WE RATE DOGS REPORT

The wrangle and analyze project was the second project in Udacity's nano degree data analyst course. The project involves *WeRateDogs* twitter wrangling data, from three sources associated with the tweets, people rated dogs humorously and gave rating of above 10. The goal of this project is to wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations.

After scrapping and merging the data, Quality and tidiness problems were evaluated and cleaned.

At the time the data was collected, the WeRateDogs account had 4 million followers at the time, this may have contributed to tweets getting more favorites and retweets. *WeRateDog* got over 5000+ tweets but I analyzed around 1500+ tweets. Finally, three visualizations were created and the insights are shown below.

1. The most rated dog was the golden retriever with 21.23% ratings. We next plotted the most common (actual) dog breeds in a pie chart. We see that according to this data, the most common dog breed in the tweets was the golden retriever with over 150 tweets of the dog breed. The next four most common breeds, in descending order, were the

Labrador retriever (14.78%), the Pembrooke (13.84%), chihuahua (12.42%, and the pug (8.46%).
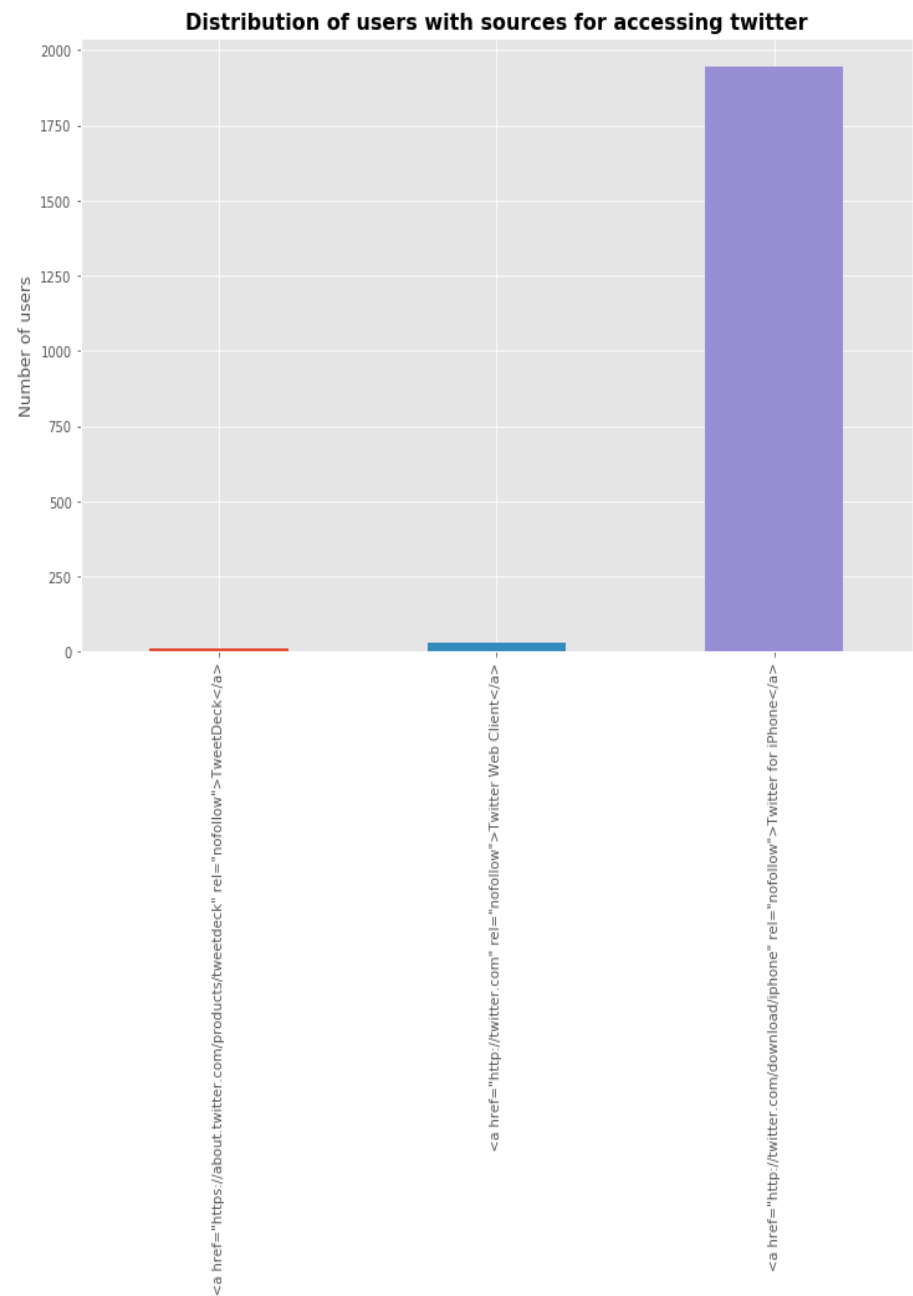
Distribution of breeds with popularity in twitter
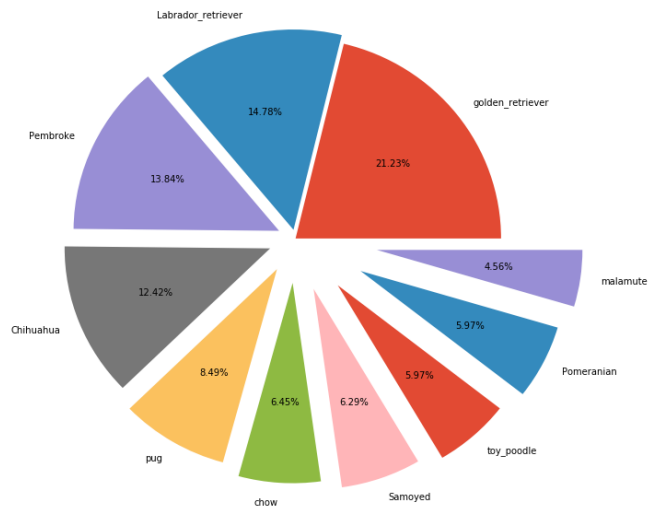
2. The most common rated dog stage was the *Pupper* at 203 ratings then *Doggo* 63, followed by *Puppo* with 22 and lastly *Flooper* with 7 ratings.
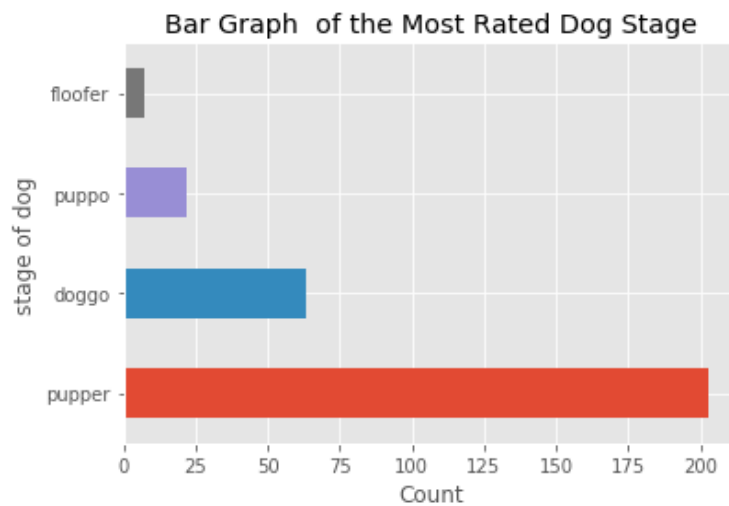
3. The most common used source to rate the dogs were iPhone followed by twitter web client and lastly tweet Deck

Distribution of users with sources for accessing twitter

In conclusion, this is real life data and imperative to know that data may not come from a single source. It may need to be extracted from various sources before analysis is done. The wrangling emphasizes on the tidiness and quality issues that occur in the dataset and how to deal with the issues.