I will be using two sets of data about the different Pokemon you can catch in Pokemon Go. One data set is organized and relatively clean while the other needs to be organized. To make this process easier I will refer to the datasets by the name of the web site I go the off of, so the organized data set will be called Kaggle and the unorganized data will be called RapidAPI. For the Kaggle dataset I will need to read the CSV file into a pandas data frame. The RapidAPI dataset will be much harder, since it is stored in an API I will need to run an API call on it but that is just the start. Since the data is stored in many different APIs I will have to run multiple API calls. I will have to create a for loop to call the different types of APIs which the data is stored in. I will also have to modify the for loop if the information is stored in a different way.

The plan is to take the Kaggle dataset, rename columns, and review the rows. If the data looks well organized, I will then pull in the Rapid API, select the columns needed, renames the columns, merge tables, remove rows(duplicates), and complete a final review on the data. I will then compare the two datasets and see if there are discrepancies. I will have to do this many times over for the Rapid API dataset, there will be many different panda data frames I will have to search through to get all the information I will need.

When I am finally ready to load the two datasets to SQL I will need to create the PokemonGo database. Then I will have to create two empty tables(RapidAPI and Kaggle) where I will push my data. Next I will create a user and password to allow the python to push data to the dataframe. Finally I will need to connect to the local database then using pandas load my data frames into the two tables on the PokemonGo database.