Visualization and Analysis

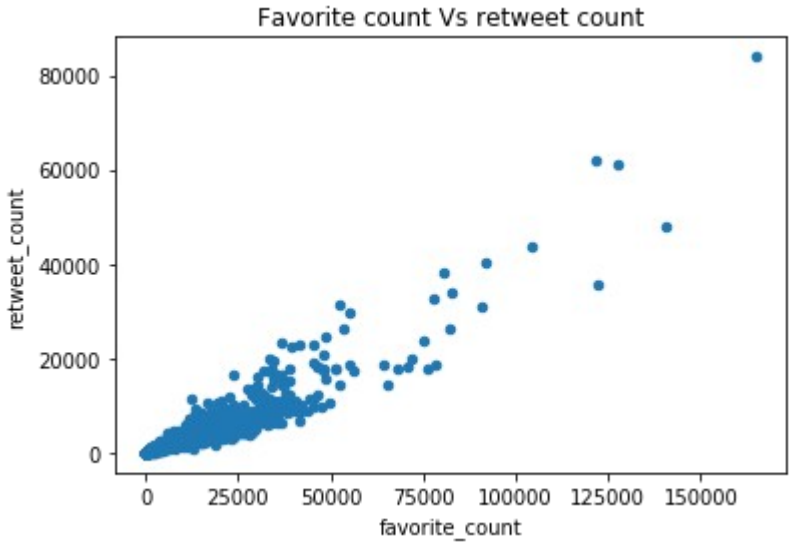
WeRateDogs Twitter archive

Nisrein Sada Dec,21th, 2018

In this report I looked into WeRateDogs Twitter Archive. The questions that i am interested to look into:

1) relationship between retweets and favorite counts:

In order to study if these two variables are correlated I used the scatter plot and calculated the pearson correlation coefficient.



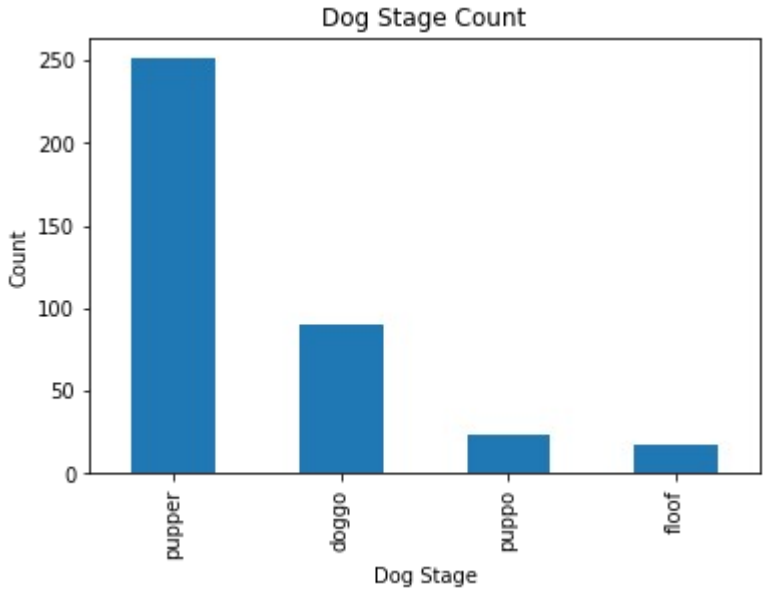
-The scatter plot shows how these two variables are correlated in what seems to be a strong correlation where an increase in any of those is associated with a higher value of the other variable. So this implies that the more retweets we have the more the favorites we will have.
-The pearson correlation value was 0.9275 which is a high correlation factor.
- other things to note is that the favorite counts has higher values than retweets with a mean a median of 4049 for favorite count and 1330 for retweet count which might imply that people might add tweets to favorite than retweet the things they like.

2) Most popular dog names.

Count the number of each dog name and get the most common names and least common.
- the most common names included: Charlie, Lucy, Oliver, Cooper, Tucker, Penny, Sadie, Lola, Winston.
- the most common name was repeated for 11 times and the least common names for one time.
- 720 names appeared only once which implies that people tend to use unique names for dogs while names that appeared 11,10 and 9 times was for 2 dog names each.

3) Dog Stages Count

- Most common dog stages in this dataset. So we used the dog stage to count the presence of each stage.
- the most prevailing dog stage is the pupper (65%)and the least present in the data is the floof(4.69%).



4) Dog stages vs favorite count.

- Group the data by the dog stage and get the average favorite count for each group.
- We can see that the puppo has the highest average favorites count though it was the 2nd least present in our data as mentioned in the previous section and the pupper which was the most present in our data had the 2nd least number of favorites count.
- so dogs appearing least in the dataset got more favorites.