Assignment no. 1. Practical Assignments on Data Science Tools and Environment, Python Programming Basics

Objective:

To develop familiarity with the data science tools and environment, understand the basics of Python programming for data science, work with Python data structures, and use data manipulation libraries like Pandas, NumPy, and Matplotlib.

Assignment no.	**Assignment statement**	**Turns**
1.1 **Introduction to the Laboratory Environment**	1. **Installation and Setup**: o Install Anaconda distribution which includes Jupyter Notebook. o Install a Python IDE (e.g., PyCharm, VS Code). 2. **Jupyter Notebook Basics**: o Create a new Jupyter Notebook. o Write a markdown cell explaining the purpose of the notebook. o Write and execute a simple Python code cell (e.g., print "Hello, Data Science!"). 3. **Exploring Python IDEs**: o Create a new Python project in the chosen IDE. o Write a simple Python script that prints the current date and time. o Execute the script from the IDE. 4. **Comparing Tools**: o Write a brief comparison (200 words) between Jupyter Notebook and the chosen Python IDE based on your experience.	1
1.2 **Introduction to Data Science**	1. **What is Data Science?**: o Write a 300-word essay on what Data Science is and its significance in today’s world. 2. **Applications of Data Science**: o Identify three real-world applications of Data Science and explain each in 200 words. 3. **Tools/Libraries for the Data Science**: o List five common tools or libraries used in Data Science. Provide a brief description of each tool/library. 4. **Case Study**: o Analyze a case study where Data Science was used to solve a significant problem. Summarize the case study in 300 words.	1
1.3 **Python Programming for Data Science**	1. **Basic Python Programs**: o Write a Python program to find the factorial of a number. o Write a Python program to check if a number is prime.	1
	2. **Control Structures**: o Write a Python program using if-else to determine whether a number is positive, negative, or zero. o Write a Python program using a loop to print the first 10 Fibonacci numbers.	
	3. **Functions in Python**: o Create a Python function to calculate the area of a circle.	1

	○ Create a Python function that takes a list of numbers and returns the largest number in the list. 4. **File Handling**: ○ Write a Python program to read a text file and count the number of words in it.	
1.4 **Python Data Structures**	1. **Lists and Tuples**: ○ Write a Python program to find the second largest number in a list. ○ Write a Python program to merge two tuples into a dictionary.	1
	2. **Dictionaries and Sets**: ○ Write a Python program to count the frequency of each word in a given string using a dictionary. ○ Write a Python program to find the union and intersection of two sets.	1
1.5 **Introduction to Data Manipulation Libraries**	1. **Pandas Basics**: ○ Create a DataFrame from a dictionary of lists. ○ Perform basic operations like selecting rows/columns, filtering, and sorting. 2. **NumPy Basics**: ○ Create a NumPy array and perform element-wise operations. ○ Use NumPy to perform matrix multiplication.	1
	3. **Data Visualization with Matplotlib**: ○ Create a simple line plot using Matplotlib. ○ Create a bar chart to visualize categorical data.	1
	4. **Data Analysis Project**: ○ Download a dataset (e.g., from Kaggle or UCI Machine Learning Repository). ○ Use Pandas to clean and preprocess the data. ○ Perform basic exploratory data analysis (EDA) and visualize the findings using Matplotlib.	1
	Total turns	9

Assignment no. 2. Practical Assignments on Data Acquisition and Cleaning

Objective:

To develop skills in data acquisition from various sources, web scraping, API integration, and data cleaning and preprocessing.

Assignment no.	**Assignment statement**	**Turns**
2.1 **Data Importing from Various Sources**	1. **Importing Data from CSV**: o Download a CSV file from the internet (e.g., a public dataset from Kaggle). o Use Python to import the CSV file into a DataFrame. o Display the first few rows of the DataFrame. 2. **Importing Data from JSON**: o Download a JSON file from the internet. o Use Python to import the JSON file into a DataFrame. o Display the first few rows of the DataFrame.	1
	3. **Importing Data from SQL Databases**: o Set up a local SQLite database and create a table with some sample data. o Use Python to import data from the SQL table into a DataFrame. o Display the first few rows of the DataFrame.	1
2.2 **Data Scraping Techniques**	1. **Web Scraping with BeautifulSoup**: o Identify a website with tabular data (e.g., a Wikipedia table). o Use BeautifulSoup to scrape the data and store it in a DataFrame. o Display the first few rows of the DataFrame. 2. **API Integration**: o Identify a public API (e.g., OpenWeatherMap, CoinGecko). o Use Python to fetch data from the API and store it in a DataFrame. o Display the first few rows of the DataFrame.	1
2.3 **Data Cleaning and Preprocessing**	1. **Handling Missing Values**: o Create a DataFrame with some missing values. o Use Python to handle missing values by dropping rows, filling with a specific value, and filling with the mean of the column. o Display the DataFrame after each operation.	1
	2. **Handling Outliers**: o Create a DataFrame with some outliers. o Use Python to identify outliers using the Z-score method. o Remove the outliers and display the cleaned DataFrame.	1
	Total turns	5

Midterm Submission 1 Turn

Assignment no. 3 Practical Assignments on Data Visualization and Exploratory Data Analysis (EDA)

Objective:

To develop skills in data visualization using libraries like Matplotlib and Seaborn and to perform Exploratory Data Analysis (EDA) using summary statistics, data distribution analysis, and correlation analysis.

Assignment no.	**Assignment statement**	**Turns**
3.1 Data Visualization Libraries	1. **Basic Data Visualization with Matplotlib**: o Use Matplotlib to create a simple line plot, bar chart, and scatter plot using a sample dataset. 2. **Advanced Data Visualization with Seaborn**: o Use Seaborn to create a histogram, box plot, and heatmap using a sample dataset.	1
3.2 Exploratory Data Analysis Techniques	1. **Summary Statistics**: o Use a sample dataset (e.g., the Iris dataset) to calculate and display summary statistics (mean, median, standard deviation, etc.).	1
	2. **Data Distribution Analysis**: o Use Seaborn to visualize the distribution of data for different variables using histograms and density plots. 3. **Correlation Analysis**: o Use a sample dataset to calculate the correlation matrix and visualize it using a heatmap.	1
	Total turns	3

Assignment no. 4. Practical Assignments on Statistical / Algorithmic Data Modeling

Objective:

To develop skills in statistical data modeling, hypothesis testing, classification and regression algorithms, model evaluation techniques, and hands-on exercises with the scikit-learn library.

Assignment no.	Assignment statement	Turns
4.1: Hypothesis Testing and Probability Distributions	1. **Hypothesis Testing**: o Perform a hypothesis test to determine if the mean of a sample differs significantly from a known population mean.	1
	2. **Probability Distributions**: o Visualize the probability distribution of a dataset using histograms and probability density functions (PDFs).	1
4.2: Basics of Classification and Regression Algorithms	1. **Classification Algorithm (Logistic Regression)**: o Implement a logistic regression model using scikit-learn to classify the Iris dataset.	2
	2. **Regression Algorithm (Linear Regression)**: o Implement a linear regression model using scikit-learn to predict house prices.	2
4.3: Model Evaluation Techniques	1. **Performance Metrics for Classification**: o Evaluate a classification model using confusion matrix, precision, recall, and F1-score. 2. **Performance Metrics for Regression**: o Evaluate a regression model using mean squared error, mean absolute error, and R-squared.	1
4.4: Hands-on Exercises with scikit-learn Library	1. **Implement a Decision Tree Classifier**: o Train and evaluate a Decision Tree Classifier on the Iris dataset.	1
	2. **Implement a Random Forest Regressor**: o Train and evaluate a Random Forest Regressor on the Boston Housing dataset.	1
	Total Turns	8

Assignment no. 5. Mini-project 4 Turns

The objective of this mini project is to provide hands-on experience with the full data science workflow, encompassing data collection, preprocessing, exploratory data analysis (EDA), modeling, evaluation, and deployment. By the end of this project, participants will have developed a comprehensive understanding of how to apply data science techniques to solve a real-world problem and present their findings effectively.

Total 30