Pre-requisites:

- Basics of Statistics
- Knowledge of Python or Java programming
- Knowledge of Basics of Statistics
- Knowledge of Customer Segmentation dataset available at Kaggle:
    - https://www.kaggle.com/code/fabiendaniel/customer-segmentation/input
    - https://www.kaggle.com/datasets/kaushiksuresh147/customer-segmentation

Note:

- There will be a pre-test. There will be MCQ of 20-30 questions on the pre-requisites mentioned above.
- There will be a post-assessment test after each module. Each test will consist of 5-10 MCQs.
- There will be a final assignment which is to be completed offline after training is over and submit that to the trainer. Trainer will evaluate and provide the score report.
- All the topics in each module will be explained via hands-on and theoretical concepts using PPT slides. There will be approx. 80% hands-on and 20% theory.
- There will be some offline assignments as well which needs to be completed by participants every day.

Day 1

1) Understanding of Python
    a) What is python?
    b) Python documentation and help
2) Python Environment Set-up and Installation
3) Jupyter Notebook Overview
4) Python Basics
    a) Data Types
        i) Numbers and Booleans
        ii) Strings
        iii) Tuples and Lists
        iv) Dictionaries
5) If, elif and else Statements
6) Loops in Python
7) Errors & Exceptions
    a) Try – except
    b) Assert, Raise
    c) Finally
8) Using NumPy Package in Python
    a) Why use NumPy?
    b) Numpy Arrays
    c) Numpy Array Indexing
    d) Numpy Array Manipulation
    e) Numpy Operations
    f) Broadcasting
    g) Numpy Statistical Functions

9) Using Pandas Package in Python
    a) Series
    b) DataFrames
    c) Missing Data Treatment
    d) Groupby
    e) Merging Joining and Concatenating
    f) Read Excel, JSON, XML files
    g) Data Input and Output
10) Statistics Concepts
    a) Measure of Central Tendency
    b) Variability of Data
        i) Quantify Spread
        ii) Outliers
        iii) IQR
        iv) Sum of Squares
        v) Standard Deviation
        vi) The Standard Deviation "Rule of Thumb”

Day 2

11) What is Machine Learning?
12) Machine Learning Basic Concepts
    a) Importing the Libraries
    b) Importing the Dataset
    c) Summary of Object-oriented programming: classes & objects
    d) Missing Data Treatment
    e) Categorical Data
    f) Splitting the Dataset into the Training set and Test set
    g) Feature Scaling
13) Analytics and Machine Learning
    a) High-Level Concepts
14) Understand the sample data to be used in our machine learning hands-on
15) Plotting
    a) Matplotlib
    b) Seaborn
16) Integration of Charts/Graphs with Web Pages in Web Projects
17) Preprocessing and Feature Engineering
    a) Formatting Models According to Use Case
    b) Transformation
    c) Preprocessing functions
    d) Working with Continuous Features
        i) Scaling and Normalization
        ii) Standard Scaling
    e) Working with Categorical Features
        i) One-Hot Encoding
    f) Feature Manipulation
    g) Feature Selection

Day 3

18) Classification
    a) Use Cases
    b) Types of Classification
    c) Classification Models
    d) Logistic Regression
        i) Logistic Regression Intuition
        ii) Sigmoid Function
        iii) Model Hyperparameters
        iv) Training Parameters
        v) Prediction Parameters
        vi) Example
    e) Decision Trees
        i) Decision Tree Regression Intuition
        ii) Pruning
        iii) Overfitting in Decision Tree
        iv) Entropy
        v) Information Gain
        vi) Model Hyperparameters
        vii) Training Parameters
        viii) Prediction Parameters
    f) Evaluators for Classification and Automating Model Tuning
19) Regression
    a) Use Cases
    b) Regression Models
    c) Linear Regression
        i) Simple Linear Regression Intuition
        ii) RMSE

iii) Model Hyperparameters

iv) Training Parameters

v) Example

vi) Training Summary

d) Decision Trees

i) Model Hyperparameters

ii) Training Parameters

iii) Example

Day 4

- Unsupervised Learning
    - Use Cases
    - Model Scalability
    - k-means
        - k-means Intuition
        - Model Hyperparameters
        - Training Parameters
        - Example
    - k-means Metrics Summary
- Model Selection
    - k-Fold Cross Validation in Python
    - Grid Search in Python
    - k-Fold Cross Validation in R
    - Grid Search in R
- Challenges of Machine Learning
    - Insufficient Quantity of data
    - Non Representative data
    - Poor Quality of data
    - Irrelevant features
    - Overfitting the training data
    - Underfitting the training data
- Model Deployment
    - Model deployment basics
    - Prediction using value
    - Save the model as pkl file
    - Serve model as API using Flask

- Tips and Tricks
    - Using Feature Scaling to Standardize Data
    - Implementing Feature Engineering with Logistic Regression
    - Extracting Data with Feature Selection and Interaction
    - Build Model Based on Real-World Problems
    - Support Vector Machines
    - Implementing kNN on the Data Set
    - Decision Tree as Predictive Model
    - Tricks with Dimensionality Reduction
    - Validation Dataset Tuning
    - Regularizing Model to Avoid Overfitting
    - Perform Metric Selection on Real Data

Day 5

- Troubleshooting
    - Splitting Your Datasets for Train, Test, and Validate
    - Persist Your Hard Earned Models by Saving Them to Disk
    - Transform Your Variable Length Features into One-Hot Vectors
    - Finding the Most Important Features in Your Classifier
    - Predicting Multiple Targets with the Same Dataset
    - Retrieving the Best Estimators after Grid Search
    - Regress on Your Pandas Data Frame with Simple Statsmodels OLS
    - Extracting Decision Tree Rules from scikit-learn
    - Finding Out Which Features Are Important in a Random Forest Model
    - Classifying with SVMs When Your Data Has Unbalanced Classes
    - Computing True/False Positives/Negatives after in scikit-learn
    - Drawing Out a Decision Tree Trained in scikit-learn
- Best Practices for Data
- Best Practices for Training
- Best Practices for Coding
- Best Practices for Deployment
- Best Practices for Team
- Best Practices for Governance
- Best practices for implementing machine learning
    - ML development
    - Data processing
    - Operationalized training
    - Model deployment and serving
    - ML workflow orchestration
    - Artifact organization
    - Model monitoring

- Case Study
    - Select a machine learning problem to solve using machine learning
    - Discuss the problem with each other
    - Implement the solution using Python