

WANYING (WINNIE) ZHANG

📞 909-699-2006 ✉️ wanying_zhang@brown.edu [in](#) [Linkedin](#) [G](#) [Github](#)

Education

Brown University

Sep. 2023 – May 2025 (Expected)

Master of Science in Data Science

Providence, RI

- **Cumulative GPA:** 4.0/4.0
- **Relevant Coursework:** Database Management & Big Data, Computational Probability and Statistics, Machine Learning, Deep Learning

University of California, Davis

Sep. 2019 – Dec. 2022

Bachelor of Science in Applied Mathematics, minor in Computer Science

Davis, CA

- **Cumulative GPA:** 3.88/4.0
- **Awards:** McCurdy Family Scholarship, Dean's Honors List, Graduated with Honors
- **Relevant Coursework:** Intro to Programming and Data Structure, Computer Vision, Math for Data Analytics, Linear Algebra, Numerical Analysis

Internship Experience

Brown University

Sep. 2023 – Present

Data Science Research Assistant

Providence, RI

- Employed **R** for advanced **statistical analysis** on healthcare datasets with **100+** features to explore the correlations between health complications arising from visceral fat, dietary habits, and various biological variables.
- Applied a range of statistical techniques (Multiple Linear Regression, Robust Linear Regression, Logistic Regression, etc) to uncover significant patterns and relationships within the data.

HireBeat

June 2022 – August 2022

Business Analytics Intern

Jersey City, New Jersey

- Processed **20K+** records from the US Department of Labor, focusing on data aggregation; constructed 3 comprehensive master tables, enhancing data organization and accessibility.
- Executed a detailed **competitor analysis** using Ubersuggest to identify key hiring keywords and referral links; led to the enhancement of internal hiring products by integrating important features and elevated product competitiveness.
- Designed **5+ OKR metrics** (impressions and CTR) to evaluate website user engagement performance; leveraged **Google Search Console** to visualize both short and long term website performance decomposed by country and query dimensions; finally led to a **10%** increase in revenue.
- Utilized **SQL** to extract fraudulent users from internal database based on active time range & email format and remove it to secure business & employer information.

Related Projects

Uber & Lyft Ride Data Analysis & Ride-hailing Prices Prediction | [Project Link](#)

Sep. 2023 - Dec. 2023

- Conducted extensive data preprocessing on **670K+** ride trip records from Kaggle, includes addressing missing values using Iterative Imputer and Reduced Feature Models, and applying appropriate preprocessing techniques to features, such as OneHot Encoder, Ordinal Encoder, and StandardScaler.
- Developed an end-to-end ML pipeline in **Python** utilizing regression models, including **Linear Regression with Regularization, KNN, SVR, Random Forest, and XGBoost** to predict ride trip prices.
- Optimized model parameters and achieved a **97%** out-of-sample testing **R² score** with the Random Forest Model; led to an **80%** reduction in **RMSE Score** compared to the baseline model.
- Conducted thorough analysis of model performance and feature significance. Utilized a variety of global and local feature importance techniques to gain insights into the key drivers of model predictions.

Medical Practices SQL Project

Oct. 2023 - Nov. 2023

- Led the design and creation of a relational database for a medical practice, integrating multiple schemas to effectively manage **2K+** patients, physician, drugs, pharmacy and pharmaceutical companies records.
- Implemented complex **SQL queries** for analytical purposes, such as identifying physicians frequently prescribing drugs with adverse interactions and analyzing drug pricing strategies across various pharmacies.
- Developed and implemented **stored procedures and triggers** to streamline data retrieval and ensure patient safety through automated alerts for drug interactions

Technical Skills

Programming Languages and Tools: Python(Pandas, Numpy, Matplotlib, Scikit-learn, Seaborn, PyTorch), R, SQL, MATLAB, NoSQL(MongoDB), PySpark

Analytics Skills: Machine Learning, A/B Testing, Data Visualization (Tableau), Data Pipelines, Google Analytics, Google Search Console, Microsoft Offices(Excel, Word, PPT), Databricks

Languages: Native in Chinese(Mandarin) and Fluent in English