

---

# **Data Mining**

## **SPSS Clementine 12.0**

### **1. Clementine Overview**

**Spring 2010**  
Instructor: Dr. Masoud Yaghini

# Outline

---

- **Introduction**
- **Types of Models**
- **Clementine Interface**
- **Projects**
- **References**

---

# Introduction

# Introduction

---

- Three of the common data mining tools
  - SPSS Clementine
  - SAS-Enterprise Miner
  - STATISTICA Data Miner
- The most popular and the oldest data mining tool package on the market today is **SPSS Clementine**

# Introduction

---

- **SPSS Clementine**
  - the most mature among the major data mining packages on the market today.
- Since 1993, many thousands of data miners have used Clementine to create very powerful models for business.
- It was the first data mining package to use the **graphical programming** user interface.
- It enables you to quickly develop data mining models and deploy them in business processes to improve decision making.

# Introduction

---

- The Clementine package was integrated with **CRISP-DM** to help guide the modeling process flow.
- Clementine supports the entire data mining process.

# Clementine Documentation

---

- **Clementine User's Guide**
  - General introduction to using Clementine.
- **Clementine Source, Process, and Output Nodes**
  - Descriptions of all the nodes used to read, process, and output data in different formats.
- **Clementine Modeling Nodes**
  - Descriptions a variety of modeling methods in Clementine
- **Clementine Applications Guide**
  - The examples in this guide provide brief, targeted introductions to specific modeling methods and techniques.
- **Clementine Algorithms Guide**
  - Descriptions of the mathematical foundations of the modeling methods used in Clementine.
- **CRISP-DM 1.0 Guide**
  - Step-by-step guide to data mining using the CRISP-DM methodology.

# Application Examples

---

- While the data mining tools in Clementine can help solve a wide variety of business and organizational problems, the application examples provide brief, targeted introductions to specific modeling methods and techniques.
- You can access the examples by choosing **Application Examples** from the **Help** menu in Clementine Client.
- The data files and sample streams are installed in the **Demos folder** under the product installation directory.



# Demos Folder

---

- The data files and sample streams used with the application examples are installed in the **Demos** folder under the product installation directory.
- This folder can also be accessed from the Clementine 12.0 program group under SPSS Inc on the Windows Start menu.

# Changing the Temp Directory

---

- Some operations performed by Clementine may require temporary files to be created.
- By default, Clementine uses the system temporary directory to create temp files.

# Changing the Temp Directory

---

- To alter the location of the temporary directory using the following steps:
  - Create a new directory called clem and subdirectory called *servertemp*.
  - Edit *options.cfg*, located in the */config* directory of your Clementine installation directory.
  - Edit the *temp\_directory* parameter in this file to read:  
*temp\_directory, "C:/clem/servertemp"*.
  - After doing this, you must restart the Clementine Server service. You can do this by clicking the Services tab on your Windows Control Panel. Just stop the service and then start it to activate the changes you made. Restarting the machine will also restart the service.
  - All temp files will now be written to this new directory.

---

# **Types of Models**

# Types of Models

---

- In Clementine 12.0, models are packaged into following modules:
  - **Classification** module
    - ◆ The Classification module helps organizations for prediction
  - **Association** module
    - ◆ Association models associate a particular conclusion (such as the decision to buy something) with a set of conditions.
  - **Segmentation** module
    - ◆ The Segmentation module is recommended in cases where the specific result is unknown

# Classification Module

---

- Classification module is included:
  - **Decision Trees**
    - ◆ C&R Tree
    - ◆ QUEST
    - ◆ CHAID
    - ◆ C5.0
  - **Decision List**
  - **Neural Networks**
  - **Regression**
    - ◆ Linear regression
    - ◆ Logistic regression
  - **Bayesian Network**
  - **Support Vector Machine (SVM)**

# Association Module

---

- Association module is included:
  - **Generalized Rule Induction (GRI)**
  - **Apriori model**
  - **CARMA model**
  - **Sequence model**

# Segmentation Module

---

- Clustering models focus on identifying groups of similar records and labeling the records according to the group to which they belong.
- The following models are included:
  - **K-Means**
  - **Kohonen**
  - **TwoStep**
  - **Anomaly Detection**



# Clementine Modules

---

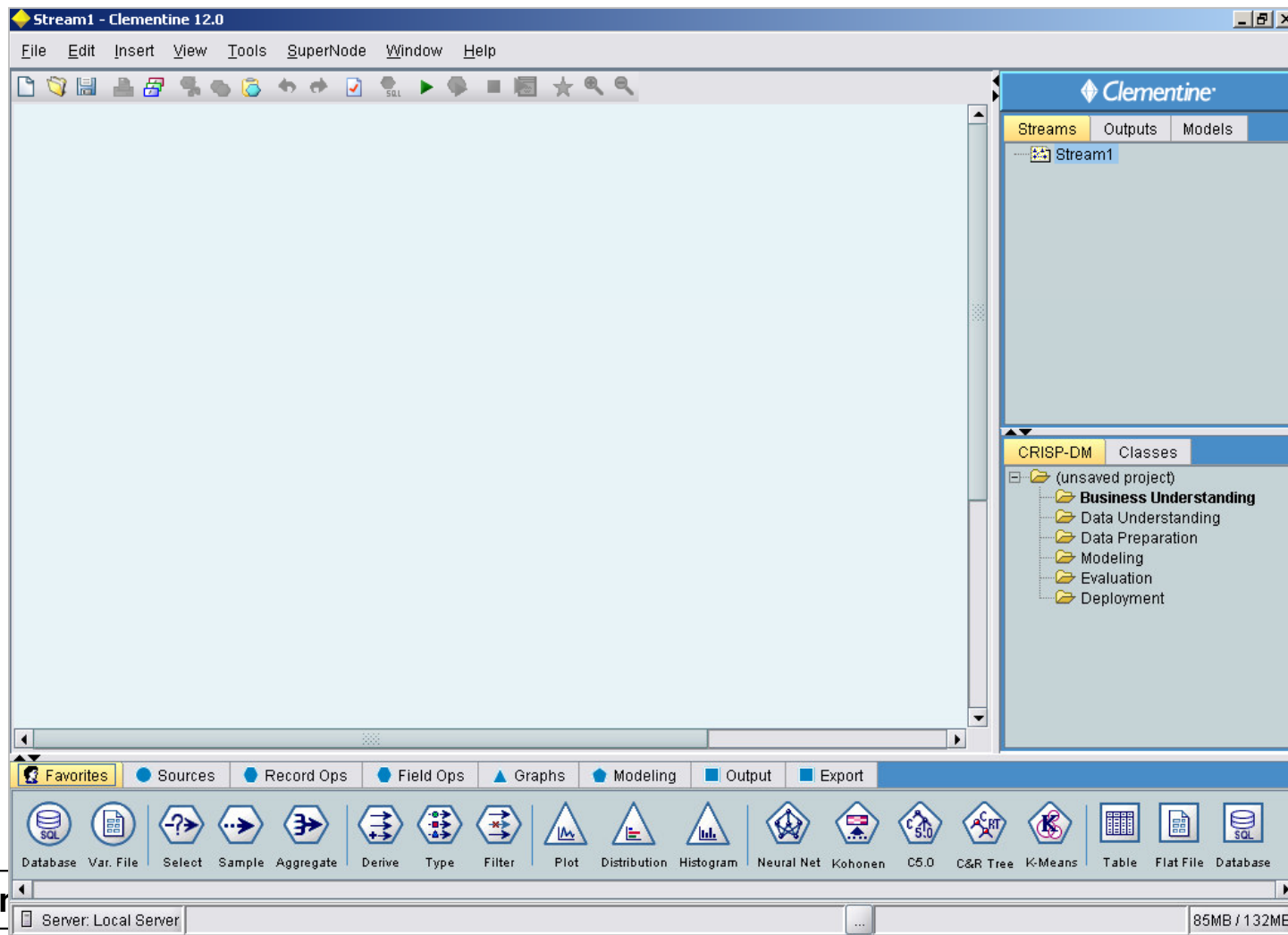
- The following are optional add-on modules for Clementine:
  - **Text Mining option** for mining unstructured data
  - **Web Mining option**
  - **Database Modeling** for integration with other data mining packages

---

# Clementine Interface

# Starting Clementine

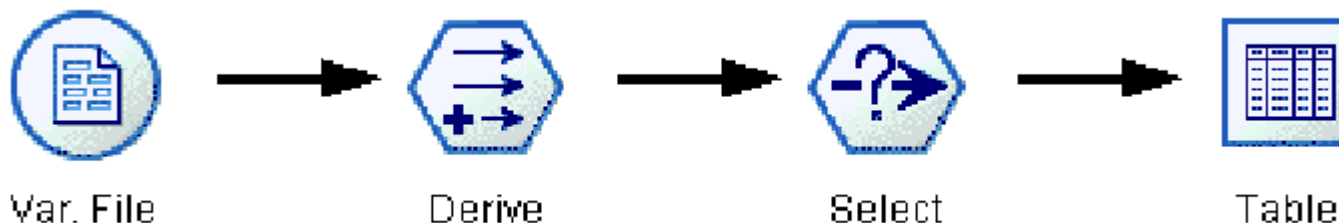
- To start the application, choose Clementine 12.0 from the SPSS Inc program group on the Windows Start menu.



Clemer

# Data Stream

- The Clementine interface employs an intuitive **visual programming** interface to permit you to draw logical data flows the way you think of them.
- Working with Clementine is a three-step process of working with data.
  - First, you read data into Clementine,
  - Then, run the data through a series of manipulations,
  - And finally, send the data to a destination.
- This sequence of operations is known as a **data stream**



# Stream Canvas

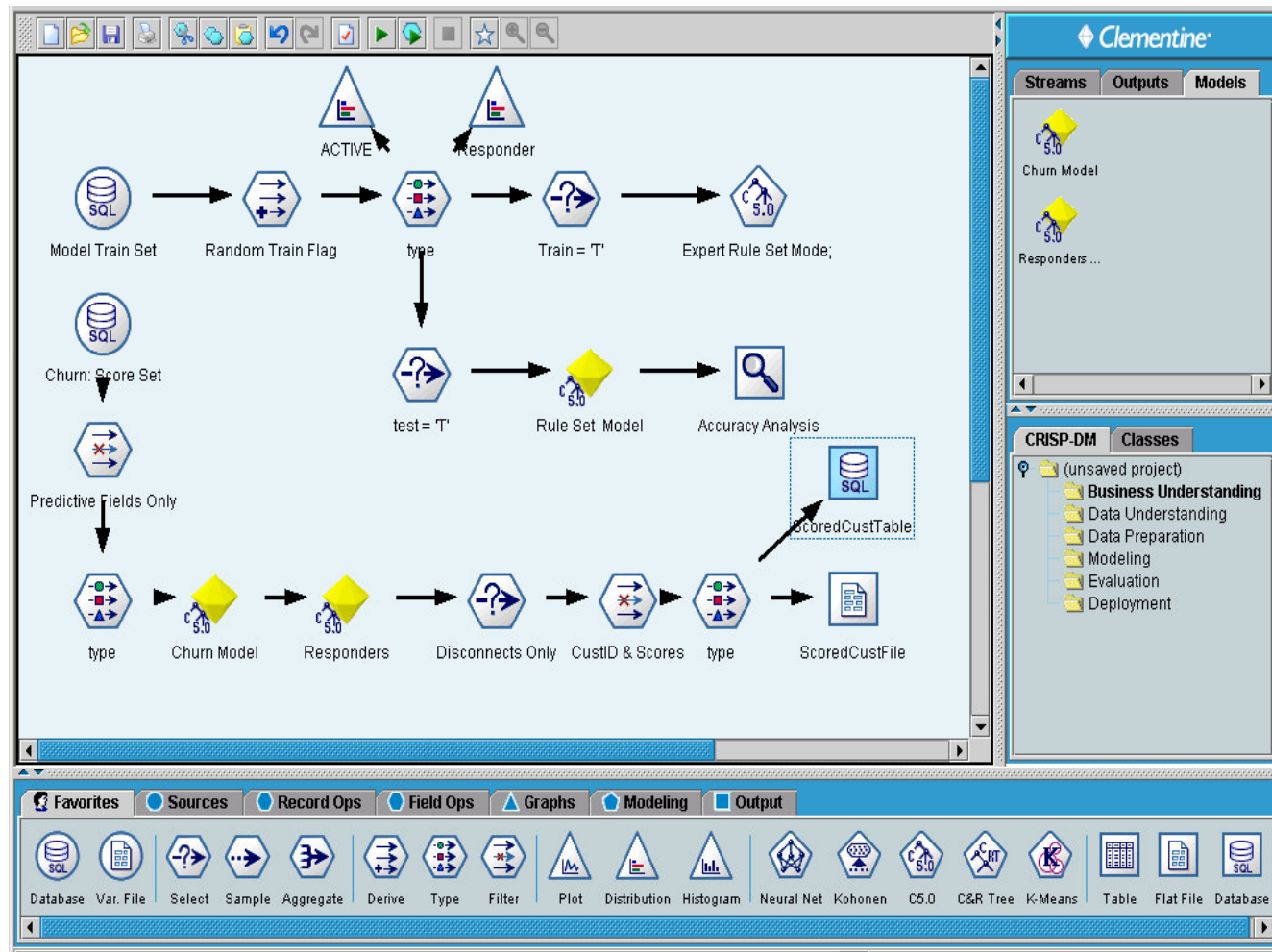
---

- **Stream Canvas**

- the largest area of the Clementine window and is where you will build and manipulate data streams.
- You can work on multiple streams in the same canvas or multiple stream files
- Streams are created by drawing diagrams of data operations on the main canvas in the interface.

# Clementine Interface

- Example of canvas:



# Clementine Interface

---

- **Node**

- Each operation is represented by an icon or **node**
- The nodes are linked together in a stream representing the flow of data through each operation.

- **Streams manager**

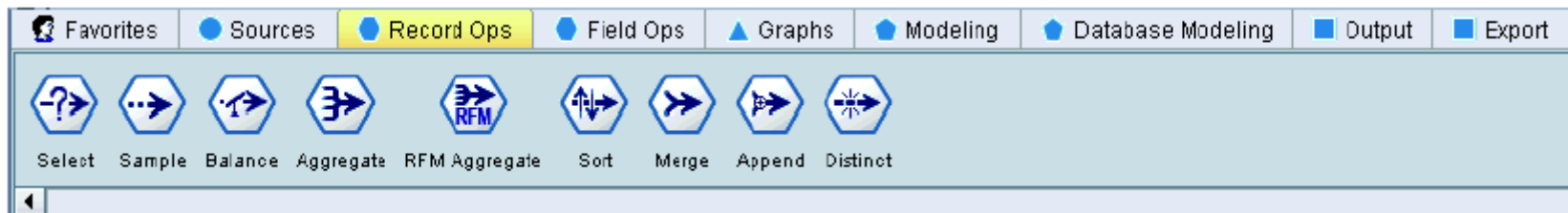
- streams are stored in the **Streams manager**, at the upper right of the Clementine window.

# Nodes Palette

- **Nodes Palettes**

- The palettes are groups of processing nodes listed across the bottom of the screen.
- The palettes include Favorites, Sources, Record Ops (operations), Field Ops, Graphs, Modeling, Output, and Export. When you click on one of the palette names, the list of nodes in that palette is displayed below.

- the **Record Ops** palette tab.



- To add nodes to the canvas, **double-click icons** from the Nodes Palette or **drag and drop** them onto the canvas.



# Nodes Palette Tabs

---

- **Nodes Palette Tabs**

- **Sources**

- ◆ Nodes bring data into Clementine.

- **Record Ops**

- ◆ Nodes perform operations on data records, such as selecting, merging, and appending.

- **Field Ops**

- ◆ Nodes perform operations on data fields, such as filtering, deriving new fields, and determining the data type for given fields.

- **Graphs**

- ◆ Nodes graphically display data before and after modeling. Graphs include plots, histograms, web nodes, and evaluation charts.

# Nodes Palette Tabs

---

- **Nodes Palette Tabs**

- **Modeling**

- ◆ Nodes use the modeling algorithms available in Clementine, such as neural nets, decision trees, clustering algorithms, and data sequencing.

- **Output**

- ◆ Nodes produce a variety of output for data, charts, and model results, which can be viewed in Clementine or sent directly to another application, such as SPSS or Excel.

# Clementine Managers

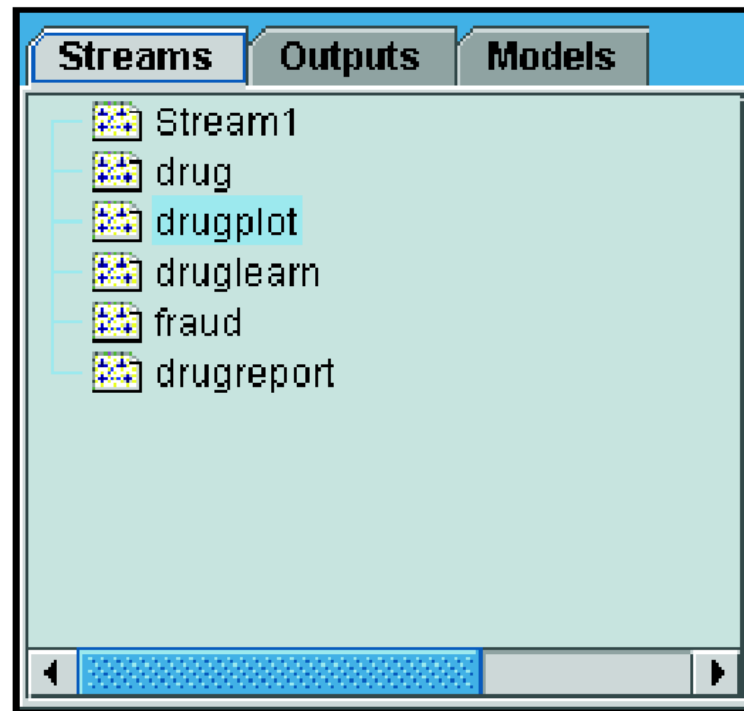
---

- The upper right window holds three managers:
  - Streams
  - Outputs
  - Models
- You can switch the display of the manager window to show any of these items.

# Clementine Managers

- **Streams tab**

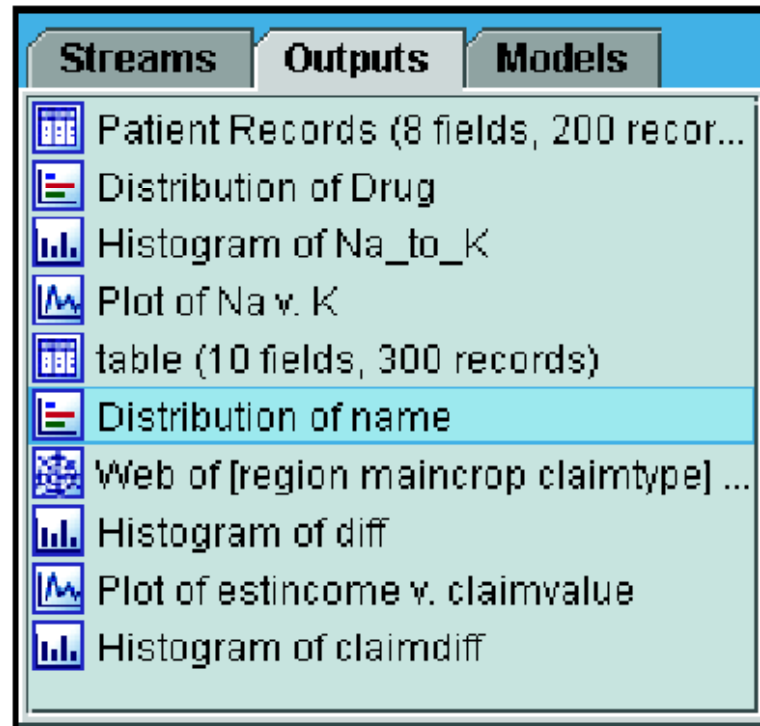
- The **Streams tab** will display the active streams, which you can choose to work on in a given session.
- You can use the **Streams tab** to open, rename, save, and delete the streams created in a session.



# Clementine Managers

- **Outputs tab**

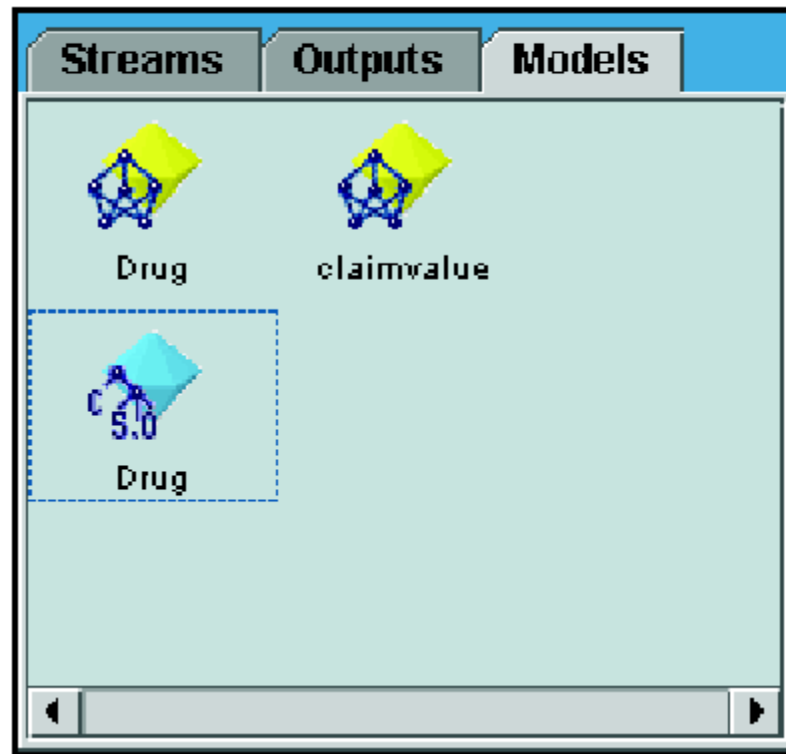
- The **Output tab** will show all graphs and tables output during a session.
- You can display, save, rename, and close the tables, graphs, and reports listed on this tab.



# Clementine Managers

- **Models tab**

- The **Models tab** will show all the trained models you have created in the session.
- These models can be browsed directly from the **Models tab** or added to the stream in the canvas.



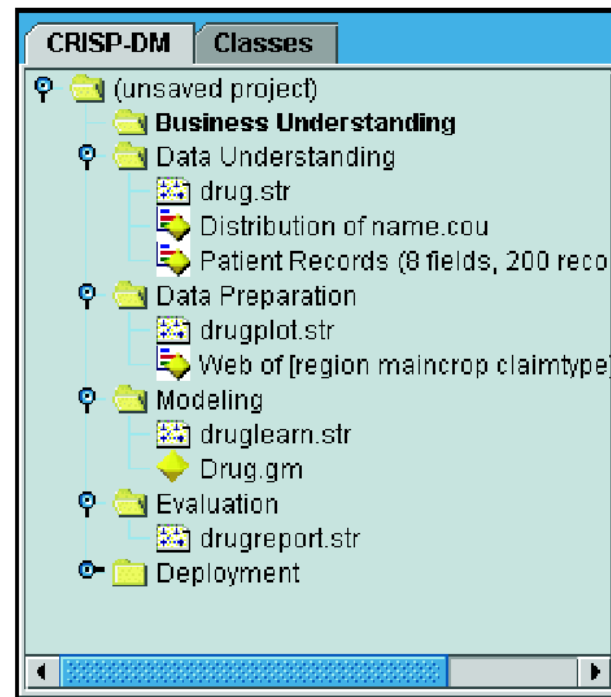
# Clementine Projects

---

- On the lower right side of the window is the **projects tool**, used to create and manage data mining projects.
- There are two ways to view projects you create in Clementine:
  - The CRISP-DM view
  - The Classes view

# Clementine Projects

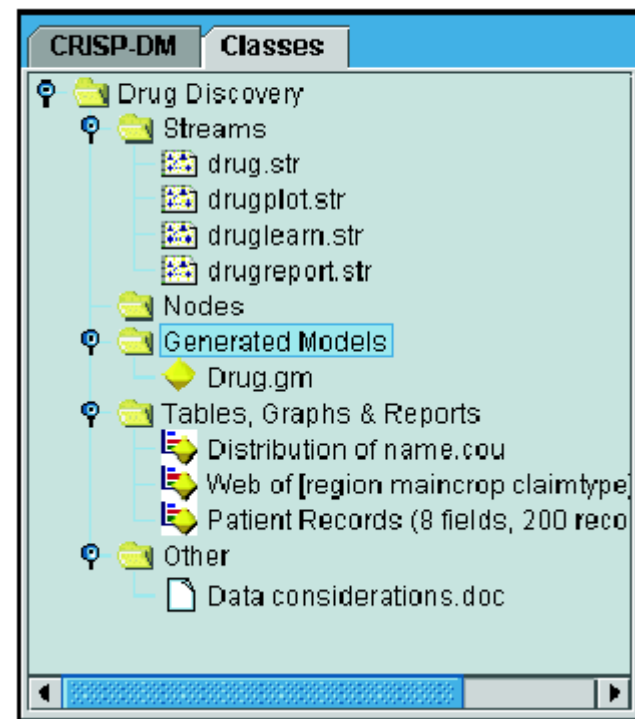
- The **CRISP-DM tab** provides a way to organize projects according to the CRISP-DM methodology.
- For both experienced and first-time data miners, using the CRISP-DM tool will help you to better organize and communicate your efforts.





# Clementine Projects

- The **Classes** tab provides a way to organize your work in Clementine categorically—by the types of objects you create.
- This view is useful when taking inventory of data, streams, and models.



# Clementine Toolbars

---



Create new stream



Save stream



Open Clementine Application  
Templates (CATs)



Copy to clipboard



Undo last action



Edit stream properties



Execute stream selection



Add SuperNode



Zoom out (SuperNodes only)



Open stream



Print current stream



Cut & move to clipboard



Paste selection



Redo



Execute current stream



Stop stream (Active only during stream  
execution)



Zoom in (SuperNodes only)

# Using Shortcut Keys

---

Shortcut Key	Function
Ctrl-A	Select all
Ctrl-X	Cut
Ctrl-N	New stream
Ctrl-O	Open stream
Ctrl-P	Print
Ctrl-C	Copy
Ctrl-V	Paste
Ctrl-Z	Undo
Ctrl-Q	Select all nodes downstream of the selected node
Ctrl-W	Deselect all downstream nodes (toggles with Ctrl-Q)
Ctrl-E	Execute from selected node
Ctrl-S	Save current stream
Alt-Arrow keys	Move selected nodes on the stream canvas in the direction of the arrow used
Shift-F10	Open the context menu for the selected node

# Automating Clementine

---

- Since advanced data mining can be a complex and sometimes lengthy process, Clementine includes several types of coding and automation support. They are:
  - **Clementine Language for Expression Manipulation (CLEM)**
  - **Scripting**
  - **Batch mode**

# Automating Clementine

---

- **Clementine Language for Expression Manipulation (CLEM)**
  - is a language for analyzing and manipulating the data that flows along Clementine streams.
  - Data miners use CLEM extensively in stream operations to perform tasks as simple as deriving profit from cost and revenue data or as complex as transforming Web-log data into a set of fields and records with usable information.
  - For more information, see What Is CLEM? in Chapter 7 in Clementine® 12.0 User's Guide.

# Automating Clementine

---

- **Scripting**

- is a powerful tool for automating processes in the user interface and working with objects in batch mode.
- Scripts can perform the same kinds of actions that users perform with a mouse or a keyboard.
- You can set options for nodes and perform derivations using a subset of CLEM.
- You can also specify output and manipulate generated models.
- For more information, see Scripting Overview in Chapter 2 in Clementine® 12.0 Scripting and Automation Guide.

# Automating Clementine

---

- **Batch mode**

- enables you to use Clementine in a non-interactive manner by running Clementine with no visible user interface.
- Using scripts, you can specify stream and node operations as well as modeling parameters and deployment options.
- For more information, see Introduction to **Batch Mode** in Chapter 7 in Clementine® 12.0 Scripting and Automation Guide.



# **Building a Project**



# Building a Project

---

- A **project** is a group of files related to a data mining task.
- Projects include
  - data streams,
  - graphs,
  - generated models,
  - reports, and
  - anything else that you have created in Clementine.
- Note:
  - If the **Projects tool** is not visible in the Clementine window, choose **Project** from the **View** menu.

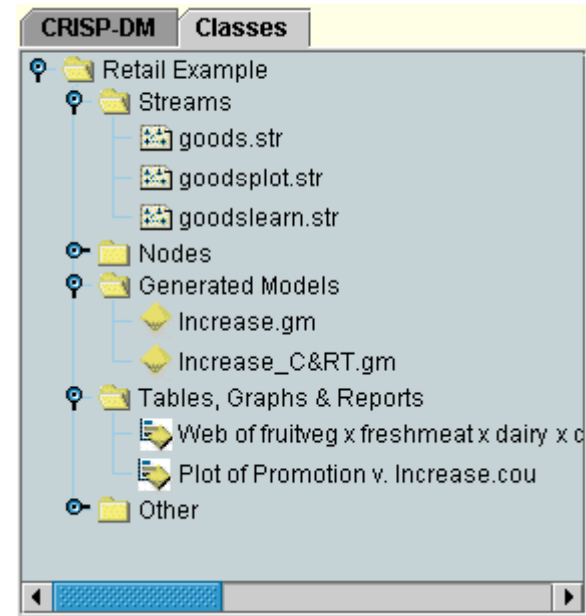
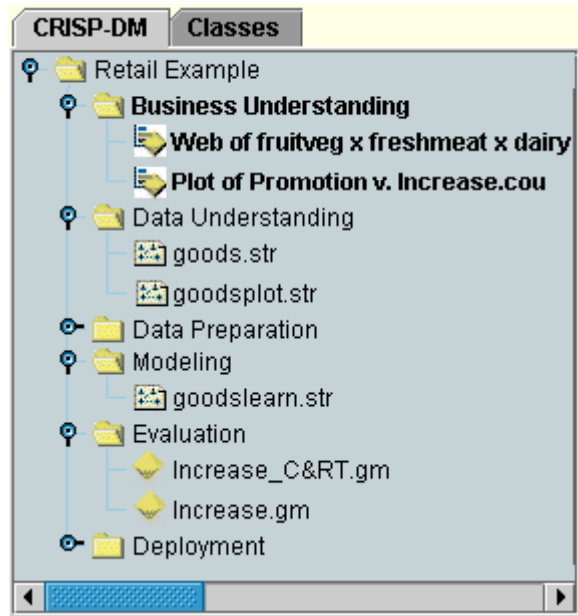
# Building a Project

---

- Using projects, you can:
  - Annotate each object in the project file.
  - Use the CRISP-DM methodology to guide your data mining efforts.
  - Add non-Clementine objects to the project, such as a PowerPoint slide show used to present your data mining goals or white papers on the algorithms that you plan to use.
  - Produce both comprehensive and simple update reports based on your annotations.

# Building a Project

- Objects that you add to a project can be viewed in two ways: **Classes view** and **CRISP-DM view**.



# Building a Project

---

- **Setting the Default Project Phase**

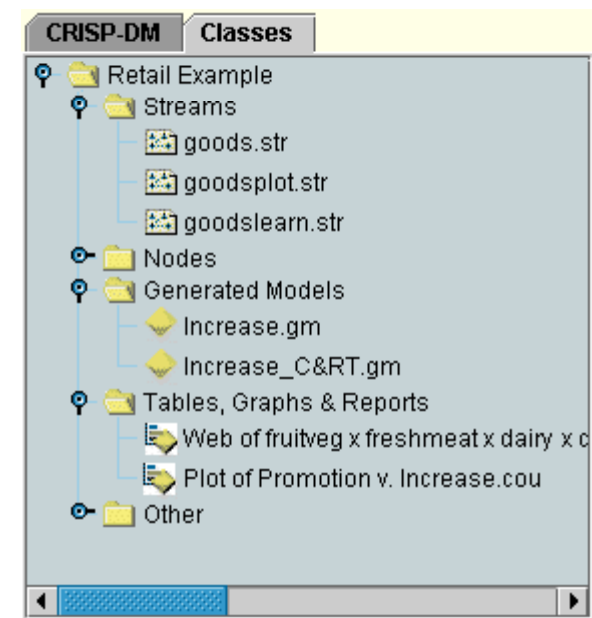
- Objects added to a project are added to a default phase of CRISP-DM.
- This means that you need to organize objects manually according to the data mining phase in which you used them.
- It is wise to set the default folder to the phase in which you are currently working.

- **To select which phase to use as your default:**

- In CRISP-DM view, right-click the folder for the phase to set as the default.
- From the menu, choose **Set as Default**.
- The default folder is displayed in bold type.

# Building a Project

- The **Classes view** in the projects tool organizes your work in Clementine categorically by the types of objects created.
- Saved objects can be added to any of the following categories:
  - Streams
  - Nodes
  - Models
  - Tables, graphs, reports
  - Other (non-Clementine files, such as slide shows or white papers relevant to your data mining work)



# Building a Project

---

- A project is essentially a file containing references to all of the files that you associate with the project.
- This means that project items are saved both individually and as a reference in the project file (.cpj).
- Because of this referential structure, note the following:
  - Project items must first be saved individually before being added to a project.
  - Objects that are updated individually, such as streams, are also updated in the project file.
  - Manually moving or deleting objects (such as streams, nodes, and output objects) from the file system will make links in the project file invalid.

# Creating a New Project

---

- **Creating a New Project**
  - You can either start building one, if none is open, or you can close an existing project and start from scratch.
- From the stream canvas menus, choose:
  - **File > Project > New Project...**

# Building a Project

---

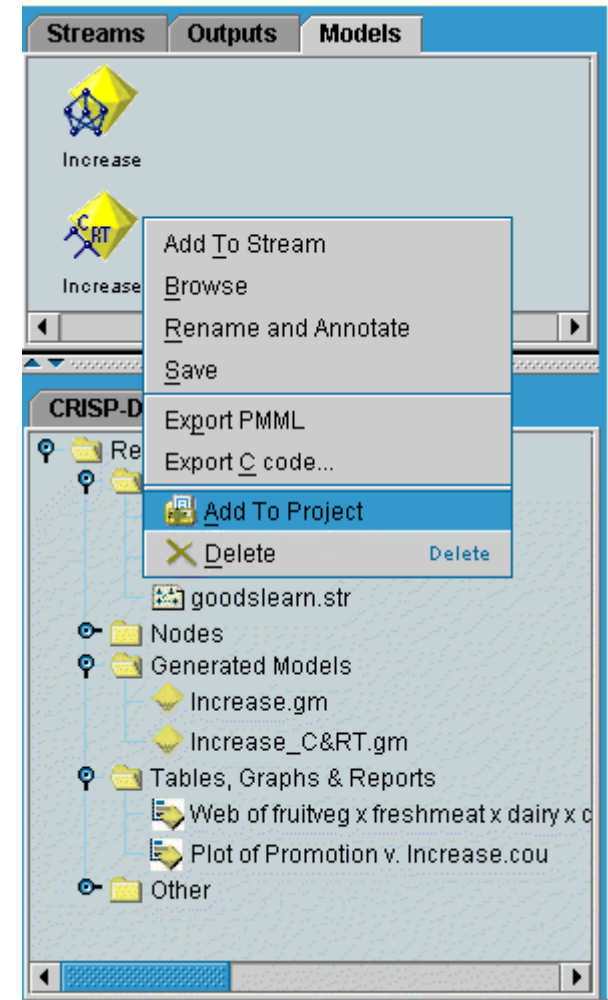
- **Adding to a Project**

- Once you have created or opened a project, you can add objects, such as data streams, nodes, and reports, using several methods.



# Adding Objects from the Managers

- **Adding Objects from the Managers**
  - Using the managers in the upper right corner of the Clementine window, you can add streams or output.
- **To Add Objects from the Managers**
  - Select an object, such as a table or a stream, from one of the managers tabs.
  - Right-click and choose **Add to Project**.
  - If the object has been previously saved, it will automatically be added to the appropriate objects folder (in Classes view) or to the default phase folder (in CRISP-DM view).
  - Alternatively, you can drag and drop objects from the managers to the project workspace.



# Adding Nodes from the Canvas

---

- **Adding Nodes from the Canvas**
  - You can add individual nodes from the stream canvas by using the Save dialog box.
- **To Add Nodes from the Canvas**
  - Select a node on the canvas.
  - Right-click and choose **Save Node**. Alternatively, from the menus choose: **Edit > Node > Save Node...**
  - In the Save dialog box, select **Add file to project**.
  - Create a name for the node and click **Save**.
- This saves the file and adds it to the project. Nodes are added to the Nodes folder in Classes view and to the default phase folder in CRISP-DM view.

# Adding External Files

---

- **Adding External Files**

- You can add a wide variety of non-Clementine objects to a project.
- This is useful when you are managing the entire data mining process within Clementine.
- For example, you can store links to data, notes, presentations, and graphics in a project.
- In CRISP-DM view, external files can be added to the folder of your choice.
- In Classes view, external files can be saved only to the Other folder.

# Adding External Files

---

- **To add external files to a project:**
  - Drag files from the desktop to the project. or
  - Right-click the target folder in CRISP-DM or Classes view.
  - From the menu, choose **Add to Folder**.
  - Select a file in the dialog box and click **Open**.
- This will add a reference to the selected object inside Clementine projects.

# Setting Project Properties

---

- **Setting Project Properties**
  - You can customize a project's contents and documentation by using the project properties dialog box.
- **To access project properties:**
  - Right-click an object or folder in the projects tool and choose Project Properties.
  - Click the Project tab to specify basic project information.

# Building a Project

- Setting project properties

The screenshot shows the 'IUST\_Project' dialog box. It has a title bar with a yellow diamond icon and the text 'IUST\_Project'. Below the title bar is a search bar with a question mark icon. The main area contains the following fields and table:

Created: December 22, 2009 7:00:10 AM IRST

Summary: [Empty text box]

Contents:

Project objects	Counts
Streams	1
Nodes	0
Supernodes	0
Generated models	0
Tables, graphs and reports	0
Other	0
TOTAL	1

Save unsaved objects as:

☒ files on the local file system

☐ objects in the Predictive Enterprise Repository

☒ Update object references when loading project

Project Report Annotations

OK Cancel Apply Reset

# Folder Properties and Annotations

---

- **Folder Properties and Annotations**

- Individual project folders (in both CRISP-DM and Classes view) can be annotated.
- In CRISP-DM view, this can be an extremely effective way to document your organization's goals for each phase of data mining.
- For example, using the annotation tool for the Business Understanding folder, you can include documentation such as "The business objective for this study is to reduce churn among high-value customers."
- This text could then be automatically included in the project report by selecting the Include in report option.

- **To annotate a folder:**

- Select a folder in the projects tool.
- Right-click the folder and choose Folder Properties.

# Object Properties

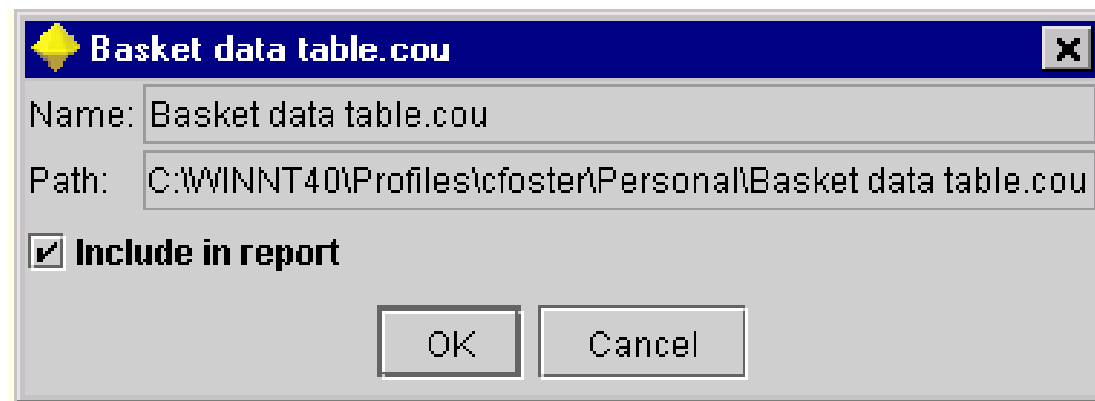
---

- **Object Properties**

- You can view object properties and choose whether to include individual objects in the project report.

- **To access object properties:**

- Right-click an object in the project window.
- From the menu, choose **Object Properties**.





# Closing a Project

---

- **Closing a Project**

- When you exit Clementine or open a new project, the existing project is closed, including all associated files.
- Alternatively, you can choose to close the project file itself and leave all associated files open.

- **To close a project file:**

- From the **File menu**, choose **Close Project**.
- If you are prompted to close or leave open all files associated with the project, click **Leave Open** to close the project file (.cpj) itself but to leave open all associated files, such as streams, nodes, or graphs.

---

# References

# References

---

- Integral Solutions Limited., **Clementine® 12.0 User's Guide**, 2007.



The end