Paper 3190-2015

Educating Future Business Leaders in the Era of Big Data

Tonya Etchison Balan, NC State University

ABSTRACT

At NC State University, our motto is "Think and Do." When it comes to educating students in the Poole College of Management, that means that we want them to not only learn to think critically but also to gain hands-on experience with the tools that will enable them to be successful in their careers. And, in the era of big data, we want to ensure that our students develop skills that will help them to think analytically in order to use data to drive business decisions. One method that lends itself well to "thinking and doing" is the case study approach. In this paper, we will discuss the case study approach for teaching analytical skills and highlight the use of SAS® software for providing practical, hands-on experience with manipulating and analyzing data. The approach will be illustrated with examples from specific case studies that have been used for teaching introductory and intermediate courses in business analytics.

INTRODUCTION

In this day and age, there is no question that big data is a big deal. We are generating data at an alarming rate. Studies show that over 90% of the world's data was created in the past 2 years. In 2014, every minute:

- Facebook users shared nearly 2.5 million pieces of content.
- Twitter users tweeted nearly 300,000 times.
- Instagram users posted nearly 220,000 new photos.
- YouTube users uploaded 72 hours of new video content.
- Apple users download nearly 50,000 apps.
- Email users sent over 200 million messages.
- Amazon generated over $80,000 in online sales.

Therefore, the question for most organizations is not where to find data – it's how to effectively extract useful information from the data. The obvious answer for tapping into that information is analytics. However, it takes more than clever applications of statistical models or machine learning algorithms to successfully compete on analytics. You also need to understand how to use the results of your analyses to drive business strategy. You can build the best data science team in the world and build the most sophisticated models. You can uncover the most interesting trends and relationships. But, if you can't use that information to drive business value – it's really just an academic exercise.

Most people have seen the 2011 McKenzie report that stated: "The United States alone faces a shortage of 140,000 to 190,000 people with analytical expertise and *1.5 million managers and analysts with the skills to understand and make decisions based on the analysis of big data*." (Emphasis mine.) This shortage of data savvy managers is what many business schools are trying to address.

In the Poole College of Management at NC State University, our goal is to ensure that our MBA students develop skills that will enable them to think analytically in order to use data to drive business decisions. In keeping with the university's motto – "Think and Do" – that means that we want them to not only learn to think critically but also to gain hands-on experience with the tools that will enable them to be successful in their careers. This paper will discuss the skill sets needed by business leaders in the era of big data and will describe a case study method that has proven effective for teaching those skills in the classroom.

WHAT DOES AN ANALYTICAL MBA NEED TO KNOW?

As an educator, having identified the need to produce quantitative MBAs, the question then becomes – what skillset does an MBA need to possess in order to lead effectively in an analytical organization? Do they need to understand statistical theory? Perhaps. Do they need to be familiar with analytical software packages? Definitely. Do they need to understand the application of analytical methods to common business problems? Absolutely.

In short, an MBA needs to be able to navigate the entire analytics process in order to lead effectively in an analytically driven organization. They don't need to be the experts in every aspect of the process, but they do need to have a solid understanding of what it takes to be successful so that they can collaborate with the IT organization and the data scientists and the operations managers in an organization.

In its simplest form, the analytics process looks something like this:

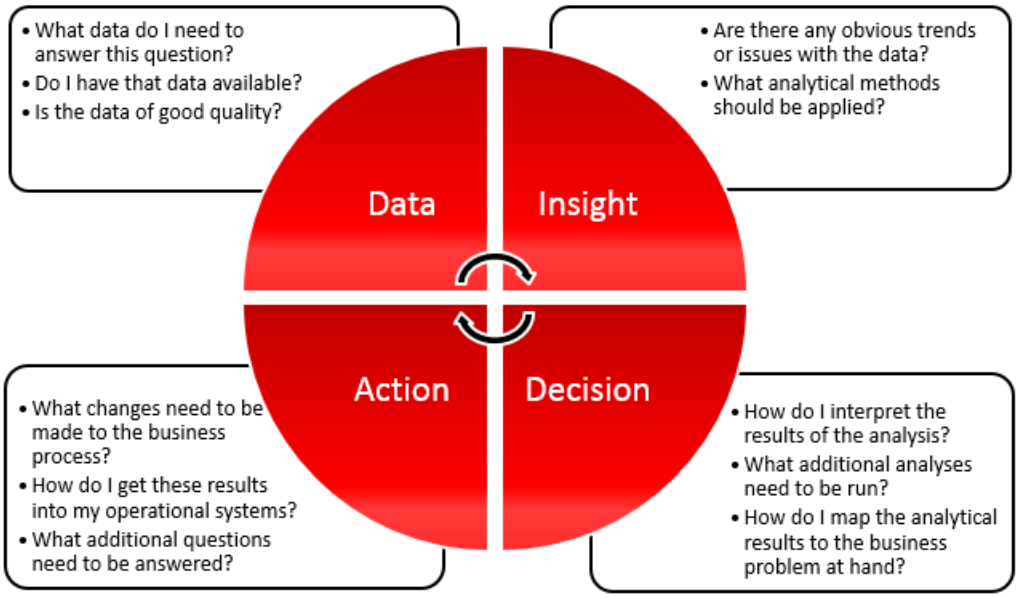


Figure 1: The Analytics Process

Being able to navigate this process relies heavily on the integration of critical thinking across the entire lifecycle. In order to achieve that goal, you need to understand the tools of the trade. You need to understand what business problems can be addressed with analytics, and you need to be able to translate business problems into discrete questions with measurable outcomes. You need a basic understanding of what data is available for answering those questions, what it takes to get access to that data, and how to maintain data privacy. You need a solid understanding of analytical methods that can be applied to big data problems - which means going beyond what you learn in a first year introductory statistics course. And, you need to understand how to interpret the results of your analysis and translate those results into actionable decisions that can be implemented as part of your operational business processes. In other words, you need to be able to follow a business problem from start to finish, and you need to be able to converse with all of the interested parties along the way.

In order to teach that kind of holistic critical thinking, you have to go beyond traditional classroom instruction. Simple textbook problems with sanitized data sets mask the complexity encountered in most real-life analytical processes. And, those problems tend to focus in a very isolated way on how to properly

execute the analytical method. Since MBA education often relies on the analysis of case studies, adapting that methodology for teaching analytics seemed like a natural approach. Following the motto of "Think and Do," this case study approach enables students to start with a business problem, explore the analytical methods and software that can be used to address the problem, and then put together an action plan identifying ways that the results of their analyses could be used to drive strategy. It allows you to simulate a real world analytics process in a controlled classroom environment.

HOW DO YOU DESIGN AN ANALYTICS CASE STUDY COURSE?

While there are many ways that you could organize the material, one approach that has worked well is to break the course up into modules based on a particular analytical method. Each module begins with an introduction to the analytical concept with an emphasis on how it is typically applied to real problems. Software is used heavily to illustrate the application of the method to realistic data sets. Class time is primarily spent working through real problems together with lots of time for discussion. At the completion of each module, the students are presented with a new case and are asked to analyze the case and present their solution along with their recommendations. Because predictive modeling is as much art as science, they are encouraged to work together on these homework problems so that they can get a broader understanding of how different people might approach the problem. A midterm project and a final project present the students with more complex business scenarios requiring them to apply multiple techniques in order to arrive at a solution. These projects are done independently in order to better assess each individual student's understanding of the material.

The most difficult aspect of this approach is finding real data sets that can be used to create the cases. The idea is to present the students with realistic data that is messy – that has outliers or missing values or requires transformation of some variables – but that also addresses a compelling business problem.

Here are a few examples that have worked well. Links to these and other resources are included at the end of this paper.

Analytical Method	Software Packages	Case Study Resource
Exploratory Data Analysis Inferential Statistics Linear and Logistic Regression	JMP	JMP Case Study Library
Linear Regression	JMP SAS®/STAT SAS® Enterprise Miner	"Multiple Regression: How Much Is Your Car Worth?" Shonda Kuiper, Grinnell College Discussion and link to data can be found in the *Journal of Statistics Education* Volume 16, Number 3 (2008)
Classification Methods	JMP SAS® Enterprise Miner	Bank Marketing Data Set UCI Machine Learning Repository Discussed in Moro et al. A Data-Driven Approach to Predict the Success of Bank Telemarketing. *Decision Support Systems*, Elsevier, 62:22-31, June 2014
Data Visualization Regression Methods Decision Trees Clustering and Segmentation	SAS® Visual Analytics SAS® Visual Statistics	Teradata University data library and case studies

Table 1: Select Resources for Developing Case Studies

EXAMPLE

In order to illustrate this approach, a summary of the "Introducing Organics" case study is provided here. This particular case is based off of the ORGANICS data set included with the "Applied Analytics Using SAS® Enterprise Miner" training materials that SAS® provides to educators.

SYNOPSIS OF THE BUSINESS SCENARIO

A supermarket is offering a new line of organic products. The supermarket's management wants to determine which customers are likely to purchase these products. They are interested in understanding the profitability of their customers as well. If they notice that customers who purchase organic products are also highly profitable customers that makes the additional cost of stocking organic products more palatable for the management.

In order to drive sales of the new organic products, the supermarket provided coupons for the products to the members of their loyalty program and collected data on whether or not these customers purchased the products. The resulting data set contains 13 variables and over 22,000 observations. Several of the variables have missing values. A list of the variables, along with a brief description, is given in the table below:

Name	Description
ID	Customer loyalty identification number
DemAffl	Affluence grade on a scale from 1 to 30
DemAge	Age, in years
DemCluster	Type of residential neighborhood
DemClusterGroup	Neighborhood group
DemGender	M = male, F = female, U = unknown
DemRegion	Geographic region
DemTVReg	Television region
PromClass	Loyalty status: tin, silver, gold, or platinum
PromSpend	Total amount spent in the store this year
PromTime	Time as a loyalty card member
TargetBuy	Organics purchased? 1 = yes, 0 = no
TargetAmt	Number of organic products purchased

Table 2: Organics Data Dictionary

STUDENT TASKS AND QUESTIONS

Students are given two major analytical tasks: (1) build a predictive model for classifying customers based on their likelihood to purchase organic products and (2) build a regression model for predicting the amount that a customer will spend at the store in a year.

At this point, the students have been introduced to JMP, SAS® Enterprise Miner, SAS® University Edition, and Excel. They have the option to choose the software package that they would like to use for the analysis, and they are encouraged to "mix and match" in order to analyze the data.

Students are asked to answer the following questions based on the data and/or their predictive models and are instructed to provide graphs and/or output from statistical analyses to support their answers:

1. How would you characterize the "profitability" of the customers who purchased organic products vs those who didn't purchase organic products? Do they spend similar amounts, or does there appear to be a significant difference? Do customers who purchase organic products spend more at your store in general than customers who don't purchase organic products (or vice versa)?

2. Are there any noticeable differences in the percentage of customers who purchase organic products across the different loyalty status groups?

JMP works well for answering questions like those posed in questions (1) and (2) above. For example, the Graph Builder makes it easy to create side-by-side box plots.

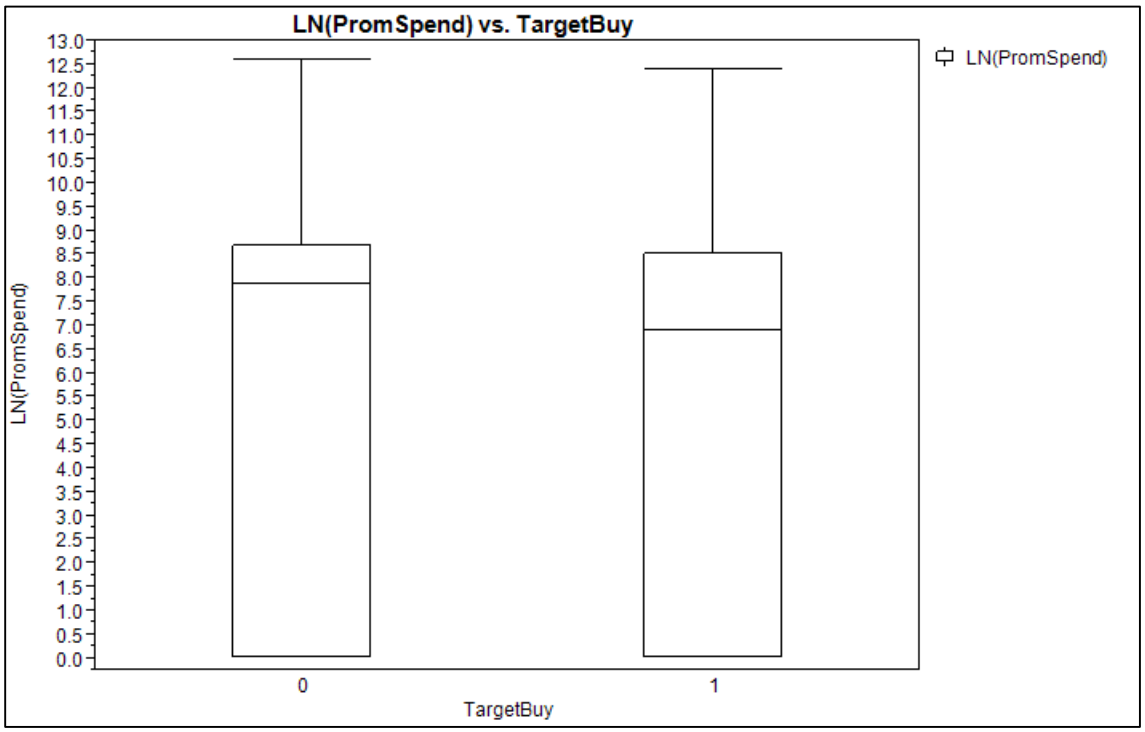
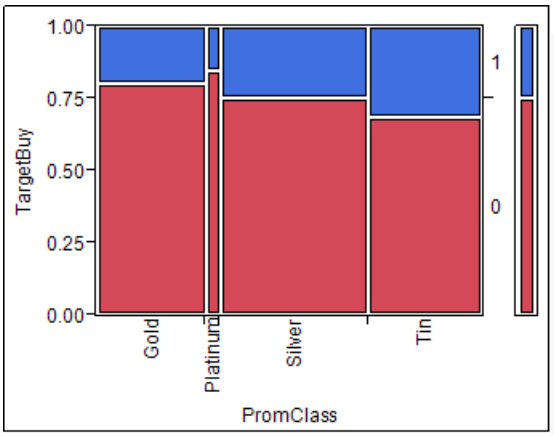


Figure 2: Side-by-Side boxplots generated by JMP

And, the Fit y-by-x platform in JMP delivers a heat map and frequency table with just a couple of clicks.



TargetBuy

PromClass	Total % Row %	0	1	
Gold		22.90 80.46	5.56 19.54	28.46
Platinum		3.20 84.76	0.58 15.24	3.78
Silver		29.07 75.36	9.50 24.64	38.57
Tin		20.06 68.72	9.13 31.28	29.19
		75.23	24.77	

Figure 3: JMP output for categorical data

Further, the interactive nature of JMP lends itself well to doing live analysis of case study data in class.

Based on the boxplots, heat map, and frequency table, students should be able to infer that there is very little difference in the profitability of customers who buy organic products and those that don't. In particular, it does not appear that the high value customers are the ones who are more inclined to buy organic products. Further, it's the customers in the lower levels of the loyalty program (silver and tin) who seem to have been influenced by the coupon to purchase organic products. This phenomenon could be due to their preference for organic products – or it could be that these are the customers who are always looking for a bargain!

3. What factors seem to have the most influence on customer profitability? Are there any actions that you might be able to take based on this information?

4. What factors seem to have the most impact on a customer's likelihood to purchase organic products? Based on your model, how would you describe the "typical" organic products customer?

Questions (3) and (4) require the students to think about the results generated by their predictive models. For example, SAS® Enterprise Miner could be used to create a decision tree for predicting which customers are most likely to purchase organic products that might look like the following:

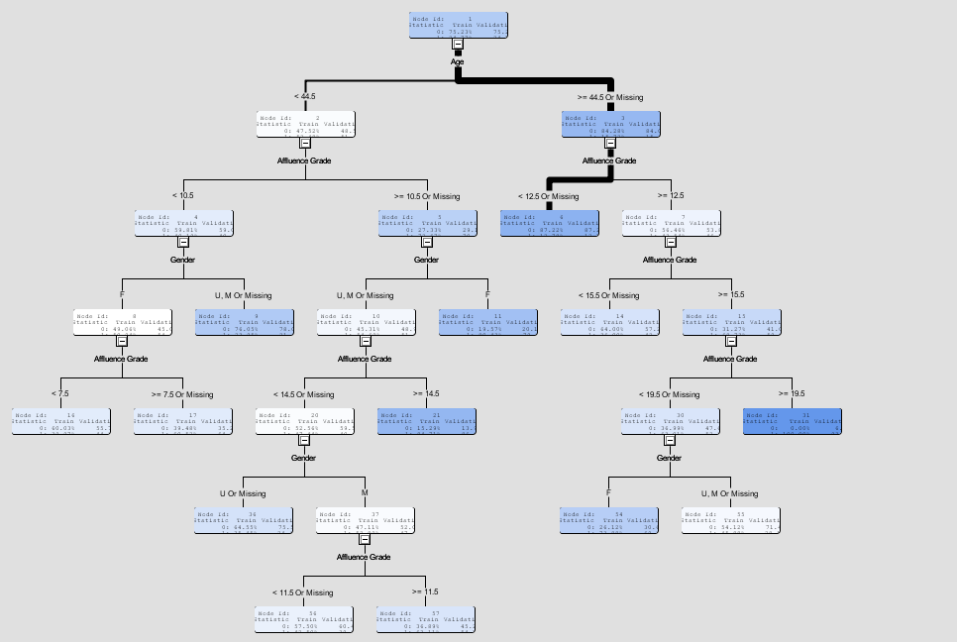


Figure 4: Decision Tree generated by SAS® Enterprise Miner

Based on this decision tree, the students should recognize that Age, Affluence Grade, and Gender are the most important factors in determining which customers will purchase organic products. In particular, female customers are more likely to purchase the organic products, regardless of age. For older customers, those that are more affluent are also more likely to purchase organic products.

Remember, the goal for these students is not to create the best possible statistical model. Ideally, when they enter the business world, they will be working with a team of statisticians or data scientists who will do that heavy lifting for them. The goal is for them to get a basic understanding of the statistical modeling methods while focusing on how to interpret the results.

5. As a result of this modeling exercise, what strategy would you suggest to the supermarket if they are interested in trying to grow their organic product line while continuing to retain profitable customers and grow their bottom line?

6. Five new customers signed up for the loyalty program this month. Three of them have recently relocated to the area and have been a part of the loyalty program for the supermarket chain in another state. Their information is provided. For each of these customers, determine whether they are a good candidate for a promotion related to organic products (e.g., are they likely to purchase organic products or not) and predict the amount that they can be expected to spend in the supermarket over the course of the year.

Questions (5) and (6) require the students to think about how the results of their analyses could impact the supermarket's strategy regarding the introduction of organic products. And, it requires them to apply the results of their models to new data in order to determine how to best treat these customers going forward.

Therefore, this relatively simple case study enables the students to navigate the full analytical process using realistic data and commonly used software while encouraging critical thinking through the application of their analyses to a specific business problem.

CONCLUSION

Educating MBAs in the era of big data requires an integrated method for teaching analytics as part of a bigger business process. A case study approach based on real world examples utilizing messy data presented in an interactive classroom environment enables students to explore and understand all aspects of the analytics process. Ultimately, this prepares them for a practicum experience which is the final requirement for students who want to receive a Certificate in Decision Analytics as part of their MBA degree from NC State. The practicum may be purely focused on analytics or it may be a project related to their concentration (e.g., marketing, finance, supply chain, etc.), but it enables them to work directly with a customer to solve a real business problem. And, ultimately, this method aligns well NC State's mission of encouraging students to "Think and Do".

REFERENCES

SAS Institute Inc. 2011. *Applied Analytics Using SAS® Enterprise Miner Course Notes.* Cary, NC. SAS Institute Inc.

Kuiper, Shonda. 2008. "Introduction to Multiple Regression: How Much Is Your Car Worth?" *Journal of Statistics Educationn*, Volume 16, Number 3.

Kuiper and Sklar. "Practicing Statistics: Guided Investigations for the Second Course eChapter." Available at http://www.pearsonhighered.com/kuiper1einfo/echapter/index.html#/1

Moro, S., Cortez, P., and Rita, P. June 2014. "A Data-Driven Approach to Predict the Success of Bank Telemarketing." *Decision Support Systems.* Elsevier. 62:22-31.

McKinsey & Company. "Big Data: The Next Frontier for Competition." Available at http://www.mckinsey.com/features/big_data

RESOURCES

JMP. "JMP Case Study Library." Available at http://www.jmp.com/en_us/academic/case-study-library.html

Teradata. "Teradata University Network." Available at http://www.teradatauniversitynetwork.com/

Center for Machine Learning and Intelligent Systems. "UCI Machine Learning Repository." Available at http://archive.ics.uci.edu/ml/

Kaggle. "The Home of Data Science." Available at https://www.kaggle.com/

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Tonya Etchison Balan
North Carolina State University
Poole College of Management
tebalan@ncsu.edu