# Data Analysis and visualization

Introduction to act report

This act eport includes the data analysis WeRateDogs twitter account data from two dataset: twitter_archive_clean and json_tweet_clean. It provides three insights from the analysis and visualization with their observation. I have chosen three different visualization histogram and box plot.

1. **What is the highest source used?**
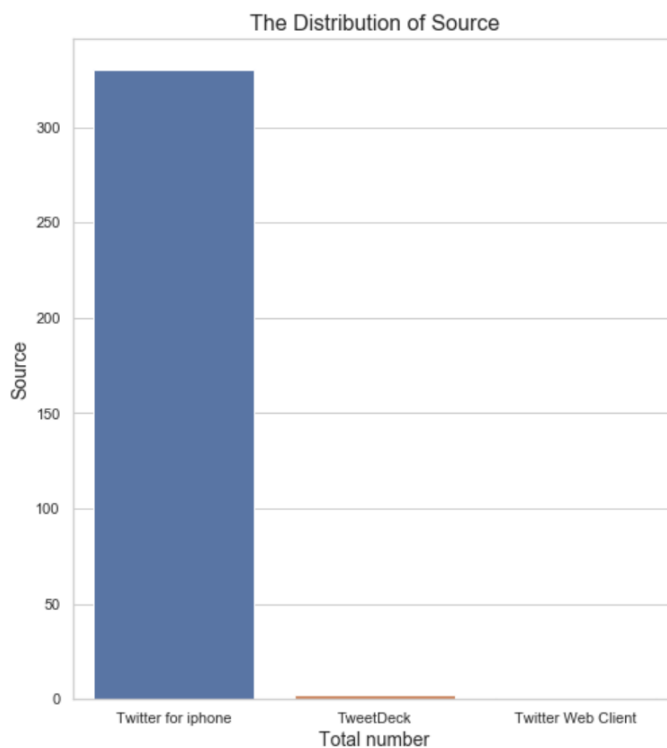
The plot above shows us the distribution of source which are iPhone twitter app, TweetDesk and Twitter Web Client. We can see that the common source of tweets is from iPhone twitter app, which is 94% in the total and that considered a high percent. That means the twitter app is the most common used for people using to tweet, retweet, favorite and post. while the TweetDesk is rare to use.

2. which is the dog category have the highest favorites rate?

Dog categories with total of favorites
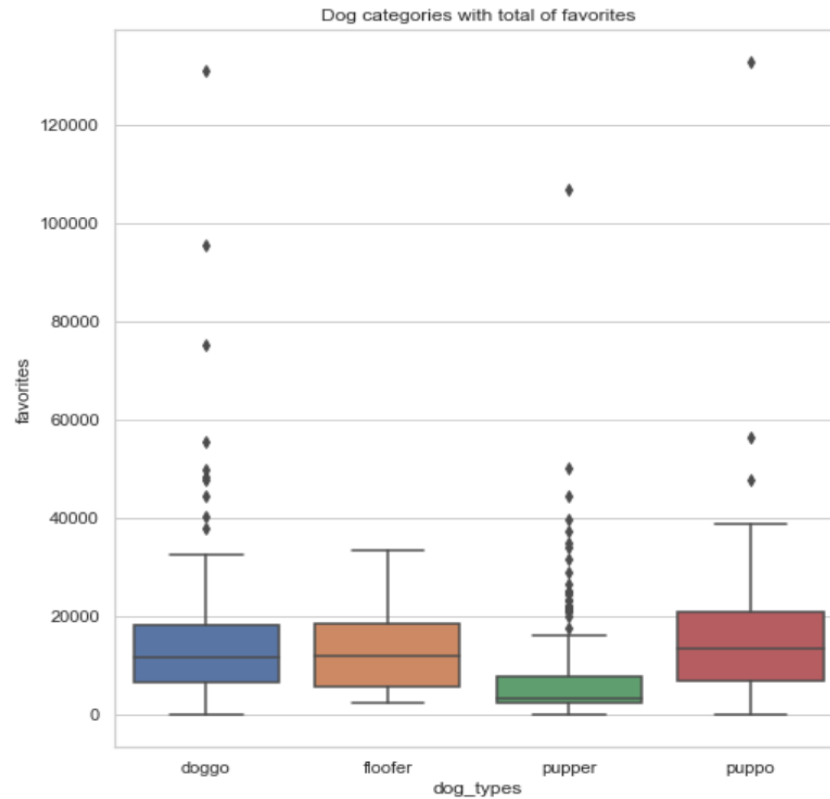
As we can see,after removing None values from the dog_types and create the visualization it is clear to us that puppo dog recorded the highest rate between all dogs, then floofer,doggo, and lowest rate are pupper dogs.

3.The correlation between retweets and favorites
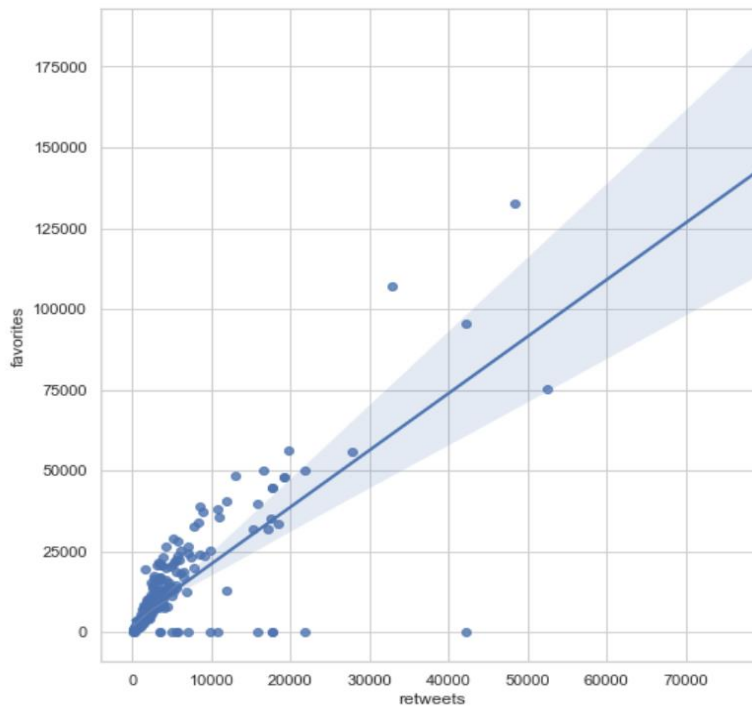
The regression plot above show us the most popular tweets usually get a large number of retweets and favorite. As we can see that I have test the correlation between the retweets and favorite and it was 0.837409, this is a high value showing a strong positive relationship between retweets and favorite. So, when the tweets get a high number of retweets there is a high chance to get favorite as well.