Wrangle Act

Presented by Ali Hesham For Udacity DAND

INTRODUCTION

This purpose of this report is to demonstrate the efforts and the steps done to complete the weRateDogs data wrangling project in Udacity DAND.

This in this project we gathered the data from three different resources all related to twitter account user@dog_rates and is used to gather insights and visualizations on data after doing data wrangling processes.

REPORT DETAILS

In this report I will present the insights found after cleaning the data provided after gathering and assessing

RETWEETS VS FAVORITES

After getting the number of retweets and favorite count from API and merging it with the twitter archive data I wanted to see if there was a relation between number favorite count and retweet count and the conclusion from this scatter plot that there is no relation
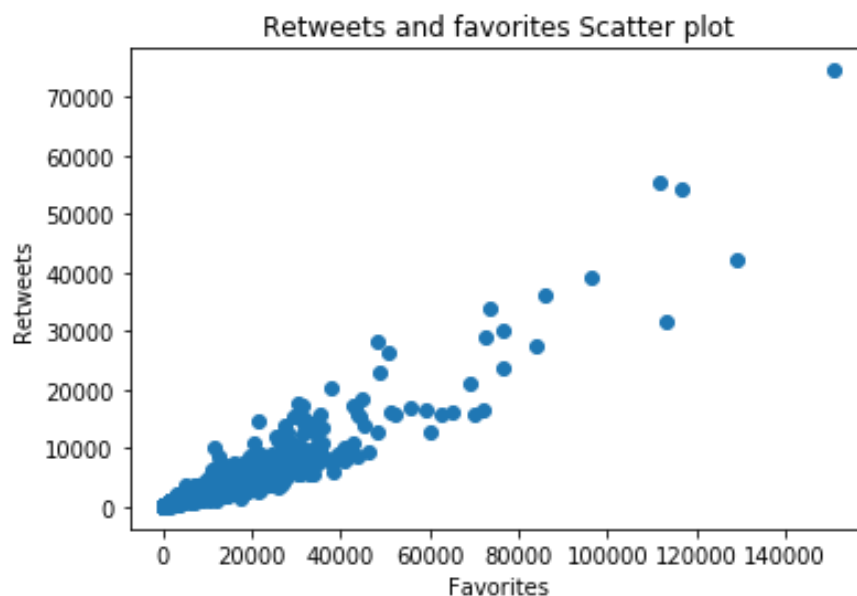
RETWEETS VS RATINGS

Amazingly many tweets have been retweeted more than 1000 times some more than even 50000 times. There is not a clear correlation between the ratings and the retweets.

Retweet Counts by Ratings Scatter Plot
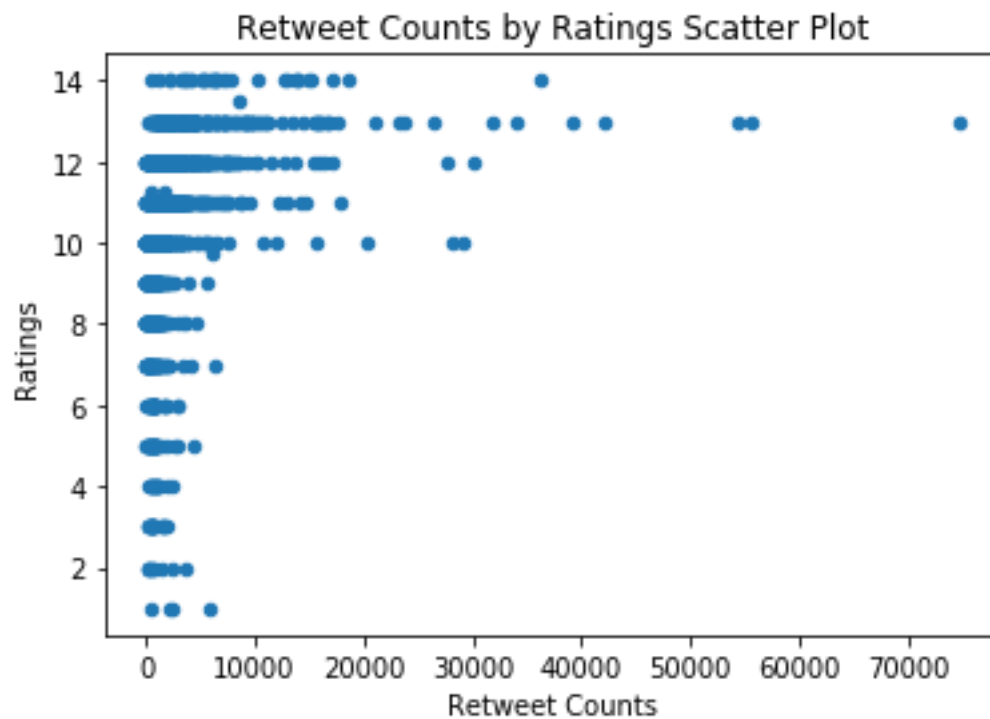
MOST RATED DOG TYPE
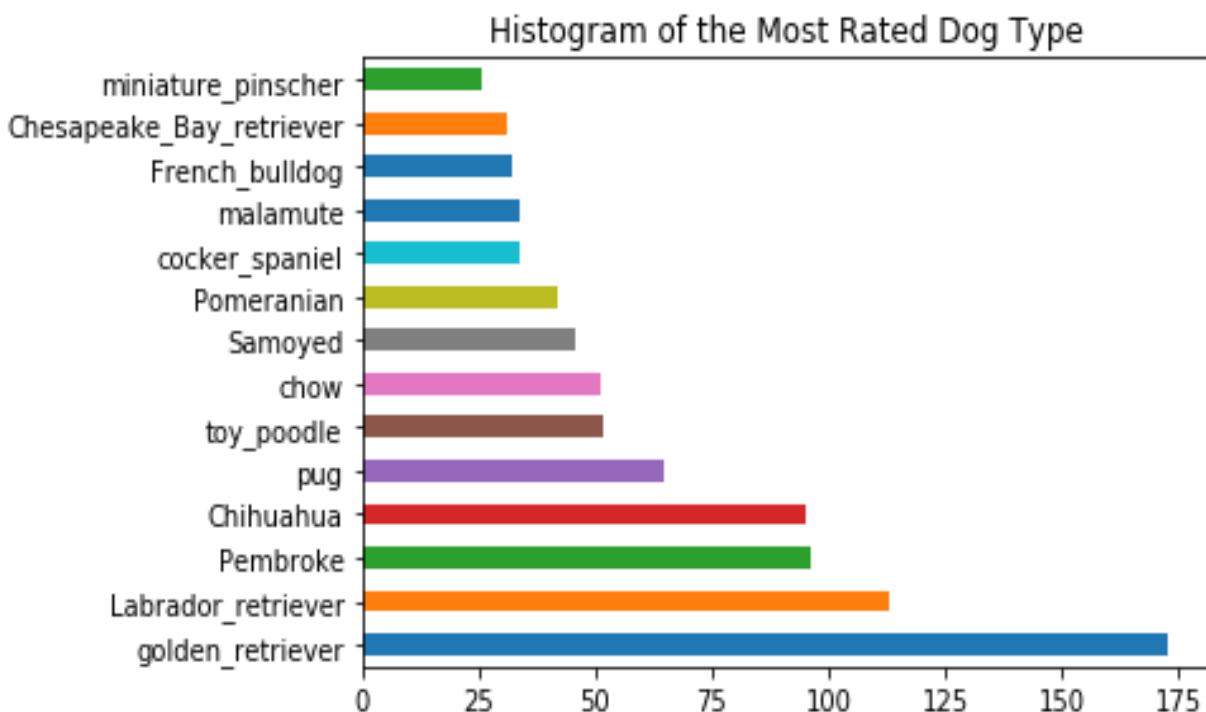
After getting the data related to the dog types from image predictions data, I had to see what was the most common dog type in the presented data set of tweets and from the presented dog types comes in the

1st place golden retriever with 175 counts

2nd is Labrador retriever with nearly 125 counts

3rd place is Pembroke with nearly 100 counts

MOST COMMON DOG STAGE

Since I extracted the stages from the archive and structured it in a new column to adhere with data tidiness, I wanted to see what it the most common dog stage and that was the result and as you can see the most common stage is pupper with 201 counts