An overview of Data Mining and its Practical Applications within Business Today

Aug. 8/2019

What is Data Mining?

- Data Mining is the use of information at a micro or granular level to make more informed decisions.
- How is this different than traditional analysis? Information has been typically been obtained to gleam some basic level of understanding or insight. These insights are typically at a macro level.
- A Knowledge Discovery Process that provides insight which can be actioned for some benefit

What is Data Mining?

Data Mining - Context

Make better informed decisions due to granularity of data

- Can reply on both advanced statistics and non advanced statistics to help you make better business decisions
- Identifies characteristics and or key areas to assist you in targeting customers or prospects
- Acquisition, Cross-Selling, Up-Selling, Retention, Loyalty, Other Due to Granularity of Data

Growing Area

More data and better tools to develop the information at more granular levels
More Accountability

- Measurement
- Effectiveness or Efficiency of Marketing $$$ spent

Current Uses of Data Mining within the Current Environment

Embedding of insight and information in new technology and devices:

At the Call Centre

One Sales Rep’s Contact Lead List

One Customer’s Laptop or iPad

At the CSR Desk

On Customer’s Mobile Phone

Some Examples of Data Mining in Action

Fraud

Risk Management

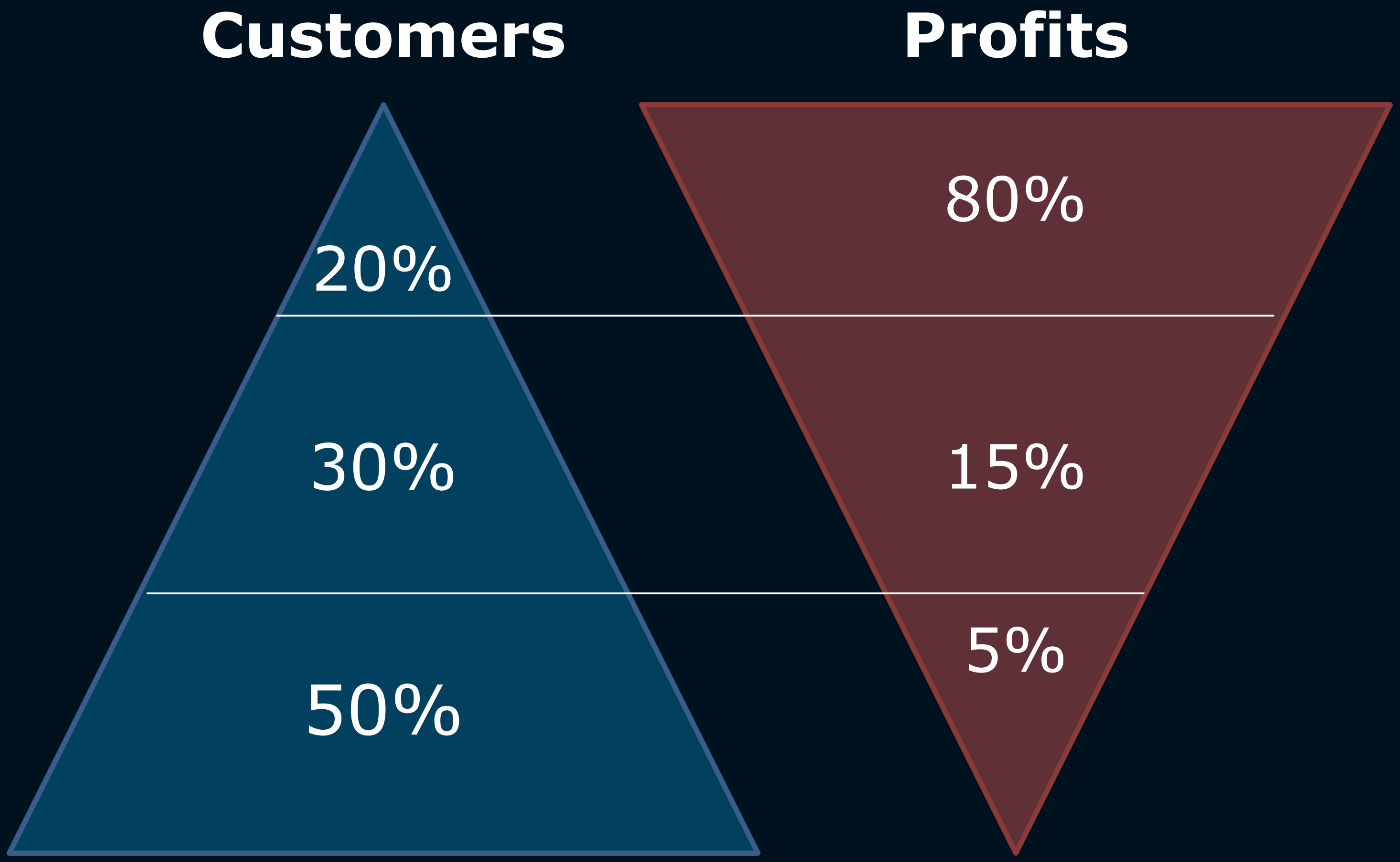
Crime Management

Health Care

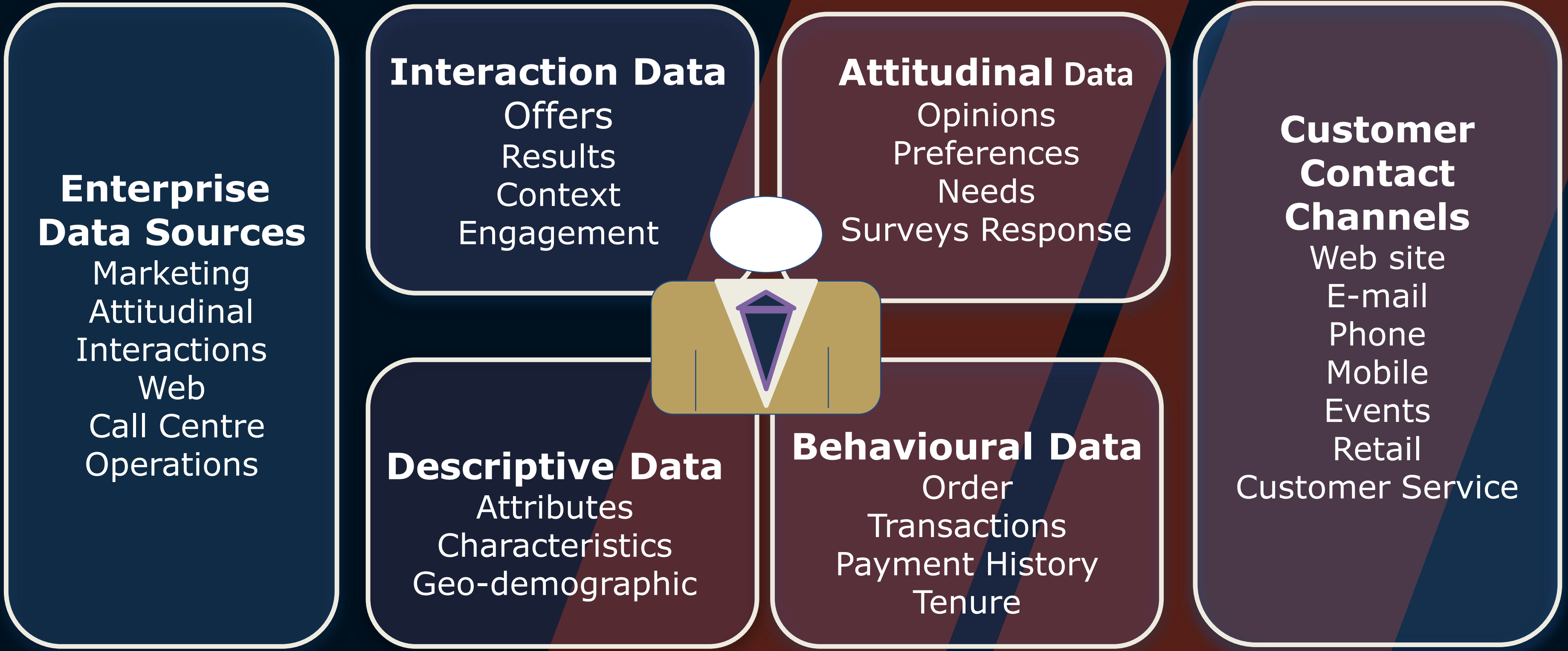
Inventory Management

It’s Main Application has traditionally been in marketing and specifically CRM (Customer Relationship Management)

DATA Mining and the Pareto Impact- The 80/20 rule-



DATA IS THE HEART OF BUSINESS INSIGHTS AND INTELLIGENCE



Enterprise Data Sources
Marketing
Attitudinal
Interactions
Web
Call Centre
Operations

Interaction Data
Offers
Results
Context
Engagement

Attitudinal Data
Opinions
Preferences
Needs
Surveys Response

Customer Contact Channels
Web site
E-mail
Phone
Mobile
Events
Retail
Customer Service

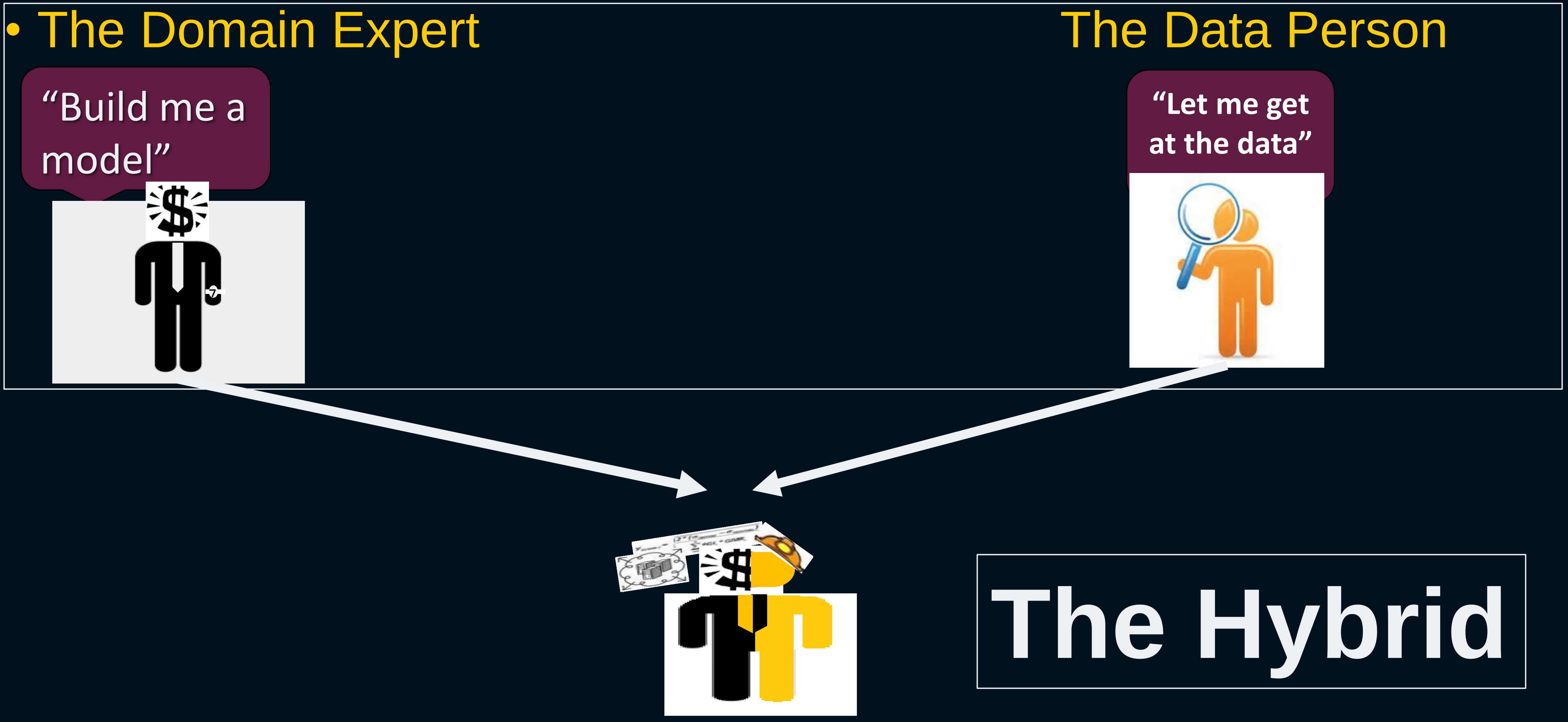
Descriptive Data
Attributes
Characteristics
Geo-demographic

Behavioural Data
Order
Transactions
Payment History
Tenure

Why is Data Mining so Important Today?

- The Explosion of Big Data and the need to become data-driven
- Increased Expectations to generate more insights quickly
- A Proliferation of new tools and technology to help empower more people
- But what is the ultimate challenge for all businesses and organizations today?

The Ultimate Challenge for all businesses and organizations today



- The Domain Expert

“Build me a model”

The Data Person

“Let me get at the data”

The Hybrid

What does this hybrid look like?

- Business Strategy

- Mathematics and statistical knowledge
  - Requirement or need is growing

- Working with data
  - Programming/Coding, Processing of Data

- Communication

Is there a Data Mining process or framework?

THE 4 STEPS PROCESS IN BUILDING A DATA ANALYTICS SOLUTIONS

A Discipline that requires STRUCTURE and PROCESS

- We utilize the following four-step process to manage projects:

T The Stakeholders

The Domain Expert

The Analytics Expert

The I/T and Data Custodian

The Four Steps

Problem Identification

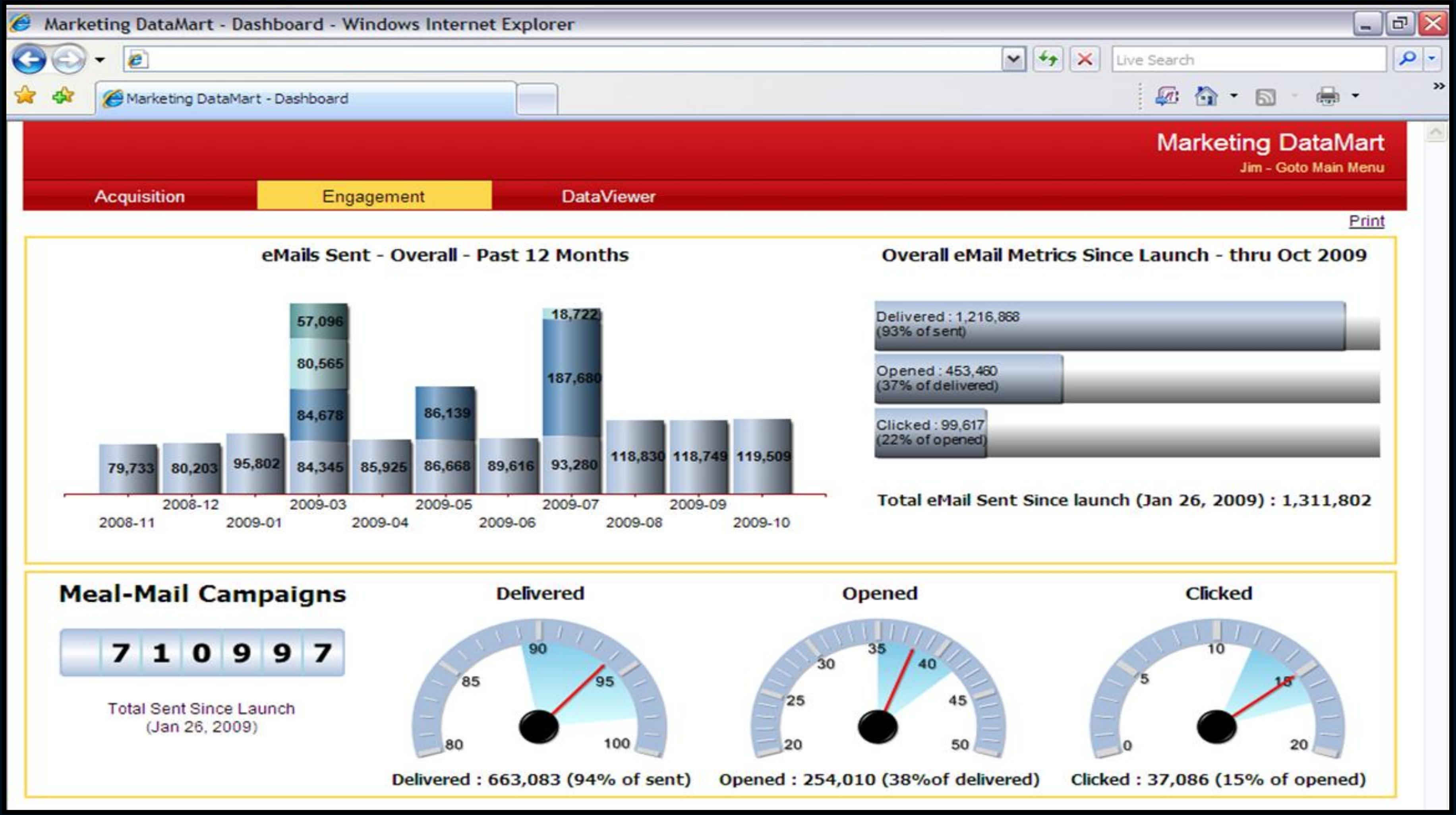
Creation of the Analytical Data Environment

Application of the Analytics Tools
Implementation and Tracking

Implementation and Tracking

Some Key Deliverables

- Reporting-KBM



Total eMail Sent Since launch (Jan 26, 2009) : 1,311,802

Delivered : 663,083 (94% of sent)

Opened : 254,010 (38%of delivered)

Clicked : 37,086 (15% of opened)

Some Key Deliverables

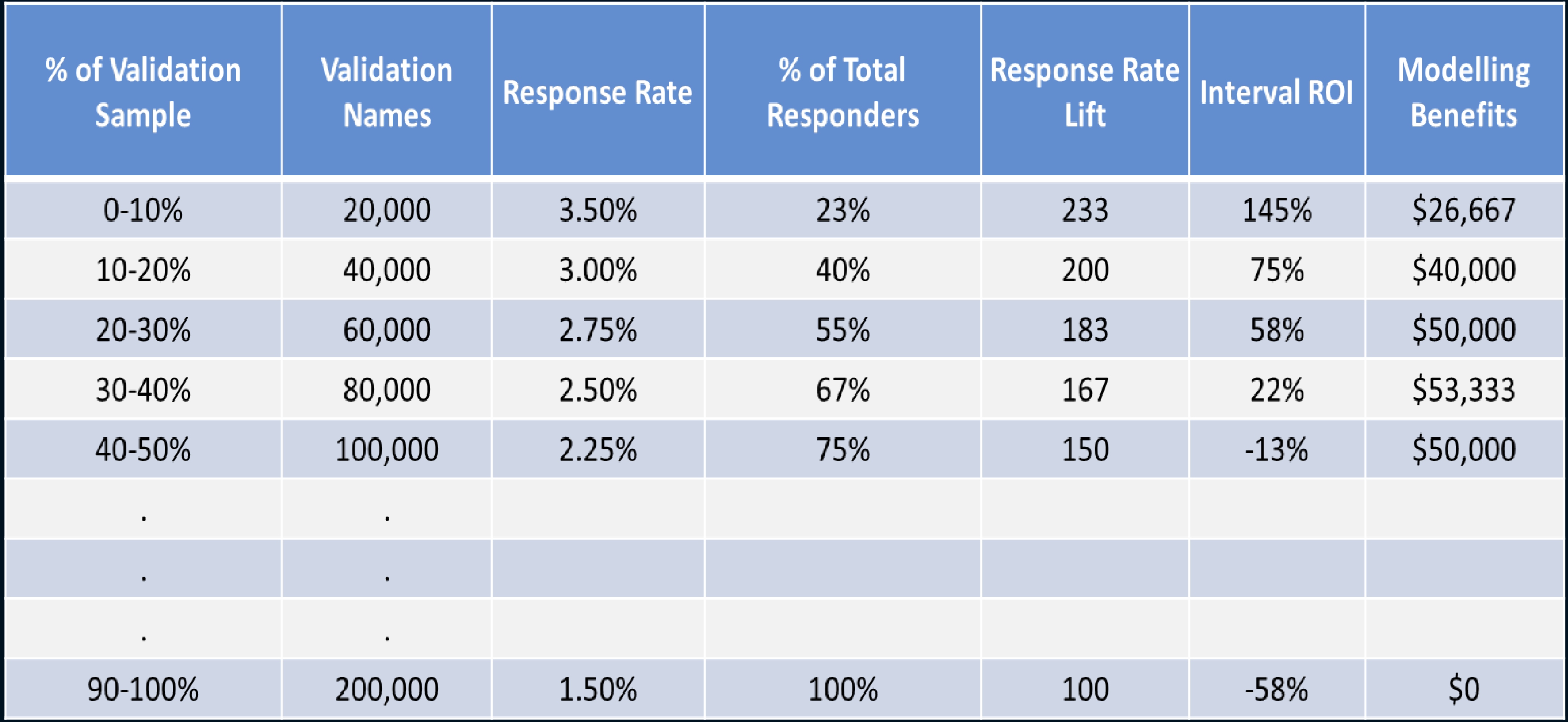
- Reporting-ADHOC (COHORT)

# Clients (New 2003)	% of Clients Retained	Cancel Rate	# of Policies		# Policies/Client	Total Premium	Average Premium/ Client	Average Premium/ Policy	Total Cumulative Premium	LTV of 2003 New Clients	Broker Revenue
Yr 1: 1,322			2,018		1.53	$3,347,447	$2,532	$1,659	$3,347,447	$2,532	**$380**
Yr2: 1,170	89.0%	11.5%	1,827		1.56	$2,928,008	$2,503	$1,603	$6,275,455	$4,747	**$712**
Yr3: 1,070	81.0%	8.5%	1,696		1.59	$2,604,706	$2,434	$1,536	$8,880,161	$6,717	**$1,008**
Yr4: 976	74.0%	8.8%	1,575		1.61	$2,396,783	$2,456	$1,522	$11,276,944	$8,530	**$1,280**
Yr5: 892	67.0%	8.6%	1,472		1.65	$2,207,536	$2,475	$1,500	$13,484,480	$10,200	**$1,530**
		9.4%	**8,588**		**1.60**	**$13,484,480**	**$2,480**		**$10,200**		

Some Key Deliverables:
THE FINAL MODEL VARIABLE REPORT:

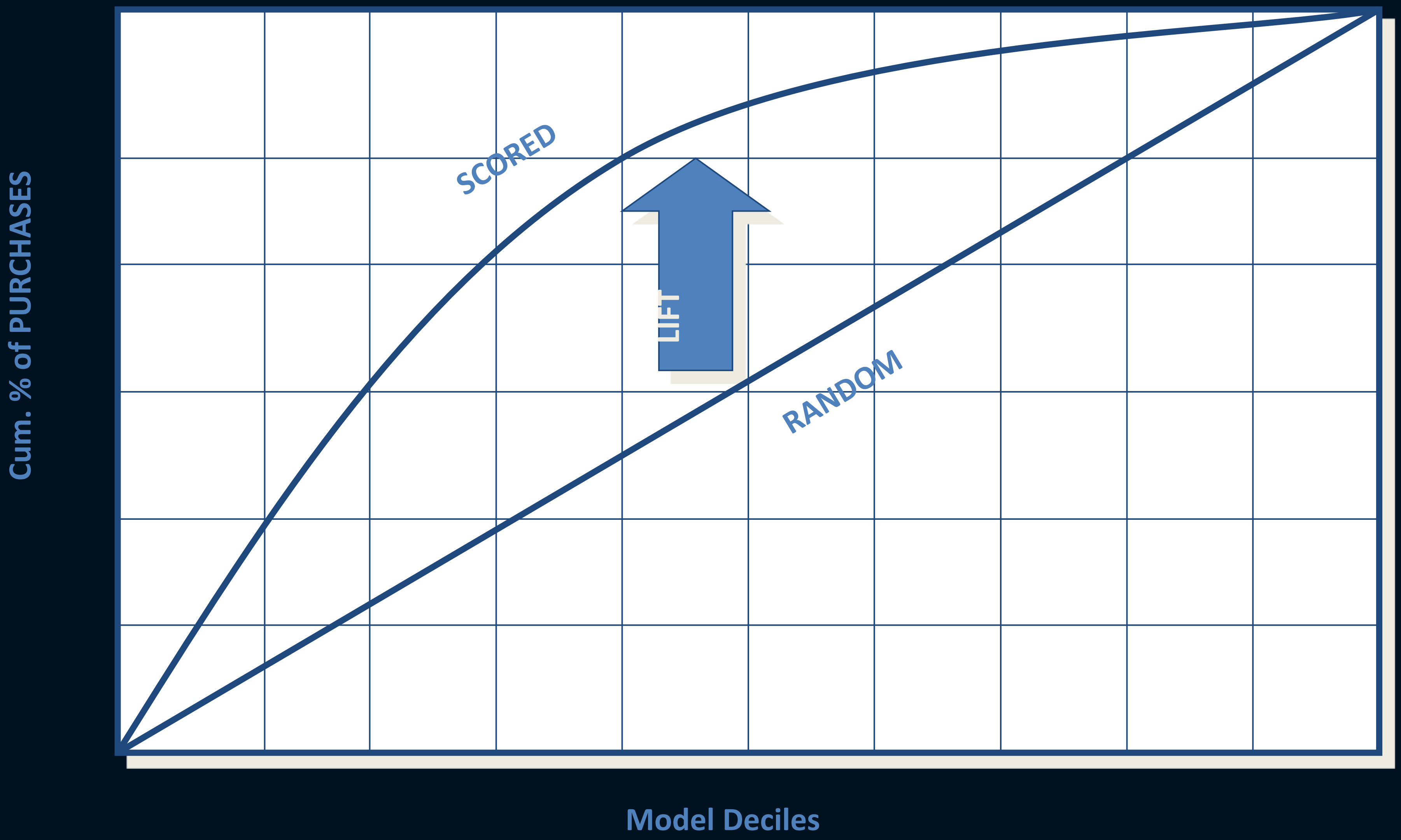
Model Variable	Impact on Response	Contribution to Overall Equation
Behaviour Score	Positive	35%
Average Score	Positive	25%
Have an RRSP Product	Negative	15%
# of Fin. Inst. Products	Positive	10%
Avg. % of Credit Limit Used	Positive	10%
Live in Prairie Provinces	Negative	5%

Some Key Deliverables:
Model Evaluation-Gains Charts



% of Validation Sample	Validation Names	Response Rate	% of Total Responders	Response Rate Lift	Interval ROI	Modelling Benefits
0-10%	20,000	3.50%	23%	233	145%	$26,667
10-20%	40,000	3.00%	40%	200	75%	$40,000
20-30%	60,000	2.75%	55%	183	58%	$50,000
30-40%	80,000	2.50%	67%	167	22%	$53,333
40-50%	100,000	2.25%	75%	150	-13%	$50,000
.	.					
.	.					
.	.					
90-100%	200,000	1.50%	100%	100	-58%	$0

Some Key Deliverables:
Model Evaluation-AUC Curve



CASE STUDY - AMERICAN EXPRESS

Data Analytics over the Long-Term

- **1980's Major Goal:**
  -acquisition of new cards

- **Results**
  -Doubled their card base over several years
  -Cost per card doubled from $100 to $200

- Cost situation was unacceptable

CASE STUDY - AMERICAN EXPRESS

Data Analytics over the Long-Term

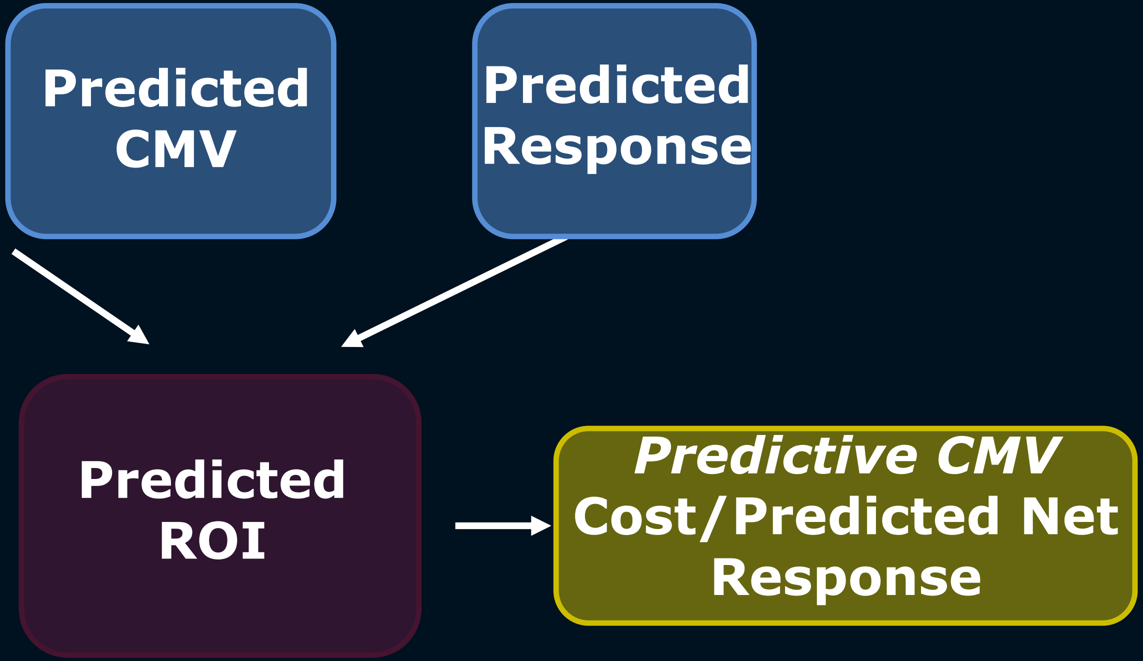
- **Began with Simple Response Model to become more cost efficient**

- **But the journey ended up where we built a series of models where we could ultimately predict ROI at the prospect level.**

CASE STUDY - AMERICAN EXPRESS

Data Analytics over the Long-Term

Predicting ROI

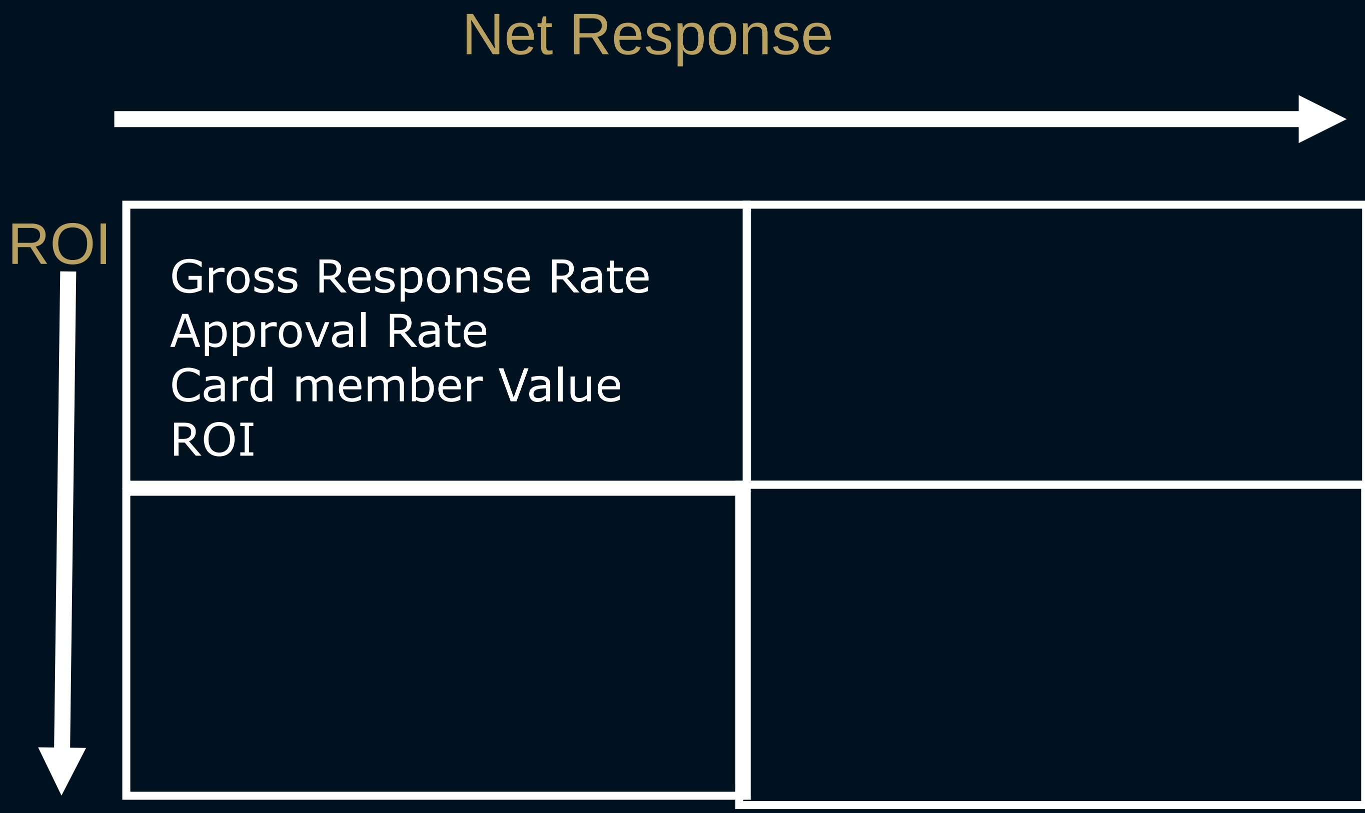


Predicted CMV

Predicted Response

Predicted ROI

Predictive CMV Cost/Predicted Net Response



Net Response

ROI

Gross Response Rate
Approval Rate
Card member Value
ROI

CASE STUDY

Financial Institution

- A conversion model was determined during the problem identification as the solution which would optimize the conversion of regular credit card holders into gold card holders

- Previous selections based on tenure were becoming ineffective. *This will be shown in a few slides*

Case Study: Financial Institution

- A series of regression routines are then run against these 30 variables. *The final results of these efforts should yield the following report:*

Model Variable	Impact on Response	Contribution to Overall Equation
Behaviour Score	positive	35.0%
Average Spend	positive	25.0%
Have a RRSP Product	negative	15.0%
# of Fin. Inst. Products	positive	10.0%
Avg. % of Credit Limit Used	positive	10.0%
Live in Prairie Provinces	negative	5.0%

Case Study: Financial Institution

- Gains Chart - Application of the Model to the Validation Sample
- Assumptions:
  - Revenue is $60 which is the card fee
  - No incremental spend is included in the revenue number
  - Cost of 1 promoted piece is $.80

% of List (Ranked by Model Score)	Validation Mail Quantity	Cum. Resp. Rate	Cum. % of Responders	ROI
0-20%	4,000	2.0%	40%	50%
20-40%	8,000	1.6%	64%	20%
40-60%	12,000	1.4%	84%	5%
60-80%	16,000	1.2%	96%	-9%
80-100%	20,000	1.0%	100%	-25%

Case Study: Financial Institution

- Quantification of Estimated $ Benefits:
- Assuming that we have to generate the same number of responders either with or without modelling, the following table can be produced

	Response Rate	# of Responders	# of Names promoted
No Modelling	1.0%	3,200	320,000
Modelling	1.6%	3,200	200,000

Saved Marketing Quantity	120,000
Estimated $ Benefits ($0.80 per promoted customer)	$96,000

CASE STUDY

Using Predictive Models to Create Better Pricing Tools for P&C Insurace

- A key challenge in auto and property insurance is the ability to effectively charge the right premium

- Historically, premiums have been based on business rules that estimate credit loss as determine by actuaries

- Cross tab reports along with statistical tests have determined the set of business rules that yield the most significant results in terms of claim loss

- There is one glaring weakness here

Case Study:
Credit Scoring - Now, contribute more factors: gender, age, and distance to work

Distance to Work	<30 km	<30 km	>30 km	>30 km	Total
Age	Under 25	Over 25	Under 25	Over 25	
Male	1.16	1.09	1.95	1.70	1.22
Female	0.61	0.49	0.97	0.91	0.73
Total	**1.16**	**1.22**	**0.88**	**1.03**	**1.00**

Female over 25 years old who drives under 30 kilometers to work would be charged:
$600 X .49 = **$294**

Male under 25 years old who drives more than 30 kilometers to work would be charged:
$600 X 1.95 = **$1170**

So why isn't this sufficient for pricing purposes?

Case Study:
Credit Scoring - Challenges with the group Differential Approach

Groups	# of Records	Differential
Male over 25 years and drives over 30 kilometers to work	100,000	1.7
Total # of policies	300,000	1

Lack of Granularity

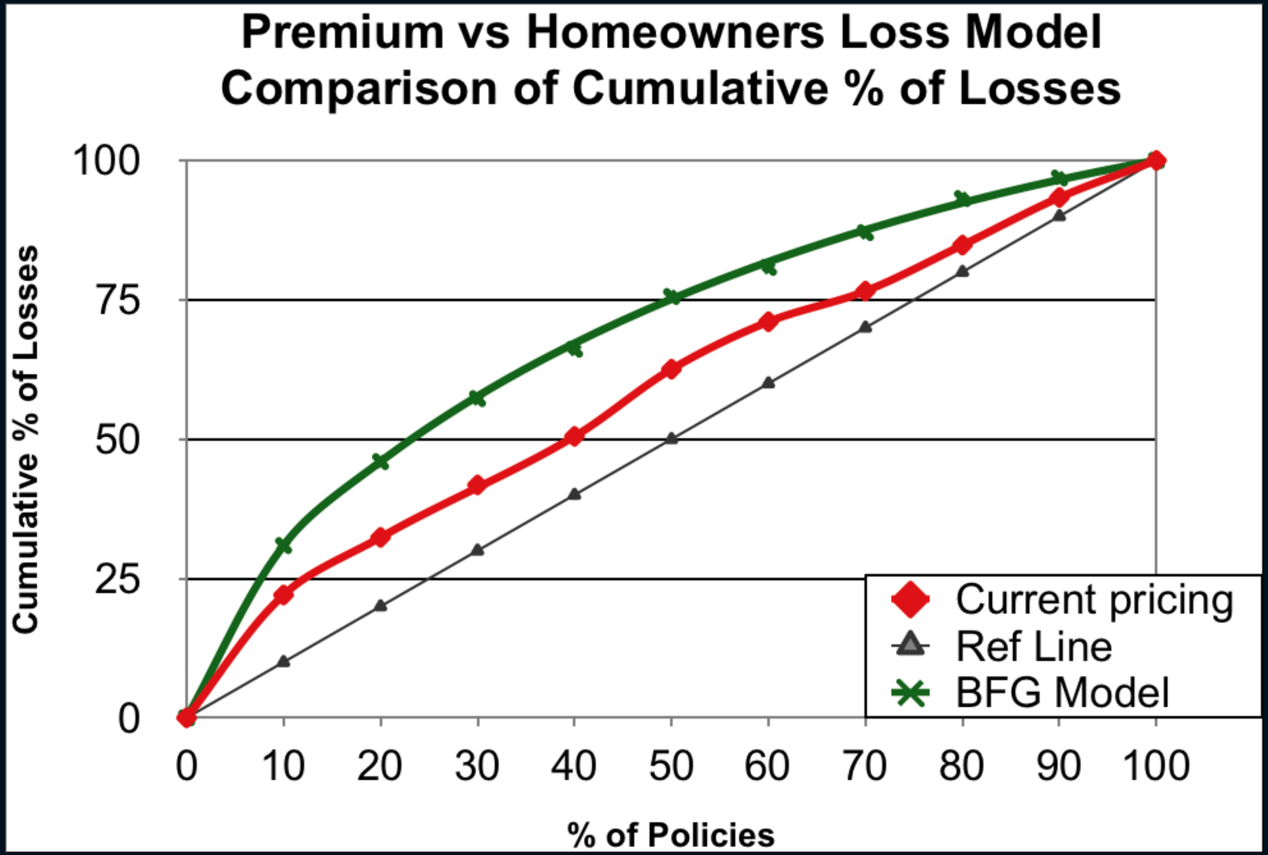
- Based on this example, 100,000 or 1/3 of the entire portfolio will obtain the same level of risk. Is it possible to get more granular in calculating risk for smaller groups of records?

No multi-collinearity or interaction between variables.

Solution: *MVA (Multivariate Analysis) or Predictive Analytics*

- *Outcome is a score for each individual*
- *Solution that takes into account the interaction between variables*

Example of Property Loss Model



- **A model developed for Homeowner’s coverage significantly outperformed existing premium as a tool to predict losses**