

WeRateDogs Project Analysis Report

By Victor Obi

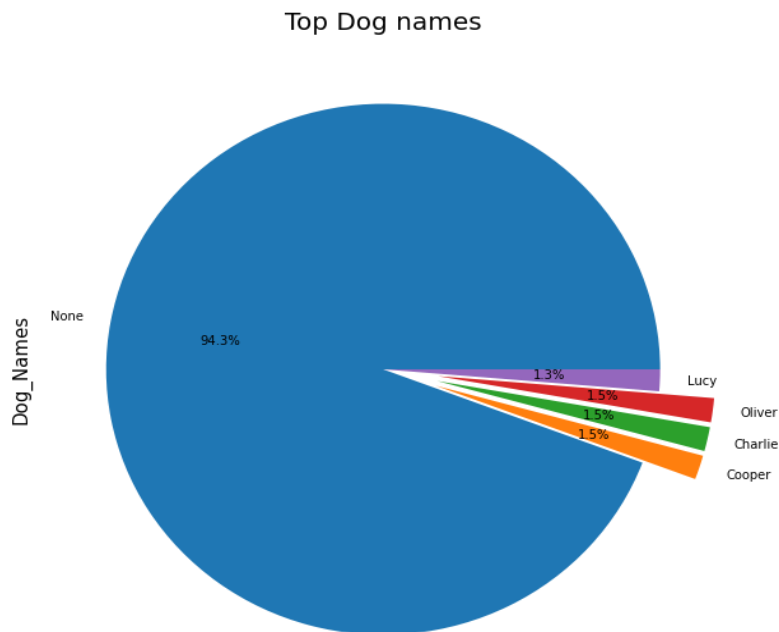
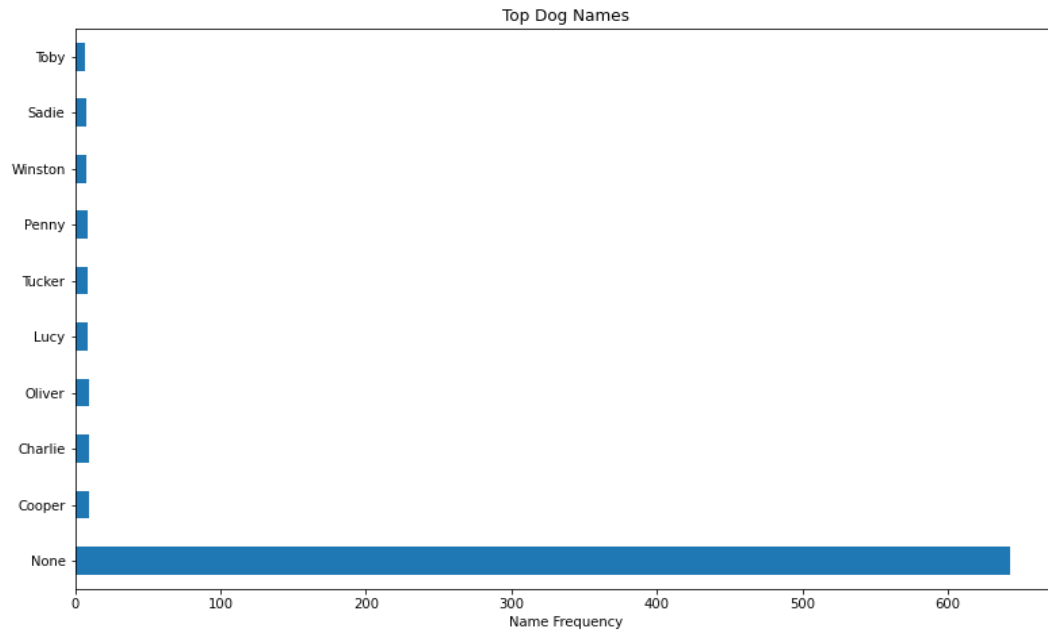
September 15, 2022

The 3-step data wrangling process yielded a much cleaner dataset ready for exploratory data. The final step in the process was merging the 3 dataframes using an inner join on the tweet_id column. The resulting columns include

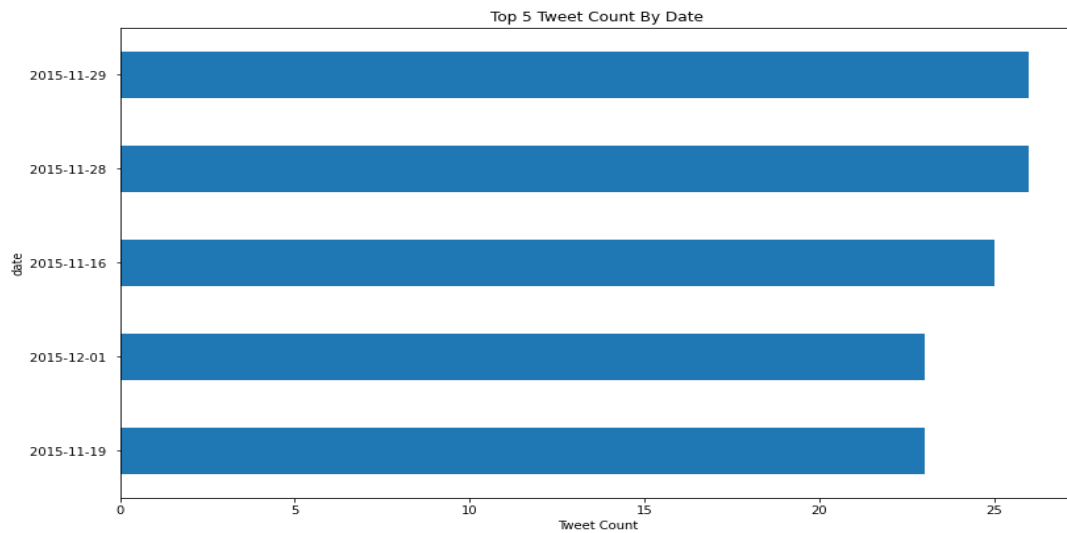
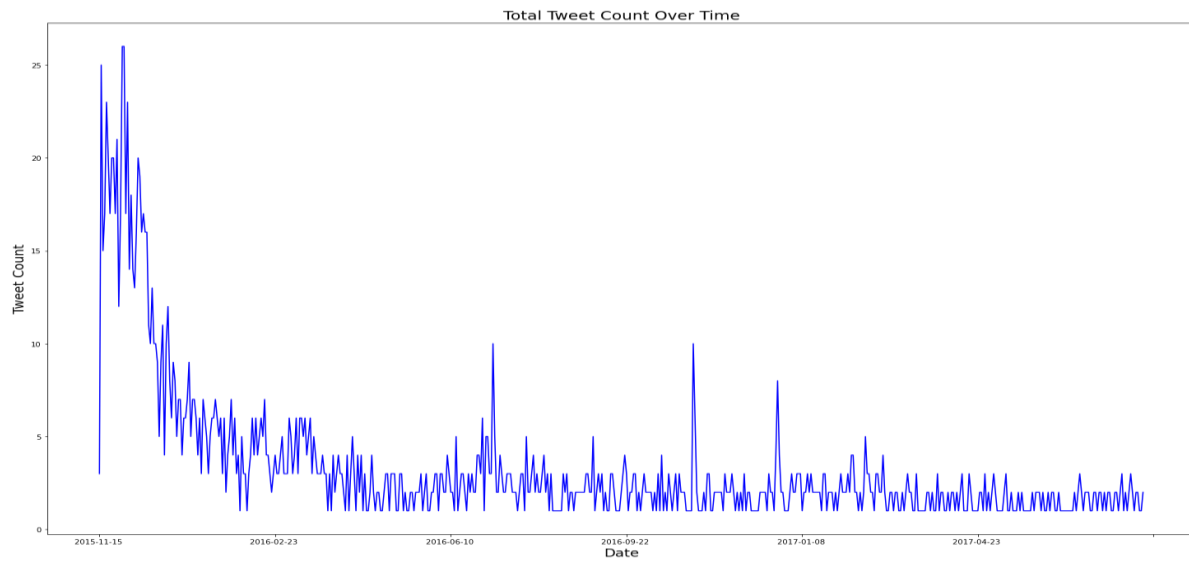
- tweet_id
- in_reply_to_status_id
- in_reply_to_user_id
- timestamp
- text
- expanded_urls
- rating_numerator
- rating_denominator
- name
- date
- source_trimmed
- total_rating
- dog_stage
- jpg_url
- img_num
- p1
- p1_conf
- p1_dog
- p2
- p2_conf
- p2_dog
- p3
- p3_conf
- p3_dog
- retweet_count

My analysis focused on dog names, Tweet count, Source, favorite count, Retweet count, ratings and sentiments.

First, the dog name column is not a unique one as several dogs shared similar name. Consequently, analyzing the most popular dog name became valid as shown below.

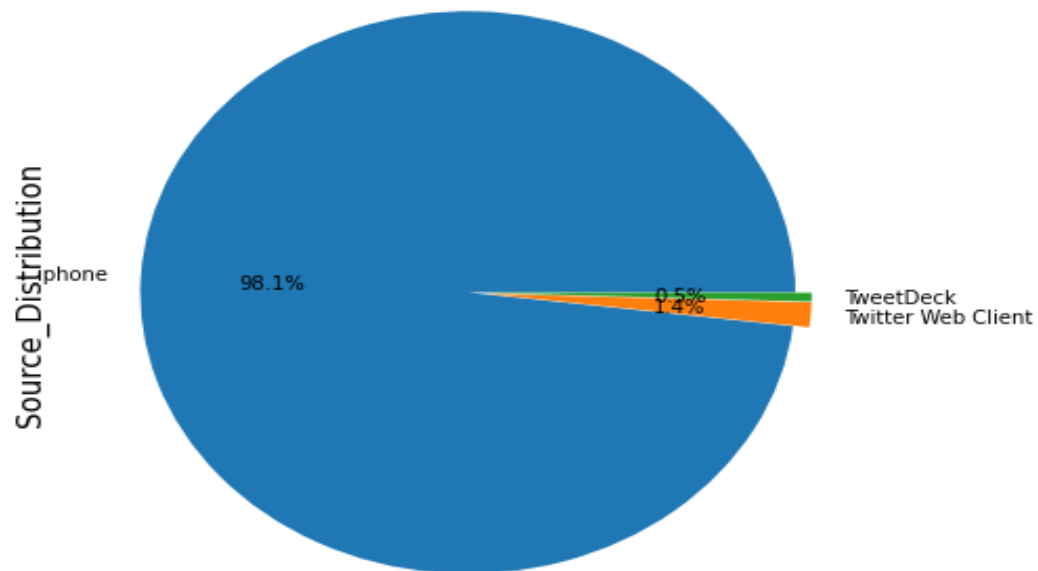


Secondly, my analysis also looked at tweet volume over the time period in question. Tweet count was at all time high between Nov - December of 2015 followed by a sharp drop afterwards. The highest daily tweet count was on 11-29 followed by 11-28 of 2015

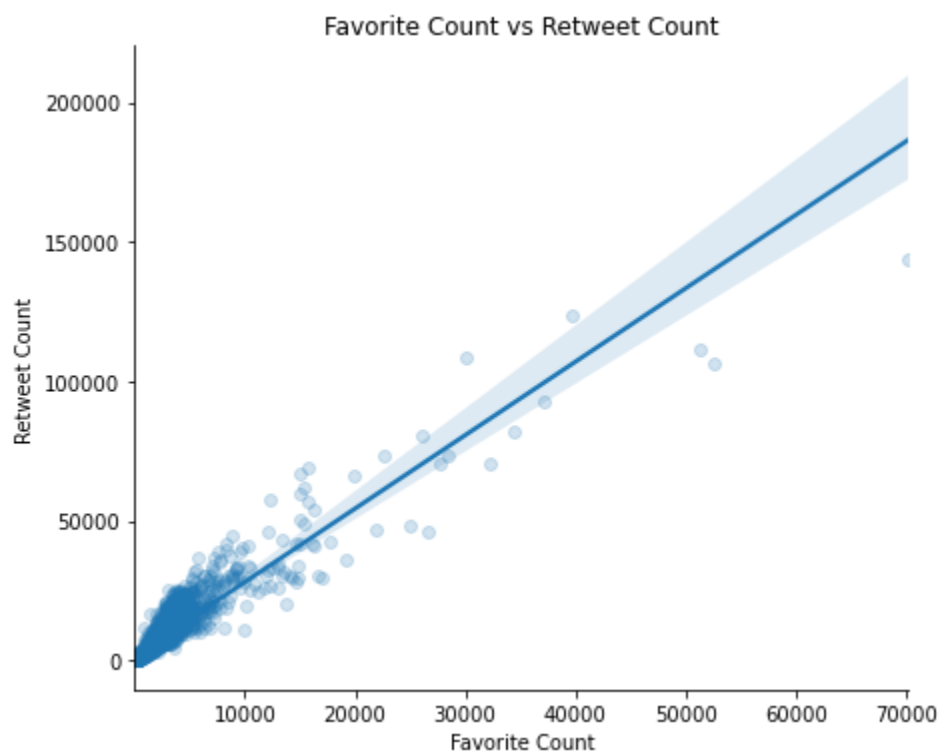


In addition, the source column was stripped to extract the specific source. Four unique sources were identified - iPhone, TweetDeck, Twitter Web Client and Vine. These reduced to 3 in the data cleaning process. Analysis shows that 98% of the tweets were made on an iPhone while TweetDeck and twitter web client contributed to 0.5% and 1.4% respectively.

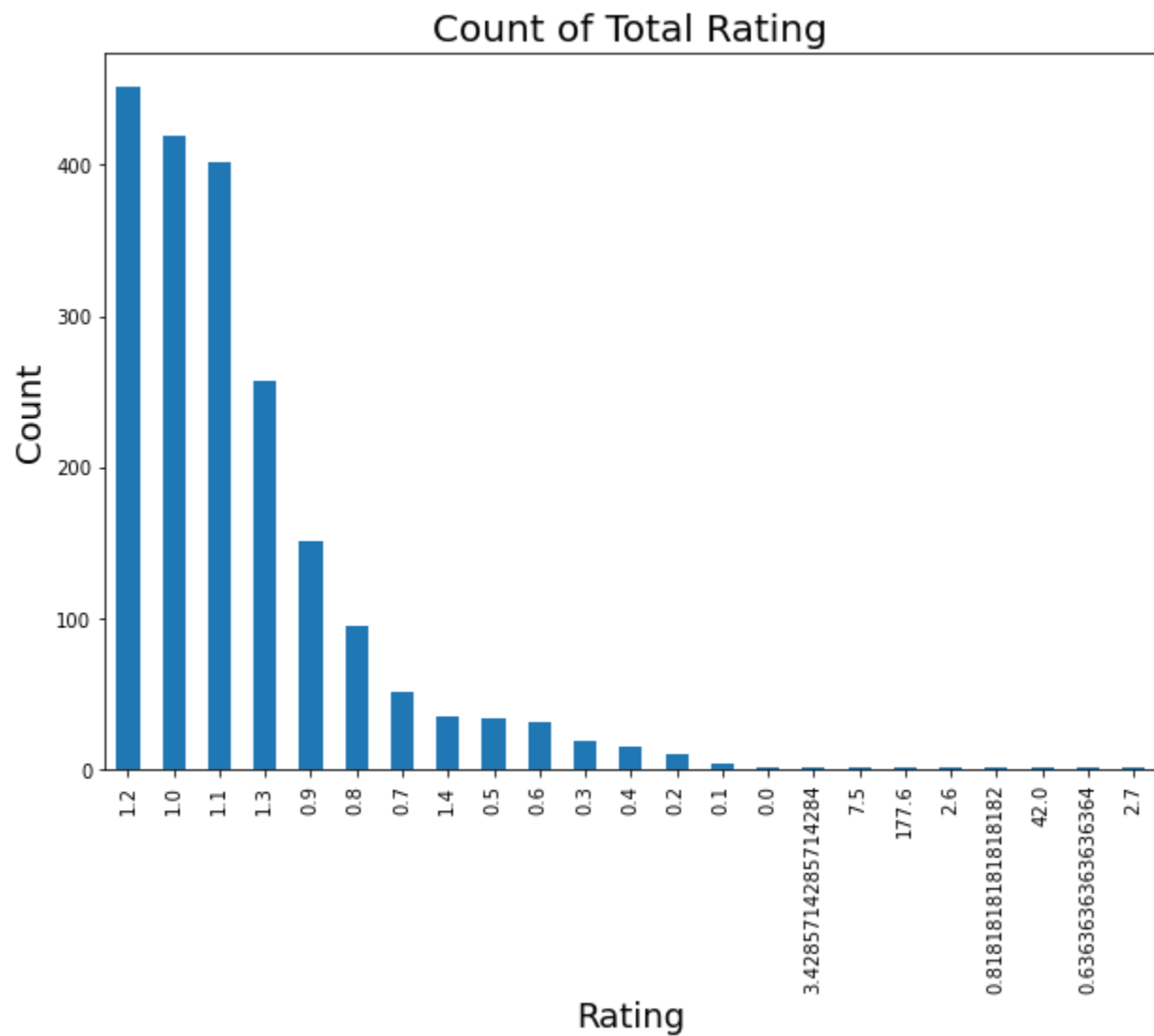
Percentage Distribution of Tweet Source



A plot of retweet count and favorite count showed a positive correlation. It is also important to mention here that correlation does not mean causation. Retweet count and favorite count both trended in the same direction.



Analysis of total rating yielded a rating count as shown below



Sentiment analysis shows a mean polarity score of 0.13 and median polarity of 0.05. A positive polarity indicates mostly positive sentiments. Mean subjectivity is 0.48 and is almost neutral on the objectivity-subjectivity scale.