

# Mohamed Ganda

San Francisco, CA | Email: [mohamedgandal66@berkeley.edu](mailto:mohamedgandal66@berkeley.edu) | [mokindacool.github.io](https://mokindacool.github.io)

## EDUCATION

**University of California, Berkeley**

*Data Science, B.A. | Certificate in Entrepreneurship & Technology*

Berkeley, CA

*May 2026*

Relevant Coursework: Multivariable Calculus, Linear Algebra and Differential Equations, Discrete Mathematics and Probability, Principles & Techniques of Data Science, Data Structures, Data Engineering

Affiliations: Computer Science Mentors (Coordinator & Python Teacher)

## EXPERIENCE

## ASUC OCFO (Data & Modeling Team)

## *Data Engineer*

*Sept. 2024 - Present*

- Built and optimized ETL pipelines in Python and SQL, automating data cleaning, transformation, and loading of financial records, improving processing efficiency by 60%.
- Implemented a PDF processor to standardize Excel inputs into 9+ unified schema fields, enhancing data consistency and accessibility.

## Vectorly

## React, Node.js, PostgreSQL

*Present*

- Engineered a full-stack web application for interview preparation, implementing reusable components and RESTful APIs, supporting 1K+ users in early testing with <250ms average response times.
- Built and optimized ETL pipelines in Python and SQL, automating data cleaning, transformation, and loading of financial records, improving processing efficiency by 60%.

## Afterwork

**San Francisco, CA**

## Data Engineering Intern

*June 2023 - Aug 2023*

- Developed machine learning techniques (k-means, decision trees, PCA, logistic regression) to perform unsupervised customer segmentation analysis on 200,000+ users, increasing product penetration by 10%.
- Built customer profiles of distinct user groups, enhancing targeted marketing strategies and improving user engagement by 15%.

## PROJECTS

## LSTM-Based Text Generation Model

# Python

*Oct 2025*

- Developed a character-level LSTM language model for next-character prediction, implementing a full training pipeline with sequence generation, hyperparameter tuning (hidden units, sequence length, dropout), and train/validation monitoring.

## Semi-Structured Data Analytics Pipeline (MongoDB + Postgres)

## Python, SQL, MongoDB, Postgres

Nov 2025

- Developed and optimized analytical pipelines across **MongoDB, Postgres, and Pandas**, transforming nested JSON Yelp documents into structured datasets, aggregating missing-value patterns, and executing cross-collection joins
- Evaluated storage strategies for 10GB+ student-athlete and sports-budget datasets by comparing JSON documents, relational schemas, and dataframes for aggregation and joining-heavy workload.

## Commodity Forecasting Challenge (Kaggle Competition)

# Python

July 2025

- Developed time-series forecasting models (XGBoost, LightGBM, LSTM) on multi-market data from LME, JPX, US Stocks, and Forex, achieving a top 15% cross-validation score on the competition leaderboard.
- Engineered lag features, rolling statistics, and volatility indicators to optimize for a Sharpe-ratio-based objective, using Python (NumPy, pandas, scikit-learn).

## Predictive Housing Analytics in Cook County

*Python, Pandas, Scikit-learn, Seaborn*

Oct 2024

- Conducted data preprocessing and exploratory data analysis on a 200K+ record housing dataset by cleaning missing values, filtering outliers, and creating new features (log-transformed price data) for better analysis.
- Designed and implemented feature engineering pipelines, including one-hot encoding categorical variables & creating derived variables by pairing the features to optimize data readiness for linear regression modeling.

## SKILLS

**Languages:** Python, Java, JavaScript, HTML, CSS, SQL, NoSQL, R

**Tools/Frameworks:** React, Bootstrap, Google Maps API, Figma, Git, Github, Kaggle, Jupyter Notebook, Docker

**Libraries:** pandas, NumPy, scikit-learn, XGBoost, LightGBM, TensorFlow, Matplotlib, NLTK, spaCy