act_report

0.1 WeRateDogs Twitter Archive - Act Report

0.2 WeRateDogs Data

February 18, 2021

TheWeRateDogs 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, 2017. 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).
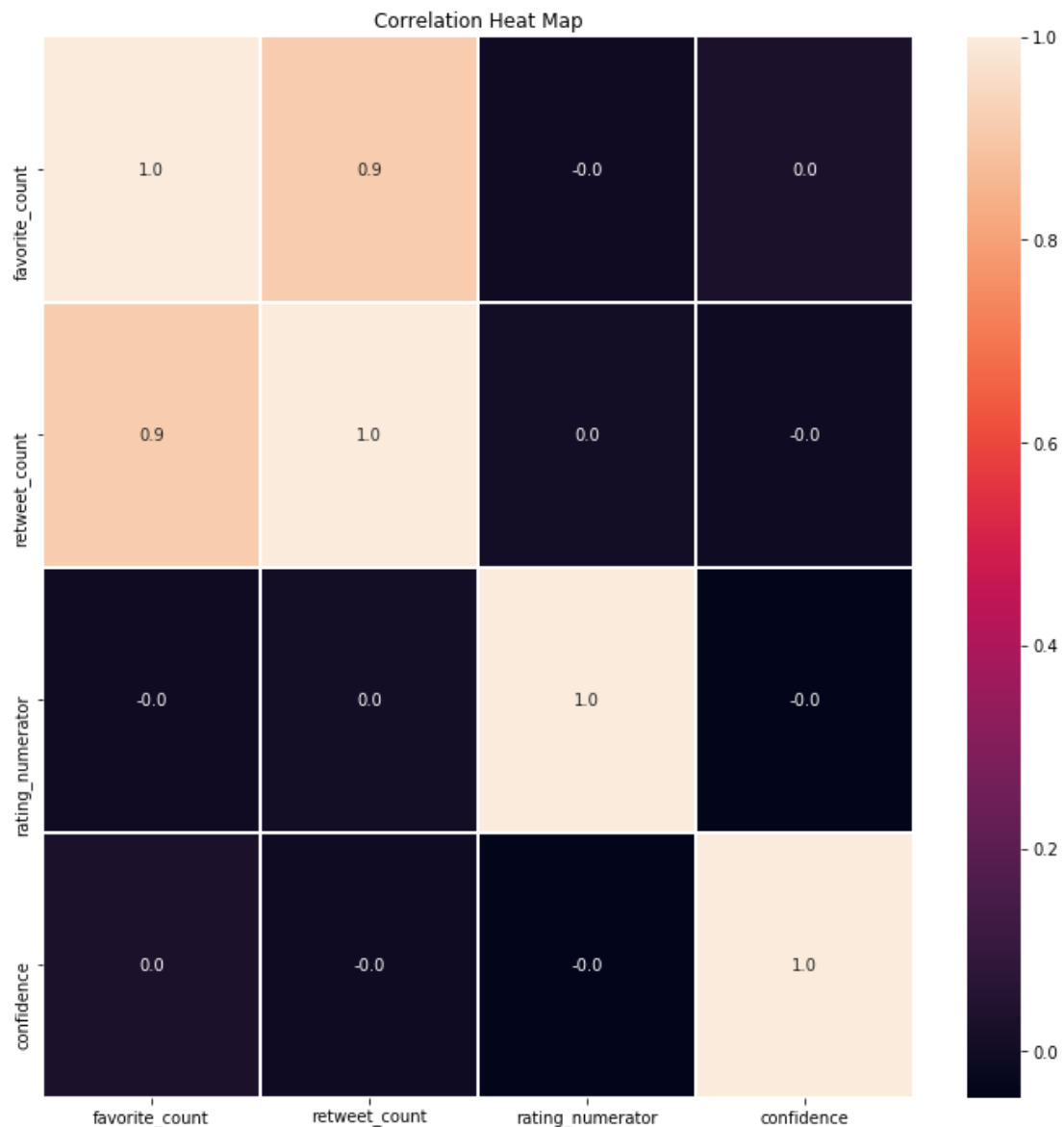
Wrangling Data

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 1955 tweets with good quality data.

Visualization 1

General Insights between different features

This type of visualization called heat map and it's presenting the correlation of different numerical features in percentage e.g the heighst percent is 1 which means the values are 100% identical and as much as the value getting closer to 1 means stronger positive correlation
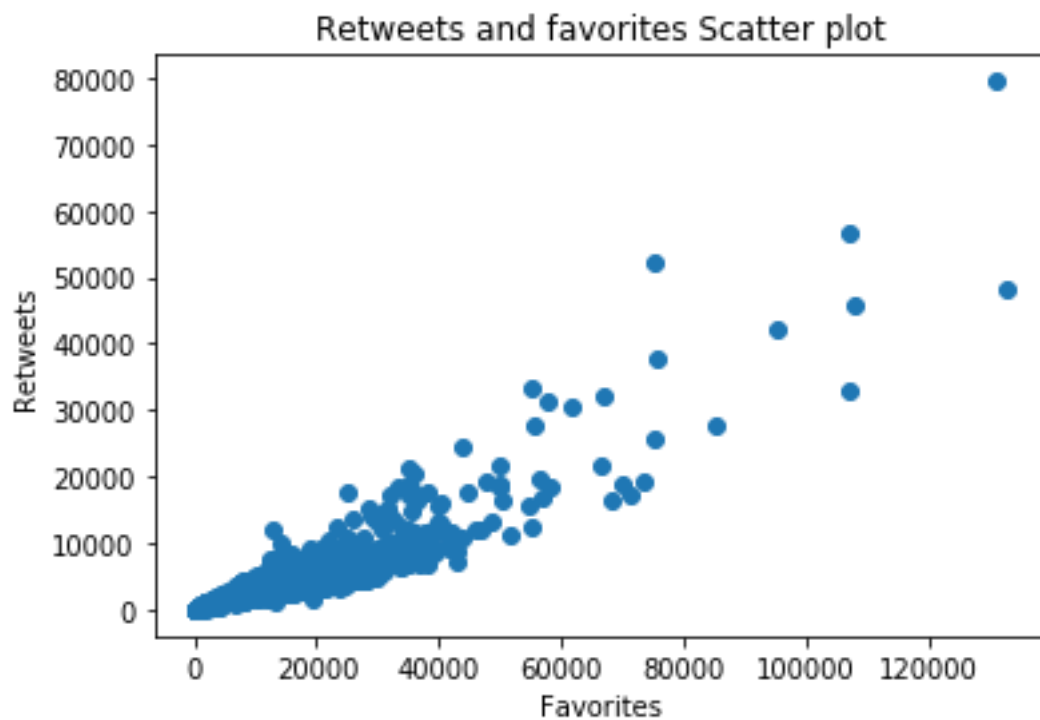
Correlation Heat Map

Strong relationship between favorite count and retweet count

There is a strong relationship between a tweet's retweet counts and favorite counts; in other words if a tweet has a high favorite count it will also have a high retweet count. Plotting these two values against each other produced the plot below, and clearly shows a high correlation by the very tight collection of points along the angled, straight line.

Visualization 2
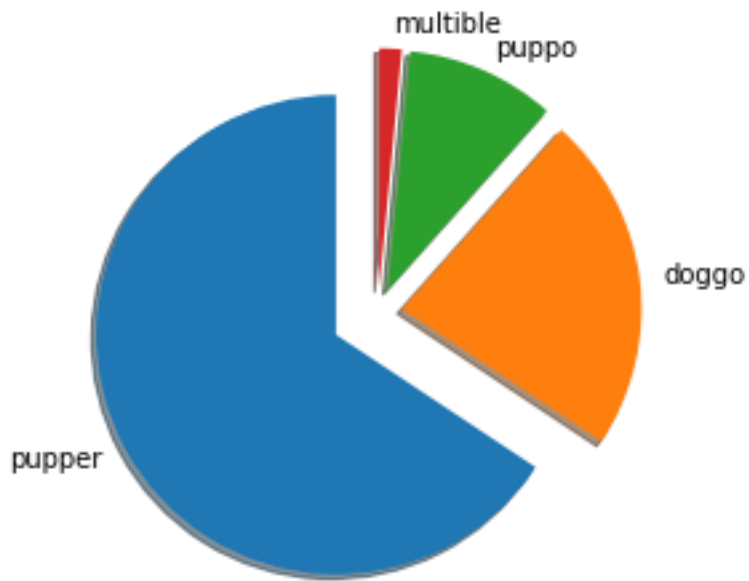
Scatter plot between favorites and retweets

Retweets and favorites Scatter plot

Analyses

As expected there is a big positive (+0.9; shown by correlation heat map) correlation between the retweets and favorites.

Visualization 3

Data partition dogs in pupper stage of dog life cycle get most tweets around two thirds of the favorite counts

Analyses

Dogs in Pupper stage of dog life cycle get most tweets, which is expected.