2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #33404

Content Analysis of Two-year and Four-year Data Science Programs in the United States

Dr. Elizabeth Milonas, New York City College of Technology

Elizabeth Milonas is an Assistant Professor with the Department of Computer Systems at New York City College of Technology -City University of New York (CUNY). She currently teaches relational and non-relational database theory and practice and Data Science courses to undergraduates in the Computer Systems Major. Her research focuses on three key computer areas: Web: research on the mechanisms used to organize big data in search result pages of major search engines, Ethics: techniques for incorporating ethics in computer curriculum specifically in data science curriculum and programs/curricula: evaluating Data Science programs in the US and China.

Dr. Duo Li, Shenyang City University

Dr. Duo Li is the chief research scientist of Green Island Hotel Industry Research Institute of Shenyang City University. Duo Li is the member of ASIST&T and his research interests are focusing on Human-Computer-Interaction, Big Data, Data Analytics, Social Networking, and Hospitality Management.

QUALIFICATIONS: Skilled professional experienced in big data, data analysis, bibliometric, social networking sites, statistic software, and online learning system. Full skilled in establishing, operating, and maintaining online course on Blackboard. Educated in data visualization, multidimensional scaling analysis, and human computer interaction. Well versed in Camtasia, and graphics processing software.

EDUCATION: Doctor of Philosophy in Information Studies, May 2017. LONG ISLAND UNIVERSITY, POST CAMPUS, Brookville, NY

Master of Science, Management Engineering, January 2010. LONG ISLAND UNIVERSITY, POST CAMPUS, Brookville, NY

Bachelor of Science, Automotive Engineering, July 2007. BEIJING INSTITUTE OF TECHNOLOGY, Beijing, P.R. China

Dr. Qiping Zhang, Long Island University

Dr. Qiping Zhang is an Associate Professor in the Palmer School of Library and Information Science at the C.W. Post Campus of Long Island University, where she also serves as director of the Usability Lab. Dr. Zhang holds a Ph.D. and an M.S. in information and library studies from the University of Michigan, Ann Arbor, and an M.S. and a B.S. in cognitive psychology from Peking University in Beijing, China. Prior to joining Long Island University in 2006, she worked at Drexel University, IBM Waterson Research Center, and Institute of Psychology at Chinese Academy of Science.

Dr. Zhang's general research areas are human-computer interaction (HCI), knowledge management (KM), social informatics and distance learning. Her primary interests lie in the areas of computer-supported cooperative work (CSCW) and computer-mediated communication. Specifically, she is interested in facilitating productive collaborations of individuals who are geographically and culturally distributed. Dr. Zhang has published numerous papers in the areas of HCI, CSCW, KM, social informatics and related disciplines.

Content Analysis of Two-year & Four-year Data Science Programs in the United States

1.0 Introduction

Data has grown exponentially in the last decade and this growth has resulted in vast challenges for both business and IT domains[1]. This data, also referred to as "Big Data," is massive and complex and cannot be stored centrally [2]. Big Data growth has accelerated the development of new smart technologies that can support the unique demands of big data. Smart technologies such as MapReduce/Hadoop, Spark, NoSQL, data virtualization, data lake, cloud computing, Artificial Intelligence (AI), Natural Language Processing (NLP) and Machine Learning (ML) have an impact on our daily lives and will continue to be an integral part of our future [3]. They have transformed the way we practice medicine, communicate, process information and make business decisions [1]. The use of smart technologies are evident in many domains including retail, finance, medical, engineering, government, penal, social media and computing [1] , [3]. Together Big Data and new smart technologies have given rise to the Data Science field which has also grown exponentially in the last few years [1] . In 2015, President Barak Obama appointed the first ever Chief Data Scientist [4]. The rise in the Data Science domain has increased the demand for skilled Data Science professionals [4].

The Data Science field has its origins in the statistics and mathematics domain [5] but is now considered a multidisciplinary field [6]. Data Science warrants knowledge of data analytics, programming, systems, applications, informatics, computing, communication, management and sociology [1],[6],[7],[8],[9]. It aims at managing large amounts of complex data and solving big data challenges [10] through implementation of tools, techniques and visualization strategies [7]. Data Scientists collect, prepare, analyze, visualize, manage and preserve large collections of information [3].

To prepare a generation of workers in the skills needed for the Data Science domain, higher educational institutions need to prepare students to support the Big Data movement and the new technologies developed as a result of this movement [3],[4]. A Data Science program is a combination of math, statistics and computing [4]. With such a combination students can develop reasoning, analytical and problem-solving skills [4] needed to gather, process, decipher and present data in a meaningful way. Many universities are already offering Data Science programs [3]. These programs vary widely in terms of core courses and electives with some concentrating more on the statistical and mathematical offerings while others on the computer and programming offerings.

The goal of this research study is to evaluate associate and undergraduate data science programs in the U.S. to identify similarities in course offerings and program structure among the colleges examined. The research questions answered as a result of this study are:

- RQ1: What are the characteristics of Data Science programs?
- RQ2: What are the common core courses and competencies of data science associate and undergraduate programs?

2.0 Literature Review

2.1 Undergraduate Data Science Programs

Several studies have focused on undergraduate data science programs and have made recommendations regarding curriculum plans. One such study [11] reported on the results of a three-week long workshop held by The Park City Math Institute which consisted of 25 undergraduate faculty from a variety of institutions in the US and from the disciplines of mathematics, statistics and computer science. The goal of the workshop was to develop guidelines for undergraduate programs in data science. The directive of the workshop was to identify the activities and skills necessary for a data science program, to identify the concentration areas and the types of courses for the major [11]. The workshop resulted in guidelines which provide some structure for institutions planning for or revising a major in data science. The guidelines identified the importance of integrating at all levels of the Data Science curriculum, mathematical foundations, computational and statistical thinking [11], [12]. The workshop participants felt this combination would provide a rich and effective series of courses and would well prepare graduates for careers in Data Science [11]. The workshop resulted in identifying key skills required for undergraduate Data Science majors to master including computational and statistical thinking, mathematical foundations, model building and assessment, algorithms and software foundation, data curation, and knowledge transference (communication and responsibility)[11]. Six main subject areas of a Data Science Major were identified; data description and curation, mathematical foundations, computational thinking, statistical thinking, data modeling, communication, reproducibility and ethics [11].

In a similar study [13], researchers monitored trends across Europe in order to assess the demands for particular Data Science skills and expertise. They [13] used automated tools for the extraction of Data Science job posts as well as interviews with Data Science practitioners. The goal of the study [13] was to find the best practices for designing Data Science curriculum which include; industry aligned, use of industry standard tools, use of real data, transferable skill set, and concise learning goals. The best practices for delivery of Data Science Curriculum include multimodality, multi-platform, reusable, cutting-edge quality, reflective and quantified, and hands-on. In addition, the researchers [13] indicated that the curriculum design should be from user, research, industry and professional recommendations and feedback. The curriculum should be flexible and adaptable, and should address the whole Data Value Chain.

The results of a study [4]which examined the core Data Science program of 5 major universities showed that courses in calculus, linear algebra, probability and statistics, discrete math, computer programming and data structures help develop the reasoning, problem-solving and analytical skills needed by Data Science professionals. The study [4] also identified that computer related courses such as; programing courses in R or Python, database management, machine learning and data mining, data visualization, should also be included in a Data Science undergraduate program. In addition, study results [4] indicated that a concentration in domain areas such as biology, business, engineering and health science is key in providing domain specific knowledge related to Data Science. Study results [4] also identified the need for capstone courses to provide students with experience in using the skills and tools learned in the Data Science program.

2.2 Undergraduate and Graduate Data Science Programs

In a study by [10], an analysis of Data Science degree programs of 279 private universities in India revealed that 52 universities offer programs in Data Science at the undergraduate and Master's level. These Data Science programs are mainly part of Computer Technology departments and were offered within Engineering and Technology tracks [10]. In a study [3] where 42 undergraduate and Master's Data Science programs were analyzed, it was discovered that most of the programs were offered by joint departments and departments of computer science. The courses most often taught in a Data Science program at the undergraduate level are Probability and Statistics and Data Mining, however, at the Master's level Information Retrieval, Information and Social Network Analysis and Text Mining are most often taught [3]

In a study by [1], 581 programs in Data Science, Data Analytics and other related fields of more than 200 universities around the world were analyzed. The study focused on the United States, on-campus based Ph.D., M.S. and B.S. programs in Data Science although Data Analytics, Data Mining and Big Data programs were also analyzed as sub-areas of Data Science. The researchers used Carnegie Classification, a framework maintained by Indiana University (http://carnegieclassifications.iu.edu) for recognizing and describing institutional diversity in U.S. higher education for the past four and a half decades [1]. The results of this study indicated that Data Science programs are growing [1]. The results showed that there are 107 on-campus Data Science programs; 49 B.S. programs, 45 M.S. programs and 13 Ph.D. programs [1] . The results showed that these programs are housed in various departments including; Computing and Informatics, Math and Computer Science, Math and Statistics, Statistics, Computer Science and Electrical Engineering, Arts and Science, Data Science, Business, Business and Management, Engineering, Independent Data Science Centers and Interdisciplinary [1]. The study concluded that there is no one solution to building a Data Science program at an institution [1] . There are several factors that contribute to planning for a Data Science program including available resources, faculty preparation and enrollment potential [1]. In a research study [9] structured content analysis was used to evaluate 30 Data Science undergraduate and Master's level programs from Arts and Sciences, Business, Computer Science, Engineering, Independent Data Science Centers, iSchools, Mathematical Science and Statistics and Professional Studies disciplines. The study results revealed credit hours in iSchool offered Data Science programs had the highest average at 71 credit hours. Independent Data Science Centers had the lowest number of required credit hours even though they had the most courses listed on their websites [9]. Only one third of the 30 programs analyzed required a capstone course or internship and 20 Data Science programs did not have this requirement [9]. The results [9] indicated that although iSchools displayed the most course offerings, they lacked in elective course offerings related to communication and visualization.

3.0 Method

3.1 Operational Definition

For the purpose of this research paper, an academic program is defined as the sequence of courses leading to an AS, AA, AAS, BA, BS, BA and BS and Certificate.

3.2 Data Sources

Two search terms were used to locate data for this study; Data science programs and undergraduate data science programs. In addition, the following websites were used to crawl the data for this paper.

- Associate Degree Programs:
  - https://www.discoverdatascience.org/programs/associate-in-data-science/
  - https://ryanswanstrom.com/colleges/
  - https://community.amstat.org/blogs/steve-pierson/2017/04/25/data-scienceanalytics-coursesprograms-in-two-year-colleges

- Undergraduate Degree Programs:
  - https://www.mastersindatascience.org/specialties/bachelor-degrees-in-data-science/
  - https://ryanswanstrom.com/colleges/
  - https://www.datasciencedegreeprograms.net/rankings/affordable-bachelors/
  - https://blog.collegevine.com/the-list-of-all-u-s-colleges-with-a-data-science-major/
  - https://www.usnews.com/best-colleges/rankings/computer-science/data-analytics-science

3.3 Datasets

A total of 171 undergraduate Data Science programs were initially crawled, but only 136 programs were evaluated for this study. 35 undergraduate programs were omitted from the analysis due to the following:

- no data science courses were evident in the list of core requirements for the program
- only a graduate degree was offered
- only a Data Science minor was offered. No Data Science major was offered.

Each institution's course offering as listed on the institution's website, department website and in the institution's academic course catalog were examined.

3.4 Program Profiles

The following program profile information was gathered from college or university websites, department webpages, and course catalogs:

1) College or university profile information: institution name, school name, type of institution (private/public)
2) Major and department profile information: major name, department name, degree
3) Credit profile information: accreditation, total number of credits for the degree and the core curriculum.
4) Instruction modality profile information: Online, In-person, Hybrid
5) Geographic profile information: city, state, geographic region

3.4.1 Department/School Clusters

- Some of the colleges examined did not have any department names associated with the Data Science Degree. In these cases, the school name was used.
- The departments or school names were examined and ten (10) clusters were created: 1) Business, 2) Computer Science, 3) Data analytics, 4) Data Science, 5) Data Science and Statistics, 6) Engineering, 7) Information Science and Technology, 8) Interdisciplinary, 9) Mathematics/Computer Science, and 10) Mathematics/Statistics.
- These department or school clusters were generated for analysis purposes as it is more effective to analyze data based on ten department or school clusters as opposed to over a hundred department or school names. The department or school names were mapped to these clusters as shown in Table 1. Department or school names that were the same for more than one college or university are only shown once for the purpose of this table.

Department or School Name	**Cluster Name**
Business	Business
Business Analytics	Business
Business and Information Management	Business
Business and Management	Business
Computer Science	Computer Science
Data Analytics	Data Analytics
Data Management/Data Analytics	Data Analytics
Center of Data Science	Data Science
Data Science	Data Science
Data Science Institute	Data Science
Data Science, Analytics and Visualization	Data Science
Department of Computational & Data Science	Data Science
Goergen Institute for Data Science	Data Science
Department of Statistics and Data Science	Data Science and Statistics
Statistics and Data Science	Data Science and Statistics
Applied Engineering & Sciences	Engineering
Dale E. and Sarah Ann Fowler School of Engineering	Engineering
College of Computing and Technology	Information science and technology
Computer and Information Science	Information science and technology
Computer and Information Systems	Information science and technology
Computer Systems Technology	Information science and technology
Computer Technology	Information science and technology
Computing and Information Science	Information science and technology
Information Studies	Information science and technology
Information Systems and Analytics	Information science and technology
Library and Information Science	Information science and technology
College of Arts & Science	Interdisciplinary
College of Computing and Digital Media	Interdisciplinary
College of Computing and Informatics	Interdisciplinary
College of Engineering and Information Science	Interdisciplinary
College of Liberal Arts and Sciences	Interdisciplinary
College of Science and Technology	Interdisciplinary
Computer Science & Department of Statistics	Interdisciplinary
Computer Science & Information Systems	Interdisciplinary
Computer Science and Engineering	Interdisciplinary
Computer Science and Statistics	Interdisciplinary
Computing, Data Science and Society	Interdisciplinary
Department of analytical, Physical & Social Sciences	Interdisciplinary
Departments of Mathematics & Statistics, Computer Science & Computer Information Systems	Interdisciplinary
Electrical Engineering & Computer Science & Dep. of Statistics	Interdisciplinary
Franklin College of Arts and Sciences	Interdisciplinary
Heavin School of Arts and Sciences	Interdisciplinary
Information Systems and Business Analytics	Interdisciplinary
Interdisciplinary (Business, Math, Science)	Interdisciplinary
Interdisciplinary (computer science & electrical eng.)	Interdisciplinary

Mathematics, Computing and Statistics	Interdisciplinary
Misher College of the Science and Arts	Interdisciplinary
School of Arts, Science & Business	Interdisciplinary
STEM	Interdisciplinary
STEM and Advanced Manufacturing	Interdisciplinary
Computing and Mathematical Sciences	Mathematics/Computer Science
Mathematics and Computer Science	Mathematics/Computer Science
Department of Statistics	Mathematics/Statistics
Mathematical Sciences	Mathematics/Statistics
Mathematical Sciences	Mathematics/Statistics
Mathematics	Mathematics/Statistics
Mathematics & Statistics	Mathematics/Statistics
Mathematics and Computational Science	Mathematics/Statistics
Mathematics and Statistics	Mathematics/Statistics
Statistics	Mathematics/Statistics
Statistics and Actuarial Sciences	Mathematics/Statistics

Table 1: Department or School Cluster mapping

3.4.2 Major Clusters

- The majors were examined and seven (7) clusters were created: 1) Big Data, 2) Business Analytics, 3) Computer Science, 4) Data Analytics, 5) Data Science, 6) Information Science and Technology and 7) Math/Statistics.
- These major clusters were generated for analysis purposes as it is more effective to analyze data based on seven major clusters as opposed to over a hundred major names.
- The major names were mapped to these clusters as shown in Table 2. Major names that were the same for more than one college or university Data Science program are only shown once for the purpose of this table.

Major name	**Cluster name**
Big Data and Analytics	Big Data
Big Data and Business Analytics	Big Data
Business Analytics	Business Analytics
Business Intelligence Specialist	Business Analytics
Computational and Data Science	Computer Science
Computational Modeling and Data Analytics	Computer Science
Computer and Data Science	Computer Science
Data Analytics	Data Analytics
Data Management and Data Analytics	Data Analytics
Data Science	Data Science
Data Science and Analytics	Data Science
Data Science and Applied Mathematics	Data Science
Data Science and informatics	Data Science
Data Science for Creative Industries	Data Science
Data Science with Computational analytics	Data Science
Data Science with Computational and Modeling	Data Science
Informatics	Information Science and Technology
Information and Data Sciences	Information Science and Technology
Information Science	Information Science and Technology
Information Studies	Information Science and Technology
Applied Data and Information Science	Information Science and Technology
Mathematics and Data Science	Math/Statistics
Statistics and Computer Science	Math/Statistics
Statistics and Data Science	Math/Statistics
Statistics and Machine Learning	Math/Statistics
Applied Statistics	Math/Statistics

Table 2: Major Clusters mapping

3.4.3 U.S. Geographical Clusters

- Geographic regions of the United States were based on National Geographic United States Reginal Map https://media.nationalgeographic.org/assets/file/us-regions-map.pdf
- Five regions are identified in the map: 1) Northeast, 2) Midwest, 3) West, 4) Southwest, and 5) Southeast. Table 3 below displays the states in each region.

Region	State abbr.	State name		Region	State abbr.	State name
Midwest	IA	Iowa		Southeast	AL	Alabama
	IL	Illinois			AR	Arkansas
	IN	Indiana			DC	District of Columbia
	KS	Kansas			DE	Delaware
	MI	Michigan			FL	Florida
	MN	Minnesota			GA	Georgia
	MO	Missouri			KY	Kentucky
	ND	North Dakota			LA	Louisiana
	NE	Nebraska			MD	Maryland
	OH	Ohio			MS	Mississippi
	SD	South Dakota			NC	North Carolina
	WI	Wisconsin			SC	South Carolina
Northeast	CT	Connecticut			TN	Tennessee
	MA	Massachusetts			VA	Virginia
	ME	Maine			WV	West Virginia
	NH	New Hampshire		Southwest	AZ	Arizona
	NJ	New Jersey			NM	New Mexico
	NY	New York			OK	Oklahoma
	PA	Pennsylvania			TX	Texas
	RI	Rhode Island				
	VT	Vermont				
West	AK	Alaska				
	CA	California				
	CO	Colorado				
	HI	Hawaii				
	ID	Idaho				
	MT	Montana				
	NV	Nevada				
	OR	Oregon				
	UT	Utah				
	WA	Washington				
	WY	Wyoming				

Table 3: United States Geographical Regions

3.5 Data Science Competencies

- 2021 ACM Report Competencies

  - The Data Science competencies used in this research are based on the findings of the Association for Computing Machinery (ACM) Data Science Task Force. The Association for Computing Machinery (ACM) is a society concerned with the education and research.
  - The ACM Data Science Task Force were tasked in 2017 with providing guidance on data science computing-specific competencies for undergraduate programs [17]. In 2018, the Task Force designed two surveys to gather information from academia and industry on the key Data Science computing competencies. In 2019, they presented these competencies at conferences and meetings and gathered the comments and suggestions from the data science community [17]. Throughout 2019 and 2020, the Task Force continued to solicit comments and suggestions from the data science community and in 2021 the revised list of competencies was published in the January 2021 Computing Competencies for Undergraduate Data Science Curricula Report [17].
  - Table 4 identifies the eleven Data Science competencies identified in the 2021 ACM report [17]:

ACM Data Science Task Force Report Competencies	
1. Analysis and Presentation ● Foundational considerations ● Visualization ● User-centered design ● Interaction design ● Interface design and development	**7. Data Privacy, Security, Integrity, and Analysis for Security** ● Data privacy ● Data security ● Data integrity ● Analysis for security
2. Artificial Intelligence ● General ● Knowledge representation and reasoning – logic based ● Knowledge representation and reasoning – probability based ● Planning and search strategies	**8. Machine learning** ● General ● Supervised learning ● Unsupervised learning ● Mixed methods ● Deep learning
3. Big Data Systems ● Problems of scale ● Big data computing architectures ● Parallel computing frameworks ● Distributed data storage ● Parallel programming ● Techniques for Big Data applications ● Cloud computing ● Complexity theory ● Software support for Big Data applications	**9. Programming, data structures and algorithms** ● Algorithmic thinking and problem solving ● Programming ● Data structures ● Algorithms ● Basic complexity analysis ● Numerical computing
4. Computing and Computer Fundamentals ● Basic computer architecture ● Storage systems fundamentals ● Operating system basics ● File systems ● Networks ● The web and web programming ● Compilers and interpreters	**10. Software development and maintenance** ● Software design and development ● Software testing
5. Data Acquisition, Management, and Governance ● Data acquisition ● Information extraction ● Working with various types of data ● Data integration	**11. Professionalism** ● Continuing professional development ● Communication ● Teamwork ● Economic considerations

● Data reduction and compression ● Data transformation ● Data cleaning ● Data privacy and security	● Privacy and confidentiality ● Ethical considerations ● Legal considerations ● Intellectual property ● On automation
6. Data Mining ● Proximity measurement ● Data preparation ● Information extraction ● Cluster analysis ● Classification and regression ● Pattern mining ● Outlier detection ● Time series data ● Mining web data ● Information retrieval	

Table 4: Sub-topic for each competency identified in the ACM Data Science Task Force report

- Cluster of Competencies

  - For the purposes of this research paper, the eleven (11) ACM Task Force Report Competencies [17] were clustered into seven (7) competencies as follows: 1) Computing fundamentals, 2) Data management, governance and privacy, 3) Data visualization, 4) Machine learning, 5) Data mining, Big data, 6) Data Science in context, and 7) Math and Statistics.
  - Table 5 presents the competencies used in this research, the corresponding competencies presented in the 2021 ACM Task Force Report[17], and sample courses for each competency.
  - This bottom up approach was used after a careful evaluation of the curriculum requirements for the Data Science programs in the 136 colleges examined. It was determined that several of the competencies presented in the 2021 ACM Task Force report [17] should be combined to align more directly with current Data Science degree programs.
  - Table 5 below maps the seven competencies used in this research to the 2021 ACM Data Science Task Force report [17] competencies.

Competencies	**ACM Data Science Task Force Report Competencies**	**Sample Courses**
Computing Fundamentals	4, 9, 10	SQL Programming, Introduction to Programming, Algorithms, Data Structures, Object Oriented Programming, Software Engineering, Systems Analysis and Design, Human-Computer Interaction
Data Management, Governance, Privacy	5, 7, 11	Data Warehousing, SQL, Databases, Security, Fraud Detection, Network Security, Ethics
Data Visualization	1	Data Visualization
Machine Learning	2, 8	Machine Learning, Data Modeling, Artificial Intelligence, Deep Learning
Data Mining, Big Data	3, 6	Data mining, Data modeling, systems analysis, Big Data, Data munging
Data Science in Context	11	Capstone, Internship, Senior Project, courses in disciplines such as physics, biology, chemistry, the humanities, or other areas

Math and Statistics		Calculus, discrete structures, probability theory, elementary statistics, advanced topics in statistics, and linear algebra.

Table 5: Competencies, their equivalent ACM Task Force Report competencies, and sample courses

3.6 Data Analysis

Content analysis

Content analysis of the Data Science programs in the United States was conducted to provide a comprehensive understanding of the characteristics of these programs including the competencies. This comprehensive knowledge can be used to inform discourse related to creating new Data Science programs as well as updating current programs. As part of the program profile analysis, department/school and major clusters were initially coded by two researchers. The evaluations were compared and Cohen's Kappa value was calculated to measure inter-coder reliability [16]. The Cohen's Kappa value for the department/school name clusters was .82 and for the major clusters was .76. The Pivot table feature in Excel was used to analyze the data and the Excel graph feature was used to visualize the results.

The program profile analysis includes; 1) College or university profile information: institution name, school name, type of institution (private/public), 2) Major and department profile information: major name, department name, degree, 3) Credit profile information: accreditation, total number of credits for the degree, total number of credits for the core, 4) Instruction modality profile information: Online, In-person, Hybrid, and 5) Geographic profile information: city, state, geographic region (RQ1). The study also presents a comparative analysis of the competencies identified by the Association of Computing Machinery (ACM) in the 2021 Computing Competencies for Undergraduate Data Science Curricula Report [17] and those competencies identified in the 136 Data Science programs evaluated as part of the study (RQ2).

4.0 Results

The purpose of this study is to conduct a content analysis of 136 two-year and four-year Data Science programs in order to acquire a deeper understanding of the undergraduate data science programs in the U.S.

4.1 Results Related to RQ1 - Program Profile

In following section, results related to the program profile were examined and analyzed.

4.1.1 College or University Profile information

Types of College and Majors

Analysis of the majors and the types of colleges (Private and Public) determined that data science curriculum is evident in seven majors; data science, data analytics, math and statistics, big data, computer science, business analytics and information science and technology. Out of these majors, the Big Data major is only available in private colleges. Analysis also showed that the most prevalent majors in both Private and Public colleges are as follows: Data Science, Data

Analytics, Math and Statistics and Information Science and Technology. Computer Science is the least prevalent major in both Private and Public colleges. Table 6 and Figure 1 below show the analysis of majors and types of colleges.

	Private	Public
Data Science	55	41
Data Analytics	11	8
Math and Statistics	5	4
Information Science and Technology	2	3
Computer Science	1	2
Big Data	2	0
Business Analytics	0	2

Table 6: Types of colleges and majors

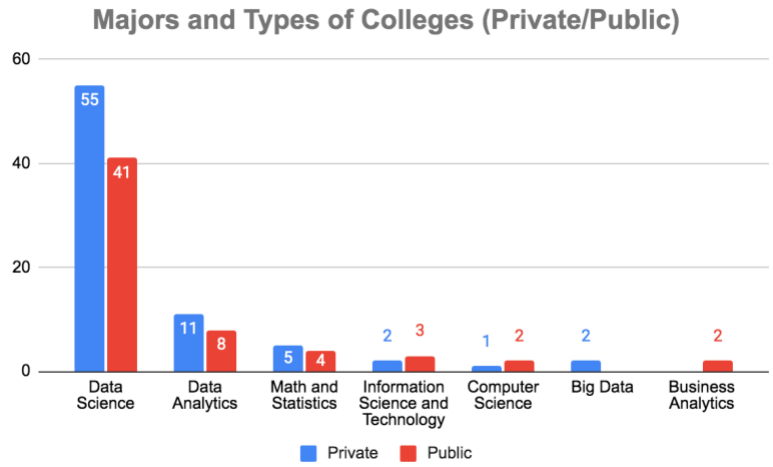

Figure 1: Types of colleges and majors

Types of college and degrees

Analysis showed that the most prevalent data science degree in both private and public colleges is the BS degree. The AS degrees is only available in public colleges and the BA degree is only available in private colleges. Table 7 and Figure 2 below show the analysis of the degrees and the type of colleges.

	Private	Public
AA	0	1
AS	0	4
AAS	1	1
BA	6	0
BS	67	52
BS and BA	2	1
Certificate	0	1

Table 7: Degrees and type of colleges

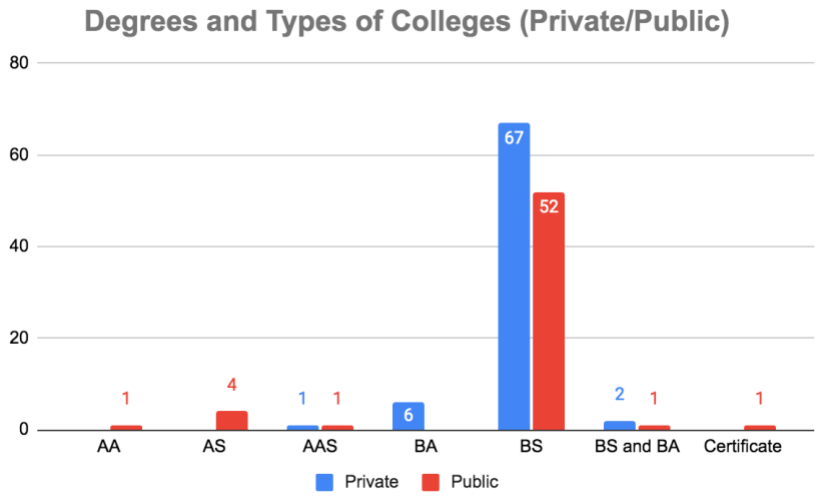

Figure 2: Degrees and type of colleges

4.1.2 Major and Department Profile Information

Majors and degrees

Analysis of the majors and the degrees associated with each of the majors showed that the BS degree is the most prominent degree in all majors with the exception of the Business Analytics major. The Business Analytics major is associated with the AS and the AAS degrees only. The analysis also showed that the Data Science major has the widest variety of degrees: AA, AAS, AS, BA, BS, BA and BS and certificate. In addition the Big Data, Computer Science and Math and Statistics majors are only associated with BS degrees and the associate degrees are associated only with data science, business analytics and data analytics degrees. The dominant majors are data science, data analytics and math and statistics. Only a small portion of degrees are in the Information Science and technology, computer science, business analytics and big data. There are almost no certificate programs in data science. Table 8 and Figure 3 below show the analysis of majors and types of degrees.

	Data Science	Data Analytics	Math and Statistics	Information Science and Technology	Computer Science	Big Data	Business
AA	1	0	0	0	0	0	0
AS	1	2	0	0	0	0	1
AAS	1	0	0	0	0	0	1
BA	4	1	0	1	0	0	0
BS	85	16	9	4	3	2	0
BA and BS	3	0	0	0	0	0	0
Certificate	1	0	0	0	0	0	0

Table 8: Majors and types of degrees

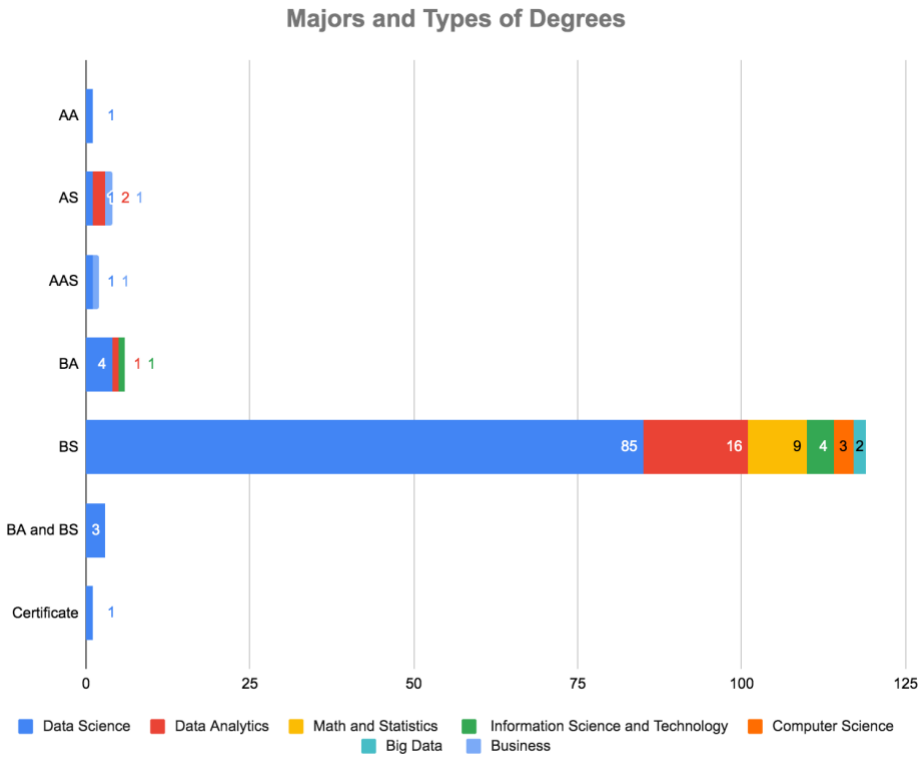

Figure 3: Majors and types of degrees

Departments and majors

Analysis of the departments and the majors associated with each of the departments showed that the data science major is the key major in departments: Computer Science, Data Science, Information Science and Technology, Interdisciplinary and Mathematics and Statistics. The data analytics major is found in computer science, data analytics, engineering, interdisciplinary, mathematics/computer science and mathematics/statistics departments. The Business and Engineering departments have an even distribution of data science, data analytics, big data and business analytics. The analysis showed that Computer Science departments do not have computer science majors. The Computer Science majors are found in the Data Science and Math and Statistics departments. Table 9 and Figure 4 below show the analysis of departments and majors.

	Business	Computer Science	Data Analytics	Data Science	Data Science and Statistics	Engineering	Information Science and Technology	Interdisciplinary	Mathematics/ Computer Science	Mathematics/ Statistics
Data Science	1	15	1	24		1	8	21	1	24
Data Analytics	1	2	3			1		6	3	3
Math and Statistics		1			3			1		4
Information Science and Technology							2	1	2	
Computer Science				1						2
Big Data	1							1		
Business Analytics	1					1				

Table 9: Departments and majors

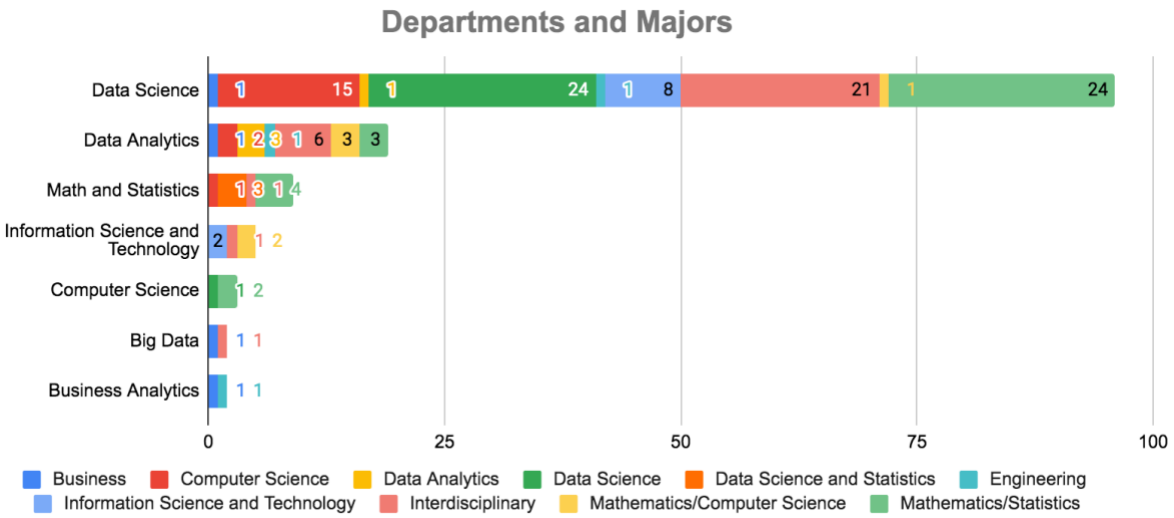

Figure 4: Departments and majors

4.1.3 Credit Profile Information

Analysis of the core credits and the majors showed that the widest range of core credits is found in the data science major where the range of core credits spans from 16 to 120. The second widest range of core credits is found in the data analytics major where the range of core credits spans from 30 to 123 credits. The third widest range of core credits is found in the math and statistics major where the range of core credits spans from 30 to 79 credits. Based on the data analyzed the most consistent range of core credits can be found in the business analytics and computer science majors. Information science and technology and big data have large jumps in the core credits requirements with information science and technology core credits at 16, 39, 40, 42, 93 and big data core credits at 27 and 48 credits. Table 10 and Figure 5 below show the analysis of the credit profile for the majors.

	15-30	31-40	41-50	51-70	71-90	91-123
Data Science			3	35	13	2
Data Analytics	1	6	5	6		1
Math and Statistics	1		3	4	1	
Information Science and Technology	1	2	1			1
Computer Science		1	2			
Big Data	1		1			
Business Analytics			2			

Table 10: Majors and core credits

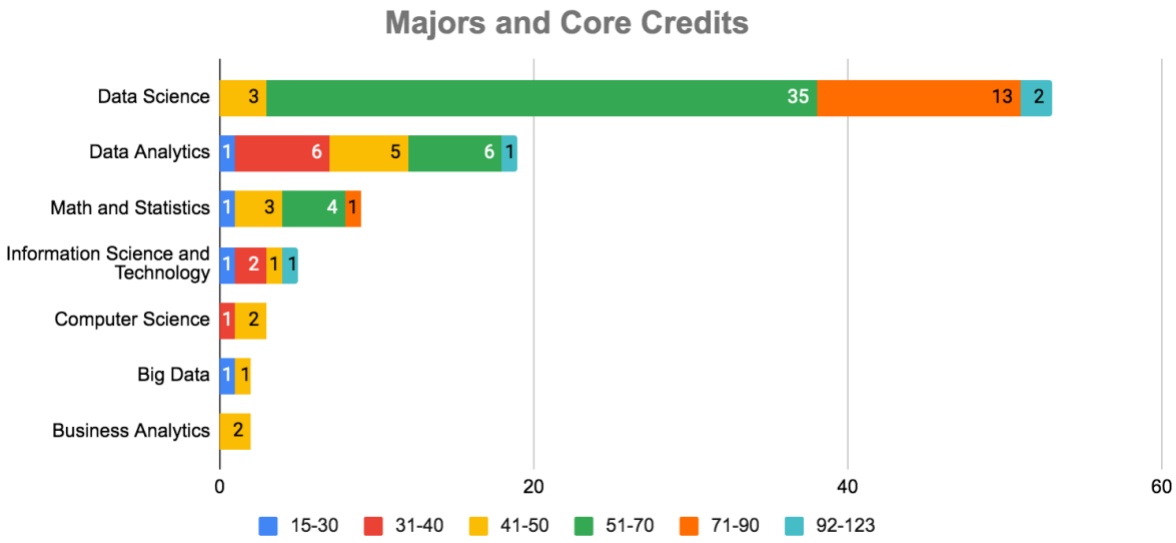

Figure 5: Majors and core credits

4.1.4 Instruction Modality Profile Information

Mode of instruction and major

Analysis of the mode of instruction showed that only two majors have online instruction: big data and data analytics. Blended and in-person instruction are nearly even in numbers with 63 colleges offering blended instruction and 71 colleges offering in-person instruction. The data science major has nearly even numbers in colleges offering blended and in-person instruction with 49 colleges offering the data science major as blended and 47 colleges offering the data science major as in-person. Table 11 and Figure 6 below show the analysis of the mode of instruction and the major.

	Blended	In-Person	Online
Data Science	49	47	0
Data Analytics	7	11	1
Math and Statistics	3	6	0
Information Science and Technology	2	3	0
Computer Scienece	1	2	0
Big Data	0	1	1
Business Analytics	1	1	0

Table 11: Majors and mode of instruction

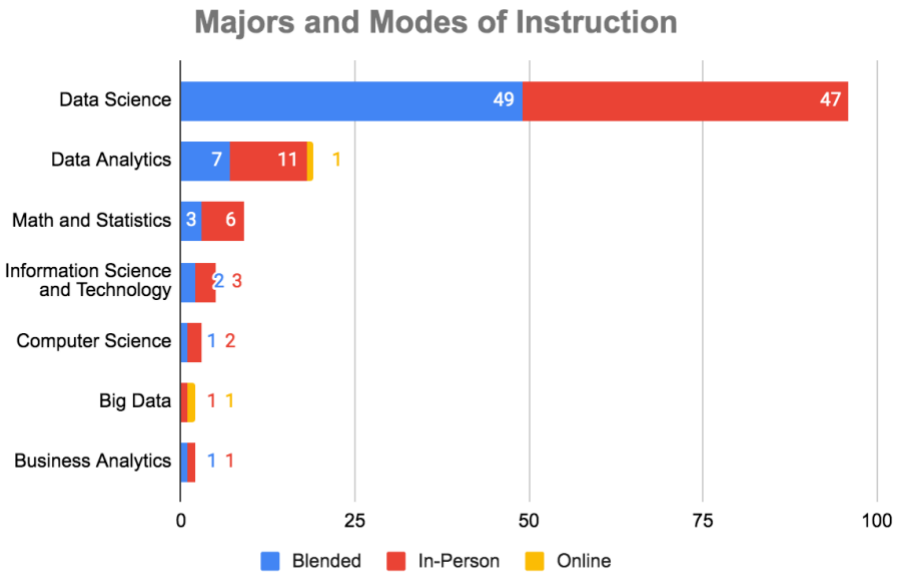

Figure 6: Majors and mode of instruction

Mode of instruction and type of college/university

Analysis shows that private colleges conduct more in-person instruction for the data science programs than public colleges. In addition, the analysis shows that online instruction in data science programs is offered only by private colleges. Table 12 and Figure 7 below show the analysis of the mode of instruction and the type of college/university.

	Blended	In-Person	Online
Private	29	45	2
Public	34	26	0

Table 12: Types of college and mode of instruction

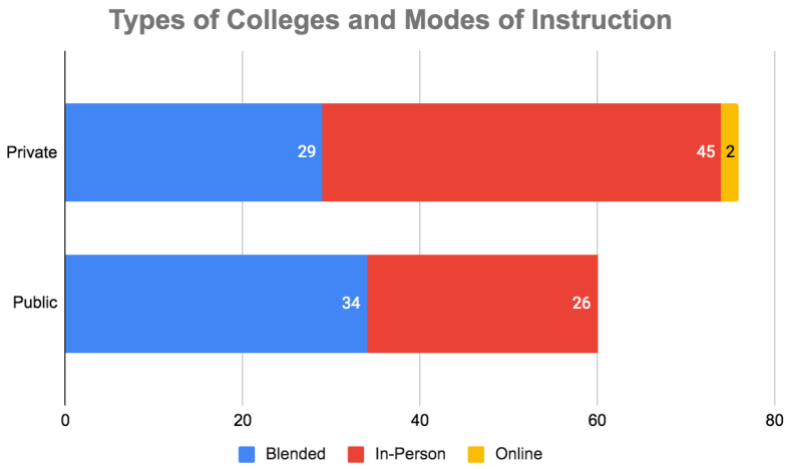

Figure 7: Types of college and mode of instruction

4.1.5 Geographical Profile Information

Regions and Majors

Analysis showed that the Northeast region has the most majors: Data Science, Data Analytics, Math and Statistics, Big Data, Computer Science and Information Science and Technology. Analysis also showed that the Western region has the least number of majors: Data Science, Data Analytics and Information Science and Technology. The Data Science major is the most prevalent major in all the five regions. Table 13 and Figure 8 below show the results of the analysis of majors and regions.

	Northeast	Midwest	West	Southwest	Southeast
Data Science	41	28	14	4	9
Data Analytics	6	6	6	1	0
Math and Statistics	4	4	0	1	0
Information Science and Technology	1	1	1	1	1
Computer Science	1	0	0	0	2
Big Data	2	0	0	0	0
Business Analytics	0	0	0	0	2

Table 13: Majors and regions

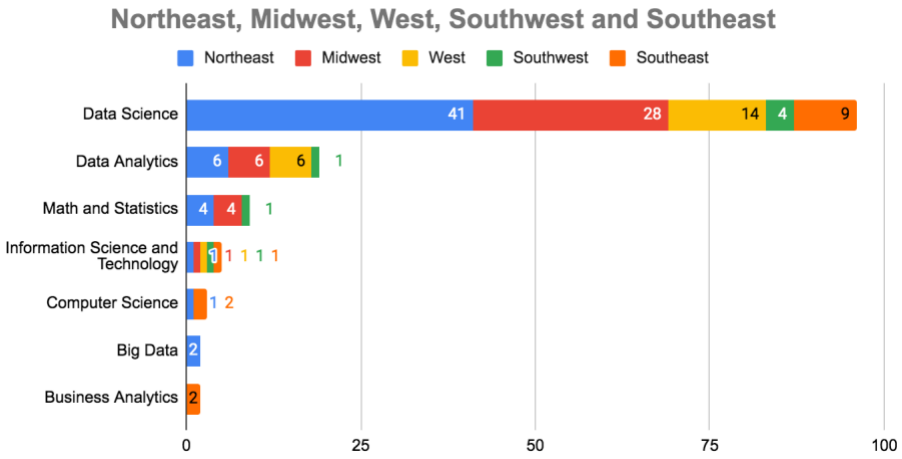

Figure 8: Majors and regions

Regions and Degrees

Analysis of degrees and regions showed that the BS degree is the most prominent degree in all 5 regions. The Northeast region has the largest concentration of BS degrees. Analysis also showed that the Northeast region has the greatest number of Associate degrees. The Southwest region only has BS degrees and has no associate or BA degrees. Table 14 and Figure 9 below show the analysis of degrees and regions.

	Northeast	Midwest	West	Southwest	Southeast
AA	1	0	0	0	0
AS	2	1	0	0	1
AAS	0	0	1	0	1
BA	4	0	2	0	0
BS	46	36	18	7	12
BS and BA	2	1	0	0	0
Certificate	0	1	0	0	0

Table 14: Degrees and regions

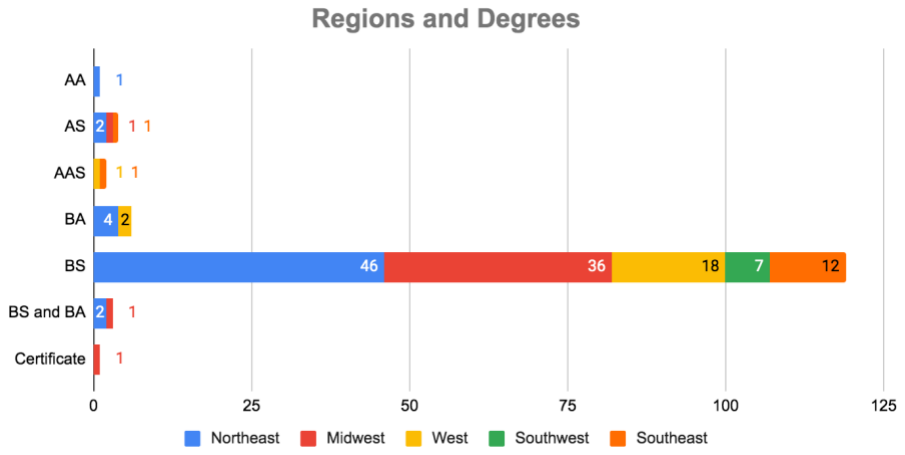

Figure 9: Degrees and regions

Regions and Type of School

Analysis of the type of school and regions showed that the Northeast region has the largest concentration of Private schools. The Midwest region has the second largest concentration of Private schools. The Western region has the third largest concentration of Private schools. The Southwest and Southeast regions have the largest concentration of Public schools. In the northern regions, there are more private schools than public schools. In the southern regions there are more public schools than private schools. Table 15 and Figure 10 below show the analysis of type of school and regions.

	Northeast	Midwest	West	Southwest	Southeast
Private	40	21	14	1	3
Public	15	18	7	6	11

Table 15: Type of school and regions

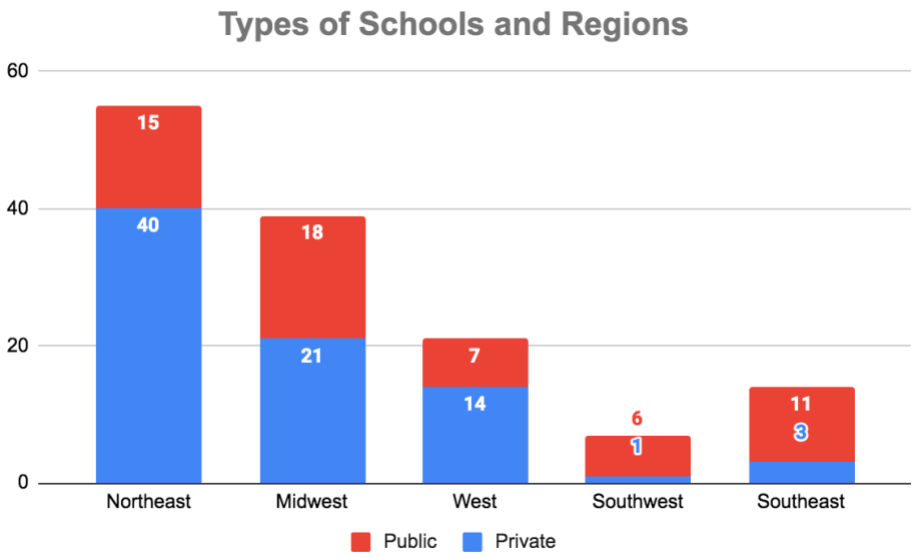

Figure 10: Type of school and regions

Regions and Departments

Analysis of departments and regions showed that in the Midwest region Mathematics/Statistics, Interdisciplinary and Data Science are the most prevalent departments. In the Northeast region, the most prevalent departments are Interdisciplinary, Data Science and Mathematics/Statistics. In the Southeast the most prevalent departments are Computer Science, Mathematics/Statistic, Information Science and Technology and Data Science. In the Southwest the most prevalent departments are Mathematics/Statistics and Interdisciplinary. In the Western region, the most prevalent departments are Mathematics/Statistics, Data Science and Interdisciplinary. Table 16 and Figure 11 below show the analysis of departments and regions.

	Northeast	Midwest	West	Southwest	Southeast
Business	2	0	1	0	1
Computer Science	6	5	1	1	5
Data Analytics	1	1	2	0	0
Data Science	12	6	5	0	2
Data Science and Statistics	3	0	0	0	0
Engineering	1	0	1	0	1
Information Science and Technology	6	1	1	0	2
Interdisciplinary	14	10	3	2	1
Math/Computer Science	3	1	1	1	0
Math/Statistics	7	15	6	3	2

Table 16: Departments and regions

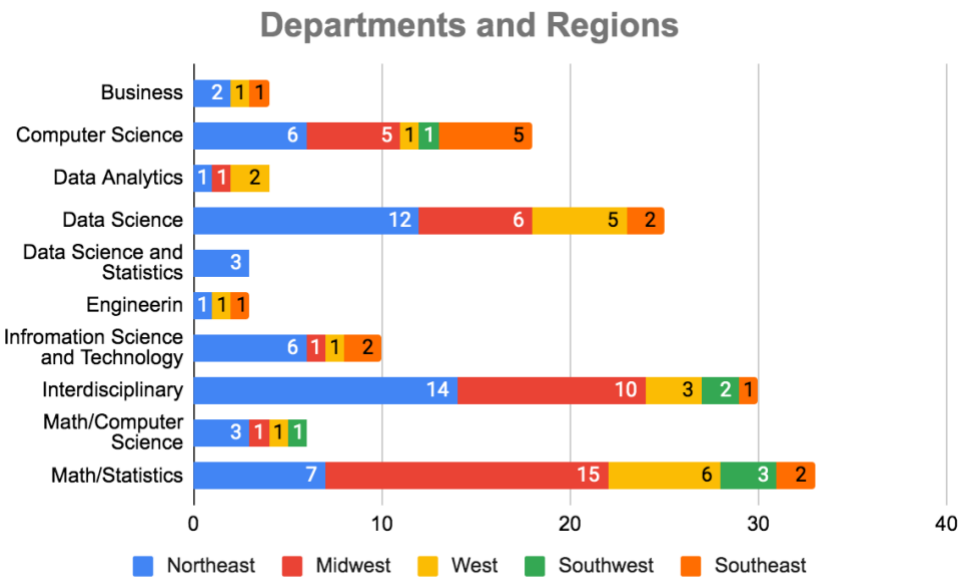

Figure 11: Departments and regions

4.2 Results Related to RQ2 - Data Science Competencies

In the following sections, results related to the data science competencies were examined and analyzed.

Number of colleges in which the competencies are found

Analysis of the competencies in the 136 colleges examined showed that Math and Statistics is the most common competency being offered in 135 out of the 136 colleges examined. Computing Fundamentals is the second most common competency found in 122 out of the 136 colleges. Data Management/Governance/Privacy is the third most common competency found in 109 out of the 136 colleges examined. Data Science in Context is the fourth most common competency found in 95 out of the 136 colleges examined. Machine learning is the fifth most common competency found in 85 out of the 136 colleges examined. Data Mining and Big Data are the sixth most common competencies found in 82 out of the 136 colleges examined. Data Visualization is the seventh most common competency found in 73 out of the 136 colleges examined. Table 17 and Figure 12 below show the results of the number of colleges and competencies.

	No. of Colleges
Computing Fundamentasls	122
Management/Governance/Privacy	109
Data Visualization	73
Machine Learning	85
Data Mining/Big Data	82
Data Science in Context	95
Math & Statistics	135

Table 17: Number of colleges and competencies

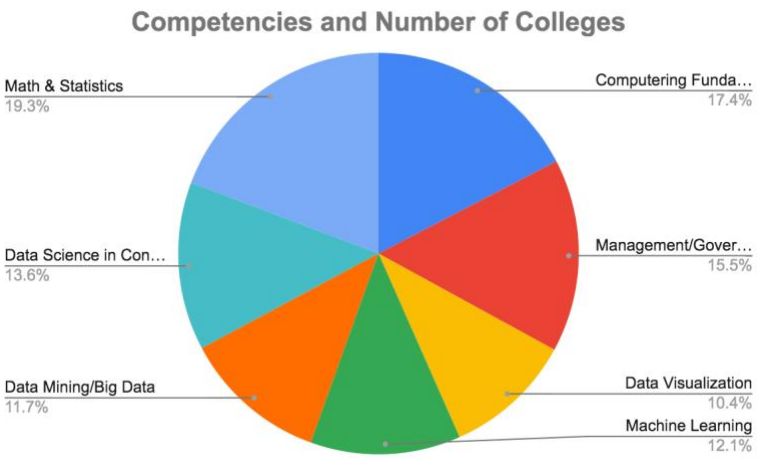

Figure 12: Number of colleges and competencies

Competencies and majors

Analysis of competencies and majors showed that the colleges that offer the Data Science major have the highest number of competencies followed by those that offer the Data Analytics major. The majors that have the least number of competencies are Math and Statistics, Computer Science, Big Data, Business Analytics, Information Science and Technology. Table 18 and Figure 13 below show the analysis of competencies and majors.

	Data Science	Data Analytics	Math and Statistics	Information Science and Technology	Compter Science	Big Data	Business Analytics
Computing Fundamentasls	85	17	9	4	3	2	2
Management/Governance/Privacy	78	17	6	3	1	2	2
Data Visualization	52	9	5	3	3	0	1
Machine Learning	64	9	7	3	1	1	0
Data Mining/Big Data	57	12	5	5	1	1	1
Data Science in Context	75	11	3	3	1	2	0
Math & Statistics	95	19	9	5	3	2	2

Table 18: Competencies and majors

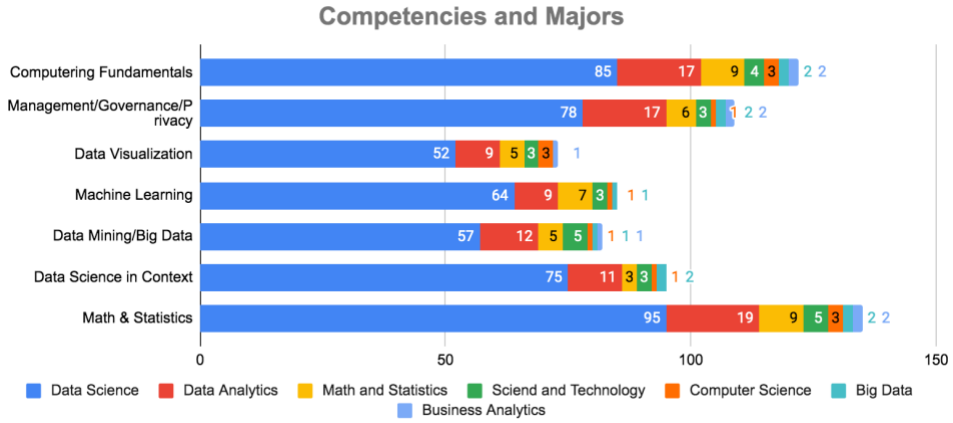

Figure 13: Competencies and majors

5.0 Discussion

The analysis of the 136 Data Science programs in the U.S. has shown that the Data Science educational domain is vibrant and thriving. The analysis showed key aspects of these programs which are discussed in the following sections.

5.1 Discussion related to RQ1: Program profile

In private and public colleges, data science curriculum is evident in the following majors including data science, data analytics, math and statistics and information science and technology. The prevalent degree in both private and public colleges is the BS degree and the BS degree is the most prominent degree in all majors with the exception of business analytics major. The Data Science major has the widest variety of degrees including AA, AAS, AS, BA, BS, BA and BS and certificate.

The data science major is prevalent in the Computer Science, Data Science, Information Science and Technology, Interdisciplinary and Mathematics and Statistics departments. The data science major has the widest range of core credits spanning 16 to 120. The business and engineering departments have an even distribution of data science, data analytics, big data and business analytics majors. Computer Science departments do not have computer science majors, this major is found in the Data Science and Math and Statistics departments.

The Big data and data analytics majors are the only majors in which online instruction is offered. Blended and in-person instruction are nearly even in terms of colleges offering these two modes of instruction and also in terms of colleges offering the data science major. Private colleges conduct more in-person instruction for the data science programs than public colleges.

The Northeast region has the most majors: Data Science, Data Analytics, Math and Statistics, Big Data, Computer Science and Information Science and Technology and the western region has the least number of majors: Data Science, Data Analytics and Information Science and Technology. The Data Science major and the BS degree are the most prevalent majors in all the five regions. The Northeast region has the largest concentration of Private schools followed by the Midwest region and the Western region. The Southwest and Southeast regions have the largest concentration of Public schools. The Midwest region, the most prevalent departments are Mathematics/Statistics, Interdisciplinary and Data Science are the most prevalent departments. In the Northeast region, the most prevalent departments are Interdisciplinary, Data Science and Mathematics/Statistics. In the Southeast the most prevalent departments are Computer Science, Mathematics/Statistic, Information Science and Technology and Data Science. In the Southwest the most prevalent departments are Mathematics/Statistics and Interdisciplinary. In the Western region, the most prevalent departments are Mathematics/Statistics, Data Science and Interdisciplinary.

It is interesting to note that private universities/colleges offer more data science undergraduate programs in the northern regions (northeast, Midwest, west), while public universities/colleges offer more data science programs in the southern regions (southwest, southeast). This may be due to the regions' corresponding economic levels.

5.2 Discussion related to RQ2: Data Science Competencies

Math and Statistics is the most common competency being offered in 135 out of the 136 colleges examined which is followed by computing fundamentals, data management/governance/privacy, data science in context, machine learning, data mining and big data and lastly data visualization.

Colleges/universities that offer the Data Science major have the highest number of competencies followed by those that offer the Data Analytics major. The majors that have the least number of competencies are Math and Statistics, Computer Science, Big Data, Business Analytics, Information Science and Technology.

5.3 Findings comparisons: Current Study and Previous Studies

By comparing the findings of this study to previous studies, it is clear that the current data science program is thriving and increasing. First, this is evident in the number of undergraduate data science programs which were identified through this study. The data science programs discovered as a result of this study were 136, an increase over the 2019 study finding [1] which identified only 107 undergraduate data science programs. Second, such growth is manifested by the number of programs offering BS degrees in data science. The current study identified 119 BS data science degrees again an increase over the 2019 study finding [1] which identified only 49 BS data science degrees. Third, the current study revealed that more departments are offering data science programs including the business, data analytics, data science and statistics and information science and technology departments. These four departments were absent from the earlier study [1] . Finally, three earlier studies identified the same competencies outlined in this study with two exceptions; governance and privacy which were absent from all three earlier studies. In addition, the data science in context competency was absent from studies [2] and [14].

6.1 Conclusion

The findings from this study show that data science undergraduate programs are thriving and increasing in concert with the demand in industry for data science. There are various private and public colleges offering data science programs across the U.S. These programs vary in core credits and degrees, however, the data science competencies as recommended by the ACM Data Science Task Force[17] are being implemented across all programs regardless of the major or departments.

The findings from this study also show that in order to pursue interest in data science domain, one should major in one of the three most prominent majors; data science, data analytics or math and statistics, and one should pursue a BS degree.

6.1 Study Limitation

The study had some limitations including, the absence of enrollment, retention and graduation data. In addition, the modality of instruction may be more indicative of the current times and not a true indication of how instruction is given traditionally because of the current Covid-19 crisis and move to online instruction. Lastly, data science program accreditation information was not available as many programs are fairly new and such accreditation may not yet be in place.

6.2 Future Research

This study surveyed the program profiles of U.S. undergraduate data science program in terms of their major names, affiliated department/school names, degree types, required credit numbers, and core competencies. Future research on curriculum details and job placement will shed light on the program outcomes. In addition, a comparison of undergraduate data science program and graduate data science programs will allow a more complete understanding of the data science education in the U.S.

Reference

[1] I. B. Hassan and J. Liu, "Data science academic programs in the US," *Journal of Computing Sciences in Colleges*, vol. *34*, pp. 56-63, 2019.

[2] X. Li, X. Fan, X. Qu, G. Sun, C. Yang, B. Zuo and Z. Liao, "Curriculum Reform in Big Data Education at Applied Technical Colleges and Universities in China," *IEEE Access*, volume *7*, pp. 125511-125521, 2019.

[3] I. Y. Song and Y. Zhu, "Big data and data science: opportunities and challenges of iSchools" *Journal of Data and Information Science*, vol. *2*, pp. 1-18, 2016.

[4] D. Debnath, "Undergraduate Degree Programs in Data Science" *In Proceedings of the 2016 ASEE North Central Section Conference March 18-19, 2016*. [Online]. Available: http://people.cst.cmich.edu/yelam1k/asee/proceedings/2016/faculty_regular_papers/2016_asee_ncs_paper_99.pdf

[5] L. Cao, "Data science: challenges and directions," *Communications of the ACM*, vol. *60*, pp. 59-68, 2017b.

[6] C.L. Aasheim, S. Williams, P. Rutner, and A. Gardiner, "Data analytics vs. data science: A study of similarities and differences in undergraduate programs based on course descriptions" *Journal of Information Systems Education*, vol. *26*, p. 103, 2016.

[7] A.P. Murillo, and K. M. Jones, "The development of an undergraduate data curriculum: A model for maximizing curricular partnerships and opportunities," In *International Conference on Information* (pp. 282-291). Springer, Cham, 2018.

[8] L. Cao, "Data science: a comprehensive overview," *ACM Computing Surveys (CSUR)*, vol. *50*, pp. 1-42, 2017a.

[9] R. Tang, and W. Sae-Lim, “Data science programs in US higher education: An exploratory content analysis of program description, curriculum structure, and course focus.” *Education for Information*, vol. *32*, pp. 269-290, 2016.

[10] P. Paul, and P.S. Aithal, “Computing Academics into New Age Programs and Fields: Big Data Analytics & Data Sciences in Indian Academics—An Academic Investigation,” *IRA-International Journal of Management & Social Sciences*, vol. *10*, pp. 107-118, 2018.

[11] R.D. De Veaux, M. Agarwal, M. Averett, B.S. Baumer, A. Bray, T. C. Bressoud, and A.Y. Kim, “Curriculum guidelines for undergraduate programs in data science”, *Annual Review of Statistics and Its Application*, vol. *4*, pp. 15-30, 2017.

[12] B. Heinemann, S. Opel, L. Budde, C. Schulte, D. Frischemeier, R. Biehler, S. Podworny, T. Wassong, “Drafting a data science curriculum for secondary schools,” In *Proceedings of the 18th Koli Calling International Conference on Computing Education Research* (pp. 1-5), November, 2018.

[13] A. Mikroyannidis, J. Domingue, C. Phethean, G. Beeston, and E. Simperl, “Designing and delivering a curriculum for data science education across Europe,” In *International Conference on Interactive Collaborative Learning* (pp. 540-550). Springer, Cham,2017..

[14] J. Saltz, F. Armour, and R. Sharda. "Data science roles and the types of data science programs," *Communications of the Association for Information Systems* vol. 43, p. 33, 2018.

[15] M. D. White and E.E. March, “Content analysis: A flexible methodology,” *Library Trends*, vol. 55, pp. 22-45, 2006.

[16] K. LaBarre, “The use of faceted analytic-synthetic theory as revealed in the practice of website construction and design,” Ph.D. dissertation, School of Library and Information Science, Indiana University, 2006.

[17] A. Danyluk and P. Leidig, “Computing competencies for undergraduate data science curricula,” Association of Computing Machinery (ACM), January 2021. [Online]. Available: https://dstf.acm.org/DSTF_Final_Report.pdf