

# ■ Data Science Fundamentals

## 1. Introduction to Data Science

- What is Data Science?
- Importance in modern industries (business, healthcare, AI, etc.)
- Data Science lifecycle: data collection → cleaning → analysis → visualization → interpretation
- Difference between Data Science, Machine Learning, and AI

## 2. Data Science Tools & Environment

- Python basics for data science
- Jupyter Notebook overview
- Popular libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
- Introduction to R and SQL for data handling

## 3. Data Collection & Cleaning

- Data sources: CSV, databases, APIs, web scraping
- Handling missing values, duplicates, and outliers
- Data type conversion and formatting
- Feature selection and engineering basics

## 4. Exploratory Data Analysis (EDA)

- Descriptive statistics (mean, median, mode, standard deviation)
- Data visualization: bar charts, histograms, scatter plots, box plots
- Identifying trends, patterns, and anomalies

## 5. Introduction to Machine Learning

- Supervised vs. Unsupervised Learning
- Simple algorithms: Linear Regression, K-Means Clustering, Decision Trees
- Model training, testing, and evaluation metrics

## 6. Data Visualization & Reporting

- Using Matplotlib and Seaborn for advanced visualizations
- Creating dashboards (intro to Power BI/Tableau)
- Preparing clear reports for stakeholders

## 7. Ethics in Data Science

- Data privacy and security
- Avoiding bias in datasets
- Responsible data usage

## Learning Outcomes

- Understand the data science workflow from collection to reporting
- Use Python libraries for data cleaning, analysis, and visualization
- Perform basic statistical analysis and EDA
- Implement simple machine learning models
- Present data insights in a clear and effective manner

## Suggested Materials for Upload

- data\_science\_fundamentals.pdf (this outline in PDF form)
- sample\_dataset.csv (for practice)
- eda\_project.ipynb (Jupyter Notebook example)
- 1-2 YouTube video links for beginner tutorials