

Business Analytics Using R	
Day One	Day Two
R Programming – Part I	Statistics – Part I
<ul style="list-style-type: none"> • Introduction to Business Analytics • The R Environment • R Packages 	<ul style="list-style-type: none"> • Introduction to Statistics • Measures of Central Tendency • Measures of Dispersion
Short Break	
<ul style="list-style-type: none"> • Basics of R • Advanced Data Structures • Reading Data into R 	<ul style="list-style-type: none"> • Exploratory Data Analysis (Histogram, Boxplot, Stem & Leaf plot, Scatterplot matrix, QQ plot) • Basic Probability Ideas • Probability Distributions (Binomial, Poisson & Normal)
Lunch Break	
<ul style="list-style-type: none"> • Basic Graphics (Column, Line, Pie, Bar, Scatter & Surface) • Writing R Functions • Control Statements • Loops, the R Way 	<ul style="list-style-type: none"> • T-tests (one-sample, paired-samples & independent-samples) • ANOVA (One-Way and Two-Way) • Specialized Graphs (Multi-axis, Multi-panel, Pareto, Radar, Bubble)
Short Break	
<ul style="list-style-type: none"> • Group Manipulation • Data Reshaping • Manipulating Strings 	<ul style="list-style-type: none"> • Categorical Data Analysis • Correlation & Covariance

Business Analytics Using R	
Day Three	Day Four
Statistics - Part II	Machine Learning - Part I
<ul style="list-style-type: none"> • An Introduction Mathematical Modelling (KNA) • Acquire And Prepare The Data <ul style="list-style-type: none"> ➤ Reading data from - CSV files, XML data, JSON data, fixed width formatted files, R data files and R libraries; Removing cases with missing values, Replacing missing values with the mean, Removing, duplicate cases, Rescaling a variable to [0,1], Normalizing data in a data frame, Binning numerical data, Creating dummies for categorical variables 	<ul style="list-style-type: none"> • Feature Selection and Dimension Reduction <ul style="list-style-type: none"> ➤ Remove redundant features using correlation, Rank features by importance using LVQ, Feature selection using RFE, Performing dimension reduction with PCA, Determining the number of principal components using a scree test, Determining the number of principal components using the Kaiser method, Visualizing multivariate data using biplot
Short Break	
<ul style="list-style-type: none"> • Linear Regression <ul style="list-style-type: none"> ➤ Introduction, Simple linear regression, Estimating regression coefficients, Multiple linear regression, Examples, Assessing models (Residual analysis, Significance tests, Performance metrics, Comparing models, Test set performance), Problems (Multicollinearity, Outliers), Feature selection, Regularization 	<ul style="list-style-type: none"> • Probabilistic And Lazy Learning <ul style="list-style-type: none"> ➤ Understanding classification using nearest neighbors, The kNN algorithm, Why is the kNN algorithm lazy?, Example ➤ Understanding naive Bayes, Basic concepts of Bayesian methods, The naive Bayes algorithm, Example
Lunch Break	
<ul style="list-style-type: none"> • Logistic Regression <ul style="list-style-type: none"> ➤ The classification problem, Introduction, Generalized linear models, Interpreting coefficients, Assumptions, Maximum likelihood estimation, Examples, Assessing models (Model deviance, Test set performance), Regularization, Classification metrics, Extensions (Multinomial, Ordinal) 	<ul style="list-style-type: none"> • Decision Trees <ul style="list-style-type: none"> ➤ Understanding decision trees, The C5.0 decision tree algorithm, Choosing the best split, Pruning the decision tree, Example
Short Break	

<ul style="list-style-type: none"> • Clustering <ul style="list-style-type: none"> ➤ Clustering data with hierarchical clustering, Cutting a tree into clusters, Clustering data with the k-means method, Drawing a bivariate cluster plot, Comparing clustering methods, Extracting silhouette information from clustering, Obtaining optimum clusters for k-means, Clustering data with the density-based method, Clustering data with the model-based method, Visualizing a dissimilarity matrix, Validating clusters externally 	<ul style="list-style-type: none"> • Evaluating And Improving Model Performance <ul style="list-style-type: none"> ➤ Measuring performance for classification, Beyond accuracy - other measures of performance, Visualizing performance tradeoffs, Estimating future performance ➤ Tuning stock models for better performance, Improving model performance with meta-learning, Understanding ensembles (Bagging, Boosting), Random forests
Business Analytics Using R	
Day Five	
Machine Learning – Part II	
<ul style="list-style-type: none"> • Black Box Methods – Neural Networks and Support Vector Machines <ul style="list-style-type: none"> ➤ Understanding neural networks, From biological to artificial neurons, Activation functions, Network topology, Training neural networks with backpropagation, Example ➤ Understanding support vector machines, Classification with hyperplanes, Finding the maximum margin, The case of linearly separable data, The case of non-linearly separable data, Using kernels for non-linear spaces, Example 	
Short Break	
<ul style="list-style-type: none"> • Clustering <ul style="list-style-type: none"> ➤ Clustering data with hierarchical clustering, Cutting a tree into clusters, Clustering data with the k-means method, Drawing a bivariate cluster plot, Comparing clustering methods, Extracting silhouette information 	
Lunch Break	
<ul style="list-style-type: none"> ➤ Advanced Data Manipulation And Writing Programs In R (SKC) ➤ Bridging IT And Analytics And Developing Models (SKC) 	
Short Break	

➤ Case Study Discussion (SKC)	
-------------------------------	--

OVERVIEW

This course covers basics of 'predictive' data analytics, mainly, from a Machine Learning perspective. 'R' programming language is used as an enabler for data analytics.

SYSTEM REQUIREMENTS

1. Windows Operating System with minimum of 4GB memory.
2. Internet connection to download packages

AUDIENCE

This course is recommended for working professionals from any stream who want to learn and practice data analytics.

PRE-REQUISITES

Basic knowledge of: mathematics (linear algebra, calculus), programming, SQL, statistics.

But overall a good analytical mindset is required.

PRE-READING

Download a basic "R" programming language introduction from:

https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

Basic statistical terms and definitions:

Lots of information can be found on internet. For a sample reference visit:

<http://documents.software.dell.com/Statistics/Textbook/Basic-Statistics>

CRISP-DM (Cross Industry Standard Process for Data Mining), a data mining process model:

https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

BUY BOOK

Essentials of Business Analytics by Jeffrey Camm et al [Price: ₹ 599.00]:

<https://www.cengage.co.in/category/academic-professional/business-economics/operation-decision-sciences/course/essentials-of-business-analytics-b2>

FREE DATA SCIENCE EBOOKS FOR YOUR READING LIST

<http://www.datasciencecentral.com/profiles/blogs/5-free-data-science-ebooks-for-your-summer-reading-list>

SOFTWARE INSTALL (OPEN SOURCE)

Before coming to the class for the first time, install R (current version is 3.3.0) for windows from the following link:

<https://cran.r-project.org/bin/windows/base/>

By default, some base packages would have been installed in R. A couple of times, during the class, we may ask you to download certain packages through internet. So WIFI needs to be enabled with the necessary system admin permissions.

COURSE DURATION

This is five weekend (Sunday's only) OR a three weekend (Saturday & Sunday) course.

CAVEAT

This is not a programming course.