Introduction & Dataset Overview

Introduction & Objective

The goal of this project is to analyze global food prices and segment markets based on price behavior using clustering techniques. This can help policymakers, NGOs, and researchers understand price trends and detect market anomalies or food insecurity zones.

Dataset Source

- **Dataset**: Global Food Prices Database
- **Source**: World Food Programme (via Kaggle)
- **Link**: https://www.kaggle.com/datasets/salehahmedrony/global-food-prices

Variables of Interest

- **Country**: Name of the country.
- **Market**: Specific locality or market name.
- **Commodity**: Type of food (e.g., rice, beans, maize).
- **Price**: Market price for the item.
- **Currency**: Currency in which the price is recorded.
- **Unit**: Quantity measure used (e.g., kg, liter).
- **Date**: Month and year of the price recording.

Coverage

- **Countries Covered**: 76
- **Time Span**: 1992–2023 (mostly post-2003)
- **Markets**: ~1,500 local markets

Data Cleaning & Processing

Selected Columns

We extracted and standardized the following variables:

- adm0_name → country
- mkt_name → market
- cm_name → commodity

- mp_price → price
- cur_name → currency
- um_name → unit
- mp_month & mp_year → date

Cleaning Steps

- Dropped rows with nulls in essential fields (price, market, commodity).
- Created a date column using month and year.
- Filtered outliers and non-standard units where needed.

Encoding for Clustering

- Used LabelEncoder for categorical variables: country, commodity
- Standardized numerical features using StandardScaler for clustering.

Exploratory Data Analysis (EDA)

Visualizations

1. **Top 10 Countries with Most Records**
    a. Nigeria, Pakistan, Ethiopia, Bangladesh, etc.
    b. Indicates frequent reporting and monitoring.
2. **Top 10 Commodities by Average Price**
    a. Fish, Meat, Wheat Flour, Lentils, Rice, etc.
    b. Protein-rich commodities tend to be most expensive.
3. **Rice Price Trend in Bangladesh**
    a. Price fluctuations show seasonal spikes.
    b. Sharp increases during certain years, potentially indicating inflation or supply chain disruptions.

Insights

- **Frequent Reporters**: African and South Asian countries dominate.
- **Volatile Commodities**: Protein-rich and imported items.
- **Inflationary Trends**: Detected in several countries over 5+ year spans.

Market Segmentation

Clustering with KMeans

- Applied KMeans with 4 clusters (chosen via elbow method).

Features Used

- Encoded country, commodity, and price
- Standardized all values before clustering

Cluster Interpretation

- **Cluster 0**: Low-price, stable commodities in developing regions
- **Cluster 1**: Mid-price range, mostly grains and beans
- **Cluster 2**: High-price, high-demand items like meat and fish
- **Cluster 3**: Volatile price regions, potential food crisis zones

Cluster Visualization

- Plotted clusters using commodity vs price and country vs price
- Clear segmentation observed among countries and commodities

Key Insights & Conclusion

Summary of Findings

- **Common Commodities**: Rice, maize, beans
- **Regional Focus**: Heavy presence of African and South Asian markets
- **Volatility**: Observed in countries facing economic instability or conflict
- **Clusters**: Help separate low-price stable markets from high-price volatile ones

Recommendations

- NGOs should monitor Cluster 3 markets for food insecurity.
- Governments can use segmentation for targeted subsidies.
- Future work can include forecasting models for price prediction.

Limitations

- Some markets lack recent or consistent data.
- Commodity units are not always standardized.
- No inflation adjustment across years.