

PYTHON FOR DATA SCIENCE 30-HOUR SYLLABUS

Day 1

Introduction to Data Science and Python Basics (6 hours)

1. Introduction to Data Science (1 hour)

- What is Data Science?
- Importance and applications of Data Science in various industries
- Overview of the data science workflow

2. Setting Up the Environment (1 hour)

- Installing Anaconda and Jupyter Notebook
- Overview of Python libraries for data science (NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn)

3. Basic Python Syntax (2 hours)

- Variables, data types, and basic operations
- Control structures (if statements, loops)
- Functions (focus on built-in functions relevant to data science)

4. Data Structures Overview (2 hours)

- Introduction to lists, dictionaries, and sets
 - Basic operations on data structures
 - Importance of data structures in data manipulation
-

Day 2

Data Manipulation with Pandas (6 hours)

1. Introduction to Pandas (1 hour)

- Overview of Pandas and its importance in data manipulation
- Understanding DataFrames and Series

2. Data Import and Export (1 hour)

- Reading data from CSV, Excel, and JSON files
- Writing data to files

3. Data Cleaning and Preparation (2 hours)

- Handling missing values (drop, fill)
- Data transformation (filtering, sorting, and renaming columns)
- Data type conversions

4. Data Analysis with Pandas (2 hours)

- Descriptive statistics (mean, median, mode, standard deviation)
 - Grouping and aggregating data
 - Using pivot tables for data summarization
-

Day 3

Data Visualization (6 hours)

1. Introduction to Data Visualization (1 hour)

- Importance of data visualization in data analysis
- Overview of Matplotlib and Seaborn

2. Basic Plotting with Matplotlib (2 hours)

- Creating line plots, bar charts, and histograms
- Customizing plots (titles, labels, legends)

3. Advanced Visualization with Seaborn (2 hours)

- Creating complex visualizations (box plots, violin plots, heatmaps)
- Customizing Seaborn plots for better insights

4. Project: Visualizing a Dataset (1 hour)

- Choose a dataset and create visualizations to explore insights
 - Present findings to the class
-

DAY-4

Introduction to NumPy and Statistical Analysis (6 hours)

1. Introduction to NumPy (1 hour)

- Overview of NumPy and its role in data science
- Understanding arrays and their advantages over lists

2. Array Operations (2 hours)

- Creating and manipulating NumPy arrays
- Basic mathematical operations on arrays

3. Statistical Analysis (2 hours)

- Basic statistics (mean, median, mode, variance, standard deviation)
- Correlation and covariance analysis

4. Project: Statistical Analysis of a Dataset (1 hour)

- Perform statistical analysis on a chosen dataset
- Present findings and insights

DAY -5

Introduction to Machine Learning (6 hours)

1. Introduction to Machine Learning (1 hour)

- Overview of machine learning concepts
- Types of machine learning (supervised vs. unsupervised)

2. Scikit-Learn Basics (2 hours)

- Installing and importing Scikit-Learn
- Understanding datasets and features

3. Building a Simple Model (2 hours)

- Linear regression example: understanding the model and its application
- Model evaluation (train/test split, metrics like RMSE and R^2)

4. Final Project: End-to-End Data Science Project (1 hour)

- Choose a dataset, perform data cleaning, analysis, visualization, and build a simple model
- Present the project to the class