

Simeng Li

804-937-8961 | s13245@cornell.edu | Ithaca, NY | Open to Relocation | www.linkedin.com/in/simeng-li-8477421b9

EDUCATION BACKGROUND

Cornell University Master of Applied Statistics, Data Science Track GPA: 4.06 / 4.00 <ul style="list-style-type: none">Coursework: Applied Time Series, Statistics for Financial Engineering, Big Data Technologies, Machine Learning, Linear Models with Matrices, Database Management and SAS High Performance ComputingSkills: SQL, Python (NumPy, scikit-learn, TensorFlow, pandas, matplotlib), SAS, R, AWS, Java, Tableau, MongoDBCertifications: SAS Certified Specialist: Base Programming, Tableau Desktop Specialist, JLPT (N3)	Ithaca, NY Aug 2024 - Dec 2025
University of Richmond Bachelor of Science, Mathematical Economics (Major), Computer Science (Minor) Major GPA: 3.93 / 4.00 Cumulative GPA: 3.87 / 4.00 (Summa Cum Laude) <ul style="list-style-type: none">Honors and Awards: multiple Dean's Lists, 2021-2022 Herman P. Scholarship, Departmental Honor GraduateRelevant Coursework: Algorithms, Data Structures, Database Systems, Advanced Econometrics, Pandas & SQL Basics, Mathematical Economics, Machine Learning, Mathematical Statistics, Probability, Statistical Learning	Richmond, VA Aug 2019 - May 2023

EXPERIENCE

T. Rowe Price <i>Modeling Unemployment Spells and Earnings Mobility (Cornell Ongoing Capstone Project)</i>	Ithaca, NY Jan 2025 – Present
<ul style="list-style-type: none">Develop statistical and machine learning model to simulate unemployment spells and earnings mobility throughout an individual's career using the Panel Study of Income Dynamics (PSID) dataset, analyzing predictors such as age, gender, earnings rank, marital status, and financial variables to assess their impact on unemployment dynamics.	
Cornell Center for Social Sciences <i>Graduate Student Data Analyst</i>	Ithaca, NY Oct 2024 – Present
<ul style="list-style-type: none">Reviewed and validated R, Python, and SAS code to ensure accurate, reproducible results and alignment with analysis goals.Developed reproducibility packages to facilitate transparency and improve research workflows.Provided detailed feedback to enhance data integrity and improve documentation processes.	
Regeneron Pharmaceuticals <i>Statistical Programming Intern</i>	Basking Ridge, NJ Jun 2024 – Aug 2024
<ul style="list-style-type: none">Built and validated statistical programs using SAS and SQL to generate reports, tables, and visualizations, ensuring high standards of accuracy and compliance with project specifications.Gained expertise in programming workflows and data management in regulated environments.Collaborated with cross-functional teams and participated in training sessions to enhance technical and industry knowledge.	
Medical Practice Evaluation Center, Massachusetts General Hospital <i>Clinical Research Coordinator, Simulation Modeling and Decision Science Pillar</i>	Boston, MA Jul 2023 – Apr 2024
<ul style="list-style-type: none">Leveraged data science techniques to create, execute and debug mathematical simulation model runs and sensitivity analyses aimed at assessing the effectiveness of interventions for managing HIV.Drove in-depth cost-effectiveness analyses for a range of treatment and prevention interventions, incorporating real-world data to inform model parameters through extensive literature review and data analysis.Utilized data-driven techniques to inform and enhance healthcare decision-making processes in resource-limited settings.	
ACADEMIC PROJECTS	
Length Bias Estimation of Small Businesses Lifetime <i>Mathematical Economics Honor Thesis</i>	Sept 2022 – Apr 2023
<ul style="list-style-type: none">Conducted in-depth analysis of lifetime data for 24 restaurants located on Carytown Commercial Street, employing advanced statistical techniques to estimate the average lifespan in the area while minimizing impacts of length bias and right-censoring.Implemented Kaplan-Meier estimators to account for right-censored observations and performed rigorous goodness-of-fit tests to assess the efficacy of the developed estimator against the exponential distribution. The results unveiled precise lifetime estimations devoid of length bias.	
Analyzing the Statistical Effectiveness of Diagnostic Tests for COVID-19 Using the ROC Curve <i>Summer Research Fellowship</i>	May 2022 – Aug 2022
<ul style="list-style-type: none">Developed a preliminary ROC curve with using patients' age as the classification thresholds based on a COVID hospitalization dataset predicting patients' survival and achieved an AUC-ROC of 0.81.Constructed multivariate logistic regression models incorporating key predictive variables such as age, comorbidities, and vital statistics to refine the predictive accuracy of survival outcomes. The updated ROC curve achieved an improved AUC of 0.86, demonstrating enhanced classifier performance and robustness.Evaluated trade-offs between sensitivity and specificity to identify optimal decision thresholds, enabling informed allocation of limited medical resources such as equipment and medications during the pandemic.	
Customer Subscription Prediction Using Machine Learning Models for a Freemium Music Business	Mar 2022 – Apr 2022
<ul style="list-style-type: none">Imported data from the university MongoDB server and created a Python pandas data frame for analyzing and building models.Designed and developed machine learning models for predicting whether a freemium music subscriber will become a paid member in Python, including KNN, logistic regression, decision tree, and random forest algorithm.Presented 4 most important features for increasing subscribers based on the decision tree model with an accuracy score of 85.3% after applying SMOTE over-sampler.	