Applied Data Science Portfolio

Andrew Krick

Master of Science in Applied Data Science

Syracuse University

Fall 2025

Executive Summary

This portfolio represents the culmination of my work in the Master of Science in Applied Data Science program at Syracuse University and serves as a comprehensive demonstration of the technical, analytical, and professional skills I developed throughout the program. The purpose of this portfolio is to document how I have met the six program learning outcomes through applied coursework and project-based learning, while also presenting a cohesive professional narrative suitable for prospective employers.

Across the program, I completed projects spanning machine learning, natural language processing, data mining, relational database management, and cloud architecture. These projects required the practical application of Python, R, and SQL to real datasets and problem contexts. Rather than focusing solely on algorithmic implementation, the work emphasized data acquisition, preprocessing, model evaluation, interpretation of results, and effective communication of insights. This portfolio highlights those applied dimensions and illustrates how individual technical skills combine to support data driven decision making.

Several projects included in this portfolio explicitly address considerations related to data governance, security, and ethical responsibility. These aspects are increasingly central to modern data science practice and are treated as integral components of applied analytics rather than as secondary concerns. By incorporating these considerations alongside technical work, the portfolio reflects a holistic understanding of what it means to operate as a professional Data Scientist in industry.

While this portfolio fulfills an academic requirement, it is also designed as a professional artifact. The structure, content, and tone are intentionally aligned with how Data Scientists are expected to communicate their experience and value in professional settings. Each project included represents work I personally completed and reflects my progression from foundational analytical skills to more advanced modeling, system design, and strategic reasoning. Collectively, the projects demonstrate readiness to contribute meaningfully in a Data Scientist role.

Introduction and Professional Context

I pursued the Master of Science in Applied Data Science program to build strong, practical capabilities in data analysis, machine learning, and analytics driven problem solving. My objective throughout the program was not only to learn individual tools and techniques, but to understand how they are applied together in real organizational contexts. This portfolio reflects that objective by presenting integrated evidence of technical proficiency, analytical judgment, and professional communication.

Throughout the program, I worked extensively with Python, R, and SQL across a range of problem domains. These included supervised machine learning, text classification, exploratory data analysis, relational database design, and cloud based system architecture. In each case, the emphasis was on applying methods to realistic datasets and producing interpretable results rather than executing isolated technical exercises. This applied orientation shaped both my learning process and the structure of this portfolio.

As I progressed through the curriculum, my approach to data science evolved. Early projects focused primarily on implementing methods correctly and understanding model mechanics. Later work emphasized evaluation rigor, reproducibility, system level thinking, and the communication of insights to non technical stakeholders. The projects included in this portfolio reflect that progression and demonstrate how skills developed across courses reinforced and extended one another.

This portfolio is organized around the six Applied Data Science program learning outcomes. Each outcome is supported by evidence drawn from multiple projects completed across different courses. Organizing the portfolio by learning outcome mirrors professional practice, where individual projects rarely map to a single skill in isolation. Instead, effective data science work integrates data acquisition, modeling, evaluation, communication, and ethical reasoning. Presenting the portfolio in this way allows for a clear demonstration of that integration.

Portfolio Organization and Project Overview

This portfolio includes five major projects completed across the Applied Data Science curriculum. Each project is referenced throughout the paper and mapped to one or more program learning outcomes. While full code, datasets, and supporting materials are maintained externally and linked where appropriate, the focus of this paper is on explaining analytical outcomes, technical decisions, and lessons learned rather than reproducing complete technical artifacts.

The projects included are intentionally selected to represent a broad and balanced set of competencies. Together, they demonstrate experience working with structured and unstructured data, applying predictive modeling techniques, designing and querying relational databases, and considering data systems and governance at scale. The five projects included in this portfolio are:

The **IST 664 Text Classification Final Project** applied natural language processing techniques to fine grained sentiment classification using the Kaggle Movie Reviews dataset. This project emphasized feature engineering, model comparison, and evaluation, and demonstrated how textual data can be transformed into structured inputs for predictive modeling.

The **IST 707 Applied Machine Learning Project** focused on building, tuning, and evaluating predictive models using structured data. This project highlighted disciplined data preparation, cross validation, and performance assessment, reinforcing best practices in applied machine learning workflows.

The **IST 736 Data Mining Project** emphasized exploratory analysis, feature engineering, and model comparison across multiple algorithms. This work reinforced analytical reasoning and the importance of selecting methods appropriate to the data and problem context.

The **IST 615 Cloud Architecture Final Project** addressed data storage, system design, security, and governance considerations within a cloud based solution. This project expanded my perspective beyond individual analyses to include system level thinking and architectural decision making.

Finally, the **IST 659 Data Administrative Concepts and Database Management Project** demonstrated relational database design and SQL based data access. Through schema development and query implementation, this project illustrated how structured data systems support analytical workflows and decision making.

Together, these projects provide comprehensive evidence of technical proficiency, analytical reasoning, and readiness for professional data science work.

Program Learning Outcomes Roadmap

The Applied Data Science program defines six learning outcomes that graduates are expected to demonstrate. This portfolio addresses each outcome explicitly using concrete evidence from the projects described above. The outcomes serve as the organizing framework for the remainder of this paper.

The six program learning outcomes addressed are:

1. Ability to collect, store, and access data using appropriate technologies
2. Ability to analyze data and generate actionable insights
3. Ability to communicate insights through visualization and narrative
4. Ability to apply predictive modeling techniques and evaluate model performance
5. Proficiency in programming languages commonly used in data science
6. Understanding of ethical, privacy, and responsible data science considerations

Each of the following sections of this paper is dedicated to one of these outcomes. Within each section, I describe how specific projects demonstrate the outcome, reference relevant analytical outputs, and reflect on lessons learned. This structure ensures that all outcomes are fully addressed while also highlighting how individual projects contributed to multiple areas of competency.

Program Learning Outcome 1: Data Acquisition, Storage, and Access

The first program learning outcome focuses on the ability to collect, store, and access data using appropriate technologies. This competency is foundational to applied data science, as all downstream analysis and modeling depend on the quality, structure, and accessibility of data. Throughout the Applied Data Science program, I demonstrated this outcome through projects involving relational database design, structured data pipelines, and cloud based system architecture.

A primary demonstration of this outcome comes from my work in **IST 659: Data Administrative Concepts and Database Management**. In this course, I designed and implemented a relational database schema intended to support structured data storage and retrieval. The associated SQL project required defining tables with appropriate primary and foreign keys, enforcing referential integrity, and writing queries to extract meaningful information from stored data. This work emphasized the importance of thoughtful schema design and normalization, as poor structural decisions can significantly limit analytical flexibility and performance.

Beyond schema design, the SQL project focused on practical data access through query development. I wrote queries that involved filtering, joining, grouping, and ordering data to support analytical use cases. These queries transformed raw stored data into structured outputs suitable for reporting and analysis. This project reinforced the role of SQL as a core skill for Data Scientists, particularly in enterprise environments where data is often housed in relational systems rather than delivered in analysis ready formats.

This learning outcome was further reinforced through my **IST 707 Applied Machine Learning Project**, which required disciplined handling of structured datasets. Although the storage mechanisms were simpler than a full database system, the project emphasized proper data loading, validation, and partitioning. Careful attention was paid to how data was split into training and evaluation sets to avoid leakage and ensure reproducibility. These practices reflect applied data acquisition and access skills that are essential for reliable modeling and analysis.

The most comprehensive illustration of this outcome occurred in my **IST 615 Cloud Architecture Final Project**. In that project, I designed a cloud-based system that addressed data ingestion, storage, access control, and security. The architecture incorporated cloud storage services, access management policies, and defined data flows to support analytics and reporting. Rather than treating data storage as an isolated technical concern, the project emphasized how storage decisions interact with governance, scalability, and security requirements.

This experience highlighted the importance of systems thinking in applied data science. Effective data acquisition and storage require an understanding of how data moves through an organization, who can access it, and how it is protected. The cloud architecture project demonstrated my ability to reason about these factors and design solutions that support analytical work while meeting organizational constraints.

Across these projects, my approach to data acquisition, storage, and access evolved significantly. Early coursework focused on obtaining usable datasets and performing basic preprocessing. Later projects emphasized long term usability, governance, and system level considerations. By integrating relational database design, disciplined data handling in machine learning workflows, and cloud-based architecture, I demonstrated the ability to select and apply appropriate technologies for managing data in applied settings.

Program Learning Outcome 2: Data Analysis and Actionable Insight

The second program learning outcome focuses on the ability to analyze data and generate actionable insights. This outcome goes beyond the mechanical application of statistical or machine learning techniques and emphasizes the interpretation of results in ways that inform decisions, explain patterns, and support conclusions. Throughout the Applied Data Science program, I developed this capability through projects that required not only building models but also understanding what those models revealed about the underlying data and how those insights could be used.

A primary demonstration of this outcome is my **IST 664 Text Classification Final Project**, which involved fine grained sentiment analysis of movie reviews from the Kaggle Movie Reviews dataset. The analytical challenge of this project extended beyond achieving high classification accuracy. I was required to explore the structure of the data, understand how sentiment is expressed linguistically, and determine which features contributed most meaningfully to model performance. Through exploratory analysis, feature engineering, and iterative experimentation, I developed insights into how different representations of text affect predictive outcomes. For example, comparing baseline bag of words features to enriched representations highlighted how incorporating additional contextual and lexical information can significantly improve model interpretability and performance.

This project reinforced the importance of treating models as analytical tools rather than black boxes. By examining misclassified examples and reviewing performance across sentiment categories, I was able to identify systematic weaknesses and strengths in the modeling approach. These analyses provided insight into where the model struggled to capture nuance, such as subtle emotional tone or ambiguous phrasing, and where it performed reliably. This process of evaluation and interpretation directly supported actionable insight by informing decisions about feature refinement and model selection.

Analytical insight generation was also central to my **IST 707 Applied Machine Learning Project**, which focused on structured data. In this project, I worked through the full analytical pipeline, including exploratory data analysis, preprocessing, model training, and evaluation. Rather than simply reporting performance metrics, I analyzed how different features influenced predictions and how preprocessing decisions affected model behavior. This included examining feature distributions, identifying potential sources of bias or noise, and interpreting evaluation results in the context of the problem domain.

A key lesson from this project was the importance of aligning analytical results with the original problem objectives. In several instances, models with marginally higher performance metrics were less desirable due to reduced interpretability or increased sensitivity to noise. By comparing models across multiple evaluation criteria, I was able to generate insights that balanced predictive performance with practical usability. This analytical reasoning reflects the type of judgment required in applied data science roles, where model outputs must be evaluated in context rather than optimized blindly.

Further evidence of this learning outcome comes from my **IST 736 Data Mining Project**, which emphasized exploratory analysis and model comparison. In this work, I applied multiple analytical techniques to the same dataset to understand how different modeling assumptions and feature engineering choices influenced results. This comparative approach provided insight into the relationships present in the data and helped identify which methods were most appropriate for the task at hand.

Through this project, I strengthened my ability to synthesize analytical findings across methods. Instead of treating each model independently, I examined patterns across results to identify consistent signals and discrepancies. This synthesis supported more robust conclusions and reduced the risk of over interpreting isolated results. The project highlighted the value of triangulating insights across multiple analytical approaches, a practice that is particularly important when working with complex or noisy data.

Across these projects, my approach to data analysis evolved from focusing primarily on technical correctness to emphasizing interpretability, context, and decision relevance. Early analytical work tended to center on whether a method worked. Later projects emphasized why it worked, where it failed, and how results should inform next steps. This shift reflects growth in analytical maturity and aligns closely with professional expectations for Data Scientists.

By integrating exploratory analysis, model evaluation, and reflective interpretation, I demonstrated the ability to generate actionable insights from data. These insights extended beyond numerical results to include explanations of patterns, limitations, and implications. Collectively, the projects supporting this outcome show that I can analyze data rigorously while maintaining a focus on meaningful, decision oriented conclusions.

Program Learning Outcome 3: Visualization and Communication

The third program learning outcome focuses on the ability to communicate analytical insights effectively through visualization and narrative. In applied data science, technical results only become valuable when they are presented in a way that supports understanding, decision making, and action. Throughout the program, I developed this competency by designing visualizations, interpreting model outputs, and translating complex analytical findings into clear explanations for both technical and non technical audiences.

A primary demonstration of this outcome is found in my **IST 664 Text Classification Final Project**. In this project, visualization played a critical role in understanding model performance and communicating results. I used visual representations such as confusion matrices, performance comparisons, and feature related summaries to evaluate and explain how different modeling approaches behaved. These visual tools allowed me to identify patterns that were not immediately obvious from numerical metrics alone, such as systematic misclassification between sentiment categories and uneven performance across classes.

Beyond model evaluation, this project emphasized narrative communication. Rather than presenting isolated figures, I structured explanations around what each visualization revealed about the data and the model. For example, performance comparisons were framed in terms of tradeoffs between precision, recall, and interpretability. This approach ensured that visualizations supported a coherent analytical story rather than serving as disconnected technical artifacts.

Visualization and communication were also central to my **IST 707 Applied Machine Learning Project**. In this work, I used exploratory data analysis plots to understand feature distributions, identify potential anomalies, and guide preprocessing decisions. Visual summaries of model performance were used to compare algorithms and justify modeling choices. These visualizations were paired with written explanations that contextualized results within the problem domain, reinforcing the importance of linking analytical output to practical implications.

A key lesson from this project was the need to tailor communication to the audience. While detailed metrics and technical plots are valuable for peer review and model validation, stakeholders often require higher level summaries that emphasize implications rather than implementation details. Through iterative refinement of visualizations and explanations, I improved my ability to balance technical rigor with clarity and accessibility.

Additional evidence of this learning outcome appears in my **IST 736 Data Mining Project**, which emphasized exploratory analysis and comparative evaluation. In this project, visualizations were used to identify relationships among variables, assess feature importance, and compare model behavior across different approaches. By visually examining patterns across multiple models, I was able to synthesize insights that would have been difficult to derive from tabular results alone. This reinforced the role of visualization as a tool for discovery as well as communication.

The **IST 615 Cloud Architecture Final Project** further expanded my communication skills by requiring me to present technical concepts at a systems level. In this project, visual diagrams were used to illustrate data flows, storage components, and security boundaries within a cloud based architecture. These diagrams supported communication with decision makers by conveying complex technical designs in an accessible format. This experience emphasized that effective communication in data science extends beyond charts and graphs to include architectural representations and conceptual visuals.

Across these projects, my approach to visualization and communication matured significantly. Early work focused on generating plots to satisfy analytical requirements. Later projects emphasized intentional design, audience awareness, and narrative coherence. I learned to view visualizations as tools for reasoning, explanation, and persuasion rather than as decorative elements.

By integrating clear visual design with structured narrative explanation, I demonstrated the ability to communicate analytical insights effectively. This capability is essential for professional data science work, where insights must be shared across interdisciplinary teams and used to inform decisions. The projects supporting this outcome show that I can translate complex analytical results into clear, actionable communication.

Program Learning Outcome 4: Predictive Modeling and Evaluation

The fourth program learning outcome focuses on the ability to apply predictive modeling techniques and evaluate model performance rigorously. Predictive modeling is a core responsibility of the Data Scientist role, but effective modeling requires more than selecting an algorithm and reporting accuracy. It involves thoughtful feature design, appropriate model selection, careful evaluation, and an understanding of how modeling decisions affect reliability and interpretability. Throughout the Applied Data Science program, I developed this competency through multiple projects that emphasized experimentation, comparison, and evaluation.

A central demonstration of this outcome is my **IST 664 Text Classification Final Project**, which involved fine grained sentiment classification of movie reviews. This project required transforming unstructured text into structured representations suitable for predictive modeling. I implemented multiple feature engineering strategies and evaluated their impact on model performance. By comparing baseline models to enhanced feature sets, I gained insight into how different representations influence predictive capability and generalization.

Model evaluation played a critical role in this work. Rather than relying on a single performance metric, I evaluated models using multiple measures to understand tradeoffs between precision, recall, and overall performance. I also examined class level behavior to identify where models performed well and where they struggled. This detailed evaluation process allowed me to move beyond surface level performance reporting and toward a more nuanced understanding of model strengths and limitations.

Predictive modeling was also a major focus of my **IST 707 Applied Machine Learning Project**, which centered on structured data. In this project, I implemented and compared multiple supervised learning algorithms within a consistent evaluation framework. Careful attention was paid to preprocessing, feature scaling, and training procedures to ensure fair comparisons between models. Cross validation was used to assess generalization performance and reduce the risk of overfitting.

This project reinforced the importance of disciplined evaluation practices. I learned that small differences in performance metrics must be interpreted cautiously and in context. In several cases, simpler models offered competitive performance while providing greater stability or interpretability. Evaluating models through multiple lenses helped inform decisions about which approaches were most appropriate for the problem domain rather than simply selecting the highest scoring option.

Additional evidence of this learning outcome comes from my **IST 736 Data Mining Project**, which emphasized model comparison and feature engineering. This project involved applying multiple predictive techniques to the same dataset and analyzing how modeling assumptions affected results. By systematically comparing models, I developed a stronger intuition for how data characteristics influence predictive performance and when certain algorithms are likely to be effective.

Through this project, I also strengthened my ability to diagnose modeling issues. Performance differences prompted deeper investigation into feature relevance, data quality, and algorithm suitability. This diagnostic mindset is critical in applied settings, where models must be refined iteratively based on evidence rather than treated as static solutions.

Across these projects, my approach to predictive modeling evolved from an initial focus on implementation to a more evaluative and reflective practice. Early work emphasized getting models to run correctly. Later projects emphasized robustness, reproducibility, and interpretability. I became more attentive to how modeling choices align with problem objectives and stakeholder needs.

By applying predictive modeling techniques across diverse datasets and contexts and by evaluating performance rigorously, I demonstrated the ability to build and assess models responsibly. The projects supporting this outcome show that I can move beyond surface level modeling to deliver predictive solutions that are analytically sound and practically relevant.

Program Learning Outcome 5: Programming Proficiency in Data Science Tools

The fifth program learning outcome focuses on proficiency in programming languages and tools commonly used in data science. In applied data science practice, programming proficiency is not limited to syntax or isolated scripts. It involves writing reliable, interpretable, and reusable code that supports data acquisition, analysis, modeling, and communication. Throughout the program, I demonstrated this competency through extensive use of Python, R, and SQL across multiple projects and problem domains.

Python served as my primary programming language throughout the program and was central to my work in **IST 664, IST 707, and IST 736**. In the **IST 664 Text Classification Final Project**, Python was used to implement the full natural language processing pipeline, including text preprocessing, feature engineering, model training, and evaluation. This required coordinating multiple libraries and writing code that could be iteratively refined as features and modeling approaches evolved. Through this project, I strengthened my ability to structure code logically, manage experimental comparisons, and interpret model outputs programmatically.

In the **IST 707 Applied Machine Learning Project**, Python was used to support a complete supervised learning workflow for structured data. This included data loading, preprocessing, feature scaling, model training, and evaluation. The project emphasized reproducibility and consistency across experiments, which required disciplined coding practices. Rather than relying on ad hoc scripts, I structured code to allow for controlled experimentation and fair model comparison. This experience reinforced the importance of writing clear and maintainable code when working with complex analytical pipelines.

The **IST 736 Data Mining Project** further expanded my programming proficiency by requiring the application of multiple modeling techniques and exploratory analyses. Python was used to perform feature engineering, implement different algorithms, and compare results systematically.

This project emphasized analytical flexibility and reinforced the need to write code that could adapt to different modeling assumptions without becoming brittle or opaque. Through this work, I gained greater confidence in using programming as a tool for exploration as well as execution.

In addition to Python, I developed practical experience with **SQL** through my work in **IST 659: Data Administrative Concepts and Database Management**. In this project, I wrote SQL scripts to define relational schemas, enforce constraints, and retrieve data through complex queries. This included the use of joins, aggregations, filtering, and ordering to transform stored data into analytically useful outputs. Working directly with SQL reinforced the importance of understanding how data is structured and accessed in relational systems, which is essential for Data Scientists operating in enterprise environments.

The SQL project complemented my Python based analytical work by highlighting how programming languages serve different roles within the data science ecosystem. While Python was used primarily for modeling and analysis, SQL supported data storage, retrieval, and preparation. Understanding how these tools interact strengthened my ability to work across the full data lifecycle rather than focusing on isolated stages.

Although Python and SQL were my primary programming tools, I also developed experience with **R** in the program, particularly for statistical analysis and exploratory work. Exposure to multiple languages reinforced the idea that programming proficiency involves selecting the right tool for the task rather than adhering rigidly to a single language. This flexibility is an important professional skill, as data science work often spans diverse technical environments.

The **IST 615 Cloud Architecture Final Project** further contextualized programming proficiency within larger systems. While this project was less code intensive, it required understanding how programming scripts, data pipelines, and analytical tools fit into broader architectures. This perspective reinforced the importance of writing code that integrates cleanly with storage systems, access controls, and organizational workflows.

Across the program, my programming skills evolved from writing functional scripts to developing structured, readable, and adaptable code. I became more attentive to clarity, organization, and reproducibility, recognizing that professional data science work often involves collaboration and long-term maintenance. By applying Python, R, and SQL across multiple projects and contexts, I demonstrated proficiency in the core programming tools expected of a Data Scientist.

Program Learning Outcome 6: Ethics, Privacy, and Responsible Data Science

The sixth program learning outcome focuses on understanding ethical, privacy, and responsibility considerations in applied data science. As data driven systems increasingly influence decisions affecting individuals and organizations, Data Scientists must be aware of the ethical implications of their work. Throughout the Applied Data Science program, I developed this awareness by examining data governance, security, bias, and accountability in both technical and organizational contexts.

A primary demonstration of this outcome is my **IST 615 Cloud Architecture Final Project**, which explicitly addressed issues of data security, access control, and governance within a cloud based system. This project required evaluating how data is stored, who has access to it, and how security mechanisms protect sensitive information. Designing the architecture involved considering authentication, authorization, and data protection practices, emphasizing that responsible data science begins at the infrastructure level rather than only at the modeling stage.

Through this project, I gained an appreciation for how architectural decisions can either mitigate or amplify ethical risk. Poorly designed systems can expose sensitive data, enable unauthorized access, or obscure accountability. By incorporating governance and security considerations into the system design, the project demonstrated how ethical responsibility is embedded in technical choices and operational practices. This perspective reinforced the idea that Data Scientists must collaborate with engineers, security teams, and organizational leaders to ensure responsible use of data.

Ethical considerations were also present in my modeling focused projects, particularly in **IST 664, IST 707, and IST 736**. In these projects, I worked with datasets that required careful interpretation and evaluation. Issues such as class imbalance, representativeness, and model bias emerged during analysis and evaluation. Addressing these issues required more than technical adjustments; it required reflection on how modeling choices could affect downstream decisions and stakeholders.

For example, evaluating model performance across different classes highlighted how aggregate accuracy metrics can obscure uneven performance. This reinforced the importance of examining results at a granular level to identify potential biases or unintended consequences. By analyzing misclassifications and performance disparities, I developed a more responsible approach to model evaluation that considers fairness and transparency alongside predictive accuracy.

The **IST 659 Data Administrative Concepts and Database Management Project** further contributed to this outcome by emphasizing data stewardship and integrity. Designing relational schemas and enforcing constraints highlighted the importance of maintaining accurate, consistent, and well governed data. Ethical data science depends on reliable data foundations, as errors or inconsistencies in stored data can propagate through analyses and lead to flawed conclusions.

Across the program, I learned that responsible data science is not a single step or checklist, but an ongoing mindset. It involves questioning assumptions, acknowledging limitations, and communicating uncertainty honestly. This mindset was reinforced through reflection on project outcomes, challenges encountered, and lessons learned. Recognizing when models may not be appropriate or when data limitations constrain conclusions is as important as producing technically correct results.

By integrating ethical, privacy, and governance considerations into my analytical and technical work, I demonstrated an understanding of responsible data science practice. The projects supporting this outcome show that I can balance technical capability with professional judgment, ensuring that data driven solutions are not only effective but also trustworthy and aligned with organizational and societal expectations.

Integrated Program Reflection and Learning Synthesis

The Applied Data Science program provided a structured yet flexible environment in which I was able to develop both technical proficiency and analytical judgment. Looking across the projects included in this portfolio, a clear progression emerges in how I approached data problems, structured analytical workflows, and interpreted results. This reflection synthesizes those experiences and highlights how my understanding of applied data science matured throughout the program.

At the beginning of the program, my focus was largely on learning individual tools and techniques. Early projects emphasized understanding syntax, implementing algorithms correctly, and becoming comfortable working with data programmatically. While these skills were essential, my early work tended to treat analysis as a sequence of tasks rather than as an integrated process. Over time, coursework and project requirements pushed me to think more critically about why specific methods were appropriate, how assumptions influenced outcomes, and how analytical decisions aligned with broader objectives.

One of the most significant areas of growth involved problem framing. In later projects, particularly in IST 664, IST 707, and IST 736, I learned to spend more time understanding the data and the problem context before selecting methods. This included examining data distributions, identifying potential sources of bias or noise, and clarifying what constituted a meaningful outcome. As a result, my analyses became more intentional and better aligned with real world decision making rather than narrowly focused on technical performance.

Another key area of development was evaluation rigor. Early projects often relied on single performance metrics or limited validation approaches. As I progressed, I adopted more disciplined evaluation strategies, including cross validation, class level performance analysis, and comparative model assessment. These practices allowed me to identify tradeoffs, diagnose weaknesses, and make informed decisions about model selection. This shift reflected a broader understanding that effective data science requires judgment and skepticism, not just computation.

The program also strengthened my ability to integrate technical work within larger systems. Projects such as the IST 615 Cloud Architecture Final Project and the IST 659 database project emphasized that data science does not occur in isolation. Data acquisition, storage, access control, and governance all shape what analyses are possible and how results are used. These projects expanded my perspective beyond individual models and highlighted the importance of systems thinking in applied data science roles.

Communication emerged as another area of significant growth. Early analytical work often focused on producing correct results without fully considering how those results would be interpreted by others. Through repeated practice with visualizations, written explanations, and presentations, I improved my ability to translate technical findings into clear narratives. This included learning how to tailor explanations to different audiences and emphasize implications rather than implementation details.

Perhaps most importantly, the program reinforced the importance of ethical and responsible data science practice. Across multiple projects, I encountered issues related to data quality, bias, security, and governance. Addressing these issues required reflection and professional judgment in addition to technical solutions. By integrating ethical considerations into analytical workflows and system designs, I developed a more holistic understanding of my responsibilities as a Data Scientist.

Taken together, these experiences illustrate a clear evolution in my approach to data science. I moved from focusing on isolated techniques to adopting an integrated, reflective, and context aware practice. This synthesis of technical skill, analytical reasoning, communication ability, and ethical awareness represents the core value of the Applied Data Science program and forms the foundation of my professional readiness.

Professional Positioning and Career Readiness

This portfolio reflects my readiness to perform in a professional Data Scientist role by demonstrating applied technical skills, analytical judgment, and the ability to communicate insights effectively. Across the projects included, I have shown the capacity to work with structured and unstructured data, apply predictive modeling techniques, design and query databases, and reason about data systems and governance. Together, these experiences position me to contribute meaningfully in data driven organizations.

My technical foundation spans Python, R, and SQL, which I used across multiple projects to support data acquisition, analysis, modeling, and evaluation. Python served as my primary language for machine learning, text analytics, and exploratory analysis, while SQL supported relational data management and access. Exposure to R further strengthened my statistical reasoning and reinforced flexibility in tool selection. This combination of skills reflects the practical expectations placed on Data Scientists working in modern analytical environments.

Beyond technical implementation, the projects in this portfolio demonstrate the ability to apply analytical reasoning in context. Rather than treating models as isolated outputs, I consistently evaluated results in relation to problem objectives, data limitations, and potential implications. This included comparing models using multiple evaluation criteria, diagnosing performance issues, and interpreting results in ways that inform decision making. These practices align with professional expectations for responsible and effective data science work.

Communication and collaboration readiness are also evident throughout the portfolio. I developed experience translating technical results into clear explanations supported by visualizations and narrative structure. This capability is essential in interdisciplinary environments where insights must be shared with stakeholders who may not have technical backgrounds. The ability to present findings clearly and confidently supports collaboration, trust, and informed decision making.

The inclusion of projects focused on data systems, cloud architecture, and governance further strengthens my professional positioning. These experiences demonstrate awareness of how data science operates within larger organizational and technical contexts. Understanding data storage, access control, security, and ethical considerations allows me to approach analytical work with a systems level perspective rather than focusing narrowly on individual tasks.

Collectively, the work presented in this portfolio reflects a balance of technical capability, analytical judgment, and professional awareness. It demonstrates readiness to enter a Data Scientist role and continue developing through applied experience. The portfolio also serves as a foundation for ongoing professional growth, providing concrete evidence of skills that can be extended and refined as new challenges arise.

Conclusion

This portfolio represents a comprehensive demonstration of my completion of the Master of Science in Applied Data Science program and my readiness to operate as a professional Data Scientist. Through the projects included, I have shown how the knowledge and skills developed across the curriculum align with the six program learning outcomes and translate into applied, real world capability.

Across the program, I developed proficiency in Python, R, and SQL and applied these tools to problems involving structured and unstructured data. I gained experience with predictive modeling, evaluation, data acquisition, database management, visualization, and cloud based system design. More importantly, I learned how to integrate these technical skills into coherent analytical workflows that support insight generation and decision making. The projects in this portfolio illustrate that integration and demonstrate growth from foundational technical execution to more reflective and intentional analytical practice.

The organization of this portfolio around program learning outcomes highlights how individual projects contributed to multiple areas of competency. Rather than treating skills in isolation, the portfolio reflects how applied data science work requires the combination of data management, modeling, evaluation, communication, and ethical awareness. This structure mirrors professional practice and reinforces the value of a holistic approach to data science.

In addition to technical growth, the program emphasized professional judgment and responsibility. Through coursework and projects, I engaged with issues related to data quality, bias, security, and governance. These experiences reinforced the importance of responsible data science and the role of Data Scientists in ensuring that analytical solutions are not only effective but also trustworthy and aligned with organizational and societal expectations.

This portfolio is intended to serve both as an academic capstone and as a professional artifact. It provides concrete evidence of my ability to perform the work expected of a Data Scientist and to communicate that work clearly to technical and non-technical audiences. As I transition from the program into professional practice, the skills and experiences documented here form a strong foundation for continued growth and contribution in data driven roles.

Appendix A — Project Index and Artifact Links

Project 1: Text Classification and Sentiment Analysis

- **Course:** IST 664 — Natural Language Processing / Text Mining
- **Project Title:** Fine-Grained Sentiment Classification of Movie Reviews
- **Description:**

  This project applied natural language processing techniques to classify movie reviews by sentiment using the Kaggle Movie Reviews dataset. The work focused on text preprocessing, feature engineering, model comparison, and evaluation. Multiple feature representations were tested, and results were analyzed to understand model strengths and limitations.
- **Primary Skills Demonstrated:**

  Text preprocessing, feature engineering, supervised learning, model evaluation, Python
- **Artifacts:**
  - Jupyter Notebook: IST664_Andrew_Krick_FinalProject.ipynb
  - Dataset: Kaggle Movie Reviews
- **Access Link:** github.com/atkrick

Project 2: Applied Machine Learning

- **Course:** IST 707 — Applied Machine Learning
- **Project Title:** Supervised Machine Learning Model Development and Evaluation
- **Description:**

  This project focused on building, tuning, and evaluating predictive models using structured data. Emphasis was placed on data preparation, feature scaling, cross-validation, and comparative model assessment. Results were interpreted in relation to problem objectives and practical usability.

- **Primary Skills Demonstrated:**

  Machine learning workflows, evaluation metrics, model comparison, Python

- **Artifacts:**
  - Jupyter Notebook: Andrew_Krick_IST_707_Project.ipynb
  - Supporting Analysis: Andrew_Krick_IST_707_HW-1 - Colab.pdf
- **Access Link:** github.com/atkrick

Project 3: Data Mining and Modeling

- **Course:** IST 736 — Data Mining
- **Project Title:** Feature Engineering and Model Comparison
- **Description:**

  This project emphasized exploratory data analysis, feature engineering, and comparative evaluation of multiple modeling techniques. The goal was to understand how different assumptions and methods influence analytical outcomes.

- **Primary Skills Demonstrated:**

  Data mining, exploratory analysis, model evaluation, Python

- **Artifacts:**
  - Jupyter Notebook: IST736_Project.ipynb
- **Access Link:** github.com/atkrick

Project 4: Cloud Architecture and Data Systems

- **Course:** IST 615 — Cloud Management

- **Project Title:** Cloud-Based Data Architecture and Governance Solution
- **Description:**

  This project involved designing a cloud-based system to support data ingestion, storage, access, and security. The architecture addressed governance, scalability, and ethical considerations related to data handling.
- **Primary Skills Demonstrated:**

  Cloud architecture, data storage, security, governance, systems thinking
- **Artifacts:**
  - Final Report: IST615 Final Project.pdf
- **Access Link:** github.com/atkrick

Project 5: Database Management and SQL

- **Course:** IST 659 — Data Administrative Concepts and Database Management
- **Project Title:** Relational Database Design and SQL Query Implementation
- **Description:** This project demonstrated relational database design, schema development, and SQL-based data access. Queries were written to retrieve, join, and aggregate data in support of analytical use cases.
- **Primary Skills Demonstrated:**

  SQL, database design, data integrity, data access
- **Artifacts:**
  - SQL Script: Krick_SQL up-down-script.sql
- **Access Link:** github.com/atkrick

Andrew Krick

Chicago, IL • atkrick@gmail.com • linkedin.com/in/andrew-krick-data-science
github.com/atkrick • YellowLabAI.com

Education

M.S., Applied Data Science – Syracuse University, 2025
GPA: 3.7 / 4.0
Coursework: Machine Learning, Causal Inference, NLP, Big Data Analytics
B.A., Political Science – The Ohio State University

Skills

Languages/Tools: Python, R, SQL, AWS, GCP, Git, Tableau, Power BI
Methods: Machine Learning, A/B Testing, Experimentation, Causal Inference, NLP, PCA, Clustering, Deep Learning, Feature Engineering

Data Science Projects

- GPT Lens (YellowLab AI) – Founded and developed a desktop AI assistant that detects screenshots and triggers GPT-powered workflows (summarization, explanation, code generation). Built with Electron, Python, and OpenAI API; deployed beta version that improved workflow efficiency by 30%.
- ESS Report Flow (Energy Safety Response Group) – Partnered with firefighters and engineers to design AI-powered pipeline converting unstructured incident data (text, PDFs, images) into compliance-ready reports. Integrated OCR, form understanding, and validation checks, reducing reporting labor by 40%.
- Legislative Bill Success Predictor – Built classification models in Python/R to forecast bill passage with >70% accuracy. Applied feature engineering, cross-validation, and interpretability methods; delivered actionable insights to policy stakeholders.
- Customer Segmentation in E-Commerce – Conducted PCA on high-dimensional features and K-Means clustering, identifying 4 distinct customer groups. Supported simulated marketing campaigns with 15% lift in engagement.
- Stakeholder Simulation via ML (AWS) – Deployed ML models on AWS to simulate stakeholder behavior. Applied experimental design to evaluate different interaction scenarios.

Professional Experience

Founder & President, YellowLab AI – Chicago, IL *(2023 – Present)*
- Delivered automation and experimentation solutions for niche service firms, including GPT-powered prototypes and agents that reduced workflows by >40%.
- Designed experiment pipelines (A/B, champion–challenger) and collaborated with engineering and to align AI with strategy.

Director of Legislative Affairs, Van Meter, Ashbrook & Associates – Columbus, OH *(2014 – 2019)*
- data analyses to guide multimillion-dollar legislative campaigns.
- Produced client-facing insights across healthcare, energy, and transportation sectors.
- Managed 50+ accounts, translating complex data into strategy briefs.