

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`.

```
[17]: html_data = requests.get("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-PY0220EN-SkillsNetwork/labs/
```



Parse the html data using `beautiful_soup`.

```
[18]: soup = BeautifulSoup(html_data,"html5lib")
```

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 using a method similar to what you did in Question 2.

► [Click here if you need help locating the table](#)

```
[19]: table = soup.find_all("tbody")[1]
      data = table.find_all("td")

      ----
      gme_revenue=pd.DataFrame({'Date': [str(entry.string) for entry in data if '$' not in entry.string],
                               "Revenue": [float(str(entry.string).replace(",","").replace("$","")) for entry in data if '$' in entry.string]})
      gme_revenue.head()
```

```
[19]:
```

	Date	Revenue
0	2020-04-30	1021.0
1	2020-01-31	2194.0
2	2019-10-31	1439.0
3	2019-07-31	1286.0
4	2019-04-30	1548.0

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

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



```
[20]:
```

	Date	Revenue
57	2006-01-31	1667.0
58	2005-10-31	534.0
59	2005-07-31	416.0
60	2005-04-30	475.0
61	2005-01-31	709.0