Data Analysis and Visualization Report

By Afnan Alabdan

Introduction

The dataset that wrangled is the tweet archive of Twitter user **@dog_rates**, also known as WeRateDogs. 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, though? 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.

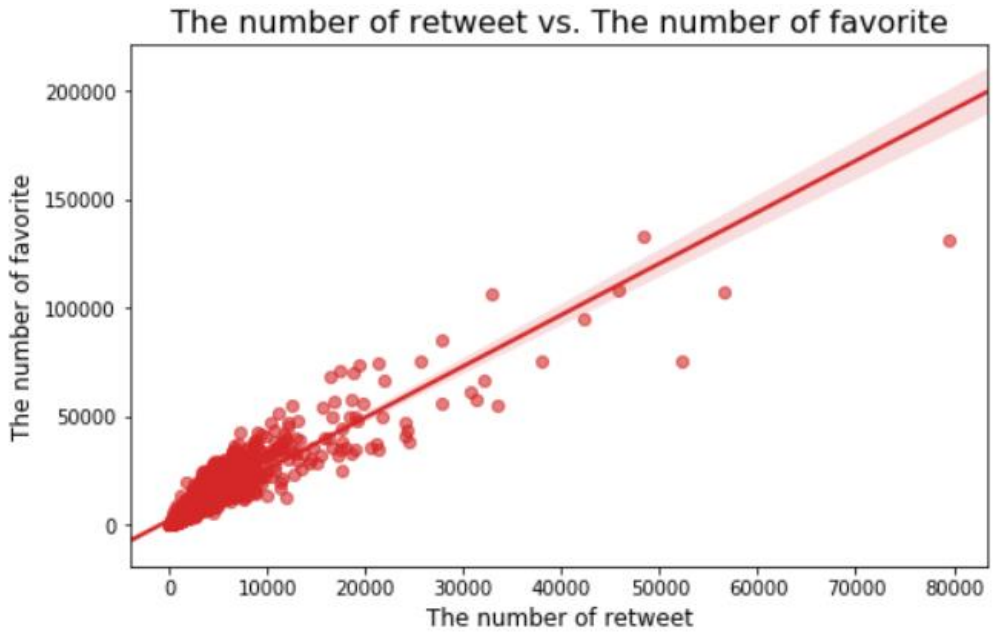
Here is an example of one of their tweets:

Source: https://twitter.com/dog_rates/status/1583219560055857152

Data Analysis and visualization

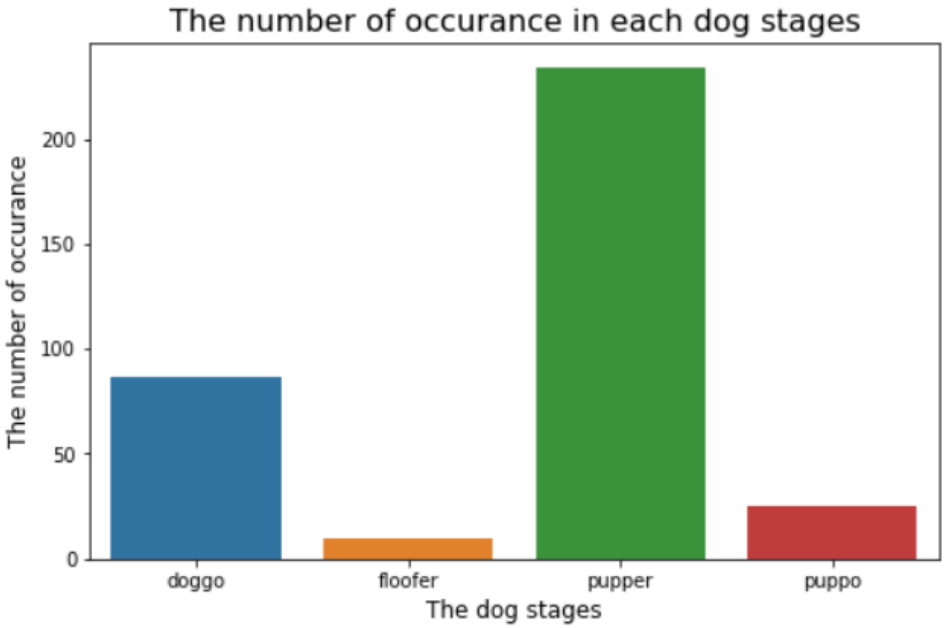
After Gathering, assessing, and cleaning the data set, I have the following Questions to get some meaningful insight and display the visualization:

1. Is there a relationship between the number of retweets and number of favorites?



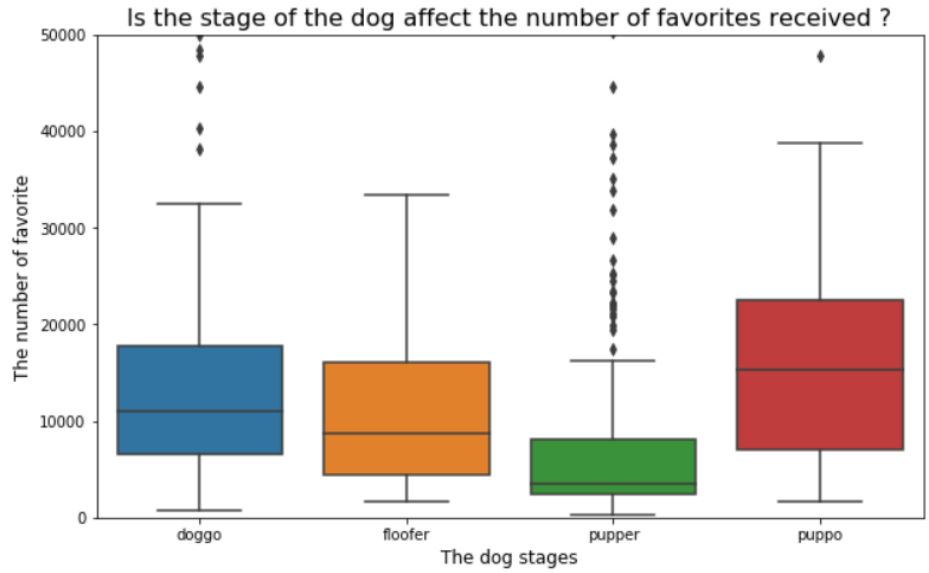
As it shown from the graph above, there is a positive relationship between the number of retweets and favorites.

2. Which dog stage was the most common one?



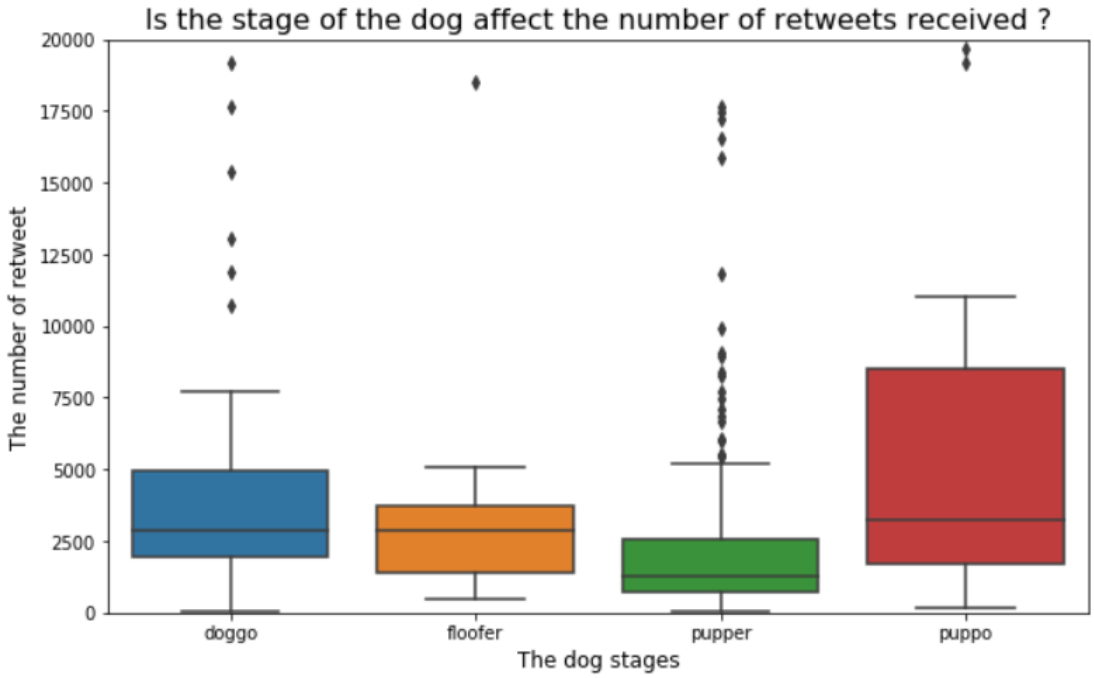
As it shows in the chart above, the 'pupper' was the most common dog stage in these tweets and the 'doggo' comes after it. While ‘floofer’ is the least.

3. is the stage of the dog affect the number of favorites received?



As it shows in the chart above, the 'puppo' stage has the highest number of favorites, while 'pupper' has the lowest.

4. is the stage of the dog affect the number of retweets received?



As it shows in the chart above, similar to the pervious chart the 'puppo' has the highest number of retweets while 'pupper' has the lowest.

Conclusion

The dataset of WeRateDogs that I have worked in this project, was from the real-world data that helped me to understand the best practices to wrangle and analyze the data set in the real life. Also, it has a lot more to work with and get more meaningful insights.