

SAS Enterprise Miner™

A horizontal bar with a black background and a yellow-to-white gradient on the left side, containing the text 'SAS Enterprise Miner' in a stylized font.

# Datenanalyse und Data Mining mit der SAS Software

Reinhard Strüby  
SAS Institute Heidelberg



# Inhalt

**Warum? - Data Mining und Anforderungen**

**Was? - Data Mining Definition**

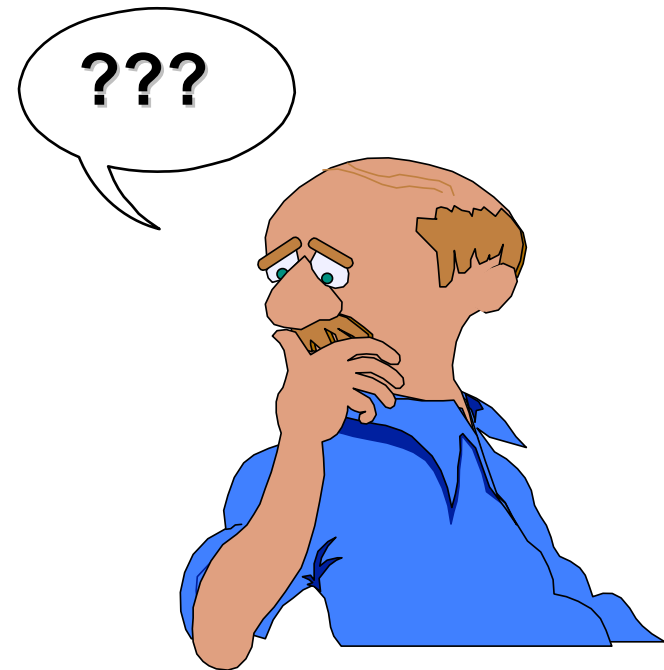
**Wer? - Anwendergruppen**

**Wie? - Erfolgsfaktoren für Data Mining**

**Wie? - Der SEMMA Prozess**

**SAS! - Die SAS Data Mining Lösung**

# DATA MINING ?



# Die Geschäftsaufgabe

- x “Kenne Deine Kunden!”
- x Wer sind sie?
- x Was wünschen sie?
- x Welche Kontakte gab es bisher?
- x Wie kann eine dauerhafte Beziehung hergestellt werden?
- x Welche Kunden könnten uns verlassen?



# Data Mining - warum jetzt?

- x Erhöhter Wettbewerbsdruck
- x Geringere Speicherkosten und höhere Rechengeschwindigkeiten
- x Data Warehouses oft vorhanden
- x Versteckte Informationen in großen Dateien
- x Data Mining - Methoden finden Muster
- x GUI Data Mining Anwendungen
- x Kundendruck auf Veränderungen
- x ROI erhöhen



# Data Mining Definition

- x Data Mining ist der Prozess des Selektierens, Erklärens und Modellierens
- x großer Datenmengen,
- x um bisher unbekannte Datenmuster für einen Geschäftsvorteil zu nutzen.



# DATA MINING

- x Data Mining ist ein Prozess.
- x Data Mining beinhaltet die enge Kooperation von IT, Fachabteilung und Data Minern.
- x Data Mining ist nicht beschränkt auf bestimmte Industriezweige oder Probleme.



# DATA MINING - INDUSTRIES

## General

Customer Segmentation  
Targeted/cross marketing  
Pricing Analysis  
Associations &  
Demography

## Insurance & Health Care

Claim Analysis  
Fraudulent Behavior

## Banking

Credit Authorization  
Credit Card Fraud  
Detection Portfolio Analysis

Cash Planning

## Telecommunications

Call Behaviour Analysis  
Churn Management

## Retail/Marketing

Market Basket Analysis  
Database Marketing  
Category Management

## Production and Utilities

Process Management  
Demand Patterns  
Capacity Planning  
Inventory Planning





# IS DATA MINING IMPORTANT?



Postbank N.V.

“50% response on first mailing payed for DM investment”

US West

“Reducing customer churn by any amount is 10 times cheaper than gaining a new customer”

ABN AMRO

“Interest earned on 40% reduction in cash in ATMs”

Neckermann Versand AG

“Increased number of good customers getting credit by 80 a day”

Gloucestershire Constabulary

“For the public, increased crime pattern identification and prevention is priceless”

# DATA MINING - Nutzer

- x Leiter von Fachabteilungen / Spezialisten
- x Data Miner



## Nutzer: Fachabteilung

- x Kennen das Fachgebiet
- x Verstehen die Inhalte der Daten
- x Suchen nach Informationen, haben aber oft geringe analytische Kenntnisse
- x Arbeiten häufig in Marketing-Abteilungen als Analyst

## Nutzer: DATA MINER

- x Quantitative Experten: statistischer/mathematischer Background oder vergleichbare Kenntnisse
- x Etwas isoliert von Geschäftsfragen
- x Vertraut mit Algorithmen und Datenanalyse-Prozess
- x Häufig im Finanzsektor, sonst eher selten

## THE DATA MINING MARKET - IN MILL \$

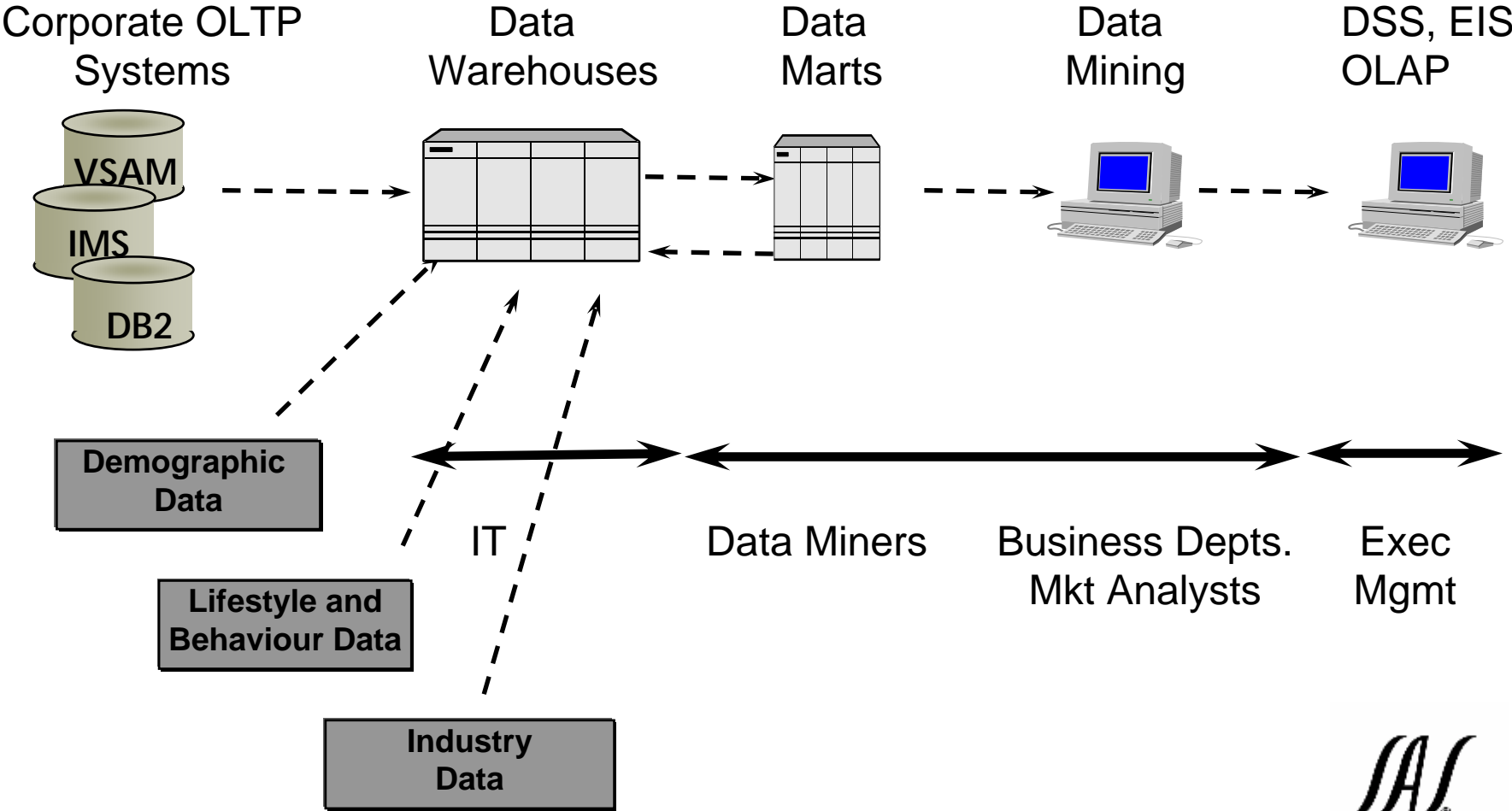
	1996	1997	1998	1999	2000
<b>Horizontal Apps*</b>	22	36	58	106	170
<b>Vertical Apps</b>	110	191	320	601	961
<b>Macro Mining*</b>	184	257	360	486	655
<b>Micro Mining*</b>	22	40	60	90	135
<b>Data Visualiz.*</b>	110	133	145	158	160

Source: META Group, Data Mining Market Trends 1997-1998

\* SAS System mentioned in this Category.



# THIRD GENERATION DATA MINING -Integrated



**Data Warehousing / Data Mining integrated**



## Erfolgsfaktoren

- x Zugriff auf alle Datenquellen - Data Warehousing
- x Skalierbarkeit: HW / SW
- x Breites Spektrum von DM Methoden: Konzentration auf Geschäftsprobleme
- x Strategie der Implementation

# Vergleich OLAP gegen Data Mining

## OLAP, Report Writing

Nutzergesteuertes  
Reporting -  
Dimensionen bekannt

Bestverkauftes Produkt  
im Jahr 1997  
in der Region X ?



## Data Mining Methodology

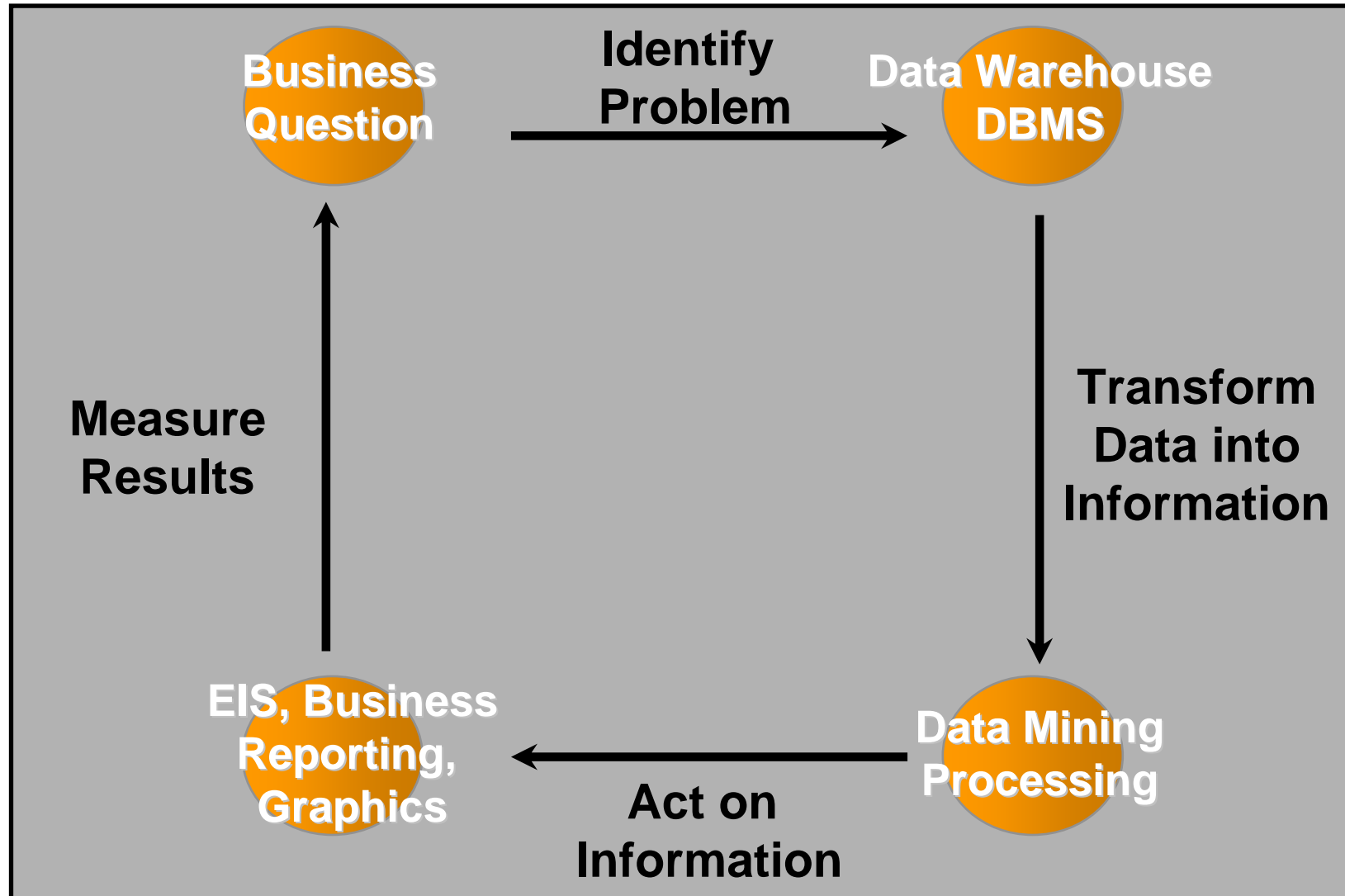
Datengesteuerte  
Exploration -  
Suche nach Dimensionen

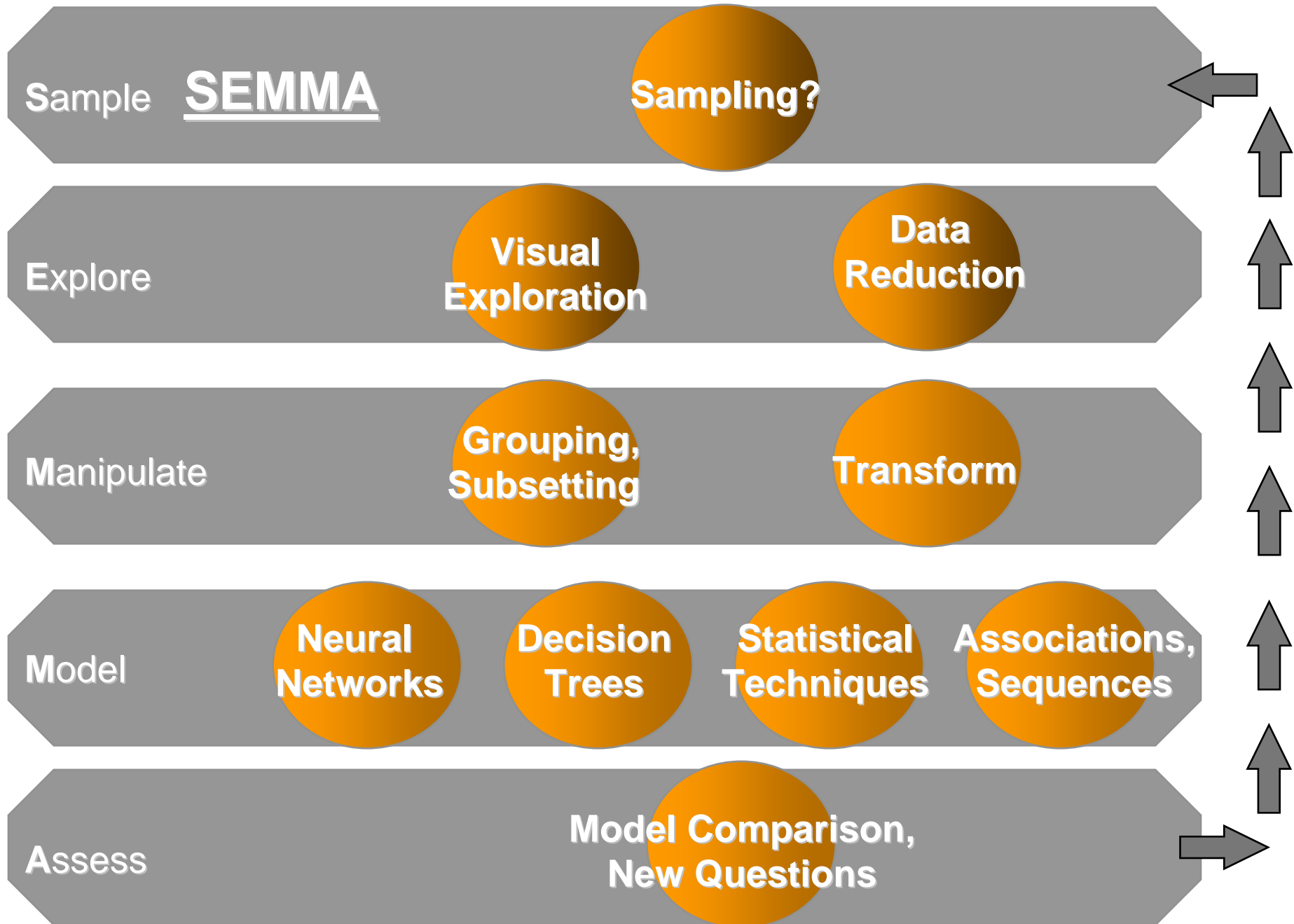
Auf welche Kunden  
sollten wir uns  
konzentrieren ?



# SAS DATA MINING SOLUTION

## Data Mining, IT and Business

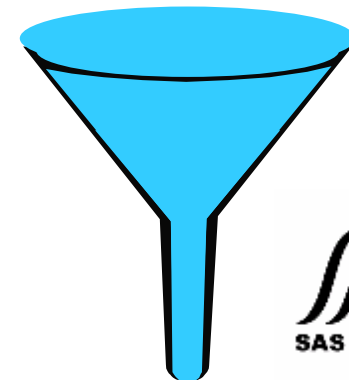




# SAMPLING ?

Empfohlen, nicht Voraussetzung:

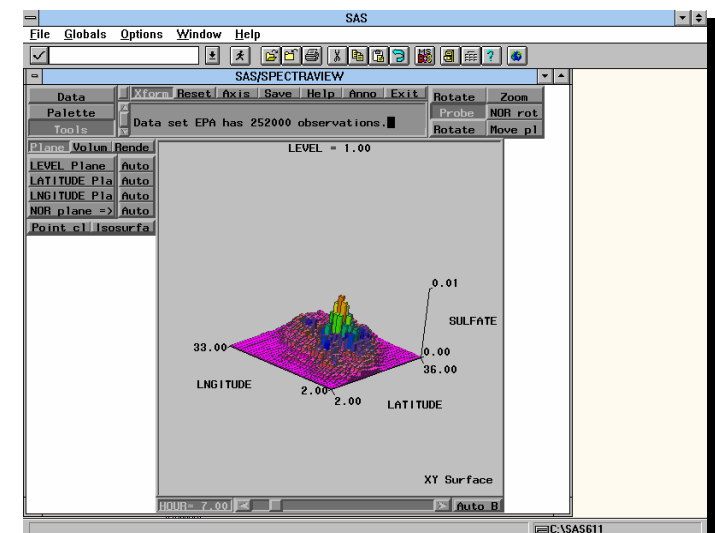
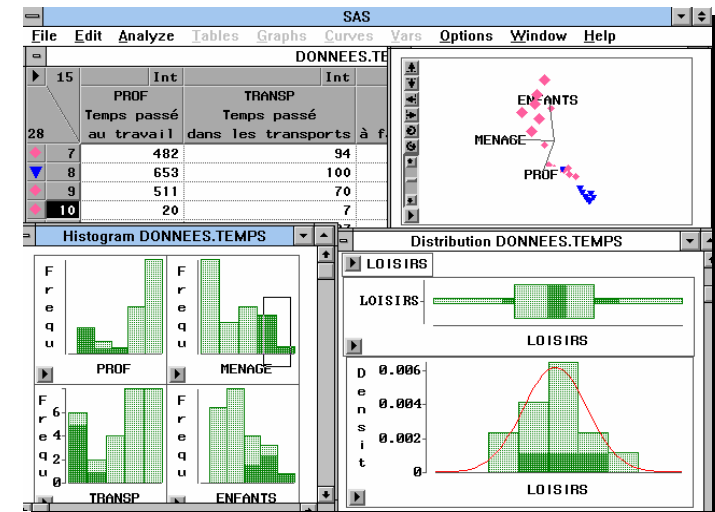
- x Inhalte gehen nicht verloren.
- x Erhebliche Performance Vorteile
- x Modellprüfung: Training, Testing, Validation Samples



# EXPLORATION

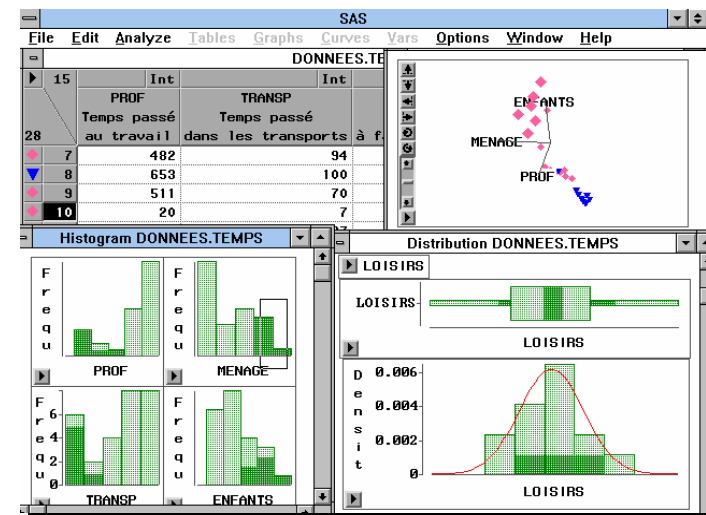
- x Erkennen von Ausreißern, Gruppen, Assoziationen ...
- x Visual Exploration:
  - w 3-dim. Charts
  - w Graphische Daten Analyse
  - w GIS
- Analytical Exploration:
  - w Cluster Analysis
  - w Correspondence Analysis
  - w PCA, Factor, MDS ...

Welche Fragen sollten gestellt werden ?



# DATA MANIPULATION

- x Welches sind wesentliche Variable?
- x Fehlende Werte ?
- x Variablentransformation ?
- x Neue Informationen hinzufügen: Groups, Labels etc.



Mit welchen Informationen sollte ich arbeiten ?



# ASSESSMENT

- x Bewertung: Wie gut ist mein Modell ?
  - w *Erklärungsbeitrag der Variablen, Ausreißer*
- x Assessment - Scoring
  - w *Klassifikation*
  - w *Lift Charts*
- x Verallgemeinerung  
für andere Daten



# SAS Data Mining Solution Currently (Feb 98)

- ◆ Data Warehousing incl. Web Technology
- ◆ Analytical Solutions
  - w *NNA - Production on Win, OS/2 and all major UNIX, ORLANDO I and II*
  - w *Tree Menue System*
  - w *Exploration: INSIGHT, SPECTRAVIEW, GIS*
  - w *Statistics*
  - w *Time Series Forecasting*
  - w *Market Research Methods*
- x EIS, Enterprise Reporter, Graphics



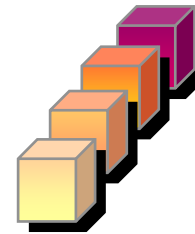
## Neue SAS DM Lösung SAS Enterprise Miner™

- x Einheitliche und voll skalierbare Business Lösung für das Data Mining
- x Füllt den Platz zwischen Data Warehousing und Endnutzer Reporting aus.
- x Das GUI schafft ein nutzerfreundliches front-end für den SEMMA Prozess.

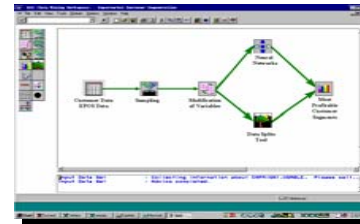
# SAS ENTERPRISE MINER

Vorteile für die Nutzer:

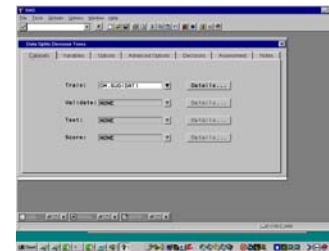
x IT: DW Zugriff, Skalierbarkeit



x Business Nutzer:  
Intuitive Oberfläche und  
Orientierung auf die  
Geschäftsfragen



x Data Miners: Analytische Tiefe  
und Flexibilität

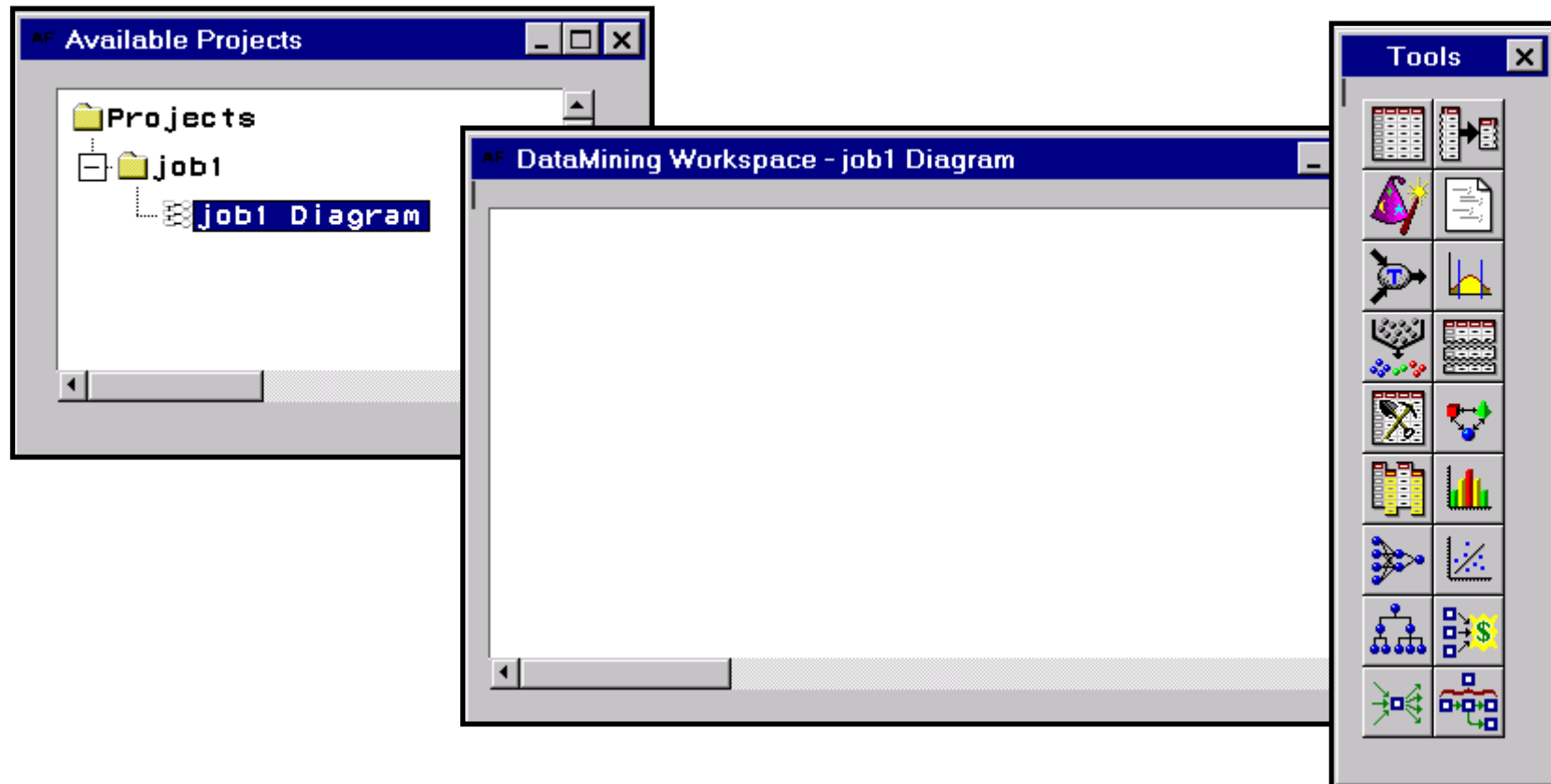


# SAS ENTERPRISE MINER Umgebung

- x Projekte/Modelle in Win95 Hierarchiestruktur
- x SEMMA Prozess in Process Flow Diagrams
- x Bestehende SAS Programme und Anwendungen können einfach integriert werden.
- x Alle Funktionalitäten des SAS Enterprise Miner wie die DMDB und alle analytischen Werkzeuge sind ausschließlich in dieser Data Mining Lösung verfügbar.

# ENTERPRISE MINER User Interface

- x 3 Hauptfenster: Projects, Data Mining Workspace, Tools Palette



# ENTERPRISE MINER

## Projekt Fenster

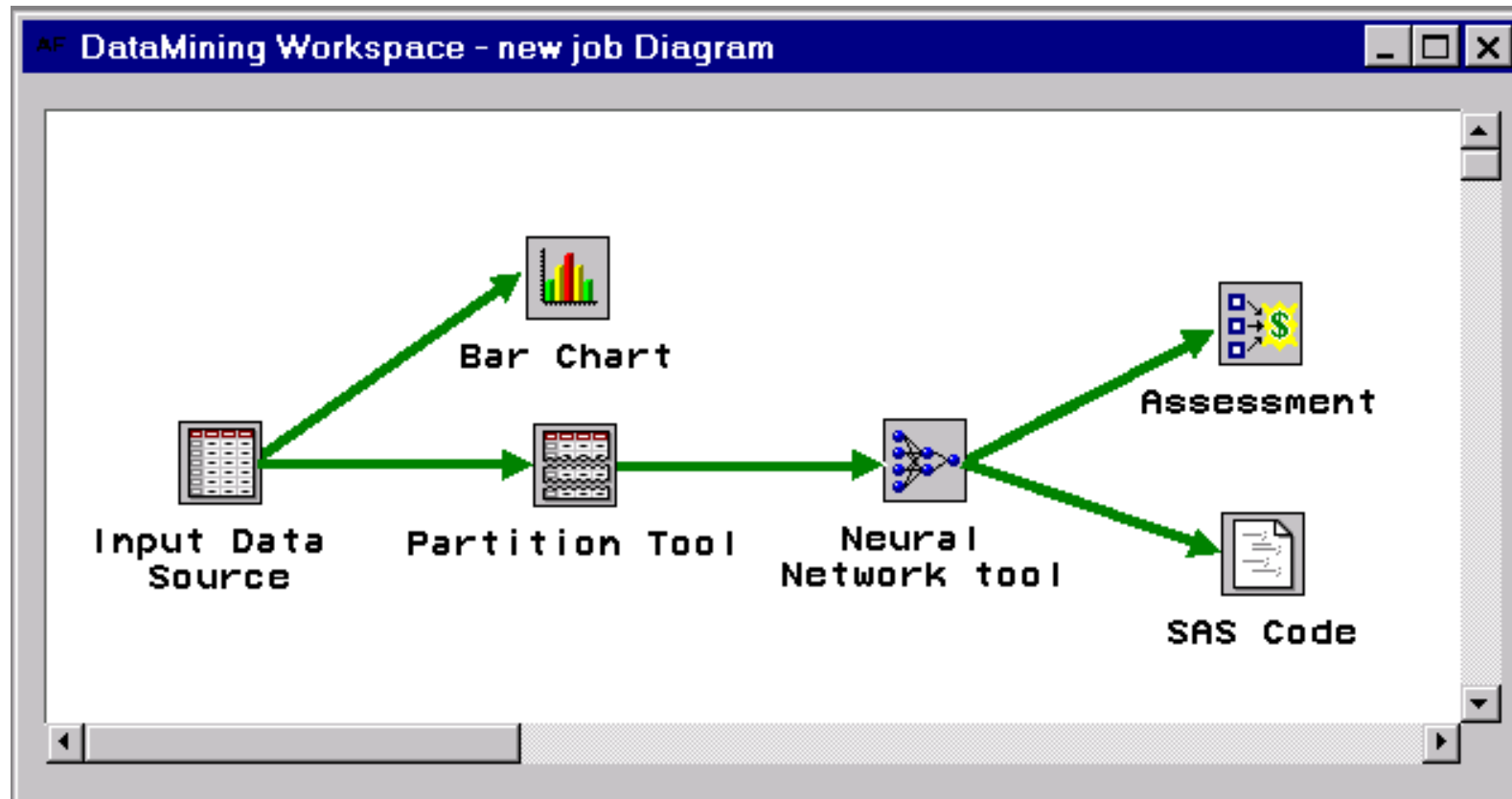
- x **Start: Doppel-click EM Icon**
- x **Fenster der verfügbaren Projekte**
- x **Maus-Steuerung**
- x **Pull-down menus: File, Edit, View, Insert, Globals, Options, Help**
- x **Toolbar: Up one level, Delete, Properties, Help**
- x **Pop-up menu: Open, Rename, Delete, Properties**
- x **Projekte: Create, Open, Save, Run, Close, Delete**

# ENTERPRISE MINER

## andere Fenster

- x Data Mining Window (DMW)
  - w *Default: open*
  - w *Build, edit, run process flow diagrams*
- x Tools Window
  - w *Default: open*
  - w *Tool palette, covers EM functionality*
  - w *D n' D tools on DMW window*
- x Message Window
  - w *Default: closed*
  - w *Messages generated when creating/running PFDs*

# ENTERPRISE MINER Process Flow Diagrams



## ENTERPRISE MINER DM Workspace Window

---

- x **Toolbar: Open, Save, Cut, Copy, Paste, Undo, Help**
- x **Pull-down menu: File, Edit, View, Actions, Globals, Options, Windows, Help**
- x **Pop-up menu: Add node, add endpoints, paste, undelete, select all, create subdiagram, refresh, up one level, top level, connect items, move and connect items**
- x **Add nodes: dnd icons or use pop-up menu**
- x **Connect, cut, delete nodes**
- x **PFD logic: tools loosely organized according to SEMMA**



# ENTERPRISE MINER

## Funktionalitäten

- x **Data:** Input Data Source, Random Sample, Partition, DMDB
- x **Explore/Modify:** Transform Data, Filter Outliers, Bar Chart, INSIGHT, Clustering, Variable Selection
- x **Modelling:** DM Regression, Neural Networks, Tree Models, Associations
- x **Assessment:** Scoring, Assessment
- x **Utilities:** Group Processing, Data Replacement, SAS Code Node, Administrator, Nodes Manager, Control Points, Subdiagrams.

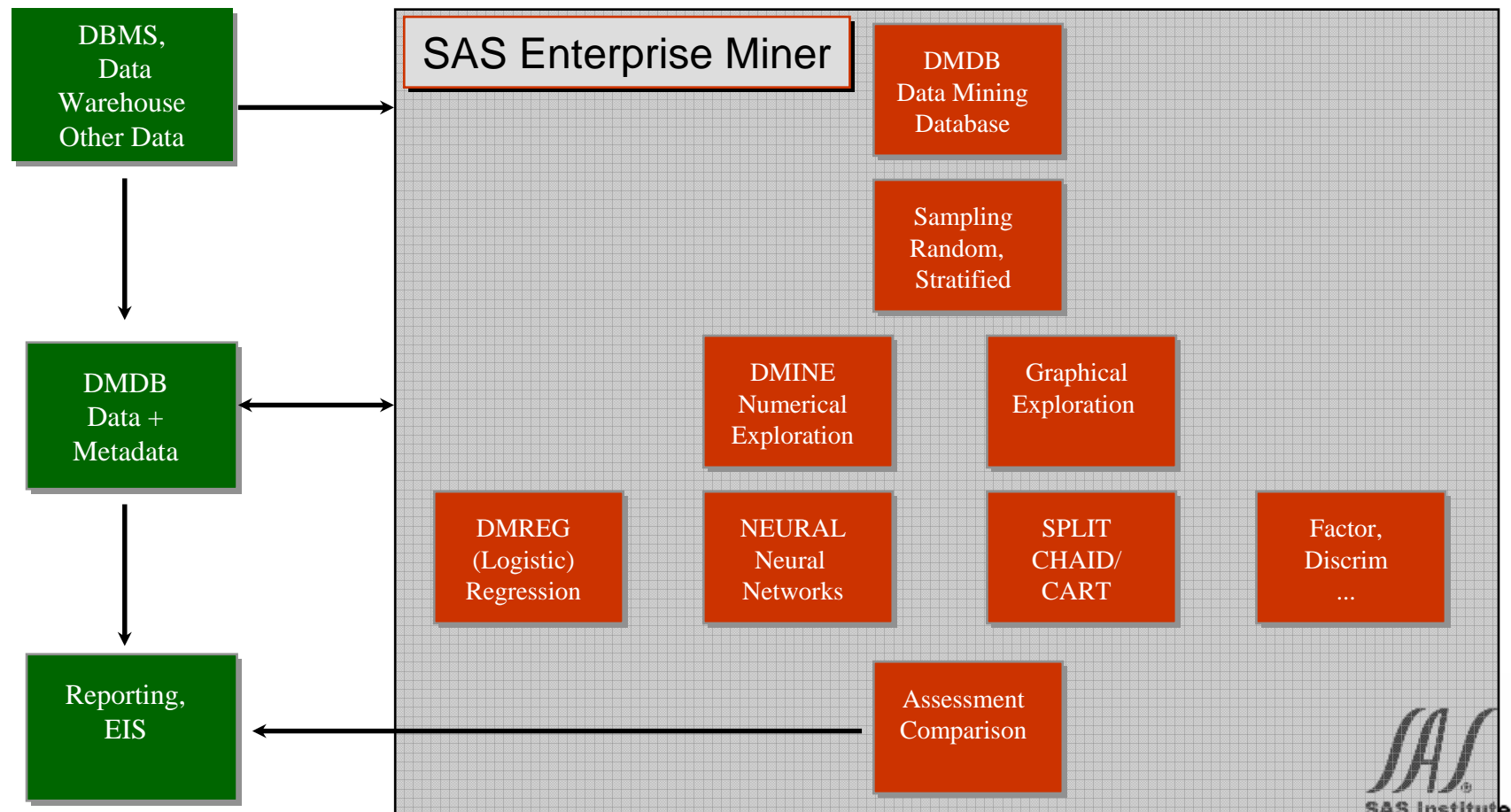
## Regeln für die Knoten

- x **Input data source node zuerst in PFD.**
- x **Sampling nach Input, dann beliebige Exploration, Modifizierung oder Modellierung**
- x **An beliebiger Stelle: Filter outliers, transform, bar chart**
- x **Nach Cluster: filter outliers, transform, bar chart, oder Modellierungen**
- x **Einem Assessment muß Modellierung vorangehen.**

## Einheitliches Erscheinungsbild der Knoten

- x **Dialog über Tabulatoren**
- x **Datendialog**
- x **Variablendialoge**
- x **Notizendialog**
- x **(einige Knoten): Browser für Resultate**

# SAS ENTERPRISE MINER Flow



# SAS ENTERPRISE MINER

## Systemanforderungen

- x Pentium PC
- x Windows NT 4.0+ or Win 95
- x 250 Mb + freier Plattenplatz
- x CD ROM Laufwerk

# SAS ENTERPRISE MINER Architektur

## Client-server Lösung:

- ◆ Clients: Win 95, Win NT
- ◆ Servers: Win NT, all major UNIX
- ◆ Mainframe als Data Server, später auch Compute Server
  
- ◆ Beta: Only Win95, Win NT initially.  
Unix: AIX, HP-UX, Solaris

# SAS ENTERPRISE MINER Beta

- x Etwa 100 EM Beta Anwendungen in USA
- x Etwa 60 EM Beta Tester in EUROPA

# Zusammenfassung

## SAS Enterprise Miner:

- ◆ Modelliert Data Mining als einen Prozess
- ◆ Ermöglicht Kooperation von IT, Business und Data Miners
- ◆ Vollständige SEMMA Implementation
- ◆ Integration von DW, DM and Reporting

➔ Wettbewerbsvorteil durch Data Mining