Khang N. Thai

khangthai03@g.ucla.edu | 11733 Ridge Run Way, CA 92131 | (858) 924-2066 | LinkedIn | Github

Objective

Obtaining B.S in Statistics and Data Science Spring 2025 and starting M.S. in Data Science Fall 2025. Seeking Summer Internship

Education

University of California, Los Angeles   *LosAngeles, CA*
- B.S in Statistics and Data Science   *September 2023 - June 2025 (Expected)*
- **Course Completed:** Statistical Programming with R, Data-Driven Mathematical Modeling, Probability, Data Analysis and Regression, Computational Statistics with R, Mathematical Statistics, Statistical Models and Data Mining, Computation and Optimization for Statistics, Linear Models, Monte Carlo Methods, Python for Data Science
- **Current Courses:** Applied Sampling, Statistical Consulting, Social Data Science

San Diego Miramar College   *San Diego, CA*
- GPA 3.95   *August 2021 - May 2023*
- Dean's List 2021 - 2023

Skills

Languages: R, Python, C++, Java, JavaScript, SQL
Technologies/Frameworks: Microsoft Excel, Tableau, Stripe API, Django, NumPy, Pandas, Jira, React, Github, Git

Work Experience

Software Developer Backend Intern, NeverEnding, Inc   *Remote*
- Work on enhancing streamline payment systems using **Python** and **Stripe API** Integration   *May 2024 - August 2024*
- Built secure and scalable systems using **Python** for managing subscriptions and gifting services.
- Worked closely with backend and frontend teams to implement features for the web platform.
- Utilized **Jira** for task management and **GitHub** for code repository management.
- **Tech: Python, Stripe API, Django, JSON, React, JavaScript, Github, Jira**

Projects

LA Airbnb Analysis
- Analyzed Airbnb listings and reviews in Los Angeles to provide insights for hosts to improve performance and profitability.

- Identified key factors for achieving superhost status using **logistic regression** and **random forest models** (74% and 76% accuracy). Key factors included response rates, acceptance rates, and ratings.

- Used **Natural Language Processing (NLP)** to analyze guest reviews and identify issues with low-rated listings. Utilized **Latent Dirichlet Allocation (LDA)** for topic modeling.

- Built an interactive dashboard with **tableau** to help guests filter listings by price, room type, and amenities.

Amazon Purchase Predictor Regression Model
- Predicted total Amazon purchases (log_total) based on explanatory variables using **regression analysis**.

- Analyzed Amazon purchase datasets, focusing on key variables like household size and order count.

- Performed **data preprocessing**, **model evaluation**, and **hyperparameter tuning** to improve **prediction accuracy**.

- Identified positive correlations between larger households, higher income, and increased order volume, refining the model to improve predictions.

The Effect of Music on Puzzle Cube Solving
- Investigated how different music genres (classical, heavy metal, no music) affect cognitive performance in puzzle solving.

- Used a **split-plot experimental design** with 36 participants, analyzing both music genre and sequence of presentation, with gender as a blocking factor.

- Data were analyzed using repeated measures ANOVA to assess the main effects of music genre and presentation order, as well as interaction effects.

- Results indicate that classical music may enhance task-related cognitive performance, whereas heavy metal may inhibit it.