ACT_REPORT

After cleaning all the three dataframes, I have merged them and created a master datadrame named as twitter_archive_master. This dataframe have been used to plot all the interesting features of the dogs. Here are few insights which I could collect:

1. First of all looks like almost all people accessed twitter through their iphones.

2. Highest rating dogs got is 12.

3. In this fig 1 given below, I have plotted the histogram of the dog breeds (as here x axis is label so I took only top 25 value counts because of the readability issues) and I see that most of the photos are of golden retrievers (looks like golden retrievers are the most famous dogs among all dog breeds). The second famous dog seems to be labrador retriever and then other breeds have smaller counts.

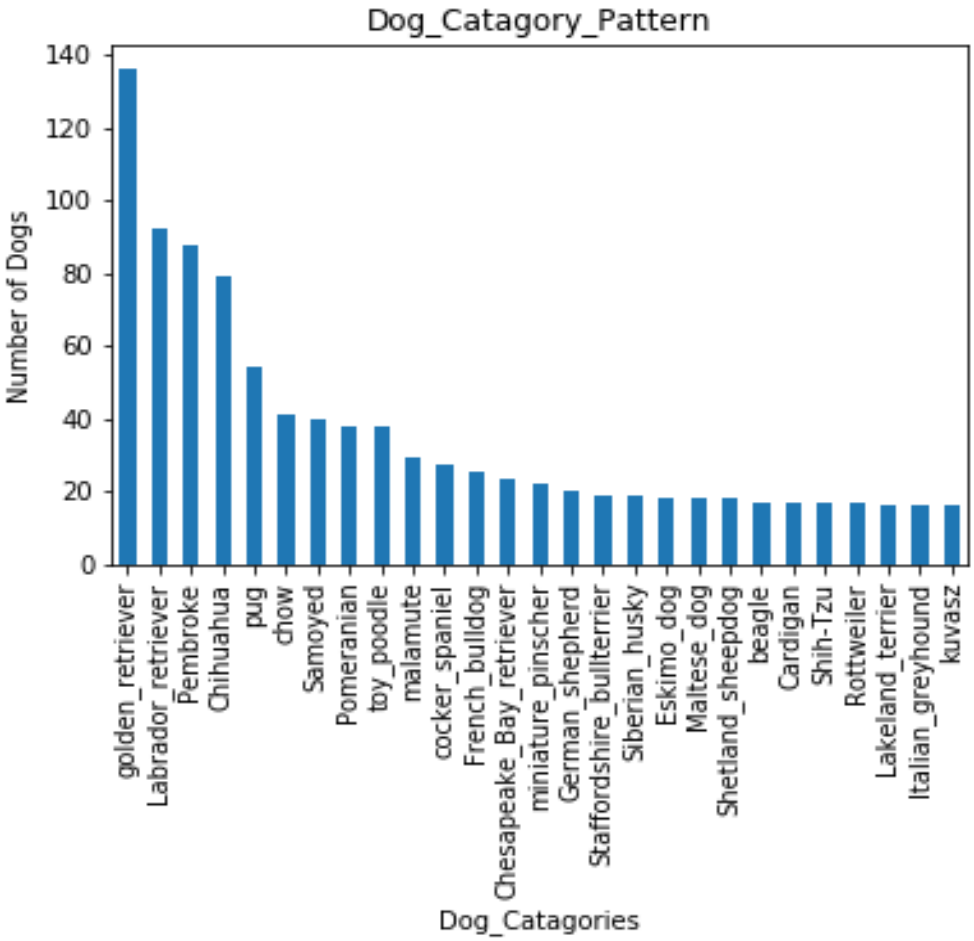

Fig 1

Here is our most famous cutie:

4. In fig 2, I have shown a scatter plot of dog rating (rating_numerator only as all the denominators are 10) with the favorite count and I notice that favorite count increases with increase in the rating_numerator meaning dogs with higher ratings are mostly liked by the people (rating 14 seems to be an exception, may be people didn't like the breed of the dog etc.)

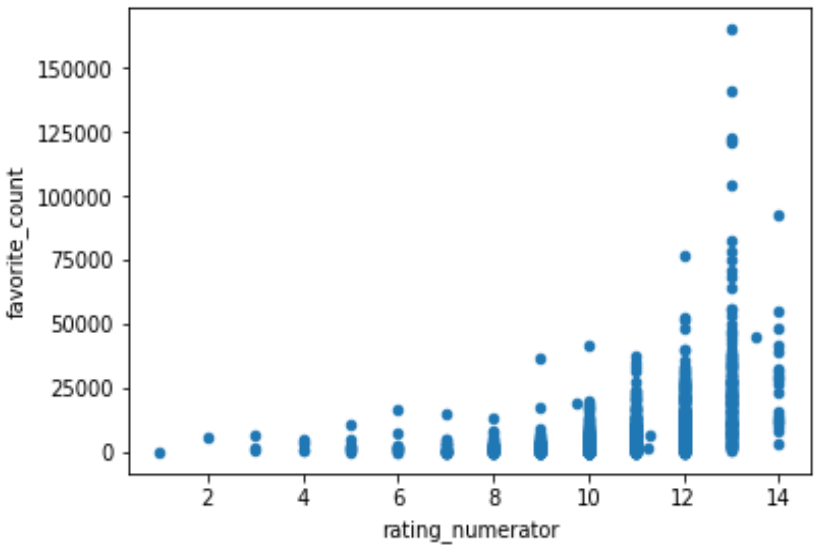

Fig 2

5. In fig 3, I have shown a scatter plot of dog rating (rating_numerator only as all the denominators are 10) with the retweet count and as with the favorite count I notice that retweet count also increases with increase in the rating_numerator meaning highly rated dogs get most retweets (rating 14 seems to be an exception here too)

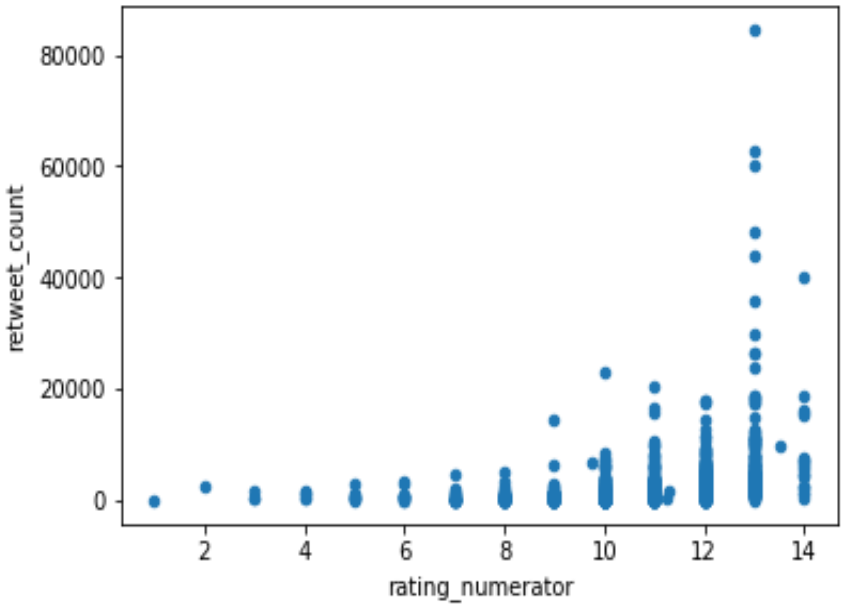

Fig 3

6. From the last two plots given above, I see that the behavior of both of the plots is same so it would be interesting to plot favorite count vs retweet count and I found that the dogs with higher numbers of favorite counts got higher number of retweets.

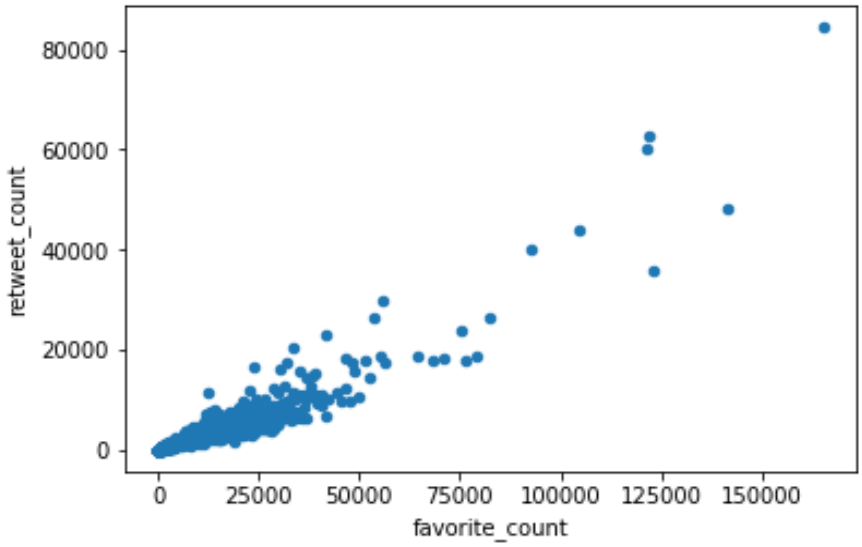

Fig 4

References:

https://pixabay.com/en/golden-retriever-dog-good-dear-642016/