

Chan Ding Hao

+65 8202 3989 | chandinghao@yahoo.com | [Github](#) | [LinkedIn](#) | [Portfolio Website](#) | [Medium](#)

Education

Singapore Management University - Bachelor of Science in Economics

Aug 2022 - July 2026 (Expected)

- **Relevant Coursework:** Machine Learning & Applications, Algorithms and Programming, Visual Analytics for Business Intelligence (Tableau), Business Data Management (SQL), Computational Thinking & Programming (Python), Data Analytics with R

Work Experience

ByteDance - Global Payments Risk Data Analytics Intern

Jan 2026 - July 2026

- Will update after the internship commences

Razer - Data Analyst Intern (Analytics Engineering)

July 2025 - Dec 2025

- Built dashboards on Apache Superset to track key metrics (Daily CTR, Daily Revenue etc) for Razer recommendation engine using SQL, reducing reporting time by 25%.
- Built transformation models (SQL) within ELT pipelines using dbt DAGs to deliver business intelligence solutions for internal stakeholders served through Apache Superset, increasing dashboard load speed by 50%.
- Developed deduplication logic for user software data through awswrangler, SQL and pandas in AWS SageMaker AI notebooks, identifying data quality issues and providing implementation specifications to data engineering.

ONE (Ocean Network Express) - Data Analyst Intern

June 2024 - Dec 2024

- Optimized truck routes using a CI/CD pipeline on GCP, leveraging Python for data transformation and analysis on VertexAI and SQL on BigQuery. Improved processing speed by 10%, enabling faster data-driven decisions and improving real-time logistics efficiency.
- Discovered important aspects related to data (e.g. Hazmat constraints), while validating discoveries with various stakeholders, reducing distance by 5%, and saving over \$500,000 USD annually with scalability potential.
- Analyzed SQL queries on BigQuery to trace undocumented data sources, enabling accurate data analysis for route optimisation.

Projects

Credit Card Fraud Detection

Aug 2025 - Dec 2025

- Engineered credit card fraud detection system using Neural Network, VAE, and CatBoost ensemble, achieving a **minimum of \$20,000 improvement over Logistic Regression baseline on cost-weighted metric**.
- Optimized for production fraud prevention with **PRAUC of 0.75, recall of 0.82, and precision of 0.71** to minimize false negatives while controlling false positive costs.
- Ensemble approach yielded **\$19,000 in additional savings vs. individual models**, deployed for real-time transaction classification while maintaining acceptable merchant decline rates.

SMU BIA Datathon - (Top 10 out of 80 teams) [Github](#)

Jan 2025- Feb 2025

- Developed a risk intelligence system for Singapore's Internal Security Department (ISD) using LLM-powered document extraction.
- Engineered LLM prompt pipelines (**Claude Haiku 3.0 on AWS Bedrock**) to structure unstructured WikiLeaks & news data, improving insight retrieval.
- Delivered interactive dashboards with extracted structured data on **Plotly Dash** to visualise trends, relationships, and risks, providing actionable insights for ISD decision-making.

SPY Analytics and Forecasting - [Github](#) - [Medium \(1\)](#), [\(2\)](#), [\(3\)](#)

Aug 2024 - Dec 2024

- Conducted extensive **data wrangling, analysis, and visualisation** of SPY's historical and inflation-adjusted returns alongside Federal Reserve inflation data to highlight long-term investment advantages and performance stability.
- Applied **time-series statistical techniques**, such as **stationarity testing (ADF) and differencing**, to preprocess SPY data for robust time series forecasting.
- Constructed a weighted ensemble model (**LSTM -PyTorch, Prophet, and Monte Carlo simulations**) for **time series forecasting**. Predicted the log returns of SPY in the next 20 years with confidence intervals.

Skills and Certifications

Technical Skills - Python, SQL, dbt, git, R

Frameworks and Libraries - pytorch, scikit-learn, xgboost, statsmodels, nltk, spacy, plotly, boto3, aws-wrangler, google-cloud, streamlit

Tools and Platforms : Apache Superset, Tableau, Github, GCP (VertexAI, BigQuery), AWS (Redshift, RDS, Bedrock, Lambda, SageMaker), Pinecone

Certifications & Training [Google Data Analytics Professional Certificate](#), [HackerRank SQL \(Advanced\)](#)