





Diego Osborn

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

EXPERIENCE

- Data Analytics Intern** | *UC San Diego Career Center* Sep 2024 – Present
- Developed an interactive Power BI dashboard analyzing career outcomes for 12,000+ graduates, providing insights on salary trends, industries, and top employers to guide student career planning
 - Built a PostgreSQL database to store and query 100,000+ Handshake accounts, creating structured views to categorize alumni by graduation year and degree type, facilitating the archival of alumni accounts for improved data organization and historical tracking
- Baseball Analytics Intern** | *UC San Diego Baseball* Feb 2023 – Present
- Automated 40+ opponent scouting reports and conducted ad-hoc analytical projects to enhance game preparation, contributing to UC San Diego's 2023 NCAA Division-1 Baseball Big West Conference Championship
 - Generated post-game pitching report graphics, summarizing pitch metrics (e.g., velocity, movement profiles) and visualizing pitch location and break patterns to support coaches in post-game analysis and 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 with a team to develop and optimize a **LightGBM** model on **2.2M+ real estate records**, incorporating geospatial and economic data to improve home price estimation
 - Merged real estate data with **six external datasets**, including census demographics, crime rates, university proximity, and property taxes, to enhance market context
 - Engineered **zip code-level features**, applying the haversine formula for university proximity and clustering market segments to improve valuation insights
 - Built a valuation framework, based on expected price and actual price, that identified **10.5% of homes** as potential investment opportunities, aiding market analysis
- Spotify Statistical Analysis** | *Python, SciPy, Statsmodels, scikit-learn, Matplotlib, Seaborn* Feb 2025-Mar 2025
- Collaborated with a team to analyze **100,000+ Spotify tracks** to identify key audio features influencing song popularity, helping artists and streaming platforms optimize advertising strategies
 - Conducted **hypothesis testing** on explicit lyrics and genre impact, while applying **PCA** to identify underlying relationships between features
 - Developed predictive models (**linear regression, logistic regression, ridge regression**) to classify songs as "popular" or "not popular."
- 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 power outages and built a predictive model for energy consumption, helping 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
 - Optimized a **RandomForestRegressor** model, achieving a **test R^2 of 0.99974** and a **relative test error of 8%**—significantly outperforming the baseline linear regression model

SKILLS

Languages: Spanish (fluent)

Programming Languages: Python, R, Java, HTML5, \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 Machine Learning, Data Management, Statistical Methods, Probability Theory, Exploratory Data Analysis & Statistical Inference

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