

CAROLYN LIU

[linkedin.com/in/carolynliu4](https://www.linkedin.com/in/carolynliu4) | [crlu4.github.io](https://github.com/crlu4) | (510) 371-1108 | crlu4@uchicago.edu

EDUCATION

University of Chicago, Harris School of Public Policy

Master of Science in Computational Analysis and Public Policy

Chicago, IL

June 2023

- Relevant coursework: CS w/ Applications (Python), Machine Learning, Databases, Data Visualization, Program Evaluation (causal inference), Time Series Analysis & Forecasting, NLP, Large-Scale Computing

Wellesley College

Bachelor of Arts in Economics, Minor in Mathematics, *Magna Cum Laude*

Wellesley, MA

June 2018

- Awarded the Natalie Bolton Faculty Prize in Econometrics (2017): Prize for the best econometrics group paper

EXPERIENCE

NYC Taxi & Limousine Commission (TLC)

Data Analyst (Full-time)

New York, NY

August 2023 - Present

- Produce and present to senior leadership high-quality, organized, and thorough data analyses, contributing to significant projects such as adjusting wait time thresholds for for-hire vehicle accessibility rules
- Utilize PowerBI and Python to develop, launch, and maintain TLC's public data dashboard: [TLC Factbook](#) (November 2023)
- Developed new methodology and ensured reproducibility for TLC's annual [Safety Honor Roll](#), reducing project time from one year to three months with minimal prior documentation
- Refactored and automated monthly data update processes for industry indicators and Vision Zero metrics, converting manual tasks in R and Excel to Python scripts

University of Chicago

Graduate Teaching Assistant (Part-time 15hrs/week)

Chicago, IL

September 2022 - May 2023

- Teaching Assistant for Advanced Statistics for Data Analysis I and Machine Learning for Public Policy
 - Prepared materials to facilitate and teach two discussion sections along with hosting weekly office hours
 - Assisted head TA with development of exams and problem sets
 - Graded and provided meaningful and timely feedback on exams and problem sets

Civic Digital Fellow

Data Science Fellow at US Census Bureau (Full-time)

Remote

June 2022 - August 2022

- Conducted exploratory record linkage and analysis in Python between the Business Formation Statistics data product and other demographic datasets (Annual Business Survey & Nonemployer Statistics by Demographics) to develop a process (matching by unique identifier; resolving duplicates to select primary record) for an improved data product
- Presented findings to external audiences and senior internal leadership on the process for assembling a more equitable & improved data product to help businesses, policymakers, regional planners, and researchers assess the current state of early entrepreneurship at the national and state levels

Harris School of Public Policy

Research Assistant (Part-time 12hrs/week)

Chicago, IL

March 2022 - June 2022

- Leveraged Stata to build infrastructure for large (>20GB) consumer expenditures data
- Explored APIs to collect Amazon ASIN codes and string-matching techniques (matching Amazon product descriptions to item descriptions in consumer expenditures data) to impute ~5 million missing UPC codes

Harvard Business School

Research Associate (Full-time)

Boston, MA

July 2018 - July 2021

- Utilized Stata and SAS to collect, clean, and apply econometric models using administrative and financial data; example projects included the effect of internet and information inequality on job flows, impact of a salary disclosure law on the gender pay gap, & creating an adjusted living wage benchmark
- Collaborated in the development of and co-authored HBS cases used in the MBA curriculum by conducting company due diligence, interviewing company staff, and creating requisite data representations

TECHNICAL SKILLS

- Programming: Python (numpy, pandas, scikit-learn, matplotlib), SQL
- Tools: Docker, Dash, Git, PowerBI, Databricks
- Statistics/econometrics: OLS, randomized controlled trials, instrumental variables, regression discontinuity

PROJECTS

Data Science Clinic: Electronic Monitoring Device Shielding Detection (Python, Docker, SQL); Available upon request

- Worked on a team of 3 with an external stakeholder; designed & conducted field tests to build sufficient ground-truth on device GPS signal loss
- Built logistic and XGBoost models to identify whether device shielding was occurring; models had accuracy of ~80% (device vendor's built-in detection had accuracy of only ~50%)

Data Visualization - The Tate Collection through the Years (D3.js, HTML/CSS, Python)

- Coded primarily in D3.js to conduct visual storytelling (used stacked bar, interactive treemap & modified beeswarm charts)

Analyzing COVID Twitter Sentiments on Healthcare Stock Price (Python: pandas, numpy, matplotlib, scikit-learn)

- Conducted sentiment analysis on COVID-related Tweets to extract polarity of tweets, preprocessed data & performed feature selection using information gain criteria
- Developed logistic regression, random forest, gradient descent & support vector classification models to predict stock price direction based on time series data for multiple train-test datasets (9-month training; the following 6-months for testing)

Chicago Food Inspection (SQL, Python, Flask); Available upon request

- Built a RestAPI via Flask and SQL to perform CRUD operations on ~1k food inspection data from the City of Chicago
- Implemented blocking and indexing in SQL to optimize data cleaning algorithm by 40%