DATA ANALYTICS NANODEGREE PROGRAM (UDACITY)

WeRateDogs Twitter Archive – Act Report

Osigah Ogedegbe | September 2022

WeRateDogs

WeRateDogs – This twitter archive contains data extracted from 2356 of the 5000+ tweets from the @dog_rates twitter account. This data contains dog ratings that were taken from the text of the tweet along with the dog name and dog classification if present. The retweet count and favorite count was not included in the archive data so it had to be extracted from a json text file provided in the lesson because I was not able to get access to twitter's api.

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.
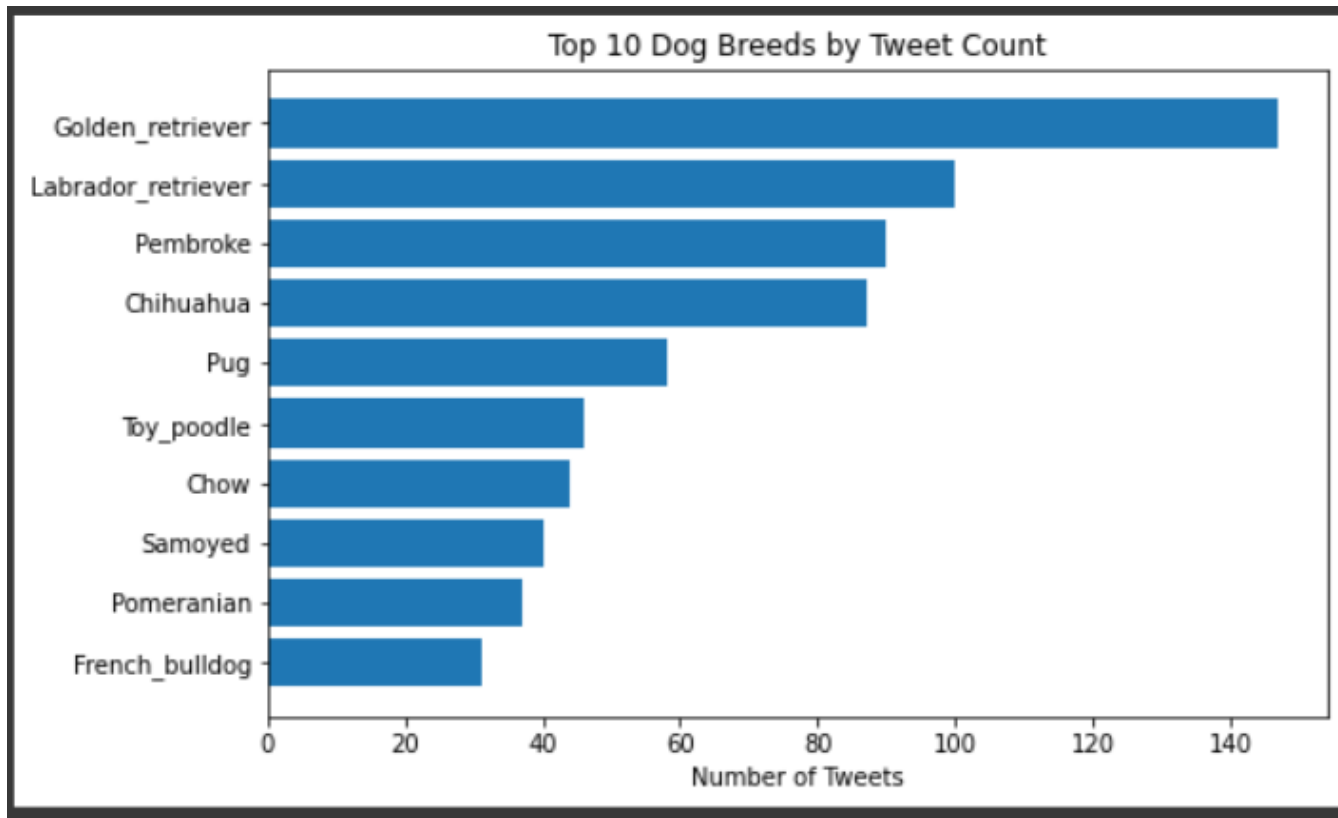
Wrangling of the Data

Prior to analysis, the data had to be wrangled and cleaned first. I assessed for quality and tidiness using both a visual and programmatic approach to ensure it was complete, accurate, valid and consistent as it should be before any analysis is done on it to provide good insights.
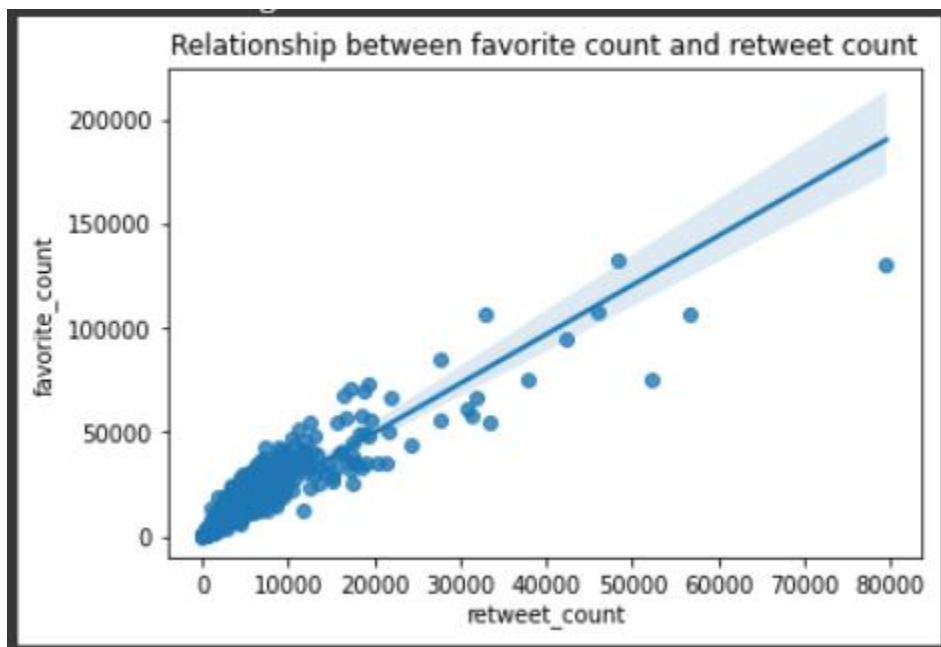
I had around 1854 with quality data after cleaning the issues found in the assessment phase.

Insights Drawn

1. There are more favorites than there are retweets on every tweet with the highest favorite value being 132,810 compared to the highest retweet value which is 79,515. I believe this is fairly normal because naturally people tend to leave likes rather than retweet from my experience on the application.

2. The most common dog people tend to always tweet about was the golden retriever with over 150 tweets about the dog alone. The 10 most common dogs tweeted about were the

   Golden retriever, Labrador retriever, Pembroke, Chihuahua, Pug, Toy poodle, Chow, Samoyed, Pomeranian and French bulldog

Top 10 Dog Breeds by Tweet Count

3. There is a strong relationship between a tweet's retweet counts and favourite counts; in other words if a tweet has a high favourite count it will also have a high retweet count. The plot below shows a high correlation with the angled, straight line.

Relationship between favorite count and retweet count

4.   There are 1482 dogs with a rating of 10 + which is over 80% of the data itself. People like to give high ratings and people just love dogs.