

Data Analysis in Python

Following on from our *Python Basics for Data Analysis* course, *Data Analysis in Python* will build on your foundational knowledge of Python and pandas.

In this course, you will learn to:

- use the extensive data manipulation capabilities of pandas DataFrames
- customise the display of the output in Jupyter Notebooks
- use the plotting capabilities of Matplotlib to plot distributions and bar charts
- use the data visualisation library, Seaborn, and
- fit a basic model using scikit-learn.

Detailed Content

Introduction

Data Analysis in Python

Modify the DataFrame Display

pandas options
Working with pandas styles
Applying styles not dependent on values
Formatting values
String_formats
Applying styles that depend on values
Built-in conditional formatting

Export Notebook As...

Export to PDF or HTML
Create slides

Copy vs View

Setting with copy warning

Working with Missing Values

Missing values
inf and -inf
Removing missing values
Replacing missing values

Importing Data

Importing into a pandas DataFrame

Manipulating Data

Summarise a dataset
Report and display multiple summary statistics
Ordering data
Working with dates
Add columns with assign()
Working with strings
Reordering and dropping columns
Selecting rows based on values
Grouping and summarising data
Replacing values
Concatenate data
Bin continuous variables into categories

Working with Relational Data

Joining data from two DataFrames

Visualising Distributions

Visual representation of distributions with Matplotlib and Seaborn
Histograms

Boxplots
Bar and column charts

Multivariate Analysis

Scatterplot matrix
Bar and column charts

Basic Modelling

Create a linear model with scikit-learn