

# GREGORY TOMY

Gregory.Tomy@colorado.edu | 917-428-6003 | [linkedin.com/in/gregorytomy/](https://www.linkedin.com/in/gregorytomy/) | [github.com/GregoryTomy](https://github.com/GregoryTomy)

## Professional Experience

---

**State of Colorado** – Statistical Analyst III Aug 2024 – Current

- Build a Property Tax Data Warehouse in Google Cloud Platform (GCP), unifying data sources to streamline reporting and analysis.
- Develop statistical models to analyze property tax data and provide forecasts on property tax law changes.
- Write reports and communicate recommendations to property tax stakeholders.

**University of Colorado at Boulder** – Graduate Teaching Assistant Aug 2023 – May 2024

- Facilitated course content for STAT 5630 Computational Bayesian Statistics.
- Facilitated course content for MSBX 5415 Advanced Data Analytics at Leeds School of Business.

**CVS Health** – Data Scientist Intern May 2023 – Aug 2023

- Developed a business case for Aetna's smoking cessation program, projecting \$3-5M annual revenue increase.
- Utilized Google BigQuery for data cleaning and analysis of the Clinical Data Repository (CDR) and Aetna databases, identifying a target population of 666,000 smokers.
- Analyzed 6+ million patient records with 50+ features in the CDR to identify individuals at risk of heart disease.

**Brean Capital** – Financial Data Analyst July 2018 – July 2021

- Compiled SQL data and built data models for mortgaged-backed securities over \$3 billion, identifying breaches of contract and quantifying material impact.
- Led a team in forecasting financial performance for low-income credit unions, successfully securing National Credit Union Administration Secondary Capital program funding for 6 clients.
- Developed Python data cleaning pipeline and forecasted mortgage lender portfolio prices, reducing data preparation time by 98% and automating error handling with 95% accuracy.
- Built dashboards in Tableau for clients to view portfolio KPIs.

## Projects

---

### Portfolio Tracking with dbt and Snowflake:

- Utilized dbt with Snowflake to build an end-to-end data platform that transforms financial portfolio source data into usable information.

### Optimizing Discount Strategy with Causal Inference:

- Utilized synthetic control, regression discontinuity, and double/debiased machine learning to assess an e-commerce company's discounting strategy, finding a \$0.25 loss per dollar increase in discounts.
- Created a causal model for targeted discounts, recommending discounts for the top 55% of customers based on a custom score.

### OpenAI Hospital LLM Chatbot:

- Developed an OpenAI-powered Retrieval-Augmented Generation chatbot for hospital data analysis using natural language querying. Built with LangChain and deployed with FastAPI, Streamlit and GitHub Actions.

### Bayesian A/B Test Analysis for a Mobile Game:

- Evaluated the impact of changing the initial game gate from level 30 to level 40 on player retention and total games played using Bayesian analysis.
- Found a -1.3% average uplift in 1-day retention and -4.3% in 7-day retention. No increase in games played

## Education

---

**University of Colorado Boulder** Aug 2022 – May 2024

Masters in Applied Mathematics | **GPA: 4.0**

**Courses:** Machine learning, Bayesian statistics, Applied deep learning, Data mining, Causal inference.

**New York University** Aug 2014 – May 2018

Bachelors in Economics and Business Studies

## Skills

---

Python, SQL, R, TensorFlow, PyTorch, Spark, Prompt Engineering, LangChain, Google BigQuery, Google AutoML, Microsoft Azure, AWS Lambda, Docker, Flask, FastAPI, Git/GitHub Actions, Tableau, Presentations.

**Certifications:** Google TensorFlow Developer – March 2024