



CALCUTTA BUSINESS SCHOOL

MANAGEMENT DEVELOPMENT PROGRAMME

ON

Data Analytics Using R

PROGRAMME DIRECTOR

Prof. Jaydip Sen

Faculty, Calcutta Business School

Date: November 19-21, 2015

Venue:

**Management Development & Research Centre,
Calcutta Business School**

CALCUTTA BUSINESS SCHOOL

Diamond Harbour Road, Bishnupur – 743 503, 24 Parganas (South)

Phone: + 91-33-2420 5200 (Extn 256); Email: parthas@calcuttabusinessschool.org; praj nab@calcuttabusinessschool.org
www.calcuttabusinessschool.org

Programme Objective:

Business Intelligence and Analytics (BI &A) and their related field of Big Data Analytics have become increasingly important for business communities over the past two decades. There has been an explosion of amount of data available in the world. The experts believe that by end of 2015, total amount of data generation will reach a gigantic volume of 7.9 Zettabyte, where 1 Zettabyte is equivalent to 1 trillion Gigabytes. Analytics is being used today and will be more used tomorrow for various applications involving large volumes of data like life science research, analysis of consumer behavior, analysis of social media usage, weather forecasting, accurate healthcare services etc. Thus, it has become very important to understand, infer and learn quickly and often in real-time from the humongous data in today's information-driven business world. It is quite natural that a large number of analytics tools and software have evolved over the last few years to handle these challenges. However, the programming language R is rapidly becoming the de facto standard among the professionals in the analytics industry because of its extremely rich set of libraries and because of it being an open source tool. This three-day management development programme covers almost all the fundamentals aspects of R language and will be ideal for professionals who are new into this area. After completion of this course, the participants will be equipped with enough expertise and knowledge to handle more complex and advanced topics in R like machine learning application, social network analysis and text mining.

Details of the Program Contents

The program will cover the following broad topics

- **Introduction to R** – reading and getting data into R, storing the results of computation, using 'scan' command for making data, reading a file data from a disk, handling missing values in data files.
- **Working with R objects** – manipulating vectors, matrix, data frames, and lists, viewing objects within objects, constructing data objects, conversions of data objects.
- **Descriptive statistics and tabulation of data** – summary commands, summarizing samples, cumulative statistics, summary statistics for data frames, matrix and lists, summary tables, cross tabulation, recreating original data form a contingency table.
- **Data distribution** – stem and leaf plots, histograms, density functions, various types of data distribution, random sampling, the Shapiro-Wilk test for normality, the Kolmogorov-Smirnov test, quantile-quantile plots.
- **Hypothesis testing** – Student's t-test for two-samples with equal/unequal variances, one-sample t-test, the Wilcoxon U-test, Paired t- and U-tests, tests for association – multiple categories chi-squared tests, Monte-Carlo simulation, Yates' correction for 2X2 tables, goodness of fit tests.
- **Graphical analysis of data** – Box-whisker plots, scatter plots, paired plots (multiple correlation plots), line charts, pie charts, Cleveland dot charts, bar charts.
- **Formula notation and complex statistics** – examples of using formula syntax for basis tests, formula notation in graphics, one-way ANOVA, -simple post-hoc testing, two-way ANOVA – graphical summary of ANOVA and post-hoc testing, extracting means and summary statistics, model tables and table commands, interaction plots.
- **Manipulating data and extracting components** – creating data for complex analysis, creating and setting data frames, matrix and factor data, making replicate treatment factors, summarizing data-simple column and row summaries, complex summary functions, summarizing using grouping variables and aggregate command.
- **Linear data modeling and regression** – simple linear regression, coefficients, fitted values, residuals, formula, best-fit lines, similarities between linear model and ANOVA, multiple regression – model building, adding terms with forward stepwise regression and removing terms with backwards deletion, curvilinear regression- logarithmic and polynomial regression, plotting linear models and curve fitting, summarizing regression models.
- **Writing R program scripts** - creating simple R functions, using default values in functions, displaying the results of customized functions and scripts.

Methodology:

All the topics will be explained both by a detailed explanation of the theoretical fundamentals and by hand-on illustrations in the lab. Appropriate data sets and case studies will be used for explanation and demonstration in the lab.

Target Group:

This programme will be useful for anyone who needs to analyze any data whatever his/her discipline or line of work. Professionals working in industry and aspiring to work in the booming and lucrative domain of analytics will find the course particularly beneficial. It will also be useful for academicians and researchers who are interested in data analytics and who intend to build their knowledge in the R programming language. Since any professional of today will have to gather data, analyze it and present the results of analysis in a very effective visualization medium and because R is a very powerful tool to perform all these activities, the course will be beneficial to professionals of all domains – science, business, medicine, or engineering.

Course Director's Profile:

Prof. Jaydip Sen obtained his Bachelor of Engineering (B.E) with honors in Mechanical Engineering from Jadavpur University, Kolkata, India in 1989 and Master of Technology (M.Tech) with honors in Computer Science from Indian Statistical Institute, Kolkata in 2001. He has around 18 of years of diverse experience in the industry, academic and research. He has worked with reputed organizations like Oil and Natural Gas Corporation Ltd., India, Oracle India Pvt. Ltd., and Tata Consultancy Services Ltd. Currently, he is the area coordinator of IT and Systems Faculty in Calcutta Business School for the last one year. His research areas include security in wired and wireless networks, intrusion detection systems, secure routing protocols in wireless ad hoc and sensor networks, secure multicast and broadcast communication in next generation broadband wireless networks, trust and reputation based systems, quality of service in multimedia communication in wireless networks and cross layer optimization based resource allocation algorithms in next generation wireless networks. He has more than 100 publications in reputed international journals and referred conference proceedings. He has delivered expert talks and keynote lectures in various international conferences and symposia across the globe. He is a member of ACM and IEEE and also a working member of IEEE 802.16 group. He has around 10 patents to his credit. Currently his area of interest includes Big Data Analytics and its various applications in business



Participation fees and payment:

Participation Fee: Rs.15,000/ (Rupees Fifteen Thousand only) + applicable service tax per participant, which includes necessary course material, food during the programme. Participation fees are to be paid by Demand Draft / Cheque drawn in favour of "Calcutta Business School" and payable on any bank in Kolkata.

About Calcutta Business School (CBS):

Promoted by a group of industrialists and academicians, CBS, an autonomous institution, commenced its journey in July 2008, offering an innovative and futuristic, fully residential 2 year Post Graduate Diploma in Management (PGDM) programme approved by AICTE. It is located on a sprawling 15 acre campus, about 7 kms south of IIM Calcutta. In addition to world-class full-time faculty, outstanding distinguished Professors and Visiting Professors from all over the world (24 from leading business schools in the USA and Australasia) provide the academic inputs. In the state-of-the-art 'intelligent campus' all students are provided with powerful laptop computing devices and high bandwidth wireless Internet connectivity. The pedagogy includes individual and team-based projects and case studies. The "Business Analytics" lab in the CBS campus is equipped with power computing workstations which are installed with most of the state-of-the-art analytics tools and software including SPSS, R, Python, CANOPY, Scientific Python, MATLAB and many SaS analytics tools. CBS has a tie-up with the global analytics leader SaS, and a six-months' programme on 'Data Science' and 'Big Data Analytics' is offered jointly by CBS and SaS institute, India.

Reaching CBS Campus:

The participants who will not stay in the campus during the programme, will be provided conveyance to and from Tollygunge Metro Station to the CBS campus on each day of the program. The timing of the service will be as follows:

Departure for Tollygunge Metro Station : 8:30 AM.

Departure from Calcutta Business School : 6:15 PM

Contact : Partha Sen, Senior Manager - Management Development Programme

CALCUTTA BUSINESS SCHOOL

Diamond Harbour Road, Bishnupur 743503, 24 Parganas (South), West Bengal

Tel: 033-2420 5200 (Extn 256), Telefax: 033 2480 9612, Mob: +919674794042

Email: parthas@calcuttabusinessschool.org

