**Word Frequency in Classic Novels**

**Abstract: -**

In the Word Frequency in Classic Novels project, you will use requests and Beautiful Soup to scrape a novel from the Project Gutenberg website. After scraping and cleaning the text data, you will use NLP to find the most frequent words in Moby Dick. The project introduces you to the world of Python web scraping and natural language processing.

For data analysts and data scientists, web scraping is an essential skill to learn. You can take a short Web Scraping with Python course to understand the tools and components of an HTML web page.

**Topic:**

Analysing word frequency in classic novels using web scraping and Natural Language Processing (NLP) techniques.

**Description of Word Frequency in Classis Novels:**

The "Word Frequency in Classic Novels" project involves using Python libraries such as requests and BeautifulSoup for web scraping to extract a classic novel from the Project Gutenberg website. By scraping and cleaning the text data, the project aims to analyse the word frequency in the novel "Moby Dick" using Natural Language Processing (NLP) techniques. This project serves as a practical application of Python web scraping and NLP, providing valuable insights into text analysis.

For data analysts and data scientists, web scraping is a fundamental skill to have. Taking a short course on Web Scraping with Python can help individuals understand the tools and components of HTML web pages, enhancing their proficiency in this area. By engaging in projects like this, professionals can further develop their expertise in web scraping and NLP, both of which are essential for roles in data analysis and data science.

**Web scraping and natural language processing (NLP) are powerful skills for data analysts and data scientist. Here’s a quick recap of the steps involved in your project.**

1. Scraping the Novel:

Use requests to fetch the HTML content of the novel from Project Gutenberg.

Parse the HTML using BeautifulSoup to extract the text.

1. Cleaning the Text:

Remove HTML tags, punctuation, and convert the text to lowercase.

Split the text into individual words.

1. Removing Stop Words:

Use the nltk library to filter out common stop words.

1. Analyzing Word Frequency:

Use collections. Counter to count the frequency of each word and identify the most common ones.

**Summary:**

In the "Word Frequency in Classic Novels" project, you will utilize the requests library and Beautiful Soup package in Python to scrape a classic novel from the Project Gutenberg website. After extracting and cleaning the text data, you will employ Natural Language Processing (NLP) techniques to identify the most frequently used words in the novel "Moby Dick." This project serves as an introductory exploration into Python web scraping and NLP, offering a practical application of these skills in analyzing literary texts.

For individuals working in data analysis and data science fields, proficiency in web scraping is considered a crucial skill to possess. By engaging in projects like this one, you can enhance your understanding of web scraping tools and processes, as well as gain insights into the fundamental components of HTML web pages. Additionally, you may consider taking a short course on Web Scraping with Python to further solidify your knowledge and capabilities in this area.