PROJECT 2 - DATA WRANGLING

Insights:

In the last data cleaning step of this project, all three (3) tables involved in the project were merged to form a master dataset named **df_all**. The description of this **df_all** table was displayed as well as the value counts of "dog_stage" and "source" columns of the **df_all** table. The following insights were revealed after these actions were carried out;

- Of all stages of dogs captured in this dataset, pupper was the most represented dog stage. This shows that, the majority of the dogs rated by WeRateDogs within the period of consideration of this dataset were child dogs (pupper).

- WeRateDogs uses the twitter app for iPhone mostly.

- The minimum rating of dogs by WeRateDogs is **zero** (0) while the maximum rating WeRateDogs ever gave a dog is **177.6** (that is, 1776/10).

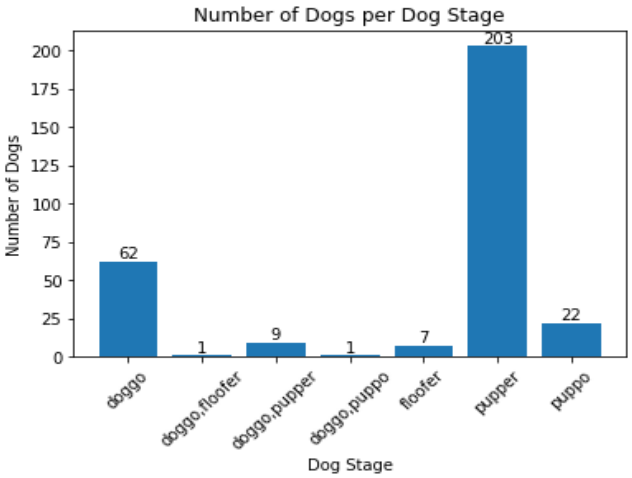
- The minimum number of retweets WeRateDogs got for a tweet is **11** while the maximum number of retweets is **70,721**.

- The minimum number of likes WeRateDogs got for a tweet is **66** while the maximum number of likes is **144,860**.

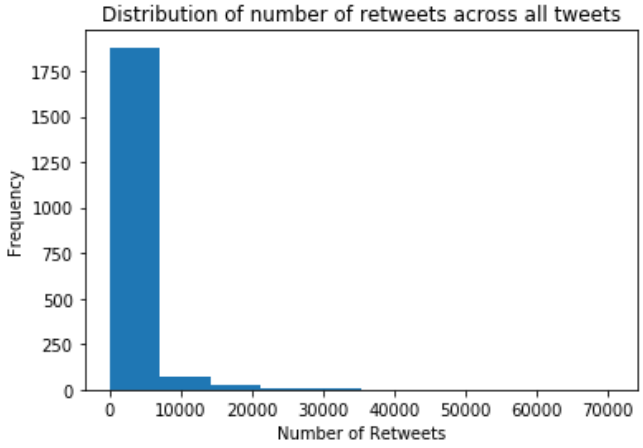
Visualizations:

The following visuals were produced from the wrangled data of this project;

1. Number of dogs for each dog Stage. This visual was created using matplotlib's plt.bar() function.



2. Distribution of retweet_count across the tweets of the dataset. This visualization shows that the distribution is skewed to the right. The visual was created using matplotlib's plt.hist() function.



3. Distribution of favorite_count across the tweets of the dataset. This visualization shows that the distribution is skewed to the right. This visual was also created using matplotlib's plt.hist() function.