

# Marvis Osazuwa

Analytics Engineer · Data Scientist



## PERSÖNLICHE DATEN

Anschrift Manchester, Vereinigtes Königreich  
Telefon +44 7349 949871  
E-Mail marvis.osazuwa@hotmail.com  
LinkedIn linkedin.com/in/marvisosazuwa  
GitHub github.com/marz1307  
Portfolio marz1307.github.io

## PROFIL

Analytics Engineer und Data Scientist mit Fokus auf produktive Datenebenen mit eingebettetem ML: aktuelle Arbeiten umfassen Survival Analysis, Causal Inference und einen Model-Context-Protocol-(MCP)-Endpoint, der die Intelligence-Schicht für agentische LLM-Workflows zugänglich macht. Sieben Jahre Erfahrung in Retail Banking, Hypothekenanalytik, Healthcare-Marketing sowie einer zuletzt ausgeübten Analytics-Engineer-Tätigkeit bei einem SaaS-Unternehmen. Ich entwerfe dimensionale Modelle in SQL, transformiere mit dbt, orchestriere mit Dagster und liefere ML-Pipelines, die sich erklären lassen. Aktuell baue ich unabhängige Enterprise-Projekte im agentischen und Analytics-Bereich und vertiefe mich auf der AI-Engineering-Seite über Zertifizierungen. Offen für Positionen in Deutschland, EMEA, im Vereinigten Königreich oder remote in EU- und UK-Zeitzone.

## BERUFSERFAHRUNG

### Analytics Engineer · Force24 · Leeds, Vereinigtes Königreich (Hybrid)

01/2026 – 04/2026

Marketing-Automation-SaaS. Leitete die Datenebene eines dreistufigen Customer-Intelligence-Platformaufbaus: alleiniger Architekt und Autor der Datenebene, primärer Autor über 8 Backend-Domänen sowie Verantwortlicher für die Customer-Surface auf dem Frontend.

- Alleiniger Architekt und Autor einer dbt- und Dagster-Datenpipeline, die 4 voneinander unabhängige Geschäftssysteme zu einem analyseergebnis-bereiten Warehouse auf einem deduplizierten Account-Spine konsolidierte. Etwa 40 dbt-Modelle über Staging-, Intermediate- und Marts-Ebenen, 5 tägliche Dagster-Schedules. Eine kanonische Kunden-ID im gesamten Unternehmen etabliert, wodurch Datensatz-Matching-Lücken beseitigt wurden, die zuvor CSM-Bücher nicht verknüpfbar machten.
- Tägliche Pipeline-Rechenzeit um etwa 95 % reduziert durch Umarchitektur hochvolumiger Revenue-Modelle als inkrementell auf einer Append-Only-Raw-Schicht mit deterministischen Hash-IDs. Plattform auf Oracle Cloud produktiv genommen mit idempotentem Bootstrap, TLS, VPN-Allowlist und dreistufigen Operator-Runbooks: 24/7-Plattform bei null Infrastrukturkosten.
- Eine Black-Box-Best-Guess-Attribution zwischen Support-Tickets und Accounts durch eine deterministische 10-Schritte-Klassifizierungskette ersetzt, wodurch Support-Kennzahlen gegenüber Auditoren und Customer Success Managern verteidigbar wurden.
- End-to-End-Korrektheit gesichert mit 123 dbt-Tests und 82 pytest-Tests, wobei 22 Datenqualitätsfehler im ersten Produktionsbuild aufgedeckt und behoben wurden. Stille Metrik-Drift wird in CI gefangen, nicht im Vorstandsmeeting.
- Customer-Backend und Angular-UI gebaut: Customer-Liste mit 10 serverseitigen Filtern, Dashboard- und Reporting-Endpoints sowie die Action-Tracker-Seite (alleiniger Autor) mit strukturierter Ergebniserfassung über 7 Ergebnistypen. Action Tracker nach dem Grundsatz konzipiert, dass Dashboards Handeln auslösen sollten, nicht nur Daten anzeigen; erfasste Ergebnisse bilden gelabelte Trainingsdaten für eine zukünftige Causal-Model-Schicht.

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

### CRM Data Specialist · Natural Clinic · Istanbul, Türkei

06/2024 – 01/2025

Healthcare-Marketing-Operation. Verantwortung für CRM-Daten und Analytik in den Bereichen Patientenakquise, Conversion und Bindung.

- CRM-Datenebene durchgängig neu aufgebaut und als Single Source of Truth für Vertrieb, Marketing und klinische Operations etabliert. Lead-Reaktionszeit -30 %, Conversion +15 %, Duplikate und Datenfehler -40 %, E-Mail-Engagement +25 %.
- Echtzeit-Dashboards und Executive-Reports in Power BI gestaltet und ausgeliefert, die wöchentliche manuelle Tabellen ersetzen und der Geschäftsleitung erstmals laufende Sichtbarkeit gaben.
- CRM mit dem Marketing-Automation-Stack (E-Mail, SMS, Paid Funnels) integriert sowie Datenflüsse zwischen CRM, Patientenmanagementsystem und Marketing-Tools entwickelt, wodurch drei Tage manueller Abstimmung pro Woche entfielen.
- Tägliche Zusammenarbeit mit Marketing, klinischen Operations und Kundenservice, um unklare Geschäftsfragen in strukturierte Datenmodelle und messbare Ergebnisse zu übersetzen.

**Stack:** SQL, Python, Power BI, Zoho CRM, Marketing-Automation, Workflow-Automation.

### Strategic Data Insights Analyst · Federal Mortgage Bank of Nigeria · Abuja, Nigeria

12/2018 – 11/2019

Bundesweite Hypothekeninstitution für den nigerianischen Wohnungsbaufonds. Kreditportfolio-Analytik und IT-Governance-Reporting in der Asset Creation Unit der Loan-Administration-Gruppe.

- Zu einer Datenintegritätsoffensive beigetragen, die die Genauigkeit der Kerndatensätze in sechs Monaten um 40 % steigerte und das Risiko fehlerhafter Kreditgenehmigungen wesentlich reduzierte.
- Python- und SQL-Analytik auf Basis von Tableau gebaut, um Rückzahlungsverhalten zu verfolgen, und damit eine 15-prozentige Effizienzsteigerung im Rückzahlungs-Monitoring sowie eine 25-prozentige Reduktion des manuellen Dateneingabe-Aufwands unterstützt.
- Die Verschlinkung von IT-Asset-Governance und Genehmigungsworkflows unterstützt, mit einer 30-prozentigen Reduktion der Bearbeitungszeit als Ergebnis.

**Stack:** Python, SQL, Tableau, Excel.

### Customer Information and Service Specialist · GTBank · Edo, Nigeria

09/2015 – 03/2016

Berufseinstieg bei einer der größten afrikanischen Privatkundenbanken.

- Kundeninteraktionsdaten analysiert, um Reibungspunkte zu identifizieren, und Erkenntnisse an funktionsübergreifende Teams weitergeben.

- Adoptionskampagnen für USSD- und Internet-Banking unterstützt und damit zur Reduktion manueller Filialtransaktionen und zur Verlagerung von Kundinnen auf Self-Service-Kanäle beigetragen.
- Verantwortung für Beschwerdebearbeitung, Logging und Ursachenklassifizierung; Zulieferung an monatliches Service-Quality-Reporting.

## AUSBILDUNG

### MSc Data Science · University of Salford · Manchester, Vereinigtes Königreich

01/2025 – 05/2026 · Schwerpunkte: Machine Learning, Big-Data-Engineering, angewandte Statistik, datengetriebene Entscheidungsfindung

### MSc Big Data Analytics and Management · Bahçeşehir Universität · Istanbul, Türkei

09/2020 – 03/2023 · Note: 3,67 / 4,00 · Masterarbeit: PySpark-Pipeline zum Ranking und zur Empfehlung von Spielern aus fünf europäischen Fußballligen anhand von UEFA-Event-Daten

### BEng Computer Engineering · University of Benin · Edo, Nigeria

09/2011 – 08/2017

## PROJEKTE

### Agentic ELT Data Platform for Customer Intelligence

MSc-Dissertation (Salford) · 2026 · Live-B2B-SaaS-Umgebung

End-to-End-ELT-Plattform und Drei-Modell-Churn-Intelligence-Stack für eine Live-B2B-SaaS-Umgebung unter NDA. Eigenentwickelte Python-Ingestion mit über 1 Mio. Datensätzen aus mehreren Vendor-APIs in ein JSONB-first PostgreSQL-Warehouse via Dagster ingestiert, durch 48 dbt-Modelle modelliert und mit einem Drei-Modell-Intelligence-Stack versehen: Survival-Analyse (kruzvalidierter C-Index ~0,94 auf Held-Out-Kohorte) für Time-to-Churn, Gradient-Boosted-Klassifikation mit SHAP-Erklärungen (AUC ~0,95 gegenüber ~0,89 Logistic-Baseline) für Churn-Wahrscheinlichkeit mit Treibern sowie Causal Inference, die reaktive Zuweisungs-Selektionsbias um ~50 Prozentpunkte auf dem naiven Schätzer korrigiert, für eine ehrliche CSM-Treatment-Effect-Schätzung. Bereitgestellt über einen FastAPI-Service mit JWT-Auth und Row-Level-Security, ein Angular-Dashboard und einen Model-Context-Protocol-(MCP)-Endpoint, abfragbar durch jeden MCP-kompatiblen LLM-Agent für Per-Account-Risiko-Profiling, Treibererklärungen und Interventionsempfehlungen. Priorisierungs-Framework identifiziert Hochrisiko-Accounts für CSM-Retention-Aktionen.

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

### Pharmaceutical Side Effect Classification

Eigenprojekt · 2025 · Healthcare-ML

Produktionsreifer Multi-Class-Klassifikator zur Zuordnung von Freitext-Beschreibungen unerwünschter Arzneimittelereignisse zu zehn klinischen Kategorien über 11.825 zugelassene Medikamente. Random Forest mit 98,5 % Genauigkeit auf der Test-Holdout-Menge, Logistic Regression als interpretierbarer Vergleich bei 97,2 %. Inference-CLI und pytest-Testsuite enthalten.

Python scikit-learn TF-IDF Random Forest Logistic Regression

### Big Data Player Scouting

MSc-Thesis · 2023 · Sport / Big Data

Verteilte PySpark-Pipeline zur Bewertung und Empfehlung von über 500 Spielern aus La Liga, Serie A, Premier League, Bundesliga und Ligue 1 anhand von UEFA-Event-Daten und der PlayeRank-Methodik. Veröffentlichte Thesis, öffentlicher Quellcode.

Python PySpark Big Data Recommender Systems

### Equity Forecasting

Eigenprojekt · 2025 · Finanzielle Zeitreihen

End-to-End-R-Analysepaket für univariate Aktienprognosen auf 5.124 NYSE-Tagesbeobachtungen. ARIMA, saisonales ARIMA und ETS über ein MODEL\_REGISTRY-Pattern. Formale Residualdiagnostik (Ljung-Box, Shapiro-Wilk) und Stationaritätstests (ADF, KPSS) zu einer einheitlichen Entscheidungsregel kombiniert.

R ARIMA ETS Zeitreihen Statistische Tests

## TECHNISCHE KENNTNISSE

**Sprachen:** SQL, Python, R, PySpark, T-SQL, Bash

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

**Transformation & Orchestrierung:** dbt, Dagster, Apache Spark, Airflow

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

**ML & Statistik:** scikit-learn, XGBoost, PostgresML, scikit-survival, econml (DR-Learner), SHAP, MLflow, RoBERTa, ARIMA, NLP, Sentimentanalyse, TF-IDF, LDA, Causal Inference, Survival Analysis

**Agentic AI:** Model Context Protocol (MCP), agentische ELT, LLM-zugängliche Datenebenen

**BI & Visualisierung:** Power BI, Tableau, Looker

**Cloud:** AWS, GCP, Azure, Oracle Cloud

**Versionskontrolle & Tooling:** Git, GitHub, GitHub Actions, pytest

## SPRACHEN

- **Englisch:** Muttersprachlich
- **Deutsch:** B1 (Mittelstufe, in Vorbereitung auf B2)

## ZERTIFIKATE

- 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