Alejandro Rodas

Berkeley, CA ● (510) 973-9134 ● alex9102017@berkeley.edu ●
Linkedin: https://www.linkedin.com/in/alejandro-rodas-/

EDUCATION

University of California, Berkeley, CA (CDSS) **Aug 2021-May 2025**
B.A. in Data Science with Domain Emphasis in Industrial Analytics

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 - GLMS, NONPARAM MODELS**)

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 **Jan 2024 - May 2024**
Lead Researcher

- 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%.
- Conducted sentiment analysis on user reviews using NLP tools (e.g., VADER, TextBlob) to quantify perception of centers.
- 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

Research Project: Causal Inference, ML for U.S. Primary Elections **Mar 2025 - May 2025**

- Applied sentiment analysis to endorsement texts using NLP to extract polarity scores, used as features in supervised models.
- Built and evaluated logistic regression and Random Forest models to predict election outcomes; performed EDA on 2,600+ candidates across race, gender, and incumbency. Presented findings in a research paper.

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.

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.