

Xinyang(Oliver) Zhou

xinyangzhou2024@u.northwestern.edu | 734-510-0762 | [linkedin.com/in/oliverzhou-data-science-analytics](https://www.linkedin.com/in/oliverzhou-data-science-analytics)

EDUCATION

Northwestern University

M.S. in Machine Learning and Data Science; Cumulative GPA: 3.94/4.0

Evanston, IL

Sep 2023 - Dec 2024

- Relevant Coursework: Cloud Engineering, Text Analytics (NLP), Deep Learning, Healthcare Analytics, GenAI

University of Michigan

B.S. in Statistics, Minor in Mathematics and Business; Cumulative GPA: 3.84/4.0

Ann Arbor, MI

August 2019 - May 2023

- Scholarship/Honors: Veeam Software Endowed Scholarship in Data Analytics; James B. Angell Scholar

TECHNICAL SKILLS

Programming: Python (pandas, NumPy, scikit-learn, TensorFlow, PyTorch, PySpark, PuLP, etc.), SQL, R, SAS, C++

Data Engineering and Production Code: AWS, GCP, Azure, Snowflake, Docker, Airflow, dbt, Hadoop, Spark, Hive

Analytics and Visualization: Tableau, Power BI, ggplot2, Matplotlib, Seaborn, Looker

Platforms and Tools: GitHub, VS Code, Conda, Linux, Databricks, PostgreSQL, Jira, AMPL (optimization)

Research/TA: Ross School of Business: Finance Course Instructor; Northwestern Medicine: Research Assistant;

Harvard Business School: Research Assistant for LLM and A/B testing

WORK EXPERIENCE

Amazon

Data Science Intern - Capstone

Greater Chicago Area, IL

Sep 2024 - Dec 2024

- Analyzed fault logs through **NLP**, created **embeddings** (Word2Vec), and generated similarity scores for AWS
- Managed the exploratory analysis and **data engineering** process to generate insights from unstructured data

CDK Global

Software Engineer Intern

Greater Chicago Area, IL

Jun 2024 - Aug 2024

- Generated propensity scores by writing production-level **ML** models, enhancing **AI marketing** efficiency
- Utilized **SQL** on **Snowflake** to fix errors in the mapping algorithm, increased customer retention rates by 10%
- Autonomized the **ELT/ETL** process of dealership data through **Airflow**, reduced data quality error rate by 15%

Little City Foundation

Data Science Intern - Practicum

Greater Chicago Area, IL

Sep 2023 - Jun 2024

- Conducted **observational studies** and exploratory analysis to understand donor behavior and identify key factors influencing satisfaction levels, leading to a 25% improvement in donor engagement
- Performed **feature engineering** and developed predictive models (XGBoost, Lasso, GLM) to categorize donors and forecast donation levels. Utilized difference-in-differences methodology to estimate potential gains
- Constructed **ML pipeline** and utilized **Docker** for deployment consistency, ensuring reliability across analyses

Nationwide Mutual Insurance Company

Business Insights Analyst Intern

Columbus, OH

May 2023 - Aug 2023

- Designed **A/B tests** to identify key factors influencing service accuracy across regions, leading to a 15% productivity increase; conducted initial **power analysis** to determine test duration and sample size
- Created visual analytics in **Tableau** and **Power BI** for 41 U.S. states, supporting pricing strategy and increasing revenue by 3.7%; minimized the financial risks for operation by data analysis and **business acumen**
- Implemented complicated **SQL** queries (CTE, window function, etc.) to extract useful data from the warehouse

Michigan Institute for Data Science

Data Analyst Assistant

Ann Arbor, MI

Apr 2022 - Apr 2023

- Developed advanced **data visualizations** using **Python** libraries (Matplotlib, Seaborn, Plotly) to enhance data visibility and accelerate data mining; leveraged data transformation and augmentation techniques to improve data quality and support insightful analysis

RESEARCH PROJECT [\(GitHub Link\)](#)

Predictive Model Analysis for Profit Optimization Strategy in Telemarketing Centers of Banks

- Developed models using **logistic regression**, **neural networks**, **KNN**, **SVM**, **decision tree**, **random forest**, and **Gradient Boosting Tree** to predict the telemarketing call result, fine-tuned each model by grid-search via **cross-validation(CV)**
- Implemented **loss matrix/function** to the previous models and produced a final deliverable containing qualitative contents analysis (**Precision**, **Recall**, **F1 Score**), along with business forecasts and solutions

Spotify Song Recommendations

- Implemented dimensionality reduction models using **PCA** and elbow plot; visualized data by using **t-SNE**
- Selected the best clustering model from **K-Prototype**, **K-Means**, and **DBSCAN** based on the **Silhouette** score
- Finalized a hybrid **recommendation system** that feeds from item-based and content-based recommendations, optimizing the best weights by minimizing the RMSE; designed a user interface for song recommendations