JOLIE PHAM

Email: phuonganh382000@gmail.com | Github: https://github.com/phuonganh-38 | LinkedIn: https://www.linkedin.com/in/phuonganh38/
Website: phuonganh-pham.com | Blog: phuonganh-pham.com/blog/ | Tableau: https://public.tableau.com/app/profile/phuong.anh.pham6362

PROFESSIONAL SUMMARY

Detail-oriented Data Analyst with a Master's degree in Data Science, skilled in Data Analysis, Modeling, and Visualization. Over the past 2 years, I have successfully completed a wide range of projects using **Tableau, Power BI, SQL**, and **Python** for data extraction, cleaning, and analysis. My solid foundation in Economics allows me to bring a unique perspective to data-driven decision-making, combining analytical rigor with practical insights. Passionate about continuous learning and knowledge sharing, I run a personal **blog** where I provide guidance for those entering the field of data analytics. Eager to apply my skills in a dynamic environment while continuing to grow as a Data Analyst.

KEY SKILLS

Visualization	Tableau, Power BI, Excel
Data Analysis	Data Wrangling, Exploratory Data Analysis (EDA), Alteryx
Programming language	SQL, Python
Mathematics	Statistics, Probability
Machine Learning	Classification, Regression, Clustering, Deep Learning
Libraries	Pandas, Numpy, scikit-learn, Matplotlib, Seaborn, Plotly
Web scraping	Beautiful Soup, Selenium
Database management	PostgreSQL, DBeaver
Others	Snowflake, Databricks, Microsoft Azure, Apache Spark, Airflow, dbt Cloud, HTML, CSS

CERTIFICATIONS

Tableau Certified Data Analyst. Issued by: Tableau 2025

WORK EXPERIENCE

Data Scientist | Data Analyst (Team Project Work) – University of Technology Sydney

A Deep Learning system for differential diagnosis of Melanoma Aug 2024 – Nov 2024

https://github.com/phuonganh-38/deep-learning-diagnosis-melanoma

- Collaborated with other data scientists and researchers for months to develop an AI-powered diagnostic tool for early Melanoma detection using Medical imaging.
- Designed and trained CNN models based on historical data, incorporating gender, age, and lesion location.
- Optimized medical image preprocessing to improve model precision and reduce misdiagnoses.
- Developed and deployed a Streamlit-based interface, allow users to upload or capture images directly, receive instant melanoma risk assessment, and get personalized recommendations.

Impact: Enhanced diagnostic accuracy, reduced doctor workload, and showcased AI's transformative role in healthcare.

Machine Learning for Transaction Analysis Apr 2024 – May 2024

https://github.com/phuonganh-38/transaction-analysis-ML

- Developed ML models for fraud detection using Python, skikit-learn, Pandas, Numpy, Matplotlib, Seaborn.
- Processed and cleaned **100+ transaction files** with **over 4.2 million records.**
- Achieved **99.7% fraud detection accuracy**, improving detection precision, reducing false positives

Impact: Lowered operational costs and enhanced customer experience by minimizing transaction disruptions.

MISA JSC Jan 2022 – Jan 2023

PROJECTS

AutoCleanAI – AI-Powered Data Preprocessing Tool https://github.com/phuonganh-38/AutoClean-tool 2025

App link: https://autoclean-tool.streamlit.app/

- Developed an AI-powered tool that automates critical data preprocessing tasks**, reducing preprocessing time by 80%**.
- Implemented robust features for missing value imputation, outlier removal, duplicate handling, improving data quality and workflow efficiency.
- Inplemented Isolation Forest for advanced anomaly detection.

Tech Stack: Python, streamlit, Isolation Forest, Scikit-learn, Pandas, Plotly

Rental Market Trend Analysis https://github.com/phuonganh-38/rental_market_trend_analysis 2024

- Led end-to-end analysis of rental market trends by scraping data from Domain website and performing Exploratory Data Analysis (EDA) with SQL and Python.
- Utilized data visualization techniques to present findings, improving stakeholders' ability to make data-driven decisions and adjust to market changes quickly.

Tech Stack: Python, SQL, Pandas, Matplotlib, Seaborn, Selenium

Global Mental Health Dashboard https://tabsoft.co/4iDKnd9 2024

Built an interactive Tableau-powered dashboard with **over 1,000 views**, unraveling the Global Mental Health crisis, guiding users from high-level trends to an in-depth disorder breakdown through dynamic storytelling.

ELT Data Pipeline for Airbnb Analytics https://github.com/phuonganh-38/airbnb-elt-pipeline 2024

- Developed and deployed an ELT pipeline processing over 1TB of data**, optimizing processing time by 40%**.
- Streamlined pipeline using Apache Spark, dbt Cloud, and Airflow, improving scalability and efficiency for real-time analysis.

Tech Stack: Apache Spark, dbt Cloud, PostgreSQL, Airflow, Google Cloud Platform (GCP), DBeaver, Python

Data Processing and Analysis of New York taxi https://github.com/phuonganh-38/nyc-taxi-data-analysis 2024

- Processed and analyzed 800 million rows of NYC taxi trip data, uncovering key insights into ridership patterns.
- Optimized data pipelines using Azure, Databricks, and Apache Spark, **achieving 45% faster processing**.

Tech Stack: Apache Spark, Databricks, Microsoft Azure, SQL, Python, Pandas

EDUCATION

University of Technology Sydney Feb 2023 – Dec 2024

Master of Data Science and Innovation. Grade: Distinction

Academy of Policy and Development 2018 - 2022

Bachelor of International Economics (Honor Programs). Grade: High Distinction