

Aidan Nguyen

(408) 345-1470 | aidannguyen27@ucla.edu | [linkedin.com/in/nguyenaidan](https://www.linkedin.com/in/nguyenaidan) | aidannguyen.netlify.app

EDUCATION

University of California, Los Angeles (UCLA)

B.S. Statistics and Data Science, Minor in Data Engineering

Expected June 2026

GPA: 3.98

Relevant Coursework: Programming with Python, Probability, Data Analysis and Regression

Statistical Programming with R, Discrete Mathematics, Linear Algebra, Data Structures, Mathematical Statistics

EXPERIENCE

AI Engineering Intern

September 2025 – Present

Corsair

- Built a local enterprise LLM chatbot with Retrieval-Augmented Generation (RAG) for instant knowledge access
- Indexed and pulled data from Atlassian services and SharePoint, processing **100TB+** of documents, tables, & slides
- Engineered **RAG** pipelines with embedding models and **Qdrant** vector databases for scalable knowledge access
- Improved internal data retrieval accuracy by **95%** with chunk reranking, source prioritization, and metadata filtering
- Reduced manual document searches from **~30 minutes** to **3-5 seconds** through deployment of enterprise LLM

Machine Learning Research Assistant

January 2024 – September 2025

Signaling Systems Lab

- Secured **\$75,000** in grant funding by submitting impactful data analyses from ongoing research initiatives and findings
- Analyzed **6M rows** of Single Cell RNA sequencing data with **R** (caret, Seurat) to classify bone marrow stroma cells
- Deployed supervised learning models, including **SVMs** and coefficient classifiers, to evaluate data quality & scalability
- Optimized hyperparameters and feature selection improving model accuracy, achieving **96.9%** classification precision

Data Science Intern

June 2024 – September 2024

Dune Road Capital

- Enhanced financial analysis using **SQL** queries for **10M+ rows** of exchange-traded fund (ETF) performance metrics
- Predicted returns for **10+** ETFs through Python time-series analysis, aiding portfolio strategy recommendations
- Produced data visualizations with **Tableau** to display key ETF metrics to prospective investors and team members

PROJECTS

CourseKata Feature Enhancement @ UCLA DataFest 2024 | R, dplyr, ggplot2, caret, tidyverse

- Achieved finalist out of **400+** participants in a 48-hour data hackathon, focused on investigating complex datasets
- Analyzed **7** spreadsheets with **100K+ rows** of user and site interaction data to uncover user engagement trends
- Identified key patterns in **1000+** users' behavior utilizing **R** and tidyverse to clean, transform, and merge datasets
- Presented visualizations recommending data-driven improvements for greater engagement & educational outcomes

AWS Covid Analysis | Amazon RDS, S3, EC2, QuickSight, MySQL

- Designed a scalable data pipeline to process COVID-19 data using **Amazon RDS & S3** for data storage & retrieval
- Applied **EC2** for running **MySQL** queries, data processing, and transforming large datasets into manageable formats
- Connected **AWS RDS** to **MySQL Workbench** for database management and visualization of real-time data
- Developed an interactive dashboard in **AWS QuickSight** to visualize key pandemic metrics across regions

YouTube Success Metrics Analysis | Python (PyTorch), transformers, pandas, numpy, googleapiclient, nltk, matplotlib

- Parsed 600,000+ YouTube comments using a fine-tuned **DistilBERT** model for multilingual sentiment analysis
- Created predictive models utilizing **OpenCV** face detection to assess how **1000+** titles & thumbnails impact views

SKILLS

Programming Languages: Python (pandas, numpy, scikit-learn, plotly, flask, tensorflow), SQL, R, Java, C++, JavaScript,

Tools: Git, Tableau, Power BI, Excel, AWS

Interests: Volleyball, Rock Climbing, Graphic Design, Catan, Cooking