We Rate Dogs: Act Report

Tim Quan

***Udacity**: Data Analyst Nanodegree Program*
***Project 03**: Wrangle and Analyze Data*

March 2022

1 Introduction

In this stage of the project, we used our previously gathered/wrangled data to interpret, analyze, come up with insights, and create visualizations.

2 Insights

- Both favorite count and retweet count have a positive correlation with rating. We were able to establish this using OLS linear regression. The associated p-values for both parameters were 0. It is likely that there is a pairwise relationship/multicollinarity between the two. We were able to visualize the pairwise relationship using the seaborn module.

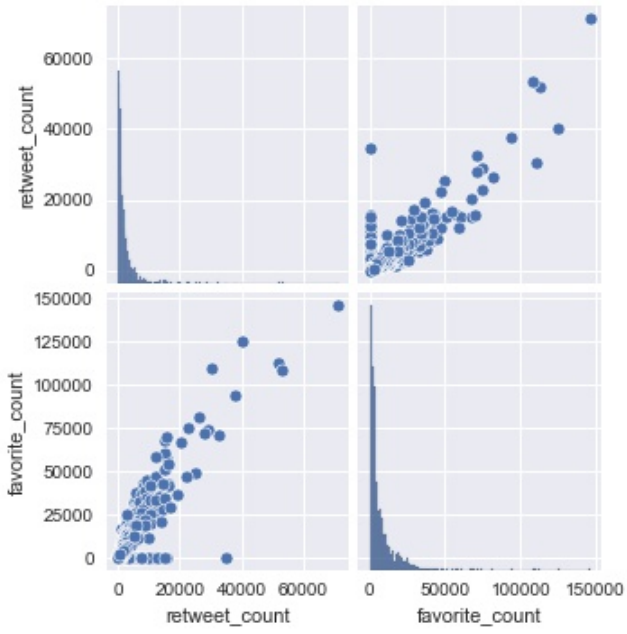

Figure 1: Retweet Count v. Favorite Count

- Using a linear regression model on the dog category/type variables with a result of rating, doggo and puppo records had sufficient p-values to indicate that they have an impact; floofers and puppers do not.
- The top 5 rated dogs (as detected) were pomeranian, Samoyed, golden retriever, kuvasz, great pyrenees, in that order. This rank order was limited to breeds that had a minimum of 10 ratings to avoid being skewed by outliers/one off ratings.

3 Visualizations

- Here are scatter charts showing the relationships between favorite counts, retweet counts, and ratings.

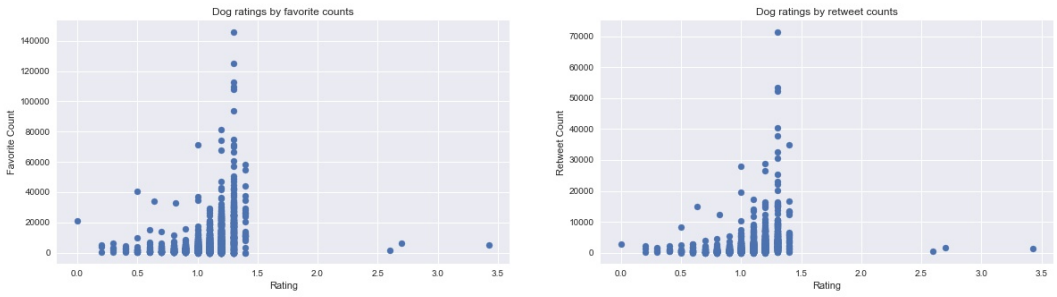

Figure 2: Retweet Count v. Favorite Count
Both charts follow the same pattern illustrating the possibility of multicollinarity.
`https://github.com/timothyquan/wrangle_and_analyze_data/blob/main/reports/scatter-retweets-favorites.jpg`

- It seemed that the source device would be interesting, but after this pie graph illustrating the source device breakdown was generated, it became apparent that the 'We Rate Dogs' account was run almost exclusively on an iphone:

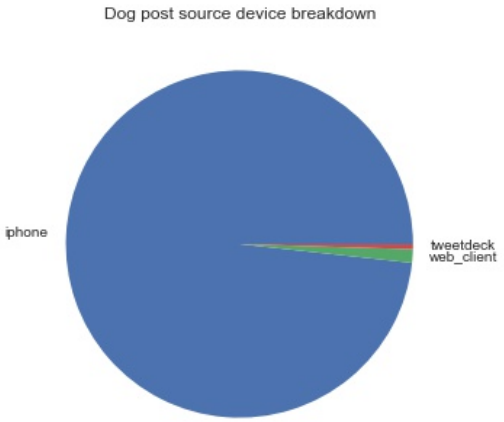

Figure 3: Source device breakdown
`https://github.com/timothyquan/wrangle_and_analyze_data/blob/main/reports/source_device_breakdown.jpg`

- Here's a wordcloud; it was a fun little process figuring this out. It turned out a fairly attractive image.

Figure 4: A beautiful dog wordcloud.
`https://github.com/timothyquan/wrangle_and_analyze_data/blob/main/reports/dogcloud.jpg`

- This one turned out great; a high-res dogsaic. Most of the heavy lifting was already done in terms of image generation code goes, but this was fairly time consuming to retrieve all 1600+ images and tweek the code to make it come out nicely.

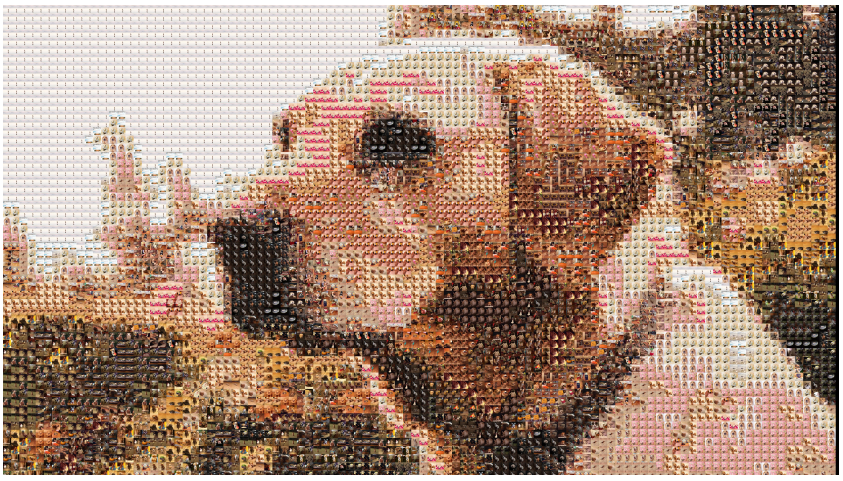

Figure 5: A mosaic generated from the images in the collection.
`https://github.com/timothyquan/wrangle_and_analyze_data/blob/main/reports/dogsaic.jpg`
Credit to `https://towardsdatascience.com/how-to-create-a-photo-mosaic-in-python-45c94f6e8308` *for most of the code required*