

Customer Profiling, Segmentation and Marketing Strategies in Telecommunications

**International Conference
on Current Advances
and Trends in Nonparametric Statistics
July 15-19, 2002 - Crete, Greece**

Bruno Scarpa
Customer Intelligence Manager



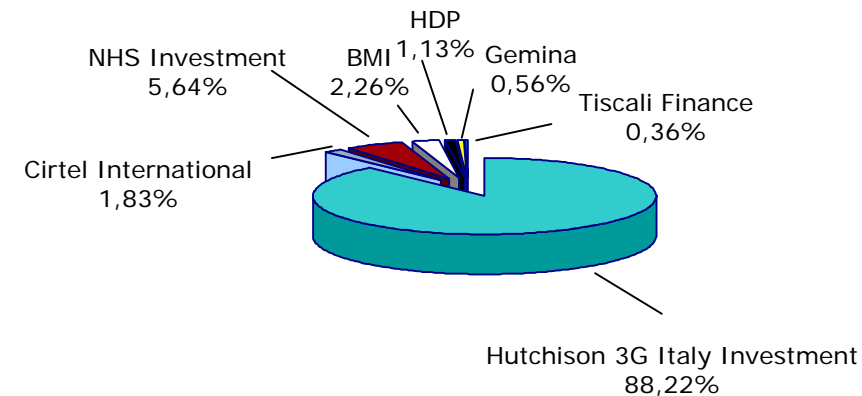
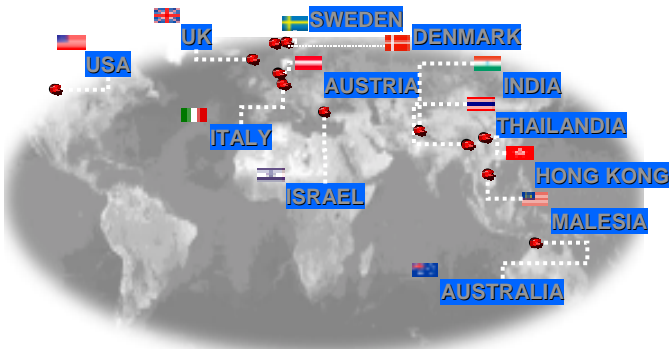


outline

- **aim of this talk:**
 - a. investigate the role of a statistician in this kind of companies**
 - b. to present the statistical problems I met in my experiences in telecom companies**
 - c. to take one (some) specific problem and present how we tried to solve it**

H3G

- H3G Italy s.p.a. is born to be the first mobile operator of 3rd generation in Italy, completely focused on the develop of innovative and mutilmmedia services in UMTS technology
- H3G is part of an international project led by Hutchison Whampoa, one of the most important player in the global economy



- H3G will be a supplier of content/services/solutions
- Customers will have a connection always on, anywhere they will be
- H3G is planning to launch services (with the brand "3") in the 4th quarter 2002

UMTS



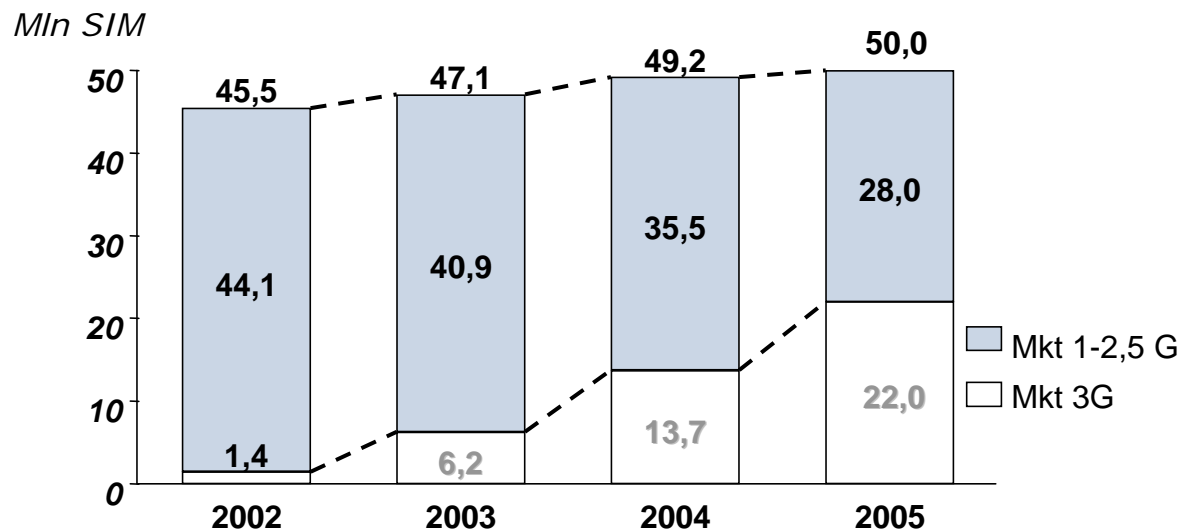
□ new mobile communication technology:

▪ **Multimedia, interactivity and high speed**

→ Data transfer speed: 10 times faster than GPRS (300kbit/s vs 30kbit/s)

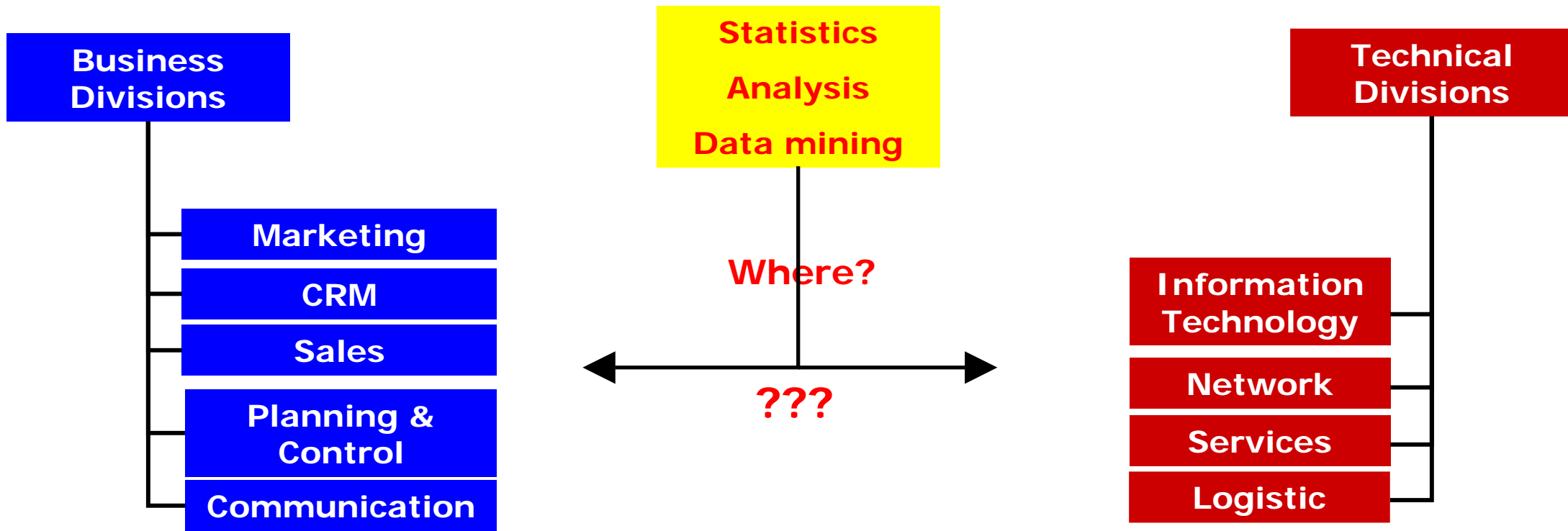
→ Not just communication, but also content provider...

Forecasts



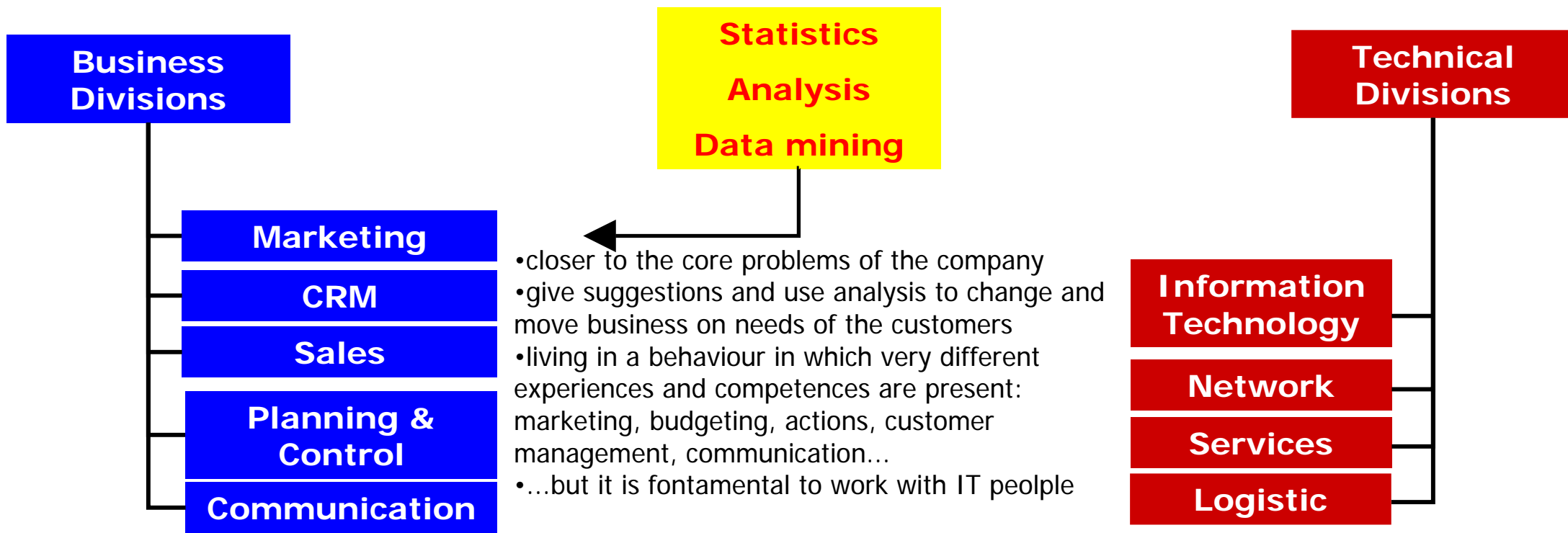
statisticians

- H3G is a start up
- H3G has no customers (not yet!)
- H3G has not yet data!
- So why do they need data miners and statisticians?
- Which is the role of statisticians in this company?



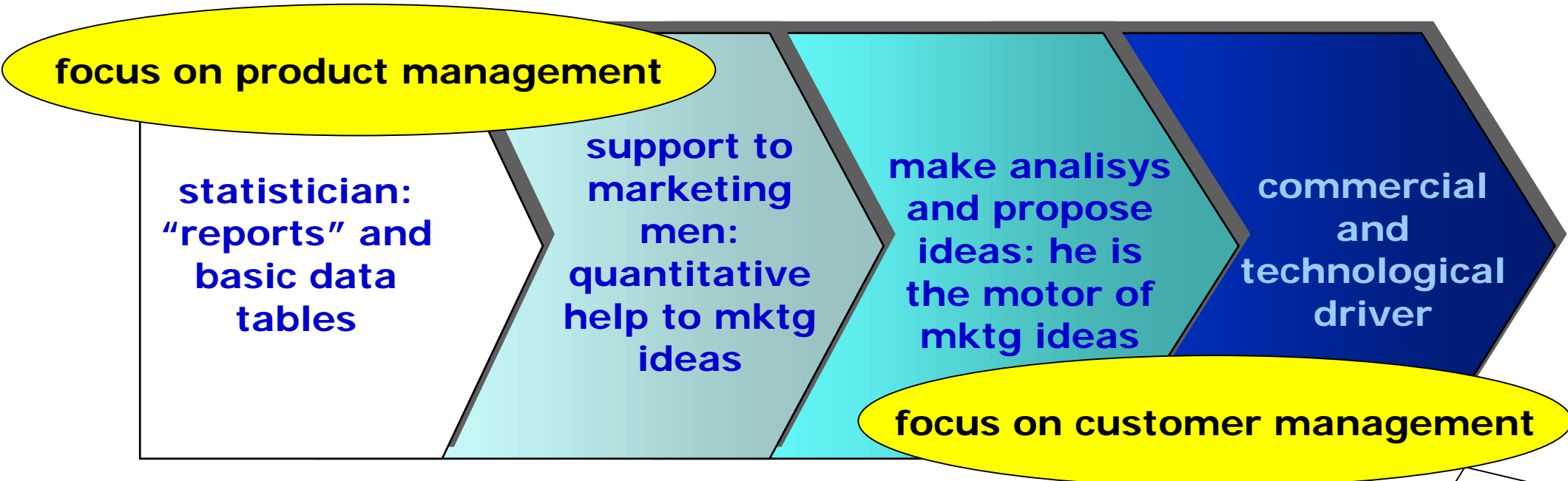
statisticians

- H3G is a start up.
- H3G has no customers (not yet!)
- H3G has not yet data!
- So why do they need data miners and statisticians?
- Which is the role of statisticians in this company?



→ **statisticians**

different level of possible involvement of a statistician in making business



- Tools:
- Reports
 - Tables
 - Simple indicators
 - Reports
 - Tables
 - Simple indicator
 - Simple models
 - Tables
 - Indicators/rate
 - Models
 - Data mining
 - Statistical models
 - Data mining
 - Reports
 - Tables
 - Indicators

H3G before launch



statisticians

- take the UMTS problems using the past experience:
 - cellular telecom world (GSM, TACS...)
 - internet world (ISP, ASP, e-commerce...)
- what kind of business questions we will have?
- what kind of data we will need to answer the new business questions?
- „design of experiment“ based on the past experiences:
 - prepare data
 - prepare tools / packages
 - prepare people



□ customer data

▪ general

- socio demographics
- activation data (e.g. Subscription date, Type of contract, Plan subscribed, ...)

▪ billing data

▪ telco data

- telco traffic data
- VAS & Killer Applications data
- other value added services (both internal and M-sites) data

▪ external researches and data

▪ market research

▪ costs data

▪ internet data

- web traffic data
- page provisioning/ Web contents (e.g. page views, unique visitors, ...)
- e-mail data
- data gathered through the web (both questionnaires and customer behaviour)
- community data

▪ customer base management data

- operational CRM data
- campaign Management & Marketing contact history data
- loyalty programs data

▪ H3G mistakes data (e.g. billing errors, ...)

▪ ...

TLC: some problems

□customer acquisition

▪**prospecting**

→achive a *minimal thereshold* of customers ASAP

→find and attract the *right* customers: how much to spend for each different customer?

▪**detecting subscription fraud**

→determine the likelihood of a fraudulent application

□customer profitability

▪**Customer value**

▪**dormancy and share of wallet**

▪**risk monitoring and managemet**

→determine and optimize risk parameters



TLC: some problems

□ Customer loyalty

- **predictive approach**
 - predictive models of churn
- **attrition and retention**
 - model and determine main drivers
- **actions: loyalty programs/campaign/
up sell-cross sell**
- **customer relationship**
 - personalization of the care and of the contacts

□ Customer profiling

- **who are customers**
- **what each customer wants**
- **how to contact each customer**

□ actions evaluation

- **no case-control experiment**
 - customers are autoselected
- **evaluate a posteriori some action**
 - estimate the effect of the action conditionally to the effect of all the others variables

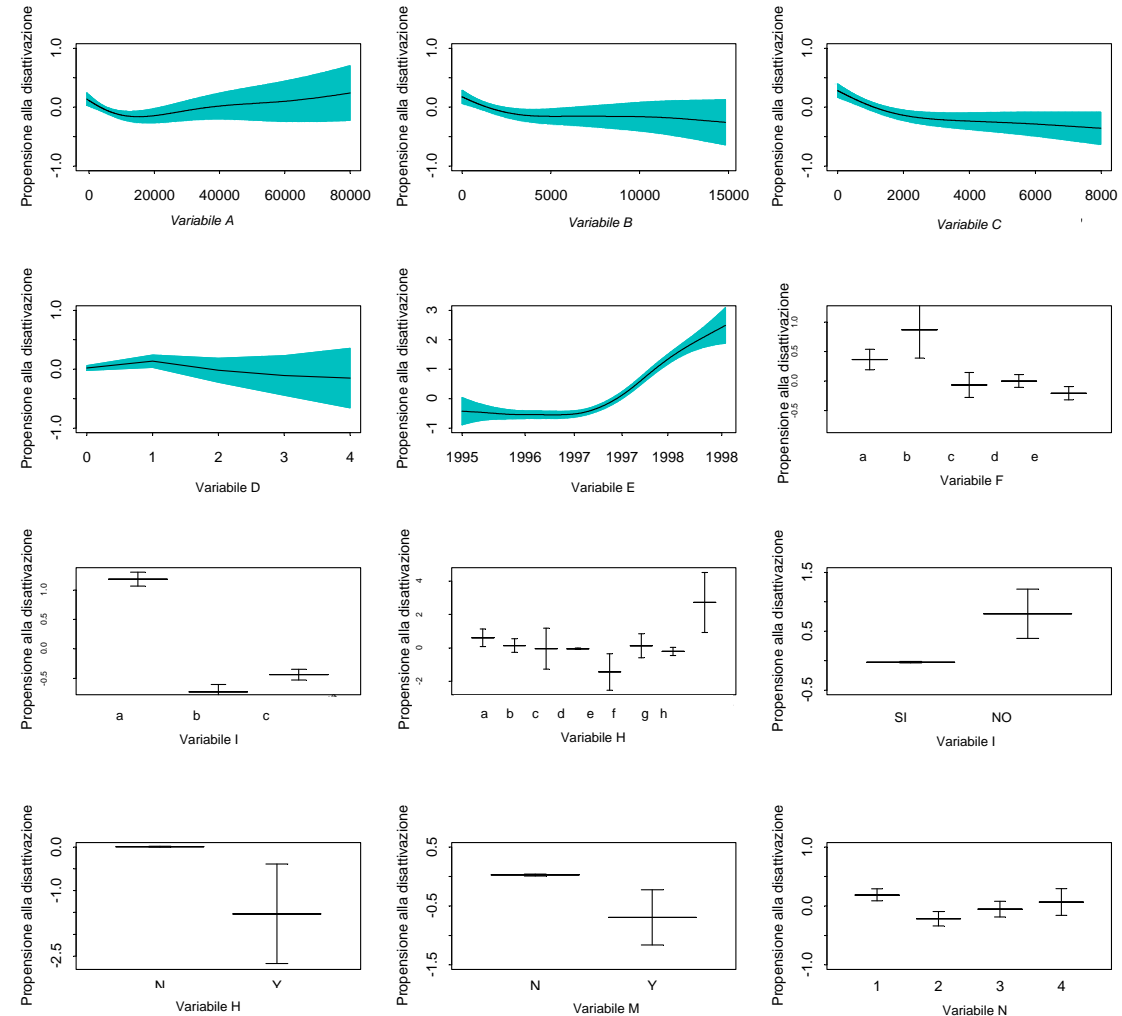
churn: GAM

□ Goal: find for each customer a score of propensity to churn

- Understand which variables have effect on the decision to churn and which is the effect of each variable considered between the others

□ It is more important to understand effects than the accuracy of prediction

- A GAM model has been fitted to a random sample of (balanced) data
- Results have been evaluated on the entire customer base





churn: CART

□ Prepaid and “traditional” customer:

- the prepaid customer IS disconnected when he has not recharged for 12 months
- the “traditional” customer ASKS to be disconnected whenever he wants

□ The prepaid customer has migrated to the competitors well before the “technical” deactivation

□ We want to anticipate another event: a **SIGNAL** of the decision not to recharge

churn: CART

□ The **SIGNAL** should be

- “Intuitive” and “simple”
- Related with the *choice* of the customer to leave
- Accurate
 - determine attrition classes
 - use models to classify existing customers
 - Determine

□ It will be obtained using

- Incoming traffic (little)
- Outgoing traffic (little)

□ **Errors of the SIGNAL:**

- Activ sims with signal on = 16%
- Churner sims with signal off = 15%

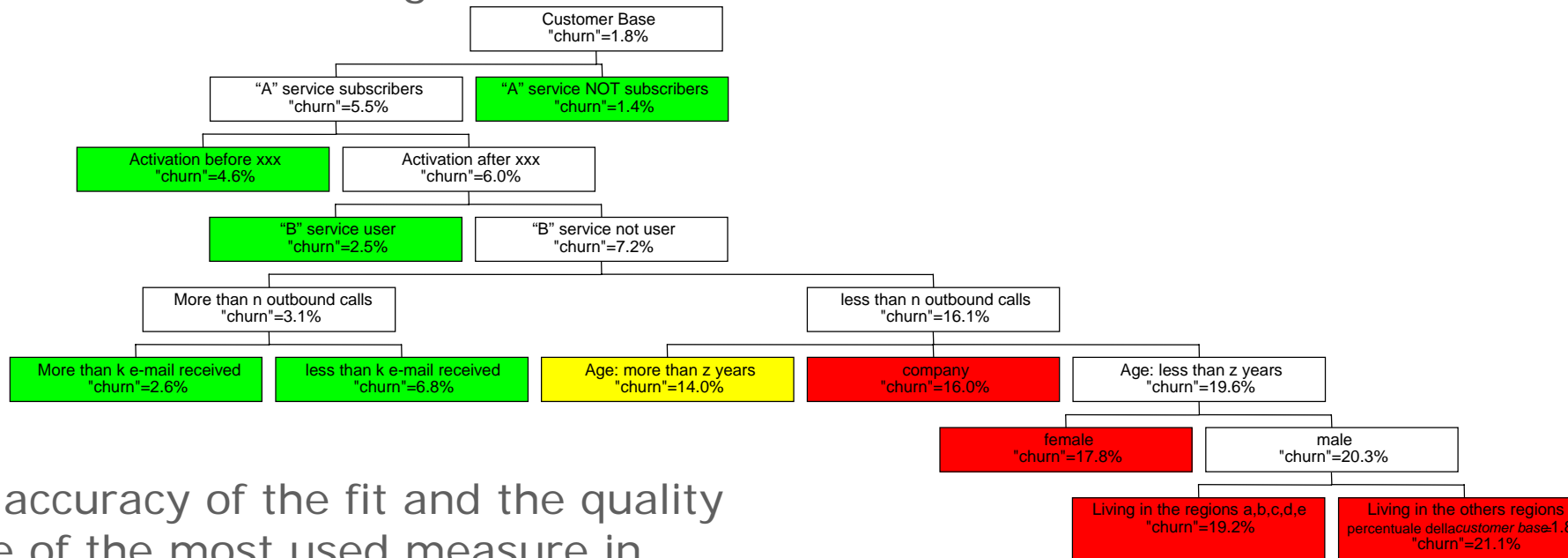


churn: CART

- A **CART** model is fitted predicting the “**signal**”
- The choice and the order of variables to be used in the model is defined a priori by the research team. The criteria of decision are based on
 - **Knowledge of the business**
 - **Actionability**
 - **Previous models estimated**
 - **Preliminary analysis and univariate estimates**
- Gini index has been used as split rule
- The Estimate of an index of propensity to churn for each node has been obtained by an “evaluation” data set which has been separated from the estimate dataset at the beginning of the analysis

churn: CART

- For each node and leaf of the tree a different index of propensity to churn (signal). The colour of the leaves of the tree are related with the risk levels: ■ low, ■ medium, ■ high



- To estimate the accuracy of the fit and the quality of the model one of the most used measure in data mining is the lift

- Lift:** in a node is the ratio between the proportion of predicted churners in the node of the model over the total population in that node and the proportion of churners in all the customer base

Percentile of the population	1%	5%	10%	20%
Lift	4.39	4.14	3.63	2.86

churn: CART

pure data mining

Black box solution where the software (IT) select the variables, choose and fit the model in a completely automatic way

**Example:
Old model**

✓ Decrease of traffic

non actionable!

data mining leaded

The statistician **lead** the analysis, the choices (at least partial) of the variables, decide how to fit models and use data mining models as analysis tools

**Example:
Model CART I**

✓ Traffic characteristic (ex. High usage in the pick hours)

✓ Services usage

✓ Decrease of traffic

✓ Complaints

**Marketing and
CRM actions**



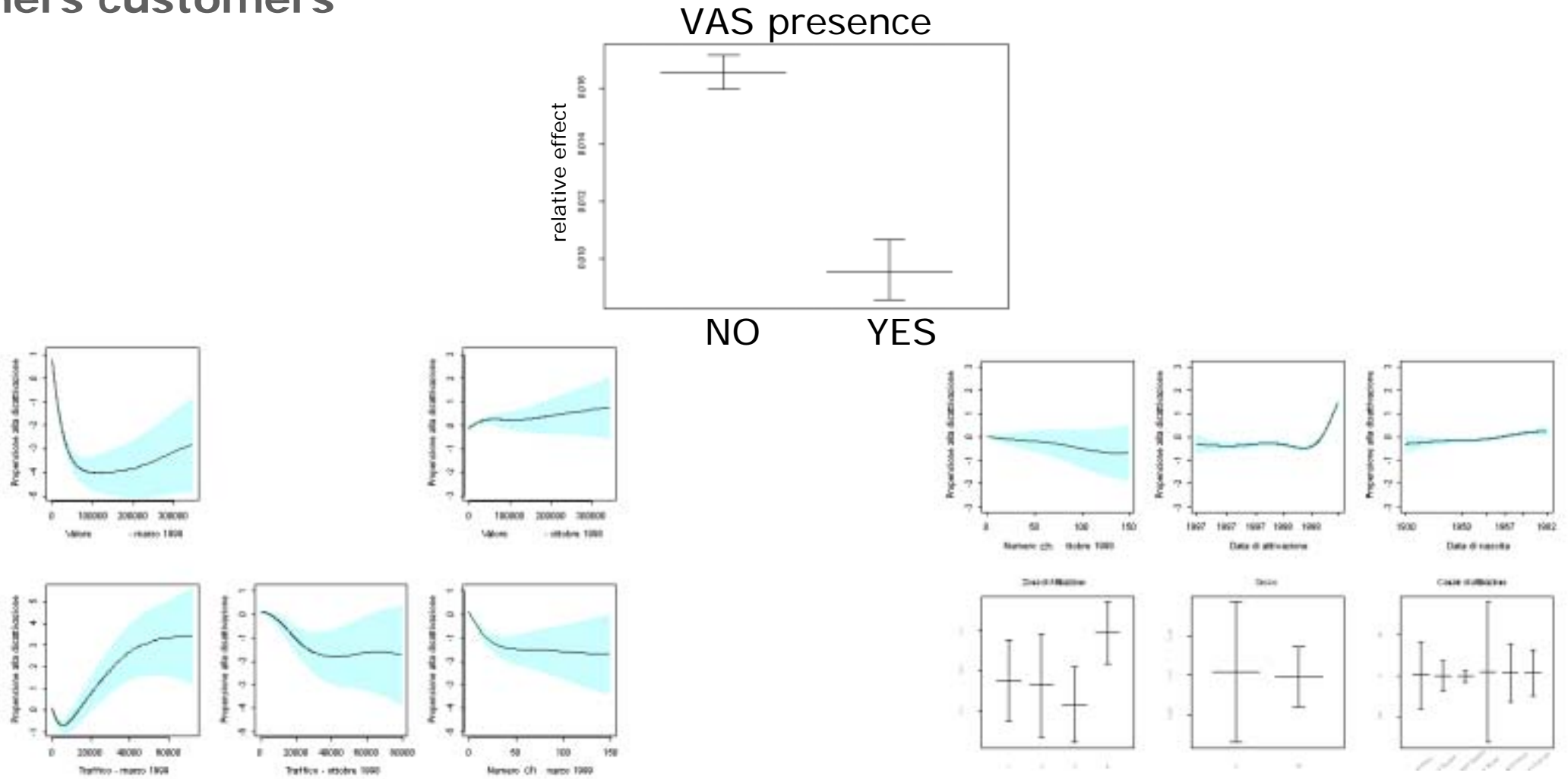
action evaluation

□ A particular Value Added Service has been prepared and sold in the last year

- All customers can subscribe it
- Question: has the subscription of this VAS some effect on churn rates?
 - It is not possible to use a case-control experiment (autoselection)
 - It is important to estimate the effect taking in account of the conjoint effect of other correlated variables
 - Use of data mining model to estimate the specific effect of the presence of that VAS
 - Explanatory variables are both continuous and discrete
 - A GAM model has been fitted to a random sample of data

action evaluation

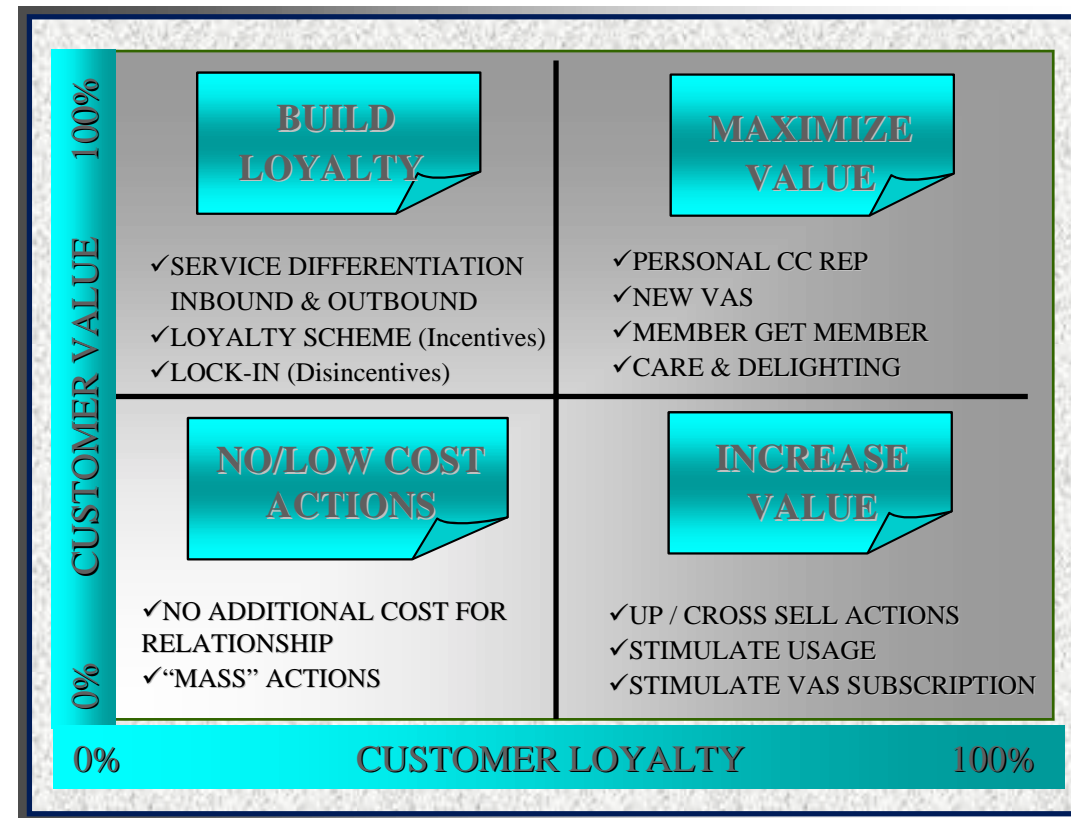
The effect is significant: te subscribers are more loyal than the others customers



customer profiling

□ The customers and the company

- plot each customer on the map
- understand his own behaviour
- analyze deeply the customer values
- apply the rules to define actions
- deploy and evaluate actions
- obtain customer lifetime value



conclusions

- business needs good statisticians!
- problems are many and interesting
- databases are huge, but we need people that look inside them, not just packages that perform algorithms
- Insight tools
 - **Non standard problems**
 - not always the solution is already in the tool
 - **commercial software proclame they solve every problem just clicking a button**
 - **not always start with huge datasets means to analyze all data and needs very fast (paarallelized) algorithm (and tools)**

Pemberton

I TRIED TO IMPROVE MY
MARKETING MODELS USING
"BLACK BOX" TECHNIQUES...

I HOPE YOU DIDN'T
LINK YOUR ANNUAL
SALARY RAISE TO THEIR
PERFORMANCE...

Thank's to

Laurence Jacobs (Kdlabs

and all the people (marketing
IT, statisticians...

I worked with in these years

Copyright © 2002 Alberto Busetto. Soggetto depositato, tutti i diritti riservati

Bruno Scarpa
Customer Intelligence

