SAS Enterprise Miner (SAS EM) 開始最重要的四步驟

by TKU IM EMBA Students (2012)

Version 2012.04.13

SAS_EM 開始最重要四步驟：

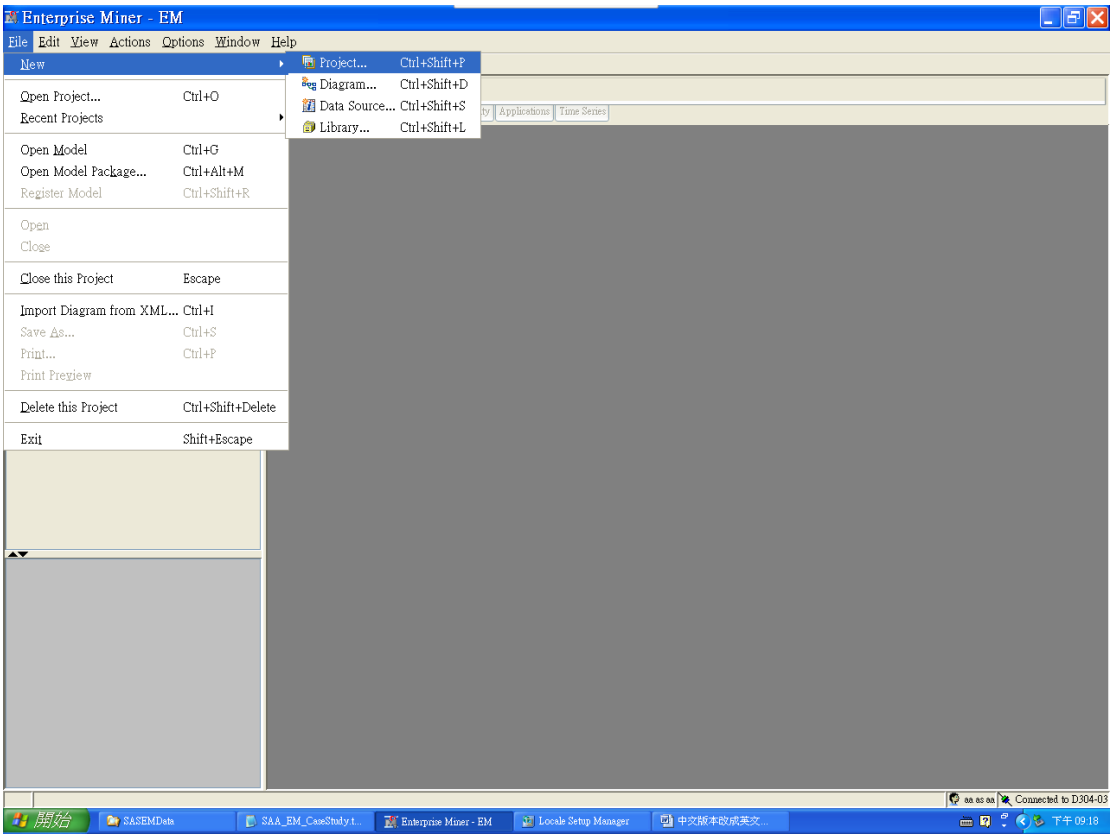
Step 1. Create Project (建 project)

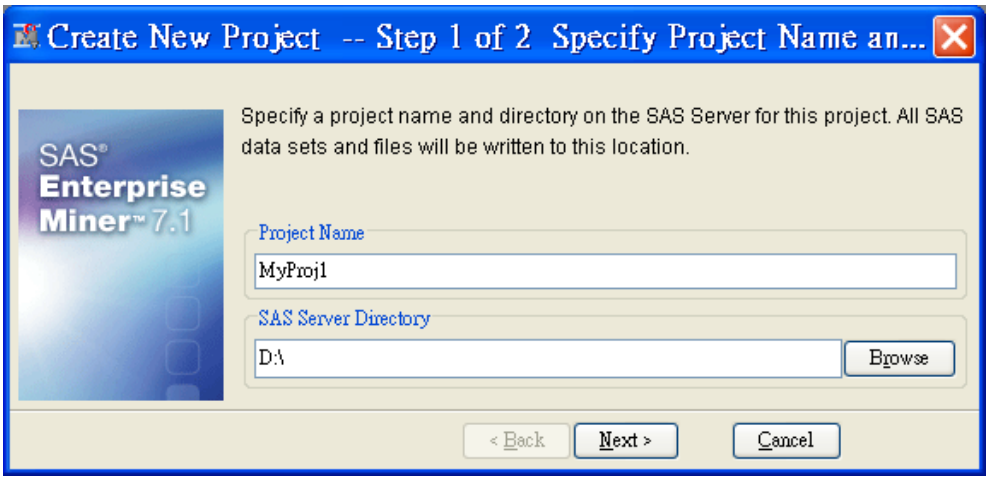
Step 2. Define Library (建 Library)

Step 3. New Data Source (Input Data Set)(指定 db source)

Step 4. Create Diagram (Diagram 流程圖)

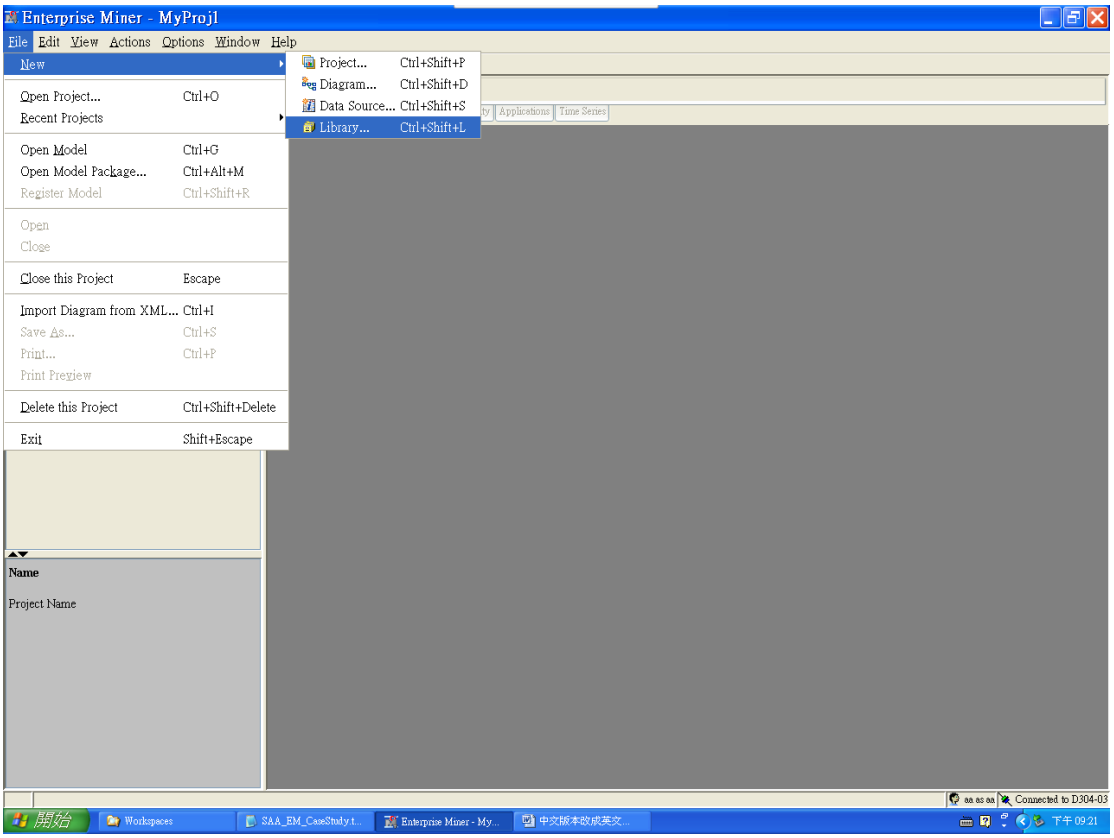
1. 建 project

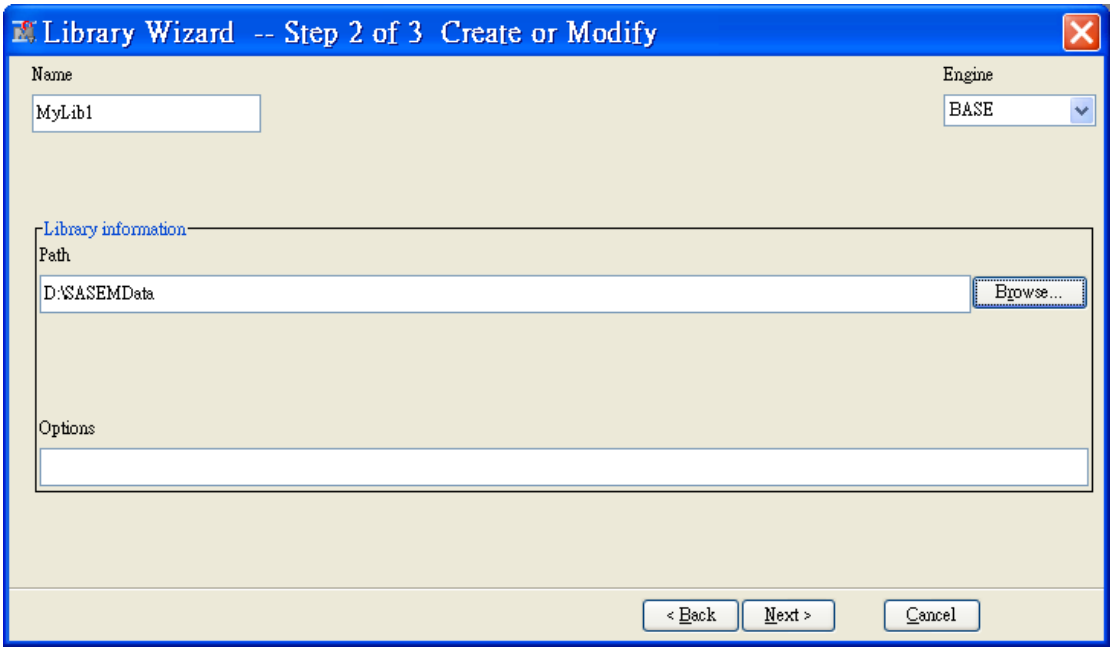




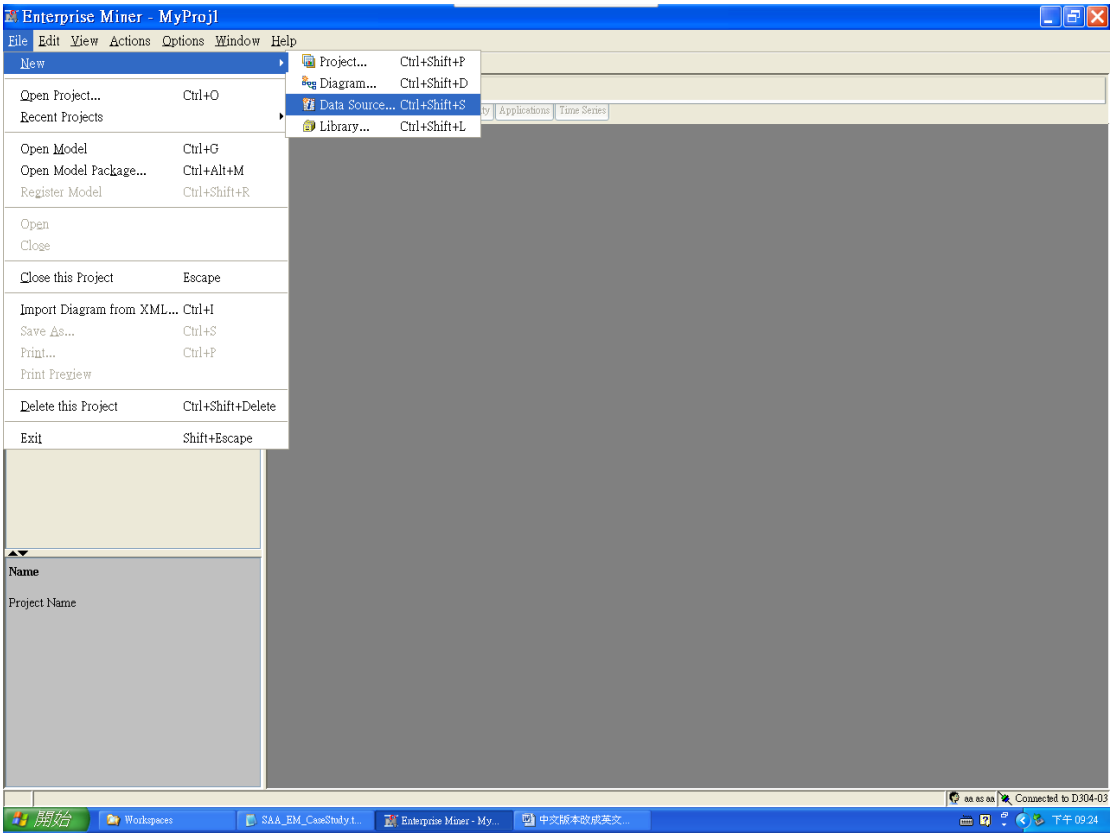
2. 建 Library（db）以 table 放在資料夾的(.sas7bdat)

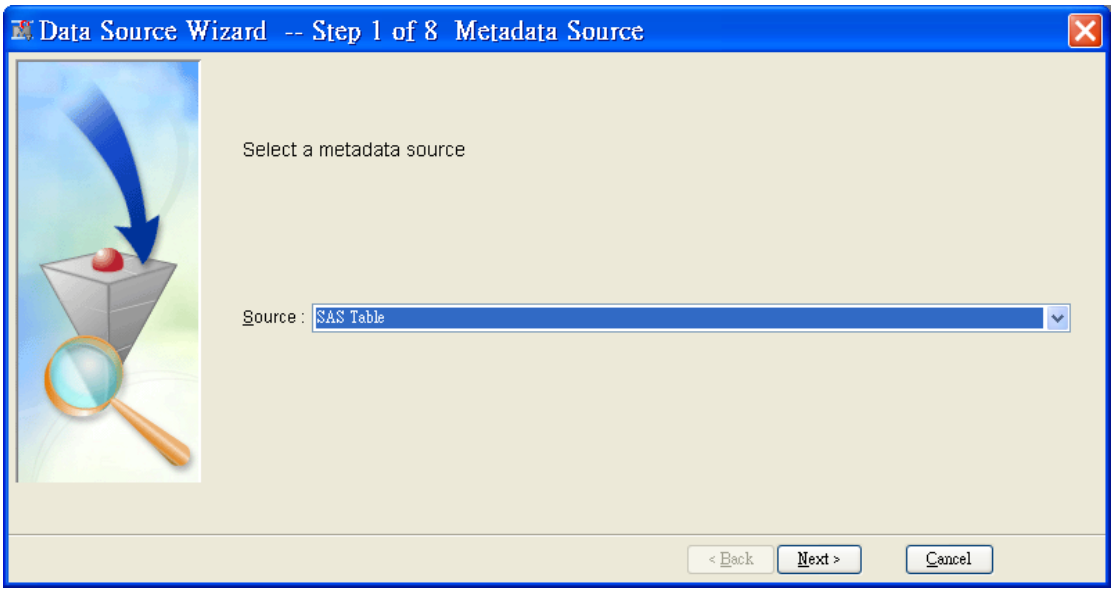
請指定資料夾，非資料夾內的檔案（D:\SASEMData）

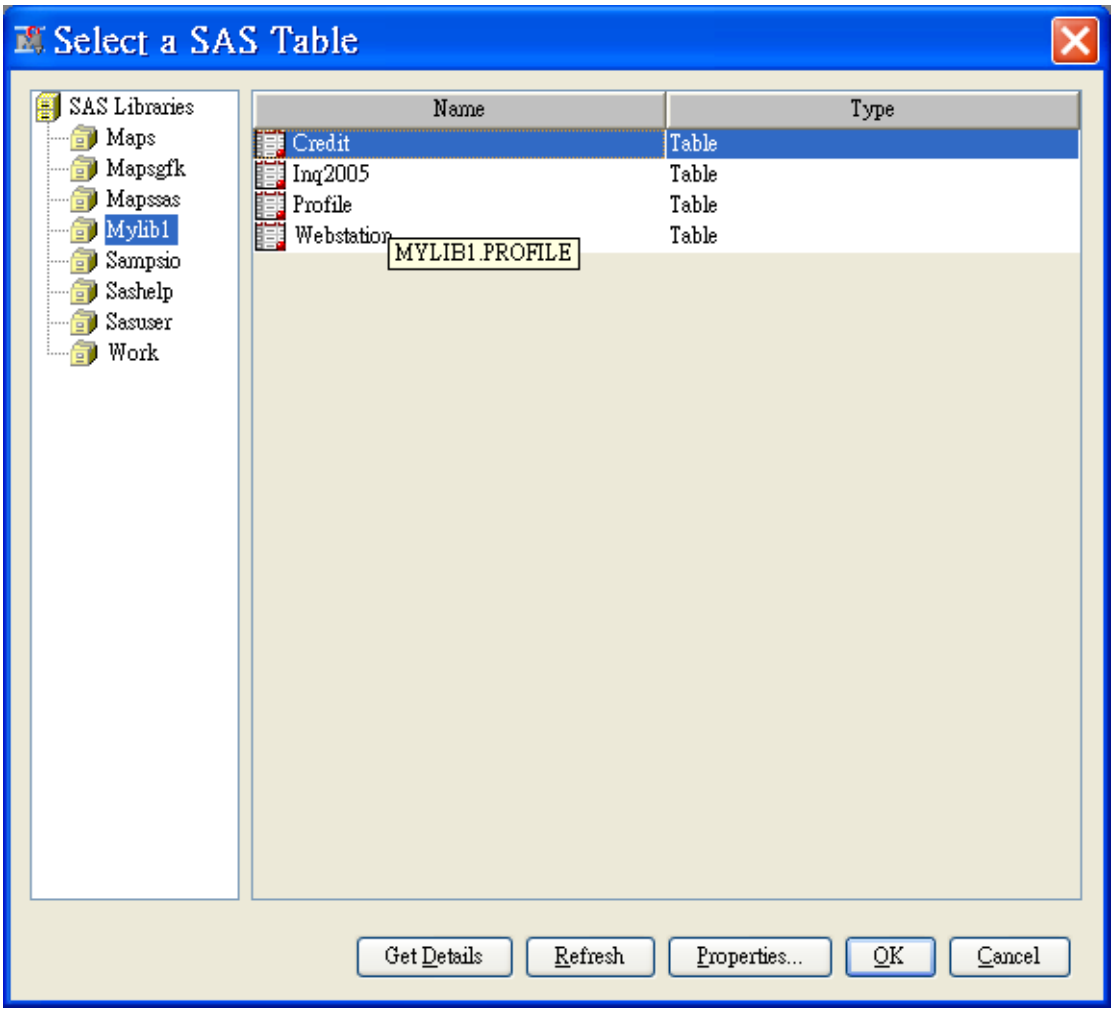


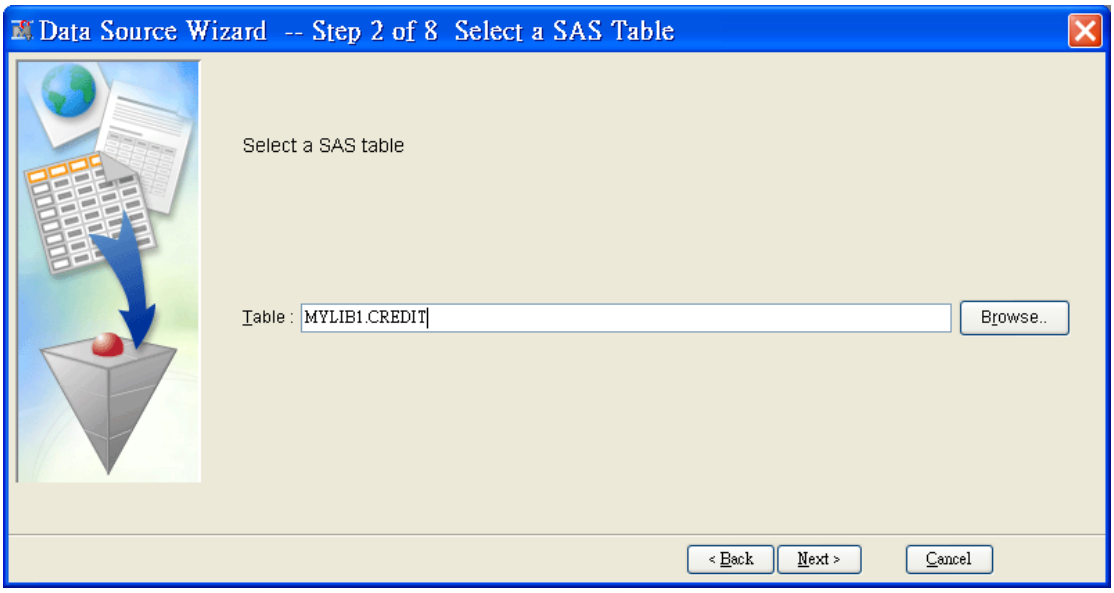


3. 指定 db source





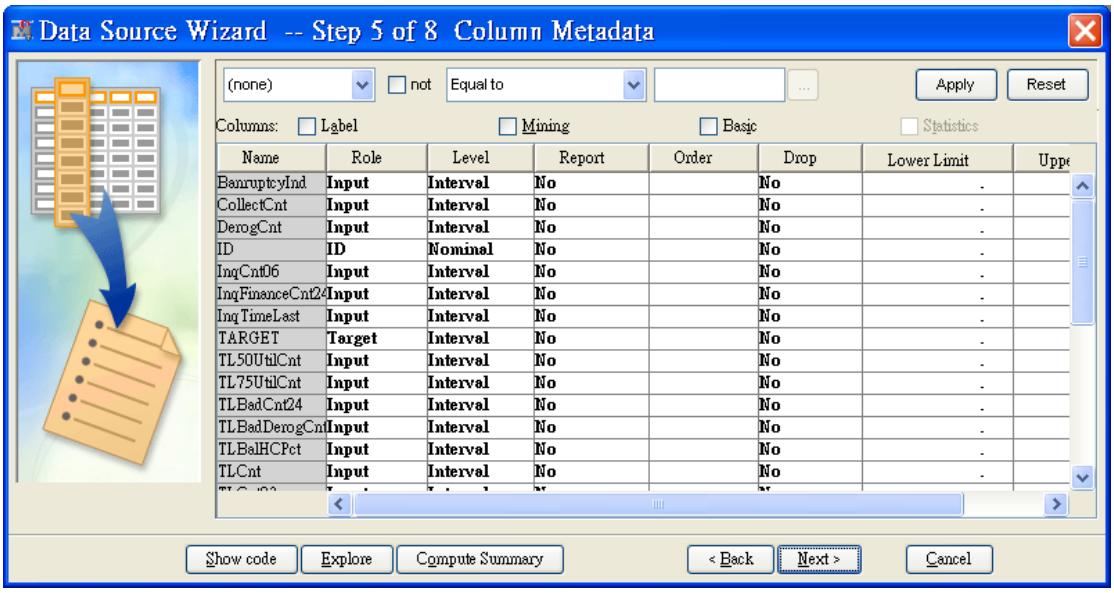




欄位名稱及角色

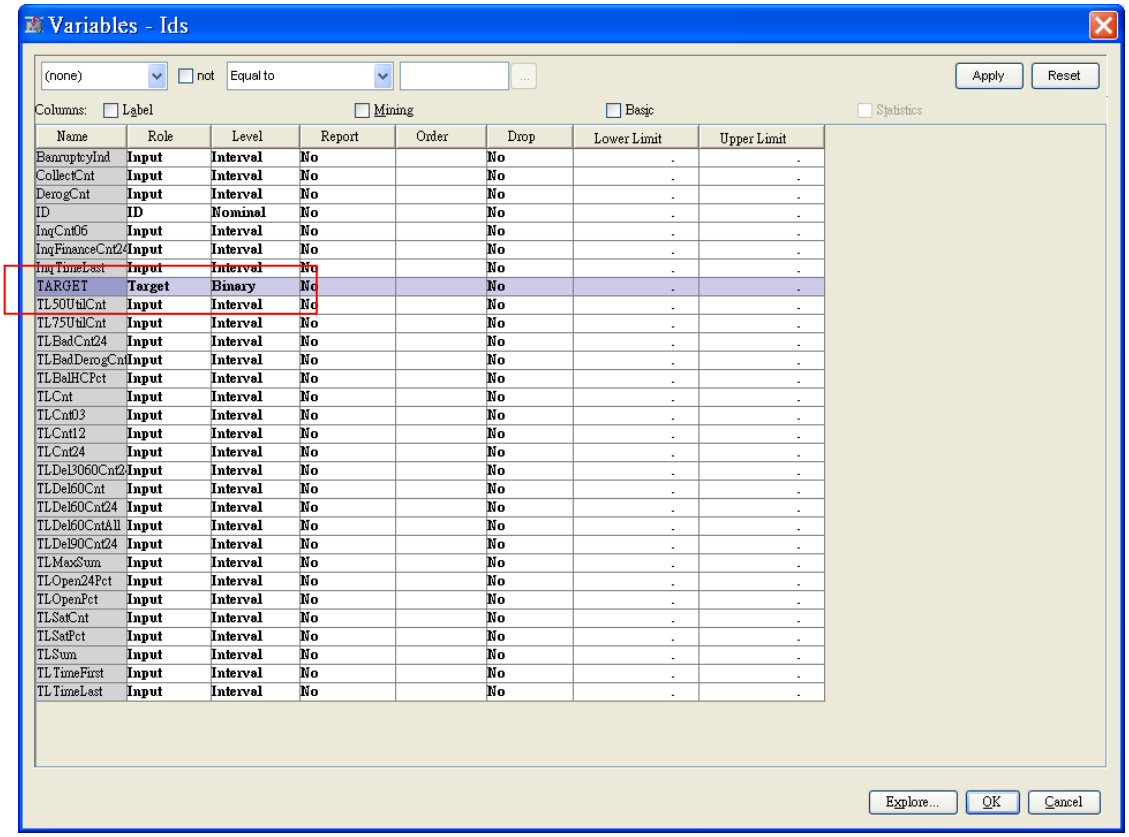
1.一定要有 Target 當 Y 變數

2.input 為 X 變數

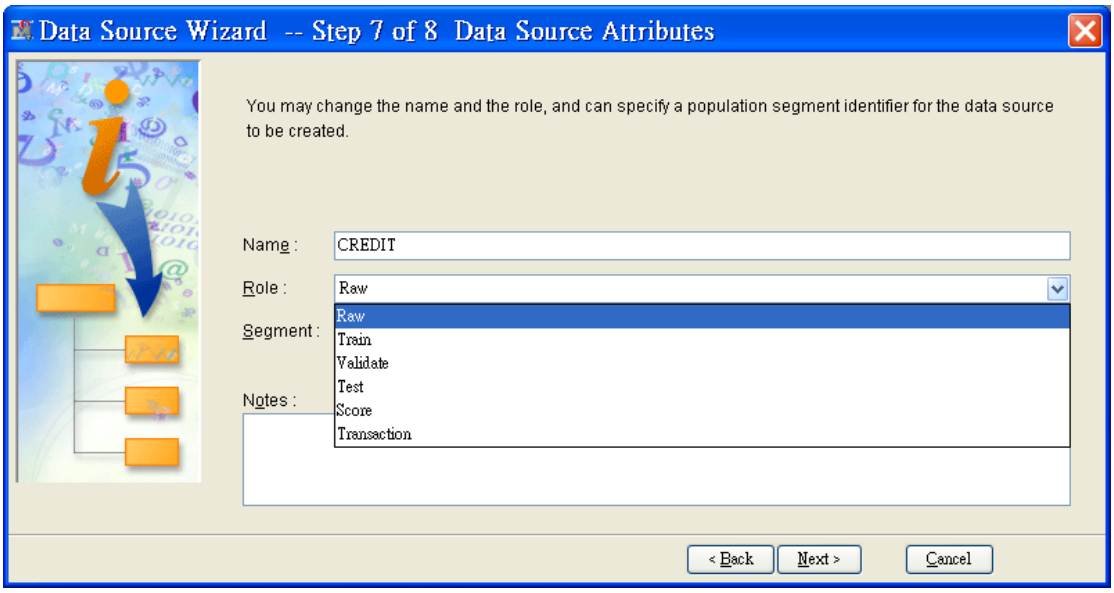


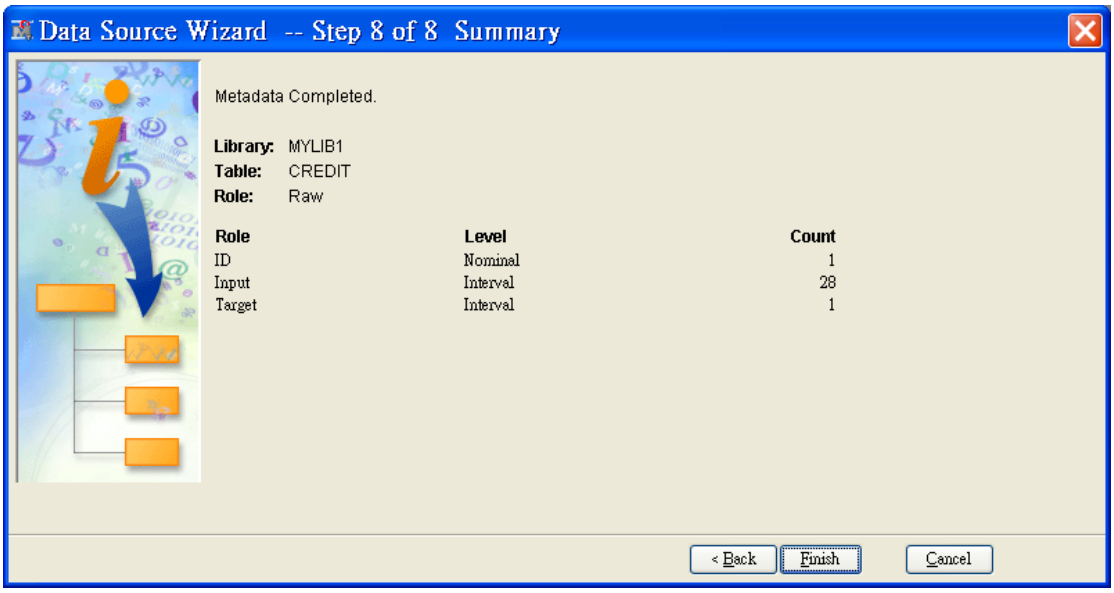
Interval：在 decision tree 會產生平均值，如：購買幾次

若改為 Binary 為 0/1 or yes/no（如：1=好 0=壞客戶、yes=會 no=不會買）

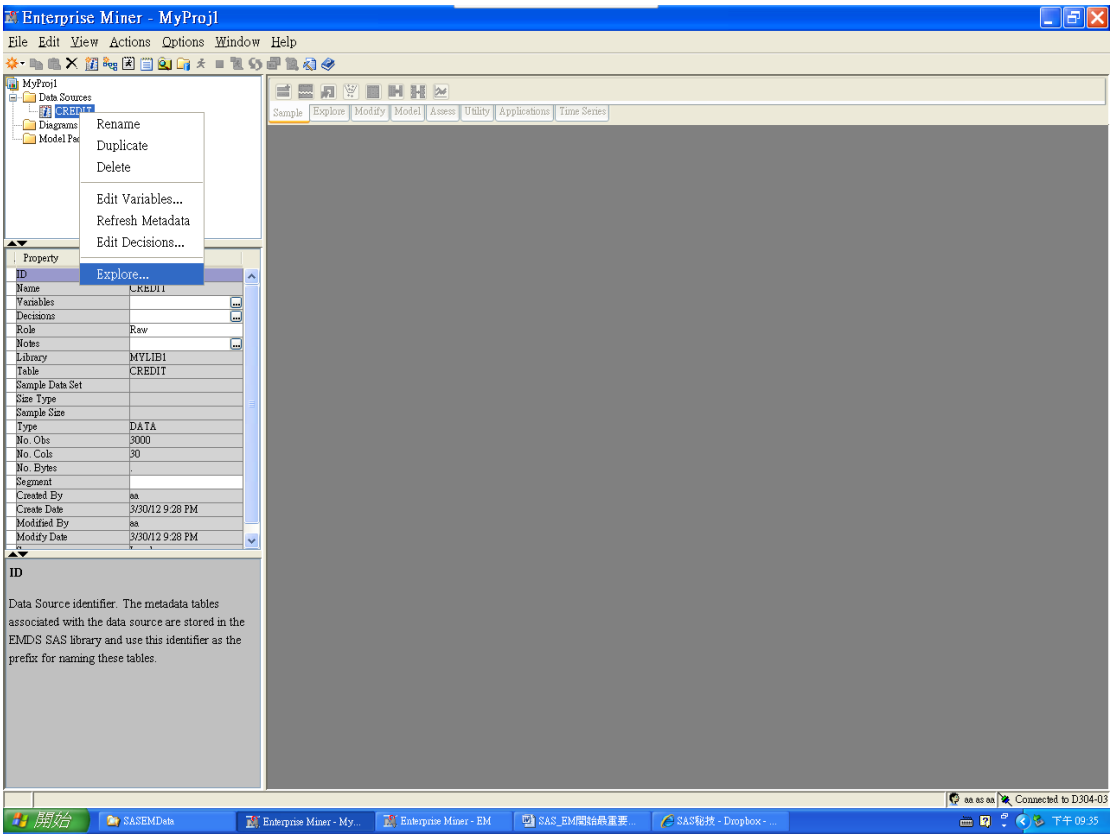


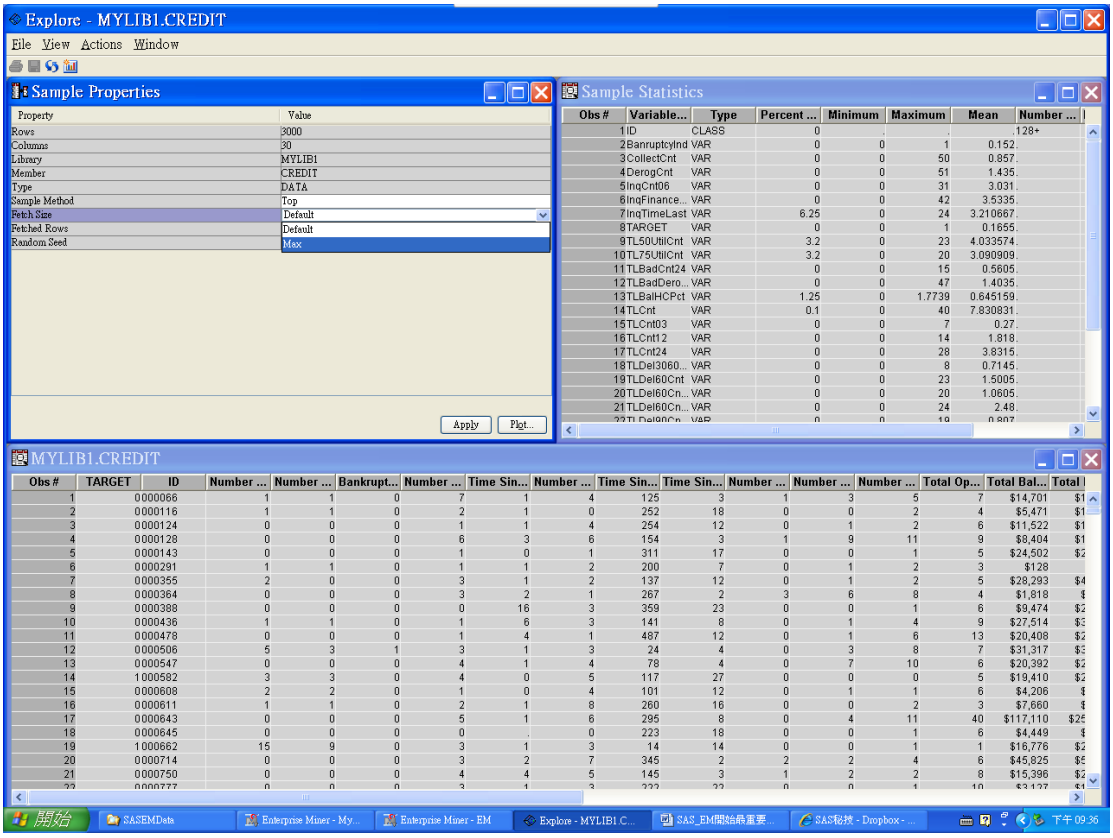
選 Raw data

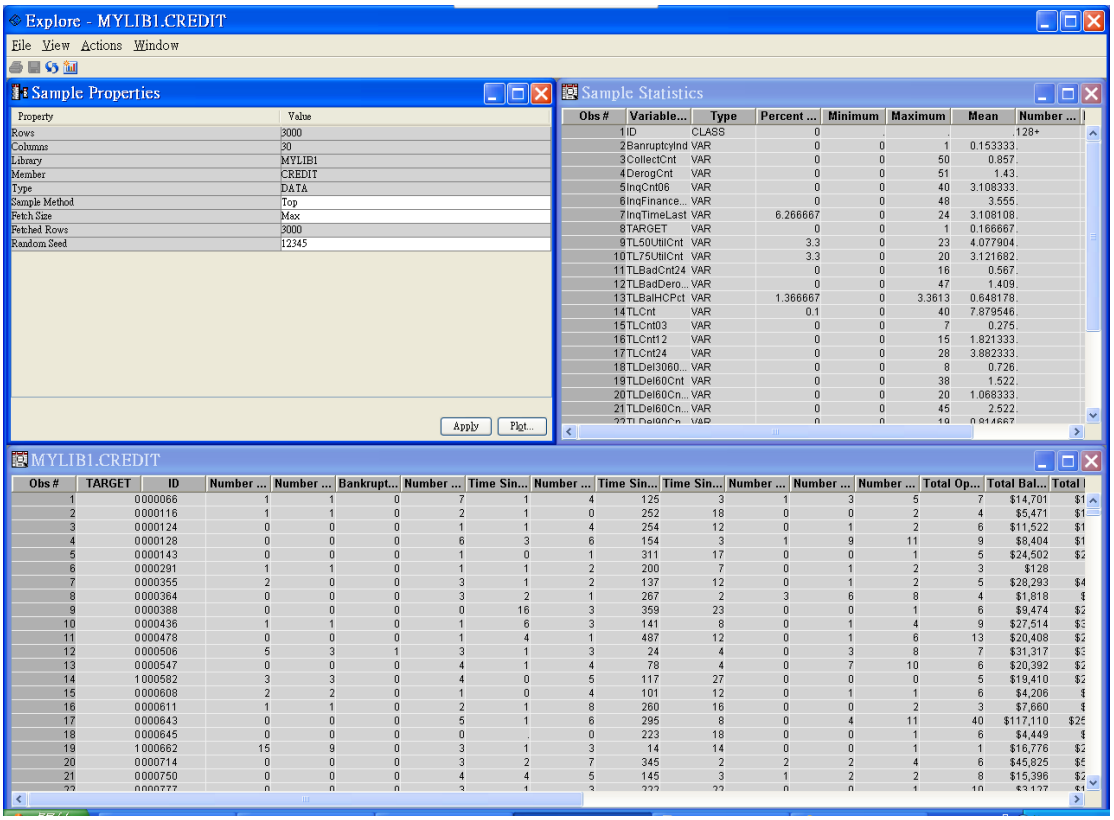




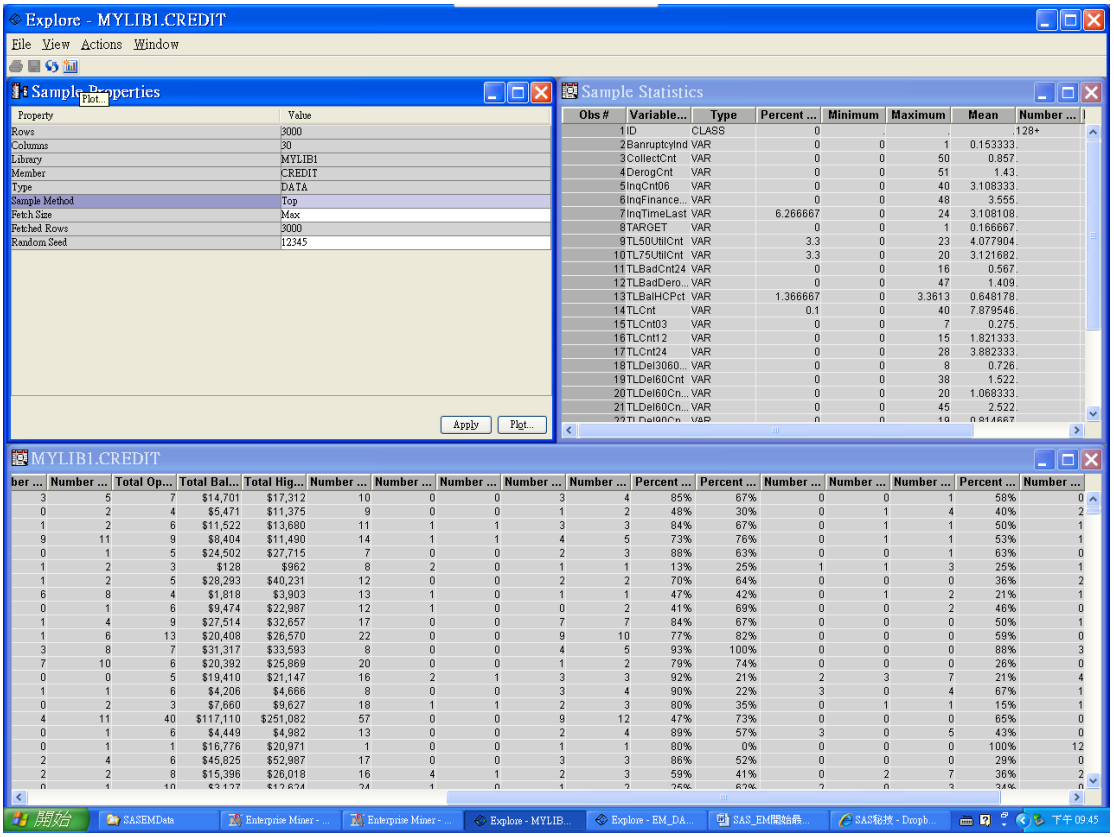
觀察資料：改最大值(Max)，講義後面章節 P.72 有資料庫欄位說明（預設為 2000 筆，改完為 3000 筆）

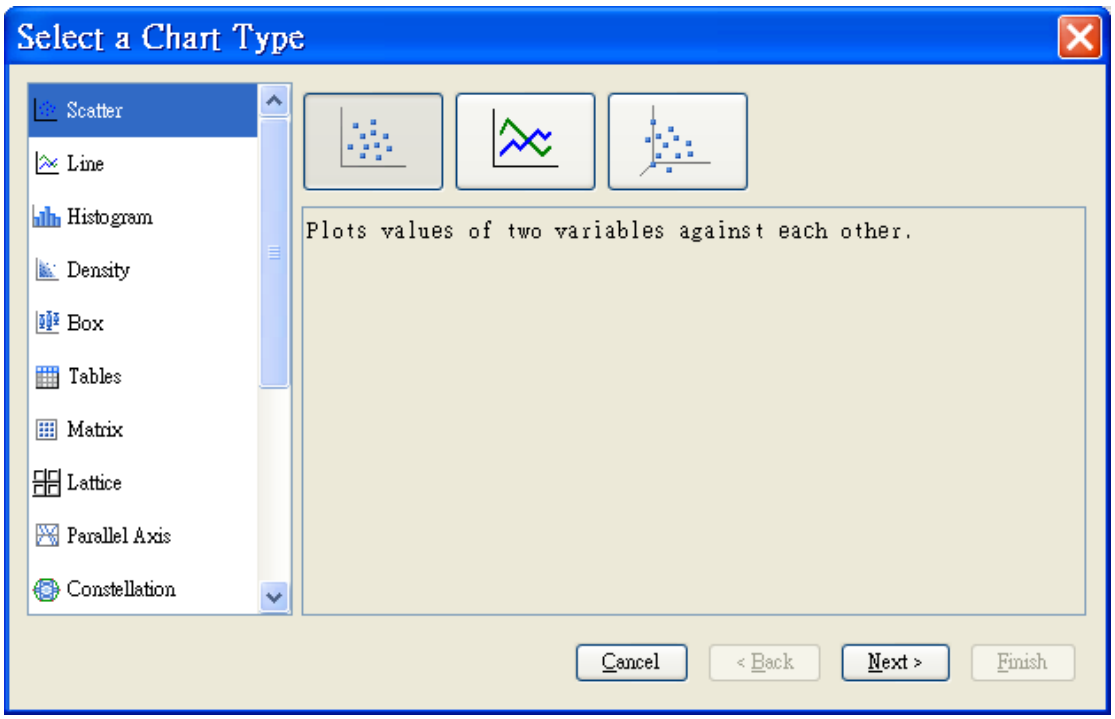

Enterprise Miner - MyProj1
File Edit View Actions Options Window Help
MyProj1
Data Sources
CREDIT
Diagrams
Model Pa
Rename
Duplicate
Delete
Edit Variables...
Refresh Metadata
Edit Decisions...
Explore...
Sample Explore Modify Model Assess Utility Applications Time Series
Property
ID CREDIT
Name CREDIT
Variables
Decisions
Role Raw
Notes
Library MYLIB1
Table CREDIT
Sample Data Set
Size Type
Sample Size
Type DATA
No. Obs 3000
No. Cols 30
No. Bytes
Segment
Created By aa
Create Date 3/30/12 9:28 PM
Modified By aa
Modify Date 3/30/12 9:28 PM
ID
Data Source identifier. The metadata tables associated with the data source are stored in the EMDS SAS library and use this identifier as the prefix for naming these tables.
Connected to D304-03
開始
SASEMData
Enterprise Miner - My...
Enterprise Miner - EM
SAS_EM開始最重要...
SAS秘技 - Dropbox - ...
下午 09:35

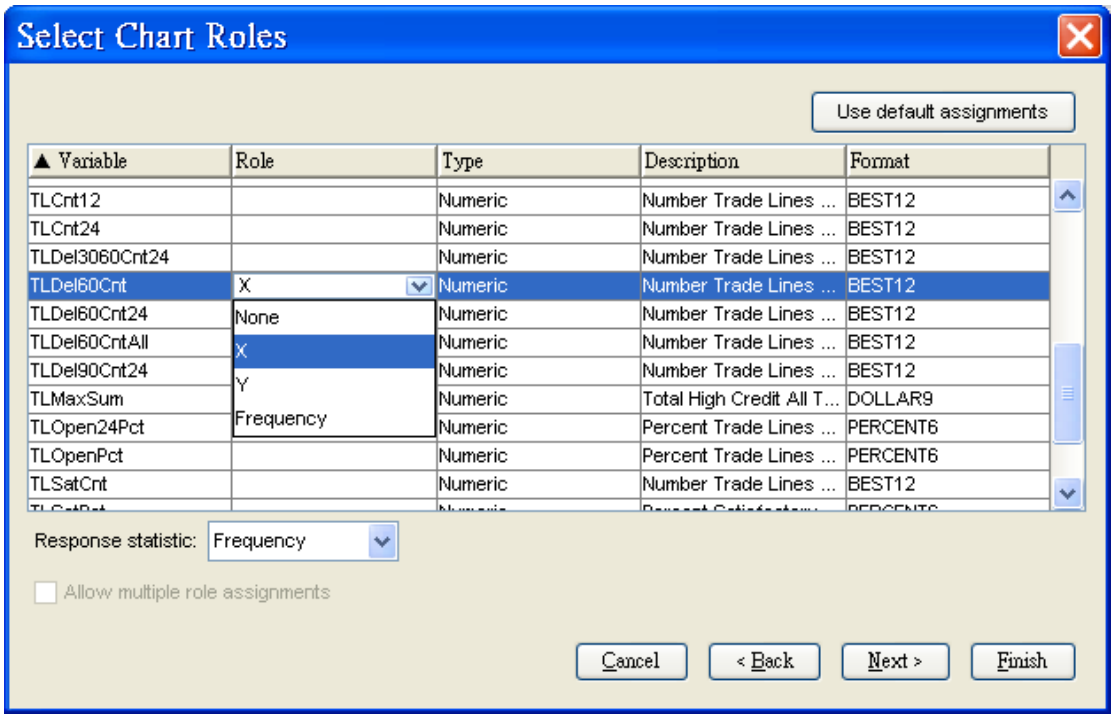

Explore - MYLIB1.CREDIT
File View Actions Window
Sample Properties
Property Value
Rows 3000
Columns 30
Library MYLIB1
Member CREDIT
Type DATA
Sample Method Top
Fetch Size Default
Fetched Rows Default Max
Random Seed
Apply Plot...
Sample Statistics
Obs # Variable... Type Percent ... Minimum Maximum Mean Number ...
1 ID CLASS 0 . . .128+
2 BanruptcyInd VAR 0 0 1 0.152.
3 CollectCnt VAR 0 0 50 0.857.
4 DerogCnt VAR 0 0 51 1.435.
5 InqCnt06 VAR 0 0 31 3.031.
6 InqFinance... VAR 0 0 42 3.5335.
7 InqTimeLast VAR 6.25 0 24 3.210667.
8 TARGET VAR 0 0 1 0.1655.
9 TL50UtilCnt VAR 3.2 0 23 4.033574.
10 TL75UtilCnt VAR 3.2 0 20 3.090909.
11 TLBadCnt24 VAR 0 0 15 0.5605.
12 TLBadDero... VAR 0 0 47 1.4035.
13 TLBalHCPct VAR 1.25 0 1.7739 0.645159.
14 TLCnt VAR 0.1 0 40 7.830831.
15 TLCnt03 VAR 0 0 7 0.27.
16 TLCnt12 VAR 0 0 14 1.818.
17 TLCnt24 VAR 0 0 28 3.8315.
18 TLDel3060... VAR 0 0 8 0.7145.
19 TLDel60Cnt VAR 0 0 23 1.5005.
20 TLDel60Cn... VAR 0 0 20 1.0605.
21 TLDel60Cn... VAR 0 0 24 2.48.
MYLIB1.CREDIT
Obs # TARGET ID Number ... Number ... Bankrupt... Number ... Time Sin... Number ... Time Sin... Time Sin... Number ... Number ... Number ... Total Op... Total Bal... Total
1 0000066 1 1 0 7 1 4 125 3 1 3 5 7 $14,701 $1
2 0000116 1 1 0 2 1 0 252 18 0 0 2 4 $5,471 $1
3 0000124 0 0 0 1 1 4 254 12 0 1 2 6 $11,522 $1
4 0000128 0 0 0 6 3 6 154 3 1 9 11 9 $8,404 $1
5 0000143 0 0 0 1 0 1 311 17 0 0 1 5 $24,502 $2
6 0000291 1 1 0 1 1 2 200 7 0 1 2 3 $128
7 0000355 2 0 0 3 1 2 137 12 0 1 2 5 $28,293 $4
8 0000364 0 0 0 3 2 1 267 2 3 6 8 4 $1,818 $
9 0000388 0 0 0 0 16 3 359 23 0 0 1 6 $9,474 $2
10 0000436 1 1 0 1 6 3 141 8 0 1 4 9 $27,514 $3
11 0000478 0 0 0 1 4 1 487 12 0 1 6 13 $20,408 $2
12 0000506 5 3 1 3 1 3 24 4 0 3 8 7 $31,317 $3
13 0000547 0 0 0 4 1 4 78 4 0 7 10 6 $20,392 $2
14 1000582 3 3 0 4 0 5 117 27 0 0 0 5 $19,410 $2
15 0000608 2 2 0 1 0 4 101 12 0 1 1 6 $4,206 $
16 0000611 1 1 0 2 1 8 260 16 0 0 2 3 $7,660 $
17 0000643 0 0 0 5 1 6 295 8 0 4 11 40 $117,110 $25
18 0000645 0 0 0 0 . 0 223 18 0 0 1 6 $4,449 $
19 1000662 15 9 0 3 1 3 14 14 0 0 1 1 $16,776 $2
20 0000714 0 0 0 3 2 7 345 2 2 2 4 6 $45,825 $5
21 0000750 0 0 0 4 4 5 145 3 1 2 2 8 $15,396 $2
開始
SASEMData
Enterprise Miner - My...
Enterprise Miner - EM
Explore - MYLIB1.C...
SAS_EM開始最重要...
SAS秘技 - Dropbox - ...
下午 09:36

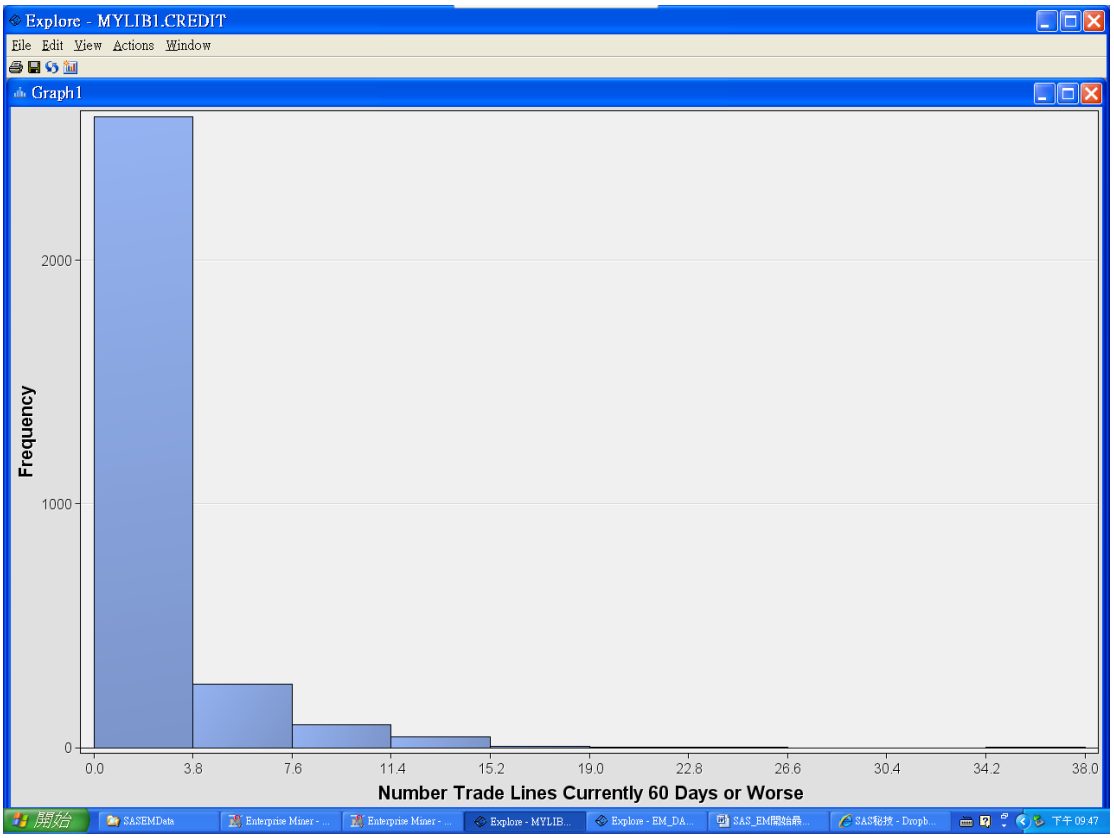

Explore - MYLIB1.CREDIT
File View Actions Window
Sample Properties
Property Value
Rows 3000
Columns 30
Library MYLIB1
Member CREDIT
Type DATA
Sample Method Top
Fetch Size Max
Fetched Rows 3000
Random Seed 12345
Apply Plot...
Sample Statistics
Obs # Variable... Type Percent ... Minimum Maximum Mean Number ...
1 ID CLASS 0 . . 128+
2 BanruptcyInd VAR 0 0 1 0.153333
3 CollectCnt VAR 0 0 50 0.857
4 DerogCnt VAR 0 0 51 1.43
5 InqCnt06 VAR 0 0 40 3.108333
6 InqFinance... VAR 0 0 48 3.555
7 InqTimeLast VAR 6.266667 0 24 3.108108
8 TARGET VAR 0 0 1 0.166667
9 TL50UtilCnt VAR 3.3 0 23 4.077904
10 TL75UtilCnt VAR 3.3 0 20 3.121682
11 TLBadCnt24 VAR 0 0 16 0.567
12 TLBadDero... VAR 0 0 47 1.409
13 TLBalHCPct VAR 1.366667 0 3.3613 0.648178
14 TLCnt VAR 0.1 0 40 7.879546
15 TLCnt03 VAR 0 0 7 0.275
16 TLCnt12 VAR 0 0 15 1.821333
17 TLCnt24 VAR 0 0 28 3.882333
18 TLDel3060... VAR 0 0 8 0.726
19 TLDel60Cnt VAR 0 0 38 1.522
20 TLDel60Cn... VAR 0 0 20 1.068333
21 TLDel60Cn... VAR 0 0 45 2.522
MYLIB1.CREDIT
Obs # TARGET ID Number ... Number ... Bankrupt... Number ... Time Sin... Number ... Time Sin... Time Sin... Number ... Number ... Number ... Total Op... Total Bal... Total
1 0000066 1 1 0 7 1 4 125 3 1 3 5 7 $14,701
2 0000116 1 1 0 2 1 0 252 18 0 0 2 4 $5,471
3 0000124 0 0 0 1 1 4 254 12 0 1 2 6 $11,522
4 0000128 0 0 0 6 3 6 154 3 1 9 11 9 $8,404
5 0000143 0 0 0 1 0 1 311 17 0 0 1 5 $24,502
6 0000291 1 1 0 1 1 2 200 7 0 1 2 3 $128
7 0000355 2 0 0 3 1 2 137 12 0 1 2 5 $28,293
8 0000364 0 0 0 3 2 1 267 2 3 6 8 4 $1,818
9 0000388 0 0 0 0 16 3 359 23 0 0 1 6 $9,474
10 0000436 1 1 0 1 6 3 141 8 0 1 4 9 $27,514
11 0000478 0 0 0 1 4 1 487 12 0 1 6 13 $20,408
12 0000506 5 3 1 3 1 3 24 4 0 3 8 7 $31,317
13 0000547 0 0 0 4 1 4 78 4 0 7 10 6 $20,392
14 1000582 3 3 0 4 0 5 117 27 0 0 0 5 $19,410
15 0000608 2 2 0 1 0 4 101 12 0 1 1 6 $4,206
16 0000611 1 1 0 2 1 8 260 16 0 0 2 3 $7,660
17 0000643 0 0 0 5 1 6 295 8 0 4 11 40 $117,110
18 0000645 0 0 0 0 . 0 223 18 0 0 1 6 $4,449
19 1000662 15 9 0 3 1 3 14 14 0 0 1 1 $16,776
20 0000714 0 0 0 3 2 7 345 2 2 2 4 6 $45,825
21 0000750 0 0 0 4 4 5 145 3 1 2 2 8 $15,396

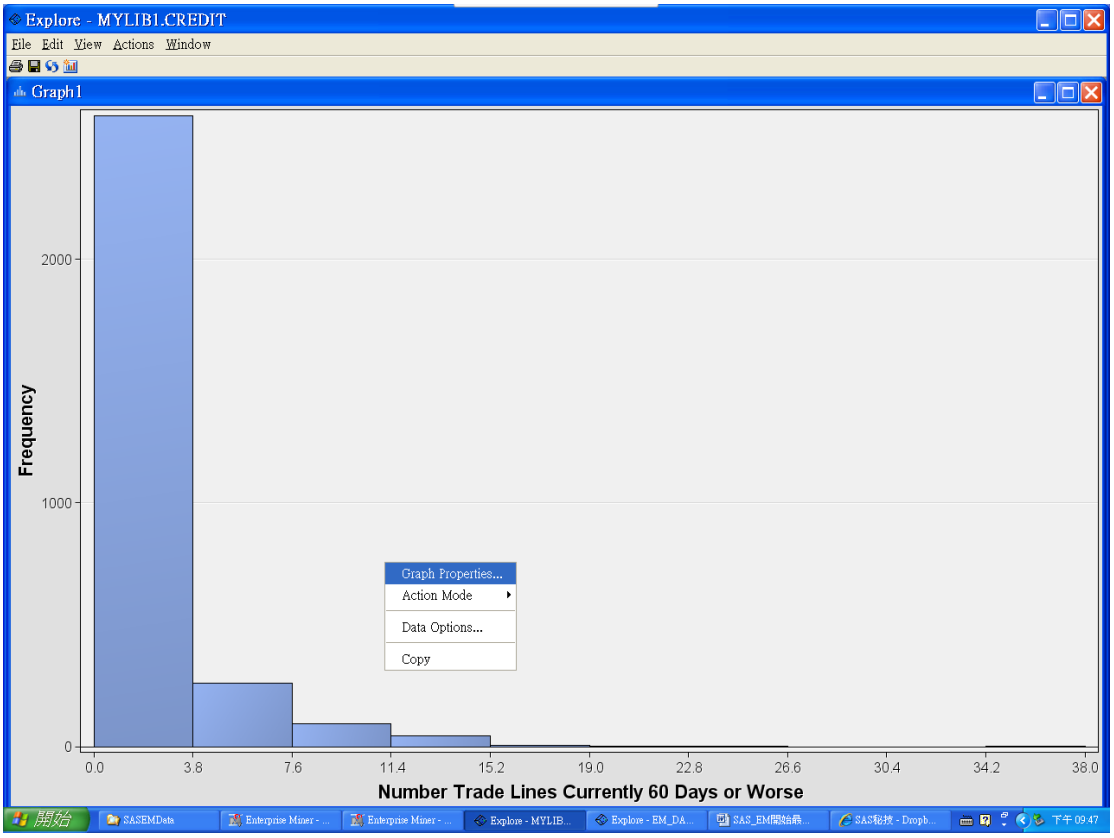
Plot


Explore - MYLIB1.CREDIT
File View Actions Window
Sample Properties
Plot
Property Value
Rows 3000
Columns 30
Library MYLIB1
Member CREDIT
Type DATA
Sample Method Top
Fetch Size Max
Fetched Rows 3000
Random Seed 12345
Apply Plot...
Sample Statistics
Obs # Variable... Type Percent ... Minimum Maximum Mean Number ...
1 ID CLASS 0 . . 128+
2 BanruptcyInd VAR 0 0 1 0.153333
3 CollectCnt VAR 0 0 50 0.857
4 DerogCnt VAR 0 0 51 1.43
5 InqCnt06 VAR 0 0 40 3.108333
6 InqFinance... VAR 0 0 48 3.555
7 InqTimeLast VAR 6.266667 0 24 3.108108
8 TARGET VAR 0 0 1 0.166667
9 TL50UtilCnt VAR 3.3 0 23 4.077904
10 TL75UtilCnt VAR 3.3 0 20 3.121682
11 TLBadCnt24 VAR 0 0 16 0.567
12 TLBadDero... VAR 0 0 47 1.409
13 TLBalHCPct VAR 1.366667 0 3.3613 0.648178
14 TLCnt VAR 0.1 0 40 7.879546
15 TLCnt03 VAR 0 0 7 0.275
16 TLCnt12 VAR 0 0 15 1.821333
17 TLCnt24 VAR 0 0 28 3.882333
18 TLDel3060... VAR 0 0 8 0.726
19 TLDel60Cnt VAR 0 0 38 1.522
20 TLDel60Cn... VAR 0 0 20 1.068333
21 TLDel60Cn... VAR 0 0 45 2.522
MYLIB1.CREDIT
ber ... Number ... Total Op... Total Bal... Total Hig... Number ... Number ... Number ... Number ... Number ... Percent ... Percent ... Number ... Number ... Number ... Percent ... Number ...
3 5 7 $14,701 $17,312 10 0 0 3 4 85% 67% 0 0 1 58% 0
0 2 4 $5,471 $11,375 9 0 0 1 2 48% 30% 0 1 4 40% 2
1 2 6 $11,522 $13,680 11 1 1 3 3 84% 67% 0 1 1 50% 1
9 11 9 $8,404 $11,490 14 1 1 4 5 73% 76% 0 1 1 53% 1
0 1 5 $24,502 $27,715 7 0 0 2 3 88% 63% 0 0 1 63% 0
1 2 3 $128 $962 8 2 0 1 1 13% 25% 1 1 3 25% 1
1 2 5 $28,293 $40,231 12 0 0 2 2 70% 64% 0 0 0 36% 2
6 8 4 $1,818 $3,903 13 1 0 1 1 47% 42% 0 1 2 21% 1
0 1 6 $9,474 $22,987 12 1 0 0 2 41% 69% 0 0 2 46% 0
1 4 9 $27,514 $32,657 17 0 0 7 7 84% 67% 0 0 0 50% 1
1 6 13 $20,408 $26,570 22 0 0 9 10 77% 82% 0 0 0 59% 0
3 8 7 $31,317 $33,593 8 0 0 4 5 93% 100% 0 0 0 88% 3
7 10 6 $20,392 $25,869 20 0 0 1 2 79% 74% 0 0 0 26% 0
0 0 5 $19,410 $21,147 16 2 1 3 3 92% 21% 2 3 7 21% 4
1 1 6 $4,206 $4,666 8 0 0 3 4 90% 22% 3 0 4 67% 1
0 2 3 $7,660 $9,627 18 1 1 2 3 80% 35% 0 1 1 15% 1
4 11 40 $117,110 $251,082 57 0 0 9 12 47% 73% 0 0 0 65% 0
0 1 6 $4,449 $4,982 13 0 0 2 4 89% 57% 3 0 5 43% 0
0 1 1 $16,776 $20,971 1 0 0 1 1 80% 0% 0 0 0 100% 12
2 4 6 $45,825 $52,987 17 0 0 3 3 86% 52% 0 0 0 29% 0
2 2 8 $15,396 $26,018 16 4 1 2 3 59% 41% 0 2 7 36% 2

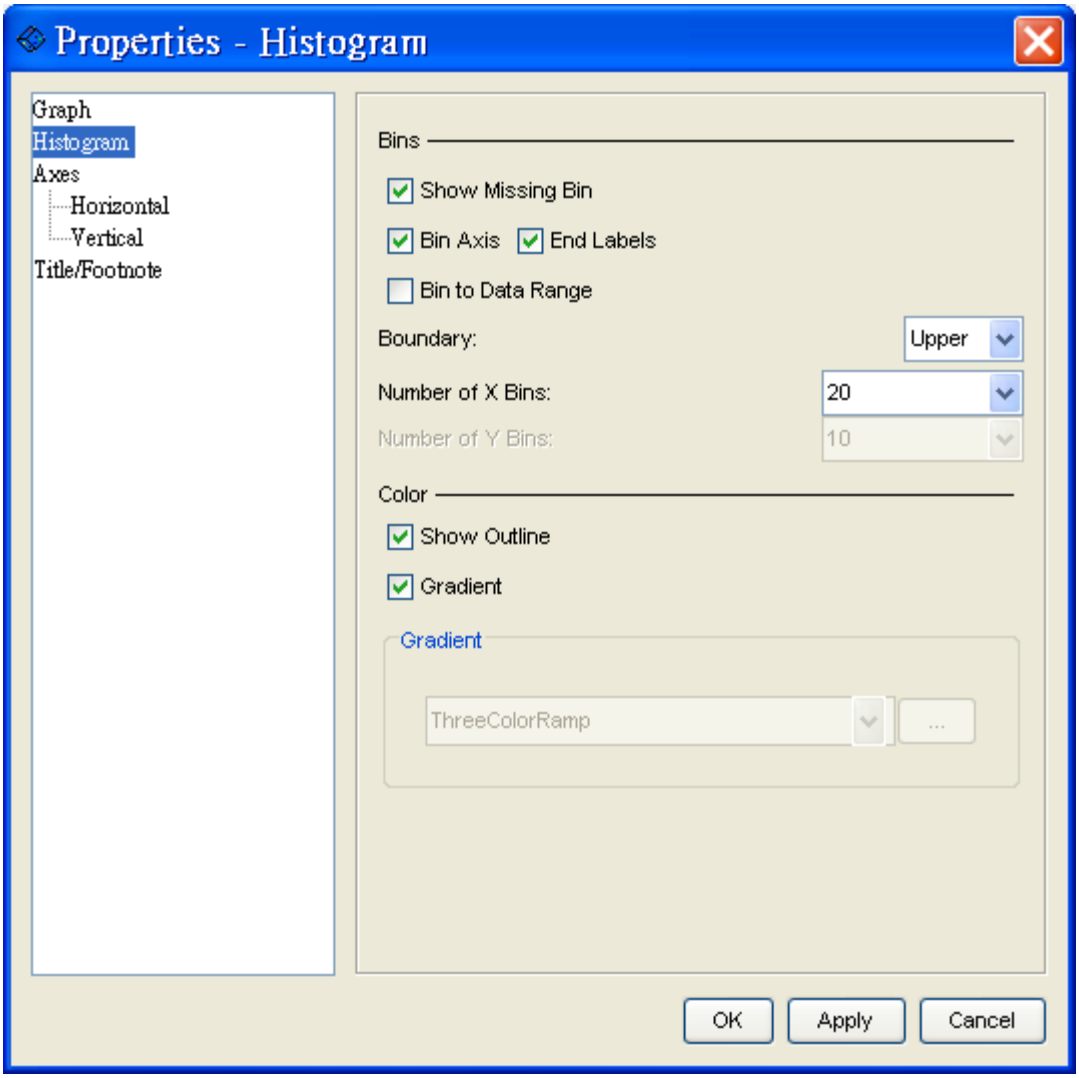

Select a Chart Type
Scatter
Line
Histogram
Density
Box
Tables
Matrix
Lattice
Parallel Axis
Constellation
Plots values of two variables against each other.
Cancel
< Back
Next >
Finish

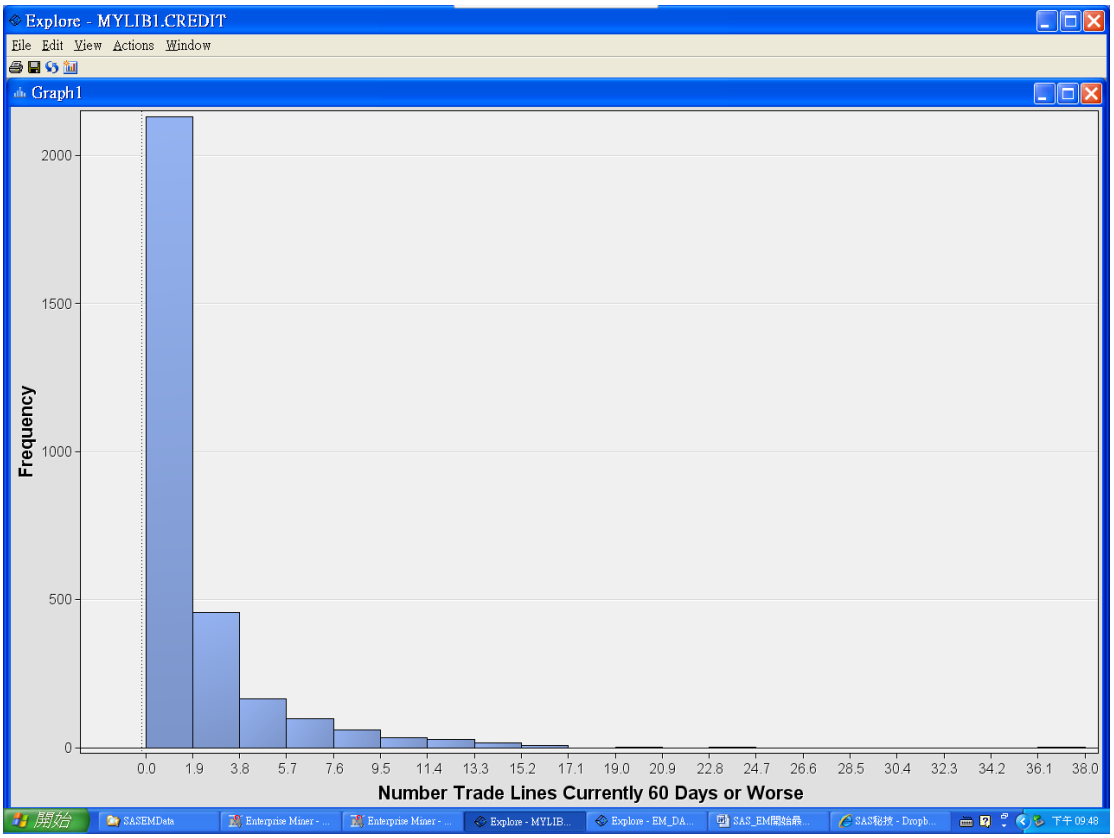

Select Chart Roles
Use default assignments
▲ Variable
Role
Type
Description
Format
TLCnt12
Numeric
Number Trade Lines ...
BEST12
TLCnt24
Numeric
Number Trade Lines ...
BEST12
TLDel3060Cnt24
Numeric
Number Trade Lines ...
BEST12
TLDel60Cnt
X
Numeric
Number Trade Lines ...
BEST12
TLDel60Cnt24
Numeric
Number Trade Lines ...
BEST12
TLDel60CntAll
Numeric
Number Trade Lines ...
BEST12
TLDel90Cnt24
Numeric
Number Trade Lines ...
BEST12
TLMaxSum
Numeric
Total High Credit All T...
DOLLAR9
TLOpen24Pct
Numeric
Percent Trade Lines ...
PERCENT6
TLOpenPct
Numeric
Percent Trade Lines ...
PERCENT6
TLSatCnt
Numeric
Number Trade Lines ...
BEST12
None
X
Y
Frequency
Response statistic:
Frequency
Allow multiple role assignments
Cancel
< Back
Next >
Finish


Explore - MYLIB1.CREDIT
File Edit View Actions Window
Graph1
Frequency
2000
1000
0
0.0
3.8
7.6
11.4
15.2
19.0
22.8
26.6
30.4
34.2
38.0
Number Trade Lines Currently 60 Days or Worse

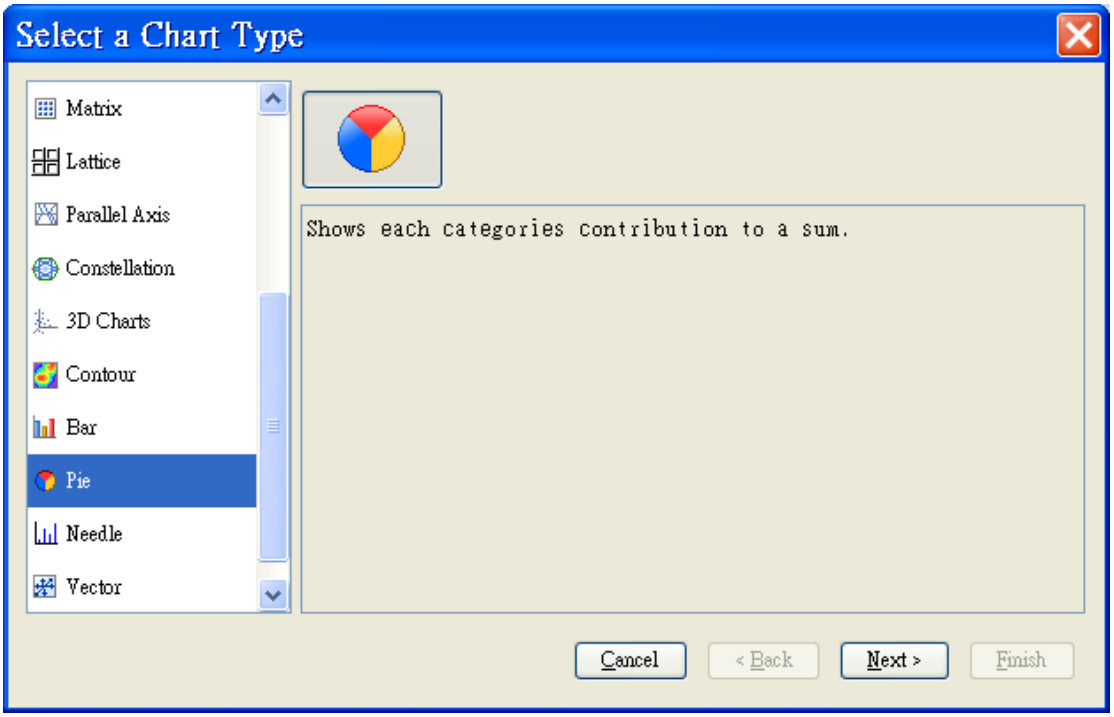

Explore - MYLIB1.CREDIT
File Edit View Actions Window
Graph1
Frequency
2000
1000
0
0.0
3.8
7.6
11.4
15.2
19.0
22.8
26.6
30.4
34.2
38.0
Number Trade Lines Currently 60 Days or Worse
Graph Properties...
Action Mode
Data Options...
Copy

Bin：區塊

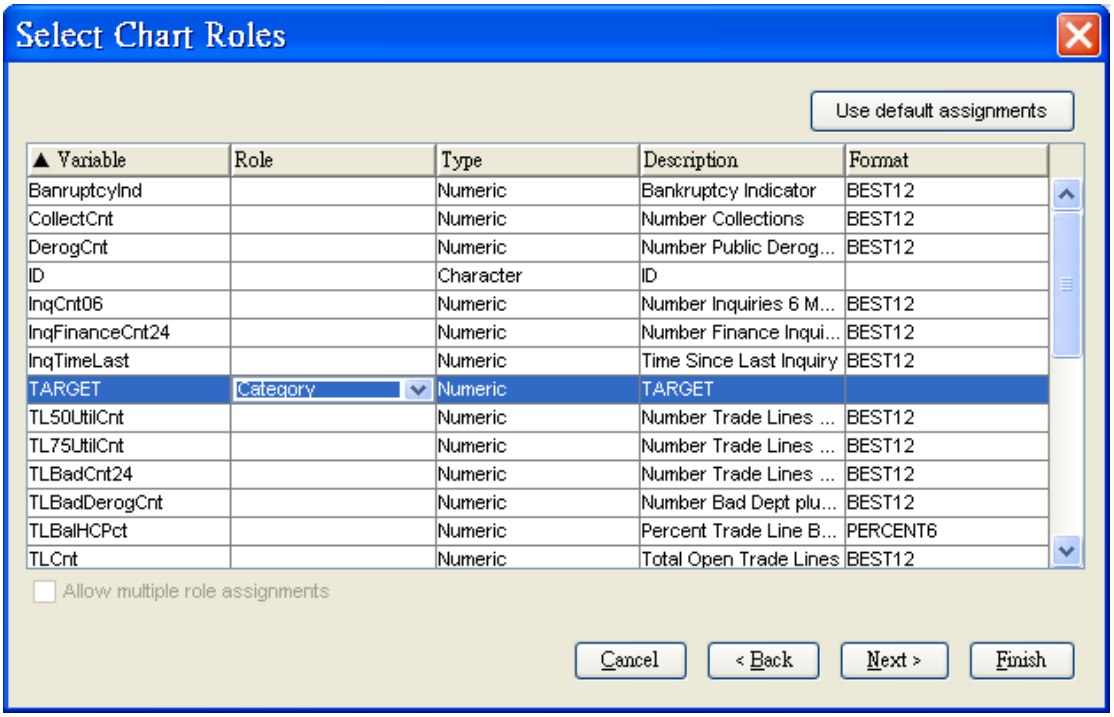

Properties - Histogram
Graph
Histogram
Axes
Horizontal
Vertical
Title/Footnote
Bins
Show Missing Bin
Bin Axis
End Labels
Bin to Data Range
Boundary:
Upper
Number of X Bins:
20
Number of Y Bins:
10
Color
Show Outline
Gradient
Gradient
ThreeColorRamp
...
OK
Apply
Cancel



Pie

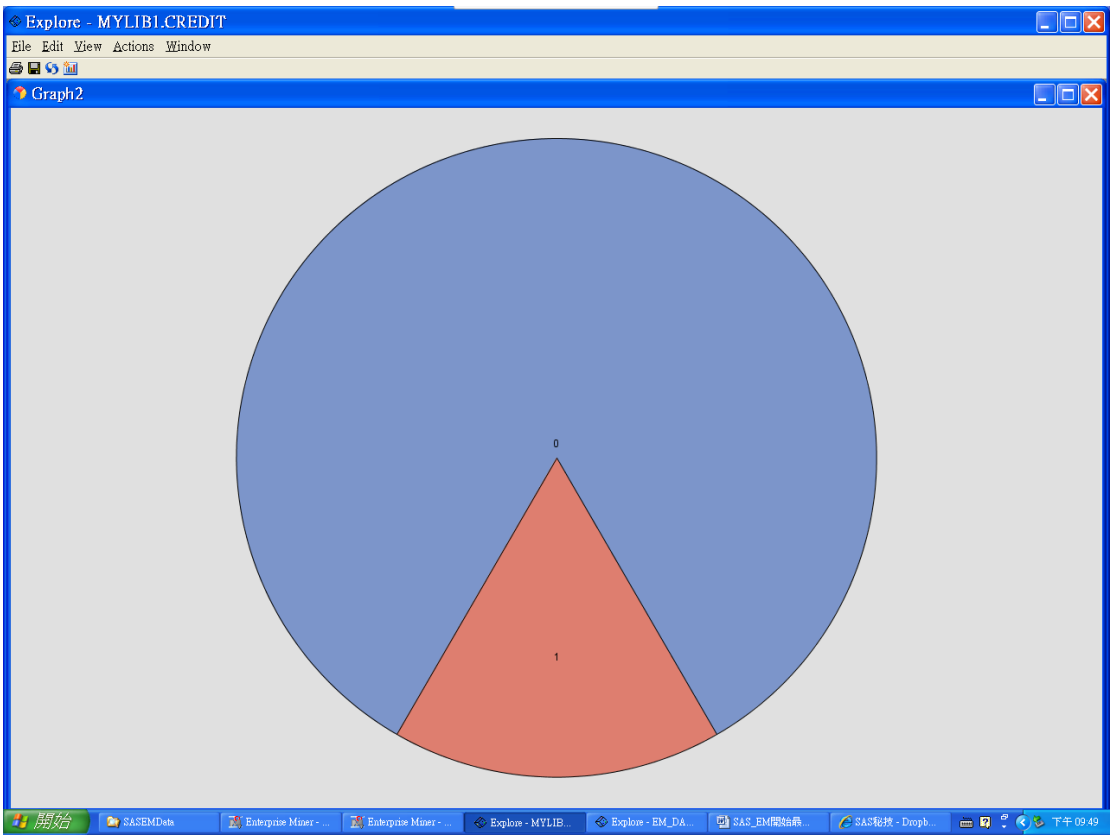


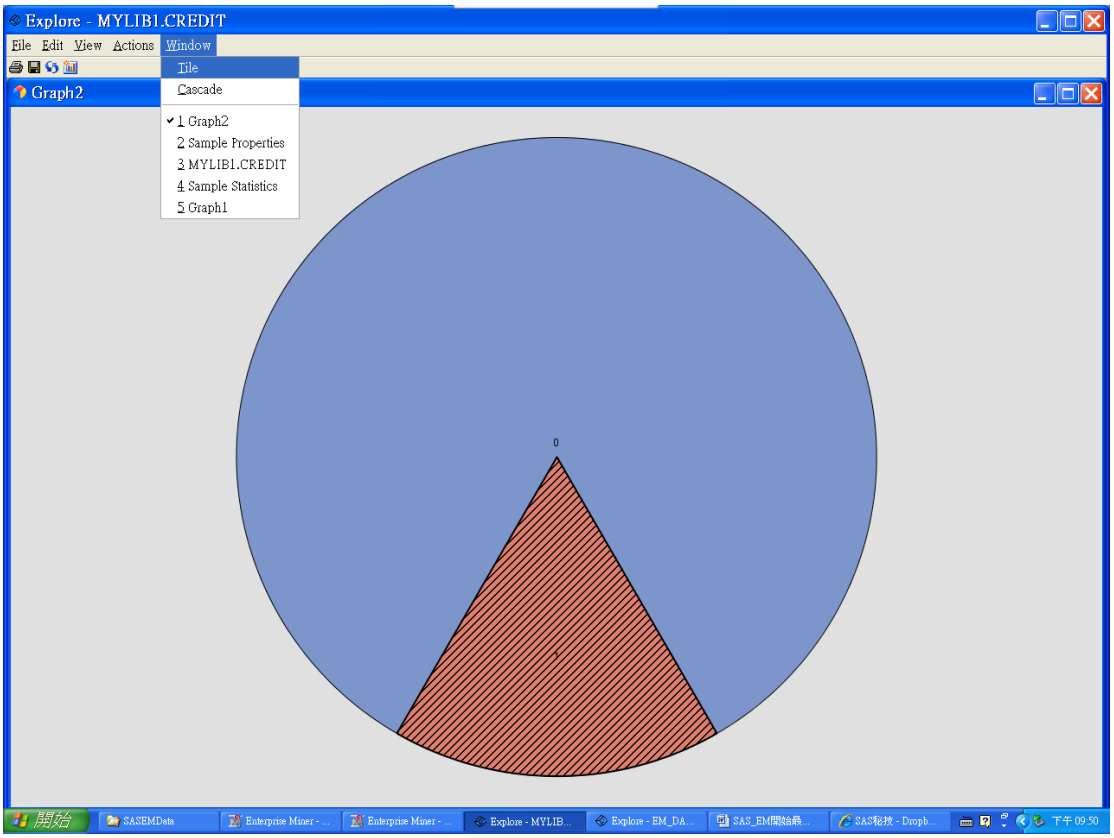
區分好壞客戶(0.1)

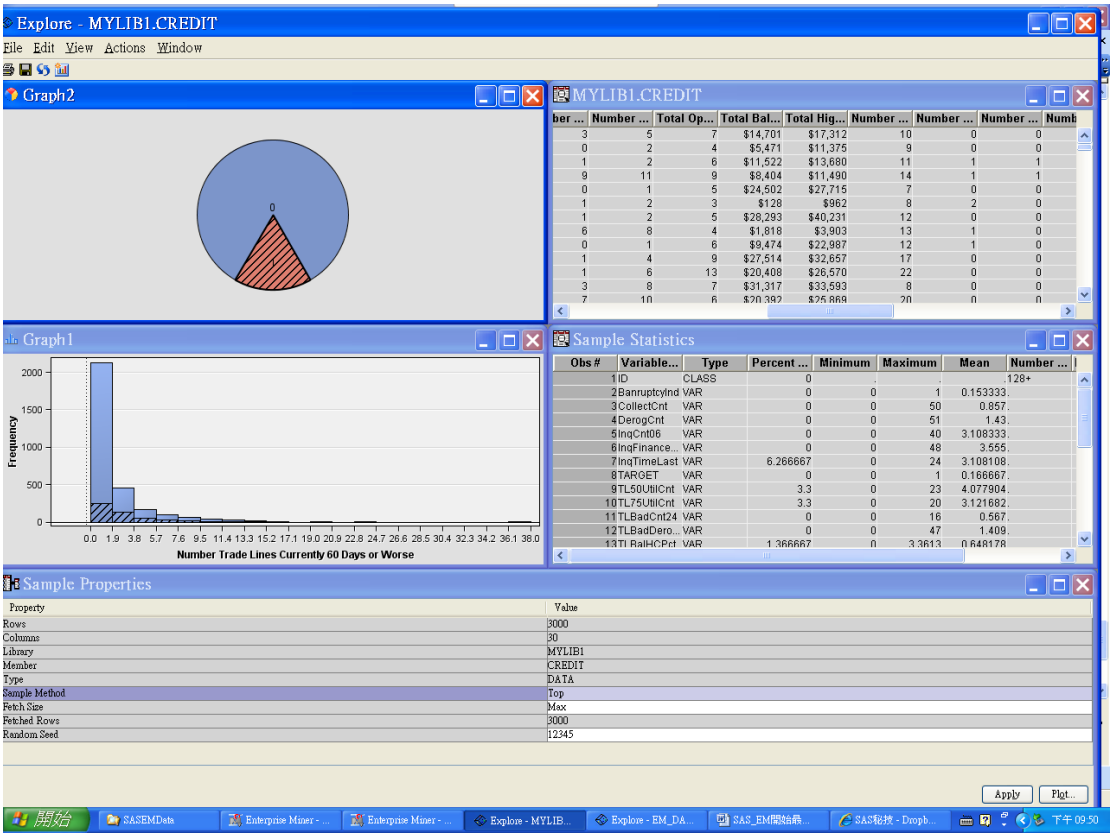


1:壞客戶:500

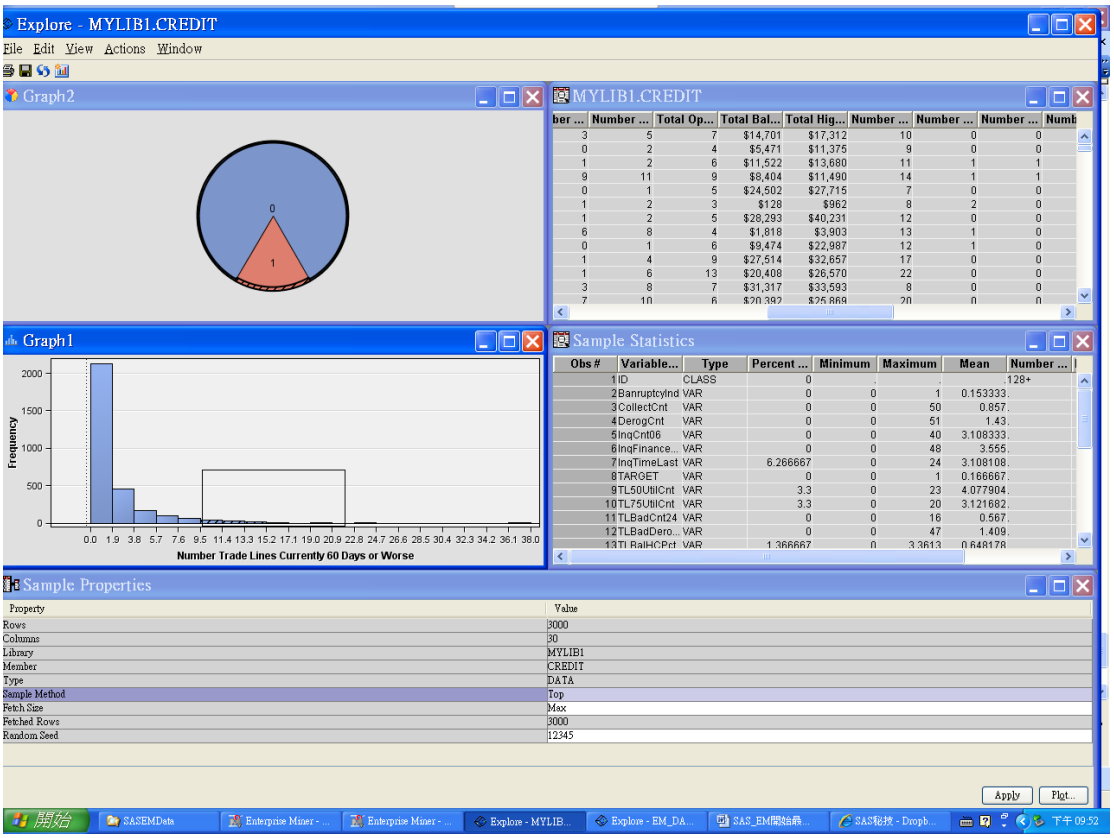
0:好客戶:2500



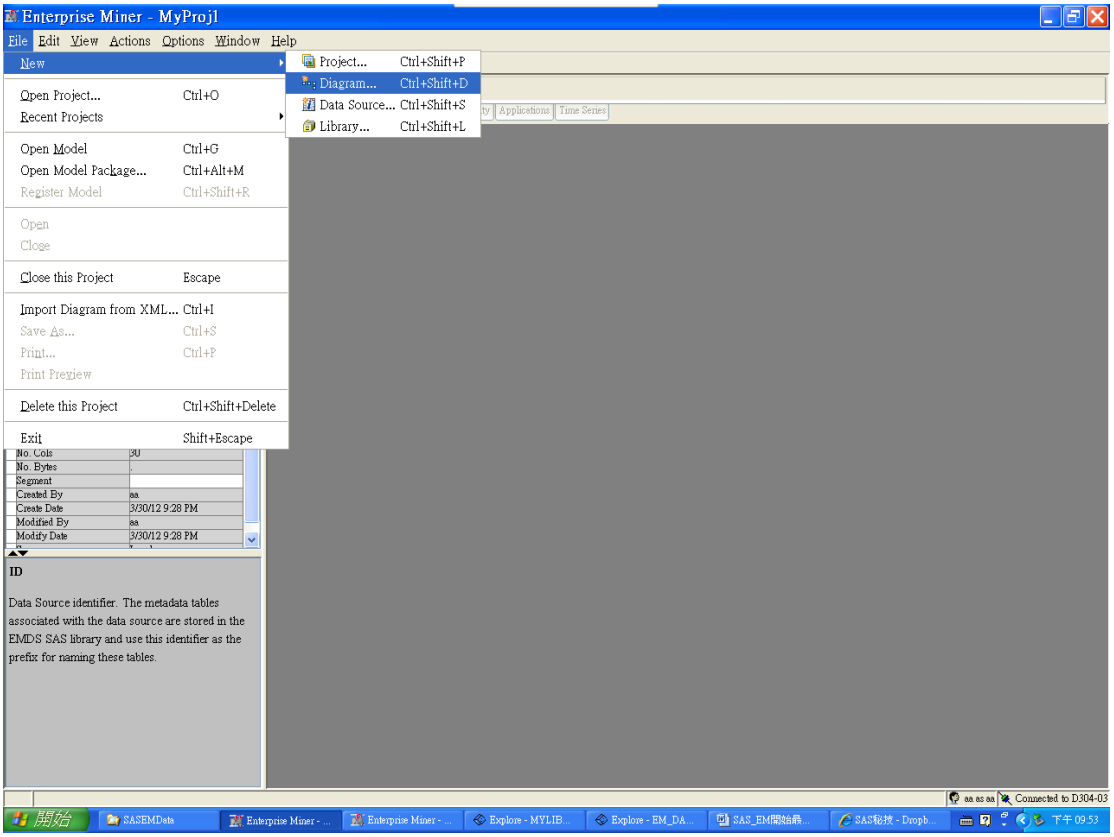


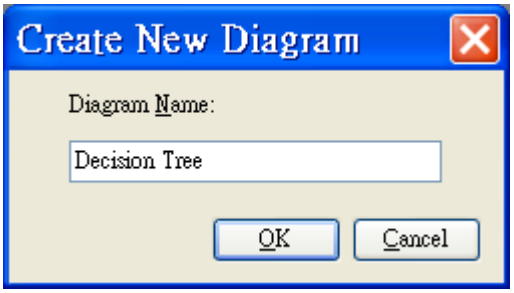


選擇部份區塊，可自動在 Pie 上看出分佈狀況

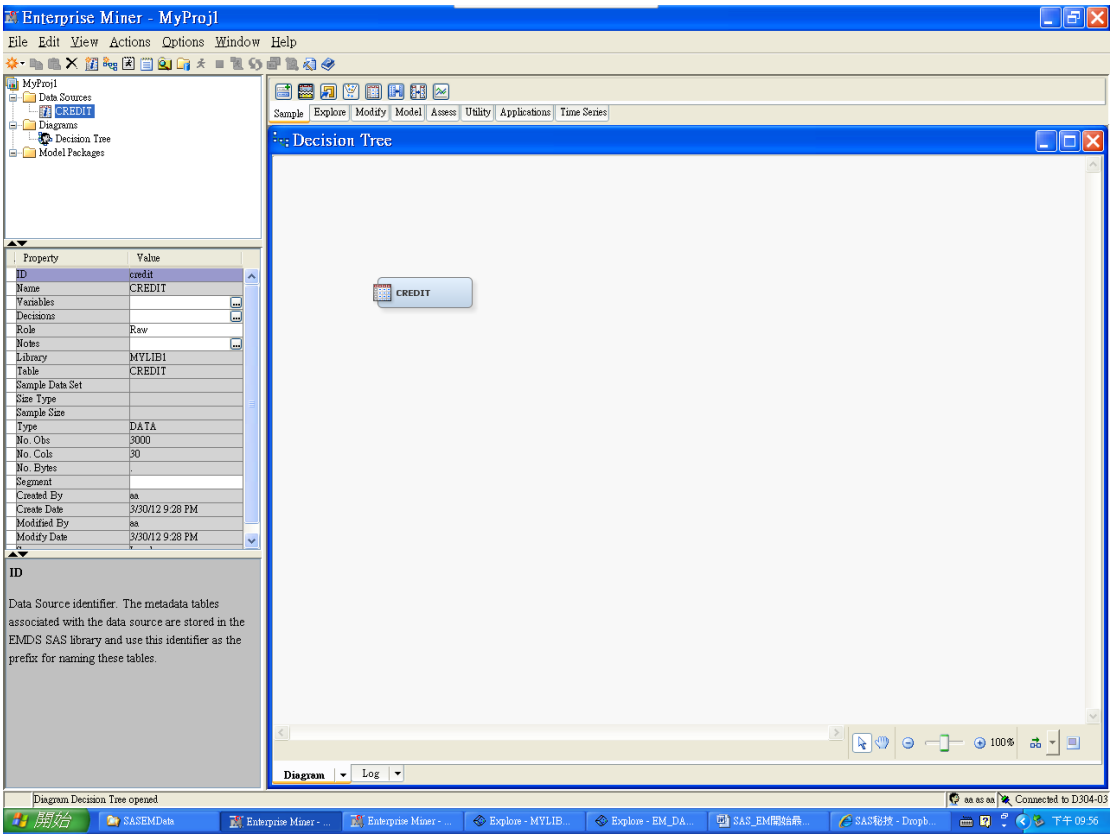

Explore - MYLIB1.CREDIT
File Edit View Actions Window
Graph2
0
1
MYLIB1.CREDIT
ber ... Number ... Total Op... Total Bal... Total Hig... Number ... Number ... Number ... Numb
3 5 7 $14,701 $17,312 10 0 0
0 2 4 $5,471 $11,375 9 0 0
1 2 6 $11,522 $13,680 11 1 1
9 11 9 $8,404 $11,490 14 1 1
0 1 5 $24,502 $27,715 7 0 0
1 2 3 $128 $962 8 2 0
1 2 5 $28,293 $40,231 12 0 0
6 8 4 $1,818 $3,903 13 1 0
0 1 6 $9,474 $22,987 12 1 0
1 4 9 $27,514 $32,657 17 0 0
1 6 13 $20,408 $26,570 22 0 0
3 8 7 $31,317 $33,593 8 0 0
Graph1
Frequency
2000
1500
1000
500
0
0.0 1.9 3.8 5.7 7.6 9.5 11.4 13.3 15.2 17.1 19.0 20.9 22.8 24.7 26.6 28.5 30.4 32.3 34.2 36.1 38.0
Number Trade Lines Currently 60 Days or Worse
Sample Statistics
Obs # Variable... Type Percent ... Minimum Maximum Mean Number ...
1 ID CLASS 0 . . 128+
2 BanruptcyInd VAR 0 0 1 0.153333
3 CollectCnt VAR 0 0 50 0.857
4 DerogCnt VAR 0 0 51 1.43
5 InqCnt06 VAR 0 0 40 3.108333
6 InqFinance... VAR 0 0 48 3.555
7 InqTimeLast VAR 6.266667 0 24 3.108108
8 TARGET VAR 0 0 1 0.166667
9 TL50UtilCnt VAR 3.3 0 23 4.077904
10 TL75UtilCnt VAR 3.3 0 20 3.121682
11 TLBadCnt24 VAR 0 0 16 0.567
12 TLBadDero... VAR 0 0 47 1.409
Sample Properties
Property Value
Rows 3000
Columns 30
Library MYLIB1
Member CREDIT
Type DATA
Sample Method Top
Fetch Size Max
Fetched Rows 3000
Random Seed 12345
Apply
Flot...
開始
SASEMData
Enterprise Miner - ...
Enterprise Miner - ...
Explore - MYLIB...
Explore - EM_DA...
SAS_EM開始篇...
SAS秘技 - Dropb...
下午 09:52

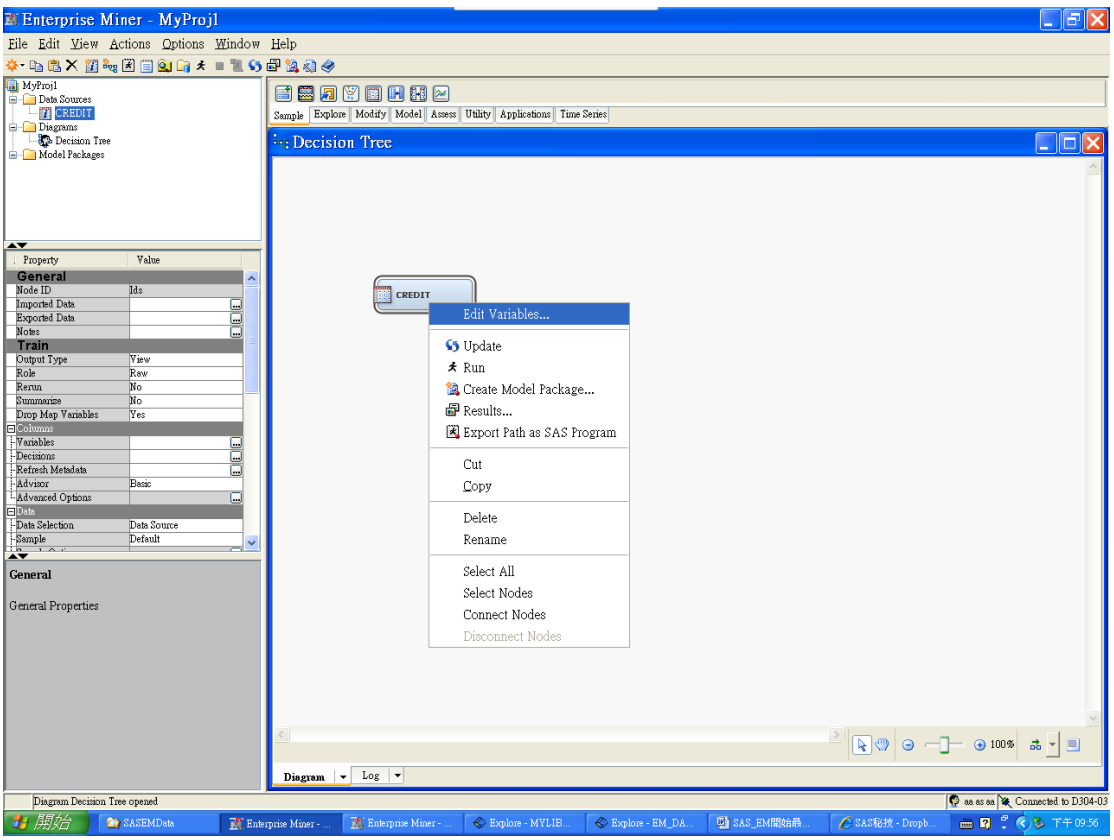
4. Diagram 流程圖（Decision Tree）

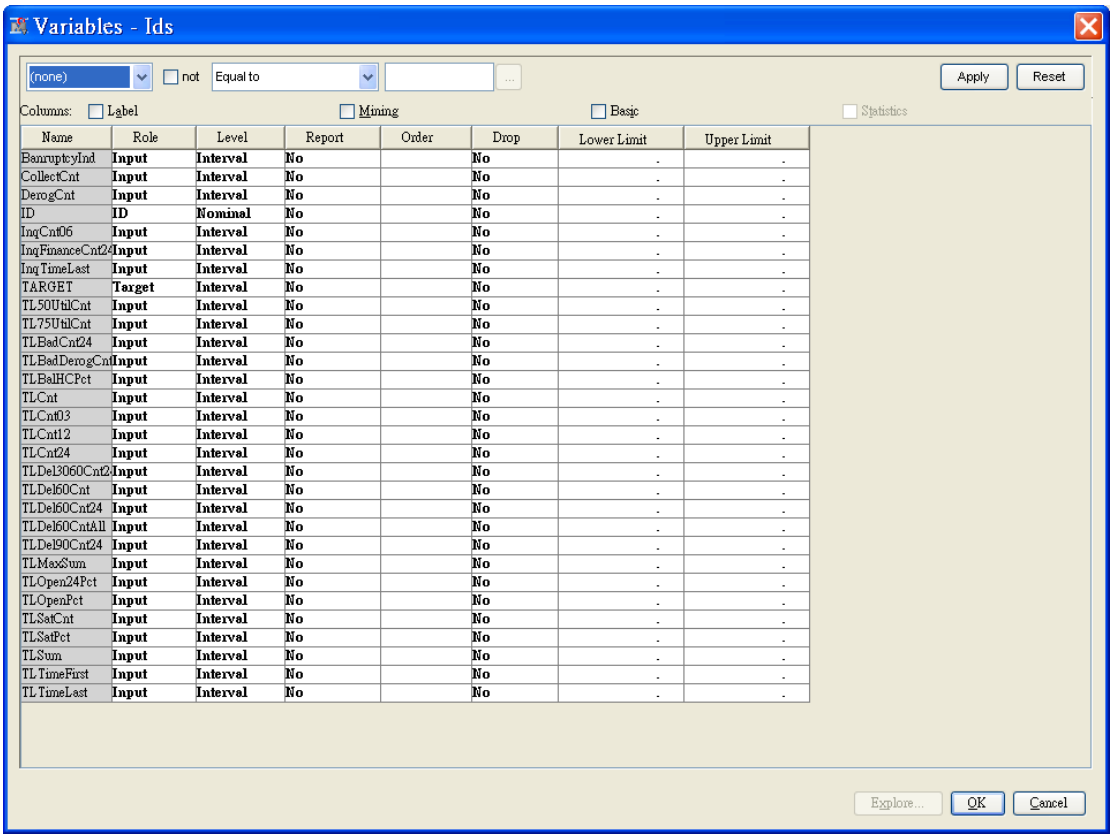


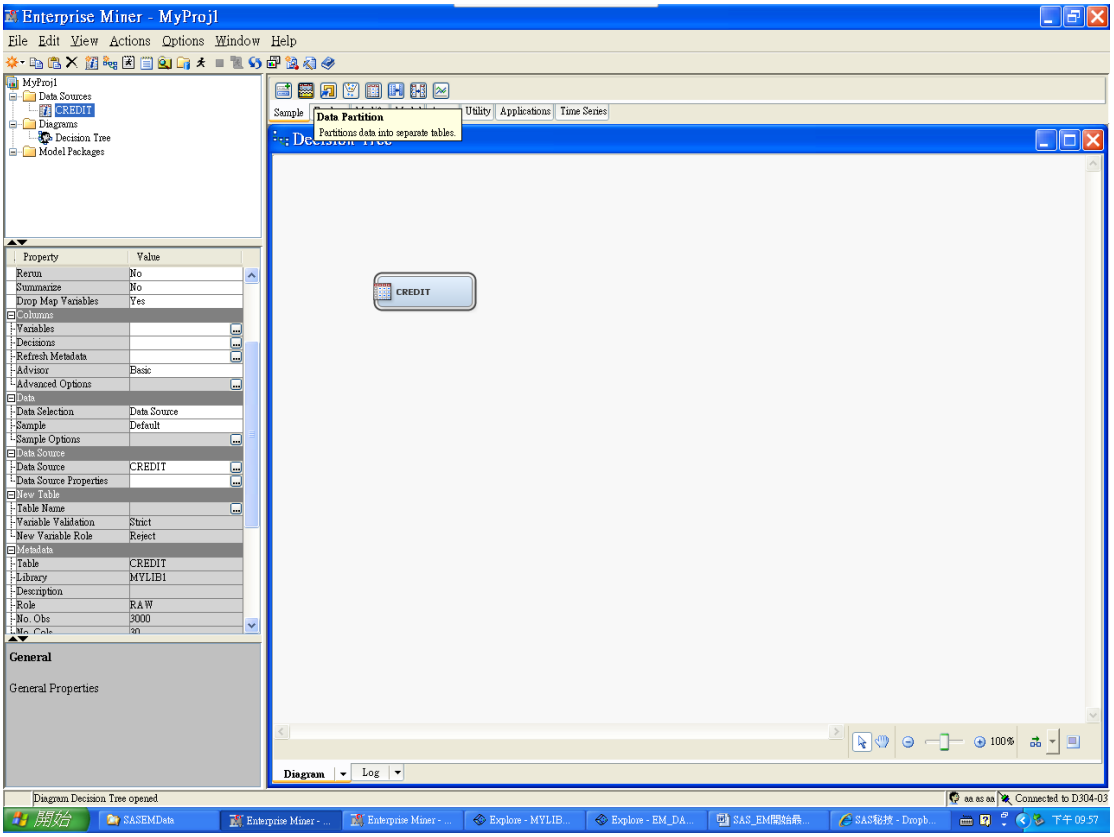


增加節點：（拖拉）

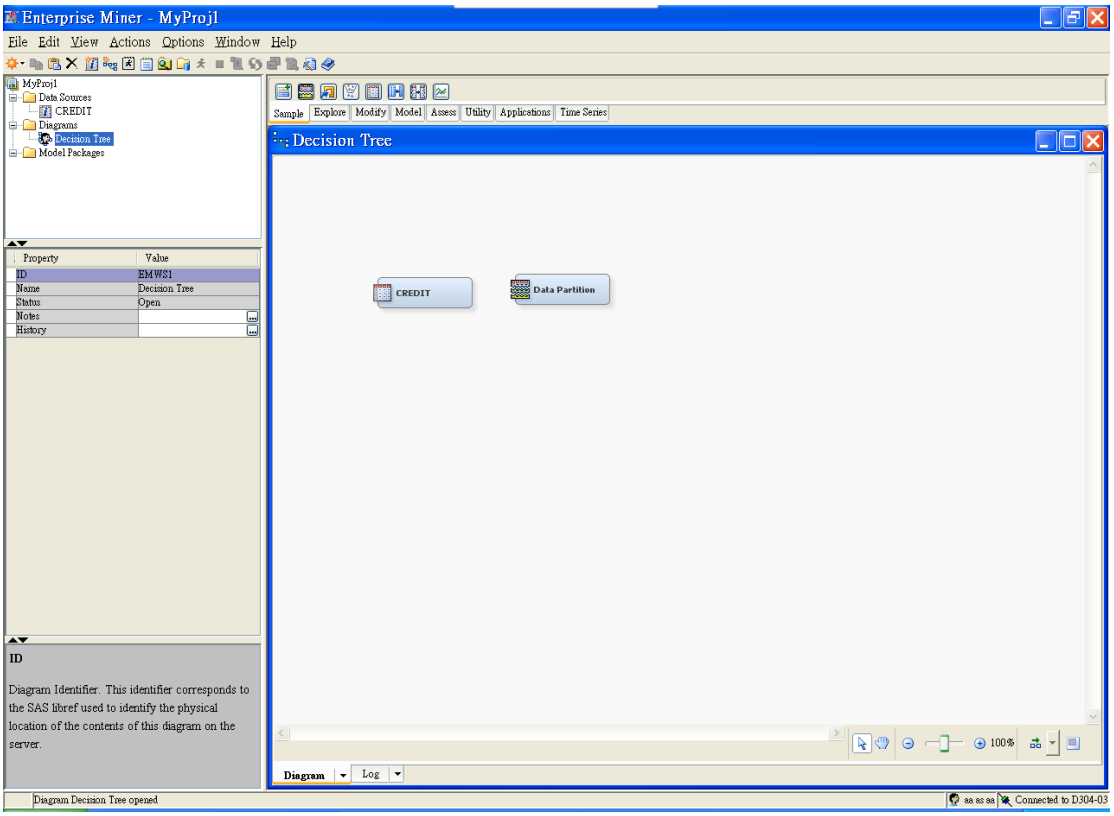

Enterprise Miner - MyProj1
File Edit View Actions Options Window Help
MyProj1
Data Sources
CREDIT
Diagrams
Decision Tree
Model Packages
Sample Explore Modify Model Assess Utility Applications Time Series
Decision Tree
CREDIT
Property Value
ID credit
Name CREDIT
Variables
Decisions
Role Raw
Notes
Library MYLIB1
Table CREDIT
Sample Data Set
Size Type
Sample Size
Type DATA
No. Obs 3000
No. Cols 30
No. Bytes
Segment
Created By aa
Create Date 3/30/12 9:28 PM
Modified By aa
Modify Date 3/30/12 9:28 PM
ID
Data Source identifier. The metadata tables associated with the data source are stored in the EMDS SAS library and use this identifier as the prefix for naming these tables.
Diagram Log
Diagram Decision Tree opened
Connected to D304-03

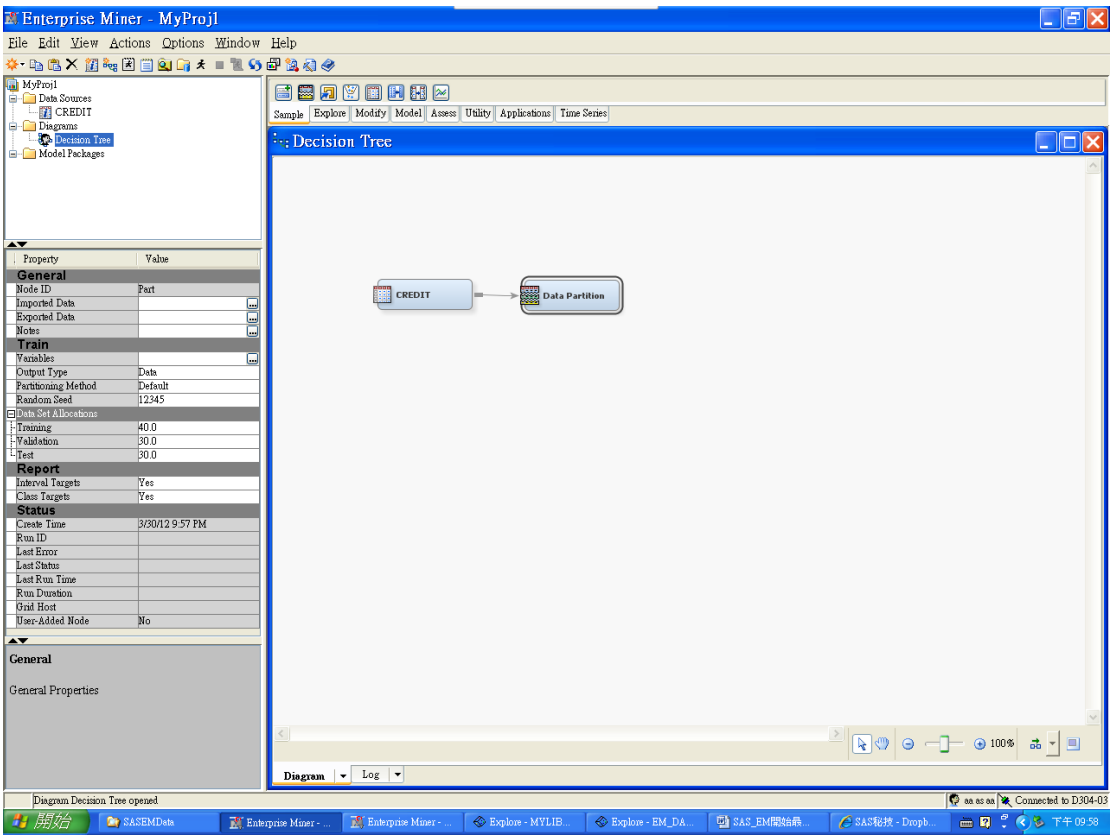

Enterprise Miner - MyProj1
File Edit View Actions Options Window Help
MyProj1
Data Sources
CREDIT
Diagrams
Decision Tree
Model Packages
Sample Explore Modify Model Assess Utility Applications Time Series
Decision Tree
CREDIT
Edit Variables...
Update
Run
Create Model Package...
Results...
Export Path as SAS Program
Cut
Copy
Delete
Rename
Select All
Select Nodes
Connect Nodes
Disconnect Nodes
Property Value
General
Node ID Ids
Imported Data
Exported Data
Notes
Train
Output Type View
Role Raw
Rerun No
Summarize No
Drop Map Variables Yes
Columns
Variables
Decisions
Refresh Metadata
Advisor Basic
Advanced Options
Data
Data Selection Data Source
Sample Default
General
General Properties
Diagram Log
Diagram Decision Tree opened
Connected to D304-03



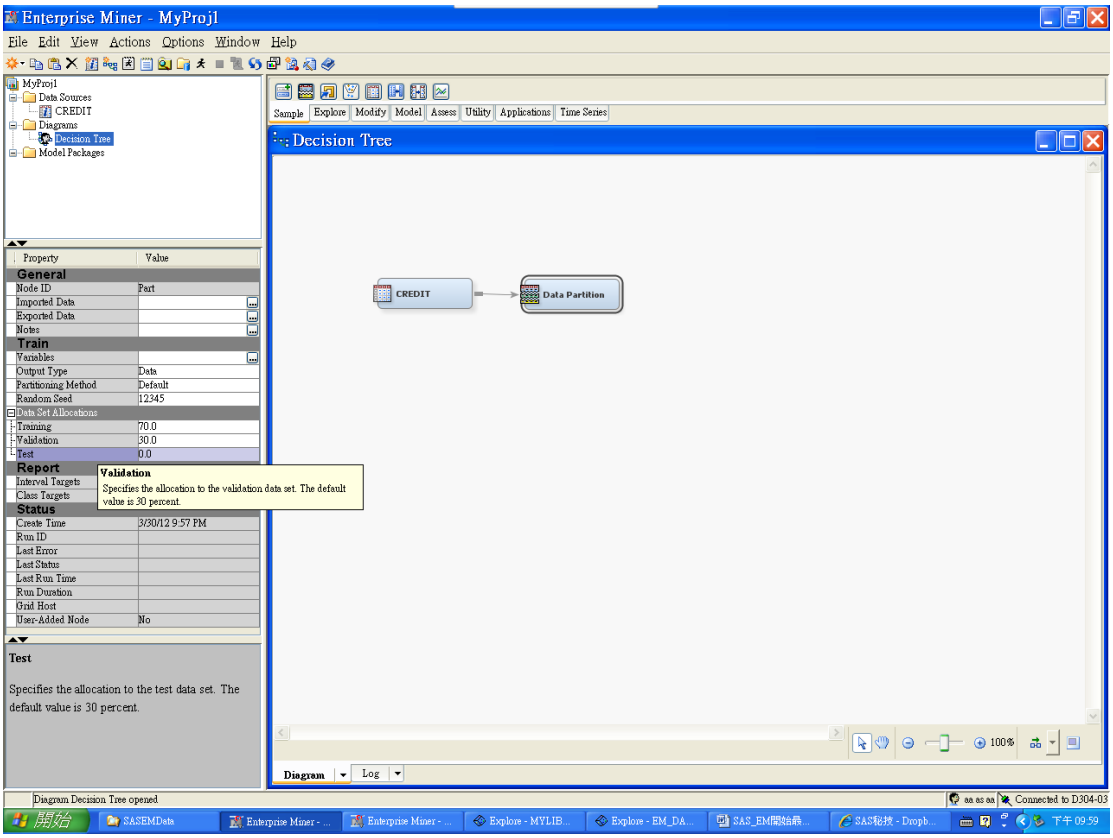


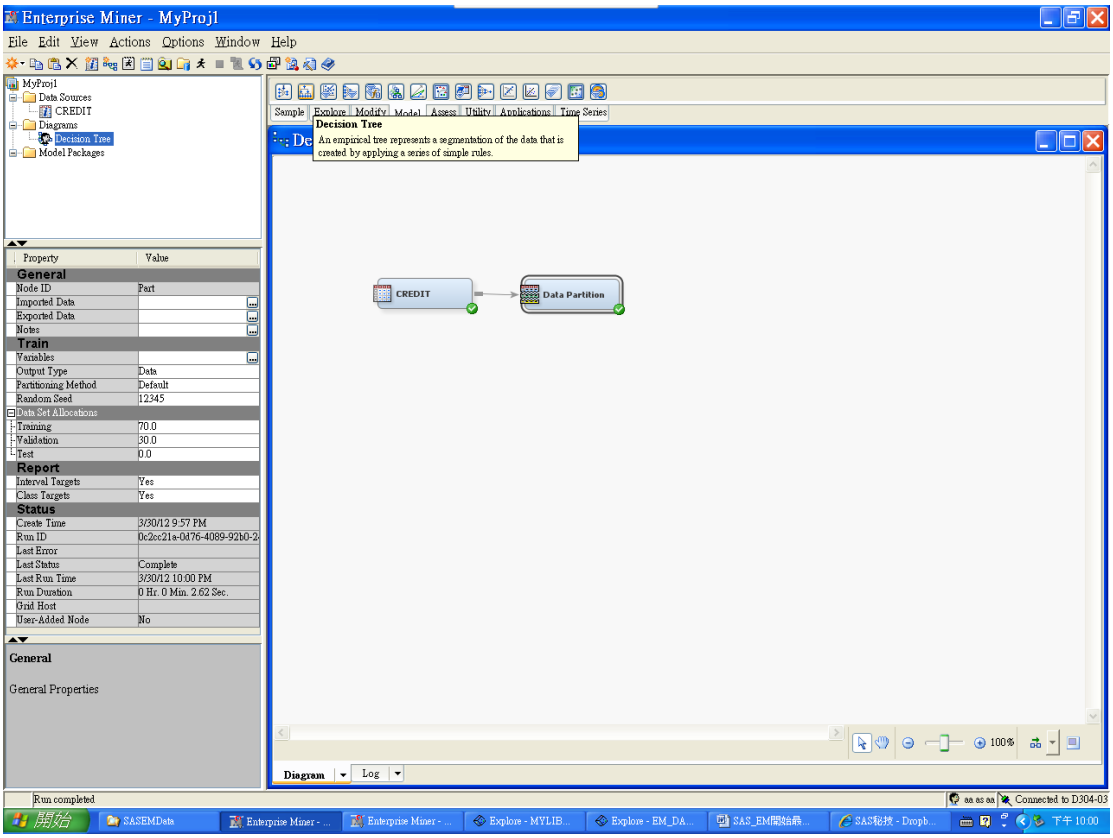
要將 3000 筆分成訓練資料 70%跟驗證資料 30%


Enterprise Miner - MyProj1
File Edit View Actions Options Window Help
MyProj1
Data Sources
CREDIT
Diagrams
Decision Tree
Model Packages
Property
Value
ID
EMWS1
Name
Decision Tree
Status
Open
Notes
History
ID
Diagram Identifier. This identifier corresponds to the SAS libref used to identify the physical location of the contents of this diagram on the server.
Sample
Explore
Modify
Model
Assess
Utility
Applications
Time Series
Decision Tree
CREDIT
Data Partition
100%
Diagram
Log
Diagram Decision Tree opened
Connected to D304-03

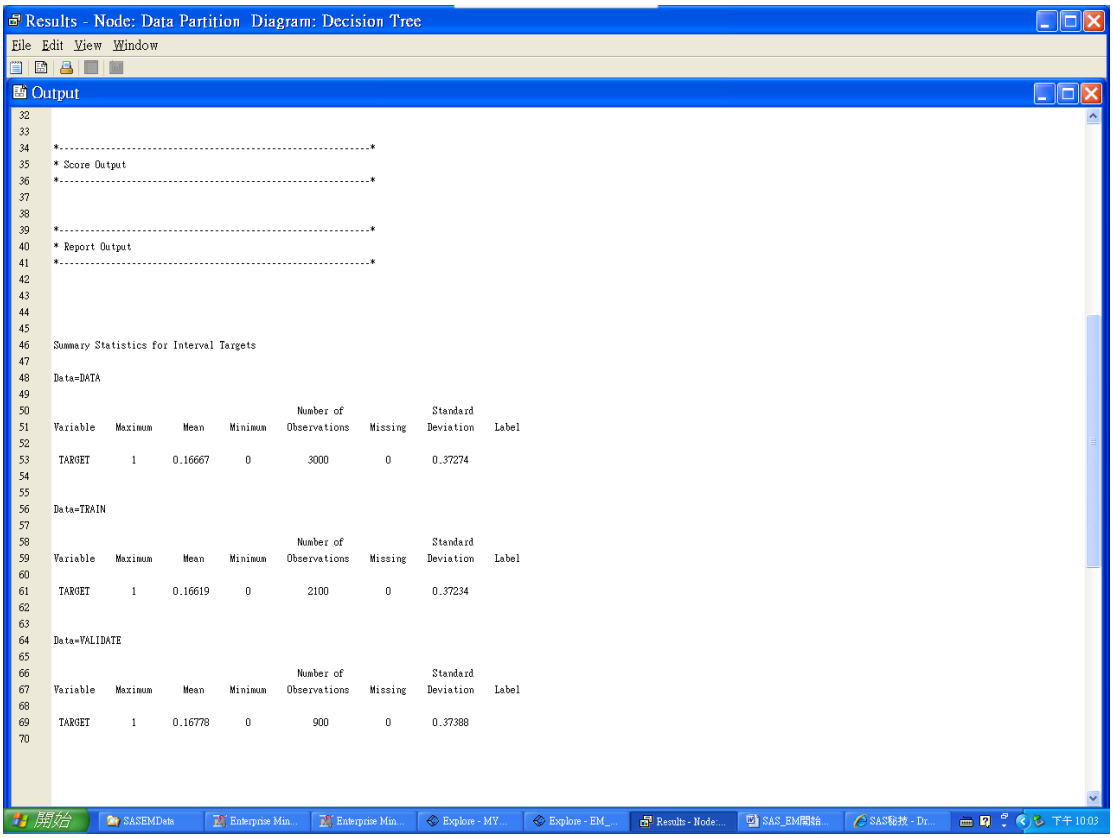

Enterprise Miner - MyProj1
File Edit View Actions Options Window Help
MyProj1
Data Sources
CREDIT
Diagrams
Decision Tree
Model Packages
Property
Value
General
Node ID
Part
Imported Data
Exported Data
Notes
Train
Variables
Output Type
Data
Partitioning Method
Default
Random Seed
12345
Data Set Allocations
Training
40.0
Validation
30.0
Test
30.0
Report
Interval Targets
Yes
Class Targets
Yes
Status
Create Time
3/30/12 9:57 PM
Run ID
Last Error
Last Status
Last Run Time
Run Duration
Grid Host
User-Added Node
No
General
General Properties
Sample
Explore
Modify
Model
Assess
Utility
Applications
Time Series
Decision Tree
CREDIT
Data Partition
100%
Diagram
Log
Diagram Decision Tree opened
Connected to D304-03

改測試筆數比例：

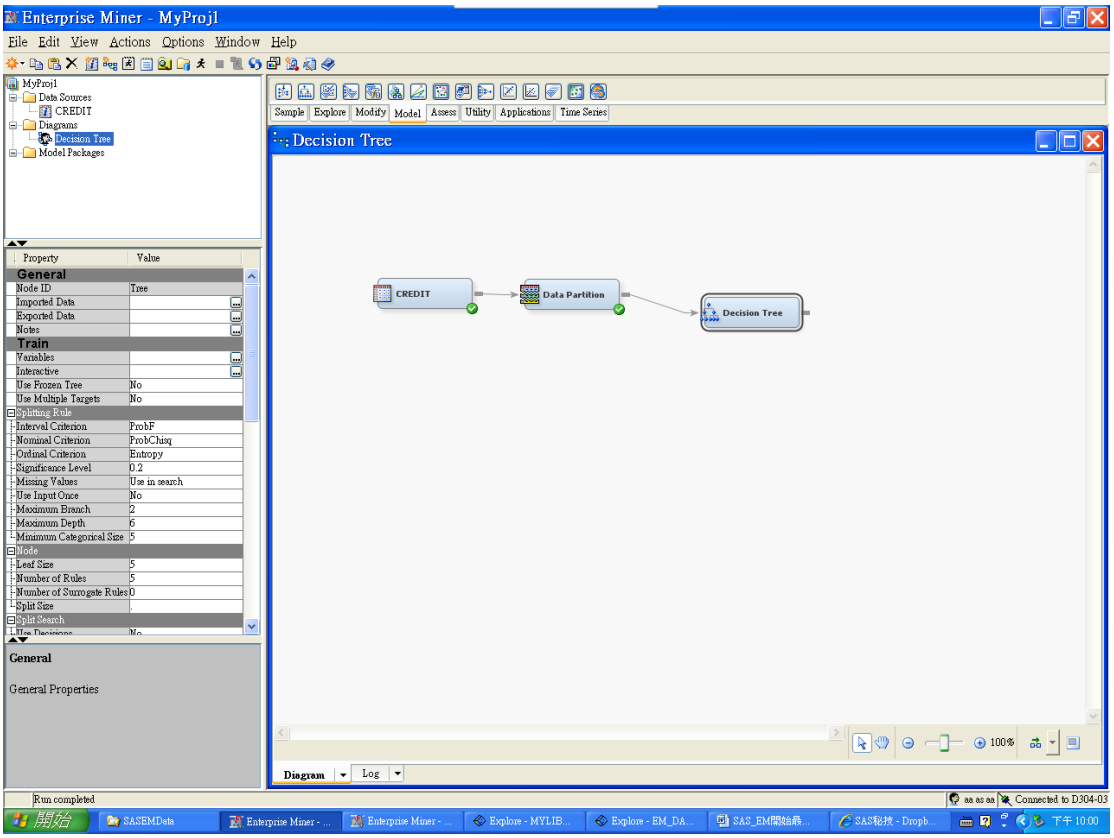

Enterprise Miner - MyProj1
File Edit View Actions Options Window Help
MyProj1
Data Sources
CREDIT
Diagrams
Decision Tree
Model Packages
Sample Explore Modify Model Assess Utility Applications Time Series
Decision Tree
CREDIT
Data Partition
Property Value
General
Node ID Part
Imported Data
Exported Data
Notes
Train
Variables
Output Type Data
Partitioning Method Default
Random Seed 12345
Data Set Allocations
Training 70.0
Validation 30.0
Test 0.0
Report
Interval Targets
Class Targets
Status
Create Time 3/30/12 9:57 PM
Run ID
Last Error
Last Status
Last Run Time
Run Duration
Grid Host
User-Added Node No
Validation
Specifies the allocation to the validation data set. The default value is 30 percent.
Test
Specifies the allocation to the test data set. The default value is 30 percent.
100%
Diagram Log
Diagram Decision Tree opened
Connected to D304-03

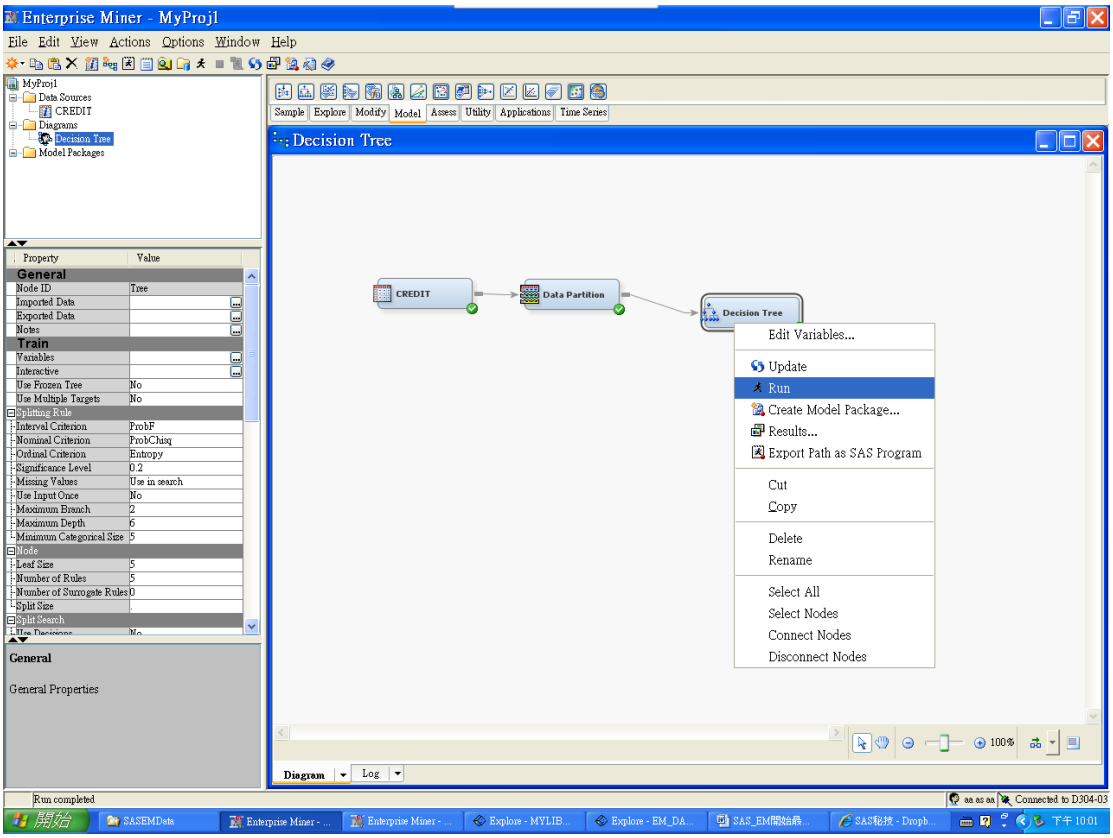

Enterprise Miner - MyProj1
File Edit View Actions Options Window Help
MyProj1
Data Sources
CREDIT
Diagrams
Decision Tree
Model Packages
Sample Explore Modify Model Assess Utility Applications Time Series
Decision Tree
An empirical tree represents a segmentation of the data that is created by applying a series of simple rules.
CREDIT
Data Partition
Property Value
General
Node ID Part
Imported Data
Exported Data
Notes
Train
Variables
Output Type Data
Partitioning Method Default
Random Seed 12345
Data Set Allocations
Training 70.0
Validation 30.0
Test 0.0
Report
Interval Targets Yes
Class Targets Yes
Status
Create Time 3/30/12 9:57 PM
Run ID 0c2cc21a-0d76-4089-92b0-2
Last Error
Last Status Complete
Last Run Time 3/30/12 10:00 PM
Run Duration 0 Hr. 0 Min. 2.62 Sec.
Grid Host
User-Added Node No
General
General Properties
100%
Diagram Log
Run completed
Connected to D304-03

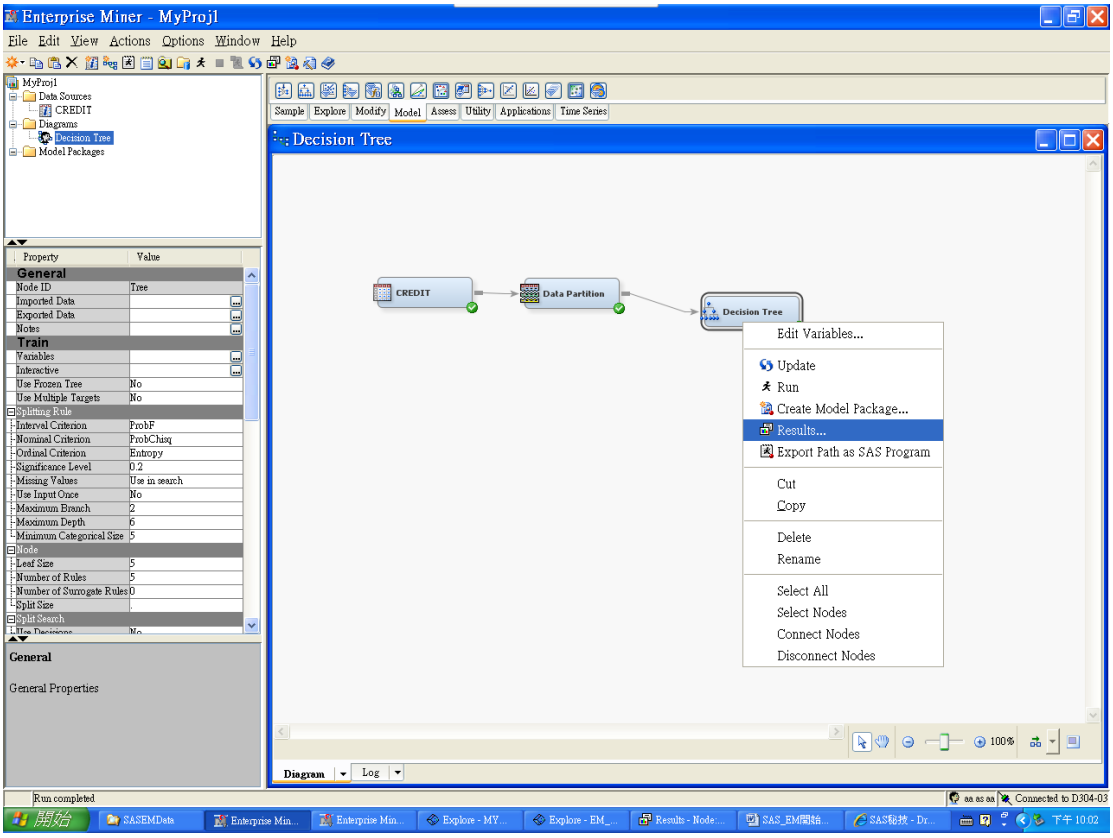
結果：


Results - Node: Data Partition Diagram: Decision Tree
File Edit View Window
Output
--
* Score Output
--
--
* Report Output
--
Summary Statistics for Interval Targets
Data=DATA
Variable Maximum Mean Minimum Number of Observations Missing Standard Deviation Label
TARGET 1 0.16667 0 3000 0 0.37274
Data=TRAIN
Variable Maximum Mean Minimum Number of Observations Missing Standard Deviation Label
TARGET 1 0.16619 0 2100 0 0.37234
Data=VALIDATE
Variable Maximum Mean Minimum Number of Observations Missing Standard Deviation Label
TARGET 1 0.16778 0 900 0 0.37388

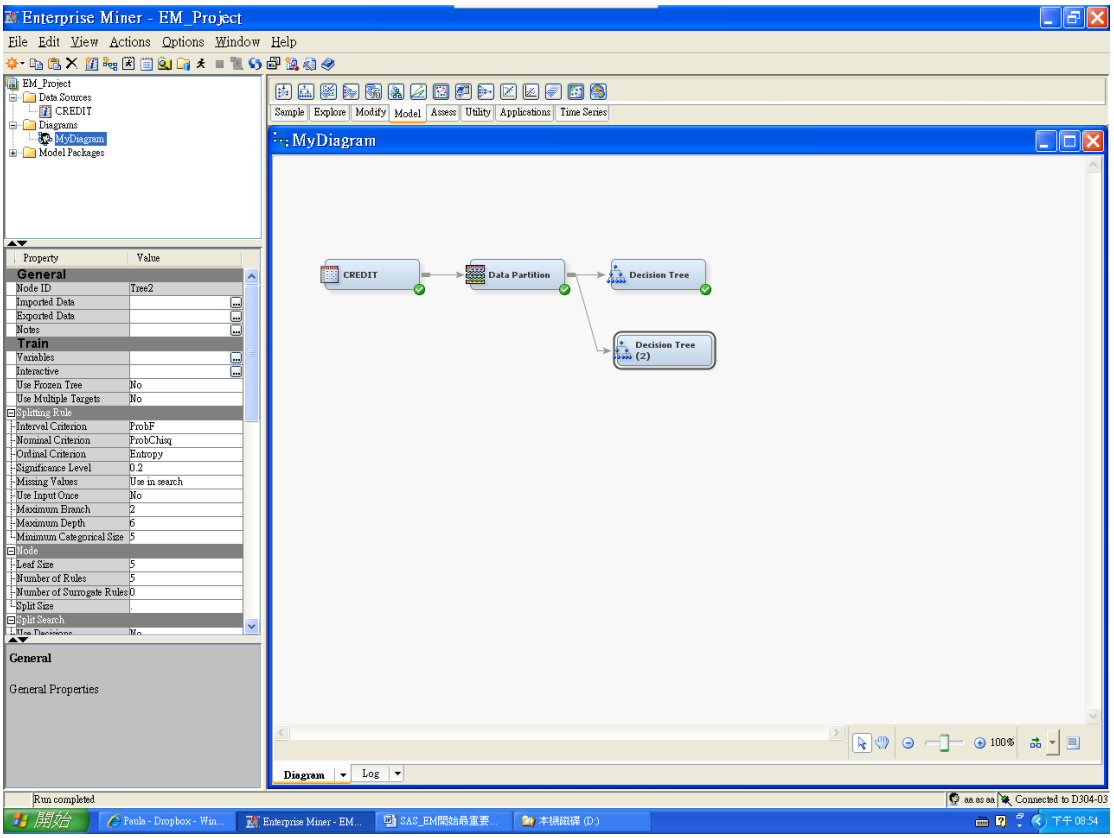
加入決策樹


Enterprise Miner - MyProj1
File Edit View Actions Options Window Help
MyProj1
Data Sources
CREDIT
Diagrams
Decision Tree
Model Packages
Sample Explore Modify Model Assess Utility Applications Time Series
Decision Tree
CREDIT
Data Partition
Decision Tree
Property Value
General
Node ID Tree
Imported Data
Exported Data
Notes
Train
Variables
Interactive
Use Frozen Tree No
Use Multiple Targets No
Splitting Rule
Interval Criterion ProbF
Nominal Criterion ProbChisq
Ordinal Criterion Entropy
Significance Level 0.2
Missing Values Use in search
Use Input Once No
Maximum Branch 2
Maximum Depth 6
Minimum Categorical Size 5
Node
Leaf Size 5
Number of Rules 5
Number of Surrogate Rules 0
Split Size
Split Search
General
General Properties
100%
Diagram
Log
Run completed
Connected to D304-03

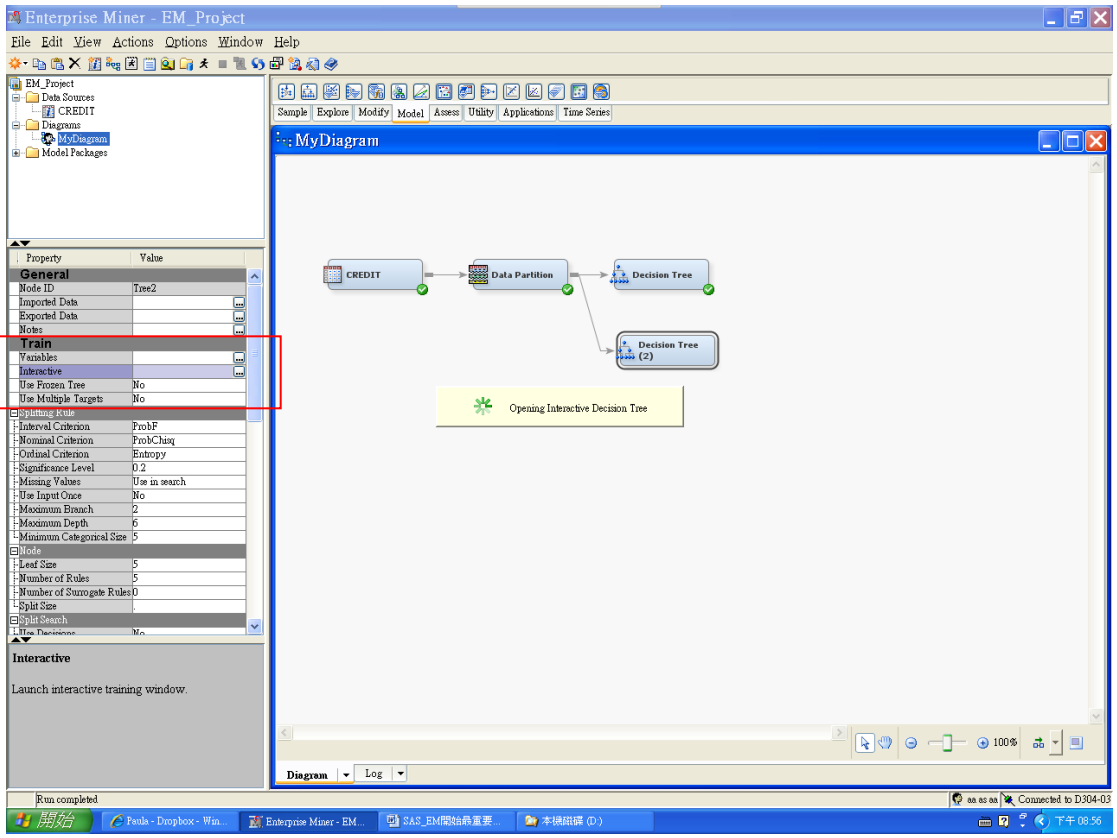

Enterprise Miner - MyProj1
File Edit View Actions Options Window Help
Decision Tree
CREDIT
Data Partition
Decision Tree
Edit Variables...
Update
Run
Create Model Package...
Results...
Export Path as SAS Program
Cut
Copy
Delete
Rename
Select All
Select Nodes
Connect Nodes
Disconnect Nodes
Diagram
Log
Run completed


Enterprise Miner - MyProj1
File Edit View Actions Options Window Help
Decision Tree
CREDIT
Data Partition
Decision Tree
Edit Variables...
Update
Run
Create Model Package...
Results...
Export Path as SAS Program
Cut
Copy
Delete
Rename
Select All
Select Nodes
Connect Nodes
Disconnect Nodes
Diagram
Log
Run completed

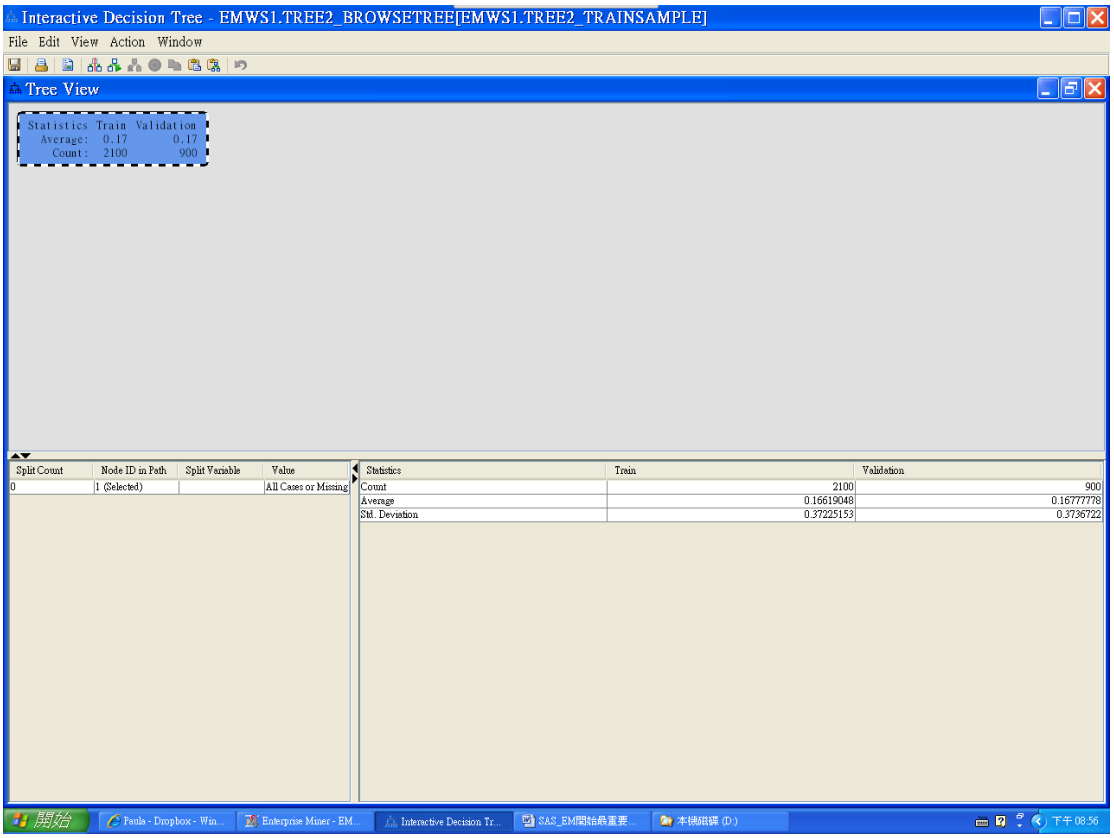
5. 自建決策樹

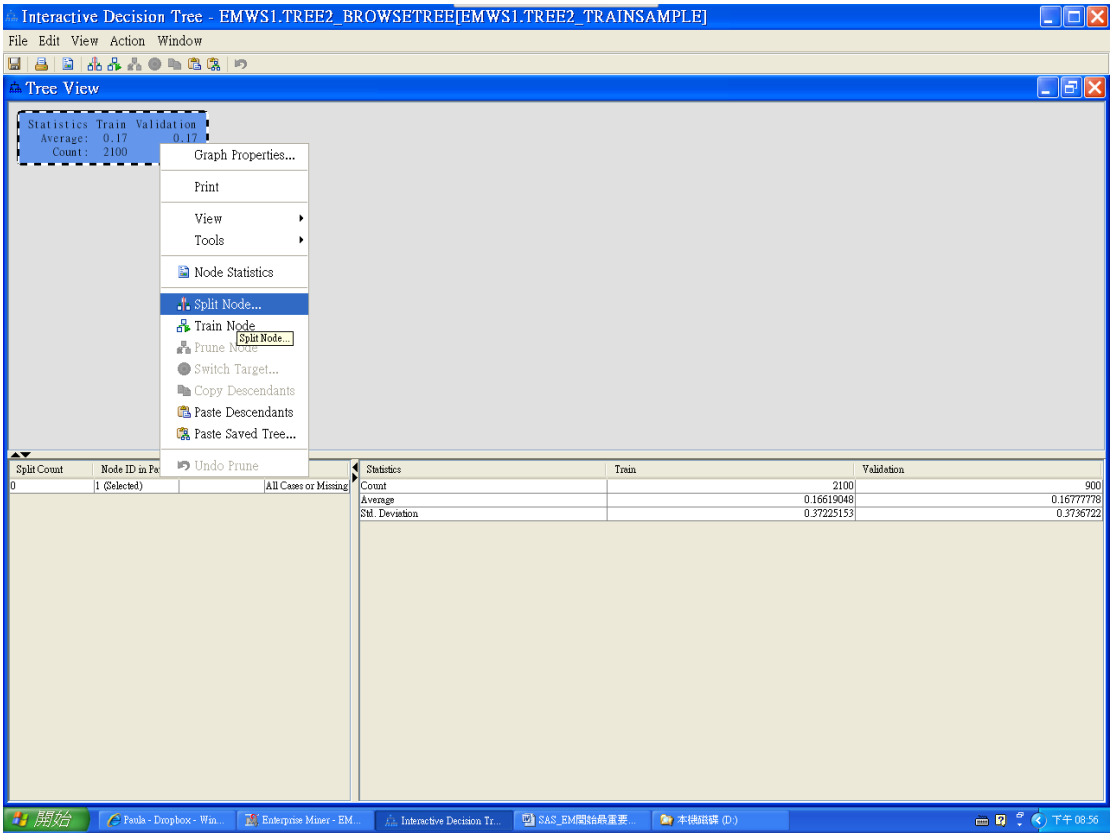


Train:Interactive

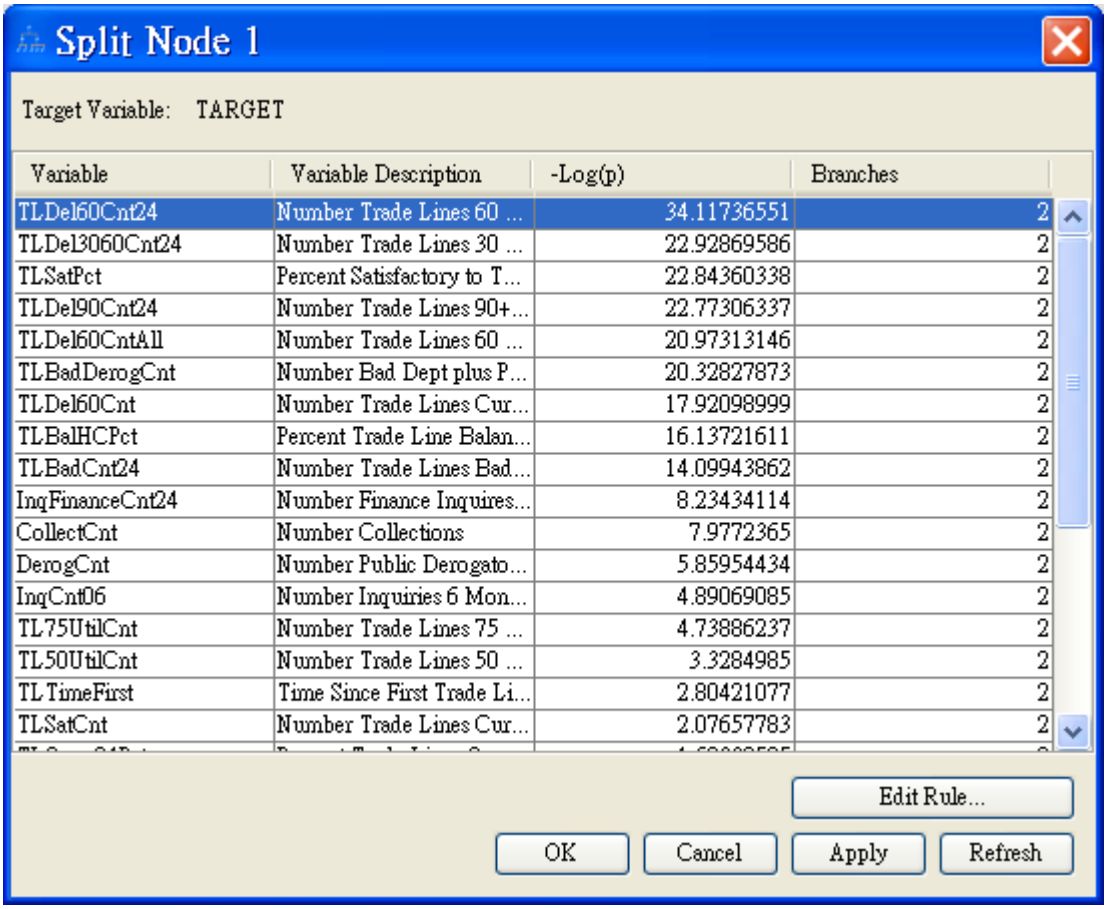


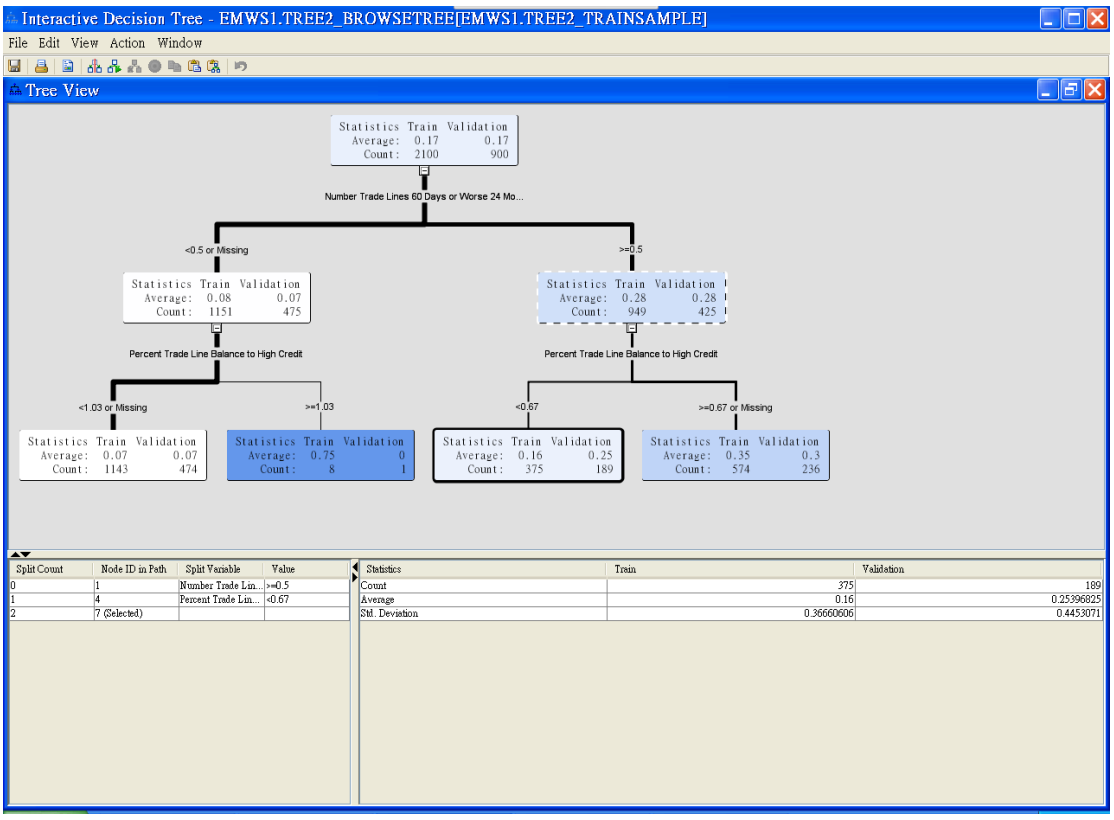
以下為：target 的 level 屬性設為：Interval：在 decision tree 會產生平均值，就會出現以下畫面





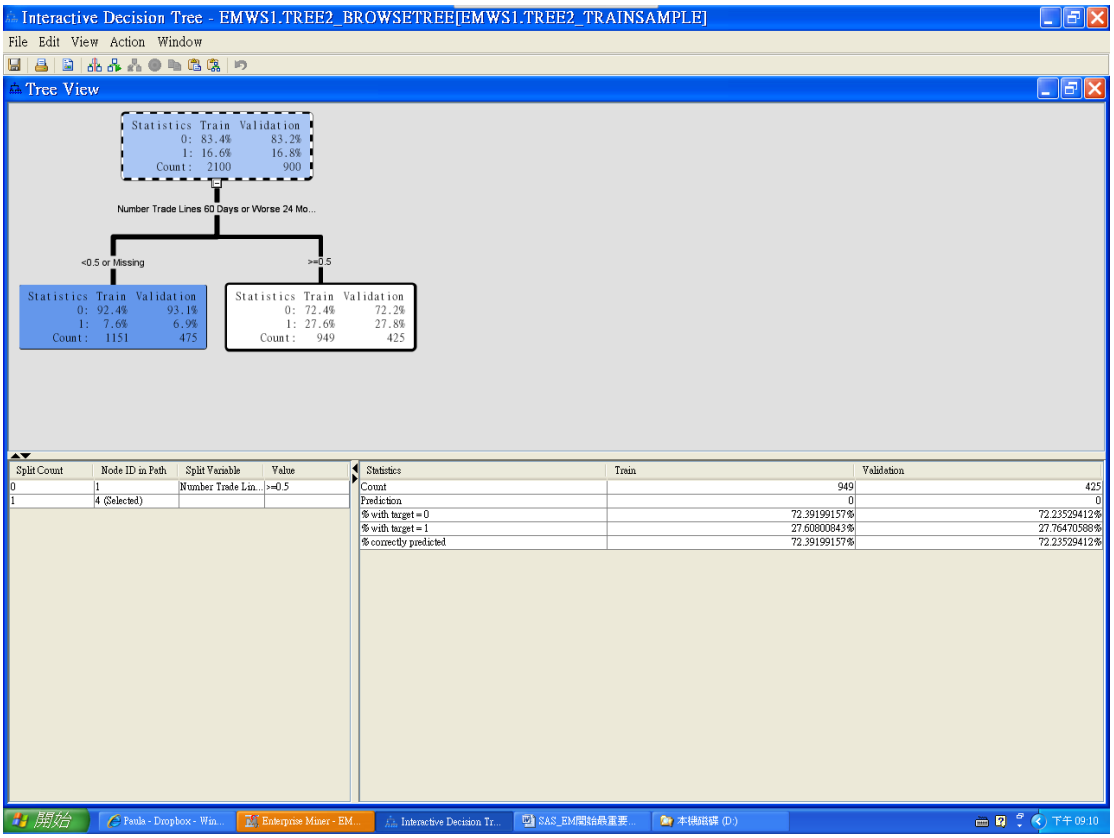
-Log(p)愈大，變數愈重要（影響力比較大）

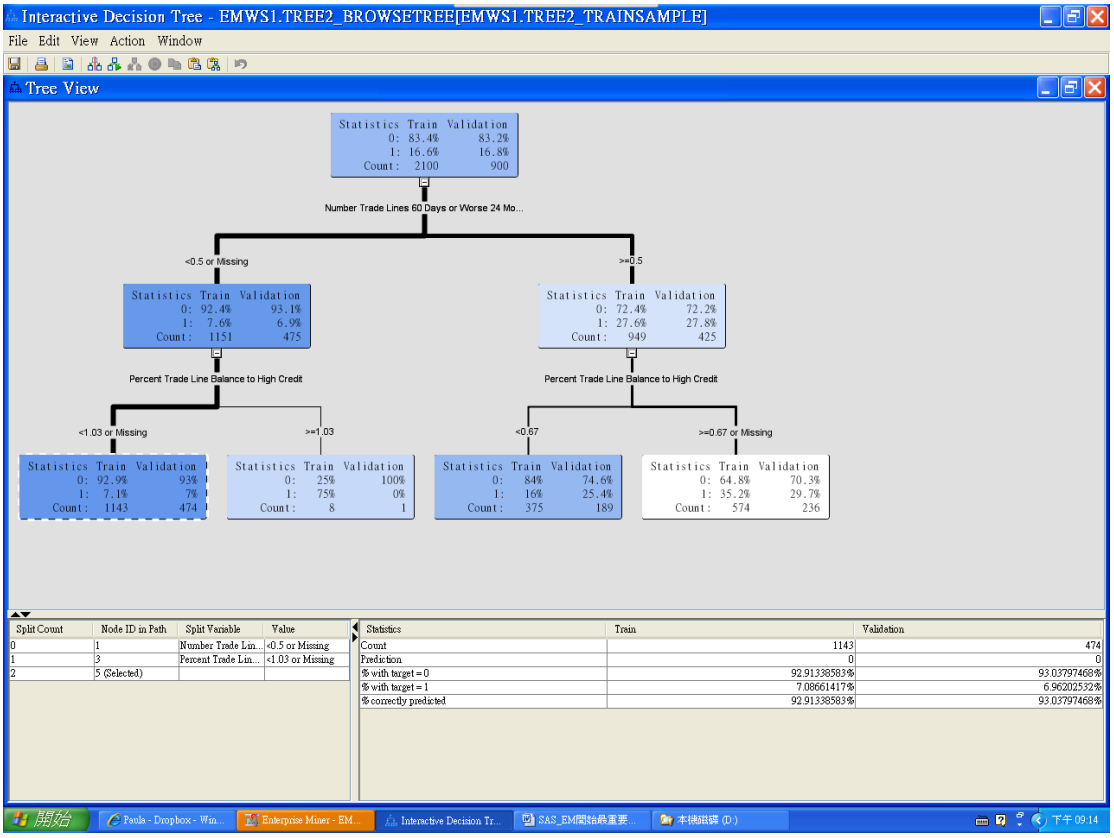




粗線代表：人數最大宗(一般性 rule)，若經費有限就是會先選擇

以下為：target 的 level 屬性設為：Binary 為 0/1 or yes/no（如：0=好 1=壞客戶、yes=會 no=不會買）

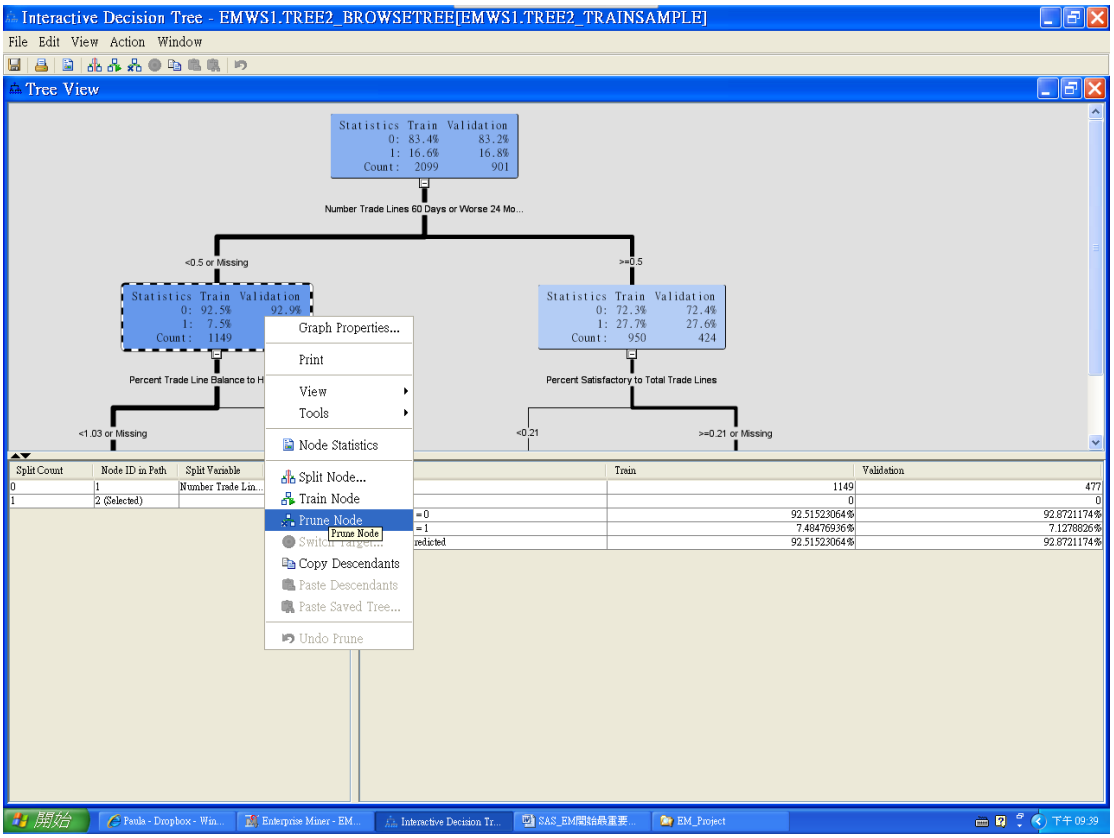


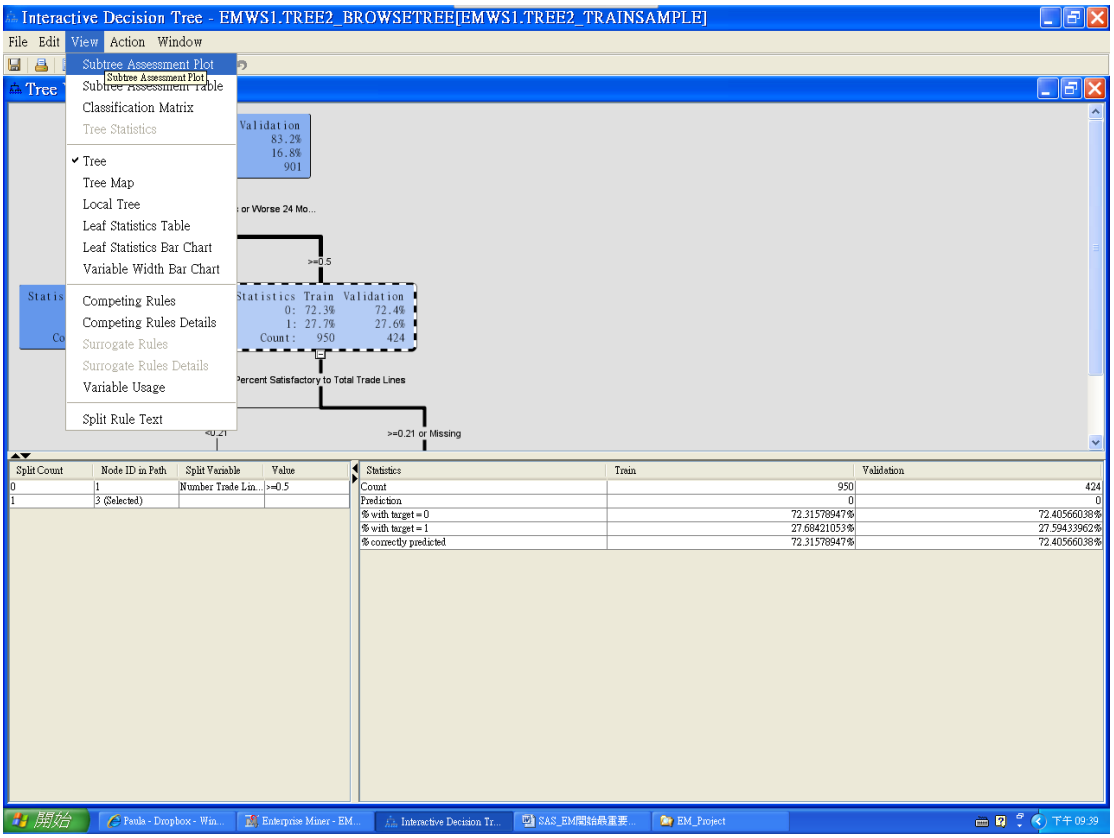


粗線代表：人數最大宗(一般性 rule)，若經費有限就是會先選擇

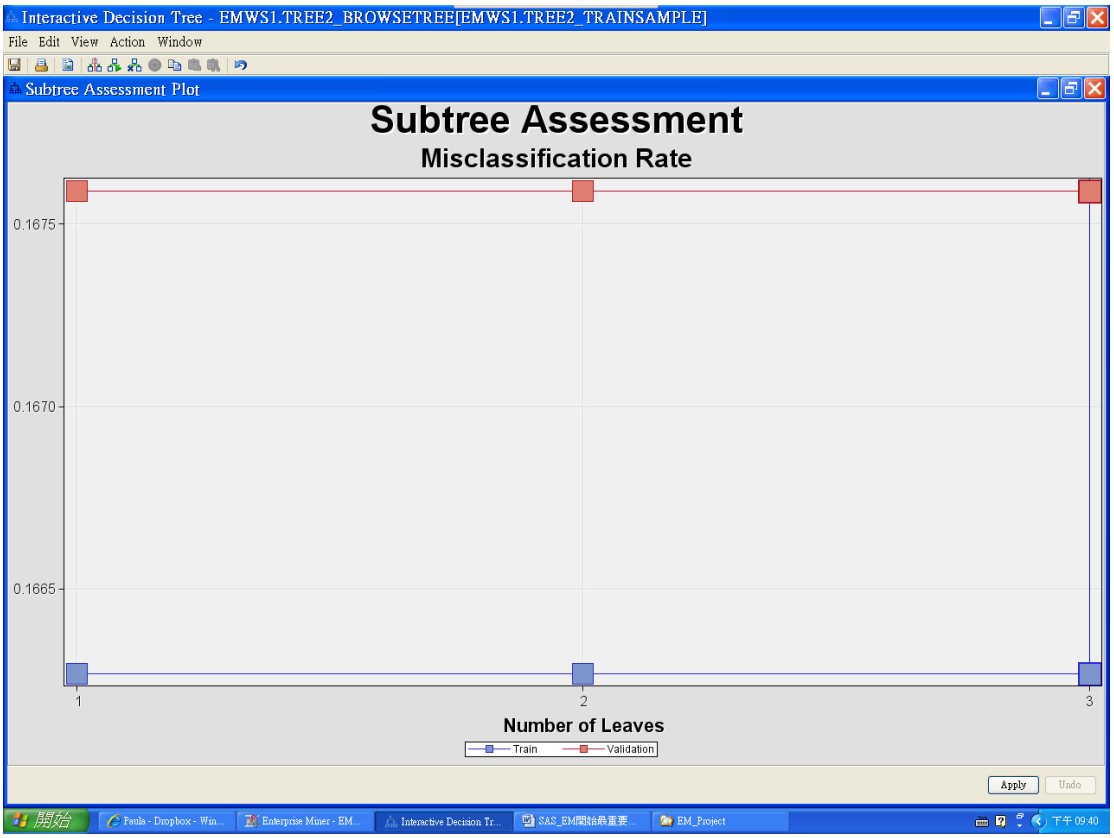
（粗細代表人數的佔比）

修技：

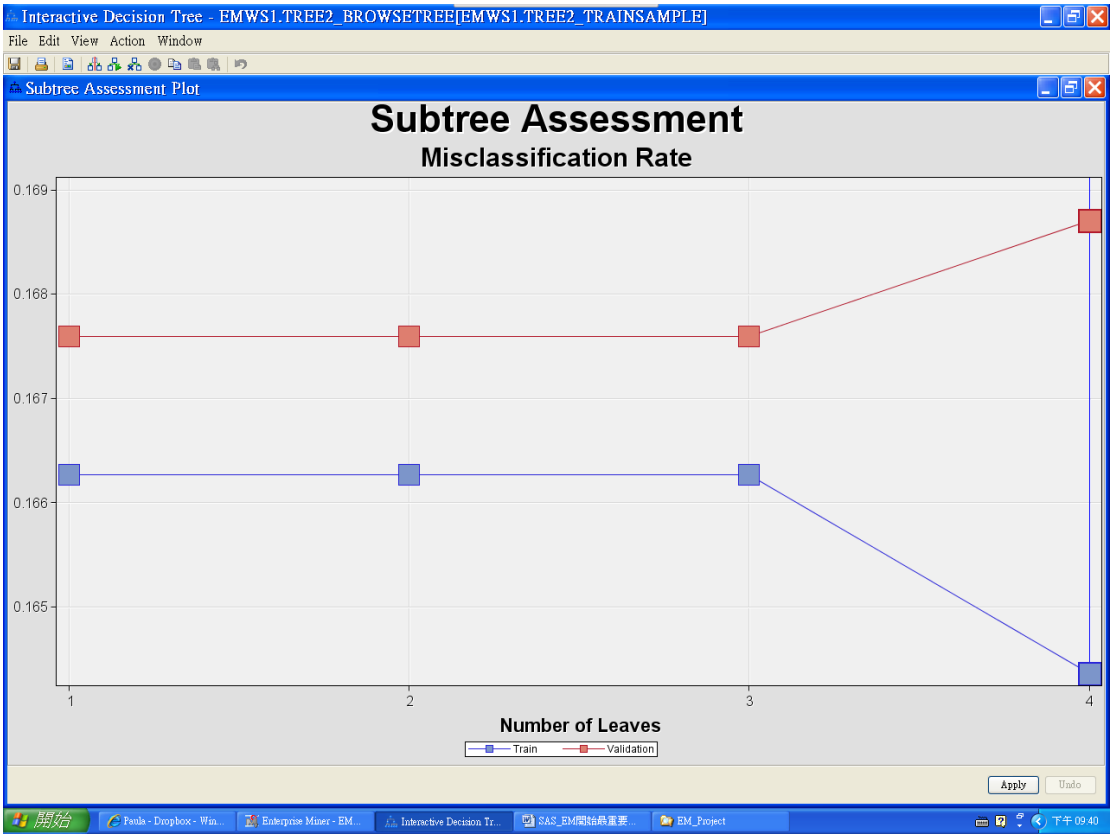

Interactive Decision Tree - EMWS1.TREE2_BROWSETREE[EMWS1.TREE2_TRAINSAMPLE]
File Edit View Action Window
Tree View
Statistics Train Validation
0: 83.4% 83.2%
1: 16.6% 16.8%
Count: 2099 901
Number Trade Lines 60 Days or Worse 24 Mo...
<0.5 or Missing
>=0.5
Statistics Train Validation
0: 92.5% 92.9%
1: 7.5%
Count: 1149
Statistics Train Validation
0: 72.3% 72.4%
1: 27.7% 27.6%
Count: 950 424
Graph Properties...
Print
View
Tools
Node Statistics
Split Node...
Train Node
Prune Node
Copy Descendants
Paste Descendants
Paste Saved Tree...
Undo Prune
Percent Trade Line Balance to H
Percent Satisfactory to Total Trade Lines
<1.03 or Missing
<0.21
>=0.21 or Missing
Split Count
Node ID in Path
Split Variable
Number Trade Lin...
Train
Validation
1149
477
92.51523064%
92.8721174%
7.48476936%
7.1278826%
開始
下午 09:39


Interactive Decision Tree - EMWS1.TREE2_BROWSETREE[EMWS1.TREE2_TRAINSAMPLE]
File Edit View Action Window
Subtree Assessment Plot
Subtree Assessment Table
Classification Matrix
Tree Statistics
Tree
Tree Map
Local Tree
Leaf Statistics Table
Leaf Statistics Bar Chart
Variable Width Bar Chart
Competing Rules
Competing Rules Details
Surrogate Rules
Surrogate Rules Details
Variable Usage
Split Rule Text
Statistics Train Validation
0: 72.3% 72.4%
1: 27.7% 27.6%
Count: 950 424
>=0.5
>=0.21 or Missing
Split Count
Node ID in Path
Split Variable
Value
Number Trade Lin...
>=0.5
Statistics
Train
Validation
Count
950
424
Prediction
0
0
% with target = 0
72.31578947%
72.40566038%
% with target = 1
27.68421053%
27.59433962%
% correctly predicted
72.31578947%
72.40566038%
開始
下午 09:39

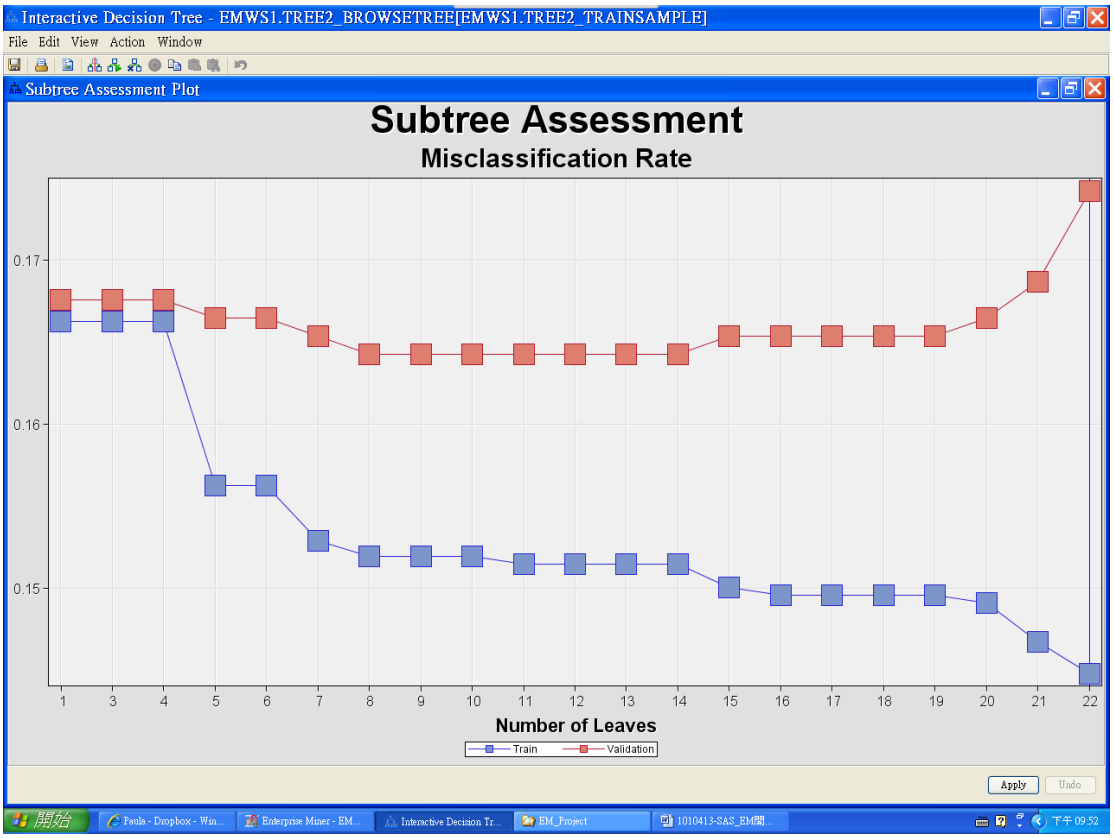
正常：



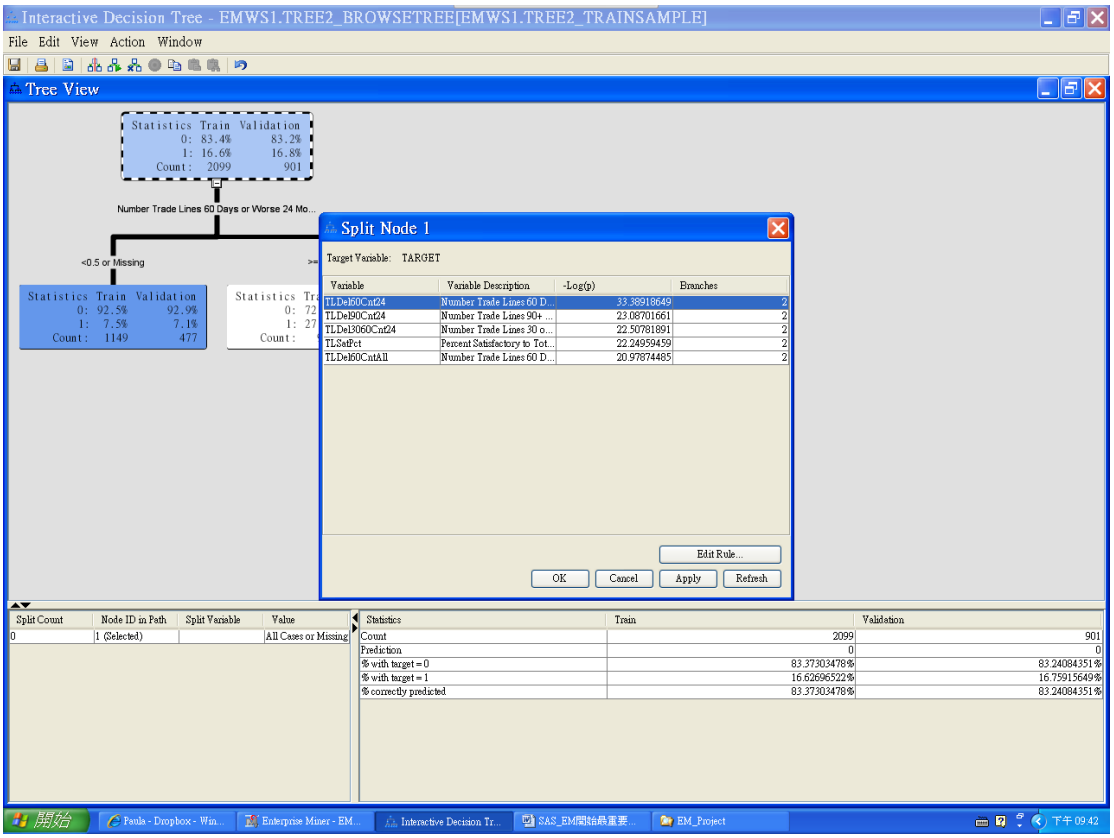
不正常或過於極端：以 Validation 為主，二個若發散就為不好，則要從發散後開始修技（保留 3 ，從 4 開始刪，Validation 要往 train 方向）

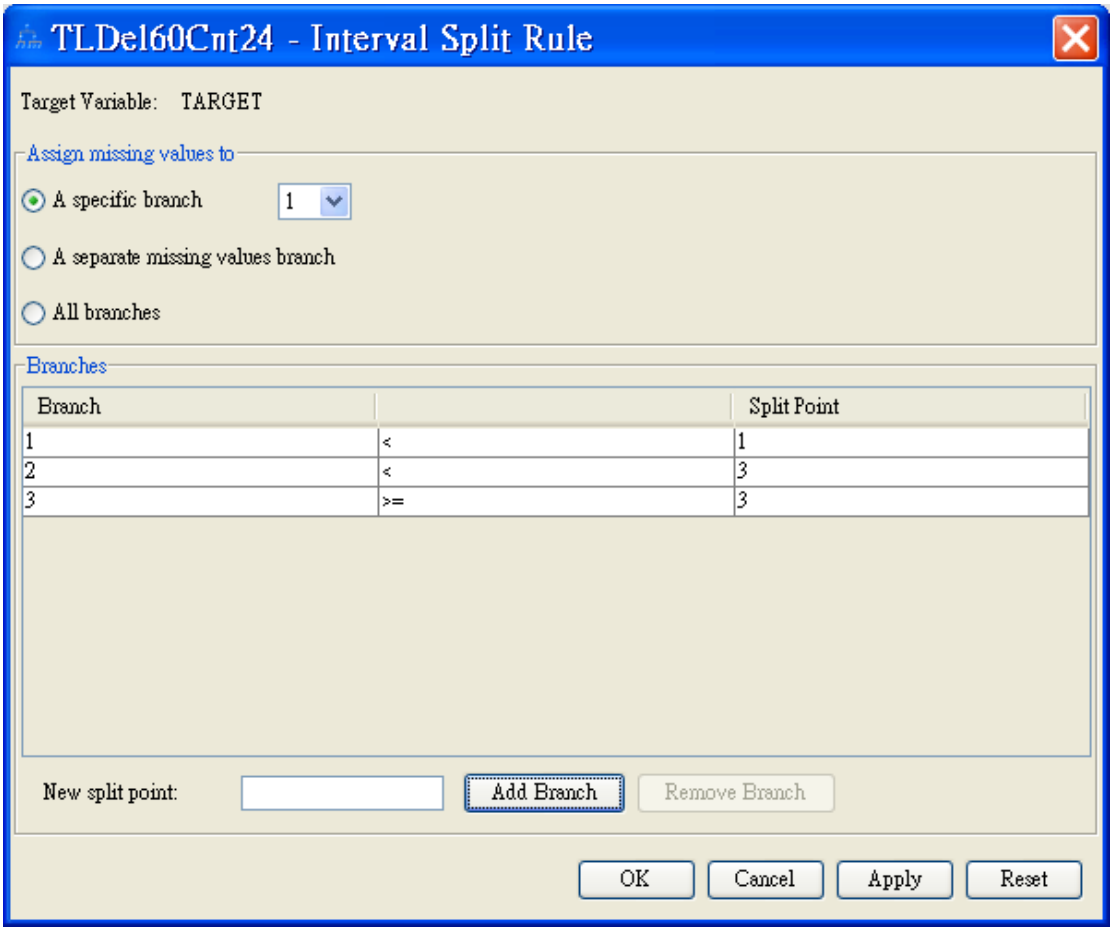


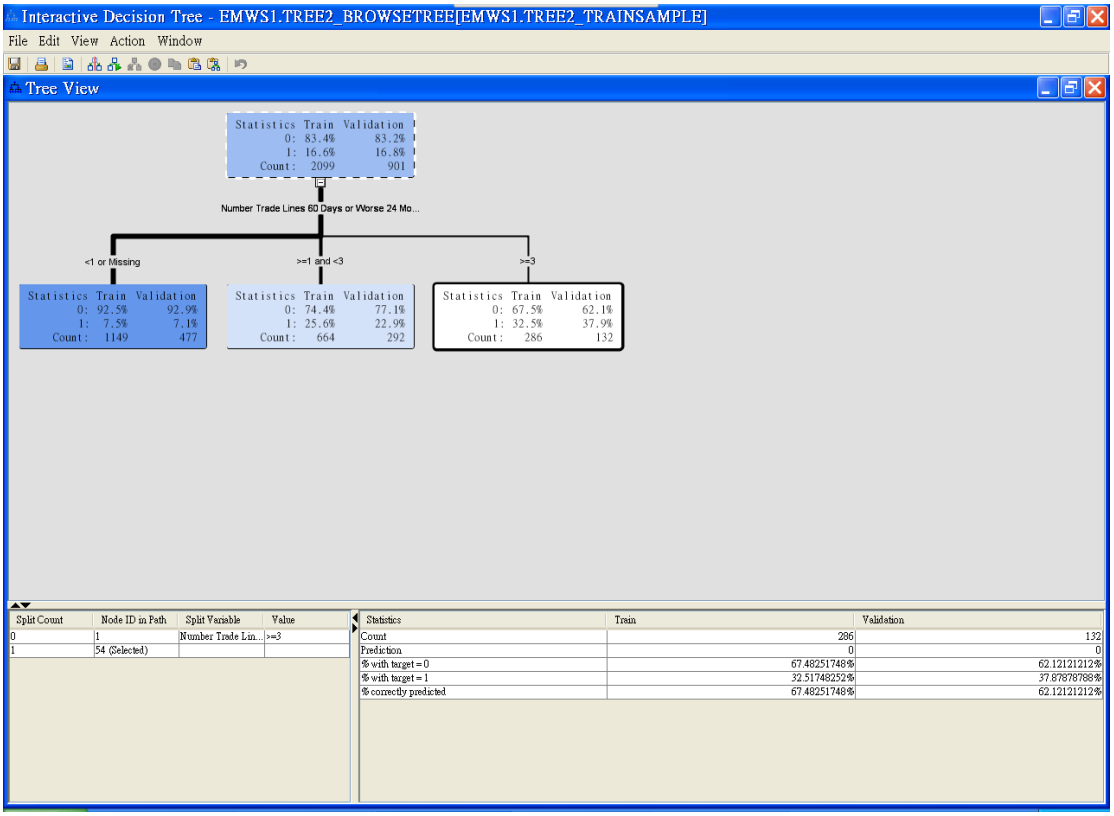
可以從發散的開始修

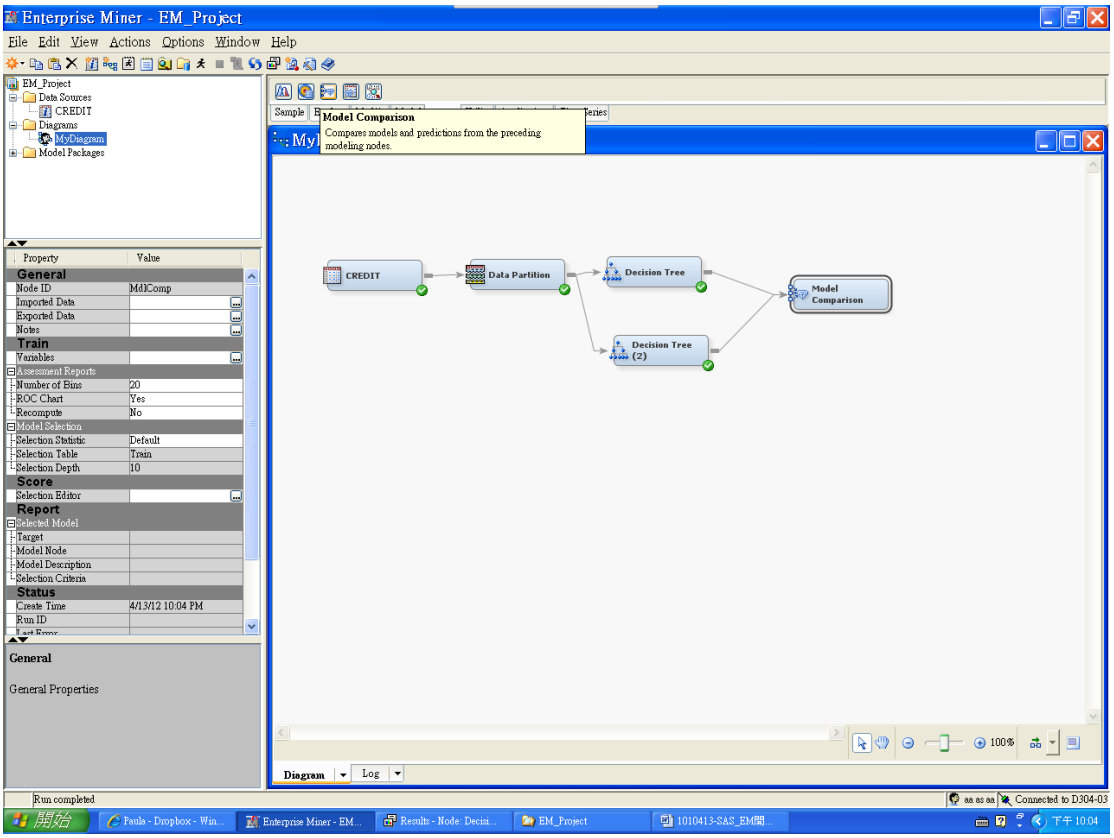

Interactive Decision Tree - EMWS1.TREE2_BROWSETREE[EMWS1.TREE2_TRAINSAMPLE]
File Edit View Action Window
Subtree Assessment Plot
Subtree Assessment
Misclassification Rate
0.17
0.16
0.15
1 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Number of Leaves
Train
Validation
Apply
Undo

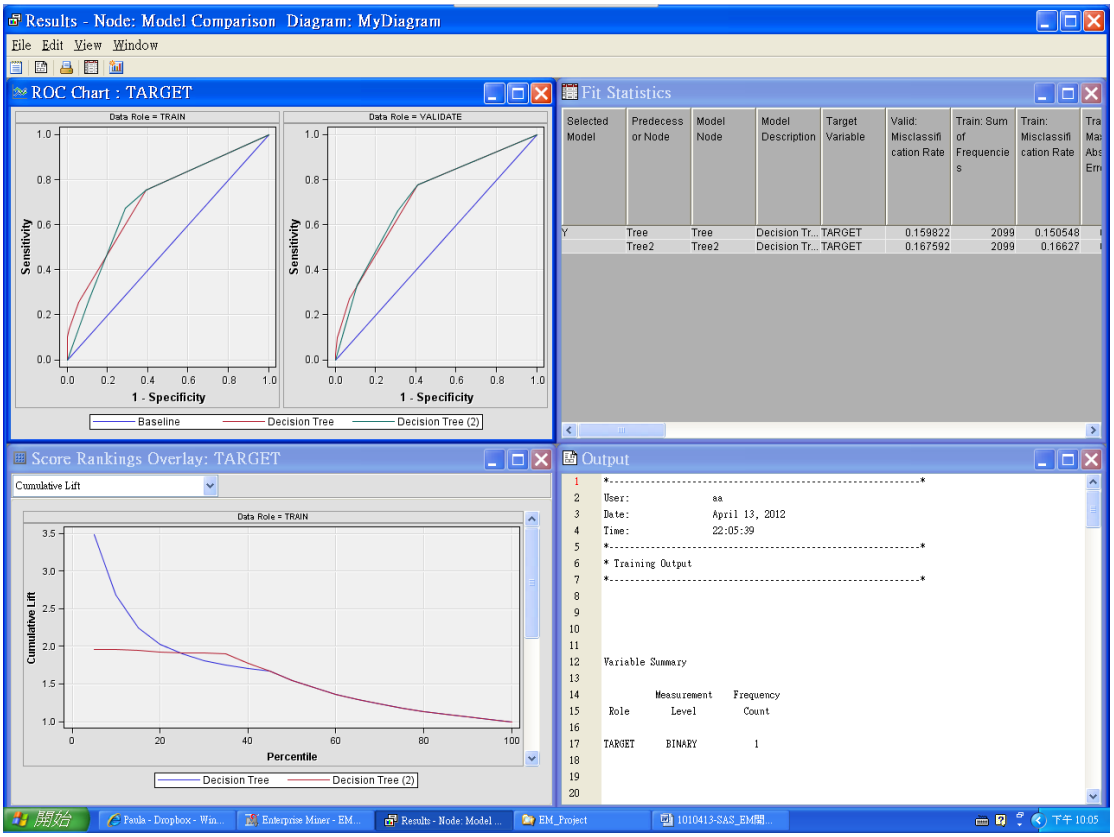
切三個：


Interactive Decision Tree - EMWS1.TREE2_BROWSETREE[EMWS1.TREE2_TRAINSAMPLE]
File Edit View Action Window
Tree View
Statistics Train Validation
0: 83.4% 83.2%
1: 16.6% 16.8%
Count: 2099 901
Number Trade Lines 60 Days or Worse 24 Mo...
<0.5 or Missing
Statistics Train Validation
0: 92.5% 92.9%
1: 7.5% 7.1%
Count: 1149 477
Split Node 1
Target Variable: TARGET
Variable
Variable Description
-Log(p)
Branches
TLDel60Cnt24
Number Trade Lines 60 D...
33.38918649
2
TLDel90Cnt24
Number Trade Lines 90+ ...
23.08701661
2
TLDel3060Cnt24
Number Trade Lines 30 o...
22.50781891
2
TLSatPct
Percent Satisfactory to Tot...
22.24959459
2
TLDel60CntAll
Number Trade Lines 60 D...
20.97874485
2
Edit Rule...
OK
Cancel
Apply
Refresh
Split Count
Node ID in Path
Split Variable
Value
0
1 (Selected)
All Cases or Missing
Statistics
Train
Validation
Count
2099
901
Prediction
0
0
% with target = 0
83.37303478%
83.24084351%
% with target = 1
16.62696522%
16.75915649%
% correctly predicted
83.37303478%
83.24084351%


TLDel60Cnt24 - Interval Split Rule
Target Variable: TARGET
Assign missing values to
A specific branch
1
A separate missing values branch
All branches
Branches
Branch
Split Point
1 < 1
2 < 3
3 >= 3
New split point:
Add Branch
Remove Branch
OK
Cancel
Apply
Reset


Interactive Decision Tree - EMWS1.TREE2_BROWSETREE[EMWS1.TREE2_TRAINSAMPLE]
File Edit View Action Window
Tree View
Statistics Train Validation
0: 83.4% 83.2%
1: 16.6% 16.8%
Count: 2099 901
Number Trade Lines 60 Days or Worse 24 Mo...
<1 or Missing
>=1 and <3
>=3
Statistics Train Validation
0: 92.5% 92.9%
1: 7.5% 7.1%
Count: 1149 477
Statistics Train Validation
0: 74.4% 77.1%
1: 25.6% 22.9%
Count: 664 292
Statistics Train Validation
0: 67.5% 62.1%
1: 32.5% 37.9%
Count: 286 132
Split Count
Node ID in Path
Split Variable
Value
0 1 Number Trade Lin... >=3
1 54 (Selected)
Statistics
Train
Validation
Count 286 132
Prediction 0 0
% with target = 0 67.48251748% 62.12121212%
% with target = 1 32.51748252% 37.87878788%
% correctly predicted 67.48251748% 62.12121212%


Enterprise Miner - EM_Project
File Edit View Actions Options Window Help
EM_Project
Data Sources
CREDIT
Diagrams
MyDiagram
Model Packages
Model Comparison
Compares models and predictions from the preceding modeling nodes.
CREDIT
Data Partition
Decision Tree
Decision Tree (2)
Model Comparison
Diagram
Log
Run completed


Results - Node: Model Comparison Diagram: MyDiagram
File Edit View Window
ROC Chart : TARGET
Data Role = TRAIN
Data Role = VALIDATE
Sensitivity
1 - Specificity
Baseline
Decision Tree
Decision Tree (2)
Fit Statistics
Selected Model
Predecessor Node
Model Node
Model Description
Target Variable
Valid: Misclassification Rate
Train: Sum of Frequencies
Train: Misclassification Rate
Score Rankings Overlay: TARGET
Cumulative Lift
Percentile
Output
User: aa
Date: April 13, 2012
Time: 22:05:39
* Training Output
Variable Summary
Role
Measurement Level
Frequency Count
TARGET BINARY 1

References:

http://mail.tku.edu.tw/myday/teaching.htm#1002BI