

JINZE WANG

+1(267) 324-9918 ◇ Philadelphia, PA

wjz000414@gmail.com ◇ <https://www.linkedin.com/in/jinze-wang-855498251/> ◇ Personal Website

EDUCATION

University of Pennsylvania

Candidate for Master of Science in Engineering, Data Science

Expected May 2024

GPA:3.95/4.0

ShanghaiTech University

Bachelor of Engineering in Computer Science and Technology (Minor: Finance)

July 2022

GPA:3.7/4.0(top 10%)

EXPERIENCE

Data Engineer Intern

Souscout

May 2023 - Aug 2023

New York, NY

- Scraped soccer player and club data from different websites using **selenium**, that can bypass Cloudflare check.
- Built a Player Id matching system that matched different player id from different data sources, by defining a metric to measure similarity between different players based on some player attributes.
- Called **Openai APIs** to verify imperfect matches, wrote prompts to normalize the output to the program.

Software Engineer Intern

Orka

Feb 2022 - July 2022

Shanghai, China

- Constructed **RESTful APIs** with **Django** for data management and third-party platform connection.
- Built an internal reimbursement approval platform using **Python** and third-party APIs.
- Applied **celery** to handle multiple backend tasks simultaneously, resulting in 50% response time reduction.

Software Engineer Intern

Deloitte

Aug 2021 - Nov 2021

Shanghai, China

- Participated in a project of constructing an automatic inquiry system for enterprise risk assessment, responsible for information gathering, data wrangling and system developing.
- Developed a data pipeline to automatically scrape data from websites using **selenium**, clean them using **pandas**, store them to **MySQL** database and used **python-docx** to generate reports based on these data.

PROJECTS

Food Classification Model

- Utilized **pandas** for data preprocessing and **seaborn** for visualization, caught insights from dataset using EDA.
- Tokenized text data into vectors using **BERT** tokenizer, and encoded some categorical features.
- Conducted grid search based on **pyspark** ML models, resulting in 80% reduction in processing time compared to sklearn models. Developed a **CNN** food classification model using **pytorch**, achieving a precision rate of 89%.

Uber Data Pipeline

- Stored Uber trip data on **GCP Storage**, and created a virtual machine instance using **Google Compute Engine**.
- Ran **Mage** on that VM and applied data transformation code onto it, connected it using **ssh**.
- Loaded data to **BigQuery** and built a dashboard on **looker** studio.

Soccer Fun Facts Web App

- Applied EDA to 3 overlapped dataset, built an **ER Diagram**, then made those table into 3NF, stored on **AWS**.
- Wrote complex SQL queries to extract fun facts from datasets. Utilized **expalin** function for **query optimization**.
- Built a web application with search and filter functions. Utilized **React** for frontend, **Node.js** for backend.

SKILLS

Programming Languages Tools

Python, SQL, MATLAB, C/C++, Java, LaTeX, HTML/CSS, Javascript
MySQL, Pandas, AWS, PyTorch, Git, Apache Spark, Django, React