SAS® Forum
Switzerland 2014
make connections • share ideas • be inspired

What's new in Statistical Analysis

Carmelo Iantosca
Practice Leader Analytics

THE POWER TO KNOW.

Copyright © 2014, SAS Institute Inc. All rights reserved.

Agenda

- What's new in **Enterprise Guide** 6.1
- What's new in **Add-in for Microsoft Office** 6.1
- What's new in **Enterprise Miner** 13.1
- SAS® Visual Statistics – eine neue Welt

Copyright © 2014 SAS Institute Inc. All rights reserved.

What‘s new in Enterprise Guide 6.1

Copyright © 2014 SAS Institute Inc. All rights reserved.

Release Zielsetzungen

- Fähigkeiten der SAS 9.3/9.4 Plattform steigern
- Steigerung des Bedienkomforts
- Steigerung der Produktivität
- Deployment Verbesserungen

Copyright © 2014 SAS Institute Inc. All rights reserved.

DEMO

Key Features

- Weitere Verbesserung des Programm-Editors:
  - Autocomplete für Dataset-Spalten
    (neu auch in data steps und bei sämtlichen keep-, drop-, by- und var-Aufrufen)
  - Autocomplete für High Performance Prozeduren
- Prozessfluss kann an beliebigen Stellen mit Notiz- bzw. Kommentar-Feldern versehen werden
- Task-Gallery

Copyright © 2014 SAS Institute Inc. All rights reserved.

Key Features

- Das Log-Fenster beinhaltet neu eine Zusammen-fassungsliste für Errors, Warnings und Notes inkl. Jump-To Verlinkungen
- Neue Tasks und Prozeduren
  - High Performance logistische und lineare Regression
  - RFM
  - Task Gallery
- Neu einen separaten MSI-Installer für den EG

Copyright © 2014 SAS Institute Inc. All rights reserved.

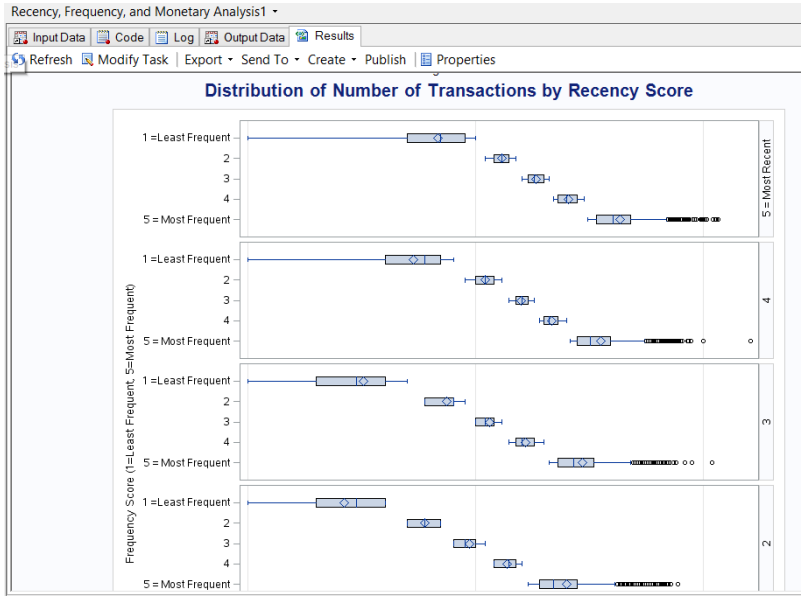
Key Features

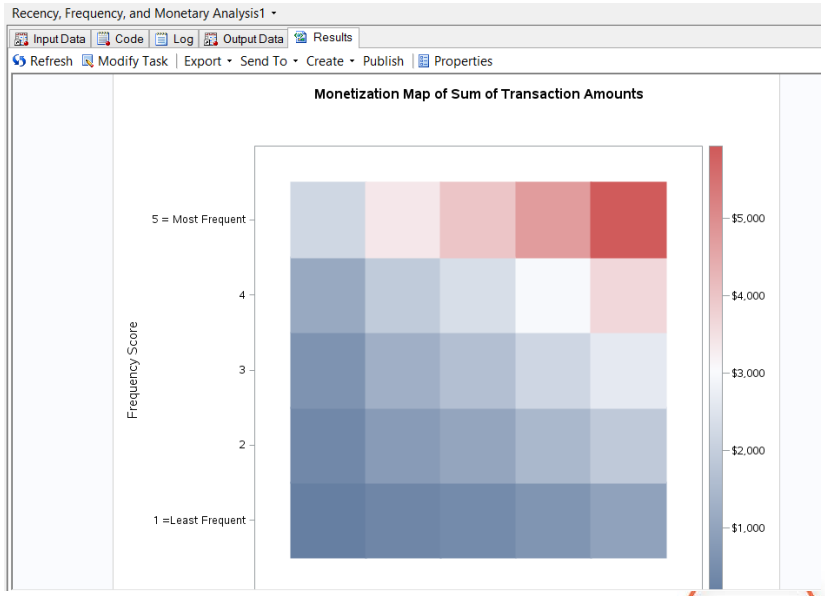
- Support für SAS 9.4 und Office 2013
- Vereinfachtes Deployment
  - .MSI Installer
  - Support für Application Streaming
- EG nur noch als 64-bit Version,
  keine Unterstützung für Windows XP
- Integration mit neuem Windows Task Scheduler

Copyright © 2014 SAS Institute Inc. All rights reserved.

Key Features

- Screenshots zum RFM Task

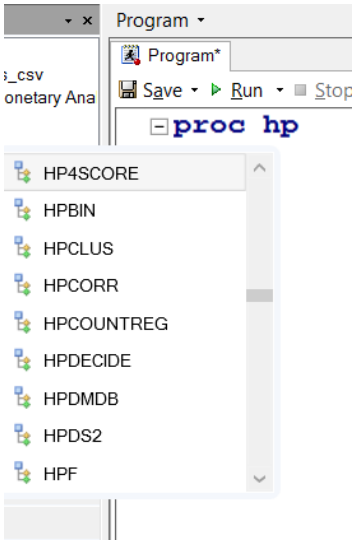




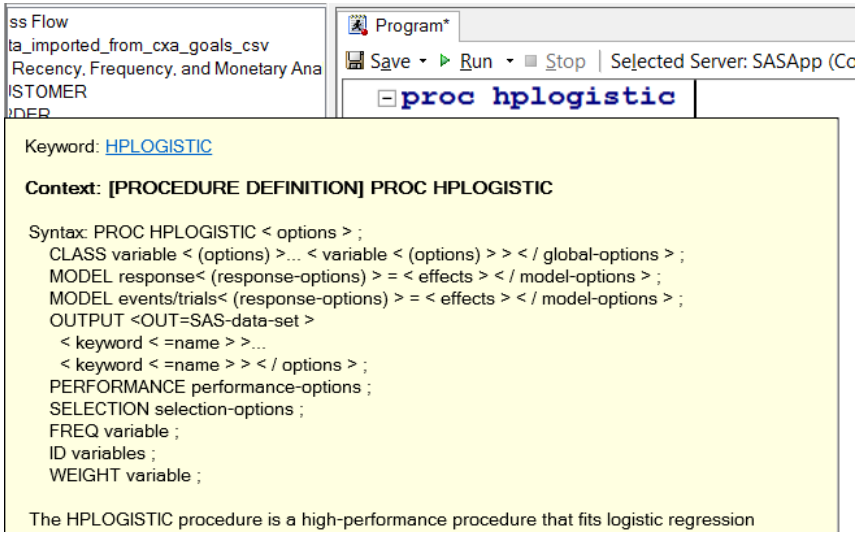
Copyright © 2014 SAS Institute Inc. All rights reserved.

Key Features

- High Performance Integration →



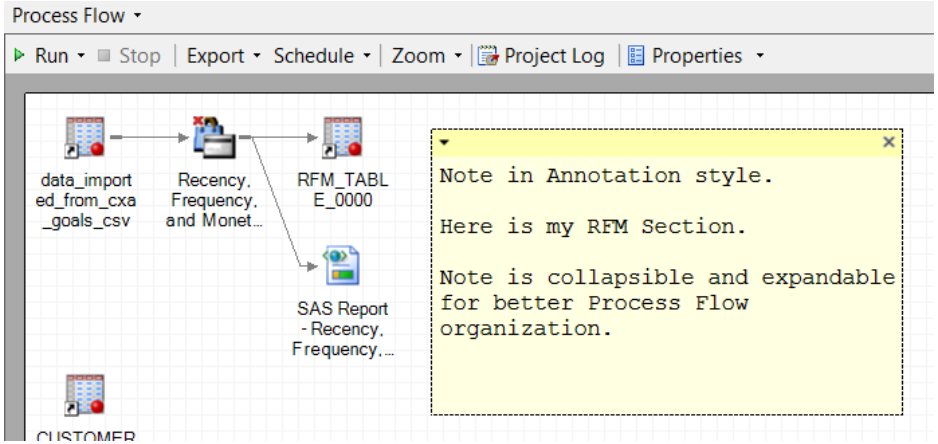
- Syntax Support für High Performance Prozeduren



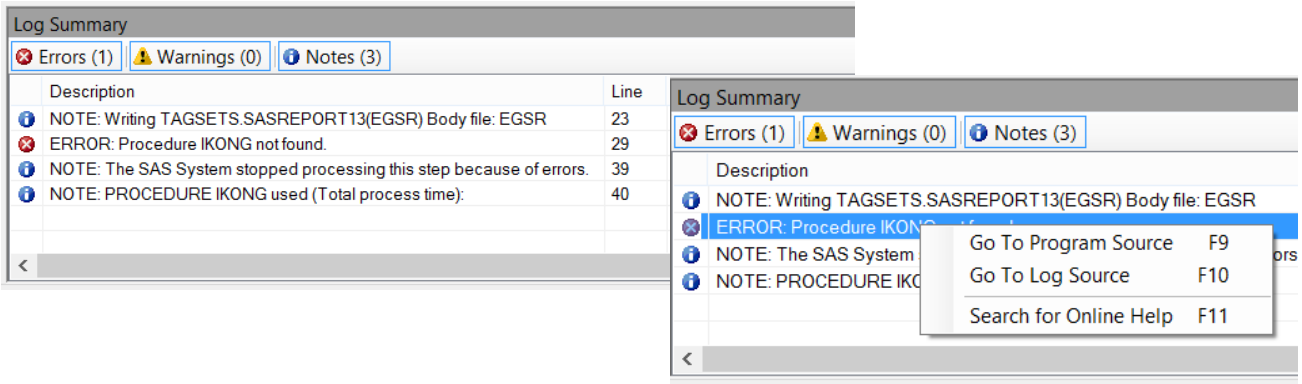
Copyright © 2014 SAS Institute Inc. All rights reserved.

Key Features

- Notizen/Kommentare im Prozessfluss



- Log Summary mit Index und Jump-To



Copyright © 2014 SAS Institute Inc. All rights reserved.

What‘s new in
Add-in for Microsoft Office 6.1

Copyright © 2014 SAS Institute Inc. All rights reserved.

Release Zielsetzungen

- Fähigkeiten der SAS 9.3/9.4 Plattform steigern
- Steigerung des Bedienkomforts
- Steigerung der Produktivität
- Steigerung der Integration

Copyright © 2014 SAS Institute Inc. All rights reserved.

DEMO

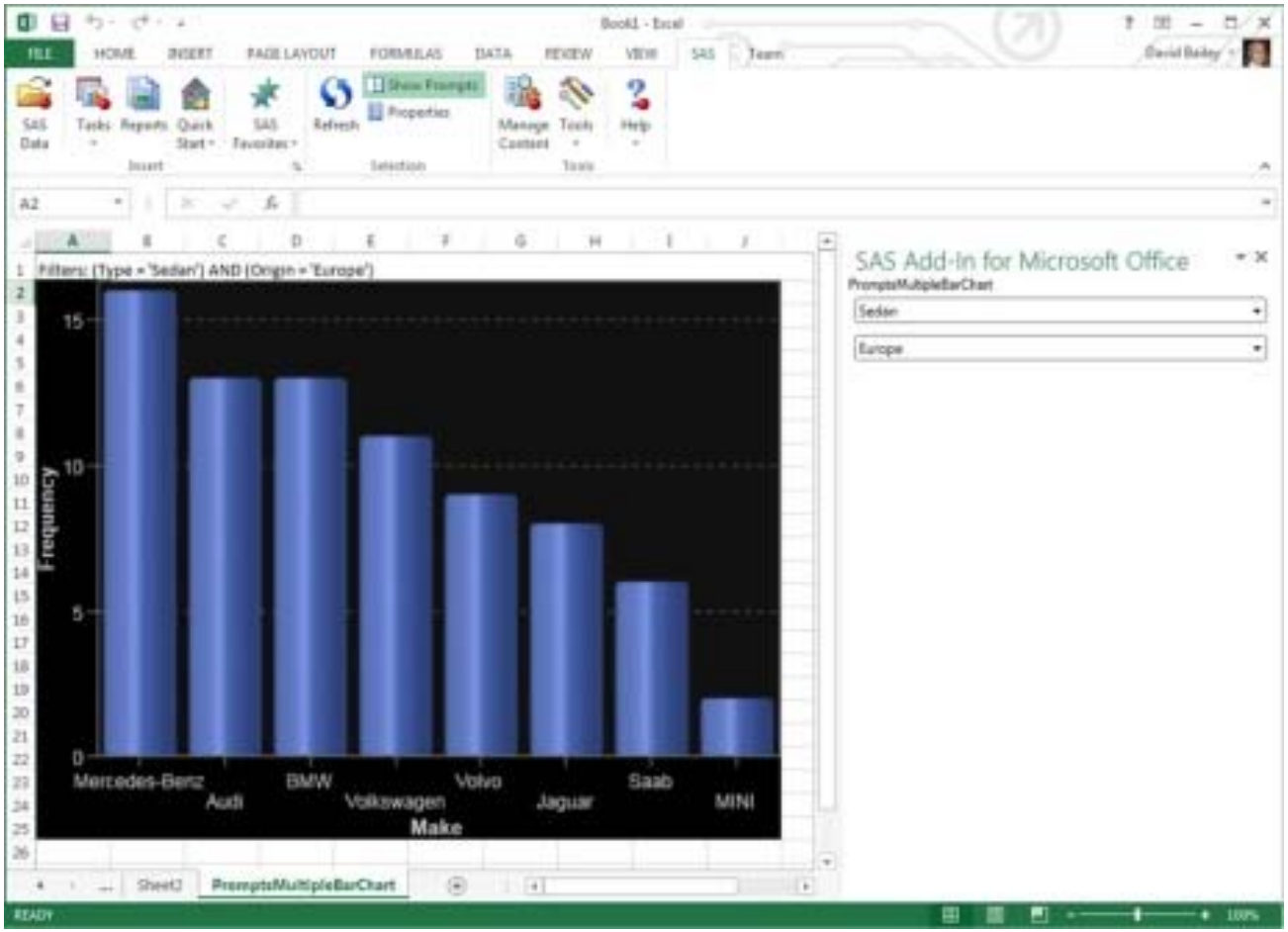
Key Features

- Support für Visual Analytics
  - Direkter Zugriff und Interaktion mit Visual Analytics Berichten (Reports)
  - Excel, Word, PowerPoint, Outlook
  - Outlook rendert dabei direkt die Flash-Inhalte (analog Web-Portal)
  - Filter in VA Berichten werden dabei in Microsoft Office Controls umgewandelt (Excel, Word und PowerPoint)
  - Listen und Kreuztabellen werden direkt in die Excel-Zellen integriert, d.h. die Werte können in gewohnter Excel-manier verlinkt und/oder für Excel-Grafiken weiter verwendet werden

Copyright © 2014 SAS Institute Inc. All rights reserved.

Key Features

- Beispiel-Screenshot



Copyright © 2014 SAS Institute Inc. All rights reserved.

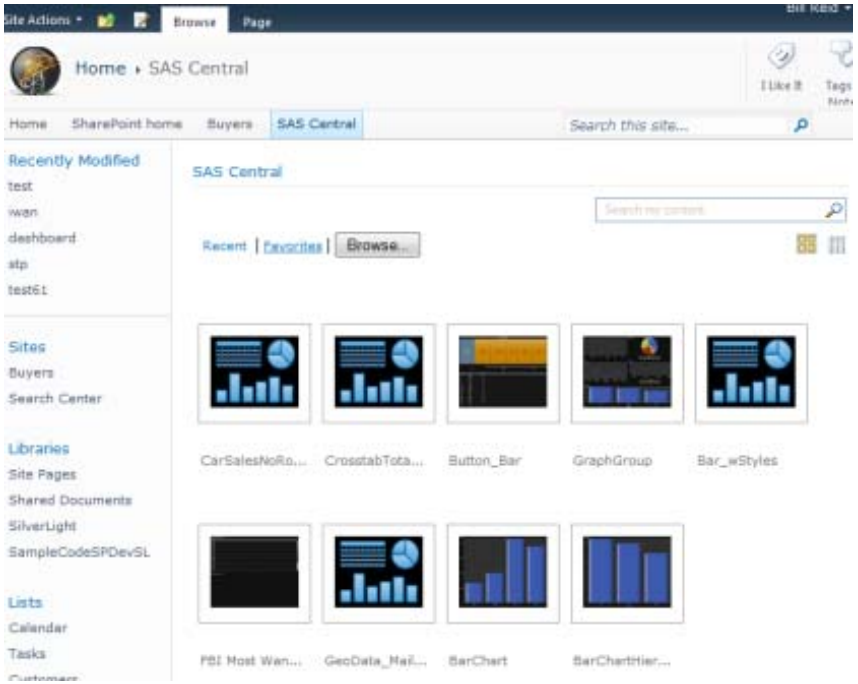
Key Features

- Verbesserungen an Benutzeroberfläche und Performance
- Verbesserungen für Pivot-Tabellen Support
  - Neu kann z.B. die Datenquelle angepasst werden und die erstellte Pivot-Struktur bleibt bestehen
- Schnellerer Outlook Start durch verzögertes Laden der Metadaten

Copyright © 2014 SAS Institute Inc. All rights reserved.

Zudem in Office Analytics 6.1
(Bundle aus SAS Server, EG und AMO)

- SharePoint-Integration:
  - SharePoint Web Part für Darstellung der Visual Analytics Berichte
  - SAS Central Navigation
  - SharePoint 2013 Support



Copyright © 2014 SAS Institute Inc. All rights reserved.

FAQ

Unterstützte Betriebssysteme

- Microsoft Windows Server 2008
- Microsoft Windows Server 2008 R2
- Microsoft Windows Server 2012
- Microsoft Windows 7 SP1
- Microsoft Windows 8

Unterstützte Office Versionen (AMO)

- Microsoft Office 2007
- Microsoft Office 2010 (32-bit oder 64-bit)
- Microsoft Office 2013 (32-bit oder 64-bit)

Unterstützte SAS Versionen

- SAS 9.2
- SAS 9.3
- SAS 9.4

Hardware Mindestanforderungen (Client)

- 400 MB Platz auf Harddisk
- 2 GB RAM

Copyright © 2014 SAS Institute Inc. All rights reserved.

FAQ

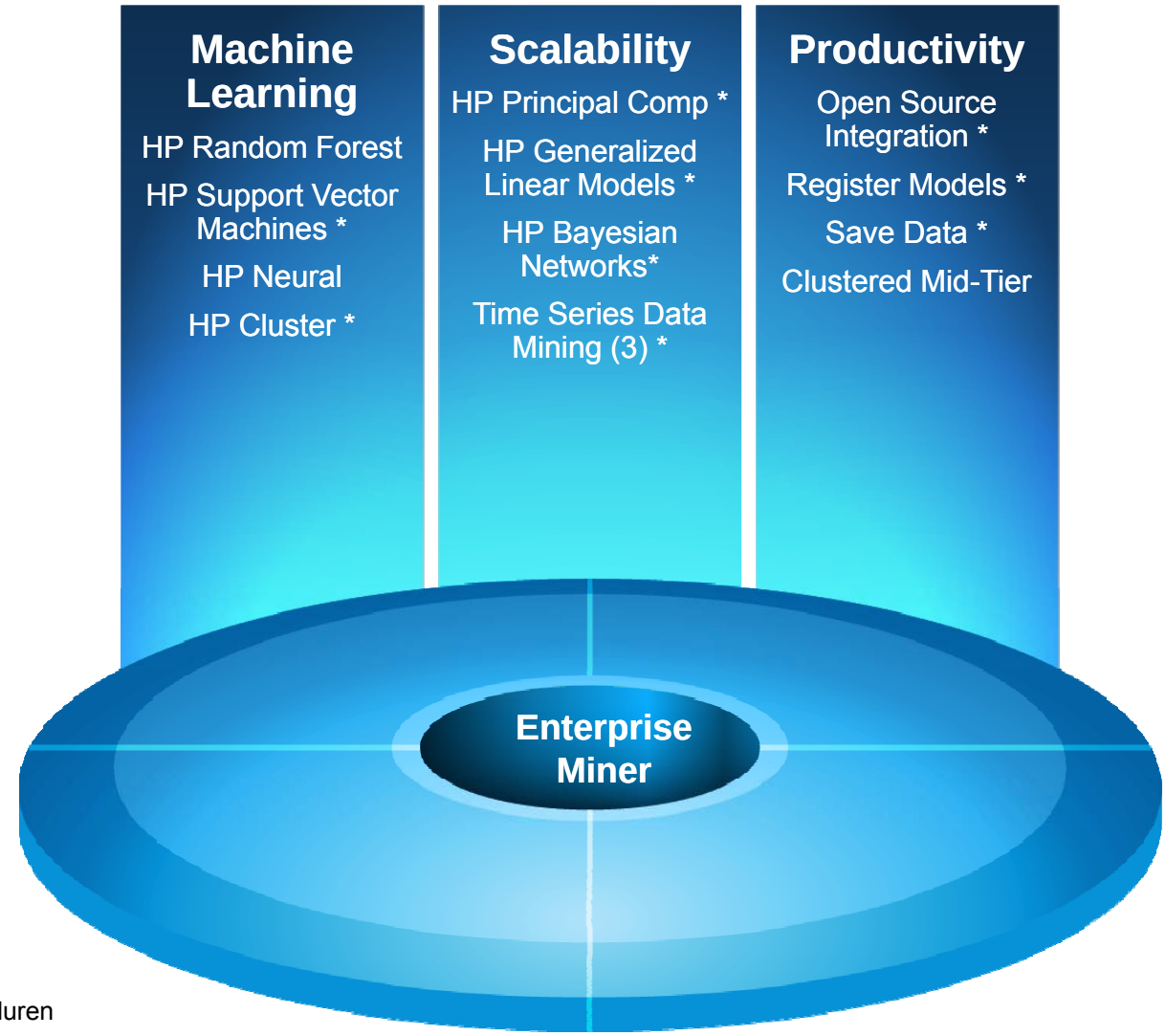
- Kann AMO / EG 6.1 mit SAS 9.2 und SAS 9.3 verwendet werden?
  - Ja 6.1 unterstützt SAS 9.2, 9.3 und 9.4
- Unterstützt AMO 6.1 VA 6.1, 6.2 und 6.3?
  - Ja
- Greift AMO oder EG auf Daten aus dem LASR-Server zu?
  - Nein, sie öffnen und interagieren direkt mit VA Reports
- Ist AMO innerhalb der VA Lizenz verfügbar
  - Nein, AMO ist Bestandteil der BI, Enterprise BI und Office Analytics Bundles
- Gibt es weiterhin 32bit-Versionen der Clients (AMO, EG)?
  - AMO gibt es weiterhin als 32bit-Version, da dies nötig ist, falls Office als 32bit-Version installiert ist
  - EG 6.1 gibt es nur als 64bit-Version, d.h. EG 6.1 ist nicht verfügbar für Windows XP und Vista

Copyright © 2014 SAS Institute Inc. All rights reserved.

What‘s new in Enterprise Miner 13.1

Copyright © 2014 SAS Institute Inc. All rights reserved.

Neue Data Mining Funktionen



* neue Knoten oder Prozeduren

Copyright © 2014 SAS Institute Inc. All rights reserved.

Komplette Liste der verfügbaren Knoten in Version 13.1

SAMPLE	Append, Data Partition, File Import, Filter, Merge, Sample, Input Data
EXPLORE	Association, Cluster, Graph Explore, Variable Clustering, DMDB, MultiPlot, Market Basket, StatExplore, Link Analysis, Path Analysis, Variable Selection, SOM/Kohonen
MODIFY	Drop, Impute, Interactive Binning, Principal Components, Replacement, Rules Builder, Transform Variables
MODEL	Decision Tree, AutoNeural, Regression, Neural Network, Partial Least Squares, Dmine Regression, DM Neural, Ensemble, Rule Induction, Gradient Boosting, LARS, MBR, Two Stage, Model Import
	Incremental Response, Survival Analysis, Credit Scoring*, TS Correlation, TS Data Prep, TS Dimension Reduction, TS Decomp., TS Similarity, TS Exponential Smoothing
	HP Explore, HP Impute, HP Regression, HP Transform, HP Variable Selection, HP Neural, HP Forest, HP Decision Tree, HP Data Partition, HP GLM, HP SVM, HP Cluster, HP Principal Components
ASSESS	Cutoff, Decisions, Model Comparison, Score, Segment Profile
UTILITY	Control Point, End Groups, Start Groups, Open Source Integration, Reporter, Score Code Export, Metadata, SAS Code, Ext Demo, Save Data, Register Metadata

* Benötigt Lizenz für "Credit Scoring for SAS Enterprise Miner Add-on"

Copyright © 2014 SAS Institute Inc. All rights reserved.

Neue Knoten/Prozeduren/Funktionen

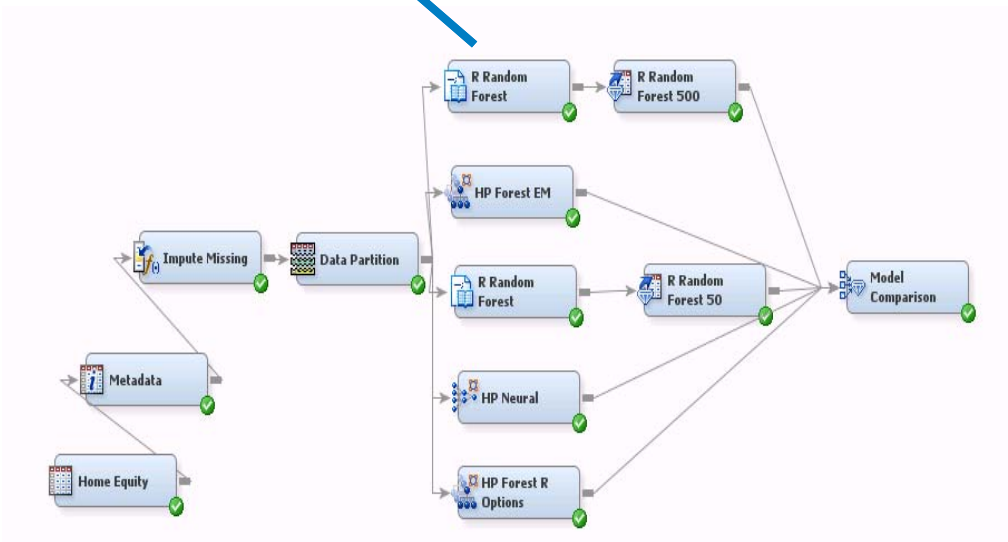
Open source integration (R Support)

- Erlaubt es den Benutzern R Code innerhalb des Enterprise Miner Prozessflusses zu integrieren (supervised und unsupervised Modelle)

- Bietet somit beispielsweise die Flexibilität die komplette Datenaufbereitung mit den mächtigen Werkzeugen des Miners umzusetzen, Modell-verfahren des Miners komplementär mit z.B. neuen Modellverfahren aus R zu kombinieren (bzw. zu vergleichen) und final das beste Modell (inkl. Scoring) für die operative Verwendung bereit zu stellen

- Modelle aus R können dabei in gewohnter EM-Manier untereinander sowie mit Modellknoten aus EM selbst verglichen werden
- Scoring-Regelwerke/Funktionen aus R-Modellen, welche den PMML Standard unterstützen, können dabei in SAS übernommen werden und Scorings somit z.B. direkt in SAS oder sogar in-database gerechnet werden (Vorteil Rechenpower)

```
library(randomForest)

&EMR_MODEL <- randomForest(&EMR_CLASS_TARGET ~  &EMR_CLASS_INPUT + &EMR_NUM_INPUT, ntree= 500, mtry= 5, data= &EMR_IMPORT_DATA, importance= TRUE)

&EMR_EXPORT_TRAIN <- predict(&EMR_MODEL, &EMR_IMPORT_DATA, type="prob")


&EMR_EXPORT_VALIDATE <- predict(&EMR_MODEL, &EMR_IMPORT_VALIDATE, type="prob")
&EMR_EXPORT_TEST <- predict(&EMR_MODEL, &EMR_IMPORT_TEST, type="prob")

&EMR_EXPORT_TRAIN[1:10,]

png("EMR_forestMsePlot.png")
plot(&EMR_MODEL, main= 'randomForest MSE Plot')
dev.off()

write.table(round(importance(&EMR_MODEL),2), file= "EMR_forestImportance.csv", sep= ",", row.names= TRUE, col.names= TRUE)

print(&EMR_MODEL)
  round(importance(&EMR_MODEL),2)
```

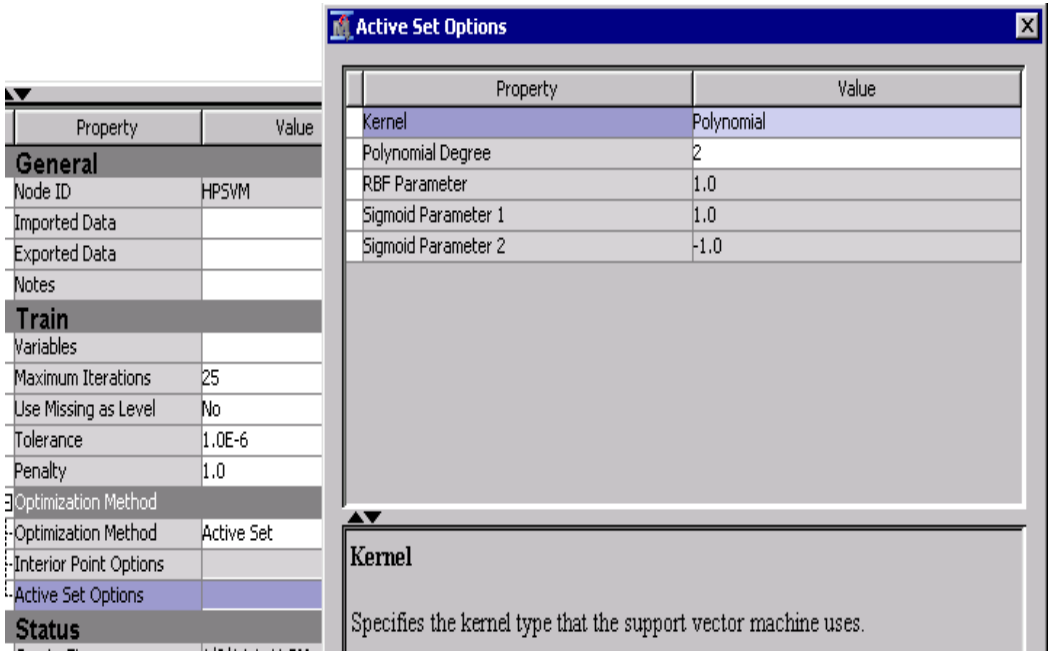


Copyright © 2014 SAS Institute Inc. All rights reserved.

Neue Knoten/Prozeduren/Funktionen

HP Support Vector Machine

- Unterstützt die Erstellung von linearen und nicht-linearen Support Vector Machine Modellen
- Konstruiert separierende Hyperebenen, welche den Abstand zwischen zwei Klassen maximieren
- Unterstützt eine Vielzahl von Kernel-Funktionen:
  linear, polynomial, radial basis function und sigmoid function.
- Als Optimierungsmethoden werden auch „interior point“ und „active set“ unterstützt



Copyright © 2014 SAS Institute Inc. All rights reserved.

Neue Knoten/Prozeduren/Funktionen

HP Generalized Linear Model

- Baut auf der Prozedur HPGENSELECT auf für die Anpassung/Berechnung generalisierter linearer Modelle
- Unterstützt Threads sowie eine verteilte Computing Umgebung
- Sämtliche Verteilungs- und Linkfunktionen sind verfügbar
- Der Knoten beinhaltet zudem Variablen-Selektionsverfahren

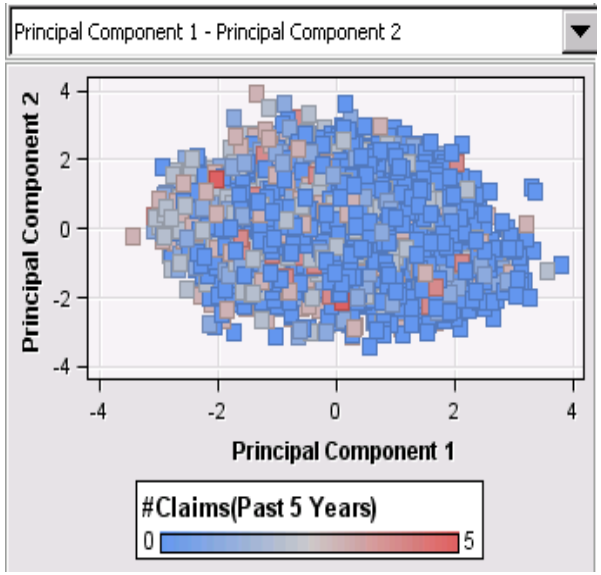
Property	Value
General	
Node ID	HPGLM
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Set Reference Level	Default
Reference Level	...
Equation	
Main Effects	Yes
Two-Factor Interactions	No
Polynomial Terms	No
Polynomial Degree	2
Suppress Intercept	No
Use Missing as Level	No
Modeling	
Interval Target Probabilit	Poisson
Interval Target Link Func	Log
Binary Target Link Functi	Logit
Optimization Options	...
Convergence Options	...
ZI Model Options	...
Tweedie Model Options	...
Model Selection	
Selection Method	Forward
Stop Criterion	DEFAULT (SL)
Selection Options	...

Copyright © 2014 SAS Institute Inc. All rights reserved.

Neue Knoten/Prozeduren/Funktionen

HP Principal Components und HP Cluster Knoten

- Auch die Hauptkomponentenanalyse und das Clustering sind nun in High Performance verfügbar, d.h. unterstützen Threads oder können über eine verteilte Computing Umgebung parallelisiert werden



Cluster Centroids

Cluster ID	AGE	BLUEBOOK	CAR_AGE	HOMEKIDS
1	47.7153	20525.15	11.38766	0.503249
2	44.57949	16567.28	8.746701	0.692308
3	46.33524	16717.82	9.018731	0.580176
4	45.41172	14163.27	7.104052	0.719665
5	43.00772	12547.33	6.02052	0.943684
6	43.23257	14626.02	7.922706	0.797468

Copyright © 2014 SAS Institute Inc. All rights reserved.

Neue Knoten/Prozeduren/Funktionen

HP Bayesian Network Prozedur

- Analysiert unterschiedliche Typen von Bayes Netzwerken: naive, treeaugmented naive (TAN), Bayesian network-augmented naive (BAN), parent-child Bayesian network und Markov blanket

- Die Prozedur ist zudem in der Lage eine effiziente Variablenselektion durchzuführen mittels Unabhängigkeitstests und es selektiert automatisch das beste resultierende Modell mit den spezifizierten Parametern

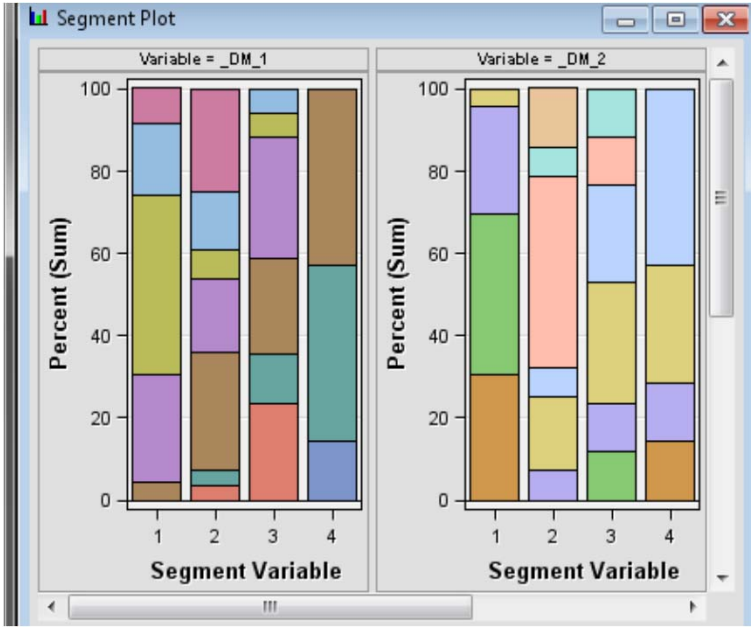
```
proc hpbnet data=sampsio.dmagecr numbin=10 alpha=0.05
         structure=Naive TAN PC MB varselect=0 1 bestmodel;
    target Good_Bad;
    input  Checking History Purpose Savings Employed Installp
           Resident Property Other Housing Existcr Job Depend
           Foreign/level=NOM;
    input Age Amount Duration/level=INT;
    partition FRACTION (VALIDATE=0.3);
    output network=network fit=fit validinfo=validinfo;
run;
```

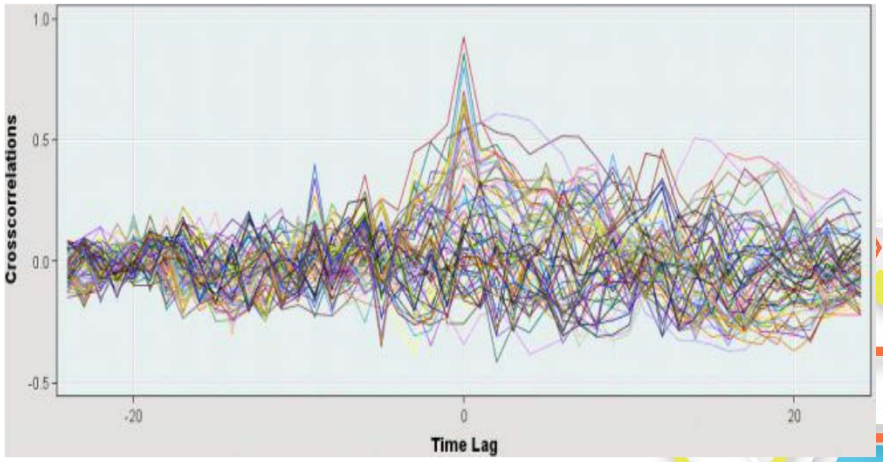
Copyright © 2014 SAS Institute Inc. All rights reserved.

Neue Knoten/Prozeduren/Funktionen

Time Series Knoten

- Der *Time Series Dimension Reduction* Knoten extrahiert Kern-Eigenschaften jeder einzelnen Zeitreihe und reduziert die Zeit-Dimension
- Der *Time Series Correlation* Knoten bietet den Benutzern die Möglichkeit Korrelationen und Kreuz-Korrelationen zu analysieren
- Der *Time Series Decomposition* Knoten bietet die Möglichkeit Zeitreihen von der Saisonalität zu bereinigen (Zerlegung)





Copyright © 2014 SAS Institute Inc. All rights reserved.

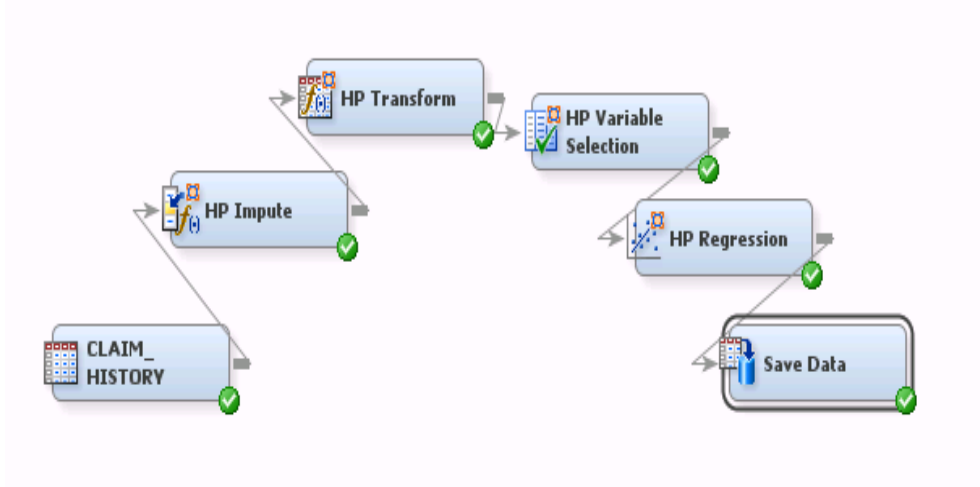
Neue Knoten/Prozeduren/Funktionen

Register Model und *Save Data* Knoten

- Erlaubt es neu ein Modell (Segmentationen, Klassifikationen und Prognosemodelle) automatisch innerhalb eines Prozessflusses im Metadaten-Server zu registrieren (statt wie bis anhin nur mittels manuellem Durchklicken)

- Registrierte Modelle können von Enterprise Guide, das Add-in für Microsoft Office, Model Manager, Enterprise Miner, Web Services und weitere Produkte/Benutzeroberflächen direkt verwendet werden

- Der *Save Data* Knoten erlaubt den automatischen Export von Tabellen (als SAS oder JMP-Tabellen, Excel-, CSV- oder TXT-Dokumente) an beliebigen Stellen eines Prozessflusses

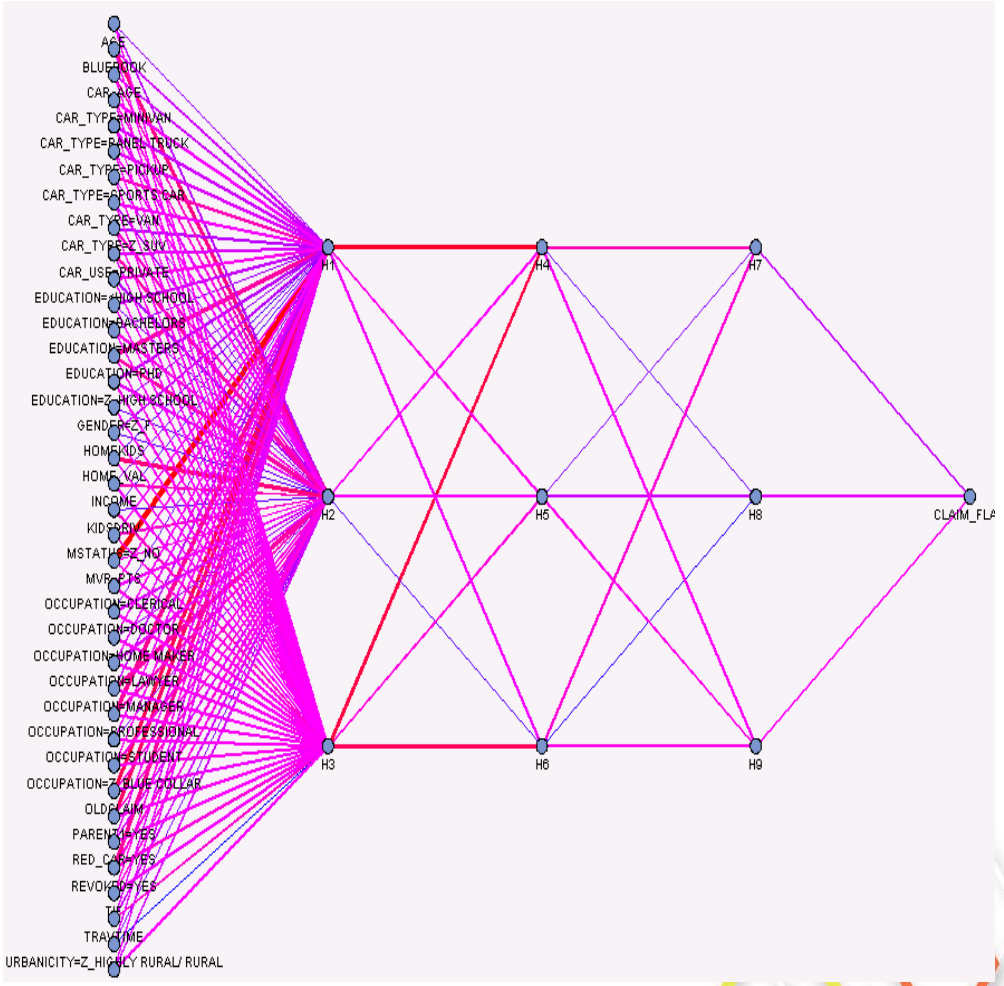


Copyright © 2014 SAS Institute Inc. All rights reserved.

Erweiterungen bestehender Knoten/Prozeduren/Funktionen

HP Neural Knoten

- Bietet eine Benutzer-definierte Architektur, welche dem Benutzer mehr Kontrolle über den Aufbau des neuronalen Netzwerks gibt
- Benutzer können die Anzahl *Hidden Layers*, die Anzahl *Hidden Neurons*, und die zugehörigen *Activation Functions* für jeden Layer bestimmen
- Benutzer können neu Standardisierungen für erklärende und Zielvariablen, *Target Error* und weitere Parameter steuern

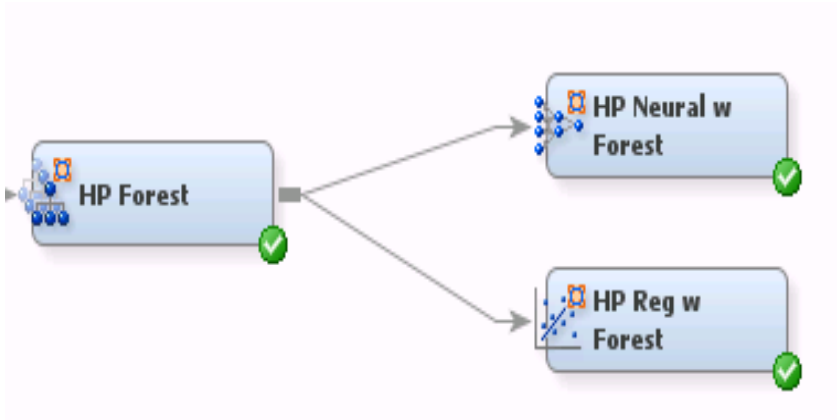


Copyright © 2014 SAS Institute Inc. All rights reserved.

Erweiterungen bestehender Knoten/Prozeduren/Funktionen

HP Forest Knoten

- Neu können Variablenselektionen mittels *OOB Average Error* für Intervallvariablen oder *OOB Marginal Reduction* für Klassenvariablen verwendet werden



Copyright © 2014 SAS Institute Inc. All rights reserved.

SAS® Visual Statistics

Copyright © 2014 SAS Institute Inc. All rights reserved.

DEMO

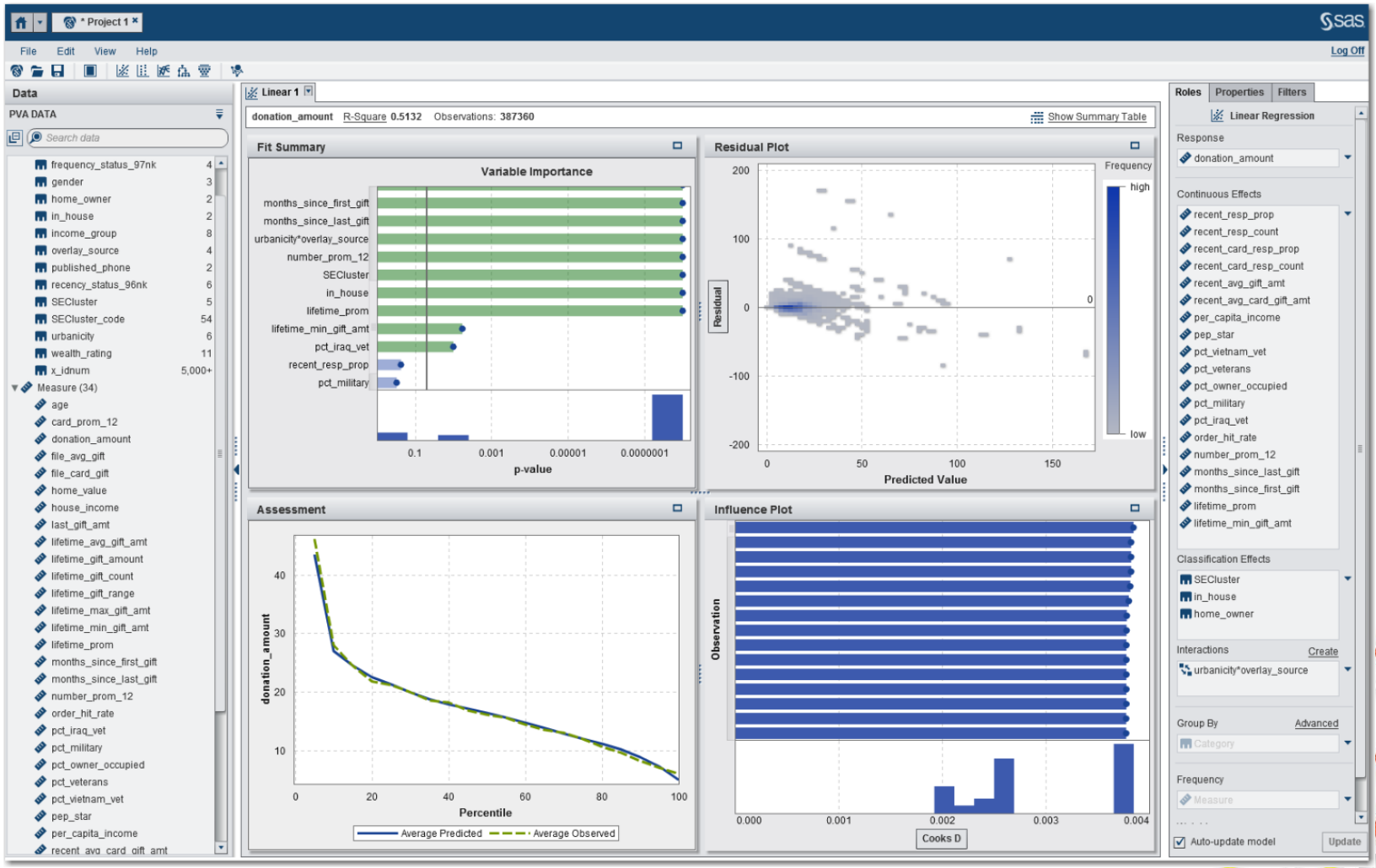
Zielsetzungen

- Erweiterung von SAS® Visual Analytics um statistische Modellierungsverfahren
- Webbrowser-basierte Bedienoberfläche (zero footprint client) für Statistiker und Analytiker
- Intuitiv zu bedienende und hochperformante Umgebung
- Vollständig integriert mit Visual Analytics und dem SAS Metadaten-Server und somit mit der gesamten SAS Welt

Copyright © 2014 SAS Institute Inc. All rights reserved.

Ein paar Screenshots zu SAS® Visual Statistics

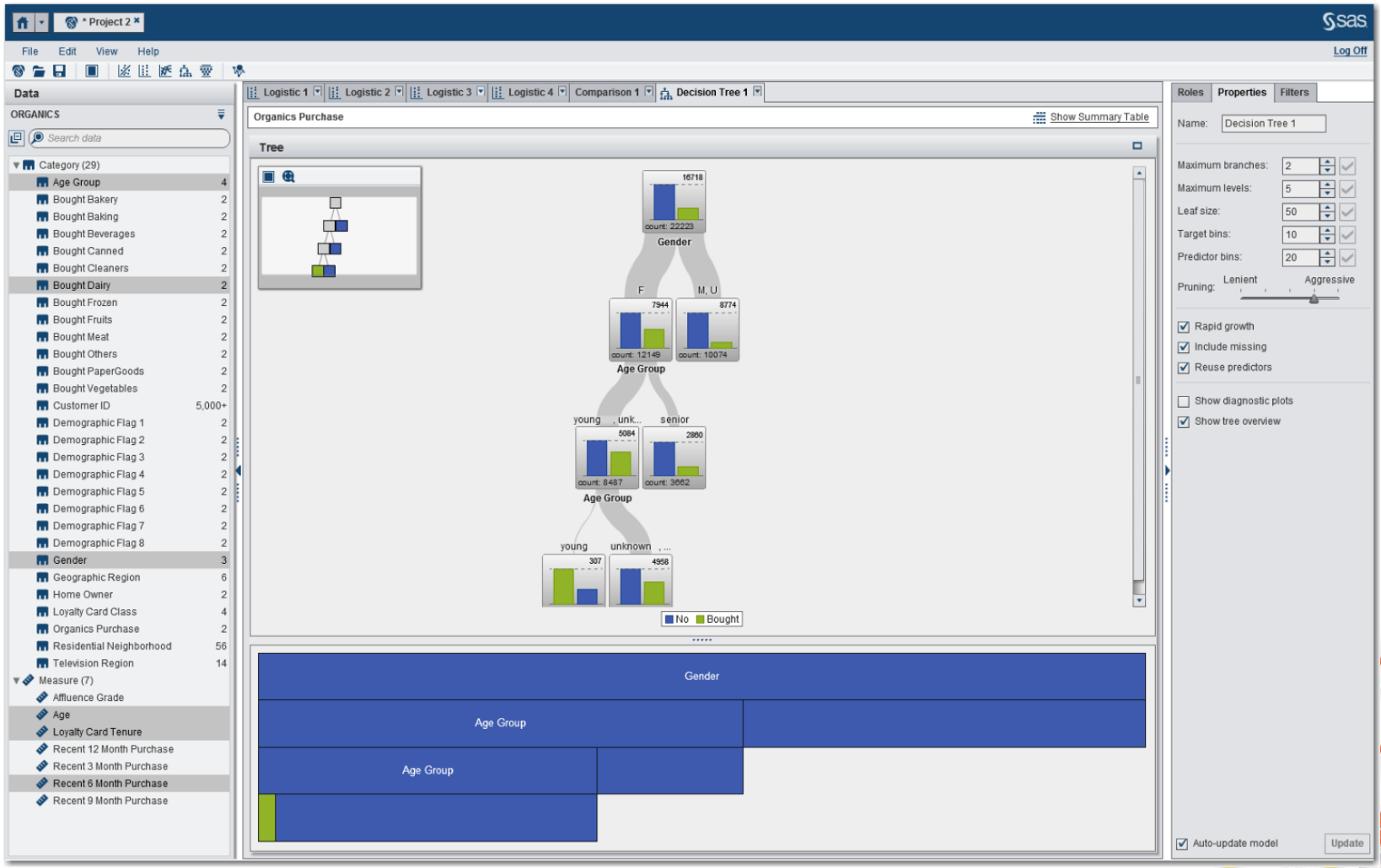
Lineare Regression



Copyright © 2014 SAS Institute Inc. All rights reserved.

Ein paar Screenshots zu SAS® Visual Statistics

Entscheidungsbaum



Copyright © 2014 SAS Institute Inc. All rights reserved.

Ein paar Screenshots zu SAS® Visual Statistics

Modellvalidierung

Selected	Model	1-ACC	C	KS	Lift	CumLift	FPR	FDR	F1	Gini	Gamma	Tau	CumPctResp	CumResp	Gain
	Logistic 1	0.247716	0.574716	0.110394	1.28165	1.28165	0	0	0	0.149431	0.167209	0.055696	31.72429	6.408248	0.28165
	Logistic 2	0.196155	0.775181	0.418813	3.466942	3.466942	0.04518	0.29003	0.45969	0.550362	0.557819	0.203903	85.09128	17.33471	2.466942
	Logistic 3	0.196035	0.775668	0.418107	3.463476	3.463476	0.045081	0.291685	0.457581	0.551336	0.5588	0.203792	84.70226	17.31738	2.463476
Y	Logistic 4	0.185199	0.811926	0.468239	3.547439	3.547439	0.055144	0.291727	0.522004	0.623851	0.630769	0.230596	86.75565	17.7372	2.547439

Copyright © 2014 SAS Institute Inc. All rights reserved.

SAS® Forum
Switzerland 2014
make connections • share ideas • be inspired

Bei Fragen einfach fragen

Practice Leader Analytics, SAS Institute AG
Carmelo.Iantosca@sas.com

Kontaktieren Sie mich via:

Copyright © 2014, SAS Institute Inc. All rights reserved.