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`.

[18]:
```
import requests

# URL of the stock revenue webpage
url2 = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-PY0220EN-SkillsNetwork/labs/project/stock.html"

# Download the webpage using requests
response2 = requests.get(url2)

# Save the text of the response as html_data_2
html_data_2 = response2.text
```

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

[19]:
```
from bs4 import BeautifulSoup

# Parse the HTML data with BeautifulSoup using html.parser
soup2 = 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.**

```
# 1. Create an empty DataFrame
gme_revenue = pd.DataFrame(columns=["Date", "Revenue"])

# 2. Find the correct table (quarterly revenue is in tbody at index 1)
table = soup2.find_all("tbody")[1]

# 3. Loop through rows and extract data
for row in table.find_all("tr"):
    cols = row.find_all("td")
    if len(cols) == 2:
        date = cols[0].text.strip()
        # Remove dollar sign and commas, then strip spaces
        revenue = cols[1].text.replace('$', '').replace(',', '').strip()
        gme_revenue = pd.concat(
            [gme_revenue, pd.DataFrame({"Date": [date], "Revenue": [revenue]})],
            ignore_index=True
        )
```

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

[21]:
```
# Display the last 5 rows of gme_revenue
gme_revenue.tail()
```

[21]:

	Date	Revenue
57	2006-01-31	1667
58	2005-10-31	534
59	2005-07-31	416
60	2005-04-30	475
61	2005-01-31	709