# Aaron Mui

(415) 601-6122 | San Francisco, California 94132 | armui@uci.edu | LinkedIn: aaron-mui

## Education

**Master of Data Science**, *University of California, Irvine* **Expected Dec 2024**
**Bachelor of Science in Statistics**, *University of California, Davis* **Jun 2022**

**Course Work:** Mathematical Statistics, Probability Theory, Data and Web Technologies, Big Data and High-Performance Statistical Computing, Time Series Analysis, Multivariate Data Analysis, Regression Analysis, Statistical Learning, Bayesian Statistics

## Skills

**Languages:** Python, R, MySQL, MATLAB, LaTeX
**Tools:** Scikit-Learn, PyTorch, Pandas, NumPy, SciPy, MatPlotLib, Seaborn, Librosa, BeautifulSoup, Shiny

## Experience

**Data Analyst** **May 2022 — Present**
*Stanford University* *Remote*

- Analyze and create interactive reports about research finance in Oracle Business Intelligence to help accountants and other departments understand their award, property, and finance data better.
- Run SQL queries to give others a more in depth understanding of the data structure.
- Using Python, I streamlined and automated two work processes that involve joining, simplifying, and cleaning multiple Excel spreadsheets.
- Responsible for creating and reporting monthly sub award spending data to the federal government under the Federal Funding Accountability and Transparency Act (FFATA).

## Projects

**Audio ML - Music Genre Classification**

- Worked in a team of four with the Librosa Python package to develop Audio ML algorithms.
- Led Feature Extraction and Dimension Reduction on single and multi-dimensional features by aggregating across frames on a given audio sample.
- Used Python's Sci-kit Learn clustering methods, such as Agglomerative Hierarchical and Mean-Shift Clustering, to classify the genres of each audio sample and assess the model.

**What Makes an Asteroid Hazardous And How to Classify Them?**

- Deployed supervised and unsupervised learning algorithms in Python to determine an asteroids danger level.
- Implemented dimension Reduction techniques such as PCA and t-SNE to check if lower dimension clustering provided better results.
- Implemented a XGBoost model in Python which achieved the highest accuracy of 99.67%.
- Created a ShinyApp in RStudio that lets users plot features for easier readability.

**Is Baseball an Unfair Sport?**

- Led web scraping on multiple websites using Python's Beautiful Soup package.
- Led statistical analyses, and used a Linear Regression model in Python to investigate big market and small market teams, batting average, and more baseball intricacies.
- Preformed case studies on the 2021 Atlanta Braves and the "Juiced Ball" theory to see if these events were unique in their respective years.

**Counter Strike Global Offensive Round Outcome Prediction**

- Implemented L1 Logistic Regression and Support Vector Classification via SGD algorithms as baseline models using Scikit-Learn in Python.
- Pruned tree models and iterated over ensemble tree models such as Random Forest to find the best and most efficient.
- Created a four layer neural network with a ReLU activation function and ADAM optimizer to compare to our baseline models.