CHAPTER 3

Secondary Data and Big Data Analytics

LEARNING OBJECTIVES

1. Understand the advantages and disadvantages of secondary data.
2. Comprehend data mining and behavioral targeting.
3. Learn the advantages of big data analytics, how to make it actionable, and the importance of data visualization.

KEY TERMS

Secondary data	Primary data
Internal database	Data visualization
Neural Network	Data mining
Geographic Information Systems (GIS)	Behavioral Targeting
Decision support system (DSS)	

CHAPTER OUTLINE

1. Nature of Secondary Data

I. Definitions

- A. Secondary Data
- B. Primary Data

II. Advantages of Secondary Data

III. Limitations of Secondary Data

- A. Lack of Availability
- B. Lack of Relevance
- C. Inaccuracy
  - 1. Guidelines for assessing the accuracy of Secondary Data
- D. Insufficiency

2. Internal Databases

I. Creating an Internal Database

- A. Internal Database

II. Data Mining

A. Neural Network

B. Data Mining

III. Behavioral Targeting

A. Definition

3. Big Data Analytics

I. Definition of Big Data and what it can offer a firm

II. Defining Relationships

III. The Big Data Breakthrough

IV. Making Big Data Actionable

V. Data Visualization

VI. Battle over Privacy

A. Criticism of Databases

B. Identity Theft

C. Governmental Actions

D. State Laws

4. Geographic Information System

A. GIS

B. Spatial Analysis

5. Decision Support Systems (DSS)

A. Interactive

B. Flexible

C. Discovery-Oriented

D. Easy to Learn and Use

6. Summary

CHAPTER SUMMARY

1. NATURE OF SECONDARY DATA

I. **Definitions**

A. **Secondary Data**–information that has already been gathered and might be relevant to the problem at hand.

1. **Internal Sources**–includes annual reports, reports to stockholders, sales data, customer profiles, purchase patterns, product testing results, internet and mobile tracking of customers, company website tracking, and house periodicals.

2. **External Sources**–federal, state, and local departments and agencies, trade and industry associations, business periodicals, and other news media that publish articles on the internet, and individual companies like Acxiom. These sources can be found on the internet.

B **Primary Data**– new surveys, observations, and experiment data collected to solve the particular problem under investigation.

II. **Advantages of Secondary Data**

A. May be obtained at a fraction of the cost, time, and inconvenience of primary data

B. May help clarify or redefine the problem during the exploratory research process

C. May provide a solution to the problem

D. May provide primary data research method alternatives

E. May alert the marketing researcher to potential problems and/or difficulties

F. May provide necessary background information and build credibility for the research report

G. May provide sample frame from which a sample is drawn

III. **Limitations of Secondary Data**

A. **Lack of availability**

1. For some research questions, there are simply no available data

B. **Lack of relevance**

1. Data is in a form that cannot be used by the researcher.

C. **Inaccuracy (Guidelines for assessing the accuracy of secondary information)**

1. Who gathered the data?–source of the secondary data is a key to their accuracy

2. What was the purpose of the study?–understanding the motivation for the research provides clues to the quality

3. What information was collected?–identify exactly what information was gathered and from whom

4. When was information collected?–if timing is important this is a need to know

5. How was the information collected?–by mail, telephone, mobile device, Internet, or personal interviews

6. Is the information consistent with other information?–lack of consistency between secondary data sets should dictate caution

D. **Insufficiency**–the available data may be relevant and accurate but still not sufficient to make a decision or bring closure to a problem

2. INTERNAL DATABASES

I. Creating an Internal Database

A. **Internal Database**–a collection of related information developed from data within the organization

B. **Starting Point**

1. Built on salesperson's "call reports"

a. Call reports provide a blueprint of a salesperson's daily activities. It details the number of calls made, characteristics of each firm visited, sales activity resulting from the call, and any information picked up from the client regarding competitors

2. Built on sales results and customer preference

3. customer preferences

4. internet, mobile and social (media) data

II. Data Mining

A. **Neural Network**–computer program that mimics the processes of the human brain and is capable of learning from examples found in patterns of data

B. **Data Mining**–use of statistical and other advanced software to discover non-obvious patterns hidden in a database. The widest applications would be the following:

III. **Behavioral Targeting**

A. Definition-Behavioral targeting is the use of online and offline data to understand a consumer's habits, demographics, and social networks in order to increase the effectiveness of online advertising.

3. BIG DATA ANALYTICS

A. Definition of Big Data: the accumulation and analysis of massive quantities of information. Big data offers a firm:

1. Deeper Insights -- Deeper Insights – Rather than looking at market segments, classifications, groups or other summary level information, big data researchers have insights into all the individual, all the products, all the parts, all the events, and all the transactions.
2. Broader Insights – Big data analytics takes into account all the data, structured and unstructured, to understand the complex, evolving, and interrelated conditions to produce more accurate insights.
3. An example of deeper and broader insights would be where a cable TV supplier showed that 95 percent of all appointments made were met on time. This sounds impressive until you know that there were 3,000 appointments in a day, so 150 customers waited at home in vain each day. If you can then tie the missed appointments to call data, survey data, and repurchase data along with tweets and Facebook comments, a manager could begin calculating how much revenue and word-of-mouth damage occurred, in addition to the extra cost of rescheduling and expediting visits.

B. Defining Relationships

1. Big Data represents a paradigm shift away from traditional hypothesis driven research, based upon well-defined parameters created by the researcher. It allows us to collect data and then see what it tells us.
2. Big Data is more about "what" than "why."

C. The Big Data Breakthrough

1. Traditional databases were downloaded and stored in rows and columns.
2. Traditional databases were limited when it came to storing strings of works, such as what is found in an email or text message. They are also unable to handle pictures of video.

3. New types of databases have emerged that allow analysts to create queries against all types of data, traditional data, text data, video, pictures, etc.

Practicing Marketing Research: Big Data Unmasks a Wealth of Knowledge

Questions:

1. **How can real-time tracking help marketing decision makers?** Answers will vary by student, but some mention of being able to track "how" and "when" consumers purchase will give marketers valuable insights.
2. **Do tracking and traditional marketing research benefit each other? If so, how?** Traditional database marketing research will always have its place in marketing, but it is limited by the data in the rows and columns. Tracking can complement traditional marketing research providing insights that were before unattainable with traditional statistical analysis.

D. Making Big Data Actionable

1. What is needed to evaluate complex output are intuitive tools that can aid everyday decision making.
2. Automated decision making via big data analytics can provide a shopper with recommendations on a real-time basis, tailored to his or her needs.

E. Data Visualization

1. Data visualization-is the use of picture visualization techniques to illustrate the relationship within data.

2. Data visualization companies, such as Gooddata, Ayasdi, Tidemark, and Platfora, turn large data sets into pictures that lead companies intuitively to the information that is more important to them.

V. **The Battle over Privacy**

A. **Criticism of Databases**–the downside to big data is the increasing concern about privacy issues–how will the information be used. Your actions catch up with you because of the records online and what you do online–Web searches, e-mail messages.

1. Other factors–toll roads, credit card transactions, vendor databases, cell-phone records, time clocks at the workplace

2. Privacy advocates are concerned because technology allows the combination of information.

3. Congress is considering laws to limit tracking.

4. An emerging technique that is upsetting many privacy advocates is called "scraping." Firms offer to harvest online conversations and collect personal details from social networking sites, resumé sites, and online forums, where people might discuss their lives.

5. Behavioral tracking has become the foundation of the online advertising industry. Online advertising is why a company like Google can spend millions and millions of dollars on free services like its search engine, Gmail, mapping tools, Google Groups, and more.

6. The Internet is in an arms race over control of personal data. Facebook's 60 billion plus value is testimony to value of tracking over a billion people. The "like" and "share" buttons enable Facebook to track people online.

Exhibit 3.2 How to Cover Your Tracks Online

B. **Identity Theft**–costs $55 billion annually (Exhibit 3.2)

1. Unreliable safeguards–allow people without the authority to access your information

C. **Governmental Actions**–laws to protect consumers from identify theft

1. **Federal Laws**

a. **Gramm-Leach-Bliley Act** (Financial Services Modernization Act)–aimed at financial companies–require that customers be told how they will use personal information

b. **Health Insurance Portability and Accountability Act**–healthcare industry–limits disclosure of individuals' medical information

c. **The Fair Credit Reporting Act (FCRA)**–enforced by the Federal Trade Commission, promotes accuracy in consumer reports and is meant to ensure the privacy of the information in them

d. **The Children's Online Privacy Protection Act (COPPA)**–gives parents control over what information is collected from their children online and how such information may be used (See page 51)

2. **State Laws**

a. **California's Notice of Security Breach Law**– if non-encrypted information has been taken by an unauthorized person, the company or agency must tell the California resident

4. GEOGRAPHIC INFORMATION SYSTEMS

A. Geographic Information System (GIS)

1**.** provides both a means of maintaining geographic databases and a tool capable of complex spatial analysis to provide information for a big data database.

2. includes a demographic database, digitized maps, and software that enables the user to add corporate data to the mix.

3**.** computer-based system that uses secondary and/or primary data to generate maps that visually display various types of data geographically. Allows researchers, managers and clients an intuitive way to organize data and to see relationships and patterns

B. Spatial Analysis - the analysis of phenomena distributed in space and having physical dimensions (the location of, proximity to, or orientation of objects, such as stores, with respect to one another).

5. DECISION SUPPORT SYSTEMS (DSS)

A. **Interactive**–give simple instructions and see the results generated on the spot

B. **Flexible**–sorts, regroups, totals, averages, and manipulates data in a variety of ways

C. **Discovery-oriented**–probes for trends, isolates problems and asks new questions

D. **Easy to learn and use**

6. SUMMARY

QUESTIONS FOR REVIEW & CRITICAL THINKING

1. Why should companies consider creating a big data database? Name some types of information that might be found in this database and sources of this information.

Big data databases are an excellent way to identify and classify customers and to monitor their purchases. This gives the firm the ability to target their market offering more effectively. The types of information that might be in the database include the customer demographics, including names, addresses, telephone numbers, age, income, family members' names and ages, and preferences, as well as purchase history. Potential customers' names might also be listed with as much demographic information as is available. This information may come from several sources, including sales records, sales call reports, credit applications, registrations for "clubs," customer panel data, and registrations for gifts and prizes.

2. Why has big data analytics become so popular with firms like United Airlines, American Express, and Ford Motor Company?

All major firms collect data related to customer shopping, transactions, and customer service feedback. Big Data analytics uses statistical and other mathematical software tools to discover non-obvious patterns of preference and behavior that might be hidden in these databases. The objective of this analysis is to identify information that marketers can use to formulate strategies and tactics to increase the firm's profitability.

3. It has been said that big data analytics turns the scientific method on its head. What does this mean?

Traditionally, marketing research has started with problem recognition, well defined problems and methods of collecting data in which the market researcher has a lot of control over. The data is put neatly into rows and columns in a statistical database, and the market researcher can analyze those relationships defined by the research study objectives. With big data analytics, however, the process is more discovery oriented. The software can analyze data patterns in the database and utilize data in a form in which was not possible with traditional statistical analysis. Hence, the traditional steps of the scientific method are not followed as they once were.

4. Why are secondary data often preferred to primary data?

Secondary data are already collected, so access should be much faster. Secondary data should be less costly to obtain than primary data. In some cases, as when investigating historical events, secondary data is all that is possible to obtain.

5. What pitfalls might a researcher encounter in using secondary data?

There may be a lack of availability of secondary data, or the data that is available might lack currency, relevance, or accuracy. Finally, the available secondary data might be insufficient to help the manager to make an accurate decision.

6. Why has behavioral targeting become so popular with marketers? Why is it controversial?

As the Internet has matured, non-targeted advertising has decreased in efficiency. One study found that only 4 percent of Internet users account for 67 percent of all display ad clicks. Another recent study by DoubleClick reported an average click-through rate of just 0.1 percent. Behavioral target has raised some issue over consumer privacy, making it controversial in some cases.

7. In the absence of company problems, is there any need to conduct marketing research or develop a decision support system?

The definition of marketing research states that information is collected not only to help find solutions to existing problems, but also to identify opportunities. Therefore, there is no enterprise that cannot benefit from marketing research. A Decision Support System (DSS) can provide the information that is needed to identify those opportunities. And, you never can tell: just because you don't have a problem today does not mean that you won't have one tomorrow. Then, a DSS could be invaluable.

8. What is data visualization? Why is it important?

Data visualization is the use of picture visualization techniques to illustrate the relationships within data. It is important because it enhances to readability and understandability of the results of analysis done using big data analysis software.

9. (Team Exercise)

Divide the class into groups of four or five. Each team should go to the Internet and look up big data analytics. Each team should then report to the class on how a specific company is effectively using big data to improve their marketing efficiency.

REAL-LIFE RESEARCH 3.1 The Interesting and Curious World of Nate Silver

Key Points

- The question on the table is whether big data – that is, the accumulation and analysis of massive quantities of information – will change our world, or whether it's just another overhyped technology with a too-good-to-be-true story line.

Questions

1. **Will big data analytics put marketing researchers out of business? Why?**
   The ability to make output usable for more managers, those without extensive training in statistical analysis will not terminate traditional marketing research. There will always be a need for people who understand the basic concepts of marketing research, and be able to translate broad management problems into more precise marketing research problems. Big Data is just a tool. There still has to be somebody with knowledge of what the tool is doing to facilitate the interpretation of the output.
2. **Go to Nate Silver's blog, www.fivethirtyeight.com and tell the class what topics his is currently discussing.**
   Student answers will vary depending upon when they read to blog.
3. **The case discusses several current non-business uses of big data. Can you think of other non-business problems that big data might solve?**
   Student answers will vary.
4. **What factors have led to the big data era?**
   There need to discover more precisely how
   consumers make decisions, when they make decisions, and how and why they buy. The questions haven't really changed that much, but the need for more efficient and accurate methods of data analysis have spurned the development of big data. From the users side, powerful analytical tools are now available and their output is understandable for persons

that don't have extensive statistical training. Big data algorithms supply information that is actionable by the person accessing the information.

