

Report: act_report

Insights

The image_file data rows with missing value were dropped due the information cannot be obtained from the main data cannot, hence supporting the decision to drop such columns. The names of the dogs have many wrong names because it was harvested programmatically to recognize the first words under the 'text' column as the name of the dog. This gave a lot of wrong name as most tweets under the 'text' column did not begin with the name of a dog. From the analysis, it was discovered that the dogs fall into more than one stage at a time; this was separated so most dog could be classified under a particular stage. After separation into the various stages, it was discovered that most dogs fall under the doggo stage.



Visualization

Visualization is an exploratory part of analyzing the dataset. it helps to explain various variable diagrammatically, the relation between three variables namely P1_conf, P2_conf and P3_conf were explored using plot. Scatter method. This plot is ideal for this visualization because it helps compares numerical variables. The outcome of the visualization showed a strong, non-linear relationship between the various stages of the dog.

Plot. Scatter

Below is the graph showing the level of interaction between P1_conf, P2_conf and P3_conf.

