SEJAL SARDA

San Francisco Bay Area | +1 (510) 599-7581 | sejal_sarda@berkeley.edu | linkedin.com/in/sejal-sarda-ucb100

EDUCATION

University of California, Berkeley — B.A. Data Science & Statistics (double major), Minor in Sustainable Design
Expected May 2026 • GPA: 3.65

SKILLS

Programming & ML: Python, SQL, JavaScript, R, scikit-learn, PyTorch, TensorFlow, transformers, LLM prototyping, RAG, fine-tuning, Bayesian inference (MCMC)
Databases & Cloud: AWS (Redshift, S3, EC2), GCP, Snowflake, MongoDB
Data Tools: Power BI, dbt, Airflow, Docker
Methods: Feature engineering (PCA, embeddings, One-Hot), A/B testing, clustering & classification, time series (LSTM/transformers), generative modeling, causal inference, NLP

EXPERIENCE

Amazon — Business Intelligence Engineer Intern (RIDE Team) | Summer 2025

- **Created internal product-analytics dashboard** to track configuration-funnel KPIs; authored/validated SQL pipelines and introduced a request identifier that eliminated duplicate joins and improved metric accuracy.
- **Led stretch DS/ML project** with Data Science team to address reproducibility gaps in dashboards; evaluated limitations of static KPI analysis and scoped **LLM-based query generation,**
- **Prototyped an LLM-powered NL→SQL system** (Amazon AI model) with context prompts that injected metric definitions/rules, **improving SQL accuracy** for complex queries and reducing analyst back-and-forth.
- Delivered **3 demos** (dashboard, NL querying, AI prototype) to PMs and engineers; prototype adopted in DS experimentation workflows.

MIT Transit Lab — Research Assistant | Sep 2024 – Present

- Conducted EDA on high-resolution AVL (Automatic Vehicle Location) data to expose bus running-time reliability gaps.
- Developed **Bayesian (MCMC) + ML forecasting framework** to predict bus running times with uncertainty quantification.
- Designed experiments comparing **deterministic models vs. deep learning** for variance capture in operational forecasts.
- Collaborated with faculty on deploying predictions into **Boston MBTA scheduling models**.

UC Berkeley HuMNet Lab — Urban Mobility Data Researcher | May 2024 – Dec 2024

- Extended **TimeGeo framework** with Path-Size Logit for realistic route choice modeling.
- Designed **transformer-based feedback loop** for trajectory imputation in sparse LBS data; explored **generative AI for mobility prediction**.
- Authored internal research proposal on **synthetic population simulation** for benchmarking trajectory models.

Protiviti — Data Science Intern | May – Aug 2024

- Built **3 Power BI dashboards** for a public-sector client; redesign improved activity by ~30%.
- Standardized Power Query/Excel transformations; collaborated on scaling reporting pipelines.

Robolabs — Data Science Intern (part-time) | Dec 2022 – Present

- Automated scoring workflows and delivered **Python ML workshops** for students.
- Supported 30+ robotics tournaments and mentored 60+ teams.

PROJECTS

Truck Parking Forecasting (Discovery Program) | Sep 2024 – Present

- Trained **LSTM models** on 2-minute interval parking data → reduced RMSE by 24% (12.54 → 9.53).
- Extended pipeline to experiment with **transformer sequence models** for longer-horizon forecasting.

Cook County Housing Prices | Feb – Mar 2024

- Built regression pipeline (500k+ records, 61 features) with One-Hot Encoding and residual diagnostics.
- **Feature engineering + fairness analysis** to flag regressive assessment patterns.

AWARDS & LEADERSHIP

Chung Leadership Prize (2025) • John Goddard Prize for Environmental Leadership (2024) • Leadership Award & Scholarship (2023) • College Ambassador, CED (2024–2025) • Student Rep, Urban Studies Committee (2024–2025)