Maggie Shao

408-431-5651 | yis082@ucsd.edu | [Linkedin](#) | [Github](#) | [Personal Website](#)

EDUCATION

University of California, San Diego   La Jolla, CA
B.S. in Data Science   *Sept 2024 - June 2028*
- Double Major: B.S. in Cognitive Science (ML Specialization); GPA: 3.81

SKILLS

Languages: Python, Java, SQL, R (basic), JavaScript (basic)
Data and Visualization: Pandas, Numpy, Matplotlib, Seaborn, Plotly, Tableau, Excel (advanced)
Soft Skills: Data Storytelling, Project Management, Team Collaboration, Problem-Solving, Communication
Certifications: Google Data Analytics Certificate, Codepath Techanical Interview Preparation Certificate

EXPERIENCE

Data Entry Specialist   Mar 2025 – May 2025
California Homebuilding Foundation   *Remote*
- Streamlined data entry, reducing errors and improving database accuracy.
- Maintained 100% compliance across 200+ monthly transaction records, improving quarterly reporting precision and reducing report review time by 15 hours.
- Analyzed data trends to identify and correct inconsistencies, strengthening data integrity.

Data Science Intern   Jan 2025 – April 2025
IDX Exchange LLC   *Remote*
- Built supervised regression models (Random Forest, Gradient Boosting) that boosted property price prediction accuracy by 12%.
- Cleaned and normalized 10k+ property datasets, and handled missing values via KNN imputation, standardized 5 key features, improving model training efficiency by 30%.
- Designed interactive market trend dashboards in Python (Plotly + Seaborn) that enabled stakeholders to explore regional pricing patterns and investment opportunities visually.

Data Science Intern   June 2024 – August 2024
Cisco Global Support Center Dalian   *Dalian, China*
- Developed and debugged Python automation scripts, reducing troubleshooting time by 25%.
- Resolved 50+ network connectivity issues and supported 30+ end users, improving response efficiency and service reliability.
- Transformed 7k+ system logs into Excel dashboards with trend charts and performance heatmaps, facilitating supervisor decision-making.

PROJECTS

California Real Estate Market Analysis | *Python, Pandas, Matplotlib, Seaborn*
- Analyzed 178,925 California real estate listings, performing data cleaning, one-hot encoding, and feature scaling to prepare structured data for price prediction.
- Built a Linear Regression model using scikit-learn to predict home closing prices, achieving $R^2 = 0.77$ and validating performance through train-test splitting (80/20).
- Conducted model diagnostics and visualization, including actual vs. predicted price and residual analysis, revealing strong overall trend capture with higher variance on high-priced properties.

Friends TV Show Dialogue Sentiment Visualization | *Python, Seaborn, Matplotlib*
- Analyzed data from 236 episodes across 10 seasons of the TV series Friends, integrating episode metadata, dialogue scripts, and emotion labels.
- Conducted exploratory data analysis and visualization on scripts from 30 full episodes, examining 7 emotion categories to identify sentiment distribution and narrative patterns across seasons.

LEADERSHIP AND AFFILIATIONS

- **Professional Event Team Board Member** — Data Science Student Society @ UC San Diego
- **Mentor** — Global Ambassador Program @ UC San Diego
- **Project Specialist** — Women in Computing @ UC San Diego
- **Community Member** — [CodePath](#)