WebScrape_Cosmetics

September 24, 2023

1 Web Scraping using BeautifulSoup

1.1 Import Libraries

[ ]:
```
from bs4 import BeautifulSoup
import requests
import pandas as pd
```

1.2 Read webpage

Top 100 buyers of beauty and related skincare imports. Ranked in descending order.

2020-1: change in international sales from 2020 to 2021.

%Total: percentage of the world's total purchases of imported cosmetics in 2021

US$: Annual sales amount

[ ]:
```
url = 'https://importedconsumerproducts.com/beauty-cosmetics/'

page = requests.get(url)

page
```

[ ]:
```
<Response [200]>
```

1.3 Load webpage content

[ ]:
```
soup = BeautifulSoup(page.content, 'html.parser')
```

[ ]:
```
table = soup.find_all("table", class_ = "tablepress tablepress-id-39")
table
```

[ ]:
```
[<table class="tablepress tablepress-id-39" id="tablepress-39">
 <thead>
 <tr class="row-1 odd">
 <th class="column-1">Rank</th><th class="column-2">Importer</th><th
class="column-3">US$</th><th class="column-4">%Total</th><th
class="column-5">2020-1</th>
```

```
 </tr>
 </thead>
 <tbody class="row-hover">
 <tr class="row-2 even">
 <td class="column-1">1</td><td class="column-2">China</td><td
class="column-3">$20,287,362,000</td><td class="column-4">27.2%</td><td
class="column-5">+17.1%</td>
 </tr>
 <tr class="row-3 odd">
 <td class="column-1">2</td><td class="column-2">Hong Kong</td><td
class="column-3">$7,280,376,000</td><td class="column-4">9.8%</td><td
class="column-5">+0.6%</td>
 </tr>
 <tr class="row-4 even">
 <td class="column-1">3</td><td class="column-2">United States</td><td
class="column-3">$5,557,871,000</td><td class="column-4">7.5%</td><td
class="column-5">+23.9%</td>
 </tr>
 <tr class="row-5 odd">
 <td class="column-1">4</td><td class="column-2">Macao</td><td
class="column-3">$3,221,960,000</td><td class="column-4">4.3%</td><td
class="column-5">+36.6%</td>
 </tr>
 <tr class="row-6 even">
 <td class="column-1">5</td><td class="column-2">Singapore</td><td
class="column-3">$3,087,914,000</td><td class="column-4">4.1%</td><td
class="column-5">+6.7%</td>
 </tr>
 <tr class="row-7 odd">
 <td class="column-1">6</td><td class="column-2">Germany</td><td
class="column-3">$2,629,789,000</td><td class="column-4">3.5%</td><td
class="column-5">+17%</td>
 </tr>
 <tr class="row-8 even">
 <td class="column-1">7</td><td class="column-2">United Kingdom</td><td
class="column-3">$2,082,046,000</td><td class="column-4">2.8%</td><td
class="column-5">+2.6%</td>
 </tr>
 <tr class="row-9 odd">
 <td class="column-1">8</td><td class="column-2">France</td><td
class="column-3">$1,906,145,000</td><td class="column-4">2.6%</td><td
class="column-5">+11.4%</td>
 </tr>
 <tr class="row-10 even">
 <td class="column-1">9</td><td class="column-2">Canada</td><td
class="column-3">$1,575,771,000</td><td class="column-4">2.1%</td><td
class="column-5">+11.9%</td>
```

```
 </tr>
 <tr class="row-11 odd">
 <td class="column-1">10</td><td class="column-2">Japan</td><td
class="column-3">$1,466,193,000</td><td class="column-4">2.0%</td><td
class="column-5">+5.6%</td>
 </tr>
 <tr class="row-12 even">
 <td class="column-1">11</td><td class="column-2">Belgium</td><td
class="column-3">$1,395,477,000</td><td class="column-4">1.9%</td><td
class="column-5">+7.8%</td>
 </tr>
 <tr class="row-13 odd">
 <td class="column-1">12</td><td class="column-2">Netherlands</td><td
class="column-3">$1,326,661,000</td><td class="column-4">1.8%</td><td
class="column-5">+33.5%</td>
 </tr>
 <tr class="row-14 even">
 <td class="column-1">13</td><td class="column-2">Russia</td><td
class="column-3">$1,291,515,000</td><td class="column-4">1.7%</td><td
class="column-5">+27.2%</td>
 </tr>
 <tr class="row-15 odd">
 <td class="column-1">14</td><td class="column-2">Poland</td><td
class="column-3">$1,171,673,000</td><td class="column-4">1.6%</td><td
class="column-5">+14.8%</td>
 </tr>
 <tr class="row-16 even">
 <td class="column-1">15</td><td class="column-2">Italy</td><td
class="column-3">$1,159,766,000</td><td class="column-4">1.6%</td><td
class="column-5">+12%</td>
 </tr>
 <tr class="row-17 odd">
 <td class="column-1">16</td><td class="column-2">Spain</td><td
class="column-3">$1,149,378,000</td><td class="column-4">1.5%</td><td
class="column-5">+18.6%</td>
 </tr>
 <tr class="row-18 even">
 <td class="column-1">17</td><td class="column-2">UAE</td><td
class="column-3">$966,135,000</td><td class="column-4">1.3%</td><td
class="column-5">+21.2%</td>
 </tr>
 <tr class="row-19 odd">
 <td class="column-1">18</td><td class="column-2">South Korea</td><td
class="column-3">$910,848,000</td><td class="column-4">1.2%</td><td
class="column-5">+3.1%</td>
 </tr>
 <tr class="row-20 even">
```

```
 <td class="column-1">19</td><td class="column-2">Czech Republic</td><td
class="column-3">$906,189,000</td><td class="column-4">1.2%</td><td
class="column-5">+10.4%</td>
 </tr>
 <tr class="row-21 odd">
 <td class="column-1">20</td><td class="column-2">Australia</td><td
class="column-3">$883,136,000</td><td class="column-4">1.2%</td><td
class="column-5">+12%</td>
 </tr>
 <tr class="row-22 even">
 <td class="column-1">21</td><td class="column-2">Taiwan</td><td
class="column-3">$867,511,000</td><td class="column-4">1.2%</td><td
class="column-5">+1.9%</td>
 </tr>
 <tr class="row-23 odd">
 <td class="column-1">22</td><td class="column-2">Mexico</td><td
class="column-3">$622,153,000</td><td class="column-4">0.8%</td><td
class="column-5">+28.3%</td>
 </tr>
 <tr class="row-24 even">
 <td class="column-1">23</td><td class="column-2">Saudi Arabia</td><td
class="column-3">$620,513,000</td><td class="column-4">0.8%</td><td
class="column-5">+4%</td>
 </tr>
 <tr class="row-25 odd">
 <td class="column-1">24</td><td class="column-2">Switzerland</td><td
class="column-3">$613,775,000</td><td class="column-4">0.8%</td><td
class="column-5">+15.5%</td>
 </tr>
 <tr class="row-26 even">
 <td class="column-1">25</td><td class="column-2">Thailand</td><td
class="column-3">$526,418,000</td><td class="column-4">0.7%</td><td
class="column-5">-14.4%</td>
 </tr>
 <tr class="row-27 odd">
 <td class="column-1">26</td><td class="column-2">Sweden</td><td
class="column-3">$524,211,000</td><td class="column-4">0.7%</td><td
class="column-5">+19.9%</td>
 </tr>
 <tr class="row-28 even">
 <td class="column-1">27</td><td class="column-2">Austria</td><td
class="column-3">$507,869,000</td><td class="column-4">0.7%</td><td
class="column-5">+17.5%</td>
 </tr>
 <tr class="row-29 odd">
 <td class="column-1">28</td><td class="column-2">Ireland</td><td
class="column-3">$450,083,000</td><td class="column-4">0.6%</td><td
```

```
class="column-5">+49.6%</td>
 </tr>
 <tr class="row-30 even">
 <td class="column-1">29</td><td class="column-2">Denmark</td><td
class="column-3">$409,357,000</td><td class="column-4">0.5%</td><td
class="column-5">+29.6%</td>
 </tr>
 <tr class="row-31 odd">
 <td class="column-1">30</td><td class="column-2">Norway</td><td
class="column-3">$393,447,000</td><td class="column-4">0.5%</td><td
class="column-5">+30.6%</td>
 </tr>
 <tr class="row-32 even">
 <td class="column-1">31</td><td class="column-2">Malaysia</td><td
class="column-3">$384,600,000</td><td class="column-4">0.5%</td><td
class="column-5">+10.1%</td>
 </tr>
 <tr class="row-33 odd">
 <td class="column-1">32</td><td class="column-2">Indonesia</td><td
class="column-3">$309,480,000</td><td class="column-4">0.4%</td><td
class="column-5">+20.4%</td>
 </tr>
 <tr class="row-34 even">
 <td class="column-1">33</td><td class="column-2">Chile</td><td
class="column-3">$307,677,000</td><td class="column-4">0.4%</td><td
class="column-5">+71.5%</td>
 </tr>
 <tr class="row-35 odd">
 <td class="column-1">34</td><td class="column-2">Kuwait</td><td
class="column-3">$303,105,000</td><td class="column-4">0.4%</td><td
class="column-5">+18.9%</td>
 </tr>
 <tr class="row-36 even">
 <td class="column-1">35</td><td class="column-2">Turkey</td><td
class="column-3">$292,394,000</td><td class="column-4">0.4%</td><td
class="column-5">+14.4%</td>
 </tr>
 <tr class="row-37 odd">
 <td class="column-1">36</td><td class="column-2">Romania</td><td
class="column-3">$269,253,000</td><td class="column-4">0.4%</td><td
class="column-5">+16.8%</td>
 </tr>
 <tr class="row-38 even">
 <td class="column-1">37</td><td class="column-2">Ukraine</td><td
class="column-3">$265,571,000</td><td class="column-4">0.4%</td><td
class="column-5">+25.6%</td>
 </tr>
```

```
 <tr class="row-39 odd">
 <td class="column-1">38</td><td class="column-2">India</td><td
class="column-3">$262,646,000</td><td class="column-4">0.4%</td><td
class="column-5">+90%</td>
 </tr>
 <tr class="row-40 even">
 <td class="column-1">39</td><td class="column-2">Portugal</td><td
class="column-3">$261,975,000</td><td class="column-4">0.4%</td><td
class="column-5">+15.6%</td>
 </tr>
 <tr class="row-41 odd">
 <td class="column-1">40</td><td class="column-2">Hungary</td><td
class="column-3">$244,499,000</td><td class="column-4">0.3%</td><td
class="column-5">+6.2%</td>
 </tr>
 <tr class="row-42 even">
 <td class="column-1">41</td><td class="column-2">Vietnam</td><td
class="column-3">$236,350,000</td><td class="column-4">0.3%</td><td
class="column-5">+34.6%</td>
 </tr>
 <tr class="row-43 odd">
 <td class="column-1">42</td><td class="column-2">New Zealand</td><td
class="column-3">$231,220,000</td><td class="column-4">0.3%</td><td
class="column-5">+18.9%</td>
 </tr>
 <tr class="row-44 even">
 <td class="column-1">43</td><td class="column-2">Slovakia</td><td
class="column-3">$228,795,000</td><td class="column-4">0.3%</td><td
class="column-5">+26.2%</td>
 </tr>
 <tr class="row-45 odd">
 <td class="column-1">44</td><td class="column-2">Lithuania</td><td
class="column-3">$228,762,000</td><td class="column-4">0.3%</td><td
class="column-5">+16.9%</td>
 </tr>
 <tr class="row-46 even">
 <td class="column-1">45</td><td class="column-2">Greece</td><td
class="column-3">$222,781,000</td><td class="column-4">0.3%</td><td
class="column-5">+4.1%</td>
 </tr>
 <tr class="row-47 odd">
 <td class="column-1">46</td><td class="column-2">South Africa</td><td
class="column-3">$193,885,000</td><td class="column-4">0.3%</td><td
class="column-5">+25.3%</td>
 </tr>
 <tr class="row-48 even">
 <td class="column-1">47</td><td class="column-2">Croatia</td><td
```

```
class="column-3">$177,886,000</td><td class="column-4">0.2%</td><td
class="column-5">+16.3%</td>
 </tr>
 <tr class="row-49 odd">
 <td class="column-1">48</td><td class="column-2">Brazil</td><td
class="column-3">$172,449,000</td><td class="column-4">0.2%</td><td
class="column-5">+5.7%</td>
 </tr>
 <tr class="row-50 even">
 <td class="column-1">49</td><td class="column-2">Israel</td><td
class="column-3">$163,768,000</td><td class="column-4">0.2%</td><td
class="column-5">+31.5%</td>
 </tr>
 <tr class="row-51 odd">
 <td class="column-1">50</td><td class="column-2">Colombia</td><td
class="column-3">$156,756,000</td><td class="column-4">0.2%</td><td
class="column-5">+26.4%</td>
 </tr><tr><td colspan="5"><span class="ezoic-autoinsert-video ezoic-
longer_content"></span><!-- ezoic_video_placeholder-
longer_content-640x360-999994-clearholder --><!-- ezoic_video_placeholder-
longer_content-640x360-999994-nonexxxclearxxxblock --><!--
ezoic_video_placeholder-longer_content-426x240-999994-clearholder --><!--
ezoic_video_placeholder-longer_content-426x240-999994-nonexxxclearxxxblock
--><!-- ezoic_video_placeholder-longer_content-384x216-999994-clearholder
--><!-- ezoic_video_placeholder-
longer_content-384x216-999994-nonexxxclearxxxblock --></td></tr>
 <tr class="row-52 even">
 <td class="column-1">51</td><td class="column-2">Finland</td><td
class="column-3">$150,173,000</td><td class="column-4">0.2%</td><td
class="column-5">+9.5%</td>
 </tr>
 <tr class="row-53 odd">
 <td class="column-1">52</td><td class="column-2">Kazakhstan</td><td
class="column-3">$149,140,000</td><td class="column-4">0.2%</td><td
class="column-5">+32.8%</td>
 </tr>
 <tr class="row-54 even">
 <td class="column-1">53</td><td class="column-2">Iraq</td><td
class="column-3">$146,273,000</td><td class="column-4">0.2%</td><td
class="column-5">+20%</td>
 </tr>
 <tr class="row-55 odd">
 <td class="column-1">54</td><td class="column-2">Philippines</td><td
class="column-3">$139,259,000</td><td class="column-4">0.2%</td><td
class="column-5">+7.2%</td>
 </tr>
 <tr class="row-56 even">
```

```
 <td class="column-1">55</td><td class="column-2">Qatar</td><td
class="column-3">$139,229,000</td><td class="column-4">0.2%</td><td
class="column-5">+18.4%</td>
 </tr>
 <tr class="row-57 odd">
 <td class="column-1">56</td><td class="column-2">Peru</td><td
class="column-3">$131,957,000</td><td class="column-4">0.2%</td><td
class="column-5">+40.9%</td>
 </tr>
 <tr class="row-58 even">
 <td class="column-1">57</td><td class="column-2">Argentina</td><td
class="column-3">$113,013,000</td><td class="column-4">0.2%</td><td
class="column-5">+15.2%</td>
 </tr>
 <tr class="row-59 odd">
 <td class="column-1">58</td><td class="column-2">Slovenia</td><td
class="column-3">$109,651,000</td><td class="column-4">0.1%</td><td
class="column-5">+18.2%</td>
 </tr>
 <tr class="row-60 even">
 <td class="column-1">59</td><td class="column-2">Morocco</td><td
class="column-3">$108,699,000</td><td class="column-4">0.1%</td><td
class="column-5">+24.3%</td>
 </tr>
 <tr class="row-61 odd">
 <td class="column-1">60</td><td class="column-2">Bulgaria</td><td
class="column-3">$105,585,000</td><td class="column-4">0.1%</td><td
class="column-5">+31.5%</td>
 </tr>
 <tr class="row-62 even">
 <td class="column-1">61</td><td class="column-2">Panama</td><td
class="column-3">$101,043,000</td><td class="column-4">0.1%</td><td
class="column-5">+11.6%</td>
 </tr>
 <tr class="row-63 odd">
 <td class="column-1">62</td><td class="column-2">Oman</td><td
class="column-3">$92,874,000</td><td class="column-4">0.1%</td><td
class="column-5">+11.5%</td>
 </tr>
 <tr class="row-64 even">
 <td class="column-1">63</td><td class="column-2">Ecuador</td><td
class="column-3">$88,525,000</td><td class="column-4">0.1%</td><td
class="column-5">+28.3%</td>
 </tr>
 <tr class="row-65 odd">
 <td class="column-1">64</td><td class="column-2">Belarus</td><td
class="column-3">$88,364,000</td><td class="column-4">0.1%</td><td
```

```
class="column-5">+10.1%</td>
 </tr>
 <tr class="row-66 even">
 <td class="column-1">65</td><td class="column-2">Luxembourg</td><td
class="column-3">$81,925,000</td><td class="column-4">0.1%</td><td
class="column-5">+6.8%</td>
 </tr>
 <tr class="row-67 odd">
 <td class="column-1">66</td><td class="column-2">Serbia</td><td
class="column-3">$74,904,000</td><td class="column-4">0.1%</td><td
class="column-5">+10.5%</td>
 </tr>
 <tr class="row-68 even">
 <td class="column-1">67</td><td class="column-2">Latvia</td><td
class="column-3">$70,928,000</td><td class="column-4">0.1%</td><td
class="column-5">+3.1%</td>
 </tr>
 <tr class="row-69 odd">
 <td class="column-1">68</td><td class="column-2">Estonia</td><td
class="column-3">$70,927,000</td><td class="column-4">0.1%</td><td
class="column-5">-0.9%</td>
 </tr>
 <tr class="row-70 even">
 <td class="column-1">69</td><td class="column-2">Guatemala</td><td
class="column-3">$64,953,000</td><td class="column-4">0.1%</td><td
class="column-5">+19%</td>
 </tr>
 <tr class="row-71 odd">
 <td class="column-1">70</td><td class="column-2">Cyprus</td><td
class="column-3">$61,739,000</td><td class="column-4">0.1%</td><td
class="column-5">+23.3%</td>
 </tr>
 <tr class="row-72 even">
 <td class="column-1">71</td><td class="column-2">Costa Rica</td><td
class="column-3">$58,507,000</td><td class="column-4">0.1%</td><td
class="column-5">+12.8%</td>
 </tr>
 <tr class="row-73 odd">
 <td class="column-1">72</td><td class="column-2">Jordan</td><td
class="column-3">$57,801,000</td><td class="column-4">0.1%</td><td
class="column-5">+18.1%</td>
 </tr>
 <tr class="row-74 even">
 <td class="column-1">73</td><td class="column-2">Bosnia/Herzegovina</td><td
class="column-3">$50,201,000</td><td class="column-4">0.1%</td><td
class="column-5">+26.2%</td>
 </tr>
```

```
 <tr class="row-75 odd">
 <td class="column-1">74</td><td class="column-2">Algeria</td><td
class="column-3">$48,641,000</td><td class="column-4">0.1%</td><td
class="column-5">-18.9%</td>
 </tr>
 <tr class="row-76 even">
 <td class="column-1">75</td><td class="column-2">Egypt</td><td
class="column-3">$48,275,000</td><td class="column-4">0.1%</td><td
class="column-5">+30.9%</td>
 </tr>
 <tr class="row-77 odd">
 <td class="column-1">76</td><td class="column-2">Dominican Republic</td><td
class="column-3">$47,039,000</td><td class="column-4">0.1%</td><td
class="column-5">+14.1%</td>
 </tr>
 <tr class="row-78 even">
 <td class="column-1">77</td><td class="column-2">Bangladesh</td><td
class="column-3">$44,628,000</td><td class="column-4">0.1%</td><td
class="column-5">+43.5%</td>
 </tr>
 <tr class="row-79 odd">
 <td class="column-1">78</td><td class="column-2">Bahrain</td><td
class="column-3">$44,586,000</td><td class="column-4">0.1%</td><td
class="column-5">+15.5%</td>
 </tr>
 <tr class="row-80 even">
 <td class="column-1">79</td><td class="column-2">El Salvador</td><td
class="column-3">$44,416,000</td><td class="column-4">0.06%</td><td
class="column-5">+42.6%</td>
 </tr>
 <tr class="row-81 odd">
 <td class="column-1">80</td><td class="column-2">Nepal</td><td
class="column-3">$43,439,000</td><td class="column-4">0.06%</td><td
class="column-5">+67.8%</td>
 </tr>
 <tr class="row-82 even">
 <td class="column-1">81</td><td class="column-2">Libya</td><td
class="column-3">$40,699,000</td><td class="column-4">0.05%</td><td
class="column-5">+41%</td>
 </tr>
 <tr class="row-83 odd">
 <td class="column-1">82</td><td class="column-2">Iran</td><td
class="column-3">$40,565,000</td><td class="column-4">0.05%</td><td
class="column-5">+183.8%</td>
 </tr>
 <tr class="row-84 even">
 <td class="column-1">83</td><td class="column-2">Georgia</td><td
```

```
class="column-3">$40,256,000</td><td class="column-4">0.05%</td><td
class="column-5">+27%</td>
 </tr>
 <tr class="row-85 odd">
 <td class="column-1">84</td><td class="column-2">Azerbaijan</td><td
class="column-3">$36,028,000</td><td class="column-4">0.05%</td><td
class="column-5">+37.2%</td>
 </tr>
 <tr class="row-86 even">
 <td class="column-1">85</td><td class="column-2">Andorra</td><td
class="column-3">$35,707,000</td><td class="column-4">0.05%</td><td
class="column-5">+48%</td>
 </tr>
 <tr class="row-87 odd">
 <td class="column-1">86</td><td class="column-2">Pakistan</td><td
class="column-3">$31,191,000</td><td class="column-4">0.04%</td><td
class="column-5">+60.3%</td>
 </tr>
 <tr class="row-88 even">
 <td class="column-1">87</td><td class="column-2">Namibia</td><td
class="column-3">$31,189,000</td><td class="column-4">0.04%</td><td
class="column-5">+6.3%</td>
 </tr>
 <tr class="row-89 odd">
 <td class="column-1">88</td><td class="column-2">Botswana</td><td
class="column-3">$30,939,000</td><td class="column-4">0.04%</td><td
class="column-5">+11.8%</td>
 </tr>
 <tr class="row-90 even">
 <td class="column-1">89</td><td class="column-2">Lebanon</td><td
class="column-3">$30,584,000</td><td class="column-4">0.04%</td><td
class="column-5">+16.3%</td>
 </tr>
 <tr class="row-91 odd">
 <td class="column-1">90</td><td class="column-2">Iceland</td><td
class="column-3">$29,730,000</td><td class="column-4">0.04%</td><td
class="column-5">+26%</td>
 </tr>
 <tr class="row-92 even">
 <td class="column-1">91</td><td class="column-2">Myanmar</td><td
class="column-3">$29,071,000</td><td class="column-4">0.04%</td><td
class="column-5">-28.1%</td>
 </tr>
 <tr class="row-93 odd">
 <td class="column-1">92</td><td class="column-2">Moldova</td><td
class="column-3">$28,335,000</td><td class="column-4">0.04%</td><td
class="column-5">+36.5%</td>
```

```
 </tr>
 <tr class="row-94 even">
 <td class="column-1">93</td><td class="column-2">Tunisia</td><td
class="column-3">$28,070,000</td><td class="column-4">0.04%</td><td
class="column-5">-14.1%</td>
 </tr>
 <tr class="row-95 odd">
 <td class="column-1">94</td><td class="column-2">Bolivia</td><td
class="column-3">$27,549,000</td><td class="column-4">0.04%</td><td
class="column-5">+16%</td>
 </tr>
 <tr class="row-96 even">
 <td class="column-1">95</td><td class="column-2">Malta</td><td
class="column-3">$25,571,000</td><td class="column-4">0.03%</td><td
class="column-5">+93%</td>
 </tr>
 <tr class="row-97 odd">
 <td class="column-1">96</td><td class="column-2">Mongolia</td><td
class="column-3">$24,861,000</td><td class="column-4">0.03%</td><td
class="column-5">+46.2%</td>
 </tr>
 <tr class="row-98 even">
 <td class="column-1">97</td><td class="column-2">Honduras</td><td
class="column-3">$23,818,000</td><td class="column-4">0.03%</td><td
class="column-5">+37%</td>
 </tr>
 <tr class="row-99 odd">
 <td class="column-1">98</td><td class="column-2">Albania</td><td
class="column-3">$23,726,000</td><td class="column-4">0.03%</td><td
class="column-5">+26%</td>
 </tr>
 <tr class="row-100 even">
 <td class="column-1">99</td><td class="column-2">North Macedonia</td><td
class="column-3">$22,232,000</td><td class="column-4">0.03%</td><td
class="column-5">+31%</td>
 </tr><tr><td colspan="5"><span class="ezoic-autoinsert-video ezoic-
longest_content"></span><!-- ezoic_video_placeholder-
longest_content-640x360-999993-clearholder --><!-- ezoic_video_placeholder-
longest_content-640x360-999993-nonexxxclearxxxblock --><!--
ezoic_video_placeholder-longest_content-426x240-999993-clearholder --><!--
ezoic_video_placeholder-longest_content-426x240-999993-nonexxxclearxxxblock
--><!-- ezoic_video_placeholder-longest_content-384x216-999993-clearholder
--><!-- ezoic_video_placeholder-
longest_content-384x216-999993-nonexxxclearxxxblock --></td></tr>
 <tr class="row-101 odd">
 <td class="column-1">100</td><td class="column-2">Paraguay</td><td
class="column-3">$20,871,000</td><td class="column-4">0.03%</td><td
```

```
class="column-5">+15.9%</td>
 </tr>
 </tbody>
 </table>]
```

1.4 Read column headings

[ ]:
```
x =soup.find_all("th")
x
```

[ ]:
```
[<th class="column-1">Rank</th>,
 <th class="column-2">Importer</th>,
 <th class="column-3">US$</th>,
 <th class="column-4">%Total</th>,
 <th class="column-5">2020-1</th>]
```

[ ]:
```
column_nm =  [title.text.strip() for title in x]
column_nm
```

[ ]:
```
['Rank', 'Importer', 'US$', '%Total', '2020-1']
```

1.5 Create dataframe and load column names

[ ]:
```
skincare=[]
skincare = pd.DataFrame(columns=column_nm)
skincare
```

[ ]:
```
Empty DataFrame
Columns: [Rank, Importer, US$, %Total, 2020-1]
Index: []
```

[ ]:
```
skincare.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 0 entries
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype 
---  ------    --------------  ----- 
 0   Rank      0 non-null      object
 1   Importer  0 non-null      object
 2   US$       0 non-null      object
 3   %Total    0 non-null      object
 4   2020-1    0 non-null      object
dtypes: object(5)
memory usage: 0.0+ bytes
```

[1]:
```
#skincare.drop(columns='', inplace=True)
```

1.6 Process HTML table to populate dataframe with rows of data

[ ]:
```
# Create a loop to read row data
# Look for tag "tr" then look for tag "td"
#  HTML table has two rows where the length is zero.  Add code to skip those␣
 ↪rows.
table_column = soup.tbody.find_all('tr')
for row in table_column:
    # Find all data for each column
    row_data = row.find_all('td', class_␣
 ↪=['column-1','column-2','column-3','column-4','column-5'])
    individual_row_data = [data.text.strip() for data in row_data]
    if len(individual_row_data) < 5:
      pass
    else:
       #print(individual_row_data, "len: ", len(individual_row_data))
       length = len(skincare)
       skincare.loc[length] = individual_row_data
```

1.7 Display populated dataframe

[ ]:
```
skincare
```

[ ]:

	Rank	Importer	US$	%Total	2020-1
0	1	China	$20,287,362,000	27.2%	+17.1%
1	2	Hong Kong	$7,280,376,000	9.8%	+0.6%
2	3	United States	$5,557,871,000	7.5%	+23.9%
3	4	Macao	$3,221,960,000	4.3%	+36.6%
4	5	Singapore	$3,087,914,000	4.1%	+6.7%
..	…	…	…	…	…
95	96	Mongolia	$24,861,000	0.03%	+46.2%
96	97	Honduras	$23,818,000	0.03%	+37%
97	98	Albania	$23,726,000	0.03%	+26%
98	99	North Macedonia	$22,232,000	0.03%	+31%
99	100	Paraguay	$20,871,000	0.03%	+15.9%

[100 rows x 5 columns]

1.8 Data Cleaning

[ ]:
```
skincare.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 100 entries, 0 to 99
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
```

```
 0   Rank      100 non-null    object
 1   Importer  100 non-null    object
 2   US$       100 non-null    object
 3   %Total    100 non-null    object
 4   2020-1    100 non-null    object
dtypes: object(5)
memory usage: 4.7+ KB
```

1.8.1 Remove % from '%Total' column

[ ]:
```
skincare['%Total']=skincare['%Total'].map(lambda x: x.rstrip('%'))
skincare
```

[ ]:

	Rank	Importer	US$	%Total	2020-1
0	1	China	$20,287,362,000	27.2	+17.1%
1	2	Hong Kong	$7,280,376,000	9.8	+0.6%
2	3	United States	$5,557,871,000	7.5	+23.9%
3	4	Macao	$3,221,960,000	4.3	+36.6%
4	5	Singapore	$3,087,914,000	4.1	+6.7%
..	…	…	…	…	…
95	96	Mongolia	$24,861,000	0.03	+46.2%
96	97	Honduras	$23,818,000	0.03	+37%
97	98	Albania	$23,726,000	0.03	+26%
98	99	North Macedonia	$22,232,000	0.03	+31%
99	100	Paraguay	$20,871,000	0.03	+15.9%

[100 rows x 5 columns]

1.8.2 Remove dollar sign from 'US$' column

[ ]:
```
skincare['US$']=skincare['US$'].map(lambda x: x.lstrip('$'))
skincare
```

[ ]:

	Rank	Importer	US$	%Total	2020-1
0	1	China	20,287,362,000	27.2	+17.1%
1	2	Hong Kong	7,280,376,000	9.8	+0.6%
2	3	United States	5,557,871,000	7.5	+23.9%
3	4	Macao	3,221,960,000	4.3	+36.6%
4	5	Singapore	3,087,914,000	4.1	+6.7%
..	…	…	…	…	…
95	96	Mongolia	24,861,000	0.03	+46.2%
96	97	Honduras	23,818,000	0.03	+37%
97	98	Albania	23,726,000	0.03	+26%
98	99	North Macedonia	22,232,000	0.03	+31%
99	100	Paraguay	20,871,000	0.03	+15.9%

[100 rows x 5 columns]

1.8.3 Remove % from '2020-1' column

[ ]:
```
skincare['2020-1']=skincare['2020-1'].map(lambda x: x.rstrip('%'))
skincare
```

[ ]:

	Rank	Importer	US$	%Total	2020-1
0	1	China	20,287,362,000	27.2	+17.1
1	2	Hong Kong	7,280,376,000	9.8	+0.6
2	3	United States	5,557,871,000	7.5	+23.9
3	4	Macao	3,221,960,000	4.3	+36.6
4	5	Singapore	3,087,914,000	4.1	+6.7
..	…	…	…	…	…
95	96	Mongolia	24,861,000	0.03	+46.2
96	97	Honduras	23,818,000	0.03	+37
97	98	Albania	23,726,000	0.03	+26
98	99	North Macedonia	22,232,000	0.03	+31
99	100	Paraguay	20,871,000	0.03	+15.9

[100 rows x 5 columns]

[ ]:
```
skincare.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 100 entries, 0 to 99
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Rank      100 non-null    object
 1   Importer  100 non-null    object
 2   US$       100 non-null    object
 3   %Total    100 non-null    object
 4   2020-1    100 non-null    object
dtypes: object(5)
memory usage: 4.7+ KB
```

[ ]:
```
skincare2 = skincare.copy(deep=True)
skincare2
```

[ ]:

	Rank	Importer	US$	%Total	2020-1
0	1	China	20,287,362,000	27.2	+17.1
1	2	Hong Kong	7,280,376,000	9.8	+0.6
2	3	United States	5,557,871,000	7.5	+23.9
3	4	Macao	3,221,960,000	4.3	+36.6
4	5	Singapore	3,087,914,000	4.1	+6.7
..	…	…	…	…	…
95	96	Mongolia	24,861,000	0.03	+46.2
96	97	Honduras	23,818,000	0.03	+37
97	98	Albania	23,726,000	0.03	+26

```
98   99  North Macedonia      22,232,000   0.03    +31
99  100         Paraguay      20,871,000   0.03  +15.9

[100 rows x 5 columns]
```

1.8.4 Convert data type Object to String or Numeric depending on column

[191]:
```
skincare2['%Total'] = pd.to_numeric(skincare2['%Total'], downcast='float')
```

[192]:
```
skincare2['2020-1'] = pd.to_numeric(skincare2['2020-1'], downcast='signed')
```

[193]:
```
skincare2['US$'] = pd.to_numeric(skincare2['US$'].str.replace('[^0-9.]', ''))
```

```
<ipython-input-193-d463c7e98e4e>:1: FutureWarning: The default value of regex
will change from True to False in a future version.
  skincare2['US$'] = pd.to_numeric(skincare2['US$'].str.replace('[^0-9.]', ''))
```

[194]:
```
skincare2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 100 entries, 0 to 99
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Rank      100 non-null    object
 1   Importer  100 non-null    object
 2   US$       100 non-null    int64
 3   %Total    100 non-null    float32
 4   2020-1    100 non-null    float64
dtypes: float32(1), float64(1), int64(1), object(2)
memory usage: 4.3+ KB
```

[212]:
```
skincare2['Rank'] = pd.to_numeric(skincare2['Rank'])
```

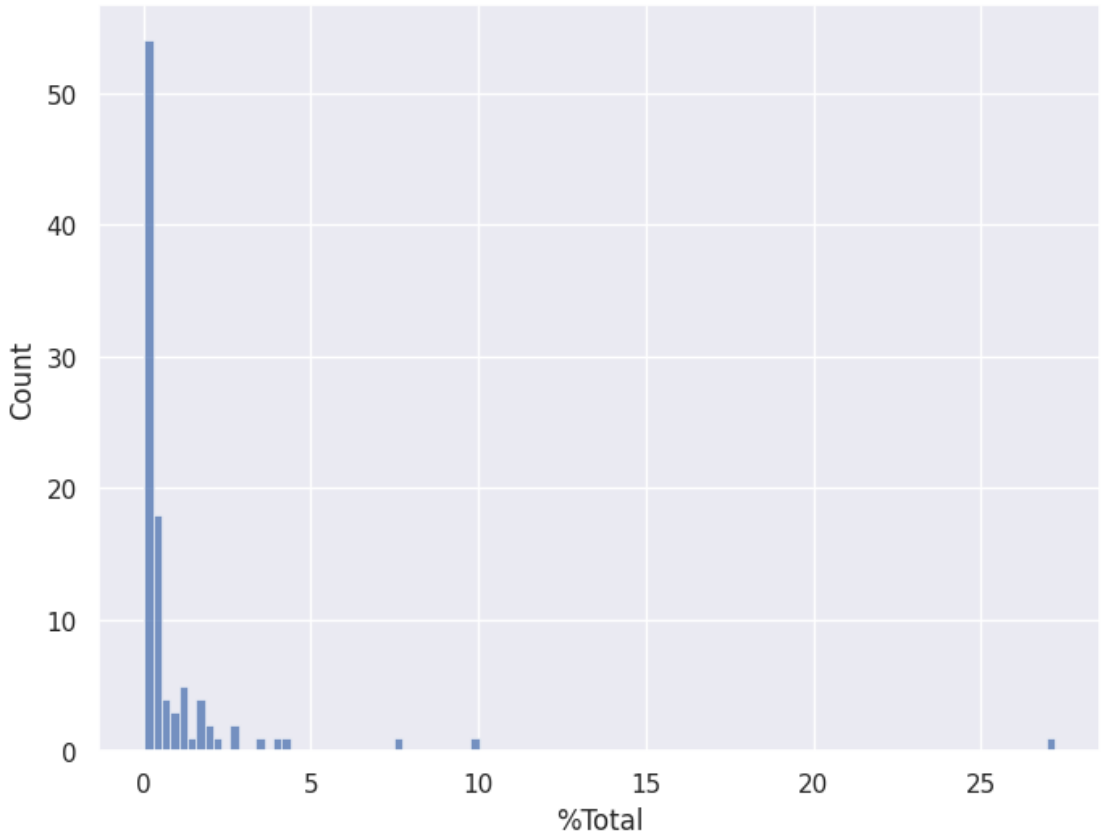
[203]:
```
skincare2['Importer'] = skincare2['Importer'].astype("string")
```

1.9 Data Visualization

[2]:
```
import seaborn as sns
import matplotlib.pyplot as plt
plt.style.use('ggplot')
```

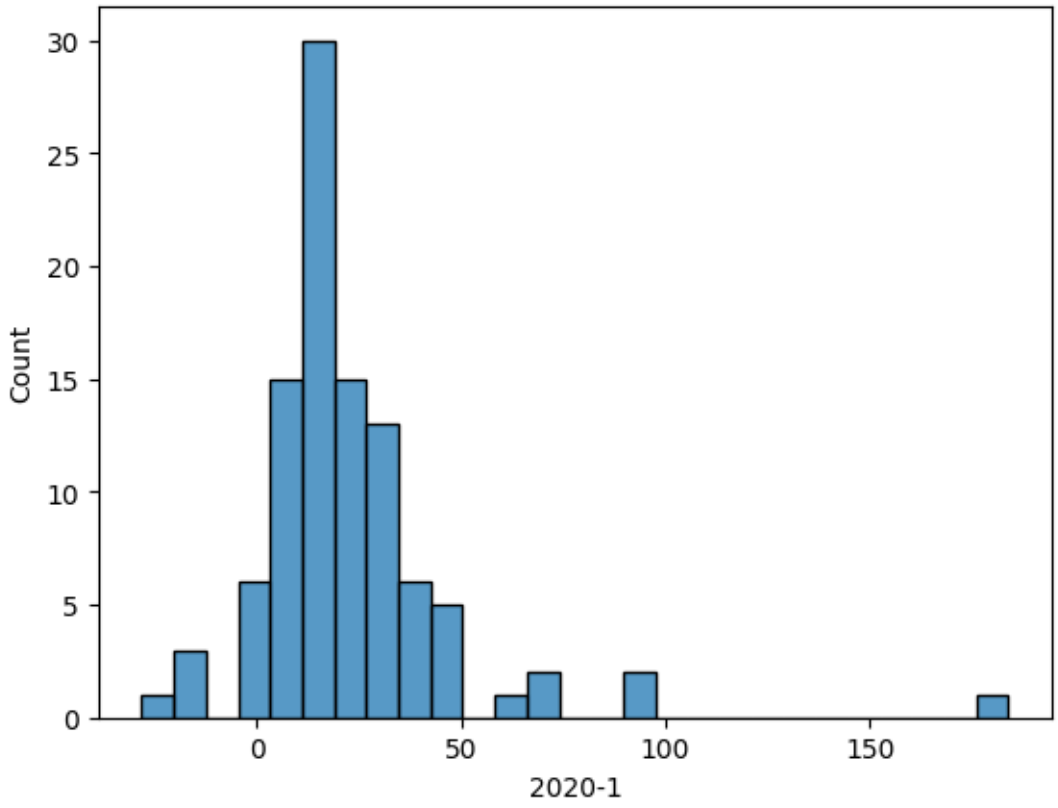
[231]:
```
#sns.set(rc={"figure.figsize":(8,6)})

sns.histplot(skincare2['%Total'])
plt.show()
```



[221]:
```
sns.histplot(skincare2['2020-1'])
```

[221]: <Axes: xlabel='2020-1', ylabel='Count'>



1.10 Alternative method: Read HTML table using Pandas

[ ]:
```
df_pandas = pd.read_html(url, attrs = {'class': 'tablepress tablepress-id-39'},␣
↪ flavor='bs4', thousands ='.')
df_pandas[0].head()
```

[ ]:

	Rank	Importer	US$	%Total	2020-1
0	1.0	China	$20,287,362,000	27.2%	+17.1%
1	2.0	Hong Kong	$7,280,376,000	9.8%	+0.6%
2	3.0	United States	$5,557,871,000	7.5%	+23.9%
3	4.0	Macao	$3,221,960,000	4.3%	+36.6%
4	5.0	Singapore	$3,087,914,000	4.1%	+6.7%

[ ]:
```
df_pandas[0].info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 102 entries, 0 to 101
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Rank      100 non-null    float64
```

```
 1   Importer  100 non-null    object
 2   US$       100 non-null    object
 3   %Total    100 non-null    object
 4   2020-1    100 non-null    object
dtypes: float64(1), object(4)
memory usage: 4.1+ KB
```