PHILIPS RHOGUNS

Toronto, ON | orhogun@gmail.com | linkedin.com/in/philips-rhoguns-266748180 | github.com/prhoguns

Data Analyst / Junior Data Engineer | SQL, Python, PySpark, Power BI, Azure | AWS Certified DevOps

SUMMARY

Computer Science student at York University. I build data pipelines and the analysis on top of them, and I can show them running: Toronto open data loaded into a PostgreSQL warehouse with dbt and Airflow, the same pipeline rebuilt on Azure, a Kafka stream of live TTC vehicle positions, and a credit-risk model tracked in MLflow. Before this I spent two years in a 24/7 operations centre, which is where I learned to write things down and to care what happens when a job fails at 3 a.m.

SKILLS

SQL and data modelling: PostgreSQL, T-SQL (Synapse), DuckDB; window functions, CTEs, star schema / dimensional modelling, data quality tests, query tuning
ETL / ELT and orchestration: Python (pandas, PySpark, psycopg, requests), dbt, Apache Airflow, Kafka (Redpanda), Data Factory; batch and streaming
Azure and Databricks: ADLS Gen2, Data Factory, Databricks (Delta Lake, medallion), MLflow, Synapse serverless, Key Vault, Bicep
AWS and DevOps: EC2, S3, Lambda, CloudTrail, IAM; Docker Compose, GitHub Actions CI/CD, Git, pytest
BI and reporting: Power BI (star schema, DAX), Tableau, Excel, Plotly; presenting findings to non-technical stakeholders

CERTIFICATIONS

CompTIA Security+ | AWS Certified DevOps Engineer – Professional | Microsoft Certified: Power Platform Developer Associate (PL-400) | Google Cybersecurity Professional Certificate

PROJECTS

Toronto Open Data Pipeline *Python, PostgreSQL, dbt, Airflow, FastAPI*

- ELT pipeline for 450K police incidents, 45K TTC delays and census data: pulled from the City's API, bulk-loaded into PostgreSQL, modelled with dbt into a star schema with 40 data-quality tests, scheduled with Airflow. A full run takes about 40 seconds.
- Added a delay-probability model that Airflow retrains after each load and FastAPI serves with live weather.
- The model did not beat the historical-rate baseline (AUC 0.752 vs 0.749). I kept the comparison in the README because that is the honest result.

Credit Default Risk on Databricks *PySpark MLlib, MLflow*

- Credit-risk model on 30K customers: feature engineering in PySpark, logistic regression vs gradient-boosted trees with cross-validation, every run logged in MLflow with parameters, AUC, KS and feature importances. The winner is registered and a batch-scoring job loads it by alias.
- GBT reached AUC 0.78 and KS 0.43. Risk bands land within two points of actual default rates.

Azure Toronto Data Platform *Data Factory, ADLS Gen2, Databricks, Synapse, Bicep*

- Same pipeline, Azure-native. Bicep deploys the storage, Data Factory, Databricks and Synapse with managed identities and no stored keys. Bronze/silver/gold notebooks in PySpark; Synapse views over Delta for Power BI.
- Ran the notebooks on plain Spark in Docker to check them: row counts match the dbt build exactly.

TTC Real-Time Pipeline *Kafka (Redpanda), GTFS-Realtime, PostgreSQL*

- Streams live TTC vehicle positions (about 1,450 every 20 seconds) through Kafka into PostgreSQL with an idempotent sink, so replays never duplicate rows. SQL views show active vehicles per route and speed from consecutive positions.

Toronto Crime SQL Analytics *SQL, DuckDB, Plotly*

- Twenty business questions on ten years of crime data, one query each, with a written findings report and a live dashboard at prhoguns.github.io/toronto-crime-sql-analytics.

SOC Alert Analytics *SQL, scikit-learn*

- Alert-tuning and SLA analysis over 45K SIEM alerts, plus Isolation Forest anomaly detection that finds a planted lateral-movement incident volume rules miss.

EXPERIENCE

Security Operations Centre Operator, then Team Lead | York University, Toronto, ON *Jul 2022 – Jan 2025*

- Ran a shift of control room operators for a campus of 50,000+. First point of escalation; reviewed and signed off the incident reports and shift logs that went to management each month; trained new operators.

EDUCATION

Bachelor of Science, Computer Science | York University, Toronto, ON *Expected December 2026*