> Data Mining Overview with IBM SPSS Modeler

This two-day course introduces you to the major steps of the data mining process. The course goal is for you to be able to begin planning or evaluate your firm's current plan for data mining. Topics include: setting goals; framing research questions; organizing, cleaning and reducing data; selecting analysis methods; presenting and automating. The course also reviews the basic operations and environment of the IBM SPSS Modeler (formerly known as Clementine®) data mining software.

Following an overview of data mining and an introduction to essential terminology and technology, you will then proceed logically through the following topics:

INTRODUCTION

- ◆ Definition of Data Mining
- ◆ Data Mining and Statistics CRISP-DM
- ◆ Data Mining Examples

BUSINESS UNDERSTANDING (SETTING GOALS)

- ◆ Business Context
- ◆ Consider Constraints
- ◆ Track and Evaluate Results

DATA UNDERSTANDING AND PREPARATION

- ◆ Data Exploration
- ◆ Visualizing Data using Graphs
- ◆ Missing Values and Outliers
- ◆ Data Selection

CHECKING AND TRANSFORMING DATA

- ◆ Data Cleaning
- ◆ Data Normalization
- ◆ Creating New Variables Transformation
- ◆ Examples Recoding (Binning)
- ◆ Data Reduction

MODELLING OVERVIEW: INFERENTIAL DATA MINING TECHNIQUES

- ◆ Linear Regression
- ◆ Analysis of Variance
- ◆ Logistic and Multinomial Regression
- ◆ Time Series Analysis

MODELLING OVERVIEW: AUTOMATED DATA MINING TECHNIQUES

- ◆ Cluster Analysis
- ◆ Market Basket or Association Analysis
- ◆ Memory-Based Reasoning Decision Trees
- ◆ Neural Networks

EVALUATION OF MODELS

- ◆ Choosing among Methods
- ◆ Evaluating Accuracy
- ◆ Evaluating Improved Performance

DEPLOYMENT OF MODELS

- ◆ Deployment of Models
- ◆ Deployment of Results
- ◆ Interpreting Results of Models

Target Audience
Business analysts and others who have heard of data mining and want to know what it entails. Those who feel data mining might be beneficial to their company, but don't know where to begin.

Prerequisites
Attendees should be familiar with basic research methods. Training in advanced statistical methods is not necessary, but those with such training will gain a deeper understanding of the models.

Objectives
By the end of this course you will have learned how to identify potential data mining applications and use data exploration techniques to model your data.

Cost
HK$6,000.00 per person

Duration
2 days

> Introduction to IBM SPSS Modeler

This course reviews the basic operations and environment of the IBM SPSS Modeler (formerly known as Clementine®) data mining software. You will learn how to read, explore and manipulate data within IBM SPSS Modeler. Within the context of data mining, you will be introduced to such machine learning techniques as rule induction, neural networks, Kohonen networks and association rules, and will apply them to business data.

The main techniques outlined on the course are:

INTRODUCING IBM SPSS MODELER

- ◆ Introduction to IBM SPSS Modeler
- ◆ Starting IBM SPSS Modeler
- ◆ Visual Programming
- ◆ Building streams with IBM SPSS Modeler

READING IN DATA FILES

- ◆ Reading data files into IBM SPSS Modeler
- ◆ Reading data from free-field ASCII Files
- ◆ Reading data from free-fixed ASCII Files
- ◆ Reading data from SPSS Files
- ◆ Viewing data files within IBM SPSS Modeler
- ◆ Defining data field types

DATA QUALITY

- ◆ Missing data in IBM SPSS Modeler
- ◆ Data Audit node
- ◆ Data Distributions
- ◆ Examining the distribution of Symbolic fields
- ◆ Examining the distribution of Numeric fields

INTRODUCTION TO DATA MANIPULATION

- ◆ A brief introduction to the CLEM Language
- ◆ Record operations and the Select node
- ◆ Field operations and the Filter node
- ◆ Executing Derive nodes simultaneously

LOOKING FOR RELATIONSHIPS IN DATA

- ◆ Studying relationships between symbolic fields
- ◆ Pairs of symbolic fields and the Matrix node
- ◆ Pairs of numeric fields and correlation
- ◆ Viewing relationships between a numeric and symbolic field

MODELLING TECHNIQUES- OVERVIEW

- ◆ Neural networks
- ◆ Rule induction
- ◆ Binary Classifier
- ◆ K-Means Clustering
- ◆ Association rules
- ◆ Sequence Detection
- ◆ Which techniques, when?

Target Audience
New and potential users of IBM SPSS Modeler

Prerequisites
Basic proficiency in Windows operations. Training in statistical methods is not necessary, but those with such training will gain a deeper understanding of the models.

Objectives
By the end of the course, you will have learned:

- ◆ How to use many of the essential features of IBM SPSS Modeler
- ◆ How to deal with missing data
- ◆ How to carry out data exploration techniques using visualization and statistics
- ◆ How to interpret results effectively using both numerical and graphical output enabling appropriate conclusions to be drawn.

Cost
HK$6,000.00 per person

Duration
2 days

> Data Manipulation with IBM SPSS Modeler

This one-day course, designed for current users of IBM SPSS Modeler (formerly known as Clementine®), reviews a number of data manipulation techniques available within IBM SPSS Modeler. Attendees will see how to combine and manipulate files, sample data, handle missing values, and work with dates and sequence data.

COMBINING DATA FROM MULTIPLE SOURCES

- ◆ Using the Append node
- ◆ Join Concepts
- ◆ Using the Merge node
- ◆ SuperNodes
- ◆ Advantages of using Supernodes

SAMPLING DATA

- ◆ Using the Distinct node
- ◆ Using the Sample node
- ◆ Data Caching with IBM SPSS Modeler

WORKING WITH DATES

- ◆ Declaring date formats in IBM SPSS Modeler
- ◆ Formatting fields for date functions
- ◆ Calculations involving dates

DEALING WITH MISSING DATA

- ◆ Missing data in IBM SPSS Modeler
- ◆ Flagging blanks using the type node
- ◆ Assessing the quality of data
- ◆ Improving the quality of data
- ◆ Replacing blanks with valid information
- ◆ Auto Checking for blanks and out of bounds values
- ◆ Advice on handling

WORKING WITH SEQUENCE DATA

- ◆ CLEM sequence functions
- ◆ COUNT and STATE variants within the Derive node
- ◆ Restructuring sequence data using the History node

FILE MANIPULATION

- ◆ Using Aggregate Node
- ◆ Using Set to Flag Node
- ◆ Combining output

Target Audience
Users of IBM SPSS Modeler who want to perform data manipulations that go beyond the basics.

Prerequisites
Experience with IBM SPSS Modeler or completion of the Introduction to IBM SPSS Modeler training course.

Objectives
By the end of the course, you will have learned:

- ◆ Learn to combine & manipulate files
- ◆ Learn to sample data
- ◆ Handle Missing values
- ◆ Work with date & sequence data

Cost
HK$4,000.00 per day per person

Duration
1 day

Advanced Modeling with IBM SPSS Modeler

This two-day course, designed for current users of IBM SPSS Modeler (formerly known as Clementine®), reviews the expert options available in the IBM SPSS Modeler modeling dialogs (neural networks, rule induction, clustering and association rules, linear and logistic regression) and introduces advanced modeling methods, such as meta modeling.

PREPARING DATA FOR MODELING

- Cleansing Data
- Balancing Data
- Numeric Data Transformations
- Flag to Range Conversion

NEURAL NET FOR CLASSIFICATION

- Training Methods
- The Multi-Layer Perception
- The Radial Basis Function
- Expert Options
- Available Algorithms
- Which Method, When
- Sensitivity Analysis
- Prevention of Over Training

ADVANCED RULE INDUCTIONS

- Comparison C5.0 and C&R
- Gain Ratio Selection Criterion
- Pruning within C5.0 Minimum Records Per Branch
- Pruning and Minimum Records per Branch
- Using C&R Tree to Predict Numeric Fields

CLUSTERING

- K-Means Clustering
- Kohonen Network
- Kohonen Node Settings
- Two-step Clustering

ADVANCED ASSOCIATION RULES

- Apriori
- Choosing a Method and Expert Options
- Missing Data with Association Rules

REGRESSION

- Linear Regression & Assumptions
- Logistic regression & Assumptions

PRINCIPAL COMPONENTS

- Use of Principal Components
- What to look for
- Principles
- Number of Components Rotations, Component Scores, Sample Size, Methods

GETTING MOST FROM MODELS

- Modifying Confidence Values for Scoring
- Meta-Level Modeling
- Error Modeling

Target Audience

Users of IBM SPSS Modeler who want a deeper understanding of the models and modeling capabilities within IBM SPSS Modeler.

Prerequisites

Experience with IBM SPSS Modeler or completion of the Introduction to IBM SPSS Modeler training course.

Objectives

By the end of the course, you will have learned:

- Learn the underlying principles behind each of DM algorithms
- Use the optimal modeling techniques for achieving organizational objectives
- Learn to manipulate the expert options to refine the model for greater precision.

Duration

HK$8,000.00 per person

2 days

TO: **SBAS (Hong Kong) Ltd.**
FAX: **+(852) 2811 3042**

<div align="center">

IBM SPSS Modeler Training Course
Enrollment Form
</div>

Please register the following person(s) for the SPSS Program:

* Please Complete The Form In Block Letters*

Company: _____   **Department:** _____

Name (Mr./Mrs./Ms.): _____   **Job Title:** _____

Tel.: _____   **Fax:** _____   **Email:** _____

Address: _____

             _____

I would like to enroll the following courses:

* Please tick the boxes *

☐   **Data Mining Overview with IBM SPSS Modeler**

☐   **Introduction to IBM SPSS Modeler**

☐   **Data Manipulation with IBM SPSS Modeler**

☐   **Advanced Modeling with IBM SPSS Modeler**

Venue: **SBAS (Hong Kong) Ltd.**

             Rm.4, 18/F, Westlands Centre, 20 Westlands Road, Quarry Bay,

             HK. (or on-site - negotiable)

Please fill in the attached enrollment form and fax back at +(852) 2811 3042 for

reservation. Application is accepted on first-come-first served basis.

●   **Training Materials are included in the enrollment fee**

For more information and other courses details, please contact Mr. Adam S. Bux

at +(852) 2811 9662 or email at abux@sbas.com.hk.