Secondary Data: Collection and Analysis - Classroom Activities for Learning

Camille P. Schuster, Beverlee Anderson and Glen Brodowsky
California State University, San Marcos

Marketing managers need to analyze markets, competitors, and consumers to create, monitor, and adapt marketing strategy. The amount of data available for these analyses is increasing exponentially. Much of the data to be analyzed is secondary data. This article presents two classroom activities designed to facilitate the collection of appropriate secondary data in one activity and analyze secondary data in the second activity. Assessment demonstrates that these activities resulted in increased learning and understanding by students. Given the requirements of marketing managers in today's marketplace, this topic needs more attention.

Key words: Secondary Data, Data Collection, Data Analysis, Assessment, Teaching Activity
Disciplines of Interest: All Business Disciplines

INTRODUCTION

The amount of digital data available to researchers through on-line data bases and the web is increasing "exponentially" (The Economist, 2010a). Researchers can become lost in a sea of data. The main problem no longer is finding information on a topic but, as The Economist (2010b) points out, the problem is ". . . laying one's hands on the relevant bits easily and quickly." As Murphy and Schlaearth (Murphy and Schlaeerth, 2010) state, "for just about any research problem a preexisting data set is available." The reasons for this explosion of available information are, of course, technology - the capability of digital devices, more people having access to more powerful tools, and a larger number of people interacting with information (The Economist, 2010c), meaning more information is shared. The Economist (The Economist, 2010d) says the amount of digital information increases tenfold every five years and that the processing power and storage capacity of computer chips double or their prices half roughly every 18 months. The shift from "information scarcity" to "information overload" has been rapid.

Big data is growing exponentially and transforming the data analysis process (Nie, 2011). For example, there were 1.8 zettabytes of data created in 2011 with the possibility of an increase of 50 times by 2020 ("Transforming the Retail," 2010)).

This plethora of data and information has created new challenges for the novice researcher. By gathering and combining data on consumer purchases, search activities on websites, comments to friends using social media, cameras that track what consumers view onscreen and in stores with survey data, companies can obtain a 360^0 view of consumers. By triangulating the analysis of data collected in different situations and for a variety of purposes, market researchers can identify what products consumers look for, what criteria they focus on for decision-making, what reviews are read, from which people advice is received, what products are purchased, and what consumers say about products.

In spite of the rapidly changing information technology landscape, little has changed in the way we prepare marketing students to function in such an information-rich environment. Few marketing research courses spend more than one or two class periods on secondary data. Marketing research generally focuses on designing the process of gathering data with the last few weeks devoted to analyzing the data - generally with descriptive statistics and regression. The statistics course focuses on sampling, interval confidence levels, regression, and other statistical tests for which there is time at the end of the semester. With big data and a 360^0 view of consumers, sampling and confidence intervals are not relevant. The data is what it is - a 360^0 view of the consumer (Nie, 2010). Much of the data are secondary data. The breadth of skills needed for data analysts and data-driven decision makers in companies is expanding. As educators, we assign research tasks to our students and are often frustrated by their inability to use reason and analytical skills to properly formulate questions and logically answer them.

Students need more than a list of basic terms, types, advantages and disadvantages of secondary data as covered in traditional marketing research courses. Recognizing, as did Castleberry (Castleberry, 2001) that our marketing graduates will be expected to be knowledge workers in the new economy, they will need skills for determining how to collect relevant secondary data and how to analyze that data.

This paper discusses two classroom activities, one designed to show students how to collect secondary data and the other designed to show them how to analyze it. The activities were designed based upon the assumption that upper-division marketing students who had completed pre-requisite statistics courses would be equipped to approach secondary marketing research assignments. The first activity was a class project focused on getting students to locate and select online secondary data to answer a few market research questions. The assumption, based upon how typical marketing research texts treat the subject of secondary data collection, was that students would be able to identify the question being asked and the type of information needed, select multiple sources of data, evaluate the quality of those data, and properly cite their sources. The second activity focused on the analysis of secondary data. The assumption was that students has the quantitative skills to analyze a secondary data set and draw conclusions that lead to actionable decisions. Unfortunately, as it turned out, the instructors' assumptions were erroneous. Simply

reading about the costs and benefits of secondary data or having completed a statistics course did not prepare students to successfully complete the assignments.

The paper begins with a literature review of the material addressing the attention teaching secondary research has received in the pedagogical and academic literature. It continues with detailed discussions of each of the two projects. It concludes with a discussion of how meaningful, in-class, interactive exercises can help prepare students to approach the process of navigating, analyzing, and interpreting the ever growing sea of data upon which they must rely for making sound business decisions.

LITERATURE REVIEW

A perusal of five standard marketing research texts (Kumar, Aaker and Day, 1999; Malhotra, Naresh and Peterson, 2006; McDaniel and Gates, 2002, Malhotra, 2010) showed that most textbooks, with the exception of Hair, et. al (McDaniel and Gates, 2002) which has two chapters, one of which focuses on database research, treat the secondary data topic in much the same matter. The first topic is the discussion of terminology definitions of primary versus secondary data, followed by a description of the advantages and disadvantages of using secondary data. While the lists may vary slightly, most consider that secondary data will help the researcher create a better research project, the costs of secondary data are generally lower than for primary data, it is usually faster, and in some cases it would be impossible for the researcher to gather some of the data that is only available through secondary sources. The disadvantages may include the currency, units of measure, relevance, and accuracy of the data among others.

Another topic usually covered in the chapters is a section on criteria for evaluating secondary data. This discussion usually includes the topics of the relevancy of the units of measure, the method used to collect the data, who gathered the data, why it was gathered, and whether it is consistent with other information.

There have been relatively few scholarly articles on the importance of secondary data in the marketing research course or on an approach to teaching how to collect or analyze secondary data. Castleberry's (2001) article, focused on melding web and off-web sources in a marketing research course project with the goal of helping students develop their own research skills with secondary data and learning the "boundaries of cyberspace." In the years since that article was written, the boundaries of cyberspace have almost vanished. While not all information is on the free web, much information that was previously available only in off-web sources, is now digitized and can be found through various data bases.

An article by Campbell and Cook (2010) focused on the benefits of collaboration between library faculty and business faculty in the development of a co-instructional plan for research. The objective of the first project was to help students spend less time searching for data so they can spend more time analyzing the data. The students also developed an appreciation for the expertise of information professionals. The plan included integrating library instructional

sessions into a marketing research class and the librarian worked with teams on their class projects. The objective of the second project was to help students stronger skills for analyzing secondary data.

PROJECT ONE: COLLECTIONS OF SECONDARY DATA

Purpose

The purpose of the marketing research assignment included multiple objectives. The first was to acquaint the students with the wealth of information that is available in digital form. The other more important objectives were designed to help students become more aware of the knowledge, skills and patience needed to become proficient in researching information on the web. The practical, hands-on experience of having to know what to search for (search terms), where to search, how to search, how to evaluate, how to validate while experiencing high levels of frustration in the process were designed to promote a significant learning experience for students.

For this project, students in an upper-level marketing research class were asked to find the answers to eight seemingly easy questions. However, as one student wrote in her follow-up assignment, ". . . I found that the all-encompassing theme of this assignment was, 'all things are not as they seem'." The simple questions were not so simple and finding the answers proved challenging to every student in the class. One student labeled this assignment, "the most difficult" of her college career.

Assignment

This assignment was developed through a collaborative relationship between the Business Librarian and the course instructor. During the term of this assignment, the Business Librarian offered a workshop in a Marketing Research class that would focus on three database (Mediamark Internet Reporter MRI+, References USA, Euromonitor's GMID) searches and the APA style [NEED REFERENCES FOR DATABASE AND APA]. The instructor also conducted a workshop on using government databases, which included the Census, Bureau of Labor Statistics, international governments' websites and the databases available through local governmental organizations. In total, four two-hour class sessions were devoted to literature reviews and secondary data. The instructor and librarian agreed that no assistance would be given to the students while completing their assignment.

The Business Librarian and the instructor collaborated in creating the 8 question assignment (Figure 1). The assignment was posted on the course website one week prior to the class sessions on secondary data. Students were told that the assignment would take quite a bit of time and that they should not wait until the night before it was due to begin the assignment. Each of the questions was designed to address one or more learning outcomes.

Figure 1. Assignment

THIS ASSIGNMENT IS TO BE DONE INDIVIDUALLY – NOT IN TEAMS
Homework Questions Based on Secondary Sources

Please answer the following questions and give the source details you used to answer the questions.

1. What are the SIC and NAICS codes for restaurants and for bars?
2. What percent of households in San Diego County have an income over $200,000 per year?
3. What are the demographics (age and income) of the people most likely to eat out at restaurants at least once a week?
4. What is the average number of times students eat out in a restaurant? Give the global figure**.**
5. On average do people in San Diego or San Francisco have the longer commute to work?
6. You are thinking of opening a chain of small kiosks in every major United States airport and want to target the largest airport first. What are the five busiest airports in the U.S?
7. What is the population of Dublin, Ireland?
8. In which region (Northeast, Midwest, South or West) do families spend the most on reading materials like books, magazines, and newspapers?

INSTRUCTIONS: **Write** an answer to each question.

- Don't just download a table or cut & paste.
- Present findings in a readable and logical form
- Be sure to include all salient information

Cite your sources in appropriate APA style.

Try to find a second source for the information
If different data are found – try to explain why different (or)
Seek a 3rd source and see which that supports.

Give some indication as to how confident you are in your answers.

Be prepared to present and discuss your search process, any difficulties or questions you had and how your resolved your questions.

Table 1. NAICS Codes for Restaurants 2007 and 2012

2007 Search Results for: Restaurants	2012 Search results for: Restaurants
Number of records found: 12	Number of records found: 12
445210 (delicatessens (except grocery store, restaurants)	445210 Delicatessens (except grocery store, restaurants)
722110 Family restaurants, full service	722511 **Restaurants,** full service
722110 Fine dining restaurants, full service	722511 Fine dining restaurants, full service
722110 Full service restaurants	722511 Full service restaurants
72110 **Restaurants**, full service	722511 Family restaurant, full service
722211 Carryout restaurants	722513 **Restaurants**, fast food
722211 Delicatessen restaurants	722513 **Restaurants,** carryout
722211 Drive-in restaurants	722513 Fast-food restaurants
722211 Family restaurants, limited-service	722513 Drive-in restaurants
722211 Fast-food restaurants	722513 Family restaurants, limited-service
722211 **Restaurants**, carryout	722513 Carryout restaurants
	722513 Delicatessen restaurants

Two examples of questions with their associated desired learning outcomes and explanations are as follows:

Question: *What are the SIC and NAICS codes for restaurants and for bars?*

Learning outcome goals:

a. One must clearly understand exactly what information is being sought.
b. One must check the dates associated with the information
c. Always seek a second source for the information

Explanation:

a. There are many different codes for different types of restaurants and bars. Broad classifications are sit-down restaurants and fast-food restaurants, but there are many finer classifications (Table 1).

b. The NAICS classification numbers were changed in 2010. If one used a source such as 2007, the number was different from that from a 2010 or 2011 source. (There is discussion of changing the codes again in the near future.)

c. By seeking a second source, it is likely that a different code will be found, which prompts the researcher to question the first finding.

Question: You are thinking of opening a chain of small kiosks in every major United States airport and want to target the largest airport first. What are the five busiest airports in the U.S?

Learning Outcome Goals:

a. Understanding what kind of information would be most appropriate for a particular marketing problem.
b. The currency of the information is very important.
c. Check the credibility of the source

Explanation:

a. The term "largest airport" could be interpreted many ways. The answer will vary depending on which interpretation is used. Given that the management decision problem relates to opening a chain of small kiosks in airports, the best information would be based on passenger traffic. However, there are other ways to interpret such as acreage of airport, number of landings and take-offs, etc.
b. Passenger traffic has shifted through the years as airlines have merged and changed their hub locations.
c. Different sources may have slightly different ordering of topics.

The assignment instructions also asked that the information not be reported as a table but in verbal form to see if students included all salient information in their answer, e.g., did they include the year of the information in the answer. In question #7, for example, did they explain that Dublin, Ireland, is both a city and a county, and did they give the information for each location, or did they explain that the number they reported was for the city for the county. In addition, did they include the margin of error associated with the reported data.

Asking students to cite sources of information in APA style was also important. The goal was to be able to give a source citation that would make it possible for others to locate the particular source that they had used.

Answers to Questions

As was expected, no two students had exactly the same answers for all eight questions. On some questions, for those who were able to find any answer, the answers were the same, such as Question #4 “What is the average number of time students eat out in a restaurant? Give the global figure.” The answer to this question could be found only in one source, Passport GMID (2011) [NEED REFERENCE], a database the University library had on a trial basis and students had been introduced to in the class session conducted by the Business Librarian the preceding week.

Figure 2. Consumer Expenditure

Table 8. Regional of residence: Average annual expenditures and characteristics, Consumer Expenditure Survey, 2010-2011 - Continued					
Item	All	Northeast	Midwest	South	West
Reading	108	127	112	84	126
Source: Consumer Expenditure Survey, U.s. Bureau of Labor Statistics, September, 2012					

On other questions, such as Question #8, "In which region (Northeast, Midwest, South or West) do families spend the most on reading materials like books, magazines, and newspapers?" the results varied widely. Some students reported the Northeast, some the South and others reported the West. The Midwest was the only region not reported by any student as an answer to this question.

On question #2, asking the percent of households in San Diego County that have an income over $200,000 per year, the answers varied from a low of 2.48% (reported using zipatlas.com) to a high of 43%, (reported using Factfinder 2 of the U.S. Census) with many different figures in between. Needless to say, this was quite a range. The sources used varied including Factfinder2 of the U.S. Census, Reference USA, Media Mark Reporter, and American Community Survey among others. In some cases the tables were misread or misinterpreted, there were also mistakes reporting income per family versus income per household (this difference is over a percentage point). In other cases, the source used sample information or the information was not the most current, and finally for some who tried to calculate the percentage, the math was incorrect.

For Question #8, which asked "In which region (Northeast, Midwest, South or West) do families spend the most on reading materials like books, magazines and newspapers?" the results were quite varied. Students had been shown the Consumer Expenditure Survey data from the U.S. Bureau of Labor in a session on government databases. By using this source and the most current data available, the information in Figure 2 could have been found, which shows that it is the Northeast that spends the most on reading materials.

However, students used a variety of sources. For example, one student used a BLS working paper, which reported price indices rather than consumer expenditures, which gave very different results. When using the MRI+ database, students could get information only on the amount spent on books, which resulted in the South spending the most. Another student used BLS data, but looked at a table that reported increases in expenditures in each category rather than the actual dollar amount spent. The result of this search was that the South increased the most. And,

finally, some students used the Reference USA database that yielded the answer that people in the West region spent the most on reading materials.

Learning

The day the students were to submit their assignments the class session was devoted to going over the students' results for each question in an open class discussion. Prior to their reporting, students were informed that they would all receive points for submitting the assignment and that the grade would not be based on the accuracy of the answers they submitted. Students reported their answers to the questions and the sources they had used. Students were then given a follow-up related assignment, which asked for them to write a 1-2 page report about what they had learned from this exercise. The purpose of this assignment was to see if the students had achieved the learning outcomes that were the objectives of the assigned task.

Overall, the most common learning cited by students was that they learned about many new and different sources of information, other than Google. Many had never used commercial databases prior to this experience and they learned about the wealth of information that is not available on the free web. They also gained an appreciation of all the information that is available through government sources, much of which can be accessed through the Census website.

In addition, one common take-away stated by almost all students was that one should use more than one source for any information. All information needs to be checked and verified. The following are representative sample excerpts from students' reports on what they learned:

Student #1 – "I learned . . . that by using the industries code for specific purposes, it can be easier to find facts

Student #2 – ". . . be careful on the accuracy and publication date of data we find in the databases. Always use three or more sources to make sure the result is accurate.

Student #3 – "Overall this exercise was a great way for me to realize that searching for relevant data may not be as simple as it sounds. There are a lot of different criteria that go into looking for current and correct information."

Student #4 – "The most important lesson I learned from the assignment is that when we do marketing research, we should use multiple sources of secondary data."

Student #5 – "Secondary sources and understanding how to use them can drastically improve my ability to research."

Student #6 – "This process isn't easy, it's time consuming and frustrating at times. But what I have learned most from this assignment is patience. It takes extreme amounts of patience, time and thoroughness to answer the problems at hand."

The objectives of this class project were met. In addition, the class sessions provided the direction and assistance necessary for expanding the students' horizons about where to find data and what information about the data was important to gather to establish the creditability of the data.

PROJECT TWO: ANALYSIS OF DATA

Purpose

This project took place in a course entitled Consumer Demand and Channel Collaboration, which has two major parts: (1) analyzing customer data and (2) collaborating between groups. The course had been taught for six years. When first designed Nielsen, P&G, and Target agreed to have Nielsen create three separate data sets for a particular product for the class project. One group of students represented P&G and received a set of point-of-sale (POS) data on the product broken out the way P&G normally requests it; the group representing Target received a set of POS data on the product broken out geographically the way Target requests it; the group representing Nielsen received a set of Nielsen consumer panel data related to the product and retailer. Each group needed to analyze its own data set, any other data they found through their own research, and Mediamark data related to this product to determine their most important current target market and the best target market for increasing sales of that product at Target. Then there was one collaboration meeting to identify joint target markets and the approach to take to increase sales of the product at Target. Then each group worked on a solution. The second collaboration meeting involved having each group present ideas and choosing a joint solution with appropriate metrics.

Analysis of the secondary data has continued to be a challenge in this course throughout the six years. Initially students were asked to review the Mediamark data and determine a best current target market and a best potential target market as a preliminary activity. An initial assumption was that after having taken research and statistics courses students were prepared to analyze the Mediamark secondary data. However, students had great difficulty with this assignment. Typical questions included "What should I look at - the rows or columns?" or "What do the index numbers mean?" or "Do I have to consider any of the other columns besides the Index number?" On the homework assignment, best target markets might be identified as 18-34 year olds because 52% of them used the product compared to 48% of those over 35. With a difference of 4% students were willing to completely ignore those over 35 because 52% is larger than 48%. Clearly they were confused about how to analyze this data. Over time explanation of how to read the data was increased and the analysis improved but only marginally.

The data provided by Nielsen is a huge spreadsheet with POS data representing sales of brands, sizes, and types of the product sold each week for 52 weeks. The Nielsen representative who prepared the dataset provided guidelines for how to

begin analyzing the data and shared the guideposts that Nielsen analysts use in their analysis. These instructions were presented to the students and they were asked to identify best selling products and/or target markets (depending upon they data set they received). At the beginning of each dataset there was a set of summary data. In general, the students reviewed the summary data, were overwhelmed by the rest of the data, and usually ignored it. For two semesters Tableau software was used as a tool to visually analyze the Nielsen data sets. Using Tableau was partially successful in that students did see some additional relationships.

The cost/benefit result of using this approach to data analysis was not positive. Taking time to present the tools necessary for the analysis of the Mediamark data, the Nielsen data, and the use of Tableau took a lot of time from the rest of the course. Analysis of data was designed to take about one-third of the class time during the semester because the emphasis was on using to inform decisions. Since the analysis activities took about half of the semester, the students no longer had time to do qualitative research testing some of their ideas and the first collaboration kept getting pushed back which left little time for the solution part of the course. If the purpose of this course was data analysis, instruction and discussion of the three types of analysis could have been expanded resulting in more thorough analysis. However, this was not the purpose of the course. The assumption that as seniors the students had skills for basic analysis of secondary data that could be used across different types of secondary data proved to be untrue. Basic analytical skills for how to interpret secondary data from a data set remained a problem all six times the course was taught. Therefore the issue of analyzing secondary data needed improvement.

Alternate Approach

With an increased emphasis on extracting meaning from secondary data sets, business and marketing analysts in companies need to understand how to analyze secondary data to provide direction for decision-making. This need in the marketplace and the difficulty students were having doing this analysis led to the conclusion that emphasizing the process of analyzing **some** secondary data was more important than providing a variety of secondary data resulting in little helpful analysis. Given the continued difficulty with using, understanding, and interpreting the Mediamark data, its use in a variety of Marketing courses, and its wide usage in the business world, analyzing Mediamark data was chosen as the focus of this project.

Learning Outcome Goals:

a. Find a place to begin the analysis in the data set.
b. Use data to answer additional questions that arise during analysis.
c. Identify variables that differentiate consumer groups.
d. Create tables to illustrate the differences.

Process

In class Mediamark data for one product was used to demonstrate the process for analysis. The emphasis was on how to approach the Mediamark data, how to create tables that help in the comparison of consumer groups across Mediamark tables, how to identify what is important in the data, and how to find a way to present relevant data in a persuasive manner.

For the data used for analysis in class, a category with which the students are likely to have only general familiarity was selected - men's skin care. Demographic information on Facial Moisturizers used by men in the last six months, Facial moisturizers by male medium users, Facial Moisturizers by male light users, Anti-Wrinkle Facial Moisturizers used by men in the last six months, Cleansing Facial Moisturizers used by men in the last six months, Moisturizing Facial Moisturizers used by men in the last six months, Night Cream Facial Moisturizers used by men in the last six months, Aftershave Lotion & Cologne for Men used in the last six months, Aftershave Lotion & Cologne for Men heavy users, Aftershave Lotion & Cologne for Men medium users, Aftershave Lotion & Cologne for Men light users were downloaded from Mediamark's online database ("Aftershave," 2010; "Facial Moisturizers," 2010). This is obviously an overwhelming amount of numbers for comparison making it impossible to determine anything by visual observation. However, this is the type of activity a marketing analyst would be expected to perform when a company begins to analyze a product category.

The class sessions begins by presenting the typical description of how to read a Mediamark data set - explaining what the columns mean, what the index numbers mean, who the sample is, and what guidelines are commonly used to determine which differences make a difference. This is where the traditional class session explanation ended. For this project the students were given the list of data sets that were downloaded and asked the following question:

> "Where do you start?," "What should you look at?," and "How can you make comparisons across these data sets?" One of the ways to begin is to identify variables that might help you distinguish between types of consumers: education, occupation, age, household income, marital status, race, and location. Making comparisons across a series of data sets even on one specific variable is confusing. This suggestion provides a framework for approaching the analysis but is not useful for making comparisons and drawing conclusions.

One possible tool for analysis is to create tables using index numbers to compare selected variables across the different data sets. This process makes it possible to keep track of which variable is being compared across which data sets. Explaining how to create these tables is important because the students will be looking at a different group of spreadsheets for a different set of products for their homework.

Table 2. Facial Moisturizers by Household Income

Income	6* Months	Anti-Wrinkle	Cleanse	Moisturizing	Night Cream
$150,000+	121	154	90	134	111
$75,000 - $149,000	106	109	64	122	109
$60,000 - $74,999	98	107	103	86	64
$50,000 - $59,999	119	121	74	112	129
$40,000 - $49,000	82	61	152	85	81
$30,000 - $39,999	91	97	105	69	149
$20,000 - $29,999	85	62	124	82	63
< $20,000	84	61	160	64	83

*Total Adult Men

They need to understand the process of identifying variables that could be relevant in distinguishing differentiation, how to create the tables, and why it is important to take the time to create these tables. The explanation related to table preparation also provides information for how to extract data from a spreadsheet and present it in a way to make the data more easily understood.

The next step was to have a discussion about the tables in class. Tables with using specific variables are shown and students are asked to identify differences and the significance of those differences. Before the discussion begins someone from each group is asked to write down all the relevant conclusions presented in class. Using the Facial Moisturizers and Usage by Income table (Table 2), the following questions were asked: which men at what income level use facial moisturizers the most? Do those preferences change across types of facial moisturizers? What conclusions can you make from this table? The same kinds of questions are asked relative to the Facial Moisturizers and Frequency by Education table (Table 3) and Facial Moisturizers and Usage by Location table (Table 4).

Aftershave and Cologne has been a man's category for a long time. Facial Moisturizers for men is a newer category so another way of understanding the usage of facial moisturizers by men is to compare the two so a set of similar tables were created for the Aftershave Lotion data sets (Aftershave Lotion & Cologne by Education - Table 5), Aftershave and Cologne by Household - Table 6 Income, and Aftershave and). Similar questions were asked and conclusions were recorded.

Table 3. Facial Moisturizers and Usage by Education

Education	6* Months	Anti-Wrinkle	Cleanse	Moisturizing	Night Cream
College +	117	140	70	128	117
Attended College	103	96	79	117	98
Graduated High School	86	84	141	79	90
Did Not Graduate High School	94	68	104	63	94
Post Graduate	119	154	58	124	103
No College	89	79	129	74	91

*Total Adult Men

Table 4. Facial Moisturizers and Usage by Location

Location	6* Months	Anti-Wrinkle	Cleanse	Moisturizing	Night Cream
Northeast	115	103	117	97	97
South	98	71	70	94	99
Midwest	86	113	113	100	96
West	105	159	111	117	118
Top 5 Media Markets	127	114	117	117	73
Next 5 Media Markets	109	143	112	115	125

*Total Adult Men

Table 5. Aftershave and Cologne by Household Income

Income	6* Months	Heavy** 8+	Medium 5-7
$150,000+	121	117	123
$75,000 - $149,000	106	116	103
$60,000 - $74,999	98	86	102
$50,000 - $59,999	119	159	115
$40,000 - $49,000	82	39	90
$30,000 - $39,000	91	55	95
$20,000 - $29,000	85	111	83
< $20,000	84	87	79

*Total Adult Men; **In Last 7 Days

Table 6. Aftershave and Cologne by Education

Education	6 Months*	Heavy** 8+	Medium 5-7	Light 0-4
College +	117	127	115	85
Attended College	103	98	104	94
Graduated High School	86	92	85	113
Did Not Graduate High School	94	72	97	112
Post Graduate	119	114	118	80
No College	89	86	89	113

* Total Adult Men; ** In Last 7 Days

Table 7. Aftershave and Cologne by Location

Location	6* Months	Heavy ** 8+	Medium 5-7	Light 0-4
Northeast	98	92	93	104
South	106	118	112	97
Midwest	103	91	104	105
West	90	85	83	96
Top 5 Media Markets	98	85	95	104
Next 5 Media Markets	102	97	113	94

* Total Adult Men; ** In Last 7 Days

The next step is to have the students reflect upon what we did (create comparative tables, review differences, draw conclusions, keep track of conclusions) with sample questions such as the following: What did you learn? What did you learn about the consumers? Can you identify variables that differentiate between consumers by product? What was important about how these tables were presented?

Then the students were told what data they had for their assignment, where they could locate it on the class website, asked to analyze the datasets in a way that is similar to what we did in class, told they can use the tables presented in class as templates, and asked to complete a homework assignment with the following questions: (1) identify the important current customers for the product category they have been assigned, (2) identify which customers might have the potential for purchasing more of the products in that category at Target, and (3) prepare the data they want to present during their first collaboration.

One other tool was used to assist the students in using this process. Using

Camtasia software the class session presenting the data analysis process (including all questions and answers) was recorded and posted on the class website. Students were able to go back and review any or all parts of the process at any time while working on their homework.

Questions

First, was it worth the time to prepare this material? In the past the assumption was that students were familiar with doing an analysis of Mediamark data from other marketing, research, or statistics courses. That assumption was not accurate. Finding a category, downloading all the data, and creating a set of 20 tables was very time consuming. Are the students better able to analyze the Mediamark data with this explanation?

Second, does the creation of sample tables enhance the presentation of data during the collaborative sessions? In the past students would cut and paste a section of Mediamark data or just pull up a Mediamark page to use during their presentations. The onslaught of numbers made it difficult for the members of other groups to focus on what was important in the mass of numbers, what it meant, and how it compared to anything else. By emphasizing the creation of tables for comparison purposes and with the discussion in class, did the students improve their presentation of data?

Third, were the students able to go beyond identifying sex, age, and race of adult users when identifying their target markets? Generally, the students will look at the Mediamark page of usage in the past six months, identify differences related to the highest index numbers, and stop their analysis. Without a way to compare heavy, medium and light usage across different products or characteristics, their analysis is superficial. Will students go beyond the traditional superficial analysis with this introduction?

Results

In the previous six times the course was taught, the analyses related to the examples used in class to describe how to read the spreadsheet. The groups typically settled on women 18-34 or parents with children under the age of 12 as their current best target market and Hispanic women with children as the growth market. With each group identifying similar consumers the process of collaboration to determine a joint consumer group focused on what age range should be used for the target market. In one instance students did some good external research and came up with a different consumer group for their growth market: newly divorced men. In the previous six courses students had access to Mediamark data, Nielsen data, and their own research.

In this project, when students had access to only Mediamark data and their own external research, each group come up with a different set of conclusions (Figure 3).

Figure 3. Conclusions from Data Analysis

P&G Group 1: 18-25 year olds were 15% less likely to use powder pre-treatment and 50% less likely to use a stick pre-treatment. They were 50% less likely to use Ultra Era and 77% less likely to use Wisk. This could be an untapped market. Tablet detergent could be marketed to those who work in the area of Natural Resources, Construction and Maintenance because they were 61% less likely to use tablets. They were 63% more likely to use Zout pre-treatment.

P&G Group 2: Consumers in the Northeast are 45% less likely to use powder detergent and 26% less likely to use powder for pre-treatment. These consumers are 84% more likely to use Wisk and 50% more likely to use Shouth Color Catcher for pre-treatment. In the South consumers are 25% more likely to use tablets. They are 51% more likely to use Gain Regular and 59% more likely to use Gain with Bleach. They are 31% less likely to use Tide Pure Essentials and 42% less likely to use Ultra Era Regular. Consumers in the Midwest are 69% more likely to use Ultra Era Regular and 66% more likely to use Ultra Era Max. They are 41% less likely to use tablets, 30% less likely to use powders, and 46% less likely to use Gain with Bleach. Consumers in the West are 36% more likely to use powder and 20% less likely to use tablets. They use Tide but are less likely to use other P&G detergents.

P&G Group 3: Men between the ages of 18-34 are 119% more likely to be Tablet users. Men between the ages of 18-34 who have never been married are light users of detergent. Men between 25-54 are less likely to use detergent.

Target Group 1: Consumers with household income of $150,000 are 22%, 54%, 24%, 26%, and 35% more likely to use wipes, stick, spray, powder, and gel stain removers. They are 73%, 46%, 22%, 30%, 41%, and 35% more likely to use Zout, Tide to Go, Spray 'N Wash, Shout, OxiClean, and Chlorox Bleach Pen brands. They are 29% and 47% more likely to be medium and heavy users of stain remover.

Target Group 2: Men represent 25% or less of the principal shoppers using fabric care products. Man 18-35 are 119% more likely to use tablet laundry detergent. Pre-treatments and stain removers are less likely to be used by men.

Target Group 3: Never married consumers are 29% less likely to be heavy users of laundry detergent while married consumers are 30% more likely to be heavy users of laundry detergent. Parents with children under the age of two are 50% more likely to use pre-treatments and stain removers. Parents with children under the age of one are 113% more likely to use wipes.

Figure 4. Joint Target Markets

Pair 1: The best current target market is married couples with kids. The best growth target market is the current market (getting more of them to purchase detergent at Target because they were purchasing detergent somewhere else).

Pair 2: The best current target market is women 18-34 in the Northeast and Midwest. The best growth market is men 18-34 in the West.

Pair 3: The best current target market is women 25-54 with kids. The best growth target market is men 18-34.

Some groups focused on the type of detergent, some focused on the section of the country, some focused on stain removers, some focused on income.

As a result, each pair of teams entered the first collaboration activity with a different idea of the current best target market and the best target market for growth. The variety of perspectives made for interesting discussions and each pair of teams agreeing to a different joint target market (Figure 4). This determination of joint target markets was more diverse than the decisions in the previous six times the course was taught. The variety of conclusions resulting from their analyses resulted in a greater depth of discussion during collaboration.

Answers to the questions asked earlier are as follows: The students did go beyond the superficial variables by doing this analysis, even though they only used Mediamark data. The students did compare different forms of the product (detergent, pre-treatment, stain removers), locations, and demographics in more depth than in previous semesters. The collaboration activities resulted in different joint target consumer groups for each Target/P&G team for the first time. In addition, none of the groups projected a Mediamark sheet of data to justify their selection of a target market during collaborations. Teams prepared tables that displayed data in formats that were easy to understand.

DISCUSSION

These were eye-opening exercises for students and instructors. As marketing professors, we are often bombarded with new and better electronic resources for our research and teaching. Few would argue that there is a dearth of data in cyberspace. The process of turning those data into information that can be used for making managerial decisions presents more of a challenge. This is especially true in the

case of secondary data that have been collected for some other purpose, at some other time, and usually, by some other researcher.

The first step in selecting the right information source – like any research endeavor – involves clearly defining the research question. Most textbooks describe the research process as a list beginning with 'define the research question' followed by 'decide which information you need.' Students usually respond to these deceptively simple steps with a single word – 'duh' -- as they seem obvious. Judging from the wide range of answers instructors get to the same question indicates that while that these two steps may be obvious and simple, they are not easy. Anybody who has completed a doctoral thesis knows that the biggest challenge is usually 'coming up with a topic.' Students, on the other hand, lack this experience and perspective. It is not until they see the wide range of interpretations of a simple question among their own peers that they realize that they really need to take the time to clearly define the question they are being asked before jumping in to answer it.

Skipping the problem definition step can lead students down the wrong path when solving a simple in-class assignment that may or may not affect their grades. However, in the real world of work, bosses may not be very clear about the type of information they ask their employees to gather. Rather, they may simply ask an employee to get some information about some general topic. If the supervisor does not clarify the question, there is a risk of receiving the wrong answer to the right question or vice versa. Such mistakes can be costly both to the employer and the employee.

Beyond the need to properly identify the question before searching for information, the students learned that not all data are created equally. Without critically evaluating the source of the data, the time period during which it is collected, or the format of how data are reported, students came up with a wide range of answers to the same question. It was only when they experienced these discrepancies that they appreciated the need to assess the quality of the data they reported. Once again, basing decisions on data that are not evaluated critically can be costly.

To the students, the exercise in the first project appeared to be a simple scavenger hunt for information using readily available electronic sources. Indeed, the Internet offered every student an answer to each question. The sharing of information among students in class and the reflective exercise revealed the true purpose of the exercise to the students: to show them what they need to consider before collecting data in terms of problem definition, and what they need to consider when evaluating the data they found BEFORE confidently reporting information back.

Allowing students to experience the frustration and failure of poorly planned secondary data research led them to conclude that although it is a powerful resource, the Internet can lead to potentially ill-informed decisions if not used properly. As the students came to realize, properly executed secondary research – while often

simpler and quicker than a full-fledged primary research study – is not easy and it takes a lot of time, effort and planning. This they did not know before the exercise. This led to frustration in completing a class exercise. What they did not know ultimately did not hurt them.

The emphasis on secondary sources through workshops and a challenging assignment is one approach to aiding students in gaining the skills and knowledge required to navigate through databases and on-line sources. The feedback from students on what they learned gives reasonable evidence that the learning goals for Project One were mostly achieved. While not all students achieved all learning outcomes, the majority of students gained valuable knowledge and skills.

Reading Mediamark and spreadsheet data are a basic Marketing class activity and assignment. However, students find it challenging and have not gone beyond a superficial analysis in the last six years of using this assignment in a particular class. The intent of Project Two is to help students look at a set of data across several related spreadsheets to better understand what the information means and how it can be used for decision-making. Since Mediamark data has presented a problem, it became the focus for this project. The purpose it to find a way for the students to better understand descriptive data, draw conclusions from it, and use it for decision-making. Understanding data is important for decision-making, creating models, and creating an effective presentation of data for managers who will be making data-drive decisions.

By presenting the steps of identifying relevant variables, creating tables to compare data across data sets, determining the significance of the data in the tables, and drawing conclusions, students will have a framework for doing their own analyses on another data set. With this introduction and explanation students did analyze the data in more depth, reached different decisions about target markets, presented data in a readable manner, and reached different determinations for a joint target market in their collaborations for the first time in six offerings of this class. This provides reasonable evidence that students need more instruction in how to analyze data sets.

CONCLUSION

Many of our marketing graduates will be expected to have the skills to find appropriate and accurate information using secondary sources. While a marketing research course is required in most marketing programs, the main focus of this course is usually on collecting and analyzing primary data. As the amount of secondary information increases at a rapid rate, the skills needed to find information through digital sources become more important. As more of this data is stored in data warehouses, the ability to analyze information from this type of secondary data has also become more important.

Learning what they did not know through the experiential exercise in Project One and appreciating and respecting the power and limitations of secondary exercise

is a positive outcome. Hopefully, the exercise impressed upon them the need to approach secondary research in their careers with clarity and caution because, in the real world of work, what you do not know can hurt you. Socrates was wise in warning us to know what we don't know.

Developing a process for approaching and analyzing a set of secondary data is increasingly important with all the data being gathered from POS data, data warehouses, and companies selling data sets. Often the amount of data available is overwhelming. Learning how to approach a vast amount of data, how to compare information, how to draw conclusions from the data, and how to create tables that present the data in a way that is easy to understand is a positive outcome.

Being able to gather appropriate secondary data and analyze data from prepared data sets are skills that are important for marketing students to learn. These in class sessions were effective as tools for helping students develop these skills. More attention needs to be given to these skills sets and more material created to facilitate the process of learning these skills.

Knowledge workers of the future need the skills to gather appropriate, accurate data from an ever-increasing array of sources. In addition, knowledge workers of the future need the skills to turn the data into useful information that can be used to facilitate decisions in organizations. Without these skills, students will find it difficult to succeed when surrounded by the ever-increasing sea of data.

REFERENCES

"Aftershave Lotion & Cologne for Men" (2010), Health & Beauty Aids, Fall, Mediamark Data Reporter Online Database, California State University San Marcos Library, Retrieved June 8, 2012. (Used in Last 6 Months, Heavy Users, Medium Users, Light Users).

Campbell, Diane K. and Ronald G. Cook (2010) "An Experiential Market Research Analysis: A Partnership between Teaching and Library Faculty" *Journal of Business & Finance Librarianship* Issue 15, pp. 171 – 178.

Castleberry, Stephen B. (2001) "Using Secondary Data in Marketing Research: A Project that melds Web and off-Web sources" *Journal of Marketing Education*, Vol. 23, no. 3, pp. 195-203.

"Facial Moisturizers" (2010), Health & Beauty Aids, Men, Fall, Mediamark Data Reporter Online Database, California State University San Marcos Library, Retrieved June 8, 2012. (Used in Last 6 Months, Medium Users, Light Users, Nivea for Men, Anti-Wrinkle, Cleansing, Moisturizing, Night Cream).

Kumar, V., David A. Aaker and George S. Day (1999) *Essential of Marketing Research* New York: John Wiley & Sons.

Kumar, V., David A. Aaker and George S. Day (1999) *Essential of Marketing Research* New York: John Wiley & Sons.

Malhotra, Naresh K. (2010) "Exploratory Research Design: Secondary Data" Chapter 4 in *Marketing Research: An Applied Orientation* 6th ed. New York: Prentice Hall. pp. 99-131.

Malhotra, Naresh K. and Mark Peterson (2006) *Basic Marketing Research: A Decision-Making Approach* 2nd ed. (Upper Saddle River, N J: Pearson/Prentice-Hall).

McDaniel, Carl and Roger Gates (2002) *Marketing Research* 5th ed. United States: South-Western/ Thomson Learning.

Murphy, John W. and Christian A. Schlaerth (2010) "Where are your Data? A Critique of Secondary Data Analysis in Sociological Research" *Humanity & Society* Vol. 34, November. pp. 379-390.

Nie, Norman H. (2011), "The Rise of Big Data Spurs a Revolution in Big Analytics," White Paper, Revolution Analytics.

The Economist (2010) "The data deluge" Downloaded on 9/28/2012 from http://www.economist.com/node/15579717/print

"The Economist (2010) "Needle in a Haystack" Downloaded on 9/28/2012 from http://www.economist.com/node/15557497

"The Economist (2010) "Data, Data Everywhere" Downloaded on 9/28/2012 from http://www.economist.com/node/15557443

The Economist (2010) "Handling the Cornucopia" Downloaded on 9/28/2012 from http://www.economist.com/node/15557507

Transforming the Retail CMO Through Data-Driven Marketing (2012), White Paper, Retail TouchPoints, http://www.retailtouchpoints.com/resource-center/51-white-papers/1666-transforming-the-retail-cmo-through-data-driven-marketing.