

1. Introduction

The main purpose of this project is giving students an opportunity to gather, assess, clean and analyze the data by using a real case. WeRateDogs™ is a community page on twitter designated to rating dogs on their appearances and stories, and which was formed by the user @dog_rates in 2015. The purpose of this project is getting insightful information about dog ratings from the twitter page WeRateDogs™. I could get insightful information about dog ratings from the twitter page WeRateDogs™ through completing this project.

2. Exploratory Analysis – Insight 1

✓ What are top 10 represented dog breeds and dog names?

→ There are top 10 represented dog breeds in the dataset.

Golden retriever	137
Labrador retriever	93
Pembroke	88
Chihuahua	74
Pug	53
Samoyed	40
Chow	39
Toy poodle	37
Pomeranian	36
Malamute	29

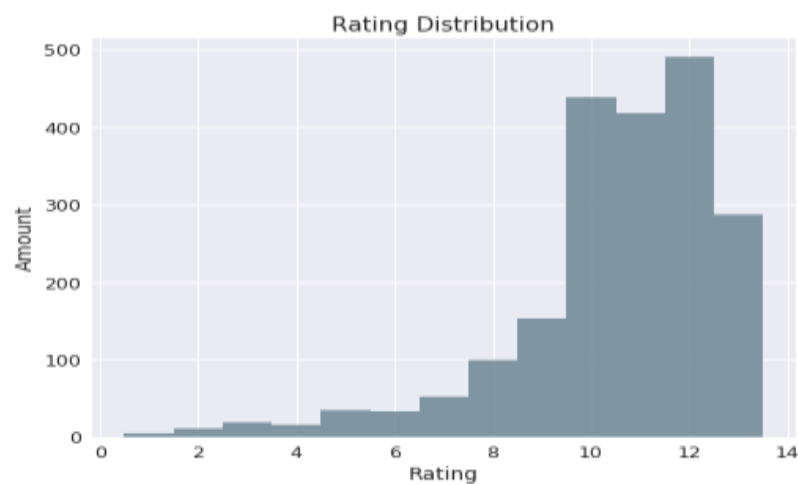
→ There are top 10 represented dog names in the dataset.

Lucy	11
Charlie	11
Oliver	10
Cooper	10
Tucker	9
Penny	9
Sadie	8

Winston	8
Lola	8
Daisy	7

3. Exploratory Analysis – Insight 2

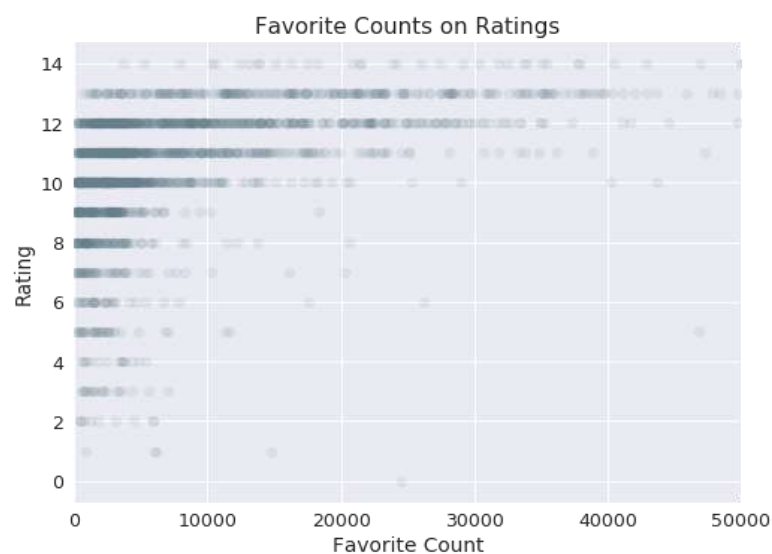
- ✓ What's the information about the dog's ratings?

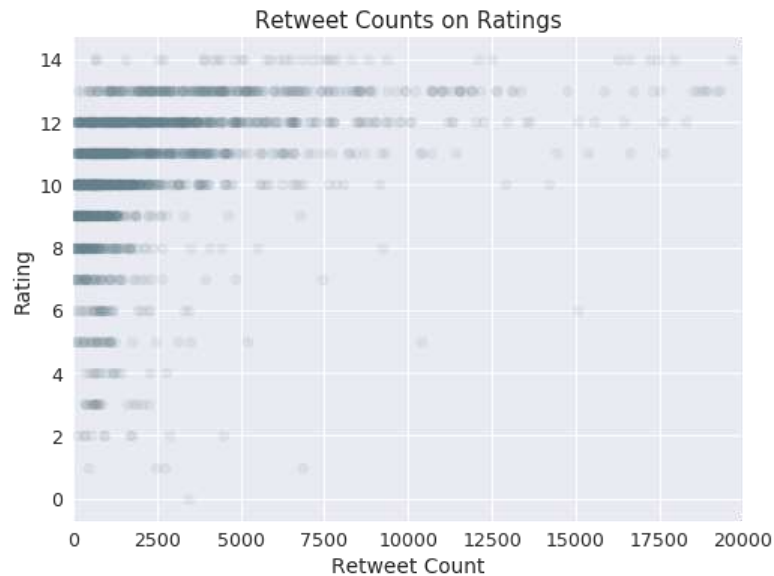


The number of dog's ratings is 2091. The average rating is 10.598278, and the median rating is 11. The lowest rating is 0, and the highest rating is 14. The rating distribution is left-skewed. Majority of the dog's ratings are over 10.

4. Exploratory Analysis – Insight 3

- ✓ What's the information about the favorite count and retweet count?



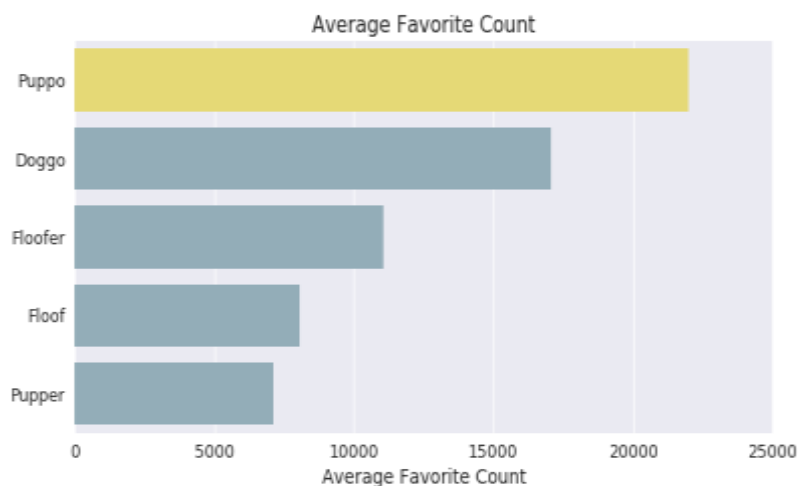


The number of 'favorite_count' is 2091. The average count is 8983.248207, and the median count is 4188. The lowest count is 81, and the highest count is 132810. The dog who received high ratings could get lots of favorite counts.

The number of 'retweet_count' is 2091. The average count is 2841.095170, and the median count is 1403. The lowest count is 16, and the highest count is 79515. The dog who received high ratings could get lots of retweet counts.

5. Exploratory Analysis – Insight 4

✓ Which dog stage got the highest and lowest favorite_counts on average?



I could find the average favorite counts of each dog stage (Puppo: 21977.785714, Doggo: 17080.164835, Floofer: 11064.444444, Floof: 8068.888889, Pupper: 7114.540984). That is, Puppo received the highest average counts of 21997, while Pupper received the lowest average counts of 7114.

6. Conclusion

I could perform the data wrangling and data analysis by completing this project by using the data from the WeRateDogs™.

Firstly, I had to gather data to get three data frame files. Then, I assessed them to find issues to be fixed. I could find 10 quality issues and 2 tidiness issues. I modified these issues through data cleaning process, then I created one combined data frame file that is more helpful to analyze data. The final data frame was saved as 'twitter_archive_master.csv' file. I interpreted this data frame through data analysis and visualization. I could find four kinds of insightful information.