

## ▼ 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`.

```
[97]: import requests

url = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-PY0220EN-SkillsNetwork/labs/project/stock.html"
response = requests.get(url)
```

```
[98]: html_data_2 = response.text
```

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

```
[99]: from bs4 import BeautifulSoup

soup = BeautifulSoup(html_data, "html.parser")
```

Using `BeautifulSoup` or the `read_html` function extract the table with `GameStop Quarterly Revenue` and store it into a dataframe named `gme_revenue`. The dataframe should have columns `Date` and `Revenue`. Make sure the

```
[108]: tbodies = soup.find_all("tbody")

GameStop_quarterly_table = tbodies[1]
```

```
[114]: import pandas as pd

gme_revenue = pd.DataFrame(columns=["Date", "Revenue"])
```

```
[115]: for row in GameStop_quarterly_table.find_all("tr"):
    cols = row.find_all("td")
    if len(cols) == 2:
        date = cols[0].text.strip()
        revenue = cols[1].text.strip()

        gme_revenue = pd.concat([
            gme_revenue,
            pd.DataFrame({"Date": [date], "Revenue": [revenue]})
        ], ignore_index=True)
```

```
[116]: gme_revenue["Revenue"] = gme_revenue["Revenue"].str.replace(',', '\$', regex=True)
```

```
[116]: gme_revenue["Revenue"] = gme_revenue["Revenue"].str.replace(',|\$', '', regex=True)
gme_revenue.dropna(inplace=True)
gme_revenue = gme_revenue[gme_revenue["Revenue"] != ""]
```

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

```
[117]: gme_revenue.tail()
```

```
[117]:
```

	Date	Revenue
49	2010-06-30	28
50	2010-03-31	21
51	2009-09-30	46
52	2009-06-30	27
53	2009-06-30	27