



Syracuse
University

IST 782 Applied Data Science Portfolio

By

Vaishnavi Meka

SUID: 893304899

Email: vameka@syr.edu

Table of Contents:

1. IST 652 – Scripting for Data Analysis
2. IST 707 – Applied Machine Learning
3. IST 719 – Information Visualization
4. References

1. IST 652 – Scripting For Data Analysis

In my course *IST652: Scripting for Data Analysis*, I tackled a project on the Chicago Divvy Bike Sharing Program. I'm using two divisions of bike users, namely premium bike users and non-premium users' data to identify key factors by which one can differentiate user segments. This analysis aimed to make key decisions in the Finance strategy of Divvy's company portfolios.

Project Overview Summary:

I conducted data analysis on the Divvy bike sharing platform's trip data to present my findings as an improvement to Divvy's improvements. During my analysis, I thought about how weather can affect bike rides taken by two segments of users, how seasonality comes into the frame, and many more. After much thinking, I finalized asking one research question that encompassed all the other sub questions I had in mind: "Are there consistent patterns or behaviors among two groups of riders between the years 2022 and 2023?"

Methodology:

My approach was comprehensive: -

First, I have understood the business context, and then I have formulated my research question. Once I had my research question, I collected data for the riders for the last 2 years 22 & 23. Later, I processed the data, standardized the data, handled nulls, removed duplicates, & Did some F.E. At last, segmented the data as per the years 2022 and 2023.

Once I had the datasets with me, I did all sorts of plots to compare various data factors in the dataset. After a lot of plotting and analysis, I chose these features: - 'time duration', 'day type', 'weekend engagement', 'hourly activity', & 'temperature'.

Conclusion:

The primary goal of my project is to identify key factors that can differentiate a regular set of biker users from a premium set of bike users. These key factors can be used by the Finance team & Marketing team at Divvy to launch promotional programs to engage non-premium bike users to convert to premium bike users and make money out of it.

Reflection:

Thanks to this course Scripting for Data Analysis, Where I had a chance to enhance my skills in data analysis, data cleaning, data preparation, feature engineering & doing data plotting using various tools. I got to know the importance of using & applying the right techniques with plots to real-world data like Divvy Bike Sharing Data. I learned to navigate the challenges of working with bike-sharing user usage data to study their patterns using tools to enable better insights for Divvy & Its Finance team.

Linking Learning Goals to Deliverables:

Scripting for Data Collection: - By using programmatic calls to download data from the Divvy website using an IPython notebook script.

Preparing & Transforming Data: - By analyzing bike-sharing data, I demonstrated my ability to work Big datasets and using scripts to perform feature transforms on data.

Real-World Problem-Solving: - By utilizing skills learned in this class enabled me to solve real-world Divvy's problems using a bunch of tools.

Preparation for Specialty area:

This project has been pivotal in preparing me for a career in data science, especially in areas intersecting Business strategy and technology. The skills gained here will be instrumental in tackling future challenges, where understanding the dynamics of people can guide effective business decisions for stakeholders of Divvy's Finance portfolio

Challenges:

1. During this phase, I faced a lot of challenges, like data gathering for all the years wasn't possible due to computing limitations.
2. Along with point 1, I also faced challenges with incomplete data, But I learned techniques to address this from my class.

2. IST 707 – Applied Machine Learning

In my course *IST707: Applied Machine Learning*, I tackled a project on the Chicago Divvy Bike Sharing Program. Basically, I'm using bike users' data to understand the daily usage counts of bike rides. This analysis was aimed at making key decisions in their strategy by the Divvy company & DOT.

For example: - Understanding the factors that drive bike usage can help the program's stakeholders make Informed decisions about station placement, bike fleet management, and service optimization

Project Overview:

I conducted data analysis to identify key data variables used for analysis, which are required. Later data preprocessing, & model selections to predict bike usage for a given request.

Methodology:

My approach was comprehensive: -

- Data Handling & Data Analysis: I gathered, bike-sharing trip data, Chicago neighborhoods shape file, holiday information, daily information. I used all this data to create a master data frame. For this master data frame, I have performed pre-processing steps, Aggregations, Geographic aggregation, and outlier removals and prepared a final dataset.
- Model Development: I applied various machine learning models like Linear regression, Random Forest, KNN, Decision Tree, SVR, XGBoost, and Neural Network models. During the modeling process, I used metrics like R-squared, and RMSE to evaluate model performance.

Conclusion:

The primary goal of my project is to predict daily bicycle usage to assist CDOT & Divvy in resource allocation and operational planning.

The models that I built, particularly the XGBoost and neural network models, were able to achieve this goal with high accuracy. These models help predict demand, which in turn helps to better schedule stations and manage the fleet.

Reflection:

Thanks to this course Applied Machine Learning, Where I had a chance to enhance my skills in data analysis, data cleaning, data preparation, feature engineering & model selection. I got to know the importance of using & applying the right techniques to real-world data like Divvy Bike Sharing Data. I learned to navigate the challenges of working with bike-sharing user usage data to enable better insights for CDOT & Divvy.

Linking Learning Goals to Deliverables:

- Data Analysis Skills: By analyzing bike-sharing data, I demonstrated my ability to work with large datasets, a key skill in data analytics for ML.
- Real World Problem-Solving: Developing models to predict Fleet management, helped me show my problem-solving skills, crucial for any data scientist.
- Employing storytelling: The way, I used a methodology-based approach to solving the problem, by showcasing important phases through factual data and graphs, This enabled me to convey to project stakeholders a compelling story.
- Technical Proficiency: Using machine learning models, various statistical skills, & more. Showcased my technical skills, preparing me for advanced work in this area.

Preparation for Specialty area:

This project has been pivotal in preparing me for a career in data science, especially in areas intersecting people operations and technology. The skills gained here will be instrumental in tackling future challenges, where understanding the dynamics of people can guide effective stakeholder communication and better policymaking for CDOT.

Challenges:

During this phase, I faced a lot of challenges, like data gathering for weather, neighborhood shape files, model generalization, identifying model validation metrics & overfitting.

3. IST 719 – Information Visualization

In my course *IST719: Information Visualization*, As a student in the USA on F1 Non-Immigrant Visa, I need to be prepared for the H1B Visa category which will allow me to work in the USA to gain work experience, So I need to be placed in a strategic location to increase my odds of securing H1B Visa.

To do this, I tackled a project on Analyzing the H1B data. This project can be scaled to fellow international students who are looking to place themselves in a strategic location for H1B Visa.

Project Overview:

I conducted data analysis by asking one research question: “Are you strategically located for your H1B Visa in 2025?”.

Methodology:

My approach was comprehensive: -

First, I understood the context and formulated key questions such as: -

- Which States hires the most H1B?
- Which Industry hires the most H1B?
- Which companies hire the most H1B?
- Which Cities Hire the most H1b?

Based on questions, I have collected the real data for the years 2023 & 2024 from the USCIS website. Followed by cleaning & data and munging the data for features of interest. From here I divided the data into two subsets.

Once I had the datasets with me, I chose to apply information visualization techniques such as bar graphs, Maps, stacked distribution plots, density-based word clouds, and many more to identify answers to my questions.

Conclusion:

The primary goal of my project is to find answers to the questions listed above so that one can use that information to strategically place themselves in the USA to higher the odds of getting an H1B Visa.

Reflection:

Thanks to this course Information Visualization, Where I had a chance to enhance my skills in data analysis, data cleaning, data preparation, Storytelling & doing data plotting using various tools. I got to know the importance of using & applying the right techniques with plots to real-world data like H1b from USCIS.

Linking Learning Goals to Deliverables:

Scripting in R: - By using a programmatic approach to do our tasks in R language, I have R to do all of the coding work, generating graphs using R.

Preparing data & Plotting Data: - By preparing & analyzing H1B data using R, I demonstrated my ability to work with data & plot the data graphs using R.

Visual Graphs: - By doing this project I have created visually appealing graphs using various tools and choose to stick with three color pallet options to enhance my graphs visually.

Real-World Problem-Solving: - By utilizing skills learned in this class enabled me to solve real-world Students' problem in terms of Job Visas.

Preparation for Specialty area:

This project has been pivotal in preparing me for a career in data science, especially in areas intersecting Consulting & Data Analytics. The skills gained here will be instrumental in tackling future challenges, where understanding the data & presenting them in a visually appealing way can guide effective business decisions in every industry.

Challenges:

During this phase, I faced a lot of challenges, like data gathering for all the years of H1B lotteries wasn't possible realistically due to computing limitations & Manual labour-intensive work.

REFERENCES:

1. Vaishnavi, Meka (2025) IST-652: MSADS Portfolio Scripting for Data Analysis
<https://github.com/vaishnavimeka27-data/IST-782-ADS-Portfolio/tree/main/IST%20652%20Scripting%20for%20Data%20Analysis>
2. Vaishnavi, Meka (2025) IST-707: MSADS Portfolio Applied Machine Learning
<https://github.com/vaishnavimeka27-data/IST-782-ADS-Portfolio/tree/main/IST%20707%20Applied%20Machine%20Learning>
3. Vaishnavi, Meka (2025) IST-719: MSADS Portfolio Information Visualization
<https://github.com/vaishnavimeka27-data/IST-782-ADS-Portfolio/tree/main/IST%20719%20%20Information%20Visualization>