

Credit Score Modeling

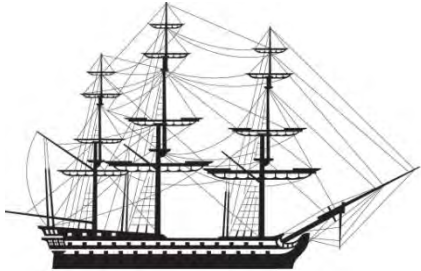
in SAS Enterprise Miner

Xiaoyuan (Tina) Zhang

What am I?



CFA



Vanguard®



Credit Score Modeling

in SAS Enterprise Miner

Xiaoyuan (Tina) Zhang



scorecard2017

- Data Sources
 - CS_ACCEPTS
 - CS_REJECTS
- Diagrams
 - 01
 - Final
- Model Packages

.. Property Value



Sample Explore Modify Model Assess Utility Credit Scoring HPDM Applications Text Mining Time Series

Final



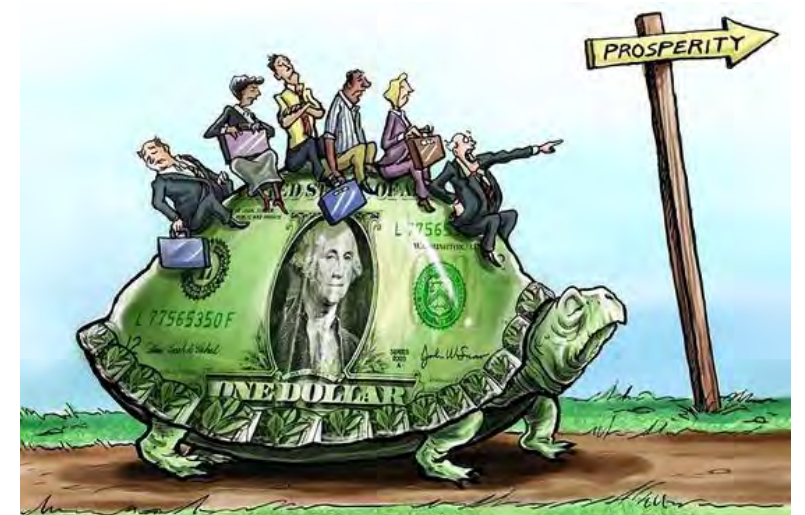
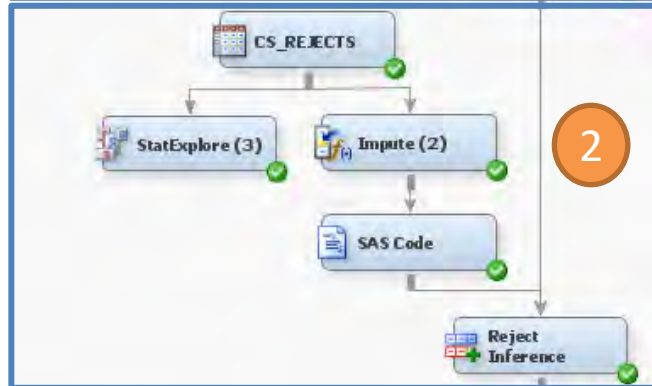
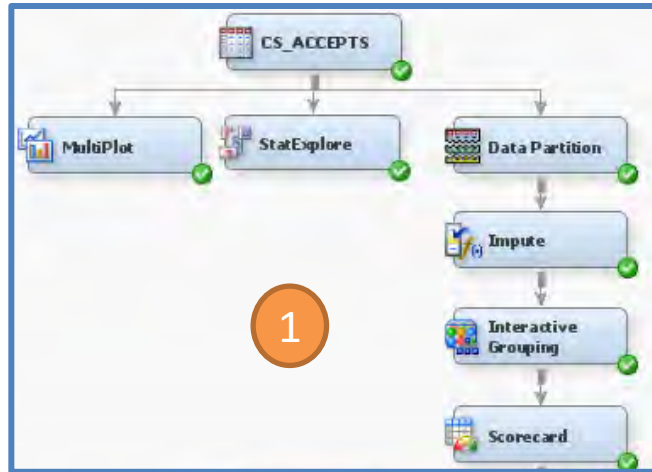
Diagram Log



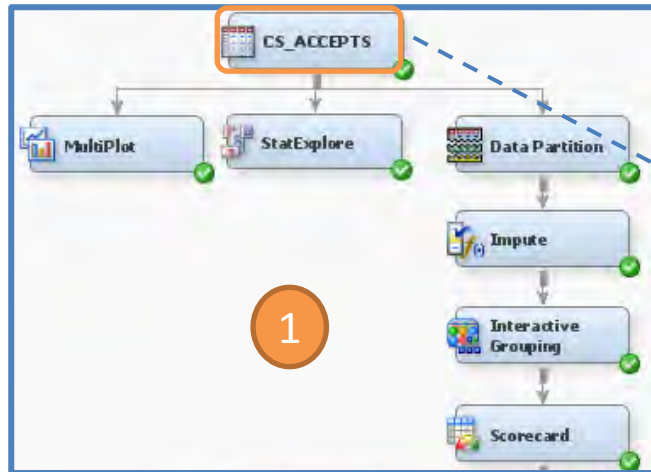
Credit Score Modeling Process

Interactive Grouping With WOE & IV

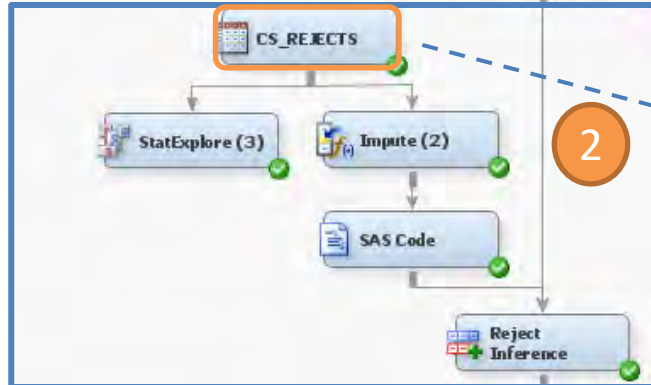
Credit Score Modeling Process



Credit Score Modeling Process



- Records of accepted loan applicants
- **3000** observations

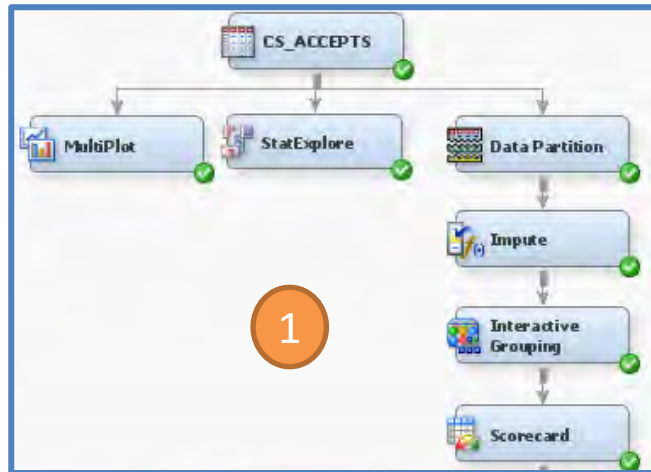


- Records of rejected loan applicants
- **1500** observations

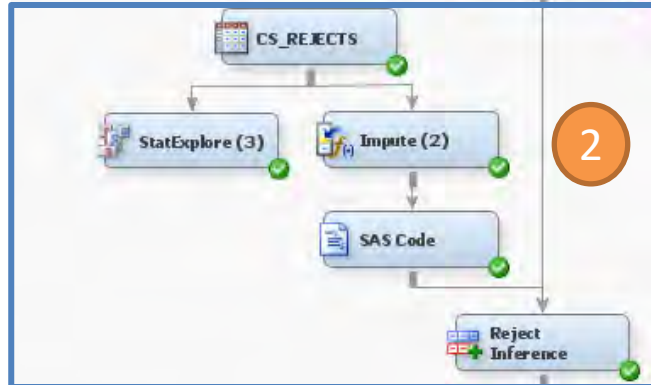


Variable Name ▼	Label	Measurement Level	Role
TMJOB1	Time at Job	Interval	Input
TMADD	Time at Address	Interval	Input
TITLE	Title	Nominal	Input
TEL	Telephone	Nominal	Input
STATUS	Status	Nominal	Input
RESID	Residence Type	Nominal	Input
REGN	Region	Nominal	Input
PROF	Profession	Nominal	Input
PRODUCT	Type of Business	Nominal	Input
PERS_H	Num in Household	Nominal	Input
NMBLOAN	Num Mybank Loans	Ordinal	Input
NAT	Nationality	Nominal	Input
LOCATION	Location of Credit Bureau	Nominal	Input
LOANS	Num of running loans	Ordinal	Input
INCOME	Income	Interval	Input
INC1	Salary+ec_card	Nominal	Input
INC	Salary	Interval	Input
GB	Good/Bad	Binary	Target
FINLOAN	Num finished Loans	Nominal	Input
EC_CARD	EC_card holders	Binary	Input
DIV	Large region	Binary	Input
CHILDREN	Num of Children	Nominal	Input
CASH	Requested cash	Interval	Input
CARDS	Credit Cards	Nominal	Input
CAR	Type of Vehicle	Nominal	Input
BUREAU	Credit Bureau Risk Class	Ordinal	Input
AGE	Age	Interval	Input

Credit Score Modeling Process



Step 1: Build initial scorecard with accepted application data

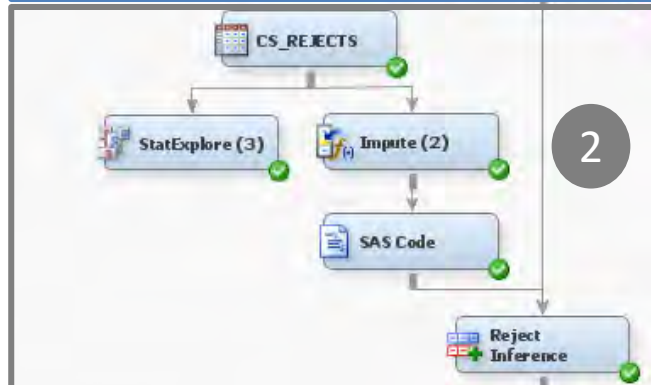
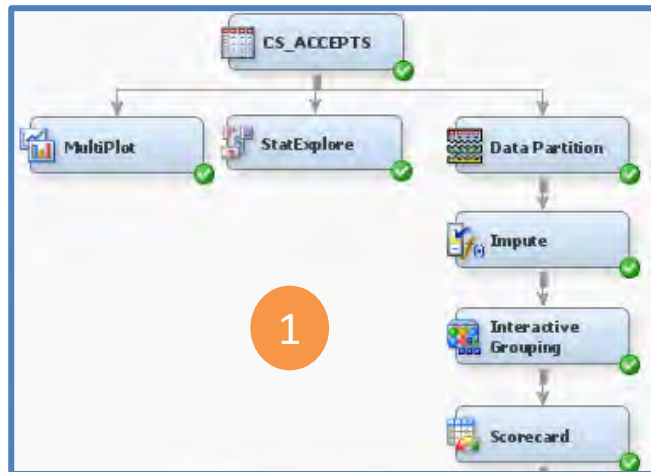


Step 2: Infer the behavior (good or bad) of the rejected applicants



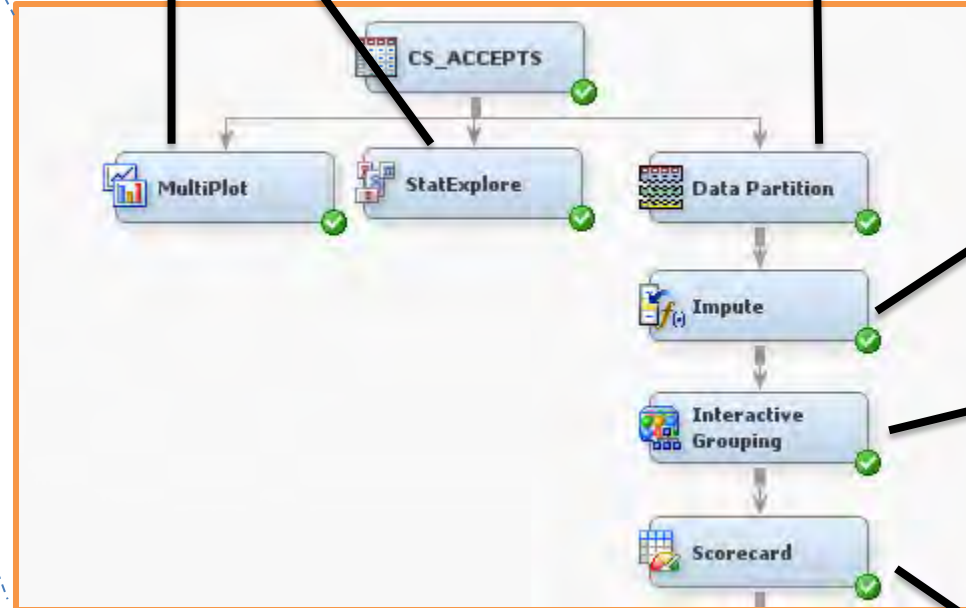
Step 3: Build the final scorecard with aggregated data combining accepted applications and rejected applications

Credit Score Modeling Process- Step 1



Exploratory data analysis

Partition the input data into training and validation sets

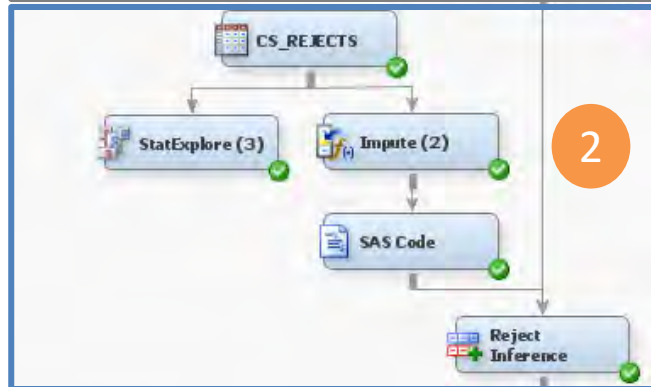
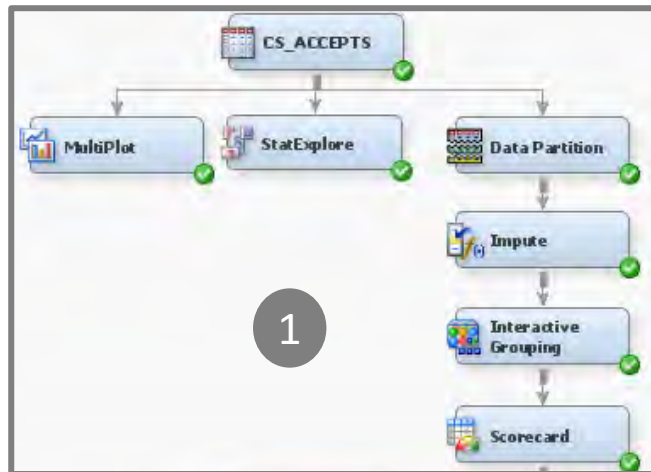


Impute missing values
Tree aggregate
Median

Group input variables
into bins;
Select variables for
scorecard

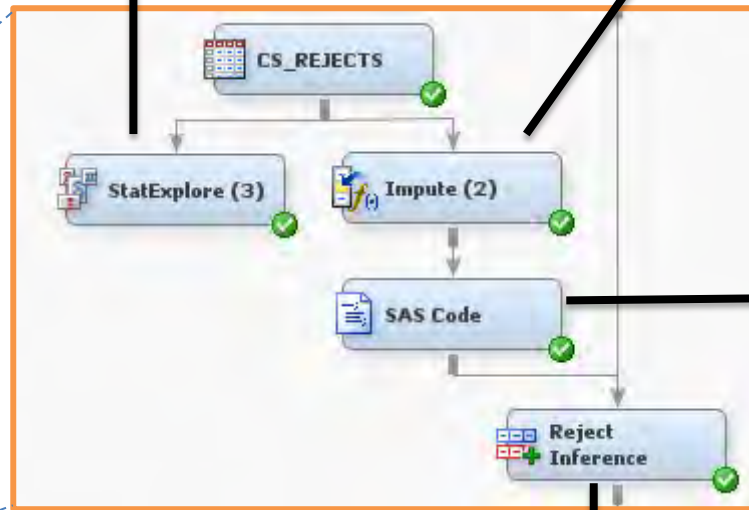
Build a logistic
regression model with
the grouped variables
as input

Credit Score Modeling Process- Step 2



Exploratory data analysis

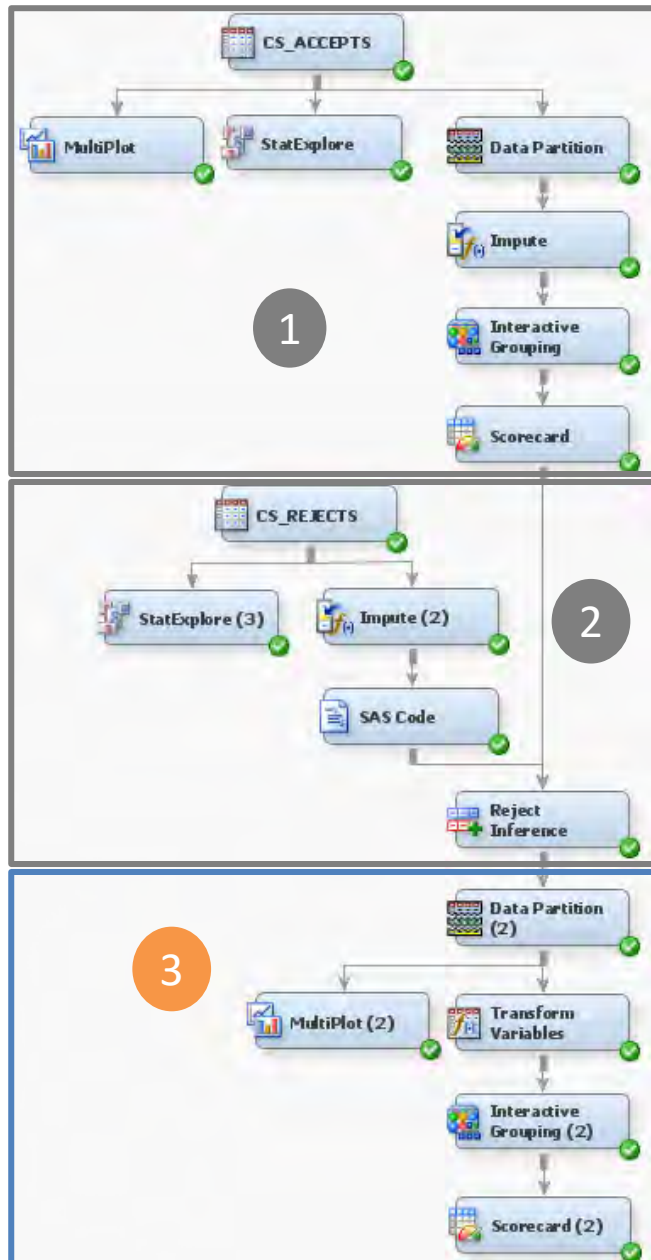
Impute values for observations with missing values



Convert data role from "raw" to "Score" so that it can be used in Reject Inference Node

Infer the performance of the rejected applicant data by the use of a model that is built on accepted applicants

Credit Score Modeling Process- Step 3



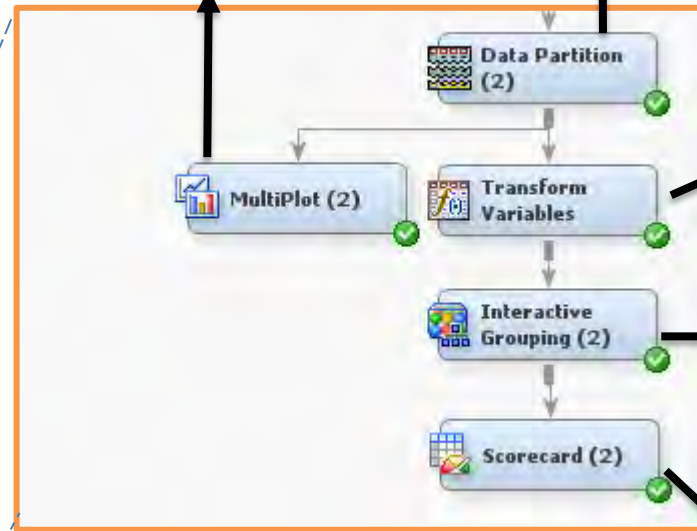
Partition the input data into training and validation sets

Exploratory data analysis

Transform input data variables to better fit the model

Group transformed input variables into bins before the credit scorecard is built; Screen variables

Build scorecard using the aggregated data as input in a logistic regression model

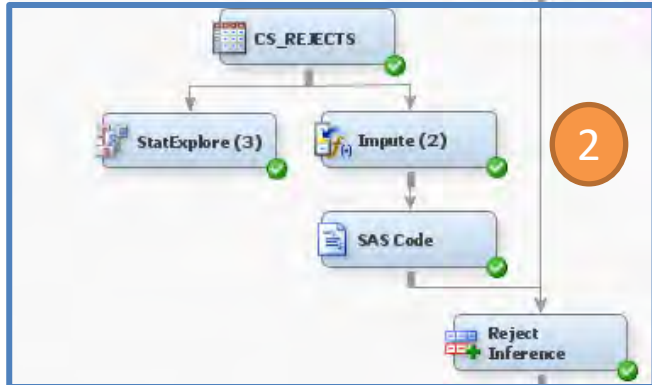
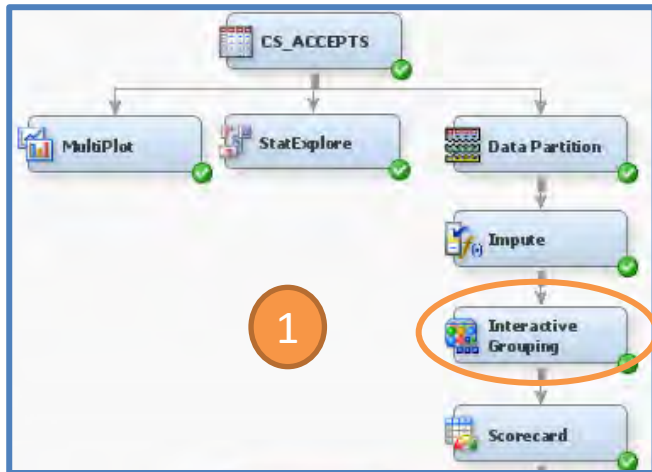


Credit Score Modeling Process



Interactive Grouping With WOE & IV

Interactive Grouping with WOE and IV



WOE



Binning/Grouping

$$\text{WOE for a bin} = \left[\ln \left(\frac{\% \text{ of total bads in bin } K}{\% \text{ of total goods in bin } K} \right) \right] * 100$$

Merge bins with similar WOE

Information Value



Variable Selection

$$\text{IV} = \sum_{k=1}^{\# \text{ of bins}} (\% \text{ of total goods in bin } K - \% \text{ of total bads in bin } K) * \ln \left(\frac{\% \text{ of total goods in bin } K}{\% \text{ of total bads in bin } K} \right)$$

Rank variables on the basis of their importance

Interactive Grouping with WOE and IV

How to pre-bin interval variables?

Interval Variable Binning Options	
Apply Level Rule	No
Binning Method	Quantile
Number of Bins	20

Method choose for different variables?

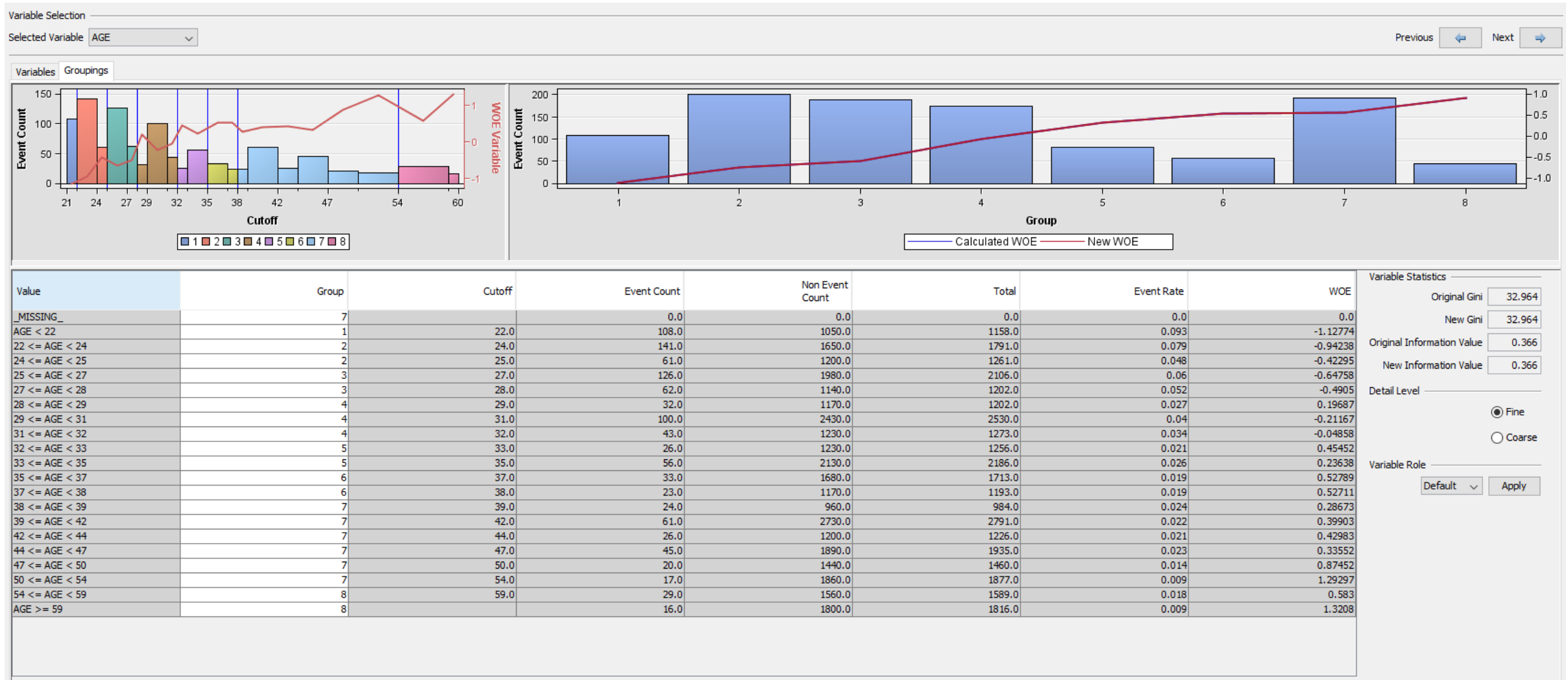
Grouping Options	
Interval Grouping Method	Monotonic Event Rate
Ordinal Grouping Method	Monotonic Event Rate
Tree Based Grouping Options	...
Constrained Optimal Options	...
Advanced Constrained Optimal	...
Maximum Number of Groups	20

Tree Based Grouping Options	
Property	Value
Criterion	Entropy
Missing Values	Separate Branch if Any
Minimum Categorical Size	10
Node Sample	2000

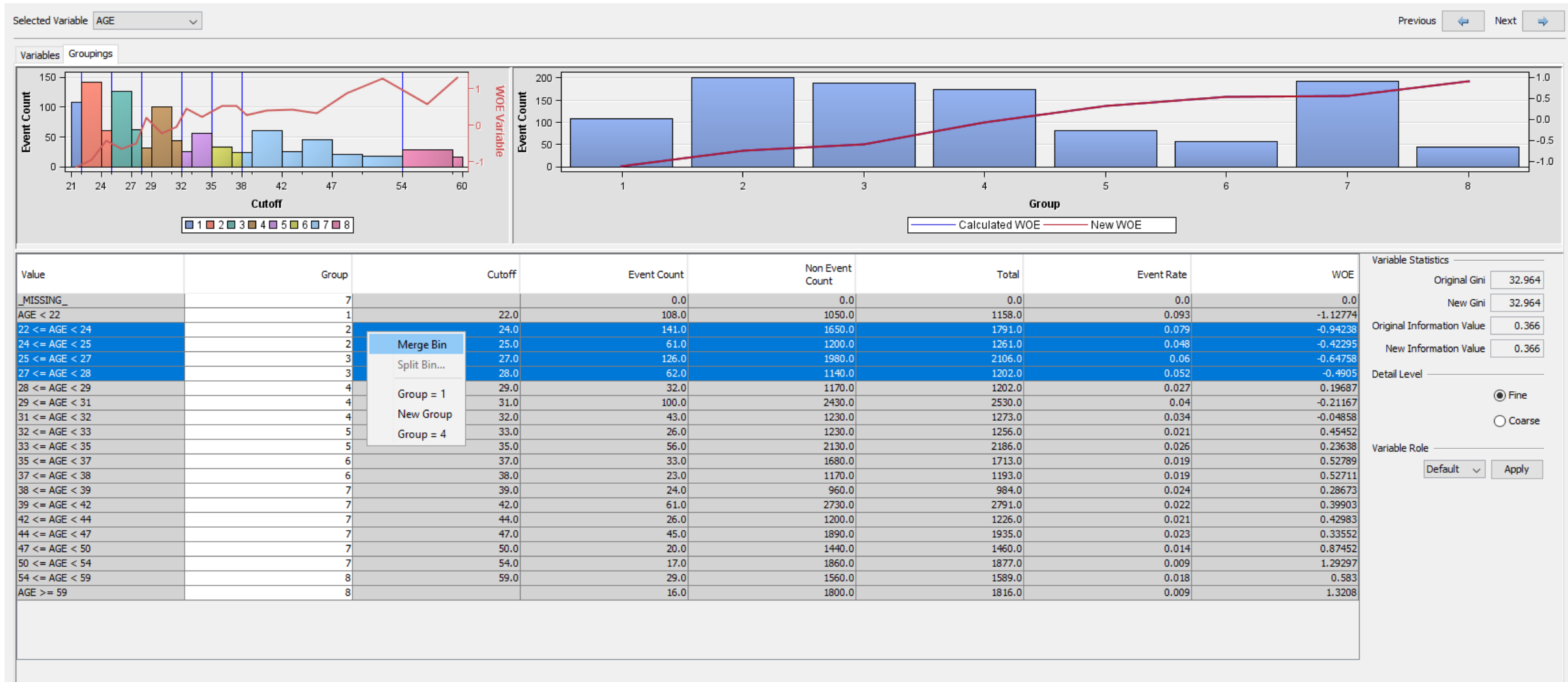
Standard for variable selection?

Score	
Group Level	Ordinal
Variable Selection Method	Information Value
Gini Cutoff	20.0
Information Value Cutoff	0.1

Interactive Grouping with WOE and IV

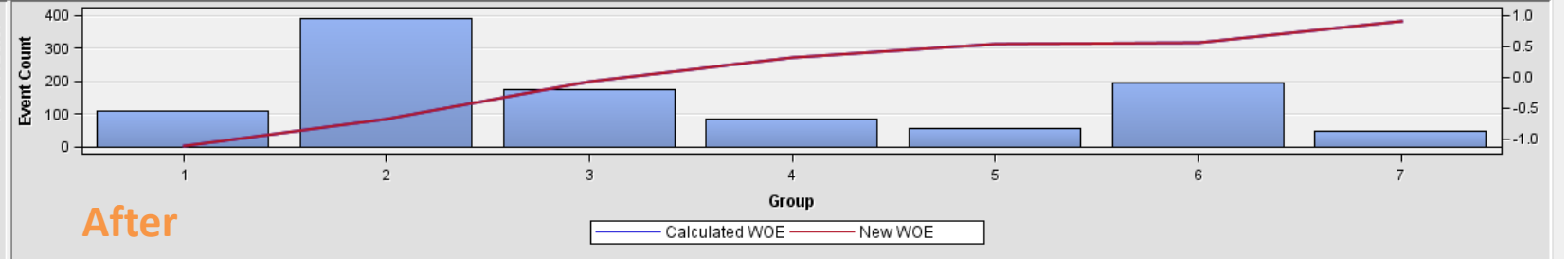
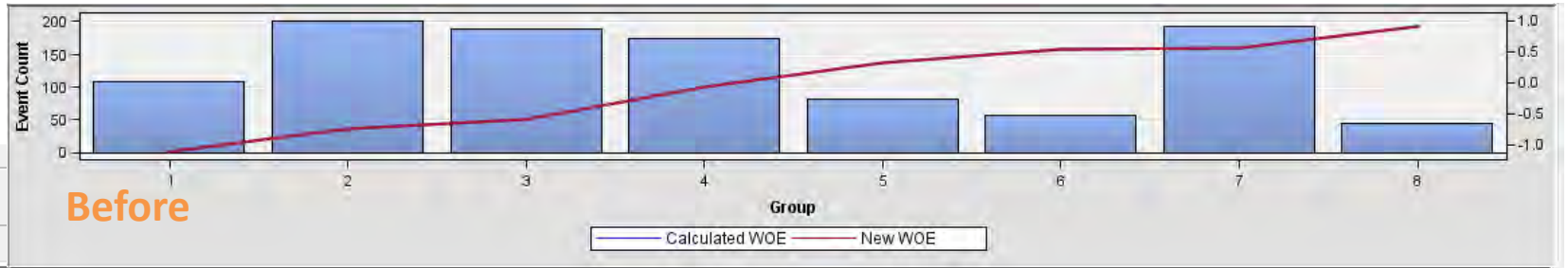
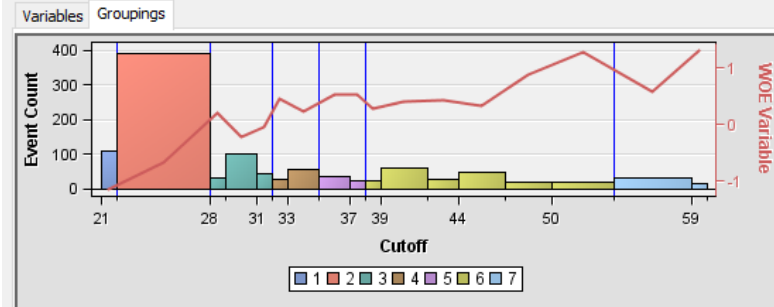


Interactive Grouping with WOE and IV



Interactive Grouping with WOE and IV

Variable Selection
Selected Variable: AGE



Value	Group	Cutoff	Event Count	Non Event Count	Total	Event Rate	WOE
MISSING	6		0.0	0.0	0.0	0.0	0.0
AGE < 22	1	22.0	108.0	1050.0	1158.0	0.093	-1.12774
22 <= AGE < 28	2	28.0	390.0	5970.0	6360.0	0.061	-0.67379
28 <= AGE < 29	3	29.0	32.0	1170.0	1202.0	0.027	0.19687
29 <= AGE < 31	3	31.0	100.0	2430.0	2530.0	0.04	-0.21167
31 <= AGE < 32	3	32.0	43.0	1230.0	1273.0	0.034	-0.04858
32 <= AGE < 33	4	33.0	26.0	1230.0	1256.0	0.021	0.45452
33 <= AGE < 35	4	35.0	56.0	2130.0	2186.0	0.026	0.23638
35 <= AGE < 37	5	37.0	33.0	1680.0	1713.0	0.019	0.52789
37 <= AGE < 38	5	38.0	23.0	1170.0	1193.0	0.019	0.52711
38 <= AGE < 39	6	39.0	24.0	960.0	984.0	0.024	0.28673
39 <= AGE < 42	6	42.0	61.0	2730.0	2791.0	0.022	0.39903
42 <= AGE < 44	6	44.0	26.0	1200.0	1226.0	0.021	0.42983
44 <= AGE < 47	6	47.0	45.0	1890.0	1935.0	0.023	0.33552
47 <= AGE < 50	6	50.0	20.0	1440.0	1460.0	0.014	0.87452
50 <= AGE < 54	6	54.0	17.0	1860.0	1877.0	0.009	1.29297
54 <= AGE < 59	7	59.0	29.0	1560.0	1589.0	0.018	0.583
59 <= AGE	7		16.0	1800.0	1816.0	0.009	1.3208

Variable Statistics

Original Gini	32.964
New Gini	32.678
Original Information Value	0.366
New Information Value	0.364

Detail Level

Fine

Coarse

Variable Role

Default

Thank you!

 Email: zhangxy.tina@gmail.com

 <https://www.linkedin.com/in/zhangxiaoyuan>