



COMSATS University Islamabad, Lahore Campus

Course Title:	Introduction to Data Science			Course Code:	CSC461	Credit Hours:	3(3,0)
Resource Person:	Dr. Muhammad Sharjeel			Programme Name:	BSSE		
Semester:	5 th	Batch:	FA21	Section:	C	Max Marks:	10

Assignment 1

Due Date: 18-09-2023

Submission: Upload the assignment solution to your GitHub account (private repository).

Important instructions: For each of the task (question) given below, please write the following information at the start of your ipython file.

Date

CSC461 – Assignment1 – Web Scraping

Your Full Name

You Complete Registration Number

A brief description of the task

Q1. [CLO-2]

Create a web scraper in Python (preferably a Jupyter Notebook) to extract the ‘title’ and ‘rating’ for 5 of your most favorite movies from the IMDB website. In order to do this, you must create a list (array) of movie URLs and iterate through the list to scrape (all IMDB movie pages have the same format) the required information for each movie. Once you have all the information you need, export the data to an excel file.

Please note that after fetching a webpage, your script must sleep for 1 second (by calling `time.sleep(seconds)`) so that you do not abuse IMDB’s servers.

Q2. [CLO-2]

Following are two websites that list historical events happened on a specific date.

- <https://www.timeanddate.com/on-this-day>
- <https://www.britannica.com/on-this-day>

Your job is to write a web scraper in Python (preferably a Jupyter Notebook) to extract the below mentioned information.

1. From the ‘timeanddate’ website, find out who you share your birthdate with?
2. From the ‘britannica’ website, find out which important event(s) happened on your birthdate?

Once you have all the information, write it to a text file with proper indentation/formatting.