XLNT consulting

Data Quality ROI
transforming soft measures into hard dollars

Tom Breur
December 2009

Introduction

Hard business cases for data quality are rare. Yet they aren't all that difficult to construct. So there must be another reason for their absence. Possibly the most prevalent explanation is that professionals are either unable or unwilling to associate "soft" outcomes with "hard" dollars. Yet without this connection, considerable ambiguity remains as to "what" the issue is, really, and how much it is costing.

What are some examples of "soft" data quality problems?

- "A single version of the truth" is the business intelligence (BI) holy grail. What we mean is that there is less confusion and useless debate among data users. Since this is non value added discussion, the time spent on it is a net and unrecoverable loss to the business.
- "More satisfied customers" are said to buy more and stay longer. And besides lower churn and better cross- and up-sell potential, they can also become ambassadors for the brand. Although customer satisfaction is "obviously" desirable, you need to dig a little deeper, do advanced research, to relate satisfaction to bottom-line figures. But how else can you arrive at a rational allocation of resources?
- "Higher employee morale" has been repeatedly shown to have a positive impact on the bottom line. And conversely, data quality issues (inaccurate or incomplete data) can make it very frustrating for staff to serve your customers properly.

Barring a sound financial underpinning, the decision which initiatives to fund and which to curtail is at risk of corporate political chicanery. You can spend every dollar only once. To improve data quality? To provide better service? Or to make your inherent offering more valuable? Where do you get the best return? Again, without hard (dollar value) business cases, it's anybody's guess where and how much to invest in data quality.

Examples of how to turn "soft" objectives into "hard" numbers

Some examples of *how* to relate "soft" measures to quantifiable business cases:

1 – *A single version of the truth*

Confusion and pointless debate over what data source to "believe" is a recurring issue. It hurts on two levels. This obviously hampers business decision making on a day-to-day basis. Olson (2003) has made a second, compelling case at a "macro" level. Businesses that control their data flows can gain access to new data sources faster, and can respond quicker and more agile to changes (and opportunities) in the market place. They will also notice *sooner* when something is changing.

To quantify the "micro" cost of poor decision making, it is necessary to make an empirical assessment of the frequency of data quality errors, and relate those to the cost of lost business. Time frittered away sorting out messy data, discussion between departments about "the right" reality, etc., can be estimated and quantified as lost employee productivity.

The "macro" value of better information provision is a bit harder to quantify. In particular in very dynamic markets this should be worth a lot. When innovations follow each other rapidly, the value of business model agility is a function of R&D investments which can be huge in some industries.

2 – *More customer satisfaction*

In his classic book, Frederick Reichheld (1996) showed how companies that managed to keep their customers happy and loyal were much more profitable (quoting research from Bain & Co). It turns out time and time again that driving up customer retention is a remarkably effective and powerful way to improve the corporate bottom line. More recent work by for instance Wayland & Cole (1997), and Gupta & Lehmann (2005) has reconfirmed these findings.

It is well established that customers with a "deeper" relation (more product holdings, more frequent usage) have a lower propensity to churn. However, discovering causality is far from straightforward. This is largely because the possibilities for true experimentation are limited in most organizations. Barring randomized experiments, you'll need to resort to quasi-random experiments or advanced modeling techniques that allow you to draw causal inferences from correlational data (Structural Equation Modeling, using tools like Lisrel, AMOS, BMDP, etc.).

One of the other interesting avenues in customer behavior modeling is scenario planning. This often involves the use of some variation of the Monte Carlo algorithm. When a relation between cross-sell and retention and/or further up-sell potential can be established, finding the exact mathematical specification would be exceedingly complex. This involves higher order Markov chains, which may or may not have an analytic solution (some problems are mathematically impossible to solve). However, using advanced analytics like Monte Carlo simulation you can still model dependencies between cross-sell options across a product line, as well as their relation to churn. This allows you to discover more beneficial “entry level” products to cross-sell, and these “best” product orderings are not necessarily intuitive. Needless to say, data of sufficient quality are required to support this.

3– *Higher employee morale*
It has been repeatedly established in research that satisfied employees are more productive. In particular at the operational level, you need to provide timely, accurate, and complete data, in order for people to perform well. Menial tasks like data entry are fully dependent on good quality data. But in particular customer facing staff (call centers, branch offices) can make or break your reputation.

Human factors research provides the scientific validation for discounting productivity based on employee satisfaction scores. Salaries are readily quantifiable, and 5-20% decrease in productivity has been observed in (several) experimental studies, so you can quickly, and easily get a range.

Also, happy employees stay longer. Besides, the (largely intangible) loss of knowledge, the cost of training and initiation programs can be multiplied by marginal turnover rates.

Relying on “bare” numbers

A word of caution seems in place here. Contrary to common wisdom, “plain” numbers don’t always speak for themselves. All too often we have observed discussions like: “We should be able to raise retention from 90% to 95%.” Unfortunately, not nearly enough attention is paid to the question how realistic that target is. Or: “We should be able to raise cross-sell from an average of 1.12 to 1.20, shouldn’t we?” How reasonable is *that* scenario? These “bare” numbers can be very deceptive.

The “surprises” when you try to estimate the likelihood of accomplishing target scenario’s are usually “hidden” in the operation somewhere. Sometimes it turns out that processing the “mere” 0.08 increase in cross-sell (going from 1.12 to 1.20) would require three times the peak capacity of your back-office for the entire year to process all the additional applications. The call center would need at least twice the staff to handle all the questions, etc. A “bare” number doesn’t necessarily show what “side effects” the achievement of realizing increased cross-sell would have.

And all these considerations are separate from the question how “likely” these accomplishments would be for marketing to pull off. In general we use history as the most reliable estimator for the future. Has your company *ever* been able to accomplish such growth numbers? Even in your most successful campaign? How many of those “homeruns” do you need to hit in order to string together this winning streak? And have you considered whether the organization at large would be able to absorb such phenomenal growth?

I *should* be able to raise my adolescent without running into conflicts. After all, I am a reasonable and mature parent, aren’t I? Well, anybody who *has* raised an adolescent should know better. And I, personally, am blessed with a very reasonable and sage teenage son. But we have our occasional clashes nonetheless!

Conclusion

Every situation, in every company has its own idiosyncrasies. But what they all have in common is that quantitative business cases for improving data quality are highly desirable. These “hard dollar value” business cases are simply the language of senior management. Even if it requires making assumptions (and *all* do!), without a measure to put urgency or importance into perspective, how can you get the proper management attention?

In this paper we have touched on just a few subject areas where quantifying a business case may be hard, but certainly not impossible. In business intelligence everybody “knows” you need a single version of the truth. Obvious as it may seem, you still need to establish a business case for it. The same holds for “soft” indicators like customer satisfaction or employee morale.

Don’t shy away from making assumptions in order to arrive at financial outcomes. Just ensure you make those assumptions explicit and transparent. If a point estimate (one single number) makes you uncomfortable, provide a range. Give a lower and upper bound for an optimistic and pessimistic estimate. But whatever you do, come up with a (financial) number. The one dimension that every manager, in every company understands so well is after all: “dollars.”

Larry English (1999) has written a great book that provides abundant examples of how you can relate business process breakdown to bottom-line numbers. He has also set out to establish a comprehensive overview of *all* ways that data quality issues can cause business process breakdown. These taken together provide excellent guidance to get a head start with establishing your business case.

References

Jack Olson (2003) Data Quality – The Accuracy Dimension. ISBN# 1558608915

Frederick Reichheld & Thomas Teal (1996) The Loyalty Effect. ISBN# 1578516870

Robert Wayland & Paul Cole (1997) Customer Connections. ISBN# 0875847994

Sunil Gupta & Donald Lehmann (2005) Managing Customers as Investments – The Strategic Value of Customers in the Long Run. ISBN# 0131428950

Larry English (1999) Improving Data Warehouse and Business Information Quality – Methods For Reducing Costs And Increasing Profits. ISBN# 0471253839