Data Analysis and Visualization Report

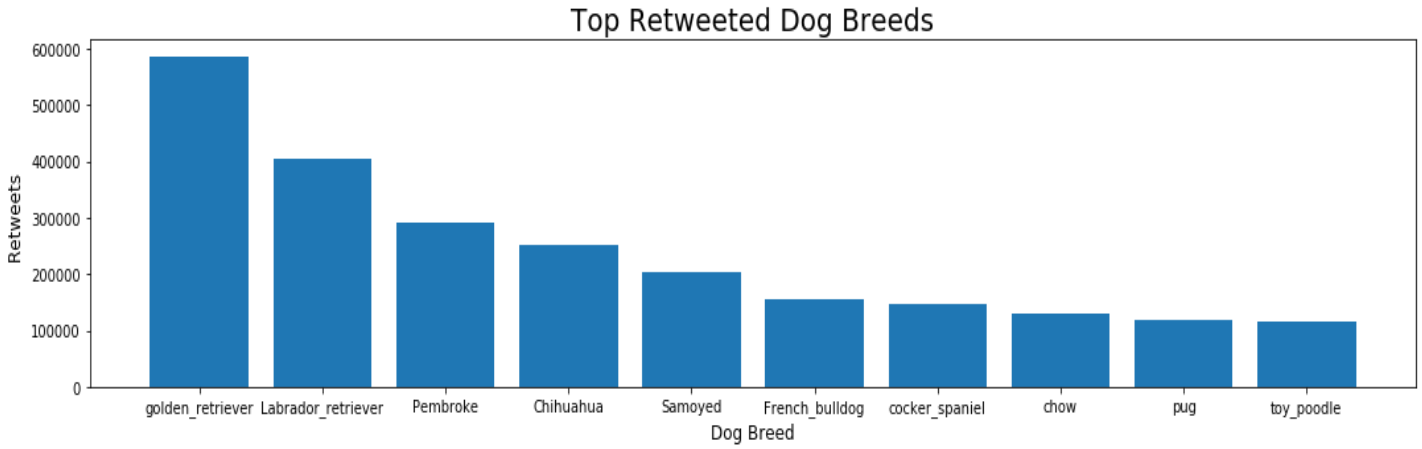
Since the rating values in for different dogs in the account is arbitrary the most important indicators that can be visualized for this data is the retweet and favorite count.

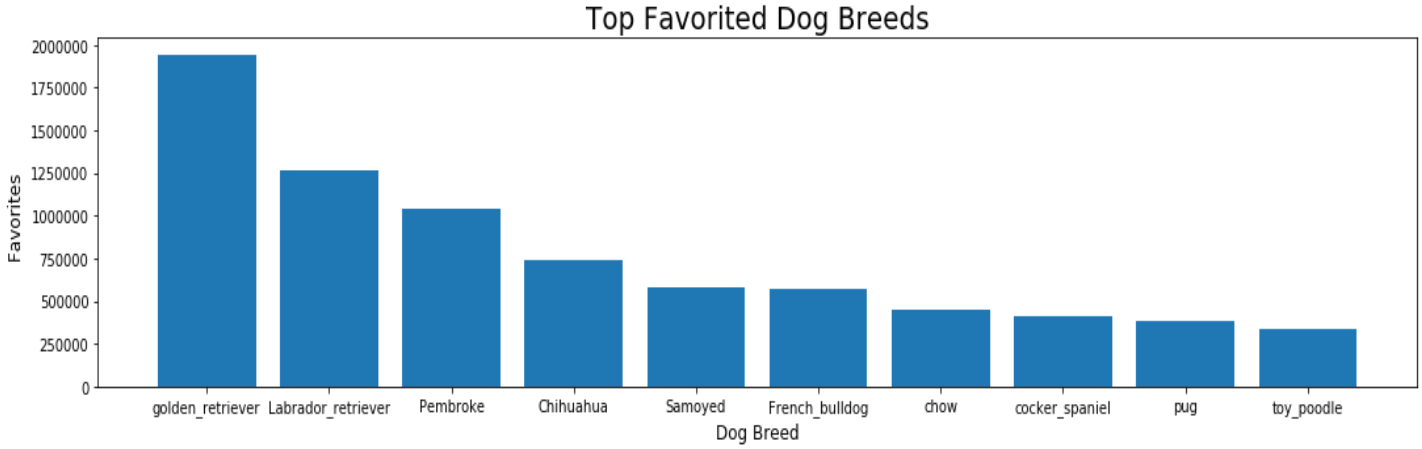
In the analysis and visualization phase, the summary statistics for the dataframe was calculated using the describe function, and visualizations for the top interacted tweets sorted by dog breeds and dog stages is presented.

	rating_numerator	rating_denominator	retweet_count	favorite_count
count	2094.000000	2094.000000	2094.000000	2094.000000
mean	12.191500	10.449379	2844.039160	8987.330468
std	40.393858	6.649800	4712.115669	12207.002968
min	0.000000	2.000000	16.000000	81.000000
25%	10.000000	10.000000	644.500000	2045.250000
50%	11.000000	10.000000	1404.000000	4195.500000
75%	12.000000	10.000000	3289.500000	11443.000000
max	1776.000000	170.000000	79515.000000	132810.000000

In the above table, resulted by the describe function we can highlight that the rating numerator have a mean of 12.2 which means that mean was not affected by the outlier values that can be clearly seen at the max value of 1776. This could be attributed to the fact there are a lot of values in the dataset and a view outliers does not affect the overall quality of the outliers. The same can be said about the denominator values as the man is 10.45 and the max value is 170.

Two things to note in the retweet and favorite count is first that none of the tweets had zero retweets or favorites, this warms my hear as none of the dogs got no love :). Second, tweets get more favorited than it get retweeted, which an important indicator of how people use their twitter accounts and explains why twitter started making the favorites of people a person follows appear in their timeline to increase tweets interaction.

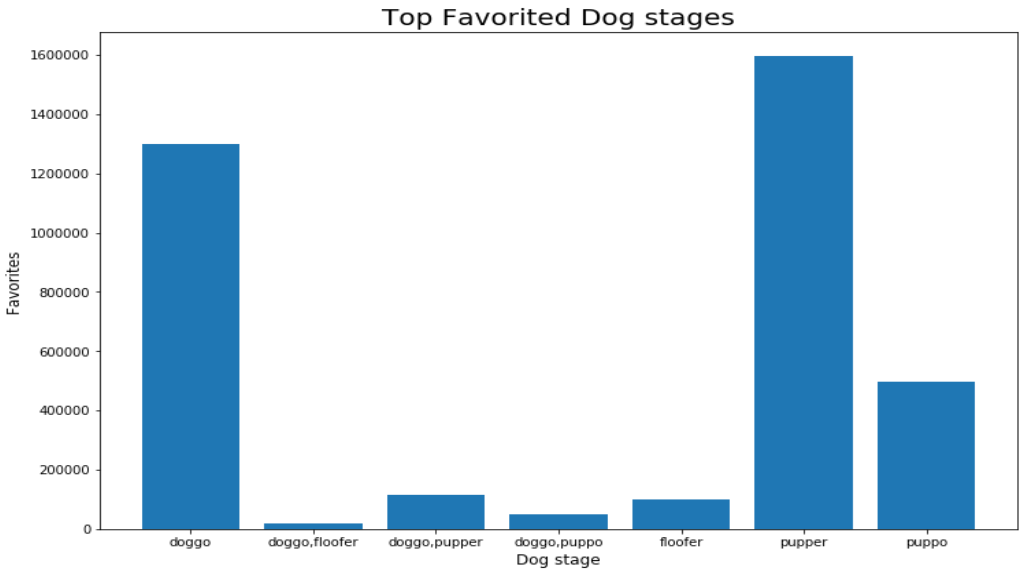


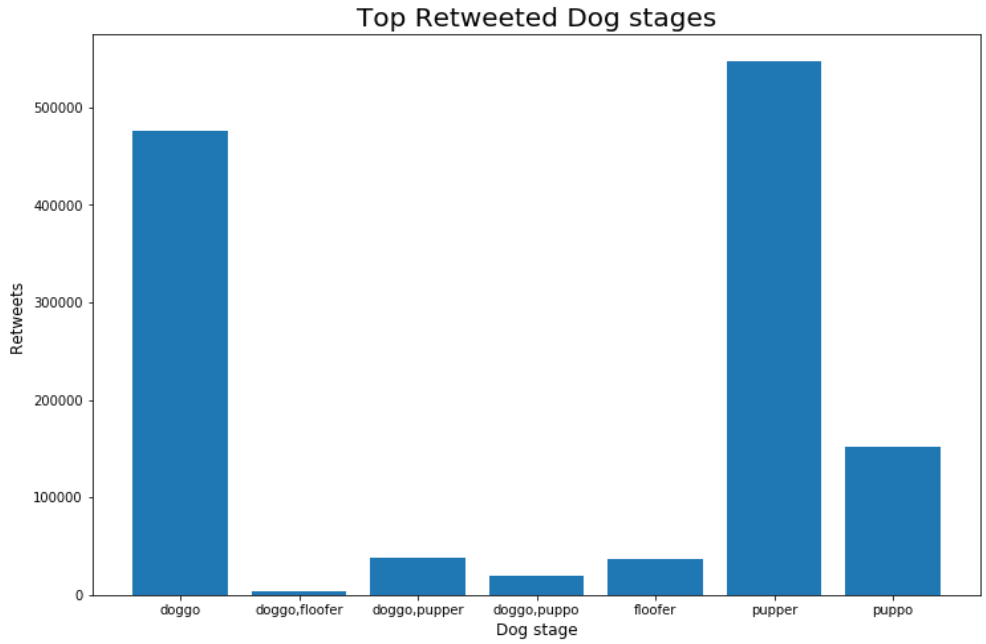


The above shown graphs shows the top interacted dog breeds in terms of retweet and favorites. The two graphs are separated because of the difference scales, we could us two scales but it could mislead the reader in terms of proportion.

We can see that the golden retriever gets a lo of love, which does not explicitly mean that people like golden retriever pictures but could just mean that there are more pictures of golden retrievers in the account.

Another thing to note is that there is high correlation between number of retweets and number of favorites. Although the amount of retweets and favorites differs, there are no tweets that are breeds that are highly retweeted by not highly favorited. Only difference in the two above charts is that the cocker spaniel preceded the chow ranking in the number of retweets while the opposite happened in the number of favorites.





As for the two above plots, it shows the number of interaction for each one of the dog stages. Before creating these plots, the entries that were not identified as any of the above stages were eliminated as they are not needed in this visualizations.

Puppers and doggos get most of the love, with a total of more than 2.5 million favorites and around 1 million retweet. According to the definitions, those two stages are usually the youngest which could explain why they are the most interacted.

Again, the two plots are almost identical, which shows that tweet retweets and favorites are highly correlated.