

Sebastian Ferragut

Data Scientist · Data Analyst · ML Engineer

ferragut.sebastian@gmail.com · (909) 343-3289 · [LinkedIn](#) · [GitHub](#) · [Portfolio](#)

PROFESSIONAL SUMMARY

Data Scientist and upcoming UCSD graduate (B.S. Data Science, GPA 3.73) with hands-on experience delivering real-world AI solutions for professional clients, building end-to-end analytical pipelines, and designing interactive data visualizations. Proficient in Python, SQL, R, machine learning, and statistical modeling. Proven ability to work across the full data lifecycle — from data cleaning and EDA through modeling, deployment, and stakeholder communication. Fluent in English and Spanish.

EXPERIENCE

Freelance Data Scientist — Dr. Juan Madrazo —

Fall 2025 – Spring 2026

Crystal View Therapy

- Contracted to design and build a local-first AI documentation system to automate clinical note generation for a licensed psychologist's practice.
- Developed a full-stack application using Python, Chainlit, and React with custom audio capture, transcript editing, and document review components.
- Implemented on-device speech-to-text transcription and a local LLM inference pipeline, ensuring all patient data remained fully private and HIPAA-aligned.
- Delivered structured document generation workflows for progress notes, intake assessments, and treatment summaries — reducing manual documentation time per session.

Social Media Editor Board — Cuban Americans for

February 2026 – Present

Cuba (CA4C)

- Produce and publish social media content for a nonprofit advocacy organization focused on Cuba, contributing to platform growth and community engagement.
- Track and analyze social media performance metrics — including reach, impressions, and follower growth — to inform content strategy and identify high-performing post formats.
- Contributed to 3,000+ follower growth and 500,000+ total impressions within 2 months by aligning content production with data-driven engagement insights.

EDUCATION

University of California, San Diego

Expected: June 2026

Bachelor of Science — Data Science | GPA: 3.73

Relevant Coursework: Data Science in Practice · Statistical Methods · Data Visualization · Probabilistic Modeling & ML · Recommender Systems & Web Mining · Text as Data · Data Management · Systems for Scalable Analytics · Data Analysis and Inference · Representation Learning

Honors: Provost Honors — Fall 2023, Winter & Spring 2024, Spring 2025, Fall 2025, Winter 2026

PROJECTS

Agentic Privacy Control Center | Python · Data Analysis · Privacy · Agentic Workflows — Senior Capstone

- Led a team building a capstone project exploring privacy-aware analytical workflows through an agentic system design.
- Architected the end-to-end system design, balancing analytical capability with strict data privacy constraints.
- Produced a public artifact site documenting project structure, methodology, and deliverables for external audiences.
- Communicated technical research outcomes clearly to both technical reviewers and non-technical stakeholders.

NYC Citibike Product Analytics Pipeline | SQL · PostgreSQL · Python · DuckDB · Docker · Data Engineering

- Engineered an end-to-end local-first data pipeline ingesting 8.8 GB of official 2025 NYC Citibike trip archives — from raw CSV extraction through DuckDB normalization, Parquet staging, and PostgreSQL loading.
- Designed a production-grade relational schema (DDL, indexes, foreign keys) and authored analytics SQL covering CTEs, window functions (LAG), date truncation, and percentile aggregations to answer core product questions.
- Built data quality validation checks against a curated schema and implemented rider-proxy metrics (DAU/WAU/MAU, churn proxies) for datasets lacking persistent user identifiers.
- Authored a reproducible Jupyter notebook with EXPLAIN ANALYZE examples and a terminal reporting script — demonstrating the full analytics engineering workflow from ingestion to insight.

Interactive Mice Dataset Visualization | [D3.js](#) · [JavaScript](#) · [HTML](#) · [CSS](#) · [Data Visualization](#)

- Built interactive D3.js visualizations to explore biological protein expression data from a scientific mouse study dataset.
- Implemented dynamic filtering, comparative views, and exploratory controls across multiple experimental conditions.
- Designed a narrative-driven interface that guides users through dataset patterns — making complex biological data accessible.
- Deployed as a public web application via GitHub Pages.

Income vs. Carbon Expenditures in San Diego | [Python](#) · [Pandas](#) · [Statistical Analysis](#) · [Data Visualization](#) · [Jupyter](#)

- Collaborated in a team to execute a complete data science workflow from raw data ingestion through statistical analysis and reporting.
- Performed data cleaning, exploratory data analysis, and hypothesis testing to evaluate the relationship between income levels and carbon expenditure patterns.
- Applied statistical techniques including correlation analysis and significance testing; produced a fully reproducible Jupyter notebook with a final written report.

TECHNICAL SKILLS

Programming Languages

Python	R	SQL	JavaScript	Java
--------	---	-----	------------	------

Data Science & Machine Learning

Pandas	NumPy	Scikit-learn	Hypothesis Testing	PyTorch/TensorFlow
--------	-------	--------------	--------------------	--------------------

Statistical Analysis & Modeling

Statistical Modeling	Model Evaluation	Feature Engineering	Data Validation	Pipeline Design
----------------------	------------------	---------------------	-----------------	-----------------

Visualization & Reporting

D3.js	Matplotlib	Seaborn	Tableau	Jupyter Notebook
-------	------------	---------	---------	------------------

Infrastructure & Frameworks

Git / GitHub	Docker	AWS EC2	Streamlit	Chainlit
--------------	--------	---------	-----------	----------

Databases & Other Tools

PostgreSQL	DuckDB	SQLite	HTML / CSS	Local LLM / AI
------------	--------	--------	------------	----------------

Languages: English (Fluent) · Spanish (Fluent)