

# Alan Liu Wu

862-485-9127 | [alanlwu@seas.upenn.edu](mailto:alanlwu@seas.upenn.edu) | [linkedin.com/in/alanlwu](https://www.linkedin.com/in/alanlwu) | [alan-wu.me](http://alan-wu.me)

## EDUCATION

### University of Pennsylvania

*Master of Science in Engineering in Data Science*

Philadelphia, PA

*Expected May 2026*

### Rutgers University- New Brunswick

*Bachelor's of Science in Computer Science and Data Science, minor in Statistics*

New Brunswick, NJ

*Sep. 2021 – May 2024*

GPA: 4.00/4.00

Awards: Summa Cum Laude, Phi Beta Kappa, Matthew Leydt Society, Paul Robeson Scholar

Activities: HackRU (Team Lead), Rutgers Men's Ultimate Frisbee (President)

## TECHNICAL SKILLS

**Languages:** Java, Python, C/C++, SQL, JavaScript, Typescript, HTML/CSS, R

**Libraries:** Pytorch, Matplotlib, Seaborn, Sksurv, Sci-kit Learn, Numpy, Pandas, Lifelines

**Developer Tools:** Git, VS Code, IntelliJ, MySQL, PostgreSQL, Docker, Azure CLI

**Frameworks:** React.js, Next.js, Node.js, Tailwind.css, Material UI

**Software:** Microsoft (Excel, Word, Powerpoint), Tableau

**Certifications:** Machine Learning Specialization (DeepLearning.ai), IBM Data Science Professional Certificate

## EXPERIENCE

### Undergraduate Research Assistant

*Rutgers University Cancer Institute*

Aug 2023 – May 2024

*New Brunswick, NJ*

- Scraped data from the Human Microbiome Project using **beautiful soup** and used principal components analysis and non-negative matrix factorization to reduce 2,000 microbes across 18 distinct body sites
- Employed Metaphlan3 bioinformatic pipeline to profile taxonomy of over 50 samples of 16srRNA sequencing data
- Utilized **Python**, **matplotlib**, **seaborn** to build heat maps, cluster diagrams, and survival curves to measure feature statistical significance. Hypothesis tested features with log-rank tests and Kaplan-Meier curves
- Analyzed viability of new DNA sequencing technique for lung cancer prognosis to model patient survival times and outcomes with Cox-proportional hazards, survival support vector machines (ssvm), and random survival forests
- Tuned L2 strength, learning rate, and kernel of ssvm model with stratified k-fold cv, yielding 0.83 C-Index

### Software Engineer Co-Op

*ASICS Digital*

Jul 2023 – Dec 2023

*Boston, MA*

- Automated previously manual process to verify metadata overlap with **Python** scripts to inform key design decisions during migration to a new Customer Data Platform. Verified over 2 million records
- Authored and optimized **SQL** queries, assisting in GDPR-user compliant deletion for over 1.7 million customers
- Utilized **Postman API** to A/B test time thresholds for email delivery in a global marketing campaign
- Directed 3 team workshops to elevate team skill sets in **Postman API**, **SFMC integrations**, **Git VCS**

## PROJECTS

### Online Retail Analysis | *Python, Pandas, Numpy, Matplotlib, Sci-kit Learn*

Nov 2023 – Dec 2023

- Conducted data pre-processing on UCI online retail dataset to ensure quality and consistency, implementing RFM (Recency, Frequency, Monetary) analysis to segment 4,000 customers based on purchase behavior
- Developed a Random Forest predictive model, fine-tuned with hyperparameter optimization, achieving a 0.76 R-squared value. This robust model was trained on a dataset with 100,000+ transactions

### Clinical Heart Failure Bayesian Analysis | *R, RMarkdown, rstanarm, ggplot2*

Nov 2023 – Dec 2023

- Executed bayesian logistic regression techniques, implementing an MCMC simulation-based model to predict patient survival from a dataset of 299 heart failure cases
- Achieved 0.87 accuracy using Bayesian logistic regression model, incorporating prior knowledge
- Applied posterior prediction and prior prediction checks to ensure validity of statistical assumptions

## INTERESTS

**Hobbies:** Cooking, Ultimate Frisbee, Reading, Golf, Learning New Languages, Running