

Madhiyogi Jain

630-397-8920 | mjjain@umass.edu | [linkedin.com/in/yougijain](https://www.linkedin.com/in/yougijain) | github.com/yougijain

EDUCATION

University of Massachusetts Amherst

B.S. Informatics & Psychology (Double Major)

Amherst, MA

Expected May 2027

PROJECTS

[NYC Motor Vehicle Collisions Dashboard](#) | *Python, SQL, SQLite, Streamlit, PyDeck* | [GitHub](#) Jan 2026

- Deployed a public Streamlit dashboard; added **caching** (600s TTL) and **pytest**/CI checks to support reliable analytics delivery.
- Built a NYC motor vehicle collisions dataset spanning **Jan 2020 – May 2025 (27,164 records)**; analyzed **13,508** injuries and **64** fatalities across **5 boroughs**.
- Authored **4 SQL** analyses for severity filtering, borough-level aggregation, hour-of-day patterns, and monthly time-series trends; validated **23** fields (**12/23** at $\geq 95\%$ completeness).

[Bank Customer Churn Prediction \(R Report\)](#) | *R, Classification, Random Forest* Dec 2025

- Built an end-to-end data science report to identify customers at risk of leaving and inform retention strategy.
- Performed **EDA** on **10,000** customers (**14** features); quantified class imbalance (**20.4%** churn) and summarized churn drivers with visualizations.
- Compared logistic regression vs. Random Forest; evaluated with **5-fold cross-validation**.
- Analyzed **credit risk** signals and decision thresholds; framed results for risk appetite and risk management decisions.

[Plug: Campus Marketplace](#) | *PostgreSQL (Supabase), SQL* Aug 2025

- **Selected for Purdue's Market Readiness Incubator**; served as project manager, supporting ERM-style project governance and program implementation milestones.
- Architected an analytics-ready backend database model enabling listings/search, transaction tracking, and reporting (inventory, demand, user activity).
- Designed and validated a Supabase (PostgreSQL) relational schema aligned to data architecture and data governance requirements.
- Implemented **SQL** for **CRUD** and search/filter queries (multi-field filters, sorting, pagination) and verified via testing and integrity checks.

EXPERIENCE

Spiritual Data

AI & Psychology Research Intern

Remote

May 2025

- Built a benchmarking harness to compare a proprietary model vs. baselines using **3+** task-appropriate metrics; automated evaluation runs and aggregated outputs for consistent comparison.
- Designed **evaluation splits** and **error slices** (outliers/edge cases) to surface failure modes and support model selection.
- Produced a concise evaluation research report for stakeholders; aligned findings to business objectives and compliance requirements.

UC Berkeley — Lawrence Hall of Science

Research Intern, Impact of Social Media

Berkeley, CA

Jun 2022

- Cleaned and prepared survey data from **700+** participants, producing analysis-ready datasets for statistical modeling and reporting.
- Conducted quantitative data analysis (EDA, statistical regression) on how video content influences public attitudes; synthesized findings for stakeholders.

TECHNICAL SKILLS

Languages: Python, SQL, R, Java

Data: Pandas, NumPy, ETL, SQLite, PostgreSQL (Supabase)

Analytics/BI: Streamlit, Excel, Tableau, Power BI, KPI design, dashboards

ML/Stats: scikit-learn, classification/regression, model evaluation, statistics

Tools: Git/GitHub, pytest, CI/CD, caching, GenAI/LLM tools (LangChain, Copilot)