



Python for Data Science/Analyst

Duration: 40-hour

Prerequisites:

This course assumes that participants have a foundational understanding of Python.

The focus is on using Python for data manipulation, analysis, and visualization.

This "Python for Data Science/Analyst" course covers essential data science concepts and tools using Python. It includes hands-on projects to reinforce learning and practical applications in various domains. Adjustments can be made based on the specific needs and pace of the participants.

Module 1: Introduction to Data Science and Python Review

- Session 1: Introduction to Data Science
 - Overview of data science and its applications
 - Python tools for data science
- Session 2: Review of Python for Data Science
 - Refreshing Python basics
 - Introduction to Jupyter Notebooks
- Session 3: Data Exploration with Pandas
 - Loading and exploring datasets
 - Basic data manipulation with Pandas

Module 2: Data Cleaning and Preprocessing

- Session 4: Handling Missing Data
 - Identifying and handling missing values
 - Imputation techniques
- Session 5: Data Cleaning Techniques
 - Outlier detection and removal
 - Data transformation and normalization
- Session 6: Feature Engineering
 - Creating new features
 - Encoding categorical variables

Module 3: Exploratory Data Analysis (EDA) and Data Visualization

- Session 7: Exploratory Data Analysis with Seaborn
 - Visualizing data distributions
 - Correlation analysis
- Session 8: Data Visualization with Matplotlib
 - Advanced plotting techniques
 - Customizing visualizations
- Session 9: Interactive Visualizations with Plotly
 - Building interactive plots for data exploration

Module 4: Statistical Analysis with Python

- Session 10: Descriptive Statistics
 - Measures of central tendency and dispersion
 - Summary statistics
- Session 11: Hypothesis Testing
 - Introduction to hypothesis testing
 - Conducting t-tests and chi-square tests
- Session 12: Regression Analysis
 - Simple and multiple linear regression
 - Model evaluation and interpretation

Module 5: Machine Learning for Data Science

- Session 13: Introduction to Machine Learning
 - Applications of machine learning in data science
 - Types of machine learning algorithms
- Session 14: Scikit-Learn for Machine Learning
 - Using Scikit-Learn for classification and regression
 - Model evaluation and validation
- Session 15: Clustering and Dimensionality Reduction
 - K-means clustering and hierarchical clustering
 - Principal Component Analysis (PCA)

Module 6: Time Series Analysis and Forecasting

- Session 16: Introduction to Time Series Analysis
 - Time series data and its characteristics
 - Decomposition and trend analysis
- Session 17: Time Series Forecasting with ARIMA
 - Autoregressive Integrated Moving Average (ARIMA) models
 - Forecasting future values
- Session 18: Hands-on Project: Time Series Forecasting
 - Applying time series analysis techniques to real-world data

Module 7: Text Analytics and Sentiment Analysis

- Session 19: Introduction to Text Analytics
 - Tokenization and text preprocessing
 - Basic text analysis with NLTK
- Session 20: Sentiment Analysis
 - Analyzing sentiment in text data
 - Practical applications in business
- Session 21: Hands-on Project: Text Analytics
 - Applying text analytics techniques to analyze sentiment



Module 8: Final Project and Review

- Session 22: Capstone Data Science Project
 - Applying data science techniques to a real-world problem
 - Project presentation and discussion
- Session 23: Review and Q&A
 - Recap of key data science concepts
 - Addressing participant questions and concerns
- Session 24: Course Conclusion and Certificates
 - Course Summary
 - Certificate distribution