Joseph Saldivar

Berkeley, California josephsaldivarg@gmail.com (559)385-3832 in/josephasaldivar josephsaldivar.dev

Education

Bachelor of Arts in Data Science — Berkeley, CA

University of California, Berkeley — *May 2025*

- Coursework: Data Structures, Data Engineering, Business Analytics, Machine Learning, Data Mining and Analytics, Industrial and Commercial Data Systems, Discrete Mathematics and Probability, Calculus, Structure & Interpretation of Programs, Principles & Techniques of Data Science

Experience

Data Annotation — Remote

Software Developer — *May 2024 - Present*

- Partnered with AI/ML engineering teams to validate and refine large language model (LLM) code outputs, improving execution accuracy by 20%+ and reducing defect rates across production-level workflows.
- Designed and maintained automated test environments leveraging virtual environments, dependency managers, and CI/CD pipelines, cutting environment setup time by 30%+ and ensuring consistent, reproducible builds.
- Debugged and optimized AI-generated applications in Python, Java/JS, and HTML, enhancing system reliability and scalability across client-facing use cases.
- Authored structured evaluation frameworks to identify recurring model errors, providing actionable insights that informed retraining cycles and reduced repeated code issues by 15%+.

Spotify Data Visualizer - trackgraph.josephsaldivar.dev — September 2025 - Present

Independent Developer/ Full-Stack Developer

- Shipped a full-stack Spotify analytics app used by 60+ users, analyzing 40k+ listens across 8 years with interactive bubble & leaderboard visualizations.
- Integrated Spotify Web API alongside custom FastAPI which performed batched fetches and an S3-backed image/metadata cache; cache-first hydration serves fresh artwork while cutting external calls by 95%+, only requesting when new data is detected.
- Optimized data efficiency with vectorized Pandas aggregations and timeframe-keyed memoization, reducing redundant recompute by 66%.
- Productionized on AWS (App Runner, ECR, S3/CloudFront, IAM least-privilege) with a one-command CI/CD pipeline, enabling safe, repeatable rollouts with automated container build/push, and cache versioning/invalidation on release.

Cal Field Hockey — Berkeley, CA

Database Design Consultant — *September 2024 - December 2024*

- Worked in an 8-member team to design and implement a relational database supporting player, event, and merchandise management; contributed to requirements gathering, schema design, build, and client delivery.
- Created 62 BCNF-normalized tables in MySQL Workbench; integrated historical and synthetic datasets to enable high-volume testing and ensure scalability.
- Developed SQL and Python analytics pipelines, including a Random Forest–based MVP prediction model and a lineup optimization tool, validated against historical results to support roster selection decisions.
- Worked directly with club stakeholders to align DB features and analytics with operational needs, ensuring usability and maintainability in a production setting.

State Center Community College District — Fresno, CA

Math Tutor/Mentor — *November 2021 - August 2022*

- Assisted students in various math academic courses and assessed learning styles of each student, tailoring a tutoring style to each individual.
- Facilitated one-on-one tutoring sessions through the cultivation of personalized connections with students, ensuring equitable tutoring.

Projects

Lock & Load – Playstyle Classification System for Borderlands Weapons

- Developed a weapon playstyle classification model using K-Means clustering and Random Forests within a Python-based environment, leveraging Scikit-Learn and Pandas for efficient data manipulation and modeling, in order to help gamers make better decisions.
- Created an end-to-end data pipeline leveraging Python and OCR tools (EasyOCR, PyTesseract) to programmatically capture and extract weapon statistics from 30+ hours of real-time gameplay video frames, then used PCA to identify key attributes influencing player playstyles.
- Processed and cleansed 4,700+ data records using Python and Pandas, correcting OCR noise, normalizing attributes, and calibrating for level scaling; accelerated data preprocessing phase, supporting advanced analytics modeling.

Skills

- **Languages :** Python, JavaScript (React), Java, SQL, HTML • Spanish (bilingual)
- **Stack :** NumPy, Pandas, Scikit-Learn, AWS (App Runner, ECR, S3, CloudFront, IAM), Docker, Plotly, CI/CD, Git, Jupyter, PostgreSQL, FastAPI, Excel
- **Concepts :** Database design, Caching, ETL/ELT Pipelines & Data Integration, Object-Oriented Programming, Unit Testing, SDLC, Adaptive Problem Solving