## **Foundations of Data Science: Practical Assignment**

Total Marks: 50

### **Task 1: Introduction to Data Science (10 marks)**

* Data Science Overview (3 marks):
  + Provide a brief explanation of what data science is and its significance in various industries.
* Setting up the Environment (2 marks):
  + Install necessary Python libraries for data science (e.g., NumPy, Pandas, Matplotlib) using Jupyter Notebook cells.
* Loading and Displaying Data (5 marks):
  + Import a dataset of your choice (e.g., CSV or Excel file) using Pandas.
  + Display the first few rows of the dataset to understand its structure.

### **Task 2: Core Concepts and Principles (15 marks)**

* Data Types and Structures (4 marks):
  + Create variables representing different data types (integer, float, string).
  + Explore and demonstrate basic operations on these variables.
* Data Manipulation (5 marks):
  + Apply data manipulation techniques using Pandas.
  + Perform tasks such as filtering, sorting, and grouping on the loaded dataset.
* Feature Engineering (6 marks):
  + Create a new feature based on existing features in the dataset.
  + Explain the rationale behind the feature engineering process.

### **Task 3: Statistical Analysis (10 marks)**

* Descriptive Statistics (4 marks):
  + Compute and interpret key descriptive statistics (mean, median, standard deviation) for relevant columns in the dataset.
* Inferential Statistics (6 marks):
  + Formulate a hypothesis related to the dataset.
  + Conduct a statistical test (e.g., t-test) to test the hypothesis.
  + Provide conclusions based on the test results.

### **Task 4: Exploratory Data Analysis (15 marks)**

* Data Visualization (8 marks):
  + Create at least three visualizations (e.g., histograms, scatter plots, box plots) to explore relationships within the dataset.
  + Include appropriate labels and titles for clarity.
* Outlier Detection (5 marks):
  + Identify and handle outliers in the dataset.
  + Explain the methodology and reasoning behind your approach.
* Correlation Analysis (2 marks):
  + Compute and interpret the correlation between two relevant variables in the dataset.

### **Submission Instructions (2 marks)**

* Jupyter Notebook Submission (2 marks):
  + Compile all tasks in a Jupyter Notebook.
  + Include comments and explanations for each step.
  + Submit the notebook file or provide a GitHub link for review.
  + You will be required to submit this in the Folder for assignments in the Assignment 1.