

Fred, a seasoned Data Analyst with **4+** years' experience in **AI** and **BI**. is proficient in **Python, Spark, SQL** and **LLM on Cloud**. With a background in **Finance, Biomedicine & Tourism**, he excels in delivering solutions and insights. Known for reliability and teamwork, Fred consistently achieves high performance.

### SPECIALTIES

- Data Science
- Large Language Model (AI LLM)
- Finance, Healthcare, Tourism, CRM
- Dashboard Analytics
- Cloud Big Data
- Graph Database
- Automation
- Agile Methodology

### TECHNICAL SKILLS

- Data Science: Research Methodology, ChatGLM-6B, Regression, Classification, Clustering, Ensembled methods, PCA
- Software: Azure, AWS, Databricks, Colab, Hive, Neo4j, Oracle, SQL Server, MySQL, Tableau, Power BI, Visio Studio Code, GitHub, Jira, Confluence, SerivceNow, Autosys, draw.io
- Programming: Python, Spark, R, SQL, Pandas, Sklearn, Linux, SAS, Power Shell, JavaScript, VBA

### PROFESSIONAL EXPERIENCE

#### Sr. Data Analyst | East Money Information Co., Ltd.

Jan 2023 – Present

- Project I – **Large Language Model - Tech & Finance Platform** - Lead AI Development - **73%** Qualification Rate
  - Business and Data requirement collection, evaluation and translation to technical Functional Design
  - Fine-Tuned the LL Model using ChatGLM-6B based LoRA with training sets of 4000 tables, Qualification Rate to 73%
  - Delivered database metadata interpretation engine with translated explanation for easy exploration by business
  - Contributed to Enterprise AI chat platform for internal finance knowledge sharing and standard compliance
- Project II – **Data Pipelines - Big Data Conversion to Markup** - ELT Development - Framework to handle **100+** jobs Daily
  - Designed pipelines to extract/transform Financial Security, Client data from configuration based Databases to Hive
  - Trained LLM model that transforms language query to a SQL query using BM25, Regex, Rules engine and Chatgpt API.
  - Launched 1 platform Orchestration function to automatically batch process 100+ data load jobs
- Project III – **Business Intelligence - Real Time Stats Monitoring** - Dashboard Solution Design - **90%** Saving time
  - Real time dashboards design to present AI training statistics to management for performance and progress tracking
  - Improved querying efficiency by 90% through middle layer design using SQL and Tableau for data asset team

#### Jr. Data Scientist | PWC Shanghai AC

Jul 2022-Nov 2022

- New Tech Research – **Knowledge Graph – Graph Database – NEO4J – Biomedical Field**
  - Assessed Graph Databases in the market and evaluation in terms of query time, stability, Visualization & Platform Maturity
  - Built a Knowledge Graph Proof of Concept Demo in NEO4J and presented to our Biomedical Client
- Quality Assurance – **Module Optimization – Automation – Python – OCR model improvement**
  - Provided functions of Similarity Calculation and JSON Consistency Check
  - Extracted tables from biomedical documents using OCR, Validated Table misclassification, spelling, text positioning and fed back to Model to improve OCR results.
  - Automated source document extraction via a framework design to loop through all table documents with predictive OCR model

#### CRM Analyst Intern | Trip.com

Jun 2021-Aug2021

- CRM Analysis – **Customer Segmentation – RFM Model** – Power BI – **Behavior Based Strategies**
  - **Data Analysis:** Conducted data collection and international order transaction analysis using SQL, Python in Hive and Excel
  - **Product Planning:** identified customer segments by analyzing browsing and ordering behaviours. Collaborated with marketing team to apply different strategies to targeted customer groups, e.g., acquisition tactics, membership discounts.
  - **Competitor Analysis:** Conducted thorough analysis comparing website interface, email campaign strategies, and membership policies with key competitors like Expedia and Huazhu. Contributed insights to strategic decision-making processes for trip.com

## Data Scientist Consultant | Beam Data

Sep 2019-Sep 2020

- **Project I – Patient Analysis** - Feature Engineering - Clustering Analysis - **200 record & 300 columns**
  - Built basic framework on patient data cleansing and clustering model for a medical team
- **Project II – Inventory Management** – Time series - improved accuracy by **18%**
  - Inventory level risk prediction using time series for Samsung Canada
  - Improved risk level prediction accuracy through a simulation by fitting the distribution model
- **Project III – Stock Price Prediction - Model Switch - AWS - 85% efficiency Improvement**
  - Switched Stock Pricing Model: Transitioned from XGBoost to LightGBM for the stock pricing model, significantly conserving memory and decreasing training time.
  - Deployed to AWS EC2: Implemented the entire process on AWS EC2 for production use.
  - Overfitting Lesson: Despite efforts, failed to enhance model accuracy. Learned the importance of avoiding overfitting by reducing features.

## ACADEMIC PROJECTS

### **P I – Game Win Rate Prediction Project** (League of Legends)

Oct 2019

- Developed a predictive model using League of Legends data to forecast real time team win rates during the gameplay.
- Collaborated on GitHub to access and refine project resources, demonstrating teamwork and resourcefulness.

### **P II – Inventory Management**

June 2022

- Led an initiative to optimize inventory management through advanced data analytics, targeted to improve distribution strategies and cost savings.
- Utilized simulation techniques to accurately predict inventory levels, driving operational efficiency.

### **P III – Bike Sharing Simulation**

Mar 2022

- Based on publicly available data from Toronto's bike-sharing system, a simulation is conducted to plan the arrangement of bikes at 9 am and 6 pm each day.
- The finding was published in university forum

## EDUCATION

### **Master of Engineering | University of Toronto**

Sep 2021-Nov 2022

Mechanical & Industrial Engineering(Analytics), GPA: 3.8 A+

### **Bachelor of Science | University of Toronto**

Sep 2015-Jun 2019

Mathematical Finance and Statistics

## CONTEST

- **Smart Beta – Portfolio Model** – Top 7 **Mar 2019**
- **TD Bank Finance Risk Management** – Top 5 **Sep 2018**

## CERTIFICATION AND TRAINING

- **Credit Risk Modeling in Python** – Online Training **Feb 2024**
- **Big Data Course** - Spark, Hadoop, Azure, Databricks, Coursera Certificate **Sep 2021 – Dec 2021**
- **Data Science Boot Camp** - IBM, Coursera Professional Certificate **Nov 2019**
- **Certificate of SAS Programming: Advanced** **Jul 2019**