

Lingyi Zhao

🏠 US, Remote/Hybrid

✉ janniezhao01@gmail.com

📞 Lingyi-Zhao-01

📠 217-305-0295

WORK EXPERIENCE

Wayfair - Advanced Analytics and Algorithms Team

Boston, MA

Senior Data Analyst

Apr 2022 - Jan 2023

- Independently owned the development of dashboards that enable Product & Operations to monitor trends in OKRs and root cause issues. Partnered with users to develop solutions and drive adoption, leading to identification of 4 major opportunity areas.
- Assessed the impact to **storefront funnel metrics** when Wayfair's merchants change the primary image for their SKUs. Analyzed **clickstream data** (Python, SQL) to select appropriate analytics technique and expose insights to Product in self-serve dashboard.
- Reduced the time to identifying data quality issues from 1 week to 1 day by implementing data quality monitors for key data sources. Defined data quality checks and designed user-friendly **Slack alerts** that streamlined resolution.

W.R.Berkley

Wilmington, DE

Data Science Intern

Jun 2021 - Sep 2021

- Extracted, transformed, and conducted feature-level analysis on commercial insurance transactions. Designed and implemented a 2-layer **anomaly detection algorithm** to effectively identify potential fraud and reduce false-positive cases.
- Solely developed a **RESTful Web API** using **Flask** to enable agents and stakeholders receive real-time email notifications about abnormal activities, shorten the case resolution time from 1 week to real time.
- Analyzed 20k data from third party vendors by flagging, normalizing, and creating compound variables, built a **GLM model** to check and filter outliers, enhanced insurance pricing accuracy by 82%.

Deloitte Consulting

Shanghai, China

Data Analytics and Information Management Intern

May 2018 - Aug 2018

- Developed **conceptual and logical data models** to assist data architecture migration for a bank loan department. Partnered with engineers to deploy and host tutorial sessions to help users smoothly transfer to the new system.
- Designed new mapping logic by standardizing 2k+ metrics and 3k+ dimensions. Constructed a **relational database** using SQL to improve the capacity and efficiency of the new business intelligence system.
- Established a health-checking dashboard with primary OKRs using Tableau to speed up the decision-making process in the loan business line, avoid weekly manual reporting errors.

ANALYTICAL PROJECTS

B2W Marketplace Merchants Future Value Prediction (Capstone Project)

Mar 2021 - Present

- Extracted historical sales data from **GCP**, analyzed metrics and performed RFM analysis with K-means to obtain seller segmentations
- Engineered time series models (**ARIMA**) and machine learning models (**Regression/ Random Forest/ Gradient Boost**) to predict seller future value, and tuned hyper parameters to improve model performance
- Built a dashboard using Tableau to deliver visualized results to business partners, enabling non-technician to easily comprehend

Banking Customer Churn Prediction Analysis

Mar 2021

- Programmed algorithms to predict customer churn probability using Python and **PySpark**, helped bank to tailor marketing strategies
- Preprocessed data by filling up missing value by median, transforming categorical features and standardizing numerical variables
- Trained models (**Logistic Regression, Random Forest, KNN**) and applied regularization with optimal parameters to overcome over-fitting
- Evaluated model performance via cross-validation and analyzed feature importance to identify top factors that influenced churn rate

NLP and Topic Modeling on Amazon User Review Dataset

Feb 2021

- Separated customer reviews into 5 product groups, preprocessed text by tokenization, lemmatization, and extracted features by TFIDF
- Trained unsupervised learning models using **K-means** and **LDA** to identify latent topics and keywords for clusters
- Built supervised models by adding important features from topic results to predict Amazon's customer rating and generat recommendations

Lego Database Analysis

Oct 2020 - Dec 2020

- Constructed a relational database using extracted and normalized 300K+ Lego's historical inventory and sales data using **SQL**
- Designed a Tableau dashboard and demonstrated to business partners about how to monitor major KPIs from 8 dynamic graphs
- Innovated marketing recommendations by forecasting future price, segmenting customer groups and comparing historical and current patterns at multiple levels

EDUCATION

University of Chicago

Chicago, IL

Master of Science in Analytics

Aug 2020 - Mar 2022

- Big Data Platforms, Machine Learning, Data Mining, DS for Consulting, Data Engineering Platforms, Time Series

University of Illinois at Urbana-Champaign

Champaign, IL

Bachelor of Science in Econometrics and Quantitative (Minor in Business and Statistics)

Aug 2016 - May 2020

SKILLS

Programming: Python, R, SQL, GCP, Google Data Studio, Git, PySpark, Hive, OpenRefine, Neo4j, Azure, Tableau

Data Science: A/B Testing, Linear/Logistic Regression, XGBoost, Random forest, LDA, K-Means, KNN, PCA, Model Evaluation