

Datenanalyse und Predictive Analytics

IBM SPSS Statistics
IBM Modeler



Wien, 27.11.2012

Predictive Analytics ist ein Prozess

Capture

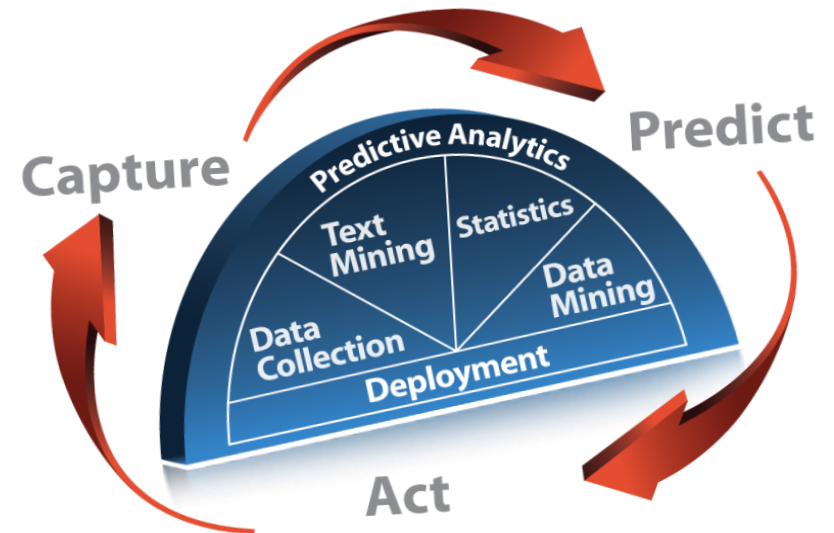
- Datenzugriff & ETL
- Befragung verschiedener Zielgruppen
- Nutzung strukturierter, halbstrukturierter und unstrukturierter Daten

Predict

- Statistische Analytik
- Data & Text Mining

Act

- Integration von Modellen in operative Systeme für bessere Entscheidungen
- Direkte Einbettung in zentrale Geschäftsprozesse

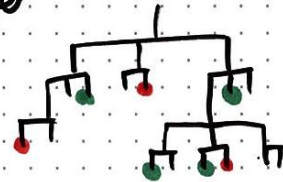


Typische Fragestellungen im Predictive Analyticsumfeld

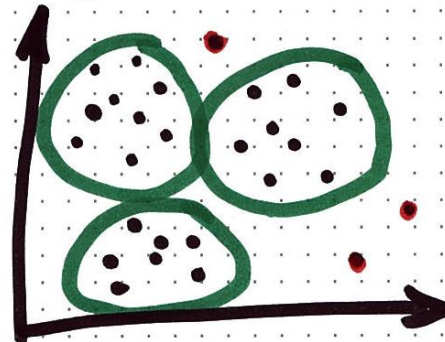
Klassifikation

ID	G	A	B	C	Ziel
1	M	30	0	0	0
2	W	38	1	0	1
3	M	50	0	0	0

Kündiger
Aquirise
Risiko
Betrag
...

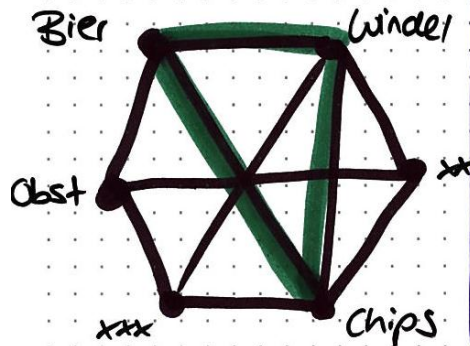


Segmentierung

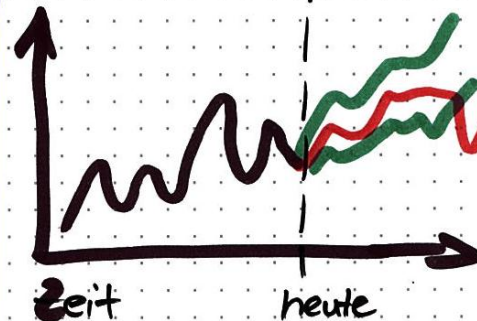


Kundensegmente
Anomalien -> Betrag

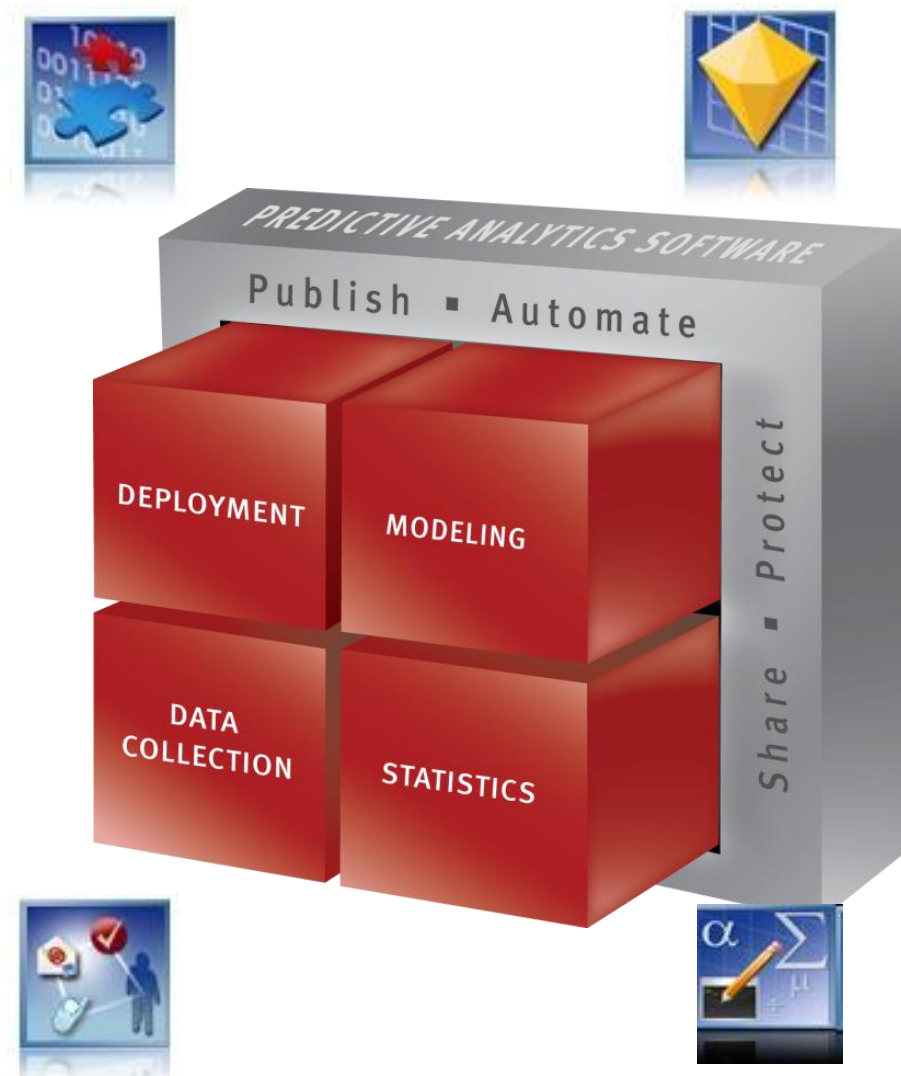
Assoziation



Forecasting



IBM SPSS Predictive Analytics Produktpalette

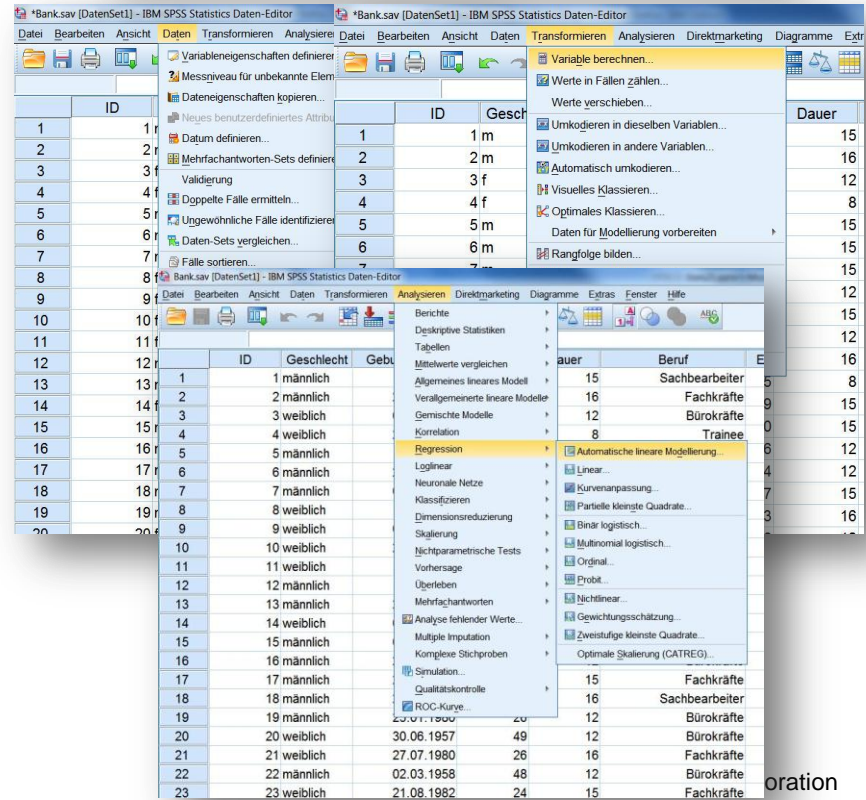
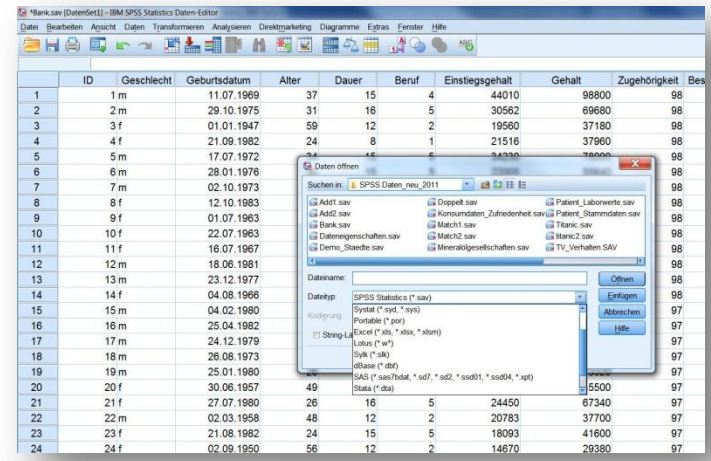


IBM SPSS Statistics

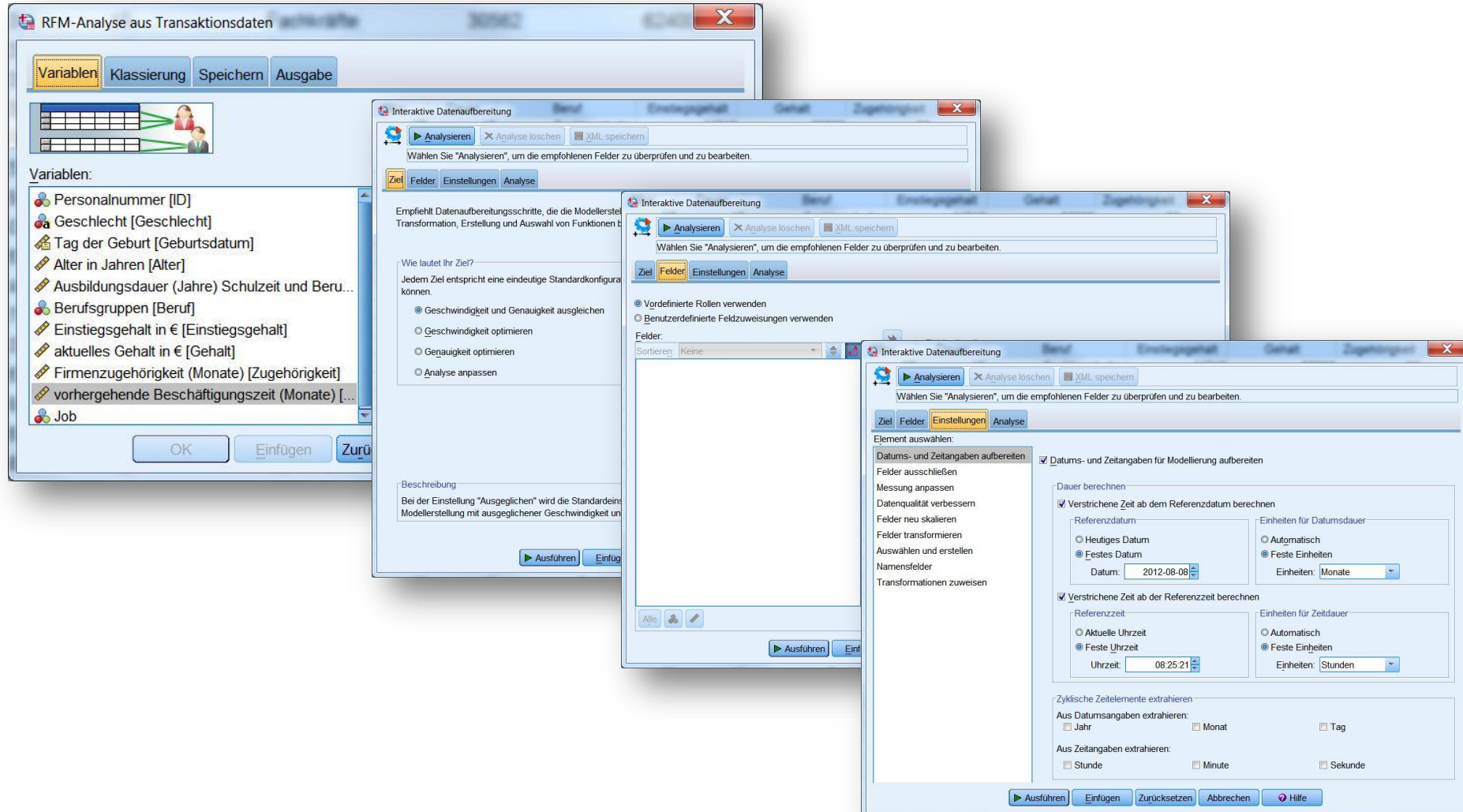


IBM SPSS Statistics 21 unterstützt Sie optimal,

- durch seinen einfachen und problemlosen Datenzugriff,
- sein umfangreiches Datenhandling,
- seine breite Palette an Datenaufbereitungsfunktionen, Validierungen und Deskriptiven Statistiken,
- sowie das mächtige Spektrum an statistischen Prozeduren, Algorithmen und Funktionen.



Zahlreiche **Assistenten** und **Editoren** machen es für Sie einfach, auch komplexe Verfahren anzuwenden und somit das breite Analyseportfolio von SPSS optimal zu nutzen.

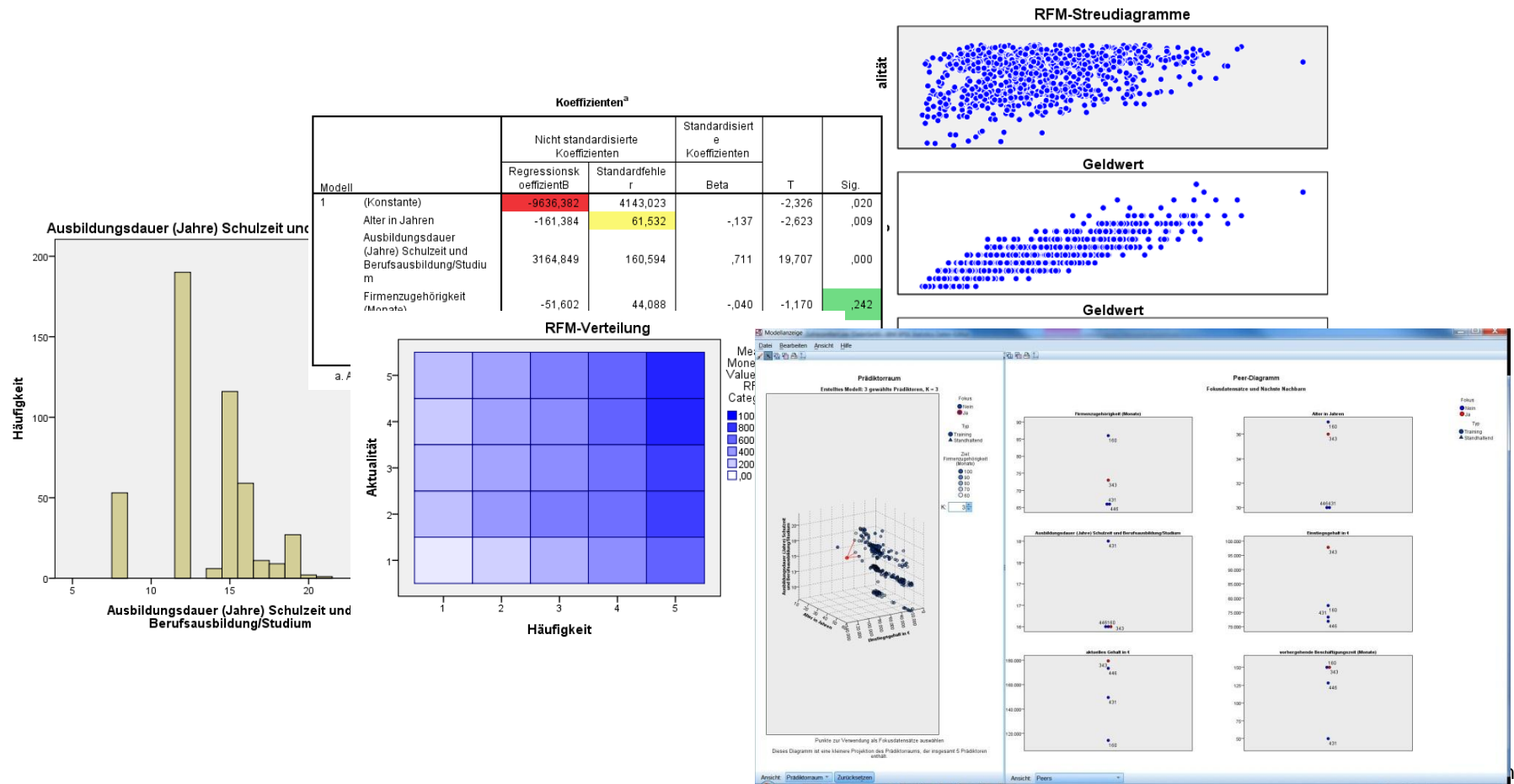


The image displays four overlapping screenshots of the SPSS Assistant and Editor interface, illustrating the workflow for RFM analysis:

- Top Left:** The main Assistant window titled "RFM-Analyse aus Transaktionsdaten". It features a navigation bar with buttons for "Variablen", "Klassierung", "Speichern", and "Ausgabe". Below the navigation bar is a diagram showing data flow. A list of variables is shown, including "Personalnummer [ID]", "Geschlecht [Geschlecht]", "Tag der Geburt [Geburtsdatum]", "Alter in Jahren [Alter]", "Ausbildungsdauer (Jahre) Schulzeit und Beru...", "Berufsgruppen [Beruf]", "Einstiegsgehalt in € [Einstiegsgehalt]", "aktuelles Gehalt in € [Gehalt]", "Firmenzugehörigkeit (Monate) [Zugehörigkeit]", "vorhergehende Beschäftigungszeit (Monate) [...]", and "Job".
- Middle Left:** An "Interaktive Datenaufbereitung" (Interactive Data Preparation) window. It prompts the user to "Wählen Sie 'Analysieren', um die empfohlenen Felder zu überprüfen und zu bearbeiten." It offers options for "Vorgefertigte Rollen verwenden" (selected) and "Benutzerdefinierte Feldzuweisungen verwenden".
- Middle Right:** Another "Interaktive Datenaufbereitung" window, showing the "Felder" (Fields) section. It includes a "Sortieren" dropdown set to "Keine" and a list of actions such as "Datums- und Zeitangaben aufbereiten", "Felder ausschließen", "Messung anpassen", "Datenqualität verbessern", "Felder neu skalieren", "Felder transformieren", "Auswählen und erstellen", "Namensfelder", and "Transformationen zuweisen".
- Bottom Right:** A third "Interaktive Datenaufbereitung" window, showing the "Einstellungen" (Settings) section. It includes options for "Dauer berechnen" (Duration calculation) and "Verstrichene Zeit ab dem Referenzdatum berechnen" (Calculate elapsed time from reference date), with settings for "Referenzdatum" (Date: 2012-08-08) and "Einheiten für Datumsdauer" (Units: Monate). It also includes settings for "Verstrichene Zeit ab der Referenzzeit berechnen" (Calculate elapsed time from reference time), with settings for "Referenzzeit" (Time: 08:25:21) and "Einheiten für Zeitdauer" (Units: Stunden). At the bottom, there are checkboxes for "Zyklische Zeitelemente extrahieren" (Extract cyclic time elements) for dates (Jahr, Monat, Tag) and times (Stunde, Minute, Sekunde).

Durch die statistische Analyse verschaffen Sie sich wichtige Einblicke in die aktuelle Situation und das momentane Geschehen.

Dies können sie durch geeignete Grafiken und Tabellen in Berichten und Dokumentationen visualisieren.



Analyse unstrukturierter Daten - Sentimentanalyse

Projekt 1 - IBM SPSS Text Analytics for Surveys

Datei Bearbeiten Ansicht Kategorien Extras Hilfe

Statement

Aufbauen Erweitern

Kategorie	Anzahl Deskriptoren	Anzahl Treffer
Alle Datensätze	-	5102
Nicht kategorisiert	-	2241
Keine Konzepte extrahiert	-	275
volumen	47	1312
wimpern	43	1089
mascara	29	455
Appeal	34	376
sinnlichkeit	7	15
schön	6	135
look	4	41
sinnlichste look	1	25
farbe	9	112
schwarze farbe		1
farbe der verpackung		2
farbe der aufmachung		1
farbe des stifts		1
farbe		88
rote farbe	4	20
finish	7	81
verführerischen finish		2
verführerishcte finish		1

Kategoriebalken Kategorienetzdiagramm Tabelle für Kategorienetzdiagramm

Gitterlayout

```

    graph TD
      gut --- Appeal_sinnlichkeit[Appeal/sinnlichkeit]
      mascara --- Appeal_sinnlichkeit
      Appeal_sinnlichkeit --- volumen
      Appeal_sinnlichkeit --- wimpern
      volumen --- wimpern
  
```

ID	Antwort	Kategorien
1	weil es was neues ist und amn das in dem rot besonders gut erkennen kann, w eil es noch keine mascaras gibt, die von aussen rot sind. passt auch sehr zu sinnlichkeit	...peal/sinnlichkeit gut mascara
2	die spielen das mit der sinnlichkeit zu hoch. das ist zu billig	...peal/sinnlichkeit
3	weil mir rot gefällt und das ziemlich viel weiblichkeit und auch sinnlichkeit ausstrahlt	...peal/sinnlichkeit
4	weil er sinnliche augen mahct und mehr länge und mehr füle gibt	...peal/sinnlichkeit
5	sinnlichkeit	...peal/sinnlichkeit
6	eine sinnliche verformung durch die anwendung	...peal/sinnlichkeit
7	weil rot was anderes ist uns mir ein gefühl von sinnlichkeit verleiht. und auch verrucht	...peal/sinnlichkeit
8	kann unglaubliche länge und va-va-volumen gleichzeitig schaffen; für einen sinnlichen wimpernaufschlag	...peal/sinnlichkeit volumen

Nicht verwendete Extrahierungen Alle Extrahierungen

Extrahieren Konzept

- wimpern (925)
- volumen (764)
- länder (753)
- nichts (425)
- mascara (392)
- high (376)
- heel (285)
- bürste (247)
- onyx (242)
- va va volumen (239)
- 70% (192)
- 80% (179)
- neu (151)
- cut (151)
- keine ahnung (142)
- canbelline (137)
- high-heel volume (130)
- schön (125)
- stretch bürste (119)

11 (29) Kategorien

Darüberhinaus können Sie mit IBM SPSS Statistics proaktiv

- Ereignisse vorhersagen,
- Muster erkennen,
- Verhaltenweisen bewerten,
- Gruppenzugehörigkeiten ermitteln und
- auffällige Transaktionen erkennen.

Dies alles ermöglicht es Ihnen und ihrer Organisation Prozesse und Aktionen optimal zu steuern und damit schneller und bequemer zu zuverlässigen Ergebnissen zu kommen.

Organisationen werden in die Lage versetzt treffgenauer zu agieren und gruppenspezifische Reaktionen und Maßnahmen zu treffen.

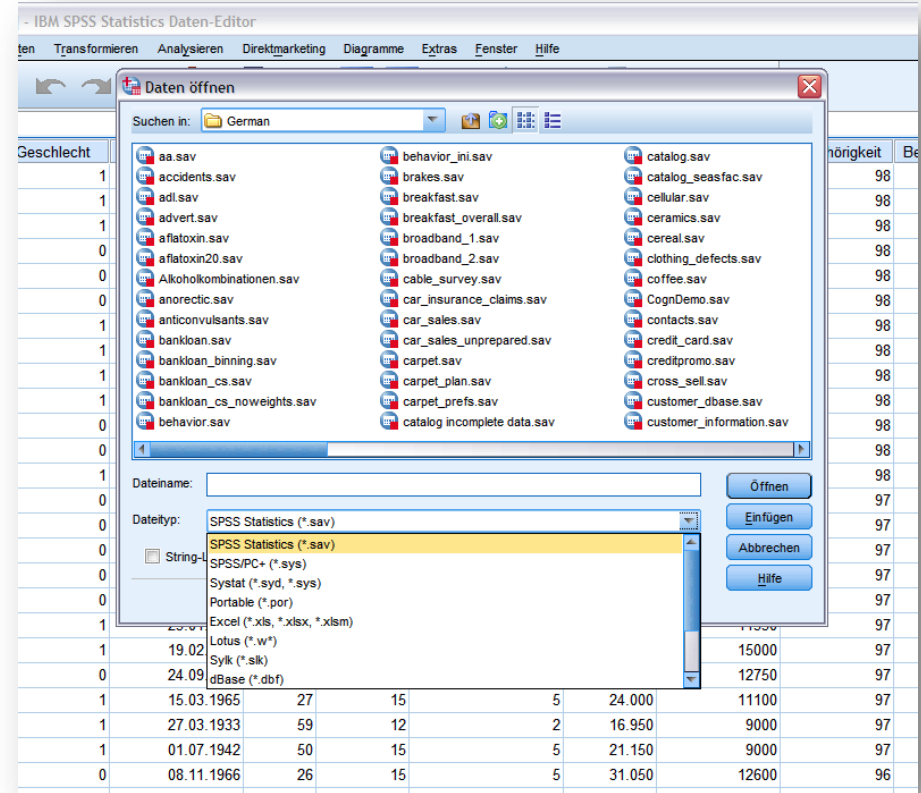
10 Gründe die für IBM SPSS Statistics sprechen



1. Einlesen verschiedener Dateiformate

IBM SPSS Statistics ermöglicht den Datenimport von verschiedenen Dateiformaten:

- Text,
- Excel, CSV,
- Lotus,
- Portable,
- Sylk,
- Systat,
- SPSS/PC+,
- SPSS Statistics,
- dBASE,
- SAS,
- Stata

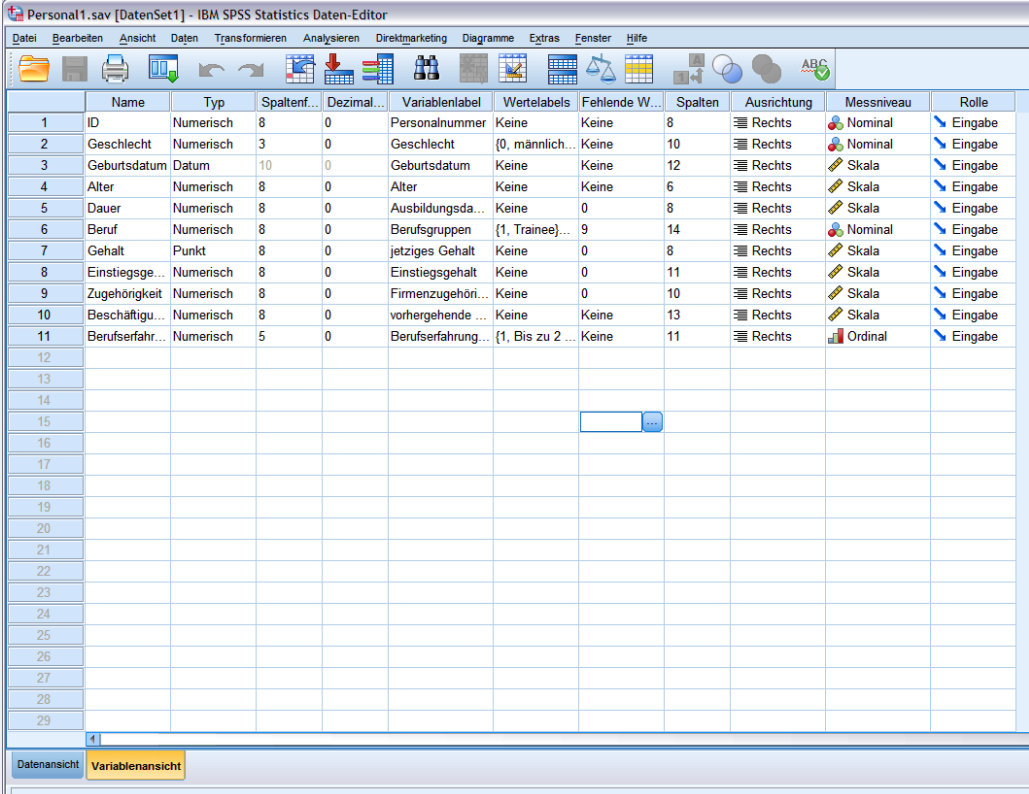


Einlesen von Datenbankdateien aus beliebigen Formaten erfolgt über den Datenbank-Assistenten

2. Variablenmanagement

Neben der Datenansicht gibt es in IBM SPSS Statistics die **Variablenansicht**. Hier wird das Variablenmanagement vorgenommen:

- Variablenname
- Variablentyp
- Variablenlabel
- Wertelabel
- Fehlende Werte
- Skalenniveaus
- etc.



The screenshot shows the 'Variable View' in IBM SPSS Statistics. The table below represents the data shown in the screenshot.

	Name	Typ	Spaltenf...	Dezimal...	Variablenlabel	Wertelabels	Fehlende W...	Spalten	Ausrichtung	Messniveau	Rolle
1	ID	Numerisch	8	0	Personalnummer	Keine	Keine	8	Rechts	Nominal	Eingabe
2	Geschlecht	Numerisch	3	0	Geschlecht	{0, männlich...	Keine	10	Rechts	Nominal	Eingabe
3	Geburtsdatum	Datum	10	0	Geburtsdatum	Keine	Keine	12	Rechts	Skala	Eingabe
4	Alter	Numerisch	8	0	Alter	Keine	Keine	6	Rechts	Skala	Eingabe
5	Dauer	Numerisch	8	0	Ausbildungsda...	Keine	0	8	Rechts	Skala	Eingabe
6	Beruf	Numerisch	8	0	Berufsgruppen	{1, Trainee}...	9	14	Rechts	Nominal	Eingabe
7	Gehalt	Punkt	8	0	jetziges Gehalt	Keine	0	8	Rechts	Skala	Eingabe
8	Einstiegsge...	Numerisch	8	0	Einstiegsgehalt	Keine	0	11	Rechts	Skala	Eingabe
9	Zugehörigkeit	Numerisch	8	0	Firmenzugehör...	Keine	0	10	Rechts	Skala	Eingabe
10	Beschäftigu...	Numerisch	8	0	vorhergehende ...	Keine	Keine	13	Rechts	Skala	Eingabe
11	Berufserfahr...	Numerisch	5	0	Berufserfahrung...	{1, Bis zu 2 ...	Keine	11	Rechts	Ordinal	Eingabe
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											
28											
29											

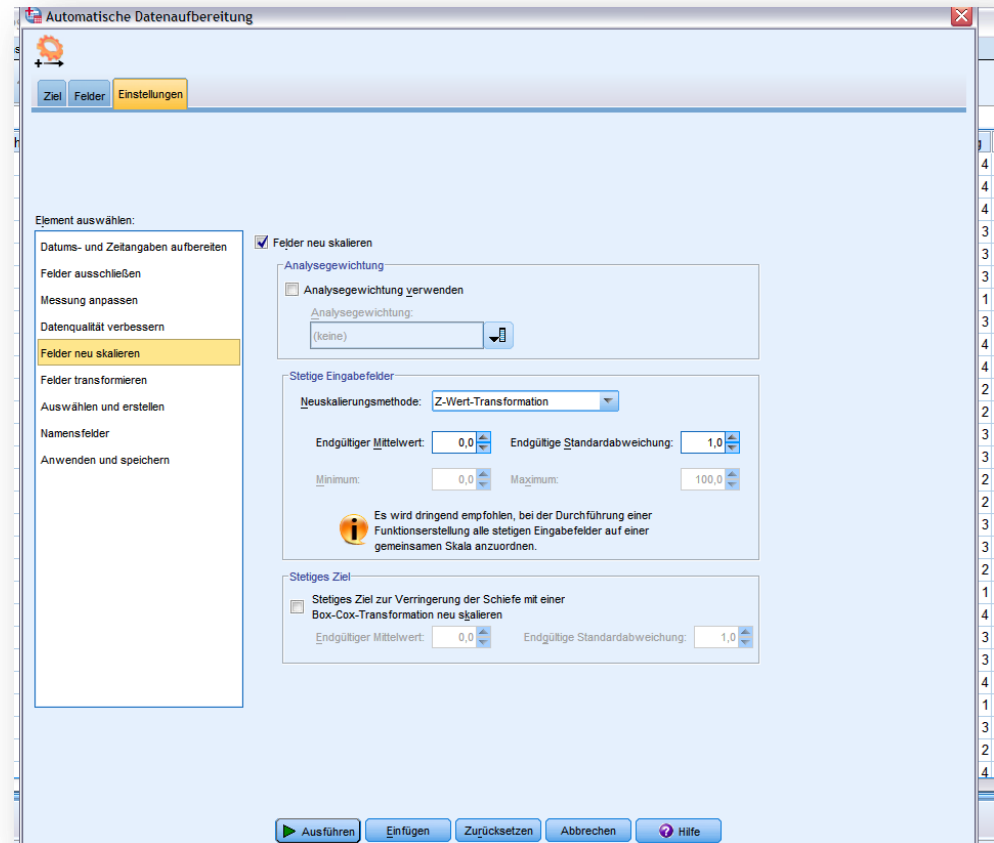
3. Automatische Datenaufbereitung

Identifiziert Problemlösungen

Findet problematische oder
wahrscheinlich nicht nützliche Felder

Leitet zum passenden Zeitpunkt neue
Attribute ab

Verbessert Leistungsfähigkeit durch
intelligente Screening-Methoden



4. Assistierte Formelerstellungen

Kann bei numerischen und String-Variablen eingesetzt werden

Zahlreiche systemeigene Funktionen stehen zur Verfügung. Bspw. String-, Statistik-, Verteilungs- und Arithmetische Funktionen

Auf Grundlage logischer Bedingungen lassen sich Werte für definierte Teilmengen berechnen

The screenshot shows the 'Daten-Editor' application with a 'Variable berechnen' dialog box open. The dialog is used to define a new variable 'Varianz_GEHALT' with the formula 'VARIANCE(Gehalt, Einstiegsgehalt)'. The background shows a data table with columns for dates, age, duration, profession, salary, starting salary, membership, and working hours.

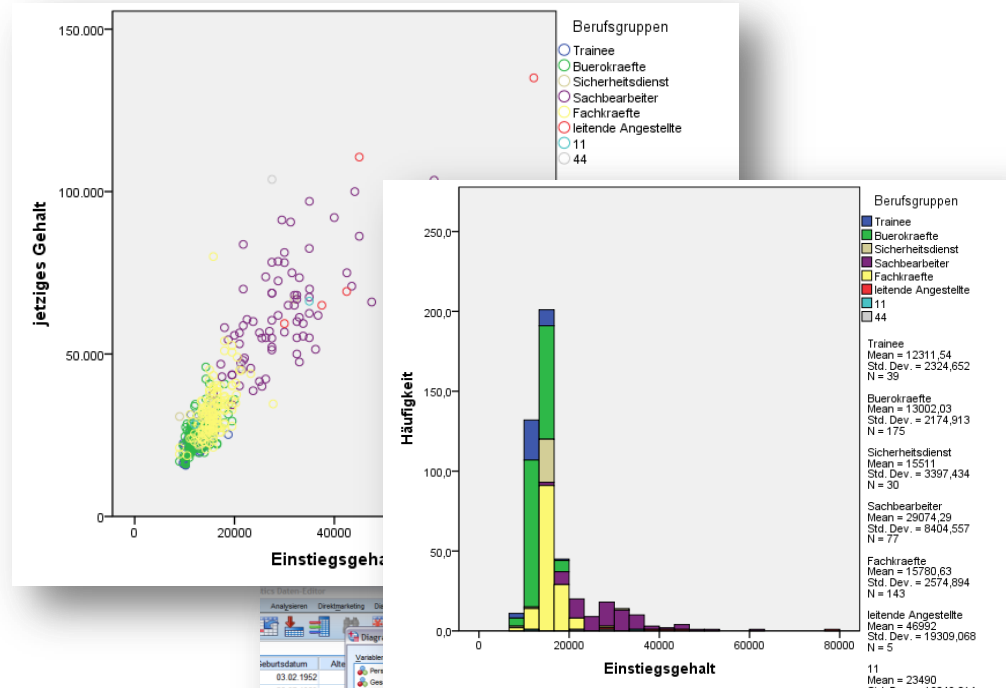
tsdatum	Alter	Dauer	Beruf	Gehalt	Einstiegsgehalt	Zugehörigkeit	Beschäftigungszeit	Berufs...
03.02.1952								144
26.07.1929								381
15.04.1965								190
09.02.1955								138
22.08.1958								67
26.04.1956								114
06.05.1966								0
23.01.1946								115
13.02.1946								244
07.02.1950								143
11.01.1964								26
17.07.1960								34
26.02.1949								137
29.08.1962								66
17.11.1964								24
18.07.1962								48
20.03.1956								70
19.08.1962								103
23.01.1940								48
19.02.1963								17
24.09.1940								315
15.03.1965								75
27.03.1933								124
01.07.1942								171
08.11.1966								14
19.03.1954	38	19	4	60.375	27480	96		96
11.04.1963	29	15	5	32.550	14250	96		43
28.01.1944	48	19	6	135.000	79980	96		199

5. Diagrammerstellung

Per Drag & Drop lassen sich Diagramme schnell und einfach umsetzen

Nachträglich im interaktiven Outputviewer veränderbar

Gruppierungsmöglichkeit



Statistik

Variablen: Berufsführung in Jahren

ganzes

Wert:

Fehlerbalken anzeigen

Fehlerbalken entsprechen:

- Ein-Minutenintervalle
- Minuten: 95
- Stundenintervalle
- Multiplikator: 2
- Stundenschiebung
- Multiplikator: 2

Subansicht:

95	3
43	2
199	4

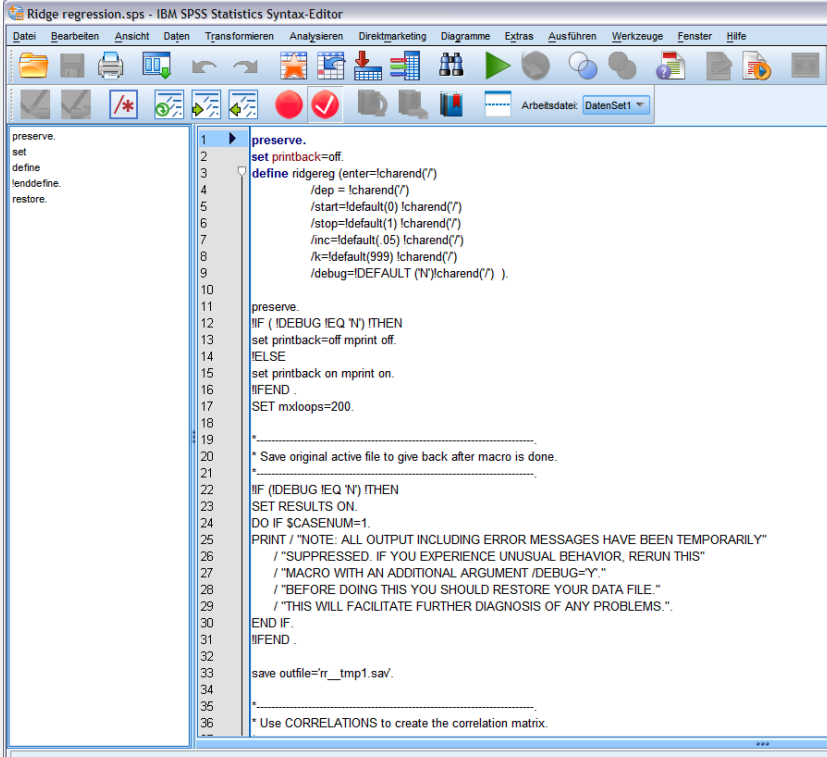
6. Automatisierung wiederkehrender Analysen

Automatisierungsmöglichkeiten mittels
Syntax

Syntax kann für Folgeanalysen
„zusammengeklickt“ und abgespeichert
werden

Syntaxfehler werden automatisch
erkannt

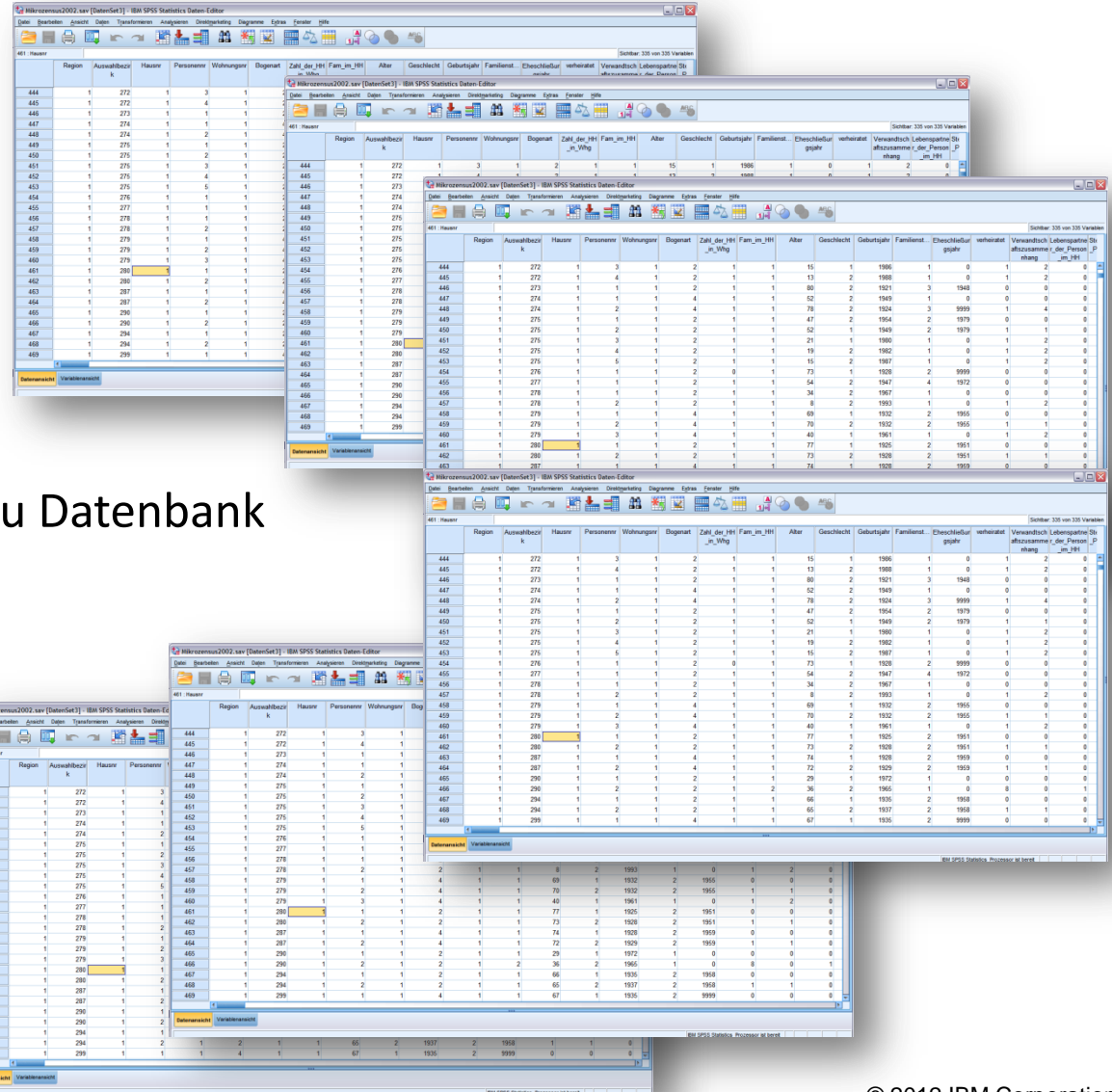
Syntaxassistent



```
Ridge regression.sps - IBM SPSS Statistics Syntax-Editor
Datei Bearbeiten Ansicht Daten Transformieren Analysieren Direktmarketing Diagramme Extras Ausführen Werkzeuge Fenster Hilfe
Arbeitsdatei: DataSet1

1 preserve.
2 set printback=off.
3 define ridgereg (enter=lcharend(?)
4 /dep = lcharend(?)
5 /start=ldefault(0) lcharend(?)
6 /stop=ldefault(1) lcharend(?)
7 /inc=ldefault(.05) lcharend(?)
8 /k=ldefault(999) lcharend(?)
9 /debug=IDEFAULT (N)lcharend(?) ).
10
11 preserve.
12 IF ( !DEBUG IEQ 'N' ) THEN
13 set printback=off mprint off.
14 ELSE
15 set printback on mprint on.
16 IFEND .
17 SET mxloops=200.
18
19 *-----
20 * Save original active file to give back after macro is done.
21 *-----
22 IF ( !DEBUG IEQ 'N' ) THEN
23 SET RESULTS ON.
24 DO IF $CASENUM=1.
25 PRINT / "NOTE: ALL OUTPUT INCLUDING ERROR MESSAGES HAVE BEEN TEMPORARILY"
26 / "SUPPRESSED. IF YOU EXPERIENCE UNUSUAL BEHAVIOR, RERUN THIS"
27 / "MACRO WITH AN ADDITIONAL ARGUMENT /DEBUG='Y.'"
28 / "BEFORE DOING THIS YOU SHOULD RESTORE YOUR DATA FILE."
29 / "THIS WILL FACILITATE FURTHER DIAGNOSIS OF ANY PROBLEMS."
30 END IF.
31 IFEND .
32
33 save outfile=tr_tmp1.sav.
34
35 *-----
36 * Use CORRELATIONS to create the correlation matrix.
37
```

7. Arbeiten mit sehr großen Datensets



Keine Limitierung

Import bzw. Export aus bzw. zu Datenbank möglich

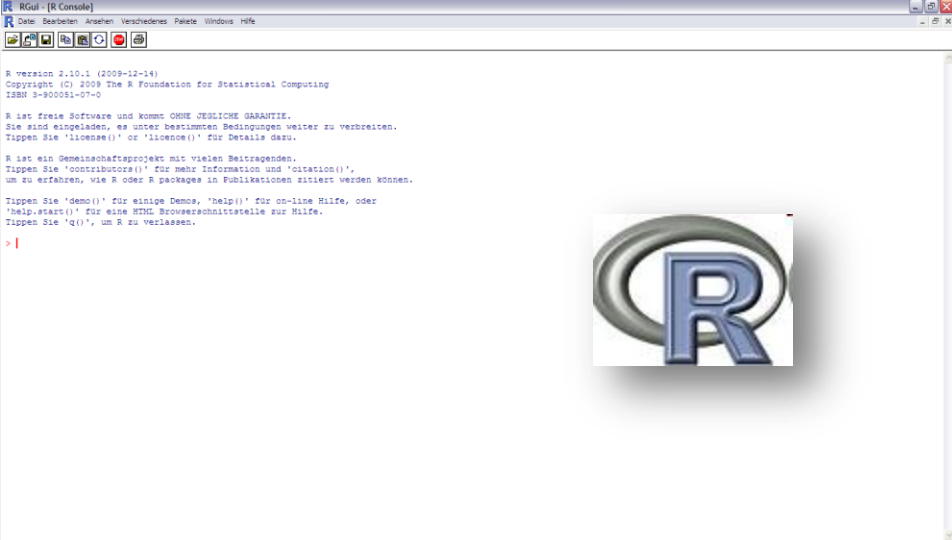
Client- / Server- Architektur

8. Erweiterungsmöglichkeiten durch R und Python

Downloadbare Plug-Ins für R und Python

„Fremder“ Code wird von IBM SPSS Statistics verarbeitet und im Outputfenster umgesetzt

Große R und Python Communities/Foren



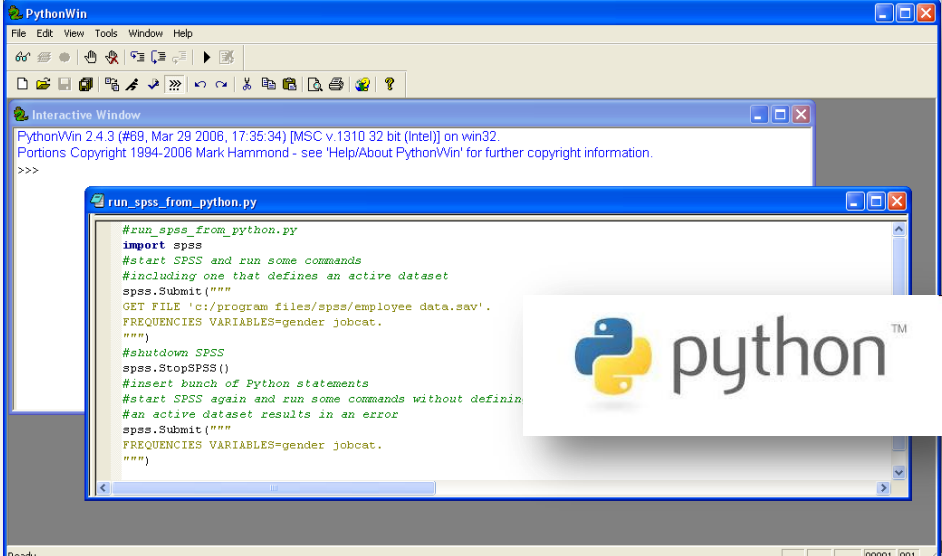
```
RGui - [R Console]
Datei Bearbeiten Ansichten Verschiedenes Paletze Windows Hilfe

R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R ist freie Software und kommt OHNE JEGLICHE GARANTIE.
Sie sind eingeladen, es unter bestimmten Bedingungen weiter zu verbreiten.
Tippen Sie 'license()' or 'licence()' für Details dazu.

R ist ein Gemeinschaftsprojekt mit vielen Beitragenden.
Tippen Sie 'contributors()' für mehr Information und 'citation()',
um zu erfahren, wie R oder R packages in Publikationen zitiert werden können.
Tippen Sie 'demo()' für einige Demos, 'help()' für on-line Hilfe, oder
'help.start()' für eine HTML-Browserschnittstelle zur Hilfe.
Tippen Sie 'q()', um R zu verlassen.

> |
```



```
PythonWin
File Edit View Tools Window Help

Interactive Window
PythonWin 2.4.3 (#69, Mar 29 2006, 17:35:34) [MSC v.1310 32 bit (Intel)] on win32.
Portions Copyright 1994-2006 Mark Hammond - see 'Help/About PythonWin' for further copyright information.
>>>

run_spss_from_python.py
#run_spss_from_python.py
import spss
#start SPSS and run some commands
#including one that defines an active dataset
spss.Submit("""
GET FILE 'c:/program files/spss/employee data.sav'.
FREQUENCIES VARIABLES=gender jobcat.
""")
#shutdown SPSS
spss.StopSPSS()
#insert bunch of Python statements
#start SPSS again and run some commands without defining
#an active dataset results in an error
spss.Submit("""
FREQUENCIES VARIABLES=gender jobcat.
""")
```

9. Umfassende Hilfefunktionen (inkl. Statistik-Assistenten)

Intelligente Hilfefunktion: aus jeder Dialogbox abrufbar
 Kontextsensitiv
 Statistik-Assistent
 Fallbeispiele

The screenshot shows the IBM SPSS Statistics 'Variable berechnen' dialog box. A red arrow points from the 'Hilfe' button in the dialog to a help window titled 'Hilfe - IBM SPSS Statistics - Mozilla Firefox: IBM Edition'. The help window displays the 'Berechnen von Variablen' section, which explains how to use the 'Berechnen' button to calculate new variables based on numerical transformations of other variables. It lists several options: 'Select Amount spent as the dependent variable', 'Select who shopping for and use coupons as fixed factors', 'Select Store ID as a random factor', and 'Click Model'. Below this, it explains how to use the 'Ausdruck' field to enter mathematical expressions or function groups.

10. Modularer Aufbau

IBM SPSS Statistics besitzt einen modularen Aufbau:

Das Basissystem als Kernstück = Grundausstattung

Verschiedene Expertenmodule zur Erweiterung

Module sind einzeln erhältlich

Nach Bedarf können einzelne Module wie bspw. „Advanced Statistics“, „Data Preparation“, etc. integriert werden

Weitere Programme stehen zusätzlich zur Verfügung:

- **Amos**
- **Textanalyse für die Auswertung von offenen Fragen**
- **SPSS Server**

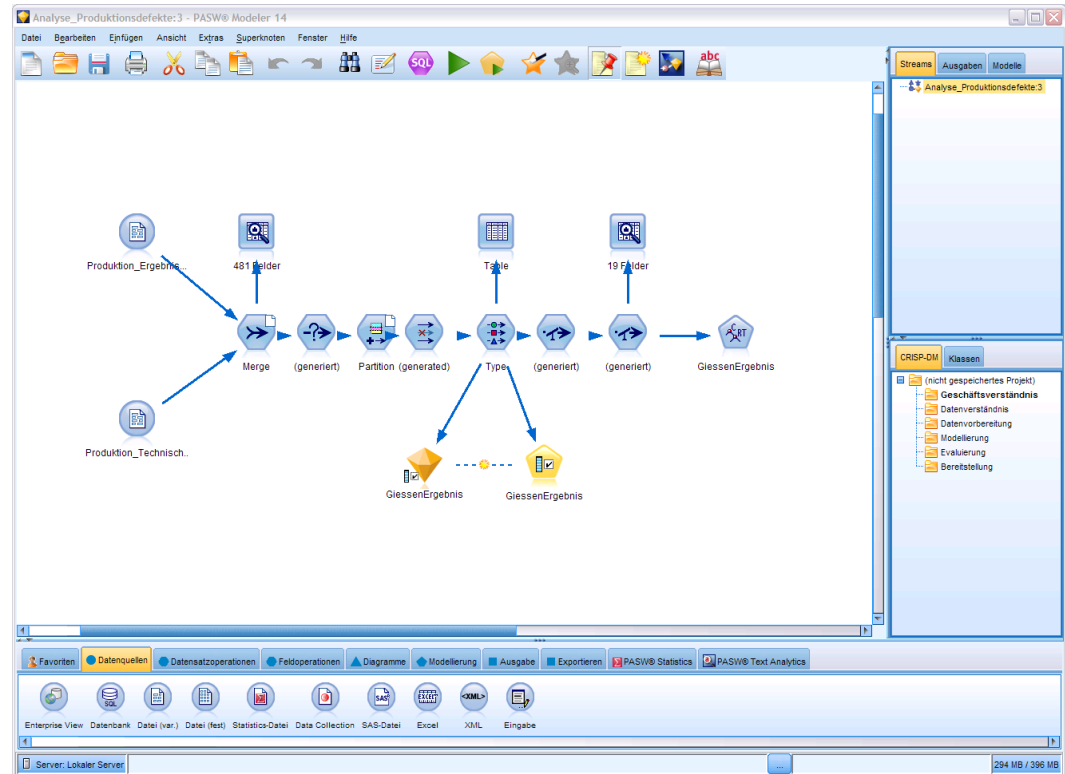


IBM SPSS Modeler

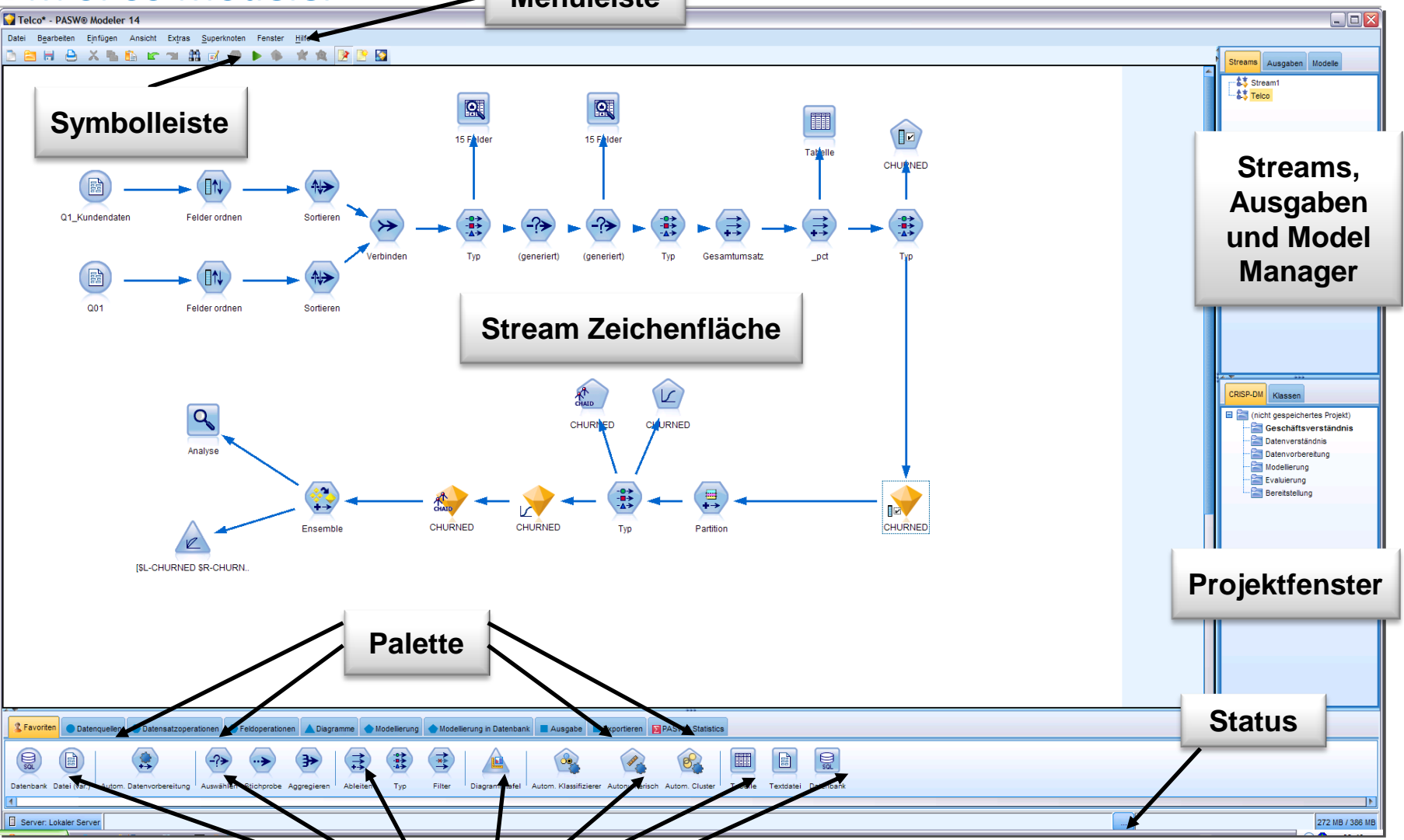


High End Analytics mit dem IBM SPSS Modeler

- Visuelles Programmieren analytischer Streams
- Hohes Maß an Interaktivität und Benutzerfreundlichkeit
- Skalierbarkeit durch Client-/Server Architektur
- Nahtlose Zusammenarbeit mit allen gängigen Datenbanksystemen
- Orientierung am CRISP-DM Modell für Data Mining



IBM SPSS Modeler



IBM SPSS Predictive Analytics – Vorteile für den Anwender

Predictive Analytics liefert schnell ROI.

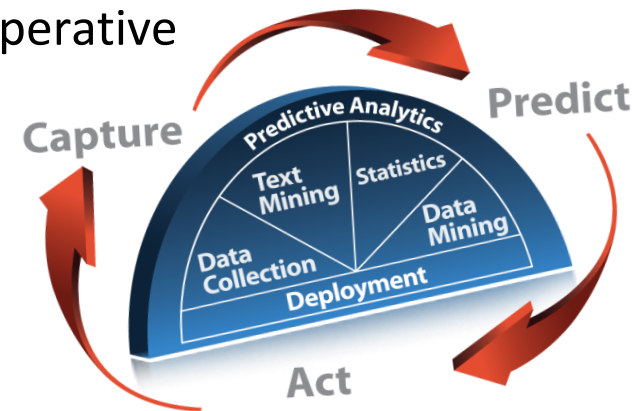
- Breites Set an Mining Algorithmen zur Vorhersage ermöglicht das Erkennen von Schlüsselindikatoren, Mustern und Trends
- Mit diesen Erkenntnissen werden anschließend operative Entscheidungen getroffen

Ermöglicht das Mining auch unstrukturierter Daten, egal aus welcher Quelle

- Verbessert die Modellgüte
- Erhöht den ROI durch noch genauere Vorhersagen

Maximale Produktivität für den Analysten

- Einfach zu lernen
- Man muss kein Programmierer oder geschulter Analyst sein



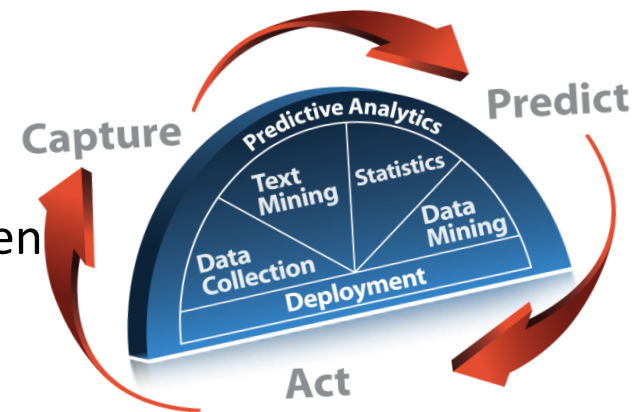
IBM SPSS Predictive Analytics – Vorteile für den Anwender

Ausgereifte Automation

- Automatisieren Sie die zeitraubende Datenaufbereitung
- Erstellen, Evaluieren und Verteilen Sie Ihre prädiktiven Modelle ebenfalls automatisiert
- „Next-best-practise“, „Next-best-Action“ Szenarios
- Modellextraktion für Drittanwendungen

Performanz und Skalierbarkeit

- Von Batchverarbeitung bis hin zu Echtzeit-Szenarien

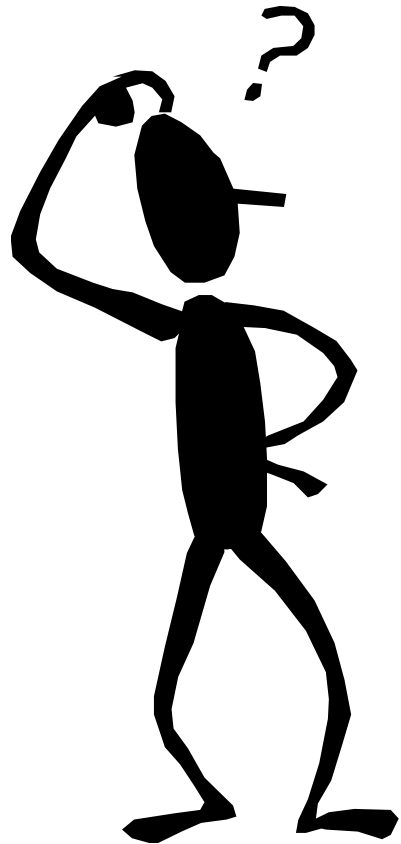


Demonstration

IBM SPSS Statistics

IBM SPSS Modeler

Vielen Dank für Ihre Aufmerksamkeit



Für weitere Informationen:

Tel: +49 (0)89 4504 2022

E-mail: spssales@de.ibm.com

Website: www.ibm.com/spss/de

