

Joseph Saldivar

📍 Berkeley, California ✉ josepha@berkeley.edu ☎ (559)385-3832 📧 in/josephasaldivar 🌐 www.josephasaldivar.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

May 2024 - Present

Software Developer

- 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%+.

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.

Cook County Housing Exploratory Analysis

- Conducted an extensive exploratory analysis of the Cook County housing market, involving the review and analysis of over 200,000 property records to find insights and draw conclusions that matter.
- Engineered a predictive linear regression model leveraging Python's Sklearn, Numpy, and Pandas libraries to accurately estimate property sale prices based on 60+ distinctive attributes.
- Implemented data visualization tools such as Matplotlib and Seaborn to effectively communicate complex data trends and insights to non-technical stakeholders who might not be familiar with the data.

Skills

- **Languages** : Python, Java, JavaScript, React, SQL, HTML, Spanish (bilingual)
- **Libraries & Tools** : NumPy, Pandas, Scikit-Learn, Seaborn, Plotly, Matplotlib, Git, Jupyter, PostgreSQL, MySQL, VS Code, Excel
- **Concepts** : Database design & Normalization, ETL/ELT Pipelines & Data Integration, Object-Oriented Programming, Unit Testing, SDLC