Thomas S. Lee

U.S. Citizen | 737-224-4010 | leethomas0182@gmail.com | linkedin.com/in/thomasleefin | Austin, TX

SUMMARY

Data scientist transitioning from academia with expertise in analytical modeling, data governance, and experimentation. Proven ability to translate business challenges into scalable, data-driven solutions that support decision-making and strategic goals. Experienced in large-scale data processing, model evaluation, and communicating insights through data visualization.

EDUCATION

The University of Texas at Austin — Austin, TX
Ph.D. in Finance — (Expected) Aug 2025

Korea University — Seoul, South Korea
B.B.A. | B.S. in Business Administration and Statistics — Feb 2018

EXPERIENCE

BGF Retail — Seoul, South Korea
Project Intern — Dec 2016 – Feb 2017

- Developed store classification using K-means clustering on customer sales data to support localized product optimization.
- Identified business data requirements and performed initial quality checks to ensure data suitability for modeling.
- Presented insights to senior leadership, enabling strategic decisions aligned with customer experience goals.

The University of Texas at Austin — Austin, TX
Graduate Researcher — Sep 2018 – Present

- Conducted data source identification and extraction from large, distributed datasets (e.g., 110GB+ CRSP, 160M+ longitudinal education records), ensuring fitness for purpose through extensive data quality checks.
- Applied data governance, storage, and accessibility principles to create scalable and sustainable analytical solutions.
- Designed and implemented statistical models (e.g., DiD, IV) for causal inference and experimentation to evaluate policy and investment outcomes.
- Automated data collection via web scraping and text parsing for over 1M+ bond features and 1.2M+ credit ratings.
- Provided actionable insights and communicated analytical findings effectively using Matplotlib and Plotly to technical and non-technical audiences at conferences (e.g., FIRS 2025).

Texas Education Agency — Austin, TX
Independent Researcher — Feb 2022 – Present

- Designed and implemented econometric models (e.g., 2SLS IV, DiD) to assess the impact of state-led debt relief programs, supporting policy strategy.
- Conducted large-scale data analysis on longitudinal education datasets (160M+ records), using Python and Stata to derive insights on investment outcomes and education policy.
- Applied experimentation techniques to identify impact of $1,000/pupil capital investment on math score improvements (+0.12 SD).

TECHNICAL SKILLS

- *Programming Languages:* Python, SQL, SAS, Stata, R, Excel
- *Quantitative Methods:* Statistical Modeling, Classification, Regression, Forecasting, Causal Inference (DiD, IV, RDD, PSM), A/B Testing, Experimental Design, Feature Selection, Model Evaluation (Chi-square, ROC, RMSE), Data Visualization (Matplotlib)