Tay Xun Yang

+65 82184129 | gratzxy@gmail.com | Singapore Citizen | Linkedin | Portfolio

WORK EXPERIENCE

SAP Asia Pte Ltd   **Oct 2023 – Present**
Data Engineer   **Singapore**
- Designed and implemented an end-to-end data pipeline to deliver data seamlessly to the Sales team, utilising SharePoint for user requests, Azure for ETL processes and datamart management and Power BI for data visualisation and access.
- Automated data workflows and processes using Airflow, ensuring timely data availability for downstream applications.
- Developed a web scraping solution to extract relevant news from online sources, utilising ML and LLM models to identify high-impact articles, resulting in a 20% increase in lead generation.
- Collaborated with business stakeholders to gather data requirements, integrating new data points into the data mart to meet evolving business needs.
- Enhanced ETL pipeline performance by conducting regular tuning and optimisations, achieving a 35% reduction in data processing time.
- Evaluated and reviewed data vendors, identifying the most valuable sources to enhance data quality and drive business insights.
- Leveraged GENAI for translation and classification of job titles by department and seniority level, enabling more precise sales targeting and significantly increasing lead generation opportunities.

Ministry of Finance   **Nov 2022 – Dec 2022**
Data Analyst Intern   **Singapore**
- Extracted, transformed, and loaded (ETL) data from multiple sources into a centralised repository, streamlining processes and eliminating manual data handling.
- Performed descriptive analysis of Singapore Government ICT expenditure using Tableau, uncovering key trends and actionable insights

PSA Corporation Limited   **May 2022 – Oct 2022**
Data Scientist Intern   **Singapore**
- Leveraged machine learning and natural language processing (NLP) techniques to predict SQL query resource consumption, reducing inefficient queries by 60% and optimising resource utilization.
- Implemented statistical and machine learning models to detect anomalies in datasets, significantly improving data quality and reliability.

Motorist Singapore   **May 2021 – Oct 2021**
Business Analyst Intern   **Singapore**
- Created a Python-based web scraping solution to extract data from 54 automotive websites, fully automating the data collection process and eliminating manual effort.
- Optimized batch dataset processing by designing an ETL pipeline with Python and Google Cloud, achieving an 80% reduction in processing time and enhancing overall efficiency.

EDUCATION

National University of Singapore   **Singapore**
Bachelor of Science in Business Analytics Honours (Distinction)   **2020 – 2023**

Yonsei University   **South Korea**
Exchange Programme - Big Data (Distinction)   **2022 - 2023**

SKILLS

Python: BeautifulSoup, Matplotlib, NumPy, Pandas, Duckdb, Polars, Scikit-learn, Selenium, TensorFlow, Pyspark
Distributed Computing: Azure, Hadoop, GCP, Spark
Programming Skills: SQL, JavaScript, Java, C#

Data Visualization: Google Data Studio, PowerBI, Tableau, D3.js
Workflow Management: Dagster, Azure Data Factory, Google Cloud Dataflow, Airflow
Containerised Deployments: Docker

PROJECTS

Automated Validator Tool   April 2023

- Designed and implemented an automated validation tool to systematically evaluate ML models, focusing on fairness, performance, statistical validity, and transparency.
- Built a user-friendly web interface to enable seamless data upload, model selection, and validation, supported by a powerful backend engine for modeling, evaluation, and detailed report generation.

SGX Stock Data Pipeline   June 2022

- Designed and built a financial data pipeline leveraging Google Cloud Platform, BigQuery, Airflow, and Python to process and transform data, delivering actionable insights for investment-focused individuals.

Instagram Influencer Analysis   Dec 2021

- Used Selenium to scrape Instagram influencer posts and extracted key features from the textual data.
- Developed and trained a machine learning model to accurately classify posts as sponsored or non-sponsored.

CERTIFICATIONS

Deep Learning by **DeepLearning.AI**   June 2022
Financial Market Analysis by **IMFx**   June 2021