BeiWatch

Economic Signal Mapper

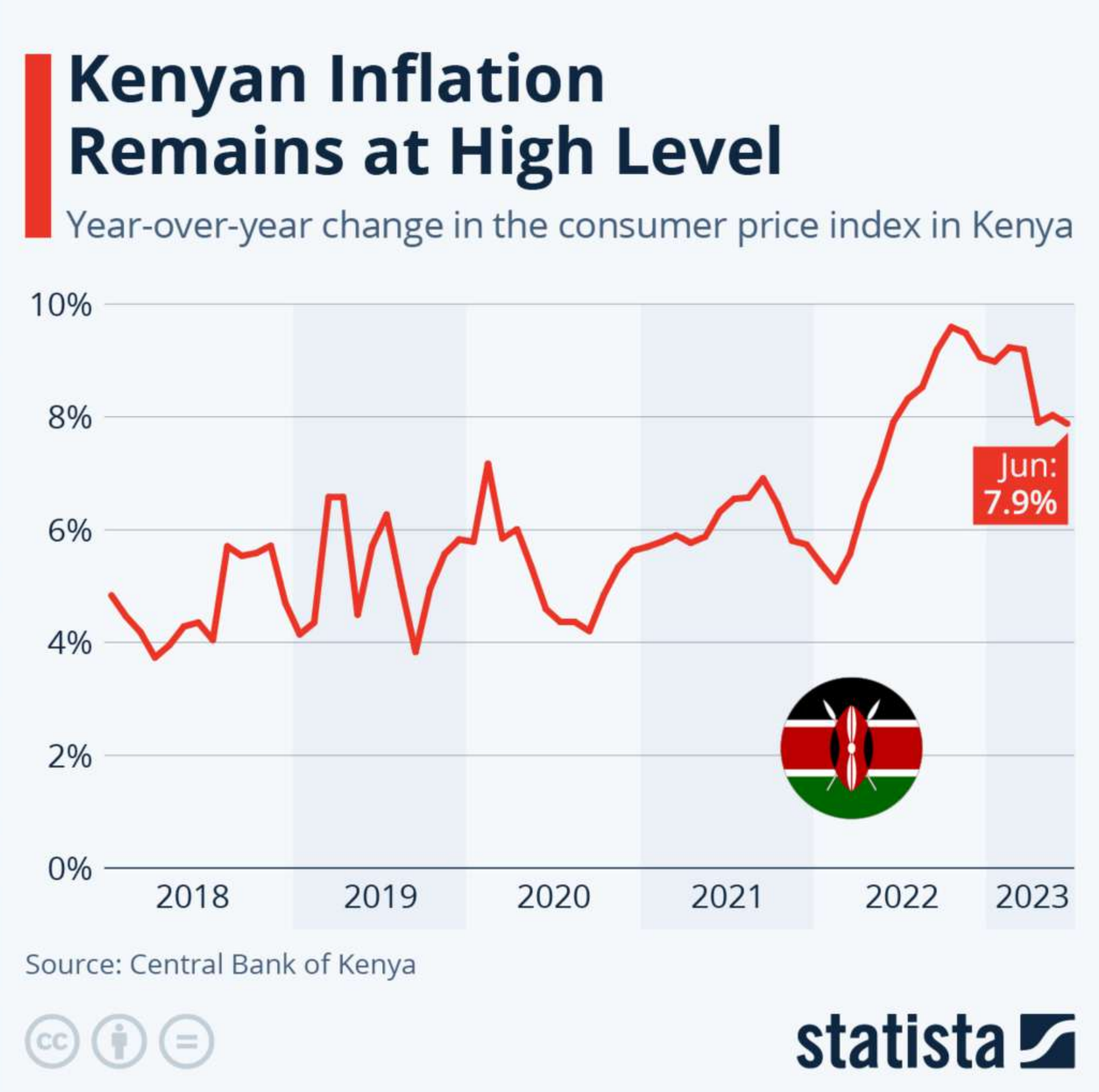
Tracking Kenya’s Price Trends

A Data Science Tool for Economic Insight and Forecasting

Presented by Group 4

Problem Statement

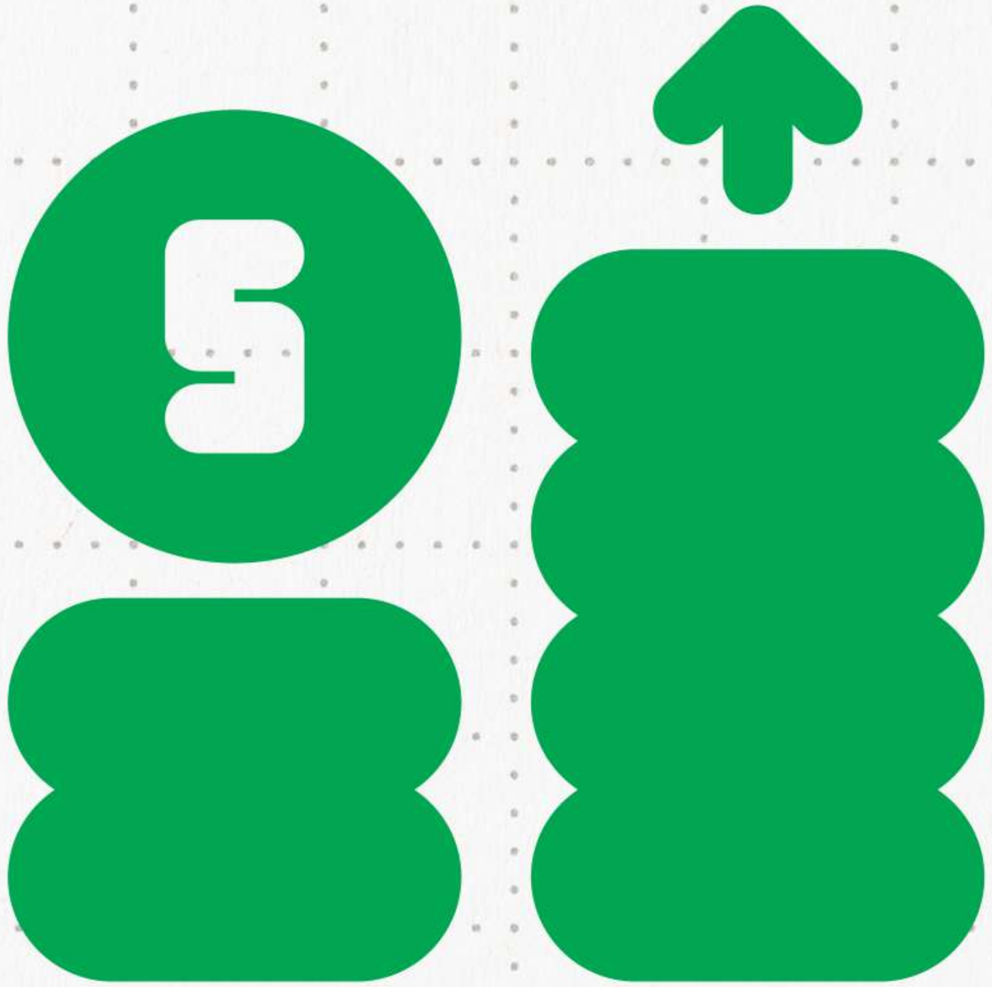
- Essential goods in Kenya (e.g., food, fuel) show frequent price shifts due to inflation, currency volatility, and global shocks.
- Price data is often delayed and buried in PDFs, leaving citizens and policymakers without real-time tools for tracking and forecasting.



Project Goal: Build a tool to track, analyze, and forecast Kenya's key price trends, empowering informed decision-making.

Project Objectives

- Collect price data from sources like KNBS & online retailers
- Clean and standardize the data for analysis
- Visualize economic trends in a dashboard
- Forecast future prices using machine learning
- Alert users of anomalies (e.g., price spikes)



Output:
An interactive platform for tracking, analyzing, and forecasting essential prices in Kenya.

Model Pipeline Overview

01

Data Scrapping

Web scraped prices from online retailers.

02

Data Cleaning

Handled missing values, outliers; normalized pricing

03

EDA

Identified trends, seasonality, and pricing patterns.

04

Forecasting

Feature Engineering: Rolling averages, date features, price change rates

Forecasting: Applied ARIMA, Prophet, and LSTM models

05

Insights

Used MAE, MAPE, RMSE for comparison

Data Sources

No.	Product	Link	Price
1	Fresh Fri Vegetable Oil 5L	https://naivas.onlir	KES 1,600
2	Avena Vegetable Oil 5L	https://naivas.onlir	KES 1,610
3	Rina Vegetable Oil 5L	https://naivas.onlir	KES 1,855
4	Chipsy Plus 3 Cooking Fat Pure Yellow 1Kg	https://naivas.onlir	KES 380
5	Chipsy Plus 3 Cooking Fat Pure Yellow 500g	https://naivas.onlir	KES 195

Online Supermarkets (Web Scraping)

Collected product prices from Kenya’s most frequented online retailers:

- Naivas Online
- Quickmart
- Carrefour

Focused on essential items: food staples, household goods, personal care, etc.

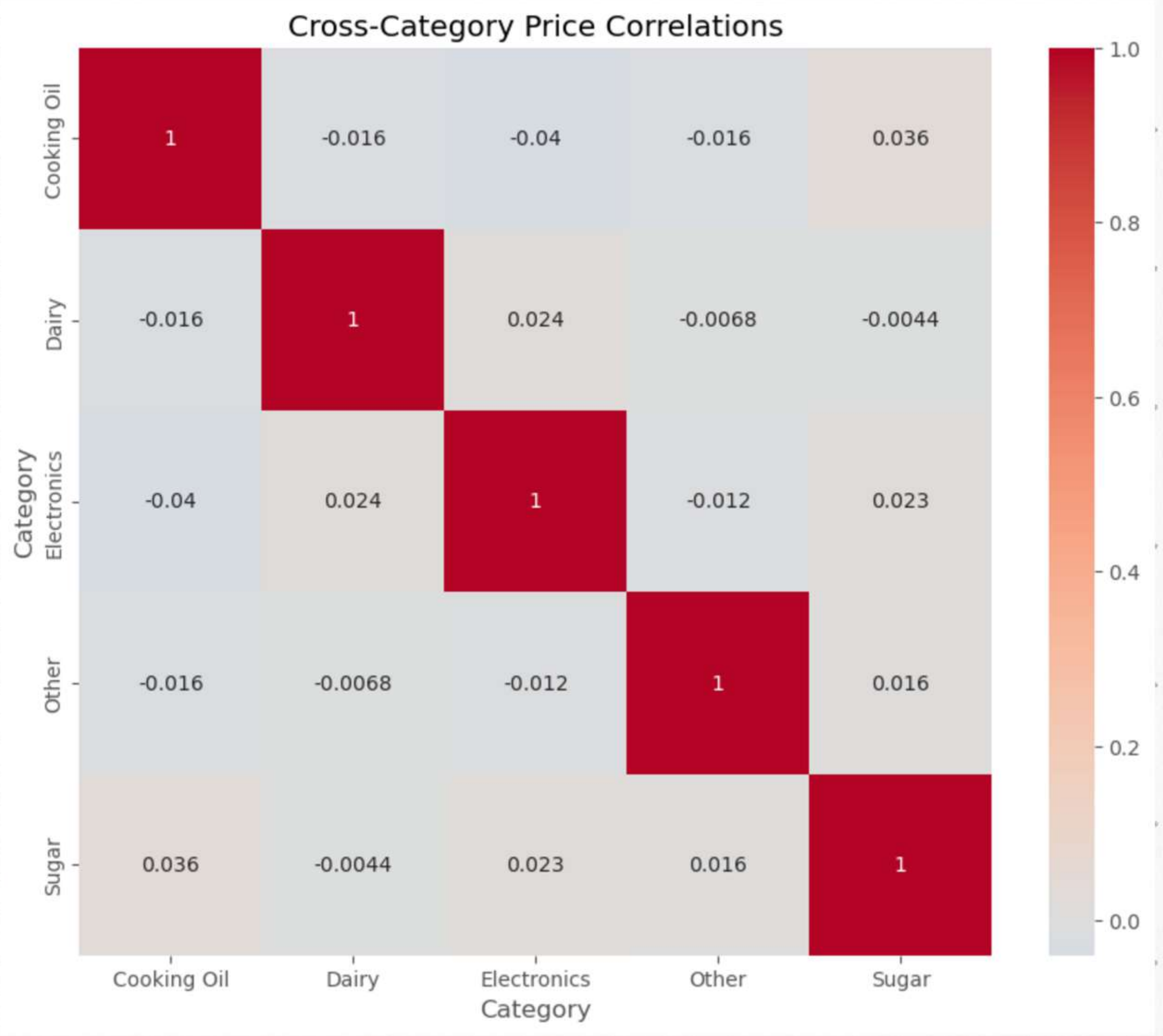
Data format: scraped and structured into CSV for analysis.

Exploratory Data Analysis

Key Findings

- **Data Coverage:**15 k daily price observations from Naivas, Carrefour & Quickmart
- **Price Distribution**: Essentials (Oil, Dairy) have tighter range; Electronics exhibit high volatility
- **Seasonal Patterns**: Monthly peaks aligned with major festivals (seen in seasonal component)
- **Anomaly Detection**: Isolation Forest flagged ~3% of outliers, cleaned before modeling
- **Category Independence:** Minimal inter-category correlation (|r|<0.05), so each series is modeled separately

Cross-Category Correlation



- **Almost zero correlation between most categories (|r|<0.05) → prices move independently.**
- **Cooking Oil & Sugar: slight positive link (r≈0.04) → shared supply / demand factors.**
- **Electronics: effectively uncorrelated with all others (|r|<0.03) → forecast each category separately.**

Forecasting Models Used

Model	MAPE	MAE	RMSE	Best For
ARIMA	5.03	45.2	61.3	Short-term, stable trends
LSTM	**6.11**	**49.8**	**65.7**	**Long-term, nonlinear patterns**
Prophet	**7.48**	**58.4**	**72.1**	**Seasonal patterns, interpretability**
Ensemble	6.08	47.1	63.2	Balanced accuracy, robust forecasts

Model Evaluation Metrics Used

MAE (Mean Absolute Error), **MAPE** (Mean Absolute Percentage Error), **RMSE** (Root Mean Squared Error)

Training Model : LSTM

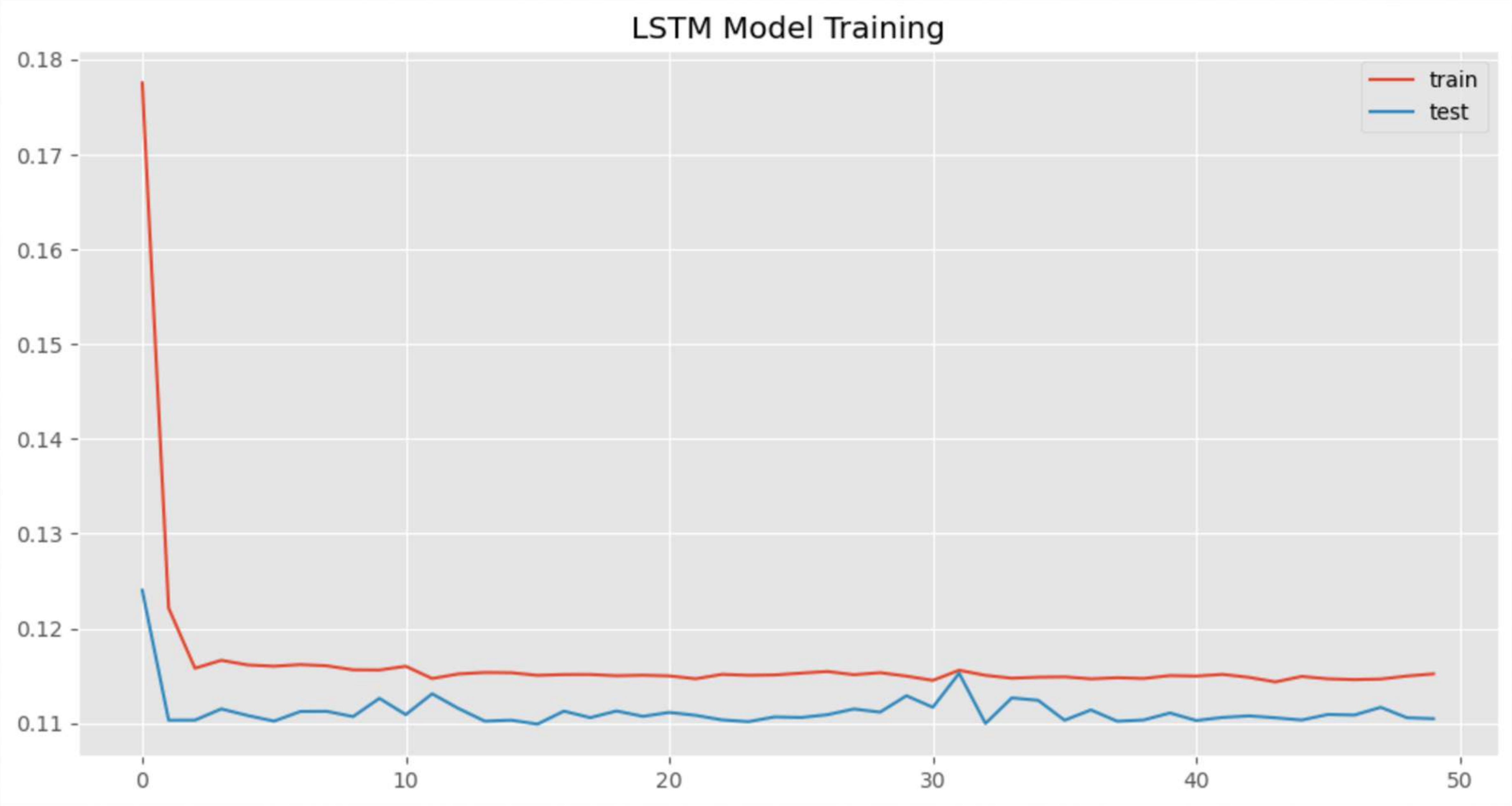
Key Reasons:

- **Handles Complexity:** Captures nonlinear price trends and long-term dependencies
- **Performs Well:** MAPE of 6.11% close to ARIMA, with strong generalization
- **Adaptable:** Scalable across different product categories and dynamic markets

Why Not ARIMA or Prophet?

- **ARIMA: Slightly better accuracy, but limited to simple, stable patterns**
- **Prophet: Higher error, better suited for seasonal data only**

LSTM offers a strong balance between performance and flexibility, making it ideal for forecasting diverse and evolving price trends.



The chart shows the model's "error" rate dropping sharply in the first few cycles, then settling at a steady, low level for both training (red) and testing (blue). This means the **LSTM learned quickly and makes similarly small errors on new data, so it's both fast and reliable without over-fitting.**

Timeline (Next Steps)

Timeframe	Action Item	Owner
Aug–Sep 2025	• Integrate LSTM model into BeiWatch backend• Validate forecasts on live data	Data Engineering
Sep–Oct 2025	• Develop & UX-test interactive dashboard• Implement date/category filters	Product & Design
Oct 2025 (1 month)	• Build alerting service (threshold set at ±5%)• Pilot alerts with select users	Dev Ops & Analytics
Nov 2025 onwards	• Automate monthly retraining pipeline• Monitor key metrics (MAPE, MAE)• Refine models based on performance	Data Science Team
Q1 2026	• Extend data sources (fuel, FX rates, new retailers)• Explore hybrid ensemble experiments	R&D

Recommendations

- Deploy BeiWatch’s LSTM-powered forecasts into production
- Build an interactive BeiWatch dashboard with live trends, confidence bands, and filters
- Automate alerts when actual prices deviate >5% from forecasted values
- Retrain the LSTM monthly and monitor MAPE/MAE for performance

For more information

Have a look at our GitHub Repo:

BeiWatch Repo

CONNECT WITH US

1. Bernice Kigochi
2. Vanessa Wambui
3. Eric Otieno
4. Lilian Kwamboka
5. Tim Kirui