

ETHAN ALLAVARPU

San Francisco Bay Area, CA · eallavar@stanford.edu

www.linkedin.com/in/ethanallavarpu · github.com/ethan-allavarpu · eallavar.web.stanford.edu

education

Stanford University · M.S. Statistics Data Science · March 2024

- Coursework: Artificial Intelligence / Natural Language Processing / Machine Learning / Deep Learning / Computer Vision / Reinforcement Learning / Time Series / Bootstrapping / Big Data

University of California, Los Angeles (UCLA) · B.S. Statistics · June 2022

- Graduated *summa cum laude*
- Completed four-year degree in three years
- Coursework: Linear Models / Data Analysis and Regression / Experimental Design / Machine Learning / Statistical Models and Data Mining / Computational Statistics with R / Computation and Optimization for Statistics / Monte Carlo Methods

skills

- **Python**: scikit-learn / nltk / fuzzywuzzy / keras / TensorFlow / PyTorch / selenium / pandas / numpy / scipy / statsmodels / matplotlib / seaborn
- **R**: tidyverse (dplyr / ggplot / readr) / lubridate / RSelenium / RShiny / ggmap
- **SQL** · **Snowflake** · **AWS** · **Tableau** · **Microsoft Excel**
- **Machine Learning** · **Statistical Analysis** · **Data Analysis and Visualization** · **Natural Language Processing (NLP)** · **Web Scraping (Selenium)**
- **Spanish** (proficient)

relevant experience

Incoming Data Science and Visualization Intern · Apple · June 2023

- Data cleaning, processing, and transformation for **statistical modeling** and **visualizations** in **Python** and **R**
- Provide product and **business insights** for use by various teams
- Member of the DataViz team within Hardware Engineering

Graduate Research Assistant · Stanford University · January 2023 - Present

- Assist randomized controlled trials for energy efficiency improvements and information on worker incentives for brick kiln owners across South Asia in **R**
- **Standardize**, process, and **QA** survey data from various questionnaires
- Perform **statistical analysis** and produce various **data visualizations**
- Member of the [Luby Lab](#) (under Dr. Stephen Luby)

Data Science Intern · Bridg · June 2022 – November 2022

- Constructed machine learning models with **AWS Sagemaker** (in **Python**) and **Snowflake (SQL)** to analyze **terabytes** of data (over **100 billion** observations)
- Developed **natural language processing (NLP)** and **machine learning** algorithms to categorize product descriptions for enhanced business analytics
- Unified and standardized attributes for over **300 million profiles** from multiple sources through **statistical analysis and imputation**
- Web scraped, visualized, and analyzed correlations between minimum wages and menu prices in various urban areas
- Assisted product recommendation and churn identification via machine learning models in **Sagemaker** and **Snowpark** (Snowflake Python API)

Data Analyst Intern · SCAN Health Plan · November 2021 – June 2022

- Gathered insights about SCAN membership experience challenges by analyzing Voice of the Consumer (VOC) data in **Python**, **R**, **Tableau**, **SQL**
- Analyzed member call data (over **1 million** calls) to create FAQ sections using **natural language processing (NLP)** and **unsupervised learning** (agglomerative clustering) in **Python** for benefit categories and disenrollment groups with visualizations in **R** and **Tableau**
- Queried disenrollment data from **SQL**, visualized in **Tableau** and **R**, analyzed demographic disenrollment rates via **chi-square**, **posthoc analyses**
- Presented findings, insights, and learnings to leadership teams and in all-teams meetings

Data Science Consultant · UCLA Library Data Science Center · November 2021 – June 2022

- Consulted to curate, transform, analyze, and visualize researcher data using **Python**, **R**, **Tableau**, **SPSS**
- Implemented and refined **machine learning models** including **K-Nearest Neighbors (KNN)**, **decision trees**, **neural networks**
- Performed statistical tests including **chi-square analyses**, **Peacock tests** (i.e., a multidimensional KS test)
- Member of the [DataSquad](#), a student-run team within the UCLA Library Data Science Center

Student Assistant · UCLA Luskin School of Public Affairs · June 2021 – March 2022

- Web scraped business-level data from over **30 million** firms with **Selenium Webdriver** in **R**
- Utilized the **Google Translate API** to convert job titles from 5+ languages to English for imputation

InStep Intern (Data Analytics and AI) · Infosys · June 2021 – August 2021

- Defined and developed **machine learning models** to analyze and detect patterns of play and their success and failure percentages for Roland Garros (the French Open) and the Australian Open to improve user experience for media, players, and coaches
- Supported creation of **natural language generation** for **AI commentary** to provide engaging point-by-point description of Roland Garros and Australian Open matches using **Python** to improve the user interfaces for tennis fans

Bruin Sports Analytics

President · June 2021 – June 2022

- **Oversaw** general members and the data journalism, research, and consulting teams for Bruin Sports Analytics
- **Coordinated** meetings and deadlines, handled club finances, and scheduled speaker events to ensure timely deliverables and club success
- **Founded** the UCLA Baseball consulting division of the club consulting and **expanded** club membership from 60 to over 100 members

Blog Chair · June 2020 – June 2021

- **Mentored** the data journalism team in the development publication of data analysis articles: [“Data Journalism Guidelines 2020-2021”](#)
- Authored data analysis articles with **R** and **Python**: [“Project Restart: Winners and Losers of the EPL’s Return”](#)

Data Journalist · October 2019 – June 2020

- Published articles applying **statistical methods** and **machine learning (R, Excel)** to analyze and predict sports trends
- Authored three articles: [“The Evolution of the Free Throw,”](#) [“The Coin Toss,”](#) [“Fantasy Football and the NFL Combine”](#)

awards

- **Graduated *summa cum laude*** · UCLA · June 2022
- **Best Visualization** (Honorable Mention) · UCLA DataFest · April 2021
- **Employee of the Quarter** · Lifetime Activities · October 2020
- **Dean’s List** · UCLA · All eligible quarters (Fall 2019 – Spring 2022)
- **National AP Scholar** · July 2019
- **U.S. Presidential Scholars Program Candidate** · February 2019