



ĐÀO HUY HOÀNG

Kỹ sư dữ liệu

LIÊN HỆ

- (+84) 914 209 065
- peachhuyhoang203@gmail.com
- Quận 7, thành phố Hồ Chí Minh
- [LinkedIn/hoanghuydao](#)
- [Github/hoanghuydao](#)

HỌC VẤN

Trường đại học Khoa Học Tự Nhiên - Đại học Quốc Gia tpHCM

2021 - hiện tại

Khoa
Toán - tin học

Chuyên ngành
Khoa học dữ liệu

Điểm hiện tại
3.2 / 4.0

KĨ NĂNG

- Máy học và mô hình dự đoán:
Pandas, Numpy, Scikit-learn
- Ngôn ngữ lập trình:
Python, C & C++, R, Matlab
- Hệ quản trị cơ sở dữ liệu:
PostgreSQL, MySQL, SQL Sever
- Trực quan hóa dữ liệu:
PowerBI, Seaborn, Matplotlib, ggplot2
- ETL và Data Warehouse:
SSIS, Databricks (Azure)
- Khác:
Microsoft Office (word, excel, ppt)
Microsoft Azure (điện toán đám mây)

NGÔN NGỮ

Tiếng việt (bản ngữ)
Tiếng anh (thành thạo)

MỤC TIÊU NGHỀ NGHIỆP

Một chuyên gia trẻ đầy nhiệt huyết và tận tâm, tập trung vào lĩnh vực phân tích dữ liệu. Thông qua quá trình học tập, tôi đã xây dựng được nền tảng vững chắc về phân tích dữ liệu, mô hình hóa thống kê, và các công cụ kinh doanh thông minh. Tôi khao khát áp dụng các kỹ năng của mình để giải quyết các vấn đề thực tiễn, không ngừng học hỏi và thích nghi với những thách thức mới. Sự đam mê học hỏi thôi thúc tôi đóng góp cho các dự án sáng tạo trong môi trường năng động.

- Mục tiêu ngắn hạn:** Nâng cao kỹ năng kỹ thuật trong các công cụ phân tích dữ liệu và đảm bảo một vị trí thực tập để áp dụng kiến thức lý thuyết vào các tình huống thực tế.
- Mục tiêu dài hạn:** Phát triển chuyên môn trong khoa học dữ liệu và trí tuệ kinh doanh, hướng tới việc đảm nhận vai trò lãnh đạo trong các dự án dựa trên phân tích dữ liệu.

DỰ ÁN HỌC THUẬT

Data Analyst:

1. Phân tích dữ liệu bán hàng cà phê (Excel):

- Sử dụng các hàm nâng cao trong Excel như XLOOKUP và INDEX MATCH để truy vấn dữ liệu, và hàm IF để tính toán và phân loại.
- Định dạng lại số và ngày tháng, đồng thời kiểm tra dữ liệu trùng lặp.
- Xây dựng dashboard tổng quan bằng Pivot Table và Pivot Chart.

2. Dự án trực quan hóa dữ liệu (Power BI):

- Làm việc với dữ liệu từ Atli Grands, một công ty quản lý nhiều khách sạn lớn.
- Thực hiện ETL nhỏ thông qua Data Model trên PowerBI desktop.
- Tạo Calculated Columns và Measures bằng các hàm DAX như SUM, COUNT, CALCULATE, FILTER, DIVIDE.
- Sử dụng Tooltips để làm rõ ràng hơn thông tin trên báo cáo.

Data Science:

1. Dự báo doanh số (Python, Pandas, Numpy, Matplotlib, Scikit-learn, Jupyter Notebook):

- Dự án nhóm phân tích dữ liệu bán hàng của Walmart (2010-2013) để dự báo doanh số.
- Làm sạch và trực quan hóa dữ liệu để xem xét tác động của các ngày lễ đến doanh số.
- Xây dựng mô hình Random Forest Regression sử dụng phương pháp WMAE (Weighted Mean Absolute Error) để dự đoán doanh số.

Data Engineering:

1. Dự án xây dựng pipeline với dữ liệu AdventureWorksLT2017 (SSMS, Microsoft Azure):

- Data Ingestion:** Thiết lập tài khoản đăng nhập trong SSMS. Sử dụng Azure Key Vault để lưu trữ an toàn username và password. Tải dữ liệu từ SSMS vào Azure Synapse Workspace (lớp Bronze) thông qua Azure Data Factory bằng các hoạt động Copy Data, ForEach, và Lookup.
- Data Transformation:** Thực hiện quy trình ETL trên Databricks, chuyển đổi dữ liệu từ Bronze sang Silver (chuyển định dạng `DateTime` thành `Date`) và từ Silver sang Gold (đổi tên các cột để thống nhất).
- Data Loading:** Xây dựng pipeline để thực hiện chuyển đổi dữ liệu từ Bronze → Silver → Gold, sau đó sử dụng Azure Synapse Analytics để tạo và kích hoạt pipeline lấy dữ liệu từ lớp Gold.
- Data Reporting:** Kết nối Power BI với Azure Synapse Analytics để trực quan hóa dữ liệu theo thời gian thực từ lớp Gold.

Bài báo cáo SQL sử dụng PostgreSQL

Thành thạo trong việc tạo cơ sở dữ liệu, bảng, và thiết lập mối quan hệ, cùng với việc truy xuất dữ liệu:

- Cơ bản: **Select From Where, Group By Having, Order By, Limit.**
- Trung cấp: Các hàm tổng hợp và **Joins.**
- Nâng cao: **CTE** (Biểu thức bảng chung) và **Window Functions** (các hàm cửa sổ).

BẰNG CẤP

IELTS - Overall band score: 6.5

IDP - August/2024

Google Data Analytics Specialization

Coursera - May/2024

SQL (Basic & Intermediate)

Hackerrank - 2023