

Marvis Osazuwa

Data Scientist · Analytics Engineer



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

Data Scientist und Analytics Engineer für erklärbares ML: Survival Analysis, DR-Learner-Causal-Inference, Gradient-Boosted-Klassifikation mit SHAP, auf produktionsreifen Datenebenen mit dbt, Dagster und Kimball-Modellierung. Aktuelle Arbeiten erschließen die Intelligence-Schicht für agentische LLM-Workflows über einen Model-Context-Protocol-(MCP)-Endpoint. Werdegang über Retail Banking und Hypothekenanalytik (Nigeria, 2015 bis 2019), Healthcare-Marketing (Türkei, 2024 bis 2025) und eine zuletzt ausgeübte Analytics-Engineer-Tätigkeit bei Force24 (UK, 2026); MSc Big Data Analytics (Bahçeşehir, 2020 bis 2023) und MSc Data Science (Salford, 2025 bis Mai 2026). Aktuell baue ich unabhängige Enterprise-Projekte im agentischen und Analytics-Bereich und vertiefe mich auf der AI-Engineering-Seite über Zertifizierungen.

BERUFSERFAHRUNG

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

01/2026 – 04/2026

Marketing-Automation-SaaS. Alleiniger Architekt und Autor der Datenebene einer dreistufigen Customer-Intelligence-Plattform, primärer Autor über 8 Backend-Domänen sowie Shipping-Engineer für das Customer-Facing-Frontend.

- Alleiniger Architekt und Autor einer dbt- und Dagster-Datenpipeline, die 4 voneinander unabhängige Geschäftssysteme zu einem analysefähigen Warehouse auf einem deduplizierten Account-Spine konsolidierte: über 40 dbt-Modelle über Staging-, Intermediate- und Marts-Ebenen, 5 tägliche Dagster-Schedules, ca. 93 % alleiniger Commit-Anteil. 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 ca. 95 % reduziert durch Umarchitektur hochvolumiger Revenue-Modelle als inkrementell auf einer Append-Only-Raw-Schicht mit deterministischen Hash-IDs. Plattform auf Oracle Cloud Infrastructure produktiv genommen mit idempotentem Bootstrap, TLS, VPN-Allowlist und dreistufigen Operator-Runbooks: 24/7-Betrieb bei null Infrastrukturkosten.
- Support-Attributionsgenauigkeit auf auditorensicher gehoben, indem eine deterministische 10-Schritte-Klassifizierungskette jeden Datensatz einem Account eindeutig zuordnet und damit eine Black-Box-Best-Guess-Logik ersetzt. Letzte offene Finding im Customer-Success-Metrikreview geschlossen.
- End-to-End-Korrektheit gesichert mit 123 dbt-Tests und 82 pytest-Tests, abgesichert über die GitHub-Actions-CI, wobei 22 Datenqualitätsfehler im ersten Produktionsbuild aufgedeckt und behoben wurden. Stille Metrik-Drift wird in CI gefangen, nicht im Vorstandsmeeting.
- Customer-Facing-Backend und Angular-UI ausgeliefert (Customer-Liste mit 10 serverseitigen Filtern, Dashboard- und Reporting-Endpoints, alleinig verantwortete Action-Tracker-Seite über 7 Ergebnistypen), damit Customer Success einen einheitlichen Workflow für Retention-Aktionen und den ersten gelabelten Outcome-Datensatz für eine zukünftige kausale Uplift-Schicht bekommt. Grundsatz: Dashboards sollen Handeln auslösen, nicht nur Daten anzeigen.

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

CRM Data Specialist · Natural Clinic · Istanbul, Türkei

06/2024 – 01/2025

Marketingbetrieb im Healthcare-Sektor. 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 %.
- Wöchentliches manuelles Tabellen-Reporting durch Echtzeit-Power-BI-Dashboards und Executive-Reports ersetzt; Reporting-Latenz von 7 Tagen auf live reduziert, ca. 6 Analystenstunden pro Woche bei der Berichterstellung freigesetzt.
- 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.
- Unklare Stakeholder-Anfragen aus Marketing, klinischen Operations und Kundenservice in 12+ ausgelieferte Datenmodelle und Dashboards übersetzt, darunter die Lead-Scoring- und Conversion-Funnel-Modelle, die den 15 % Conversion-Uplift trieben.

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.
- Bearbeitungszeit der IT-Asset-Governance-Genehmigungen um 30 % reduziert durch Neukartierung des Approval-Workflows und Instrumentierung des Stage-Level-Reportings in Tableau, wodurch die Median-Zeit-bis-Genehmigung in der Loan-Administration-Gruppe geschrumpft ist.

Stack: Python, SQL, Tableau, Excel.

Customer Information and Service Specialist · GTBank · Edo, Nigeria

09/2015 – 03/2016

Berufseinstieg bei einer der größten afrikanischen Privatkundenbanken.

- Die drei größten Customer-Drop-off-Punkte im Filial-Service-Funnel identifiziert durch Analyse von 6 Monaten Interaktions-Log-Daten in SQL und Excel; lieferte die Priorisierungsgrundlage für die USSD- und Internet-Banking-Migrationskampagne.

- Die Migration von Filialkunden auf USSD und Internet-Banking durch gezielte Onboarding-Sessions im Rahmen einer mehrmonatigen Kampagne unterstützt und damit zum Kanalmigrationsziel des regionalen Betriebs beigetragen.
- Verantwortung für Beschwerdebearbeitung, Logging und Ursachenklassifizierung im Filialnetz; den monatlichen Service-Quality-Datensatz aufgebaut, der vom regionalen Management für SLA-Reporting konsumiert wurde.
- Aufeinanderfolgende Kohorten von Customer-Service-Praktikanten in KYC, USSD-Kanal-Troubleshooting und Beschwerde-Logging eingearbeitet.

AUSBILDUNG

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

01/2025 – 05/2026 · Dissertation: Agentic ELT Data Platform for Customer Intelligence · Schwerpunkte: ML, Big-Data-Engineering, angewandte Statistik

MSc Big Data Analytics and Management · Bahçeşehir Üniversitesi · Istanbul, Türkei

09/2020 – 03/2023 · Note: 3,67 / 4,00 (ca. 1,7 nach deutschem System) · Masterarbeit: PySpark-Pipeline für Spielerempfehlungen in fünf europäischen Fußballligen (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.

- **Ingestion und Warehouse:** Eigene Python-Extraktoren ingestierten über 1 Mio. Datensätze aus mehreren Vendor-APIs in ein JSONB-orientiertes PostgreSQL-Warehouse, orchestriert mit Dagster.
- **Serving und Agent-Zugriff:** FastAPI-Service mit JWT-Auth und Row-Level-Security, Angular-Dashboard sowie Model-Context-Protocol (MCP)-Endpoint, abfragbar durch jeden MCP-kompatiblen LLM-Agent für Per-Account-Risiko-Profilung, Treibererklärungen und Interventionsempfehlungen. Die Priorisierung der Ausgaben hob Hochrisiko-Accounts für CSM-Retention-Aktionen hervor.
- **Modellierung:** 48 dbt-Modelle über Raw-, Staging-, Intermediate- und Marts-Schichten; mehrschichtiges dimensionales Design nach Kimball.
- **Intelligence-Stack:** Survival-Analyse (kruzvalidierter C-Index ca. 0,94 auf Held-Out-Kohorte) für Time-to-Churn; Gradient-Boosted-Klassifikation mit SHAP-Erklärungen (AUC ca. 0,95 gegenüber ca. 0,89 Logistic-Baseline) für Churn-Wahrscheinlichkeit und Treiber; DR-Learner-Causal-Inference (econml), die reaktive Zuweisungs-Selektionsbias um ca. 50 Prozentpunkte auf dem naiven Schätzer korrigiert, für eine ehrliche CSM-Treatment-Effect-Schätzung.

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

Masterarbeit · 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 (pandas, NumPy, scikit-learn), R, PySpark, T-SQL, Bash

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

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

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)

BI & Visualisierung: Power BI, Tableau, Looker

Cloud: AWS, GCP, Azure, Oracle Cloud

Versionskontrolle & CI: 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, AI & Development · Coursera
- Generative AI for Business Leaders · Coursera
- Enterprise Design Thinking Practitioner / Team Essentials for AI / Co-Creator · IBM

ENGAGEMENT

Microsoft Learn Student Ambassador · Bahçeşehir Üniversitesi · Istanbul, Türkei

01/2021 – 09/2023 · Universitätsübergreifende Azure-Workshops und Community-Hackathon mit Ambassadors und Lernenden.

Verfügbarkeit: ab Mai 2026 für Positionen in Deutschland, UK und remote in EMEA-Zeitzone.