Question 4: Use Webscraping to Extract GME Revenue Data

Use the `requests` library to download the webpage https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-PY0220EN-SkillsNetwork/labs/project/stock.html. Save the text of the response as a variable named `html_data_2`.

[5]:
```
from bs4 import BeautifulSoup
import requests

# Step 1: Download the HTML data using requests (assuming the URL from previous tasks)
url = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-PY0220EN-SkillsNe
response = requests.get(url)

# Step 2: Parse the HTML content using BeautifulSoup
html_data_2 = response.text

# Using 'html.parser' or 'html5lib' parser
soup = BeautifulSoup(html_data_2, 'html.parser')  # You can also use 'html5lib'

# Step 3: Print out the parsed HTML (first 500 characters) to check the content
print(soup.prettify()[:500])  # Prints the first 500 characters in a pretty format
```

```
<!DOCTYPE html>
<!-- saved from url=(0105)https://web.archive.org/web/20200814131437/https://www.macrotrends.net/stocks/charts/GME/g
amestop/revenue -->
<html class="js flexbox canvas canvastext webgl no-touch geolocation postmessage websqldatabase indexeddb hashchange
```

Parse the html data using `beautiful_soup` using parser i.e `html5lib` or `html.parser`.

[6]:
```
soup = BeautifulSoup(html_data_2, 'html.parser')
```

Using `BeautifulSoup` or the `read_html` function extract the table with `GameStop Revenue` and store it into a dataframe named `gme_revenue`. The dataframe should have columns `Date` and `Revenue`. Make sure the comma and dollar sign is removed from the `Revenue` column.

> **Note: Use the method similar to what you did in question 2.**

▼ Click here if you need help locating the table

Below is the code to isolate the table, you will now need to loop through the rows and columns like in the previous lab

```
soup.find_all("tbody")[1]
```

If you want to use the read_html function the table is located at index 1

[15]:
```
import requests
import pandas as pd
from bs4 import BeautifulSoup

# Step 1: Download the HTML content of the webpage
```

Using `BeautifulSoup` or the `read_html` function extract the table with `GameStop Revenue` and store it into a dataframe named `gme_revenue`. The dataframe should have columns `Date` and `Revenue`. Make sure the comma and dollar sign is removed from the `Revenue` column.

> **Note: Use the method similar to what you did in question 2.**

▶ Click here if you need help locating the table

```
import requests ●●●
```

Display the last five rows of the `gme_revenue` dataframe using the `tail` function. Take a screenshot of the results.

[16]:
```
# Display the last 5 rows of the gme_revenue DataFrame
gme_revenue.tail()
```

[16]:

	Date	Revenue
11	2009	8806
12	2008	7094
13	2007	5319
14	2006	3092
15	2005	1843