Diego Osborn

🌐 [portfolio/diego-osborn](#) ⦿ [github/d2osborn](#) 💼 [linkedin/diego-osborn](#) ✉ [diegoisaacosborn@gmail.com](#) 📞 [+1(928) 723-9215](#)

EDUCATION

University of California, San Diego — Sep 2022 - Jun 2026

Bachelor of Science - BS, Data Science; Minors in Economics and Mathematics

Organizations: Data Science Student Society, Triton Ball Sports Analytics Club
Honors: UCSD Eleanor Roosevelt College Honors Program, *Provost Honors* (4x)
Relevant Coursework: Stochastic and Probabilistic Modeling, Machine Learning, Data Management and Scalable Systems

EXPERIENCE

Baseball Analytics Intern | *UC San Diego Baseball* — Feb 2023 – Present

- Automated previously time-consuming opponent scouting report process, saving coaches *tens* of hours each season and improving game preparation.
- Designed an automated pitch classification pipeline in Python using *gradient boosted trees* and *expectation maximization* to identify and classify new pitches.
- Developed a Stuff+ framework to quantify pitch "*nastiness*" using *gradient boosted trees* on pitch-tracking features for use in scouting and player development.
- Built a *Monte Carlo simulation model* to forecast 2026 season wins under several scenarios to support potential scheduling decisions.
- Developed the team's first structured data storage pipeline by migrating raw tracking CSVs into *PostgreSQL* for reliable team SQL access.

Data Analytics Intern | *UC San Diego Career Center* — Sep 2024 – Jun 2025

- Performed EDA and data cleaning on career outcomes data for 12,000+ graduates to support the development of an interactive Power BI dashboard; *highlighting insights* on salary trends, industries, and top employers for student career planning.
- Developed a *PostgreSQL* database to store and query 100,000+ Handshake accounts; eased the archival of alumni accounts for improved data organization and historical tracking.

Baseball Analytics Intern | *Palm Springs Power Baseball* — Jun 2023 – Jul 2023

- Operated the league's pitch tracking system (Flightscope), and tracked and analyzed statistics for the PSCL's (Palm Springs Collegiate League) league website.
- Built an interactive Looker dashboard to provide a *user-friendly interface* for coaches and players to track their performances.

PROJECTS

A Bayesian Approach to Modeling Arm-Injury Risk in Pitchers (in development) — Nov 2025 – Jan 2026

- Developing a *Bayesian* model to predict season-ahead arm-injury risk for MLB pitchers.
- Constructed a longitudinal dataset of 6,300 pitcher-seasons by integrating large-scale baseball and injury data sources.
- Applying partial pooling in a *hierarchical framework* to improve predictions for pitchers with limited histories.

[Quantifying Defensive Aggression With a Bayesian Hierarchical Model](#) — Jun 2025 – Aug 2025

- Built a probabilistic defensive baseball metric, Aggression Over Expected (AOE), that quantifies the defensive aggressiveness of outfielders from spatiotemporal tracking data using *Bayesian hierarchical* modeling.
- Developed a catch-probability model using a logistic GAM, achieving 0.162 test log-loss and a 95.7% test F1 score.
- Communicated technical concepts to non-technical judges using clear visuals and intuitive explanations.
- Earned honorable mention recognition as a finalist in the 2025 Sports Media Technology (SMT) Data Challenge, ranking *top 10* among 50 teams (114 students).

[The Effect of Two Strikes on a Hitter's Swing](#) — May 2025 – Jun 2025

- Built an interactive web-based analysis exploring how swinging behavior of batters changes under specific conditions.
- Included league-level trends, a deep-dive case study, and a 3D visualization sandbox for exploratory analysis.

SKILLS

Programming & Scripting Languages: Python, SQL, R, JavaScript, HTML, CSS
Data/Platforms: PostgreSQL, SQLite, Apache Spark, AWS
BI/Visualization: Tableau, Power BI, Looker
Spoken Languages: English, Spanish (Native)