Pierre Zhang

Phone: (647) 809-0880 **Email:** pierre4spirit@gmail.com
LinkedIn: www.linkedin.com/in/pierre-z/
Website: www.pierrezhang.online

ABOUT MYSELF

Curious by nature and rigorous by training, I grow at the intersection of data, systems, and human impact. From optimizing fraud detection at Weihai Bank to building automated tools at Clutch, I've realized that data science is as much about innovation and empathy as algorithms. What I can bring is not just technical fluency, but a builder's mindset: eager to learn deeply, collaborate widely, and turn insight into meaningful changes.

EDUCATION

Bachelor of Science (Honours), Mathematics, Statistics **09/2021 - 06/2025**
University of Toronto | Toronto, ON

WORK EXPERIENCE

Quality Control Associate **07/2025 – Present**
Clutch Technologies Inc | Mississauga, ON

- Designed an automated QC-reporting tool with JavaScript that **optimizes** generation time **from 15 minutes to 3 minutes (80% reduction)** and ensures reproducible outputs
- Collaborated with cross-functional teams to use, discuss and improve real-time dashboards (**SQL**)
- Maintained documentation in **Google Sheets** and published monthly analytical summaries using **Pivot Tables**

Power BI Developer Intern **09/2024 – 12/2024**
Shandong Eagle Software Technology | China

- Developed 3 end-to-end visual analytics dashboards (**Power BI + DAX**) that shortened financial-analysis cycles by 20%
- Wrote **Python ETL** scripts to automate daily ledger reconciliation, halving error rate and lowering manual processing time by 30%. The process was fully documented for reproducibility
- Implemented **row-level security (RLS)** policies, ensuring client-specific data segregation and regulatory compliance
- Delivered weekly analytical reports that led to 3 major process optimizations with a **25% faster** turnaround for claims

Data Analyst Intern **05/2024 - 08/2024**
Weihai Bank | China

- Queried and analyzed 10,000+ daily transactions (MySQL), improving rule-based fraud-detection recall by 20%
- Built 3 interactive heat-map dashboards (**Power BI + MySQL**) for risk profiling
- Cleaned a **1.2M-record** MySQL database via outlier detection & schema normalization, reducing missing-value rate to approximately 1%

HIGHLIGHTED PROJECTS

Project Researcher & Data Analyst **05/2025 – Present**

Project Name: New York City Yellow Taxi Power BI Dashboard Report

- Engineered a 1.3 GB TLC trip-record pipeline using **Power Query**, transforming 8.5 M rows into a **star-schema** model that cut **DAX** query latency
- Built interactive **KPI** suite (trips, revenue, distance, duration, tip rate) and calendar heat-map; optimized measures to support auto-refresh in **Power BI** Service with row-level security

Project Researcher & Data Analyst **06/2025 – 08/2025**

Project Name: Canada Income & Tax Calculator (2025)

- Architected a single-page, responsive calculator that delivers **real-time** 2025 federal + provincial tax estimates for all 13 Canadian jurisdictions, incorporating CPP/QPP, EI, RRSP caps and Ontario/Quebec surtaxes
- Encoded **70+ bracket rules** in a functional engine. **Unit-tested** edge cases up to $1M annual income with sub-second response, demonstrating algorithmic rigour

Project Lead & Data Analyst **09/2024 – 04/2025**

Project Name: Campus Fridge Network

- Integrated **mixed-methods data** with 104-response survey on food insecurity, campus foot-traffic counts, and existing fridge logs covering 60,000 + students, built a consolidated demand dataset
- Used **k-means clustering** method that ranked candidate locations by student density, walking distance and turnover, producing a prioritized map that MealCare supervised and adopted, raising utilization and significantly cutting monthly food waste

Project Researcher & Statistical Modeler **06/2024 – 08/2024**

Project Name: Heart Attack Model Analysis Report

- Curated and cleaned a **1,000**-record cardiovascular dataset (14 predictors); split 75 / 25 to guard against over-fitting and reproducibly seeded all analysis
- Conducted comparative modeling study across **OLS, Ridge, and Lasso** regressors with **AIC/BIC** model selection, emphasizing generalization and statistical interpretability

SKILLS

Data Engineering & Databases: SQL, data modeling & ETL, Power Query

Exploratory Data Analysis: Python (pandas, NumPy) and R (tidyverse); data cleaning, outlier/missing handling, feature engineering, distribution/correlation analysis; produce fully reproducible EDA reports

Statistical Learning & Machine Learning:

Supervised learning: regression (OLS, Ridge, Lasso) and classification (with model selection and diagnostics)
Unsupervised learning: K-means clustering
Time series: decomposition and ARIMA forecasting; emphasis on interpretability and statistical rigor

Data Visualization: Power BI (DAX, RLS, geospatial visuals), Chart.js, ggplot2; build interactive dashboards and monitoring panels for stakeholders

Optimization & Quantitative Modeling: Objective/constraint modeling for business metrics; apply optimization principles in regularized regression, clustering, and resource allocation; systematic hyperparameter tuning

Software, Tools & Reproducibility: Python, SQL, R, Power BI, Excel/VBA, JavaScript; versioned documentation and reproducible pipelines (scripted ETL, parameterized reports, standardized data dictionaries)

Data Governance & Ethics: Data quality management, access control (RLS), compliance-aware design; careful consideration of ethical implications in data use and model deployment

Cross-functional Collaboration: Requirements clarification, metric alignment, and insight delivery; monthly analytical summaries and visuals to support multi-department decision-making