Act Report

Overview

This report includes valuable insights and visualizations discovered from WeRateDogs twitter data wrangling project.

Image via Boston Magazine

This project comprises of three datasets

- The first dataset `twitter-archive-enhanced.csv` which is a csv file was downloaded manually and contains basic tweet data (tweet ID, timestamp, text, etc.) of about 2356 @dog_rates tweets downloaded by WeRateDogs from their twitter archive and sent to Udacity exclusively for this project

- The second dataset `image_predictions.tsv` is a tsv file and was downloaded programmatically from Udacity servers. It contains a table full of image predictions (the top three only) which comprises of about 2075 tweet ID, image URL, and the image number that corresponded to the most confident prediction (numbered 1 to 4 since tweets can have up to four images).

- The third dataset which comprises of about 2327 observations was obtained using Tweepy to query Twitter's API . This dataset which I queried using Twitter's API to the file `tweet_json.txt` houses data such as tweet ID, retweet count, and favorite count which corresponds to the tweet ID in the `twitter-archive-enhanced.csv` dataset.
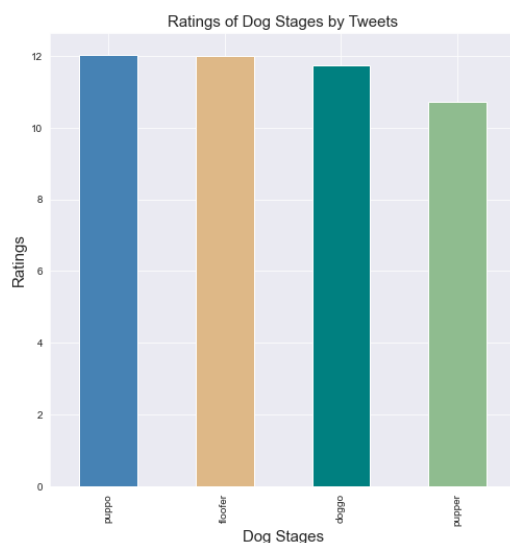
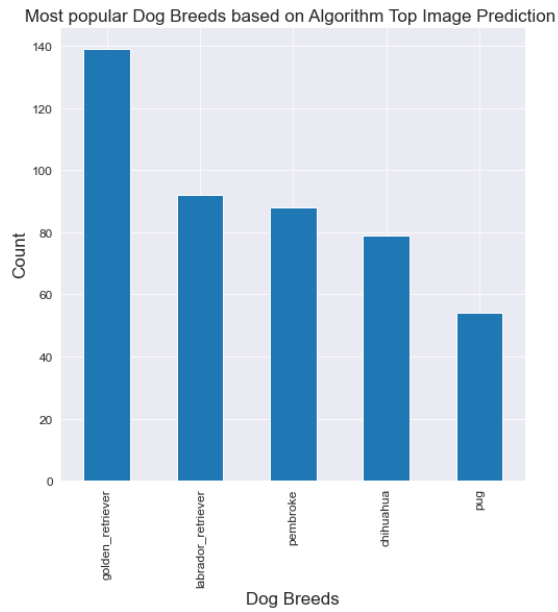### Insights and Visulaizations

Insights

1. The highest rated dog stage by tweet is **Puppo**

2. Based on the algorithm's #1 prediction, the most popular dog breed is the **golden retriever**

3. There is a **high positive correlation of about 0.92** between retweet count and favorite count

4. The minimum retweet count is 11, mean is 2243, and the maximum retweet count is 70717

5. The minimum favorite count is 66, mean is 7709, and the maximum favorite count is 144774
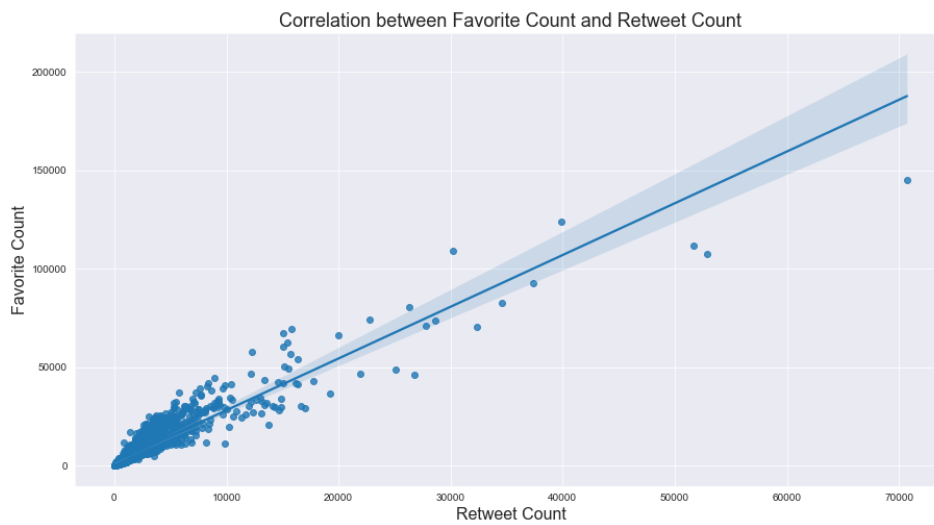
Visualizations

1. The highest rated dog stage (doggo, floofer, pupper, and puppo) by tweet is **Puppo**

1. Based on the algorithm's most confident image prediction, the most popular dog breed by occurence is the **golden retriever**

Most popular Dog Breeds based on Algorithm Top Image Prediction

1. There is a **strong positive correlation of about 0.928** between retweet count and favorite count

The Correlation between Retweet Count and Favorite Count is approximately 0.928, indicating that retweet count and favorite count have a strong positive correlation. However, we cannot mistake correlation for causation as although retweet count and favorite count are strongly correlated, we cannot imply there is a causal relationship.