Jason Gu

jaesongu@gmail.com • (626) 277-8952 • LinkedIn • Website • Github

EDUCATION

University of California, San Diego | B.S. Data Science — Expected Graduation: Mar 2025

- Practice of Data Science, Probabilistic Modeling & Machine Learning, Statistical Methods, Data Management, Data Analysis & Inference

Certifications: Google Data Analytics (*Coursera*)

WORK EXPERIENCE

Data Science Fellow | San Diego Gas & Electric - *San Diego, CA* — Sep 2024 - Present

- Conducted time series analysis and geospatial analysis using **Python** to evaluate EV charger density and growth across 1100+ chargers in San Diego, identifying underserved areas to initiate infrastructure development using **Pandas**, **Scikit-Learn**, **GeoPandas**, and **Folium** libraries**.**
- Projected a 57% increase in EV adoption in 2024 through regression modeling on DMV vehicle registration data using **Python** and **SQL**.
- Applied statistical modeling with **Statsmodels** to evaluate the correlation between EV ownership and charger availability in San Diego.
- Authored a findings report on EV adoption trends and charger optimization opportunities and presented it to the Director of Data Science.

Data Science Intern | Mercury Alert AI - *San Diego, CA* — Jun 2023 - Oct 2023

- Independently drove the development of an internal quality assurance dashboard using **Python**, designed to monitor over 50+ devices and provide real-time reports on temperature, empty frames, and device capture errors.
- Performed anomaly detection using **AWS QuickSight** to identify and analyze time-stamped image captures with low confidence scores.
- Reviewed and updated the **Jupyter Notebook** data management system by relabeling mispredictions and annotating low confidence score images identified through time-stamped analysis, improving the retraining efficiency of AWS Lambda by 30%.

Data Analyst Intern | Redrock Biometrics - *San Francisco, CA* — Jun 2022 - Sep 2022

- Implemented a custom image processing pipeline using **OpenCV** and **NumPy** Python libraries to efficiently load, preprocess, and analyze 1,200 palm print images, enabling accurate edge detection and feature extraction for biometric analysis.
- Optimized palm print recognition accuracy by identifying the ideal Top-N predictions, reducing the False Rejection Rate (FRR) by 63.6%.
- Showcased results of analysis using **Python**, **SQL**, and **Tableau** to influence software engineers' decisions on the ideal Top-N prediction.

PROJECTS AND LEADERSHIP EXPERIENCE

Recipes Protein Analysis and Predictor | Data Science / Machine Learning Project (https://jingchenggu.github.io/Protein_Prediction/)

- Built a machine-learning model to predict protein content in recipes using **Scikit-Learn**, **Random Forest Regressor**, and **GridSearchCV**.
- Performed exploratory data analysis, feature engineering, model training, and hyperparameter tuning to improve the original model, achieving an R^2 of 0.92 and reducing RMSE from 24.7 to 8.4.
- Conducted fairness analysis using hypothesis and permutation tests, revealing significant RMSE disparities between meat and non-meat recipes.

Academic Performance Analysis at UCSD | Data Science Research Paper (github link)

- Performed a data-driven analysis of student grades at UCSD to identify factors influencing academic performance, utilizing multiple regression models in **Scikit-Learn** to quantify the impact of predictors such as class size and departments, achieving an R-squared value of 0.702.
- Enhanced prediction model performance through feature engineering, creating 5 new variables and converting 3 categorical variables for effective inclusion in regression models using **Python**, resulting in improved interpretability and accuracy.

Data Science Student Society |VP Internal (https://www.ds3ucsd.com/)

- Spearheaded a dynamic interdisciplinary group of over 300 undergraduate data science students, conducting weekly meetings for all members, ensuring cohesive management and effective implementation of the organization's diverse initiatives.
- Initiated a mentorship program, pairing 30+ junior members with senior mentors, boosting member engagement by 50%.

TECHNICAL SKILLS

Languages: Python, Java, JavaScript, HTML, SQL, R

Frameworks / Libraries: Pandas, NumPy, PyTorch, TensorFlow, OSMnx, Cenpy, GeoPandas, OpenCV, JUnit, Beautiful Soup, Scikit-Learn

Tools: Amazon Web Services (Lambda, DynamoDB, QuickSight), Github, Excel, Tableau, Power BI, Docker, IntelliJ, D3.js, Svelte