

LOH YEN SHEN

lohyenshen01@gmail.com | +60176792282 | Kuala Lumpur, Malaysia
<https://lohyenshen.github.io/> | github.com/lohyenshen | [Shawn Loh](#) | [LinkedIn](#)

WORKING EXPERIENCE

MoneyLion

Malaysia

Data Engineer Intern

February 2023 - Present

- Collaborated with cross-functional teams to design and develop fault-tolerant ELT pipelines in Python, handling over 100GB of data across multiple sources including internal OLTP databases (PostgreSQL, MySQL, MongoDB), API services, and S3 storage, resulting in improved data availability for analytics.
- Orchestrated ELT pipelines using Apache Airflow to automate data extraction and loading into Snowflake, reducing manual intervention by 90% and sending automated failure alerts to Slack, leading to a higher reliability rate.
- Leveraged dbt to optimize data transformation workflows, providing cleansed, transformed, and documented data sets to stakeholders, allowing accelerated data-driven decision-making.
- Provisioned and managed data infrastructure resources (AWS, Snowflake, Periscope) using Terraform, aligned with SRE's GitOps principles.

DoctorOnCall

Malaysia

Data & Tech Intern

July 2022 - October 2022

- Assisted the marketing team in price analysis by extracting 3+ competitors' product prices with multiple Puppeteer web scrapers, saving over 20+ hours of manual work weekly.
- Proposed and automated the marketing team's manual data extractions from 6 government websites with multiple Selenium web scrapers, saving over 50+ hours of manual work.
- Deployed a dengue data scraper on Google Cloud Functions, scheduled the recurring scraping process daily with Google Cloud Scheduler and stored the latest dengue data into Firebase, saving 30 minutes daily.
- Worked closely under senior software engineers' guidance on developing and maintaining DoctorOnCall's E-Commerce platform using ExpressJS and NextJS.
- Developed a landing page using HTML, CSS, JavaScript and Owl Carousel that facilitates AIA customers' appointments with DoctorOnCall's medical specialists.
- Assisted data analysts in aggregating the latest in-house stock-keeping units and analyzed campaigns' performances with Google BigQuery to drive business decisions.

TECHNICAL SKILLS

- **Data Engineering:** Python, SQL, NoSQL, Apache Airflow, data build tool (dbt), Snowflake Cloud
- **DevOps:** AWS, Terraform, Docker, Codefresh CI/CD
- **Data Science:** Numpy, Pandas, Scikit-Learn
- **Data Visualization:** Microsoft Power BI, Matplotlib, Seaborn
- **Web Development:** HTML, CSS, JavaScript, ExpressJS, Streamlit, R Shiny, Firebase, Java, Git, Github
- **Automation/Scraping:** Puppeteer, Selenium, BeautifulSoup

EDUCATION

University of Malaya

Malaysia

Bachelor of Computer Science (Data Science)

October 2020 - February 2024

- Dean's List, CGPA 3.95

CERTIFICATIONS

- [Machine Learning](#) Stanford Online
- [Introduction to Machine Learning in Production](#) DeepLearning.AI
- [Machine Learning Data Lifecycle in Production](#) DeepLearning.AI
- [Machine Learning Modeling Pipelines in Production](#) DeepLearning.AI

SOFTWARE PROJECTS

- **Divorce Rate Predictor**
 - Deployed a data product using R Shiny to forecast a couple's divorce rate.
 - Incorporated a binomial-response generalized linear model with 97.62% accuracy in forecasting divorce rates.
- **Issue Tracker System**
 - Developed a system in Java that helps software developers actively track, discuss, and resolve software applications' issues.
 - Incorporated data visualization features using R scripts that generate insightful statistical plots from queried data in MySQL.
- **Clinic Management System**
 - Designed and implemented a simple clinic management system in Java and MySQL for Poliklinik Mahmood to facilitate patients' data storage and retrieval.