

Summer Long

618-600-5000 | salong@uchicago.edu | [LinkedIn](#) | [GitHub](#) | [Publication](#)

EDUCATION

The University of Chicago Chicago, IL
Master of Science in Computational Analysis and Public Policy June 2024 (Expected)
Relevant Coursework: Big Data Platforms, Machine Learning, Data Visualization, Topics in Software Engineering

The University of Chicago Chicago, IL
Bachelor of Arts in Public Policy Studies with a Specialization in Statistics June 2023
Relevant Coursework: Linear Algebra, Analysis of Categorical Data, Experimental Design
Awards: Public Policy Leaders Program, Honors

PROFESSIONAL EXPERIENCE

Bureau of Labor Statistics | Python Washington, D.C.
Data Analytics Fellow June 2023 - August 2023

- Utilized SQLAlchemy to manipulate data from 1987 to present from encrypted Microsoft SQL Server
- Automated the creation of visualizations complying with organizational standards, eliminating manual creation of 100+ visualizations and saving 40+ hours per report
- Consulted with domain area experts to implement a formula to order generated visualizations by importance
- Designed intuitive interface with Tkinter and packaged into an executable, ensuring accessibility for 10+ staff

Chapin Hall | R Chicago, IL
Research Assistant June 2021 - September 2021

- Performed linear regression analysis on ~5700 observations to determine statistically significant factors of community college matriculation for youth in foster care
- Designed resource graphics about pregnancy for dissemination to target populations at outreach events

Be Strong Families | R, Tableau Chicago, IL
Youth Advisory Board Intern January 2021 - June 2021

- Examined various state and college-level resources for youth in foster care that improve retention
- Produced interactive data visualization of density of school attendance among youth in foster care for audience of non-technical stakeholders
- Created a set of recommendations based on youth input and research findings to present to director of DCFS

PROJECT / ACADEMIC / RESEARCH EXPERIENCE

Matching Between Humans & Artificial Generation of the 'Perfect Match' | Python January 2024 - March 2024

- Used SentenceBERT to vectorize text and computed cosine similarity between ~60000 individuals for matching
- Fine-tuned GPT-2 model on computed matches to generate an ideal match with user-submitted text
- Deployed Detoxify to screen generated matches, ensuring text is not harmful prior to display

Graduation Rates & Predictors for Youth in Foster Care | Python, HTML, CSS September 2022 - April 2023

- Cleaned ~26000 survey responses and made decisions on how to handle null and erroneously inputted values
- Tested various machine learning models to explore predictive ability of survey responses at age 17 on graduation by age 19
- Published accompanying webpage using GitHub Pages with interactive visualizations of EDA and background

Citation Network Analysis | PySpark, GCP January 2023 - March 2023

- Parsed through 18GB dataset to remove duplicate papers and erroneously inputted information
- Utilized big data tools to compute top categories by year, distribution of topic count, and most common terms
- Built models to classify paper topic, predict number of citations, and define intersections of fields over time

SKILLS

Technical Languages: Python (Scikit-learn, Pandas, Numpy, PyTorch, Hugging Face), C, SQL, R, HTML, CSS, d3.js

Technologies: Tableau, GCP, AWS, GitHub, Docker, Postgres

Extracurriculars: Student Body President (2022-2023); Tournament Organizer (2019-Present)