

Data Science with Python Syllabus

Week 1: Python Fundamentals for Data Science

Day 1: Introduction to Python

Installing Python and IDEs (Anaconda, Jupyter Notebook)

Basic syntax, variables, and data types

Operators and expressions

Day 2: Data Structures and Control Flow

Lists, tuples, dictionaries, and sets

Conditionals (if-else), loops (for, while)

List comprehensions and lambda functions

Day 3: Functions and Modules

Writing and calling functions

Scope and namespaces

Importing libraries (`math`, `os`, `random`)

Day 4: Introduction to NumPy

Arrays and array operations

Indexing, slicing, and reshaping

Basic mathematical operations

Day 5: Working with Pandas

DataFrames and Series

Reading/writing data (CSV, Excel, JSON)

Data exploration: `head()`, `info()`, `describe()`

Week 2: Data Manipulation and Visualization

Day 1: Data Cleaning with Pandas

Handling missing data

Data filtering and selection

Grouping and aggregating data

Day 2: Exploratory Data Analysis (EDA)

Descriptive statistics

Identifying patterns in data

Day 3: Data Visualization with Matplotlib

Line plots, bar charts, and scatter plots

Customizing plots (titles, legends, and colors)

Day 4: Advanced Visualization with Seaborn

Pair plots, heatmaps, and distribution plots

Customizing themes and palettes

Day 5: Mini Project

Conduct EDA and create visualizations on a sample dataset (e.g., Titanic dataset or a sales dataset)

Week 3: Probability, Statistics, and Data Wrangling

Day 1: Probability Basics

Introduction to probability concepts

Conditional probability and Bayes' theorem

Day 2: Descriptive Statistics

Measures of central tendency (mean, median, mode)

Measures of dispersion (variance, standard deviation)

Day 3: Inferential Statistics

Hypothesis testing

Confidence intervals

Day 4: Advanced Data Wrangling

Pivot tables and cross-tabulations

Time series analysis basics

Day 5: Introduction to APIs

Using Python to fetch data from APIs

Parsing JSON data

Week 4: Introduction to Machine Learning

Day 1: Overview of Machine Learning

Types of ML (Supervised, Unsupervised, Reinforcement)

The ML pipeline

Day 2: Supervised Learning with Scikit-learn

Regression models: Linear Regression

Classification models: Logistic Regression

Day 3: Model Evaluation

Metrics: MAE, MSE, RMSE, accuracy, precision, recall, F1-score

Train-test split and cross-validation

Day 4: Unsupervised Learning

Clustering: K-Means, DBSCAN

Dimensionality reduction: PCA

Day 5: Mini Project

Build and evaluate a simple ML model (e.g., predict housing prices or classify iris species)

Week 5: Advanced Topics in Data Science

Day 1: Feature Engineering

Handling categorical variables (encoding)

Scaling and normalizing data

Day 2: Time Series Analysis

Decomposing time series

ARIMA model basics

Day 3: Natural Language Processing (NLP)

Text preprocessing (tokenization, stemming, lemmatization)

Bag-of-Words and TF-IDF

Day 4: Big Data and PySpark Introduction

Introduction to Big Data concepts

Basics of PySpark for large-scale data processing

Day 5: Advanced Python for Data Science

Introduction to web scraping with BeautifulSoup

Automation with Python (e.g., scheduling tasks)

Week 6: Capstone Project and Deployment

Day 1: Capstone Project Introduction

Select a real-world dataset and define objectives

Day 2: Data Preparation

Clean and preprocess the dataset for analysis

Day 3: Model Building and Evaluation

Train, validate, and test ML models

Day 4: Deployment Basics

Using Flask or Streamlit for creating a simple web app

Deploying the app on platforms like Heroku

Day 5: Presentation and Feedback

Present the capstone project results

Get feedback and discuss improvements

Additional Notes

Prerequisites: Familiarity with basic programming concepts is recommended.

Tools Used: Jupyter Notebook, Google Colab, Python libraries (NumPy, Pandas, Matplotlib, Seaborn, scikit-learn).

Resources Provided: Datasets, templates for assignments, and project guidance.

Learning Outcome: By the end of the course, students will have a strong foundation in Python for data science, hands-on experience in data manipulation and visualization, and the ability to build and deploy machine learning models.