# Alejandro Rodas

New York City, NY | +1 (510) 973-9134 | LinkedIn | Portfolio Website | alejandro.rodas.1944@gmail.com

## EDUCATION

•**University of California, Berkeley (CDSS)** *Aug 2021 – May 2025*
*Bachelor of Arts in Data Science (Industrial Analytics Emphasis)* Berkeley, CA
  **Relevant Coursework:** Data Engineering, Data Mining & Analytics, Probability for Data Science, Data Inference & Decisions, Principles & Techniques of Data Science, Data Structures & Algorithms.

  **Awards & Scholarships:** Generation Change Scholar, James Hjul Scholar, Allmond Scholar, Albert Job Scholar.

## TECHNICAL SKILLS

**Languages / Libraries**: Python, Java, SQL, HTML, CSS, JavaScript, pandas, PyTorch, GeoPandas, matplotlib, scipy
**DS/DA Skills**: A/B testing, bootstrapping, EDA, dashboards (Tableau, Sigma), AWS, ETL, Bayesian Inference, Databases, Git, Advanced Stats

## EXPERIENCE

•**UC Berkeley Financial Aid & Scholarships** *Mar 2022 – May 2025*
*Data Science Intern* Berkeley, CA
  – Designed Python and SQL-based **data pipelines** and monitoring tools to track usage of scholar spaces.
  – Cleaned and integrated multi-year student and aid datasets into **relational schemas**, optimizing **SQL** queries for large-scale reporting.
  – Designed and analyzed **A/B tests** on outreach emails and student portals, improving FAFSA completion and critical financial-aid form response rates by **12%**.
  – Built **AWS/GCP dashboards** on aid usage, unmet need, retention, and first-gen/low-income outcomes to support policy decisions by financial officers.

•**U.S. Health Department (HHS)** *May 2024 – Aug 2024*
*Data Science Intern* Washington, D.C. (Remote)
  – Developed regression, optimization models and **A/B tests** on CIL funding and utilization data (2019 vs. 2022) to quantify post-pandemic underperformance.
  – Helped to automate **ETL pipelines** using Python and APIs, reducing data processing time by **30%**.
  – Built **BigQuery** and **Tableau** dashboards summarizing funding and usage trends for non-technical decision-makers.
  – Synthesized pre-pandemic and post-pandemic funding and utilization findings with executive directors, contributing statistical analyses and visualizations to a federal report scheduled for release in 2026.

•**CDSS Discovery Research Program** *Jan 2024 – May 2024*
*Lead Researcher* Berkeley, CA
  – Led a research team analyzing federal reports for **350+** Centers for Independent Living, cleaning and standardizing multi-year CSVs into a state-level dataset.
  – Applied Python-based **NLP** sentiment analysis (NLTK/VADER), correlation analysis, and visualizations (**pandas, matplotlib, seaborn, Plotly**) to relate funding, coverage, and reported achievements/challenges.
  – Merged CIL data with U.S. Census disability statistics to engineer coverage and funding-per-disabled-resident metrics capturing "good" CIL performance.
  – Built **Sigma and Tableau** dashboards and presented findings at the 2024 CDSS Data Discovery Spring Symposium to inform discussions on HHS resource allocation and to our policy team based in Princeton (Mathematica, Inc.)

## HIGHLIGHTED PROJECTS

•**Causal Inference and ML (Python) for U.S. Primary Elections** *Spring 2025*
  – Conducted **causal analysis and Bayesian statistics** (e.g. Propensity Score Matching) to estimate endorsement effects.
  – Built and evaluated logistic regression and Random Forest models; **performed EDA on 2,600+ candidates.**
  – Tuned models; best Random Forest achieved **85% accuracy** and 0.78 F1-score; presented findings using **Matplotlib/Seaborn** in a 23-page paper.

•**Yelp Insights and NoSQL Data Processing with MongoDB** *Fall 2024*
  – Queried and analyzed semi-structured JSON data from Yelp business, review, and user datasets using PyMongo.
  – Designed **data pipelines and ETL pipelines** for aggregating and extracting user behavior insights.
  – Applied indexing and query optimization techniques to improve performance on unstructured datasets.

•**Campus Sensor Data Cleaning and Time-Series Interpolation in PostgreSQL** *Fall 2024*
  – Standardized HVAC and energy sensor data in **PostgreSQL**, resolving inconsistent units and building metadata tables.
  – Engineered 15-minute analysis-ready time-series tables using **SQL window functions**, median/MAD outlier handling, and interpolation to fill missing readings.
  – Produced clean, structured datasets used for **exploratory analysis** of campus energy usage.

•**Predicting Streaming Service Churn with Machine Learning** *Fall 2024*
  – Built machine learning models (NNs, Random Forests, Decision Trees) to predict churn rates with **85% accuracy**.
  – Used logistic regression to model churn probability and estimate marginal effects of key engagement features.
  – Evaluated models with ROC curves and AUC.