



# Diego Osborn

 [github.com/d2osborn](https://github.com/d2osborn)  [linkedin.com/diego-osborn](https://linkedin.com/diego-osborn)  [diegoisaacosborn@gmail.com](mailto:diegoisaacosborn@gmail.com)  +1(928) 723-9215

## EXPERIENCE

---

- Data Analytics Intern** | *UC San Diego Career Center* Sep 2024 – Present
- Performed EDA and 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
  - Built a PostgreSQL database to store and query 100,000+ Handshake accounts; categorized alumni by graduation year and degree type, and facilitated the archival of alumni accounts for improved data organization and historical tracking
- Baseball Analytics Intern** | *UC San Diego Baseball* Feb 2023 – Present
- Automated 60+ opponent scouting reports to enhance game preparation, contributing to UC San Diego's 2023 NCAA Division-1 Baseball Big West Conference Championship
  - Developed a QDA-based model to automatically classify pitch types, improving manual tagging accuracy and correcting tracking inconsistencies in away games
  - Generated post-game pitching report graphics for our team's pitchers, summarizing each pitch type's metrics (e.g., velocity, movement profiles, whiff%) and visualizing pitch location and break patterns to support coaches in player development
- Baseball Analytics Intern** | *Palm Springs Power Baseball* Jun 2023 – Jul 2023
- Developed an interactive [Looker dashboard](#) with visualizations to assist players and coaches in evaluating pitch effectiveness and batted ball metrics by analyzing live in-game pitch data from FlightScope

## PROJECTS

---

- Home Price Evaluator: Determining Market Value** | *Python, Pandas, scikit-learn, Matplotlib, Seaborn* Feb 2025–Mar 2025
- Collaborated to develop and optimize a LightGBM regression model on 2.2M+ real estate listings, integrating geospatial, demographic, and economic data to enhance price predictions
  - Engineered zip code-level features by calculating proximity to universities using the haversine formula and applying geodemographic segmentation to classify homes into home type (e.g., ranch, luxury home, tiny home, etc.)
  - Created a valuation framework comparing listed vs. predicted prices, identifying 10.5% of properties as undervalued, enabling investors to target high-potential opportunities
- Spotify Statistical Analysis** | *Python, SciPy, Statsmodels, scikit-learn, Matplotlib, Seaborn* Feb 2025–Mar 2025
- Collaborated to analyze 100,000+ Spotify tracks with hypothesis testing and PCA to identify key audio features influencing song popularity, helping artists and streaming platforms optimize advertising strategies
  - Built a logistic regression model to determine the factors that differentiate "popular" from "not popular" songs, with an AUC score of 0.6226—slightly better than random guessing
- Power Outages Statistical Analysis** | *Python, Pandas, scikit-learn, Plotly* Jun 2024
- Analyzed U.S. power outage data (2000–2016) to identify key factors affecting major outages; built a RandomForestRegressor model for energy consumption to help power companies improve outage response
  - Engineered features related to seasonality, climate, economic indicators, and region; conducted hypothesis testing and missingness analysis; implemented a machine learning pipeline for regression modeling

## SKILLS

---

**Languages:** Spanish (fluent)  
**Programming Languages:** Python, R, Java, HTML5,  $\text{\LaTeX}$   
**Tools/Databases:** SQL, PostgreSQL, SQLite, Tableau, Power BI, Looker  
**Libraries:** Pandas, NumPy, Matplotlib, scikit-learn, Seaborn, Plotly, SciPy, Psycopg2, Statsmodels

## EDUCATION

---

**University of California, San Diego** Sep 2022 - Jun 2026  
*B.S. Data Science; Minors: Economics, Mathematics*

**Relevant Coursework:** Probabilistic Modeling and ML, Relational Databases, Statistical Methods, Probability Theory, Exploratory Data Analysis & Statistical Inference, Representation Learning

**Honors:** UCSD Eleanor Roosevelt College Honors Program, Provost Honors (2x)