

Tay Xun Yang

+65 82184129 | gratzxy@gmail.com | Singapore Citizen | [Linkedin](#) | [Portfolio](#)

WORK EXPERIENCE

SAP Asia Pte Ltd

Oct 2023 – Present

Data Engineer

Singapore

- Scraped news sites using Python to improve customer targeting, resulting in a 25% increase in sales leads.
- Leveraged Natural Language Processing and Machine Learning to pinpoint high-impact articles, significantly boosting lead generation and driving sales growth.
- Collaborated with the Data Engineer team to migrate local MySQL databases to Microsoft Azure, ensuring 100% data integrity and laying the foundation for scalable solutions to meet future business needs.
- Automated extraction of vital business data from B2B providers like ZoomInfo using Python, leading to the elimination of manual contact sourcing per account and a significant increase in sales team efficiency.
- Streamlined ETL processes by developing Python scripts to automate the loading of weekly data into Azure Blob storage.
- Improved overall performance of the ETL pipeline by conducting regular performance tuning on data processing jobs, resulting in a 30% decrease in processing time

Ministry of Finance

Nov 2022 – Dec 2022

Data Analyst Intern

Singapore

- Extracted and transformed data from various sources and loaded it into a central repository, eliminating the need for manual data processing.
- Conducted descriptive analysis of Singapore Government ICT expenditure using Tableau, identifying key trends and actionable insights to reduce costs by 15% and improve efficiency.

PSA Corporation Limited

May 2022 – Oct 2022

Data Scientist Intern

Singapore

- Utilised machine learning and natural language processing (NLP) techniques to predict SQL query resource consumption, resulting in a 60% reduction in inefficient queries and optimizing resource allocation.
- Applied statistical and machine learning methods to identify anomalies in datasets, increasing data quality.

Motorist Singapore

May 2021 – Oct 2021

Business Analyst Intern

Singapore

- Developed web scraping script using Python to extract data from 54 automotive websites, eliminating the need for manual data collection.
- Streamlined batch mode dataset processing by building an ETL pipeline with Python and Google Cloud, resulting in a 80% reduction in processing time.

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 Job Search

June 2024

- Developed a Python tool using BeautifulSoup to scrape job listings from websites like CareerFuture.
- Utilised machine learning and NLP to accurately identify relevant job postings.
- Utilised Generative AI to generate tailored cover letters and resumes based on user experience and job descriptions.

Automated Validator Tool

April 2023

- Developed and implemented an automated validator tool that provides a comprehensive and systematic evaluation of ML models, with a focus on fairness, performance, statistical validity, and transparency
- Developed a web interface that facilitated data upload, model selection, and validation, backed by a robust backend engine for modeling, evaluation, and report generation.

SGX Stock Data Pipeline

June 2022

- Developed a financial data pipeline using Google Cloud Platform, BigQuery, Airflow, and Python to transform and deliver actionable insights for individuals keen on investing.

Instagram Influencer Analysis

Dec 2021

- Scraped Instagram influencer posts with Selenium and built features from the textual data
- Trained a Machine Learning Model to classify posts as sponsored or not sponsored post

CERTIFICATIONS

[Deep Learning](#) by DeepLearning.AI

June 2022

[Financial Market Analysis](#) by IMFx

June 2021