THIS IS THE IMPLEMENTATION DOCUMENTATION FOR Integrated Cold Chain Availability Platform (ICAP) Scraping procedure

Agency name - National Horticulture Board(NHB)

Website Link: http://nhb.gov.in/Iccap.aspx

Step 1: Using the export to excel button, get the excel file (excel readable HTML file) on your local computer for every state.

Step 2: Now this file is a HTML file which is being viewed as an excel file. Thus, you cannot import this file directly to work on it.

This file must be converted to a csv file for working. We can simply save each file as a .csv file using Microsoft Excel which would do the trick.

Step 3: Now that the files are saved a comma-separated-value files, we can import them into a pandas Dataframe. The files are encoded in "ISO-8859-1" and thus we need to provide the encoding while importing the CSV files.

Along with that, the CSV files contain a bunch of information not required by us. To skip those rows, we use skipheader and skipfooter and get the final Dataframe.

Step 4: Once we get the Dataframe we can start the process of collating the CSV files by reading them one by one and appending them to the Dataframe object.

Step 5: Finally, we can name the columns as per the format and export the data to a csv file.

Step 6: The final CSV files contains some rows which do not have a project name and only contains miscellaneous extra data which makes the table imperfect. We remove those rows. We get the final collated data as per the requirements.