Yuxuan (Susan) Wang

Ann Arbor, MI 48105 | susanwyx@umich.edu | 413-552-9752 | Linkedin | Github | Website

EDUCATION

University of Michigan — Ann Arbor, MI
Masters of Data Science — Aug. 2023 - May 2025
Coursework: Databases Management Systems, Data Mining, Machine Learning (IP), Multivariate Analysis (IP)

Mount Holyoke College (MHC) — South Hadley, MA
Bachelor of Arts in Mathematics, Data Science (GPA: 3.98/4.00) — Aug. 2019 - May 2023
Honor: ICERM Fellowship (Mar. 2022), Sarah Williston Scholar (Oct. 2021), Mildred L Sanderson Prize Math (May 2020)

SKILLS

Programming Languages and Tools: Python, R, SQL, MatLab, Tableau, Excel
Machine Learning: Logistics Regression, KNN, Decision Trees, Random Forests, K-means Clustering, Topic Modeling
Math and Statistics: Latex, Overleaf, SageMath, Bayesian Statistics, Principal Component Regression, Linear Regression
Publication: "Leveraging Financial News Analysis to Predict Stock Price Movement," *Frontiers in Economics and Management*, ISSN: 2692-7608, Volume 2 Issue 7, 2021, DOI: 10.6981

WORK EXPERIENCE

Mount Holyoke College | *Data Infrastructure Construction Assistant* | South Hadley, MA — March 2022 - May 2023
Data Science Infrastructure Construction using AWS and Docker Container
- Developed data science docker container with **R** working environment based on existing images
- Assisted in setting up the Littlest JupyterHub on **AWS** for MHC campus to serve 5+ courses 3+ semesters
- Tested out the JupyterHub functions of user login, folder permissions, backup processes, SSN authedications

ICERM | *Research Assistant* | Providence, RI — June - August 2022
Computational Combinatorics on Parking Functions: Unit Interval Parking Functions and r-Fubini Numbers
- Conducted literature reviews on various combinatoric patterns of parking functions with different statistics subsets
- Enumerated parking functions with fixed statistics sets by coding in **Python** Sagemath
- Extended previous studies by proving theorems and conjectures about the association between Fubini, R-Fubini numbers and unit interval parking functions with certain tie subsets
- Presented at Joint Mathematics Meetings 2023 conference and co-authored the study with team members

Mount Holyoke College Math and Statistics Department | *Research Intern* | South Hadley, MA — June - December 2021
Determination of Average Time that A Particle Escapes from a Channel with Diverse Parameters
- Conducted literature reviews on probability distribution laws and random walks derived from billiards escaping a channel
- Performed simulations in **R** and proposed conjectures about the influence mechanism of geometric microstructures in a channel on the mean escape time of a billiard following the specular-diffuse random collision law with various parameters
- Designed an R-shiny app to present our findings, allowing users to observe the changes on the mean escape time by altering the parameters, such as the portions of specular or diffuse collisions and the incoming angles of the billiards

PROJECTS

Exploratory Analysis on Starbucks Customer Data (SQL, MySQL) — Nov. 2023 – Jan. 2024
- Preprocessed 306,534+ Starbucks customer profile and transaction data with 10+ features by resolving invalid input problems, label encoding, replacing empty string values
- Conducted exploratory data analysis on the metadata of the Starbucks rewards app simulation data
- Analyzed and examined the customer demographic features of those who actively respond to offer types from the Starbucks rewards app simulation data

Prediction on Avocado Price (R, Random Forest, PCR) — Sep. - Dec. 2022
- Cleaned and transformed the avocado price data according to various regression assumptions
- Examined and compared the predictive regression models based on the test **MSE** to select the final model
- Leveraged and interpreted Random Forest Regression model (test MSE = 0.000475) in real life context

Prediction of Stock Price Movement (Python) — May - July 2020
- Collected and cleaned stock market data for 3538 stock tickers over the period of 2017-2020 using NLTK tools
- Performed sentiment analysis using **VADER** and fitted the models to predict stock market price data involving Logistic Regression, Naive Bayes model and Random Forest
- Forecasted stock prices with sentiment data and stock price data and co-authored the paper entitled "Leveraging Financial News Analysis to Predict Stock Price Movement," contributing to the sections of Frontiers in Economics and Management