

M.S. in applied Data Science Portfolio

Syracuse University School of Information Studies



December 6, 2020

Nina Williams – SUID- 481553807 – nowillia@syr.edu

How I got here, What I have learned, & Where I will go…

Contents

[About Me: 2](#_Toc57986845)

[How I Got Here: My Road to Syracuse: 3](#_Toc57986846)

[Learning Goals: 4](#_Toc57986847)

[Project 1: IST 719 – Information Visualization, Prof. Gary Krudys: 5](#_Toc57986848)

[Project Name: Final Poster (Williams, IST 719 - Information Visualization, 2020) 5](#_Toc57986849)

[Project Summary: 6](#_Toc57986850)

[Learning Goals Achieved: 6](#_Toc57986851)

[Project 2: MBC 638 – Data Analysis & Decision Making, Prof. Lynn Gill 7](#_Toc57986852)

[Project Name: Process Improvement Project with Storyboard (Williams, MBC - 638 - Data Analysis & Decision Making, 2020) 7](#_Toc57986853)

[Project Summary: 7](#_Toc57986854)

[Learning Goals Achieved: 8](#_Toc57986855)

[Project 3: IST 707 – Data Analytics, Prof. Amy Gates 9](#_Toc57986856)

[Project Name: Data Mining & Machine Learning Final Project (Williams, IST 707- Data Analytics, 2020) 9](#_Toc57986857)

[Project Summary: 10](#_Toc57986858)

[Learning Goals Achieved: 10](#_Toc57986859)

[Real Life Action: 11](#_Toc57986860)

[Tomorrow and Beyond: 11](#_Toc57986861)

[References 11](#_Toc57986862)

# About Me:

My name is Nina Ourania Williams. I was born on June 9th, 1984 in Warren, Michigan. I am a first generation American. My mother (Aristea meaning “Ace”) came to the United States at 14 years old for 14 days and 14 nights on a ship across the Atlantic Ocean from Corinth, Greece. My father (Asaad, goes by George) was an international musician and celebrity in Baghdad, Iraq. He came to America with his group to perform in the 1970’s and never went back.

My Mother and Father met in Detroit Michigan in the late 70’s, early 80’s. In 1987, we came to San Diego, California which is where is still currently reside. My parents wanted the same for both my sister and I. That was to live the best life they possibly could, full of health, love, opportunities, happiness, and success. Education was very important to my family. My sister and I both graduated from San Diego State University, and both earned a master’s degree from Pepperdine University.



Figure - Mom, Mary, Myself & Dad – August 2016 in Fairfield, Connecticut. I had just baptized my nephew Reid.

# How I Got Here: My Road to Syracuse:

Math has always been my absolute favorite subject, I enjoyed solving the problems. I decided that I wanted to pursue a degree in Finance. After college, I entered the defense industry working in Program Finance. I soon learned that I had a passion for excel and automation. I enjoyed the work that I did very much. I was able to handle a tremendous workload and was even given two computers at Northrop Grumman! One to pull data, and the other to work on so I did not crash the data pull. I was almost instantly the excel go to. People would ask me to automate several reports and that is exactly what I did. I loved it.

Later in my career I realized that there had to be more out there than just Excel. That is when I discovered data science at Syracuse. I felt that I needed a new challenge, I needed something hard to occupy my mind and challenge my brain. I joined Syracuse pregnant with our first baby girl. During the program I married the love of my life Dobber, gave birth to our sweet girl Domino Rose, had my Thyroid removed after she was delivered due to Thyroid Cancer, moved into a new home, and took a new role at Teradata as a Business Analytics Manager in Product Operations. Domino is now 15 months old and has been rooting for me the whole way!

Figure - Dobber and I Surprise Wedding at the Baby Shower! July 20th, 2019

Figure - Domino Rose Williams 9.4.2019 - Syracuse Baby!

# Learning Goals:

The following learning goals were designed for the Applied Data Science Program. All these goals have been achieved through my participation in the program. I will use the following list throughout my paper to reference the specific goals that were achieved in the highlighted projects. The ability to reach these goals was a result of the completion of all courses, lectures, readings and the various projects that I have worked on during my time here at Syracuse and in the Data Science program.

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.

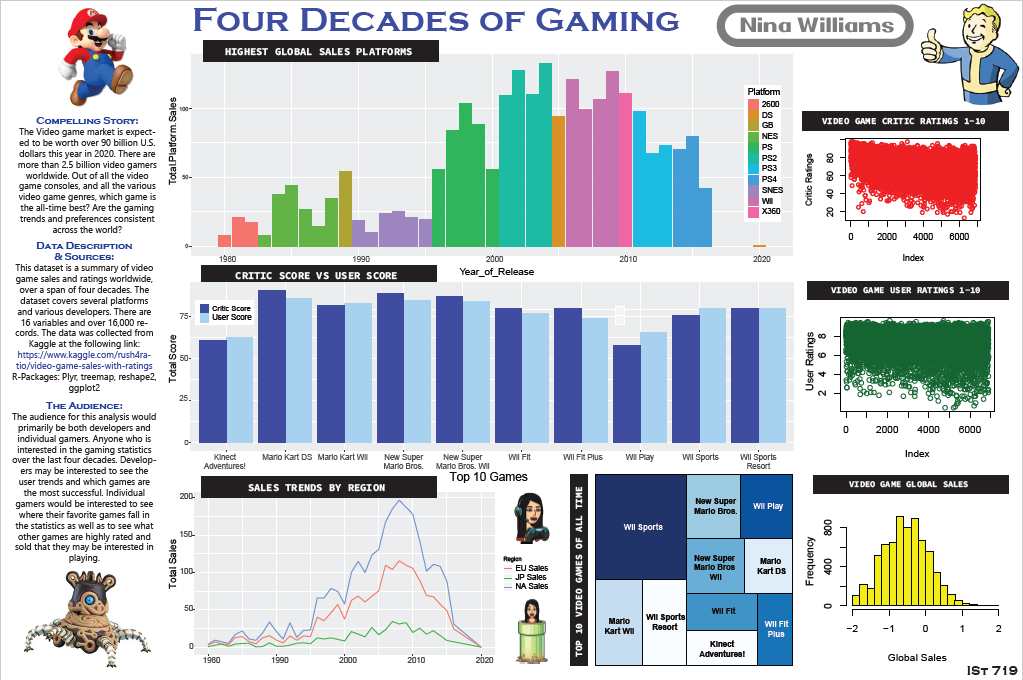
5. Develop a plan of action to implement the business decisions derived from the analyses.

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).

# Project 1: IST 719 – Information Visualization, Prof. Gary Krudys:

## Project Name: Final Poster (Williams, IST 719 - Information Visualization, 2020)



## Project Summary:

The Final Poster in IST 719 Information Visualization is one of my most favorite projects in the entire program. This project and course contributed to the accomplishment of every single learning goal in the course. The assignment was to select a substantial data set that could be transformed into visuals that would answer questions and ultimately tell the story of the data.

For my project I chose a data set describing 40 years of the video game market and industry. The analysis for this project was done in R as well as the visuals that were created. I also used Adobe Illustrator to clean and edit some of my visualization. For this project, I was able to clean and scrub the data, and create several single and multi-dimension visualizations to successfully communicate the data.

## Learning Goals Achieved:

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

To create meaningful visualizations, there are several steps that need to be taken to have meaningful data. These steps can include data mining, data cleaning, business intelligence, strategy, machine learning, data visualization etc. All of which are major practice areas in data science that were part of this project and this course.

2. Collect and organize data.

Data was collected from Kaggle and cleaned and scrubbed in R.

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

Patterns were identified within the various visualization within the poster to communicate messages and findings. For example, you can see that PlayStation holds the record of highest video game sales of all time occurring in the early 90’s.

4. Develop alternative strategies based on the data.

After looking at the data I was able to learn which variables were significant and where to focus my charts and graphs to support my analysis.

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

As a Business Executive, after the analysis is complete, I would be able to make strategic business decisions based on the data to reach the business goals.

6. Demonstrate communication skills regarding data and its analysis for managers, IT professionals, programmers, statisticians, and other relevant professionals in their organization.

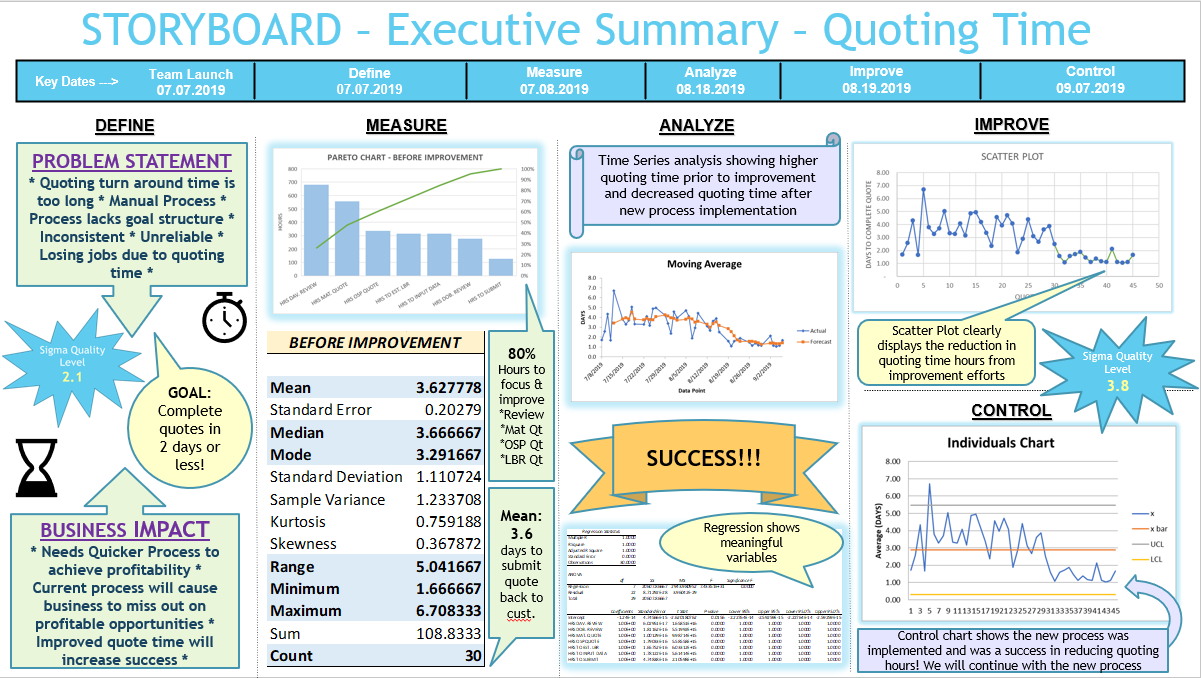
The final poster taught me how to consolidate the various aspects of the analysis to communicate the final messages about the subject. Through the visualizations I was able to communicate my findings and ultimately tell a story about the data.

7. Synthesize the ethical dimensions of data science practice (e.g., privacy).

In this case the data was public and derived from Kaggle. If the data had been proprietary, I would be sure to implement security and access privileges to sensitive data so only those authorized were able to view.

# Project 2: MBC 638 – Data Analysis & Decision Making, Prof. Lynn Gill

## Project Name: Process Improvement Project with Storyboard (Williams, MBC - 638 - Data Analysis & Decision Making, 2020)



## Project Summary:

The goal of the Process Improvement Project was essentially to select a Business Case, identify a current process that is problematic, determine a plan to improve the process, implement that plan, track the same data after the process improvement plan was implemented, then apply various analytic techniques to determine if the process improvement plan was successful in improving the business process or not.

For my project I analyzed a machine shop and their quoting process. I created a Data Measurement Plan, collected the data both before and after the improvement efforts were implemented. I calculated a before and after SQL score, calculated the sample size, performed a multiple regression analysis as well as scatter plots surrounding the data. I also performed and XMR chart analysis as well as a time-series analysis on the data. In conclusion at the end of the experiment the analysis proved that the implementation plan had been effective! The process had in fact improved and the quoting turn-around time had showed significant results!

## Learning Goals Achieved:

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

This class was essential is providing an overview of statistical methods and measures that are a huge part of data science.

2. Collect and organize data.

The raw data used to produce the final product of this project was collected over time to build the official data set behind the analysis. The data needed to be recorded and collected over time as well as organized and clean.

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

There were several data visualizations and statistical analysis explored throughout this project. Examples include: Multiple Regression, XMR Charts, Descriptive Statistics and Time-Series Analysis.

4. Develop alternative strategies based on the data.

All the strategies developed to improve the quoting process were driven from the analysis performed on the data collected surrounding the process prior to the improvements. The original process was used to develop strategies on overall process improvements and showed us where the process was suffering and where opportunities were to improve.

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

The plan of action to improve the quoting process was developed from analyzing the original process. This was a perfect example of how this learning goal has been achieved.

6. Demonstrate communication skills regarding data and its analysis for managers, IT professionals, programmers, statisticians, and other relevant professionals in their organization.

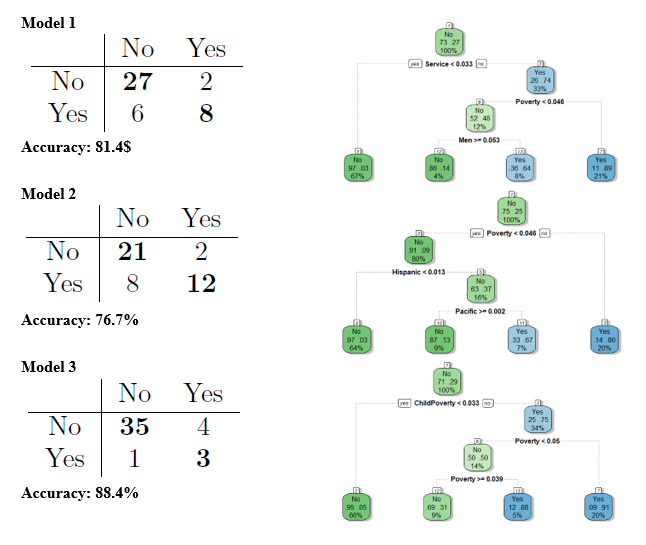
The final PowerPoint deliverable and “Story Board” in general is an excellent example of how the learning of demonstrating communication was achieved. I learned how to organize the story of the data and present that story in a clear, clean, and understandable way.

7. Synthesize the ethical dimensions of data science practice (e.g., privacy).

Prior to executing this experiment, I obtained written permission from the company that I was using to use their data for this project. Data confidentiality is very important and full permission was requested and granted.

# Project 3: IST 707 – Data Analytics, Prof. Amy Gates

## Project Name: Data Mining & Machine Learning Final Project (Williams, IST 707- Data Analytics, 2020)



## Project Summary:

The goal of this project was to analyze the locations of the United Way and see if we could find correlations of those locations within the Census data. From there we wanted to see if we could successfully predict where we would expect to see the next United Way location. The two data sets used for this analysis included the United Way location data as well as the 2015 Census. For the final project question, we wanted to see if we could use data from Kentucky and Tennessee to predict the next United Way location in Ohio.

Both data sets were scrubbed and cleaned. The following models were created: Association Rules, Clustering, Naïve Bayes, Decision Trees, Support Vector Machines, K-Nearest Neighbors, and Random Forest. Overall, we found that our model with the highest accuracy was the Support Vector Machine Model just over 55%. This was not what we had expected. The results showed us that further data would be required to make a more accurate and meaningful model in predicting the next United Way location.

## Learning Goals Achieved:

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

The elements of data science represented in this course and project specifically are my absolute favorite part of data science. I was able to learn predictive analysis and machine learning that is ultimately the area I would love to focus on and work in one day.

2. Collect and organize data.

For this project, the census data was a direct download, but the United Way data took more work as we had to collect the locations from their website and build the data set. From there we merged the data with our census data to create our models and perform our analysis.

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

I learned of many different data visualizations in this course and project specifically. These visualizations included Association Rule Mining plots, Clustering Visuals including Dendrograms, Cluster Plots, Box plots, Confusion Matrix’s and Decision Trees.

4. Develop alternative strategies based on the data.

Based on the results of our data we determined that the alternative strategy would be to implement additional data that would assist in improving our models. We had determined that we did not have enough data to result in the accuracy of our expectations.

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

If our analysis was able to show higher rates of accuracy and the executive decision makers of United Way were able to utilize our models, then they could absolutely use these tools to make a decision on what location would be best to put the next United Way.

6. Demonstrate communication skills regarding data and its analysis for managers, IT professionals, programmers, statisticians, and other relevant professionals in their organization.

The final written report is a demonstration of communications skills that were learned because of this project. The analysis was clearly explained, and the results and visuals were provided to support our overall findings.

7. Synthesize the ethical dimensions of data science practice (e.g., privacy).

In this project the data that we used was available to the public. As always, if the data was sensitive, we would have taken the necessary measure to keep the data private while being able to successfully communicate the final message.

# Real Life Action:

The first project I had worked on in the real world that was a direct result of what I was taught and had learned in this program was a clustering analysis that I had done in R surrounding the material grouping while I was working at Cobham. I had received a lot of excellent feedback after presenting the data and I was proud of what I had learned in this program.

In July I had accepted a new position at Teradata as a Business Analytics Manager. I know that this opportunity would have never happened without the Applied Data Science Program at the iSchool. I have already performed and NPS analysis and created a sentiment analysis in Python based on the 700+ comments from the survey. I was also able to create word clouds to highlight the key messages from each of the NPS categories. I had great feedback after presenting my results and full analysis. This achievement was a direct result of what I have learned here at Syracuse.

# Tomorrow and Beyond:

I am very happy with my recent move out of the Defense industry and into the Tech industry. Currently I am working on a broad spec of different sort of data analytics. My goal will be to grow in my current role and eventually be able to focus more on Predictive Analytics and Machine Learning Algorithms.

Making the decision to attend Syracuse in the Applied Data Science Program has been one of the best decisions of my life. It has already provided so much benefit to myself and my family. I am finding that I am happier with my professional life and career. I am really enjoying the work that I am doing and am excited about the endless growth and possibilities that I have moving forward.

# References

Williams, N. (2020, 12). *IST 707- Data Analytics*. Retrieved from GitHub: https://github.com/Nwill0609/Applied-Data-Science-Final-Project-Portfolio/tree/main/IST%20707%20-%20Data%20Analytics

Williams, N. (2020, 12). *IST 719 - Information Visualization*. Retrieved from GitHub: https://github.com/Nwill0609/Applied-Data-Science-Final-Project-Portfolio/tree/main/IST%20719%20-%20Information%20Visualization

Williams, N. (2020, 12). *MBC - 638 - Data Analysis & Decision Making*. Retrieved from GitHub: https://github.com/Nwill0609/Applied-Data-Science-Final-Project-Portfolio/tree/main/MBC%20-%20638%20-%20Data%20Analysis%20%26%20Decision%20Making