JOLIE PHAM

Github: https://github.com/phuonganh-38 | **LinkedIn**: https://www.linkedin.com/in/phuonganh38/
Portfolio: https://phuonganhpham.com/ | **Email**: phuonganh382000@gmail.com
Blog: https://data-tales.com/ | **Tableau**: https://public.tableau.com/app/profile/phuong.anh.pham6362/vizzes

PROFESSIONAL SUMMARY

Results-driven **Data Analyst** with a Master's degree in **Data Science**, skilled in **data analytics, statistical modeling, machine learning** and **data visualization**. Proficient in **SQL, Python, Tableau, and Power BI**, with hands-on experience in **ETL processes, big data processing**, and **predictive analytics**. Demonstrated success in building **AI-driven automation tools**, optimizing **data pipelines**, and improving **decision-making** through **data storytelling**. Passionate about **leveraging AI to enhance business insights** and actively contributing to the **data community** through **blog writing** and open-source projects.

KEY SKILLS

Data Analytics & Visualization	Tableau, Power BI, Excel, Plotly, Matplotlib, Seaborn
Programming & Database	SQL, Python (Pandas, NumPy, Scikit-learn), PostgreSQL, Snowflake, DBeaver
Machine Learning & AI	Classification, Regression, Clustering, Deep Learning (CNN), Anomaly Detection (Isolation Forest)
Big Data & Cloud Technologies	Apache Spark, Airflow, dbt Cloud, Microsoft Azure, Google Cloud Platform (GCP)
ETL & Data Pipelines	Data Wrangling, Feature Engineering, Data Preprocessing, Web Scraping (Selenium, Beautiful Soup)
Business Intelligence & Reporting	A/B Testing, Business Insights, KPI Tracking, Data Storytelling
Additional skills	Alteryx (Data prep and workflow automation), Statistics & Probability, HTML, CSS

CERTIFICATIONS

- **Tableau Certified Data Analyst** (Issued 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
- Developed an **AI-powered diagnostic tool** for **early melanoma detection** using deep learning and medical imaging.
- Engineered **CNN models**, optimizing **medical image preprocessing** for improved accuracy.
- Built a **Streamlit-based interface** enabling **real-time risk assessment** and **personalized recommendations**.
- **Impact**: Increased diagnostic accuracy, reduced physician workload, and demonstrated AI's role in healthcare.

Machine Learning for Transaction Analysis | Apr 2024 – May 2024
https://github.com/phuonganh-38/transaction-analysis-ML
- Designed and deployed **fraud detection models** achieving **99.7% accuracy**, reducing false positives.
- Processed and cleaned **4.2 million transactions**, optimizing **data pipelines** for high-performance analytics.
- **Impact**: Enhanced fraud detection, reduced operational costs, and improved customer experience.

MISA JSC
Business Analyst Intern | Jan 2022 – Dec 2022
- Conducted **exploratory data analysis (EDA)** to identify trends, patterns, and business insights.
- Developed dashboards in **Power BI** to track business KPIs and optimize decision-making.

PROJECTS

AutoCleanAI – AI-Powered Data Preprocessing Tool (2025)
GitHub: https://github.com/phuonganh-38/AutoClean-tool | App: https://autoclean-tool.streamlit.app/
- Developed an **AI-powered data preprocessing tool**, automating tasks like missing value imputation, outlier removal, and duplicate handling.
- Implemented **Isolation Forest** for advanced anomaly detection, improving data quality.
- **Impact**: **Reduced data preprocessing time by 80%**, increasing workflow efficiency for analysts.
Tech Stack: Python, Streamlit, Isolation Forest, Scikit-learn, Pandas, Plotly

ELT Data Pipeline for Airbnb Analytics (2024)
GitHub: https://github.com/phuonganh-38/airbnb-elt-pipeline
- Designed and deployed an **ELT pipeline** processing **1TB+ of data** using **Apache Spark, dbt Cloud, and Airflow**.

- Optimized **data transformation and orchestration**, reducing processing time by **40%**.
- **Impact**: Enhanced **real-time analytics**, improving business insights for property market trends.
Tech Stack: Apache Spark, dbt Cloud, PostgreSQL, Airflow, Google Cloud Platform (GCP), DBeaver, Python

Data Processing and Analysis of New York taxi (2024)
Github: https://github.com/phuonganh-38/nyc-taxi-data-analysis
- Processed **800 million rows** of taxi trip data, optimizing data pipelines using **Azure, Databricks, Apache Spark**.
- **Impact**: Achieved **45% faster processing**, uncovering insights into ridership patterns.

Rental Market Trend Analysis (2024)
GitHub: https://github.com/phuonganh-38/rental_market_trend_analysis
- 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 (2024)
Tableau: https://tabsoft.co/4iDKnd9
- Designed an **interactive Tableau dashboard** visualizing **global mental health trends** with **1,000+ views**.
- Applied **data storytelling** to provide actionable insights into mental health issues in Australia.

EDUCATION

University of Technology Sydney | Feb 2023 – Dec 2024
Master of Data Science and Innovation (**Distinction**)

Academy of Policy and Development | 2018 - 2022
Bachelor of International Economics (Honor Programs) (**High Distinction**)