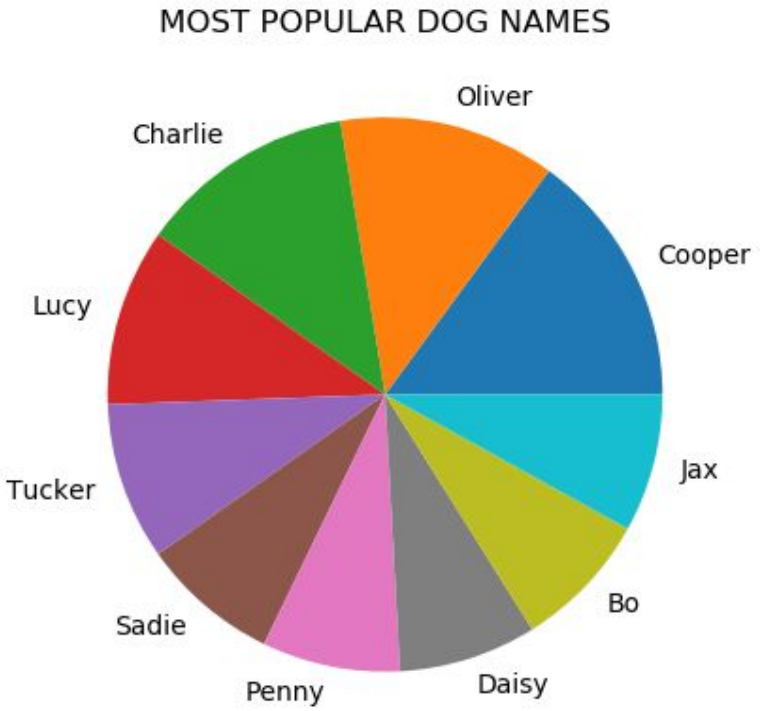
DATA ANALYSIS AND VISUALIZATION

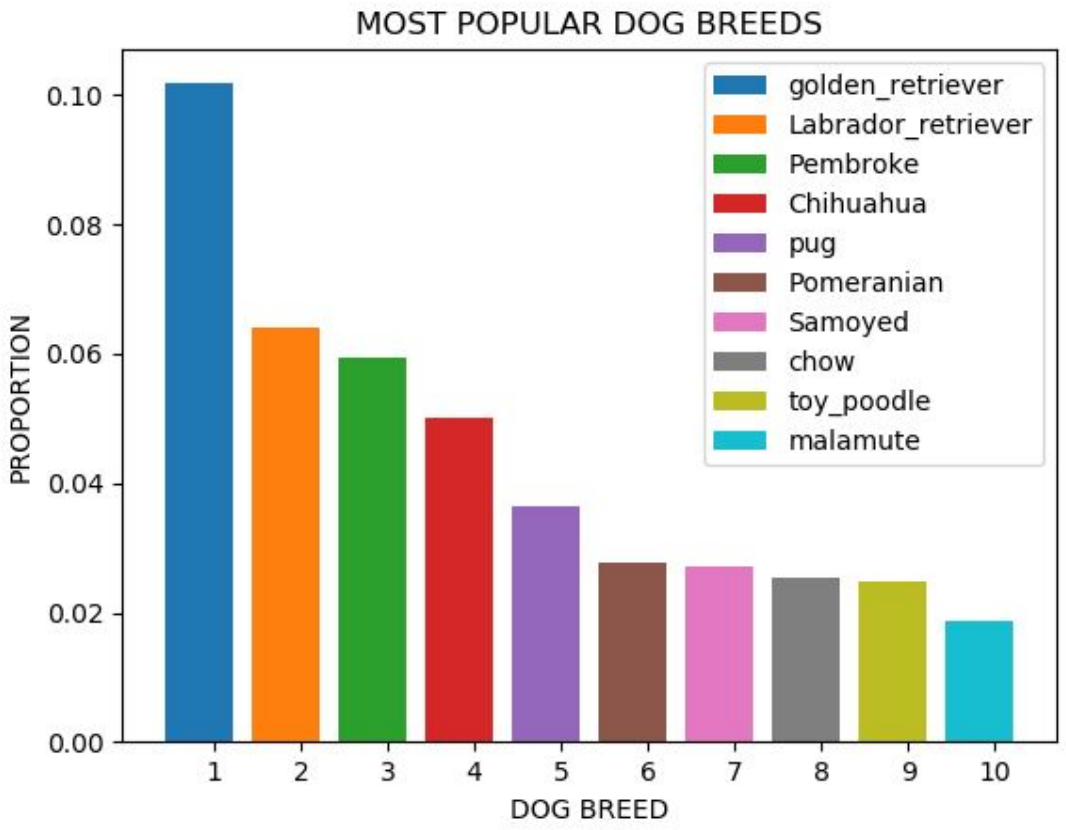
First, we tried to find out what are the most popular dog names in our archive. Using value_counts() function on our name column, we got the count of each name. Then, using slicing, we picked up the names with the most counts. In our results, the word ‘the’ also crept up. This may be due to the fact that many of the tweets didn’t contain any specific names. We removed it using drop(). Then using matplotlib.pyplot’s pie function, we created a pie chart displaying the top dog names and their share. We used the index of the series obtained from value_counts() to mark dog names and the values to create pie slices.



Cooper turned out to be the most popular dog name.

We then tried to find out the most popular breeds in the WeRateDogs Twitter archive. The value_counts() function on the predicted_breed column of our DataFrame gave us counts of occurrences of different dog breeds in our archive. We then divided the counts of each breed by the total number of records to obtain the proportion of each dog breed. Then we sliced the value_counts() series of proportions to get the top ten breeds. Using bar() function of

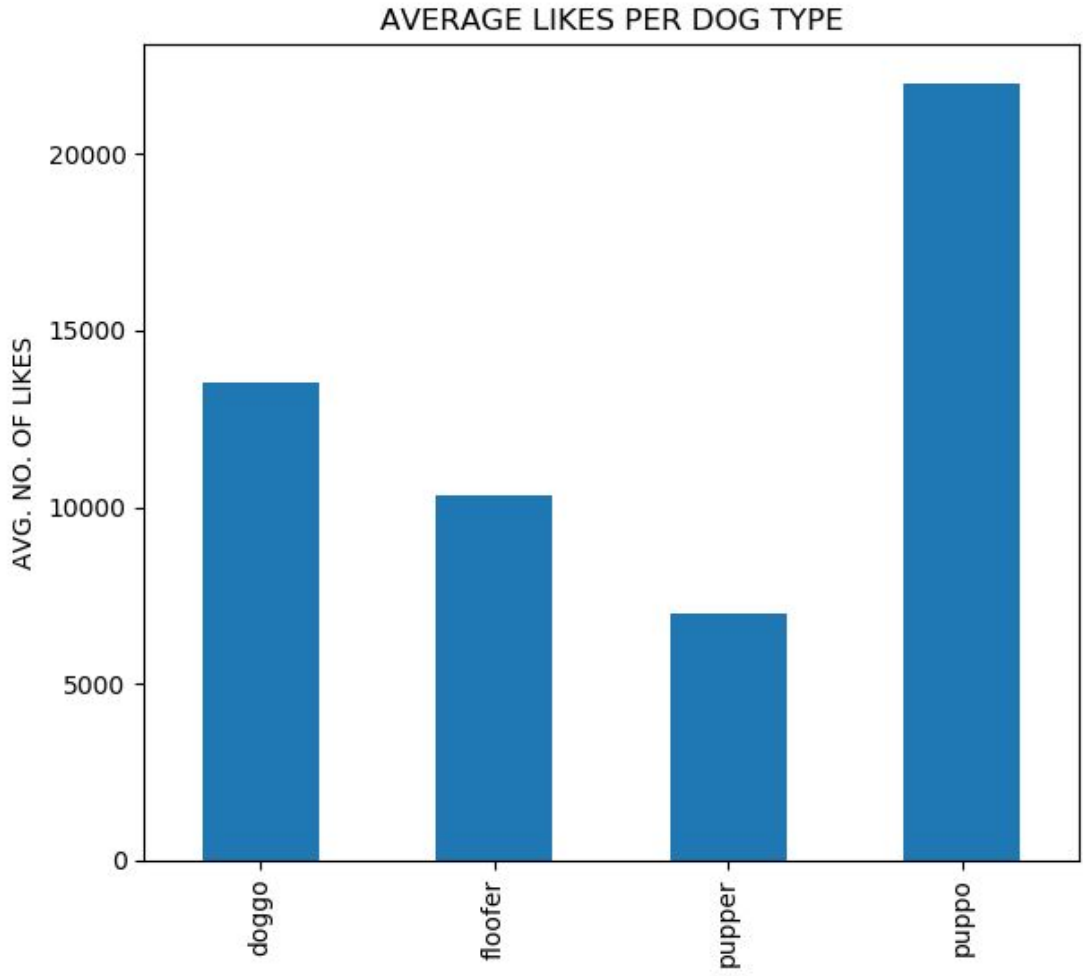
pyplot, we created a bar chart representing our results.



Golden Retriever turned out to be the most popular breed.

We then tried to find the relation between the dog type and the average number of likes(favorite count) it received. We used groupby() to aggregate our data according to the four dog types (doggo, floofer, pupper and puppo) and calculated the total sum of favorite count for each type. Then, to get the average, we divided these sums by the number of occurrences for each respective dog type. We then created a bar graph using Pandas' plot()

function to observe the average likes per dog type.



Thus, we can see that out of the four dog types, puppo received the maximum likes while pupper received the least.