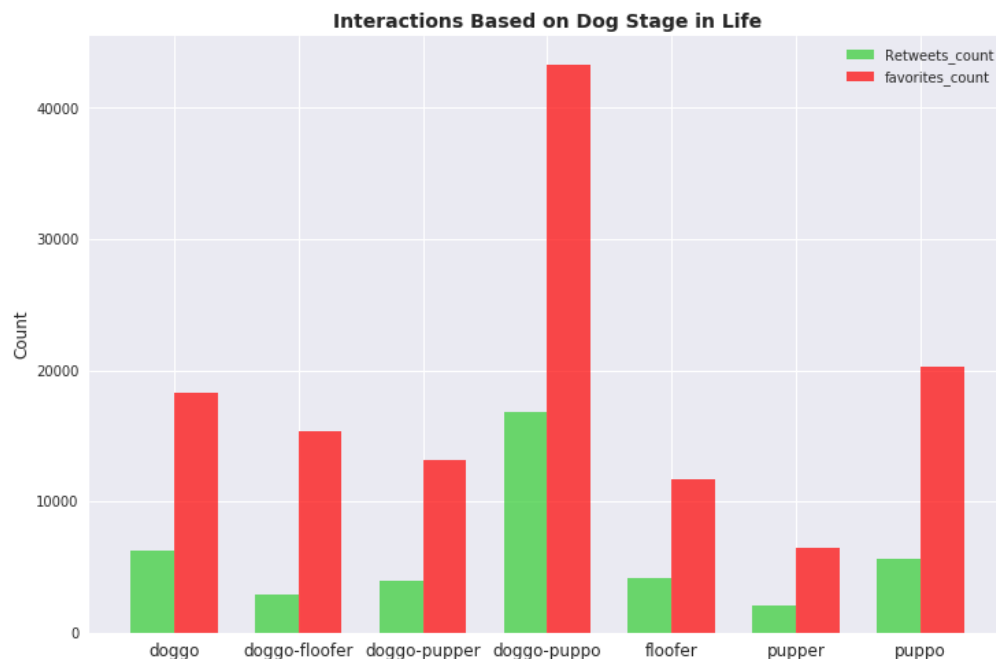
Most Loveable Dog Species
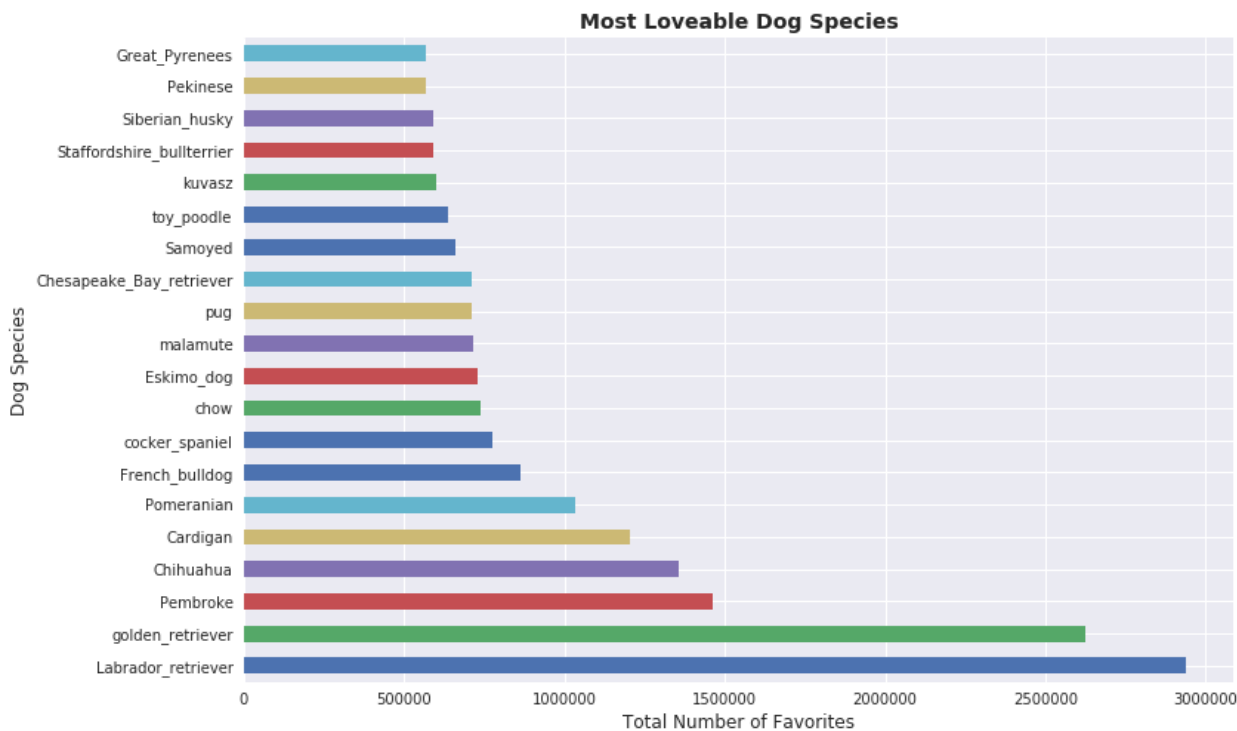
You guessed it! Most Loveable dog species is Labrador Retriever as shown by the figure above. Next, comes the second Golden Retriever. The least popular dog species among followers of the WeRateDogs account is Great_Pyrenees. Golden Retrievers and Labrador Retrievers are also highly rated dogs but not the highest even though they gain they are the followers' favorite species.

Bouvier Des Flandres dogs have the highest average rating. As we can see, not all highly rated dogs gain the level of interaction anticipated. The followers always have a mind of their own!

Highest Rated Dogs

References:

https://nfpdiscussions.udacity.com/