

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/Analytics Manager/Actuarial Scientist/Business Analytic Practitioners**. The course gives an overview of the data, questions, and tools that data analysts and data scientists work with. The course is a combination of various data science concepts such as machine learning, visualization, data mining, programming, data munging, etc. There are three components to this course. The first is a conceptual introduction to the ideas behind turning data into actionable knowledge. The second is manual calculations will be shown on how formulae's are used behind the logics. The third is a practical introduction to the tools that will be used in the program like R Programming and EXCEL.

Course features:

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

Pre-Requisite / Qualification:

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

Duration of the course:

- ✓ 90 Hours (On working days-one hour and weekends-3hrs).

Mode of course delivery:

- ✓ Classroom/Online Training

INTRODUCTION

- What is Data Science? – Introduction.
- What background is required?
- Why Data Science?
- Importance of Data Science.
- Demand for Data Science Professional.
- Brief Introduction to Big data and Data Analytics.
- Lifecycle of data science.
- Tools and Technologies used in data Science.
- What is Machine Learning?
- Different types of Data Science Tasks.

BUSINESS STATISTICS

- Descriptive statistics and Inferential Statistics
- Sample and Population
- Variables and Data types
- Percentiles
- Measures of Central Tendency
- Measures of Spread
- Skeweness, Kurtosis
- Degrees of freedom
- Variance, Covariance, Correlation
- Standardization/Scaling
- Probability
- Expected of 'x'
- Sampling Distribution
- Standard Probability Distribution Functions
- Bernoulli, Binomial, Normal distributions
- Standard Normal Deviate
- Decision Making Rules
- Test of Hypothesis
- One sample t-Test, Chi-square
- Two sample t-Test Analysis of Variance (ANOVA)

EXPLORATORY DATA ANALYSIS AND VISUALIZATION

- Summary Statistics
- Data Transformations
- Outlier Detection and Management
- Charts and Graphs
- One Dimensional Chart
- Box plots
- Bar graph
- Histogram
- Scatter plots
- Multi-Dimensional Charts
- Fancy Charts - Bubble charts

DATA PRE-PROCESSING

- Data Types and Conversions
- Binning, Scaling, Standardization, Normalization
- Min-max Scaling
- Missing values Treatment
- Imputation

QSHORE TECHNOLOGIES

Reach us at 9030821111

Email: Info@qshore.com

PREDICTION ANALYTICS

- Simple Linear Regression
- Multiple Linear Regression
- Estimation of Model Parameters
- Hypothesis Testing in Multiple Linear Regression
- Extra sum of squares
- R – Square, R- Square Adjusted
- Variable Selection
 - a. All Possible Regressions
 - b. Sequential Selection (Forward, Backward, Stepwise)
- Multicollinearity – VIF
- Residual Analysis/Regression Diagnostics.
- Polynomial Regression
- Transformations
 - a. Bulging Rules
 - b. Box Tidwell
 - c. Box cox
 - d. Weighted Least Square
- Dummy variables
 - a. General Concepts of Indicator variables.
- Predicted Error sum of squares (PRESS)
- Assessing Performance
 - a. Variance Biased Trade-off
 - b. Resampling Methods
 - c. Cross Validation
 - d. Leave one out Cross validation
 - e. k-Fold Cross Validation
 - f. Bootstrap
- Logistic Regression
 - A Case Study will be presented on Logistic Regression

MACHINE LEARNING

Introduction to Supervised and unsupervised Learning

➤ Neural Networks

- a. Network Topology
- b. Single Layer Perceptron
- c. Multi-Layer perceptron
- d. Feed forward and Back propagation Models

➤ Introduction to Deep Learning

➤ Association Rules

- a. Market Basket Analysis
- b. APRIORI
- c. Support, Lift, Confidence

➤ Nearest-Neighbour Methods (KNN – Classifier)

- a. Euclidian Distance
- b. Hamming Distance

➤ Decision Tree

- a. Finding Root Node, Intermediate Nodes, Terminal Nodes
- b. Construction of Rules
- c. Miss classification
- d. Gini Index
- e. Overfitting and Pruning
- f. Regression Trees

➤ Boosting, Bagging and Random Forest

- a. Resampling Methods
- b. Resampling methods with Replacement
- c. Resampling methods without Replacement
- d. Random Forest

➤ Dimensional Reduction Techniques

1. Principle Component Analysis
 - a. Eigen values and Eigen Vectors
2. Cluster Analysis
 - a. Hierarchal Clustering
 - b. Linkage Methods
 - c. Non- Hierarchal Clustering
 - d. K-Means Clustering

➤ **Text Mining / Natural Language processing**

- a. Unstructured Data
- b. Text Analytics
- c. Cleaning Text data
- d. Tokenization
- e. Pre-processing
- f. Word counts and word clouds
- g. Sentiment Analysis
- h. Text classification
- i. Distance measures

➤ **Introduction to probabilistic methods Introduction**

- a. Naive Bayes
- b. Joint and Condition probabilities
- c. Classification using Naive Bayes Approach

➤ **Support Vector Machines**

- a. Maximum Margin Classifier
- b. Support vector Classifier
- c. Support vector machines
- d. Kernels – Linear and Non Linear

➤ **PYTHON - PROGRAMMING**

- How to install python (Anaconda)
- How to install sciKit Learn (Anaconda)
- How to work with Jupyter Notebook
- How to work with Spyder IDE
- Strings
- Lists
- Tuples
- Sets
- Dictionaries
- Control Flows
- Functions
- Formal/Positional/Keyword arguments
- Predefined functions (range, len, enumerates etc...)
- Data Frames
- Packages required for data Science in Python
- Lab/Coding

➤ **Introduction to NumPy**

- One-dimensional Array
- Two-dimensional Array
- Pr-defined functions (arrange, reshape, zeros, ones, empty)
- Basic Matrix operations
- Scalar addition, subtraction, multiplication, division
- Matrix addition, subtraction, multiplication, division and transpose
- Slicing
- Indexing
- Looping
- Shape Manipulation
- Stacking

➤ **Introduction to Pandas**

- Series
- DataFrame
- df.GroupBy
- df.crosstab
- df.apply
- df.map

➤ **Apache Spark Analytics**

What is Spark

Introduction to Spark RDD

Introduction to Spark SQL and Dataframes

Using R-Spark for machine learning

Hands-on:

installation and configuration of Spark

Hands on Spark RDD programming

Hands on of Spark SQL

Dataframe programming

Using R-Spark for machine learning programming

➤ R – PROGRAMMING

1. Getting R
 - 1.1 Downloading R
 - 1.2 R Version
 - 1.3 32-bit versus 64-bit
 - 1.4 Installing

2. The R Environment
 - 2.1 Command Line Interface
 - 2.2 RStudio

3. R Packages
 - 3.1 Installing Packages
 - 3.2 Loading Packages

4. Reading Data into R
 - 4.1 Reading CSVs
 - 4.2 Excel Data
 - 4.3 Clipboard

5. Advanced Data Structures
 - 5.1 Data.frames
 - 5.2 Lists
 - 5.3 Matrices
 - 5.4 Arrays
 - 5.5. Factors

6. Basics of R
 - 6.1 Basic Math
 - 6.2 Variables
 - 6.3 Data Types
 - 6.4 Vectors
 - 6.5 Calling Functions
 - 6.6 Function Documentation
 - 6.7 Missing Data

7. Control Statements
 - 7.1 if and else
 - 7.2 switch
 - 7.3 ifelse

8. Loops
 - 8.1 for Loops

- 8.2 while Loops
- 8.3 Controlling Loops

- 9. Group Manipulation
 - 9.1 Apply Family
 - 9.2 aggregate

- 10. Data Reshaping
 - 10.1 cbind and rbind
 - 10.2 Joins
 - 10.3 Reshape2

- 11. String Theory
 - 11.1 paste
 - 11.2 sprintf
 - 11.3 Extracting Text/ Regular Expressions

- 12. Graphs with R and GGPlot2
 - 12.1 Basic and Interactive Plots
 - 12.2 Dendrograms
 - 12.3 Pie Chart and Its Alternatives
 - 12.4 Adding the Third Dimension
 - 12.5 Visualizing Continuous Data

- 13. Basic Statistics
 - 13.1 Summary Statistics
 - 13.2 Correlation and Covariance
 - 13.3 T-Tests
 - 13.4 ANOVA

- 14. Probability Distributions
 - 14.1 Normal Distribution
 - 14.2 Binomial Distribution

Course Highlights

- ✓ A Dedicated Portal For Practicing.
- ✓ Real Time Project Data Models to Work
- ✓ 1-1 Mentorship
- ✓ Internship Offers for Freshers.
- ✓ Weekly Assignments.
- ✓ Weekly Doubt Sessions\
- ✓ Resume Preparation Tips
- ✓ Interview Guidance And Support.
- ✓ Dedicated HR Team for Job Support And Placement Assistance.