SANNIDHYA DAS

ST. XAVIER'S COLLEGE (AUTONOMOUS),KOLKATA

DATE: 31.03.2025

TOPIC: Analysis and Data Visualization of Food & Beverage Industry Trends: Insights into Consumer Preferences and Market Dynamics

: Methodology :

Data Cleaning:

Step 1: Identifying Missing Values

To ensure data quality, we first identified missing values across all columns using the colSums(is.na(df)) function. This provided an overview of the extent of missing data in each variable.

Step 2: Dropping Irrelevant Columns

Certain columns were deemed unnecessary for analysis and were removed from the dataset using:

This step reduces redundancy and enhances model efficiency by retaining only relevant features.

Step 3: Understanding Data Distribution

To gain insights into the dataset, we performed a summary and structure check using:

This helped in identifying the data types, distributions, and potential inconsistencies in the dataset.

Step 4: Filtering Data for Arunachal Pradesh

To address missing values specific to the state of Arunachal Pradesh, we extracted its subset using:

From this, we observed the following:

- The **Urban** column contained only NA values.

- The **Combined** column contained only values from the **Rural** column.

Step 5: Imputation of Missing Values

Based on our observations, we implemented targeted imputation strategies:

1. **Replacing NA values in the Urban column**
   Since Arunachal Pradesh had no recorded values for the Urban category, all missing values were replaced with 0, assuming no significant urban population contribution.
2. **Adjusting the Combined column**
   As the **Combined** column only reflected rural values, we replaced it with the corresponding values from the **Rural** column.

Analyzing and Visualizing All India CPI Data and State Comparisons :

This analysis aims to explore and compare the Consumer Price Index (CPI) across Indian states, with a particular focus on the "ALL India" CPI and the disparities between rural and urban areas. The methodology involves data extraction, processing, and visualization to uncover trends, disparities, and insights related to inflation dynamics.

1. Data Preparation:

- **Filtering Data for "ALL India" CPI**: CPI data for "ALL India" was isolated from the dataset, and a new Date column was created by combining the Year and Month columns.
- **State Comparison**: CPI data from individual states was joined with the "ALL India" CPI to allow for direct comparisons. The CPI difference between each state and the national average was computed and visualized.
- **Date Formatting**: A proper Date column was created for better time-series analysis.

2. Visualizing CPI Trends:

- **All India CPI Line Plot**: A line plot was created to visualize the trend of "ALL India" CPI over time, showcasing the national inflationary trend.
- **State vs. All India CPI Comparison**: Line plots were generated for select states (Bihar, Delhi, Gujarat, West Bengal) to visualize how their CPI compares with "ALL India," highlighting over- or under-performance.

3. Identifying Top Performing States:

- **Top States Based on CPI Differences**: The states with the highest and lowest CPI differences from the national average were identified. Year-wise rankings of top states were also created to track consistent performers.

4. Rural vs Urban CPI Analysis:

- **CPI Differences Between Rural and Urban Areas**: The rural-urban CPI differences were computed, with positive values indicating higher rural inflation. The average differences were analyzed at the state level.
- **Trends and Outliers**: Line plots tracked rural and urban CPI trends over time, while boxplots helped identify outliers. The volatility and correlation between rural and urban CPI were analyzed.

5. Additional Insights:

- **Purchasing Power Analysis**: The impact of inflation on purchasing power was assessed by calculating how ₹100's value changed over time, reflecting inflation's effect on consumer spending power.
- **Inflation During Festive Seasons and COVID-19**: Seasonal and pandemic-related inflationary effects were visualized to understand how inflation fluctuates during festive months and global crises.
- **Cost of Food and Inflation**: The rise in food expenses was tracked through the Cost_Index, which measures CPI's impact on food costs over time.

6. Inflation Rate and Volatility:

- **Inflation Rate**: Year-over-year CPI inflation rates were calculated and visualized to understand inflationary patterns over time.
- **Volatility**: The standard deviation of CPI was computed for each state to identify areas with the highest inflation instability.

7. Heatmaps and Correlation:

- **CPI Heatmap**: A heatmap visualized the variation of CPI across states and years, providing a color-coded view of inflation trends.
- **Rural vs Urban CPI Gap**: Bar plots compared the rural-urban CPI gap by state, revealing the disparities in inflation between these areas.

This methodology provides a comprehensive approach to understanding CPI dynamics across India, offering valuable insights into regional inflation trends, volatility, and the impact of inflation on different sectors like food and consumer purchasing power.

Inflation Trends and CPI Analysis :

This methodology utilizes multiple analytical techniques to examine inflation trends across Indian states, focusing on categorical analysis, time series forecasting, CPI volatility, stationarity, and clustering of states based on similar inflation behavior. The goal is to explore whether inflation trends are state-dependent, forecast CPI trends, analyze CPI stability, and identify groups of states exhibiting similar inflation patterns.

1. Chi-Square Test for Inflation Trends

Objective:

The goal of this test is to examine whether inflation trends (whether CPI is increasing or decreasing) are independent of the state.

Approach:

A categorical analysis was performed where inflation trends were categorized as either "Increase" or "Decrease" based on whether the Combined CPI of the current period was greater than the previous period. This trend was then analyzed using the Chi-Square test to assess whether there is a dependence between inflation trends and states.

The null hypothesis (H0) is that inflation trends are independent of the state, while the alternative hypothesis (H1) posits that inflation trends depend on the state.

Result:

- The p-value obtained from the Chi-Square test is **less than 0.05** (p-value < 2.2e-16), which leads to rejecting the null hypothesis.
- Conclusion: Inflation trends **do depend on the state**, indicating that inflation dynamics are not uniform across states.

2. Time Series Analysis for CPI Trend Forecasting

Objective:

To forecast future CPI trends for "ALL India" using ARIMA (AutoRegressive Integrated Moving Average) and analyze the time series properties of CPI.

Approach:

- **Data Preparation:** A time series object was created using the CPI data for "ALL India," starting from January 2013 and having a monthly frequency.
- **ARIMA Modeling:** The auto.arima() function was applied to the time series to automatically select the best ARIMA model based on AIC (Akaike Information Criterion) and other performance measures.
- **Forecasting:** The fitted ARIMA model was used to forecast CPI values for the next 12 months.

Result:

- The ARIMA model produced a forecast of the future CPI, and the plot displayed the predicted CPI trend.
- Conclusion: The ARIMA model provides a forecast of inflation for the next year, which helps in understanding potential future inflation trends.

3. CPI Stability Analysis

A. Standard Deviation of CPI Over Years

Objective:

To identify which states exhibit the most volatile CPI by calculating the standard deviation of CPI for each state.

Approach:

- **Data Grouping:** The data was grouped by State, and the standard deviation of the Combined CPI was calculated for each state.
- **Result Interpretation:** States with a higher standard deviation of CPI exhibit greater volatility, suggesting that these states experience more instability in inflation.

Result:

- States like **Manipur**, **Lakshadweep**, **Andaman**, **Telangana**, and **West Bengal** were found to have high CPI volatility.
- Conclusion: High volatility in these states indicates inconsistent inflation, potentially driven by local economic conditions or external factors like policy changes or supply chain disruptions.

B. CPI Stationarity Test (ADF Test)

Objective:

To test whether the CPI time series is stationary, as stationarity is a prerequisite for certain time series models.

Approach:

- The **Augmented Dickey-Fuller (ADF) test** was conducted on the CPI time series to assess its stationarity.
- **Null Hypothesis (H0)**: The time series has a unit root (i.e., the series is non-stationary).
- **Alternative Hypothesis (H1)**: The time series is stationary.

Result:

- The p-value from the ADF test was **0.7091**, which is greater than 0.05.
- Conclusion: The CPI series is **non-stationary**, indicating the presence of trends or seasonal components. Differencing the series may be required to achieve stationarity for certain forecasting models.

4. Clustering Analysis: Identifying Similar States

Objective:

To group states based on similar inflation patterns, using K-means clustering based on the average CPI of each state.

Approach:

- **Data Preparation:** The data was grouped by State, and the mean CPI for each state was computed.
- **K-means Clustering:** The states were clustered using the kmeans() algorithm, with the number of clusters set to 3, reflecting high, medium, and low inflation groups.
- **Cluster Visualization:** A scatter plot was generated to visualize the states and their respective clusters.

Result:

- **Cluster 1 (High inflation)** includes states such as **Andaman & Nicobar**, **Kerala**, **Manipur**, and **Puducherry**.

- **Cluster 2 (Medium inflation)** includes states like **Andhra Pradesh**, **Karnataka**, **Telangana**, **West Bengal**, and others.
- **Cluster 3 (Low inflation)** includes states like **Assam**, **Bihar**, **Delhi**, **Gujarat**, and others.

Conclusion:

Clustering groups states into zones of high, medium, and low inflation, which can guide targeted policy interventions. States in the high inflation cluster may require fiscal measures to control price volatility, while states in the low inflation group may indicate stable economic conditions.

Overall Insights and Conclusion

1. **Inflation Trends and State Dependency:** The Chi-Square test revealed that inflation trends are not independent of the state, suggesting that each state has unique inflation characteristics influenced by local factors.
2. **Time Series Forecasting:** The ARIMA model allowed for the forecasting of CPI trends, providing valuable information for future policy planning and economic forecasting.
3. **CPI Volatility:** The standard deviation analysis identified states with high CPI volatility, indicating areas where inflation is unpredictable and may pose challenges for economic stability.
4. **Stationarity of CPI Series:** The ADF test confirmed that CPI is non-stationary, suggesting that further transformations or differencing may be needed for more accurate modeling.
5. **Clustering of States:** The K-means clustering approach categorized states into high, medium, and low inflation zones. This segmentation aids in understanding regional inflation dynamics and can inform state-specific economic policies.

These analyses provide a comprehensive understanding of inflation dynamics across India, highlighting areas of concern, forecasting future trends, and suggesting targeted actions for managing inflation stability.