Sihui (Sylvie) LYU

San Jose, CA (open to relocate) | 216-703-0651 | Homepage | LinkedIn | lyu.sihui96@gmail.com

EDUCATION

Northeastern University - Silicon Valley, *M.S. in Computer Science*, **GPA: 4.0/4.0** **08/2024 - Present**

Case Western Reserve University, *M.S. in Management*, GPA: 3.60/4.0 **08/2019 - 01/2021**

Wuhan University of Technology, *B.S. in Logistics Engineering*, GPA: 3.54/4.0 **09/2015 - 06/2019**

EXPERIENCE

Bettaway Supply Chain Services **South Plainfield, New Jersey**

Logistics Data Engineer **09/2021 - 05/2024**

Automatic Trailer Usage & Reporting System *(Python - Pandas and Numpy, SQL, ETL pipelines)*

- Designed and built scalable Python-based ETL pipelines to automate trailer usage and expense reporting, reducing manual workload by 90% and enhancing analytics accuracy for fleet management.
- Engineered data ingestion workflows from a SQL-based transportation management system using pyodbc, ensuring real-time data consistency across location, cargo, and order status datasets.
- Developed data transformation logic using Pandas and Numpy, integrating multi-source data (5+ relational tables) via SQL for downstream payment reconciliation and analytics workflows.

Analytics Infrastructure Cloud Migration *(SQL, BI tools, data visualization, data modeling, schema optimization)*

- Led the migration of analytics infrastructure from QlikView (on-premises) to QlikSense (cloud), refactoring scripts and visual components to ensure reliability, compatibility, and performance at scale.
- Created interactive data visualizations using Matplotlib and Seaborn to monitor fleet activities and identify operational inefficiencies in real time.
- Collaborated with operations, accounting, and fleet teams to develop 50+ report pages and 200+ tables/charts, updating KPIs and providing critical data support for decision-making across four departments.

PROJECTS

Balloon Flight Prediction - live demo *(Python - sklearn, Flask, REST API, relational database)* **03/2025 - 05/2025**

- Built an ML-powered data pipeline in Python to ingest and clean real-time telemetry (Windborne) and weather data via public APIs (OpenSky and OpenWeather) for balloon trajectory prediction.
- Trained and validated a Random Forest Regressor (scikit-learn) on historical flight and wind data to forecast balloon speed and direction, achieving 98% validation accuracy.
- Implemented a risk detection engine using KD-Tree nearest-neighbor search to identify potential airspace conflicts (e.g., FAA no-fly zones and military bases), enabling automated safety alerts.
- Deployed a Flask-based web application with APScheduler to perform hourly automated predictions and update a live dashboard with risk indicators and trajectories.

Blackjack-ACE - live demo *(Java, Q-learning, AI agent, OOP, unit tests)* **01/2025 - 03/2025**

- Designed and implemented a Java-based Blackjack game with MVC architecture, featuring real-time AI strategy suggestions via reinforcement learning.
- Modeled game entities using object-oriented principles, creating reusable classes (e.g., Player, Dealer, Deck).
- Integrated a Q-learning agent that simulated 1M+ self-play games, generating an optimal state-action Q-table for strategy selection.
- Achieved 100% JUnit test coverage across the program to ensure correctness and maintainability.

SKILLS

Programming & Data Processing: Python, Java, SQL, Bash/Shell scripting, PySpark, Apache Airflow

Tools & Modules: scikit-learn, TensorFlow, Git, Docker, AWS (S3, Lambda, Redshift), MySQL, Kafka, REST API