Udacity Nanodegree - Project 4
Wrangle and Analyze Data | WeRateDogs Twitter

Context & Introduction

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about dogs. In this project, what we are going to do is wrangle and analyze data from this tweet archive of Twitter user @dog_rates.

For additional context, here's a sample Tweet from one of our data points.
https://twitter.com/twitter/status/892420643555336193

In terms of data points, we were working through datasets from 3 different sources, (1) .csv file (2) downloaded .tsv file online (3) Twitter API - going through the data wrangling process of gathering, assessing, and cleaning of data.

Please feel free to reference my Colab Notebook titled "wrangle_act.ipynb" linked out here
https://colab.research.google.com/drive/1jEHMNRY00IJp70aaM-CCAhgBgXi5L-Xt?usp=sharing
for reference :)

Insights & Visualization

After going through the wrangling process (well documented in the above Colab Notebook), here are the questions that we seeked to answer and insights that i've generated from the data:

Question 1: What is the most popular dog breed?

The most popular dog breed is the golden retriever, with a labrador retriever coming in as the second most popular breed.
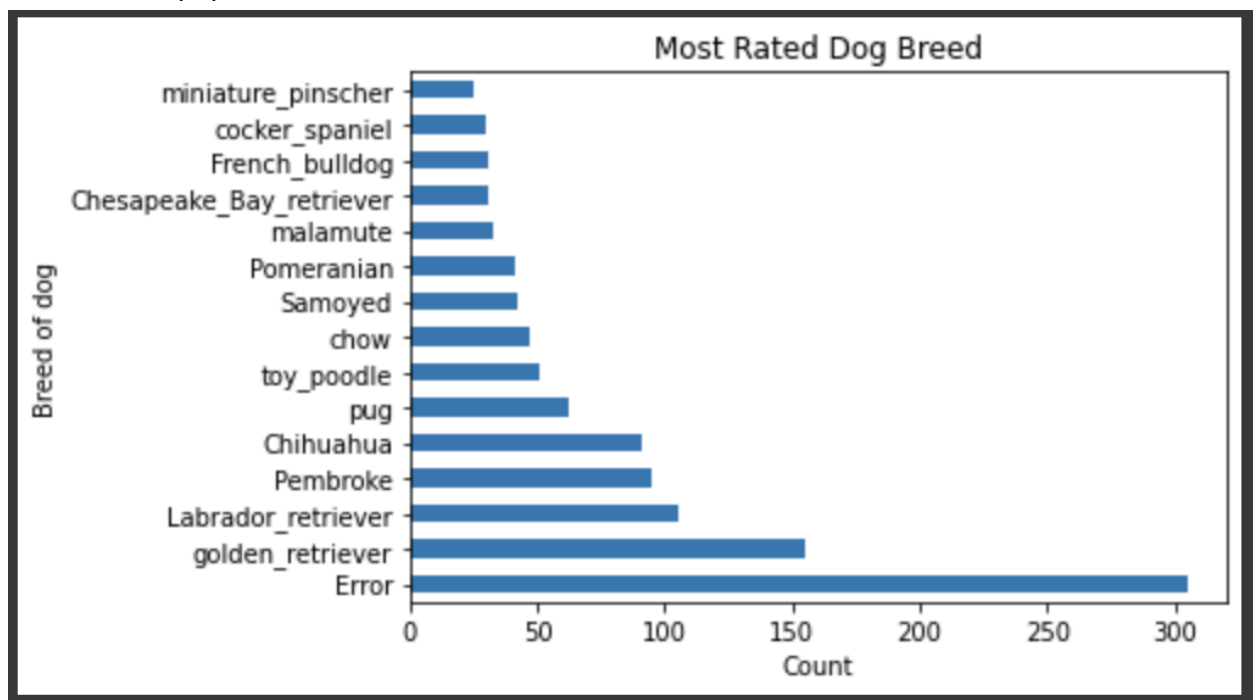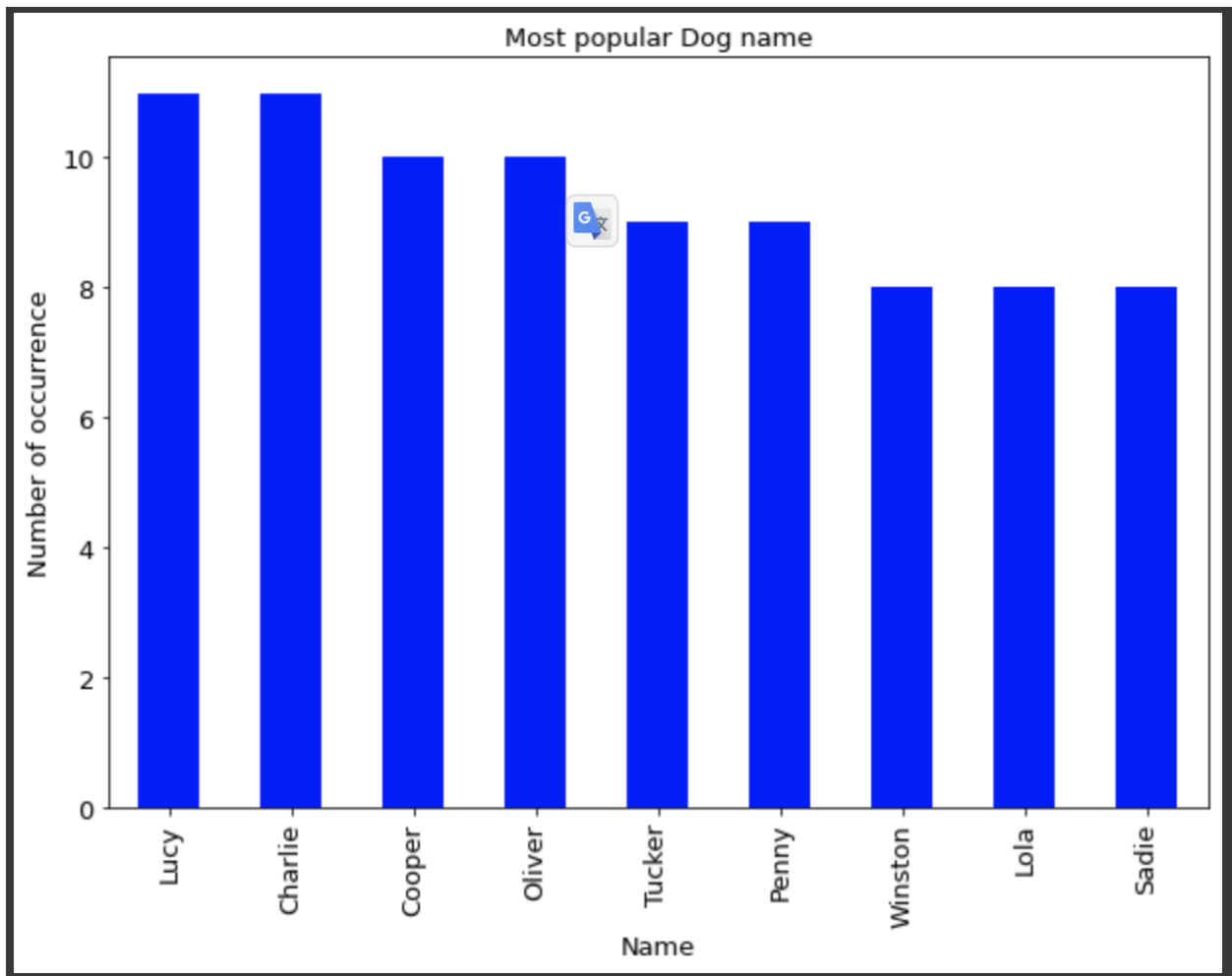

Question 2: What is the most popular dog name?

The most popular dog name is Lucy, followed by Charlie.

Most popular Dog name

Question 3: Which is the common stage?

Most dogs are classified within the 'Pupper' stage.
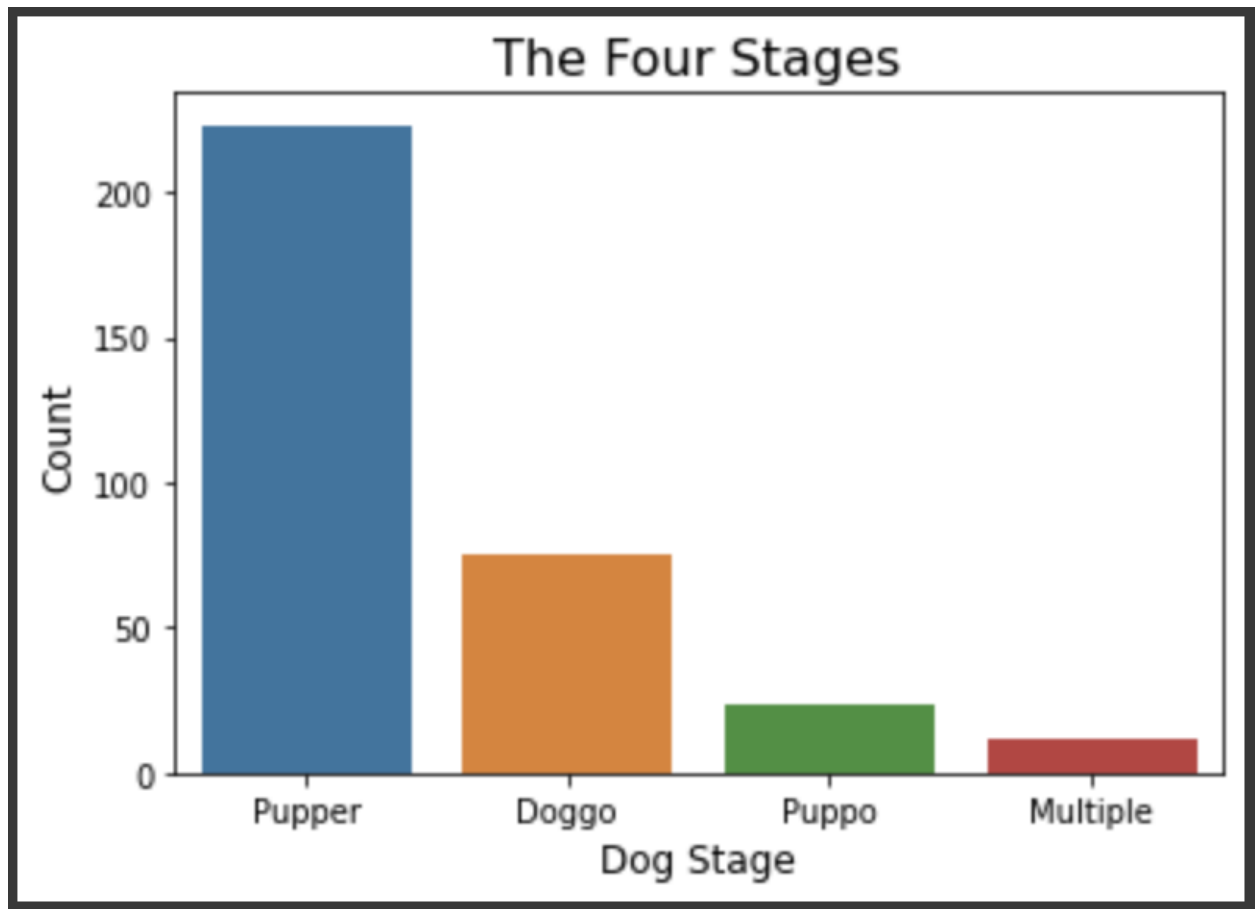
Question 4: What is the rating distribution?

After removing anomalous data the rating distribution is as such with 12 being the most number of Tweets followed by 9.

Udacity Nanodegree - Project 4

Wrangle and Analyze Data | WeRateDogs Twitter

Links to colab notebook:
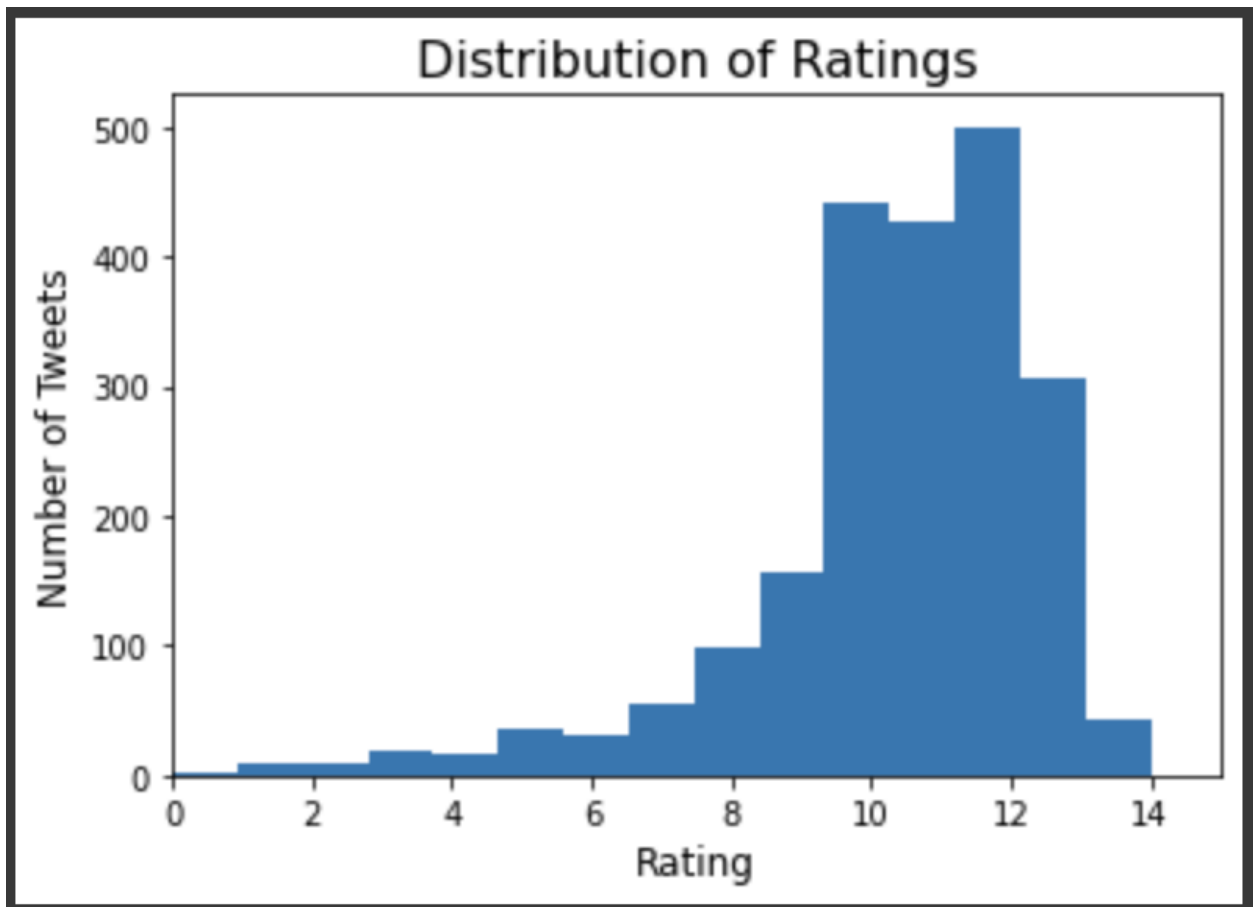
https://colab.research.google.com/drive/1jEHMNRY00lJp70aaM-CCAhgBgXi5L-Xt?usp=sharing