Junnan Lyu

Email: junnanl18@g.ucla.edu | Phone: (626) 996-0445 | Los Angeles, CA

EDUCATION

University of California, Los Angeles — **Los Angeles, CA**
Major: B.S. in Statistics and Data Science — *Sept 2024 – Expected Apr 2026*
GPA:3.73/4
Related Courses: Machine Learning, Experimental Design, Optimization, Statistical Modeling, Data Visualization

Mt. San Antonio College — **Walnut, CA**
Major: Transfer only in Computer Science — *Aug 2022 – Jun 2024*
GPA:3.93/4

RESEARCH EXPERIENCE

UCLA Complex Networks Lab — **Los Angeles, CA**
Data Engineer & Statistical Analyst | Advisor: Prof. VwaniP. Roychowdhury — *Aug 2025 – Nov 2025*

- Collaborated with clinicians to annotate 400+ seizure videos; achieved ≥80% inter-annotator accuracy.
- Built Python and R pipelines for metadata integration, data validation, and stratified random sampling.
- Developed an LLM-based Python tool to compare annotator justifications, boosting efficiency by 60%.
- Ensured dataset consistency at 85% feature-level verification and improved collaboration between technical and clinical teams.

UCLA Directed Research (PS199) — **Los Angeles, CA**
Research Assistant | Advisor: prof. Susanne Lohmann — *Sept 2025 – Dec 2025*

- Analyzed manual and automated Schelling-type segregation models to study how individual tolerance thresholds and population density relationship with people region clustering;
- Implemented an AI-assisted NetLogo version of the Iannaccone MARAS model, made an accessible agent-based simulation with 3 to 4 important factors in the theory.
- Designed coding-free simulation prompts that enable students with no programming background to use AI to generate working NetLogo model and understand theory.

PROJECTS EXPERIENCE

U.S. County Election Outcome Classifier — **Los Angeles, CA**
Team Member — *Aug 2025 – Sept 2025*

- Situation: Classify U.S. counties' presidential outcomes using demographic and economic features.
- Task: Train and compare multiple machine learning models under reproducible conditions.
- Action:
  - Engineered datasets in R (tidymodels), applied k-fold cross-validation, and tested Random Forest, KNN.
  - Tuned hyperparameters and normalized features to stabilize variance across folds.
  - Visualized learning curves and accuracy comparisons for eight candidate models.
- Result: Achieved 93.5% test accuracy; developed internal "Folder Assurance" structure for reproducible experiments.

Amazon Monthly Order Forecasting (2018–2022) — **Los Angeles, CA**
Team Lead — *Aug 2025 – Sept 2025*

- Situation: Forecast Amazon's regional order volume for 2023 using multi-year data.
- Task: Build and evaluate predictive models with high accuracy and stability.
- Action:
  - Tuned hyperparameters and normalized features to stabilize variance across folds.
  - Led 4-member team; developed Random Forest and LightGBM models in R.
  - Coordinated team data-cleaning and feature-engineering strategies.
  - Integrated codebases into a unified reproducible pipeline with clear documentation.
- Result:

- Achieved 93% accuracy and ranked Top 3 on course leaderboard.
- Pipeline later shared for instructional use.

ASA DataFest — Savills Real Estate Challenge **Los Angeles, CA**
Team Leader, Data Analytics Assistant *Apr 2025 – Apr 2025*

- Situation: Analyze national leasing data to identify office market trends for Savills.
- Task: Recommend strategies for corporate office expansion and risk resistance post-COVID.
- Action:
  - Directed 5-person team; allocated roles in modeling, visualization, research, and presentation.
  - Applied regression and heatmap visualization to assess location stability and growth potential.
  - Presented findings on high-resilience submarkets and expansion opportunities to judges.
- Result:
  - Earned commendation for innovation in visualization and model interpretation.
  - Deepened understanding of translating statistical insights into actionable business advice.

Westwood Housing Price Analysis **Los Angeles, CA**
Team Member *Feb 2025 – Mar 2025*

- Situation: Quantify key drivers of Westwood housing prices using a dataset of 345 properties; deliver an interpretable model for non-technical stakeholders.
- Action:
  - Cleaned and prepared features; fit a log-price regression with curated structural and lot variables.
  - Verified assumptions with residual and Q–Q plots and checked multicollinearity (VIF ≈ 1.4–2.4).
- Result:
  - Achieved $R^2 \approx 0.859$ with residual SE ≈ 0.267.
  - Identified 5 significant predictors — Square Feet, single-family, HOA, Year Built, Lot Size — as primary drivers.
- Impact: Produced clear stakeholder visuals highlighting elasticities and effect sizes, enabling data-backed pricing discussions and location comparisons.

EXTRACURRICULAR LEADERSHIP

Associated Student Government **Walnut, CA**
Student Service Senator *Jun 2023 – Jun 2024*

- Represented student concerns in legislative sessions.
- Contributed to funding allocation and policy drafting.

Math Club, MT.SAC **Walnut, CA**
Vice President *Apr 2023 – Nov 2023*

- Revived club within one week; organized math competitions and tutoring sessions.
- Led collaboration with professors to host final-exam review workshops for 100+ students.

TECHNICAL SKILLS

- Programming: Python, R, C++
- Machine Learning: Random Forest, LightGBM, XGBoost, Neural Networks
- Statistics: Bayesian Inference, Causal Inference, Experimental Design, Regression Modeling
- Tools: RStudio, VS Code, GitHub, Excel, PowerPoint
- Data Techniques: Sampling, Feature Engineering, Data Cleaning, Visualization

PUBLICATION

Coauthor , “Seizure-Semiology-Bench: Benchmarking Multimodal Large Language Models on Fine-Grained Spatio-Temporal Clinical Video Understanding” Manuscript submitted for review at (CVPR 2026).