Ryan Leem

Los Angeles, CA | rleem8381@gmail.com | (213) 284-9361 | LinkedIn | GitHub

Professional Summary

Data Science student at UCLA with hands-on experience building ML-adjacent systems, including semantic search pipelines, fraud analytics models, and AI evaluation frameworks. Proficient in Python, SQL, and R with a track record of delivering quantified results across NLP, predictive analytics, and data engineering projects.

Education

University of California, Los Angeles (UCLA)
B.S. Data Science (Intended). *Expected 2029*

Experience

AI Benchmark Developer February 2026 – Present
Handshake AI

- Developed 30+ coding benchmark tasks for AI system evaluation using Python and shell scripting
- Built Docker-based environments and automated verifier tests to evaluate correctness and reproducibility
- Debugged evaluator failures and revised task logic based on reviewer feedback and specification requirements

Projects

GitHub Bug Finder August 2026 – August 2026

- Built a personal portfolio project demonstrating end-to-end semantic search: indexed 11,336 GitHub bug reports using Python, PostgreSQL, pgvector, FastAPI, and sentence-transformers to enable natural-language bug lookup across thousands of existing issues
- Improved top-result retrieval accuracy from 70% to 80% by adding title-aware reranking to a baseline semantic search model
- Integrated live GitHub issue retrieval via the REST API and evaluated retrieval quality on 20 paraphrased bug descriptions

Bank Account Fraud Risk Analytics May 2026 – July 2026

- Analyzed 1M bank account applications from the Feedzai Bank Account Fraud dataset in PostgreSQL to identify fraud patterns and high-risk segments
- Built and validated a rule-based fraud risk score, identifying a segment with an 11.74% fraud rate and more than 10x baseline lift
- Applied CTEs, views, indexing, and EXPLAIN ANALYZE to organize analysis and optimize query performance, with an indexed risk query running in approximately 6-7 ms

Banking Performance Report Analyzer March 2026 – April 2026

- Built an automated monthly banking report in Excel using Power Query, PivotTables, PivotCharts, formulas, slicers, and VBA
- Aggregated multi-million-row transaction data in Power Query before loading into Excel to work around worksheet row limits
- Designed a reusable CSV mapping sheet and VBA refresh workflow to update source data without manually editing Power Query paths

E-Commerce Funnel and Conversion Analysis January 2026 – February 2026

- Analyzed 500K+ e-commerce interactions in R to compare funnel behavior and brand-level conversion performance
- Applied regression analysis to test the relationship between traffic and conversion ($R^2 = 0.0845$, $p = 0.335$)
- Built visualizations and filtered high-traffic brands to improve reliability of conversion comparisons

Activities & Leadership

Coding Club 2024-2025
Vice President

- Mentored club members in Python programming and guided project development across multiple student-led initiatives

PAVA 2022-2024
District President

- Led volunteer teams and coordinated community service initiatives across the district

Skills

Programming: Python, SQL, R, Luau

Databases & Data: PostgreSQL, Pandas, tidyverse, data cleaning, exploratory data analysis, regression analysis

AI / Evaluation: semantic search, embeddings, sentence-transformers, benchmark development, automated testing, model-output evaluation

Analytics & Reporting: Excel, Power Query, PivotTables, PivotCharts, VBA, ggplot2, Matplotlib

Development: Git, GitHub, Docker, FastAPI, REST APIs

Languages: English, Korean