

## Wrangle and Analyze WeRateDogs



The dataset that you will be wrangling (and analyzing and visualizing) is the tweet archive of

Twitter user @dog\_rates, also known as WeRateDogs.

WeRateDogs is a Twitter account that rates

people's dogs with a humorous comment about the dog

The WeRateDogs Twitter archive contains basic tweet data for +5000 all of their tweets, but not everything.

One column the archive does contain though:

each tweet's text, which I used to extract rating, dog name, and dog "stage" (i.e. doggo, floofer, pupper, and puppo) to make this "enhanced Twitter archive"

Download dog prediction images programmatically using the Requests Library

Additional Data via the Twitter API

Use Tweepy library to query the Twitter API for each tweet's

Visual Assessment

JSON file

The JSON file has 2337 rows and I have chosen 4 columns of interest to include in the DataFrame.

Image Predictions File

This file has 2075 rows and 12 columns

The Twitter Archive File

This file has 2356 rows and 17 columns

Programmatic Assessment

JSON File

Twitter Archive File

Image Prediction file

Clean data :-

Tidiness Issues

- Tidiness Issue #1

Use `.merge()` to create one large dataframe from the three available dataframe

- Tidiness Issue #2

To create Tidy Data, the four columns to indicate the stage of life for a dog should

be moved to only one column, named 'dog\_stages'

Quality Issues

- Quality Issue #1

Delete extraneous columns

- Quality Issue #2

Change data types for 'tweet\_id' to string

- Quality Issue #3

Change 'timestamp' to datetime

- Quality Issue #4

Change 'rating\_numerator' to float

- Quality Issue #5

Change data types to category

- Quality Issue #6

All dog names are actually names

- Quality Issue #7

Remove Retweets

- Quality Issue #8

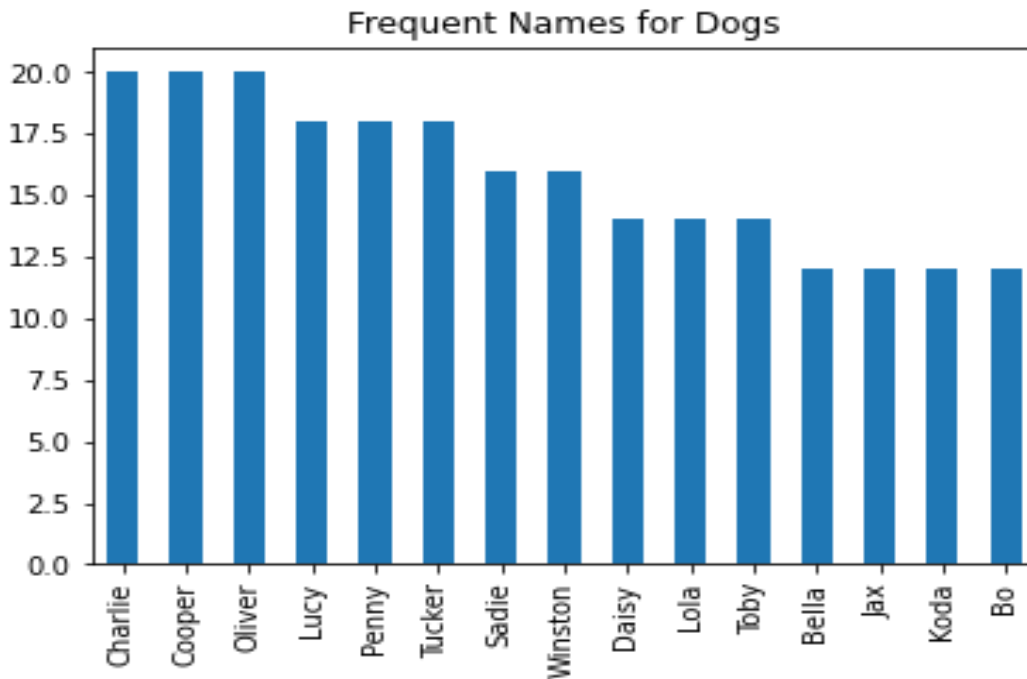
Remove Tweets without Images

- Quality Issue #9

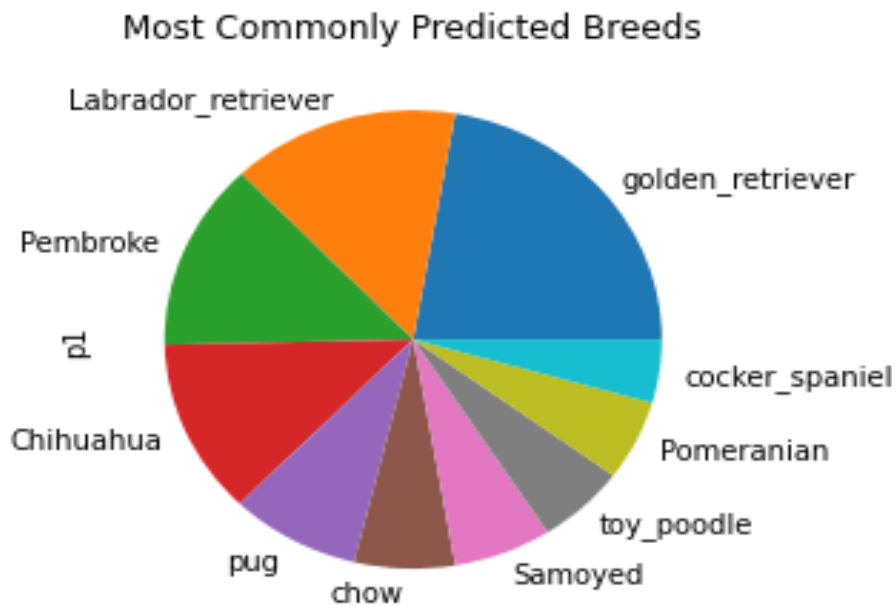
Incorrect Numerators 10 Storing Cleaned Dataframe

# Visualizations

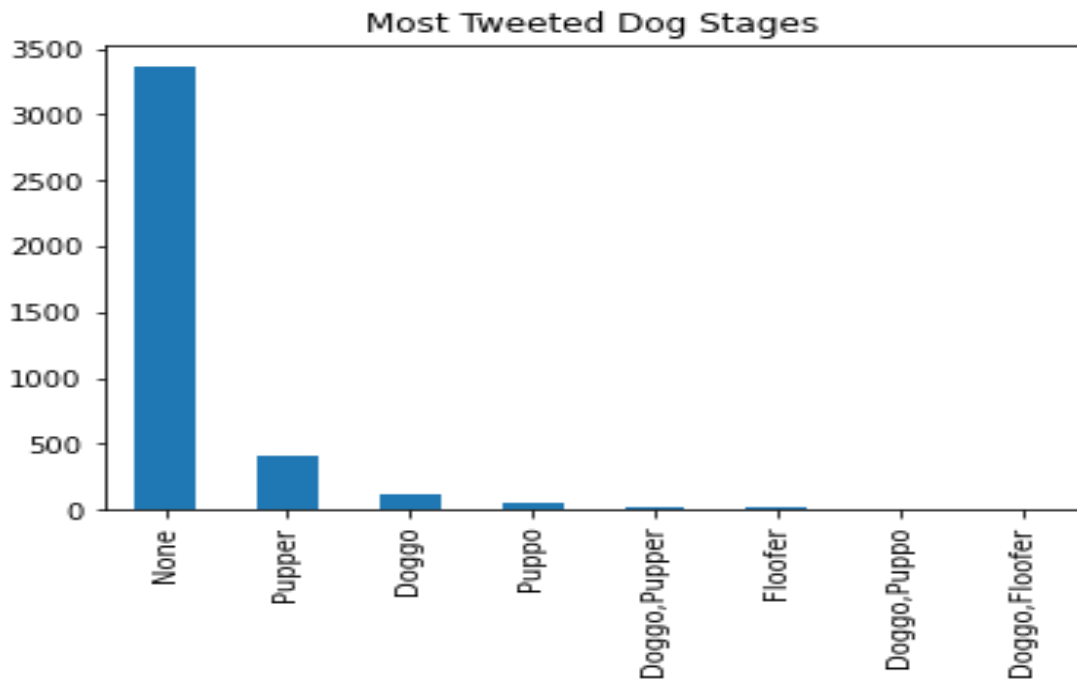
- Distribution of dog names



- Most commonly predicted breeds



- Distribution of Dog Stages



- Retweet Count versus Favorite Count

