

# **Certificate in Business Analytics**

## **Course outline for Collaborative Program**

**Program Audience:** Undergraduate Students, Graduate Students

**Course Mode** : Lecture, Tutorial, Practical

### **Course Objectives**

This course is designed to enable learners with the crucial step in the Data Lifecycle: using data to make decisions!

- Possess the modeling skills needed by companies all over the world to go beyond storing data to understanding data
- Learn how to use these skills to make decisions such as cancer detection, fraud detection, customer segmentation and predicting machine downtime.
- Get introduced to the data mining process and modeling techniques using one of the most popular software, IBM's SPSS Modeler.
- Learn how to build models on trained data, test the model with historical data, and use qualifying models on live data or other historical untested data.
- Save or earn companies millions of dollars with your decisions!

### **Course Pre-requisite**

- 3.1. A familiarity with the basic concepts in statistics will be useful for Predictive Analytics sessions
- 3.2. Participants must have working knowledge of Windows , MS Excel etc
- 3.3. Concepts of diverse data set will be useful

## **Program Content**

### **Basics of Statistics**

This module is divided into two parts –

Foundation of Statistics & Mathematical Optimization for Business Problems

### **Foundation of Statistics**

This is a beginner's course covering the fundamentals of statistics. Start with mean, mode, and median. Then learn about standard deviation using examples from basketball. Learn about probability with dice. Learn what it means to group data by categorical variables, and how you can transform your data into appropriate graphs and charts.

This course is taught using SPSS Statistics. No prior experience necessary.

#### **Module 1 - Welcome to Statistics!**

- Welcome to Statistics
- Data visualization
- All about data
- SPSS Statistics
- SPSS Statistics in 15 minutes
- Lab exercises

#### **Module 2 - Basic Statistics**

- Types of data
- Measures of dispersion
- Mean, median, mode
- Statistics by datatype
- Probability
- Lab exercises

#### **Module 3 - Summarizing data**

- Statistics by groups
- Visualization of group statistics
- Pivoting
- Cross-tabulations
- Correlation
- Lab exercises

#### **Module 4- Data Visualization**

- Visualization fundamentals
- Descriptive and statistical charts
- Scatterplots

- Statistical charts
- Time series charts
- Lab exercises

#### Module 5 – Descriptive Statistics

- Weighted means, standard deviations
- Data wrangling
- Descriptive Statistics
- Reproducibility with syntax in SPSS Statistics
- Lab exercises

### **Mathematical Optimization for Business Problems**

Mathematical Programming is a powerful technique used to model and solve optimization problems. This training provides the necessary fundamentals of mathematical programming and useful tips for good modeling practice in order to construct simple optimization models. In this training, students will explore several aspects of mathematical programming to start learning more about constructing optimization models using IBM Decision Optimization technology, including:

- Basic terminology: operations research, mathematical optimization, and mathematical programming
- Basic elements of optimization models: data, decision variables, objective functions, and constraints
- Different types of solution: feasible, optimal, infeasible, and unbounded
- Mathematical programming techniques for optimization: Linear Programming, Integer Programming, Mixed Integer Programming, and Quadratic Programming
- Algorithms used for solving continuous linear programming problems: simplex, dual simplex, and barrier
- Important mathematical programming concepts: sparsity, uncertainty, periodicity, network structure, convexity, piecewise linear and nonlinear

These concepts are illustrated by concrete examples, including a production problem and different network models.

#### Module 1 - The Big Picture

- What is Operations Research?
- What is Optimization?
- Optimization models

#### Module 2 - Linear Programming

- Introduction to Linear programming
- A production problem : Part 1 - Writing the model
- A production problem : Part 2 - Finding a solution
- A production problem : Part 3 - From feasibility to unboundedness
- Algorithms for solving linear programs : Part 1 - The Simplex and Dual Simplex Algorithm
- Algorithms for solving linear programs : Part 2 - The Simplex and Barrier methods

#### Module 3 - Network Models

- Introduction to Network Models
- The Transportation problem
- The Transshipment problem
- The Assignment problem
- The shortest path problem
- Critical path analysis

#### Module 4 - Beyond simple LP

- Nonlinearity and Convexity
- Piecewise linear programming
- Integer programming
- The branch and bound method
- Quadratic Programming

#### Module 5 - Modelling Practice

- Modelling the real world
- The importance of Sparsity
- Tips for better models

### **Foundation program in R Programming**

R is a powerful language for data analysis, data visualization, machine learning, and statistics. Originally developed for statistical programming, it is now one of the most popular languages in data science. In this course, participants will be learning about the basics of R, and will end with the confidence to start writing your own R scripts. But this isn't a typical textbook introduction to R. Not just learning about R fundamentals, they shall be using R to solve problems related to movies data. Using a concrete example makes the learning painless. Learn about the fundamentals of R syntax, including assigning variables and doing simple operations with one of R's most important data structures -- vectors! From vectors, you'll then learn about lists, matrix, arrays and dataframes. Then jump into conditional statements, functions, classes and debugging.

Further the module guides on how "A picture is worth a thousand words". We are all familiar with this expression. It especially applies when trying to explain the insight

Obtained from the analysis of increasingly large data sets. Data visualization plays an essential role in the representation of both small and large scale data.

One of the key skills of a data scientist is the ability to tell a compelling story, visualizing data and findings in an approachable and stimulating way. Learning how to leverage a software tool to visualize data will also enable you to extract information, better understand the data, and make more effective decisions.

The main goal of this course is to teach you how to take data that at first glance has little meaning and present that data in a form that makes sense to people. Various techniques have been developed for presenting data visually but in this course, we will be using the open source language R.

#### Module 1 - R basics

- Math, Variables, and Strings
- Vectors and Factors
- Vector operations

#### Module 2 - Data structures in R

- Arrays & Matrices
- Lists
- Dataframes

#### Module 3 - R programming fundamentals

- Conditions and loops
- Functions in R
- Objects and Classes
- Debugging

#### Module 4 - Working with data in R

- Reading CSV and Excel Files
- Reading text files
- Writing and saving data objects to file in R

#### Module 5 - Strings and Dates in R

- String operations in R
- Regular Expressions
- Dates in R

#### Module 6 - Basic Visualization Tools

- Bar Charts
- Histograms
- Pie Charts

#### Module 7 - Basic Visualization Tools Continued

- Scatter Plots
- Line Plots and Regression

#### Module 8 - Specialized Visualization Tools

- Word Clouds
- Radar Charts
- Waffle Charts
- Box Plots

#### Module 9 - How to create Maps

- Creating Maps in R

### **Digital Analytics & Regression**

Follow a case study where you define the business objective, establish the data required to address that objective, and use R, the programming language, to derive insights from the data. As with any business challenge, you will be required to articulate your findings to a business audience.

- Learn basic concepts in statistics with step-by-step guidance on how to conduct an analysis to solve the business problem.
- Data Science is like triathlon. Programming is cycling, by far the biggest investment is required in hardware and software. Running is domain expertise and communication skills and, swimming is mathematics, statistics and modelling. There are competitions in each of these disciplines, cycling, running and swimming (and there always will be), but the need for super athletes who can do all 3 is growing. An athlete who is brilliant at one discipline can learn the other two and succeed in the triathlon.

#### Module 1 - A Case Study Approach to Analytics

- Understand the business context
- Formulate the business objective
- State the hypothesis
- Assess available data
- Assign data for use

#### Module 2 - Data Scientist Workbench

- Using Data Scientist Workbench
- What is R?
- Loading data into R with Data Scientist Workbench
- Upload a CSV data file into Data Scientist Workbench and RStudio

#### Module 3 - Google Trends Data in R

- Access Google Trends data in R

#### Module 4 - Simple Linear Regression in R

- Regression and Google Trends Data in R
- Box Plots and Histograms in R
- Scatter Plots & Lines of best fit in R
- Simple Linear Regression in R

#### Module 5 - Presenting Data Analytics in Business

- Using data to answer a business question
- Summarizing the data analytics process
- Presenting data insights

### **Cognitive analytics using Watson**

Analyzing data has no value if you cannot tell a story of insight out of it. This is where tools like Watson Analytics can take your analysis to the next level. Watson Analytics offers you the benefits of advanced analytics without the complexity.

- Use Excel to get you started and put you in the right analytical mindset.
- Learn, in a few hours, how to perform data exploration, predictive analytics and effortless dashboard and infographic creation.
- Get answers and new insights to make confident decisions in minutes—all on your own!

#### Module 1 - Overview of Watson Analytics

- Identify Watson Analytics components and the user interface (Overview)

#### Module 2 - Data

- Identify data structure limitations and requirements
- Upload data from flat files
- Shape data before uploading
- Upload data from Twitter

#### Module 3 - Work with data connections

- Create a secure connection
- Create a secure gateway

#### Module 4 - Refine data

- Join data sets
- Add a calculation to the data set
- Add a data group to the data set
- Add a hierarchy

#### Module 5 - Discover patterns, relationships, and predictive insights

- Change column headings
- Sort data
- Perform analysis using natural language questions
- Edit targets in a predictive analysis
- Navigate the spiral in a predictive analysis

- View decision rules and the decision tree

Module 6 - Provide added value to the analysis

- Share the analysis through email

Module 7 - Assemble a display

- Change visualization types in a display
- Modify properties of a display
- Filter visualizations in a display

Module 8 - Control access

- Share assets and set access permissions

References - Create and Use expert Storybooks

- Author an expert storybook
- Use an expert storybook

## **Predictive Analytics Modeling**

The Predictive Analytics Modeler will learn the essential analytics models to collect and analyze data efficiently. This will require skills in predictive analytics models, such as data mining, data collection and integration, nodes, and statistical analysis. The Predictive Analytics Modeler will use tools for market research and data mining in order to predict problems and improve outcomes.

Module 1 - Introduction to Data Mining

- CRISP-DM Methodology
- Introduction to SPSS Modeler - predictive data mining workbench
- SPSS Modeler Interface

Module 2 - The Data Mining Process

- Business Understanding
- Data Understanding
- Data Preparation

Module 3 - Modeling Techniques

- Introduction to Common Modeling Techniques
- Cluster Analysis (Unsupervised Learning)
- Classification & Prediction (Supervised Learning)
- Classification - Training & Testing
- Sampling Data in Classification
- Predictive Modeling Algorithms in SPSS Modeler
- Automated Selection of Algorithms

Module 4 - Model Evaluation

- Metrics for Performance Evaluation
- Accuracy as Performance Evaluation tool
- Overcoming Limitations of Accuracy Measure

- ROC Curves

#### Module - Predictive Analytics Modeler

- Integrating data (methods, options, merging, and sampling)
- Deriving and reclassifying fields (CLEM)
- Looking for relationships (matrix, distribution, means, histogram, statistics and plot)
- Functions (conversion, string, and statistical)
- Data transformation
- Statistical. graphical and sample nodes
- Automated data mining and modeling
- Predictive models and customer segmentation