

# Lukas Amare

Email: [lukasamaree@gmail.com](mailto:lukasamaree@gmail.com)  
Linkedin: <https://www.linkedin.com/in/lukas-amare/>

---

## EDUCATION

### **Masters of Data Science**

University of San Francisco

**July 2024 - August 2025**

San Francisco, CA

Relevant Coursework: Data Acquisition, Statistics, SQL, Data Analysis, Pytorch, Pandas, Machine Learning

### **Bachelor of Science in Biological Science**

University of California, Irvine

**August 2020 - June 2024**

Irvine, CA

Relevant Coursework: Cell Biology, Human Physiology, RNA Sequencing

## WORK EXPERIENCE

### **UCSF Radiology/Oncology Cancer Research**

**September 2024 - June 2025**

*Data Scientist*

*San Francisco, CA*

- Fine-tuned a BERT model to predict hospital readmissions based on doctor notes, achieving a ROC AUC score of 0.71
- Created and utilized an LLM pipeline for literature mining to expand PhosphoAtlas database, integrating diverse sources to simulate protein-protein interactions using graph neural networks (GNNs).
- Identified therapeutically actionable kinase networks, contributing to the development of targeted cancer therapies.

### **Jiffy Lube**

**February 2024 - June 2025**

*Data Scientist*

*Remote*

- Conducted demographic and socioeconomic analysis of Southern California to identify optimal locations for Jiffy Lube expansion, refining the focus from a 100-mile radius to five high-potential zip codes.
- Utilized AI-driven agents to search for open leases and properties for sale, streamlining site selection for new locations.

### **Lung Cancer RNA Sequencing Analyst**

**April 2024 - May 2024**

*University of California, Irvine*

*Irvine, CA*

- Using violin plots, PCA, and k-mean clustering I cleaned and visualized thousands of data points into organized clusters in R.
- Discovered 7 different cells associated with lung carcinoma differentiated through k-means clustering.
- Discovered which lung cancer differentiated cells communicate with one another through cell signaling

## PROJECT EXPERIENCE

### **Due Diligence App**

**October 2024**

*Personal*

*San Francisco, CA*

- Developed an ETL pipeline to identify undervalued stocks based on key financial metrics, including SIR and insider transactions
- Automated data extraction from 100+ financial webpages to enhance stock summaries and investment insights.
- Designed and implemented a Retrieval-Augmented Generation (RAG) pipeline and leveraged prompt engineering with LLMs

### **Football SQL Database Project**

**October 2024**

University of San Francisco

*San Francisco, CA*

- Designed a schema for multiple tables containing thousands of rows of world wide football (soccer) data.
- Queried the database to identify the most aggressive leagues and highest-scoring leagues using CTEs and joins.
- Deployed the database on Snowflake and retrieved queried data for further exploratory data analysis in Jupyter Notebook.

## LEADERSHIP EXPERIENCE

### **UC Irvine Cross Country and Track and Field (Division 1 Athlete)**

**August 2020 - May 2024**

*Team Leader*

*Irvine, CA*

- Achieved top finishes, earning a spot on the All Big West Conference First Team in Cross Country for two straight years.
- Finished top on the team, resulting in third-place rankings in school history for both the 10 km and 8 km races in Cross Country

## SKILLS

- Python, Pandas, Numpy, R, Web Scraping, Data Acquisition, Statistics, SQL, Data Analysis Machine Learning, Fine Tuning