

Sammy Cayo

New York, NY | sacayo@berkeley.edu | (929) 210-3594
[LinkedIn/sammy-cayo](#) | [Portfolio](#) | [GitHub Repo](#)

Education

University of California, Berkeley - Master of Information and Data Science **Expected: 12/2025**
Relevant Courses: Statistics for Data Science, Fundamentals of Data Engineering, Experiments and Causal Inference, Applied Machine Learning, Generative AI

Syracuse University - Bachelor of Science, Finance **05/2018**

Technical Skills

Programming Languages: Python, R | **Databases:** SQL(PostgreSQL, BigQuery), NoSQL(Neo4j) | **Libraries:** Pandas, NumPy, PyTorch, TensorFlow, Keras, scikit-learn, Matplotlib, Seaborn, ggplot2 | **Developer Tools:** AWS, Docker, Git, GitHub, Linux CLI, Jupyter, VSCode, RStudio | **Machine Learning:** Classification, Regression, K-Means Clustering, NLP, Deep Learning | **Data Analysis:** A/B testing, Causal Inference, Statistical Analysis, Experimental Design, Inferential Statistics

Professional Experience

Chimera Securities, New York, NY **07/2019 - 08/2023**
Senior Equity Trader Associate

- Achieved a 22% return on a \$1.5 million portfolio by analyzing high-frequency, multi-dimensional streaming data and translating real-time market insights into trading decisions.
- Elevated personal KPI by 31% through a systematic analysis using historical trade data, optimizing individual risk management strategies.
- Enhanced company market intelligence by collaborating with cross-functional stakeholders to share daily insights compiled from industry research with 100+ traders, boosting firm-wide decision-making.
- Mentored 20+ junior traders by developing a data-driven performance tracking system, resulting in a 35% increase in new trader retention and improved overall team performance metrics.

Data Science Projects

Google Store Customer Return Prediction

- Improved user retention prediction recall rate from 85% to 93.72% by progressing from a baseline logistic regression model to an advanced LSTM model in TensorFlow.
- Demonstrated proficiency in large-scale data analysis by leveraging SQL (BigQuery) to process 35GB of user behaviors across 13 months, enabling collaboration on feature engineering for faster loss optimization.
- Refined model performance by fine-tuning hyperparameters such as dropout rate, neuron configurations, and learning rates, resulting in a 50,255-parameter model capable of making predictions in approximately 2 milliseconds per user.

Refugee Networks Analysis using NoSQL

- Identified Moldova as the most influential node for Ukrainian War refugee asylum seekers by implementing a graph centrality algorithm using Neo4j within an AWS Linux environment.
- Engineered an ETL (Extract, Transform, Load) process using PostgreSQL to streamline refugee data for exploratory data analysis (EDA).
- Optimized humanitarian aid distribution by identifying key transit countries for aid allocation through application of PageRank and Personalized PageRank algorithms to refugee movement data.

YouTube Engagement A/B Test

- Designed and executed an A/B test on 357 YouTube videos to analyze the causal impact of Bitcoin-related comments on video engagement (views, likes, and comments).
- Strengthened causal inference validity by implementing automated data collection via YouTube Data API and conducting intent-to-treat and non-attrition analyses in Python, interpreting treatment effects across different video categories.
- Analyzed and presented results from a study involving simulation-based power analysis and Bonferroni correction for multiple comparisons, demonstrating understanding of advanced statistical techniques.