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
Analytics : A/B testing, bootstrapping, statistical inference, EDA, dashboarding (Tableau, Sigma), AWS, ETL, bayesian inference, Databases

Experience

•UC Berkeley Financial Aid & Scholarships *Mar 2022 – May 2025*

Data Science Intern Berkeley, CA

- Designed pipelines and monitoring tools to determine usage of scholar spaces.
- Standardized data into d **relational formats** and **SQL** queries for performance at scale.
- Built **AWS/GCP dashboards** providing insights for financial officers and supporting policy evaluation.

•U.S. Health Department (HHS) *May 2024 – Aug 2024*

Data Science Intern Washington, D.C. (Remote)

- Developed optimization models and **A/B tests** to evaluate resource allocation and to verify if HHS was underperforming compared to pre-pandemic levels.
- Aided development of automated **ETL pipelines** using Python and APIs, reducing processing time by **30%**.
- Created dashboards in **BigQuery and Tableau** for non-technical decision-makers.
- Collaborated with executive directors on a federal report releasing in 2025.

•CDSS Discovery Research Program *Jan 2024 – May 2024*

Lead Researcher Berkeley, CA

- Led a team analyzing federal datasets, restructuring unorganized records from 350+ centers nationwide.
- Applied **hypothesis testing, A/B testing**, GLMs, NLP, and Geopandas to assess resource allocation.
- Presented findings using **matplotlib, seaborn, Sigma** at the 2024 CDSS Data Discovery Symposium.

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 **Matplolib/ 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 (neural networks, 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.