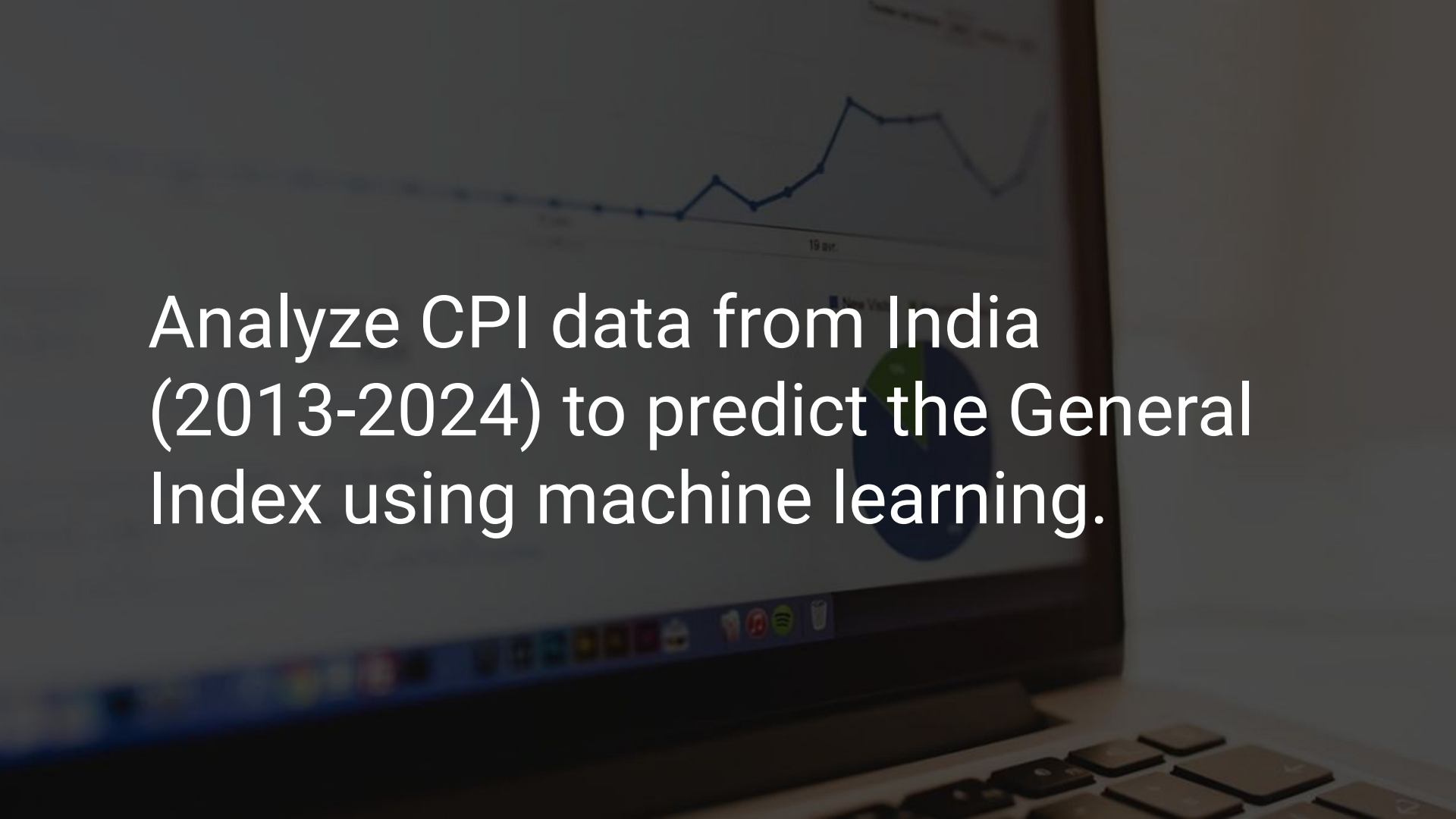

Analyze CPI data from India (2013-2024) to predict the General Index using machine learning.

Objective

- Analyze price trends in India’s Consumer Price Index (CPI).
- Predict the General Index using machine learning models.

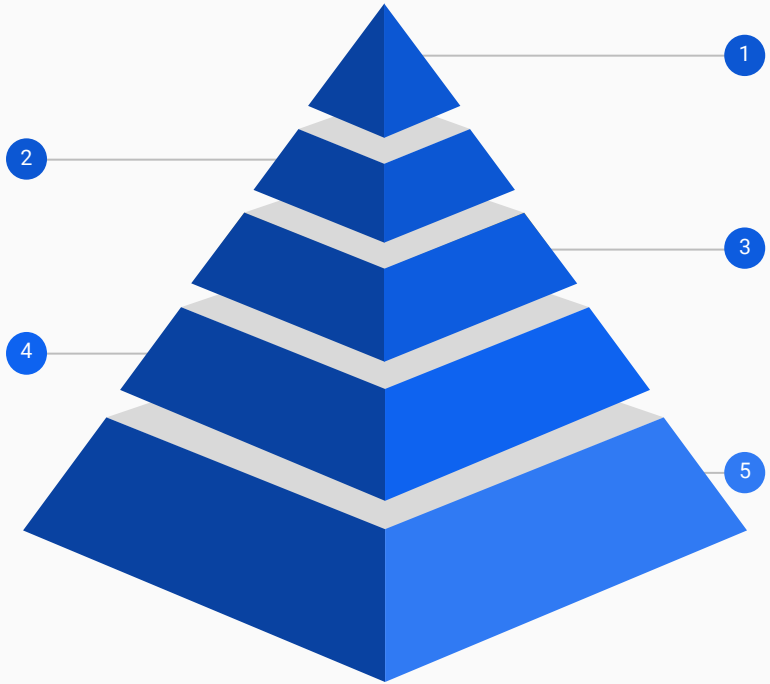
Key Actions:

Data Preparation

Loaded the CPI dataset, converted data types, and handled missing values using **linear interpolation** and **backward fill**. Removed outliers from relevant columns and merged redundant columns.

Model Building and Evaluation

Trained **Linear Regression** and **Decision Tree** models to predict the General Index. Evaluated models using **Mean Squared Error (MSE)** and **R-squared**, with Linear Regression outperforming Decision Tree.



Data Import and Library Setup:

Loaded libraries like pandas, matplotlib, seaborn, and scikit-learn. Data was imported using pandas from a CSV file.

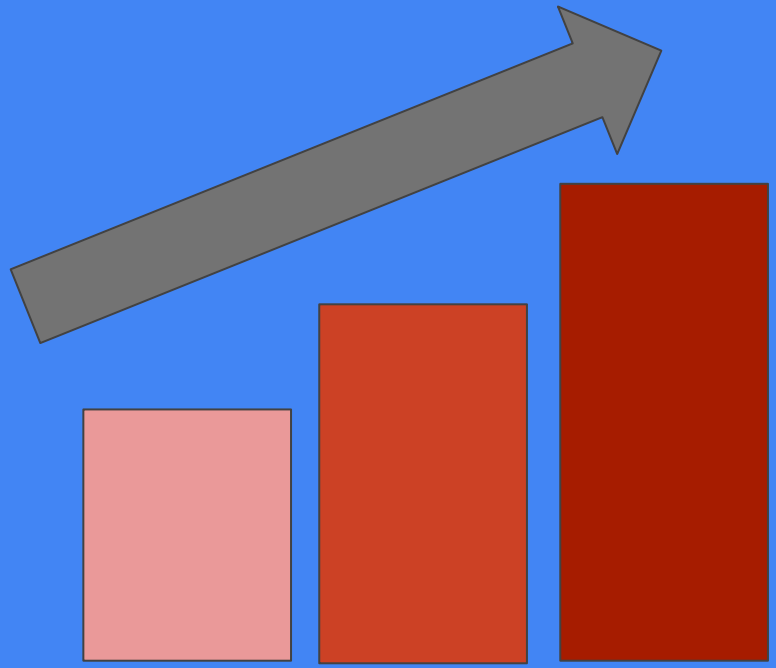
Exploratory Data Analysis (EDA)

Conducted **descriptive analysis**, created **scatter plots**, and **analyzed yearly** and **monthly trends**. A **correlation matrix** was generated to reveal relationships between CPI components and the General Index

Feature Importance and Insights

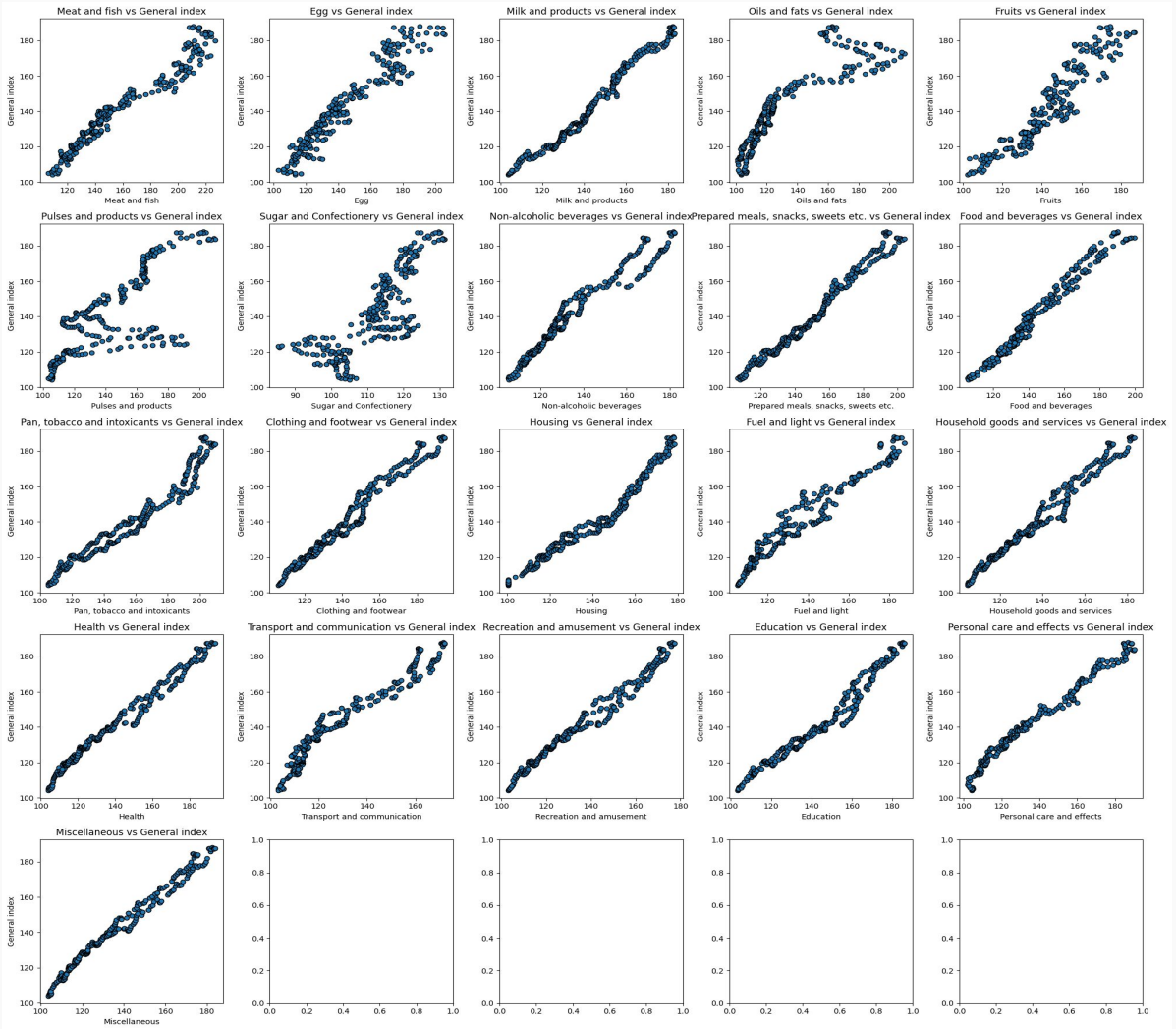
Analyzed the coefficients from Linear Regression to identify key factors impacting the General Index, such as 'Food and beverages', 'Housing', and 'Fuel and light'. **Visualized feature importance** and trends.

Key Insights from EDA



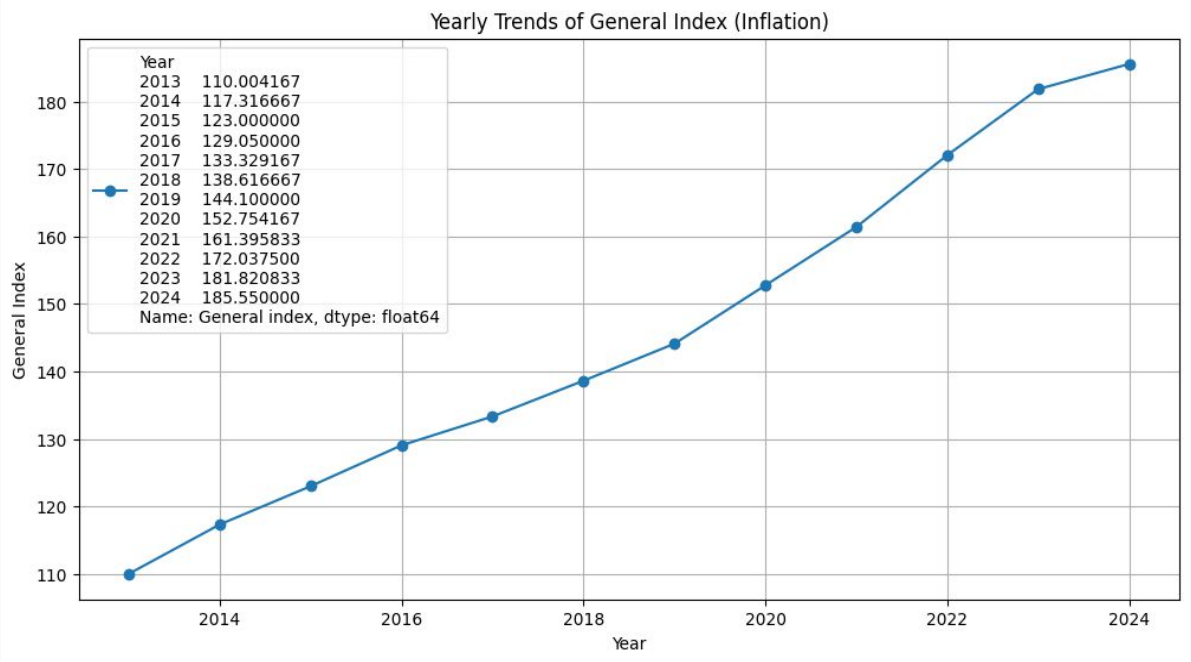
Scatter Plot

- Several features, such as **Housing**, **Miscellaneous**, **Health**, **Clothing and Footwear**, and **Fuel and Light**, show a strong positive linear relationship with the General Index. This indicates that as the values of these categories increase, the General Index also increases proportionally, suggesting these are key drivers of inflation.
- Features like **Pulses and Products**, **Prepared Meals, Snacks, and Sweets**, and **Non-Alcoholic Beverages** show more scatter, suggesting that their relationship with the General Index is weaker or non-linear. These categories may have less predictive power or exhibit different price behavior over time.



Yearly Trends

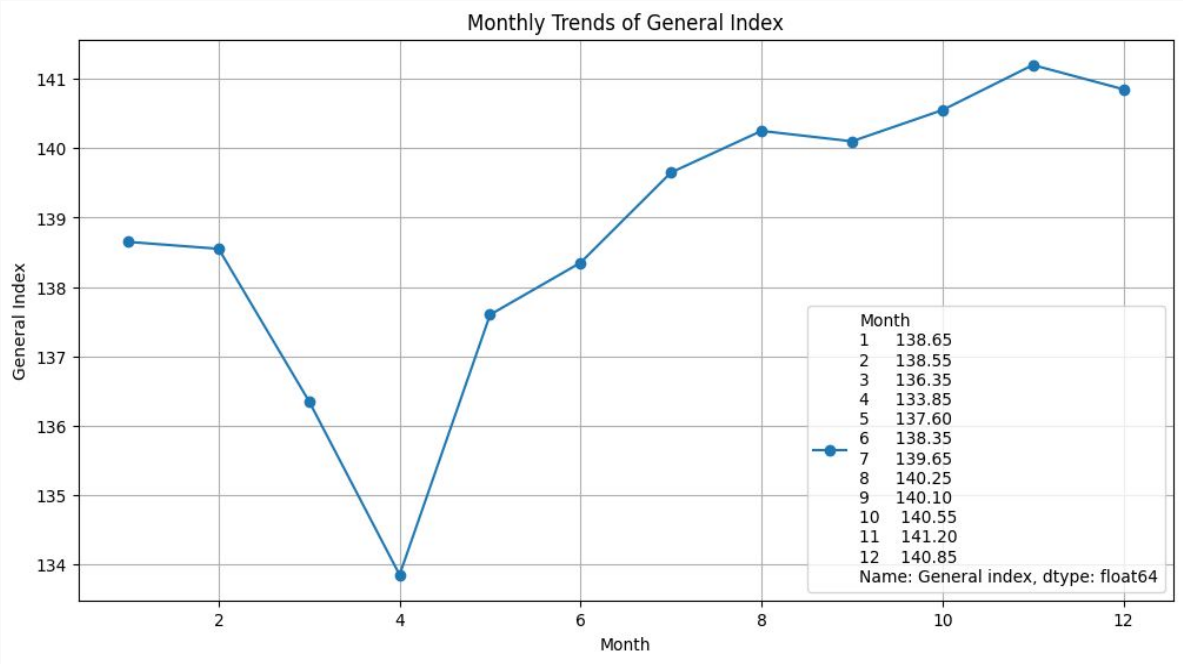
- **Steady Increase in the General Index**:The General Index shows a steady upward trend from 2013 to 2024, indicating a consistent rise in prices over time. This reflects ongoing inflation, with the General Index rising from around 110 in 2013 to nearly 186 by 2024.
- **Economic Influence**:The consistent rise in the General Index across years suggests the overall price level of goods and services continues to increase, influencing both economic policies and household purchasing power. This may lead to adjustments in fiscal policies, interest rates, or market strategies to mitigate inflation.



Monthly Trends

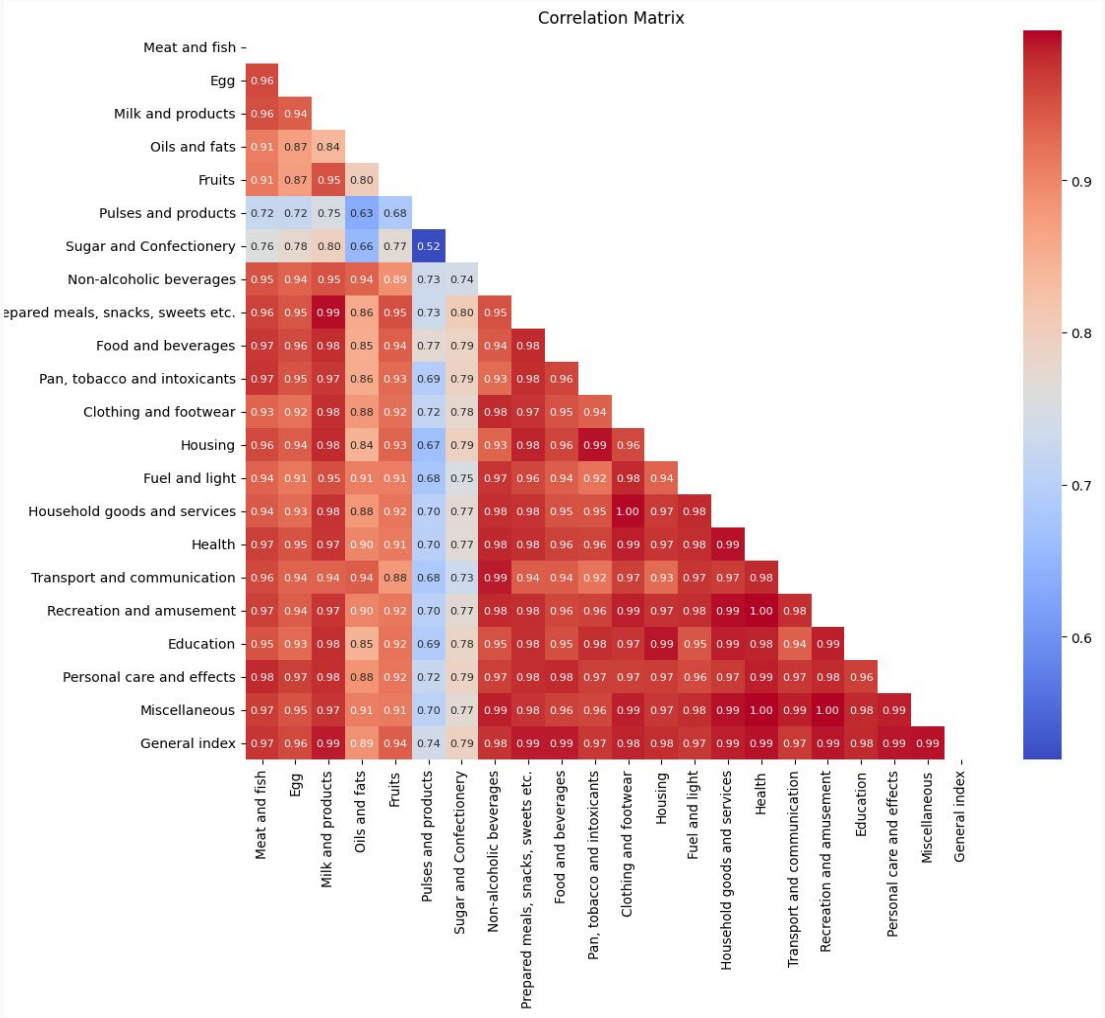
This chart reflects **persistent seasonal patterns** in the General Index across multiple years:

- **Early-year stability**, followed by a **spring decline**.
- A **strong recovery in summer**, with **peak inflation** in the fall.
- A slight **easing at year-end**.



Correlation Matrix

This matrix reveals that **housing, food, and essential services** play a dominant role in influencing the General Index. The strong interconnections between categories suggest that inflationary trends in one area, particularly housing or food, are likely to impact other areas as well, reinforcing their importance in the overall CPI calculation.



Models Training & Evaluation

Linear Regression

- Mean Squared Error (MSE) - `0.19`
- R-squared (R^2) - `0.9996`

Decision Tree

- Mean Squared Error MSE - 1.67
- R-squared (R^2) - 0.997

Feature Importance

Top Influential Features:

- *Housing, Fuel & Light, Food & Beverages, and Clothing & Footwear were the most impactful components on the General Index.*
- *Scatter plots confirmed a linear relationship between these features and the General Index.*

About Me :

Name: Sachin Guria
Linkedin: https://www.linkedin.com/in/sachin-guria-data-analyst/
Background: Data Science enthusiast with experience in machine learning, data visualization, and statistical analysis. Passionate about solving real-world problems using data.
Key Skills:

- Python, SQL, Pandas, Matplotlib, R
- Machine Learning
- Data Visualization(Tableau) & Analysis
- Additional Tools : Google Docs, Google Sheets,Google Slides,Google Cloud Platform

Future Aspirations: Committed to continuous learning and professional development in data analytics, with particular interest in:

- Exploring new machine learning techniques
- Enhancing forecasting accuracy
- Developing more effective data visualization methods
- Contributing to impactful data-driven projects