

<b>Course Title</b>	Data Science and Machine Learning using Python
<b>Target Audience</b>	This course is ideal for aspiring and current data scientists, career switchers, and professionals looking to expand their skills
<b>Prerequisites (if any)</b>	Participants should have a basic working knowledge of Microsoft Excel and familiarity with handling data.
<b>Language of Instruction</b>	English and Urdu

Course Description
<p>This course seamlessly blends essential Python programming, hands-on data exploration, and practical machine learning concepts for a comprehensive learning experience. Students will explore libraries like Pandas, NumPy, Matplotlib, and scikit-learn to manipulate, understand, and build predictive models with your data, and will learn to leverage generative AI support for code generation, troubleshooting, and concept understanding.</p> <p>In this course students will focus on:</p> <p><b>Module 1: Python and Data Fundamentals</b>  In this module, students will dive into the essentials of Python for data science. They will learn about variables, data types, how to control the flow of their code with conditionals and loops, and how to build modular code using functions. Additionally, they'll be introduced to the cornerstone libraries of data science – NumPy for numerical operations and Pandas for working with tabular data in Data Frames. Students will learn how to load datasets, perform basic data cleaning, and transformations. To tie it all together, students will be guided on setting up a Jupyter Notebook, the preferred working environment for data scientists.</p> <p><b>Module 2: Exploratory Data Analysis (EDA)</b>  The power of data lies in understanding the story it tells. In this module, students will master the art of Exploratory Data Analysis (EDA). Students will learn techniques for handling missing data and outliers, and how to convert data into appropriate formats. Students will calculate essential summary statistics with NumPy and Pandas, uncovering measures like mean, median, and standard deviation. The focus then shifts to visualization. Students will harness Matplotlib and Seaborn to create histograms, scatterplots, and boxplots, learning to interpret these to glean insights from their data. They will solidify these skills with an EDA mini project, where they'll take a dataset from start to finish.</p> <p><b>Module 3: Introduction to Machine Learning</b>  Students will embark on the exciting world of machine learning! This module introduces the fundamentals. They will understand the differences between supervised and unsupervised learning, as well as classification and regression tasks, illustrated with real-world examples. Students will deep dive into linear</p>

regression, learning how this model works, implementing it with scikit-learn, and interpreting the results. Next, they will explore decision trees - how they are built, visualized, and understood. Finally, model selection concepts like train/test splits, overfitting, and cross-validation will be introduced.

**GENERATIVE AI Support Throughout**

Students will have seamless support through integrated GENERATIVE AI assistance. They will get tailored code examples for common operations, help with troubleshooting errors, and easy-to-understand explanations for complex concepts, making their learning journey smoother.

Course Learning Outcomes	
	By the end of this course, the students should be able to:
LO1:	Python Fluency: Demonstrate proficiency in core Python concepts (variables, data types, control flow, functions) for data science tasks.
LO2:	Data Handling Expertise: Utilize Pandas to effectively import, clean, transform, and manipulate datasets for analysis and modelling.
LO3:	Exploratory Analysis Mastery: Employ NumPy, Matplotlib, and Seaborn to calculate summary statistics and create informative visualizations, extracting meaningful insights from data.
LO4:	Machine Learning Foundations: Understand the principles of supervised learning and build basic linear regression and decision tree models using scikit-learn. Evaluate model performance using appropriate metrics.
LO5:	Process-Oriented Mindset: Apply a structured workflow to a data science project encompassing data cleaning, exploratory analysis, model selection, and result interpretation.

Course Summary			
Week	Module Name	Key Concepts/Topics Covered	Assessments

Week 1	Python and Data Fundamentals	<ul style="list-style-type: none"> <li>• Introduction to Python, Data Types, Variables, Operators, NumPy Basics</li> <li>• Introduction to Pandas</li> </ul>	Short coding quizzes Mini-data cleaning exercise
Week 2	Python and Data Fundamentals (Cont.)	<ul style="list-style-type: none"> <li>• Control Flow, Functions</li> <li>• Pandas data selection and transformation</li> </ul>	Practice project: Data cleaning and manipulation with Pandas
Week 3	EDA	<ul style="list-style-type: none"> <li>• Understanding statistical data analysis concepts</li> <li>• Summary Statistics</li> <li>• Data aggregation</li> </ul>	Quiz on EDA concepts
Week 4	EDA (Cont.)	<ul style="list-style-type: none"> <li>• Confidence interval and hypothesis testing</li> <li>• Visualization</li> </ul>	EDA mini-project progress check-in
Week 5	Intro to Machine Learning	<ul style="list-style-type: none"> <li>• Machine learning workflows</li> <li>• Supervised learning: classification and regression, including linear and logistic regression, decision trees</li> <li>• Unsupervised learning, including k-means clustering.</li> </ul>	Quiz on fundamentals
Week 6	Model Building	<ul style="list-style-type: none"> <li>• Model selection concepts, training, validation and testing</li> <li>• Capstone Project</li> </ul>	Capstone Project: Peer feedback and evaluation

### Supplementary Material/Reading Material

- **Python:**
  - Learn Python (<https://www.learnpython.org/>)
  - "Automate the Boring Stuff with Python" (<https://automatetheboringstuff.com/>)
- **Data Analysis and Visualization**
  - Kaggle Datasets (<https://www.kaggle.com/datasets>)
  - Python Data Science Handbook (<https://jakevdp.github.io/PythonDataScienceHandbook/>)
  - "Storytelling with Data" (<https://www.storytellingwithdata.com/>)
  - Flourish (<https://flourish.studio/>)

- **Machine Learning**

- "Introduction to Statistical Learning" (<https://www.statlearning.com/>)
- Towards Data Science Blog (<https://towardsdatascience.com/>)

- **Generative AI**

- OpenAI API Documentation (<https://beta.openai.com/docs>)
- "Coding with ChatGPT" (<https://medium.com/@tanyamarleytsui/coding-with-chatgpt-b50ab3fcb45f>)
- Democratizing access to AI-enabled coding with Colab (<https://blog.google/technology/ai/democratizing-access-to-ai-enabled-coding-with-colab>)