

# Han Liu

646-431-0908 | [hliu55@uw.edu](mailto:hliu55@uw.edu) | [www.linkedin.com/in/hanliuuu](http://www.linkedin.com/in/hanliuuu)

## EDUCATION

### University of Washington

*Master of Science in Biostatistics*

Seattle, WA

Sep 2021 – Mar 2023

### University at Buffalo, The State University of New York

*Bachelor of Science in Statistics, Minor in Mathematics, Dean's List*

Buffalo, NY

Jan 2019 – May 2021

## EXPERIENCE

### Data Scientist, Intern

*Aeyesafe Inc.*

May 2022 – Present

Seattle, WA

- Developed and implemented Convolutional Neural Networks to launch an AI senior safety monitoring system.
- Designed and built data pipeline for live streaming image collection, processing, object detection and classification.

### Research Assistant

*BERD Winter Institute, University at Buffalo*

Dec 2019 – Jan 2020

Buffalo, NY

- Applied appropriate hypothesis testing to determine the correlation of an oncogene with treatment methods, race, cell, type, age at diagnosis, etc. Optimized chi-square, fisher's exact test, Non-Parametric ANOVA for evaluating data, Using R.
- Performed logistic regression models over 1000 observations and 30 variables to analyze the relationship between resistance of chemotherapy treatment and predict the impact of the patient's various features for the tumor.
- Summarized research results which is concluded high MTDH expression by IHC is associated with high stage/grade, resistance to chemotherapy and poor outcomes on GOG 177, and presented analysis reports to faculty.

### Teaching Assistant & Project Coordinator

*Helpyourmath website, The City University of New York*

Sep 2018 – May 2021

Manhattan, NY

- Hosted study groups to obtain student feedback on class activities, provided problem-solving, test-taking strategies, and tutoring students in the math lab and statistical programming.
- Made instructional videos and uploaded them to Helpyourmath website for all students to learn, compiled statistics question notebooks, contributed to constructing the layout and design of the website, Using R and Javascript.

## PROJECTS

### Mental Health Survey In Tech | R, Cross-sectional study and Supervised Learning

Feb 2022 – May 2022

- Determined the type of study and identified possible bias before wrote statistical analysis plan.
- Preprocessed data set by data cleaning, categorical feature transformation and standardization.
- Trained supervised learning models including Logistic Regression, Random Forest and K-Nearest Neighbors, and applied regularization with optimal parameters to overcome overfitting.
- Evaluated model performance of classification via 5-fold cross-validation technique and analyzed feature importance to identify top factors that influenced the results with 0.83 precision, and presented to faculty and cohorts.

### Towed Vehicle Records in Buffalo | Python and Javascript

Jan 2020 – May 2020

- Crawled more than 100,000 tow records from city website by using python package Urllib and BeautifulSoup.
- Extracted features and labels, cleaned up data for different implement names to standardize the entire dataset.
- Created Heatmap over buffalo maps to visualize the towed car location distribution using Matplotlib.
- Performed correlation and regression analysis to identify the features associated with towed cars.

### House Price Prediction | R

June 2020 – Aug 2020

- Implemented dplyr to modify original data, built logistic, multivariate regression and random forest based on house price and various influencing factors, applied time series and ordinary least squares to examine models.
- Concluded factors that ultimately affect the price mainly come from nitric oxide concentration and transportation, achieved a prediction with accuracy of 82% based on decision tree model.

## TECHNICAL SKILLS

**Statistical Computing Languages:** Python, R (tidyverse, ggplot2), SQL (Postgres), SAS Certification, JavaScript, Tableau

**Other Software/Tools:** PyTorch, Jupyter, Github, Minitab, Excel, Stata

**Languages:** English, Mandarin, Cantonese

**Analytics Skills:** Predictive modeling (Logistic Regression, Random Forest, Boosting Tree, SVM, Non-Parametric Regression), Experiment Design, Bayesian Inferences, Time Series Analysis, Pandas, NumPy, Matplotlib, A/B Testing