Question 2: Use Webscraping to Extract Tesla 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/revenue.htm Save the text of the response as a variable named `html_data`.

```
url="https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-PY0220EN-SkillsNetwork/labs/project/revenue.htm"
html_data  = requests.get(url).text
print(html_data)
```

```
<!DOCTYPE html> ●●●
```

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

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

Using `BeautifulSoup` or the `read_html` function extract the table with `Tesla Revenue` and store it into a dataframe named `tesla_revenue`. The dataframe should have columns `Date` and `Revenue`.

▶ Step-by-step instructions

▶ Click here if you need help locating the table

```
# Step 2: Locate the correct table (Quarterly Revenue)
tables = soup.find_all("table")  # Find all tables
tesla_revenue_table = tables[1]  # The second table contains revenue data

# Step 3: Create an empty DataFrame
tesla_revenue = pd.DataFrame(columns=["Date", "Revenue"])
```

```
# Step 4: Extract rows from the table body
table_rows = tesla_revenue_table.find("tbody").find_all("tr")

# Step 5: Loop through rows and extract Date and Revenue
for row in table_rows:
    cols = row.find_all("td")  # Find all columns in the row
    if len(cols) == 2:  # Ensure it has both Date and Revenue columns
        date = cols[0].text.strip()
        revenue = cols[1].text.strip()

        # Append to DataFrame
        tesla_revenue = pd.concat([tesla_revenue, pd.DataFrame({"Date": [date], "Revenue": [revenue]})], ignore_index=True)

# Step 6: Print the first few rows
print(tesla_revenue.head())
```

```
         Date  Revenue
0  2022-09-30  $21,454
1  2022-06-30  $16,934
2  2022-03-31  $18,756
3  2021-12-31  $17,719
4  2021-09-30  $13,757
```

```
tesla_revenue.tail()
```

	Date	Revenue
49	2010-06-30	$28
50	2010-03-31	$21
51	2009-12-31	

Execute the following line to remove the comma and dollar sign from the `Revenue` column.

[36]:
```
tesla_revenue["Revenue"] = tesla_revenue['Revenue'].str.replace(',|\$', "", regex=True)
```

Execute the following lines to remove an null or empty strings in the Revenue column.

[37]:
```
tesla_revenue.dropna(inplace=True)

tesla_revenue = tesla_revenue[tesla_revenue['Revenue'] != ""]
```

[38]:
```
print(tesla_revenue.head())
```

```
         Date Revenue
0  2022-09-30   21454
1  2022-06-30   16934
2  2022-03-31   18756
3  2021-12-31   17719
4  2021-09-30   13757
```

Display the last 5 row of the `tesla_revenue` dataframe using the `tail` function. Take a screenshot of the results.

[39]:
```
tesla_revenue.tail()
```

[39]:

	Date	Revenue
48	2010-09-30	31
49	2010-06-30	28
50	2010-03-31	21
52	2009-09-30	46
53	2009-06-30	27