Data Analytics vs. Data Science: A Study of Similarities and Differences in Undergraduate Programs Based on Course Descriptions

Cheryl L. Aasheim
Susan Williams
Department of Information Technology
Georgia Southern University
Statesboro, GA 30460, USA
caasheim@georgiasouthern.edu, rebstock@georgiasouthern.edu

Paige Rutner
Adrian Gardiner
Department of Information Systems
Georgia Southern University
Statesboro, GA 30460, USA
prutner@georgiasouthern.edu, agardine@georgiasouthern.edu

ABSTRACT

The rate at which data is produced and accumulated today is greater than at any point in history with little prospect of slowing. As organizations attempt to collect and analyze this data, there is a tremendous unmet demand for appropriately skilled knowledge workers. In response, universities are developing degree programs in data science and data analytics. As a contribution to the design and development of these programs, this paper presents findings from a review of the descriptions of courses offered in a small sample of undergraduate programs in data science and data analytics. Our investigation clarifies and illustrates the similarities and differences between undergraduate data analytics and data science programs.

Keywords: Data analytics, Job skills, Emerging technologies, Program improvement

1. INTRODUCTION

Inexpensive data storage and the ever-growing flow of data from a variety of sources increase the amount of data available to organizations. Competing in the era of big data will require analytically-focused employees with the specialized knowledge and skills to extract useful information from this data. Some have expressed great concern that the demand for employees with this skill set will far outstrip supply (see, e.g., Davenport and Patil, 2012). A widely cited report by McKinsey and Company concluded, "The United States alone faces a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyze big data and make decisions based on their findings" (Manyika et al., 2011, p. 3).

Universities are responding to this call to educate the next generation of entry level data savvy professionals. There are a growing number of degree programs, specializations, and certificates in data science and data analytics at both the graduate and undergraduate levels (Davenport and Patil, 2012; Dumbill et al., 2013). However, there are still relatively few full degree programs at the undergraduate level. A recent review of undergraduate degree programs in data analytics and data science identified thirteen such programs across the United States (Aasheim et al., 2014). Since that time, it is likely that more such programs have been developed. However, as more universities expand into this area, little is currently known about the specifics of skills covered in those degree programs or the extent to which skills coverage is comparable across different programs. This paper fills that gap by examining the course offerings of a small sample of undergraduate data analytics and data science programs to determine what similarities and differences exist across programs. In addition, discrepancies between skills in the literature and those offered in degree programs are identified. This examination will contribute to the goal of identifying important topics for an undergraduate program in data analytics and data science. The focus will be data analytics programs specifically and how they relate to the traditional information systems program.

2. LITERATURE REVIEW

Organizations have collected and analyzed data in an attempt to gain strategic advantage in the market place for many years. However, in recent years, the amount and complexity of available data has exploded, making it more difficult to gain insights from data to improve business decision making. This section presents a review of relevant literature addressing (1) the growth of big data, (2) the evolution of data analytics as a field of study, (3) legal and ethical issues surrounding big data, and (4) implications for academia.

2.1 The Growth of Big Data

A number of factors have contributed to the explosion of data. In the latter part of the 20th century, organizations emphasized integrating transactional databases into data warehouses that could then be analyzed to improve business decisions (Eckerson, 2011). As organizations began to realize benefits from this analysis, this trend accelerated. In one example, Walmart was able to identify bestselling products in hurricane-prone areas when storms were approaching; as a result, prior to storm season, Walmart stores stocked-up not only on obvious high-demand staples such as batteries but also the less obvious number two best-selling item – Pop-Tarts (Preimesberger, 2011).

Growth in e-commerce and social media has contributed to the increase in data accumulation, particularly as organizations utilize clickstream data and social media comments to track customer sentiment and understand consumer behavior. Organizations also collect unstructured data through sources such as bar codes, QR codes, RFID tags, and sensors. United Parcel Service (UPS) installed sensors on more than 46,000 delivery trucks to monitor location, safety, and efficiency related data including speed, direction, and mechanical performance (Davenport, 2013). Similarly, data is collected through devices such as smart phones equipped with GPS and accelerometers that can track and store detailed information about users' whereabouts and movements. These and related technologies have created a rapidly growing set of networked data collection sources (termed the Internet of Things (IoT) (Ashton, 2009)), which will only serve to further accelerate the accumulation of data.

The trend of increasing data collection shows no signs of slowing. EMC estimates the total amount of stored information will increase by 300 times from 2005 to 2020 (Wittman, 2013). However, volume is only one of the defining characteristics of big data. Volume, along velocity and variety, comprise the 3 V's of big data first identified by Laney (2001). The term 'velocity' describes the rate at which data streams in and out, with much of it arriving at or near real-time (Laney, 2001). Additionally, it indicates the need to rapidly extract useful information from data (Kelly, 2014). The term 'variety' describes the plethora of incompatible formats and inconsistent semantics across data arriving from different sources. The combination of these characteristics pushes traditional data management systems to their limits, eventually rendering them ineffective for storing and managing big data. Based on these characteristics, big data can be defined as the "data sets that are so large that they require advanced and unique data storage, management, analysis, and visualization technologies to pinpoint opportunities for improving decision making and adding value" (Aasheim et al., 2014).

2.2 The Evolution of Data Analytics and Data Science

In addition to collecting data for the traditional purposes of record keeping and transactional processing, modern organizations wish to extract useful information from data to improve decision making. However, the techniques for accomplishing this goal have changed considerably. Traditional statistical techniques evolved predominantly during times when acquiring data was costly and time-consuming (Viswanathan, 2014). Therefore, statistics was important for dealing with the 'small data' problem. However, as the cost of collecting data became less prohibitive, and data became a natural derivative of many automated business processes, organizations started to accumulate large data sets. Consequently, organizations started facing a 'big data' problem.

One approach to address the big data problem is through data analytics. While traditional statistical analysis and data analytics are often used interchangeably—and share many similarities—data analytics extends the focus to the analysis of larger data sets gathered from a wide variety of data sources (Davenport, 2013; Viswanathan, 2014). In expanding the scope of traditional statistics, many now consider machine learning and data mining technologies - which are grounded in statistics - as important aspects of data analytics. It follows from the definition of data analytics that 'business analytics' is data analytics, but applied within a business environment to address business issues. Closely related to applying data analytics to address the big data problem has been the rise of the data science field.

'Data science' has been defined as "a set of fundamental principles that support and guide the principled extraction of information and knowledge from data" (Provost and Fawcett, 2013, p. 52). Data science takes a multidisciplinary approach to the big data problem: not only is the data scientist skilled in data analytics, s/he may also be able to develop (code/program) analytical algorithms, systems and applications (Dumbill et al., 2013). In addition, a 'talented' data scientist would normally be expected to possess sound business skills to be able to judge the value of generated insights and to ensure applications address important business problems (Provost and Fawcett, 2013).

2.3 Legal and Ethical Issues Surrounding Big Data

The ability to link information from multiple sources, especially when that information is collected by different organizations, or for different purposes, raises significant legal and ethical questions. Privacy issues, in particular, take on a new dimension. Data brokers collect and summarize large amounts of information from various sources and sell it to their customers, typically for use in marketing efforts (Crawford, 2014). In the course of doing so, they construct detailed and alarmingly intimate profiles of consumers including demographic data, health issues, life events, social media posts, shopping habits, and personal issues such as gun ownership and political leanings (Timberg, 2014). An OfficeMax customer recently received a letter from the company with the words "Daughter Killed in Car Crash" following his name; the company blamed the insensitive wording on a "mailing list rented through a third-party provider" (Crawford, 2014). The government also collects and stores massive amounts of personal data. Former intelligence worker Edward Snowden's public release of U.S. government secret documents surprised many by revealing the extent of digital spying practiced by the National Security Agency on the communications of U.S. citizens and foreign governments (Finkbeiner, 2013).

Critics of the accumulation of data by data brokers and government raise several concerns. One is that consumers are unaware of the breadth and depth of personal data collected and stored about them. They have little or no control over how this data is used and minimal opportunity to review it or correct any errors (Timberg, 2014). Another concern is that accumulated data might be used to discriminate against individuals in the areas of housing, healthcare, financial matters or employment (Wigan and Clarke, 2013). Critics have called for the development of new legal structures and business practices to protect the rights of individuals. These authors point out that many collections of consolidated data may exhibit data quality issues, and can be potentially used to facilitate dataveillance - a term introduced in 1988, which "offers a more economical method for monitoring individuals than physical and electronic surveillance" (Wigan and Clarke, 2013, p. 48). Given such concerns and potential misuse, an awareness of the issues surrounding ethical use of data is likely to be of critical importance for professionals entering this industry.

2.4 Implications of Big Data for Academia

Employees with skills in the area of big data, data science, data analytics, and data-driven decision making are already very important to organizations, and will become even more important in the future. McKinsey and Company (Manyika et al., 2011) predicts a shortage of employees, managers and analysts who can effectively analyze big data for purposes of decision making. Davenport and Patil, describing data scientist as "The Sexiest Job of the 21st Century" (2012, p. 70), have suggested industry's biggest constraint in capitalizing on big data will be a shortage of talent. Accordingly, academia is challenged to develop programs that produce appropriately skilled graduates. In this section, we identify big data skills specifically addressed in the literature.

Davenport and Patil (2012) claim that "more than anything, what data scientists do is make discoveries while swimming in data" (p. 73). This view reaffirms the importance of analytically focused employees who possess solid skills in data management principles and technologies. Topic areas should include traditional relational database management systems (RDBMS), as well as extract, transform, load (ETL) processes, online analytical processing (OLAP), and other emerging data mining technologies. As datasets grow in size and complexity, knowledge of distributed file systems such as Hadoop and Cassandra become critical (Manyika et al., 2011). Capturing and integrating data from multiple sources will also be a challenge for data science and analytics professionals. Technologies such as metadata management solutions and advanced indexing techniques will be similarly important for these professionals (Laney, 2001). In addition, technologies related to data cleansing and preparation are critical. One industry survey found that analytics professionals spend up to 75% of their time retrieving data from multiple sources and cleaning it up for analysis (Balboni et al., 2013).

Programming is also important. According to Davenport and Patil (2012), "Data scientists' most basic, universal skill is the ability to write code" (p. 73). Although analytical tools are becoming more sophisticated, data scientists may still need to develop their own customized solutions to fit specific unique problems or situations (Eckerson, 2011). Skills with statistics, mathematics, and modeling techniques for analytical decision making will also be important in the big data era. Regression analysis, hypothesis testing, clustering, anomaly detection, and predictive modeling are some of some of the fundamental quantitative skills necessary to extract useful information from data (Chen, Chiang, and Storey, 2012). Time series analysis may be used for sales forecasting or for predicting the spread of disease (Manyika et al., 2011). Other analysis techniques including optimization, neural networks, and genetic algorithms may also be employed (Chen, Chiang, and Storey, 2012).

In addition to statistical and technical skills, data scientists will need to cope with ethical, legal and governance issues. Ethical awareness is critical given the highly sensitive nature of some data and the potential for its misuse (Timberg, 2014). Data scientists will need appropriate communication and data visualization skills to help create value from big data. It is also important to have sufficient contextual knowledge of the business to understand the organization's needs and identify areas where big data will be most likely to create value (Dumbill et al., 2013; Kelly, 2014; Provost and Fawcett, 2013). Many analysts and data scientists will work in cross-functional teams which demands strong teamwork and communication skills (Davenport, 2013). Data scientists will require an understanding of business needs to effectively communicate their results through compelling business cases and data visualizations (Kelly, 2014; Van Dyke, 2013). Finally, "the dominant trait among data scientists is an intense curiosity - a desire to go beneath the surface of a problem, find the questions at its heart, and distill them into a very clear set of hypotheses that can be tested" (Davenport and Patil, 2012, p. 73). Beyond the skills identified in the literature, this problem-solving approach and sense of curiosity are important characteristics to cultivate.

The array of relevant skills identified here for knowledge workers in the areas of big data, data analytics, and data science is extensive and diverse. It is unreasonable to expect that all of these skills and technologies could be given full coverage within an undergraduate degree program. The challenge for academics will be to select which of these skills should be covered, and to what extent. Facing budget and time constraints, faculty must carefully consider the extent to which new technology will be integrated into the curriculum. Expertise with the latest data storage and management technologies is helpful in raising the profile of both graduating students and the program that produces those graduates. However, given the rapid rate of change in this industry, faculty must be cautious about chasing the latest technologies. It is important to remember that today's technology may become obsolete tomorrow and that graduating students will face a long career of coping with rapid change. Students must develop a habit of lifelong learning that will enable them to discover and evaluate new technologies and how they should be applied.

The goal of this paper is to determine how choices have been made in existing data analytics and data science programs. In the following sections of this paper, data science and data analytics programs are identified and examined to determine the skills covered in those programs' course offerings. Commonalities and differences across programs are identified.

3. METHODOLOGY

This study uses a direct survey method to obtain a sample of course descriptions for content analysis. Direct survey methods make use of available material that exists online (e.g., web sites) or in printed format (Stefanidis and Fitzgerald, 2014). Advantages of the direct survey method, include being able to focus on a specific program of interest (i.e., undergraduate degree programs in data analytics or data science), allowing collection of data in a systematic way, and facilitating standard quantification of data (Kung, Yang, and Zhang, 2006). This technique has been used to gain understanding of curricula and programs in a variety of disciplines, ranging from information architecture (Zhang et al., 2002) to music education (Mishra et al., 2011). A number of studies have focused on review of Information Systems programs and curriculum (see, e.g.: Bell, Mills, and Fadel, 2013; Kung, Yang, and Zhang, 2006; Lifer, Parsons, and Miller, 2009). Our application of course description content analysis follows in this tradition by using this publicly available data to better understand coverage of big data topics in a variety of programs.

To generate the initial sample for the direct survey, the authors used a list from Aasheim et al. (2014) of 86 universities within the United States that might be expected to offer undergraduate degree programs in the areas of data analytics or data science. After reviewing the web sites of each of the 86 universities, thirteen universities were found to offer full data analytics or data science programs at the undergraduate level (as opposed to those with concentrations or minors) as identified by the name of the program including terms like analytics or data science. The final sample of the thirteen universities used in this study is listed in Table 1, along with the abbreviations that will be used throughout the remainder of this paper to refer to these institutions. In addition to identifying universities with programs in data analytics and data science, Aasheim et al. (2014) explored the similarities and differences between the programs based on an examination of the curriculum requirements for each program. A summary of the findings in Aasheim et al. (2014) can be found in Tables 2 and 3.

University	Program	Abbreviation
Arizona State University	Analytics	ASU
Creighton University	Analytics	CU
Drexel University	Analytics	DU
Miami University	Analytics	MU
Old Dominion University	Analytics	ODU
Rutgers University	Analytics	RU
University of Kentucky	Analytics	UK
University of Tennessee	Analytics	UT
College of Charleston	Data Science	CC
Northern Kentucky	Data Science	NKU
The Ohio State University	Data Science	OSU
University of Rochester	Data Science	UR
University of San Francisco	Data Science	USF

Table 1: Universities with Undergraduate Data Analytics or Data Science Programs and Abbreviations

As shown in Table 2, data analytics programs require at least brief/business calculus (or calculus) and at least one traditional statistics course, with some programs requiring additional statistics courses. All but one requires a database or data warehousing course. Almost all have some course relating to data analytics or modeling and half require a course in data mining. Some programs require courses in visualization or big data. All of the examined data analytics programs are in business schools – with some located in information systems departments and others located in departments that house quantitative methods/statistics programs. Some programs require co-majors or concentration areas. Table 3 shows that the included data science programs are usually housed in computer science departments (rather than business schools), or are interdisciplinary. The majority of interdisciplinary programs require a concentration area.

Characteristics		% of Schools (# of Schools)	Credit Hours
Traditional Course Requirements	Math		
	• Brief Calculus	75% (6)	3
	• Calculus 1	25% (2)	3
	Programming	25% (2)	3
	Statistics	100% (8)	3-6
	Statistics beyond normal requirement	37.5% (3)	5-6
	Database	75% (6)	2-3
	Data Warehousing	12.5% (1)	3
Courses Specific to Analytics or Data Science	Visualization	25% (2)	1.5-3
	Data Mining	50% (4)	3
	Big Data	25% (2)	1.5-3
	Analytics/Modeling	87.5% (7)	3-15
Other Characteristics	In IS program	37.5% (3)	
	In Quant program	25% (2)	
	In B-School	100% (8)	
	Co-Major	25% (2)	
	Concentration Area Required	12.5% (1)	

Table 2: Data analytics program characteristics

Characteristics		% of Schools (# of Schools)	Credit Hours
Traditional Course Requirements	Math		
	• Discrete Math	80% (4)	3
	• Calculus 2	100% (5)	3
	• Linear Algebra	100% (5)	3
	Programming	100% (5)	9-15
	Statistics	100% (5)	6-15
	Database	40% (2)	6
Courses Specific to Analytics or Data Science	Visualization	60% (3)	3
	Data Mining	100% (5)	3
	Big Data	20% (1)	3
	Data Science/Analytics/Modeling	80% (4)	3-9
Other Characteristics	In CS program	40% (2)	
	Interdisciplinary	60% (3)	
	Concentration Area Required	60% (3)	

Table 3: Data science program characteristics

A natural extension of the work done by Aasheim et al. (2014) is to refine the key differences between the programs by examining the content of the courses as opposed to the curriculum requirements and to compare the content to skills and competencies identified in the literature. To determine the content of courses offered in the data analytics or data science area, the authors reviewed the websites of each of the universities listed in Table 1 to obtain more detailed information regarding the content of courses included in those degree programs. Specifically, course descriptions were obtained for courses related to data analytics or data science. Tables 4 and 5 and Figure 1 provide a summary of the findings. The actual course titles can be found in the Appendix, while the full course descriptions can be provided upon request.

Area	# (%) of Universities with Coverage
Visualization	2 (25.0%)
Data Mining	5 (67.5%)
Big Data	2 (25.0%)
Analytics/ Modeling	7 (87.5%)

Table 4: **Summary of Coursework Specific to Data Analytics Programs**

Area	# (%) of Universities with Coverage
Visualization	3 (60.0%)
Data Mining	5 (100.0%)
Big Data	1 (20.0%)
Analytics/ Modeling	4 (80.0%)

Table 5: **Summary for Coursework Specific to Data Science Programs**

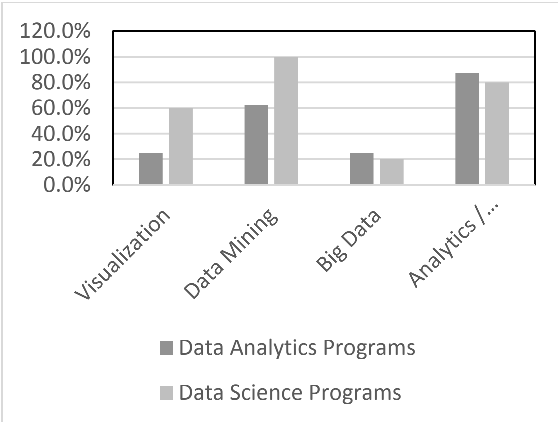

Figure 1: **Percentage of Programs with Coverage of Topic Areas**

The competencies and skills cited in literature were reviewed and are listed below along with a definition and abbreviation for each.

- *Mathematics/statistics/probability (MS)*
  Coverage of mathematics, statistical techniques (including hypothesis testing and regression), and/or probability (Chen, Chiang, and Storey, 2012; Davenport and Patil, 2012; Manyika et al., 2011; Eckerson, 2011)
- *Data mining techniques (DMIN)*
  Coverage of data mining techniques including classification, text and web mining, stream mining, knowledge discovery, anomaly detection, associations, and other techniques (Chen, Chiang, and Storey, 2012; Manyika et al., 2011; Eckerson, 2011)
- *Other modeling/analytics techniques (OMA)*
  Coverage of analytical/quantitative techniques such as machine learning, neural networks, decision trees, linear programming, integer programming, goal programming, queuing, etc. (Manyika et al., 2011; Eckerson, 2011)
- *Visualization techniques (VIS)*
  Coverage of the visual presentation of data (Manyika et al., 2011)
- *Programming skills (PROG)*
  Coverage of computer programming skills (Davenport and Patil, 2012)
- *Data management (DMGT)*
  Coverage of data management topics including SQL, data models, and entity-relationship diagrams (Chen, Chiang, and Storey, 2012; Laney, 2001; Balboni et al., 2013)
- *Understanding of big/ unstructured data (BIG)*
  Coverage of topics related to very large data sets. May include mention of unstructured data, variety, or velocity (Manyika et al., 2011; Laney, 2001).
- *Data capture techniques/technologies (CAP)*
  Coverage of topics related to the storage of data including distributed storage technologies such as Hadoop (Manyika et al., 2011; Balboni et al., 2013; Eckerson, 2011)
- *Data storage techniques/technologies (STOR)*
  Coverage of topics related to data acquisition or collection (Chen, Chiang, and Storey, 2012; Manyika et al., 2011; Provost and Fawcett, 2013; Balboni et al., 2013)
- *Data security (SEC)*
  Coverage of data security issues (Manyika et al., 2011; Balboni et al., 2013)
- *Data preparation/quality (PREP)*
  Coverage of data cleaning, data preparation, and data quality (Balboni et al., 2013)
- *Ethical considerations (ETH)*
  Coverage of topics related to ethical use of data and/or confidentiality (Nash, 2012; Tallon, 2013; Wigan and Clarke, 2013; Poole, 2013; Manyika et al., 2011; Balboni et al., 2013)
- *Data governance policies (GOV)*
  Coverage of data governance issues including decision rights, accountability, and/or use of data (Tallon, 2013; Manyika et al., 2011; Balboni et al., 2013)
- *Decision making skills (DEC)*
  Coverage of tools/techniques for decision-making and/or decision-making process (Lavalle et al., 2011; Manyika et al., 2011)
- *Communication skills (COMM)*
  Coverage of or practice with communication skills; may include presentation of project results or reports (Kelly, 2014)

Two of the authors independently reviewed the course descriptions and assigned the above codes/abbreviations to the skills and competencies found in each. Any differences in the codes initially assigned by the two authors were identified, and each coder then independently reviewed the courses and assigned codes for a second time. Following the second round of coding, the two authors met to discuss the remaining differences. This process was repeated until agreement was reached.

During the coding process, two additional topics emerged. First, the use of case studies was treated as a separate topic as it was mentioned in many course descriptions. The use of case studies has been associated with greater student engagement and the development of higher level cognitive skills and can be an effective method for introducing emerging technologies and concepts (Pridmore, Bradley, and Mehta, 2010). Second, a topic named 'evaluation' was added that represents the mention of higher-order cognitive skills such as determining the applicability of a quantitative method to a given problem in a course description. These two additional topics are shown below.

- *Case studies (CASE)*
  Use of case studies within a course
- *Evaluation (EVAL)*
  Evaluate applicability of tools or techniques for a given problem. May include topics such as alignment of organizational strategy or goals.

4. RESULTS

Table 6 provides the results of the coding and highlight areas of overlap and distinction between the two types of programs. The actual mapping of the course titles to the content/topic area is provided in the Appendix.

Topic	# (%) of Universities with Analytics Programs with Coverage	# (%) of Universities with Data Science Programs with Coverage
Mathematics/ statistics/ probability	6 (75.0%)	5 (100.0%)
Programming	2 (25.0%)	3 (60.0%)
Data management	2 (25.0%)	2 (40.0%)
Visualization techniques	4 (50.0%)	4 (80.0%)
Data mining techniques	5 (62.5%)	5 (100.0%)
Understanding of big/unstructured data	3 (37.5%)	3 (60.0%)
Other modeling/ analytics techniques	7 (87.5%)	4 (80.0%)
Data capture techniques/ technologies	0 (0.0%)	1 (20.0%)
Data governance policies	1 (12.5%)	0 (0.0%)
Data preparation/ quality	1 (12.5%)	5 (100.0%)
Data security	1 (12.5%)	1 (20.0%)
Data storage techniques/ technologies	1 (12.5%)	2 (40.0%)
Decision making skills	4 (50.0%)	0 (0.0%)
Communication skills	2 (25.0%)	1 (20.0%)
Ethical considerations	0 (0.0%)	0 (0.0%)
Case studies	2 (25.0%)	1 (20.0%)
Evaluation	6 (75.0%)	2 (40.0%)

Table 6: Mapping of competencies to programs in data analytics and data science

Observations about the programs in data analytics include:

- All programs include coverage of math and statistics beyond what is required as part of the normal curriculum (as described under traditional course requirements in Table 2);
- Some programs include additional coverage of programming and data management;
- Almost half of the programs cover big data and visualization;
- The majority of programs cover data mining, and all but one cover other modeling/analytics techniques;
- Half of the programs cover decision making skills;
- Some programs include explicit coverage of, or practice with, communication skills, with a few explicitly mentioning the use of case studies;
- Most programs require students to evaluate of the applicability of tools or techniques for a given problem; and
- Programs lack courses with coverage of data capture, and there is limited inclusion of data preparation, data storage, data security, and data governance.

All data science programs include additional coverage of math and statistics beyond what is required as part of the normal curriculum (as described under traditional course requirements in Table 3); most include additional coverage of programming and a few have additional coverage of data management. All programs cover data mining and data preparation, and most programs cover other modeling/analytics techniques, as well as big data and visualization. A few programs require students to evaluate the applicability of tools or techniques for a given problem. Overall, data science programs lack courses with coverage of decision making skills and data governance, and are limited with respect to data capture, data storage, security, communication skills, and the use of case studies.

Perhaps most notably, Table 6 draws attention to fact that there is no explicit mention of the coverage of ethical/legal considerations in any of the courses related to data analytics or data science in the thirteen programs examined, although these topics are cited repeatedly in the literature as an area of importance.

5. KEY FINDINGS

Based upon the analysis of the sampled courses, below is a summary of the key similarities and differences between programs.

5.1 Similarities between Programs

We identified a number of similarities between programs that reflect the expectations identified from our literature review of requisite skills. Reflecting the previously discussed evolution of traditional statistics into data analytics and data science, both types of programs, place a heavy emphasis on math and statistics. For both types of programs, the number of math and stats courses requires is greater than normal in the respective discipline where the program is offered. In addition, they tend to have additional coverage of traditional data management. They also both tend to offer courses in data mining, data visualization, and other modeling and analytics techniques. In this respect, academia appears to be meeting industry expectations. We also identified a potentially troubling similarity between data analytics and data science programs. In the sample of programs we examined, we found little or no coverage of ethics, data governance, or data capture. With increased scrutiny on damaging data breaches and potential misuse of data (Tallon, 2013), these are critical issues for professionals entering the field. There are two possible reasons we did not identify coverage of these topics. One is that the topics are actually included in existing courses, but were not incorporated in course descriptions. Another, of greater potential concern, is that the topics are not covered in the programs included in our sample. Given the nature of the current study, content analysis of course descriptions, we are unable to state conclusively which of these possible explanations applies. However, we note the potential lack of coverage in these areas as a possible area of concern.

5.2 Differences between Programs

Our study also identified differences between the two types of programs. As noted in the Aasheim et al. (2014) study, data science programs differ from data analytics programs in at least two respects with regards to traditional course requirements (see Tables 2 and 3). First, data science programs require additional mathematics courses – at least through linear algebra, which is typically after Calculus II, and most require discrete math. Second, they all require at least nine hours of programming courses and at least two statistics courses.

Based on an examination of Table 6, there are several additional key distinctions that can be made between the two types of programs. Some data analytics programs use case studies in their data analytics courses while data science programs do not. In addition, in data analytics programs there is a greater emphasis on the evaluation of tools and techniques, while data science programs tend to emphasize implementation of the tools and techniques (programming).

Based on an examination of the course descriptions, several additional distinctions can be made in terms of the nature and content of the courses offered in the areas of visualization, data mining and other types of modeling or data analytics courses. One of the two courses in visualization in the data analytics programs heavily emphasizes visualization as a means of effective communication, while the courses in data science tend to emphasize the types of visualizations learned and data used. The data mining courses in the data analytics programs focus on the application of the data mining techniques, while data science programs focus on the algorithms taught in the course. The same observation can be made regarding the data analytics and modeling courses.

Many of the similarities and differences highlighted can be attributed to the location of the programs within the university structure. Data analytics programs have typically arisen in colleges or schools of business and data science programs in other areas such as computer science. The emerging new programs appear to reflect the heritage of the disciplines from which they emerged and to draw on the expertise of the faculty in those disciplines.

6. CONCLUSION, LIMITATIONS AND FUTURE RESEARCH

The purpose of the study was to identify the similarities and differences between data analytics and data science programs. The authors developed a list of skills and competencies identified in the literature and compared the competencies and skills to courses offered in data analytics and data science programs. There were a few gaps between the skills and competencies identified in the course descriptions and those identified in literature. Data analytics programs lack coverage of data capture and ethics as it relates to data and have limited coverage of data preparation, data storage, data security and data governance.

The study is limited in that only course descriptions were examined. By their nature, course descriptions provide a brief, high level overview but do not necessarily identify every topic covered within a course or how those topics are implemented. However, given that the intent of this research was to identify basic differences between data science and data analytics programs in terms of overall content and location within the university, an examination of course titles and descriptions is sufficient for this level of understanding. A more detailed analysis of data science and data analytics curricula could be conducted by analyzing course syllabi. These documents will contain greater detail about both topics covered in a course and pedagogical approaches. Such an analysis is a potential topic for future research.

A natural extension of this study is to use the identified gaps and structure of the programs identified in this paper to make specific curriculum recommendations. An area of future research is to make curriculum recommendations for creating data analytics programs as an extension of current IS curricula and data science programs as an extension of current CS curricula.

Further future research may wish to go beyond this comparison of programs to evaluate the programs' effectiveness. How successful have program been in recruiting and graduating sustainable numbers of students? How successful have those graduates been in finding relevant jobs and how prepared are they for the demands of those jobs? Other potentially fruitful research opportunities may be to directly assess the skills demanded by industry. The world of technology and big data change rapidly. Academic programs will need to be mindful of this rate of change and of current industry expectations in order to successfully prepare graduates to enter this world.

Finally, the authors would like to express a profound appreciation to those faculty and institutions who have taken the lead in developing and offering courses and programs in the areas of data science and data analytics. The authors are grateful that they have provided models for the structure of these degrees that other institutions may adopt and build upon.

7. REFERENCES

Aasheim, C., Williams, S. R., Rutner, P., & Gardiner, A. (2014). Big Data Analytics and Data Science Undergraduate Degree Programs. *2014 Annual Meeting of the Decision Sciences Institute Proceedings (DSI 2014)*, 1-22, Tampa, Florida.

Ashton, K. (2009). That "Internet of Things" Thing. *RFiD Journal*, Retrieved June 20, 2014, from http://www.rfidjournal.com/articles/view?4986

Balboni, F., Finch, G., Reese, C. R., & Shockley, R. (2013). Analytics: A Blueprint for Value [White paper]. Retrieved April 23, 2014, from IBM Global Business Services. http://www-935.ibm.com/services/us/gbs/thoughtleadership/ninelevers/

Bell, C., Mills, R., & Fadel, K. (2013). An Analysis of Undergraduate Information Systems Curricula: Adoption of the IS 2010 Curriculum Guidelines. *Communications of the Association for Information Systems*, 32(1), 73-94.

Chen, H. C., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165-1188.

Crawford, K. (2014). Big Data Stalking. *Scientific American*, 310(4), 14.

Davenport, T. H. (2013). Analytics 3.0. *Harvard Business Review*, 91(12), 64-72.

Davenport, T. H. & Patil, D. J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*, 90(10), 70-76.

Dumbill, E., Liddy, E. D., Stanton, J., Mueller, K., & Farnham, S. (2013). Educating the Next Generation of Data Scientists. *Big Data*, 1(1), 21-27.

Eckerson, W. (2011). Big Data Analytics: Profiling the Use of Analytical Platforms in User Organizations [White Paper]. Retrieved July 23, 2013, from SAS Institute Inc. http://www.sas.com/content/dam/SAS/en_us/doc/research2/big-data-analytics-105425.pdf

Finkbeiner, A. (2013). Researchers Split over NSA Hacking. *Nature*, 502(7470), 152.

Kelly, J. (2014). Big Data: Hadoop, Business Analytics and Beyond. *Wikibon*. Retrieved November 2, 2014, from http://wikibon.org/wiki/v/Big_Data:_Hadoop,_Business_Analytics_and_Beyond

Kung, M., Yang, S. C., & Zhang, Y. (2006). The Changing Information Systems (IS) Curriculum: A Survey of Undergraduate Programs in the United States. *Journal of Education for Business*, 81(6), 291-300.

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety [White Paper]. Retrieved April 23, 2014, from META Group, Inc. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, 52(2), 21-32.

Lifer, J. D., Parsons, K., & Miller, R. E. (2009). A Comparison of Information Systems Programs at AACSB and ACBSP Schools in Relation to IS 2002 Model Curricula. *Journal of Information Systems Education*, 20(4), 469.

Manyika, J., Chui, M., Brown, B., Bughin, J., Richard, D., Roxburgh, C., & Byers, A. H. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity [White Paper]. Retrieved December 31, 2013, from McKinsey Global Institute. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Mishra, J., Day, K., Littles, D., & Vandewalker, E. (2011). A Content Analysis of Introductory Courses in Music Education at NASM-Accredited Colleges and Universities. *Bulletin of the Council for Research in Music Education*, (190), 7-19.

Nash, K. S. (2012). Big Data, Big Brother, Big Bucks. *CIO*. Retrieved March 8, 2013, from: http://www.cio.com/article/706457/Equifax_Eyes_Are_Watching_You_Big_Data_Means_Big_Brother

Poole, S. (2013). The Digital Panopticon. *New Statesman*, 142, 22-25.

Preimesberger, C. (2011). Big Ideas About "Big Data." *eWeek*, 28, 34-37.

Pridmore, J. L., Bradley, R. V., & Mehta, N. (2010). Methods of Instruction and Learning Outcomes: A Theoretical Analysis of Two Approaches in an Introductory Information Technology Course. *Decision Sciences Journal of Innovative Education*, 8(2), 289-311.

Provost, F. & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51-59.

Stefanidis, A. & Fitzgerald, G. (2014). IS Education Research: Review of Methods of Surveying the IS Curriculum to Support Future Development of IS Courses. In *Information Systems: Education, Applications, Research*, Springer, 1-11.

Tallon, P. P. (2013). Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost. *Computer*, 46(6), 32-38.

Timberg, C. (2014). Brokers Use "Billions" of Data Points to Profile Americans. *The Washington Post*. Retrieved July 12, 2015, from: http://www.washingtonpost.com/business/technology/brokers-use-billions-of-data-points-to-profile-americans/2014/05/27/b4207b96-e5b2-11e3-a86b-362fd5443d19_story.html

Van Dyke, M. (2013). Predictive Analytics: Pinpointing How to Best Allocate Patient Resources. *HFM Magazine*. Retrieved April 11, 2014, from: https://www.hfma.org/Content.aspx?id=16069

Viswanathan, V. (2014). *Data Analytics with R: A Hands-on Approach (1e)*. Infivista Inc.

Wigan, M. R. & Clarke, R. (2013). Big Data's Big Unintended Consequences. *Computer*, 46(6), 46-53.

Wittman, A. (2013). Brave the Big Data Wave. *InformationWeek*. Retrieved April 11, 2014, from http://reports.informationweek.com/abstract/81/11935/Business-Intelligence-and-Information-Management/Brave-the-Big-Data-Wave.html

Zhang, X. M., Strand, L., Fisher, N., Kneip, J., & Ayoub, O. (2002). Information Architecture as Reflected in Classrooms. *Asist 2002: Proceedings of the 65th Asist Annual Meeting*, 74-82.

AUTHOR BIOGRAPHIES

Cheryl Aasheim is a Professor of Information Technology at Georgia Southern University. She received her B.S. and M.S.T. degrees in Mathematics from the University of Florida. She received her Ph.D. in Decision and Information Sciences from the University of Florida. Her research interests include I.T. skills research, big data and data analytics, and IT education, curriculum development, and assessment.

Adrian Gardiner is an associate professor of Information Systems at Georgia Southern University. He earned his doctorate in Information Systems from the University of New South Wales, Sydney, Australia in 2004. His research interests are in the areas of Big Data and Social Media Analytics.

Paige S. Rutner is an associate professor of Information Systems at Georgia Southern University. She received her PhD from the University of Arkansas in 2009. Her research interests include issues related to big data and data analytics as well as IS/IT workforce issues including motivation, retention, and work exhaustion.

Susan Williams is Chair and Professor of Information Technology at Georgia Southern University. She received her Ph.D. in Information Systems from Oklahoma State University. She has twice been selected as a Distinguished Professor by the Computing Research Association for Women. Her research interests include IT education, big data and data analytics, and knowledge and skill requirements for entry-level IT workers.

APPENDIX

Area	Univ	Course Names
Visualization	ASU	CIS 415: Big Data Analytics and Visualization in Business
	MU	STA 404/504 Advanced Data Visualization
Data Mining	ASU	CIS 375: Business Data Mining
	DU	MIS 349: Predictive Business Analytics Using Relational Databases
	MU	ISA 491: Introduction to Data Mining in Business
	UK	AN 420G: Data Mining
	UT	STAT 474: Data Mining and Business Analytics
Big Data	ASU	CIS 415: Big Data Analytics and Visualization in Business
	MU	ISA 414: Managing Big Data
Analytics/ Modeling	ASU	CIS 315: Introduction to Business Data Analytics
		CIS 450: Enterprise Analytics
	CU	BIA 499: Practicum in Business Intelligence and Analytics
		BIA 479: Seminar in Decision and IT (Required that Analytics be the topic for BIA)
	DU	BUSN 460: Business Analytics Senior Project
		OPR 320: Linear Models for Decision Making
	ODU	BNAL 406: Spreadsheet Modeling and Analysis for Business Decisions
		BNAL 407: Business Analysis
		BNAL 476: Simulation Modeling and Analysis for Business Systems
	RU	33:136:485: Time Series Modeling
		33:136:400: Business Decision Analytics under Uncertainty
	UK	AN 306: Analytics: Models and Methods
		AN 450G: Analytics: Technologies
	UT	STAT 471: Business Analytics Capstone

Table 7: **Course Titles Specific to Data Analytics Programs**

Area	Univ	Course Names
Visualization	NKU	DSC 321: Data Visualization
	OSU	CSE 5544: Intro to Scientific Visualization OR ISE 5760: Visual Analytics and Sensemaking**
	USF	CS 360: Data Visualization
Data Mining	CC	CSCI 334: Data Mining
	NKU	DSC 411: Data Mining
	OSU	CSE 5243: Introduction to Data Mining
	UR	CSC 297: Introduction to Data Mining
	USF	CS 451: Data Mining
Big Data	NKU	DSC 421: Big Data
Data Science/ Analytics/ Modeling	CC	DATA 101: Introduction to Data Science
		DATA 210: Dataset Organization and Management
		DATA 495: Data Science Capstone
	NKU	DSC 101: Intro to Data Science
		DSC 311: Data Analytics
		DSC 496: Data Science Capstone
	OSU	ISE 3230: Systems Modeling and Optimization for Analytics**
	USF	MATH 345: Mathematical Modeling
**Courses under development or not in current catalog		

Table 8: **Course Titles Specific to Data Science Programs**

Univ	Course Name	MS	DMIN	OMA	PROG	CAP	STOR	DMGT	SEC	BIG	PREP	GOV	DEC	COMM	VIS	ETH	CASE	EVAL
ASU	Big Data Analytics and Visualization in Business		x		x					x					x		x	
	Business Data Mining	x	x	x														
	Big Data Analytics and Visualization in Business		x		x					x					x		x	
	Introduction to Business Data Analytics		x	x						x								
	Enterprise Analytics								x			x						x
CU	Practicum in Business Intelligence and Analytics													x				x
	Seminar in Decision and IT												x					x
DU	Predictive Business Analytics Using Relational Databases	x	x	x				x										
	Business Analytics Senior Project												x					x
	Linear Models for Decision Making	x		x									x					
MU	Advanced Data Visualization	x												x	x			x
	Introduction to Data Mining in Business	x	x	x						x								x
	Managing Big Data		x				x	x		x					x			
ODU	Spreadsheet Modeling and Analysis for Business Decisions	x		x									x					
	Business Analysis			x													x	
	Simulation Modeling and Analysis for Business Systems	x		x														
RU	Time Series Modeling	x		x														
	Business Decision Analytics under Uncertainty	x		x	x													
UK	Data Mining		x							x					x			
	Analytics: Models and Methods.			x						x			x					x
	Analytics: Technologies			x									x					
UT	Data Mining and Business Analytics	x	x	x							x				x			x
	Business Analytics Capstone	x									x							

Table 9: Mapping of competencies to courses in data analytics programs

Univ	Course Name	MS	DMIN	OMA	PROG	CAP	STOR	DMGT	SEC	BIG	PREP	GOV	DEC	COMM	VIS	ETH	CASE	EVAL
CC	Data Mining	x	x	x	x		x				x							
	Introduction to Data Science	x	x	x						x								
	Dataset Organization and Management						x	x	x		x							
	Data Science Capstone		x							x								x
NKU	Data Visualization														x			
	Data Mining	x	x		x													x
	Big Data						x	x		x								
	Intro to Data Science									x								
	Data Analytics			x										x	x			
	Data Science Capstone			x							x			x	x			
OSU	Intro to Scientific Visualization														x			
	Introduction to Data Mining	x	x	x							x							
UR	Introduction to Data Mining	x	x								x				x			
USF	Data Visualization				x	x				x					x			
	Data Mining	x	x	x	x					x	x							
	Mathematical Modeling			x													x	

Table 10: Mapping of competencies to courses in data science programs

STATEMENT OF PEER REVIEW INTEGRITY

All papers published in the Journal of Information Systems Education have undergone rigorous peer review. This includes an initial editor screening and double-blind refereeing by three or more expert referees.

ISSN 1055-3096