 <b>Marwadi University</b>	<b>Marwadi University</b> <b>Faculty of Engineering and Technology</b> <b>Department of Information and Communication Technology</b>	
<b>Subject: Data Visualization and Dashboard (01CT0410)</b>	<b>Aim: Web page scraping using beautifulsoup in python</b>	
<b>Experiment No: 10</b>	<b>Date: 19-02-2024</b>	<b>Enrollment No: 92200133030</b>

**Aim:** Web page scraping using beautifulsoup in python

**IDE:** Tableau

**In Lab :-**

Task-1:- Collect the data of the top 100 big companies by revenue of USA


**Code :-**

```
import pandas as pd
import requests
from bs4 import BeautifulSoup
url = "https://en.wikipedia.org/wiki/List_of_largest_companies_in_the_United_States_by_revenue"
Page = requests.get(url)
Soup = BeautifulSoup(Page.text,"html")
Table = Soup.find_all("table")
Table = Table[1]
Headers = Table.find_all("th")
Titles = [titles.text.strip() for titles in Headers]
Dataset = pd.DataFrame(columns = Titles)
Rows = Table.find_all('tr')
for row in Rows[1:] :
    Data = row.find_all('td')
    Row_Data = [row.text.strip() for row in Data]
    length = len(Dataset)
    Dataset.loc[length] = Row_Data
print(Dataset)
Dataset.to_excel("D:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-11
Web Scrapping Using Python/Largest Companies In USA by Revenue.xlsx",index=False)
print("File Saved")
```

**Output :-**

```
Soup = BeautifulSoup(Page.text,"html")
Rank      Name      Industry Revenue (USD millions) Revenue growth Employees      Headquarters
0 1 Walmart Retail 611,289 6.7% 2,100,000 Bentonville, Arkansas
1 2 Amazon Retail and cloud computing 513,983 9.4% 1,540,000 Seattle, Washington
2 3 ExxonMobil Petroleum industry 413,680 44.8% 62,000 Spring, Texas
3 4 Apple Electronics industry 394,328 7.8% 164,000 Cupertino, California
4 5 UnitedHealth Group Healthcare 324,162 12.7% 400,000 Minnetonka, Minnesota
.. ... ..
95 96 Best Buy Retail 46,298 10.6% 71,100 Richfield, Minnesota
96 97 Bristol-Myers Squibb Pharmaceutical industry 46,159 0.5% 34,300 New York City, New York
97 98 United Airlines Airline 44,955 82.5% 92,795 Chicago, Illinois
98 99 Thermo Fisher Scientific Laboratory instruments 44,915 14.5% 130,000 Waltham, Massachusetts
99 100 Qualcomm Technology 44,200 31.7% 51,000 San Diego, California

[100 rows x 7 columns]
File Saved
PS C:\Users\online>
```

 <b>Marwadi University</b>	<b>Marwadi University</b> <b>Faculty of Engineering and Technology</b> <b>Department of Information and Communication Technology</b>	
<b>Subject: Data Visualization and Dashboard (01CT0410)</b>	<b>Aim: Web page scraping using beautifulsoup in python</b>	
<b>Experiment No: 10</b>	<b>Date: 19-02-2024</b>	<b>Enrollment No: 92200133030</b>

**Post Lab :-**

Task-1:- Collect the data of the top 100 big companies by revenue of USA by selenium

**Code:-**

```
import pandas as pd
from selenium import webdriver
url = "https://en.wikipedia.org/wiki/List_of_largest_companies_in_the_United_States_by_revenue"
driver = webdriver.Chrome() # You need to have ChromeDriver installed and in PATH
driver.get(url)
table = driver.find_elements("xpath" , "//*[@table]")[1]
headers = table.find_elements("xpath" , ".//th")
titles = [header.text.strip() for header in headers]
dataset = pd.DataFrame(columns=titles)
rows = table.find_elements("xpath" , ".//tr")[1:]
for row in rows:
    data = row.find_elements("xpath" , ".//td")
    row_data = [cell.text.strip() for cell in data]
    dataset.loc[len(dataset)] = row_data
print(dataset)
dataset.to_excel("D:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-11 Web
Scrapping Using Python/Largest Companies In USA by Revenue by selenium.xlsx", index=False)
print("File Saved")
driver.quit()
```

**Output :-**

```
PS C:\Users\online> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Data Visulization and Dashboards/Lab Manual/Exp-11
Web Scrapping Using Python/Web_Scrapping.py"

DevTools listening on ws://127.0.0.1:59137/devtools/browser/3e43ea32-3607-47c2-9723-b9d166d54e7b
Rank      Name      Industry Revenue\n(USD millions) Revenue growth Employees      Headquarters
0 1 Walmart Retail 611,289 6.7% 2,100,000 Bentonville, Arkansas
1 2 Amazon Retail and cloud computing 513,983 9.4% 1,540,000 Seattle, Washington
2 3 ExxonMobil Petroleum industry 413,680 44.8% 62,000 Spring, Texas
3 4 Apple Electronics industry 394,328 7.8% 164,000 Cupertino, California
4 5 UnitedHealth Group Healthcare 324,162 12.7% 400,000 Minnetonka, Minnesota
.. ... ..
95 96 Best Buy Retail 46,298 10.6% 71,100 Richfield, Minnesota
96 97 Bristol-Myers Squibb Pharmaceutical industry 46,159 0.5% 34,300 New York City, New York
97 98 United Airlines Airline 44,955 82.5% 92,795 Chicago, Illinois
98 99 Thermo Fisher Scientific Laboratory instruments 44,915 14.5% 130,000 Waltham, Massachusetts
99 100 Qualcomm Technology 44,200 31.7% 51,000 San Diego, California

[100 rows x 7 columns]
File Saved
PS C:\Users\online>
```