

# Diego Osborn

[diegoisaacosborn@gmail.com](mailto:diegoisaacosborn@gmail.com) [linkedin.com/in/diego-osborn](https://www.linkedin.com/in/diego-osborn) [d2osborn.github.io](https://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, D3.js, HTML, CSS, Streamlit), Java, Git, Spanish (Fluent)*

## EXPERIENCE

### UC San Diego Baseball

Feb 2023 – Present

*Data Science Intern*

La Jolla, CA

- Automated the end-to-end generation of opponent scouting reports using Python, reducing weekly preparation time by 90% to streamline pre-game strategy.
- 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.
- Developed a *Monte Carlo simulation* running 10,000+ iterations on 3 years of historical data to forecast expected win ranges across 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 Analytics Intern*

La Jolla, CA

- Developed a reproducible Python *ETL workflow* integrating 2,000+ Qualtrics responses with LinkedIn data for 12,000+ graduates, establishing a single source of truth for outcome reporting.
- Authored technical SOPs for version control and environment management, defining *Git* protocols to reduce code conflicts and standardize development for a *4-person analytics team*.
- Migrated 100,000+ student career Handshake profiles into a local PostgreSQL database, using SQL to ensure compliance with university data retention policies and to deliver actionable datasets.

### Palm Springs Power Baseball

Jun 2023 – Jul 2023

*Data Analytics Intern*

Palm Springs, CA

- Oversaw data integrity and validation across 100+ league games, maintaining the central repository to ensure accurate reporting for the official league website.
- Deployed an interactive Looker Studio dashboard to visualize 10+ KPIs from 7,400+ tracked pitch events, transforming raw data into actionable insights for 11 coaches and 260+ players.

## PROJECTS

### Quantifying Defensive Aggression With a Bayesian Hierarchical Model

Jun 2025 – Aug 2025

- Formulated a defensive baseball metric using a Bayesian hierarchical framework to estimate latent *aggression* rates, leveraging partial pooling to stabilize posteriors for ~90% of outfielders with limited observations.
- Trained a Logistic 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*, utilizing 700,000+ MLB Statcast records, to identify and analyze relationships between deviations in swing mechanics and offensive outcomes 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.