

Caleb Miller

Philadelphia, PA - caleb10miller@gmail.com
calebzmiller.com – linkedin.com/in/calebzmiller - github.com/caleb10miller

Professional Summary

Data Engineer specializing in scalable Snowflake and dbt data platforms that power executive and client-facing analytics. Experienced in architecting production ELT systems and operationalizing machine learning workflows using Snowpark. Strong at translating complex business requirements into reliable, revenue-impact data solutions.

Professional Experience

NeuroFlow

Philadelphia, PA

Data Engineer

June 2025 - Present

- Lead Data Engineer for NeuroFlow’s inaugural BHIQ product, architecting the claims-based analytics platform powering a \$3M, multi-year Pay-for-Performance contract used to calculate provider incentive payments.
- Re-architected legacy entity resolution system post-acquisition of Quartet Health, replacing a multi-file Python framework with a Snowflake-native SQL solution that improved patient recall and reduced system complexity; productionized via stored procedures, secure service accounts, and automated monitoring.
- Designed and productionized Snowpark-based ML workflows in Snowflake, enabling automated training and inference within the data warehouse.
- Developed quarterly forecasting models to inform executive capacity planning and contract renewal strategy across client segments.
- Drove \$60k annual tooling cost reduction by piloting dbt Project Objects in Snowflake as a Cloud alternative.

Data Engineer Associate

June 2024 – June 2025

- Built and maintained core ELT pipelines in Snowflake using dbt, Fivetran, and RudderStack to power internal analytics and customer-facing reporting across HIPAA-regulated healthcare data.
- Architected secure client data share and reporting infrastructure for PHI datasets, provisioning Snowflake managed accounts and automating S3 exports via External Stages, Stored Procedures, and Tasks to meet enterprise SLAs.
- Stabilized production reporting infrastructure by diagnosing and correcting data model inconsistencies impacting clinical and executive decision-making.
- Developed Python-based Snowflake automation tools to create users, assign network and authentication policies, enforce MFA, and streamline platform governance.

Education

Drexel University – *M.S. Data Science*

2025

The University of Tennessee – *B.S. Business Analytics*

2022

Skills

- Data Platforms: Snowflake, dbt, Fivetran, RudderStack
- Data Engineering: ELT/ETL, Data Modeling, Warehouse Architecture, Stored Procedures, External Stages, Snowflake Tasks, Data Quality & Monitoring
- Programming: Python, SQL, Snowpark
- Machine Learning: Scikit-Learn, TensorFlow, PyTorch, XGBoost
- Governance & Infrastructure: CI/CD, Git, HIPAA Data Controls