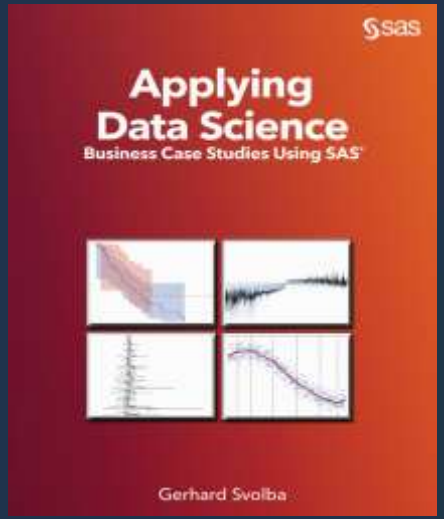
27. SAS Club



Buchpräsentation:
Applying Data Science Business Case Studies Using SAS

Gerhard Svolba, Franz Helmreich, Gernot Engel,
Matthias Svolba, Mihai Paunescu

Wien, 23. November 2017 – ARES Tower, Wien

sas
THE POWER TO KNOW.

SAS Tipps und Tricks Session

Mihai Paunescu (Bundesministerium für Finanzen)

Gerhard Svolba (SAS)

Listen to Your Data!

Unsupervised Machine Learning Techniken zeigen Ihnen Zusammenhänge in Ihren Daten

(Mihai Paunescu)

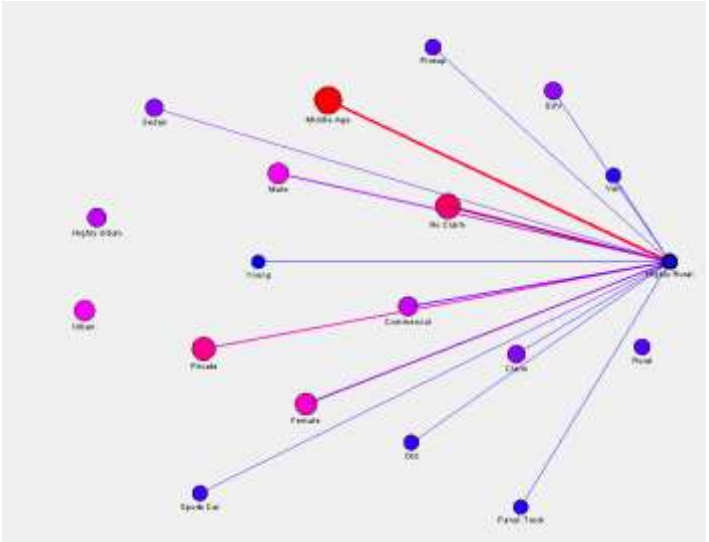
Lassen Sie ihre Daten sprechen!

Auffinden von Zusammenhängen in Ihren Analysedaten

- Daten aus der KFZ-Versicherung mit 6 Eigenschaften pro Versicherungsnehmer

Variable	Feature
AGE	YOUNG, MIDLIFE, OLD
GENDER	MALE, FEMALE
DENSITY	HIGHLY URBAN, URBAN, HIGHLY RURAL, RURAL
CAR_TYPE	VAN, SPORTS CAR, SUV, SEDAN, PICK UP
CAR_USAGE	PRIVATE, COMMERICIAL
CLM_FLAG	CLAIM, NO CLAIM

- Anwenden von unsupervised machine learning (Assoziationsanalyse) um Zusammenhänge zwischen den Eigenschaften aufzudecken.



Vorbereitung der Daten: von one-row-per subject in eine multiple-row-per-subject Struktur

POLICYNO	CLM_FLAG	CAR_USE	CAR_TYPE	AGE	GENDER	DENSITY
160	No	Private	Sedan	60	M	Highly Urban
24836	No	Commercial	Sedan	43	M	Highly Urban
28046	No	Private	Van	48	M	Urban
28960	No	Private	SUV	35	F	Highly Urban
40933	No	Private	Sedan	51	M	Highly Urban
55277	No	Private	SUV	50	F	Urban
63212	Yes	Commercial	Sports Car	34	F	Highly Urban
69651	No	Private	SUV	54	F	Highly Urban
88070	Yes	Private	Sedan	40	M	Urban
93553	No	Commercial	SUV	44	F	Rural
127444	Yes	Commercial	Van	37	M	Highly Urban
141509	Yes	Private	SUV	34	F	Highly Urban
145326	No	Commercial	Van	50	M	Rural
146809	Yes	Private	Sports Car	53	F	Urban
148250	No	Private	Sedan	43	F	Rural
157851	No	Commercial	Van	55	M	Urban

POLICYNO	Feature
160	Highly Urban
160	No Claim
160	Sedan
160	Private
160	Male
160	Old
24836	Highly Urban
24836	No Claim
24836	Sedan
24836	Commercial
24836	Male
24836	Middle Age

Vorbereitung der Daten: von one-row-per subject in eine multiple-row-per-subject Struktur: SAS Code

SAS Datastep

```
data claims_feature(keep = policyno
                           feature);
 set claims_nodup;
 format Feature $40.;
 *** 1. Gender;
 if gender = 'M' then Feature = 'Male';
  else Feature = 'Female';
 output;
*** 2. Age;
 if 0 < Age < 26 then feature = 'Young';
  else if 26 <= age <= 55 then feature =
   'Middle Age';
  else feature = 'Old';
 output;
 *** 3. Density;
 feature = Density; output;
 *** 4. Car Type;
 feature = Car_type; output;
 *** 5. Car Use;
 feature = Car_use; output;
 *** 6. Claim Flag;
 if clm_flag = 'Yes' then feature =
   'Claim';
 else feature = 'No Claim';
 output;
run;
```

oder

PROC TRANSPOSE

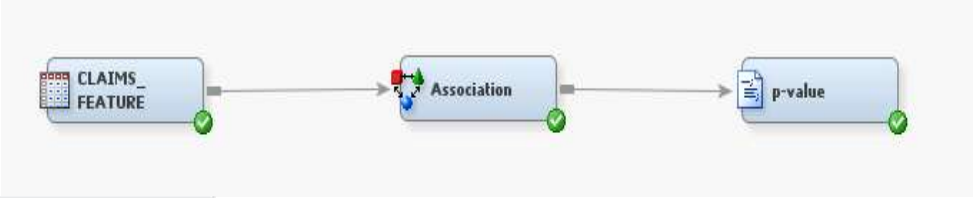
```
data claims_nodup2;
 set claims_nodup;
 Age=round(Age,10);
run;


proc transpose data=claims_nodup2
               out=claims_Long;
 by policyno ;
 var age gender Density car_type car_use
     clm_flag;
run;

data Key_Value(drop =_label_ col1
               rename=(_name_ = Key));
 set claims_Long;
  Value = strip(col1);
  Feature = catx('=',_name_,Value);
run;
```

Association Analysis zur Auffinden der Kombinationen

Unsupervised Machine Learning mit dem SAS Enterprise Miner



ID	claimsfeature
Name	CLAIMS_FEATURE
Variables	
Decisions	
Role	Transaction

Variables - CLAIMS_FEATURE

(Ohne) not Ist gleich

Columns: Label

Name	Role	Level
Feature	**Target**	**Nominal**
POLICYNO	**ID**	**Interval**

Association	
Maximum Items	2
Minimum Confidence Level	1
Support Type	Count
Support Count	1
Support Percentage	5.0
Sequence	
Rules	
Number to Keep	10000
Sort Criterion	Default
Number to Transpose	5000
Export Rule by ID	No
Recommendation	No

Lassen Sie ihre Daten sprechen!

Männer fahren keine Sportwägen?

Rule 278 shows that sports cars are only driven in 2.54% of the cases by men, whereas this was expected in around 46% of the cases.

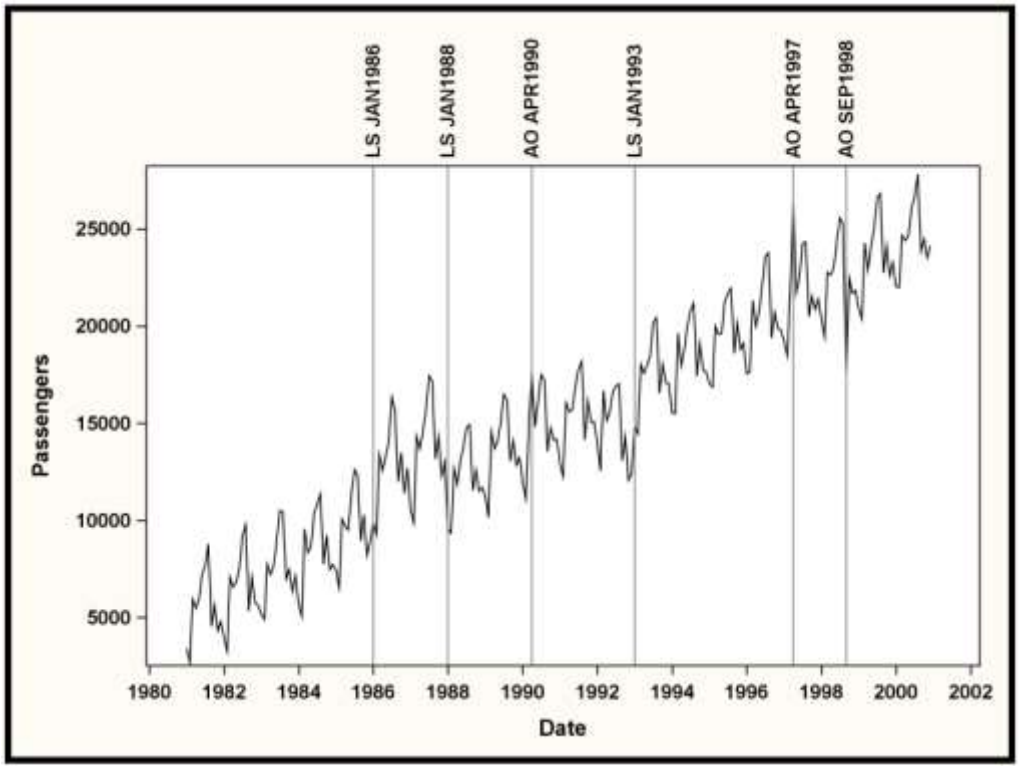
index	RULE	_LHAND	_RHAND	COUNT	SUPPORT	EXP_CONF	CONF	LIFT
267	Commercial ==> Sports Car	Commercial	Sports Car	200.00	1.94	11.44	5.28	0.46
268	Rural ==> Claim	Rural	Claim	102.00	0.99	26.66	6.52	0.24
269	Claim ==> Rural	Claim	Rural	102.00	0.99	15.18	3.71	0.24
270	Young ==> Highly Urban	Young	Highly Urban	10.00	0.10	34.93	8.33	0.24
271	Highly Rural ==> Claim	Highly Rural	Claim	32.00	0.31	26.66	6.30	0.24
272	Claim ==> Highly Rural	Claim	Highly Rural	32.00	0.31	4.93	1.17	0.24
273	Van ==> Female	Van	Female	117.00	1.14	53.82	12.70	0.24
274	Female ==> Van	Female	Van	117.00	1.14	8.94	2.11	0.24
275	Panel Truck ==> Female	Panel Truck	Female	40.00	0.39	53.82	4.69	0.09
276	Male ==> SUV	Male	SUV	99.00	0.96	27.98	2.08	0.07
277	SUV ==> Male	SUV	Male	99.00	0.96	46.18	3.43	0.07
278	Sports Car ==> Male	Sports Car	Male	30.00	0.29	46.18	2.54	0.06

- This might indicate a situation that for the customer base, sports cars are really predominantly driven by women.
- It could be a trigger to an investigation of the quality status of your data.
- A business interpretation could be that in a family, the sports car is the 2nd or 3rd car that is registered in the wife's name for financial reasons.
- The competitor is offering a policy to men for a much more attractive price.

Erzeugen Sie Ihre individuellen Simulationsdaten mit SAS

Erzeugen Sie Ihre individuellen Simulationsdaten mit SAS

- Saisonalität
- Trend
- Level Shifts
- Ausreißer

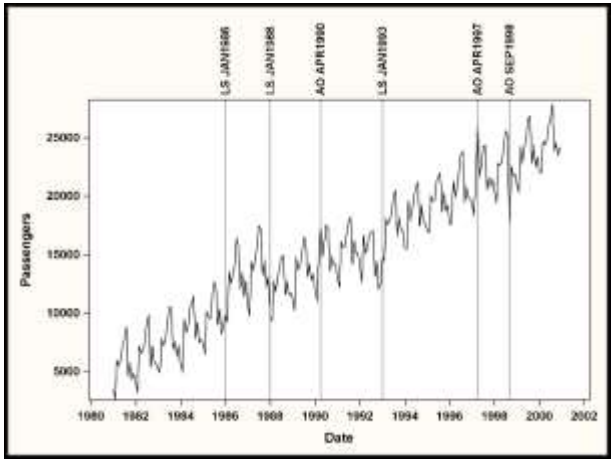


http://www.sascommunity.org/wiki/A_simple_and_powerful_way_to_simulate_your_individual_time_series_data

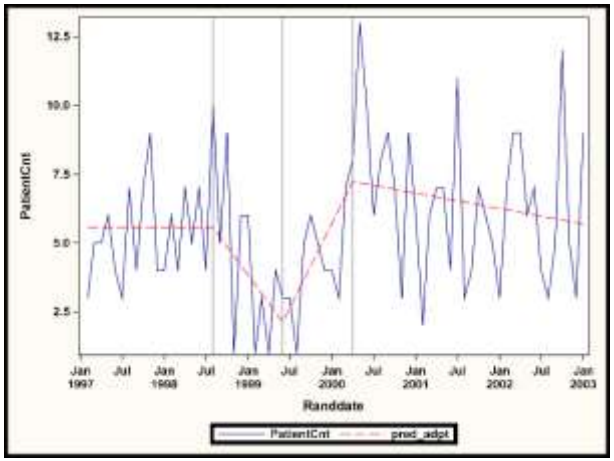
Automatisches Erkennen von Ausreißern und Break-Points mit SAS Analytics

Automatisches Erkennen von Breakpoints und Ausreißern

Anwenden von analytischen Methoden zum Erkennen von Zeitpunkten wo der Verlauf der Daten vom „normalen“ Muster abweicht.



Erkennen von Shifts und Pulse Events mit ARIMA Modellen



Verwenden von Multivariaten Regression Splines zum Auffinden von Bruchpunkten

Coding Tipp: Automatisches Anzeigen der vertikalen Referenz-Linien bei den jeweiligen Breakpoints (3 Schritte)

1

```
proc adaptivereg data=patients_1997_2002
                 plots=all details=bases ;
 format randdate date9.;
 model PatientsCnt = randdate  / maxbasis=100;
 output out=recruit_adpt predicted=pred;
 ods output BWDParams=BWDParams;
run;
```

	Name	Coefficient	Parent	Variable	Knot
1	Basis0	5.6580		Intercept	.
2	Basis1	0.02808	Basis0	Randdate	14396
3	Basis3	-0.01830	Basis0	Randdate	14701
4	Basis5	-0.01131	Basis0	Randdate	14092

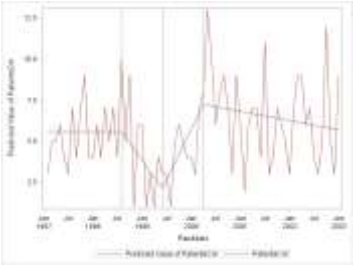
2

```
filename reflines 'c:/tmp/reflines.sas';
data _NULL_;
 set bwdparams;
 where upcase(variable) eq upcase('randdate');
 format knot 8.;
 file reflines;
  put @04 "refline " knot " / axis = x;";
run;
```

```
refline 14396  / axis = x;
refline 14701  / axis = x;
refline 14092  / axis = x;
```

3

```
proc sgplot data=recruit_adpt;
 series x=randdate y=pred;
 series x=randdate y=PatientsCnt;
 %include reflines;
run;
```



Key Takeaways

Analytics und Data Science sind da um Ihnen zu helfen!

- Sie sehen ein klareres, objektiveres Bild Ihrer Daten und Analyse-Subjekte
- Sie erhalten explizite Ergebnisse anstatt die Nadel im Heuhaufen zu suchen
- Die Daten sprechen zu Ihnen und Sie erhalten die Ergebnisse automatisch statt manuell
- Do it again! – Behandeln Sie Ihre Modelle als “Asset” und wiederholen Sie Ihre Analyse

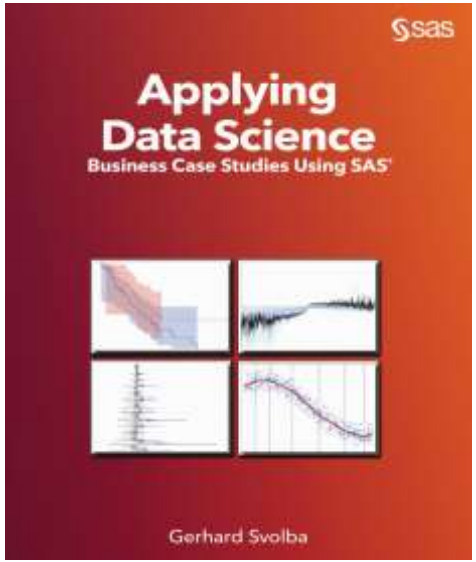
Machine Learning and Data Science sind das Kernstück der SAS Analytic Platform

- Umfassendes Set an Methoden – Entdecken und Produktivstellen
- Offen für unterschiedliche Benutzertypen (Coding, Point&Click, SAS, R, Python, …)

More Information

Gerhard Svolba – Principal Analytic Solutions Architect

sastools.by.gerhard@gmx.net



- Applying Data Science – Business Case Studies Using SAS, SAS Press 2017
- Eight Case Studies showing how Data Science and Analytics can be applied to provide insight into yout data and improve your business decisions
- http://www.sascommunity.org/wiki/Applying_Data_Science_-_Business_Case_Studies_Using_SAS