ACE IT CAREERS
BEST CAREER BUILDER

DATA & ANALYTICS · COURSE SYLLABUS

Data Science with Python

From Python fundamentals and data analysis to statistics, visualization and machine learning, ending with a complete end-to-end project. 10 modules and about 40 to 50 hours of live training.

40–50 Hours | Live online | Placement support

Syllabus

MODULE 1

Introduction to data science

- What is data science?
- Data science vs data analytics vs machine learning vs AI
- Applications in banking, healthcare, retail, education and e-commerce
- Structured, semi-structured and unstructured data
- The data science lifecycle
- Datasets, features, labels and observations
- Setting up the Python environment

Practical: Explore a student-marks dataset and identify patterns in performance.

MODULE 2

Python programming fundamentals

- Variables, data types, operators and type conversion
- Conditional statements and loops
- Strings, lists, tuples, sets and dictionaries
- Functions, arguments, return values and lambda functions
- List comprehensions and exception handling
- Modules and packages
- Reading and writing CSV and text files
- Introduction to classes and objects

Practical: Calculate student grades, find highest and average marks, count word frequencies and summarize salaries from a CSV file.

MODULE 3

NumPy for numerical computing

- One-dimensional and multidimensional arrays
- Indexing, slicing and reshaping
- Vectorized operations and broadcasting
- Aggregate functions: sum, mean, min, max, std
- Boolean indexing
- Random numbers and basic matrix operations

Practical: Analyze examination marks with NumPy arrays, without repetitive loops.

MODULE 4

Pandas for data analysis (key module)

- Series and DataFrames
- Importing CSV and Excel files
- Exploring data with head, tail, info and describe
- Selecting, filtering, sorting and ranking
- Handling missing values and duplicates
- apply, groupby and aggregation
- merge, join and concat
- Dates, categorical data, pivot tables and cross-tabulation
- Exporting to CSV and Excel

Practical: Analyze an employee dataset: headcount by department, average salary, top earners and missing data.

MODULE 5

Data cleaning and exploratory data analysis

- Assessing dataset quality
- Missing, inconsistent, duplicate and invalid records
- Outliers and their causes
- Univariate, bivariate and multivariate analysis
- Correlation, covariance and frequency distributions
- Turning business questions into analytical tasks

Practical: Analyze retail sales: best sellers, monthly trends and underperforming categories.

MODULE 6

Data visualization

- Matplotlib line, bar, histogram, pie, scatter and box plots
- Titles, labels, legends, annotations and subplots
- Seaborn count, distribution, box and violin plots
- Heatmaps and correlation matrices
- Choosing the right chart and avoiding misleading visuals

Practical: Build a sales-analysis report with at least five different charts.

MODULE 7

Statistics and probability

- Population and sample
- Mean, median, mode, range, variance and standard deviation
- Percentiles, quartiles and the normal distribution
- Probability rules, conditional probability and Bayes' theorem
- Sampling bias and the central limit theorem
- Confidence intervals and hypothesis testing
- p-values, correlation vs causation and A/B testing

Practical: Compare the average spending of two customer groups and test whether the difference is significant.

MODULE 8

Introduction to machine learning

- Supervised and unsupervised learning
- Regression, classification and clustering
- Training and testing data, features and targets
- Overfitting and underfitting
- Scikit-learn and the machine learning workflow

Practical: Train a simple model to predict house prices.

MODULE 9

Machine learning algorithms with Scikit-learn

- Linear and polynomial regression; MAE, MSE, RMSE and R^2
- Logistic regression, KNN, decision trees and random forests
- Confusion matrix, accuracy, precision, recall and F1-score
- K-means clustering and PCA
- Feature scaling and encoding categorical variables
- Cross-validation, hyperparameter tuning and data leakage

Practical: Predict house prices, predict customer churn and segment customers by purchasing behavior.

MODULE 10

End-to-end project and career preparation

- Understanding the business problem
- Cleaning, exploring and visualizing a real-world dataset
- Selecting, training and evaluating a model
- Documenting findings and limitations
- Presenting results to a non-technical audience
- Building a project portfolio and discussing it in interviews

Practical: Final deliverable: a complete Jupyter Notebook with code, visualizations, analysis and a project report.

What you will be able to do

- Write Python programs for data processing and analysis
- Work with NumPy arrays and Pandas DataFrames
- Import, clean, transform and analyze real-world datasets

- Create meaningful visualizations with Matplotlib and Seaborn
- Apply descriptive statistics and introductory probability
- Build and evaluate machine learning models with Scikit-learn
- Complete an end-to-end data science project and explain the findings

Included with every course

- Live instructor-led classes in US time zones, with session recordings
- Real-time project work and a certificate
- Resume preparation and marketing, mock interviews and placement support