Getting started with Python-

Introduction to Libraries in Python

What are Libraries?

- In Python, a library is used loosely to describe a collection of the core modules.
- A module is a .py file that contains certain functions and classes that can be used when called.
- The term ‘standard library‘ in Python language refers to the collection of exact syntax, token and semantics of the Python language which comes bundled with the core Python distribution. There are more than 200 modules that forms a part of core Python. This library is written in the C language.
- “Additional libraries” are those libraries that can be added additionally to help with specific uses in Python. As an open source language, Python has a lot of additional libraries available.

Contents

`

Importing Libraries in Python

- You need to import a library into an environment before you can call it's functions in your program.
- To import any library just use any of the following codes -

```
import math as m
from math import *
```

Useful Libraries for Data Science

- There are a lot of Libraries available for scientific computations and data analysis in Python. Following is a list of Libraries that we will be using during this course :
- NumPy stands for Numerical Python. It is the fundamental package for scientific computing and is used to manipulate homogenous array based data. The most powerful feature of NumPy is n-dimensional array. This library also contains basic linear algebra functions, Fourier transforms, advanced random number capabilities and has an easy to use C API.
- Pandas contains high level data structures and manipulation tools which are used extensively for data munging and preparation. Pandas was added relatively recently to Python and have been instrumental in boosting Python's usage in data scientist community. It is built on top of NumPy.

Useful Libraries for Data Science

- SciPy stands for Scientific Python. SciPy is built on NumPy. It is one of the most useful libraries for variety of high level science and engineering modules like discrete Fourier transform, Linear Algebra, Optimization and Sparse matrices.
- Matplotlib is used for plotting vast variety of graphs, starting from histograms to line plots to heat plots. You can also use Latex commands to add math to your plot.
- Scikit Learn is used for machine learning. Built on NumPy, SciPy and matplotlib, this library contains a lot of efficient tools for machine learning and statistical modelling including classification, regression, clustering and dimensionality reduction.
- Statsmodels for statistical modelling. Statsmodels is a Python module that allows users to explore data, estimate statistical models, and perform statistical tests. An extensive list of descriptive statistics, statistical tests, plotting functions, and result statistics are available for different types of data and each estimator.

Quick Recap

NumPy	• Manipulate homogenous array based data
Pandas	• High level data munging and preparation
SciPy	• High level science and engineering modules. • Discrete Fourier transform, Linear Algebra, Optimization and Sparse matrices.
Matplotlib	• Plotting vast variety of graphs • Histograms, line plots, heat plots
Scikit Learn	• Used for machine learning and statistical modeling. • Classification, regression, clustering and dimensionality reduction
Statsmodels	• Explore data, estimate statistical models, and perform statistical tests