Introduction

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, 13.5/10 etc.

We were to wrangle and analyze the data from their twitter account with details such as (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017.

Data Wrangling

Gather
We gathered data from the WeRateDogs Twitter archive : twitter_archive_enhanced.csv - this was downloaded manually.

We then gathered data with image predictions for their tweets according to a neural network - this was hosted on Udacity's servers. So we proceeded to download the file programmatically using the requests python library

We then gathered data on their retweet and favourite counts from Twitter API using python tweepy access library. We placed the JSON data on `tweet_json.txt`

Access

We proceeded to assess the data visually and programmatically for quality and tidiness issues.

Clean
We first made copies of the data, then we proceed to clean assessed data by defining, coding and testing to see if the data had been cleaned. Starting first with the tidiness issues.

Storing

After cleaning the data we proceed to make a clean copy of data named : `twitter_archive_master.csv`