sYRACUSE University

m.s. IN APPLIED DATA SCIENCE

aNNIE tITUS\_portfolio

**SUID: 345602079**

**atitus@syr.edu**

**GRADUATION DATE**: SPRING 2022

# Introduction

Thank you for taking the time to review my portfolio for the Master of Science in Applied Data Science at Syracuse. Being part of this program has been an experience that’s both challenging and rewarding. Earlier on into the program, my professor from IST 687, told me that frustration is a power source of learning and if I can push myself through to the ‘other side’, then I can and will ultimately learn a greater deal and get around that source of frustration. This statement has stayed with me throughout the course of this program and have gotten me to a higher level of understanding of the Data Science field. My exposure to data science was minimal, prior to the start of the program, with most of my knowledge in the data analytics space. This program provided more insight into the evolving field of Data Science, and I do look forward to continuing that learning even after the program ends.

# Professional Background

I received my bachelor’s degree from the University of Colorado-Boulder for Business Administration, with an emphasis in Finance. From there, I attended Southern Methodist University in Dallas, TX and received my Project Management Professional (PMP) certification.

My current profession is in the Aerospace industry, working for a major aerospace provider in Fort Worth, TX. My daily roles and responsibilities consist of utilizing my PMP certification, on large scale enterprise projects. These projects varied in scope, with a general theme in process improvement, for current and/or future state of the company. This includes working with large and complex datasets, that requires machine learning for data analysis, visualization, development of strategy and deployment. The initial interest in Data Science came from the large and complex datasets we would work on, and at times, the Data Scientist was the only individual to make sense of the dataset, causing a potential point of failure if he/she was not available or busy to provide the specific information needed. The led to myself learning more about the field and the skills needed to perform.

The additional knowledge that I have gained from the Data Science education from Syracuse allowed me to be a source of assistance in data analysis, data munging, visualization, and develop of an overall strategy and deployment to a successful project, I add on now as a valued asset, not only as the project manager but as an emerging data scientist of my projects,

# Learning Goals contained in Educational Projects

Outlined below are the learning objectives achieved through the program. Each item is mastered through project based learning.

1. **Major Practice areas** contained in project overview are:
   1. Data transformation: munging, linking, processing, aggregation, data modeling and use of data normalization techniques, clustering analysis, time series analysis, Geo Data, working with structure and unstructured data, and data statistics
2. All projects contain an element of **Collecting and Organizing** data.
   1. Not all data is ready to use or easy to get.
   2. Part of a data scientist skill must be preparing to cleaning the dataset.
3. **Identify patterns through visualization.**
   1. A visual of the data can explain more and justified certain action taken
4. Show an **alternative strategy** based on the failure, the original plan for analysis might not always work out
5. Using the results of the analysis, effectively communicate intelligent **business decisions** to the stake holder that will help in a path towards the desired goal.
6. Who is the audience for each **method of communication**?
   1. Business Leader/Manager
   2. Project Stakeholders
   3. IT Professional
   4. And so on.
7. Monitoring of any type of data has specific **ethical boundaries**, i.e. protect the privacy of the use of data and what/who the data represents

# Project base learning

One of the emphases of the Data Science Program at Syracuse is project-based learning. Below is a list of the major projects that was achieve throughout my time in the program. These projects reflect good examples of real-world problems, the Data Science techniques used demonstrates a general knowledge of Data Science, a glimpse of the different programming languages (R, SQL, Python) used, and a level of complexity and understanding from the beginning of the programming to the end of the programming.

**IST 687- AIRLINE SATISFACTION DATASET**

**Github link:** <https://github.com/AnnieTitus/Syracuse_Portfolio/blob/main/IST_687_Arline_Statisfaction_Final%20Project.zip>

This was my first project in the Applied Data Science program, and first exposure to the R programming language.

In this project, the objective was to analyze an airline dataset that consisted of responses from surveys of customers flying within the United States. The goal of this project and throughout was to clean and analyze the data by using various R coding techniques such as descriptive statistics, modeling, and visualization to identify visible trends and patterns in predicting customers with low satisfactions. Through this prediction, provide actionable insight that would improve future customer satisfaction results and in turn help increase airline profits.

*The following Business questions were established as the guide to the analysis of this project:*

1. *Is there an attribute or a combination of attributes that correlate to higher satisfaction or lower satisfaction score?*
2. *Is there a correlation between time spent at airport to satisfaction ranking?*
3. *Based on the satisfaction survey, what airline is performing well, and which one is performing poorly?*
4. *Does the demographic group differ between each level of satisfaction? Who are the satisfied customers and who are the unsatisfied customers?*

**Methods:**

This project was written in Rscript using R-studio. As a novice to R, I learned quite a bit with data cleaning, visualization, and statistical analysis. I also covered two models in this project: Linear Model Regression and Support Vector Machines, using KSVM function.

**Learning Objectives Completed:** Data transformation, munging and mining, Collecting and organizing data, Identifying patterns in data via visualization, Statistical analysis, Business Decisions, Data modeling of test and training datasets, and Well defined audience of Airline owners and executives

**IST 659- PHOTO ORGANIZATION DATABASE (POD)**

<https://github.com/AnnieTitus/Syracuse_Portfolio/blob/main/IST_659_POD_FinalProject_Deliverable.docx>

The Photo Organization Database project was a great exposure project to the world of SQL. This project proposed the implementation of a photo organization database (POD) to streamline photographs. This dataset used was REAL data pulled from a personal tracking system of my own. First step of the project was the creation of the conceptual and then the logical model before creating the tables and uploading the data within SQL. Through the development of the tables, POD assisted in addressing the business questions and assisted in improving the photo organization method and photo business.

*The following Business questions were established as the guide to the analysis of this project:*

1. *Is there an area of artistic interest (ie. Photo type) that stands out within the photo database?*
2. *Who had the most photos taken?*
3. *Are there any visible trends or patterns on when or where the photos are taken to determine the photographer’s preference?*
4. *When is the most photos taken? Is there a ‘busy’ month/season?*

**Methods:**

Using SQL code and creating good conceptual and logical models was key to understandin the tables and relationships within SQL. Being able to understand the various SQL syntax was very important and I quickly learn that DELETE, ALTER, and DROP were important SQL commands.

**Learning Objectives Completed:** Collecting and organizing real world data, Business Decision, Well defined audience of the photographer and business partner, and data modeling

**MBC 638 – INCREASE SIGHT WORD RECOGNITION**

<https://github.com/AnnieTitus/Syracuse_Portfolio/blob/main/MBC638_Increase_Sight_Word_Recognition_FinalProject.pptx>

My MBC 638 (Data Analysis & Decision Making) course was unique in that, no programming language was utilized but there was a heavy emphasis on being able to identify and apply the appropriate statistical techniques for a given set of conditions to answer a particular question.

In the MBC 638 project, an analysis and experiment was conducted in increasing the number of sight word recognition for my son. This was live collection of data as it was happening and real world decision making through a period of performance.

**Methods:**

This was a statistical analysis project, following the methodology of how to properly identify problems starting with: Problem Statement, Business Impact, Goals, Project Scope, Team, Project Plan, and a Process Map of the existing problem. I touched on the following statistical techniques in this project: Descriptive Statistics, Working with the appropriate sample size, Chi-Square Test and analysis, Confidence Interval, and Control Charts

**Learning Objectives Completed:** Collecting and organizing real world data, alternative strategies, Business Decisions, Well Defined audience of myself and my son, time series analysis, data statistics, ethical boundaries, Identify patterns through visualization.

**MAR 653 – HULU ORIGINAL FEATURE FILM**

<https://github.com/jdcpate/Syracuse_Portfolio/tree/master/Onondaga_Lake_Water_Quality>

In this project, it addresses a business problem for an existing brand using marketing research techniques. Looking at answering the question “*What are the movie attributes that are tied to a success movie rating, and what would my recommendation be to an audience such as HULU if they were interested in producing original film content?”*

This project focuses heavily on R and R modeling and visualization. It allowed me to really look at a dataset and produce a business problem and from there meaningful results through a clear and visual story.

**Methods:**

This project was done in R Studio specifically using the following techniques: cluster, regression, logistic, and/or conjoin analysis, perceptual maps, collaborative filtering, text analytics, and ordinal logit. Through these techniques, create support through data and visualization to make a recommendation to the applicable audience.

**Learning Objectives Completed:**

Collecting and Organizing data, Data Transformation: munging, linking, processing, aggregation, data modeling, Identify patterns through visualization, alternative strategy, Business Decision, Well Defined audience of Hulu.

**IST 652 – PREDICTING SONG POPULARITY USING PYTHON**

The predicting song popularity project was a great insight into using Python. It allowed me to tap

**Github links to :** [.ipynb file](https://github.com/AnnieTitus/Syracuse_Portfolio/blob/main/IST652_PREDICTING_SONG_POPULARITY_IPYNB.ipynb), [Report](https://github.com/AnnieTitus/Syracuse_Portfolio/blob/main/IST652_PREDICTING_SONG_POPULARITY_FinalProjectReport.docx), [Presentation](https://github.com/AnnieTitus/Syracuse_Portfolio/blob/main/IST652_PREDICTING_SONG_POPYULARITY_FinalPresentation.pptx)

into the other programming languages I was already exposure to and see familiarity in the programming. I believe this made it easier for me to understand Python. This project looks at the evolution of the music industry and how song attributes changed over time, and the effects it has on determining its popularity.

*The following were analysis questions in exploring the datasets with the primary goal in mind of can a song’s popularity be predicted based off certain song attributes.*

1. *What song genre has the most popular songs?*
2. *How have popular songs changed over time? Is there a particular period (time) where song popularity was higher?*
3. *How has song attributes change over time?*
4. *Are there certain song attributes/qualities that are correlated with popular songs?*
5. *Can a song’s popularity be predicted from its attributes?*

fantastic combination of predictive modeling and using GGPLOT for mapping the dataset for a visual representation of data exploration and results. This project involved a large dataset of solar panel installation by geographical scans from a satellite image. The goal of the project was to identify based on local demographics and poverty levels where a good market for future solar panel sales would be.

**Methods:**

This project was completing in Python, following the methodology of Data Importation, Data Preparation, Data Analysis, and Data Visualization to answering the analysis questions . This project was heavily focused on data import and clean, due to multiple data sources being used. Two datasets from Kaggle, and two text data, obtained through web scrapping the web. This project did allow for programming languages to cross over, so the web scrapping was completed in R Studio and merged into Python. This provided challenges on it own, but very rewarding.

**Learning Objectives Completed:** Data Transformation, Data Cleanup, Collecting and organizing real world data, alternative strategies, Business Decisions, identify patterns through visualization, Well Defined audience of song artists.

**IST 718- MARCH MADNESS PREDICTION**

In my final term, I look forward to diving deeper into Python and learning how to Handle Big Data. This course will also refine the Data Science methodology of how data is obtain and explaining data structures and data elements, Scrubbing data by applying scripting methods, Exploring the data using qualitative techniques, modeling relationships between data, and finally interpreting the data, model, analysis, and finding.

In this course, I am to look at a data focused problem to which I will look at the March Madness Tournament and pull multiple data sources together, in hopes of answering the project question of “*Who will win the March Madness Tournament?”*

# CONCLUSION:

The Syracuse Masters in Applied data science program has taught so much about the field of Data Science and the techniques used. However, even more than that, I have learned how to tackle problems in a methodical approach with the understanding that not all datasets, problems can be resolved through the same 1,2,3 steps. The program equipped me with various tools in my toolbox, creating a strong foundation to solve any problems though through some trials and errors, but in the process also building onto that foundational knowledge. I am now mroe equipped to succeed further because of this degree. Thank you Syracuse!