The Data Wrangling involved Gathering, Assessing, Cleaning and Saving the cleaned data.

Gathering Phase: The project made use of 3 sources for data. I was provided with one which I only had to upload and read into my project workspace. The second was downloaded programmatically from Udacity server using the request library. The third was gathered using the tweet_ids with the aid of twitter API which was generated with the Tweepy library after successfully setting a Twitter Developer Account.

Assessing Phase: I employed both Visual and Programmatic Assessment to detect quality and tidiness issues with the dataset. Visual Assessment was using Microsoft Excel and Programmatic Assessment was done in my project workspace (jupyter notebook). I was able to detect 8 quality issues and 3 tidiness issues.

Cleaning Phase: The Cleaning Phase proved to be the most tedious aspect of the project. After making a copy of each table, I proceeded to cleaning the first table.
The first cleaning I did was to make remove the tweets which were retweets. This was one of the guidelines I was given for the project. This was done by using a line of code that filters the dataset of retweets.
In the dataset I was given, the names of dogs were not extracted properly. So, I used regular expressions (re library) to extract the names to a new column which I named dog_names.
I also used a regex to extract the numerators of ratings which were decimal numbers. This was because rating_numerators for decimal numbers were incorrect. From my assessment, I also discovered that for denominators which were greater than 10, there contained 2 or more dogs. What I did was to create a new column: rating, that normalizes all the rating_numerators.
Then I melted the doggo, floofer, pupper & puppo columns into a single column and removed unnecessary columns from the table. This was for the first table
In the second table, I removed duplicates and I removed columns I felt weren't germane in the third table.

After I addressed all the issues I detected from the Assessing phase, I merged the datasets using the tweet_ids and then saved it to a csv file named twitter_archive_master.csv as was instructed.