



# Data Analysis Using Python

## Introduction:

Data Science and analysis is playing the most significant role today covering every industry in the market. For e.g. finance, e-commerce, business, education, government. Now organizations play a 360-degree role to analyze the behavior and interest of their customers to make decisions in favor of them. Data is analyzed through a programming language such as python which is one of the most versatile languages and helps in doing a lot of things through it. Netflix is a pure data science project that reached the top through analyzing every single interest of their customers. Keywords: Data Visualization, Anaconda Jupyter Notebook, Exploratory Data Analysis, Machine Learning.

## Duration:

- 12 Hours.

## Content:

Days	Topic Name	Sub Topics	Duration(12 Hours)
1	Introduction to Data Analysis and Data Manipulation with NumPy	Introduction NumPy Arrays NumPy Basics Math Random Indexing Filtering Statistics Aggregation Saving Data	2hr



2	Data Analysis with pandas	Introduction Series DataFrame Combining Indexing File I/O Grouping Features Filtering Sorting Stastical Plotting	2hr
3	Data Preprocessing with Scikit-Learn	Introduction Standardizing Data Data Range Robust Scaling Normalizing Data Data Imputation	2hr
4	Cleaning Data in Python	Working with Duplicates and Missing Values Which values should be replace with missing values based on data Identifying and Eliminating Outliers Dropping duplicate data Filling missing data Applying on raw dataset and introduction to Kaggle and other data sources	2hr
5	Introduction to Data Visualization Matplotlib	Introduction to Visualization and Python packages Matplotlib history Introduction to plotting Line Plot Scatter Plot Bar Graph Histogram Pie Chart Box Plot Tasks	2hr



6	Data Visualization using Seaborn	Using Seaborn Styles Setting the default style Color Palettes Creating Custom Palettes stripplot() and swarmplot() boxplots, violinplots and lvplots barplots, pointplots and countplots Regression Plots Binning data Matrix plots Creating heatmaps	2hr
---	----------------------------------	---	-----

### Course Objectives:

- The main goal of this course is to help students or Faculty to learn, understand, and practice data analysis and machine learning approaches, which include the study of modern computing data technologies and scaling up machine learning techniques focusing on industry applications. Mainly the course objectives are conceptualization and summarization of Data Analysis and machine learning computing technologies, machine learning techniques, and scaling up machine learning approaches.

### Entry Requirements (Pre-requisites):

- Faculty&Students must have Knowledge on Python Programming and Statistics.

### Hardware Requirements:

- i3 or above Processor is required
- 4 GB or above RAM is recommended
- Good Internet Connectivity
- OS-Windows 10 is Preferable