WeRateDogs

Analysis and Visualization report

Introduction:

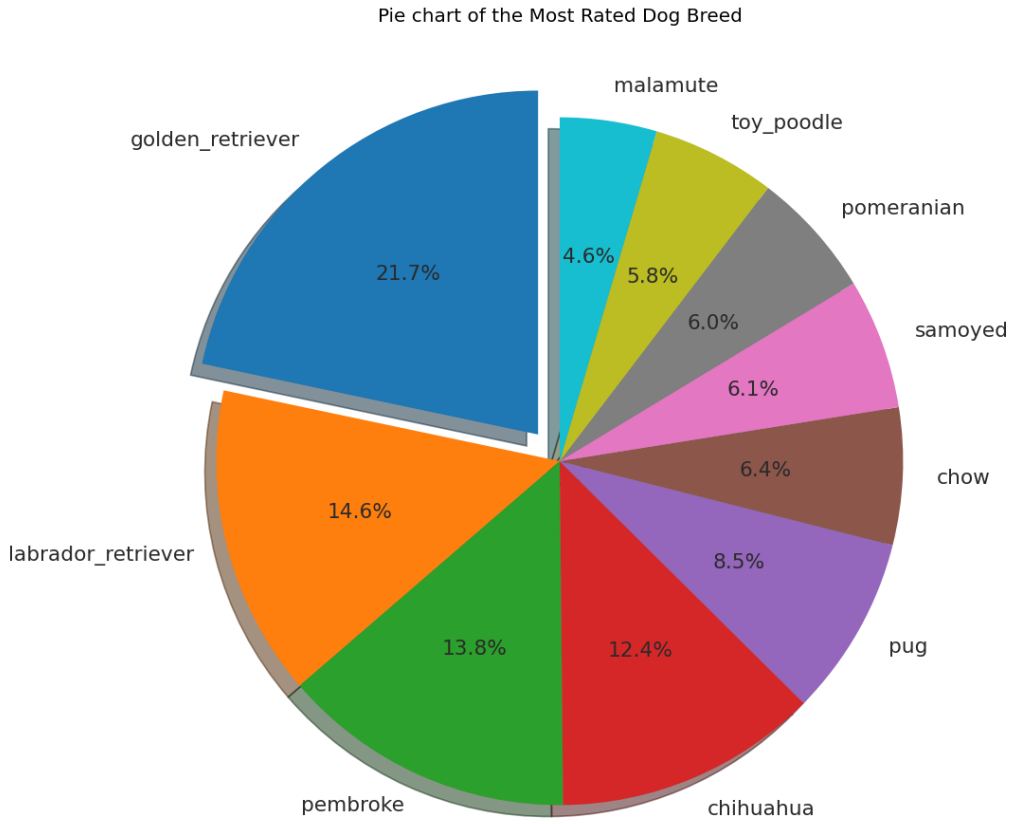
WeRateDogs is a twitter account that has 8.8 million followers, they rate people's dogs with funny comments, it started at 2015, it was growing so fast, it is known with its strange rating system, the rating is out of 10, but most of rates are above 10. It is unique and interesting. There was a lot of work to do to gather, assess and clean the data to make it ready for analysis and virtualization. I end with 1677 record in the dataset.

Let's see what we got from the analysis and visualization, and we have some question that we need to answer.

Analysis and Visualization:

1) The most common dog breed in the dataset

There are a lot of breeds in this dataset but of course there are some breeds is frequently rated in WeRateDogs account. You can know what breed is more popular. The next figure will answer is question.

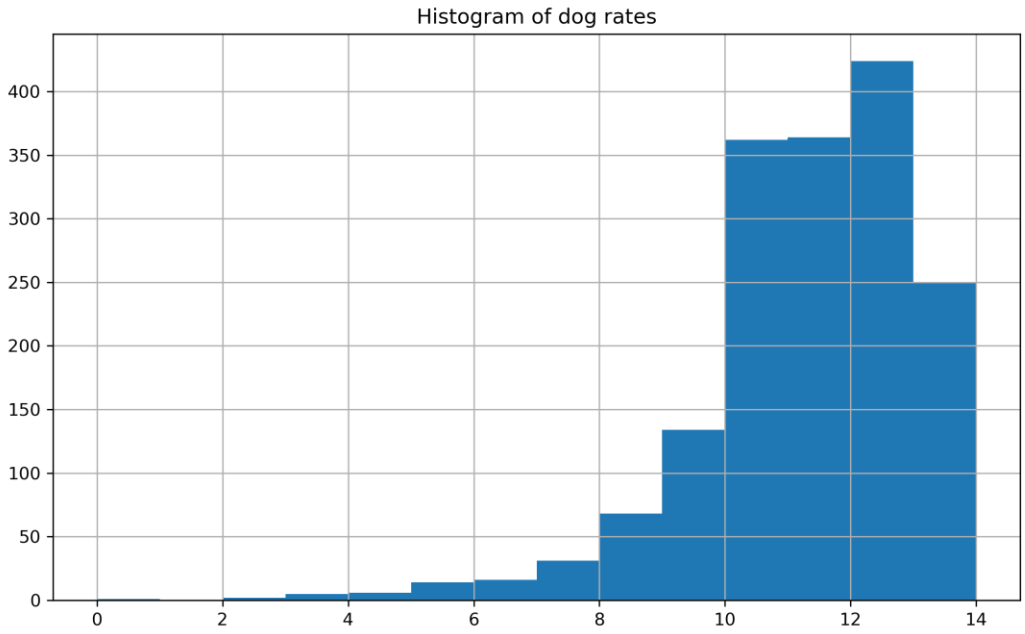


Golden Retriever is the most common dog breed in the dataset with 21.7%, and the second dog is Labrador retriever with 14.6%.

2) The rates majority

As we know there is a strange and unique rating system, what is the percent of rates above 10, is they are large amount or it is rarely happened to get a rate above 10. The next figure illustrate that most of rates is higher than 10, therefor the rates higher than 10 are not dedicated for very special dogs.

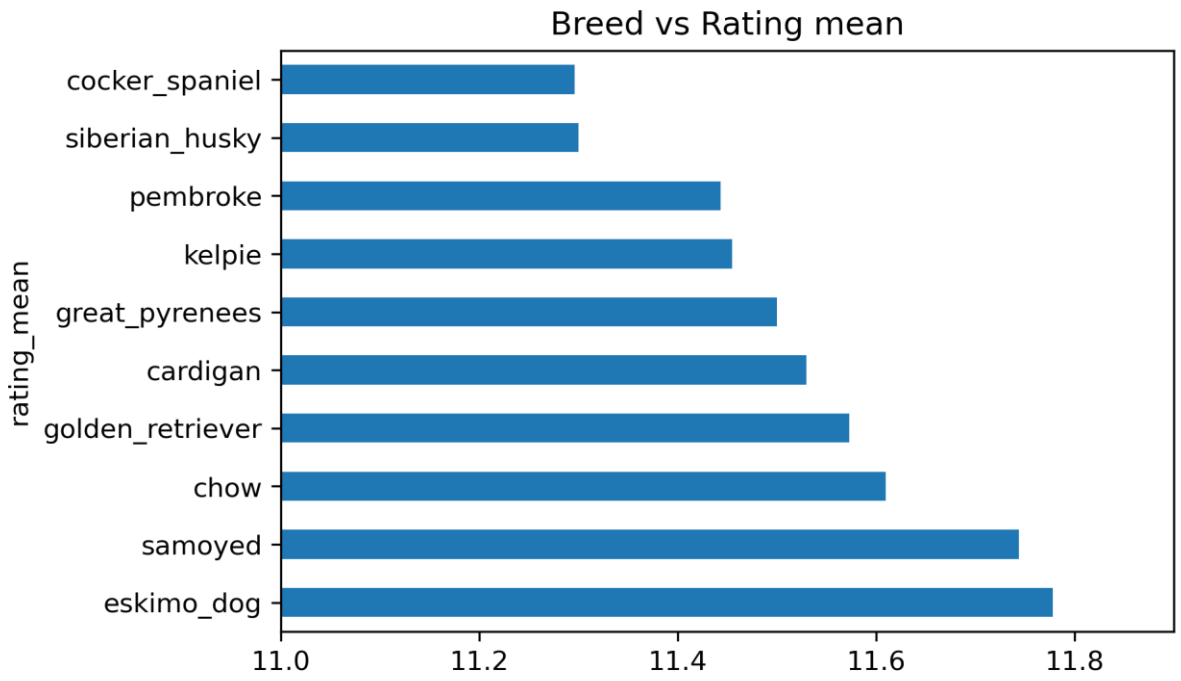
61.8% of rates are above 10, the account is biased to give a dog a high rate. The dog has a high chance to high rate.



In this histogram chart explains that the most of dogs have rates above 10.

3) The breed that has the highest mean value of its rates

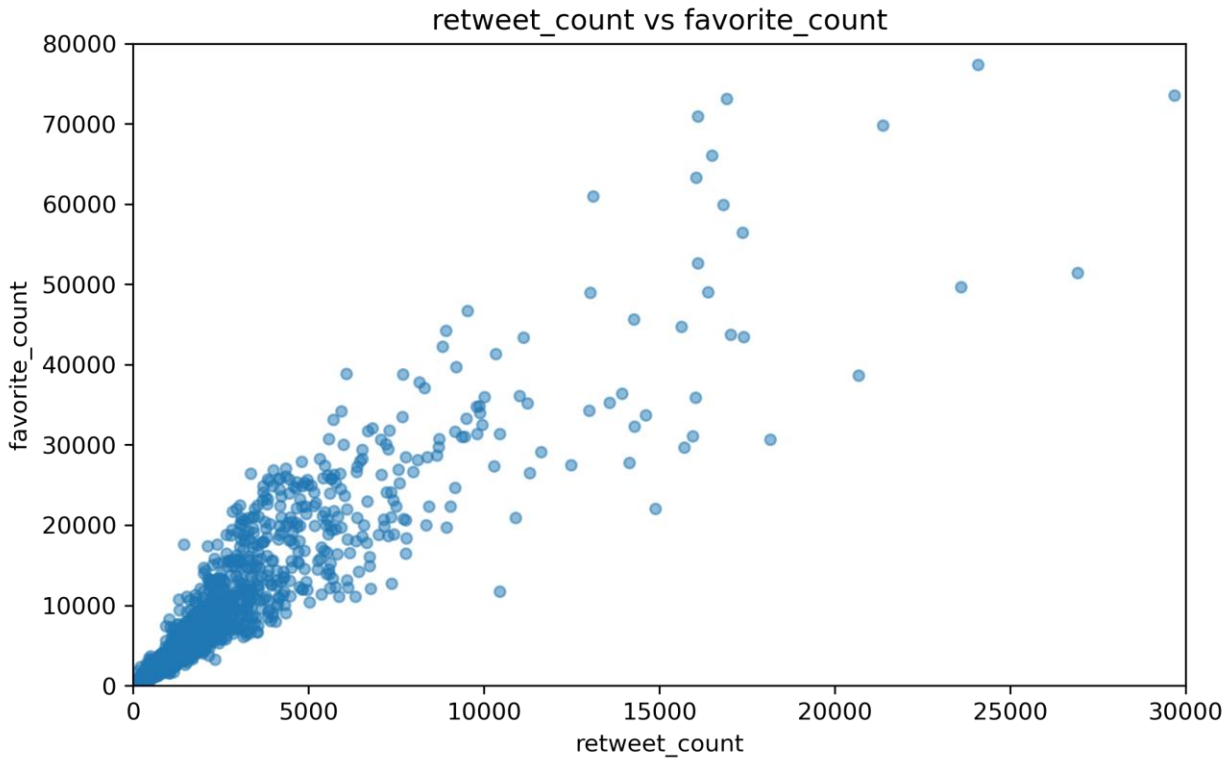
The rates is good indicator for which breed is popular and loved from the people, next figure has top 10 breeds that have highest mean of its rates



Eskimo dog has the highest mean rate value with 11.7, he is so cute btw.
The difference between each one is very small.

4) The relation between retweet_count and favorite_count

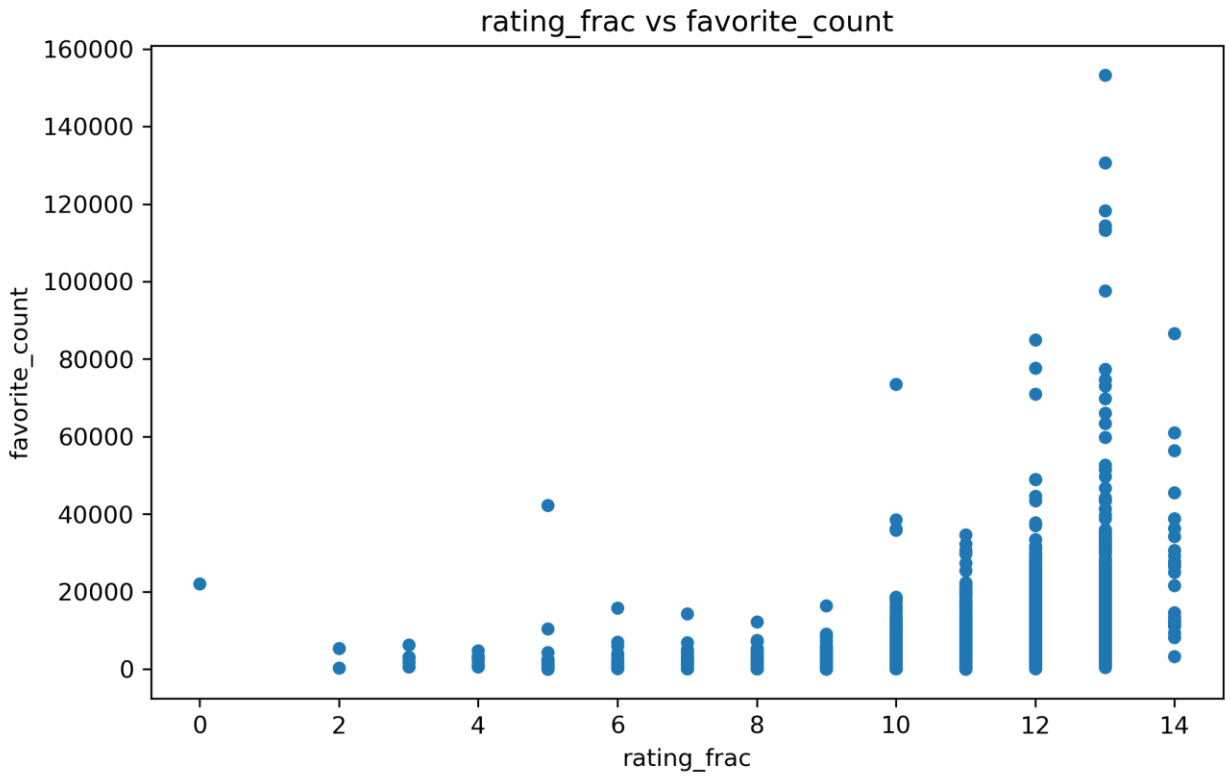
Is there any correlation between number of retweets and number of likes on a tweet? Let's see in the next figure.



As expected, there is a strong relation between the number of retweets and number of likes, it is positive correlation, when the retweets increase, therefore the likes increase on the same tweet.

5) The relation between Rates and likes count

We need to know if there are a strong relation between the dog rate and the number of likes on this rate. Is the rate obviously effect on the number of likes on the tweets or not? Let's see the next figure.



There is weak positive correlation between them, there are some tweets with high and low rates have the same number of likes. Therefore, a high rate does not guarantee a high number of likes by 100%.