LE DUC HAU

Ho Chi Minh City, Vietnam | haule.work@gmail.com | (+84) 866923231 | linkedin.com/in/hauld2304 | https://github.com/hauledata

CAREER OBJECTIVE

A dynamic Data & AI Engineer with over 3 years of experience specializing in data analysis, interactive dashboard creation, and process automation. Proven ability to architect automated data pipelines, engineer AI-powered Retrieval-Augmented Generation (RAG) agents, and deliver data-driven insights that significantly reduce manual effort by up to 95% and improve strategic decision-making. Passionate about transforming raw data into actionable intelligence and building robust, scalable data solutions.

EDUCATION & CERTIFICATIONS

University of Economics Ho Chi Minh City Sep 2021 - Oct 2024

Bachelor of **Data Science**

- **GPA: 3.7/4.0**

EXPERIENCE

ATHENA STUDIO - Automation & Prompt Engineer Apr 2025 – Present

- **Built a closed-loop automated pipeline:** Use **n8n** to fetch newly released mobile games by genre from market intelligence APIs, ingest them into a centralized database, and filter daily for top publisher tags. This workflow processes **100+ new game entries per genre each day**, updates the dataset **up-to-date**, and reduces manual curation effort by **90%**.
- **Designed a Retrieval-Augmented Generation (RAG) agent**, and **configured long-term memory** to retain user context across sessions. By integrating **LLMs and document-parsing pipelines** (PDF, Docs, Excel, CSV)**.** This agent powers an internal domain-specific chatbot with 85% accuracy, cutting cross-team information search time by 40%.
- **Enhanced the domain-specific chatbot with automatic follow-up messaging capabilities**, and **configured long-term memory** to retain user context across sessions, shortening **average response times** by **30%** and boosting **engagement** by **25%**.
- **Refined prompt engineering strategies (system prompts, context injection, dynamic templating) to optimize AI response accuracy for use cases** such as game discovery assistance, user support Q&A, and internal prototype validation. These improvements raised **task-relevant response rates** by **35%**.

MAZHOCDATA COMMUNITY - Data Analyst cum Data Engineer Aug 2022 - Sep 2024

Project: Learning Website for Automated Payment Checking (Mar 2024 - June 2024)

- Objective: Create an **eLearning website** integrated with **automated payment verification**, immediately feeding new registrations into Google BigQuery and sending timely email notifications.
- Developed a **responsive WordPress site** (Landing page + 3 Product pages).
- Built a **Zapier workflow** to connect the **payment gateway** to the site's registration form, loading data into BigQuery within **under 2 minutes of each transaction.**
- Configured automatic email notifications status payment with a **99.5% delivery success rate**, **reducing manual intervention** by **95%**.
- **Outcome**: System operates **24/7, cutting manual reconciliation by 95%,** raising data accuracy to 99%, and reducing average registration confirmation time from 1 days to under 5 minutes.

Project: Registration Form Operations Dashboard (Sep 2023 – Oct 2023)

- **Objective:** Implement an **interactive Metabase dashboard** to monitor **25+ livestream programs** in **real time**, tracking **daily registrations**, **conversion rates**, and other **key performance indicators**.
- Designed a **Metabase dashboard** with **overview**, **time-series charts**, and **per-program KPI snapshots**.
- **Connected Google Sheets** data to **BigQuery** and utilized the **BigQuery API** to push updated datasets to a **Dockerized Linux server** hosting Metabase.
- Configured automated email alerts triggered when **daily registrations** fluctuated by **>20%** compared to the previous week.
- **Outcome:** Elevated **data accuracy to 98%**, enabled **faster decision-making**, boosted **registration processing efficiency by 35%**, and **reduced manual reporting time by 90%**.

PROJECTS

Analyzing Street Tree Health Data In New York | Python *(View more)* Nov 2023 - Dec 2023

Objective: Evaluate NYC Open Data on street tree health to assess species distribution and overall urban tree well-being.

Key metrics:

- Tree health distribution percentages (alive, good, fair, poor);
- Regional risk index based on the presence or absence of tree guards;
- Projected impact of interventions

Approach:

- Utilized Pandas and NumPy for data cleaning—handling missing values with imputation and addressing outliers using the Modified Z-score method—and ETL
- Employed Matplotlib and Seaborn for multi-level visual analysis (city to borough).

Outcome:

- Revealed that 95.3% of trees are alive—with 77.1% in good health and 14.2% in fair health—highlighting critical vulnerabilities in Brooklyn and Queens.
- Recommended targeted maintenance and installation of tree guards in high-risk areas, a solution projected to reduce future tree health issues by 10% and enhance urban canopy longevity.

Pet Store Data Management | MS SQL *(View more)* Nov 2022 - Dec 2022

Objective: Optimize inventory management and enhance customer insights through robust data analysis and automation.

Key metrics:

- Inventory turnover rate and stockout frequency
- Overstock reduction percentages
- Reorder point accuracy and cycle time efficiency

Approach:

- Designed a comprehensive data and inventory management system using data modeling and ERD;
- Executed advanced SQL queries with aggregate functions; and implemented stored procedures and triggers to automate inventory updates.

Outcome:

- Developed a dynamic inventory solution that optimized reorder points and automated real-time updates, resulting in a 30% reduction in overstock and a 25% decrease in stockouts.
- These targeted adjustments enabled precise, data-driven decision-making, directly enhancing customer satisfaction and driving revenue growth.

CERTIFICATIONS

HackerRank - Python (Basic) Certificate *(View)* Apr 2024

HackerRank - SQL (Advanced) Certificate *(View)* Mar 2023

TOEIC: 620 *(View)* Sep 2023

TECHNICAL SKILLS

- **Automation & Integration:** n8n, Zapier
- **Languages**: SQL, Python, DAX, R, C#.
- **Data Preprocessing**: SQL, Python (Numpy, Pandas).
- **Data Analytics**: Descriptive Statistics, Time Series, Cluster Analytics.
- **Data Visualization**: Matplotlib, Seaborn, Power BI.
- **Data Management & Cloud**: Google Cloud Platform (GCP), Supabase (Postgres, Vector Store Database), Azure Data Studio.