Marvis Osazuwa

Analytics Engineer · Data Scientist

Manchester, UK · +44 7349 949871 · marvis.osazuwa@hotmail.com · linkedin.com/in/marvisosazuwa · github.com/marz1307 · marz1307.github.io

PROFILE

MSc Data Science graduate (Salford, May 2026) who has already shipped what most graduates have only studied: production data layers end to end, with ML that explains itself. I build dbt and Dagster pipelines on Kimball-modelled warehouses, gate correctness in CI, and put applied ML (survival, gradient-boosted with SHAP, causal inference) directly on top, most recently exposed to LLM agents through a Model Context Protocol (MCP) endpoint. At Force24 I owned the data layer of a team-built customer intelligence platform; my dissertation was a separate solo build, an end-to-end churn intelligence stack designed and written alone against the same live environment. UK Graduate visa: full right to work in the UK, no sponsorship needed; EU Blue Card eligible for Germany and the EU.

EXPERIENCE

Analytics Engineer · Force24 · Leeds, UK (Hybrid) · Fixed-term contract, January 2026 - April 2026

Marketing automation SaaS. A 16-week collaborative build of a 3-tier customer intelligence platform. Within the team delivery I owned the data layer outright as sole architect and author, was primary author across 8 backend domains, and shipped the customer-facing frontend.

- Sole architect and author of a dbt and Dagster data pipeline that consolidated 4 disconnected business systems into one analytics-ready warehouse on a deduplicated account spine (over 40 dbt models, 5 daily schedules, ~93% solo commit share). Established a single canonical customer ID across the business, which closed the record-matching gaps that had made Customer Success books unjoinable.
- Cut daily pipeline compute by ~95% by re-architecting high-volume revenue models as incremental on an append-only raw layer with deterministic hash IDs. Productionised the platform on Oracle Cloud with idempotent bootstrap, TLS, VPN allowlist, and three-level operator runbooks, so it runs 24/7 at zero infrastructure cost.
- Made support attribution auditor-defensible with a 10-step deterministic classification chain that maps every ticket to exactly one account. This replaced a black-box best-guess join and closed the last open finding in the Customer Success metrics review.
- Gated the build behind 123 dbt tests and 82 pytest tests in GitHub Actions CI, which caught 22 data-quality bugs during the first production run. Silent metric drift now fails a test instead of surfacing in the boardroom.
- Shipped the customer-facing backend and Angular UI (customer list with 10 server-side filters, dashboard and reporting endpoints, sole-author Action Tracker across 7 outcome types), which gave Customer Success a single workflow for retention actions and produced the first labelled outcomes dataset for a planned causal uplift model.

Stack: `Dagster · dbt · PostgreSQL · Python · FastAPI · Redis · Angular · Docker · Caddy · Oracle Cloud · GitHub Actions · Git`

CRM Data Specialist · Natural Clinic · Istanbul, Türkiye · June 2024 - January 2025

Healthcare marketing organisation. CRM data and analytics ownership across patient acquisition, conversion, and retention.

- Rebuilt the CRM data layer end to end as the single source of truth for sales, marketing, and clinical operations. Lead response time dropped 30%; conversion rose 15%; duplicates and data errors fell 40%; email engagement rose 25%.
- Replaced weekly manual spreadsheet reporting with real-time Power BI dashboards and exec-ready reports, cutting leadership reporting latency from 7 days to live and freeing ~6 analyst-hours per week on report assembly.
- Integrated the CRM with the marketing automation stack (email, SMS, paid funnels) and engineered customer data flows between CRM, patient management, and marketing tooling, which removed three days a week of manual reconciliation.
- Translated ambiguous stakeholder requests from marketing, clinical operations, and customer service into 12+ shipped data models and dashboards, including the lead-scoring and conversion-funnel models that drove the 15% conversion lift.

Stack: `SQL · Python · Power BI · Zoho CRM · Marketing automation · Workflow automation`

Strategic Data Insights Analyst · Federal Mortgage Bank of Nigeria · Abuja, Nigeria · December 2018 - November 2019

Federal mortgage institution managing Nigeria's national housing fund. Loan portfolio analytics and IT-governance reporting in the Asset Creation Unit of the Loan Administration group.

- Contributed to a data-integrity overhaul that improved core dataset accuracy by 40% in six months, materially reducing the risk of faulty loan approvals.
- Built Python and SQL analytics on top of Tableau to track loan repayment behaviour, supporting a 15% lift in repayment-monitoring efficiency and a 25% reduction in manual data entry.
- Cut IT-asset-governance approval turnaround by 30% by re-mapping the approval workflow and instrumenting stage-level reporting in Tableau, shrinking median time-to-approve across the Loan Administration group.

Stack: `Python · SQL · Tableau · Excel`

Customer Information and Service Specialist · Guaranty Trust Bank · Edo, Nigeria · September 2015 - March 2016

Early-career customer service role at a Tier-1 African retail bank: supported the migration of in-branch customers to USSD and Internet banking through targeted onboarding sessions, owned complaint-resolution logging and root-cause categorisation for the branch's monthly service-quality report, and trained incoming customer-service interns on KYC and complaint handling.

SELECTED PROJECTS

Agentic ELT Data Platform for Customer Intelligence · MSc Dissertation (Salford) · 2026 · Live B2B SaaS Environment

Solo research build, end to end: I designed and wrote every layer myself, from ingestion to the serving API, against a live B2B SaaS environment under NDA. It ran on the same environment as my Force24 contract but is independent work; the Force24 platform itself was a team delivery.

- Ingestion and warehouse: custom Python extractors ingested over 1M records from multiple vendor APIs into a JSONB-first PostgreSQL warehouse, orchestrated with Dagster. - Serving and agent access: FastAPI service with JWT auth and Row-Level Security, Angular dashboard, and a Model Context Protocol (MCP) endpoint queryable by any MCP-compatible LLM agent for per-account risk profiling, driver explanations, and intervention recommendations. - Modelling: 48 dbt models across raw, staging, intermediate, and marts; layered Kimball-style dimensional design. - Intelligence stack: survival analysis (cross-validated C-index ~0.94), gradient-boosted classification with SHAP explanations (AUC ~0.95 vs ~0.89 logistic baseline), and DR-Learner causal inference (econml) correcting reactive-assignment selection bias by ~50 percentage points.

`Python PostgreSQL JSONB dbt Dagster PostgresML XGBoost SHAP Causal Inference Survival Analysis FastAPI Angular MCP`

Account Intelligence MCP Server · Side Project · 2026 · Agentic AI

FastMCP server exposing a full customer-intelligence stack (survival hazards, XGBoost health scores, DR-Learner causal effects, SHAP attributions) to any MCP-compatible LLM agent: 12 tools across churn-risk, account, and CSM-priority domains over a shared database layer. Dashboards inform; agents act.

`Python FastMCP Model Context Protocol Anthropic SDK SHAP SQL`

AirWave Express - Relational Database Design · Side Project · Aviation / SQL Server

SQL Server schema for an airport ticketing platform: seven core entities with sequence-and-trigger key generation, foreign keys enforcing real-world invariants (no ticket without its reservation), CHECK constraints on date ordering, and UNIQUE constraints on passports and flight codes. The schema is the spine; get it wrong and every workflow above it corrupts differently. Public case study on the portfolio.

`SQL Server T-SQL Schema Design Triggers Sequences`

Pharmaceutical Side Effect Classification · Side Project · 2025 · Healthcare ML

Production-shape multi-class classifier mapping free-text adverse-event descriptions to ten clinical categories across 11,825 marketed medicines. Random Forest at 98.5% accuracy on held-out test, Logistic Regression baseline at 97.2%. Above 95% F1 across every retained category. Inference CLI and pytest test suite shipped.

`Python scikit-learn TF-IDF Random Forest Logistic Regression pytest`

EDUCATION

MSc Data Science · University of Salford · Manchester, UK · January 2025 - May 2026 Dissertation: Agentic ELT Data Platform for Customer Intelligence.

MSc Big Data Analytics and Management · Bahcesehir University · Istanbul, Türkiye · September 2020 - March 2023 · GPA 3.67 / 4.00 Thesis: distributed PySpark pipeline for football scouting.

BEng Computer Engineering · University of Benin · Edo, Nigeria · September 2011 - August 2017

TECHNICAL SKILLS

Production and project experience: Languages: SQL · Python (pandas, NumPy, scikit-learn) · R · PySpark · Bash **Warehouse & pipelines:** PostgreSQL (JSONB + GIN) · dbt · Dagster · Kimball dimensional modelling · ELT/ETL design **ML & statistics:** XGBoost · SHAP · scikit-survival · econml (DR-Learner) · PostgresML · causal inference · survival analysis · RoBERTa and transformer sentiment (NLP) · TF-IDF · LDA · ARIMA/ETS time series **Serving & agentic access:** FastAPI · REST APIs · Model Context Protocol (MCP) endpoints for LLM agents · Docker · Caddy · Oracle Cloud **BI:** Power BI · Tableau **CI & source control:** Git · GitHub Actions · pytest **Working knowledge (coursework and certifications):** Snowflake · BigQuery (Google Cloud certified) · Databricks · SQL Server / T-SQL · Airflow · Looker · AWS · GCP · Azure · MLflow

LANGUAGES

English: Native · **German:** B1 (intermediate, working toward B2)

CERTIFICATIONS

Engineer Data for Predictive Modeling with BigQuery ML (Google Cloud) · Python for Data Science, AI & Development (Coursera) · Generative AI for Business Leaders (Coursera) · Enterprise Design Thinking Practitioner / Team Essentials for AI / Co-Creator (IBM)