

# DATA ANALYSIS WITH PYTHON

## Part 1: Tools For Data Analysis

### 1. Python Refresher

- Pre-Requisite
- Ipython Interpreter
- Jupyter Notebook
- Basic Data Type
- Collection Type - List
- Collection Type – Dictionaries
- Collection Type – Sets
- Collection Type – Tuples
- Functions
- Classes And Object

### 2. NumPy Core Concept for Data Analysis

- NumPy Introduction
- Arrays Revisited
- Ways to Create Array
- Concept + Lab: NumPy Array Internals
  - Dimensions
  - Shape
  - Strides
- Concept + Lab: Data Type & Casting
- Concept + Lab: Slicing & Indexing
  - Understand Slicing & Indexing 1-D Array
  - Understand Slicing & Indexing Multidimensional Array
- Concept + Lab: Array Operations
  - Common Operations on Array
  - Commonly Used Functions for NumPy Array Operations
- Concept + Lab: Broadcasting
  - Array Broadcasting Principle
  - Understand Usage of Broadcasting
- Concept + Lab: Understand Vectorization

### 3. Pandas Core Concept for Data Analysis

- What is Pandas
- Data Frames
- Data Frames Basic
- Handling Missing Data
- GroupBy
- Aggregation
- Transform
- Window Functions
- Filter
- Join Merge & Concat
- Apply Method
- Data Frames Reshape

- Calculate Frequency Distribution

## **Part 2: Data Analysis Core Concept**

- 1. What is Data**
- 2. What is Data Set**
- 3. Types of Variables**
  - Definition
  - Types of Data Types
    - Categorical
    - Numeric
  - Why Data Types are Important
- 4. How do you collect Information for Different Data**
- 5. Descriptive Statistics Concepts**
  - Type of Statistics
  - What it is
  - Concept: Understand Normal Distribution
  - Concept: Central Tendency
  - Concept Measures of Variability
  - Concept: Variance & Standardized Score
  - Concept: Z-score or Standardized Score
  - Concept: Modality
  - Concept: Skewness
  - Concept: Kurtosis

## **Part 3: Tools for Data Visualization**

- 1. Matplotlib Introduction**
- 2. Matplotlib Architecture**
- 3. Seaborn Plot Overview**
- 4. Parameters of Plot**
- 5. Types of Plot by purpose**
  - Correlation
    - What it is
    - Type of Graphs in Correlation Category
      - Scatter Plot
      - Scatter Plot with Line of Best Fit
      - Counts Plot
      - Marginal Boxplot
      - Correlogram
      - Pairwise Plot
  - Deviation
    - Diverging Bars
    - Diverging Dot Plot
  - Ranking
    - Ordered Bar Chart

- Dot Plot
- Distribution
  - Histogram for Continuous Variable
  - Histogram for Categorical Variable
  - Density Curves with Histogram
  - Box Plot
  - Dot + Box Plot
  - Categorical Plots
- Composition
  - Pie Chart
  - Treemap
  - Bar Chart
- Change
  - Time Series Plot
  - Time Series Decomposition Plot

## **Part 4: Step by Step EDA & Data Preparation Workflow with Project**

- 1. What is EDA**
- 2. Value of EDA**
- 3. Steps of Data Exploration & Preparation**

- Step 1: Variable Identification
- Step 1: Univariate Analysis
- Step 1: Bi-Variate Analysis
- Step 1: Missing Value Treatment
- Step 1: Outlier Detection & Treatment
- Step 1: Variable Transformation
- Step 1: Variable Creation