Certified Associate Data Analyst with Python

(PCAD)

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

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.)

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

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

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