Udacity Data Analyst Nanodegree - Project IV: Wrangle & Analyze Data - Wrangle Report

For this project, data needed to be wrangled from three different sources:

1. twitter_archive_enhanced.csv - This file was provided from the twitter account WeRateDogs that had information regarding ratings for dogs, and "stage" of dog, referred to as "doggo", "floofer", "pupper", and "puppo", as well as specific tweet ID's and other information from a range of tweets.

2. Image_predictions.tsv - This file was hosted on Udacity's server and was accessed and downloaded programmatically before being read into pandas as a CSV file in order to read it into a dataframe. This file contained information from a neural network that analyzed photos posted on the WeRateDogs Twitter account in an attempt to identify the breed of a dog based on the photo in a tweet. The information in the .tsv file contained confidence levels, whether the image was of a dog or not, and what the breed of dog the model predicted was in the image.

3. A third file was downloaded via the Twitter API programmatically using the API documentation in order to download information from the WeRateDogs Twitter account.

   After creating a developer account for Twitter and obtaining access to auth keys that are required in order to use the API, we were able to use pandas to programmatically access information on the Twitter servers regarding the WeRateDogs Twitter account. This information was downloaded into memory into a JSON file that was then dumped into a .txt file and read into pandas as a dataframe.

   The information provided via the Twitter API gave us access to a range of information for the WeRateDogs Twitter account, though it could be used for accessing any account on the platform. Some of the information that was selected to be obtained were favorite & retweet counts for each tweet ID, and a 'created at' timestamp.