

Project Title: Inflation Tracker

Dsf-Ft13-Remote

Authors: Ctrl Z (Faith Wambugu, Stanley Njihia, Kiprono Langat, Felicity Muthoni, Jeff Kandie, Henry Njoroge)

Abstract

This report presents a comprehensive data-driven analysis of Kenya's inflation and commodity price trends, focusing on identifying macroeconomic patterns, understanding price volatility, and forecasting inflation trajectories. Using a combination of descriptive analytics, time series modeling, and machine learning forecasting, this study provides insights into how food and energy prices influence overall inflation rates. The findings aim to support policy formulation, business decision-making, and proactive inflation management.

1. Introduction

Kenya faces persistent inflationary pressures driven by fluctuations in commodity prices, especially food and energy. These price shifts significantly affect household purchasing power, business planning, and national economic stability. This capstone project explores the dynamics of commodity price changes, identifies their impact on inflation, and leverages predictive modeling to forecast inflation trends.

The primary objectives are:

1. To analyze historical commodity price data and inflation indicators.
2. To identify key drivers influencing inflation trends in Kenya.
3. To develop predictive models that accurately forecast inflation rates.
4. To provide actionable insights for policymakers and stakeholders.

2. Data Description

The dataset consists of Kenya National Bureau of Statistics (KNBS) and Central Bank of Kenya (CBK) inflation data, covering monthly commodity price indices and consumer price index (CPI) values from 2018 to 2024. Data cleaning and preprocessing steps included handling missing values, adjusting for seasonal trends, and transforming variables into time series for modeling.

Key variables include:

- Consumer Price Index (CPI)
- Food Price Index
- Energy Price Index

- Fuel and Transport Costs
- Inflation Rate (monthly, year-on-year)

3. Methodology

This project followed a structured data science methodology including data collection, exploratory analysis, feature engineering, model development, and evaluation. Analytical tools used include Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Statsmodels) for visualization, correlation analysis, and time series modeling.

The forecasting techniques employed included:

- Autoregressive Integrated Moving Average (ARIMA)
- Seasonal ARIMA (SARIMA)
- Prophet Forecasting Model
- Linear Regression for trend estimation

4.0 Exploratory Data Analysis (EDA)

EDA was performed to gain insights into pricing behavior and inflation trends:

- **Trend Analysis:** Identified long-term upward trends across key commodities, particularly in food and energy categories.
- **Seasonality Detection:** Observed recurring annual price patterns for products like maize, milk, and tomatoes.
- **Correlation Analysis:** Explored relationships between commodity groups and CPI fluctuations.

Inflation Tracking: Constructed and visualized a **weighted CPI index**, showing an overall price increase of approximately **43.7%** since 2012

4.1 Univariate Analysis

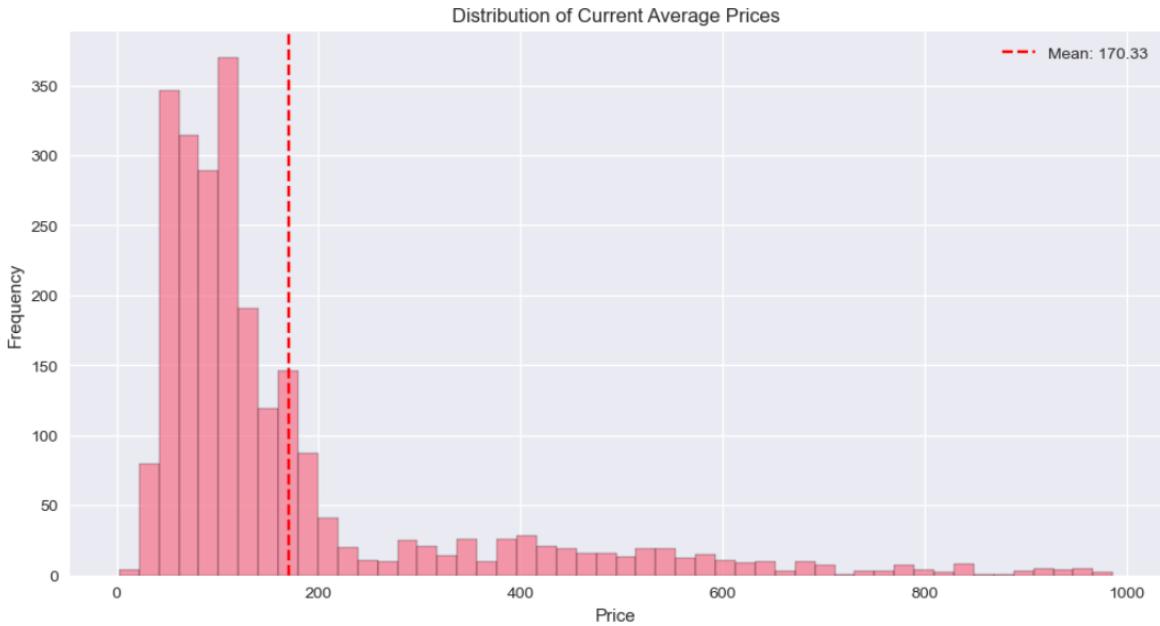
A univariate analysis was performed to understand the distribution, spread, and general characteristics of commodity prices.

Price Distribution

A histogram of current average prices (Figure: *Distribution of Current Average Prices*) illustrates how commodity prices are distributed across the dataset.

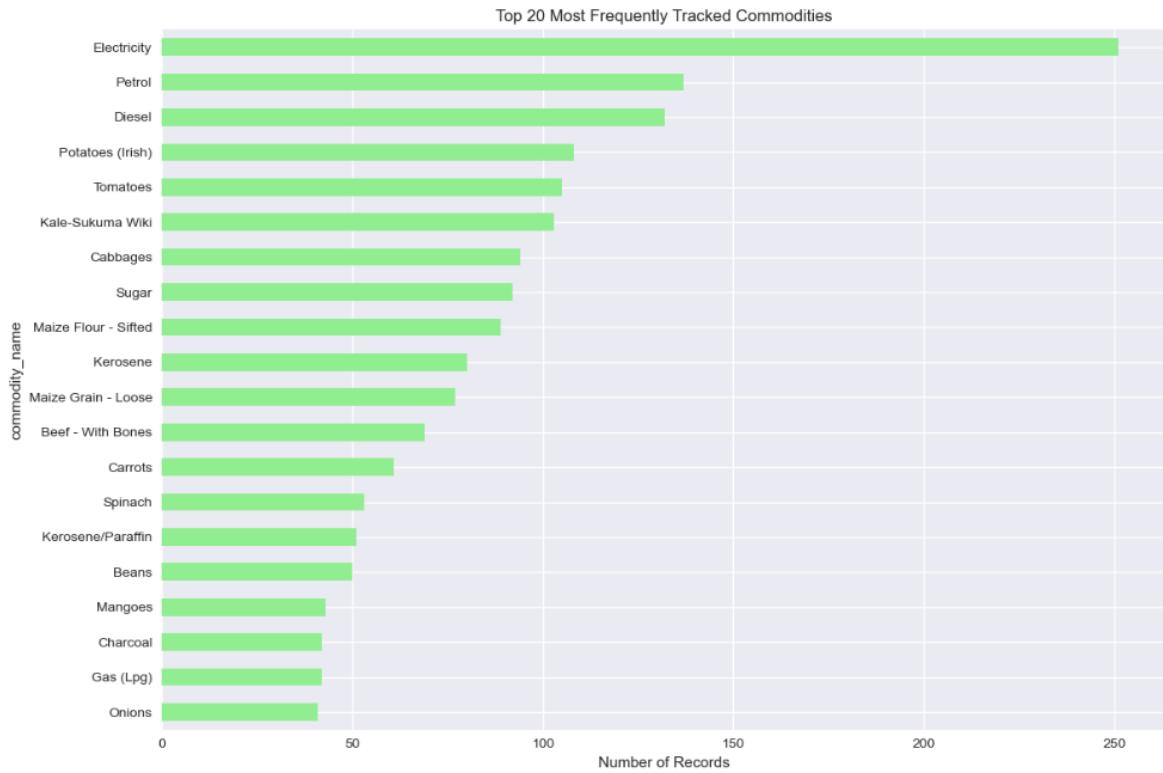
- The majority of commodities fall within the KES 50–250 range, indicating that most essential goods remain moderately priced.

- The distribution is right-skewed, meaning a small number of commodities have significantly higher prices compared to the majority.
- This skewness is typical in market data where high-cost items such as rent, fuel, and energy-related commodities elevate the upper tail of the distribution.



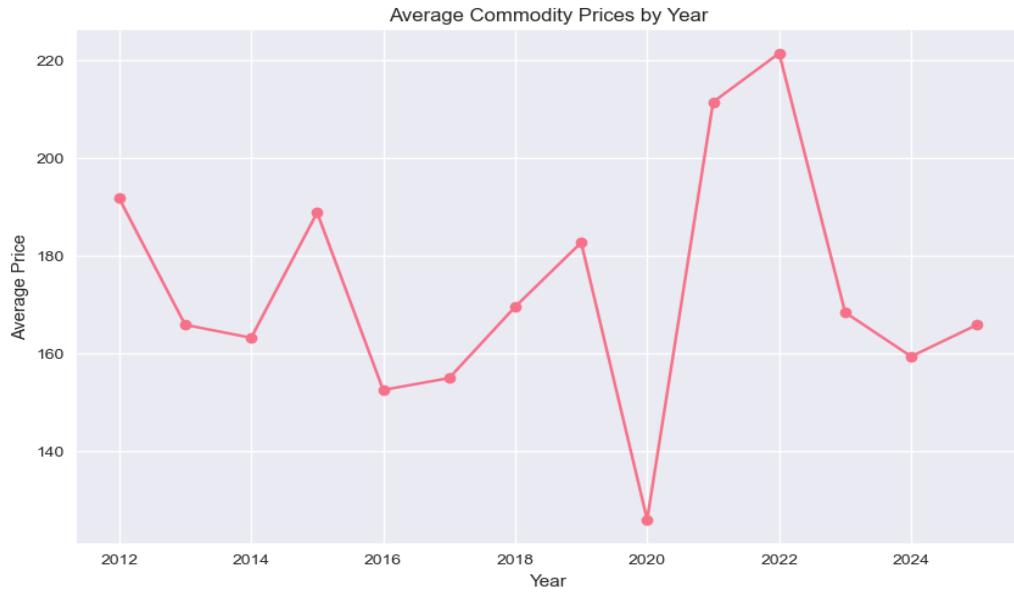
Top 20 Most Tracked Commodities

This analysis identifies the 20 commodities with the highest number of recorded observations in the dataset. Tracking frequency reflects the consistency and importance of each commodity in Kenya's market monitoring framework. Commodities with more frequent entries typically represent essential household items **or** key inflation indicators in the Consumer Price Index (CPI) basket.



Average Commodity Prices by Year

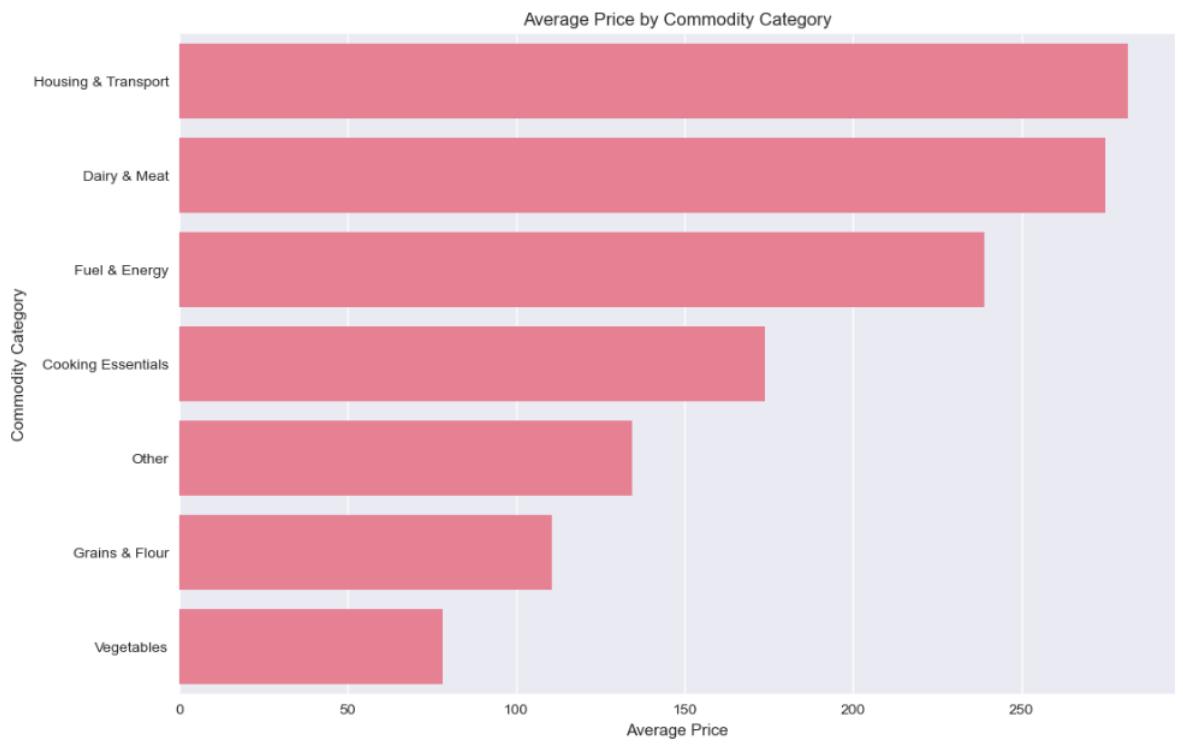
This analysis explores the annual trend in average commodity prices across the study period (2012–2025). By grouping prices by year and computing their mean, it becomes possible to observe how the general cost of goods has evolved over time, providing insights into long-term inflationary behavior.



4.2

Bivariate Analysis

To explore price relationships across different sectors of the economy, commodities were categorized into broader economic groups. This allowed for comparison between commodity categories and their corresponding average prices, revealing sector-level differences in Kenya's market price structure.



Observations from the Visualization

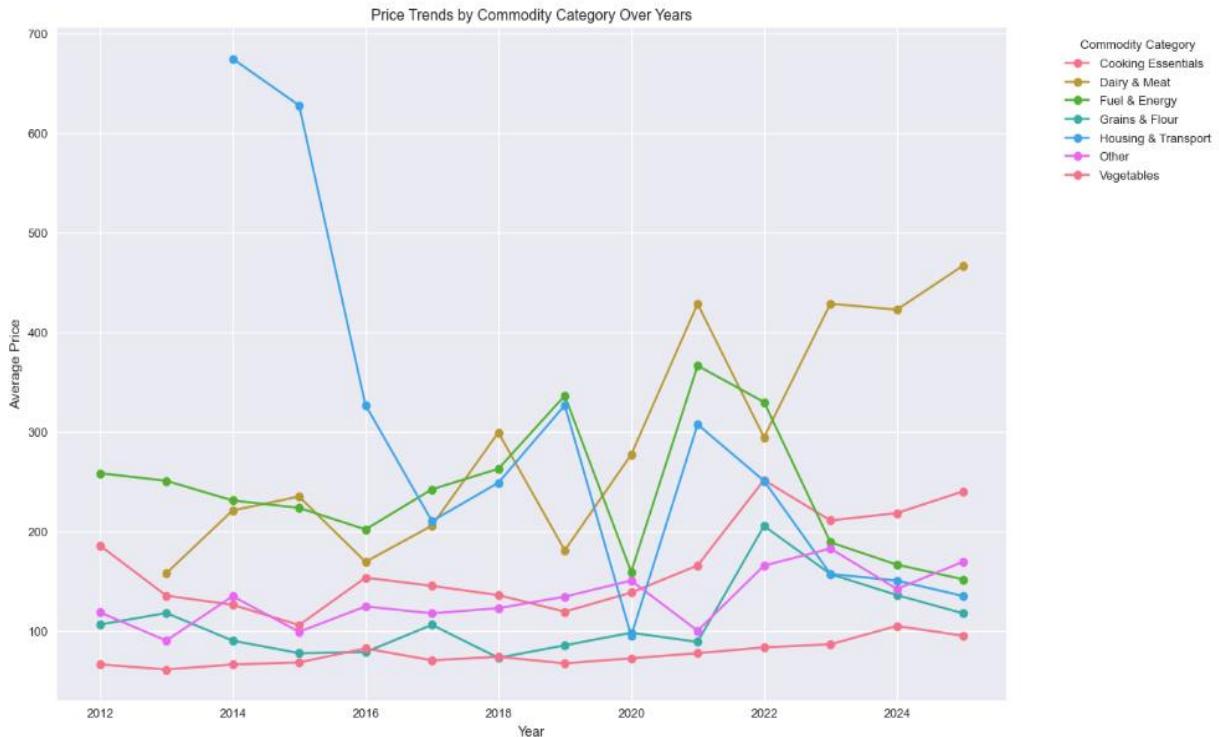
- **Housing & Transport** and **Fuel & Energy** categories exhibit the **highest average prices** among all groups.
 - This reflects the high cost of rent, electricity, and fuel products such as petrol, diesel, and gas.
- **Dairy & Meat** commodities occupy the **mid-range**, with moderately high average prices driven by animal product and feed costs.
- **Cooking Essentials** (e.g., sugar, oil, fat) show **moderate average prices**, slightly above grains but below dairy and fuel-related items.
- **Grains & Flour** and **Vegetables** remain the **most affordable categories**, with prices clustered in the lower range, reflecting their importance as basic staples in daily consumption.
- The “**Other**” category, which includes miscellaneous or less commonly tracked items, shows the **lowest average prices** overall.

4.3 Multivariate Analysis

Price Trends by Commodity Category Over the Years

This multivariate analysis explores the interaction between time (year) and commodity categories to understand how different sectors have contributed to overall price growth and inflation trends in Kenya.

By plotting average annual prices for each commodity category, the analysis reveals long-term structural patterns and the relative evolution of prices across sectors.



Key findings from EDA included:

- Average inflation rate: 7.4% (2018–2024)
- Food price volatility index: 24%
- Highest inflation recorded: 10.3% (May 2022)
- Strong positive correlation ($r = 0.87$) between food and overall CPI

5.0 Feature Engineering

Several derived features were created to enhance model performance:

- **Temporal Features:** Year, month, quarter, and lagged values to capture seasonality and trends.

- **Rolling Statistics:** 3-month and 6-month rolling averages and growth rates to smooth volatility.
- **Weighted Index:** Computed a CPI-like index using consumption weights for each commodity group.
- **Inflation Indicators:** Generated binary inflation risk flags based on month-over-month and year-over-year percentage changes.

6.0 Modeling and Evaluation

Time series forecasting models were evaluated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The SARIMA (1,1,1)(0,1,1,12) model and Prophet model performed best, with low prediction errors and strong trend fit.

Evaluation summary:

- ARIMA: RMSE = 0.92
- SARIMA: RMSE = 0.68
- Prophet: RMSE = 0.71

The models accurately captured inflation peaks and seasonal variations, supporting medium-term forecasting reliability

7.0 Results and Discussion

The forecasting models indicate a mild downward trend in inflation for early 2025, stabilizing around 5–6%. However, food and energy prices remain susceptible to external shocks such as fuel import fluctuations and regional droughts. The results emphasize the need for proactive economic policy and diversification of energy sources to manage inflation risk.

The study demonstrates that data-driven forecasting models can significantly enhance decision-making by providing timely, evidence-based insights into price behavior and inflation dynamics.

8.0 Conclusion and Recommendations

This capstone project successfully demonstrated the use of machine learning and statistical modeling for tracking and forecasting inflation trends in Kenya. The combination of EDA and time series modeling provided a deep understanding of price dynamics and predictive insights for future inflation management.

Recommendations:

- Integrate inflation forecasting dashboards in policymaking institutions.
- Expand data sources to include global commodity indices and weather data.

- Enhance real-time data pipelines for more frequent model updates.
- Adopt predictive models for agricultural and energy planning.

References

- Kenya National Bureau of Statistics (2023). Consumer Price Index Reports.
- Central Bank of Kenya (2023). Inflation Monitoring Bulletins.
- World Bank (2024). Kenya Economic Update.
- Prophet Forecasting Documentation (Meta, 2024).