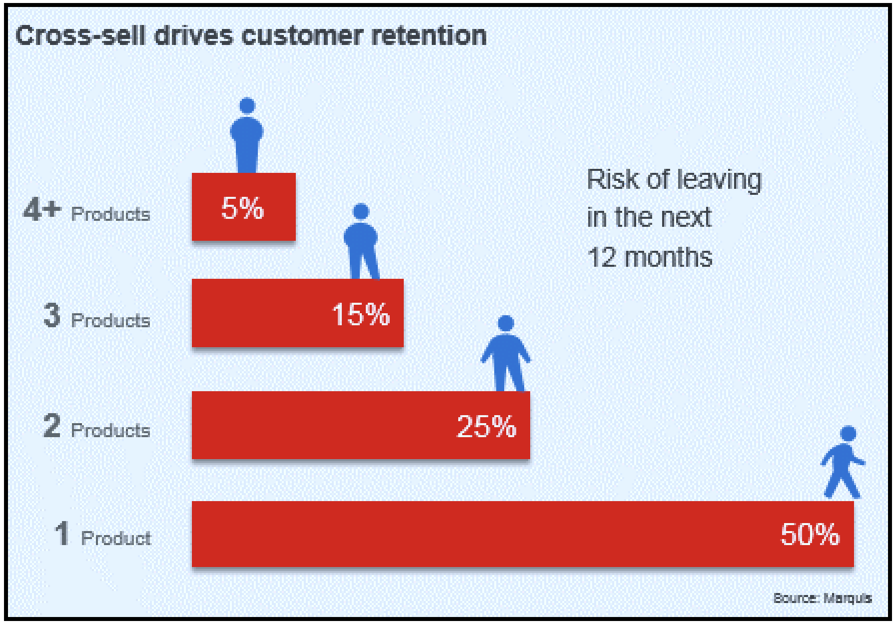
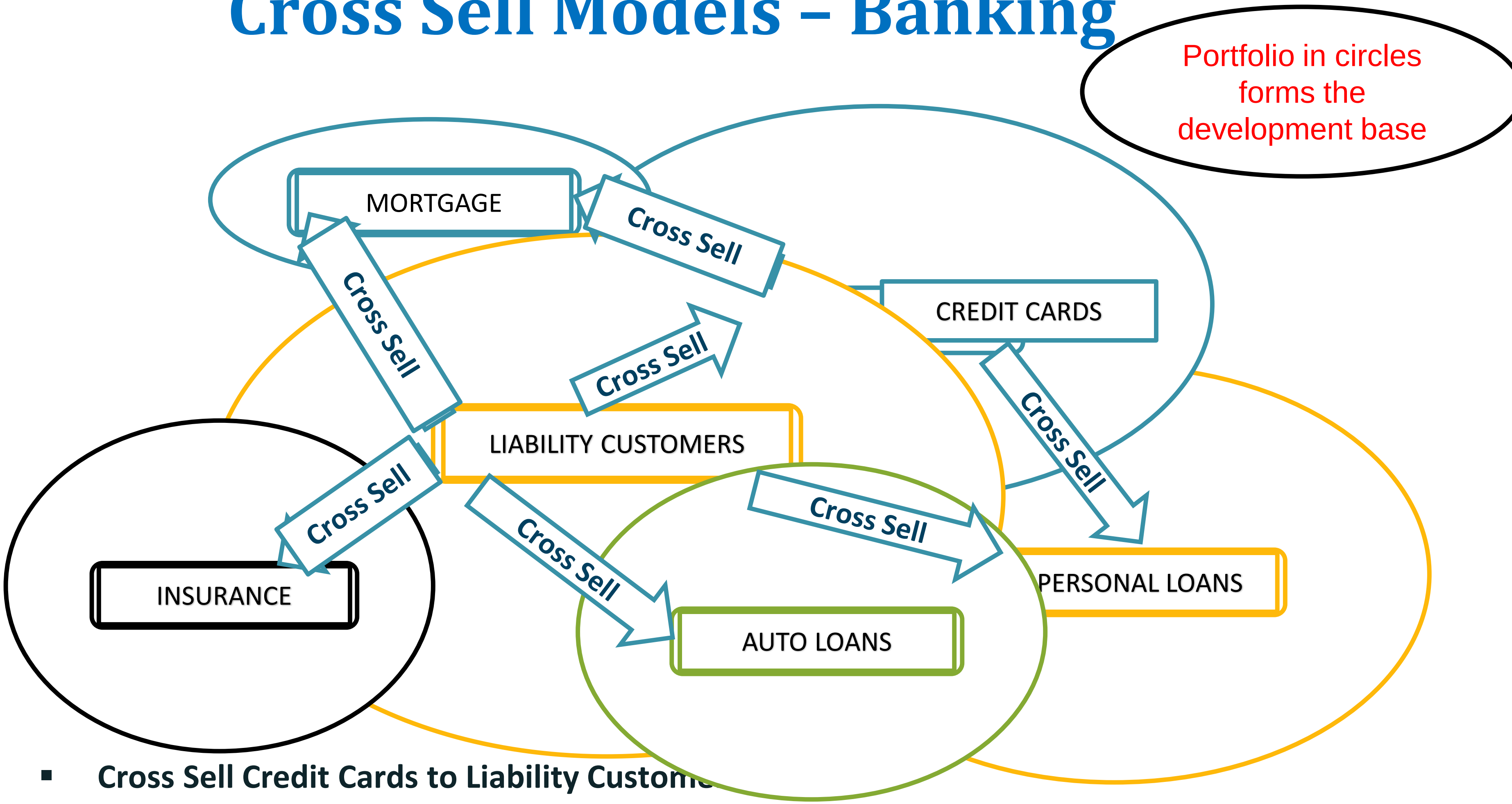
CRM



Module 4.2 – How to structure Propensity/Cross-sell/Look-alike/Attrition Models

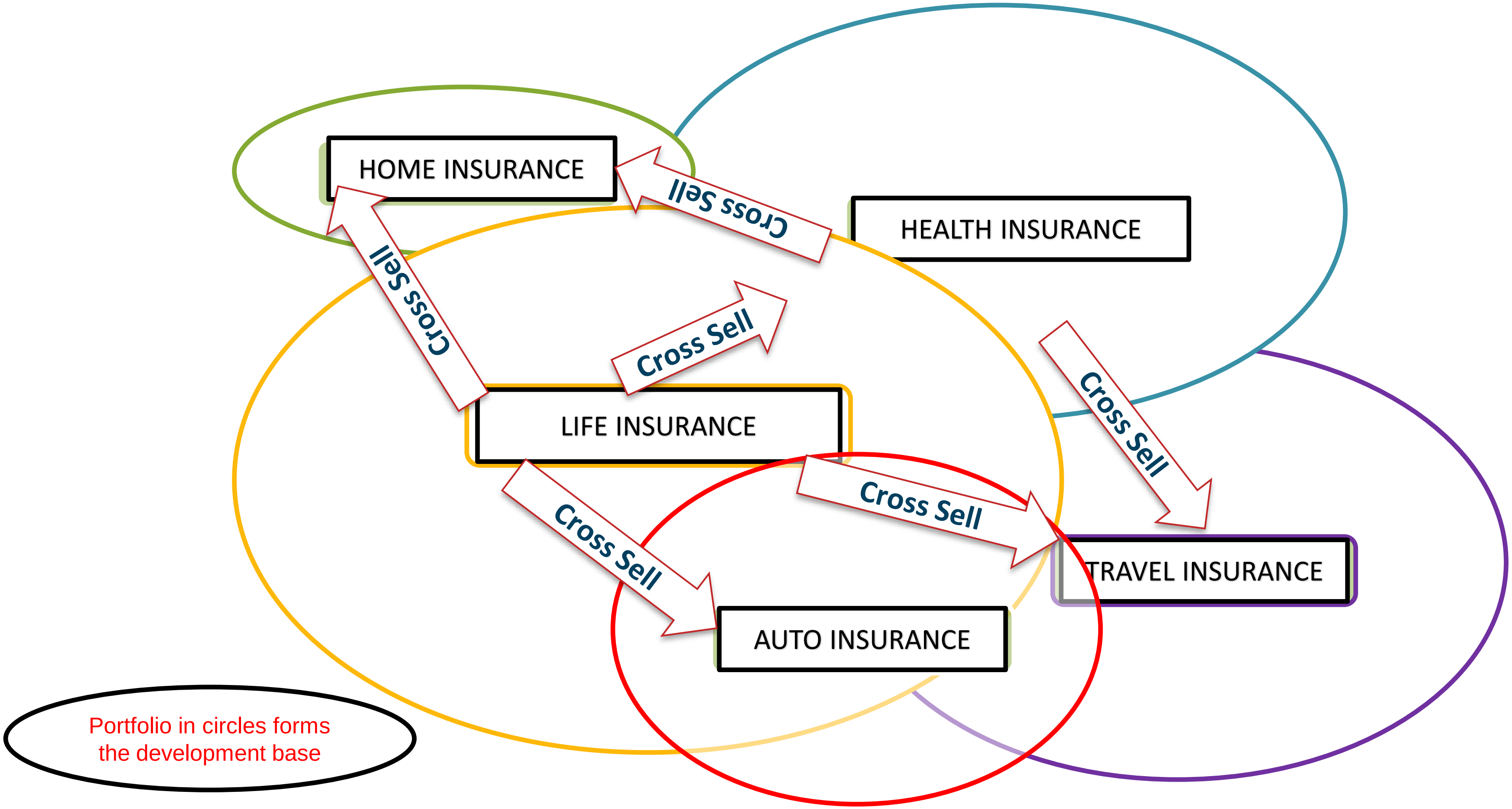
Dr Rita Chakravarti
Institute of Systems Science
National University of Singapore
Email: rita@nus.edu.sg

Cross Sell Models – Banking



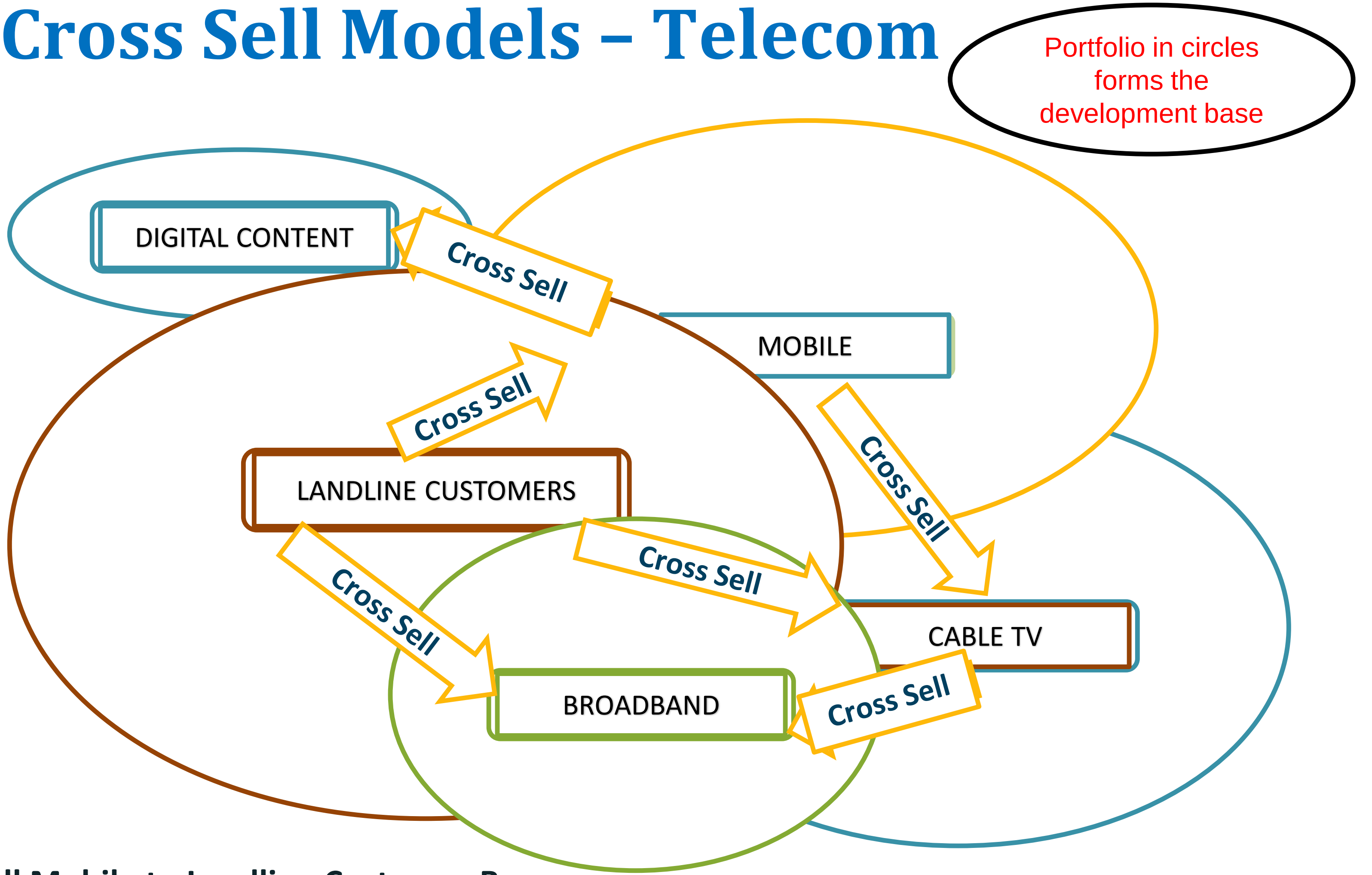
- Cross Sell Credit Cards to Liability Customers
- Cross Sell Mortgage to Credit Cards or Liability Customer Base
- Cross Sell Personal Loans to Credit Cards or Liability Customer Base
- Cross Sell Auto Loans to Liability Base
- Cross Sell Insurance to Liability Base

Cross Sell Models – Insurance



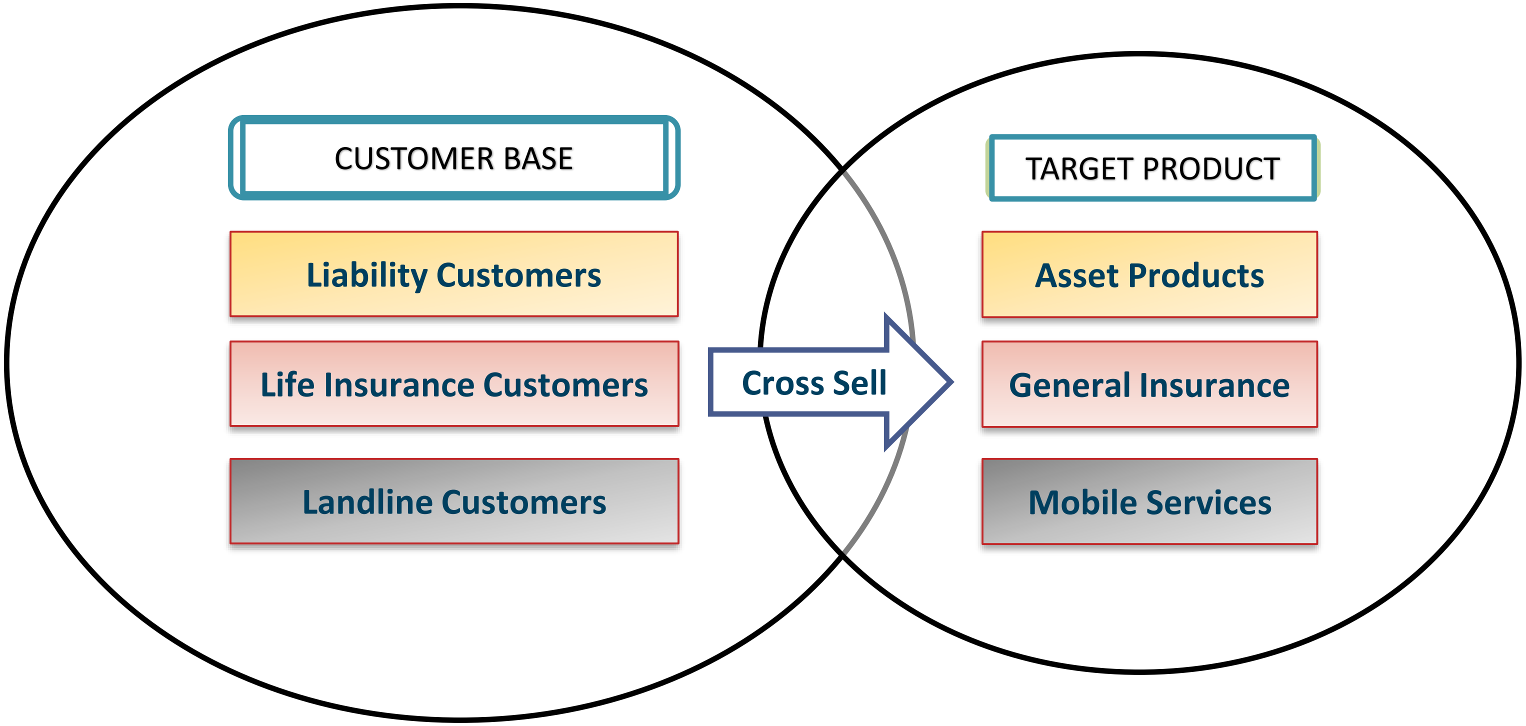
- **Cross Sell Health Insurance to Life Insurance Customers**
- **Cross Sell Home Insurance to Life or Health Insurance Customer Base**
- **Cross Sell Travel Insurance to Life or Health Insurance Customer Base**
- **Cross Sell Auto Insurance to Life Insurance Customer Base**

Cross Sell Models – Telecom



- Cross Sell Mobile to Landline Customer Base
- Cross Sell Digital Content to Mobile Customer Base
- Cross Sell Cable TV to Mobile or Landline Customer Base
- Cross Sell Broadband to Landline or Cable TV Customer Base

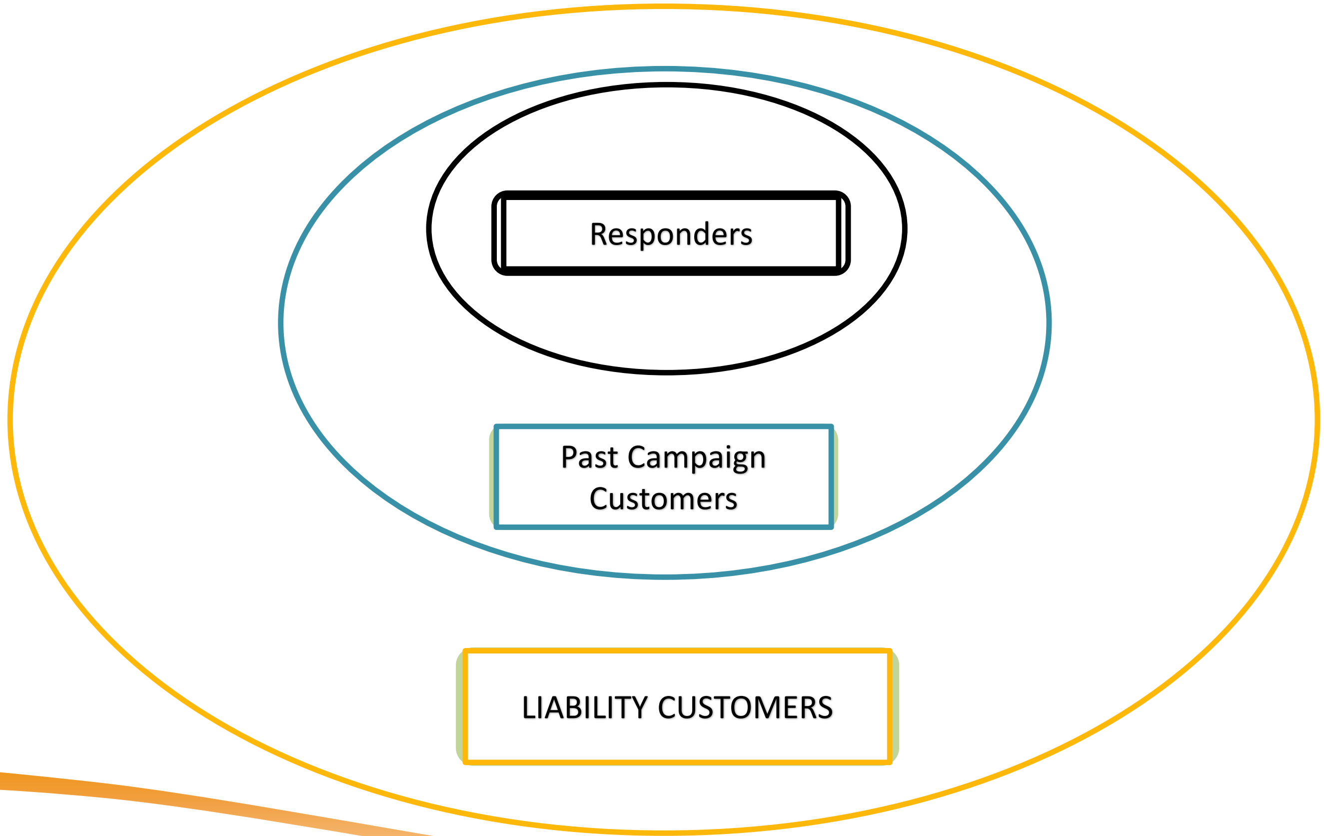
Cross Sell Models



- It is a Behavior Score
- Predict Product Purchase Inclination
- Higher Score implies higher propensity to purchase

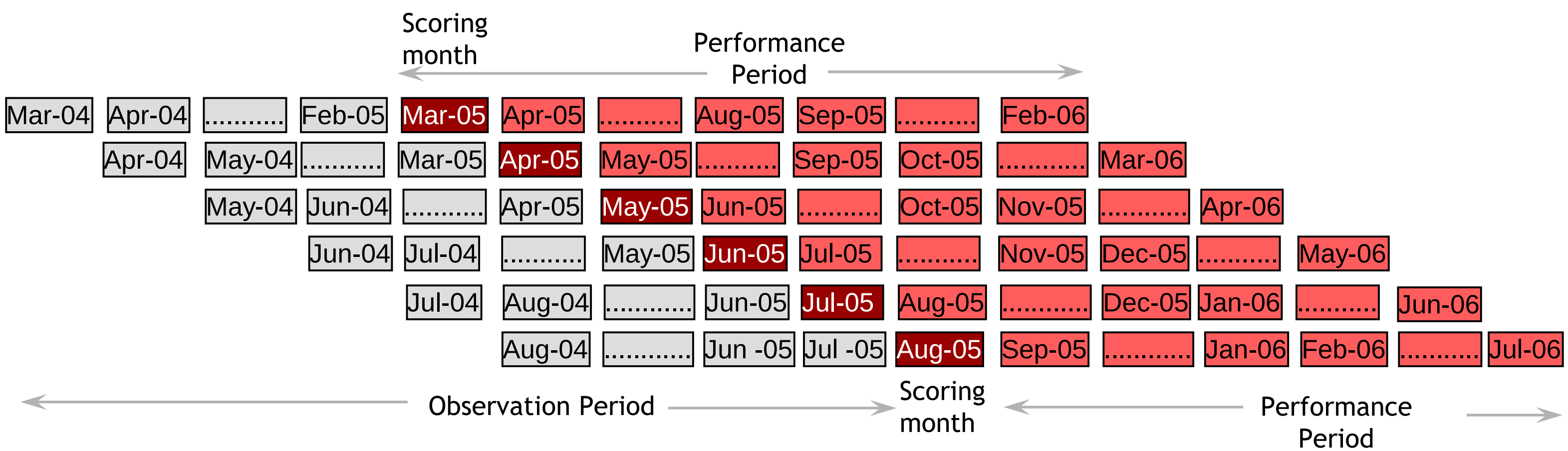
Cross Sell Models – Response Models

- Developed when Past Campaign Data is Available
- Past Campaign Population serves as Scorecard Development Sample
- Responders are the target (good) customers



Development Base & Target Variable Creation

Attrition/Purchase Happens During Performance Period



Applicable to : Anti-attrition/ Cross-sell/ Propensity Models

Multiple vintages/snapshots increases sample size and reduces seasonal effects

Sample Scorecard - Attrition Score – when not much behavior is available

S. No.	Variable Description	Weight	Condition	Interpretation
1	Gender	50	Female	Female customers have a higher likelihood of attrition.
		0	Male	
2	Education Level	50	College	Customers with college education level have a higher likelihood of attrition.
		40	University or higher	
		0	Others	
		30	Missing & Remaining Cases	
3	Age	60	≤ 24	The younger the customer, the higher the likelihood of attrition.
		40	25 – 30	
		20	31 – 40	
		0	> 40	
4	Card Type			
5	Bill-Mail-To Indicator			
6	Annual Income			

Sample Scorecard - Attrition Score – when Behavior is available

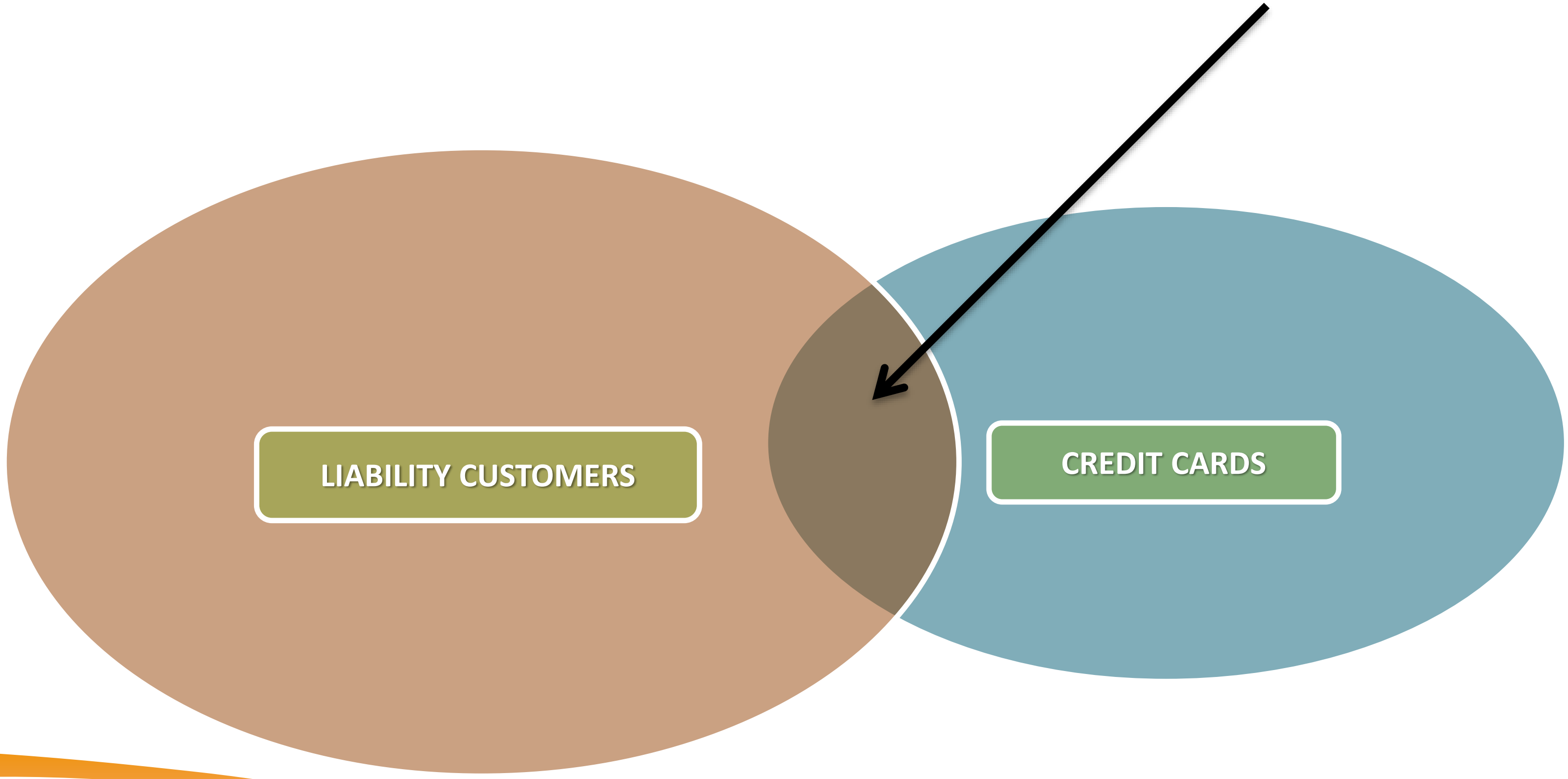
S. No.	Description	Weight	Condition	Interpretation
1	Age	64	≤ 30	The younger the customer, the higher the likelihood of attrition.
		29	31 – 33	
		0	> 34	
2	Month End Number of Local Retail Transactions	40	≤ 0	Accounts without transactions have a higher likelihood of attrition
		21	1	
		0	> 1	
3	Last 6 Months Average Number of Month End Local Cash Transactions			
4	Month End Number of Local Cash Transactions			
5	Month End Total Interest Amount			
6	Ratio of Month End Total Balance Amount to Credit Limit			
7	Marital Status			
8	Education Level			
9	Last 6 Months Average Ratio of Month End Total Amount to Credit Limit			
10	Gender			
11	Ratio of Month End Retail Balance to Credit Limit			
12	Last 3 Months Average Month End Total Number of Transactions			

Sample Scorecard - Mutual Fund X-sell Model

S. No.	Variable Description	Coeff	Condition	Interpretation
1	Number of open Gold cards.	0	0	Better if a customer holds a gold card.
		33	> 0	
2	Total cards avg financial charge over last 12 months.	191	0	Lower the better.
		162	0 < & ≤ 100	
		80	100 < & ≤ 190	
		0	> 190	
3	Customer income.			
4	Total bankcards avg credit limit over last 12 months.			
5	Occupation code			
6	Education code.			
7	Total bankcards average purchase amount over last 12 months			
8	Customer age as of scoring month.			
9	Customer marital status.			
10	Total number of open cards (bankcards and Diners) as of scoring month.			
11	Alignment constant			

Cross Sell Models –> Look Alike Models

- Developed when Past Campaign Data is not available
- Customer Profiling based on product overlap
- Customers holding both the products are the target (good) customers



Sample Scorecard - Look-a-like Model

S. No.	Variable Description	Coeff	Condition	Interpretation
1	Last 12 month's avg of interst / financial charges billed.	100	0	Lower the better.
		50	0 < & ≤ 5	
		0	> 5 or < 0	
2	Last 12 month's avg of bankcard credit limit.	0	≤ 6300	Higher the better.
		70	6300 < & ≤ 19M	
		100	> 19M	
3	Last 12 month's avg of total retail sales.	0	≤ 550	Higher the better.
		40	550 < & ≤ 1280	
		70	> 1280	
4	Last 12 month's avg of utilization.	65	0	Positive utilization of up to 10% is best; above 90% is neutral.
		115	0 < & ≤ 10	
		90	10 < & ≤ 25	
		55	25 < & ≤ 90	
		0	> 90 or < 0	
5	Occupation code	0	004, 007, 010, 012, 013, 016, 009	Unskilled workers are worst; Retirees are best.
		130	003, 005, 008, 011, 015, 017, 018, 023, 006	
		190	002	
		220	Otherwise.	
		350	020	
6	Residential type.	120	1	1: Own is best; 3: Buying is worst.
		80	Otherwise.	
		0	3	
7	Alignment constant	344	added for all instances	

Class Exercise

Scoring month, performance period

Define Dependent Variable

Good/Bad (Attrition Score)

Good

Bad

Good/Bad (Propensity/Cross Sell/Response Score)

Good

Bad

Good/Bad (Look-alike Model)

Good

Bad

Proper understanding of the business objective/ scenario is the key to develop a good working model

Key issues	Reason
Reason for Model Development Exercise	To understand the business need and decide on appropriate solution
History of past policy changes/marketing activity .Future plans in terms of policy/ marketing activity	To help design the model correctly and select appropriate variables
Past Model Usage Methodology	To understand whether there was any loop hole in the past model execution
Relevant MIS	To understand success target post model implementation
Existing Good/bad definition	To help decide new Good/bad definition

Key issues in data collection

Key issues	Reason
Availability of Various Data Tables	To understand the availability of the raw material for model development
Process of accessing the tables	To understand the effort required to prepare the database for model development
Merge Key	To understand the linking logic for various data tables
Data Dictionary	To interpret the data

There are five main steps to model development

Define Development, OOT validation window, Performance Period

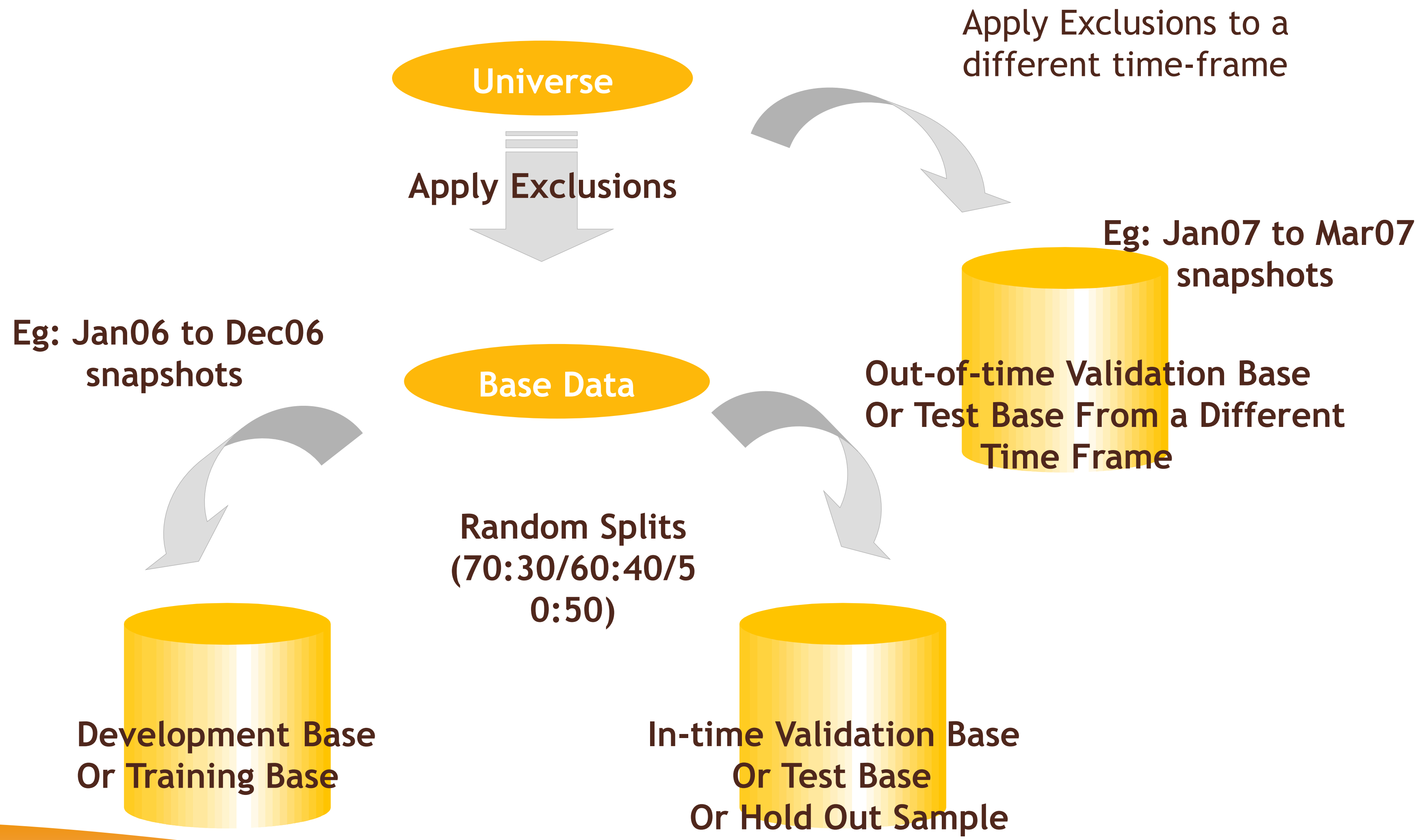
Development Base Creation

Split the data in Development & Holdout Sample Base & Create OOT Base; Define Dependant variable

Create Explanatory Variables for the model development

Build the Model;
Validate on holdout and OOT

STEP3 : Split the data in Development & Holdout Sample Base & Create OOT Base



Scorecards – Tools for Marketing Analytics

X-Sell Score/ Propensity Score

Population Take up Rate in a campaign 11.21%

Bins	Good Rate	% Customers
1	49.95%	2%
2	41.25%	1%
3	34.02%	2%
4	32.78%	2%
5	30.05%	1%
6	28.51%	1%
7	28.40%	2%
8	23.00%	2%
9	22.00%	3%
10	7.37%	84%

X Sell Score

Increasing Propensity

Decreasing Scores

Decreasing Propensity

Increasing Risk

Scorecards – Tools for Marketing Analytics – Raw MIS

Score Band	Total Customers	% Customers	Cum % Customers	Good Customers	% Good Customers	Cum% Good Customers	Bad Customers	Good Rate	Lift
> 620	13,506	2%	2%	6,746	8.9%	9%	6,760	49.95%	4.46
571-620	6,753	1%	3%	2,786	3.7%	13%	3,967	41.25%	3.68
511-570	13,506	2%	5%	4,595	6.1%	19%	8,911	34.02%	3.03
481-510	13,506	2%	7%	4,427	5.8%	25%	9,079	32.78%	2.92
466-480	6,753	1%	8%	2,029	2.7%	27%	4,724	30.05%	2.68
451-465	6,753	1%	9%	1,925	2.5%	30%	4,828	28.51%	2.54
441-450	13,506	2%	11%	3,836	5.1%	35%	9,670	28.40%	2.53
421-440	13,506	2%	13%	3,106	4.1%	39%	10,400	23.00%	2.05
391-420	20,259	3%	16%	4,457	5.9%	45%	15,802	22.00%	1.96
<=390	567,255	84%	100%	41,807	55.2%	100%	525,448	7.37%	0.66
Total	675,303	100%		75,714			599,589	11.21%	

Increasing Risk

Scorecards – Tools for Marketing Analytics

Portfolio Attrition Rate 5.28%

Population Avg Revenue 132.2

ATTRITION SCORE

Decreasing Scores → *Decreasing Attrition*

Bins	Bad Rate	% Customers
1	14.30%	10%
2	11.10%	10%
3	8.30%	10%
4	6.10%	10%
5	4.80%	10%
6	3.20%	10%
7	2.20%	10%
8	1.50%	10%
9	0.90%	10%
10	**0.40%**	10%

Decreasing Attrition Rates

Attrition Score

Revenue Score

Increasing Revenue

REVENUE SCORE

Bins	Revenue	% Customers
1	335	10%
2	280	10%
3	234	10%
4	185	10%
5	128	10%
6	70	10%
7	50	10%
8	30	10%
9	10	10%
10	0	10%

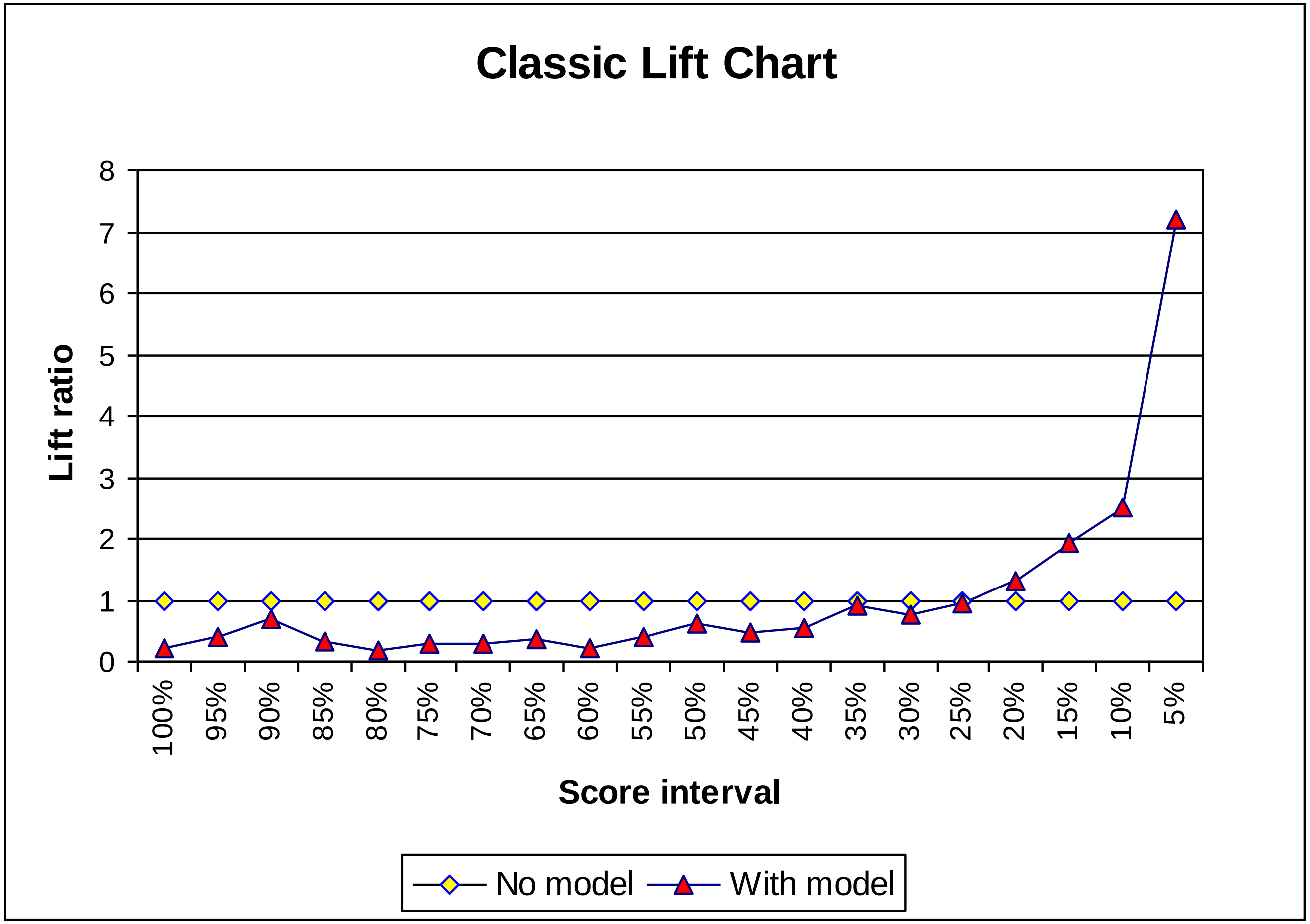
Decreasing Scores → *Decreasing Revenue*

MIS – Attrition

	Marginal		*Cumulative Distribution*			
Bin #	***Attrition %***	***Lift***	***Attrition %***	***Attrition Coverage***	***Total***	***Lift***
1	9.16%	5.54	9.16%	28%	5%	5.54
2	3.95%	2.39	6.54%	40%	10%	3.95
3	2.82%	1.70	5.30%	48%	15%	3.20
4	2.28%	1.38	4.57%	55%	20%	2.76
5	1.86%	1.12	4.00%	61%	25%	2.42
6	1.74%	1.05	3.65%	66%	30%	2.21
7	1.67%	1.01	3.36%	71%	35%	2.03
8	1.31%	0.79	3.11%	75%	40%	1.88
9	1.12%	0.68	2.88%	78%	45%	1.74
10	1.25%	0.75	2.72%	82%	50%	1.64
11	0.89%	0.54	2.55%	85%	55%	1.54
12	0.99%	0.60	2.42%	88%	60%	1.46
13	0.97%	0.59	2.31%	91%	65%	1.40
14	0.69%	0.42	2.19%	93%	70%	1.33
15	0.69%	0.41	2.09%	95%	75%	1.27
16	0.44%	0.27	1.99%	96%	80%	1.20
17	0.49%	0.30	1.90%	98%	85%	1.15
18	0.28%	0.17	1.81%	99%	90%	1.09
19	0.29%	0.18	1.73%	99%	95%	1.05
20	0.18%	0.11	**1.65%**	100%	100%	1.00

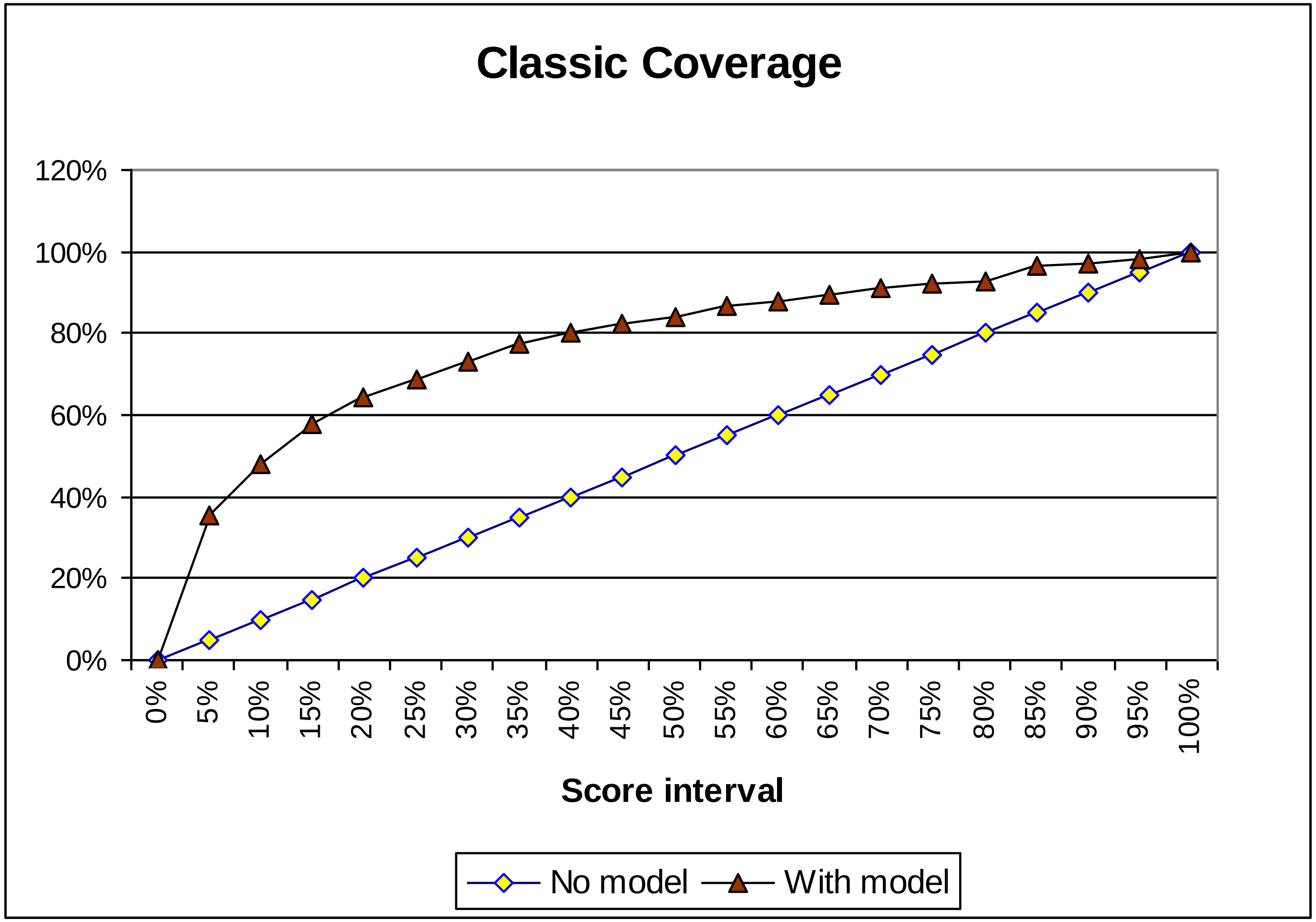
KS = 37

Lift Chart



Classic Lift Chart
Lift ratio
8
7
6
5
4
3
2
1
0
100%
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
Score interval
No model
With model

Coverage


Classic Coverage
120%
100%
80%
60%
40%
20%
0%
0%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
55%
60%
65%
70%
75%
80%
85%
90%
95%
100%
Score interval
No model
With model

Mutual Fund X-Sell Model: Two-way Score/Bin Distribution

1st row = Lift rate - Recurring payment mode   2nd row = Lift rate - Lump-Sum payment mode

BINLS / BINRSP	1	2	3	4	5	6	7	8	9	10	ROW RATE
1	4.8 5.6	5.0 2.5	6.3 1.4	4.1 0.5	3.6 0.5	3.8 0.9	1.5 0.1	2.1 –	2.4 –	– –	4.6 2.8
2	1.4 4.3	1.7 3.2	2.5 1.4	2.7 1.0	2.0 1.0	1.8 0.2	0.5 0.1	3.2 0.1	5.4 –	– –	2.0 1.9
3	0.3 4.5	0.9 2.1	1.3 1.5	0.9 1.0	0.4 0.4	0.3 0.1	2.5 0.1	0.6 0.2	3.6 –	– –	0.9 1.4
4	0.9 3.3	0.3 1.9	0.3 0.3	1.1 0.4	0.6 0.7	0.8 0.5	0.2 0.1	0.3 0.2	0.3 0.1	– –	0.6 0.8
5	0.8 2.3	0.8 1.5	– 1.2	0.3 0.4	0.9 0.7	0.9 0.4	0.6 0.1	0.3 0.3	1.1 0.1	1.1 –	0.7 0.6
6	– 4.3	0.9 2.2	0.5 1.1	– 0.6	0.5 0.5	0.2 0.6	0.2 0.0	0.3 0.1	– 0.2	0.7 0.2	0.3 0.7
7	0.4 3.4	– 2.1	0.9 1.3	0.6 0.4	0.5 0.4	0.4 0.1	– 0.1	0.1 0.2	0.4 –	– 0.1	0.3 0.5
8	– 3.2	1.2 1.3	0.4 1.1	0.9 1.6	0.2 0.2	0.4 0.1	0.4 0.2	0.1 0.1	0.2 –	0.1 0.0	0.3 0.3
9	– 8.0	– 1.5	– 1.3	0.4 0.3	– 0.6	0.2 0.8	0.8 0.5	0.4 0.2	0.3 0.1	0.1 0.1	0.2 0.5
10	– –	– 2.0	0.5 0.8	0.2 0.7	– 0.9	– 0.3	– 0.0	– –	0.3 0.2	– –	0.1 0.4
COLUMN RATE	2.2 4.5	1.8 2.2	1.6 1.2	1.1 0.7	0.8 0.6	0.7 0.4	0.5 0.1	0.4 0.2	0.6 0.1	0.1 0.1	1.0 1.0

X-sell MF through combined mode
X-sell MF through Lump-Sum
X-sell Mf through RSP mode

Mutual Fund X-Sell Model: Design

Segment 1 : Card Only

Segment 2 : Card + Banking

Why ?? : The take-up rate of the Mutual Fund among the card only customers and card + banking customers are entirely different.

	Observation Period											Score	Performance period			
Date yymm	9801	9803	9804	9805	9806	9807	9808	9809	9810	9811	9812	9901	9902	9903	9904	9905
Month index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

12 months' historical data (inclusive of scoring month) is used to develop the models

Total 4 models : Recurring Payment Mode & Lumpsum Payment Mode for 2 segments

Dependent Variable Creation

Take-up group: Customers who subscribe to at least one mutual fund through payment mode X during the performance period.

Non-Take-up group: customers who do not subscribe to any mutual fund during the performance period.

Indeterminate: customers who subscribe to mutual fund through payment different from payment method X during the performance period.

Mutual Fund X-Sell Model: Data Set

Sample Description

Scoring month 12 : January 1999 - October 1999 (inclusive)

Distribution of Take-up and Non-Take-up

	Segment1-RSP			Segment1-Lump Sum		
	Sample Size	Weighted Size	Weighted %	Sample Size	Weighted Size	Weighted %
Non take-up	10000	4909147	99.98%	10000	4909147	99.95%
Take-up	837	837	0.02%	2348	2348	0.05%
Total	10837	4909984		12348	4911495	

	Segment2-RSP			Segment2-Lump Sum		
	Sample Size	Weighted Size	Weighted %	Sample Size	Weighted Size	Weighted %
Non take-up	10000	463218	99.76%	10000	463218	99.01%
Take-up	1123	1123	0.24%	4650	4650	0.99%
Total	11123	464341		14650	467868	

Note: Segments 1 and 2 had 312 and 414 customers respectively who had subscribed to Mutual Fund through combined RSP and Lump-Sum payment mode. These customers were excluded at the model development stage to eliminate noise from the data but included when the profiling and characteristic analysis were performed and model performance were measured.

Mutual Fund X-Sell Model: Data Set

Distribution of Take-up and Non-Take-up for the Development and Validation Samples

	Segment1		Segment2	
	RSP	LS	RSP	LS
Development				
Non take-up	7507	7502	7503	7501
Take-up	621	1759	839	3486
Total	8128	9261	8342	10987
In-Time Validation				
Non take-up	2493	2498	2497	2499
Take-up	216	589	284	1164
Total	2709	3087	2781	3663

Modeling Technique

Logistic regression technique was used to develop the models for the two segments X two payment modes

Recall Regression & Logistic Regression

Linear Regression : $Y = a + b_1X_1 + b_2X_2 \ldots + b_kX_k + \epsilon$ where the error term ϵ is i.i.d Normal N $(0, \sigma^2)$

Logistic Regression : P (Y = 1) = *p*, P (Y = 0) = 1- *p*

- where $p(X) = \dfrac{\exp^{a + b_1X_1 + b_2X_2 \ldots + b_kX_k}}{1 + \exp^{a + b_1X_1 + b_2X_2 \ldots + b_kX_k}}$ *and* $X_1, X_2, \ldots X_k$ are predictor variable which can be either numerical or categorical.

Logistic regression is used for the prediction of the probability of occurrence of an event by fitting data to a **logistic** curve

It is classified as log-linear model as by doing the following transformation (logit) it becomes a linear model :

- $\log\left(\frac{p}{1-p}\right) = a + b_1X_1 + b_2X_2 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + b_kX_k$

Why Logistic

Regression methods have become an integral component of any data analysis concerned with the relationship between a dependent variable (Y) and one or more explanatory/independent variables (X1, X2…..) – however Y is typically a continuous variable and the error term is normally distributed with constant error variance

It is often the case that the outcome variable(Y) is discrete taking 2 or more possible values

Over the last 20 years the logistic regression model has become, in many fields, the standard method of analysis in this situation

Special Problems when Outcome (Y) variable is binary or dichotomous

1. Non-normal Error Terms

2. Non-constant Error Variance

3. Constraints on response function

One Of The 4 Models (Partial)

Card-only segment (MF through RSP payment mode)

Higher the weight → Higher probability of MF take-up

	Variables	Weights	Interpretation
1	**Profession**		
	Manager, Owner/Chairman	0	
	General staff	36	
	Others	38	
	Senior staff, Government employees, Educator	43	
	Executive, Shareholder, M.D., Engineer	137	
2	**Age**		**Younger, the better**
	>=21-<40	99	
	>=40	0	
3	**Behavior score**		**Higher, the better**
	<=31	0	
	>500-<=653	38	
	>653-<=658	63	
	>658	111	
4	**Latest 12 months' average Cycle Balance Utilization**		**Best in the range 7.8%-10%**
	<0	48	
	>=0-<=0.5%	0	
	>0.5%-<=5%	48	
	>5%-<=7.8%	54	
	>7.8%-<=10%	60	
	>10%-<=17%	56	
	>17%	34	
5	**Latest 12 months' average Credit limit**		**Best in the range $148M-$431M**
	<=$148M	0	
	>$148M-<=$431M	61	
	>$431M	38	
6	**Gold card indicator**		**Having Gold Card, better**
	No Gold Card	0	
	Has Gold Card	100	
7	**Income**		**Best in the range $330M-$800M**
	<=$279M	1	
	>$279M-<=$330M	60	
	>$330M-<=$800M	76	
	>$800M	0	
8	**Gender**		**Female, better**
	Female	49	
	Male	0	
9	**Latest 12 months' average Retail Sales**		**Higher, the better**
	<0	17	
	>=0-<=$7M	0	
	>$7M-<=$9.2M	12	
	>$9.2M	29	

Sample MIS for Strategy Formulation - Score/Bin Distribution – Card only customers – MF with Recurring Payment model - Development

Development																
	Marginal						Cumulative Distribution									
Bin #	Take-up% Combined	Take-up% RSP	Take-up% Lump Sum	Lift Combined	Lift RSP	Lift Lump Sum	Take-up% Combined	Take-up% RSP	Take-up% Lump Sum	Take-up Coverage Combined	Take-up Coverage RSP	Take-up Coverage Lump Sum	Total	Lift Combined	Lift RSP	Lift Lump Sum
1	0.023%	0.109%	0.175%	3.69	6.41	3.65	0.023%	0.109%	0.175%	18.5%	32.1%	18.3%	5%	3.7	6.4	3.7
2	0.018%	0.049%	0.098%	2.91	2.89	2.05	0.021%	0.079%	0.136%	33.3%	46.8%	28.8%	10%	3.3	4.6	2.8
3	0.017%	0.044%	0.110%	2.73	2.56	2.30	0.020%	0.067%	0.127%	46.8%	59.5%	40.2%	15%	3.1	4.0	2.7
4	0.012%	0.026%	0.075%	1.83	1.51	1.58	0.018%	0.057%	0.115%	55.9%	67.0%	47.9%	20%	2.8	3.3	2.4
5	0.008%	0.014%	0.056%	1.22	0.80	1.17	0.016%	0.048%	0.103%	62.2%	71.1%	54.0%	25%	2.5	2.8	2.1
6	0.008%	0.017%	0.076%	1.23	0.99	1.60	0.014%	0.043%	0.098%	68.5%	76.2%	62.2%	30%	2.3	2.5	2.1
7	0.009%	0.009%	0.049%	1.35	0.54	1.01	0.014%	0.038%	0.092%	74.8%	78.7%	66.9%	35%	2.1	2.3	1.9
8	0.005%	0.010%	0.032%	0.75	0.56	0.68	0.013%	0.035%	0.084%	78.4%	81.4%	70.2%	40%	2.0	2.0	1.8
9	0.006%	0.010%	0.035%	0.88	0.59	0.73	0.012%	0.032%	0.079%	82.9%	84.4%	73.9%	45%	1.8	1.9	1.6
10	0.003%	0.013%	0.024%	0.42	0.77	0.51	0.011%	0.030%	0.074%	84.7%	87.8%	76.1%	49%	1.7	1.8	1.5
11	0.001%	0.006%	0.036%	0.08	0.33	0.74	0.010%	0.028%	0.070%	85.1%	89.7%	80.4%	55%	1.6	1.6	1.5
12	0.002%	0.004%	0.030%	0.36	0.25	0.62	0.009%	0.026%	0.067%	86.9%	91.0%	83.5%	60%	1.5	1.5	1.4
13	0.002%	0.005%	0.020%	0.25	0.27	0.41	0.009%	0.024%	0.063%	88.3%	92.4%	85.7%	65%	1.4	1.4	1.3
14	0.003%	0.005%	0.025%	0.49	0.31	0.53	0.008%	0.023%	0.060%	90.5%	93.8%	88.1%	70%	1.3	1.3	1.3
15	0.002%	0.003%	0.009%	0.25	0.15	0.19	0.008%	0.021%	0.057%	91.9%	94.6%	89.1%	75%	1.2	1.3	1.2
16	0.002%	0.007%	0.022%	0.36	0.41	0.46	0.007%	0.021%	0.055%	93.7%	96.7%	91.4%	80%	1.2	1.2	1.1
17	0.001%	0.004%	0.017%	0.18	0.23	0.35	0.007%	0.020%	0.052%	94.6%	97.8%	93.1%	85%	1.1	1.2	1.1
18	0.005%	0.004%	0.028%	0.80	0.25	0.59	0.007%	0.019%	0.051%	98.7%	99.0%	96.1%	90%	1.1	1.1	1.1
19	0.001%	0.001%	0.019%	0.18	0.06	0.40	0.007%	0.018%	0.049%	99.6%	99.4%	98.1%	95%	1.0	1.0	1.0
20	0.001%	0.002%	0.018%	0.09	0.13	0.38	0.006%	0.017%	0.048%	100.0%	100.0%	100.0%	100%	1.0	1.0	1.0

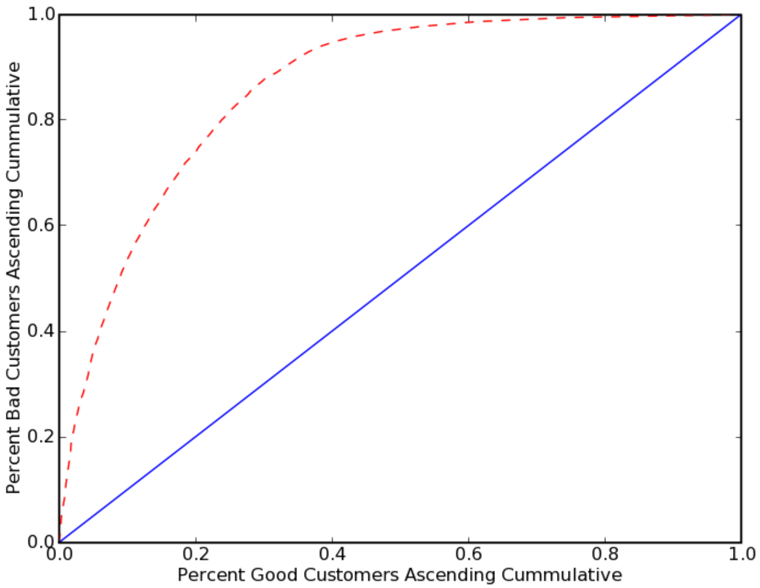
Sample MIS for Strategy Formulation - Score/Bin Distribution – Card only customers – MF with Recurring Payment model - Validation

Validation

	Marginal						Cumulative Distribution									
Bin #	Take-up% Combined	Take-up% RSP	Take-up% Lump Sum	Lift Combined	Lift RSP	Lift Lump Sum	Take-up% Combined	Take-up% RSP	Take-up% Lump Sum	Take-up Coverage Combined	Take-up Coverage RSP	Take-up Coverage Lump Sum	Total	Lift Combined	Lift RSP	Lift Lump Sum
1	0.029%	0.127%	0.177%	4.58	7.47	3.69	0.029%	0.127%	0.177%	22.2%	36.2%	17.9%	5%	4.6	7.5	3.7
2	0.013%	0.035%	0.099%	1.97	2.06	2.07	0.020%	0.078%	0.135%	33.3%	47.8%	29.6%	10%	3.2	4.6	2.8
3	0.011%	0.034%	0.074%	1.69	1.99	1.54	0.017%	0.061%	0.111%	44.4%	60.9%	39.7%	17%	2.6	3.6	2.3
4	0.015%	0.034%	0.095%	2.44	2.01	1.98	0.016%	0.055%	0.108%	55.6%	70.0%	48.7%	22%	2.6	3.2	2.3
5	0.004%	0.018%	0.054%	0.65	1.04	1.12	0.014%	0.048%	0.097%	58.9%	75.4%	54.4%	27%	2.2	2.8	2.0
6	0.009%	0.012%	0.064%	1.36	0.68	1.34	0.013%	0.042%	0.091%	66.7%	79.2%	62.1%	32%	2.1	2.4	1.9
7	0.010%	0.012%	0.058%	1.60	0.70	1.21	0.013%	0.038%	0.088%	73.3%	82.1%	67.1%	37%	2.0	2.2	1.8
8	0.009%	0.014%	0.029%	1.38	0.80	0.61	0.012%	0.035%	0.081%	80.0%	86.0%	70.1%	41%	1.9	2.1	1.7
9	0.003%	0.009%	0.047%	0.50	0.54	0.98	0.011%	0.033%	0.078%	82.2%	88.4%	74.4%	46%	1.8	1.9	1.6
10	0.003%	0.008%	0.024%	0.53	0.46	0.50	0.011%	0.031%	0.073%	84.4%	90.3%	76.5%	50%	1.7	1.8	1.5
11	0.000%	0.008%	0.024%	-	0.44	0.51	0.010%	0.028%	0.068%	84.4%	92.8%	79.3%	55%	1.5	1.7	1.4
12	0.004%	0.005%	0.042%	0.70	0.30	0.87	0.009%	0.027%	0.066%	87.8%	94.2%	83.5%	60%	1.5	1.6	1.4
13	0.001%	0.006%	0.015%	0.22	0.37	0.30	0.009%	0.025%	0.062%	88.9%	96.1%	85.0%	65%	1.4	1.5	1.3
14	0.003%	0.002%	0.018%	0.48	0.10	0.37	0.008%	0.023%	0.059%	91.1%	96.6%	86.8%	70%	1.3	1.4	1.2
15	0.000%	0.004%	0.018%	-	0.23	0.37	0.008%	0.022%	0.057%	91.1%	97.6%	88.3%	74%	1.2	1.3	1.2
16	0.003%	0.002%	0.026%	0.47	0.10	0.55	0.008%	0.021%	0.055%	93.3%	98.1%	91.0%	79%	1.2	1.2	1.2
17	0.001%	0.003%	0.015%	0.22	0.19	0.31	0.007%	0.020%	0.053%	94.4%	99.0%	92.5%	84%	1.1	1.2	1.1
18	0.001%	0.003%	0.033%	0.22	0.19	0.68	0.007%	0.019%	0.051%	95.6%	100.0%	96.0%	89%	1.1	1.1	1.1
19	0.000%	0.000%	0.018%	-	-	0.38	0.006%	0.018%	0.050%	95.6%	100.0%	98.1%	95%	1.0	1.1	1.0
20	0.005%	0.000%	0.017%	0.82	-	0.35	0.006%	0.017%	0.048%	100.0%	100.0%	100.0%	100%	1.0	1.0	1.0

Performance Reports

- Receiver Operator Characteristic (ROC) Curve for each model
- Divergence, KS and ROC stats for Both Train and Test Datasets



	Train			Test		
Model	**Div**	**ROC**	**KS**	**Div**	**ROC**	**KS**
1	1.643	0.862	0.577	1.655	0.863	0.577
2	1.055	0.845	0.542	1.049	0.845	0.54
3	0.969	0.876	0.596	0.986	0.877	0.601
4	1.221	0.853	0.557	1.251	0.857	0.568
5	0.961	0.977	0.852	0.987	0.977	0.855

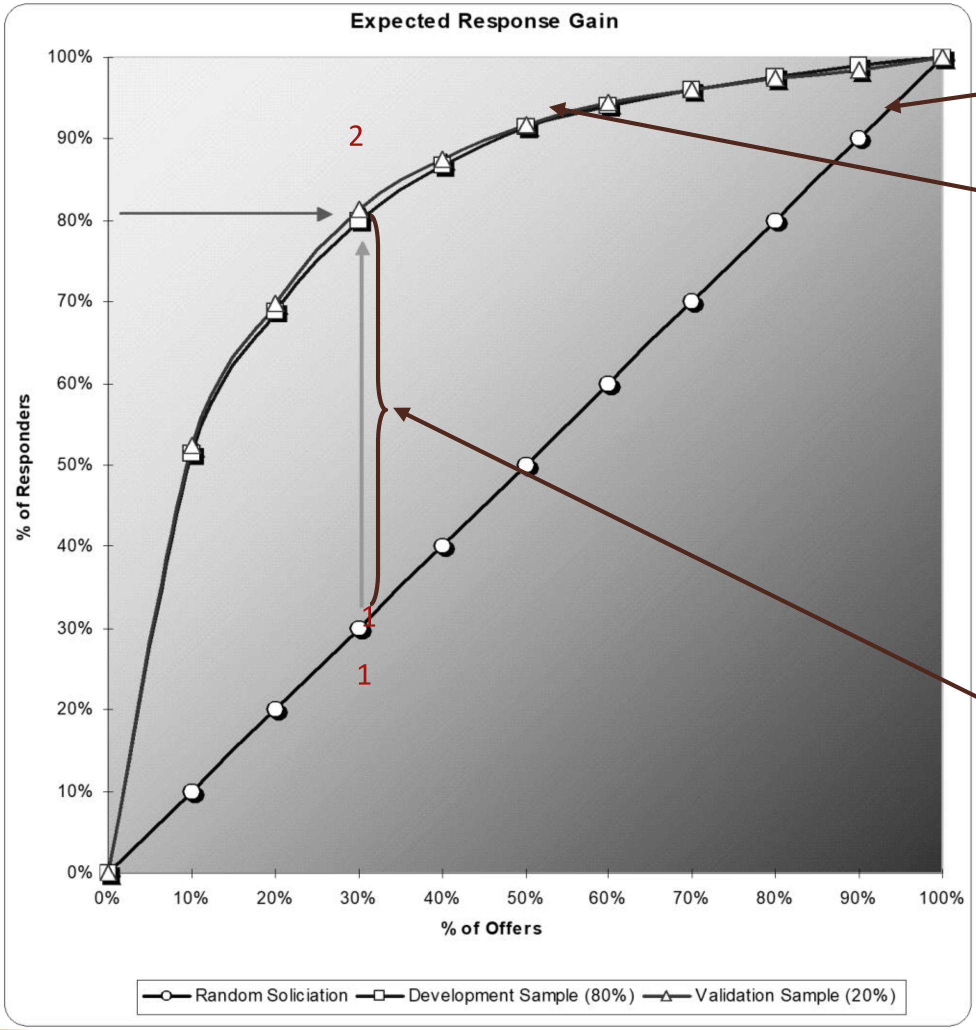
Cross Sell Models Success Stories

Western World

- At the 3rd Decile, the model captured over 66% more applicants when compared to the random selection.
- The highest scoring 10% of the population has an application rate of 340% greater than average.
- Over 75 percent of the applicants on the development file are identified at a 40% mailing depth
- Predictive models provide a 25% performance lift per campaign vs. other selection means and 60-70% lift for marginal buyers
- Response lift of 70% for a cross-sell program conducted for a Financial Services portfolio
- Response lift 70% for a Retail chain, where models for prospects were based on existing customer behaviours and combined with improved messaging

Look-Alike Model Results

Increase in Precision for Identifying Responders



Response expected by not applying the model (Black Diagonal line)

Response expected by applying model (Curved line)

Selecting a random 30% of the list, we would find 30% of total responders. With the model, selecting the top 30% (Top 3 Deciles), we expect to reach over 80% of the total responder population (point 2).

Using the model to select 30% of the list, we reach **170% more responders** than random selection.

Response Model Case Study: American Software Giant

Business Problem:

- The dominant player in the software industry, with tens of billions of dollars in sales through multiple channels
- Their divisions target specific customer groups or support field marketing efforts
- Create statistical models that will increase customer intelligence, improve product targeting, increase attendance at compnay events and enhance customer contact strategy

Solution:

- Develop event promotion response models for distinct segments
- Develop nine response models for product re-launch
- Conducted focus group and quantitative survey research to understand customer attitudes and opt-out behavior
- Customer segmentation and valuation for multiple group

Further Sophistication With Propensity Models

- TTE Modeling : Concept Borrowed from Survival Analysis

- TTE Modeling Recognizes :
  - Importance of Time
  - That your chance of default/churn/attrite/upsell depends not only on your attributes but also your tenure or your position in the typical customer life cycle

Leveraging The Propensity Model Concept : What would you do if you could anticipate how customers will use their credit card?

Predict who will … in the near future	Action	Benefit
Make a $4000+ balance transfer		
Spend $500+ on car repair expenses, large deposit in checking account		
Start making large home improvement expenses		
Stop using card for petrol purchases		
Start making large college related expenses (life change?)		

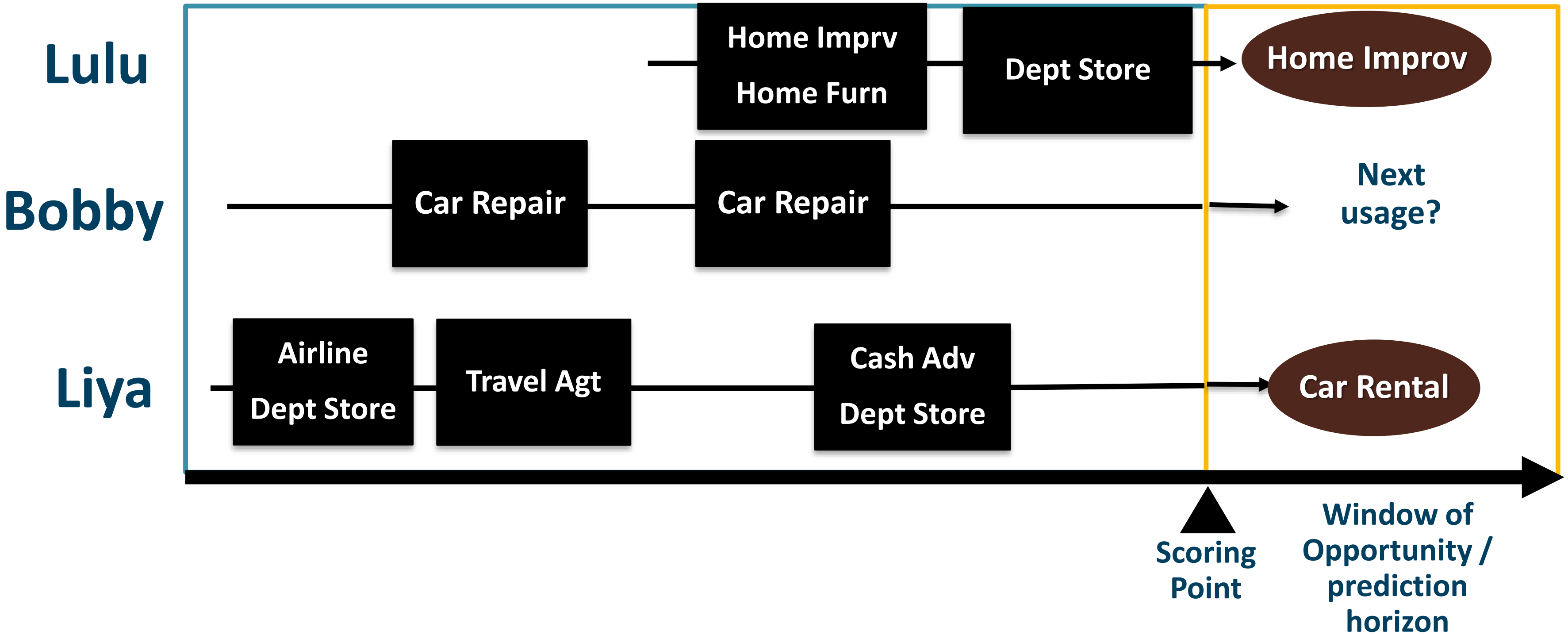
Offer Next Best Product ?

Predict who will … in the near future	Action	Benefit
Make a $4000+ balance transfer	Increase credit limit. Offer incentives for larger balance transfer	Grow revenues
Spend $500+ on car repair expenses, large deposit in checking account	Offer auto loan Offer auto insurance (if loan is also taken)	Cross-sell lender products
Start making large home improvement expenses	Offer 6 months no-interest on home improvement expenses. Offer incentives for higher spending at a partner retailer	Keep customers from taking the retailer's financing Grow revenues Strengthen partnerships
Stop using card for petrol purchases	Offer cash-back on petrol purchases	Minimize merchant category-level attrition
Start making large college related expenses (life change?)	Offer credit card for students	Hook the next generation users early

Answer : Time-to-Event (TTE) Modeling

- **Current environment:**
  - Competition for wallet share is intense and will further intensify given the current market volatility
  - Requires highly accurate estimations of near term customer usage patterns
    - Informed timely marketing opportunities
    - Fine tuned account management decision making
- **Challenge:**
  - Customer usage patterns are highly dynamic (“moving target”)
  - Understand multi-dimensional intentions of a consumer
- **Solution: Time-to-Event Modeling**
  - Predicts hundreds of actionable future events
  - Helps identify a seller’s best next action (who, what and when)
    - Cross sell products
    - Make merchant offers

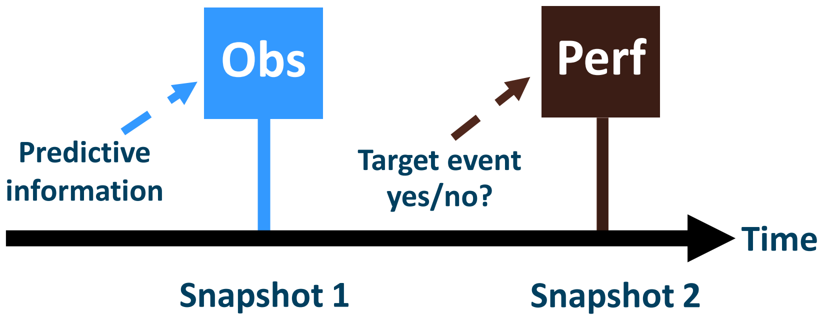
Sophistication of Use of Cross Sell Models

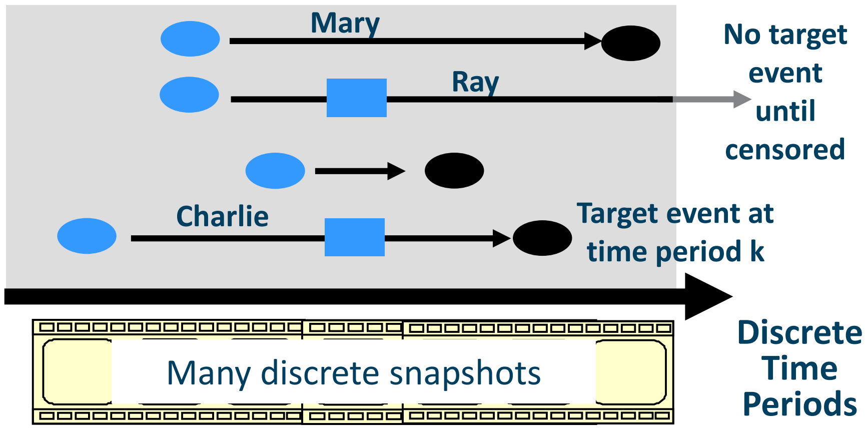


- Determine events of interest, flexible definitions possible (e.g. home improvement purchase over $200).
- Build scorecard for each event to predict the propensity of customers to experience that event in the near future (e.g. 2 months later)
- Make short-term predictions but update them frequently

Why Time To Event Scorecards??

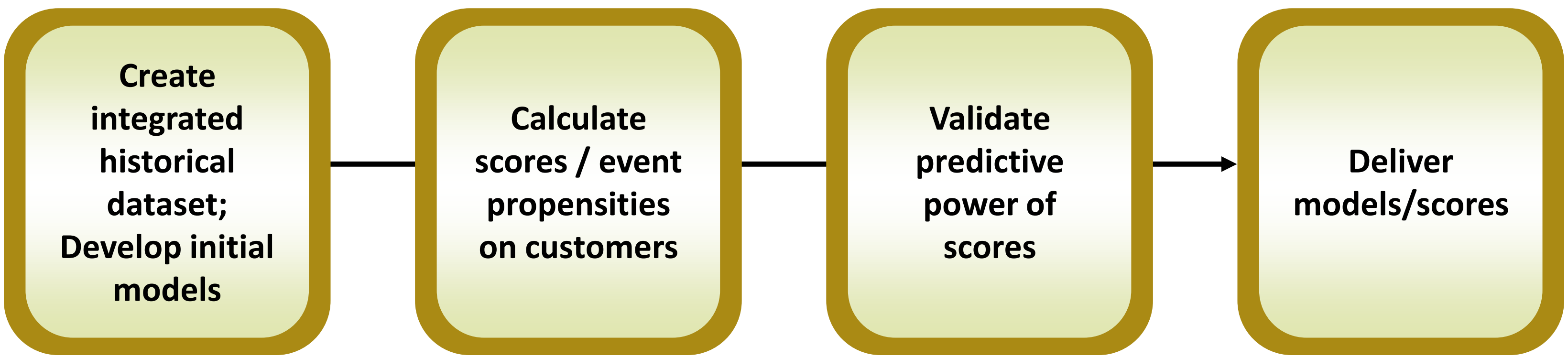
Traditional models A cross-sectional data analysis paradigm	Time-to-Event Scorecards A longitudinal data analysis paradigm
» **Focus is on *whether* event will happen, not *when*** » Predictive information is **summarized** at observation point. » Predictive info between obs point and perf is ignored » Fixed Scorecard Building Period	» **Focus is on *when event will happen*** » Predictive information is exploitable right up to the time period prior to the target event or until right censoring » Moving Time Window





Modeling Objective

- Create customer and event-level propensity scores from transaction history, account summary and demographic data
    - Predict events in a short period of time in future (performance period)
- Sophisticated model development methodology for powerful empirical models



Modeling Events

- Types of Events Covered
  - Purchase of a coupon offer
  - Utilization level of credit card in near future
  - Cross sell of different products
- Coupon Offers Based on mapping of items to higher aggregation
- Modeling at a pre-defined aggregation level of events for merchant offers
  - Many similar transaction activities grouped in merchant categories
  - Benefits of this type of aggregation
  - Reduces data fragmentation
    - E.g., one model for all types of departmental stores
  - Increases model robustness and signal pickup in the model

Real Time Scoring for Making Coupon Offers

- Profile maturation at the time of deployment
  - Compute characteristics of each customer – called profile
  - Store the profiles in a database
- Incremental updates in real time
  - When a new transaction is received:
  - Fetch the stored profile for the customer
  - Use new transaction to update characteristics in the profile
  - Store the updated profile in the database replacing the stale one
- Scoring in real time
  - Use updated profile to score through the TTE models
  - Make relevant offers

Score File for Targeted Decisioning

TTE Scorecards

Score File Allows Targeted Decisioning

Customer-Event Propensity Matrix

Events

Customers

	Pr {Life Insurance}	Pr {Home Loan}	. . . . .	Pr {Bal Transfer}
James	0.0158	0.6564	0.2067	0.0439
Charlie	0.0971	0.0363	0.0329	0.0382
Rachel	0.0957	0.8499	0.2773	0.0766
Drew	0.0485	0.0934	0.0467	0.0795
Adam	0.0008	0.6790	0.0972	0.0187
John	0.0142	0.7586	0.4239	0.0049
Mary	0.0422	0.7430	0.2958	0.0446
…	0.0916	0.3923	0.1708	0.0646

WHAT to offer

WHO to target

Time
in 0-1 months,
in 1-3 months, ...

WHEN to offer

Typical Data Requirements in Banking – Account Summary Data

- Customer account information across selected accounts
  - <u>Mandatory if a cross sell model is required for the product</u>
  - Checking, saving account etc.
  - Loans - Credit card, mortgage, car, personal, overdraft etc.
  - Insurance, demat, asset management etc.
- For each account provide account level details
  - Customer id - Match key
  - Account id
  - Account type
  - Date since open
  - Monthly activity summary
    - Total cash flow (debit/credit), payments, penalties, interest, missed payment (where applicable) etc.
  - Product specific information – e.g., # of add on cards, debt to income

Typical Data Requirements – Transactional data with special format : A Real Case

- Data requirement is heavy
- Minimum 24 months of customer transactional data for all the customers for <u>all accounts</u>
- For each account a separate file
  - Each line item must have the following fields:
    - Customer id/account id
    - Date of transaction
    - Amount
    - Type of transaction
      - Customer initiated - purchase, cash advance, premium payment, ATM withdrawal etc.
      - Bank initiated – fees, penalties, waivers etc.
    - For credit card
      - Item code
      - Merchant name
- Requirements:
  - Each line item in this data must represent a single transaction
  - Data exploration may expose additional data needs

Data Requirements – Demographic Data : A Real Case

- A single demographic file to be provided across all accounts
  - Each line in the file must contain demographic information for one customer only
  - There should not be any duplicate entries for a single customer
  - Identify each customer by the match key (e.g., customer id)
  - If different accounts have same fields, use single latest view
- Requirements:
  - Customer id should match with the transactional data
  - The dataset must not contain any special character other than alpha-numeric and the delimiter

Cross sell Assumptions

- Prediction window can be up to 3 months (business decision)
  - Unlike in Merchant Offer where this has to be more granular
- Risk must be factored in cross-sell decisions
  - Use risk as a pre-filter
  - As a micro-segmentation
  - Include CB data in predictors

Bank Card Attrition Model: Design

Problem Statement

To build a model which could predict voluntary account attrition, 3 to 5 months ahead. This is to provide the business sufficient time to launch pro-active retention programs. The attrition score is expected to be used in conjunction with revenue estimates to design profitable retention strategies.

Criteria applied to the base data to filter off non-qualifying data :

1. Open Bankcard primary account, and
2. Current or Bkt 1 delinquency, and
3. Not part of the Bulk Sales program, and
4. MOB >= 6, and
5. Fee anniversary date in next 6 months

For modeling purpose, further segments created based on expected apriori behaviors and attrition rates in these segments.

1. Segment 1: Active, MOB >= 12
2. Segment 2: Inactive, MOB >= 12
3. Segment 3: Active, MOB 6-11
4. Segment 4: Inactive, MOB 6-11

Inactive is defined as having zero sales in the last 3 months (months -2, -1 & 0).

Bank Card Attrition Model: Design

Notation of month index

The following table illustrates how the months are annotated with reference to the development data scoring month of August 1997 (development base).

	Observation Period											Performance Period					
																Attition peri	
Date yymm	9609	9610	9611	9612	9701	9702	9703	9704	9705	9706	9707	9708	9709	9710	9711	9712	9801
Mth Index	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5

Good / Bad Definitions

Bad
Account that voluntarily attrited in months 4 - 6.

Good
No voluntary attrition; delinquency < Bkt 2 in months 1 - 6

Indeterminate
The rest of the base that includes

- Voluntary attrition in months 1 - 3 (I1)
- Involuntary attrition in months 1 - 6 (I2)
- Delinquency >= Bkt 2 in months 1 - 6 (G2)

Bank Card Attrition Model: Data Set

Base Size

Segment	Bad	Good	Indt (G2+I1+I2)	Total	Bad Rate Bad/(Good+Bad)	Attrition rate
active, MOB >=12	8,038	243,483	40,756	292,277	3.2%	11.2%
inactive, MOB >=12	6,877	50,820	19,930	77,627	11.9%	35.6%
active, MOB 6-11	5,770	64,788	10,154	80,712	8.2%	28.4%
inactive, MOB 6-11	4,466	12,147	7,325	23,938	26.9%	74.8%

Development Sample Size

Segment	Bad	Good	Indt (G2)	Indt (I1)	Indt (I2)	Total
active, MOB >=12	4,817	8,736	821	860	623	15,857
inactive, MOB >=12	4,124	8,768	831	870	693	15,286
active, MOB 6-11	3,462	6,000	300	300	300	10,362

Bank Card Attrition Model: Development

The data comprised of financial, demographic and relationship data. Several computed variables were generated by taking various statistical operations, i.e. weighted average, simple average, standard deviation, slope, ratio of slope, log of variables, etc., from the last 12 months financial data, and by taking into account the customer level relationship, i.e. no of gold cards, total customer credit line, MOB as a customer, etc.

Bank Card Attrition Model: Score Distribution MIS

- Bins are score groupings that have been organized into size of approximately 5%.
- The bin is assigned so that the higher the bin number, the more likely to attrite

Score Distribution MIS - Development

Development Base - Active, MOB 6-11

	Marginal %				Cumulative %		
Bin	Bad	Good	Indt	Total	Bad	Good	Bad Rate
1	0.5%	5.7%	3.1%	5.0%	100.0%	100.0%	0.8%
2	0.9%	5.7%	2.8%	5.0%	99.5%	94.3%	1.3%
3	0.8%	5.5%	4.5%	5.0%	98.6%	88.6%	1.2%
4	1.3%	6.7%	3.0%	5.8%	97.8%	83.1%	1.7%
5	1.1%	4.5%	3.4%	4.1%	96.5%	76.4%	2.1%
6	1.4%	5.5%	3.7%	5.0%	95.5%	71.9%	2.2%
7	2.1%	5.7%	3.7%	5.2%	94.0%	66.4%	3.2%
8	2.2%	5.3%	3.6%	4.9%	91.9%	60.7%	3.6%
9	2.5%	5.0%	6.2%	5.0%	89.7%	55.3%	4.3%
10	3.0%	5.6%	5.4%	5.4%	87.2%	50.4%	4.6%
11	3.7%	4.8%	4.0%	4.6%	84.2%	44.8%	6.5%
12	4.7%	4.9%	6.9%	5.2%	80.4%	40.0%	7.8%
13	4.9%	5.1%	4.8%	5.0%	75.7%	35.1%	7.9%
14	4.0%	3.3%	4.1%	3.5%	70.8%	30.0%	9.6%
15	8.4%	5.9%	7.7%	6.3%	66.8%	26.7%	11.3%
16	7.0%	4.6%	6.7%	5.0%	58.5%	20.8%	12.1%
17	9.5%	4.2%	7.5%	5.0%	51.4%	16.2%	16.7%
18	10.3%	4.3%	6.7%	5.0%	41.9%	12.0%	17.7%
19	13.7%	3.9%	7.1%	5.0%	31.7%	7.8%	23.8%
20	17.9%	3.8%	5.2%	5.0%	17.9%	3.8%	29.3%
Total	100.0%	100.0%	100.0%	100.0%			8.2%

Score Distribution MIS - Validation

Development Base - Active, MOB -11

	Marginal %				Cumulative %		
Bin	Bad	Good	Indt	Total	Bad	Good	Bad Rate
1	0.6%	6.4%	3.0%	5.6%	100.0%	100.0%	0.8%
2	0.8%	5.4%	3.0%	4.7%	99.4%	93.6%	1.3%
3	0.8%	4.3%	3.5%	3.9%	98.7%	88.2%	1.7%
4	1.3%	6.0%	3.4%	5.3%	97.8%	84.0%	1.8%
5	1.1%	4.6%	3.7%	4.2%	96.6%	78.0%	2.1%
6	1.1%	5.7%	3.3%	5.0%	95.5%	73.4%	1.7%
7	1.8%	5.9%	4.7%	5.4%	94.4%	67.7%	2.6%
8	1.5%	5.2%	4.3%	4.8%	92.6%	61.8%	2.6%
9	2.3%	5.8%	4.4%	5.4%	91.1%	56.7%	3.5%
10	3.0%	4.9%	3.9%	4.7%	88.8%	50.9%	5.1%
11	3.6%	4.6%	6.1%	4.7%	85.8%	46.0%	6.4%
12	4.2%	5.5%	5.8%	5.5%	82.2%	41.4%	6.3%
13	5.3%	4.9%	6.8%	5.2%	78.1%	35.8%	8.7%
14	4.4%	3.5%	4.8%	3.7%	72.8%	30.9%	10.2%
15	8.9%	5.7%	7.0%	6.1%	68.4%	27.4%	12.1%
16	9.1%	4.0%	5.7%	4.6%	59.5%	21.6%	16.9%
17	10.3%	4.4%	6.4%	5.1%	50.4%	17.7%	17.2%
18	10.3%	4.3%	5.8%	4.9%	40.2%	13.3%	17.5%
19	14.1%	5.0%	8.2%	6.1%	29.9%	9.0%	20.0%
20	15.8%	3.9%	6.1%	5.1%	15.8%	3.9%	26.2%
Total	100.0%	100.0%	100.0%	100.0%			8.2%

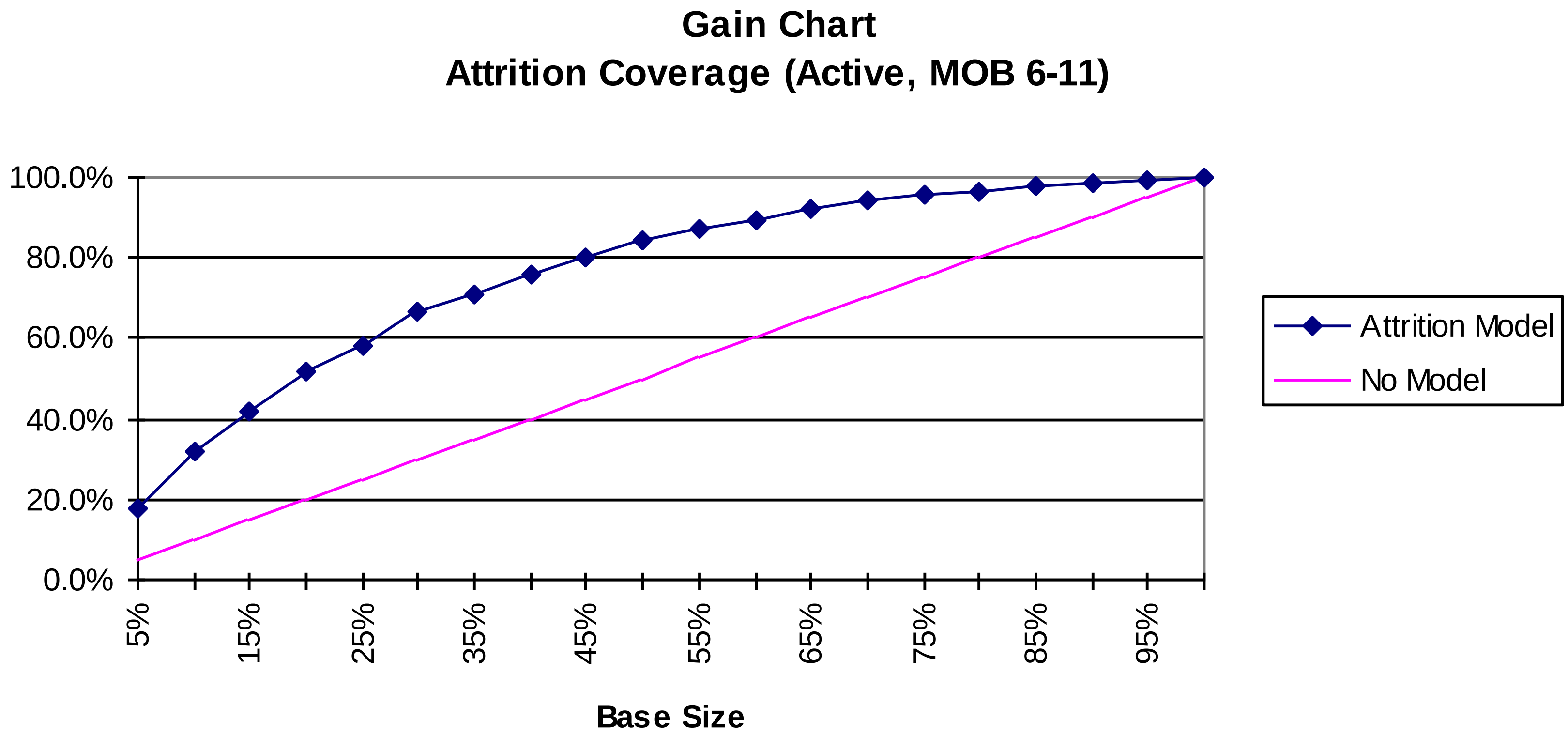
Bad Coverage

The purpose of the model is to proactively do retention programs to the group who is the most likely to attrite. We illustrate in the table below the percentage of Month 4 - 6 attritors covered in the Top 20 %.

Segment	Development (Month 0 = Aug 97)	In-time Validation (Month 0 = Aug 97)	Out-time Validation (Month 0 = Dec 97)
Active, MOB >= 12	45.2%	45.8%	45.0%
Inactive, MOB >= 12	40.7%	39.8%	34.1%
Active, MOB 6 - 11	51.4%	50.4%	55.5%

What the table says, for example is that by concentrating on the top 20% (Active, MOB>=12), 45.2% of the voluntary attritors in Month 4 - 6 can be captured (Based on development base).

ROC Curve or Gain Chart



Goodness of Fit

Separation measure

We give below the KS (Kolmogorov-Smirnoff) statistics to illustrate the measure of separation of Good and Bad, and also as a measure of shift between the development and validation period score distributions.

Good vs. Bad

Segment	Development (Month 0 = Aug 97)	In-time Validation (Month 0 = Aug 97)	Out-time Validation (Month 0 = Dec 97)
Active, MOB >= 12	33.0%	32.5%	34.2%
Inactive, MOB >= 12	33.6%	33.0%	29.8%
Active, MOB 6 - 11	40.8%	42.3%	42.1%

Population Stability vs Development Period

Segment	In-time Validation (Month 0 = Aug 97)	Out-time Validation (Month 0 = Dec 97)	Out-time Validation (Month 0 = Jun 98)
Active, MOB >= 12	0.8%	3.3%	3.1%
Inactive, MOB >= 12	1.9%	6.0%	5.0%
Active, MOB 6 - 11	1.3%	4.2%	3.1%

Revenue by Attrition – Active Mob 6-11 – Population Distribution – For Final Strategy

Account CNR Last 6 Months

Attrition Bin

Bin	<=760	760<-1000	1000<-1300	1300<-1650	1650<-2200	2200<-3000	3000<-4300	4300<-5800	5800<-8500	>8500	All
1	0.0%	0.0%	0.0%	0.0%	0.0%	3.6%	11.4%	29.6%	30.7%	24.6%	100.0%
2	0.1%	0.1%	0.0%	0.0%	0.4%	10.4%	38.2%	27.9%	13.6%	9.3%	100.0%
3	0.0%	0.0%	0.1%	0.1%	0.8%	7.5%	17.3%	29.3%	29.4%	15.4%	100.0%
4	0.0%	0.0%	0.1%	0.2%	6.0%	17.3%	34.9%	27.7%	10.4%	3.4%	100.0%
5	0.0%	0.0%	0.0%	1.2%	10.2%	13.4%	20.9%	23.4%	19.4%	11.5%	100.0%
6	0.0%	0.1%	1.0%	4.5%	9.8%	14.5%	23.5%	21.6%	19.0%	6.1%	100.0%
7	0.0%	0.2%	1.6%	4.6%	7.5%	15.6%	31.2%	26.4%	9.8%	3.0%	100.0%
8	0.0%	0.2%	6.6%	16.9%	21.8%	25.0%	13.3%	8.1%	4.8%	3.3%	100.0%
9	0.0%	0.4%	12.6%	16.3%	18.9%	18.8%	14.8%	9.0%	4.8%	4.4%	100.0%
10	0.2%	3.5%	35.0%	12.0%	17.2%	16.5%	7.3%	3.3%	3.7%	1.1%	100.0%
11	0.1%	5.4%	18.6%	13.3%	28.7%	20.4%	6.7%	2.8%	2.3%	1.6%	100.0%
12	0.2%	10.0%	25.0%	18.2%	21.4%	16.4%	5.3%	1.5%	1.1%	0.8%	100.0%
13	0.1%	9.5%	18.7%	27.5%	27.1%	12.0%	3.2%	0.7%	0.8%	0.4%	100.0%
14	0.2%	11.9%	17.0%	41.1%	21.1%	6.8%	1.8%	0.1%	0.0%	0.0%	100.0%
15	0.1%	9.7%	38.2%	29.8%	15.3%	4.8%	1.8%	0.3%	0.0%	0.0%	100.0%
16	0.2%	14.2%	31.2%	30.2%	15.4%	6.2%	2.2%	0.3%	0.0%	0.0%	100.0%
17	0.1%	10.6%	40.0%	28.6%	12.9%	6.4%	1.2%	0.2%	0.0%	0.0%	100.0%
18	0.1%	11.1%	60.6%	16.3%	9.2%	2.6%	0.1%	0.0%	0.0%	0.0%	100.0%
19	0.3%	28.3%	50.6%	10.9%	7.7%	1.9%	0.2%	0.1%	0.0%	0.0%	100.0%
20	0.5%	53.7%	35.7%	4.6%	4.1%	1.1%	0.3%	0.1%	0.0%	0.0%	100.0%
ALL	0.1%	7.6%	18.4%	13.7%	12.6%	11.2%	12.7%	11.4%	7.9%	4.4%	100.0%

Question & Answer Session