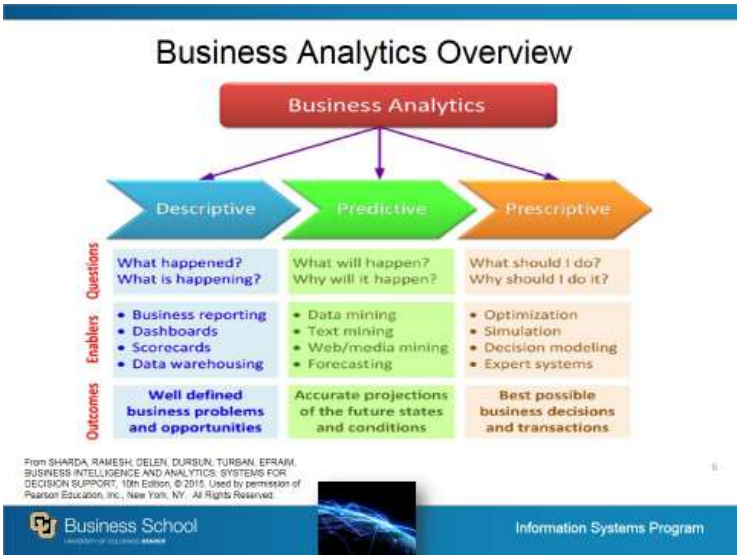
Data Warehousing for Business Intelligence
Course 4: Business Intelligence Concepts, Tools, and Applications

Module 4 Bonus Materials

Lesson 4: Business Analytics

We've arranged for students in this MOOC to purchase at a very low cost digital versions of chapters 1, 2, and 4 of the authoritative textbook *Business Intelligence and Analytics: Systems for Decision Support*, 10th edition, 2015 by Sharda, R., Delen, D., and Turban, E. See the optional text book link under course overview to purchase (US$4 for one chapter, US$10 for all three; the regular price for students is $15 per digital chapter).

- The word "analytics" has replaced the previous individual components of computerized decision support technologies that have been available under various labels in the past. Indeed, many practitioners and academics now use the word *analytics* in place of BI. The Institute for Operations Research and Management Science (INFORMS) has created a major initiative to organize and promote analytics.
- According to INFORMS, analytics represents the combination of computer technology, management science techniques, and statistics to solve real problems. These eight levels of analytics are described in more detail in a white paper by SAS (http://www.sas.com/news/sascom/analytics_levels.pdf).
- The majority of analytic offerings available today fall into one of the first four areas, which report historical data on what happened in the past but no insight about the future. For simple business problems, these analytic solutions will be all you need. But if you're asking more complex questions or looking for predictive insight, you need to look at the second half of the spectrum. Even better, if you can learn to use these technologies together and identify what type of analytics to use for every individual situation, you'll really be increasing your chances for true business intelligence.



Tools used in descriptive analytics include data warehouses and visualization applications. A significant technology that has become a key player in this area is visualization. Using the latest visualization tools in the marketplace, we can now develop powerful insights into the operations of our organization.

- Standard reports: Answer the questions: What happened? When did it happen? Example: Monthly or quarterly financial reports. We all know about these. They're generated on a regular basis and describe just "what happened" in a particular area. They're useful to some extent, but not for making long-term decisions.
- Ad hoc reports: Answer the questions: How many? How often? Where? Example: Custom reports that describe the number of hospital patients for every diagnosis code for each day of the week. At their best, ad hoc reports let you ask the questions and request a couple of custom reports to find the answers.
- Query drilldown (or OLAP): Answers the questions: Where exactly is the problem? How do I find the answers? Example: Sort and explore data about different types of cell phone users and their calling behaviors. Query drilldown allows for a little bit of discovery. OLAP lets you manipulate the data yourself to find out how many, what color and where.

Predictive modeling

Answers the questions: What will happen next? How will it affect my business? Example: Hotels and casinos can predict which VIP customers will be more interested in particular vacation packages. If you have 10 million customers and want to do a marketing campaign, who's most likely to respond? How do you segment that group? And how do you determine who's most likely to leave your organization? Predictive modeling provides the answers.

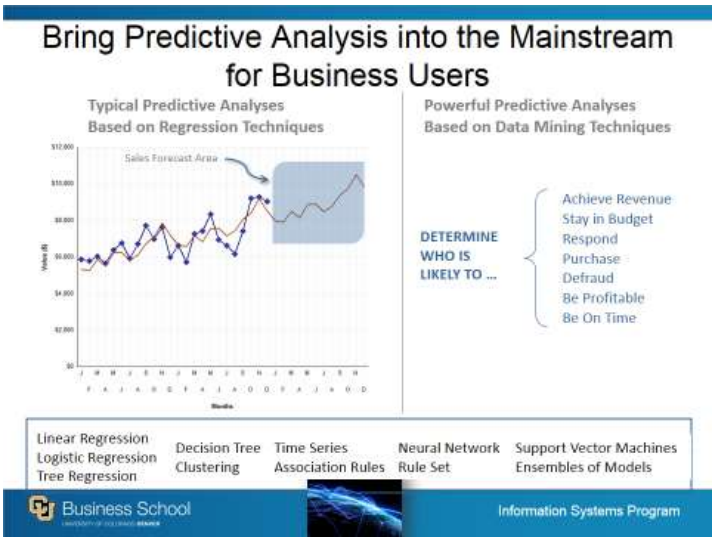
A number of techniques are used in developing predictive. For example, we can use classification techniques such as decision tree models and neural networks to predict how well a motion picture will do at the box office using analytical applications, including various classification algorithms. We can also use clustering algorithms for segmenting customers into different clusters to be able to target specific promotions to them. Finally, we can use association mining techniques to estimate relationships between different purchasing behaviors. That is, if a customer buys one product, what else is the customer likely to purchase? Such analysis can assist a retailer in recommending or promoting related products. For example, any product search on Amazon.com results in the retailer also suggesting other similar products that may interest a customer.

- Alerts: Answer the questions: When should I react? What actions are needed now? Example: Sales executives receive alerts when sales targets are falling behind. With alerts, you can learn when you have a problem and be notified when something similar happens again in the future. Alerts can appear via e-mail, RSS feeds or as red dials on a scorecard or dashboard.
- Statistical analysis: Answers the questions: Why is this happening? What opportunities am I missing? Example: Banks can discover why an increasing number of customers are refinancing their homes.
- Here we can begin to run some complex analytics, like frequency models and regression analysis. We can begin to look at why things are happening using the stored data and then begin to answer questions based on the data.
- Forecasting: Answers the questions: What if these trends continue? How much is needed? When will it be needed? Example: Retailers can predict how demand for individual products will vary from store to store. Forecasting is one of the hottest markets – and hottest analytical applications – right now. It applies everywhere. In particular, forecasting demand helps supply just enough inventory, so you don't run out or have too much.

Analytics is the sophisticated analysis and use of business data to enable fact-based decision making at every level of your organization. Analytics can drive superior performance – from customer management to

the supply chain to product/service development to strategic planning. For top companies such as Marriott International, Harrah's Entertainment, Capital One, Amazon.com, Netflix and the Boston Red Sox, analytics creates a competitive advantage. Quick examples:

- Amazon uses market basket analysis
- Mars uses linear regression to predict candy sales, particularly during peak holidays like Halloween



- Prescriptive analytics is a set of techniques that use descriptive data and forecasts to identify the decisions most likely to result in the best performance. Usually, an organization uses prescriptive analytics to identify the decisions or actions that will optimize the performance of a system.
- Optimization: Answers the question: How do we do things better? What is the best decision for a complex problem? Example: Given business priorities, resource constraints and available technology, determine the best way to optimize your IT platform to satisfy the needs of every user. Optimization supports innovation. It takes your resources and needs into consideration and helps you find the best possible way to accomplish your goals.

> Organizations have used prescriptive analytics to set prices, create production plans, and identify the best locations for facilities such as bank branches. The goal here is to provide a decision or a recommendation for a specific action. These recommendations can be in the forms of a specific yes/no decision for a problem, a specific amount (say, price for a specific item or airfare to charge), or a complete set of production plans. The decisions may be presented to a decision maker in a report or may directly be used in an automated decision rules system (e.g., in airline pricing systems). Thus, these types of analytics can also be termed decision or normative analytics.

Today Big Data refers to almost any kind of large data that has the characteristics of volume, velocity, and variety. Examples include data about Web searches, such as the billions of Web pages searched by Google; data about financial trading, which operates in the order of microseconds; and data about consumer opinions measured from postings in social media. One computer, even a powerful one, could not handle the scale of Big Data. The solution is to push computation to the data, using the MapReduce programming paradigm. This was a new paradigm, and it gave rise to a whole new way of processing data. This is what we know today as the MapReduce programming paradigm, which made processing Big Data a reality. MapReduce was originally developed at Google, and a subsequent version was released by the Apache project called Hadoop MapReduce. Today, when we talk about storing, processing, or analyzing Big Data, HDFS and MapReduce are involved at some level. Other relevant standards and software solutions have been proposed. Although the major toolkit is

available as open source, several companies have been launched to provide training or specialized analytical hardware or software services in this space. Some examples are HortonWorks, Cloudera, and Teradata Aster.

> Over the past few years, what was called Big Data changed more and more as Big Data applications appeared. The need to process data coming in at a rapid rate added velocity to the equation. One example of fast data processing is algorithmic trading. It is the use of electronic platforms based on algorithms for trading shares on the financial market, which operates in the order of microseconds. The need to process different kinds of data added variety to the equation. Another example of the wide variety of data is sentiment analysis, which uses various forms of data from social media platforms and customer responses to gauge sentiments. Today Big Data is associated with almost any kind of large data that has the characteristics of volume, velocity, and variety.

Business and Visual analytics Platforms

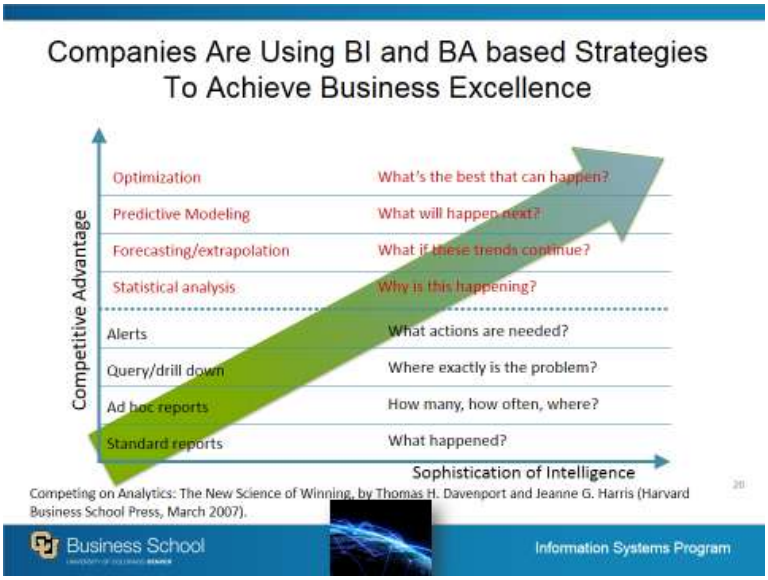
These systems help to empower larger numbers of users, solve complex problems more quickly, and improve collaboration and information sharing. By enabling end-users, IT staff are freed up. In addition, these tools allow for growth at a self-determined pace. Right now, most of the activity in the business intelligence and analytics platform market is from organizations that are trying to mature their visualization capabilities and to move from descriptive to diagnostic (i.e., predictive and prescriptive) analytics. The vendors in the market have overwhelmingly concentrated on meeting this user demand. If there were a single theme in 2012, it would be that data discovery/visualization became a mainstream architecture.
For years, data discovery/visualization vendors—such as QlikTech, Salient Management Company, Tableau Software, and Tibco Spotfire—received more positive feedback than vendors offering OLAP cube and semantic-layer-based architectures.

In 2012, the market responded:

- MicroStrategy significantly improved Visual Insight.
- SAP launched Visual Intelligence.
- SAS launched Visual Analytics.
- Microsoft bolstered PowerPivot with Power View.
- IBM launched Cognos Insight.
- Oracle acquired Endeca.
- Actuate acquired Quiterian.

This emphasis on data discovery/visualization from most of the leaders in the market — which are now promoting tools with business-user-friendly data integration, coupled with embedded storage and computing layers (typically in-memory/columnar) and unfettered drilling—accelerates the trend toward decentralization and user empowerment of BI and analytics, and greatly enables organizations' ability to perform diagnostic analytics. SAS Visual Analytics is an example of such an environment. These systems help to empower larger numbers of users, solve complex problems more quickly, and improve collaboration and information sharing.

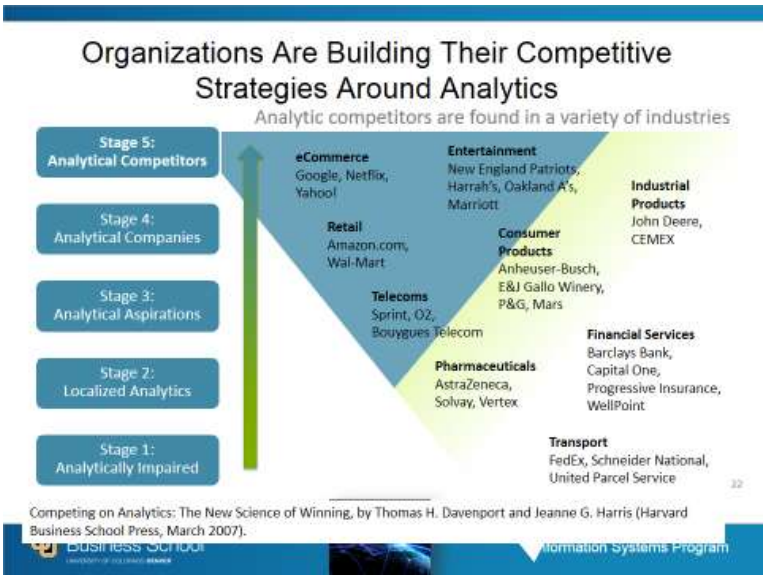
Below are the type of analytical questions the BI system should be able to answer



The competitive landscape is changing, and for many firms, BI has evolved from being a "nice to have," to being a requirement for competing in the marketplace. Such firms are *BI-based organizations* because of the role that BI plays in their operations and overall business success.

One view is that data analyst is just another term for professionals who were doing business intelligence in the form of data compilation, cleaning, reporting, and perhaps some visualization. Their skill sets included Excel, some SQL knowledge, and reporting.

- In contrast, a data scientist is responsible for predictive analysis, statistical analysis, and more advanced analytical tools and algorithms. They may have a deeper knowledge of algorithms and may recognize them under various labels —data mining, knowledge discovery, machine learning, and so forth. Some of these professionals may also need deeper programming knowledge to be able to write code for data cleaning and analysis in current Web-oriented language such as Java and Python. Again, you should recognize by now that these as falling under the predictive and prescriptive analytics umbrella.
- The distinction between analytics and data science is more of a degree of technical knowledge and skills sets than the functions. It may also be more of a distinction across disciplines. Computer science, statistics, and applied mathematics programs appear to prefer the data science label, reserving the analytics label for more business-oriented professionals. The graduates of this certificate program tend to be responsible for tasks more in line with data science professionals (as defined by some circles) than just reporting analytics.



- New England Patriots – New England Patriots were able to win 3 super bowls in 4 years. In-depth analytics help the team select its players and stay below the salary cap (recently the team was ranked 24th highest payroll in the NFL). Team selects players rating potential draft choices on non-traditional factors such as intelligence, willingness to subsume personal ego for the benefit of the team. The team also has an extensive use of analytics for on-the-field decisions. They employ statistics to decide whether to punt or "go for it" on a 4th down, whether to try for 1 point or 2 after a touchdown, and whether to challenge a referee's ruling. Off the field, the team sues detailed analytics to assess and improve the "total fan experience". At every home game 20-25 people have specific assignments to make quantitative measurements of the stadium food, parking, personnel, bathroom cleanliness, and other factors.
- Capital One - 1980's, two financial services consultants identified a major problem in the credit card industry, as well as a potential solution. They proposed an "information based market strategy" to banks, and Signet Bank (in Virginia) hired them to work in it's bank card division. Over two years they ran thousands of analytical tests on the customer database and discovered the most profitable customers where people who borrowed large amounts quickly and then paid off the balances slowly. At the time the credit card industry treated such customers just as they treated people who made small purchases and paid their balances off in full every month. Recognizing an opportunity, the team created the industry's first balance-transfer card. It targeted debtors as valued customers, and quickly took off in the industry. The success with analytics continued and led Signet to spin off the bank card division as a company called Capital One. Through analysis of CD Interest rates, rollover incentives, minimum balances they determined the true impact of these metrics on retention rates and new money coming into the bank. Through such analyses, the savings business increased retention by 87% and lowered the cost of acquiring a new account by 83%.

See more information in the web sites below:

- The differences between data analysts and data scientists
- Intelligent People but Stupid Choices – Try Using Analytics
- 7 Traits a Big Data Scientist Shouldn't Have
- Magic Quadrant for Business Analytics Services, Worldwide
- Democratizing Big Data with Interactive Visual Analytics