# 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 by analyzing every single interest of their customers. Keywords: Data Visualization, AnacondaJupyter Notebook, Exploratory Data Analysis, Machine Learning.

## Duration:

- 18 Hours.

## Content:

| Days | Topic Name | Sub Topics | Duration |
|------|-----------|-----------|----------|
| 1 | Introduction to Data Analysis | Introduction to Data<br>Types of Data in Statistics (Numerical & Categorical)<br>Overview of Python Concepts | 2hr |
| 2 | Data Manipulation with NumPy | Introduction<br>NumPy Arrays<br>NumPy Basics<br>Math<br>Indexing<br>Random | 2hr |
| 3 | | Filtering<br>Statistics<br>Aggregation<br>Saving Data<br>Introduction to Data Analysis using Pandas | 2hr |

| | | Pandas Series | |
|---|---|---|---|
| 4 | Data Analysis with pandas | Pandas DataFrame<br>Combining<br>Indexing<br>File I/O<br>Grouping<br>Features<br>Filtering<br>Sorting<br>Statistical<br>Plotting | 2hr |
| 5 | Data Preprocessing with Scikit-Learn | Introduction<br>Standardizing Data<br>Data Range<br>Robust Scaling<br>Normalizing Data | 2hr |
| 6 | Cleaning Data in Python | Working with Duplicates and Missing Values<br>Which values should be replaced with missing values based on data<br>Identifying and Eliminating Outliers<br>Dropping duplicate data<br>Filling missing data using Data Imputation | 2hr |
| 7 | Introduction to Data Visualization Matplotlib | Introduction to Visualization and Python packages<br>Matplotlib history<br>Introduction to plotting<br>Line Plot<br>Scatter Plot<br>Bar Graph<br>Histogram<br>Pie Chart<br>Box Plot<br>Tasks | 2hr |
| 8 | Data Visualization using Seaborn | Using Seaborn Styles<br>Setting the default style<br>Color Palettes<br>Creating Custom Palettes<br>stripplot() and swarmplot()<br>boxplots, violin plots | 2hr |

| | | Bar plots, point plots, and count plots | |
|---|---|---|---|
| 9 | Data Visualization using Seaborn | Regression Plots<br>Binning data<br>Matrix plots<br>Creating heatmaps<br>Applying on raw dataset and introduction to Kaggle and other data sources | 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 Knowlege 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