Emergence of Data Analytics in the Information Systems Curriculum

Musa J. Jafar
Musa.jafar@manhattan.edu
Manhattan College
Riverdale, NY 10471 USA

Jeffry Babb
jbabb@wtamu.edu

Amjda Abdullat
aabdullat@wtamu.edu

West Texas A&M University
Canyon, TX 79016, USA

Abstract

As a phenomenon of interest, impact, and import, there is little doubt that the pervasive expansion of data is upon us as Information Systems educators. Concerns and topics such as Data Science, Data Analytics, Machine Learning, Business Analytics, and Business Intelligence are now ubiquitous and often situated as being the "next big thing." Educators and practitioners who identify and resonate with information systems, as a discipline, are watching these developments with interest. With data being both input and output to so many concerns that intersect with the information systems discipline, several themes emerge when considering what curriculum and pedagogy are appropriate. The role, position, location, and shape of data science topics are considered. Curricular approaches are also discussed with an eye to breadth and depth. Fundamental and existential questions are raised concerning the nature of data science and what role the Information Systems discipline can play. We also discuss evidence from cases. Case one involves a student business analytics competition and case two investigates how information systems knowledge areas can appropriate data science as an integral component of many competencies that exist solidly within the canon of Information Systems (IS) topics.

Keywords: IS Curriculum, Data Science, Data Analytics, Machine Learning, Business Analytics.

1. INTRODUCTION

In the late 1990's, the data mining discipline was viewed then as a "single phase in a larger life cycle" of Knowledge Discovery in Databases (Collier et. al., 1998). The earliest we could trace back the offering of datamining courses is to Guo (1998) and Lopez (2001). Since then, data mining has evolved from appearing as elective course in an Information Systems (IS) Curriculum (Lenox 2002, Patel 2003, Goharian 2004, Musicant 2006, Jafar 2008, Asamoah 2015) to a minor area of study, a co/dual-major, or even as a fully-independent degree program. Today, this degree program will typically be referred to as Data Analytics, Business Intelligence or Business Analytics. Although the content of the curriculum (or even the individual courses) is emergent, and therefore as stream-lined as in the case with the more mature disciplines of finance, accounting, marketing, we do see the need for an extended minor, a co/dual-major and/or a undergraduate program in data analytics in the IS curriculum. In this paper we extend our consideration of the data analytics subject by articulating the case and the requirements for a Master's degree in data analytics as an IS degree program.

In the past 10 years, terms and concepts popularly known as Big Data, Data Science, Data Analytics, Machine Learning, Business Analytics and Business Intelligence have become lexically normalized in the discussion of unfolding horizons that impact organizations, their use of information technology, and their expected utility of their information systems. In both the academic and corporate worlds, these terms are somewhat elusive as they mean different things to different constituencies (O'Neil 2014).

We use the terms machine learning, and data science to highlight the Statistics-Mathematic-algorithmic and the Computer Science aspect of the discipline where the theory is established and the algorithms are coded. We use the terms Data Analytics, Business Analytics or Business Intelligence to emphasize the applications side of the discipline where algorithms are understood, the underlying computing software is comprehended and utilized to solve business problems, reveal patterns and extract insights from the data. We use the term Big Data to emphasize the volume, variety, velocity and veracity of data and the need for fault-tolerant computing platforms that can manage large amounts of unstructured data where daily tasks need to be parallelized, distributed, load-balanced, processed and results are combined. It is the layer of abstractions created to hide the infrastructure code and manage these tasks. It is the Map-Reduce model and its derivatives. This paper focuses on curricular issues as they relate to data (business) analytics. We will use Data Analytics to mean both Data Analytics and Business Analytics.

Simon (2013) describes data analytics as "…the combination of statistics, mathematics, programming, problem solving, capturing data in ingenious ways the ability to look at things differently and the activity of cleansing, preparing and aligning data." Conway (2010) summarized it as the intersection of hacking skills, mathematics & statistics combined substantive expertise. Conway's (2010) Venn diagram elegantly and colorfully draws the boundaries between machine learning, traditional research, data science and the danger zones. Figure 1 is a testimony to the many intersections of the data analytics disciplines. For example, a self-driving car or a pattern recognition system are examples of machine learning; however an recommender system or a system (such as what drives the consumer experience when shopping on Amazon's website) that is used to reveal patterns constitutes data analytics. Figure 3 is an illustrative attempt to disambiguate the problem space.

The sum of these innovations in data management and use are both prescient and compelling in a contemporary dialog on the themes and content that define the IS discipline. As an inter-discipline, IS has commonly absorbed innovations over its history. Thus, it is quite normal for IS to develop an existential conversation when new waves of innovation impact its shores. However, as a bridging discipline, between organizations, people, information and computing technology, data is foundational the discipline's identity in an acute sense.

Normalizing the Discipline

Data is in the very bloodstream of an organization. Every aspect of business, government, science, humanities, medicine, etc. has both a data and an analytics component. Further having emerged out of the various schools of business concerns, IS has always had a central focus on data and its processing. Rhetorically, we can ponder "what do transaction processing and analytical processing have in common and where do they bifurcate?" Arguably, their intersection and union revolves around data. Transaction processing uses business models (rules) to manipulate the data. Simply put. it is SQL-based data warehousing analytics. Analytical processing on the other hand, relies on wider ranges of data including transaction processing and digital sensors (social media, web, apps, government, etc.). Analytical processing uses statistical models to sift through and extend the use of transactional data to produce insights and reveal patterns. We may even argue that analytical processing would not have existed if it was not for the maturity and openness of transaction processing systems.

Rush to Discipline

All of the business disciplines are exuberant over the prospects of data analytics. Further, this exuberance is leading to a rapid refashioning towards data analytics, up to and including the development of new programs. We could see a situation where no two Data Analytics degrees are the same as the tenets of the discipline are not focused, defined, or agreed upon. In many cases, this problem persists with IS discipline at large.

This leads to questions regarding who "owns" the data analytics topic. In academia, we can posit that one common point of contention in academic institutions these days is the issue of ownership and where should the different courses be housed. We can even ask where interdisciplinary programs such as text analytics and social media analytics should be housed. In the Kuhnian (2012) sense: the ebb and flow, and evolution and emergence of disciplines is per paradigmatic shifts. Disciplines flourish or flounder according to need and environment, but also fad and fashion. Thus the ownership question remains. Whether computer science calls it machine learning or data science, or mathematics-statistics call it data science or data analytics, or IS programs call it data analytics or business analytics, we acknowledge the fact that although the different disciplines have the same concerns, they have different focuses.

Fundamentals

We now examine the nature of the concerns, trends, emerging disciplines, and other phenomena surrounding the rush to data. We consider data analytics in terms of its first principles, and contrast and compare these first principles to those of IS.

Let us reflect on the circular relationship between transaction processing systems and analytic systems. To build transaction processing applications, data, business logic, reporting and presentation are core concerns. To build analytical processing applications, data, algorithms, model validation, discovery, insights and presentation are core concerns. Further, we can characterize transaction and business processing algorithms as a transparent and open box, whereas analytic processing can be seen as a translucent black box. In both cases, core concerns remain data design, gathering, repurposing, conceptualizing, storing, retrieval, manipulation and presentation. Some common present day use cases come to mind which belie the complexity and systems knowledge required to function effectively. Take the case of an association Rules (Recommender) System:

- Using R (programming language and software environment for statistical computing and graphics) and its packages, prepare the data so it is a transaction.
- Clean up the data to establish non-duplicate items in a basket. Try to interpret the results.
- Grapple with incompatibilities between package versions and R-Versions as the project evolves.
- Often, the data is poorly organized and error-prone as given/found/extracted. Null values, type mismatches, and other data quality issues may take hours or even days to correct.
- Data consolidation and exploratory data analysis is challenging and may require a variety of tools (R, Python, SQL, Excel, Excel pivot tables and tableau).
- Multi-tasking and high task saturation commonly accompanies the above steps in an iterative nature where jumps among the steps are common.
- We may also think that (which is most likely the case) that a transaction processing system provided input data to the recommender system.

Thus, actionable results are not simply a matter of firing up a data analytics computing engine, connect to data, and display tidy results on a dashboard. Rather, it is a matter of integration. Classic problems related to software engineering and IS development are readily evident. Basically a data analytics project is just another software engineering project where different people with different skills work together to produce a software product.

This narrative gives way to the central theses of this paper:

- What is data analytics?
- What does data analytics hold for IS, and vice versa?
- What makes data analytics different? (we already transitioned through the knowledge-engineering era and phasing-out the data warehousing era)
- How have colleges and universities met the growing demand for data analytics skills?
- What is a good IS approach to incorporate data analytics into the curriculum?
- What have we learned from early experiences?

These questions come full circle to our context of IS education as we must decide how, and to what extent, will data analytics pervade the discipline. There are even questions related to the appropriate level for engaging data analytics. Should data analytics gravitate more towards a graduate-level concern? At the graduate level, given the demographics of the students, we may focus on skill building and expanding the boundaries of their technical skill knowledge, contrasted with their foundations in a given subject matter area of expertise (medical, financial, marketing, education, learning, etc.). At the undergraduate level, the needs of foundational systems development topics may “crowd out” the data analytics topics such that the subject may not be feasibly or fully explored. Concomitantly, there is also a requisite level of intuition, driven by tacit knowing and acumen in subject matter expertise wrought through long-term exposure to data, necessary for success.

This is evident from the diversity of the degree programs in the area.

2. ARTICULATING A DATA ANALYTICS CURRICULUM FOR IS

As there are heterogeneous inputs to and consumers of data analytics, it is natural for many disciplines to become involved with data analytics and otherwise appropriate its benefits to suit their ongoing core dialog. As such, given the nascence and emergence of the "field" it is difficult to know where data analytics should call home. Or, should we accept it as a new discipline, as was afforded to IS at one time.

As such, degrees in data analytics are elusive as they cross-cut multiple concerns (Figure 1). Upon early inspection, one would find that there is no uniformity in the course offering of such degrees across Universities (content, course descriptions, requirements, prerequisites and transferability). Whereas more established disciplines - Marketing, Finance, Computer Science, and Political Science - have established a reasonable degree of consistency in their curricula.

While data analytics is certainly consistent with most Universities' mission statements, there are no clear imperatives that compel data analytics to locate in one area or another. Furthermore, even the modalities of delivery are neither suggestive nor limiting. Accordingly, digital learning, on-line programs, graduate programs and service to the local community – to attract professionals, continuing and life-long learners, or even alumni – are all also neither prohibitive nor suggestive. This paper's authors attempt to articulate a proposal of what a master's program in data analytics might look like. This section will proceed to share the broad strokes of the proposal and otherwise share our discoveries.

Program Modalities

Increasingly, accommodating a wide variety of delivery modalities is necessary. Thus, any of the following modalities is recommended:

- An on the ground program,
- An On-line program,
- A hybrid of on-line and on the ground program.

Proposed Degree Requirements

For the purpose of this paper, we surveyed a representative sample of more than 13 graduate programs and 8 undergraduate programs that offer degrees in Data Analytics, Business Intelligence, or Business Analytics. We observed that there is more uniformity in the course offerings at the graduate level as compared to the undergraduate. We compared the content of their courses, pulled as many syllabi as we could, read and compared their course descriptions. We summarized and compared their course contents. Although all of these programs share a common theme, the structure of the curriculum, the course requirements, prerequisites, course descriptions, computing technologies, focus and depth of offering widely varied. We offer this as an indication of a discipline in flux and under formation. The findings of this analysis are shown in Figure 2. Note that the MIT Sloan School of Management just introduced a One-Year Masters of Business Analytics program that is modular in nature (MIT-Sloan 2016). We were not able to collect complete descriptions of the Modules (beyond the Optimization Methods, Intro to Applied Probability, Data Mining, the Analytics lab courses and the Analytics Edge which is also a coursers.org course). Since we did not have complete information about the degree program, we opted not to include it in the graduate programs list.

From our review of these graduate programs, the following curricular patterns emerge.

- Graduate programs typically require a total of 30-36 credit hours as follows: 21-24 core credits plus 6-9 approved elective credits and a 3-credit Capstone Course.
- There is no uniformity of offerings across the undergraduate programs, some programs just rebranded and renamed their data warehousing ETL courses into Big Data-I and Big Data-II, or the Business Statistics courses into Data Analytics-I and Data Analytics-II.

Students from a computing discipline such as Math, CIS, CS, and Statistics may have 6 approved credits waived if they have completed the equivalent course work with a B grade or higher. Candidate courses are Programming, Data Management and Statistical Data Analysis. The courses for the degree are grouped into five separate categories as follows:

- Core Data Management: 3 Courses
- Core Statistical Data Analysis: 2 Courses
- Core Data Analytics: 2-3 Courses
- Core Capstone Course: One Course
- Electives: 2-3 Courses

Pre-Requisite Knowledge

Fundamental prerequisite knowledge from an undergraduate (or graduate if 2nd masters) education would include knowledge of programming, statistics, and calculus. For students who do not have the background knowledge, they would either take leveling courses or use identified bridging courses which may include approved Massive Open Online Courses (MOOC) offerings where a certificate of completion satisfies the prerequisites that can be obtained.

The Graduate Education Component

In this section we explicate a provisional design for a graduate curriculum in data analytics.

1. Three Core Data Management Courses

1.1 Programming for Data Analytics
Pre-req: Basic Programming Knowledge

Description: Reading data from different data sources & streams such as files, web-searching, etc. Organizing, manipulating and repurposing data. Parsing-in and storing-out data JSON formats, string manipulation using RegEx libraries, Sets, Arrays and dictionaries.

Technology: Python, RegEx, DOM, JSON.

1.2 Data Management for Data Analytics
Pre-req: Programming

Description: Fundamentals of sound database design, storing, manipulating & retrieving data, Conceptual Data Modeling, SQL, Functional Dependencies, Data Normalization.

Technology: SQL, MySQL DBMS, and ER-Modeling tools.

1.3 Data Visualization for Data Analytics
Pre-req: Programming
Co-req: Statisitcs-1

Description: Provide an understanding of the different data types and their encoding schemes. Learn data visualization principles and Mantra(s). Learn how to tell a story through data. Learn how to detect insight through data, learn and utilize current technologies to visualize large data sets for the purpose of providing insight.

Technology: Excel, Tableaux, Web-GL, D3.JS, R-ggplot and R-Shiny.

2. Two Core Statistical Data Analysis

2.1 Statistics-I for Data Analytics
Pre-req: Business Calculus

Description: Probability distributions and their applications (geometric, binomial, Poisson, uniform, normal, exponential, t, F and Ki-Squared), Sampling Statistics, Confidence Intervals, Hypothesis Testing, Analysis of Variance and Linear Regression Modeling

Technology: Excel, R and R-Packages.

2.2 Statistic-II for Data Analytics
Pre-req: Statistics-I

Description: The different types of Regression Models and Time Series forecasting. The course emphasizes statistical computing.

Technology: Excel, R and R-Packages.

3. Three Core Data Analytics

3.1 Principles of Data Analytic
Pre-req: Stats-I, Data Management
Co-req: Data Management

Description: Exploratory Data Analysis, Statistical Models, Classification and Prediction, Clustering Analysis, Similarity Measures, Fitness of Models, learn how to use machine learning technologies to analyze data for the purpose of decision support.

Technology: Excel, R and R-Packages.

3.2 Advanced Data Analytics
Pre-req: Principles of Data Analytics, Stats-II

Description: Optimization Models, Social media analytics, Text analytics, advanced Analytics Algorithms like SVM, Neural Networks, bootstrap models, Model Validation.

Technology: R, R-Packages, Gephi and NodeXL.

3.3 Big Data Analytics
Pre-req: Advanced Data Analytics

Description: Big Data Meets Data Science, Meets Data Analytics. Map-Reduce Model, Apache Spark, Mongo DB, NoSQL Model and Scale out Models.

Technology: Cloud Computing, Apache Spark and MongoDB.

4. Capstone Course

4.1 Capstone Course
Co-req: Big Data Analytics

Description: Special Topics in Business Analytics with a comprehensive real world project, the project usually extends over 2-terms, a summer and the fall or the spring and the summer. Although the topics of the course are taught by one faculty and during one semester, a student might be working with another faculty from the program or from an approved discipline on their project. The project may be in collaboration with a partner organization or a business.

Technology: Whatever is needed for a successful project portfolio.

5. Three Elective Courses (More electives can be added)

5.1 Legal and Ethical Issues of Data
Prereq: Principles of Data Analytics

Description: Provide an understanding of the legal and ethical issues as it relates to data storage, retrieval access and sharing as well as analytics models. Copyright law, legal and ethical ramifications are at issue when recommending choices that have economic, social, environmental or legal impacts.

5.2 Project Management
Prereq: Principles of Data Analytics

Description: A data analytics project is just another software engineering project that will culminate in a software product. Deliverables, artifacts, timelines and resources need to be tracked and managed.
Technology: Project Management Software.

5.3 Decision Modeling
Prerq: Statistics-II

Description: Understand, formulate, solve and analyze optimization problems in the business domain and its operations. Utilize excel functions and macros to perform what-if analysis. Understand the geometric interpretation of linear optimization problems, Use solver family of packages to solve linear optimization problems. Formulate and solve Network graph types of problem especially in the social media domain.

Technology: Excel, Solver, Gephi.

In figure 4 a visual of the course sequences and their dependencies is presented.

3. TESTING THE ASSUMPTIONS – LESSONS LEARNED FROM TWO CASES

Student Data Analytics Competition Case

We continue our inquiry with a case description that illustrates some of the points made thus far about the nature of a data analytics curriculum and its relationship to the IS curriculum. The case is situated about a student data analytics competition held in New York City in the spring of 2016. The objective of the two-day student competition was to provide insights into a data-set that was made available from an online discount retailer that aggregates luxury brands and offers them at discounted prices. The student team had ten weeks to produce a poster that explains the nature of the data, the problems that surround it, and their insights into the data. The team's mentor (a coauthor of this paper) provided guidance and acted as a chief architect, coach, and project manager for the team. At the competition, the students presented their poster to a team of judges. Subsequently, after this initial round of judging, the students are given a new related data set that the team has not seen previously. Each team then had twelve hours in a reserved space to analyze the new data set to develop new insights, prepare and submit a presentation the next morning (No communication with the mentor was allowed). The team would then deliver a ten-minute presentation to a team of judges. Both the new analysis and preparation must have been completed within the twelve-hour overnight period. There were 18 teams competing from various North American universities.

The crux of the competition is to use the given data, and the context of the company providing the data, to bring technical solutions to business problems. During the 10-weeks period, the students and mentor approached the problem as though it were an IS development, project which requires: planning; skills development; computing resources; produce artifacts and deliver a product. The team of interest consisted of juniors and seniors, three members in total. Collectively, the team demonstrated sound business, analytical, technical skills and dedication. Salient to the competition, two students were enrolled in a data mining course during the semester of the competition. Relevant course topics included data mining applications, algorithms and technologies (using R, R packages, and Tableau). Further, at the start of the project, the team was already familiar with Python and its APIs, Data Modeling and SQL, SAS, R, Tableau, and everyone knew Excel very well. From an Analytical perspective, the team was familiar with data modeling concepts, business statistics, statistical modeling, Finance, Marketing, etc.

As the project progressed, the team's depth of knowledge matured as well as an increasing awareness that more concepts, knowledge, and technologies were yet to be learned. Concomitantly, as their semester in data mining wore on, the team acquired more in-depth knowledge of R's data mining packages, ggplot (a graphing and plotting package for R), and Tableau (for visualization and visual analytics) Furthermore, the students learned SQL-Server Analysis Services Business Intelligence platform on the side and heavily utilized its capabilities into their analysis.

The poignant aspect of this case is the wonderful outcome whereupon this team won first prize in the competition. Thus, we present the case not only to characterize and celebrate the student accomplishment, but to highlight what can be learned from this experience on these main grounds:

- What can we generalize from the experience about a successful data analytics project?
- What are the implications of these generalizations on a data analytics curriculum?
- How can IS the discipline at large, incorporate data analytics into the curriculum?
- Lastly, what does IS have to offer to data analytics?

Upon reflection, the ingredients of the students' success can be attributed to the mix of student talent, mentoring, technology infrastructure, competency in underlying computing technology, and the utilization of data analytics skills with an eye for the foundational business problem. That is, the combination of these factors put these students in a good position to prevail.

As we generalize from the student's experience, it is important to note that their accomplishment was entirely their own work. Grappling with the new data set, the twelve-hour overnight drill, the insights produced, and the quality of the presentation was accomplished with no input from the mentor. While orientation, training, coaching and mentoring all occurred in the lead up, the student team ran the relay race.

Lessons from the Case

While the lessons from the case are myriad, a standout lesson certainly lies in the blend of skills utilized, the quality of the individuals involved, and the overarching perspective assumed throughout the competition.

First, the success of the project can be primarily accorded to focus on utilization of the correct computing tools underscoring the true nature of a data analytics project – it is still largely a software and systems problem. Knowledge of computing tools to analyze, clean, prepare and repurpose the data were required in addition to knowledge about the nature of data. Furthermore, the data is never encountered in a perfect state: that the data requires review, cleaning and preparation.

Further, a business and subject-matter orientation was adopted, which is often characteristic of many contemporary software and systems approaches – early and iterative delivery of working artifacts in close cooperation and partnership with stakeholders.

Next, the team was aware of a blend of competencies that we can generalize away from the competition setting and to several emergent concerns. The team understood the importance of APIs – addressing them, consuming them, integrating their outputs into their analysis system, understanding how API inputs could emanate from their system, etc. Often, these APIs are consumed through calls to REST services where JSON is the carrier of the data. The team understood the general processes surrounding software and systems development as well as the concepts, tools and techniques of data analytics. For modeling, statistics were important as well as the modeling techniques inherent in data mining. Finally, business acumen – an awareness of the business context of the problem and the need to integrate business-oriented decision-making into and out of organizational management IS – was a significant success factor. Through the poster and the presentation, they were able to tell a story, reveal patterns and provide insights.

Some of the final lessons from this case pertain to curricular and disciplinary concerns. In a previous section, the outline for a graduate curriculum in data analytics was supported by the outcomes of the case. The curriculum for data analytics must blend data management, traditional system development, data and statistical modeling, and the business context. As is the case with computing, software, and systems, these elements span a spectrum where a universal curriculum of data analytics may not coalesce cleanly. Some of these concerns may lean considerably into traditional computer science and software engineering and yet others may lean towards the spaces occupied by IS. Additionally, some of the modeling may lean toward operations research. In the discussion, we shall elaborate further on how the bridging and spanning nature of the IS discipline are well-suited to provide leadership as data analytics emerges as an ongoing concern.

Twitter Text Analytics Case

The second case is an exposition on the requirements for a twitter text analytics research project that culminated in the case being published in The Case Research Journal of the North American Case Research Association. This is shared to highlight typical challenges faced when undertaking a type of research project typically used to characterize the power of data analytics. We highlight the challenges in what seems, on the surface, to be a simple and straightforward endeavor – using well-documented APIs to bring Twitter data into an application or data analysis context. It is a common misconception that Twitter data are ready, as is, for text and social networks analytics. Some software packages like *NodeXL (Hansen 2011)*, attempt to provide easy and straightforward access in the social network of a tweets dataset. However, in our experience, the matter is not simple unless the requirements for analysis are superficial. That is, repurposing Twitter data sets for analytics requires a wide variety of skills. Skills in programming languages like Python, in statistical packages like R and text analytics packages like R-tm and R-weka. Occasionally web-development skills are required for DOM scrapping of web-pages. This does not even account for the set of skills required to build effective systems for reporting, integration, and other interactions required to make use of this knowledge. Commonly, Twitter allows access to their data in multiple modalities:

1. ***Through the web-interface***: Users search content based on screen-name, hashtags, mentions, body of text, etc. They can read content, take notes, print pages, type counts (favorite, retweet, etc.) into an excel spreadsheet. Content is not captured into structures in digital formats where it is easy to manipulate. This has disadvantages based on how cumbersome this process is when large datasets are desired.

2. ***Through the Twitter-API(s)***: These are well documented and afforded programmers with extensive real-time access to Twitter's data. For instance, a programmer may use Python to retrieve JSON data structures which can be passed on to R for analysis. Some of the challenges with this modality would be learning the protocols of the Twitter-API, communicate, request and store the content. It is important to note that these are significant computing skills that need to be acquired in order to utilize this modality.

3. ***Subscription and purchase***: Based on usage agreements, users can purchase content from Twitter subsidiaries. Usually content is delivered in file fragments where each file may contain a bundle of tweets in JSON format plus some other trailing file meta-data that needs to be cleaned-up. These file bundles often need to be combined (and cleaned-up) in order to be useful.

The implication here is that the data will not move into and out of our data analytics systems without curation and coaxing.

Case Details

We continue with a discussion on how the Twitter data may be used in a typical data analytics context. In the project, we used Twitter data to analyze a crisis situation in a market-research study. We needed to have the full data set, we purchased 53,900 tweets from Gnip (a twitter subsidiary) that comprehensively covered 40 consecutive days of the hashtags "XXXXX" or "YYYYYY". During the analysis, we learned more about the twitterers, their agendas, and their narratives. The primary twitterer, and the creator of one of the hashtags that went viral is Mr. John Dow. In this study, we wanted to know what John Dow's social media interests on Twitter were the day before he brought these hashtags to life. The challenge lies in the fact that the Twitter timeline API allows access to a limited number of historical tweets from a user's timeline and Mr. DOW is an avid twitterer the limited number of historical tweets was insufficient. Other than purchasing the content, it seemed that we could not answer that question. Initially, our only option was to scrape the data from twitter.com search. Using Google-Chrome with the advanced search option, we were able to display the tweets on the web-page. With Chrome Inspector, we pinned the content and saved it as html text document. Using Python, we wrote code that parsed the content of the file for the id(s) of the tweets, and then we used the Twitter API functions we iterated through the id(s), extracted the full content of the tweets. We were then able to perform the needed text analytics for the paper.

Lessons from the Case

The most poignant lesson to emerge from the case described above is the issue of federation. Data will not always cleanly answer questions, and data analytics fundamentally exists to facilitate questioning the leads to information and decision-making. There are two important "gravities" here: the need to work with computing tools for data manipulation, and the need for discretion in subject-matter expertise that informs the extent to which we will strive for the needed data. These worked together in the case in that the subject-matter was not the social data, but the knowledge of how data can be obtained and maintained.

4. CONCLUSION

In summary, to conceive of and design a Master's of Science in Data Analytics is to engage an emergent and multi-disciplinary phenomenon. We must embrace multiple imperatives and paradigms: the objective of which is for students to master, at least, the tools and techniques of both data science methods & their computing.

This objective requires addressing knowledge areas from multiple disciplines. Data Analytics represents the intersection of Computing, Statistics, and other disciplines from Humanities, Business, Science, Bio-Informatics, Learning Analytics, Natural Language Processing, Social Networks Analysis, etc. At the same time, a data Analytics project is a software engineering project that requires different people from different disciplines with different skills working together to build a data product that is useful, usable and maintainable. Although we do not expect everyone in the project to be as skillful, knowledge of analysis and design and project management principles is essential.

To build data products and to transfer data into insights that can be utilized, data engineers take large data sets from different sources that need to be repurposed, cross-referenced, analyzed and presented. The same project might require multiple computing machinery and machine learning algorithms to exploit the underlying structures of the data. Exploratory data analysis requires data visualization tools and dashboards, statistics and machine learning algorithms for the purpose of finding patterns that can be exploited within the underlying disciplines. Different data types require different types of repurposing techniques, analysis and discovery methods. Numeric data are statistical, social networks data are graphs, trees and associations' centric, textual data require natural language processing, named-entity recognitions and sentiment analysis. Being able to insure a reliable computing and hardware infrastructure is a challenge.

In conclusion, we hope that we made the case that a Master's degree in data analytics is the right place for IS educators start the assimilation of the data analytics phenomenon into their discipline, departments, and programs. The masters level is where the most holistic picture of what data analytics will mean for our discipline and its premise. We also hope that we provided a template and road map of a model curriculum for a Master's degree in data analytics based on our experience, research of the programs we analyzed. However, at the undergraduate level an extended minor or a co/dual-major (Programming. Data Management, Advanced Statistics, Principles of Business Analytics plus another elective) gives undergraduate students reasonable amount of knowledge and awareness of the application of data analytics into their area(s) of specialization.

Finally we even created a poster that highlights the different aspects of the degree program (Figure 5).

5. REFERENCES

Asamoah, D., Doran D., Schiller, S. (2015). Teaching the Foundations of Data Science: An Interdisciplinary Approach. Pre-ICIS SGDSA Workshop. Retrieved Aug 15, 2016 from https://works.bepress.com/daniel_asamoah/14/

Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Chiang, R., Goes, P., & Stohr E. (2012). Business Intelligence and Analytics Education, and Program Development: A Unique Opportunity for the Information Systems Discipline. ACM Transactions on Management Information Systems 3(3). 12:1-12.13

Collier, K. Carey B., Grusy, E. & Marjaniemi, C. (1998). A Perspective on Data Mining. Retrieved Dec. 18, 2016 from: http://www.insight.nau.edu/downloads/DM%20Perspective%20v2.pdf

Conway, Drew (2010). The Data Science Venn Diagram. Retrieved Aug 20, 2016 from http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Guo, Y., (1998). Data Mining: Theory and Practice. Retrieved Dec. 20, 2016 from http://www.doc.ic.ac.uk/~yg/course/dmml/

Hansen, D., Shneiderman, B., & Smith, M. (2011). Analyzing Social Media Netowrks with NodeXL. Elsevier Inc.

Hirschheim, R., & Klein, H. K. (2012). A glorious and not-so-short history of the information systems field. Journal of the Association for Information Systems, 13(4), 188.

Goharian, N., Grossman, D., & Raju, N. (2004). Extending the undergraduate computer science curriculum to include data mining. Proceeding of the International Conference on Information Technology: Coding and Computing, 2, P.251

Jafar, M. J., Anderson, R. R., & Abdullat, A. (2008). Data mining methods course for computer information systems students. Information Systems Education Journal, 6(48).

Kuhn, T. S. (2012). The structure of scientific revolutions. University of Chicago press.

Lopez, D. and Ludwig, L, Data Mining at the Undergraduate Level, Proceedings of the Midwest Instruction and Computing Symposium, 2001.

Musicant D. R. (2006). A data mining course for computer science: primary sources and implementations. SIGSE '2006 Proceedings of the SIGCSE technical symposium on computer science education.

Lyytinen, K., & Yoo, Y. (2002). Ubiquitous computing. Communications of the ACM, 45(12), 63-96.

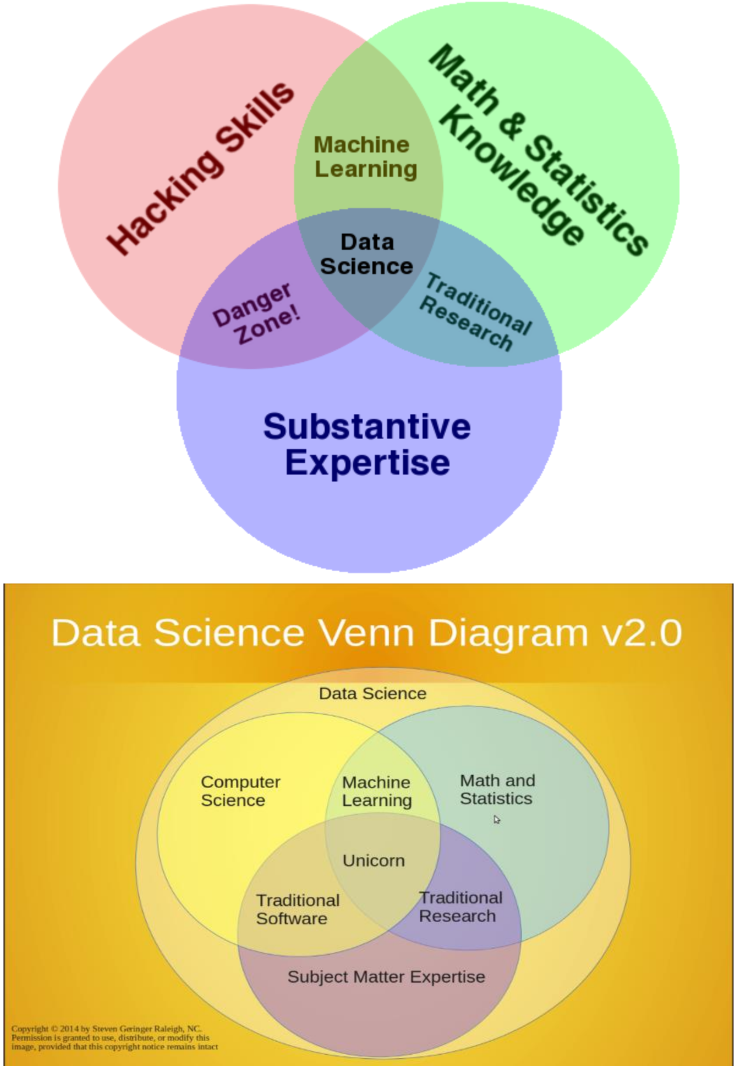
O'Neil C., & Schutt, R. (2014). Doing Data Science. Oreilly, USA.

Patel, N. (2003). 15.062 Data Mining, Spring 2003. Retrieved Dec. 15, 2016 from: http://ocw.nur.ac.rw/OcwWeb/Sloan-School-of-Management/15-062Data-MiningSpring2003/CourseHome/index.htm

Simon, P. (2013). Too Big to Ignore: The Business Case for Big Data (Vol. 72). John Wiley & Sons.

Terri L. Lenox (2002) Development of a Data Mining Course for Undergraduate Students. ISECON-2002.

Appendix



Copyright © 2014 by Steven Geringer Raleigh, NC.
Permission is granted to use, distribute, or modify this image, provided that this copyright notice remains intact

Figure 1: (Conway 2010, Germiger 2014) Data Analytics is Cross Cutting.

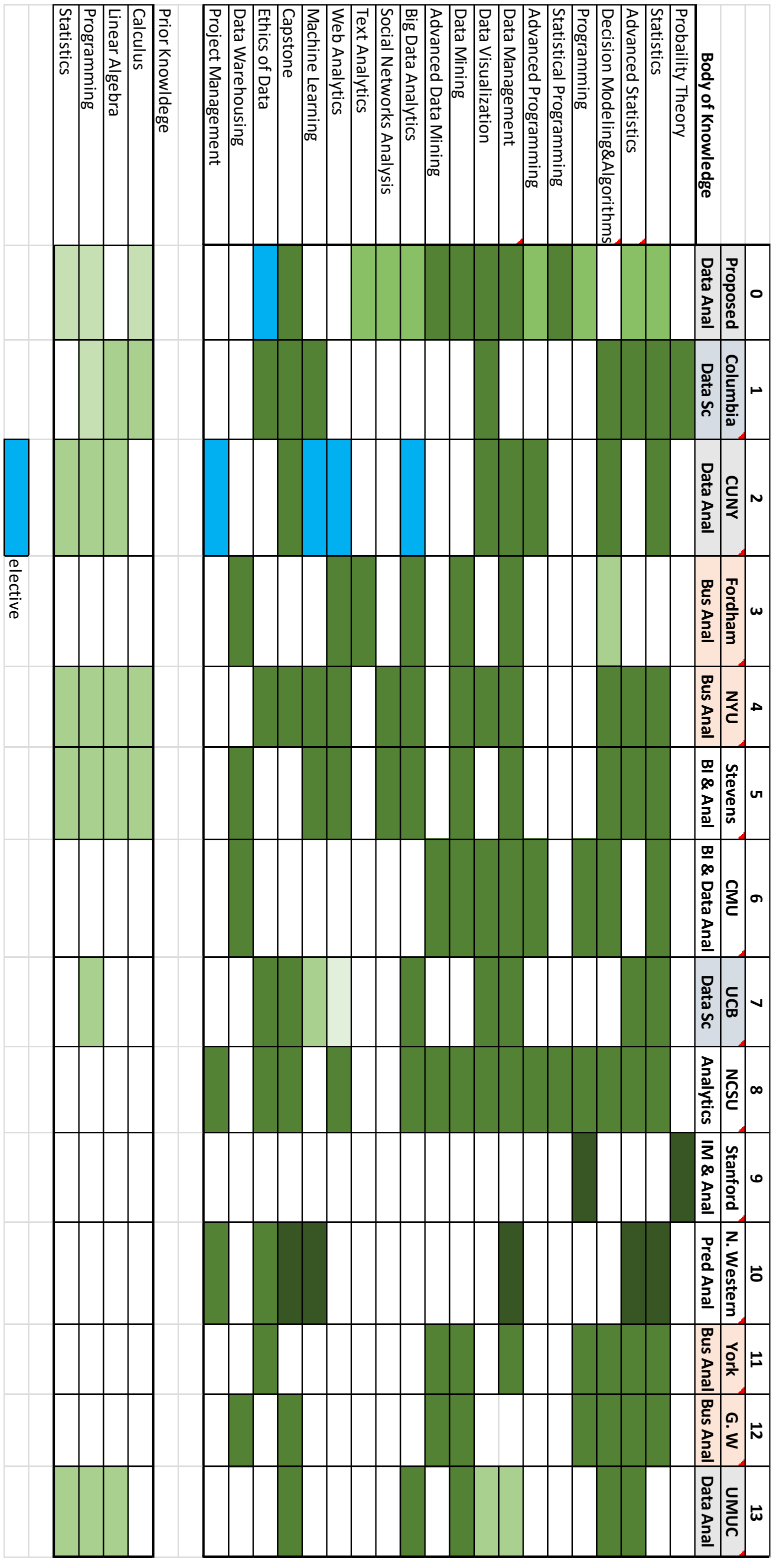
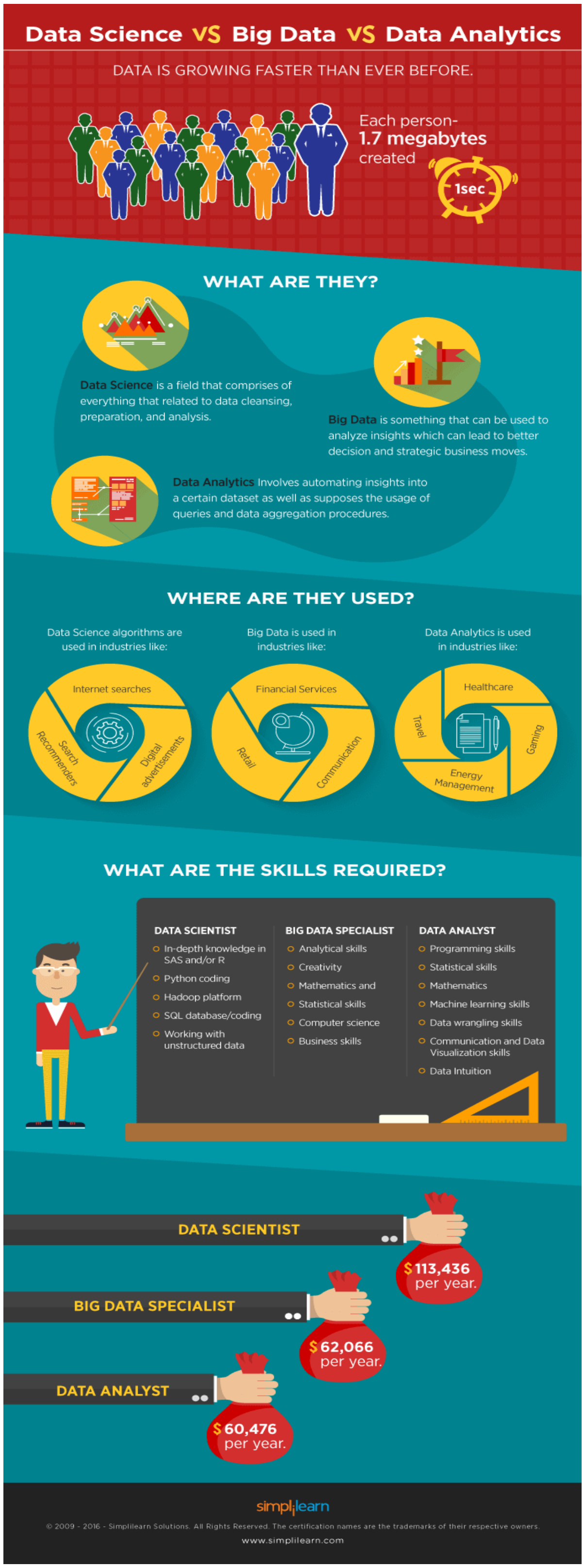


Figure 2. Topic map against existing programs, the darker the color the heavier the emphasis



Data Science VS Big Data VS Data Analytics

DATA IS GROWING FASTER THAN EVER BEFORE.

Each person- **1.7 megabytes** created

WHAT ARE THEY?

Data Science is a field that comprises of everything that related to data cleansing, preparation, and analysis.

Big Data is something that can be used to analyze insights which can lead to better decision and strategic business moves.

Data Analytics Involves automating insights into a certain dataset as well as supposes the usage of queries and data aggregation procedures.

WHERE ARE THEY USED?

Data Science algorithms are used in industries like:

Big Data is used in industries like:

Data Analytics is used in industries like:

WHAT ARE THE SKILLS REQUIRED?

DATA SCIENTIST

- In-depth knowledge in SAS and/or R
- Python coding
- Hadoop platform
- SQL database/coding
- Working with unstructured data

BIG DATA SPECIALIST

- Analytical skills
- Creativity
- Mathematics and
- Statistical skills
- Computer science
- Business skills

DATA ANALYST

- Programming skills
- Statistical skills
- Mathematics
- Machine learning skills
- Data wrangling skills
- Communication and Data Visualization skills
- Data Intuition

DATA SCIENTIST $113,436 per year.

BIG DATA SPECIALIST $62,066 per year.

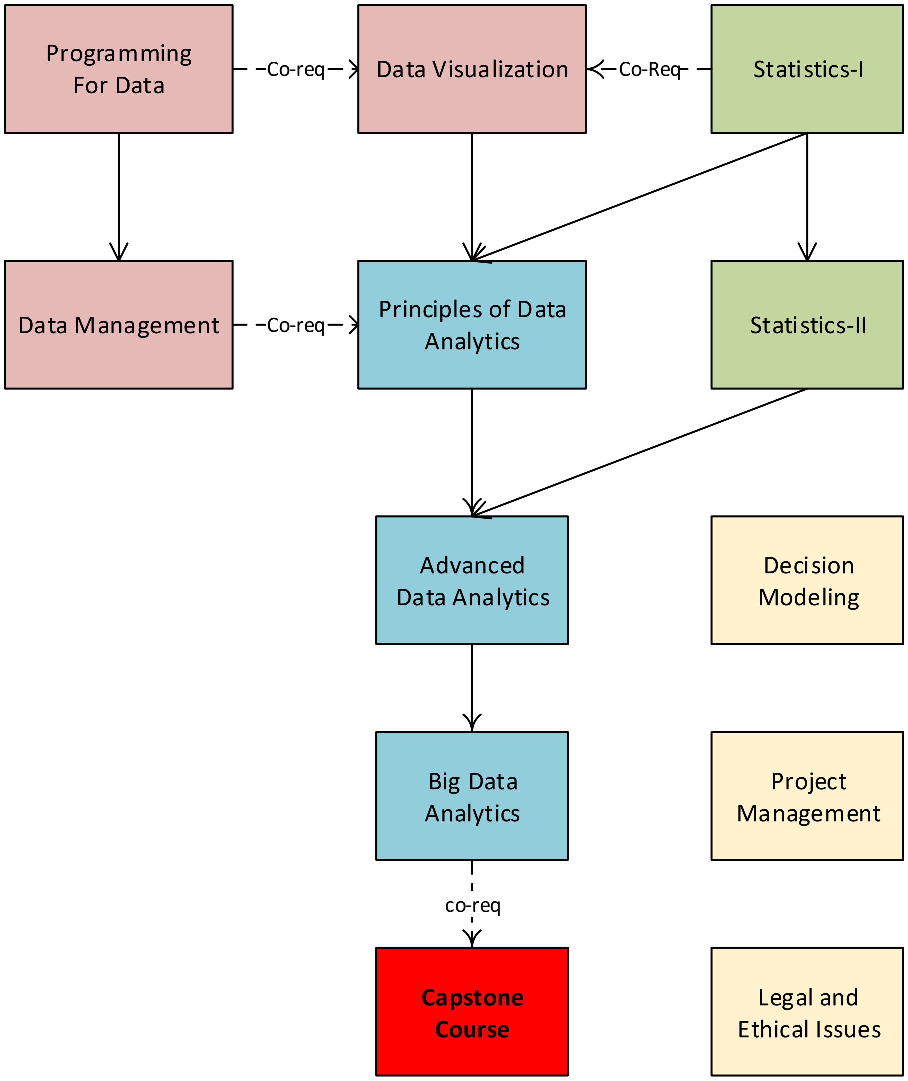
DATA ANALYST $60,476 per year.

simplilearn

© 2009 - 2016 - Simplilearn Solutions. All Rights Reserved. The certification names are the trademarks of their respective owners.

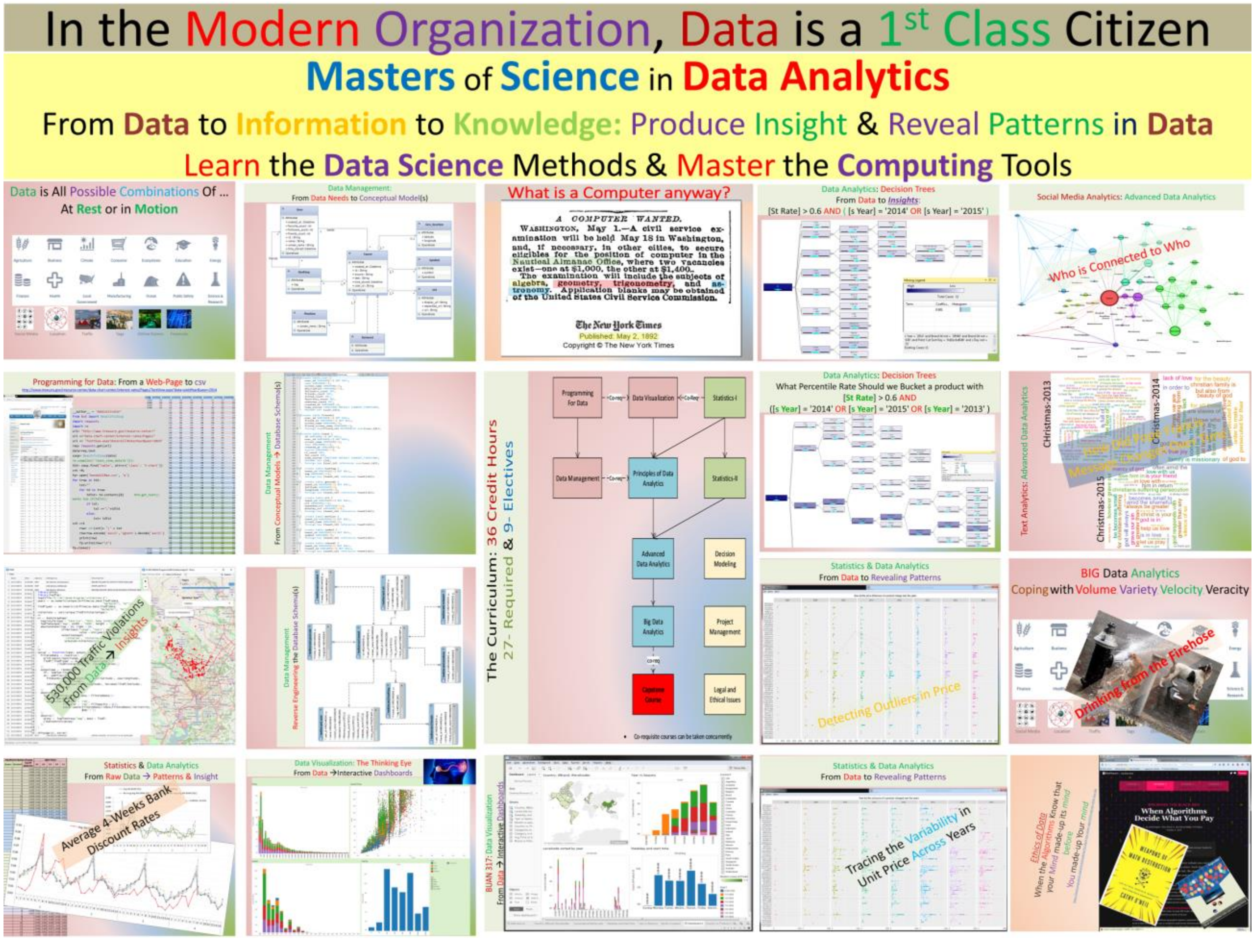
www.simplilearn.com

Figure 3. Differentiating the Problem Space
(http://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article)



- Co-requisite courses can be taken concurrently

Figure 4. Course Sequences and Dependencies



In the Modern Organization, Data is a 1st Class Citizen

Masters of Science in Data Analytics

From Data to Information to Knowledge: Produce Insight & Reveal Patterns in Data

Learn the Data Science Methods & Master the Computing Tools

What is a Computer anyway?

A COMPUTER WANTED.

WASHINGTON, May 1.—A civil service examination will be held May 18 in Washington, and, if necessary, in other cities, to secure eligibles for the position of computer in the Nautical Almanac Office, where two vacancies exist—one at $1,000, the other at $1,400..

The examination will include the subjects of algebra, geometry, trigonometry, and astronomy. Application blanks may be obtained of the United States Civil Service Commission.

The New York Times
Published: May 2, 1892
Copyright © The New York Times

The Curriculum: 36 Credit Hours
27- Required & 9- Electives

Data is All Possible Combinations Of ...
At Rest or in Motion
Data Management:
From Data Needs to Conceptual Model(s)
Data Analytics: Decision Trees
From Data to Insights:
[St Rate] > 0.6 AND ([s Year] = '2014' OR [s Year] = '2015')
Social Media Analytics: Advanced Data Analytics
Who is Connected to Who
Programming for Data: From a Web-Page to csv
Data Management
From Conceptual Models → Database Schema(s)
Programming For Data
Co-req
Data Visualization
Co-Req
Statistics-I
Data Management
Co-req
Principles of Data Analytics
Statistics-II
Advanced Data Analytics
Decision Modeling
Big Data Analytics
Project Management
co-req
Capstone Course
Legal and Ethical Issues
Co-requisite courses can be taken concurrently
Data Analytics: Decision Trees
What Percentile Rate Should we Bucket a product with
[St Rate] > 0.6 AND
([s Year] = '2014' OR [s Year] = '2015' OR [s Year] = '2013')
Text Analytics: Advanced Data Analytics
CHristmas-2013
Christmas-2014
Christmas-2015
530,000 Traffic Violations
From Data → Insights
Data Management
Reverse Engineering the Database Schema(s)
Statistics & Data Analytics
From Data to Revealing Patterns
Detecting Outliers in Price
BIG Data Analytics
Coping with Volume Variety Velocity Veracity
Drinking from the Firehose
Statistics & Data Analytics
From Raw Data → Patterns & Insight
Average 4-Weeks Bank Discount Rates
Data Visualization: The Thinking Eye
From Data → Interactive Dashboards
BUAN 317: Data Visualization
From Data → Interactive Dashboards
Statistics & Data Analytics
From Data to Revealing Patterns
Tracing the Variability in Unit Price Across Years
Ethics of Data
When the Algorithms Know that your Mind made-up its mind before You made-up Your mind
When Algorithms Decide What You Pay
WEAPONS OF MATH DESTRUCTION
CATHY O'NEIL

Figure 5. The Poster