Data Wrangle:

In this project we are going to learn how to gather the data, assess the data clean the data and then analyze the data that are learned in the course of Data Wrangling which is one of the important aspects of Data Analysis.

Data Gathering: Here in this project we have gathered data from 3 sources which are twitter_archive, image_predictions and tweet_df which was pulled from twitter using twitter_api. The data from twitter api is in JSON format then converted into pandas dataframe using pandas function.

Data Assessing: I have assessed the data to check for various columns in the data sets, their data types and list of all possible values that are there in the data set. I observed that there are some data records that have wrong values in the columns like an example would be in one column I observed there is letter 'O' instead of number 0 etc. This made me to clean the data before I do the analysis of it.

Data Cleaning: Once I observed there are some anomalies in the data during the data assessment phase, I did the data cleaning using pandas. Then we look for issues in the data like tidiness issues and quality issues.11 quality issues and 1 tidiness issues like retweet which are not needed ,four variables needed to be merged. Some of the cleaning I have done are replacing the values in the columns with correct values and dropping unnecessary columns that are not required for analysis. The datatype of rating_numerator is Int so there was issue with numerator value its only taking the decimal part not whole rating in numerator.To resolve the issue ran the code snippet given by Reviewer then assigned the ratings numerator with the ratings_numerator in twitter_archieve_copy.
Updated the dogstage column with multiple which use to have a dog name instead of Nan.Used replace function to replace the dog name with multiple.

Analyzing : I analyzed the data to see the distribution of tweets by platform (web, iphone, android). I also analyzed the distribution of tweets among various breeds of dogs. I also analyzed the distribution of dog rating across various months of 2017. I found out that dogs had highest rating in August of 2017. I also analyzed retweet count and favorite count with timestamp.