New Challenge: Separating Good and Bad Data

Just because data is accessible and plentiful does not mean it is all good. Inadvertently using bad data can cost a business in many ways, including poor decision making and missed insights.

The quantity and type of data available for capture is endless, but not all data is good. There is good data and bad data. Successful companies know how to collect and sort through the data, extract the data useful to business needs, convert it into insightful analytics offering, and finally turn insights into actionable opportunities and effective decision-making. Collecting the data is fairly easy today because of the software developed by innovative companies, but the rest of the process is more challenging. The massive amounts of data available is overwhelming for many business leaders, and the result is they avoid getting involved in what continues to be viewed as an "IT issue." The process for data-based decision-making begins with separating good and bad data and ends with eventually democratizing analytics so that people throughout the organization can question and/or use the data and analytics. The bottom line is that the final data analytics should make sense for the organization.

Understanding Good and Bad Data

Good data is data that is accurate and relevant to the business. Relevant means it can provide information business decision-makers need to do activities like identify new opportunities and trends, accurately measure internal performance, make decisions about future processes, and make critical problem solving decisions. Data is structured and unstructured. Structured data is data that has a high degree of organization, is relational, and stored in a database that is easily searchable. This type of data usually flows from corporate systems. Unstructured data does not fit into predefined databases. It is the type of data that is presenting the greatest challenge for businesses because it is voluminous and flows from a myriad of sources that include emails, webpages, social media, videos, podcasts, graphics, images, word processing documents, and so on.

With an estimated 90 percent of data now falling into the unstructured category, it is imperative for organizations to have a system for sorting through the data to determine what is good and relevant to the business. Good data produces good data analytics that help business leaders make the right decisions. Bad data produces data analytics that can lead decision-makers down the wrong path. Often organizations learn they are using bad data when decision-makers cannot seem to get desired results or make poor decisions despite reliance on what is believed to be good data.

Getting Scientific: Identify the Need or Problem

Data should be collected based on standards established by business leaders collaborating with technical professionals. It takes both sets of people to define and understand data relevant to the business. Good analytics uses valid scientific methods to address the identified problem or need (hypothesis in the scientific world). For example, a sales department wants to assess the impact of a promotion on consumers in different geographic areas. To answer the question of promotion effectiveness, the consumer groups compared need to be based on a random selection process. If the selection process is not random, and the people in the groups were only those already exposed

to the promotion because they were interested in the deal earlier, the data analytics are bad. How will the business know if the promotion can convince people to buy a product or service if they have already made a decision based on prior exposure?

Some data is obvious as to its importance to the company, like data concerning customer buying or browsing behaviors. Other data may not be collected if the business has not decided upfront the problem it wants to solve. Good data helps solves problems like identifying new markets, improving marketing effectiveness, identifying new industry trends, or expanding product lines. Different data is needed to make decisions in each of these situations. If the relevant data is not collected and used to produce good analytics, decisions are made using inadequate or irrelevant data. One of the most important questions to ask is: What measures are needed to make a good decision or to solve a problem?

Democratized and Self-Service Data Analytics

Collecting good data leads to good analytics, but the process is not completed. Democratizing analytics means embedding the measurements in operations so that people throughout the organization can review and utilize them. The more people looking at the numbers, the better for the organization. This means breaking down information silos and providing access to the measures across functional units or business operations and at all levels. It can also include engaging with others like customers and suppliers, depending on the problem needing a solution or the strategic goal executives have in mind. Giving people at all levels of the organization access to analytics empowers the workforce and increases the likelihood managers will get significant input concerning their relevancy, adequacy, and usefulness. It will also increase the likelihood that problems needing solving will be discovered.

To get the best results from data democratization, the business needs to create a culture that encourages the workforce to be data and analytics literate. However, it is important to develop the self-service analytics approach. Self-service analytics refers to analytics that can be produced by business users as needed and do not require a decision science or statistical background. The real-time access to analytics presented in an understandable format enable people to assess their reasonableness, spot business opportunities, and make valid decisions. A business can develop these analytics in-house or utilize one or more of the software firms that now specialize in assisting businesses with sorting through good and bad data, presenting good analytics throughout the organization, and providing real-time access to data.

Data is of critical importance to every organization today, but it has to be good data that decisions are based on. The rapid growth in data flows has been overwhelming, but it is now time to begin harnessing its power to inform business leaders. The time when data was managed to meet a single department need or was hoarded by senior leaders are gone. Data analytics must adapt to the new business models which break down hierarchies and enable networking. The challenge is making sure the data used is only good data.

DIVERSITY
EDUCATION
CENTER