Jason Gu

jaesongu@gmail.com | (626) 277-8952 | U.S. Citizen | [LinkedIn](#) | [Website](#) | [Github](#)

TECHNICAL SKILLS

Languages: *SQL, Python, Java, HTML, CSS*
Frameworks / Libraries: *Pandas, NumPy, PyTorch, HuggingFace, TensorFlow, OSMnx, XGBoost, GeoPandas, OpenCV, Scikit-Learn, Plotly*
Tools: *Amazon Web Services (S3, Lambda, DynamoDB, QuickSight), Git, Excel, Tableau, Power BI, Docker, Apache Airflow, Apache Spark*
Others: *Data Analytics, Data Manipulation, Machine Learning, Data Extraction, Data Visualization, Statistical Analysis, Query Optimization*

EXPERIENCE

Data Science Fellow | San Diego Gas & Electric | *San Diego, CA* — Sep 2024 - Mar 2025

- Analyzed EV charger density and growth across 1100+ chargers in San Diego using **Python**, identifying underserved areas using **Pandas**, **Scikit-Learn**, **GeoPandas**, and **Folium**, uncovering 15+ underserved zip codes and informing data-driven city planning decisions.
- Forecasted a 57% rise in EV adoption in 2025 by leveraging statistical modeling, advanced **SQL** queries and **Python** on DMV data, enabling proactive infrastructure scaling and strategic investment decisions.
- Authored weekly **Tableau** and **PowerPoint** reports on EV trends and actionable insights to influence stakeholders and the Director of Data.

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

- Developed a comprehensive internal quality assurance dashboard in **Python** to support product supply chain, automating an **ETL** pipeline to transform raw device data into real-time insights across 50+ devices, reducing manual error checks by 40%.
- Performed anomaly detection in **AWS QuickSight** to flag low-confidence mispredicted images, improving model retraining effectiveness.
- Reviewed and updated the **Jupyter Notebook** data management dashboard 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** on biometric data (palm print images) to efficiently load, preprocess, and analyze 1,200 samples, enabling accurate edge detection and feature extraction for biometric analysis.
- Improved recognition accuracy by optimizing Top-N predictions, reducing False Negative Rate (FNR) by 63.6% based on biometric **KPIs**.
- Performed business strategy support using **Python**, **SQL**, and **Tableau** to influence software engineers' decisions on the ideal Top-N value.

PROJECTS

[Real-Time Stock News Sentiment Dashboard Pipeline](#)

- Built a daily **batch ETL pipeline** to automate news and price API pulls for 10 popular tech stocks using **Airflow**, **Docker**, and **PostgreSQL**.
- Applied FinBERT-based **NLP** sentiment scoring with **PyTorch** to analyze financial news headlines and descriptions, generating daily sentiment scores from -1 to 1 to quantify stock-specific news sentiment.
- Designed a dynamic **Tableau** dashboard to compare sentiment trends with stock price movements, supporting exploratory financial analysis.

[EV Charger Fault Reporting Application](#)

- Built a proof-of-concept mobile app for EV owners to report charger faults, providing data driven product solutions for the EV communities.
- Curated a two-step **Python** pipeline using NVIDIA's **mit-b3** for semantic segmentation of EV charger components, followed by a **binary classifier** to determine component health; hosted all machine learning models via **Hugging Face API**.
- Presented the product to 200+ SDG&E employees, showcasing the pipeline's impact on fault detection and repair service improvement.

[Recipes Ratings Analysis and Predictor](#)

- Built a machine-learning model to predict protein content in 234,429 different recipes from Food.com using **Scikit-Learn**, **Random Forest Regressor**, and **GridSearchCV**.
- Performed **exploratory data analysis**, **A/B testing**, feature engineering, and hyperparameter tuning to improve the original model.
- Conducted fairness analysis using hypothesis and permutation tests, revealing RMSE disparities between meat and non-meat recipes.

EDUCATION

University of California, San Diego | B.S. Data Science

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

Certification: [Google Data Analytics](#) (*Coursera*)