

Nguyen Hoang Thien

Van Ninh commune,
Khanh Hoa province
0868301039
Nhthien30kh@gmail.com
Data Analyst

Skills

Databases: MySQL, BigQuery, Spreadsheets, Excel.

Visualization tools: Tableau, R, Looker Studio, Plotly, Dash.

Statistical Methods: Hypothesis Testing (t-test, chi-square), A/B Testing, ANOVA

Languages: C++, Java, Python (Numpy, Pandas, Scipy, Seaborn, Matplotlib, Statsmodels, Scikit-learn).

Machine Learning Models: Regression (Linear, Logistic), Naive Bayes, Decision Trees, Random Forest, XGBoost.

Projects

CASTNET/NADP Chemical Exposure (DA)

Built ETL pipeline for 25-year multi-source dataset (reducing 70%). Detected cross-network missing-pattern correlations between 2 monitoring systems, revealing temporal anomaly clusters.

Cyclistic Bike-Share (DA)

Processed and cleaned >1M trips, 92% improvement in data quality consistency. Identified casual riders take 66% longer trips. Revealing 50k-use high-traffic hubs near Lake Michigan and numerous low-activity stations.

BellaBeat Health (DA)

Identified that ~60% of users fall into low-moderate activity. Detected abnormal HR patterns in ~8% of users. Estimated that targeted personalization (step goals, nudges) could increase daily activity consistency by 12–18%.

Japan Restaurant (DA)

Identified top 20% products contributing 80% revenue. Detected low-performing items (bottom 10%). Found peak hours (used to design staffing recommendations). Found that promo users have +90% higher AOV.

Employee Churn Prediction (DS)

Built an end-to-end automated ML pipeline (automated 85%). Identified the key churn drivers, high-risk employee groups. Improved model accuracy from 0.8 to 0.98 using XGBoost and deployed the final model with FastAPI.

Education

HCMC University of Transport / Bachelor's Degree in Electronics and Telecommunications Engineering
9/2018 – 9/2023 – Thanh My Tay Ward, HCMC, VN

FPT Academy / Automation Test

2022 – Sai Gon Hitech Park, HCMC, VN

IIG / Toeic 555

2023 – HCMC, VN

Coursera / Google Data Analytics + Google Advanced Data Analytics

4/2025 – now – Online

Portfolio

https://github.com/nhthien306AuT/Portfolio_DA