

# Home Credit Default Risk

The background of the slide features a dark blue grid. Overlaid on this grid are two light blue data visualizations. A line chart with circular markers at each data point spans the width of the slide, showing a fluctuating trend that generally rises from left to right. Below the line chart, a bar chart with numerous vertical bars of varying heights is visible, also spanning the width of the slide.

Team Member:

Chow Sai Yat Morris

# AGENDA

- ❏ Company Background
- ❏ Data Overview
- ❏ Objective 1: Customer segmentation
- ❏ Objective 2: Predict overdue chance
- ❏ Objective 3: Predict interest rate level

# COMPANY BACKGROUND

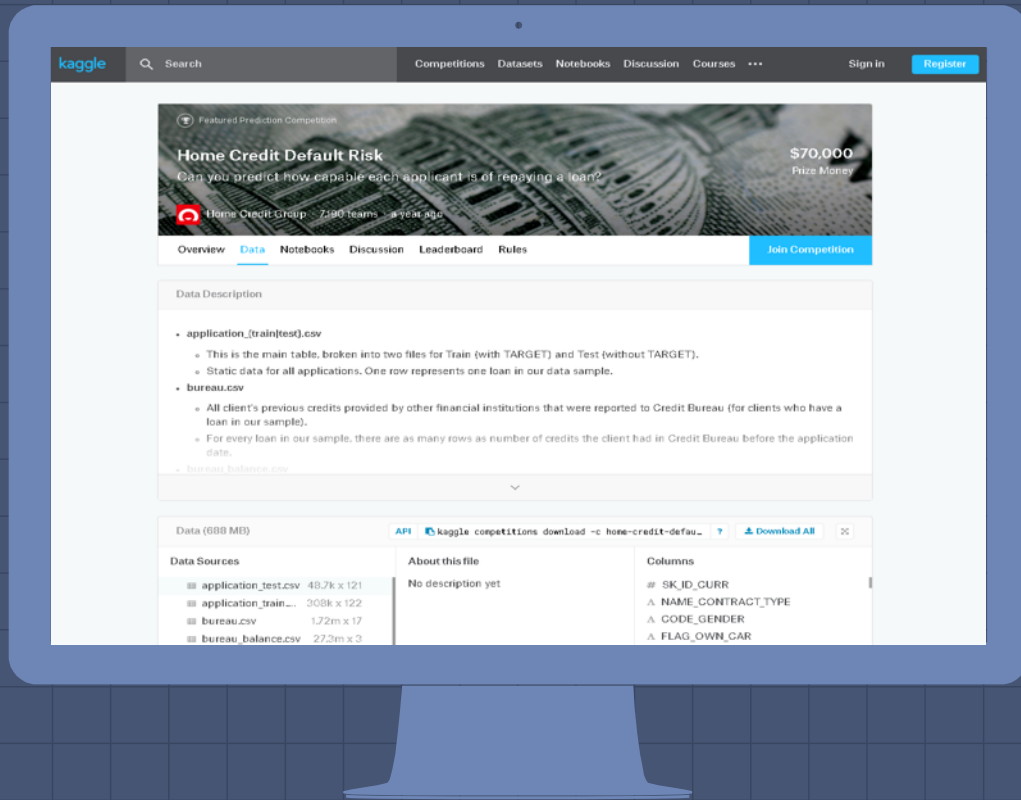
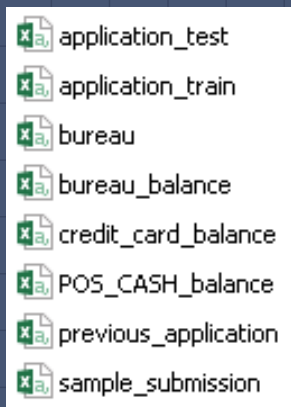
- Home Credit Group
  - Provide loan products to the unbanked population
  - Use data to predict client's repayment abilities to **minimize risk**

The logo for Home Credit, featuring the words "HOME" and "CREDIT" in a bold, red, sans-serif font, stacked vertically. The logo is set against a white rectangular background.

**HOME  
CREDIT**

# DATA OVERVIEW

- Data source
  - 9 datasets



# OBJECTIVE 1:

# Customer segmentation

Aims:

- Propose business strategies to different customer segments based on their characteristics

1

# METHODS

- Clustering

- Method: Average, Centroid, Ward
- Internal standardization: Standardization
- Ordinal: Index, Rank
- Nominal: GLM, Deviation

Biased data

- SOM

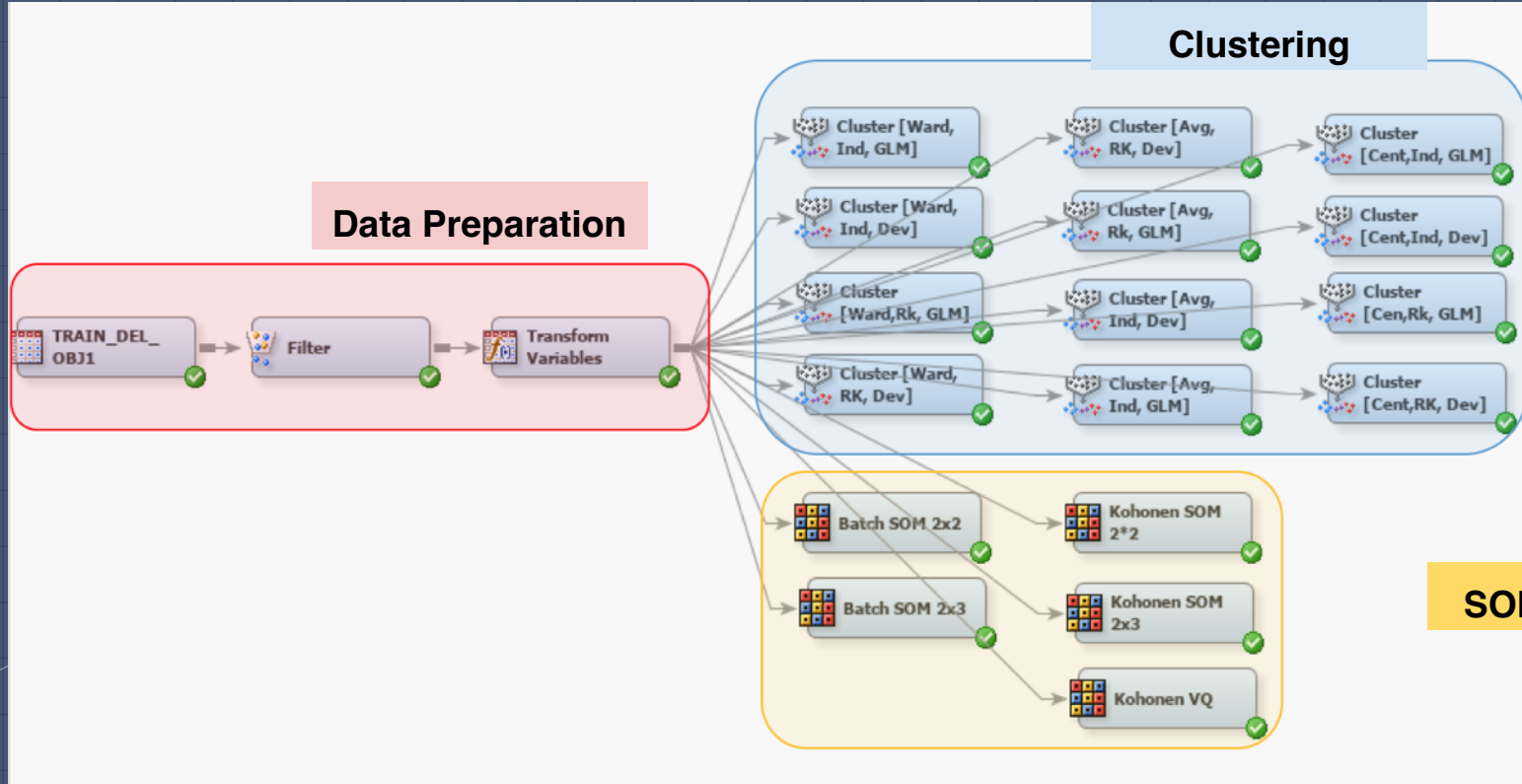
- Batch SOM, Kohonen SOM, Kohonen VQ
- 2x2, 2x3, 10

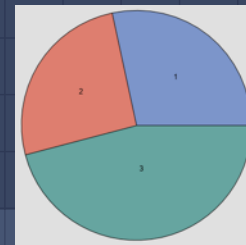
- Variables:

- ID
- Age
- Monthly income
- Monthly payment
- Credit score (FICO score)
- Own car
- Gender
- No. of children

**Selection criteria: CCC, PSF, RSQ, Within STD**

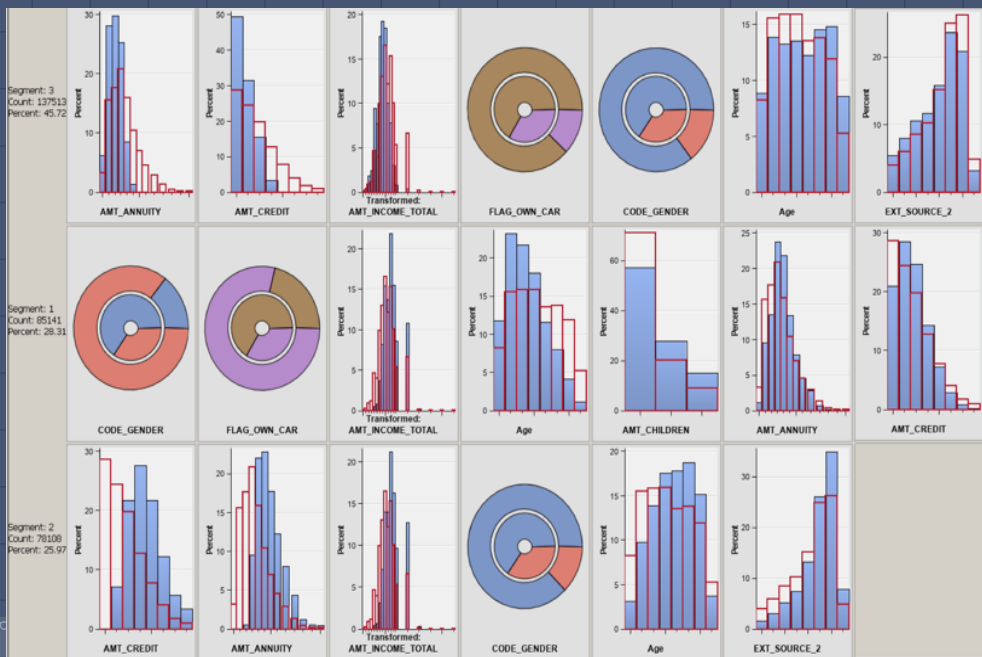
# EM DIAGRAM





# CLUSTER ANALYSIS RESULTS

- 12 clustering models
- Same result for 6 models
- Randomly select “**Average Standardization Index Deviation**” for later comparison



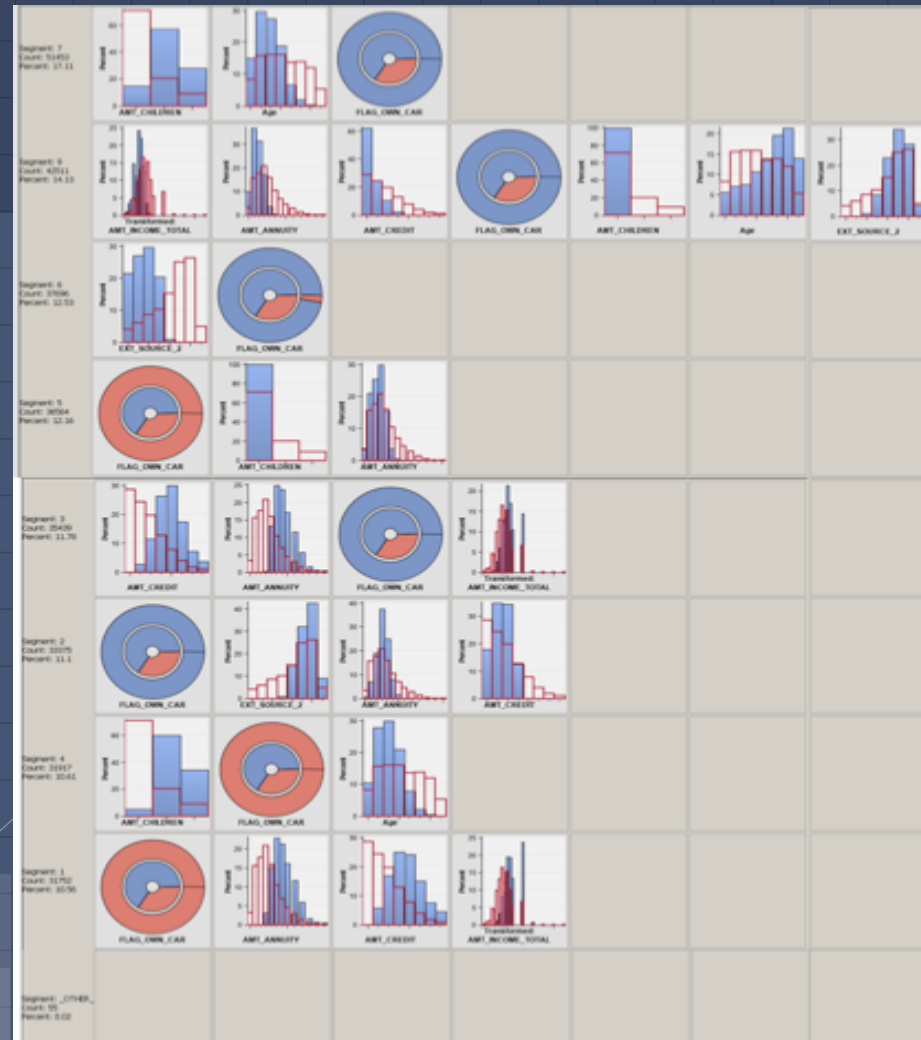
Cluster Method	Internal Standardization	Ordinal Encoding	Nominal Encoding	No. of Cluster (K)	CCC	PSF	RSQ	Within-STD
Average	Standardization	Index	Deviation	3	637.0984	57754.59	0.277487	0.803726
Average	Standardization	Rank	Deviation	3	637.0984	57754.59	0.277487	0.803726
Centroid	Standardization	Index	Deviation	3	637.0984	57754.59	0.277487	0.803726
Centroid	Standardization	Rank	Deviation	3	637.0984	57754.59	0.277487	0.803726
Ward	Standardization	Index	Deviation	3	637.0984	57754.59	0.277487	0.803726
Ward	Standardization	Rank	Deviation	3	637.0984	57754.59	0.277487	0.803726
Average	Standardization	Index	GLM	5	428.4803	42384.77	0.380494	0.685056
Average	Standardization	Rank	GLM	5	428.4803	42384.77	0.279012	0.685056
Centroid	Standardization	Index	GLM	5	428.4803	42384.77	0.380494	0.685056
Centroid	Standardization	Rank	GLM	5	428.4803	42384.77	0.380494	0.685056
Ward	Standardization	Index	GLM	5	428.4803	42384.77	0.380494	0.685056
Ward	Standardization	Rank	GLM	5	428.4803	42384.77	0.380494	0.685056



# SOM RESULTS

- Best model:  
Kohonen VQ with Standardization
- Not consider model with CCC < 3

SOM Method	Internal Standardization	No. of Cluster	CCC	PSF	RSQ	Within-STD
Kohonen VQ	Standardization	10	594.647	33748.22	0.502467	0.602957
Kohonen SOM	Standardization	2x3	480.6263	42157.81	0.412063	0.657167
Kohonen SOM	Standardization	2x2	397.3765	43821.32	0.304158	0.714542
Batch SOM	Standardization	2x2	-112.447	26061.84	0.206325	0.755468
Batch SOM	Standardization	2x3	-411.461	18229.4	0.232575	0.827179

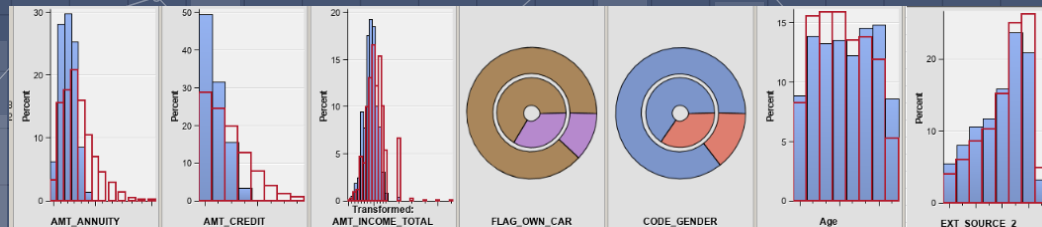


# INTERPRETATION & RECOMMENDATIONS

Best model  
chosen:

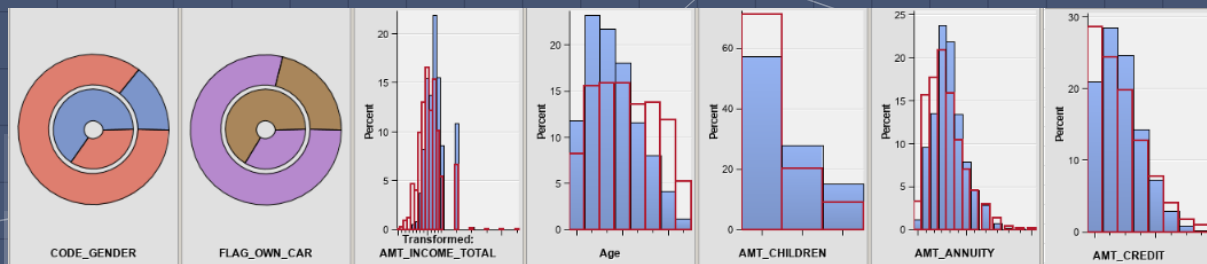
Method	Internal Standardization	No. of Cluster (K)	CCC	PSF	RSQ	Within-STD
Average	Standardization	3	637.0984	57754.59	0.277487	0.803726
Kohonen VQ	Standardization	10	594.647	33748.22	0.502467	0.602957

Segment	Characteristics	Recommendations
3	<ul style="list-style-type: none"> <li>- Female, Old (mean: 60)</li> <li>- Low monthly income</li> <li>- Low credit score</li> <li>- Car owner</li> </ul>	<ul style="list-style-type: none"> <li>- Poorer score: Higher interest rate</li> <li>- Car owner: Car loan offers</li> <li>- Retiree:               <ul style="list-style-type: none"> <li>a. reverse mortgage offer (if is property owner)</li> <li>b. limited potential for the business</li> </ul> </li> </ul>



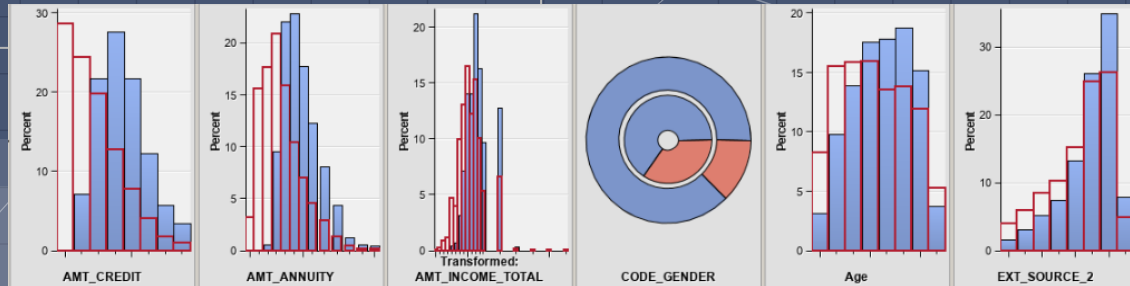
# INTERPRETATION & RECOMMENDATIONS

Segment	Characteristics	Recommendations
1	<ul style="list-style-type: none"> <li>- Relatively young male (mean: 36)</li> <li>- Have children</li> <li>- Relatively high monthly income</li> <li>- Do not have car</li> </ul>	<ul style="list-style-type: none"> <li>- Working class: tax loan offers</li> <li>- Family planning: housing loan offers (potential property buyers)</li> <li>- Insufficient financial/credit score data: Consultation required before loan approval</li> </ul>



# INTERPRETATION & RECOMMENDATIONS

Segment	Characteristics	Recommendations
2	<ul style="list-style-type: none"> <li>- Relatively old female (mean: 48)</li> <li>- High monthly income</li> <li>- High credit amount of loan</li> <li>- High credit score</li> </ul>	<ul style="list-style-type: none"> <li>- Working class: tax loan offers</li> <li>- Good score: lower interest rate</li> <li>- Large potential: Provide premium service to build long-term relationship (e.g. accelerate processing time)</li> </ul>



# OBJECTIVE 2:

## Predict whether overdue will occur

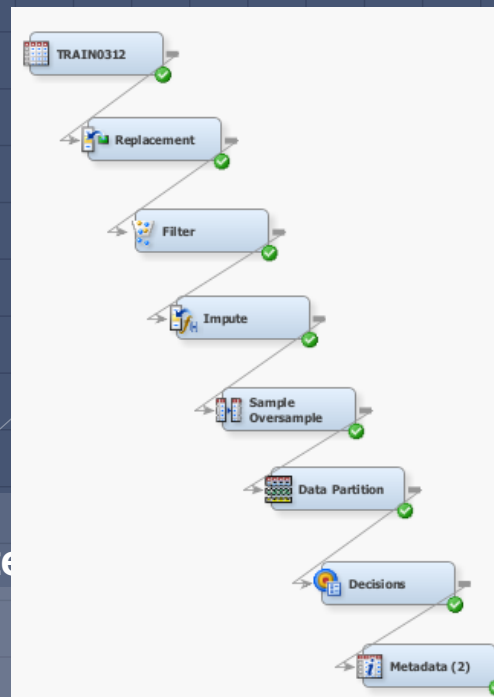
Aims:

- Allocate resources to customers/segments  
with lower default risk

2

# METHODS

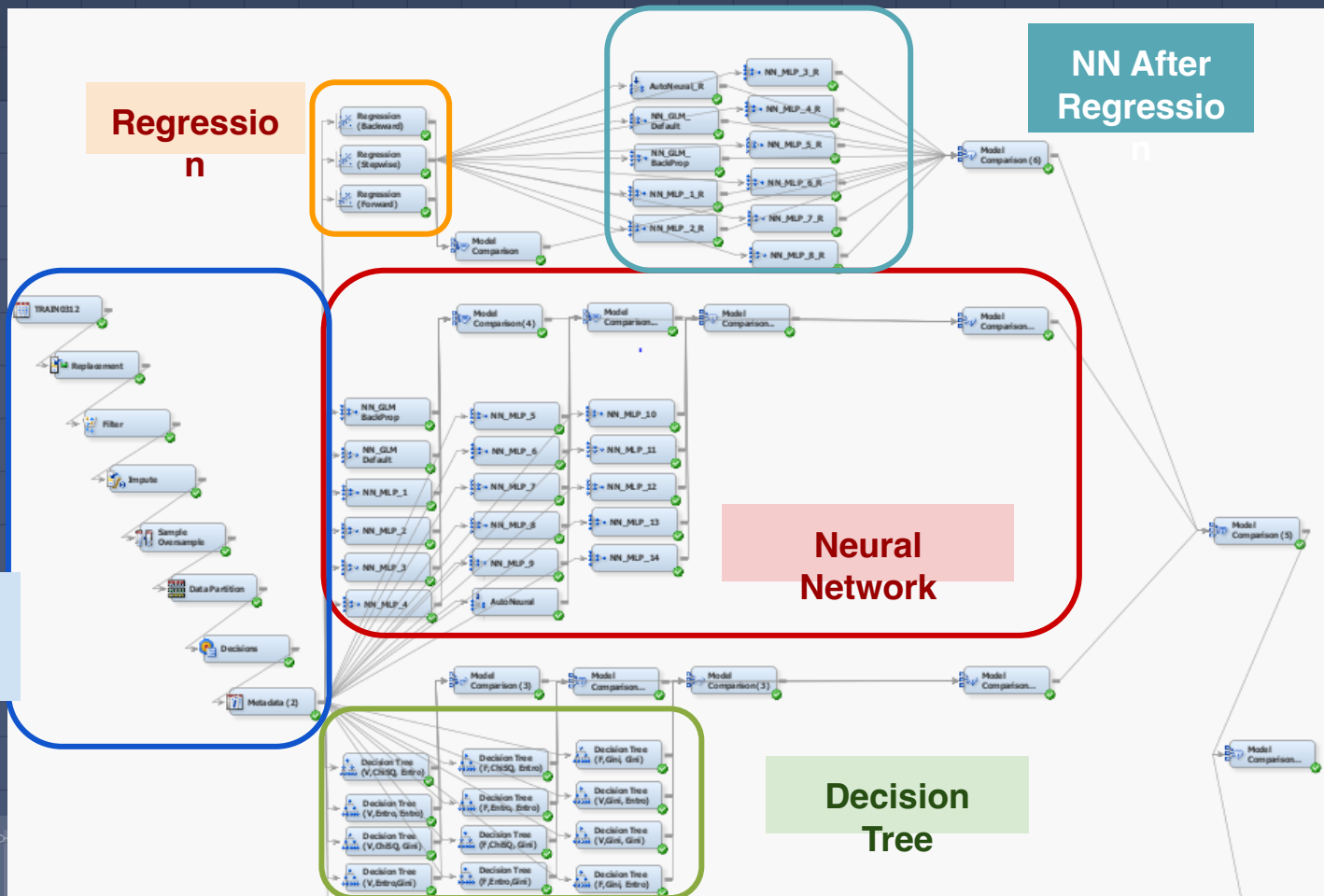
- Logistic regression: Stepwise, Backward, Forward
- Decision tree:
  - Interval: Variance, ProbF
  - Nominal: ProbChisq, Entropy, Gini
  - Ordinal: Entropy, Gini
- Neural network:
  - Network: GLM, MLP
  - AutoNeural
- Neural network after regression
- **Selection criterion: Misclassification rate**
- Target: Overdue (1) / No overdue (0)



Sample+Decision  
:  
Oversampling  
(biased target)

# EM DIAGRAM

15



# EM DIAGRAM

16

Transform  
variables

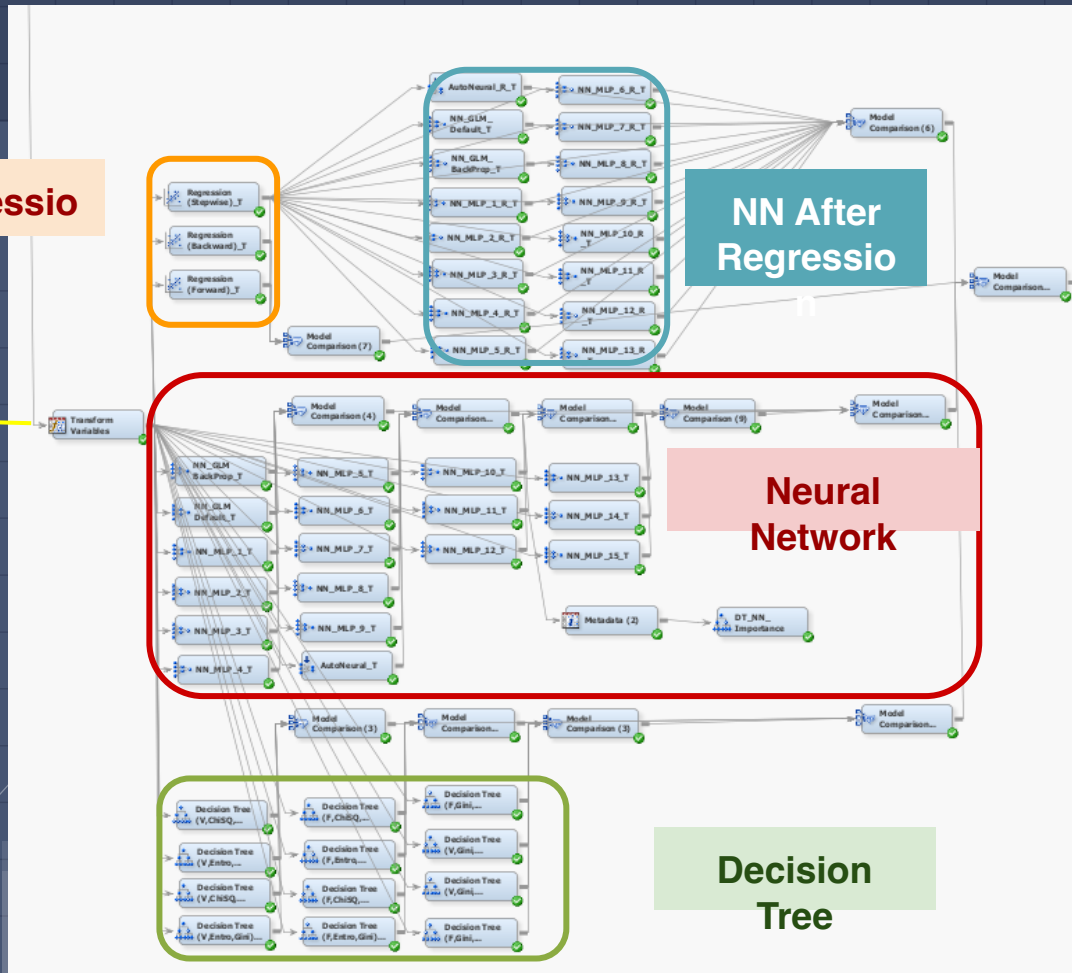
Regression

n

NN After  
Regression

Neural  
Network

Decision  
Tree





# LOGISTIC REGRESSION RESULTS

Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate
Without transformation			
Regression (Stepwise)	TARGET	TARGET	0.326017
Regression (Forward)	TARGET	TARGET	0.326017
Regression (Backward)	TARGET	TARGET	0.32642

Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate
With transformation			
Regression (Stepwise)_T	TARGET	TARGET	0.328534
Regression (Backward)_T	TARGET	TARGET	0.328534
Regression (Forward)_T	TARGET	TARGET	0.328534

- Variables without transformation give lower misclassification rate
- 3 logistic regression models (Stepwise, Backward, Forward)
- Same misclassification rate (0.326017) for **Stepwise & Forward regression**

# LOGISTIC REGRESSION RESULTS

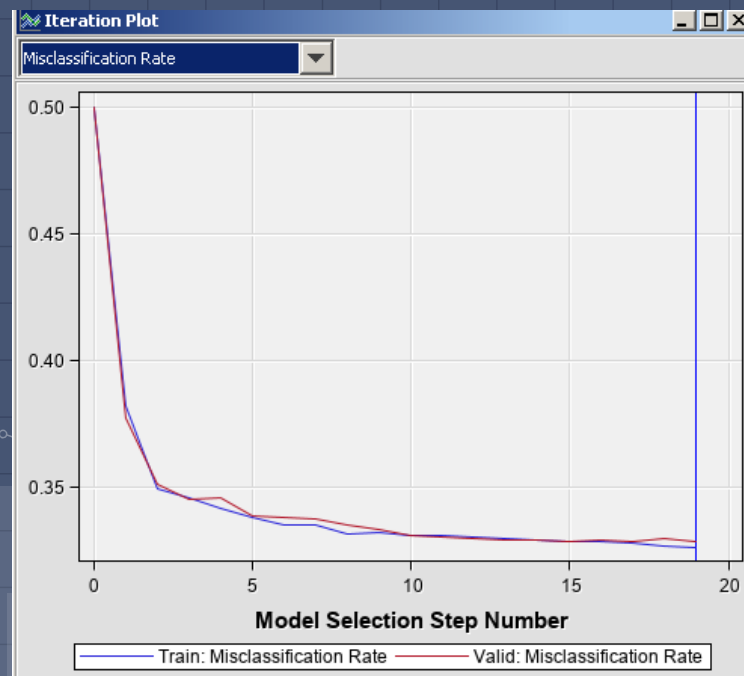
Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood Intercept Only	Likelihood Intercept & Covariates	Likelihood Ratio Chi-Square	DF	Pr > ChiSq
55060.839	48206.659	6854.1806	23	<.0001

Type 3 Analysis of Effects

Significant model

Effect	DF	Wald Chi-Square	Pr > ChiSq
CODE_GENDER	1	258.1709	<.0001
FLAG_OWN_CAR	1	117.1507	<.0001
IMP_EXT_SOURCE_2	1	1318.0357	<.0001
IMP_EXT_SOURCE_3	1	1999.5627	<.0001
IMP_REP_DAYS_EMPLOYED	1	226.1478	<.0001
IMP_payToincome	1	75.1134	<.0001
NAME_CONTRACT_TYPE	1	99.5106	<.0001
NAME_EDUCATION_TYPE	4	204.9952	<.0001
NAME_FAMILY_STATUS	4	56.2425	<.0001
NAME_INCOME_TYPE	6	173.9331	<.0001
REGION_RATING_CLIENT_W_CITY	2	70.2262	<.0001



# LOGISTIC REGRESSION RESULTS

Variables	Prob(overdue)
Gender	M > F
Own car	No car > own car
Contract type	Cash loan > Revolving loan
Education type	Lower secondary > Secondary > Incomplete higher > Higher education > Degree
Family status	Civil marriage > Separated > Single > Widow > Married
Income type	Maternity leave > Unemployed > Working > Commercial associate > State servant > Pensioner > Student
Region rating	2 > 3 > 1

Odds Ratio Estimates		
Effect		Point Estimate
CODE_GENDER	F vs M	0.674
FLAG_OWN_CAR	0 vs 1	1.316
IMP_EXT_SOURCE_2		0.123
IMP_EXT_SOURCE_3		0.059
IMP_REP_DAYS_EMPLOYED		1.000
IMP_payToincome		2.924
NAME_CONTRACT_TYPE	Cash loans vs Revolving loans	1.517
NAME_EDUCATION_TYPE	Academic degree vs Secondary / secondary special	0.214
NAME_EDUCATION_TYPE	Higher education vs Secondary / secondary special	0.677
NAME_EDUCATION_TYPE	Incomplete higher vs Secondary / secondary special	0.803
NAME_EDUCATION_TYPE	Lower secondary vs Secondary / secondary special	1.184
NAME_FAMILY_STATUS	Civil marriage vs Widow	1.179
NAME_FAMILY_STATUS	Married vs Widow	0.935
NAME_FAMILY_STATUS	Separated vs Widow	1.092
NAME_FAMILY_STATUS	Single / not married vs Widow	1.080
NAME_INCOME_TYPE	Commercial associate vs Working	0.874
NAME_INCOME_TYPE	Maternity leave vs Working	109.119
NAME_INCOME_TYPE	Pensioner vs Working	0.658
NAME_INCOME_TYPE	State servant vs Working	0.779
NAME_INCOME_TYPE	Student vs Working	0.008
NAME_INCOME_TYPE	Unemployed vs Working	5.265
REGION_RATING_CLIENT_W_CITY	1 vs 3	0.675
REGION_RATING_CLIENT_W_CITY	2 vs 3	0.833

# NEURAL NETWORK AFTER REGRESSION

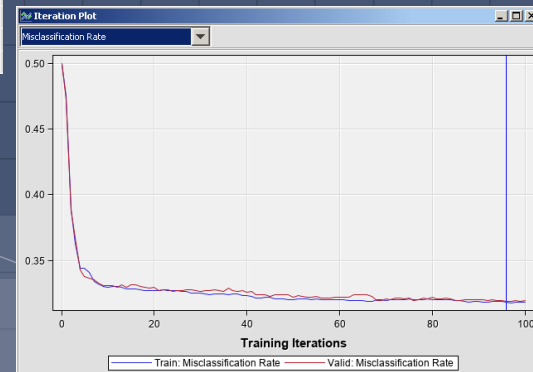
Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate
Without transformation			
NN_MLP_8_R	TARGET	TARGET	0.322795
NN_MLP_7_R	TARGET	TARGET	0.323399
NN_MLP_2_R	TARGET	TARGET	0.3236
NN_MLP_3_R	TARGET	TARGET	0.324104
NN_MLP_1_R	TARGET	TARGET	0.324205
NN_MLP_5_R	TARGET	TARGET	0.324507
AutoNeural_R	TARGET	TARGET	0.324507
NN_MLP_4_R	TARGET	TARGET	0.324607
NN_GLM_Default	TARGET	TARGET	0.325614
NN_MLP_6_R	TARGET	TARGET	0.326017
NN_GLM_BackProp	TARGET	TARGET	0.499899

Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate
With transformation			
NN_MLP_13_R_T	TARGET	TARGET	0.318566
NN_MLP_10_R_T	TARGET	TARGET	0.319875
NN_MLP_6_R_T	TARGET	TARGET	0.319875
NN_MLP_9_R_T	TARGET	TARGET	0.319976
NN_MLP_8_R_T	TARGET	TARGET	0.320177
NN_MLP_12_R_T	TARGET	TARGET	0.320681
NN_MLP_11_R_T	TARGET	TARGET	0.321788
NN_MLP_5_R_T	TARGET	TARGET	0.322191
NN_MLP_4_R_T	TARGET	TARGET	0.323802
NN_MLP_2_R_T	TARGET	TARGET	0.324909
NN_MLP_7_R_T	TARGET	TARGET	0.325614
NN_MLP_3_R_T	TARGET	TARGET	0.327024
AutoNeural_R_T	TARGET	TARGET	0.327628
NN_GLM_Default_T	TARGET	TARGET	0.328937
NN_MLP_1_R_T	TARGET	TARGET	0.330346
NN_GLM_BackProp_T	TARGET	TARGET	0.5

- No. of hidden unit = 8
- 11 NN models

→ NN after regression (transformed, 13 hidden units, MLP) obtains lowest Misclassification Rate (0.318566)

- No. of hidden unit = 13
- 16 NN models



# NEURAL NETWORK RESULTS

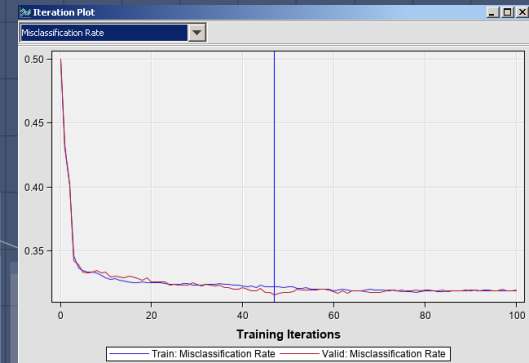
Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate
Without transformation			
NN_MLP_10	TARGET	TARGET	0.320681
NN_MLP_8	TARGET	TARGET	0.322292
NN_MLP_4	TARGET	TARGET	0.323701

- No. of hidden unit = 14
- 17 NN models

Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate
With transformation			
NN_MLP_11_T	TARGET	TARGET	0.316049
NN_MLP_7_T	TARGET	TARGET	0.317358
NN_MLP_13_T	TARGET	TARGET	0.321184
NN_MLP_2_T	TARGET	TARGET	0.324003

- No. of hidden unit = 15
- 18 NN models

→ NN (transformed, 11 hidden units, MLP)  
obtains lowest Misclassification Rate (0.316049)

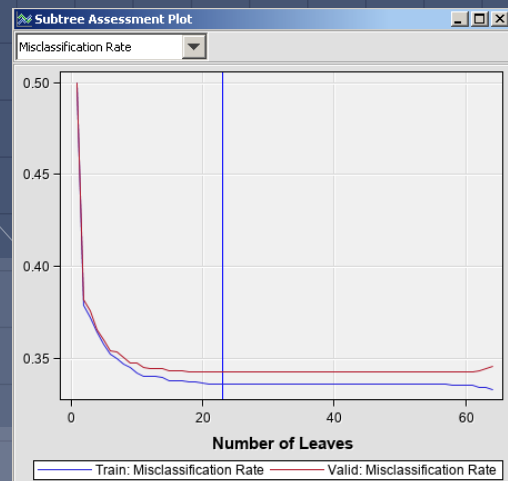


# DECISION TREE RESULTS

Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate
Without transformation			
Decision Tree (V,Gini, Entro)	TARGET	TARGET	0.342328
Decision Tree (V,Entro, Entro)	TARGET	TARGET	0.346355
Decision Tree (F,Entro, Entro)	TARGET	TARGET	0.346355

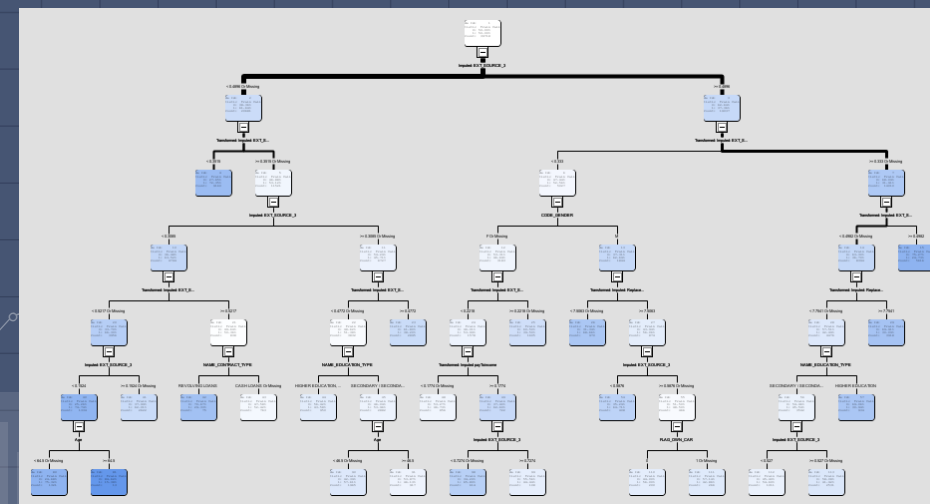
Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate
With transformation			
Decision Tree (V,Gini, Entro)_T_/M	TARGET	TARGET	0.342328
Decision Tree (V,Entro, Entro)_T_/M	TARGET	TARGET	0.346355
Decision Tree (F,Entro, Entro)_T_/M	TARGET	TARGET	0.346355

- 12 Decision Tree models respectively
- With & without transformed variables give the **same** misclassification rate (0.342328)



# DECISION TREE RESULTS

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
IMP_EXT_SOURCE_3	Imputed: EX...	6	1.0000	1.0000	1.0000
LOG_IMP_EXT_SOURCE_2	Transforme...	6	0.8035	0.7986	0.9940
LOG_IMP_REP_DAYS_EMPLOYED	Transforme...	2	0.2358	0.1889	0.8009
CODE_GENDER	CODE_GEN...	1	0.2054	0.2226	1.0838
NAME_EDUCATION_TYPE	NAME_EDU...	2	0.1819	0.2018	1.1096
Age	Age	2	0.1347	0.1250	0.9278
LOG_IMP_payToincome	Transforme...	1	0.1180	0.0599	0.5078
NAME_CONTRACT_TYPE	NAME_CON...	1	0.0717	0.0337	0.4697
FLAG_OWN_CAR	FLAG_OWN...	1	0.0527	0.0470	0.8917

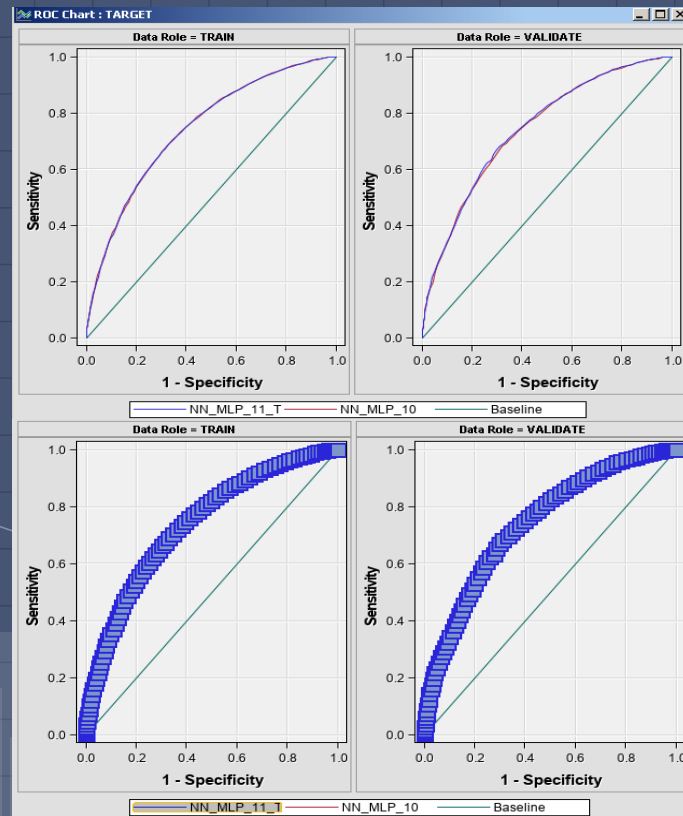


# OVERALL MODEL COMPARISON

Final Model comparison

Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Misclassification Rate
NN_MLP_11_T	TARGET	TARGET	0.316049	0.322297
NN_MLP_10	TARGET	TARGET	0.320681	0.32129

- Best model: NN (transformed, 11 hidden units, MLP) obtains lowest misclassification rate (0.316049)





# INTERPRETATION & RECOMMENDATIONS

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance
IMP_EXT_SOURCE_3	Imputed: EXT_SOURCE_3	8	1.0000	1.0000
LOG_IMP_EXT_SOURCE_2	Transformed: Imputed: EXT_SOURCE_2	5	0.7372	0.7082
NAME_EDUCATION_TYPE	NAME_EDUCATION_TYPE	3	0.2847	0.2681
CODE_GENDER	CODE_GENDER	3	0.2816	0.2652
LOG_IMP_REP_DAYS_EMPLOYED	Transformed: Imputed: Replacement: DAYS_EMPLOYED	2	0.1848	0.2026
Age	Age	2	0.1638	0.1872
FLAG_OWN_CAR	FLAG_OWN_CAR	2	0.1532	0.1286
TI_REGION_RATING_CLIENT_W_C1	REGION_RATING_CLIENT_W_CITY:1	1	0.0928	0.1110
LOG_IMP_payToincome	Transformed: Imputed payToincome	1	0.0863	0.0863
TI_REGION_RATING_CLIENT_W_C3	REGION_RATING_CLIENT_W_CITY:3	1	0.0859	0.0851
NAME_CONTRACT_TYPE	NAME_CONTRACT_TYPE	1	0.0498	0.0461

Variable Importance:  
 EXT\_SOURCE\_3(TU  
 Score)  
 EXT\_SOURCE\_2(FICO  
 Score)  
 EDUCATION TYPE  
 GENDER  
 DAYS EMPLOYED

Variables	Prob(overdue)
Gender	M > F
Contract type	Cash loan > Revolving loan
Education type	Lower secondary > Secondary > Incomplete higher > Higher education > Degree

- The regression model might provide some insight.

- Tighten cash loan policies to reduce overdue probability
- Allocate more resources in revolving loans

# OBJECTIVE 3:

## Predict interest rate for new customer

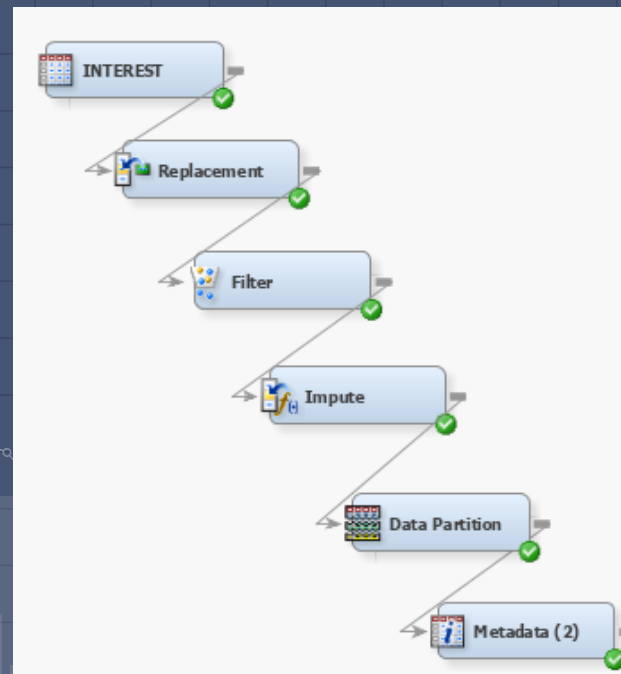
Aims:

- Maximize profit

3

# METHODS

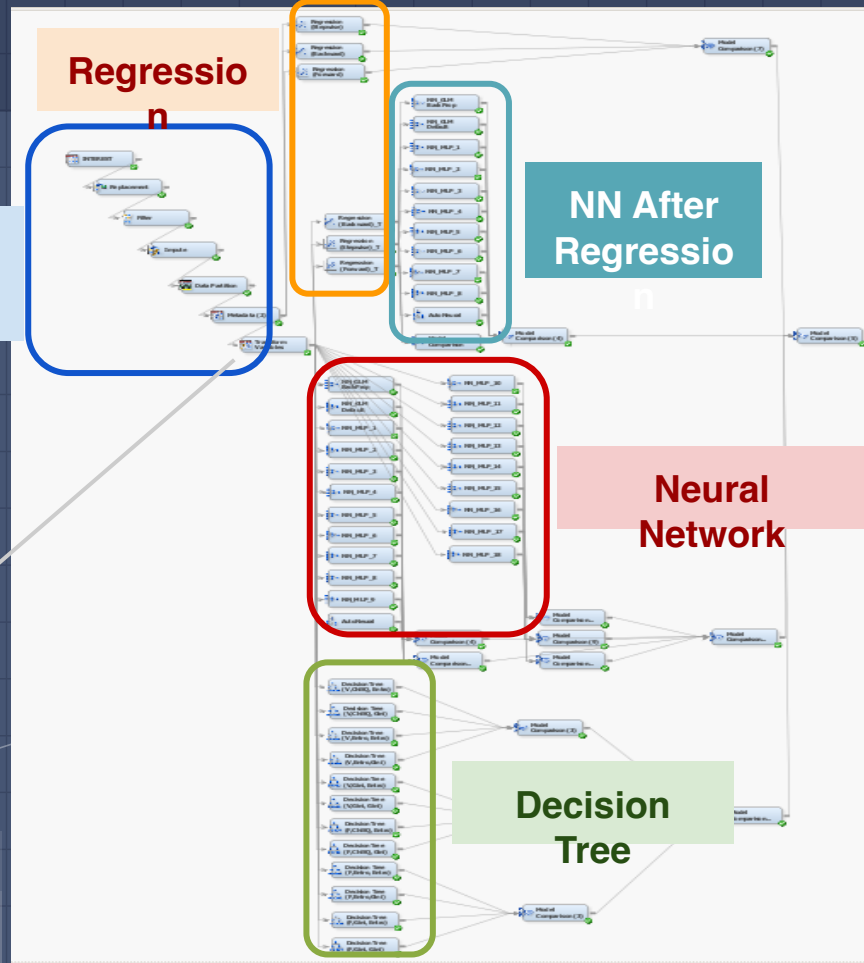
- Linear regression: Stepwise, Backward, Forward
- Target: Interest rate
- Decision tree:
  - Interval: Variance, ProbF
  - Nominal: ProbChisq, Entropy, Gini
  - Ordinal: Entropy, Gini
- Neural network:
  - Network: GLM, MLP
  - AutoNeural
- Neural network after regression
- Selection criterion: Average squared error



# EM DIAGRAM

Transform  
variables

Data  
preparation

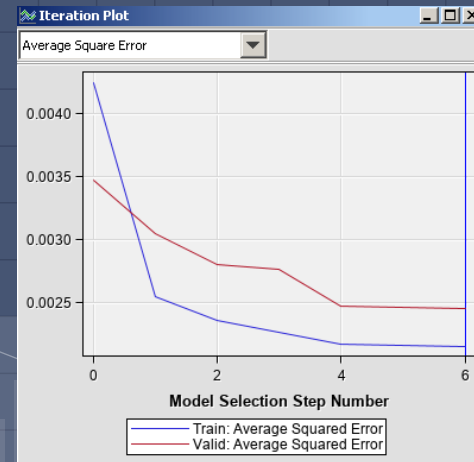


# LINEAR REGRESSION RESULTS

Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error
Without transformation			
Regression (Backward)	RATE_INTE...		0.005952
Regression (Stepwise)	RATE_INTE...		0.00603
Regression (Forward)	RATE_INTE...		0.00603

Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error
With transformation			
Regression (Stepwise)_T	LOG_RATE...	Transforme...	0.002453
Regression (Forward)_T	LOG_RATE...	Transforme...	0.002453
Regression (Backward)_T	LOG_RATE...	Transforme...	0.002745

- 3 linear regression models with and without transformation respectively (Stepwise, Backward, Forward)
- Regression with transformed variables give smaller ASE
- **Stepwise & Backward (transformed variables) obtain same ASE (0.002453)**



# LINEAR REGRESSION RESULTS

30

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	7.038345	0.306015	140.94	<.0001
Error	3337	7.245222	0.002171		
Corrected Total	3360	14.283568			

Significant model

Model Fit Statistics

R-Square	0.4928	Adj R-Sq	0.4893
AIC	-20587.3689	BIC	-20590.0865
SBC	-20440.4890	C(p)	412.6875

Type 3 Analysis of Effects

Effect	DF	Sum of Squares	F Value	Pr > F
LOG_AMT_APPLICATION	1	0.0221	10.17	0.0014
LOG_RATE_INTEREST_PRIVILEGED	1	0.5737	264.26	<.0001
M_REP_DAYS_FIRST_DUE	1	0.0124	5.71	0.0170
M_REP_DAYS_LAST_DUE_1ST_VERSION	1	0.0828	38.13	<.0001
NAME_GOODS_CATEGORY	16	0.3461	9.96	<.0001
PRODUCT_COMBINATION	3	1.0314	158.35	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr >  t
Intercept	1	0.2057	0.0200	10.28	<.0001
LOG_AMT_APPLICATION	1	-0.00452	0.00142	-3.19	0.0014
LOG_RATE_INTEREST_PRIVILEGED	1	0.2592	0.0159	16.26	<.0001
M_REP_DAYS_FIRST_DUE	1	0.00457	0.00191	2.39	0.0170
M_REP_DAYS_LAST_DUE_1ST_VERSION	1	0.0160	0.00260	6.18	<.0001
NAME_GOODS_CATEGORY	1	0.0844	0.0189	4.47	<.0001
NAME_GOODS_CATEGORY	1	-0.0175	0.0239	-0.73	0.4638
NAME_GOODS_CATEGORY	1	0.1174	0.0165	7.11	<.0001
NAME_GOODS_CATEGORY	1	0.00514	0.0113	0.45	0.6497
NAME_GOODS_CATEGORY	1	0.0392	0.0154	2.55	0.0108
NAME_GOODS_CATEGORY	1	0.0399	0.0187	2.14	0.0327
NAME_GOODS_CATEGORY	1	0.0514	0.0135	3.80	0.0001
NAME_GOODS_CATEGORY	1	0.1867	0.0335	5.57	<.0001
NAME_GOODS_CATEGORY	1	-0.0527	0.0176	-3.00	0.0027
NAME_GOODS_CATEGORY	1	-0.00789	0.0451	-0.17	0.8613
NAME_GOODS_CATEGORY	1	-0.1884	0.0497	-3.79	0.0002
NAME_GOODS_CATEGORY	1	-0.1681	0.0457	-3.67	0.0002
NAME_GOODS_CATEGORY	1	-0.00468	0.0111	-0.42	0.6720
NAME_GOODS_CATEGORY	1	-0.00131	0.0127	-0.10	0.9181
NAME_GOODS_CATEGORY	1	-0.00897	0.0269	-0.33	0.7387
NAME_GOODS_CATEGORY	1	-0.0268	0.0287	-0.93	0.3511
PRODUCT_COMBINATION	1	0.0779	0.0114	6.86	<.0001
PRODUCT_COMBINATION	1	0.0473	0.0136	3.48	0.0005
PRODUCT_COMBINATION	1	-0.1560	0.00974	-16.03	<.0001

# NEURAL NETWORK AFTER REGRESSION

Model Description	Target	Target Label	Selection Criterion: Valid: Average Squared Error
With transformation			
NN_MLP_6	LOG_RATE...	Transforme...	0.002286
NN_MLP_7	LOG_RATE...	Transforme...	0.002328
NN_MLP_5	LOG_RATE...	Transforme...	0.002348
NN_MLP_8	LOG_RATE...	Transforme...	0.002368
NN_MLP_4	LOG_RATE...	Transforme...	0.002409
NN_MLP_3	LOG_RATE...	Transforme...	0.002416
NN_GLM Default	LOG_RATE...	Transforme...	0.002491
AutoNeural	LOG_RATE...	Transforme...	0.002713
NN_MLP_2	LOG_RATE...	Transforme...	0.00272
NN_MLP_1	LOG_RATE...	Transforme...	0.002742
NN_GLM BackProp	LOG_RATE...	Transforme...	0.00344

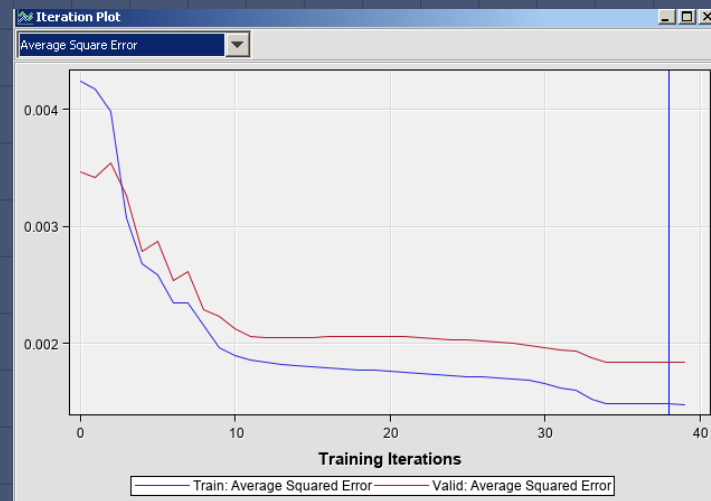
- No. of hidden unit = 8
- 12 NN after regression models
- **NN after regression (6 hidden units, MLP)** obtains smallest ASE (0.002286)



# NEURAL NETWORK RESULTS

Model Description	Target	Target Label	Selection Criterion: Valid: Average Squared Error
With transformation			
NN_MLP_2	LOG_RATE...	Transforme...	0.001838
NN_MLP_9	LOG_RATE...	Transforme...	0.002021
NN_MLP_17	LOG_RATE...	Transforme...	0.002066
NN_MLP_10	LOG_RATE...	Transforme...	0.002113
NN_MLP_16	LOG_RATE...	Transforme...	0.002569

- No. of hidden unit = 18
- 21 NN models
- **NN (2 hidden units, MLP) obtains smallest ASE (0.001838)**

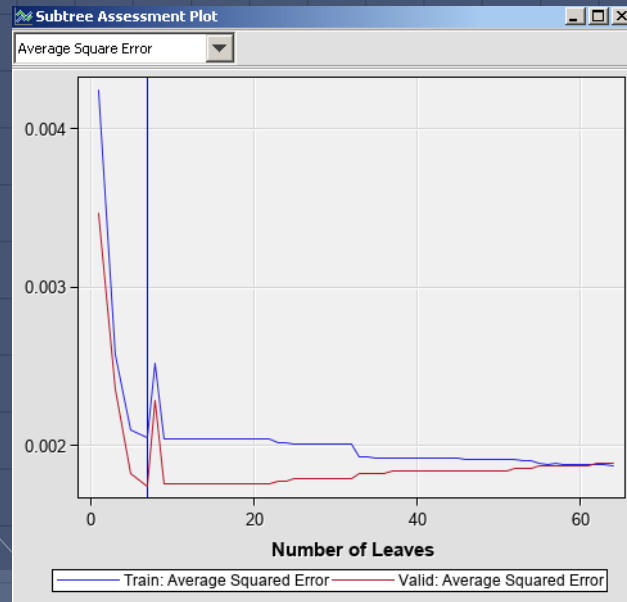




# REGRESSION TREE RESULTS

Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error
With transformation			
Decision Tree (V,ChiSQ, Entro)	LOG_RATE...	Transforme...	0.001745
Decision Tree (V,Gini, Entro)	LOG_RATE...	Transforme...	0.002498
Decision Tree (F,Entro, Entro)	LOG_RATE...	Transforme...	0.002498

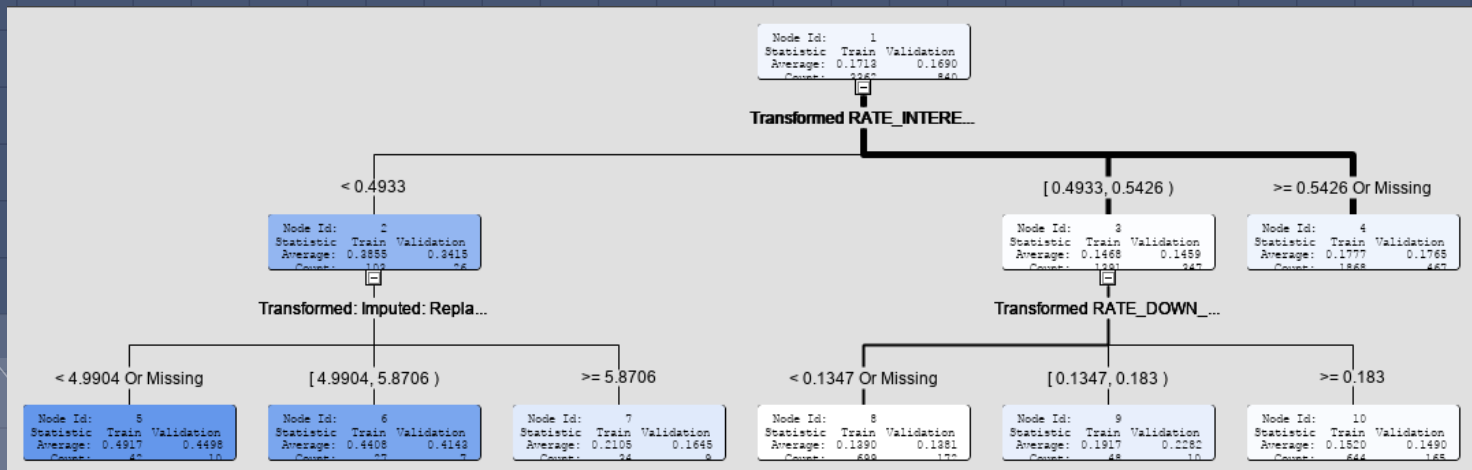
- 12 Regression Tree models
- **Decision Tree (Variance, ProbChisq, Entropy)** model gives smallest ASE (0.001745)



# OVERALL MODEL COMPARISON & RECOMMENDATIONS

Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error	Train: Average Squared Error
With transformation				
Decision Tree (V,ChiSQ, Entro)	LOG_RATE...	Transforme...	0.001745	0.002051
NN_MLP_2	LOG_RATE...	Transforme...	0.001838	0.001482
NN_MLP_6	LOG_RATE...	Transforme...	0.002286	0.001941

- Best model: **Regression Tree (Variance, ProbChisq, Entropy)** model give smallest ASE (0.001745)



# OVERALL MODEL COMPARISON & RECOMMENDATIONS

Important factors

Variable Name	Label	Number of Splitting Rules	Number of Surrogate Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
LOG_DAYS_DECISION	Transformed DAYS_DECISI...	0	3	1.0000	1.0000	1.0000
LOG_IMP_REP_DAYS_FIRST_DUE	Transformed: Imputed: Repl...	0	3	0.9991	0.9992	1.0001
LOG_IMP_REP_DAYS_TERMINATION	Transformed: Imputed: Repl...	0	3	0.9690	0.9646	0.9955
LOG_RATE_INTEREST_PRIVILEGED	Transformed RATE_INTERE...	1	1	0.9685	0.9145	0.9442
LOG_RATE_DOWN_PAYMENT	Transformed RATE_DOWN_...	1	1	0.7789	0.7437	0.9549
NAME_CLIENT_TYPE		0	1	0.7556	0.6998	0.9262
LOG_IMP_REP_DAYS_LAST_DUE_1ST_VE	Transformed: Imputed: Repl...	1	1	0.5264	0.6412	1.2181
LOG_PAYTOINC	Transformed PAYTOINC	0	1	0.3579	0.4291	1.1988
LOG_SELLERPLACE_AREA	Transformed SELLERPLAC...	0	1	0.3509	0.4206	1.1988

## Recommendations:

- Apart from demographic info
- Use the above variables as they impose significant impact
- Maximize profit with predicted interest rate

e.g. Days decision, previous loan due days, Pay-to-income ratio

# THANKS!

## Q&A

