

Data Mining in Business Analytics for Decision Support

Albert He
Customer Insight & Measurement
Scotiabank International Banking Marketing

Contents

- An Example
- Concept of Data Mining & Business Analytics
- Modeling for Decision Support
- Automatic Campaign Tracking & Reporting



Who is more likely to be a lead analyst?

■ Data sample

Please answer the following questions.

1. What is your level of SAS skills (1-10)?
2. Do you have good education?
3. Do you like SAS?
4. Do you feel old?
5. Do you feel young?
6. Are you mature?
7. What is your birth date?



Who is more likely to be a lead analyst?

■ Profile

Segment	# Analysts	Avg Age	% Young	% Old	SAS Level	% Good Edu	% Mature	% Lead Analyst
Hard Worker	1,685	37.0	74.2%	13.3%	5.1	85%	94%	0%
Retiring	77	59.9	76.6%	100.0%	5.8	87%	92%	3%
Sr. Analyst	1,187	37.0	100.0%	16.3%	6.7	100%	90%	52%
Tech Manager	2,051	37.8	85.3%	16.3%	8.3	100%	100%	72%
All	5,000	37.7	84.9%	16.6%	6.8	95%	96%	42%

■ Insight

- If you are not a technical manager or a sr. analyst, you are very unlikely to be a lead analyst.
- All sr. analysts consider themselves young.
- Few people would think themselves not mature.
- ...



Who is more likely to be a lead analyst?

■ Predictive Model

```
LOGIT =  -15.21243801
          + 0.0142918147 * age
          + 4.971641269 * I_OLD
          - 0.1156662831 * ageI_OLD
          + 1.6239612562 * SAS_LEVEL
          + 2.0553092137 * i_segment
          ;
```

```
Score = 1 / (1 + exp(-logit)) ;
```

Percent
Concordant
89.1

■ Insight

- Those who are older and consider themselves old are more likely to become lead analysts.
- Segment is a key driver.



The Data

```
data dm_sample;
  do id=1 to 5000;
    age=30+ranuni(1)*ranuni(1)*40-ranuni(2)*ranuni(2)*10;
    I_Young=(ranuni(3)*50/age>0.2);
    I_OLD =(ranuni(4)*age>30);
    SAS_LEVEL=int(((1-ranuni(5))*ranuni(5))*8+1.3);
    Good_Edu=(ranuni(6)>0.05);
    I_Mature=(ranuni(7)>0.05);
    if sas_level>7 and good_edu=1 and i_mature=1
      then segment="Tech Manager";
    else if sas_level>5 and good_edu=1 and 27<age*i_young<55
      then segment='Sr. Analyst ';
    else if age*i_old>55 then segment='Retiring  ';
    else segment='Hard Worker ';

    Lead_Analyst=(sas_level>6)*good_edu*i_mature*i_young*(28<age<56);

    i_segment=(segment="Tech Manager")+(segment='Sr. Analyst ')*2;
    output;
  end;
run;
```



What are the issues?

- Data don't always have the truth.
- Answers are subjective.
- The target definition is highly correlated to a derived variable.
- Statistics can be used to fool.
- Lastly (or firstly), are we doing the right thing?



What is Data Mining?

An Old Question

■ Some Definitions/Descriptions

- The process of analyzing data from different perspectives and summarizing it into useful information.
 - The nontrivial extraction of implicit, previously unknown, and potentially useful information from data.
 - The science of extracting useful information from large data sets or databases.
 - Exploration & analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns & rules.
- The science and engineering process of obtaining knowledge and concentrated information from data repositories.



Process and Techniques

- **Deposit Exploration**

Meta data, tables/columns related, data type, value distribution, missing values

- **Construction and Extraction**

Oracle, DB2, SQL Server, SAS, ...

- **Concentration and Smelting**

Classification (decision tree), clustering, profiling, regression, neural network,...

- **Shipping**

Model score, report, profile, learning, insight, ...

- **Who are the best data miners?**



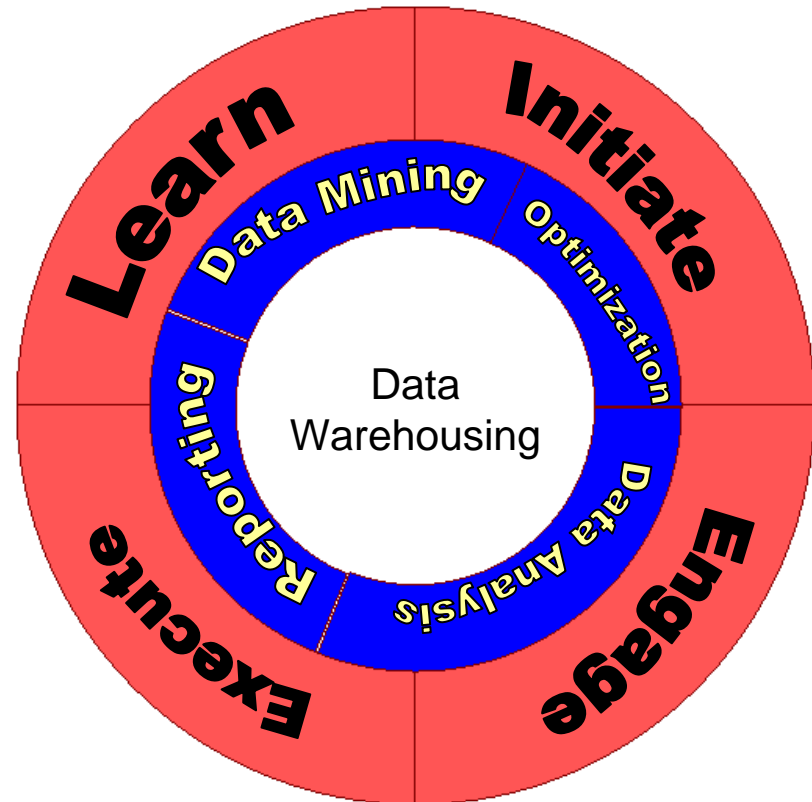
What is business analytics?

■ Business analytics

Provide **effective decision support** by leveraging the analytical capacity in data mining, optimization, data analysis, and reporting.

■ Keys

- Initiate
- Engage
- Execute
- Learn



Modeling for Decision Support

■ Build a model to support direct marketing

➤ Initiate

- Needs for targeting
- Requirement analysis

➤ Engage

- Communication
- Project charter
- Sign off

➤ Execute

- Data preparation
- Model development
- Model validation
- Model production/scoring
- Campaign optimization

➤ Learn

What are the issues with our sample model and some real models?

- Why did we decide to build the model?... (initiate)
- Who were involved? ... (engage)
- How was the model developed? ... (execute)
- What did we learn? ... (learn)

What model to build?

- What is a good model?

Decision background, business needs, effective support

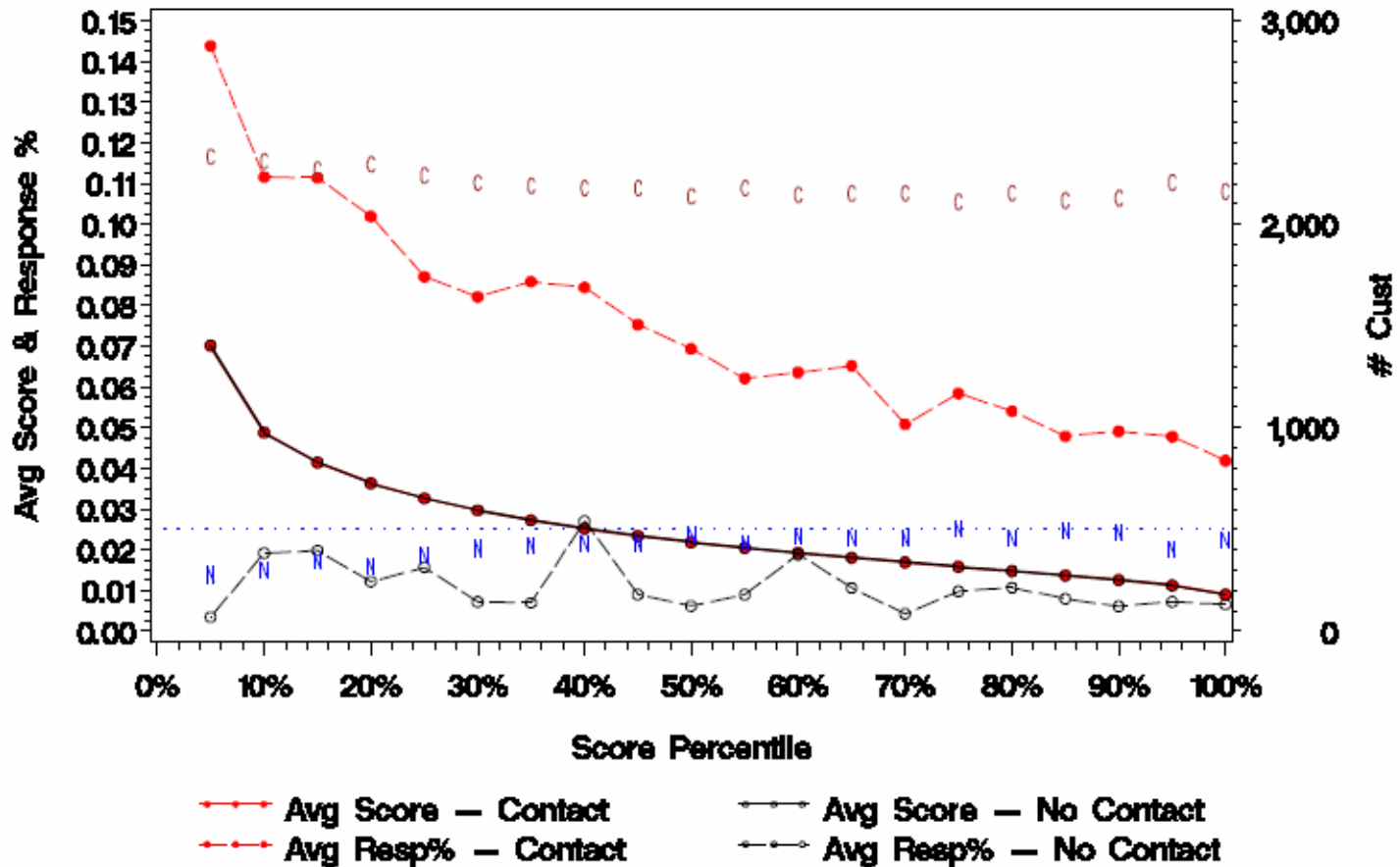
- What is not a good model?

Not needed, inappropriately defined, questionable development

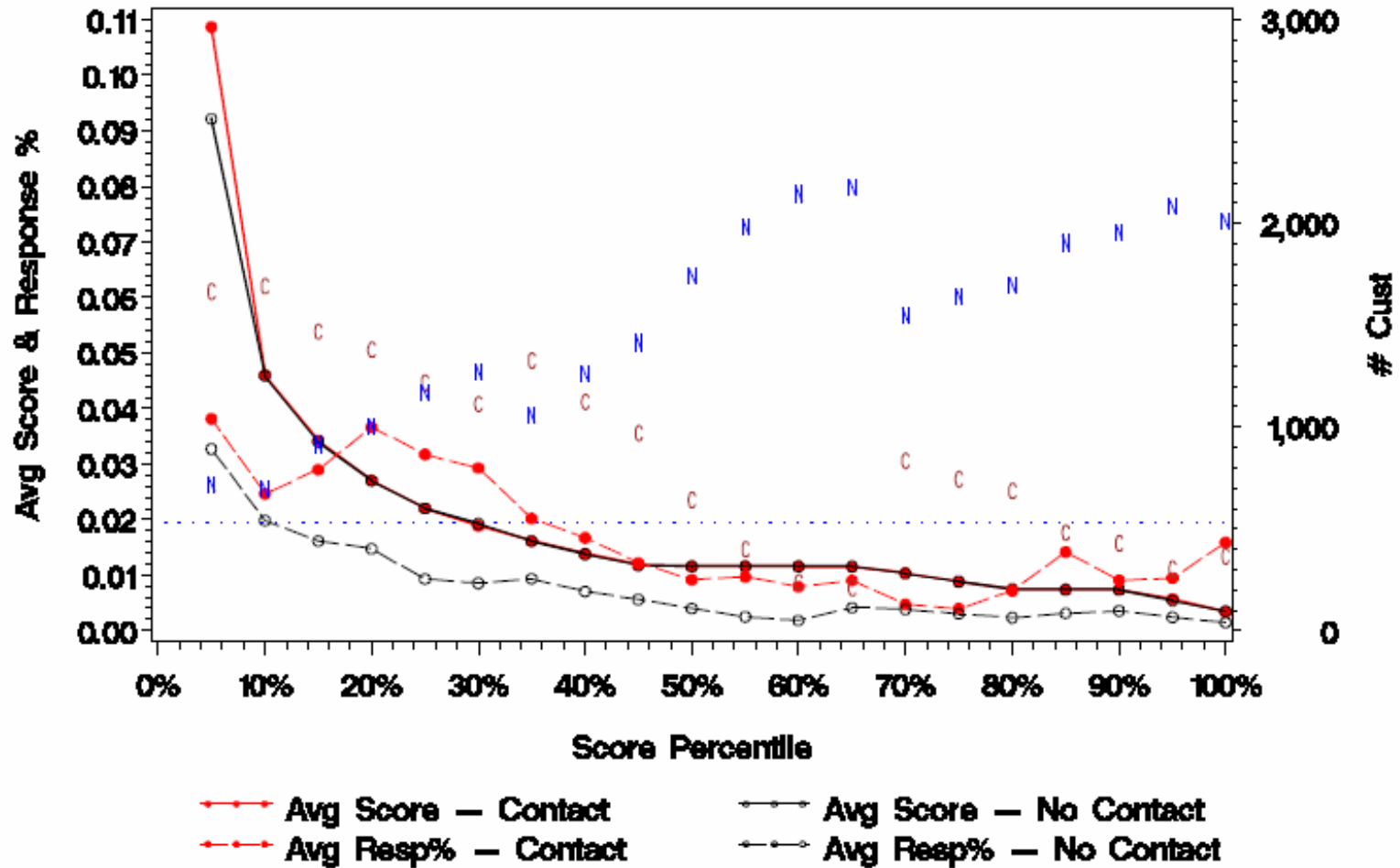
- What if models are competing?

Integration? Optimization?

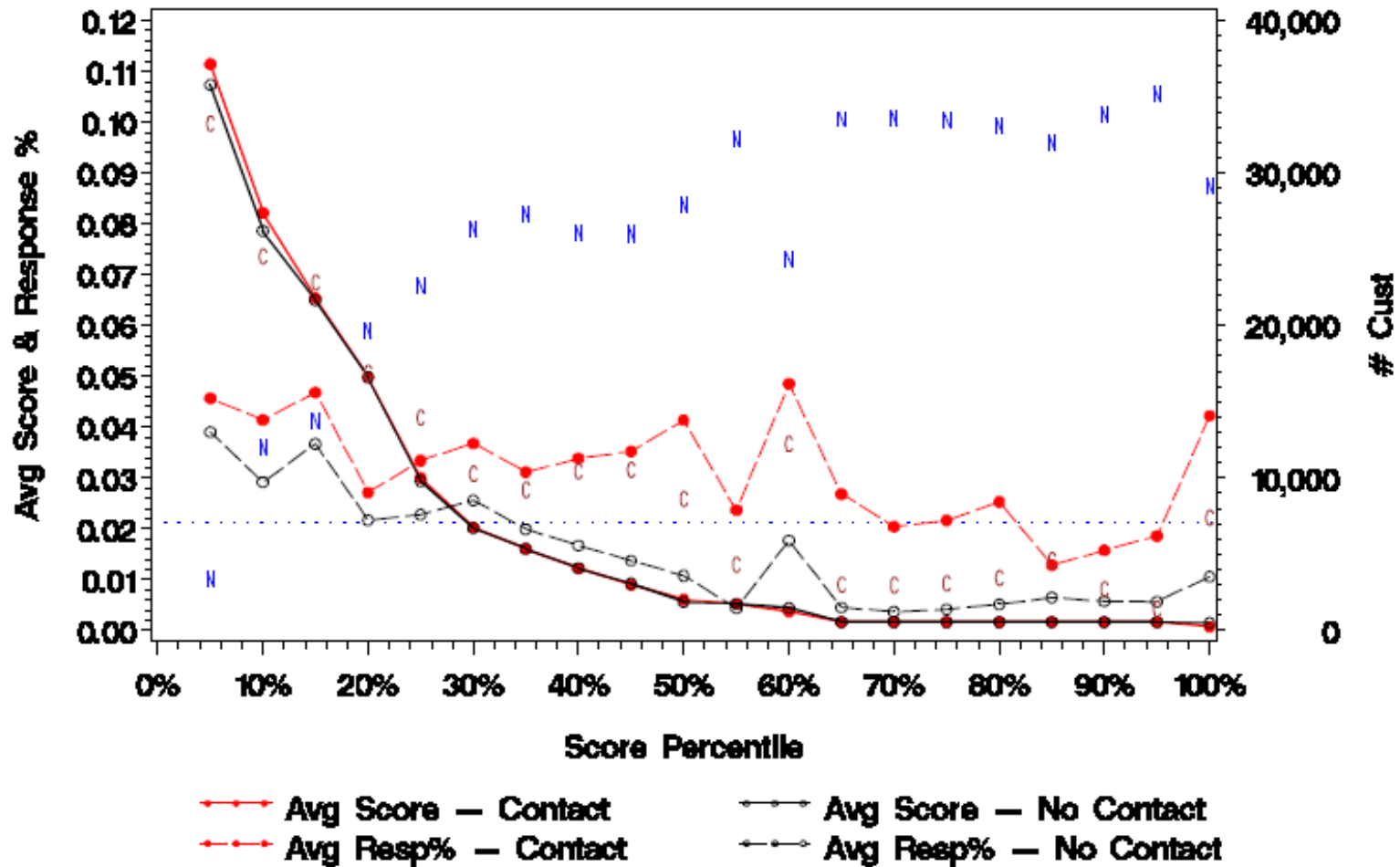
A Good Response Model



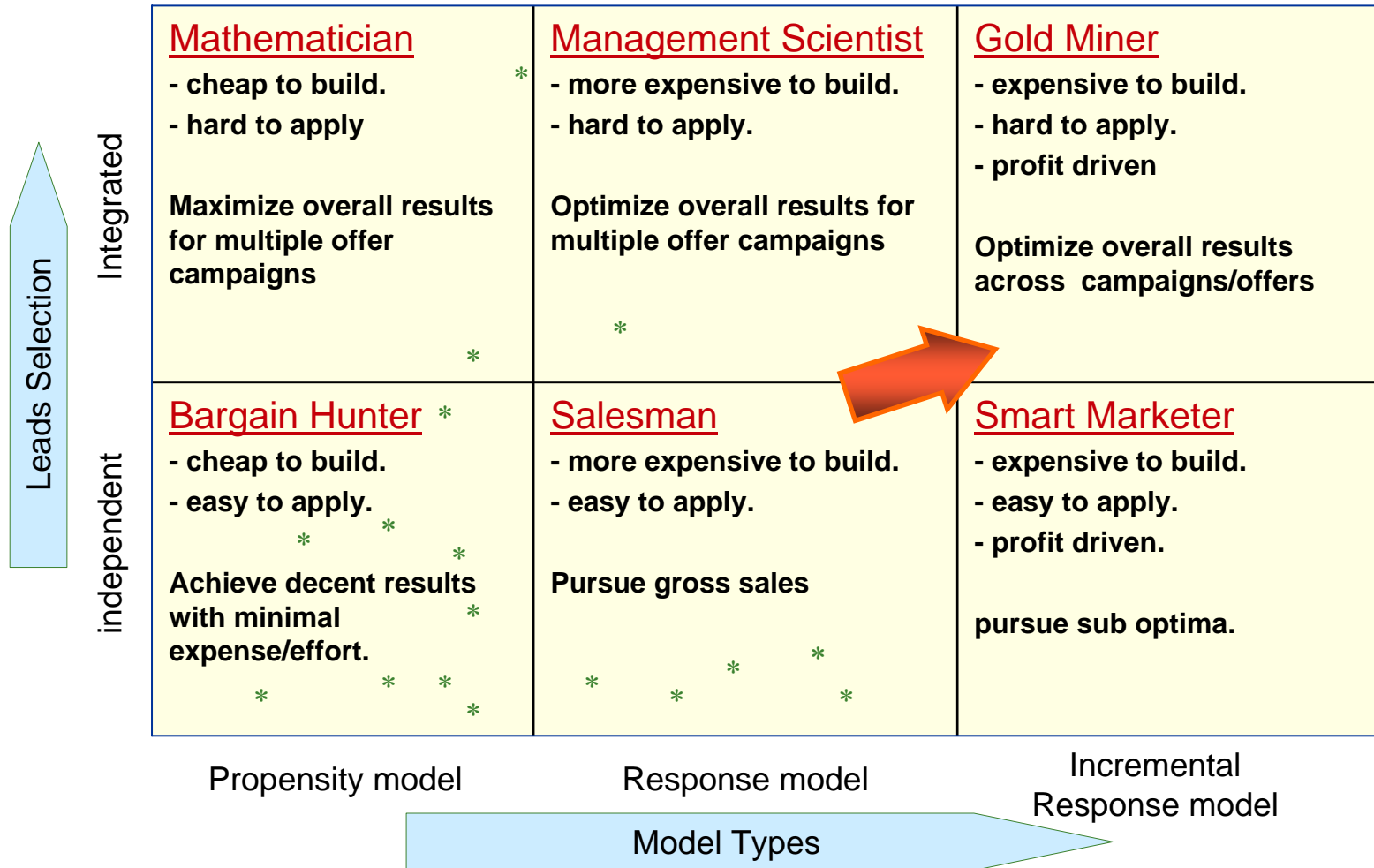
A Good Propensity Model?



What About This Model?



Modeling Strategies



Automatic Campaign Tracking and Reporting

■ Background

- Current structure
- Needs
- Challenges

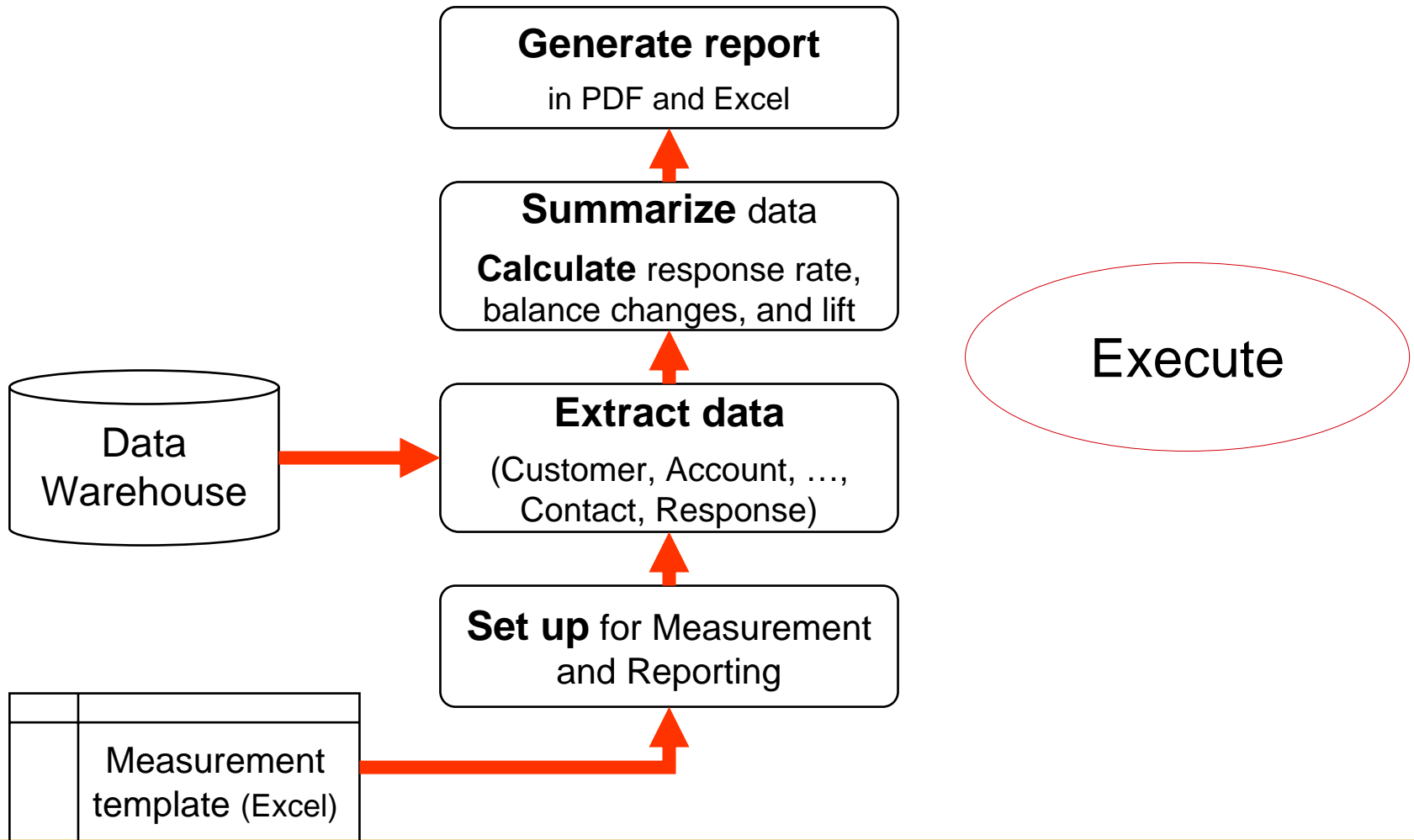
Initiate

■ Sell to stakeholders

- Ideas
- Methodology
- Processes
- Outcome

Engage

Automatic Campaign Tracking and Reporting



A Peek at the Standard Report

CA_Report200711.pdf - Adobe Reader

File Edit View Document Tools Window Help

4 / 13 71.2%

Find

Bookmarks

- Lift by Campaign
 - CA200711: Lift in Balance
 - CA200711: Lift in Account
 - CA200711: Lift in Service
- Lift by Offer
 - CA200711: Lift in Balance by Offer**
 - CA200711: Lift in Account by Offer
 - CA200711: Lift in Service by Offer
- CA200711: Response Summary - On-going Campaigns
- Offer Response Summary by Campaign
 - Campaign Name
 - Blueprint/Mut Fund SIS Event Triggers
 - ETD4017_Ultimate_GIC
 - INDIRECT LOAN CUSTOMER WELCOME
 - Investment Opportunities - MTV
 - Last Loan Payment
 - Maturing Opportunities
 - Mortgage Continuous Campaign
 - MvVault Investment

CA200711: Lift in Balance by Offer
From 01NOV2007 to 30NOV2007

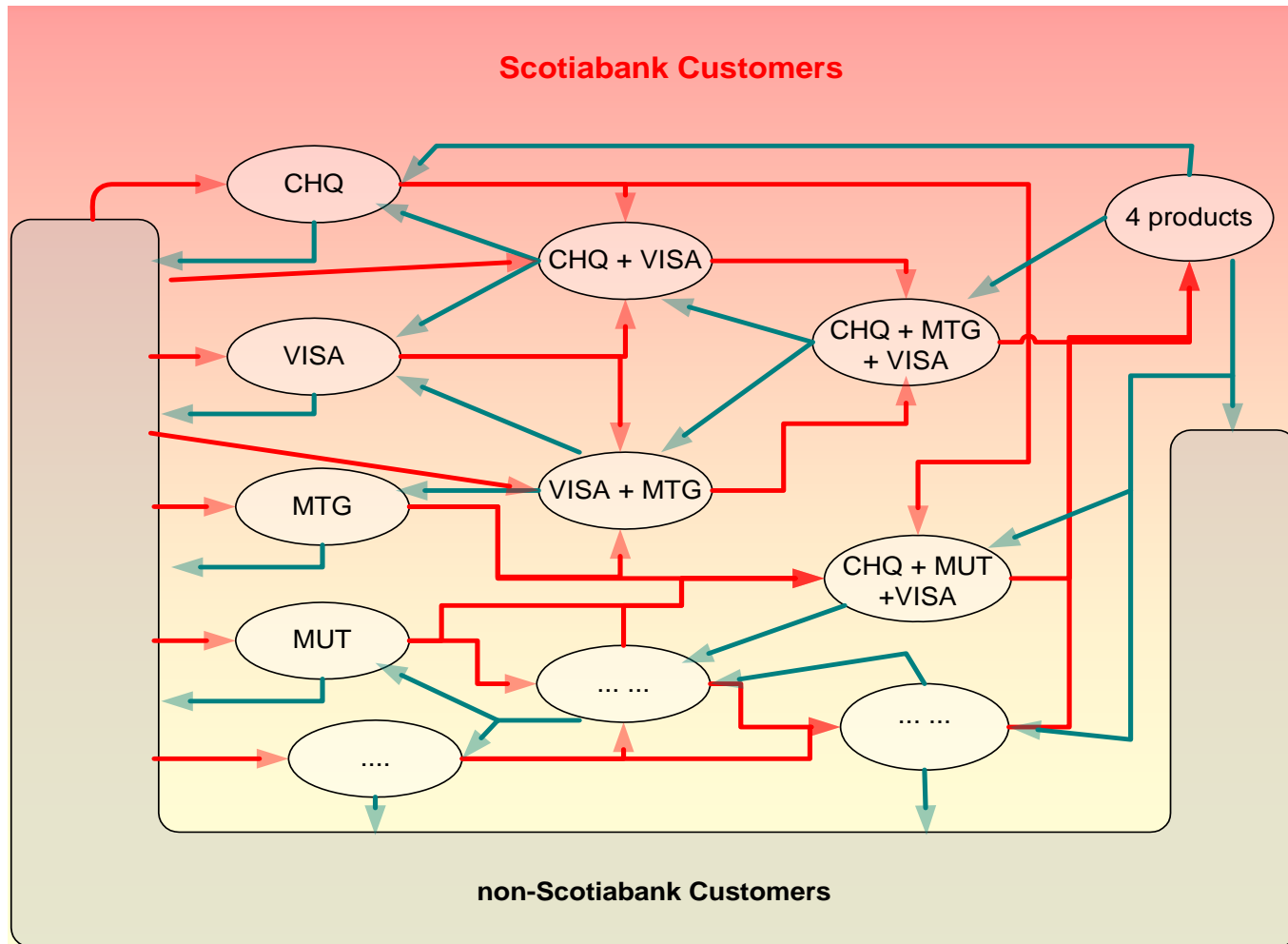
OFFER_DESC	# Leads	Balance Lift			Balance Lift by Product										
		Total	Average	Chequing	Money Mstr	Savings	GIC	Mutual Fund	RSP	RIP	Mortgage	VISA	Scotia Line	Loans	
*** All Campaigns ***	2,034,453	\$-28,233,743	\$-14	\$75,650,858	\$24,460,800	\$-20,914,678	\$-164,643,867	\$-11,105,617	\$1,688,217	\$-2,076,074	\$700,102	\$618,551	\$11,216,628	\$-3,742,170	
Mutual Fund Retention	7,444	\$1,139,733	\$151	-	-	-	-	\$3,139,733	-	-	-	-	-	-	
GIC Renewal-Ultimate	4,323	\$1,784,344	\$413	-	\$98,629	-	\$-978,082	\$1,089,235	\$233,650	\$-116,488	-	-	-	-	
Relationship Call	1,014	\$1,323,544	\$1,305	\$657,640	-	\$184,582	-	\$-31,268	\$18,908	\$-17,654	\$-8,663	\$596,441	\$-494,400	\$-191,914	\$9,734
RIP Acq. / Conv.	1,292	\$459,869	\$356	-	\$-92,647	-	-	\$-7,214	-	\$859,670	-	-	-	-	
STEP Usage	6,787	\$307,851	\$46	-	-	-	-	-	\$-659	\$1,004	\$-259,933	\$502,490	\$-69,620	\$43,810	
Scotiline Usage	237	\$236,788	\$979	-	-	-	-	-	-	-	-	\$-16,083	\$36,437	-	
Indirect Auto Loan Welcome Call-PA	679	\$277,366	\$410	\$-31,116	\$-1,189	\$2,867	\$-106,921	\$11,797	\$-13,928	\$-63	\$514,963	\$-29,334	\$-67,264	\$-3,467	
ET-Parked Funds Conversion	3,543	\$265,894	\$76	-	\$-1,247,328	-	\$1,983,480	\$-45,760	\$216,895	\$53,557	-	-	-	-	
VISA Usage-Scotiline	474	\$94,192	\$199	-	-	-	-	-	-	-	-	\$31,103	\$63,889	-	
RSP Consolidation & RIP Conversion	1,445	\$3,783	\$3	-	\$-143,291	-	-	\$-198,394	\$283,204	\$62,264	-	-	-	-	
Market Power GIC Acq.	22,422	\$0	\$0	-	-	-	\$0	\$0	\$0	\$0	-	-	-	-	
VISA Usage-Scotiline, RSP, or RIP	15,119	\$-46,019	\$-3	-	-	-	-	-	-	-	-	\$-41,070	\$-4,949	-	
Indirect Auto Loan Welcome Call-NBA	475	\$-452,333	\$-952	\$-20,298	\$7,149	\$-9,067	\$-31,403	\$-1,871	\$1,299	-	\$-241,933	\$13,080	\$-82,772	\$-86,591	
ICM Retention	11,553	\$-1,412,814	\$-122	\$-3,897,436	\$2,024,635	\$-91,523	\$-1,580,264	\$1,405,680	\$393,296	\$16,787	-	-	-	-	
Relationship Call: Blue Print Prospects	1,196	\$-2,389,634	\$-1,970	\$141,362	\$1,196,312	\$-145,216	\$-243,698	\$-1,454,819	\$129,216	\$39,254	\$-1,519,976	\$-33,774	\$-268,146	\$-178,139	
ET-Significant Deposit	48,502	\$-31,785,235	\$-650	\$76,620,354	\$22,160,500	\$-21,056,028	\$-164,272,209	\$-15,343,185	\$2,450,142	\$-2,679,431	\$1,619,939	\$89,503	\$13,551,796	\$-5,527,623	

Lift in Account by Offer
From 01NOV2007 to 30NOV2007

OFFER_DESC	# Leads	Account Lift			Account Lift by Product									
		Total	Average	Chequing	Money Mstr	Savings	GIC	Mutual Fund	RSP	RIP	Mortgage	VISA	Scotia Line	Loans
*** All Campaigns ***	2,034,453	-1,374	0.0007	34	-121	-100	-1,021	-474	-19	9	-93	269	125	3
Indirect Auto Loan Welcome Call-PA	675	118	0.1755	5	5	0	1	-7	-1	0	2	96	4	-1
RIP Acq. / Conv.	1,292	106	0.0820	-	9	-	-	-	1	-109	-	-	-	-
ICM Retention	11,553	92	0.0080	72	-3	21	-18	17	2	3	-	-	-	-
ET-Parked Funds Conversion	3,543	19	0.0052	-	1	-	60	-12	-28	7	-	-	-	-
Indirect Auto Loan Welcome Call-NBA	475	12	0.0253	8	4	1	-1	-1	0	-	-1	6	-3	-1
STEP Cross-sell	624	11	0.0174	-	-	-	-	-	-	-	-	1	8	1
ET-Last Loan Payment PA	189	9	0.0458	1	0	0	5	0	4	0	0	1	2	-4
Mortgage Acquisition	305,820	6	0.0020	-	-	-	-	-	-	-	6	-	-	-
Market Power GIC Acq.	22,422	0	0.0000	-	-	-	-	0	-	-	-	-	-	-
Deposit Account - Powerchequing, 12 mth	32,076	0	0.0000	0	0	-	-	-	-	-	-	-	-	-
Moneyback VISA PA: Double moneyback for	5,602	0	0.0000	-	-	-	-	-	-	-	-	0	-	-

Scotiabank Confidential Customer Knowledge & Insights Page 3

Anybody Likes to Have a Markov Transition Report?



Summary & Conclusion

- Data mining is the process of turning data into knowledge and information with clear understanding of business needs.
- Business analytics is to provide effective decision support using data mining, optimization, data analysis, and reporting.
- Successful business analytics often requires continuous “Initiate, Engage, Execute, and Learn”.

