Yi (Oliver) Ouyang

LinkedIn | ouyang2066@gmail.com | (+1) 4157430103 | Berkeley, CA | Data Scientist
Professional Data Scientist Increased customer satisfaction by **10%** through business analysis and data analysis

SKILLS

Data Science: Statistical Analysis, Machine Learning, Exploratory Data Analysis, Data Wrangling, Data Analytics, A/B Testing, Data Mining
Data Visualization: Databoarding, Tableau, Power BI, Excel, Matplotlib, Seaborn
Database: SQL, MySQL, SQL Server
Programming & Tools: Python, R, NumPy, pandas, SciPy, Scikit-Learn, AWS, Git

WORK EXPERIENCE

POIZON Inc. | E-commerce **Shanghai, China**
Data Scientist Intern, Searching Algorithms Team Mar 2024 - June 2024

- Boosted satisfaction metrics by **10%** via data-driven insights and cross-functional collaboration, directly influencing marketing intelligence efforts
- Developed an interactive, real-time Tableau dashboard for detailed feedback tracking, e.g., counting users' ratings about categories by bar chart, enhancing user satisfaction strategies for weekly stakeholder meetings
- Built robust data pipelines with Python and SQL, leveraging advanced query optimization to ensure stable and reliable data quality for dashboards while demonstrating project management and maintenance abilities

LALAMOVE | On-Demand Logistics **Shanghai, China**
Data Scientist Intern, Carpooling Team Aug 2023 - Dec 2023

- Increased driver engagement by **20%** through data analysis of driver behavior metrics, identifying key trends in Tab usage and improving process efficiency for match success
- Analyzed survey data alongside cohort and retention analysis using SQL, uncovering business insights that shaped targeted driver retention strategies.
- Developed interactive heatmaps to visualize transportation capacity through Python, boosting strategic planning efficiency by **40%** and reflecting strong interpretation skills

BOSCH | Automotive **Shanghai, China**
Data Scientist Intern, Automotive Team Feb 2023 - Aug 2023

- Created a standardized workflow for data modeling, using statistical filtering to improve accuracy and help managing projects, displaying strong listening and interpret skills in stakeholder feedback
- Used statistical methods like Gaussian filtering and thresholding to refine steady-state data, enabling clearer trend detection in experimental results and improving precision in decision-making
- Built a Tkinter-based tool to automate data cleaning, increasing efficiency by **30%** with a focus on responsibility and leadership

PROJECTS

Kaggle Competition: Fairness-Aware Post-HCT Survival Prediction Dec 2024 - Feb 2025

- Achieved a top 10% ranking (**Silver Medal**) by developing an ensemble model that improved survival prediction fairness across racial groups while maintaining a high C-index of **0.67**
- Enhanced model interpretability through statistical visualization (survival curves, hazard functions) and feature importance analysis, providing actionable insights for healthcare decision-making
- Implemented advanced survival analysis techniques, and ensemble models (LightGBM, XGBoost), to effectively handle censored medical data and enhance predictive accuracy
- Developed a fairness-focused modeling approach, using SurvivalGAN to generate synthetic survival data and reduce biases linked to race, socioeconomic status, and geography in predictions
- Actively contributed to the Kaggle community, sharing methodology improvements via discussion posts and notebooks to refine approaches for fairness-aware survival prediction

Streaming Platform Data Pipeline & Analytics Aug 2024 - Oct 2024

- Analyzed trends in video categories and view distributions, uncovering factors behind variations and predicting future audience preferences to support strategic decisions
- Scraped JSON data from a streaming platform's must-watch series using Scrapy and Xpath, bypassing anti-crawling measures to gather 12 fields from 25,104 videos over 100 weeks
- Automated data storage in MySQL and Excel using Python, implementing filtering and cleaning to ensure data quality

EDUCATION

University of California, Berkeley **Berkeley, U.S.**
Master of Analytics, IEOR (GPA: 3.86/4.0) Aug 2024 - Aug 2025
Shanghai University of Finance and Economics **Shanghai, China**
Bachelor's in Data Science and Big Data Technology (GPA: 3.77/4.0, ranked top 10%) Sep 2020 - Jun 2024