Khang N. Thai

11733 Ridge Run Way, CA 92131 | khangthai348@gmail.com | (858)-924-2066 | Linkedin | Github | Portfolio

Education

University of Southern California, MS Applied Data Science — August 2025 – May 2027 (Expected)

University of California, Los Angeles, BS Statistics and Data Science — September 2023 – June 2025

- **Relevant Coursework:** Statistical Programming with R, 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, Applied Sampling, Statistical Consulting, Databases

San Diego Miramar College — August 2021 – May 2023

- GPA 3.95
- Dean's List 2021 - 2023

Experience

Software Development Backend Intern, NeverEnding, Inc. – Remote — May 2024 - August 2024

- Work on enhancing streamline payment systems using **Python** and **Stripe API** Integration.
- 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 — github.com/Airbnb Analysis

- Analyzed Airbnb listings and reviews in Los Angeles to provide insights for hosts to improve performance and profitability.
- **R** was used to identify 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 — github.com/Amazon 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.
- Used **R** to perform **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.

Technologies

Languages: R, Python, C++, Java, JavaScript, SQL

Technologies: Microsoft Excel, Tableau, Jira, React, Github, Git, Stata, PostgreSQL