The wrangle effort was split up into multiple parts to effectively gather, assess, clean, and analyze the WeRateDogs Twitter data analysis.

Along with the data available from the WeRateDogs twitter archives, an effort was made to also include data from Udacity neural network project which involved determining the breed of a dog from an image. Gathering data process involved getting data from three different sources. An original enhanced data file was provided in the form a CSV file and consisted of the WeRateDogs data filtered to only include data that have images associated with them. For the retrieval of important 'retweet' and 'favorite' frequency, the data was retrieved using a twitter API called Tweepy. Using the tweet id as the reference key, the retweet count and favorite count per tweet id was retrieved using the twitter API. Before the data could be retrieved using the API, a development account was set up with twitter and using their provided keys to directly access their archives to retrieve all the information. The process of retrieving data took a long time due to restrictions in twitter itself, so the gathering function includes exceptions to wait for retrieval if a timeout stage is reached. The dog breed information was made available using a url and was easily accessed using the the read_csv pandas function. Older versions of python would probably require using the request function, as that was the suggestion made for this course

Main focus of the assessment was to look at the enhanced twitter data. Various quality and tidiness issues were uncovered using pandas functions for viewing, filtering. It was evident that the consists of redundant variables that are not be helpful for finding any insight with provided dog ratings. It could also be seen that a lot of the postings were not of ratings but posts through WeRateDogs that happened to captured in the dataset

It was possible to find patterns for what would constitute as a post vs what is an actual rating and other pandas functions were used to effectively siphon out redundant posts and variables and leave only dog rating posts. Other cleaning efforts were utilized to ensure 'completeness', 'validity', 'accuracy', and 'consistency'. The data was reassessed many times as other tidiness and quality issues were discovered.