EducateNXT

Data Science

Comprehensive Training & Live Projects

Objectives of the Program

- To enable participants, understand how Data Science is used in every aspect of our daily lives and businesses
- To enable the participants to learn basics of Python.
- To enable the participants to learn Data Science using Python.
- To empower the participants with sufficient knowledge to use latest technologies of Data Science (Hands-On).
- To make the participants understand overview of how Exploratory Data Analysis works.
- To provide the participants Hands-on experience in Data science concepts.
- To empower the participants with sufficient knowledge to use latest technologies of Data Science (Hands-On).

Overview of the Program Content

- Introduction to Data Science
- Basics of Python and Hands-on on Python
- Introduction to Data Science using Python
- Hands on session on Data Science with Python [Database Connectivity with PYTHON: Performing Database Transactions (Inserting, Deleting, and Updating the Database)
- Project in Data Science

Introductory Session of the Program and Analytical Problem Solving

- Introduction to Data Science - What is Data science
- Evolution of Data Science
- Why Data Science
- Data Science Use Cases
- Ways of Doing Data Science
- Pre-requisites (Skills/Tools/Math/Stats - Introduction/Names - Briefly)
- Job Opportunities in DS

Python for Data Science

Chapter 1: Basic Building Blocks

- Introduction to Python
- History of Python and its Features
- Installation of Anaconda Tool
- Introduction to Jupyter Notebook
- Basic Arithmetic and String Operations
- Variable
- Identifiers
- Print and Input Function

Python for Data Science

Chapter 2: Data Types

- Numeric Data Types
- Integer
- Float
- Complex Number
- Boolean
  - True
  - False
- String
- List
  - All List Operation
- Tuple
  - All Tuple Operation
- Dictionary
  - All Dictionary Operation
- Set
  - All Set Operation
- Type Casting: Data Conversion from one type to Another

Python for Data Science

Chapter 3: Operators

- Logical Operators
- Bitwise Operators

Chapter 4: Control Structure

- Sequential
- Alternative or Branching
  - If
  - If…else
  - If…elif…else
- Iterative or Loop
  - For loop
  - While loop
  - Break, Pass and Continue statements
  - Range functions

Python for Data Science

Chapter 5: Functions

- Pre-defined Function
- User Defined Function
- Lambda Function
- Recursive Function

Chapter 6: Data Structures

- Linked List
- Stack and Queue
- Binary Tree
- Binary Search Tree
- Heap
- Sorting
- Searching
- Hashing
- Asymptotic Analysis
- Big 0 Notation
- Time Complexity

Python for Data Science

Chapter 7: String Operations and Regular Expressions

- String object basics
- String methods
- Splitting and Joining Strings
- String format functions

Chapter 8: File Handling

- Reading File
- Writing Files

Chapter 9: Exception Handling

- Try
- Except
- Finally

Chapter 10: The Object-Oriented Approach

- Classes
- Methods
- Objects
- Inheritance, Multiple Inheritance

Maths Behind Data Science

Chapter 11: Linear Algebra, Calculus & Probability

- Introduction to Linear Algebra
- Introduction to Calculus
- Introduction to Probability

Chapter 12: Statistical Analysis

1. Types of Data
   - Quantitative
   - Qualitative
2. Data Source
   - Primary
   - Secondary
3. Sampling
   - Probability Sampling
   - Non-Probability Sampling
4. Types of Scale
   - Nominal
   - Ordinal
   - Ratio
   - Interval

Maths Behind Data Science

Descriptive Statistics

1. Measure of Central Tendency

- Mean
- Mode
- Median

2. Measure of Dispersion (Spread)

- Range
- IQR
- Variance
- Standard Deviation

3. Coefficient of Variation

4. Covariance and Correlation

5. Measure of Shape

- Skewness
- Kurtosis

6. Frequency Distribution

7. Histogram

8. Five Number Summary QQ Plots, Quantile Plot, Scatter Plot

9. Empirical and Chebyshev rule

Maths Behind Data Science

Inferential Statistics

- Sample & Population- Terminology
- Sampling Techniques
- Sampling Errors
- Continuous & Discrete Random Variables
- Probability Basics
- Mutually Exclusive Events
- Independent Events
- Conditional Probability
- Bayes Theorem
- Probability Distribution Functions- Continuous & Discrete
- Cumulative Distribution Functions
- Normal Distribution
- Uniform Distribution
- Exponential Distribution
- Bernoulli / Binomial Distribution
- Poisson Distribution
- Chi-Square etc...
- Goal of Statistical Inference
- Inferences about population parameter from sample statistic using standard normal distribution
- Sampling Distribution
- Central Limit Theorem (CLT)
- Z-score, Confidence level, Confidence Interval, Significance Level

Maths Behind Data Science

Chapter 13: Hypothesis Testing

- What is Hypothesis
- Null Hypothesis & Alternate Hypothesis
- Why Hypothesis Testing
- Types of Hypothesis Tests
- One sample t-test
- Paired t-test
- Z-test
- Chi-square test
- One sample proportion test
- F-ratio test of variance
- Anova test etc…
- Types of errors- Type 1 & Type2
- A/B Testing

Data Science (Exploratory Data Analysis / Data Wrangling / Data Munging / Data Cleaning Techniques / Data Preparation / Feature Engineering)

Chapter 14: Data Science Library

NumPy

- **Basics of NumPy array**
- **1D and 2D Array Operation**
- **Broadcasting an array**
- **Array Slicing**
- **Mathematical and Trigonometric Operation**
- **Concatenation and Split**
- **Random Sampling**
- **Matrix Indexing**
- **Selection Techniques**
- **Saving and loading arrays**

Pandas

- Series
- Indexing elements of a series
- Dictionaries
- Data frame
- Different ways of indexing in a data frame
- Conditional indexing in a data frame
- loc and iloc
- Dropna and fillna
- Append Function
- Joining, Concat, Groupby, merging similar to SQL logic

Data Science (Exploratory Data Analysis / Data Wrangling / Data Munging / Data Cleaning Techniques / Data Preparation / Feature Engineering)

EducateNXT

Panda (Cont...)

- Filtering, sorting and indexing
- Pivot
- Loops and functions
- Date time index
- Saving and loading a csv

Matplotlib

- Histogram
- Line Plot
- Pie Chart
- Scatter Plot
- Box Plot
- Count Plot
- Dist Plot
- Others Plot

Seaborn

- Pairplot
- Heatmap

Scipy

Sklearn

Data Science (Exploratory Data Analysis / Data Wrangling / Data Munging / Data Cleaning Techniques / Data Preparation / Feature Engineering)

EducateNXT

Chapter 15: Data Exploration

- Identification of variables (I/P and O/P) and data types
- Treating Duplicate Data
- Univariate, bivariate, multivariate analysis

Chapter 16: Missing value treatment

- Removing Missing Value
- Mean/Median
- Mode
- Others Method

Chapter 17: Outlier treatment

- Box Plot
- Percentile
- Standard Deviation
- Z Score
- IQR

Chapter 18: Categorical to Numerical

- **Dummy Variable**
- **One hot Encoding**
- **Label Encoding**

Data Science (Exploratory Data Analysis / Data Wrangling / Data Munging / Data Cleaning Techniques / Data Preparation / Feature Engineering)

Chapter 19: Feature Rescaling

- Normalization
- Standardization

Chapter 20: Feature (Variable) transformations

- Log Transform
- Exponential Transform

Chapter 21: Correlation Analysis

- VIF
- Heatmap

Chapter 22: Feature selection Methods

- Backward
- Forward
- Stepwise
- RFE
- Using Statistical methods
- PCA
- Lasso Regression etc

Data Visualization using Tableau

- Introduction to Tableau
- Creating Different type of Charts
- Creating Dashboards
- Use Cases

Basics of SQL

- Introduction to Database
- Introduction to SQL
- Basic Operations of SQL
  - Joins
  - Groupby
  - Orderby
  - Where and having
  - Partition by
  - Windows Function-lead,lag
  - Sub queries

Tools, Languages & Libraries Used

- Python
- Anaconda
- Jupyter Notebook
- Google Colab
- NumPy
- Pandas
- Matplotlib
- Seaborn
- Sklearn

EducateNXT

Thank You

Gopinath Puralachetty

Head - Institutional Alliances

Email: gopinath.p@educatenxt.co

Ph.: 7794931347 | 9121000729

Website: www.educatenxt.co