

ZENUS INFOTECH INDIA PVT. LTD.

Curriculum of Data science

Course Duration: 4/6 Weeks

- ❖ What is Data science
- ❖ Introduction to python data science
- ❖ Installation of Pandas, numpy, sklearn, Matplotlib
- ❖ Basic terminologies of DS
 - a. Data science
 - b. Data scientist
 - c. Data set
 - d. Data mining
 - e. Data visualization
 - f. Data modeling
 - g. Data wrangling
 - h. Big data
 - i. Machine learning
 - j. Algorithms
 - k. Deep learning

Hands on with Pandas – Data Analysis library [Data Processing]

- ❖ Why Pandas?
- ❖ Features of Pandas
- ❖ Data structures in Pandas
 - a. Series
 - b. DataFrame
 - c. Panel
 - d. Panel4D
- ❖ Series creation
 - a. Using ndarray
 - b. Using dict
 - c. Using scalar values
 - d. Using list
- ❖ Accessing elements of Series
 - a. Using indexing
 - b. Using slicing
 - c. Using ranging
 - d. Using iloc method
 - e. Using loc method
- ❖ Vectorizing operations
 - a. Vector operations using same index

- values
 - b. Vector operations using different index values
- ❖ DataFrame creation
 - a. Using list
 - b. Using dict
 - c. Using ndarray
 - d. Using series
 - e. Using DataFrame
- ❖ Viewing DataFrame elements
 - a. Using describe function
 - b. Using column name
 - c. Using iloc method
 - d. Using iat method
 - e. Using head()
 - f. Using tail()
 - g. Using index method

Working with Pandas Data

- ❖ Handling missing values
 - a. Using Dropna()
 - b. Using Fillna()
 - c. Using add between 2 vector series
- ❖ Data operations with customized functions
 - a. Using groupby()
 - b. Using sorting
 - c. Using merge
 - d. Using duplicate
 - e. Using concatenation
- ❖ Statistical functions in data operations
 - a. Max()
 - b. Min()
 - c. Mean()
 - d. Std()
- ❖ Data Processing
 - a. Processing CSV data
 - b. Processing JSON data
 - c. Processing XLS data
 - d. Date and time in data

Numpy – Mathematical Computation

- ❖ Why numpy?
- ❖ Powerful properties of numpy
- ❖ Types of arrays
 - a. One dimensional
 - b. Two dimensional
 - c. Three dimensional
- ❖ Attributes of ndarray
 - a. Using .ndim
 - b. Using .shape
 - c. Using .size
 - d. Using .dtype
- ❖ Basic operations
 - a. (+, -, *, /, %, //, &, |, ~, <, <=, >, >=, ==, !=)
 - b. Accessing array elements using axis values
 - c. Indexing with Boolean array
- ❖ Creating functions for arrays
 - a. Using arange()
 - b. Using linspace()
 - c. Using ones()
 - d. Using zeros()
 - e. Using diag()
 - f. Using random.rand()
 - g. Using random.randn()
 - h. Using random.seed()
- ❖ Copy and view
 - a. Deep copy
 - b. Shallow copy
 - c. Simple assignment
- ❖ Universal functions
 - a. Sqrt
 - b. Cos
 - c. Floor
 - d. Exp
- ❖ Shape manipulation
 - a. Using flatten
 - b. Using reshape
 - c. Using resize
 - d. Using split

- e. Using stack
- Broadcasting
- Using tile()
- Using ones()
- Using newaxis()

Hands on with Matplotlib library - [Basic Data Visualization]

- ❖ Chart properties
 - a. Creating a chart
 - b. Labeling the axes
 - c. Formatting line style and color
 - d. Saving the chart in a file
- ❖ Styling the chart
 - a. Adding annotations
 - b. Adding legends
 - c. Presentation style
- ❖ Types of presentation styles
 - a. Scatter plots
 - b. Heat maps
 - c. Bubble chart
 - d. Bar chart
 - e. Pie chart
 - f. XKCD style
 - g. 3D chart
 - h. Box and whisker plots
 - i. Time series plot
 - j. Graph data / line graph
 - k. Geographical data

Advanced Data Visualization using SEABORN

- ❖ Visualization techniques used
 - a. Histogram
 - b. Histogram with grid
 - c. Distplot
 - d. Pairplot
 - e. Scatterplot
 - f. Lmplot
 - g. box plot