Documentation for the “shell_pull” and “shell_looped” Python Files

shell_pull.py:

- Libraries needed: requests & os
- Function: get_dataset_name:
    - This function sends a post request to the “url” variable, in this case the url of NDGIS with the station id inserted in
    - The function returns the response text with no quotes, this response is used as the dataset name
- The dataset name from the get_dataset_name is used to create the url that the csv file is downloaded from; “waterchem_data_url”
- Function: download_file:
    - The first two lines use the os library to set a location in the files app for the downloaded csv file, this location can be changed to its own folder, it is currently set to put the csv files in the downloads folder
    - Then a GET request is sent to the “waterchem_data_url” to get the content of the file
    - If statement to check if the response status code ensures that the request worked, as if the status code is not 200 that means some error occurred with the request
        - If response.content checks to make sure the file gotten from the request isn’t empty, and as long as the file isn’t empty the csv is created

shell_looped.py:

- This file uses the functions created in shell_pull to loop through every station id in the masterlist.
- Make sure to import shell_pull and specifically the functions created, as well as the pandas library as that is what is currently used to read the excel master list of station ids
- First, an empty list is set for “station_ids”
- The excel file is read, and for every row in the column containing the station ids, the content of that row is added to the empty “station_ids” list
    - The df.iloc[:,0] reads EVERY row for the specified column, the 0 means column 1 of the excel file, so if the list of station ids is in another column, change 0 accordingly
- Then, a loop occurs for each station id in the list of station ids to download the csv for each station
    - The url is set for the NDGIS website for each station id
    - Run the get_dataset_name with the url being the NDGIS waterchem_url
    - The dataset returned is added into the waterchem_data_url
    - The download_file function is run for the waterchem_data_url, downloading the csv for the station id