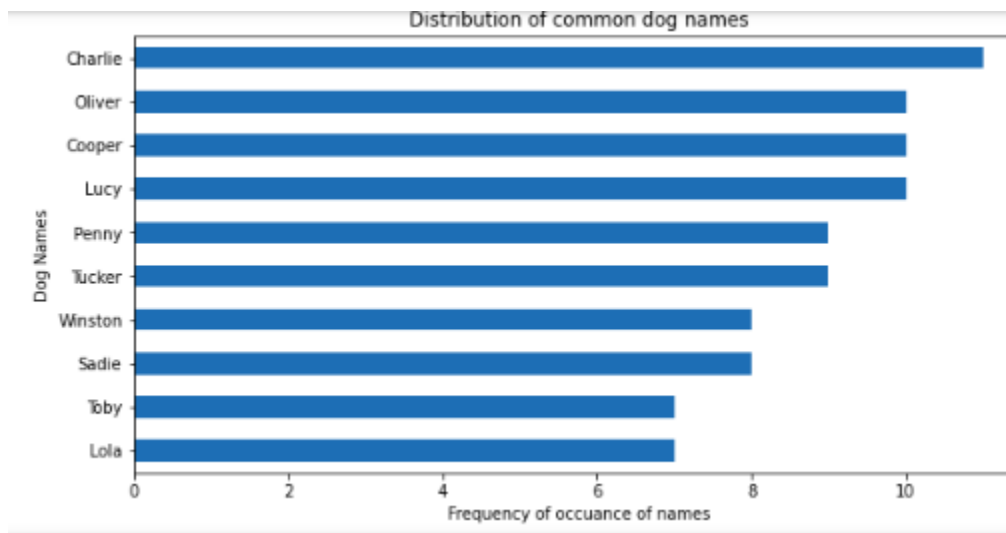# Data Analysis Report

Introduction

This report describes the process of data analysis and visualizing part of the master merged dataset from the earlier steps of this project. Once the data analysis is finished, we will move on to answering the questions that will be raised as the result of the analysis part.
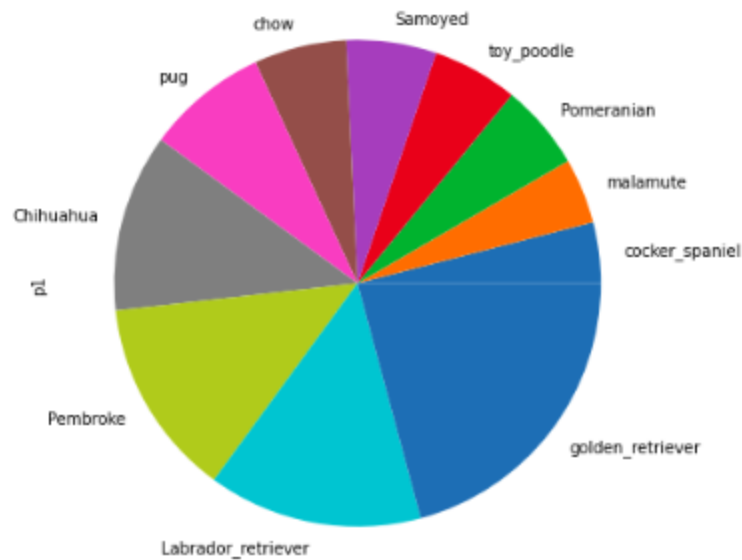
Observation 1: Which dog name is the most common one?

In order to find out the most commonly used dog name, we will first count the number of common names from the dataset by using the value_counts() method. Next, it is plotted by using the matplotlib barplot. This kind of plot is used to visualize the distribution of a variable. From this plot, it is concluded that the name "Charlie" is the most common dog name in this dataset

Observation 2: Which Breed is Common?

p1 variable is used to find out the most common breed of dog. Pie chart is used to visualize the distribution of the most common dog breed in the dataset. Since we have multiple variables of breeds, it is good to use different colors to portray the breeds of dogs from the twitter data.Based on the analysis and visualization, golden retriever is the most common breed of dog.

Observation 3: What is the common rating number that is given to a tweet?

To answer this question, the rating_numerator and rating_denominator variables are used. Since the rating denominator is the same for all tweets, we will use the rating_numerator variable to find out the common rate that is given to tweets. This analysis shows most tweets are given ratings of 12/10.

Remark: Denominator is 10