Kai-Yin (Carolyn) Chan

Los Angeles, CA | chankaiyin24@gmail.com | github.com/carolyntw | linkedin.com/in/kaiyinchan | kaiyinchan.com | (213) 618-7101

EDUCATION

University of Southern California   Los Angeles, CA
Master of Science in Applied Data Science   01/2022 – 12/2023
> Core Courses: Machine Learning (Scikit-Learn / TensorFlow), Data Mining (Spark), Database Systems (SQL / NoSQL), Research Methods and Analysis for User Studies (A/B Testing)

Fu Jen Catholic University   Taipei, Taiwan
Bachelor of Arts in Economics (Minor: Statistics and Information Science)   09/2015 – 07/2019

SKILLS

Programming Languages: Python, R, SAS, JavaScript
Data Analytics: Tableau, Power BI, Excel, Plotly, Streamlit, SQL, NoSQL
Technology: AWS, GCP, Azure Data Factory, DataBricks, Git, Jupyter Notebook, Databases (PostgreSQL, MongoDB), API

WORK EXPERIENCE

Ark Biotech (Bioprocessing startup revolutionizing lab-grown meat production)   Boston, MA
Software Engineer Intern   05/2023 – 08/2023
- Developed 5 dashboards tracking over 30 environmental metrics, enabling real-time client monitoring and delivering actionable insights for production optimization
- Extracted 100K+ time series data points from TimescaleDB using SQL and Python (Pandas) to identify contamination in cell cultures, enhancing production processes, reducing risks, and yielding $20K+ cost savings
- Engineered a real-time anomaly detection system using Databricks and TimescaleDB after identifying data inconsistencies with Python (NumPy), boosting operational agility
- Architected data models with optimized structures and types and devised an automated data integration solution in Databricks, improving data quality by 15% and meeting tight deadlines

NEC Taiwan Ltd. (Global leader in IT and network technologies, $30 billion in revenue)   Taipei, Taiwan
System Engineer   10/2020 – 04/2021
- Led pre-launch customer training sessions on system functionality and best practices, achieving a 97% user satisfaction rate and fostering strong stakeholder relationships through clear and effective communication
- Managed feature rollouts timelines and coordinated with cross-functional teams to meet milestones ahead of schedule, utilizing Agile methods and clear documentation
- Reduced system downtime by 10% through log analysis and resolution of discrepancies caused by user behavior and code bugs, enhancing system reliability

Taipei Medical University Hospital   Taipei, Taiwan
Data Science Intern (IT Department)   07/2019 – 08/2019
- Cleaned 2M+ patient records using SQL String/Window functions and Subqueries, ensuring accurate data for evaluating medication impacts on weight gain and metabolic syndrome
- Built Logistic Regression model in Python (Scikit-learn) to assess metabolic syndrome risk, offering treatment insights
- Co-developed an alert system for high-risk osteoporosis patients using Python and SQL, generating targeted messages for doctors based on patient risk factors to optimize care efficiency

First Securities Inc. (Leading financial services firm with a net profit of $21 million)   Taipei, Taiwan
Data Analyst Intern   2/2019 – 6/2019
- Spearheaded development of an R Shiny data visualization tool, enhancing SQL data retrieval and interdepartmental reporting, accelerating business intelligence workflows by 15%
- Improved decision-making by creating detailed Power BI reports with securities analysts, offering investment strategies based on market trend analysis

PROJECTS

YouTube Video Trends Analytics (Tableau / Python / AWS / Airflow)   12/2023
- Created an automated ETL data pipeline for YouTube analytics with Apache Airflow on AWS EC2, enhancing data processing and scalable storage with Amazon S3
- Designed a Tableau dashboard to analyze viewer engagement and content popularity, providing business insights and recommending strategies to increase engagement

Decoding Lifestyle's Influence on Success Metrics (Python / Statistical Interpretation / Research Design)   11/2023
- Cleaned and preprocessed data, addressing outliers and missing values to ensure reliable findings on lifestyle impacts
- Analyzed lifestyle and success correlations using Python with t-tests and regression, identified sleep hours as a key predictor, and made specific lifestyle recommendations to improve success metrics