# Denzel Kanyeki Kinyua Data Analyst [denzelkinyua11@gmail.com](mailto:denzelkinyua11@gmail.com) [LinkedIn](https://www.linkedin.com/in/denzel-kinyua-a73597277)[GitHub](https://github.com/dkkinyua)[Blog](https://dev.to/dkkinyua)[Twitter](https://x.com/nyekssotrilll?t=K_I8odMIo9bigyxO1StVnQ&s=09)



# ABOUT

Data Analyst with expertise in building scalable data pipelines, developing interactive dashboards, and applying statistical and predictive analytics to uncover trends, assess risk, and drive data-informed decision-making. Skilled in Python, SQL, and BI tools, that is, Power BI and Grafana, with hands-on experience in ETL/ELT workflows, data modeling, and hypothesis testing. Adept at analyzing structured and unstructured datasets to deliver actionable insights that improve efficiency, support compliance and risk mitigation, and guide strategic business outcomes. Familiar with applying machine learning and forecasting techniques to real-world datasets, with a strong interest in leveraging data for financial services, insurance, and other regulated industries.



# SKILLS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **LANGUAGES AND TOOLS** | **SKILLS** | **DATA STORAGE PLATFORMS** | **DATA TOOLS** | **SOFT SKILLS** |
| Python | Data Modelling | AWS, Azure | Apache Airflow | Communication |
| SQL | Machine Learning | Snowflake, BigQuery | Apache Spark | Analytical skills |
| Git, GitHub Actions, Docker, Linux, Microsoft Excel | ETL and ELT Development | PostgreSQL, MySQL | dbt (data build tool) | Interpersonal skills |
| Grafana, Power BI | Data Cleaning, Visualization | Apache Cassandra | Apache Kafka | Team Collaboration |

PROJECTS

**Kenyan Food Prices and Inflation Project -** [**GitHub**](https://github.com/dkkinyua/FoodPricesMonitoring)**,** [**Grafana Dashboard**](https://deecodes.grafana.net/public-dashboards/21fed46cfba74c1c887547b2e8f8d2dd), [PowerPoint Presentation](https://docs.google.com/presentation/d/1VO8mjxU8ZAOtcj_4GM0UavXe1nc1Iul3/edit?usp=sharing&ouid=112113118836508124063&rtpof=true&sd=true)

* Designed and implemented a **Food Prices and Inflation Monitoring Pipeline**, automating ingestion of food price datasets from the **Humanitarian Data Exchange (HDX)**.
* **Applied Python (pandas, NumPy) transformations to clean, structure, and validate data for accuracy and reliability.**
* Developed **ELT workflows orchestrated with Apache Airflow**, ensuring timely updates and reproducibility of food price data.
* Produced datasets that support **inflation analysis and trend monitoring**, providing actionable insights for **food security and economic resilience**.
* Automated ETL process and email notifications using **Apache Airflow** by running the script on an hourly basis.
* Visualized food prices and inflation data, for example, prices for different food categories in various Kenyan counties in **Grafana**, with dynamic layers based on magnitude categories.
* Documented data pipeline processes and maintained data quality checks to ensure **governance, security, and usability**.

**Fraud Detection System with Isolation Forest ML Model -** [**GitHub**](https://github.com/dkkinyua/FraudDetectionSystem)**,** [**Grafana Dashboard**](https://deecodes.grafana.net/public-dashboards/ac5cf512ae80435c8bdb1adad4d1a27c)

* Designed and implemented a **real-time fraud detection pipeline** leveraging **Kafka hosted on Redpanda** for event streaming of synthetic transactions generated via Faker.
* Built and managed **Kafka topics** (transactions, fraud) to handle transaction data ingestion with a one-week retention policy.
* Developed **Kafka consumers** to stream data into **Postgres tables**, separating datasets for model training and data validation.
* Trained an **Isolation Forest anomaly detection model** to identify potentially fraudulent transactions and stored the trained job for real-time inference.
* Integrated **Streamlit dashboards** to visualize fraud detection performance and model accuracy in real time.
* Deployed **Grafana dashboards** for continuous monitoring and analytics, enabling real-time insights with auto-refresh functionality.
* Applied best practices in **data engineering, stream processing, and observability**, ensuring scalability and transparency of fraud detection workflows and reduced risk exposure by enabling real-time anomaly detection, supporting fraud prevention decisions.

**ETL Data Pipeline for Cryptocurrency Data Analytics -** [**GitHub**](https://github.com/dkkinyua/BinanceCryptoPipeline)**,** [**Grafana Dashboard**](https://deecodes.grafana.net/public-dashboards/f7c9b0e801d0496987ccd3f38ac92013)

* Built an end-to-end crypto data pipeline using Python, PostgreSQL, Debezium, Apache Cassandra, PySpark, and Apache Airflow to automate data extraction from the Binance Websocket API.
* Orchestrated ETL workflows using Airflow to schedule, monitor, and manage data extraction and transformation tasks.
* Stored extracted real-time data in a PostgreSQL database.
* Used Debezium for a Change Data Capture (CDC) mechanism to improve data consistency, integrity, and data processing by using log-based CDC.
* Stored the data in Apache Cassandra hosted on an Azure Virtual Machine as a data warehouse for downstream data consumption, i.e., data analysis.
* Used a PySpark Streaming job to stream data from the Cassandra database for ease of visualization in Grafana.
* Visualized data on a Grafana dashboard for insights like volume of cryptocurrency traded per hour, and number of trades per coin in the Binance order book.

**eazyetl, an ETL pipeline Python Package -** [**PyPI**](https://pypi.org/project/eazyetl/)

* Built and published eazyetl, a modular Python ETL package using Python OOP, pandas, requests, sqlalchemy, and psycopg2.
* Released on PyPI, allowing easy installation and use in custom data workflows via pip install.
* Implemented reusable components for extraction, transformation, and loading across databases.
* Added logging, config management, and error handling to support reliable production use and assist in efficient pipeline development



EXPERIENCE

**Intern - Data Engineer | Data Science East Africa | Nairobi, Kenya | May 2025 – Present (Full Time)**

* Supported **ETL workflows** and assisted in integrating various **data sources** for inventory and logistics analytics.
* Collaborated with team members to **optimize data pipelines**, hence increasing operational efficiency by 15%.
* Utilized Grafana and other Business Intelligence (BI) tools, such as Tableau, for dashboarding and real-time data analysis in streaming pipelines.
* Performed data validation to ensure accuracy and consistency across datasets, and collaborated with data analysts and data scientists to support the development of internal dashboards and analytical reports by providing clean, structured data from data pipelines and increasing data integrity by 20%.
* Participated in weekly reviews to interpret data trends and support risk mitigation decisions.
* Implemented data governance, validation, and access controls to ensure secure, reliable analytics pipelines.

### Intern - Operations | Herdy Fresh | Nairobi, Kenya | January 2020 – November 2020 (Full Time)

* Supported daily operational activities, including inventory tracking, order fulfillment, and logistics coordination, working under the Chief Operating Officer (COO).
* Assisted in streamlining supply chain processes to improve delivery efficiency and reduce operational delays by 25%.
* Contributed to data entry and basic analysis to support decision-making in procurement and distribution.



## EDUCATION

Bachelor of Science ( Civil Engineering ) - Machakos University

**CERTIFICATIONS**

Data Engineering Certificate - Lux Tech Academy (Nairobi)  
Data Protection Course (Ongoing) - Strathmore School, Center For Intellectual Property and Information Technology Law (CIPIT)



## INTERESTS AND ACTIVITIES

* I write and publish technical articles on my [blog](https://dev.to/dkkinyua) page, where I document my journey into data and share insights on tools, techniques, and technologies in the data engineering space.