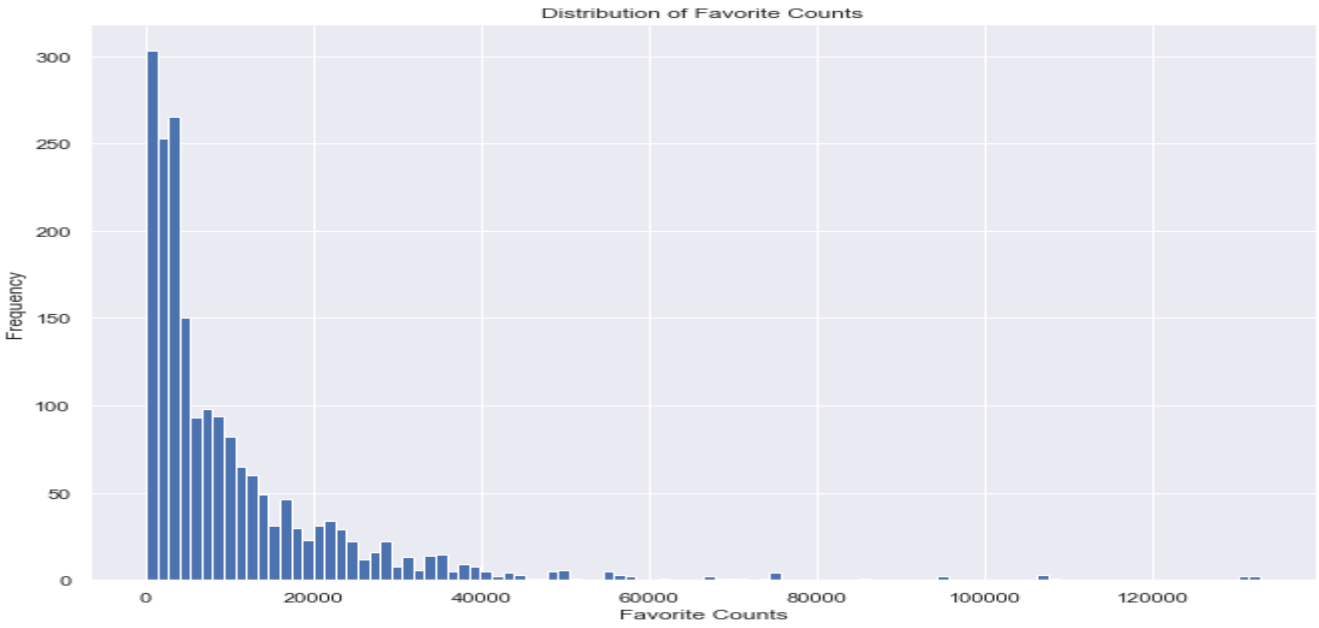
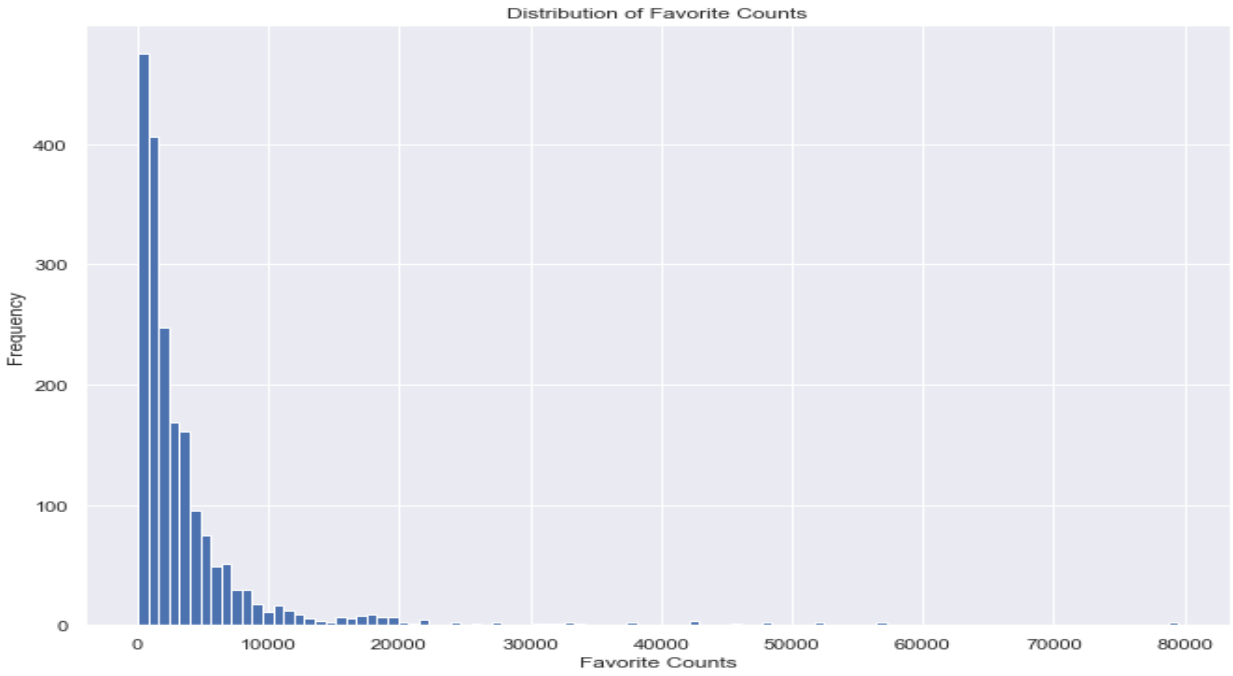
Data Analysis and Visualization

This data analysis and visualization report includes the basic data analysis of WeRateDogs twitter account data from three datasets: (twitter_archive_clean) , (image_predictions), and (tweets_clean) data frames. It provides four insights from the analysis and visualization results. We RateDogs is a twitter account that, as its name implies, it rates dogs. It receives dog pictures, it posts and comments them and provides a rating for the dogs.

Data Analysis and Visualizations:

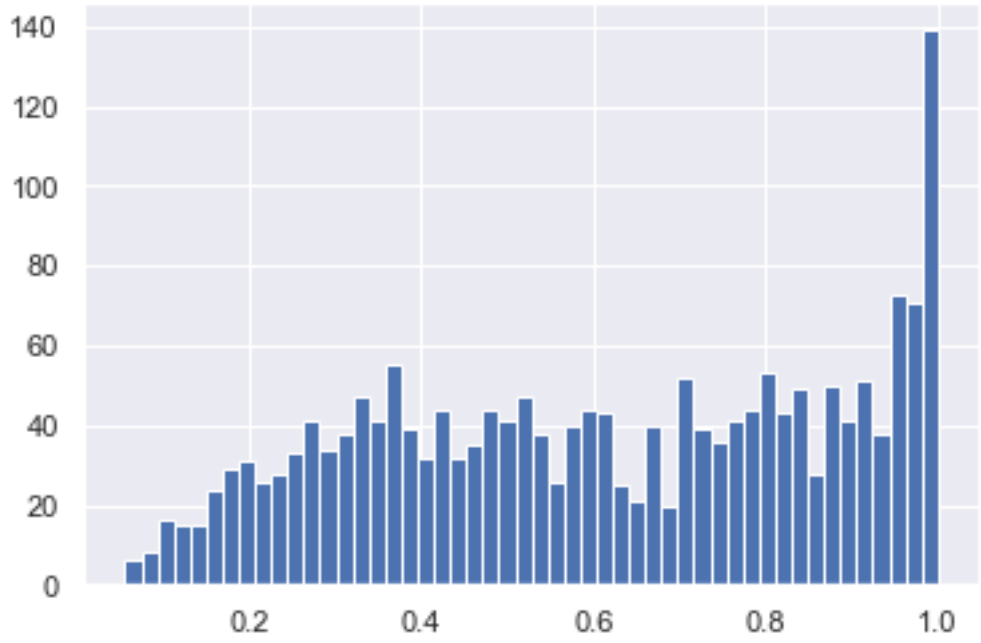
1- **Distribution of Favorite Counts and retweet counts**

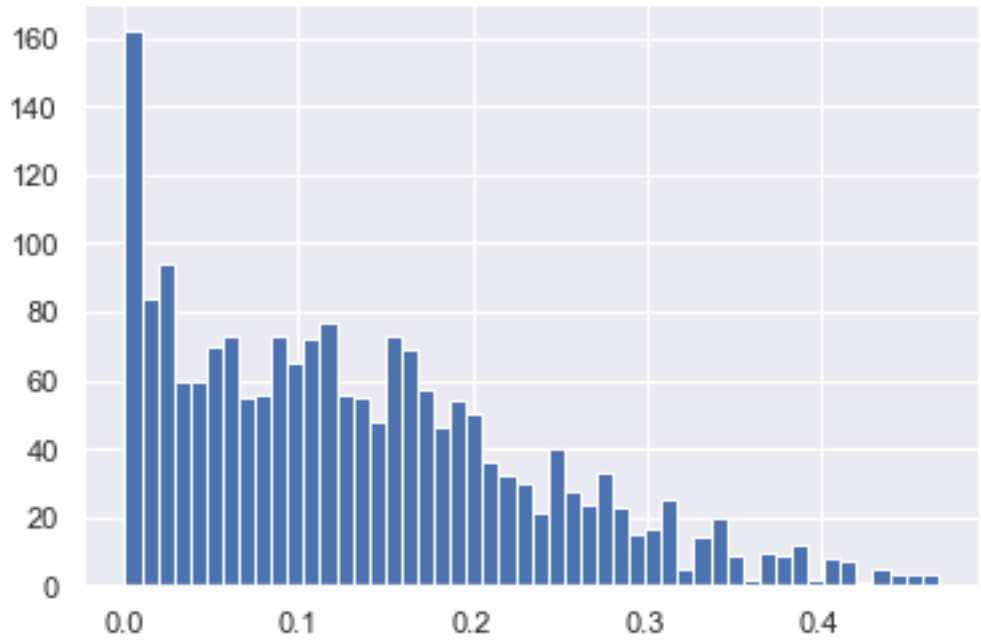




The distribution of retweet_count and favorite_count look similar, but retweet_count is lower than favorite_count. I would guess that these numbers are related to timestamp. Tweets that were tweeted after the account went launched , and verified on twitter, would have higher retweet_counts and favorite_counts.

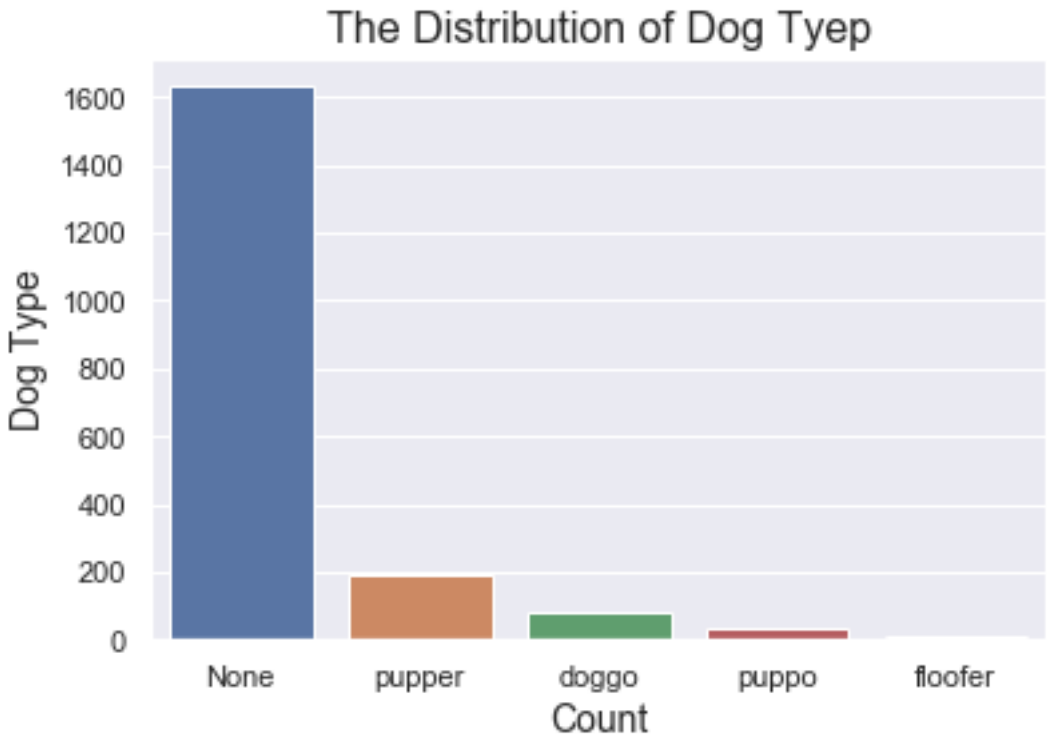
2- First and second confidence:





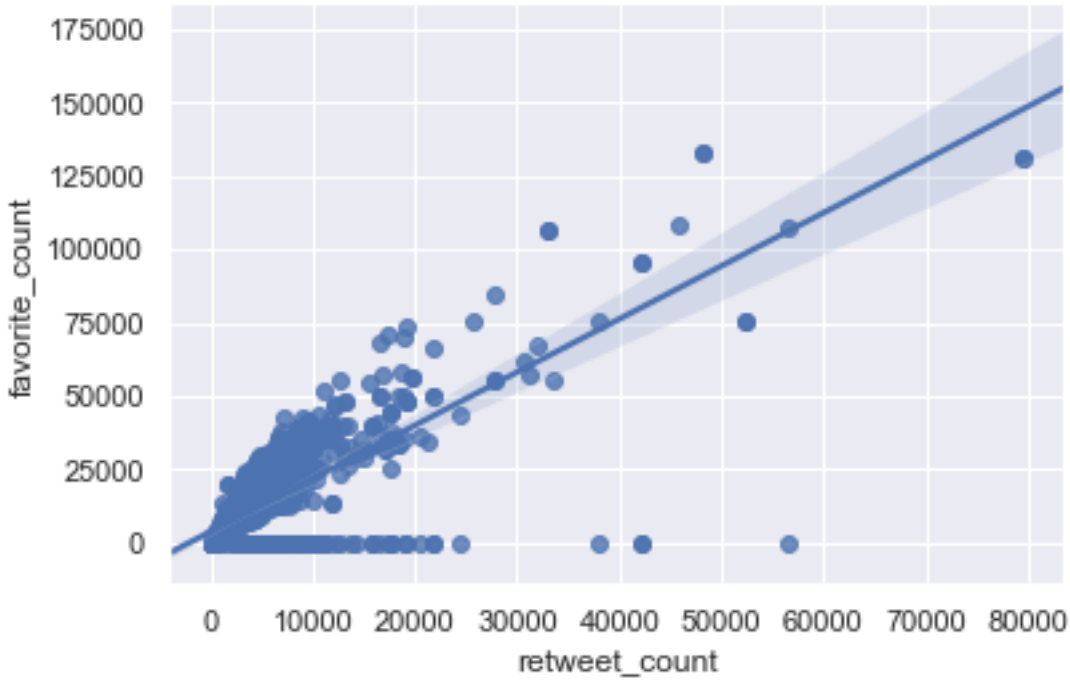
It looks like there is a higher frequency of first confidence values that are close to 1.0, but an even greater frequency of second confidence values that are closer to 0.0. This makes sense, it is likely that these high frequency values for first confidence and first confidence are from the same tweets.

3- The distribution of doge types



Even we have missing data but diagram It shows that pupper which is a small and adorable , usually younger) is the most popular dog type , followed by ‘doggo’ and ‘puppo’. It could be due to the young and unmatured dog is usually cuter than the adult dog.

4- Retweet and favorite relationship.



The diagram illustrates there is a There is significant relationship between retweet and favorite. That’s mean the popular tweet has most of retweet and favorite.