The following act report summarizes the steps in the data analysis process that was performed during the project.

For this project I worked with 3 datasets with data that pertains to the WeRateDogs twitter account

The first dataset that Udacity provided was a .csv file that contains the basic details about WeRateDogs users' tweets. It was downloaded manually.

The second dataset was an image prediction file, which was hosted on Udacity's server. It contained over 2,000 predictions made by an unsupervised neural network.

The third dataset contains data that was scraped using the Tweepy Twitter API. An attempt was made to obtain this information programmatically but due to authorization issues with the Twitter API I was provided a text file by an instructor which I loaded into a pandas data frame.

After loading all 3 data I discovered several quality and tidiness issues which was addressed both visually and programmatically to assess and clean. The three datasets were then merged into one and saved as a master file.

Some insights I was able to gain from the data was that the most popular source of tweets were:

Most tweets were made from the Twitter for iPhone app
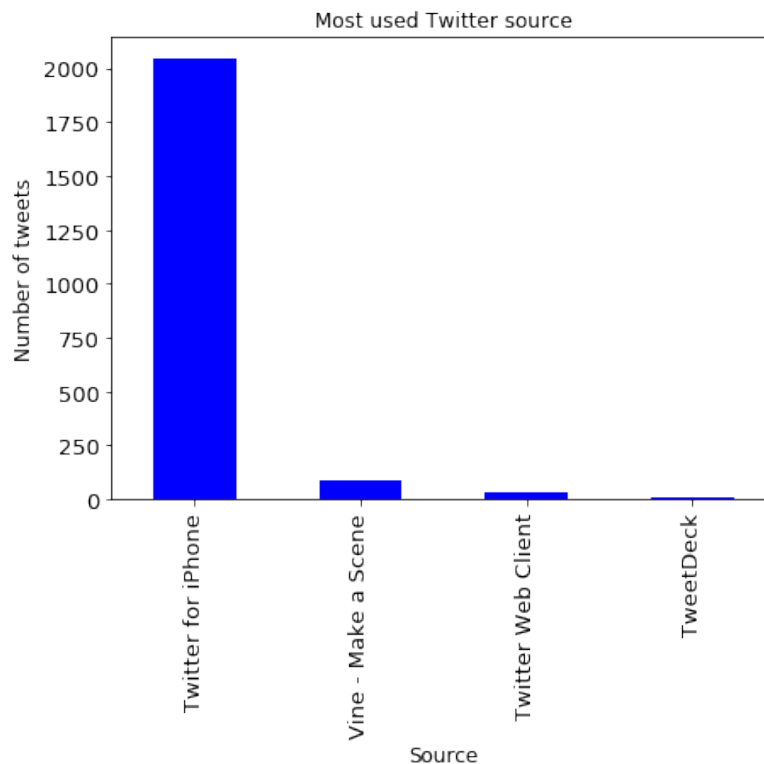
The most popular dog name was Charlie with 11 occurrences followed by Oliver and Cooper with 10 occurrences each.
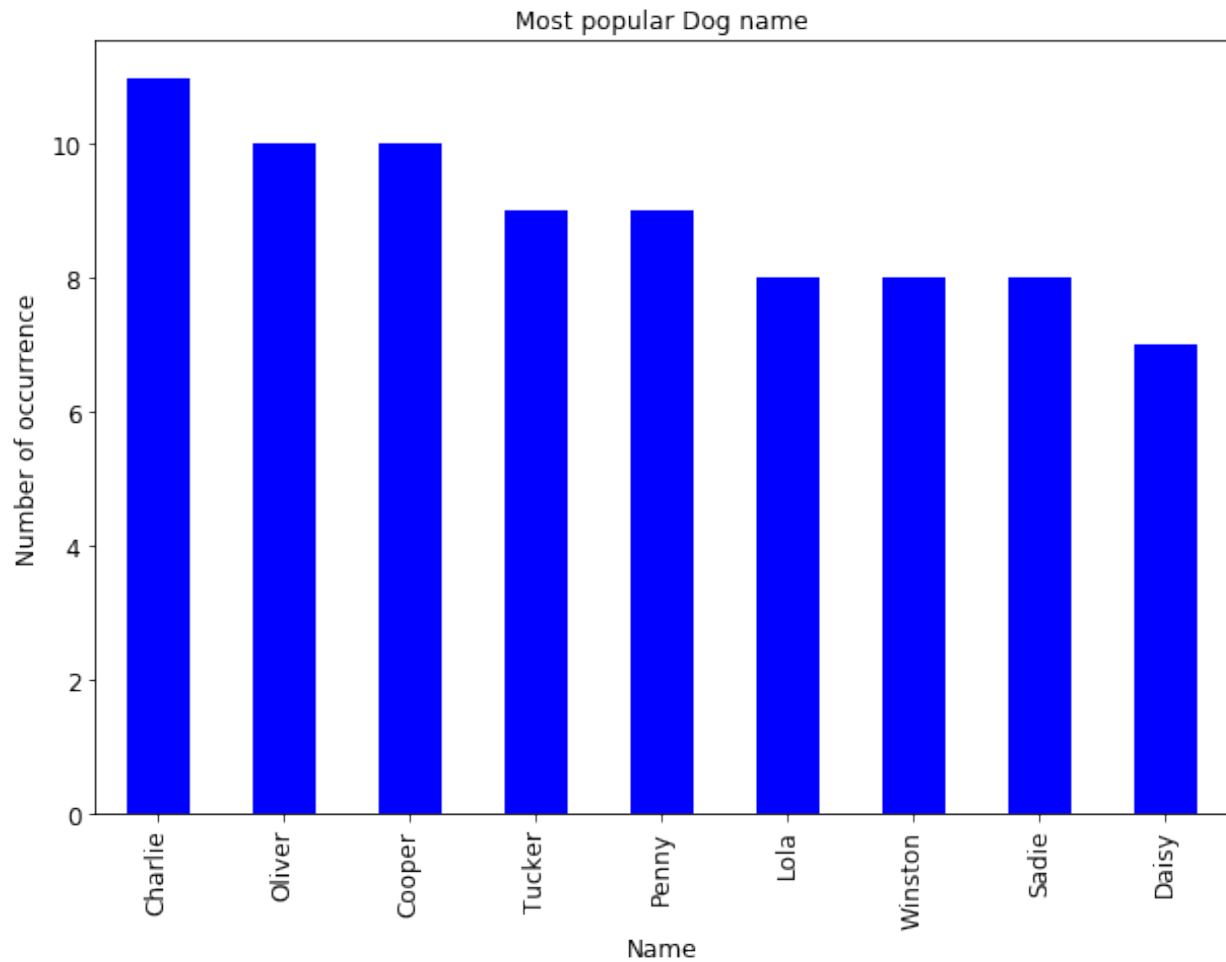
The most commonly predicted dog breeds were Golden Retrievers, Labrador Retrievers and Chihuahuas.