Thomas Lee

510-230-7619 | thomascl@berkeley.edu | linkedin.com/in/thomaslee24 | PUT GITHUB PORTFOLIO HERE

Professional Profile

Early-career data professional with two years of experience in data science and machine learning, leveraging analytical thinking and technical proficiency to deliver impactful insights. Skilled in Python and machine learning with a strong foundation in statistical analysis for data-driven decision-making and a growing portfolio of applied data science projects.

Education

University of California, Berkeley Berkeley, CA

Master of Information and Data Science | 4.0 GPA August 2025

- **Relevant Coursework:** Machine Learning Systems Engineering (Docker, Kubernetes), Natural Language Processing with Deep Learning, Research Design and Applications for Data and Analysis, Statistics for Data Science

B.A. in Computer Science, Minor in Data Science | 3.9 GPA May 2024

- **Upsilon Pi Epsilon Member**: International Computing and Information honor society (top 30% of undergraduates)

Experience

East Bay Municipal Utility District Oakland, CA

Data Science Intern May 2024 – Present

- Improved accessibility to water for **30,000** low-income families, reducing water bills for low-income families by **31%** by constructing an optimized water rate structure using SciPy optimization functions.
- Conducted cost-benefit analysis for advanced metering technology initiative projected to affect **400,000** customers, guiding long-term investment strategy and saving **$25M** per year.
- Streamlined vendor data access by implementing an Amazon Redshift-based pipeline, replacing costly **$1,500 per request** extractions with automated data streaming.
- Designed executive-facing dashboards in Power BI and Tableau to visualize rate equity and water usage trends across **1.4M** customers, reducing turnaround time for policy analysis from weeks to days.
- Managed data access and privacy workflows for company collaborations with three UC Berkeley research groups, providing mentorship with data analysis and ensuring regulatory compliance.

Eikon Therapeutics Hayward, CA

Machine Learning Intern May 2022 – Aug 2023

- Identified **20** promising drug treatment candidates from over **200,000** compounds using an extended isolation forest machine learning model for anomaly detection.
- Created a neural network architecture with TensorFlow, classifying protein agonists and antagonists with over **84%** accuracy from a highly noisy dataset.
- Engineered a Python-based ETL data pipeline to accelerate data extraction, cleaning, and preprocessing, boosting data request efficiency by **22%** for a team of **10+** scientists.

Projects

Predicting Card Power and Synergy in Pokemon Trading Card Game (TCG) April 2025

- Enabled early price estimation for newly released cards by modeling latent card strength prior to observed gameplay performance and predicted novel card combinations with high synergy to give players a competitive edge.
- Fine-tuned BERT and RoBERTa-based NLP regression models and created custom card embeddings to efficiently identify and isolate relevant information from raw card text.

San Francisco Crime Interactive Dashboard March 2025

- Designed and constructed interactive Tableau data visualizations for temporal, geographic, and trend analysis of San Francisco crime.
- Created Flask web application to embed data visualizations into a user-friendly dashboard.

Skills

Programming Languages: Python (Tensorflow, Keras, NumPyro, XGBoost, NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn), SQL (MySQL, Amazon Redshift), R, NoSQL (Neo4j)

Data Visualization and Analysis Tools: Tableau, Power BI, Dash

Other Tools: Kubernetes, Docker, AWS, Git/GitHub, HuggingFace, Visual Studio, RStudio