



PERSÖNLICHE DATEN

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

PROFIL

MSc-Data-Science-Absolvent (Salford, Mai 2026), der bereits produktiv ausgeliefert hat, was die meisten Absolventen nur aus dem Studium kennen: produktive Datenschichten end-to-end, mit ML, das sich selbst erklärt. Ich baue dbt- und Dagster-Pipelines auf Kimball-modellierten Data Warehouses, sichere Korrektheit in der CI und setze angewandtes ML (Survival-Analyse, Gradient-Boosting mit SHAP, kausale Inferenz) direkt darauf, zuletzt über einen Model-Context-Protocol-(MCP)-Endpoint für LLM-Agenten zugänglich gemacht. Bei Force24 verantwortete ich die Datenschicht einer im Team gebauten Customer-Intelligence-Plattform; meine Masterarbeit war davon unabhängig ein Solo-Projekt: ein End-to-End-Churn-Intelligence-Stack, den ich allein gegen dieselbe Live-Umgebung entworfen und gebaut habe. UK Graduate Visa: volles Arbeitsrecht im Vereinigten Königreich, kein Sponsoring erforderlich. Für Deutschland und die EU EU-Blue-Card-berechtigt.

BERUFSERFAHRUNG

Analytics Engineer · Force24 · Leeds, Vereinigtes Königreich (Hybrid) · Befristeter Vertrag, Januar 2026 bis April 2026
Marketing-Automation-SaaS. Ein 16-Wochen-Teamprojekt: eine dreistufige Customer-Intelligence-Plattform. Innerhalb der gemeinsamen Lieferung verantwortete ich die Datenschicht vollständig als alleiniger Architekt und Autor, war primärer Autor über 8 Backend-Domänen und lieferte das kundenseitige Frontend aus.

- 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, sodass sie rund um die Uhr bei null Infrastrukturkosten läuft.
- 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. Die erfassten Ergebnisse bilden den ersten gelabelten Trainingsdatensatz für eine geplante 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, unter Senior-Supervision.

- 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 im Kundenservice bei einer der größten afrikanischen Privatkundenbanken: Kundinnen und Kunden beim Umstieg auf USSD- und Internet-Banking begleitet, die Beschwerdebearbeitung samt Ursachenklassifizierung für das monatliche Service-Quality-Reporting verantwortet und neue Kundenservice-Praktikanten eingearbeitet.

AUSGEWÄHLTE PROJEKTE

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

Eigenständiges Solo-Projekt von Anfang bis Ende: jede Schicht selbst entworfen und geschrieben, von der Ingestion bis zur Serving-API, gegen eine Live-B2B-SaaS-Umgebung unter NDA. Über 1 Mio. Datensätze aus Vendor-APIs in ein JSONB-first PostgreSQL-Warehouse (Dagster), 48 dbt-Modelle in Kimball-Schichten, Drei-Modell-Intelligence-Stack (Survival-Analyse, C-Index ~0,94; XGBoost mit SHAP, AUC ~0,95; DR-Learner Causal Inference) und ein Model-Context-Protocol-(MCP)-Endpoint für agentische LLM-Workflows.

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

Pharmaceutical Side Effect Classification · Eigenprojekt · 2025 · Healthcare-ML

Produktionsreifer Multi-Class-Klassifikator: Freitext-Beschreibungen unerwünschter Arzneimittelereignisse auf zehn klinische Kategorien über 11.825 Medikamente. Random Forest 98,5 % Genauigkeit (Holdout), Logistic Regression 97,2 % als interpretierbarer Vergleich. Inference-CLI und pytest-Testsuite.

Python scikit-learn TF-IDF Random Forest pytest

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, mit formaler Residualdiagnostik (Ljung-Box, Shapiro-Wilk) und Stationaritätstests (ADF, KPSS).

R ARIMA ETS Zeitreihen Statistische Tests

Clinical Trial Data Analysis · Eigenprojekt · 2024 · Healthcare-Statistik

Reproduzierbare Notebook-Analyse klinischer Studienergebnisse: Schema-Audit, dokumentierte Missing-Value-Regeln, deskriptive Statistik über Behandlungs- und Kontrollarme, Hypothesentests je nach Datenform, Effektstärken neben p-Werten. Öffentliche Fallstudie im Portfolio.

Python Pandas Statistische Tests Jupyter

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

TECHNISCHE KENNTNISSE

Produktions- und Projekterfahrung: Sprachen: SQL · Python (pandas, NumPy, scikit-learn) · R · PySpark · Bash **Warehouse & Pipelines:** PostgreSQL (JSONB + GIN) · dbt · Dagster · Kimball-Dimensionsmodellierung · ELT/ETL-Design **ML & Statistik:** XGBoost · SHAP · scikit-survival · econml (DR-Learner) · PostgresML · Causal Inference · Survival Analysis · RoBERTa- und Transformer-Sentimentanalyse (NLP) · TF-IDF · LDA · ARIMA/ETS-Zeitreihen **Serving & agentischer Zugriff:** FastAPI · REST APIs · Model-Context-Protocol-(MCP)-Endpoints für LLM-Agenten · Docker · Caddy · Oracle Cloud **BI:** Power BI · Tableau **CI & Versionskontrolle:** Git · GitHub Actions · pytest **Grundkenntnisse (Kurse und Zertifikate):** Snowflake · BigQuery (Google-Cloud-zertifiziert) · Databricks · SQL Server / T-SQL · Airflow · Looker · AWS · GCP · Azure · MLflow

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 (IBM) · Tableau for Data Scientists (LinkedIn Learning) · Statistics Foundations 1-4 (LinkedIn Learning)

Verfügbarkeit: Sofort verfügbar für Rollen im Vereinigten Königreich; ab Juli 2026 für Deutschland und die EU (Blue-Card-Bearbeitung). MSc abgeschlossen Mai 2026. Ziel: Analytics-Engineer- und Data-Scientist-Rollen im Vereinigten Königreich und in Deutschland/EU. UK Graduate Visa: kein Sponsoring nötig; EU-Blue-Card-berechtigt.