Data Science with Python Programming

- **Presentation By Uplatz**
- **Contact us: https://training.uplatz.com**
- **Email: info@uplatz.com**
- **Phone: +44 7836 212635**

Uplatz

Learning outcomes:

Data Science Libraries

Libraries for Data Processing and Modeling

- **Pandas**
- **Numpy**
- **SciPy**
- **Scikit-learn**

Libraries for Data Visualization

- **Matplotlib**
- **Seaborn**
- **Plotly**

Uplatz

Data Science Libraries

Python is one of the most popular languages used by data scientists and software developers alike for data science tasks. It can be used to predict outcomes, automate tasks, streamline processes, and offer business intelligence insights. Most data scientists are already leveraging the power of Python programming every day.

Data science is an extremely important field in current times!

Uplatz

Data Science Libraries

The reason for this huge success of **Python** in **Data Science** is its extensive library support for data science and analytics. There are many Python libraries that contain a host of functions, tools, and methods to manage and analyse data. Each of these libraries has a particular focus with some libraries managing image and textual data, data mining, neural networks, data visualization, and so on. Here we have divided the top **Python libraries for Data Science** into those focusing on data processing and data visualization respectively. So let's check out these libraries now!

Uplatz

Libraries for Data Processing and Modeling

Pandas:

Pandas is a free Python software library for data analysis and data handling. It was created as a community library project and initially released around 2008. Pandas provides various high-performance and easy-to-use data structures and operations for manipulating data in the form of numerical tables and time series. Pandas also has multiple tools for reading and writing data between in-memory data structures and different file formats.

Uplatz

Libraries for Data Processing and Modeling

Pandas:

In short, it is perfect for quick and easy data manipulation, data aggregation, reading, and writing the data as well as data visualization. Pandas can also take in data from different types of files such as CSV, excel etc. or a SQL database and create a Python object known as a data frame. A data frame contains rows and columns and it can be used for data manipulation with operations such as join, merge, groupby, concatenate etc.

Libraries for Data Processing and Modeling

Pandas:

Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data — load, prepare, manipulate, model, and analyse.

Uplatz

Libraries for Data Processing and Modeling

Numpy:

NumPy, which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy, mathematical and logical operations on arrays can be performed.

NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.

Libraries for Data Processing and Modeling

Numpy:

NumPy is a free Python software library for numerical computing on data that can be in the form of large arrays and multi-dimensional matrices. These multidimensional matrices are the main objects in NumPy where their dimensions are called axes and the number of axes is called a rank. NumPy also provides various tools to work with these arrays and high-level mathematical functions to manipulate this data with linear algebra, Fourier transforms, random number, etc.

Libraries for Data Processing and Modeling

Numpy:

Some of the basic array operations that can be performed using NumPy include adding, slicing, multiplying, flattening, reshaping, and indexing the arrays. Other advanced functions include stacking the arrays, splitting them into sections, broadcasting arrays, etc.

Uplatz

Libraries for Data Processing and Modeling

SciPy:

SciPy is a free software library for scientific computing and technical computing on the data. It was created as a community library project and initially released around 2001. SciPy library is built on the NumPy array object and it is part of the NumPy stack which also includes other scientific computing libraries and tools such as Matplotlib, SymPy, pandas etc.

Libraries for Data Processing and Modeling

SciPy:

This NumPy stack has users which also use comparable applications such as GNU Octave, MATLAB, GNU Octave, Scilab, etc. SciPy allows for various scientific computing tasks that handle data optimization, data integration, data interpolation, and data modification using linear algebra, Fourier transforms, random number generation, special functions, etc. Just like NumPy, multidimensional matrices are the main objects in SciPy, which are provided by the NumPy module itself.

Uplatz

Libraries for Data Processing and Modeling

Scikit-learn:

Scikit-learn is a free software library for Machine Learning coding primarily in the Python programming language. It was initially developed as a Google Summer of Code project by David Cournapeau and originally released in June 2007. Scikit-learn is built on top of other Python libraries like NumPy, SciPy, Matplotlib, Pandas, etc. and so it provides full interoperability with these libraries.

Libraries for Data Processing and Modeling

Scikit-learn:

While Scikit-learn is written mainly in Python, it has also used Cython to write some core algorithms in order to improve performance. You can implement various **Supervised** and **Unsupervised** Machine learning models on Scikit-learn like Classification, Regression, Support Vector Machines, Random Forests, Nearest Neighbors, Naive Bayes, Decision Trees, Clustering, etc. with Scikit-learn.

Uplatz

Libraries for Data Visualization

Matplotlib:

Matplotlib is a data visualization library and 2-D plotting library of Python It was initially released in 2003 and it is the most popular and widely-used plotting library in the Python community. It comes with an interactive environment across multiple platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers etc.

Uplatz

Libraries for Data Visualization

Matplotlib:

It can be used to embed plots into applications using various GUI toolkits like Tkinter, GTK+, wxPython, Qt, etc. So you can use Matplotlib to create plots, bar charts, pie charts, histograms, scatterplots, error charts, power spectra, stem-plots, and whatever other visualization charts you want! The Pyplot module also provides a MATLAB-like interface that is just as versatile and useful as MATLAB while being totally free and open source.

Uplatz

Libraries for Data Visualization

Matplotlib:

One of Matplotlib's most important features is its ability to play well with many operating systems and graphics backend. Matplotlib supports dozens of backend and output types, which means you can count on it to work regardless of which operating system you are using or which output format you wish. This cross-platform, everything-to-everyone approach has been one of the great strengths of Matplotlib.

Uplatz

Libraries for Data Visualization

Seaborn:

[Seaborn](#) is a Python data visualization library that is based on Matplotlib and closely integrated with the numpy and pandas data structures. Seaborn has various dataset-oriented plotting functions that operate on data frames and arrays that have whole datasets within them. Then it internally performs the necessary statistical aggregation and mapping functions to create informative plots that the user desires.

Libraries for Data Visualization

Seaborn:

It is a high-level interface for creating beautiful and informative statistical graphics that are integral to exploring and understanding data. The Seaborn data graphics can include bar charts, pie charts, histograms, scatterplots, error charts, etc. Seaborn also has various tools for choosing color palettes that can reveal patterns in the data.

Uplatz

Libraries for Data Visualization

Plotly:

Plotly is a free open-source graphing library that can be used to form data visualizations. Plotly (plotly.py) is built on top of the Plotly JavaScript library (plotly.js) and can be used to create web-based data visualizations that can be displayed in Jupyter notebooks or web applications using Dash or saved as individual HTML files.

Uplatz

Libraries for Data Visualization

Plotly:

Plotly provides more than 40 unique chart types like scatter plots, histograms, line charts, bar charts, pie charts, error bars, box plots, multiple axes, sparklines, dendrogram, 3-D charts, etc. Plotly also provides contour plots, which are not that common in other data visualization libraries. In addition to all this, Plotly can be used offline with no internet connection.

Thank you

Uplatz