First the data wrangling efforts including the steps of Gathering, Assessing and Cleaning were conducted on data. Then once a cleaner master dataframe was achieved it was written to a master archive ‘csv’ file. The cleaned final dataframe was used to answer three questions with the help of data analysis and data visualization.

Questions answered were as follows:

1)Out of the records where the stage of the dog is recognized which stage of the dog is most common?
2)Out of the predicted dog breeds which is the most common breed ?
3)Out of the records where dog breed gets predicted- which breed gets retweeted and is the most favorite?

Question 1: the modified column dog_cl was analyzed and visualized using seaborn library. On visualizing using horizontal bar plot it is evident that “pupper” is the most common state among all the records in which it was possible to identify the state.

Question 2: the modified column dog_breed was analyzed and visualized again using seaborn library. On visualizing using horizontal bar plot it is evident that “golden retriever” and “Labrador retriever” are the most common state among all the records in which it was possible to identify the breed.

Question 3: the retweet count and favourite count are found out using seaborn library’s count plot. The retweet count and favourite count parameters are counted using groupby method using dog breed. Once the parameters are grouped, they are sorted properly. Then finally the sorted retweet count and favourite count data is plotted for each dog breed is visualized using seaborn’s bar plot. This again shows “golden retriever” breed has the highest retweet count and favourite count.