

## **Jean-François Michaud (Cont')**

**Partner/Family:** Small family with my girlfriend and me.

**Pets:** None for the moment, but I'll get a parrot one day

**Sports/Hobbies:** Hockey, roller hockey, salsa, swimming.

**What my ideal weekend would be:** Weekend -- that would have more than two days :)  
-- in the Eastern Townships at my lake cottage.

**Favourite Foods:** It's sad I have to choose... but since I have to, I'll say sushi with a buttery chardonnay.

**If I could be anything at all (besides a SAS consultant),** I would be.... An archeologist, digging dust instead of data. :)

**When I'm not involved on SAS consulting projects, I like to .....** Rollerblade, go to the movies, dance salsa and go to wine tastings.

**Something I've learned out in the field doing consulting work for SAS that I feel would benefit other SAS users is ....**

Starting a first data mining project is always an exciting challenge. Whether you are a member of the customer intelligence team, a statistician, a business analyst or the director of the analytics department, as a future data miner, you better ensure and prepare so that you have all the necessary resources.

In data mining, success really depends on integration. Three components are essential: business objective, tools and data.

Too often, the focus of data mining projects turns quickly to what is available in terms of explanatory variables. The project is approved and the race is on to gather all the data, aggregate it and bring it into our magic data set. Unfortunately, the time necessary for this process is often underestimated. Scores deployment is also an extensive process. Deployment will need to involve the IT group for data preparation, and also for the distribution of scores to the business units that will act upon it.

So many resources are involved at every stage of the process that the project must begin with a clear focus, or it will quickly veer off track.

Before the project begins, and before the development of any data, many questions need to be answered.

For example, if a company decides to address churn, how do you define churn, attrition or retention? What are the dimensions that qualify that phenomenon? Voluntary churn?

Involuntary churn? Length of association? Should we look at this year's churn only, since it's on the rise? Hmm! Does that impact the way we should develop our data set? It sure does!

Data preparation is an extensive project for any company. The information needed is typically on different systems and not at the level of aggregation needed for mining. Given all the work involved in these transformations it is critical to properly define the business question to address and translate those questions into a proper target.

The business objective must be determined from the beginning, and made clear to all parties taking part in the project. We usually start projects with a kick-off meeting with each of the groups involved (IT, database marketing, credit). The objective of this meeting is to give each party a clear view of the global process involved and their inputs to the project.

The initial goals for the data miner include grasping the business problem the company is trying to address, understanding what information is available and the format the information is in.

The information is usually available at the transactional level, which is not the right format for mining. For transactional data to be useful it must be aggregated at a customer level that makes sense given the nature of the business and the business question at stake.

Once the miner has a good idea of the business objective, he or she should document the past evolution of the event of interest or target. The observation of its evolution and importance from month to month helps in the selection of the proper time frame for the event of interest.

If historic data shows the event is rare in the population, the miner might choose customers who had the event within different months. There are different scenarios for targets construction. One common pitfall is to use a variable as the target because it is a variable available in the systems -- the idea being that it has the seal of approval of finance or of any other group within the company. If your target is constructed around the way the company process works, it will tell you how you treat the customers and how your company works. If that is the business question of interest, then you are on the right path; but aren't you more interested in the customer's behavior and how you can influence it?

With the proper understanding of the business goal, the data miner is now in a good position to come up with a recommendation of what we typically call in the mining world "the target." The target is the point of equilibrium of the data mining exercise.

If the target doesn't reflect the business goal, you can mine as much as you want. You might find something, but it will be of no use to the business.