# 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 :-

# 

# 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 :-

# 