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 w/ Applications (Python), Machine Learning, Databases, Data Visualization, Program Evaluation (causal inference), Time Series Analysis & Forecasting, Advanced ML/NLP (Spring), Large-Scale Computing (Spring)

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

TECHNICAL SKILLS

- Programming: Python (numpy, pandas, scikit-learn, matplotlib), SQL, D3.js, R, Stata, SAS, HTML/CSS
- Tools: Flask, Docker, Dash, Git
- 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
- 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%

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%

EXPERIENCE

Civic Digital Fellow   Remote
Data Science Fellow at US Census Bureau   *June 2022 - August 2022*
- Conducted exploratory record linkage and analysis in Python between the Business Formation Statistics data products and various 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   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; 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
- 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