Hayley Baek

hayleybaek@ucsd.edu
(858) 248-3908
San Diego, CA

[GitHub](#)
[LinkedIn](#)
[Portfolio](#)

Education

MAS Data Science and Engineering, UCSD — Fall 2024 - June 2026
- In Progress: Data Science Capstone.
- Completed: Machine Learning, Database Management Systems, Data Integration & ETL, Scalable Data Analysis, Probability and Statistics, Data Visualization.

B.S. Mathematics, Computer Science Minor, NYU — Fall 2015 - May 2019
- Mathematics coursework: Data Analysis, Discrete Mathematics, Financial Mathematics, Numerical Analysis.
- Computer Science coursework: Databases, Design & Analysis of Algorithms, Data Structures, Object-Oriented Programming.

Experience

Data Analyst Intern, American Institute for Behavioral Research and Technology — October 2025 – February 2026
- Validated results for acceptance within the scientific community of AIBRT's Universal Website Rater tool, a machine learning model that evaluates the political biases of any news article text.
- Collaborated within a multidisciplinary research group and reported directly to the founder and senior research directors.
- Evaluated political biases of around 250 news articles across 30 news sources.

Data Scientist Lead Intern, IDX Exchange — September 2025 – December 2025
- Managed intern team of four people for project direction, code assistance and presentation creation.
- Developed predictive models and machine learning algorithms to forecast home property prices and other key metrics.
- Presented findings and insights to internal teams and clients in a clear and compelling manner.
- Trained machine learning model on around 500k data points.

Data Analyst, County of San Diego — April 2021 – August 2023
- Upgraded clinical reports and databases to a new system platform to support the inpatient hospital's operational reporting workflow and decrease patient intake waiting times for the entire San Diego behavioral health unit.
- Created visualizations that communicate findings to capitalize on physicians' reports metadata collected using BI software and tools: SQL, Tableau, and Business Objects.
- Developed data sets and structures following HIPAA compliance specific to behavioral health care.
- Maintained ongoing cross-platform data extract uploads and ensured data integrity for up to five active reports using SQL, Snowflake, and Oracle Cerner's cloud platform.

Recent Projects

Opioid Knowledge Graph | Jupyter Notebook
- Built an end-to-end ETL pipeline that translates documents into a knowledge graph to connect the main actors and events involved in the opioid epidemic.
- Engineered prompts for two LLMs to structure and enrich raw data.
- Psycopg, OpenAI, spaCy, Qwen, Neo4j.

Amazon Product Reviews Analysis | Jupyter Notebook
- Applied machine learning algorithms like regression and classification to predict successful ratings for Amazon products.
- Handled large-scale unbalanced data, engineered features for fitting machine learning models, evaluated performance metrics.
- Correlated price and rating for trend discovery in predicting popular products and ratings to boost future sales.
- Spark, Pandas, SciPy, MatPlotLib.

Skills

Tools: Python, SQL, Excel, Tableau, Oracle Cerner EHR software, Snowflake.

Approaches: visualizations, storytelling with data, statistical analysis, logical problem-solving.

Additional: UCSD Women in Computing, Academic Integrity Review Board member, HAM radio club.