



## Course in Data Science

### About the Course:

In this course you will get an introduction to the main tools and ideas which are required for **Data Scientist/Business Analyst/Data Analyst**. The course gives an overview of the data, questions, and tools that data analysts and data scientists work with. There are two components to this course. The first is a conceptual introduction to the ideas behind turning data into actionable knowledge. The second is a practical introduction to the tools that will be used in the program like R Programming, SAS, MINITAB and EXCEL.

### Course features:

- ✓ Exclusive doubt clarification session on every weekend
- ✓ Real Time Case Study driven approach
- ✓ Live Project
- ✓ Placement Assistance

### Qualification

- ✓ Any Graduate. No programming and statistics knowledge or skills required

### Duration of the course:

- ✓ 4 months(On working days-one hour and weekends-2hrs) .

### Mode of course delivery

- ✓ Classroom/Online Training

### Faculty Details:

- ✓ A team of faculty having an average 20 + years' experience in the data analysis across various industries and training.

## **Module:1 - Descriptive & Inferential Statistics:(35 Hrs)**

### **1. Turning Data into Information**

- ✓ Data Visualization
- ✓ Measures of Central Tendency
- ✓ Measures of Variability
- ✓ Measures of Shape
- ✓ Covariance, Correlation
- ✓ Using Software-Real Time Problems

### **2. Probability Distributions**

- ✓ Probability Distributions: Discrete Random Variables
- ✓ Mean, Expected Value
- ✓ Binomial Random Variable
- ✓ Poisson Random Variable
- ✓ Continuous Random Variable
- ✓ Normal distribution
- ✓ Using Software-Real Time Problems

### **3. Sampling Distributions**

- ✓ Central Limit Theorem
- ✓ Sampling Distributions for Sample Proportion,  $p$ -hat
- ✓ Sampling Distribution of the Sample Mean,  $\bar{x}$
- ✓ Using Software-Real Time Problems

### **4. Confidence Intervals**

- ✓ Statistical Inference
- ✓ Constructing confidence intervals to estimate a population Mean, Variance, Proportion
- ✓ Using Software-Real Time Problems

### **5. Hypothesis Testing**

- ✓ Hypothesis Testing
- ✓ Type I and Type II Errors
- ✓ Decision Making in Hypothesis Testing
- ✓ Hypothesis Testing for a Mean, Variance, Proportion
- ✓ Power in Hypothesis Testing
- ✓ Using Software-Real Time Problems

### **6. Comparing Two Groups**

- ✓ Comparing Two Groups
- ✓ Comparing Two Independent Means, Proportions
- ✓ Pairs wise testing for Means
- ✓ Two Variances Test(F-Test)
- ✓ Using Software-Real Time Problems

### **7. Analysis of Variance (ANOVA)**

- ✓ One-Way and Two-way ANOVA
- ✓ ANOVA Assumptions
- ✓ Multiple Comparisons (Tukey, Dunnett)
- ✓ Using Software-Real Time Problems

### **8. Association Between Categorical Variables**

- ✓ Two Categorical Variables Relation
- ✓ Statistical Significance of Observed Relationship / Chi-Square Test
- ✓ Calculating the Chi-Square Test Statistic
- ✓ Contingency Table
- ✓ Using Software-Real Time Problems

## **Module:2 – Prediction Analytics (35Hrs)**

### **1. Simple Linear Regression**

- ✓ Simple Linear Regression Model
- ✓ Least-Square Estimation of the Parameters
- ✓ Hypothesis Testing on the Slope and Intercept
- ✓ Coefficient of Determination
- ✓ Using Software-Real Time

### **2. Multiple Linear Regression**

- ✓ Multiple Regression Models
- ✓ Estimation of Model Parameters
- ✓ Hypothesis Testing in Multiple Linear Regression
- ✓ Multicollinearity
- ✓ Using Software-Real Time Problems

### **3. Model Adequacy Checking**

- ✓ Residual Analysis
- ✓ The PRESS Statistic
- ✓ Detection and Treatment of Outliers
- ✓ Lack of Fit of the Regression Model
- ✓ Using Software-Real Time Problems

### **4. Transformations**

- ✓ Variance-Stabilizing Transformations
- ✓ Transformations to Linearize the Model
- ✓ Box-Cox, Tidwell Transformations
- ✓ Generalized and Weighted Least Squares
- ✓ Using Software-Real Time Problems

### **5. Diagnostics for Leverage and Influence**

- ✓ Leverage/ Cook's D /DFFITS/DFBETAS
- ✓ Treatment of Influential Observations
- ✓ Using Software-Real Time Problems

### **6. Polynomial Regression**

- ✓ Polynomial Model in One/ Two /More Variable
- ✓ Using Software-Real Time Problems

### **7. Dummy Variables**

- ✓ The General Concept of Indicator Variables
- ✓ Using Software-Real Time Problems

### **8. Variables Selection and Model Building**

- ✓ Forward Selection/Backward Elimination
- ✓ Stepwise Regression
- ✓ Using Software-Real Time Problems

### **9. Generalized Linear Models**

- ✓ Concept of GLM
- ✓ Logistic Regression
- ✓ Poisson Regression
- ✓ Negative Binomial Regression
- ✓ Exponential Regression

### **10. Autocorrelation**

- ✓ Regression Models with Autocorrelation Errors

## **Module:3 – Applied Multivariate Analysis (20hrs)**

### **1. Measures of Central Tendency, Dispersion and Association**

- ✓ Measures of Central Tendency/  
Measures of Dispersion
- ✓ Using Software-Real Time Problems

### **2. Multivariate Normal Distribution**

- ✓ Exponent of Multivariate Normal  
Distribution
- ✓ Multivariate Normality and Outliers
- ✓ Eigenvalues and Eigenvectors
- ✓ Spectral Decomposition
- ✓ Single Value Decomposition
- ✓ Using Software-Real Time Problems

### **3. Sample Mean Vector and Sample Correlation**

- ✓ Distribution of Sample Mean Vector
- ✓ Interval Estimate of Population Mean
- ✓ Inferences for Correlations
- ✓ Using Software-Real Time Problems

### **4. Principal Components Analysis (PCA)**

- ✓ Principal Component Analysis (PCA)  
Procedure
- ✓ Using Software-Real Time Problems

### **5. Factor Analysis**

- ✓ Principal Component Method
- ✓ Communalities
- ✓ Factor Rotations
- ✓ Varimax Rotation

### **6. Discriminant Analysis**

- ✓ Discriminant Analysis (Linear/Quadratic)
- ✓ Estimating Misclassification Probabilities
- ✓ Using Software-Real Time Problems

### **7. MANOVA**

- ✓ MANOVA
- ✓ Test Statistics for MANOVA
- ✓ Hypothesis Tests
- ✓ MANOVA table
- ✓ Using Software-Real Time Problems

## Module:4 - Machine Learning(30hrs)

### 1. Introduction

- ✓ Application Examples
- ✓ Supervised Learning
- ✓ Unsupervised Learning

### 2. Regression Shrinkage Methods

- ✓ Ridge Regression
- ✓ Lasso Regression
- ✓ Using Software-Real Time Problems

### 3. Classification

- ✓ Variance-Bias Tradeoff
- ✓ Gradient Descent/Ascent Procedure
- ✓ Maximum Likelihood Method
- ✓ Logistic Regression
- ✓ Bayes Law
- ✓ Naïve Bayes
- ✓ Nearest-Neighbor Methods (K-NN Classifier)
- ✓ Using Software-Real Time Problems

### 4. Tree-based Methods

- ✓ The Basics of Decision Trees
- ✓ Regression Trees
- ✓ Classification Trees
- ✓ Ensemble Methods
- ✓ Bagging, Bootstrap, Random Forests, Boosting
- ✓ Using Software-Real Time Problems

### 5. Neural Networks

- ✓ Introduction
- ✓ Single Layer Perceptron
- ✓ Multi-layer Perceptron
- ✓ Forward Feed and Backward Propagation
- ✓ Using Software-Real Time Problems

### 6. Support Vector Machine

- ✓ Maximum Marginal Classifier
- ✓ Support Vector Classifier
- ✓ Kernel Trick
- ✓ Support Vector Machine
- ✓ SVMs with More than Two Classes
- ✓ Using Software-Real Time Problems

### 7. Cluster Analysis

- ✓ Agglomerative Hierarchical Clustering
- ✓ K-Means Procedure
- ✓ Medoid Cluster Analysis
- ✓ Using Software-Real Time Problems

### 8. Dimensionality Reduction

- ✓ Principal Component Analysis
- ✓ Using Software-Real Time Problems

### 9. Association rules

- ✓ Market Basket Analysis
- ✓ Apriori/Support/Confidence/Lift
- ✓ Using Software-Real Time Problems

## Module:5 - R Programming (20hrs)

### 1. R Programming

- ✓ R Basics
- ✓ Numbers, Attributes
- ✓ Creating Vector
- ✓ Mixing Objects
- ✓ Explicit Coercion
- ✓ Formatting Data Values
- ✓ Matrices, List, Factors, Data Frames, Missing Values, Names
- ✓ Reading and Writing Data
- ✓ Using Dput/DDump
- ✓ Interface to the Outside world
- ✓ Sub setting R objects
- ✓ Vectorized Operations
- ✓ Dates and Times
- ✓ Managing Data Frames with the DPLYR package
- ✓ Control Structures
- ✓ Functions
- ✓ Lexical /Dynamic Scoping
- ✓ Loop Functions
- ✓ Debugging

### 2. Data Analytics Using R

- ✓ Module 1-4 demonstrated using R programming