

Beginning Data Analysis with Python and Jupyter

Course Summary

Description

Data science is becoming increasingly popular with more and more industries beginning to value its importance, while recent advancements in open-source software have made the discipline accessible to a wide range of people. Python is a popular choice for most data scientists, owing to its ease of use and versatile nature.

In this course, we show how Jupyter Notebooks can be used with Python for various data-science applications. Aside from being an ideal "virtual playground" for data exploration, Jupyter Notebooks are equally suitable for creating reproducible data processing pipelines, visualizations, and prediction models.

We will look at various data modelling concepts using Jupyter Notebooks, and we will see the full power of Jupyter Notebooks as we work through this course.

This fast-paced practical single-day course focuses on solving challenges presented by data science in a manner that is simple to conceptualize and easy to implement.

This is achieved by leveraging Python libraries that offer abstractions to complicated underlying algorithms. The result is that data science becomes very approachable for beginners, a fact which is reflected by the strength and growing popularity of the Python ecosystem. In this course, we will cover every aspect of the standard data workflow process, including collecting, cleaning, investigating, visualizing, and modeling data.

The Jupyter Notebook acts as an add-on tool - a virtual playground - that allows you to create and share live code, equations, visualizations, and text.

By touching on a variety of topics within the discipline, you'll be exposed to many interesting examples with real-world data.

This course focuses on creating reproducible data analyses using Python and Jupyter, and is intended for an audience with a background in Python. As such, we do not cover the basics of Python in this course. However, we will take a brief tour of the Jupyter interface.

Topics

- Jupyter Fundamentals
- Data Cleaning and Advanced Machine Learning
- Web Scraping and Interactive Visualizations

Audience

If you're a Python programmer stepping out into the hugely popular world of data science, opting for this course is the right way to get started.

Prerequisites

For the best experience in this course, you should have knowledge of programming fundamentals and some experience with Python. In particular, having some familiarity with the Python libraries Pandas, Matplotlib, and scikit-learn will be useful.

Duration

One day

Beginning Data Analysis with Python and Jupyter

Course Outline

- I. Jupyter Fundamentals*
 - A. Basic Functionality and Features
 - B. Our First Analysis - The Boston Housing Dataset
- II. Data Cleaning and Advanced Machine Learning*
 - A. Preparing to Train a Predictive Model
 - B. Training Classification Models
- III. Web Scraping and Interactive Visualizations*
 - A. Scraping Web Page Data
- IV. Interactive Visualizations*