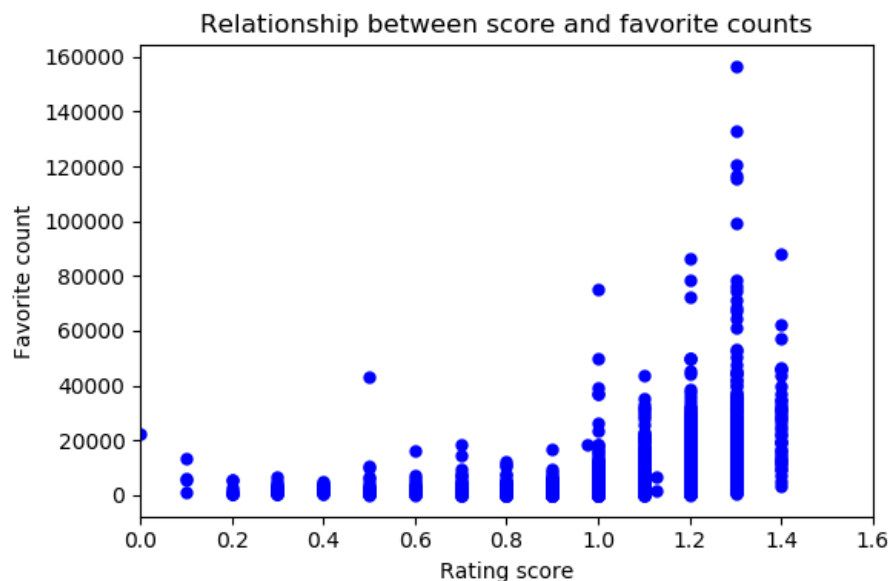# Data insight report
WeRateDog Project

The collected data is a combination of tweets from WeRateDogs that contain posts about dogs. The data was combined with the output of an image recognition algorithm used to identify the dog breed. Also, the retweet and favorite counts were added to the dataset to determine which post are the most popular ones. Below are the few insights based on the clean dataset:

1. Plotting the rating score against the favorite and retweet counts showed that higher rating score usually translated to higher retweet and favorite counts as seen in Figure 1.
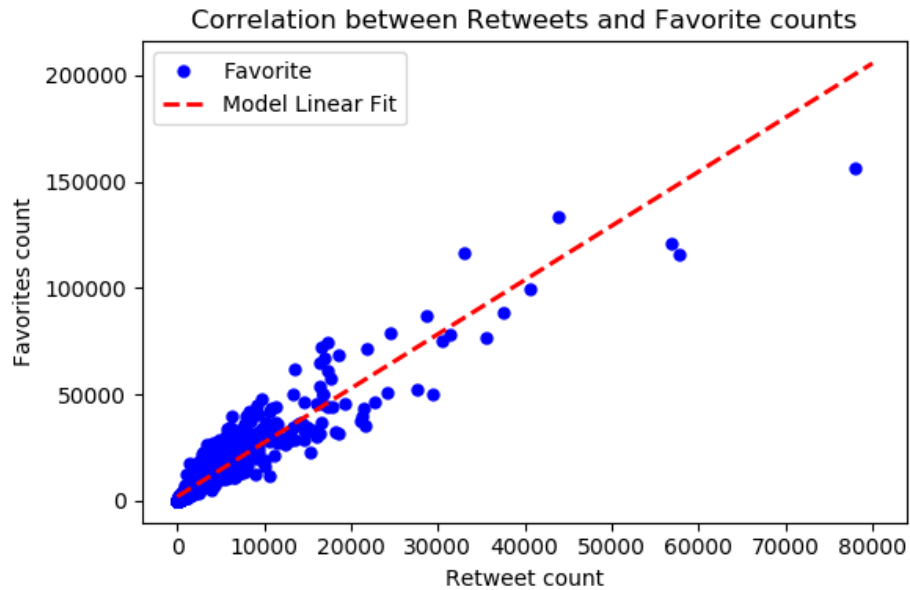

Figure 1: Relationship between rating score and favorite counts
Note dog with highest score (1776/10) is not shown since is beyond the rating scale the scale
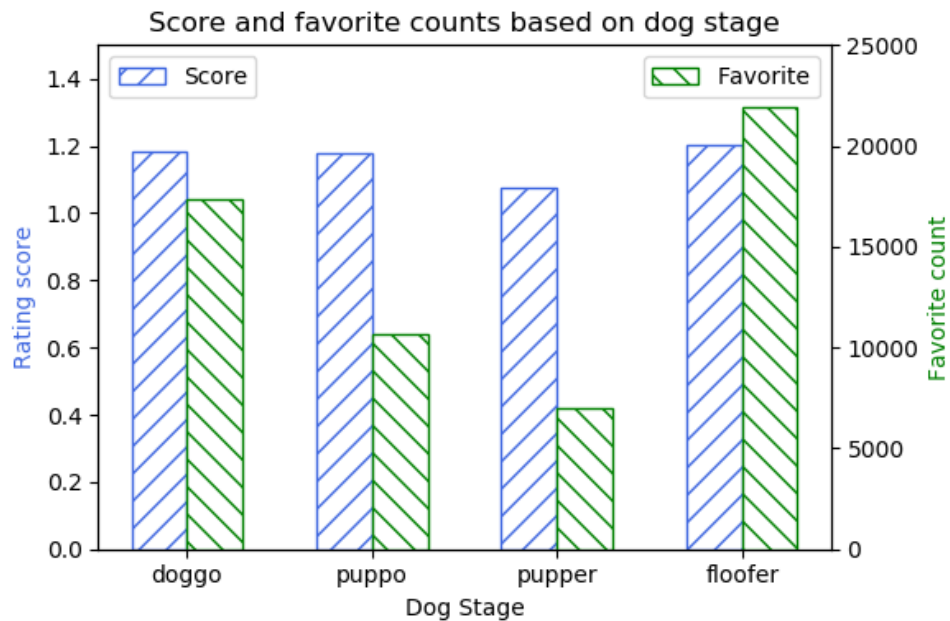
2. The data also showed that tweets with high retweet counts also have high favorite counts following a liner trend (as seen in Figure 2) with a R-squared of 0.859. Even though this seems to be a strong correlation, it is important to keep in mind that correlation does not imply causation. There are other factors not considered in this study that could have influence the correlation.

    Since the retweets and favorite counts have a linear correlation, the rest of the analysis will focus on the favorite counts only.
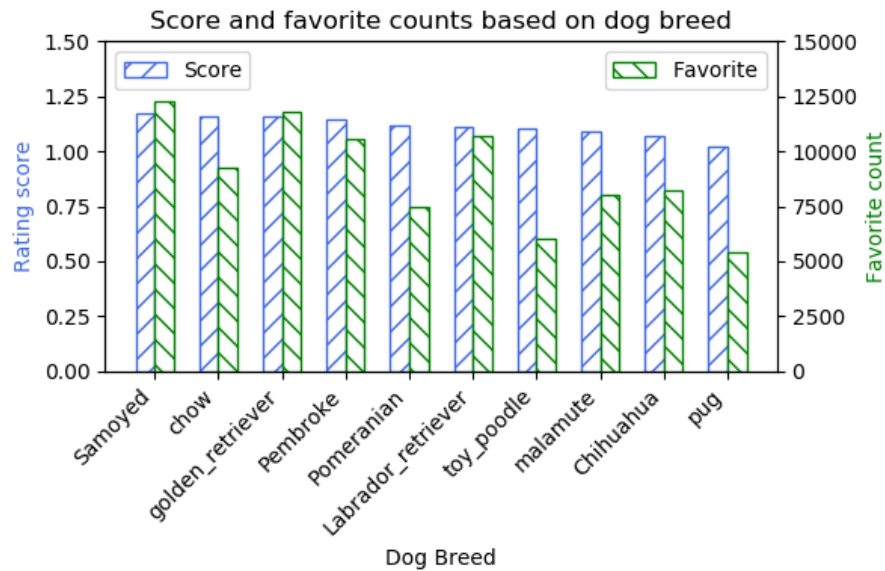
Figure 2: Linear correlation between the retweet counts and the favorite counts

3. Looking at the average rating score for the different dog stages reveals that puppers tend to have slightly lower rating score while the others dog stages seems to be around the same value, with floofer slightly higher. The data also showed that pupper seems to have the lowest favorite ratings while floofer have the highest, similar trend as the rating score.

Figure 3: Rating score and favorite counts for the different dog stages.

4. From the 10 dog breeds with that larger number of "mentions" within the data, the Samoyed has the highest favorite count as well as the highest ratings as seen in Figure 4. Pug seems to have the lowest score and favorite counts.

Figure 4: Rating score and favorite counts for different dog breed.
Breeds were selected as the 10 dog breeds with the larger number of "mentions"

5. The dog with the highest score is Atticus (Figure 5) with a score of 1776/10. However, that is not the post of the highest retweet and favorite counts.

Figure 5: Atticus, dog with highest rating: 1776/10

6. The tweet with the highest favorite and retweet counts was about a doggo in a pool with a score of 13/10 (Figure 6). The retweet and favorite counts were 77,862 and 156,327 respectively.

Figure 6: Dog with the highest favorite and retweet counts