

# Yuxuan (Susan) Wang

Ann Arbor, MI 48105 | [susanwyx@umich.edu](mailto:susanwyx@umich.edu) | 413-552-9752 | [LinkedIn](#) | [Github](#) | [Website](#)

## EDUCATION

---

|  |  |
|--|--|
| <b>University of Michigan</b><br><i>Masters of Data Science</i><br>Coursework: Databases Management Systems, Data Mining, Machine Learning (IP), Multivariate Analysis (IP)  | Ann Arbor, MI<br>Aug. 2023 - May 2025    |
| <b>Mount Holyoke College (MHC)</b><br><i>Bachelor of Arts in Mathematics, Data Science (GPA: 3.98/4.00)</i><br>Honor: ICERM Fellowship (Mar. 2022), Sarah Williston Scholar (Oct. 2021), Mildred L Sanderson Prize Math (May 2020) | South Hadley, MA<br>Aug. 2019 - May 2023 |

## 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

---

|   |                       |
|---|-----------------------|
| <b>Mount Holyoke College</b>   <i>Data Infrastructure Construction Assistant</i>   South Hadley, MA<br><i>Data Science Infrastructure Construction using AWS and Docker Container</i> <ul style="list-style-type: none"><li>Developed data science docker container with <b>R</b> working environment based on existing images</li><li>Assisted in setting up the Littlest JupyterHub on <b>AWS</b> for MHC campus to serve 5+ courses 3+ semesters</li><li>Tested out the JupyterHub functions of user login, folder permissions, backup processes, SSN authentications</li></ul>  | March 2022 - May 2023 |
| <b>ICERM</b>   <i>Research Assistant</i>   Providence, RI<br><i>Computational Combinatorics on Parking Functions: Unit Interval Parking Functions and r-Fubini Numbers</i> <ul style="list-style-type: none"><li>Conducted literature reviews on various combinatoric patterns of parking functions with different statistics subsets</li><li>Enumerated parking functions with fixed statistics sets by coding in <b>Python</b> Sagemath</li><li>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</li><li>Presented at Joint Mathematics Meetings 2023 conference and co-authored the study with team members</li></ul>   | June - August 2022    |
| <b>Mount Holyoke College Math and Statistics Department</b>   <i>Research Intern</i>   South Hadley, MA<br><i>Determination of Average Time that A Particle Escapes from a Channel with Diverse Parameters</i> <ul style="list-style-type: none"><li>Conducted literature reviews on probability distribution laws and random walks derived from billiards escaping a channel</li><li>Performed simulations in <b>R</b> 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</li><li>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</li></ul> | June - December 2021  |

## PROJECTS

---

|  |                       |
|--|-----------------------|
| <b>Exploratory Analysis on Starbucks Customer Data (SQL, MySQL)</b> <ul style="list-style-type: none"><li>Preprocessed 306,534+ Starbucks customer profile and transaction data with 10+ features by resolving invalid input problems, label encoding, replacing empty string values</li><li>Conducted exploratory data analysis on the metadata of the Starbucks rewards app simulation data</li><li>Analyzed and examined the customer demographic features of those who actively respond to offer types from the Starbucks rewards app simulation data</li></ul>  | Nov. 2023 – Jan. 2024 |
| <b>Prediction on Avocado Price (R, Random Forest, PCR)</b> <ul style="list-style-type: none"><li>Cleaned and transformed the avocado price data according to various regression assumptions</li><li>Examined and compared the predictive regression models based on the test <b>MSE</b> to select the final model</li><li>Leveraged and interpreted Random Forest Regression model (test MSE = 0.000475) in real life context</li></ul>  | Sep. - Dec. 2022      |
| <b>Prediction of Stock Price Movement (Python)</b> <ul style="list-style-type: none"><li>Collected and cleaned stock market data for 3538 stock tickers over the period of 2017-2020 using NLTK tools</li><li>Performed sentiment analysis using <b>VADER</b> and fitted the models to predict stock market price data involving Logistic Regression, Naïve Bayes model and Random Forest</li><li>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 <i>Frontiers in Economics and Management</i></li></ul> | May - July 2020       |