

AI Report

Introduction

I found AI to be a very helpful tool throughout this project. While I was still able to make my own visualisations and observations, using AI to generate the specific code needed to scrape data from certain HTML elements was especially useful. The HTML used on the website I was scraping from stored data in a `<tbody>` element with certain classes that I had never seen before, I was able to use AI to tweak the code I already had to make it work properly.

The main challenge I faced when using AI was that the initial outputs were often overly complex and included libraries or functions, I hadn't seen or used before such as lambdas. I had to rework my prompts several times because some of the responses were either too advanced for my understanding or too simplified to be useful. I knew I would not actually understand what the code was doing and would therefore struggle to explain it.

Prompt	Output
<p>How can I scrape something stored within a HTML <code><td class td></code></p> <pre><td class="smaller">9</td> ... <td class="bl mp smaller"> == \$0 <div>Netherlands</div></pre>	<pre>country_cell = linode.find("td", class_="bl mp smaller") print("top country:", country_cell.get_text())</pre>
<p>Its giving me this error:</p> <pre>def getInfo(linode): #See AI Report print("artist:", linode.find_all("a")[0].get_text()) country_cell = linode.find("td", class_="bl mp smaller") print("top country:", country_cell.get_text()) ulnode = soup.find("table", {"id": "artists"}) for linode in ulnode.find_all("tr"):## why tr getInfo(linode) print()</pre> <p>-----</p> <p>IndexError Traceback (most recent call last)</p> <p>/tmp/ipython-input-79772136.py in <cell line: 0>()</p> <p>6 ulnode = soup.find("table", {"id": "artists"})</p> <p>7 for linode in ulnode.find_all("tr"):## why tr</p> <p>----> 8 getInfo(linode)</p> <p>9 print()</p> <p>/tmp/ipython-input-79772136.py in getInfo(linode)</p> <p>1 def getInfo(linode): #See AI Report</p> <p>----> 2 print("artist:", linode.find_all("a")[0].get_text())</p> <p>3 country_cell = linode.find("td", class_="bl mp smaller")</p> <p>4 print("top country:", country_cell.get_text())</p> <p>5</p> <p>IndexError: list index out of range</p>	<pre>def getInfo(linode): print("artist:", linode.find_all("a")[0].get_text()) country_cell = linode.find("td", class_="bl mp smaller") print("top country:", country_cell.get_text()) ulnode = soup.find("table", {"id": "artists"}).find("tbody") for linode in ulnode.find_all("tr"): getInfo(linode) print()</pre>
<p>What command can I use to change a dictionary to dataframe</p>	<pre>df2 = pd.DataFrame(dictionary_name) print(df2)</pre>
<p>How can I create a seperate dataset with only these countries listed:</p>	<pre>df_latin = merged[merged["top country"].isin(latin_countries)]</pre>

How can I group by and get averages	
How can I group by and get the count of each entry	<pre>top_genre = (merged.groupby(["top country", "genre"]) .size() .reset_index(name="count"))</pre>