Seungho (Samuel) Lee

slee19@cmc.edu • linkedin.com/in/samuellee19 • github.com/samuellee19

TECHNICAL SKILLS

- Python (Pandas, Polars, Scikit-learn, Keras), R, SQL (PostgreSQL/pgvector, ClickHouse, SingleStore, MySQL, Snowflake), Machine Learning/NLP (LLM, RAG, Vector Embeddings, Rerankers, LLM Evals, Clustering, Graph Analysis), AWS Stack (EC2, ECS, Lambda, S3), dbt, Apache Airflow, Prefect, Coiled, MLflow, Docker, OpenTofu, Tableau, Claude Code, Cursor, DuckDB

WORK EXPERIENCE

Data Scientist, **The Tie Inc**, New York, New York **Feb 2022 – Sep 2026**

- Built and redesigned the Snowflake-to-SingleStore pipeline decoding Polymarket and Opinion Trade smart-contract events into 23M+ hourly candlesticks across ~1M markets via an S3-Parquet architecture on Prefect/Coiled with idempotent catch-up (MERGE-based reruns) and memory-bounded hourly export cycles
- Designed a shared, three-tier data-validation framework for the data platform, covering pre-merge SQL tests in DuckDB, post-write checks, and Snowflake-to-SingleStore parity, with the candlestick pipeline as its pilot
- Fixed a bug that listed exchange contracts as top sellers in on-chain trading data, rebuilt 400M+ rows of trader history, and built a dashboard table that cut cohort and retention queries about 8x
- Authored and published a study of 5,203 digital-asset partnership announcements, using a BTC-adjusted event study with weekday-matched placebo windows to show announcements lift tweet volume, not price, for cryptocurrency tokens under $2B market cap

Grad Fellow, **Murty Sunak Quantitative and Computing Lab**, Claremont, California **Jun 2021 – Jan 2022**

- Designed, moderated, and ran workshops on Stata, R, Tableau, and other quantitative tools for over 40 undergraduate students and faculty at Claremont Colleges
- Utilized Tableau to analyze over 2 years of workshop data to find trends and areas of improvement across 6 different undergraduate segments and retention by weekly cohorts
- Developed and maintained MySQL database with 200+ million transactions for Data Science Capstone program and deployed software applications in Docker

PROJECTS

Markets Terminal and Data Platform, Personal Project **Oct 2026 – Present**

- Built a markets data platform that runs 27 scheduled jobs pulling prices, fundamentals, macro series, 57 news feeds, SEC filings, CFTC COT, and prediction-market odds, producing a citation-checked daily brief
- Launched an AI research assistant on the Claude API with hybrid keyword and vector search plus reranking, citing a source for every claim and finding 98% of hand-labeled items in its top 30 results
- Developed a point-in-time security master from 154,000+ SEC filings that tracks which company held each ticker on each date across 21,500+ entities, verifying model-extracted merger terms against filing text

EDUCATION

Claremont McKenna College – Claremont, California **May 2021**

Bachelor of Arts in Quantitative Economics with Data Science Sequence

- Relevant Coursework: Applied Microeconometrics, Big Data, Corporate Finance, Data Mining, Data Science, Game Theory, Intermediate Micro/Macroeconomics, Linear Algebra, Mergers & Acquisitions, Topics in Investments and Valuation
- Thesis: "Feature Investigation for Stock Returns Prediction Using XGBoost and Deep Learning Sentiment Classification"