

# Aidan Nguyen

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

## EDUCATION

### University of California, Los Angeles (UCLA)

Expected June 2026

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

GPA: 3.97

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

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

## EXPERIENCE

### Machine Learning Research Assistant

January 2024 – Present

*Signaling Systems Lab*

- Secured grant funding by visualizing cluster **t-SNE plots** that verified data from ongoing research initiatives
- 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 **95.8%** classification precision

### Data Analyst Intern

September 2024 – December 2024

*Improper Etiquette*

- Modeled polyester structure cost using market research, public data, & regression for **50%** profit margin targets
- Evaluated **15+** production variables to identify break-even points and optimal sales targets for hats in Los Angeles
- Created data visualizations with **Python** and detailed reports with **10+** research papers sourced to support findings
- Presented regression-based insights on cost modeling and profit optimization to guide strategic client decisions

### 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 a gradient boosting regressor and predictive models to assess how **1000+** titles & thumbnails impact views
- Implemented **OpenCV** for face detection in video thumbnails, correlating findings with viewership metrics

## 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