Nicholas Rovelli

IST 782

Portfolio Paper

Throughout my time as a master's student at Syracuse University studying Applied Data Science, I have been able to learn through my classes so many valuable technical skills. These include how to plan a project, where to find reliable data, how to clean data, how to explore and learn about your data, how to model data, and how to effectively communicate findings to a variety of audiences. These experiences have helped build upon what I have learned while studying towards a bachelor's degree in sport analytics. I have also been able to apply many of the skills I have learned while a master's student to my internship at DC United as a data analyst over the summer and towards my work with the Syracuse University Men's Soccer Team as the lead data analyst. As a part of my portfolio, I have included completed work from four of my classes that I have completed before the spring semester of 2025. These classes are IST 687: Introduction to Data Science, IST 659: Data Administration Concepts and Database Management, IST 652: Scripting for Data Analysis, and IST 719: Information Visualization.

In IST 687: Introduction to Data Science, the class was based around the introduction of R Studio as well as the introduction of the concepts of many skills that were to be later honed on through classes taken as a later part of the degree. The first part of this project was planning the format of how we wanted the project to progress and the roles of each person in the group. After that, we explored the data sets to learn more about them and then we cleaned and merged them. Once we had an extensive knowledge of the data, we came up with a question we wanted to answer. From this we used more specific exploratory analysis to answer introductory questions and then figure out what modelling techniques we wanted to incorporate. Once doing this, we landed on a XGBoost model, and from the results of that, we were able to do more exploratory analysis using more scientific findings. Once we have completed all of the work with the data, we needed to find the right way to effectively communicate our findings to the audience through a PowerPoint presentation. We decided to first display a road map of how we manipulated the data, followed by some exploratory analysis so our audience would become more familiar with the data we are using, then our modelling techniques, built upon by more visualizations that used results of the models to provide more answers, and concluding by a synopsis of our results and findings.

In IST 659: Data Administration Concepts and Database Management, we learned how to organize complex data sets through SQL and then create an application of that data. Our group first decided which direction in sports we wanted to go into as we all had sports related backgrounds. Once we decided on hockey, we knew which reliable sources we could get trustworthy data from. Once we did that, we produced questions that we believe stakeholders would be asking and based how we wanted to design our application around that. We first had to decide which platform we wanted to use, and we landed on Tableau. There, we created a dashboard that contained a variety of visualizations for the user as well as allowing the user to create their own visualizations which evaluated the different relationships

our data sets had. Once we completed the Tableau dashboard, we created a presentation for the class that walked through the different relationships our data sets had, questions for potential stakeholders that we were trying to answer, and the inner working of how to navigate our application.

In IST 652: Scripting for Data Analysis, the primary goal of this class is to get meaningful insights out of data using code in an efficient way. For our final project, we made a poster using data we found to analyze the play of FC Barcelona. First, with my group, we discussed our strengths and weaknesses as well as our domain knowledge. Once we discussed, we were able to find data from a reliable source that we believed we could apply many of the concepts that we learned in class to. Once we read in the APIs that were required for our research, we did some exploratory analysis on our own to learn more about the different data set. Once we knew how to attack our questions, we were able to manipulate and merge data in an effective way in python. Once we had a data set that we believed to be easily used, we exported that and brought it into R Studio. There, we were able to create complex soccer specific visualizations to try and answer our questions. Once we were able to tell a story through those visualizations, we picked out the most important once and made a poster that told the story in a way that was made to be digestible by a variety of audiences. In addition to the poster, we wrote further analysis in greater detail on both about the process of how we processed data in addition to more findings to add to our story in a notebook. This project both helped reinforce what I have learned in previous classes, apply much of what I learned through my internship with DC United, and implement new processes that I learned throughout the semester in this class.

In IST 719: Information Visualization, the class was primarily built through R Studio and how to use that tool to make unique and informative graphics and prepared us to present in front of a variety of audiences. In this class, I wanted to use my knowledge of ACC men's soccer from experience with the Syracuse Men's Soccer Team and apply it through a mid-season report on the conference and how different teams stacked up. I first was able to grab data from a resource that I knew to be reliable. After doing some exploratory analysis to learn more about the data, I was able to find a story and knew which direction I wanted to go in. I then created a variety of graphics using various metrics to try and tell the story of how the season has played out so far. Keeping the Syracuse University audience in mind, I then added to the story by making a small deep dive into how the season for the Syracuse University Men's Soccer Team has gone so far. Through this project I was able to incorporate a variety of the visualization methods taught in class as well as using my data processing background taught in other classes in addition to applying my experience that I have gained through working for the Syracuse University Men's Soccer Team.

In addition to the project for which I have already completed, I have four more projects for classes that are currently ongoing. In IST 707: Applied Machine Learning, my group is working on a project where we are analyzing the impact the weather has on the prices of commodities. We predict crop prices using weather and climate data, training a Random Forest model on six years of monthly records. To handle unexpected disruptions, we add a Markovian Momentum Model (MMM) that monitors shocks like price spikes, volatility, and model errors. By blending the base model with MMM

adjustments, the system makes more reliable forecasts, even during events like COVID or supply chain crises. In IST 718: Big Data Analytics, we are quantifying the value of a batter's swing decision by measuring how it changes run expectancy. Using an XGBoost model, we predict the expected change based on game state, pitch quality, location, and swing choice. For each pitch, we compare the predicted outcome with the actual decision versus a hypothetical no-swing (or swing) scenario. The difference between these predictions is the batter's Swing Decision Rating. In IST 736: Text Mining, we developed a system to detect COVID-19 from unstructured clinical notes using advanced NLP techniques. A two-stage pipeline first uses NER models like Clinical BERT and spaCy to extract key symptoms, severity markers, and time references. These entities are standardized and fed into a transformer-based classifier (e.g., BioBERT) to predict COVID-19 likelihood. The goal is to accelerate case identification, support clinical decision-making, and reduce the burden of manual chart review, with results delivered through an interactive ChatBot. In IST 737: Visual Analytic Dashboards, my group is using Tableau to visualize analysis of video game sales, across the world, different demographics, types of games, and different consoles. These projects will help tie together everything that I have learned so far while trying to attain this degree and the work I have done outside of the classroom and prepare me for the professional world.