

PROFILE

- Full-stack data scientist with expertise in machine learning, cloud computing, and data engineering. Proven track record of developing production-grade ML models and data pipelines, with experience in healthcare, e-commerce, and biotech domains. Strong foundation in both technical implementation and business impact, delivering scalable solutions that drive measurable outcomes.

EXPERIENCE

- Qventus** San Francisco
Data Scientist, Intern February 2025 - Present
 - Developing machine learning models to predict case response status and optimize healthcare operations using AWS SageMaker
 - Building risk prediction models for case cancellations using healthcare datasets
 - Implementing data pipelines using Snowflake and AWS (S3, Lambda) to process and analyze healthcare data
- Raydiant** San Francisco
Data Scientist, Intern October 2024 - February 2025
 - Developed and optimized ETL pipelines using AWS (S3, Glue, Lambda, Athena) to process customer data, reducing processing time by 30% for 3M+ records
 - Built customer churn prediction model using Python and scikit-learn, achieving 85% accuracy and identifying key factors driving customer retention
 - Created interactive executive dashboards using Python and Tableau, providing real-time analytics for cross-functional teams
- Axaitech** Cape Town, South Africa
Data Scientist, Intern July 2022 - August 2022
 - Collaborated with team to optimize ML model, reducing gene expression analysis from 58,000 to 16 genes, leading to 75% cost reduction and 60% faster processing
 - Built cancer classification model with 95% accuracy using minimal gene expressions, enabling faster diagnosis and reducing false positives by 40%
 - Conducted data cleaning, visualization, and feature selection using Python, improving model performance by 25%
- DataRes** Los Angeles
Data Scientist, Consultant October 2022 - May 2023
 - Built Python Streamlit dashboard to provide real-time feedback on home descriptions, analyzing 10,000+ listings to identify key factors for content improvement, resulting in 30% higher engagement rates
 - Developed and tuned OCR model for reading wine labels, processing images for accuracy, and creating visualizations to communicate results to stakeholders
- Matt's Tutoring Services** Los Angeles, CA
Math Instructor September 2021 - Present
 - Tutored 25+ students in subjects including Algebra, Geometry, Calculus, Linear Algebra, Statistics, and Data Science, improving grades by average of 20-30%
 - Managed and delivered tutoring sessions both in-person and virtually via Zoom, maintaining high student engagement

PROJECTS

- Kaggle Competition: Advanced Machine Learning 2025 (USF Course Project)**
 - Secured 5th place out of 90 students in Advanced ML competition using Matrix Factorization (Funk SVD) and Neural Networks for predicting childhood growth curves
 - Achieved 92% accuracy through advanced feature engineering, hyperparameter optimization, and model stacking techniques
- Fingerhut: Cart Abandonment Prediction (UCLA Capstone Project)**
 - Led data preparation and cleaning for ML model to predict cart abandonment, achieving 85.7% accuracy with Light Gradient Boosting model
 - Provided detailed insights and recommendations to stakeholders on reducing cart abandonment and boosting revenue
- AI-Powered Weight Loss Chatbot Web App**
 - Designed and built AI-driven weight loss chatbot using FastAPI, PostgreSQL, and React with LLM-based natural language processing
 - Implemented macro/calorie estimation, user profile tracking, and interactive visualizations for weight/nutrition trends

EDUCATION

- University of San Francisco** San Francisco, CA
Master's of Science in Data Science (MSDS) Expected July 2025
 - Coursework: Linear Regression, Relational Databases, Computation in Python, Statistics, Data Structures and Algorithms, Distributed Computing, Machine Learning, Deep Learning, Experimental Design, Distributed Data Systems, Data Engineering
- University of California, Los Angeles (UCLA)** Los Angeles, CA
Bachelor of Science in Data Theory June 2023
 - Roles: Data Scientist at UCLA DataRes and Vice President at UCLA's Club Sampling Club
 - Coursework: Machine Learning, Math Modeling, Python and Technologies for Data Science, Data Analysis and Regression, Statistical Programming with R, Algorithms, SQL

SKILLS

- Cloud & Infrastructure:** AWS (EC2, Lambda, S3, Glue, Redshift, Athena, SageMaker, ECS, ECR), Snowflake, Postgres, MongoDB
- Programming:** Python (NumPy, Pandas, Seaborn, Matplotlib, Scikit-Learn, PyTorch, FastAPI, Flask), R, SQL, Git
- ML & AI:** Deep Learning, NLP, Computer Vision, LLMs, Feature Engineering, Predictive Modeling, A/B Testing, Time Series Analysis, Ensemble Methods, Gradient Boosting, Random Forests, Neural Networks, MLFlow
- Data Engineering:** ETL/ELT Pipelines, Data Warehousing, Data Modeling, Data Quality, MLOps, Model Deployment & Monitoring, Data Governance
- Big Data:** Apache Spark, Distributed Computing, Parallel Processing, Real-time Analytics, Stream Processing
- Visualization:** Tableau, Plotly, Interactive Dashboards, Storytelling with Data