

M I N I N G

A
R
T



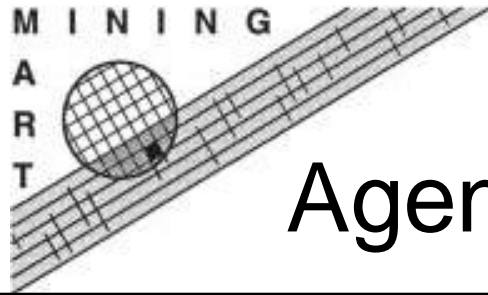
Analyzing Churn of Customers

Marco Richeldi

Alessandro Perrucci

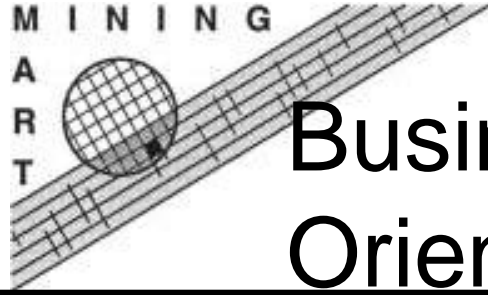
TELECOM ITALIA LAB

Via G. Reiss Romoli 274, 10148 Torino – Italy
{Marco.Richeldi, Alessandro.Perrucci@tilab.com}



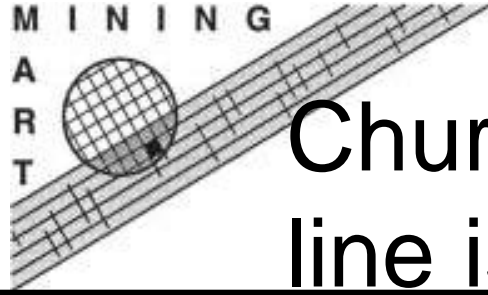
Agenda

- **Churn management in Telcos**
- **A Churn Analysis system for wireless network services**
- **The MiningMart solution**
- **Conclusions**



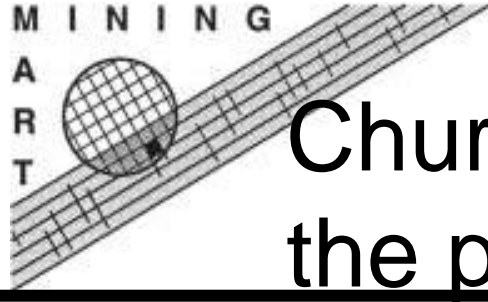
Business Scenario: Customer Orientation is key for Telcos

- **Most Telcos' products and services: commodities (no longer relevant for competitive advantage)**
- **Telcos: evolving a process-oriented organization (CRM, SCM)**
 - CRM application architectures: integrate front-office / back-office applications
 - Through 2005, telcos: mktg automation applications + call centers => unified customer interaction frameworks
- **Europe: Analytical CRM solutions market growing rapidly**
 - CAGR: ~ 50% (from \$0.5 billion in 1999 to \$3.5 billion in 2004)
- **Telco's investment in Analytical CRM moderate due to investments in 2.5G and 3G (UMTS) technology, but relevant**



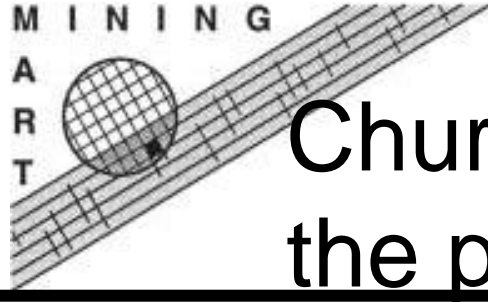
Churn management: a bottom line issue

- **Attracting thousands of new subscribers is worthless if an equal number are leaving**
- **Minimizing customer churn provides a number of benefits, such as:**
 - **Minor investment in acquiring a new customer**
 - **Higher efficiency in network usage**
 - **Increase of added-value sales to long term customers**
 - **Decrease of expenditure on help desk**
 - **Decrease of exposure to frauds and bad debts**
 - **Higher confidence of investors**



Churn management: scooping the problem (1)

- **Churn can be defined and measured in different ways**
 - **“Absolute” Churn.** number of subscribers disconnected, as a percentage of the subscriber base over a given period
 - **“Line” or “Service” Churn.** number of lines or services disconnected, as a percentage of the total amount of lines or services subscribed by the customers
 - **“Primary Churn”.** number of defections
 - **“Secondary Churn”.** drop in traffic volume, with respect to different typology of calls

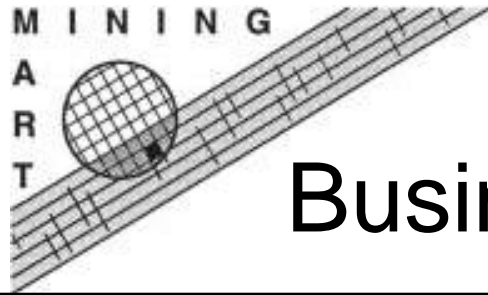


Churn management: scooping the problem (2)

- **Measuring churn is getting more and more difficult**
 - Growing tendency for Business users to split their business between several competing fixed network operators
 - Carrier selection enables Residential customers to make different kind of calls with different operators
 - Carrier pre-selection and Unbundling of the Local Loop makes it very difficult to profile customers according to their “telecommunication needs”
- **Other frequent questions for Fixed Network Services**
 - What if a customer changes his type of subscription, but remains in the same telco? What if the name of a subscriber changes? What if he relocates?

The case study: Churn Analysis for wireless services

- **The framework**
 - A major Italian network operator willing to establish a more effective process for implementing and measuring the performance of loyalty schemes
- **Objectives of the “churn management” project**
 - Building a new corporate Customer Data Warehouse aimed to support Marketing and Customer Care areas in their initiatives
 - Developing a Churn Analysis system based upon data mining technology to analyze the customer database and predict churn



Business understanding

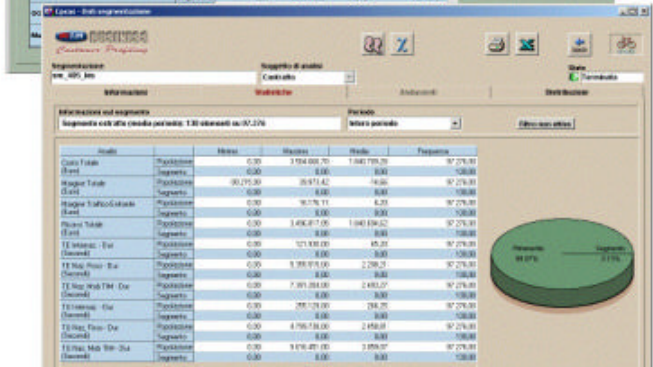
- **Sponsors**
 - Marketing dept., IT applications, IT operations
- **Analysis target**
 - Residential Customers, subscriptions
- **Churn measurement**
 - Absolute, primary churn
- **Goal:**
 - Predict churn/no churn situation of any particular customer given 5 months of historical data

Solution scope



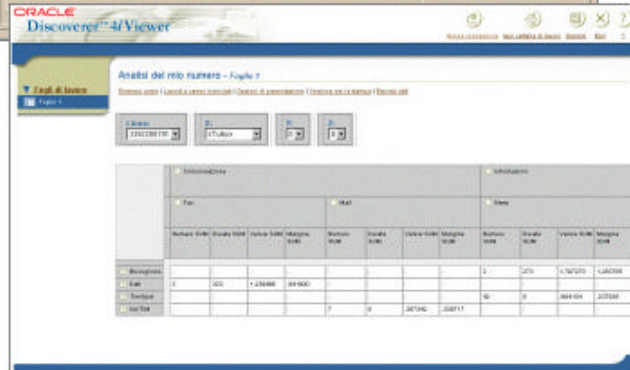
Customer Profiling Consumer:
21 millions of residential customers

Usage patterns analysis of Voice Services by single subscriber line



Customer Profiling Business:
2 millions of business customers

Usage patterns analysis of Voice Services by subscriber line, contract, company, etc.

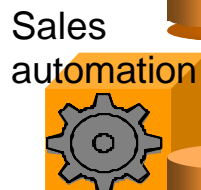
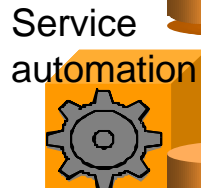
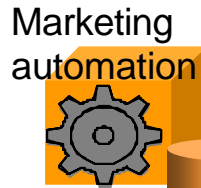
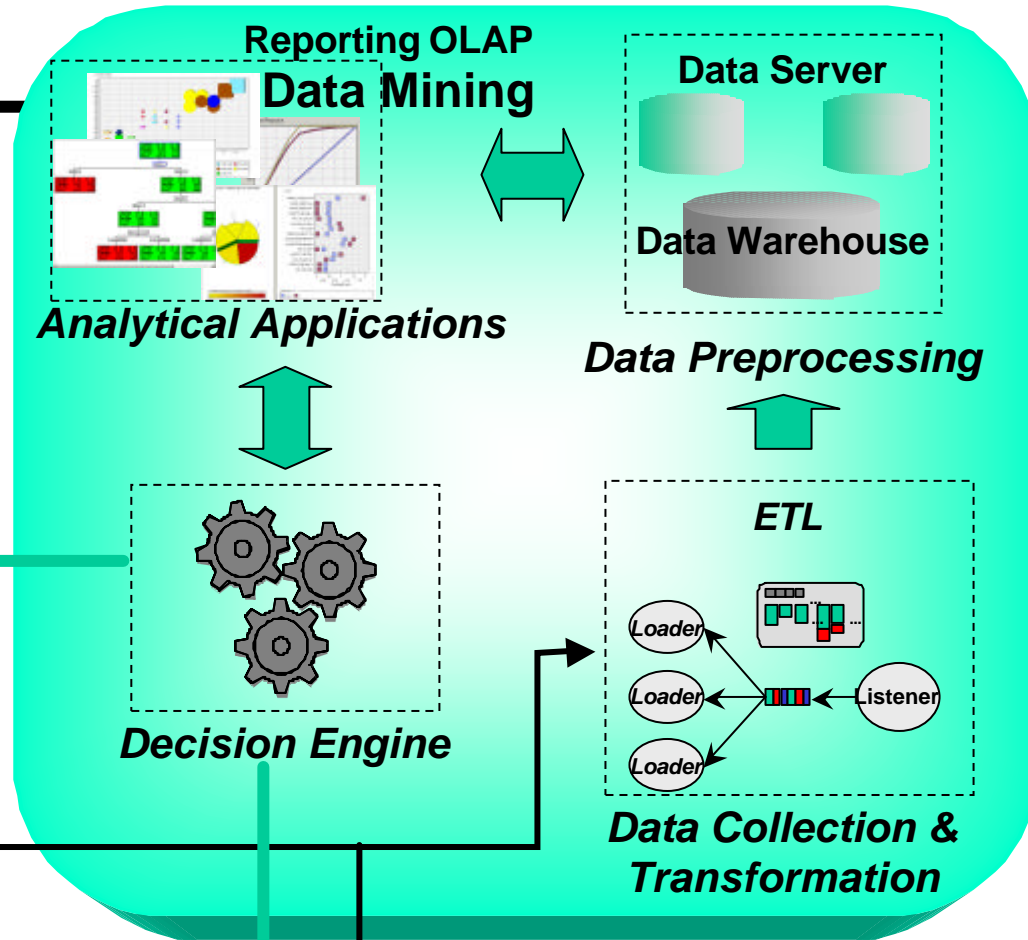


Customer Profiling VAS:
23 millions of customers

Usage patterns analysis of VAS by single subscriber line

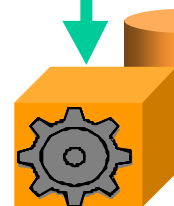
Application framework

- Campaign Targets
- New product / services
- Loyalty schemes
- Performance analysis



Front-office Systems

- Customer data
- Market data
- Sales data
- Customer service contacts

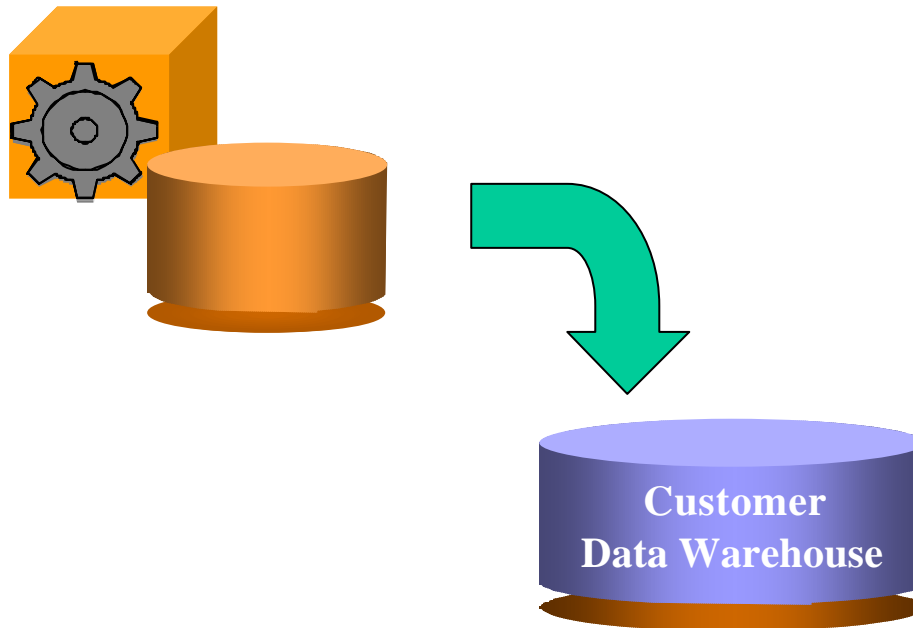


Back-office Systems

- Contracts
- Tariff plans
- Billing data
- Accounts data
- Fraud / Bad debts data

Data understanding

13 operational systems



- More than 500 indicators per customer
- Extraction delay: 2 months
- Loading: on a monthly basis
- Size: 1.5 Tb

Input Data

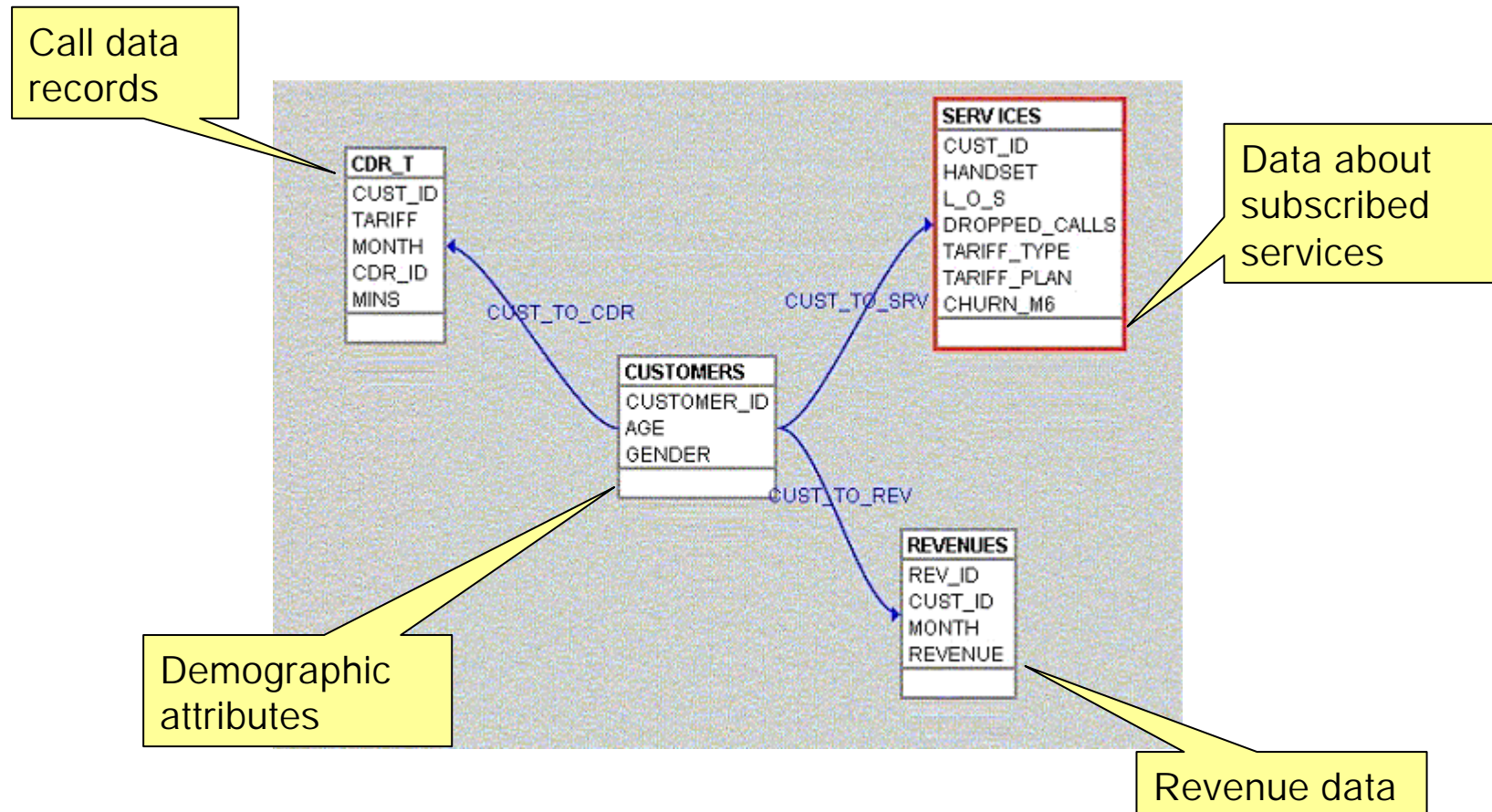
- Customer demographics
 - Basic customer information
- Service Profile
 - Products/services purchased by each customer.
- Tariff plans
 - Details of the tariff scheme in use
- Extra service information
 - Special plans / rates
 - Service bundles
- Call data aggregated by month
- Billing data aggregated by month
- *Complaint information*
- *Fraud and bad debts data*
- *Customer service contacts*
- *Sales force contacts*
- *Market data*

Modeling with Mining Mart

The screenshot shows the MiningMart - ChurnPrediction application. The main window displays a workflow diagram for 'Step 1 - Treat missing values in CDR'. The workflow starts with 'Sel customers having miss vals', followed by 'join', 'Sel incomplete CDRs', 'Segm by customer', 'Segm by tariff', 'Build miss vals estimation', 'UnSegm by tariff', 'UnSegm by customer', 'Merge complete and incomplete CDR', and 'Assign estimated n'. A yellow callout box highlights the main steps: 'Define Concepts, Attributes, Relationships ...', 'Select Operators', and 'Build the execution workflow'.

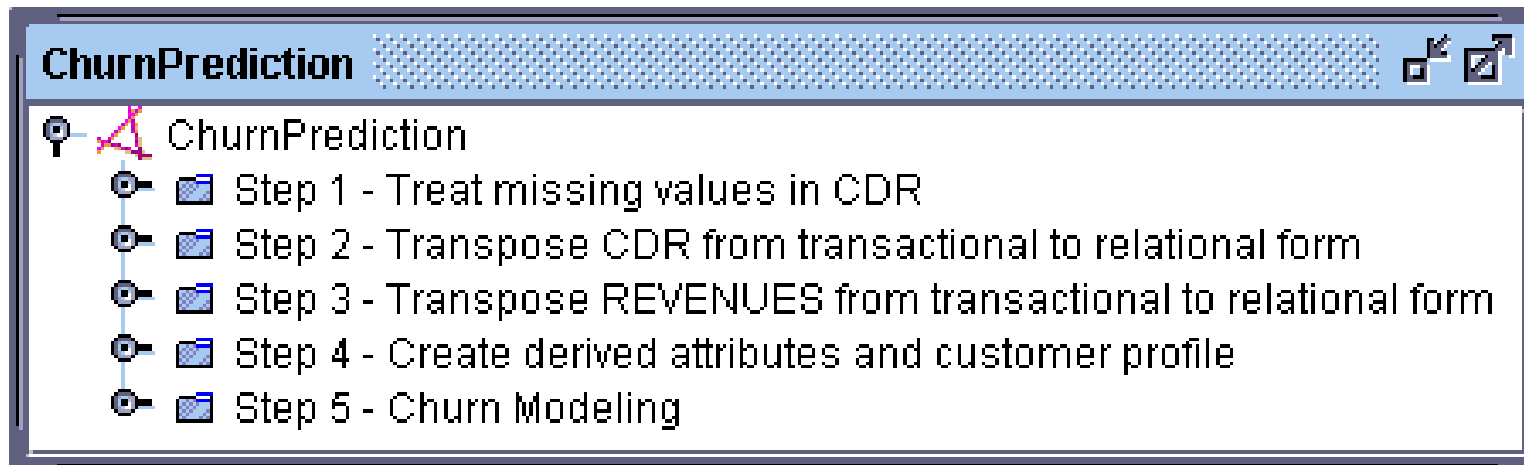
The Concept Editor window shows a list of concepts on the left and a central workspace. The workspace contains two tables: 'CDR_T' and 'REVENUES'. The 'CDR_T' table has columns: CDR_ID, CUST_ID, TARIFF, MONTH, MINS. The 'REVENUES' table has columns: CUST_ID, CUST_AGE, CUST_GENDER. Blue arrows labeled 'CUST_TO_CDR' and 'CUST_TO_REV' indicate relationships between the tables.

Concepts, Attributes, Relationships

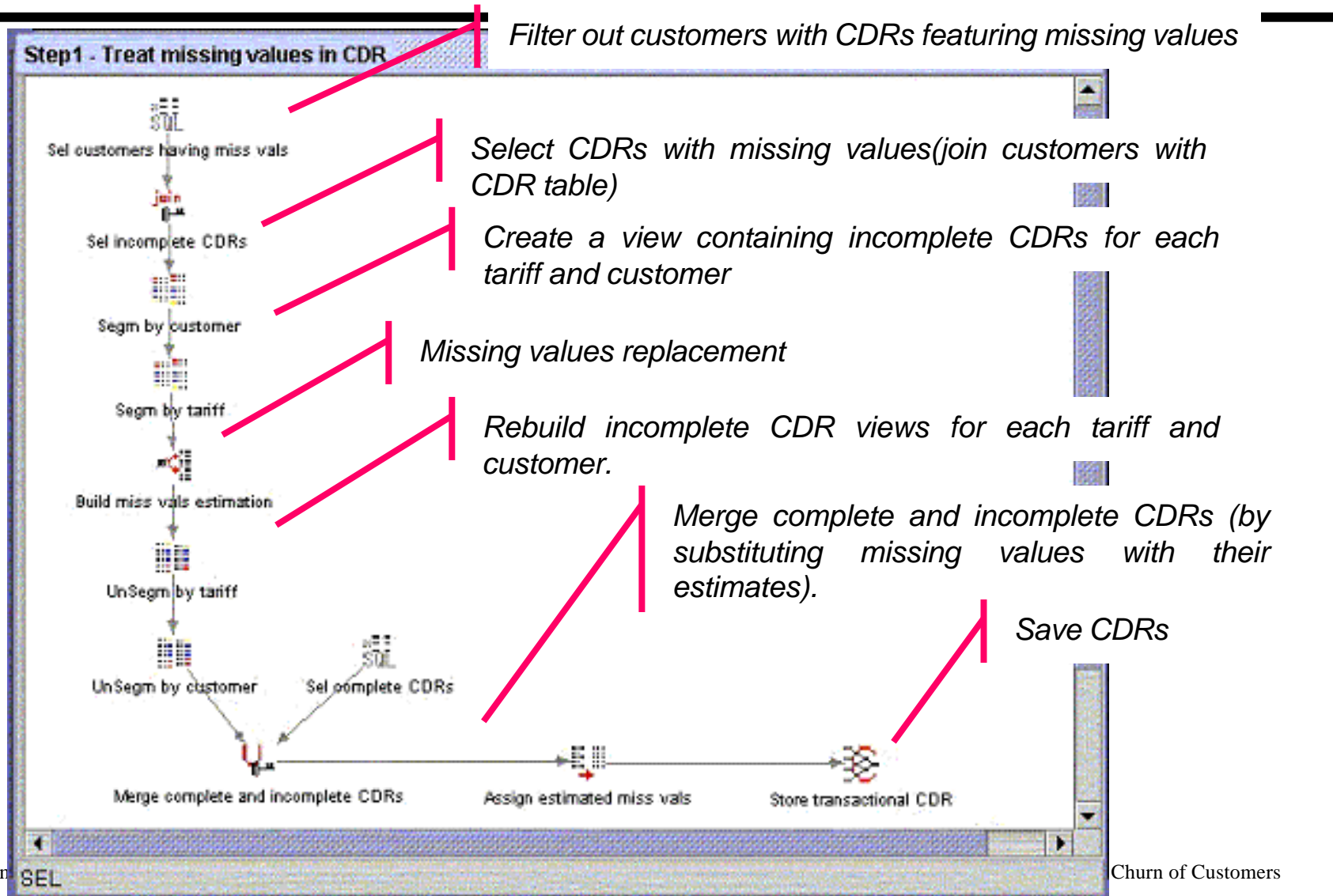


Pre-processing chains

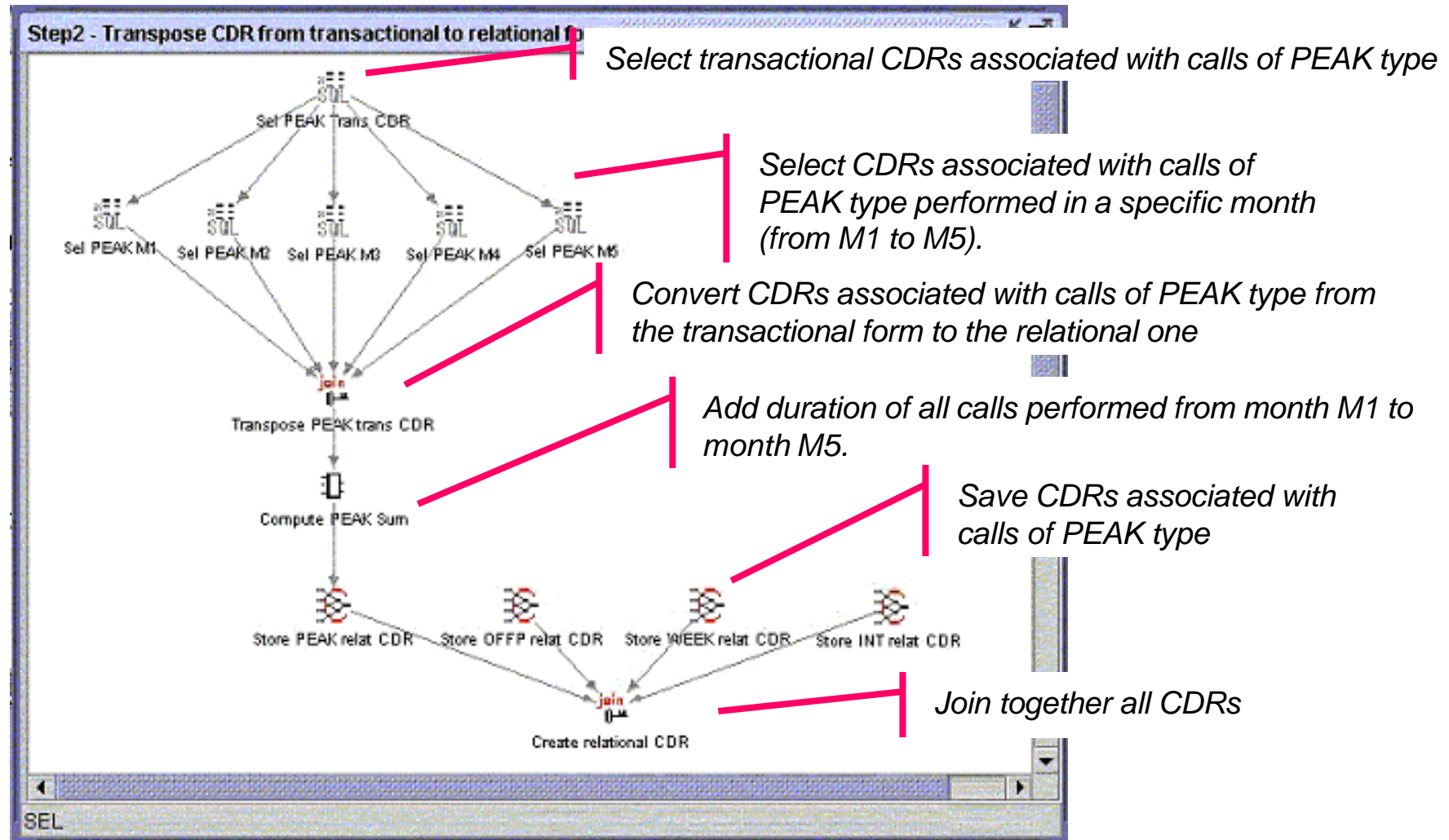
The data mining process has been divided into five tasks as follows:



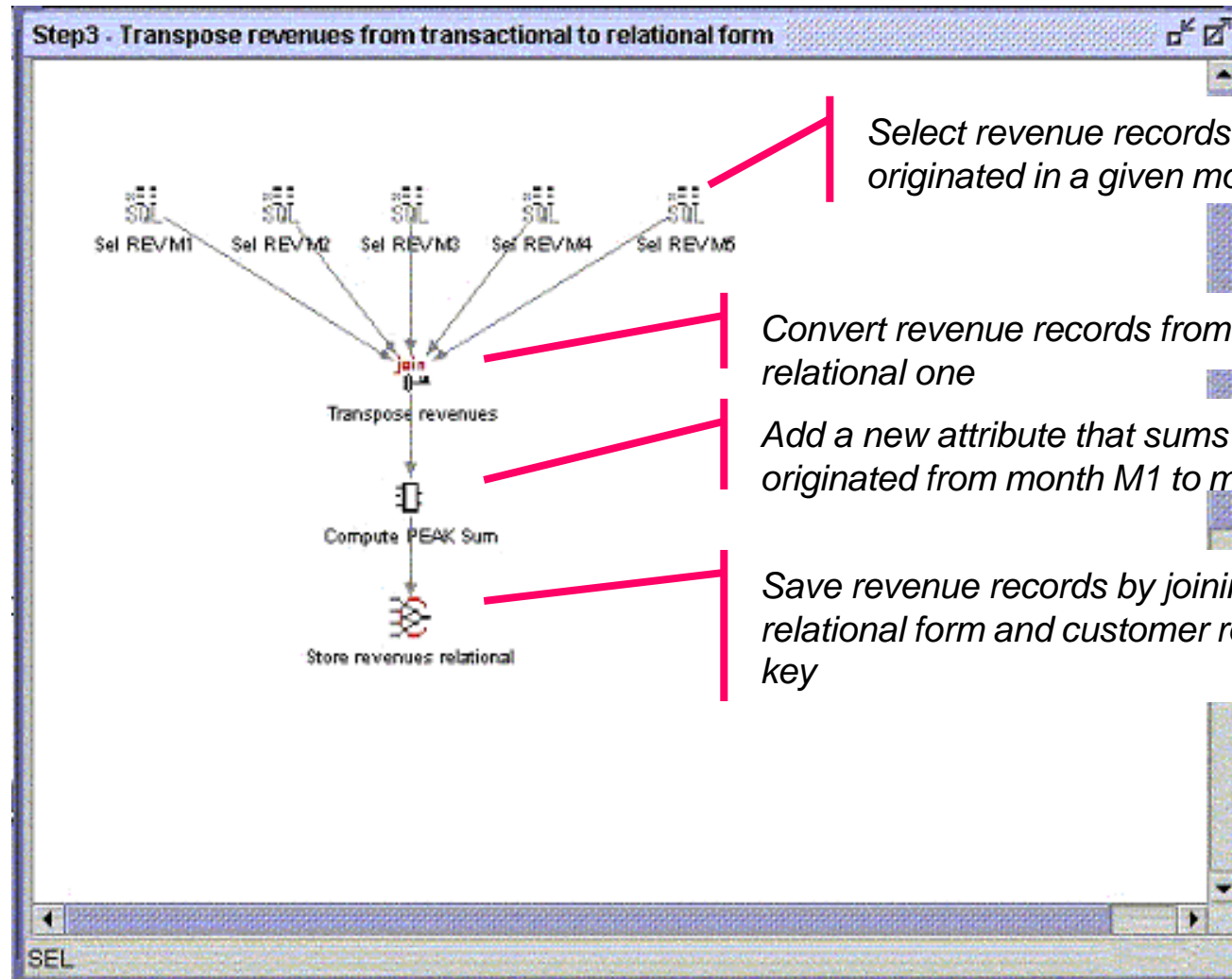
Handle missing values in CDRs



Transpose CDR from transactional to relational form



Transpose REVENUES from transactional to relational form



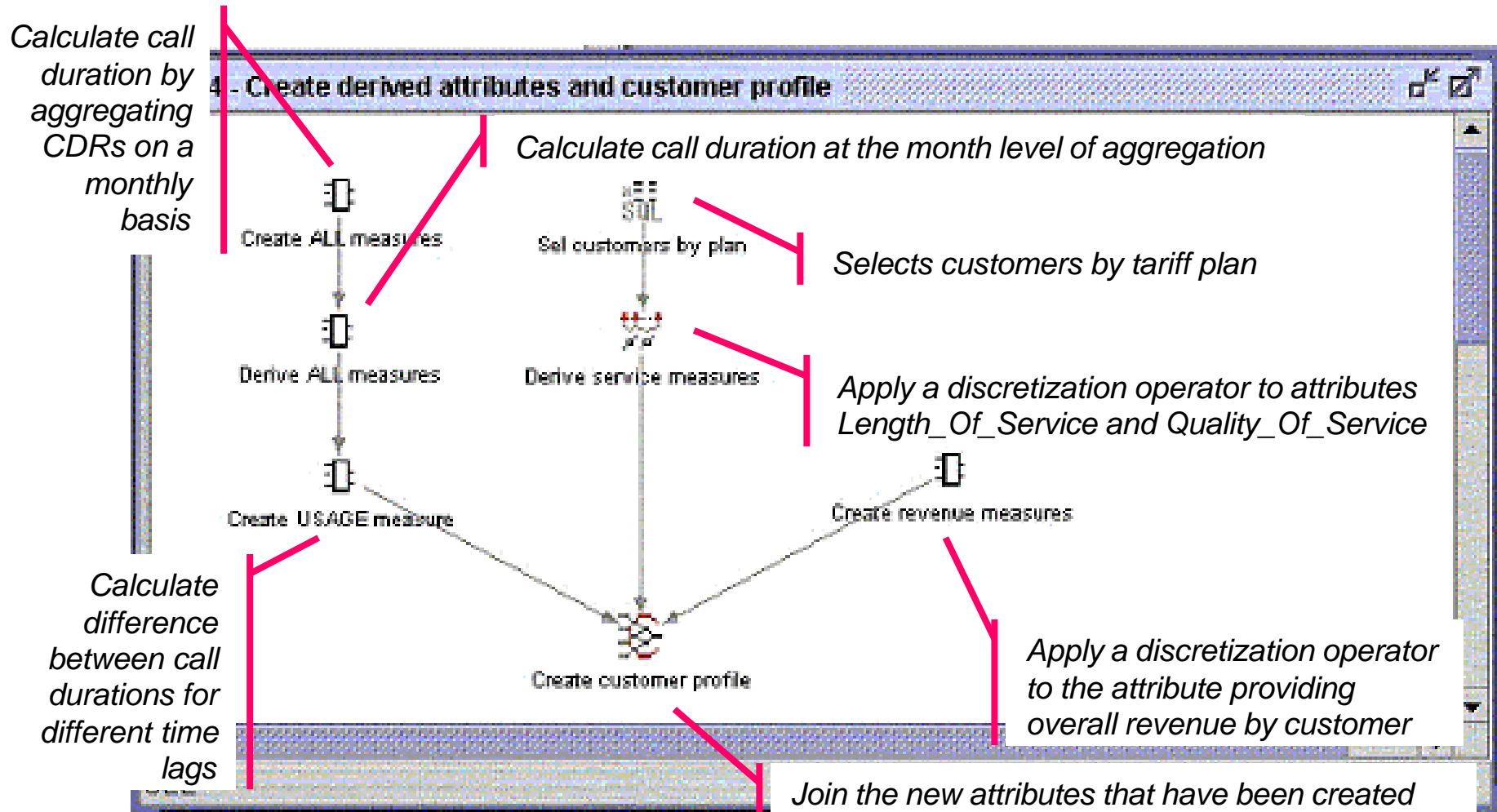
Select revenue records associated with calls originated in a given month (from M1 to M5)

Convert revenue records from a transactional form into a relational one

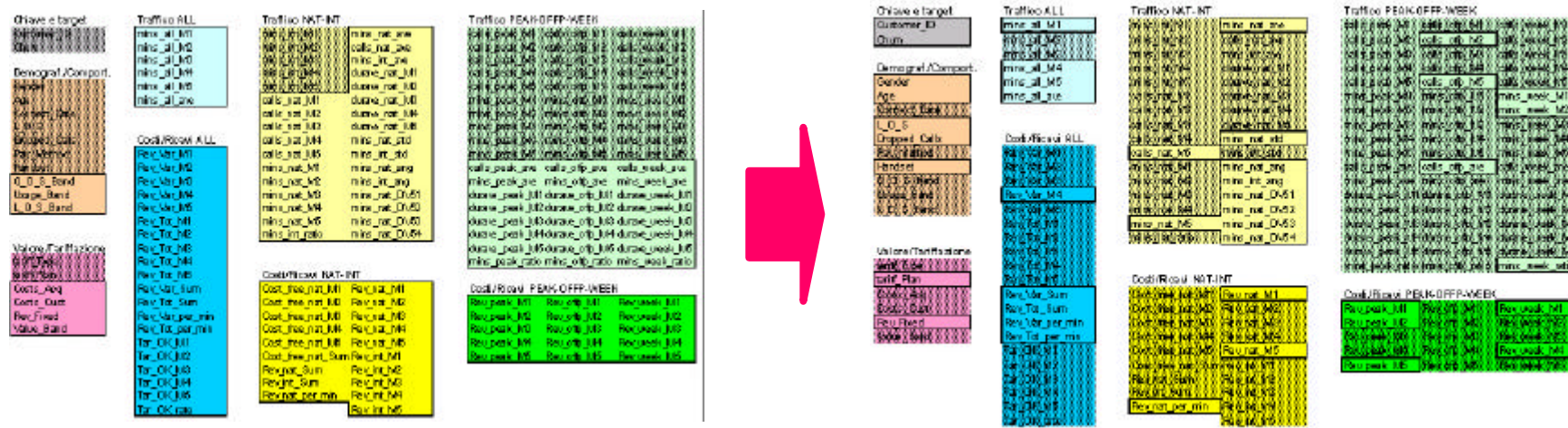
Add a new attribute that sums up the revenue of calls originated from month M1 to month M5

Save revenue records by joining revenue records in relational form and customer records by customer key

Create derived attributes and customer profile

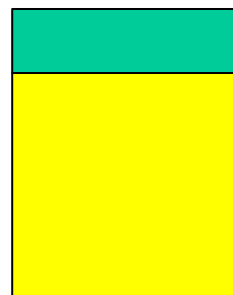


Construction stage output



Data Construction

Feature Selection

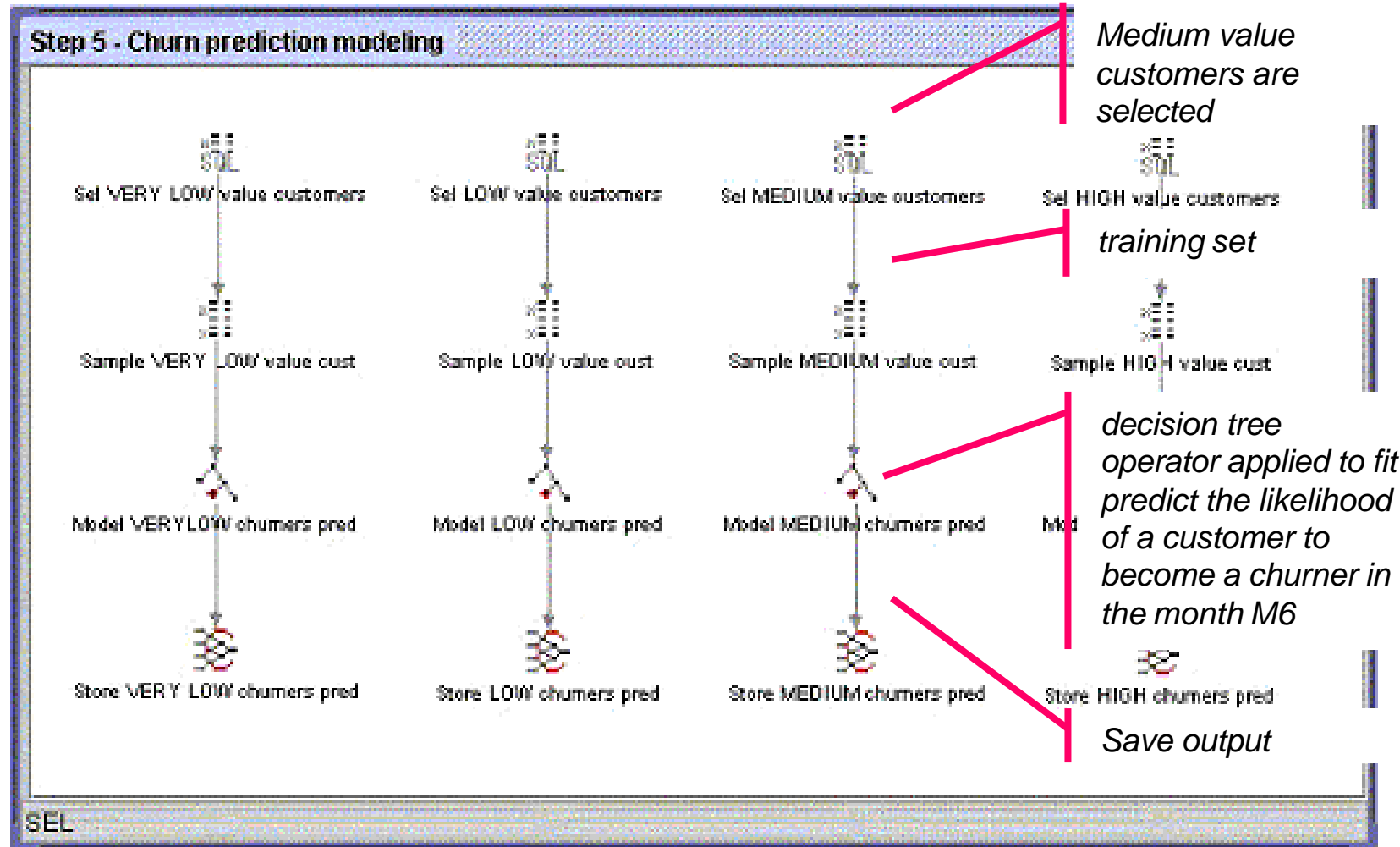


16 Raw attributes

45 Derived attributes

Churn modeling chain

4 Predictive models,
one for each
customer segment



The resulting model

The screenshot displays a database schema browser for 'MMART_BD@TLAB.CSELT.IT'. The 'Functions' folder is expanded, showing a list of functions including 'DT_100002845_CHURN_1_FCT', which is selected. The main window shows the PL/SQL code for this function, which is a complex conditional logic for predicting churn state based on various metrics like MINS_NAT_DV51, MINS_NAT_ANG, MINS_INT_M4_IS, HANDSET, REV_FIXED, DURAVE_NAT_M4, DURAVE_NAT_M1, MINS_INT_AVE, MINS_NAT_DV52, MINS_INT_AVE, MINS_NAT_M1_IS, and HANDSET.

Below the code, a window titled 'View data for mmart_bd.CS_100002841_3 (991 rows)' shows the execution results. The data is presented in a table with the following columns: CUSTOMER_ID, GENDER, AGE, HANDSET, REV_NAT_SUM, REV_INT_SUM, MINS_INT_M1_LS, CHURN, and CHURN_STATE_PRED.

CUSTOMER_ID	GENDER	AGE	HANDSET	REV_NAT_SUM	REV_INT_SUM	MINS_INT_M1_LS	CHURN	CHURN_STATE_PRED
K119125	M	45	BS110	0	2.83629529	12.63562520196	CHURNED	CHURNED
K119133	M	30	S50	6.463874901	14.18148054	9.662545407712	CHURNED	CHURNED
K119137	F	43	BS110	7.063034084	5.224260478	34.75469890578	ACTIVE	CHURNED
K119184	F	13	BS110	2.248565661	26.08142769	28.70986103953	CHURNED	CHURNED
K119195	F	22	CAS30	1.235638085	6.328907671	1.440765316548	CHURNED	CHURNED
K119248	F	44	BS110	0	31.59619603	0.306711193056	ACTIVE	ACTIVE
K119300	M	44	BS110	7.209183717	10.02681376	19.05957740729	CHURNED	CHURNED
K119353	F	22	CAS30	0	50.05748082	2.422178630615	CHURNED	CHURNED
K119354	F	37	BS110	0	5.060881659	4.618084760024	ACTIVE	CHURNED
K119366	M	16	CAS30	0	25.36608805	8.006565004430	CHURNED	ACTIVE
K119397	F	16	BS110	4.429011292	13.69428609	8.721588400716	CHURNED	CHURNED
K119409	M	34	SOP20	0	68.34176719	5.793478623954	CHURNED	CHURNED
K119448	F	27	BS110	0	9.239293953	34.43225233958	ACTIVE	CHURNED
K119478	M	63	BS110	0	46.12988683	10.88935309473	CHURNED	CHURNED
K119537	F	47	S50	2.067208396	2.381026472	1.768965908566	CHURNED	CHURNED
K119658	M	26	CAS60	12.909804112	11.93698689	16.14004740886	ACTIVE	ACTIVE
K119683	F	23	CAS30	3.935751905	6.968798472	13.98927541817	CHURNED	CHURNED
K119689	F	28	BS110	0	34.88081486	8.971911281864	ACTIVE	CHURNED
K119670	M	30	S50	0	13.18055217	5.061986871416	ACTIVE	ACTIVE
K119752	F	12	MC85	0	65.80230336	2.390545445222	ACTIVE	ACTIVE
K119828	M	28	S50	0	12.29398487	5.414179868706	CHURNED	CHURNED
K119972	M	53	SOP10	7.094335433	61.37452005	7.837699178657	CHURNED	CHURNED
K119975	F	34	BS110	0	12.23959893	7.130817519913	ACTIVE	ACTIVE
K119981	M	37	S50	6.893898546	61.08607817	8.727065210481	CHURNED	CHURNED
K119996	F	12	CAS30	1.36633408	7.460340289	8.301669234075	CHURNED	CHURNED
K120067	M	43	BS110	11.36479682	37.24471686	0.613685317199	CHURNED	CHURNED
K120097	F	26	BS110	4.002053298	1.060680081	4.877163974607	CHURNED	CHURNED
K120111	F	26	BS110	0	5.296896335	9.096747686869	ACTIVE	CHURNED
K120128	M	43	S50	0	1.873781791	20.34914884600	CHURNED	CHURNED
K120138	M	33	SOP20	0	8.354436533	15.93758728468	CHURNED	CHURNED

The decision tree - excerpt

BEGIN

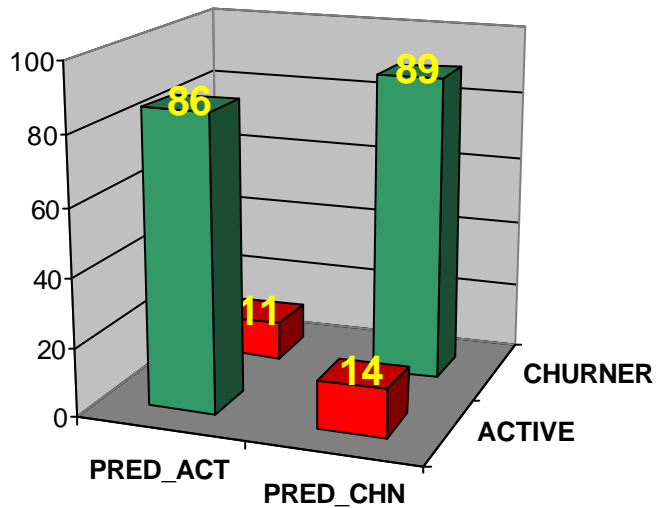
```

if ALL_M5 <= 483.526001 then
  if HANDSET = 'ASAD1' then
    return 'ACTIVE';
  elsif HANDSET = 'ASAD9' then
    if PEAK_M1 <= 139.363846 then
      if OFFP_M3 <= 106.607796 then
        return 'ACTIVE';
      else
        return 'CHURNED';
      end if;
    else
      return 'CHURNED';
    end if;
  elsif HANDSET = 'S50' then
    if PEAK_M3 <= 144.418304 then
      return 'CHURNED';
    else
      if REV_SUM <= 294.393341 then
        if L_O_S_BAND = 'HIGH' then
          return 'ACTIVE';
        elsif L_O_S_BAND = 'MEDIUM' then
          return 'ACTIVE';
        end if;
      end if;
    end if;
  end if;
end if;

```

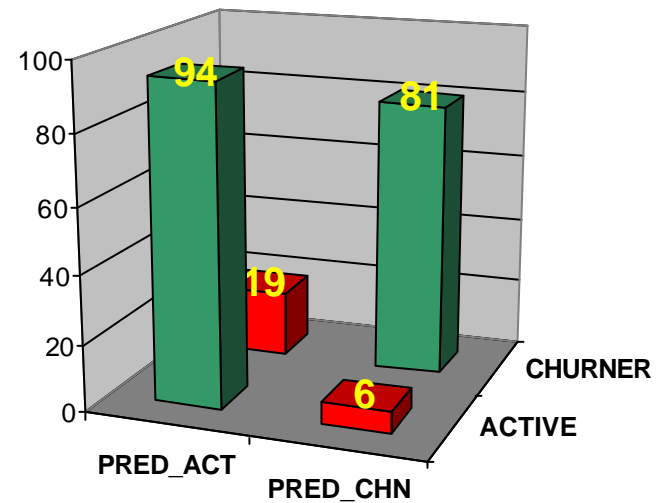
Predictive performance

MEDIUM customer model performance

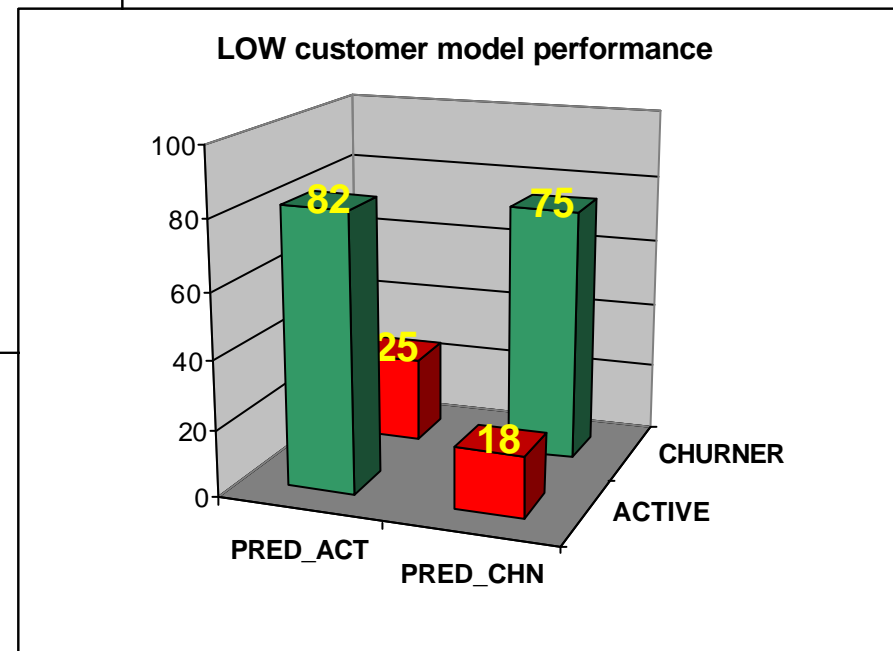
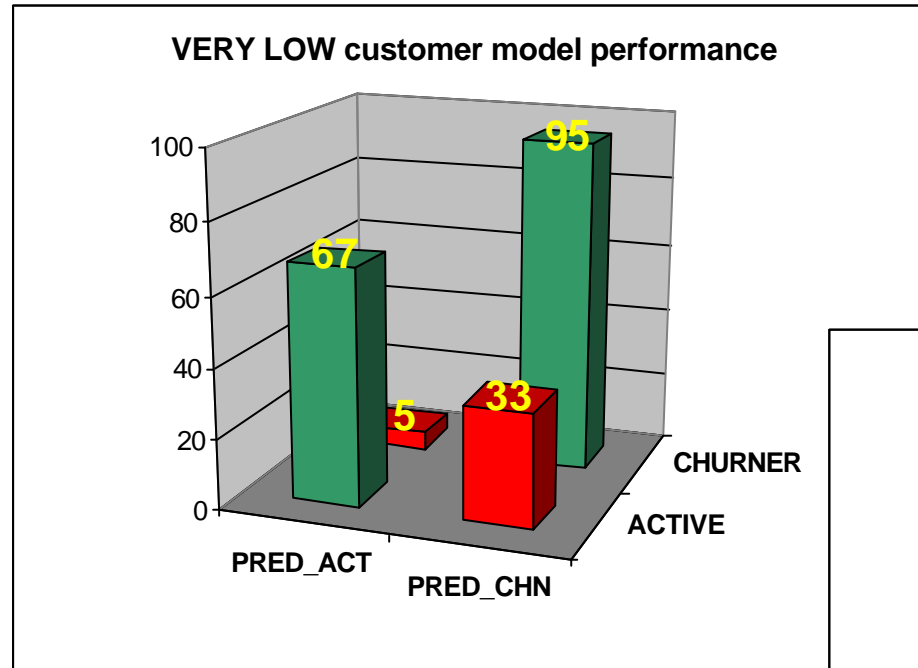


Training / test set: 70% / 30%

HIGH customer model performance

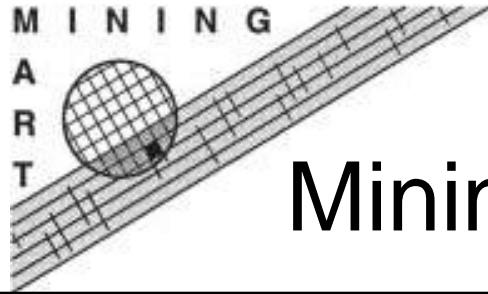


Predictive performance



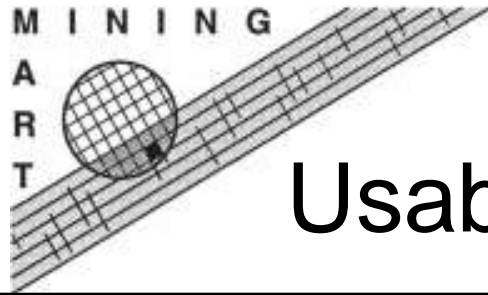
Execution Time

Data Set Size (num. records)	Pre-processing Time (mins)	Modeling Time (hours)
8,000	17.3	4.3
800,000	27.8	13.5



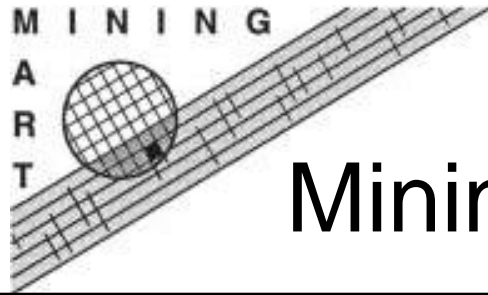
Mining Mart evaluation

- **Usability**
- **Mining process speed-up**
- **Mining process quality**
- **Integration (into the business processes)**



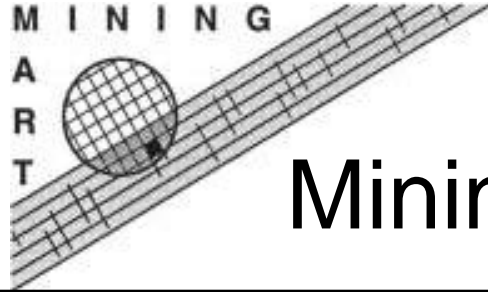
Usability

- **Human Computer Interface is user-friendly and effective. Few steps required to implement any data mining process**
- **Interface quality compares to the ones of leading commercial tools (SPSS, SAS). Improves on IBM Intelligent Miner's interface with respect to a number of features**
- *Suggestions for future work*
 - **Definition of concepts can be further simplified (db attributes defined by directly editing table column names)**



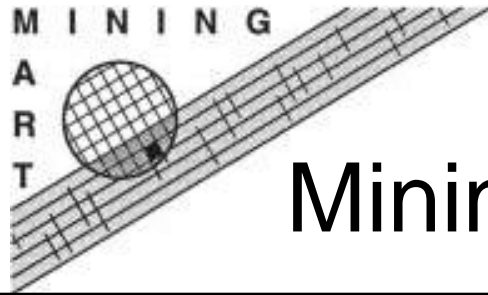
Mining process speed-up

- **Preprocessing operators show quite good scalability on large data set:**
 - **MMart leverages Oracle scalability when carrying out preprocessing tasks. Overhead due to parsing of operators is negligible (unless for very small datasets)**
 - **Modeling operators are not optimized**
- **Processing chains can be quickly tested during chain set-up**
- **Multistep and loopable operators enable users to define parallel mining tasks consistently and effectively**
- **Processing chains can be saved and restored, allowing versioning**



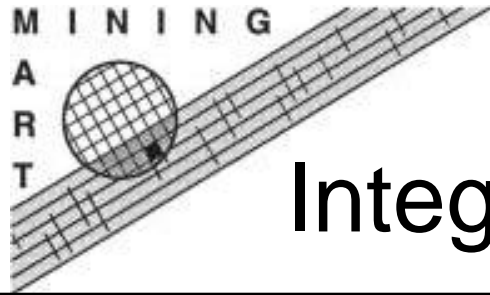
Mining process speed-up

- **Less trials required to develop the data mining solution**
 - **Operator constraints drive unskilled users to build correct and effective analytical applications**
 - **Users achieve a better understanding of data structure by:**
 - Browsing source and processed data
 - Computing descriptive statistics
 - **Operator chains makes it possible to implement data mining best-practices**
- *Suggestions for future work*
 - **Improve graphical investigation features**
 - **Improve workgroup enabling features: multiple users capabilities, definition of user roles and access rights**



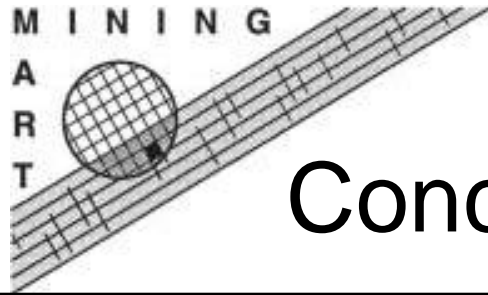
Mining process quality

- **Best practices may be easily pre-packaged**
- **Libraries of data mining applications may be developed and customized to satisfy new business requirements**
- **MMart framework ensures chain consistence and correctness, avoiding potential conceptual mistakes**
- **Users can focus their effort on modeling tasks rather than on preprocessing tasks**
- **Domain knowledge improves and extend usability of pre-packaged data mining applications**



Integration

- **The Mining Mart system may be integrated into the Analytical CRM platform as the analytical extension of either the enterprise data warehouse or the business-oriented data marts**



Conclusions

- **Speed up for some preprocessing tasks increased by 50% at least**
- **Power users may find Mining Mart as much easy to use as the leading commercial dm platforms**
- **It enables building libraries of predefined data mining applications that can be easily modified**
- **MMart guarantees the highest scalability, since it exploits leading commercial db tools features**
- **Quality of data mining output increases as the number of preprocessing trials decrease in number**
- **Bottom line: Mining Mart supports efficiently and effectively the preprocessing stage of a data mining process**