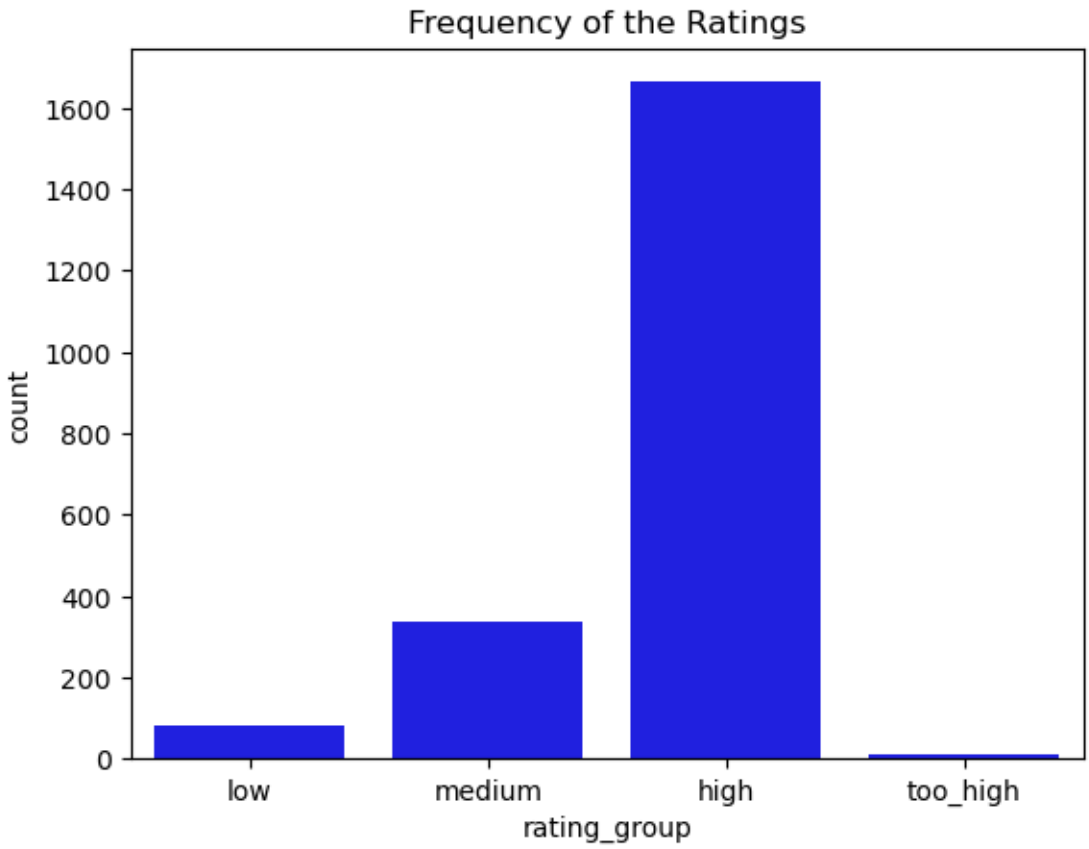
Analysis of the Dog Rating Dataset

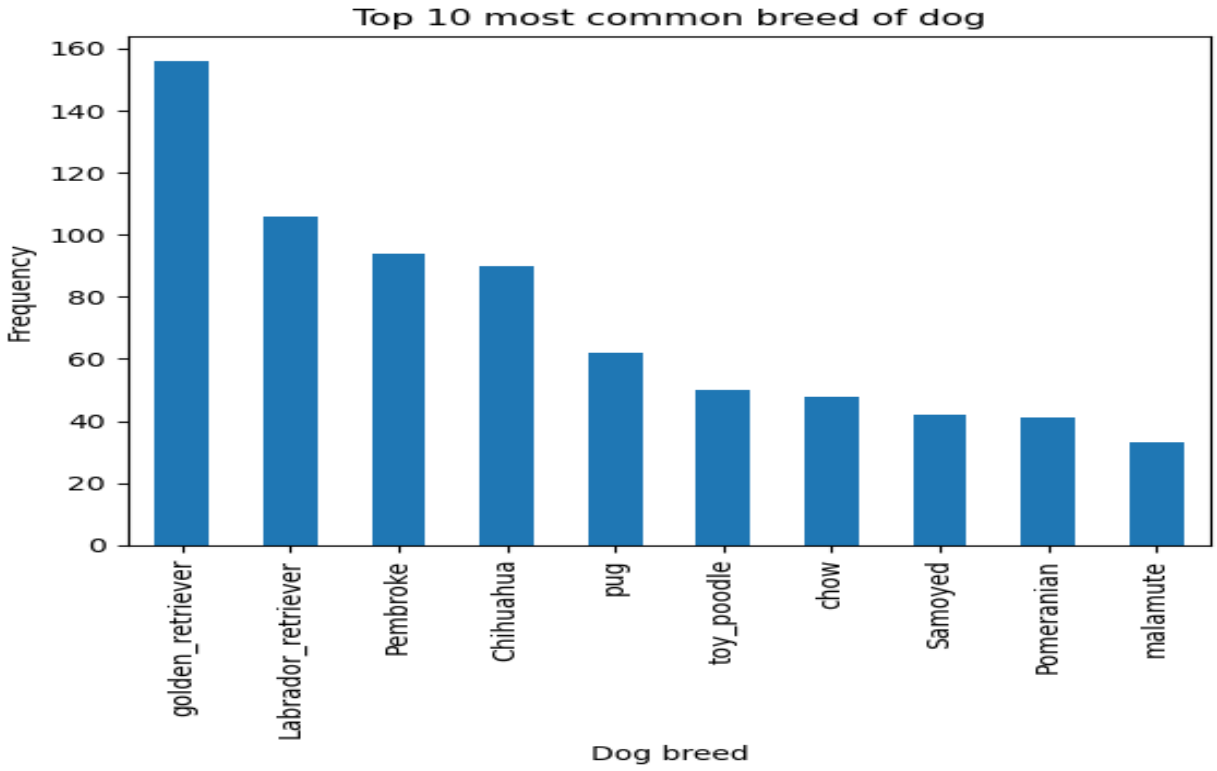
RESULT REPORTING

After wrangling of the dataset, analysis was done to get insights from the dataset. The analysis was also visualized.

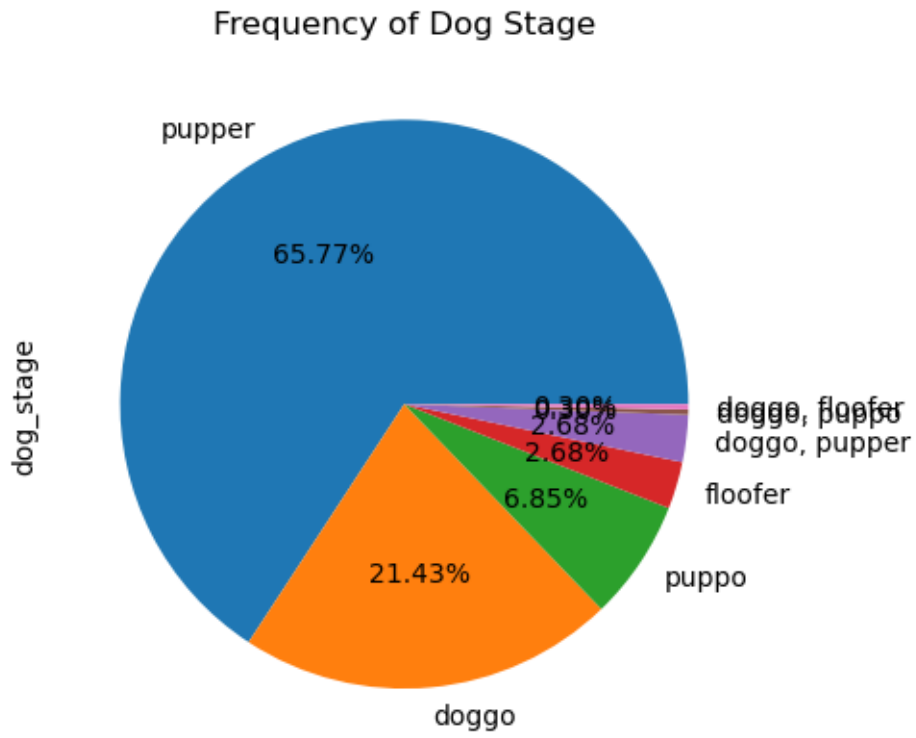
Most of the ratings were greater than or equal to 10, which shows a good percentage of the dogs were rated highly by the twitter user.



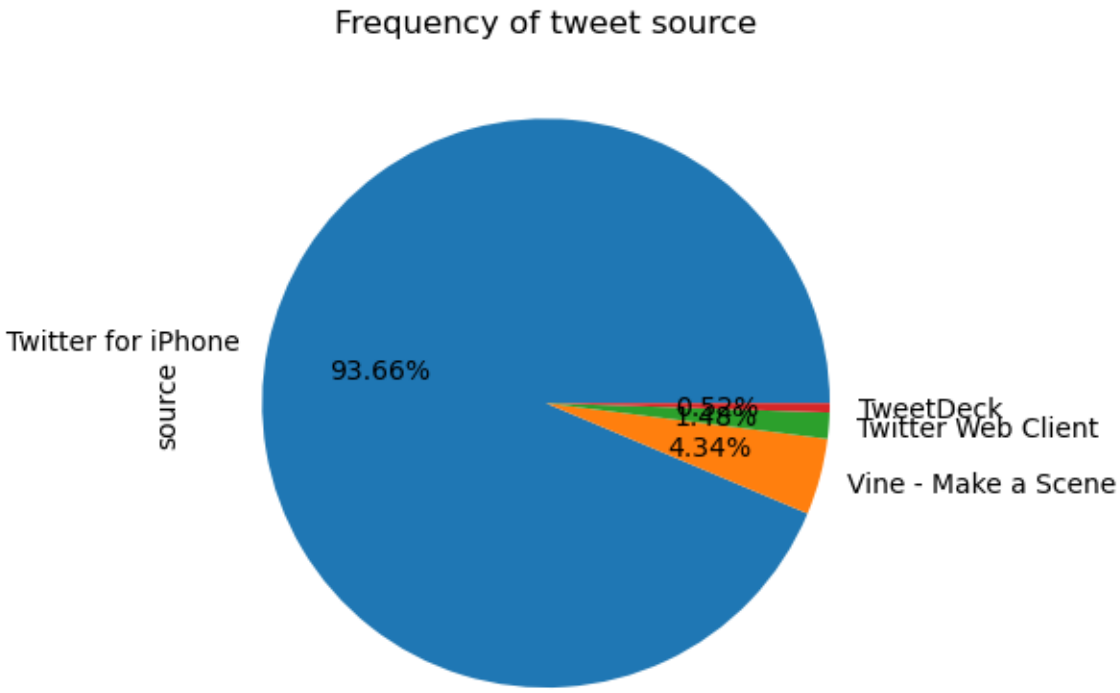
During analysis, we saw that 2 particular dog breeds were the most common, those 2 breeds also happened to be the breeds with the highest rating, with ratings greater or equal to 10.



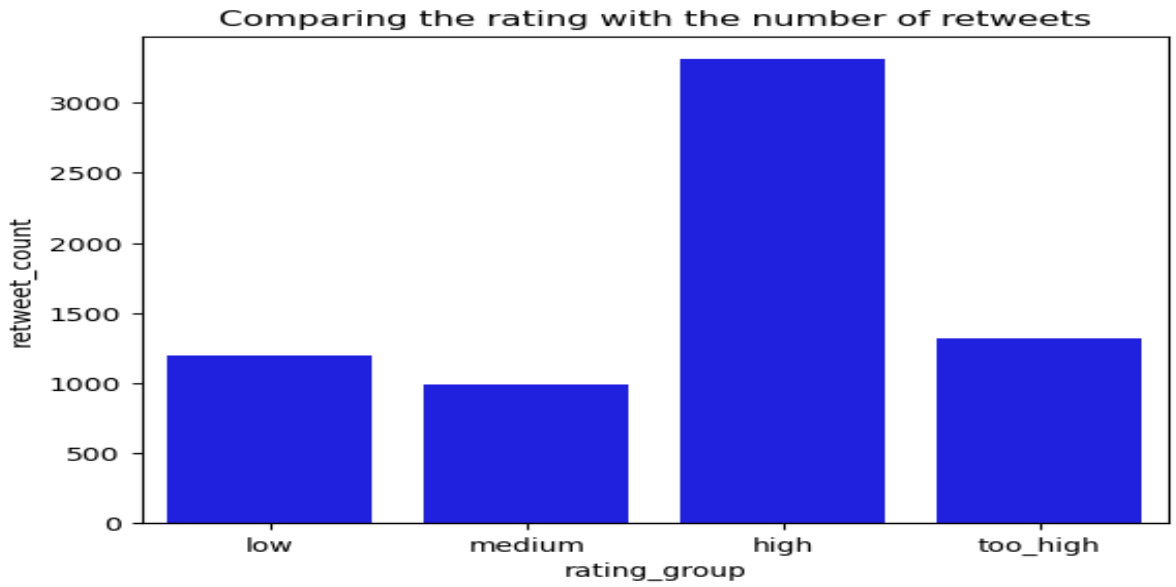
The dogs are of different stages, there are 4 stages of dog, the doggo, puppo, pupper and floofer. More than 65% of dogs in the dataset were puppers, followed by the doggo comprised over 21% . The pupper dog stage also had the highest ratings.

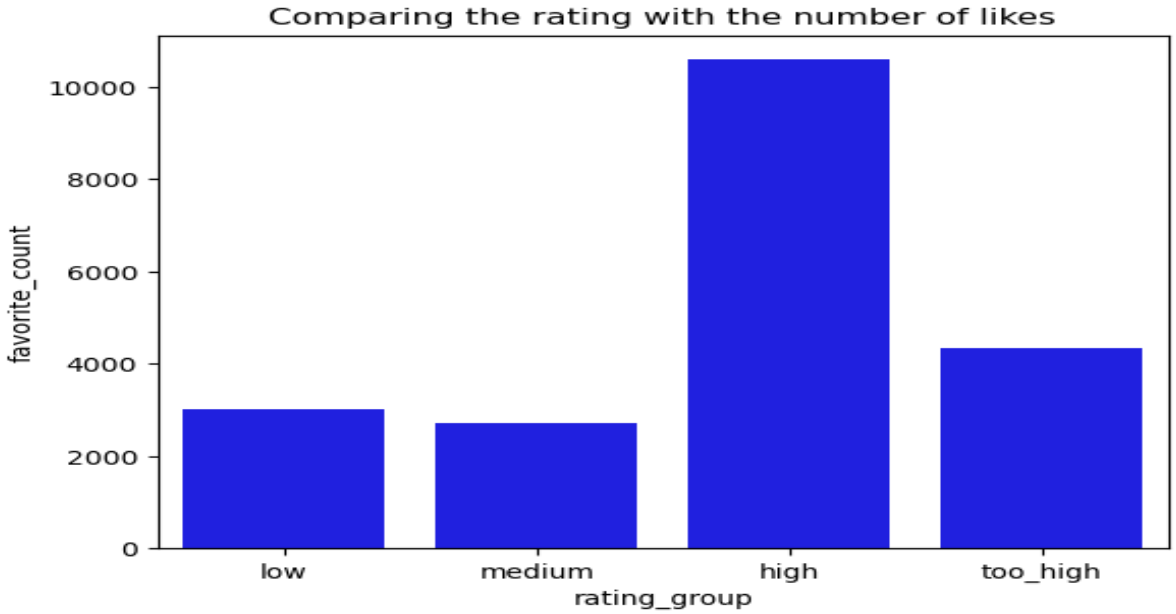


Most of the tweets of dogs which were rated were made by iPhone users. iPhone users account for over 90% of the tweets



Comparing the number of likes and retweets with the ratings, we can see that there is not a bad correspondence. This shows that the ratings of the twitter user WeRateDogs can be trusted

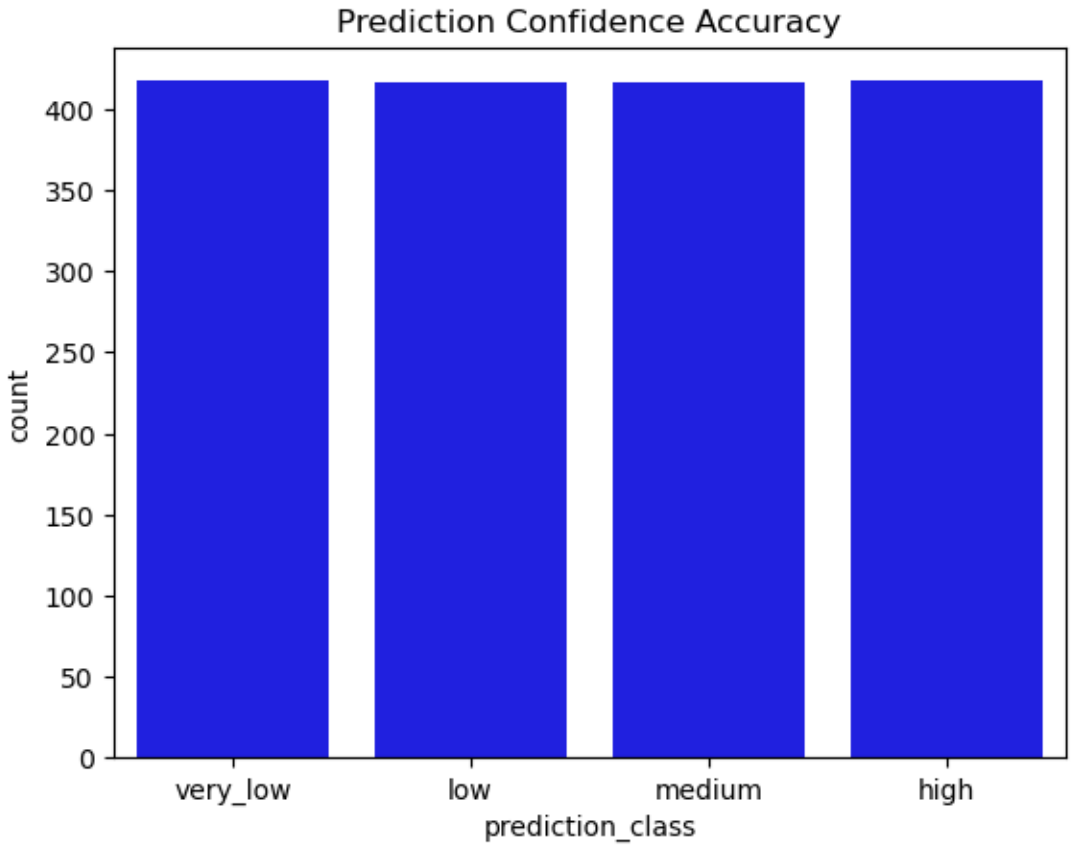




The accuracy of the algorithm used to predict the breed of dog cannot be 100% trusted as its confidence interval was a bit low.

Very low predictions fall in the range 0 - 0.305955
Low predictions fall in the range 0.305955 - 0.550914
Medium predictions fall in the range 0.550914 - 0.822939
High predictions are greater or equal to 0.822939,with the maximum being 0.999956

The prediction confidence of all the classes has virtually the same count, this brings doubt about the accuracy of the algorithm



CONCLUSION

Most of the dogs were rated high by the twitter user which was backed up by the likes and retweet counts. The golden retriever and the labrador retriever were the breeds with the highest ratings.

LIMITATION

The accuracy of the algorithm in predicting the dog breed. This could have been as a result of unclear images.