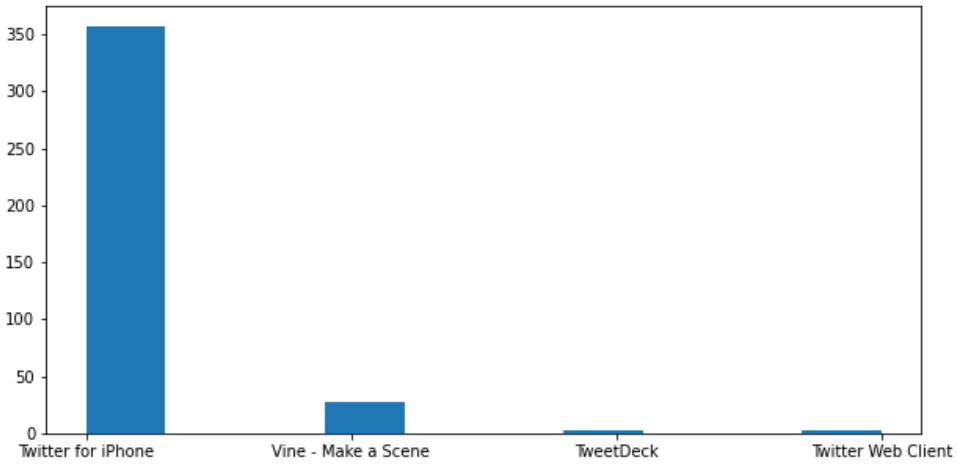
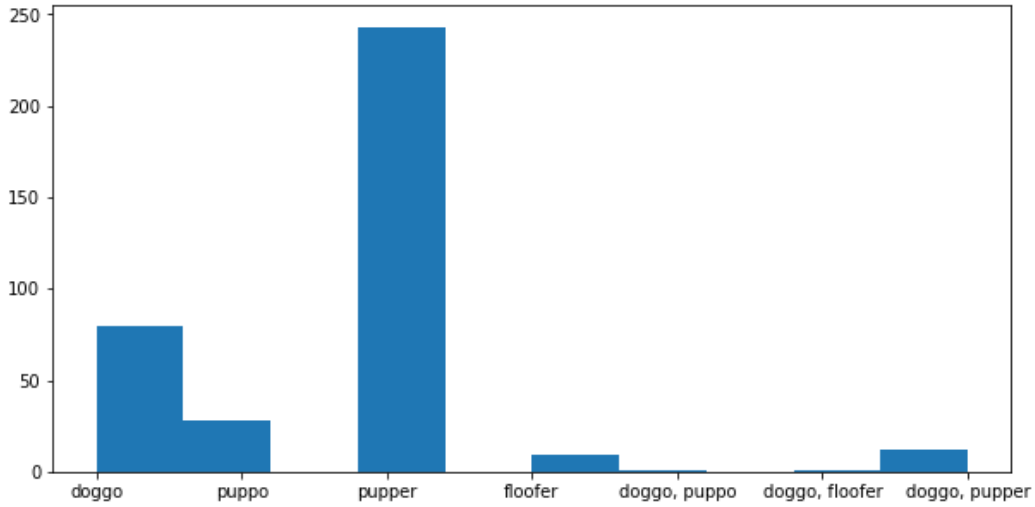
Report for Insights and Visualizations part of twitter project

In twitter archive table:

- Plotted a histogram of the source column in twitter archive table to find the most used source for tweeting. As shown the most frequent source is Twitter for iPhone while the least one is Twitter web client. An interesting fact is that twitter android application isn't included in the sources.

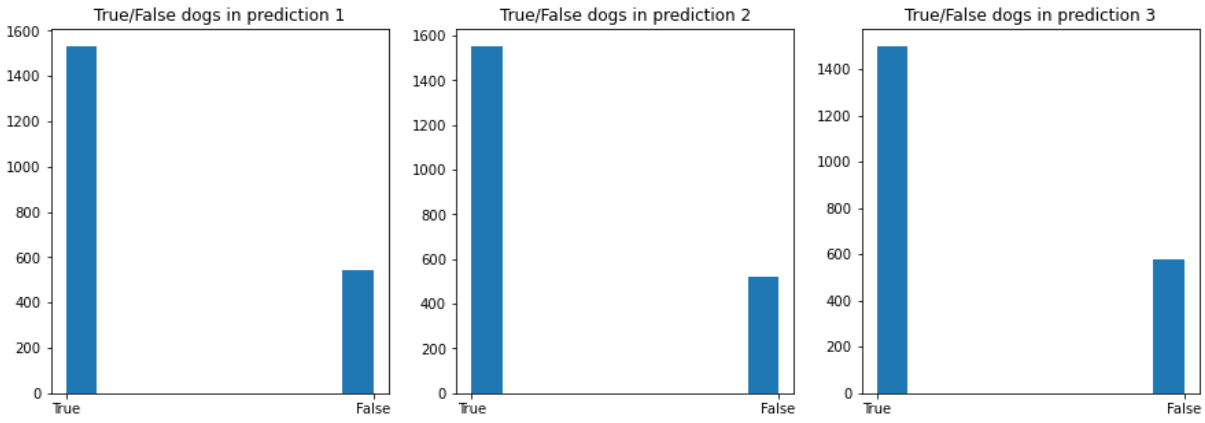


- Plotted a histogram of the dog stages. As shown by the histogram the most frequent dog stage is the pupper stage followed by the doggo stage. According to the numbers on the y-axis compared to the number of tweets in the original dataset, only a small subset of the dogs have their stage recorded correctly.



In image predictions table:

- Plotted a histogram of p1_dog, p2_dog, p3_dog variables to compare the number of predictions that were actually dogs with the ones that were incorrectly classified as another type of animal. As shown by the histogram for each prediction variable there were 500-600 dogs that were falsely identified as another type of animal. This gives us an approximate ratio of 3:1, therefore, for each 3 predictions that are correctly classified as dogs there is a dog that is miscalssified as another animal.



- Plotted a density distribution graph of each prediction variable confidence (p1_conf, p2_conf, p3_conf) using seaborn distplot function to get an overview of the density distribution of each prediction variable confidence. From the distributions in the graph we notice that the first prediction (p1) often have higher confidence values and that the confidence values for the third prediction (p3) are very low.