

EDUCATION

Johns Hopkins University | Baltimore, MD

M.S.E in Data Science | Dec 2024

Macalester College | Saint Paul, MN

B.A. in Statistics, Computer Science | May 2023

GPA: 3.73/4.0 | ASA DataFest 2023 Awards – Best Interactive Visualization @ College

Related Courses: Statistical Modeling, Machine Learning, Database Management, Data Structure

INTERNSHIP EXPERIENCE

Student Researcher | Zhengzhou University, Department of Computer Science

Zhengzhou, China | May 2023 - Aug 2023

- Collaborated with multidisciplinary team in developing real-time ECG recognition and diagnosis application, responsible for sorting out and preparing collected ECG data for 80+ patients with multiple diseases using Python packages (NeuroKit) for future analysis and trend studying

Data Analyst Intern | Midwest Reliability Organization

Saint Paul, MN | Jun 2022 - Aug 2022

- Curated and managed 5-year operational records, refined and redesigned impactful charts and graphs for 3 different energy data systems, proposed novel way of visualizing the data that enhanced data comprehension and facilitated decision-making, presented findings and recommendations within company

PROJECT EXPERIENCE

AI Text Comparison and Analysis Project | Macalester College

Saint Paul, MN | Apr 2023 – May 2023

- Developed project comparing features and homogeneity of AI-generated text with human-generated text using Natural Language Processing (NLP) models (BERT, FLESCH score)
- Conducted organization and quality control of collected human-written texts, employed ChatGPT to generate several texts of three different categories and themes for comparative analysis
- Leveraged findings to build range of informative visualizations, showcased results from various perspectives and precipitated comprehensive understanding of text comparison

Transport Ridership-Gas Price Analysis Project | Macalester College

Saint Paul, MN | Sep 2022 - Dec 2022

- Cleaned and combined three years of daily ridership data for 100+ routes and retail gasoline price data, transformed massive shape data using ArcGIS Pro for geographical analysis
- Employed linear regression, time series, and Linear Mixed Models to run numerous single route and route type models, unearthed gas price-ridership relationship from coefficients and residuals
- Constructed interactive R Shiny App with Leaflet Map, accessible data table, and multi-line plot to provide insights into impact of gas price on public transport ridership, augmented understanding of transportation dynamics

Mapping and Spatial Analysis of Crime | Macalester College

Saint Paul, MN | Apr 2022

- Combined and cleaned 8-year crime report data for 17 neighborhoods in Saint Paul using RStudio, utilized Lasso regression to identify significant predictors for further analysis
- Performed exploratory data analysis on demographic data to assist feature selection, created visualizations for variable classification, and trend analysis in context of crime mapping and spatial analysis
- Utilized GEE (Generalized Estimating Equations) model to minimize residuals and p-values, performed Moran's I test on residuals to identify potential spatial correlations

SKILLS

Programming: R-Studio, Python (pandas, NumPy, Matplotlib, Altair), Java, SQL, GitHub

Data Visualization: Tableau, Power BI, Microsoft Excel (pivot table, conditional formatting, etc), Power Point

Methods/Industry Skills: Linear Regression, Logistic Regression, Mixed Effect Model, Time Series, Spatial Analysis, Classification, Clustering, Decision Tree, Random Forest, Principal Component Analysis, KNN Algorithm