

# AIR MILES

un caso di studio di  
customer segmentation

G. Saarenvirta, "Mining customer data", DB2  
magazine on line, 1998

<http://www.db2mag.com/98fsaar.html>

# Clustering & segmentazione dei clienti

- ◆ Obiettivo: analizzare i dati di acquisto dei clienti per
  - Comprendere i comportamenti di acquisto
  - Creare strategie di business
  - Mediante la suddivisione dei clienti in **segmenti** sulla base di variabili di valore economico:
    - ◆ volume di spesa
    - ◆ margine
    - ◆ frequenza di spesa
    - ◆ "recency" di spesa (distanza delle spese più recenti)
    - ◆ misure di rischio di defezione (perdita del cliente, churn)

# Segmenti

## ◆ Clienti **high-profit, high-value, e low-risk**

- In genere costituiscono dal 10% al 20% dei clienti e creano dal 50% all'80% del margine
- Strategia per il segmento: **ritenzione!**

## ◆ Clienti **low-profit, high-value, e low-risk**

- Strategia per il segmento: **cross-selling** (portare questi clienti ad acquistare altri prodotti a maggior margine)

# Segmenti di comportamento di acquisto

- ◆ All'interno dei segmenti di comportamento di acquisto, si possono creare sottosegmenti demografici.
- ◆ I dati demografici non sono usati, di solito, insieme a quelli economici per creare i segmenti
- ◆ I sottosegmenti demografici invece usati per scegliere appropriate **tattiche** (pubblicità, canali di marketing, campagne) per implementare le **strategie** identificate a livello di segmenti.

# The Loyalty Group in Canada

- ◆ Gestisce lo AIR MILES Reward Program (AMRP) per conto di più 150 compagnie in tutti i settori - finanza, credit card, retail, gas, telecom, ...
- ◆ coinvolge il 60% delle famiglie canadesi
- ◆ è un programma **frequent-shopper**:
  - Il consumatore accumula punti che può redimere con premi (biglietti aerei, hotel, autonoleggio, biglietti per spettacoli o eventi sportivi, ...)

# Acquisizione dei dati

- ◆ Le compagnie partner catturano i dati di acquisto e li trasmettono a The Loyalty Group, che
- ◆ immagazzina le transazioni in un DW e usa i dati per iniziative di marketing, oltre che per la gestione dei premi.
- ◆ Il DW di The Loyalty Group conteneva (al 2000)
  - circa 6.3 milioni di clienti
  - circa un 1 miliardo di transazioni

# Stato dell'arte prima del data mining

- ◆ The Loyalty Group impiega tecniche analitiche standard per la segmentazione dei clienti
  - Recency, Frequency, Monetary value (RFM) analysis
- ◆ In sostanza, un modello fatto di regole generali che vengono imposte ai dati per creare i segmenti
- ◆ Analogo delle regole di classificazione dei soci Unicoop:
  - Socio costante: ha fatto almeno 2 spese al mese per almeno 3 degli ultimi 4 mesi

# Una esperienza di Data mining

## ◆ Obiettivo:

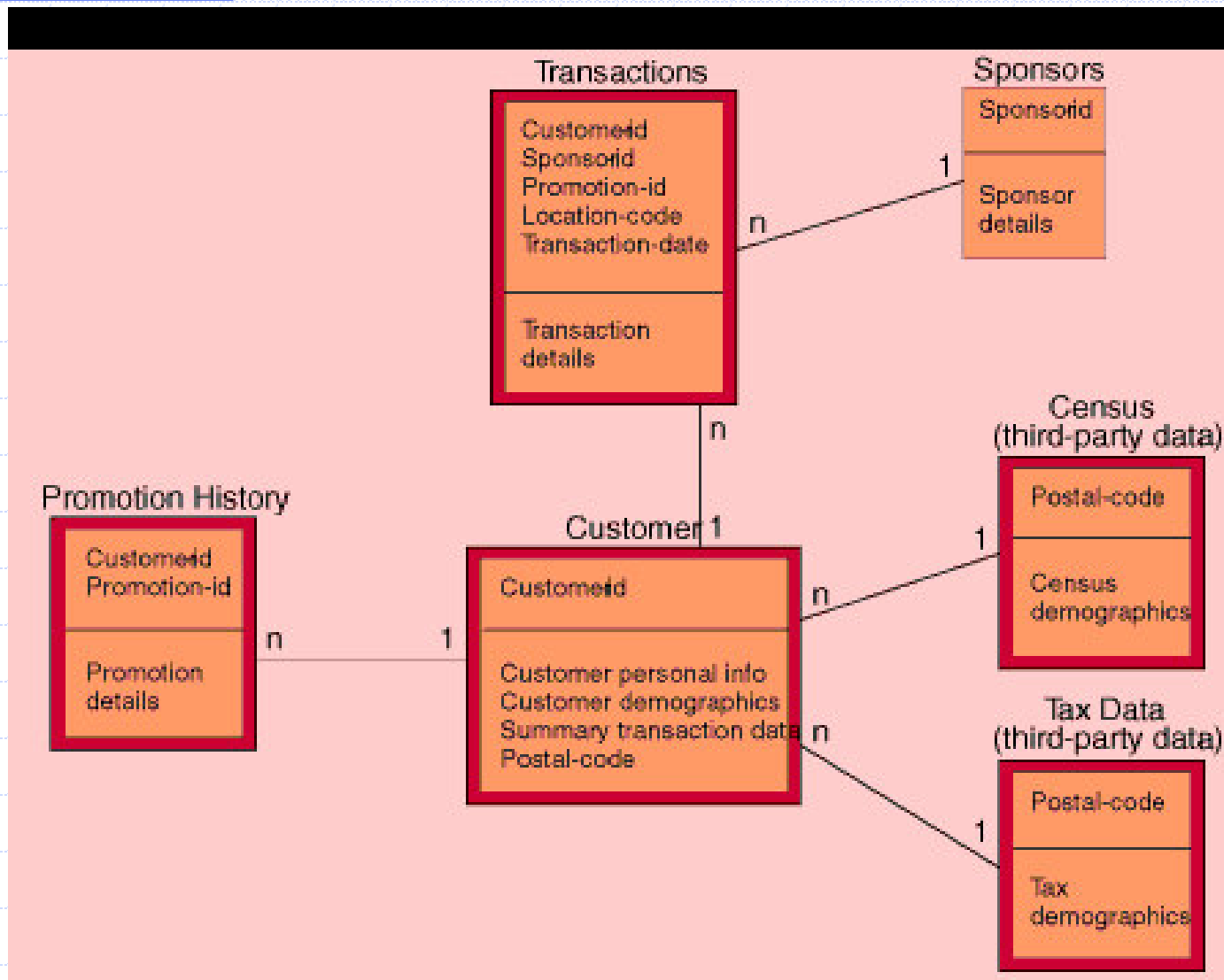
- creare una segmentazione dei clienti
- a partire dai dati su clienti e loro acquisti nel DW
- usando il **clustering**, una tecnica di data mining
- e confrontare i risultati con la segmentazione esistente sviluppata con l'analisi RFM.

◆ ... lasciare che **i segmenti emergano direttamente dai comportamenti di acquisto simili effettivamente riscontrati nella realtà**, senza imporre un modello preconfezionato ...

◆ ... e vedere che succede!



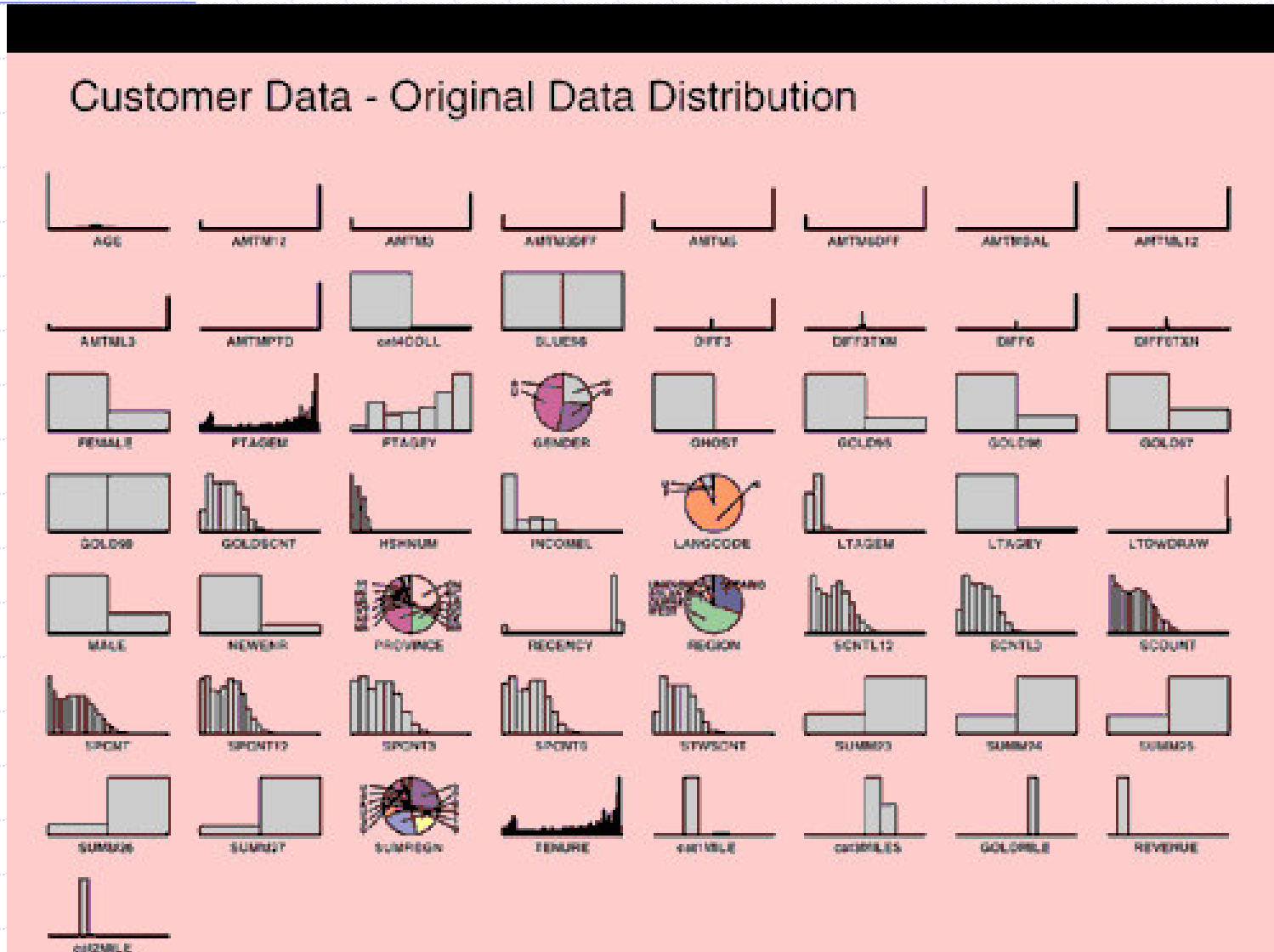
# Sorgente dei dati nel DW



# Preparazione dei dati

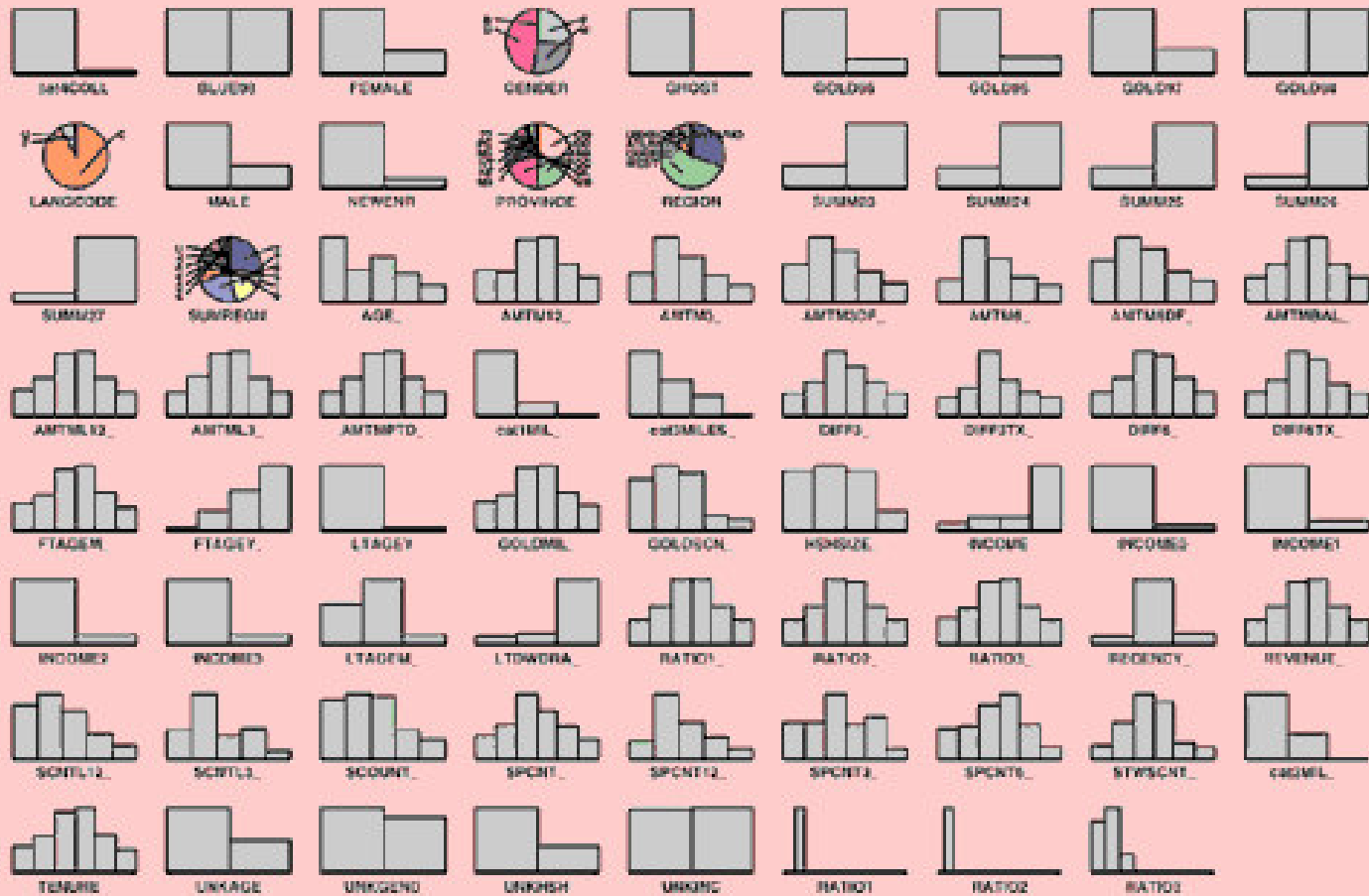
- ◆ Creazione delle variabili economiche di ciascun **cliente**, mediante aggregazione dei propri acquisti
  - Volume di spesa
  - Durata del suo ciclo di vita
  - Numero di compagnie sponsor in cui ha acquistato
  - Numero di compagnie sponsor in cui ha acquistato negli ultimi 12 mesi
  - Distanza (in mesi) dall'ultimo acquisto
  - ...
- ◆ Circa 100 variabili economiche derivate dai dati di acquisto nel DW!

# I dolori della pulizia dei dati: prima ...



# .... e dopo la cura

## Customer Data - Discretized



# Prima e dopo la cura

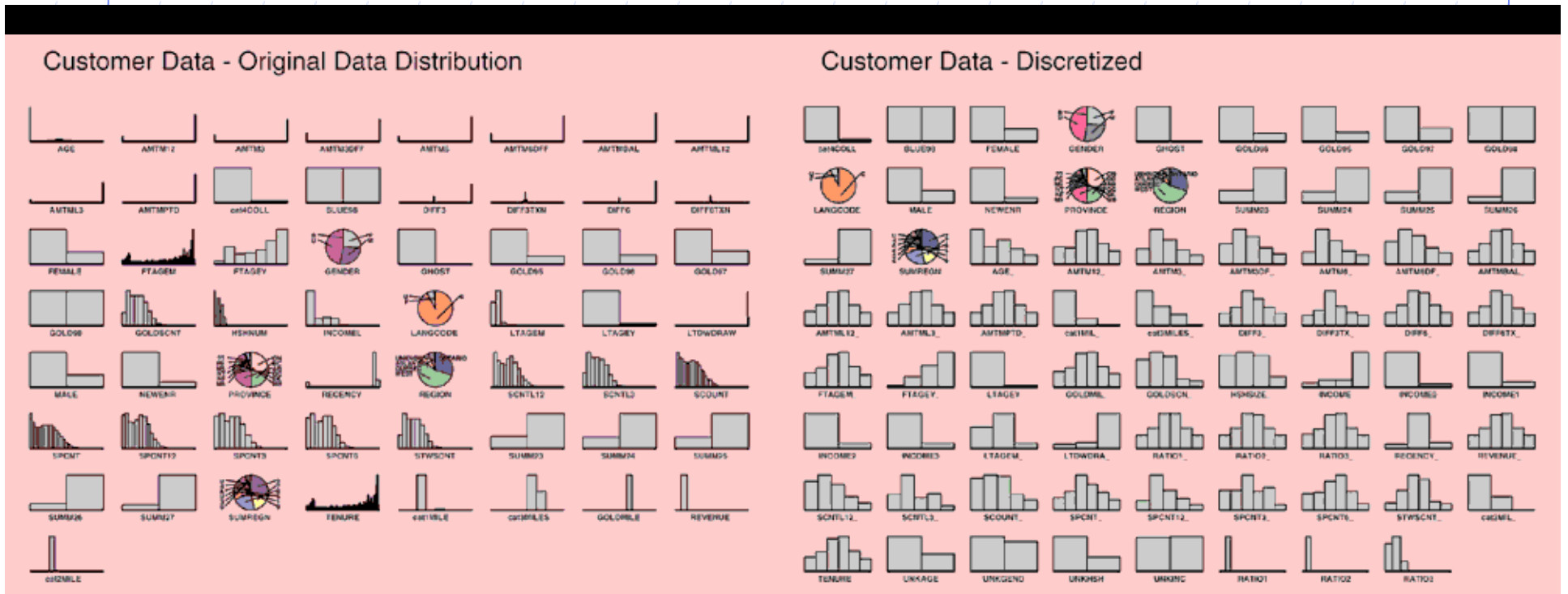
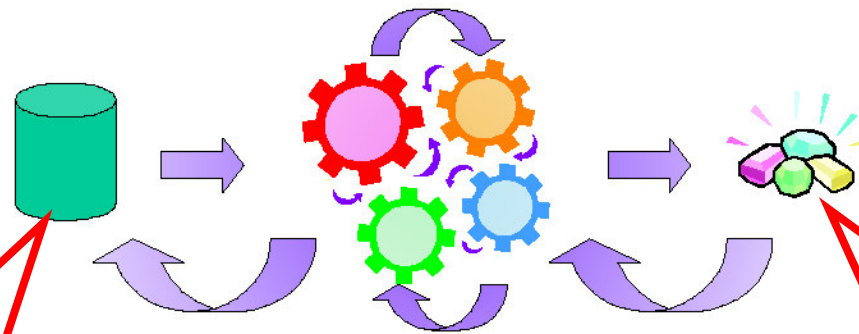


Figure3. Original data.

Figure 4. Discretized data.

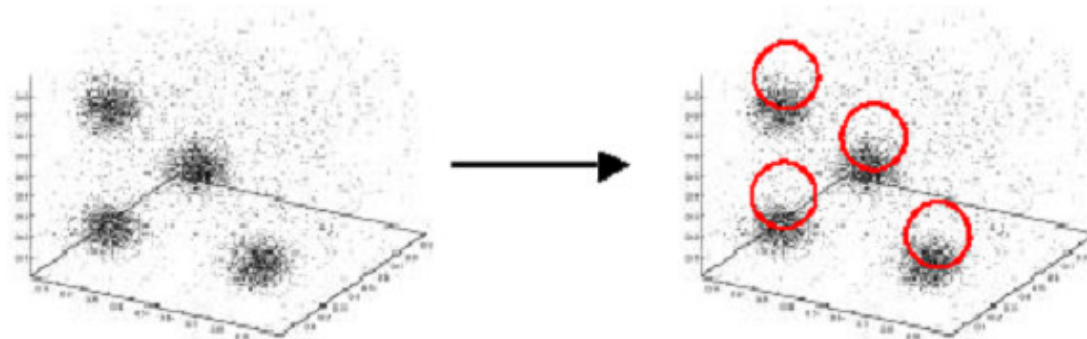
# Estrazione del modello di clustering

Clustering = raggruppamento di oggetti simili in gruppi omogenei



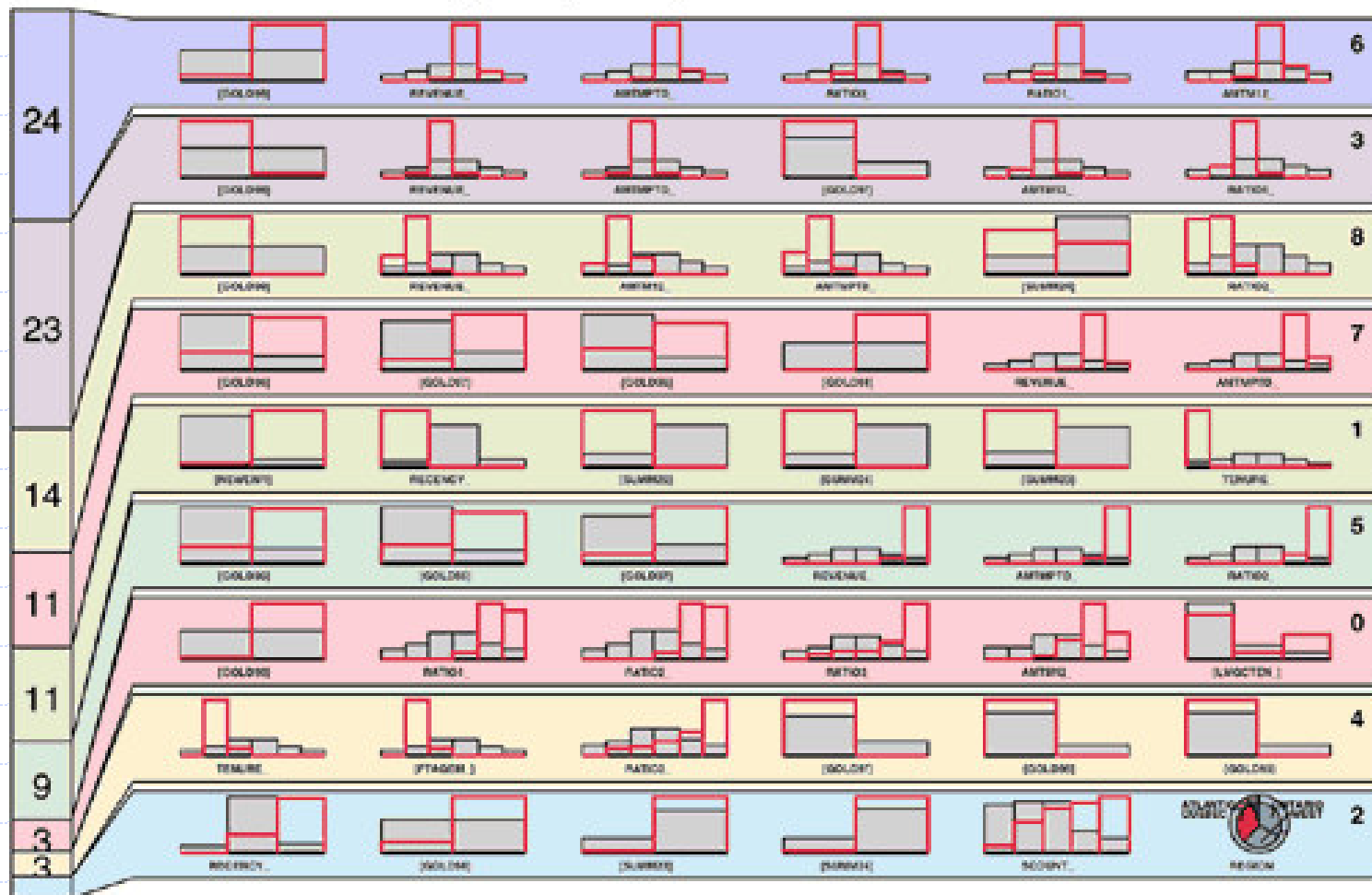
Dati in input:  
variabili economiche  
di ciascun cliente

Modello in output:  
segmentazione dei  
clienti simili in clusters



# Output del clustering

Customer Clustering(DG) - Layer 1



# Analisi qualitativa dei cluster

- ◆ La variabile **Gold98** indica se il cliente è o meno uno migliore clienti, secondo la segmentazione preesistente creata con le tecniche RFM.
- ◆ Nel clustering non viene usata: serve solo a "spiegare" i clienti del cluster.
- ◆ Il modello di clustering conferma la definizione esistente: tutti i cluster hanno quasi tutti clienti Gold oppure non Gold.



# Analisi qualitativa dei cluster

◆ Ma il risultato non si limita a validare il concetto esistente di cliente Gold:

- Crea un sottosegmento dei clienti Gold, raffinando la conoscenza preesistente
- In pratica, è stato scoperto un sottosegmento di clienti **Platinum**

## ◆ **Cluster 5**

- Quasi tutti clienti Gold98, con molte variabili economiche nei percentili alti

# Analisi del cluster 5 – clienti Platinum

- ◆ 9 % della popolazione
- ◆ volume di spesa totale e mensile, durata, punti redenti, ... sono tutti al di sopra del 75esimo percentile, alcuni addirittura sopra il 90esimo
- ◆ Mette in luce un segmento di clienti molto redditizio

# Vista dettagliata del cluster 5

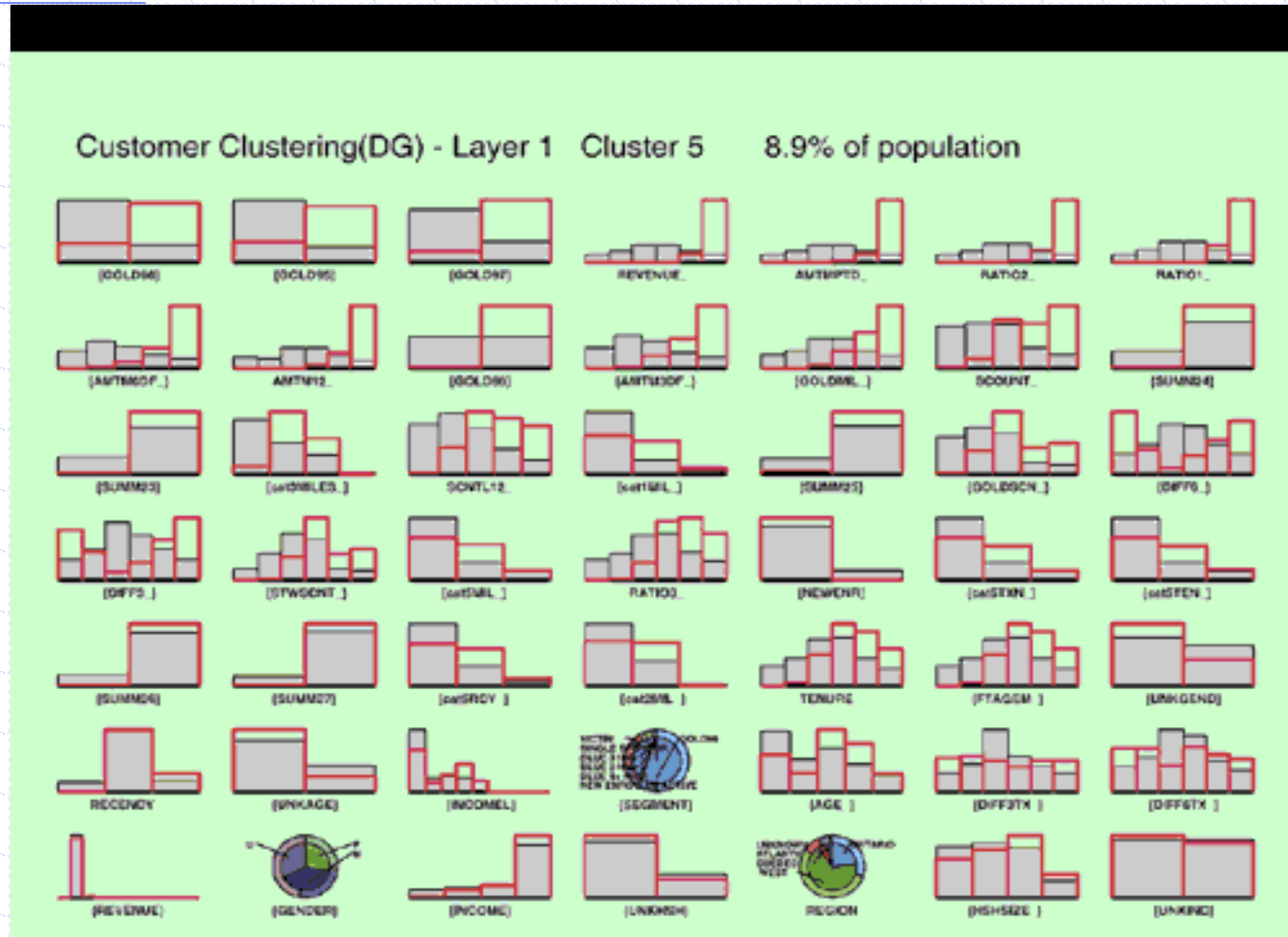


Figure 8. Cluster 5 output.

# Analisi dei cluster

- ◆ Obiettivo: un rapporto che valuti quantitativamente il valore potenziale dei cluster trovati mediante indicatori calcolati per aggregazione sui clienti di ciascun cluster.

CLUSTERID	REVENUE	CUSTOMERS	PRODUCT INDEX	LEVERAGE	TENURE
5	34.74%	8.82%	1.77	3.94	60.92
6	26.13%	23.47%	1.41	1.11	57.87
7	21.25%	10.71%	1.64	1.98	63.52
3	6.62%	23.32%	.73	.28	47.23
0	4.78%	3.43%	1.45	1.40	31.34
2	4.40%	2.51%	1.46	1.75	61.38
4	1.41%	2.96%	.99	.48	20.10
8	.45%	14.14%	.36	.03	30.01
1	.22%	10.64%	.00	.02	4.66

**Table 1.** *Profiling a cluster.*

# Analisi dei cluster

- ◆ **leverage** = rapporto fra
  - *revenue* (ricavo) e
  - popolazione del cluster.
- ◆ Il cluster 5 il più redditizio.
- ◆ **product index** = rapporto fra
  - numero medio di prodotti acquistati dai clienti del cluster e
  - numero medio di prodotti acquistati dai clienti in generale
- ◆ La redditività del cliente aumenta con la *tenure* (durata)
- ◆ NOTA: questa non è altro che analisi OLAP con la nuova dimensione della segmentazione appena scoperta!!

# Opportunità di business

- ◆ Migliori clienti (clusters 2, 5 e 7):
  - indicazione: **ritenzione!!**
- ◆ Clusters 6 e 0
  - indicazione: **cross-selling**
  - Goal: cercare di convertire i clienti dei clusters 6 e 0 ai clusters 2, 5 o 7.
  - Si può procedere a studiare quali siano i prodotti maggiormente acquistati nei vari clusters per trovare prodotti candidati al cross-selling ...

# Opportunità di business (2)

## ◆ Clusters 3 e 4

- indicazione: **cross-selling** verso i clusters 2, 6 e 0

## ◆ Cluster 1

- indicazione: **attendere**, potrebbe essere un nuovo segmento di clienti

## ◆ Cluster 8

- indicazione: **nessun investimento** di marketing (maledetti cherry-peakers!)

# Una buona pratica di mining

- ◆ Reazioni di The Loyalty Group ai risultati del progetto
  - La visualizzazione dei risultati supporta un livello di analisi significativa e utile alle decisioni.
  - La segmentazione preesistente viene confermata, ma anche raffinata attraverso sottosegmenti sconosciuti a priori, e potenzialmente utili e proficui.
  - Decisione di intraprendere nuovi progetti di mining:
    - ◆ Messa a regime della segmentazione usando clustering su dati più completi sui comportamenti di acquisto,
    - ◆ Modelli predittivi per **direct mail targeting**,
    - ◆ Identificazione di opportunità di cross selling usando **regole di associazione frequenti** nei segmenti scoperti.