ACT REPORT

For my analysis, tables and plots, are created with the cleaned data from the previous phases documented in my wrangling report. The variables that are used primarily - rating, retweet_count,dog_stage,stage_count and favorite_count from the archive and the prediction table - have all undergone cleaning to various levels.

I first started by making a clean copy of the dataset.I then explored it using the describe function in pandas to see the summary statistics.

First Insight;

After using group by function and plotting in a bar chart. We saw that the ratings are from 10 and above, the dogs are 'good dogs' and ideally are not supposed to be rated less than 10.

Where I plotted retweet_count on the y axis and ratings on the x axis.

The Rating with the highest retweet is 14 and the rating with the lowest retweet is 26.

Second Insight:

Using the stage_count column. We saw that the 'None' value in the stage_count column has 7973 this is indicates that there were alot of None values before the melting and further investigation needs to be done on why the 'None' values are high.

From this dataset we can conclude that the dog_stage with the highest count besides the 'None' value is the Pupper.

Third Insight;

After grouping by and plotting the stage_count and favorite_count.

Where I plotted favorite_count on the y axis and dog_stage_count on the x axis.

From this visualization we can conclude that the dog stage with the highest number of likes is the Pupper.

Fourth Insight:

For this insight I wanted to see what source most tweets came from using unique tweet_ids and I grouped by source and used unique function in pandas.

Twitter for IPhone has the most number of unique tweets, followed by Vine- Make a Scene,Twitter Web Client and then TweetDeck with the least.

Insight Five;

Using the prediction dataset. 1780 twitter users had img_num '1' which is the image number that corresponded to the most confident prediction. Followed by 2,3 and 4**.**