Kaitlyn Widjaja

(650) 501-8006 • kaitlynwidjaja@berkeley.edu • linkedin.com/in/kaitlynwidjaja • Portfolio

EDUCATION

University of California, Berkeley, GPA: 3.6 Berkeley, CA
Bachelor of Arts in Data Science Expected May 2026
Relevant Courses: Machine Learning, Data Mining & Analytics, Structure and Interpretation of Computer Programs, Probability & Statistics, Data Science with Python, Business Analytics, Linear Algebra, Differential Equations, Macroeconomics, Microeconomics

Skyline College, GPA: 4.0 San Bruno, CA
Associate in Science in Mathematics and Economics June 2024

SKILLS

Programming & Data Technologies: Python, SQL, Java, C++, Apache Spark (PySpark & Spark SQL), Pandas, NumPy, Scikit-Learn, PyTorch, FastAPI, Django, REST APIs, NLP, Model Evaluation

Data Engineering & Infrastructure: ETL Development, Data Pipeline Design, AWS (RDS, S3), Azure AI Speech, Apache Hive, ClickHouse, MySQL, PostgreSQL, MongoDB

Data Visualization & Tools: Tableau, Matplotlib, Seaborn, Grafana, Excel, SAS, Jupyter Notebook, Git, Google Workspace

Certification: IBM Data Science Professional Certificate (Coursera)

Languages: English (Fluent), Indonesian (Native), Chinese (Conversational)

RELEVANT EXPERIENCE

TikTok US Data Security San Jose, CA
Data Engineer Intern May 2025–August 2025
- Led the design of **two** high-impact dashboards for critical storage infrastructure, collaborating with stakeholders to define requirements and developing end-to-end ETL pipelines (Spark, SQL) to integrate multi-source data
- Reduced SRE on-call resolution time by **96%** (1 hour to 2 minutes) and provided insights driving millions in cost savings
- Automated compliance and legal inquiry workflows, improving visibility and saving **40–100** manual hours per request

Qlay Technologies Berkeley, CA
Software Development Intern February 2025–May 2025
- Engineered an AI-powered speech analysis system (NLP, ML) to evaluate pronunciation, fluency, and behavioral patterns, aiding real-time anomaly and potential fraud detection
- Developed backend APIs (FastAPI, PostgreSQL) for model integration, data collection, speech and behavioral analysis
- Created analytics frameworks to process multi-modal inputs (speech, gaze) and generate actionable performance insights

Excidia AI Customs Clearance and Compliance Berkeley, CA
Data Scientist November 2024–May 2025
- Built Python scripts to scrape, validate, and structure datasets of **5,000+** HS codes, ensuring accuracy and deduplication
- Evaluated ML model performance, collaborated with engineers to design a chatbot for automated HS code classification
- Selected as one of 50 startups in Berkeley SkyDeck's Batch 19 Pad-13 Incubator Program, meeting with importers and exporters to gather requirements and present AI-driven customs clearance solutions

Pawty Berkeley, CA
Data Science & Operations Strategy Intern September 2024–December 2024
- Conducted EDA with Python and SQL to uncover trends, optimize retention strategies and boost user acquisition by **10%**
- Built and managed Tableau dashboards, integrating SQL queries to deliver actionable insights for decision-making
- Developed an AI platform leveraging Instagram API and diffusion models to analyze dog–parent dynamics, generate personalized insights, which increased engagement by **15%**

PROJECTS AND LEADERSHIP

Impact of Educational Games, Phi Theta Kappa Honor Society Beta Theta Omicron San Bruno, CA
Research Lead and Co-President June 2022–February 2024
- Directed **11**-member research team for Phi Theta Kappa Honor Society, boosting student interaction by **40%** and 'A' grades by **200%**; earned Distinguished Honors in Action award and ranked Top **41** of **1,200+** chapters globally

Enhancing Resource Access for Students to Navigate College San Bruno, CA
Research Project Lead August 2023–May 2024
- Developed a QR code keychain system and optimized website UX/UI to improve access to academic and wellness resources for **250+** students, in collaboration with campus organizations

Road Safety Improvement Datathon San Bruno, CA
2nd Place Winning Team April 2024
- Analyzed **100+** data points with Pandas/NumPy in **48** hours, identifying patterns in crash types, locations, and lane conditions; presented **10+** actionable road safety recommendations to judges