# Python for Data Analysis - Training Outline

**Course Title**

Python for Data Analysis

**Target Audience**

- Aspiring data analysts
- Excel users transitioning to Python
- Business analysts, researchers, or professionals working with data

**Course Duration**

4-6 Weeks | 2-3 Sessions per Week | 2 Hours per Session

**Module 1: Introduction to Python for Data Analysis**

- Why Python? Benefits in data analysis
- Installing Anaconda and setting up Jupyter Notebooks
- Python IDEs: Jupyter, VS Code, Google Colab
- Python syntax essentials: variables, data types, and operators

Hands-On: Writing your first Python script in Jupyter Notebook

**Module 2: Python Fundamentals for Data Manipulation**

- Data structures: Lists, Tuples, Dictionaries, and Sets
- Loops and control flow (for, while, if-else)
- Functions and Lambda expressions
- List comprehensions

Mini Project: Simulate a basic sales log and analyze totals

**Module 3: Working with Libraries - NumPy**

- Understanding arrays and their advantages
- Creating and manipulating NumPy arrays
- Array operations and broadcasting
- Indexing, slicing, reshaping

Exercise: Compare Excel formulas with NumPy operations

**Module 4: Data Analysis with Pandas**

- Introduction to Pandas Series and DataFrames
- Reading and writing data (CSV, Excel, JSON)
- Filtering, sorting, and subsetting data
- Handling missing data and duplicates

Hands-On: Cleaning a sample customer dataset

**Module 5: Data Aggregation and Grouping**

- GroupBy and aggregation techniques
- Pivot tables and cross-tabulations
- Multi-indexing in Pandas

Exercise: Analyze sales by region, product, and category

**Module 6: Data Visualization with Matplotlib & Seaborn**

- Introduction to data visualization
- Plot types: line, bar, histogram, scatter, boxplot
- Styling and customizing plots
- Correlation and distribution plots

Project: Create visual dashboards using Seaborn

**Module 7: Real-World Data Cleaning Techniques**

- Text and date/time data handling
- String operations in Pandas
- Parsing and formatting datetime
- Regular expressions basics

Case Study: Clean a messy HR dataset (names, dates, departments)

**Module 8: Exploratory Data Analysis (EDA)**

- What is EDA and why it matters

- Descriptive statistics
- Outlier detection
- Data transformation (scaling, binning)

Mini Project: Perform EDA on a public dataset (e.g., Titanic or Superstore)

**Module 9: Intro to Working with APIs and Web Data**

- What are APIs?
- Requests and JSON parsing
- Importing data from websites and APIs

Exercise: Import live COVID-19 or weather data using APIs

**Module 10: Capstone Project and Presentation**

- Capstone project: Choose from 2-3 real-world datasets (e.g., Sales analysis, customer churn, HR attrition)
- Prepare and present findings using EDA + visuals
- Peer review and trainer feedback

**Final Deliverables**

- Completed Jupyter notebooks
- One fully documented data analysis project
- Optional: GitHub portfolio update

**Optional Add-ons (Advanced or Custom Topics)**

- Introduction to SQL with Pandas
- Basics of Machine Learning with Scikit-learn
- Integrating Python with Excel using openpyxl/xlwings