

LIHENG HUANG

liheng.huang@outlook.com | Hong Kong | github.com/BestYlay

Education

City University of Hong Kong — *MSc in Biostatistics*, GPA: 3.83/4.3 Sep 2025 – Aug 2026
East China Jiao Tong University — *B.Eng. in Data Science*, GPA: 84.3/100 Sep 2020 – Jul 2024

- Outstanding Graduation Thesis Award (Top 1% in School)

Work Experience

Hogenton Limited — *Web Developer* May 2026 – Aug 2026

- Adapted an open-source bookkeeping system into a clinic income-and-expense application through AI-assisted development, translating business requirements into data fields and validating daily, monthly, yearly, and project-level Excel reports.
- Delivered a responsive e-commerce website (hogenton.com) for a Hong Kong retailer.

Shenzhen Qianhai Xinxin Digital Technology — *Data Analyst Intern* Jan 2024 – Jun 2024

- Built XGBoost and Logistic Regression models with WOE/IV scorecards for SMS targeting, improving AUC by 0.03 and improving customer prioritization across low-risk segments.
- Mapped PostgreSQL schemas and wrote SQL queries across customer, transaction, and credit tables to deliver 10+ financial reports.

Projects

Automated FinTech KPI Reporting Pipeline Aug 2026 – Sep 2026
[GitHub] [Live Dashboard] *Synthetic data*

- Built a Python/PostgreSQL credit reporting pipeline with consistent definitions for approval rates, disbursement amounts, and delinquent balances; implemented incremental processing, data reconciliation, and Streamlit visualizations.
- Automated daily processing and cloud synchronization of aggregate metrics for credit funnel, loan portfolio, and origination cohort analysis; passed 19 tests and verified repeatable publication, transaction rollback, and read/write permissions.

Email Marketing A/B Test Analysis and Targeting Strategy Evaluation [GitHub] Aug 2026 – Sep 2026

- Analyzed 64,000 records from the historical Hillstrom randomized email experiment, checking data quality and baseline balance; estimated a US\$0.77 increase in two-week spending per customer for men's email versus no email (95% CI: US\$0.49–1.05).
- Compared three-action targeting strategies using nested cross-validation and AIPW, assessing stability across five random seeds; offline results did not establish an advantage for personalized over uniform targeting.

Breast Cancer Prognosis Modeling and Disease Progression Simulation [GitHub] Apr 2025 – Aug 2025

- Integrated longitudinal clinical records from the MSK-CHORD breast cancer cohort (5,368 patients) into a landmark survival analysis pipeline and Transformer model to predict subsequent survival risk at multiple follow-up time points.
- Achieved a C-index of 0.887 ± 0.009 (mean $\pm$ SD) in five-fold internal validation; used SHAP and simulated progression events to examine changes in predicted risk, interpreting simulations as model sensitivity rather than causal effects.

Skills

Data Analysis: Python, SQL, R, PostgreSQL, pandas, NumPy, Streamlit
Statistics & Experimentation: A/B Testing, Causal Inference, Survival Analysis, Cox Regression, Hypothesis Testing, Multiple Testing
Machine Learning & Tools: scikit-learn, XGBoost, PyTorch, Feature Engineering, Model Evaluation, SHAP, Git
Languages: Mandarin (Native), Cantonese (Native), English (IELTS 7.0)