**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/)

SKILLS

**Programming Languages**: Python, SQL, R Programming, C, C++, HTML, CSS, Javascript, Matlab

**Libraries**: NumPy, Pandas, PySpark, Flask, Django, Scikit-Learn, Tensorflow, Keras, PyTorch, MLflow

**Database/Cloud**:MS SQL Server, PostgreSQL, Mongo DB, AWS, S3, GCP, BigQuery, Vertex AI, Azure

**Technology**: Power BI, Tableau, Linux, Git, Docker, Hugging Face **Languages**: Cantonese, English, Chinese

PROFESSIONAL EXPERIENCE

**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.
* 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 Pandas, SciPy and Scikit-Learn.
* Delivered 30+ reports and presentations for cross-functional teams and stakeholders, effectively communicating technical findings derived from 30+ rigorous statistical analyses on 10+ datasets with Python and SQL
* Reduced time spent for retrieving 30m+ rows by 90%, by automating data pipelines with GridFS and MongoDB.

EDUCATION

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

Master of Science in Data Science 2021 – 2023

* GPA: 3.83 (Distinction); Was selected as the graduation representative of Class 2023 from 100 graduates.
* Relevant Courses: Machine Learning, CNN, RNN, LSTM, Bayesian Statistics, Numerical Optimization, Time-Series

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

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

DATA SCIENCE PROJECTS

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

* Conducted data cleaning and ETL on a large dataset with 2.9m+ rows and 141 columns, and designed a dashboard using DAX and Power Query to identify the risk factors and KPIs such as default rate, ROI and IRR for the loans.
* Modelled the probability of default for credit risk analysis, materially boosting ROI of a portfolio by 58%.

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

* Developed a value-at-risk model based on Quantile Regression, improving the original computational speed by 90%.
* Created a web app dashboard on AWS EC2, enabling real-time visualization, A/B testing and forecasting.

**NLP for Fact Checking** – [Link](https://yatshunlee.super.site/projects/llm-for-fact-checking) Requests-HTML, PyTorch, Transformers, Langchain, OpenAI API

* Developed a fact-checking system utilizing GPT 3.5, leveraging webscraping to get relevant news content and employing RAG to improve the prompt’s quality via natural language inferencing and sentence embedding by LLMs.

**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.