

Python Associate Data Analyst with AI and Data Visualization

Duration: 7 Days

Prerequisites: Knowledge of Python Programming

Day 1

Module 1: Data Analytics Overview

- Data Analytics Process
- Data Analysis Stages
- Data Analytics Process
- Python for Data Analysis
- NumPy
- Pandas
- Matplotlib

Module 2: Exploratory Data Analysis (EDA)

- What is EDA
- EDA – Quantitative Technique
- EDA – Graphical Technique
- Data Analysis Predictions
- Data Analysis Communication
- Data Types for Plotting

Module 3: Data Ingestion and Preparation

- Reading data from various sources (CSV, Excel, databases)
- Data cleaning
- Handling missing values
- Data transformation
- Data normalization
- Data Standardization
- Handling Outliers
- Feature Engineering
- Feature Transformation

Day 2

Module 4: Statistical Analysis

- Descriptive Statistics
- Data visualization with Matplotlib and Seaborn
- Understanding data distribution
- Identifying patterns and trends
- Statistical Analysis Considerations
- Data Distribution

Module 5: Data Manipulation and Aggregation

- Grouping and aggregation using Pandas
- Pivot tables and cross-tabulation
- Handling categorical data
- Data merging and joining

Module 6: Time Series Analysis

- Introduction to time series data
- Time series indexing and slicing
- Time series visualization
- Stationarity and differencing
- Forecasting techniques (ARIMA, etc.)

Day 3

Module 7: Introduction to Machine Learning

- Supervised vs unsupervised learning
- Model evaluation metrics
- Overfitting and underfitting

Module 8: Linear Regression

- Simple linear regression
- Multiple linear regression
- Model evaluation and interpretation

Module 9: Logistic Regression

- Logistic regression model
- Model evaluation and interpretation
- Applications of logistic regression

Day 4

Module 10: Decision Trees and Random Forests

- Decision trees
- Random forests
- Model evaluation and interpretation

Module 11: Model Evaluation and Selection

- Model evaluation metrics
- Hyperparameter tuning
- Cross-validation
- Model selection techniques

Module 12: Data Analysis Case Study

- Real-world data analysis project
- Data cleaning
- Data Exploration,
- Data Visualization
- Model Building
- Model Evaluation
- Insights

Day 5

Module 13: Python for Data Visualization

- Advanced data visualization with Matplotlib
- Interactive visualizations with Plotly
- Geospatial data visualization
- Dashboard creation

Module 14: Big Data with Python

- Introduction to big data concepts
- PySpark
- Big data processing
- Big Data Analysis

Module 15: Clustering and Dimensionality Reduction

- Support Vector Machines (SVM)
- Clustering algorithms
- K-means Clustering
- Hierarchical clustering)
- Dimensionality reduction (PCA)
- Applications in data exploration and visualization

Day 6:

Module 16: PlotLine and GGPlot 2

- Data Visualization using Plotnine and ggplot2
- What is ggplot and its Use?
- Creating a Scatter Plot
- Customizing Plot Aesthetics
- Changing Colors and Shapes
- Adjusting Axis Labels and Titles
- Adding Legends and Annotations
- Customizing Themes and Templates
- Applying Predefined Themes
- Creating Custom Themes
- Saving and Sharing Plots

Module 17: Natural Language Processing

- Introduction
- Example Sentiment Analysis

Module 18: Large Language Models

- Introduction to LLM and its applications
- Importance of LLM
- Working of LLM
- Future of LLM

Day 7:

Module 19: Agentic AI

- Introduction
- Evolution of Agentic AI
- Components of Agentic AI
- Use Cases
- Risk and Future of Agentic AI

Module 20: Prompt Engineering

- Introduction to Prompt Engineering
- What is prompt in AI
- Need of prompt engineering

- Types of Prompts
- Use Cases and examples of Prompt Engineering
- Strategies for writing better prompts