Wrangle Report

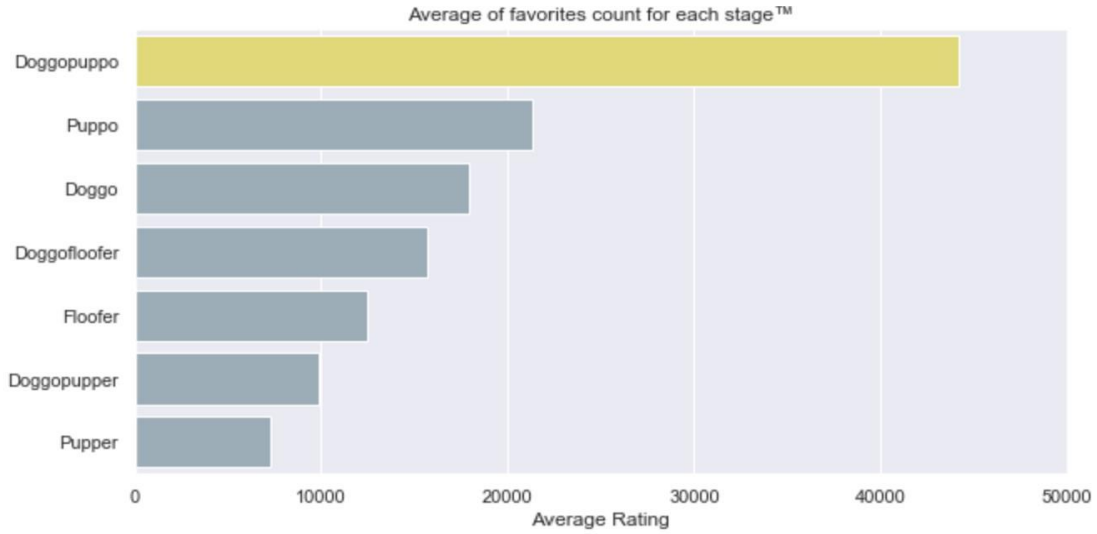
Introduction

This is a report trying to analyze the 3 basic insights of WeRateDogs twitter account and displays the visualizations produced from the wrangled data derived from twitter_archive and image predictions dataset. WeRateDogs is a twitter account where you can send a picture of your dog. Then, WeRateDogs selects some and tweets these pictures with a funny comment and rating. These ratings should vary between 0 to 10 numerators on a 10-point denominator however they often exceed 10 because “They’re good dogs Brent”

Data Analysis and Visualizations

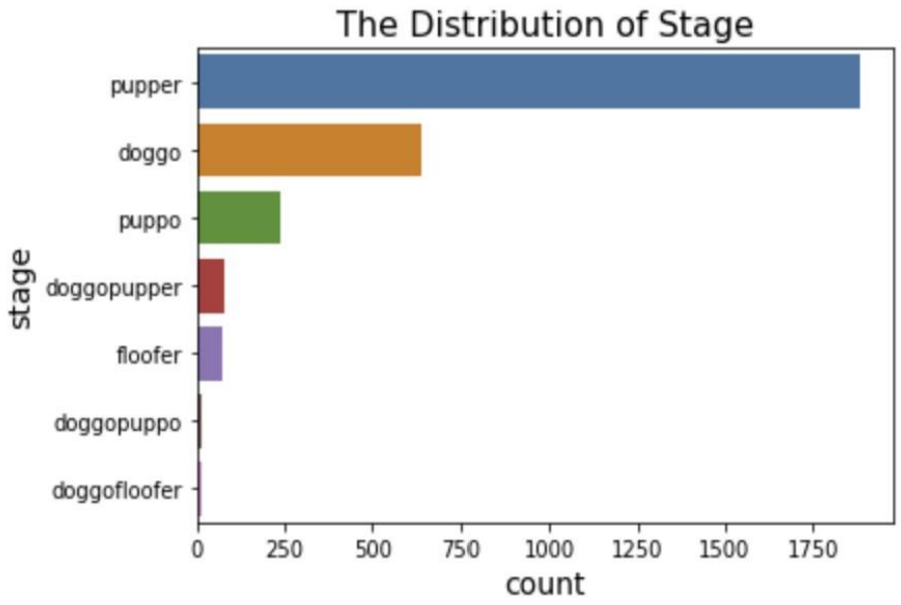
1. **Highest and Lowest Favorite Count on Average?**

To calculate that which dog stages received the highest and lowest favorite count on average, I grouped the dataset by dog stages and computes the average favorite count for each group. Then, I sorted the stage groups by descending order. Afterwards, I plot the bar chart as it shows below. The result illustrates that the doggopuppo and pupper have received the most and the lowest favorite count on average between the other stages with 44217 and 7312 favorite counts respectively.



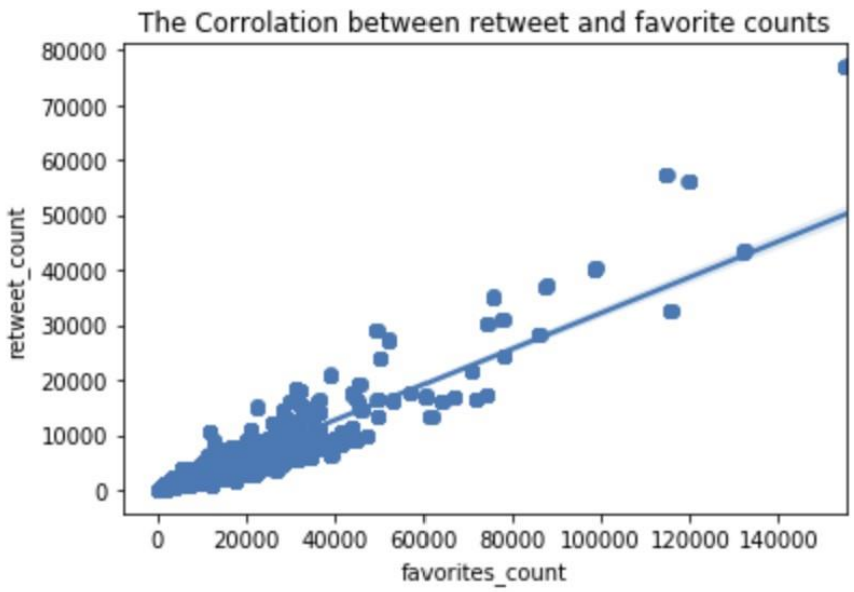
2. Distribution of Dog Stages

The below plot shows the distribution of stage. We can see that opposite the favorite count number pupper has the highest distribution among others which means the higher standard deviation.



3. Correlation between Retweets and Favorites

The below diagram shows that there is a strong positive relationship between favorite count and retweet count. Also, the r square of two variables is 0.92 which proves the high correlation between the two variables. However, it is suggested that to have look to those variables with zero favorite count and considerable retweet numbers, since hypothetically and based on our r square value most popular tweets usually get a larger number of retweets and favorite account.


The Corrolation between retweet and favorite counts
retweet_count
favorites_count
80000
70000
60000
50000
40000
30000
20000
10000
0
0
20000
40000
60000
80000
100000
120000
140000