Act Report

After wrangling the datasets, I analyzed and visualized them to get insights from the datasets. With the help of some pandas functions, I was able to get information on the final datasets. For instance, getting basic information about the final dataframes such as the shape, datatypes of each column, missing column values, etc. Moving on with my analysis, I wanted to learn more about the datasets.

Looking at the number of tweets on the WeRateDogs page, the number of likes and retweets tells us if people like the contents of their tweets or how interactive their page is. Even though their page has millions of followers, the likes and retweets also informs you how many active followers they have. The highest likes and retweets are about 132 thousand and 79 thousand respectively. Their percentage difference is almost half of favorite count, concluding to the fact that a lot of users tend to like a tweet more than they retweet a tweet.

The dataframe containing information about the image prediction model also provides insights on how this model works and its accuracy. Visualizing the confidence column (accuracy score on a scale of 0-1) shows how accurate the model is on all three pictures. The model has a higher accuracy score on the first pictures of a tweet and a lower accuracy score on the other two pictures of a tweet on the WeRateDogs page. In general, the model's effectiveness of about 0.7 which is not excellent but good.

Since WeRateDogs is mostly about rating dogs, finding out what kind of rating dogs get on this page is important. The project description tells us that ratings are on a scale of 10 but some of the rating is greater than 10. So I queried the dataframe to find out if more or fewer tweets have ratings greater than 10. Way more than half of the tweets have a rating greater than 10. Hence, most dog tweets have good ratings which may result from many different factors such as these dog tweets are adorable, huge dog lovers, etc.

Lastly, for the visualization, I plotted a bar graph to show the number of dogs in each dog type category. This gives us a clear picture of what percentage is in each group. Pupper has the highest height in the graph, which means there are more puppers than any other dog type in the dataset.