Project: Investigating WeRateDogs Twitter Data

Data Wrangling Project Report

Exploratory Data Analysis

Explore data to know if dogs with higher rating had the greatest retweet counts

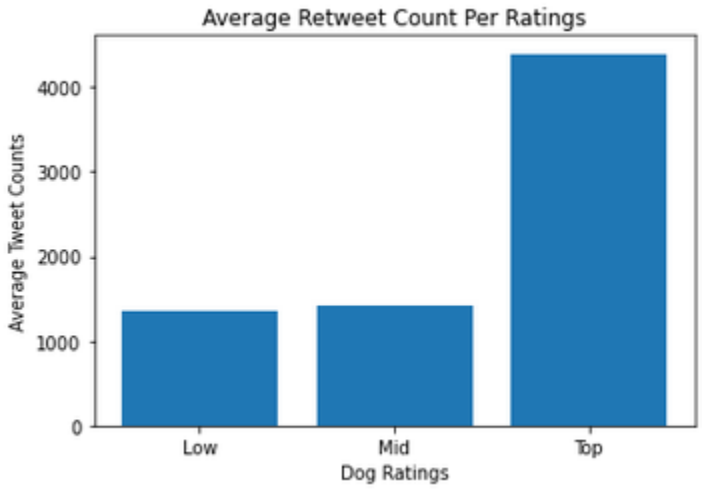
The dogs will be categorized into three levels:

- dogs rated above 10 (Top_rated)
- dogs rated between 6 and 10(Mid_rated)
- dogs rated less than or equal to 5 (Low_rated)

The average retweet count was calculated for each of the category. The average retweet count is shown below:

```
Rating : Average Retweet
TOP:  4395.979951884523
MID:  1428.56640625
LOW:  1362.060975609756
```

Illustrating the outcome of calculations in a Bar Chart



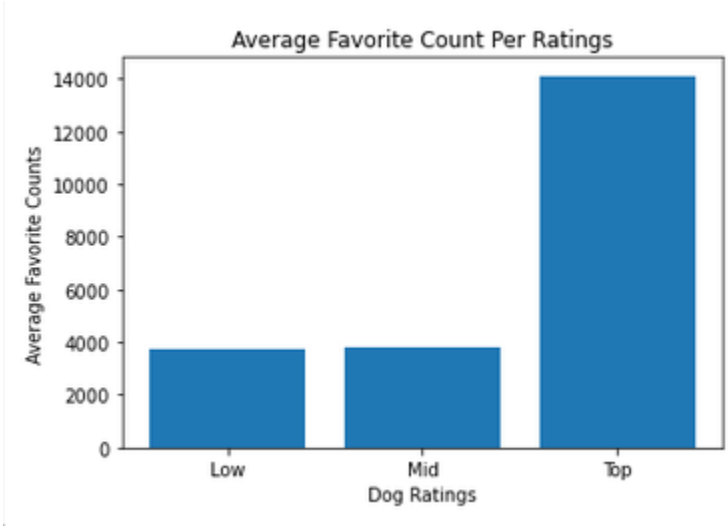
From the results shown above, it can be shown that the Top ratings i.e. those above 10 had the greatest retweets with an average retweet count of about 4400. This may suggest that the model for ratings may be customer centric. Customers will only retweets what they love, and this is somewhat indicative by the ratings.

Explore data to know if dogs with higher rating had the greatest favorite count

The average favorite count was calculated for each category of the dog ratings. The average favorite count is shown below:

```
Rating : Average Favorite Count
TOP:  14116.128307939054
MID:  3836.7135416666665
LOW:  3703.939024390244
```

Illustrating the outcome of calculations in a Bar Chart



The result above also shows that the high rated dogs were the most favorited by the customers. This also indicative that the scoring model is somewhat authentic seeing that the customers favorites these high rated dogs.

It is also seen that the average retweet per dog rating category is higher than the average favorite counts. This may show customers likeliness to retweet a post they love than favoriting by a click.

Explore Extent of Model Consistency

The consistency of the model is assessed by checking the exactness of the model to predict a True or a False over a set of runs.

From the calculation done, the model shows a 75.52% reliability as a function of its consistency.