6-2017

The Characteristics of Universities Offering Data Analytics Programs: An Analysis of US Regional Masters Universities

Troy Strader
Drake University, Troy.Strader@drake.edu

Andrew Bryant
Drake University, Andrew.Bryant@Drake.edu

The Characteristics of Universities Offering Data Analytics Programs: An Analysis of US Regional Masters Universities

Troy J. Strader
Drake University
Troy.Strader@Drake.edu

Andrew Bryant
Drake University
Andrew.Bryant@Drake.edu

ABSTRACT

A number of colleges and universities have developed new undergraduate and graduate data analytics related programs in the past five years. The purpose of this study is to identify the unique resource-related characteristics of the universities that have developed these programs when compared to those that have not. An analysis of 391 regional masters universities in the US finds that universities with data analytics programs are more likely to be in larger cities and have larger student enrollments, better educational quality rankings, and an existing statistics and/or actuarial science program. These findings support the idea that data analytics programs are more likely to be developed when universities have access to a larger number of businesses and governmental organizations, and have larger financial, student, and faculty resources. The study concludes with a discussion of implications for educational administration and future research directions.

Keywords

Data analytics education, business analytics, data science, curriculum development, regional masters universities, university rankings, educational administration

INTRODUCTION

It is estimated that by 2020 about 1.7 megabytes of data will be created each second for every person on earth (Marr, 2015). This data includes pictures and videos to document one's life, the details of every transaction across every global organization's supply chain, and every interaction a government agency has with its citizens. Businesses and government agencies are increasingly viewing their data as an asset that can be utilized to better serve their customers and improve their competitive position, but identifying value in vast amounts of data requires skills that may not be readily available in their current employees. This situation has led to increased demand for data analysts across a wide range of industries and governmental functions. It is difficult to forecast future demand for data analysts, but one prediction is that that there will be a 50-60% shortfall for supply of data analysts in the US by 2018 (Orihuela and Bass, 2015). This scenario provides opportunities for current and future students as well as educational institutions at all levels.

Growth in data analytics jobs has led to an increase in student interest for analytics degree programs. The natural response has been an increase in development and implementation of undergraduate and graduate analytics programs in both small and large US colleges and universities (Chiang et al., 2012). Given that this is a new educational area, there is an opportunity to now reflect on how the development of data analytics programs has evolved over the past decade. The purpose of this study is to address the following questions:

1. What are the characteristics of the colleges and universities that have data analytics related programs?
2. What are some of the reasons that motivated colleges and universities to develop their undergraduate data analytics programs?

This study addresses issues in two research areas – the current state of business intelligence and analytics education, and the design and development of business intelligence and analytics curriculum (Wang, 2015). The following sections describe the study's background and findings. First, the study methodology and data are described. This is followed by a discussion of the study's background, hypotheses and findings. Finally, the paper concludes with a discussion of overall conclusions, implications, and directions for future research.

STUDY METHODOLOGY

To address the issues described above, a representative sample of US colleges and universities was identified. The sample includes all of the regional masters universities identified in the US News rankings for 2016 that were ranked 100 or better in each of four regions (Usnews.com Midwest, 2016; Usnews.com North, 2016, Usnews.com South, 2016; USnews.com West, 2016). Universities in this category offer a full range of undergraduate degree programs and some masters degrees, but few doctoral-level programs. The US News ranking web pages provided information about each universities region, ranking, and undergraduate enrollment. Next, each university website was reviewed to identify which universities had a data analytics related undergraduate or graduate degree program. This data was typically found through links on an academics page or through a search using the terms data analytics, business analytics, or data science. In addition, it was also determined whether each university had existing majors in actuarial science, statistics, or information systems. Finally, population data was identified for each university's city (or metropolitan area) from a US census website (Census.gov, 2013). The combination of data collected for each ranked university was compiled and analyzed in SAS v9.4. The sample characteristics are summarized in Table 1.

	North	South	Midwest	West	Overall
Number of Universities in Sample	102	97	103	89	391
University Undergraduate Enrollment					
Minimum	136	560	1,010	326	136
Average	4,983	6,204	5,945	7,862	6,195
Maximum	18,807	49,744	25094	35,616	49,744
University City Population (thousands)					
Minimum	2	1	2	9	1
Average	4,073	687	1,997	2,954	2,431
Maximum	19,567	6,538	9,461	12,829	19,567
Universities with Related Undergrad Majors					
Actuarial Science and/or Statistics	26	18	33	13	90
Information Systems	64	71	75	55	265

Table 1. University Sample Demographics

An analysis of the data collected for each university showed that 57 (15%) of the universities have an undergraduate data analytics related major. The most common major names appear to be data science, business analytics, and bioinformatics, but there are a wide range of other names used to describe these programs. A review of the websites also showed that there are 53 (14%) schools with graduate (masters level) programs. Again, some of the most common program names are data science or business analytics, but there are also a number of bioinformatics programs. Interestingly, only 11 (less than 3%) of the 391 universities reviewed have both an undergraduate and graduate data analytics program. It appears that the most common strategy is to focus on either the undergraduate or graduate program, but not both.

BACKGROUND AND STUDY HYPOTHESES

After identifying which universities have data analytics programs, the next step was to identify the characteristics of the schools that have data analytics programs when compared with those that do not. Five hypotheses describe the potential resource-related factors that may have increased the likelihood that a data analytics program was developed at a particular school.

The first characteristic relates to overall university academic quality. Universities with better academic rankings (lower ranking numbers) are often more innovative than universities that have lower overall academic quality. This makes it more

likely that they would develop new curricula to keep up with changing market needs and student demand. This leads to the first hypothesis.

H1. Universities with better overall academic rankings are more likely to have data analytics programs.

The next two issues considered relate to the resources that universities have to support the development of data analytics programs. Larger cities will have more businesses and government agencies to provide support for data analytics educational programs and jobs. Some examples would include opportunities for program funding, data, and funding for real-world projects, and opportunities for student internships and full-time employment. Also, universities with larger student enrollments will have greater financial and personnel resources to develop new programs when compared with very small universities. This leads to the second and third hypotheses.

H2. Universities in larger cities are more likely to have data analytics programs.

H3. Universities with larger undergraduate enrollments are more likely to have data analytics programs.

The final two hypotheses relate to whether universities have resources available from existing majors in areas associated with data analytics such as statistics or computer information systems. If schools already have faculty and degree programs in these areas, then there is a shorter path to developing new data analytics programs. If a school does not have either of the programs then it would be much more difficult to start a new program given the need for funding, administrative support, and hiring new faculty. Most universities have math and computer science programs so these related areas were not considered as a differentiator when comparing universities. This leads to the final two hypotheses.

H4. Universities with existing statistics or actuarial science majors are more likely to have data analytics programs.

H5. Universities with existing information systems majors are more likely to have data analytics programs.

FINDINGS

A logistics regression model was used to assess H1 through H5. The independent variable parameter estimates are shown in Table 2. The modeled dependent variable is a binary variable where one indicates that a school has either an undergraduate major in data analytics, or a graduate level master's degree program in data analytics, or both. Zero indicates that a university does not have a data analytics program at any educational level.

Parameter (modeled DV = 1)	Estimate	p-value
Intercept	-1.18	< 0.001***
University Ranking	-0.02	< 0.001***
City Population (in millions)	0.09	< 0.001***
Undergraduate Enrollment (in thousands)	0.04	0.071*
Statistics and/or Actuarial Science Major	0.97	< 0.001***
Information Systems Major	0.43	0.156

***p < .01; **p < .05; *p < .10

Table 2. Parameter Estimates to Test Hypotheses

In total, the variables in our model enhance the overall model fit. A likelihood ratio test of the full model against an intercept only model indicates that, as a set, these variables significantly increase the model fit (χ^2 = 63.76, d.f. = 5, p < 0.001). Further, a Receiver Operator Characteristic (ROC) curve, shown in Figure 1, was used to examine the model fit by plotting the true positive classification (sensitivity) against false positives (1 - specificity). The area under our ROC curve was 0.759 indicating a good model fit. For comparison sake, the area under a perfect model ROC curve equals one. Finally, Tjur's coefficient of discrimination (2009) was calculated (d = 0.163). This statistic indicates a benefit associated with our model when comparing predicted probabilities between universities with and without data analytics programs.

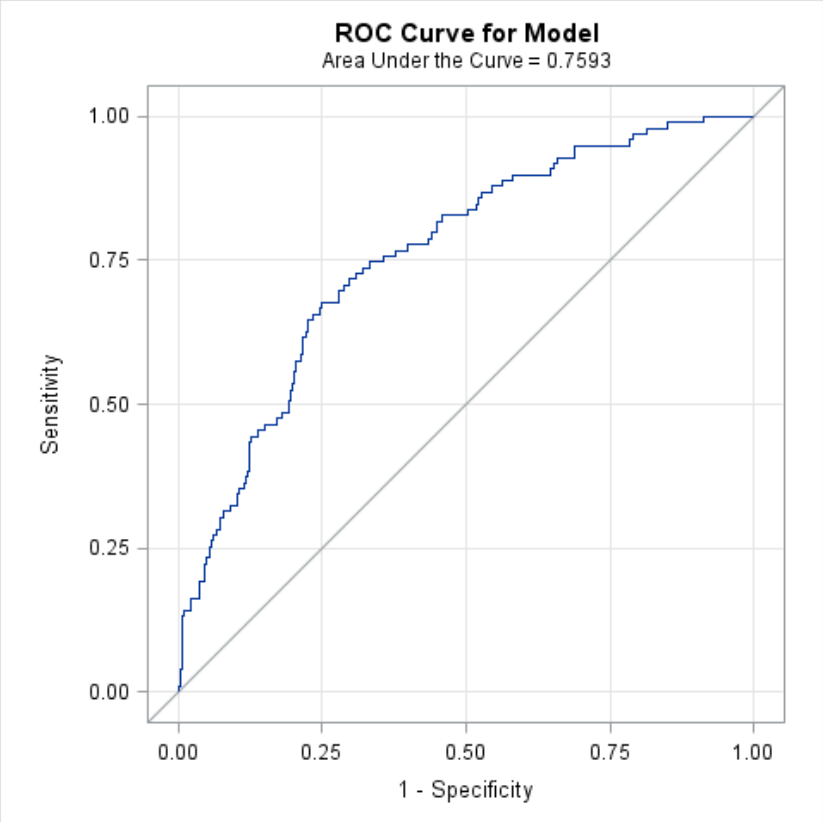

Figure 1. Receiver Operator Characteristics

The model shows that there is a statistically significant negative relationship ($\beta = -0.02$, Wald $\chi^2 = 20.35$, d.f. = 1, $p < 0.001$) between university rank and the existence of a data analytics programs at a university. Universities with better overall academic quality (lower ranking numbers) are more likely to have some type of data analytics program (either undergraduate, graduate, or both). This provides support for H1.

In order to examine the robustness of our findings, we test our finding under an alternate specification. Specifically, as university ranking are related to the regional location of the university, we include dummy variables to account for region location and compare model fits. Using the Akaike Information Criterion (AIC) to compare models with a different number of independent variables, we find a superior fit (smaller is better) when region location is excluded from the model (AIC = 390.71) compared to a model where region location is included in the model (AIC = 391.84). Regardless, even when the model with a regional variable is used, university ranking is still negative and significant ($\beta = -0.02$, Wald $\chi^2 = 22.04$, d.f. = 1, $p < 0.001$). The consistency of these results further support H1.

The model also shows that universities with larger city populations ($\beta = 0.09$, Wald $\chi^2 = 12.47$, d.f. = 1, $p < 0.001$) and larger undergraduate enrollments ($\beta = 0.04$, Wald $\chi^2 = 3.24$, d.f. = 1, $p < 0.05$, one-tail test) are significantly more likely to have a data analytics program. This supports H2 and H3. Finally, it is found universities with statistics and/or actuarial science majors are significantly ($\beta = 0.97$, Wald $\chi^2 = 11.95$, d.f. = 1, $p < 0.001$) more likely to have data analytics programs. This supports H4.

Interestingly, the relationship between information systems majors and data analytics programs is positive but not statistically significant ($\beta = 0.43$, Wald $\chi^2 = 2.01$, d.f. = 1, $p = 0.156$). H5 is not supported. One explanation could be that some information systems programs have evolved into an analytics related program. Information systems was available prior to the development of the data analytics program, but it no longer offered as a separate major program. Thus, information

systems did provide resources to support the development of a data analytics program, but the data only provides a snapshot of the situation at this one point in time.

The overall findings are that universities with better overall academic rankings, larger university or city populations, and existing majors in related statistical and quantitative areas are more likely to have a data analytics program, ceteris paribus.

An additional issue addressed in this study is to dig deeper into the reasons why some universities have developed undergraduate data analytics majors. To answer this question a short questionnaire was sent to faculty responsible for administering undergraduate data analytics majors. The focus was on the more common data science and business analytics majors. Questionnaires were not sent to administrators for bioinformatics programs. About 35 questionnaires were sent to program administrators and seven responses were received. The three main questions related to who initiated the development of the undergraduate data analytics program, what factors motivated the development, and when the program started.

Qualitatively examining the seven responses, it appears that there are three groups that initiated the development of the programs. It was evenly split between university administrators, college administrators, and faculty. Industry partner companies may support the programs, but they did not initiate the idea for these programs. There are also two primary motivators for the development of these programs. The primary reasons stated for why programs were created was that there was a need to increase undergraduate student enrollment and a need to maintain an up-to-date curriculum. Of the seven schools that returned the questionnaire, one program started in Fall 2013, two in Fall 2014, one in Fall 2015, and three in Fall 2016.

CONCLUSIONS AND IMPLICATIONS

Data analytics is a very new field of study that incorporates a number of long standing academic areas including math, statistics, computer science, information systems, and application areas in business and the sciences. The programs fill a need in industry and government for people that can analyze vast amounts of data to support organizational decision making and improve their competitive position. The findings from this study provide some insights into the current practices for data analytics programs in the US regional masters universities.

Data analytics undergraduate majors and graduate degree programs have almost all been introduced in the past five years. In the sample of universities evaluated in this study, about 25% have a program (99 out of 391). It is likely that the schools that intend to develop a data analytics program have already done so. Growth in new programs should slow over the next five years although it can be expected that it is more likely that new programs would come from schools with high overall academic quality, larger enrollments, larger cities, and existing majors in statistics and related quantitative areas. New programs can also be expected to be initiated by university and college administrators and faculty who see an opportunity to attract new students and keep their curriculum up-to-date.

Several directions for future research arise from this study. First, this study's hypotheses could be addressed using a sample of larger research focused universities or a sample of universities outside the US. Additionally, the motivations for data analytics program development could be addressed in more detail across a wider range of universities. This could provide additional insights for educational administrators to identify the best processes to use when development new technology-related interdisciplinary undergraduate and graduate programs.

The amount of data collected by organizations will continue to grow in the future so the need for data analytics experts can also be expected to continue into the foreseeable future. Meeting this need will require partnerships between universities, businesses, and governmental agencies.

REFERENCES

1. Census.gov (2013, March) Population change for metropolitan and micropolitan statistical areas in the United States and Puerto Rico: 2000 to 2010, accessed from http://www.census.gov/population/www/cen2010/cph-t/CPH-T-5.pdf.

2. Chiang, R. H. L., Goes, P. and Stohr, E. A. (2012) Business intelligence and analytics education, and program development: A unique opportunity for the information systems discipline, *ACM Transactions on Management Information Systems*, 3, 3, 12.

3. Marr, B. (2015, September 30) Big data: 20 mind-boggling facts everyone must read, *Forbes*, accessed from http://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/.

4. Orihuela, R. and Bass, D. (2015, June 4) Help wanted: Black belts in data, *Bloomberg Businessweek*, accessed from http://www.bloomberg.com/news/articles/2015-06-04/help-wanted-black-belts-in-data.

5. Tjur, T. (2009) Coefficients of determination in logistic regression models — A new proposal: The coefficient of discrimination, *The American Statistician*, 63, 4, 366-372.

6. Usnews.com (2016) Regional universities Midwest rankings, *U.S. News & World Report*, accessed from http://colleges.usnews.rankingsandreviews.com/best-colleges/rankings/regional-universities-midwest.

7. Usnews.com (2016) Regional universities North rankings, *U.S. News & World Report*, accessed from http://colleges.usnews.rankingsandreviews.com/best-colleges/rankings/regional-universities-north.

8. Usnews.com (2016) Regional universities South rankings, *U.S. News & World Report*, accessed from http://colleges.usnews.rankingsandreviews.com/best-colleges/rankings/regional-universities-south.

9. Usnews.com (2016) Regional universities West rankings, *U.S. News & World Report*, accessed from http://colleges.usnews.rankingsandreviews.com/best-colleges/rankings/regional-universities-west.

10. Wang, Y. (2015) Business intelligence and analytics education: Hermeneutic literature review and future directions in IS education, in *Proceedings of the Twenty-first Americas Conference on Information Systems*, August 13-15, Puerto Rico, Association for Information Systems, 1-10.