Career path for data science professional and data scientist

Mr. Prashant Kumar Koshta

Department of Computer Science & Engineering

Gyan Ganga College of Technology

Jabalpur, India

prashantkumarkoshta@gmail.com

Dr. Vimmi Pandey

Department of Computer Science & Engineering

Gyan Ganga College of Technology

Jabalpur, India

vimmipandey@ggct.co.in

ABSTRACT

Careers in data science typically include a variety of career options. Data scientists collect and analyze data using statistical tools and techniques to derive meaningful information. Data science encompasses principles, processes, and techniques for understanding phenomena through the (automated) analysis of data. If you are considering a career in this field, understanding the career path of a data scientist can help you evaluate the profession. Data Science Career Paths, shows you how to become a data scientist, tells you the average salaries in each country, and describes the basic skills for this profession. A career in data science is exciting, fun, interesting, promising, and rewarding.[1][3] It is important to note that unlike other traditional careers, you do not need to have a degree or specific educational background like other traditional jobs to begin your journey into data science.

Many of the elements of data science were developed in related fields such as machine learning and data mining. In fact, the terms data science, machine learning, and data mining are often used interchangeably. The commonality of these disciplines is that they focus on improving decision making through the analysis of data. Although data science borrows from these other fields, it is broader in scope. Machine learning (ML) focuses on developing and evaluating algorithms to extract patterns from data.[2] Data mining is generally concerned with the analysis of structured data and often includes a focus on commercial applications. Data science takes all of these considerations into account, but also addresses other challenges such as the collection, cleaning, and transformation of unstructured data from social media and the Internet, the use of Big Data technologies to store and process large, unstructured data sets, and issues related to data ethics and regulation.

Keywords—Big data, Data analytics, Visualization, Business intelligence, Machine learning

#  INTRODUCTION

 Data science involves the interpretation of data. People who work in data science can extract, store, sort, and analyze data to obtain valuable information. Most businesses collect data as part of their daily operations. This may include data on consumer behavior, demographics, and purchasing power. This data can fall into the category of structured or unstructured data, depending on how you obtain it. Structured data is precise data that is stored in a specific format, while unstructured data is a group of different types of data that is stored in its original format.[23] A data scientist collects, organizes and interprets the data to give companies insight into the market they interact with. This helps companies make smarter, data-driven decisions.Data scientists obtain and manage data for businesses to help them improve their operations.[4][6] Here are some of the most common tasks of a data scientist:

* Gathering data from a variety of sources, such as business records and market news
* Compiling and organizing data into new formats to facilitate analysis easier
* Developing programs to automate the collection and organization of data
* Extracting valuable information and insights
* Analyzing data to find and record patterns
* Producing detailed reports that can inform management decisions

 Specializing in data science specialization is simply about focusing on one area within data science to focus on and expanding your knowledge and skills in that area until you are a professional. Specializing helps you as a data scientist because it establishes you as a professional and resource for that field, making you highly sought after by companies looking for someone to work in that area of data science. There are many areas of data science that you can specialize in, from data analytics to database management to Big Data to machine learning.

Data science itself is a field that applies scientific principles and advanced analytics techniques to analyze data sets and extract useful information to help develop business strategies and guide important business decisions.[7][9] It spans multiple industries and has evolved as interest in data analytics has grown to the extent that each area of data science has gained prominence in its own right. A company looking for a data scientist may now want someone with more experience in a particular area of data science, such as statistics, machine learning, or database management.[23][24]



**Figure 1. Data Science Career Path**



**Figure 2. Data Scientist**

**Data Scientists Skills Required: Career Path**

 By now, we know from Apple, Google, Twitter, Spotify, Swiggy, and Ola that each is seeking dominance in data science and machine learning. There's no denying that Data Science is one of the fastest growing fields.

However, there is a massive shortage of qualified Data Scientists in the Data Science industry! Even though the number of data science jobs is steadily increasing, there is a noticeable shortage of data scientists with the right skills. Below is a list of skills you ideally need to become a successful data scientist. It is a long and difficult process to acquire all of these skills, but it is not impossible. With time and dedicated practice, you can learn and master them figure1 and figure2.

**Understanding the basic concepts of Data Science**

You can only become a master in this field if you know its roots and fundamentals. Therefore, it's crucial that you understand the basics of the field.

**Statistics**

Machine learning algorithms are developed based on statistics and mathematics. You must have a good basic understanding of statistics and mathematics. You don't need a PhD or Masters in statistics, but a general understanding is a must.

**Programming skills**

To instruct computers to put your analysis into action, you must have exceptional programming skills. You must love computers and their language. The language most commonly used in the industry is Python, so you must be proficient in Python. In addition to Python, you'll also need to learn other languages, such as R, C, C++, shell scripting, and SQL. These languages play a crucial role in your journey as a Data Scientist.

**Data manipulation and analysis**

You must have an experimental mindset that allows you to find and explore different ways to manipulate and make the most of the available data. To do this, you will need to learn various data preprocessing operations. You can start this with SQL, which is an essential prerequisite for your journey into data science.

**Data Visualization**

The adage "a picture is worth a thousand words" is ideal for data science. You need to create effective and meaningful graphs/charts of data that convey the pattern by itself. There are various paid and free tools available in the market for you to choose from. Some examples are PowerBI, Tableau, QlikSense, etc. You can also try open source Python libraries like Matplotlib and Seaborn.

**Machine Learning**

ML is at the heart of Data Science. Therefore, you need to acquire exceptional knowledge of different types of algorithms, how they work on deployed datasets, how to evaluate the effectiveness of algorithms, and finally, which algorithms to use and when.

**Deep Learning**

Deep Learning is an advanced version of machine learning that takes inspiration from the human brain; for complex use cases and data sets. To be a good Data Scientist, you need to learn and understand the complex concepts of Deep Learning.

**Big Data**

The scale and variability of data has changed dramatically since the last decade. Data Scientists are expected to understand the journey of data and how we can effectively manage it for any task
Artificial Intelligence

**Software Development**

Application development skills are very useful when envisioning the end-to-end function of any ML application. You understand how data and operations will progress from one stage to another.

**Model deployment**

Creating an accurate model is only part of the process. You must have skills to put that model into action. You'll need to learn and execute various strategies for deploying your model in real-time production systems.

**Communication Skills**

I rank this skill as THE MOST important because you need to communicate your results and analysis in a simple and effective way to a broader audience that is typically not from a technical or data solutions background.

**Structured thinking**

Data science is a field of experimentation. Therefore, you need to learn and apply a clear and structured thought process to evaluate and test different approaches. If you want to proceed randomly, you'll be lost.

**Curiosity and a willingness to learn**

Because the field of data science is constantly evolving, we see a tremendous amount of progress every day. To stay on the cutting edge, you need to learn and apply new things every day. Therefore, you must be open to learning new things. This is by no means a perfect guide or an exhaustive list of skills you need to become a Data Scientist. However, you can consider this a foundation, and as you learn these skills, you'll encounter and learn new skills.

**Current statistics on job growth in Data Science careers**

If you're here, I'm assuming you've chosen a career path or are in the process of choosing one. Let me draw your attention to some other important factors that may help you in your decision.

In 2022 alone, there were more than 137,000 open data science jobs. Analytical jobs also saw a staggering 47% growth over last year.[1][12]

The statistics above show the growth and demand for data science professionals in various business sectors, geographic locations, and even experience areas. As more and more companies adopt data-based solutions, we'll see a continued upward trend in the demand for data science jobs.

It further distinguishes the role of data scientist from three other roles:

• Data authors: the scientists, educators, students, and others involved in research that produces digital data. They include domain scientists, educators and students who have a vested interest in the research generated from the data

• Data managers: the organizations and data scientists [our emphasis, to denote the potential for confusion] responsible for database operation and maintenance and a reliable and competent partner in data archiving and preservation

• Data users: the larger scientific and education communities, including their representative professional and scientific communities

Having studied data management in practice in the research community in the UK, especially with respect to practice in HEIs as required by the project sponsor, we would not make these distinctions in this way. In our view the roles most clearly distinguish themselves as:

Data creators or data authors: researchers with domain expertise who produce data. These people may have a high level of expertise in handling, manipulating and using data, gained through experience and as a result of need or personal interest.

Data scientists: people who work where the research is carried out – or, in the case of data center personnel, in close collaboration with the creators of the data – and conduct all or a number of the functions described in the NSF’s definition above including, in many cases, being data creators themselves. In origin and training they may be domain experts; computer scientists or information technologists and their career development may have required them to assimilate skills from a discipline from which they did not originate.[14] So, a data scientist in systems biology may be a biologist by origin who has acquired very considerable computing skills and a data scientist whose background is in software engineering may have acquired a considerable degree of biological knowledge. Some data scientists told us that an important part of their role is to be a ‘translator’, communicating the needs of the data creators to data managers (see below) and working with the data managers to ensure that data are stored and accessible in a usable way.[17]

Data managers: people who are computer scientists, information technologists or information scientists and who take responsibility for computing facilities, storage, continuing access and preservation of data. They liaise extremely closely with data scientists, ensuring that the right technological facilities are available for the research group to be able to carry out its work effectively. Some data managers described their role as data ‘plumber’, piping data from one place to another, ensuring data flows operate properly and that valuable data are not lost

Data librarians: originating from the library community, trained and specializing in the curation, preservation and archiving of data. Originally, the term data librarian seemed to be confined to librarians dealing with social science data, but the title now encompasses people with data skills in all disciplines.[4] It is a particularly important area as institutions begin to develop digital repositories for the collection and curation of their research outputs. Datasets are part of those outputs; an institutional repository is a natural home for them and the repository is usually in the care of the library. Whilst ‘big science’ has its (international) data centers and some research councils in the UK provide national data storage facilities, ‘small science’ will need to be provided for by institutions. Even if a third player in the form of a national data service does materialize, there will still be the need for local facilities for data that do not qualify, for one reason or another, for inclusion in that data center, and data librarians will be the custodians of that body of data.

The big data and analytics domain is large and a frame work can be helpful in evaluating specific curricular needs[5]. Recent research by Kang and his colleagues (2015) identified four pillar so analytics and suggested related skills for each pillar(See Figure 3).The four pillars include:

1)data preprocessing, storage, and retrieval; 2) data exploration; 3) analytical models and algorithms, and 4) data product. Other research has also identified a variety of areas related to big data (i.e. 19 big data content considerations and 10 big data skill areas) (Columbus, 2014; Gefen et al.,2011).Drawing on this frame work, we identify three objectives and five research questions for this study[[4][6][8].



Analytical Models & Algorithms

Data Exploration

Data

Preprocessing, Storage &Retrieval

Data Product

## **Figure3. Skills Required by Pillars of Analytics(Kang,Holden, andYu,2015**)

Objective 1: Using the pillars of analytics as a framework, what changes have occurred in Big Data/analytics curricula between 2011 and 2016?

Question 1: What percentage of programs added additional Big Data/analytics courses between 2011 and 2016?

Question 2: What are the most common analytics offerings?

Objective 2: Does the department's course offerings in 2011 affect the changes in analytics course offerings in 2016?

Question 3: What is the impact of advanced database course offerings in 2011 on analytics course offerings in 2016?

Question 4: What impact does adherence to the 2010 Model Curriculum recommendations have on analytics course offerings in 2016?

Objective 3: Do departmental resources impact changes in analytics courses offered in 2016?

Question 5: What is the impact of tuition on analytics courses offered in 2016?

# LITERATURE REVIEW

**2.1 Big Data/Analytics Research in the Curriculum from IS**

In 2011, an ICIS panel examined the curriculum MIS and found a mismatch between what is taught in universities and the needs of industry. Specifically, the panel called for additional emphasis and courses on business analytics, data mining, SQL, and Big Data (Gefen et al., 2011). SQL was likely mentioned because Big Data and analytics from a data-centric approach have their roots in the database field (Chen, Chiang, and Storey, 2012)[11].

Universities have attempted to fill the gaps in the field[3][8]. IS Groups have particularly responded to the opportunity to offer academic programs specializing in data and business analytics to train data scientists. Such programs are spreading rapidly (Goes, 2014, p. iii).

|  |  |
| --- | --- |
| **PillarsofAnalytics** | **Skills** |
| Data Preprocessing, Storage, and Retrieval | No SQL, Data Modeling, Data Warehousing & Distribution /Parallel Computing |
| Data Exploration | Statistical Analysis &Visualization |
| Analytical Models &Algorithms | Machine Learning/Data Mining, Natural Language Processing, Information Retrieval |
| DataProduct | Data and Information Organization, Knowledge Representation &Application Development |

## **Table 1. Skills Required by Pillars of Analytics (Kang,Holden, andYu,2015)**

A third Big Data/analytics approach is offered by Anderson and colleagues (Anderson et al., 2014). They focus on a comprehensive program for undergraduates in predictive analytics, machine learning, and data mining, noting that Big Data and analytics education at the undergraduate level has been feasible for over 10 years (Anderson et al., 2014). Although there are similarities with Kang, Holden, and Yu's (2015) four pillars, Anderson et al.'s (2014) design addresses additional topics outside of the traditional Big Data domain by including topics such as ethics and corporate communications.

Table 2 summarizes the preceding discussion and the recommended skills required for courses in Big Data and analytics. While the 2010 IS curriculum guidelines do not specifically address course offerings in analytics, aside from database/SQL as part of the core, the ability to offer career options that allow students to focus on a specific area such as data analytics is highlighted (Topi et al., 2010)[10].

"As a community of scientists, we would be remiss if we did not take full advantage of the scientific opportunities created by the availability of Big Data, sophisticated analytic tools, and powerful computing infrastructures" (Agarwal and Dhar, 2014, p. 447).

|  |  |  |
| --- | --- | --- |
| **Research** | **#ofAreas** | **Areas/Pillars/Topics** |
| Andersonet al.,2014 | Eight Areas for Big Data/Analytics | 1) Large data sets: Create/design, access, clean, analyze, aggregate, organize, visualize; 2) Database: design, storage, query, modeling; 3) AI techniques: genetic algorithms, neural networks, machine learning, pattern matching; 4) Software and algorithms: Design, programming, testing; 5) Information retrieval: information theory, data mining, text mining; 6) Mathematics: logic and counting, discrete structures, statistics, modeling and simulation; 7) Oral and written communication; 8) Social, ethical and legal issues: privacy and security |
| Chiang, Goes, and Stohr,2012 | Three Areas for Big Data Curriculum | 1) Analytical skills (e.g., data mining, neural networks); 2) IT knowledge and skills (e.g., relational databases, ETL, OLAP, visualization ); and 3) Business skills (e.g., understanding business problems and functional business areas |
| Gupta, Goul, andDinter,2015 | Eighteen Topic Areas(Undergrad &Grad) | 1) Introduction to BI; 2) DBMS; 3)Dimensional modeling; 4) BI infrastructure (e.g., data warehouse); 5) BI infrastructure (e.g., dashboards); 6) Data visualization; 7) Data/text mining; 8) EIS; 9) BI applications; 10) Business justification for BI applications;11) BI management; 12) Strategic use of BI; 13) Data security; 14) Ethical issues in BI; 15) Web-based BI; 16) Future trends; 17) Business performance management; 18) BI and organizational issues (e.g., culture) |
| Kang,Holden,and Yu,2015 | Four Pillars ofAnalytics | 1) Data preprocessing, storage, and retrieval (e.g., NoSQL, data modeling ); 2) Data exploration (e.g., visualization ); 3) Analytical models and algorithms (e.g., machine learning, data mining); 4) Data product (e.g., application development) |
| Andersonet al.,2014 | EightAreasfor BigData/Analytics | 1) Large data sets: Create/design, access, clean, analyze, aggregate, organize, visualize; Databases: design, storage, query, modeling; 3) AI techniques: genetic algorithms, neural networks, machine learning, pattern matching; 4) Software and algorithms: Design, programming, testing; 5) Information retrieval: information theory, data mining, text mining; 6) Mathematics: logic and counting, discrete structures, statistics, modeling and simulation; 7) Oral and written communication; 8) Social, ethical and legal issues: privacy and security |
| Chiang,Goes,andStohr,2012 | Three Areas for Big Data Curriculum | 1) Analytical skills (e.g., data mining, neural networks); 2) IT knowledge and skills (e.g., relational databases, ETL, OLAP, visualisation); and 3) Business skills (e.g., understanding business problems and functional business areas |
| Gupta,Goul, and Dinter,2015 | Eighteen Topic Areas(Undergrad &Grad) | 1) Introduction to BI; 2) DBMS; 3)Dimensional modeling; 4) BI infrastructure (e.g., data warehouse); 5) BI infrastructure (e.g., dashboards); 6) Data visualization; 7) Data/text mining; 8) EIS; 9) BI applications; 10) Business justification for BI applications;11) BI management; 12) Strategic use of BI; 13) Data security; 14) Ethical issues in BI; 15) Web-based BI; 16) Future trends; 17) Business performance management; 18) BI and organizational issues (e.g., culture) |
| Kang, Holden and Yu,2015 | Four Pillars of Analytics | 1) Data preprocessing, storage,and retrieval (e.g., NoSQL, data modelling); 2) Data exploration (e.g., visualisation);3) Analytical models and algorithms (e.g., machine learning, data mining); 4) Data product (e.g., application development) |

**Table 2. Skills Required for Big Data and Analytics**

**2.2 Shortage of Analytics Expertise in Industry**

 With global data growing at 40-50% per year, professionals with Big Data and analytics skills are in high demand (Gordon, 2013; Manyika et al., 2011). Data scientists with degrees in information systems-related fields are also in high demand, and hiring shortages are expected for those with deep Big Data and analytics skills (Manyika et al., 2011)[10]. A survey of 153 professionals from IT found that technological skills such as SQL, computer languages, and web design are critical to the future needs of the industry (Downey, McMurtrey, and Zeltmann, 2008). In addition, the demand for graduates with SQL skills continues to increase as it remains a standard method of accessing Big Data (Soat, 2014). The high salaries and high demand are exerting upward pressure, and average salaries for professionals in Big Data and analytics exceed $100,000(Columbus,2014).

The demand for graduates with expertise in Big Data and analytics goes far beyond MIS. "Big Data is the biggest opportunity for marketing and sales since the Internet went mainstream nearly 20 years ago" (Davis et al., 1997, p. 1). He further argues that companies that use Big Data and analytics effectively are more than five percent more profitable than their competitors. In marketing, for example, databases can help develop a comprehensive picture of customers so that companies can personalize and address their needs (Gordon, 2013).

Database marketing improves profitability, increases sales, improves marketing communications, and improves product development (Duval, 2013). Similarly, it is estimated that analysis of Big Data stored in healthcare could create a potential annual value of more than $300 billion (Manyika et al., 2011). Wall Street investment banks and security firms are also looking for analysts with database skills (Taft, 2012).

Internet- and data-driven companies are driving demand for professionals with skills in predictive modeling and machine learning (Dhar, 2013). The most in- demand Big Data professionals include specific skills in Python programming (96%), Linux (76%), and SQL (76%) (Columbus, 2014).

Table 3 shows a list of other in-demand skills related to Big Data and analytics. Currently, nearly one in 20 careers in the U.S. relates to software development, where employers are looking for individuals with programming languages such as Python and SQL (Gallagher, 2015).

|  |  |
| --- | --- |
| **Skill** | **% Growth inDemand OverPreviousYear** |
| Python | 96 |
| Structured QueryLanguage(SQL) | 76 |
| Linux | 76 |
| Data warehousing | 69 |
| Java | 63 |

**Table 3. Industry Demand Increases for Big Data Professionals (Columbus, 2014) [14]**

**2.3 Research Methodology and Data Collection**

As described in more detail in the next two sections, data for our analyzes were collected over three months in the fall of 2011 and over two months in late 2015 and early 2016. Our sample was a randomly selected group of AACSB programs at universities in the United States. Specific questions were created to answer our research questions and objectives, reviewed by other faculty who are experts in the field of MIS curricula, and revised accordingly [15]. Data to answer these questions came from an examination of college websites, catalogs of course descriptions, and, in some cases, telephone interviews with academic advisors to identify data that did not come from online sources. We describe our methods in detail below[13][14].

**2.3.1 Population and Sample:** The population for this study included undergraduate information systems courses at AACSB-accredited institutions in the United States. Our baseline data came from the same 118 programs (out of approximately 485 AACSB programs) that were randomly selected and used by Bell, Mills, and Fadel (2013) in their 2011 analysis. As reported in their manuscript, the data were collected over three months in fall 2011. Consistent with Yamane's (1967) formula based on a 90% confidence interval, at least 74 of the 485 AACSB programs were required for the sample size to provide sufficient statistical power for the statistical analyzes. Our sample size of 118 exceeds this minimum. One hundred four programs (80%) were public institutions and 25 were private institutions[16]**.**

**2.3.2 Data Collection Procedures:** The survey instrument (see Appendix A) was developed primarily based on the literature review presented earlier in this thesis and focused on the analytic pillars (Kang, Holden, and Yu, 2015), analytic skills (Columbus, 2014), and program clusters (Mills et al., 2012). An initial set of questions to answer our research questions and objectives were prepared by the first author. They were then reviewed by two faculty members with backgrounds in information systems and IS curriculum design to ensure appropriate data were collected. Based on their feedback, minor changes were made indicating that the final set of questions met content validity requirements.

The baseline data for 2011 came from the Bell, Mills, and Fadel (2013) data set, which was initially collected over three months in the fall of 2011. The second data set was collected directly from college websites over two months in fall 2015 and spring 2016. Primary sources were a department's website, which described course offerings, and online course catalogs, which provided information about course content. When important data could not be found on a department's website, we called academic advisors. As described in Bell et al. (Bell, Mills, and Fadel, 2013), two researchers collected the 2011 data set, and an additional researcher examined a random subset of 20 programs to ensure that the data were correctly collected and interpreted[22].

**III. DATA ANALYSIS AND RESULTS**

**3.1 Research Question 1**

What percentage of degree programs have added Big Data/analytics courses? The results of the analysis of the direct survey data show that over 60% of the study programs added at least one new Big Data/Analytics course between 2011 and 2016. Thirty-five percent of programs added one additional course, 15% added two additional courses, about 7% added three courses, and 3% of programs added four courses. Table 4 illustrates the percentage of programs that added a Big Data/Analytics course offering [21].

|  |  |  |
| --- | --- | --- |
| **CourseOfferings** | **Frequency** | **Percent** |
| 0 | 47 | 39.8 |
| 1 | 42 | 35.6 |
| 2 | 18 | 15.3 |
| 3 | 8 | 6.8 |
| 4 | 3 | 3.4 |
| Total | 118 | 100 |

**Table 4. Frequency and Percent of IS Programs adding Analytics Courses 2011-2016**

**3.2 Research Question 2**

What are the most common analytics citations? For this question, course averages for each Big Data/analytics pillar were calculated for all schools in 2011 and 2016. These averages were then compared, representing the movement toward Big Data and analytics over the past four years (see Tables 5-8). The results show that a course in Big Data/Analytics was the most frequently added course, followed by courses in Visualization, Business Data Analytics, and Business Intelligence [9].

|  |  |  |  |
| --- | --- | --- | --- |
| **Pillar1Offerings** | **2011** | **2016** | **%Change** |
| DatabaseManagement | 113 | 114 | 0% |
| Advanced DatabaseManagement | 17 | 19 | 11% |
| OtherDatabase/Administration | 5 | 9 | 88% |

**Table 5. Pillar 1 – Data Preprocessing, Storage, and Retrieval Comparison – Mean Averages of Courses**

|  |  |  |  |
| --- | --- | --- | --- |
| **Pillar2Offerings** | **2011** | **2016** | **%Change** |
| Visualization | 1 | 3 | 300% |
| Business DataAnalysis | 9 | 26 | 289% |
| BusinessIntelligence | 10 | 26 | 260% |

**Table 6. Pillar 2 – Data Exploration Comparison – Mean Averages of Courses**

.

|  |  |  |  |
| --- | --- | --- | --- |
| **Pillar3Offerings** | **2011** | **2016** | **%Change** |
| DataMining | 13 | 22 | 69% |
| DataWarehousing | 7 | 10 | 43% |

**Table 7. Pillar 3 – Models and Data Mining Comparison– Mean Averages of Courses**

|  |  |  |  |
| --- | --- | --- | --- |
| **Pillar4Offerings** | **2011** | **2016** | **%Change** |
| Big Data Analytics | 6 | 34 | 583% |
| Decision Support and Expert Systems | 7 | 9 | 29% |

**Table8.Pillar4–Product Comparison – Mean Averages of Courses**

**3.3 Research Question3**

What is the impact of offering an advanced database course in 2011 on analytics courses offered in 2016? To examine this research question, a one-way between-group analysis of variance (ANOVA) was conducted using the 2011 advanced database course offering as the independent variable and the dependent variable included the number of new analytics courses offered in 2016. The independent variable included a 0 (no advanced database course in 2011) or 1 (advanced database course in 2011). The dependent variable included a range of 0 to 4 new analytics courses in 2016). Assumptions of homo scedasti city were tested using Levene's test for homogeneity of variances (Levene statistic 0.255, df1 1, df2116, Sig.0.614)[12].

There was a significant effect of departments offering advanced database courses in 2011 on offering new analysis courses in 2016 at the <.05level[F(1,116)=6.219,p=0.014](SeeTables 10and11).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **New Analytics Course Offerings** | **N** | **Mean** | **Std.****Deviation** | **Std. Error** | **95%****Confidence Interval for Mean** |
| No(0) | 101 | 0.87 | 0.997 | 0.099 | 0.067 | 1.07 |
| Yes(1) | 17 | 1.53 | 1.068 | 0.259 | 0.98 | 2.08 |
| Total | 118 | 0.97 | 1.029 | 0.095 | 0.78 | 1.15 |

**Table 10. 2011 Advanced Database Impact on 2016 Analytics Course Offerings**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Sum of Squares** | **Df** | **Mean Square** | **F** |
| Between Groups | 6.302 | 1 | 6.302 | 6.219\*\* |
| Within Groups | 117.562 | 116 | 1.013 |  |
| Total | 123.864 | 117 |  |  |
| \*\*P<.05 |

**Table 9.ANOVA for the Regression Equation, Data base on Analytics Offerings**

**IV. DISCUSSION**

Many of the components of data science are not new, but they are becoming increasingly important as organizations agree that analytics is critical to better decision making and thus better performance (Agarwal and Dhar, 2014).It has become apparent that the old components of data science, combined with new data technologies and models, have formed a new, independent discipline - data science[4].

This research provides the first empirical examination of IS programs shifting to Big Data and analytics, and the results confirm that there is a dramatic increase from 2011. Pillar 2 (Data Exploration) and Pillar 4 (Product) saw the largest growth, including courses in visualization, business data analytics, and business data analytics, supporting earlier claims that WIS programs are driving curriculum development in this area (Brandon, 2015). At the same time, nearly 40% of IS programs did not add a course in Big Data or analytics, and 36% added only one course (see Table 4)[10].

One explanation could be that many departments at IS did not have the faculty they needed or were unable to hire new faculty to teach more advanced computer science courses. Another explanation is that departments could not add new courses without dropping existing courses in their IS curricula and were unwilling or unable to make this trade-off. Further research is needed to clarify whether this explanation is correct.

Nevertheless, it is difficult to understate the extent of the shift in MIS toward Big Data and analytics. For example, the number of business intelligence courses increased from 10 in 2011 to 26 in 2016, similar to the number of business data analytics courses (see Table 6)[7][11]. Both courses are part of Pillar 2 (Data Exploration), which includes statistical methods. One explanation for the exceptional growth in the pillar is that HIS faculty and faculty who teach quantitative methods are sometimes in the same department. IS Programs may have found it easier to expand course offerings in the pillar because the faculty members who are proficient in quantitative methods could more easily retool to teach courses on the analysis of economic data.

This research shows that programs with established database offerings, including an advanced database course (Pillar 1 - Data Preprocessing), are significantly more likely to offer Big Data/analytics offerings. This seems reasonable, as the first pillar of data pre-processing logically serves as the foundation for building other data science offerings. Subject matter experts teaching advanced database courses were likely among the first IS academics to recognize the growing importance of data to businesses. At the same time, as companies realized the benefits, the demand for new employees trained in data science grew rapidly.

Analytical Models &Algorithms

-Big Data Analytics

-Decision Support/Expert Systems

Data Exploration

-Data Visualization

-Business Intelligence

-Business Data Analysis

Statistical Methods

Data Preprocessing, Storage, & Retrieval

-Data & Information Management

-Advanced Database Management

-Programming with Python

Data Product

-Data Mining

-Data Warehousing

**Figure 4. Skills Required by Pillar of Analytics based on Current Study Data (Kang, Holden ,and Yu, 2015)**

**V. CONCLUSION**

Data Science is the most in-demand profession of this decade and will continue to be so in the next. As awareness of the field grows, competition between professionals for jobs is at its peak. If you follow this guide and do a true self-assessment, I am sure you will make the right choice to find the best path for you. Just remember that choosing the right career path is only the beginning; your journey starts from there.

## **VI. REFERENCES**

[1] AACSB International. (2013). AACSB Assurance of Learning Standards: An Interpretation. Retrieved from http://www.aacsb.edu/~/media/AACSB/Publications/whit e-papers/wp-assurance-of-learning-standards.ashx.

[2] Agarwal, R. & Dhar, V. (2014). Big Data, Data Science, and Analytics: Opportunity and Challenge for IS Research. Information Systems Research, 25(3), 443-448.

[3] Anderson, P., Bowring, J., McCauley, R., Pothering, G., & Starr, C. (2014). An Undergraduate Degree in Data Science: Curriculum and a Decade of Implementation Experience. ACM Special Interest Group on Computer Science Education (SIGCSE), Atlanta, GA.

[4] Bell, C. C., Mills, R. J., & Fadel, K. (2013). An Analysis of Undergraduate Information Systems Curricula: Adoption of the IS 2010 Curriculum Guidelines. Communications of the Association for Information Systems, 12(4), 73-94.

[5] Brandon, D. (2015). Teaching Data Analytics Across The Computing Curricula. Journal of Computing Sciences in Colleges, 30(5), 6-12.

[6] Breslin, S. (2016). Data Scientists: The Myth and the Reality. Retrieved from http://www.datanami.com/2016/01/22/data-scientists-themyth-and-the-reality/.

[7] Chaudhuri, S., Vivek, D., & Narasayya, V. (2011). An Overview of Business Intelligence Technology. Communications of the ACM, 54(8), 88-98.

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

[9] Chiang, R. H. L., Goes, P., & Stohr, E. A. (2012). Business Intelligence and Analytics Education, and Program Development: A Unique Opportunity for the Information Systems Dicipline. ACM Transactions on Management Information Systems, 3(3), 12.11-12.13.

[10]Columbus, L. (2014). Where Big Data Jobs Will Be In 2015. Forbes. Retrieved from http://www.forbes.com/sites/louiscolumbus/2014/12/29/w here-big-data-jobs-will-be-in-2015/.

[11] Davenport, T. H., Barth, P., & Bean, R. (2012). How 'Big Data' IS Different. Sloan Management Review, 22-24.

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

[13] Davis, G. B., Gorgone, J. T., Couger, J. D., Feinstein, D. L.,& Longenecker, H. E. (1997). IS '97 Model Curriculum and Guidelines for Undergraduate Degree Programs in Information Systems. The DATA BASE for Advances in Information Systems, 26(1), 1-94.

[14] Dhar, V. (2013). Data Science and Prediction.Communications of the ACM, 56(12), 64-74.

[15] Downey, J. P., McMurtrey, M. E., & Zeltmann, S. M. (2008). Mapping the MIS Curriculum Based on Critical Skills of New Graduates: An Empirical Examination of IT Professionals. Journal of Information Systems Education, 19(3), 351-363.

[16] Duval, J. (2013). What is Database Marketing and When Can it Be Used? Customer Think. Retrieved from http://customerthink.com/what\_is\_database\_marketing\_and\_when\_can\_it\_be\_used/.

[17] Gallagher, S. (2015). Software is Eating the Job Market. Retrieved from http://techcrunch.com/2015/06/09/software-is-eating-thejob-market/#.qe7szzq:2jgD.

[18] Gefen, D., Ragowsky, A., McLean, E. R., Markus, M. L., Rivard, S., & Rossi, M. (2011). ICIS 2011 Panel Report: Are We on the Wrong Track and do MIS Curricula Need to be Reengineered? Communications of the Association of Information Systems, 30(11), 161-170.

[19] Glassdoor. (2016). 25 Best Jobs in America. Retrieved from https://www.glassdoor.com/Best-Jobs-in-AmericaLST\_KQ0,20.htm?utm\_medium=email&utm\_source=new sletter&utm\_campaign=January2016\_US&utm\_content= BESTJOBS16US.

[20] Goes, P. B. (2014). Big Data and IS Research. MIS Quarterly, 38(3), iii-viii.

[21] Gordon, J. (2013). Big Data, Analytics and the Future of Marketing and Sales. McKinsey: Digital Advantage.

[22] Gupta, B., Goul, M., & Dinter, B. (2015). Business Intelligence and Big Data in Higher Education: Status of a Multi-Year Model Curriculum Development Effort for Business School Undergraduates, MS Graduates, and MBAs. Communications for the Association of Information Systems, 36, 449-476.

[23] Web Site <https://www.knowledgehut.com/blog/data-science/data-science-skills>

[24] Web Site https://www.quora.com