Allen Arriaga

Boston, MA | allenarriaga10@gmail.com | linkedin.com/in/allen-arriaga-a7b08b324 | github.com/allena3103

Education

University of Massachusetts Amherst — Amherst, MA
Bachelor of Science in Informatics — *Expected Dec 2026*

- Relevant coursework: Applied Statistics, Statistical Modeling, Statistics in Psychology, Intro to Data Science, Object-Oriented Programming

Experience

Solo Developer and Product Analyst — Aug 2026 – Present
Lual, iOS app live on the App Store — *Remote*

- Built and operate a group calendar app with **2,900+ registered users** and **1,700+ monthly active users**, and analyze its production PostgreSQL data to decide what to build next
- Designed the tracking and a nightly SQL rollup that records daily active users, signups, onboarding completion, plans, invites and notifications in one daily metrics table
- Wrote a SQL cohort view for Day 1, Day 7 and Day 30 retention across 2,300+ signups, measuring **33% Day 1** and **12% Day 7** retention
- Found that users who joined a friend or group on day one came back the next day at **3x the rate** of unconnected users (43% vs 13%), which led to a redesigned first-session onboarding step
- Measured the invite funnel (39% of 2,200+ invites accepted) and built a private Next.js dashboard that tracks growth, retention and AI feature usage

Projects

Customer Shopping Behavior Analysis | *Python, PostgreSQL, Power BI* — Nov 2025

- Cleaned 3,900 retail transactions in pandas, filled missing ratings with category medians, engineered age-group and purchase-frequency features, and loaded the result into PostgreSQL
- Answered 10 business questions in SQL, covering revenue by segment, discount dependence, top products per category and subscriber versus non-subscriber spend
- Built a Power BI dashboard and recommended stronger subscription incentives and a tighter discount strategy, based on which segments and products drove revenue

Steam Game Popularity Analysis | *Python, MySQL, scikit-learn* — Nov 2025

- Prepared 339 Steam game records in MySQL and pandas, removed outliers with an IQR filter (270 games modeled), and trained Random Forest models to predict ownership and positive reviews (R^2 of 0.86 for reviews)
- Showed through feature importance that review counts and concurrent players predict ownership far better than price or discount size, so deeper discounts alone are a weak growth lever

Hospital Infection Risk Analysis | *R* — Oct 2025

- Fit two multiple linear regression models on 113 US hospitals and compared them with 10-fold cross-validation (RMSE of 1.14)
- Found length of stay to be the strongest predictor of infection risk ($p < 0.001$), while patient age added little once stay was included

Technical Skills

Analysis: SQL (PostgreSQL, MySQL), Python (pandas, scikit-learn, Jupyter), R, Excel
Visualization: Power BI, Tableau, Recharts
Methods: Cohort retention, funnel analysis, KPI design, linear regression, random forest, cross-validation, data cleaning, feature engineering
Other: Git, Supabase, TypeScript, Next.js