Li Wen (Wendy) Hu

202-560-6942 | liwen881129@gmail.com | https://www.linkedin.com/in/liwenhu1999/

EDUCATION

Georgetown University, Washington, DC, USA Aug 2024 – May 2026
MS in Data Science & Analytics (GPA 3.89/4.0)

National Taiwan University, Taipei, Taiwan Aug 2018 – June 2023
BS in Social Work (GPA 3.8/4.3)

TECHNICAL SKILLS

- **Programming Language:** Python, R, SPSS, SQL, GIS, Quarto
- **Big Data & Data Processing:** Apache Spark (PySpark), ETL pipelines, AWS (S3, EC2), MySQL
- **Analytics & Modeling:** Exploratory Data Analysis, Statistical Learning, Hypothesis Testing, Machine Learning, Dimensionality Reduction , Clustering, Natural Language Processing
- **Visualization Skills**: Tableau, Power BI, Looker Studio, GIS, Quarto, MS Office

PROFESSIONAL EXPERIENCE

Data Analyst Intern, Operations **Chicago, USA (Remote)**
Vosyn June 2025 – Aug 2025

- Built a complete **Python-based ETL pipeline** to extract raw interaction logs from internal databases, clean and normalize text using regex/NLTK, engineer sentiment features, and load analytics-ready tables for reporting.
- Developed a custom **sentiment analysis module** (tokenization, stopword removal, lemmatization, lexicon scoring) that improved text quality and reduced manual preprocessing by 40%.
- Designed and deployed interactive **Power BI dashboards** connected to ETL outputs, enabling trend analysis and KPI monitoring, increasing operational efficiency by 66%.

Data Scientist / Founding Member **Taipei, Taiwan**
Data Hub Aug. 2023 – Aug 2025

- Conducted survey-based analytics on gender-friendly government services using **Pearson correlation + PCA**, identifying "environmental score" as the strongest driver of inclusivity, translating results into actionable policy recommendations.
- Led **factor analysis** on behavioral/psychological datasets in **R**, reducing manual analysis effort by **72%** through automated workflows and improving result reliability.
- Built visual storytelling outputs using **ggplot2** and delivered an interactive **Looker Studio dashboard** to communicate findings and support stakeholder decision-making.

Data Analyst Intern **Taipei, Taiwan**
Taiwan Mobile Co., Ltd. Apr. 2023 – Sep. 2023

- Developed **Selenium** web scraping tool integrated OpenAI moderation models, detecting and filtering potential malicious messages from a dataset of 5,000+ inputs, increasing content moderation efficiency by 60%.
- Applied **Natural Language Processing (NLP)** using the ckiptagger API to identify key words, entities, and frequently mentioned names in Chinese fake news, contributing to the company's anti-fraud.

PROJECTS

Reddit as a Travel Guide: Trends, Sentiment, and Preferences Sep. 2025 – Dec 2025

- Processed and analyzed **10M+ Reddit travel post** using **distributed Spark workflows**, transforming large-scale unstructured text into structured datasets for downstream analysis.
- Conducted **temporal and destination-level exploratory analysis** by aggregating activity over time, revealing **seasonal planning cycles and emerging destination trends**, including momentum patterns that preceded peak interest.
- **Applied NLP and predictive modeling** (topic modeling, sentiment analysis, engagement and trend forecasting) to generate actionable insights, achieving **AUC = 0.81** for destination trend prediction and personalized travel suggestions.

Emergency Department Patient Disposition Prediction Mar. 2025 – May 2025

- Analyzed **16,000+ emergency department records** from the 2022 NHAMCS dataset, consolidating 11 raw disposition labels into 5 clinically meaningful outcomes to model patient flow and support emergency care operations.
- Engineered and selected features across **clinical, procedural, and utilization domains**, applying **Lasso-regularized logistic regression** and **stratified 5-fold cross-validation** to manage high-dimensional, imbalanced healthcare data.
- Built and evaluated **multiclass ML models** (Logistic Regression, Random Forest, XGBoost, SVM), achieving up to **92% AUC** and **74% accuracy**, and identified that **vital signs, medications, procedures, and length of visit** were substantially more predictive than demographic factors.

Predict Urban Parking Violation Trends using Machine Learning Oct. 2024 – Dec.2024

- Conducted **large-scale exploratory analysis on 1.2M+ urban parking violation records**, uncovering spatial and temporal patterns to support enforcement planning and hotspot identification.
- Applied **clustering methods (K-Means, DBSCAN) and PCA** to segment violations by geographic and behavioral characteristics, and built **spatial visualizations** to translate patterns into actionable insights.
- Engineered 34 features and evaluated a Random Forest Classifier (64% accuracy) to validate that identified patterns carried predictive signal, reinforcing analytical findings.