

Analytics in Business Environment

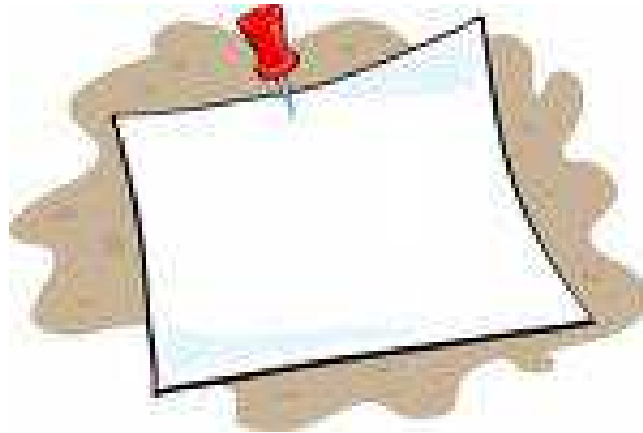
dr. Aleksander Pivk

SAS Adriatic Region
(aleksander.pivk@adr.sas.com)

Jozef Stefan Institute
(aleksander.pivk@ijs.si)


Ljubljana, 09.03.2012

Who knows



- A piece of paper is **0.193 mm** thick
- If the paper is folded **40 times**, use your “gut feel” **to forecast** the thickness of the folded sheet
- **ANSWER:** $0.193 \times (2^{40}) = 212.205,74 \text{ km}$
(Halfway to the moon!)

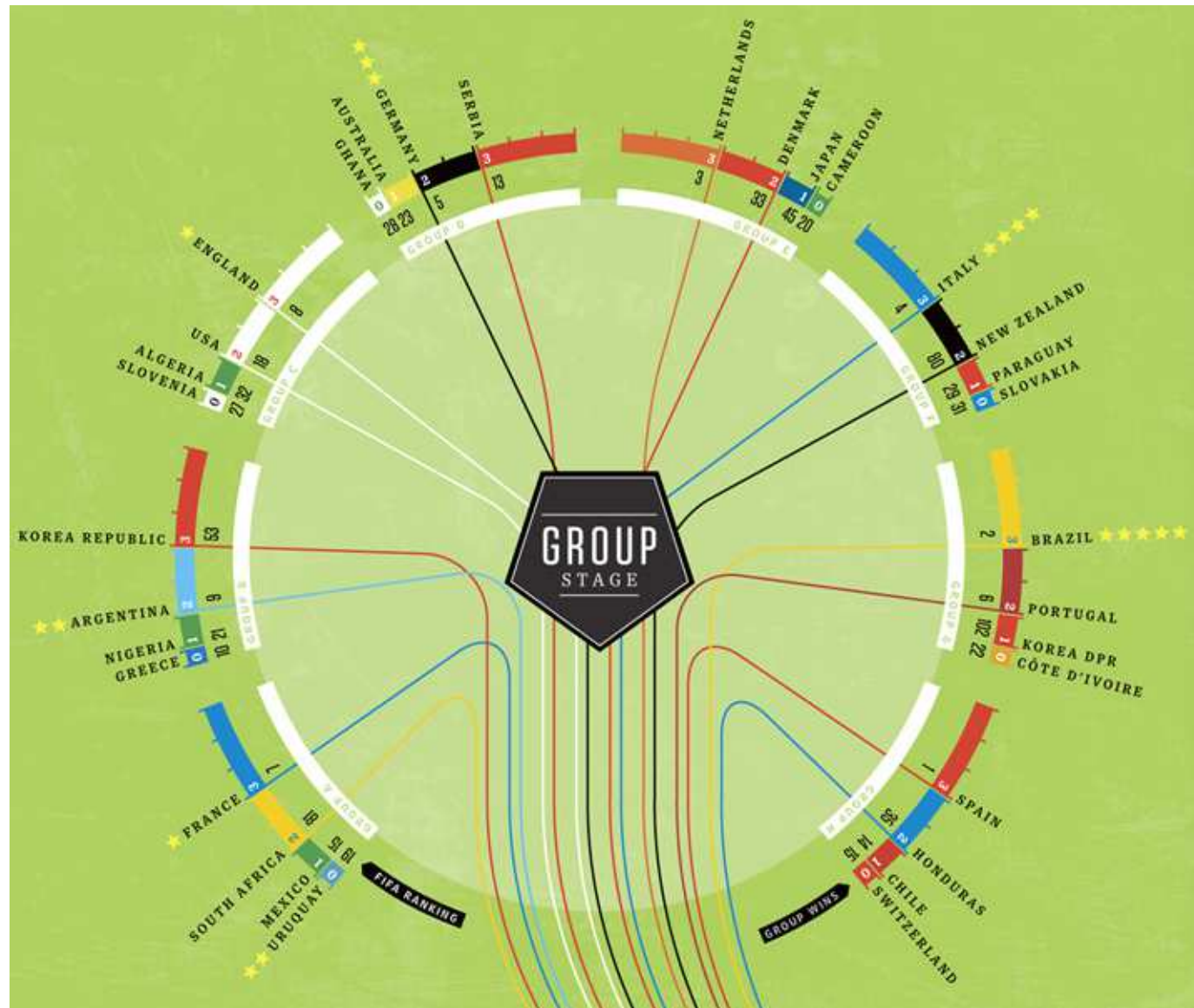
Svetovno prvenstvo v nogometu, JAR 2010

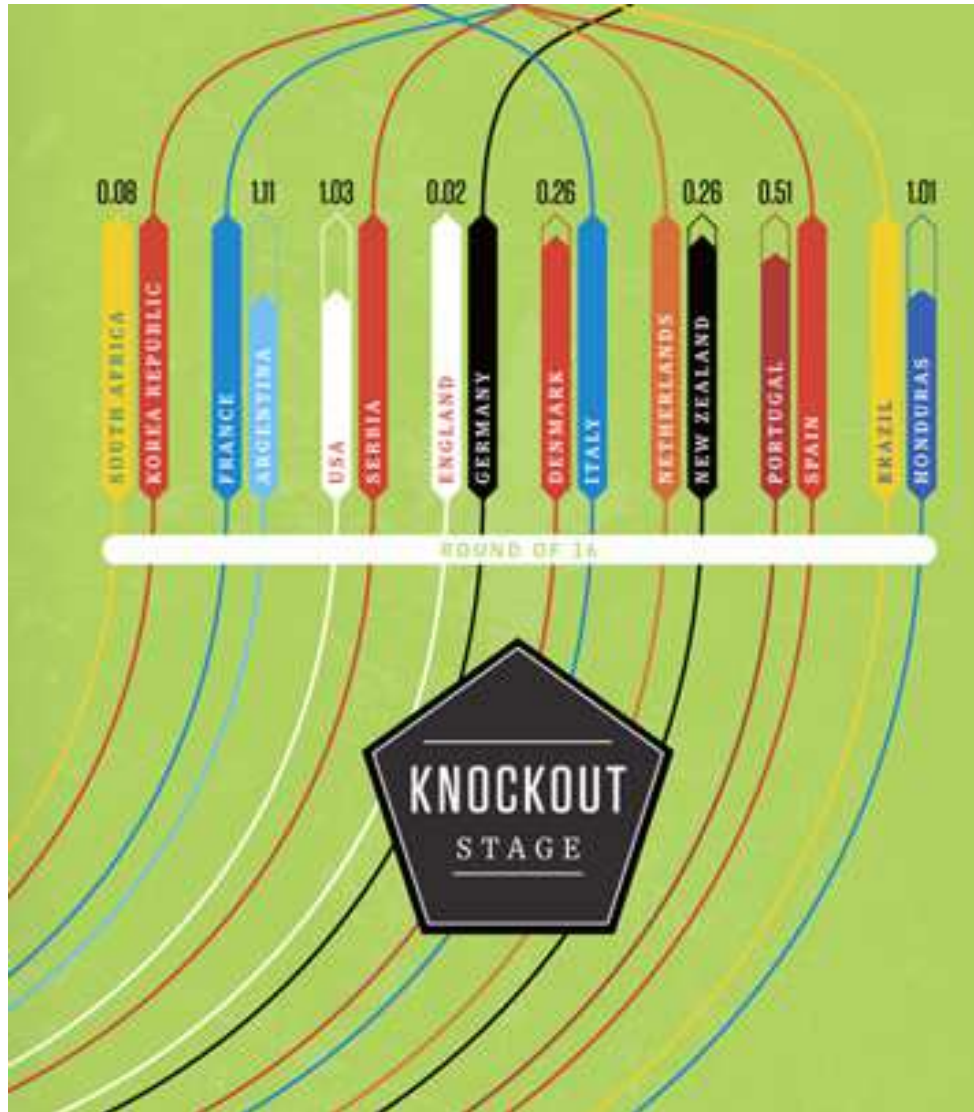
Group A	Group B	Group C	Group D	Group E	Group F	Group G	Group H
Team	Team	Team	Team	Team	Team	Team	Team
South Africa	Argentina 	England 	Germany 	Netherlands 	Italy 	Brazil 	Spain 
Mexico	Nigeria	USA	Australia	Denmark	Paraguay	Korea DPR	Switzerland
Uruguay	Korea Repu	Algeria	Serbia	Japan	New Zealand	Côte d'Ivoire	Honduras
France 	Greece	Slovenia	Ghana	Cameroon	Slovakia	Portugal	Chile

- 32 ekip / 8 skupin
- Napredujeta najboljši ekipi (2) iz vsake skupine
- Kateri ekipi bosta igrali v **finalu**?
- Katera ekipa bo **novi svetovni prvak**?

Kaj predvideva analitični model?

- avtorja: Simon Kuper, Stefan Szymanski
- vhodni podatki (predvsem ekonomski faktorji)
 - Populacija države
 - BDP na prebivalca
 - Izkušnost ekipe
- model
 - $RG(i, j) = 0,137 \log\left(\frac{pop(i)}{pop(j)}\right) + 0,145 \log\left(\frac{bdp(i)}{bdp(j)}\right) + 0,739 \log\left(\frac{izk(i)}{izk(j)}\right) (+0,657)$
 - točnost: 72%
- model prilagojen glede na vse mednarodne tekme med letom 1980 in 2001



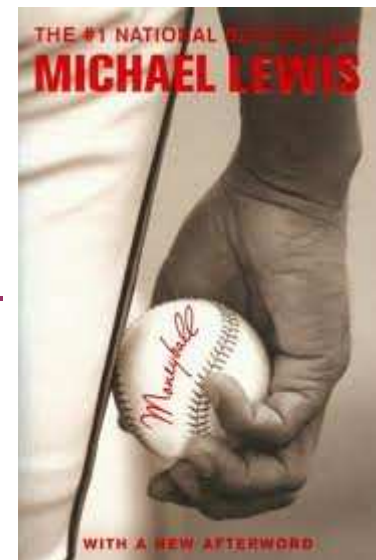
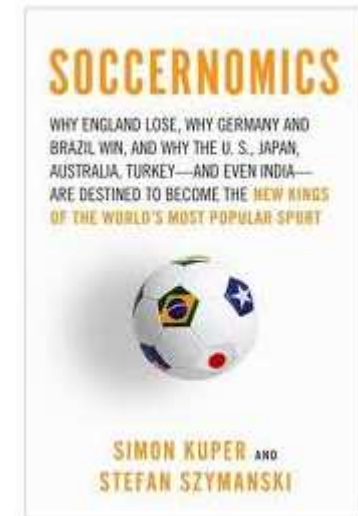


Actual results



Interesting sports analytics literature

- Soccernomics: Why England Loses, Why ...,
- Moneyball: The Art of Winning an Unfair Game
- Beyond 'Moneyball': The rapidly evolving world of sports analytics, Part I

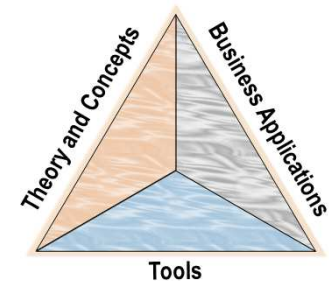


Analytics

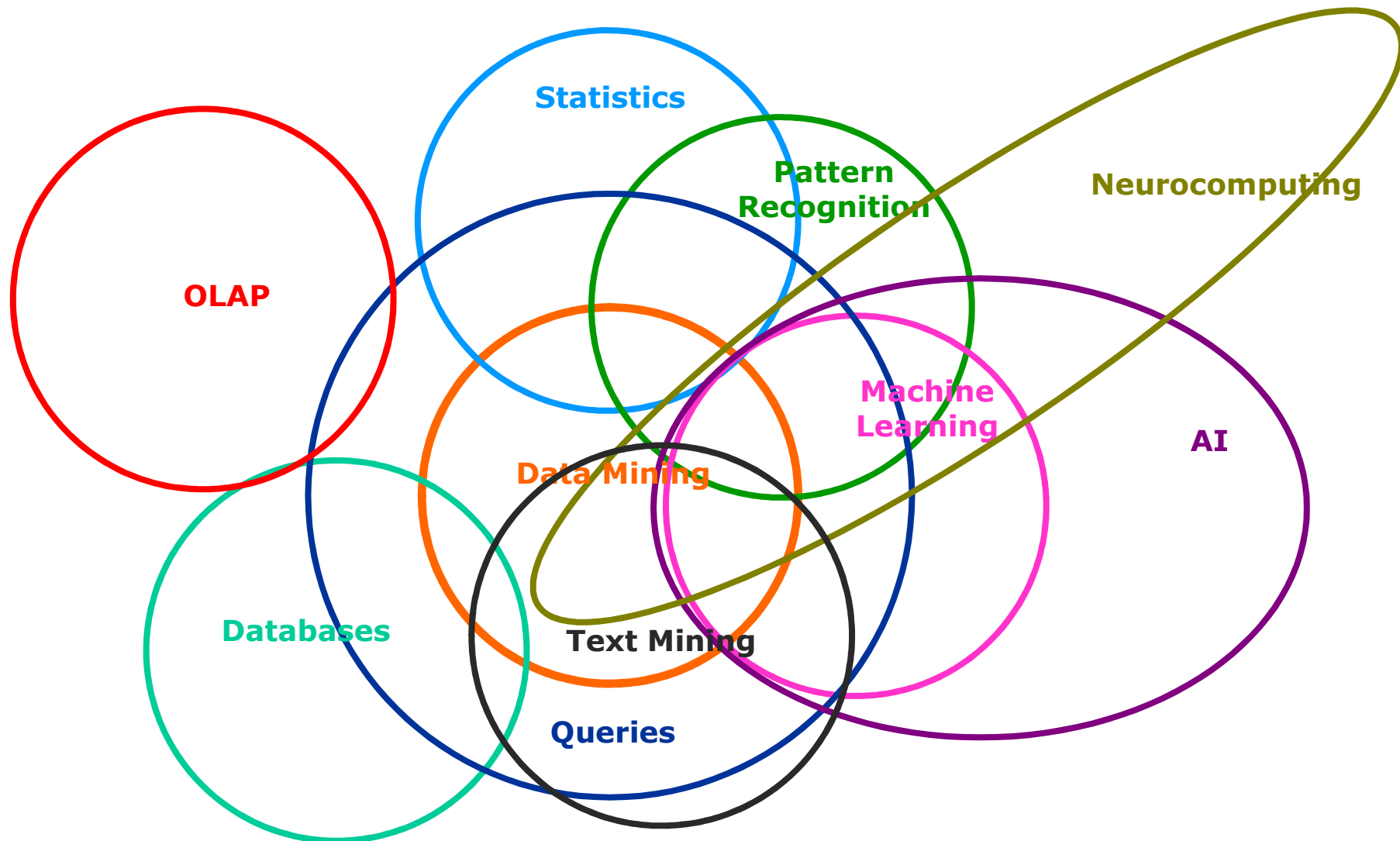
- **the goal** for employing analytics is to be able to **make better and more accurate decisions quickly**
- is **the process** of obtaining an optimal or realistic **decision** based on extensive usage of **existing data, statistical and quantitative analysis, explanatory and predictive modeling, and fact-based management**
- refers to the **skills, technologies, applications and practices** for continuous iterative exploration and investigation of past performance to gain insight and drive future actions

Those who ~~KNOW~~ first, WIN

Those who **KNOW and ACT** first, WIN
(provided they act intelligently)



What is “Analytics”?





EIGHT LEVELS OF ANALYTICS

1



Number	Report Date	Expenditures
25-Oct-06		\$739.55
10-Oct-06		\$449.59
29-Sep-06		\$1,031.95
22-Sep-06		\$455.09
13-Sep-06		\$170.88
23-Aug-06		\$1,009.03
21-Aug-06		\$154.96
17-Aug-06		\$812.91
27-Jun-06		\$510.76
24-Jun-06		\$699.14
25-May-06		\$991.20
3-May-06		\$994.25
23-Feb-06		\$789.63
24-Jan-06		\$375.00
19-Dec-05		\$84.00
15-Dec-05		\$590.11
24-Nov-05		\$1,333.54
22-Nov-05		\$200.10
19-Nov-05		\$15.00
13-Nov-05		\$495.00
grand Total		\$11,902.09

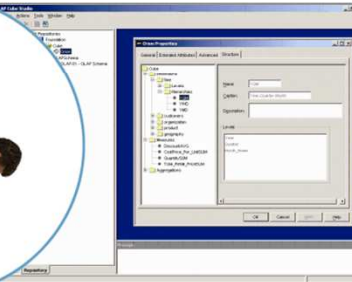
STANDARD REPORTS

Answer the questions: What happened? When did it happen?

Example: Monthly or quarterly financial reports.

We all know about these. They're generated on a regular basis and describe just "what happened" in a particular area. They're useful to some extent, but not for making long-term decisions.

2



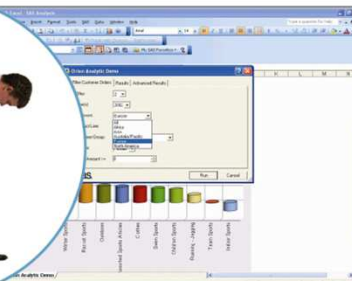
AD HOC REPORTS

Answer the questions: How many? How often? Where?

Example: Custom reports that describe the number of hospital patients for every diagnosis code for each day of the week.

At their best, ad hoc reports let you ask the questions and request a couple of custom reports to find the answers

3



QUERY DRILLDOWN (OR OLAP)

Answer the questions: Where exactly is the problem? How do I find the answers?

Example: Sort and explore data about different types of cell phone users and their calling behaviors.

Query drilldown allows for a little bit of discovery. OLAP lets you manipulate the data yourself to find out how many, what color and where.

4



ALERTS

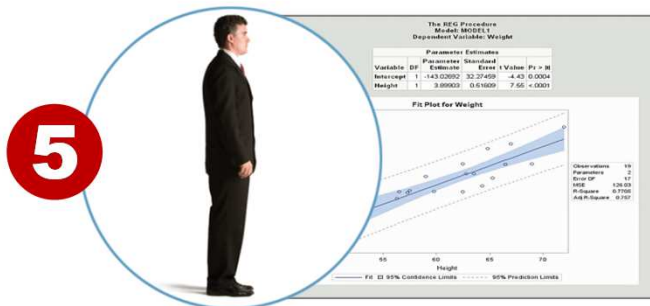
Answer the questions: When should I react? What actions are needed now?

Example: Sales executives receive alerts when sales targets are falling behind.

With alerts, you can learn when you have a problem and be notified when something similar happens again in the future. Alerts can appear via e-mail, RSS feeds or as red dials on a scorecard or dashboard.



...reporting...



STATISTICAL ANALYSIS

Answer the questions: Why is it happening? What opportunities am I missing?

Example: Banks can discover why an increasing number of customers are refinancing their homes.

Here we can begin to run some complex analytics, like frequency models and regression analysis. We can begin to look at why things are happening using the stored data and then begin to answer questions based on the data.

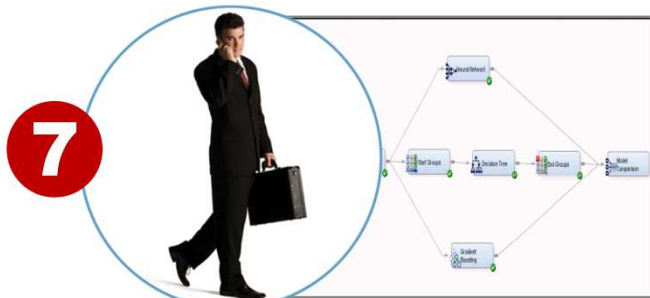


FORECASTING

Answer the questions: What if these trends continue? How much is needed? When will it be needed?

Example: Retailers can predict how demand for individual products will vary from store to store.

Forecasting is one of the hottest markets – and hottest analytical applications – right now. It applies everywhere. In particular, forecasting demand helps supply just enough inventory, so you don't run out or have too much.

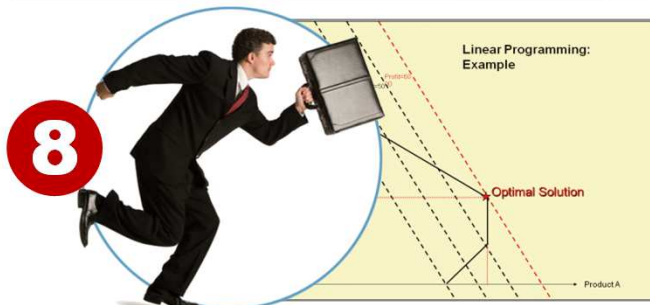


PREDICTIVE MODELING

Answer the questions: What will happen next? How will it affect my business?

Example: Hotels and casinos can predict which VIP customers will be more interested in particular vacation packages.

If you have 10 million customers and want to do a marketing campaign, who's most likely to respond? How do you segment that group? And how do you determine who's most likely to leave your organization? Predictive modeling provides the answers.



OPTIMIZATION

Answer the questions: How do we do things better? What is the best decision for a complex problem?

Example: Given business priorities, resource constraints and available technology, determine the best way to optimize your IT platform to satisfy the needs of every user.

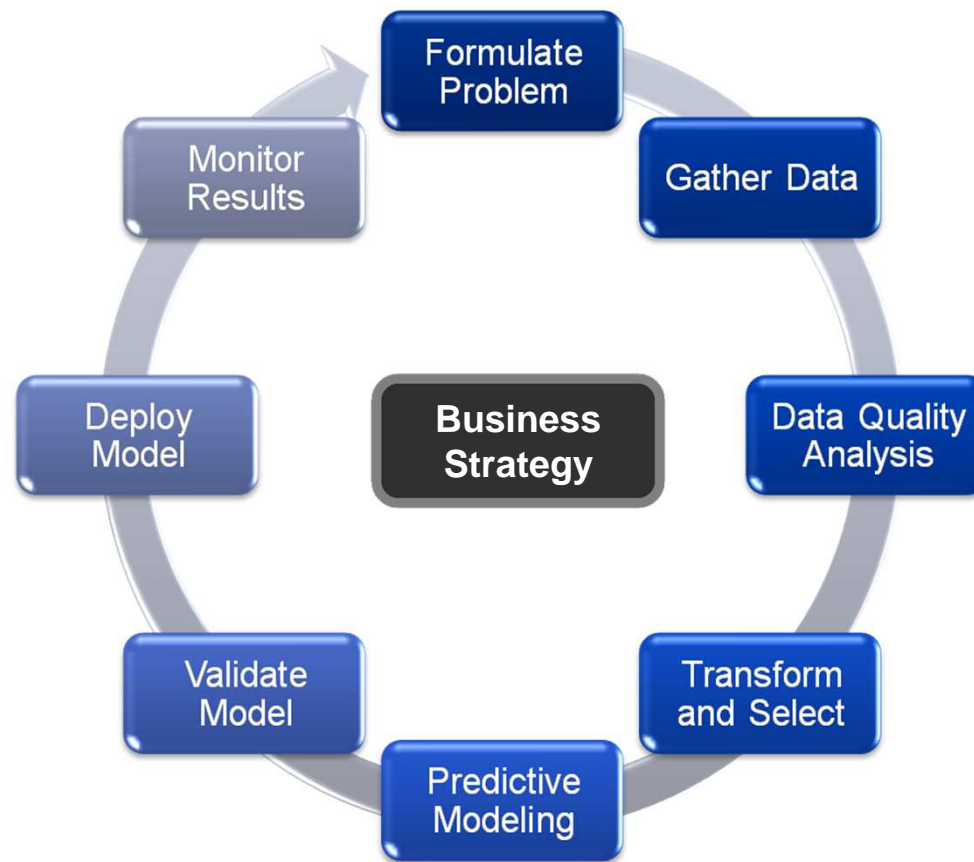
Optimization supports innovation. It takes your resources and needs into consideration and helps you find the best possible way to accomplish your goals.

... analytics ...

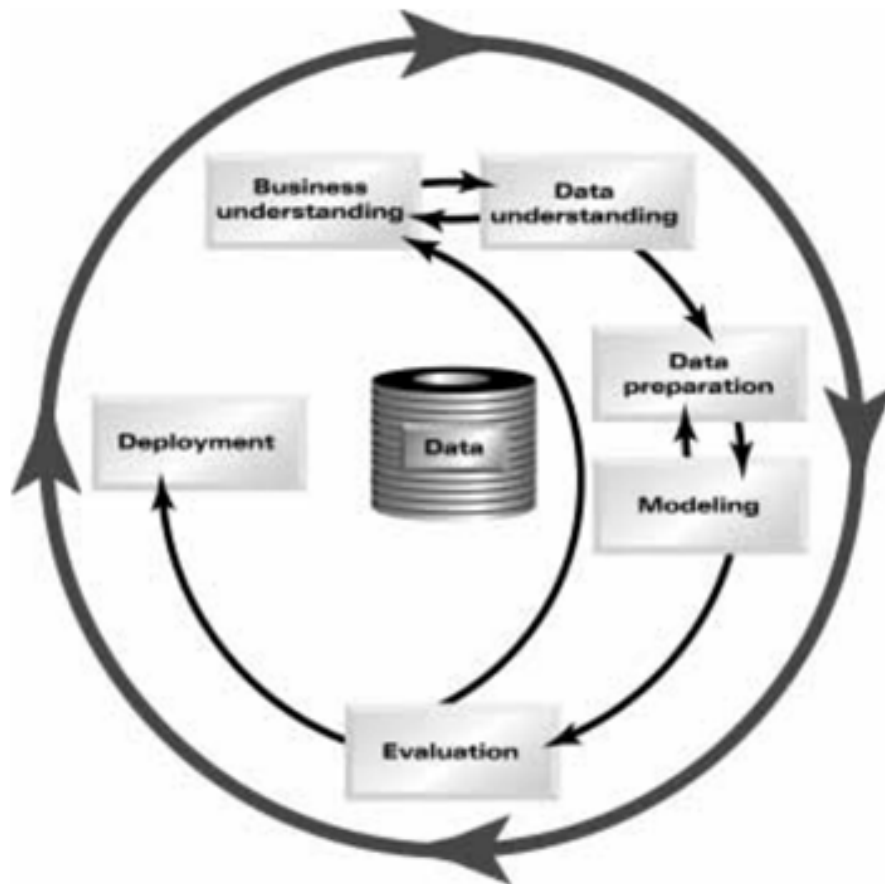


Successful Analytics

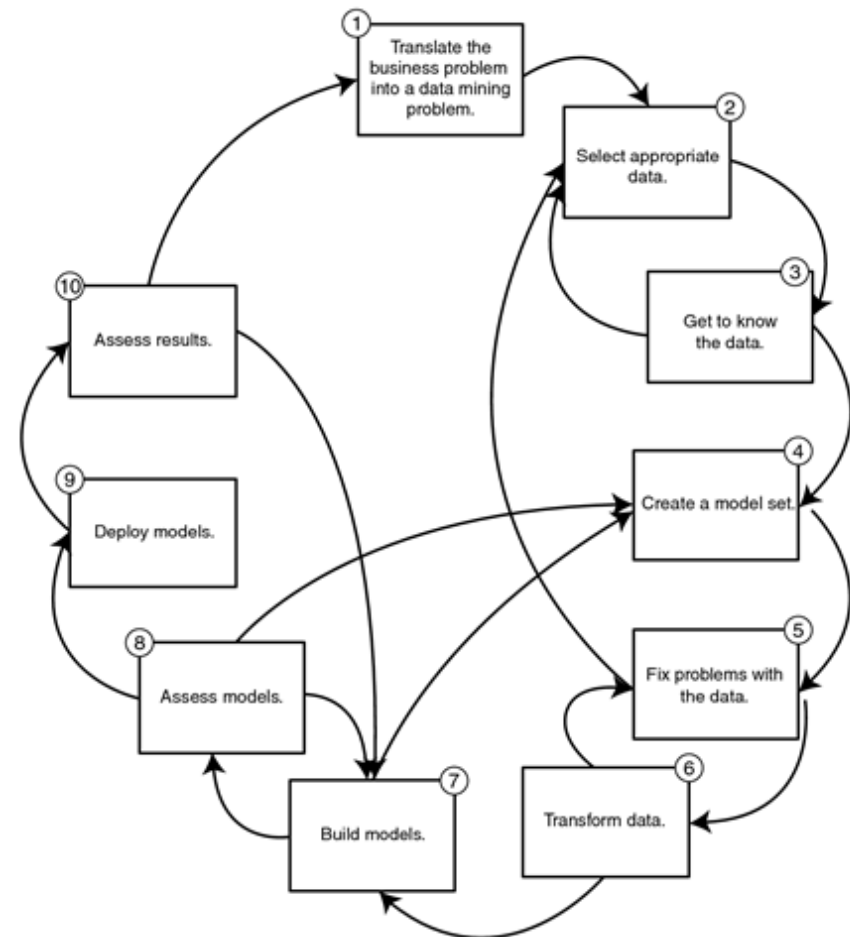
Integration of People, Processes, and Technology
aligned with Business Strategy



Other relevant methodologies

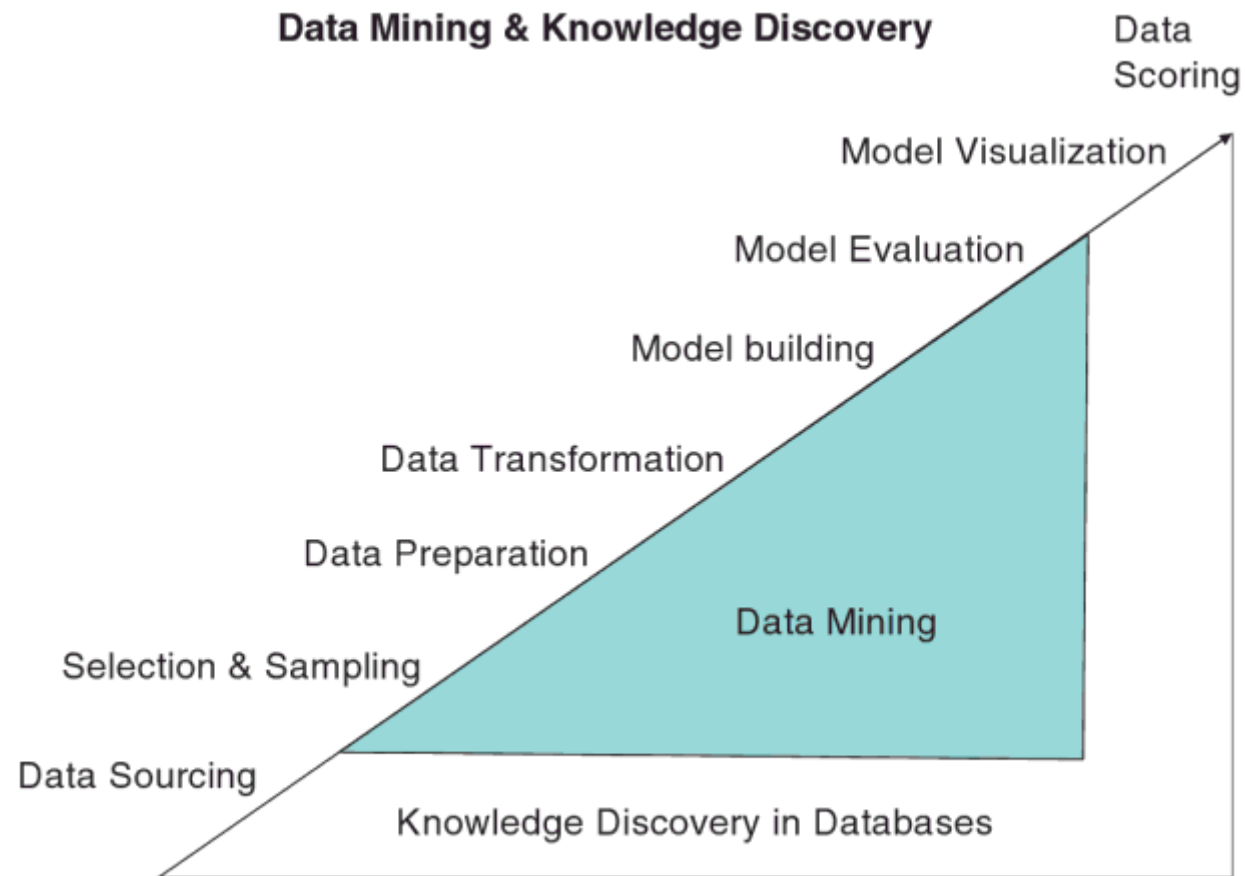


Shearer C., The CRISP-DM model: the new blueprint for data mining. J Data Warehousing 2000;5:pp13-22.

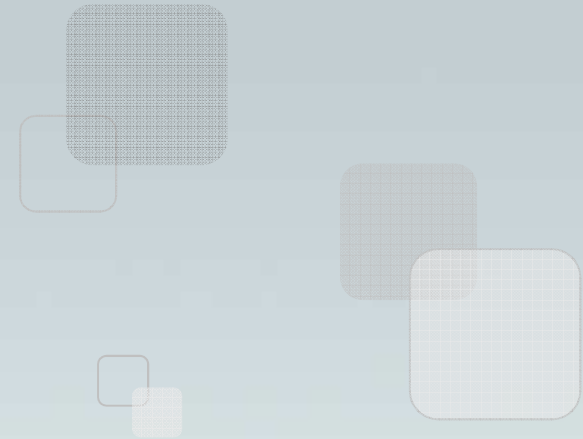


Linoff, Berry: Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, 2011, ISBN 978-0470650936

Other relevant methodologies



Nisbet R., Elder J., Miner G., Handbook of Statistical Analysis and Data Mining Applications, Academic Press, 2009, ISBN 978-0123747655



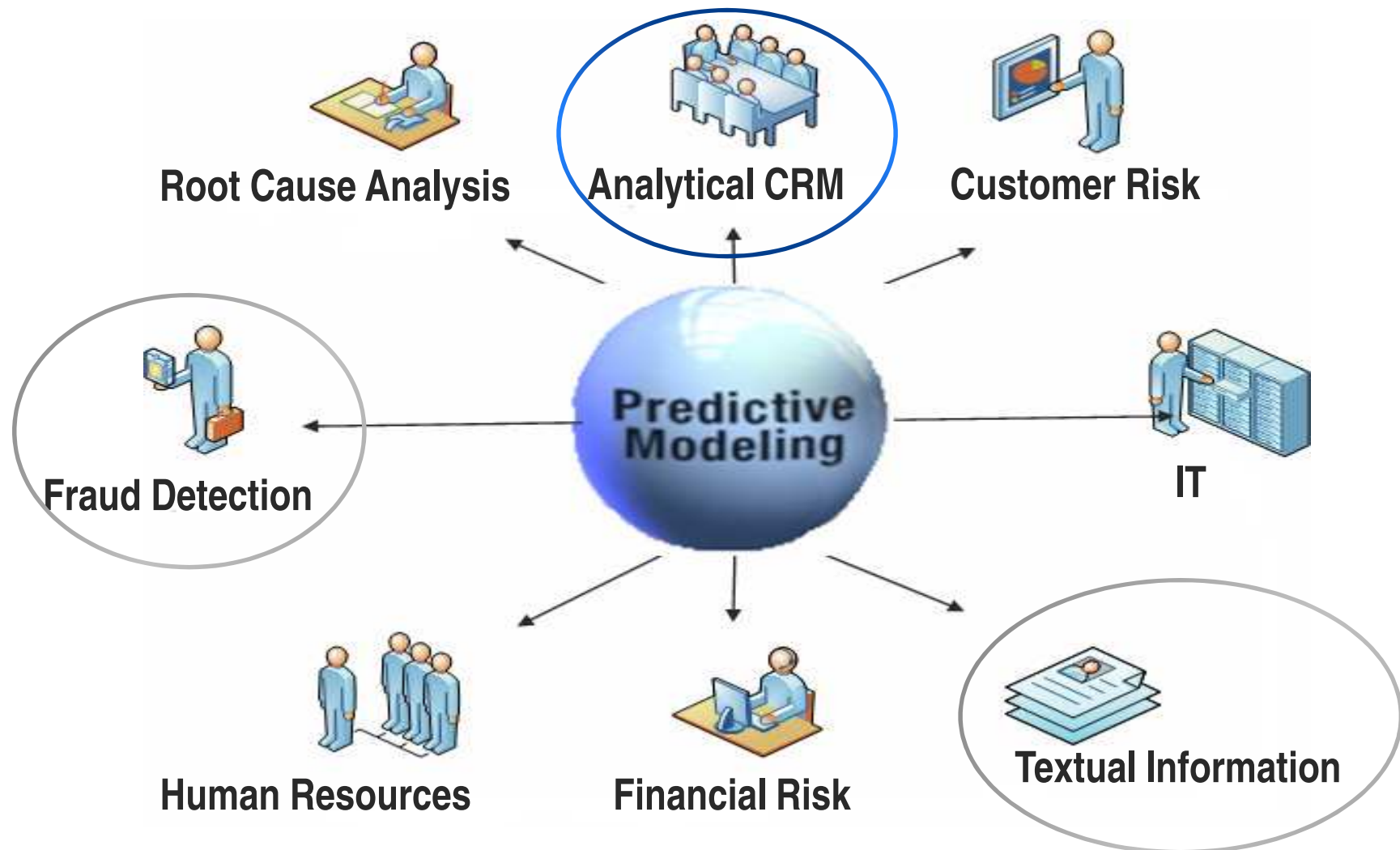
Business Applications



Business Application Areas

What do you think are some of the most important and widespread **business applications** of Analytics?

Analytical Needs by Business Area



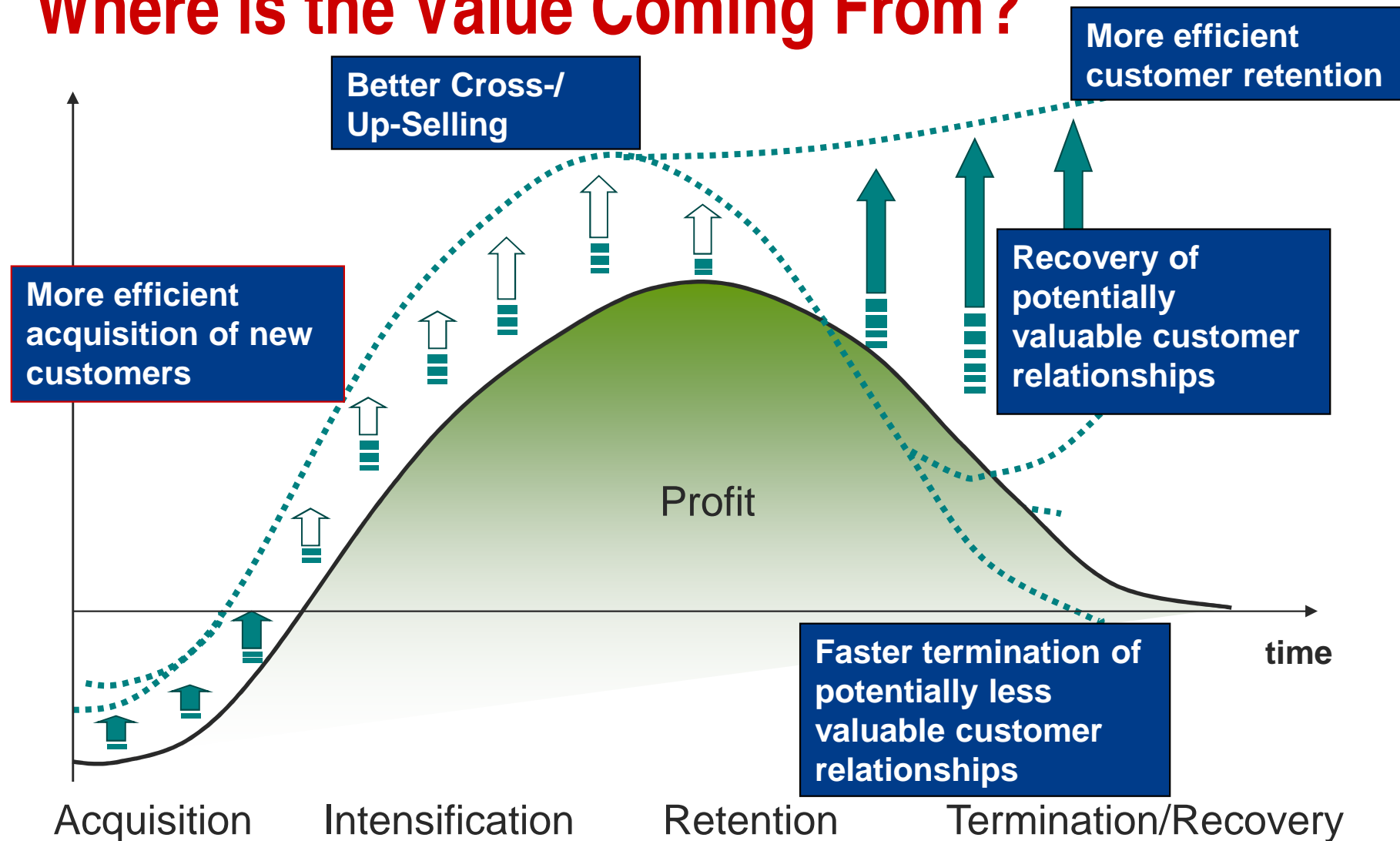
Customer Intelligence (aCRM), **Definition**

Customer Intelligence is the concept of **gathering, analyzing and exploiting information** on company's customers with an aim to create and maintain a mutually beneficial relationship, whereby the company aims to maximize the **Value** of the relationship.

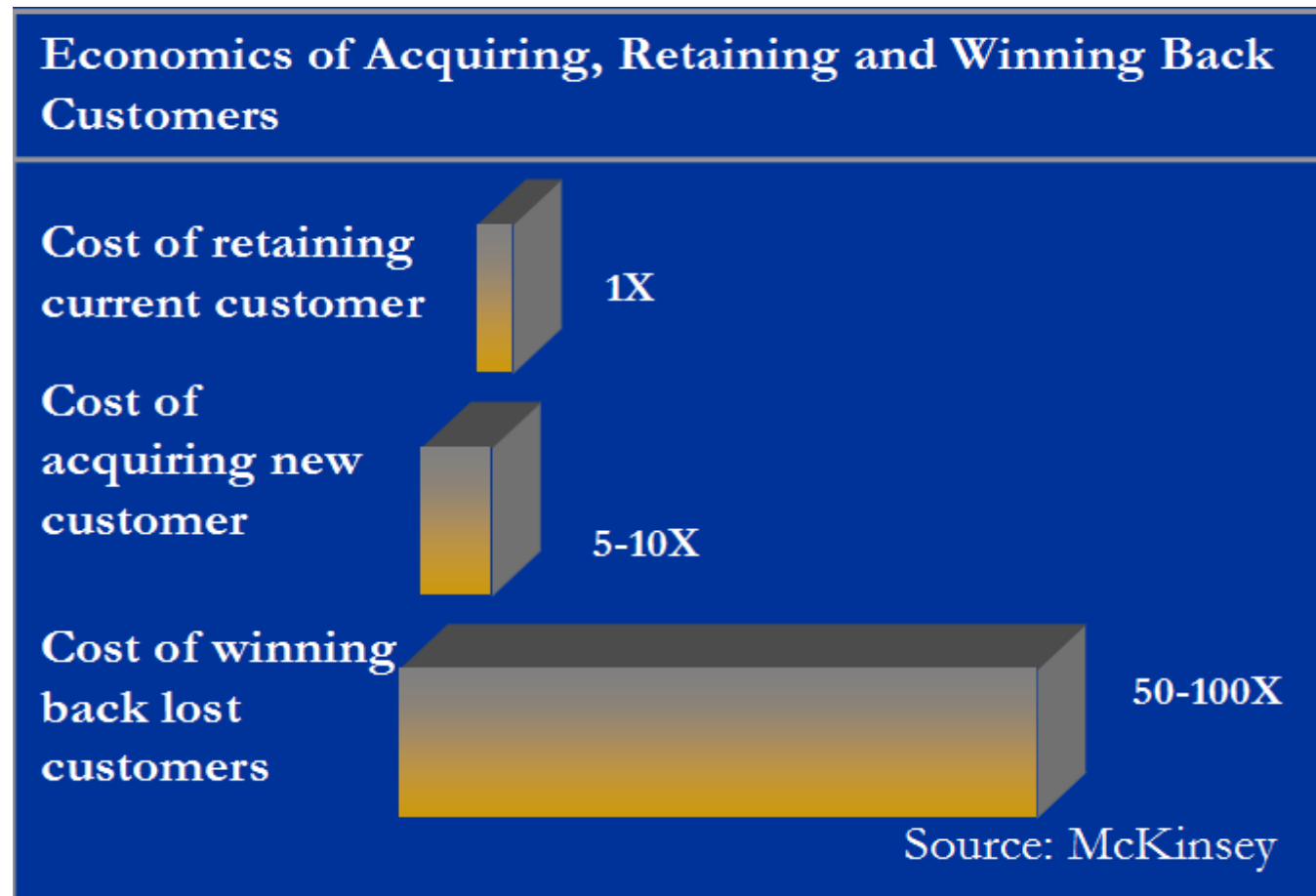
This Value must always correspond to an amount expressed in real **Money** (i.e. earned or saved, profit or income).

Hence, the process must always be **Measured**.

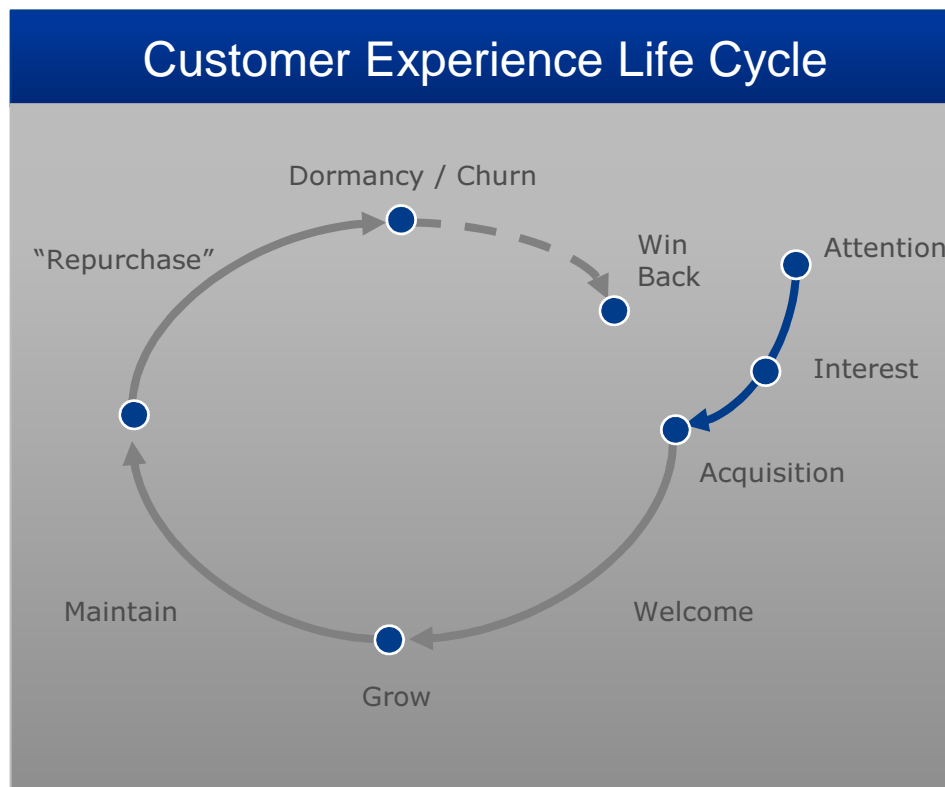
Customer Intelligence, Where is the Value Coming From?



Economics of Acquiring, Retaining and Win-Back Customers



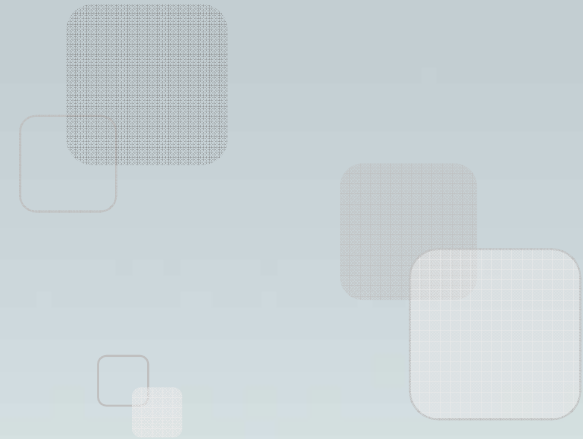
Customer Life Cycle



- A complete customer experience process mimicking the individual life cycle
- Prioritizations, actions and campaigns predefined and highly automated
- Achieve the right effort at the right time for the right customer
- Be proactive and let customer events trigger actions
- Managed through a BI and marketing platform driven by analytical customer insight

Customer Intelligence, **Key Principles**

- One to One marketing strategy
- Collaboration and dialogue with customers
- Customer differentiation, not just product differentiation
- Share of customer, not just share of market
- Taking products to customers, not customers to products



Descriptive modeling / Segmentation

Customer Segmentation, **Definition**

Segmentation is in essence the process by which items or subjects are categorized or classified into **groups with similar characteristics**. That characteristic could be one or more attributes.

Randal S. Collica

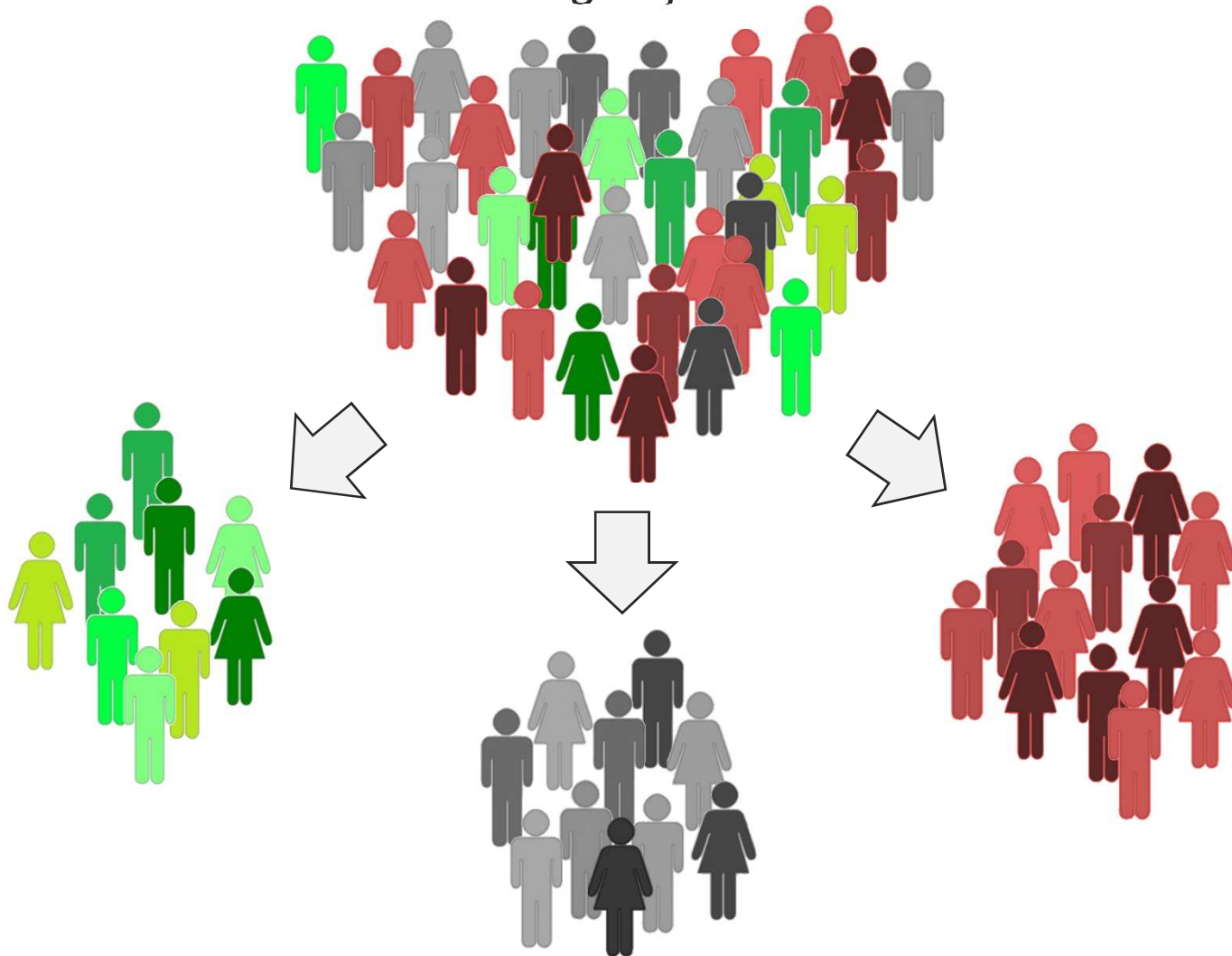
book: CRM Segmentation and Clustering

Types of segmentations

Type	Definition	Typical Variables	Method
Existing	Single or multiple dimension view of the customer	Income, Loyalty, Activity, Channel, Cross sell indicator	Frequency distribution, Complex filtering rules
Behavioural	Behave in a similar manner in relation with the business	Volumes of loans Number of transactions Number of products, Usage of limits...	Clustering
Marketing	Share characteristics that cause them to demand similar product and/or services.	Age, Occupation, Family status, Living area, Hobbies	Clustering
Tactical Operational	Groups of customers related to an action (xsell, upsell, retain)	Behavioural segment Marketing segment, Single dimensions	Predictive modelling, Filtering

Customer Segmentation, **Example**

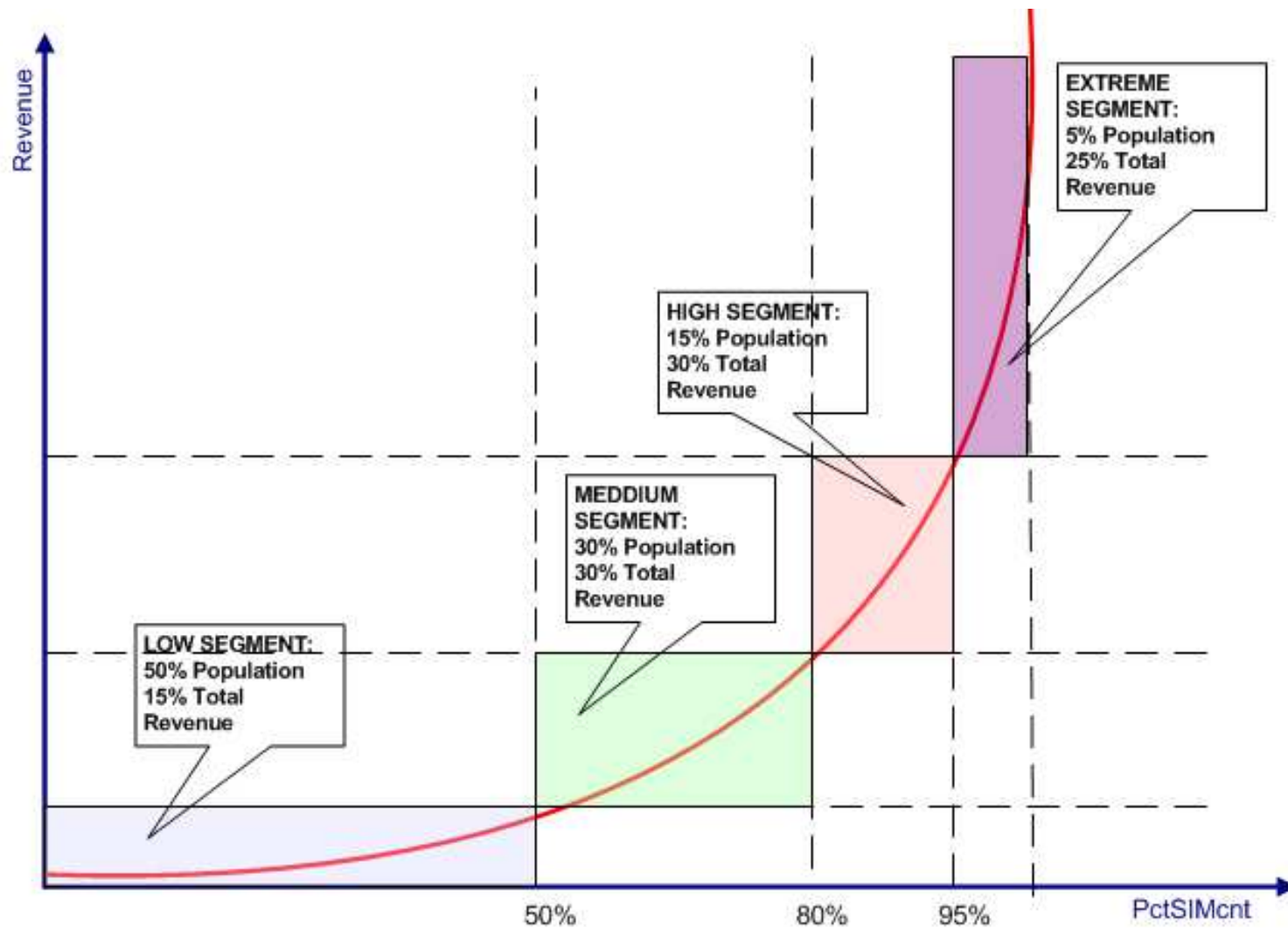
“Create groups that have similar characteristics. Also a measure of how ‘different’ each group is from the others”



Customer Segmentation, **Methods**

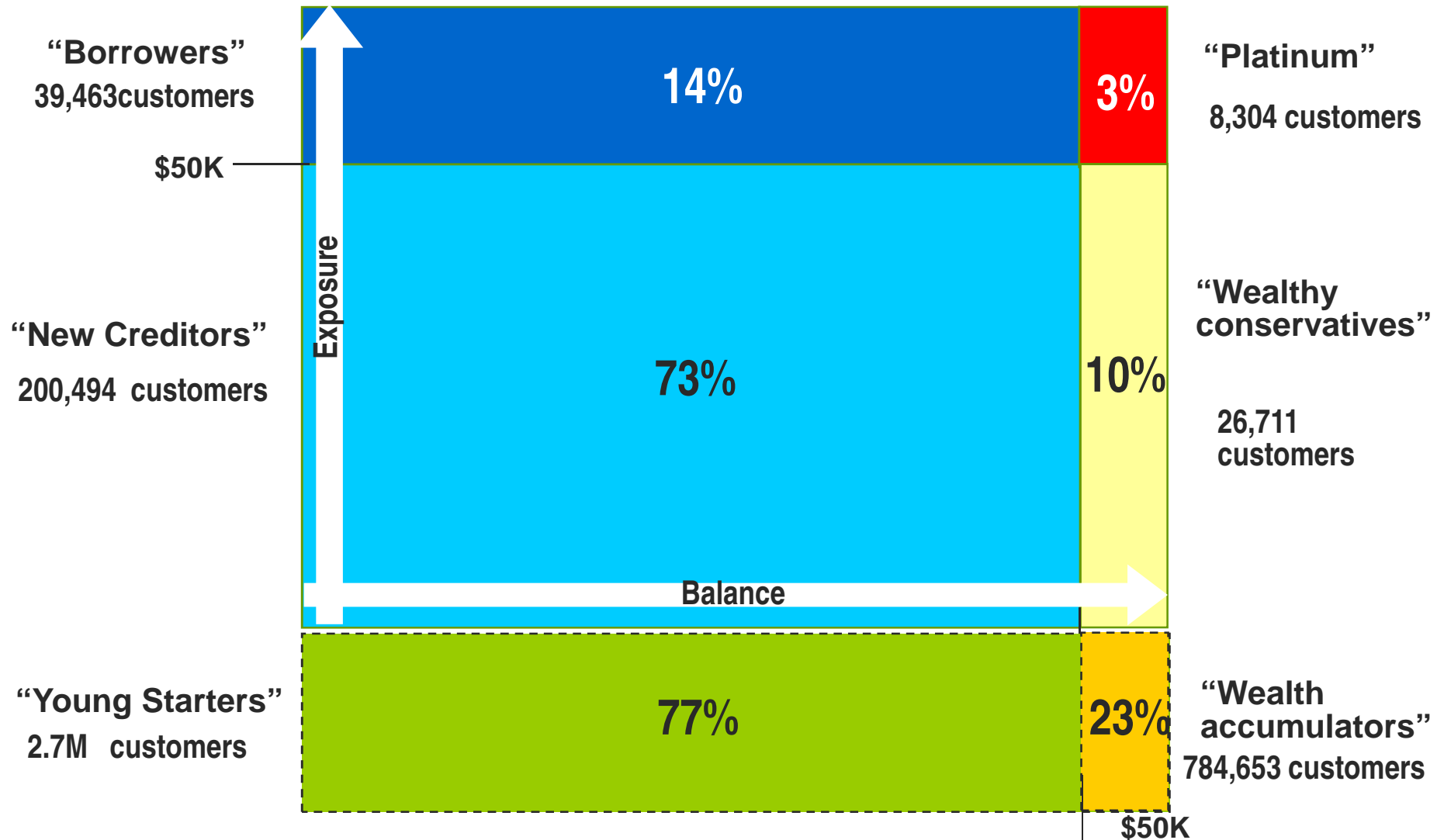
- **Supervised (directed) Segmentation** – business analyst defines one or more target variables that should drive the segmentation
- **Unsupervised (undirected) Segmentation** – analytical algorithm uncovers hidden patterns that may be significant and useful for the given purpose.

Revenue segmentation



Directed Segmentation – Example

Financial Relationship Segments



Good segment definition

- **Meaningful and Actionable**

Easy to explain and name

Homogeneous (clearly distinguish between clients in different segments, but not too discriminative: alike on most / not all of the criteria)

Align to business purpose (i.e. Assist in preparing the most adequate offer for targeting, e.g. dormancy prevention)

Easy to define segment development strategy (favorable and unfavorable migrations)

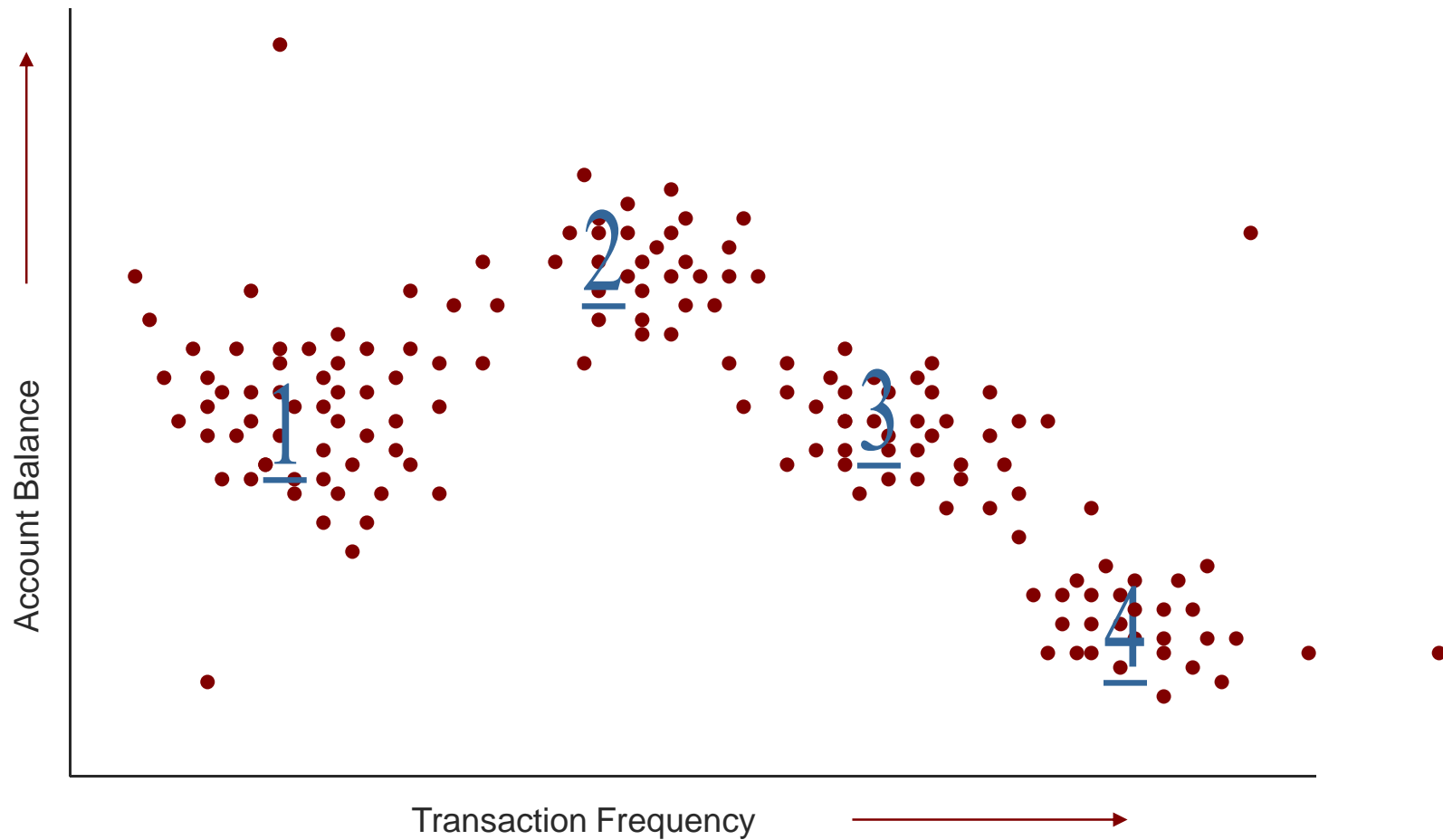
- **Not too Granular**

Additional splits within segments rather than large number of segments. Assist in viewing of large number of clients – Not pin-point a single client.

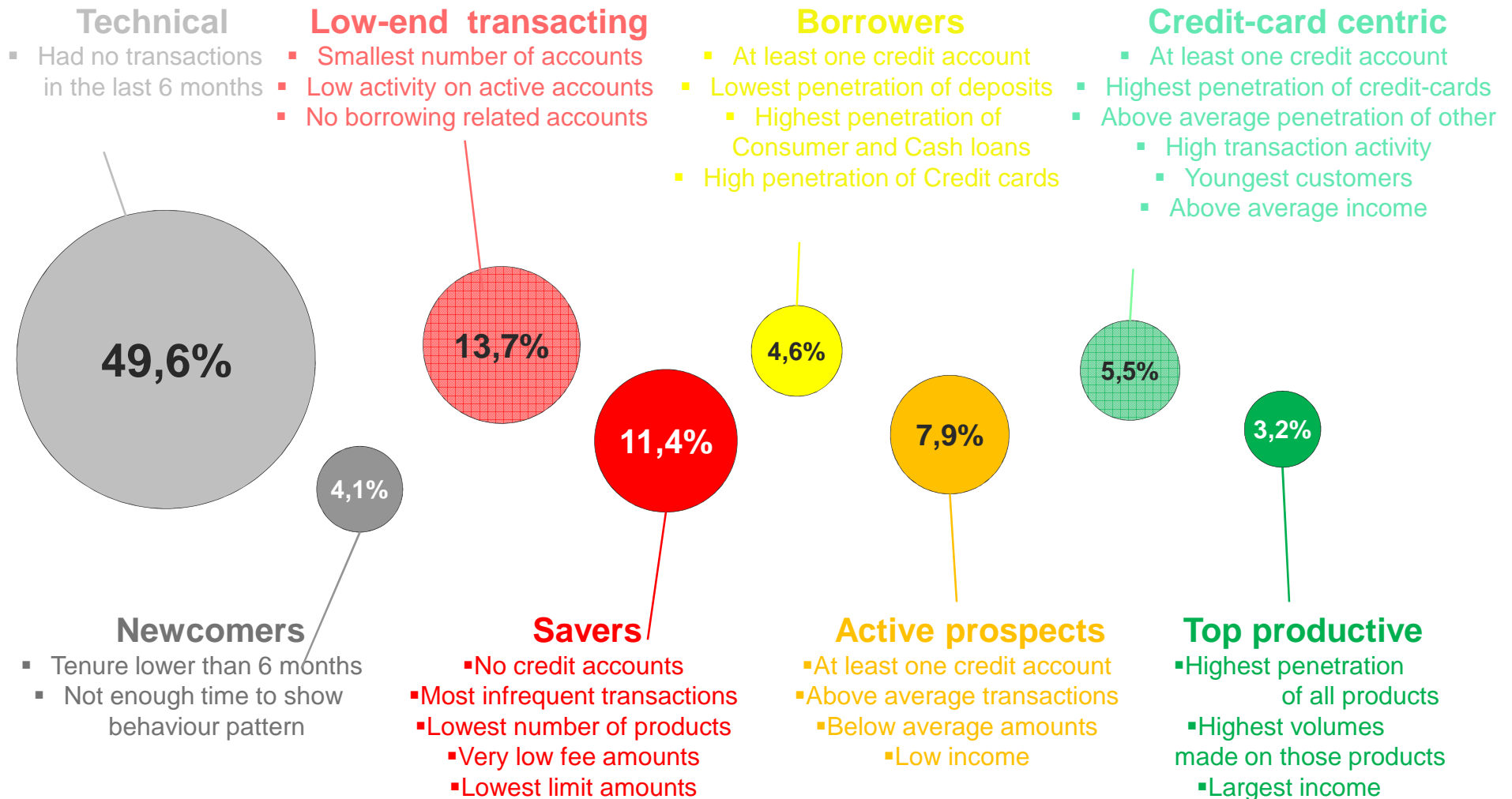
- **Stable over Time**

All migrations among profiles can be explained in business language (i.e. The client opened additional products and is now receiving salary into another bank)

Customer Segmentation, **Clustering Example**

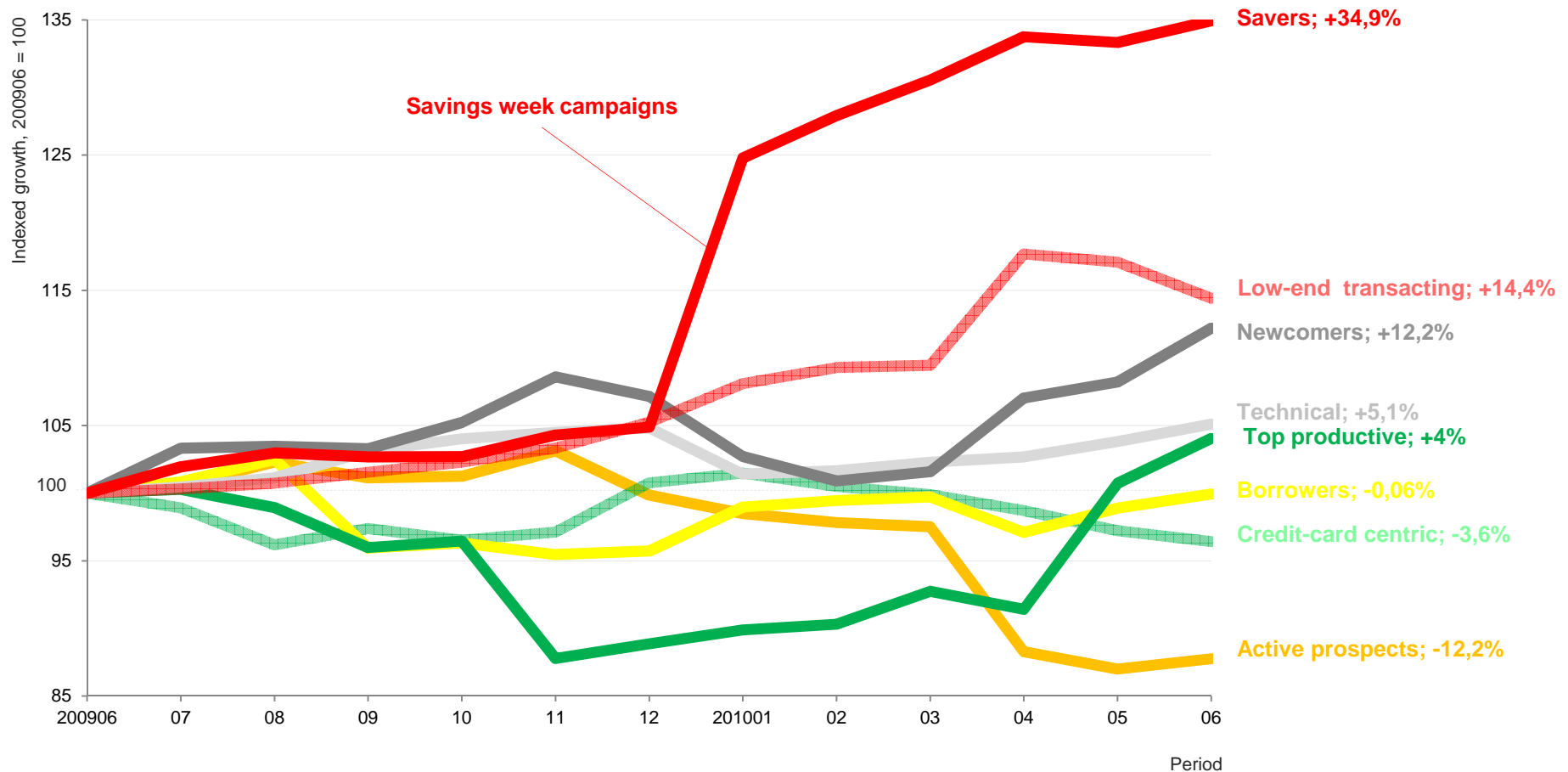


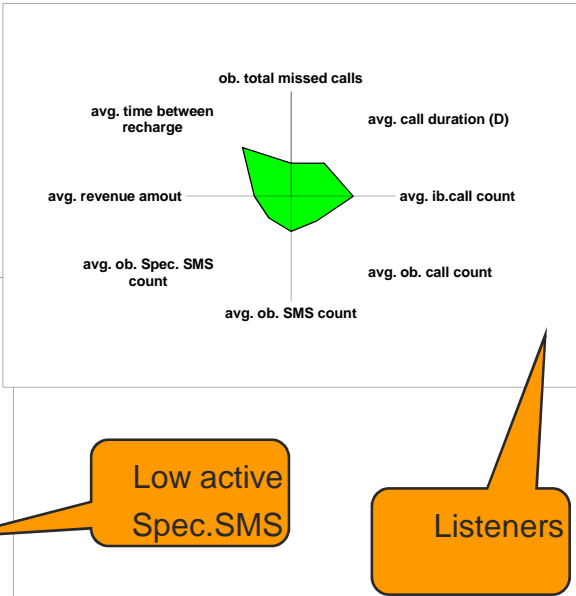
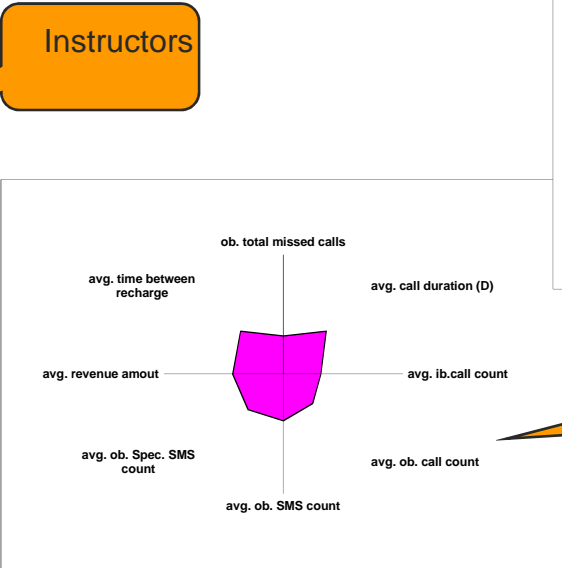
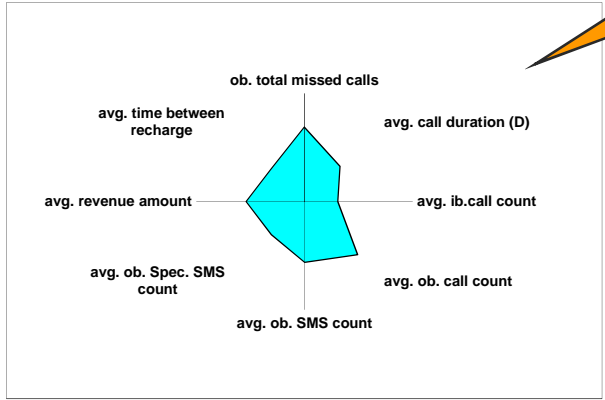
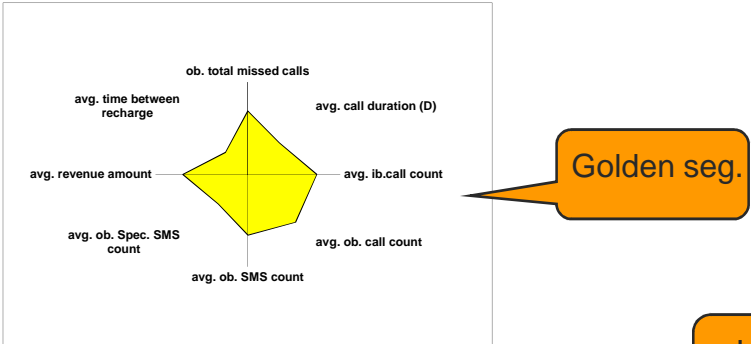
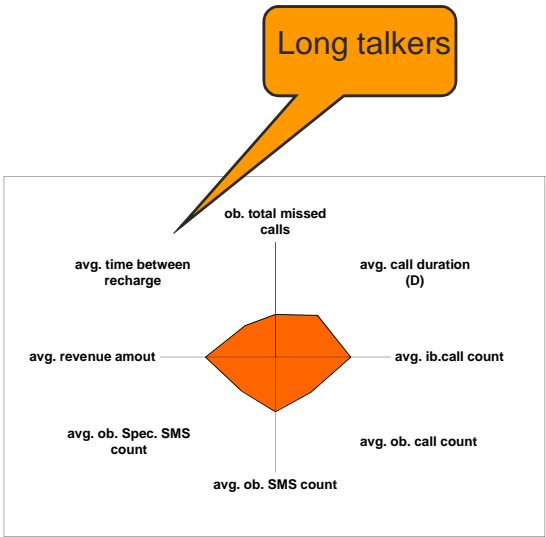
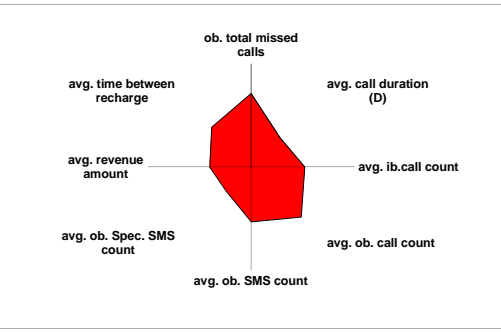
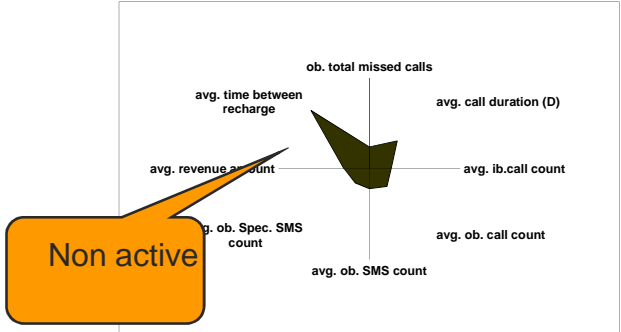
Having this in mind, we produced...



Who grew the most in the last year?

- After losing them for almost the entire year, Top productives managed to recuperate at the end of the observed period.
- By doing so, they joined Low-ends, Savers and Newcomers in the positive territory.
- A drop of 13% in Active prospects is the most worrying.

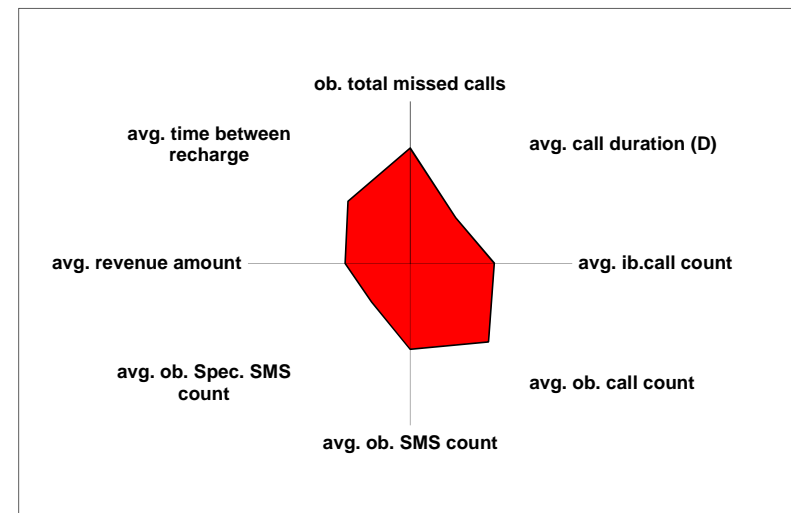




Description of segments

7. NIGHT BIRDS (x16.764 customers)

- Outbound activity:
 - very short duration of outbound calls
 - very high amount of missed calls, but
 - the number of calls with duration is lower then average
 - more active during the night periods (18h-01h)
 - use local and mobile network
 - use a lot of SMS's, but not special SMS's
- They have higher inbound activity
- They have lower revenue.



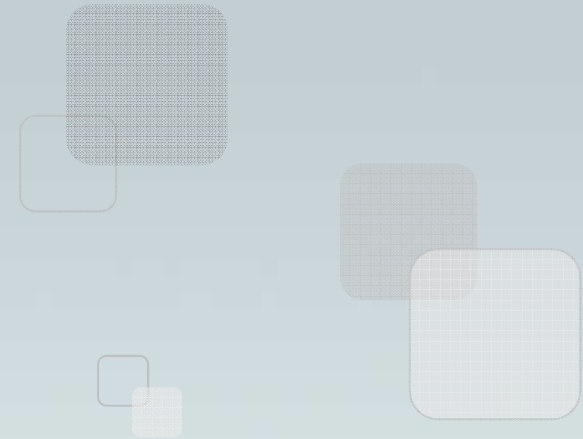
FREQ	Average Number Of Outbound Calls	Total Number Of Missed Outbound Calls	Total Number Of Regular Outbound Calls	Average Duration Of Regular Outbound Calls	Average Number of Outbound SMSs	Average Number Of Outbound Special SMSs	Average Number Of Inbound Calls	Total Recharge Amount	Average Time Between Recharges	Total Number Of Recharges	Average Recharge Amount	Average Revenue Amount
16764	382.678	237110913	14565688	21.9281	44.9923	0.58092	29.1883	169068410.24	33.8836	691710	245.7638	245.7638

xxx

xxx

Customer Segmentation, **Where is it Used?**

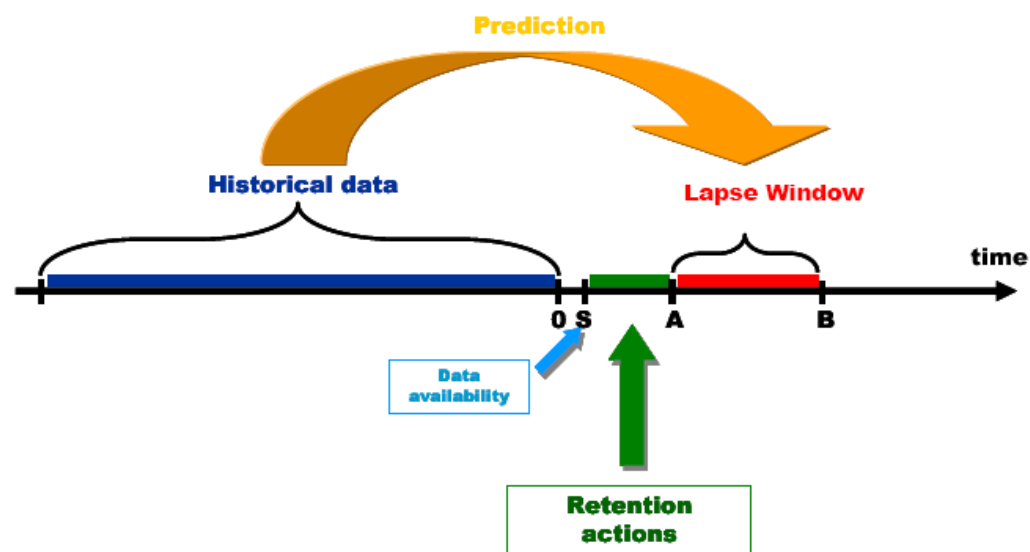
- As a foundation to predictive modeling
- For purposes of reactive campaigns
- For planning and strategy directions
- In offer definitions



Predictive modeling

Predictive data mining

- What are the chances of an event occurring at some point in the future?

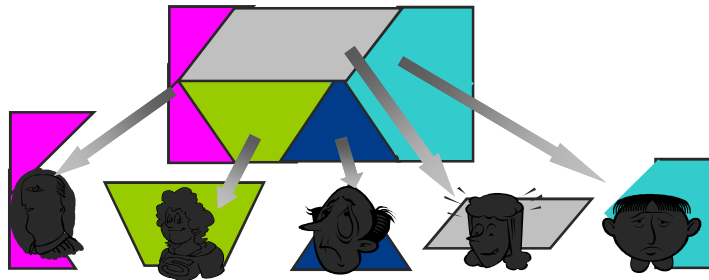


Customer reference	Chance of leaving %	Customer Value
M42511	80	High
A41125	77	High
P41256	77	Low
P12523	75	High
M17522	73	Low
A14123	72	Low

Predictive data mining

- examples:
 - Which people are most likely to respond to a marketing campaign promoting a new service / product?
 - Predict likelihood of a fraudulent card transaction to prevent the transaction occurring before any loss is incurred
 - Identify customers most likely to churn and use this to differentiate customer service and / or target retention offers
 - Propensity to default on a credit line?
 - and many more...
- AKA:
 - Propensity modelling, logistic regression, decision trees, neural networks, ensemble models, probabilistic modelling, scorecard building

Customer Retention, **Process Flow**



Commensurate Buckley	Incentive Customers	Self Core
(#)	(#)	(%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	40	10
7	30	10
8	20	10
9	10	10

Commensurate Buckley	Incentive Customers	Self Core
(#)	(#)	(%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	40	10
7	30	10
8	20	10
9	10	10

Commensurate Buckley	Incentive Customers	Self Core
(#)	(#)	(%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	40	10
7	30	10
8	20	10
9	10	10

Commensurate Buckley	Incentive Customers	Self Core
(#)	(#)	(%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	40	10
7	30	10
8	20	10
9	10	10

Commensurate Buckley	Incentive Customers	Self Core
(#)	(#)	(%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	40	10
7	30	10
8	20	10
9	10	10



Standardize and consolidate customer information

Segment customer base

Develop business cases and select target event

Develop predictive model

Design and deploy retention campaign

Business Case, **Definition**

- A business case captures the **reasoning for initiating a project or task**, the logic of the business case is that, whenever resources such as money or effort are consumed, they should be in support of the business. **Returns** must be estimated and set **as targets**.
- Consideration should also be given to the option of doing nothing including **the costs** and risks **of inactivity**. From this information, the justification for the project **is derived**.

Predictive Modelling, **Typical Business Cases**

Attrition

- Close all accounts

Dormancy Prevention

- No client-initiated transactions
- Decline in client-initiated transactional amounts
- No balance / Low balance
- Decline in balance amounts
- Low profitability
- Decline in profitability ...



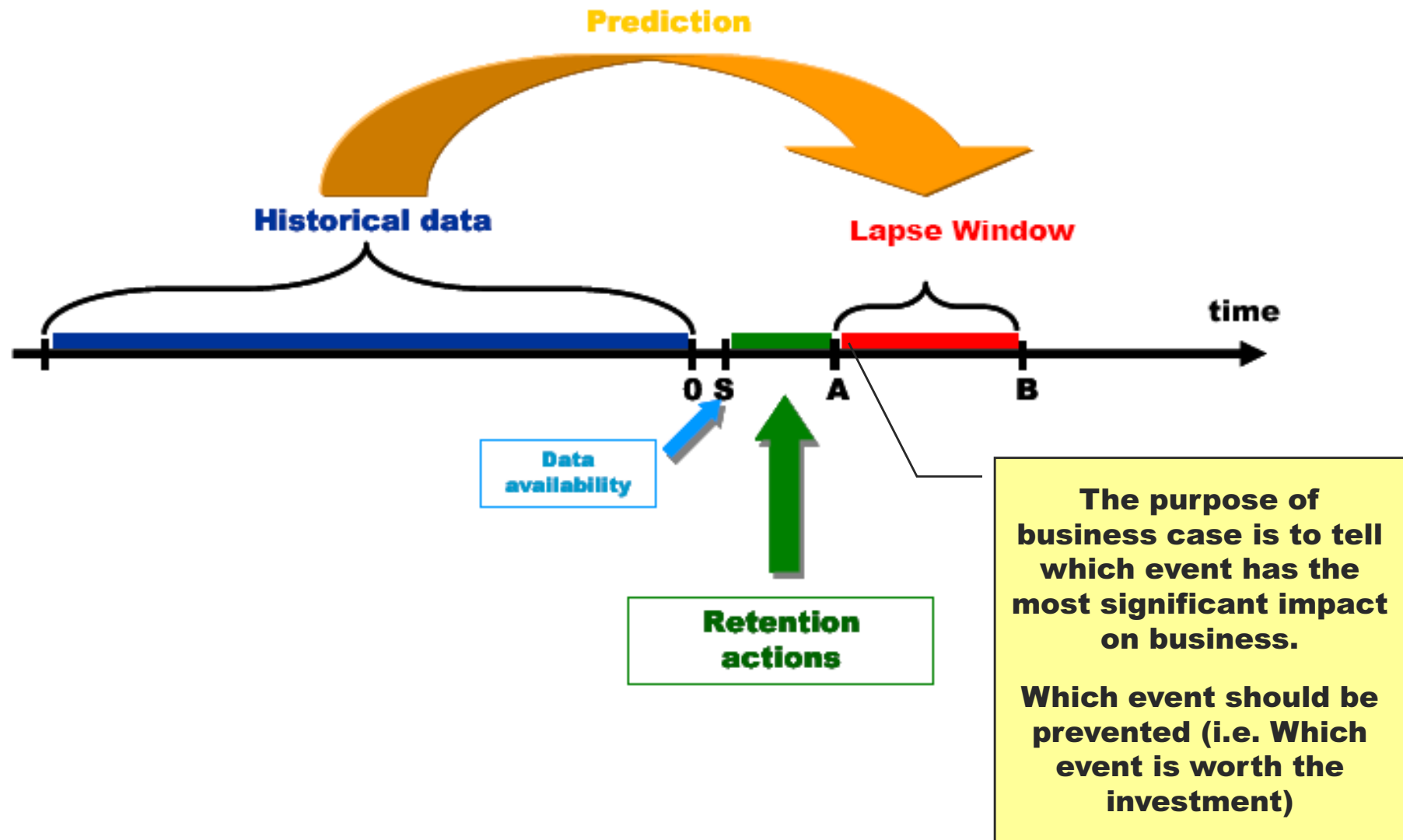
For how long?



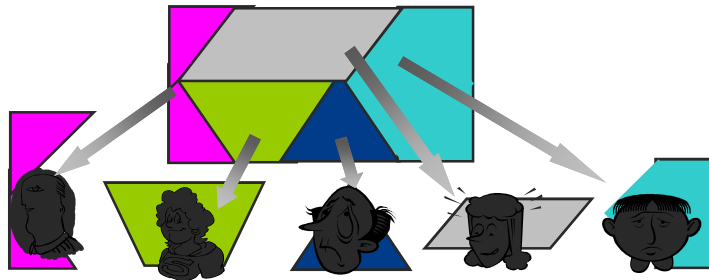
Customer Retention Business Case

Average Current Churn	8%
Current Number of Customers	500,000
Number of Customers leaving	40,000 a year 110 a day
Average Profit pre Customer	EUR 19
Savings by 2% decrease in churn	EUR 190.000 a year

Business Case, **Purpose**



Customer Retention, **Process Flow**



Commensurate Hustle (#)	Incentive Customers (#)	Self Core (%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	40	10
7	30	10
8	20	10
9	10	10

Commensurate Hustle (#)	Incentive Customers (#)	Self Core (%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	40	10
7	30	10
8	20	10
9	10	10

Commensurate Hustle (#)	Incentive Customers (#)	Self Core (%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	40	10
7	30	10
8	20	10
9	10	10

Commensurate Hustle (#)	Incentive Customers (#)	Self Core (%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	40	10
7	30	10
8	20	10
9	10	10

Commensurate Hustle (#)	Incentive Customers (#)	Self Core (%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	40	10
7	30	10
8	20	10
9	10	10



Standardize and consolidate customer information

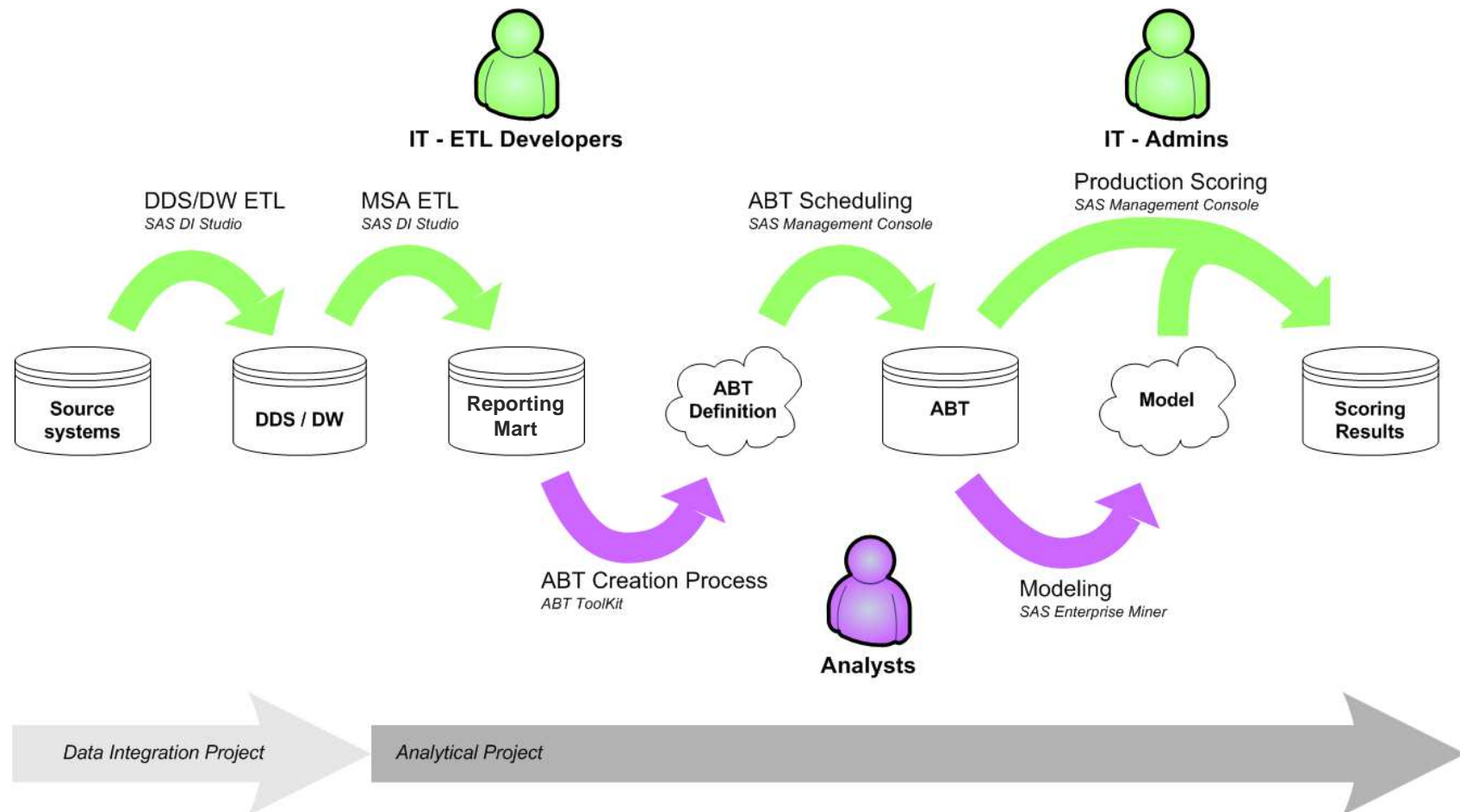
Segment customer base

Develop business cases and select target event

Develop predictive model

Design and deploy retention campaign

Data preparation process



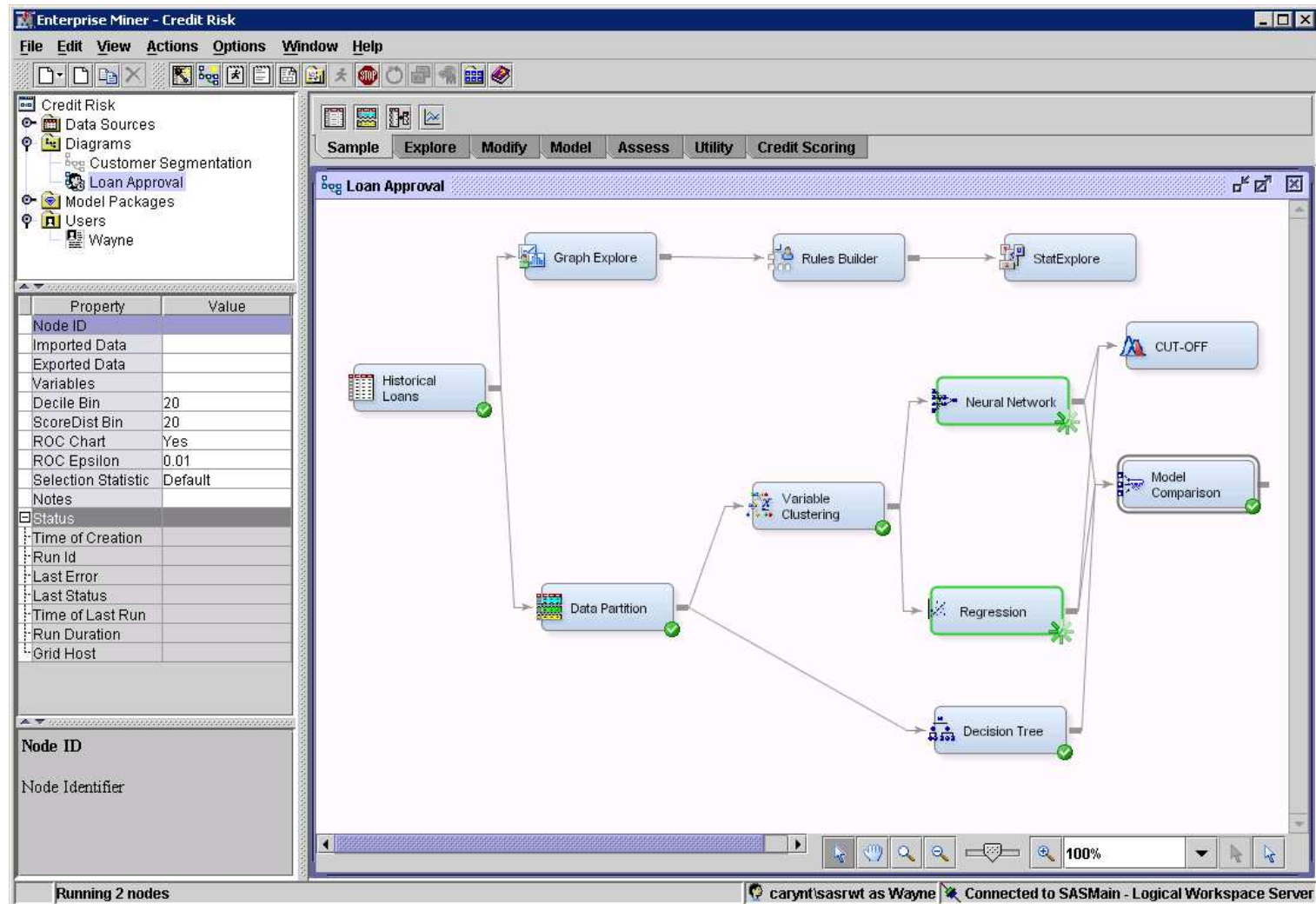
Data preparation – ABTs

CLIENT_ID	_M6	...	_M1	Derived Variables				Target
id1								
id2								
...								
...								
...								

M6	M5	M4	M3	M2	M1	ABT 200908		SLA_FLG_m4				
	M6	M5	M4	M3	M2	M1	ABT 200909	INCOME_AMT_m1				
		M6	M5	M4	M3	M2	M1	ABT 200910				
			M6	M5	M4	M3	M2	M1	ABT 200911			
				M6	M5	M4	M3	M2	M1	ABT 200912		
					M6	M5	M4	M3	M2	M1	ABT 201001	
						M6	M5	M4	M3	M2	M1	ABT 201002
												LAST_OPEN_PRODUCT
												LEN_AVG_BAL_AMT_m1
												CC_TOT_CR_TXN_CNT
												CMB_190_CWI
												SA_EOM_BAL_AMT_m1

MONTHLY_PAYMENT_AVG_AMT	Average of monthly paym
CWI_AVG_AMT	Average monthly Weight
CWI_AVG_BAL_m1	Weighted Income in curr
MATURITY_MTH	Months to maturity
DAYS_PAST_DUE_m1	Days-Past-Due at end of c
TOT_TXN_CNT	Number of client initiated
MB_190_CWI	Months below RSD 190 W
DAYS_SINCE_FIRST_ACCOUNT_OPEN	Days since first account o
SA_TOT_TXN_CNT	Number of transactions o
SA_MTH_SINCE_LAST_OPEN	Months since Savings Ac
NII_AVG_AMT	Average monthly Non-Int
PAST_DUE_AMT_m1	Past-Due Amount in curre
TOT_CR_TXN_CNT	Number of credit transact
MO_500_CWI	Months above RSD 500 W
CC_EOM_BAL_AMT_m1	Credit Card accounts bal
SALARY_FLG_m1	Salary received into RBR
CMO_500_CWI	Consecutive months abov
SLA_MTH_SINCE_LAST_OPEN	Months since last Salary
CMO_0_PDUE	Consecutive months with
TD_AVG_BAL_AMT_m1	Term Deposit accounts b
SLA_TENURE_MTH	Salary account tenure (m
TOT_TXN_AMT	Transaction turnover in la
TOT_DB_TXN_AMT	Debit transaction turnove
TENURE_MTH	Customer tenure (months
DAYS_SINCE_LAST_ACCOUNT_OPEN	Days since last account v
SLA_FLG_m4	Salary received into RBR
INCOME_AMT_m1	Customer income in curre
LAST_OPEN_PRODUCT	Last opened product
LEN_AVG_BAL_AMT_m1	Lending accounts balanc
CC_TOT_CR_TXN_CNT	Number of credit transact
CMB_190_CWI	Consecutive months belo
SA_EOM_BAL_AMT_m1	Savings accounts balanc

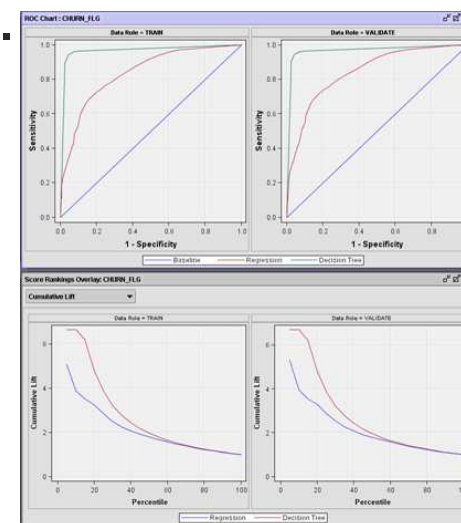
Streamline process in SAS EM



Modeling, **Champion model selection**

- The models need to be compared in terms of:
 - their business value/understanding (reasonable predictors)
 - statistical accuracy measures
 - typical measure is lift, captured response
 - other measures: ROC, fit statistics, ...

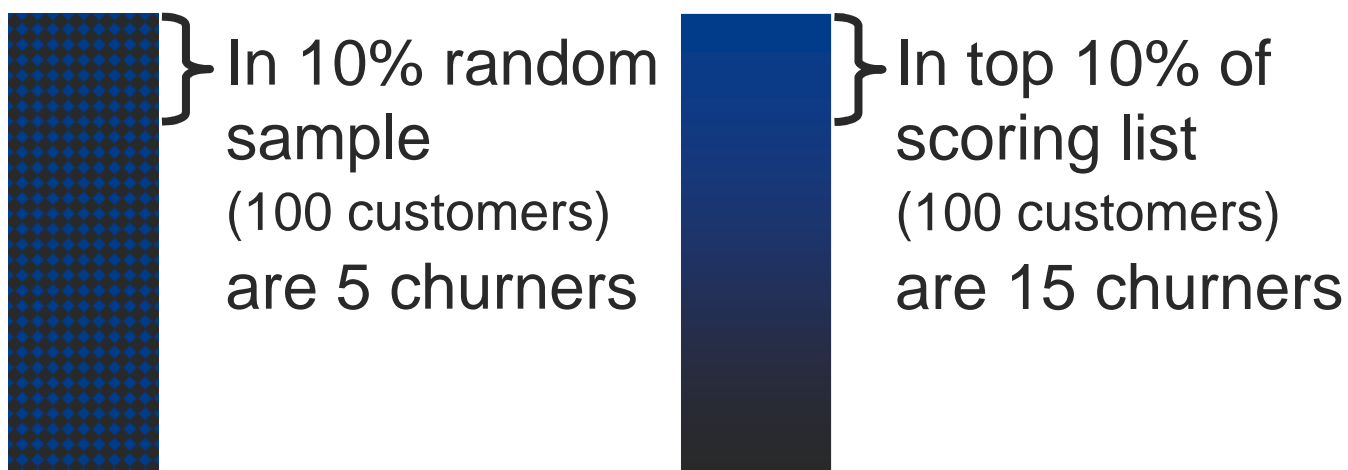
$$\text{LIFT}_{10\%} = \frac{\text{number of events in the top 10\%}}{\text{expected number of events in 10\%}}$$



Lift as measure of prediction efficiency

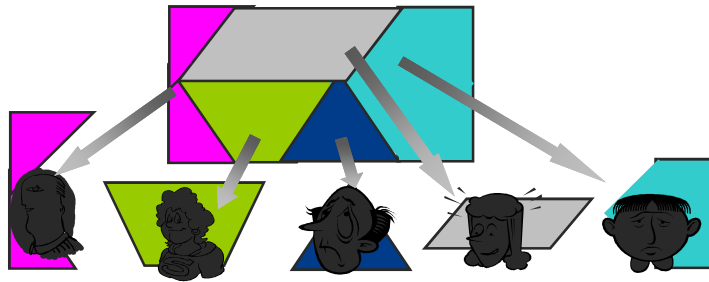
- Example:

- 1000 Customers
- 5% Churn-Rate → 50 Churners



$$\text{LIFT}_{10\%} = 3 (=15/5)$$

Predictive modelling, **Process Flow**



Standardize and consolidate customer information

Segment customer base

Connective Buddy (#)	Inactive Customers (#)	Self Care (%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	30	20
7	20	10
8	20	10

Connective Buddy (#)	Inactive Customers (#)	Self Care (%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	30	20
7	20	10
8	20	10

Connective Buddy (#)	Inactive Customers (#)	Self Care (%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	30	20
7	20	10
8	20	10

Connective Buddy (#)	Inactive Customers (#)	Self Care (%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	30	20
7	20	10
8	20	10

Connective Buddy (#)	Inactive Customers (#)	Self Care (%)
1	100	10
2	90	10
3	70	20
4	60	10
5	50	10
6	30	20
7	20	10
8	20	10

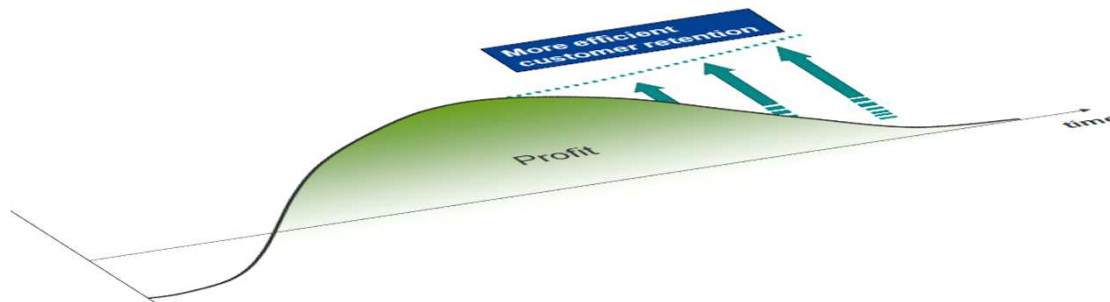
Develop business cases and select target event



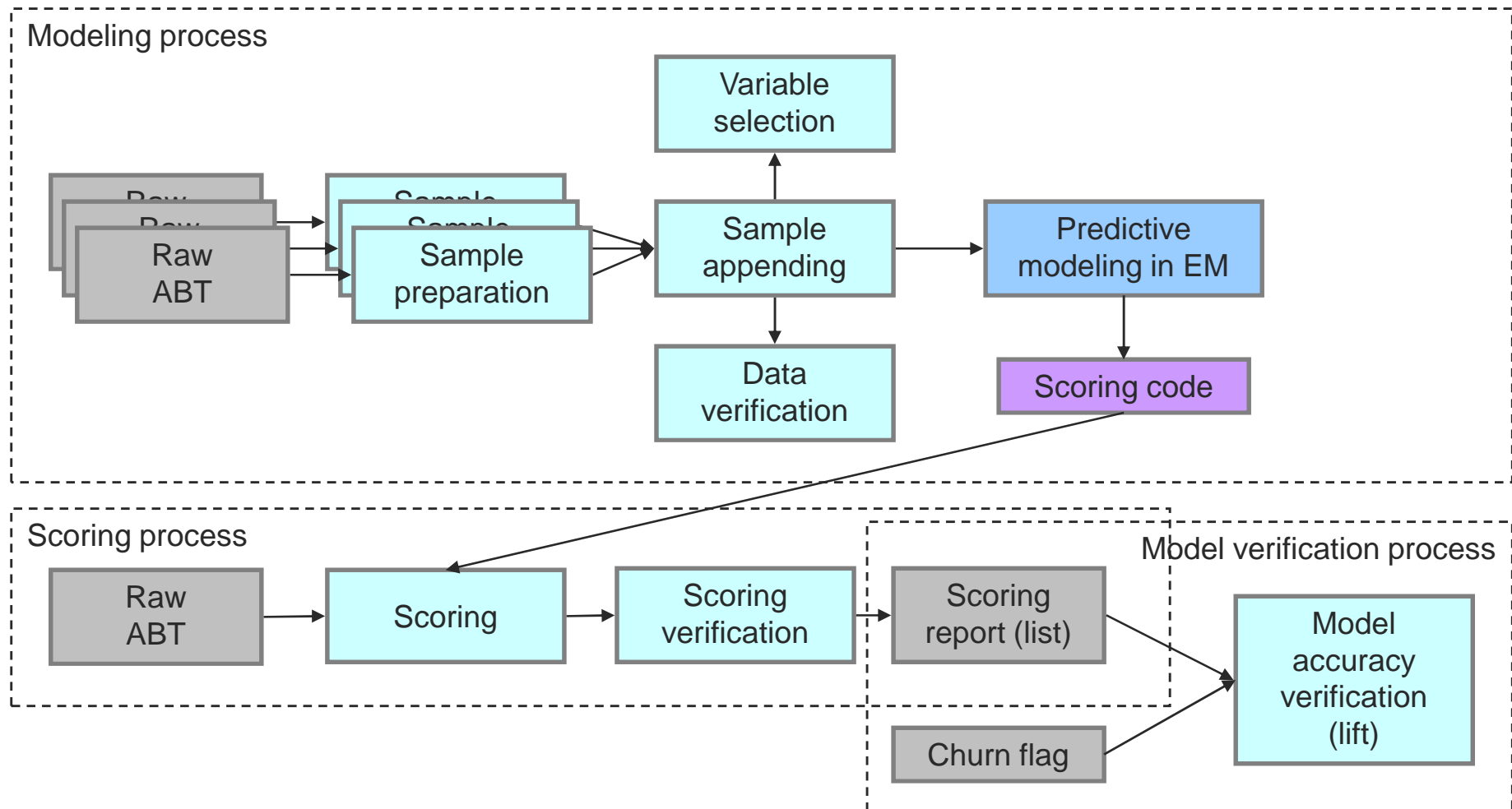
Develop predictive model



Design and deploy focused campaign(s)



Model performance



Retention Campaign, **Select Target Population**

Model Scores

Customer	Propensity for Dormancy	Dormancy Cost	Money at Risk
1	20%	€ 100,00	€ 20,00
2	30%	€ 1.200,00	€ 360,00
3	40%	€ 100,00	€ 40,00
4	10%	€ 1.000,00	€ 100,00
5	60%	€ 200,00	€ 120,00
6	50%	€ 500,00	€ 250,00
7	20%	€ 300,00	€ 60,00
8	30%	€ 100,00	€ 30,00
9	40%	€ 1.000,00	€ 400,00
10	10%	€ 500,00	€ 50,00
11	60%	€ 100,00	€ 60,00
12	50%	€ 200,00	€ 100,00

- Threshold: **40%**
- Campaign Capacity: **4 Customers**

Strategy: **Market Penetration**

Customer	Propensity for Dormancy	Dormancy Cost	Money at Risk
5	60%	€ 200,00	€ 120,00
11	60%	€ 100,00	€ 60,00
6	50%	€ 500,00	€ 250,00
12	50%	€ 200,00	€ 100,00

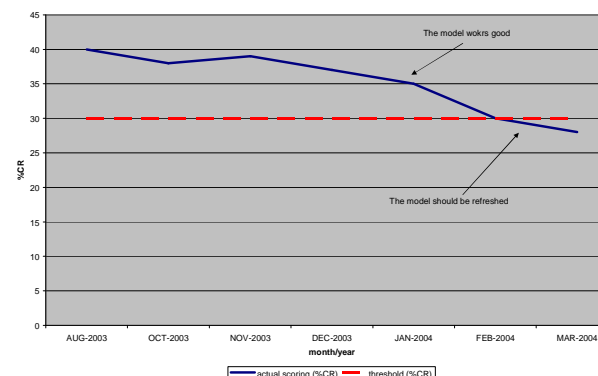
Strategy: **Profitability**

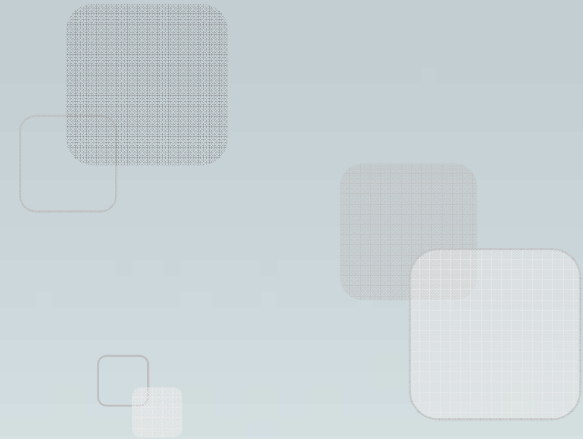
Customer	Propensity for Dormancy	Dormancy Cost	Money at Risk
9	40%	€ 1.000,00	€ 400,00
6	50%	€ 500,00	€ 250,00
5	60%	€ 200,00	€ 120,00
12	50%	€ 200,00	€ 100,00

Commonly also crossed with other campaign's expected return, product-mix, theoretical monthly repayments and propensity to buy (Xsell scores)

Model **refreshment** strategy

- Depending on assumptions the model should be „refreshed”:
 - Periodically, e.g., every 3 or 6 months by default
 - Monitor the accuracy of model predictions and re-build then model when their performance falls down below a threshold value
 - Increase in score warnings also indicates the need for re-training
 - When a market is changed due to introduction of new products or services





Next best offer

Cross-sell / Up-sell

“Up” Selling

Upgrades, higher-end services, premium products, additional features

Example: “Would you like the large combo for only \$0.50 more?”

“Cross” Selling

Additional products & services, complementary “add-ons”

Example: “Would you like a muffin to go with that coffee?”

Whatever we call it, the topic is...

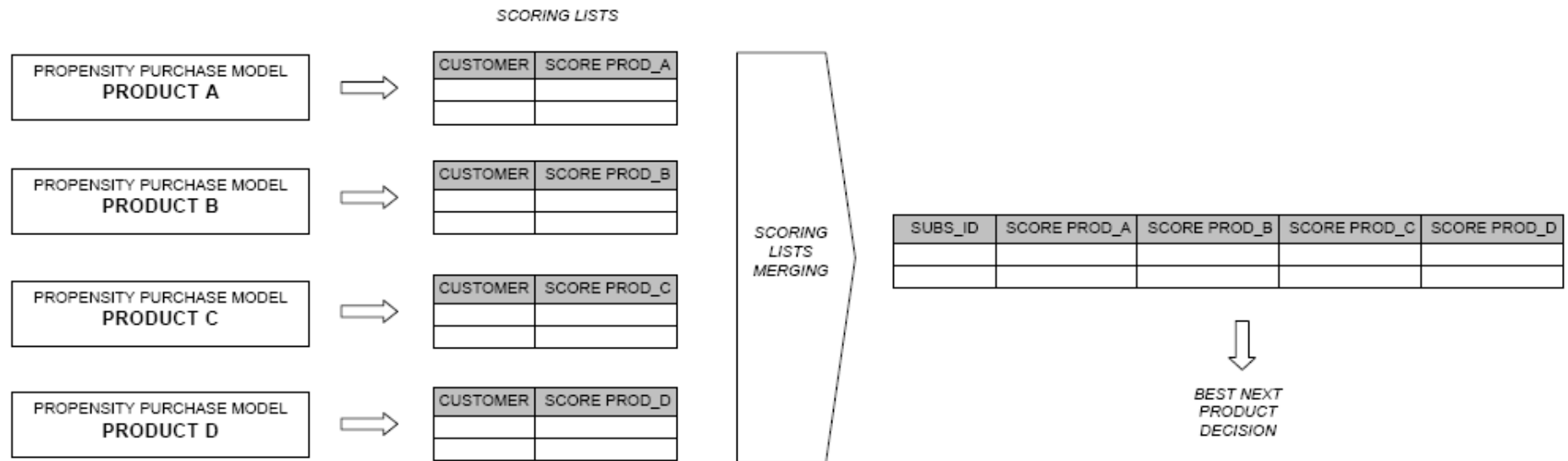
**getting more revenue (*profit? economic value?*)
from your existing customers.**



Next Best Offer - Methods

- **What products are usually bought together?**
 - Association (Market Basket Analysis)
- **What is the usual order of products purchase?**
 - Sequence Analysis

Different Approach to Next Best Offer Modeling



Customer Scoring

CUSTOMER_RK	Score			
	A	B	C	D
6902	0.97	0.8	0	0.01
6903	0.1	0.4	0.9	0.1
6904	0.55	0.9	0.3	0.2
6905	0.55	0.2	0.4	0.3
6906	0.6	0.4	0.2	0.1
6907	0.1	0.1	0.7	0
6909	0.05	0.07	0.1	0

Selection of the Best Product

Customer Preference Based Selection

CUST_SUBS_ID	Score			
	A	B	C	D
6902	0.97	0.8	0	0.01
6903	0.1	0.4	0.9	0.1
6904	0.55	0.9	0.3	0.2
6905	0.55	0.2	0.4	0.3
6906	0.6	0.4	0.2	0.1
6907	0.1	0.1	0.7	0
6909	0.05	0.07	0.1	0

Selection of the Best Product

Selection of Product with the Highest Average Profit

CUSTOMER_RK	Score * Average Profitability			
	A	B	C	D
6902	48.5	80	0	0.6
6903	5	40	63	6
6904	27.5	90	21	12
6905	27.5	20	28	18
6906	30	40	14	6
6907	5	10	49	0
6909	2.5	7	7	0

Selection of the Best Product

Optimization of Total Profit

CUST_SUBS_ID	Score * Individual Profitability			
	A	B	C	D
6902	65.5	70	0	3
6903	5	60	23	4
6904	15	95	10	15
6905	30	29	35	25
6906	40	30	20	8
6907	20	30	50	0
6909	9	7	5	0

Association “Rules”

Key Components / Terminology for Associations:

Rule	Expected Confidence	Confidence	Lift	Support
------	---------------------	------------	------	---------

Examples:

A → B	20%	40%	2.0	5%
A → C	10%	30%	3.0	1%

Interpretation of the first rule:

Customers who bought A, also bought B	success rate for randomly selling B	success rate for “selling B to A”	twice as likely	5% of “market basket”
--	--	--	--------------------------------	--------------------------------------

NBO Results Example - Banking

	Relations	Lift	Support(%)	Confidence(%)	Transaction Count	Rule
1	2	1.04	41.47	72.93	58319	SA ==> FD
2	2	1.04	41.47	59.36	58319	FD ==> SA
3	2	1.23	10.56	85.61	14848	UT ==> FD
4	2	1.23	10.56	15.11	14848	FD ==> UT
5	2	2.38	8.89	63.84	12504	CC ==> CA
6	2	2.38	8.89	33.14	12504	CA ==> CC
7	2	1.22	8.56	69.43	12041	UT ==> SA
8	2	1.22	8.56	15.06	12041	SA ==> UT
9	2	1.90	6.25	50.95	8790.0	MG ==> CA
10	2	1.90	6.25	23.30	8790.0	CA ==> MG
11	2	2.61	4.46	36.34	6270.0	MG ==> CC
12	2	2.61	4.46	32.01	6270.0	CC ==> MG
13	2	1.14	3.77	30.54	5297.0	UT ==> CA
14	2	1.14	3.77	14.04	5297.0	CA ==> UT
15	3	1.01	7.61	13.38	10696	SA ==> FD & CA
16	3	1.01	7.61	57.45	10696	FD & CA ==> SA
17	3	1.46	7.44	60.34	10465	UT ==> SA & FD
18	3	1.24	7.44	13.09	10465	SA ==> UT & FD
19	3	1.24	7.44	10.65	10465	FD ==> UT & SA
20	3	1.24	7.44	86.91	10465	UT & SA ==> FD
21	3	1.24	7.44	70.48	10465	UT & FD ==> SA
22	3	1.46	7.44	17.94	10465	SA & FD ==> UT
23	3	2.30	4.24	30.47	5968.0	CC ==> FD & CA
24	3	2.30	4.24	15.82	5968.0	CA ==> FD & CC

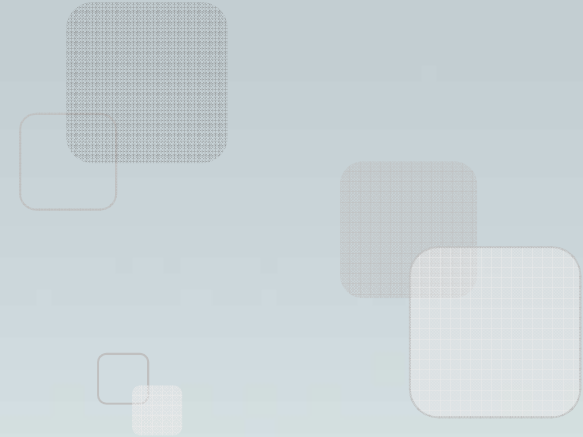
- **41.47%** of all the Gold customers having both Saving Account and FD (this is equivalent to **58,319** customers)

- For customers who have Saving Account, **72.93%** of them also have FD in the bank

Cross selling CA to CC customers will yield 2.38X better success rates than if cross sold randomly to other customers (with 63.84% confidence)

Association Analysis

- <http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>



Conclusion

Literature / blogs / links

- Hand, Mannila, and Smyth: Principles of Data Mining
- Rud, Data Mining Cookbook
- www.kdnuggets.com
- Dataminingblog.com
- The-Data-Mine.com
- Stanford free online [courses](#)
- [Top analytic blogs and websites, with trending information](#)



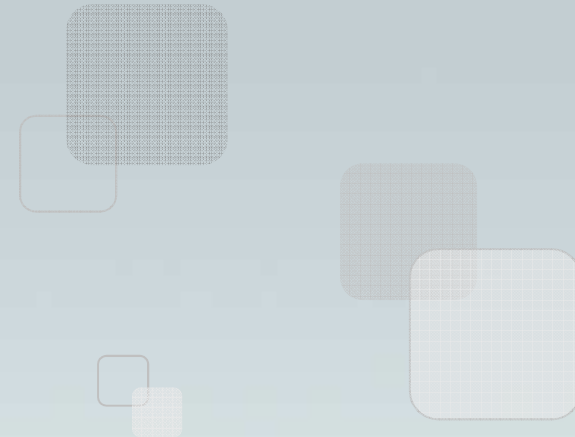
Five Common Mistakes in Analytic Projects

- Underestimating the time required to get the data
- There is not a good plan for deploying the model
- Working backwards, instead of starting with an analytic strategy
- Trying to build the perfect model
- The predictions of the model are not actionable



MARKETING NPV Rini

*Izgradili smo model, ki projecira vaše rezultate
na zaključnem testu. Na žalost ste pogrnili...*



Analytics in Business Environment

dr. Aleksander Pivk

SAS Adriatic Region
(aleksander.pivk@adr.sas.com)

Jozef Stefan Institute
(aleksander.pivk@ijs.si)