

Alejandro Rodas

New York City, NY • (510) 973-9134 • alex9102017@berkeley.edu •
LinkedIn: <https://www.linkedin.com/in/alejandro-rodas-/>
Website: <https://alejandro-rodas.github.io/>

EDUCATION

University of California, Berkeley, CA (CDSS)

B.A. in Data Science with Domain Emphasis in Industrial Analytics

Aug 2021-May 2025

RELEVANT COURSEWORK: Data Structures and Algorithms (**DATA STRUCTURES/GITHUB/GRAPHS**); Foundations of Data Science (**EDA, PANDAS, NUMPY**); Principles and Techniques of Data Science (**PIPELINES, SQL**); Probability for Data Science (**STATISTICAL MODELING**); Data Engineering (**ETL, MONGODB, OLAP/DATABASES**); Data Mining and Analytics (**ML ALGS**), Data Inference and Decisions (**ADVANCED STATISTICAL MODELING**)

SKILLS: Advanced SQL, Query Optimization, Window Functions, Joins, CTEs, Subqueries, Data Modeling, Advanced Python, Scikit-learn, TensorFlow, PyTorch, Regression Analysis, A/B Testing, Statistical Inference, Bayesian Inference, NLP, GLMs, Optimization Modeling, AWS, GCP, OLAP, Databases, PySpark, Snowflake, MongoDB (and PyMongo), ETL pipelines, NumPy, Matplotlib, Seaborn, GeoPandas, Sigma, Model Development (Random Forests, Neural Networks), Snowflake, Databricks, etc.

EXPERIENCE

CDSS Discovery Research Program

Lead Researcher

Jan 2024 - May 2024

- Led a team analyzing federal financial datasets, structuring unorganized records from 350+ centers all over the country.
- Designed ETL pipelines and regression models in order to determine efficiency of centers.
- Applied model selection techniques, hyperparameter tuning, and evaluation metrics(AUC, precision/recall) to assess impact.
- Presented findings to policy and research teams unfamiliar with machine learning and data science terms, guiding decision-making through clear visuals. Relying on libraries like matplotlib seaborn and platforms like Sigma.

U.S. Health Department

May 2024 - Aug 2024

Data Science Intern

- Built machine learning models (logistic regression, decision trees) to predict post-pandemic changes in service usage
- Automated ETL pipelines using Python and APIs, reducing processing time for large federal datasets by 30%.
- Created dashboards in BigQuery & Tableau to present risk insights to non-technical decision-makers. Collaborated with executive directors to release a federal report coming in late 2025.

UC Berkeley's Financial Aid & Scholarships

Mar 2022 - May 2025

Data Science Intern

- Wrote and optimized complex SQL queries using joins, CTEs, and window functions to extract and aggregate financial data.
- Standardized raw data into relational schemas and built monitoring tools to track scholar space usage and data quality.
- Developed and deployed ML models (logistic regression, random forests) to predict student aid utilization.

HIGHLIGHTED PROJECTS

Yelp Insights and NoSQL Data Processing with MongoDB

Nov 2024 - Dec 2024

- Queried and analyzed semi-structured JSON data from Yelp's business, reviews, and user datasets using PyMongo, and designed data pipelines for aggregating and extracting user behavior insights.
- Leveraged indexing and performance optimization techniques to enhance query efficiency for unstructured datasets.

Campus Sensor Data Cleaning and Time-Series Interpolation in PostgreSQL

Oct 2024 - Nov 2024

- Standardized campus HVAC and energy sensor data in PostgreSQL, resolving inconsistent units and building metadata
- Engineered analysis-ready 15-minute time-series tables using SQL window functions, robust outlier handling (median/MAD), and interpolation to fill missing readings.

LEADERSHIP

Latin American Leadership Society at Berkeley

Jan 2022 - Jun 2022

Forum Committee Vice President

- Organized events for 100+ attendees using data-driven tools to optimize scheduling and participation. Streamlined event logistics through technology platforms, enhancing communication while addressing complex organizational challenges.