HTML Module 11 Challenge

Hi and thank you for reaching out to [AskBCS.My](http://askbcs.my/" \t "_blank) name is Dinh(he/him)



**Tyler Evans**  [1:43 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674330236979249)

Hi Dinh - thank you for offering to help!

[](https://app.slack.com/services/B045KMQ3G1X)

**dchauAPP**  [1:44 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674330291596729)

which step are you on? I am gonna pull up the instructions on my end



**Tyler Evans**  [1:46 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674330417879939)

I have extracted all rows of data using the following code:  
# Extract all rows of data  
table=html\_soup.find('table', class\_='table')  
table  
I'm now on:  
# Create an empty list  
mars\_facts=[]  
# Loop through the scraped data to create a list of rows  
Then: # Create a Pandas DataFrame by using the list of rows and a list of the column names

[](https://app.slack.com/services/B045KMQ3G1X)

**dchauAPP**  [1:47 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674330466620839)

okay give me a moment to look at the code you shared and type something out



**Tyler Evans**  [1:47 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674330478738929)

Thank you :slightly_smiling_face:

[](https://app.slack.com/services/B045KMQ3G1X)

**dchauAPP**  [1:54 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674330859316759)

So I will paste your code here and then provide some hints/advice on what to change. Due to AskBCS policies, I can not give you straight answers (such as giving you an entire piece of code) so if some parts are vague I apologize beforehand.

table=html\_soup.find('table', class\_='table')

table

1. So to extract all rows of data, it means just that. In the webpage there is a section called "data-row" you want to extract that.  
    a. so you need to alter the code a bit. So the things inside the parentheses specify where to extract from on an web page  
    b. so what do you need to swap the first table with?  
    c. what do you put for class\_ = ?

rows = table.find\_all('tr')

for row in rows:

row\_dict= {}

row\_heading = row.find('th').text

row\_data = row.find('td').text.strip()

row\_dict[row\_heading] = row\_data

1. So you almost got it. for this part. So first thing,  comment out (rows = table.find\_all('tr'))  
    a. move row\_dict= {} and place it abovee for row in rows. Also in order to make an empty list it has to be square brackets not curly  
    b. for row\_heading remove the .text and replace th with td  
    c. for row\_data erase the current code. you need to make code that looks for column text. Essentially you will specify what you are looking for and make a for loop (extra hint: the code here has to be placed in square brackets to make the output clean)  
    d. the last line is where you would append row\_dict (edited)



**Tyler Evans**  [1:55 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674330903108999)

No apologize I appreciate any help offered. I'll probably ask some questions while I'm digesting this.

[1:58](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674331116768379)

For number 1 that makes sense replace html\_soup.find('table', class\_='table') with  
html\_soup.find('tr', class\_='data-row')

[](https://app.slack.com/services/B045KMQ3G1X)

**dchauAPP**  [1:59 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674331162754849)

yes

[2:00](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674331200603279)

so you need that code will ensure you get all the data you need to loop through



**Tyler Evans**  [2:00 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674331206432349)

But it would actually be html\_soup.find\_all('tr', class\_='data-row') correct?

[](https://app.slack.com/services/B045KMQ3G1X)

**dchauAPP**  [2:00 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674331247137159)

yes that is correct

[2:01](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674331262133019)

so I realize something else

[2:02](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674331346680379)

since your variable is named table for your html soup.find\_all codeyou need to plug that variable in the for loop codeSo instead of

for row in rows:

you would instead make it like this

for row in table:



**Tyler Evans**  [2:06 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674331590425449)

Okay so I would I need to do a nested for loop? Where I would go through the table and grab each tr (row) and then grab all the data for each data (td)?

[](https://app.slack.com/services/B045KMQ3G1X)

**dchauAPP**  [2:09 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674331744446779)

reread what I wrote out. in tandem with this link  
<https://www.w3schools.com/tags/tag_col.asp> (edited)

**w3schools.com**

[**HTML col tag**](https://www.w3schools.com/tags/tag_col.asp)

W3Schools offers free online tutorials, references and exercises in all the major languages of the web. Covering popular subjects like HTML, CSS, JavaScript, Python, SQL, Java, and many, many more. (6 kB)

<https://www.w3schools.com/tags/tag_col.asp>

[2:10](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674331857496199)

for the link i shared what is the tag the link focuses on



**Tyler Evans**  [2:11 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674331861060799)

Good deal - I'll reread your hints - sorry I think I missed reading some of them regarding the for loop. Please standby for questions if you can...



**Tyler Evans**  [2:17 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674332269102249)

So the link you shared focuses on the column tag...

[](https://app.slack.com/services/B045KMQ3G1X)

**dchauAPP**  [2:19 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674332341989759)

yes



**Tyler Evans**  [2:20 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674332450520899)

That website is showing how it's imbedded, but not really how I retrieve it. Are you saying I need to try to retrieve the data based on the th headers?

[](https://app.slack.com/services/B045KMQ3G1X)

**dchauAPP**  [2:23 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674332596599589)

so the code itself is really simple  
I rewrote my hint a bit  
Hint:  
    a. for row\_data erase the current code. you need to make code that looks for column text. Essentially you will specify you are looking for column text and then apply a for loop that looks for columns (extra hint: the code here has to be placed in square brackets to make the output clean) (edited)

[2:24](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674332671419339)

my other hint is the .text function



**Tyler Evans**  [2:26 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674332808622859)

Okay I'm going to need to step away - I'm getting pretty lost and frustrated. I'll have to reread your hints with fresh eyes. Thank you for your help

[](https://app.slack.com/services/B045KMQ3G1X)

**dchauAPP**  [2:27 PM](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674332864993879)

no problem. Sorry that my hints left you frustrated

[2:28](https://kuvirtdatapt1-ojo1897.slack.com/archives/C04KZ38DY2E/p1674332933937889)

So I will close the question. If you work on something and find yourself stuck you can always reach out again. After I close the question if you can give some feedback that would be appreciated. Have a great day and/or night