

# SAMREENA SIDDIQUI

samreena@berkeley.edu || LinkedIn || Portfolio || (408) 835 - 3864

## EXECUTIVE SUMMARY

With a strong background in data science and analytics, I am seeking internship opportunities to collaborate with engineering teams on designing software solutions that positively impact society and enhance daily living. I am passionate about using data science to develop innovative solutions that increase efficiency and streamline processes across industries, leveraging my experience in machine learning, data visualization, and big data analytics with teamwork and collaboration to drive data-informed decision-making.

## EDUCATION

### University of California, Berkeley

Graduation: May 2026

B.A. Data Science, B.A. Cognitive Science | GPA 3.96

#### Relevant Coursework

Artificial Intelligence, Data Structures, Foundations of Data Science, Linear Algebra and Differential Equations, Principles & Techniques of Data Science, Probability of Data Science and Statistics, The Structure and Interpretation of Computer Programs

## EXPERIENCE

### Applied Materials

Berkeley, CA

Data Science Intern

September 2024 - Present

- Designing and implementing a data cleaning and extraction pipeline utilizing OCR and computer vision technologies (Tesseract, PaddleOCR) to extract relevant data from various forms of papers for semiconductor performance analysis.
- Training and deploying a Large Language Model to generate summaries of key semiconductor performance data, aiding company decision-making.
- Developing a twin signal model of semiconductor signals to predict change with 85% accuracy, improving testing safety and operational efficiency.

### Jacob's Ladder Therapies

San Jose, CA

Data Science Intern

June 2024 - August 2024

- Developed and implemented advanced data visualization and analytical models on client company review datasets, resulting in a 30% improvement in predictive accuracy for service performance and operational efficiency.
- Optimized survey design using statistical techniques such as feature engineering and modeling, increasing response rate by 15%, and applied data analysis to generate actionable insights that informed business strategies for over 100 clients.
- Powered data mining and analysis on 200+ Qualtrics legacy surveys, uncovering key trends and patterns that drove data-driven decision-making in marketing and project management across the organization.

### UC Berkeley College of Computing, Data Science, and Society

Berkeley, CA

Data Science Academic Tutor

August 2024 - Present

- Collaborate with data science instructor and team of 20 other tutors to support a class of 1600 students, providing one-on-one tutoring to 10 students per section on data science fundamentals, including linear regression, distributions (hyper-geometric, binomial, normal) and classifiers.
- Conduct lab sections and grade students based on accuracy, ensuring comprehensive understanding of course material.
- Facilitate weekly office hours to address individual student questions using personalized worksheets and mini-lectures.

### The Gomez Lab || Computational and Molecular Biology

Berkeley, CA

Undergraduate Researcher

May 2023 - April 2024

- Utilized R for RNA sequencing, converting biological files (FASTA, FASTQ) into amino acid sequencing data.
- Shadowed and was mentored by in lab computational biologist.
- Designed over 30 + experiments to assess the strength of neuronal connections, conducting extensive genotyping and PCR tests to test results.

## LEADERSHIP

### Neurotechnology at Berkeley || Software

Berkeley, CA

Dementia Detection

August 2024 - Present

- Actively developing machine learning algorithms [Decision-trees, Logistic regression, PCA] aimed at predicting the presence of dementia utilizing voice biomarkers with 92% accuracy through IDE of VSCode.
- Engaging in the process of cleaning voice recording data to effectively compress and diminish noise, facilitating subsequent analysis and further experimentation.
- Conducting comprehensive literature reviews, identifying key variables utilized in published papers to inform the foundation of project on Github.

### Brain Exercise Initiative

Berkeley, CA

President

August 2022 - Present

- Expanded club membership from 25 to 60 active participants within a single semester, fostering a more vibrant and engaged community.
- Conducted weekly team meetings, facilitating discussion, creating weekly summaries, and coordinating tasks for other executive members.
- Volunteering with seniors with dementia to enhance cognitive abilities through reading, writing, and conversation.

## SKILLS

**Coding Languages:** Advanced - Python, Java, SQL, Pandas, NumPy, SciKit, Matplotlib, Seaborn, PySpark; Proficient - R, NodeJS

**Skills:** Statistical Modeling, Data Wrangling, Data Mining, Power BI, Tableau, Data Modeling, Data Visualization, Cloud Computing, Data Cleaning

**Interests:** Machine Learning, Big Data Analytics, AI, Deep Learning, Business Intelligence, Behavioral Data Analytics, Natural Language Processing