



Business Analytics and Data Mining for DW/BI Professionals: Jumpstart work-through Workshop

Audience: Experienced Professionals

Experienced DW / BI Professionals, Business Analysts, Data Architects, Team Leads and Project Managers, Information Excellence professionals

Business Success is all about being proactive, learning from the past, anticipating the future, positioning right and actioning the right way at the right time.

That holds very much true (and necessary) in the arena of [Competing with Analytics](#) and [Information Excellence](#). Business metrics do a great job summarizing the past, *Business analytics* helps you predict, anticipate and act proactively with your customers, partners and employees and systemic dependencies. The customer predictions generated by predictive analytics deliver more relevant content to each customer, improving response rates, click rates, buying behaviour, retention and overall profit.

Business Analytics and Data Mining Workshop is the jumpstart, skill oriented work-through workshop. This course is custom designed for the experienced professionals from DWH / BI / Decision Support / Business Analyst backgrounds. A good understanding of the Decision project lifecycle / ETL / KPIs and Metrics is expected, with at least one full life cycle project execution. No prior knowledge of statistics or modeling is expected.

This is a three day intensive workshop. It provides the foundations, fundamentals, techniques and methodologies with relevant hands on case studies for real world simulation. The course covers the most important skills required for building business models for analytics.

The Workshop includes hands on work through of the business cases during the course of the workshop. The workshop uses open source tool kits, towards hands-on sessions and work-through examples. Ensuring you will walk away with the working understanding of Classification, Clustering and Regression Analysis, and workable tools that you can put to use in your company.

At the end of this workshop, we expect the participants to have a good foundational jumpstart understanding of the methods, methodologies and techniques spanning from the basics of statistics and distribution to the specific techniques in the space of analytics.

Further, we always customize the workshop based on the mix of participant profiles and the underlying business driving needs.

Welcome to the journey into the mysterious world of "Learning to Compete with Analytics".

The best way to put distance between you and the crowd is to do an outstanding job with information. How you gather, manage, and use information will determine whether you win or lose.

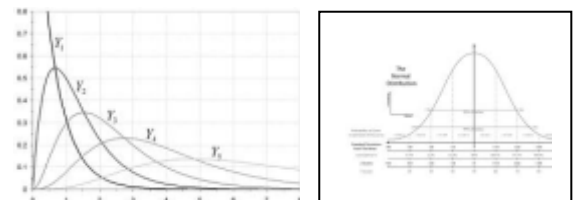
Bill Gates in *Business @ The Speed of Thought*

Foundational understanding

- Basic Data Analysis from the world of DW/BI Reporting
 - 80-20, ABC, RFM, Top N, Rule Based, RFM;
 - Data Explosion and Limitations
- Learning from the DWH and BI World
- Unlearning from the DWH and BI World
- Basic principles and terminology for data mining
 - What is Data Mining
 - Gain Charts
 - Decile Analysis
 - Relating to Analytics: Analogies
- Leveraging the Power of Data: What Can we do with the Data
 - Visualization
 - Description
 - Classification
 - Estimation
 - Prediction
 - Association
 - Clustering

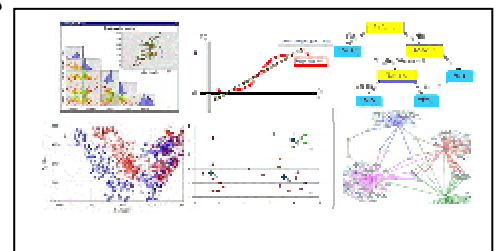
Essential Statistics

- Variable Type
 - Continuous Vs Categorical (Discrete)
 - Univariate vs Bivariate and Multivariate
- Mean, Mode, Median – measures of central tendency
- Variance and Standard Deviation – measures of deviation



Exploratory Data Analysis (Visual Analysis):

- Single variable – Histograms, Bar Charts, Trend Charts, Box plots
- Two variables – Scatter plots, Correlation Plots, Bar Charts
- Multiple variables – Parallel coordinates, Mosaic plots
- Other exploratory visualization techniques
- Hands On Work-through Case



Analytical Methodology and Business Modeling

- Business data model
- Evaluating the need for modeling and Justifying the project

- Building Analytical Data Mart

- CRISP DM Methodology for Analytical Project Lifecycle
 - Business understanding
 - Data understanding
 - Data Preparation
 - Modeling Methods
 - Model Building
 - Model Evaluation: Evaluating and Validating the Model

- SEMMA for Analytical Model Building
 - Defining factors for model
 - Selection of type of model



6. Pre-processing

- Source data mapping
- Source data analysis
- Derived Variables
- Testing distributions
- Testing outlier, multicollinearity
- Binning/Classing
- Treating missing values
- Creating standard data sets – training, testing and validation
- Data Sampling and Sampling methods:
 - Sampling - Data reduction
 - Sampling - Modeling
 - Sampling – Balancing
 - Over-sampling / Under-sampling
- Feature selection and creation
- **Hands On Work through**



Statistical Modeling Methodologies

- Linear Regression
- Multi Linear Regression

Undirected Modeling (Unsupervised Modeling)

- Clustering and segmentation
- K-means clustering
- Case Study work Through



Directed Modeling Techniques (Supervised Modeling)

- Classification by Decision Tree
- Case Study and work Through
- Predictive Modeling by Logistic Regression
- Case Study and work Through

Quick Summary of Model Validation and Deployment Methods

- Validation Methods Overview
 - Prediction accuracy
 - Confusion matrix
 - ROC curve
 - Lift
 - KS test
- Deployment Methods
 - Model scoring
 - PMML
 - SQL/C based scoring engines



11. Illustrative Applications and Cases:

Discussion and White boarding to help bring out the similar set of business scenarios that call for the decision techniques that can be applied in the participant context:

- Customer Analytics
- Marketing Analytics
- Financial Analytics
- Automotive Analytics
- Retail Analytics