

JOHN ELIFURAHA MZIRAY

Dar es Salaam, Tanzania | jem.mziray@gmail.com | +255 743 924 467 |
<https://www.linkedin.com/in/john-mziray/> | <https://github.com/jemmziray-tech>

PROFESSIONAL SUMMARY

Highly analytical and results-driven Data Analyst and Python Developer specializing in extracting actionable business intelligence from complex, real-world datasets. Expert in Python, SQL, and modern data visualization (Streamlit, Plotly, PowerBI). Proven ability to architect end-to-end data pipelines, curate large-scale datasets, and deploy predictive forecasting models that solve significant economic and operational challenges. Passionate about transforming raw telemetry into high-impact, executive-level insights.

TECHNICAL EXPERTISE

- **Languages:** Python, SQL, R
- **Analytics & Data Engineering:** Data Visualization, ETL Pipelines, Exploratory Data Analysis (EDA), Statistical Forecasting, Business Intelligence, Data Cleaning, Geospatial Analysis.
- **Libraries & Frameworks:** Pandas, NumPy, Plotly, Streamlit, Scikit-Learn, Pydeck (Geospatial), NLP Libraries.
- **Tools & Cloud Infrastructure:** Google Cloud Platform (Firestore/Firebase), GitHub Actions (for Cron Scheduling/Automation), Git, APIs (Google Maps Enterprise, Open-Meteo), PowerBI.

ANALYTICS & ENGINEERING PROJECTS

Dar es Salaam Smart City Traffic Engine | Creator & Lead Analyst (*Python, Streamlit, Pandas, Scikit-Learn, GitHub Actions, Firebase*)

- **Data Pipeline Engineering:** Architected an autonomous, cloud-native ETL pipeline using GitHub Actions to scrape, clean, and ingest real-time geospatial traffic and meteorological data across 21 major city arteries every 20 minutes.
- **Business Intelligence Dashboard:** Designed an enterprise-grade interactive dashboard using Streamlit and Plotly to visualize network efficiency, spatial bottlenecks, and weather-related infrastructure vulnerability.
- **Economic Impact Analytics:** Developed a "Cost of Congestion" mathematical model that converts live traffic variance into tangible business metrics, calculating daily capital and fuel lost to gridlock.

- **Predictive Modeling:** Engineered and trained a Random Forest regression model (achieving an R^2 score of 0.83) to accurately forecast city-wide traffic delays based on cyclical hour, day, and precipitation states.

Netflix Trend Analysis | Data Analyst (*Python, Pandas, Matplotlib*)

- Conducted comprehensive Exploratory Data Analysis (EDA) on a massive public Netflix dataset.
- Cleaned and transformed unstructured data to investigate and visualize specific entertainment trends and duration distributions for 1990s content.

Daktari Msaidizi (Healthcare NLP) | Data Architect (*Python, NLP, Data Curation*)

- Curated, cleaned, and structured a proprietary dataset of 1,000 Swahili medical dialogues, demonstrating strict attention to data quality and intelligence structuring for the Tanzanian context.
- Architected a localized healthcare chatbot leveraging Natural Language Processing to serve underrepresented demographics.

Atmos-Live-API | Developer (*Python, REST APIs*)

- Engineered a lightweight command-line application that seamlessly extracts, parses, and visualizes real-time meteorological data using external global APIs.

Open-Source Contribution | MalariaGEN

- Contributed to the open-source genomics community by setting up a development environment and reviewing issues for the [malaria-gen-data-python](#) repository, supporting public health data initiatives.

CERTIFICATIONS

- **Google Data Analytics Professional Certificate** | Coursera (Apr 2026)
- **Google AI Essentials Specialization** | Coursera (Apr 2026)
- **Google Data analysis with Python professional Certificate** | Coursera (Mar 2026)
- **Python (Basic) Certification** | HackerRank (Apr 2026)
- **Google AI Professional Certificate** | Coursera (Mar 2026)
- **Secure Against Network Intrusions** | Coursera (Mar 2026)
- **Data Structures in Python** | Coursera (Mar 2026)
- **Explore Raw Data** | Coursera (Mar 2026)
- **Hello, Python!** | Coursera (Mar 2026)
- **Loops and Strings** | Coursera (Mar 2026)
- **Data Manipulation with pandas** | DataCamp (Feb 2026)
- **Introduction to SQL** | DataCamp (Feb 2026)

- **Intermediate SQL** | DataCamp (May 2026)

EDUCATION

University of Limerick | Limerick, Ireland

- *Prospective Student (Commencing Sept 2026)* - Bachelor of Science in Artificial Intelligence and Machine Learning