|  |  |
| --- | --- |
| **SESSION** | **July 2023** |
| **PROGRAM** | **MASTER OF BUSINESS ADMINISTRATION (MBA)** |
| **SEMESTER** | **IV** |
| **course CODE & NAME** | **DADS404 – DATA Scrapping** |
| **CREDITS** | **04** |
| **nUMBER OF ASSIGNMENTS & Marks** | **02**  **30 MARKS EACH** |

**Assignment Set – 1**

**1 a. Define data scraping and explain its significance in the digital age.**

**b. List and briefly describe three tools used for data scraping. Discuss one advantage for each.**

**Ans 1a.**

Data scraping, also known as web scraping or data harvesting, is the process of automatically extracting information from websites or online sources. It involves using software or programming scripts to navigate web pages, locate specific data elements, and then collect and store that data in a structured format, such as a spreadsheet or database. Data scraping is significant in the digital age for Its Half solved only

Buy Complete from our online store

<https://smuassignment.in/online-store/>

MUJ Fully solved assignment available for**session July 2023.**

Lowest price guarantee with quality.

Charges**INR 200 only per assignment.**For more information you can get via mail or Whats app also

Mail id is [aapkieducation@gmail.com](mailto:aapkieducation@gmail.com)

Our website www.smuassignment.in

After mail, we will reply you instant or maximum

1 hour.

Otherwise you can also contact on our

whatsapp no 8791490301.

**2 a. Outline the ethical considerations a data scraper must keep in mind. Why is respecting robots.txt important?**

**b. Differentiate between data scraping and data wrangling. Provide two tasks unique to each process.**

**Ans 2a.**

**Ethical Considerations for Data Scraping:**

1. **Respecting Robots.txt**: This is crucial because it signals the website's owner's

**3 a. Given a sample website structure, identify the potential challenges in scraping data and suggest solutions.**

**b. Explain the process of manual scraping using Python. Include a brief code snippet as an example.**

**Ans 3a.**

**Potential challenges in scraping data from a website:**

**Website Structure Changes**: Websites frequently undergo structural changes, such as layout modifications, updated CSS classes, or reorganized HTML elements. This can break your scraping co

**Assignment Set – 2**

**1 a. Explain what is API-based scraping. Why is it often preferred over traditional web scraping methods?**

**b. Describe how rate limits and authentication mechanisms work in API-based scraping, giving an example of a popular API that employs these.**

**Ans 1a.**

1. API-based scraping, also known as web scraping using APIs (Application Programming Interfaces), involves extracting data from websites or online services by interacting with their APIs instead of directly parsing the web pages' HTML content. This method is often preferred over traditional web

**2. Using Twitter as an example, discuss how you would access and scrape data using R. What challenges might arise?**

**Ans 2.**

Accessing and scraping data from Twitter using R involves a series of steps, including authentication, querying the Twitter API, and extracting the desired data. Here's a high-level overview of the process, along with potential challenges:

1. **Set up a Twitter Developer**

**3a. Describe the process of scraping cryptocurrency data. Highlight the importance of putting this data in a standard format and the potential challenges in doing so.**

**b. Define 'Data Quality'. Discuss two dimensions of data quality and explain how automated data quality checks can be beneficial.**

**Ans 3a.**

**Scraping Cryptocurrency Data:**

Scraping cryptocurrency data involves extracting information about various cryptocurrencies, such as their prices, trading volumes, market capitalization, historical data, and more from websites or APIs. This data is valuable for traders, investors, analysts, and researchers to make informed decisions in the volatile cryptocurrency market. Here's a step-by-step description of the