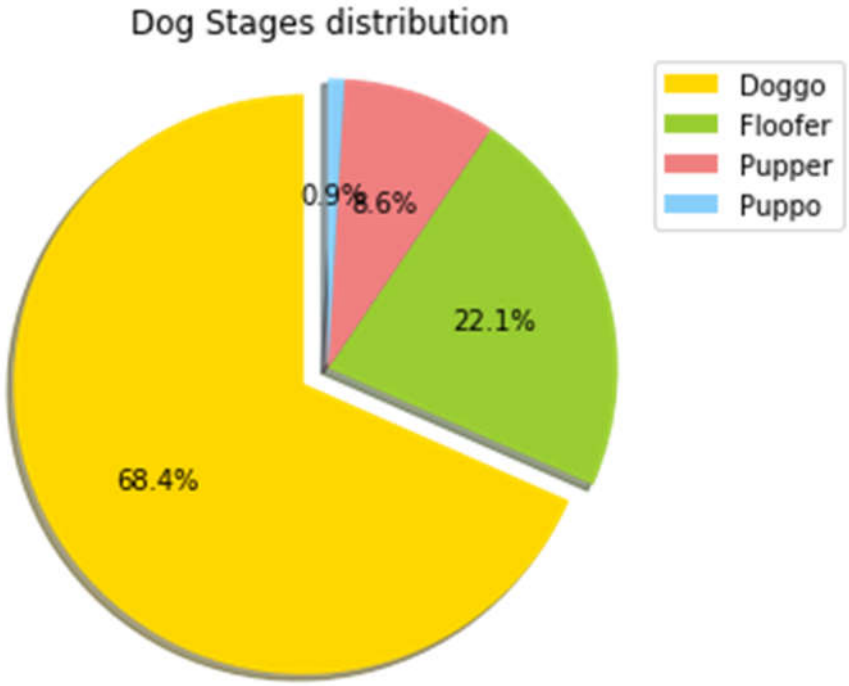
We Rate Dogs Twitter Feed Analysis

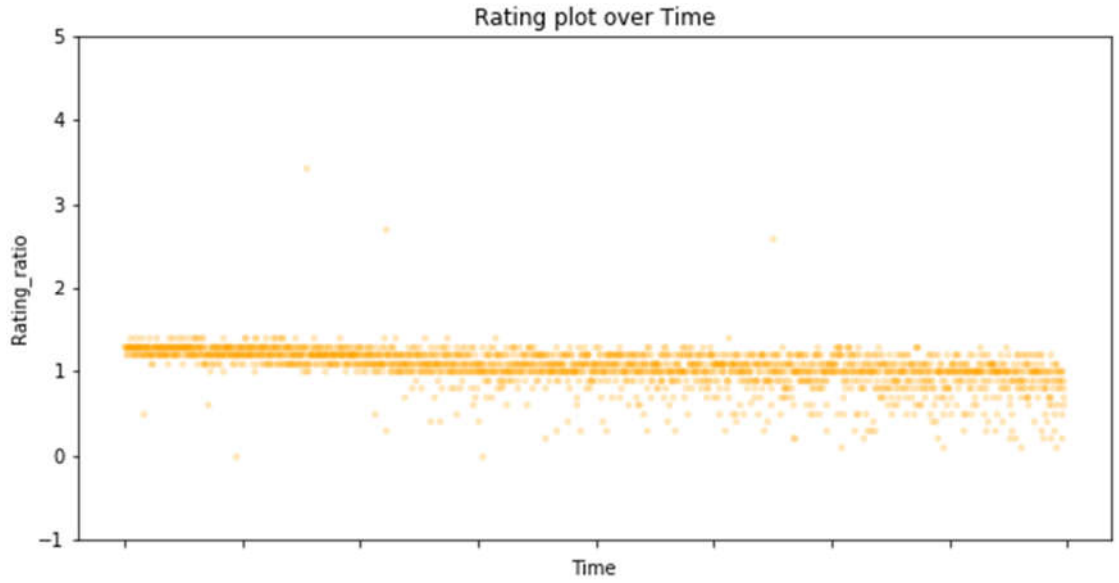
This project is all about gathering, assessing, cleaning, storing, analyzing and then visualizing the tweet history of the famous twitter account WeRatesDogs.

The first chart below demonstrates the difference in the proportion of dog stages gathered in the dataset.

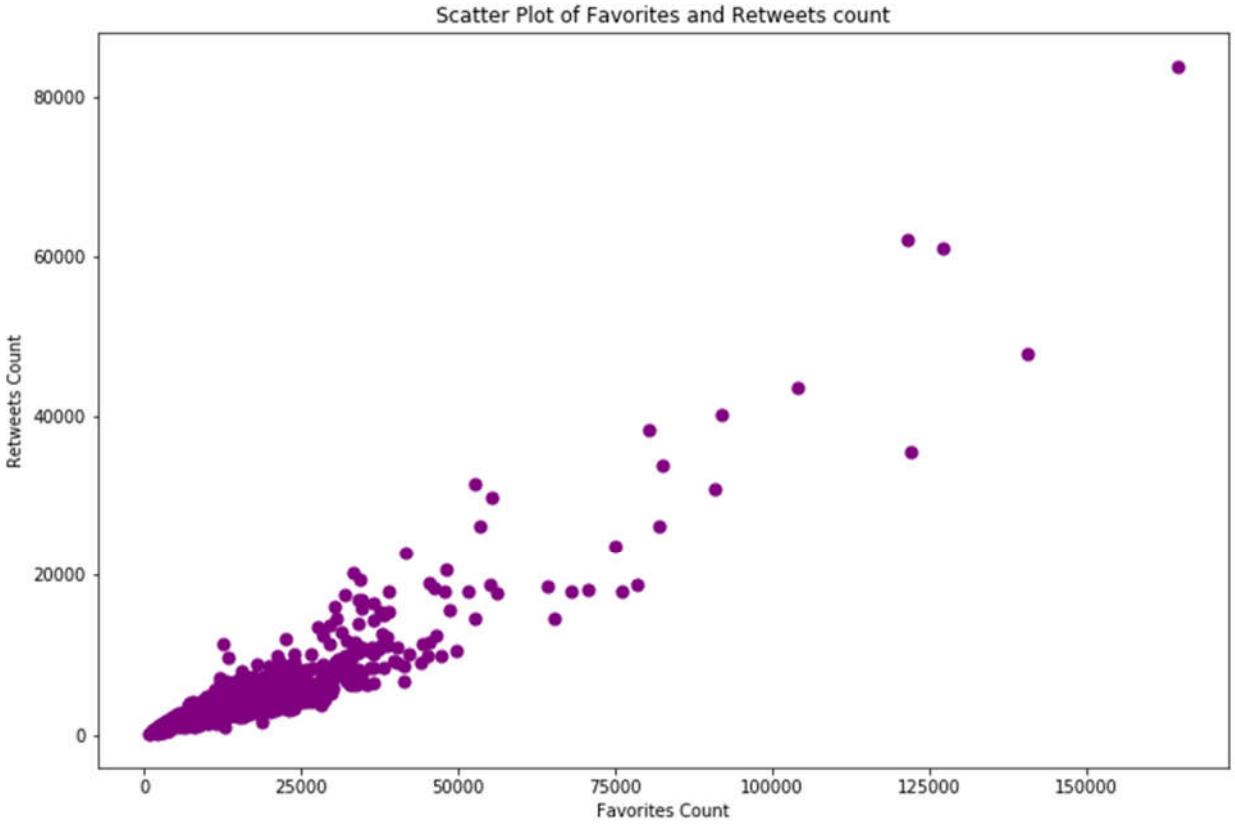


We can clearly see that 'Doggo' is the most popular dog stage among the survey. It accounted for nearly 70% of all stage which is two times higher that of the rest's combination. The second most popular is ‘Floofer’ which accounted for 22.1%, followed by ‘Pupper’ and ‘Puppo’.

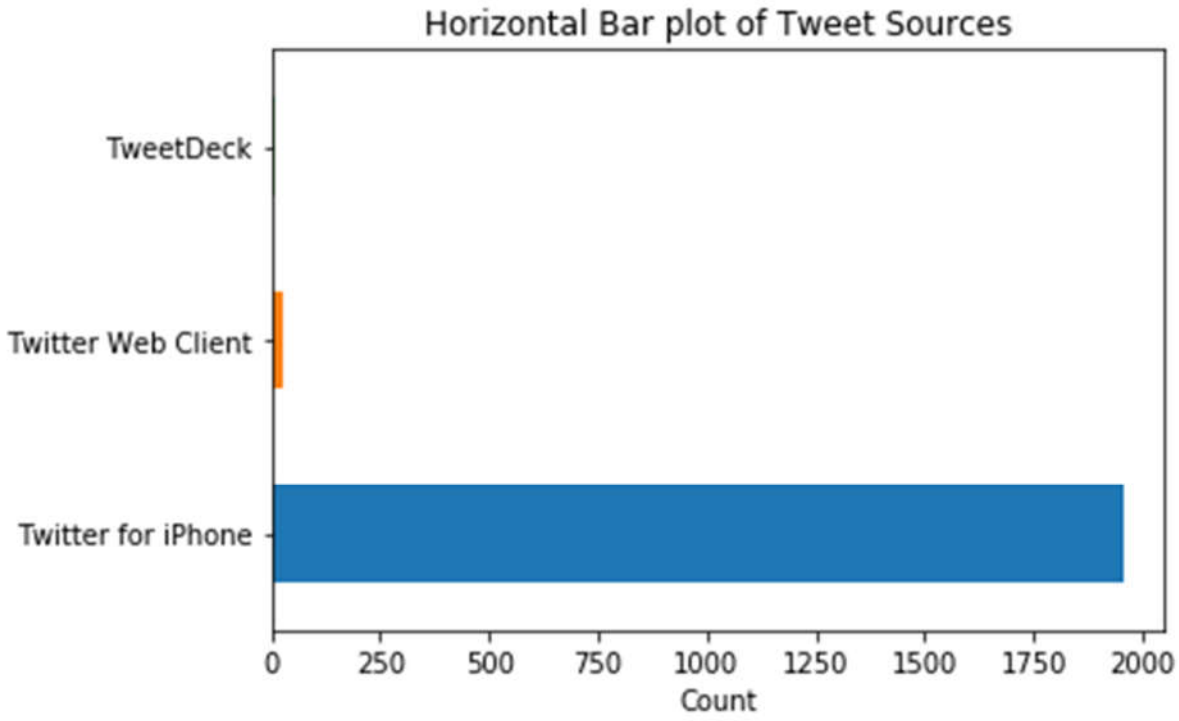
The second chart describes how rating ratio distributed over time. One can not deny that rating ratio mostly stay between 1.0 and 1.2. In detail, More than 75% of the data has more than 12/10 as rating. More than 50% of the data has more than 11/10 as rating. Furthermore, the rating range seems to be more diversity than that of the past



The third chart, on the other hand, shows how retweets count and favorite count correlated to each other. According the chart, retweet count has a close relationship with favorite count as they are very similar in number. The more retweets count is, the more favorite count is. However, the majority of the dots stay in the range of 0-20000, indicating a large number of the counts stay in that range



Last but not least, **in the final chart,** the number of tweets submitted from Iphone is ridiculously higher than that of all other source, indicating that people tent to tweet by their Iphone instead of other means when using Twitter.



Conclusion

The Twitter account WeRateDog is a good source to collect information regarding user's beloved dogs' stages, rating, favorite and retweets counted over time. However, we need to put effort in cleaning data retrieved from Twitter API as the raw data is having a lot of problems making it difficult to read and research.

If you want a specific row of data to analyze, my advise is to validate its own and related columns data before using them for visualization.