Project: Wrangle and analyze data from WeRateDogs Twitter account – Insights report

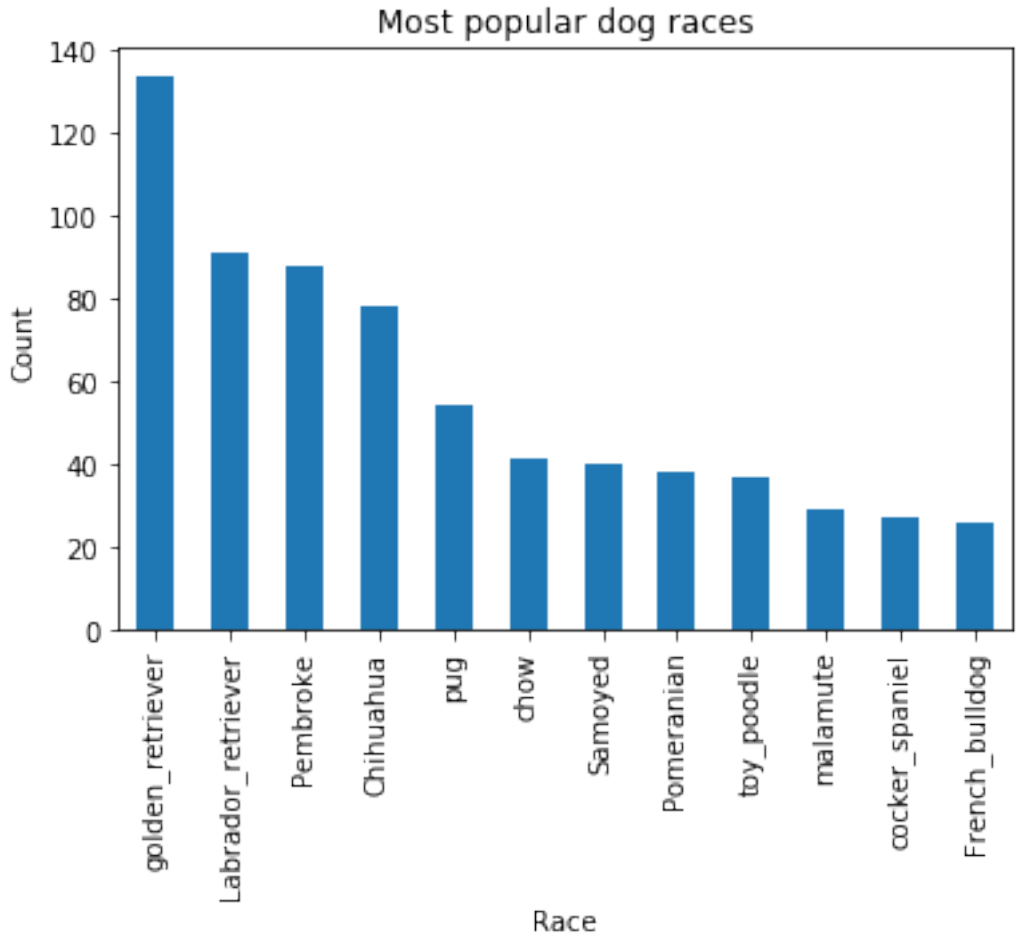
The purpose of this project is to wrangle data from WeRateDogs Twitter account to create interesting and trustworthy analyses and visualizations.

The analysis has been conducted on three data sources

1. "Twitter_archive_enhanced.csv" - The WeRateDogs Twitter archive.
2. The tweet image predictions (image_predictions.tsv), i.e. what breed of dog (or another object, animal, etc.) is present in each tweet according to a neural network.
3. Using the tweet IDs in the WeRateDogs Twitter archive, query the Twitter API for each tweet's JSON data using Python's Tweepy library and store each tweet's entire set of JSON data in a file called tweet_json.txt file.

The three data sets have been assessed visually and programmatically to find issues related with data quality and tidiness. The cleaned dataset has been merged into a master file.

From the analysis of the final dataset, emerged that the most common dog's races are:



Race	Count
golden_retriever	134
Labrador_retriever	91
Pembroke	88
Chihuahua	78
pug	54
chow	41
Samoyed	40
Pomeranian	38
toy_poodle	37
malamute	29
cocker_spaniel	27
French_bulldog	26

There are four possible dog's maturity stages:

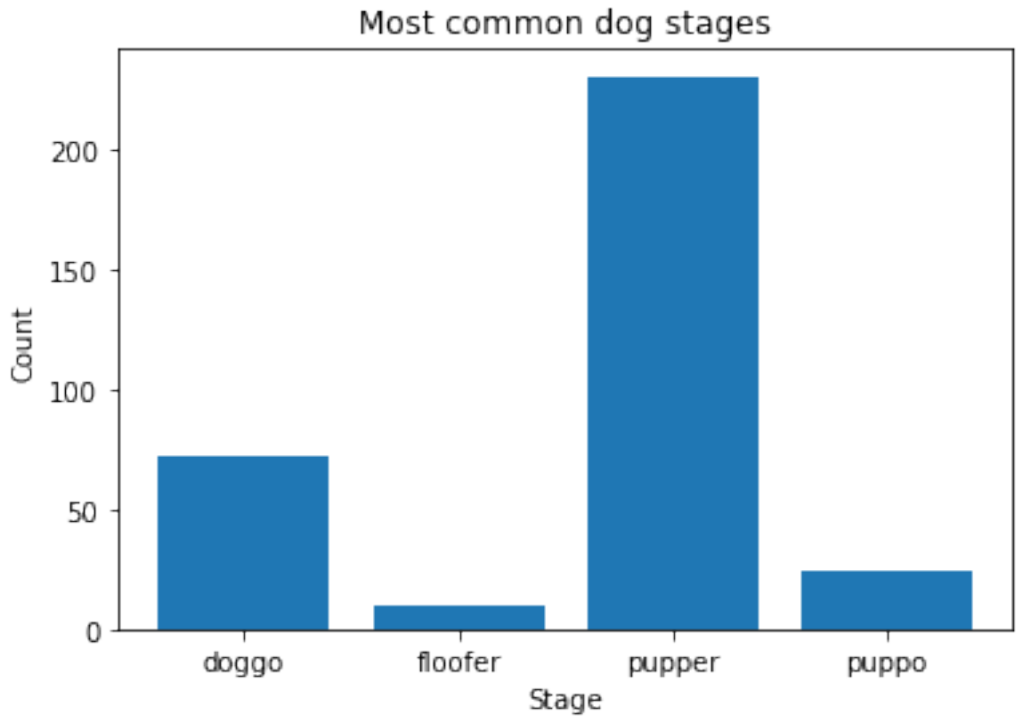
Doggo: a big and usually older pupper

Floofer: a fluffy dog

Pupper: another word for pup

Puppo: a transitional phase between pupper and doggo, the dog equivalent of a teenager

From the analysis of the dataset, emerged that tweets with puppers are the most common:



Stage	Count
None	1744
pupper	230
doggo	72
puppo	24
floofer	10

However, puppo dogs seems to obtain higher votes on average

Stage	Mean
None	11.802752
doggo	11.847222
floofer	11.800000
pupper	10.817391
puppo	12.041667