

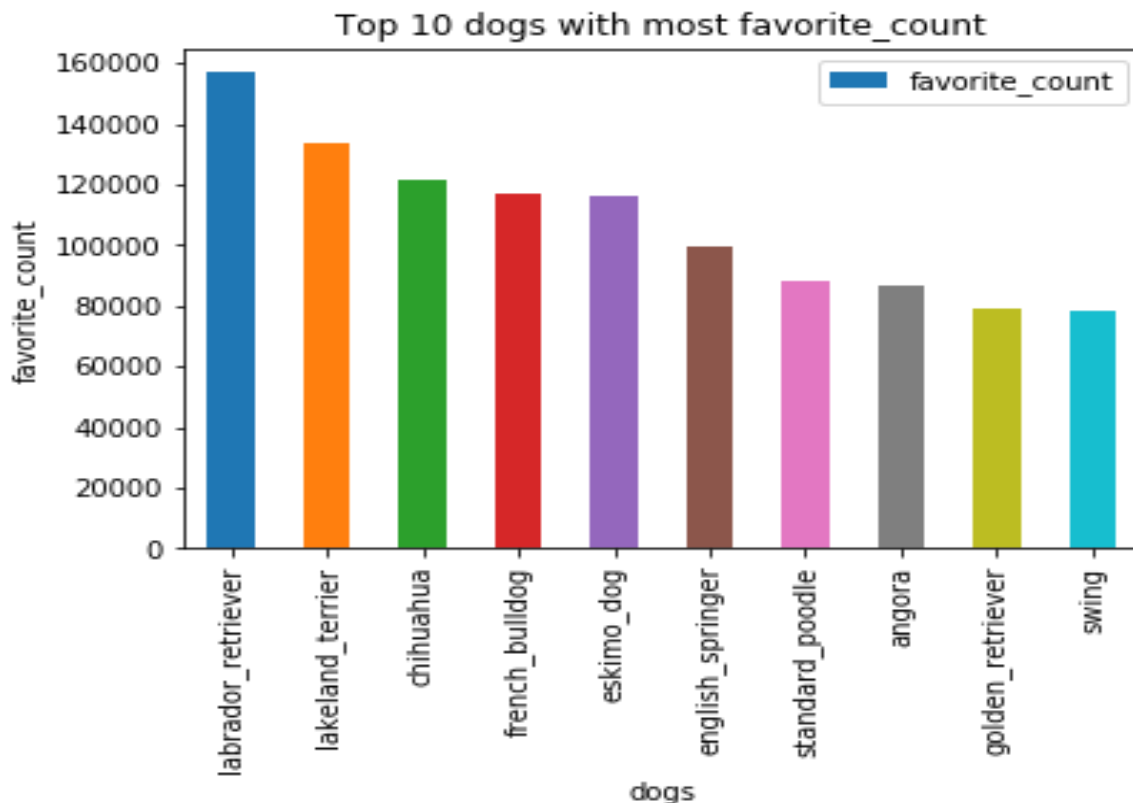
Introduction

The dataset that will be wrangling and analyzing is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs has over 4 million followers and has received international media coverage.

The following are some visualizations and insights that found by analyzing this dataset.

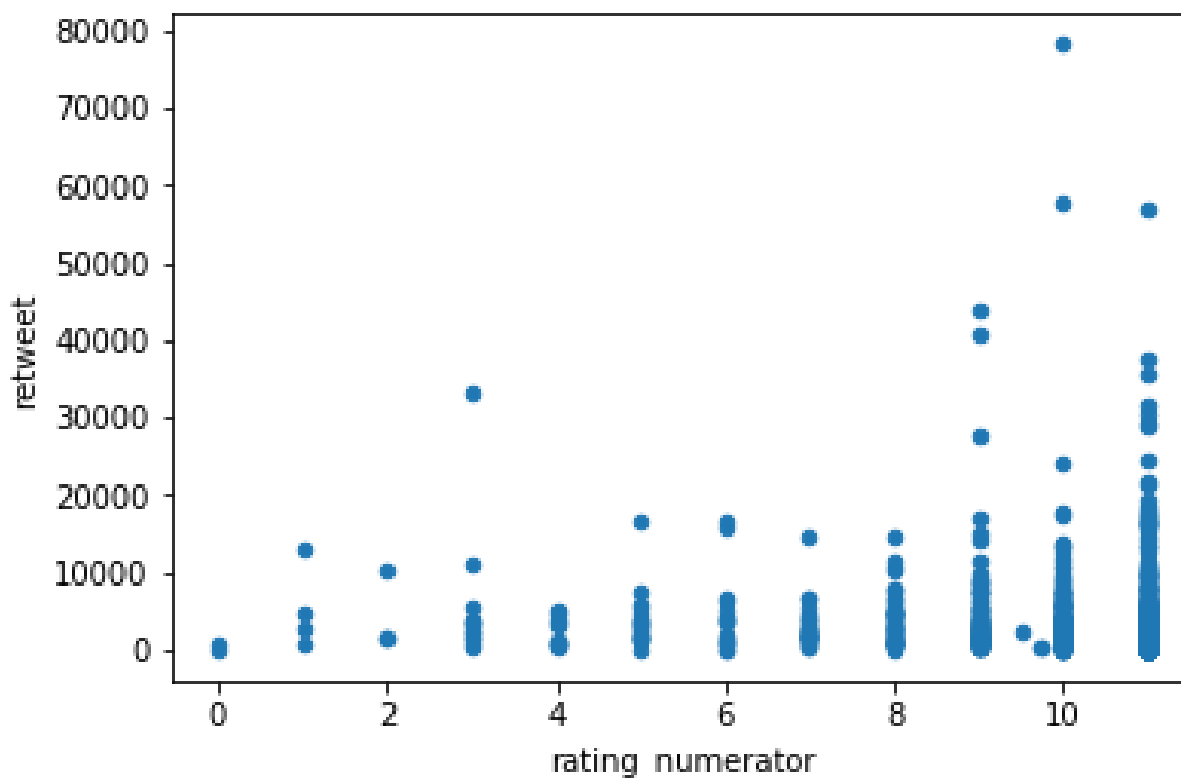
1. Which breeds of dog got the most favorites? Top 10

| | p1 | favorite_count |
|------|--------------------|----------------|
| 1221 | labrador_retriever | 156871.0 |
| 1690 | lakeland_terrier | 133604.0 |
| 1599 | chihuahua | 121177.0 |
| 1898 | french_bulldog | 116837.0 |
| 1186 | eskimo_dog | 115864.0 |
| 1950 | english_springer | 99619.0 |
| 1666 | standard_poodle | 88434.0 |
| 1870 | angora | 86849.0 |
| 1914 | golden_retriever | 78803.0 |
| 571 | swing | 78652.0 |



From the bar chart and table, we can see that the Labrador Retriever got the most favorites, which is around 156971 favorites. It got 23267 more favorites than the second place (Lakeland Terrier). There are small differences between the No.9 and No.10, with only 151.

2. Do higher rated pets have more retweets?



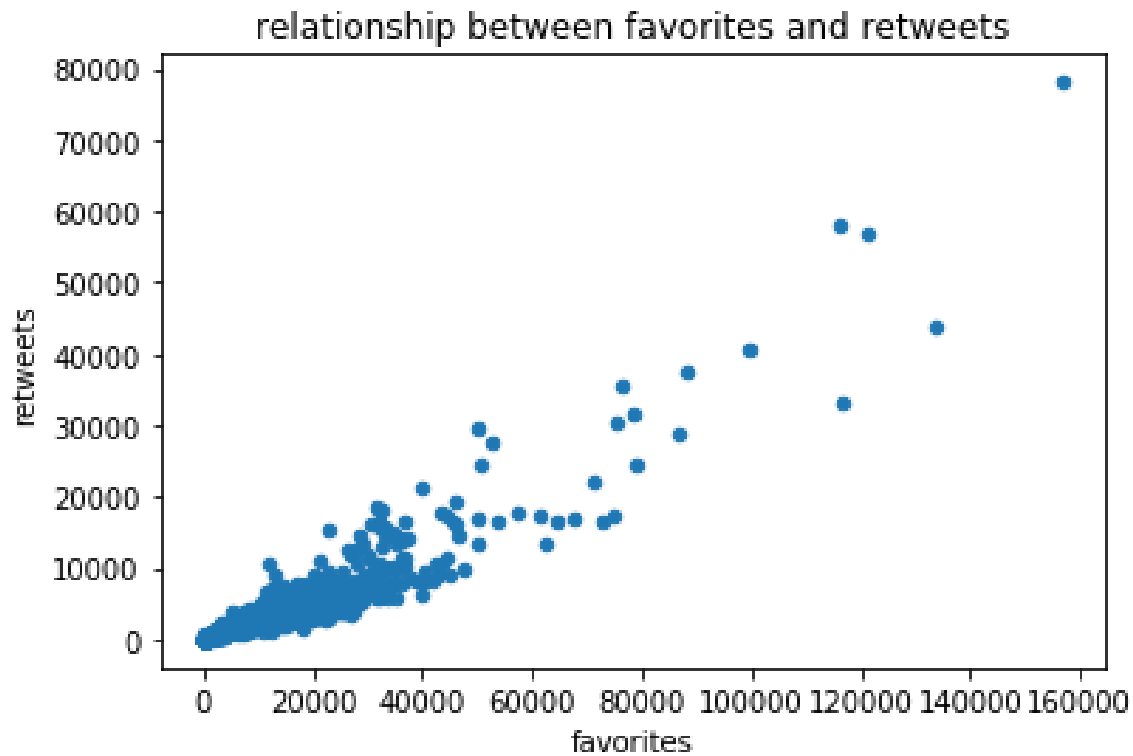
As can be seen from the scatter plot, higher rating numerator is likely to have higher retweets. The rating numerator is positively related to the retweet_account.

3. Which stages of dog have more favorites?

| | |
|---------|--------------|
| doggo | 3052.129032 |
| floofer | 2603.500000 |
| pupper | 10570.268199 |
| puppo | 3092.972973 |

As can be seen from the table, the stage Pupper got the most favorites compare to other stages. The differences between doggo and Puppo is very small. Floofer is the stage that has lowest favorites.

4. Do higher favorites are likely producing higher retweet?



According to the scatter graph, we can see that higher favorites will likely to have more retweets. Favorites is positively related to retweets.

Limitations:

1. There are some outliers in the dataset, it may affect the accuracy of result
2. There are some rating numerators are very high, so I put those in a group and set the rating to 11 to avoid outliers.
3. There are some missing data, so the result may not covered all dataset.