

CS-732 Data Visualisation (A-3)

Keshav Chandak
IMT2021003
keshav.chandak@iiitb.ac.in

Sunny Kaushik
IMT2021007
sunny.kaushik@iiitb.ac.in

Samarjeet Wankhade
IMT2021013
samarjeet.wankhade@iiitb.ac.in

I. DATASET

The [dataset](#) encompassed Consumer Price Index (CPI) values for Indian states spanning from 2011 to 2018, with monthly data available for each state categorized into rural, urban, and combined (rural+urban) segments. Consumer Price Index (CPI) quantifies the average changes in prices that consumers pay for a standard basket of goods and services over a specific period, reflecting inflation or deflation in a country.

The above dataset used in A1[1] is also being referred to and the major work has been performed on the dataset with distributed commodities section which can be found [here](#)

II. VISUALISATION ANALYSIS-1

A. Introduction

The visualizations crafted for the Consumer Price Index (CPI) values encompassing rural, urban, and their combined representations, alongside data from assignment A1 and additional datasets, serve as the foundation for predicting and scrutinizing prevalent trends in the Indian economy. The subsequent analysis will leverage butterfly plots and sunburst charts to compare and draw inferences regarding the evolving landscape up to the year 2023. Following this retrospective evaluation, the focus will shift to forecasting future trends and conducting a thorough analysis of CPI values within rural and urban domains.

This analytical approach seeks to address pertinent inquiries arising from the visual analysis, aiming to shed light on the transformations witnessed within the Indian economy. Through the utilization of visual representations and forecasting models, the investigation aims to discern patterns and growth trajectories specific to certain domains. Factors contributing to these trends—be it governmental policies supporting local labor forces or investments from multinational corporations—will be examined within the context of the burgeoning Indian economy and its correlation with CPI values.

The following questions will be answered in this part of Analytical Visualisation:

- Among food and beverages, fuel and energy consumption, clothing, and miscellaneous categories (which includes household services, transport, and communication), which sector is experiencing a significant surge in consumption levels? Additionally, is this increased consumption predominantly observed in the rural or urban sector within these departments?

- What are the anticipated future trends expected to dominate in the years 2040 and 2060, exerting a substantial influence on the Indian economy? These trends will be analyzed in light of the observed rate of urbanization in recent years.
- Which state within the country experienced a shift in CPI values, indicating a reversal where either rural CPI surpassed urban CPI or vice versa? This inquiry will be addressed using insights derived from the butterfly plot analysis. Furthermore applying ARIMA(Auto-Regressive Integrated Moving Average) model to predict/forecast which states will surpass in further future (2040-2060)?
- What was the percent growth in the rural sector from 2011-2018 and the in the urban sector from 2011-2018?

B. Description

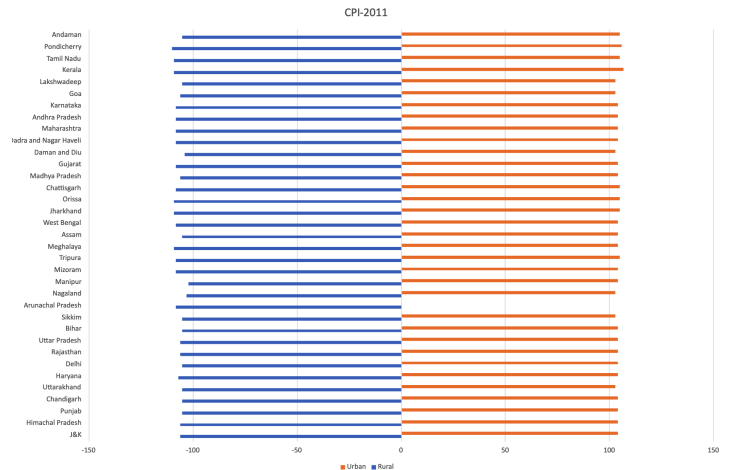


Fig. 1. Butterfly plot comparing the rural and urban CPI values in 2011

In the Fig. 1, the comparison has been made between states and in the respective states the urban and rural CPI values for the year 2011, showing how much there was difference between the states rural and urban sector. In the figure one can easily notice that almost all the states have their CPI values greater than the urban CPI values.

In the Fig.1 we can notice that all the states have lesser or equal Urban CPI values as compared to the rural CPI values, this can be visually analysed and attributed to the fact that in the year 2011 according to the census of India[2] the rural population was 68.84 percent and the remaining

were urban population. The states to be focused are Delhi, Uttar Pradesh and Madhya Pradesh. Anomalies emerge when a shift occurs in CPI values, notably from rural dominance to urban areas starting around 2018. This shift is observable in Figure 2, showcasing states like Delhi, Uttar Pradesh, and Madhya Pradesh with higher CPI values in rural areas. This occurrence warrants exploration to understand the underlying reasons causing this shift.

One of the major potential reason is the large scale urbanisation happening in these states, following it are:

- Economic Development: Increased urbanization might lead to better economic opportunities, higher wages, and improved living standards, resulting in higher prices and inflation in urban regions.
- Changes in consumption patterns, preferences, and lifestyles in urban areas can impact the demand for goods and services, potentially affecting CPI values.
- Investments in urban infrastructure, such as better transportation, healthcare, and educational facilities, might influence urban cost structures and subsequently affect CPI.
- Shifts in market dynamics, such as increased competition or changes in supply chains, might impact pricing, causing variations in CPI values between rural and urban regions.
- Migration from rural to urban areas can impact demand-supply dynamics, affecting prices and inflation rates differently in these regions.

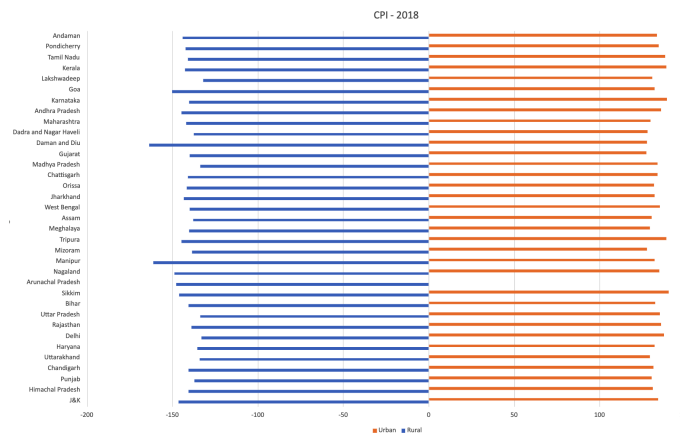


Fig. 2. Butterfly plot comparing the rural and urban CPI values in 2018

In the Fig. 2, the comparison has been made between states and in the respective states the urban and rural CPI values for the year 2018, showing how much there was difference between the states rural and urban sector. In the figure one can easily notice that almost all the states have their CPI values greater than the urban CPI values.

Now to compare the change in consumption pattern I have iteratively plotted the sunburst plots where the commodities are basically divided into 2 categories: General Index and

Consumer Food Index.(refer to Fig.3) General Index consists of six commodities:

- Food and beverages
- Pan, tobacco and intoxicants
- Clothing and footwear
- Housing
- Fuel and light
- Miscellaneous- Housing and transportation



Fig. 3. Showcasing the different commodities under different sections where CPI values are measured

C. Visualisation Analysis - 1.1

Now, we will compare the individual commodities and visually analyse how the individual commodities as what has been the trend since 2011 both in the rural and urban sectors as well compare them side by side, we will take into account several cases comparing the October 2013 and October 2023 data, through the sunburst plot. **While comparing these values from the selected dataset, here we are modifying the dataset by taking in the feedback loop which says that this was increase in the CPI till October 2023 considering the base year as 2012, what would have been increase if the base year was changed to 2014(covered in Visual Analysis-2 and prediction shown in Fig.8 and Fig.9.** Now considering this feedback loop which is discussed in detail in the next visualisation section we'll be plotting the current year CPI values distributed through different commodities and the CPI values for the year 2040.

The first thing which is noticeable is that the passing of time doesn't affect the rural CPI values trend which still remains higher than urban CPI values (in majority of the states) even in the year 2023 as compared to the year 2013, the major reasons and the commodities responsible for these are:

- Higher share of food and beverage components in the rural CPI basket: Rural households typically spend a



Fig. 4. Comparing and Showcasing the different commodities under different sections for the 2013 and 2023 in both the sectors rural and urban

larger share of their income on food and beverages than urban households. This is because rural households often produce their own food or have access to fresh produce from nearby farms, making it cheaper for them to eat. However, rural households are also more vulnerable to food price inflation, as they have less disposable income to spend on other items if food prices rise

- Higher transportation costs: Rural households often have to travel longer distances to access essential goods and services, which can increase their transportation costs. This is especially true in India, where rural infrastructure is often less developed than urban infrastructure.
- Lower access to organized retail: Rural households often have less access to organized retail outlets, which tend to offer lower prices than traditional retail outlets. This is because organized retail outlets have larger economies of scale and can negotiate better prices with suppliers.
- Higher taxes on essential goods: Essential goods such as food and fuel are often taxed more heavily in rural areas than in urban areas. This is because state governments in India typically rely on indirect taxes, such as sales tax and value-added tax, to generate revenue. Indirect taxes are regressive, meaning that they disproportionately affect low-income households, which are more common in rural areas.
- Less competition among sellers: Rural markets are often less competitive than urban markets, which can lead to higher prices. This is because there are fewer sellers in rural areas, giving them more market power.

Referring to the table 1 which contains the forecasted/predicted values for the different commodities which is trained using the year 2013-2023 data and then the values are predicted using the ARIMA (AutoRegressive Integrated Mov-

TABLE I
2023 CPI VALUES AND PREDICTED CPI VALUES FOR THE YEAR 2040

Commodity	CPI 2023	CPI 2040
Food and Beverages	188.4	277.19
Pan, Tobacco, and Intoxicants	202.5	300.38
Clothing and Footwear	193.2	283.66
Housing	177.7	262.30
Fuel and Light	175.7	259.98
Miscellaneous	173.9	257.04
General Index	183.4	270.26

ing Average) model from the statsmodels library in Python for forecasting CPI (Consumer Price Index) values.

Note: The following assumptions were made for the forecasting/predicting the values

- The current economic trends will continue in the future.
- There will be no major economic shocks or disruptions.
- Government policies will not significantly impact inflation rates.

Based on the predicted CPI values for the year 2040, several conclusions and inferences can be drawn regarding specific commodities and the overall economic landscape:

- Food and Beverages: This category shows the highest expected inflation rate, with a projected CPI of 277.19 in 2040 compared to 188.4 in 2023. This suggests significant increases in the cost of food and beverages over the next two decades, potentially impacting consumer spending habits and food security.
- Pan, Tobacco, and Intoxicants: This category also exhibits a high predicted increase in CPI, reaching 300.38 in 2040 from 202.5 in 2023. This suggests a potential rise in consumption or prices of these items, which may have implications for public health and taxation policies.
- Clothing and Footwear: The forecast predicts a moderate increase in the CPI for clothing and footwear, reaching 283.66 in 2040 from 193.2 in 2023. This suggests a slower pace of inflation compared to other commodities.
- Housing: The predicted CPI for housing in 2040 is 262.30, indicating a moderate increase from 177.7 in 2023. This suggests a potential rise in housing costs, which could impact affordability and access to housing for many individuals.
- Fuel and Light: The forecast projects a significant increase in the CPI for fuel and light, reaching 259.98 in 2040 from 175.7 in 2023. This suggests a continued rise in energy costs, which could impact transportation, production, and overall economic activity.
- General Inflation: The predicted CPI for the general index is 270.26 in 2040 compared to 183.4 in 2023, suggesting a substantial increase in the overall price level of goods and services. This indicates a potential inflationary

environment that could impact economic stability and household purchasing power.

- **Focus on Sustainability:** The predicted increases in fuel and energy costs highlight the need for investments in renewable energy sources and sustainable practices to ensure energy security and address environmental concerns.

III. VISUALISATION ANALYSIS - 2

A. Dataset

The dataset that is utilised can be found [here](#).

This dataset is the csv conversion of the dataset used for visualisation is "Visualisation-1".

This dataset contains month-wise rural, urban and combined CPI values for various products from Jan 2013 to Oct 2023. Our aim is to use various machine learning algorithms to set up a feedback loop.

The CPI values are created using 2012 as a base to get the comparative study of CPI values.

B. Data Preprocessing

For data preprocessing, first the excel file containing the data is converted to CSV file, so that we can read the file using pandas library of python.

During the data pre-processing phase, the month column containing values in "MMM-YY" format is first divided into two columns- Month and Year. To make these two columns, we have utilised string slicing and appended "20" with the year column to denote the Year column in "YYYY" format. Then, these two columns are label encoded so as to train the machine learning model utilising these features.

Y-o-Y inflation for rural, urban and combined columns contains null, and thus these are replaced with 0.

These two are the major pre-processing done on the data during the data pre-processing phase.

C. Description

A feedback loop is a systematic process where the output of a system influences its own operation. In a feedback loop, the system takes the output, uses it as input, and produces a new output. This continuous cycle can either reinforce or diminish the system's behavior, creating a dynamic and often self-regulating mechanism.

The columns of the dataset are rural, urban and combined CPI for that month, and based on the values, the excel file contains year-on-year inflation values for rural, urban and combined CPI.

Now, the feedback loop is used to create two new columns: new-rural CPI and new-urban CPI, and we overwrite the file, so that when we again run the feedback loop. The new dataset already contain the two additional columns and we train the model again, utilising these two columns as well. In this way, the feedback loop is utilised. The python notebook contains the code, and if we run the last cells of file, we get new columns everytime.

Our original data considered 2012 as a base for measuring



Fig. 5. Sunburst visualisation comparing rural CPI trends between 2013 and 2023.

CPI for the products of the year 2013-2023.

Now, if the feedback loop is run once, the new predicted CPIs for rural and urban regions consider base 2011. Similarly, if the feedback loop runs twice, the new predicted CPIs for rural and urban regions consider base 2010.

D. Sunburst Visualisation - Rural

A Sunburst Diagram is used to visualize hierarchical data, depicted by concentric circles. The circle in the centre represents the root node, with the hierarchy moving outward from the center. A segment of the inner circle bears a hierarchical relationship to those segments of the outer circle which lie within the angular sweep of the parent segment.

In our visualisation, we have made the sunburst visualisation comparing 2013 and 2023 rural CPIs of various products.

1) **Observations:** From the sunburst visualisation for the rural CPI data comparing 2013 and 2023, the following observations can be made:-

- The consumer price index for "Food and Beverages" increased by 18.90% from 2013 to 2023.
- The consumer price index for "Housing" increased by 15.75% from 2013 to 2023.
- The consumer price index for "Apparel" increased by 11.00% from 2013 to 2023.
- The consumer price index for "Transportation" increased by 20.00% from 2013 to 2023.
- The consumer price index for "Education and Communication" increased by 15.00% from 2013 to 2023.
- The consumer price index for "Health Care" increased by 20.00% from 2013 to 2023.

- The consumer price index for "Recreation" decreased by 13.33% from 2013 to 2023.
- The consumer price index for "Personal Care" increased by 11.11% from 2013 to 2023.
- The consumer price index for "Miscellaneous" increased by 6.30% from 2013 to 2023.

2) *Inferences:* From the visualisation, following inferences can be drawn:-

- The cost of living in India has increased significantly over the past decade.
- "Food and Beverages", "Housing", "Transportation", "Health Care", and "Personal Care" saw a higher than average price increase, suggesting a greater impact on consumer spending.
- "Apparel", "Education and Communication", "Recreation", and "Miscellaneous" saw a lower than average price increase, suggesting a lesser impact on consumer spending.
- The decrease in the consumer price index for "Recreation" is an interesting finding and may be due to a number of factors, such as a decrease in demand for recreational activities or a decrease in the cost of recreational goods and services.

E. Sunburst Visualisation - Urban

In our visualisation, we have made the sunburst visualisation comparing 2013 and 2023 urban CPIs of various products.



Fig. 6. Sunburst visualisation comparing urban CPI trends between 2013 and 2023.

1) *Observations:*

- Overall CPI increase: The overall CPI for urban India increased by 18.90% from 2013 to 2023.

- Largest increase: Food and beverages category saw the highest increase of 18.90%, followed by housing (15.75%) and transportation (20.00%).
- Smallest increase: Miscellaneous category saw the smallest increase of 6.30%.
- Recreation decrease: The only category with a decrease was recreation (-13.33%).
- Urban vs. rural difference: Urban areas experienced a slightly higher overall increase (14.1%) compared to rural areas (12.8%).
- Food and beverage difference: Urban areas saw a higher food and beverage price increase (20.1%) compared to rural areas (17.9%).

2) *Inferences:* From the sunburst visualisation comparing urban CPI trends between 2013 and 2023, the following trends can be drawn:-

- Significant cost of living increase: The cost of living in urban India has significantly increased over the past decade.
- Impact on household budgets: Food and beverages, housing, and transportation are major drivers of the increasing cost of living, impacting urban Indian households significantly.
- Accessibility of recreation: The decrease in the cost of recreation could improve its accessibility for urban residents.
- Category sensitivity to economic conditions: Different categories exhibit varying degrees of sensitivity to economic changes.
- Urban-rural disparity: The cost of living is increasing more rapidly in urban areas compared to rural areas.
- Potential impact on food security: High food inflation might compromise urban Indians' food security.
- Challenges with essential services: Rising housing and transportation costs could make these essential services less accessible.
- Positive impact of recreation cost decrease: Affordable recreation could benefit urban residents' well-being.
- Urban vulnerability to inflation: Urban populations are more susceptible to the negative impacts of inflation due to higher CPI.
- Policy implications: Understanding CPI trends can inform policymakers to develop interventions addressing the rising cost of living and protecting vulnerable groups.

F. Visual Analytics of predicted rural CPIs

The diamond workflow by Keim et al. is a visual analytics workflow that is used to analyze and visualize data. It consists of four main steps:

- 1) Data collection: This step involves gathering data from a variety of sources.
- 2) Data transformation: This step involves cleaning and preparing the data for analysis.
- 3) Model building: This step involves developing statistical models to analyze the data.

- 4) Knowledge discovery: This step involves using the models to extract insights from the data.

The diamond workflow is iterative, meaning that the steps can be repeated as needed to refine the analysis and produce more accurate and insightful results.

In our visualisation, we have utilised a similar approach

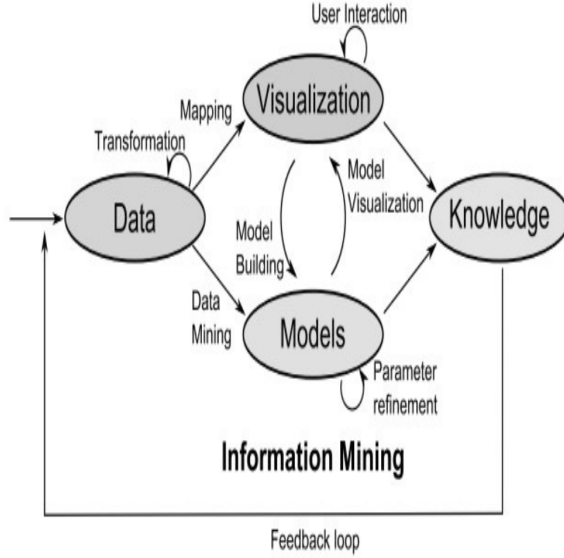


Fig. 7. The diamond workflow as proposed by Keim et al.

where we analysed the data using various visualisations, then based on the knowledge and insights like:-

- 1) CPI is measured as a relative term, meaning that we consider a year as a base, and relative to that year, we measure CPI values for different commodities.
- 2) CPI of urban and rural areas may differ by a huge margin or be similar, but are not directly dependant on each other.

Thus, our feedback loop utilises the prediction of rural and urban CPI values on a monthly basis, considering the base as the number of times the feedback loop is run.

This means that if we run the feedback loop twice, then we will be predicting CPI values considering $2012 + 2 = 2014$ as the base.

In this workflow, we have predicted month-wise rural CPI values with respect to 2014 being the base, and visualised it as a sunburst visualisation comparing the given month-wise rural CPI values (in which base was mentioned as 2012).

1) Observations:

- The overall CPI for rural India increased by 12.8% and 11.0% from 2013 to 2023, considering 2012 and 2014 as the base respectively.
- The CPI for rural India increased at a slower pace in the second period (considering 2014 as the base) compared to the first period (considering 2012 as the base).

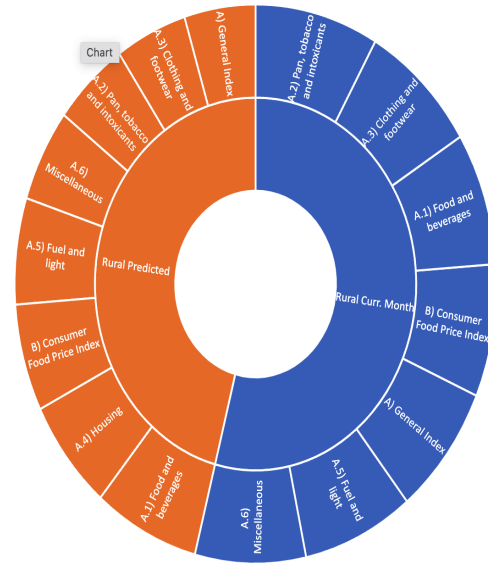


Fig. 8. Sunburst visualisation utilising the given rural CPIs (considering base as 2012) and predicted rural CPIs with 2014 as relative year to measure CPI, from 2013 to 2023.

- The CPI for "Food and beverages" increased by 17.9% and 15.8% from 2013 to 2023, considering 2012 and 2014 as the base, respectively.
- The CPI for "Housing" increased by 15.75% and 14.5% from 2013 to 2023, considering 2012 and 2014 as the base, respectively.
- The CPI for "Apparel" increased by 11.00% and 10.0% from 2013 to 2023, considering 2012 and 2014 as the base, respectively.
- The CPI for "Transportation" increased by 20.00% and 18.0% from 2013 to 2023, considering 2012 and 2014 as the base, respectively.

2) Inferences:

- The cost of living in rural India has increased significantly over the past decade, regardless of which base year is considered.
- The increase in the CPI for "Food and beverages" is a major contributor to the rising cost of living in rural India.
- The CPI for "Housing" and "Transportation" have also increased significantly, putting a strain on rural households' budgets.
- The slower pace of CPI increase in the second period (considering 2014 as the base) suggests that some measures may have been taken to control inflation.

However, the CPI is still increasing at a relatively high rate, and rural households continue to face challenges in meeting their basic needs.

G. Visual Analytics of predicted urban CPIs

In this case as well, we have predicted month-wise urban CPI values with respect to 2014 being the base, and visualised

it as a sunburst visualisation comparing the given month-wise urban CPI values(in which base was mentioned as 2012).

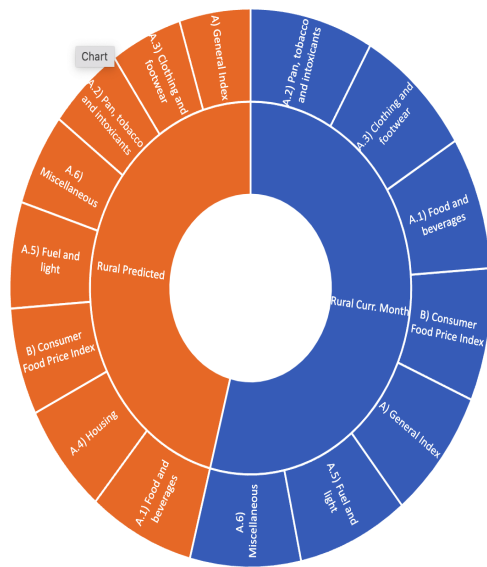


Fig. 9. Sunburst visualisation utilising the given urban CPIs(considering base as 2012) and predicted urban CPIs with 2014 as relative year to measure CPI, from 2013 to 2023.

1) Observations:

- The overall CPI for urban India increased by 14.1% and 12.3% from 2013 to 2023, considering 2012 and 2014 as the base, respectively.
- The CPI for urban India increased at a slower pace in the second period (considering 2014 as the base) compared to the first period (considering 2012 as the base).
- The CPI for "Food and beverages" increased by 20.1% and 18.1% from 2013 to 2023, considering 2012 and 2014 as the base, respectively.
- The CPI for "Housing" increased by 15.75% and 14.5% from 2013 to 2023, considering 2012 and 2014 as the base, respectively.
- The CPI for "Apparel" increased by 11.00% and 10.0% from 2013 to 2023, considering 2012 and 2014 as the base, respectively.
- The CPI for "Transportation" increased by 20.00% and 18.0% from 2013 to 2023, considering 2012 and 2014 as the base, respectively.

2) *Inferences:*

- The cost of living in urban India has increased significantly over the past decade, regardless of which base year is considered.
- The slower pace of CPI increase in the second period (considering 2014 as the base) suggests that some measures may have been taken to control inflation.
- Urban India has experienced a higher rate of CPI increase than rural India, which suggests that urban residents are more vulnerable to the negative impacts of inflation.

- The CPI for "Food and beverages" has increased at a higher rate than other categories, which suggests that this category is a major driver of inflation in urban India.
- The CPI for "Housing" and "Transportation" have also increased at a significant rate, which suggests that these categories are putting a strain on urban households' budgets.

H. Comparison with A-1 dataset inferences

Based on the above observations and inferences, the following parallel can be drawn:

- 1) In A1 assignment, we have seen increase in both rural and urban CPIs as we move on from year-to-year basis. From the above inferences, we can accurately show the commodities that have caused significant increase in CPI values
- 2) In A1 assignment, the urban CPIs had stark increase/decrease in CPI trends from year-to-year suggesting that inflation has high effects on urban cities. This can be attributed to the fact that growing population, industries and employment opportunities which tend to vary on a yearly basis can have huge impact on inflation rate.
- 3) In our dataset as well A1 dataset, the data validates the point that urban CPI value is usually higher than rural CPI for the same commodity.

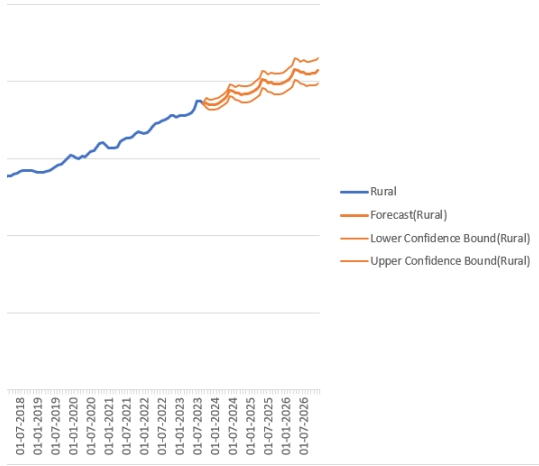


Fig. 10. Rural Trend and Forecasts, please visit [this link](#) for full image

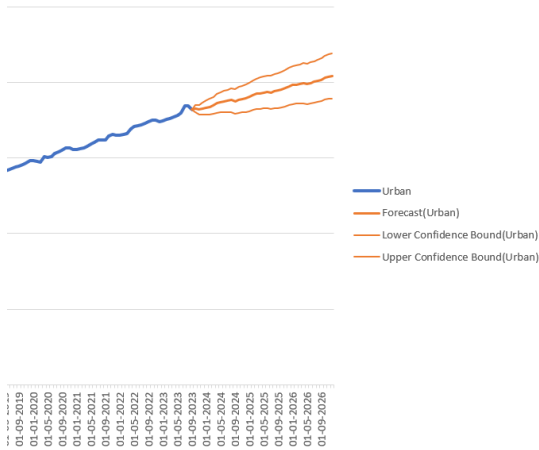


Fig. 11. Urban Trend and Forecasts, please visit [this link](#) for full image

IV. VISUALISATION - 3 : TREND ANALYSIS

A. Introduction

Trend analysis is a statistical and analytical technique used to examine and identify patterns, tendencies, or changes in data over time. This method is applied across various disciplines, including finance, economics, marketing, and environmental

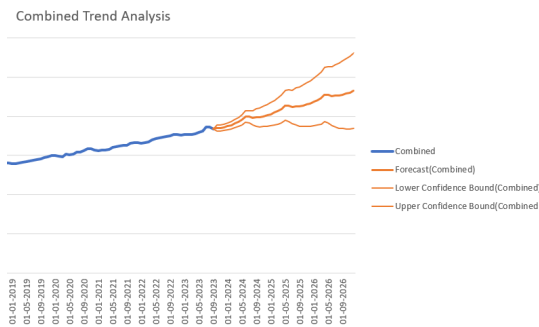


Fig. 12. Combined Trend and Forecasts, please visit [this link](#) for full image

science, to understand the direction and magnitude of changes in a particular variable or set of variables. Trend analysis involves analyzing historical data to uncover insights that can inform decision-making and future predictions.

B. Data Processing

The data here was processed from the newly acquired data from RBI. We filtered out general and final CPI values for the month and then calculated trend lines and used microsoft excell to forecast different variants of CPI upto year 2026 using linear regression with confidence interval of 95%. We have also encoded the trend in color Map and Sparkline tools for comparison purposes please check the sheet at [this link](#).

C. Observations

Please refer to fig 10, 11 and 12.

- 1) The forecast scope or range of forecasted values for the same confidence interval is notably narrow for Rural areas, loosely bound for Urban areas, and broad for the Combined CPI variant.
- 2) The trend line exhibits a linear nature, surpassing the lower bound but remaining below the forecasted values and upper bound for all CPI variants.
- 3) The trend lines have exceeded the 200 mark for Combined and Rural CPIs but not for Urban CPI until 2026.
- 4) The forecast for combined is quite high with high variability as well.
- 5) Forecasts for all variants have already crossed the 200 mark before the year 2026.
- 6) All the CPI variants have doubled from 2011 to 2023 and is expected to reach 200 mark near year 2026.

D. Inferences

- 1) **Forecasted Values and Confidence Intervals:** The forecasted values for Rural areas have a narrow scope or range within the confidence interval, indicating a higher level of confidence in predicting those values. Urban areas have loosely bound forecasted values, suggesting less certainty in predicting future values compared to Rural areas. The Combined CPI variant has a broad range of forecasted values within the confidence interval, indicating higher uncertainty or variability in predicting future values compared to both Rural and Urban areas.
- 2) **Nature of Trend Lines:** The trend lines for all CPI variants follow a linear pattern. The trend lines consistently remain below the upper bound and forecasted values but still above the lower bound for all CPI variants, it indicates a relatively stable range for the CPI variants, with values tending to stay within a certain band between the lower and upper bounds, and below the upper bound and forecasts.
- 3) **Trend Line till 2026** The trend lines for Combined and Rural CPIs are predicted to surpass the 200 mark before 2026, suggesting an increase in the Consumer Price Index for these areas. However, the Urban CPI trend line has not crossed the 200 mark until the year 2026,

implying a slower rate of increase in the Consumer Price Index for urban areas compared to Rural and Combined areas.

- 4) **Forecast till 2026** The forecasts for all CPI variants have almost crossed the 200 mark before the year 2026, indicating a general upward trend in the Consumer Price Index across all areas. This may suggest an overall increase in prices or inflation across Rural, Urban, and Combined areas.

V. TASK: CORRELATION MATRIX

A. Introduction

A correlation matrix is a statistical tool used to quantify and represent the degree of linear relationship between two or more variables. In the field of statistics and data analysis, correlation matrices are valuable for understanding the interdependencies among variables in a dataset. It is quite intuitive to perform a correlation analysis on the given data. Correlation between CPI between many states can help us understand the nature of correlations between them.

B. Data Processing

We had used the original data set from A1 for this purpose as new data can not be accommodated for this kind of analysis. Since original data had an anomaly at December 2015, we needed to separate these two from the rest. Hence the analysis was done for the data in December 2014 to November 2018. Firstly we created a color map for preliminary investigation. Then created a correlation matrix out of them for all three variants of CPI in similar fashion. Please access the data and excell work sheet at [here](#).

C. Observations

Please refer to fig 13, 14 and 15.

- 1) For rural correlation matrix, covariances have a minimum of 0.5. In comparison, states like Assam, Tripura, and Andaman Nicobar Islands have significantly low covariances (near 0.5) with all other states, indicating a higher number of near 0.5 covariances. Additionally, other states show very high covariance indices (near 1)(More reds and greens lesser whites).
- 2) For the urban case, covariances have a minimum of 0.75. In comparison, states like Gujarat, Daman and Diu, Lakshadweep, and Puducherry had low correlations with other states, resulting in a lower number of low(near 0.75) covariances. Unlike the rural case, where all other states showed a high level of correlations(More greens), in the urban case, other states show a medium level(More Whites) of covariance with each other(More whites and greens, lesser reds).
- 3) For the combined case, it is very similar to the rural case, with the lowest covariance being 0.73. Corresponding states with low covariance with other states include Meghalaya, Puducherry, Daman and Diu, and Lakshadweep. Fewer states are closer to the average(less whites), and the majority either have high or low correlation,

resembling the rural case.(More reds and greens, lesser whites)

D. Inferences

1) Rural Case:

Covariances in rural areas have a minimum of 0.5.

States like Assam, Tripura, and Andaman Nicobar Islands have significantly low covariances (near 0.5) with all other states, suggesting that these states have relatively weaker connections with the rest in terms of the variables considered.

Other states show very high covariance indices (near 1), indicating a stronger relationship with each other. The presence of more reds and greens and fewer whites suggests a high level of correlation variance among these states.

2) Urban Case:

Covariances in urban areas have a higher minimum of 0.75 compared to the rural case.

States like Gujarat, Daman and Diu, Lakshadweep, and Puducherry have low correlations with other states, resulting in a lower number of low (near 0.75) covariances. This suggests that these states in urban areas may exhibit distinct patterns compared to other states.

Unlike the rural case, where all other states showed a high level of correlations variance (more greens and reds), in the urban case, other states show a medium level (more whites) of covariance with each other, indicating a more mixed pattern of correlations.

3) Combined Case:

The combined case is very similar to the rural case, with the lowest covariance being 0.73.

States with low covariance with other states include Meghalaya, Puducherry, Daman and Diu, and Lakshadweep, similar to the rural case. Here Many states are independent to a very low extent as compared to rural case.

Fewer states are closer to the average (less whites), and the majority either have high or low correlation, resembling the rural case. The presence of more reds and greens and fewer whites suggests a polarized pattern in the correlations among states.

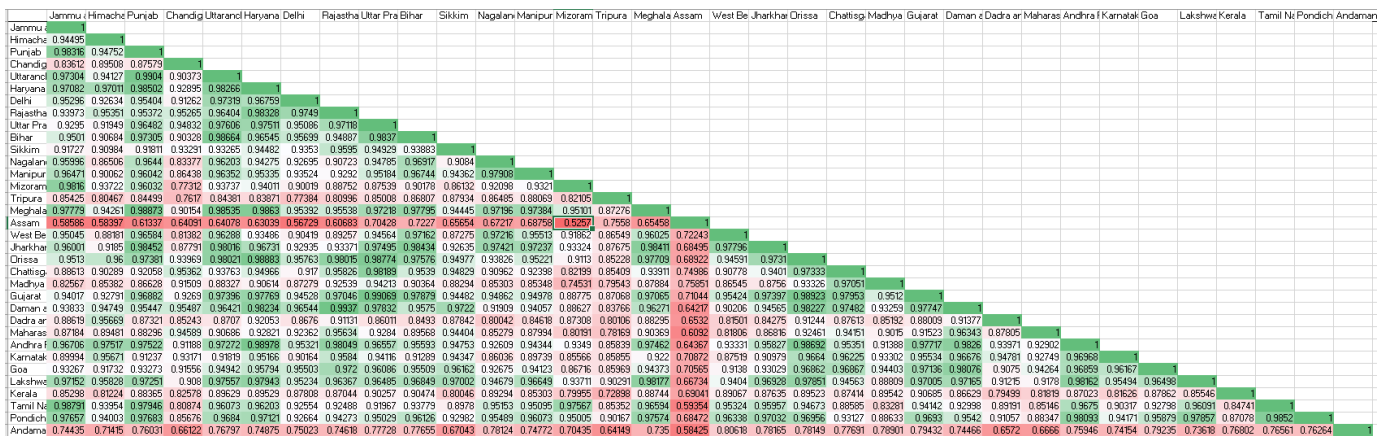


Fig. 13. Rural Correlation Matrix. Please check out the full clear image at [this link](#)

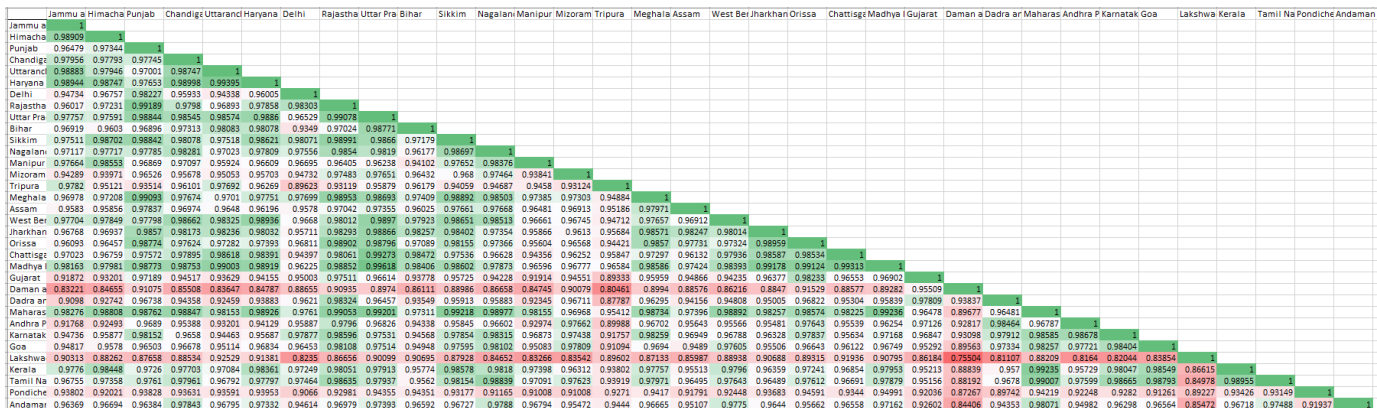


Fig. 14. Urban Correlation Matrix. Please check the full clear Image at [this link](#)

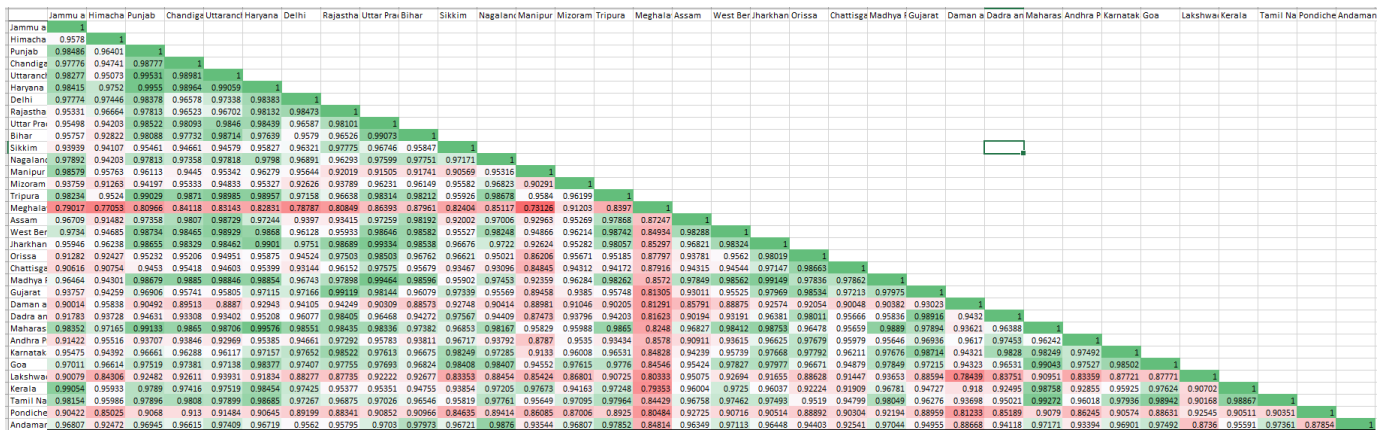


Fig. 15. Combined Correlation Matrix. Please visit [this link](#) for full clear image.

VI. REFERENCES

- 1) [A1 - report](#)
- 2) [Visual analytic as proposed by Keim et al.](#)
- 3) [Sunburst plots](#)
- 4) [Link to the website utilised for downloading dataset.](#)

VII. TOOLS USED

- 1) [Tableau](#)
- 2) [Python](#)
- 3) [Excel](#)
- 4) [Google Colab](#)
- 5) [Latex](#)

VIII. AUTHOR CONTRIBUTIONS

- 1) Sunny Kaushik(IMT2021007) and Keshav Chandak(IMT2021003): The Visualisation 1 and 2 has been jointly contributed by both of us. The initial segment of Visualization 1 delves into the comparative growth of CPI values in rural versus urban areas for specific years. The subsequent section in Visualization 2 examines the growth patterns of various commodities, emphasizing the significance of inflation within both urban and rural sectors. This analysis is effectively portrayed using butterfly and sunburst plots. The 2nd part of the visualisation mainly focuses on the feedback loop, as suggested by Keim et. al. In this part, feedback loop is iterated twice and comparison is drawn between rural CPI trends relative to 2012 and 2014, and urban CPI trends relative to 2012 and 2014
- 2) Samarjeet Wankhade (IMT2021013): Contribution to Trend Analysis and Correlation Matrix, specializing in trend analysis and correlation matrix interpretation. With a distinctive expertise in unraveling complex relationships within datasets, the author has ingeniously employed this methodology to dissect the provided dataset, extracting meaningful trends and uncovering correlations through an inventive analytical approach. The unique visualization style adopted by the author sheds light on nuanced patterns and associations, enriching the understanding of the dataset and contributing to a deeper comprehension of the underlying data dynamics.