

Introduction to Python for Data Analysis

Python for Data Analysis

Outline

- What is Data Analysis?
- Why Data Analysis using Python?
- Most Commonly used Libraries for Data Analysis
- Data Analysis using NumPy
- Data Analysis using Pandas



**What is Data
Analysis?**

What is Data Analysis?

- Data Analysis is the discovery and communication of meaningful patterns in data.
- Especially, valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming, and operation research to qualify performance.
- Data Analysis often favors data visualization to communicate insight.
- In a nutshell, analytics is the scientific process of transforming data into insight for making better decisions.
- The goal of Data Analysis to get actionable insights resulting in smarter decisions and better business outcomes.





Why Data Analysis using
Python?

Why Data Analysis using Python?

- Python is open source, interpreted, high level language and provides great approach for object-oriented programming.
- It is one of the best language used by data scientist for various data science projects/application.
- Python provide great functionality to deal with mathematics, statistics, scientific function, and libraries to deals with data science application.
- One of the main reasons why Python is widely used in Data Analysis is because of its ease of use and simple syntax which makes it easy to adapt for people who do not have an coding background.





Most Commonly used Libraries for Data Analysis

Most Commonly used Libraries for Data Analysis



NumPy

NumPy is Python library that provides mathematical function to handle large dimension array.



Pandas

Pandas is one of the most popular Python library for data manipulation and analysis.



Data Analysis using NumPy

Data Analysis using NumPy



- NumPy provides various method/function for Array, Metrics, and linear algebra.
- NumPy stands for Numerical Python and provides lots of useful features for operations on n-arrays and matrices in Python.
- This library provides vectorization of mathematical operations on the NumPy array type, which enhance performance and speeds up the execution.
- It's very easy to work with large multidimensional arrays and matrices using NumPy.



Data Analysis using Pandas

Data Analysis using Pandas



- Pandas provide easiest method to perform analysis and provide useful functions to manipulate large amount of structured data.
- Pandas is a perfect tool for data wrangling and designed for quick and easy data manipulation, aggregation, and visualization.
- There two data structures in Pandas :
 - 1) **Series** – It Handle and store data in one-dimensional data.
 - 2) **DataFrame** – It Handle and store Two dimensional data.