

Marvis Osazuwa

Analytics Engineer · Data Scientist

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

Analytics Engineer and Data Scientist building production data layers with embedded ML: recent work spans survival analysis, causal inference, and a Model Context Protocol (MCP) endpoint exposing the intelligence layer to agentic LLM workflows. Seven years across retail banking, mortgage analytics, healthcare marketing, and a most recent Analytics Engineer role at a SaaS company. I design dimensional models in SQL, transform with dbt, orchestrate with Dagster, and ship ML pipelines that explain themselves. Currently building independent enterprise projects in the agentic and analytics space and deepening on AI engineering via certifications. Open to roles in Germany, EMEA, the UK, or remote across EU and UK time zones.

EXPERIENCE

Analytics Engineer · Force24 · Leeds, UK (Hybrid)

January 2026 – April 2026

Marketing automation SaaS. Led the data layer for a 3-tier customer intelligence platform, sole architect and author of the data layer, primary author across 8 backend domains, and owner of the customer-facing surface area on the 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: 39 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.
- Replaced a black-box best-guess attribution join between support tickets and accounts with a 10-step deterministic classification chain, making support metrics defensible to auditors and Customer Success Managers.
- Locked correctness end-to-end with 123 dbt tests and 82 pytest tests, surfacing and resolving 22 data-quality bugs during the first production build. Silent metric drift now caught in CI instead of the boardroom.
- Built the customer-facing backend and Angular UI: customer list with 10 server-side filters, dashboard and reporting endpoints, and the Action Tracker page (sole author) with structured outcome capture across 7 outcome types. Designed the Action Tracker on the principle that dashboards should drive action, not just display data; captured outcomes form labelled training data for a future causal-model layer.

Stack: Dagster, dbt, PostgreSQL, Python, FastAPI, Redis, Angular, Docker, Caddy, Oracle Cloud, Git.

CRM Data Specialist · Natural Clinic · Istanbul, Türkiye

June 2024 – January 2025

Healthcare marketing operation. 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%.
- Designed and shipped real-time Power BI dashboards and exec-ready reports that replaced weekly manual spreadsheets, giving leadership live visibility for the first time.
- 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.
- Partnered daily with marketing, clinical operations, and customer service to translate ambiguous business questions into structured data models and measurable outcomes.

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.
- Supported the streamlining of IT asset governance and approval workflows, delivering a 30% reduction in approval turnaround time.

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.

- Analysed customer interaction data to identify drop-off and friction points, feeding insights into cross-functional teams.
- Supported adoption campaigns for USSD and Internet banking, helping shift customers off in-branch manual transactions and onto self-service channels.
- Owned complaint resolution logging and root-cause categorisation, feeding monthly service-quality reporting.

EDUCATION

MSc Data Science · University of Salford · Manchester, UK

January 2025 – May 2026

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

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. Built a custom Python ingestion layer that landed over 1M records from multiple vendor APIs into a JSONB-first PostgreSQL warehouse via Dagster, modelled the data through 48 dbt models, and shipped a three-model 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 with drivers, and causal inference correcting reactive-assignment selection bias by ~50 percentage points on the naïve estimator for honest CSM treatment effect. Surfaced via a FastAPI service with JWT auth and Row-Level Security, an 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. Prioritisation framework identified high-risk accounts for CSM retention action.

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-grade multi-class classifier mapping free-text adverse-event descriptions to ten clinical categories across 11,825 marketed medicines. Random Forest at 98.5 percent accuracy on held-out test, Logistic Regression baseline at 97.2 percent (interpretable comparison within two points). Above 95 percent 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 plus 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, R, PySpark, T-SQL, Bash

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)

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

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

BI & Visualisation: Power BI, Tableau, Looker

Cloud: AWS, GCP, Azure, Oracle Cloud

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

LANGUAGES

- **English:** Native or bilingual proficiency
- **German:** B1 (intermediate, working toward B2)

CERTIFICATIONS

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