

Peter Mangoro

New York, NY · petermangoro@gmail.com · +1 (201) 780-8979

linkedin.com/in/peter-mangoro-a98121147 · github.com/PeterMangoro · petermangoro.com

Summary

Data Engineer with a Master's degree in Data Analytics and Visualization, Yeshiva University, May 2026 and 3+ years building production data systems, APIs, and SQL-backed applications, including healthcare software environments. Designs, builds, and maintains data pipelines and ETL/ELT workflows for analytics, reporting, and machine learning; develops queries, views, and database objects for reliable BI access; and collaborates with stakeholders on high-quality, efficient data solutions. Strong in Python, SQL, Git, and cloud data platforms; experienced supporting AI/ML data preparation and model evaluation. Authorized to work in the U.S., open to relocating to Columbia, SC.

Core Skills

Data engineering: Design, build, and maintain data pipelines; ETL/ELT automation; data integration from diverse sources; monitor and troubleshoot pipeline reliability; improve performance and efficiency

SQL & databases: Relational database design; advanced SQL (PostgreSQL, Amazon Athena); queries, views, tables, and stored-procedure-style logic for reporting; data warehouses / analytical storage patterns

Cloud platforms: AWS data platforms (S3, Glue, Lambda, Athena, DynamoDB, API Gateway)—transferable to Microsoft Azure patterns (Azure SQL, Data Factory—style orchestration, cloud SQL warehouses); evaluating Azure Data Engineer pathways (DP-203)

Languages & APIs: Python (ETL, automation, data processing); API development and integration for third-party data extraction and management; Git and CI/CD-friendly workflows

BI & analytics: Collaborate on BI models and reporting; Tableau, Streamlit, QuickSight, R Shiny; data preparation for machine learning and AI workflows

Healthcare data context: Experience in healthcare software and analytics (patient/operational data workflows, HIPAA-conscious handling, peer-reviewed QI research)

Projects

AWS Serverless ETL / Analytics Pipeline : [Demo](#) | [GitHub](#) | 2025

Designed, built, and maintained end-to-end data pipelines for analytics and reporting (S3 ingest → Lambda triggers → Glue ETL → curated Parquet → Athena SQL → dashboards). Implemented automation, data-quality checks, and troubleshooting for reliable data flow; optimized storage (~75% cost reduction via Parquet).

AWS API-Driven Data Ingestion System : [Demo](#) | [GitHub](#) | 2025

Created and managed API Gateway + Lambda integrations for third-party-style data processing and extraction; automated recurring ingestion into DynamoDB/S3 and delivered operational reporting dashboards.

Large-Scale Analytical Pipeline (ML-ready) : [GitHub](#) | [Live demo](#) | 2026

Built Python pipelines over 4M records with strict validation, versioning, and documentation—data preparation and evaluation support for machine learning model comparison (F1 ~0.906, ROC-AUC ~0.966).

Hospital Readmission Analytics (Healthcare BI) : [Shiny app](#) | [GitHub](#) | 2025

Prepared multi-hospital datasets and SQL/analytical workflows for BI-style exploration; supported predictive modeling and stakeholder-facing dashboards for healthcare decision support.

Traffic Violation ETL & Reporting : [Tableau](#) | [GitHub](#) | 2025

Developed Python ETL over 2.8M+ records; structured data for consistent querying and maintained dashboards that improved accessibility of analytics across non-technical audiences.

GraphRAG Data Platform (APIs + Pipelines) : [Demo](#) | [GitHub](#) | 2026

Built ingestion-to-retrieval pipelines and API-backed services for structured/unstructured research data—autonomous ownership of technical work product from extract through serving.

Professional Experience

Software Developer (Healthcare) | SoliderMed | Jan 2023 – Aug 2024

- Built healthcare applications and data workflows handling sensitive patient/operational data with confidentiality and production-grade database design.
- Developed SQL-backed reporting, cleaning/reconciliation processes, and integrations that replaced ad hoc spreadsheets; automated recurring processes (~30% less manual work).
- Collaborated with clinical and operations stakeholders on reliable data access for day-to-day and reporting needs.

Software Developer | Propertybook | Aug 2024 – Dec 2024

- Designed SQL/PostgreSQL data layers, APIs, and reporting automation for multi-source operational data.
- Improved data capture and accessibility for business reporting (~20% lead-conversion lift; ~30% process automation).
- Partnered with non-technical stakeholders to translate requirements into maintainable data solutions.

Software Engineer | DataConexions | Jan 2022 – Dec 2024

- Owned technical work products—APIs, relational schemas, SQL access layers, and deployments—for platforms used by 10,000+ users, with significant autonomy on delivery.
- Built and maintained data integration and reporting components; documented systems and troubleshooted production data issues.
- Used Git-based development workflows and clear written/oral communication with clients and team members.

Education

M.S., Data Analytics and Visualization : Yeshiva University, New York, NY | May 2026,
Coursework in data engineering concepts, SQL, Python, ETL/analytics pipelines, BI reporting, and
AI/ML data preparation. GPA: 4.0

B.Eng., Telecommunications Engineering : National University of Science and Technology
(NUST), Zimbabwe, Related quantitative / systems foundation (Computer Science—adjacent
professional path via software and data engineering work)

Certifications

Google Data Analytics : Ask Questions to Make Data-Driven Decisions; Foundations: Data, Data,
Everywhere (Jan 2025)

Neo4j Graph Academy : Fundamentals, Cypher, Graph Data Modeling, Graph Data Science