Kaibo Zhang

+1 (438)8650693 │ kb.zh@outlook.com │ 1188 Saint Antoine O H3C1B4 │ [LinkedIn](#) │ [Personal Website](#) │ [GitHub](#)

EDUCATION

MCGILL UNIVERSITY — Montreal, Canada
*B.Com. in **Business Analytics**, Desautels Faculty of Management* — Aug. 2022 – May 2025 (Expected)

- *Minor in **Statistics**, Department of Mathematics and Statistics* — GPA: **3.96/4**
- Coursework: ETL Workflow, Data Mining, Optimization, Data Structure, Machine Learning, Statistical Modelling, Simulation, Time Series Forecasting, Linear Algebra, Multivariate Calculus, Probability, etc.
- Scholarships and Awards: Dean's Honour List – McGill (top 10%), Accenture Prize in Management ($1,590)

PROFESSIONAL EXPERIENCE

Research Assistant at Professor Kyunghee Lee's Lab — Montreal, Canada
ML/DL Model for Patent and Tech Company in the U.S. — Jan. 2024 – Now

- Constructed a Regression model based on patent data and the corresponding vector space from the extracted coefficients to realize mapping firms' relative spatial relationships for comparison analysis.
- Collected patents and tech companies' data to create Entity Embedding and built a vector database and 2 ETL pipelines for data preparation and model building.
- Coded pre-processing Python scripts for data cleaning and joining individual patent dataframes, aggregating company names as unique IDs and storing patent-holding information in dictionaries.
- Collaborated with Professor Lee to research and design novel validation methods, boosting the efficiency and effectiveness of the resulting vector space.

Machine Learning and AI Experimental Lab Intern — Shenzhen, China
Bosera Asset Management Co. Limited — May. 2024 – Aug. 2024

- Designed and implemented relational database and APIs with SQLAlchemy and FastAPI to facilitate efficient LLM data retrieval, achieving 85% critical information retrieval accuracy and outperforming sample responses.
- Researched and enhanced Retrieval Augmented Generation performance with Landmark Embedding and tree-organized Natural Language Retrieval workflow, improving semantic and answering accuracy by 45%.
- Integrated K-Means algorithms into Multi-Agent workflow using Python to extract and group performance metrics from research reports to compare companies' performance in different time horizons with LLM.
- Collaborated with 4 senior Data and AI Engineers in drafting AI and Agent Engineering Cookbook and leading 12 peers for task arrangements and progress communications.

Data Analyst Intern — Remote
VariFlight Technology Ltd — Sept. 2023 – Dec. 2023

- Utilized Mixed-Integer Optimization and sensitivity analysis on resource constraints to discover revenue management insights for airlines and airports, projecting a 2-4% expected increase in simulated profit.
- Designed 3 MySQL relational databases and built 4 ETL pipelines with Python to load API data into databases for model training and deployment, extracting, managing, and cleaning data for weekly report publications.
- Implemented interactive features for 2 existing visualization dashboards with Power BI and Python, simplifying graphs and enhancing readability.

Data Analyst Intern at Online Ride-hailing Area (Urban Central Platform) — Tianjin, China
Beijing DiDi Infinity Technology and Development Co. — May. 2023 – Aug. 2023

- Built data automatic processing models that execute SQL queries to update datasets from source data at preset times, resulting in a 30% increase in efficiency.
- Cleaned 5 datasets with SQL and built data visualization and Regression models with Excel for online ride-hailing operational needs, unifying data formats and helping to put 13% of stale data into use.
- Led and delegated tasks for colleagues to call online ride-hailing to collect first-hand data manually and used hypothesis testing for measuring price-related market standings with competing brands.

PROJECTS AND COURSEWORK

Develop Test Cases and Redesign Content for INSY 446 Data Mining for Business Analytics
Machine Learning | Python | Course Coordination

- Redesigned course content to combine theoretical mathematics behind algorithms and analytical applications.
- Developed test case functions in Python for automatic grading and instant feedback for assignments and exams.

Traffic Prediction and Optimization of Bike-sharing System in New York City — [GitHub](#)
Time Series | Bayes Rules | Network Analysis | KNN | Regression | Optimization

- Developed a Regression model to analyze bike-sharing demand dynamics, building on prior HKUST research.
- Utilized the Adjacency Matrix to capture relationships between individual stations and applied a KNN model to group stations into 3 clusters to enhance result generalizability.
- Conducted Time Series forecasting using Conditional Probabilities to predict people's bike travel habits.
- Implemented a Linear Programming model to optimize resource allocation for the bike-sharing system in NYC.

Yelp Star Rating Initiative [GitHub]
Sentiment Analysis | KNN | Chi-Sqaured Test | Logistic Regression
- Utilized Machine Learning and Sentiment Analysis to analyze the Yelp dataset, providing a roadmap for affiliated businesses who desired to improve their ratings on the platform.
- Implemented a KNN model to cluster businesses, improving interpretability and insight generation.
- Conducted a Chi-Squared test and leveraged Logistic Regression to confirm generalizability across continents and extract actionable strategic insights.

LinkedIn Database Replication Design [GitHub]
Relational Database Design | DDL | DML | MySQL
- Conducted a comprehensive analysis of LinkedIn's business model to replicate its underlying database and replicated 85% of its original functionalities.
- Developed conceptual, relational, and physical database models using a top-down approach with MySQL.
- Crafted 20 advanced SQL queries to demonstrate the database's capability to support complex user operations.

Social Speculation for Harnessing Reddit to Forecast Bitcoin Fluctuations [GitHub]
ETL | API | NLP | MLP
- Built two ETL workflows with Python using Alpha Vantage and Reddit APIs to gather and store relevant data.
- Transformed textual comment data into sentiment and subjectivity scores with TextBlob for deeper analysis.
- Trained an MLP model using Pytorch to predict Bitcoin close price, improving baseline MSE by 22.62%.

Regression Analysis on the Factors Affecting the Monthly Spending of Desautels Students [GitHub]
Data Collection | Exploratory Analysis | Multiple Regression
- Conducted a statistical analysis to identify factors influencing students' monthly spending for improved financial management.
- Collected more than 300 records of first-hand data using a self-administered questionnaire on Qualtrics.
- Performed data preprocessing, exploratory analysis, hypothesis testing, and multiple regression with Excel and concluded that dine-out, groceries, and subscriptions were the significant factors influencing monthly spending.

LEADERSHIP AND EXTRACURRICULAR

Teaching Assistant Montreal, Canada
MGCR 271 Business Statistics &. MGSC 404 Founds of Decision Analytics Aug. 2023 – Now
- Prepared sample final exam solution explanations, helping students apply different probability distribution models under the right context.
- Held tutorials and graded assignments and exams, innovating an intuitive way to explain the Z-score notion by making analogies with the unit transformation between meters and centimeters.
- Streamlined communication between students and the professor with an online forum, efficiently managed inquiries, and resolved issues, saving supervisors significant time and effort in administrative tasks.
- Scrutinized and modified existing final project instructions to avoid misunderstandings about using Hypothesis Testing and Multiple Regression.

VP of Finance Montreal, Canada
McGill Data Network May. 2023 – Now
- Spearheaded the development and management of the club's annual budget, ensuring financial sustainability and responsible allocation of resources for cost-saving and expenditure optimization.
- Maintained accurate financial records and facilitated data-driven decision-making for future projects and activities.
- Collaborated with 7 members to plan and hold panel discussions in Analytics and other interactive teaching events for basic analytics techniques to bridge the gap between the student body and the Analytics industry.

SKILLS&ACTIVITIES

- Technical Skills: MySQL, Neo4j, R, Python (Pandas, Sktlearn, Requests, Matplotlib, etc.), Minitab, Git, JavaScript, Java, HTML, CSS, Canva, Figma, Microsoft Office Suites
- Language Skills: English and Mandarin
- Interests: Photography, Guitar, Auto-saxophone, and Piano