

Sarah He

(626) 233-1717 | sarahhe725@gmail.com | sarahhe05.github.io/sarah-he/

EDUCATION

University of California, San Diego
Bachelor of Science in Data Science

Expected Graduation: June 2026
Provost Honors, GPA: 3.90/4.00

EXPERIENCE

Data Analyst Intern

June 2025 - August 2025

Capital One

McLean, VA

- Designed and implemented SQL-based compliance controls, automating previously manual audit procedures and integrating them into a centralized platform
- Diagnosed and resolved issues in legacy compliance logic, removing 85% of false positives and streamlining data validation
- Consolidated fragmented datasets into a centralized data mart to support transparent, consistent reporting and simplified downstream analysis
- Collaborated with cross-functional partners to align control outputs with business logic, ensuring accurate reflection of policy requirements and compliance goals

Data Science Intern

August 2024 - February 2025

A Round Entertainment

Jersey City, NJ

- Developed and optimized predictive models (CatBoost, LightGBM and Ensemble methods) to estimate concert revenue, capturing over 90% of the patterns in the data and uncovering key factors influencing revenue generation
- Developed interactive Tableau dashboards that allowed non-technical stakeholders to explore data by artist and time period, effectively visualizing key metrics and trends
- Scraped and parsed HTML data using the BeautifulSoup library to build robust datasets with over 705,000 rows from online sources, enabling in-depth analysis of concert industry trends

Human Resources Intern

August 2024 - October 2024

A Round Entertainment

Jersey City, NJ

- Automated attendance tracking by setting up new employee attendance system in Excel, reducing manual effort and increasing efficiency in performance analysis
- Facilitated effective communication between employees and management

Starry Sticker Shop

June 2020 - September 2023

Co-Founder

Temple City, CA

- Managed product design, customer service, and financial planning, growing the Etsy business to 270+ global orders
- Analyzed and visualized sales patterns and customer sentiment for actionable insights using Python and Tableau

PROJECTS

Major Power Outage Analysis | *Python (Pandas, Scikit-learn, Matplotlib), HTML, CSS*

- Conducted in-depth analysis of power outage data to identify regional and infrastructural factors influencing outage severity
- Applied feature engineering and hyperparameter tuning to develop a Logistic Regression model that forecasts severe power outages with 85% accuracy
- Conducted permutation testing to assess missingness dependency, model fairness, and analyze key trends in the dataset

EcoMenu | *Python (Pandas, Matplotlib, Flask), Node.js, Tableau, HTML, CSS, JavaScript*

- Sourced, compiled, and cleaned datasets on the carbon footprint of over 350 food ingredients and 2100 recipes
- Developed web application enabling users to select ingredients and receive low-carbon recipe recommendations
- Presented visualizations to showcase insights from the recipe and ingredient data
- Boost Hacks II: Most Creative Hack, CirFin CREATE: Civic Achievement Award

CrimeMap SD | *JavaScript (Mapbox, Turf.js), Python (Pandas), HTML, CSS*

- Developed an interactive map dashboard intended for patrol vehicles, providing situational awareness by displaying real-time GPS data and historical crime records from the San Diego Police Department to track incidents and neighborhood trends
- Designed and implemented a route risk awareness feature to analyze historical crime density along selected paths using spatial buffers and geospatial querying
- Optimized for field use with a responsive UI, dark/light mode toggle, and touch-friendly controls; deployed via GitHub Pages for lightweight, device-agnostic access

SKILL SET

Technical: Python, SQL, Tableau, Java, R, HTML/CSS, JavaScript, Excel, Git, Snowflake, Jupyter Notebook, Figma, Flask

Production: Adobe Lightroom, Adobe Photoshop, CapCut, PowerPoint