Python for Data Science

Overview

This course – Python for Data Science is an introductory course meticulously crafted to empower students with the indispensable skills required for not just data analysis but for crafting data-driven models that can unlock hidden insights and drive informed decisions.

It starts by exploring Pandas, a powerful library for data manipulation and analysis, enabling students to effectively manage and process data.

As we venture further, we'll paint vivid stories through data visualization using Matplotlib and navigate the intricate world of exploratory data analysis (EDA) to unveil the narratives within datasets.

Beyond that, we'll equip you with the tools to handle missing values and outliers, preserving the integrity of your data. And as the final stroke, we'll delve into data normalization techniques and the creative artistry of feature engineering, transforming your data into a masterpiece optimized for machine learning.

By the end of the course, you'll not only have a solid Python foundation but the creative prowess to sculpt data into actionable insights. This course also acts as the first steppingstone for aspiring data scientists. This course is a part of the program titled **Fractal Data Science Professional Certificate.**

Requirements/Prerequisites

There are no requirements or prerequisites for the course. To succeed in this course, all you need to have is a curious and logical mind. While having a basic understanding of programming would be beneficial, it is not mandatory.

Course Structure

The course consists of the following materials:

- ***Videos:*** The lectures are delivered via videos, which are broken into small chunks, usually between five and seven minutes each.
- ***Expert talks*****:** Valuable insights and expertise into the intricacies of real-world project dynamics and processes in a concise 15 to 20-minute presentation by industry experts.
- ***Reading:*** We have made available pdf files of all the reading materials. There can be some external links or references for further reading.
- ***Ungraded Practice Quizzes:*** There will be *non-graded* short "quiz" questions that will follow some of the videos to help you gauge your understanding.

Python for Data Science

- ***Ungraded Assignment:*** There will be ungraded assignments which will help you with the hands-on practice for every lesson which will give you enough practice to attempt the graded assignments.
- ***Course materials:*** We have included the Python notebooks and datasets that are utilized in each module to facilitate hands-on practice and learning and help you refer to the concepts taught through the course..
- ***Graded Assignment:*** There will be graded assignments for submission part of every module.

Course Topics

Week 1: Introduction to Python for Data Science

Introduction, Kickstarting with Python and Jupyter, How Python Solved a Business Problem, The Business Problem for the Course,

Week 2: Data Wrangling with Python

Getting started with a data set, Dataframe Fundamentals and Indexing, Exploring and Querying on Columns, Manipulation and Grouping of Data, Date Time Operations in Python, Putting Together Data from Multiple Sources

Week 3: Exploratory Data Analysis

Descriptive Statistics, Understanding Data Distributions, Relationships Between Features, Enhancing Data Visualizations, Comparing Two Groups: Hypothesis Testing

Week 4: Data pre-processing

Treating Missing Data, Data Transformation, Understanding Data Normalization and Outlier Detection

Week 5: Feature Engineering

Creating Derived Features in Python, Gen AI in Python

Grading

Your grade in the course will be based solely on the graded labs. Ungraded labs and practice quizzes – will help. You are free to follow the course without completing the graded assignments and quizzes, but then you will not receive a certificate of completion.