

## **Introduction to Data Science**

### **Learning objectives**

Understand the basics of Data Science and gauge the current landscape and opportunities. Get acquainted with various analysis and visualization tools used in data science.

### **Topics**

- What is Data Science?
- Data Analytics Landscape
- Life Cycle of a Data Science Project
- Data Science Tools and Technologies

## **Mastering Python**

### **Learning objectives**

The Python module will equip you with a wide range of Python skills. You will learn to:

- To Install Python Distribution - Anaconda, basic data types, strings, and regular expressions, data structures and loops, and control statements that are used in Python
- To write user-defined functions in Python
- About Lambda function and the object-oriented way of writing classes and objects
- How to import datasets into Python
- How to write output into files from Python, manipulate and analyse data using Pandas library
- Use Python libraries like Matplotlib, Seaborn, and ggplot for data visualization

### **Topics**

- Python Basics
- Data Structures in Python
- Control and Loop Statements in Python
- Functions and Classes in Python
- Working with Data
- Data Analysis using Pandas
- Data Visualisation

- Case Study

### **Hands-on**

- How to install Python distribution such as Anaconda and other libraries
- To write python code for defining as well as executing your own functions
- The object-oriented way of writing classes and objects
- How to write python code to import dataset into python notebook
- How to write Python code to implement Data Manipulation, Preparation, and Exploratory Data Analysis in a dataset

### **Probability and Statistics**

#### **Learning objectives**

In the Probability and Statistics module you will learn:

- Basics of data-driven values - mean, median, and mode
- Distribution of data in terms of variance, standard deviation, interquartile range
- Basic summaries of data and measures and simple graphical analysis
- Basics of probability with real-time examples
- Marginal probability, and its crucial role in data science
- Bayes' theorem and how to use it to calculate conditional probability via Hypothesis Testing
- Alternate and Null hypothesis - Type1 error, Type2 error, Statistical Power, and p-value

#### **Topics**

- Measures of Central Tendency
- Measures of Dispersion
- Descriptive Statistics
- Probability Basics
- Marginal Probability
- Bayes Theorem
- Probability Distributions
- Hypothesis Testing

## **Hands-on**

- How to write Python code to formulate Hypothesis
- How to perform Hypothesis Testing on an existent production plant scenario

## **Advanced Statistics and Predictive Modelling I**

### **Learning objectives**

Explore the various approaches to predictive modelling and dive deep into advanced statistics:

- Analysis of Variance (ANOVA) and its practicality
- Linear Regression with Ordinary Least Square Estimate to predict a continuous variable
- Model building, evaluating model parameters, and measuring performance metrics on Test and Validation set
- How to enhance model performance by means of various steps via processes such as feature engineering, and regularisation
- Linear Regression through a real-life case study
- Dimensionality Reduction Technique with Principal Component Analysis and Factor Analysis
- Various techniques to find the optimum number of components or factors using screen plot and one-eigenvalue criterion, in addition to a real-Life case study with PCA and FA.

### **Topics**

- Analysis of Variance (ANOVA)
- Linear Regression (OLS)
- Case Study: Linear Regression
- Principal Component Analysis
- Factor Analysis
- Case Study: PCA/FA

## **Hands-on**

- With attributes describing various aspect of residential homes for which you are required to build a regression model to predict the property prices

- Reducing Dimensionality of a House Attribute Dataset to achieve more insights and better modelling

## **Advanced Statistics and Predictive Modelling II**

### **Learning objectives**

Learning Data Science with Python will help you to understand and execute advanced concepts. Take your advanced statistics and predictive modelling skills to the next level in this module covering:

- Binomial Logistic Regression for Binomial Classification Problems
- Evaluation of model parameters
- Model performance using various metrics like sensitivity, specificity, precision, recall, ROC Curve, AUC, KS-Statistics, and Kappa Value
- Binomial Logistic Regression with a real-life case Study
- KNN Algorithm for Classification Problem and techniques that are used to find the optimum value for K
- KNN through a real-life case study
- Decision Trees - for both regression and classification problem
- Entropy, Information Gain, Standard Deviation reduction, Gini Index, and CHAID
- Using Decision Tree with real-life Case Study

### **Topics**

- Logistic Regression
- Case Study: Logistic Regression
- K-Nearest Neighbour Algorithm
- Case Study: K-Nearest Neighbour Algorithm
- Decision Tree
- Case Study: Decision Tree

### **Hands-on**

- Building a classification model to predict which customer is likely to default a credit card payment next month, based on various customer attributes describing customer characteristics

- Predicting if a patient is likely to get any chronic kidney disease depending on the health metrics
- Building a model to predict the Wine Quality using Decision Tree based on the ingredients' composition

## **Time Series Forecasting**

### **Learning objectives**

All you need to know to work with time series data with practical case studies and hands-on exercises.

You will:

- Understand Time Series Data and its components - Level Data, Trend Data, and Seasonal Data
- Work on a real-life Case Study with ARIMA.

### **Topics**

- Understand Time Series Data
- Visualizing Time Series Components
- Exponential Smoothing
- Holt's Model
- Holt-Winter's Model
- ARIMA
- Case Study: Time Series Modelling on Stock Price

### **Hands-on**

- Writing python code to Understand Time Series Data and its components like Level Data, Trend Data and Seasonal Data.
- Writing python code to Use Holt's model when your data has Constant Data, Trend Data and Seasonal Data. How to select the right smoothing constants.
- Writing Python code to Use Auto Regressive Integrated Moving Average Model for building Time Series Model
- Use ARIMA to predict the stock prices based on the dataset including features such as symbol, date, close, adjusted closing, and volume of a stock.

## **Capstone Project**

### **Learning objectives**

This industry-relevant capstone project under the experienced guidance of an industry expert is the cornerstone of this applied **Data Science with Python course**. In this immersive learning mentor-guided live group project, you will go about executing the data science project as you would any business problem in the real-world.

### **Hands-on**

- Project to be selected by candidates.