Yi-Li Chen

+1 (323)336-2382 | leochenusc@gmail.com | https://www.linkedin.com/in/yi-li-chen-leo/

PROFESSIONAL SUMMARY

Detail-oriented professional with over 5 years of hands-on data analytics experience in SQL, Python, R, and Tableau. Proficient in translating business requirements into solutions and collaborating with cross-functional teams to drive data-driven decision-making.

WORK EXPERIENCE

SmartBI LLC (getBIsmart.com), Irvine, California

Lead Analyst   *Jul 2022 – Present*
Data Analyst   *Jul 2021 – Jul 2022*

- Acted as an analytics consultant for entertainment companies, including Sony Pictures, Crunchyroll and Scopely
- Optimized the ETL process for multiple data sources and **90M+** user accounts, resulting in a **30%** increase in efficiency
- Designed KPIs and conducted cohort analysis on in-product data, leading to a **15%** increase in conversion rate
- Automated ad-hoc analytic reports in **SQL** to provide marketing performance insights for monthly ROAS
- Implemented and evaluated A/B tests with product teams to fine-tune features and achieve a **10%** increase in retention
- Developed models in **Python** to forecast CLTV and classified users using XGBoost for campaign performance
- Provided insights through interactive **Tableau** dashboards, synthesizing complex analytical tasks into stories.
- Collaborated with Marketing, Product, and Engineering teams, offering expertise for the applied analytics team
- Assisted with new hire training processes and documented onboarding program setup and learning manuals

Academia Sinica, Taipei, Taiwan   *Apr 2019 – Jul 2019*
Research Assistant (Data Analysis)

- Preprocessed noisy wireless signal data through cleaning, integration, and transformation
- Conducted **Python** and **SQL** analyses on over **1-million** data points from noisy wireless signals
- Provided quantitative, visual explanations and critical features for machine learning processes and predictions
- Utilized visualization techniques such as PCA and FastMap to project high-dimensional features onto a 2D space

EDUCATION

University of Southern California (USC), Los Angeles, CA   *Aug 2019 – May 2021*
Masters in Analytics
Relevant Coursework: Machine Learning for Data Science, Data Management, Data Mining, Optimization Methods for Analytics, Text Analytics, Integrative Analytics, Predictive Analytics

PROFESSIONAL SKILLS

- Programming: SQL, Python (*scikit-learn, pandas, numpy*), R (*dplyr, tidyverse*), JavaScript
- Computer Skills: Jupyter, PyCharm, R Studio, Git, MySQL, pgAdmin, Snowflake, DBeaver, AWS
- Visualization: MS Excel, Tableau, ggplot2, Matplotlib, Seaborn
- Expertise: NLP, Text Mining (*NLTK*), Web Scraping, Regression, Clustering, Classification, Exploratory Data Analysis, Statistical Modeling, A/B Testing, Time-Series Forecasting, Data Mining, Data Visualization

PROJECTS

A linked knowledge base of crime data (Python, SQL, Tableau)   *Fall 2019*

- Built ETL flow and integrated crime dataset with PostgreSQL; analyzed over **2 million** entries using Python
- Created data-driven solutions and visualized the results by using Seaborn and Matplotlib

Recipe Recommendation System (Python, Web scraping)   *Spring 2020*

- Constructed a recipe recommendation system enabling users to customize meals based on preferences in Python
- Scraped hundreds of thousands of recipes and reviews using Selenium and BS4
- Optimized user-based weekly meal planning system with Pyomo library, achieving **~90%** satisfaction

NLP-based comedian recommender system (Python, NLTK, TextBlob, Hypothesis Testing)   *Fall 2020*

- Recommended stand-up comedians based on script content using Machine Learning
- Implemented content-based filtering analyses on transcripts and vectorized comedians with GMM
- Applied sentiment analysis and matrix completion on comments, enhancing user preference prediction accuracy by **10%**

Airbnb Data Mining (R)   *Fall 2020*

- Analyzed and visualized **hundreds of thousands** of crime and housing data using tidyverse, ggplot, and rpart libraries
- Achieved over **85%** accuracy in predicting Airbnb prices with cluster analysis and decision trees

Hospital Performance Analysis and Visualization (SQL, Execl, Tableau)   *Spring 2021*

- Collaborated with Keck hospital, supporting ad-hoc analytics using SQL/Excel and interactive charts in Tableau
- Visualized relationships between healthcare KPIs, such as wait time and equipment utilization, and seasonality