Performa for Project 2023

Course Code: **Group No.:-** 03

Group Detail					
S.No.	Name of Candidate	BTU Roll No.	Branch	E-mail	Mobil No.
1.	Anjali Sharma	20EBKCS011	CS-1	sharmanji20@gmail.com	7023943101
2.	Jatin	20EBKCS052	CS-1	jatinbkbiet@gmail.com	8003398009

Title of the Project: Web Scrapping for Data Analysis

Abstract:

Our project focuses on harnessing the power of Python for web scraping to gather pertinent information from diverse online sources and transform it into actionable insights.

The Python Web Scraping for Data Analysis project explores web scraping techniques using Python, emphasizing BeautifulSoup and Selenium libraries. It focuses on finding and selecting elements, scraping live sites, and analyzing data. Participants practice finding elements using BeautifulSoup and Selenium, learning HTML parsing and dynamic content scraping. They extract data from live sites, storing it using CSV or pandas. Analysis involves data cleaning, transformation, aggregation, and visualization with pandas, NumPy, and Matplotlib. This project develops practical scraping skills and data analysis proficiency, essential for business intelligence, market research, and data-driven decision-making.

Concepts Practised :-

Finding and Selecting Particular Elements with BeautifulSoup: Leveraging the BeautifulSoup library, we efficiently locate specific elements within the HTML structure of web pages, enabling precise data extraction.

Finding and Selecting Elements on a Website with Selenium: With Selenium, we navigate through dynamic web pages, interact with elements, and extract relevant data, expanding our scraping capabilities.

Use BeautifulSoup to Scrape Website Data: We utilize BeautifulSoup's robust parsing functionalities to extract structured data from HTML documents with ease and precision.

Use Selenium to Scrape Website Data: Selenium empowers us to scrape data from websites that employ JavaScript and dynamic content, ensuring comprehensive data collection.

Scraping a Live Website: By scraping live websites, we capture real-time data updates, ensuring the freshness and relevance of our dataset

The extracted data can be stored in CSV files or pandas DataFrames, facilitating easy access and manipulation. Professionals can then employ advanced data analysis techniques, including cleaning messy data, transforming formats, aggregating information, and visualizing trends and patterns using pandas, NumPy, and Matplotlib. These insights aid decision-making processes in business intelligence, market research, and competitive analysis, empowering organizations to stay ahead in their respective fields.

Keywords: Python, Web Scraping, BeautifulSoup, Selenium, Data Analysis, Data Visualization, CSV, Pandas, NumPy, Matplotlib.

References: 1. Dr. Angela Yu, Developer and Lead Instructor for "100 Days of Code: The Complete Python Pro Bootcamp".

2. Udemy, for learning purposes.

3. Python.org, for documentation of the required libraries.

Guides Declaration

I certify that the topic selected by Ms. Anjali Sharma & Mr. Jatin **, in my opinion is fully adequate in scope and quality as a project for the degree of B.Tech in** Computer Science

Approval Date:- 16 March 2024 **Approval of Guide(YES/NO):** Yes

Approved by (Name Of Guide): Dr. Soram Mittal

Have you join whatsapp group(of your guide YES/NO): Yes

Project and Seminar Evaluation Cell