

Yiwen Li

Boston, MA | +1 857-270-8656 | yiwenli@fas.harvard.edu | [linkedin.com/in/yiwen-li-8671b7209](https://www.linkedin.com/in/yiwen-li-8671b7209)

EDUCATION

Harvard University Sept 2024 - Jun 2026

Master of Science in Data Science

- **Relevant Courses:** Machine Learning (ML), Data Science, Natural Language Processing, Multilevel Modeling

University of Toronto Sept 2019 - Jun 2023

Bachelor of Science in Statistics and Economics (double majors) | GPA: 3.99/4.00

- **Relevant Courses:** Time Series Analysis, Python Programming, Advanced Data Analysis, Financial Economics
- **Awards:** Science & Mathematics Scholarship (Top 1%), Reuben Wells Leonard Scholarship (Top 1%), C. L. Burton Open Scholarship, U of T Special Admission Scholarships, Dean's List Scholar (3 years)

INTERNSHIPS

Jd.Com, Inc Aug 2022 - Dec 2022

Data Analyst

Beijing, China

- Implemented automatic workflows in **Power Query** to standardize large-scale data and generate weekly reports within 5 seconds, empowering faster data-driven marketing decisions through 20+ KPIs like conversion rate
- Optimized processing efficiency for 100K+ eCommerce data rows using advanced Excel functions (**VLOOPUP**, **Pivot Tables**) and Python **Pandas**, enabling accurate data extraction and reducing human error rates by 30%
- Created 10+ **Tableau** validation dashboards to ensure consistency between **SQL** queries and reporting logic
- Increased business sales by 25% through optimizing logistics resource allocation with an in-depth quantitative analysis of regional gross profit and customer engagement, supporting cross-functional teams in establishing a new warehouse

UnionPay May 2022 - Jul 2022

Data Analyst

Shenzhen, China

- Designed 20+ **ETL** (Extract, Transform, Load) data pipelines in **SQL** to process 600K transactional data from financial institutions, increasing reporting efficiency by 40% and saving the team 8 hours per week
- Assessed annual company performance in **R** by benchmarking 200+ business indicators, with Exploratory Data Analysis (**EDA**) to examine distributions and the Kruskal-Wallis test to identify statistically significant indicators
- Built a risk prediction model (**Logistic Regression**) to detect suspected cash-out activities with a 90% accuracy, informing the executive board of high-risk merchants and recommended solutions through clear **presentation**

PROJECT EXPERIENCES

Surrogate Assisted Positive Unlabeled Learning on EHR data May 2023 – May 2024

Research Assistant | Supervisor: Prof. Jessica Gronsbell

Toronto, Canada

- Developed a **semi-supervised ML** algorithm for phenotype prediction with an **AUC** score exceeding 93 that outperformed all 7 baseline models, offering a substantial and accurate solution to reduce manual chart-review efforts for data labeling
- Achieved robust feature selection with adaptive **LASSO** and automated hyperparameter tuning with **R** (*glmpath*)
- Conducted **NLP** analysis on real doctor notes to extract disease-indicative terms using the Unified Medical Language System, improving the model's predictive accuracy by 10%
- Demonstrated model effectiveness in handling **high-dimensionality** through extensive model robustness testing on 42K Electronic Health Records (EHR) from the MIMIC database with a generation of 1100 covariates
- Presented research poster to 300+ professionals and published the model as an **R package** (*SAPUL*), contributing to the open-source development and allowing the research community to reproduce statistical findings

Monty Hall Meets AI: The Influence of AI on Decision-Making Apr 2023 – Present

Research Assistant | Supervisor: Prof. Boris Babic

Toronto, Canada

- Coded survey questions in Qualtrics to collect **multi-level data**, facilitating the evaluation of participants' trust in AI-generated suggestions for the Monty Hall problem via quantitative and qualitative metrics
- Established a connection between **Prolific** and **Qualtrics**, streamlining the recruitment process of 2K participants
- Enabled dynamic survey design by leveraging predefined hypotheses based on **Bayes theorem** to find optimal solutions and implementing logic structures to target different participant segments

Canadian Fitness Watch Market Analysis Feb 2022 – Apr 2022

Data Consultant

Toronto, Canada

- Collaborated with a consulting team of fellow students to provide comprehensive **market analysis** for a virtual fitness watch company targeting the Canadian market
- Built regression models with R and performed advanced data wrangling on **20K+** consumer records provided by the client
- Conducted **web scraping** to gather up-to-date market data, enabling sophisticated **data visualizations** and trend analysis to support the client in refining their market positioning and strategy
- Delivered a comprehensive report with an **executive summary** and **technical analysis**, providing actionable insights on product performance, customer segments, and market competition

SKILLS

Programming & Software: R (4 years), Python (2 years), SQL (2 years), Microsoft Office Suite (Excel, PowerPoint, Word)

Languages: English, Mandarin, Cantonese