Ingrid Yeung

628-226-8260 | ingridyeung2027@u.northwestern.edu | LinkedIn

EDUCATION

Northwestern University **Evanston, IL**

Bachelor of Arts (Major in Computer Science, Data Science and Statistics) ***GPA: 3.8/4.0*** Expected 06/2027

Relevant Coursework: Data Structures & Algorithms, Statistical Computing, Regression Analysis, Data Visualization

PROFESSIONAL EXPERIENCE

Live Nation Entertainment **Chicago, IL**

Incoming Ticketing & Pricing Analytics Intern *01/2026 - present*

DeepManifold inc **New York, NY**

Machine Learning Engineer Intern *09/2025 - 01/2026*

- Generated embeddings for **5M+ active users**, enabling **daily batch** and **near real-time (<10min**) updates with **98% cross-ID coverage** across cookies, devices, and logins, using **BigQuery** storage.
- Implemented six advanced **ML models (Word2Vec, SimCLR/MoCo, GraphSAGE, Spectral Clustering, LDA, DIN)**, building a full training pipeline with WebSocket-based progress monitoring for near real-time feedback.
- Delivered insights through **interactive dashboards** (Tableau), reducing cohort analysis turnaround time and improving marketing budget allocation accuracy **by 20%**, directly informing executive decision-making.
- Collaborated **cross-functionally** with **engineering and product teams** to operationalize segmentation and recommendation insights, ensuring **analytical outputs** directly informed **scalable product and growth decisions**.

Pixocial | Human-First AI **Shenzhen, China**

Data Science Intern *06/2025 - 08/2025*

- Designed **an A/B test** for screen brightness optimization in **BeautyPlus (2M+ DAU)** using **BigQuery regression modeling**, achieving a **+0.3% retention lift** and driving full rollout across **iOS and Android**.
- Cleaned and processed **20 M+ Firebase user event logs**, engineered **user- and brightness-level cohorts** for experiment analysis, and reduced group variance to improve **model stability and result reliability**.
- Conducted **feature analytics** for the **AI Eraser launch**, analyzing core and guardrail metrics in **SQL/BigQuery** to reveal trade-offs between **subscription conversion and user experience**; insights informed product optimization and release strategy.
- Rebuilt **GenAI Creative dashboards** in **Looker Studio**, adding filter and portrait-conversion funnels, standardizing geo-filters, and refactoring Airflow DAG logic — reducing dashboard load time by **30 %** and improving reliability for analytics teams.

Promed Health + **Evanston, IL**

Data Analyst Intern 10/2024 - 01/2025

- Developed **an ETL lead-generation pipeline** using **SQL** to clean and integrate **1K+ provider records** from **Definitive Healthcare**, creating new data fields (region, specialty, ownership) to support segmentation and performance tracking.
- Automated **data processing and reporting workflows** in **Excel and Python**, reducing manual preparation time by **80%** and delivering **25–30 qualified leads per week** that sustained a **50%+ sales conversion rate**.
- Performed **exploratory data analysis (EDA)** and segmentation on healthcare claims using **SQL and Python (pandas, matplotlib)**; built **data visualization dashboards** and geo-heatmaps that identified underserved clinic clusters and improved **lead-targeting efficiency by 15%**.
- Collaborated **cross-functionally** with the sales team to translate analytical insights into **region-specific outreach strategies**, directly contributing to new account growth in the TN market.

PROJECT EXPERIENCE

Dynamic Pricing System 09/2025 - 12/2025

- Designed an **end-to-end dynamic pricing system**, building **a feature engineering framework** with **63+ engineered features** (temporal, spatial, market, behavioral, interaction) from **50K+ historical orders.**
- Built **a full ML pipeline** covering data ingestion, feature engineering, model training, evaluation, and deployment, achieving **R^2 = 0.881** in pricing prediction using **Random Forest**
- Developed **model ensemble system** combining demand forecasting, price elasticity modeling, and reinforcement learning, using **a weighted ensemble (30% / 30% / 40%)**.
- Performed **feature importance analysis** to identify key pricing drivers (demand–supply ratio, time-of-day, zone type), translating model insights into **actionable pricing strategy recommendations** for business stakeholders.

Analysis of Pit Stop Errors on F1 Race Results 03/2025 - 05/2025

- Processed and merged two **F1 datasets** (~9,800 race records) using **Python (pandas, NumPy)** to engineer pit stop error metrics normalized by race averages and **features such as grid position, circuit complexity, and weather** for regression modeling.
- Built and compared multiple **regression models** to evaluate **pit stop impact on race outcomes**, **improving model fit (R^2: 0.012 → 0.311)** and revealing that team-level variance dominated performance differences.

SKILLS

Programming & Tools: C, Python(Pandas, Numpy, Scikit-learn), SQL (BigQuery, MySQL), Airflow, Hive, Git, Jupyter, Flask
Analytics & Experimentation: Looker Studio, Tableau, Matplotlib, Seaborn, Plotly, A/B Testing, Regression Modeling
Activities & Interests: Fusion Dance Company, Refresh Dance Crew Manager, K-Pop concerts, running, choreography