Sally Lan

Sydney, Australia s.allylan043@gmail.com sallylan Portfolio

Technical Skills

Languages: Python, R, HTML, CSS, JavaScript, SQL

Dev Tools: Jupyter Notebook, VSCode, GitHub, MySQL, PostgreSQL, Tableau, Power BI, Excel

Technologies: Pandas, Numpy, Scikit-Learn, Matplotlib, Seaborn, Plotly, React, Django, Hugging Face, Airflow, Docker, Kafka, Spark

Professional Experience

UNICEF Australia *Spring 2025*
Data Scientist

- Built predictive model on **661,000+** donor records to predict donation actions within next 3-month windows
- Used **Python/SQL** to conduct data cleaning, manipulation, exploration, and visualization
- Performed RFM, EDA, advanced statistical analysis, and hypothesis testing to inform fundraising decisions
- Achieved **80% accuracy** on **ML models** in predicting donor behavior, informing data-driven fundraising strategies
- Collaborated cross-functionally to deliver tailored solutions, driving a **10% increase** in donations

The University of Sydney *Winter 2024*
Data Scientist

- Analyzed **15,000+ tweets** using **Python NLP / SQL** to translate findings for content moderation improvement
- Performed data manipulation, tokenization, sentiment analysis, and topic modeling to reveal hidden patterns in texts
- Achieved **95% accuracy** on **ML models** in extracting features, driving a **12% boost** in engagement
- Presented to various stakeholders with visualizations by translating findings into 3 actionable recommendations

Cardinal Blue Software Inc. (PicCollage) *Winter 2023*
Product Tester

- Performed **quality assurance** for PicCollage Apps, ensuring performance standards for **200+ million users**
- Identified **over 150 bugs** and conducted **A/B testing** on user interface to deliver high-quality performance
- Collaborated with cross-functional team of 3 engineers, UI designer, and PM to optimize the app's performance

Personal Projects

Churn Prediction and Retention Strategy *2025*
Machine Learning — High Distinction

- Analyzed **40,000+** customers records by using **R**, identifying churn drivers and **top 3000** users with high churn risk
- Built and validated **ML models** with **75% accuracy** to support data-driven retention strategies
- Segmented high-churn-risk customers into **4 distinct behavioral clusters** using **hierarchical clustering** and **PCA**, identifying top 5 churn drivers per segment to enable targeted retention strategies
- Deployed a **Shiny web app** to predict real-time churn probability for new customers

Key Drivers of Rental Perceived Quality *2025*
Machine Learning — High Distinction

- Analyzed **1,000+ property reviews** to identify key quality drivers, informing tenant engagement strategies
- Applied **GLM** to uncover 2 factors as key determinants of tenant satisfaction
- Authored a research report with 2 recommendations for marketing management and enhance satisfaction

Education

University of Sydney *2025*
MS in Data Analytics

University of New South Wales *2023*
BC in Data Analytics