Paper 132-2009

The Path to Profitable Analytics: Via Predictive Analytics to Marketing Automation

Petr Ptáčník, Česká spořitelna, Prague, Czech Republic

ABSTRACT

The SAS BI solutions enabled us to increase the number of direct marketing campaigns from 20 to 1000+ a year. Now with SAS Marketing Automation implemented we are able to further increase the number of campaigns while improving the quality of target list selections.

The paper focuses on data handling methodology in preparation of direct marketing campaigns and how the business needs were reflected in SAS Marketing Automation.

INTRODUCTION

No doubt there are many possible ways to profitable analytics. This paper aims to describe the way to profitable analytics we chose in Česká spořitelna, the largest Czech retail bank, which serves 5.3 mil. customers. We started six years ago setting up the data mining team of three core members. Today the team consists of eight people, has wide acceptance within the bank and was recognized by the SAS Customer Intelligence User Award in Cary, NC in October 2008. We believe that the success was fueled by the close cooperation between the data mining team and the direct marketing campaign management team in the bank's retail department which resulted in a successful SAS Marketing Automation implementation.

DATA MINING LIMITATIONS

Before explaining the path to profitable analytics we chose in Česká spořitelna, let's stress some limitations which analytics has in the banking business environment:

- Data mining needs data from the past to be analyzed. It is obvious but with important consequences. Where data is missing, inaccessible or of a poor quality, no analytics makes sense and the organization has to do the homework first to prepare a source of quality data.
- An ideal data mining playground in the direct marketing area means large enough, repeated campaigns in a stable environment. Unfortunately, quite often it isn't the case as new products emerge, new benefits are offered, new client groups addressed, new communication channels used, client's behavior or sales staff incentives change or the data history is insufficient. In such cases it is often necessary to consider test campaigns first to build the data history and real life experience. Sufficient time and money for testing is not always available because the business environment may force fast reactions to market changes and immediate full campaign launch. However, it is not always that urgent and test campaigns may be incorporated in the campaign planning. Of course the testing period must be long enough to allow time for test campaign responses to arrive and for analysts to look at them and analyze their structure before the main campaign preparation starts. Only then the main campaign may benefit from the knowledge gained from the test campaigns fully. The main campaign communication and/or target group selection can be adjusted to reach the best results.
- Another limitation has its roots in minds of some product and segment managers. They strive to meet their selling targets. All means are the right means and the larger campaign for their particular product or segment the better. They want to address as many clients as possible not seeing the potential benefits to the company as a whole if some (or even most of the clients) were addressed by another more appropriate offer. In fact this kind of "we want them all" thinking makes data mining useless. Predictive models can be developed to find the most likely responders even taking into account the expected volume sold. However, if all clients eligible are really addressed by the direct marketing campaign then no space for knowledge derived by predictive analytics remains. It isn't always wrong as sometimes the campaign is so powerful that even among the least likely responders enough sales happen. In my experience this is rather an exception though.
- Data mining can discover new knowledge about the clients' past behavior. In this extent it can support but not lead development of new benefits, communication messages, products and services for the clients. Imagine a snowball rolling down the hill. Using analytics you can predict its direction, speed and even how its size grows on the way. Using this knowledge you may become excellent in predicting the behavior of future snowballs. However, you can hardly predict the thaw that will melt all snow and change the conditions completely.

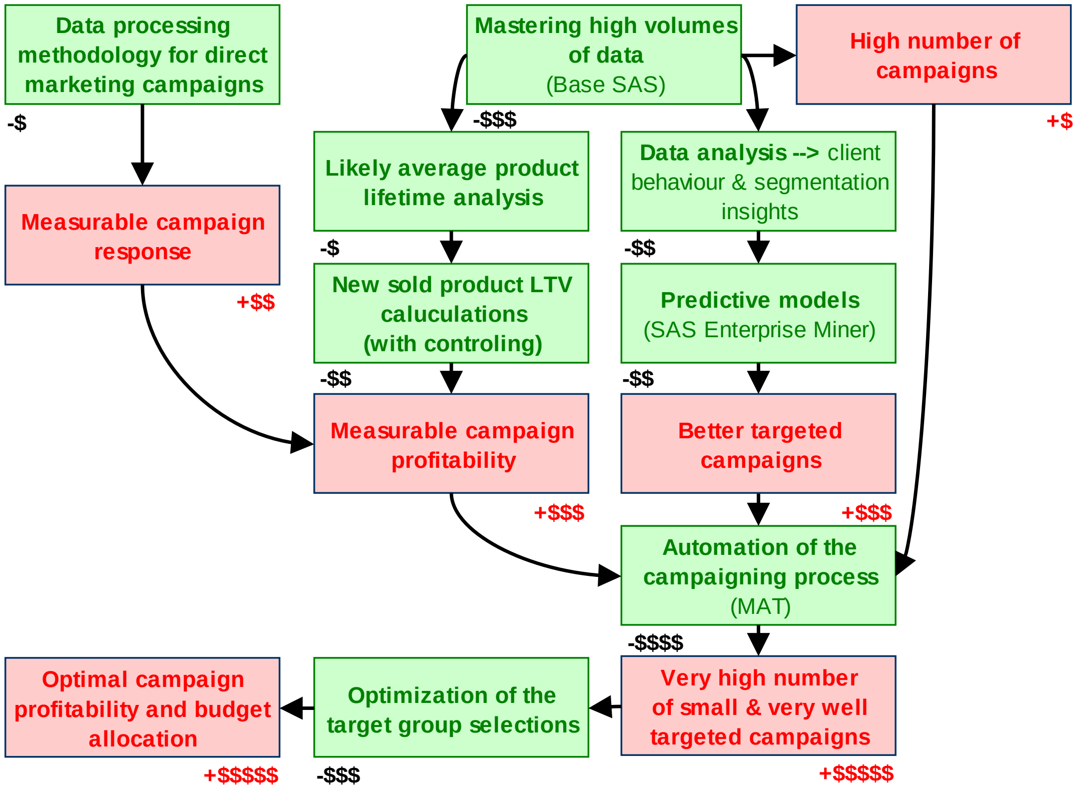
Unfortunately for analysts, the market isn't just a simple snowball and one of the strategies to achieve a competitive advantage is based on being the first to try something new on the market. Then inspiration from different markets, from market research, R&D or elsewhere should play the leading role and analytics support it with predictions, calculations, what-if scenarios etc. Analytics is powerful but not almighty.

- The nature of data mining implies that its outputs are tricky and difficult to explain quite often. The business people are excellent in understanding and influencing the client's behavior. They are not always statistics fans. The business without analytics would be blind and the analytics without business empty. This is why a certain level of trust from the business side is necessary. Even when the trust is there the analysts have to be as simple in their explanations as possible to make sure that the newly mined wisdom is used for the business purposes correctly. Finally the quality of the complicated predictive models can be measured quite simply using the campaign response rate and profitability calculations. This is why campaign evaluation as well as its presentation to all stakeholders is very important.
- We don't need to argue that data mining is more time consuming and requires a certain level of expertise. The time requirement has to be taken into account in planning of analytically aided business activities. The expertise may be purchased from specialized companies initially but should be built internally for the sake of continuity and incremental knowledge development.

THE PATH TO PROFITABLE ANALYTICS

On our path to profitable analytics we could observe several important milestones. They are described in the picture below. The analytical and data processing activities as well as the SW implementation activities are the cost drivers. They are represented by the green boxes with negative dollar signs in the chart. The negative dollar signs stand for the expenses and difficulty of completing the tasks. The red boxes with positive dollar signs stand for the benefits that the analytics brings to the bank. The positive dollar signs visualize the size of benefits that can be achieved. The black and red dollars aren't of the same volume. The red dollars on the benefits side are much heavier than the green ones on the costs side. The individual milestones on the path are described in more detail in the following sections of this paper.

Figure 1. The path to profitable analytics



MASTERING HIGH VOLUMES OF DATA

There is a variety of data available in the banking business. Banks can make use of the rich internal data. There are or can be made many client, product and transaction level data items available. New customer level data may be captured in the bank's CRM system. However, adding interesting data fields into CRM has to be complemented by a relevant motivation for the front office employees to fill in and update the data properly. Another potential data source is the internet. Identification of the bank's web page visitors opens new space for data based customer knowledge discovery with a great business benefit. External data sources are also relevant. External databases on individual clients have different quality and accessibility in different countries from being prohibited by law to comprehensive personal information collected, updated and sold by specialized companies. External data provided by market research may be useful for generic analyses providing soft information about client behavior but cannot be collected and updated for the majority of customers because it would be too costly.

Not only the data availability but the data quality is crucial as well. Data should be as accurate as possible for reasonable costs. The good luck for business analytics is that for most purposes 99 per cent data correctness is sufficient. Of course 100 per cent is better. But in some cases the strive for 100 per cent correctness may be too costly and the distortion caused to the analytical results by 1 per cent incorrect data is acceptable and can be interpreted in analysis conclusion. Business analytics isn't accounting. The results usually don't need to be 100 per cent correct in exact numbers but have to be very precise in business interpretation of data. E.g. a business question about who should be addressed by a campaign for mortgages could be answered from the raw data point of view so that 70 per cent of mortgage owners are men thus male customers are the right target. Business interpretation must go deeper demonstrating that women are mortgage co-debtors usually. Thus the customer's sex doesn't matter that much.

For deeper analysis the raw data should be as detailed as possible. This means in the banking business the granularity level of individual transactions with further details about them. However, for the start monthly aggregates on client and account level are sufficient.

Fresh data is important for up to date campaign evaluation, analysis of the latest developments and timely identification of business relevant events derivable from the clients' behavior. For other analytical purposes monthly aggregates are sufficient.

Data history is an issue when analyzing trends in time, preparing predictive models, etc. For most purposes 4 years of data history are sufficient. However, 6 months of data history should be available at least and 2 years of data are the minimum for analyzing yearly seasonality.

Our starting point was on a low level of data handling and analytical capabilities in 2002. All data requirements were passed over to the IT department. The data sourcing and testing took two weeks at least. The same process applied for data selections for direct marketing campaigns. The Oracle data warehouse and operational Siebel CRM system were under construction. By that time a wise decision was taken to set up the data mining team immediately to allow enough time for its establishment and know how development. Sufficient data was available in .txt files extracted from the pre-CRM/DWH systems.

The SAS tools were evaluated as the best of the breed and the server version of Base SAS, SAS Enterprise Miner and SAS Enterprise Guide were installed on a dedicated server. These powerful tools proved to be capable of handling the large multigigabyte data tables effectively. With this expertise we were technically enabled to produce a high number of direct marketing campaigns as well as high quality analytical outputs. But the newborn analytical team consisted of three people only, had no reputation from the past, campaigning processes were quite inefficient and campaigning methodology medieval.

DATA ANALYSIS

Banks produce and store large quantities of data because of the nature of the banking business. Storing the data is a must. Analyzing the data provides new insights that can be used by the campaign, segment and product management. Good direct marketing campaigns must be based on intimate customer knowledge. In this respect data analysis is a rich source of competitive advantage for those who do it right.

The SAS analytics enabled us to provide fairly new customer insights especially in the following areas:

- Survival analysis of bank's products based on Cox regression brought important insights into the survival of bank's products in time. One of the outputs – the mean product lifetime – is used for product lifetime value (LTV) calculations.
- Predictive modeling of client's attrition and purchases deserves a dedicated section – see it below.

- Based on predictive models and LTV calculations we have developed the Next Best Product model. We score the whole client base of 5.3 million clients weekly using Base SAS and export the three most likely products to be sold for every client to the CRM system. The bank advisors started using this functionality spontaneously and like it a lot. Every month approximately 40 000 sales hints (based on our next best product model) are processed by the bank advisors. The success rate is about 10 percent of sold products.
- Behavioral segmentation of the bank's private clients provided a usable alternative to the previous purely demographic segmentation.
- Analysis of clients' households enables us to use the householding information in direct marketing campaigns and in analytics.
- Many smaller analyses of client's behavior and product usage.

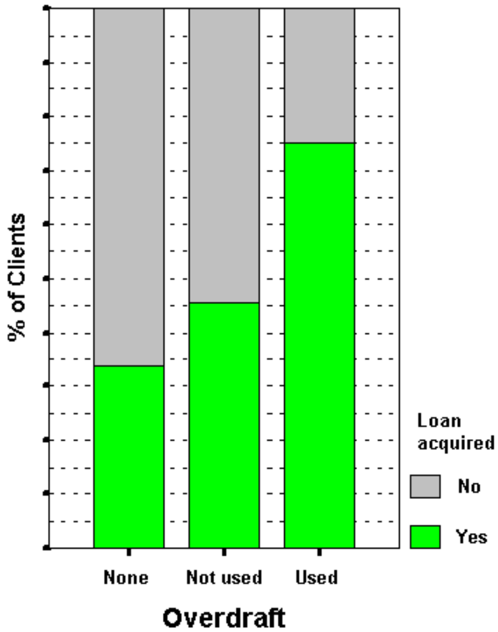
PREDICTIVE MODELS

Predictive modeling is the core competence of any data mining team. For predictive modeling a proven methodology should be used, for example SEMMA (Sample, Explore, Modify, Model, Assess) method developed by SAS and supported by the SAS Enterprise Miner data mining software. Before SEMMA is used data have to be prepared. The initial data preparation can be done in a data mining data mart or ad hoc. Either the data mart preparation or programming macros for ad hoc data preparation is a subject of a small to medium project. The regular data preparation for modeling takes approximately as long as the data processing and analysis using SEMMA. But it can be shorter when using a data mining data mart.

For every new predictive model several parameters have to be set in cooperation with business experts:

- The target behavior that should be modeled. Most typically it is purchase of a new product or leaving the bank. It could also be activation of a product or the expected volume purchased. Seasonality of purchases has to be taken into account if significant. Sufficient number of past target behaviors must be available in the data. The minimum is 200 cases but at least 500 cases are recommended.
- Special variables that could be useful for prediction of the particular target. See an example of a useful predictor in the figure below. It shows the relationship between consumer loan purchases and overdraft to current account. Customers who don't have or have but don't use the overdraft buy consumer loans with half the frequency of customers who use the overdraft.

Figure 2. Example of a predictive variable



- Group of clients for whom the model should be prepared. It should be large enough to allow client prioritization for campaign. But it shouldn't contain clients who aren't eligible for the campaigns envisioned. Including such clients could disturb the model performance unnecessarily.
- The prediction lag between the predicted behavior and predictive data. We usually use one month.

Predictive models can be widely used to increase the effectiveness of direct marketing campaigns by suggesting the most likely buyers of the banking products, buyers with the highest expected volumes or the most likely churners. In Ceska sporitelna we maintain over 40 predictive models in production covering all major product lines of the whole financial group.

The performance of predictive models degrades over time with changes in customer behavior, selling priorities and data quality. This is why permanent monitoring of model performance is important. When the model performance degrades remodeling is necessary.

DATA PROCESSING METHODOLOGY FOR DIRECT MARKETING CAMPAIGNS

The target isn't just a high quantity of direct marketing campaigns based on customer insight. Reaching maximum profitability requires addressing the most appropriate clients as well as measuring the campaign response and profitability. Let's have a look at the direct marketing campaigning process and describe the role that analytics should play in it:

1. Campaign initiation

Analysts have a great insight into data and aren't directly motivated on sale of particular products or services. When cooperating with the business people on product and segment analyses they may discover interesting customer segments or underestimated products and product features that could be subject of future campaigns. Based on this knowledge they can suggest new campaign ideas.

2. Campaign design

Campaign design aims to address the right customers at the right time with the right message by the right communication channel. Analytics may support all of these elements and it plays the major role in the selection of the right clients.

Predictive models should be used for improved campaign targeting. Before using them the campaign and product specific selection criteria must be applied. E.g. exclusion of clients who do already possess the product offered, who don't meet the envisioned age profile for the communication used, etc. Obligatory exclusions have to be applied as well, e.g. exclusion of clients who opted out from campaigns, who don't meet the obligatory credit risk criteria, etc. All of these criteria shouldn't be too restrictive if a good predictive model is used. Otherwise the predictive power of the model is underutilized or not used at all.

Applying a well defined client contact policy is also good advice. Too frequent contacts to the same customers diminish their satisfaction with the bank's service, increase their resistance to the bank's direct communication and opt out likelihood. Besides that, not communicating with other customers on a reasonable frequency is a likely missed selling opportunity. This is why frequency of contacts via the same communication channel or with the offer of a similar product should be set.

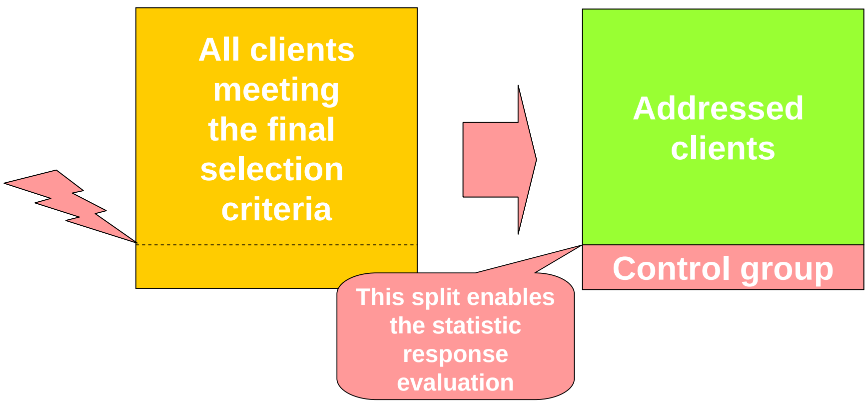
Optimization of which clients to target with which campaigns is more important the more campaigns a company produces that are subject to the client contact policies. Where many campaigns compete for the same clients at the same time someone has to decide about their priority. When not optimized, campaigns take clients as they go. If clients for a low priority campaign are selected first, then a later more important campaign cannot address them within the period restricted by the client contact policies. On a higher level of optimization awareness the prioritization of campaigns can be made and clients for more important campaigns are selected first. A further step towards high quality optimization is selecting clients for the future more profitable campaigns in advance. The clients can be booked for the future campaigns preventing their ineffective utilization by earlier campaigns. The analytical team can help in prioritizing and booking future campaigns. The top approach involves utilization of a specialized campaign optimization module. This is an opportunity for further development in Česká spořitelna.

3. Target list creation (including control groups)

For the purpose of statistically sound campaign evaluation a statistically correct campaign design is necessary. It encompasses creation of control groups which are representative for the target audience.

After the final selection criteria are applied to the client base, the control group has to be separated randomly. The control group won't be addressed by the campaign to measure the spontaneous buying behavior. Besides that the further handling of the addressed clients and the control group has to be the same (including the client contact policy). The size of control group necessary for statistically correct campaign evaluation depends on the expected response rate difference between the control group and the targeted group. The lower the difference the larger control group is necessary. The real campaign effect can be measured as the difference between the response rate in the addressed group and the rate of spontaneous purchases in the control group. Separating some clients into a control group that isn't addressed by the campaign means that for the clients in the control group no additional sales caused by the campaign can be achieved. This is a lost business but it is inevitable. Because the spontaneous sales are sometimes surprisingly high and ignoring them could lead to supporting completely ineffective campaigns.

Figure 3. Random control group separation



Typically the control group isn't targeted by the campaign. However, sometimes it is also useful to create further control groups randomly selected from clients who wouldn't be selected by the predictive model and addressing them to be able to measure the power of the predictive model. In this way the selling effect of the predictive model can be separated from the selling impact of the campaign message and benefits and learning for further model improvement can be derived. These client selection techniques are the base for statistical campaign evaluation. On this base and using the LTV of newly sold products we calculate the profit and ROI in our campaigns. Using this knowledge we decide about continuing or modifying the campaigns.

The logic of sales comparison in the addressed and the control group can be used for techniques like champion/challenger testing as well. The challenger may be defined by different selection criteria or the same selection criteria but different offers or unique selling propositions for the test groups.

4. Direct mail/telephone script preparation

This step is the task of the marketing and communication specialists. They should be using analytical outputs describing the target audience of the campaign to make the communication as relevant as possible. More relevance can be achieved by a creative usage of the information on customers or their products extracted from the bank's systems.

5. Marking clients in CRM

This is a technical step aiming to spread the information about which clients were or should be addressed by which campaign throughout the organization. Optimally it should be automated by a marketing automation tool.

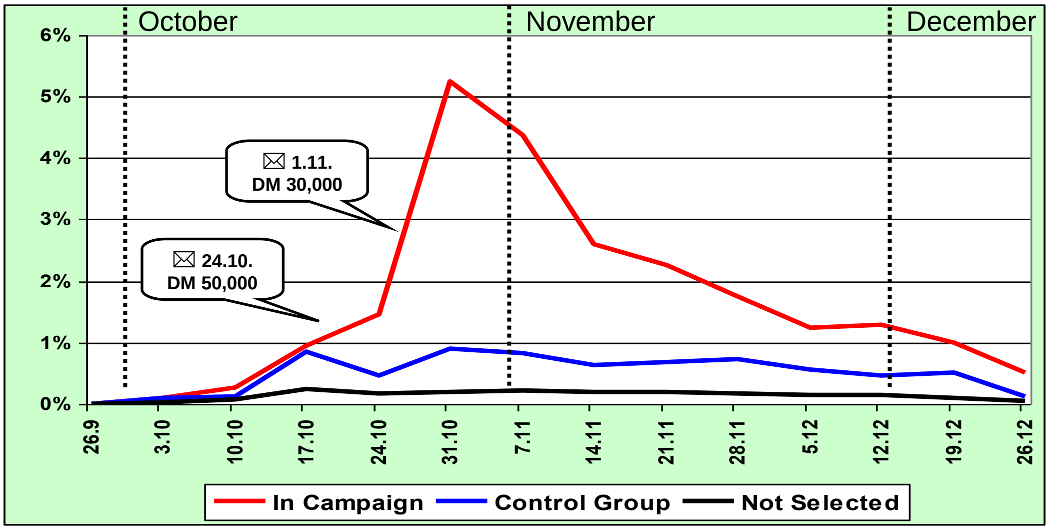
6. Campaign execution

The campaign success or failure depends on this step. In case of direct mails, the right address and name, correct and up to date additional client specific information printed in the letter have to be used. The message has to attract the customers to the bank branches or motivate them to buy the product offered in another way. The simpler and the less time consuming the purchase is for the customer the more likely it happens.

7. Campaign evaluation

The evaluation of the real campaign business contribution is based on the sales comparison in the addressed and control group of clients. The chart below shows the development of a consumer loans campaign weekly response rate visually:

Figure 4. Campaign weekly response rate development



We can observe how both the target group of clients in the campaign and the control group behaved similarly until 17.10. Then the clients got direct mails and the difference in buying behavior caused by the campaign became obvious. The red line remained about 0.5 per cent above the control group even after Christmas. The black line stands for clients who were neither selected by the predictive model for the campaign nor addressed by the campaign. It can be compared with the blue line. The difference shows the different spontaneous buying behavior of the most likely responders selected by the model and the average remaining clients. It clearly demonstrates that the campaign was targeted at the clients with higher level of spontaneous purchases.

When assessing the campaign result using this incremental measure even a longer period after the campaign end can be used. We are using the period from two to three months after the campaign end for the final campaign evaluations. Summing up all the differences between the red and blue line provides the best estimation of the real campaign impact to the client's behavior. These sales multiplied by the lifetime value of the products sold are equal to the campaign earnings. The campaign costs consist of costs for addressing the clients, costs for campaign preparation and costs for the campaign benefits provided to the clients in form of gifts, discounts, etc. This enables us to calculate the campaign profit and ROI.

Not only the main product in the campaign but also the other bank's products sold during the campaign should be considered in the campaign evaluation. Some products support sales of complementary products whereas other products cannibalize the sales of similar products. These effects are sometimes obvious but often just suspected or even completely illusory because of statistical insignificance of the differences.

The campaign responders should be looked at to compare the initial selection criteria with the characteristics of the clients who really bought the products offered. For large enough and repeated campaigns specialized predictive models may be developed.

The campaign evaluation should answer the question about campaign profitability and provide important insights useful for further campaigns. If the campaign performance monitoring is already done during the campaign then corrective measures can be taken to influence its results positively.

We support the optimization of the direct marketing budget by calculating the optimal number of clients to be addressed by different product offers and campaigns. Our input is based on the calculation of expected profits from addressing clients in individual client deciles when ordering the client base by the probability to buy using our predictive models. The offers are recommended only for clients from the profitable deciles.

AUTOMATING THE CAMPAIGNING PROCESS

Mastered high volumes of data, established campaigning process, statistically sound campaign profitability evaluation and very good targeting based on predictive models constitute the environment for effective direct marketing. The technology of today allows another kick – automating the whole thing.

Our motivation to implement the SAS Marketing Automation Tool originated in the still growing internal demand for direct marketing campaigns. Using just Base SAS and SAS Enterprise Miner we were able to generate hundreds of campaigns a year but it still wasn't enough. Requirements for monthly repeated permanent campaigns emerged and occupied our capacity to work on analytical tasks even more. In the year preceding Marketing Automation implementation we already produced more than 1000 target lists for direct marketing campaigns a year. All the data preparation had to be done by the data mining team as ordered by the campaign managers. Planning and management of the campaigns was supported by programmed Excel forms and tables only.

This is why the implementation of SAS Marketing Automation was a logical further development step to consider. We had quite a clear idea about the campaigning methodology and didn't want to compromise it. So we were quite a demanding customer for SAS. The implementation was successful and brought especially the following benefits to us:

- It enabled us to empower the campaign managers enabling them to prepare the target groups for campaigns. The initial concerns about teaching campaign managers who aren't programmers to select data in the Marketing Automation GUI turned out unjustified. The data mining team has the capacity to do the analytical work again.
- The data usage for campaigns remains acceptably flexible. The tool has a practical GUI where more than 500 data items are ready to be used for target group selections on all granularity levels including individual transactions. In case that wouldn't be enough a simple import of external data extends the possibilities. We are confident to be able to perform even the most difficult data selections as also a code node may be included in the selection diagram containing just any piece of SAS code.
- Our current client contact policies on client, communication channel and product level were implemented without substantial changes.
- One of the main benefits is seen in the permanent campaigns which run on a regular schedule, further increasing the high number and profitability of direct marketing campaigns.
- Support for campaign approval process is provided, improving the managerial control over the campaigns.
- Standard interfaces to communication channels and CRM increase our speed to market and stabilize the output quality. We are communicating to our customers via the ordinary communication channels like direct mails, call center and SMS as well as the bank specific communication channels like ATMs, Internet banking and account statements.
- The automatic weekly campaign evaluations are accessible by the campaign stakeholders. They were configured to reflect the campaign methodology using control group based response evaluation fully. The campaign definition and evaluations remain accessible in the tool for informative purposes after the campaign ends. A kind of wide campaigning knowledge base is being developed in this way.

- Our bet on a simple administration of the tool was a good one. Only one administrator located within the data mining team is necessary to support the solution. We are purchasing 40 mandays a year from the Czech SAS office for the necessary system maintenance and as a support to our continuous internal development. This enabled us to avoid higher personnel costs of standard IT support departments which only maintain the Windows installation and the servers technically. Another benefit is seen in our flexibility to develop always the most urgently needed functionalities, interfaces and additional data items.

The implementation itself was done as a project supported by an integration partner. We have benefited from our clear idea about the campaigning methodology and campaigning processes a lot. One of the most time consuming implementation tasks was testing all the data imported from our data warehouse.

CONCLUSION

The paper comments on the path to profitable analytics of the large Czech retail bank Česká spořitelna. Start-up of the data mining team as a part of the CRM/DWH project was a wise decision which enabled further development. Using the results of high quality analytics and the computing power of the SAS server we have been able to boost the number and quality of direct marketing campaigns enormously. The number of target lists increased 40 times within 6 years from 30 a year in 2002 to more than 1200 in 2007. The share of Direct Marketing sales of total retail sales has grown from close to zero to 30 percent in consumer loans, to 33 percent in overdrafts on current accounts, and to 60 percent in credit cards. This means a yearly profit of more than 30 million USD. This was supported by the establishment of a dedicated campaign management team responsible for all direct marketing campaigns in the bank and the most important daughter companies.

A substantial qualitative improvement was achieved by the successful SAS Marketing Automation implementation. A marketing automation implementation project is a challenge. It has the potential to advance the bank to another level of direct marketing effectiveness. It is an adventure like everything new and not so easy to achieve. It is quite hard to imagine how it will work in a particular organization context before it is implemented. The adventure can be lived through more safely when the organization is prepared for it. Then the new horizons of higher effectiveness can be discovered. The decision to implement the marketing automation is binding because the investment and internal effort is significant. When successful, the organization can become the direct marketing leader in its market and think about even more advanced software implementations in the direction towards campaign optimization, detection of events, real time marketing, etc.

RECOMMENDED READING

Bauer, J.L. a Kelly, Ch.: Database Marketing for Retail Banks and Financial Services Companies, Lafferty publications, Dublin, Ireland, 1997, ISBN 1-86068-0550

Lo, Victor S. Y.: The True Lift Model – A Novel Data Mining Approach to Response Modeling in Database Marketing, ACM SIGKDD Explorations Newsletter, Volume 4, Issue 2, ACM, New York, USA, 2002, ISSN 1931-0145

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Petr Ptáčník
Enterprise: Česká spořitelna, a.s.
Address: Olbrachtova 62
City, State ZIP: Praha 4, 140 00, Czech Republic
Work Phone: +420 724 160 213
Fax:
E-mail: pptacnik@csas.cz
Web: www.csas.cz