Executive Summary

Objective:

The objective is to maximize profit for Trojan financial services company by minimizing the amount of loan defaults for its Home Equity line of credit service. The company interests to know customer's likelihood to default by studying the variables that pertain to the individuals. The company hope to find the best model to predict the best loan amount and prevent future default. The objective is to identify the high risk default line of credit.

JMP Model:

- Trojan financial services created the JMP best logistic regression model based on statistical significance of the variables
- Model: Log(BAD)= 4.447-0.8201(Derog)-0.6778351(Delinq)+0.0077(Clage)-0.0813(Debtinc)
- 4 independent variables:

1. "Derog": **Number of major derogatory reports**.
2. "Delinq": **Number of delinquent credit lines**.
3. "Clage": **Age of oldest credit line in months**.
4. "Debtinc": **Debt-to-income ratio**.

- This model explained 21.78% of variation in the dependent variable.
- 4 independent variables are all significantly significant at alpha level of 0.05.
- The projected revenue calculated using the algorithm generated from JMP is $24,520,000, which exceeds the company's expected minimum profit of $20,000,000.

Key Insights: The model was built based on statistical significance of the variable

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	4.44718476	0.682225	42.49	<.0001*
DEROG	-0.8201803	0.1613065	25.85	<.0001*
DELINQ	-0.6778351	0.1236336	30.06	<.0001*
CLAGE	0.00770365	0.0019194	16.11	<.0001*
DEBTINC	-0.0813458	0.0162045	25.20	<.0001*

For log odds of 0/1

Your Best Model:

- Trojan financial services created the JMP best logistic regression model based on statistical significance of the variables
  Model:
- **"VALUE" : Value of current property**
- "Derog": **Number of major derogatory reports**.

- “Delinq”: **Number of delinquent credit lines**.
- “Clage”: **Age of oldest credit line in months**.
- “Debtinc”: **Debt-to-income ratio**.
- Apart from this we have
- This model explained 30.35% of variation in the dependent variable.
- All the independent variables are all significantly significant at alpha level of 0.05.
- The projected revenue calculated using the algorithm generated from JMP is $26,560,000, which exceeds the JMP’s expected profit of $24,800,000.

Key Changes Made: We considered new independent variables include the “Value” which indicates the value of current property, “CLNO”, which indicates the number of trade lines, and “Loan”, which indicates the loan amount.

Key Insights: A more detailed valuation of the credit line seem to offer higher chances to correctly identify and decrease the amount of loan defaults. The JMP only included four impendent variables which is not enough in this case.

Why your model is better?

The new model is better with a higher Rsquare value. The parameters have the least **Prob>ChiSq** score which makes our model a better fit when compared to the original model. Also the overall model also has a **Prob>ChiSq score which is less then 0.01 which confirms that the model we have built is optimal.** It also generates more profit than the JMP model.

What is the lift (as a ratio) provided by your model compared to Baseline Model for both training and testing? What is the increase in net dollar amount compared to the Baseline Model for both training and testing?

The lift has been considerably high. With the baseline model having a score of $24,520,000and our model has $26,560,000, we have a lift of 1:1.0832 for the testing and for training we have a lift 1:1.01575. The net dollar amount has increased by $2,400,000 for testing and $80,000 in training.

JMP Logistic Model

Build the Logistic Model using JMP (Go option) on the following conditions,
Y = BAD
X = All predictors
Cutoff Probability for mailing = 0.14

i) Statistical KPIs of JMP Model – From JMP Printout

Measure	Training	Validation	Definition
Entropy RSquare	0.3137	0.1889	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.3845	0.2379	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.2185	0.2454	∑ -Log(ρ[j])/n
RMSE	0.2372	0.2538	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.1135	0.1232	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0640	0.0740	∑ (ρ[j]≠ρMax)/n
N	1000	1000	n

Statistical KPIs of JMP Model – From Excel Printout

	Training	Validation
Accuracy %	**90.80%**	**90.30%**
True Positive Rate	**54.64%**	**44.44%**
False Positive Rate	**5.32%**	**5.16%**
Sensitivity (True Positive Rate)	**54.64%**	**44.44%**
Specificity (True Negative Rate)	**94.96%**	**94.84%**

ii) a) Business KPIs of JMP Model – Training

Predicted number of Good Loans	=	8990
Upper limit for Loans	=	10000
Actual number of approved loans	=	8990

Propensity of Good Loan	=	95.106%
Propensity of Bad Loan	=	4.894%

Total Profit	=	$ 25,400,000

b) Business KPIs of JMP Model – Testing

Predicted number of Good Loans	=	9130
Upper limit for Loans	=	10000
Actual number of approved loans	=	9130

Propensity of Good Loan	=	94.524%
Propensity of Bad Loan	=	5.476%

Total Profit	=	$ 24,520,000

ii) Interpret the Model (decision tree) – From Business Point of view & Statistical Point of view

- **"Derog": Number of major derogatory reports. A higher number usually indicates a higher probably of default based on past history;**
- **"Delinq": Number of delinquent credit lines. A higher number usually indicates a higher probably of default based on past history;**
- **"Clage": Age of oldest credit line in months. A higher number usually indicates a good credit, which leads to a lower probably of default based on past history**
- **"Debtinc": Debt-to-income ratio. A higher number usually indicates a higher probably of default based on the individual's ability to repay the debt.**

iv) Confusion Matrix for Training

	GoodLoan	BadLoan	
GoodLoan	855	48	903
BadLoan	44	53	97
	899	101	1000

iv) Confusion Matrix for Testing

	GoodLoan	BadLoan	
GoodLoan	863	47	910
BadLoan	50	40	90
	913	87	1000

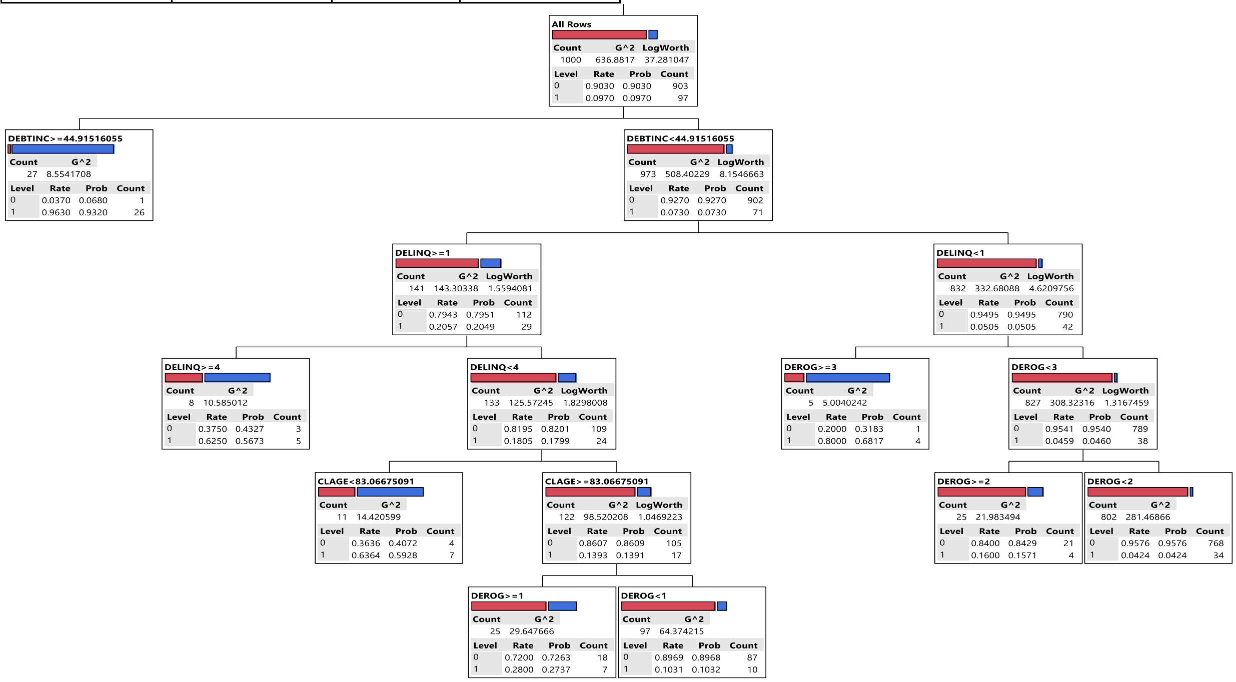
v) Lift Table **(copy & paste from Excel)**

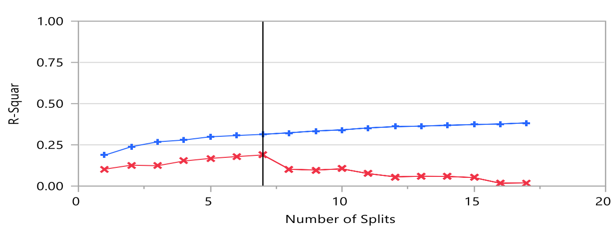
Lift Table in Dollars	Training	Testing
Lift with respect to Baseline - JMP Model	13.84591391	10.18091079
Lift with respect to Baseline - My Best Model	12.01809108	11.519243
Lift with respect to JMP Model - My Contribution	0.867988286	0.831959745
Overall Lift with respect to Baseline -My Best Model	12.01809108	11.519243

Lift Table in Propensity	Training	Testing
Lift with respect to Baseline - JMP Model	5.409819332	4.739898092
Lift with respect to Baseline - My Best Model	2.849298037	2.755945698

vi) Attach JMP Printout

	RSquare	N	Number of Splits
Training	0.314	1000	7
Validation	0.189	1000	





My Best Logistic Model

Build the Logistic Model using JMP (Go option) on the following conditions,

Y = BAD

X = VALUE, REASON[DebtCon], DEROG, DELINQ, CLAGE, CLNO, DEBTINC, Log (Loan) , Log (Value) , Log (Value)

Cutoff Probability for mailing = 0.14

Note: It may not be possible to obtain some values for Validation data in that case ignore it.

iii) Statistical KPIs of Best Logistic Model – From JMP Printout

Measure	Training	Validation	Definition
Entropy RSquare	0.3035	**0.2744**	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.3731	**0.3369**	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.2218	**0.2195**	∑ -Log(ρ[j])/n
RMSE	0.2443	**0.2416**	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.1203	**0.1153**	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0720	**0.0690**	∑ (ρ[j]≠ρMax)/n
N	1000	**1000**	n

Statistical KPIs of the Best Logistic Model – From Excel Printout

	Training	Validation
Accuracy %	**87.8%**	92.90%
True Positive Rate	**65.56%**	84.21%
False Positive Rate	**10.00%**	6.76%
Sensitivity (True Positive Rate)	**65.56%**	84.21%
Specificity (True Negative Rate)	**90.00%**	93.24%

ii) a) Business KPIs of the Best Logistic Model – Training (copy & paste from Excel)

Predicted number of Good Loans	=	8610
Upper limit for Loans	=	10000
Actual number of approved loans	=	8610

Propensity of Good Loan	=	95.819%

Propensity of Bad Loan	=	4.181%

Total Profit	=	$ 25,800,000

b) Business KPIs of the Best Logistic Model – Testing (copy & paste from Excel)

Predicted number of Good Loans	=	8500
Upper limit for Loans	=	10000
Actual number of approved loans	=	8500

Propensity of Good Loan	=	96.353%
Propensity of Bad Loan	=	3.647%

Total Profit	=	$ 26,560,000

iii) Interpret the Model (decision tree) – From Business Point of view & Statistical Point of view

- **Derog": Number of major derogatory reports. A higher number usually indicates a higher probably of default based on past history;**
- **"Loan" : Amount of loan request**
- **"Value" : Value of current property**
- **"Delinq": Number of delinquent credit lines. A higher number usually indicates a higher probably of default based on past history;**
- **"Clage": Age of oldest credit line in months. A higher number usually indicates a good credit, which leads to a lower probably of default based on past history**
- **"Debtinc": Debt-to-income ratio. A higher number usually indicates a higher probably of default based on the individual's ability to repay the debt**

iv) Confusion Matrix for Training (copy & paste)

825	89	913
36	50	87
861	139	1000

iv) Confusion Matrix for Testing (copy & paste)

	GoodLoan	BadLoan	
GoodLoan	819	91	910
BadLoan	31	59	90
	850	150	1000

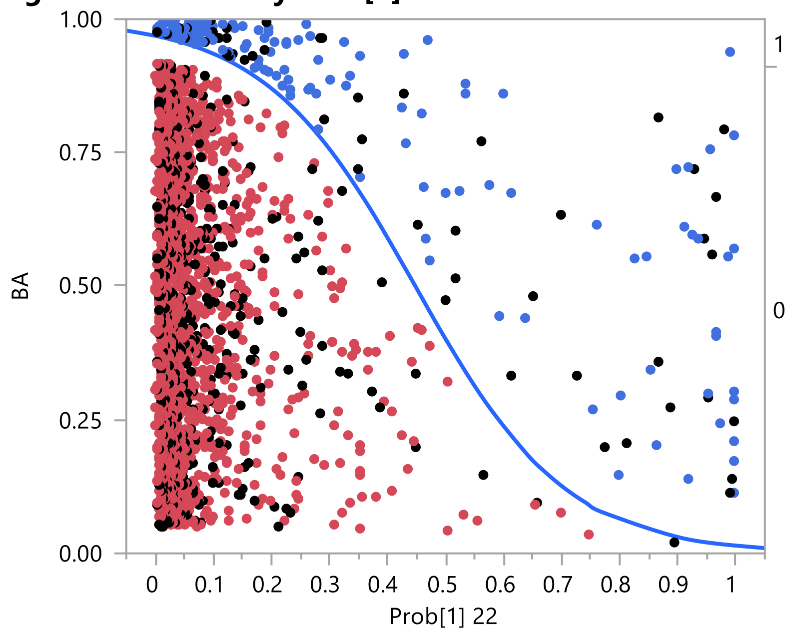
v) Lift Table **(copy & paste from Excel)**

Lift Table in Dollars	Training	Testing
Lift with respect to Baseline - JMP Model	31.3050571	37.8548124
Lift with respect to Baseline - My Best Model	31.4274062	30.12291441
Lift with respect to JMP Model - My Contribution	1.003908286	0.962237964
Overall Lift with respect to Baseline -My Best Model	31.4274062	30.12291441

Lift Table in Propensity	Training	Testing
Lift with respect to Baseline - JMP Model	4.134623336	4.521072797
Lift with respect to Baseline - My Best Model	3.176803558	3.072721065

vi) Attach JMP Printout (Remove unwanted parts – Copy and Paste then edit it.)

Logistic Fit of BAD By Prob[1] 22



Whole Model Test

Model	**-LogLikelihood**	**DF**	**ChiSquare**	**Prob>ChiSq**
Difference	83.00488	1	166.0098	<.0001*
Full	219.53294			
Reduced	302.53782			

RSquare (U)	0.2744
AICc	443.078
BIC	452.881
Observations (or Sum Wgts)	1000

Measure	Training	Definition
Entropy RSquare	0.2744	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.3369	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.2195	∑ -Log(ρ[j])/n
RMSE	0.2416	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.1153	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0690	∑ (ρ[j]≠ρMax)/n
N	1000	n

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	3.4393461	0.1813232	359.79	<.0001*
Prob[1] 22	-7.6791467	0.7989179	92.39	<.0001*

For log odds of 0/1

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	99.07251	12	198.145	<.0001*
Full	203.46531			
Reduced	302.53782			

RSquare (U)	0.3275
AICc	433.3
BIC	496.731
Observations (or Sum Wgts)	1000

Measure	Training	Definition
Entropy RSquare	0.3275	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.3960	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.2035	∑ -Log(ρ[j])/n
RMSE	0.2338	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.1105	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0700	∑ (ρ[j]≠ρMax)/n
N	1000	n

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	987	203.46531	406.9306
Saturated	999	0.00000	Prob>ChiSq
Fitted	12	203.46531	1.0000

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Saturated	999	0.00000	Prob>ChiSq
Fitted	12	203.46531	1.0000

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept[0]	-36.253816	9.8892804	13.44	0.0002*
VALUE	-1.2421e-5	6.647e-6	3.49	0.0617
REASON[DebtCon]	0.23284482	0.1576488	2.18	0.1397
JOB{Office&ProfExe-Self&Other&Mgr&Sales}	0.08023248	0.147395	0.30	0.5862
DEROG	-0.697103	0.1915348	13.25	0.0003*
DELINQ	-0.7719048	0.1273723	36.73	<.0001*
CLAGE	0.00923648	0.0022252	17.23	<.0001*
NINQ	-0.1797981	0.0721581	6.21	0.0127*
CLNO	0.01397287	0.0151952	0.85	0.3578
DEBTINC	-0.3380778	0.0661574	26.11	<.0001*
Log (Loan)	0.71659771	0.6138844	1.36	0.2431
Log (Value)	4.68956713	1.839139	6.50	0.0108*
Log(Debt)	16.0256744	4.4714031	12.85	0.0003*

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
VALUE	1	1	2.75152385	0.0972
REASON	1	1	2.13157075	0.1443
JOB{Office&ProfExe-Self&Other&Mgr&Sales}	1	1	0.29917636	0.5844
DEROG	1	1	12.8997384	0.0003*
DELINQ	1	1	49.4642538	<.0001*
CLAGE	1	1	20.344261	<.0001*
NINQ	1	1	5.82009496	0.0158*
CLNO	1	1	0.86813736	0.3515
DEBTINC	1	1	49.3374083	<.0001*
Log (Loan)	1	1	1.38257833	0.2397
Log (Value)	1	1	5.47531416	0.0193*
Log(Debt)	1	1	21.1312947	<.0001*

Key information → Cutoff Probability for mailing = 0.14