Emily Chen

Dublin, CA | 925-623-2900 | emilyluo.chen@gmail.com | www.linkedin.com/in/emilylchen2002

EDUCATION

University of California, **San Diego – Bachelor of Science (B.S.)**   2020 - 2024
Major: Data Science, Minor: Business   Major GPA: 3.740

PROFESSIONAL EXPERIENCE

Deloitte | **Data Analyst**, Sacramento, CA   Jul. 2024 – Present
- Developed data pipelines for scalable reporting in Databricks using Python and SQL to perform data quality analysis and data reconciliation on $156B+ in transactional activity from financial systems to the general ledger.
- Implemented a Python script for data mining of 400K+ U.S. Air Force Civil Engineer Center real properties and conduct cost estimation for hazardous material cleanup and environmental disposal liabilities.
- Created Power BI dashboard to monitor expenditure of U.S. Air Force base operations and built a financial model for asset valuation, reducing balance sheet material weakness by $20.7B.
- Communicated with cross-functional teams to align practices with federal government accounting standards for audit remediation and improved financial reporting accuracy.

IBM | **Data Engineering Intern**, New York, NY   Jun. 2023 – Sept. 2023
- Constructed an ETL pipeline for automation in Apache Airflow using SQL and Python to process Salesforce engagement data and 300K+ daily product reports from Google Ads API to load into Amazon Redshift data warehouse.
- Designed a Qlik reporting dashboard to monitor campaign key performance indicators and user segmentation insights to identify marketing opportunities for potential customer bases, increasing monthly revenue by 5%.
- Collaborated with product managers and internal stakeholders to understand product goals and translate complex data to actionable, data-driven insights for decision-making.

U.S. Bank | **Data Analytics Intern**, Cincinnati, OH   Jun. 2022 – Aug. 2022
- Created Microsoft Power BI dashboard from workstation database system to monitor real time status and progress of internal upgrades, improving IT visibility and enabling secure operations for customer-facing financial technology and services.
- Optimized an internal application to assess workstation health, schedule fixes, and automate OS upgrades across 10,000+ enterprise devices, enhancing security compliance and reducing system vulnerabilities to protect sensitive financial data in regulated banking environments.

Scripps Research Institute | **Data Analytics Intern**, San Diego, CA   Jun. 2021 – Mar. 2022
- Built automated Python pipelines for data processing of demographic and survey participant data in the All of Us Research Program, improving operational efficiency by 30 minutes.
- Conducted data analysis, A/B testing, and regression models on user engagement data to identify content preferences, recommend personalization strategies, and increase subscriber interaction by 10%.

PROJECTS

Trip Destination Prediction – San Diego Association of Governments   Sept. 2023 – Mar. 2024
- Applied statistical analysis and machine learning algorithms to an Activity Based Model (ABM) for simulating resident travel pattern and travel demand forecasting.
- Implemented model pipeline orchestration for decision trees classification of 12M+ trip destinations based on individual demographic and past travel data, improving computational efficiency of SANDAG's statistical inference based ABM from 40 minutes to 5 minutes, while maintaining predictive performance.

GoFundMe Predictive Analysis – Data Science Student Society   Oct. 2020 – Mar. 2021
- Evaluated performance for Bag of Words model, TF-IDF vector representation, and BERT transformer based large language model (LLM) for Natural Language Processing (NLP) and sentiment analysis to predict achievement of GoFundMe campaign goals based on keyword patterns.
- Established an auto-text generation model for GoFundMe campaign descriptions using TensorFlow, Keras, and Long Short-Term Memory (LSTM) networks.

SKILLS & CERTIFICATIONS

Skills: Python, Java, SQL, R, Databricks, AWS S3, AWS EC2, Google Cloud Platform (GCP), Data Visualization Software (Qlik, Microsoft PowerBI, Tableau), Spark, MapReduce/Hadoop, Apache AirFlow, Kubernetes, Docker, MySQL, NoSQL, Amazon RedShift, Microsoft SQL Server, D3.js, HTML/CSS, IBM SPSS, Stata, Jupyter Notebook, PySpark, Keras, Tensorflow, PyTorch, NumPy, Pandas, Matplotlib, scikit-learn, NLTK