

# EMMANUEL KOECH

manureservations@gmail.com ❖ +254727639047 ❖ Nairobi  
medium.com/@manureservations? ❖ www.linkedin.com/in/emmanuelkoech- ❖ github.com/emmanuelkipnetich  
source=user\_profile----- a223a2171

## PROFESSIONAL SUMMARY

As a skilled Data Analyst with experience in statistical analysis, data visualization, and machine learning techniques, I am eager to leverage my expertise to make a significant impact in the field of data science. I possess a detail-oriented and analytical approach, which enables me to identify trends, patterns, and insights that inform strategic decision-making. My proficiency in data cleaning, transformation, and visualization using tools such as Python, SQL, Excel, and Power BI makes me a valuable asset to any organization. I have experience collaborating with interdisciplinary teams to achieve project goals, and my excellent communication skills allow me to convey complex data-driven insights to diverse audiences. I am dedicated to staying up-to-date with the latest industry trends and continuously expanding my skills and knowledge in the data science field. I am excited to bring my passion and expertise to a dynamic and innovative organization where I can contribute to the success of the team and help drive business growth.

## ACCOMPLISHMENTS

- Completed the KPMG Virtual Internship Program which involved: Data Quality Assessment, Data Insights, Data Presentation
- Doing the BCG Virtual Internship Program which involves: Market Research, Data Analysis, Modelling, Understanding Consumer Needs, and Client Recommendations

## EDUCATION

|   |                       |
|---|-----------------------|
| <b>Azubi Africa</b>   | <b>October 2023</b>   |
| <i>Data Analytics Professional Certificate</i>  | <i>Accra, Ghana</i>   |
| <b>Jomo Kenyatta University of Agriculture and Technology</b>   | <b>June 2024</b>      |
| <i>Bsc. Financial Engineering</i>   | <i>Nairobi, Kenya</i> |
| <ul style="list-style-type: none"><li>Coursework: Probability &amp; Statistics, Time Series Analysis, Calculus, Stochastic Processes, R programming, Linear Algebra</li></ul> |                       |
| <b>Mang'u High School</b>   | <b>November 2017</b>  |
| <i>HIGH SCHOOL DIPLOMA</i>  | <i>Nairobi, Kenya</i> |
| <ul style="list-style-type: none"><li>GPA: B+, 71/84points</li><li>Coursework: Maths, Business, Physics, Geography</li></ul>  |                       |
| <b>Sally Ann School</b>   | <b>November 2013</b>  |
| <i>Primary Education</i>  |                       |
| <ul style="list-style-type: none"><li>GPA: A, 428/500</li></ul>   |                       |

## WORK EXPERIENCE

|   |                            |
|---|----------------------------|
| <b>DataSeal Software Ltd</b>  | <b>July 2023 – Present</b> |
| <i>Data Architect</i>   | <i>Nairobi, Kenya</i>      |
| <ul style="list-style-type: none"><li>Responsible for delivering and coordinating data-profiling, data-analysis, data-governance, data-models (conceptual, logical, physical), data-mapping, data-lineage and reference data management.</li><li>Participated in meetings &amp; JAD sessions to gather and collect requirements from the stakeholders as well as reporting requirements. Worked with Business Analysts, SQL Developers, end users and stakeholders to understand the requirements and signoff on the various deliverables.</li><li>Create team strategies and establish project scopes of work; communicate project deliverable time frames and benchmarks to project sponsors.</li><li>Created process flows and data flow diagrams of the current and future systems.</li></ul> |                            |

- Suggest architectural improvements, design and integration solutions, and formulate methodologies to optimize database development.
- Performed Data profiling and exploration using SQL and Informatica IDQ
- Performed gap analysis for different data sources.
- Created Source to Target mapping to include transformations and data quality rules needed for EDW Integration
- Created Data Models using Data Vault methodology to support enterprise reporting.
- Maintained metadata, version controlling of the data model.
- Created Materialized views & Summary tables for Reporting
- Integrated physical data model with business glossary.
- Implemented data governance. Worked with the business users to populate business glossary.
- Worked in Agile environment to help teamwork toward our goals
- Performance Tuning on SQL Queries for efficiency and performance
- Build and Manage AWS Data Lake for data storage and analytics
- Data Migration from Hadoop data lake to Amazon Web Services cloud and Hive platforms.

## Zalego Academy

August 2021 – December 2021

### Data Analytics Attache

Nairobi, Kenya

- Collaborated in the development of a digital credit scoring model built around open-source software. The process involved data collection, data review, data cleansing, feature selection and engineering, Model selection and fit, model validation and model deployment. The major python libraries utilized were
- Matplotlib, Pandas, Numpy and Skicit-learn.

## SKILLS

**Excel:** V Lookup, Conditional Formatting, Pivot Tables

**SQL:** Oracle SQL Developer, MySQL, DBeaver, Informatica Powercenter, Toad Oracle, SQL Server

**Visualization:** Power BI, Tableau, Power Query

**Python:** Pandas, NumPy, Scikit-learn, Matplotlib, SciPy, FastAPI, TensorFlow

**Microsoft Azure:** Azure Synapse Analytics, Azure Data Factory, Azure SQL Database, Azure Stream Analytics, Azure Data Lake Storage

**AWS:** Amazon S3 (Simple Storage Service), Amazon Redshift, AWS Glue, Amazon Managed Workflows for Apache Airflow (MWAA)

## PROJECTS

### COVID-19 VACCINE SENTIMENT ANALYSIS AND DEPLOYMENT USING GRADIO | Gradio, August 2023 –

Fine Tuning Pretrained Hugging Face Models, Google Colab, NLP

September 2023

- This project aimed to develop a Machine Learning model to assess positive, negative, or neutral Twitter posts related to vaccinations. The solution will help the governments and other public health actors monitor public sentiment towards COVID-19 vaccinations and help improve public health policy, vaccine communication strategies, and vaccination programs worldwide. In this project, I worked with text data; text data was initially non-numerical and I had to process it before it could be fed into a Machine Learning algorithm. Finally deployed the model as an app using Gradio.

### CLASSIFICATION ML PROJECT: A CASE STUDY OF CUSTOMER CHURN AT VODAFONE |

May 2023 –

Microsoft SQL Server, pyodbc Python: pandas, numpy, matplotlib, scipy, sklearn

June 2023

- The project aimed at training a machine learning model using supervised learning with an end goal of having the model predict the customer churn and also to define the reasons that lead to a customer churning. Some models utilized include: Logistic regression, Random Forest, Gradient Boosting, Support Vector Machine, Gaussian Naive Bayes, K-Nearest Neighbors, Decision Tree, XGBoost and LightGBM

### INDIAN START-UP EDA AND DEPLOYMENT TO POWER BI | Python: Exploratory Data Analysis &

April 2023 –

Data Cleaning, pandas, NumPy, Seaborn, Matplotlib Power BI

May 2023

- The goal of this project was to use data to explore the Indian Start-up ecosystem, uncover insights, and recommend the best course of action to a team attempting to enter the ecosystem.

## CERTIFICATIONS

Data Analysis using Python

April 2023

*IBM*

Microsoft Power BI masterclass

June 2023

*Zindua School*

## PUBLICATIONS

---

Medium Articles on Data Science/Analytics & Architecture

Medium

[medium.com/@manureservations? source=user\\_profile](https://medium.com/@manureservations?source=user_profile)-----

- I have published several articles on different aspects relating to data and the data science career that are useful to professionals across the field. The articles have been made publicly available via Medium.