Report: act_report

- Create a **250-word-minimum written report** called "act_report.pdf" or "act_report.html" that communicates the insights and displays the visualization(s) produced from your wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

Insights

The names of the dogs were programmatically extracted from the sentences in which the dog's name is suppose to follow after this format ('This is'), but this extracted a lot of words which are not dog's name. This none dog name was possible because some text follow a different format which doesn't conform to the line of code use to extract the dog names. This was resolved by extracting only the names that correspond to a dog name using the pandas.series.str.extract method after which the wrong names were dropped. The extraction for the stages of the dogs that fell under 'doggo','floofer', 'puppo', 'pupper' columns contain 'None' value where NaN should be used. The None values were changed while there are dogs have that can be classified under more than one breed but that were not properly separated was not well defined and separated, this was corrected by assigning each as a separated name while creating the stage column. The image prediction data has missing data which cannot be completed because of lack of access to the original data and hence these rows that lack the missing columns were dropped.

Visualization

Visualization of the Data obtained helped is an essential part of explorative analysis which helps understand the underlying structure within the Dataset or explore the relationships between variables. To get a visual understanding between two of the variable(retweet_count and favorite_count) in the Dataset, a seaborn regplot function was used.

seaborn.regplot()

Below is a graph of retweet_count against favorite_count. This graph showed a positive correlation between booth variable.