



Diapers, Pop-Tarts, and dog food

The case for data mining

By John P. Girard

In the early 1990's, the retail industry watched with great interest as something called data mining gained popularity. For the first time, extremely powerful computers – well, at least powerful in 1990's terms – were available to analyze vast quantities of retail data. The talk of the town was that this pioneering technology would revolutionize the industry. Fifteen years later, many retailers have forgotten about the hype over data mining. Was this a missed opportunity or a wise use of scarce resources?

One of the first stories about data mining emerged from England. In 1991, a team of data gurus worked with a large British grocery chain to try to find unusual quirks in their data, quirks that could be used to create a strategic advantage.

After taking a deep dive into the data collected at check-outs across the country, the team found something very unexpected.

The story goes something like this.

It seemed that many people who purchased diapers on Friday evening also purchased beer. The team continued their research and discovered that often before leaving work on Friday evenings husbands would call their wives and ask if they needed anything. Invariably, the wife would ask her husband to pick up a package of diapers on his way home. The caring husband would stop off at the local grocer to purchase the diapers. And, while he was there, he might just as well grab a six-pack of beer –

no sense wasting an opportunity. Some even suggested the call to the wife was not really a call from a caring husband, but rather, more about finding an excuse to stop at the store.

Even today, it is not clear if this so-called 'beer and diapers' story ever took place. Many people believe this is yet another urban myth. Whether it was real or imaginary, the story created lots of interest in how organizations could or should use data mining. For a couple of years following the story there was much interest and a few organizations invested in the technology but then, almost on cue, the interest faded.

Fast forward a decade or so and lo and behold data mining resurfaces. This time it is not a UK chain but instead the world's largest retailer boasting about its use of data mining. One thing is very different this time - this is no urban myth. Wal-Mart's use of data mining is anything but imaginary and anything but hype.

With more than 100 million customers checking-out of Wal-Mart stores every week, they have collected an estimated 500 terabytes of data. To put that number into perspective, 500 terabytes is roughly the equivalent of 800,000 CD-ROMs of data.

So what do they do with the data? Simply put, they use it to create a competitive advantage. One thing we know for sure is Wal-Mart guards this data. To Wal-Mart this data is extremely valuable because they know or can

find out much more about consumers than any other organization.

Although Wal-Mart is very secretive about how they use this data, every once in a while they share one of their successes. Take for example Wal-Mart's discovery, through data mining, that U.S. consumers like to purchase Pop-Tarts just before a hurricane hits. Not just any Pop-Tarts though, American consumers prefer strawberry Pop-Tarts. Wal-Mart's data mining efforts show that their stores sell seven times more strawberry Pop-Tarts just before a hurricane hits than any other time.

Such knowledge allows Wal-Mart to ensure Pop-Tarts are in the right place and at the right time. Now THAT is a competitive advantage. It turns out that Wal-Mart's investment in data collection and analysis paid off.

There is an interesting lesson here. It seems that the over-hype of data mining in its early days drove away many retailers. And yet a decade or so later we discover that it is possible to create an advantage with data mining.

Why the change? In fact, this is very common with new technology.

Bill Gates, cofounder of Microsoft, wrote in his book *The Road Ahead*, "we always overestimate the change that will occur in the next two years and underestimate the change that will occur in the next ten."

Fast forward another couple of years and suddenly retailers find themselves in the midst of a national pet-food scare. It seems almost

every day another batch of dog or cat food is being taken off the shelves. But how can store owners or retailers possibly stay on top of the many brands and lot numbers affected? The answer, once again, has to do with managing data. Some retailers are relying on staff to ensure that the tainted food is not sold; other retailers are turning to technology to help.

Wal-Mart has centralized the control of the suspect food. As soon as Wal-Mart headquarters in Bentonville, Arkansas confirms the UPC Code or barcode for a bad batch, they immediately put a 'stop sale' message into their system. This message prevents a customer from accidentally purchasing a dangerous dose of dog food for Fido. The system does not rely on sales associates in 4000 stores 'remembering' which types of food are safe.

The moral of this story is that leaders can achieve a competitive advantage by having access to high-quality data that supports data-driven decisions.

Compare this to the recent well-known fiasco with an Enterprise Resource Planning (ERP) system in North Dakota. ERPs, so the consultants tell us, are supposed to integrate all of the data and processes of organization, presumably to help inform decision makers. So why is it that the decision makers are so frustrated? Why is it that the decision makers – who, by the way, allocate very scarce resources on the ERPs – are unable to access the data they need? Perhaps it is because some ERPs do NOT help decision makers, just the consultants. Sometimes an ERP is just an excuse to Expend Resources Poorly.

(Dr. John Girard is an Associate Professor of Business at Minot State University in North Dakota. He is the lead professor for knowledge management and is actively researching the relationships that exist between information anxiety, organizational memory loss, and contemporary knowledge management theories. He can be reached at John.Girard@minotstateu.edu.)