Project : WeRateDogs Data Analytics
By Hester Lim Tze Hung

The purpose of this project is to wrangle the tweet archive of Twitter user @dog_rates also know as WeRateDogs to recreate interesting and trustworthy analyses and visualizations. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators are almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

With that, I have developed a few simple yet interesting analysis on WeRateDogs Twitter data.

1. What is the type of the dog that has the highest retweet?

I am really curious about the dog breed of the most popular dog on Twitter. Who knows? Surprisingly based on my research, Labrador Retriever was the name that has the Twitter world going buzz for accounting for 83752 retweets (Imagine people watching Dogs picture instead of cats!). Other than that, the images are all fed through a neural network (an AI prediction technique) and it was a Labrador Retriever with 82.5 confidence level
Then, I started to ponder the common dog names and asked :

2. What is the most common dog names?

Surprisingly, there were 4 dogs that tie for the number 1 spot

Here is the result :-

- Oliver
- Cooper
- Charlie
- Lucy

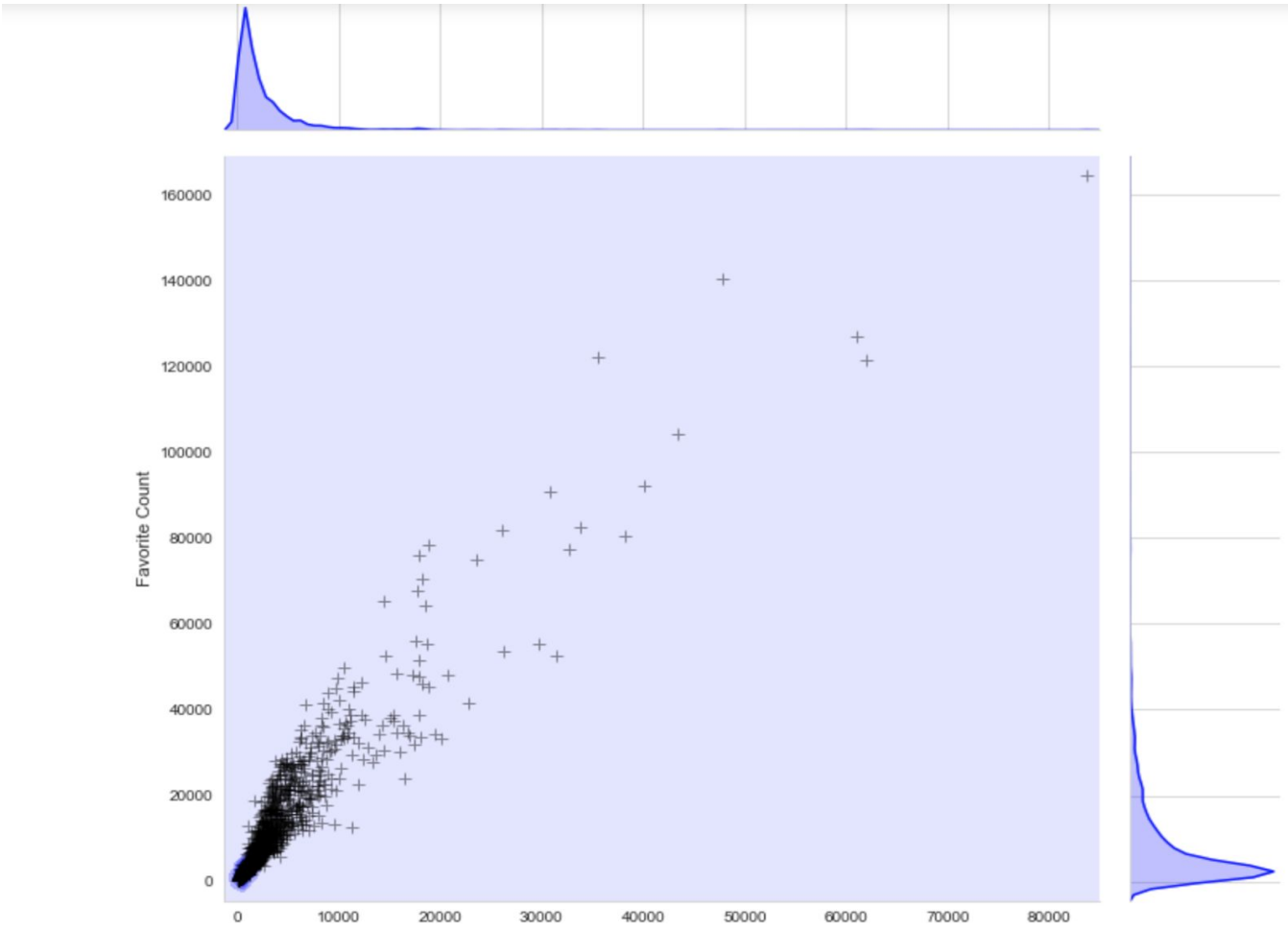
Since I already know the most common dog names , I wanted to know the most common dog ratings by @dog_rates aka WeRateDogs.

3. What are the most common dog rating?

Here, we found out that 12 is the most common dog rating with approximately 449 times rated. So, you should not be surprised when your dog picture was given 12 rating.

Then, I was thinking does the increase in retweet_count increases the favourite_count since it is logical to assume that the more people retweet it, the more people favour it. Hence, I wanted to prove my hypothesis

4. Does the increase in retweet_count increases the favourite_count?



Here, I found out to my surprise that there is a strong correlation between these two when the retweet count is below 10, 000 but then the correlation are marginal when it is above it.

Hence, my data analysis hence proven that my hypothesis was half-correct and half-wrong since the retweet count is directly proportional to the favorite count when it is below 10,000 .I was also false because there are marginal correlation between these two once the retweet count is above 10, 000.