

Python Programming Level 3: Data Analysis Using Python

Overview

The widespread use of the World Wide Web and social media has resulted in the creation of and access to an enormous amount of data becoming available. The data needs to be analysed to be able to apply the information in useful ways in many fields including business, science, and social science.

The course will teach you to apply your Python programming skills to complex data analysis problems. You will learn to use Pandas for data analysis and Seaborn for data visualization, with JupyterLab as your IDE. Additionally, you'll learn how to get, clean, prepare, and analyse data, including time-series data. Moreover, you'll learn to use linear regression models to predict unknown and future values.

Duration

3 days

COURSE OUTLINE

Module 1: Introduction to Python for data analysis

- What data analysis is
- The Python skills that you need for data analysis
- How to use JupyterLab as your IDE
- How to split the screen between two Notebooks
- How to use Magic Commands

Module 2: The Pandas essentials for data analysis

- Introduction to the Pandas DataFrame
- How to examine the data
- How to access the columns and rows
- How to work with the data
- How to shape the data
- How to analyse the data

The Pandas essentials for data visualization

- Introduction to data visualization

- How to create 8 types of plots
- How to enhance a plot

The Seaborn essentials for data visualization

Introduction to Seaborn

How to enhance and save plots

How to create relational plots

How to create categorical plots

How to create distribution plots

Other techniques for enhancing a plot

How to get the data

- How to find the data that you want to analyse
- How to import data into a DataFrame
- How to get database data into a DataFrame
- How to work with a Stata file
- How to work with a JSON file

How to clean the data

- Introduction to data cleaning
- How to simplify the data
- How to find and fix missing values
- How to fix data type problems
- How find and fix outliers

How to prepare the data

- How to add and modify columns
- How to apply functions and lambda expressions
- How to work with indexes
- How to combine DataFrames
- How to handle the SettingWithCopyWarning

How to analyse the data

- How to create and plot long data
- How to group and aggregate the data
- How to create and use pivot tables
- How to work with bins
- More skills for data analysis

How to analyse time-series data

- How to reindex time-series data
- How to resample time-series data
- How to work with rolling windows
- How to work with running totals

How to make predictions with a linear regression model

- Introduction to predictive analysis
- How to find correlations between variables
- How to use Scikit-learn to work with a linear regression
- How to plot regression models with Seaborn

How to make predictions with a multiple regression model

- A simple regression model for a Cars dataset
- How to work with a multiple regression model
- How to work with categorical variables
- How to improve a multiple regression model