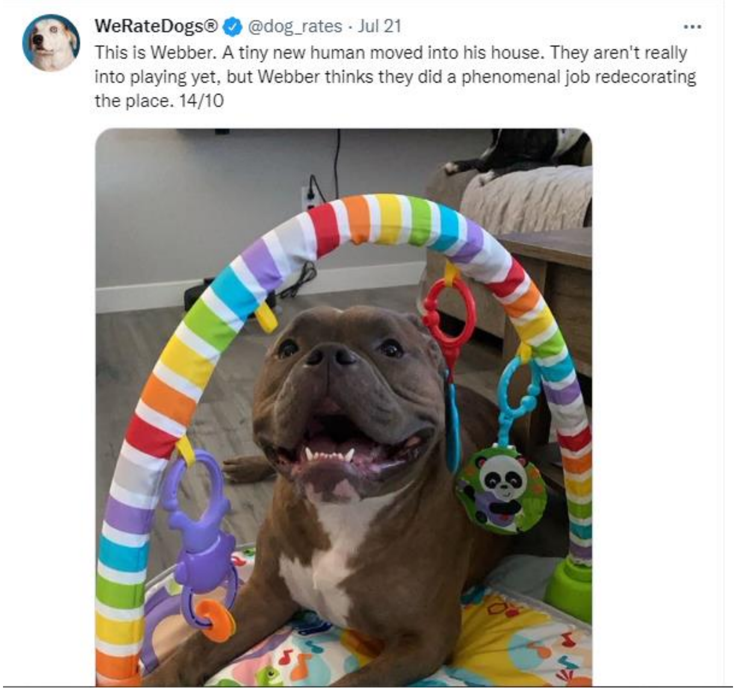
The Role of Correlation Analysis in Predictive Modeling

Correlation analysis is one of the important techniques for pattern tracing in data analytics as it can be used to check for relationship between variables(predictor and response).This can aid scientists/researchers in their predictive modeling process such that they don't have to always exhaust all variables in a dataset on bad quality models.Rather,correlation analysis can be used to check for closely related variables that can be used for forecasting even if it means gathering more datasets that could aid that.

One of the qualities of a good predictive model is using multiple datasets which I would be demonstrating from the analysis of three datasets : WeRateDogs twitter archive dataset, WeRateDogs image prediction file(which contains top three predictions of dog breeds for unique tweet ids) and a new dataset I generated using this text file which contains number of likes and retweets for each tweet id.



WeRateDogs® @dog_rates · Jul 21

This is Webber. A tiny new human moved into his house. They aren't really into playing yet, but Webber thinks they did a phenomenal job redecorating the place. 14/10

Figure 2: *Image of a dog tweet and its rating of 14/10*

After undergoing the data analysis processes of cleaning and visualizing my datasets, I realized that we have 179 correct out of 220 predictions of an image containing a dog which shows that the predictive model (neural networks) might be able to predict future images with the question - 'Does this image contain a dog?'. There is also a very strong correlation of magnitude 0.93 which shows that, the higher the number of likes, the higher the number of retweets. This information can be used in the designing the next model since predictive modeling is an iterative process.

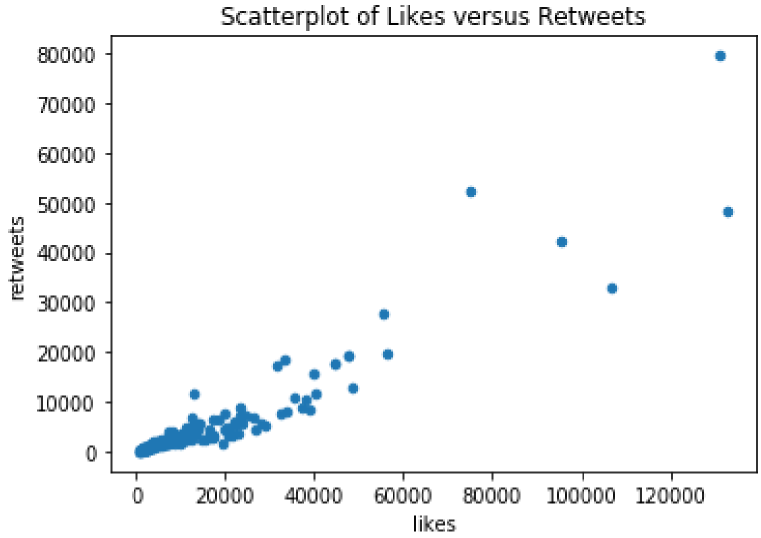


Figure 2 : *Diagram showing the strong relationship between number of likes and number of retweets*

We have no consistent predictions of dog breeds which may indicate that the neural network or kind of neural network might not be able to predict dog breed from future images.