# Xinyang(Oliver) Zhou

xinyangzhou2024@u.northwestern.edu | 734-510-0762 | linkedin.com/in/oliverzhou-data-science-analytics

## EDUCATION

**Northwestern University** **Evanston, IL**
*M.S. in Machine Learning and Data Science;* Cumulative GPA: 3.94/4.0 Sep 2023 - Dec 2024
- Relevant Coursework: Generating Business Values by Analytics, Text Analytics (NLP), Deep Learning, GenAI

**University of Michigan** **Ann Arbor, MI**
*B.S. in Statistics, Minor in Mathematics and Business* August 2019 - May 2023
- Scholarship/Honors: Veeam Software Endowed Scholarship in Data Analytics; James B. Angell Scholar

## TECHNICAL SKILLS *(GitHub Link)*

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

## WORK EXPERIENCE

**UBS** **Chicago, IL**
*Data Scientist* Feb 2025 – Present
- Oversaw anti-money laundry (**AML**) model development and optimization across multiple lines of business
- Collaborated with stakeholders to define success criteria, ensured **risk governance** by building dashboards
- Created Python scripts to automate transaction model and monthly reports, accelerated processing time by 40%

**Amazon** **Greater Chicago Area, IL**
*Data Science Intern - Capstone* Sep 2024 – Dec 2024
- Analyzed fault logs through **NLP**, created **embeddings** via **LSTM**, and generated similarity scores for AWS
- Managed the exploratory analysis and **data engineering** process to generate insights from unstructured data

**CDK Global** **Greater Chicago Area, IL**
*Software Engineer Intern* 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%
- Led discussions with **Product Managers**, demonstrating key functionalities and use cases of the product

**Little City Foundation** **Greater Chicago Area, IL**
*Data Science Intern - Practicum* Sep 2023 – Jun 2024
- Conducted **observational studies** to understand donor behavior and identify key factors influencing satisfaction levels, providing **data consulting** to clearly explain methodologies to **stakeholders**
- 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** **Columbus, OH**
*Business Insights Analyst Intern* 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

## RESEARCH EXPERIENCE

**Harvard University, Harvard Business School** **Remote**
*Research Assistant*
- Used **prompt engineering** to build a pipeline that takes in biographical information of participants to predict users' behavior through different **LLM APIs**
- Collected, cleaned, and visualized the outcomes, utilizing several approaches to enhance the **LLM performance**, including **fine-tuning,** and demonstrating the use cases for marketing purposes in the industry
- Define **business metrics** to showcase the reliability and robustness comparing to traditional ML methods

**s**

**Michigan Institute for Data Science** **Ann Arbor, MI**
*Data Analyst Assistant*
- 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