

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, Stochastic Processes, Relational Databases, Cloud Computing, Scalable ML

SKILLS

Python (PyTorch, Dask, scikit-learn, PyMC, Pandas, NumPy), R, SQL (PostgreSQL, BigQuery, SQLite), AWS, Google Cloud, 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.
- Engineered a *Python* data ingestion pipeline to migrate 1.5M+ rows (1.7GB) of raw TrackMan CSVs into *Google Cloud Storage* and *BigQuery*, establishing the team's first centralized cloud data warehouse.
- Authored technical SOPs to streamline database access and workspace setup, standardizing *BigQuery* workflows for a 10-person analytics team.

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 survey 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

- 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

Sean McVAI: Beating Two-High Coverage

Feb 2026

- Engineered a *Python* quantile regression pipeline using gradient boosted trees to forecast expected yardage across 5 distinct percentiles (2.5th to 97.5th), quantifying strategic risk and upside of play designs.
- Developed an interactive **Streamlit** dashboard integrating 2 predictive models, enabling end users to dynamically evaluate custom play designs.
- Presented to NFL analysts at the 2026 Football Analytics Blitz, competing against 15 teams (~70 students).

Quantifying Defensive Aggression With a Bayesian Hierarchical Model

Jun 2025 – Aug 2025

- Implemented a Bayesian hierarchical framework in *PyMC* to formulate a defensive baseball metric, estimating latent aggression rates and leveraging partial pooling to stabilize posteriors for ~90% of outfielders from spatiotemporal tracking data.
- 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

- Created an interactive website using *Python* and *D3.js*, utilizing 700,000+ MLB Statcast records, to analyze relationships between deviations in swing mechanics and offensive outcomes in high-pressure contexts.