CAROLYN LIU

linkedin.com/in/carolynliu4 | crliu4.github.io | (510) 371-1108 | crliu@uchicago.edu

EDUCATION

Harris School of Public Policy, University of Chicago Chicago, IL
Master of Science in Computational Analysis and Public Policy *Expected June 2023*
- Relevant coursework: CS with Applications (Python), Math for Machine Learning & Machine Learning, Databases, Data Visualization, Program Evaluation (causal inference)

Wellesley College Wellesley, MA
Bachelor of Arts in Economics, Minor in Mathematics, *Magna Cum Laude* *June 2018*
- Awarded the Natalie Bolton Faculty Prize in Econometrics (2017): Prize for the best econometrics group paper

SKILLS

- Programming: Python (numpy, pandas, scikit-learn, matplotlib), SQL, D3.js, Stata, SAS
- Statistics/econometrics: OLS, randomized controlled trials, instrumental variables, regression discontinuity

EXPERIENCE

Civic Digital Fellow Remote
Data Science Fellow at US Census Bureau *June 2022 - August 2022*
- Conducted exploratory linking and analysis in Python/Jupyter Notebook between the Business Formation Statistics data products and various demographic datasets (Annual Business Survey & Nonemployer Statistics by Demographics), developing a process for how an improved data product will be assembled
- Presented findings to an external audience and senior internal leadership on the process for assembling improved data product

Harris School of Public Policy Chicago, IL
Research Assistant *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 Boston, MA
Research Associate *July 2018 - July 2021*
- Utilized Stata and SAS to collect, clean, and apply econometric models using administrative and financial data
- Aided 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
- Example projects include 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

PROJECTS

Analyzing COVID Twitter Sentiments on Healthcare Stock Price
- Conducted sentiment analysis on COVID-related Tweets to extract polarity of tweets (positive or negative), preprocessed data and performed feature selection using information gain
- Adopted logistic regression, random forest, gradient descent & support vector classifiers to predict stock price direction based on time series data for multiple train-test groups (9-month training period with the following 6-months for testing)

Chicago Food Inspection (SQL, Flask)
- Wrote SQL queries to interact with DBMS and acquire information from ~1k restaurant records; cleaned data and linked restaurants that are the same; implemented blocking and indexing in SQL to optimize the data cleaning algorithm by ~40%
- Coded with Flask to set up web service endpoints for food inspections in Chicago; endpoints received http POST and GET requests to store and read data, i.e., restaurant location, latest inspection date