**Explain how Python data analysis libraries are used?**

Python is a computer programming language that is frequently used to create websites and software, automate tasks, and analyze data.

## **Data Analysis**

Data analysis is defined as the process of cleaning, transforming, and modeling data in order to find useful information for business decisions. The goal of data analysis is to extract useful information from data and make decisions based on that information.

In this article, we will explain how python data analysis libraries are used

## **NumPy - Fundamental Scientific Computing**

NumPy is an abbreviation for Numerical Python. The n-dimensional array is NumPy's most powerful feature. This library also includes basic linear algebra functions, Fourier transforms advanced random number capabilities, and integration tools for Fortran, C, and C++.

NumPy is a popular Python data analysis package. NumPy allows you to speed up your workflow and interact with other Python ecosystem packages such as scikit-learn, which use NumPy under the hood. NumPy was created in the mid-2000s as an offshoot of an even older package called Numeric. Because of its longevity, almost every data analysis or machine learning package for Python makes use of NumPy in some way.

### **Applications**

* Used extensively in data analysis
* Makes a strong N-dimensional array
* It serves as the foundation for other libraries such as SciPy and scikit-learn.
* MATLAB replacement when combined with SciPy and matplotlib

## **Scipy - Fundamental Scientific Computing**

SciPy is a Python library that can be used to solve a variety of mathematical equations and algorithms. It is built on top of the Numpy library, which provides more options for finding scientific mathematical formulas such as Matrix Rank, Inverse, polynomial equations, LU Decomposition, and so on. Using its high-level functions significantly reduces the complexity of the code and aids in better data analysis. SciPy is an interactive Python session that serves as a data-processing library, competing with competitors such as MATLAB, Octave, R-Lab, and others. It has a wide range of user-friendly, efficient, and simple-to-use functions that aid in the resolution of problems such as numerical integration, interpolation, optimization, linear algebra, and statistics.

The advantage of using the SciPy library in Python to create ML models is that it also provides a powerful programming language for developing less complex programmes and applications.

### **Applications**

* Multidimensional image operations.
* Optimization algorithms for solving differential equations and the Fourier transform.
* Linear algebra.

## **Pandas - Data Manipulation and Analysis**

Pandas is one of these packages, and it greatly simplifies data import and analysis.

Pandas aim to combine the functionality of NumPy and matplotlib to provide a user-friendly data analytics and visualization tool. Aside from the integration, it also improves usage significantly.

Pandas is used to perform structured data operations and manipulations. It is widely employed in data munging and preparation. Pandas were added to Python relatively recently and have been instrumental in increasing Python's usage among data scientists.

### **Applications**

* Data wrangling and cleaning in general
* Because it has excellent support for loading CSV files into its data frame format, it is ideal for ETL (extract, transform, load) jobs for data transformation and data storage.
* Statistics, finance, and neuroscience are just a few of the academic and commercial applications.
* Date range generation, moving window, linear regression, and date shifting are examples of time-series-specific functionality.

## **Matplotlib – Plotting and Visualization**

Data visualization is one of the essential skills required of data scientists. Visualization techniques can be used to understand and address the majority of business problems. Exploratory Data Analysis (EDA) and Graphical Plots are the two main components of visualisation. Effective visualization assists users in understanding data patterns and solving business problems more effectively. Another advantage of visualization is that it reduces complex data to a more understandable format.

Matplotlib can be used to create a wide range of graphs, from histograms to line plots to heat maps. To use these plotting features inline, use the Pylab feature in ipython notebook (ipython notebook -pylab = inline). If you ignore the inline option, pylab converts the ipython environment to a Matlab-like environment.

### **Applications**

* Variable correlation analysis
* Display the models' 95 percent confidence intervals.
* Outlier detection with a scatter plot etc.
* Visualize data distribution to gain instant insights.

## **Scikit-learn – Machine Learning and Data Mining**

SciPy Toolkits, also known as scikits, are widely used for machine learning. A scikit is a specialised toolkit that is used for specific tasks such as machine learning or image processing. Scikit-learn and Scikit-image are the two specialised packages used for this. The package contains a collection of useful algorithms for dealing with the processes involved in machine learning and image processing.

Scikits are widely used by programmers and software developers. Scikit-learn is even regarded as one of the pillars of Python-based machine learning. This can be used to create various models, prepare and evaluate data, and even perform post-model analysis.

### **Applications**

* Clustering
* Classification
* Regression
* Model selection
* Dimensionality reduction

## **StatsModels – Statistical Modeling, Testing, and Analysis**

Statsmodels are statistical modelling tools. It's a Python module that lets you explore data, estimate statistical models, and run statistical tests. For each type of data and estimator, a comprehensive set of descriptive statistics, statistical tests, plotting functions, and result statistics is available.

## **Seaborn – For Statistical Data Visualization**

Seaborn is a Matplotlib-based free and open-source data visualization library. Because of its high-level interface for drawing attractive and informative statistical graphics, many data scientists prefer seaborn over matplotlib.

Seaborn has simple functions that allow you to concentrate on the plot and learn how to draw it. Seaborn is a must-have library that you must master.

Seaborn's objective is to make visualisation a central part of data exploration and comprehension.

## **Conclusion**

This article explains how different types of libraries are used for data analysis in Python. We also learned about their applications.