# Ziteng Cheng

Los Angeles, CA zitengcheng3@gmail.com LeonZitengCheng LinkedIn

## Education

University of California, Los Angeles, B.S. Statistics and Data Science — Class of 2026

## Employment

**Math Tutor** — Fed 2024 - Jun 2024

Delivered over 40 personalized tutoring sessions, analyzing students' performance data to identify knowledge gaps and adjust lesson plans.

Designed tailored study plans and practice sets, leading to an average 15% improvement in exam scores among students over a 16-week period.

**Operation data Analyst** — Jun 2024 - Sep 2024

Designed and implemented SQL queries to track daily KPIs (on-time rate, average delivery time, cost per shipment), enabling data-driven decision-making for a 3-person dispatch team.

Developed a predictive scheduling model based on historical demand data, reducing average delivery delays by 23.1% and improving resource allocation efficiency

**Data Analysis Intern** — Jun 2025 - Aug 2025

Built and automated Tableau dashboards and client-facing reports that visualized real-time occupancy data from 100+ IoT sensors, enhancing operational visibility and decision-making efficiency by 30%.

Developed and maintained end-to-end data pipelines transforming 1M+ raw IoT data points daily into cleaned, aggregated datasets used for executive dashboards

## Project

Federated Lasso Regression with Coordinate Descent — Spring 2025

Implemented from scratch a coordinate descent algorithm for Lasso regression in R, including custom soft-thresholding and convergence criteria.
Tuned the regularization parameter $\lambda$ via validation loss across three independent datasets (600 predictors each); reported non-zero coefficient indices and test loss for each model.
Designed a federated learning framework where three data owners iteratively performed local Lasso updates and shared only model coefficients with a trusted aggregator, ensuring data privacy.

Regression Modeling of Amazon Order Totals — Summer 2025

Built supervised learning models to predict log-total Amazon order amounts using a dataset of ~5,000 customers across states and months
Produced a fully reproducible R pipeline with annotated scripts and a technical report detailing diagnostics, limitations, and potential improvements

Memory Study — Experimental Design and ANOVA Analysis — Spring 2025

Designed and conducted a 2×3 factorial experiment to study how beverage temperature (cold milk, warm milk, hot tea) and emotional state (happy, neutral, sad) affect short-term memory performance.

Implemented a Python-based random sampling scheme and determined required sample size (power = 0.8, $\alpha$ = 0.05, effect size = 0.25) using G*Power.

Collected and analyzed data from 261 participants; applied two-way ANOVA and Tukey HSD post-hoc tests in R to assess main and interaction effects.

Verified model assumptions via residual and Q–Q plots; interpreted that emotional state had a statistically significant but distributed effect on memory performance.

Savills Commercial Real Estate Market Analysis UCLA DataFest 2025

Built a multiple linear regression model to quantify the impact of economic indicators and lease characteristics on market growth and leasing activity.
Incorporated qualitative insights (e.g., political, social factors) to complement the quantitative model and propose actionable recommendations to maximize return on investment.
Delivered data-driven strategic guidance to Savills on high-opportunity regions despite limitations in post-project impact assessment due to competition constraints.

Regression Modeling of Amazon Order Totals Summer 2025

Built supervised learning models to predict log-total Amazon order amounts using a dataset of ~5,000 customers across states and months.
Performed extensive feature engineering, preprocessing, and EDA (8+ visualizations) to handle categorical variables, missing data, and seasonality.
Compared and tuned 5+ regression algorithms (linear regression, random forest, XGBoost, etc.) with v-fold cross-validation, selecting the best model based on RMSE.
Produced a fully reproducible R pipeline with annotated scripts and a technical report detailing diagnostics, limitations, and potential improvements.

Classification of U.S. County Election Outcomes Summer 2025

Developed models to classify the 2020 U.S. presidential winner (Biden vs. Trump) for 3,111 counties using demographic and education predictors.
Engineered features, explored interactions, and implemented data preprocessing tailored to highly imbalanced classes.Built supervised learning models to predict log-total Amazon order amounts using a dataset of ~5,000 customers across states and months.
Evaluated and tuned multiple classification methods (logistic regression, random forest, SVM, XGBoost, neural networks) using cross-validation, achieving high Accuracy/AUC on the private Kaggle leaderboard.
Delivered a reproducible R workflow and concise technical report summarizing methodology, model comparison, and practical insights.

Federated Lasso Regression with Coordinate Descent Spring 2025

Implemented from scratch a coordinate descent algorithm for Lasso regression in R, including custom soft-thresholding and convergence criteria.
Tuned the regularization parameter $\lambda$ via validation loss across three independent datasets (600 predictors each); reported non-zero coefficient indices and test loss for each model.
Designed a federated learning framework where three data owners iteratively performed local Lasso updates and shared only model coefficients with a trusted aggregator, ensuring data privacy.

## SKILLS

**Programming Language**: R, Tableau, SQL, C++, Java, Python.
**Analytical Skills**: Regression modeling, Statistical testing, Data visualization.
**Technical Skills**: Microsoft(Excel, Word, PowerPoint).
**Leadership Skills**: Thoughtful acting, Patient listening, Problem solving, Teamwork, Written and verbal Communication.
**Languages**: English (fluent), Mandarin(native).