Wrangle Report

My wrangling efforts included gathering three datasets: Archive, Ratings and Images, each requiring their own means of wrangling. To gather the Archive data, was retrieved from the csv provided by Udacity called "twitter_enhanced_archive" which is the WeRateDogs archive of tweets, and it includes Tweet ID, Tweet Source and Dog stages. This JSON data from this file was then read and saved in a .txt file and later on read into a dataframe entitled "archive" for further cleaning and analysis. The wrangling process for the Ratings data included creating a developer account on Twitter to be able to access the Twitter API. I used the Tweepy library to access the API, thus allowing me to gather data from the WeRateDogs archive. This data includes, most importantly, each tweet's retweet and favorite count. And finally, to gather the Images data, I had to programmatically download a file using the Requests library. I then had to open and write the data into a file which was entitled "image-predictions.tsv". To load the data, I read the .tsv file into a dataframe called "images" for cleaning and analysis. The data includes info on the image predictions of one of the photos in each tweet, meaning it was passed through a Convolution Neural Network to predict what the picture is. There were three predictions, each with their own confidence values and whether or not the image was a dog or not.