

Marvis Osazuwa

Data Scientist · Analytics Engineer

Location: Manchester, UK · **Phone:** +44 7349 949871 · **Email:** marvis.osazuwa@hotmail.com · **LinkedIn:** linkedin.com/in/marvisosazuwa

GitHub: github.com/marz1307 · **Portfolio:** marz1307.github.io

PROFILE

Data Scientist and Analytics Engineer who ships ML that explains itself: survival analysis, DR-Learner causal inference, gradient-boosted classification with SHAP, on production data layers built end to end with dbt, Dagster, and Kimball-style dimensional modelling. Most recent work exposed the intelligence layer to agentic LLM workflows via a Model Context Protocol (MCP) endpoint. Career spans retail banking (Nigeria, 2015 to 2016), mortgage analytics (Nigeria, 2018 to 2019), healthcare marketing (Türkiye, 2024 to 2025), and a most recent Analytics Engineer role at Force24 (UK, 2026); MSc Big Data Analytics (Bahçeşehir, 2020 to 2023) and MSc Data Science (Salford, 2025 to May 2026). Currently building independent enterprise projects in the agentic and analytics space, deepening on AI engineering via certifications.

EXPERIENCE

Analytics Engineer · Force24 · Leeds, UK (Hybrid)

January 2026 – April 2026

Marketing automation SaaS. Sole architect and author of the data layer for a 3-tier customer intelligence platform, primary author across 8 backend domains, and shipping engineer for 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, eliminating record-matching gaps that had previously 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: 24/7 platform at zero infrastructure cost.
- Lifted support-attribution accuracy to auditor-defensible by designing a 10-step deterministic classification chain mapping every ticket to a single account, replacing a black-box best-guess join and closing the last open finding in the Customer Success metrics review.
- Locked correctness end-to-end with 123 dbt tests and 82 pytest tests gated through GitHub Actions CI, surfacing and resolving 22 data-quality bugs during the first production build. Silent metric drift now caught in CI instead of 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) giving Customer Success a single workflow for retention actions and generating the first labelled outcomes dataset for a planned causal uplift model. Designed on the principle that dashboards should drive action, not just display data.

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, eliminating 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 role at a Tier-1 African retail bank.

- Identified the top 3 customer drop-off points in the branch-service funnel by analysing 6 months of interaction-log data in SQL and Excel, feeding the prioritisation that informed the USSD and Internet banking migration campaign.
- Supported the migration of in-branch customers to USSD and Internet Banking by running targeted onboarding sessions during a multi-month campaign, contributing to the channel-migration target set by regional operations.
- Owned complaint-resolution logging and root-cause categorisation across the branch network, building the monthly service-quality dataset consumed by regional management for SLA reporting.
- Onboarded successive cohorts of customer-service interns on KYC, USSD-channel troubleshooting, and complaint-logging procedures.

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

PROJECTS

Agentic ELT Data Platform for Customer Intelligence

MSc Dissertation (Salford) · 2026 · Live B2B SaaS Environment

End-to-end ELT platform and three-model churn intelligence stack for a live B2B SaaS environment under NDA.

- **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. Output prioritisation surfaced high-risk accounts to CSMs for retention action.
- **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 on held-out cohort) for time-to-churn; gradient-boosted classification with SHAP explanations (AUC ~0.95 vs ~0.89 logistic baseline) for churn likelihood and drivers; DR-Learner causal inference (econml) correcting reactive-assignment selection bias by ~50 percentage points on the naïve estimator for honest CSM treatment effect.

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

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% (interpretable comparison within two points). Above 95% F1 across every retained category. Inference CLI and pytest test suite shipped.

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

Big Data Player Scouting (MSc Thesis)

Bahcesehir Thesis · 2023 · Sports / Big Data

Distributed PySpark pipeline ranking and recommending footballers across five top European leagues using UEFA event data and PlayeRank methodology. 500+ players, hypothesis-driven research. Published thesis, public codebase.

Python PySpark Big Data Recommender Systems

Equity Forecasting

Side Project · 2025 · Financial Time Series

End-to-end R analysis package for univariate equity forecasting on 5,124 daily NYSE observations. ARIMA, seasonal ARIMA, and ETS through a MODEL_REGISTRY pattern, with formal residual diagnostics (Ljung-Box, Shapiro-Wilk) and stationarity tests (ADF, KPSS) combined into a single decision rule.

R ARIMA ETS Time Series Statistical Testing

TECHNICAL SKILLS

Languages: SQL, Python (pandas, NumPy, scikit-learn), R, PySpark, T-SQL, Bash

ML & Statistics: scikit-learn, XGBoost, PostgresML, scikit-survival, econml (DR-Learner), SHAP, MLflow, RoBERTa, ARIMA, Time Series Analysis, NLP, Sentiment Analysis, TF-IDF, LDA, Causal Inference, Survival Analysis, A/B testing / experimentation

Agentic AI: Model Context Protocol (MCP), agentic ELT, LLM-accessible data layers

Warehousing & Storage: PostgreSQL (JSONB + GIN), Snowflake, BigQuery, Databricks, SQL Server

Transformation & Orchestration: dbt, Dagster, Apache Spark, Airflow

Service Layer: FastAPI, REST APIs, Docker, Caddy, Model Context Protocol (MCP)

BI & Visualisation: Power BI, Tableau, Looker

Cloud: AWS, GCP, Azure, Oracle Cloud

Source Control & CI: Git, GitHub, GitHub Actions, pytest

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

COMMUNITY & LEADERSHIP

Microsoft Learn Student Ambassador · Bahçeşehir University · Istanbul, Türkiye

January 2021 – September 2023 (2 yrs 9 mo) · Cross-university Azure workshops and a community hackathon with fellow ambassadors and interested learners.

Availability: from May 2026 for roles in Germany, the UK, and remote across EMEA timezones.