

Hanyu (Yuki) Liu

[Tableau Portfolio](#) | +1 (646)-290-1094 | hanyuliu041@gmail.com | New York, NY, 10021

WORK EXPERIENCE

TomTom - Maps Technology

New York, NY

Data Scientist Intern

Oct 2023 - Jan 2024

- Enhanced map search user satisfaction by 2.1% and accelerated weekly data delivery by 73% and by designing and developing automated tagging of 1M geographical locations using LLM BERT transformer
- Expedited data ingestion time from 1 week to 1 hour by mapping supplier data headers to company standards using machine learning classifiers (Random Forest, XGBoost), improving model performance by fixing class imbalance
- Created a web scraping tool using Selenium to capture timely updates from Open Street Map community, and wrapped it in RESTful API with front-end interface to make it easily accessible to business team, saving 40% of their labor time
- Leveraged SQL to clean and transform 10K nested JSON data, building tables to enable analytics team to conduct map quality evaluation and competitive market analysis

WelSpot - FinTech Startup

New York, NY

Data Scientist Intern

Aug 2023 - Oct 2023

- Streamlined ETL processes with Airflow, integrating data from multiple sources into Snowflake to update personal loan records, set unique identifiers to ensure integrity constraints and maintain data accuracy
- Utilized Power BI to visualize loan performance metrics with 15K data, providing real-time insights into risk profiles
- Engineered a chatbot integrating OpenAI API, enabling domain-specific Q&A and web searches by prompt engineering

IEEE - Technical Professional Organization

New York, NY

Data Science Advisor

Apr 2023 - June 2023

- Enhanced search precision in research database by improving keyword extraction from unstructured full-text articles
- Developed ETL pipeline with PySpark to retrieve 500+ articles from AWS S3 to Redshift and process on EC2, achieving a 21% reduction in batch processing time via optimizing job partitioning and caching
- Conducted text preprocessing and fine-tuned a BERT model, resulting in 13% increase in database keywords quality

NielsenIQ - Top 3 Marketing Consulting Company

Shanghai, China

Data Analyst Intern

Apr 2022 - Sep 2022

- Used Tableau to develop a monitoring platform for e-commerce clients including 30+ interactive dashboards with 100+ sheets to track key sales, price, and review metrics, launching the platform successfully within three months
- Enabled automated reporting by custom SQL queries to link PostgreSQL to Tableau, boosting 63% operational efficiency
- Enhanced dashboard functionality through direct client engagement, refining 40+ UI features, and administered user access controls to optimize the analytics solution's security and collaborative utility
- Utilized Python to analyze 30K+ sales data using descriptive and statistical methods to identify emerging trends, and formulated operational strategies aligning with current trends, incorporating factors such as price and promotion effects

Ruisheng Investment - Private Equity firm

Qingdao, China

Data Analyst Intern

Jan 2020 - Apr 2020

- Retrieved 10K historical data from Wind for back-testing and calculated win and earnings yield to analyze trading strategy
- Conducted intraday simulation trading for equity options based on market signals
- Crafted correlation matrix using Python to analyze stock index closing prices, informing subsequent stock selections

PROJECT

Substance Abuse Treatment Disparities - Merck Datathon

Apr 2023 - May 2023

- Conducted EDA (Exploratory Data Analysis) on 4K clinical data with Pandas; performed feature engineering and optimized a neural network to reach 85% prediction accuracy, identifying key factors influencing treatment outcomes
- Generated a comprehensive report, highlighting critical factors and presenting results through 2 interactive [Tableau dashboards](#); delivered actionable strategies from individual, socioeconomic, and research perspectives

EDUCATION

M.S. in Data Science | Cornell University | GPA: 3.96

Sep 2022 - Aug 2023

- Coursework: Data Science (Python & R), Data Management (SQL), Machine Learning, NLP, Statistics

B.S. in Applied Mathematics | University of Nottingham | GPA: 3.93

Sep 2018 - Jul 2022

- Dean's Scholarship for 2019, 2020, 2021

SKILLS

- Programming:** Python, SQL, R, MATLAB, SAS (Certified Programmer)
- Technical:** Tableau, AWS S3, AWS EC2, PySpark, PowerBI, Snowflake, Airflow, dbt, Git
- Packages:** Pandas, NumPy, Matplotlib, Seaborn, TensorFlow, scikit-learn, NLTK, SpaCy, LangChain, FastAPI