



"turning data into dollars"

Tom's Ten Data Tips – March 2010

### Value From Transaction Data

Transaction data are atomic building blocks that create value in the exchange between consumer and corporation. Extracting value from transaction data is a bit like trying to sip from a firehose. A value stream often consists of multiple (many) transactions. You need to consider these transactions *together* in order to understand the relation. And unless you understand *the relation*, you cannot interpret separate transactions.

Individual transactions are surrounded by "context" which consists of all the descriptive attributes. Think of descriptions like time and place of transaction, sales channel, employee handling the transaction, etc. Your data model surfaces tacit knowledge within the organization about business dynamics and value creation.

#### 1. No Value Without The Appropriate Data Model

Transaction data only become valuable when you understand the context in which they should be interpreted. In terms of star schema modeling this implies ferreting out "the right" dimensions surrounding the central fact table with individual transactions. Because a transaction processing system (OLTP – On-Line Transaction Processing) is tuned for performance (*fast* processing), the transactions tend to carry very few descriptive characteristics. So all these surrounding dimensions need to be constructed (derived) from a deep understanding of the system.

Constructing the context of transactions, typically, is hard BI (data modeling) work. But unless these dimensions are tied in with the transaction, you will not get the full value from your transaction data. Examples of descriptors that are not (evidently) contained in the transaction might be: customer segment, sales channel, branch office, sales person, sales package (offering/campaign), competing alternatives available, etc. Unless you "add" all these descriptors (dimensions) you will not properly understand the context (and thus, "meaning") of the transaction.

## 2. Let (*Help*) Your Customers Vote With Their Feet

What customers are willing to pay for, and how much, is extremely valuable business information. Of course a sales ticket only gains meaning after you define the context surrounding it. And that context could be other products in stock at the time of purchase (and their prices), repeat purchases, discounts, etc. This is the most valid evidence of customer preferences, it's the *actual behavior* they are displaying, and the *money* they are spending.

In a broader sense, the context of a transaction could also be sales happening in other (similar) stores, and how those stores or customers compare. Part of the context can be which products are purchased together, often referred to as market basket analysis. When you combine many, many sales tickets (with multiple items), you can infer a clustering of your assortment that is truly customer driven. It represents the way *they* perceive an (often implicit) structure within your product offerings. The possibilities are endless.

## 3. Transactions Initiate Trigger-Based Marketing

Event-based marketing are actions that are triggered by changes in the customer's life. The term trigger-based marketing is also commonly used. We would consider an "event" a complex, multi-faceted occurrence in the customer's life. "Some" (unusual) transaction will be a signal this event has taken place. For instance, a customer who has bought a house will subsequently change address. Or a (female) customer who gets married changes her name. A customer reports a stolen credit card, etcetera. Sometimes it is clear from the transaction which event has taken place (like in the case of a female changing her name after getting married), and sometimes it isn't.

If you understand the implications of an event to the customer's life, it can help you in servicing the customer better. Or possibly selling additional products. This can become a very efficient means of interacting if the campaign or dialogue follows *automatically* from the transaction that triggers identification of the event.

## 4. Automated Processing Of Unstructured Transactions Requires Textmining

Transactions can come in as structured, semi-structured, or unstructured content. When processing volume exceeds even modest amounts, it quickly makes sense to automate some of the reading,

sorting and interpreting steps. Let's look at some examples. Mail (or email) coming into the customer contact center (CCC) needs to be sorted, for instance (classification). After scanning the content (optical character reading – OCR), incoming letters might need to be routed as information requests, change of details (people moving, for instance), and complaints. Complaints might need to be further sorted into "ordinary" and "severe" complaints. Etcetera.

All these are text mining efforts that have matured in past years to the point where automated algorithms can be tuned to outperform humans. The main reason for this is inter observer reliability. When the work is spread between more than a handful of humans (as will be the case in a good size CCC) it becomes exceedingly difficult to manage consistency *across* employees. Although the algorithm may not outperform *one* employee, it will outperform five of them because of inconsistencies in their judgment. As an added bonus to automating these processes (with human intervention for exception handling) you gain control and facilitate monitoring of back-office processing. These non-value adding activities are perennial sources of efficiency improvement.

## 5. Feature Extraction Provides Data Value

The term "feature extraction" is commonly used to describe the process of parsing unstructured fields, and deriving meaningful information from it. In clickstream analysis (see also tips# 7 & 8), for example, information about the user's browser is derived from text fields lacking a standard layout. This means that you need to parse these fields, and glean the crucial information from it. Then you can code that information into a limited number of distinct browser types/versions. Likewise for the screen settings the user has for surfing the internet. This is how websites "know" you are accessing their site with a PDA, for instance, so they can serve up a dedicated version of their site that was optimized for handheld devices.

In the context of analyzing transaction data, we refer to feature extraction as the process of appending meaningful, useful context information to atomic transactions. In a star-schema, there are theoretically an almost infinite number of fields that could be derived. The exact number of potential derived variables is the product of the number of unique levels for all attributes multiplied, which for even a moderate star schema will equate to more than the number of atoms in the universe. Infinite, for all intents and purposes. To deal with the "curse of dimensionality" (winding up with *too many* columns) you

need to be selective when calculating these fields, very selective indeed. Inclusion should be driven by value to the business, cost of calculation (which can easily become prohibitive), and overlap/correlation with already available attributes.

## 6. How Much History Should You Keep?

Because of the sheer disk volume transaction data use, questions will be raised when data can be archived or disposed. An alternative intermediate strategy might be to partially aggregate or summarize transactions at a coarser grain. For instance, roll up transactions into a total for the day, week, or month, etc. That way you can choose how much space you need to save.

The risk of aggregating prematurely is that cannot “undo” it. So all detail that is lost in that process is lost forever. And you can’t positively “know” now what future use might be required for those details. Bill Inmon and Dan Linstedt have coined a term “temperature” of data: as soon as less and less requests are made for data, they “cool off”, and moving the data to a near line or archival sector comes under consideration. See Inmon et al, DW 2.0 (2008).

## 7. Clickstream Data Are A Special Flavor Of Transaction Data

Clickstream data (“log files”) are the elementary transactions that describe at the finest grain how people interact with a web site. There are, fundamentally, three ways to capture clickstream data (see also a whitepaper we wrote on this topic) server logs, user side, or ASP. User side data capture has a few unique advantages. However, because it has become so closely associated with spyware, that option is really only available for captive audiences like intranets, etc. Server logs provide the most detail, up to the level of individual gifs that make up a web page. ASP’s have the advantage that data capture can be geared (beforehand) to the intended analytic use. So all three methods have pros and cons.

Individual mouseclicks make up the atomic level elements that together comprise a user session. Page elements (text, gifs, etc.) roll up to page views, and page views strung together form a path through your website. One session can contain several paths, several “tasks” a user is trying to accomplish. Together they make up the “value creating” exchange between a visitor and a website. Some context (see also tip# 1) can be derived from the user’s machine as this

information about his configuration is exchanged with the server in the process of delivering the requested pages.

### 8. Log-File Analysis Is An Experimental Goldmine

When customers interact with you through your website, this provides unique opportunities. Clickstream analysis (see also tip# 7) allows you to analyze intricate details of their behavior. When you serve up dedicated “experimental” pages to sub-groups of your visitors, you can manipulate what customers see, and test their reactions. Since customers vote with their feet anyway (see also tip# 2), you help them “teach” you what works for them, and what doesn’t. It is hard to overestimate the value of these marketing “playgrounds.”

### 9. Privacy Is A Thorny Issue

When you attempt to extract value from transaction data, the question of privacy protection should be explicitly considered. Analyzing these data can be so revealing, that the analyst might feel a bit like a “voyeurist.” Needless to say, data protection acts have considered these aspects, too. Although legislation is different across the globe, even the way privacy laws are *structured* (prescriptive versus normative), there are some general tenets.

One of the principles is whether use of data (transaction data in this case) serves the customer’s needs. If you can rightfully make that claim, far more liberties are permitted. Safety, security, the customers’ best interests are such potential applications. A second recurring principle pertains to “sufficient care”, which translates into “careful processing.” That usually means ensuring auditable working practices, and *demonstrating* (by empirical tests) that your queries are accurate. Of course there’s a difference between legal and publicity risk here, so you’ll probably want to stay firmly within bounds.

### 10. Transaction Data Are Both A Challenge And An Opportunity

Data are often said to be a company’s most valuable asset. Because almost any other asset can be copied or bought by the competition (*including* talent), it is hard to overestimate the value of exploiting one’s data in creating sustainable competitive advantage. If that is true, then *transaction* data are the real cornerstone. They can be aggregated in an infinite number of ways, and combined with myriad qualitative resources. This quickly becomes daunting, both technically as well as conceptually.

The keyword is learning how to *exploit* one's transaction data. Because of the sheer volume, volatility, and innate fickleness, harnessing the value of transaction data remains one of the frontiers of intelligent data analysis.