Yiyun (Eloise) Xu

Seattle, WA | yiyunxu0127@gmail.com | (808)219-1747 | LinkedIn | Work Portfolio

WORK EXPERIENCE

University of Washington **Seattle, Washington**
Data Analyst | Techniques: Python, Data Pipeline, Machine Learning April 2023 – Present

- Collaborated with Finance department and improved financial data accuracy up to **98%**, utilizing **Python** for outlier detection and value correction, optimizing data quality and refining processes.
- Developed **Time Series Analysis** on payroll data using **Python** by integrating historical trends and economic factors, applied **ARIMA modeling** for precise payroll forecasting and optimized resource allocation within Food & Housing department.
- Achieved **90%** accuracy in predicting turnover rates with a **logistic regression** model, leveraging **feature engineering** with factors such as salary, job satisfaction, years of experience, and role, thereby aiding proactive budget planning.
- Enhanced revenue by **15%** through **K-Means clustering** on business data, identifying student preferences and financial behaviors, customizing offerings, and optimizing housing and meal plans for a more efficient university experience.

RSM US Alliance KMH LLP **Honolulu, Hawaii**
Assurance Data Analyst | Techniques: SQL, Snowflake, ETL, VBA August 2021 – November 2022

- Developed, validated, and implemented cross-departmental **relational databases** utilizing **Snowflake** to manage **100,000+** financial records with **80+ table structured** for the Hawaii Department of Health.
- Achieved a **25%** operational efficiency enhancement which automated **10 hours** human-operated process and improved **20%** data analysis accuracy through the strategic design **ETL data pipeline** based on **300,000+** financial records.
- Established **linear regression model** to estimate financials and incorporated dummy variables to assess the impact of COVID-19 towards the hotel company, carrying out the time-series analysis reports for the upcoming **3** years.
- Automated repetitive tasks through the implementation of **VBA** programming in Excel, performing unique calculations with a specific focus on the PPE area, achieved through the development of **custom macros**.

Verity Certified Public Accountants **Honolulu, Hawaii**
Audit Intern | Techniques: Tableau, IDEA, EXCEL July 2020 – August 2021

- Developed **10+ interactive Tableau dashboards** for consulting department to visualize and monitor financial metrics (operation expense, revenue, working capital), leading to real-time insights and enhanced decision-making.
- Leveraged **CaseWare IDEA** to analyze **100,000+** financial data, detect irregularities, and prevent fraud, resulting in a **10%** improvement in time efficiency and a **35%** increase in data accuracy through collaborative preprocessing.

SKILLS

- Programming: Python(Pandas, NumPy, scikit-learn, matplotlib, seaborn), SQL(MySQL, PostgreSQL), Spark, R, JSON, VBA
- Techniques: Machine Learning, Data pipeline(ETL), A/B Testing, LLM, Web Scraping, Statistics, Snowflake, AWS, NoSQL
- Data Visualization: Power BI, Tableau, Advanced Excel: Solver, TreePlan, Oracle Crystal Balls

EDUCATION

University of Washington – Michael G. Foster School of Business **Seattle, WA**
Master of Science in Information Systems *(STEM)* | GPA:3.71 March 2023 - May 2024
University of Hawaii at Manoa **Honolulu, HI**
Bachelor of Business Admission in Accounting and Finance | GPA:3.82 August 2016 - December 2021

TECHNICAL PROJECT EXPERIENCE

The Purdue University Data 4 Good Case Competition **September 2023 – November 2023**
Open AI LLM Training | Tools: Python, API, Llama, Git, JSON

- Optimized chat-based question results accuracy by **25%** through **Python** development for **Open-AI API** interaction, parameter adjustments, and utilization of **Large Language Model (GPT-3.5, Llama)**.
- Implemented data processing mechanism, adeptly managing **JSON data structures**, appending results to **Pandas DataFrames**, and ensuring transparent progress updates for enhanced efficiency throughout execution.

Stroke Prediction ML Model **February 2024**
Machine Learning Model | Tools: Python, Machine Learning, Statistics

- Conducted z-score statistical analysis to eliminate **2.11%** of outlier data points and utilized the resample function to augment the 'no stroke' dataset, enhancing robustness in model training.
- Delivered an **82%** accuracy and **81.7%** F1 score in stroke prediction and identified **XG Boost** as the optimal machine learning model based on comprehensive data analysis and model evaluation.