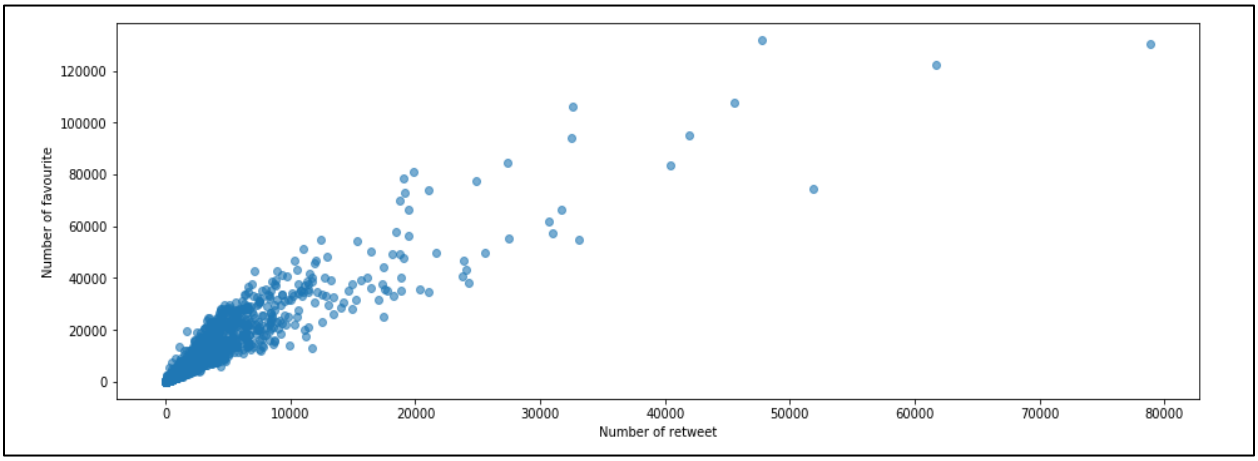
Insight report

After the analysis toward the dataset given. There were some insights found. The dataset only remained 2175 rows after removing the retweeted post during data cleaning process. The analysis is based on these 2175 rows.

The analysis is started with looking for the relationship between number of retweet and number of favorite (like). Figure 1 shows the scatter plot with number of retweets on x-axis and



number of favorites on y-axis. The chart shows that there is a positive correlation between both variables.

Figure 1: Scatter plot between number of retweet and number of favorites

Besides using the scatter plot to visualize the relation, linear regression also applied to find the relationship. Based on the model generated using both variables, it shows that the variables have a strong positive correlation coefficient with value of approximately 0.9149 and the number of retweet is statistically significant in relating to number of favorite proved with the p-value of 0. Another finding of this model is, we can expect the number of favorite increases by 2.3875 as for each retweet increment for the Twitter post. Q

The second analysis is to find out the ratio of the models predict the image as dog or failed to predict as dog. Figure 2 shows the portion of images that was fail or successful predicted as dog.

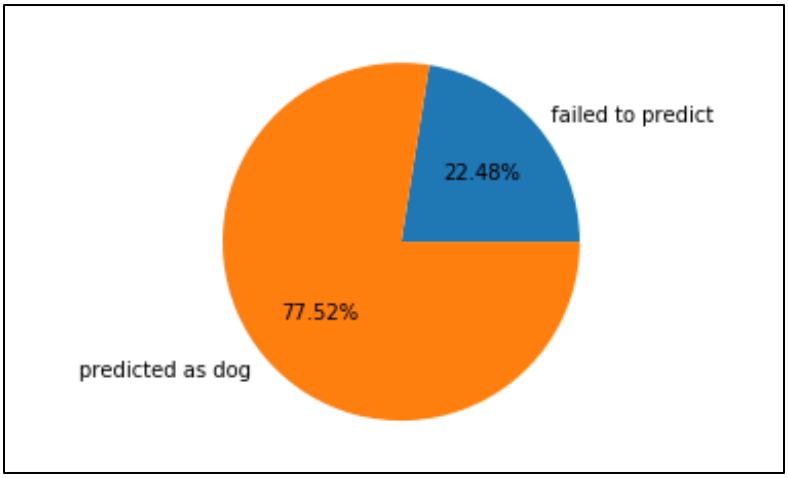


Figure 2: The portion of success/fail predict as dog

The chart shows that there were 77.52% of images were predicted as dog and 22.48% remained were failed.

The third analysis made is to find out the top 10 breed of dog from the dataset. Figure 3 shows the histogram of top 10 breed with breed on x-axis and number of image on y-axis.

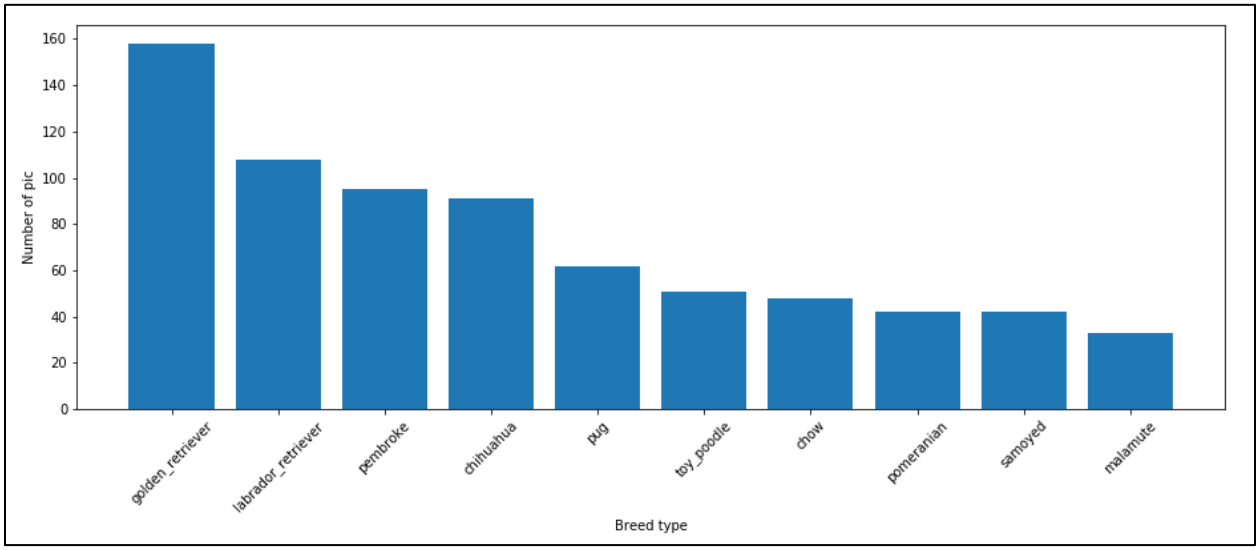


Figure 3: Histogram of top 10 breed

The first in th3e list is golden retriever with the number of 158 then followed by labrador retriever (108), pembroke (95), chihuahua (91), pug (62), toy_poodle (51), chow (48), pomeranian (42), samoyed (42), malamute (33).