Faysal Aziz

Python for Data Science

From Fundamentals to Linear Regression

FA-0382 · four-day

Software Development · Data & Analytics

Introduction

This four-day intensive course offers a deep dive into Python programming with a focus on its application in data science. It covers everything from the fundamentals of Python programming to specific data science skills like data acquisition using APIs, exploratory data analysis, visualization techniques, and building predictive models with simple linear regression.

Google Colaboratory will serve as the primary learning platform, providing a hands-on experience through interactive notebooks. The course is structured to equip participants with the necessary knowledge and skills to utilize Python for real-world data science projects effectively.

Learning Outcome:

- **Understand and Utilize Python Basics**: Master the syntax, control structures, and basic data types in Python. Gain proficiency in using Python for various programming tasks including file operations and handling exceptions.
- **Apply Advanced Python Concepts**: Utilize advanced data structures like list comprehensions and dictionaries effectively. Implement lambda functions and understand the use of modules and packages to organize code.
- **Navigate the Data Science Workflow**: Demonstrate a clear understanding of the data science process, from data acquisition to processing and analysis, using Python tools.
- **Acquire and Clean Data**: Efficiently acquire data using APIs and web scraping, and perform data cleaning and preparation tasks using pandas to make data ready for analysis.
- **Conduct Exploratory Data Analysis (EDA)**: Apply statistical methods and pandas functionalities to explore and analyze datasets, handling missing data and engineering features to uncover insights.

- **Master Data Visualization**: Create compelling and informative visualizations using Matplotlib, Seaborn, and Plotly to communicate findings effectively. Design interactive dashboards that provide actionable insights.
- **Apply Machine Learning Basics**: Gain a foundational understanding of machine learning concepts and apply simple linear regression models using scikit-learn. Evaluate model performance using appropriate metrics.
- **Develop and Evaluate Predictive Models**: Implement simple linear regression to develop predictive models. Understand the assumptions behind the model and how to evaluate its performance for real-world applications.
- **Implement Projects with Real-world Data**: Complete hands-on projects that simulate real-world data science problems, integrating skills from data acquisition to machine learning model evaluation.
- **Collaborate and Share Findings**: Use Google Colab for collaborative coding and sharing of Jupyter notebooks. Learn best practices for presenting data science findings to stakeholders.

These outcomes are designed to provide participants with a solid foundation in Python programming and data science, equipping them with the skills necessary to tackle real-world data analysis challenges.

Prerequisites

Basic computer literacy and an understanding of programming concepts

- variables,
- loops,
- conditional statements

- A foundational knowledge of algebra and statistics.
- No prior experience with Python or Google Colab is required, but a gmail account is mandatory.
- Unrestricted access to the internet.
- VS Code (free) installation.

Training outline

1. Introduction to Python Programming

1. Setting up the Python Environment: Introduction to Google Colab.
2. Python Basics: Syntax, Variables, Data Types, Operators.
3. Control Structures: If-Else Conditions, For and While Loops.
4. Functions: Defining Functions, Arguments, Return Values.
5. Data Structures: Lists, Tuples, Dictionaries, Sets.
6. File Operations: Reading and Writing Files.
7. Exception Handling: Try-Except Blocks.

8. Basics of Object-Oriented Programming: Classes and Objects.

2. Advanced Python Concepts

1. Understanding List and Dictionary Comprehensions.
2. Exploring Lambda Functions and the Use of Map, Reduce, and Filter.
3. Modules and Packages: How to Import and Utilize Them.
4. Introduction to Virtual Environments and Their Importance.

3. Python for Data Science Basics

1. The Data Science Workflow: An Overview.
2. NumPy for Numerical Data Processing: Arrays and Operations.
3. Pandas for Data Analysis: Working with DataFrames.

4. Data Visualizations

1. Introduction to Data Visualization with Matplotlib.
2. Enhancing visualizations with Seaborn
3. Data manipulation for visualizations.
4. Classical vs OOP methods for data visualizations.

5. Data Acquisition Techniques

1. Fundamentals of Web Scraping: Techniques and Legal Considerations.
2. Using APIs for Data Acquisition: Understanding REST APIs, Making Requests.
3. Processing JSON Data: Parsing and Transformation.
4. Data Cleaning and Preparation with pandas.

6. Exploratory Data Analysis (EDA)

1. The Role of EDA in Data Science.
2. Descriptive Statistics and Their Interpretation.
3. Detecting and Handling Missing Data.
4. Feature Engineering and Data Transformation.

7. Data Visualization Techniques

1. Principles of Effective Data Visualization.
2. Advanced Usage of Matplotlib and Seaborn for Creating Visuals.
3. Creating Interactive Visualizations with Plotly.
4. Best Practices for Designing and Implementing Dashboards.

8. Introduction to Machine Learning

1. Machine Learning Overview: Types and Applications.
2. Simple Linear Regression: Concept, Implementation.
3. Data Preprocessing for Machine Learning.
4. Evaluating Model Performance: Metrics and Interpretations.

9. Fun Project

Building a recommendation engine using a movie database.

1. Data acquisition
2. Data Cleaning
3. Data wrangling and munging
4. Process
5. Outcome
6. Visualization

10. Practical Exercises and Project Work

1. Interactive Python Exercises using Google Colab for Hands-on Learning.
2. Data Acquisition Project: Implementing API Calls and Data Cleaning.
3. EDA Project: Analyzing a Dataset to Uncover Insights.
4. Visualization Project: Designing Visualizations and Dashboards.
5. Linear Regression Project: Building and Evaluating a Predictive Model.

The course wraps up with a final review session covering the key concepts and methodologies learned throughout the course. Participants will have the opportunity to present their projects, facilitating a collaborative review and feedback session. This course aims to empower participants with the Python programming and data science skills necessary to undertake their data analysis and visualization projects, laying the foundation for further exploration and specialization in the field of data science.

© 2026 Faysal Aziz · Kuala Lumpur

https://faysalaziz.com/courses/python-for-data-science/