top

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.

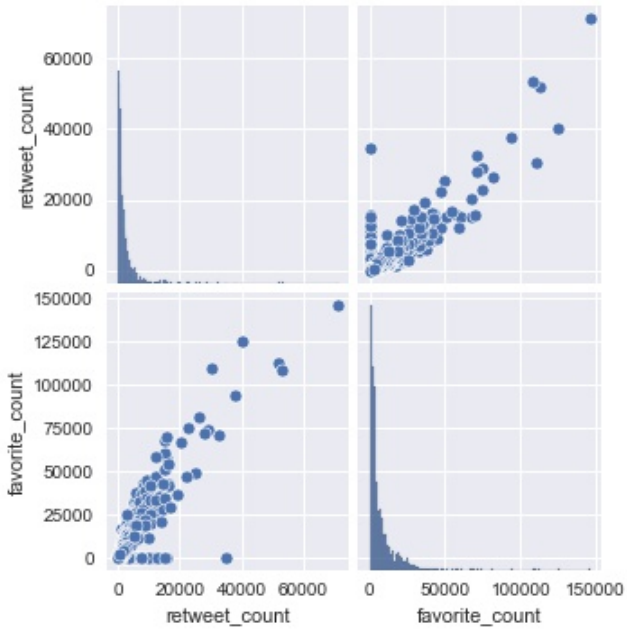

Figure 1: Retweet Count v. Favorite Count

- Of the dog category/type, doggo and puppo records sufficient p-values to indicate that they have an impact rating; floofers and puppers do not. We were able to establish this using linear regression.
- 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

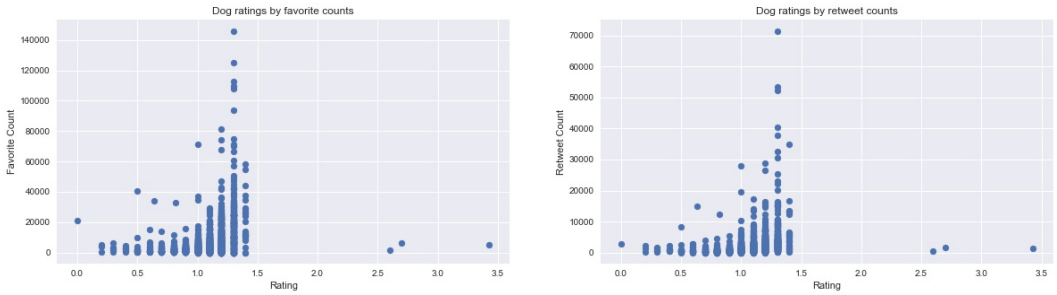

Figure 2: Retweet Count v. Favorite Count
Both charts follow the same pattern illustrating the possibility of multicollinarity.
`https://github.com/timothyquan/wrangle_and_analyze_data/blob/main/reports/scatter-retweets-favorites.jpg`

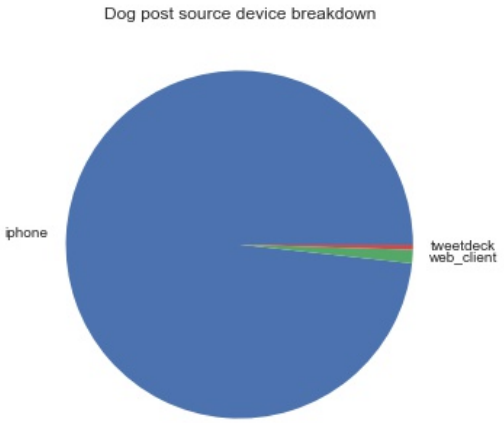

Figure 3: Source device breakdown
`https://github.com/timothyquan/wrangle_and_analyze_data/blob/main/reports/source_device_breakdown.jpg`