Due April 18th:

Your task will be to download data from a website, clean up the dataframe, and then store it as an excel file.

You should web scrap a table from https://en.wikipedia.org/wiki/MonsterVerse

that includes the different characters and the movies those characters appeared in. On Wikipedia, the table looks like this:

Table

Description automatically generated

You should web scrap using pd.read\_html.

To clean up the dataframe you web scrap, please do the following:

1. make the index the different characters instead of 0, 1, 2, 3. That is, it should be Godzilla, MUTO, King Kong,….
2. change the column names to just be the names of the movies (e.g., it should just be Godzilla instead of Films, Godzilla, Monsters. Also, change the index name to also just be “character” instead of Character, Character, Monster
3. remove the row that has all duplicates (i.e., “human”, “human”, “human”), since that doesn’t seem helpful to us in this dataframe.
4. save the file as a .csv with the delimiter being tabs

Your blocks of code should be: (1) importing the libraries, (2) web scraping the data, (3) doing steps 1-3 above, and (4) doing step 4 above.

For extra credit (1 point), you should convert steps 1-3 above into a function that is able to do the same kinds of cleaning for the cast of characters and movies for the following link: https://en.wikipedia.org/wiki/List\_of\_Fast\_%26\_Furious\_characters

For the extra credit, you may have to be clever with how you do steps 1-3 to make it account for the different numbers of columns (though, there will be partial credit for any attempts!). If you are doing the extra credit, your blocks of code should be (1) importing libraries, (2) web scraping data, (3) your function, (4) running the function, (5) writing the .csv file, and (6) repeating 4-5 but with the Fast and Furious link.

See below for examples of how the tables looked at the start and how they looked at the end (based on my solution):

Here is the original table when loaded into python: df.head(7)Table

Description automatically generated with medium confidence

And this is the table I expect: cleaned\_df.head(7)Table

Description automatically generated

It may not be a perfect table. There is still some cleaning up to do inside of the cells (e.g., the super and subscripts look strange), but at least the tables are more manageable now.