



IBM Cognos Forum

Ignite knowledge, ideas, connections

Predictive Analytics

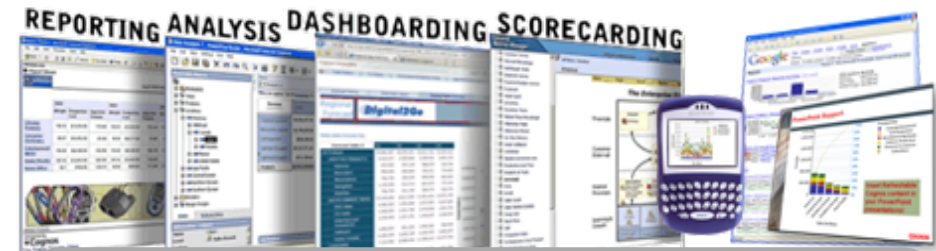
Information Management

Cognos.
software

Agenda

- > Understand a bit about Data Mining/ Predictive Analytics
- > Understand how IBM Cognos software and IBM InfoSphere Warehouse work together to answer your needs
- > Use customer case studies to illustrate how
- > Answer your questions

Cognos Synergy with InfoSphere Warehouse



Cognos.
software

Data Mining

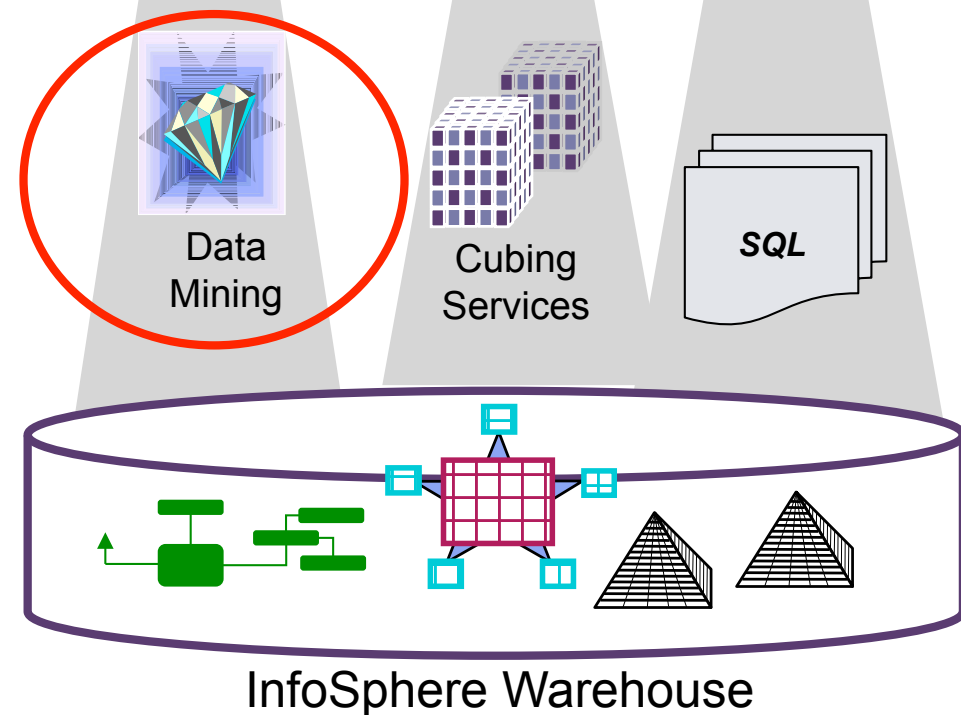
- Batch or Dynamic process integration and visualization
- Advanced DB2 features for data mining and unstructured (text) analytics

Cubing Services

- Open cube access through XMLA

Query Optimization

- Cognos Optimization
- Query optimization
- Performance tuning for AIX



IBM Cognos 8 Integration with IBM InfoSphere Warehouse

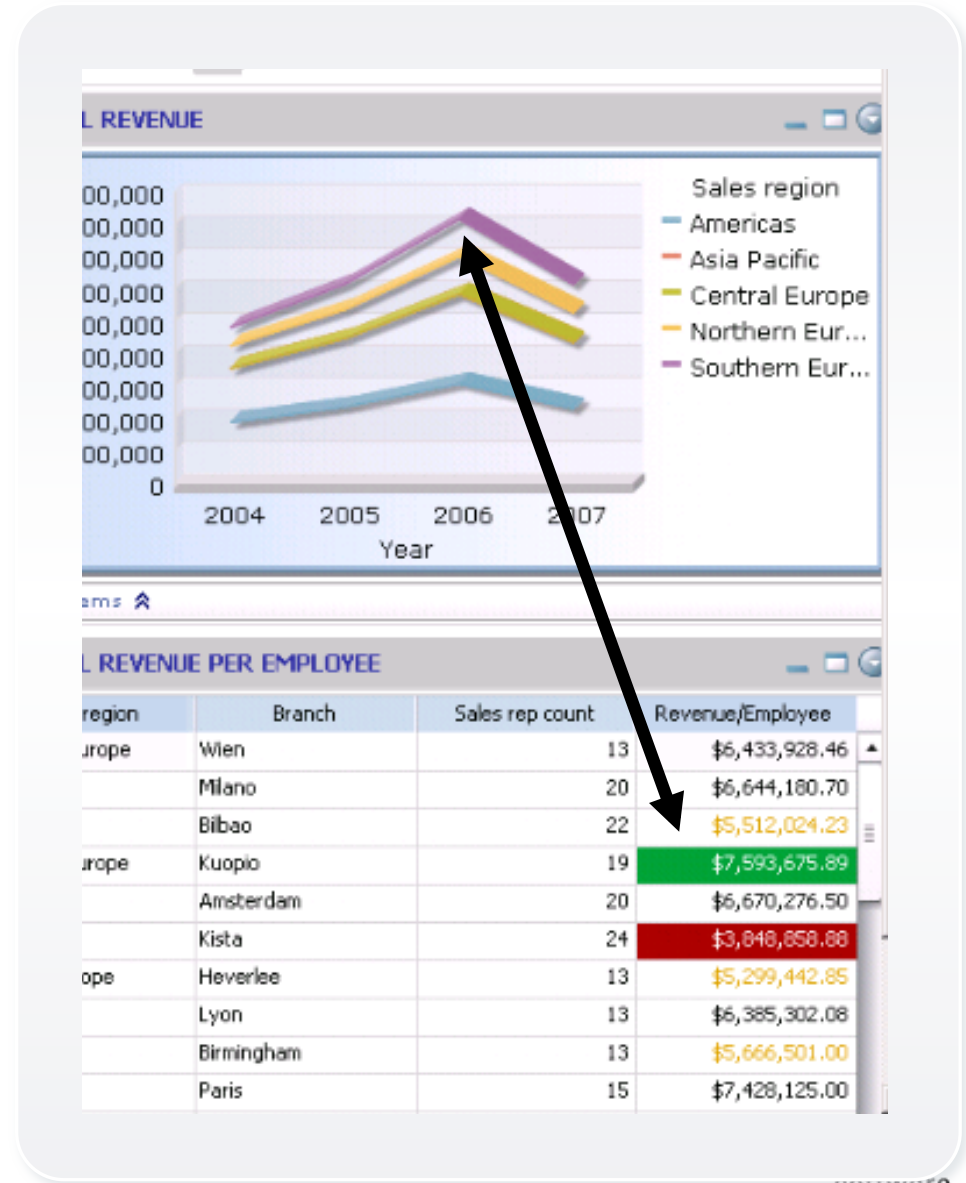
Challenge: Enable more sophisticated and predictive BI applications

Solution

- > Deliver data mining results to a broader audience of users with IBM Cognos 8 BI

Benefits

- > Integration and leverage of predictive models within IBM Cognos software provides necessary answers



The Four Styles of Analysis with IBM Cognos 8

Broad Usage
(Consumers)

Focused Usage
(Specialist)



Analytical Reporting
Drill

- Top down view
- Drillable reports
- Sort top & bottom
- Review then query
- Market shifts
- Product ranking

Trending
Slice and Dice

- Personal exploration
- Compare & contrast
- Rotate and nest
- Work disconnected
- Sales trend analysis
- Market analysis

Scenario Modeling
What-if

- Model scenarios
- Reorganize, reshape
- Compare scenarios
- Save versions
- Financial analysis
- Profitability analysis

Predictive Analytics
What might be

- Uncover patterns
- Statistical algorithms
- Mine data and text
- Predict outcomes
- Fraud prevention
- Churn analysis



Casual Users



Executive



BUSINESS MANAGER



FINANCIAL & BUSINESS ANALYST



Statistical Analyst

The most important ingredient for meaningful predictive analytics:

~~University Degree in Statistics~~

~~Computer Science Education~~

Business Understanding

Two Types of Data Mining – Discovery & Predictive

Discovery

- Automatically find trends and patterns
- Answer unasked questions
- Relatively undirected analysis
- Tool reports on findings
- In a word – “Easier”
- Useful for non-statisticians

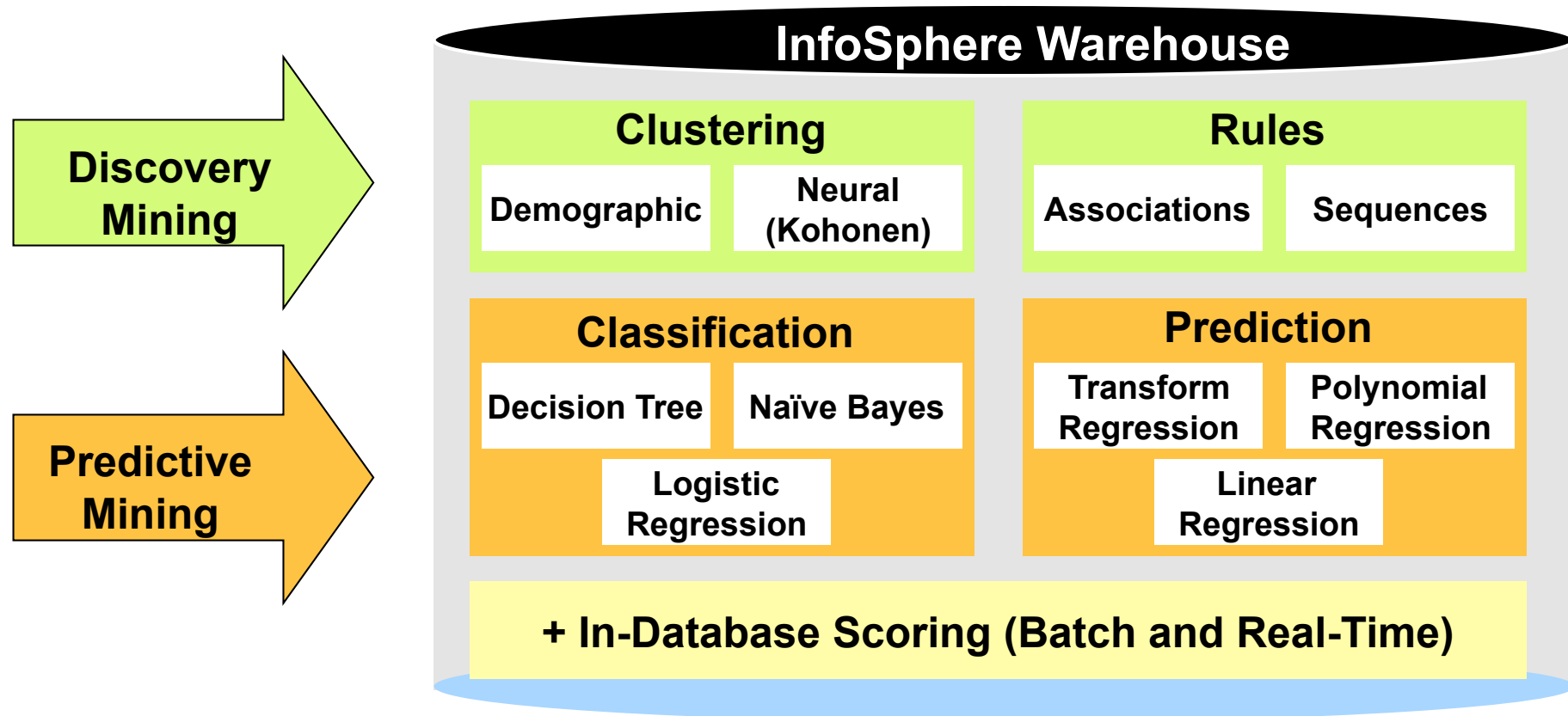


Predictive

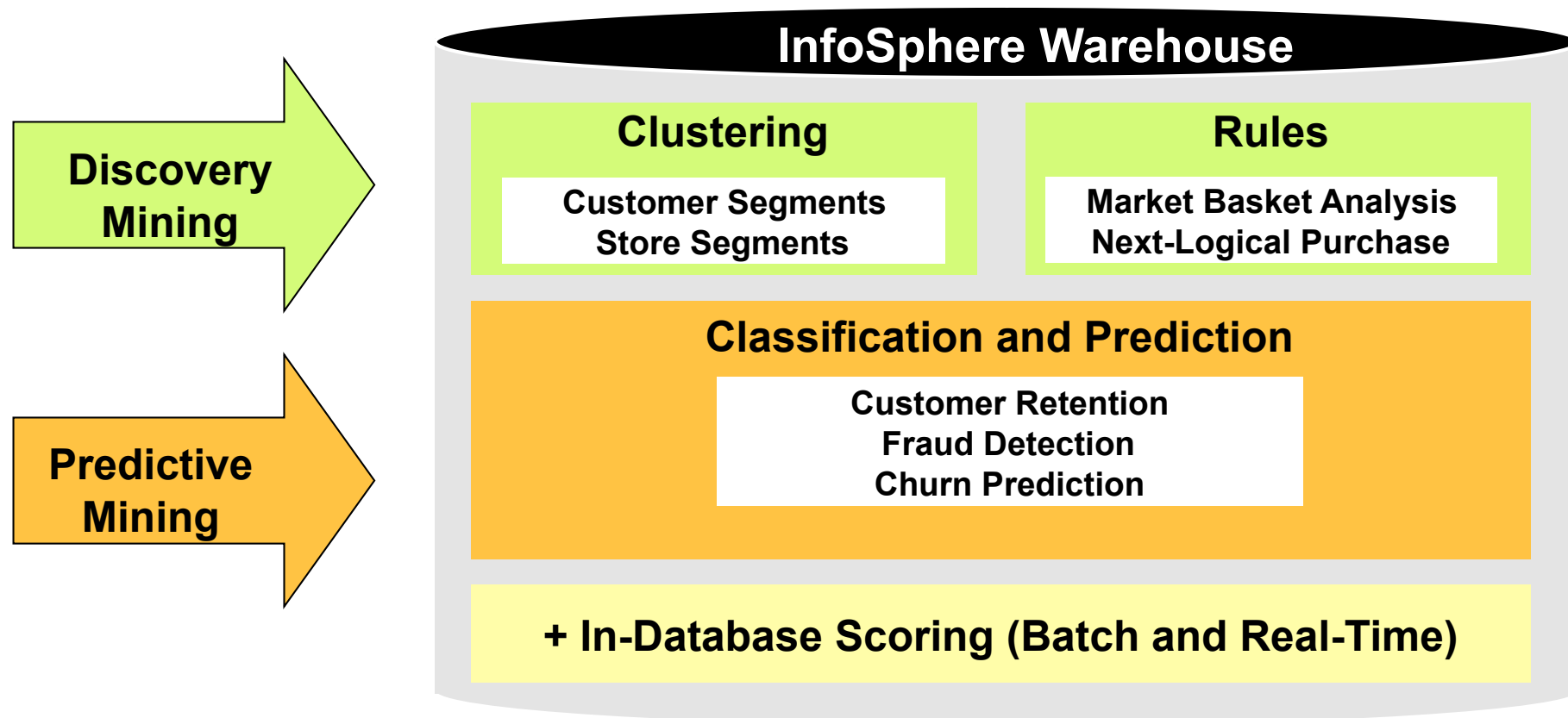
- Specific question
- Probability associated with outcomes
- Directed analysis
- Iterative process
 - Train, Test, Apply
- Apply model in database at customer touch points



InfoSphere Data Mining Methods and Algorithms

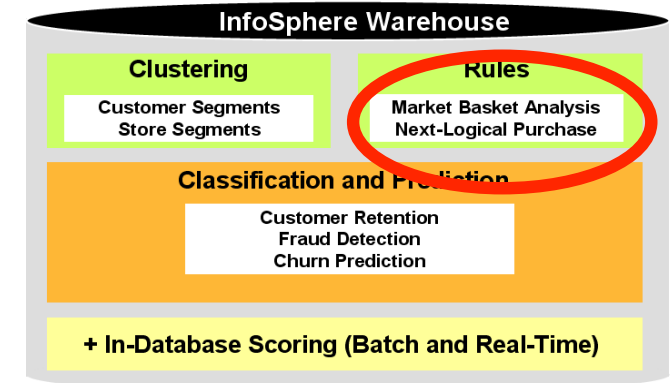


InfoSphere Data Mining Application Examples



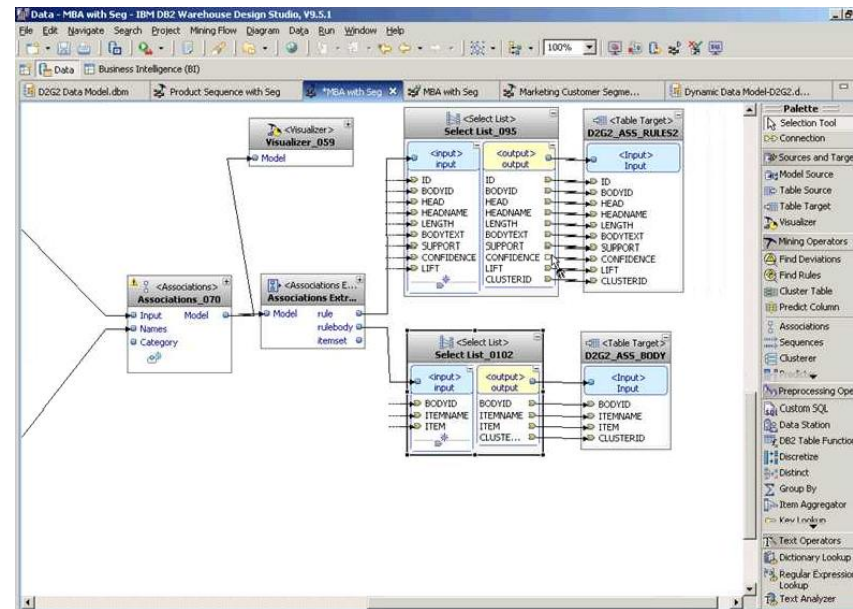
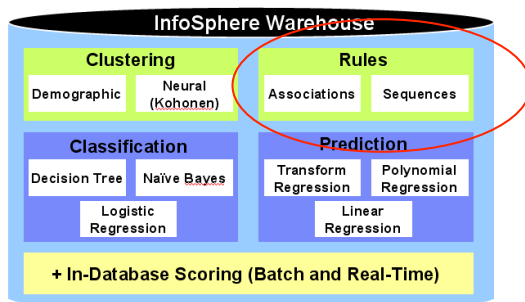
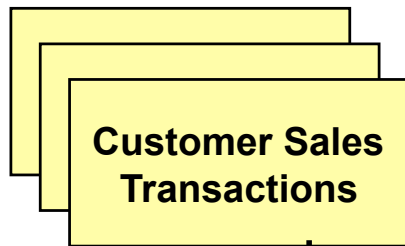
Market Basket Analysis and Next Logical Purchase

- > Customer Background
 - Electronics Retailer
- > Business Problem
 - Need to better understand what products are sold together and
 - predict the next, most-logical product a customer will purchase
- > Business Goals
 - Develop a data mining flow to use advanced analytics to show driver and implied items within a Market Basket analysis
 - Develop a platform for delivering the “next purchase” prediction results in support of marketing
- > Results
 - Created a set of tables with the results of data mining to allow more team members within the retailer to take proper action
 - Report shows the average amount of time from one purchase to the next of a given set of items and driver items

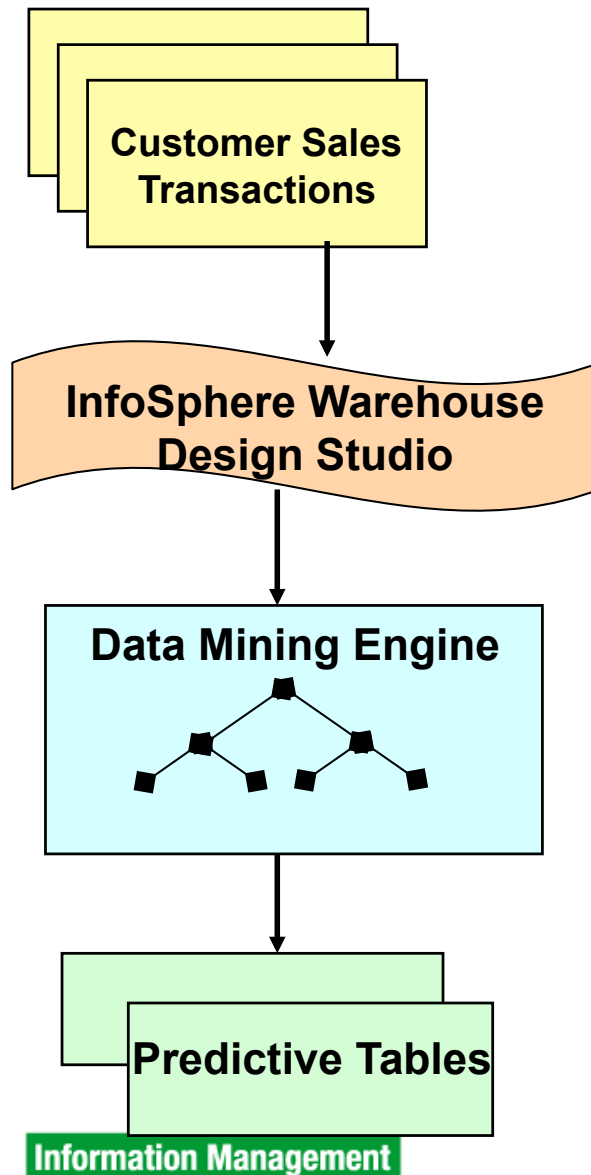


Process Overview

- > Step 1 – Create Mining Flow in InfoSphere Warehouse using Design Studio operators via a drag-and-drop palette approach. In this case we use the associations and sequences algorithms.

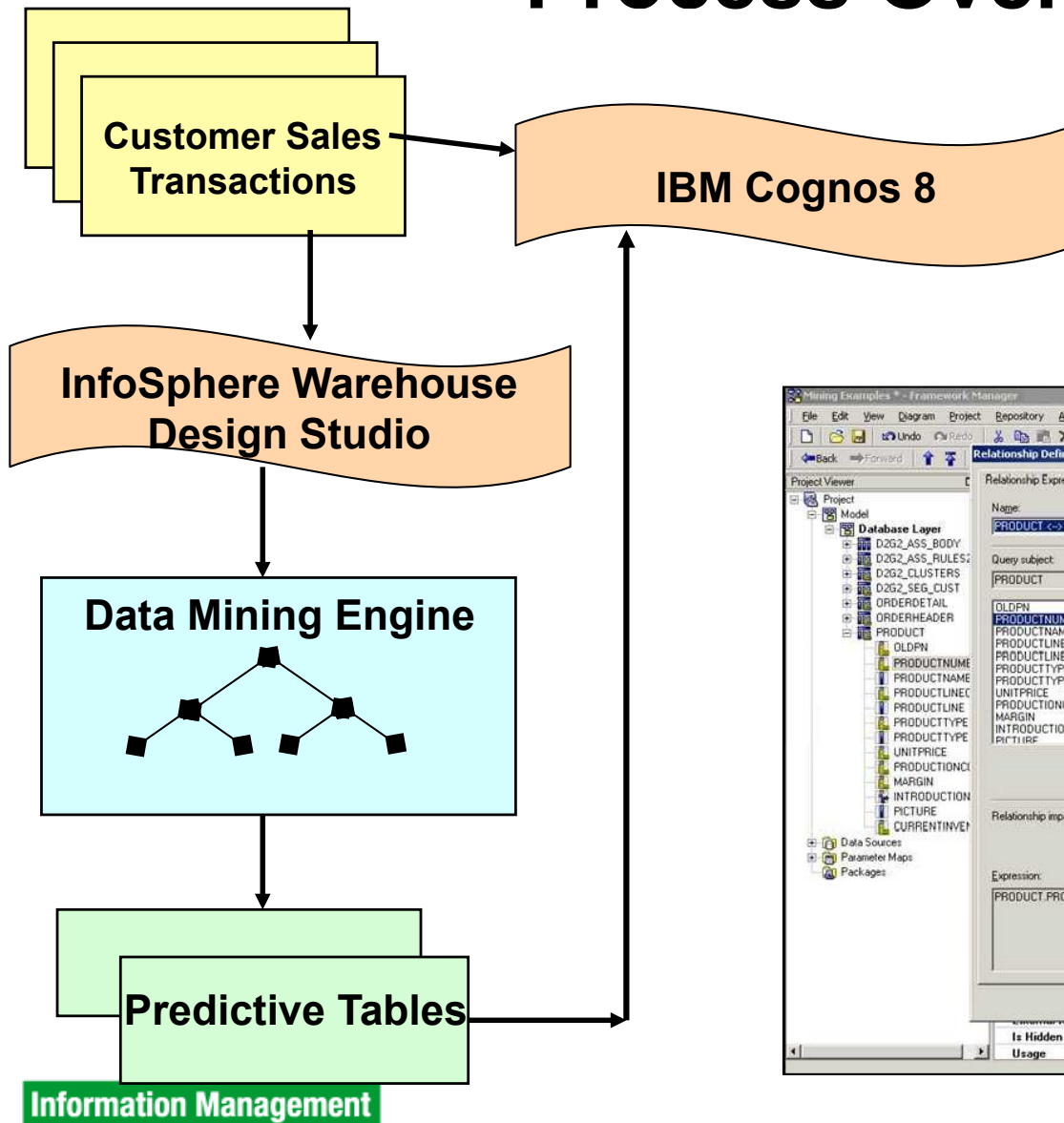


Process Overview

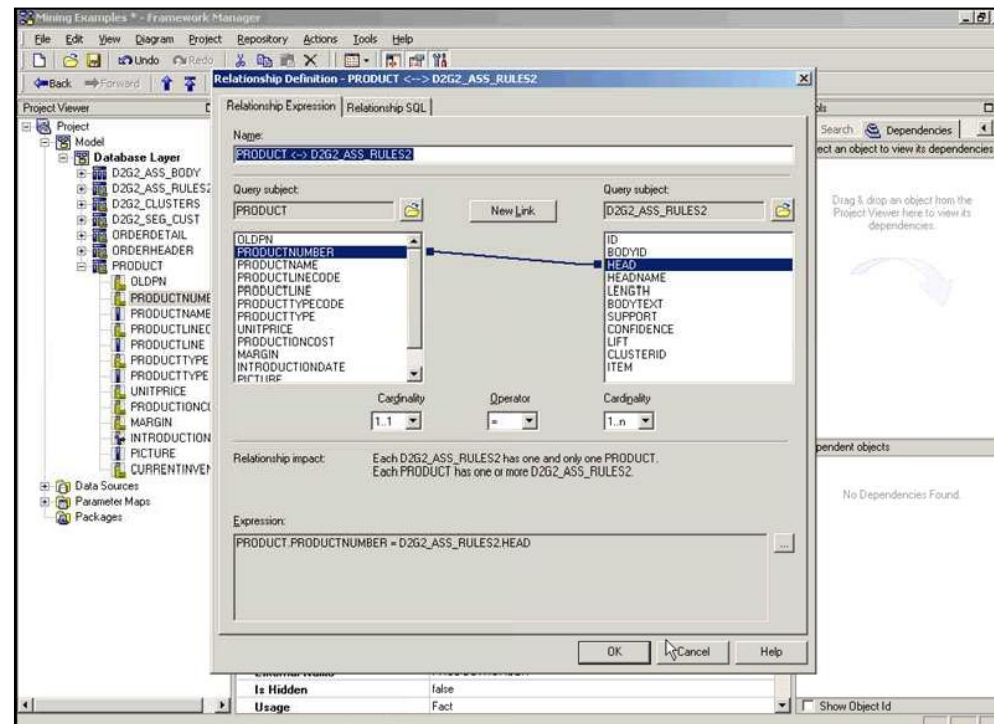


- > Step 1 – Create Mining Flow in InfoSphere Warehouse using Design Studio operators via a drag-and-drop palette approach
- > Step 2 – Run the data mining engine to create the predictive tables containing information about which products are purchased together (market basket analysis) or over time (sequencing)

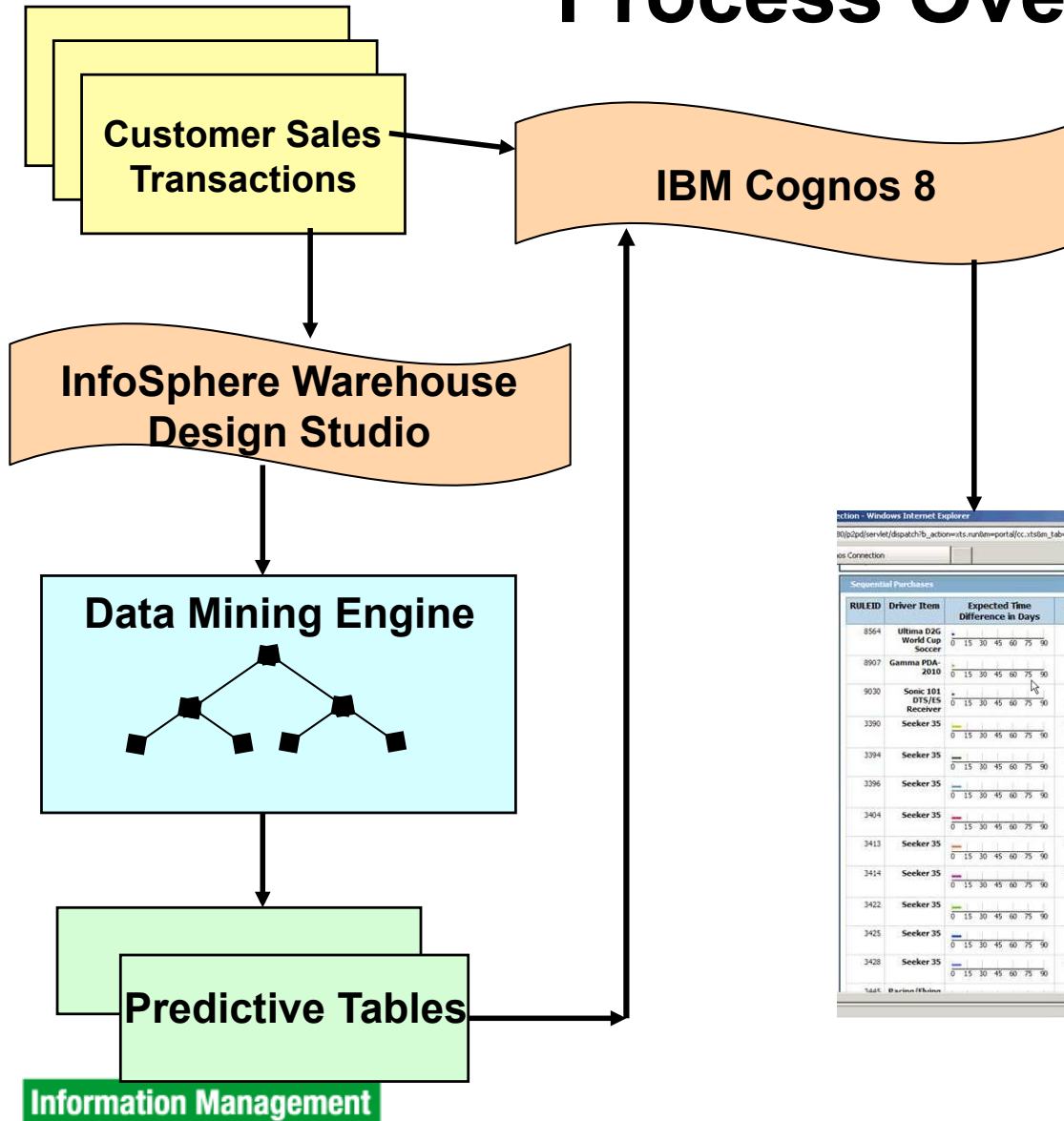
Process Overview



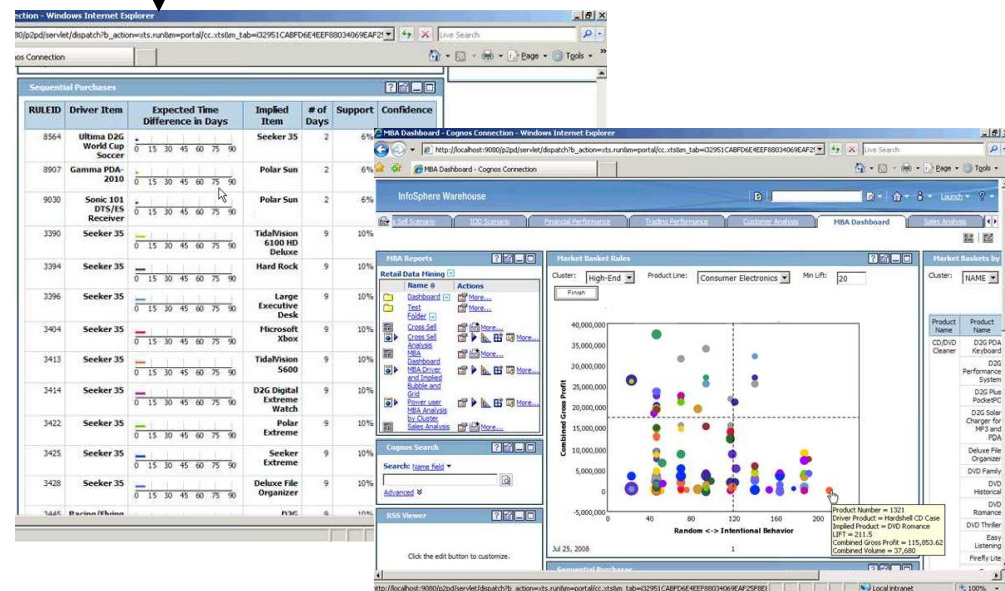
- > Step 3 - Use IBM Cognos 8 Framework Manager to setup table relationships with existing tables and the predictive results tables. Can also define functions and stored procedures as req'd.



Process Overview



> Step 4 - Use IBM Cognos 8 to create Dashboards, Reports etc to communicate the information broadly



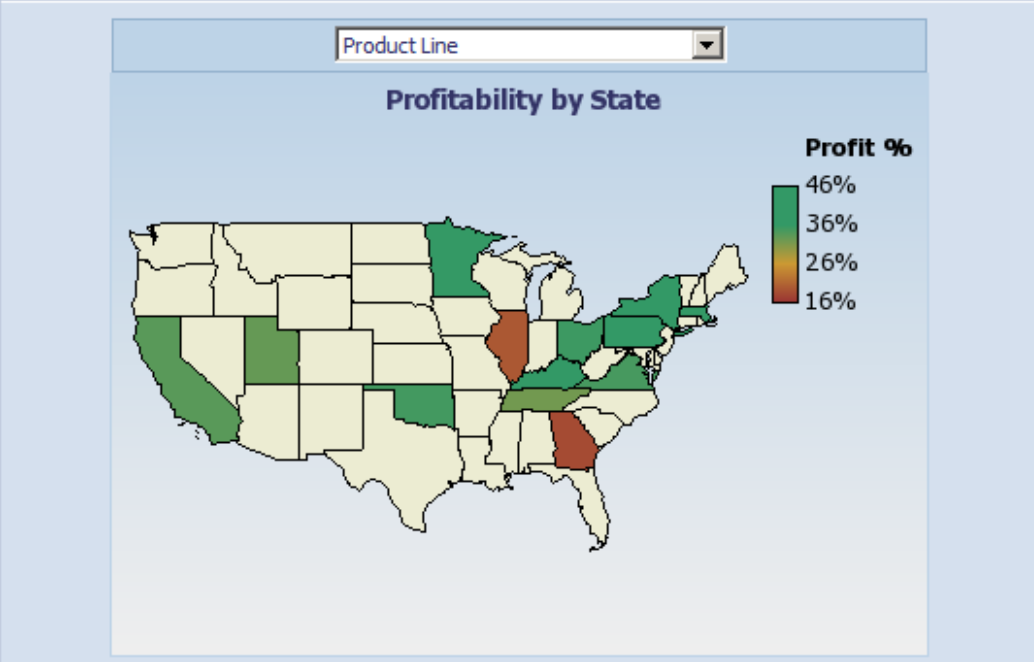
D2G SALES ANALYSIS

TOP SELLING PRODUCTS

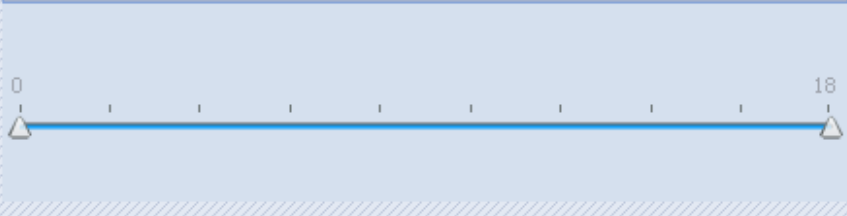
Top: * 3

Rank	Product Name	MBA	SEQ	Sales [\$ 000]
1	Sonic DTC Digital Advance	☰	→	\$85,849
2	Sonic 101 DTS/ES Receiver	☰	→	\$47,091
3	Sonic TLR Thunder Receiver	☰	→	\$46,184
Consumer Electronics				\$179,123
1	D2G Blast Em	☰	→	\$25,938
2	Microsoft Xbox	☰	→	\$6,385
3	Sega Dreamcast	☰	→	\$5,632
Entertainment Media				\$37,955
1	Computer Extended Desk	☰	→	\$51,852
2	Computer Desk Ultra	☰	→	\$41,539
3	Standard Stool	☰	→	\$38,477
Home Office				\$131,867
1	Wavestation 4200 FP	☰	→	\$384,344

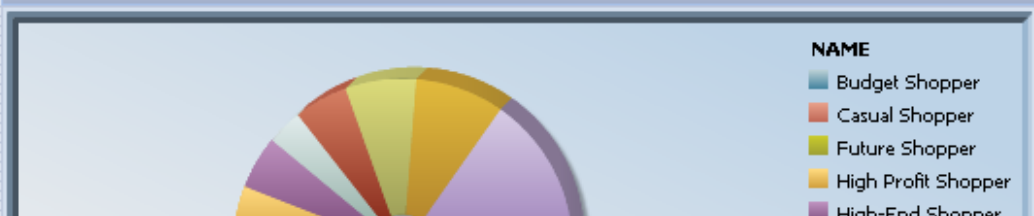
PROFITABILITY BY STATE

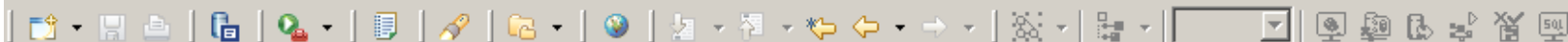


RESTRICT CUST SEGMENTCOUNTS TO...



SEGMENTATION COUNTS



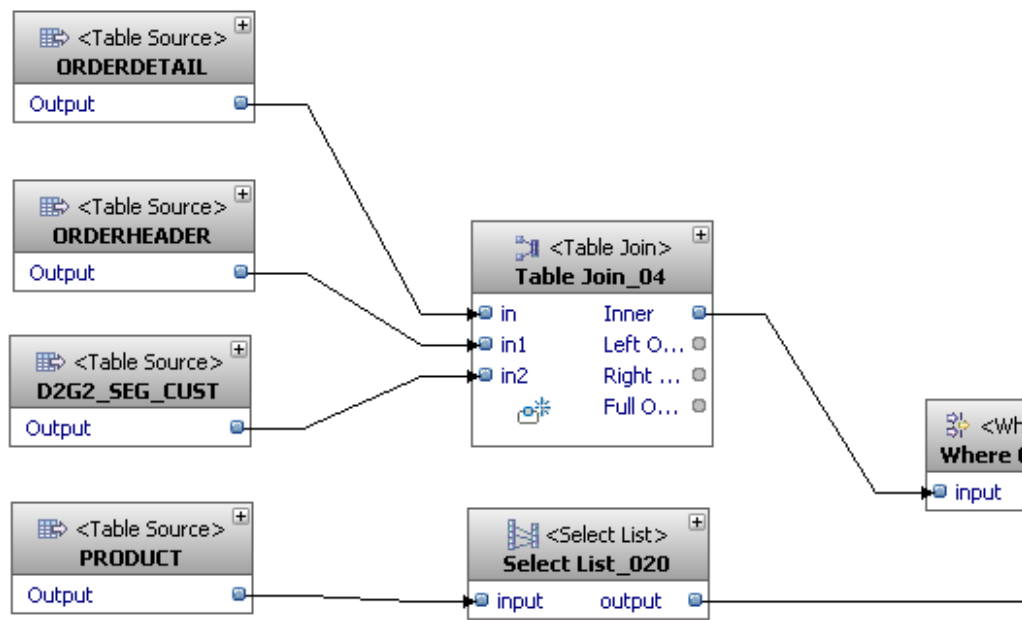


Data Project Explorer

- D2G2 Data Project (D2G2:jdbc:db2://loca
- D2G2 OLAP
 - Data Diagrams
 - Data Models
 - Other Files
 - SQL Scripts
- D2GO
 - Control Flows
 - Data Diagrams
 - Data Flows
 - Data Models
 - D2G2 Data Model.dbm
 - Dynamic Data Model-D2G2.dbm
 - File Definitions
 - Miningblox Profiles
 - Mining Flows
 - Churn Prediction

Database Explorer

- DBO
 - Aliases
 - Dependencies
 - Federated Stored Pro
 - MQTs
 - Nicknames
 - Packages
 - Sequences
 - Stored Procedures
 - Tables
 - User-Defined Functio
 - User-Defined Types
 - Views
 - AP_AGED_BALAN
 - AP_INVN_OFFST



Palette

- Selection Tool
- Connection
- Sources and Targets
- Mining Operators
 - Find Deviations
 - Find Rules
 - Cluster Table
 - Predict Column
- Associations
- Sequences
- Clusterer
- Predictor
- Rule Filter
- Scorer
- Tester
- Associations Extractor
- Sequences Extractor
- Cluster Extractor
- Correlations Extractor
- Fields Extractor
- Tree Rules Extractor
- Gains Extractor
- Quality Extractor
- Preprocessing Ope...
- Text Operators

Launch Design Studio and select a solution plan

Solution Plan Selection

A **Flow Solution** helps you to create a flow for your business problem or scenario.
 A **Solution Plan** defines the steps to be completed to create the flow solution.

1. Select a Problem Domain: 2. Select a Problem: 3. Select a Solution Plan:

Problem Domains

- Advanced Analytics Solutions
- Industry Solutions
- Warehouse Transformations

Advanced Analytics Problems

- Association Rules
- Clustering
- Prediction

Solution Plans

- General association

General Associations solution plan

You can use the General Associations solution plan to create mining flows that complete most tasks that involve discovering association rules.

The analysis finds ... in your data. Example of this ... mining is market ...

Associations creates a mining flow that organizes your data to discover the following ...

- A table that contains the elements of the bodies of the association rules
- A table with all the items contained in the data together with properties such

Go To:

- All Topics
- Search
- Related Topics
- Bookmarks
- Index

Context-sensitive help throughout the solution

To be added: dedicated plan for market-basket analysis ("associations" is the underlying approach)

Select the data source with customer purchases

The screenshot shows two overlapping dialog boxes in the IBM Cognos interface. The top dialog, titled "Connection Selection", is at "Step 1 / 26" and shows a list of database connections: DWECTRL, DWECUST, DWESAMP (highlighted), NHTSA, SAMPLE, and TORO. The bottom dialog, titled "Select a source table", is at "Step 2 / 26" and shows a tree view of tables under the "DB2ADMIN" connection. The "RETAIL" table is selected. To the right of the tree view, the "Selected source table:" field contains "DB2ADMIN.RETAIL" and the "Sampling rate in percent:" field contains "100.0".

Specify how rules should be generated

The screenshot shows a configuration window for an association operator. The window title is "Configure the Association operator" and it is part of a multi-step process (Step 7 / 26). The current step is "7. Configure the A...", with previous steps being "6. Remove columns" and the next being "8. Specify name ta...".

The configuration is divided into two main sections:

- Specify the transaction format:** This section explains that if several rows of the source table can contain data for a single transaction, a unique transaction column must be specified. Two radio buttons are present: "A transaction is stored in one row" (unselected) and "A transaction is stored in several rows" (selected). A "Group column:" dropdown menu is set to "CUSTOMER_ID".
- Specify mining settings:** This section allows specifying characteristics of the association rules. It includes:
 - Maximum rule length:** Radio buttons for "Unlimited" (unselected) and "Specify:" (selected) with a numeric input field set to "2".
 - Minimum confidence in percent:** A text input field set to "25.0".
 - Minimum support in percent:** A text input field set to "5.0".

At the bottom of the window, there are navigation buttons: "Back" and "Next".

The market-basket solution plan would use the term "purchase" instead of "transaction"

Create (and run) “analytic flow” that generates rules

The screenshot shows a web browser window titled '*MyMBA' with a 'Summary' tab. The page is at 'Step 26 / 26' and displays the following information:

- Flow Solution Summary**
 - Problem Domain: Advanced Analytics Solutions
 - Problem Type: Association Rules
 - Solution Plan: General association
 - Source Table(s): DB2ADMIN.RETAIL
 - Target Table(s): DB2ADMIN.RETAIL_RULES
- Target Flow Creation**
 - Name of the target mining flow: MyMBA
 - Overwrite existing flow
 - Buttons: [Create Flow](#) [Create and Run Flow](#)

Navigation buttons at the bottom include 'Back' and 'Next'. The footer shows 'Solution Editor' and 'Overview' tabs.

Keep track of where you are

Overview - General association

The Overview page lists the steps of the solution plan that you selected. Click the current step or a step that you have completed to display the step in the solution editor, or click the button in the top-right corner to return to the current step in the solution editor.

Data Pre-processing

1. Connection selection
2. Select a source table
3. Decision: Transform your source data
4. Add columns
5. Select rows
6. Remove columns

➔

Association Rule Model Cr...

7. Configure the Association operator
8. Specify name tables
9. Specify taxonomy tables
10. Decision: Post-process your new mining model
11. Decision: Save association rules and related information
12. Save association rules: Select columns
13. Save association rules: Select target table

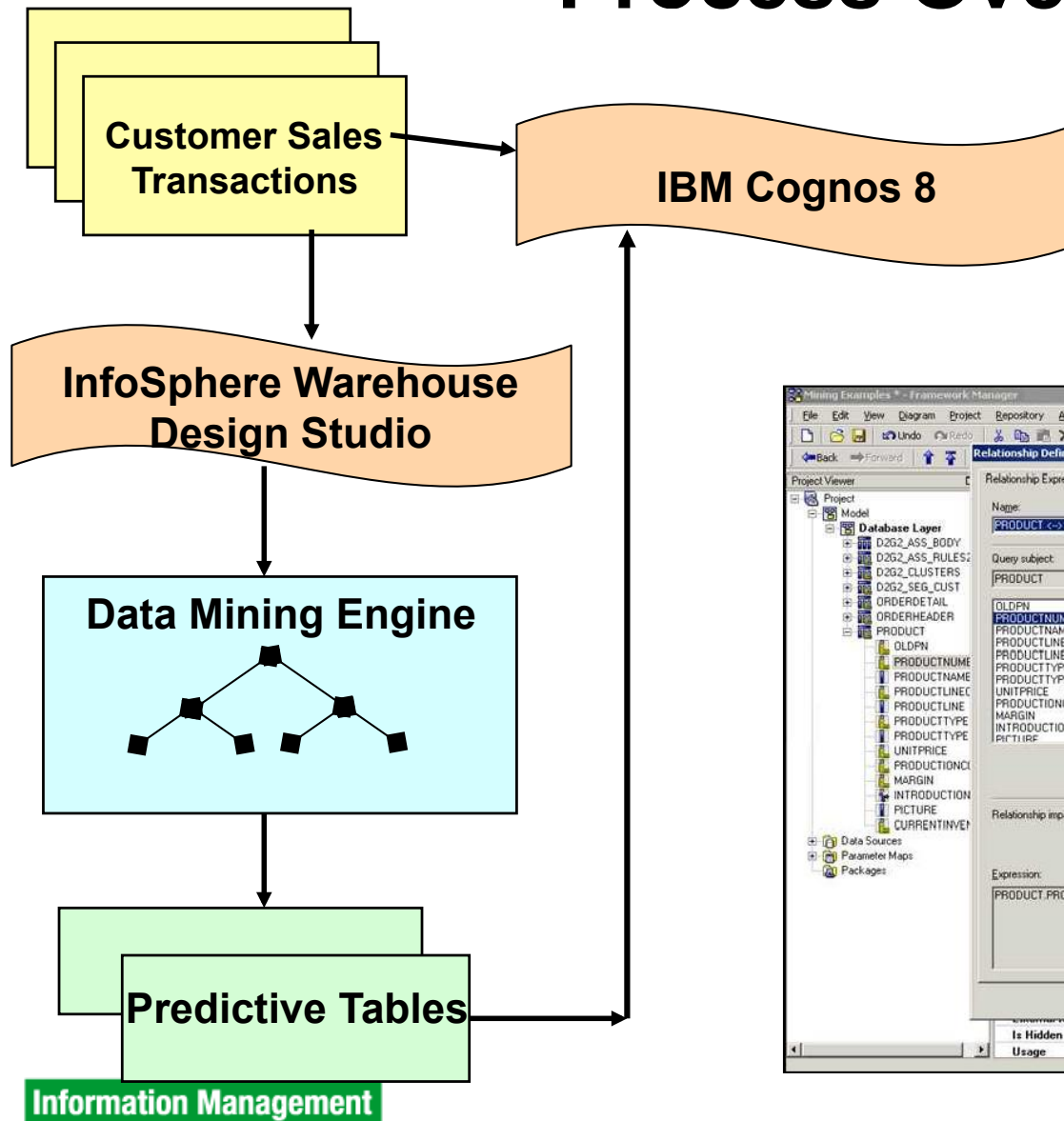
➔

Scoring New Item Sets

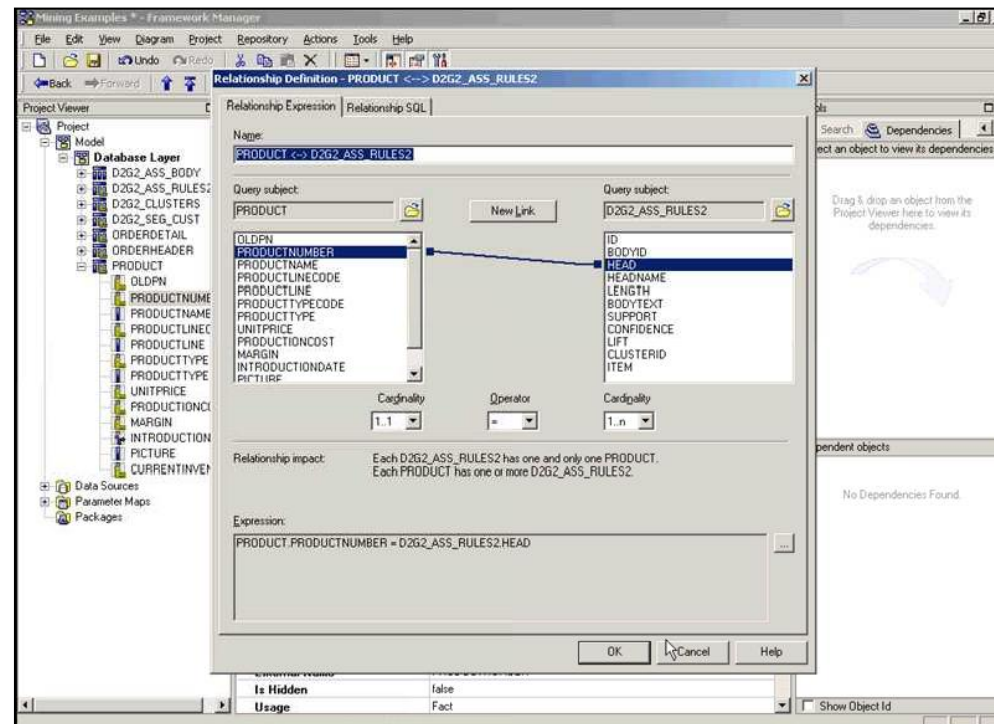
16. Select source table for scoring
17. Decision: Column and row filtering for scoring table
18. Select or add columns to scoring table
19. Choose rows in scoring table
20. Scorer configuration
21. Decision: Column and row filtering for scoring result
22. Add columns to scoring result

Irrelevant steps are hidden, based on user decisions

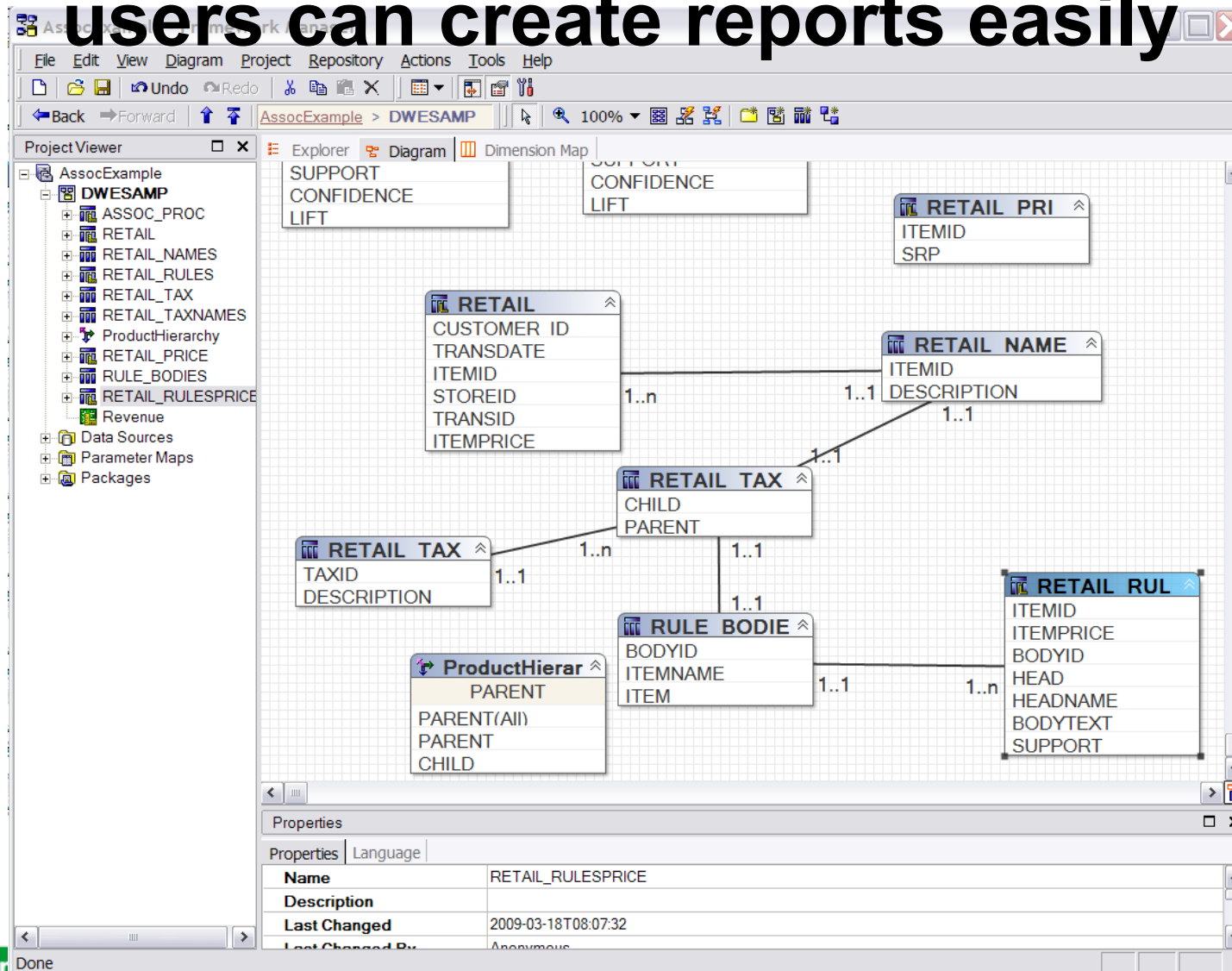
Process Overview



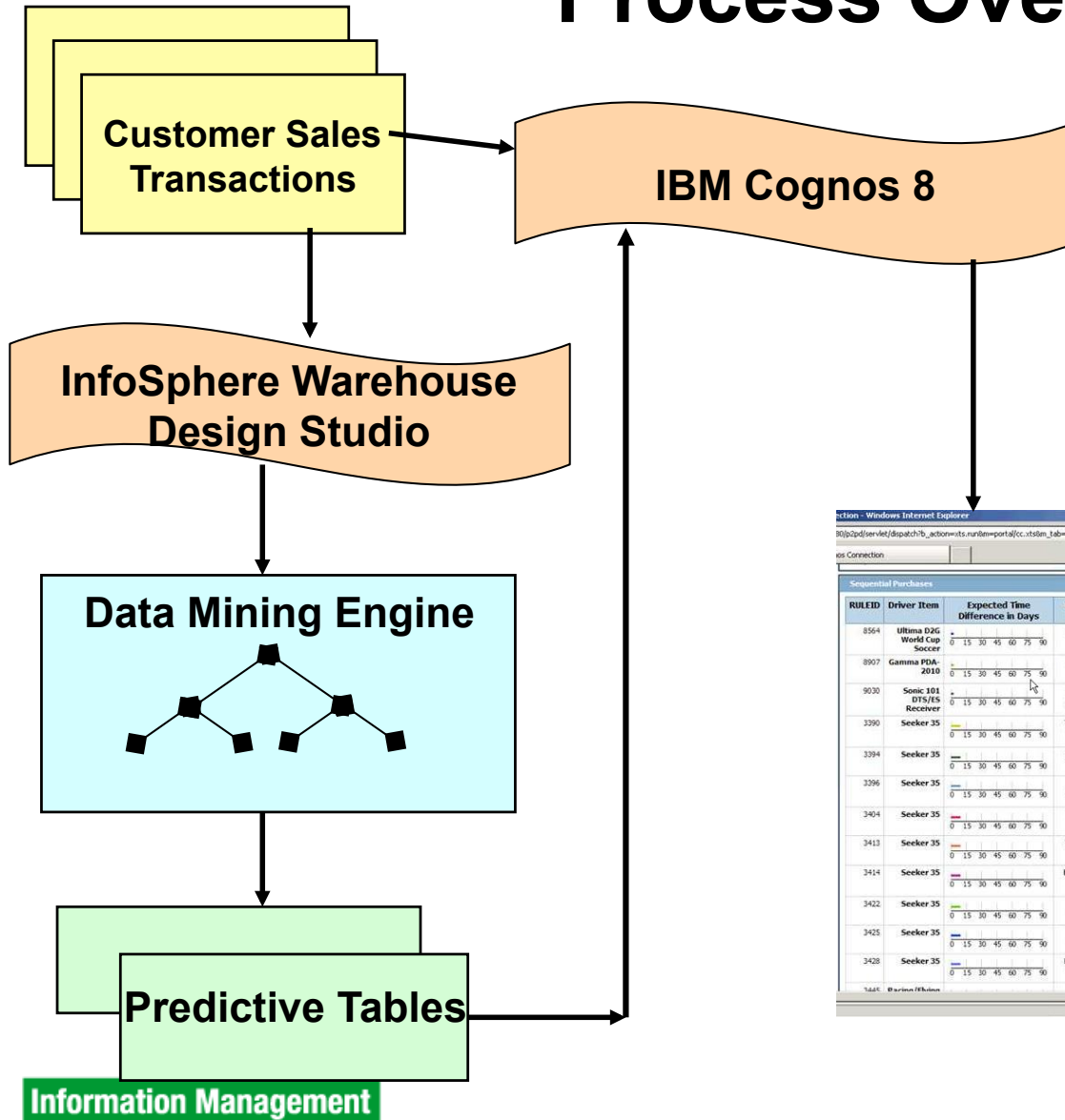
- **Step 3 - Use IBM Cognos 8 Framework Manager to setup table relationships with existing tables and the predictive results tables. Can also define functions and stored procedures as req'd.**



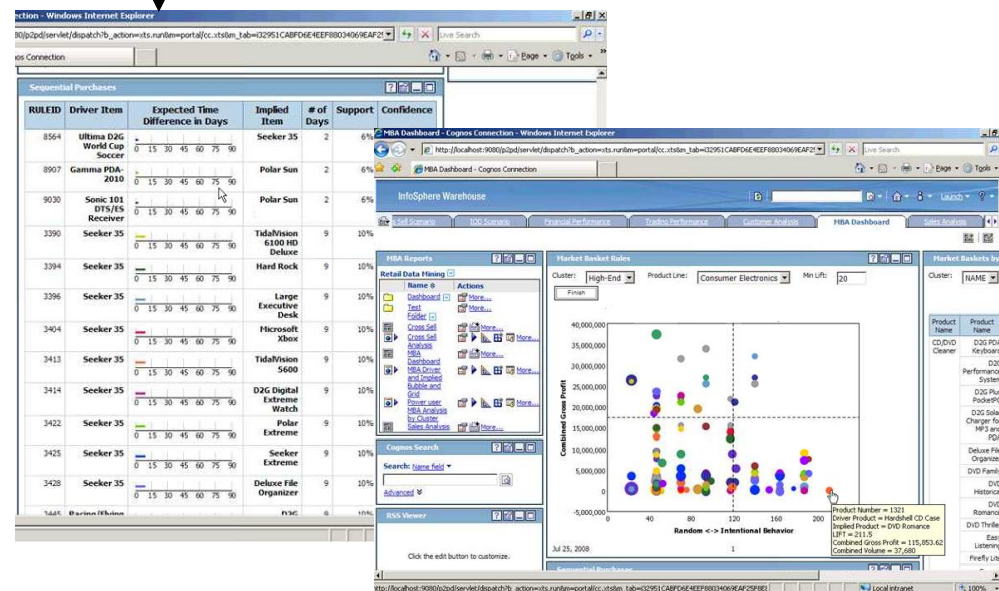
Framework Manager: Model the data so users can create reports easily



Process Overview



➤ Step 4 - Use IBM Cognos 8 to create Dashboards, Reports etc to communicate the information broadly



Build Reports, Dashboards, etc

The screenshot shows the IBM Cognos Report Studio interface within a Microsoft Internet Explorer browser window. The title bar reads "DraggedRevenueForProduct - Report Studio - Microsoft Internet Explorer". The interface includes a menu bar (File, Edit, View, Structure, Table, Data, Run, Tools, Help) and a toolbar with various icons for report design.

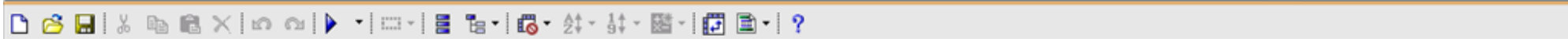
On the left, the "Insertable Objects" pane lists a hierarchy of objects: AssociationRulePackage, ASSOC_PROC, RETAIL, RETAIL_NAMES, RETAIL_RULES, RETAIL_TAX, RETAIL_TAXNAMES, ProductHierarchy, RETAIL_PRICE, RULE_BODIES, RETAIL_RULESPRICE, ITEMID, ITEMPRICE, BODYID, HEAD, HEADNAME, BODYTEXT, SUPPORT, and Revenue.

The main workspace displays a report design titled "Drag-on revenue for <%ParamDisplay...%>". It features a bar chart with the following configuration:

- Default measure (y-axis):** <ITEMPRICE>
- Series:** A dropdown menu with a "Drop item here" option.
- Axis titles:**
 - Category (x-axis):** <#HEADNAME#>

The chart shows four groups of bars on the x-axis, each labeled "abc". The y-axis ranges from 0 to 100. Each group contains three bars of different colors (blue, red, and yellow).

At the bottom of the report design, there are three fields: "Date", "1", and "Time".



Insertable Objects

- Great Outdoors Company
 - Years
 - Products
 - Sales region
 - Retailers
 - Margin range
 - Measures
 - Currency

Page layers: Drop members here to create page layers

Context filter: Drop members here to create a context filter (slicer)

Double click to edit text

Columns

Rows

Measures

New - IBM Cognos 8 Go! Dashboard - Window | D2G8.4JanRefresh | VM | http://cognosd2g/cognos8/cgi-bin/cognos.cgi/gdb/cognos/dashboard/html/?frag-mode=edit&ui.backURL=%2fcognos8%2f | Live Search

File Edit View Favorites Tools Help

New - IBM Cognos 8 Go! Dashboard | Page | Tools | Anonymous | My Area | Log On

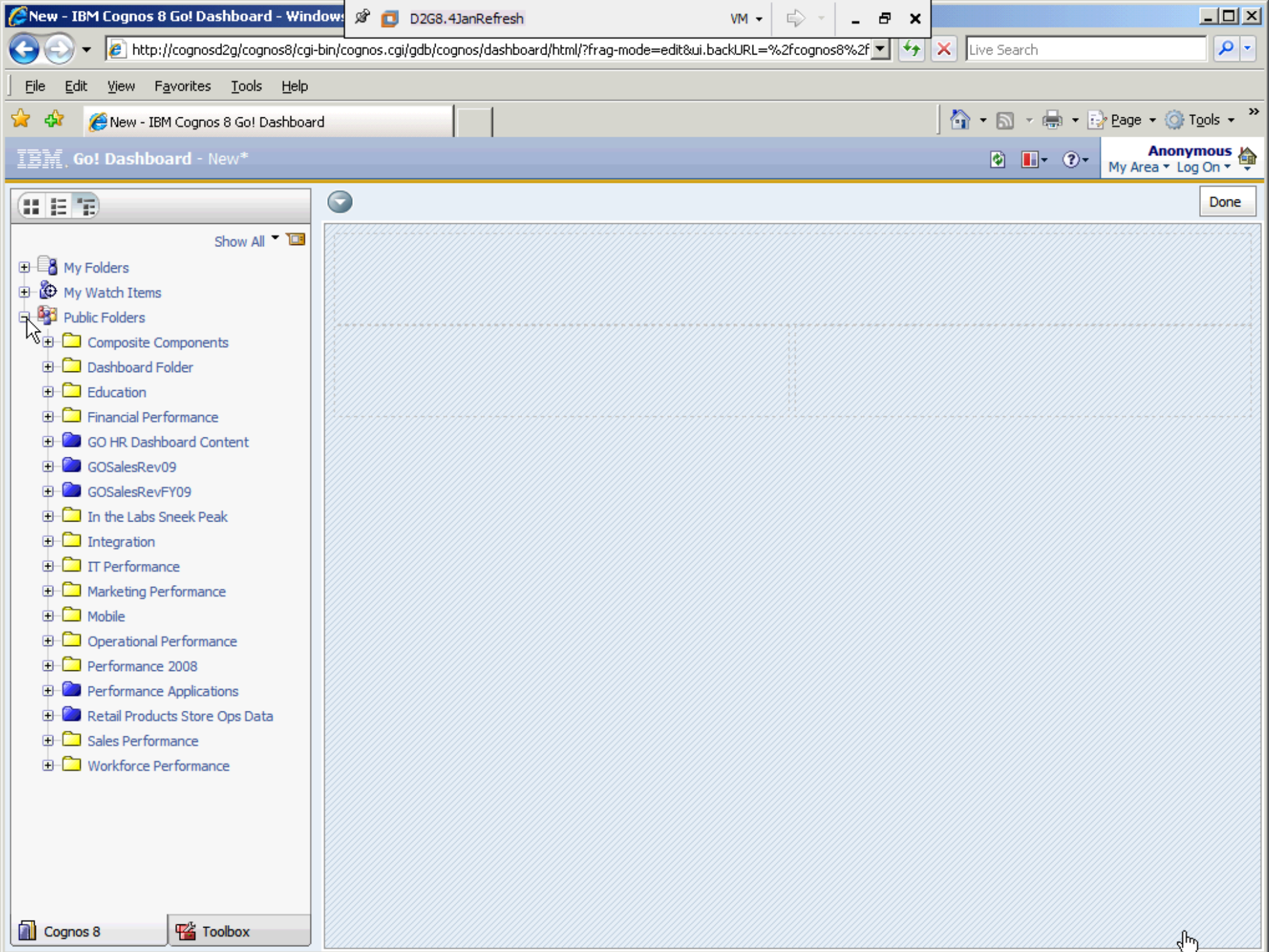
IBM Go! Dashboard - New*

Done

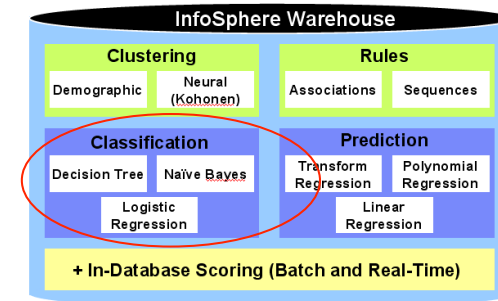
Show All

- My Folders
- My Watch Items
- Public Folders
 - Composite Components
 - Dashboard Folder
 - Education
 - Financial Performance
 - GO HR Dashboard Content
 - GOSalesRev09
 - GOSalesRevFY09
 - In the Labs Sneek Peak
 - Integration
 - IT Performance
 - Marketing Performance
 - Mobile
 - Operational Performance
 - Performance 2008
 - Performance Applications
 - Retail Products Store Ops Data
 - Sales Performance
 - Workforce Performance

Cognos 8 | Toolbox



Sales Example – Account Abandonment Prediction



> Customer Background

- Multi-billion \$ bank

> Business Problem

- Some customers who develop negative account balances subsequently abandon their accounts, resulting in a loss (charge-off) to the bank.

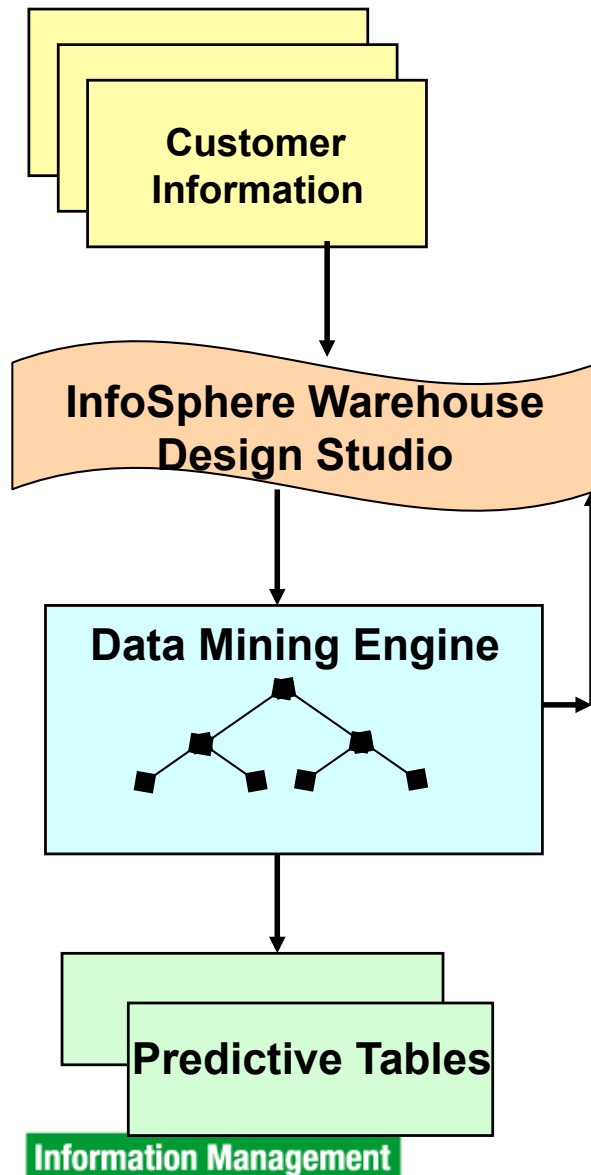
> Business Goals

- Develop a data mining model to predict account abandonment
- Develop a platform for delivering the prediction results in support of bank operations

> Results

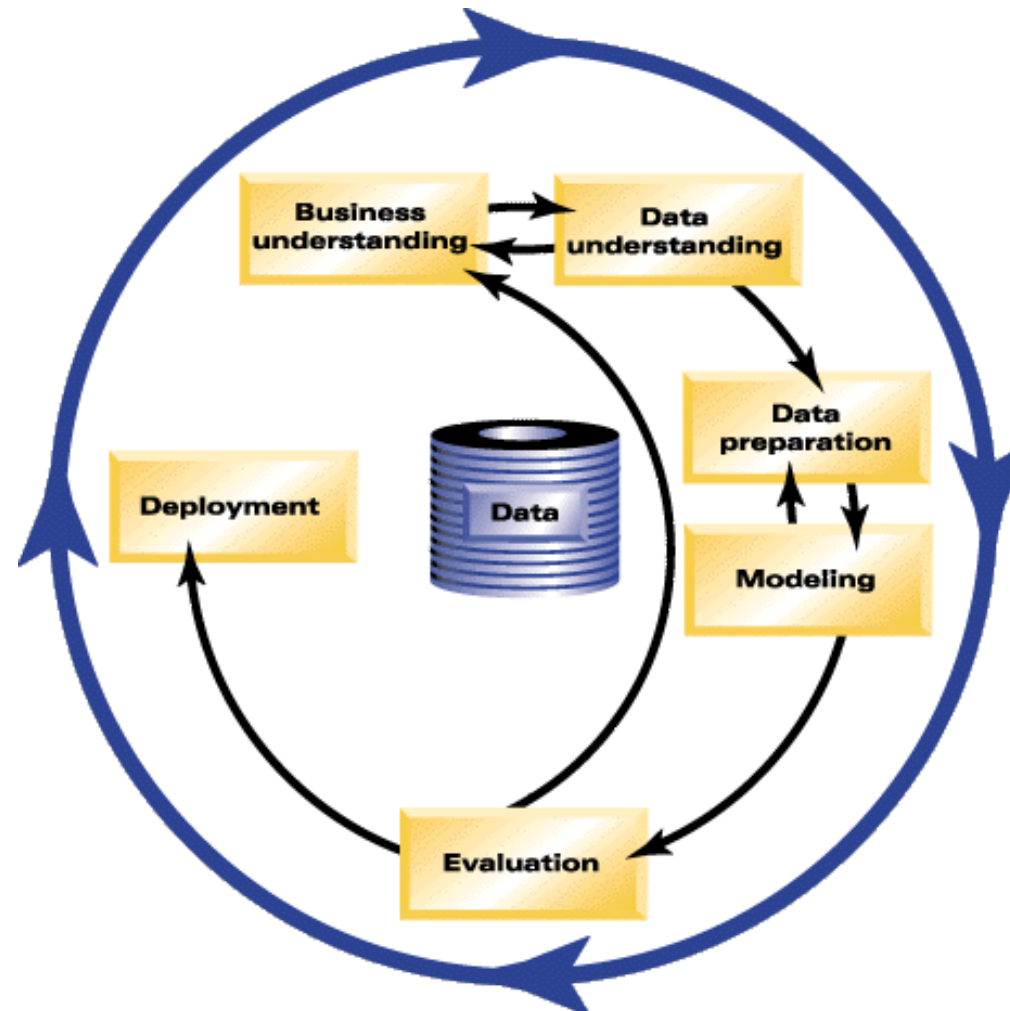
- Created a classification model to predict account abandonment
- Model shows that account balances and customer interactions with the bank are good predictors of account abandonment

Process Overview



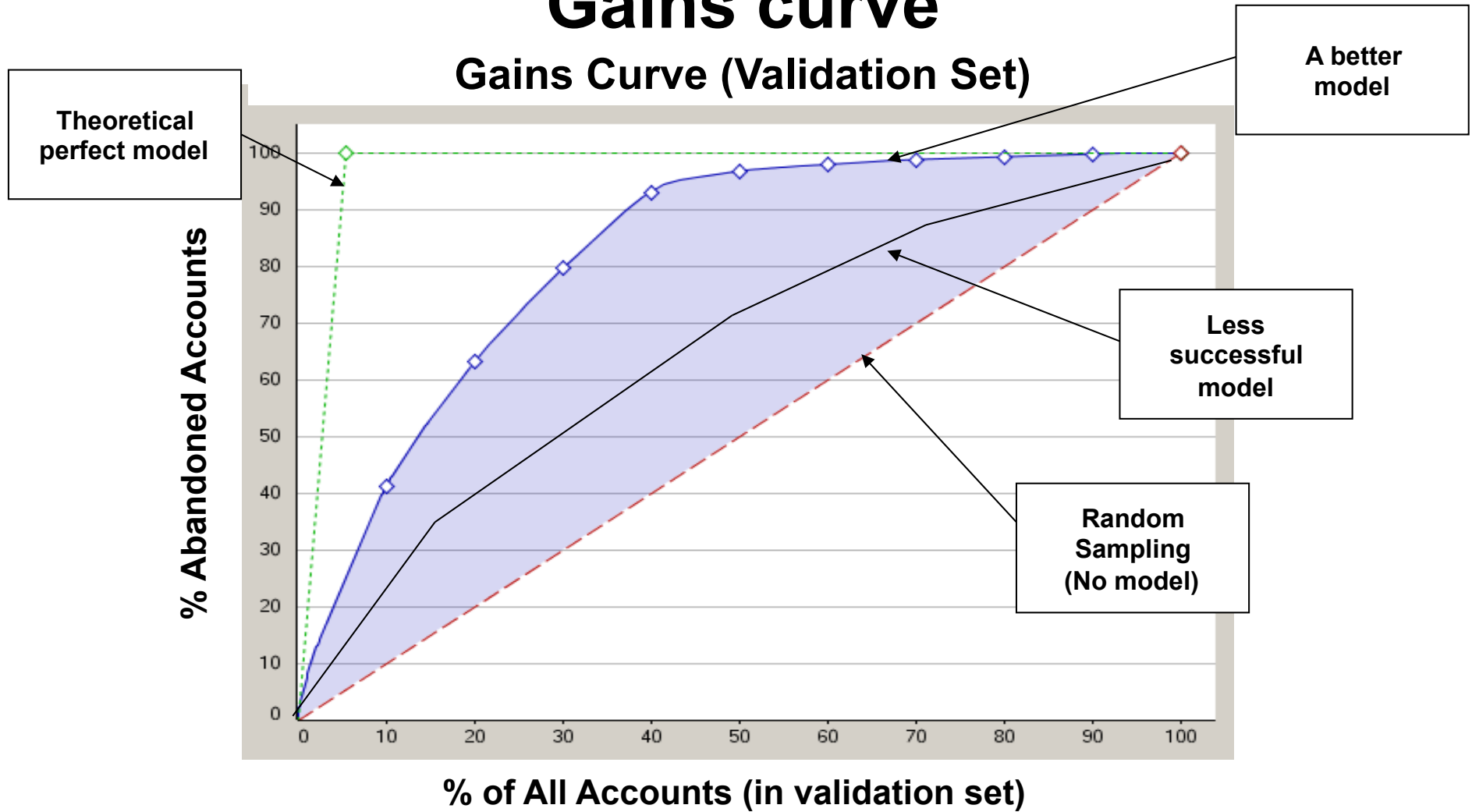
- > Step 1 – Create Mining Flow in InfoSphere Warehouse using Design Studio operators via a drag-and-drop palette approach
- > Step 2-N Iterate to develop a useful model
- > Step N+1 – Run the data mining engine to create the predictive tables containing information about which customers are likely to abandon their accounts (scoring)

CRoss Industry Standard Process for Data Mining (CRISP-DM)



Account Abandonment Prediction – Gains curve

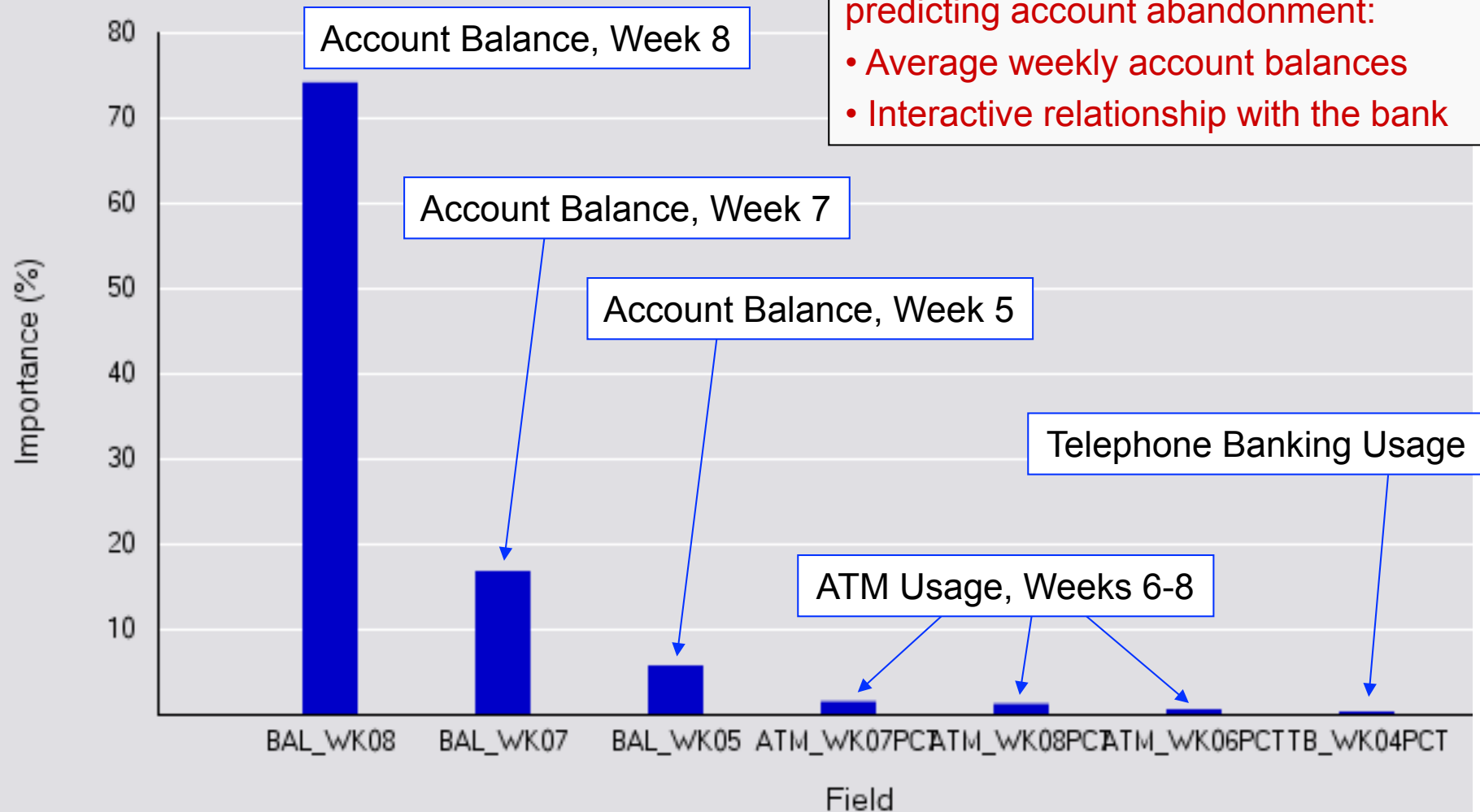
Gains Curve (Validation Set)



Example: Account Abandonment Prediction

▼ Field Importance Chart

Target Field *CO_WK15*



Data mining model found the most important customer behaviors for predicting account abandonment:

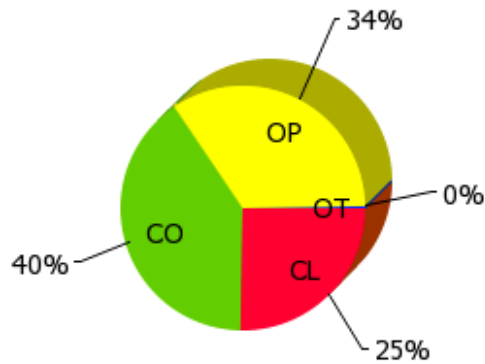
- Average weekly account balances
- Interactive relationship with the bank

Account Summary for Region: Dallas Region

* Dallas Region

Account Status

- CL
- CO
- OP
- OT



% of Accounts

OP-Open, CL-Closed, CO-Charged Off, OT-Other

Total No. Accounts	3,673
# Accounts Open	1,254
# Accounts Charged Off	1,471
# Accounts Closed	933
# Accounts At Risk	37

[Account Summary](#)

Branch Region	Account Status	# Accounts	% Accounts	Fees Generated	Alabama	All Regions
All Regions	CL	7,485	22%	\$118,755.30	60%	
	CO	14,072	40%	\$828,523.90		

REGION
■ All Regions
■ Alabama

Customer references

- > **German Car Manufacturer (Automotive)**
 - Identify problematic combination of parts/defects early
 - ⇒ Reduce warranty expenses
 - ⇒ Improve customer satisfaction
 - Association Rule Mining (Sequential Pattern Mining)
- > **United Health Group (Healthcare)**
 - Analysis of End Stage Renal Disease (ESRD)
 - Sequential Pattern Mining, Clustering
- > **Dillards (Retail)**
 - Increase revenue by identifying distinct segments of high-profit customers (and others who behave like them) and targeting them with more personalized offers
 - Clustering, Association Rule Mining
- > **University Hospital Freiburg (Clinical Research)**
 - Analysis of relations between weather and mental symptoms
 - Clustering (Classification)
- > **Woodforest Bank (Finance)**
 - *Better understand the characteristics of charge-off customers as a first step to developing predictive models for use in the approval process*
 - Clustering, (Classification)

Two Types of Data Mining – Discovery & Predictive

Discovery

- Automatically find trends and patterns
- Answer unasked questions
- Relatively undirected analysis
- Tool reports on findings
- In a word – “Easier”
- Useful for non-statisticians



Predictive

- Specific question
- Probability associated with outcomes
- Directed analysis
- Iterative process
 - Train, Test, Apply
- Apply model in database at customer touch points



Summary

- > You can get more value from your existing information assets
- > Start with discovery
- > Work towards Predictive as you get more comfortable
- > We can help

Qs and As

