

CHAO GAO

300 Worchester Lane, Allen, TX 75002

✉ chaogao686@gmail.com

🌐 [linkedin.com/in/chaogao](https://www.linkedin.com/in/chaogao)

🐙 github.com/Chaoccz

🌐 [My Portfolio](#)

EDUCATION

Boston University

Sep 2022 – May 2023

Master of Science in Applied Data Analytics

Boston, MA

- Coursework: Machine Learning, Data Mining, Big Data Analytics and Database Management

University of Victoria

Sep 2018 – May 2022

Bachelor of Science in Mathematics and Statistics

Victoria, Canada

- Coursework: Multivariate Analysis, Time Series Analysis, Generalized Linear Models and Survival Analysis

EXPERIENCE

Walmart

May 2023 – Present

Data Engineer

Dallas, Texas

- **Data Transformation and Integration:** Extract data from identified databases. Creates data pipelines and transform data to a structure that is relevant to the problem by selecting appropriate techniques. Develops knowledge of current analytics trends by performing Exploratory Data Analysis.
- **Data Modeling:** Develops the Logical Data Model and Physical Data Models including data warehouse and data mart designs. Defines relational tables, primary and foreign keys, and stored procedures to create a data model structure. Analyses data-related system integration challenges and proposes appropriate solutions.
- **Code Development and Testing:** Writes code to develop the required solution and application features. Creates test cases to review and validate the proposed solution design. Contributes code documentation, maintains playbook, and provides timely progress updates.

Bank of China

May 2022 – Aug 2022

Operation Analyst Intern

Beijing, China

- **Formulate Flow Strategy:** Participated in setting up A/B test for different flow strategies. Cleaned and analyzed video performance data using R, collaborating with 3 data scientists to conduct traffic strategy, boosted DAU by 10k.
- **Optimize Marketing Strategy:** Consolidated, cleaned, and transformed customer transaction data, optimizing strategic marketing model using python. Gained a lift of 5 basis points in ROL.
- **Build Expenditure Dashboard:** Automated computation of 20+ business metrics using MySQL and visualized patterns in an interactive Tableau dashboard, reduced manual reporting efforts by 60%.

Peking Union Medical College Hospital

Sep 2021 – Dec 2021

Data Analyst Intern

Beijing, China

- **Support Keyword Optimization:** Extracted and analyzed million-level clickstream data to identify top search keywords for the e-commerce platform using Google BigQuery, increasing conversion rate by 3.5%.
- **Predictive Modelling:** Provided exploratory data analysis with sale predictions using Machine Learning. Realized 0.93 in testing performance scores, and projected 30%+ boost in optimal sales.
- **Time Series Analysis:** Analyzed the multivariate sale and promotion lags using the VAR model. Automated recommender functions that generate optimal promotion strategies. Projected 40% decrease in advertising costs.

PROJECTS

Survival Analysis of Heart Failure Patients | R, MySQL, Weka

Jan 2023 – Apr 2023

- Cleaned 75k+ patient records using MySQL and summarized information into statistics to gain meaningful insights.
- Implemented Logistics regression, the Kaplan Meier model, the Cox Proportional-Hazards model and Random forest to predict future heart failure cases.
- Compiled relevant performance metrics and leveraged retrospectives to continuously improve.

Weather Data ETL Pipeline Automation | Python, Airflow, AWS EC2

Sep 2022 – Dec 2022

- Extracted current weather data from open weather map API, transformed the data and loaded the data into an S3 bucket using Apache Airflow.
- Developed Airflow DAGs to orchestrate ETL workflows, using PythonOperators for task execution and dependencies.
- Configured the scheduler for automatic and scheduled DAG execution. Set up monitoring tools and logging mechanisms to track pipeline health and performance.

SKILLS SUMMARY

Programming Languages and Platform: Python, Java, R, SQL, MATLAB, Google Cloud, Spark, MongoDB

Data Analytics: Database Management, Big Data, Data Cleaning & Modelling & Visualization, Query Optimization

Tools: Tableau, PowerBI, Excel, Docker, Airflow, Databricks, AWS EC2, Microservices Architecture, Google BigQuery