

# Andres Gutierrez

axndy.g@gmail.com · meet-andy.me · LinkedIn

## EDUCATION

### University of California, Berkeley

B.A Data Science (Emphasis in Linguistic Sciences)

GPA: 3.96

May 2026

**Coursework:** Evidence & Uncertainty; Machine Learning; Principles & Techniques of Data Science; Probability for Data Science; Foundations of Data Science; Computational Structures in Data Science; Data Structures; Linear Algebra

**Honors & Awards:** Dean's List - College of Comp Data Science & Society

## EXPERIENCE

### UC Berkeley College of Computing, Data Science, and Society

Berkeley, CA

*Undergraduate Student Instructor (uGSI) - Foundations of Data Science*

Aug 2025 - Dec 2025

- Led weekly 2-hour discussion sections of 45 students, providing instruction on keystone topics in the world of Data Science.
- Designed 15+ Jupyter Notebook demos to convey statistical concepts (e.g., Linear Regression, CLT) using real-world datasets manipulated with Pandas and seaborn visualizations to ensure intuitive understanding through interactive data exploration.
- Facilitated 4 weekly office hours for 40+ students, adapting course fundamentals to diverse learning paces.

*Undergraduate Researcher - StrangLab*

May 2025 - Dec 2025

- Engineered a scalable Wavelet-transform pipeline to generate 10M+ coefficients across 1,000+ audio files, creating a foundational dataset for hypothesis testing to validate a Quasi-Sparse Hyperprior assumption with 82% statistical model alignment.
- Developed procedural top-down coefficient splits by 30 frequency groups, effectively reducing NumPy data processing latency.

### NFL+

Berkeley, CA

*Marketing Analytics Student Consultant*

Jan 2025 - May 2025

- Led 500+ consumer surveys and 20+ person focus groups to shape digital strategy and audience segmentation.
- Leveraged Decision Tree Analysis to identify high-purity user profiles (.23 Gini-Impurity index) for targeted marketing.

### Major League Soccer (MLS)

Berkeley, CA

*Data Science Student Consultant*

Aug 2024 - Dec 2024

- Identified 3 critical key performance indicators (KPIs) for player valuation with 90% confidence by conducting permutation testing on 10,000+ entries from Premier League Data, replacing intuition-based scouting with data-driven metrics.
- Visualized high-dimensional player data using Matplotlib & seaborn to make abstract statistical findings interpretable.

## PROJECTS

### Fine-tuned MCQ Qwen-Instruct

*LLM Supervised Fine-tuning*

Nov 2025 - Dec 2025

- Curated a specialized training set of 985 MCQ samples using a blend of synthetic data and selected public sources to improve a pre-trained Qwen2.5-0.5B-Instruct model's capacity for machine learning and general logic reasoning in MCQ format.
- Achieved 1st in model accuracy score (57.6%) in a 417-student cohort, demonstrating strength in supervised-finetuning methods.

### GMU Speech Accent Archive Detection

*Spectrogram CNN Classification*

Nov 2025 - Dec 2025

- Transformed raw audio data with TorchAudio into STFT Spectrograms for downstream accent "image" classification.
- Implemented a custom PyTorch CNN & ResNet Architecture to stabilize vanishing gradients across 400,000 trainable parameters.
- Optimized training via AdamW & LR-Scheduler, achieving 62% validation accuracy in Automatic Speech Recognition (ASR) tasks.

### BearTensor

*PyTorch Imitation*

Nov 2025

- Engineered a NumPy wrapper as a tensor-like class from scratch to implement automatic differentiation for gradient tracking.
- Parallelized Adam, Momentum, and Stochastic Gradient Descent (SGD) within the custom BearTensor framework to model a simplistic single-layer neural network (NN) to achieve a .93 RMSE on Red Wine Quality Scores dataset.

### Noise Invariant Fashion-MNIST

*Augmented Image Classification*

Oct 2025

- Improved model generalization by performing in-memory data augmentation (rotations, blurs, shifts) on NumPy-represented Fashion-MNIST images, scaling the training set from 50K to 250K and effectively boosting MLP classifier accuracy by 20%.
- Benchmarked data-quality applying KMeans to a 10,000 image sample, revealing categorical risks of classification error.

## SKILLS & INTERESTS

**Skills:** Exploratory Data Analysis (EDA), Data visualization, Statistical Modeling, Natural Language Processing (NLP)

**Tools and Languages:** Python, SQL, Pandas, PyTorch, scikit-learn, seaborn, git, C++, Tableau, Microsoft Office

**Interests:** Pickle-ball, Pickup Soccer, Live Music, Mario Kart, Reading, Language Learning, Travel, Vinyl