Siyu Wu

Los Angeles | siywtom@gmail.com | https://github.com/Siyu-Wu | https://www.linkedin.com/in/siyw/

EDUCATION

University of Southern California Expected May 2024
Master of Science, Analytics (Data Engineering) **GPA: 4.0/4.0**

University of Wales Trinity Saint David Aug 2022
Bachelor of Engineering, Automotive Engineering **First Class Honours**

Wuhan University of Technology Jun 2022
Bachelor of Engineering, Vehicle Engineering **GPA: 3.7/4.0**

SKILLS

Tools: Git, Tableau, Power BI, Excel, SAS, ArcGIS, JIRA **Languages:** SQL, Python (Spark, Kafka, Hadoop, Airflow), Bash, NoSQL
Databases: SQL Server, MySQL, Cassandra, MongoDB **Cloud:** AWS, Snowflake, DataBricks, Google Cloud Platform, Azure

EXPERIENCE

Apollo Auto, Analytics Engineer May 2023–Present
- Improved **SQL query** efficiency, reducing **retrieval time by 20%** and **delivering real-time** used car market analysis dashboards.
- Developed **ETL pipeline**, incorporating vehicle config and price data from **APIs and S3 Bucket**, **reducing processing time by 10%.**
- **Designed data architectures**: **data lake** for HTML files, **data warehouse** for historical vehicle data, and **databases** for market analysis.

USC Viterbi School of Engineering, Research & Teaching Assistant Jan 2023–Present
- Conducted **data modeling** for various business cases, including the creation of **logical data models**, **entity-relationship diagrams**.
- Deployed **IE model** to generate user profiles based on search queries, resulting in **20% improvement in search results relevance.**
- Constructed a solar panel market dashboard with **BigQuery and Tableau**, resulting in a **25% increase in the diversity of KPIs.**

Emily Shane Foundation, Data Analyst Jan 2023–May 2023
- Utilized **NLTK/Spacy library** to extract questionnaire data, **expanding feature sets by 40%** in structured tables.
- Managed **ETL workflows** using **Airflow** to process data from students' report cards and questionnaires.
- Created **real-time dashboards** to track student grades and **set notifications** for abnormal performance, ensuring timely intervention.

Kiana Analytics, Data and Analytics Consultant (USC Practicum) Jan 2023–May 2023
- Used **distributed computing** through **DataBricks** and **PySpark** to optimize the query of **IoT GIS** with a **25% efficiency improvement**.
- Utilized **Apache Kafka** for real-time IoT **data streaming** to enable a GIS visualization demo through **ArcGIS**.
- Implemented a **semi-supervised ML** model to classify device status and employee type, achieving **30% enhancement in F-measure.**

Tesla, Integration Engineer Jan 2022–Jun 2022
- Worked in **cross-functional collaboration** for prototype, resulting in **20% decrease in integration issues** during the buck trial.
- Managed the prototype building life cycle and maintained project timelines, achieving a **99% milestone completion rate**.
- Collaborated across diverse teams to **respond to change actions**, **reducing costs by 15%** and minimizing project timeline impact.

PROJECTS

City Transportation Data Analysis (Github Link) May 2023–Present
- Deployed ETL pipeline on **AWS EC2/GCP** using **Airflow** to integrate ride data from multiple sources with **relational schema mapping**.
- Implemented **Snowflake dynamic table** with automated refreshes, ensuring consistently updating, **reduced by 30% response time**.
- Utilized a stack that includes **Airflow, Kafka, Zookeeper, Spark, and Cassandra**, all efficiently **containerized with Docker**.

NBA Game Attendance Optimization Apr 2023–May 2023
- Utilized **web scraping** to gather external data and conducted SQL analysis in **PostgreSQL** to assess attendance patterns.
- Conducted **A/B testing**, presented to stakeholders, with the potential to **booste revenue by 15%** through optimized strategies.
- Developed and implemented **regression models** for NBA game **attendance forecasting**, achieving an **R-squared value of 0.85**.

FSAE Racing Car Tire Data Analysis (Github Link) Dec 2021–Jun 2022
- Using **PySpark and Plotly** to analyze data from **7 different tires across 49 tests**, **improving analysis efficiency by 50%.**
- Built **regression models** to understand tire behavior under different testing conditions and **developed KPIs** for tire evaluation.
- Conducted **tire selection** for the vehicle based on tire models, resulting in a **15% improvement in lap time.**