

Diego Osborn

diegoisaacosborn@gmail.com linkedin.com/in/diego-osborn d2osborn.github.io

EDUCATION

University of California, San Diego

Jun 2026

B.S. Data Science; Minors in Economics and Mathematics

- Graduate (cross enrollment): Bayesian Inference, Hierarchical & Probabilistic Modeling
- Principles & Techniques of Data Science, Monte Carlo Methods, Relational Databases, Cloud Computing, Scalable ML

SKILLS

Python (*PyTorch, scikit-learn, PyMC, Pandas, NumPy*), R, SQL (*PostgreSQL, SQLite*), Dask, AWS, Web (*JavaScript, HTML, CSS, Streamlit*), Java, Git, Spanish (Fluent)

EXPERIENCE

UC San Diego Baseball

Feb 2023 – Present

Data Science Intern

La Jolla, CA

- Designed a semi-supervised pipeline using *Expectation–Maximization* and *Gaussian Mixture* models to resolve pitch label noise, training *gradient-boosted trees* for robust pitch classification.
- Implemented a conditional *gradient-boosting ensemble* to quantify Stuff+ (pitch quality), using probability weighting to isolate the *intrinsic signal* from stochastic environmental noise.
- Developed a *Monte Carlo simulation* utilizing historical scoring data to forecast expected win ranges under varying schedule scenarios to support strategic planning.
- Built a Python data ingestion pipeline to migrate 1.1M+ rows of raw Trackman CSVs into *PostgreSQL*, creating a centralized repository to eliminate manual file merging.

UC San Diego Career Center

Sep 2024 – Jun 2025

Data Engineering Intern (Data Analyst)

La Jolla, CA

- Developed a reproducible Python *ETL workflow* in Jupyter Notebooks to integrate Qualtrics survey data with manual LinkedIn records, establishing a single source of truth for graduate outcomes reporting.
- Authored technical SOPs for version control and environment management, defining *Git* protocols (commit standards, conflict resolution) to standardize development across the analytics team.
- Migrated 100,000+ student career Handshake profiles to a local PostgreSQL database, utilizing SQL to ensure compliance with university data retention policies and deliver actionable datasets.

Palm Springs Power Baseball

Jun 2023 – Jul 2023

Data Analytics Intern

Palm Springs, CA

- Managed daily tracking data operations, maintaining the league repository, and ensuring up-to-date statistical reporting for the official league website.
- Deployed an interactive Looker Studio dashboard to visualize KPIs, transforming raw game data into actionable insights to support coaching and player development.

PROJECTS

Quantifying Defensive Aggression With a Bayesian Hierarchical Model

Jun 2025 – Aug 2025

- Formulated “Aggression Over Expected” (AOE) by engineering a Bayesian hierarchical framework that contrasts individual player posteriors against positional counterfactuals, utilizing MCMC sampling and partial pooling to regularize estimates across limited sample sizes.
- Trained a Logistic Generalized Additive Model (GAM) to quantify catch probability from spatiotemporal tracking data, achieving 0.162 test log-loss and 95.7% test F1.
- Earned honorable mention recognition as a Top 10 finalist among 50 teams (114 students) in the 2025 SMT Data Challenge.

The Effect of Two Strikes on a Hitter’s Swing

May 2025 – Jun 2025

- Constructed an interactive dashboard using *Python* and *D3.js* to process MLB Statcast tracking data, conducting exploratory analysis on swing mechanic shifts in high-pressure contexts.

What Is a Home Worth? Large-Scale Ensemble Learning for U.S. Housing Prices

Feb 2025 – Mar 2025

- Architected a predictive pricing pipeline for 2.2M+ U.S. properties, synthesizing heterogeneous data sources (crime, tax, demographics) into a gradient-boosting framework to capture complex market drivers.