

Data and Technology

Business Analytics Data

Structured and unstructured data is needed to generate analytics. As a beginning for organizing data into an understandable framework, statisticians usually categorize data into meaning groups.

Categorizing Data

There are many ways to categorize business analytics data. Data is commonly categorized by either internal or external sources. Typical examples of internal data sources include those presented in Table 3.4. When firms try to solve internal production or service operations problems, internally sourced data may be all that is needed. Typical external sources of data (see Table 3.5) are numerous and provide great diversity and unique challenges for BA to process. Data can be measured quantitatively (for example, sales dollars) or qualitatively by preference surveys (for example, products compared based on consumers preferring one product over another) or by the amount of consumer discussion (chatter) on the Web regarding the pluses and minuses of competing products.

Type of Internal Data	Description
Billing and Reminder Systems	Billing systems and reminder systems print bills and monitor customer payment information on valued-based customer segments.
Business	Industry codes, accounting information, personnel information, and so on are routinely collected in the course of business.
Customer	Names, addresses, returns, special contracts, segmentations, and so on are obtained when customers sign for or pay for products or services.
Customer Relationship Management Systems	<i>Customer relationship management</i> (CRM) systems collect and provide data on customer history, behavior on matters like complaints, the end of a relationship with a firm, and so on.
Human Resources	Information about employees, salaries, competencies, and so on is recorded by routine efforts over the history of employment.
Information from Enterprise Resource Planning Systems	<i>Enterprise resource planning</i> (ERP) systems are used to communicate internal business transactions to provide a direct feed of information on management issues and concerns, as well as other operations activities required to produce and sell products.
Product	Information is collected from procurement through post sales to monitor profitability, durability, and quality.
Production	Information that can be used to optimize production, inventory control, and supply chain delivery of the product to the customers is collected during the production processes.
Questionnaires	Information on customer behavior is obtained by customer questionnaires to measure customer service and product quality, among other things.
Web Logs	Information is collected on the firm's website usage via cookies and other means to learn customer navigation behavior and product interests.

Table 3.4 Typical Internal Sources of Data on Which Business Analytics Can Be Based

Type of External Data	Measured By
Customer Satisfaction	<ul style="list-style-type: none"> • Revenue, profit • Market share, sales • Product/service survey data • Loyalty • Brand awareness • Average spend per customer
Customer Demographics	<ul style="list-style-type: none"> • Geographic location (distance from market) • Income level • Market size
Competition	<ul style="list-style-type: none"> • Market share • Competitor profitability • Advertising/promotion efforts • Preference surveys • Web chatter on products
Economic	<ul style="list-style-type: none"> • Population statistics • Income distribution statistics

Table 3.5 Typical External Sources of Data on Which Business Analytics Can Be Based

A major portion of the external data sources are found in the literature. For example, the US Census and the International Monetary Fund (IMF) are useful data sources at the macroeconomic level for model building. Likewise, audience and survey data sources might include Nielsen (www.nielsen.com/us/en.html) for psychographic or demographic data, financial data from Equifax (www.equifax.com), Dun & Bradstreet (www.dnb.com), and so forth.

Data Issues

Regardless of the source of data, it has to be put into a structure that makes it usable by BA personnel. We will discuss data warehousing in the next section, but here we focus on a couple of data issues that are critical to the usability of any database or data file. Those issues are data quality and data privacy. Data quality can be defined as data that serves the purpose for which it is collected. It means different things for different applications, but there are some commonalities of high-quality data. These qualities usually include accurately representing reality, measuring what it is supposed to measure, being timeless, and having completeness. When data is of high quality, it helps ensure competitiveness, aids customer service, and improves profitability. When data is of poor quality, it can provide information that is contradictory, leading to misguided decision-making. For example, having missing data in files can prohibit some forms' statistical modeling, and incorrect coding of information can completely render databases useless. Data quality requires effort on the part

of data managers to cleanse data of erroneous information and repair or replace missing data.

Data privacy refers to the protection of shared data such that access is permitted only to those users for whom it is intended. It is a security issue that requires balancing the need to know with the risks of sharing too much. There are many risks in leaving unrestricted access to a company's database. For example, competitors can steal a firm's customers by accessing addresses. Data leaks on product quality failures can damage brand image, and customers can become distrustful of a firm that shares information given in confidence. To avoid these issues, a firm needs to abide by the current legislation regarding customer privacy and develop a program devoted to data privacy.

Collecting and retrieving data and computing analytics requires the use of computers and information technology. A large part of what BA personnel do is related to managing information systems to collect, process, store, and retrieve data from various sources.

Business Analytics Technology

Firms need an information technology (IT) infrastructure that supports personnel in the conduct of their daily business operations. The general requirements for such a system are stated in Table 3.6. These types of technology are elemental needs for business analytics

Type of Technology	Description
Computer Hardware	This is physical equipment used for input, processing, and output activities in an information system. Hardware can include computers of various sizes; various input, output, and storage devices; and telecommunications devices that link computers, including mobile handheld devices.
Computer Software	These are the preprogrammed instructions that control and coordinate the computer hardware components in the information system. They include system-wide software like ERP and smaller <i>apps</i> (computer software applications) for mobile devices.
Networking and Telecommunications Technology	Physical devices and software link the various pieces of hardware and transfer data from one physical location to another. They include the computers and communications equipment connected in networks for sharing voice, data, images, sound, and video. They also include the Internet, <i>intranets</i> (internal corporate networks based on Internet technology with limited access to employees within the firm), and <i>extranets</i> (private intranets extended to authorized users outside the organization).
Data Management Technology	Software governs the organization of data on physical storage media. It includes database management systems, data warehouses, data marts, and online analytical processing, as well as data, text, and Web mining technologies.

Table 3.6 General Information Technology (IT) Infrastructure

Operations Of particular importance for BA is the data management technologies listed in Table 3.6. Database management system (DBMS) is a data management technology software that permits firms to centralize data, manage it efficiently, and provide access to stored data by application programs. DBMS usually serves as an interface between application programs and the physical data files of structured data. DBMS makes the task of understanding where and how the data is actually stored more efficient. In addition, other DBMS systems can handle unstructured data. For example, object-oriented DBMS systems are able to store and retrieve unstructured data, like drawings, images, photographs, and voice data. These types of technology are necessary to handle the load of big data that most firms currently collect.

DBMS includes capabilities and tools for organizing, managing, and accessing data in databases. Four of the more important capabilities are its data definition language, data dictionary, database encyclopedia, and data manipulation language. DBMS has a data definition capability to specify the structure of content in a database. This is used to create database tables and characteristics used in fields to identify content. These tables and characteristics are critical success factors for search efforts as the database grows. These characteristics are documented in the data dictionary (an automated or manual file that stores the size, descriptions, format, and other properties needed to characterize data). The database encyclopedia is a table of contents listing a firm's current data inventory and the data files that can be built or purchased. The typical content of the database encyclopedia is presented in Table 3.7. Of particular importance for BA are the data manipulation language tools included in DMBS. These tools are used to search databases for specific information. An example is structure query language (SQL), which allows users to find specific data through a session of queries and responses in a database.

Types of Information	Description	Example
Associations	Occurrences linked to a single event.	An ad in a newspaper is associated with greater sales.
Classification	Recognizes patterns that describe the group an item belongs to by examining previous classified existing items and by inferring a set of rules that guide the classification process.	Identify customers who are likely to need more customer service than those who need less.
Clustering	Similar to classification when no groups have yet been defined, helps to discover different groupings within data.	Identify groups that can be differentiated within a single, large group of customers. An example would be distinguishing tea drinkers from those who would drink other beverages offered in flight on an airline.
Forecasting	Predicts values that can identify patterns in customer behavior.	Estimate the value of a future stream of dollar sales from a typical customer.
Sequence	Links events over time.	Identify a link between a person who buys a new house and subsequently will buy a new car within 90 days.

Table 3.8 Types of Information Obtainable with Data Mining Technology

Text mining is a software application used to extract key elements from unstructured data sets, discover patterns and relationships in the text materials, and summarize the information. Given that the majority of the information stored in businesses is in the form of unstructured data (e-mails, pictures, memos, transcripts, survey responses, business receipts, and so on), the need to explore and find useful information will require increased use of text mining tools in the future.

Web mining seeks to find patterns, trends, and insights into customer behavior from users of the Web. Marketers, for example, use BA services like Google Trends (www.google.com/trends/) and Google Insights for Search (<http://google.about.com/od/i/g/google-insights-for-search.htm>) to track the popularity of various words and phrases to learn what consumers are interested in and what they are buying.