

# Thomas Grover

tgrover317@gmail.com • (781) 258-6832 • [linkedin.com/in/thomas-grover2](https://www.linkedin.com/in/thomas-grover2)

## EDUCATION

North Carolina State University, Raleigh, NC

May 2027

Bachelor of Science in Statistics | Minors: Data Science, Psychology, and Philosophy

GPA: 3.619 | Dean's List

## EXPERIENCE

Institute for Transportation Research and Education, Raleigh, NC

May 2025–August 2025

Data Collection Intern

- Conducted large-scale data collection across **50+ collection sites** in North Carolina, capturing thousands of observations on seatbelt usage while assuring accuracy through standardized procedures.
- Collaborated with a Ph.D. statistician on route optimization, increasing daily site coverage and efficiency, reducing travel and labor costs.
- Contributed to preliminary statistical reporting and gained practical exposure to advanced statistical methods including logistic regression and survey analysis.

Sigma Nu Fraternity, Raleigh, NC

December 2024–December 2025

President, Vice President

- Developed an alumni engagement system in SAS to separate the chapter's alumni base into industry specific lists, providing members with a networking resource to improve professional relations and support networking initiatives.
- Worked with executive leadership to manage a **\$100,000 budget**, reduced chapter debt by ~15,000 by implementing a structured repayment plan and improving fiscal discipline.
- Led executive committee operations for 100+ members, incorporating structured decision making and accountability.

## PROJECTS

Bank Marketing Classification (R)

November 2025

- Developed and evaluated a **binary classification model** using bank marketing data to predict term deposit subscription.
- Engineered features** to incorporate engagement and financial context (e.g. long call indicator, top 20% balance flag)
- Implemented stratified train and test splits to address class imbalance in the dataset (only about 12% positive rate).
- Compared GLM, Lasso, Random Forest, and XGBoost models with **cross-validation**.
- Prioritized Precision-Recall AUC instead of ROC-AUC due to imbalance in target.
- Evaluated calibration and threshold sensitivity and assessed model fairness across demographic groups.
- Validated final model performance on a held-out test dataset.

Large Language Model News Assistant (Python)

November 2025

- Fine-tuned base BERT as a four-class news topic classifier with the AG News dataset (Accuracy = 0.927, Macro F1 = 0.926).
- Built a Retrieval Augmented Generation (**RAG**) pipeline using **MiniLM** embeddings and a **FAISS** similarity index.
- Used **vector-based retrieval** and **flan-t5-small** to integrate topic classification to generate answers grounded in context to users prompts.
- Evaluated performance and documented system limitations related to retrieval quality and corpus size.

Beverage Industry Analysis (SAS)

April 2025

- Cleaned and integrated **8 raw datasets** (CSV, TXT, Access DB) into a unified dataset for analysis by applying data import features, logic, variable engineering, and transformation procedures.
- Performed statistical analysis with SAS procedures (**MEANS, FREQ**) to summarize sales and compare products across states and product categories through cross tabulation.
- Developed visualizations (**SGPLOT, SGPanel, REPORT**) to clearly illustrate trends across regions and product types.
- Generated a polished PDF report with SAS ODS, combining summaries, tables, and visualizations into a deliverable.

## SKILLS AND CERTIFICATIONS

**Certifications:** SAS Base Programming Specialist (v9.4); Hands on Training for AI/LLM Applications for Research (14 Learning Hours)

**Programming/Tools:** R; Python; SAS; GitHub; Microsoft Office; Google Workspace

**Statistical Methods:** Regression Modeling; Classification; Cross-Validation; Hypothesis Testing; Survey/Experimental Design

**Machine Learning:** Feature Engineering; Model Evaluation; Calibration; Threshold Optimization; Assessing Fairness

**AI & Natural Language Processing:** Transformer Fine Tuning; RAG; Vector Search