Exploratory Data Analysis
Basic

Duration – 1.5 Days / 12 Hours

Program Description

This course introduces the fundamental concepts of Exploratory Data Analysis (EDA) for data science. Participants will learn data types, structures, and descriptive statistics to understand data distributions. The course covers data cleaning techniques, handling missing values, and detecting outliers for better data preprocessing. Basic visualization techniques, such as histograms, boxplots, and scatter plots, are used to analyse patterns and trends. By the end of this module, learners will gain hands-on experience in structuring and preparing data for further analysis.

Learning Goals

- ❖ Understand EDA principles and its role in data science
- ❖ Identify different data types and structures
- ❖ Perform data cleaning and handle missing values
- ❖ Apply summary statistics (mean, median, mode, standard deviation, skewness, kurtosis)
- ❖ Visualize data using histograms, boxplots, and scatter plots
- ❖ Detect and manage outliers and anomalies

Course Topics

- ❖ Introduction to EDA and Its Importance in Data Science
- ❖ Understanding Data Types & Structures
- ❖ Data Cleaning and Handling Missing Values & Outliers
- ❖ Descriptive Statistics & Summary Statistics
- ❖ Data Visualization

Exploratory Data Analysis
Advanced

Duration – 2 Days / 16 Hours

Program Description

This advanced EDA course builds upon fundamental concepts, focusing on data transformation, feature engineering, and multivariate analysis. Participants will explore correlation techniques, dimensionality reduction (PCA, t-SNE, LDA), and hypothesis testing for data validation. The course covers automated EDA techniques using Python libraries like Pandas, Seaborn, and Matplotlib to streamline the analytical process. By the end of this module, learners will be able to extract deep insights from data and prepare it for predictive modelling.

Learning Goals

- ❖ Master advanced data cleaning and feature engineering
- ❖ Apply multivariate analysis and correlation techniques
- ❖ Perform dimensionality reduction using PCA, t-SNE, and LDA
- ❖ Conduct hypothesis testing for data-driven decision-making
- ❖ Automate EDA using Python libraries (Pandas, Seaborn, Matplotlib)
- ❖ Apply SQL for statistical and analytical operations in data science

Course Topics

- ❖ Advanced Data Cleaning & Transformation
- ❖ Feature Engineering & Data Preprocessing
- ❖ Multivariate Analysis & Correlations
- ❖ Dimensionality Reduction Techniques (PCA, t-SNE, LDA)
- ❖ Hypothesis Testing for Data Validation

Note: All modules are enriched with real-world contextualization, using data and challenges from UPS's own operations to ensure immediate relevance and application.