HAI NAM VU

✉ namhaivu97@gmail.com • ☎ 857-326-7638 • linkedin.com/in/hai-vu • github.com/namhaivu173

PROFESSIONAL EXPERIENCE

Data Analyst | Seattle, Washington, USA | Remote **Jun 2023 – Current**
Institute for Health Metrics and Evaluation (IHME)

- Collaborated with lead researchers to develop diagnostic plots, produce **100+ publication-ready** figures/tables/slides and support the fact-checking process for the Global Burden of Diseases (GBD) studies and The Lancet climate paper series
- Designed and automated data pipelines in R and Python to extract, clean, and validate large-scale survey results, study findings and health records, improving reproducibility and efficiency across 6+ project teams
- Scaled R Markdown workflows to dynamically generate 50+ U.S. state health briefings, reducing production time by 70%
- Created algorithms in R to classify and analyze Long Covid clusters, integrating cohort data with scientific literature from 19 countries and adapting meta-regression models to estimate symptom duration for newly defined symptom groups.

Machine Learning Data Associate | Boston, Massachusetts, USA **Apr 2023 – Jun 2023**
Amazon

- Handled complex prompting processes and provided high-quality data creation, annotation, and analytics services to support the development, testing, and training of Alexa AI
- Received training in Reinforcement Learning from Human Feedback (RLHF) that guides AI models to align with human values
- Collaborated with the Quality Assurance team to verify and ensure that training inputs are useful, factual, and indiscriminate

Case Team Assistant Data Analyst | Ho Chi Minh, Vietnam | Remote **Sep 2020 – Dec 2020, Jun 2021 – Aug 2021**
Boston Consulting Group (BCG)

- Collaborated with 5 cross-functional teams at the client site to define long & short-term KPIs and redesign pricing & brand activation schemes accordingly at 4 distribution channels, resulting in a **5% gain of market shares** from local competitors
- Developed generalized **Python scripts** that extracted and transformed data from periodic sales reports; designed interactive Excel dashboards to measure performance, monitor compliance, progress & impacts of newly launched schemes, and present to non-technical stakeholders on a monthly basis
- Analyzed data from multiple work streams, performed maintenance and bug fixes on-demand, and visualized key findings to accurately report market status and recommend next-step actions during bi-weekly board meetings

DATA PROJECTS

Graduate Capstone: KDI Media – Boosting grocery stores performance (R, Python)

- Collaborated in a team of five to perform extensive data cleaning, data wrangling, and exploration data analysis on a dataset consisting of over 9 million rows of grocery sales data in the US
- Conducted time-series and market basket analysis (Apriori algorithm) to identify trendy products and correlated item sets, thus optimizing store layout and cross-selling strategies
- Executed in-depth product and store segmentation analysis, generating actionable insights that guided strategies for customer acquisition via reward card program, personalized product recommendations, and flash sales initiatives

Stock Index Interactive Webapp design using Streamlit (Python – app deployed to Streamlit Community Cloud)

- Performed web-scraping to extract real-time information of world major stock indices from Yahoo Finance
- Provided historical information of Stock indices, visualized the Efficient Frontier curve via portfolio simulations, estimated the portfolio Value at Risk, and allowed users to generate price predictions for different stock indices

Uber & Lyft ride-share prices in Boston area analysis & predictions (Python)

- Performed data mining, data cleaning and conducted exploratory analysis along with hypothesis tests on roughly 700k rows of data to explore time, distance, and fair price patterns of 13 Uber & Lyft products in Boston
- Developed machine learning pipeline using 3 regression models (Lasso, Ridge, Random Forest) while controlling for multicollinearity to estimate fare prices based on relevant predictors such as time, weather, pickup/drop-off locations, etc.
- Constructed 4 classification models (Naive Bayes, Logistic Regression, Decision Tree, Random Forest) to forecast whether ride-share prices would surge. The best model was evaluated based on F1, accuracy, and AUC scores

Housing prices in Nashville, Tennessee analysis & predictions (R, R Markdown)

- Imputed missing/irrelevant data, removed outliers and regrouped categorical features to allow for more precise analysis on 56k rows of data. Created relevant visualizations to extract insights on the main factors affecting house prices
- Trained 4 predictive models (Linear Regression, KNN, Random Forest, Gradient Boosting) on partitioned data set to estimate housing prices using 15 relevant predictors. The best model was evaluated using RMSE, MAPE, R-squared and adjusted R-squared

EDUCATION

Master of Professional Studies in Analytics - Statistical Modeling (GPA 3.6) | Northeastern University **Dec 2022**
Master of Finance in Alternative Investments (GPA 3.88) | University of Massachusetts – Amherst **Sep 2020**
BBA in Operations & Information Management (GPA 3.67) | University of Massachusetts – Amherst **May 2019**