Udacity Data Analyst Nanodegree Program

Wrangle and Analyze Data Project

Act Report

Name: *Soyoung Cho*

Introduction

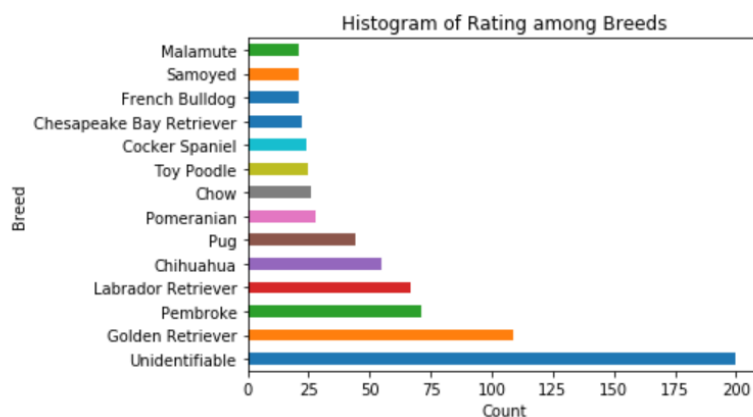I have pulled out 4 insights from the datasets including 2 visualizations.

Descriptive Statistical Analysis

	rating	stage	favorite_count	retweet_count	img_num	confidence
count	1297.000000	0.0	1297.000000	1297.000000	1297.000000	1297.000000
mean	1.223538	NaN	8035.976870	2420.283732	1.187355	0.460126
std	5.039810	NaN	11254.015039	3923.190768	0.541113	0.337762
min	0.100000	NaN	77.000000	11.000000	1.000000	0.000000
25%	1.000000	NaN	1658.000000	555.000000	1.000000	0.140394
50%	1.100000	NaN	3714.000000	1211.000000	1.000000	0.456047
75%	1.200000	NaN	9907.000000	2867.000000	1.000000	0.767563
max	177.600000	NaN	125439.000000	60190.000000	4.000000	0.999956

Table 1. Statistical Analysis table

I will highlight some of the numbers in the above table. First, the mean of rating is 1.22. Second, the mean of favourite count is 8036. Third, the mean of retweet count is 2420.
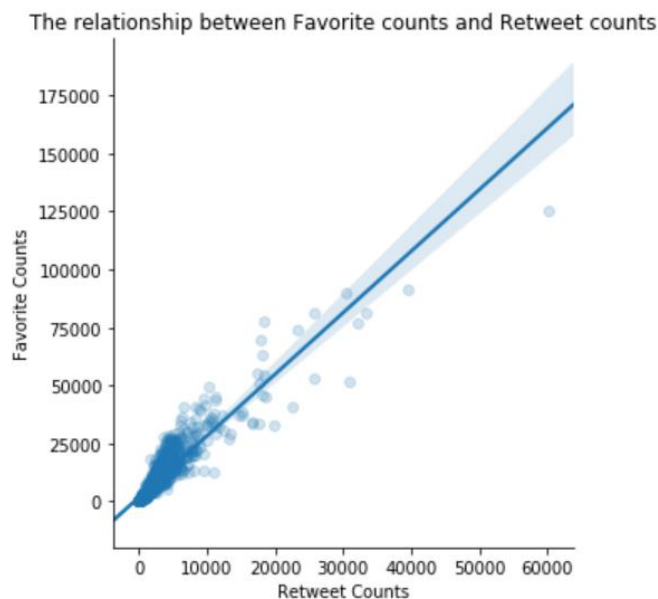
Insight 1. Which type of breeds has highest ratings?

Graph 1. Histogram of Rating among types of Breeds

If you see the histogram above, many of the type of dogs are unidentifiable. However, next to 'identifiable type', Golden Retriever has highest count. The ranking of the breed does not mean they are top breeds. The higher ranking, they are, it refers that they are more commonly shown in the WeRateDogs page.

Insight 2. What is the relationship between the retweet and favourite counts?

Graph 2. The relationship between Favourite counts and Retweet counts

If you see the scatterplot above, the graph is trying to figure the relationship between retweets counts and favourite counts. The line of the graph is heading towards right top corner which refers that the relationship between those two variables are positive. Most of retweets and favourite counts are positioned along and below 20000 and 50000. This can be interpreted as every 1 retweet has approximately 2.5 favourite.

Insight 3. What are the most popular name?

428 tweets were unable to recognize the name of the dog. The top 11 popular dog names are Oliver, Winston, Tucker, Penny, Lucy, Cooper, Lola, Bella, Sadie and Bailey.

Insight 4. Who has received the highest and lowest rating?

The dog name 'Atticus' has received the highest rating of 177.6 (outlier). And three dogs who cannot be identified with their name has lowest rating of 0.1.

Conclusion

In conclusion, I have pulled out four insights by analyzing the data about WeRateDogs. First, Golden Retriever has highest count on the ranking; this refers that they are more commonly shown in the WeRateDogs page. Second, every 1 retweet counts as approximately 2.5 favourite. Third, Oliver is most popular name used in WeRateDogs page. And lastly, the dog called 'Atticus' has received the highest rating of 177.6 while the rating is done from 0 to 10.