Analysis Report - WeRateDogs

Introduction -

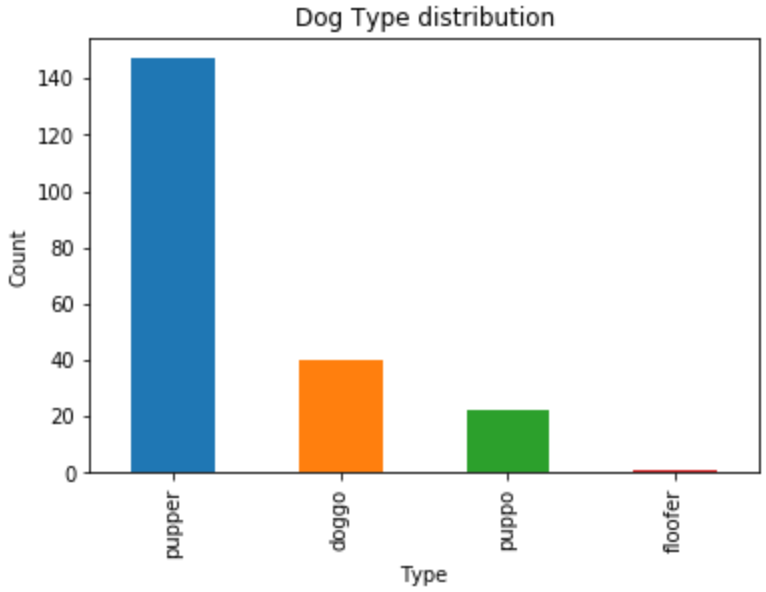
WeRateDogs gives rating to dogs in a unique way and also provide review for them in tweets. The dataset have been analysed briefly using some of the key variables and are explained below with the posed question.

Following questions have been kept in focus before analysing dataset -

1. How distribution of dog type looks over dataset available ?
2. How ratings distribution looks for overall dataset ?
3. How distribution of sources for dataset looks ?
4. Popularity of WeRateDogs over time using Favorite Count and Retweet Count ?
5. How favorite and retweet count related with each other ?

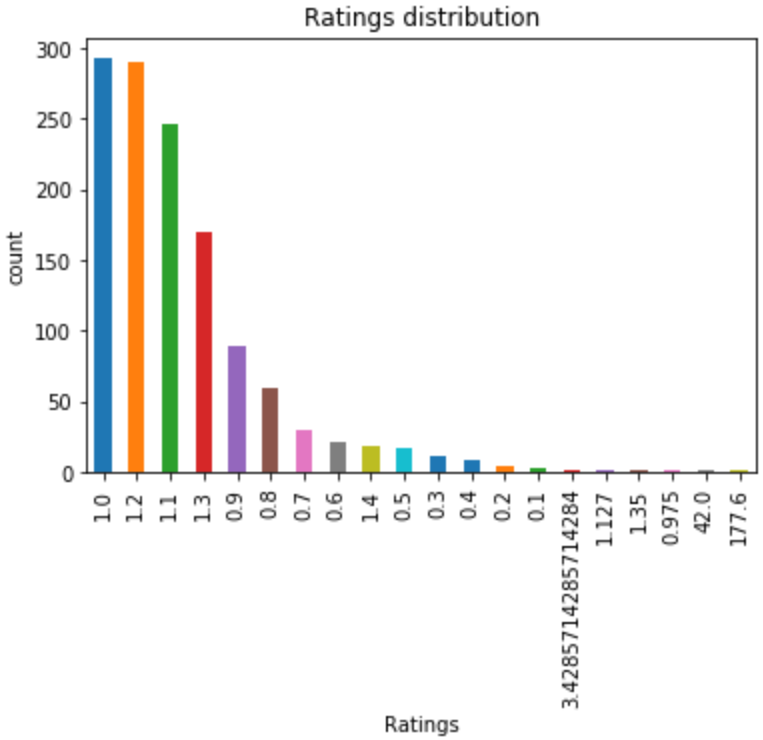
1. Dog type distribution -

Dataset contains majority of dogs of type “Pupper” while least is “Floofer”. This shows the unequal distribution of these type which can be because of two reason either Pupper is famous or there has been biasing about this type of dog.



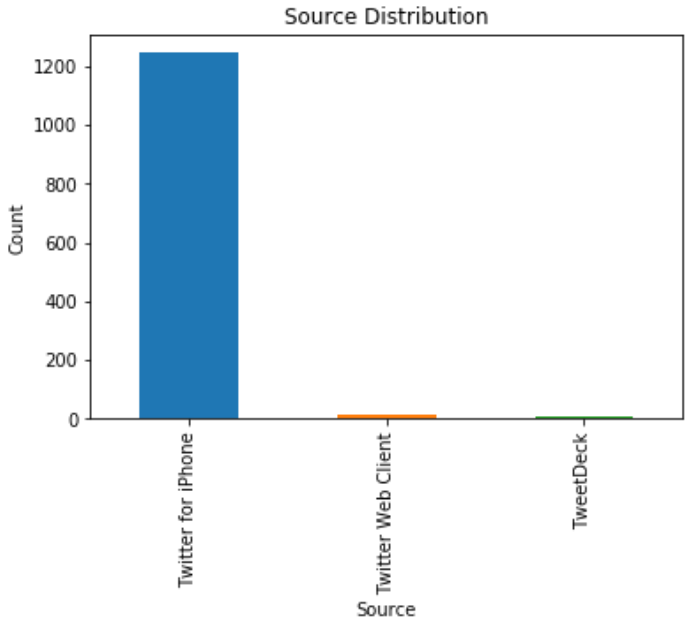
2. Ratings Distribution -

Each dog has been rated in a unique way and the ratio has been calculated and plot. It was visible that 1.2 rating is most occurred and rating less than 1 and greater than 1.5 is less.



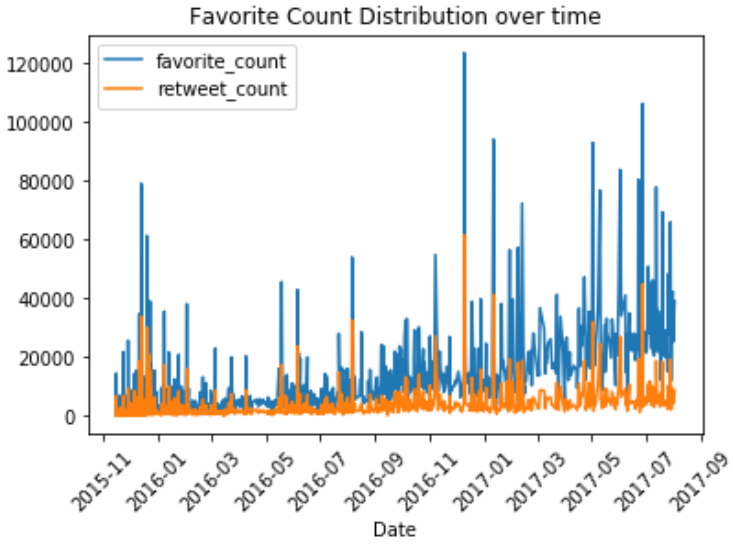
3. Sources Distribution -

Dataset provides the source of the tweet like IOS, web, etc. It is observed that the poster used Twitter for IOS for majority of the tweets.



4. Popularity of WeRateDogs -

One way to look how popularity of WeRateDogs has been over the time is by looking in to Favorite count and Retweet Count distribution.



Its clearly visible that over time popularity has been increased ,as both mean Favorite and retweet count overtime is increasing by date(Excluding Outliers).

5. Relation between Favorite Count and Retweet Count

It looks like favorite and retweet are related to each other and the scatter plot I made confirms the same. The plot shows high correlation between these two .