JUSTIN LEE

+61 412-550-307 | justinlee61@hotmail.com | Github | Linkedin
- PORTFOLIO WEBSITE -

PROFESSIONAL SUMMARY

I am a Data Science graduate from Queensland University of Technology (QUT), having completed my degree in November 2024. With a passion for data analysis and problem-solving, I specialise in extracting insights, building data-driven solutions, and creating impactful visualisations. Proficient in Python, SQL, and Tableau, I am seeking entry-level Data Analyst or Data Engineer roles where I can contribute to innovative teams and grow in the field.

WORK EXPERIENCE

Project Intern (Samsung Bixby's Algorithm Model) | Adsound, Seoul   **MAR 2021 - SEP 2021**
- Recorded and processed voice data for algorithm training, ensuring high-quality datasets.
- Coordinated with international teams to coordinate recording process and data cleaning.

Retail Assistant | Galleria, Toronto   **DEC 2018 - FEB 2020**
- Provided excellent customer service while supporting inventory management and daily operational tasks.
- Assisted a multicultural clientele, strengthening interpersonal and communication skills.

PROJECTS

Telco Churn Prediction & Analysis (Python, scikit-learn, Streamlit, SHAP). **AUG 2025 - AUG 2025**
- **Engineered** domain features and built a leakage-safe **scikit-learn** pipeline (ColumnTransformer with one-hot encoding and scaling).
- Trained and compared **Logistic Regression**, **Random Forest**, and **XGBoost** with stratified 5-fold CV, selecting the best by **PR-AUC 0.662** (**ROC-AUC 0.851**).
- **Optimised** the decision threshold via a profit curve (margin, incentive, contact cost) to **maximise** retention ROI.
- **Explained** predictions with **SHAP**, top drivers were Month to Month contracts, short tenure, Electronic Check, and lack of Tech Support, and deployed a **Streamlit** churn-probability app backed by a versioned model.pkl and a clean predict_proba_df() interface.

Healthcare ETL Pipeline & Dashboard (Python, Streamlit, PostgreSQL).   **JUN 2025 - JUL 2025**
- **Scraped** and extracted structured datasets from the **AIHW** website, automating updates from multiple web pages and document types.
- **Cleaned** and **transformed** Excel and CSV files, standardising formats for time-series analysis.
- Automated **statistical reporting** and generated visual summaries of key metrics such as health spending, hospital activity, and workforce trends.
- Built a **Streamlit dashboard** with filters by financial year, state, and demographics to enable interactive exploration of public health insights.

Blockchain ETL Pipeline (Python, PostgreSQL, Bitquery API)   **MAR 2025 - MAR 2025**
- Developed an ETL pipeline to extract, transform, and load Ethereum blockchain transactions into a **PostgreSQL** database.
- Implemented API data extraction using **GraphQL** and optimised transformation processes for structured storage.
- Designed a custom database schema to store transaction details, including sender, receiver, gas fees, and timestamps..
- Utilised **Python** (pandas, SQLAlchemy, psycopg2) for data transformation and loading.

London Bike Sharing Analysis(Python, Pandas) & Dashboard(Tableau)   JAN 2025 - FEB 2025

- Preprocessed bike-sharing data by handling missing values, renaming columns, and restructuring features for analysis.
- Conducted exploratory data analysis (EDA) using Pandas and visualised trends in bike usage.
- Created **Tableau Dashboard** to highlight key insights such as bike demand trends, weather impact, and seasonal patterns.

Data Cleansing Project (MySQL)   APR 2024 - MAY 2024

- Implemented a four-step data cleansing process: removing duplicates, standardising data formats, handling null values, and dropping unnecessary columns.
- Ensured clean, structured datasets to support accurate and efficient downstream analysis.

VOLUNTEER

Industry Officer | QUT BANDS: Business Analysis & Data Science   DEC 2023 - NOV 2024
- Brisbane

- Collaborated with executive team members to organise events connecting students with industry professionals for students.
- Managed communications and logistics to ensure smooth and successful event execution.

Secretary | QUT KSA: Korean Student Association - Brisbane   FEB 2024 - NOV 2024

- Organised and documented meeting agendas, minutes, and event run sheet.
- Coordinated logistics for student events, ensuring seamless execution and active participation.

Marketing Manager | QUT BANDS: Business Analysis & Data Science   JUL 2023 - DEC 2023
- Brisbane

- Created and promoted event content across social media platforms.
- Boosted engagement and attendance through targeted, strategic marketing campaigns.

TECHNICAL SKILLS

- ***Python*, *R*, *Matlab*, *Tableau*, *PowerBI*, *MySQL*, *SQLITE*, *Microsoft SQL Server*, *PostgreSQL*, *Mongodb*, *NoSQL*, *C#*, *Flask*, *API*, *React.js*, *Javascvript*, *Node.js*, *Express.js*, *AWS*, *EC2*, *GIT*, *Docker*, *MS Azure*, *Databricks***

EDUCATION

Bachelor of Data Science   FEB 2022 - NOV 2024
Queensland University of Technology

- Computational and Simulation Science & Data Mining and Artificial Intelligence Minor
- Society: QUTBANDS, CodeNetwork, KSA

AWARDS & ACHIEVEMENTS

- **Best Community Project - 2024 Winter Hackathon (AI Flashcard App)**
- **Volunteer - QLD AI Summit 2024 (Queensland AI Hub)**

REFEREES

- **Joseph Jung - Senior Validation and Certification Test Engineer, Tritium | Email**: jjung@tritium.com.au
- **Toni Peggrem - Chief Executive Officer, Queensland AI Hub | Email:** info@qldaihub.com
- **Sehyeong Lee - Data Analyst, Transport and Mains Road | Email:** sehyeong.v.lee@tmr.qld.gov
- **Sam Park - Managing Director, One-on-One Education Consulting | Email:** sam.oec1@gmail.com