***DRAFT***

IST 782 Applied Data Science

Portfolio Milestone Report

Michael Farrell

SUID: 245166296

SU Email: [mfarrell@syr.edu](mailto:mfarrell@syr.edu)

*“The goal is to turn data into information and information into insight”*

*- Carly Fiorina, Former CEO of Hewlett-Packard*

Contents

[Resume 1](#_Toc103009760)

[Introduction 7](#_Toc103009761)

[Program Learning Goals 8](#_Toc103009762)

[Describe a broad overview of the major practice areas in data science. 10](#_Toc103009763)

[IST 652 Scripting for Data Analysis 12](#_Toc103009764)

[IST 707 Data Analytics 15](#_Toc103009765)

[IST 718 Big Data Analytics 20](#_Toc103009766)

[SCM 651 Business Analytics 24](#_Toc103009767)

[Conclusion 25](#_Toc103009768)

[Other courses are included at the end for safekeeping and reference but will not be included in the final version. 26](#_Toc103009769)

[IST 659 Database Administration Concepts and Database Management 26](#_Toc103009770)

[IST 664 Natural Language Processing 26](#_Toc103009771)

[IST 687 Introduction to Data Science 28](#_Toc103009772)

[IST 719 Information Visualization 28](#_Toc103009773)

[IST 722 Data Warehouse 29](#_Toc103009774)

[ACC 652 Accounting Analytics 31](#_Toc103009775)

[FIN 654 Financial Analytics 31](#_Toc103009776)

[MBC 638 Data Analysis and Decision Making 32](#_Toc103009777)

# Resume

**Core Competencies**

|  |  |  |
| --- | --- | --- |
| Data Analytics | Leadership/Management | Problem Solving |
| Team Building | Strategic Planning | Risk Management |
| Training |  |  |

**Leadership/Management**

* Led 35+ staff in the Charitable Bingo Operations Division in regulating and auditing charitable bingo throughout Texas. Result: Greatly improved relations and education of stakeholders.
* Led 35+ staff department in the facility maintenance and infrastructure support for the Austin area Texas Workforce Commission. Result: Well maintained, safe and secure facilities.
* Represented the U.S. Government at international fora and functions. Established and managed relationships within the Mexican Government to meet U.S. Government goals. Result: Unprecedented cooperation between the U.S. and international partners on maritime issues.
* Over 27 years of experience leading, supervising, managing, and training diverse organizations. Result: Met/exceeded readiness and organizational requirements.
* Commanded large Coast Guard operational unit, led headquarters division, and operational unit departments and divisions. Result: Received awards for outstanding performance.
* Led hundreds of hazardous operational aviation missions. Result: Numerous lives saved under adverse conditions.

**Problem Solving/Team Building/Risk Management**

* Developed project and the subsequent tasks needed to implement the changes to state-wide charitable bingo as a result of the 86th Legislature law-making. The short time and significant changes made this a high-risk project. Result: Increased compliance, revised numerous forms, educated stakeholders and decreased resubmissions from stakeholders.
* Developed and nurtured cooperative international information-sharing relationships. Result: Landmark cooperation on counternarcotics interdictions with Mexico supporting US goals and initiatives, accomplished what was unattainable for the past 15 years.
* As Group Commander, developed, planned, coordinated and convinced local responders and commercial operators the benefits of a regional mass casualty exercise. Result: Successful mass casualty exercise with hundreds of diverse participants.
* As a pilot and military officer, routinely weighed risk versus a gain of missions assigned. As Group Commander, evaluated assigned rescue and law enforcement missions and determined how to manage/mitigate risk to accomplish the goal. Result: Unit flew over 5400 accident-free flight hours, hundred plus boardings, Columbia River safe for commerce.

**Strategic Planning**

* As Director of Charitable Bingo, planned and implemented division reorganization to better meet internal and external needs.
* Led Texas Ranger Boat planning, from the development of standard operating procedures to refinement for tactical use. Result: Rangers receiving training from US Coast Guard on small boats.
* As Group Commander and Air Station Executive Officer, developed, reviewed and assessed major command first-ever business plan with metrics. Result: Increase in efficiency and effectiveness in mission accomplishment while training subordinates to be better leaders and planners.
* As Chief, Migrant Interdiction Division, developed, reviewed and assessed Coast Guard-wide policy and measures of effectiveness for Coast Guard migrant interdiction program. Result: Metrics that could be used to measure the effectiveness of the Coast Guard migrant interdiction program.

**Work History**

**Management Analyst, Texas Depart of Family Protective Services, Adult Protective Services (5/2020 – Present)**

Analyzes work problems identified by data analysis and makes recommendations for program improvement. Uses extensive datasets to develop and monitor solutions to organizational issues as directed by the Performance Manager. Contributes to the successful management of the APS program through the analysis, interpretation and recommended implementation of policy and data. Creates monthly, quarterly and ad hoc reports for upper management use, decision-making, and/or forwarding to state government elected officials. Tools include Excel (advanced skills), Tableau, R, python and Adobe.

**Adjunct Professor, St. Edward's University, Austin, TX (1/2015 – present)**

Instruct core required computer science classes to freshmen through seniors. Responsible for curriculum development and refinement. Topics include the use of the computer as a tool for problem-solving and quantitative reasoning in a variety of disciplines; introduction to computer programming, Internet resources and tools, including browsers, searching, and web page design and construction; network basics; personal online safety and social issues involved with the use and misuse of computers; an introduction to web-oriented programming; the history, development and application of information-processing systems, including an overview of the need for and roles of computer information systems in business. Excel, Access, HTML, and Python are used to develop student projects.

**Director, Charitable Bingo Operations Division, Texas Lottery Commission, Austin, TX (7/2018 – 2/2020)**

Led and managed the Charitable Bingo Operations Division of the Texas Lottery Commission. Responsible for all aspects of licensed bingo activities and assisting charities in their navigation of the myriad requirements to conduct charitable bingo fairly and legally in Texas, which allows charities to fund their charitable purposes. The Charitable Bingo Operations Division provides the strict control and close supervision of bingo conducted in this state so that bingo is legally and fairly. This is done through education and audits. Our goal is to achieve voluntary compliance and assist licensees through training and education. The division consists of the Licensing, Accounting, Compliance, Education, and Audit departments.

Experience in recruiting, employing and managing diverse staff to effectively and efficiently assist the organization. Establishes and maintains effective relationships with all levels of staff, from Lottery Commissioners to junior support staff, along with external stakeholders and other government entities.

As the director, the face of the Charitable Bingo Operations Division. Travel extensively throughout Texas to visit with stakeholders to observe bingo operations and hear and address concerns. The division ensures that bingo games are conducted fairly, and the proceeds are used for authorized purposes. This is done through education, enforcement of applicable laws and rules, and auditing. The division consists of the Licensing and Accounting Department, Compliance and Education Department, and Audit Department.

Worked directly for the commissioners. Responsibilities include strategic, operational and tactical planning, developing strategies, creation of measures and metrics for division evaluation and feedback, using results to implement the division, and Texas Lottery Commission's goals and objectives.

**Director, Risk and Security Management (RSM), Texas Workforce Commission, Austin, TX (9/2014 – 7/2018)**

Provided TWC customers and employees a safe, secure and professional business environment through a program of robust Risk Management (including state-wide security and incident response), Business Continuity, Security and Emergency Response Plans and resources; Safety Services, privacy governance program, budget creation (operations and personnel), management, tracking, and implementation.

Responsibilities included strategic, operational and tactical planning, developing strategies, staff training, creation of measures and metrics for program(s) evaluation and feedback, using results to implement department, division, and Texas Workforce Commission's goals and objectives.

Required to analyze situations, evaluate risk and priorities, and make decisions under stressful, time-limited situations. Decisions made are followed up with written reports documenting situations and decisions.

Additional responsibilities included maintaining the security of 100+ offices worth over $31MM located throughout the state. Oversaw the RSM department. Led a complex privacy governance program dealing with personally identifiable information (PII), federal tax information (FTI), education information (FERPA), personal health information (PHI), and agency sensitive information. Recruited, employed and managed a diverse staff to effectively and efficiently assist the organization. Established and maintained effective relationships with all levels of staff, from Workforce Commissioners to junior support staff, along with external stakeholders and other government entities. Responsible for contract management oversight of over 300 data exchange contracts.

Effectively used all tools available, including process improvement/analysis tools, such as Rapid Process Improvement and Throughput Operating System.

**Adjunct Professor, University of Phoenix, Austin, TX (1/2012 – 1/2021)**

Instruct and facilitate courses in mathematics, statistics, and those courses related to rhetoric, critical thinking and creative problem-solving. Topics include focusing on the business uses of mathematics and statistics, developing the critical and creative thinking skills necessary to analyze and solve problems, make decisions, implement strategies, and formulate well-supported points of view on key academic, social, and professional issues.

**Lead Faculty, University of Phoenix, TX (9/2016 – 8/2017)**

Taught, assessed faculty, performed classroom performance reviews, mentoring, content interviews, meetings, faculty certification and training, consultation, faculty governance meetings, and student advisement. Assigned content areas in the College of Humanities and Sciences: Mathematics, Communication, English.

**Director, Infrastructure Services & Risk Management, Texas Workforce Commission, Austin, TX (7/2011 – 8/2014)**

Responsibilities included strategic, operational and tactical planning, developing strategies, staff training, creation of measures and metrics for program evaluation and feedback, using results to implement department, division, and Texas Workforce Commission's goals and objectives.

Tools included practicing Rapid Process Improvement, analyzing Throughput Operating System, and other process improvement and analysis tools.

Provided TWC customers and employees a safe, secure and professional business environment through a program of preventive and restorative maintenance; facilities contract management; robust Business Continuity and Emergency Response Plans; and voice telecommunications systems that meet customer and employee needs. Responsible for Facility Planning, Contract Management, and Maintenance; Headquarters Services; Infrastructure Support & Safety Services, Complex Security, Safety Officer, Safety & Health Committee, Risk Management, Business Continuity Program Resources, budget creation (operations and personnel), management, tracking, and implementation.

Required to analyze situations, evaluate risk and priorities, and make decisions under stressful, time-limited situations. Decisions made are followed up with written reports documenting situations and decisions. Developed, maintained and was responsible for 400K square feet of multiuse buildings worth over $31MM and approximately $1MM budget. Oversees 19 contracts ranging from elevators to fire suppression systems to security.

Experienced in recruiting, employing and managing diverse staff to effectively and efficiently assist the organization. Establishes and maintains effective relationships with all levels of staff, from Workforce Commissioners to junior support staff, along with external stakeholders and other government entities.

**Senior Analyst/Consultant, Abrams Learning & Information Systems, Inc., Austin, TX (4/2011 – 12/2015)**

Senior-level consultant on Emergency Management, Homeland Security (Border and Port), Strategic and Operational Planning, Transformation Management, Training, and Process Improvement. Provides accurate and timely high-level analyses, development, review, and assessment of current and proposed strategies and options using extensive leadership, strategic planning, emergency management, performance measurement and analysis, process improvement, maritime and law enforcement experience. This includes mission definition and refinement, gap analysis, policy guidance, training requirements, and strategic, operational and tactical manual development. FEMA qualified to instruct/conduct/facilitate Incident Management and Continuity of Operations courses.

**Senior Analyst, Abrams Learning & Information Systems, Inc., Austin, TX (01/2010 – 12/2010)**

Requirements, Concepts, and Plans Officer, in the Texas Rangers’ Border Security Operations Center (BSOC). Member of high-performance teams of senior-level advisers responsible for providing accurate and timely high-level analyses, development, review, and assessment of current and proposed strategies and options, while providing on-site support as required. Spearheaded DPS Strategic Plan, worked side-by-side with DPS counterparts in the drafting and revision of the Strategic Plan, included measures and metrics used to assess DPS efforts and success, as well as for the biannual budgeting process. Applies extensive maritime and law enforcement experience to assist the Texas Rangers in the implementation and operational use of their newly acquired small boat. Involved in all aspects of boat operations, from mission definition and refinement, gap analysis, policy guidance, training requirements, and manual development. Results: Delivered completely revised finished DPS Strategic Plan ahead of schedule.

**US Coast Guard Attaché, US Embassy Mexico City, Mexico – DAO (05/2007 – 06/2009)**

Senior US Coast Guard representative in Mexico. Provided Ambassador and other embassy staff insight, military-political observations, developments, and trends into Mexican military and civilian affairs. Coordinated and facilitated operations and interaction between U.S. Coast Guard, Department of Defense and Embassy sections with military and civilian departments of the Government of Mexico regarding counter-narcotics, illegal immigration, port security, infrastructure protection, search and rescue, and environmental protection. Routinely worked with western hemisphere military representatives. Required developing professional relationships with all domestic and foreign stakeholders. Extensive travel throughout Mexico. Results: Unparalleled cooperation between the US Coast Guard and the Mexican Navy, Secretariat of Communications and Transportation, and Secretariat of Foreign Relations.

**Group Commander, US Coast Guard Group / Air Station Astoria, OR (06/2003 - 07/2006)**

Led and executed all U.S. Coast Guard missions in the area of responsibility along the Washington and Oregon coasts. This included all aspects of aviation and maritime operations, including mission training, planning (long-term, emergency, and exercise), logistics, medical, infrastructure, budget and personnel support for over 300 Coastguardsmen and families at a multi-unit Coast Guard command in the Pacific Northwest ($3+MM budget, $.5B infrastructure, helicopters, and boats). Ensured the safety, security, environmental protection and efficient flow of commerce affecting a multi-billion-dollar maritime transportation system, over 7 million ton grain export, 1000+ foreign vessel arrivals, 40+ cruise passengers, 290K recreational boats, 2K commercial fishing vessels, and numerous Maritime Transportation Safety Act facilities. Support infrastructure included a medical clinic and 160 housing units. The Group also provided operational, logistical, and administrative support to two Astoria-based medium endurance cutters, a 225’ buoy tender also home-ported in Astoria, the National Motor Lifeboat School, the Advanced Helicopter Rescue School, and local National Guard units (800+ personnel). Coordinated operations and exercises with local, county and state authorities. Helicopter Pilot, Aircraft Commander HH-60J Jayhawk. Results: Successful accomplishment of Coast Guard missions, lives saved. First mass casualty exercise involving local, county, and state responders and volunteers from Oregon and Washington.

**Executive Officer, US Coast Guard Air Station Barbers Point, HI (6/1999 - 07/2002)**

Managed the day-to-day operations, administration, and multi-million dollar budget of a large US Coast Guard Air Station. Provided support for over 200 Coastguardsmen and their families. Air station assigned to protect 12.2 million square miles of open ocean, atolls, and island nations with long-range patrol and logistical support capabilities, as well as quick and versatile  
search and rescue response. Additionally responsible for a military dining facility and all-hands club/recreation facility. Helicopter pilot, Aircraft Commander HH-65A Dolphin. Results: Created the first unit business plan. Successful accomplishment of Coast Guard missions, lives saved. The unit received a readiness award. All hands club profitable providing financial support to morale, welfare, and recreational events for air station personnel and guests.

**Adjunct Professor, Hawai’i Pacific University (1/2000 – 5/2002)**

Instruct and facilitate courses in mathematics, computer programming in the C language, and general sciences for undergraduate students. Topics supported students’ learning on the uses of mathematics and statistics, developing the critical and creative thinking skills necessary to analyze and solve problems, make decisions, and formulate well-supported points of view on key academic, social, and professional issues.

# Introduction

I have had a long and varied career in both the military and civilian sectors. During my 32 years in the U.S. Coast Guard and 12 years working in the public sector, I have had numerous positions where I have used data and analytics to successfully accomplish mission and agency goals. For example, creating search areas uses predictive analytics using factors such as wind direction, current direction and search object size. These techniques were and are standardized. When I first started my career, they were not automated, and we accomplished the task using charts and vector math (although they did not call it that).

Further in my career, I was part of teams that created measures and metrics to assist in determining strategic goals and predicting requirements to maintain operational status – 3 helicopters ensure a 98% probability that there will be at least 1 operational; and process improvement – how do we ensure that we can get airborne in the 30-minute window when it takes 35+ minutes normally to get airborne? These skills have also been used in my roles as an emergency response manager and continuity of operations planner.

In the civilian side, I have used basic measures for facility maintenance, work team assignments and metrics to meet the government organizations strategic plan measures.

I have also used measures and taught the basics in my role as an adjunct professor at three different universities.

My personal learning goals were to learn to better use the tools available to better understand data and better equip me in my current role as a management analyst. I have learned and/or been exposed to a variety of process analyses and wanted to have a more in-depth understanding of these techniques. I also wanted to expand my knowledge. During this course of study, I have been able to directly transfer the skills learned for use in my “day job.”

In my current position as a Management Analyst at the Texas Department of Family and Protective Services (Agency), I found myself wanting to do more than an Excel spreadsheet – pie charts and bar charts. I wanted to help my agency look forward strategically while being able to look at current performance through a more developed statistical lens using more advanced tools such as regression analysis, forecasting, and graphic presentations. For many of the project and assignments, I used “real-world” data from my agency to answer meet course requirements and answer an agency question, create graphics and/or do analysis.

Additionally, and to be honest, I also wanted to use my GI Bill in an area that interests me, data science. I already have a BS in Mathematics and MA in computer science, so Applied Data Science seemed like a natural fit. My agency would allow me time off to study, so it appeared to be a win-win.

# Program Learning Goals

Offered jointly by the School of Information Studies and the Martin J. Whitman School of Management, the Master of Applied Data Science degree program is designed to be a professional program of study, with a strong emphasis on the applications of data science to enterprise operations and processes, particularly in the areas of data capture, management, analysis and communication for decision making.

Demonstrate Achievement of Program Learning Goals:

The following details the preparation and assessment of the Project Portfolio for the MS in Applied Data Science. The overall goal of the Project Portfolio is to demonstrate to the faculty expert overseeing the portfolio that the student is able to:

1. Describe a broad overview of the major practice areas in data science.
2. Collect and organize data.
3. Identify patterns in data via visualization, statistical analysis, and data mining.
4. Develop alternative strategies based on the data.
5. Develop a plan of action to implement the business decisions derived from the analyses.
6. Demonstrate communication skills regarding data and its analysis for managers, IT professionals, programmers, statisticians, and other relevant professionals in their organization.
7. Synthesize the ethical dimensions of data science practice (e.g., privacy).

Graduates of the MS Applied Data Science program at Syracuse University School of Information Studies, in collaboration with the Whitman School of Management, must be able to demonstrate that they have been able to master each fundamental aspect of this discipline, while also being able to synthesize their individual ability to analyze, interpret and recommend actions to stakeholders in organizations when challenged with new operational problems to solve.

The intent of this Portfolio Milestone is for the student to assemble evidence and reflect on how each course they have taken has contributed to their acquisition of the cognitive strategies defined in the program learning outcomes, and how this has enabled them to become professionally prepared in their chosen area of specialty. Within the Learning Objectives section of this document, the student will use references from their course projects or assignments and reflect on how their work has demonstrated their mastery of these concepts.

First, the Program Learning Goals (Goals) for this program cross-referenced with the courses taken.

Then, use the referenced courses taken (**bold**) and reference the program goals by number then provide an explanation/description on the linkage.

The course project is listed once after the course name. Courses are explained further and their linkages to Goals.

1. Describe a broad overview of the major practice areas in data science.
   1. All courses contributed to this learning goal.
2. Collect and organize data
   1. IST 664 Natural Language Processing - Sentiment analysis of State of the US Coast Guard speeches
   2. **IST 707 Data Analytics – Analyze errors on Adult Protective Services Cases**
   3. **IST 718 Big Data Analytics – Adult Protective Services Intakes and Workload Analyses**
   4. **IST 652 Scripting for Data Analytics – Movie: Budgets, Popularity and Revenue**
   5. IST 659 Database Administration Concepts and Database Management – Student Classes and Grades
   6. MBC 638 Data Analysis and Decision Making
3. Identify patterns in data via visualization, statistical analysis, and data mining
   1. **IST 652 Scripting for Data Analytics**
   2. **IST 707 Data Analytics**
   3. **IST 718 Big Data Analytics**
   4. IST 719 Information Visualization – Adult Protective Services Client Intake Demographics
4. Develop alternative strategies based on the data
   1. **IST 707 Data Analytics – Adult Protective Services Case reviews – Is there a relationship between errors and how to improve**
   2. **SCM 651 Business Analytics**
5. Develop a plan of action to implement the business decisions derived from the analysis
   1. **IST 652 Scripting for Data Analytics**
   2. IST 659 Database Administration Concepts and Database Management
   3. MBC 638 Data Analysis and Decision Making
   4. **SCM 651 Business Analytics**
6. Demonstrate communication skills regarding data and its analysis for managers, IT professionals, programmers, statisticians, and other relevant professionals in their organization
   1. **IST 718 Big Data Analytics**
   2. MBC 638 Data Analysis and Decision Making - Purchase Order Process Improvement
7. Synthesize the ethical dimensions of data science practice (e.g., privacy)
   1. IST 659 Database Administration Concepts and Database Management
   2. **IST 707 Data Analytics**
   3. **IST 718 Big Data Analytics**

**Note**: I have a history of working with sensitive information and/or data. This data included sensitive personal identifiable information (PII), federal tax information (FTI), Federal Education Rights and Privacy Act (FERPA), Health Insurance Portability and Accountability Act (HIPPA), and government classified information. The degree program made me more sensitive to how what can appear to be innocuous data can be turned into personalized information using multiple datasets. That has made me much more sensitive to how data is being used and can be abused.

## Describe a broad overview of the major practice areas in data science.

I digress here from my stated format as all courses taken contributed towards this goal but not all courses are referenced in this report. It seemed to me to make more sense to shift formats for this Goal.

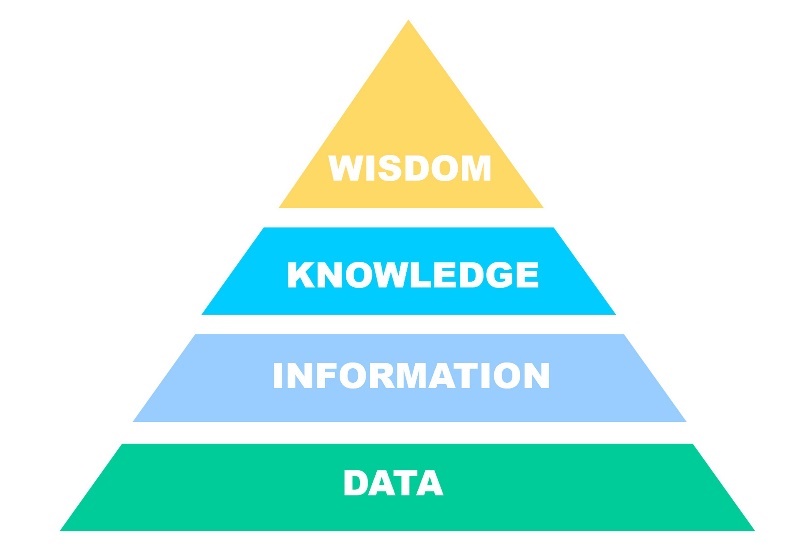
The course of study is focused on how to understand data, how to synthesize actionable insight, and then communicate that insight to business managers and others that might gain knowledge from that insight.

It ranges from understanding and practicing data science programming using real-world tools: R, python, Excel®, Adobe Illustrator® and IDEA®. Along with this is the need to understand clients and their requirements and how to understand the data available and/or needed to fulfil those requirements.

The Syracuse University program is wide-ranged encompassing the hard skills – technical and programming – to the softer skills - communication, and understanding, and client engagement. [[1]](#footnote-1)

Data Science is the field that attempts to turn data into insight. A data scientist certainly needs to understand that full spectrum of skills and knowledge one needs to really kind of turn data into insight[[2]](#footnote-2) and wisdom.

The Data, Information, Knowledge and Wisdom (DIKW) pyramid[[3]](#footnote-3)



Graphically explains the high-level goals of the various projects.

Data analysis is the foundation of all business analytics.[[4]](#footnote-4) Data analytics can be broken into four key types:

* Descriptive, which answers the question, “What happened?”
* Diagnostic, which answers the question, “Why did this happen?”
* Predictive, which answers the question, “What might happen in the future?”
* Prescriptive, which answers the question, “What should we do next?”

The course of study addressed these 4 types.

Descriptive analytics goal is to understand trends and evaluate metrics over time – tell the data’s story. It touches the surface of the data. It is useful for seeing trends and relationships between variables. It also encompasses visualization of that data.

Diagnostic Analytics uses data a step further to determine correlations, causes and trends between variables. This is where hypothesis testing, correlation v. causation, and regression analysis is used.

Predictive analytics takes data further to make predictions. It uses historical data and analytics to predict the future. Trends such as: forecasting, predicting the values of missing fields in a data set (supervised learning), classifying unlabeled data (unsupervised learning), and probable impact of data changes on future trends (semi-supervised learning).

Prescriptive analytics has been called “the future of data analytics.”[[5]](#footnote-5) It is the final known step of the data analysis. Using prescriptive analytics provides for next steps.

The following techniques were used to perform data analytics

* 1. Data Mining – structured, unstructured and text
  2. Statistical Analysis – using various models and tools
  3. Database Architecture and Management – using various tools
  4. Business Intelligence, Operations and Strategy – Business, Accounting and Financial Analytics
  5. Machine Learning – using various tools
  6. Data Visualization and Presentation – using charts, graphs, posters

***Goals 2 – 7 will be listed under the referenced courses***

## [IST 652 Scripting for Data Analysis](https://2su.datascience.syr.edu/ap/courses/1405/sections/50870bfe-a94c-4e6c-8235-1067a5697ec0/coursework)

*Goals 2, 3, and 5*

**Course Description***:* Scripting for the data science pipeline. Acquiring, accessing, and transforming data in the forms of structured, semistructured, and unstructured data.

*Additional Course Description:* The goal of this class is to teach students the tools and skills of scripting needed to solve problems of accessing and preparing data in a variety of formats and situations, sometimes known as data wrangling. The scripting will provide the skills needed to form data science pipelines, from acquiring and cleaning data to accessing data and transforming data for analysis or visualization. The main content focus is on information access and processing tasks on the types of structured, semistructured, and unstructured data in current use in information applications. For these three types of data, the course includes the use of structured numeric and text data such as that from a spreadsheet or database, the use of data obtained through standard data exchange formats such as HTML or XML from web pages or JSON from web-based APIs, and the use of data obtained by pattern matching from text or log files. The scripting language Python was chosen because of its ease of use and available packages to work with data in many information applications. The skills learned in this class are intended to complement the analytical and visualization skills learned in other data science courses. The scripting language Python will be taught, but it will be assumed that students already have a programming background, either through course work or through online study.

Learning Objectives: Upon successful completion of this course, the student will be able to write scripts to access and amass data from fields in structured data, access fields in semistructured data, and define and find patterns of data in unstructured data; prepare and transform data to produce data summaries, lists, and networks.

* analyze and solve data access problems for the three types of data and to find and deploy appropriate software packages that can be integrated into the problem solution; and
* frame real-world data questions and show how they can be answered from data.

**Project Description**

This was a team project

*Goal*: Determine/Answer the question, “Does budget a blockbuster movie make? (aka highly profitable)

*Data and Preparation*: The datasets used were downloaded from the IMDb website. The separate datasets contained data about movies to include release date, title, budget, genres, production companies, revenue, profit, popularity, vote average, and vote count.

One of the most tedious tasks of data science is the collection and organization of data. Datasets needed to be collected, cleaned and standardized using uniform naming conventions and formatting, removing erroneous data, converting dates, groupings, adding fields, etc. This may need to be done on numerous datasets with the eye towards linking them together.

Rows were removed that contained missing values using the pandas’s dropna() function. The new data frame replaced the original one.

Movie budget information is important for the analysis, so any movies that have a 0 in the budget column will be removed. To do this we are going to get a list of index names for movies that have a budget equal to 0, again revising the data frame by removing these movies.

Similarly, it is necessary to remove movies that have a 0 revenue, as it appears that this data just has not be collected, and therefore is not accurate. These entries were removed using the same method that was used to drop the movies that had a 0 in the budget column.

The data type used for the release data needed to be changed to date type. This was done using the pd\_to\_datetime function, allowing extraction of each element of the date. The month, day, and year were extracted, and new columns created named month, day and year.

**Conclusions**

After organizing, cleaning, identifying patterns and analyzing the team concluded that

* Percent profits are not uniformly distributed across budget levels.
* Movies with an extremely high budget are the least likely to have a negative percent profit.
* Movies with an extremely low budget are the most likely to have an extremely high percent profit.

Highly popular movies earned a larger percentage for the studios. The most popular movies are Disney and Warner movies. Disney has the Marvel and Star Wars franchises – all money makers. Warner Brothers has the DC comic franchises. In both cases, action adventure and comic books coming to life make money at the box office. One of the graphics generated follows.

Chart, scatter chart

Description automatically generated

The recommended plan of action to implement business decisions derived from these analyses is for studios to follow Warner Brothers and Disney - produce movies with an extremely high budget and veer away from productions with an extremely low budget.

## IST 707 Data Analytics

*Goals 2, 3, 4, and 7*

**Description**: Introduction to data mining techniques, familiarity with particular real-world applications, challenges involved in these applications, and future directions of the field. Hands-on experience with open-source software packages.

*Additional Course Description*: This course will introduce popular data mining methods for extracting knowledge from data. The principles and theories of data mining methods will be discussed and will be related to the issues in applying data mining to problems. Students will also acquire hands-on experience using state-of-the-art software to develop data mining solutions to scientific and business problems. The focus of this course is in understanding of data and how to formulate data mining tasks in order to solve problems using the data. The topics of the course will include the key tasks of data mining, including data preparation, concept description, association rule mining, classification, clustering, evaluation and analysis. Through the exploration of the concepts and techniques of data mining and practical exercises, students will develop skills that can be applied to business, science or other organizational problems. The format of the class meetings will be a combined lecture and lab format, with lectures and class discussions to cover material and lab time to investigate small examples for the topic of the week. There will be weekly readings

**Learning Objectives**: After taking this course, the students will be able to:

* Document, analyze, and translate data mining needs into technical designs and solutions.
* Apply data mining concepts, algorithms, and evaluation methods to real-world problems.
* Employ data storytelling and dive into the data, find useful patterns, and articulate what patterns have been found, how they are found, and why they are valuable and trustworthy.

**Project Description**

This was a team project.

*Background:* The Department of Family and Protective Services – Adult Protective Services Division (Agency) is charged with protecting the most vulnerable among us. Allegations are received, and investigations are conducted to validate the allegation, ensure alleged victim safety, and put processes in place to prevent reoccurrence. Post case closure, a sample of cases is reviewed to ensure the investigation was conducted correctly, ensuring the alleged victim is safe and receiving any needed services.

The Agency is concerned that there are connections between errors, that certain errors produce other related errors, and challenge alleged victim safety and wellbeing. Additionally, these connections may be causing cases to be scored lower than they should and provide false indications as to the quality of care being provided. This affects metrics on the strategic plan that needs to be submitted to the state legislature. It also affects staff morale, some of whom do not see any way to improve scoring.

Using this information, the Agency would like to determine if there are relationships between categories and if there are gaps in services provided and/or redundancies in scoring that if eliminated will improve staff morale and provide for a more accurate “picture” to the state legislature.

Using this information, the agency wants to ensure a high standard of care and if indicated, target certain categories for quality-of-care improvements. An additional benefit may be an indication as to whether the current pandemic affects the scoring.

*Goal*: Are there relationships between scoring categories that may indicate redundancy in scoring that if eliminated will improve staff morale and provide for a more accurate “picture” to the state legislature.

*Data and Preparation*: The data was not located in one file. Two files were needed, one containing individual errors associated with cases and one with comments associated with those cases.

The multiple datasets being used contain data provided by the agency over six months (June – Nov 2020). The dataset contained almost 24,000 lines of data covering 3.200+ individuals and encompasses state-wide cases. A six-month period was used as the case intakes are seasonal. It contains data for all case scorings, over errors given to each case.

There were over 60,000 cases from June – November 2021. The case readings sampled 4.8% of the cases. Very few cases did not receive at least one error.

The datasets required extensive cleaning, removal of case identification information and other personally identifying information. During the cleaning process, it was discovered that there were a significant number of non-standard responses - other possible words being used, such including Unknown and other caseworker-provided text. These needed to be standardized. Additional categories were added for Month\_Yr, Standard (1.1, 1.2, 2.1, etc.) and as to whether the case met standards or not (Yes\_No). Excel was used for the majority of the cleaning. This allowed the two datasets to be linked for analysis.

Cases being reviewed are scored on 11 categories for alleged victim safety and quality containing subcategories which are scored. The scoring uses a Met; Did Not Meet; Exceeded; Unknown/NA ratings. Comments are required for any category that receives a Did Not Meet, Exceeds and Unknown.

Frequency charts, descriptive statistics, associative rule mining using Apriori algorithm, relationship plots, and cluster dendrogram were used. Predictive analytics using Naïve Bayes, confusion matrices, SVM, Knn, random forest and decision tree models were explored and used.

Basic text mining techniques were also used creating a word cloud of error codes and frequencies of words used in the comments.

Identifying patterns via visualization, analysis and data mining provided interesting results. Numerous visualization tools were used, and some are presented below.

Chart, bar chart

Description automatically generated

Using R Studio’s apriori associative rule mining package produced charts and tables similar to the following

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **lhs** |  | **rhs** | **confidence** | **count** | **% of cases** |
| {2.1} | => | {1.2} | 0.779 | 335 | 33% |
| {5.1} | => | {1.2} | 0.726 | 312 | 31% |
| {2.2} | => | {1.2} | 0.810 | 310 | 31% |
| {2.1} | => | {3.1} | 0.600 | 258 | 26% |
| {3.1} | => | {2.1} | 0.524 | 258 | 26% |
| {5.1} | => | {3.1} | 0.565 | 243 | 24% |
| {2.1} | => | {5.1} | 0.502 | 216 | 21% |
| {5.1} | => | {2.1} | 0.502 | 216 | 21% |
| {6.1} | => | {3.1} | 0.522 | 187 | 19% |
| {4.3} | => | {1.2} | 0.815 | 119 | 12% |
| {4.2} | => | {1.2} | 0.800 | 113 | 11% |
| {2.3} | => | {1.2} | 0.750 | 102 | 10% |

There were some interesting combinations with high confidence but as the percentages indicate, did not happen very often. 1.2 errors were in 6 of 9 cases listed above. There was no easy way to pull out 1.2 errors in this analysis.

Clustering was also used

**Chart, histogram

Description automatically generated**

From clustering we looked at predictive modeling using Naïve Bayes, SVM, Knn, Confusion matrices and decision trees.

***Timeline

Description automatically generated***

Ethical dimensions of data science practice were a concern in this project. Data privacy was a real concern as the initial datasets contained personally identifying information (PII) which violated the APS clients’ privacy.

I cleaned and organized those datasets that contained the PII so that my teammates who did not work for the Agency could use the data.

**Conclusions**

As the data owners suspected, the results produced correlations between the errors. In plain English, errors in documenting the cases are related to, and can possibly produce, errors in the investigation or vice-a-versa – counting errors twice. These results appear to support their initial hypothesis that improving the quality of the investigation documentation should directly affect other documentation errors.

An important note is that when dealing with people and their evaluations, there is the propensity for a wide range of results as the data deals with humans with varying levels of experience and abilities, at timers not generating high accuracy rates.

To completely understand the balance between causation and correlation in this project will require a more in-depth analysis of the comments attached to the errors.

## IST 718 Big Data Analytics

*Goals 2, 3, 6 and 7*

**Description:** A broad introduction to analytical processing tools and techniques for information professionals. Students will develop a portfolio of resources, demonstrations, recipes, and examples of various analytical techniques.

**Learning Objectives:**

**During the course, we will emphasize:**

* 1. Experiential learning through reading and practical exercises.
  2. Collaborative learning through online discussions between instructors and peers.
  3. Self-learning with appropriate instructional support and timely feedback using analytical case studies.

**After taking this course, the students will be able to:**

1. **O**btain data and explain data structures and data elements.
2. **S**crub data by applying scripting methods, to include debugging, for data manipulation in Python, R or other languages.
3. **E**xplore data by analyzing using qualitative techniques including descriptive statistics, summarization, and visualizations.

**Final Project**

This was a team project.

*Goal*: For the final project, students will identify a data-focused problem, bring together different data sources, conduct analysis, draw conclusions, and produce a report explaining the results. Maximum points are possible if the submission is on-time, complete, and demonstrates the student’s ability to select the appropriate analytical methods to the chosen problem; interpret the data, model, analysis, findings; draw appropriate conclusions; and present the results in a meaningful way.

4. **M**odel relationships between data using the appropriate analytical methodologies matched to the information and the needs of clients and users.

5. I**N**terpret the data, model, analysis, and findings. Communicate the results in a meaningful way.

6. Select an applicable analytical methodology for real problems in areas such as business, science, and engineering.

The question asked “What Does the Future Hold For Texas Protective Services? “

*Data and Preparation*: The datasets used were “big”.

* Dataset 1 – Type of allegations:Size: 856,264 x 4
* Dataset 2 – Demographics: Size: 497,026 x 7
* Dataset 3 – Source of Report: Size: 35,626 x 5

We used data from the Agency, which in its raw form contained personal identifiable information and other sensitive personal information. I was allowed to use that data only if I cleaned it and anonymized it. Additionally, some of the data points while not sensitive by themselves but because of the small sample size and/or linkages to the other datasets could be traced to a person with some minor additional digging.

Cleaning was and organizing was extensive. For example, dataset 1 - after filtering out the NA values, non-standard spellings on the 254 Texas counties, so after filtering out counties and removing questionable rows it ended up with 855,300 rows.

Additionally, there were non-standard entries for allegations, which needed to be filtered and standardized. After filtering out all those we could not correctly group into one of the 5 categories (Physical, Mental, Sexual, Exploitation, and Emotional) we ended up with 855,257 rows of data.

Similar cleaning and organization needed to be conducted on the other datasets.

Data exploration and visualizations produced some of the following

*Graphical user interface

Description automatically generated*

We identified patterns in the data

*Chart, histogram

Description automatically generated*

*Chart, bar chart

Description automatically generated*

And used the data to make predictions so the Agency could look at staffing and workloads

*Chart, line chart

Description automatically generated*

Jupyter (Python) was used for data analysis and graphics, Microsoft PowerPoint for presentations. VARMAX modeling was used for analysis, forecasting and graphing those forecasts. plotly package for python was used to create the chloroplast of Texas.

**Conclusion**:

There are no significant differences between women and men when it comes to intakes and services

What we did not look at, which should be explored, is what percentage of the eligible population is male and female and if there is/are any cultural biases that would make one gender more likely to ask for assistance than the other. Anecdotally, it appears that women are more likely to accept assistance than men.

## SCM 651 Business Analytics

*Goals 3, 4 and 5*

**Course Overview and Learning Objectives**

This course is intended for the graduate student who is interested in developing a portfolio of skills in business analytics.

The course learning objectives include:

1. Data collection: using tools to collect and organize data (e.g., Google Analytics)
2. Data analysis: identify patterns in the data via visualization, statistical analysis, and data mining
3. Strategy and decisions: develop alternative strategies based on the data
4. Implementation: develop a plan of action to implement the business decisions

Class discussions will be based on case situations and on articles from business and technical publications.

The class will include substantial hands-on work in data collection, analysis and interpretation.

**Course Format**

Classes will include a mixture of hands-on lab sessions and case discussions. The course readings will serve as the basis for live discussion on basic business analytics. Lab sessions focus on learning skills required for data analysis.

This course was different from most of the other courses in the program, as it was an elective and taken at the Whitman School of Business. It used Excel and Tableau as some of the tools. There was no final project.

*Data and Preparation and Analysis*: Various provided datasets were used. What I found most interesting was the extent that Excel has evolved in the data analysis area. I had a wealth of experience using Excel, but not in the ways demonstrated in class.

We used generated ANOVA, Probit, Logit and Neural Networks using Excel.

Diagram

Description automatically generated with medium confidence

Conducted sensitivity analyses, generated dashboards, Google analytics, and 3d mapping using Excel.

**Conclusion**

I found this course very interesting as the primary general staff data analytic tools at the Agency is Excel. Tableau is in limited use. The course provided tools and techniques that I could use directly in my “day-job.” Using what I learned about 3D mapping, I created 3D mapping visualizations for the Agency, bringing a new perspective to the data that my agency uses.

# Conclusion

I am an atypical graduate student. In most, if not all of the classes, I was the oldest student by decades. I have an extensive background in process improvement (TQM, Theory of Constraints, Throughput Operating Systems, Six Sigma), continuity planning, management of risk, project management, maintenance planning, military operations, aviation, and maritime law enforcement. My varied background helped me look at some things differently that I might have earlier in my career. I think at times it allowed me to look at the tools presented and see applications in other areas than being presented. This is a success of the program.

In all the assignments and projects, I have been challenged to go out of my comfort zone. Although I enjoy programming, it had been years since I programmed in depth. I was a beginner in python, and had no R or SQL background, which was challenging. I learned how to prepare and clean data, import packages and modules, and produce results that I did not think I could.

I have always been an analytical type and problem solver. I have created business plans and strategic plans, both of which used metrics to help determine success. What was lacking was the prescriptive and proscriptive skills. The courses helped hone those skills and whet my appetite for more.

The teamwork aspect of many courses was a two-edged sword. Being part of a team presented some of its own challenges, The differing of knowledge and skill levels, remote work, and time zone differences made some projects more challenging than others. Workload distribution was generally divided based on expertise – and many times it was evenly divided. But is not that just like the “real world?” I learned from teammates, but at times saw my teammates move ahead leaving me behind…good learning experiences to take with me.

There was a direct benefit to my current position as a management analyst. I was able to take what I learned in many of the classes and directly apply the techniques and skills learned to conduct analyses on real-world data in some unique ways. I have provided trend analyses and predictive analyses, as well as created some models that have been used for staffing, hiring and mitigating the negative effects of turnover. (Unfortunately, much of what I did for the Agency used sensitive data, which cannot be shared in this document.)

I have only used four courses in this paper, but the learning did not stop with these four. IST 664 Natural Language Processing was one of the more difficult courses requiring, in my opinion, requiring more work than the 3 credits earned. That said, it was a new experience and I learned and can use some unique tools when processing language data – which can be a significant part of my “day job.”

# Other courses are included at the end for safekeeping and reference but will not be included in the final version.

### IST 659 Database Administration Concepts and Database Management

*Description*: This course is aimed at information professionals and serves as a foundational introduction to the field of data and database management. \*Description: IST 659 is an introductory course to database management systems. This course examines data structures, file organizations, concepts, and principles of database management systems (DBMS) as well as data analysis, database design, data modeling, database management, and database implementation. More specifically, it introduces hierarchical, network, and relational data models; entity-relationship modeling; basics of Structured Query Language (SQL); data normalization; and database design. Using Microsoft’s Access and SQL Server DBMSs as implementation vehicles, this course provides hands-on experience in database design and implementation through assignments, lab exercises, and course projects. This course also introduces advanced database concepts such as transaction management and concurrency control, distributed databases, multitier client/server architectures, web-based database applications, data warehousing, and NoSQL.

*Learning Objectives*: After taking this course, the students will be able to: Describe fundamental data and database concepts Explain and use the database development lifecycle Create databases and database objects using popular database management system products Solve problems by constructing database queries using Structured Query Language (SQL) Design databases using data modeling and data normalization techniques COURSE SYLLABUS IST 659 Database Administration Concepts and Database Management 2 Syllabus Template Revised 12/17/2015 Syllabus last updated 06/27/2019 Develop insights into future data management tool and technique trends Recommend and justify strategies for managing data security, privacy, audit/control, fraud detection, backup and recovery Critique the effectiveness of DBMS in computer information systems

*Project*. The major portion of grading in this course is the course project. There are two deliverables for this project. This first will be due roughly halfway through the course, the second due at the end. Consult the project instructions and the instructor for specific due dates. This project is a comprehensive demonstration of your mastery of the content of this course.

### [IST 664 Natural Language Processing](https://2su.datascience.syr.edu/ap/courses/1679/sections/3adeac32-2a1b-4ffa-9bac-b2a62490c684/coursework)

*Course Description*:

This course is designed to develop an understanding of how natural language processing (NLP) can process written text and produce a linguistic analysis that can be used in other applications. This goal will be achieved by:

* Readings, lectures, and class discussions of the multiple levels of linguistic analysis required for a computer to accept natural language input, interpret it, and carry out a particular application.
* Lab exercises and assignments in using some of the computational techniques required to perform these levels of natural language processing of text.
* Studies of real-world applications that incorporate substantive NLP modules.

The course primarily covers the techniques of NLP in the levels of linguistic analysis, going through tokenization, wordlevel semantics, partofspeech tagging, syntax, semantics, and on up to the discourse level. It also includes the use of the NLP techniques, such as information retrieval, question answering, sentiment analysis, summarization, and dialogue systems, in applications.

*Learning Objectives*

At the end of the course the student will be able to:

* Demonstrate the levels of linguistic analysis, the computational techniques used to understand text at each level, and what the challenges are for those techniques.
* Process text through the language levels using the resources of the Natural Language Toolkit (NLTK) and some rudimentary use of the programming language Python.
* Describe how NLP is used in many types of real-world applications.

*Project/Final Assignment.* The final assignment will be a final project where students will focus on a text classification task, which may be something like sentiment analysis, and conduct and report on a series of experiments.

Processing language is a difficult task due to the fluidity of languages. Tome of voice, sarcasm, idiomatic expressions all complicate the ability of a machine to correctly nd successfully identify patterns an identify trends.

Naïve Bayes classifier, SKLearn, cross-validation, NLTK. The course primarily covers the techniques of NLP in the levels of linguistic analysis, going through tokenization, word-level semantics, part-of-speech tagging, syntax, semantics, and on up to the discourse level. It also includes the use of the NLP techniques, such as information retrieval, question answering, sentiment analysis, summarization, and dialogue systems, in applications.

While it is a daunting task, there is a need as more and more data are unstructured, and text based – in applications/programs/repositories such as Twitter® and Facebook®. This data needs to be processed into structured data so it can be used effectively by data scientists.

### [IST 687 Introduction to Data Science](https://2su.datascience.syr.edu/ap/courses/1252/sections/a40bc046-4cd7-47d2-80bd-3441bfc85901/coursework)

*Course Description*: The course introduces students to applied examples of data collection, processing, transformation, management, and analysis to provide students with hands-on introduction to data science experience. Students will explore key concepts related to data science, including applied statistics, information visualization, text mining and machine learning. “R”, the open source statistical analysis and visualization system, will be used throughout the course. R is reckoned by many to be the most popular choice among data analysts worldwide; having knowledge and skill with using it is considered a valuable and marketable job skill for most data scientists

*1.2. Learning Objectives* At the end of the course, students are expected to understand:

* Essential concepts and characteristics of data
* Scripting/code development for data management using R and R-Studio
* Principles and practices in data screening, cleaning, and linking
* Communication of results to decision makers

At the end of the course, students are expected to be able to:

* Identify a problem and the data needed for addressing the problem
* Perform basic computational scripting using R and other optional tools
* Transform data through processing, linking, aggregation, summarization, and searching
* Organize and manage data at various stages of a project lifecycle
* Determine appropriate techniques for analyzing data

*Final Project* – Team

### [IST 719 Information Visualization](https://2su.datascience.syr.edu/ap/courses/1600/sections/794fe514-4421-4874-a057-bca15b39f02c/coursework)

*Description:* A broad introduction to data visualization for information professionals. Students will develop a portfolio of resources, demonstrations, recipes, and examples of various data visualization techniques. Additional Course Description: Introduction to skills and techniques related to information visualization, through the R programming language and Adobe Illustrator. These skills include using data cleaning techniques, controlling the R graphics environment, developing custom plots, visually exploring data, using design concepts to visually communicate the story in the data, and discussing issues related to the ethics of data visualization. Conceptual themes will be presented alongside technical aspects of data visualization. Additional work and higher grading expected of graduate students.

*Learning Objectives*: After taking this course, the students will be able to

* 1. Use R to do basic data cleaning and preparation on a wide range of datasets
     1. Use functions to summarize and compare fields
     2. Find missing values
     3. Use subsets or filter data
     4. Retype data into correct format
  2. Identify stories in datasets through exploration
     1. Use R to create appropriate rough plots to identify distributions and relationships in the data
     2. Use data subsetting and filtering to home in on questions of interest
  3. Create rich visual artifacts that communicate data stories
     1. Identify the optimal type of visualization to minimize viewer cognitive overload and maximize image interpretability
     2. Enhance viewer cognition through context cues
     3. Use basic design principles to enhance viewer receptivity and convey meaning
     4. Use Adobe Illustrator to combine R data visualizations, design elements, and context cues into a single artifact



### IST 722 Data Warehouse

*Description*

Introduction to concepts of business intelligence (BI) and the practice/techniques in building a BI solution. Focuses are on how to use data warehouses as a BI solution to make better organizational decisions.

*Additional Course Description*

This course provides concepts, principles, and tools for designing, implementing, and using Data Warehouses. More specifically, we introduce database constructs such as Operational Data Store (ODS), Data Warehouse, and Data Mart, and related components. We study the differences between Ralf Kimball’s and Bill Inmon's approaches, roles and responsibilities in the design and implementation of a Data Warehouse, project management guidelines and techniques, requirements gathering, dimensional modeling, Extract Transform and Load (ETL) architecture, analytical reporting concepts, data governance and recent trends in the data warehouse domain.

The course provides students with hands-on lab assignments, course readings, and selected videos that serve as the basis for live class discussions. The daily discussions will include assessing business process needs, designing the data warehouse, building the data warehouse with ETL, and analysis services and business intelligence. The course will leverage a portfolio of SQL Server tools that include SQL Server DBMS, SQL Server Integration Services (SSIS), and SQL Server Analysis Service (SSAS).

There will be approximately one week of pre-work prior to the start date, including reading articles and watching videos. The prework will provide a foundation to jump start the live discussions. Each day during the four-day course, there will be class exercises and homework assignments that must be completed prior to the start of the next morning’s class. There will also be a group project that will need to be submitted approximately one week after the last live session, as there is no final exam.

The course will meet from 10:00AM EST - 6:00PM EST each day. Late arrival and early dismissal will not be permitted so please arrange accordingly.

*Learning Outcomes*

Taking this course will provide the following learning outcomes:

1. Technical Knowledge

You will gain technical knowledge and comprehension about data warehouses. You will develop the ability to apply these technologies to solve information problems at the individual and organizational levels. After completing this course, you will be able to:

* Define analytics requirements based on business process understanding
* Describe various database constructs - Data Warehouse, Data Mart, ODS
* Describe the components of a data warehouse
* Differentiate between Ralf Kimball’s and Bill Inmon's approaches • Know how to apply various integration approaches in a merger - ETL, EII, EAI
* Describe a Master Data Management (MDM) solution
* Create database objects using popular database management system products
* Design and implement data warehouse and business intelligence components
* Gain extensive hands-on with SSIS (ETL), SSAS (Cube) and Power BI tools

1. Management of Solution Development You will gain knowledge and comprehension of the disciplines used in the development of data warehouse solutions. You will develop the ability to apply these disciplines in developing solutions for certain organizational and business problems. After completing this course, you will be able to:

* Define the roles and responsibilities in the design and development of data warehouses
* Differentiate various requirements gathering and dimensional modeling techniques • Relate business processes with objects in the data warehouse
* Define project management guidelines

1. Management of Information Technology You will be able to integrate technical and solution development concepts with the principles of data governance, strategic alignment and information analysis. You will be able to apply these concepts in the analysis of complex management case studies and problems. After completing this course, you will be able to:

* Identify business processes in the data warehouse
* Describe the data governance concepts
* List some of the recent trends in Data Warehouse

Complete a Group Project, which demonstrates your ability to work in a team to build a data warehouse and the business intelligence solution around it. An outline of the group project instructions will ap…

### [ACC 652 Accounting Analytics](https://2su.onlinebusiness.syr.edu/ap/courses/745/sections/40357f6a-9aa2-435d-b938-e2e3003b156d/coursework)

*Learning Objectives*: The course learning objectives include:

1. Perform tests to assess data conformity, summation, and duplication using Benford’s Law
2. Test the internal diagnostics of current and prior period data
3. Identify accounting anomalies using subset tests, growth tests, and relative size factor tests
4. Identify accounting fraud using correlation and time-series analysis
5. Detect financial statement and financial transaction biases

Analytics Project (100 points, 15%): Each student will complete an analytics project utilizing data from a fictitious beverage distributor. Requirements, grading rubric and necessary data files will be provided to students during week 7 of the course. Final submissions will be due prior to the week 10 Live Session.

### [FIN 654 Financial Analytics](https://2su.onlinebusiness.syr.edu/ap/courses/707/sections/1c14edc5-1174-417b-b4e3-b4667f51b7a6/coursework)

*Course Description*

An introduction to models, methods and tools useful in decision-making in the financial industry. Topics include Capital Asset Pricing Model, portfolio sorting and optimization, and the application of machine learning techniques in financial decision making. Computer statistics programs are widely used with emphasis on the assumptions, limitations and interpretation of the methods.

*Learning Objectives*

Upon successful completion of the course, students will be able to:

* Use R to download, analyze and interpret time series data on financial securities,
* Employ Capital Asset Pricing Model and make investment decisions,
* Optimize investment portfolios using R and based on historic data,
* Apply machine learning methods to classification problems such as credit risk rating and bankruptcy prediction,
* Develop artificial neural network models for quantitative prediction problems such as real-estate pricing and revenue prediction.

### [MBC 638 Data Analysis and Decision Making](https://2su.onlinebusiness.syr.edu/ap/courses/693/sections/3d8055f0-1df4-49de-94fb-1544e8f62d30/coursework)

*Course Description:* This course will familiarize students with the assumptions underlying various statistical techniques and assist in identifying their appropriateness in a variety of situations. The student should be able to perform statistical analysis and interpret results in a meaningful way. Students are expected to relate results of such analyses to become an information-based decision maker. Learning Objectives:

Help students understand the value of data collection and analysis in acquiring knowledge and making decisions in today’s business environment.

Students will be able to identify and apply the appropriate statistical technique for a given set of conditions in order to answer a particular question.

*Project*: Process improvement storyboard

1. Adapted from Welcome to the Program, 1.5 Applied Data Science [↑](#footnote-ref-1)
2. IST 687 Introduction to Data Science, Introduction 1.2 Data Science: Many Skills [↑](#footnote-ref-2)
3. “DIKW Pyramid”, accessed 06 May 2022, http://shonellerton.com/wp-content/uploads/posts/2019/20191016-dikw-pyramid/y2019m10d16\_dikw\_pyramid\_003.jpg [↑](#footnote-ref-3)
4. Conrad, Andrew. “What Is Descriptive Analytics? A Definition.” *GetApp, GetApp*, 20 Apr. 2022, https://www.getapp.com/resources/descriptive-analytics-definition/. [↑](#footnote-ref-4)
5. Cote, Catherine. “What is Prescriptive Analytics? 6 Examples.” *Harvard Business School Online*, 02 Nov , https://online.hbs.edu/blog/post/prescriptive-analytics/. [↑](#footnote-ref-5)