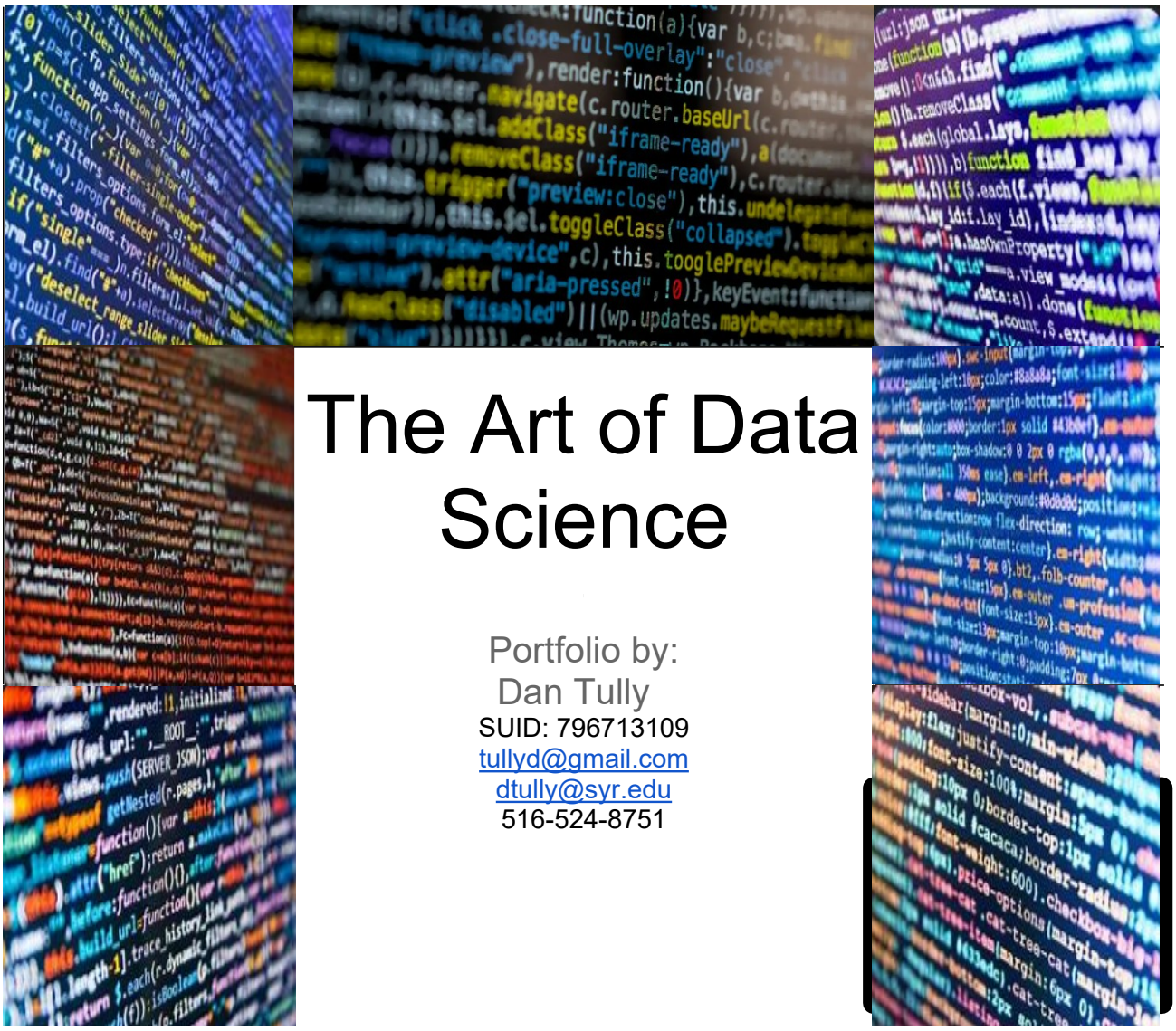

The Art of Data Science

Portfolio by:
Dan Tully
SUID: 796713109
tullyd@gmail.com
dtully@syr.edu
516-524-8751

Table of Contents

About Why I Started This Journey

My fascination with data science started a few years ago. My company offered a data science 101 week long boot camp, which was conducted by data society. The immersive classes cover a wide range of data science principles mostly utilizing R programming to clean, normalize and visualize data. This experience solidified my interest in learning more about data science. I researched different programs offered by the top universities and felt that the iSchool (Syracuse University) had exactly what I was looking for in their Applied Data Science (ADS) Master's (MS) Degree program. This online program offered the flexibility I need to work full time, raise my two children, and still attend to my professional developmental needs.

The ADS MS Learning Outcomes

These are the program's learning goals. The ability to:

[1] 💾 Collect, store, and access data by identifying and leveraging applicable technologies

[2] 🔃 Create actionable insight across a range of contexts (e.g. societal, business, political), using data and the full data science life cycle

[3] 👩‍🏫 Apply visualization and predictive models to help generate actionable insight

[4] 👨‍💻 Use programming languages (i.e. such as R and Python) to support the generation of actionable insight

[5] 📢 Communicate insights gained via visualization and analytics to a broad range of audiences (including project sponsors and technical team leads
[6] 🔏 Apply ethics in the development, use and evaluation of data and predictive models (e.g., fairness, bias, transparency, privacy)

My Courses

ACC652 Accounting Analytics

This course as it relates to the program learning goals includes: 💾 🔃 🧑‍🏫 🧑‍💻 📢 🔏
📁 [Portfolio Project Folder (ACC652)]

| ACC652_ Project - Data Definitions - 2.17.22.xlsx
| Dan_Tully_ACC652_Project.xlsx
| ACC652_Document_File.txt

The objectives of this course helped me to realize and articulate how analytics have changed the accounting profession. Since data is no longer stored in handwritten/typed book ledgers but rather digitally in databases. Store transactions are automated and easily accessible. The financial performance (i.e 10-K reports) for publicly-traded companies are downloadable from the internet (https://www.sec.gov/) making research and data analysis much easier.

Analytic tests, consistent with sales suppression fraud (zapper fraud), included: analyzing transaction sequencing for caps, analyzing cash vs. credit payment trends over time and industry benchmarks, reviewing average bill amount trends and others. Register disbursement consist of False entries allowing the removal of funds from the cash register (e.g. False refunds, voids, or coupons)

Common fraud analytical anomalies include: outliers/inliers, Too many/few transactions, unusual relationships between items, unexpected timing of transaction, unusual accounts, gaps/duplicates and unreasonable items. There are common fraud detection principles such as Benford's law, duplication/z-score tests, the same-same-same test, correlation, trends, Gestalt Element Analysis (GEL), Relative Size Factor (RSF).

The part I really enjoyed in this course was the use of different analytical tools that are used in the accounting profession (such as Excel, Tableau, IDEA, and Alteryx) to visualize and communicate insight from the data. The instructor encouraged us to become familiar with these tools to answer the questions and explore the data. We used these tools to construct ad hoc and automated analytics. My project demonstrates

how I utilized the various tools and techniques taught in this class to find and report on the anomalies in the data sets.

IST644 Managing Data Science Projects

This course as it relates to the program learning goals include: 🔃 📢 🔒

📂 Portfolio Project Folder (IST644)

|- IST644_ReflectionPaper.pdf

|- IST644_Document_File.txt

We discussed how to leverage agile concepts within a data science project. Within this reflection paper, I summarized all that we learned in the course. We learned that key differences between data science projects and software development projects is primarily task estimation and decoupling meetings from interactions, allowing capability-based iterations. We examine several data science life cycles and standard workflow frameworks, such as CRISP-DM and TDSP. We explained the responsibilities of the different roles within a data science project as well as examined key aspects of coordination frameworks, such as Kanban and Scrum.

This course also challenged us to identify and minimize the potential ethical implications of a data science project. While reflecting on these different lifecycles and frameworks we were asked to select and use the most appropriate team process framework for a specific project. I grew an appreciation for all the different lifecycle/frameworks that are available. This was enlightening and prepared me well to engage in future data science project planning.

IST652 Scripting for Data Analysis

This course as it relates to the program learning goals include: 💾 🔃 🤷 🤦 📢 🔒

📂 Portfolio Project Folder (IST652)

|- Tully_Dan_Final_Project.pdf

|- Tully_Dan_Final_Project.ipynb

|- IST652_Document_File.txt

This course challenged me on writing scripts to access and amass data from fields in structured and semistructured data as well as defining patterns in unstructured data. We reviewed all types of data while preparing and transforming it to produce data summaries, lists, and networks. I really enjoyed learning more about jupyter notebook and python as we reviewed and munged data.

The final project in the folder shows how I analyze and solve data problems using different types of data and integrated solutions. I reviewed real-world questions and showed how they can be answered from data. This class showed us how to collect, access and store data while creating actionable insight. I think this class was the first time I accessed mongoDB.

We created visuals from the data using python to support actionable insight. The final report communicates my findings. Although it was not heavily emphasized, the textbook and course slides related to social media discussed bias and privacy in data sharing and use. My final project was chosen to help me better understand and connect with my son and his current interests.

IST659 Database Admin Concepts and Database Mgmt

This course as it relates to the program learning goals include: 💾 🔃 🧑‍🏫 🧑‍💻 📢 🔐
📂 Portfolio Project Folder (IST659)
|- 8263_Tully_Project2.pdf
|- IST659_Document_File.txt

This was the first class I took in this program and it really helped me appreciate the database development lifecycle. I created procedures, tables, and solved problems by constructing database queries using Structured Query Language (SQL). We create databases and database objects using popular database management system products (e.g., SSMS and MS Access) . Described fundamental data, database concepts, and design databases using data modeling and data normalization techniques.

While designing the database we discussed developing insights into future data management tools and technique trends as well as recommending strategies for managing data security, privacy, audit/control, fraud detection, backup and recovery. I really enjoyed this class as we critiqued the effectiveness of a database management system (DBMS) in computer information systems.

We achieved all of the program learning objectives as demonstrated in the final project located in the folder. There wasn't a lot of visualization but we did model the database using the entity relationship (ER) diagram which visually shows the tables and relationships of the data. Working through this class I grew an appreciation for relational databases and how they are constructed.

IST687 Intro to Data Science

This course as it relates to the program learning goals include: 💾 🔃 👷 👨‍💻 📢 🔐

📂 Portfolio Project Folder (IST687)

|- Dan_Tully_Team3_IST687_FinalProject.pptx
|- IST687_Document_File.txt

This introduction course provided the essential concepts and characteristics of data. It introduced scripting/code development for data management using R and R-Studio. Starting from the basics of dataframes and data munging through more sophisticated discussions in support vector machine (SVM), data mining, as well as big data.

We applied the principles and practices in data screening, cleaning, and linking as demonstrated in the team project in the folder. In this project, my team created a presentation to brief executives in the Hyatt Hotel company on how to increase traffic to their hotels. We prepared various visualizations to emphasize story points we found in the data. Throughout this exercise we learned the importance of communicating our findings effectively to decision makers.

IST707 Data Analytics

This course as it relates to the program learning goals include: 💾 🔃 👷 👨‍💻 📢 🔐

📂 Portfolio Project Folder (IST707)

|- Dan_Tully_HW6-7.docx
|- IST707_Document_File.txt
|- 21044_FedPapers.RMD

The project folder artifact shows how we document, analyze, and translate data mining needs into solutions. It allowed me to apply data mining concepts, algorithms, and evaluation methods to real-world problems. Early in the course we even discussed some of the challenges of data mining such as privacy preservation. This class relied on R for the data analysis but also introduced me to other software packages such as Weka Explorer.

We used Weka and R to explore the federal papers of Hamilton, Jay, and Madison. This was interesting to me mostly because of the significance in history. It really opened my eyes to how technology has evolved and we are able to analyze (using K-means clustering and decision tree) these writings to better determine the authors based on the words.

After the evaluation was complete - the data storytelling began - we dove into the data, finding useful patterns, and articulating those patterns found, how they are found, and why they are valuable and trustworthy. This homework research assignment in the folder exemplifies the journey of a data scientist. It leverages various models and examines their strengths and weaknesses.

IST718 Big Data Analytics

This course as it relates to the program learning goals include: 💾 🔃 👷 👨‍🏭 📢 🔐

📂 Portfolio Project Folder (IST718)

|- 24119_Dan_Tully_IST718_HW1.pdf
|- IST718_Document_File.txt

This course I am currently in and we will address all of the program goals. We started with discussing trends in big data including data governance and privacy. We are using Python and Apache Spark to build big data analytics pipelines and learn classic and state-of-the-art machine learning (ML) techniques. We are then going to explain how advanced analytics can be forecasting or predictive to create competitive advantages. In this course, I was also introduced to Google Colab which is a free Jupyter notebook environment that runs entirely in the cloud. This allowed us to run big data analysis without crashing our local computers.

In the project folder is a homework assignment. We were looking at how a business challenge can easily translate into an analytics challenge. The homework reviews and explains how advanced analytics can be leveraged to create a competitive advantage. Analyzing big data and identifying insights that can lead to actionable results. At the end of the course in our final project, we will be creating visualizations and communicating our result analysis with a different dataset that has a similar big data problem.

IST737 Visual Analytic Dashboards

This course as it relates to the program learning goals include: 💾 🔃 👷 👨‍🏭 📢 🔐

📂 Portfolio Project Folder (IST737)

|- Final_DashboardData.png
|- Final_DashboardDetail.png
|- Final_DashboardMain.png
|- Final_Q&A.png
|- IST737_Document_File.txt
|- Tully_Dan_Final.twbx

In this class we used Tableau to describe best practices in using (interactive) data visualization to support human reasoning that can deliver actionable insights (visual analytics). We were heavily focused on the design and critiqued the dashboards that support visual analytics using the best supported visual analytics to tell the data story.

In my project I created three dashboards and a storyboard (FAQ). The main dashboard view for the animal shelter pointed out trends and seasonality. The detail dashboard view dove more into the specific animal characteristics. The data dashboard view was for data exploration and pointed out blanks and errors in the data. Finally, I used the storyboard (FAQ) to highlight specific business questions and suggested answers.

The final product in the folder demonstrates actionable insight from the data that could help the business. The text for course discusses privacy and biases in data visualizations and recommends user awareness and protecting data. This course forced us to look not just at the data, but the presentation.

IST772 Quantitative Reasoning for Data Science

This course as it relates to the program learning goals include: 💾 🔃 🧑‍🏫 🧑‍💻 📢 🔏

📁 Portfolio Project Folder (IST772)

|- HW10_Tully.pdf
|- HW10_Tully.RMD
|- HW8_Tully.pdf
|- HW8_Tully.RMD
|- IST772_Document_File.txt

In this course we demonstrate knowledge of contemporary inferential statistical concepts and data analysis strategies. We practice effective data science analytics to demonstrate competence of the skills needed for use of popular statistics. The data management platform used to conduct sound and reproducible analysis was mainly R.

The homework in the folder demonstrates the use of statistics in data analytics. We explored and interpreted the data in our analysis of the data sets to determine inferential statistical concepts. In the folder examples it shows how we used analysis of variance (ANOVA), diagnostic for multicollinearity, and AMOC ("at most one change") for data analysis.

MBC638 Data Analysis and Decision Making

This course as it relates to the program learning goals include: 💾 🔃 🧑‍🏫 🧑‍💻 📢 🔏

📁 Portfolio Project Folder (MBC638)

|- 20344 DAN TULLY FINAL PROJECT v2.pptx
|- MBC638_Document_File.txt

Learning Objectives:
This class explored DMAIC (Define, Measure, Analyze, Improve, and Control) for process improvement. We discussed pareto and control charts as well as chi squared and confidence intervals. We also discussed correlation, regression and descriptive statistics. This helped me understand the value of data collection and analysis in acquiring knowledge while making decisions in today's business environment. I was ultimately able to identify and apply the appropriate statistical technique for a given set of conditions in order to answer a particular question.

In my example, I applied what I learned to a real-world example at work and was able to identify no value add (NVA) steps in my process by utilizing process mapping. I was able to measure the defects, sigma quality level, and cycle time. I reduced errors, increased productivity, and ultimately improved the user experience. I still use this new process at work today and I am very happy with this improvement.

SCM651 Business Analytics

This course as it relates to the program learning goals include: 💾 🔃 👷 👨‍💻 📢 🔒
📂 Portfolio Project Folder (SCM651)
|- HW4_Team2.docx
|- HW4_Team2.xlsx
|- SCM651_Document_File.txt

This class challenged me with using various tools to collect and organize data (MS Excel, MS Access, Google Analytics, R, and Tableau). We used these tools to identify patterns in the data via visualization, statistical analysis, and data mining. We developed strategies based on the data and put a plan in action to influence business decisions.

We covered many items in this class but I think the overall idea was for us to see that there are many tools at our disposal for data analysis and visualization. Using data in real-world scenarios to solve problems including correlations and exponential regression through fraud detection identifying outliers with Benford's Law. The homework example shows our review/analysis using regression, neural network, and sensitivity analysis, to determine which coefficients would most impact receiving a loan.

Conclusion

Completing this master's degree in applied data science is a significant achievement for me that reflects both my passion for the field and my commitment to professional growth. Throughout my academic journey, I have refined my analytical skills, explored complex datasets, and developed innovative solutions. This program has equipped me with the knowledge and practical experience needed to tackle real-world challenges, from predictive modeling to optimizing business processes. As I move forward, I am excited to apply my expertise in data science to make meaningful contributions in industry and research.

Syracuse University is a great school academically, but also personally. I enjoyed my time working in 2SU, Blackboard, and utilizing the vast resources available to me as a remote student. My academic advisors were always a phone call away and were with me throughout my journey. The other students were wonderful to get to know and to collaborate with on projects and assignments. Overall, my experience both academically and personally was positive and I will cherish these memories for a lifetime. Thank you.