|  |
| --- |
| Explain the word data wrangling. Identify and explain the significance of data wrangling in the context of preparing datasets for analysis. |
| Demonstrate how you would use Python to open and read the data from a Comma Separated Values(CSV) file. |
| Demonstrate how does it facilitate the process of cleaning and transforming raw data into a usable format? |
| Imagine you need to process an XML file containing information about the person. How would you read this XML data and display using python code. |
| Demonstrate how you would use Python to open and read the data from a Comma Separated Values(CSV) file. |
| Apply your knowledge to import a JSON file named "data.json" into a Python script and display the contents. |
| Outline importance of data wrangling process. |
| Imagine you need to process an XML file containing information about the person. How would you read this XML data and display using python code. |
| Explain different data wrangling tools available for cleaning, data manipulation, analysis and organizing the data sets. |
| Explain the different basic data types available in python with examples for each. |
| Illustrate the use of list and its methods for adding, removing the values and extending the list with examples. |
| Construct a Python dictionary and give an example for different methods of dictionary. |
| Explain the different data wrangling tools for cleaning, data manipulation, analysis and organizing the data sets. |
| Illustrate different data wrangling tools available for cleaning, data manipulation, analysis and organizing the data sets. |
| Contrast the word data wrangling. Identify and explain the significance of data wrangling in the context of preparing datasets for analysis. |
| Outline the basic data types and data containers available in python. |
| Apply your knowledge to import a JSON file named "data.json" into a Python script and display the contents. |
| Summarize how does data wrangling facilitate the process of cleaning and transforming raw data into a usable format? |
| Develop a python script to read the xml document containing a person information. |
| Explain mutable and immutable data containers available in the python with examples for each. |
| Elaborate the process of cleaning and transforming raw data into a usable format? |
| Explain the significance of data wrangling in the context of preparing datasets for analysis. |
| Explain type(), help() and dir() methods in python. |

Module-1



Module-2

|  |
| --- |
| Demonstrate how you would use a Python script to open and parse text data from a PDF file. |
| Write a python script to read the excel file **data.xlsx** and display the contents of it using suitable library. |
| Imagine you have pdf file “EN-Final table9.pdf” and converted this to text file.Write a python script to read the country names from the converted text file “EN-Final table9.txt” |
| Demonstrate how to convert pdf file to text file using pdfminer and write a script to display the contents of file. |
| Compare The differences between relational database and non-relational database with examples. |
| Make use of openpyxl library and write a python script to read the excel files. |
| Write the Python script to read and parse the text data from a PDF file. Outline the procedure to read text data from PDF. |
| Demonstrate how to convert pdf file to text file using pdfminer and write a script to display the contents of file |
| Write a python script to read the country names from the converted text file “EN-Final table9.txt”(consider pdf file “EN-Final table9.pdf” is converted to text file) |
| Write a python script to read the excel file(.xlsx) and display the contents of it using suitable library. |
| Explain the use of range method and indexing concept along with slicing method in python. |
| Consider the text file table.txt containing tabular data with country names and corresponding totals. Write a python script that reads this file cleans the data and transform it into dictionary where each country name is paired with its total value. |
| the use of range() method and indexing concept along with slicing method in python. |
| Imagine you have a text file table.txt containg tabular data with country names and corresponding totals .write a python script that reads this file cleans the data and transform it into dictionary where each country name is paired with its total value. |
| Imagine you have excel file ”child labor and marriage.xlsx” .Write a python script to extract only child labor and marriage information of all countries. |
| Compare and explain the differences between relational database and non-relational database with examples. |
| Write a python script to read the country names from the converted text file “EN-Final table9.txt” |
| Write a python script to read the excel file data.xlsx and display the contents of it using suitable library. |
| Imagine you have excel file ”child labor and marriage.xlsx” .Write a python script to extract only child labor and marriage information of all countries. |
| Compare The differences between relational database and non-relational database with examples. |

Module 3

|  |
| --- |
| Compare and contrast with example fuzzy.ratio() and fuzzy.partial\_ratio() functions of fuzzy matching concept in python. |
| Develop a Python script that interacts with an SQLite database to store and retrieve your data. |
| Explain what an f-string is in Python and provide an example of how it can be used to format a string with variables. |
| Write Python code and explain how to print the current date and time in the format YYYY-MM-DD HH:MM:SS |
| Illustrate the concept of regular expressions to match, find, or eliminate patterns in strings or data using python code. |
| Outline the purpose of using fuzz.ratio() and fuzz.partial\_ratio() functions from the fuzzywuzzy library. What do these functions measure in terms of string similarity? |
| Illustrate the role of regular expressions. Describe the difference between re.findall() and re.search() methods in terms of their functionality and the type of results they return. |
| Discuss the significance of normalizing scores to a range between 0 and 1 and how it facilitates fair comparison of team performance across different categories. |
| Develop a Python script that interacts with an SQLite database to store and retrieve your data. |
| What is f-string in Python and explain how to format a string with variables. |
| Apply your knowledge to replace the short headers with longer headers using python code. Consider the following mn.csv and mn-headers.csv files. |
| Explain the normalization and standardization process with example. |
| Construct a Python script to store and view the data in SQLite database. |
| Given a list colors = ["red", "green", "blue"], write code to output these colors as a comma-separated string. |
| Write Python code to format the following dictionary data in order to get the given output:  data = {  'price': 199.99,  'quantity': 1000,  'discount\_rate': 0.15  }  output:  Price: $199.9900  Quantity: 1,000  Discount Rate: 15.000% |
| OR |
| Illustrate the purpose of np.where(np.abs(z\_scores)>threshold)[0]. What does it identify in the dataset? |
| Compare and contrast using Python's built-in set and NumPy's unique method for identifying unique elements in datasets. |
| How does cleaning data facilitate easier storage, search, and reuse. Describe the process of identifying and replacing values for data cleanup. |
| Develop Python code to format the current date and time in the following form YYYY-MM-DD HH:MM:SS |
| OR |
| Given a list fruits = ["apple", "banana", "orange"], write code to output these fruits as a comma-separated string. |
| Explain the use of z-score in identifying the outliers from dataset. |
| Demonstrate the concept of fuzzy ratio and fuzzy partial ratio concept with example. |

Module 4

|  |
| --- |
| Explain the steps to determine the number of sheets and their names and display the sheet\_row\_values in the workbook using the xlrd library, and describe how you would proceed with the agate library to explore and analyze the data further. |
| Make use of agate data analysis library to create agate table and explain how it will access the data according to its documentation with example. |
| Utilize different agate data analysis library functions to perform different operations on the agate table. |
| Elaborate on joins and Compare different methods to join datasets with examples for each. |
| Elaborate on joins and Compare different methods to join datasets with examples for each. |
| Identify and explain the different tools that can helps in digital storytelling and data presentation. |
| Construct a python script to provide insights into the distribution and characteristics of the data within each category using aggregate functions. |
| Make use of agate data analysis library to create agate table and explain how it will access the data according to its documentation with example. |
| Explain the steps to determine the number of sheets and their names and display the rows in the particular sheet using the xlrd library, and mention how the agate library reads data. |
| Identify and contrast different tools that can helps in digital storytelling and data presentation. |
| Make use of agate data analysis library to create agate table and explain how it will access the data according to its documentation with example. |
| Develop a python script to give insights about the data within each category using aggregate functions. |
| Construct a python script to provide insights into the distribution and characteristics of the data within each category using aggregate functions. |
| Construct print bars using python script to visualize the aggregated data by taking sample example. |
| Identify and explain the different tools that can helps in digital storytelling and data presentation. |
| Develop a python script for visualizing the dataset in bar chart using matplotlib.pyplot library. |
| Write a python script to provide insights into the distribution and characteristics of the data within each category using aggregate functions. |
| Identify and explain the different tools that can helps in digital storytelling and data presentation. |
| Define joins and explain the different methods of joins with examples for each. |
| Utilize different agate data analysis library functions to perform different operations on the agate table. |

Module 5

|  |
| --- |
| Write HTML code for a webpage that meets the following criteria:   1. Button Functionality:    * When the button is clicked, the background color of the entire webpage should change to green. 2. Button Styling:    * Padding: 20px 40px    * Background Color: Blue    * Text Color: White    * Font Size: 18px 3. Page Styling:    * The default background color of the page should be light gray.   Provide the complete HTML code, including inline CSS for styling and JavaScript for handling the button click event. |
| As part of a project, you are tasked with building a Scrapy spider to collect information from a news website. Demonstrate how you would structure your Scrapy spider code to navigate the site, locate relevant data, and store it in a structured format. |
| Demonstrate how you would use Beautiful Soup to extract text content from a specific HTML element. |
| You are tasked with extracting specific information from an XML file in a Python project. Identify how you would use LXML to parse the XML and extract the desired data. Provide a snippet of code to showcase this application. |
| Explain how does understanding the hierarchical relationship of HTML tags help in creating an effective web scraper and Write a simple structure for html code. |
| As part of a project, you are tasked with building a Scrapy spider to collect information from a news website. Demonstrate how you would structure your Scrapy spider code to navigate the site, locate relevant data, and store it in a structured format. |
| Write a python script to send request for the web page through request library. |
| Apply your knowledge to write a python script to read the web page using the inbuilt lxml library and explain in detail. |
| Demonstrate how does the Network tab in browser developer tools reveal about a web page's loading process. |
| Explain what data do you want to scrape from a website, and how will you decide what’s appropriate to collect? |
| Apply your knowledge to scrap the web using beautifulsoup library . |
| Write a python script that uses urllib.request and urllib.parse libraries when sscraping web page. Explain the code. |
| Contrast how can inspecting the Network tab help you identify which resources are dynamically loaded and might require additional scraping techniques? |
| Explain the concept of browser based parsing with example python code. |
| Outline what steps do you take to make sure you’re scraping data legally and ethically? |
| Demonstrate Selenium interacts with a website through any selenium supported browsers. |
| As part of a project, you are tasked with building a Scrapy spider to collect information from a news website. Demonstrate how you would structure your Scrapy spider code to navigate the site, locate relevant data, and store it in a structured format. |
| You are given two scripts—one using urllib and another using a different library for URL requests. Analyze the advantages and disadvantages of using urllib in terms of simplicity, flexibility, and features for making HTTP requests. |
| Demonstrate how you would use an XPath expression to locate and select a specific HTML element on a web page. |
| Write a html code that includes both internal CSS and javascript, and the webpage for that code should be as given below and when you click the button background color of the page should change. |
| Demonstrate how you would use Beautiful Soup to extract text content from a specific HTML element. |
| Explain how do you check if a website allows scraping, and where can you find rules about it? |
| Write a python script for scraping the web page using urllib library. Explain each step in the code. |
| Demonstrate how can inspecting the Network tab help you identify which resources are dynamically loaded and might require additional scraping techniques? |