

# Kyle Shiroma

kyleshiroma77@gmail.com • 909-378-0684 • [GitHub](#) • [LinkedIn](#)

---

## Education

University of California, San Diego – B.S. in Data Science  
Norco College – Associate of Science for Transfer (Mathematics)

Expected June 2027  
August 2022 – June 2025

## Certifications & Ongoing Learning

- Apache Airflow: The Hands-On Guide (Udemy) – Currently building scalable data pipelines using Airflow DAGs, Docker, and storing data in AWS S3.

## Technical Skills

**Programming:** Python, SQL, C++

**Data Analysis & ML:** Pandas, NumPy, Scikit-learn, XGBoost, imbalanced-learn

**Databases:** PostgreSQL, MySQL, DBEaver

**Visualization:** Tableau, R Shiny, Matplotlib, Seaborn

**Tools & Cloud:** Git, Docker, Jupyter, VS Code, Astro CLI, AWS (basic), AWS (basic) including AWS Glue Crawler, Apache Airflow (in progress), Flask (in progress)

## Projects

**Pulsepanion – AI4Purpose NYC Hackathon Winner (Team Lead)** July 2025 – July 2025

- Led product direction and team coordination for a **patient-facing health summary tool**.
- Designed and implemented backend Python scripts to **de-identify sensitive medical data**.
- Integrated **OpenAI's LLM API** to generate clear, patient-readable summaries from visit transcripts.
- Built a **PDF export pipeline** using rmarkdown for formatted report generation.

**Customer Segmentation with RFM Analysis**

April 2025 – May 2025

- Developed Tableau dashboards to visualize customer behavior segments from **500K+** retail transactions, uncovering that **<20% of customers drove >70% of revenue**.
- Analyzed seasonal sales trends and customer engagement patterns to inform targeted marketing and promotional strategies.

## Experience

**Norco College Learning Resources Center**

Jan. 2023 – June 2025

Peer Tutor

- Tutored **100+** students in Calculus, Statistics, and C++ programming.
- Maintained accurate documentation and timesheets using **Microsoft Word, Excel, and Outlook**.
- Strengthened **communication** skills by delivering personalized academic support.

**CIC-Summer Research Program, California State University, Fullerton**

May 2024 – July 2024

Data Science Research Intern

- Cleaned and structured a dataset of 5,903 European football matches for predictive modeling.
- Engineered Elo-based team strength ratings, boosting model performance by 10%.
- Built and evaluated **Logistic Regression** and **XGBoost** models, achieving 61% accuracy in multi-class outcome prediction.
- Enhanced recall for the minority class ("Draw") by 20% through advanced resampling techniques (e.g., SMOTE, undersampling).