

# Nikita Jayaprakash

Pittsburgh, PA | (925) 353-6120 | [nikita.jp@gmail.com](mailto:nikita.jp@gmail.com) | <https://github.com/nikita-jaya> | [www.linkedin.com/in/nikita-jaya](https://www.linkedin.com/in/nikita-jaya)

## EDUCATION

<b>Incoming Ph.D. in Demography</b>   University of California, Berkeley	Expected Start Date: August 2026
<b>Masters of Science in Applied Data Science</b>   Carnegie Mellon University	2025 - 2026
<b>Bachelor's of Arts in Statistics</b>   University of California, Berkeley	2022 - 2025
<i>Cluster in Mathematics, Minors in Demography and English</i>	
<b>Pembroke Cambridge Summer Programme</b>   University of Cambridge	Summer 2023

## SKILLS & QUALIFICATIONS

- **Programming Languages:** Python (pandas, numpy, matplotlib, xgboost, scikit-learn, pytorch, streamlit), R (dplyr, tidyr, lubridate, ggplot2, caret, xgboost, randomForest, purrr, Shiny), SQL, GitHub, LaTeX, Microsoft Suite (Office, Excel, Powerpoint)
- **Technical Skills:** practical applications of statistics and AI, data-driven best practices, machine learning, analytical problem-solving
- **Soft Skills:** interpersonal skills, written and oral communication skills, public speaking, organized, leadership skills, creativity
- **Domain Experience:** academic research, education, academic writing (IMRad), open-source software, nonprofit environments

## ACADEMIC EXPERIENCE

**Fellow at NSF AI Institute for Societal Decision Making** | Carnegie Mellon University | Pittsburgh, PA **Aug 2025 - PRESENT**

- Received one of two inaugural NSF AI-SDM Fellowships for CMU master's of applied data science students
- Presented at and attended workshops for 50+ hybrid participants (for industry and general public) on generative AI Literacy
- Prepared research and collaborated with researchers at CMU and affiliated programs with weekly meetings to provide updates

**Research Assistant** | UC Berkeley | Berkeley, CA **Aug 2025 - PRESENT**

- Assisted in experiment that compares grading accuracy across experts (faculty), student graders and multimodal LLMs (AI)
- Developed workflow to streamline study design while maintaining student privacy (CITI program: Social & Behavioral Research)
- Implemented Python scripts to remove identifiable data and R package to compare the grading accuracy across the grading methods

**Instructor of Data Science Lab with R** | UC Berkeley | Berkeley, CA **June 2025 - July 2025**

- Taught 6-week summer course to 16 advanced high schoolers via UC Berkeley's Academic Talent Development Program
- Modified curriculum to include a final project using R Shiny while maintaining STEM course and program objectives
- Employed flipped-classroom learning environment to maximize hands-on coding and statistical experience with real-world datasets

**Data Science Student Researcher** | Aspiranet | Berkeley, CA **Sept 2024 - Dec 2024**

- Presented in collaboration with 5 students about therapy attendance of foster youth at UC Berkeley's Discovery Symposium
- Utilized data analytics and statistical software to identify patterns in cancellation behavior over 3 years of data of 2000+ clients

**Statistics Student Researcher** | UC Berkeley | Berkeley, CA **Feb 2023 - May 2025**

- Co-founded *NemoGB*, an instructor-facing learning management system deployed on main Berkeley software repository
- Received \$34,000 grant from Statistics Department, impacting ~10,000 students across large data science and statistics courses
- Developed R software and UI with R Shiny while ensuring user flexibility and confidentiality of student data

**Course Teaching Assistant and Tutor** | UC Berkeley | Berkeley, CA **Jan 2023 - May 2025**

- Appointed to help students via Statistics Scholar program (intended to assist students of diverse backgrounds within the course) by communicating with students, planning group tutoring sessions and conducting quiz reviews
- Supported 100+ students and coordinated with professor in interactive lectures and assignments in an introductory statistics course

## PROFESSIONAL EXPERIENCE

**Development Research Intern** | KQED | San Francisco, CA **Jan 2025 - June 2025**

- Conducted qualitative research of 100+ donor profiles and strategic event briefings for Development and Executive Management
- Performed routine database updates — including uploads of 286 bios, 100+ donor pictures, and 85 inter-donor relationships — to optimize document management; provided data interpretation of reports of financial summaries of public media grants

**Data, Audits & Contracts Intern** | Fortune Society | Queens, NY **June 2024 - Aug 2024**

- Audited various Excel reports in Education & Employment Services department — including data about reported demographics, job placements and client assessments — for the fiscal year in order to assess contract compliance
- Verified education attendance sheets and finance submission logs to assess accuracy of reporting and record keeping of client data

**AI / Data Science / ML / Developer Intern** | Hyphenova | Remote **Jan 2024 - May 2024**

- Led exploratory data analysis team in guiding decision-making with quantitative analysis of 700+ social media creators
- Built a Shiny app (in Python) to allow for quick analysis of creators in order to assess accuracy of statistical modeling

## PUBLICATIONS

- **Jayaprakash, Nikita**, Voutov, Monika, Turner, Zach, and Andrew Bray. NemoGB: An R Package for Calculating Course Grades. *Journal of Open Source Education*, 8(94), 303. <https://doi.org/10.21105/jose.00303>.

# Nikita Jayaprakash

Pittsburgh, PA | (925) 353-6120 | [nikita.jp@gmail.com](mailto:nikita.jp@gmail.com) | <https://github.com/nikita-jaya> | [www.linkedin.com/in/nikita-jaya](http://www.linkedin.com/in/nikita-jaya)

## PROFESSIONAL & ACADEMIC PROJECTS

### **Vooks Behavioral Signals in User Engagement Capstone Project** **Spring 2026**

- Predicted trial conversion and subscription churn with machine learning models (LASSO logistic regression, XGBoost, balanced random forest) and survival analysis techniques (time-varying cox proportional hazard model) for a storybook streaming platform
- Presented actionable business insights based on stakeholder input, data-driven recommendations, and qualitative industry research
- Developed skills in project planning, stakeholder management, and real-world applications of predictive machine learning models

### **Deep Learning and Clustering Analysis of Gene Expression Project** **Spring 2026**

- Utilized applications of deep learning (variable auto-encoders), statistical genomics, and high-dimensional inference
- Partition the 3005 cells into biologically meaningful groups that align with known cell types and generated low-dimensional embeddings that provide insight into the underlying structure and separability of the single-cell RNA sequencing of mice

### **U.S. Weekly Gasoline Prices Time Series Project** **Spring 2026**

- Utilized vector autoregressive methods in order to analyze the relationship of American gasoline prices with crude oil prices and comparing regional gasoline prices within the United States in order to model, forecast and understand US gasoline price dynamics

### **ETL (Extract, Transform, Load) Data Pipeline Project** **Fall 2025**

- Designed database structure and table schema for Postgres SQL database design with a team of 2 other graduate students using large data sets from Health and Human Services and hospital quality information from Centers for Medicare and Medicaid Services
- Develop reports and data analytics using Python's Streamlit to summarize data with benchmarking

### **Goodreads Textual Analysis Project** **Fall 2025**

- Developed a web scraper to scrape 2713 Goodreads descriptions of books banned in the United States from 2023 to 2024
- Wrote a full IDMRaD report that used multi-dimensional factor analysis of the Biber features and keyness in order to determine literary features and literary differences between the Goodreads descriptions of more-banned and less-banned books

### **HIV Treatment Data Analysis Project** **Fall 2025**

- Utilized linear regression models to analyze 2139 observations from a randomized clinical trial for HIV treatment data
- Investigated prolonged clinical benefit from drug intervention and responses to HIV therapy according to progression of disease and prior HIV drug exposure using observational causal inference methodologies

### **KQED Donor Portfolio Analysis Project** **Spring 2025**

- Utilized data-driven approach and data visualizations of 1200+ donors for cross-portfolio comparison and continued donor giving
- Developed an interactive 15-tab Excel spreadsheet to Development team in order to guide future business strategy and feasibility
- Identified key indicators of a donor's potential for increased giving, retention and acquisition rates within an MGO's portfolio

### **Demography Course Feedback** **May 2025**

- Advised professor on previously-taken course (Demog C126 - Sex, Death & Data), including syllabus feedback and strategies to increase attendance and participation and minimize academic dishonesty via AI usage

### **Fortune Society Demographic Data Analysis Capstone Project** **Summer 2024**

- Analyzed 1450+ formerly incarcerated clients' data (e.g. demographics, career milestones) over the past two years that resulted in a focused effort to address areas of outreach to increase program retention