CYRILE NNOKO MBEBI

Columbus, OH | 614-668-5133 | nnokombebicyrile@gmail.com
GitHub: github.com/cyrile-data-engineer | LinkedIn: linkedin.com/in/cyrilennokombebi

PROFESSIONAL PROFILE

Entry-Level Data Engineer with a Master of Science in Data Analytics – Data Engineering and hands-on experience building ETL pipelines, cloud data platforms, data warehouses, and scalable analytics solutions using Python, SQL, AWS, Azure, Apache Airflow, Spark, and Tableau. Passionate about delivering reliable, high-quality data solutions for business decision-making.

INFORMATION TECHNOLOGY COMPETENCIES

- **Languages:** Python, SQL, R
- **Cloud:** AWS, Microsoft Azure, Google Cloud
- **Databases:** PostgreSQL, MySQL, SQL Server, SQLite, Snowflake
- **Data Engineering:** ETL/ELT, Data Pipelines, Data Warehousing, Data Lake, Data Modeling, Apache Airflow, Spark, PySpark
- **Visualization:** Tableau, Power BI
- **Tools:** Git, GitHub, Docker

EDUCATION

- Western Governors University (WGU)
  Master of Science, Data Analytics – Data Engineering (2026)
- Udacity Nanodegree – Data Engineering with AWS
- Udacity Nanodegree – Data Engineering with Microsoft Azure

PROFESSIONAL EXPERIENCE

Data Analytics – Springboard, Apr 2022-Sept 2026

❖ **Southern Water Corp. Financial Analysis**

- Analyzed **revenue performance** across Kootha, Surjek, and Jutik desalination plants using **Microsoft Excel**.
- Evaluated **Residential, Public, and Private customer segments** and Hard/Soft water demand.
- Identified **financial trends and revenue patterns** across the three plants.
- **Generated insights to support** capacity planning, revenue optimization, and operational decisions**.**

❖ **ChemCorp Inc. Market Analysis & Product Divestment**

- Analyzed **market segments and growth opportunities** across major U.S. chemical industries using **Tableau**.
- Developed interactive **Tableau dashboards** to evaluate market share, industry performance, and product trends.
- Identified **high-growth market opportunities** and potential products for divestment.
- Generated insights to support **market-share growth, product strategy, and data-driven decision-making**.

PROJECTS EXPERIENCE

Time-Series Inventory Replenishment Pipeline (Capstone)

- Designed and implemented an end-to-end inventory pipeline to automate ETL workflows, predict stockout events, and improve supply chain decisions using Python, SQL, SQLite, Apache Airflow, Pandas, Scikit-learn, Tableau, Git
- Built ETL workflows to ingest, clean, validate, and transform 10,000+ inventory records.
- Engineered time-series features and developed Logistic Regression and Random Forest models.
- Designed relational database schema and Tableau dashboards.
- Achieved 75%+ prediction accuracy.
- Automated reporting and enabled proactive replenishment planning.

AWS Airflow Data Pipeline

- Built a cloud ETL pipeline to automate ingestion and loading of transactional data using Apache Airflow, AWS S3, Amazon Redshift, PostgreSQL, Python, SQL.
- Developed Airflow DAGs, loaded data from S3 into Redshift, and Implemented SQL transformations and data quality validation.
- Processed 50K+ records, reduced manual processing by 60%, and detected 95% of data inconsistencies.

AWS Data Lake Pipeline

- Designed a scalable AWS data lake for semi-structured data using AWS S3, AWS Glue, Amazon Athena, Python, SQL.
- Designed Raw, Trusted, and Curated layers, Built Glue ETL jobs, and Created Athena external tables.
- Processed 100K+ JSON records, Reduced transformation time by 50%, and Improved query performance by 30%.

Azure Data Engineering Pipeline

- Built a cloud-native ETL solution using Microsoft Azure using Azure Data Factory, Azure Blob Storage, Azure Databricks, PySpark, SQL
- Developed Azure Data Factory pipelines.
- Integrated Databricks for distributed transformations.
- Prepared analytics-ready datasets.
- Automated cloud ingestion, and Improved scalability and reduced manual processing.

Spark Big Data Processing

- Developed distributed data processing workflows for large datasets using Apache Spark, PySpark, Spark SQL, Python.
- Built Spark ETL pipelines, Performed joins, aggregations, and transformations, and Benchmarked Spark against Pandas.
- Processed 100K+ records, and Improved scalability and execution efficiency.

Apache Cassandra Data Modeling & ETL Pipeline

- Built an **ETL and NoSQL solution** using **Apache Cassandra, Python, Pandas, and CSV**.
- Developed a **Python ETL pipeline** to process and load music streaming data.
- Designed **query-driven Cassandra tables** using partition and clustering keys.
- Implemented efficient **INSERT and SELECT queries** without `ALLOW FILTERING`.
- Created **analytics-ready data models** for scalable song and user activity analysis.