



UCIRVINE | EXTENSION



Credit Risk Modeling Project

Subatra Devi Selvaraj

Winter 2015 / 2016

March 17, 2016

Table of Contents

- Lending Club Business Model
- Statement of Problem
- Objectives
- Executive Summary
- Modeling Methodology
 - Overview
 - Data Discovery – Cluster Analysis
 - Classification Model - Regression Analysis
- Conclusion
- Potential Future Analysis
- Questions / Discussion



Lending Club Business Model

Lending Club (www.LendingClub.com) is a San Francisco based online provider of loans for debt consolidation, small business expenses, moving costs, home improvements, medical expenses or other select major one time purchase or expenditures.

Lending Club's website advertises consolidating your variable, high interest debt with a single, low, fixed rate loan with APRs ranging from 5.99% to 35.96% (depending on credit score, loan amount, loan terms, and credit usage and history).



Business Problem Statement and Objective

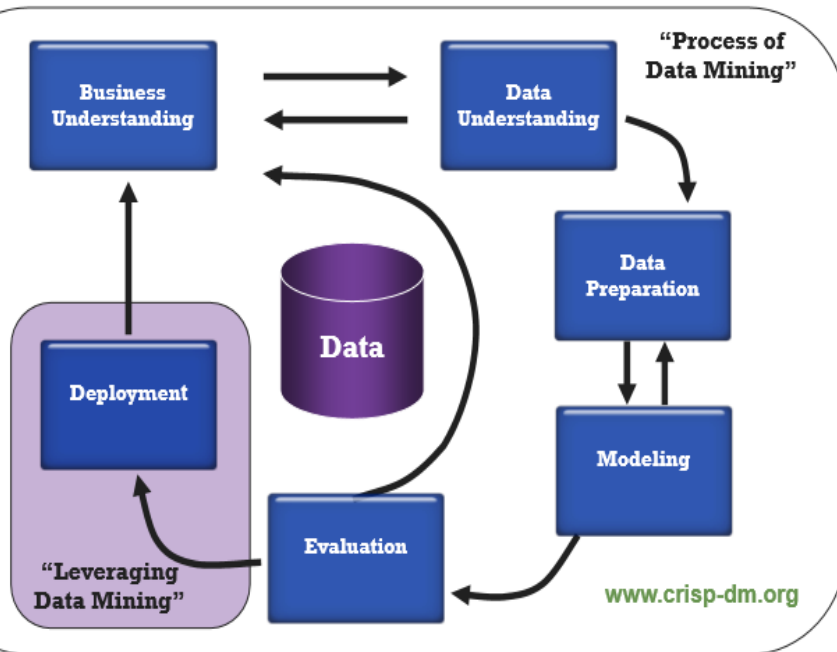
Statement of Problem

- Improve on Lending Club's existing risk management and risk pricing strategy in order to maximize economic value to the owners of the organization.

Objectives

- Utilize historical loan default data to credit risk segmentation used in credit scoring so proper risk pricing component can be utilized to ensure returns are achieved across entire credit spectrum during pricing process.
- This will be accomplished if the team can build a credit scoring model that will utilize the existing independent variables that will have higher predictive power over the legacy model, and result in a lower default rate, which will increase operating profits (and shareholder value).

CRISP-DM Approach



- Exploratory Data Analysis
- Created CAR Customer Analytics Record
- Excluded Rejects
- Classification / Regression Model using CAR
- Evaluated the Model based on Probability Scores
- Built Model 2 – Weighted Regression
- Appended the Probability Score to the CAR
- Created Supervised Clustering Model based on CAR



Lending Club – Data Understanding

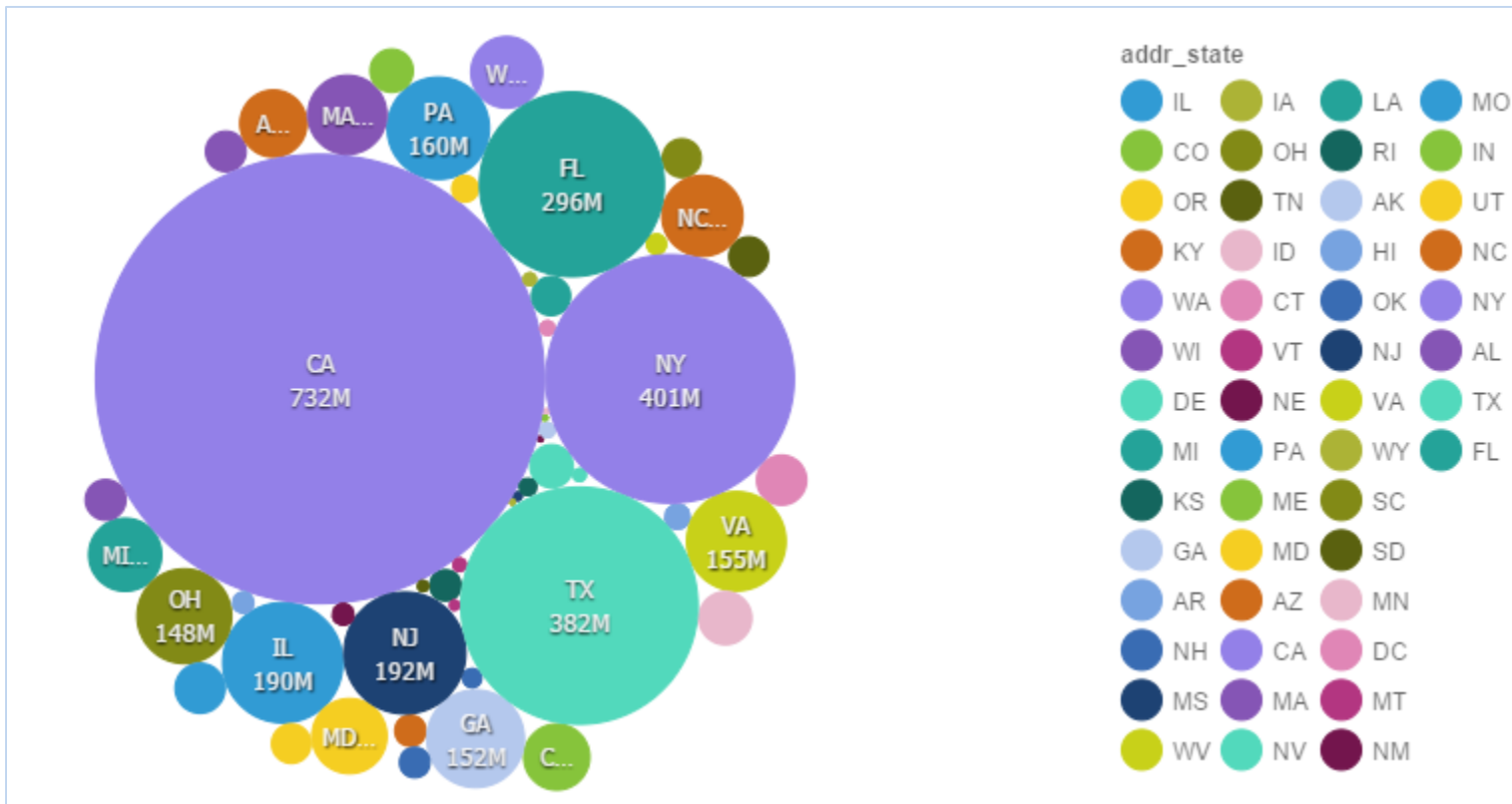
Class and Years			
Class	Year	Loan Count	Method
A	2007-2011	42,535	FICO
B	2012-2013	188,123	FICO & Variable Measures
C	2014-2015	656,724	FICO & Revised Variable Measures

Variable Measures				
Borrower Assessment	Loan Characteristic	Borrower Assessment	Credit History	Borrower Indebtedness
Grade	Loan Purpose	Annual Income	Credit History Length	Loan Amount to Annual Income
Sub- Grade	Loan Amount	Housing Situation	Delinquency 2 years	Annual Installment to Income
Interest Rate		Emp Length	Inquires Last 6 mo	Debt to Income
			Public Records	
			Revolving Utilization	
			Open Accounts	
			Months Since Last Delinq	

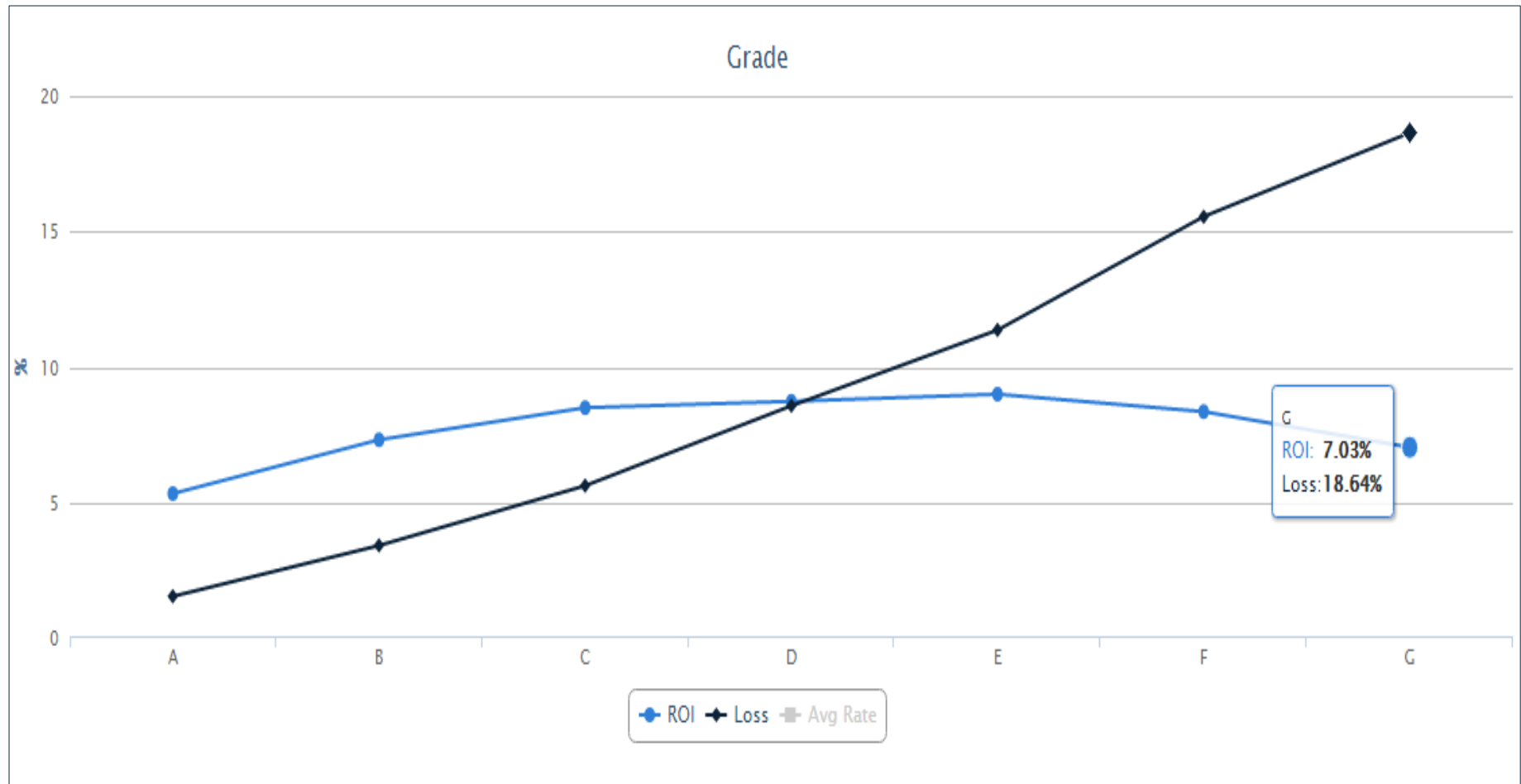
Lending Club Interesting Facts				
Variable	Best	ROI	Worst	Loss
Grade:	E	8.99%	G	18.65%
Purpose:	Wedding	8.07%	Small Business	13.01%
Year:	2013	8.29%	2007	14.80%
FICO:	>674 <= 679	8.33%	<=664	8.54%
Annual Income:	>150K	8.74%	<= 28K	8.10%
Inquires:	<=0	8.08%	>3	10.94%



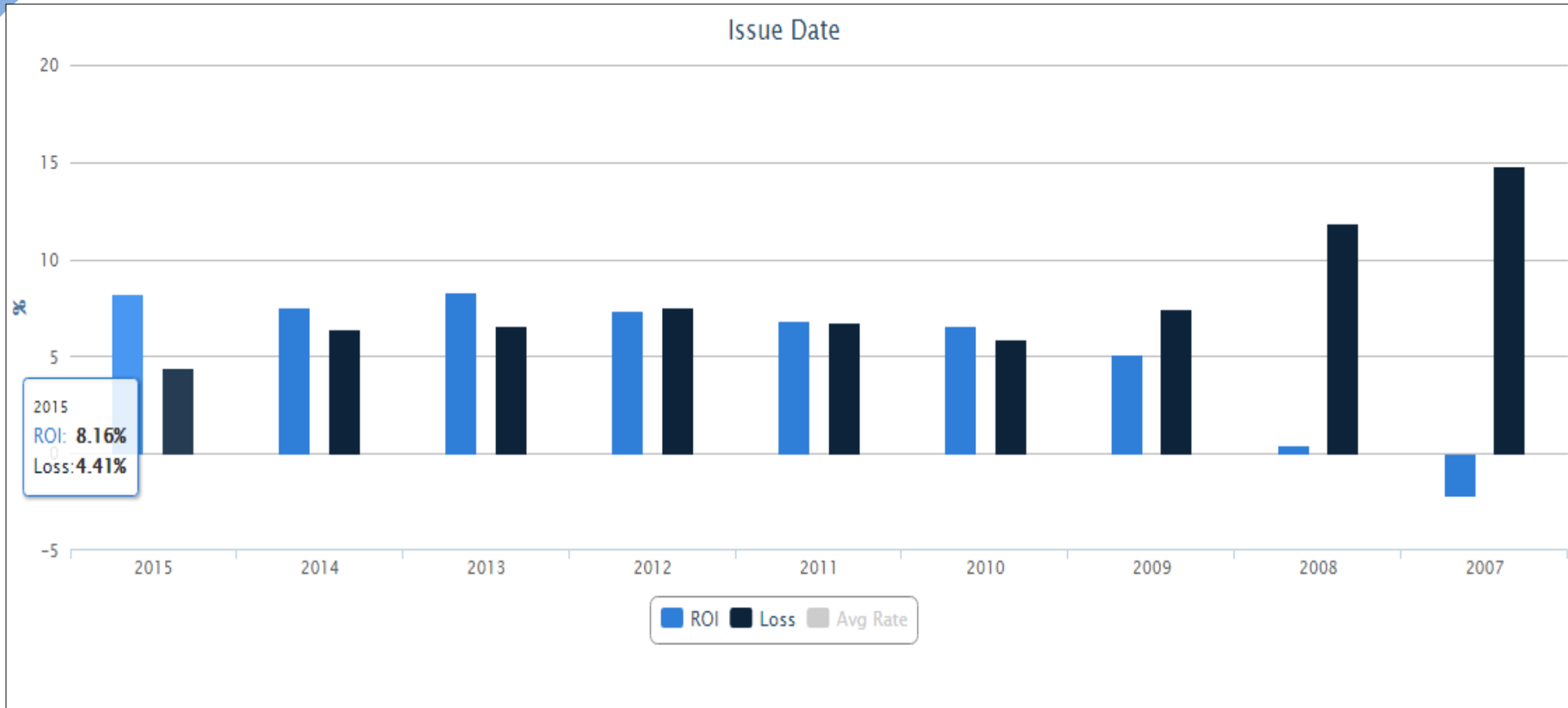
Data Understanding



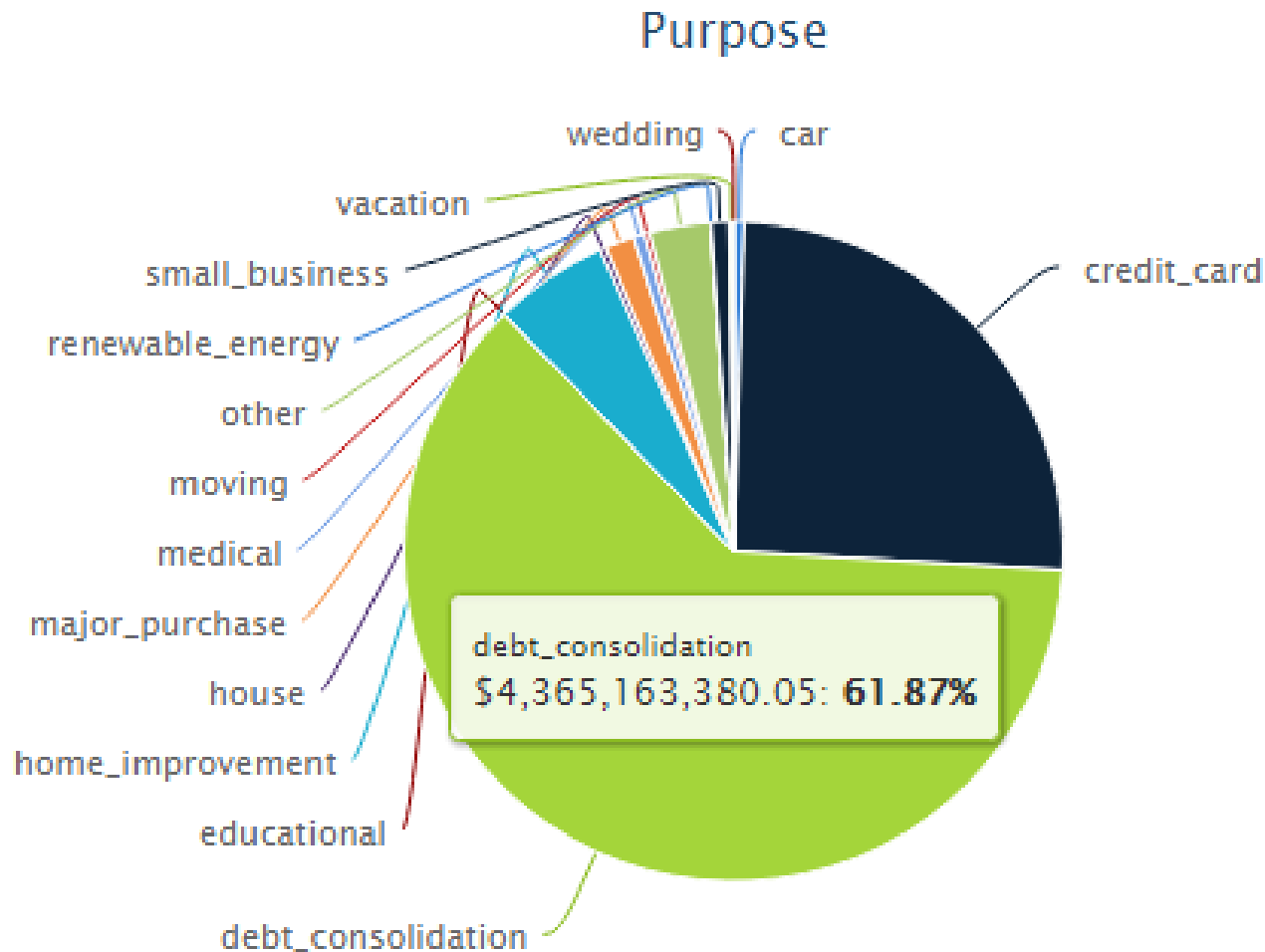
Lending Club – Data Understanding



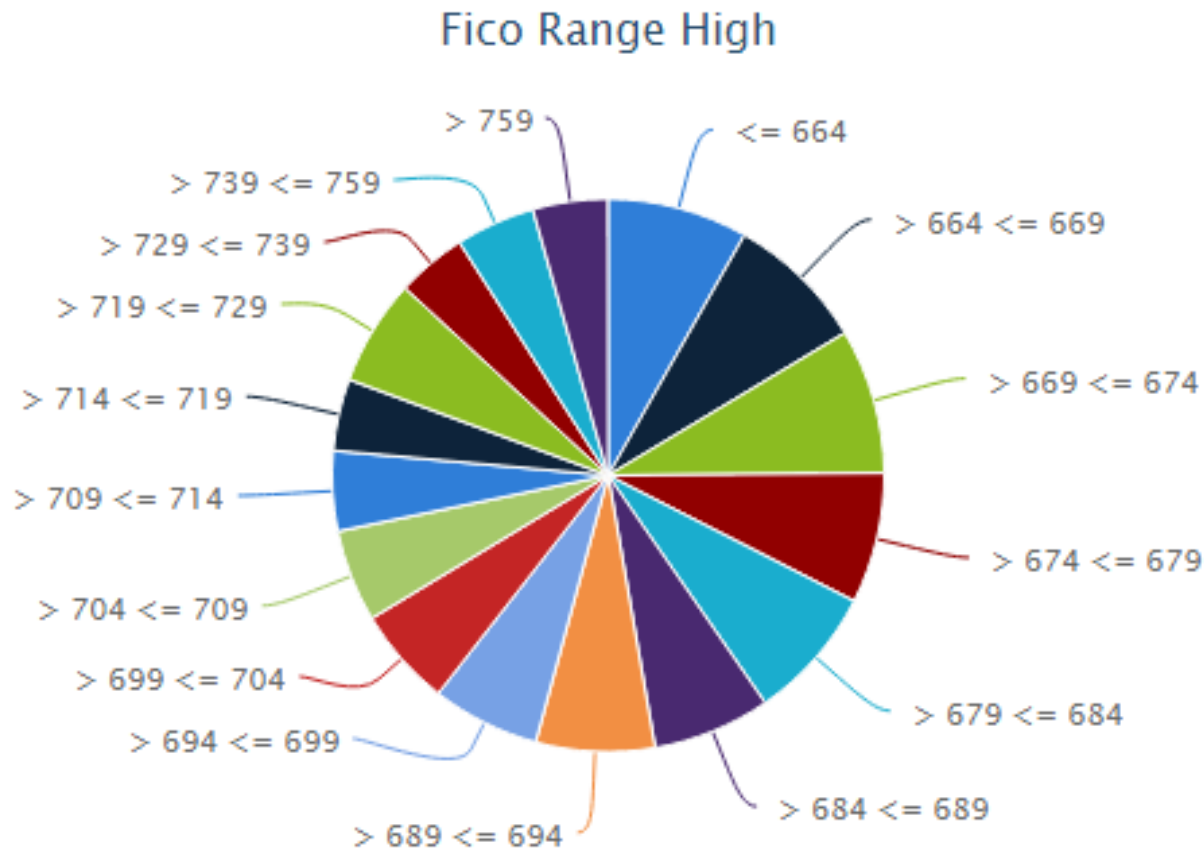
Lending Club – Data Understanding



Lending Club – Data Understanding



Lending Club – Data Understanding



Classification Model – Executive Summary

- The credit risk modeling team utilized SAP Predictive Analytics to perform data discovery on Lending Club's historical loan portfolio and then create a classification model using the statistically relevant independent variables to improve the prediction of a borrower's probability of default.
- This modeling strategy yielded an improved identification and segmentation of the independent variables and resulted in a measurable improvement to the existing probability of default rate of 49.89%



Classification Model 1 - Summary

Eligible Population

Number of records : 333,162

- Description of the population on which the score is computed

Target Population

Observed Target Frequency : 5.73%

- Description of the Modeled Phenomena

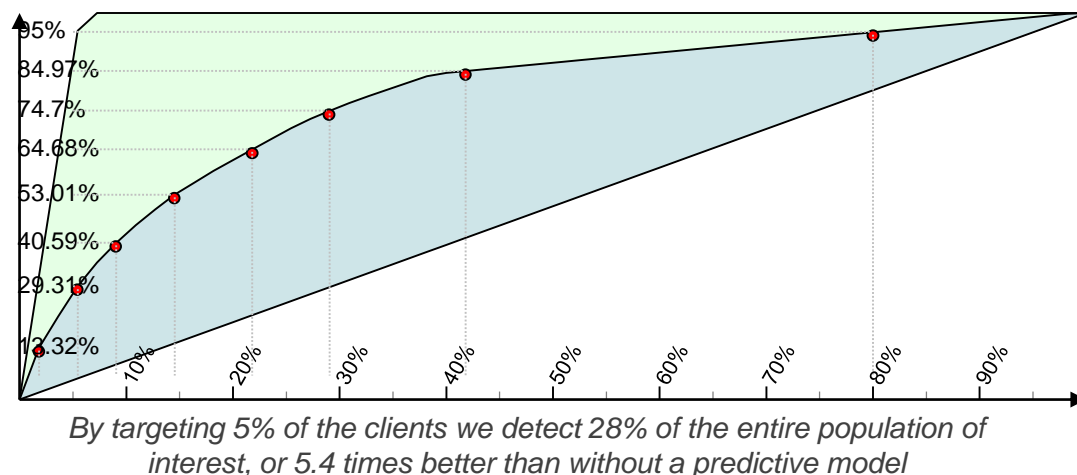
Predictive Power (KI) = 0.601



Prediction Confidence (KR) = 0.995



Performance and Quality



Variable contribution

Less contributive	Variable Name	Most contributive
{c}	loan_type	{a}
{163;168;169;171;172;174; ...}	mo_sin_old_rev_tl_op	KxMissing
{A1}	sub_grade	{KxOther}
{10;11;3;4}	num_bc_sats	KxMissing
[13 ; 211]	mo_sin_rcnt_tl	KxMissing
{0;KxOther}	num_tl_30dpd	KxMissing
[0 ; 37]	num_actv_rev_tl	KxMissing
]96000 ; 7.4464e+006]	annual_inc	[1896 ; 26700] & KxMissing
[52900 ; 560800]	total_bc_limit	KxMissing
{0}	num_tl_op_past_12m	KxMissing



UCIRVINE | EXTENSION



Classification Model 2 Summary

Eligible Population

Number of records : 38,136

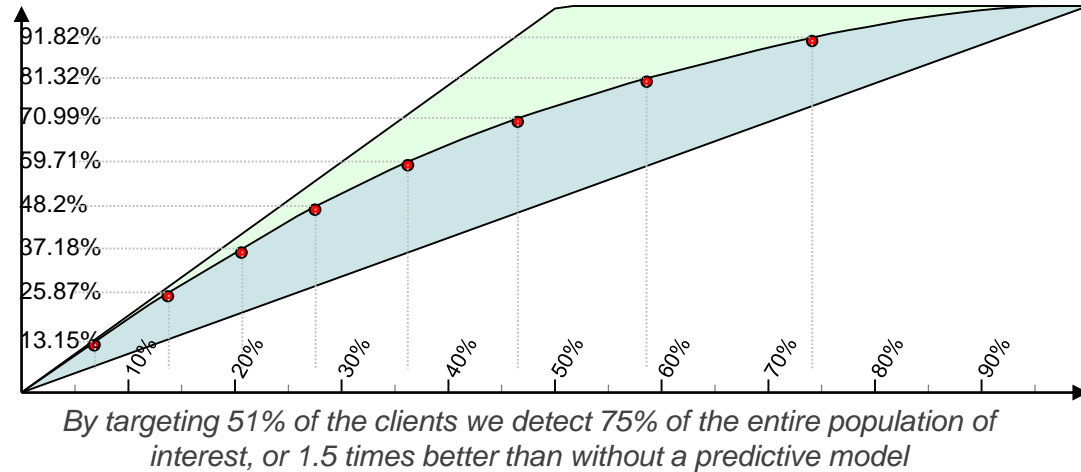
•Description of the population on which the score is computed

Target Population

Observed Target Frequency : 49.89%

•Description of the Modeled Phenomena

Performance and Quality



Variable contribution

Less contributive	Variable Name	Most contributive
{a}	loan_type	{c}
KxMissing	num_tl_120dpd_2m	{KxOther;0}
KxMissing	num_actv_rev_tl	[0 ; 29]
{E4;F2;F3;F4;F5}	sub_grade	{A1;A2}
[1896 ; 22440] & KxMissing	annual_inc]94000 ; 136500]
{f}	met	{t}
KxMissing	mths_since_recent_bc]62 ; 538]
KxMissing	tot_cur_bal]273322 ; 2.54717e+006]
KxMissing	acc_open_past_24mths	[0 ; 2[
KxMissing	num_actv_bc_tl	[1 ; 18]

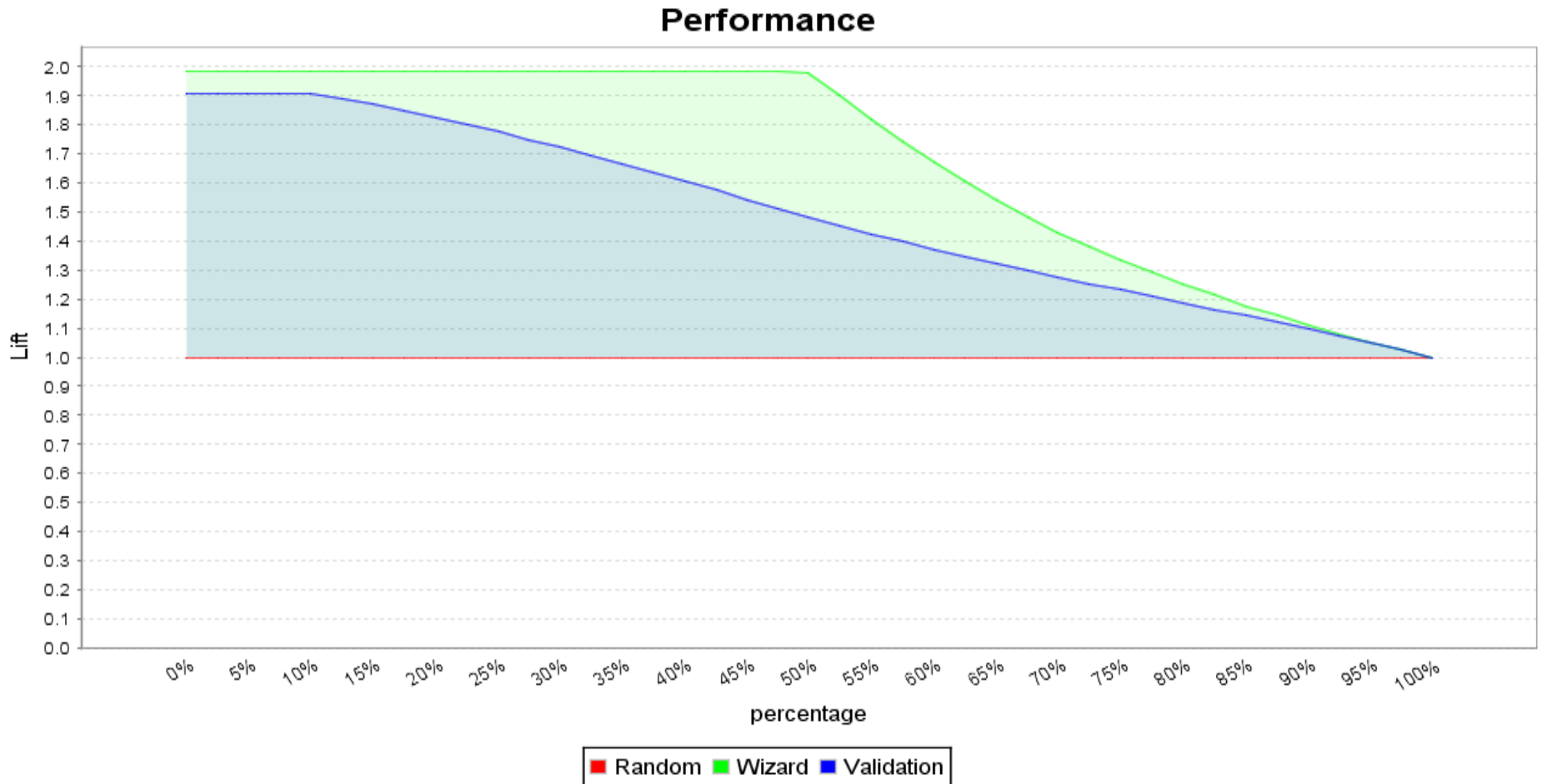
Predictive Power (KI) = 0.652



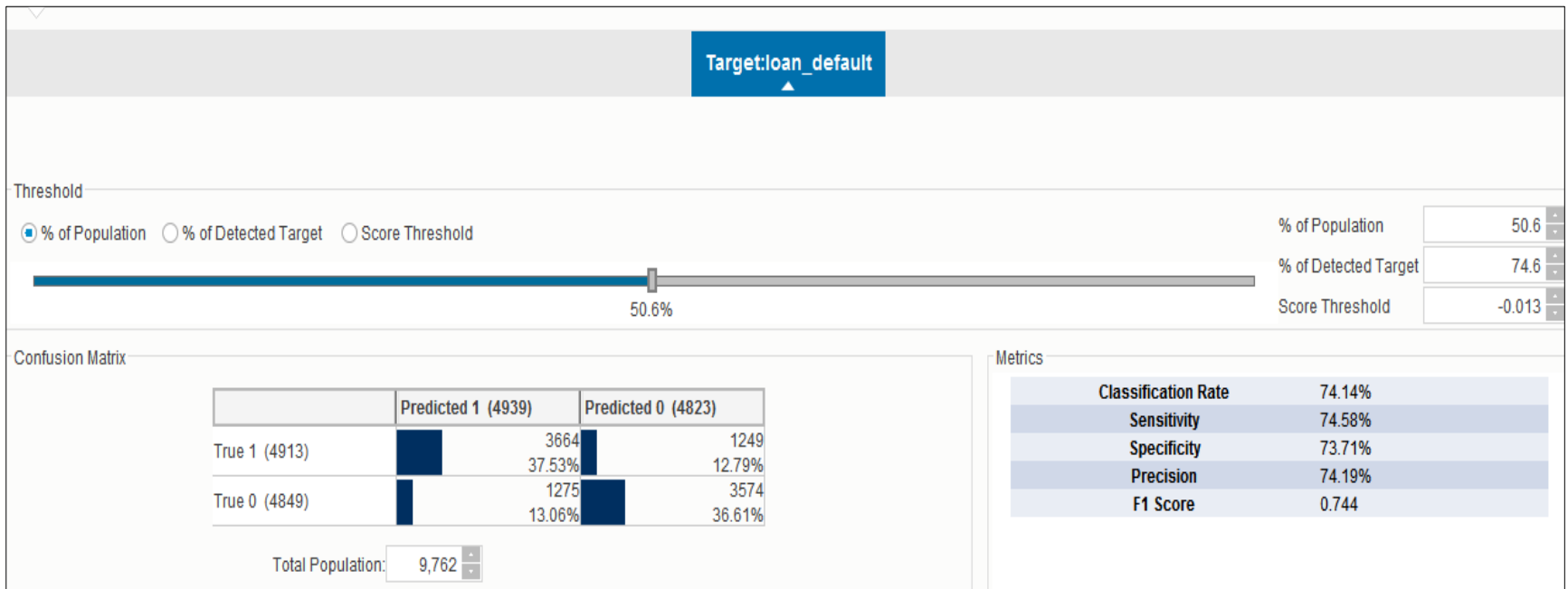
Prediction Confidence (KR) = 0.989



Lift Curve– Model 2



Confusion Matrix





Model Evaluation

Probability Scores - Model 1 (KI 0.601)	Probability Scores - Model 2 (KI 0.652)	TRUE Positives	False Positives	Total records	Loan Amount gain (\$)	Loan Amount Loss (\$)	Net Profit in \$
> 90%		179	10	189	1,630,650	142,875	1,487,775
> 80%		268	20	288	2,259,050	215,475	2,043,575
> 70%		379	56	435	3,143,900	554,100	2,589,800
> 60%		418	78	496	3,413,775	842,100	2,571,675
<10%		9338	275576	284914	134,529,650	3,993,112,350	(-3,858,582,700)
	> 95%	120	61169	61289			
	< 10%	2453	732	3185	24,085,175	9,946,550	14,138,625
	<20%	5689	10609	16298	66,726,925	1,44,479,400	-77752475

Defaults identified by Model 2: 2453

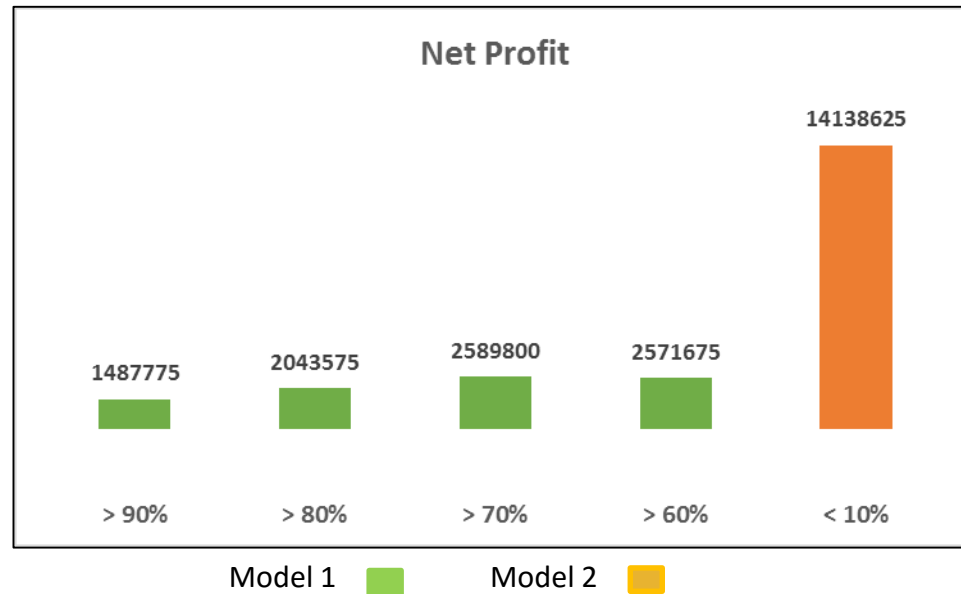
Total Defaults(Baseline): 19068



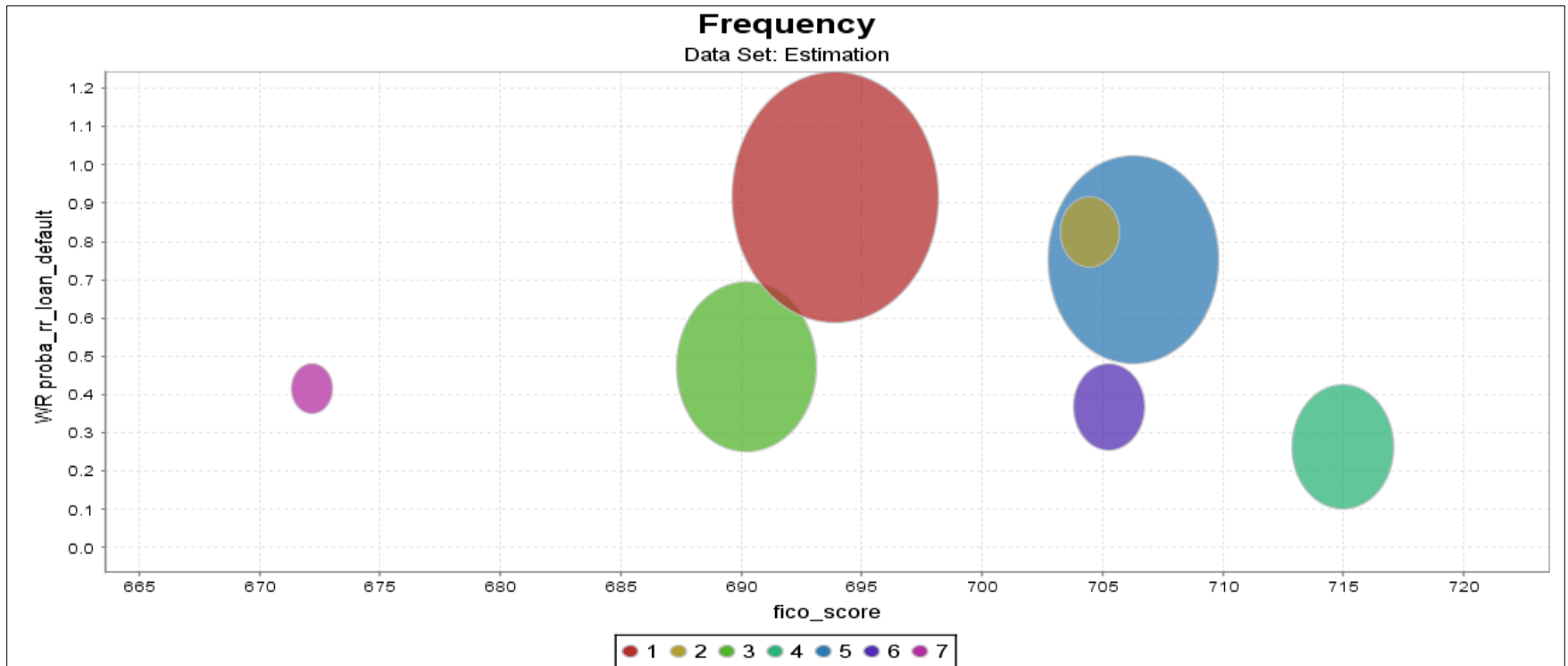
Recommendations

- Classification Model 2 can be used to predict the Loan Defaults in future. Borrowers with Probability Score $< 10\%$ can be Rejected.
- Based on the Evaluation, the Profit with Model 2 is

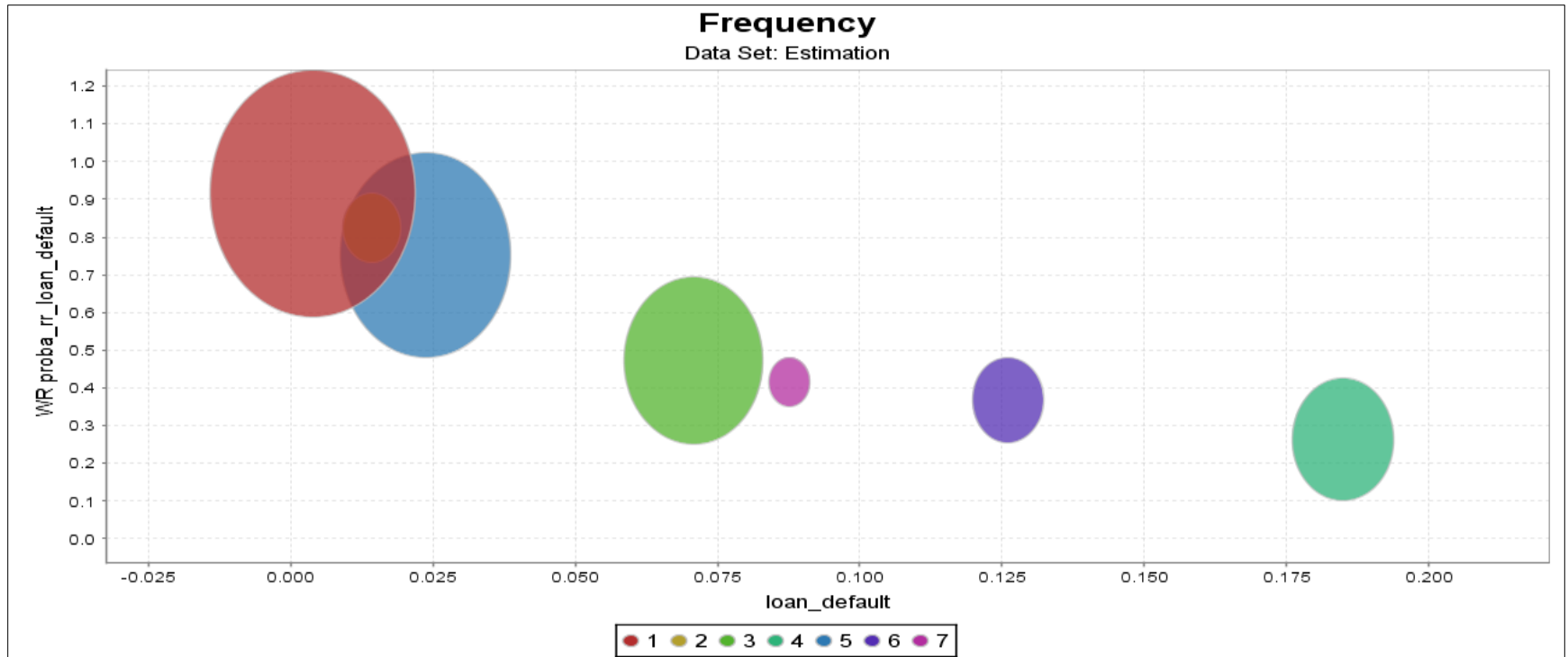
\$ 14,138,625



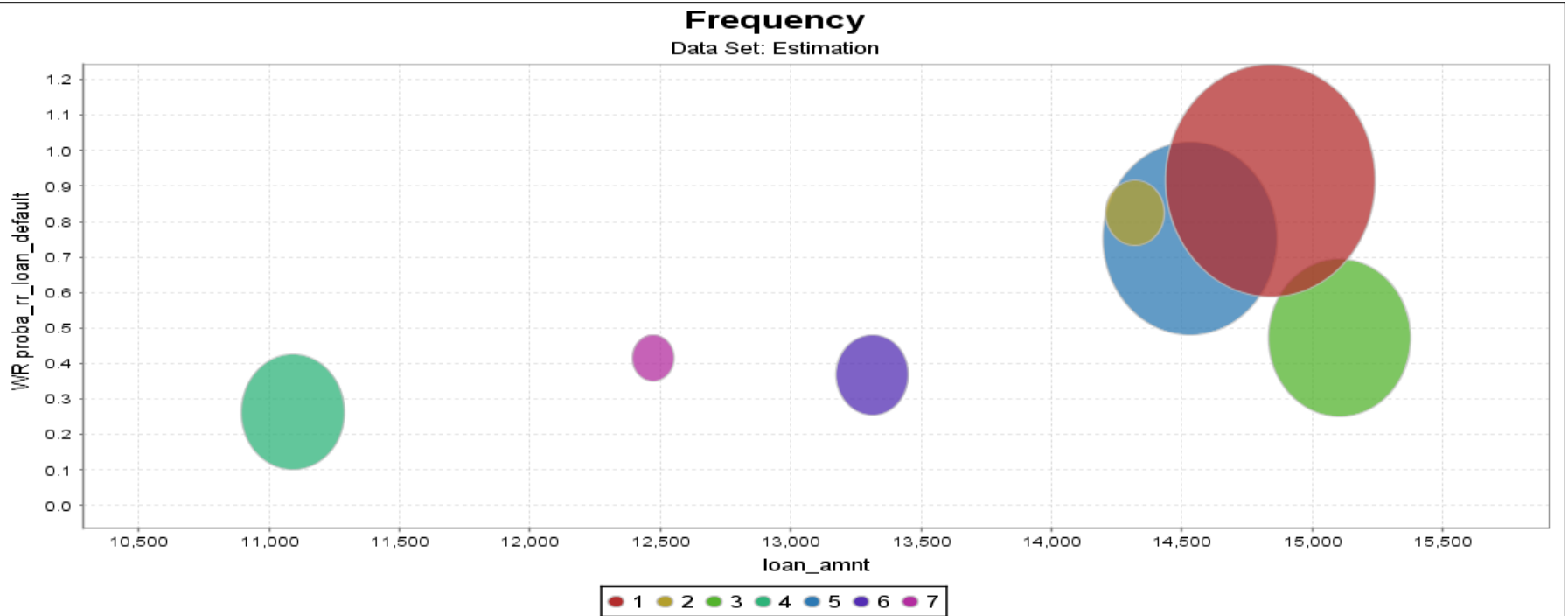
Clustering Model – 2 (fico_score)



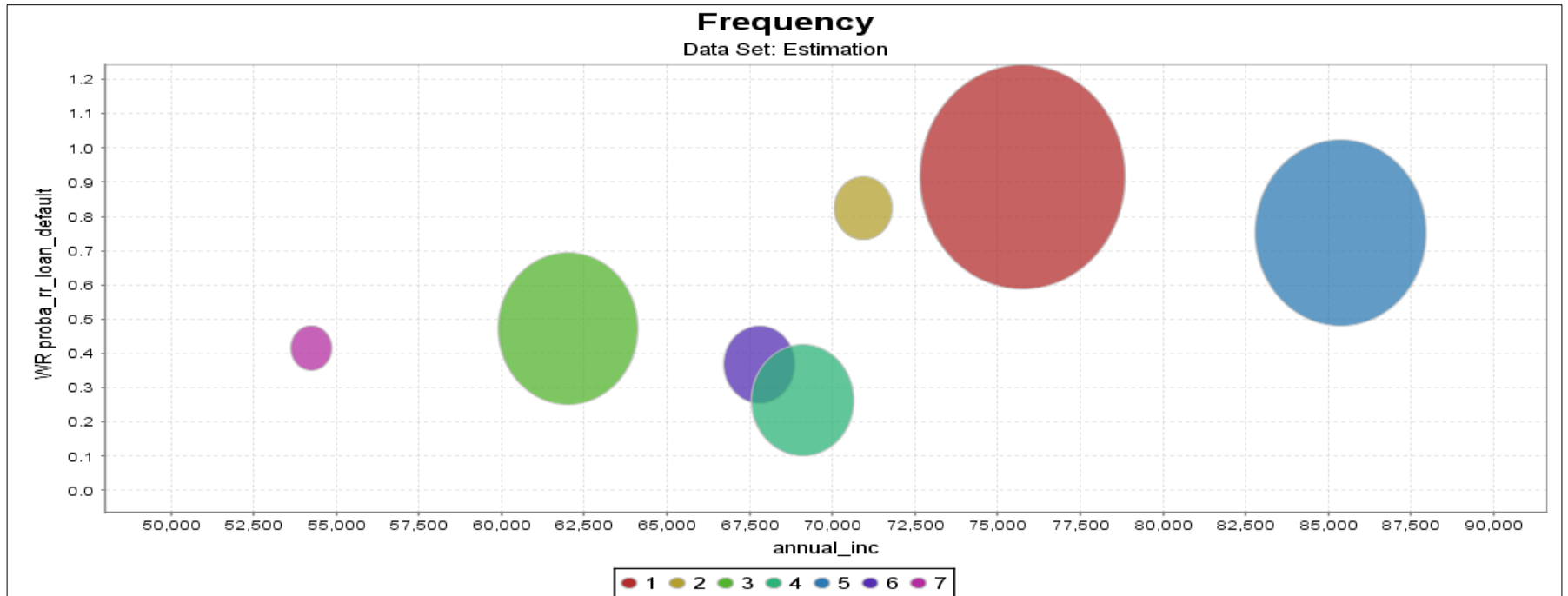
Loan Default



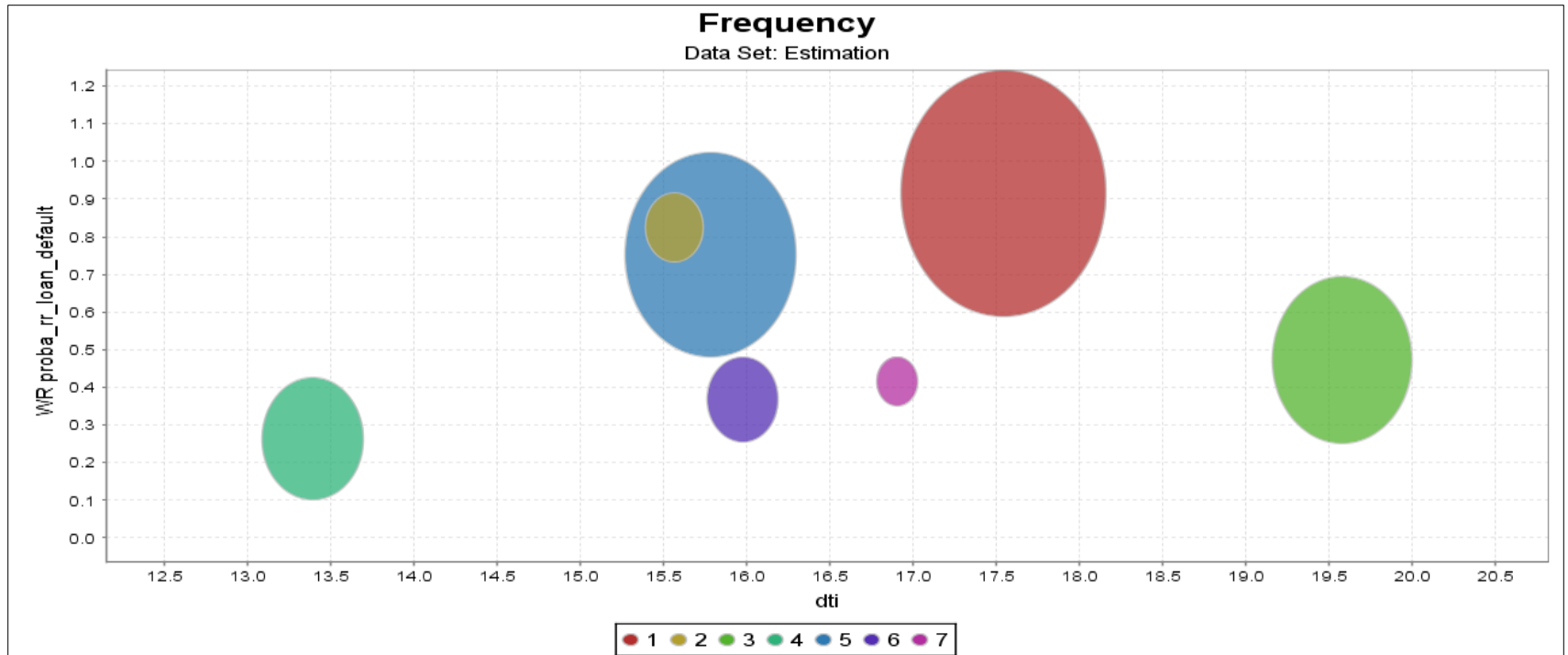
Loan Amount



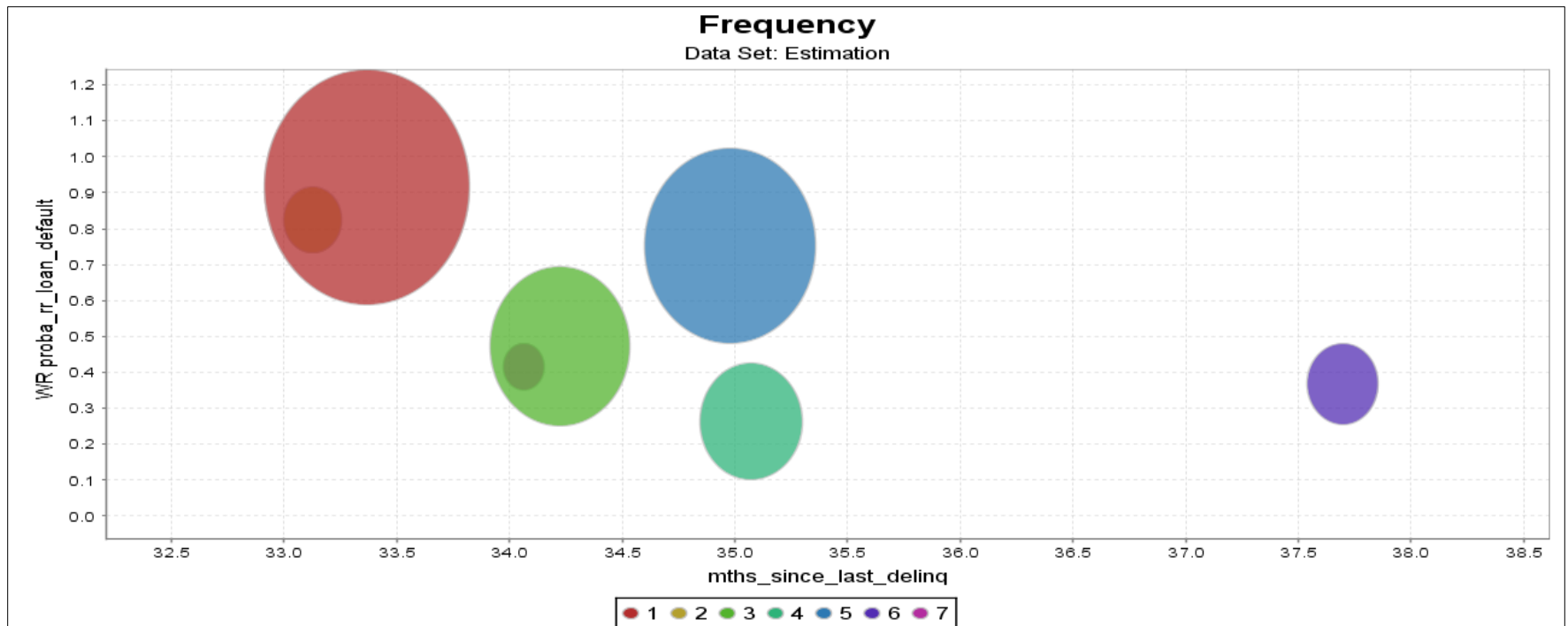
Annual Income



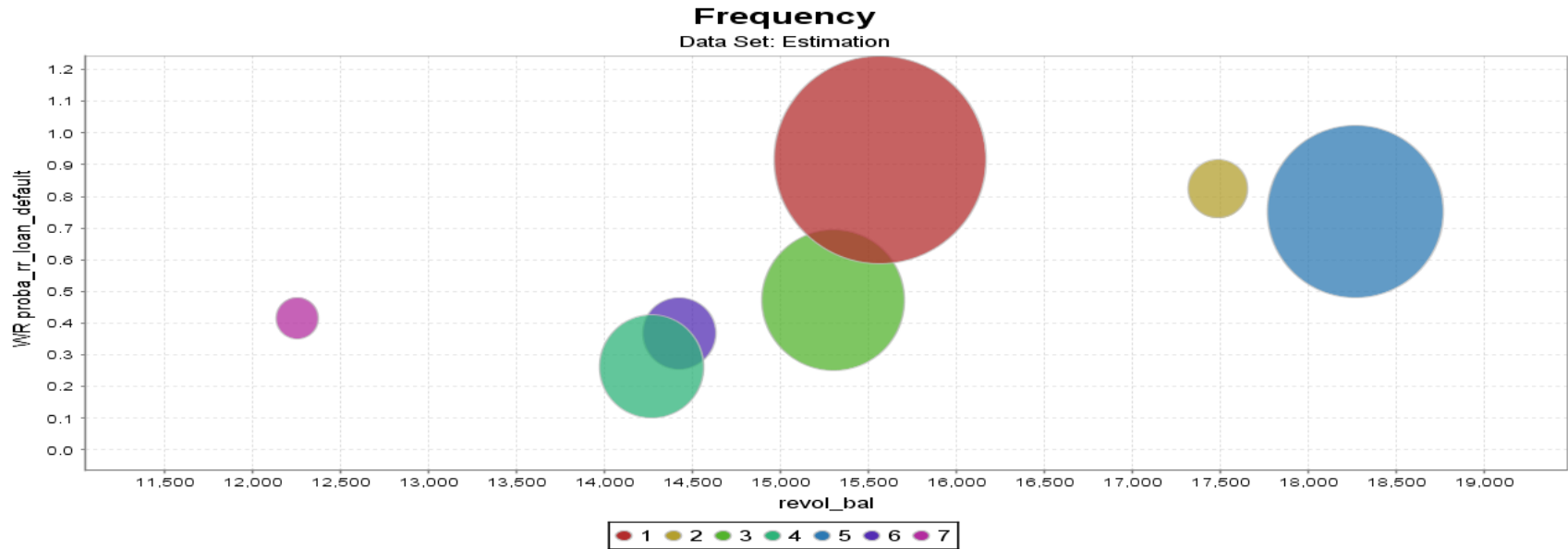
Debt to Income Ratio



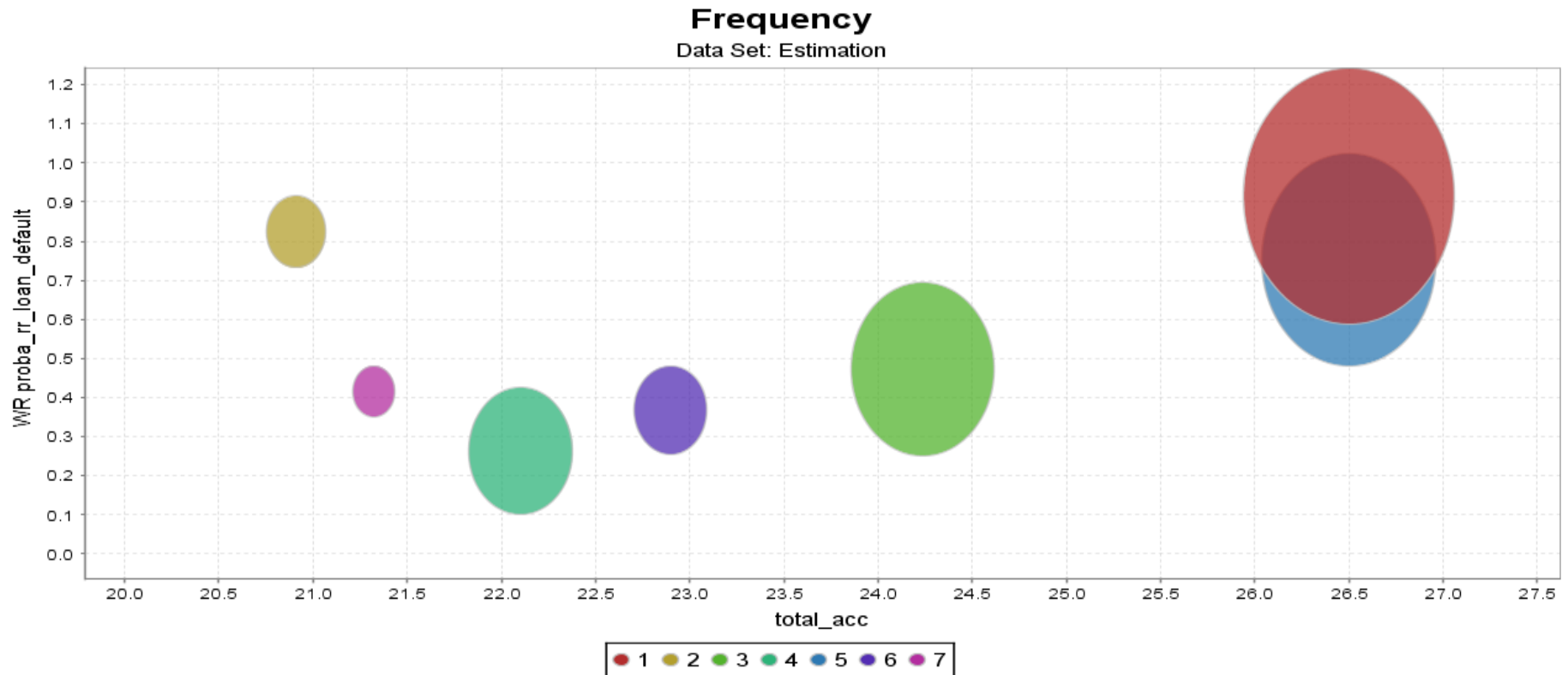
Months since last delinquency



Total credit revolving balance



Total Accounts



Appendix



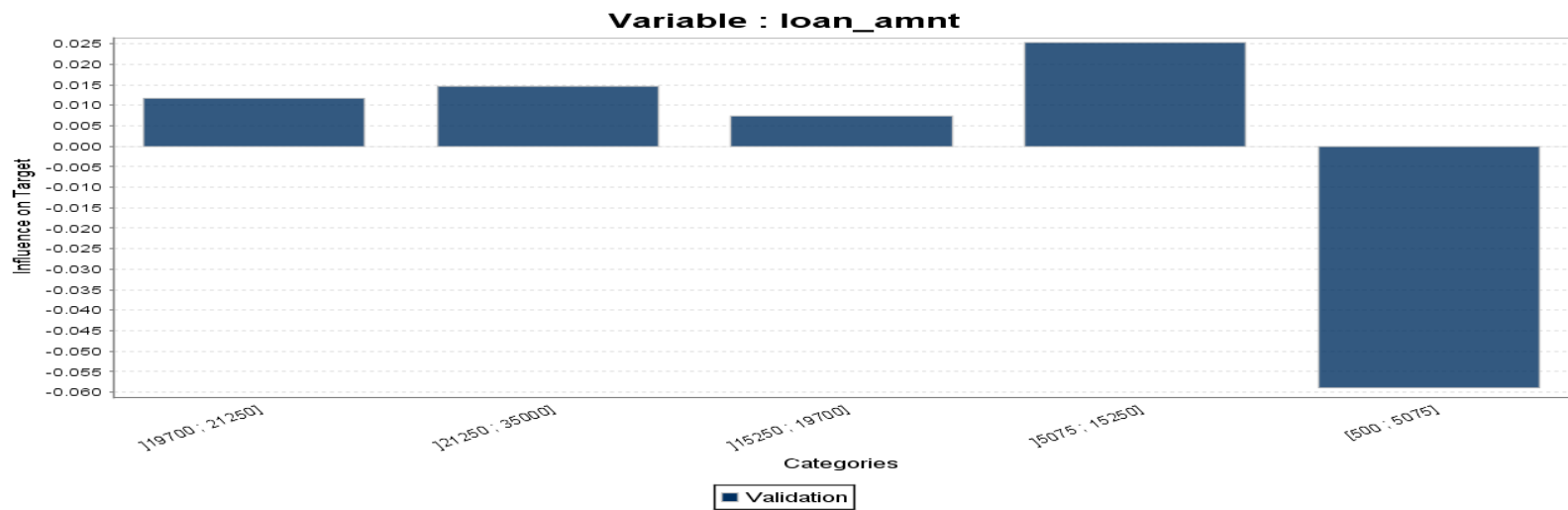
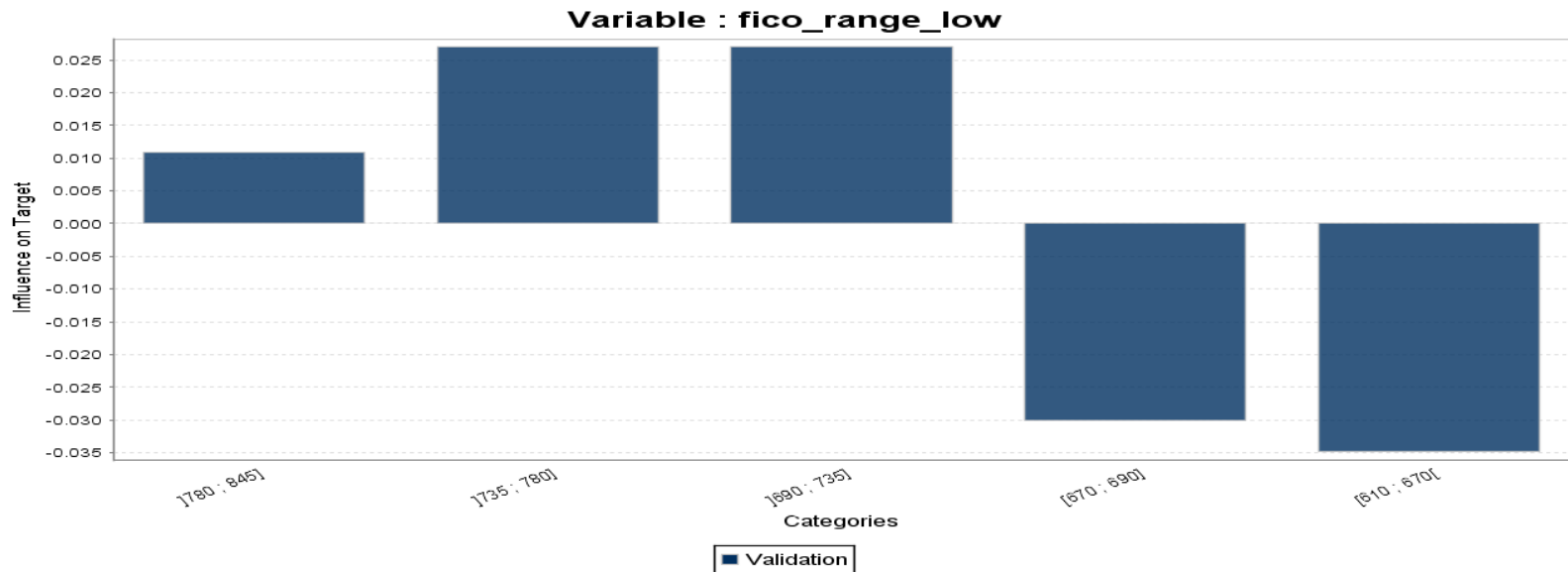
CAR - Sample

id	member_id	loan_default	fico_range_low	fico_range_high	fico_score	loan_amnt	annual_inc	dti	home_ownership	zip_code	addr_state	grade	sub_grade	is_inc_v	acc_now
54734	80364	1	735	739	737	25000	85000	19.48	RENT	941xx	CA	B	B4	Verified	0
55521	107577	0	640	644	642	1000	30000	23.84	RENT	600xx	IL	F	F2	Not Verified	0
55742	114426	1	705	709	707	7000	65000	14.29	RENT	112xx	NY	B	B5	Not Verified	0
56413	129814	0	645	649	647	7000	189500	22.47	MORTGAGE	956xx	CA	F	F2	Verified	0

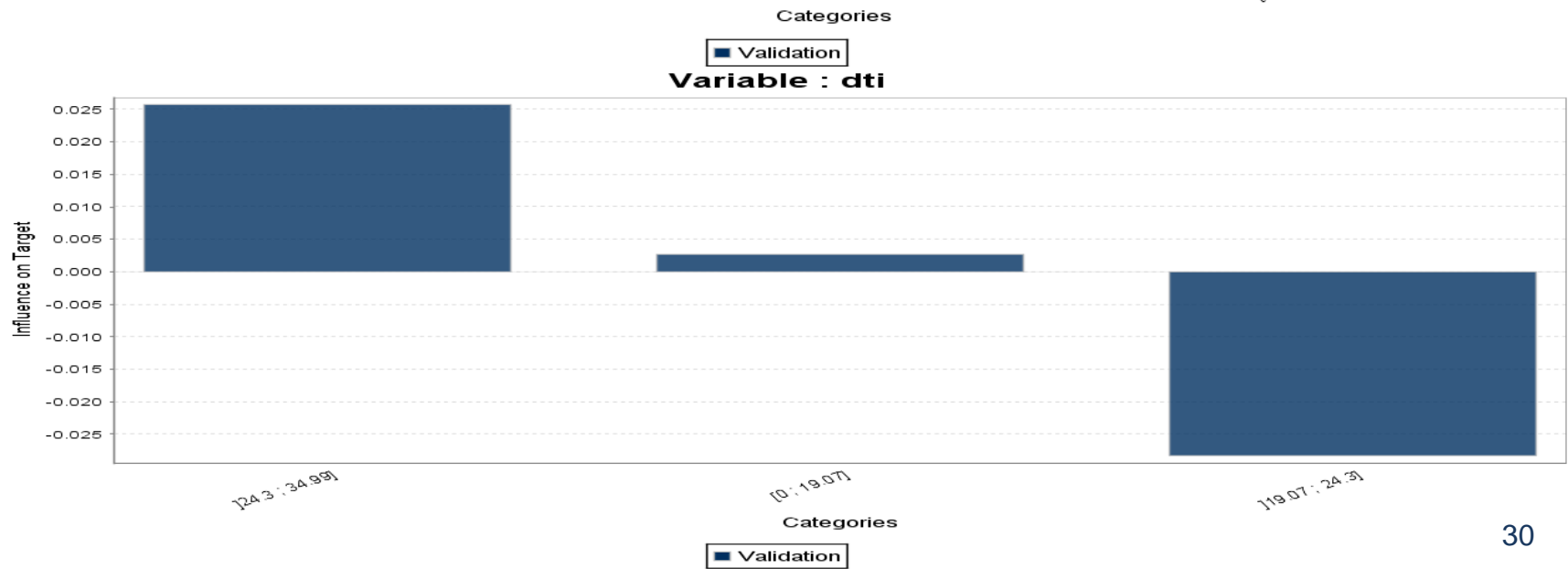
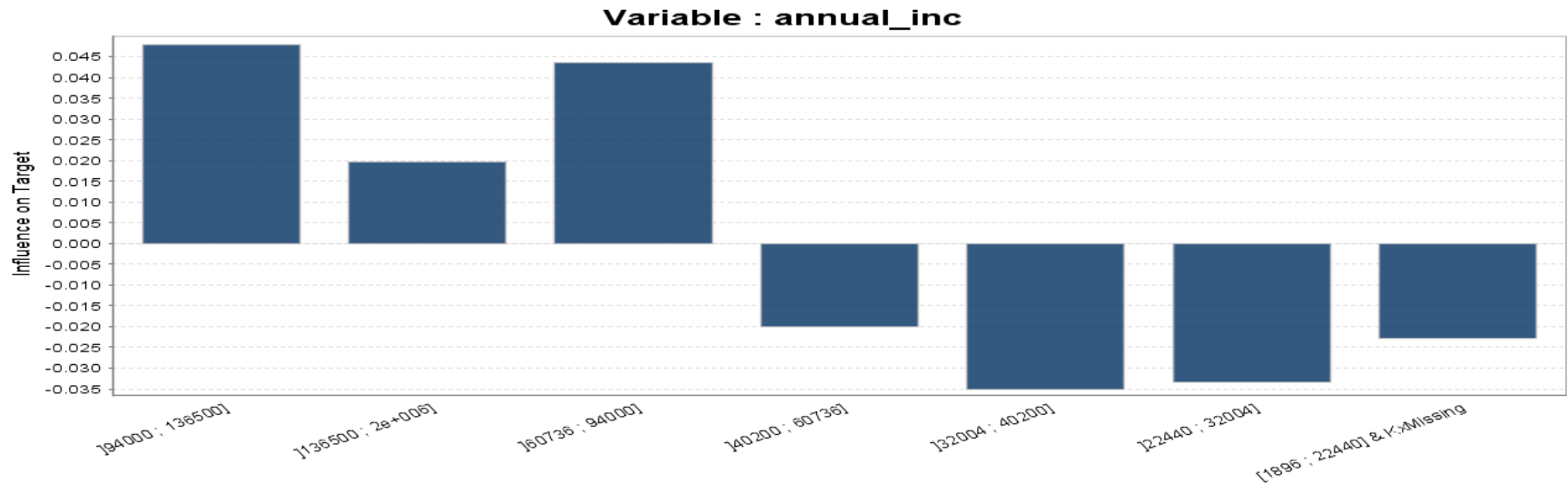
mths_since_last_maj or derog	num_sats	num_tl_op_ past 12m	mo_sin rcnt tl	tot_hi_c red lim	tot_cur bal	avg_cur bal	num_bc tl	num_ac tv bc tl	num_bc sats	pct_tl_n vr dlq	mo_sin old_re v tl op	mo_sin rcnt_r ev_tl_o p	total_re v_hi_li m	num_re v_tl_bal gt 0	num_op rev tl	tot_coll amt	loan_ty pe	met
																	a	t
																	a	f



Variable Contribution

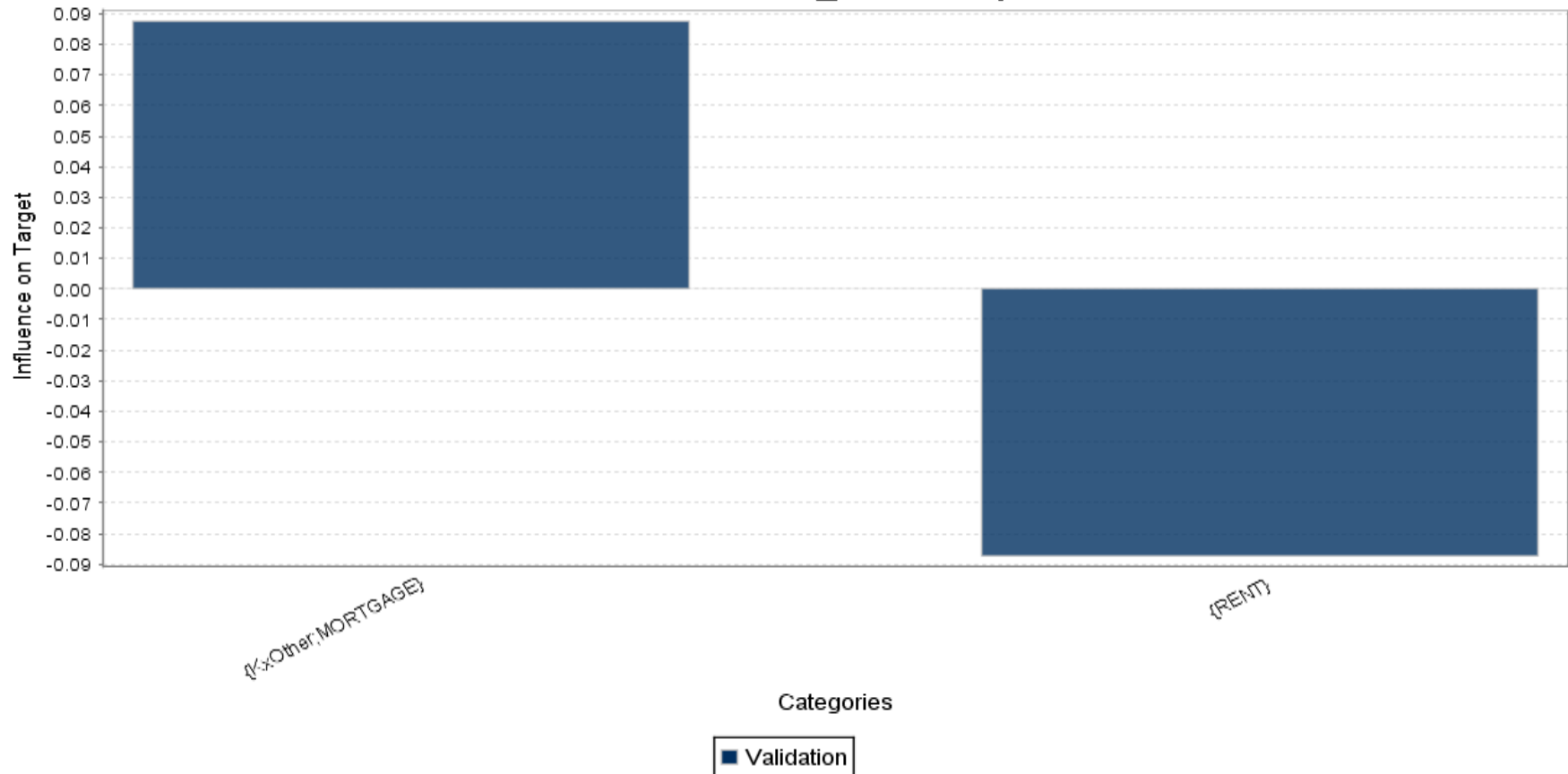


Variable Contributions - Contd



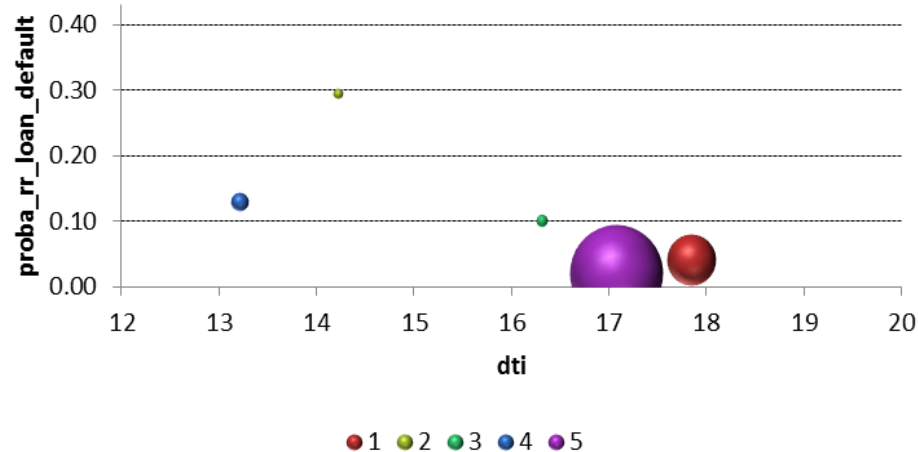
Variable Contribution - Contd

Variable : home_ownership

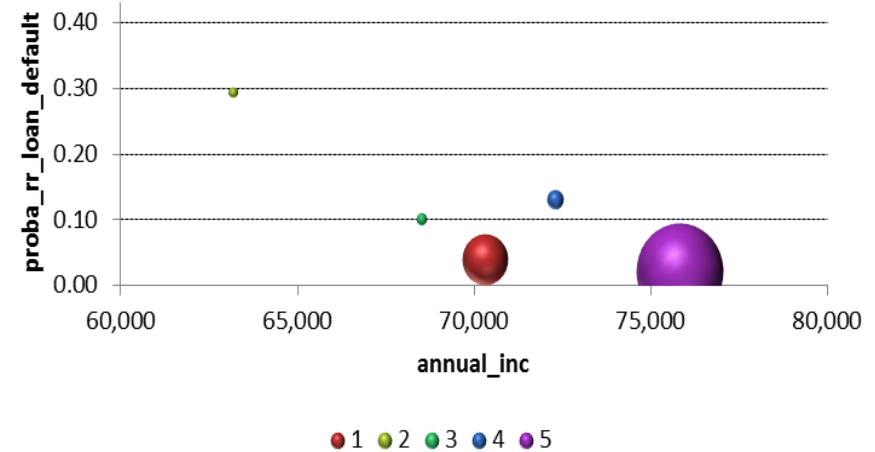


Data Discovery - Cluster Analysis Model 1

Probability of Default vs. Debt to Income Ratio

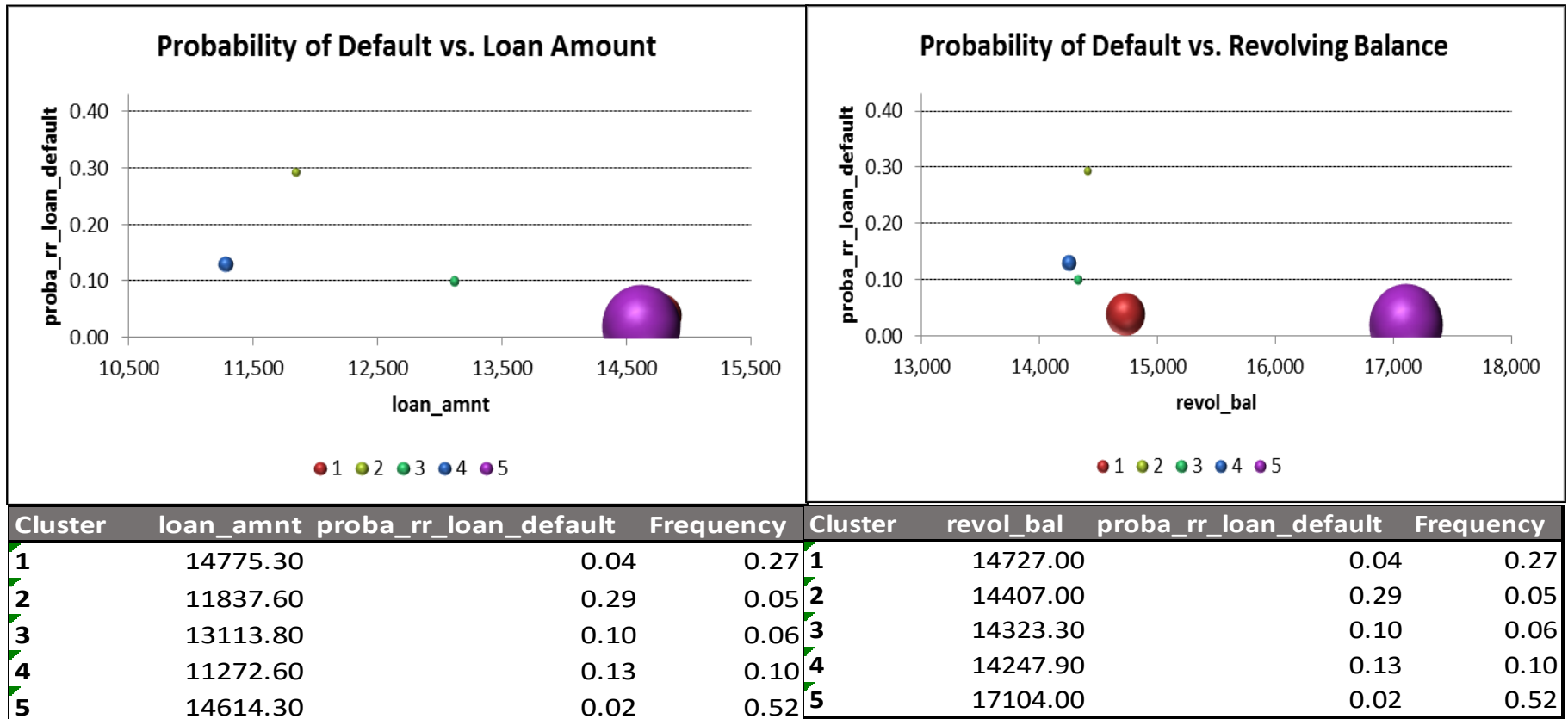


Probability of Default vs. Annual Income

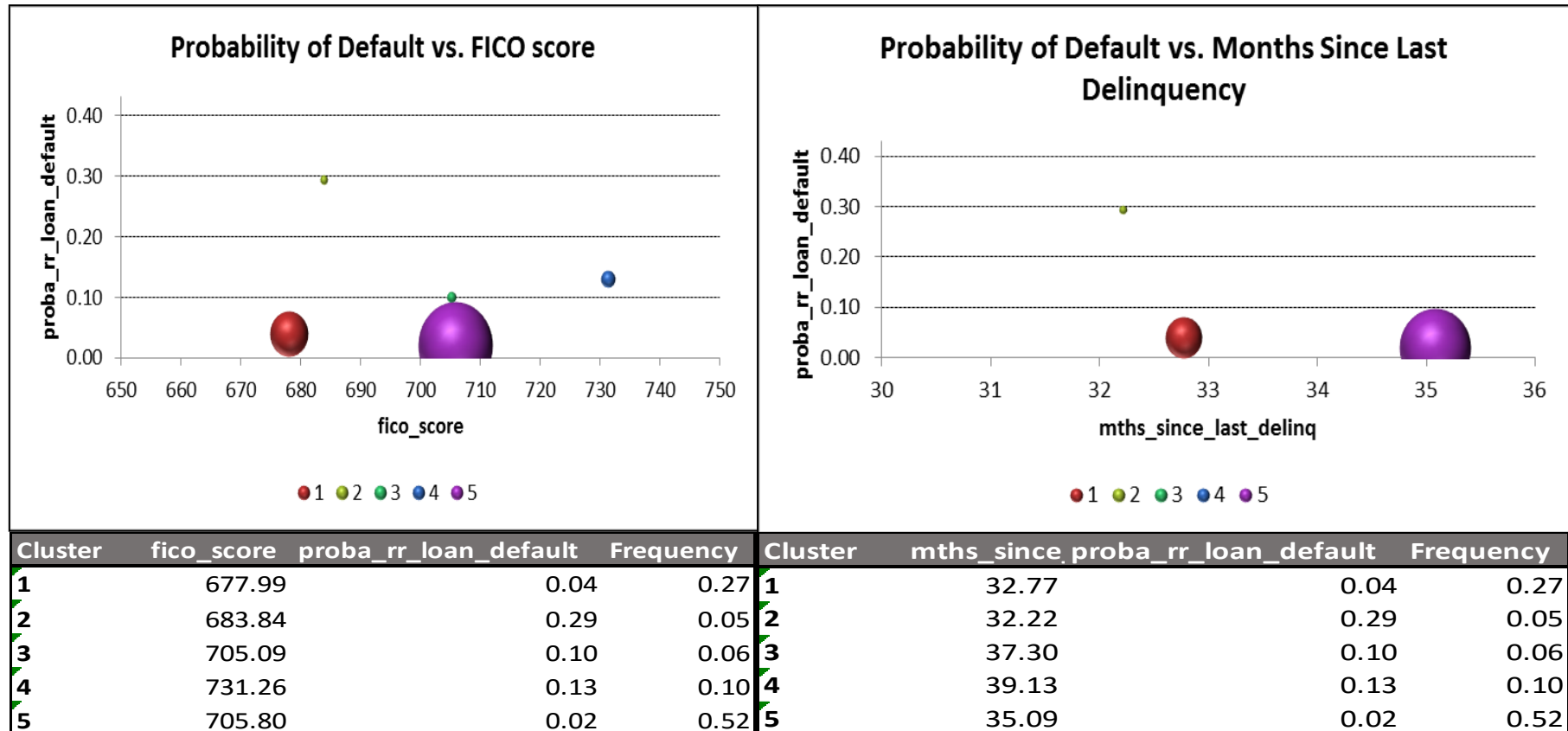


Cluster	dti	proba_rr_loan_default	Frequency	Cluster	annual_inc	proba_rr_loan_default	Frequency
1	17.84	0.04	0.27	1	70312.90	0.04	0.27
2	14.22	0.29	0.05	2	63172.30	0.29	0.05
3	16.31	0.10	0.06	3	68506.90	0.10	0.06
4	13.21	0.13	0.10	4	72297.70	0.13	0.10
5	17.08	0.02	0.52	5	75830.20	0.02	0.52

Data Discovery - Cluster Analysis Model 1..



Data Discovery - Cluster Analysis Model 1..





UCIRVINE | EXTENSION



Questions / Discussion