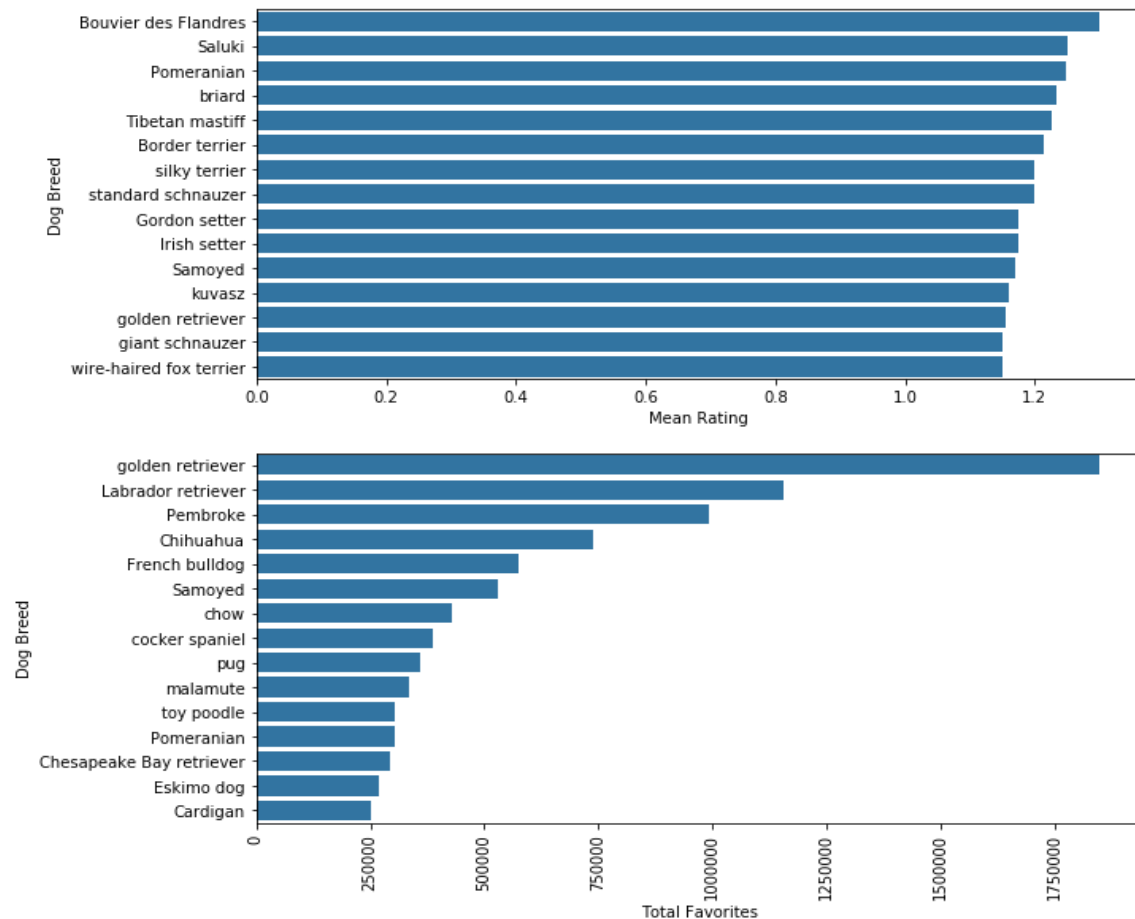


Act Report

Analysis and Visualization

- Insight 1

Which dog breed in the dataset is the most favored?

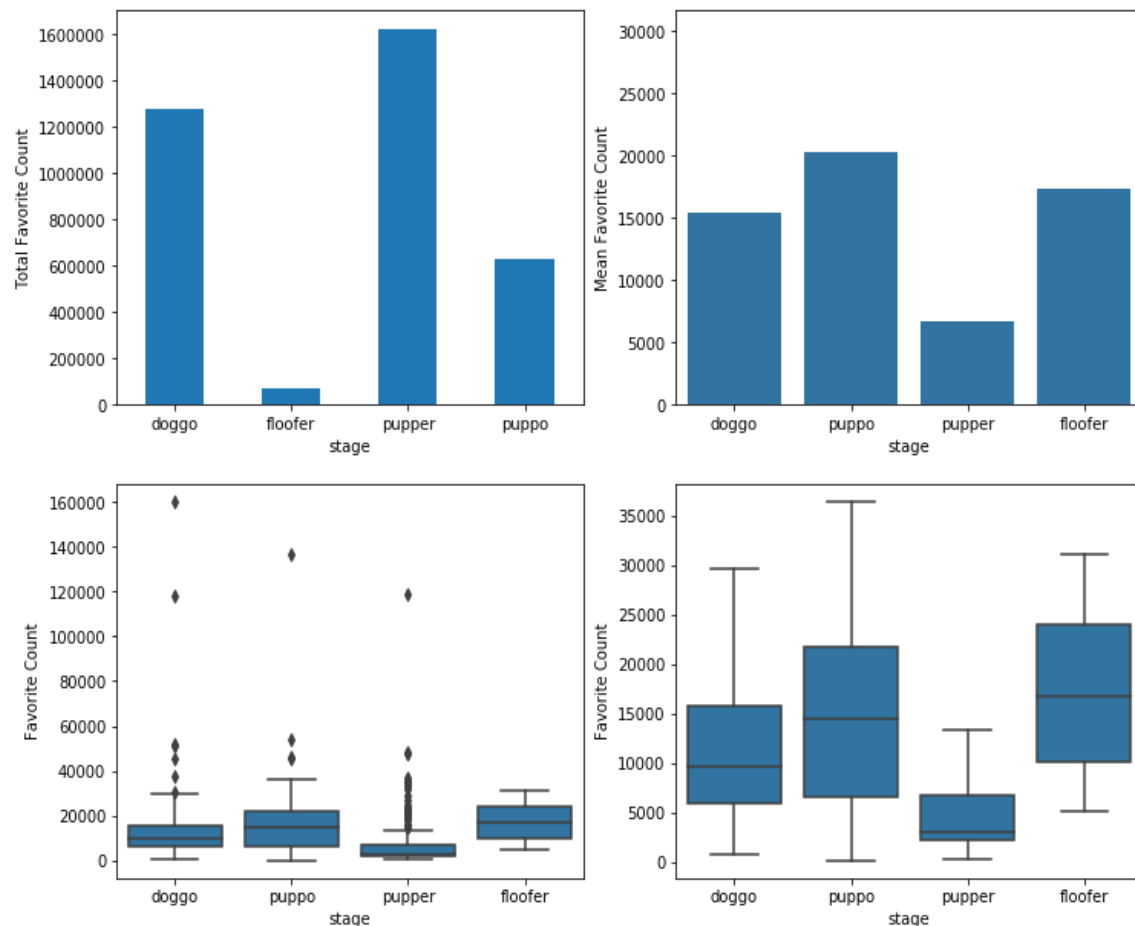


From the above visual, of all the dog breeds in the dataset, the **Bouvier des Flandres** had the highest mean rating. Also, the **Golden retriever** breed had, by far, the highest number of total favorites of any other breed.

However, since the rating system in WeRateDogs is totally arbitrary and doesn't conform a defined set of rules of rating, it makes more sense to use the total number of favorites as a statistical metric to determine the popularity of each dog breed. So it is more convenient to consider the **Golden retriever** breed to be the most favored breed in the dataset.

- **Insight 2**

Which of the dog stages in the dataset is the most favored?

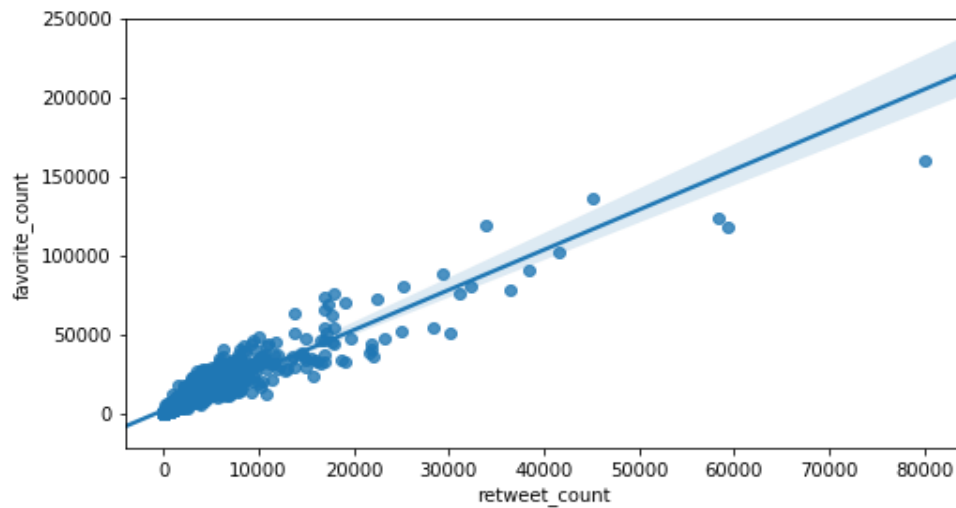
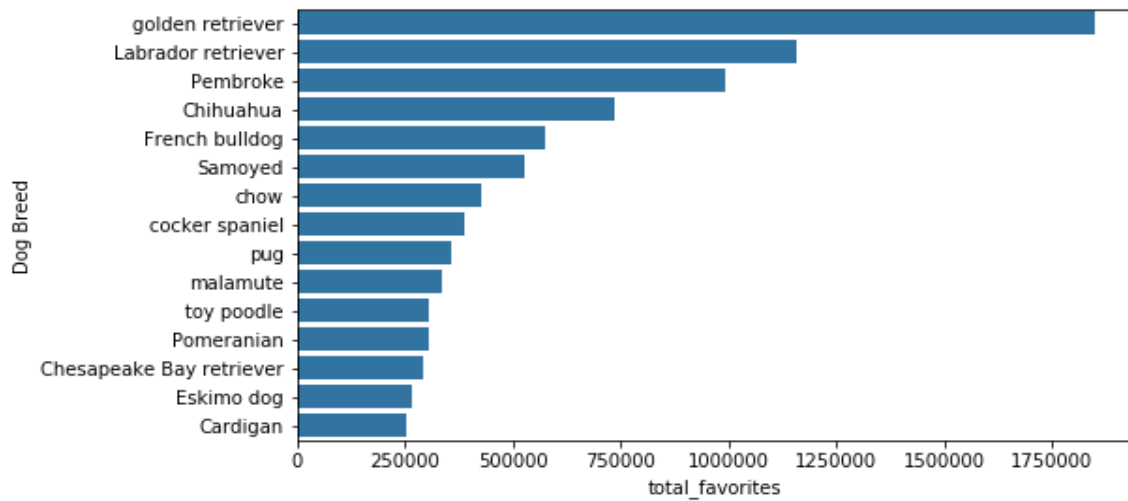
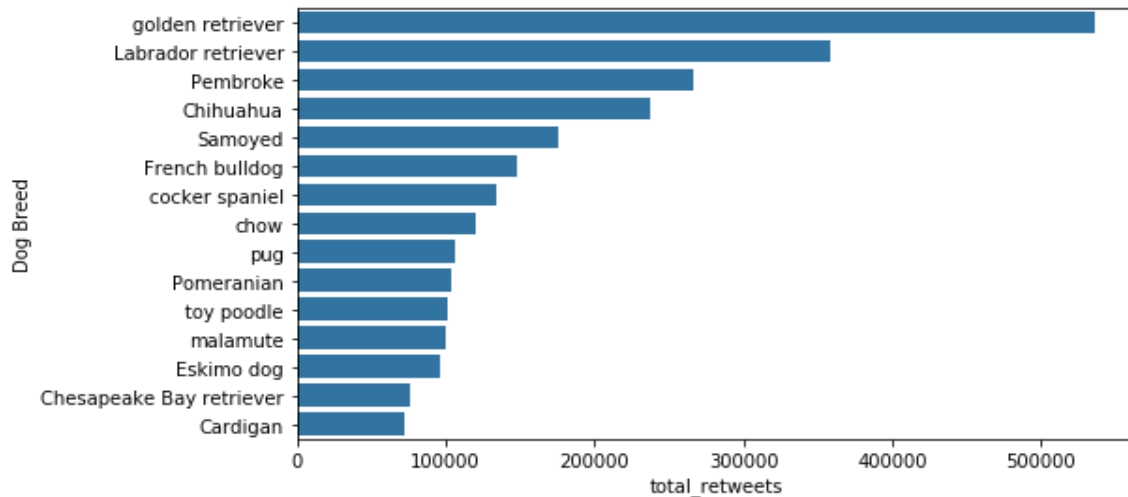


At first glance, looking at the total favorite count in the first visual, it looks like **pupper** dogs are the most favored ones with total number of favorites approaching 1600000. However, further investigation of data reveals that this is probably due to the presence of outliers that causes the total number of favorites to be so high. This is confirmed when the stage of dogs is plotted against the mean favorite count in the second visual, and it shows clearly that **pupper** dogs are actually the least favorite, while the most favored dogs are the **puppo** dogs.

The effect of outliers is clear in the box plots, as the one on the right confirms that the high number of the total favorite count of the **pupper** dogs was mostly due to outliers, which when removed, the **puppo** dogs turned out to be the most favored dogs with the **pupper** being the least.

- Insight 3

What is the relationship between the retweet count and the favorite count?



From the above visual, there seem to be a direct positive relationship between the number of retweets and the number of favorites. The **Golden retriever** dog breed had the highest number of favorites as well as the highest number of retweets. This pattern stands true for, at least, the top 15 dog breeds. This correlation is also confirmed by the scatter plot which clearly demonstrates the positive and somewhat strong relationship between the number of retweets and the number of favorites.