BA4206 BUSINESS ANALYTICS

Learn to

1. Use business analytics for decision making
2. To apply the appropriate analytics and generate solutions
3. Model and analyse the business situation using analytics.

UNIT I INTRODUCTION TO BUSINESS ANALYTICS (BA) 9

Business Analytics - Terminologies, Process, Importance, Relationship with Organisational DecisionMaking, BA for Competitive Advantage.

UNIT II MANAGING RESOURCES FOR BUSINESS ANALYTICS 9

Managing BA Personnel, Data and Technology. Organisational Structures aligning BA. ManagingInformation policy, data quality and change in BA.

UNIT III DESCRIPTIVE ANALYTICS 9

Introduction to Descriptive analytics - Visualising and Exploring Data - Descriptive Statistics - Samplingand Estimation - Probability Distribution for Descriptive Analytics - Analysis of Descriptive analytics

UNIT IV PREDICTIVE ANALYTICS 9

Introduction to Predictive analytics - Logic and Data Driven Models - Predictive Analysis Modeling andprocedure - Data Mining for Predictive analytics. Analysis of Predictive analytics

UNIT V PRESCRITIVE ANALYTICS 9

Introduction to Prescriptive analytics - Prescriptive Modeling - Non Linear Optimisation -Demonstrating Business Performance Improvement.

TOTAL: 45 PERIODS

UNIT I INTRODICTION TO BUSINESS ANALYTICS

Introduction

The word analytics has come into the foreground in last decade or so. The proliferation of the internet and information technology has made analytics very relevant in the current age. Analytics is a field which combines data, information technology, statistical analysis, quantitative methods and computer-based models into one. This all are combined to provide decision makers all the possible scenarios to make a well thought and researched decision. The computer-based model ensures that decision makers are able to see performance of decision under various scenarios.

Application

Business analytics has a wide range of application from customer relationship management, financial management, and marketing, supply-chain management, human-resource management, pricing and even in sports through team game strategies.

IMPORTANCE OF BUSINESS ANALYTICS

- **Business analytics is a methodology or tool to make a sound commercial decision**. Hence it impacts functioning of the whole organization. Therefore, business analytics can help improve profitability of the business, increase market share and revenue and provide better return to a shareholder.
- Facilitates better understanding of available primary and secondary data, which again affect operational efficiency of several departments. Provides a competitive advantage to companies. In this digital age flow of information is almost equal to all the players. It is how this information is utilized makes the company competitive. Business analytics combines available data with various well thought models to improve business decisions.

- Converts available data into valuable information. This information can be presented inany required format, comfortable to the decision maker.

EVOLUTION OF BUSINESS ANALYTICS

Business analytics has been existence since very long time and has evolved with availability of newer and better technologies. It has its roots in operations research, which was extensively used during World War II. Operations research was an analytical way to look at data to conduct military operations. Over a period of time, this technique started getting utilized for business. Here operation's research evolved into management science. Again, basis for management science remained same as operation research in data, decision making models, etc.

As the economies started developing and companies became more and more competitive, management science evolved into business intelligence, decision support systems and into PC software.

Data for Analytics

Business analytics uses data from three sources for construction of the business model. It uses business data such as annual reports, financial ratios, marketing research, etc. It uses the databasewhich contains various computer files and information coming from data analysis.

Challenges

Business analytics can be possible only on large volume of data. It is sometime difficult obtain large volume of data and not question its integrity.

TYPES OF BUSINESS ANALYTICS

There are four main types of business analytics companies can employ to better understand and grow their business.

- **Descriptive analytics—**is one of the most basic forms of analytics, providing insights on what has happened or is currently happening. Sales reports and social media engagement are examples of descriptive analytics.

employees on service, their age, gender, everything. This can help in the selection of applicants while interviewing.

WHY ANALYTICS IS IMPORTANT

Business analytics can help companies make better, more informed decisions and achieve avariety of goals. By leveraging data, businesses can:

- Better understand consumer behavior
- Gain insight into their competitors
- Identify market trends
- Measure accomplishments against goals
- Optimize operations

Business analytics has helped many companies navigate tough times, especially regarding the COVID-19 pandemic. According to a survey by business intelligence company Sisense, 50% of companies reported using analytics "more often" or "much more often," during the pandemic. The increased use of analytics was even more pronounced among smaller companies. Helping organizations navigate through crises is just one of the many reasons why analytics is important to business. Data suggest companies that use analytics to their advantage are twice as likely to rank in the top quarter for financial performance, five times more likely to make timely decisions and three times more likely to execute their decisions and plans.

FUTURE OF BUSINESS ANALYTICS

Big data and analytics will shape the business of the future in remarkable ways. Global spending on business analytics was projected to reach $215.7 billion in 2021, a 10% increase from 2020 according to the IDC. Furthermore, that number is expected to grow even more over the next five years as the world economy recovers from the COVID-19 pandemic.

Organizations and their leaders will depend on business analytics to cut operating costs, improve asset use, and increase the agility and reliability of their operations. Analytics will be key to delivering real-time performance insight, real-time logistics management and production quality analysis.

The emergence of AI and machine learning technology is key to the future of business analytics. This technology will allow companies to interpret and draw insights from massive datasets to gain a better understanding of their customers and overall operations. Additionally, advanced data visualization tools will make it easier for businesses to identify and communicate these insights, and allow everyone throughout the organization to understand them.

Big data used to merely be a buzzword, with futuristic-sounding connotations, but the future of business analytics and big data has arrived. Businesses that want to compete in the 21st century need to continue to embrace business analytics for their operations. They need to also equip their teams with individuals who have the requisite skills to help translate data into actionable intelligence.

MAIN COMPONENTS

The main components of a typical business analytics dashboard include:

- **Data Aggregation**: prior to analysis, data must first be gathered, organized, and filtered, either through volunteered data or transactional records
- **Data Mining**: data mining for business analytics sorts through large datasets using databases, statistics, and machine learning to identify trends and establish relationships
- **Association and Sequence Identification**: the identification of predictable actions that are performed in association with other actions or sequentially
- **Text Mining**: explores and organizes large, unstructured text datasets for the purpose of qualitative and quantitative analysis
- **Forecasting**: analyzes historical data from a specific period in order to make informed estimates that are predictive in determining future events or behaviors
- **Predictive Analytics**: predictive business analytics uses a variety of statistical techniquesto create predictive models, which extract information from datasets, identify patterns, and provide a predictive score for an array of organizational outcomes
- **Optimization**: once trends have been identified and predictions have been made, businesses can engage simulation techniques to test out best-case scenarios
- **Data Visualization**: provides visual representations such as charts and graphs for easy and quick data analysis

- The essentials of business analytics are typically categorized as either descriptive analytics, which analyzes historical data to determine how a unit may respond to a set of variables; predictive analytics, which looks at historical data to determine the likelihood of particular futureoutcomes; or prescriptive analytics, the combination of the descriptive analytics process, which provides insight on what happened, and predictive analytics process, which provides insight on what might happen, providing a process by which users can anticipate what will happen, when it will happen, and why it will happen.

- Some business analytics examples include the operation and management of clinical information systems in the healthcare industry, the tracking of player spending and development of retention efforts in casinos, and the streamlining of fast food restaurants by monitoring peak customer hours and identifying when certain food items should be prepared based on assembly time.
- Modern, high quality business analytics software solutions and platforms are developed to ingestand process the enormous datasets that businesses encounter and can exploit for optimal businessoperations.
- **PREPARE FOR A CAREER IN BUSINESS ANALYTICS**
- Companies across virtually every industry are increasingly recognizing the importance of business analytics. As a result, analytics experts are in high demand for their ability to extract meaningful insights from data that inform strategic business decisions.
- The Villanova University online Master of Science in Analytics program offers a comprehensivecurriculum that addresses the entire continuum of business analytics, from data collection and analysis to implementation.
- Learn more about how the program can help you develop the skills to become a business analytics leader.

FUTURE FOR BUSINESS ANALYTICS PROFESSIONALS

The future for business analytics is very bright, so professionals looking to build a career in this domain are in for the ride of a lifetime. This also means that the demand for such professionals will also be on the rise, so to get an edge, you need to have some aces up your sleeves. One way to hone your skills and give yourself the edge is by pursuing online analytics courses. Online analytics courses from credible platforms can equip you the latest tools and technologies that other applicants might not know about. One such online program you can pursue is the business analytic online course from IIM Kozhikode. Offered on the Talent edge platform, the business analytic online course from IIM Kozhikode helps you inherit the skill-set and training you need to succeed in this domain.

As a new-age business, operating in a world full of social media updates and news, there is so much information that comes to you, about your consumers or even potential consumers. All data elements that reach you, whether through formal sources or informal ones, need to be converted into information that you can use, for decision making. Data is forming the biggest part of an organization's business strategy and future goals right now.

When the data is converted into information, it takes the form of intelligence or analytics. Courses for business analytics talk about these concepts in detail.

ROLES AND RESPONSIBILITIES IN BUSINESS ANALYTICS

Business analytics professionals' main responsibility is to collect and analyze data to influence strategic decisions that a business makes. Some initiatives they might provide analysis for include the following:

- Identifying strategic opportunities from data patterns;
- Identifying potential problems facing the business and solutions;
- Creating a budget and business forecast;
- Monitoring progress with business initiatives;
- Reporting progress on business objectives back to stakeholders;
- Understanding KPIs; and
- Understanding regulatory and reporting requirements.

Business analysts must have a mixture of hard and soft skills. A business analyst does not need a deep understanding of IT but does need to understand how systems work together. Some business analysts choose to move from an IT-centric role into a BA role.

When recruiting for these jobs, employers typically look for the following capabilities:

- Ability to perform cost-benefit analysis;
- Familiarity with process modeling;
- Understanding stakeholder analysis;
- Analytical problem-solving;
- Oral and written communications skills;
- A basic understanding of IT systems, including databases;
- Detail-oriented;
- Experience with BA tools and software; and
- Ability to create visual representations of data.

Here we will take a brief look at what differences exist between these.

Business Analytics is usually carried out using statistical analysis, to be able to determine why certain trends are showing up or specific things are happening. The analysis is oriented towards finding answers to certain concerns, based on past or present data. Business Intelligence, on the other hand, is related to using the data to understand how it can be used and what decisions can itenable the company to make.

Analytics does help in predicting trends for the future. A lot of the Business Analytics is mainly about predictive modelling and sharing insights on what you might face, and therefore preparation for the same. Intelligence, on the other hand, gives a great depth of information for the data at hand. So business intelligence is what sets the foundation for the decision making that an organization undertakes. It shifts the focus from intuitive based decisions or judgement based approaches, to actual fact-based decisions.

Business Analytics is mainly related to making changes to the business or its direction. Business intelligence, on the other hand, is focused on sustaining the business. Therefore, if as an organization, you plan to switch gears and move into another direction, the analytics are what will help you to make crucial decisions and validate your approach. If you are planning to build stability in your existing venture, intelligence is what will help you in choosing how to define steps to do that.

Both these terms have often been used interchangeably. While there are some fundamental similarities between them, there are several differences too. The orientation for Business Intelligence has always been towards reporting the actual analysis that has been undertaken. It is the logical next step after Business Analytics since it shares implications that analyses have uncovered. So knowledge of both these areas is equally important for all business professionals, irrespective of their functional areas.

EXAMPLES OF BUSINESS ANALYTICS

Business analytics has applications in a wide array of different businesses. Some companies are developing innovative ways to use big data in order to improve their customer's experience and maximize profits. Here is a real life example of business analytics:

Fast-food companies have begun to implement business analytics to streamline their restaurants. Who wants to have a slow experience in a fast-food drive-thru? By monitoring how busy the drive-thru is these businesses can increase efficiency during peak hours. When the line gets long, the digital order boards change. They begin to highlight items that can be prepared quickly. This leads to more simple orders that can be completed quickly. When the lines are short, slower items with higher margins are featured. In this way, the store can respond to real-time needs to improve efficiency.

Other types of business analytics applications do more than just respond to the current situation. These techniques help businesses predict which customers are less likely to return. They can thentarget advertising and promotions to these customers to improve retention. Here are some examples of predictive analytics in business:

Casinos use business analytics to improve their profits and keep customers coming back. Though the house wins most of the time, players typically need to win enough to enjoy themselves and keep playing. Otherwise, players may lose interest and stop coming back. By tracking players spending, casinos can learn which customers they make the most money from. They can offer greater incentives to these big spenders to keep them coming back. The collected data also helps these resorts understand which amenities are most popular.

BUSINESS ANALYTICS TOOLS

There are data analytics tools that can be used in business analytics to streamline the big data pipeline.

Tools for use in business analytics range substantially in complexity. Self-service analytics tools provide a simplified interface, often are paid services that can do basic data analytics tasks in a user-friendly way. Alternatively, advanced statistical analysis tools require programming and software engineering skills to use effectively. Many of these tools are open-source and available for free to users.

The most well-known tools in both data analytics and business analytics are open source programming languages that provide statistical tools. Two widely used options are R and Python (with the pandas library). Any data processing or analysis task can be automated using these languages. R and Python both have large communities that provide support and many packages and libraries which provide added functionality and statistical methods. These include data visualization tools, advanced statistical algorithms, data scraping tools and much more.

There are also paid statistical programming languages. These include SAS, SPSS and MATLAB. These languages have the advantage of paid support and professional development. However, they are typically used less often than open source solutions.

Not all statistical analysis tools require programming. There are many options for statistical analysis with a graphical user interface (GUI). These tools are generally paid and include Tableau, Qlik, Sisense and SAP. These are self-service analytics tools that can take raw data and turn it into user-friendly charts with the click of a button. This user-friendly workflow allows the most useful insights to quickly be visualized.

Selecting the right tool involves balancing financial costs, time costs, the complexity of the data and the ease of use.

BENEFITS OF BUSINESS ANALYTICS

Business analytics can help provide a wide array of benefits

- Enable data-driven decision making that has the potential to increase profits and improveefficiency
- With predictive analytics, allow businesses to plan for the future in ways that werepreviously impossible
- Helps a company make informed business decisions
- By modeling the outcomes and understanding the past, guesswork is minimized
- Present meaningful, clear data to support decision making and convince stakeholder

Business analytics provides a way for businesses to plan for the future. By modeling the trends ina businesses' sales, profits and other key metrics, these indicators can be projected into the future. Understanding the changes that are likely to occur seasonally, annually or on any scale allow businesses to better prepare. This may mean decreasing spending in preparation for a slow season or investing in new marketing campaigns to compensate. Large suppliers can use this data to predict order volume and minimize waste in their warehouses. Planning for future events provides a huge advantage to all businesses.

Business analytics can also enable new types of marketing campaigns. The data collected by businesses give insights into customer behavior which helps businesses understand the effectiveness of advertising campaigns with different audiences. Targeting audiences that are more likely to respond to specific campaigns or products increases efficiency overall. In addition, understanding consumer habits can help businesses improve customer retention. By identifying customers who are less likely to return, businesses can offer targeted promotions. This provides a cost-effective way to gain customer loyalty.

The applications of business analytics are wide-ranging and the benefits of business analytics areclear – data-driven businesses have a competitive edge in almost every industry.

COMMON CHALLENGES OF BUSINESS ANALYTICS

Businesses might encounter both business analytics and business intelligence challenges when trying to implement a business analytics strategy:

- **Too many data sources.** There is an increasingly large spectrum of internet-connected devices generating business data. In many cases, they are generating different types of data that must be integrated into an analytics strategy. However, the more complex a data set becomes, the harder it is to use it as part of an analytics framework.
- **Lack of skills.** The demand for employees with the data analytic skills necessary to process BA data has grown. Some businesses, particularly small and medium-sized businesses (SMBs), may have a hard time hiring people with the BA expertise and skills they need.
- **Data storage limitations.** Before a business can begin to decide how it will process data, it must decide where to store it. For instance, a data lake can be used to capture large volumesof unstructured data.

BUSINESS ANALYTICS EXAMPLES AND TOOLS

There are several BA and BI tools that can automate advanced data analytics functions and require few of the specialized skills or deep knowledge of the programming languages used in data science.

These tools help businesses organize and make use of the massive amounts of data that modern internet of things and enterprise cloud applications generate. These applications may be part of supply chain management, enterprise resource planning and customer relationship management applications.

Below are some business analytics tools on the market:

- *Dundas* BI, with automated trend forecasting and a user-friendly interface;
- *Knime Analytics Platform*, which has high-performance data pipelining and machinelearning;
- *Qlik's* QlikView with data visualization and automated data association features;
- *Sisense*, known for its dynamic text-analysis features and data warehousing;
- *Splunk*, which has intuitive user interface and data visualization features;
- *Tableau,* which has advanced unstructured text analysis and natural languageprocessing capabilities; and
- *Tibco Spotfire,* which offers powerful, automated statistical and unstructured text analysis.

BA tools are used in many ways. For example, they can identify customers who are likely to cancel a service offering subscription. A company would first use aggregate data from enterprise applications, using a Data Ops analytics platform like Data Kitchen. Then it would use a BA tool to present that data to employees. The BA tool would help employees identify customers at risk of canceling and let them take steps to keep those customers.

When choosing a business analytics tool, organizations should consider the following:

- the sources which their data comes from;

- the type of the data to be analyzed; and
- the usability of the tool.

A good business analytics tool is intuitive and user-friendly. It also provides a full suite of features for more advanced analytics.

Career and salary trends in business analytics

There are several career paths for a person with a BA background. Some common job titles and annual salaries as of 2021, according to PayScale, include the following:

- senior business analyst -- $86,050
- business systems analyst -- $70,155
- business analyst -- $69,785
- business intelligence analyst -- $69,639
- junior business analyst -- $51,009

TERMINOLOGIES

Analytics Terminology

Analytics, Business analytics, Predictive modeling, Advanced analytics, Big Data Analytics, Data Mining, Knowledge Discovery, Artificial Intelligence, Machine learning, Business Intelligence, OLAP, Reporting, Data warehousing, Statistics

There are many terms that get thrown around in the field of analytics. This article is an attempt to list the subtle differences or similarities between the common terms.

Analytics – Analytics can simply be defined as the process of breaking a problem into simpler parts and using inferences based on data to drive decisions. Analytics is not a tool or a technology; rather it is a way of thinking and acting.

Analytics has widespread applications in spheres as diverse as science, astronomy, genetics, financial services, telecom, retail, marketing, sports, gaming and health care.

Business analytics – This term refers to the application of analytics specifically in the sphere of business. It includes subsets like –

- Marketing analytics
- Risk analytics
- Fraud analytics
- CRM analytics
- Loyalty analytics
- Operations analytics
- HR analytics

Industries which rely extensively on analytics include –

- Financial Services (Banks, Credit Cards, Loans, Insurance etc.)
- Retail
- Telecom
- Health care
- Consumer goods
- Manufacturing
- Sports
- Hotels
- Airlines
- Any industry where large amounts of data is generated

Predictive Analytics – Predictive analytics is one of the most popular analytics terms. Predictive analytics is used to make predictions on the likelihood of occurrence of an event or determine some future patterns based on data. Remember it does not tell whether an event will happen. It only assigns probabilities to the future events or patterns.

Google Trends analysis of "Predictive Analytics"

The term emphasizes the predictive nature of analytics (as opposed to, say the retrospective nature of tools like OLAP). This is one of those terms that is designed by sales people and marketers to add glamour to any business. "Predictive analytics" sounds fancier than just plain "analytics". In practice, predictive analytics is rarely used in isolation from descriptive analytics.

Descriptive analytics – Descriptive analytics refers to a set of techniques used to describe or explore or profile any kind of data. Any kind of reporting usually involves descriptive analytics. Data exploration and data preparation are essential ingredients for predictive modeling and these rely heavily on descriptive analytics.

Inquisitive analytics – Whereas descriptive analytics is used for data presentation and exploration, inquisitive analytics answers terms why, what, how and what if. Ex: Why have the sales in the Q4 dropped could be a question based on which inquisitive analysis can be performed on the data

Advanced analytics – Like "Predictive analytics", "Advanced analytics" too is a marketing driven terminology. "Advanced" adds a little more punch, a little more glamour to "Analytics" and is preferred by marketers.

Big data analytics – When analytics is performed on large data sets with huge volume, variety and velocity of data it can be termed as big data analytics. The annual amount of data we have is expected to grow from 8 zettabytes (trillion gigabytes) in 2015 to 35 zettabytes in 2020.

Growing data sizes would inevitably require advanced technology like Hadoop and Map Reduce to store and map large chunks of data. Also, large variety of data (structured, unstructured) is flowing in at a very rapid pace. This would not only require advance technology but also advanced analytical platforms. So to summarize, large amounts of data together with the technology and the analytics platforms to get insights out of such a data can be called as the Big data analytics.

Data Mining – Data mining is the term that is most interchangeably used with "Analytics". Data Mining is an older term that was more popular in the nineties and the early 2000s. However, data mining began to be confused with OLAP and that led to a drive to use more descriptive terms like "Predictive analytics".

According to Google trends, "Analytics" overtook "Data mining" in popularity at some point in 2005 and is about 5 times more popular now. Incidentally, Coimbatore is one of the only cities inthe world where "Data mining" is still more popular than "Analytics".

Data Science – Data science and data analytics are mostly used interchangeably. However, sometimes a data scientist is expected to possess higher mathematical and statistical sophistication than a data analyst. A Data scientist is expected to be well versed in linear algebra, calculus, machine learning and should be able to navigate the nitty-gritty details of mathematics and statistics with much ease.

Artificial Intelligence –During the early stages of computing, there were a lot of comparisons between computing and human learning process and this is reflected in the terminology. The term "Artificial intelligence" was popular in the very early stages of computing and analytics (in the 70s and 80s) but is now almost obsolete.

Machine learning – involves using statistical methods to create algorithms. It replaces explicit programming which can become cumbersome due to the large amounts of data, inflexible to adapt to the solution requirements and also sometimes illegible.

It is mostly concerned with the algorithms which can be a black box to interpret but good models can give highly accurate results compared to conventional statistical methods. Also, visualization, domain knowledge etc. are not inclusive when we speak about machine learning. Neural networks, support vector machines etc. are the terms which are generally associated with the machine learning algorithms

Algorithm – Usually refers to a mathematical formula which is output from the tools. The formula summarizes the model Ex: Amazon recommendation algorithm gives a formula that can recommend the next best buy.

OLAP – ***Online analytical processing*** refers to descriptive analytic techniques of slicing and dicing the data to understand it better and discover patterns and insights. The term is derived from another term "OLTP" – online transaction processing which comes from the data warehousing world.

Reporting – The term "Reporting" is perhaps the most unglamorous of all terms in the world of analytics. Yet it is also one of the most widely used practices within the field. All businesses use reporting to aid decision making. While it is not "Advanced analytics" or even "Predictive analytics", effective reporting requires a lot of skill and a good understanding of the data as well as the domain.

Data warehousing – Ok, this may actually be considered more unglamorous than even "Reporting". Data warehousing is the process of managing a database and involves extraction, transformation and loading (ETL) of data. Data warehousing precedes analytics. The data managed in a data warehouse is usually taken out and used for business analytics.

Statistics – Statistics is the study of the collection, organization, and interpretation of data. Data mining does not replace traditional statistical techniques. Rather, it is an extension of statistical methods that is in part the result of a major change in the statistics community. The development of most statistical techniques was, until recently, based on elegant theory and analytical methods that worked quite well on the modest amounts of data being analyzed. The increased power of computers and their lower cost, coupled with the need to analyze enormous data sets with millions of rows, have allowed the development of new techniques based on a brute-force exploration of possible solutions.

Analytics platform – Software that provides for the computation required to carry out the statistical methods, descriptive and inquisitive queries, machine learning, visualization and Big data (which is software plus hardware). Ex: SAS, R, Tableau, Hadoop etc.

Click stream analytics/ Web analytics – Analysis on user imprints created on the web Ex: Number of clicks, probability to buy based on search times of a particular word etc.

Text analytics – Usually refers to analyzing unstructured (not tabulated) data in the form of continuous text. Ex: Face book data analysis, twitter analysis etc.

Location analytics – With advanced GPS and location data available location analytics has become quite popular, Ex: Offers based on customer location, insurance risk calculations based on proximity to hazards

Sports analytics – Analysis of sports data using analytical tool and methods. Performance as well as revenue data can be subjected to analytical procedures to achieve better results

BUSINESS ANALYTICS PROCESS - THE 7-STEP

Real-time analysis is an emerging business tool that is changing the traditional ways enterprises do business. More and more organisations are today exploiting business analytics to enable proactive decision making; in other words, they are switching from reacting to situations to anticipating them.

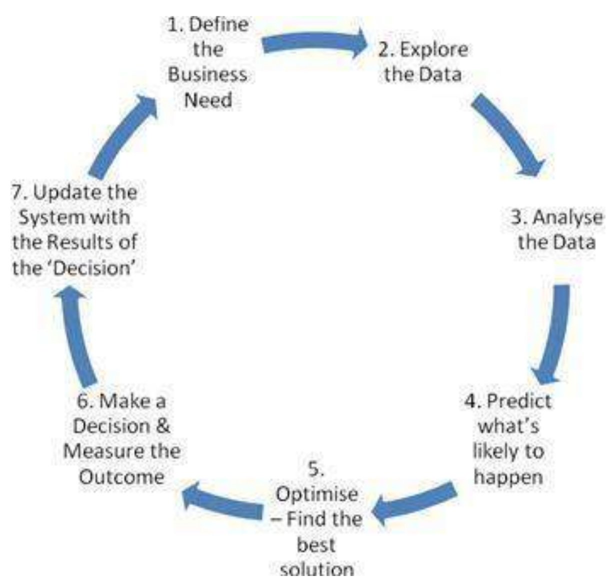
One of the reasons for the flourishing of business analytics as a tool is that it can be applied in any industry where data is captured and accessible. This data can be used for a variety of

reasons, ranging from improving customer service as well improving the organization's capability to predict fraud to offering valuable insights on online and digital information.

However business analytics is applied, the key outcome is the same: The solving of business problems using the relevant data and turning it into insights, providing the enterprise with the knowledge it needs to proactively make decisions. In this way the enterprise will gain a competitive advantage in the marketplace.

So what is business analytics? Essentially, business analytics is a 7-step process, outlined below.

Step 1. Defining the business needs



The first stage in the business analytics process involves understanding what the business would like to improve on or the problem it wants solved. Sometimes, the goal is broken down into smaller goals. Relevant data needed to solve these business goals are decided upon by the business stakeholders, business users with the domain knowledge and the business analyst. At this stage, key questions such as, "what data is available", "how can we use it", "do we have sufficient data" must be answered.

Step 2. Explore the data

This stage involves cleaning the data, making computations for missing data, removing outliers, and transforming combinations of variables to form new variables. Time series graphs are plottedas they are able to indicate any patterns or outliers. The removal of outliers from the dataset is a very important task as outliers often affect the accuracy of the model if they are allowed to remain in the data set. As the saying goes: Garbage in, garbage out (GIGO)!

Once the data has been cleaned, the analyst will try to make better sense of the data. The analyst will plot the data using scatter plots (to identify possible correlation or non-linearity). He will visually check all possible slices of data and summarise the data using appropriate visualisation and descriptive statistics (such as mean, standard deviation, range, mode, median) that will help provide a basic understanding of the data. At this stage, the analyst is already looking for generalpatterns and actionable insights that can be derived to achieve the business goal.

Step 3. Analyze the data

At this stage, using statistical analysis methods such as correlation analysis and hypothesis testing, the analyst will find all factors that are related to the target variable. The analyst will also perform simple regression analysis to see whether simple predictions can be made. In addition, different groups are compared using different assumptions and these are tested using hypothesis testing. Often, it is at this stage that the data is cut, sliced and diced and different comparisons are made while trying to derive actionable insights from the data.

Step 4. Predict what is likely to happen

Business analytics is about being proactive in decision making. At this stage, the analyst will model the data using predictive techniques that include decision trees, neural networks and logistic regression. These techniques will uncover insights and patterns that highlight relationships and 'hidden evidences' of the most influential variables. The analyst will then compare the predictive values with the actual values and compute the predictive errors. Usually, several predictive models are ran and the best performing model selected based on model accuracy and outcomes.

Step 5. Optimize (find the best solution)

At this stage the analyst will apply the predictive model coefficients and outcomes to run 'what-if' scenarios, using targets set by managers to determine the best solution, with the given constraints and limitations. The analyst will select the optimal solution and model based on the lowest error, management targets and his intuitive recognition of the model coefficients that are most aligned to the organization's strategic goal.

Step 6. Make a decision and measure the outcome

The analyst will then make decisions and take action based on the derived insights from the model and the organisational goals. An appropriate period of time after this action has been taken, the outcome of the action is then measured.

Step 7. Update the system with the results of the decision

Finally the results of the decision and action and the new insights derived from the model are recorded and updated into the database. Information such as, 'was the decision and action effective?', 'how did the treatment group compare with the control group?' and 'what was the return on investment?' are uploaded into the database. The result is an evolving database that is continuously updated as soon as new insights and knowledge are derived.

TALKING ABOUT THE PROCESS OF BUSINESS ANALYTICS

Business Analytics techniques can be deployed in any industry where data is conquered and handy to obtain business solutions through concerned data and curve it into understanding and knowledge to make valuable decisions. Multiple BI tools are implemented that helps an organization to obtain a competing asset in the market.

Business Analytics in Action: 7-steps Process outlined below;

Step 1: Address the Business Problems

Initially, business problems need to be addressed, the purpose of applying analytics is sometimes designated categorically or broken into parts. So, relevant data is selected to address these business problems by business users or business analysts equipped with domain knowledge.

Some examples are: keeping modeling for a postpaid subscription, fraud detection for credit cards, or customer analysis of a mortgage portfolio. Business experts define perimeters for the analytical process which is crucial for assuring general understanding of the goal.

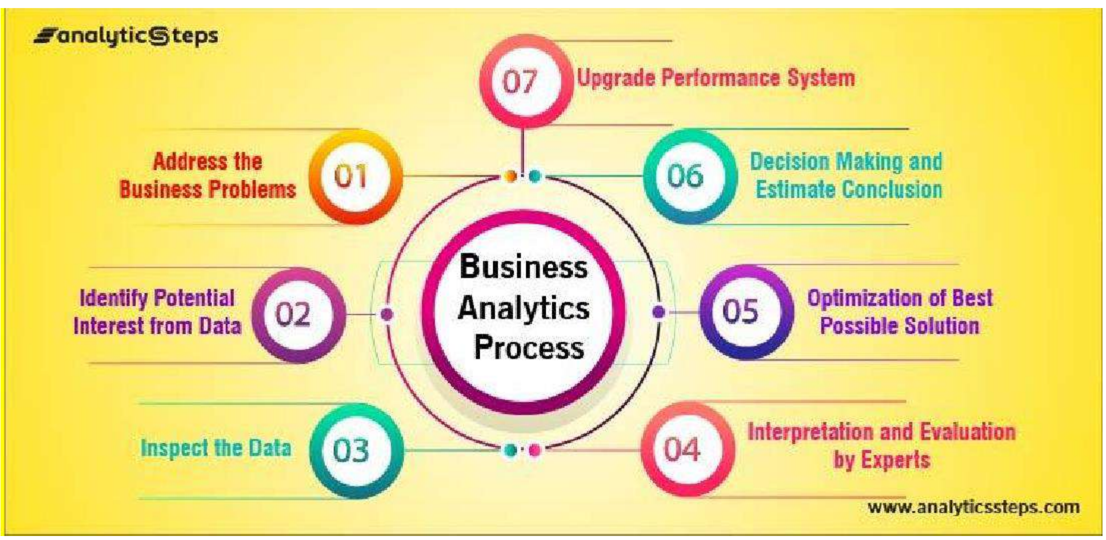
Step 2: Identify Potential Interest from Data

All sources of data having potential interest are required to identify. The key asset in this step is the more the data, the better it is. All the data will then be accumulated and consolidated in a datawarehouse or data mart or at a spreadsheet file. Some exploratory data analysis is executed to do the computation for missing data, removing outliers, and transforming variables.

For example, time-series analysis graphs are plotted to figure out some patterns or outliers, scatter plots are used to find correlation or non-linearity, OLAP system for multidimensional analysis.

<u>Step 3: Inspect the data</u>

Once moving to the analytics step, an analytical model will be predicted on the prepared and transformed data using statistical analysis techniques like correlation analysis and hypothesis testing. The analyst figures out all parameters in connection with the target variable. The business expert also performs regression analysis to make simple predictions depending upon thebusiness objective. In this step, data is also often reduced, divided, crumbled and compared with various groups to derive powerful insights from data.



7-step representation of Business Analytics Process

<u>Step 4: Interpretation and Evaluation by Experts</u>

Finally, after obtaining model results, business experts interpret and evaluate them. Results may be clusters, rules, relations, or trends known as analytical models derived from applying analytics. Experts use predictive techniques like decision trees, neural networks, logistics regression to reveal the patterns and insights that show the relationship and invisible indication of the most persuasive variables.

Several prediction models are executed to select the best performing model on the basis of modelaccuracy and consequences. But yet, to explore unknown though engaging and tribal patterns arechallenging that can add value to data and convert into new turnout opportunities.

Step 5: Optimization of Best Possible Solution

Once the analytical model has been validated and approved, the analyst will apply predictive model coefficients and conclusions to drive "what-if" conditions, using the defined to optimize the best solution within the given limitations and constraints.

Necessary considerations are how to serve model output in a user-friendly way, how to it, how to confirm the monitoring of the analytical model accurately. An optimal solution is chosen based on the lowest error, management objectives, and identification of coefficients that are associated with the company's goals.

Step 6: Decision Making and Estimate conclusions

Analysts then would make decisions and endure action based on the conclusions derived the model in accordance with the predefined business problems. Spam of period is accounted the estimation of conclusion, all the favorable and opponent consequences are measured in duration to satisfy the business needs.

Step 7: Upgrade performance system

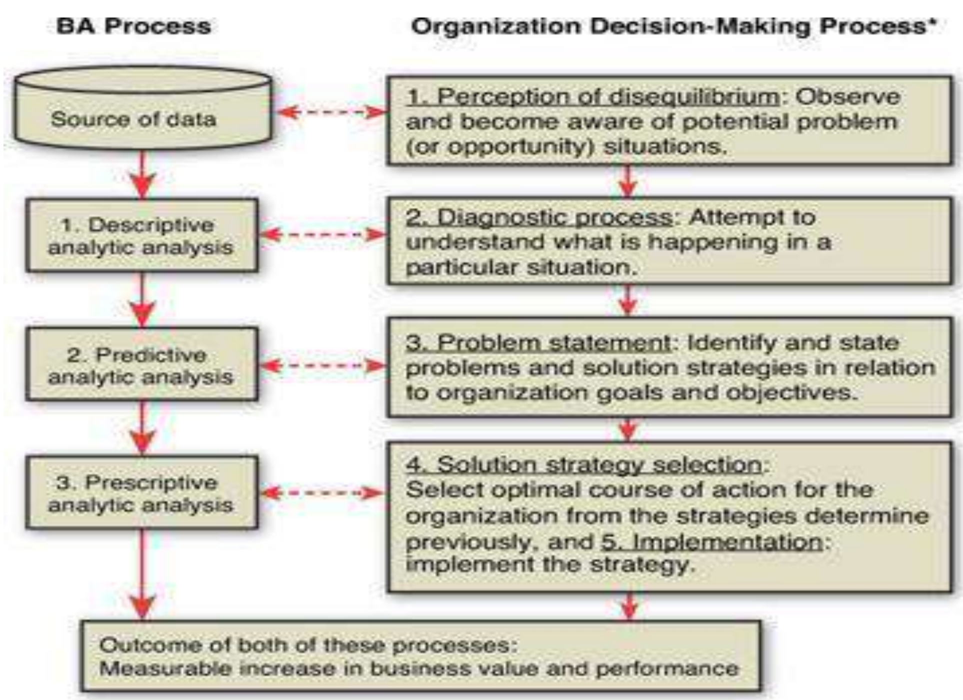
At last, the outcome of decision, action and the conclusion conducted from the model documented and updated into the database. This helps in changing and upgrading performance of the existing system.

Some queries are updated in the database such as " were the decision and action impactful?" what was the return or investment ?","how was the analysis group compared with the class?". The performance-based database is continuously updated once the new insight knowledge is extracted.

Relationship of BA Process and Organization Decision-Making Process

The BA process can solve problems and identify opportunities to improve business performance.In the process, organizations may also determine strategies to guide operations and help achieve competitive advantages. Typically, solving problems and identifying strategic opportunities to follow are organization decision-making tasks. The latter, identifying opportunities, can be viewed as a problem of strategy choice requiring a solution. It should come as no surprise that the BA process described in Section 1.2 closely parallels classic organization decision-making

processes. As depicted in Figure 1.2, the business analytic process has an inherent relationship tothe steps in typical organization decision-making processes.



*Source: Adapted from Figure 1 in Elbing (1970), pp. 12–13.

Figure 1.2 Comparison of business analytics and organization decision-making processes

The *organization decision-making process* (ODMP) developed by Elbing (1970) and presented in Figure 1.2 is focused on decision making to solve problems but could also be applied to finding opportunities in data and deciding what is the best course of action to take advantage of them. The five-step ODMP begins with the perception of disequilibrium, or the awareness that a problem exists that needs a decision. Similarly, in the BA process, the first step is to recognize that databases may contain information that could both solve problems and find opportunities to improve business performance. Then in Step 2 of the ODMP, an exploration of the problem to determine its size, impact, and other factors is undertaken to diagnose what the problem is. Likewise, the BA descriptive analytic analysis explores factors that might prove useful in solvingproblems and offering opportunities. The ODMP problem statement step is similarly structuredto the BA predictive analysis to find strategies, paths, or trends that clearly define a problem or opportunity for an organization to solve problems. Finally, the ODMP's last steps of strategy selection and implementation involve the same kinds of tasks that the BA process requires in the final prescriptive step (make an optimal selection of resource allocations that can be implemented for the betterment of the organization).

The decision-making foundation that has served ODMP for many decades parallels the BA process. The same logic serves both processes and supports organization decision-making skills and capacities.

What role does business analytics play in organizational decision-making?

Better business decisions

Data analytics **allows Executives to make decisions based on statistical facts**. Those facts can be used to guide choices about future company growth by evaluating a long-term view of themarket and competition.

How Business Analytics impacts Decision Making in Businesses

Business analytics enables managers to understand their company's dynamics, forecast market developments, and manage risks.

Companies are adopting analytics and rigorous statistical reasoning to make decisions that enhance efficiency, risk management, and profits, rather than "going with gut" when keeping inventories, pricing solutions, or employing people.

Data is provided for all essential company activities, including industry trends, consumer behavior, productivity, inventory, and thorough financial analysis.

Business intelligence software collects information and transforms it into clear insights to enable actionable and strategic decision-making, allowing employees to easily achieve their objectives.

What is Business Analytics?

Big data is driving a lot of advancement in a variety of businesses. If you haven't already figured out how it works to make businesses and organizations more efficient through its many procedures, now is the time to do so!

One way Big Data may help with this is through strategic usage in Business Analytics. Don't worry if you've never heard of the phrase Business Analytics; it's very closely related to a term you've probably heard of — Data Analytics.

Data analytics is the process of analyzing data using various approaches, some quantitative and others predictive, in order to get insight into the data. So, what exactly is Business Analytics? Simply said, it's Data Analytics, but it's used to help businesses achieve their goals and make business decisions.

Business Analytics, like Data Analytics, may utilize past data to make sense of current data,discover trends, and forecast which consequence is likely to occur (if at all) and when.

Finally, it may provide crucial insight into what will be the greatest conclusion for your organization — and as a consequence, it becomes an indispensable tool, particularly when it comes to making profitable judgments.

As technology has slowly but steadily advanced beyond our expectations, there are now countless data sources and many more ways to categorize and analyze them. Every person, gadget, or organization generates an almost unmanageable quantity of data practically every day,which causes issues with logistical storage.

Businesses use a variety of methods to ensure that issue is resolved, with cloud storage being a significant one. Beyond that, the more immediate problem is the expertise necessary to sift through these mounds of data and uncover the insight that can cause a ripple effect across a business function or branch.

Because that's the beauty of data: it doesn't only help in one area, even if that's how its benefits are typically perceived. It has an immediate influence and then gradually and progressively altersthe system around which the data point was constructed.

Business Analytics, in particular, uses Data Analytics to enhance efficiency in terms of production and expenses, as well as to determine whether or not the general structural systems that are applied are successful.

Furthermore, Business Analytics may aid in broad strategic direction as well as the formalization of decision making procedures.

Business Analytics for Competitive

Advantage The Competitive Advantage of

Business Analytics

Business analytics is the process of gathering data, measuring business performance, and producing valuable conclusions that can help companies make informed decisions on the future of the business, through the use of various statistical methods and techniques.

Analytics has become one of the most important tools at an organization's disposal. When data and analytics work hand in hand, the benefits become obvious. Companies can leverage data to improve cost savings, redefine processes, drive market strategy, establish competitive differentiators and, perhaps most importantly, build exceptional and truly personalized customer experience.

Business analytics for organisations is becoming a competitive advantage and is now necessaryto apply business analytics, particularly its subset of predictive business analytics. The use of business analytics is a skill that is gaining mainstream value due to the increasingly thinner margin for decision error. It is there to provide insights, predict the future of the business and inferences from the treasure chest of raw transactional data, that is internal and external data that many organizations now store (and will continue to store) as soft copy.

Business analytics enables differentiation. It is primarily about driving change. Business analytics drives competitive advantage by generating economies of scale, economies of scope, and quality improvement. Taking advantage of the economies of scale is the first way organizations achieve comparative cost efficiencies and drive competitive advantage against their peers. Taking advantage of the economies of scope is the second-way organizations achievecomparative cost efficiencies and drive competitive advantage against their peers.

Business analytics improves the efficiency of business operations. The efficiencies that accumulate when a firm embraces big data technology eventually contributes to a ripple effect ofincreased production and reduced business costs. In the modern world, the vast quantities of dataproduced by corporations make their study and management practically impossible.

One can make the case that increasing the primary source of attaining a competitive advantage will be an organization's competence in mastering all flavors of analytics. If your management team is analytics-impaired, then your organization is at risk. Predictive business analytics is arguably the next wave for organizations to successfully compete. This will result not only from being able to predict outcomes but also to reach higher to optimize the use of their resources, assets and trading partners. It may be that the ultimate sustainable business strategy is to foster analytical competency and eventually mastery of analytics among an organization's workforce.

Analytics gives companies an insight into their customers' behaviour and needs. It also makes it possible for a company to understand the public opinion of its brand, to follow the results of various marketing campaigns, and strategize how to create a better marketing strategy to nurture long and fruitful relationships with its customers.

Business analytics helps organisations to know where they stand in the industry or a particular niche provides the company with the needed clarity to develop effective strategies to position itself better in the future.

For a company to remain competitive in the modern marketplace that requires constant change and growth, it must stay informed on the latest industry trends and best practices. Not only does

business analytics provide the needed knowledge for companies to survive in today's constantly changing business environment, but it also makes room for growth and improvement, providing a detailed look into various opportunities and challenges that companies face on a day-to-day basis.

The retention of company employees has been a concern for business enterprises although it is taken more seriously in some niches that it is in other industries. A recent study that was conducted by IBM infers that a business enterprise had over 5,000 job applications reviewed but only hired 200 employees monthly. Big data has made it possible for companies to quickly analyses long time worker's histories to identify the job traits for long-term employment prospects.

As a result, corporations and small business enterprises are revamping their recruitment process which reduces employee turnover significantly. Companies can dedicate resources that are newlyavailable to activities that are of more productive value to the business and increase their levelsof service delivery. The retention of an experienced pool of employees can significantly assist a business enterprise to outperform its competitors using their long-term experiences.

UNIT II MANAGING RESOURCES FOR BUSINESS ANALYTICS

Managing BA Personnel, Data and Technology. Organisational Structures aligning BA. Managing Information policy, data quality and change in BA.

Introduction

To fully understand why business analytics (BA) is necessary, one must understand the nature of the roles BA personnel perform. In addition, it is necessary to understand resource needs of a BA program to better comprehend the value of the information that BA provides. The need for BA resources varies by firm to meet particular decision support requirements. Some firms may choose to have a modest investment, whereas other firms may have BA teams or a department of BA specialists. Regardless of the level of resource investment, at minimum, a BA program requires resource investments in BA personnel, data, and technology.

Business Analytics Personnel

One way to identify personnel needed for BA staff is to examine what is required for certification in BA by organizations that provide BA services. *INFORMS* (www.informs.org/Certification-Continuing-Ed/Analytics- Certification), a major academic and professional organization, announced the startup of a *Certified Analytic Professional* (CAP) program in 2013.

Another more established organization, *Cognize* (www.cognizure.com/index.aspx), offers a variety of service products, including business analytic services. It offers a general certification *Business Analytics Professional* (BAP) exam that measures existing skill sets in BA staff and identifies areas needing improvement (www.cognizure.com/cert/bap.aspx). This is a tool to validate technical proficiency, expertise, and professional standards in BA. The certification consists of three exams covering the content areas listed in Table 3.1.

Exam	Topic	Specific Content Areas Covered	Examples
I	Statistical Methods	1. Visualizing and Exploring Data 2. Descriptive Statistics 3. Probability Distributions 4. Sampling and Estimation 5. Statistical Inference 6. Regression Analysis 7. Predictive Modeling and Analysis	1. Graphs and charts 2. Mean, median, mode 3. Normal distribution 4. Confidence intervals 5. Hypothesis testing 6. Multiple regression 7. Curve fitting of models and functions to raw data
II	Operations Research Methods	1. Linear Optimization 2. Integer Optimization 3. Nonlinear Optimization 4. Simulation 5. Decision Analysis 6. Forecasting	1. Linear programming 2. Integer programming 3. Quadratic programming 4. Monte Carlo method 5. Expected value analysis 6. Exponential smoothing
III	Case Studies	Practical knowledge of real-world situations	Application of the methods above to solve a real world problem

**Source*: Adapted from Cognize Organization website (www.cognizure.com/cert/bap.aspx).

Table 3.1 Cognize Organization Certification Exam Content Areas

Most of the content areas in Table 3.1 will be discussed and illustrated in subsequent chapters and appendixes. The three exams required in the Cognize certification program can easily be understood in the context of the three steps of the BA process (descriptive, predictive, and prescriptive) discussed in previous chapters. The topics in Figure 3.1 of the certification program are applicable to the three major steps in the BA process. The basic statistical tools apply to the descriptive analytics step, the more advanced statistical tools apply to the predictive analytics step, and the operations research tools apply to the prescriptive analytics step. Some of the tools can be applied to both the descriptive and the predictive steps.

Likewise, tools like simulation can be applied to answer questions in both the predictive and the prescriptive steps, depending on how they're used. At the conjunction of all the tools is the reality of case studies. The use of case studies is designed to provide practical experience whereall tools are employed to answer important questions or seek opportunities.

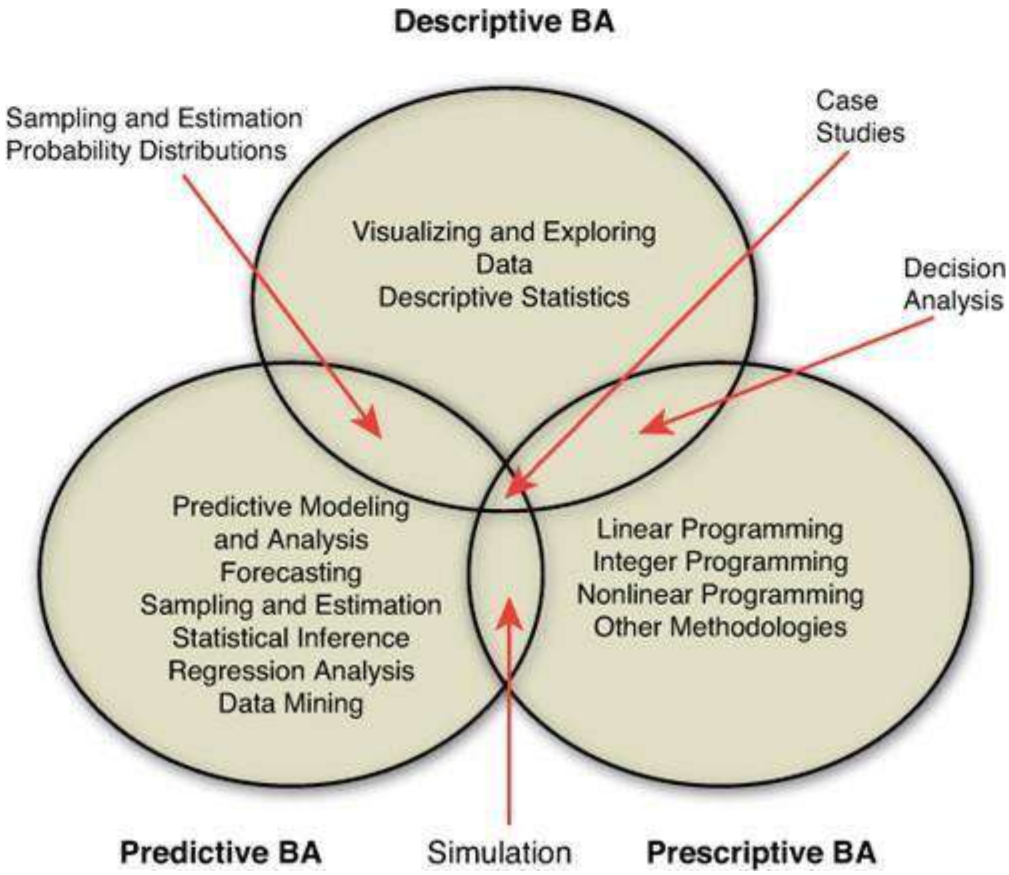


Figure 3.1 Certification content areas and their relationship to the steps inBA

Other organizations also offer specialized certification programs. These certificationsinclude other areas of knowledge and skills beyond just analytic tools. IBM, for example, offers a variety of specialized BA certifications (www- 03.ibm.com/certify/certs/ba_index.shtml). Although these include certifications in several dozen statistical, information systems, and analytic methodologies related to BA, they also include specialized skill sets related to BA personnel (administrators, designers, developers, solution experts, and specialists), as presented in Table 3.2.

BA Personnel Specialty	Description
Administrators	Within the context of the IBM BA and business intelligence (BI) software platforms, administrators manage servers (their load balancing, installation, and configurations). They manage reports from computer portals, manage dispatchers, and perform troubleshooting for technology. They are also in charge of user authorization and authentication for security.
Designers	As members of a team, designers are responsible for building reports using relational data models, as well as enhancing, customizing, and managing professional reports.
Developers	As members of a team, the skills required for developers are closely tied to the BA process and involve the application of analytics, data warehousing, model building, use of operations research and statistical methodologies, and real-time monitoring of data flow to users.
Solution Experts	As members of a team, solution experts analyze, plan, design, deploy, and operate BA applications using an appropriate methodology and development approach. This requires knowledge in many differing BA software applications, including statistical, information systems, and operations research methods.
Technical Specialists	As members of a team, they are responsible for the installation and configuration of BA and BI applications.

**Source*: Adapted from IBM website (www-03.ibm.com/certify/certs/ba_index.shtml).

Table 3.2 Types of BA Personnel*

With the variety of positions and roles participants play in the BA process, this leads to the question of what skill sets or competencies are needed to function in BA. In a general sense, BA positions require competencies in business, analytic, and information systems skills. As listed in Table 3.3, business skills involve basic management of people and processes. BA personnel must communicate with BA staffers within the organization (the BA team members)and the other functional areas within a firm (BA customers and users) to be useful. Because they serve a variety of functional areas within a firm, BA personnel need to possess customer service skills so they can interact with the firm's personnel and understand the nature of the problems they seek to solve. BA personnel also need to sell their services to users inside the firm. In addition, some must lead a BA team or department, which requires considerable

Type of Skill or Competency	Description of Possible Roles
Business	• Leadership • People-related management and communication skills

Type of Skill or Competency	Description of Possible Roles
Business	• Leadership • People-related management and communication skills • Manage BA projects (prioritize, schedule, and so on) • Manage BA processes (rules, procedures governing acceptance and use) • Determine project requirements • Train BA personnel to develop competencies
Analytic	• Know how to use statistical methodologies • Know how to use operations research methodologies • Know how to use data mining for quantitative data and text analytics for unstructured data
Information system	• Maintain and use computer portals • Identify and extract data • Maintain quality data

Table 3.3 Select Types of BA Personnel Skills or CompetencyRequirements

Fundamental to BA is an understanding of analytic methodologies listed in Table 3.1 and others not listed. In addition to any tool sets, there is a needto know how they are integrated into the BA process to leverage data (structured or unstructured) and obtain information that customers who willbe guided by the analytics desire.

In summary, people who undertake a career in BA are expected to know how to interact with people and utilize the necessary analytic tools to leverage data into useful information that can be processed, stored, and shared in information systems in a way that guides a firm to higher levels ofbusiness performance.

Business Analytics Data

Structured and unstructured data (introduced in Chapter 2, "Why Are Business Analytics Important?") 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 (Bartlett, 2013, pp. 238–239). 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	*Customer relationship management* (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	*Enterprise resource planning* (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 Web site 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 AnalyticsCan Be Based

Type of External Data	Measured By
Customer Satisfaction	• Revenue, profit • Market share, sales • Product/service survey data • Loyalty • Brand awareness • Average spend per customer
Customer Demographics	• Geographic location (distance from market) • Income level • Market size
Competition	• Market share • Competitor profitability • Advertising/promotion efforts • Preference surveys • Web chatter on products
Economic	• Population statistics • Income distribution statistics

Table 3.5 Typical External Sources of Data on Which Business AnalyticsCan 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), psychographic or demographic data sourced from *Claritas* (www.claritas.com), 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 tothe 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, butthere 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. We will discuss some of these quality data measures in later chapters.

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 toa 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 supportspersonnel 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 operations.

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 *apps* (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, *intranets* (internal corporate networks based on Internet technology with limited access to employees within the firm), and *extranets* (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

Of particular importance for BA is the data management technologies listed in Table 3.6. *Database management systems* (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 thetask 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 retrieveunstructured 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 in size. 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* isa table of contents listing a firm's current data inventory and what data filescan be built or purchased. The typical content of the database encyclopedia is presented in Table 3.7. Of particular importance for BA is the *data manipulation language* tools included in DMBS. These tools are used to searchdatabases 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.

Database Content Item	Description
Purpose	Why the database exists, including any additional reports or analyses used in leveraging the data.
Time	Window of time period when the data is collected or will be useful.
Source	Internal (auditing, accounting, and so on) and external (customers, and so on) sources.
Schematics	Diagrams illustrating the connections between tables and other data files.
Cost	Expense of collecting data, including purchasing prices.
Availability of Data	Window of time when the data may be available.
Collection Techniques	Methods of collection, including observation, data mining, census, and focus groups.
Collection Tools	Web, customer generated, e-survey, and so on.

Table 3.7 Database Encyclopedia Content

Data warehouses are databases that store current and historical data of potential interest to decision makers. What a data warehouse does is make data available to anyone who needs access to it. In a data warehouse, the data is prohibited from being altered. Data warehouses also provide a set ofquery tools, analytical tools, and graphical reporting facilities. Some firms use intranet portals to make data warehouse information widely available throughout a firm.

Data marts are focused subsets or smaller groupings within a data warehouse. Firms often build enterprise-wide data warehouses where a central data warehouse serves the entire organization and smaller, decentralized data warehouses (called data marts) are focused on a limited portion of the organization's data that is placed in a separate database for a specific population of users. For example, a firm might develop a smaller database on just product quality to focus efforts on quality customer and product issues. A data mart can be constructed more quickly and at lower cost than enterprise-wide data warehouses to concentrate effort in areas ofgreatest concern.

Once data has been captured and placed into database management systems, it is available for analysis with BA tools, including online analytical processing, as well as data, text, and Web mining technologies. *Online analytical processing* (OLAP) is software that allows users to view data in multiple dimensions. For example, employees can be viewed in terms of their age, sex, geographic location, and so on. OLAP would allow identification of the number of employees who are age 35, male, and in thewestern region of a country. OLAP allows users to obtain online answers toad hoc questions quickly, even when the data is stored in very large databases.

Data mining is the application of a software, discovery-driven process that provides insights into business data by finding hidden patterns and relationships in big data or large databases and inferring rules from them topredict future behavior. The observed patterns and rules are used to guide decision-making. They can also act to forecast the impact of those decisions. It is an ideal predictive analytics tool used in the BA process mentioned in Chapter 1, "What Are Business Analytics?" The kinds of information obtained by data mining includethose in Table 3.8.

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 identifying tea drinkers who choose that beverage from others offered in flight from 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 (mentioned in Chapter 2) 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. Giventhat the majority of the information stored in businesses is in the form of unstructured data (emails, 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.

In addition to the general software applications discussed earlier, there are focused software applications used every day by BA analysts in conducting the three steps of the BA process (see Chapter 1). These include *Microsoft Excel*® spreadsheet applications, SAS applications, and SPSS applications. *Microsoft Excel* (www.microsoft.com/) spreadsheet systems have add-in applications specifically used for BA analysis. These add-in applications broaden the use of Excel into areas of BA. *Analysis ToolPak* isan Excel add-in that contains a variety of statistical tools (for example, graphics and multiple regression) for the descriptive and predictive BA process steps. Another Excel add-in, *Solver*, contains operations research optimization tools (for example, linear programming) used in the prescriptive step of the BA process.

SAS® Analytics Pro (www.sas.com/) software provides a desktop statistical toolset allowing users to access, manipulate, analyze, and present information in visual formats. It permits users to access data from nearly any source and transform it into meaningful, usable information presented in visuals that allow decision makers to gain quick understanding of critical issues within the data. It is designed for use by analysts, researchers, statisticians, engineers, and scientists who need to explore, examine, and present data in an easily understandable way and distribute findings in a variety of formats. It is a statistical package chiefly useful in the descriptiveand predictive steps of the BA process.

IBM's *SPSS software* (www-01.ibm.com/software/analytics/spss/) offers users a wide range of statistical and decision-making tools. These tools include methodologies for data collection, statistical manipulation, modeling trends in structured and unstructured data, and optimizing analytics. Depending on the statistical packages acquired, the software can cover all three steps in the BA process.

Other software applications exist to cover the prescriptive step of the BA process. One that will be used in this book is LINGO® by Lindo Systems (www.lindo.com). LINGO is a comprehensive tool designed to make building and solving optimization models faster, easier, and more efficient. LINGO provides a completely integrated package that includes an understandable language for expressing optimization models, a full-featured environment for building and editing problems, and a set of built-in solvers to handle optimization modeling in linear, nonlinear, quadratic, stochastic, and integer programming models.

In summary, the technology needed to support a BA program in any organization will entaila general information system architecture, including database management systems and progress in greater specificity down to the software that BA analysts need to compute their unique contributions tothe organization. Organizations with greater BA requirements will have substantially more technology to support BA efforts, but all firms that seek to use BA as a strategy for competitive advantage will need a substantial investment in technology, because BA is a technology-dependent undertaking.

Summary

Why BA is important is directly proportional to what it costs. In this chapter, we have explored costs, but also many of the benefits of BA as a means to justify why a BA program isnecessary. This chapter discussed what resources a firm would need to support a BA program.From this, three primary areas of resources were identified: personnel, data, and technology. Having identified BA personnel and needed skill sets, a reviewof content in BA certification exams was presented. Types of personnel specialties also were discussed. BA data internal and external sources were presented as a means of data categorization. Finally, BA technologywas covered in terms of general, organization-wide information systems supportto individual analyst support software packages.

In this chapter, we focused on the investment in resources needed to havea viable business analytics operation. In Chapter 4, we begin Part III, "How Can Business Analytics Be Applied?" Specifically, in the next chapter we will focus on how the resources mentioned in this chapter are placed into anorganization and managed to achieve goals.

Discussion Questions

1. How does using BA certification exam content explain skill sets for BA analysts?
   What skill sets are necessary for BA personnel?
2. Why is leadership an important skill set for individuals looking tomake a career inBA?
3. Why is categorizing data from its sources important in BA?
4. What is data quality, and why is it important in BA?
5. What is the difference between a data warehouse and a datamart?

4. How Do We Align Resources to Support Business Analytics within an

Organization?

Organization Structures Aligning Business Analytics

According to Isson and Harriott (2013, p. 124), to successfully implement business analytics (BA) within organizations, the BA in whatever organizational form it takes must be fully integrated throughout a firm. This requires BA resources to be aligned in a way that permits a viewof customer information within and across all departments, access to customer information from multiple sources (internal and external to the organization), access tohistorical analytics from a central repository, and making technology resources align to be accountable for analytic success. The commonality of these requirements is the desire for an alignment that maximizes the flow of information into and through the BA operation, which in turn processes and shares information to desired users throughout the organization. Accomplishing this information flow objective requires consideration of differing organizational structures and managerial issues that help align BA resources to best serve an organization.

Organization Structures

As mentioned in Chapter 2, "Why Are Business Analytics Important?",most organizations are hierarchical, with senior managers making the strategic planning decisions, middle-level managers making tactical planning decisions, and lower-level managers making operational planningdecisions. Within the hierarchy, other organizational structures exist to support the development and existence of groupings of resources like those needed for BA. These additional structures include programs, projects, andteams. A *program* in this context is the process that seeks to create an outcome and usually involves managing several related projects with the intention of improving organizational performance. A program can also bea large project. A *project* tends to deliver outcomes and can be defined as having temporary rather than permanent social systems within or across organizations to accomplish particular and clearly defined tasks, usually under time constraints. Projects are often composed of teams. A *team* consists of a group of people with skills to achieve a common purpose.

Teams are especially appropriate for conducting complex tasks that have many interdependent subtasks.

The relationship of programs, projects, and teams with a business hierarchy is presented in Figure 4.1. Within this hierarchy, the organization's senior managers establish a *BA program* initiative to mandatethe creation of a BA grouping within the firm as a strategic goal. A BA program does not always have an end-time limit. Middle-level managers reorganize or break down the strategic BA program goals into doable *BA project* initiatives to be undertaken in a fixed period of time. Some firms have only one project (establish a BA grouping) and others, depending on the organization structure, have multiple BA projects requiring the creation of multiple BA groupings. Projects usually have an end-time date in which to judge the successfulness of the project. The projects in some cases are further reorganized into smaller assignments, called *BA team* initiatives, to operationalize the broader strategy of the BA program. BA teams may have a long-standing time limit (for example, to exist as the main source of analytics for an entire organization) or have a fixed period (for example, to work on aspecific product quality problem and then end).

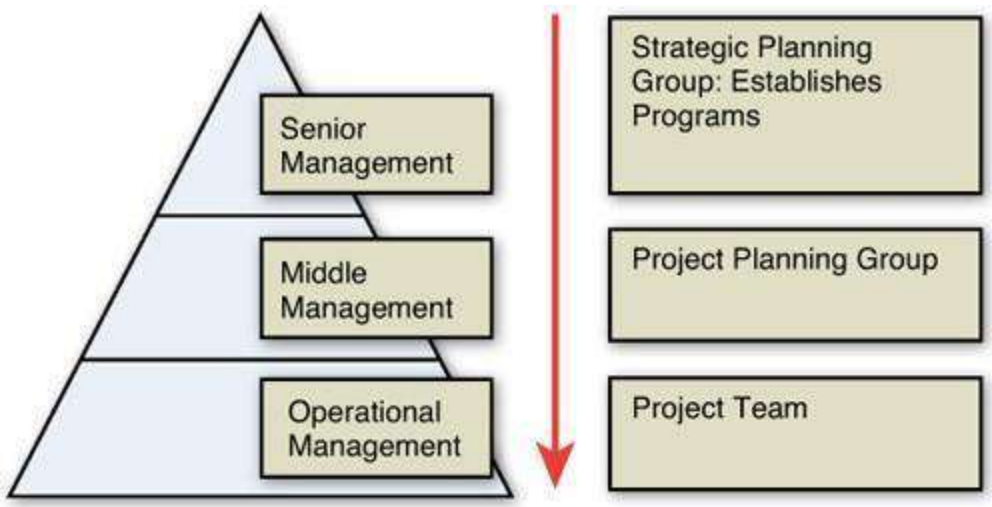


Figure 4.1 Hierarchal relationships program, project, and team planning

In summary, one way to look at the alignment of BA resources is to view it as a progression of assigned planning tasks from a BA program, to BAprojects, and eventually to BA teams for implementation. As shown in Figure 4.1, this hierarchical relationship is a way to examine how firms align planning and decision-making workload to fit strategic needs and requirements.

BA organization structures usually begin with an initiative that recognizes the need to use and develop some kind of program in analytics. Fortunately, most firms today recognize this need. The question then becomes how to match the firm's needs within the organization to achieve its strategic, tactical, and operations objectives within resource limitations. Planning the BA resource allocation within the organizational structure of a firm is a starting place for thealignment of BA to best serve a firm's needs.

Aligning the BA resources requires a determination of the amount of resources a firm wants to invest. The outcome of the resource investment might identify only one individual to compute analytics for a firm. Becauseof the varied skill sets in information systems, statistics, and operations research methods, a more common beginning for a BA initiative is the creation of a BA team organization structure possessing a variety of analytical and management skills. (We will discuss BA teams in Section 4.1.2.) Another way of aligning BA resources within an organization is to use a project structure. Most firms undertake projects, and some firms actually use a project structure for their entire organization. For example, consulting firms might view each client as a project (or product) and align their resources around the particular needs of that client. A project structure often necessitates multiple BA teams to deal with a wider variety of analytic needs. Even larger investments in BA resources might be requiredby firms that decide to establish a whole BA department containing all the BA resources for a particular organization. Although some firms create BA departments, the departments don't have to be large. Whatever the organization structure that is used, the role of BA is a staff (not line management) role in their advisory and consulting mission for the firm.

In general, there are different ways to structure an organization to align its BA resources to serve strategic plans. In organizations where functional departments are structured on a strict hierarchy, separate BA departments orteams have to be allocated to each functional area, as presented in Figure

This *functional organization structure* may have the benefit of stricter functional control by the VPs of an organization and greater efficiency in focusing on just the analytics within each specialized area. On the other hand, this structure does not promote the cross-department access that is suggested as a critical success factor for the implementation of a BA program.

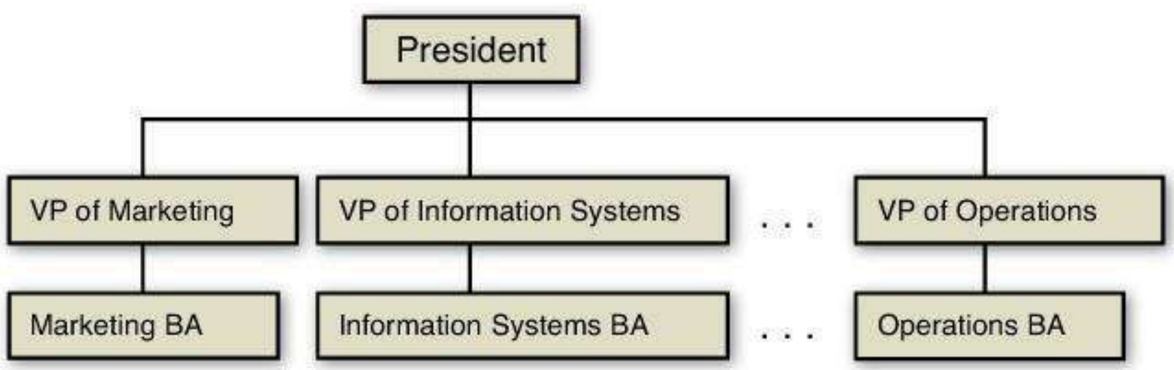


Figure 4.2 Functional organization structure with BA

The needs of each firm for BA sometimes dictate positioning BA within existing organization functional areas. Clearly, many alternative structures can house a BA grouping. For example, because BA provides information to users, BA could be included in the functional area of management information systems, with the *chief information officer* (CIO) acting as boththe director of information systems (which includes database management) and the leader of the BA grouping.

An alternative organizational structure commonly found in large organizations aligns resources by project or product and is called a *matrix organization*. As illustrated in Figure 4.3, this structure allows the VPs some indirect control over their related specialists, which would include the

BA specialists but also allows direct control by the project or product manager. This, similar to the functional organizational structure, does notpromote the cross-department access suggested for a successful implementation of a BA program.

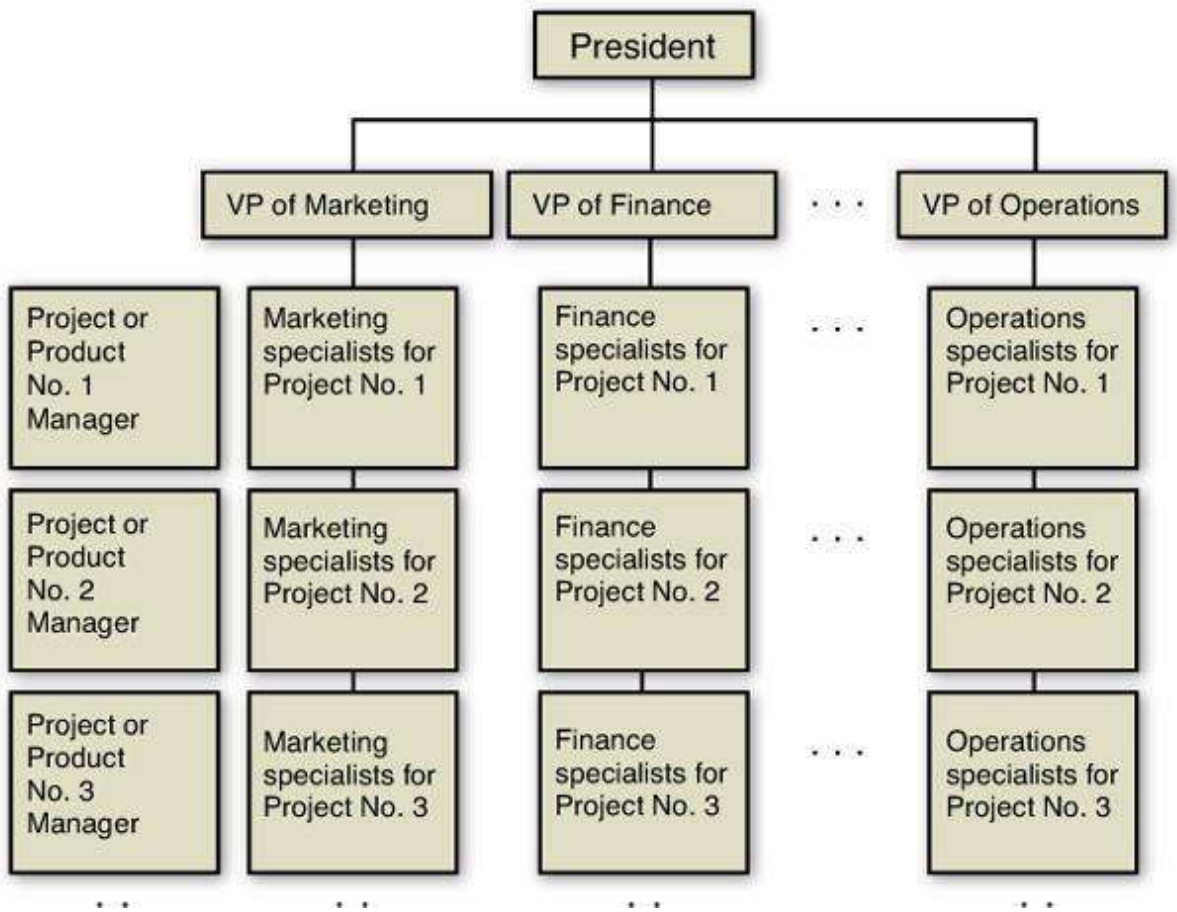


Figure 4.3 Matrix organization structure

The literature suggests that the organizational structure that best aligns BA resources is one in which a department, project, or team is formed in a staff structure where access to and from the BA grouping of resources permits access to all areas within a firm, as illustrated in Figure 4.4 (Laursen and Thorlund, 2010, pp. 191–192; Bartlett, 2013, pp. 109–111; Stubbs, 2011, p.

68). The dashed line indicates a staff (not line management) relationship. This *centralized BA organization structure* minimizes investment costs by avoiding duplications found in both the functional and the matrix styles of organization structures. At the same time, it maximizes information flow between and across functional areas inthe organization. This is a logical structure for a BA group in its advisory role to the organization. Bartlett

(2013, pp. 109–110) suggests other

advantages of a centralized structure like the one in Figure 4.4. These include a reduction in the filtering of information traveling upward throughthe organization, insulation from politicalinterests, breakdown of the *siloed functional area* communication barriers, a more central platform for reviewing important analyses that require a broader field of specialists, analytic- based group decision-making efforts, separation of the line management leadership from potential clients (for example, the VP of marketing would not necessarily come between the BA group working on customer service issues for a department within marketing), and better connectivity between BA and all personnel within the area of problem solving.

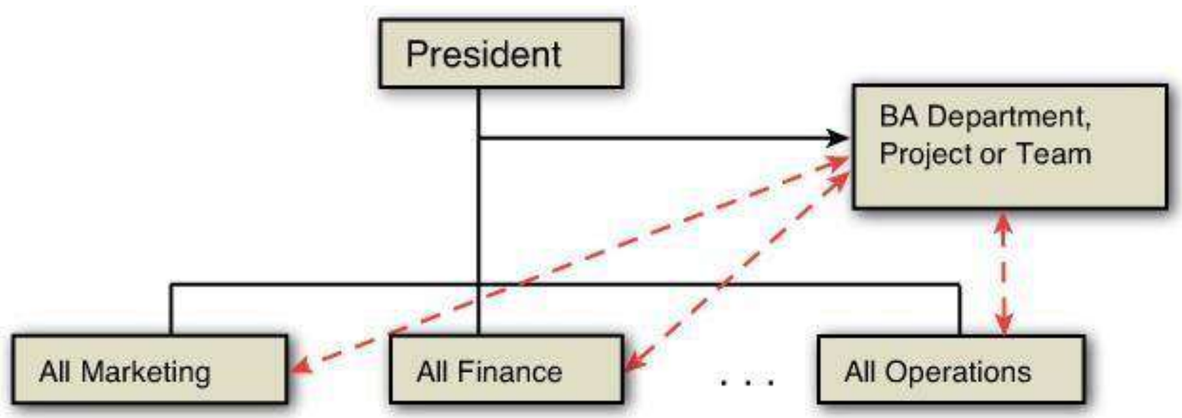


Figure 4.4 Centralized BA department, project, or team organizationstructure

Given the advocacy and logic recommending a centralized BA grouping, there are reasons for all BA groupings to be centralized. These reasons helpexplain why BA initiatives that seekto integrate and align BA resources into any type of BA group within the organization sometimes fail. The listing in Table 4.1 is not exhaustive, but it provides some of the important issues to consider in the process of structuring a BA group.

Reason	Description
Lack of Executive Sponsorship	Senior executive failure to recognize the value of BA and its importance eventually leads to a reduction in resources and eventual failure.
Limited Context Perception	There is an incorrect perception that analytics must be applied within a particular functional area in order to have the necessary validity to be applied to that area. Example: Financial regression analysis can only be applied correctly in the context of the finance area.
Belief of Physical Proximity	There is misperception that it takes physical proximity of the BA grouping in the business application area to be valid.
Lack of Leadership in BA Groupings	Without an advocate leader in the organization, as well as leaders in BA projects and teams to move the analysis to achieve desired goals, the entire BA effort will lead to eventual failure.
Lack of support	Without support for needed personnel, collecting data and technology to process the data will lead to failure.
Lack of Collaboration Across All Organizational Groups	Analytics that solve problems across multiple, functional areas are more likely to be accepted and successful than those that lack the cross-over into multiple organizational groups.
Lack of Skilled and Human Resources	BA departments, projects, or teams that don't have the skilled personnel to deal with the execution of analysis will eventually cause the failure of BA.
Inability to Delegate Responsibility	There is a desire to delegate responsibility to solve problems locally (a matter of trusting your own) rather than seeking help throughout the organization. This impedes the flow of problem solving efforts by an external BA department and impedes communication of information needed to successfully apply BA.
Lack of Integrated Processes	Information that is stored in silos and not shared makes it more difficult for BA analysis to succeed.

Table 4.1 Reasons for BA Initiative and Organization Failure

In summary, the organizational structure that a firm may select for the positioning of theirBA grouping can either be aligned within an existing organizational structure, or the BA grouping can be separate, requiring fullintegration within all areas of an organization. While some firms may start with a number of small teams to begin their BA program, other firms may choose to start with a full-sized BA department. Regardless of the size of the investmentin BA resources, it must be aligned to allow maximum information flow between and across functional areas to achieve the most benefits BA can deliver.

4.1.2. Teams

When it comes to getting the BA job done, it tends to fall to a BA team. For firms that employ BA teams the participants can be defined by the roles they play in the team effort. Some of the roles BA team participants undertake and their typical background are presented in Table 4.2.

Title or Function	Role Description	Background or Skills of Participant
Analytics Modeler	Develop and maintain predictive and forecasting models to provide insight.	Statistics, operation research, analytic modeling.
Analytics Process Designer	Develop and enforce reusable processes to reduce BA execution time.	Management consultant, process mapping, systems design.
Analytics Analyst	Respond to BA inquiries from functional areas within the firm to gain insight.	Reporting, problem solving, communicating, and providing customer service.
BA Team Head	Provide leadership to BA team, define strategies and tactics to ensure improved business performance, and interface with management.	BA manager or administrator.
Business Domain Expert	Provide business experience to ensure relevance of insight, help interpret business measures and the meaning of data.	Business experience in the area where the problem or opportunity exists.
Data Manager	Ensure data availability and access while minimizing costs.	Data modeling or warehousing, experience in data quality processes.
Implementation Specialist	Ensure rapid and robust model deployment to reduce time in interface.	Information system and data warehousing expertise, enterprise architecture experience.
Monitoring Analyst	Identify, establish, and enforce common analytics to be used to measure value and optimize effort.	Management and BA expert, predictive and financial modeling, process design, and team mentoring.

*Source: Adapted from Stubbs (2013), pp.137–149; Stubbs (2011) Table3.3; Laursen andThorlund (2010), p.15.

Table 4.2 BA Team Participant Roles*

Aligning BA teams to achieve their tasks requires collaboration efforts from team members and from their organizations. Like BA teams, *collaboration* involves working with people to achieve a shared and explicitset of goals consistent with their mission. BA teams also have a specific mission to complete. Collaboration through teamwork is the means to accomplish their mission.

Team members' need for collaboration is motivated by changes in the nature of work (no more silos to hide behind, much more open environment, and so on), growth in professions (for example, interactive jobs tend to be more professional, requiring greater variety in expertise sharing), and the need to nurture innovation (creativity and innovation are

fostered by collaboration with a variety of people sharing ideas). To keep one's job and to progress in any business career, particularly in BA, team members must encourage working with other members inside a team and out. For organizations, collaboration is motivated by the changing nature ofinformation flow (that is, hierarchical flows tend to be downward, whereas in modern organizations, flow is in all directions) and changes in the scopeof business operations (that is, going from domestic to global allows for a greater flow of ideas and information from multiple sources in multiple locations).

How does a firm change its culture of work and business operations to encourage collaboration? One way to affect the culture is to provide the technology to support a more open, cross-departmental information flow. This includes e-mail, instant messaging, *wikis* (collaboratively edited works, like Wikipedia), use of social media and networking through *Facebook* and *Twitter*, and encouragement of activities like collaborative writing, reviewing, and editing efforts. Other technology supporting collaboration includes webinars, audio and video conferencing, and even the use of iPads to enhance face-to-face communication. These can be toolsthat change the culture of a firm to be more open and communicative.

Reward systems should be put into place to reward team effort. Teamsshould be rewarded for their performance, and individuals should be rewarded for performance in a team. While middle-level managers build teams, coordinate their work, and monitor their performance, senior management should establish collaboration and teamwork as a vital function.

Despite the collaboration and best of intentions, BA teams sometimes fail. There are many reasons for this, but knowing some of the more common ones can help managers avoid them. Some of the more common reasons for team failure are presented in Table 4.3. They also represent issues that can cause a BA program to become unaligned and unproductive.

Reason for Failure	Descriptions
Lack of Communication	It is not enough to come up with valuable information for decision-making and to find business opportunities in data. That information must be shared with users, clients, and everyone within a firm for benefit to come from it. It is only when analytics show a tangible and beneficial outcome that they are considered business analytics (BA). If those results are not communicated on a continual basis, BA teams can be perceived to provide less value to the organization.
Failure to Deliver	Not every BA team will be able to deliver valued information if the team lacks the ability or resources to deliver needed answers and information. The greater the number of BA team failures, the greater are the chances that the team will be eliminated.
Lack of Justification	BA teams require resource allocations. Those allocations come from other departments that supposedly benefit from BA contributions. Without the role of BA and its potential contributions to a firm being clearly spelled out, users might not associate the ongoing efforts of a BA team as being worth the money spent on them.
Fail to Provide Value	BA teams have to sell their roles and suggested solutions or ideas. Without a clear understanding of value for potential users, the team faces a hard sell.
Inability to Prove Success	BA teams need to document and measure the impact of their ideas and suggestions. Without that proof, potential users might not support future BA efforts.

Source: Adapted from Flynn (2008) pp. 99–106 and Stubbs (2011) p. 89.

Table 4.3 Reasons for BA Team Failures*

Management Issues

Aligning organizational resources is a management function. There are general management issues that are related to a BA program, and some arespecifically important to operating a BA department, project, or team. The ones covered in this section include establishing an information policy, outsourcing business analytics, ensuring data quality, measuring business analytics contribution, and managing change.

Establishing an Information Policy

There is a need to manage information. This is accomplished by establishing an *information policy* to structure rules on how information and data are to be organized and maintained and who is allowed to view the dataor change it. The information policy specifies organizational rules for sharing, disseminating, acquiring, standardizing, classifying, and inventorying all types of information and data. It defines the specific procedures and accountabilities that identify which users and organizationalunits can share information, where the information can be distributed, and who is responsible for updating and maintaining the information.

In small firms, business owners might establish the information policy. For larger firms, *data* *administration* may be responsible for the specific policies and procedures for data management (Siegel and Shim, 2003, p. 280). Responsibilities could include developing the information policy, planning data collection and storage, overseeing database design, developing the data dictionary, as well as monitoring how information systems specialists and end user groups use data.

A more popular term for many of the activities of data administration is *data governance*, which includes establishing policies and processes for managing the availability, usability, integrity, and security of the data employed in businesses. It is specifically focused on promoting data privacy, data security, data quality, and compliance with government regulations.

Such information policy, data administration, and data governance must be in place to guard and ensure data is managed for the betterment of the entire organization. These steps are also important in the creation of database management systems (see Chapter 3, "What Resource Considerations Are Important to Business Analytics?") and their support ofBA tasks.

Outsourcing Business Analytics

Outsourcing can be defined as a strategy by which an organization chooses to allocate some business activities and responsibilities from an internal source to an external source (Schniederjans, et al., 2005, pp. 3–4). Outsourcing business operations is a strategy that an organization can use toimplement a BA program, run BA projects, and operate BA teams. Any business activity can be outsourced, including BA. Outsourcing is an important BA management activity that should be considered as a viable alternative in planning an investment in any BA program.

BA is a staff function that is easier to outsource than other line management tasks, such as running a warehouse. To determine if outsourcing is a useful option in BA programs, management needs to balance the advantages of outsourcing with its disadvantages. Some of the advantages of outsourcing BA include those listed in Table 4.4.

Advantage of Outsourcing BA	Description
Less Expensive	Maintaining a fully functioning BA department when analytics might only be useful periodically may be more expensive than occasionally hiring an outside consulting BA firm to solve a problem.
Superior Analytics	The pool of analytic capabilities is always going to be greater outside a firm.
More Staffing Flexibility	Staff positions are often the first cut in economy downturns. Using consultants is easier and less expensive than hiring full-time BA staff. Outsourcing permits more flexibility to add and reduce BA services as needed.
New Knowledge	Experienced BA consultants bring a variety of knowledge and experience from having worked with many other firms. That type of experience may be of great competitive advantage.

Table 4.4 Advantages of Outsourcing BA Nevertheless, there are disadvantages of outsourcing BA. Some of the disadvantages to outsourcing are presented in Table 4.5.

Disadvantages of Outsourcing BA	Description
Loss of Control	Once outsourced, most of the control of a BA project is under the control of the outsourcing firm. The client firm might lose not only control, but also opportunities for new and unique information, which the outsourcing firm may not share with the client.
Difficulties in Managing the Relationship	Client firms may find it difficult to communicate with the outsourcing firm because of distance, differing culture, language issues, and more. The lack of management could cause substantial problems with customer service and product quality.
Weakens Innovation	Having outsourced a client firm's internal experts, the remaining collaboration within the firm's personnel is reduced, and that reduces the opportunity for innovation efforts through shared collaboration.
Risk of Information	Outsourcing staff are exposed to client proprietary information, including innovations in analytics. This information could be shared with other competing firms, placing the client firm at risk.
Worthless Analytics	Sometimes outsourcing partners are less capable than internal analysts, wasting time and money.

Table 4.5 Disadvantages of Outsourcing BA

Managing outsourcing of BA does not have to involve the entire department. Most firms outsource projects or tasks found to be too costly to assign internally. For example, firms outsource cloud computing services tooutside vendors (Laudon and Laudon, 2012, p. 511), and other firms outsource software development or maintenance of legacy programs to offshore firms in low-wage areas of the world to cut costs (Laudon and Laudon, 2012, p. 192).

Outsourcing BA can also be used as a strategy to bring BA into an organization (Schniederjans, et al., 2005, pp. 24–27). Initially, to learn how to operate a BA program, project, or team, an outsource firm can be hired for a limited, contracted time period. The client firm can then learn from theoutsourcing firm's experience and instruction. Once the outsourcing contract is over, the client firm can form its own BA department, project, orteam.

Ensuring Data Quality

Business analytics, if relevant, is based on data assumed to be of high quality. *Data quality* refers to accuracy, precision, and completeness of data. High-quality data is considered to correctly reflect the real world in which it is extracted. Poor quality data caused by data entry errors, poorly maintained databases, out-of-date data, and incomplete data usually leads tobad decisions and undermines BA within a firm. Organizationally, the database management systems (DBMS, mentioned in Chapter 3) personnel are managerially responsible for ensuring data quality. Because of its importance and the possible location of the BA department outside of the management information systems department (which usually hosts the DBMS), it is imperative that whoever leads the BA program should seek to ensure data quality efforts are undertaken.

Ideally, a properly designed database with organization-wide data standards and efforts taken to avoid duplication or inconsistent date elements should have high-quality data. Unfortunately, times are changing, and more organizations allow customers and suppliers to enter data into databases via the Web directly. As a result, most of the quality problems originate from data input such as misspelled names, transposed numbers, or incorrect or missing codes.

An organization needs to identify and correct faulty data and establish routines and procedures for editing data in the database. The analysis of data quality can begin with a *data quality audit*, where a structured survey or inspection of accuracy and level of completeness of data is undertaken. This audit may be of the entire database, just a sample of files, or a survey of end users for perceptions of the data quality. If during the data quality audit files are found that have errors, a process called *data cleansing* or *data scrubbing* is undertaken to eliminateor repair data. Some of the areas in a data file that should be inspected in the audit and suggestions on how tocorrect them are presented in Table 4.6.

Data Inspection Items	Description and Cleansing/Scrubbing Recommendation
Current Data	Check to make sure the data is current. If it is out of date, remove it.
Completeness	Check to see if there is missing data. If more than 50% is missing, remove the entire file from the database.
Relevance	Check to see if the data is no longer relevant for the purpose for which it was collected. If it's no longer relevant, consider removing it from the database.
Duplication	Check to see if duplicate data files exist in the database. Remove duplicate data.
Outliers	Check for extreme values (outliers) in quantitative data files for possible errors in data coding. Remove from the data file any suspected of being in error, or repair the data.
Inconsistent Values	If data fields contain both characters and real numbers data where only characters or numbers should be, explore repairing the data.
Coding	If suspicious or unknown coding of data exists in data files, remove from the database or repair the coding of data.

Table 4.6 Quality Data Inspection Items and Recommendations

Measuring Business Analytics Contribution

The investment in BA must continually be justified by communicating the BA contribution to the organization for ongoing projects. This means that performance analytics should be computed for every BA project and BA team initiative. These analytics should provide an estimate of the tangible and intangible values being delivered to the organization. This should also involve establishing a communication strategy to promote thevalue being estimated.

Measuring the value and contributions that BA brings to an organization is essential to helping the firm understand why the application of BA is worth the investment. Some BA contribution estimates can be computed using standard financial methods, such as *payback period* (how long it takesfor the initial costs are returned by profit) or return on investment (ROI) (see Schniederjans, et al., 2010, pp. 90–132), where dollar values or quantitative analysis is possible. When intangible contributions are a major part of the contribution being delivered to the firm, other methods like cost/benefit analysis (see Schniederjans, et al., 2010, pp. 143–158), which include intangible benefits, should be used.

The continued measurement of value that BA brings to a firm is not meant to be self- serving, but it aids the organization in aligning efforts to solve problems and find new business opportunities. By continually running BA initiatives, a firm is more likely to identify internal activities that should and can be enhanced by employing optimization methodologies during the Prescriptive step ofthe BA process introduced in Chapter 1, "What Are Business Analytics?" It can also helpidentify underperforming assets. In addition, keeping track of investment payoffs for BAinitiatives can identifyareas in the organization that should have a higher priority for analysis. Indeed, past applications and allocations of BA resources that have shown significant contributions can justify priorities established by the BA leadership about where there should be allocated analysis efforts within the firm. They can also help acquire increases in data support, staff hiring, andfurther investments in BA technology.

Managing Change

Wells (2000) found that what is critical in changing organizations is organizational culture and the use of change management. *Organizational culture* is how an organization supports cooperation, coordination, and empowerment of employees (Schermerhorn 2001, p. 38). *Change management* is defined as an approach for transitioning the organization (individuals, teams, projects, departments) to a changed and desired future state (Laudon and Laudon, 2012,pp. 540–542). Change management is a means of implementing change in an organization, such as adding a BA department (Schermerhorn 2001, pp. 382–390). Changes in an organization can be either planned (a result of specific and planned efforts at change with direction by a change leader) or unplanned (spontaneous changes without direction of a changeleader). The application of BA invariably will result inboth types of changes because of BA's specific problem-solving role (a desired, planned change to solve a problem) and opportunity finding exploratory nature (i.e., unplanned new knowledge opportunity changes) of BA. Change management can also target almost everything that makes up an organization (see Table 4.7).

Change Target	Description
Culture	This represents the changing values and norms of the individuals and groups that make up the organization. BA has to sell itself in some situations, build trust, and alter decision-making. It often requires a different culture of thinking about decision-making.
Organization Structure	This is the changing organizational lines of authority and communication. The cross-departmental nature of BA positions may provide information that changes the organization and alters relationships and tasks.
Personnel	BA information about the need for human resource changes in attitudes and skills can mandate changes that permit an organization to achieve higher business performance levels.
Tasks	BA analysis might find that some job designs, specifications, and descriptions that employees perform need to have their objectives and goals changed to achieve higher business performance levels.
Technology	BA analysis might find information system technology used in the design and workflow that integrate employees and equipment into operating systems and require change to achieve higher business performance levels.

**Source*: Adapted from Figure 7 in Schniederjans and Cao (2002), pp. 261.

Table 4.7 Change Management Targets*

It is not possible to gain the benefits of BA without change. The intent is change that involves finding new and unique information on which change should take place in people, technology systems, or business conduct. By instituting the concept of change management within an organization, a firm can align resources and processes to more readily accept changes that BA may suggest. Instituting the concept of change management in any firm depends on the unique characteristics of that firm. There are, though, a number of activities in common with successful change management programs, and they apply equally to changes in BA departments, projects, or teams. Some of these activities that lead to change management success are presented as best practices in Table 4.8.

Best Practice	Description
Champion	Change is scary business for some, and a strong leader for change can champion the change effort, calming fears and explaining the need for change. The champion also helps direct efforts, motivate change, and keep the change activities on track.
Clearly Stated Goals	Any type of change should be clearly defined, including what the changes are, which personnel have to change, and what the processes involve and how they affect technology. This would also include deadlines needed to keep the change effort on track.
Good Communication	To avoid resistance to change (a natural norm to anything that is new), it is useful to help those facing the change understand its value through effective and repeated communications, keeping them informed on progress and easing fears.
Measured Performance	Any goals stated prior to the launch of change can be used to measure performance during the changeover period. Seeing business performance improve with changes can motivate further change and support by those impacted.
Senior Management Support	Critical to all BA departments, projects, or teams is the need for senior management to support change efforts. Sometimes that support is in direct dollars, and sometimes it's in lending authority to get resources needed for BA work.

Discussion Questions

1. The literature in management information systems consistently suggests that a decentralized approach to resource allocation is the most efficient. Why then do you think the literature in BA suggests that the opposite—a centralized organization—is the best structure?
2. Why is collaboration important to BA?
3. Why is organization culture important to BA?
4. How does establishing an information policy affect BA?
5. Under what circumstances is outsourcing BA good for thedevelopment ofBA in an organization?
6. Why do we have to measure BA contributions to an organization?
7. How does data quality affect BA?
8. What role does change management play in BA?

UNIT III DESCRIPTIVE ANALYTICS

Introduction to Descriptive analytics - Visualizing and Exploring Data - Descriptive Statistics - Sampling and Estimation - Probability Distribution for Descriptive Analytics - Analysis of Descriptive analytics

What Are Descriptive

Analytics? Introduction

In any BA undertaking, referred to as *BA initiatives* or *projects*, a set of objectives is articulated. These objectives are a means to align the BA activities to support strategic goals. The objectives might be to seek out andfind new business opportunities, to solve operational problems the firm is experiencing, or to grow the organization. It is from the objectives that exploration via BA originates and is in part guided. The directives that come down, from the strategic planners in an organization to the BA department or analyst, focus the tactical effort of the BA initiative or project. Maybe the assignment will be one of exploring internal marketing data for a new marketing product. Maybe the BA assignment will be focused on enhancing service quality by collecting engineering and customer service information. Regardless of the type of BA assignment, the first step is one of exploring data and revealing new, unique, and relevant information to help the organization advance its goals. Doing this requires an exploration of data.

This chapter focuses on how to undertake the first step in the BA process: *descriptive analytics*. The focus in this chapter is to acquaint readers with more common descriptive analytic tools used in this step and available in SPSS and Excel software. The treatment here is not computational but informational regarding the use and meanings of these analytic tools in support of BA. For purposes of illustration, we will use the data set in

Figure 5.1 representing four different types of product sales (Sales 1, Sales2, Sales 3, and Sales

	A	B	C	D
1	Sales 1	Sales 2	Sales 3	Sales 4
2	23	1234	1	1
3	31	943	2	5
4	48	896	3	9
5	16	12	4	12
6	28	15	5	18
7	29	15	6	19
8	31	23	6	19
9	35	21	6	21
10	51	25	6	21
11	42	27	7	21
12	34	27	8	21
13	56	29	9	21
14	24	20	10	21
15	34	18	11	19
16	43	13	12	19
17	56	8	13	18
18	34	7	14	12
19	38	6	15	9
20	23	4	16	5
21	27	1	17	1

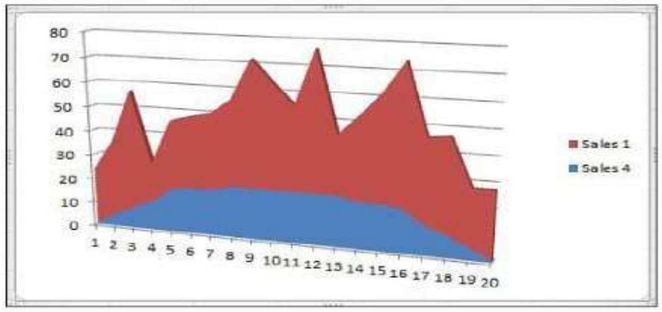
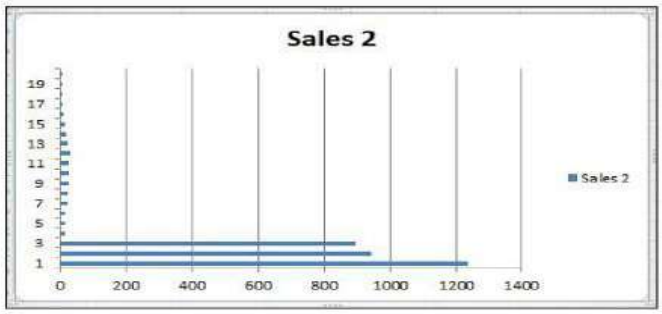
Figure 5.1 Illustrative sales data sets

Visualizing and Exploring Data

There is no single best way to explore a data set, but some way of conceptualizing what the data set looks like is needed for this step of the BA process. Charting is often employed to visualize what the data mightreveal.

When determining the software options to generate charts in SPSS or Excel, consider that each software can draft a variety of charts for the selected variables in the data sets. Using the data in Table 5.1, charts can be created for the illustrative sales data sets. Some of these charts are discussed in Table 5.1 as a set of exploratory tools that are helpful in understanding the informational value of data sets. The chart to select depends on the objectives set for the chart.

Statistics	Computation (in Data Set)	Application Area	Example	Application Notes
N or Count	Number of values.	Any.	Sample size of a company's transactions during a month.	Useful in knowing how many items were used in the statistics computations.
Sum	Total of the values in the entire data set.	Any.	Total sales for a company.	Useful in knowing the total value.
Mean	Average of all values.	Any.	Average sales per month.	Useful in capturing the central tendency of the data set.
Median	Midpoint value in the data set arranged from high to low.	Finding the midpoint in the distribution of data.	Total income for citizens of a country.	Useful in finding the point where 50 percent of the data is above and below.
Mode	Most common value in the data set.	Where values are highly repeated in the data set.	Fixed annual salaries where a limited number of wage levels are used.	Useful in declaring a common value in highly repetitive data sets.
Maximum/ Minimum	Largest and smallest values, respectively.	To conceptualize the spread of the data's distribution.	Largest and smallest sales in a day.	Useful in providing a scope or end points in the data.

Type of Chart	Application Notes	Chart Example
Area	• Overlay more than one variable at a time. • Ideal for contrasting two variables. • Example: Overlaying different product sales to show improvement. • Note also the 3D effect, which is possible with most of the charts listed in this table.	
Bar	• Can be horizontal, vertical, cone, or cyclically shaped and multidimensional with overlaying variables. • Ideal for showing comparative improvement over time. • Example: Bars showing productivity of one person versus another.	

The charts presented in Table 5.1 reveal interesting facts. The area chart is able to clearly contrast the magnitude of the values in the two variable data sets (Sales 1 and Sales 4). The column chart is useful in revealing the almost perfect linear trend in the Sales 3 data, whereas the scatter chart reveals an almost perfect nonlinear function in Sales 4 data. Additionally, the cluttered pie chart with 20 different percentages illustrates that all chartscan or should be used in some situations. The best practices suggest charting should be viewed as an exploratory activity of BA. BA analysts should run a variety of charts and see which ones reveal interesting and useful information. Those charts can be further refined to drill down to more detailed information and more appropriate charts related to the objectives of the BA initiative.

Of course, a cursory review of the Sales 4 data in Figure 5.1 makes the concave appearance of the data in the scatter chart in Table 5.1 unnecessary.But most BA problems involve big data—so large as to make it impossible to just view it and make judgment calls on structure or appearance. This is why descriptive statistics can be employed to view the data in a parameter-based way in the hopes of better understanding the information that the datahas to reveal.

Column	• Same as a bar chart. • Note how this chart clearly reveals a positive trend upward.	
Line	• Ideal for showing linear trend and other linear or non-linear appearance. • Best applied with time-series data with time as the X-axis.	
Pie	• Useful in conceptualizing proportions. • Various other versions, like the donut chart (with a hollow center), can also be used. • Useful in situations where the number of variables is limited (not like the illustration to the right).	

The charts presented in Table 5.1 reveal interesting facts. The area chart is able to clearly contrast the magnitude of the values in the two variable data sets (Sales 1 and Sales 4). The column chart is useful in revealing the almost perfect linear trend in the Sales 3 data, whereas the scatter chart reveals an almost perfect nonlinear function in Sales 4 data. Additionally, the cluttered pie chart with 20 different percentages illustrates that all chartscan or should be used in some situations. The best practices suggest charting should be viewed as an exploratory activity of BA. BA analysts should run a variety of charts and see which ones reveal interesting and useful information. Those charts can be further refined to drill down to more detailed information and more appropriate charts related to the objectives of the BA initiative. Of course, a cursory review of the Sales 4 data in Figure 5.1 makes the concave appearance of the data in the scatter chart in Table 5.1 unnecessary.But most BA problems involve big data— so large as to make it impossible to just view it and make judgment calls on structure or appearance. This is why descriptive statistics can be employed to view the data in a parameter- based way in the hopes of better understanding the information that the datahas to reveal.

Descriptive Statistics

When selecting the option of *descriptive statistics* in SPSS or Excel, a number of useful statistics are automatically computed for the variables in the data sets. Some of these descriptive statistics are discussed in Table 5.2 as exploratory tools that are helpful in understanding the informational value of data sets.

Range	Difference between the max and min values.	A crude estimate of the spread of the data's distribution.	Spread of sales in units during a month.	Useful as a simple estimate of dispersion.
Standard deviation	Square root of the average of the differences squared between the mean and all other values in the data set.	A precise estimate of the spread of the data's distribution from a mean value in terms of the units used in its computation.	Standard deviation in dollars from mean sales.	The smaller the value, the less the variation and the more predictable using the data set.
Variance	Average differences squared between the mean and all other values.	A variance estimate of the spread of the data's distribution from a mean value, not in terms of the units used in its computation.	Measure of variance that is best used when compared with another variance computed on the same data set.	The smaller the value, the less the variation and the more predictable the data set.
(Coefficient of) Skewedness	Positive or negative values. If value sign is +, distribution is positively skewed; if -, it is negatively skewed. The larger the value, the greater it is skewed.	Measure of the degree of asymmetry of data about a mean.	As the age of residents in a country becomes older, the population age distribution becomes more negatively skewed.	The closer the value is to 0, the better the symmetry. A positively skewed distribution has its largest allocation to the left, and a negative distribution to the right.

	N	Range	Min	Max	Sum	Mean		Std. Dev.	Variance	Skewedness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Sales 1	20	40	16	56	703	35.15	2.504	11.198	125.397	.490	.512	–.429	.992
Sales 2	20	1233	1	1234	3344	167.20	83.686	374.254	140065.853	2.241	.512	3.636	.992
Sales 3	20	16	1	17	171	8.55	1.065	4.763	22.682	.272	.512	–.988	.992
Sales 4	20	20	1	21	292	14.60	1.603	7.170	51.411	-.824	.512	–.825	.992
Valid N	20												

	A	B	C	D	E
1	*Statistics*	*Sales1*	*Sales2*	*Sales3*	*Sales4*
2					
3	Mean	35.15	167.2	8.55	14.6
4	Standard Error	2.503970531	83.68567758	1.064931428	1.603286099
5	Median	34	19	7.5	18.5
6	Mode	34	15	6	21
7	Standard Deviation	11.19809664	374.2537276	4.762518131	7.17011341
8	Sample Variance	125.3973684	140065.8526	22.68157895	51.41052632
9	Kurtosis	-0.429071744	3.636015862	-0.988175871	-0.825197874
10	Skewness	0.490453685	2.240917321	0.27209513	-0.824089881
11	Range	40	1233	16	20
12	Minimum	16	1	1	1
13	Maximum	56	1234	17	21
14	Sum	703	3344	171	292
15	Count	20	20	20	20

Fortunately, we do not need to compute these statistics to know how to use them. Computer software provides these descriptive statistics where they're needed or requested. The SPSS descriptive statistics for the illustrative sales data sets are presented in Table 5.3, and the Excel descriptive statistics are presented in Table 5.4.

Table 5.3 SPSS Descriptive Statistics

Table 5.4 Excel Descriptive Statistics

	A	B	C	D	E
1	*Statistics*	*Sales1*	*Sales2*	*Sales3*	*Sales4*
2					
3	Mean	35.15	167.2	8.55	14.6
4	Standard Error	2.503970531	83.68567758	1.064931428	1.603286099
5	Median	34	19	7.5	18.5
6	Mode	34	15	6	21
7	Standard Deviation	11.19809664	374.2537276	4.762518131	7.17011341
8	Sample Variance	125.3973684	140065.8526	22.68157895	51.41052632
9	Kurtosis	-0.429071744	3.636015862	-0.988175871	-0.825197874
10	Skewness	0.490453685	2.240917321	0.27209513	-0.824089881
11	Range	40	1233	16	20
12	Minimum	16	1	1	1
13	Maximum	56	1234	17	21
14	Sum	703	3344	171	292
15	Count	20	20	20	20

Looking at the data sets for the four variables in Figure 5.1 and at the statistics in Tables 5.3 and 5.4, there are some obvious conclusions based onthe detailed statistics from the data sets.

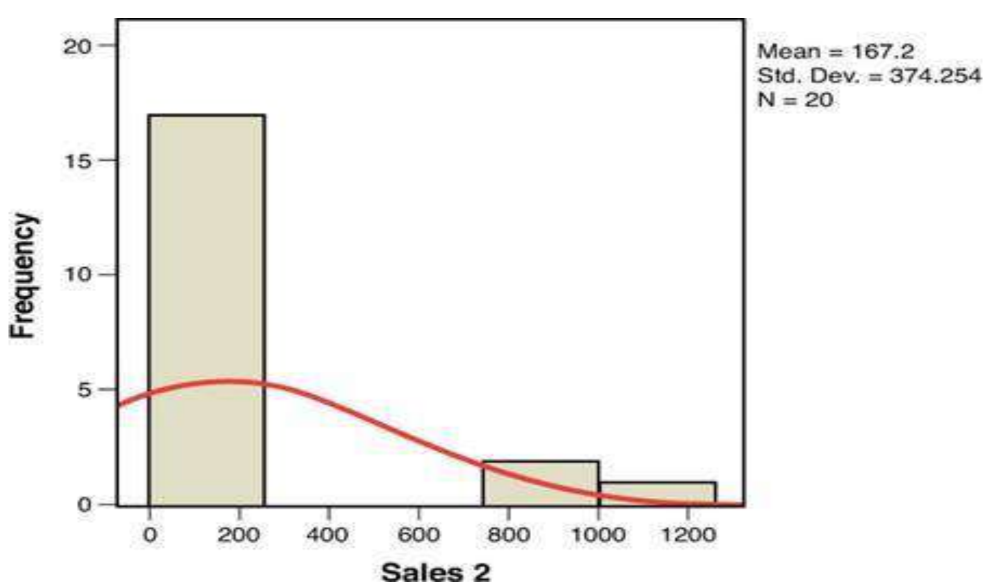
It should be no surprise that Sales 2, with a few of the largest values and mostly smaller ones making up the data set, would have the largest variance statistics (standard deviation, sample variance, range, maximum/minimum). Also, Sales 2 is highly, positively skewed (Skewedness > 1) and highly peaked (Kurtosis >3). Note the similarity of the mean, median, and mode in Sales 1 and the dissimilarity in Sales 2. These descriptive statistics provide a more precise basis to envision the behavior of the data. Referred to as *measures of central tendency,* the mean, median, and mode can also be used to clearly define the direction of a skewed distribution. A negatively skewed distribution orders these measures such that mean < median < mode, and a positive skewed distribution orders them such that mode < median < mean.

So what can be learned from these statistics? There are many observations that could be drawn from this data. Keep in mind that, in dealing with the big data sets, one would only have the charts and the statistics to serve as a guide in determining what the data looks like. Yet, from these statistics, one can begin describing the data set. So in the case of Sales 2, it can be predicted that the data set is positively skewed and peaked. Note in Figure 5.2 that the histogram of Sales 2 is presented. The SPSS chart also overlays a normal distribution (a bell-shaped curve) to reflect the positioning of the mean (highest point on the curve, 167.2) and how the data appears to fit the normal distribution (not very well in this situation). As expected, the distribution is positively distributed with a substantial variance between the large values in the data set and the many more smaller valued data points.

Table 5.2 Descriptive Statistics Useful in BA

(Coefficient of) Kurtosis	Value where less than 3 means a flat distribution and more than 3 means a peaked distribution.	Measure of the degree of spread vertically in a distribution about a mean. Also, it reveals a positive and a negative symmetry depending on its sign.	Distribution of customers at lunch and dinner times peaks and then flattens out.	The closer the value is to 2, the less is the kurtosis (peaking or flattening in the distribution).
Standard Error (of the Mean)	Mean of the sample standard deviation (that is, a standard deviation adjusted to reflect a sample size).	Standard deviation of a sampling distribution.	Standard deviation in dollars from mean sales based on a sample.	The smaller the value, the less the variation and the more predictable the sample data set.
Sample Variance	Same as variance, but adjusted for sample sizes.	Variance estimate of the spread of the sampling data distribution.	Measure of variance when sampling is used for collection purposes.	The smaller the value, the less the variation and the more predictable the sample data set.

Figure 5.2 SPSS histogram of Sales 2 data



We also know that substantial variance in the data points making up the data set is highly diverse—so much so that it would be difficult to use this variable to predict future behavior or trends. This type of information maybe useful in the further steps of the BA process as a means of weeding out data that will not help predict anything useful. Therefore, it would not helpan organization improve its operations.

Sometimes big data files become so large they cannot be manipulated bycertain statistical software systems. In these instances, a smaller but representative sample of the data can be obtained if necessary. Obtaining the sample for accurate prediction of business behavior requires understanding the sampling process and estimation from that process.

Sampling and Estimation

The estimation of most business analytics requires sample data. In thissection we discuss various types of sampling methods and follow that upwith a discussion on how the samples are used in sampling estimation.

Sampling Methods

Sampling is an important strategy of handling large data. If data files are too big to be run by software or just too large to work with, the number of items in the data file can be sampled to provide a new data file that seeks to accurately represent the population from which it comes. In sampling data, there are three components that should be recognized: a population, a sample, and a *sample element* (the items that make up the sample). A firm's collection of customer service performance documents for one year could be designated as a population of customer service performance for that year.From that population, a sample of a lesser number of sample elements (the individual customer service documents) can be drawn to reduce the effort ofworking with the larger data. Several sampling methods can be used to arrive at a representative sample. Some of these sampling methods are presented in Table 5.5.

Sampling Method	Description	Application	Application Notes
Simple Random	Allows each sample element in a population to have an equal chance of selection.	Selecting customers based on their percentage of occurrence as a member of a particular race.	Sample size must be sufficient to avoid sampling bias.
Systematic Random (or Period)	Selects sample elements from a population based on a fixed number in an interval.	Selecting every fifth person leaving an airport to interview.	Assumes the sample elements order in the interval is presented in a random fashion; otherwise, it can result in sampling bias.
Stratified Random	Stage 1: Divide a population into groups (called strata); Stage 2: Apply simple random sampling.	Randomly selecting an equal number of people in each of three different economic strata.	Strata must be representative of the population, or it can result in sampling bias.
Cluster Random	Stage 1: Group sample elements geographically (called clusters); Stage 2: Apply simple random sampling.	Randomly selecting an equal number of people from voting districts.	Cluster must be representative of the population, or it can result in sampling bias.
Quota	Based on a fixed quota or number of sample elements.	Selecting the first 200 people who enter a store.	• Mainly used to save time and money. • Sample size must be sufficient to avoid sampling bias.
Judgment	Selects sample elements based on expert judgment.	Selecting candidates for an interview with a special offer based on their appearance.	Prone to bias without defined criteria for selection because of dependency on interviewer experience.

Table 5.5 Sampling Methods

The simple, systematic, stratified, and cluster random methods are based on some kind of probability of their occurrence in a population. Quota and judgment methods are non-probability-based tools. Although the randomization process in some methods helps ensure representative samples being drawn from the population, sometimes because of cost or time constraints, non-

probability methods are the best choice for sampling.

Which sampling method should be selected for a particular BA analysis?

It depends on the nature of the sample. As mentioned in the application notes in Table 5.5, the size of the population, the size of the sample, the areaof application (geography, strata, ordering of the data, and so on), and even the researchers running the data collection effort impact the particular methodology selected. A best practices approach might begin with a determination of any constraints (time allowed and costs) that might limit the selection of a sample collection effort. That may narrow the choice to something like a quota method. Another best practices recommendation is to start with the objective(s) of the BA project and use them as a guide in the selection of the sampling method. For example, suppose the objective of a BA analysis is to increase sales of a particular product. This might lead to random sampling of customers or even a stratified sample by income levels, if income is important to the end results of the analysis. Fortunately, there issoftware to make the data collection process easier and less expensive.

Data file software can be used with the methods mentioned earlier to quickly collect a sample. For example, SPSS permits simple, systematic, stratified, and cluster random methods, among others. Using this software requires a designation of the number of sample elements in each stratum (for example, selected 2 for each stratum in this example). In Table 5.6, SPSS has defined seven strata for the Sales 4 data. The logic of this stratification can be observed by looking at the Sales 4 data in Figure 5.1. The additional SPSS printout in Figure 5.3 shows the specific sample elements that were randomly selected in each stratum, as well as totals and their percentages in the resulting sample. For example, only 0.33, or 33 percent, of the “21” strata sample elements were randomly selected by the SPSS program.

Summary for Stage 1

		Number of Units Sampled		Proportion of Units Sampled	
		Requested	Actual	Requested	Actual
Sales4 =	1	2	2	100.0%	100.0%
	5	2	2	100.0%	100.0%
	9	2	2	100.0%	100.0%
	12	2	2	100.0%	100.0%
	18	2	2	100.0%	100.0%
	19	2	2	50.0%	50.0%
	21	2	2	33.3%	33.3%

Table 5.6 SPSS Stratifications of Sample for Sales 4 Variable

Sales4	InclusionProbability _1_	SampleWeight Cumulative_1_	PopulationSize _1_	SampleSize _1_
1	1.00	1.00	2	2
5	1.00	1.00	2	2
9	1.00	1.00	2	2
12	1.00	1.00	2	2
18	1.00	1.00	2	2
19	.	.	.	.
19	.50	2.00	4	2
21	.	.	.	.
21	.	.	.	.
21	.33	3.00	6	2
21	.	.	.	.
21	.33	3.00	6	2
21	.	.	.	.
19	.50	2.00	4	2
19	.	.	.	.
18	1.00	1.00	2	2
12	1.00	1.00	2	2
9	1.00	1.00	2	2
5	1.00	1.00	2	2
1	1.00	1.00	2	2

Figure 5.3 SPSS stratified sample of Sales 4 variable identified Excel also permits simple random and periodic sampling. For example, Figure 5.4 shows the Excel input and printout results for the Sales 4 data. Inthis example, a random sample of 5 values is requested from the sample elements of 20. The resulting 5 sample elements that were randomly selected are presented in the lower-right side of Figure 5.4.

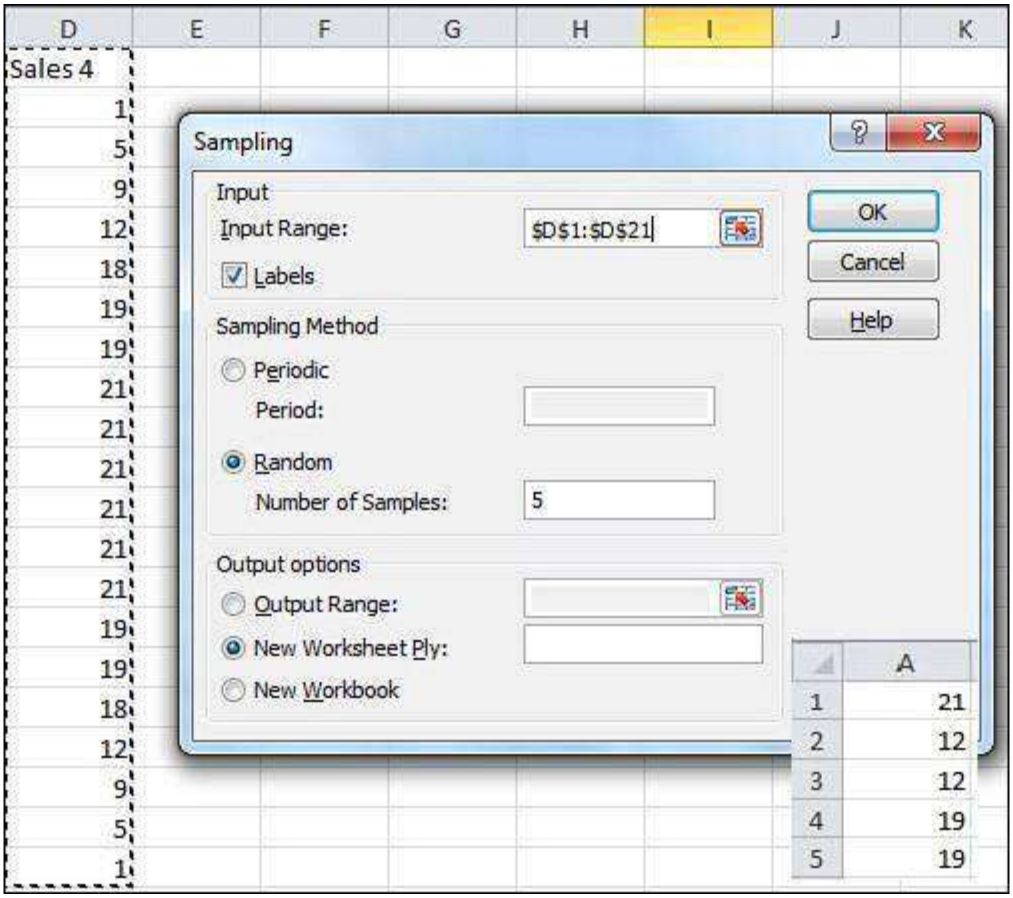


Figure 5.4 Excel random sample of Sales 4 variable

Sampling Estimation

Invariably, using any sampling method can cause errors in the sample results. Most of the statistical methods listed in Table 5.2 are formulated forpopulation statistics. Once sampling is introduced into any statistical analysis, the data must be treated as a sample and not as a population. Manystatistical techniques, such as standard error of mean and sample variance, incorporate mathematical correction factors to adjust descriptive analysis statistical tools to compensate for the possibility of sampling error.

One of the methods of compensating for error is to show some degree of confidence in any sampling statistic. The confidence in the sample statisticsused can be expressed in a *confidence interval*, which is an interval estimate about the sample statistics. In general, we can express this interval estimate as follows:

Confidence interval = (sample statistic) ± [(confidence coefficient) ×(standard error of the estimate)]

The *sample statistic* in the confidence interval can be any measure orproportion from a sample that is to be used to estimate a population parameter, such as a measure of central tendency like a

mean. The *confidence coefficient* is set as a percentage to define the degree of confidence to accurately identify the correct sample statistic. The larger the confidence coefficient, the more likely the population mean from the sample will fall within the confidence interval. Many software systems set a95 percent *confidence level* as the default confidence coefficient, although any percentage can be used. Both SPSS and Excel permit the user to enter adesired percentage. The *standard error of the estimate* in the preceding expression can be any statistical estimate, including proportions used to estimate a population parameter. For example, using a mean as the sample statistic, we have the following interval estimate expression:

Confidence interval = mean ± [(95 percent) × (standard error of themean)]

The output of this expression consists of two values that form high and low values defining the confidence interval. The interpretation of this interval is that the true population mean represented by the sample has a 95 percent chance of falling in the interval. In this way, there is still a 5 percent chance that the true population mean will not fall in the interval due to sampling error. Because the standard error of the mean is based on variationstatistics (standard deviation), the larger the variance statistics used in this expression, the wider the confidence interval and the less precise the sample mean value, which results in a good estimate for the true population mean.

Both SPSS and Excel compute confidence intervals when analyzing various statistical measures and tests. For example, the SPSS printout in Table 5.7 is of the 95 percent confidence interval for the Sales 1 variable. With a sample mean value of 35.15, the confidence interval suggests there is a 95 percent chance that the true population mean falls between 29.91 and 40.39. When trying to ascertain if the sample is of any value, this kind of information can be of great significance. For example, knowing with 95 percent certainty there is at least a mean of 29.91 might make the differencebetween continuing to sell a product or not because of a needed requirementfor a breakeven point in sales.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean	95 Percent Confidence Interval of the Difference	
					Lower	Upper
Sales 1	20	35.15	11.198	2.504	29.91	40.39

Table 5.7 SPSS 95 Percent Confidence Intervals for Sales 1 Variable Confidence intervals are also important for demonstrating the accuracy of some forecasting models. For example,

confidence intervals can be createdabout a regression equation model forecast to see how far off the estimatesmight be if the model is used to predict future sales. For additional discussion on confidence intervals, see Appendix A, “Statistical Tools.”

Introduction to Probability Distributions

By taking samples, one seeks to reveal population information. Once a sample is taken on which to base a forecast or a decision, there is the possibility that it may not accurately capture the population information. No single sample can assure an analyst that the true population information has been captured. Confidence interval statistics are employed to reflect the possibility of error from the true population information.

To utilize the confidence interval formula expressed in Section 5.4, a confidence coefficient percentage (95 percent) is set as a way to express thepossibility that the sample statistics used to represent the population statistics may have a potential for error. The confidence coefficient used in the confidence interval is usually referred to as a *Z value*. It is spatially related to the area (expressed as a percentage or frequency) representing theprobability under the curve of a distribution. The *sample standard normal distribution* is the bell-shaped curve illustrated in Figure 5.5. This distribution shows the relationship of the Z value to the area under the curve. The Z value is the number of standard error of the means.

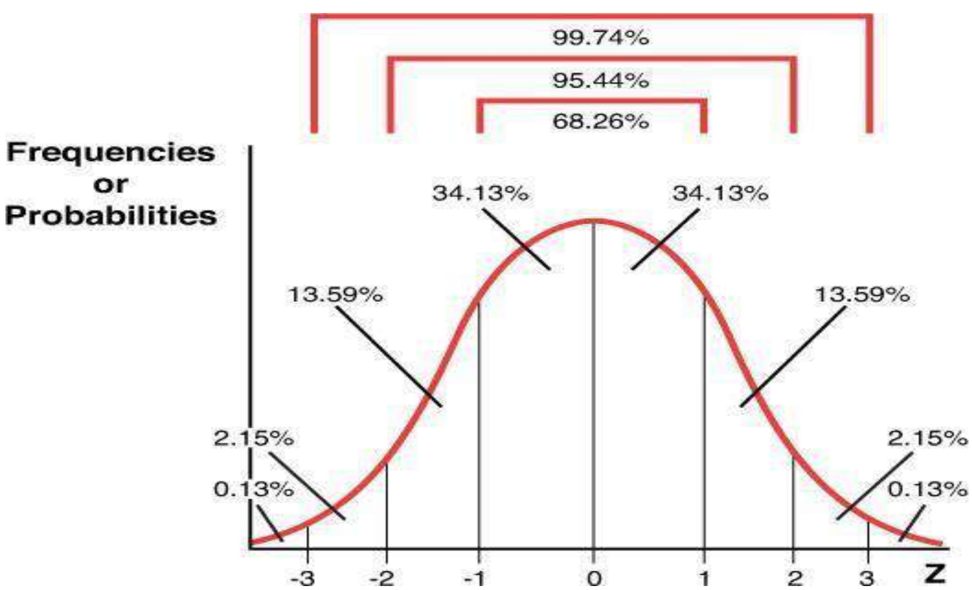


Figure 5.5 Standard normal probability distribution

The *confidence coefficient* is related to the Z values, which divide the area under a normal curve into probabilities. Based on the *central limit theorem*, we assume that all sampling distributions of sufficient size are normally distributed with a standard deviation equal to the standard error ofthe estimate. This means that an interval of plus or minus two standard errors of the estimate (whatever the estimate is) has a 95.44 percent chance of containing the true or actual population parameter. Plus or minus three standard errors of the estimate has a 99.74 percent chance of containing the true or actual population parameter. So the Z value represents the number of standard errors of the estimate. Table 5.8 has selected Z values for specific confidence levels representing the probability that the true population parameter is within the confidence interval and represents the percentage ofarea under the curve in that interval.

Confidence Level	Related Z-Value
0.60	0.253
0.70	0.524
0.80	0.842
0.90	1.282
0.95	1.645
0.99	2.327
0.999	3.080

Table 5.8 Selected Z-values and Confidence Levels

The important BA use of the probability distributions and confidence intervals is that they suggest an assumed parameter based on a sample that has properties that will allow analysts to predict or forecast with some assessed degree of statistical accuracy. In other words, BA analysts can, with some designated confidence level, use samples from large databases to accurately predict population parameters.

Another important value to probability distributions is that they can be used to compute probabilities that certain outcomes like success with business performance may occur. In the exploratory descriptive analytics step of the BA process, assessing the probabilities of some events occurring can be a useful strategy to guide subsequent steps in an analysis. Indeed, probability information may be very useful in weighing the choices an analyst faces in any of the steps of the BA process. Suppose, for example, the statistics from the Sales 1 variable in

Table 5.8 are treated as a sample todiscover the probability of sales greater than one standard error of the mean above the current mean of 35.15. In Figure 5.6, the mean (35.15) and standard error of the mean (2.504) statistics are included at the bottom of the standard sampling normal distribution. By adding one standard error of the mean to the sample mean, the resulting value is 37.654. The sum of the area (the shaded region in Figure 5.6) representing the total probability beyond 37.654 is a probability of 15.87 (13.59+2.15+0.13). So there is onlya 15.87 percent probability that sales will exceed 37.654 based on the sample information for the Sales 1 variable.

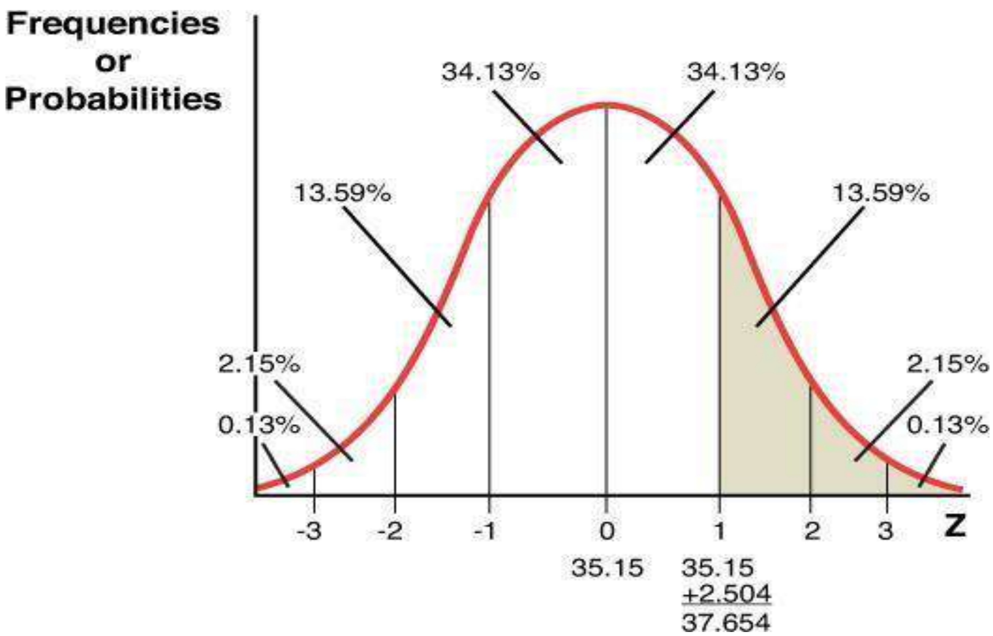


Figure 5.6 Probability function example

The ability to assess probabilities using this approach is applicable to other types of probability distributions. For a review of probability conceptsand distributions, probability terminology, and probability applications, see Appendix A.

Marketing/Planning Case Study Example: DescriptiveAnalytics Step in the BA Process

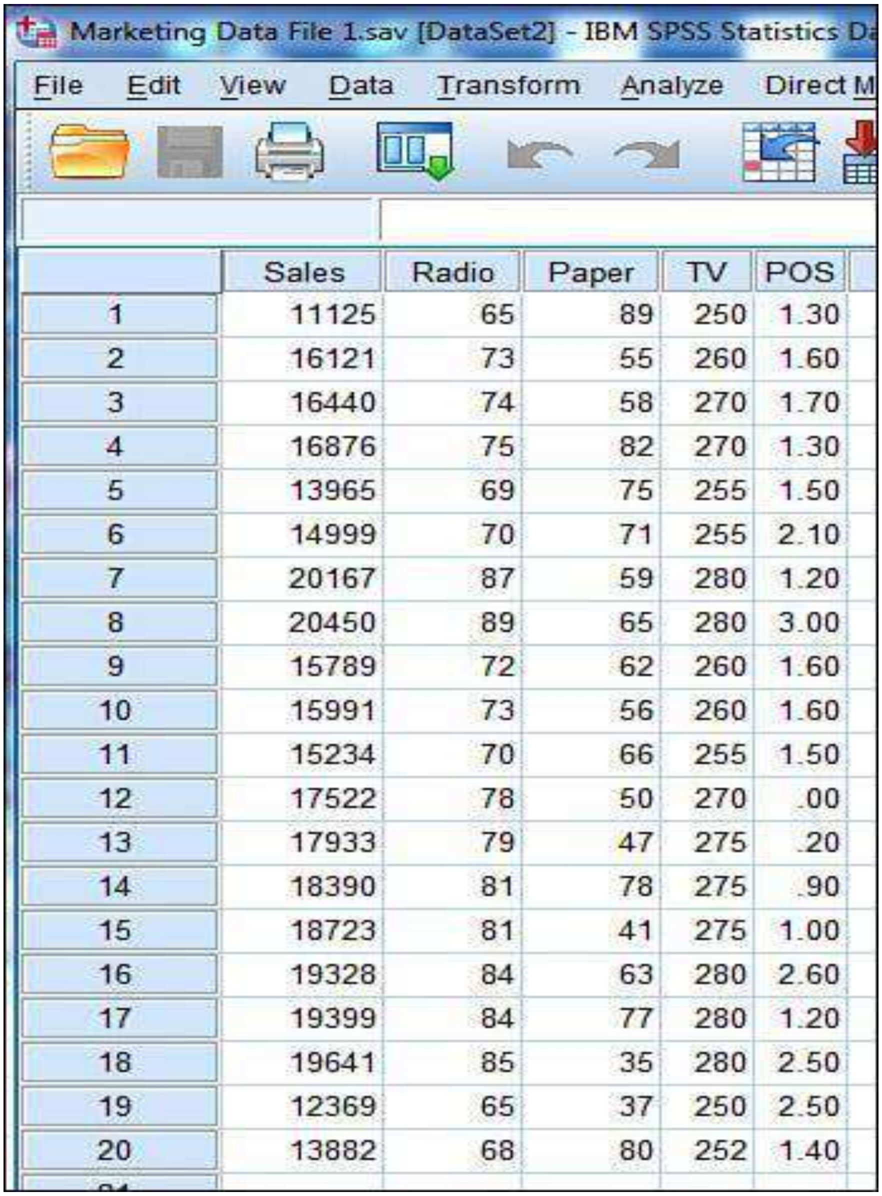
In the last section of this chapter and in Chapters 6, "What Are Predictive Analytics?" and 7, "What Are Prescriptive Analytics?" an ongoing marketing/planning case study of the relevant BA step discussed in those chapters will be presented to illustrate some of the tools and strategies used in a BA problem analysis. This is the first installment of the case study dealing with the descriptive analytics step in BA. The predictive analytics step (in Chapter 6) and prescriptive analytics step (in Chapter 7) will continue with this ongoing case study.

Case Study Background

A firm has collected a random sample of monthly sales information on a service product offered infrequently and only for a month at a time. The sale of this service product occurs only during the month that the promotionefforts are allocated. Basically, promotion funds are allocated at the beginning or during the month, and whatever sales occur are recorded for that promotion effort. There is no spillover of promotion to another month, because monthly offerings of the service product are independent and happen randomly during any particular year. The nature of the product does not appear to be impacted by seasonal or cyclical variations, which preventsforecasting and makes planning the budget difficult.

The firm promotes this service product by using radio commercials, newspaper ads, television commercials, and point-of-sale (POS) ad cards. The firm has collected the sales information as well as promotion expenses. Because the promotion expenses are put into place before the sales take place and on the assumption that the promotion efforts impact products, the four promotion expenses can be viewed as predictive data sets (or what will be the predictive variables in a forecasting model). Actually, in terms of modeling this problem, product sales is going to be considered the dependent variable, and the other four data sets represent independent or predictive variables.

These five data sets, in thousands of dollars, are present in the SPSS printout shown in Figure 5.7. What the firm would like to know is, given a fixed budget of $350,000 for promoting this service product, when offered again, how best should budget dollars be allocated in the hope of maximizing future estimated months' product sales? This is a typical question asked of any product manager and marketing manager's promotion efforts. Before allocating the budget, there is a need to understand how to estimate future product sales. This requires understanding the behavior of product sales relative to sales promotion. To begin to learn about the behavior of product sales to promotion efforts, we begin with the first step in the BA process: descriptive analytics.



	Sales	Radio	Paper	TV	POS
1	11125	65	89	250	1.30
2	16121	73	55	260	1.60
3	16440	74	58	270	1.70
4	16876	75	82	270	1.30
5	13965	69	75	255	1.50
6	14999	70	71	255	2.10
7	20167	87	59	280	1.20
8	20450	89	65	280	3.00
9	15789	72	62	260	1.60
10	15991	73	56	260	1.60
11	15234	70	66	255	1.50
12	17522	78	50	270	.00
13	17933	79	47	275	.20
14	18390	81	78	275	.90
15	18723	81	41	275	1.00
16	19328	84	63	280	2.60
17	19399	84	77	280	1.20
18	19641	85	35	280	2.50
19	12369	65	37	250	2.50
20	13882	68	80	252	1.40

Figure 5.7 Data for marketing/planning case study

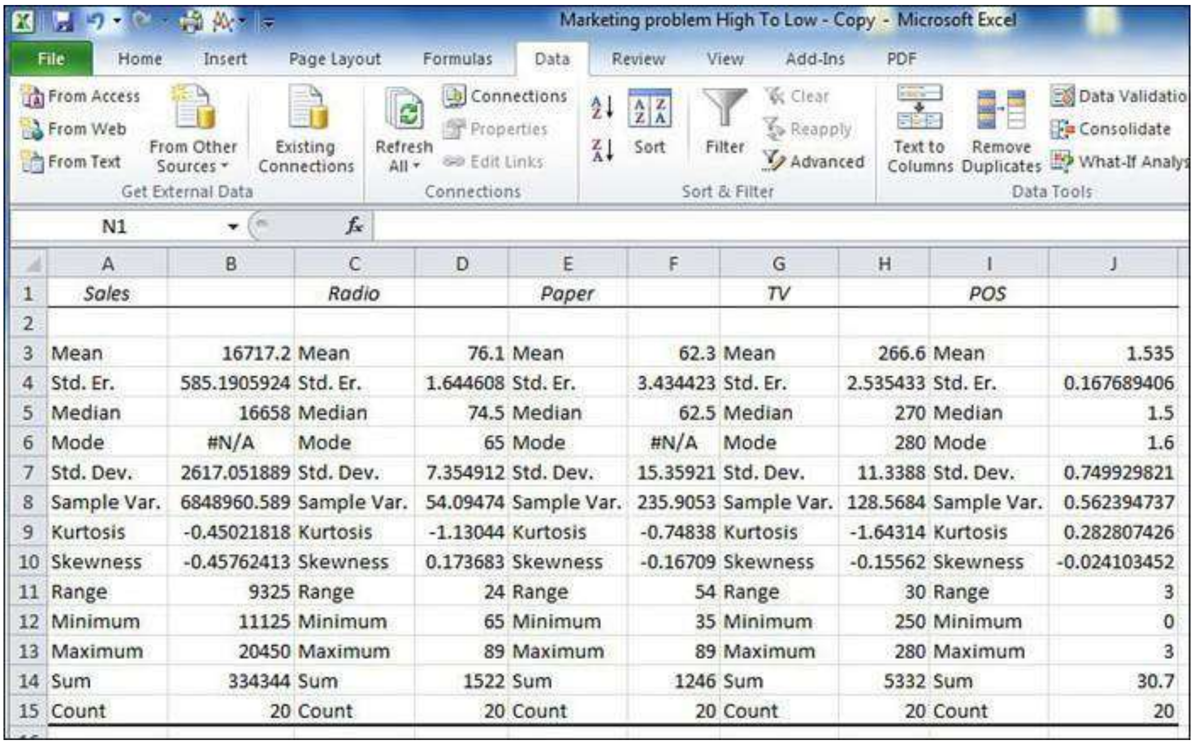
Descriptive Analytics Analysis

To begin conceptualizing possible relationships in the data, one might compute some descriptive statistics and graph charts of data (which will endup being some of the variables in the planned model). SPSS can be used to compute these statistics and charts. The SPSS software printout in Table 5.9 provides a typical set of basic descriptive statistics (means, ranges, standarddeviations, and so on) and several charts. Similarly, Excel's printout in Figure 5.8 provides a basic set of descriptive statistics. Where values cannot be computed, a designation of #N/A is provided.

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Radio	20	24	65	89	76.10	7.355	54.095
Paper	20	54	35	89	62.30	15.359	235.905
TV	20	30	250	280	266.60	11.339	128.568
POS	20	3	0	3	1.54	.750	.562
Sales	20	9325	11125	20450	16717.20	2617.052	6848960.589
Valid N (listwise)	20						

Table 5.9 SPSS Descriptive Statistics for the Marketing/Planning CaseStudy



	A	B	C	D	E	F	G	H	I	J
1	Sales		Radio		Paper		TV		POS	
2										
3	Mean	16717.2	Mean	76.1	Mean	62.3	Mean	266.6	Mean	1.535
4	Std. Er.	585.1905924	Std. Er.	1.644608	Std. Er.	3.434423	Std. Er.	2.535433	Std. Er.	0.167689406
5	Median	16658	Median	74.5	Median	62.5	Median	270	Median	1.5
6	Mode	#N/A	Mode	65	Mode	#N/A	Mode	280	Mode	1.6
7	Std. Dev.	2617.051889	Std. Dev.	7.354912	Std. Dev.	15.35921	Std. Dev.	11.3388	Std. Dev.	0.749929821
8	Sample Var.	6848960.589	Sample Var.	54.09474	Sample Var.	235.9053	Sample Var.	128.5684	Sample Var.	0.562394737
9	Kurtosis	-0.45021818	Kurtosis	-1.13044	Kurtosis	-0.74838	Kurtosis	-1.64314	Kurtosis	0.282807426
10	Skewness	-0.45762413	Skewness	0.173683	Skewness	-0.16709	Skewness	-0.15562	Skewness	-0.024103452
11	Range	9325	Range	24	Range	54	Range	30	Range	3
12	Minimum	11125	Minimum	65	Minimum	35	Minimum	250	Minimum	0
13	Maximum	20450	Maximum	89	Maximum	89	Maximum	280	Maximum	3
14	Sum	334344	Sum	1522	Sum	1246	Sum	5332	Sum	30.7
15	Count	20	Count	20	Count	20	Count	20	Count	20

Figure 5.8 Excel descriptive statistics for the marketing/planning casestudy

Remember, this is the beginning of an exploration that seeks to describe the data and get a handle on what it may reveal. This effort may take someexploration to figure out the best way to express data from a file or database, particularly as the size of the data file increases. In this simple example, the data sets are small but can still reveal valuable information if explored well.

In Figure 5.9, five typical SPSS charts are presented. Respectively, these charts include a bar chart (sales), an area chart (radio), a line chart (paper), apie chart (TV), and a dot chart (POS). These charts are interesting, but they're not very revealing of behavior that will help in understanding future sales trends that may be hiding in this data. Figure 5.10 presents the comparable Excel charts. The only real difference is that the pie chart is called a donut chart.

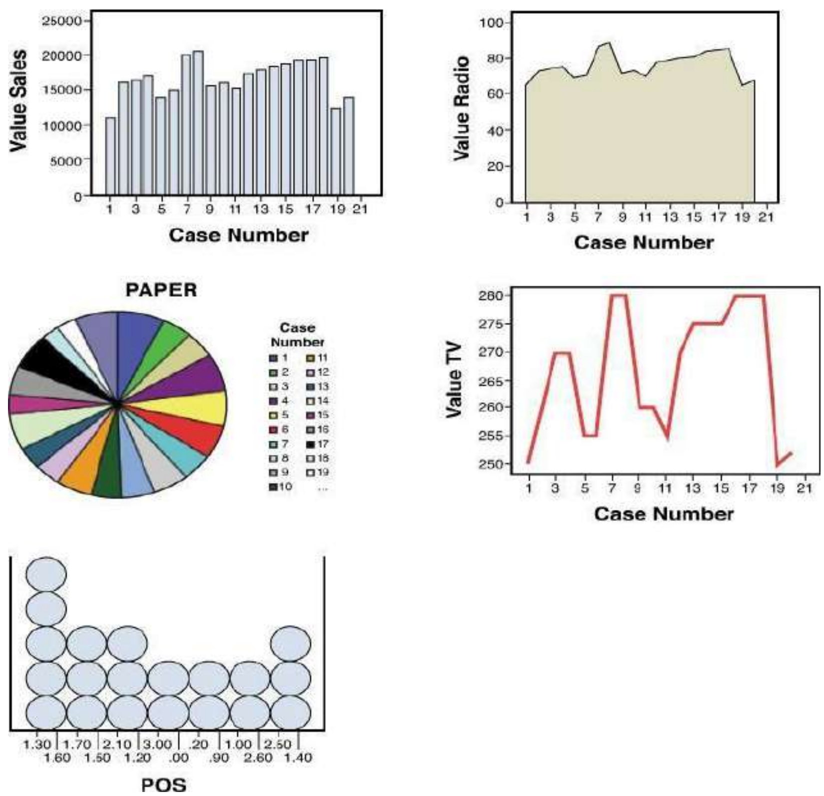
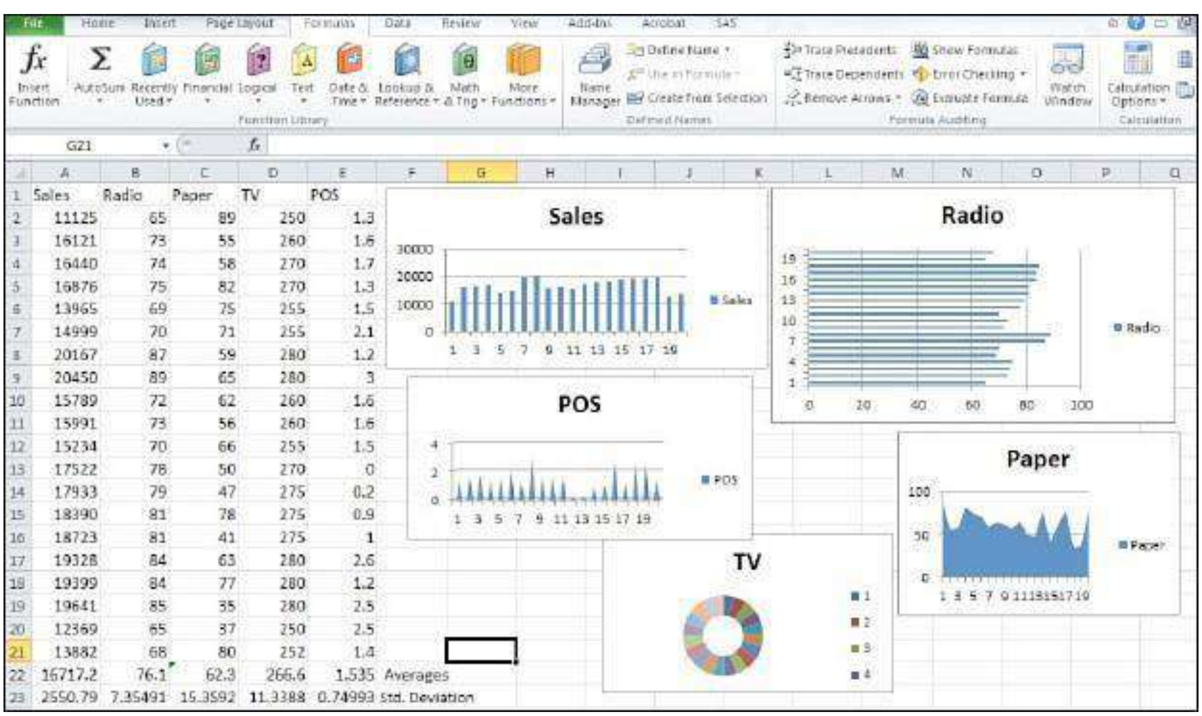


Figure 5.9 Preliminary SPSS charts for the marketing/planning case study



	A	B	C	D	E	F
1	Sales	Radio	Paper	TV	POS	
2	11125	65	89	250	1.3	
3	16121	73	55	260	1.6	
4	16440	74	58	270	1.7	
5	16876	75	82	270	1.3	
6	13965	69	75	255	1.5	
7	14999	70	71	255	2.1	
8	20167	87	59	280	1.2	
9	20450	89	65	280	3	
10	15789	72	62	260	1.6	
11	15991	73	56	260	1.6	
12	15234	70	66	255	1.5	
13	17522	78	50	270	0	
14	17933	79	47	275	0.2	
15	18390	81	78	275	0.9	
16	18723	81	41	275	1	
17	19328	84	63	280	2.6	
18	19399	84	77	280	1.2	
19	19641	85	35	280	2.5	
20	12369	65	37	250	2.5	
21	13882	68	80	252	1.4	
22	16717.2	76.1	62.3	266.6	1.535	Averages
23	2550.79	7.35491	15.3592	11.3388	0.74993	Std. Deviation

Figure 5.1O Preliminary Excel charts for the marketing/planning casestudy

To expedite the process of revealing potential relational information, think in terms of what one is specifically seeking. In this instance, it is to predict the future sales of the service product. That means looking for a graph to show a trend line. One type of simple graph that is related to trendanalysis is a line chart. Using SPSS again, line charts can be computed for each of the five data sets. These charts are presented in Figure 5.11. The vertical axis consists of the dollar values, and the horizontal axis is the number ordering of observations as listed in the data sets. The comparable Excel charts are presented in Figure 5.12.

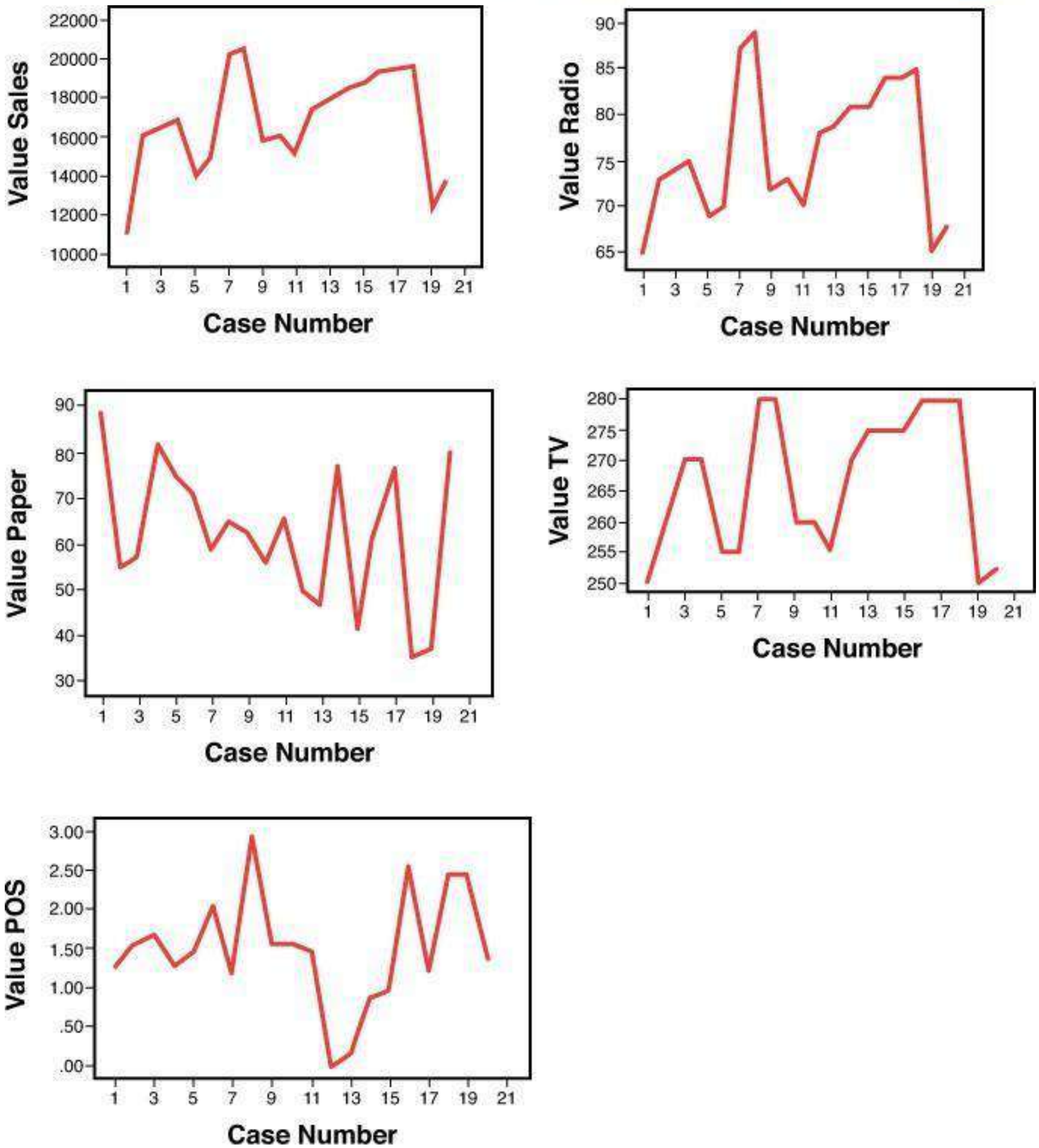
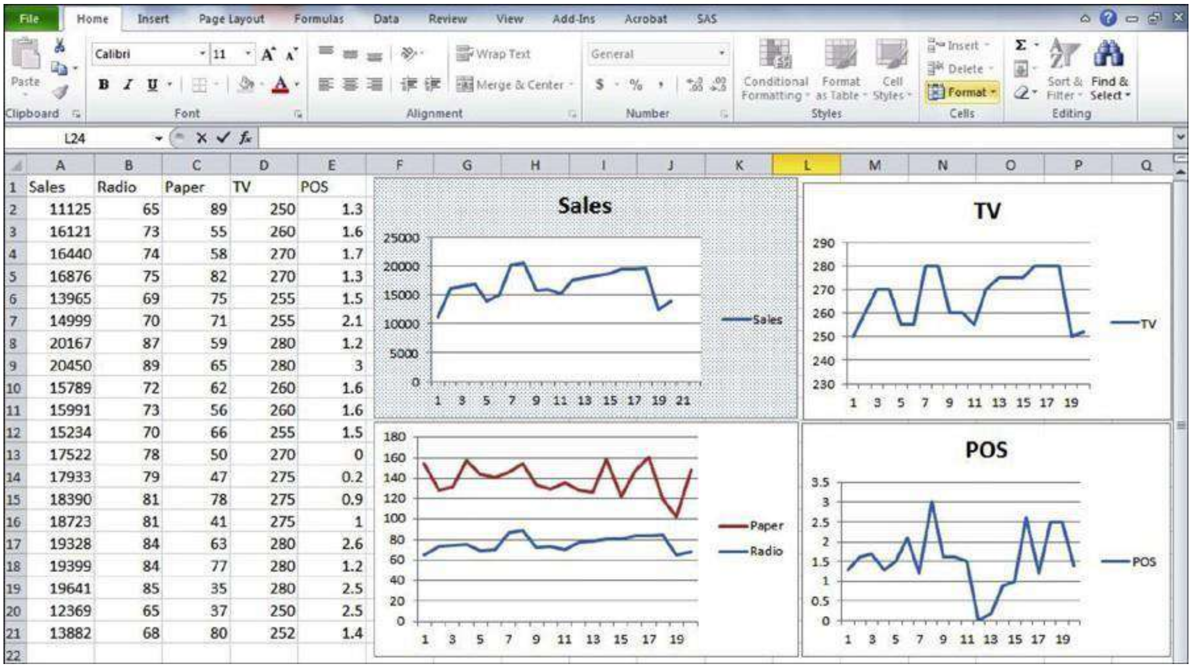


Figure 5.11 Preliminary SPSS line charts for the marketing/planning casestudy

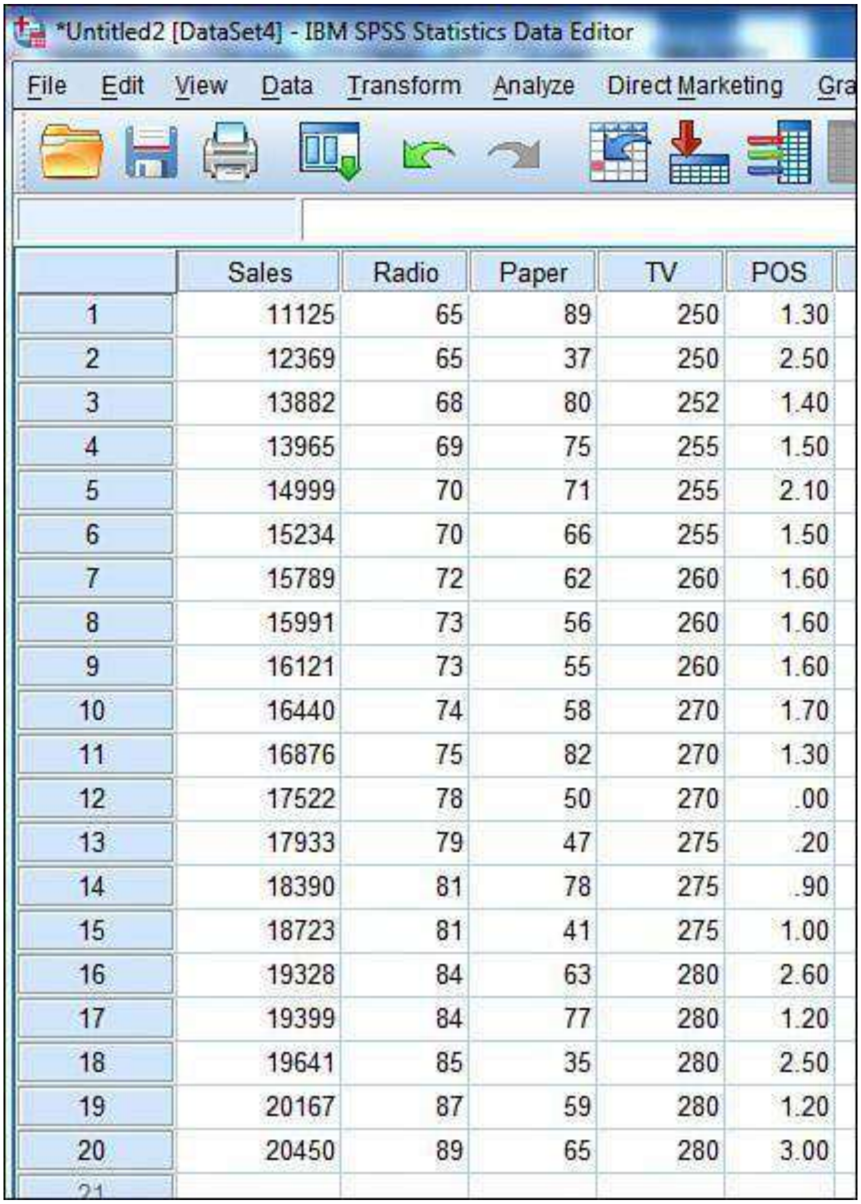


Sales	Radio	Paper	TV	POS
11125	65	89	250	1.3
16121	73	55	260	1.6
16440	74	58	270	1.7
16876	75	82	270	1.3
13965	69	75	255	1.5
14999	70	71	255	2.1
20167	87	59	280	1.2
20450	89	65	280	3
15789	72	62	260	1.6
15991	73	56	260	1.6
15234	70	66	255	1.5
17522	78	50	270	0
17933	79	47	275	0.2
18390	81	78	275	0.9
18723	81	41	275	1
19328	84	63	280	2.6
19399	84	77	280	1.2
19641	85	35	280	2.5
12369	65	37	250	2.5
13882	68	80	252	1.4

Figure 5.12 Preliminary Excel line charts for the marketing / planning case study

While providing a less confusing graphic presentation of the up-and- down behavior of the data, the charts in these figures still do not clearly reveal any possible trend information. Because the

20 months of data are not in any particular order and are not related to time, they are independent values that can be reordered in any way. Reordering data or sorting it can bea part of the descriptive analytics process when needed. Because trend is usually an upward or downward linear behavior, one might be able to observe a trend in the product sales data set if that data is reordered from low to high (or high to low). Reordering the sales by moving the 20 rows ofdata around such that sales is arranged from low to high is presented in Figure 5.13. Using this reordered data set, the SPSS results are illustrated in the new line charts in Figure 5.14. The comparable Excel charts are presented in Figure 5.15.



	Sales	Radio	Paper	TV	POS
1	11125	65	89	250	1.30
2	12369	65	37	250	2.50
3	13882	68	80	252	1.40
4	13965	69	75	255	1.50
5	14999	70	71	255	2.10
6	15234	70	66	255	1.50
7	15789	72	62	260	1.60
8	15991	73	56	260	1.60
9	16121	73	55	260	1.60
10	16440	74	58	270	1.70
11	16876	75	82	270	1.30
12	17522	78	50	270	.00
13	17933	79	47	275	.20
14	18390	81	78	275	.90
15	18723	81	41	275	1.00
16	19328	84	63	280	2.60
17	19399	84	77	280	1.20
18	19641	85	35	280	2.50
19	20167	87	59	280	1.20
20	20450	89	65	280	3.00
21					

Figure 5.13 Reordered data in line charts for the marketing/planning casestudy

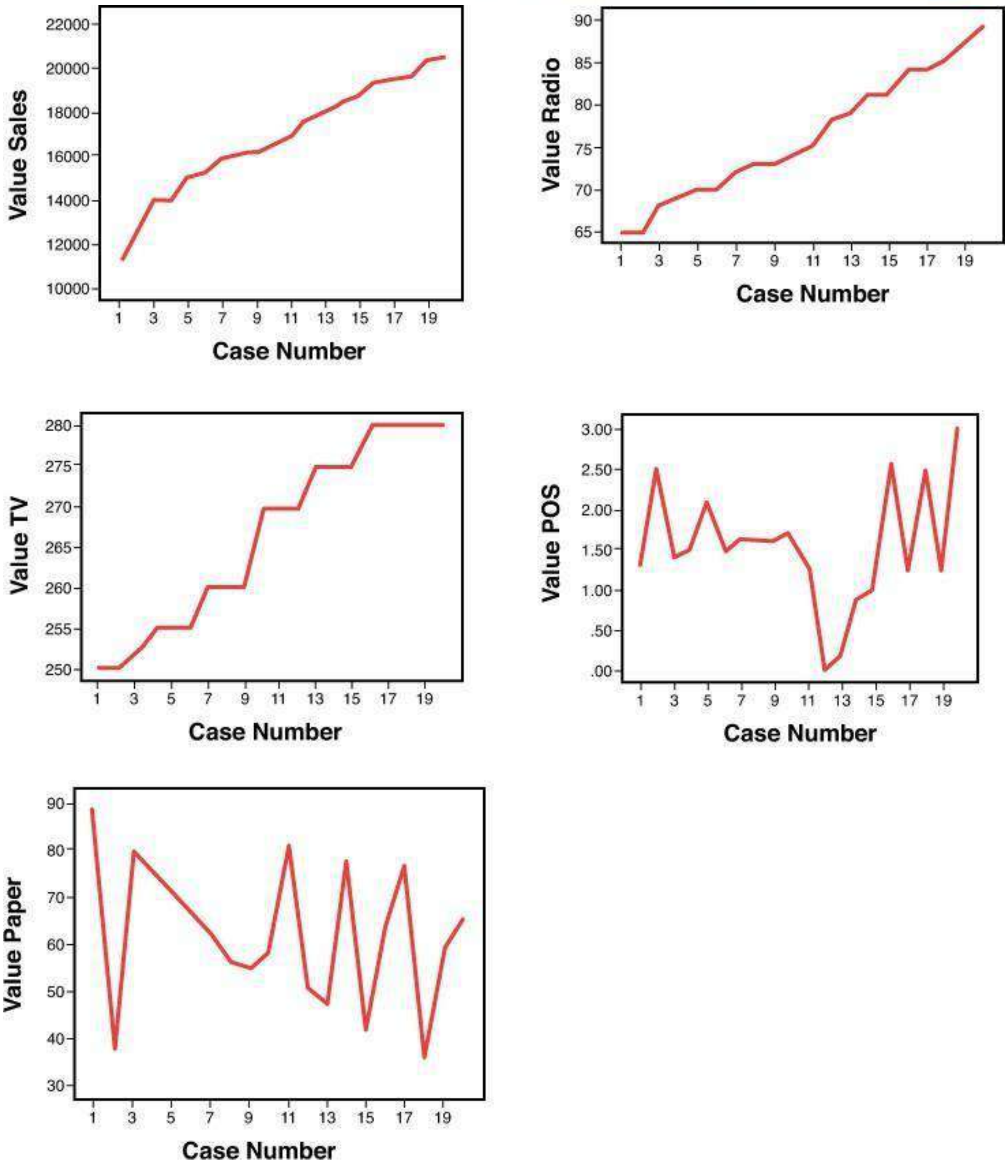
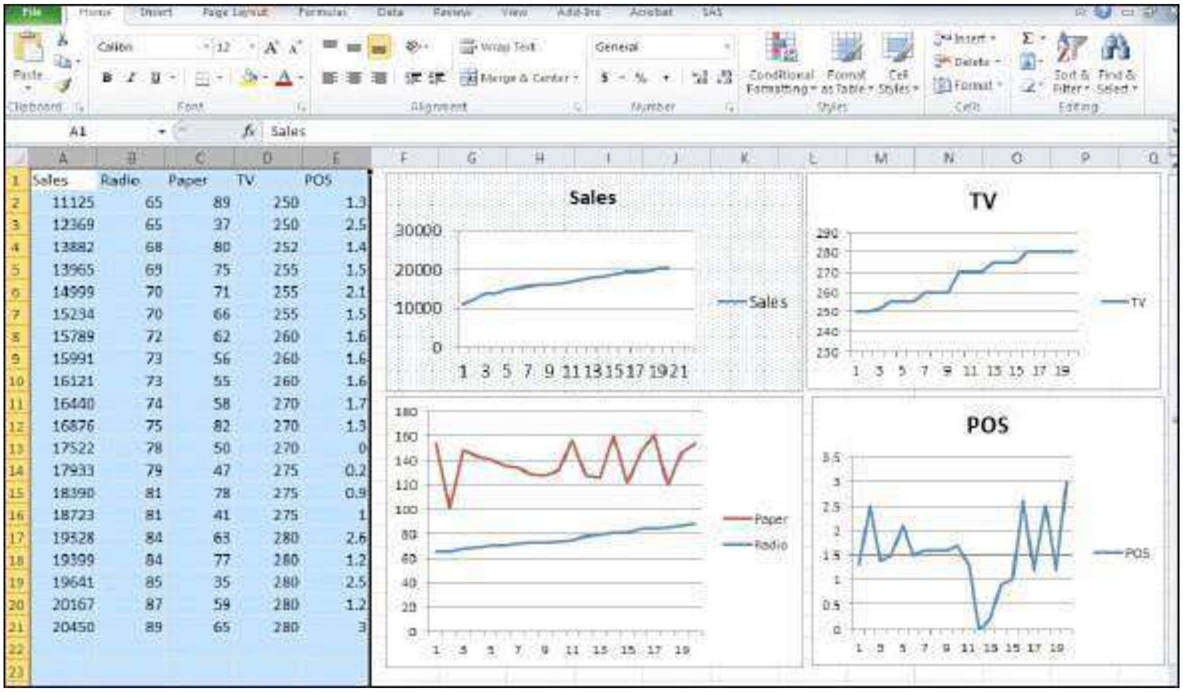


Figure 5.14

SPSS line charts based on reordered data for themarketing/planning case study



Sales	Radio	Paper	TV	POS
11125	65	89	250	1.3
12369	65	37	250	2.5
13882	68	80	252	1.4
13965	69	75	255	1.5
14999	70	71	255	2.1
15234	70	66	255	1.5
15789	72	62	260	1.6
15991	73	56	260	1.6
16121	73	55	260	1.6
16440	74	58	270	1.7
16876	75	82	270	1.3
17522	78	50	270	0
17933	79	47	275	0.2
18390	81	78	275	0.9
18723	81	41	275	1
19328	84	63	280	2.6
19399	84	77	280	1.2
19641	85	35	280	2.5
20167	87	59	280	1.2
20450	89	65	280	3

Figure 5.15 Excel line charts based on reordered data for the marketing/planning case study

Given the low to high reordering of the product sales as a guide, some of the other four line charts suggest a relationship with product sales. Both radio and TV commercials appear to have a similar low to high trending relationship that matches with product sales. This suggests these two will begood predictive variables for product sales, whereas newspaper and POS ads are still considerably volatile in their charted relationships with product sales. Therefore, these two latter variables might not be useful in a model seeking to predict product sales. They cannot be ruled out at this point in the analysis, but they are suspected of adding little to a model for accurately forecasting product sales. Put another way, they appear to add unneeded variation that may take away from the accuracy of the model. Further analysis is called for to explore in more detail and sophistication the best setof predictive variables to predict the relationships in product sales.
In summary, for this case study, the descriptive analytics analysis has revealed a potential relationship between radio and TV commercials and future product sales, and it questions the relationship of newspaper and POS ads to sales. The managerial ramifications of these results might suggest discontinuing investing in newspaper and POS ads and more productively allocate funds to radio and TV commercials. Before such a reallocation can be justified, more analysis is needed. The next step in the analysis, predictive analytics, will be presented in the last section of Chapter 6.

Summary

This chapter discussed data visualization and exploration. In particular, this chapter described and illustrated graphic and statistical methods usefulin the descriptive analytics step of the BA process. Illustrations of both SPSS and Excel printouts of graphs, charts, and statistical methods were presented. In addition, sampling methods were described, along with the available software applications from SPSS and Excel. Sampling estimation was also discussed, as was its connection to sampling distributions for purposes of error estimation in measures of central tendency. Finally, this chapter presented the first installment of a case study illustrating the descriptive analytics step of the BA process. The remaining installments will be presented in Chapters 6 and 7.

Several of the appendixes of this book are designed to augment the chapter material by including technical, mathematical, and statistical tools.For both greater understanding of the methodologies discussed in this chapter and a basic review of statistical and other quantitative methods, a review of the appendixes mentioned in this chapter is recommended.

The results of the descriptive analytics step of the BA process create an exploratory foundation on which further analysis can be based. In Chapter 6, we continue with the second step of the BA process: predictive analytics.

Discussion Questions

1. Why is it important to explore data with graphs and charts?
2. What is the difference between skewedness and kurtosis?
3. Why would we ever want to use a sample if we have populationinformation?
4. Is there a way to determine skewedness from the ordering of themean, median, and mode measures of central tendency?
5. Which of the sampling methods listed in Table 5.5 is the best, andwhy?
6. In setting the confidence level, why not just set one that is low enough for the population parameter to be assured of inclusion?

Problems

1. Using either SPSS or Excel, draw a line graph of the Sales 2 distribution from the data in Figure 5.1. Does the kurtosis statistic in Table 5.4 make sense given your graph? Does the positioning of themean, median, and mode support the skewedness statistic in Table 5.4? Explain the answer to both questions.
2. Using either SPSS or Excel, draw a scatter diagram of the Sales 3 distribution from the data in Figure 5.1. Based on the statistics in Table 5.4, how would you

judge its skewedness: highly or slightly? Does the positioning of the mean, median, and mode support the skewedness statistic in Table 5.4? Explain the answer to both questions.

3. Using SPSS or Excel, make a random sample on Sales 3 distribution from the data in Figure 5.1. Using the software, determine four items from the data set for sampling purposes. Which specific values shouldbe selected from the data set?
4. Using SPSS or Excel, make a random sample on Sales 2 distribution from the data in Figure 5.1. Using the software, determine six items from the data set for sampling purposes. Which specific values shouldbe selected from the data set?
5. With a mean value of 50 and a standard error of the mean of 12, whatis the 90 percent confidence interval for this problem?
6. With a mean value of 120 and a standard error of the mean of 20,what is the 99 percent confidence interval for this problem?
7. A firm has computed its mean sales for a new product to be 2,000 units for the year, with a standard error of the mean of 56. The firmwould like to know if the probability of its mean sales for next year (based on this year) will be above 2,112. What is the probability?
8. The Homes Golf Ball Company has made a number of different golf products over the years. Research on thousands of balls revealed the mean flight distance of its Maximum Fly golf ball product to be 450 yards, with a standard error of the mean of 145 yards. The company is hoping to improve the product to fly an additional 290 yards. What is the probability of the improvement from 450 to 740 yards?

UNIT IV PREDICTIVE ANALYTICS

Introduction to Predictive analytics - Logic and Data Driven Models - Predictive AnalysisModeling and procedure - Data Mining for Predictive analytics. Analysis of Predictive analytics

What Are Predictive Analytics?

Introduction

In Chapter 1, “What Are Business Analytics?” we defined predictive analytics as an application of advanced statistical, information software, or operations research methods to identify predictive variables and build predictive models to identify trends and relationships not readily observed in the descriptive analytic analysis. Knowing that relationships exist explains why one set of independent variables (predictive variables) influences dependent variables like business performance. Chapter 1 further explained that the purpose of the descriptive analytics step is to position decision makers to build predictive models designed to identify and predict future trends.

Picture a situation in which big data files are available from a firm’s sales and customer information (responses to differing types of advertisements, customer surveys on product quality, customer surveys on supply chain performance, sale prices, and so on). Assume also that a previous descriptive analytic analysis suggests there is a relationship between certain customer variables, but there is a need to precisely establish a quantitative relationship between sales and customer behavior. Satisfying this need requires exploration into the big data to first establish whether a measurable, quantitative relationship does in fact exist and then develop a statistically valid model in which to predict future events. This is what the predictive analytics step in BA seeks to achieve.

Many methods can be used in this step of the BA process. Some are justto sort or classify big data into manageable files in which to later build a precise quantitative model. As previously mentioned in Chapter 3, “What Resource Considerations Are Important to Support Business Analytics?” predictive modeling and analysis might consist of the use of methodologies, including those found in forecasting, sampling and estimation, statistical inference, data mining, and regression analysis. A commonly used methodology is multiple regression. (See Appendixes A, “Statistical Tools,” and E, “Forecasting,” for a discussion on multiple regression and ANOVA testing.) This methodology is ideal for establishing whether a statistical relationship exists between the predictive variables found in the descriptive analysis and the dependent variable one seeks to forecast. An example of itsuse will be presented in the last section of this chapter.

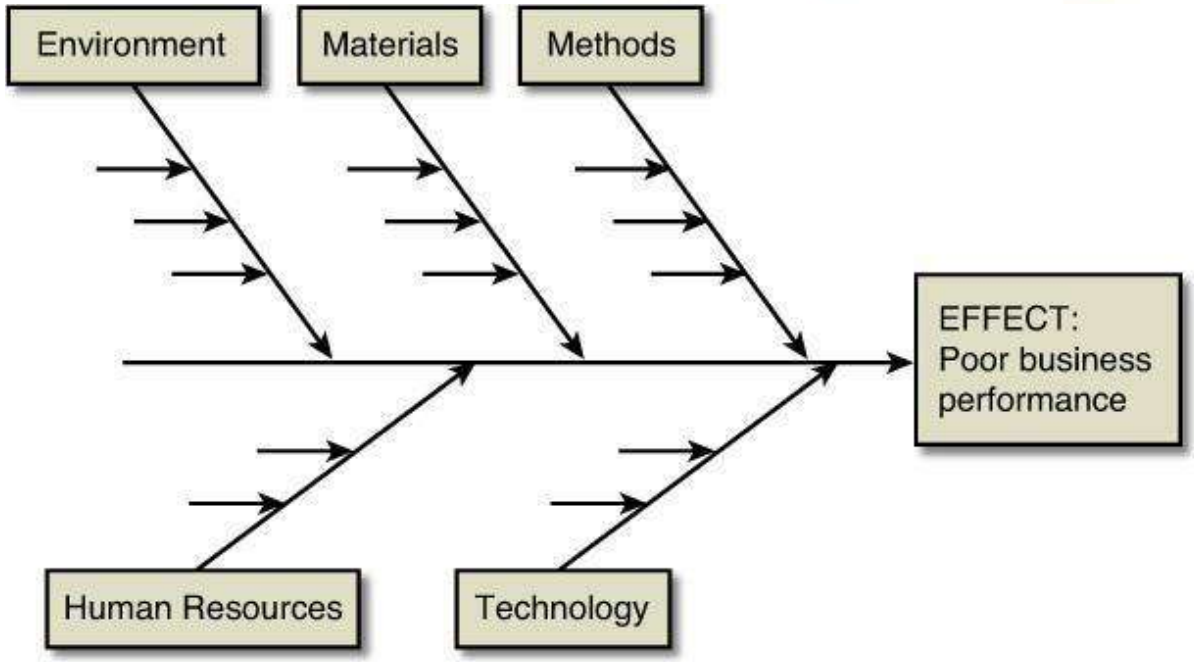
Although single or multiple regression models can often be used to forecast a trend line into the future, sometimes regression is not practical. Insuch cases, other forecasting methods, such as exponential smoothing or smoothing averages, can be applied as predictive analytics to develop needed forecasts of business activity. (See Appendix E.) Whatever methodology is used, the identification of future trends or forecasts is the principle output of the predictive analytics step in the BA process.

Predictive Modeling

Predictive modeling means developing models that can be used to forecast or predict future events. In business analytics, models can bedeveloped based on logic or data.

Logic-Driven Models

A *logic-driven model* is one based on experience, knowledge, and logicalrelationships of variables and constants connected to the desired business performance outcome situation. The question here is how to put variables and constants together to create a model that can predict the future. Doing this requires business experience. Model building requires an understanding of business systems and the relationships of variables and constants that seek to generate a desirable business performance outcome. To help conceptualize the relationships inherent in a business system, diagramming methods can be helpful. For example, the *cause-and-effect diagram* is a visual aid diagram that permits a user to hypothesize relationships between potential causes of an outcome (see Figure 6.1). This diagram lists potentialcauses in terms of human, technology, policy, and process resources in an effort to establish some basic relationships that impact business performance. The diagram is used by tracing contributing and relational factors from the desired business performance goal back to possible causes, thus allowing the user to better picture sources of potential causes that could affect the performance. This diagram is sometimes referred to as a *fishbone diagram* because of its appearance.



**Source*: Adapted from Figure 5 in Schniederjans et al. (2014), p. 201.

Figure 6.1 Cause-and-effect diagram*

Another useful diagram to conceptualize potential relationships with business performance variables is called the *influence diagram*. According to Evans (2013, pp. 228–229), influence diagrams can be useful to conceptualize the relationships of variables in the development of models. An example of an influence diagram is presented in Figure 6.2. It maps the relationship of variables and a constant to the desired business performanceoutcome of profit. From such a diagram, it is easy to convert the information into a quantitative model with constants and variables that define profit in this situation:

Profit = Revenue − Cost, or
Profit = (Unit Price × Quantity Sold) − [(Fixed Cost) + (Variable Cost ×Quantity Sold)],or
P = (UP × QS) − [FC + (VC × QS)]

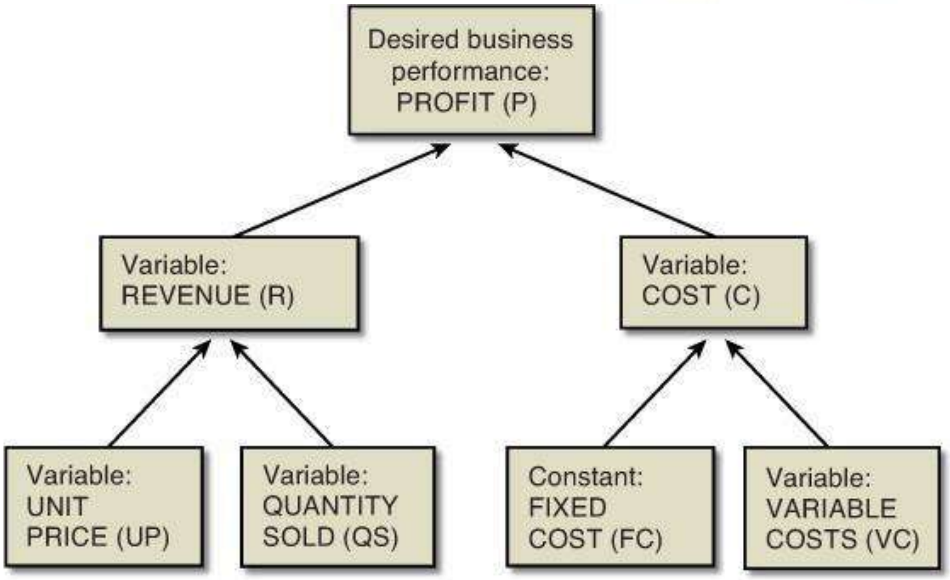


Figure 6.2 An influence diagram

The relationships in this simple example are based on fundamental business knowledge. Consider, however, how complex cost functions mightbecome without some idea of how they are mapped together. It is necessary to be knowledgeable about the business systems being modeled in order to capture the relevant business behavior. Cause-and-effect diagrams and influence diagrams provide tools to conceptualize relationships, variables, and constants, but itoften takes many other methodologies to explore and develop predictive models Data-Driven Models

Logic-driven modeling is often used as a first step to establish relationships through *data-driven models* (using data collected from many sources to quantitatively establish model relationships). To avoid duplication of content and focus on conceptual material in the chapters, most of the computational aspects and some computer usage content are relegated to the appendixes. In addition, some of the methodologies are illustrated in the case problems presented in this book. Please refer to the Additional Information column in Table 6.1 to obtain further information onthe use and application of the data-driven models.

Data-Driven Models	Possible Applications	Additional Information
Sampling and Estimation	Generate statistical confidence intervals to define limitations and boundaries on future forecasts for other forecasting models.	Chapter 5, "What Are Descriptive Analytics?," Appendix A, "Statistical Tools," Appendix E, "Forecasting."
Regression Analysis	(1) Create a predictive equation useful for forecasting time series forecasts. (2) Weed out predictive variables in forecasting models that add little to predicting values. (3) Generate a trend line for forecasting.	Chapter 6, "What Are Predictive Analytics?," Chapter 8, "A Final Case Study Illustration," Appendix E.
Correlation Analysis	(1) Assess variable relationships. (2) Weed out predictive variables in forecasting models that add little to predicting values.	Chapter 6, Appendix E.
Probability Distributions	(1) Estimate trend behavior that follows certain types of probability distributions. (2) Conduct statistical tests to confirm significance of variables.	Chapter 5, Appendix A.
Predictive Modeling and Analysis	Fit linear and nonlinear models to data to use the models for forecasting.	Appendix A, Appendix E.
Forecasting Models	Those listed in this table and others such as smoothing models can be used to forecast values.	Appendix E.
Simulation	Project future behavior in variables by simulating the past behavior found in probability distributions.	Appendix F, "Simulation."

Table 6.1 Data-Driven Models

Data Mining

As mentioned in Chapter 3, "What Resource Considerations are Important to Support Business Analytics?" *data mining* is a discovery- driven software application process that provides insights into business databy finding hidden patterns and relationships in big or small data and inferring rules from them to predict future behavior. These observed patterns and rules guide decision-making. This is not just numbers, but text and social media information from the Web. For example, Abrahams et al. (2013) developed a set of text-mining rules that automobile manufacturers could use to distill or mine specific vehicle component issues that emerge on the Web but take months to show up in complaints or other damaging media. These rules cut through the mountainous data that exists on the Web and are reported to provide marketing and competitive intelligence to manufacturers, distributors, service centers, and suppliers. Identifying a product's defects and quickly recalling or correcting the problem before customers experience a failure reduce customer dissatisfaction when problems occur.

A Simple Illustration of Data Mining

Suppose a grocery store has collected a big data file on what customers put into their baskets at the market (the collection of grocery items a customer purchases at one time). The grocery store would like to know if there are any associated items in a typical market basket. (For example, if a customer purchases product A, she will most often associate it or purchaseit with product B.) If the customer generally purchases product A and B together, the store might only need to advertise product A to gain both product A's and B's sales. The value of knowing this association of productscan improve the performance of the store by reducing the need to spend money on advertising both products. The benefit is real if the association holds true.

Finding the association and proving it to be valid requires some analysis.

From the descriptive analytics analysis, some possible associations may have been uncovered, such as product A's and B's association. With any size data file, the normal procedure in data mining would be to divide the file into two parts. One is referred to as a training data set, and the other as avalidation data set. The *training data set* develops the association rules, andthe *validation data set* tests and proves that the rules work. Starting with the training data set, a common data mining methodology is *what-if analysis*

using logic-based software. Excel and SPSS both have what-if logic-based software applications, and so do a number of other software vendors (see Chapter 3). These software applications allow logic expressions. (For example, if product A is present, then is product B present?) The systems can also provide frequency and probability information to show the strengthof the association. These software systems have differing capabilities, which permit users to deterministically simulate different scenarios to identify complex combinations of associations between product purchases in a market basket.

Once a collection of possible associations is identified and their probabilities are computed, the same logic associations (now considered association rules) are reran using the validation data set. A new set of probabilities can be computed, and those can be statistically compared using hypothesis testing methods to determine their similarity. Other software systems compute correlations for testing purposes to judge the strength and the direction of the relationship. In other words, if the consumer buys product A first, it could be referred to asthe *Head* and product B as the *Body* of the association. (Nisbet et al., 2009, p. 128). If the same basic probabilities are statistically significant, it lends validity to the association rules and their use for predicting market basket item purchases based on groupings of products.

Data Mining Methodologies

Data mining is an ideal predictive analytics tool used in the BA process. We mentioned in

Chapter 3 different types of information that data mining can glean, and Table 6.2 lists a small sampling of data mining methodologies to acquire different types of information. Some of the same tools used in the descriptive analytics step are used in the predictive step but are employed to establish a model (either based on logical connections or quantitative formulas) that may be useful in predicting the future.

Types of Information	Description	Sample of Data Mining Methodologies
Association	Occurrence linked to a single event.	Association rules (for example, if-then analysis), correlation analysis, neural networks.
Classification	Pattern that describes the group an item belongs to. Found by examining previous classified existing items and inferring a set of rules that guide the classification process.	Discriminant analysis, logistics regression, neural networks.
Clustering	Similar to classification when no groups have yet been defined. Helps discover different groupings within data.	Hierarchical clustering, K-mean clustering.
Forecasting	Used to predict values that can identify patterns in customer behavior.	Regression analysis, correlation analysis.
Sequence	Event that is linked over time.	Lag correlation analysis, cause-and-effect diagrams.

Table 6.2 Types of Information and Data Mining Methodologies Several computer-based methodologies listed in Table 6.2 are briefly introduced here. *Neural networks* are used to find associations where connections between words or numbers can be determined. Specifically, neural networks can take large volumes of data and potential variables and explore variable associations to express a beginning variable (referred to as an *input layer*), through middle layers of interacting variables, and finally toan ending variable (referred to as an *output*). More than just identifying simple one-on-one associations, neural networks link multiple association pathways through big data like a collection of nodes in a network. These nodal relationships constitute a form of classifying groupings of variables as related to one another, but even more, related in complex paths with multiple associations (Nisbet et al., 2009, pp. 128–138). SPSS has two versions of neural network software functions: *Multilayer Perceptron* (MLP) and *Radial Basis Function* (RBF). Both procedures produce a predictive model for one or more dependent variables based on the values of the predictive variables. Both allow a decision maker to develop, train, and use the

software to identify particular traits (such as bad loan risks for abank) based on characteristics from data collected on past customers).

Discriminant analysis is similar to a multiple regression model except that it permits continuous independent variables and a categorical dependent variable. The analysis generates a regression function whereby values of the independent variables can be incorporated to generate a predicted value for the dependent variable. Similarly, *logistic regression* is like multiple regression. Like discriminant analysis, its dependent variable can be categorical. The independent variables, though, in logistic regression can be either continuous or categorical. For example, in predicting potentialoutsource providers, a firm might use a logistic regression,in which the dependent variable would be to either classify an outsource provider as rejected (represented by the value of the dependent variable being zero) or classify the outsource provider as acceptable (represented by the value of one for the dependent variable).

Hierarchical clustering is a methodology that establishes a hierarchy of clusters that can be grouped by the hierarchy. Two strategies are suggested for this methodology: agglomerative and divisive. The *agglomerative strategy* is a bottom-up approach, where one starts with each item in the data and begins to group them. The *divisive strategy* is a top-down approach, where one starts with all the items in one group and divides the group into clusters. How the clustering takes place can involve many different types of algorithms and differing software applications. One method commonly used is to employ a Euclidean distance formula that looks at the square root of the sum of distances between two variables, their differences squared. Basically, the formula seeks to match up variable candidates that have the least squared error differences. (In other words, they're closer together.)

K-mean clustering is a classification methodology that permits a set of data to be reclassified into *K groups*, where *K* can be set as the number of groups desired. The algorithmic process identifies initial candidates for the *K* groups and then interactively searches other candidates in the data set to be averaged into a mean value that represents a particular *K* group. The process of selection is based on maximizing the distance from the initial *K* candidates selected in the initial run through the list. Each run or iteration through the data set allows the software to select further candidates for eachgroup.

The K-mean clustering process provides a quick way to classify data into differentiated groups. To illustrate this process, use the sales data in Figure

6.3 and assume these are sales from individual customers. Suppose a company wants toclassify the sales customers into high and low salesgroups.

Time	Sale
1	13444
2	12369
3	15322
4	13965
5	14999
6	15234
7	12999
8	15991
9	16121
10	18654
11	16876
12	17522
13	17933
14	15233
15	18723
16	13855
17	19399
18	16854
19	20167
20	18654

Figure 6.3 Sales data for cluster classification problem

The SPSS K-Mean cluster software can be found in Analyze > Classify >K-Means Cluster Analysis. Any integer value can designate the *K* number of clusters desired. In this problem set, *K*=2. The SPSS printout of this classification process is shown in Table 6.3. The solutionis referred to as a *Quick Cluster* because it initially selects the first two high and low values.

The Initial Cluster Centers table listed the initial high (20167) and a low (12369) value from the data set as the clustering process begins. As it turns out, the software divided the customers into nine high sales customers witha group mean sales of 18,309 and eleven low sales customers with a groupmean sales of 14,503.

Quick Cluster

Initial Cluster Centers

	Cluster 1	Cluster 2
Sale	20167	12369

Final Cluster Centers

	Cluster 1	Cluster 2
Sale	18309	14503

Number of Cases in each Cluster

Cluster	1	9.000
	2	11.000
Valid		20.000
Missing		.000

Table 6.3 SPSS K-Mean Cluster Solution Consider how large big data sets can be. Then realize this kind of classification capability can be a useful tool for identifying and predictingsales based on the mean values.

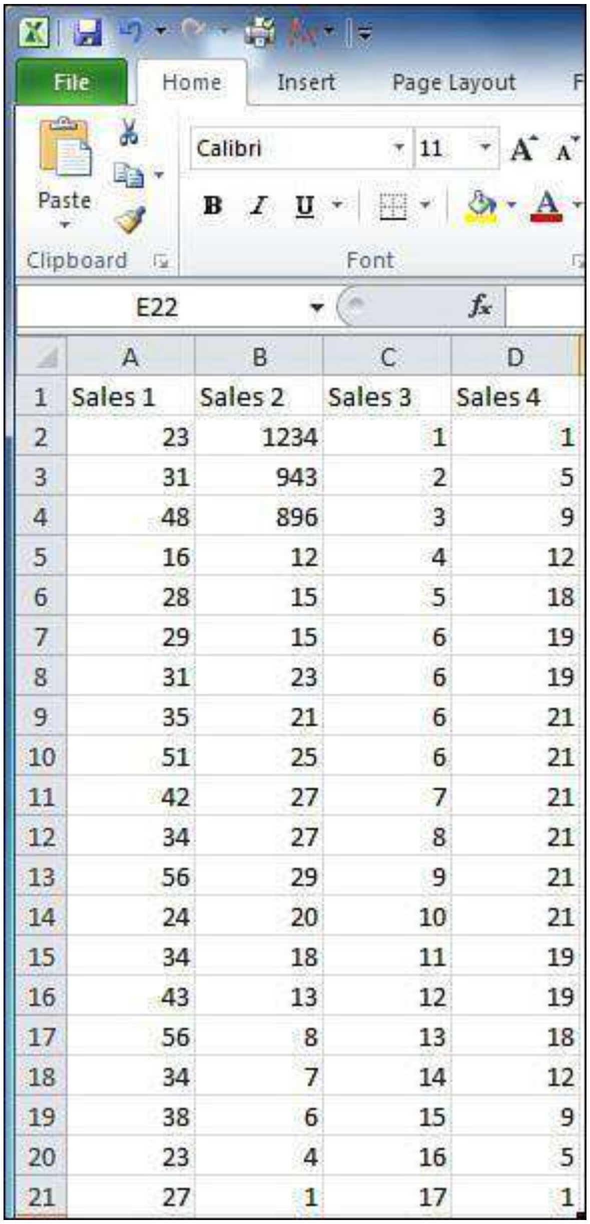
There are so many BA methodologies that no single section, chapter, oreven book can explain or contain them all. The analytic treatment and computer usage in this chapter have been focused mainly on conceptual use. For a more applied use of some of these methodologies, note the casestudy that follows and some of the content in the appendixes.

Continuation of Marketing/Planning Case Study Example:Prescriptive Analytics Step in the BA Process

In the last sections of Chapters 5, 6, and 7, an ongoing marketing/planning case study of the relevant BA step discussed in those chapters is presented to illustrate some of the tools and strategies used in a BA problem analysis. This is the second installment of the case study dealing with the predictive analytics analysis step in BA. The prescriptive analysis step coming in Chapter 7, “What Are Prescriptive Analytics?” will complete the ongoing case study.

Case Study Background Review

The case study firm had collected a random sample of monthly sales information presented in Figure 6.4 listed in thousands of dollars. What the firm wants to know is, given a fixed budget of $350,000 for promoting this service product, when offered again, how best should the company allocate budget dollars in hopes of maximizing the future estimated month’s product sales? Before making any allocation of budget, there is a need to understand how to estimate future product sales. This requires understanding the behavior of product sales relative to sales promotion efforts using radio, paper, TV, and point-of-sale (POS) ads.



	A	B	C	D
1	Sales 1	Sales 2	Sales 3	Sales 4
2	23	1234	1	1
3	31	943	2	5
4	48	896	3	9
5	16	12	4	12
6	28	15	5	18
7	29	15	6	19
8	31	23	6	19
9	35	21	6	21
10	51	25	6	21
11	42	27	7	21
12	34	27	8	21
13	56	29	9	21
14	24	20	10	21
15	34	18	11	19
16	43	13	12	19
17	56	8	13	18
18	34	7	14	12
19	38	6	15	9
20	23	4	16	5
21	27	1	17	1

Figure 6.4 Data for marketing/planning case study

The previous descriptive analytics analysis in Chapter 5 revealed a potentially strong relationship between radio and TV commercials that might be useful in predicting future product sales. The analysis also revealed little regarding the relationship of newspaper and POS ads to product sales. So although radio and TV commercials are most promising, amore in-depth predictive analytics analysis is called for to accurately measure and document the degree of relationship that may exist in the variables to determine the best predictors of product sales.

Predictive Analytics Analysis

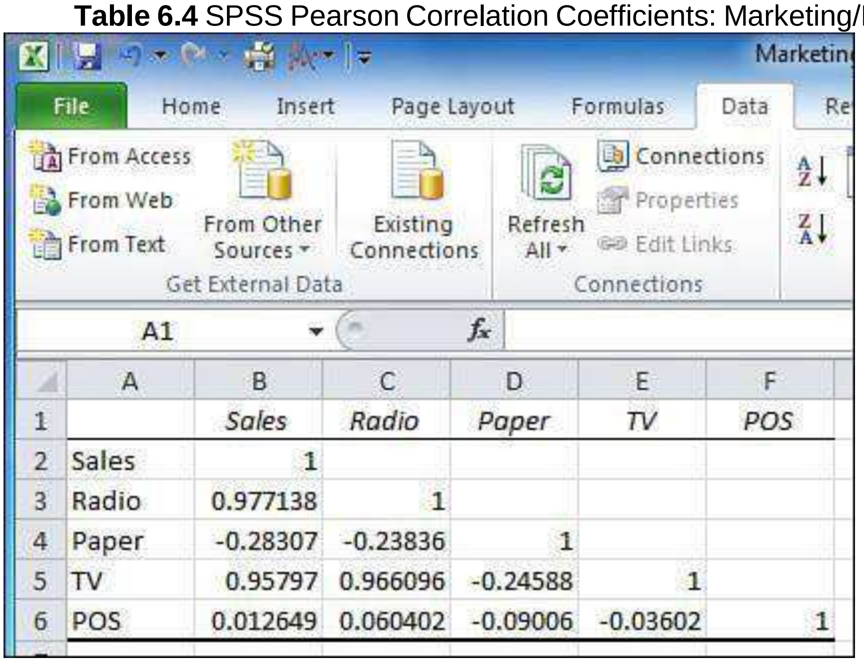
An ideal multiple variable modeling approach that can be used in this situation to explore variable importance in this case study and eventually lead to the development of a predictive model for product sales is correlation and multiple regression. We will use both Excel and IBM's SPSS statistical packages to compute the statistics in this step of the BA process.

First, we must consider the four independent variables—radio, TV, newspaper, POS—before developing the model. One way to see the statistical direction of the relationship (which is better than just comparing graphic charts) is to compute the Pearson correlation coefficients *r* between each of the independent variables with the dependent variable (product sales). The SPSS correlation coefficients and their levels of significance are presented in Table 6.4. The comparable Excel correlations are presented in Figure 6.5. Note: They do not include the level ofsignificance but provide correlations between all the variables being considered. The larger the Pearson correlation (regardless of the sign) and the smaller the *Significancetest* values (these are t-tests measuring the significance of the Pearson *r* value; see Appendix A), the more significant the relationship. Both radio and TV are statistically significant correlations, whereas at a 0.05 level of significance, paper and POS are not statistically significant.

Statistic	Radio	Paper	TV	POS
Pearson Correlation *r* with Product Sales	.977	–.283	.958	.013
Significance Test (1-Tailed)*	.000	.113	.000	.479

*Values of 0.05 or less would designate a significant relationship with product sales

Table 6.4 SPSS Pearson Correlation Coefficients: Marketing/PlanningCase Study



	A	B	C	D	E	F
1		Sales	Radio	Paper	TV	POS
2	Sales	1				
3	Radio	0.977138	1			
4	Paper	-0.28307	-0.23836	1		
5	TV	0.95797	0.966096	-0.24588	1	
6	POS	0.012649	0.060402	-0.09006	-0.03602	1

Figure 6.5 Excel Pearson correlation coefficients: marketing/planning casestudy

Although it can be argued that the positive or negative correlation coefficients should not automatically discount any variable from what will be a predictive model, the negative correlation of newspapers suggests that as a firm increases investment in newspaper ads, it will decrease product sales. This does not make sense in this case study. Given the illogic of sucha relationship, its potential use as an independent variable in a model is questionable. Also, this negative correlation poses several questions that should be considered. Was the data set correctly collected? Is the data set accurate? Was the sample large enough to have includedenough data for this variable to show a positive relationship? Should it be included for further analysis? Although it is possible that a negative relationship can statistically show up like this,it does not make sense in this case. Based on this reasoning and the fact that the correlation is not statistically significant, this variable (i.e., newspaper ads) will be removed from further consideration in this exploratory analysis to develop a predictive model.

Some researchers might also exclude POS based on the insignificance (p=0.479) of its relationship with product sales. However, for purposes ofillustration, continue to consider it a candidate for model inclusion. Also, the other two independent variables (radio and TV) were both found to be significantly related to product sales, as reflected in the correlation coefficients in the tables.

At this point, there is a dependent variable (product sales) and three candidate independent variables (POS, TV, and Radio) in which to establish a predictive model that can show the relationship between product sales and those independent variables. Just as a line chart was employed to reveal the behavior of product sales and the other variables in the descriptive analytic step, a statistical method can establish a linear model that combines the three predictive variables. We will use multiple regression, which can incorporate any of the multiple independent variables, to establish a relational model for product sales in this case study. Multiple regression also can be used to continue our exploration of the candidacy of the three independent variables.

The procedure by which multiple regression can be used to evaluate which independent variables are best to include or exclude in a linear model is called *step-wise multiple regression*. It is based on an evaluation of regression models and their validation statistics— specifically, the multiple correlation coefficients and the F-ratio from an ANOVA. SPSS software and many other statistical systems build in the step-wise process. Some are called *backward step-wise regression* and some are called *forward step-wise regression*. The backward step-wise regression starts with all the independent variables placed in the model, and the step-wise process removes them one at a time based on worst predictors first until a statistically significant model emerges. The forward step-wise regression starts with the best related variable (using correction analysis as a guide), and then step-wise adds other variables until adding more will no longer improve the accuracy of the model. The forward step-wise regression process will be illustrated here manually. The first step is to generate individual regression models and statistics for each independent variable with the dependent variable one at a time. These three models are presentedin Tables 6.5, 6.6, and 6.7 for the POS, radio, and TV variables, respectively. The comparable Excel regression statistics are presented in Tables 6.8, 6.9 and 6.10 for the POS, radio, and TV variables, respectively.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.013[a]	.000	-.055	2688.55013	.000	.003	1	18	.958

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	16649.445	1398.322		11.907	.000
	POS	44.140	822.471	.013	.054	.958

a. Dependent Variable: Sales

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20819.162	1	20819.162	.003	.958[b]
	Residual	130109432.038	18	7228301.780		
	Total	130130251.200	19			

a. Dependent Variable: Sales / b. Predictors: (Constant), POS

Table 6.5 SPSS POS Regression Model: Marketing/Planning Case Study

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.977[a]	.955	.952	571.64681	.955	380.220	1	18	.000

a. Predictors: (Constant), Radio

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-9741.921	1362.939		-7.148	
	Radio	347.689	17.831	.977	19.499	

a. Dependent Variable: Sales

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	124248209.906	1	124248209.906	380.220	.000[b]
	Residual	5882041.294	18	326780.072		
	Total	130130251.200	19			

a. Dependent Variable: Sales / b. Predictors: (Constant), Radio

Table 6.6 SPSS Radio Regression Model: Marketing/Planning Case Study

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.958[a]	.918	.913	$771.31951	.918	200.731	1	18	.000

a. Predictors: (Constant), TV

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-42229.208	4164.121		-10.141	.000
	TV	221.104	15.606	.958	14.168	.000

a. Dependent Variable: Sales

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	119421442.977	1	119421442.977	200.731	.000[b]
	Residual	10708808.223	18	594933.790		
	Total	130130251.200	19			

a. Dependent Variable: Sales/ b. Predictors: (Constant), TV

Table 6.7 SPSS TV Regression Model: Marketing/Planning Case Study

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.0126486
R Square	0.00015999
Adjusted R Square	-0.0553867
Standard Error	2688.55013
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	20819.162	20819.16	0.00288	0.957791029
Residual	18	130109432	7228302		
Total	19	130130251.2			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	16649.4448	1398.322032	11.90673	5.72E-10	13711.67923	19587.21	13711.6792	19587.21039
POS	44.14019	822.4712269	0.053668	0.957791	-1683.807738	1772.0881	-1683.80774	1772.088118

Table 6.8 Excel POS Regression Model: Marketing/Planning Case Study

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.977138075
R Square	0.954798817
Adjusted R Square	0.95228764
Standard Error	571.646807
Observations	20

ANOVA

	df	SS	MS	F	ignificance F
Regression	1	124248209.9	124248210	380.2197	1.492E-13
Residual	18	5882041.294	326780.072		
Total	19	130130251.2			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-9741.92148	1362.939419	-7.1477289	1.17E-06	-12605.35	-6878.492	-12605.351	-6878.49202
Radio	347.68885	17.83090866	19.4992222	1.49E-13	310.2275	385.150199	310.227501	385.150199

Table 6.9 Excel Radio Regression Model: Marketing/Planning Case Study

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9579703
R Square	0.917707
Adjusted R Square	0.9131352
Standard Error	771.31951
Observations	20

ANOVA

	df	SS	MS	F	ignificance F
Regression	1	119421443	1.19E+08	200.7306	3.336E-11
Residual	18	10708808.22	594933.8		
Total	19	130130251.2			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-42229.21	4164.121037	-10.1412	7.19E-09	-50977.7	-33480.714	-50977.702	-33480.7145
TV	221.10431	15.60595543	14.16794	3.34E-11	188.31741	253.8912	188.317411	253.8912023

Table 6.1O Excel TV Regression Model: Marketing/Planning Case Study

The computer printouts in the tables provide a variety of statistics for comparative purposes. Discussion will be limited here to just a few. The R-Square statistics are a precise proportional measure of the variation that isexplained by the independent variable's behavior with the dependent variable. The closer the R-Square to 1.00, the more of the variation is

explained, and the better the predictive variable. The three variables' R- Squares are 0.000

(POS), 0.955 (radio), and 0.918 (TV). Clearly, radio is the best predictor variable of the three, followed by TV and, without almost any relationship, POS. This latter result was expected based on the prior Pearson correlation. What it is suggesting is that only 0.082 percent (1.000– 0.918) of the variation in product sales is explained by TV commercials.

From ANOVA, the F-ratio statistic is useful in actually comparing the regression model's capability to predict the dependent variable. As R- Square increases, so does the F-ratio because of the way in which they are computed and what is measured by both. The larger the F-ratio (like the R- Square statistic), the greater the statistical significance in explaining the variable's relationships. The three variables' F-ratios from the ANOVA tables are 0.003 (POS), 380.220 (radio), and 200.731 (TV). Both radio and TV are statistically significant, but POS has an insignificant relationship. Togive some idea of how significant the relationships are, assuming a level of significance where $\alpha=0.01$, one would only need a cut-off value forthe F-ratio of 8.10 to designate it as being significant. Not exceeding that F-ratio (as in the case of POS at 0.003) is the same as saying that the coefficient in the regression model for POS is no different from a value of zero (no contribution to Product Sales). Clearly, the independent variables radio and TV appear to have strong relationships with the dependent variable. The question is whether the two combined or even three variables might provide a more accurate forecasting model than just using the one best variable like radio.

Continuing with the step-wise multiple regression procedure, we next determine the possible combinations of variables to see if a particular combination is better than the single variable models computed previously. To measure this, we have to determine the possible combinations for the variables and compute their regression models. The combinations are (1) POS and radio, (2) POS and TV, (3) POS, radio, and TV, and (4) radio andTV.

The resulting regression model statistics are summarized and presented inTable 6.11. If one is to base the selection decision solely on the R-Square statistic, there is a tie between the POS/radio/TV and the radio/TV combination (0.979 R-Square values). If the decision is based solely on the F-ratio value from ANOVA, one would select just the radio/TV combination, which one might expect of the two most significantly correlated variables.

To aid in supporting a final decision and to ensure these analytics are the best possible estimates, an additional statistic can be considered. That tie breaker is the R-Squared (Adjusted) statistic, which is commonly used in multiple regression models.

Variable Combination	R-Square	R-Square (Adjusted)	F-Ratio
POS/radio	0.957	0.952	188.977
POS/TV	0.920	0.911	97.662
POS/radio/TV	0.979	0.951	123.315
Radio/TV	0.979	0.953	192.555

Table 6.11 SPSS Variable Combinations and Regression Model Statistics: Marketing/Planning Case Study

The *R-Square Adjusted* statistic does not have the same interpretation as R-Square (a precise, proportional measure of variation in the relationship). It is instead a comparative measure of suitability of alternative independent variables. It is ideal for selection between independent variables in a multiple regression model. The R-Square adjusted seeks to take into account the phenomenon of the R-Square automatically increasing when additional independent variables are added to the model. This phenomenon is like a painter putting paint on a canvas, where more paint additively increases the value of the painting. Yet by continually adding paint, there comes a point at which some paint covers other paint, diminishing the value of the original. Similarly, statistically adding more variables should increasethe ability of the model to capture what it seeks to model. On the other hand, putting in too many variables, some of which may be poor predictors, might bring down the total predictive ability of the model. The R-Square adjusted statistic provides some information to aid in revealing this behavior.

The value of the R-Square adjusted statistic can be negative, but it will always be less than or equal to that of the R-Square in which it is related. Unlike R-Square, the R-Square adjusted increases when a new independentvariable is included only if the new variable improves the R-Square more than would be expected in the absence of any independent value being added. If a set of independent variables is introduced into a regression model one at a time in forward step-wise regression using the highest correlations ordered first, the R-Square adjusted statistic will end up being equal to or less than the R-Square value of the original model. By systematic experimentation with the R-Square adjusted recomputed for each added variable or combination, the value of the R-Square adjusted will reach a maximum and then decrease. The multiple regression model with the largest R-Square adjusted statistic will be the most accurate combinationof having the best fit without excessive or unnecessary independent variables. Again, just putting all the variables into a model may add unneeded variability, which can decrease its accuracy. Thinning out the variables is important.

Finally, in the step-wise multiple regression procedure, a final decision on the variables to be included in the model is needed. Basing the decision on the R-Square adjusted, the best combination is radio/TV. The SPSS multiple regression model and support statistics are presented in Table 6.12,and the Excel model is shown in Table 6.13.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.979[a]	.958	.953	568.87547	.958	192.555	2	17	.000

a. Predictors: (Constant), TV, Radio

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-17150.455	6,965.591		-2.462	.025
	Radio	275.691	68.728	.775	4.011	.001
	TV	48.341	44.580	.209	1.084	.293

a. Dependent Variable: Sales

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	124628723.140	2	62314361.570	192.555	.000[b]
	Residual	5501528.060	17	323619.298		
	Total	130130251.200	19			

a. Dependent Variable: Sales / b. Predictors: (Constant), TV, Radio

Table 6.12 SPSS Best Variable Combination Regression Model andStatistics: Marketing/Planning Case Study

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.97863319
R Square	0.95772291
Adjusted R Square	0.95274914
Standard Error	568.875468
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	124628723.1	62314362	192.5545	2.09842E-12
Residual	17	5501528.06	323619.3		
Total	19	130130251.2			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-17150.4554	6965.590997	-2.46217	0.024791	-31846.56777	-2454.343	-31846.5678	-2454.342976
Radio	275.69065	68.72801022	4.011329	0.000905	130.6872233	420.694077	130.687223	420.6940765
TV	48.3405736	44.5804165	1.084345	0.293351	-45.71588363	142.397031	-45.7158836	142.3970308

Table 6.13 Excel Best Variable Combination Regression Model andStatistics: Marketing/Planning Case Study

Although there are many other additional analyses that could be performed to validate this model, we will use the SPSS multiple regressionmodel in Table 6.12 for the firm in this case study. The forecasting model can be expressed as follows:

$$Y_p = -17150.455 + 275.691\, X_1 + 48.341\, X_2$$

where:

Y_p = the estimated number of dollars of product sales
X_1 = the number of dollars to invest in radio commercials X_2 = the number of dollars to invest in TV commercials

Because all the data used in the model is expressed as dollars, the interpretation of the model is made easier than using more complex data. The interpretation of the multiple regression model suggests that for everydollar allocated to radio commercials (represented by X_1), the firm will

receive $275.69 in product sales (represented by Y_p in the model). Likewise, for every dollar allocated to TV commercials (represented by X_2), the firm will receive $48.34 in product sales.

A caution should be mentioned on the results of this case study. Many factors might challenge a result, particularly those derived from using

powerful and complex methodologies like multiple regression. As such, the results may not occur as estimated, because the model is not reflecting past performance. What is being suggested here is that more analysis can always be performed in questionable situations. Also, additional analysis to confirm a result should be undertaken to strengthen the trust that others must have in the results to achieve the predicted higher levels of business performance.

In summary, for this case study, the predictive analytics analysis has revealed a more detailed, quantifiable relationship between the generation of product sales and the sources of promotion that best predict sales. The best way to allocate the $350,000 budget to maximize product sales might involve placing the entire budget into radio commercials because they givethe best return per dollar of budget. Unfortunately, there are constraints and limitations regarding what can be allocated to the different types of promotional methods. Optimizing the allocation of a resource and maximizing business performance necessitate the use of special business analytic methods designed to accomplish this task. This requires the additional step of prescriptive analytics analysis in the BA process, which will be presented in the last sectionof Chapter 7.

Summary

This chapter dealt with the predictive analytics step in the BA process. Specifically, it discussed logic-driven models based on experience and aided by methodologies like the cause-and-effect and the influence diagrams. This chapter also defined data-driven models useful in the predictive step of the BA analysis. A further discussion of data mining waspresented. Data mining methodology such as neural networks, discriminant analysis, logistic regression, and hierarchical clustering was described. An illustration of K-mean clustering using Excel was presented. Finally, this chapter discussed the second installment of a case study illustrating the predictive analytics step of the BA process. The remaining installment of the case study will be presented in Chapter 7.

Once again, several of this book’s appendixes are designed to augment the chapter material by including technical, mathematical, and statistical tools. For both a greater understanding of the methodologies discussed in this chapter and a basic review of statistical and other quantitative methods,a review of the appendixes is recommended.

As previously stated, the goal of using predictive analytics is to generatea forecast or path for future improved business performance. Given this predicted path, the question now is how to exploit it as fully as possible.

The purpose of the prescriptive analytics step in the BA process is to serve as a guide to fully maximize the outcome in using the information providedby the predictive analytics step. The subject of Chapter 7 is the prescriptiveanalytics step in the BA process.

Discussion Questions

1. Why is predictive analytics analysis the next logical step in any businessanalytics (BA) process?
2. Why would one use logic-driven models to aid in developing data-driven models?
3. How are neural networks helpful in determining both associations andclassification tasks required in some BA analyses?
4. Why is establishing clusters important in BA?
5. Why is establishing associations important in BA?
6. How can F-tests from the ANOVA be useful in BA?

Problems

1. Using the equation developed in this chapter for predicting dollar product sales (note below), what is the forecast for dollar product sales if the firm could invest $70,000 in radio commercials and $250,000 in TV commercials?

   Y_p = -17150.455 + 275.691 X_1 + 48.341 X_2

   where:

   Y_p = the estimated number of dollars of product sales
   X_1 = the number of dollars to invest in radio commercials X_2 = the number of dollars to invest in TV commercials
2. Using the same formula as in Question 1, but now using an investment of $100,000 in radio commercials and $300,000 in TV commercials, what is the prediction on dollar product sales?
3. Assume for this problem the following table would have held true for the resulting marketing/planning case study problem. Which combination of variables is estimated here to be the best predictor set?Explain why.

1. Assume for this problem that the following table would have held true for the resulting marketing/planning case study problem. Which of the variables is estimated here to be the best predictor? Explain why.

Image

2. Given the coefficients table that follows, what is the resulting regression

model for TV and product sales? Is TV a good predictor of product sales according to this SPSS printout? Explain.

Image

References

Abrahams, A.S., Jiao, J., Fan, W., Wang, G., Zhang, Z. (2013). "What's Buzzing in the Blizzard of Buzz? Automotive Component Isolation in Social Media Postings." *Decision Support Systems*. Vol. 55, No. 4, pp. 871–882.

Evans, J.R. (2013). *Business Analytics*. Pearson Education, Upper SaddleRiver, NJ.

Nisbet, R., Elder, J., Miner, G. (2009). *Handbook of Statistical Analysis& Data Mining Applications*, Academic Press, Burlington, MA.

Schniederjans, M.J., Cao, Q., Triche, J.H. (2014). *E-CommerceOperations Management*, 2nd ed.World Scientific, Singapore.

UNIT V PRESCRITIVE ANALYTICS

Introduction to Prescriptive analytics - Prescriptive Modeling - Non Linear Optimization - Demonstrating Business Performance Improvement.

1. What Are Prescriptive Analytics?

Chapter objectives:

- List and describe the commonly used prescriptive analytics in the businessanalytics (BA) process.
- Explain the role of case studies in prescriptive analytics.
- Explain how curve fitting can be used in prescriptive analytics.
- Explain how to formulate a linear programming model.
- Explain the value of linear programming in the prescriptive analytics stepof BA.

Introduction

After undertaking the descriptive and predictive analytics steps in the BA process, one should be positioned to undertake the final step: prescriptive analytics analysis. The prior analysis should provide a forecast or prediction of what future trends in the business may hold. For example, there may be significant statistical measures of increased (or decreased) sales, profitability trends accurately measured in dollars for new market opportunities, or measured cost savings from a future joint venture.

If a firm knows where the future lies by forecasting trends, it can best plan to take advantage of possible opportunities that the trends may offer. Step 3 of the BA process, prescriptive analytics, involves the application of decision science, management science, or operations research methodologies to make best use of allocable resources. These are mathematically based methodologies and algorithms designed to take variables and other parameters into a quantitative framework and generate an optimal or near-optimal solution to complex problems. These methodologies can be used to optimally allocate a firm's limited resources to take best advantage of the opportunities it has found in the predicted future trends. Limits on human, technology, and financial resources prevent any firm from going after all the opportunities. Using prescriptive analyticsallows the firm to allocate limited resources to optimally or near-optimally achieve the objectives as fully as possible.

In Chapter 3, "What Resource Considerations Are Important to Support Business Analytics?" the relationships of methodologies to the BA process were expressed as a function of certification exam content. The listing of the prescriptive analytic methodologies as they are in some cases utilized inthe BA process is again presented in Figure 7.1 to form the basis of this chapter's content.

Image

Figure 7.1 Prescriptive analytic methodologies

Prescriptive Modeling

The listing of prescriptive analytic methods and models in Figure 7.1 is buta small grouping of many operations research, decision science, and management science methodologies that are applied in this step of the BA process. The explanation and use of most of the methodologies in Table 7.1 are explained throughout this book. (See Additional Information column inTable 7.1.)

Table 7.1 Select Prescriptive Analytic Models

Nonlinear Optimization

The prescriptive methodologies in Table 7.1 are explained in detail in the referenced chapters and appendixes, but nonlinear optimization will be discussed here. When business performance cost or profit functions becometoo complex for simple linear models to be useful, exploration of nonlinear functions is a standard practice in BA. Although the predictive nature of

exploring for a mathematical expression to denote a trend or establish a forecast falls mainly in the predictive analytics step of BA, the use of thenonlinear function to optimizea decision can fall in the prescriptive analytics step.

As mentioned previously, there are many mathematical programing nonlinear methodologies and solution procedures designed to generate optimal business performance solutions. Most ofthem require careful estimation of parameters that may or may not be accurate, particularly given the precision required of a solution that can be so precariously dependent upon parameter accuracy. This precision is further complicated in BA by the large data files that should be factored into the model-buildingeffort.

To overcome these limitations and be more inclusive in the use of large data, regression software can be applied. As illustrated in Appendix E, Curve Fitting software can be used to generate predictive analytic modelsthat can also be utilized to aid in making prescriptive analytic decisions.

For purposes of illustration, SPSS's Curve Fitting software will be used in this chapter. Suppose that a resource allocation decision is being faced whereby one must decide how many computer servers a service facility should purchase to optimize the firm's costs of running the facility. The firm's predictive analytics effort has shown a growth trend. A new facilityis called for if costs can be minimized. The firm has a history of setting uplarge and small service facilities and has collected the 20 data points in Figure 7.2. Whether there are 20 or 20,000 items in the data file, this SPSS function fits the data based on regression mathematics to a nonlinear line that best minimizes the distance from the data items to the line. The software then converts the line into a mathematical expression useful for forecasting.

Image

Figure 7.2 Data and SPSS Curve Fitting function selection window

In this server problem, the basic data has a u-shaped function, as presentedin Figure 7.3. Thisis a classic shape for most cost functions in business. In this problem, it represents the balancing of having too few servers (resulting in a costly loss of customer business through dissatisfaction and complaints with the service) or too many servers (excessive waste in investment costs as a result of underutilized servers). Although this is an overly simplified example with little and nicely ordered data for clarity purposes, in big data situations, cost functions are considerably less obvious.

Image

Figure 7.3 Server problem basic data cost function

The first step in using the curve-fitting methodology is to generate the best-fitting curve to the data. By selecting all the SPSS models in Figure 7.2, thesoftware applies each point of data using the regression process of minimizing distance from a line. The result is a series of regression models and statistics, including ANOVA and other testing statistics. It is known from the previous illustration of regression that the adjusted R-Square statistic can reveal the best estimated relationship between the independent (number of servers) and dependent (total cost) variables. These statistics are presented in Table 7.2. The best adjusted R-Square value (the largest) occurs with the quadratic model, followed by the cubic model. The more detailed supporting statistics for both of these models are presented in Table

The graph for all the SPSS curve-fitting models appears in Figure 7.4.Image

From Table 7.3, the resulting two statistically significant curve-fittedmodels follow:

$Y_p = 35417.772 - 5589.432\ X + 268.445\ X^2$ [Quadratic model]

$Y_p = 36133.696 - 5954.738\ X + 310.895\ X^2 - 1.347\ X^3$ [Cubic model]

where:

Y_p = the forecasted or predicted total cost, and

X = can be the number of computer servers.

For purposes of illustration, we will use the quadratic model. In the next step of using the curve-fitting models, one can either use calculus to derive the cost minimizing value for X (number of servers) or perform a deterministic simulation where values of X are substituted into the model to compute and predict the total cost (Y_p). The calculus-based approach is presented in the "Addendum" section of this chapter.

As a simpler solution method to finding the optimal number of servers, simulation can be used. Representing a deterministic simulation (see Appendix F, Section F.2.1), the resulting costs of servers can be computed using the quadratic model, as presented in Figure 7.5. These values were computed by plugging the number of server values (1 to 20) into the Y_p quadraticfunction one at a time to generate the predicted values for each ofthe server possibilities. Note that the lowest value in these predicted values occurs with the acquisition of 10 servers at $6357.952, and the next lowest is at 11 servers at $6415.865. In the actual data in Figure 7.2, the minimumtotal cost point occurs at 9 servers at $4533, whereas the next lowest total cost is
$4678 occurring at 10 servers. The differences are due to the estimation process of curve fitting. Note in Figure 7.3 that the curve that is fitted does not touch the lowest 5 cost values. Like regression in general, it is an estimation process, and although the ANOVA statistics in the quadratic model demonstrate a strong relationship with the actual values, there is some error. This process provides a near-optimal solution but does not guarantee one.

Figure 7.5 Predicted total cost in server problem for each server alternative

Like all regression models, curve fitting is an estimation process and has risks, but the supporting statistics, like ANOVA, provide some degree of confidence in the resulting solution.

Finally, it must be mentioned that many other nonlinear optimization methodologies exist. Some, like quadratic programming, are considered constrained optimization models (like LP). These topics are beyond the scope of this book. For additional information on nonlinear programming, see King and Wallace (2013), Betts (2009), and Williams (2013). Other methodologies, like the use of calculus in this chapter, are useful in solving for optimal solutions in unconstrained problem settings. For additional information on calculus methods, see Spillers and MacBain (2009), Luptacik (2010), and Kwak and Schniederjans (1987).

Continuation of Marketing/Planning Case Study Example: Prescriptive Stepin the BA Analysis

In Chapter 5, "What Are Descriptive Analytics?" and Chapter 6, "What Are Predictive Analytics?" an ongoing marketing/planning case study was presented to provide an illustration of some of the tools and strategies usedin a BA problem analysis. This is the third and final installment of the casestudy dealing with the prescriptive analytics step in BA.

Case Background Review

The predictive analytics analysis in Chapter 6 revealed a statistically strong relationship between radio and TV commercials that might be useful in predicting future product sales. The ramifications of these results suggest a better allocation of funds away from paper and POS ads to radio and TV commercials. Determining how much of the $350,000 budget should be allocated between the two types of commercials requires the application of an optimization decision-making methodology.

Prescriptive Analysis

The allocation problem of the budget to purchase radio and TV commercials is a multivariable (there are two media to consider), constrained (there are some limitations on how one can allocate the budget funds), optimization problem (BA always seeks to optimize business performance). Many optimization methods could be employed to determinea solution to this problem. Considering the singular objective of maximizing estimated product sales, linear programming (LP) is an ideal methodology to apply in this situation. To employ LP to model this problem, use the six-step LP formulation procedure explained in Appendix B.

Formulation of LP Marketing/Planning Model

In the process of exploring the allocation options, a number of limitations or constraints on placing radio and TV commercials were observed. The total budget for all the commercials was set at a maximum of $350,000 for the next monthly campaign. To receive the price discount on radio commercials, a minimum budget investment in radio of $15,000 is required, and to receive the price discount on TV commercials, a minimum of $75,000 is necessary. Because the radio and TV stations are owned by the same corporation, there is an agreement that for every dollar of radio commercials required, the client firm must purchase $2 in TV commercials. Given these limitations and the modeled relationship found in the previous predictive analysis, one can formulate the budget allocation decision as an LP model using a five-step LP formulation procedure (see Appendix B, Section B.4.1):

1. **Determine the type of problem**—This problem seeks to maximize dollar product sales by determining how to allocate budget dollars over radio and TV commercials. For each dollar ofradio commercials estimated with the regression model, $275.691 will be received, and for each dollar of TV commercials, $48.341 will be received. Those two parameters are the product sales values to maximize. Therefore, it will be a maximization model.

2. **Define the decision variables**—The decision variables for the LP modelare derived from the multiple regression model's independent variables. The only adjustment is the monthly timeliness of the allocation of the budget:

X_1= the number of dollars to invest in radio commercials for the next monthly campaign

X_2=the number of dollars to invest in TV commercials for the next monthlycampaign

3. **Formulate the objective function**—Because the multiple regression model defines the dollar sales as a linear function with the two independent variables, the same dollar coefficients from the regression model can be used as the contribution coefficients in the objective function. This results in the following LP model objective function:

Maximize: $Z = 275.691 X_1 + 48.341 X_2$

4. **Formulate the constraints**—Given the information on the limitations in this problem, there are four constraints:

Constraint 1—No more than $350,000 is allowed for the total budget to allocate to both radio (X_1) and TV (X_2) commercials. So add $X_1 + X_2$ and set it less than or equal to 350,000 to formulate the first constraint as follows:

$X_1 + X_2 \leq 350000$

Constraint 2—To get a discount on radio (X_1) commercials, a minimum of $15,000 must be allocated to radio. The constraint for this limitationfollows:

$X_1 \geq 15000$

Constraint 3—Similar to Constraint 2, to get a discount on TV (X_2)commercials, aminimum of $75,000 must be allocated to TV. The
constraint for this limitation follows:

$X_2 \geq 75000$

Constraint 4—This is a blending problem constraint (see Appendix B,Section B.6.3). What is needed is to express the relationship as follows:

Imag

which is to say, for each one unit of X_1, one must acquire two units of X_2.Said differently, the ratio of one unit of X_1 is equal to two units of X_2.

Given the expression, use algebra to cross-multiply such that:$2 X_1 = X_2$

Convert it into an acceptable constraint with a constant on the right sideand the variableson the left side as follows:

$2 X_1 - X_2 = 0$

5. State the Nonnegativity and Given Requirements—With only two variables, thisformal requirement in the formulation of an LP model is expressed as follows:

$X_1, X_2 \geq 0$

Because these variables are in dollars, they do not have to be integer values. (They can beany real or cardinal number.) The complete LP modelformulation is given here:

Imag

Solution for the LP Marketing/Planning Model

From Appendix B, one knows that both Excel and LINGO software can be used to run the LP model and solve the budget allocation in this marketing/planning case study problem. For purposes of brevity, discussion

will be limited to just LINGO. As presented in Appendix B, *LINGO* is a mathematical programming language and software system. It allows the fairly simple statement of the LP model to be entered into a single windowand run to generate LP solutions.

LINGO opens with a blank window for entering whatever type of model is desired. After entering the LP model formulation into the LINGO software, the resulting data entry information is presented in Figure 7.6.

Image

Figure 7.6 LINGO LP model entry requirements: marketing/planning casestudy

There are several minor differences in the model entry requirements over the usual LP model formulation. These differences are required to run a model in LINGO. These include (1) using the term “Max” instead of “Maximize,” (2) dropping off “Subject to” and “and” in the model formulation, (3) placing an asterisk and a space between unknowns and constant values in the objective and constraint functions where multiplication is required, (4) ending each expression with a semicolon, and

(5) omitting the nonnegativity requirements, which aren’t necessary.

Having entered the model into LINGO, a single click on the SOLVE optionin the bar at the top of the window generates a solution. The marketing budget allocation LP model solution isfound in Figure 7.7.

Image

Figure 7.7 LINGO LP model solution: marketing/planning case study

As it turns out, the optimal distribution of the $350,000 promotion budget is to allocate $116,666.70 to radio commercials and $233,333.30 to TV commercials. The resulting Z value, which in this model is the total predicted product sales in dollars, is 0.4344352E+08, or $43,443,524.

Comparing that future estimated month's product sales with the average current monthly product sales of $16,717,200 presented in Figure 7.7, it does appear that the firm in this case study will optimally maximize future estimated monthly product sales if it allocates the budget accordingly (that is, if the multiple regression model estimates and the other parameters in the LP model hold accurate and true).

In summary, the prescriptive analytics analysis step brings the prior statistical analytic steps into an applied decision-making process where a potential business performance improvement is shown to better this organization's ability to use its resources more effectively. The management job of monitoring performance and checking to see that business performance is in fact improved is a needed final step in the BA analysis. Without proof that business performance is improved, it's unlikelythat BA would continue to be used.

Final Comment on the Marketing/Planning Model

Although the LP solution methodology used to generate an allocation solution guarantees an optimal LP solution, it does not guarantee that the firm using this model's solution will achieve the results suggested in the analysis. Like any estimation process, the numbers are only predictions, notassurances of outcomes. The high levels of significance in the statistical analysis and the added use of other conformational statistics (R-Square, adjusted R-Square, ANOVA, and so on) in the model development providesome assurance of predictive validity. There are many other methods and approaches that could have been used in this case study. Learning how to use more statistical and decision science tools helps ensure a better solutionin the final analysis.

Summary

This chapter discussed the prescriptive analytics step in the BA process. Specifically, this chapter revisited and briefly discussed methodologies suggested in BA certification exams. An illustration of nonlinear optimization was presented to demonstrate how the combination of software and mathematics can generate useful decision-making information. Finally, this chapter presented the third installment of a marketing/planning case study illustrating how prescriptive analytics canbenefit the BA process.

We end this book with a final application of the BA process. Once again, several of the appendixes are designed to augment this chapter's content by including technical, mathematical, and statistical tools. For both a greater understanding of the methodologies discussed in this chapter and a basic review of statistical and other quantitative methods, a review of the appendixes and chapters is recommended.

Addendum

The *differential calculus* method for finding the minimum cost point on thequadratic function that follows involves a couple of steps. It finds the zero slope point on the cost function (the point at the bottom of the u-shaped curve where a line could be drawn that would have a zero slope). There are limitations to its use, and qualifying conditions are required to prove minimum or maximum positions on a curve. The quadratic model in the server problem follows:

$Y_p = 35417.772 - 5589.432\ X + 268.445\ X^2$ [Quadratic model]

Step 1. Given the quadratic function above, take its first derivative: $d(Y_p) = -5589.432 + 536.89\ X$

Step 2. Set the derivative function equal to zero and solve for X.$0 = -5589.432 + 536.89\ X$

$X = 10.410758$

Slightly more than ten servers should be purchased at the resulting optimally minimized cost value. This approach provides a near-optimalsolution but does not guarantee one. For additional information on the application of calculus, see Field, M.J. (2012) and Dobrushkin, V.A. (2014).

Discussion Questions

1. How are prescriptive and descriptive analytics related?

2. How can we use simulation in both predictive and prescriptive analytics?

3. Why in the server problem were there so few statistically significantmodels?

4. Does it make sense that the resulting quadratic model in Figure 7.4 did not touch thelowest cost data points in the data file? Explain.

5. What conditions allowed the application of LP?

Problems

1. A computer services company sells computer services to industrial users. The company's analytics officer has predicted the need for growth to meet competitive pressures. To implement this strategy, upper management has determined that the company would tactically expand its sales and service organization. In this expansion, new districts would be defined and newly hired or appointed managers would be placed in charge to establish and runthe new districts. The first job of the new district managers would be to select the sales people and staff support employees for their districts. To aidthe new district managers in deciding on the number of sales people and staffers to hire, the company researched existing office operations and made a number of analytic-based observations, which they passed on to thenew district managers. A new manager's district should, at the very least, have 14 sales people and 4 staffers to achieve adequate customer service. Research has indicated that a district manager could adequately manage theequivalent of no more than 32 employees. Sales people are twice as time consuming to manage as staffers. The district manager was assigned part ofthe floor inan office building for operations. This space could house no more than 20 sales people and staffers. The district manager had some discretion regarding budgetary limitations. A total payroll budget for sales people and staffers was set at $600,000. The manufacturing company's policy in developing a new territory would be to pay sales people a fixed salary instead of commissions and salary. The yearly salary of a beginning sales person would be $36,000, whereas a staffer would receive $18,000.
All the sales people and staffers being hired for this district would be newwith the company, and as such, would start with the basic salaries mentioned. Finally, the source of prospective sales people and staffers would be virtually unlimited in the district and pose noconstraint on the problem situation. What is the LP formulation of this model?

2. (This problem requires computer support.) What is the optimal answer to the problem formulated in Problem 1?

3. A trucking firm must transport exactly 900, 800, 700, and 1,000 units ofa product to four cities: A, B, C, and D. The product is manufactured and supplied in two other cities, X and Y, in the exact amounts to match the total demand. The production of units from the two cities is 1,900 and 1,500 units, respectively, to X and Y. The cost per unit to transport the
product between the manufacturing plants in cities X and Y and the demand market cities A, B, C, and D are given here:

Image

For example, in the table, $0.65 is the cost to ship one unit from Supply Plant X to Demand Market A. The trucking firm needs to know how many units should be shipped from each supply city to each demand city in sucha way that it minimizes total costs. Hint: This is a multidimensional decision variable problem (see Section B.6.4 in Appendix B). What is the LP model formulation for this problem?

4. (This problem requires computer support.) What is the optimal answer to the problem formulated in Problem 3?

References

Adkins, T.C. (2006). C*ase Studies in Performance Management: A Guidefrom the Experts*. Wiley, New York, NY.

Albright, S.C., Winston, W.L. (2014). *Business Analytics: Data Analysis &Decision Making*. Cengage Learning, Stamford, CT.

Betts, J.T. (2009). *Practical Methods for Optimal Control and EstimationUsing NonlinearProgramming*, 2nd ed., Society for Industrial & AppliedMathematics. London.

Cooper, W.W., Seiford, L.M., Zhu, J. (2013). *Handbook on Data EnvelopmentAnalysis*. Springer, New York, NY.

Dobrushkin, V.A. (2014). *Applied Differential Equations: An Introduction*. Chapman andHall/CRC, New York, NY.

Field, M.J. (2012). *Differential Calculus and Its Applications*. Dover Publishing,Mineola, NY.

Hillier, F.S. (2014). *Introduction to Operations Research*, 10th ed., McGraw-HillHigher Education, Boston, MA.

King, A.J., Wallace, S.W. (2013). *Modeling with Stochastic Programming*, Springer, NewYork, NY.

Kwak, N.K, Schniederjans, M.J. (1987). *Introduction to Mathematical Programming*. Kreiger Publishing, Malabar, FL.

Liebowitz, J. (2014). *Business Analytics: An Introduction*. Auerbach Publications,New York, NY.

Luptacik, M. (2010). *Mathematical Optimization and Economic* Analysis. Springer, NewYork, NY.

Rothlauf, F. (2013). *Design and Modern Heuristics: Principles and Application*.Springer, New York, NY.

Sekaran, U., Bougie, R. (2013). *Research Methods for Business: A Skill-BuildingApproach.* Wiley, New York, NY.

Spillers, W.R., MacBain, K.M. Williams, H.P. (2013). *Modeling Building in Mathematical Programming*.Wiley, New York,NY.

8. A Final Business Analytics Case Problem

Chapter objectives:

• Provide a capstone business analytics (BA) overview within a case studyproblem.

• Show the step-wise connections of the descriptive, predictive, and prescriptivesteps in the BA process.

Introduction

In Parts I, "What Are Business Analytics?" and II, "Why Are Business Analytics Important?" (Chapters 1 through 3), this book explained what BA is about and why it is important to business organization decision- making. In Part III, "How Can Business Analytics Be Applied?" (Chapters4 through 7), we explained and illustrated how BA can be applied using a variety of different concepts and methodologies. Completing Part III, we seekin this chapter a closing illustration of how the BA process can be applied by presenting a final case study. This case study is meant as a capstone learning experience on the business analytics process discussed throughout the book. Several of the concepts and methodologies presentedin prior chapters and the appendixes will once again be applied here.

As will be seen in this case study, unique metrics and measures are sometimes needed in a BA setting to affect a solution to a problem or answer a question. Therefore, the methodologies and approach used in this chapter should be viewed as just one approach in obtaining the desired information.

Undertaking the analytic steps in the BA process (see Chapter 1, "What Are Business Analytics?") requires a beginning effort that preempts data collection efforts. This prerequisite to BA is to understand the business systems that are a part of the problem. When BA effort has been outsourced(see Chapter 4, "How Do We Align Resources to Support Business Analytics within an Organization?") or when it is completely performed in house by a BA team (Chapter 3, "What Resource Considerations Are Important to Support Business Analytics?"), experienced managers must bebrought into the process to provide the necessary systems behavior and general knowledge of operations needed to eventually model and explain how the business operates. In this case study, it is assumed that the staff or information is available. Based on this information, a BA project can be undertaken.

Case Study: Problem Background and Data

A Midwest US commercial manufacturing firm is facing a supply chain problem. The manufacturer produces and sells a single product, a general- purpose small motor as a component part to different customers who incorporate the motor into their various finished products. The manufacturer has a supply chain network that connects production centers located in St. Louis, Missouri, and Dallas, Texas, with six warehouse facilities that serve commercial customers located in Kansas City, Missouri; Chicago, Illinois; Houston, Texas; Oklahoma City, Oklahoma; Omaha, Nebraska; and Little Rock, Arkansas.

Part of the supply chain problem is the need to keep the cost of shipping motors to the customers as low as possible. The manufacturer adopted a lean management philosophy that seeks to match what it produces with what is demanded at each warehouse. The problem with implementing this philosophy is complicated by the inability to forecast the customer demand month to month. If the forecast of customer demand is too low and not enough inventory is available (an underage of inventory), the manufacturer has to rush order motors that end up being costly to the manufacturer. If theforecast is too high and the manufacturer produces and ships unwanted inventory (an overage of inventory), the warehouse incurs wasteful storage costs. The management of the manufacturing firm has decided that an analytics-based procedure needs to be developed to improve overall business performance. This would be a procedure that analysts could use each month to develop an optimal supply chain schedule of shipments fromthe two supply centers to the six warehouse demand destinations that would minimize costs. A key part of this procedure would be to include a means to accurately forecast customer demand and an optimization process for shipping products from the manufacturing centers to the warehouse demand destinations.

The manufacturing firm created a small BA team to develop the procedure (see Chapter 4, Section 4.1.1). The BA team consists of a BA analyst (whowould be responsible for using the procedure and heads the BA team), the supply chain general manager, the shipping manager (responsible for drafting the shipping schedule), and a warehouse manager (whose job it is to develop monthly forecasts).

Descriptive Analytics Analysis

Determining a procedure by which analyst teams can determine optimal shipments between supply sources and demand destinations requires differing types of data. There is supply, demand, and cost data required to plan shipments. The total manufactured supply of motors produced at the St. Louis and Dallas plants is determined once the forecast demand is established. The BA team established that there is ample capacity between both plants to satisfy the forecasted customer demand at the six warehousedemand destinations.

The BA team determined that the cost data for shipping a motor from the production centers to the customers depends largely on distance between the cities, where the items are trucked directly by the manufacturer to the warehouses. The cost data per motor shipped to a customer is given in Table 8.1. For example, it costs the manufacturer $4 per motor to ship from St. Louis to Kansas City. These cost values are routinely computed by the manufacturer's cost accounting department and are assumed by the BA team to be accurate.

Image

Table 8.1 Estimated Shipping Costs Per Motor

The present system of forecasting customer demand usually results in costly overages and underages shipped to the warehouses. In the past, themanufacturer would take a three valuesmoothing average to estimate the monthly demand. (See Section E.6.1 in Appendix E, "Forecasting.") This

evolved by taking the last three months of actual customer motor demand and averaging them to produce a forecast for the next month. The process was repeated each month for each of the six warehouses. Not making products available when customers demanded them caused lost sales, so the manufacturer would rush and ship products to customers at a loss. On the other hand, producing too much inventory meant needless production, inventory, and shipping costs.

To deal with the variability in customer demand forecasting, models for each warehouse's customer demand would need to be developed. The customer demand data on which to build the models was collected from prior monthly demand in motors. To determine which data to include in a final sample and which to exclude, a few simple rules were adopted to eliminate potentially useless and out-of-date data. Going back more than 27 months invited cyclical variations caused by changes in the economy that were no longer present, so that data was removed. Unfortunately, some of the data files were incomplete and required cleansing (see Chapter 4). The resulting time series data collected on warehouse customer monthly demand files is presented in Table 8.2. It was decided that the most recent three months (darkened months of 25, 26, and 27) would not be included inthe model development, but instead would be used for validation purposes to confirm the forecasting accuracy of the resulting models. This is similar to what was referred to as a training data set and a validation data set (see Section 6.3.1 in Chapter 6, "What Are Predictive Analytics?").

Image
Image

Table 8.2 Actual Monthly Customer Demand in Motors

As a part of the descriptive analysis, summary statistics were generated from both Excel (Table 8.3) and SPSS (Table 8.4). The mean values provide some basis for a monthly demand rate, but at this point consideration of overall behavior within data distributions is required to more accurately capture relevant variation. To that end, other statistics can provide some picture of the distribution of the data. For example, the Kurtosis coefficient (see Chapter 5, "What Are Descriptive Analytics?") forOmaha's demand suggests a peaked distribution. This indicates that the variance about the mean is closely grouped toward the mean, implying a lack of variability in forecast values (a good thing). Note that the Standard Error statistic (see Chapter 5, Section 5.3) for Omaha is the smallest. Other statistics such as the Skewedness Coefficients suggest most of the distributions are negatively skewed. The median value peaks at a larger value than the mean and implies that the mean and mean-related statistics might not be as accurate in measuring the entire distribution's behavior as other measures (like the median).

Image

Table 8.3 Excel Summary Statistics of Actual Monthly Customer Demandin Motors

Image

Table 8.4 SPSS Summary Statistics of Actual Monthly Customer Demandin Motors

To better depict the general shape of the data and to understand their behavior, line graphs (see Chapter 5, Section 5.2) of the six customer demand files are graphed using SPSS in Figures 8.1 to 8.6. (The Excel versions look the same and will not be displayed.) As expected based on the summary statistics and now visually from the graphs, some of the customer demand functions look fairly linear, others are clearly nonlinear, and some possess so much variation they are unrecognizable. Identifying the almost perfect linear customer demand behavior in the warehouses in Chicago (Figure 8.2) and Oklahoma City (Figure 8.4) suggests the use of a simple linear regression model for forecasting purposes. The very clear, bell-shaped, nonlinear functions for Houston (Figure 8.3) and Little Rock (Figure 8.6) suggest that a nonlinear regression model should be determined by the BA team to find the best- fitting forecasting model.

Finally, the excessively random customer demand behavior for Kansas City (Figure 8.1) and Omaha (Figure 8.5) suggests that considerable effort is needed to find a model that may or may not explain the variation in the datawell enough for a reliable forecast. There appear to be many time series variations (see Appendix E, Section E.2) in customer demand for the warehouses in these two cities.

Figure 8.1 Graph of Kansas City customer

demandImage**Figure 8.2** Graph of Chicago customer demandImage **Figure 8.3** Graph of Houston customer

demandImage

Figure 8.4 Graph of Oklahoma City customer demandImage

Figure 8.5 Graph of Omaha customer demandImage

Figure 8.6 Graph of Little Rock customer demand

The fact that two of the four warehouse time series data files have more time series variations than the other four warehouse files does not prevent in this case (and in most others) a fairly accurate forecast. Because four of the six customer demand warehouses appear to have a fairly observable pattern of behavior, they will help improve the overall accuracy even withthe substantial variations of the other two warehouses adding in some forecast error.

Predictive Analytics Analysis

In this section, we continue with our illustrative example. Here we use the predictive analytics analysis step that requires model development effort and then model validation for the example. To complete the predictive analytics analysis, forecasts of warehouse demand are determined.

Developing the Forecasting Models

The descriptive analytics analysis has suggested a course of action in identifying appropriate forecasting models in this next step of the BA process. To ensure the best possible forecasting models and confirm the descriptive analytics analysis results, the curve-fitting feature (Curve Estimation function) of SPSS will be utilized. Each of the six customerdemand data files is analyzed through the SPSS program to generate potential regression models, as presented in Tables 8.5 through 8.10.

Image

Table 8.5 SPSS Curve-Fitting Analysis for Kansas City Motor Demand Forecasting Model: Model Summary and Parameter Estimates

Image

Table 8.6 SPSS Curve-Fitting Analysis for Chicago Motor Demand Forecasting Model: Model Summary and Parameter Estimates

Image

Table 8.7 SPSS Curve-Fitting Analysis for Houston Motor Demand Forecasting Model: Model Summary and Parameter Estimates

Image

Table 8.8 SPSS Curve-Fitting Analysis for Oklahoma City Motor Demand Forecasting Model: Model Summary and Parameter Estimates

Image

Table 8.9 SPSS Curve-Fitting Analysis for Omaha Motor Demand Forecasting Model: Model Summary and Parameter Estimates

Table 8.1O SPSS Curve-Fitting Analysis for Little Rock Motor Demand ForecastingModel: Model Summary and Parameter Estimates

Reviewing the R-Square values for each of the potential curve-fitting models, it turns out that the cubic model is the best fitting for all six data files. It is not surprising that in the cases of Houston and Little Rock, wherethe descriptive analytics graphs clearly show typical cubic (or quadratic) function behavior, that the only significant (F-ratio, p<.000) models were cubic or quadratic (see Chapter 6, Section 6.4.2). In other cases (Chicago and Oklahoma City), it is surprising that a nonlinear cubic model does a slightly better job than the descriptive analytics step linear model. On the other hand, note that for both locations, the linear model, according to the R-Square statistics, is either the next best choice or the next to the next best choice. Indeed, in both cases, the F-ratio clearly shows that the resulting linear model can provide a statistically significant forecasting capability.
Other models also have significant (p<.000) F-ratios, suggesting the possibility of accurate forecasting. Because the objective of this case study is to develop a procedure that analysts could use each month to develop an optimal supply chain schedule of shipments and to

accurately forecast customer demand, the BA analyst can use the highest R-Square statistic asa means to determine the most accurate forecasting model from those fittedwith the data.

In this case study the resulting cubic regression models estimated by the SPSS program based on the parameters from the curve-fitting effort are presented in Table 8.11.

Image

Table 8.11 Resulting Cubic Forecasting Models from SPSS Curve-FittingAnalysis

The generalized formula for a cubic regression model follows: $Y_p = a + b_1 X + b_2 X^2 + b_3 X^3$

By inserting the curve-fitted parameters for Little Rock, the resulting cubicregression model for forecasting warehouse customer demand is this:

$Y_p = 4426.153 + 100.999 X - 6.347 X^2 + 0.801 X^3$

where:

X = month number in the form of the time series data file (26, 27, and 28)

Validating the Forecasting Models

One of the fundamental requirements of a BA analysis is to show or prove the possibility of improving business performance (see Chapter 1, Section 1.1). One criterion for improving forecasting is to improve forecasting accuracy. To compare the current forecasting method with the newly devised one, each cubic model is used to forecast the respective location of customer demand. Substituting the numbered time values (25, 26, and 27) for X in each cubic model, the analyst is able to compute the three forecast values. These forecasts are then compared with the actual values in Table 8.1. The resulting comparison is expressed in the MAD statistics (see Appendix E, Section E.8), as presented in Table 8.12.

Image

Table 8.12 Resulting Cubic Model Forecasts and MAD Statistics (RoundedUp to Next Integer Value)

The MAD statistics for all warehouse facilities except Oklahoma City are extremely small, suggesting the cubic models for these locations are very accurate. The Oklahoma City MAD statistic, on the other hand, is so great relative to the other MADs that it suggests further analysis is needed to finda better forecasting model for Oklahoma City.

To explore this forecast exception in Oklahoma City, the next two best models (based on R-Square) from the SPSS curve-fitting effort in Table 8.8are examined. These two include the linear regression model (R-Square 0.986):

$Y_p = 15229.746 - 488.963X$

and the quadratic regression model (R-Square 0.987): $Y_p =$

$15420.511 - 532.986\,X + 1.761\,X^2$

In Table 8.13, the resulting warehouse forecasts of customer demand for each of the three years are presented along with their MAD statistics for both the linear and the quadratic models. Clearly, the linear regression model's small MAD suggests that it is the better model for forecasting than either the quadratic or the cubic models. This result is not surprising, giventhe prior descriptive analytics step, which appeared to suggest that a linear model would be the best type of forecasting model.

Image

Table 8.13 Resulting Linear and Quadratic Forecasts with MAD Statistics for Oklahoma City (Rounded Up to the Next Integer Value)

Having found the models that provide low error rates, it is now necessary to validate them by demonstrating they can improve forecasting accuracy and, therefore, enhance business performance by minimizing costly shipping efforts.

To validate the forecasting accuracy and demonstrate forecasting improvement of the cubic and linear models, a comparison with the currently used smoothing average method is undertaken by the analyst. Utilizing a similar smoothing average formula to that mentioned in

Section

E.6 in Appendix E, the forecast values for warehouse customer demand canbe computed using the simple formula below:

Image = $(Y_{t-1} + Y_{t-2} + Y_{t-3})/3$where:

Image = the forecast value in time period *t*

Y_{t-1} = the actual value in the time period just prior to time period *t*

Y_{t-2} = the actual value of two time periods prior to time period *t*

Y_{t-3} = the actual value of three time periods prior to time period *t*

Using the formula, the resulting smooth average forecast values are presented in Table 8.14 along with their respective MAD statistics.

Image

Table 8.14 Resulting Smooth Average Forecasts and MAD Statistics(Rounded Up to Next Integer Value)

These smoothing average forecasts and their MADs can be compared with the forecasts and MADs for the cubic and linear models. Comparing the MADs in Table 8.14 with the MADs in Tables 8.12 and 8.13, several pointsabout forecasting improvement can be made. In the case ofthe locations forKansas City, Chicago, Houston, and Little Rock, the cubic regression models are the lowest; therefore, they have more accurate forecasting results. For those four locations, the cubic model is recommended. In the case of Oklahoma City, the linear regression model results in the lowest MAD value, reflecting improved forecasting accuracy over the other models. Finally, the MADs for both the cubic regression (Table 8.12) and the smoothing average methods (Table 8.14) result in the same MAD value(1.66) for Omaha, which suggests either method is accurate in forecasting this location's customer demand. Because either method can be used, the manufacturer's BA analyst selected to employ the cubic regression model for forecasting Omaha's warehouse customer demand.

Resulting Warehouse Customer Demand Forecasts

The selected forecasting models and their forecast values for the future 28th month (X = 28 in the models) are presented in Table 8.15. The direction of movement from the 27th to the forecast of the 28th month appears mainly to be downward for most of the warehouse locations. The resulting forecast values for the six locations are generally consistent with graphs from the descriptive analytics analysis, although some of the timeseries variation behavior (for example, Kansas City) can hardly bepredicted up ordown in movement.

Image

Table 8.15 Resulting Forecasts for the Future 28th Month (Rounded Up to Next Integer Value)

Prescriptive Analytics Analysis

Based on the predictive analytics analysis, the total forecast demand (see Table 8.15) of 21,826 motors for all six warehouse locations has to be balanced out by product capacity of the two production facilities. The BA team decided that the St. Louis production center would produce 10,000 motors for the 28th month, and the Dallas production center would produce the remaining 11,826 motors.

Selecting and Developing an Optimization Shipping Model

In terms of data, the analyst now possesses the supply, forecast demand, and cost information on which to begin selecting a modeling approach to achieve an optimal shipping schedule. Reviewing the requirements of the problem setting at this point in the analysis, the BA team is looking at a multivariable (number of motors to ship from two supply sources to three demand destinations), multidimensional (scheduling motor shipments fromtwo supply sources to six demand markets or supply and demand), constrained (the exact number of motors required is deterministic at this point), integer (shipping whole motors, not motor parts), and optimal solution (seeking to minimize cost of shipping). The ideal BA methodology to satisfy these requirements is *integer programming* (IP). (See Appendix D, "Integer Programming.")

To conceptualize a two-dimensional problem, a *transportation method* (an operations research methodology) table that combines location cost per unit shipped and supply and demand information is presented. The table for this case study problem is shown in Table 8.16. The decision variables for the model are also added. So, for example, X_{11} represents the number of

manufacturer motors to ship from the St. Louis production center to meet the forecast customer demand at the Kansas City warehouse. In this table, the sum of the motors produced in St. Louis and shipped to any of the sixcustomer demand locations must add up to 10,000. Likewise, for Dallas, the shipments must equal 11,826. Also, for each column, the sum of the motors shipped must equal the forecast demand in that column. For example, the sum of X_{11} and X_{21} for the Kansas City warehouse must equal the forecast demand of 2,923 motors.

Image

Table 8.16 Transportation Method Table for Conceptualization of Supply Chain Shipping Problem

With the transportation method table as a framework, the IP model can be developed. In this type of shipping problem, there are two supply-side constraints and six demand-side constraints required to ensure the supply is allocated to meet the demand. The same formulation procedure for LP models in Appendix B, "Linear Programming," and for integer programming in Appendix D is applied here to generate the following integer model:

Image

Determining the Optimal Shipping Schedule

To run this model, LINDO (see Appendix B, Section B.5.3) software is utilized. As it turns out, in this situation the unique formulation of the transportation method model mathematically forces an all-integer solution without the need for using the IP software algorithm. This permits the regular LP software to be used to solve this problem, although LINDO has both IP and LP solution software. The LINDO LP model input is presented in Figure 8.7, and the results are presented in Figure 8.8.

Image

Figure 8.7 LINDO input for supply chain shipping model problem

Image

Figure 8.8 LINDO output for supply chain shipping model problem

Extracting the shipping schedule for the supply chain problem, the number of motors to be shipped from the two supply source locations to the six demand destinations is presented in Table 8.17 (the bold numbers in the rows and columns in the table). For example, to achieve a cost-minimized shipping schedule, the manufacturer has to ship 286 motors from St. Louis to the Kansas City warehouse in the 28th month. Likewise, all the other eight scheduled shipments in Table 8.17 must be shipped exactly as scheduled to ensure the optimization of the total costs. Note in Table 8.17 that the allocation of motors exactly adds up to the last column supply values and the bottom row demand values.

Image

Table 8.17 Shipping Schedule for 28th Month Supply Chain ShippingProblem

Also, the value of 104,266 in Figure 8.2 is the total optimized cost for thisshipping schedule (taking the units shipped in each cell of Table 8.17 andmultiplying them by the number of units in those cells). The resulting shipping schedule for the supply chain problem in month 28 is detailed inTable 8.18.

Image

Table 8.18 Resulting Shipping Schedule for Month 28

Summary of BA Procedure for the Manufacturer

The intent of this BA application is to develop a BA procedure for the manufacturer to utilize every month in planning the supply chain problemof setting up an optimal shipping schedule in the supply chain network.

This BA procedure based on the BA team analysis presented here involves both data collection efforts: statistical analysis and the application of optimization software. Specifically:

1. Collect shipping cost information from the firm's cost accountingdepartment.

2. Collect and update monthly actual demand values from warehousecustomers.

3. Collect supply center supply capacity to ensure sufficient supply capacityexists to handlemonthly demand.

4. Rerun curve-fitting software on new and old actual demand data to determine bestforecasting model based on R-Square and other statistics asneeded.

5. Forecast warehouse customer demand and affirm through analysis that the resulting estimates are truly based on the best forecasting model. Revise as needed to select the best model.

6. Incorporate the cost, supply, and forecast demand information into a linear programming model similar to what was developed in Section 8.5.1.

7. Run the IP or LP model and extract the shipping schedule from themodel output.

Demonstrating Business Performance Improvement

A BA is not complete without showing that business performance can or will be improved. This case study has a basis for comparison. In comparingthe MAD statistics from the present forecasting procedure and the BA proposed procedure, a potential for improvement in shipping can be observed. In Table 8.19, the MAD values based on the three months (months 25, 26, and 27) used for model validation are presented. The MADstatistics (see Appendix E, Section E.8) represent the average monthly overage or underage of motors that could have been avoided if the BA proposed procedure would have been in place. Such needless shipments waste effort and add to cost inefficiencies for the manufacturer.

Establishing a procedure that lowers the MAD statistics would represent an opportunity for improving business performance.

Image

Table 8.19 Comparisons of MAD Statistics Between Present and BA Proposed Forecasting Procedures

As can be seen in Table 8.19, the current procedure of using a smoothed averaging to generate a forecast results in fairly large MAD statistics compared to the proposed BA procedure. The total of the MADs in Table 8.19 clearly shows a significant reduction in monthly overages or underages when using the proposed BA procedure in forecasting motor customer demand. Reducing the inaccuracy in forecasting also translates into minimizing wasted costs of shipping the motors that are either not needed in the warehouses during low customer demand periods or rush- ordered when shortages occur. These results reveal that the implementationof the proposed BA procedure for the supply chain shipping schedule problem could have improved business performance for the manufacturer over what was previously used to forecast the last three months.

As a final recommendation from the BA team on the prescriptive analyticsstep, the analyst or BA team responsible for utilizing the new BA procedure should continuously run updates to check and confirm the benefits of using the BA procedure on a monthly basis. Continuously showing the worth of BA is recommended for the success of BA in firms (see Table 2.2 in Chapter 2, "Why Are Business Analytics Important?").

Summary

This chapter presented a case study illustrating the use of BA to solve a supply chain shipping problem. The case study utilized the three-step BA process to develop a BA procedure that could be repeated monthly to improve a manufacturer's business performance.

The particular use of methodologies in this case study could have been different andcould have highlighted the fact that BA is meant as a step-wise guide in the application of statistical, information system, and management science methodologies. Like a walk in a forest, there may be many paths, but the goal is to reach the other side using knowledge and information (from a BA analysis) to support your steps.

This chapter ends the text material of this book, but the appendixes offer readers a rich foundation of methodologies useful in BA. Some have been demonstrated in the text material, and others have not, but all can be useful for differing analyses. The more methodologies that BA analysts know, themore likely they are to utilize the right one in the right situation. The appendixes are a starting point on which to build a foundation of methodological tools to strengthen and continually augment BA knowledge.

Discussion Questions

1. Some of the graphs (for example, Chicago going up) in the descriptive analytics analysis tended to show fairly linear behavior, yet a cubic model, rather than a linear model, was used in most cases. Did it make sense to usethe cubic model instead of the linear model? Why?

2. The SPSS curve-fitting process involves the development of 11 different regression models. Should all the regression models have been tested with the validation data in months 25, 26, and 27?

3. What other methodologies in this book might have been applied to analyze thiscase study problem?

4. Why is it necessary to show and continuously demonstrate business improvement inthe analysis when it is clear that forecasting results areimproved?

Problems

1. How was the Kansas City cubic regression forecast of 2,983 for month 25 computed?Show the formula with input values.

2. How was the Chicago cubic regression forecast of 4,561 for month 27 computed?Show the formula with input values.

3. How was the Oklahoma City quadratic regression forecast of 2,754 for month 26computed? Show the formula with input values.

4. How was the MAD statistic for the Oklahoma City linear regression modelcomputed on the three-month validation data? Show the formulawith input values.