**Jasper Yat Shun Lee**

Hong Kong | +852 9304 9832 | [yatshunlee@gmail.com](mailto:yatshunlee@gmail.com) | [github.com/yatshunlee](https://www.github.com/yatshunlee) | [linkedin.com/in/yatshunlee](https://www.linkedin.com/in/yatshunlee/) | [Portfolio](https://yatshunlee.super.site/)

PROFESSIONAL EXPERIENCE

**Blue Cross (Asia-Pacific) Insurance Limited, an AIA Subsidiary** **Hong Kong**

Data Analyst, Full-Time September 2024 – Present

* Automated 10+ data pipelines to conduct data cleaning and ETL on 60+ tables to Azure SQL database using Qlik Replicate and PySpark, allowing scheduled and live data refreshments to save 2 hours for manual refreshing daily.
* Migrated 10+ dashboards and 5+ datamarts from Tableau to Power BI and successfully managed access to data for sales, claims and marketing departments, resolving 100% issues ahead of the deadline in 2 months.
* Interviewed 10+ users to understand business requirements and build 10+ reports including patterns, charts and KPIs.

**ASRC at the Hong Kong Polytechnic University** **Hong Kong**

Data Scientist, Full-Time February 2021 – June 2023

* Developed and deployed an AI software by Python with YOLOv5 deep learning model, generating $1.4M revenue and achieving an F1 score of 0.93 (a 130% enhancement) via fine-tuning with 2k+ images and data augmentation.
* Improved the text extraction accuracy by 200% in an industrial project, by recommending the use of PaddleOCR, a state-of-the-art OCR model, and synthetic data generation with generative AI to train the model.
* Enhanced the performance of U2Net image segmentation model by nonlinear brightness scaling and image augmentation, ensuring its ability remains effective 99% of the times across varying lighting conditions.
* Achieved a 75% reduction in aerostructure distortion via analysis of material properties and machining parameters in a $5M project, employing A/B testing, visualization, and regression with Matplotlib, SciPy and Scikit-Learn.

EDUCATION

**City University of Hong Kong** **Hong Kong**

Master of Science in Data Science 2021 – 2023

* GPA: 3.83 (Distinction); Selected as the graduation representative of Class 2023 from 100 graduates.

**The Hong Kong Polytechnic University** **Hong Kong**

Bachelor of Engineering (Honors) in Mechanical Engineering 2015 – 2019

PERSONAL PROJECTS

**Quantitative Risk Modelling** – [Link](https://yatshunlee.super.site/projects/quantitative-risk-analysis) | Ranked 1st in class, 10+ Stars GitHub Repo NumPy, SciPy, AWS EC2

* Developed a value-at-risk model based on Quantile Regression, improving the original computational speed by 90%.

**Movie Library Website** – [Link](https://yatshunlee.super.site/projects/movie-library-website-with-recommendation-engine) | A grade project HTML, Django, SQL, SQL Server, Scikit-Learn

* Led a team of 5 to develop a website with a recommendation system, including front-end, back-end and database.

**Car Park Availability Analysis** – [Link](https://yatshunlee.super.site/projects/car-park-availability-analysis-and-predictive-modelling) | Top 10% in the hackathon Pandas, Matplotlib, Scikit-Learn

* Headed a team of 4 to perform exploratory data analysis on 1m rows and train a regression model with an R2 of 0.99.

**NumPy Financial Internal Rate of Return (IRR)** – [Link](https://github.com/numpy/numpy-financial/pull/99) | Open Source Contribution NumPy

* Replaced the single-root solver with a multi-roots solver by computing the eigenvalues of the companion matrix.

**LendingClub Loan Analysis** – [Link](https://yatshunlee.super.site/projects/lendingclub-loan-analysis) Pandas, AWS RDS (MySQL), Power BI, Scikit-Learn, MLflow

* Modelled the probability of default for credit risk analysis, materially boosting ROI of a portfolio by 58%.

SKILLS

**Programming**: Python (NumPy, Pandas, Tensorflow, PyTorch), SQL **Languages**: Cantonese, English, Mandarin

**Technology**: Power BI, Tableau, Linux, Git, Docker, PostgreSQL, MS SQL Server, MongoDB, AWS, GCP