Analyzing and Visualizing WeRateDogs Twitter Data

The tweet archive of Twitter user @dog_rates, also known as WeRateDogs was analyzed and visualized in this report. The data obtained from three different sources were assessed and cleaned prior to analyzing. Matplotlib and Seaborn libraries were used for the visualization. The following insights were obtained after the data analysis:

Favorite count vs Retweet count

The Favorite count is plotted against Retweet count in a scatterplot in Figure 1 to observe the correlation between the two. It can be seen from Figure 1 that the Favorite count is positively correlated with the Retweet count. The majority of the data falls between 40000 favorites and 10000 retweets. So, for every 4 favorites there is 1 retweet.

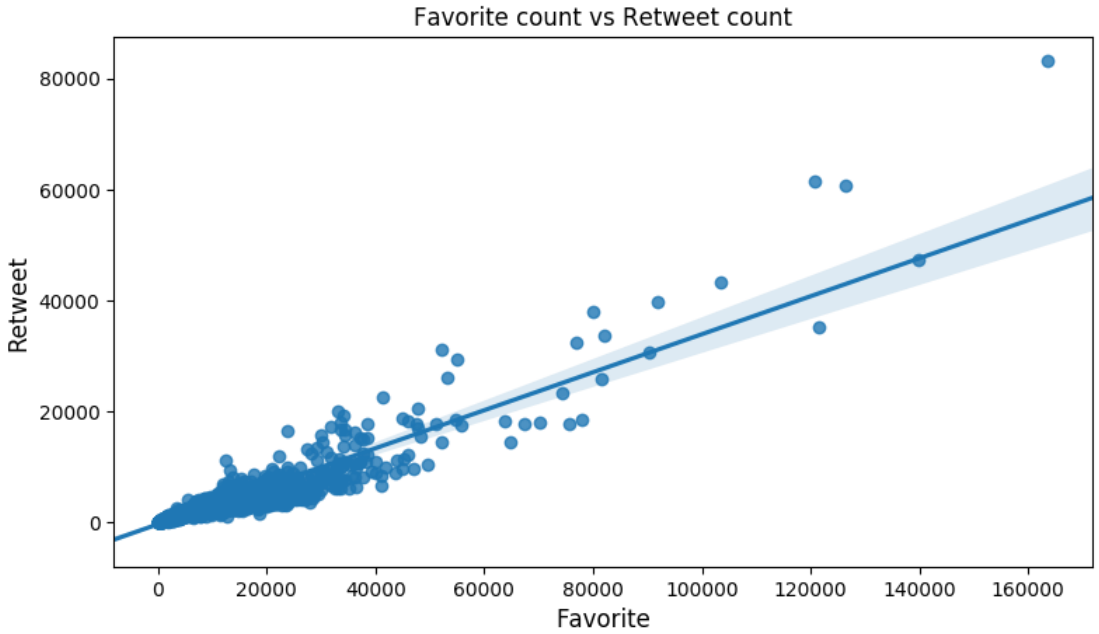

Fig.1: Favorite count vs Retweet count from 887 tweets from Twitter user @dog_rates, also known as WeRateDogs.

Table 1: Most frequent ratings obtained by different dog stage

dog stage	count	unique	top	frequency
doggo	52	6	12/10	25
floofer	3	3	11/10	1
pupper	192	19	11/10	55
puppo	21	3	12/10	9

The above table obtained by analysis of the rating of dog obtained in different stages shows that the maximum number of rating doggo and puppo obtained is 12/10. Similarly, for pupper it is 11/10. Floofer only has three entries with three different ratings.

Retweet counts and Favorite counts on the basis of dog stages

The box-plot of the retweet counts and favorite counts obtained by four different dog stages (doggo, puppo, pupper and floofer) are shown in Figure 2 and Figure 3 below respectively.

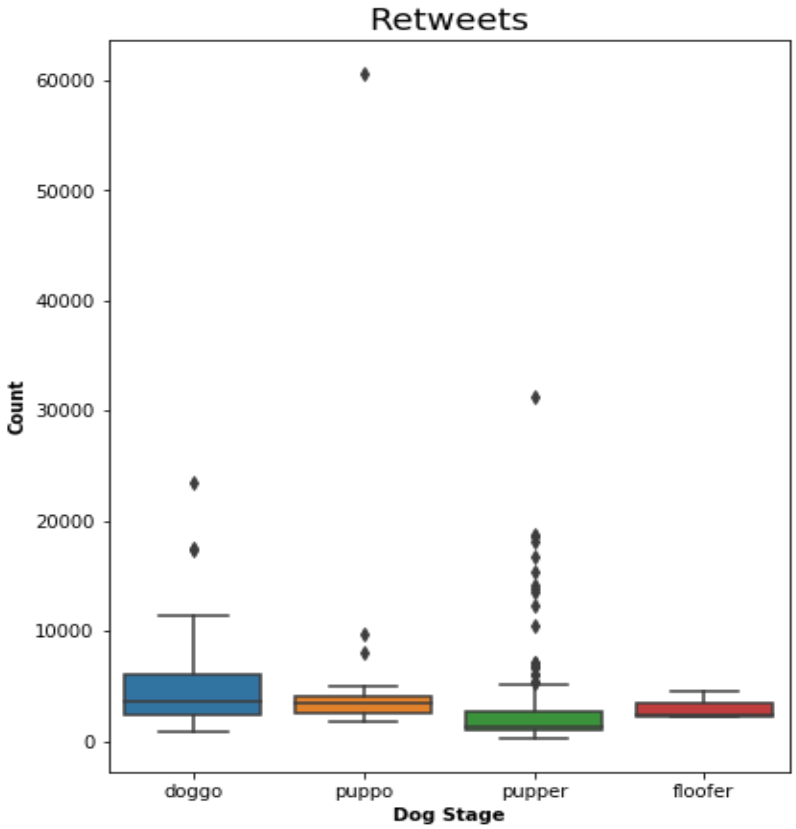

Fig.2: Box-plot of the Retweet counts obtained by four different dog stages.

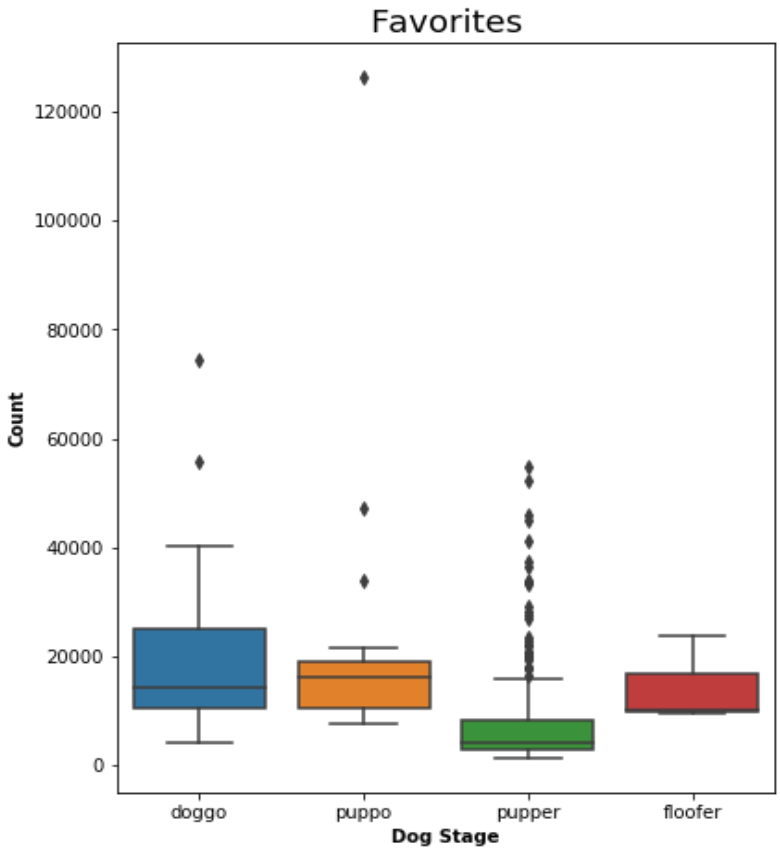

Fig.3: Box-plot of the Favorite counts obtained by four different dog stages.

From the above plots of retweet count and favorite count for different dog stages shows that the mean retweet count and favorite count for doggo and puppo are higher than that of pupper and floofer. Although pupper obtained the lowest minimum count of mean favorites and retweets, the outliers are maximum for this stage of dog. The reason behind this can be that the maximum number of dogs enlisted by WeRateDogs is in pupper stage followed by doggo.

Most common dog breeds

The 12 most common dog breeds according to the data of WeRateDogs are shown in Figure 4.

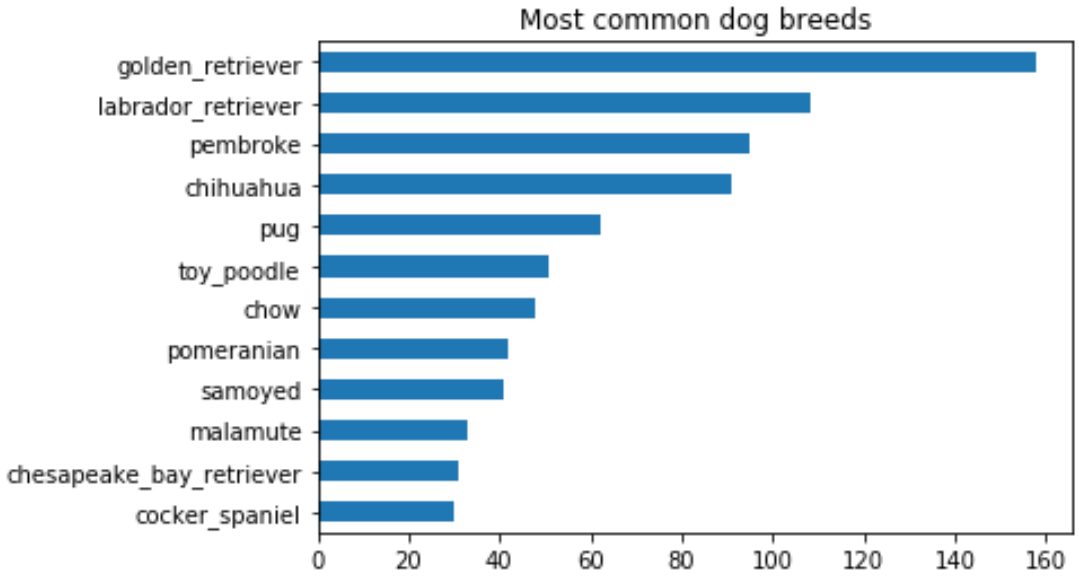

Fig.4. Twelve most common dog breeds according to the data of WeRateDogs.

It was found from the above analysis that golden retriever is the most common dog breed, followed by labrador retriever, pembroke, chihuahua, chow and so on. These five most common dog breeds match well with what we see on the internet. So, the distribution of most common dog breeds obtained from the data of WeRateDogs seems to be acceptable.