

Portfolio Milestone

An assessment of learning for Syracuse University's
Applied Data Science program



By: Vladimir Dinolov
M.S. in Applied Data Science Candidate

Introduction

A description of the author's background, location, and industry experience

My name is Vladimir Dinolov and I'm currently living in Dallas, Texas as an educator turned data professional. I spent the greater half of the last decade teaching in underserved communities as a Teach for America and City Year alumni. Armed with experience in both public education and the nonprofit sector, I transitioned into the data science industry as a graduate student at Syracuse University. Since participating in this rigorous program, I have worked as a Policy Analyst and Marketing Analytics professional. The Applied Data Science program from the School of Information Studies, known also as the iSchool, allowed me to gain the necessary skills and experience to transform data into insights.

During my time in the program, I completed courses from both the Whitman School of Management and the iSchool. This interdisciplinary approach to information management and data analysis ensures working professionals have exposure to solving real-world problems from a variety of industries. The following coursework was completed as part of the Applied Data Science curriculum:

School of Information Studies	Applied Data Science Database Administration Data Mining and Machine Learning Big Data Analytics Text Mining Natural Language Processing Scripting for Data Pipelines Quantitative Reasoning and Statistical Analysis Information Visualization
Whitman School of Management	Business Analytics Data Analysis and Decision Making Marketing Analytics

This portfolio serves as a summary and assessment of learning from that coursework as it aligns to the program's core competencies:

Core Competency	Course	Project
Describe a broad overview of the major practice areas of data science	Applied Data Science	All Projects
Collect and Organize Data	Database Administration	Educator Effectiveness Database
Identify patterns in data via visualization, statistical analysis, and data	Information Visualization	Erasing the Opportunity Gap in New York Public Schools
Develop alternative strategies based on the data	Big Data Analytics	SAM Labs Customer Profiling
Develop a plan of action to implement the business decisions derived from the analyses	Big Data Analytics	SAM Labs Customer Profiling
Demonstrate communication skills regarding data and its analysis for relevant professionals	Data Analytics	Optimizing Hydroelectric Energy Supply Chains
Synthesize the ethical dimensions of data science practice	Text Mining Information Visualization	Death Row Final Statements Erasing the Opportunity Gap in New York Public Schools

I look forward to sharing my data science journey at Syracuse with you. Let us begin!

All Projects

Describe a broad overview of the major practice areas of data science

Data science is a unique, interdisciplinary field that is built on foundational knowledge and skills like statistics and information management. According to my introductory data science course, a successful data scientist should have a strong background in the “*collection, preparation, analysis, visualization, management, and preservation of large collections of information*”. To remain competitive in this ever changing industry, data science practitioners must continuously grow and adapt to new technologies. The artifacts in this portfolio showcase both traditional and novel data science techniques in the context of this program’s core competencies.

Educator Effectiveness Database

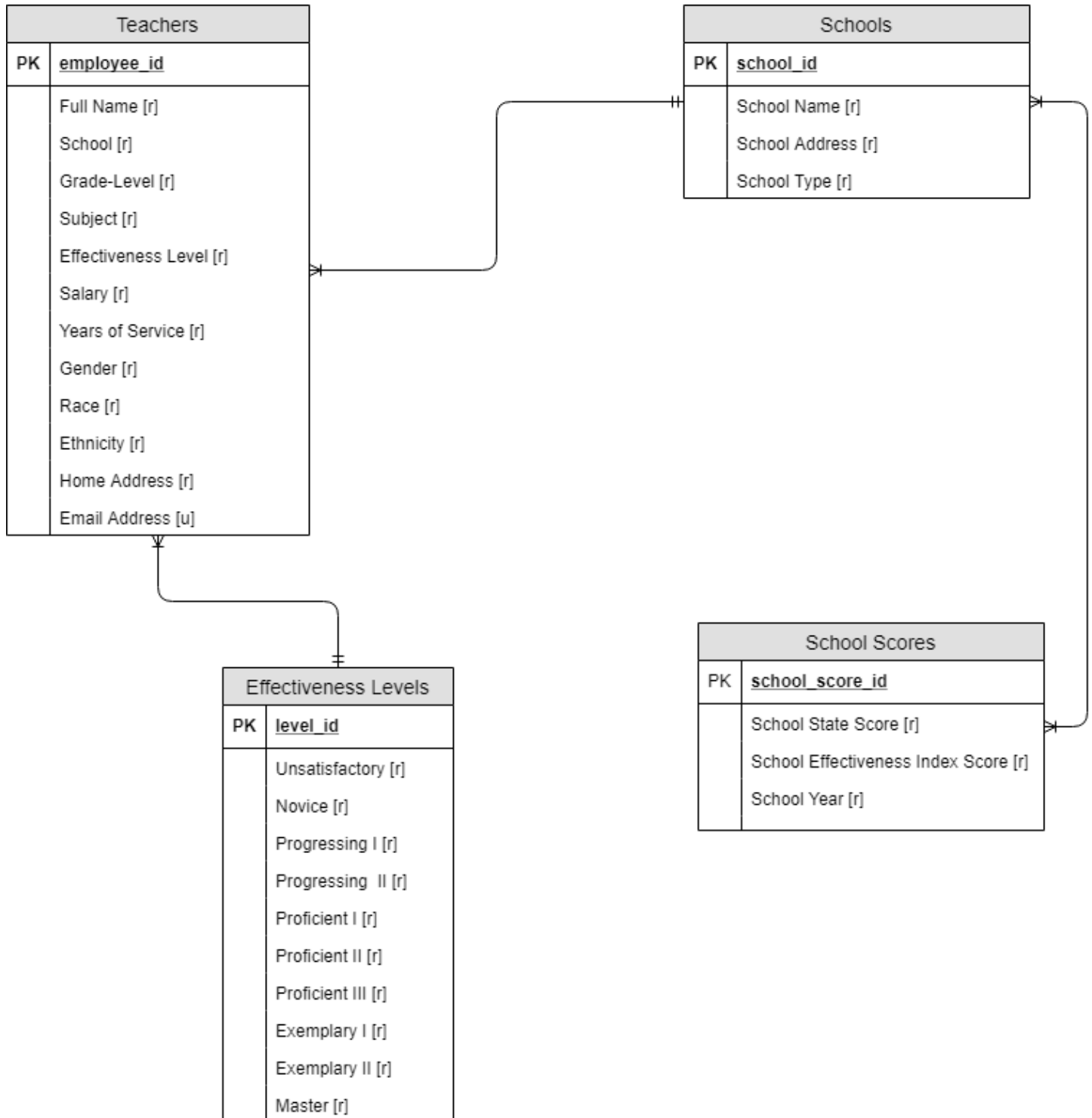
Collect and Organize Data

Much of my data science applications focus on public education. One such application was the creation of a prototype for the Dallas Independent School District (ISD) teacher retention program. Unfortunately, many U.S. school districts have teacher shortages that stem from many issues, the most salient being teacher compensation. In an effort to mitigate this shortage locally, Dallas ISD created the “Teacher Excellence Initiative”. This strategic initiative rewards high-performing teachers with increased compensation up to \$100,000. To understand the success of the program, I created a teacher effectiveness database that gathers, stores and allows others to extract information on Dallas ISD educators.

The Educator Effectiveness Database creates a centralized location for teacher retention data ranging from effectiveness level to school site. Much of Dallas ISD’s data is in spreadsheets or Excel files but cannot be accessed in a database. This database helped inform district trustees and school leaders on the Teacher Excellence Initiative’s success in recruiting and retaining educators. The district most recently reported a 97% teacher retention rate.

This project showcased my ability to collect, manage, and preserve data over time. Learning how to create a relational database is one of the many data science skills I picked up in the program. Creating a wireframe, logic model, data dictionary, and user interface were essential steps to make the database usable and actionable.

Competency Two Artifact



VDDINOLO@SYR.EDU

SUID: 2075564

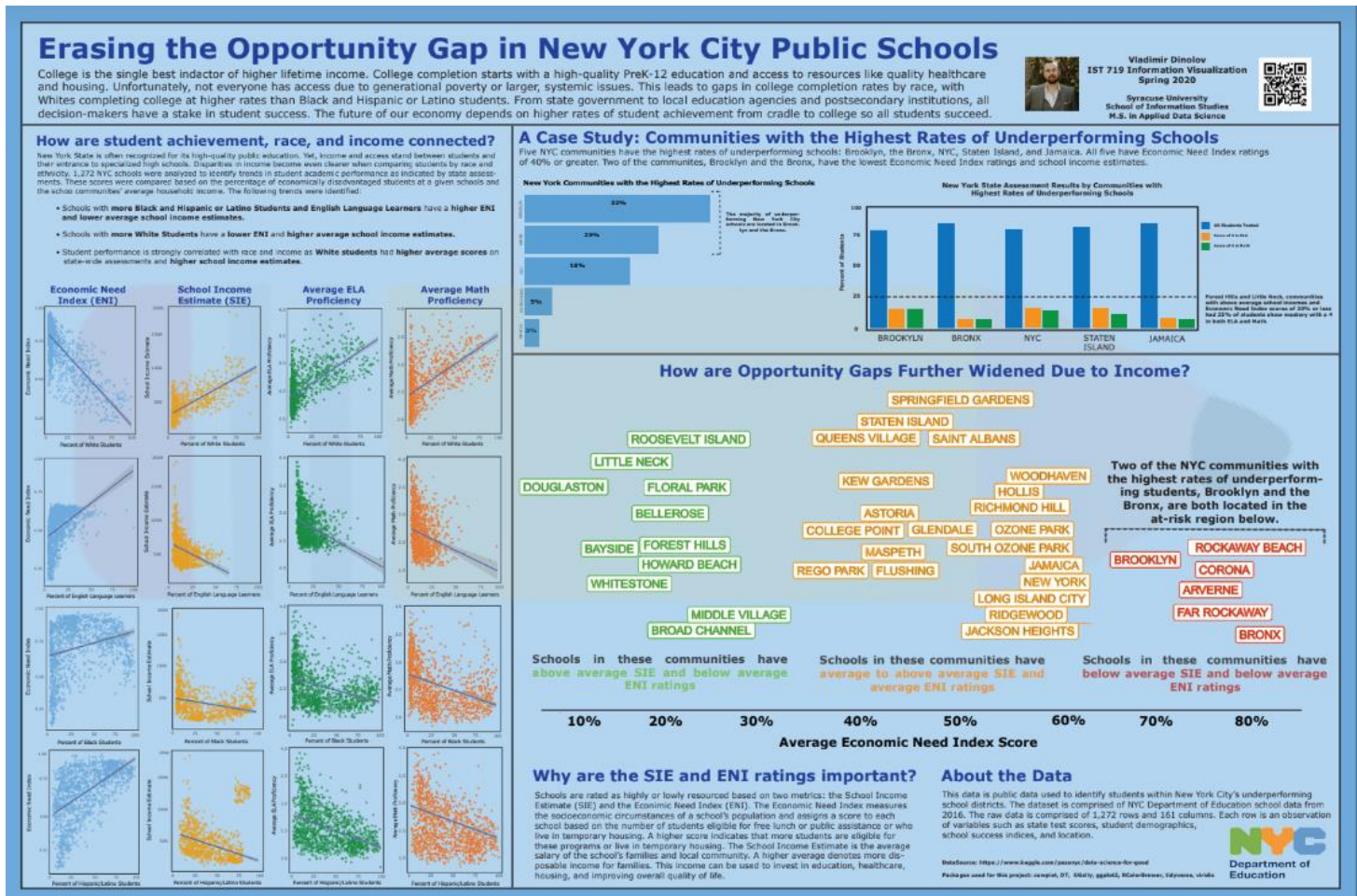
Erasing the Opportunity Gap in NYC Schools

Identify patterns in data via visualization, statistical analysis, and data mining

Continuing the theme of education, the goal of this project was to examine the opportunity gap in New York City public schools in relation to students' socioeconomic backgrounds, race, and geographies. I **gathered, prepared, analyzed, and visualized** the supplied geodemographic and achievement data to inform nonprofit decision makers on the most optimal location for afterschool programs. I was able to cluster schools by two engineered features: the economic need index (ENI) and school income estimate (SIE). I then identified which schools had the most at risk students based on performance, ENI scores, and SIEs.

I appreciated the opportunity to use R as a statistical programming language and visualization software. It has many robust packages for data visualization that can be enhanced by Adobe Illustrator. Learning about the grammar of graphics and how the brain translates color, shape, size, and position has added significant value to my company.

Competency Three Artifact



SAM Labs Customer Profiling

Develop alternative strategies based on the data

I currently work as a Marketing Analyst for an educational technology company called 'SAM Labs'. The company specializes in hands-on STEAM and coding solutions for K-8 educators. As a Marketing team, we are focused on account-based marketing and segmenting our customers based on behavior, historical purchases, demographics, and finances. Data was gathered from our customer relationship management system, the National Center for Education Statistics, and GovSpend, a spending database that aggregates government agency purchase orders. Statistical models were used to **identify** and **develop alternative marketing strategies** based on competitor purchases and their significance in predicting a SAM Labs purchase.

This research produced a list of 100 school districts meeting the criteria for all or some of the statistically significant variables. Examples include purchases from specific competitors, the category of the item purchased, the type of school district purchasing, and the school district's

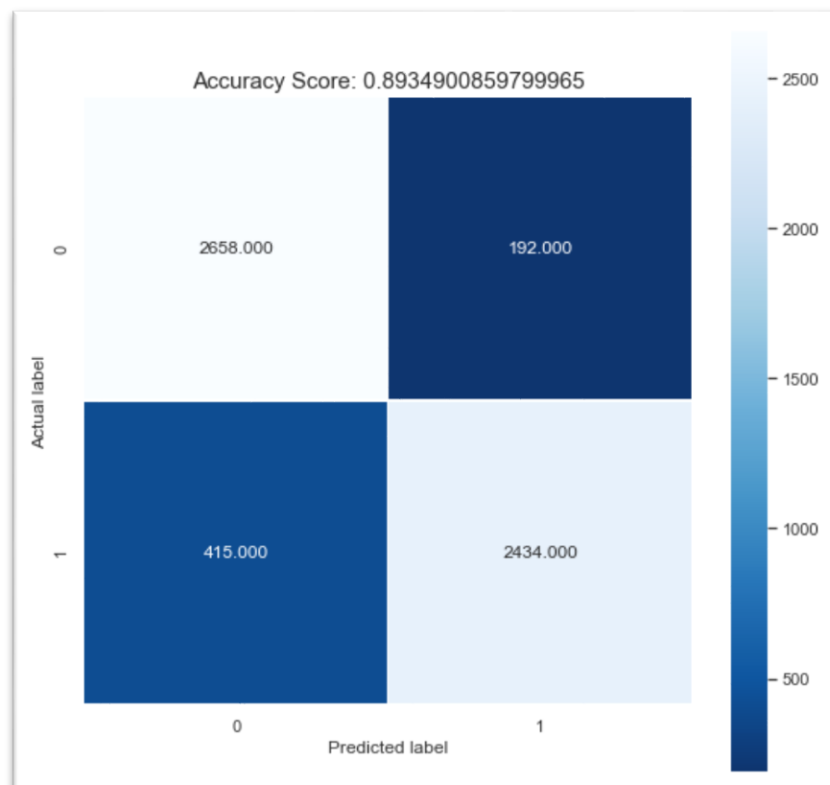
VDDINOLO@SYR.EDU

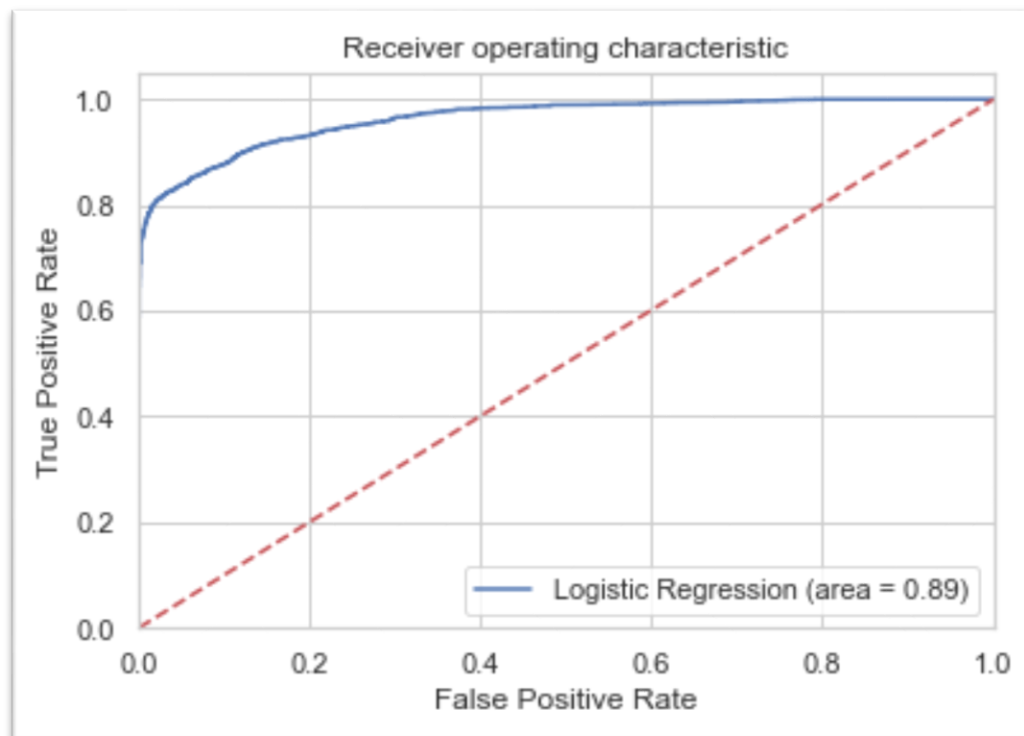
SUID: 2075564

expenditures and revenue.

The Marketing team has used this insight to shift its approach in nurturing leads, creating smart content, and reengage segments of customers that meet our customer profiles.

Competency Four Artifacts





SAM Labs Customer Profiling Redux

Develop a plan of action to implement the business decisions derived from the analyses


After identifying customer profiles, I assisted the SAM Labs Marketing team in creating a 'Solutions' gateway on our company website. Each solution caters to a different subset of our customers. We also began to target school districts that receive federal STEM grants, as these grants were identified as of the most statistically significant variables associated with a SAM Labs purchase.

This project taught me the value in applying the results of statistical models versus reporting on them only. While communicating model results is important, communicating *how to use the results* is central to data-driven decision making.

Competency Five Artifacts



GRANT APPLICATION GUIDE



'Doing the Right Thing. - To be a teacher is doing the right thing. For the Toshiba America Foundation, doing the right thing means supporting STEM education by funding classroom teachers as they seek a better way to help their students learn.' ~ Toshiba Values

Grant name: Toshiba Grants for Grade 6-12 & K-5

Website: <http://www.toshiba.com/taf/612.jsp> <http://www.toshiba.com/taf/k5.jsp>

Link to apply:
https://www.cybergrants.com/pls/cybergrants/quiz.display_question?x_gm_id=4184&x_quiz_id=4698&x_order_by=1

Grant value: \$1k - \$5k+

Grant level: Classroom

Region: USA

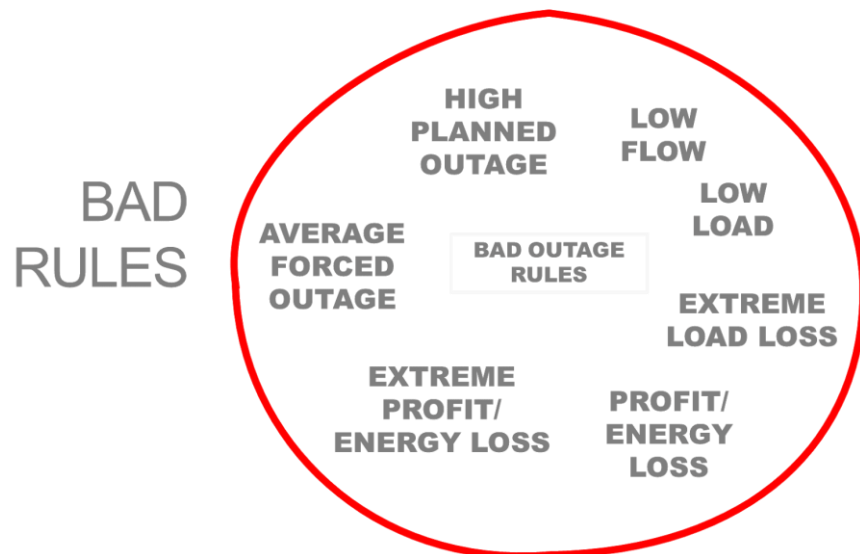
Optimizing Hydroelectric Energy Supply Chains

Demonstrate communication skills regarding data and its analysis for relevant professionals

For this project, I summarized multiple machine learning models and presented the results to a hydroelectric company's board of directors. My team was tasked with identifying how to improve operational efficiency at the company's hydro plants. We identified the ideal outage schedule that would result in maximizing revenue and minimizing costs. Variables such as water supply flow, load units of energy, and the rate of forced or planned outages impacted the amount of energy produced.

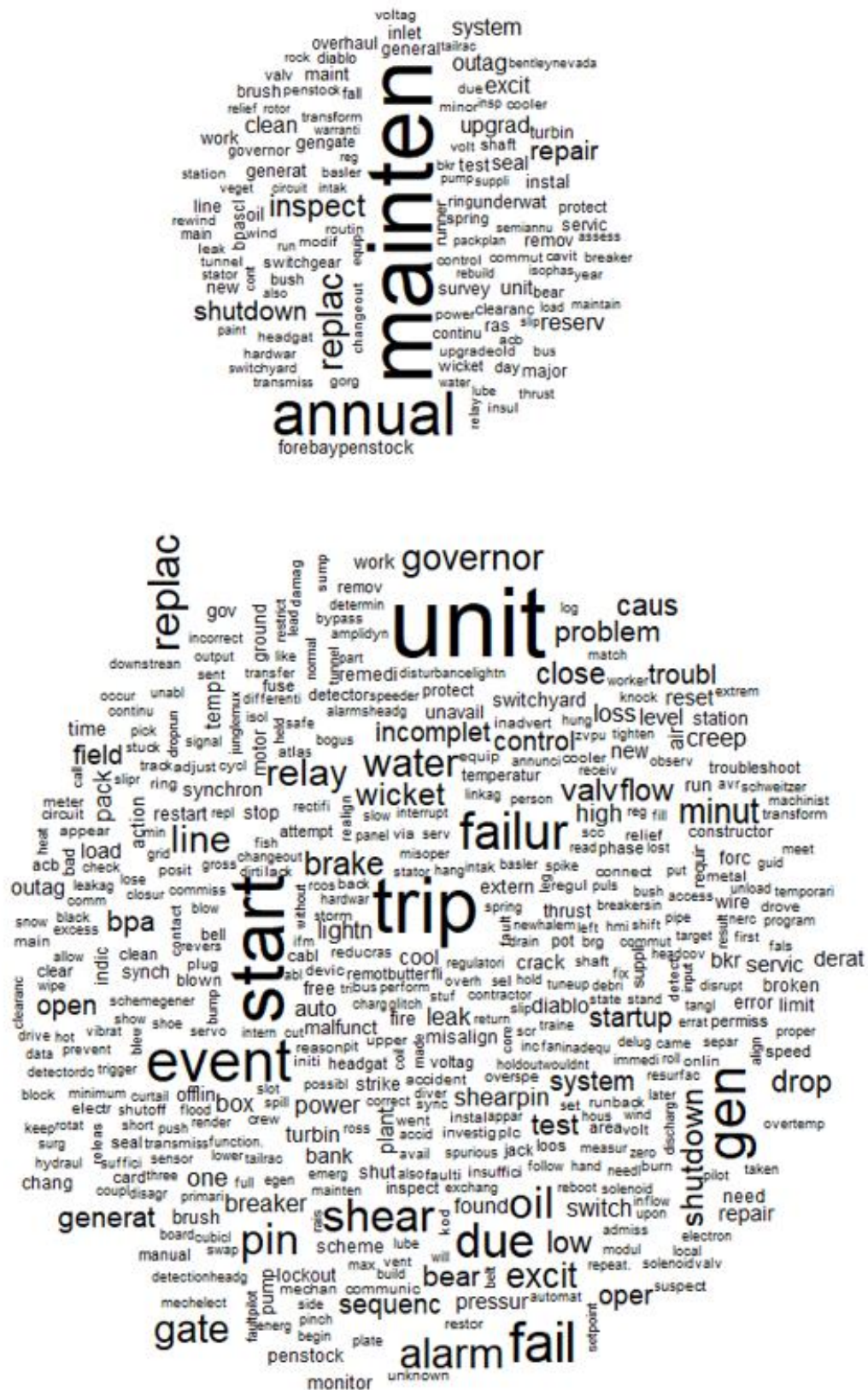
I was particularly proud of my ability to use word clouds and association rules to identify the root issue behind energy deficits and revenue loss. By making the information easily digestible, decision makers were able to move swiftly in finding optimization solutions for the company's energy supply chain.

Competency Six Artifacts



VDDINOLO@SYR.EDU
SUID: 2075564

Word clouds showing frequent words from description of planned (top) and forced (bottom) outages



Multiple Projects

Synthesize the ethical dimensions of data science practice

There is inherent power in the acquisition and usage of data. Data governance and ethics are widely debated across all industries, especially when data is applied in the social sciences. All data, especially data that summarizes populations, behaviors, and viewpoints, must be carefully examined for possibilities of bias or error. As data professionals, we must address our own biases and examine how we apply data to participate in self-serving bias on behalf of our employer or clients.

Furthermore, data science practitioners are obligated to accurately represent data in every step of the pipeline. We must consider how data can be used to positively change the world and ensure data is not used to further harmful rhetoric. This is most important in fields like education, medicine, economics, environmental science, and politics.

Professors like Dr. Jeffrey Saltz and Dr. Ami Gates cite how data science can be used for social good. I couldn't agree more. Data is a tool, a currency, and most importantly, a vessel for change.

Closing Thoughts

Reflecting on the author's Syracuse experience

As I reflect on my time in the Applied Data Science program at Syracuse, I find myself feeling grateful, excited, and inquisitive. I have developed foundational skills that will further my career and enhance my understanding of the world around me.

I have learned about the applications of Bayesian and frequentist statistics, the usage of programming languages to extract insights, the design principles of data visualization, and how to best communicate results to drive impact. With these concepts and skills, I will continue my journey as a data professional that collects, prepares, analyzes, visualizes, manages, and preserves data.