

Lia Wang

Jersey City, NJ | (805)-637-7329 | rw2618@nyu.edu | [Github](#)

EDUCATION

New York University

M.S. in Data Science | GPA: 3.7/4.00

New York City, NY

September 2023 - May 2025

University of California, Santa Barbara

B.S. Statistics and Data Science | GPA: 3.88/4.00 (High Honors)

Santa Barbara, CA

September 2019 - June 2023

Relevant Courses: Reinforcement Learning | Big Data | Machine Learning (Random Forest, KNN, Neural Networks, Clustering, etc) | Natural Language Processing | Computational Cognitive Model | Database Systems Management

SKILLS

Tools & Frameworks: Git, Jupyter, MySQL, SQLite3, Hadoop, MapReduce, Spark, Power BI, Excel, Tableau

Programming Languages: SQL, Python (scikit-learn, matplotlib, tensorflow, etc), SAS, R(ggplot, dplyr, tidyverse, etc)

Analytical Skills: Data Visualizations, Data Mining, Machine Learning, NLP, Business Reporting

WORK EXPERIENCES

International Rescue Committee

January 2024 - June 2024

Data Analytics Management Intern

Elizabeth, NJ

- Designed data structures for multiple departments and visually presented using flowcharts in Visio to improve process efficiency with clearer guidance and facilitate future collaboration on data manipulation.
- Analyzed **100k+** client participation data and visualized client activities using Power BI, delivering regular reports and results to department managers with clients feedbacks and suggestions for future project plans.
- Collaborated closely with data science team and refined data entry process by leveraging SQL queries on client enrollment data (removing duplicates, standardizing data formats, etc) to ensure data accuracy for future analysis.
- Managed data quality checks and reconciled data discrepancies with Excel pivot tables and functions (INDEX-MATCH, IFERROR, etc) to streamline data validation and maintain data consistency across teams.
- Coordinated weekly meetings to develop a SQL-based classification model in MySQL that automatically assigning clients to proper programs based on predefined eligibility criteria and improving efficiency by **15%**.

China Huarong Asset Management Co., Ltd. Shanghai Branch

June 2021 - September 2021

Data Analyst Intern

Shanghai, China

- Developed SQL queries (SQLite3) to structure a database on **20,000+** property records (addresses, valuations, etc) from background investigation, elevating efficiency for data extraction by **17%** and boosting department productivity.
- Conducted data cleaning and exploratory analysis on property records with Python and designed interactive visualizations to identify high-risk areas and trends of property values.
- Organized meetings with clients to review program-related materials and resolved discrepancies in collaboration with internal teams through effective communication, ensuring expectations alignment and successful delivery of results.

PROJECTS

Probability of Default Prediction for Banca Massiccia | Python

September 2024 - December 2024

- Developed a Probability of Default prediction model for Banca Massiccia using XGBoost and achieved a ROC-AUC score of **0.85** from **1000k+** financial statement records (total assets, COGS, etc) to enhance bank loan underwriting.
- Addressed data challenges like imputing missing values and handling highly skewed data distribution, and leveraging financial metrics (e.g. ROA, Gross Profit Margin) in modeling for accurate and interpretable predictions.
- Strengthened risk-based strategies and credit decision-making for Banca Massiccia by predicting the one-year default likelihood with the model, enabling more effective future bank portfolio risk management.

Optimizing Blackjack Strategies with Reinforcement Learning | Python

January 2024 - May 2024

- Implemented Reinforcement Learning methods (Q-learning, SARSA, Monte Carlo) to optimize winning strategies in Blackjack with **900k** hands and achieved a highest winning rate of **44.32%** with Double Q-Learning algorithm.
- Revealed Reinforcement Learning's limitations in high-randomness environments by evaluating results against human intuition and proposed the AI-human collaboration for better performance in high-stakes settings.

Identifying Anomalous Transactions | Python

September 2022 - June 2023

- Developed outlier detection models with several algorithms and statistical methods (SVM, Isolation Forest, Tukey's Fences) to identify anomalous transactions in property management data across **50,000+** transaction records, improving anomaly detection by **10%**.
- Enhanced company's anomaly detection by delivering a final Isolation Forest model of **74%** accuracy after implementing feature engineering and evaluating model with the injection of artificial outliers.