Toby Liu

Data Scientist | CA, NY, WA | 713-294-5792 | tobyliu2004@gmail.com | linkedin.com/in/toby-liu-b45090257

EDUCATION

B.S. of Statistics and Data Science & B.A of Economics University of California - Santa Barbara, Letters & Sciences College Honors Program, Expected June 2026

B.A of Economics University of Barcelona, Study Abroad Semester, September 2024 - December 2024

Relevant Coursework: Data Science Principles, Probability & Statistics, Linear Algebra, Design of Experiments, Econometrics, Vector Calculus, Regression Analysis, Statistical Machine Learning(Graduate Level)

TECHNICAL SKILLS

Programming Languages: Python, SQL, R, Java, VBA, Bash

Data Analysis & Visualization: Pandas, NumPy, Matplotlib, Tableau, PowerBI, Seaborn, Tidyverse, ggplot2

Databases & Querying: PostgreSQL, MySQL, AWS, Google Cloud, Microsoft Excel, Data Engineering

Machine Learning & Statistical Modeling: Scikit-learn, TensorFlow, PyTorch, XGBoost, Time Series, Clustering, Classification, Regression, A/B Testing, Causal Inference, Natural Language Processing

Tools & Software: Git, Jupyter Notebook, VSCode, Linux

EXPERIENCE

Boston Children's Hospital (Computational Biology Department), Boston, MA, Data Science Intern, July 2024 – August 2024

● Analyzed **10,000+** genomic sequences using DANPOS(Python) to assess environmental impacts on nucleosome dynamics and protein-DNA occupancy

● Designed and implemented exploratory data analysis protocols that **improved pipeline efficiency by 8%**

● Created R visualizations that highlighted statistical deviations and trends in sequencing data for the head of the Computational Biology Department at Boston Children's Hospital

Baylor College of Medicine (Bioinformatics Research Lab), Houston, TX, Student Intern, June 2023-August 2023

● Developed interactive R dashboards for Genetic Simulation Resources data, leading to a **5% increase in trend detection accuracy for predictive models**

● Engineered data pipelines using Python and R, **reducing data wrangling time by 10%**, enabling faster iteration cycles for statistical testing

PROJECTS

Diamond Pricing Analysis |Python| December 2024 - January 2025

● Built a linear regression model using Python to predict diamond prices using variables such as carat, cut, color, clarity, and dimensions, achieving **89% accuracy**

● Performed exploratory data analysis on **15,000+ records** to identify key predictors and visualize trends, assessing model performance through residual analysis, R squared values, and refinement of the model through variable transformation

ADDITIONAL

Languages: Fluent English; Conversational Mandarin

Certifications & Training: Complete SQL+Databases Bootcamp(Zero to Mastery), Machine Learning Specialization (Coursera, Andrew Ng)

Awards: 1st Place in Houston Hackathon w/ Python

Interests: Health and Wellness, Machine Learning, Healthcare Analytics, Surfing, Running, Muay Thai, Cooking