Act Report

Introduction:

Data wrangling project based on 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. The account was started in 2015 by college student Matt Nelson, and has received international media attention both for its popularity and for the attention drawn to social media copyright lawwhen it was suspended by Twitter for breaking these aforementioned laws.

In this project the goal is to wrangle and analyze tweet archive of Twitter user @dog_rates.

Gathering Data:

I've gathered data from three different sources: - Twitter_Archive_Enhanced.csv Which was provided on hand. - File image_prediction.tsv I've extracted the data programmatically using Python request library. - tweet_json.txt extracted from Twitter API using Python tweepy library.

Assessing Data:

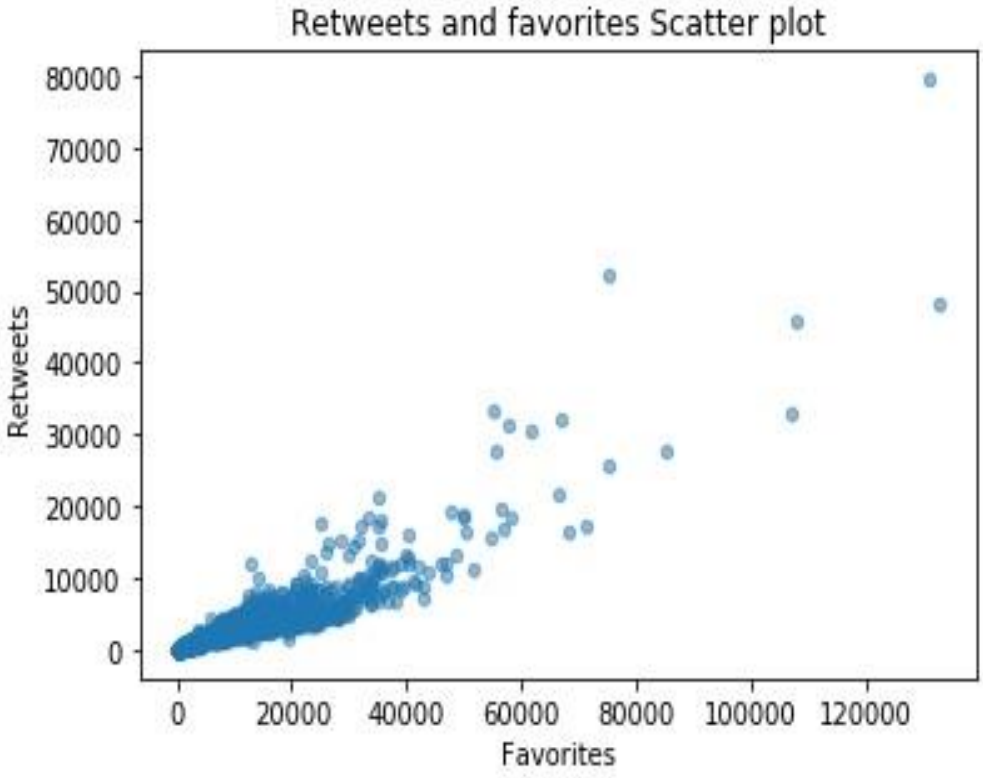
I've asses the data frames Visually and Programmatically to combine quality issues and tidiness issues.

Cleaning Data:

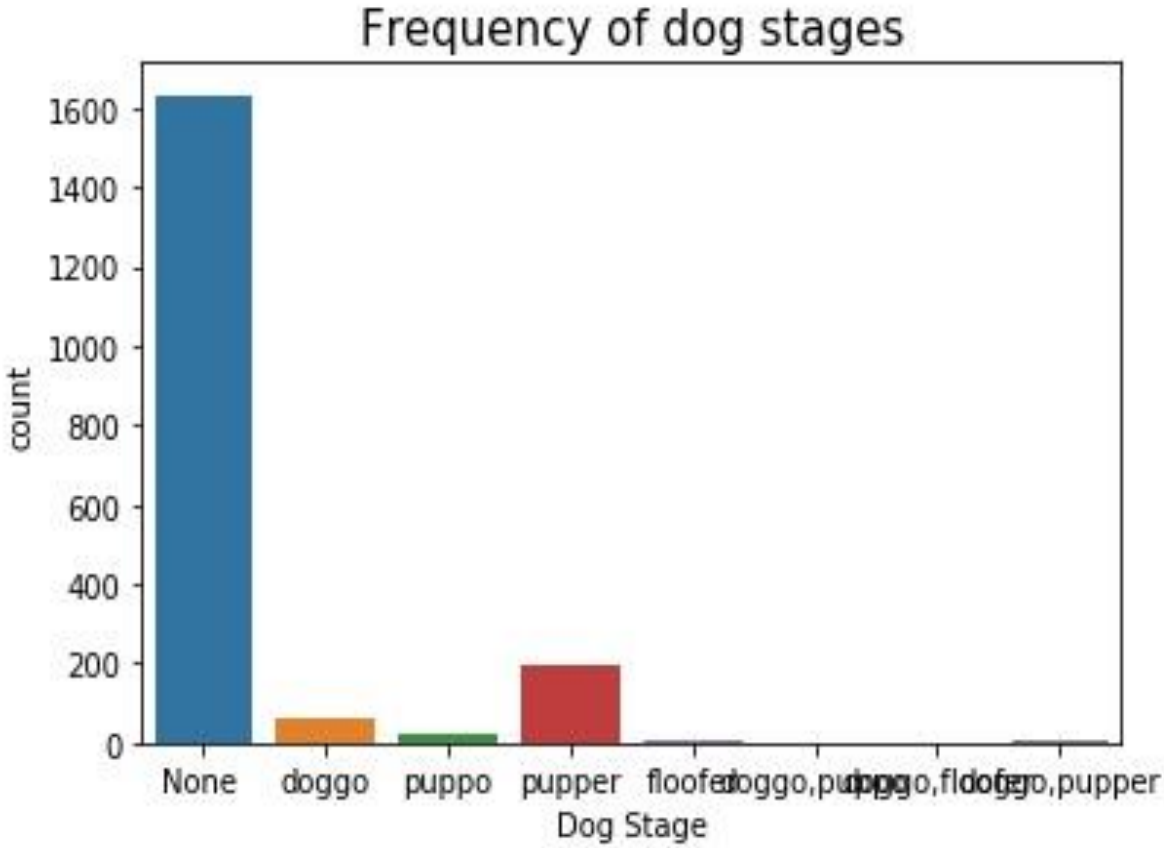
Finally, I've used basic python function to clean and fix all the qualities and tidiness issues like the drop, info, and value count.

Visualizing:

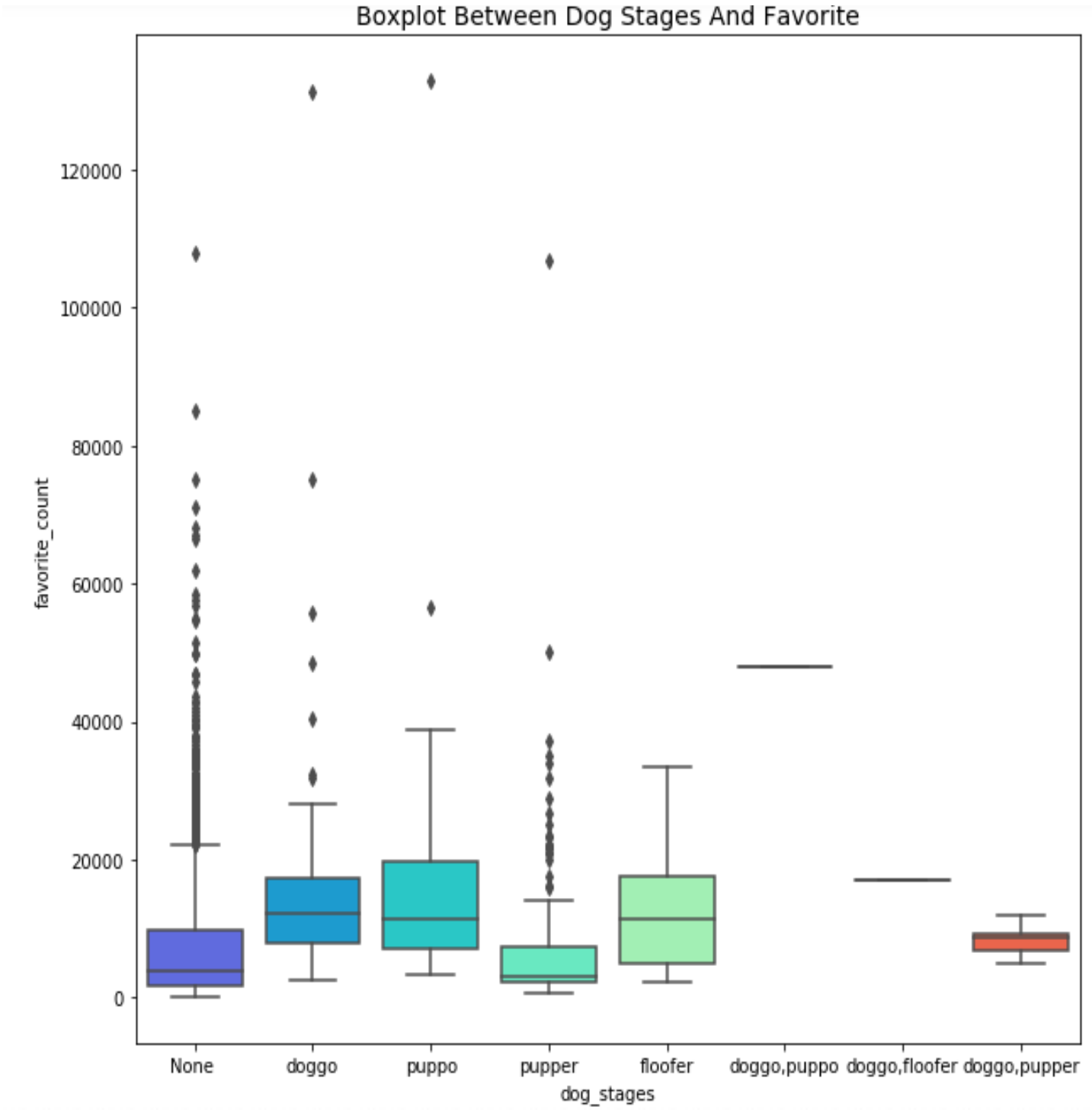
During cleaning step, I found some interesting analytics after finishing the cleaning process I made some charts to discover the data even more.



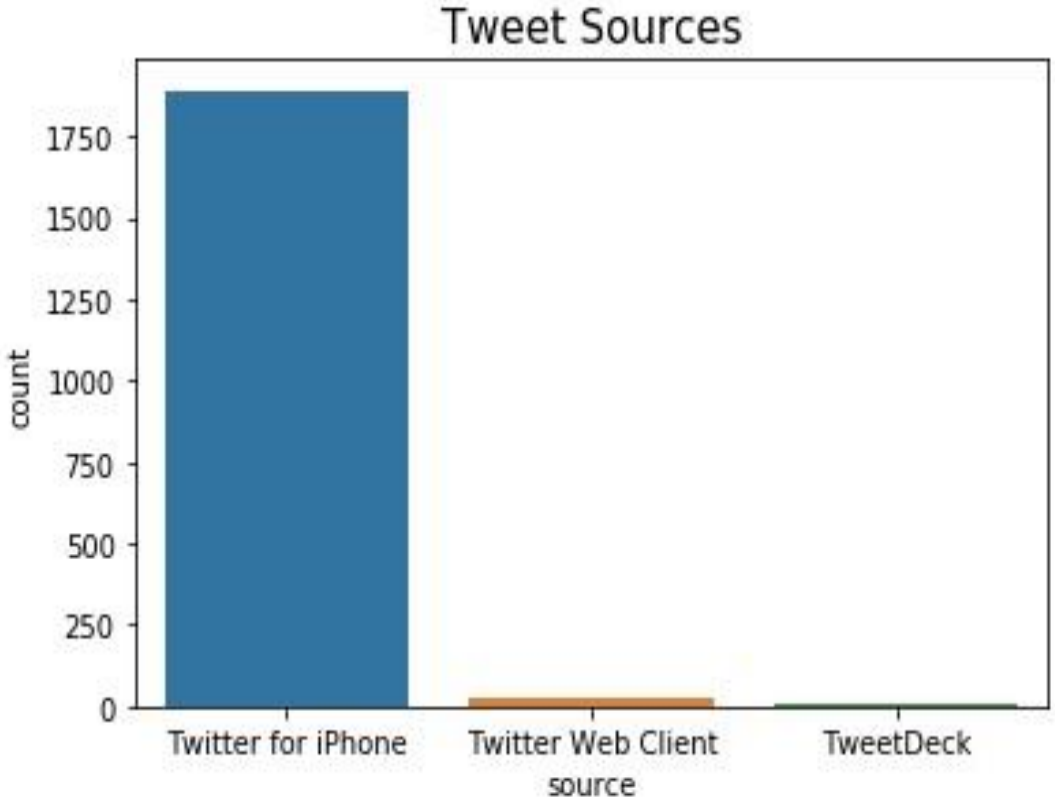
as shown in graph retweet is highly correlated with favorites with positive correlation.



as shown in graph Pupper is the most frequent type.



The Box plot shown that the Puppo has the most favorite count



The bar graph shows that Twitter for iphone has the highest count of sources.