Samuel Mai

samicmof@gmail.com | linkedin.com/in/samuel-mai | samuelmai.com

Education

University of California, San Diego
Bachelor of Arts in Economics; Minor in Data Science — *San Diego, CA*

- Specialization in Data Analysis. Relevant coursework: Econometrics, Statistics, Data Structures and Algorithms, Database Systems in SQL, Data Science in Practice, Linear Algebra, and Recommender Systems.

Projects

FantasyDex | *Python, scikit-learn, TypeScript, Next.js, PostgreSQL*

- Built an end-to-end NFL fantasy platform: a Python modeling pipeline producing full-PPR season projections for 762 players and defenses, served through a Next.js and TypeScript web app backed by PostgreSQL.
- Assembled 16 seasons (2010–2025) of player, team, roster, and play-by-play data into checksummed snapshots with a cross-source player identity crosswalk and schedule-normalized point and availability labels.
- Trained position-specific ridge and random-forest models using rolling time-based backtests to prevent data leakage; models cut point error 9–10% for QBs and WRs, but failed pre-set subgroup checks, so calibrated baselines shipped.
- Shipped projections through a validated JSON data contract with one-command rollback, and set up CI on GitHub Actions running typechecks, unit, database, and Playwright end-to-end tests on every pull request.

Inside MIPS | *TypeScript, three.js, Vite, Vitest, Playwright*

- Built a browser-based, MARS-compatible MIPS R2000 emulator and assembler in strict TypeScript (13.8K lines), supporting all 61 MIPS I integer instructions, 62 pseudo-instructions, and 15 MARS syscalls.
- Modeled single-cycle and 5-stage pipelined CPUs with toggleable forwarding, load-use and branch hazard stalls, flushes, and precise exceptions, plus set-associative L1/L2 caches with LRU, FIFO, and random replacement.
- Animated every clock cycle on an interactive three.js 3D datapath driven by real emulator traces, with 10 guided lessons, reverse stepping via deterministic replay, and 287 unit and 18 end-to-end tests.

Student Survey Data Platform | *UCI Science of Learning Lab, Python, DuckDB, SQL*

- Built a survey data pipeline under the mentorship of Dr. Joseph Wong, consolidating pre/post questionnaire surveys, Canvas grades, and course-analytic data into a unified student-level dataset.
- Automated ingestion, cleaning, and schema standardization of raw survey files, linking survey responses to Canvas records and de-identifying data prior to analysis.
- Stored processed data in DuckDB, wrote SQL to compute question-level response rates and automate participation scoring, and visualized pre/post trends to evaluate student engagement.

Student Dropout Analysis | *Python, pandas, scikit-learn*

- Implemented a reusable scikit-learn pipeline combining categorical encoding, feature scaling, and logistic regression to classify students at risk of dropping out.
- Conducted hypothesis tests on population proportions to identify six significant financial, academic, and personal factors associated with dropout.

Experience

Sales Representative — October 2023 – November 2023
Aspire Team — *San Diego, CA*

- Generated leads through face-to-face outreach, recommending Verizon products against weekly sales targets and managing follow-ups through the sales cycle.

Squad Leader — January 2020 – February 2020
Chinese Chamber of Commerce — *San Francisco, CA*

- Led volunteer preparation and on-site operations for San Francisco's annual Chinese New Year Parade, coordinating setup with 20+ vendors and resolving attendee issues in real time.

Technical Skills

Languages: Python, TypeScript, SQL, Java
Frameworks & Libraries: Next.js, React, three.js, Tailwind CSS, pandas, NumPy, scikit-learn
Data & Visualization: PostgreSQL, Drizzle ORM, DuckDB, Streamlit, Tableau, Excel
Methods: Statistical Analysis, Hypothesis Testing, Supervised Learning, Time-Series Backtesting
Developer Tools: Git, GitHub Actions, Vite, Vitest, Playwright, Vercel, Jupyter