

Course Structure: Data Science & Analytics

Fundamentals

Key Areas Covered

Introduction to Data Science & Analytics

- Understanding the role of Data Science in modern business decision-making.
- Key differences between Data Science, Data Analytics, and Data Engineering.
- Real-world use cases across industries.

Data Types and Formats

- Structured, Semi-Structured, and Unstructured Data.
- Common data formats: CSV, JSON, Excel, SQL databases.

The Data Science Lifecycle

- Data Collection: Gathering data from various sources (databases, APIs, web scraping).
- Data Cleaning: Removing inconsistencies, handling missing data, and ensuring quality.
- Data Analysis: Identifying trends and patterns.
- Data Visualization: Presenting insights through graphs, charts, and dashboards.
- Interpretation and Communication of Results.

Tools for Data Science and Analytics

- Programming Languages: Python, R.
- Data Manipulation Libraries: Pandas, NumPy.
- Visualization Libraries: Matplotlib, Seaborn, Tableau.
- SQL for database querying and manipulation.

Introduction to Statistics

- Basic statistical measures (mean, median, mode, standard deviation).
- Probability basics and its application in data analysis.

Problem-Solving with Data

- Formulating data-driven questions.
- Exploratory Data Analysis (EDA) to uncover insights.

Ethics in Data Science

- Understanding bias in data.
- Privacy and confidentiality in handling datasets.

Python for Data Science

Basics:

- Introduction to Python Programming
- Variables, Data Types, and Operators
- Control Structures (if-else, loops)
- Functions and Modules

Core:

- Introduction to Python Libraries: NumPy, Pandas, and Matplotlib
- Data Manipulation using Pandas
- Working with Arrays using NumPy
- File Handling and Exception Handling

Intermediate:

- Exploratory Data Analysis (EDA) Techniques
- Data Visualization with Seaborn and Matplotlib
- APIs and JSON Data Processing
- Basic Statistics using Python

Advanced:

- Advanced Data Wrangling
- Database Integration with Python (SQLAlchemy, SQLite)
- Automation and Script Development
- Real-World Applications and Case Studies

Statistics and Probability

Basics:

- Introduction to Statistics
- Measures of Central Tendency (Mean, Median, Mode)
- Measures of Dispersion (Variance, Standard Deviation)

Core:

- Probability Concepts and Distributions
- Hypothesis Testing
- Correlation and Regression Analysis

Intermediate:

- Sampling Methods and Techniques
- Analysis of Variance (ANOVA)
- Time Series Analysis

Advanced:

- Multivariate Statistics
- Bayesian Statistics
- Statistical Model Validation

Data Visualization**Basics:**

- Introduction to Data Visualization
- Fundamentals of Visualizing Data
- Creating Basic Plots (Bar, Line, Scatter)

Core:

- Data Visualization with Matplotlib
- Advanced Visualizations with Seaborn
- Dashboard Creation with Tableau

Intermediate:

- Interactive Visualizations using Plotly
- Storytelling with Data Visualizations
- Heatmaps, Pair Plots, and Other Advanced Charts

Advanced:

- Building Data Dashboards
- Best Practices for Effective Visualization
- Real-World Visualization Projects

Machine Learning Basics**Basics:**

- Introduction to Machine Learning
- Types of Machine Learning (Supervised, Unsupervised, Reinforcement)

- Preparing Data for Machine Learning

Core:

- Linear Regression and Logistic Regression
- Clustering Techniques (K-Means, Hierarchical)
- Feature Engineering and Scaling

Intermediate:

- Decision Trees and Random Forests
- Model Evaluation and Metrics (Confusion Matrix, ROC Curve)
- Dimensionality Reduction (PCA)

Advanced:

- Introduction to Neural Networks
- Basics of Deep Learning
- Building End-to-End Machine Learning Models

SQL for Data Analysis

Basics:

- Introduction to SQL
- Writing Basic Queries (SELECT, INSERT, UPDATE, DELETE)
- Filtering and Sorting Data

Core:

- Joins (Inner, Outer, Cross, Self)
- Aggregations and Grouping
- Subqueries and Nested Queries

Intermediate:

- Window Functions
- Common Table Expressions (CTEs)
- Advanced Query Optimization

Advanced:

- Integration with Python for Analysis
- Building and Managing Databases
- Real-World SQL Scenarios and Use Cases

Big Data Tools

Basics:

- Introduction to Big Data
- Basics of Distributed Systems
- Overview of Hadoop Ecosystem

Core:

- Working with HDFS
- Basics of Apache Spark
- Data Processing with MapReduce

Intermediate:

- Real-Time Processing with Spark Streaming
- Hive and Pig for Big Data Analysis
- Introduction to Cloud Platforms for Big Data (AWS, Azure, GCP)

Advanced:

- Handling Large Datasets with Spark MLlib
- Optimizing Big Data Pipelines
- Big Data Project Implementation