In [3]:
```
import requests
from bs4 import BeautifulSoup as bs
```

In [8]:
```
r = requests.get("https://keithgalli.github.io/web-scraping/example.html")
soup = bs(r.content)
print(soup.prettify())
```

```
<html>
 <head>
  <title>
   HTML Example
  </title>
 </head>
 <body>
  <div align="middle">
   <h1>
    HTML Webpage
   </h1>
   <p>
    Link to more interesting example:
    <a href="https://keithgalli.github.io/web-scraping/webpage.html">
     keithgalli.github.io/web-scraping/webpage.html
    </a>
   </p>
  </div>
  <h2>
   A Header
  </h2>
  <p>
   <i>
    Some italicized text
   </i>
  </p>
  <h2>
   Another header
  </h2>
  <p id="paragraph-id">
   <b>
    Some bold text
   </b>
  </p>
 </body>
</html>
```

In [9]:
```
first_header = soup.find('h2')
first_header
```

Out[9]:
```
<h2>A Header</h2>
```

In [10]:
```
headers = soup.find_all('h2')
headers
```

Out[10]:
```
[<h2>A Header</h2>, <h2>Another header</h2>]
```

In [11]:
```
list_header = soup.find(['h1','h2'])
list_header
```

Out[11]:
```
<h1>HTML Webpage</h1>
```

In [12]:
```
list_header2 = soup.find_all(['h1','h2'])
list_header2
```

Out[12]:
```
[<h1>HTML Webpage</h1>, <h2>A Header</h2>, <h2>Another header</h2>]
```

In [13]:
```
paragraph = soup.find_all('p', attrs={'id':'paragraph-id'})
paragraph
```

Out[13]:
```
[<p id="paragraph-id"><b>Some bold text</b></p>]
```

In [14]:
```
body = soup.find('body')
body
```

Out[14]:
```
<body>
<div align="middle">
<h1>HTML Webpage</h1>
<p>Link to more interesting example: <a href="https://keithgalli.github.io/web-scraping/we
bpage.html">keithgalli.github.io/web-scraping/webpage.html</a></p>
</div>
<h2>A Header</h2>
<p><i>Some italicized text</i></p>
<h2>Another header</h2>
<p id="paragraph-id"><b>Some bold text</b></p>
</body>
```

In [15]:
```
div = body.find('div')
div
```

Out[15]:
```
<div align="middle">
<h1>HTML Webpage</h1>
<p>Link to more interesting example: <a href="https://keithgalli.github.io/web-scraping/we
bpage.html">keithgalli.github.io/web-scraping/webpage.html</a></p>
</div>
```

In [16]:
```
dive_header = div.find('h1')
dive_header
```

Out[16]:
```
<h1>HTML Webpage</h1>
```

In [17]:
```
print(soup.prettify())
```

```
<html>
 <head>
  <title>
   HTML Example
  </title>
 </head>
 <body>
  <div align="middle">
   <h1>
    HTML Webpage
   </h1>
   <p>
    Link to more interesting example:
    <a href="https://keithgalli.github.io/web-scraping/webpage.html">
     keithgalli.github.io/web-scraping/webpage.html
    </a>
   </p>
  </div>
  <h2>
```

```
  </h2>
  <p>
   <i>
    Some italicized text
   </i>
  </p>
  <h2>
   Another header
  </h2>
  <p id="paragraph-id">
   <b>
    Some bold text
   </b>
  </p>
 </body>
</html>
```

In [19]:
```
soup.find_all('p',string = (some))#or else u can do this
```

Out[19]:
```
[]
```

In [20]:
```
import re

para = soup.find_all('p', string = re.compile('some'))
para
```

Out[20]:
```
[]
```

In [21]:
```
header3 =  soup.find_all('h2', string = re.compile('(h|H)eader'))
header3
```

Out[21]:
```
[<h2>A Header</h2>, <h2>Another header</h2>]
```

In [22]:
```
c = soup.select('p')
c
```

Out[22]:
```
[<p>Link to more interesting example: <a href="https://keithgalli.github.io/web-scraping/webpage.html">keithgalli.github.io/web-scraping/webpage.html</a></p>,
 <p><i>Some italicized text</i></p>,
 <p id="paragraph-id"><b>Some bold text</b></p>]
```

In [23]:
```
i = soup.select('div p')#only print paragraph inside div
i
```

Out[23]:
```
[<p>Link to more interesting example: <a href="https://keithgalli.github.io/web-scraping/webpage.html">keithgalli.github.io/web-scraping/webpage.html</a></p>]
```

In [24]:
```
soup.body

paragraph2 =soup.select('h2 ~p')#p that are directly after h2
paragraph2
```

Out[24]:
```
[<p><i>Some italicized text</i></p>,
 <p id="paragraph-id"><b>Some bold text</b></p>]
```

In [25]:
```
bold_text = soup.select('p#paragraph-id b')
bold_text
```

Out[25]:
```
[<b>Some bold text</b>]
```

In [26]:
```
soup.select('title')
```

Out[26]:
```
[<title>HTML Example</title>]
```

In [27]:
```
para2 = soup.select('body > p')#body followed by a direct p element (p-paragraph)
print(para2)
```

```
[<p><i>Some italicized text</i></p>, <p id="paragraph-id"><b>Some bold text</b></p>]
```

In [30]:
```
for paragraph in para2:
   print(para2.find('i'))
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
<ipython-input-30-b3bafedc2f8c> in <module>
      1 for paragraph in para2:
----> 2    print(para2.find('i'))

c:\users\andre\appdata\local\programs\python\python39\lib\site-packages\bs4\element.py in __getattr__(self, key)
   2171     def __getattr__(self, key):
   2172         """Raise a helpful exception to explain a common code fix."""
-> 2173         raise AttributeError(
   2174             "ResultSet object has no attribute '%s'. You're probably treating a list of elements like a single element. Did you call find_all() when you meant to call find()?" % key
   2175         )

AttributeError: ResultSet object has no attribute 'find'. You're probably treating a list of elements like a single element. Did you call find_all() when you meant to call find()?
```

In [31]:
```
soup.select("[align=middle]")
```

Out[31]:
```
[<div align="middle">
 <h1>HTML Webpage</h1>
 <p>Link to more interesting example: <a href="https://keithgalli.github.io/web-scraping/webpage.html">keithgalli.github.io/web-scraping/webpage.html</a></p>
 </div>]
```

In [32]:
```
h3 = soup.find('h2')
h3.string
```

Out[32]:
```
'A Header'
```

In [33]:
```
print(div.prettify())
```

```
<div align="middle">
 <h1>
  HTML Webpage
 </h1>
 <p>
  Link to more interesting example:
  <a href="https://keithgalli.github.io/web-scraping/webpage.html">
   keithgalli.github.io/web-scraping/webpage.html
  </a>
 </p>
</div>
```

In [34]:
```
print(div.string)
```

```
None
```

In [35]:
```
print(div.get_text())# if multiple element use get_text
```

```
HTML Webpage
Link to more interesting example: keithgalli.github.io/web-scraping/webpage.html
```

In [36]:
```
#getting specific property from the element
link = soup.find('a')
link
```

Out[36]:
```
<a href="https://keithgalli.github.io/web-scraping/webpage.html">keithgalli.github.io/web-scraping/webpage.html</a>
```

In [37]:
```
link['href']
```

Out[37]:
```
'https://keithgalli.github.io/web-scraping/webpage.html'
```

In [39]:
```
paragraphs = soup.select('p#paragraph-id')
paragraphs[0]['id']
```

Out[39]:
```
'paragraph-id'
```

In [40]:
```
paragraphs
```

Out[40]:
```
[<p id="paragraph-id"><b>Some bold text</b></p>]
```

In [42]:
```
paragraphs[0]
```

Out[42]:
```
<p id="paragraph-id"><b>Some bold text</b></p>
```

In [43]:
```
# code navigation

soup.body
```

Out[43]:
```
<body>
<div align="middle">
<h1>HTML Webpage</h1>
<p>Link to more interesting example: <a href="https://keithgalli.github.io/web-scraping/webpage.html">keithgalli.github.io/web-scraping/webpage.html</a></p>
</div>
<h2>A Header</h2>
<p><i>Some italicized text</i></p>
<h2>Another header</h2>
<p id="paragraph-id"><b>Some bold text</b></p>
</body>
```

In [44]:
```
soup.body.div
```

Out[44]:
```
<div align="middle">
<h1>HTML Webpage</h1>
<p>Link to more interesting example: <a href="https://keithgalli.github.io/web-scraping/webpage.html">keithgalli.github.io/web-scraping/webpage.html</a></p>
```

In [45]:
```
soup.body.div.h1
```

Out[45]:
```
<h1>HTML Webpage</h1>
```

In [46]:
```
soup.body.div.h1.string
```

Out[46]:
```
'HTML Webpage'
```

In [47]:
```
## know the terms : paraents,siblings,child
# in here the BODY is the parent of div and div is the child of the body
# and the elements of the same level are sibling for example:body and the narrow down div
print(soup.body.prettify())
```

```
<body>
 <div align="middle">
  <h1>
   HTML Webpage
  </h1>
  <p>
   Link to more interesting example:
   <a href="https://keithgalli.github.io/web-scraping/webpage.html">
    keithgalli.github.io/web-scraping/webpage.html
   </a>
  </p>
 </div>
 <h2>
  A Header
 </h2>
 <p>
  <i>
   Some italicized text
  </i>
 </p>
 <h2>
  Another header
 </h2>
 <p id="paragraph-id">
  <b>
   Some bold text
  </b>
 </p>
</body>
```

In [48]:
```
soup.body.find('div').find_next_siblings()
```

Out[48]:
```
[<h2>A Header</h2>,
 <p><i>Some italicized text</i></p>,
 <h2>Another header</h2>,
 <p id="paragraph-id"><b>Some bold text</b></p>]
```

In [49]:
```
soup.body.find('h2').find_previous_siblings()
```

Out[49]:
```
[<div align="middle">
 <h1>HTML Webpage</h1>
 <p>Link to more interesting example: <a href="https://keithgalli.github.io/web-scraping/w
ebpage.html">keithgalli.github.io/web-scraping/webpage.html</a></p>
 </div>]
```

In [50]:
```
soup.find('h1').find_parent()
```

Out[50]:
```
<div align="middle">
<h1>HTML Webpage</h1>
<p>Link to more interesting example: <a href="https://keithgalli.github.io/web-scraping/we
bpage.html">keithgalli.github.io/web-scraping/webpage.html</a></p>
</div>
```

In [51]:
```
soup.find('div').find_parent()
```

Out[51]:
```
<body>
<div align="middle">
<h1>HTML Webpage</h1>
<p>Link to more interesting example: <a href="https://keithgalli.github.io/web-scraping/we
bpage.html">keithgalli.github.io/web-scraping/webpage.html</a></p>
</div>
<h2>A Header</h2>
<p><i>Some italicized text</i></p>
<h2>Another header</h2>
<p id="paragraph-id"><b>Some bold text</b></p>
</body>
```

In [52]:
```
# excercise 1
## task1
## find all links

r2=requests.get('https://keithgalli.github.io/web-scraping/webpage.html')
soup1 = bs(r2.content)
print(soup1.prettify())
```

```
<head>
 <title>
  Keith Galli's Page
 </title>
 <style>
  table {
    border-collapse: collapse;
  }
  th {
    padding:5px;
  }
  td {
    border: 1px solid #ddd;
    padding: 5px;
  }
  tr:nth-child(even) {
    background-color: #f2f2f2;
  }
  th {
    padding-top: 12px;
    padding-bottom: 12px;
    text-align: left;
    background-color: #add8e6;
    color: black;
  }
  .block {
  width: 100px;
  /*float: left;*/
    display: inline-block;
    zoom: 1;
  }
  .column {
  float: left;
  height: 200px;
  /*width: 33.33%;*/
  padding: 5px;
  }
```

```
    content: "";
    clear: both;
    display: table;
  }
 </style>
</head>
<body>
 <h1>
  Welcome to my page!
 </h1>
 <img src="./images/selfie1.jpg" width="300px"/>
 <h2>
  About me
 </h2>
 <p>
  Hi, my name is Keith and I am a YouTuber who focuses on content related to programming,
data science, and machine learning!
 </p>
 <p>
  Here is a link to my channel:
  <a href="https://www.youtube.com/kgmit">
   youtube.com/kgmit
  </a>
 </p>
 <p>
  I grew up in the great state of New Hampshire here in the USA. From an early age I alway
s loved math. Around my senior year of high school, my brother first introduced me to prog
ramming. I found it a creative way to apply the same type of logical thinking skills that
I enjoyed with math. This influenced me to study computer science in college and ultimatel
y create a YouTube channel to share some things that I have learned along the way.
 </p>
 <h3>
  Hobbies
 </h3>
 <p>
  Believe it or not, I don't code 24/7. I love doing all sorts of active things. I like to
play ice hockey &amp;amp; table tennis as well as run, hike, skateboard, and snowboard. In add
ition to sports, I am a board game enthusiast. The two that I've been playing the most rec
ently are
  <i>
   Settlers of Catan
  </i>
  and
  <i>
   Othello
  </i>
  .
 </p>
 <h3>
  Fun Facts
 </h3>
 <ul class="fun-facts">
  <li>
   Owned my dream car in high school
   <a href="#footer">
    <sup>
     1
    </sup>
   </a>
  </li>
  <li>
   Middle name is Ronald
  </li>
  <li>
   Never had been on a plane until college
  </li>
  <li>
   Dunkin Donuts coffee is better than Starbucks
  </li>
```

```
  A favorite book series of mine is
  <i>
   Ender's Game
  </i>
 </li>
 <li>
  Current video game of choice is
  <i>
   Rocket League
  </i>
 </li>
 <li>
  The band that I've seen the most times live is the
  <i>
   Zac Brown Band
  </i>
 </li>
</ul>
<h2>
 Social Media
</h2>
I encourage you to check out my content on all social media platforms
<br/>
<ul class="socials">
 <li class="social instagram">
  <b>
   Instagram:
  </b>
  <a href="https://www.instagram.com/keithgalli/">
   https://www.instagram.com/keithgalli/
  </a>
 </li>
 <li class="social twitter">
  <b>
   Twitter:
  </b>
  <a href="https://twitter.com/keithgalli">
   https://twitter.com/keithgalli
  </a>
 </li>
 <li class="social linkedin">
  <b>
   LinkedIn:
  </b>
  <a href="https://www.linkedin.com/in/keithgalli/">
   https://www.linkedin.com/in/keithgalli/
  </a>
 </li>
 <li class="social tiktok">
  <b>
   TikTok:
  </b>
  <a href="https://www.tiktok.com/@keithgalli">
   https://www.tiktok.com/@keithgalli
  </a>
 </li>
</ul>
<h2>
 Photos
</h2>
Here are a few photos from a trip to italy I took last year
<div class="row">
 <div class="column">
  <img alt="Lake Como" src="images/italy/lake_como.jpg" style="height:100%"/>
 </div>
 <div class="column">
  <img alt="Pontevecchio, Florence" src="images/italy/pontevecchio.jpg" style="height:10
0%"/>
  </div>
<div class="column">
```

```
   <img alt="Riomaggiore, Cinque de Terre" src="images/italy/riomaggiore.jpg" style="heigh
t:100%"/>
  </div>
 </div>
 <div>
 </div>
 <h2>
  Table
 </h2>
 My MIT hockey stats :)
 <br/>
 <table class="hockey-stats">
  <thead>
   <tr>
    <th class="season" data-sort="">
     S
    </th>
    <th class="team" data-sort="team">
     Team
    </th>
    <th class="league" data-sort="league">
     League
    </th>
    <th class="regular gp" data-sort="gp">
     GP
    </th>
    <th class="regular g" data-sort="g">
     G
    </th>
    <th class="regular a" data-sort="a">
     A
    </th>
    <th class="regular tp" data-sort="tp">
     TP
    </th>
    <th class="regular pim" data-sort="pim">
     PIM
    </th>
    <th class="regular pm" data-sort="pm">
     +/-
    </th>
    <th class="separator">
    </th>
    <th class="postseason">
     POST
    </th>
    <th class="postseason gp" data-sort="playoffs-gp">
     GP
    </th>
    <th class="postseason g" data-sort="playoffs-g">
     G
    </th>
    <th class="postseason a" data-sort="playoffs-a">
     A
    </th>
    <th class="postseason tp" data-sort="playoffs-tp">
     TP
    </th>
    <th class="postseason pim" data-sort="playoffs-pim">
     PIM
    </th>
    <th class="postseason pm" data-sort="playoffs-pm">
     +/-
    </th>
   </tr>
  </thead>
  <tbody>
   <tr class="team-continent-NA">
    <td class="season sorted">
```

```
    </td>
    <td class="team">
     <i>
      <img src="images/flag.png"/>
     </i>
     <span class="txt-blue">
      <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2014-201
5?tab=stats">
       MIT (Mass. Inst. of Tech.)
      </a>
     </span>
    </td>
    <td class="league">
     <a href="https://www.eliteprospects.com/league/acha-ii/stats/2014-2015">
      ACHA II
     </a>
    </td>
    <td class="regular gp">
     17
    </td>
    <td class="regular g">
     3
    </td>
    <td class="regular a">
     9
    </td>
    <td class="regular tp">
     12
    </td>
    <td class="regular pim">
     20
    </td>
    <td class="regular pm">
    </td>
    <td class="separator">
     |
    </td>
    <td class="postseason">
     <a href="https://www.eliteprospects.com/league/acha-ii/stats/2014-2015">
     </a>
    </td>
    <td class="postseason gp">
    </td>
    <td class="postseason g">
    </td>
    <td class="postseason a">
    </td>
    <td class="postseason tp">
    </td>
    <td class="postseason pim">
    </td>
    <td class="postseason pm">
    </td>
   </tr>
   <tr class="team-continent-NA">
    <td class="season sorted">
     2015-16
    </td>
    <td class="team">
     <i>
      <img src="images/flag.png"/>
     </i>
     <span class="txt-blue">
      <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2015-201
6?tab=stats">
       MIT (Mass. Inst. of Tech.)
      </a>
     </span>
    </td>
    <td class="league">
```

```
      <a href="https://www.eliteprospects.com/league/acha-ii/stats/2015-2016">
       ACHA II
      </a>
     </td>
     <td class="regular gp">
      9
     </td>
     <td class="regular g">
      1
     </td>
     <td class="regular a">
      1
     </td>
     <td class="regular tp">
      2
     </td>
     <td class="regular pim">
      2
     </td>
     <td class="regular pm">
     </td>
     <td class="separator">
      |
     </td>
     <td class="postseason">
      <a href="https://www.eliteprospects.com/league/acha-ii/stats/2015-2016">
      </a>
     </td>
     <td class="postseason gp">
     </td>
     <td class="postseason g">
     </td>
     <td class="postseason a">
     </td>
     <td class="postseason tp">
     </td>
     <td class="postseason pim">
     </td>
     <td class="postseason pm">
     </td>
    </tr>
    <tr class="team-continent-NA">
     <td class="season sorted">
      2016-17
     </td>
     <td class="team">
      <i>
       <img src="images/flag.png"/>
      </i>
      <span class="txt-blue">
       <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2016-201
7?tab=stats">
        MIT (Mass. Inst. of Tech.)
       </a>
      </span>
     </td>
     <td class="league">
      <a href="https://www.eliteprospects.com/league/acha-ii/stats/2016-2017">
       ACHA II
      </a>
     </td>
     <td class="regular gp">
      12
     </td>
     <td class="regular g">
      5
     </td>
     <td class="regular a">
      5
```

```
 <td class="regular tp">
  10
 </td>
 <td class="regular pim">
  8
 </td>
 <td class="regular pm">
  0
 </td>
 <td class="separator">
  |
 </td>
 <td class="postseason">
 </td>
 <td class="postseason gp">
 </td>
 <td class="postseason g">
 </td>
 <td class="postseason a">
 </td>
 <td class="postseason tp">
 </td>
 <td class="postseason pim">
 </td>
 <td class="postseason pm">
 </td>
</tr>
<tr class="team-continent-EU">
 <td class="season sorted">
  2017-18
 </td>
 <td class="team">
  Did not play
 </td>
 <td class="league">
  <a href="https://www.eliteprospects.com/stats">
  </a>
 </td>
 <td class="regular gp">
 </td>
 <td class="regular g">
 </td>
 <td class="regular a">
 </td>
 <td class="regular tp">
 </td>
 <td class="regular pim">
 </td>
 <td class="regular pm">
 </td>
 <td class="separator">
  |
 </td>
 <td class="postseason">
  <a href="https://www.eliteprospects.com/stats">
  </a>
 </td>
 <td class="postseason gp">
 </td>
 <td class="postseason g">
 </td>
 <td class="postseason a">
 </td>
 <td class="postseason tp">
 </td>
 <td class="postseason pim">
 </td>
 <td class="postseason pm">
 </td>
```

```
   <tr class="team-continent-NA">
    <td class="season sorted">
     2018-19
    </td>
    <td class="team">
     <i>
      <img src="images/flag.png"/>
     </i>
     <span class="txt-blue">
      <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2018-201
9?tab=stats">
       MIT (Mass. Inst. of Tech.)
      </a>
     </span>
    </td>
    <td class="league">
     <a href="https://www.eliteprospects.com/league/acha-iii/stats/2018-2019">
      ACHA III
     </a>
    </td>
    <td class="regular gp">
     8
    </td>
    <td class="regular g">
     5
    </td>
    <td class="regular a">
     10
    </td>
    <td class="regular tp">
     15
    </td>
    <td class="regular pim">
     8
    </td>
    <td class="regular pm">
    </td>
    <td class="separator">
     |
    </td>
    <td class="postseason">
     <a href="https://www.eliteprospects.com/league/acha-iii/stats/2018-2019">
     </a>
    </td>
    <td class="postseason gp">
    </td>
    <td class="postseason g">
    </td>
    <td class="postseason a">
    </td>
    <td class="postseason tp">
    </td>
    <td class="postseason pim">
    </td>
    <td class="postseason pm">
    </td>
   </tr>
  </tbody>
 </table>
 <h2>
  Mystery Message Challenge!
 </h2>
 <p>
  If you scrape the links below grabbing the &lt;p&gt; tag with id="secret-word", you'll d
iscover a secret message :)
 </p>
 <div width="50%">
  <div align="left" class="block">
   <ul>
```

```
     <a href="challenge/file_1.html">
      File 1
     </a>
    </li>
    <li>
     <a href="challenge/file_2.html">
      File 2
     </a>
    </li>
    <li>
     <a href="challenge/file_3.html">
      File 3
     </a>
    </li>
    <li>
     <a href="challenge/file_4.html">
      File 4
     </a>
    </li>
    <li>
     <a href="challenge/file_5.html">
      File 5
     </a>
    </li>
   </ul>
  </div>
  <div align="center" class="block">
   <ul>
    <li>
     <a href="challenge/file_6.html">
      File 6
     </a>
    </li>
    <li>
     <a href="challenge/file_7.html">
      File 7
     </a>
    </li>
    <li>
     <a href="challenge/file_8.html">
      File 8
     </a>
    </li>
    <li>
     <a href="challenge/file_9.html">
      File 9
     </a>
    </li>
    <li>
     <a href="challenge/file_10.html">
      File 10
     </a>
    </li>
   </ul>
  </div>
 </div>
 <h2>
  Footnotes
 </h2>
 <p id="footer">
  1. This was actually a minivan that I named Debora. Maybe not my dream car, but I loved
her nonetheless.
 </p>
</body>
```

In [53]:

```
soup1.select('a')
```

Loading [MathJax]/extensions/Safe.js

```
/www.youtube.com/kgmit">youtube.com/kgmit</a>,
```

Out[53]:
```
 <a href="#footer"><sup>1</sup></a>,
 <a href="https://www.instagram.com/keithgalli/">https://www.instagram.com/keithgalli/</a
>,
 <a href="https://twitter.com/keithgalli">https://twitter.com/keithgalli</a>,
 <a href="https://www.linkedin.com/in/keithgalli/">https://www.linkedin.com/in/keithgalli/
</a>,
 <a href="https://www.tiktok.com/@keithgalli">https://www.tiktok.com/@keithgalli</a>,
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2014-2015?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>,
 <a href="https://www.eliteprospects.com/league/acha-ii/stats/2014-2015"> ACHA II </a>,
 <a href="https://www.eliteprospects.com/league/acha-ii/stats/2014-2015"> </a>,
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2015-2016?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>,
 <a href="https://www.eliteprospects.com/league/acha-ii/stats/2015-2016"> ACHA II </a>,
 <a href="https://www.eliteprospects.com/league/acha-ii/stats/2015-2016"> </a>,
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2016-2017?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>,
 <a href="https://www.eliteprospects.com/league/acha-ii/stats/2016-2017"> ACHA II </a>,
 <a href="https://www.eliteprospects.com/stats"> </a>,
 <a href="https://www.eliteprospects.com/stats"> </a>,
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2018-2019?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>,
 <a href="https://www.eliteprospects.com/league/acha-iii/stats/2018-2019"> ACHA III </a>,
 <a href="https://www.eliteprospects.com/league/acha-iii/stats/2018-2019"> </a>,
 <a href="challenge/file_1.html">File 1</a>,
 <a href="challenge/file_2.html">File 2</a>,
 <a href="challenge/file_3.html">File 3</a>,
 <a href="challenge/file_4.html">File 4</a>,
 <a href="challenge/file_5.html">File 5</a>,
 <a href="challenge/file_6.html">File 6</a>,
 <a href="challenge/file_7.html">File 7</a>,
 <a href="challenge/file_8.html">File 8</a>,
 <a href="challenge/file_9.html">File 9</a>,
 <a href="challenge/file_10.html">File 10</a>]
```

In [54]:
```
links=soup1.select('ul.socials a')
print(links)
```

```
[<a href="https://www.instagram.com/keithgalli/">https://www.instagram.com/keithgalli/</a
>, <a href="https://twitter.com/keithgalli">https://twitter.com/keithgalli</a>, <a href="h
ttps://www.linkedin.com/in/keithgalli/">https://www.linkedin.com/in/keithgalli/</a>, <a hr
ef="https://www.tiktok.com/@keithgalli">https://www.tiktok.com/@keithgalli</a>]
```

In [55]:
```
actual_links =[link['href'] for link in links]
actual_links
```

Out[55]:
```
['https://www.instagram.com/keithgalli/',
 'https://twitter.com/keithgalli',
 'https://www.linkedin.com/in/keithgalli/',
 'https://www.tiktok.com/@keithgalli']
```

In [56]:
```
ulist = soup1.find("ul", attrs = {"class":"socials"})
print(ulist)
```

```
<ul class="socials">
<li class="social instagram"><b>Instagram: </b><a href="https://www.instagram.com/keithgal
li/">https://www.instagram.com/keithgalli/</a></li>
<li class="social twitter"><b>Twitter: </b><a href="https://twitter.com/keithgalli">http
s://twitter.com/keithgalli</a></li>
<li class="social linkedin"><b>LinkedIn: </b><a href="https://www.linkedin.com/in/keithgal
li/">https://www.linkedin.com/in/keithgalli/</a></li>
<li class="social tiktok"><b>TikTok: </b><a href="https://www.tiktok.com/@keithgalli">http
s://www.tiktok.com/@keithgalli</a></li>
</ul>
```

In [57]:
```
nd_all("a")
```

```
print(links)
```

```
[<a href="https://www.instagram.com/keithgalli/">https://www.instagram.com/keithgalli/</a
>, <a href="https://twitter.com/keithgalli">https://twitter.com/keithgalli</a>, <a href="h
ttps://www.linkedin.com/in/keithgalli/">https://www.linkedin.com/in/keithgalli/</a>, <a hr
ef="https://www.tiktok.com/@keithgalli">https://www.tiktok.com/@keithgalli</a>]
```

In [58]:
```
actual_links =[link['href'] for link in links]
actual_links
```

Out[58]:
```
['https://www.instagram.com/keithgalli/',
 'https://twitter.com/keithgalli',
 'https://www.linkedin.com/in/keithgalli/',
 'https://www.tiktok.com/@keithgalli']
```

In [60]:
```
links = soup1.select("li.social a")
print(links)
```

```
[<a href="https://www.instagram.com/keithgalli/">https://www.instagram.com/keithgalli/</a
>, <a href="https://twitter.com/keithgalli">https://twitter.com/keithgalli</a>, <a href="h
ttps://www.linkedin.com/in/keithgalli/">https://www.linkedin.com/in/keithgalli/</a>, <a hr
ef="https://www.tiktok.com/@keithgalli">https://www.tiktok.com/@keithgalli</a>]
```

In [61]:
```
actual_links =[link['href'] for link in links]
actual_links
```

Out[61]:
```
['https://www.instagram.com/keithgalli/',
 'https://twitter.com/keithgalli',
 'https://www.linkedin.com/in/keithgalli/',
 'https://www.tiktok.com/@keithgalli']
```

In [62]:
```
# scrape table

soup1.select('tbody')
```

Out[62]:
```
[<tbody>
 <tr class="team-continent-NA">
 <td class="season sorted">
                  2014-15
               </td>
 <td class="team">
 <i><img src="images/flag.png"/></i>
 <span class="txt-blue">
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2014-2015?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>
 </span>
 </td>
 <td class="league"> <a href="https://www.eliteprospects.com/league/acha-ii/stats/2014-201
5"> ACHA II </a> </td>
 <td class="regular gp">17</td>
 <td class="regular g">3</td>
 <td class="regular a">9</td>
 <td class="regular tp">12</td>
 <td class="regular pim">20</td>
 <td class="regular pm"></td>
 <td class="separator"> | </td>
 <td class="postseason">
 <a href="https://www.eliteprospects.com/league/acha-ii/stats/2014-2015"> </a>
 </td>
 <td class="postseason gp">
 </td>
 <td class="postseason g">
 </td>
 <td class="postseason a">
```

```
 </td>
 <td class="postseason tp">
 </td>
 <td class="postseason pim">
 </td>
 <td class="postseason pm">
 </td>
 </tr>
 <tr class="team-continent-NA">
 <td class="season sorted">
                  2015-16
               </td>
 <td class="team">
 <i><img src="images/flag.png"/></i>
 <span class="txt-blue">
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2015-2016?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>
 </span>
 </td>
 <td class="league"> <a href="https://www.eliteprospects.com/league/acha-ii/stats/2015-201
6"> ACHA II </a> </td>
 <td class="regular gp">9</td>
 <td class="regular g">1</td>
 <td class="regular a">1</td>
 <td class="regular tp">2</td>
 <td class="regular pim">2</td>
 <td class="regular pm"></td>
 <td class="separator"> | </td>
 <td class="postseason">
 <a href="https://www.eliteprospects.com/league/acha-ii/stats/2015-2016"> </a>
 </td>
 <td class="postseason gp">
 </td>
 <td class="postseason g">
 </td>
 <td class="postseason a">
 </td>
 <td class="postseason tp">
 </td>
 <td class="postseason pim">
 </td>
 <td class="postseason pm">
 </td>
 </tr>
 <tr class="team-continent-NA">
 <td class="season sorted">
                  2016-17
               </td>
 <td class="team">
 <i><img src="images/flag.png"/></i>
 <span class="txt-blue">
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2016-2017?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>
 </span>
 </td>
 <td class="league"> <a href="https://www.eliteprospects.com/league/acha-ii/stats/2016-201
7"> ACHA II </a> </td>
 <td class="regular gp">12</td>
 <td class="regular g">5</td>
 <td class="regular a">5</td>
 <td class="regular tp">10</td>
 <td class="regular pim">8</td>
 <td class="regular pm">0</td>
 <td class="separator"> | </td>
 <td class="postseason">
 </td>
 <td class="postseason gp">
 </td>
 <td class="postseason g">
```

```
<td class="postseason a">
</td>
<td class="postseason tp">
</td>
<td class="postseason pim">
</td>
<td class="postseason pm">
</td>
</tr>
<tr class="team-continent-EU">
<td class="season sorted">
                2017-18
            </td>
<td class="team">
                Did not play
            </td>
<td class="league"> <a href="https://www.eliteprospects.com/stats"> </a> </td>
<td class="regular gp"></td>
<td class="regular g"></td>
<td class="regular a"></td>
<td class="regular tp"></td>
<td class="regular pim"></td>
<td class="regular pm"></td>
<td class="separator"> | </td>
<td class="postseason">
<a href="https://www.eliteprospects.com/stats"> </a>
</td>
<td class="postseason gp">
</td>
<td class="postseason g">
</td>
<td class="postseason a">
</td>
<td class="postseason tp">
</td>
<td class="postseason pim">
</td>
<td class="postseason pm">
</td>
</tr>
<tr class="team-continent-NA">
<td class="season sorted">
                2018-19
            </td>
<td class="team">
<i><img src="images/flag.png"/></i>
<span class="txt-blue">
<a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2018-2019?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>
</span>
</td>
<td class="league"> <a href="https://www.eliteprospects.com/league/acha-iii/stats/2018-20
19"> ACHA III </a> </td>
<td class="regular gp">8</td>
<td class="regular g">5</td>
<td class="regular a">10</td>
<td class="regular tp">15</td>
<td class="regular pim">8</td>
<td class="regular pm"></td>
<td class="separator"> | </td>
<td class="postseason">
<a href="https://www.eliteprospects.com/league/acha-iii/stats/2018-2019"> </a>
</td>
<td class="postseason gp">
</td>
<td class="postseason g">
</td>
<td class="postseason a">
</td>
<td class="postseason tp">
```

```
 </td>
 <td class="postseason pim">
 </td>
 <td class="postseason pm">
 </td>
 </tr>
 </tbody>]
```

In [63]:

```
soup1.select('table td')
```

Out[63]:

```
[<td class="season sorted">
                    2014-15
                </td>,
 <td class="team">
 <i><img src="images/flag.png"/></i>
 <span class="txt-blue">
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2014-2015?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>
 </span>
 </td>,
 <td class="league"> <a href="https://www.eliteprospects.com/league/acha-ii/stats/2014-201
5"> ACHA II </a> </td>,
 <td class="regular gp">17</td>,
 <td class="regular g">3</td>,
 <td class="regular a">9</td>,
 <td class="regular tp">12</td>,
 <td class="regular pim">20</td>,
 <td class="regular pm"></td>,
 <td class="separator"> | </td>,
 <td class="postseason">
 <a href="https://www.eliteprospects.com/league/acha-ii/stats/2014-2015"> </a>
 </td>,
 <td class="postseason gp">
 </td>,
 <td class="postseason g">
 </td>,
 <td class="postseason a">
 </td>,
 <td class="postseason tp">
 </td>,
 <td class="postseason pim">
 </td>,
 <td class="postseason pm">
 </td>,
 <td class="season sorted">
                    2015-16
                </td>,
 <td class="team">
 <i><img src="images/flag.png"/></i>
 <span class="txt-blue">
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2015-2016?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>
 </span>
 </td>,
 <td class="league"> <a href="https://www.eliteprospects.com/league/acha-ii/stats/2015-201
6"> ACHA II </a> </td>,
 <td class="regular gp">9</td>,
 <td class="regular g">1</td>,
 <td class="regular a">1</td>,
 <td class="regular tp">2</td>,
 <td class="regular pim">2</td>,
 <td class="regular pm"></td>,
 <td class="separator"> | </td>,
 <td class="postseason">
 <a href="https://www.eliteprospects.com/league/acha-ii/stats/2015-2016"> </a>
 </td>,
 <td class="postseason gp">
 </td>,
 <td class="postseason g">
```

```
 </td>,
 <td class="postseason a">
 </td>,
 <td class="postseason tp">
 </td>,
 <td class="postseason pim">
 </td>,
 <td class="postseason pm">
 </td>,
 <td class="season sorted">
                    2016-17
                </td>,
 <td class="team">
 <i><img src="images/flag.png"/></i>
 <span class="txt-blue">
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2016-2017?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>
 </span>
 </td>,
 <td class="league"> <a href="https://www.eliteprospects.com/league/acha-ii/stats/2016-201
7"> ACHA II </a> </td>,
 <td class="regular gp">12</td>,
 <td class="regular g">5</td>,
 <td class="regular a">5</td>,
 <td class="regular tp">10</td>,
 <td class="regular pim">8</td>,
 <td class="regular pm">0</td>,
 <td class="separator"> | </td>,
 <td class="postseason">
 </td>,
 <td class="postseason gp">
 </td>,
 <td class="postseason g">
 </td>,
 <td class="postseason a">
 </td>,
 <td class="postseason tp">
 </td>,
 <td class="postseason pim">
 </td>,
 <td class="postseason pm">
 </td>,
 <td class="season sorted">
                    2017-18
                </td>,
 <td class="team">
                    Did not play
                </td>,
 <td class="league"> <a href="https://www.eliteprospects.com/stats"> </a> </td>,
 <td class="regular gp"></td>,
 <td class="regular g"></td>,
 <td class="regular a"></td>,
 <td class="regular tp"></td>,
 <td class="regular pim"></td>,
 <td class="regular pm"></td>,
 <td class="separator"> | </td>,
 <td class="postseason">
 <a href="https://www.eliteprospects.com/stats"> </a>
 </td>,
 <td class="postseason gp">
 </td>,
 <td class="postseason g">
 </td>,
 <td class="postseason a">
 </td>,
 <td class="postseason tp">
 </td>,
 <td class="postseason pim">
 </td>,
 <td class="postseason pm">
```

```
 </td>,
 <td class="season sorted">
                  2018-19
              </td>,
 <td class="team">
 <i><img src="images/flag.png"/></i>
 <span class="txt-blue">
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2018-2019?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>
 </span>
 </td>,
 <td class="league"> <a href="https://www.eliteprospects.com/league/acha-iii/stats/2018-20
19"> ACHA III </a> </td>,
 <td class="regular gp">8</td>,
 <td class="regular g">5</td>,
 <td class="regular a">10</td>,
 <td class="regular tp">15</td>,
 <td class="regular pim">8</td>,
 <td class="regular pm"></td>,
 <td class="separator"> | </td>,
 <td class="postseason">
 <a href="https://www.eliteprospects.com/league/acha-iii/stats/2018-2019"> </a>
 </td>,
 <td class="postseason gp">
 </td>,
 <td class="postseason g">
 </td>,
 <td class="postseason a">
 </td>,
 <td class="postseason tp">
 </td>,
 <td class="postseason pim">
 </td>,
 <td class="postseason pm">
 </td>]
```

In [72]:

```
import pandas as pd
table =soup1.select('table.hockey-stats')[0]

print(table)
```

```
<table class="hockey-stats">
<thead>
<tr>
<th class="season" data-sort="">S</th>
<th class="team" data-sort="team">Team</th>
<th class="league" data-sort="league">League</th>
<th class="regular gp" data-sort="gp">GP</th>
<th class="regular g" data-sort="g">G</th>
<th class="regular a" data-sort="a">A</th>
<th class="regular tp" data-sort="tp">TP</th>
<th class="regular pim" data-sort="pim">PIM</th>
<th class="regular pm" data-sort="pm">+/-</th>
<th class="separator"> </th>
<th class="postseason">POST</th>
<th class="postseason gp" data-sort="playoffs-gp">GP</th>
<th class="postseason g" data-sort="playoffs-g">G</th>
<th class="postseason a" data-sort="playoffs-a">A</th>
<th class="postseason tp" data-sort="playoffs-tp">TP</th>
<th class="postseason pim" data-sort="playoffs-pim">PIM</th>
<th class="postseason pm" data-sort="playoffs-pm">+/-</th>
</tr>
</thead>
<tbody>
<tr class="team-continent-NA">
<td class="season sorted">
                  2014-15
              </td>
```


<a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2014-2015?tab=
stats"> MIT (Mass. Inst. of Tech.)

</td>
<td class="league"> <a href="https://www.eliteprospects.com/league/acha-ii/stats/2014-201
5"> ACHA II </td>
<td class="regular gp">17</td>
<td class="regular g">3</td>
<td class="regular a">9</td>
<td class="regular tp">12</td>
<td class="regular pim">20</td>
<td class="regular pm"></td>
<td class="separator"> | </td>
<td class="postseason">

</td>
<td class="postseason gp">
</td>
<td class="postseason g">
</td>
<td class="postseason a">
</td>
<td class="postseason tp">
</td>
<td class="postseason pim">
</td>
<td class="postseason pm">
</td>
</tr>
<tr class="team-continent-NA">
<td class="season sorted">
2015-16
</td>
<td class="team">

<a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2015-2016?tab=
stats"> MIT (Mass. Inst. of Tech.)

</td>
<td class="league"> <a href="https://www.eliteprospects.com/league/acha-ii/stats/2015-201
6"> ACHA II </td>
<td class="regular gp">9</td>
<td class="regular g">1</td>
<td class="regular a">1</td>
<td class="regular tp">2</td>
<td class="regular pim">2</td>
<td class="regular pm"></td>
<td class="separator"> | </td>
<td class="postseason">

</td>
<td class="postseason gp">
</td>
<td class="postseason g">
</td>
<td class="postseason a">
</td>
<td class="postseason tp">
</td>
<td class="postseason pim">
</td>
<td class="postseason pm">
</td>
</tr>
<tr class="team-continent-NA">
<td class="season sorted">
2016-17

```
                </td>
<td class="team">
<i><img src="images/flag.png"/></i>
<span class="txt-blue">
<a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2016-2017?tab=
stats"> MIT (Mass. Inst. of Tech.) </a>
</span>
</td>
<td class="league"> <a href="https://www.eliteprospects.com/league/acha-ii/stats/2016-201
7"> ACHA II </a> </td>
<td class="regular gp">12</td>
<td class="regular g">5</td>
<td class="regular a">5</td>
<td class="regular tp">10</td>
<td class="regular pim">8</td>
<td class="regular pm">0</td>
<td class="separator"> | </td>
<td class="postseason">
</td>
<td class="postseason gp">
</td>
<td class="postseason g">
</td>
<td class="postseason a">
</td>
<td class="postseason tp">
</td>
<td class="postseason pim">
</td>
<td class="postseason pm">
</td>
</tr>
<tr class="team-continent-EU">
<td class="season sorted">
                    2017-18
                </td>
<td class="team">
                    Did not play
                </td>
<td class="league"> <a href="https://www.eliteprospects.com/stats"> </a> </td>
<td class="regular gp"></td>
<td class="regular g"></td>
<td class="regular a"></td>
<td class="regular tp"></td>
<td class="regular pim"></td>
<td class="regular pm"></td>
<td class="separator"> | </td>
<td class="postseason">
<a href="https://www.eliteprospects.com/stats"> </a>
</td>
<td class="postseason gp">
</td>
<td class="postseason g">
</td>
<td class="postseason a">
</td>
<td class="postseason tp">
</td>
<td class="postseason pim">
</td>
<td class="postseason pm">
</td>
</tr>
<tr class="team-continent-NA">
<td class="season sorted">
                    2018-19
                </td>
<td class="team">
<i><img src="images/flag.png"/></i>
-blue">
```

```
<a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2018-2019?tab=
stats"> MIT (Mass. Inst. of Tech.) </a>
</span>
</td>
<td class="league"> <a href="https://www.eliteprospects.com/league/acha-iii/stats/2018-201
9"> ACHA III </a> </td>
<td class="regular gp">8</td>
<td class="regular g">5</td>
<td class="regular a">10</td>
<td class="regular tp">15</td>
<td class="regular pim">8</td>
<td class="regular pm"></td>
<td class="separator"> | </td>
<td class="postseason">
<a href="https://www.eliteprospects.com/league/acha-iii/stats/2018-2019"> </a>
</td>
<td class="postseason gp">
</td>
<td class="postseason g">
</td>
<td class="postseason a">
</td>
<td class="postseason tp">
</td>
<td class="postseason pim">
</td>
<td class="postseason pm">
</td>
</tr>
</tbody>
</table>
```

In [73]:

```
columns = table.find('thead').find_all('th')
print(columns)
```

```
[<th class="season" data-sort="">S</th>, <th class="team" data-sort="team">Team</th>, <th
class="league" data-sort="league">League</th>, <th class="regular gp" data-sort="gp">GP</t
h>, <th class="regular g" data-sort="g">G</th>, <th class="regular a" data-sort="a">A</th
>, <th class="regular tp" data-sort="tp">TP</th>, <th class="regular pim" data-sort="pim">
PIM</th>, <th class="regular pm" data-sort="pm">+/-</th>, <th class="separator"> </th>, <t
h class="postseason">POST</th>, <th class="postseason gp" data-sort="playoffs-gp">GP</th>,
<th class="postseason g" data-sort="playoffs-g">G</th>, <th class="postseason a" data-sort
="playoffs-a">A</th>, <th class="postseason tp" data-sort="playoffs-tp">TP</th>, <th class
="postseason pim" data-sort="playoffs-pim">PIM</th>, <th class="postseason pm" data-sort
="playoffs-pm">+/-</th>]
```

In [74]:

```
column_names = [c.string for c in columns]
print(column_names)
```

```
['S', 'Team', 'League', 'GP', 'G', 'A', 'TP', 'PIM', '+/-', '\xa0', 'POST', 'GP', 'G',
'A', 'TP', 'PIM', '+/-']
```

In [76]:

```
table_rows = table.find('tbody').find_all('tr')
table_rows
```

Out[76]:

```
[<tr class="team-continent-NA">
 <td class="season sorted">
                  2014-15
              </td>
 <td class="team">
 <i><img src="images/flag.png"/></i>
 <span class="txt-blue">
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2014-2015?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>
 </span>
 </td>
```

```
 <td class="league"> <a href="https://www.eliteprospects.com/league/acha-ii/stats/2014-201
5"> ACHA II </a> </td>
 <td class="regular gp">17</td>
 <td class="regular g">3</td>
 <td class="regular a">9</td>
 <td class="regular tp">12</td>
 <td class="regular pim">20</td>
 <td class="regular pm"></td>
 <td class="separator"> | </td>
 <td class="postseason">
 <a href="https://www.eliteprospects.com/league/acha-ii/stats/2014-2015"> </a>
 </td>
 <td class="postseason gp">
 </td>
 <td class="postseason g">
 </td>
 <td class="postseason a">
 </td>
 <td class="postseason tp">
 </td>
 <td class="postseason pim">
 </td>
 <td class="postseason pm">
 </td>
 </tr>,
 <tr class="team-continent-NA">
 <td class="season sorted">
                  2015-16
              </td>
 <td class="team">
 <i><img src="images/flag.png"/></i>
 <span class="txt-blue">
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2015-2016?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>
 </span>
 </td>
 <td class="league"> <a href="https://www.eliteprospects.com/league/acha-ii/stats/2015-201
6"> ACHA II </a> </td>
 <td class="regular gp">9</td>
 <td class="regular g">1</td>
 <td class="regular a">1</td>
 <td class="regular tp">2</td>
 <td class="regular pim">2</td>
 <td class="regular pm"></td>
 <td class="separator"> | </td>
 <td class="postseason">
 <a href="https://www.eliteprospects.com/league/acha-ii/stats/2015-2016"> </a>
 </td>
 <td class="postseason gp">
 </td>
 <td class="postseason g">
 </td>
 <td class="postseason a">
 </td>
 <td class="postseason tp">
 </td>
 <td class="postseason pim">
 </td>
 <td class="postseason pm">
 </td>
 </tr>,
 <tr class="team-continent-NA">
 <td class="season sorted">
                  2016-17
              </td>
 <td class="team">
 <i><img src="images/flag.png"/></i>
 <span class="txt-blue">
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2016-2017?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>
```

```
 </span>
 </td>
 <td class="league"> <a href="https://www.eliteprospects.com/league/acha-ii/stats/2016-201
7"> ACHA II </a> </td>
 <td class="regular gp">12</td>
 <td class="regular g">5</td>
 <td class="regular a">5</td>
 <td class="regular tp">10</td>
 <td class="regular pim">8</td>
 <td class="regular pm">0</td>
 <td class="separator"> | </td>
 <td class="postseason">
 </td>
 <td class="postseason gp">
 </td>
 <td class="postseason g">
 </td>
 <td class="postseason a">
 </td>
 <td class="postseason tp">
 </td>
 <td class="postseason pim">
 </td>
 <td class="postseason pm">
 </td>
 </tr>,
 <tr class="team-continent-EU">
 <td class="season sorted">
                  2017-18
               </td>
 <td class="team">
                  Did not play
               </td>
 <td class="league"> <a href="https://www.eliteprospects.com/stats"> </a> </td>
 <td class="regular gp"></td>
 <td class="regular g"></td>
 <td class="regular a"></td>
 <td class="regular tp"></td>
 <td class="regular pim"></td>
 <td class="regular pm"></td>
 <td class="separator"> | </td>
 <td class="postseason">
 <a href="https://www.eliteprospects.com/stats"> </a>
 </td>
 <td class="postseason gp">
 </td>
 <td class="postseason g">
 </td>
 <td class="postseason a">
 </td>
 <td class="postseason tp">
 </td>
 <td class="postseason pim">
 </td>
 <td class="postseason pm">
 </td>
 </tr>,
 <tr class="team-continent-NA">
 <td class="season sorted">
                  2018-19
               </td>
 <td class="team">
 <i><img src="images/flag.png"/></i>
 <span class="txt-blue">
 <a href="https://www.eliteprospects.com/team/10263/mit-mass.-inst.-of-tech./2018-2019?tab
=stats"> MIT (Mass. Inst. of Tech.) </a>
 </span>
 </td>
 <td class="league"> <a href="https://www.eliteprospects.com/league/acha-iii/stats/2018-20
                  a> </td>
```

```
<td class="regular gp">8</td>
<td class="regular g">5</td>
<td class="regular a">10</td>
<td class="regular tp">15</td>
<td class="regular pim">8</td>
<td class="regular pm"></td>
<td class="separator"> | </td>
<td class="postseason">
<a href="https://www.eliteprospects.com/league/acha-iii/stats/2018-2019"> </a>
</td>
<td class="postseason gp">
</td>
<td class="postseason g">
</td>
<td class="postseason a">
</td>
<td class="postseason tp">
</td>
<td class="postseason pim">
</td>
<td class="postseason pm">
</td>
</tr>]
```

In [80]:

```
l =[]
for tr in table_rows:
    td =tr.find_all('td')
    row =[str(tr.get_text()).strip() for tr in td]
    l.append(row)
    print(l[0])
```

```
['2014-15', 'MIT (Mass. Inst. of Tech.)', 'ACHA II', '17', '3', '9', '12', '20', '', '|', '', '', '', '', '', '', '']
['2014-15', 'MIT (Mass. Inst. of Tech.)', 'ACHA II', '17', '3', '9', '12', '20', '', '|', '', '', '', '', '', '', '']
['2014-15', 'MIT (Mass. Inst. of Tech.)', 'ACHA II', '17', '3', '9', '12', '20', '', '|', '', '', '', '', '', '', '']
['2014-15', 'MIT (Mass. Inst. of Tech.)', 'ACHA II', '17', '3', '9', '12', '20', '', '|', '', '', '', '', '', '', '']
['2014-15', 'MIT (Mass. Inst. of Tech.)', 'ACHA II', '17', '3', '9', '12', '20', '', '|', '', '', '', '', '', '', '']
```

In [93]:

```
df = pd.DataFrame((0,1),columns=column_names)
df
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
c:\users\andre\appdata\local\programs\python\python39\lib\site-packages\pandas\core\internals\managers.py in create_block_manager_from_blocks(blocks, axes)
   1674                 blocks = [
-> 1675                     make_block(
   1676                         values=blocks[0], placement=slice(0, len(axes[0])), ndim=2

c:\users\andre\appdata\local\programs\python\python39\lib\site-packages\pandas\core\internals\blocks.py in make_block(values, placement, klass, ndim, dtype)
   2731
-> 2732     return klass(values, ndim=ndim, placement=placement)
   2733

c:\users\andre\appdata\local\programs\python\python39\lib\site-packages\pandas\core\internals\blocks.py in __init__(self, values, placement, ndim)
    141         if self._validate_ndim and self.ndim and len(self.mgr_locs) != len(self.values):
--> 142             raise ValueError(
    143                 f"Wrong number of items passed {len(self.values)}, "

Loading [MathJax]/extensions/Safe.js g number of items passed 1, placement implies 17
```

```
During handling of the above exception, another exception occurred:

ValueError                                Traceback (most recent call last)
<ipython-input-93-9402710deda4> in <module>
----> 1 df = pd.DataFrame((0,1),columns=column_names)
      2 df

c:\users\andre\appdata\local\programs\python\python39\lib\site-packages\pandas\core\frame.
py in __init__(self, data, index, columns, dtype, copy)
    582                     mgr = arrays_to_mgr(arrays, columns, index, columns, dtype=dty
pe)
    583                 else:
--> 584                     mgr = init_ndarray(data, index, columns, dtype=dtype, copy=cop
y)
    585             else:
    586                 mgr = init_dict({}, index, columns, dtype=dtype)

c:\users\andre\appdata\local\programs\python\python39\lib\site-packages\pandas\core\intern
als\construction.py in init_ndarray(values, index, columns, dtype, copy)
    236         block_values = [values]
    237
--> 238     return create_block_manager_from_blocks(block_values, [columns, index])
    239
    240

c:\users\andre\appdata\local\programs\python\python39\lib\site-packages\pandas\core\intern
als\managers.py in create_block_manager_from_blocks(blocks, axes)
   1685         blocks = [getattr(b, "values", b) for b in blocks]
   1686         tot_items = sum(b.shape[0] for b in blocks)
-> 1687         raise construction_error(tot_items, blocks[0].shape[1:], axes, e)
   1688
   1689

ValueError: Shape of passed values is (2, 1), indices imply (2, 17)
```

In [ ]: