Research Analysis Report

Prepared by: Selline Owiti
Date: June, 2025
Department: Data Science | Research Analysis
Project: Tracking Livestock Production Trends in Kenya (2021 - 2023)

---

1. Introduction

This report presents a comprehensive analysis of livestock production trends in Kenya between 2021 and 2023. The project aims to uncover long-term production patterns across various animal products, identify key changes over time, and generate data-driven insights to inform sustainable and climate-resilient livestock systems.

The dataset used in this study originates from FAO-like structured agricultural data, including fields such as **Element, Item, Year, Unit, Value, and Data Quality Flags**. The scope of this report covers data preprocessing, exploratory data analysis, and interpretation of key findings across the three-year span.

Report Structure:

- **Section 2:** Data Preprocessing – Cleaning, handling missing values, and structuring for analysis.
- **Section 3:** Exploratory Data Analysis (EDA) – Visualizing trends, comparisons, and insights by year and animal product.
- **Section 4:** Key Insights & Conclusions – Summary of findings and recommendations.

---

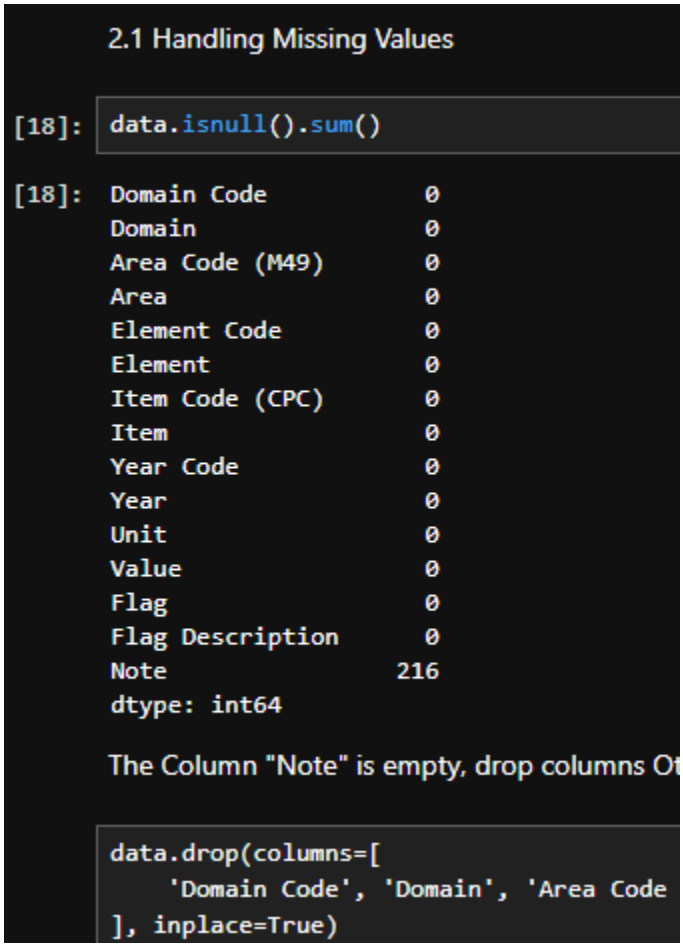
2. Data Preprocessing

2.1 Data Loading and Initial Exploration

The dataset was imported from a CSV file, comprising a multi-year time series of livestock data. Initial observations revealed:

- **Total Records:** 216 rows

- **Initial Features:** 15 columns, including multiple identifiers, categorical firleds, and quantitative values.
- **Missing Data:** The column ‘Note’ contained missing values in all rows.
- **Data Types:** A mix of text (categorical) and numerical fields



2.2 Cleaning and Transformation

To streamline the dataset for analysis, the following steps were executed:

- **Dropped Redundant Columns:** 'Domain Code', 'Domain', 'Area Code (M49)', 'Area', 'Year Code', and 'Note'.

- **Handled Missing Values:** The only column with 100% missing data (Note) was removed.

- **Validated Data Consistency:** Remaining values were confirmed for consistency and cleaned for formatting errors.

2.3 Output

The cleaned dataset was saved as **FOSTAS Data Cleaned_alt.csv**, ready for downstream analysis and dashboard development in Power BI.

---

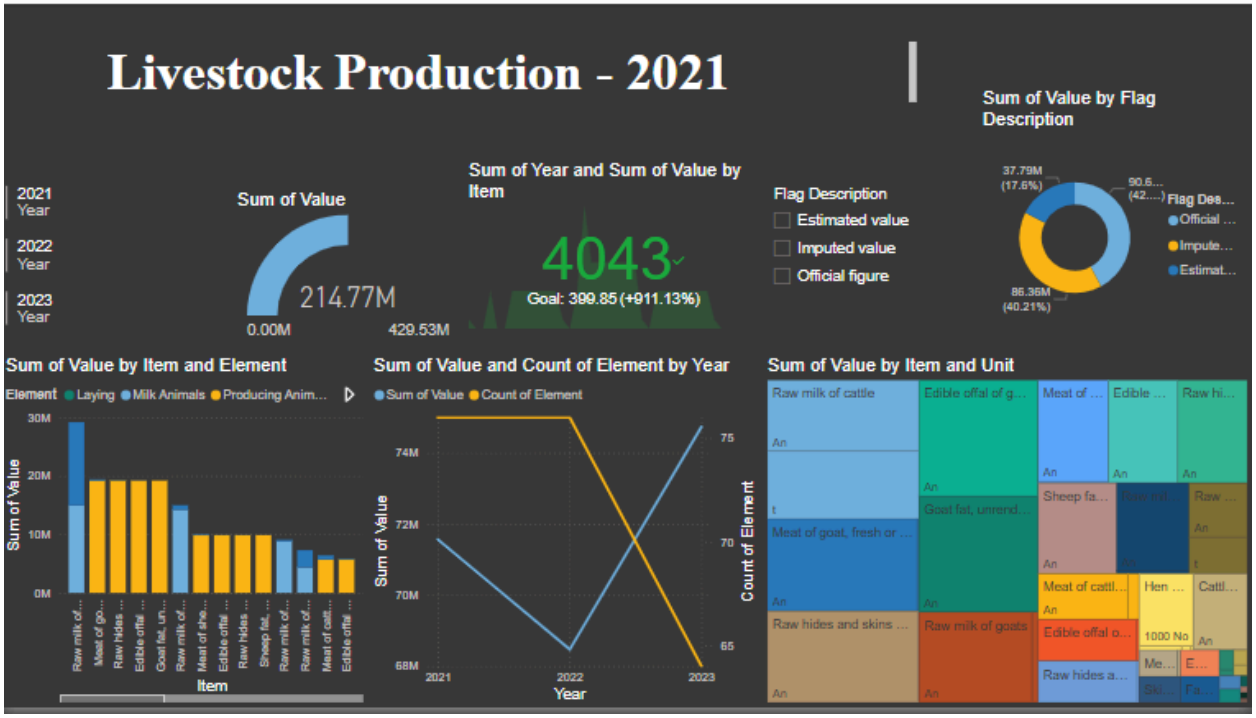
3. Exploratory Data Analysis (EDA)

3.1 Animal Product Trends by Year

Production values were analyzed across three consecutive years (2021–2023), focusing on key dairy and meat products, with data quality assessed via the Flag Description.

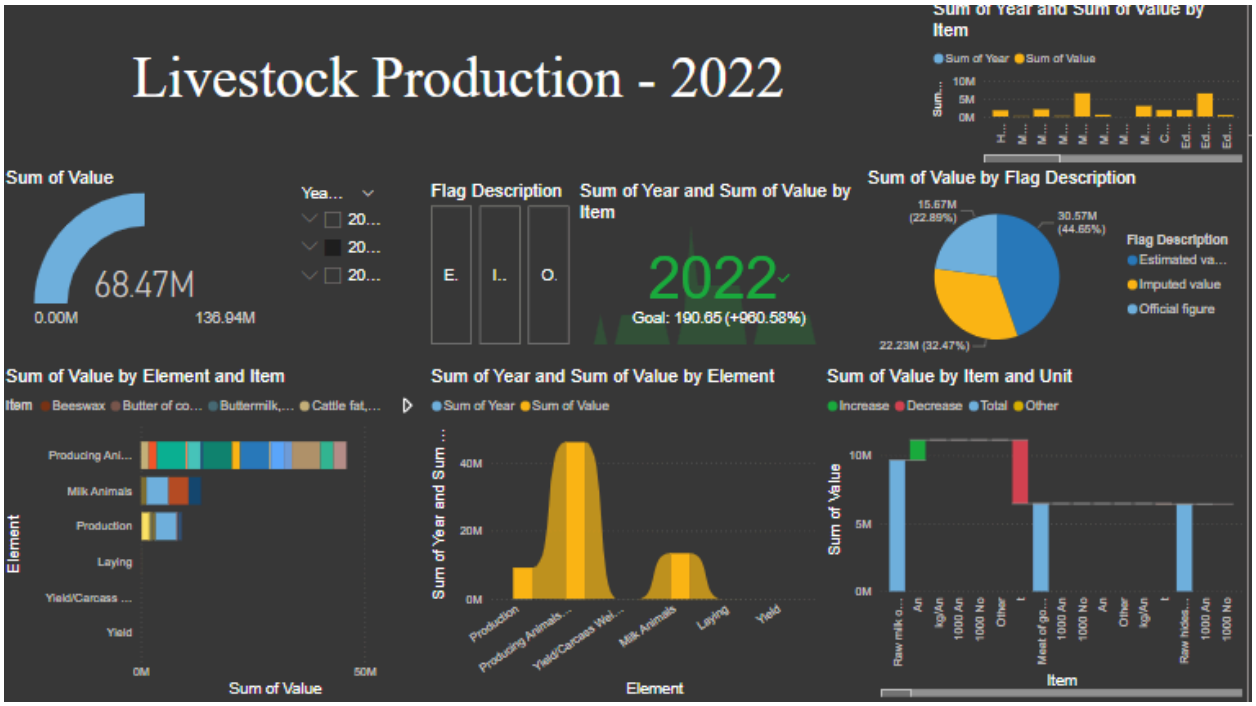
2021 Highlights:

- **Raw milk of cattle:** 5,017,991 (Animals, Official figure)
- **Raw milk of cattle:** 4,780,090 (Tonnes, Official figure)



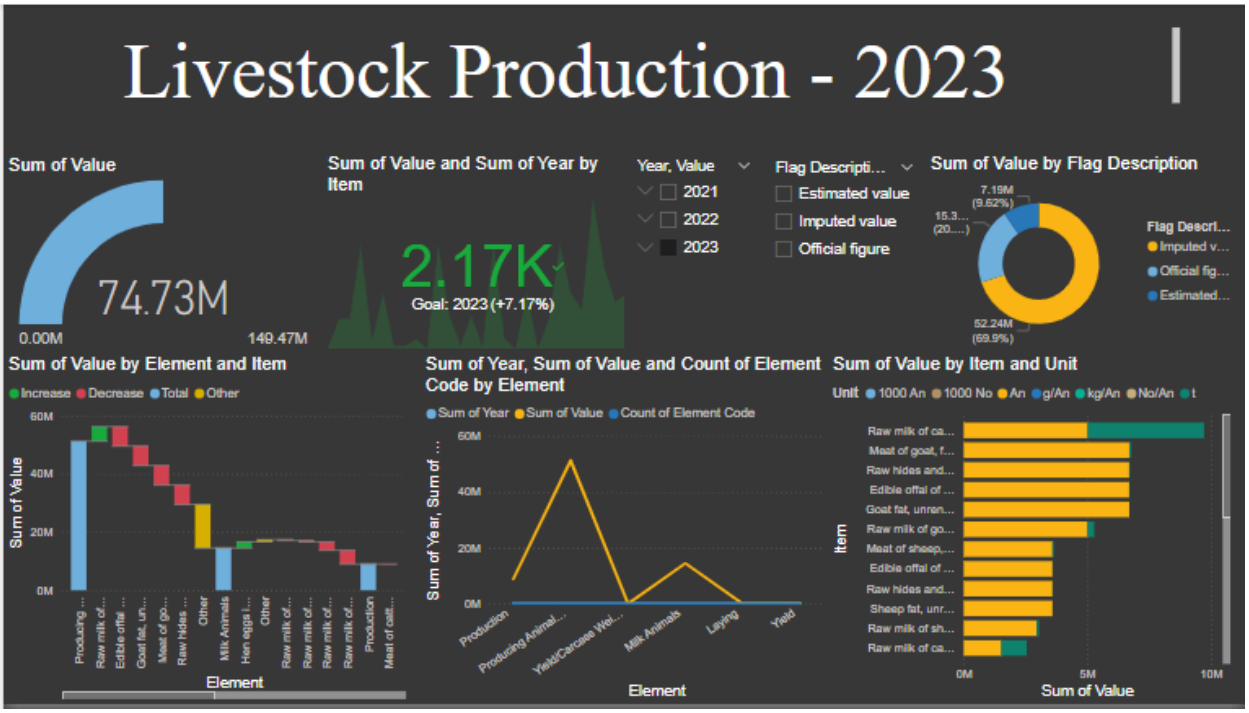
2022 Highlights:

- **Raw milk of cattle:** 4,915,833 (Animals, Estimated value)
- **Raw milk of cattle:** 4,742,430 (Tonnes, Official figure)
- **Raw milk of goats:** 4,480,390 (Animals, Imputed value)



2023 Highlights:

- **Raw milk of cattle:** 5,015,388 (Animals, Estimated figure)
- **Raw milk of goats:** 5,011,382 (Animals, Imputed value)
- **Raw milk of cattle:** 4,705,660 (Tonnes, Official figure)



3.2 Quantifiable Insights

2022

- **Raw milk of cattle increase by 0.79%**

2023

- **Raw milk of cattle increase by 0.78%**
- **Raw milk of goats increase by 11.85%**

---

4. Conclusions & Key Takeaways

Cattle milk remains the dominant livestock product, showing stable production trends over three years.

Goat milk production shows significant growth, particularly in 2023, which may reflect shifts in farming preferences or adaptive practices under climate stress.

Data Quality: The presence of a high number of *Estimated* and *Imputed* figures in 2022 and 2023 suggests either incomplete official reporting or increased modeling reliance.

Year-on-Year Comparisons revealed mild fluctuations but no catastrophic drops, suggesting overall sector stability.
