INFORMATION SYSTEMS EDUCATION JOURNAL

**Volume 19, No. 6**
December 2021
ISSN: 1545-679X

In this issue:



iscap
Information Systems & Computing Academic Professionals

The **Information Systems Education Journal** (ISEDJ) is a double-blind peer-reviewed academic journal published by **ISCAP** (Information Systems and Computing Academic Professionals). Publishing frequency is six times per year. The first year of publication was 2003.

ISEDJ is published online (https://isedj.org). Our sister publication, the Proceedings of EDSIGCON (https://proc.iscap.info) features all papers, panels, workshops, and presentations from the conference.

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the EDSIGCON conference. At that point papers are divided into award papers (top 15%), other journal papers (top 25%), unsettled papers, and non-journal papers. The unsettled papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in the ISEDJ journal. Currently the target acceptance rate for the journal is under 40%.

Information Systems Education Journal is pleased to be listed in the Cabell's Directory of Publishing Opportunities in Educational Technology and Library Science, in both the electronic and printed editions. Questions should be addressed to the editor at editor@isedj.org or the publisher at publisher@isedj.org. Special thanks to members of ISCAP/EDSIG who perform the editorial and review processes for ISEDJ.

# INFORMATION SYSTEMS EDUCATION JOURNAL

# Moving to Business Analytics: Re-Designing a Traditional Systems Analysis and Design Course

James J. Pomykalski
pomykalski@susqu.edu
Sigmund Weis School of Business
Susquehanna University
Selinsgrove, PA 17870, USA

## Abstract

Many traditional Information Systems (IS) programs are either redesigning current courses to incorporate business/data analytics or expanding curricular offerings to include business /data analytics; our IS program chose the former route to meet the demand (from employers) for business/data analytics. In that transition, a traditional systems analysis and design (SA&D) course was redesigned to focus on the procedures and skills needed to perform business/data analytics; this required changes to the focal topical areas. In 2017-18, curricular changes transitioned a traditionally focused database systems analysis and design course to incorporate business/data analytic skills. This paper focuses on the course redesign and the new business/data analytic topics, activities, and technologies; we highlight the important shifts created to the topical areas of this course. The paper emphasizes the rationale and student outcomes generated.

**Keywords:** Data/Business Analytics, Pedagogy, Systems Analysis and Design, Structured Data, SQL

### 1. INTRODUCTION

Over the past decade, Information Systems (IS) programs have been adjusting to meet the growing demand by industry to produce business analytics/data science professionals (Mills, Chudoba, & Olson, 2016). This has required a change/addition of courses that meet the need for the skills of these professionals (Radovilsky, Hegde, Acharya, & Uma, 2018). In addition, "there are a growing number of degree programs, specializations, and certificates in data science and data analytics at both the graduate and undergraduate levels" (Davenport & Patil, 2012; Dumbill, 2013; Aasheim, Williams, Rutner, & Gardiner, 2015).

Our undergraduate program decided to revamp the IS curriculum to address the change in skills. The first change occurred in existing courses in the business foundation. Our business foundation is unique, we did not have a traditional three credit hour Introduction to IS course, but instead all students complete three IS-related courses. The first course in this sequence was a formal Systems Analysis and Design course. This course culminated in students designing a simple database in Access.

The sophomore level Systems Analysis and Design (SA&D) course was re-focused and re-titled Data Collection and Modeling; this paper focuses on the changes made this SA&D course.

When designing this course change certain areas of the SA&D process were preserved due their importance in capturing the structured data used in business analytics; namely requirements analysis and data modeling; we discuss each area in a later section of the paper. In addition, new areas—data literacy, data curation/metadata, data quality and cleansing and SQL—were added.

This paper is structured in the following manner. In the next section, we describe the primary learning objectives of this newly designed course; including the areas of SA&D that remain due to their importance in the data/business analytics process. In section three, the importance of focusing on data literacy is examined. Section four continues the data literacy discussion with the introduction of the two major models that

form the basis for this newly developed course; the data life cycle model of Chisholm (2015) and the CRISP-DM model for data analytic problem solving (Chapman, Clinton, Kerber, Khabaza, Reinartz, Shearer & Wirth, 2000). In section five, we address the primary pedagogical approach used to introduce the skills-based modules of the course. In section six, we describe the major topics explored throughout the course. These topics include project management, data curation and the use of metadata, data cleansing, data modeling and data structuring, data analysis using the CRISP-DM model, and data retrieval using basic SQL commands. In the seventh section, we examine the major project that is undertaken (in teams) by the students to gain hands-on experience. In the final section, we state some concluding remarks by discussing the impact and major outcomes of the course for the students. In addition, we address some future work that needs to be undertaken.

## 2. COURSE LEARNING OBJECTIVES

The newly designed Data Collection and Modeling course required a complete reworking of the course learning objectives. For this redesign effort the work of L. Dee Fink (2013) was utilized. Fink's taxonomy of significant learning includes six critical areas to consider when designing a course for significant learning impact. These areas are: foundational knowledge, application, integration, human dimension, caring, and learning how to learn. Fink believes that a course that addresses each of these categories will create a lasting change in the learner that will be important to the learner's life. In particular, we, when designing this course, envisioned that change to be one of understanding the importance of leveraging data to solve business problems.

A primary goal of the course is to begin to introduce students to the basic skills necessary to become a data literate member of a data-driven society. The concept of data literacy is not new and not limited to data analytics. The primary definition of data literacy used for the course is given by Wolff et al. (2016). Data literacy is "the ability to ask and answer real-world questions from large and small data sets through an inquiry process, with consideration of ethical use of data" (Wolff, Gooch, Cavero Montaner, Rashid, & Kortuem, 2016, p. 23). Students "need to have at least a basic understanding of the concept of data, and they need to be able to understand and engage with data fitting their role and start talking the language of data" (Goodhardt, Lambers, & Madlener, 2018, p. 1).

The new course learning objectives are:

1. Identify the data literacy skills necessary for your given profession.
2. Utilize Microsoft Project, Microsoft Visio, Microsoft Access and basic SQL commands to answer business questions using data from a project domain.
3. Develop a multi-part report that describes how your data activities addressed the selected business questions
4. Employ and advance your written communication skills in conveying technical information.
5. Demonstrate and advance your teamwork skills in working through a data-intensive project.

In order to achieve the first learning objective, we strive to produce students that at least meet the minimum data literacy requirements of their chosen major field. This includes acquiring the basic foundational knowledge necessary to function productively in their first employment opportunity. The second learning objective deals with the application (skills) that the student will obtain through the course. This learning objective deals primarily with Fink's (2013) areas of providing foundational knowledge, the application of that knowledge and the ability to learn how to learn especially in dealing with new skill development. These areas were chosen based on the work of authors who had addressed the specific skills necessary for today's data scientist. (Dumbill, 2013; Goodhardt, et al., 2018; Mills, et al., 2016; Radovilsky, et al., 2018). The third learning objective deals with the team project assigned in the course. This project work will encompass the integration of the knowledge and skills through application, human dimension in understanding how to function as a productive team member, caring, about the work and the people on your team and learning how to learn category through the work of the project. The final two learning objectives are included to fulfill of both the University's Central Curriculum and the business school's AACSB learning objectives. These objectives are most concerned with integration of knowledge and the human dimension in terms of both written communication and teamwork skills.

In addition to these learning objectives, we identified a set of conceptual and constructive take-aways for each student; these take-aways are shown in Appendix 2. These take-aways highlight more specific skills learned in the course. The learning objectives were developed

using the current literature on analytics skills (Dumbill, 2013; Mills, et al., 2016; Radovilsky, et al., 2018) and in consultation with recent alumni in the business analytics and related fields. The skills were deemed appropriate for a first course in business analytics, taught to all business students, in the business foundation of the program.

## 3. DATA LITERACY

In 1992, Peter Drucker, in an article in the Wall Street Journal (Drucker, 2005)), explained that data users—executive or professional—need to be data-literate and decide what data to use, what to use it for and how to apply the data to solve problems. For organizations to prosper in a data rich environment, "*every organization, in every industry, should adopt a data-literate culture" (Smith, n.d.).*

The development of data literacy requires individuals to acquire skills in two areas. First, individuals must be able to understand and appropriately use tools and technologies for data analysis and decision-making. Second, individuals "need to learn to think critically and analyze data to choose the correct data and the suitable analytical and presentation methods for the situation" (Smith, n.d.).

This course module begins with two separate readings on data literacy to familiarize the students with the need for data literacy and a data culture, to provide a basic understanding of data literacy, and to show how data literacy is germane to the various majors with the business curriculum. The essence of this discussion is summarized by the figure in Appendix 1 which is taken from Goodhardt et al. (2018).

To further enhance data literacy skills assignments are used so -students practice using tools—Microsoft Project to create and update a project plan, Excel for data cleaning, Microsoft Access to structure data, and SQL to create queries—that allow students to fulfill the technology component of data literacy.

To improve the critical thinking and data analysis skills of the students, individual assignments in data cleaning, database development, and SQL statement development, and project assignments problem statement development, selection and matching of potential data sets for analysis, data cleaning, database development, and SQL statement creation.

Another aspect of data literacy that is covered in this module is the life cycle and use of data. Students are introduced to the two models which are described in more detail in the next section.

Through data life cycle (Chisholm, 2015), the students are presented with a model of how data is typically handled within the business environment; the model describes seven stages within the life cycle of data. The CRISP-DM model (Chapman, et al., 2000) is also introduced—although to lesser extent. The model illustrates the application of data within a problem-solving process. The concurrent examination of these models exhibits the important business processes undertaken within overlapping intervals of these models; as stated earlier the discussion on the overlap of the models is a work in progress.

One final discussion area in this module is the development of a data strategy as proposed by DalleMule and Davenport (2017). The article reveals the two intertwined sides of a data strategy: data offense and data defense. Data offense, which is the primary focus of their project work, deals with how an enterprise uses data to "support business objectives such as increasing revenue, profitability, and customer satisfaction. Data defense activities ensure compliance with regulations and minimizing downside risk. Data defense shows the ethical implications that are addressed in data literacy (Wolff, et al., 2016) and includes discussion of data governance issues.

Data literacy, the data life cycle model, the CRISP-DM model and data strategy are topics that will be revisited throughout the course as they set the foundation for the student's understanding of how data "should be" viewed and handled within and enterprise.

The primary assessment vehicle for the understanding of data literacy is a short reflective paper (two to four pages). The students are asked to identify the activities and topics that were covered in the course that contribute to building their data literacy skills foundation. The students are asked to put themselves, in their chosen major, along the scale shown in Appendix 1 and discuss the level of proficiency and their projected data role. The reflective assignment is used as the final individual learning assessment to gauge each student's depth of knowledge attained through the course and work on the team project.

## 4. DATA AND DATA-DRIVEN DECISION-MAKING

When examining data and data-driven decision-making it is often helpful to examine existing models that describe these processes. In particular the examination of the data life cycles within an organization and the data-driven decision-making process.

In examining the data life cycle, one particular model that is most useful is the data life cycle (DLC) model proposed by Chisholm (2015). Chisholm proposed a seven-stage model of the data life cycle: capture, maintenance, synthesis, usage, publication, archival, and purging. The data life cycle does not necessarily define all the specific processes involved in handling data, it does provide "high-level", i.e., strategic, understanding of the activities within that stage regarding enterprise data; the stages are shown in Figure 1 and discussed in more detail in Chisholm (2015) and Pomykalski (2020).

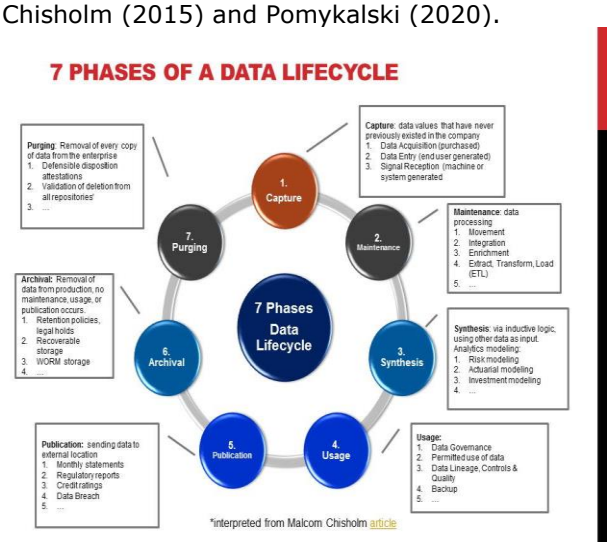



**Figure 1: Stages in the Data Life Cycle**

**On the other hand, the analytics problem-solving process is best described by CRISP-DM model which was developed through a partnership led by DaimlerChrysler. The CRoss-Industry Standard Process for Data Mining (CRISP-DM) methodology, developed in 1996, is "based on the practical, real-word experience of how people conduct data-mining projects" (Chapman, et al., 2000, p. 3). The CRISP-DM methodology (see Figure 2) consists of six stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment.**

These two models, blended together, form the basis of the course. The primary objective of which is to introduce students to the use of data to solve business questions. While a preliminary understanding of the integration of these models has been developed the formal description of the integration is considered future work.

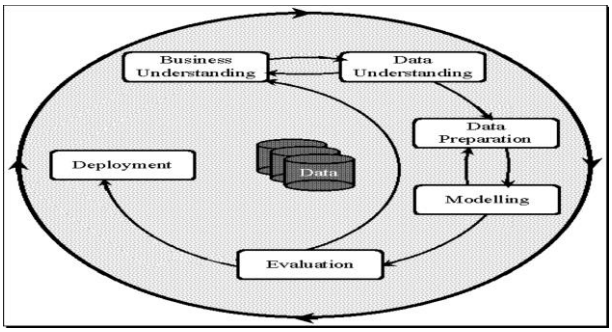



**Figure 2: Stages in CRISP-DM**

## 5. GENERAL PEDAGOGICAL APPROACH

As described in the previous sections, the course covers several topical areas relevant for understanding and sufficiently comprehending the business/data analytics field. In the sections below, we describe the topical areas and justify their inclusion in the course.

Before examining the course topics, a brief discussion of the pedagogical approach taken to enhance the learning of the students is essential. For the major, skill-based modules within the course (project management, data cleansing, data modeling, data structuring and data retrieval), the course is designed to follow a similar pattern in the student learning experience. Each module is designed and introduced using a similar pattern:

1. The module is introduced through a reading(s) with an in-class lecture and discussion to clarify the topic;
2. Students are then engaged in a low-stakes, in-class assignment as practice (discussion of the frequent, low-stakes (FLS) assignments is given below);
3. Students are then given an individual assignment—with a more complex problem as a graded element; finally
4. Students then perform the same assignment, within the context of their team project.

For the skill-based modules, the method of FLS graded assignments is used (Warnock, 2012). This is done, as stated by Warnock (2012) as a means for student to build confidence that they

are performing the task correctly and as a means to communicate and clarify any misunderstands or missteps apparent in the assignments. This is important because after the low-stakes assignments, the individual and team-based assignments carry larger weight in the overall grading of the course.

## 6. TOPICAL AREAS

Within the current course there are two topical areas that are largely addressed as general topic areas (data literacy and data curation/metadata) and five skill-based areas. The five skill-based areas all follow the pedagogical approach outlined in the section four. While data curation/metadata can be viewed as a skill-based area, the pedagogical approach is focused more on the work done as a project team and not individually.

**Project management**

Project management has been deemed as an important skill for all students within the business foundation and was made part of the original SA&D course; it has carried over into the business analytics curriculum. Knowledge of project management is vital for students to manage the significant number of course projects that they have each semester within their business foundation courses and for their work in their career.

Within this module, students gain basic knowledge through readings, chapters from Heagney (2016) and FSL assignment in developing both a work breakdown structure and the Gantt chart for short case study. In addition, the students create a work breakdown structure for their team project, based on the overall project description, and create a Gantt chart in Microsoft Project that must be updated throughout the semester; this is done to track the project deliverables.

This learning module and work on the project allows each student to not only understand the vocabulary and basic activities performed in the management of project work, but gain understanding as to how their individual contribution to that work enhances not only their individual learning but their team development.

**Data curation and metadata**

Data curation is a data management activity that plays an important role in the data life cycle. Data curation is "the work of organizing and managing a collection of datasets to meet the needs and interests of a specific groups of people" (Wells, 2019, p. np). Data curation is performed prior to the start of the data analysis process; it is the identification and structuring of the data pertinent to the data analysis process. The primary purpose, according to Wells (2019), of data curation is to make datasets easy to find, understand and access. Data curation activities are performed primarily by problem domain experts and those ensuring metadata quality.

Metadata, or "data about data", is vital in the use of data driven solutions in enterprises. Metadata contains information that not only describes the data, but includes the type, length, and previous use information; this is captured in many modern data catalogs (Villanova University, 2019). Data curation and metadata development are important in the metadata management process.

The students develop knowledge of the importance of metadata to the business problem-solving process primarily in the team project by examining data dictionaries to find pertinent data fields. They must differentiate between descriptive, structural, and administrative metadata and apply the three types of metadata to their project work by reading and utilizing a data dictionary that is provided (Villanova University, 2019).

**Data quality/data cleaning**

"In the case of business analytics, or the study of data and what information can be gained from the data, the 80/20 rule becomes: 80% of the time spent by a data scientist is on gathering, cleansing, and storing the data, while 20% of the time is spent on analyzing the data" (Snyder, 2019, p. 23). While gathering and storing are explicit parts of the data life cycle process, Snyder (2019) points out that data cleansing has often been granted a "lower status" in the data quality activities within business analytics. However, given the importance of data quality in the business analytics process it should have a more prominent role in the overall data life cycle.

In this course we dedicate an entire module to data quality and data cleansing. Through a reading and a class discussion the importance of high-quality data for decision-making is stressed. The students are then introduced to a number of basic Excel techniques that can be used in the data cleansing process with any size data sets. Excel is used, in this course and subsequent business foundation courses, as part of a business school wide plan to incrementally build the student's knowledge and proficiency with using Excel as a tool in solving business problems. Again, a low-stakes in-class assignment is used to assess their initial understanding.

A data cleansing activity highlights the work in this module and the students are expected to repeat the data cleansing process within their data analytics project; see section seven. The data cleansing activity—a specific tutorial in Monk, Brady, and Mendelssohn (2017)—focuses on finding missing data, inconsistent data (based on data type), duplicate records, formatting issues, incorrect entries based on spelling mistakes, and simple logic errors. The students create an error log which utilizes a different type of entry for each particular error type. A class discussion that focuses on the remedies for each type of error follows the work on the class activity. An individual assignment with a larger dataset is given to determine their skill development in Excel and understanding of the data cleansing process; a key output is the development of the well documented error log.

**Data modeling**

Within a data analytics project, the data model is most critical model for the understanding of the structured data necessary for solving the business problem. Data modeling is a "technique for organizing and documenting a system's data. A method for the representation of organizational data" (Whitten & Bentley, 2008, p. 270).

The primary model that is used to capture the entities and the business relationships between the entities is an entity-relationship diagram (ERD). The entity-relationship diagram is "a data model utilizing several notations to depict data in terms of the entities and relationships described by that data" (Whitten & Bentley, 2008, p. 271). The students use traditional SA&D readings to understand and develop an ERD through the in class exercise. Coverage of all the elements of the ERD—entities, attributes, relationships and cardinality—are addressed. The individual assignment used is a carryover from the SA&D course in which the students develop both the model and a business memo to describe the major elements of the ERD (Pomykalski, 2006).

The students, within the context of their project also design and develop a data model specific to the data selected to address the chosen business problems and create a new memo, using the same structure, to discuss the ERD development process.

**Data structuring/analysis**

Having the ability to structure data in a simple database application is important to all students regardless of the particular business discipline. In this course, Microsoft Access is used to structure the project data.

The students are introduced to Microsoft Access though a set of in-class and individual assignments taken from the Monk, et al. (2017). These assignments give the students the basic skills necessary to create relational database tables, forms, queries and reports. The students learn to move Excel data into tables, and then perform basic data analysis functions.

One of the fundamental Excel skills introduced in this module is the use of Excel pivot tables as a means for further exploration and analysis of the data. This again is covered through tutorial in Monk, et al. (2017). The basic skills developed in this module also followed the prescribed methodology laid in section five.

The CRISP-DM model is also examined in this module of the course. The students examine the particular activities that occur in the later stages of the model dealing with data preparation, modeling, and evaluation.

**Data retrieval—SQL**

The final learning module of the course is focused on the development of an understanding of basic SQL commands. This is the first introduction to SQL commands that students receive, and it was implemented in this course due to feedback from alumni that were obtaining first jobs in a variety of different majors. The knowledge of SQL to find and extract data for their new teams was seen as an enhancement to the role they could fill directly after graduation.

This three-week module includes introduction to a basic SELECT statements and the ability to perform many simple extraction and summarization functions. In particular, the module covers the initial half of the Forta (2013) text on SQL. In particular, we introduce commands for: data retrieval, sorting, filtering, manipulation, summarizing, grouping and table joins. This module does not cover the development of a database structure using SQL.

Again, this skills-based module follows the same pedagogical structure described in section five, however, since each new function needs to practice, a series of in-class assignments, each with newly understood functions, are provided in a low-stakes assignment format.

## 7. PROJECT WORK

One of the primary learning assessments in the course is the work done through the team project. The team project is a full semester project which has multiple intermediary deliverables that help

comprise the overall grade for this component. Teams of three to four students work through the various activities that mimic the skill-based modules in the course. The project idea and the necessary business questions and data are taken from the Teradata Analytics challenge. Each of the two projects that are currently available have been used without significant overlap due to the richness of the business questions and the vast amount of data that is available to the students. The project, the description of which is given as Appendix 3, includes a number of activities for the project team to undertake.

The students begin the project by creating an overall project management plan for the undertaking of the project. The first activity is to review and select two business questions, these are provided as part of the Teradata Analytics Challenge; the students are encouraged to review the data that is associated with the particular business questions. An in-class review and discussion of the data sets (along with the corresponding data dictionary) is undertaken.

The first project deliverable includes (1) a written description of the selection of the business questions and their significance, (2) a work breakdown structure including the rationale for many of the project management decisions, and (3) a Gantt chart that schedules each of the particular activities on a semester long timeline.

In part two of the project, the students turn their attention to closer examination of the data sets that they have chosen for their particular business questions. The students are asked to explore and cleanse the data sets. They must provide a detailed account of the activities that they undertook in the examination and data cleansing work. The final stage for this work is to structure the selected data fields into different data sheets within Excel to facilitate the import of the data to a Microsoft Access database structure.

In part three, the students are asked to model the data and create an ERD that shows the data in clear logical, well organized structure. The students must define the data entities, attributes, relationships, and cardinality. The deliverable for this stage is the memo and ERD model for their project data similar to the deliverable for the individual assignment (Pomykalski, 2006).

The final activity that is undertaken within the project is the development of the SQL queries needed to examine the data and answer the business questions. The final deliverable—the final written report—incorporates the discussion on the development of the SQL queries as well as summarizing all of the work undertaken in the project.

The project gives the students the experience of providing a final client driven report that summarizes all of their activities over the course of the semester. The students also gain experience creating intermediate, progress reports that can be provided to the client and the project team as a means for documenting the work. The students are allowed to utilize the previous (corrected) reports to create the final project deliverable.

## 8. CONCLUSIONS/FUTURE WORK

The transition of a traditional IS curriculum to a business analytics curriculum requires a substantial amount of rework and new thinking as to the topics and pedagogies used. However, not all of the traditional IS concepts need to be removed. As we have shown, both project management and data modeling have been revised to focus on the data aspects of the analytics process.

In this paper, we examined and described this transitional work—we continue to tinker with many areas—and this new structure seems to serve the students well. We have had a number of students go into both internships and employment opportunities with these new skills and that feedback from these students has been positive.

Analytics is a field that continues to evolve and therefore the evolution of this course and subsequent courses in the business foundation will require close scrutiny and monitoring in the coming years. The integration of the data life cycle and the CRISP-DM models are future projects so that students can investigate the complexities of the business/data analytics process.

We are pleased with the progress we have made thus far and we believe that these courses will better serve our students are they undertake new employment opportunities as both interns and permanent employees.

## 9. ACKNOWLEDGEMENTS

The author would like to acknowledge the contributions of the other faculty, both current and recently retired, in the development of this course redesign and the establishment of a recently created major in business data analytics. As new faculty have recently joined our ranks we look forward to their contributions to the continual development of this course and expect

to be able to provide updates of these development efforts in the near future. In addition, we appreciate the comments and feedback of the reviewers.

## 10. REFERENCES

Aasheim, C. L., Williams, S., Rutner, P., & Gardiner, A. (2015). Data Analytics vs. Data Science: A Study of Similarities and Differences in Undergraduate Programs Based on Course Descriptions. *Journal of Information Systems Education, 26*(2), 102-115.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide.* Somers, NY: IBM. Retrieved June 2017, from http://www.crisp-dm.org/CRISPWP-0800.pdf

Chisholm, M. (2015, July 9). Seven Phases of a Data Life Cycle. *Information Management*, p. n.p. Retrieved May 28, 2018, from https://www.information-management.com/news/7-phases-of-a-data-life-cycle

DalleMule, L., & Davenport, T. (2017, May-June). What's Your Data Strategy? *Harvard Busines Review*, 112-121.

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

Drucker, P. (2005, November 15). Be Data Literate—Know What to Know. *Wall Street Journal.*

Dumbill, E. L. (2013). Educating the Next Generation of Data Scientists. *Big Data, 1*(1), 21-27.

Dykes, B. (2019). *Effective Data Storytelling: How to Drive Change with Data, Narrative and Visuals .* New York, NY: John Wiley & Sons, Inc.

Fink, L. D. (2013). *Creating Significant Learning Experiences: An Intergrated Approach to Desgining College Courses.* San Francisco, CA: Jossey-Bass.

Forta, B. (2013). *SAMS Teach Yourelf SQL in 10 Minutes* (Fourth ed.). Indianapolis, IN: Pearson Education, Inc.

Goodhardt, B., Lambers, E., & Madlener, J. (2018). *How to Become Data Literate and Support a Data- Driven Culture.* Netherlands: KPMG IT Advisory. Retrieved from https://www.compact.nl/en/articles/how-to-become-data-literate-and-support-a-data-driven-culture/#ref

Heagney, J. (2016). *Fundamentals of Project Management* (Fifth ed.). New York, NY.: AMACOM.

Mills, R. J., Chudoba, K. M., & Olson, D. H. (2016). IS Programs Responding to Industry Demands for Data Scientists: A Comparison between 2011 – 2016. (L. Freeman, Ed.) *Journal of Information Systems Education, 27*. Retrieved May 9, 2020

Monk, E. F., Brady, J. A., & Mendelsohn, E. I. (2017). *Problem-Solving Cases in Microsoft Access & Excel* (15th ed.). Boston, MA, USA: Cengage Learning.

Nussbaumer Knaflic, C. (2015). *Storytelling with Data: A Data Visualization Gudie for Business Professioanals.* New York, NY: John Wiley & Sons, Inc.

Pomykalski, J. J. (2006). Teaching Tip: Interleaving Modeling and Writing Activities in Systems Analysis and Design. *Journal of Information Systems Education, 17*(3), 249-252.

Pomykalski, J. J. (2020, March). Privacy and the Data Life Cycle. *Journal of Applied Information Systems Research (JISAR), 13*(1), 40-48.

Radovilsky, Z., Hegde, V., Acharya, A., & Uma, U. (2018, March). Skills requiremens for Business Data Analytics and Data Science Jobs: A Comparative Analysis. *Journal of Supply Chain and Operations Management, 16*(1). Retrieved May 9, 2020

Smith, A.M. (2020). Fundamental of Data Literacy. Retrieved from http://www.ewsoultions.com/fundamentals-of-data-literacy.

Snyder, J. (2019, December). Data Cleansing: An Ommission from Data Analytics Coursework. *Information Systems Education Journal, 17*(6), 22-29. Retrieved from http://isedj.org/2019-17/

Villanova University. (2019, October 24). Metadata and Its Importance in a Data Driven World. Philadelphia, PA, USA. Retrieved from https://www.villanovau.com/resources/bi/metadata-importance-in-data-driven-world/

Warnock, S. (2012). Frequent, Low-Stakes Grading: Assessment for Communication, Confdence. *Online Classroom, 12*(3), pp. 5,7. Retrieved from https://www.facultyfocus.com/articles/educational-assessment/frequent-low-stakes-grading-assessment-for-communication-confidence/

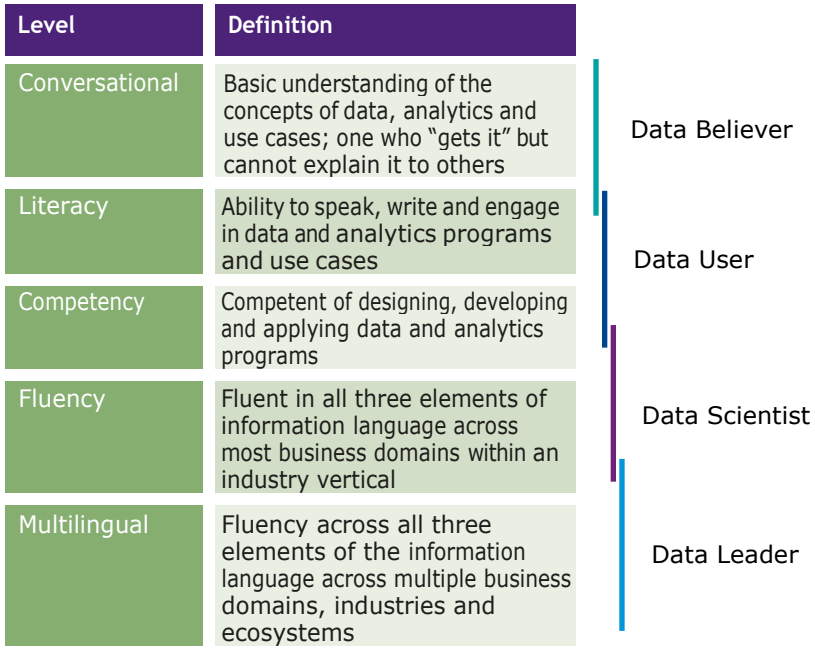
Wells, D. (2019, May 30). *What is Data Curation?* Retrieved from Alation.com: https://www.alation.com/blog/what-is-data-curation/

Whitten, J., & Bentley, L. (2008). *Introduction to Systems Analysis and Design* (1st ed.). New York, NY: McGraw-Hill Irwin.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *4th international conference on the practical applications of knowledge discovery and data mining*, (pp. 29-39). Retrieved June 2017, from https://pdfs.semanticscholar.org/48b9/293cfd4297f855867ca278f7069abc6a9c24.pdf

Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an Understanding of Data Literacy for a Data-driven Society. *The Journal of Community Informatics, 12*(3), 9-26.

**APPENDIX 1: Relation between Data Literacy Levels of Proficiency and Data Roles**
(Goodhardt, Lambers, & Madlener, 2018)




**APPENDIX 2: Course Take-Aways**

| **Conceptual Take-Aways** | **Constructive Take-Aways** |
|---|---|
| Develop a data literacy mindset | Use Microsoft Project to manage a data analysis project |
| Examine data through a forma data life cycle | Compile a large dataset, clean and structure the data |
| Relate the roles and responsibilities involved in data management | Develop/enhance your Microsoft Excel skills |
| | Load data into a Microsoft Access database |
| | Write SQL queries to produce answers to business questions |
| | Produce a formal written report explaining the results of your queries in addressing the business problems. |

**APPENDIX 3: Course Project Description**

# Analytics Project

**Objective:** The goal of this project is to create a data analysis report. To be able to do this, you will chose two business questions, identify, organize, clean and normalize data relevant to the business questions, load the data into a database, develop SQL queries, create charts, and detail your analysis in a written report. You will choose your own teams of 2-3 people.

Project: The general description of this project is given at:
https://www.teradatauniversitynetwork.com/Community/Student-Competitions/2019/2019-Data-Challenge

The business questions for this project reside in a file on Blackboard; under the Analytics Project folder. There are five major categories of questions listed: Client Services, Volunteer Programs, Development, Employer Partnerships and Opportunities, and Serving Spouses Program. Your team is to pick a category and within that category, they will select at least two questions to investigate further.

**The General Task (From Teradata Challenge submission template)**:

1. State and describe your understanding of the business question(s) you are addressing.
2. Choose the datasets provided by the Hiring Heroes USA organization from Teradata that are most appropriate to address your business question(s). Similar to an SQL query:
    a. Identify the field names and values that you need to address each business question.
    b. Identify the data sets that have related fields to those that you need to address the question.
    c. Clean the datasets to make them consistent in form and format.
3. Provide an Entity Relationship diagram of your relevant data.
4. Design SQL queries to address each question from the cleaned and structured data.
5. If you cannot run your queries, describe what you expect to get (estimate your results)

**The Specific Deliverables**

**I.** **Team Sign-Up**: This is the simplest of the deliverables for this project. Each person enrolled in the course must get into a team with one or two other people currently enrolled in the course; preferably in the same section. Once the team members have agreed to become a team, one member of the team must send me an EMail that lists the members of the team.

**Due Date**: Friday, September 20, 2019

**II.** **Plan**: There are two deliverables associated with this stage.

a. A document describing the selection of the category and the two business questions your team will address and your rationale for your choices. You should clearly state the questions and then describe, in your own words, your understanding of those questions.

b. A Microsoft Project file containing the work breakdown structure of tasks, precedence of tasks, who completed each lowest-level task (resource) and duration (time) to complete each task. In addition, you need to provide a document that describes your decision making process in completing this project management task.

**Tentative Due Date**: Wednesday, October 9, 2019

**III. Microsoft Excel File(s):** The data with each worksheet cleaned and normalized as a table suitable for importing to Microsoft Access. There will be two deliverables associated with this activity.

a. Your team will produce a **progress report**, which will outline the selection of the data sets, field names and values that you have identified, and the activities you have undertaken, to date, on the files and the work that still needs to be completed.

**Tentative Due Date** of Friday, October 18, 2019.

b. Upon completion of the work with the data files your team will create a document which describes all of the activities you performed on the data sets to put the data into a consistent form and format.

**Tentative Due Date** of Monday, November 4, 2019.

**IV. Microsoft Visio ERD:** There will be two deliverables associated with this stage.

a. Your team will produce a business memo that explains the entities, attributes, including primary keys, relationships and cardinality depicted in your ERD. Make sure that the relations are normalized to at least Third Normal Form (resolve all many-to-many relationships with associative entities).

b. Create an entity relationship diagram using Visio. For full credit, make sure cardinalities are valid and display with crow's feet for the "many" side.

**Tentative Due Date**: Wednesday, November 20, 2019

**V. Microsoft Word Report**: Report summarizing what was done that includes:

a. The questions being addressed and your current understanding of these questions.
b. A corrected, updated version of your Microsoft Excel file activity report.
c. A report that describes the SQL queries that your team created, the rationale for each of those queries and the results (or expected results) from those queries. You are to include a discussion on how your team expected the queries to address the particular business questions.

**Tentative Due Date**: Monday, December 9, 2019

**VI. Individual Files**: Send both as attachments to an e-mail.

a. Microsoft Excel Peer Evaluation: Download the file from Blackboard. **Be aware**, awarding all members all 4s for all criteria will be worth 0 points.
b. Microsoft Word Reflection: The personal statement should include a half-page (in length) to one page document reflecting on the experience of doing this project (what parts of the project were your responsibility, what you learned, what parts of the project would you improve, …).

**Tentative Due Date**: Wednesday, December 11, 2019

**Grading**:

| Requirement: | | Poss. Points: |
|---|---|---|
| I. | Team Sign-Up | 2 |
| II. | Plan (MS Project File) | 17.5 |
| III. | Cleaned and normalized Data [MS Excel File(s)]—Including both reports | 24.5 |
| IV. | ERD (MS Visio File) | 15 |
| V. | Analysis Report with Charts/Graphs (MS Word File) | 26 |
| TEAM GRADE | | **85** |
| VI. | Individual files (to be e-mailed) | |
| | 1 Personal Statement (MS Word) | |
| | 2 Peer Evaluation (MS Excel) | |
| INDIVIDUAL GRADE | | **15** |

Please note: that while each team member will likely receive the same score for the team based portion of this project, individuals grades for the team component are adjustable (mostly downward) based on an individual's contribution to the team deliverables. The impetus for this point deduction will come from discussions with team members and the two evaluation documents shown above.