IDENTIFYING LIVING INCOME AND LIVING WAGE ZONES OF INDIA

Anker Research Institute
Working Paper Series

Number 7

Authors

Sandip Sarkar
Balwant Mehta

Anker Research Institute

IDENTIFYING LIVING INCOME AND LIVING WAGE ZONES OF INDIA

March 2023

Anker Research Institute

- Sandip Sarkar
- Balwant Mehta

ABSTRACT

This paper tries to answer is whether it is possible to map India into Living Income/Living Wage zones, given its size and diversity. We approached this task by using cluster analysis to separate geographic areas into clusters with similar socio-economic and demographic characteristics associated with the cost of a decent standard of living for a family.

This involved (i) first identifying the most relevant socio-economic and demographic characteristics and the most relevant datasets, and (ii) carrying out cluster analysis using those indicators. The cluster analysis was based on K-means methodology and was validated by unweighted composite index and hierarchical methodologies. We identified 24 broad rural living income/wage zones, and 25 urban living income/wage zones. These broad classifications will help conducting representative living income/wage studies in India in future.

KEYWORDS: Living income, living wages, Anker Methodology, Demographic factors, Cluster Analysis, India

JEL CODES: J10, J30, J50, J80.

- Sandip Sarkar – Institute for Human Development, New Delhi, E-mail: delhisandip@gmail.com.

ACKNOWLEDGEMENTS

This study immensely benefited from the excellent guidance of Martha Anker and Richard Anker, Directors, Anker Research Institute. They have also taken considerable efforts in editing the final version of the report. Helpful reviews on the draft report and inputs from the following are duly acknowledged: Dr. Anup Karan (Additional Professor at the Indian Institute of Public Health, Delhi); Prof. Arup Mitra (Professor at Institute of Economic Growth, Delhi); Kuriakose Mamkoottam (Consultant, Bengaluru, Karnataka); Dr. Nidhi Kaicher (Assistant Professor at Ambedkar University, New Delhi). The support provided by Dr. Prashant Kumar Arya (Associate Fellow at Institute for Human Development, Ranchi) in preparing Living Income Zones (LIZ) maps was invaluable. We also extend our gratitude to the secretarial staff at Anker Research Institute, Joseph Varun and others at Laudes Foundation for their valuable comments and other support.

This report was supported by Laudes Foundation. However, all results and opinions reflected in this report are completely that of the authors and do not represent that of the Foundation. For any queries on the report, please reach out to the leadership of the Anker Research Institute <Email:marthaandrichard@ankerinstitute.org>

Sandip Sarkar
Visiting Professor, Institute for Human Development

Balwant S. Mehta
Associate Professor, Institute for Human Development

FOREWORD

Living wages and living incomes differ by location in every country, because of differences in living costs and expectations about what people consider decent. This is most apparent for rural areas compared to urban areas and for smaller cities or towns compared to large cities. It is also quite apparent that living wages and living incomes differ across regions of countries, especially for large and diverse countries. This variation is acknowledged in the widely used Global Living Wage Coalition definition of living wage and the widely used Living Income Community of Practice definition of living income – as these definitions specify that living wage and living income are "location specific".

But the fact that living wage and living income are location-specific leads to an important question and issue: How many different living wages and living incomes make sense for a country? This paper by Professors Sandip Sarkar and Balwant Mehta is a first step to tackling this very important question for India. The answer to this question is essential for the living wage and living income movement in India since it is clear that estimating separate living wages and living incomes for every village or city or even every taluka or district in India would not make sense for either practical reasons or policy uses. It is worth noting that the Anker Research Institute is currently also exploring this question of how many different living wages are needed in a country in some other countries – including Brazil and Mexico which although quite large are not nearly as large or diverse as India.

The facile answer to the question posed in the previous paragraph is "as many as possible", since if every location is different, every location then seemingly should have its own living wage and living income. And perhaps for this reason, some serious organizations follow this approach. For example, the MIT living wage calculator for the United States has approximately 3,000 estimates, because there are approximately 3,000 counties in the United States, while Living Wage 4 US has 709 estimates because there are said to be 709 commuting zones in the United States. At the global level, WageIndicator has more than 2,000 estimates (and growing) for regions in 148 countries, and Fair Wage Network has more than 1,500 estimates for regions and cities in nearly 200 countries. What all these efforts have in common is a desire to maximize the number of living wage estimates - with the number of estimates provided determined by data availability. There has been little recognition of the practical difficulties associated with estimating and having so many living wages and living incomes for each country. This includes problems with data availability to estimate living costs and expectations (especially outside of high-income countries which are rich in published data series). Nor is there been any recognition that many estimates for different locations within any given country are quite similar and so in reality not truly different. Neither has there been much recognition of the policy implications of having very many estimates for a country for example for trade unions in collective bargaining for wages, or for governments in minimum wage setting, or for organization in setting floor prices for agricultural products.

Other organizations and governments take quite a different approach to maximizing the number of living wage and living income estimates for countries. The World Bank uses 4 poverty lines for the entire world. Almost all governments use one or a few minimum wages and poverty lines for their country. The Anker Research Institute uses two living wage and living income values per country (rural and urban) for its Reference Values – with a clear indication and recognition that these are approximate average values for rural and urban areas to be used for assessing risk for companies and countries. Although crude, these parsimonious approaches are practical. They implicitly or explicitly recognize

data problems, and policy and other limitations of having many estimates for a country or sector.

Our view is that neither of the two approaches discussed above is optimal from either a policy point of view or a practical point of view – neither the facile approach of estimating and using as many living wage and living income estimates per country as possible given data limitations nor the parsimonious approach of using very few estimates per country. Our view is that it is best to be in the middle - with as few living wages and living incomes as possible that capture all substantial differences in living costs and living standards within a country. The absurdity of the first approach of having as many estimates as possible approach is evident for India which has more than 640,000 villages and 4,000 towns and cities. On the other hand, the limiting nature of one or two or very few living wages and living incomes approach for India, which is very big and very diverse, is equality evident.

The above discussion brings us to the present report by Professors Sandip Sarkar and Balwant Mehta. We asked Sandip and Balwant to determine how many living wage and living income zones are needed to represent differences in living costs and living conditions across India. This is a daunting task, because India is so large and so diverse. India has over 1.4 billion people and almost 18% of the world's population. It has hundreds of languages with 22 major languages recognized in its constitution, and its geography and climate go from small islands to a large land mass and from plains and tropical areas to the Himalayas and cold weather. Even food habits differ greatly across India with large parts of the country vegetarian and other large parts of the country non-vegetarian.

Professors Sandip Sarkar and Balwant Mehta took this daunting task to map India into a parsimonious number of living wage and living income zones very seriously, and as a result, this report represents an impressive amount of work and effort. Their task was made more difficult by data limitations, because the latest available NSS household expenditure survey is from 2010/11. This meant that this most important data source was too old to be useful, since economic growth in India between 2010 and the 2020 pandemic has been high with a real annual growth GDP per capita growth rate of around 5 percent. This required Sandip and Balwant to turn to alternative data sets and proxy variables. They ended up investigating, measuring and using some 50 socio-economic-demographic-cultural variables to understand differences across India. And to identify living wage and living income zones, they used three different aggregating methodologies (K-means method, Hierarchal method, and unweighted index method). In the end, after starting with 88 rural and 88 urban NSS regions, they identified 24 urban zones and 25 rural zones based on a combination of sophisticated statistical analyses and common sense such as analysing rural and urban areas separately and taking into consideration whether areas are or are not geographically contiguous. Thus, this report and its conclusions present a sensible middle ground between having too few living wage and living income zones to reflect the diversity of India and having too many zones for practical purposes.

Our thanks go out to Sandip and Balwant for their excellent and thoughtful report which we hope will significantly move forward work on living wages and living incomes in India and shedding light on different methods to divide India into living wage and living income zones.

Richard Anker and Martha Anker

February 2023

IDENTIFYING LIVING INCOME AND LIVING WAGE ZONES OF INDIA

March 2023

Anker Research Institute

TABLE OF CONTENTS

1. INTRODUCTION

1.1 DEFINITION OF LIVING WAGE AND LIVING INCOME

A living wage is the wage needed by a full-time worker in a typical size family to be able to afford a basic but decent living standard for his/her family. It has a long and distinguished history (Anker and Anker 2017) and is now gaining in popularity because all workers should be paid a decent wage, so they do not have to live in poverty.

The Anker Methodology is commonly used to estimate a living wage and/or living income n specific locations and at a specific time and is widely seen as the gold standard for measuring living wage and living income. A living wage is defined by the Global Living Wage Coalition (GLWC) as:

> ***"The remuneration received for a standard workweek by a worker in a particular place sufficient to afford a decent standard of living for the worker and her or his family. Elements of a decent standard of living include food, water, housing, education, health care, transportation, clothing, and other essential needs including provision for unexpected events."***

A living income is defined by the Living Community of Practice. It is mostly applicable to farm families and families relying on self-employment. A living income is:

> ***"The net annual income required for a household in a particular place to afford a decent standard of living for all members of that household. Elements of a decent standard of living include food, water, housing, education, health care, transportation, clothing, and other essential needs including provision for unexpected events."***

1.2 HOW TO ESTIMATE A LIVING WAGE – THE ANKER METHODOLOGY

The Anker Methodology estimates a living wage and living income in a particular location at a particular time based on normative standards for nutrition food, decent housing, adequate health care, and education for children through secondary school.

Since the living wage and living income is location specific, a question often asked is whether an estimate of living wage or living income for one location can be used in a different location. It is clear that not every small location should have its own living wage or living income estimate – since this would be impractical. Therefore, it is necessary to understand the geographic zones that each estimate is for. Furthermore, for large diverse countries such as India, it is necessary to map the country into living wage and living income zones that are likely to have similar living wages and living incomes throughout the entire zone. In this way, an estimate for one part of a living wage/living income zone could be applied to other parts of the zone as well. The central question that this paper tries to answer is whether it is possible to map India into such zones, given its size and diversity.

1.3 THE CHALLENGE OF MAPPING INDIA INTO LIVING WAGE AND LIVING INCOME ZONES.

Mapping India into living wage/living income zones is a challenging objective. India is a federal democratic republic with twenty-eight states

and eight union territories. It has the second largest population in the world, and three Indian States - Uttar Pradesh, Maharashtra and Bihar - have a population of more than 100 million each. Therefore, it is not surprising that India is very diverse in terms of cultures, demographics, living conditions, incomes, living costs, poverty rates, dietary pattern, and dominant industries, etc. For this reason, it is not appropriate to apply a living wage or living income estimate in one part of the country to other parts of country, which may be distinct and different in several factors. The wide variety of cultures, labour market conditions, living conditions, and living costs across the country means that one living wage for of the whole country is meaningless. It is also clear that a living wage and living income for one part of the country is not necessarily relevant for another part of country.

With the above background, this study aims to divide India into different living wage and living income zones where each zone is expected to have relatively similar living costs throughout the zone. This will help in selecting places to carry out Anker living wage/living income studies in the future and estimate how much living wage and living incomes are in each of these zones.

1.4 APPROACH TO MAPPING LIVING WAGE AND LIVING INCOME ZONES

The way in which we approached this task was to use cluster analysis to separate geographic areas into clusters with similar socio-economic and demographic characteristics associated with the cost of a decent standard of living for a family. This involved (i) first identifying the most relevant socio-economic and demographic characteristics and the most relevant datasets, and (ii) carrying out cluster analysis using those indicators. The final cluster analysis was based on K-means methodology and was validated by unweighted composite index and hierarchical methodologies. We identified 24 broad rural living income zones, and 25 urban living income zones. These broad classifications we feel should help conducting representative living income/wage studies in India in future.

1.5 OUTLINE OF THE REST OF THE PAPER

The remainder of this paper is organised in various sections. Section 2 focuses on indicators and database, followed by discussion on methodology in section 3. Sections 4 and 5 analyse the results from cluster methodologies for rural and urban areas, and the identification of living income zones in rural and urban India is undertaken in sections 6 and 7. Last section 8 concludes the paper with summary of the findings and implications of the study. Annex I describes the Multiple Poverty Index (MPI). Annexes II and III indicate NSS regions in text, table, and map. Annexes IV and V provide definitions and list of all of the variables used in this study. Annex VI provides details of results of the unweighted composite index method, and Annex VII provides details of results of the Hierarchical cluster method analysis.

2. INDICATORS AND DATABASE

This section begins with an explanation of why multiple indicators are required for mapping India into living income zones. To select the most appropriate indicators for the analysis, several factors come into play including (i) the level of disaggregation required, (ii) the most relevant data source to use, (iii) the need for different variables for rural and urban areas, and (iv) ways to normalize (standardize) variables selected for statistical analysis.

2.1 HOW DATA WAS SELECTED TO IDENTIFY LIVING INCOME AND LIVING WAGE ZONES

In previous paragraphs, we indicated how large and diverse India is and why this poses a significant challenge to identifying living incomes and living wages for India in a practical way. For example, it would not be practical nor manageable to have a separate living wage or living income estimate for each and every village in India. Therefore, this paper aims to identify a manageable number of distinct living income zones of India with similar living incomes throughout the zone. To do this, requires aggregating data to a manageable number of zones on the basis of suitable indicators. The indicators for this focus on measures of economic welfare, because living wages and living incomes are associated in part with the level of economic welfare in countries and the same would be expected for India. However, the degree of association is not absolute, as many factors come into play when considering living wages and living incomes. Therefore, there is no straightforward immutable association between the level of economic welfare and living wages of specific areas. Some typical possible indicators are discussed below.

2.1.1 Difficulty of using monthly per capita consumption expenditure as an indicator for India

The most common indicator that is used to represent economic welfare in a specific region is monthly per capita consumption expenditure (MPCE). However unfortunately, the last available detailed estimation of MPCE for India pertains to the year of 2011-12, and so it is about one decade old. Considerable changes have taken place in India in the last one decade in terms of economic welfare. Hence, the MPCE for 2011-12, cannot be used as an indicator of the current level of economic welfare.

2.1.2 Multidimensional Poverty Index

The Government of India think-tank, NITI Aayog released Multidimensional Poverty Index (MPI) in 2021 based on 12 indicators grouped into three dimensions, namely health, education and standard of living to examine the impact of economic welfare policies. The findings of MPI can serve as a useful source for measuring the situation at baseline i.e., before the large-scale rollout of government important welfare schemes in the country (see Annex I for details).

2.1.3 Other relevant indicators

There are many other indicators that are associated with living wages and living incomes that we consider in this analysis. These factors include administrative delineation (e.g., state, district, etc.); economic conditions (such as income/consumption expenditure per capita, types of industries, and unemployment rates); social and cultural factors (such as. veg or non-veg diet, female status and female labour force participation rates, and literacy rates); and demographic factors (such as total fertility rate, average household size, and population density).

2.2. LEVEL OF DISAGGREGATION OF DATA USED IN ANALYSIS

As indicated earlier, India is endowed with considerable diversity in culture, labour market characteristics, sources and level of income & earnings, dietary patterns, demographic characteristics and composition of industries and employment.

Therefore, the living wage and living income estimated for one part of India cannot be used for other parts of India. The geographical boundaries of living wage/living income estimation would need to be based on various factors such as administrative boundaries (district, state or combination of districts), and economic, social & cultural and demographic factors.

The issue is at what level of disaggregation does the data used to identify living income zones need to be. We considered different levels of disaggregation such as the district, administrative regions, and NSS regions and concluded that the level of analysis should be NSS regions because they can provide the most accurate data for our purposes. The pros and cons of using each of various possible levels of disaggregation are discussed below:

2.2.1 District level

Although the most detailed disaggregated administrative region of India is the district, with 718 at present, districts vary substantially in terms of size of geographical area and population. The number of districts in a state or union territories is decided by state or union territories authorities. The decision of creating new districts varies substantially from one state to another.

2.2.2 States and union territories

The next higher level of aggregation is states, and union territories, which are 28 and 8 in number respectively. The relative differences in size of land area and population across state and union territories is even greater compared to districts.

2.2.3 NSS regions

In India, most socio-economic and demographic data are collected through sample surveys. There is a national census and the last census was undertaken in 2011. Decadal census covers a few indicators that are mostly demographic and a few indicators that cover economic well-being. Sample surveys provide reliable estimates at the state and union territory level, but reliable estimates at the district level for various indicators are not available. In this case, we have used an alternative level of data disaggregation i.e., at regional level. This lies between state and district levels. They are known as NSS regions.

At the regional level, various socio-economic estimations can be done with reasonable accuracy on the basis of sample surveys. Each NSS region is a combination of districts falling within states and union territories. Smaller states and union territories contain mostly a single NSS region, while larger states have diverse geographical characteristics and contain multiple NSS regions.

NSS regions were originally formed having similar geographical factors, rural population densities and crop-pattern. The number of NSS regions in 1977 was 73, which increased marginally during mid-1970s to mid-1990s, and was 78 in 1993. But in recent years, the number of regions has increased substantially due to creation of many new states, which were created in order to provide greater coverage to distant regions. The number of NSS regions was 88 in 2017-18.

Annex II provides the list of NSS regions falling within various states and union territories. More importantly, it provides the names of districts that fall within each of NSS regions.

Annex III provides a map of NSS regions that shows geographical boundary of each NSS region falling within each state.

2.3 NEED TO ANALYSE RURAL AND URBAN AREAS SEPARATELY

2.3.1 Major differences between factors that affect living wage and living income between urban and rural locations

India is a very diverse country, which requires separate analysis for rural and urban areas. There are several reasons for this. First, in rural areas, the primary sector (related to agriculture) provides employment to more than half of the workforce, whereas in urban areas, corresponding employment share of primary sector is one-twentieth. Second, the share of regular workers in the workforce is as high as 42 per cent in urban areas, compared to only 13 per cent share in rural areas. Third, housing cost contributes a greater proportion of living costs and so to the living wage and living income in urban areas compared to rural areas. For example, the Anker Living Wage Report of Rural Nilgiris of 2018 estimated that the rent of decent house for a family is Rs. 2,500 per month, which is far less than similar housing facility in urban areas. Therefore, we undertake separate disaggregated analyses in this paper for rural and urban areas using NSS regions.

2.3.2 Further breakdown of urban areas based on influence of organized manufacturing.

However, we have additional indicators for the urban analysis due to following reasons. The vast pockets of some urban areas are much less influenced than other pockets by concentration of organised manufacturing, and economic hub of major metropolitan cities. Some metropolitan areas are increasingly dominated by organised service activities, while simultaneously manufacturing activities are being pushed out to their peripheries. Therefore, urban areas have two distinct categories: i) urban areas relatively uninfluenced by organised manufacturing activities or metropolitan cities, and ii) urban areas in industrial zones/corridors and under the direct influence of major metropolis.

2.3.3 Aggregation of NSS zones into living wage and living income zones

On the basis of suitable statistical methodology, the regions/zones in rural areas, and in urban areas taking into the above two categories will be further aggregated from the 88 NSS regions into a smaller reasonable number of zones in the analysis in this paper.

2.4 DATA SOURCES

One of the largest sample surveys on employment and unemployment in India is conducted by National Sample Survey Organisation (NSSO), and the latest survey is available for the year 2018-19. In our analysis, this NSSO survey is supplemented by National Health and Family Welfare Survey (NFHS-4) for the year 2015-16, which provides data on food habits and health indicators. It is also supplemented by fertility rate data from the latest round of latest SRS (Sample Registration System).

2.5 SELECTION OF VARIABLES

The selection of the variables has been done from these databases in the following manner. First, we categorised variables into three broad factors: economic conditions, social and cultural factors, and demographic factors. We collected more than 50 variables separately for both rural and urban areas at the NSS regional level. The broad categorisation of collected variables are given in the Table 1. Annex IV provides the definitions and concepts of the variables generated for this study, while Annex V presents the list of variables created for each of 88 NSS regions, separately for rural, urban and all areas.

Table 1: Broad Composition of Factors and Variables for Rural and Urban Areas

Sl. No.	Broad Category of Variables	Specific Variables
Economic factors		
1	Consumption expenditure	Per capita consumption expenditure Household consumption expenditure
2	Sources of household earnings	Self-employed in agriculture and non-agriculture Regular salary earnings Casual labour in agriculture and non-agriculture Other sources of income
3	Income	Household and per capita income
4	Wages & earnings	Wages of casual workers Earnings of regular workers
5	Distribution of employment	Employment in primary, secondary and tertiary sector Share of manufacturing Organised sector employment Organised manufacturing sector employment
6	Location of work	Rural and urban location of work
Demographic factors		
7	Family size	Average size of households
8	Population density	Population, area and population density
Social and Cultural factors		
9	Literacy rate	Female literacy rate aged seven & above
10	Labour supply, workforce and unemployment	UPS - Female labour participation rates UPS and CWS work participation rates UPS unemployment rate
11	Vegetarianism	Proportion of vegetarians

Notes: UPS indicates usual principal status. CWS indicates current weekly status.

At least one variable is chosen from each category/factor for our analysis for rural areas and separately for urban areas. The choice of variables in each group is dependent on the level of inter-correlation between the chosen variables. We chose variables that are comparatively less correlated with each other; otherwise, it would lead to redundancy among the chosen variables. In the following sections, the choice of variables, and justifications are provided.

2.5.1 Choice of Variables for Rural Areas

Economic factors

1. From the first broad economic factor category consumption expenditure in rural areas, we have two choices either per capita consumption expenditure or household consumption expenditure. The per capita consumption expenditure

was chosen, because we intend to keep size of household as independent variable. There are two reasons for doing so. Firstly, household size directly impacts the calculation of living wage, and secondly, it is a proxy for total fertility rate (TFR) as TFR is not available separately for rural and urban areas at NSS regional level.

2. The main source of household earnings factor has three broad components: casual wage earnings, regular wage earnings, and self-employment earnings. These components are further divided between agriculture and non-agriculture. With the development of a region, the dependency of rural households on casual labour work declines. At the same time, casual labour households often earn from both agriculture and non-agriculture at different points of time in the same year. Therefore, it is prudent to consider casual households earnings from both agriculture and non-agriculture together. Another reason is that inter-correlation of share of casual labour households' earnings with other chosen variables is low. Therefore, the share of casual labour households' earnings variable is chosen.

3. For income we selected, per capita income but it is dropped from the analysis because per capita income and per capita consumption are highly correlated.

4. For wages and earnings, wage of casual workers is chosen as we have already chosen the share of casual workers from the household earnings category.

5. The share of employment from manufacturing sector is chosen from the distribution of employment category. The choice of manufacturing sector is guided by the consideration that the share of employment in manufacturing sector goes up with economic development.

6. For location of work, there was only one variable available.

Demographic factors

1. From demographic factor, the average size of household is the only variable available.

Social and cultural factors

1. From the first social & cultural factor, female literacy rate and female labour force (usual principal status) participation rates are chosen. First reason is that both female literacy rate and female labour force participation rate vary substantially across regions and secondly, they also capture women's role and status factor, which largely indicate socio-economic development status of that region.

2. Lastly, the prevalence of vegetarianism among population variable is chosen. As a large proportion of vegetarian population in some regions can impact total expenditure on food and to that extent can influence living wage and living income estimates.

To summarize, the following eight variables were selected for rural area analysis:

i. Per capita consumption expenditure,

ii. Size of household,

iii. Share of casual workers income in household income,

iv. Casual wage rate,

v. Female literacy rate,

vi. Female labour force participation rate,

vii. Share of manufacturing sector employment in total employment, and

viii. Share of vegetarian population.

Correlation between selected indicator variables for rural areas

Table 2 indicates the correlation between the eight selected variables for rural areas.

Table 2: Pearson Correlation of Chosen Variables in Rural Areas across 88 NSS Region

Variables	MPCA_RU	CL_R	HH_SIZER	Lit_rate_FR	UPS_LS_FR	WAG_CA_PR	MAN_SH_PR	all_veg_ru
MPCA_RU	1	-.209	-.245*	.442**	.047	.564**	.301**	-.099
CL_R	-.209	1	-.275**	-.230*	.040	-.141	-.121	-.026
HH_SIZER	-.245*	-.275**	1	-.184	-.457**	-.016	-.101	.486**
Lit_rate_FR	.442**	-.230*	-.184	1	-.089	.572**	.003	-.431**
UPS_LS_FR	.047	.040	-.457**	-.089	1	-.140	-.123	-.104
WAG_CA_PR	.564**	-.141	-.016	.572**	-.140	1	.074	-.302**
MAN_SH_PR	.301**	-.121	-.101	.003	-.123	.074	1	-.134
all_veg_ru	-.099	-.026	.486**	-.431**	-.104	-.302**	-.134	1

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).
Notes: MPCA_RU – Per capita consumption expenditure in rural areas
CL_R – The share of casual workers' income in household income
HH_SIZER – Size of household
Lit_rate_FR – Female rural literacy rate
UPS_LS_FR – Female labour force participation rate (UPS) in rural area.
WAG_CA_PR – Wage rate of casual workers in rural areas
MAN_SH_PR – The Share of manufacturing sector in total employment
All_veg_ru – The share of vegetarian population.

It can be observed from the correlation matrix that highest correlation is between female literacy rate and wage rate of casual workers. It shows that higher female literacy rate is positively correlated with casual wage rate of all workers. It is closely followed by per capita consumption expenditure, which reflects that higher casual wage rate indicating higher income level positively affects per capita consumption expenditure. In contrast, high negative correlation between household size and female labour force participation rate indicates that higher family size reflecting larger number of dependent children and elderly population may be discouraging female to participate in the labour market or maybe higher female labour force participation encourages women to family plan.

2.5.2 Choice of Variables for Urban Areas

A similar procedure was followed for the selection of variables in urban areas. Some of the variables in urban areas are same as selected for rural areas. Selected variables were:

i. Monthly per capita consumption expenditure

ii. Size of households

iii. In place of share of casual worker households, we have chosen regular worker households since the share of regular salaries as source of earning is substantially higher in urban areas.

iv. In a similar fashion, we have included earnings of regular workers for urban areas in place of earnings of casual workers in rural areas.

v. Female literacy rate (FLR) variable is dropped because in urban areas, FLR is fairly high in all regions and it is no longer a differentiating variable across NSS regions. We have kept female labour force participation rate (usual principal status) as it also represents women's role and status. Secondly, as virtually all males in working age participate in the labour market, the extent of second regular earning source as female participation in the labour market also affects the estimation of living wage per family.

vi. We have also chosen share of vegetarian population as in case of rural areas.

We considered the option of bifurcating urban areas into metropolitan and non-metropolitan areas. But after examination of our database, such distinction was not clearly seen. We have included two more variables that would differentiate these two areas. The first variable is location of work in urban areas. NSS regions with large metropolitan or industrial areas would have much larger share of workers working in the urban areas for the whole region (rural and urban combined) due to higher urbanisation rate and a larger share of commuting workers from neighbouring rural areas. The second additional variable for urban areas is the share of organised sector workers, as in large urban areas, there is a concentration of organised sector workers. However, the share of organised sector workers shows high correlation with regular wage rate, but still we included this in the estimation because of its importance.

In summary, the following eight variables have been chosen for urban areas.

i. The monthly per capita consumption expenditure

ii. The size of households

iii. The share of regular workers households

iv. The monthly earnings of regular wage workers

v. The female labour force participation rate (UPS)

vi. The share of vegetarian population

vii. The proportion of workers working in urban location of work

viii. The share of organised sector workers

Correlation between selected variables in urban areas

Table 3 indicates correlations between the eight variables for urban areas.

We have already mentioned the high positive correlation between salary/wages level of regular workers with the share of organised sector worker. Another variable, which shows high positive correlation, is share of workers in urban areas with per capita consumption level in urban areas. It indicates that NSS regions that contain metropolitan cities have comparatively higher share of urban workers and higher level of per capita consumption. Other variables that show positive correlation are share of manufacturing employment and level of regular salary. This indicates that regular workers salary in unorganised manufacturing sector is comparatively higher than those engaged in unorganised service activities. High negative correlation between household size and female labour force participation rate indicates that higher family size reflecting larger number of dependent children, and elderly population

which may discourage women from participating in the labour market in urban areas as they need to engage more in care activities.

Table 3: Pearson Correlation of Chosen Variables in Urban Areas in 88 NSS Regions

	MPCA_UR	REG_SALU	HH_SIZEU	UPS_LS_FU	WAG_RE_PU	ORG_SH_PU	MAN_SH_PU	WRK_PLC_PU	all_veg_ur
MPCA_UR	1	.450**	-.488**	.372**	.375**	.421**	.088	.526**	.023
REG_SALU	.450**	1	-.324**	.347**	.017	.725**	.483**	.466**	.101
HH_SIZEU	-.488**	-.324**	1	-.609**	-.038	-.326**	-.208	-.253*	.397**
UPS_LS_FU	.372**	.347**	-.609**	1	-.002	.252*	.051	.295**	-.329**
WAG_RE_PU	.375**	.017	-.038	-.002	1	.372**	-.525**	.009	-.231*
ORG_SH_PU	.421**	.725**	-.326**	.252*	.372**	1	.275**	.350**	-.017
MAN_SH_PU	.088	.483**	-.208	.051	-.525**	.275**	1	.412**	.299**
WRK_PLC_PU	.526**	.466**	-.253*	.295**	.009	.350**	.412**	1	.008
all_veg_ur	.023	.101	.397**	-.329**	-.231*	-.017	.299**	.008	1

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Notes: MPCA_UR – Per capita consumption expenditure in urban areas.
REG_SALU – The share of regular workers' income in household income
HH_SIZEU – Size of household
UPS_LS_FU – Female labour force participation rate (UPS, usual status) in urban area
WAG_RU_PU – Salary earnings of regular workers in urban areas
ORG_SH_PU – The share of organised sector workers in urban areas
MAN_SH_PU – The share of manufacturing sector in total employment
WRK_PLC_PU – The proportion of workers working in urban location of work
All_veg_ur – The share of vegetarian population

2.6 STANDARDISATION OF SELECTED VARIABLES

After choosing the variables, we used min-max procedure to normalize the variables as they differ in units, values and ranges. Min-max normalization is one of the most common ways to normalize data. For every feature, the minimum value of that feature gets transformed into a 0, the maximum value gets transformed into a 1, and every other value gets transformed into a decimal between 0 and 1. In the normalisation process, we have kept minimum value very close to zero as in statistical procedure if value of a variable is zero then that observation is dropped from the estimation. After normalization, the influence of differences in unit or range has been substantially reduced.

The next step after standardization of variables is to identify the methodology that will be used to reduce the number of observations (88 in both rural and urban areas). This is discussed in detail in the next section.

3. AGGREGATING DATA

As the database contains 88 NSS regions each for both rural and urban areas and these are substantially large in number, we need to find ways to aggregate NSS regions to smaller number of living wage zones separately for rural and urban areas. It would have been easier if per capita household income or per capita consumption expenditure data were available at NSS regional level, which is not available beyond the year 2011-12. The alternative is to cluster NSS regions to smaller number of Living Wage and Living Income Zones based on various available economic, social and cultural, and demographic factors. This exercise requires us to perform an exercise to explore alternative methodologies to arrive at a suitable methodology, which can be adopted in the identification of Living Wage and Living Income Zones.

Aggregation can be done using statistical analyses such as composite index, cluster analysis techniques, principal component analysis (PCA), etc. based on factors such as economic conditions, social and cultural factors, and demographic factors.

We decided to not use principal component analysis, where the impact of a single or a few variables become dominant, and the importance of other chosen variables are greatly diminished. In the analysis of rural areas for example, we have observed that female literacy rate captures the major part of the impact in the principal component analysis.

After considering several options, we decided to explore using the following three methods:

1. Hierarchical Cluster Analysis
2. K-Means Cluster Analysis
3. Unweighted/Composite Index

The first two are categories of cluster analysis. Hierarchical and K-means cluster analysis belong to a class of unsupervised learning method of machine learning and are used to classify cases or observations into homogeneous groups or distinct clusters having similar characteristics on the basis of defined set of variables. In other words, the clusters are nothing but the grouping of data points such that the distance between the data points within the clusters is minimal, and regions where the density of similar data points is high.

There are several ways to measure the distance between clusters in order to decide the rules for clustering, and they are often called linkage methods. Some of the common linkage methods are: (a) Single linkage (nearest neighbour): the distance between the two clusters is the *shortest* distance between two points in each cluster; (b) Complete linkage (furthest neighbour): the *longest* distance between two points in each cluster, (c) Average linkage: the distance between the two clusters is the average distance between each point in one cluster to every point in the other cluster, and (d) Centroid-linkage which finds the centroid of cluster 1 and centroid of cluster 2, and then calculates the distance between the two before merging. The choice of linkage method entirely depends on choice and there is no hard and fast method that will always give good results. Through trial and error, we chose average linkage and used Euclidean distance as measurement of distance.

Here the distance means Euclidean distance, which is defined as:

$$D_{ij}=\sqrt{\sum_{k=1}^{n}(x_{ki}-x_{kj})^2}$$

D_{ij} distance between cases i and j

X_{ki} value of variable X_k for case i

X_{kj} value of variable X_k for case j

Substantial differences exist between Hierarchical and K-means methods. These two methods are discussed one after another and then these are compared.

3.1 HIERARCHICAL CLUSTER ANALYSIS

Hierarchical cluster analysis groups data points into clusters. There are two types: Agglomerative Hierarchical Clustering, and Divisive Hierarchical Clustering.

3.1.1 Agglomerative Hierarchical Clustering or bottom-up approach

This is the most common type of hierarchical clustering used to group observations into clusters based on their similarity. In this approach, each data point acts as a cluster initially, and then it groups the clusters one by one. This method begins with as many clusters as there are observations and ends with a single cluster containing all observations. It's also known as Agglomerative Nesting and "bottom-up" approach as each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. The process of clustering of data points and how the hierarchical clustering works is represented by a tree showing hierarchical relationships between different sets of data through linkages called dendrogram. By looking at the Dendrogram, one can tell the number of clusters formed.

3.1.2 Hierarchical divisive clustering or top-down approach

This is the opposite of Agglomerative; it starts off with all the points in one cluster and divides them to create more clusters. As it begins with a single cluster, it ends with as many clusters as there are observations. In other words, one can say that Agglomerative (start from n cluster, to get to 1 cluster) and Divisive (start from 1 cluster to get to n cluster). We tried both of these methods and finally chose Agglomerative Nesting Approach to also get the possible number of clusters in our analysis.

3.2 K-MEANS CLUSTER ANALYSIS

K-means clustering is one of the most widely used non-hierarchical method employed for quick clustering. K-means clustering approach explores for a pre-planned number of clusters in an unlabelled multidimensional dataset; it concludes this via an easy interpretation of how an optimized cluster can be expressed. It partitions the data points into k clusters based upon the distance metric used for the clustering. The value of 'k' is defined by the user. Primarily the concept would be in two steps: firstly, the cluster centre is the arithmetic mean (AM) of all the data points associated with the cluster, and secondly, each point is adjoint to its cluster centre in comparison to other cluster centres. These two interpretations are the foundation of the K-means clustering model. It utilizes an iterative procedure to yield its final clustering based on the number of predefined clusters, as per need according to the dataset and represented by the variable K. For instance, if K is set to 3 (k3), then the dataset would be categorized in 3 clusters; if k is equal to 4, then the number of clusters will be 4, and so on.

This is the centroid-based algorithm such that each cluster is connected to a centroid while following the objective to minimize the sum of distances between the data points and their corresponding clusters. The first step is to create c new observations among our unlabelled data and locate them randomly, called centroids. The number of centroids represents the number of output classes. The first step of the iterative process for each centroid is to find the nearest point (in terms of Euclidean distance) and assign them to its category. The distance is calculated between the data points and the centroids of the clusters. The data point which is closest to the centroid of the cluster gets assigned to that cluster. Next, for each category, the average of all the points attributed to that class is computed. The output is the new centroid of the class. After

an iteration, it computes the centroids of those clusters again and the process continues until a pre-defined number of iterations are completed or when the centroids of the clusters do not change after an iteration.

Presented below is a broad comparison between Hierarchical and K-Means clusters.

Hierarchical clustering	K-means clustering
For a large number of variables and data-set, Hierarchical clustering is slow, and produces less reliable results.	For a large number of variables and dataset, K-means operates quicker or faster and produce more reliable results.
If number of clusters cannot be determined based on prior beliefs, Hierarchical clustering should be used to determine the number of clusters.	Need to specify the number of clusters.
It is easier to determine the number of clusters by Hierarchical clustering's dendrogram. Hierarchical clustering is a very useful way of segmentation.	If there is a specific number of clusters in the dataset, but the group they belong to is unknown, choose K-means.
Hierarchical clustering is less sensitive to noise in a dataset.	K-means are highly sensitive to noise in the dataset and performs better than Hierarchical clustering when there is considerable noise in the dataset.

Therefore, it is suggested widely in the research literature that it is best to first perform a Hierarchical method to define the number of clusters, and then use the K-means method to actually form the clusters. In the analysis in this report, we have adopted a similar approach to determine the number of clusters.

3.3 UNWEIGHTED/ COMPOSITE INDEX

In this estimation technique, the variables need to be unidirectional. As we discussed earlier, since some of the variables are negatively correlated with other variables in the correlation matrix (e.g., size of household and casual wage rate in rural areas, and size of household in urban areas), they are therefore made unidirectional.

Further for the composite index method, we used a simple average of our 8 variables (normalised and unidirectional) both in rural and urban areas.

3.4 DETERMINING NUMBER OF CLUSTERS

As mentioned above, the Hierarchical cluster analysis (scree plot of coefficient)[1], which starts from n (88 NSS regions in our case) clusters

1 The Hierarchical analysis in our database is based on 88 observations and so the agglomerative schedule consists of 87 stages of clustering process. The coefficient of each stage represents the distance between two clusters combined. When there is large difference between the coefficients of two consecutive stages, it suggests that clusters being merged are increasing in heterogeneity and that would be ideal to stop clustering process. The scree plot of coefficients by stages clearly shows these breaks.

and then aggregates different observations to arrive at 1 cluster, is the most suitable method to identify the number of clusters for our analysis. Accordingly, from 88 NSS regions, clear breaks are visible at 10 and 5 clusters in rural areas, and 10 and 6 clusters in urban areas. Hence, as suggested from the scree plot breaks at 10 cluster and 5 or 6 cluster grouping and the great diversity of India, we have decided to use the 10 grouping variants for all three methods (Hierarchical; K-Means; and Composite Index) for both rural and urban areas in the next section.

4. ANALYSIS OF CLUSTERS IN RURAL AREAS

In this section, we discuss results of the three methods starting with the unweighted composite index followed by Hierarchical and the K-means. We conclude this section with a detailed discussion of results from the K-means analysis since we conclude that this method provides the best result for dividing India into living income and living wage zones.

In a country as diverse as India, it is important that clusters are geographically next to one another. This would give them much greater face validity. Stakeholders and others would be much more willing to accept that two areas fall into the same living wage/living income zone if they are next to one another than if they are thousands of miles apart, because locations bordering on one another are much more likely to have similar socio-demographic and economic characteristics, resulting in similar lifestyles and similar costs of living. Therefore, a key criterion we use for evaluating results from the different cluster analyses is the number of contiguous NSS regions in each cluster compared to the number of NSS regions in the cluster. This is a more important criterion for rural than for urban clusters.

A second way in which we evaluate the clustering method is how evenly NSS regions fall into clusters. Ideally, we would like to avoid that most NSS regions fall in a small number of clusters.

A third way to evaluate clusters is to see if more NSS developed regions tend to be in the same cluster. This is used more for urban than for rural areas, because contiguity in urban areas is less important than for rural areas.

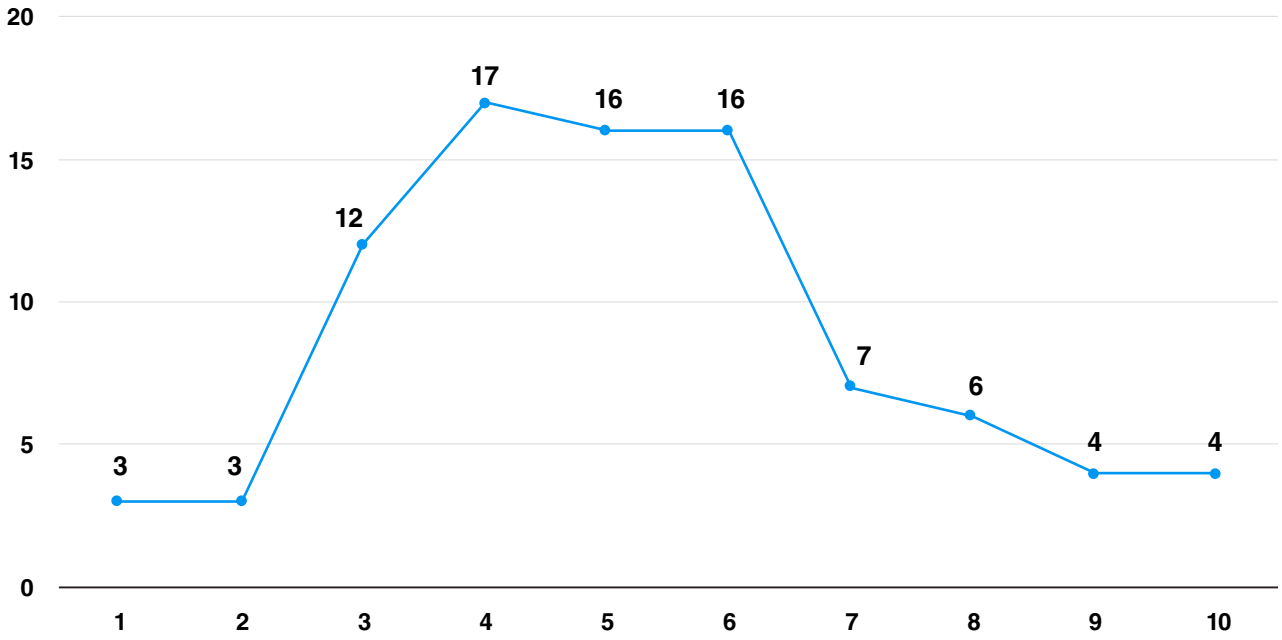
4.1 UNWEIGHTED COMPOSITE INDEX

The unweighted composite index is the average of the eight variables selected in section 2 above for indicating living wage and living income zones in rural areas (slightly different variables were selected for urban areas). The 10-clusters grouping discussed below has been done by equally dividing the range of index values into 10 groups after ranking index values in ascending order. The index allows us to rank the clusters.

Cluster 1 contains NSS regions with lowest set of index values and cluster 2 contain NSS regions with second lowest grouping of index values, and so forth. Graph 1 shows the distribution of 88 rural NSS regions across 10 cluster groups. It shows that the distribution of observations (NSS regions) is not uniformly spread across these clusters. The values are concentrated in clusters 4, 5 and 6 with cluster 4 having highest modal number of 17 regions followed by 16 regions each in clusters 5 and 6 (Graph 1). These 3 clusters contain 50 NSS regions out of total 88 NSS regions. In all the clusters below cluster 4 and above cluster 6, the

number of NSS regions in each cluster declines sharply. It shows that most rural NSS regions are around mid-range reflecting dominance of moderately developed states.

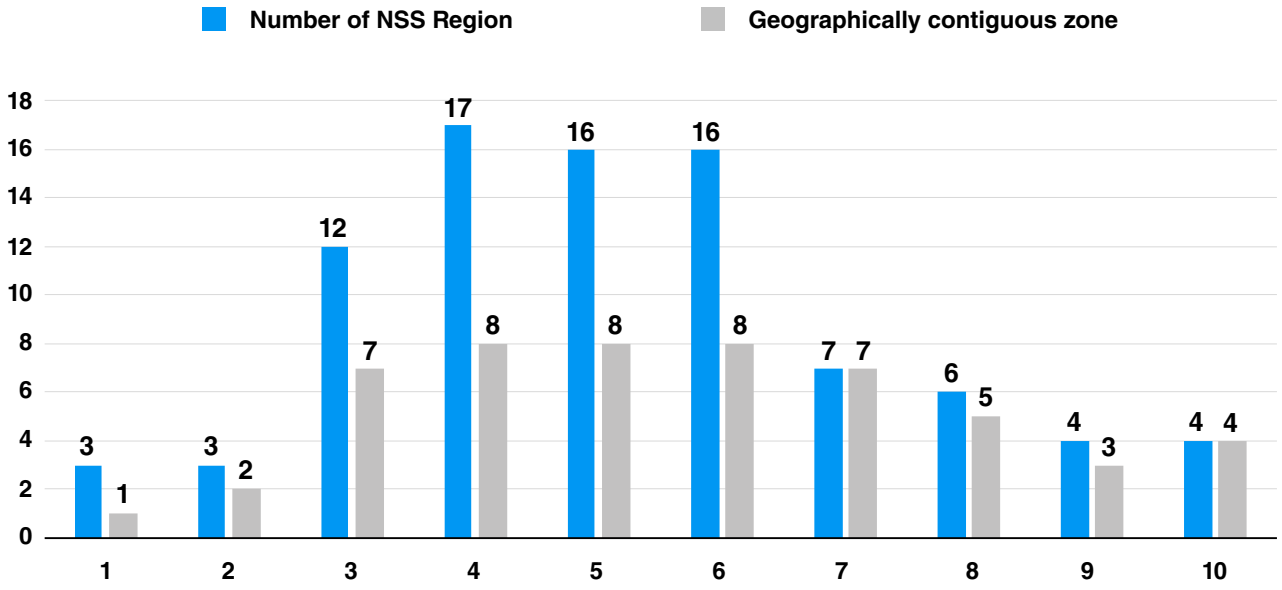
Graph 1: Number of rural NSS Regions in Each of 10 Clusters in Unweighted Composite Index Method



The 1st cluster with lowest index values contains two NSS regions of Bihar and one NSS region of Jharkhand state of Eastern India. These are regions with highest poverty level and they are geographically contiguous regions. In contrast, cluster 10 representing group of four NSS regions with highest index values are geographically far apart. It contains NSS regions belonging to Himalayan mountainous state of Himachal Pradesh in Northern India, coastal state of Goa in Western India, Coastal NSS region of most Southern state of Kerala, and island union territory of Andaman & Nichobar that is 800 kilometres away from mainland. None of these regions are agriculturally prosperous regions but heavily dependent on tourism. The least developed cluster is geographically clustered but the most developed cluster regions are at least one thousand kilometres away from one another.

The next most developed cluster 9 also includes 4 NSS regions -- three mountainous regions and one coastal region. These two mountainous regions fall in Northern Indian states of Jammu & Kashmir and Himachal Pradesh that are geographically contiguous and one falls in Eastern Indian state of Sikkim. The coastal region is Dadra and Nagar Haveli union territory in Western part of India. In contrast, the three NSS regions belonging to 2nd cluster has two NSS regions -- Western Plain of West Bengal and Southern Region of Orissa – which are poor areas and are geographically contiguous in Eastern India. The other region in this cluster is northern upper Ganga plain that lies next to capital Delhi.

Graph 2: Number of rural NSS regions in each cluster (blue bars) along with number of geographically contiguous regions in each cluster (orange bars) using Unweighted composite index method



Graph 2 shows the number of geographically contiguous regions in each cluster. When a cluster contains fewer contiguous NSS regions than the total number of NSS regions, this means that one or more NSS regions share a border with each other). It is clear from the graph 2 that a certain degree of geographical contiguity exists from cluster 1 to cluster 6 (since there are several geographically contiguous regions in these living income zone clusters). Geographical contiguity is either absent or quite weak from cluster 7 onwards that contain rural NSS regions with higher composite index values representing higher socio-economic developed states. These top four clusters that contain 21 NSS regions belong to 18 geographical contiguous zones. In contrast, cluster 4 to cluster 7 contain altogether 61 NSS regions (out of a total of 88 NSS regions,), but they belong to 31 geographical contiguity zones.

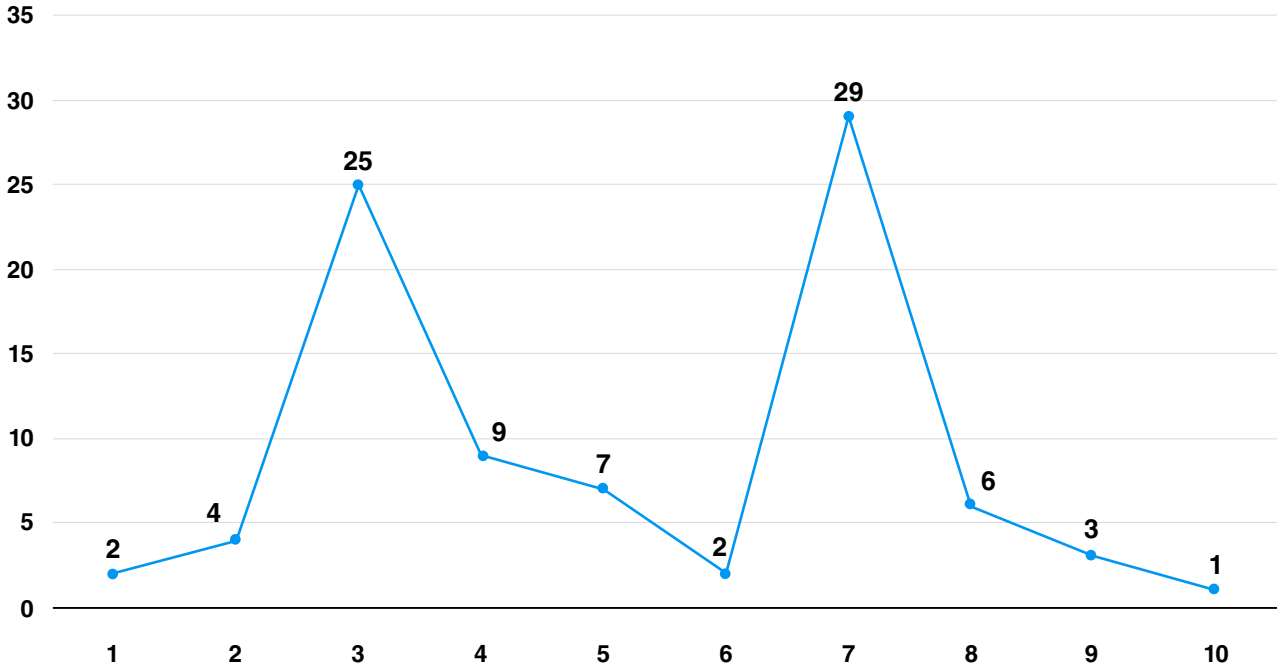
In the Annex VI, the Annex Table 1 presents the composition of each cluster in terms of specific NSS regions along with their location in states or union territories, and also specific geographically contiguous zones. The Annex VI Map 1 gives the geographical composition of each of the 10 clusters.

Our conclusion is that the unweighted composite index method does not provide very good results in terms of a fairly even distribution of NSS regions in clusters in terms of the number of contiguous regions within clusters – and therefore is not used to determine living income zones in this report.

4.2 HIERARCHICAL METHOD

The Hierarchical methodology assigns observations (in our case NSS regions) into different clusters on the basis of similarity/ dissimilarity among them. This method identifies similarity on the basis of weighted value obtained from Euclidian distance, and it does not consider evenness of the distribution of observations across different clusters. It is the extent of dissimilarity obtained from Euclidian distance that directs to put the values in different clusters. In the case of rural India, two clusters include 29 and 25 NSS regions respectively (Graph 3). These two clusters together take in more than 60 per cent of all NSS regions. As a result, the number of NSS regions vary from 1 to 4 in five other clusters. This is not a good attribute.

Graph 3: Number of Rural NSS Regions in Each Cluster in Hierarchical Method



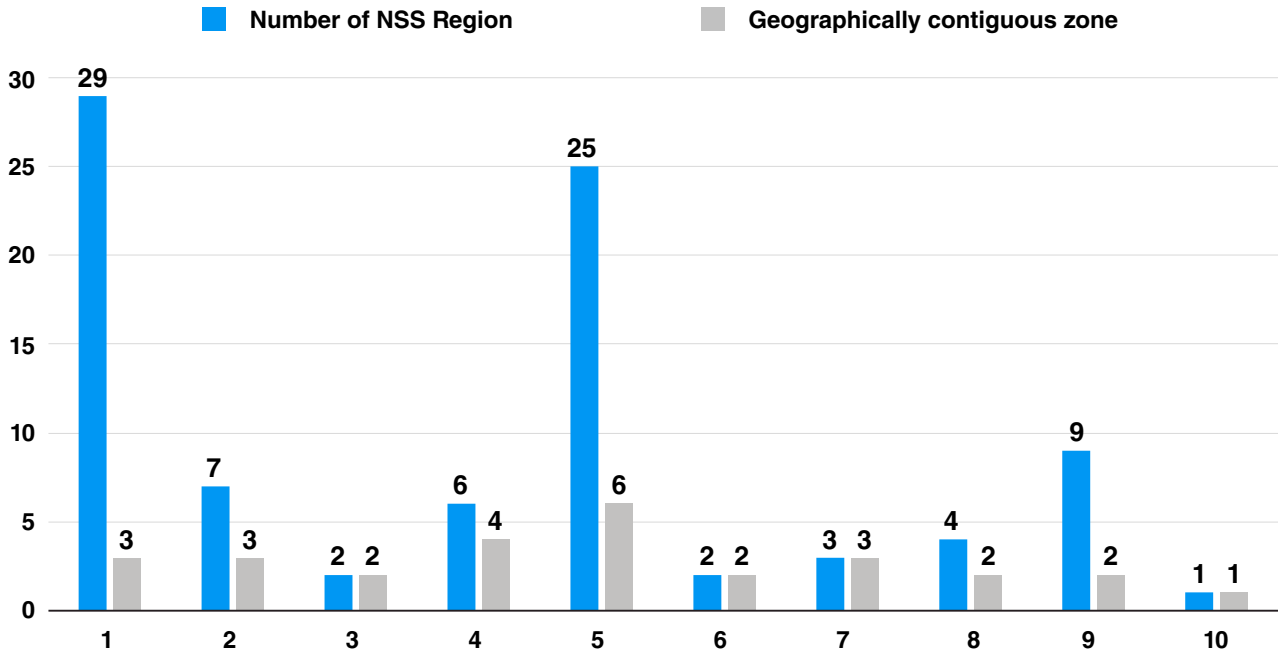
Since clusters in Hierarchical methodology do not give any mean cluster values, it does not rank clusters in terms of socio-economic development. We have ranked them on the basis of derived cluster value using the average individual unweighted composite index value of different NSS regions within each of these clusters. Following this pattern of ranking, it can be observed that lowest ranked cluster contains two NSS regions and highest ranked cluster includes only one NSS region (see Graph 3). According to Hierarchical methodology, these NSS regions are quite dissimilar from other NSS regions. This methodology shows the highest ranked cluster is 10th cluster, and it includes Andaman & Nichobar Island that is located far away from mainland and has distinct characteristics. Two other clusters (1st and 6th) include two NSS regions each. In the 1st cluster, the two NSS regions are located in North-Western India in the Valley of Himalayan Mountain and the other one just below Himalayan Mountain region. The other cluster (6th) includes two mostly geographically small NSS regions that are mostly urban. These NSS regions are Daman & Diu in Western India and Capital region of Delhi.

The number of NSS regions across clusters varies substantially with a range from 1 to 29, but the dissimilarity is much less when we consider geographically contiguous zones within each cluster separately (see Graph 4). The range of number of contiguous zones is 1 to 6 across different clusters. In largest two clusters with 29 and 25 NSS regions, the corresponding number of contiguous zones are 3 and 6 respectively. In the four clusters having 1 to 3 NSS regions, geographical contiguity exists.

In Annex VI, Annex Table 2 presents the composition of each cluster in terms of specific NSS regions along with their location in states or union territories and specific geographically contiguous zones in Hierarchical method. Also, Annex Map 2 gives the geographical composition of each of 10 clusters.

Our conclusion is that results from Hierarchical method do not produce a sufficiently robust basis for dividing India into living income/living wage zones, because the majority of NSS regions fall within only 2 clusters, although we do find them useful for helping to inform results from the K-means method (see Section F below).

Graph 4: Number of NSS Regions (blue bars) and Geographically Contiguous Zones (orange bars) in each Cluster in Rural India (Hierarchical method)

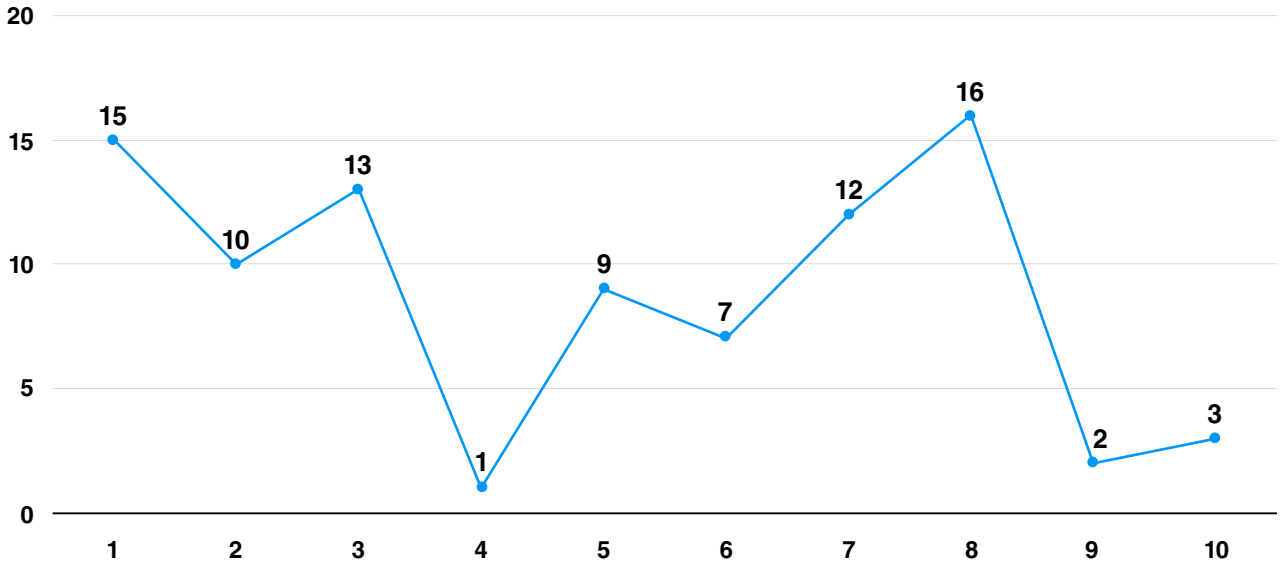


4.3 K-MEANS METHOD

The K-means cluster methodology does not give any mean cluster value, therefore, the ranking of different clusters in terms of their socio-economic development level is not provided. Similar to the analysis of Hierarchical cluster, we calculated mean cluster values using our unweighted composite index for NSS regions falling within each cluster, and we ranked clusters on this basis. Even after this ranking, the distribution of NSS regions across clusters does not show any pattern in terms of development. However, the distribution of NSS regions across different clusters is more even in the K-means methodology compared to the Hierarchical methodology. At least five clusters contain at least 10 NSS regions each. After ranking, highest number of NSS regions fall in the 8th cluster having 16 NSS regions. It is closely followed by cluster 1 having 15 NSS regions. There are only 3 clusters that have less than four NSS regions each (see Graph 5).

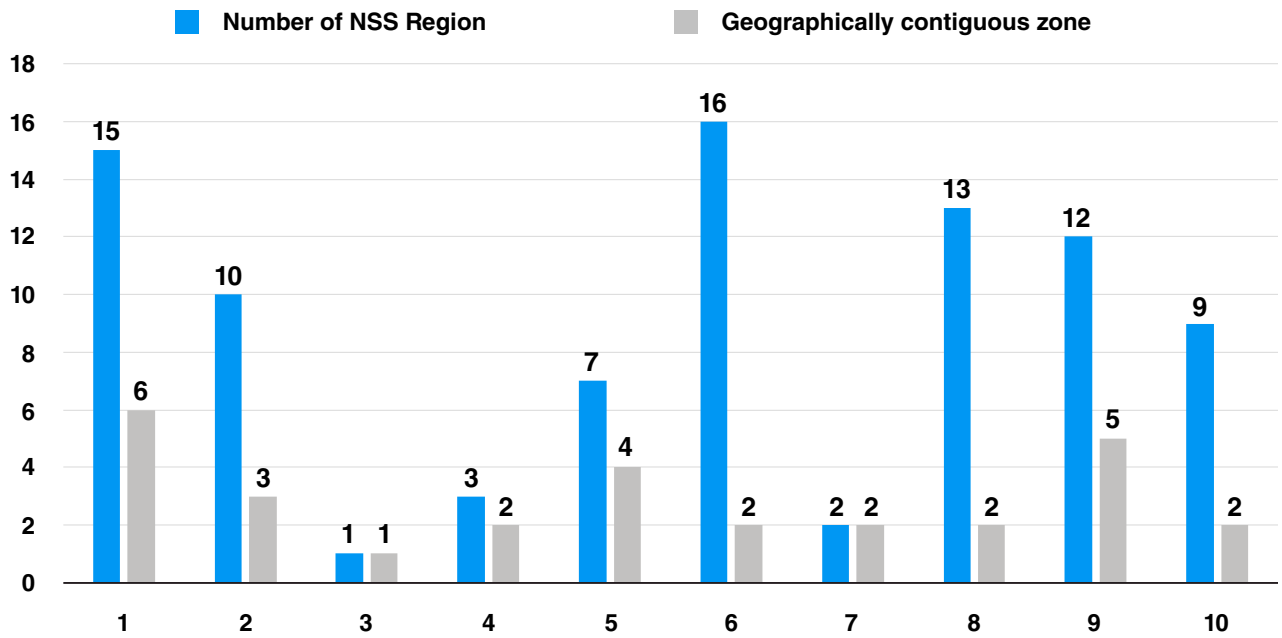
The reason for less diverse composition of the K-means cluster methodology compared to Hierarchical methodology is that it follows a different procedure. In K-means methodology, NSS regions are initially divided into 10 different clusters as we have fixed the number of clusters to be 10. In each cluster, it calculates the sum of squared deviation of all observations from respective cluster mean. (Note that the cluster mean is the mean of the 8 variables selected in section B.) It adds these squared deviations for all 10 clusters to arrive at overall sum of squared deviation. Then it follows an iterative process to shift NSS regions to neighbouring clusters to minimise the overall sum of squared deviations. In other words, it partitions N (88) observations into K (10) clusters in which each observation belongs to the cluster with nearest mean. The basic idea is to make each cluster more homogeneous and differences between clusters more heterogeneous. Conceptually K-means is similar to Hierarchical cluster analysis, but it is achieved through different mechanism. In Hierarchical cluster method, this similarity is examined for each pair of observations not as a group.

Graph 5: Number of NSS Regions in Each Cluster in K-Means Method for Rural Areas



Even though the number of NSS regions across clusters varies from 1 to 16, the dissimilarity is much less when we consider the degree to which NSS regions are geographically contiguous within each cluster (see Graph 6). The number of contiguous zones within each cluster varies from 1 to 6 across different clusters. In largest two clusters having 16 and 13 NSS regions, the corresponding number of contiguous zones are only 2 in both cases. In two other clusters having 15 and 10 NSS regions, there are only 3 and 6 contiguous zones respectively. Thus, clusters having large number of NSS regions often have large areas that are geographically contiguous. In smaller clusters, geographical contiguity necessarily does not exist.

Graph 6: Number of NSS Regions (blue bars) and Number of Geographically Contiguous Zones (orange bars) in each Cluster in K-means Method, Rural India



Comparing all three methods for rural India, results from the K-means method are much more suitable for identifying living income zones. Clusters in the K-means method are much more geographically contiguous regarding NSS regions compared to Hierarchical method and Unweighted composite index method, and K-means method more evenly distributes NSS regions in the clusters compared to the Hierarchical method. This is discussed in more detail below where living income zones in rural India are identified.

4.4 DETAILED DESCRIPTION OF CLUSTERS DERIVED FROM THE K-MEANS METHOD

As K-means cluster is the more suitable method for identification of living income zones, the composition of clusters within them is discussed more in detail here. A similar analysis of unweighted means and Hierarchical cluster is given in Annex VI.

Table 4 presents the composition of each K-means cluster in terms of specific NSS regions along with their location in state or union territory and specific geographically contiguous zones. Map 1 also gives geographical composition of each of the 10 clusters.

The 1st cluster (that contains 15 NSS regions) is divided into 6 contiguous zones. The largest number of NSS regions is in zone 4 that covers whole of Jharkhand state and 2 NSS regions of Orissa state both of them are located in Eastern part of India. The contiguous zone 2 is more populous as it covers whole Bihar state of Eastern India and one NSS region of Northern Indian state of Uttar Pradesh. The 2nd cluster (having 10 NSS regions) includes a somewhat large contiguous zone 7 that includes whole state of West Bengal and one NSS region of Orissa state in Eastern India.

In contrast, the 6th cluster (that contain 16 NSS regions), the 18th geographically contiguous zone comprises 15 NSS regions spreading over whole of Punjab and Haryana state of Northern India and substantial part of most populous state of Uttar Pradesh in Northern India and also substantial part of Rajasthan and Gujarat state of Western India. The 2nd cluster (without ranking that contains 10 NSS regions) has somewhat large contiguous zone 7 with 6 NSS regions having whole state of West Bengal and one NSS region of Orissa state in Eastern India.

Among the smaller clusters, the 3rd cluster has one island union territory of Lakshadweep located away from mainland. The other smaller cluster 7th includes two small NSS regions that are mostly urban. These NSS regions are Dadra & Nagar Haveli in Western India and Capital region of Delhi. Another small cluster (4th) contains two NSS regions of state of Kerala in Southern India and the island union territory of Andaman & Nichobar.

Once again, note that compared to Hierarchical cluster method, in K-means cluster method the number of NSS regions assigned to different clusters is much less diverse in the sense that 5 out of 10 clusters contain at least 10 NSS regions each.

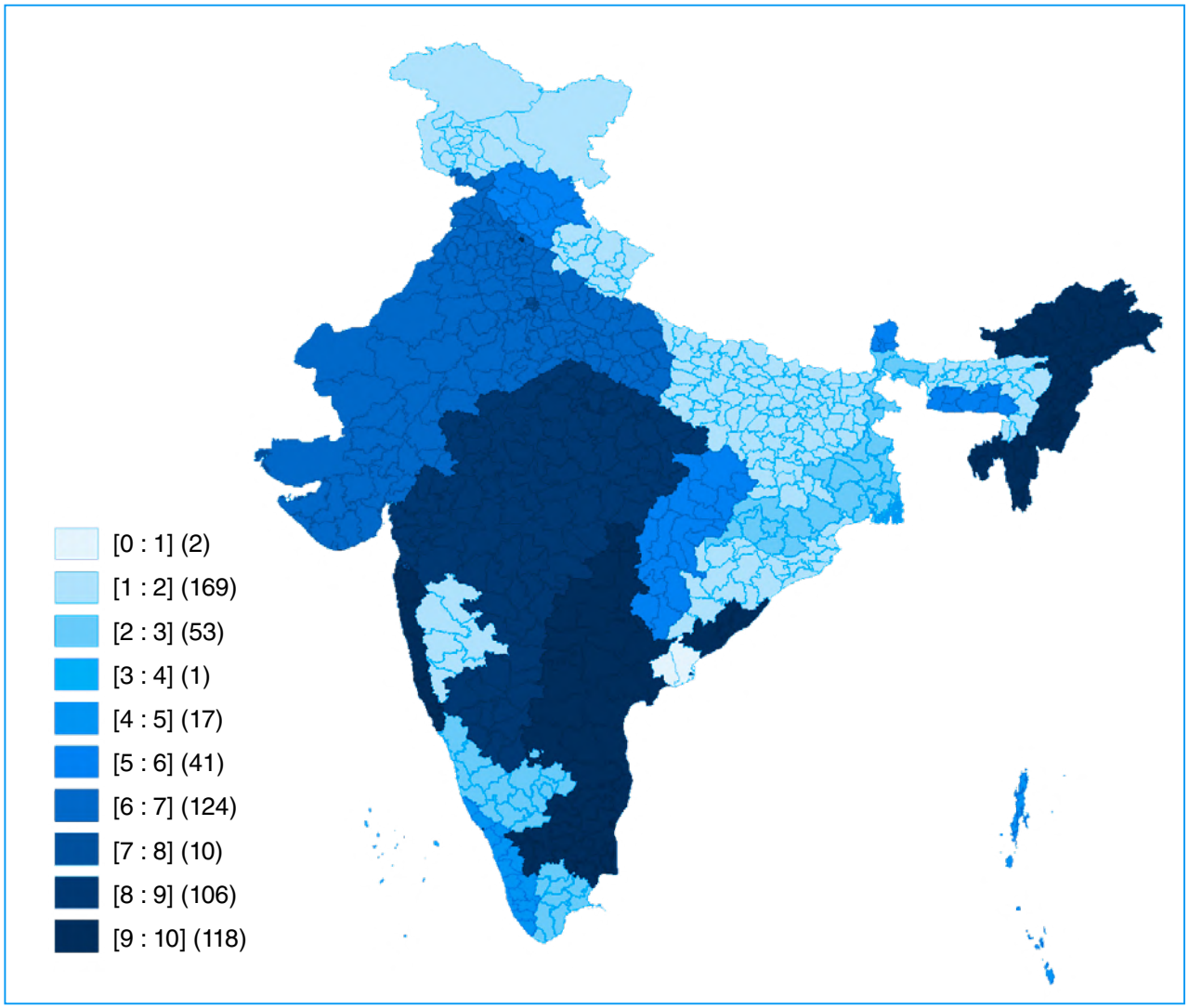
Table 4: Composition of Clusters in Rural India in K-Means Cluster Methodology

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
1st Cluster, unweighted value= 0.307	1	Uttarakhand	All (1)	0.370
	2	Bihar	Northern Plain, Southern Plain (1)	0.234
	2	Uttar Pradesh	Eastern (1)	
	3	Assam	Plain Western, Cachar Plain, Central Brahmaputra Plain (1)	0.320
	4	Jharkhand	Ranchi Plateau, Hazaribagh Plateau (1)	0.266
	4	Orissa	Coastal, Southern (1)	
	5	Maharashtra	Inland Western (1)	0.350
	6	Jammu & Kashmir	Outer Hills, Ladakh, Jhelum Valley (2)	0.304
2nd Cluster, unweighted value= 0.311	7	Orissa	Northern (2)	0.289
	7	West Bengal	Eastern Plain, Southern Plain, Central Plain, Himalayan, Western Plain (2)	
	8	Karnataka	Coastal & Ghats, Inland Eastern, Inland Southern (2)	0.344
	9	Tamil Nadu	Southern (2)	0.389
3rd Cluster, unweighted value=0.348	10	Lakshadweep	All (3)	0.348
4th Cluster, unweighted value= 0.510	11	Kerala	Northern, Southern (4)	0.479
	12	Andaman & Nichobar	All (4)	0.542
5th Cluster, unweighted value= 0.419	13	Himachal Pradesh	Central Plain, Southern, Trans Himalayan (5)	0.502
	14	Chhattisgarh	Mahanadi Basin, Southern Chhattisgarh, Northern Chhattisgarh (5)	0.365
	15	Sikkim	All (5)	0.492
	16	Meghalaya	All (5)	0.406

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
6th Cluster, unweighted value= 0.442	17	Jammu & Kashmir	Mountainous (6)	0.506
	18	Punjab	Northern, Southern (6)	0.377
	18	Haryana	Eastern, Western (6)	
	18	Rajasthan	Western, North-Eastern, Southern, Northern (6)	
	18	Gujarat	Plain Northern, Dry Areas, Kachchh, Saurastra (6)	
	18	Uttar Pradesh	North Upper Ganga Plain, Central, South Upper Ganga Plain (6)	
7th Cluster, unweighted value= 0.468	19	Delhi	All (7)	0.432
	20	Dadra & Nagar Haveli	All (7)	0.504
8th Cluster, unweighted value= 0.322	21	Uttar Pradesh	Southern (8)	0.336
	21	Madhya Pradesh	Vindya, Central, Malwa, South, South-Western, Northern (8)	
	21	Rajasthan	South-Eastern (8)	
	22	Gujarat	South Eastern (8)	0.307
	22	Maharashtra	Inland Northern, Inland Central, Inland Eastern (8)	
	22	Karnataka	Inland Northern (8)	
9th Cluster, unweighted value= 0.429	23	Chandigarh	All (9)	0.468
	24	Arunachal Pradesh	All (9)	0.395
	24	Nagaland	All (9)	
	24	Manipur	Hills, Plain (9)	
	24	Mizoram	All (9)	
	24	Assam	Plain Eastern (9)	
	24	Tripura	All (9)	
	25	Maharashtra	Coastal (9)	0.380
	25	Goa	All (9)	0.445
	26	Puducherry	All (9)	0.469
	27	Daman & Diu	All (9)	0.359

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
10th Cluster, unweighted value= 0.362	28	Andhra Pradesh	Coastal Northern, Coastal Southern, Inland Southern (10)	0.342
	28	Telangana	Inland North Western, Inland North Eastern (10)	
	28	Maharashtra	Eastern (10)	
	29	Tamil Nadu	Coastal Northern, Coastal, Inland (10)	0.389

Map 1: K-Means Clusters, Rural India (number of districts in each cluster in brackets)



Note: Cluster numbers (from 1 to 10) do not indicate any ranking in K-means method. It indicates number of different NSS regions.

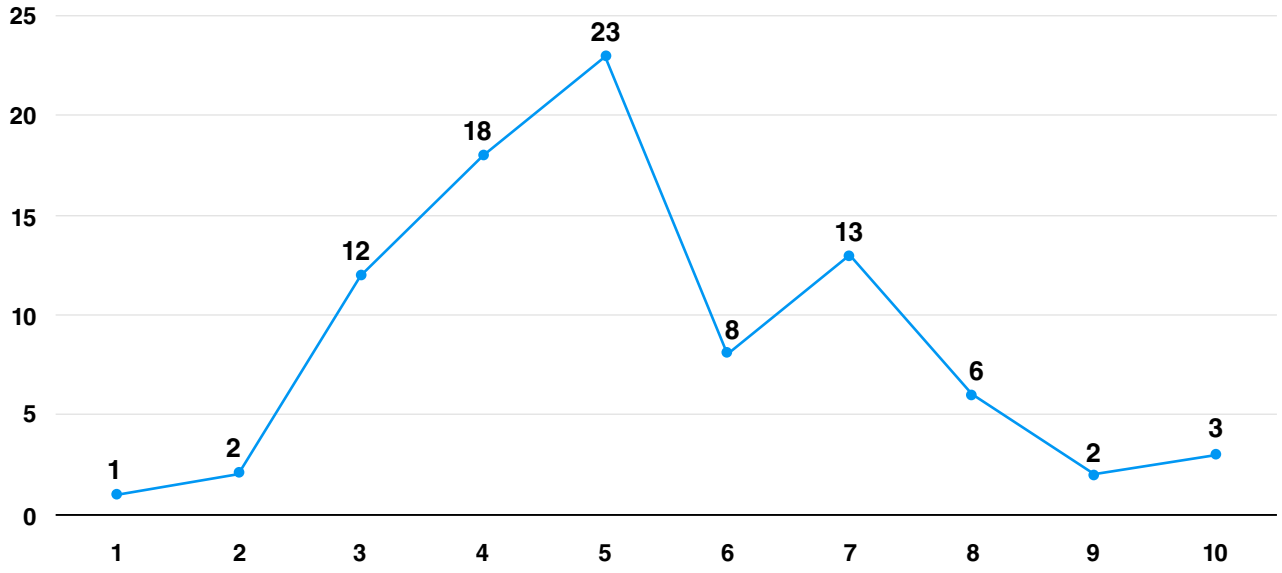
5. ANALYSIS OF CLUSTERS IN URBAN AREAS

Similar to the analysis for rural areas, the cluster analysis for urban areas uses three aggregation methodologies – with the conclusion that the K-means method is best. These are discussed below.

5.1 UNWEIGHTED COMPOSITE INDEX METHODOLOGY FOR URBAN AREAS

Similar to rural areas, the unweighted composite index is calculated for urban areas. The 10-cluster urban grouping is done in a similar fashion. Graph 7 indicates the distribution of the 88 NSS regions across 10 clusters. The fifth cluster contains 23 NSS regions. If the 4th and 5th clusters are taken together, they account of 47 per cent of all NSS regions. In contrast, the first two clusters together contain only three NSS regions and the 9th and 10th clusters contain only 5 NSS regions. These four clusters are clearly outliers.

Graph 7: Number of NSS Regions in Each Cluster in Unweighted Index in Urban India



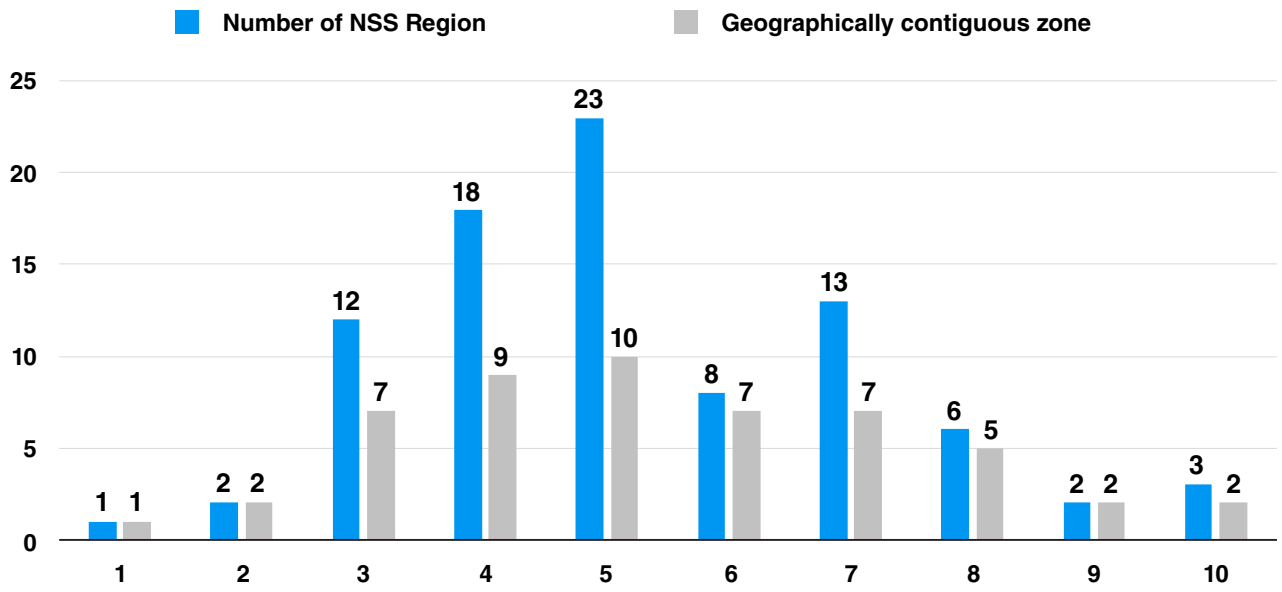
The first cluster has the single NSS region of Northern of the state of Bihar. It is very poor region with very low urbanisation and with negligible presence of industrial activity. The second cluster includes two NSS regions of neighbouring states of Bihar and West Bengal with Central Plain of Bihar and Western Plain of West Bengal. All these three NSS regions belonging to cluster 1 and 2 have low urbanisation rates and high levels of urban poverty.

In contrast, all three NSS regions in the 10th cluster with highest composite index values are all small union territories. Daman & Diu and Dadra & Nagar Haveli are situated close by in the western part of India close to coastal region of the state of Gujarat. Recently, they have been combined together and converted into single union territory. These two NSS regions have very high presence of organised manufacturing and more than half of all workers are engaged in organised manufacturing sector. The third

NSS region in cluster 10 is the union territory of Chandigarh and it is the capital city of two neighbouring states of Punjab and Haryana located in Northern India. Employment in this capital city is dominated by the organised service sector with large presence of government servants belonging to both states of Punjab and Haryana.

Graph 8 presents numbers of NSS regions in each urban cluster along with number of geographically contiguous zones in the urban areas of each cluster.

Graph 8: Number of NSS Regions (blue bars) and Number of Geographically Contigous NSS Regions (orange bars) in each Cluster in Urban India (Unweighted Index Method)



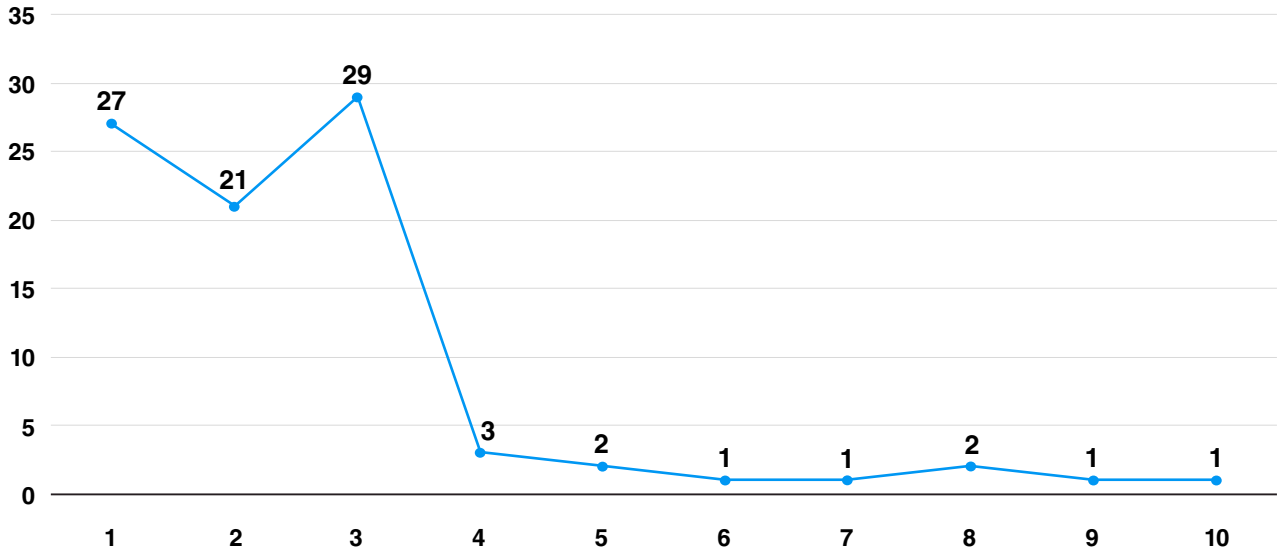
The argument for having geographical contiguity for NSS regions for urban areas is not as strong as for rural areas, because urban areas are comparatively smaller geographical area with higher concentration of population and economic activities. But it does show whether level of development has spread over wider areas of cities and towns that are in neighbouring NSS regions. This broader area of development or non-development is limited to four clusters (3rd, 4th, 5th and 7th) that also have larger number of NSS regions within them.

In Annex VII, the Annex Table 3 presents the composition of each cluster in terms of specific NSS regions along with their location in states or union territories and specific geographically contiguous zones. Also, Annex VII Map 3 gives geographical composition of each of the 10 clusters.

5.2 HIERARCHICAL CLUSTER METHODOLOGY FOR URBAN AREAS

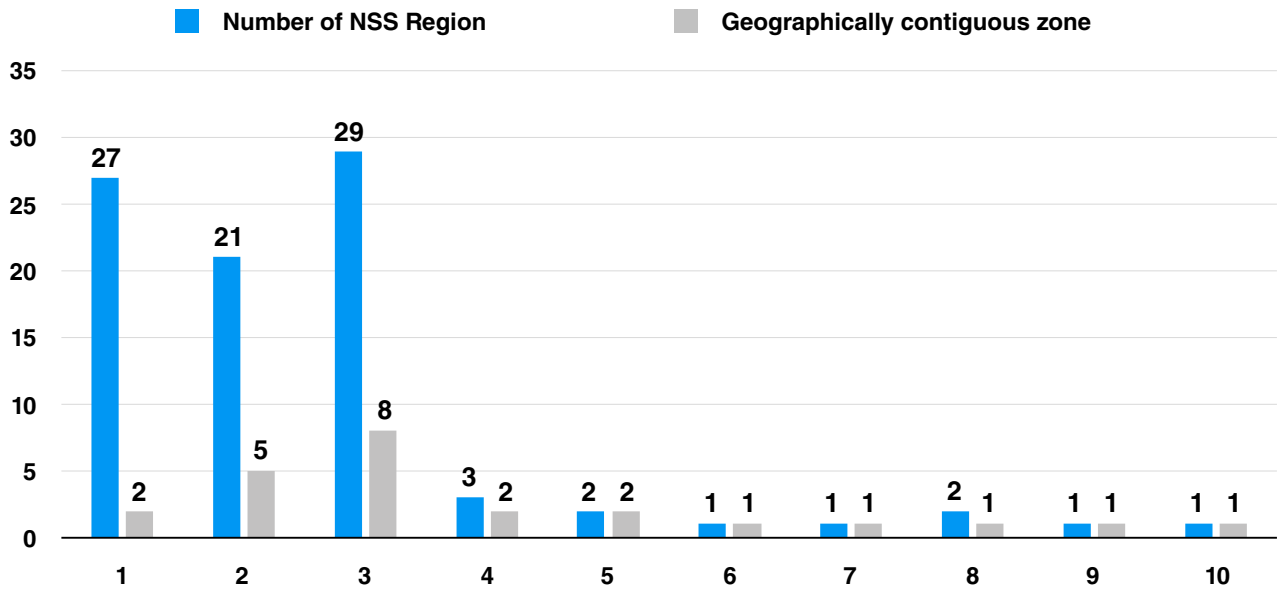
In urban areas, Hierarchical method results show a large concentration of NSS regions in the first three clusters as they contain seventy-seven out of total 88 NSS regions (see Graph 9). The remaining 7 clusters contain only 11 NSS regions with 4 clusters having single NSS region within them. In other words, clusters 4 to 10 are outliers in the sense that these are quite dissimilar to first three clusters.

Graph 9: Number of NSS Regions in Each Cluster in Hierarchical Method in the Urban Areas



Clusters 4 to 10 are also dissimilar from the first three clusters as there are much fewer geographical contiguous zones (see Graph 10). For example, cluster 1 contains 27 NSS regions but these are divided into 2 contiguous zones compared cluster 4 that has 3 NSS regions but has 2 contiguous zones.

Graph 10: Number of NSS Regions (blue bars) and Number of Geographically Contiguous Zones (orange bars) in Each Cluster in Urban India (Hierarchical method)

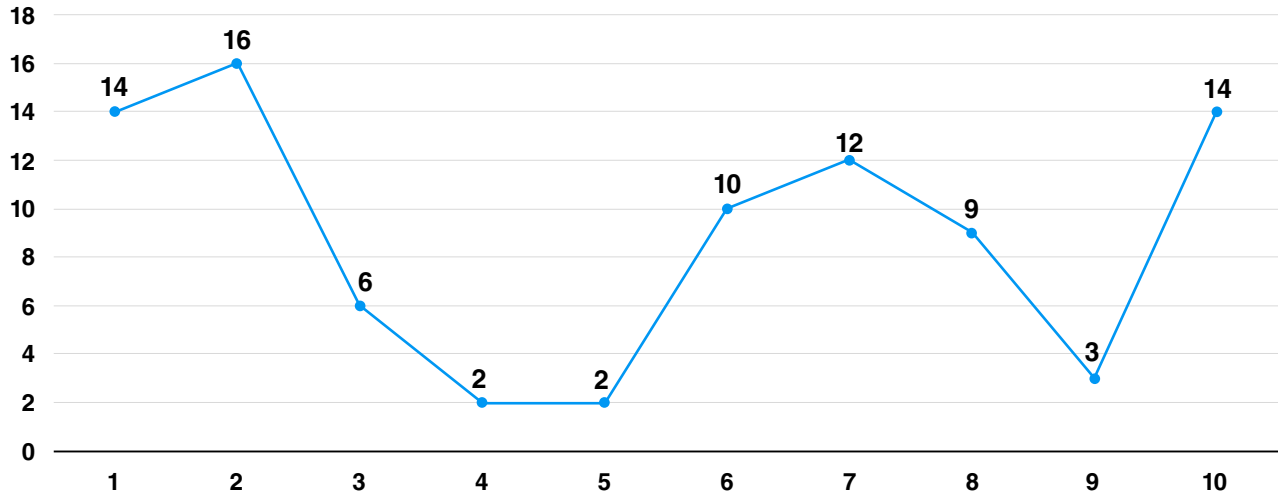


Our conclusion is that results from the Hierarchical method are not suitable by themselves for determining living income/living wage clusters for urban India - although we also conclude above that they are useful for advising on the appropriateness of results from K-means method.

5.3 K-MEANS CLUSTER METHODOLOGY FOR URBAN AREAS

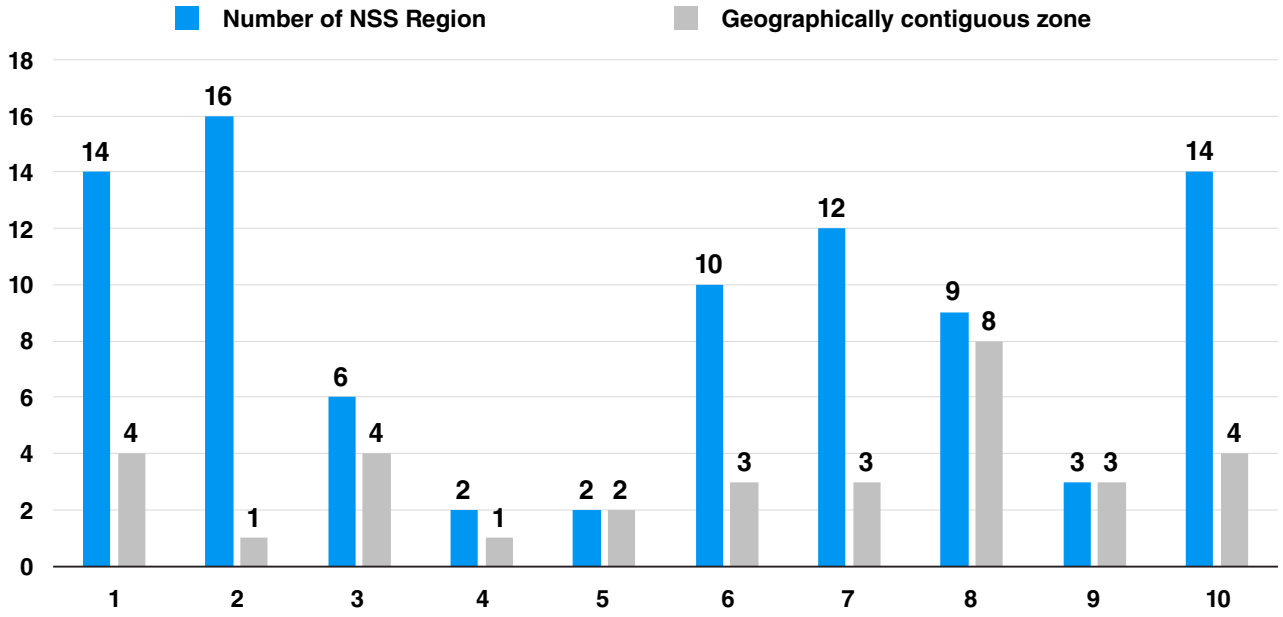
In K-means cluster methodology for urban India, the number of NSS regions is highest in 1, 2, 6, 7, 8, and 10 clusters (see Graph 11). However, the magnitude of variations in the number of NSS regions in each cluster varies between 2 and 16. Five clusters, with minimum of 10 NSS regions each, together comprise 75 per cent of all NSS regions.

Graph 11: Number of NSS Regions in Each Cluster in K-Means Method for Urban Areas



The number of geographical contiguous zones shows less variation except of 8th cluster where there are 8 geographically contiguous zones with nine NSS regions. In other clusters with 10 or more NSS regions, geographical continuity exists. In particular, the largest cluster (2nd cluster) has all 16 NSS regions in one contiguous zone (see Graph 12).

Graph 12: Number of NSS Regions (blue bars) and Number of Geographically Contiguous Zones in each Cluster (orange bars) in Urban India (K-Means method)



The unweighted index shows 52 contiguous zones but in Hierarchical and K-means methods, there are 24 and 33 zones respectively. In urban areas, contiguous geographical zones are comparatively higher in K-means method but the differences in the size of clusters is much higher in Hierarchical method as its three largest clusters include 29, 27 and 21 NSS regions whereas in K-means method, 5 clusters have more than 10 NSS regions each. As will be seen in this section, this means that the K-means analysis more effectively classifies less developed urban areas in specific clusters and more developed urban areas with different characteristics in other clusters compared to Unweighted index and Hierarchical methods.

Comparing all three methods for urban India, K-means method is found to be more suitable for identification of living income zones. This aspect will be discussed in detail in the section where living income zones in urban India are identified.

As K-means clusters are more suitable as the basis of identification of living income zones, the composition of clusters within them is discussed more in detail here. Similar analyses of unweighted composite index and Hierarchical cluster are shifted to Annex VII.

Table 5 presents the composition of each cluster in terms of specific NSS regions along with their location in state or union territories and specific geographically contiguous zones. Map 2 gives geographical composition of each of the 10 clusters.

The 5th geographical contiguous zone in cluster 2 contains not a single whole state but contains 16 NSS regions belonging to 6 states (see Graph 12). These NSS regions belong to Uttar Pradesh, Punjab and Haryana of northern India, Rajasthan and Gujarat of Western India and Madhya Pradesh of Central India. In cluster 1, there is one large contiguous zone having five NSS regions that cover whole of two South Indian states of Andhra Pradesh and Telangana. Similarly in 10th cluster, the 32nd contiguous zone has 5 NSS regions, three Eastern Indian states of Bihar, Jharkhand and West Bengal and one region of North Indian state of Uttar Pradesh. This contiguous zone is one of the poorest urban areas with much less presence of organised manufacturing and service activities.

The 17th contiguous zone of the 7th cluster includes seven NSS regions of North-Eastern India mostly consisting of less developed small urban towns in furthest corner of India with less developed transport system development.

Cluster 4 has two union territories of Daman & Diu and Dadra & Nagar Haveli in Western India with high share of organised manufacturing activities. Cluster 9 is also a specialised cluster with three NSS regions that contain Cities of Mumbai, Delhi and Chandigarh with high concentration of organised service activities in government and corporate sector.

In nutshell, K-means analysis separate less developed urban areas in specific clusters and more developed urban areas with different characteristics in separate clusters more effectively than unweighted index and Hierarchical methodologies.

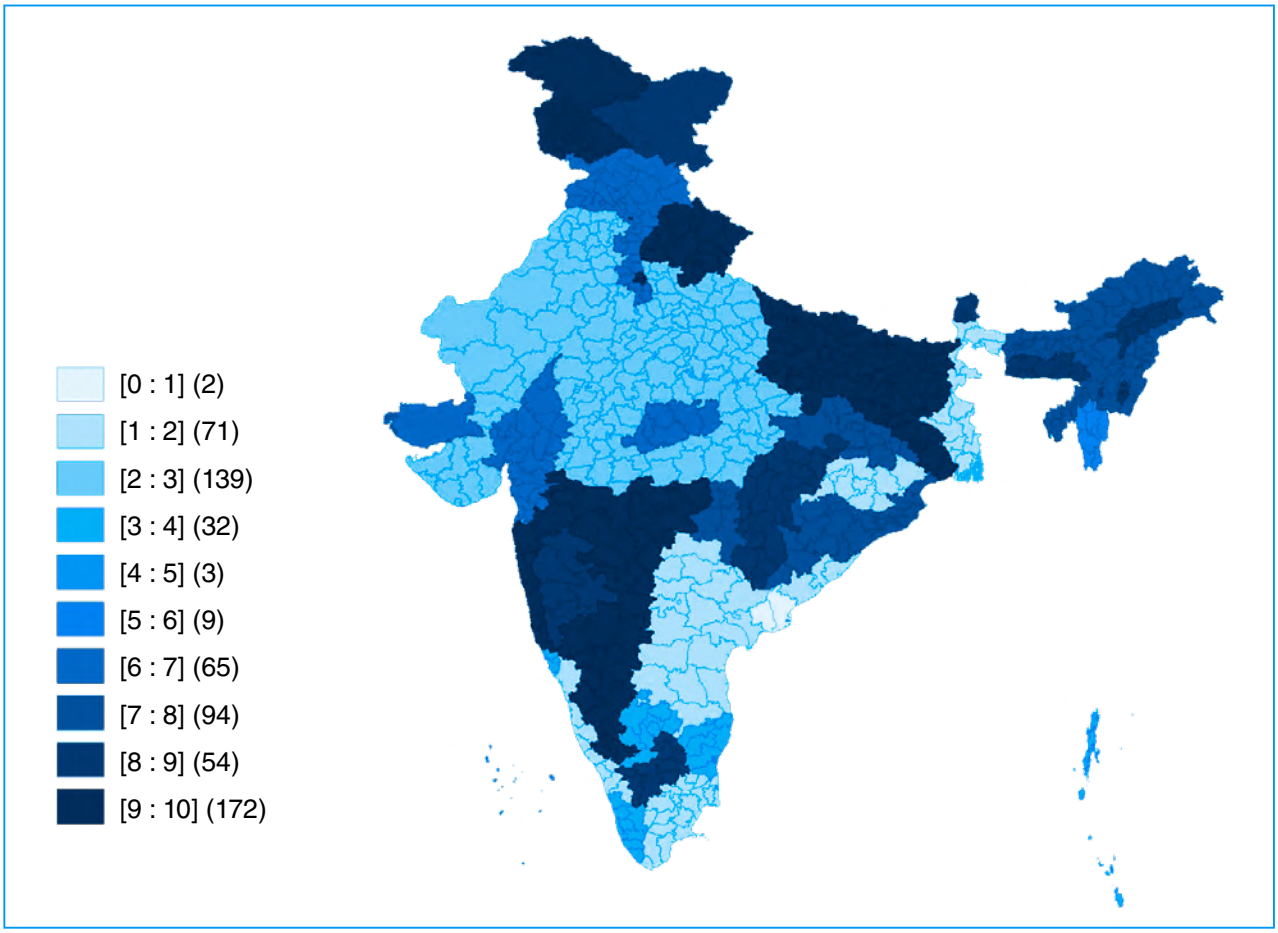
Table 5: Composition of Clusters in Urban India in K-Means Methodology

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
1st Cluster (unweighted index value= 0.319)	1	Andhra Pradesh	Coastal Northern, Coastal Southern, Inland Southern (1)	0.341
	1	Telangana	Inland North -Western, Inland North-Eastern (1)	
	2	Karnataka	Coastal & Ghats (1)	0.238
	2	Kerala	Northern (1)	
	3	Tamil Nadu	Coastal, Southern (1)	0.361
	4	Orissa	Northern (1)	0.289
	4	West Bengal	Himalayan, Eastern Plain, Southern Plain, Central Plain (1)	
2nd Cluster (unweighted index value= 0.327)	5	Punjab	Southern (2)	0.327
	5	Haryana	Western (2)	
	5	Rajasthan	Western, North-Eastern, South-Eastern, Northern (2)	
	5	Gujarat	Dry Areas, Saurastra (2)	
	5	Uttar Pradesh	Central, Southern, South Upper Ganga Plain (2)	
	5	Madhya Pradesh	Vindya, Malwa, South, South-Western, Northern (2)	
3rd Cluster (unweighted index value= 0.458)	6	Karnataka	Inland Southern (3)	0.449
	6	Tamil Nadu	Coastal Northern (3)	
	6	Puducherry	All (3)	
	7	Goa	All (3)	0.448
	8	Kerala	Southern (3)	0.386
	9	Andaman & Nichobar Island	All (3)	0.548
4th Cluster (unweighted index value= 0.585)	10	Daman & Diu	All (4)	0.585
	10	Dadra & Nagar Haveli	All (4)	
5th Cluster (unweighted index value= 0.401)	11	Lakshadweep	All (5)	0.411
	12	Mizoram	All (5)	0.391

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
6th Cluster(unweighted index value= 0.436)	13	Jammu & Kashmir	Mountainous (6)	0.456
	13	Himachal Pradesh	Central Plain, Trans Himalayan & Southern (6)	
	13	Punjab	Northern (6)	
	13	Haryana	Eastern (6)	
	14	Rajasthan	Southern (6)	0.462
	14	Gujarat	South-Eastern, Plain Northern, Kachchh (6)	
	15	Madhya Pradesh	Central (6)	0.391
7th Cluster (unweighted index value= 0.302)	16	Arunachal Pradesh	All (7)	0.281
	16	Nagaland	All (7)	
	16	Manipur	Hills (7)	
	16	Assam	Plains Western, Cachar Plain, Brahmaputra Plain (7)	
	16	Tripura	All (7)	
	17	Orissa	Coastal, Southern (7)	0.292
	17	Jharkhand	Ranchi Plateau (7)	
	17	Chhattisgarh	Northern (7)	
	18	Maharashtra	Eastern (7)	0.336
8th Cluster (unweighted index value= 0.348)	19	Jammu & Kashmir	Ladakh (8)	0.339
	20	Sikkim	All (8)	0.419
	21	Manipur	Plains (8)	0.258
	22	Meghalaya	All (8)	0.320
	23	Assam	Plains Eastern (8)	0.327
	24	Chhattisgarh	Mahanadi Basin, Southern (8)	0.288
	25	Maharashtra	Inland Western (8)	0.386
	26	Tamil Nadu	Inland (8)	0.447
9th Cluster (unweighted index value= 0.531)	27	Chandigarh	All (9)	0.606
	28	Delhi	All (9)	0.488
	29	Maharashtra	Coastal (9)	0.500

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
10th Cluster (unweighted index value= 0.264)	30	Jammu & Kashmir	Outer Hills, Jhelum Valley (10)	0.260
	31	Uttarakhand	All (10)	0.323
	31	Uttar Pradesh	North Upper Ganga Plain (10)	
	32	Uttar Pradesh	Eastern (10)	0.194
	32	Bihar	Northern Plain, Central Plain (10)	
	32	Jharkhand	Hazaribagh Plateau (10)	
	32	West Bengal	Western Plain (10)	
	33	Maharashtra	Inland Northern, Inland Central, Inland Eastern (10)	0.275
	33	Karnataka	Inland Eastern, Inland Northern (10)	

Map 2: K-Means Clusters, Urban (number of districts in each cluster in bracket)



Note: Cluster numbers (from 1 to 10) do not indicate any ranking in K-means method. It shows clusters are different from one another.

6. IDENTIFYING LIVING INCOME ZONES FOR RURAL INDIA

6.1 DIVIDING RURAL CLUSTERS INTO CONTIGUOUS NSS REGIONS

In a previous section, we discussed in detail the results of three cluster analysis methods. It would have been better if the composition of NSS regions in different clusters were similar across these three methods. In the absence of this, some criteria are used to identify rural living income zones (LIZ).

Out of these methods, unweighted index is a supervised method. Composite index is developed through a simple average of standardised values of chosen 8 indicators. The other two methods (Hierarchical and K-means) are unsupervised methods. In unsupervised methods after providing the standardised values to the models, these methods on their own identify NSS regions to be included in which clusters. In that sense, unsupervised methods are supposedly better, because they are free from additional assumptions made to construct a composite index.

In the previous section. we identified 10 K-means clusters. In this section, we subdivide these clusters into sets of contiguous NSS regions. India is a large and diversified country with 1.34 billion population mostly resident of rural areas. There exists a large difference in terms of level of living, food habits, female empowerment, labour market conditions, energy use and public provisioning of education and health facilities across India. These factors are more homogeneous in neighbouring NSS regions of the same state or neighbouring states because geographically contiguous areas or areas in close proximity usually share similar socio-demographic and economic conditions. For this reason, NSS regions within a cluster, in close geographical proximity, are combined into living income zones. All together we identified 24 rural living income zones.

We tabulated in earlier tables the geographically contiguous zones for 10 clusters from original 88 NSS rural regions for all three methods. The unweighted index method shows 53 geographically contiguous zones in 10 clusters from original data of 88 NSS regions. This means that NSS regions falling in the same clusters largely belong to areas either far from one another or at least located some distance away from one another with less homogeneity in the level of living, food habits, energy use, etc. In contrast, in Hierarchical and K-means methods, 10 clusters belong to 28 and 29 geographically contiguous zones with relatively more similarity among them. These two considerations led us to prefer for rural areas the Hierarchical and K-means cluster groupings over the unweighted index method.

In the previous section, it has been observed that size of clusters (in terms of number of NSS regions within them) is quite dissimilar in Hierarchical clusters. In Hierarchical method, there are two large clusters with 29 and 25 NSS regions within them respectively. All of the remaining 8 clusters have 34 NSS regions within them. Again, in the Hierarchical method, in the 1st cluster, 28 out of 29 NSS regions are geographically contiguous covering multiple states in Northern, Central and Western India. This large landmass includes agricultural prosperous regions, desert & dry land areas and large rainfed agricultural areas. They are very different in terms of level of living, food habits (vegetarianism), energy use and public provisioning of different state governments- and so reduces our confidence in this method.

In K-means cluster configuration, 5 out of 10 clusters contain at least 10 NSS regions within them and the size of different clusters are less divergent. The largest geographically contiguous zone in 6th cluster includes 15 NSS

regions compared to 28 in largest Hierarchical cluster configuration.

Under these considerations, we prefer the K-means cluster configuration as the base for identifying living income zones of rural India. However, before recommending use of these living income zones (LIZ), we will cross check and inform whether combination of NSS regions in single LIZ in the K-means method belong to the same cluster in either of the unweighted index or the Hierarchical methods, because, ideally, the composition of NSS regions in different clusters should be broadly similar from all three methods. It would increase the robustness of the results obtained from cluster analysis exercise. At least in case of recommended LIZs, the composition of NSS regions within each/ most of them should belong to the same cluster in one of the other two methods (Hierarchical cluster or unweighted index). It would further strengthen our case for recommended LIZs.

In most cases, the NSS regions in a particular LIZ are mostly geographically contiguous, but in a few cases they are located a short distance from one another and have similar food habits, energy use, etc. and so we decided to combine them.

6.2 RURAL LIVING INCOME ZONES

Below we discuss the distribution of rural Living Income zones across clusters and the composition of NSS regions within each rural Living Income Zone. We also highlight similarities with the Hierarchical clustering method or unweighted index method – and when similar clusters exist in both methods.

1. The 1st LIZ is state of Uttarakhand in the mountainous region of Northern India with single NSS region.

2. The 2nd LIZ includes 2 NSS regions of Bihar state in Eastern India and one region of Eastern Uttar Pradesh. These regions are geographically contiguous and are located in the Gangetic plain with similar food habits. Note that both regions of Bihar state belong to the same cluster in both unweighted index and Hierarchical methods.

3. The 3rd LIZ includes three regions of North-Eastern state of Assam. These are contiguous regions and they are part of the same cluster in Hierarchical method.

4. The 4th LIZ zone includes all two NSS regions of Jharkhand state of Eastern India and 2 NSS regions of Eastern state of Orissa. Jharkhand state and 2 regions of Orissa are located at short distances. These four regions also belong to the same cluster in Hierarchical method.

5. The 5th LIZ zone includes Inland Western region of Maharashtra and union territory of Daman & Diu located in Western India. These two regions are located at short distance and belong to the same cluster in unweighted index method.

6. The 6th LIZ zone and the last LIZ in the 1st cluster contains 3 regions from the mountainous state of Jammu & Kashmir in Northern India. Two of the three NSS regions (Outer Hills and Ladakh) are part of the same cluster in both unweighted index and Hierarchical method.

7. The 7th LIZ of the 2nd cluster includes all 5 NSS regions from the Eastern Indian states of West Bengal and 1 geographical contiguous NSS region of state of Orissa. All these 6 regions belong to the same cluster in Hierarchical method.

8. The 8th LIZ zone and the last one in the 2nd cluster includes 3 NSS regions of Karnataka state and one NSS region of state of Tamil Nadu. The region of Tamil Nadu is not geographically contiguous to other 2 regions of Karnataka as they

are separated by 1 NSS region of Tamil Nadu. Both these states are neighbouring states belonging to Southern India. Two regions of Karnataka belong to the same cluster in Hierarchical methodology. To this, Lakshadweep Island is added, the single NSS region of the 3rd cluster. This island and the Coastal & Ghat region of Karnataka state included in this LIZ is the part of the same cluster in Hierarchical method.

9. The 4th cluster has two LIZ zones, the 9th and 10th. The 9th LIZ covers all 2 NSS regions of the state of Kerala. They are also part of the same cluster in Hierarchical method. The 10th LIZ has single NSS region. It is the union territory of Andaman & Nichobar Island. It is not merged with other LIZ since in Hierarchical method also it forms a separate cluster.

10. The 5th cluster is divided into 3 LIZs. The 11th LIZ covers the whole of mountainous state Himachal Pradesh having 2 NSS regions. To this, we have added geographically contiguous Mountainous region of Jammu & Kashmir state because it belongs to the same cluster in Hierarchical method. The 12th LIZ covers whole of Central Indian hilly state of Chhattisgarh having 3 NSS regions. Two of the three regions fall in the same cluster in both unweighted index and Hierarchical method. The 13th LIZ includes 2 hilly/mountainous small states of North-Eastern India with both having single NSS region each. They are also part of the same cluster in Hierarchical method.

11. The 14th LIZ of 6th cluster includes 4 NSS regions and covers agriculturally prosperous and neighbouring states of Punjab and Haryana in Northern India. These 4 NSS regions also falls in the same cluster in Hierarchical method.

12. The 15th LIZ includes 4 NSS regions each of the Western Indian states of Gujarat and Rajasthan. They are geographically contiguous. 6 out 8 regions belonging to this LIZ are also part of the same cluster in Hierarchical method.

13. The 16th LIZ includes 3 contiguous NSS regions of Northern Indian state of Uttar Pradesh. It is the last LIZ of 6th cluster. 2 out of 3 NSS regions belonging to this LIZ are part of the same cluster in Hierarchical method.

14. The 7th cluster includes Delhi and union territory of Dadra & Nagar Haveli. Delhi is almost fully urbanised with negligible section working in a rural occupation like agriculture. Dadra & Nagar Haveli also has very small rural population of around 15 thousand in 2021. Therefore, these two NSS regions are not included in any rural LIZ.

15. The 17th LIZ of 8th cluster includes 4 NSS regions with 2 NSS regions of Madhya Pradesh (South-Western and Northern) and 1 NSS region each from Rajasthan (South-Eastern) and Uttar Pradesh (Southern). These NSS regions are contiguous and fall in the same cluster in Hierarchical method.

16. The 18th LIZ contains the rest of the 4 regions of Central Indian state of Madhya Pradesh. They also belong to the same cluster in Hierarchical method.

17. The 19th LIZ consists of 1 NSS region each from Western Indian state of Gujarat and Southern India state of Karnataka and 3 NSS region of Western Indian state of Maharashtra. These are geographically contiguous. 4 out of 5 NSS regions are part of the same cluster in Hierarchical method.

18. The 9th cluster is divided into 3 LIZs. The 20th LIZ covers 4 NSS regions

belonging to 4 states in North-Eastern part of India. These are all hilly areas and they are part of the same cluster in Hierarchical method. These regions are geographically contiguous. The 21st LIZ include 3 NSS regions of North-Eastern India and these are mostly plain areas of states of Manipur, Assam and Tripura. Two of the three regions belong to the same cluster in both Hierarchical and unweighted index methods. The 22nd LIZ includes two neighbouring NSS regions of Goa and Maharashtra states. There are 2 other NSS regions in this cluster. Pondicherry as well as Chandigarh are basically urban settlement. We have excluded these two NSS regions, because they are almost exclusively urban. Daman & Diu is already included in 5th LIZ.

19. The 10th cluster contains two LIZ. The 23rd LIZ includes all 5 regions fully covering two South Indian states of Andhra Pradesh and Telangana. It also includes one NSS region of Maharashtra state. These are contiguous regions. In Hierarchical method, all 6 NSS regions are part of the same cluster. The last 24th LIZ include 3 NSS regions of Southern India state of Tamil Nadu. These regions are part of the same cluster in Hierarchical method.

Table 6: Recommended Composition of Living Income Zones (24) in Rural India

Cluster	Living Income Zone	State	NSS Regions
1st Cluster	1	Uttarakhand	All (1)
	2	Bihar	Northern Plain, Southern Plain (1)
		Uttar Pradesh	Eastern (1)
	3	Assam	Plain Western, Cachar Plain, Central Brahmaputra Plain (1)
	4	Jharkhand	Ranchi Plateau, Hazaribagh Plateau (1)
		Orissa	Coastal, Southern (1)
	5	Maharashtra	Inland Western (1)
		Daman & Diu	All (9)
	6	Jammu & Kashmir	Outer Hills, Ladakh, Jhelum Valley (2)
2nd Cluster	7	Orissa	Northern (2)
		West Bengal	Eastern Plain, Southern Plain, Central Plain, Himalayan, Western Plain (2)
	8	Karnataka	Coastal & Ghats, Inland Eastern, Inland Southern (2)
		Lakshadweep	All (3)
		Tamil Nadu	Southern (2)

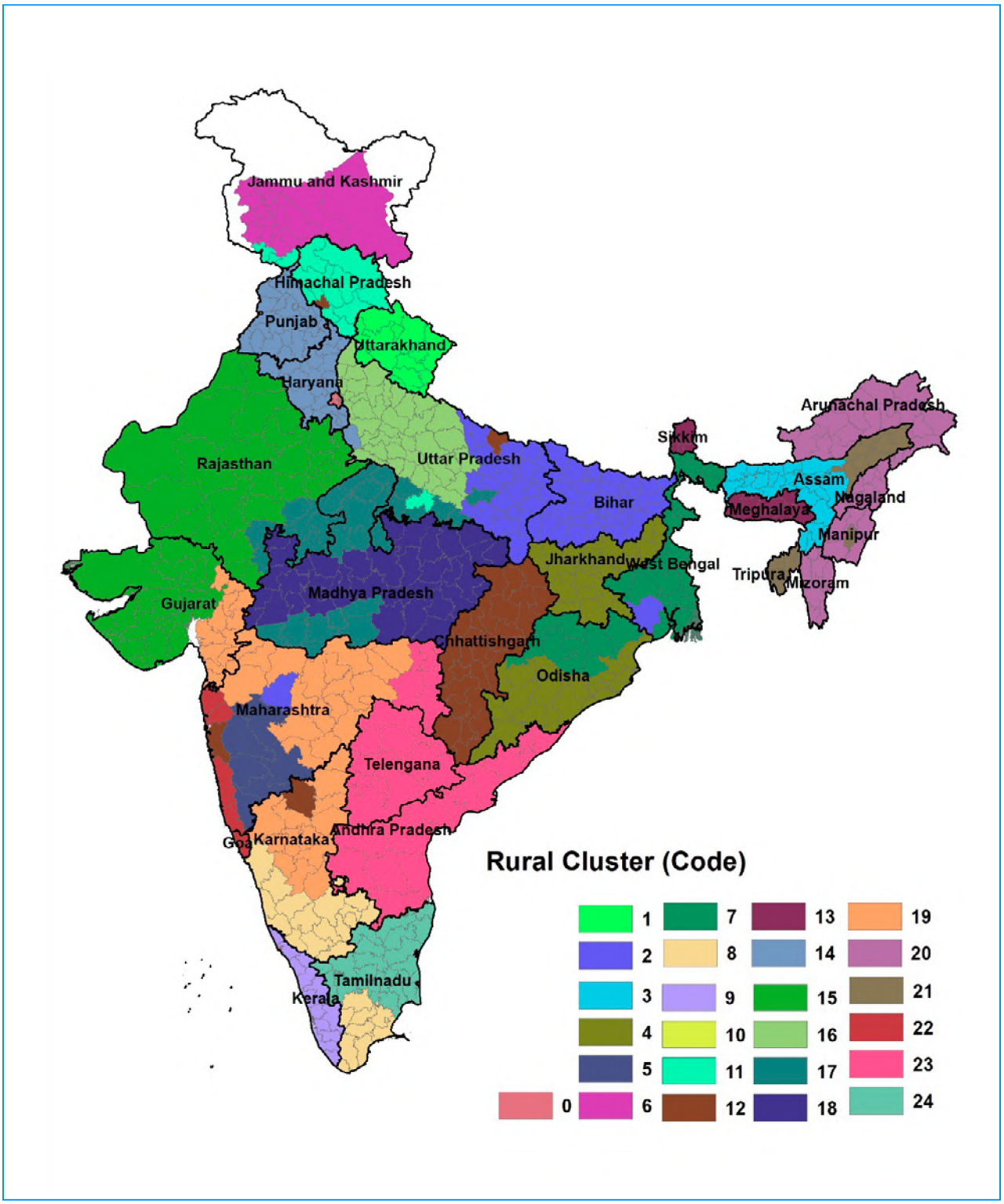
Cluster	Living Income Zone	State	NSS Regions
4th Cluster	9	Kerala	Northern, Southern (4)
	10	Andaman & Nichobar	All (4)
5th Cluster	11	Himachal Pradesh	Central Plain, Trans Himalayan and Southern (5)
		Jammu & Kashmir	Mountainous (6)
	12	Chhattisgarh	Mahanadi Basin, Southern Chhattisgarh, Northern Chhattisgarh (5)
	13	Sikkim	All (5)
		Meghalaya	All (5)
6th Cluster	14	Punjab	Northern, Southern (6)
	14	Haryana	Eastern, Western (6)
	15	Rajasthan	Western, North-Eastern, Southern, Northern (6)
	15	Gujarat	Plain Northern, Dry Areas, Kachchh, Saurastra (6)
	16	Uttar Pradesh	North Upper Ganga Plain, Central, South Upper Ganga Plain (6)
8th Cluster	17	Uttar Pradesh	Southern (8)
		Madhya Pradesh	South-Western, Northern (8)
		Rajasthan	South-Eastern (8)
	18	Madhya Pradesh	Vindya, Central, Malwa, South,
	19	Gujarat	South Eastern (8)
	19	Maharashtra	Inland Northern, Inland Central Inland Eastern (8)
	19	Karnataka	Inland Northern (8)
9th Cluster	20	Arunachal Pradesh	All (9)
		Nagaland	All (9)
		Manipur	Hills (9)
		Mizoram	All (9)

Cluster	Living Income Zone	State	NSS Regions
	21	Manipur	Plain (9)
		Assam	Plain Eastern (9)
		Tripura	All (9)
	22	Goa	All (9)
		Maharashtra	Coastal (9)
10th Cluster	23	Andhra Pradesh	Coastal Northern, Coastal Southern, Inland Southern (10)
		Telangana	Inland North-Western, Inland North-Eastern (10)
		Maharashtra	Eastern (10)
	24	Tamil Nadu	Coastal Northern, Coastal, Inland (10)

Note: Four NSS regions are excluded from rural LIZs. These are mostly or fully urban NSS regions where hardly any rural jobs exist. These include Capital Delhi and 3 union territories of Chandigarh, Dadra & Nagar Haveli and Pondicherry.
The 3rd and 7th clusters of K-means contain single NSS regions each - Lakshadweep and Daman & Diu respectively. These two regions have very small rural population (less than 100 thousand). These are merged with 8th and 5th LIZ respectively that are geographically close to them.

Map 3 provides a graphical presentation of our rural living income zones of India.

Map 3: Rural Living Income Zones (LIZs) of India



Note 1: Composition of the districts in NSS regions included in each rural living income zone (LIZ) is given in Table 10.
Note 2: '0' refers to four NSS regions that are excluded from the rural LIZ analysis. These are mostly or fully urban NSS regions where hardly any rural jobs exist. These include Capital Delhi and 3 union territories of Chandigarh, Dadra & Nagar Haveli and Pondicherry.

7. IDENTIFYING LIVING INCOME ZONES FOR URBAN INDIA

7.1 HOW WE COMPARED RESULTS FROM CLUSTER METHODOLOGIES FOR URBAN AREAS – TO DECIDE ON WHICH CLUSTER METHOD TO USE FOR URBAN INDIA

We analysed the results of the three clustering methodologies for urban regions as we did for rural regions, with much less emphasis on the need to have contiguous regions because urban areas are as not geographically contiguous as in rural India.

Urban India is composed of towns and cities of varying sizes with preponderance of small towns (< 50,000 population). The total geographical area under urban settlement is much smaller but with substantially higher levels of population density. Cost of living differences across urban areas depends greatly on the size of urban settlement especially in terms of differences in housing and transportation cost. In small towns with less than 50,000 population, there is often a lack of public transportation and individual transportation is used by people.

7.2 RESULTS OF THREE CLUSTER ANALYSIS METHODS FOR URBAN INDIA

One of the criteria we use to select the best cluster analysis method to use for creating Living Income Zones in urban India is the distribution of NSS regions among clusters. The differences in the size of clusters (in terms of number of NSS regions within them) is much higher in Hierarchical method as the three largest zones with 29, 27 and 21 NSS regions cover 77 of the 88 NSS regions. The remaining 11 NSS regions are spread over 7 different clusters. In K-means method, there are 5 zones with more than 10 NSS regions each. So, K-means has a much more even distribution of NSS zones among clusters.

In addition, K-means analysis distinctly puts less developed urban areas in specific clusters and more developed urban areas with different characteristics in other clusters more effectively compared to unweighted index method and Hierarchical method.

Under these considerations, we prefer the K-means cluster configuration as the basis for identifying living income zones. However, before we finalize our urban living income zones (LIZ), we cross check and inform regarding whether combinations of NSS regions in a single LIZ belong to the same cluster in either the unweighted index method and/or the Hierarchical method. As explained in the previous section, ideally the composition of NSS regions in different clusters obtained from all three of these methods should be broadly similar. This would increase our confidence in the robustness of the results obtained from cluster analysis exercise and in particular results from the K-means method that we use as the main basis for identifying urban living income and living wage zones. At least in case of recommended LIZs, the composition of NSS regions within each/most clusters should belong to the same cluster in at least one of the other two methods. This would, methodologically, further strengthen our case for recommended LIZs.

7.3 CREATION OF URBAN LIVING INCOME ZONES WITHIN CLUSTERS

We created Living Income Zones within clusters partly based on their proximity to one another.

For most clusters, the LIZs are selected in such a way that they consist of neighbouring NSS

regions, although in few cases NSS regions within a LIZ are located a short distance from one another, with some other towns located between them.

In addition, we made sure that all NSS regions in each LIZ are part of the same cluster according to two methods (the unweighted cluster method, Hierarchical clustering method, or K-means method) This meant that a few NSS regions were shifted from one cluster to another cluster.

In most LIZs, various size class of towns exist and the cost of living is expected to differ among them. However, the range of cost of living estimated in the large and smaller cities is likely to be within a small acceptable range of around10 percent. In larger cities, the cost of housing is likely to be higher but in small cities lack of public facilities including transportation is likely to make transportation and other cost relatively higher.

Table 7 shows the recommended distribution of Living Income Zones across different clusters and the composition of NSS regions within each Living Income Zone in urban areas.

The 1st Living Income Zone (LIZ) includes 5 NSS regions covering whole states of Andhra Pradesh and Telangana in Southern India. These are contiguous regions and these regions also belong to the same cluster in both unweighted and hierarchical methods. In this LIZ, Hyderabad city can represent it.

The 2nd LIZ also covers 5 NSS regions from three Southern Indian states of Karnataka, Kerala and Tamil Nadu. 3 NSS regions of Karnataka and Kerala are contiguous zones but 2 NSS regions of Tamil Nadu are located short distance away. All these are part of the same cluster in Hierarchical method. To this, we have added 3 NSS regions – 1 each from Kerala and Tamil Nadu and small state of Goa. That makes 8 NSS regions in this zone and it becomes contiguous zone. Still all 8 NSS regions are part of the same cluster in Hierarchical method.

The 3rd LIZ, the last one from 1st cluster, includes 1 NSS region from Eastern Indian state of Orissa and 4 NSS regions from another Eastern Indian state of West Bengal. These are contiguous zones. They are part of same cluster in Hierarchical method. In this LIZ, large city of Kolkata is located and it can represent the whole LIZ.

The 4th LIZ, the first one from 2nd cluster includes 8 NSS regions containing 1 each of Northern Indian states of Punjab and Haryana, and 4 and 2 NSS regions from two Western Indian states of Rajasthan and Gujarat respectively. These are contiguous regions and 7 NSS regions are part of the same cluster in Hierarchical method.

The 5th LIZ, a contiguous zone, contains 3 NSS regions of Northern Indian state of Uttar Pradesh. It includes two large cities of Kanpur and state capital of Lucknow, any of them could represent this zone. It is part of the same cluster in Hierarchical method.

The 6th LIZ includes 5 out of 6 NSS regions of the Central Indian state of Madhya Pradesh. We have added Central region of Madhya Pradesh in this LIZ that is part of 6th cluster to make a LIZ for the whole state of Madhya Pradesh. These are contiguous NSS regions and are part of the same cluster in Hierarchical method.

The 7th LIZ belonging to 3rd cluster contains altogether 3 NSS regions from Southern Indian states of Karnataka, Tamil Nadu and union territory of Puducherry. It includes the two largest cities of Southern India – Bangalore and Chennai. Both these cities can represent this zone. It is contiguous zone and they are all part of the same cluster in Hierarchical method.

The 8th LIZ has 1 NSS region of Andaman & Nichobar Island and it is kept separate as it forms an independent cluster in Hierarchical method.

The 9th LIZ is the only LIZ in 4th cluster and it includes the 2 NSS regions of Daman & Diu and Dadra & Nagar Haveli which now are part

of the same union territory. They also form an independent cluster in Hierarchical method and part of the same cluster in unweighted index method. The capital city of Daman can represent this LIZ.

The 10th LIZ in the 5th cluster has 1 NSS region of Lakshadweep Island. It also forms an independent cluster in Hierarchical method. It is a small island having total population of only 65 thousand with capital city Kavaratti having 11 thousand population. It can be excluded from a future survey because of its size. Another NSS region, that was originally part of this cluster, comprises whole state of Mizoram has been shifted to 7th cluster.

The 11th LIZ in the 6th cluster includes 1 NSS region each from Jammu & Kashmir and Punjab and 2 NSS regions covering whole of Himachal Pradesh. All these states are from Northern India and these are contiguous regions. 3 of the 4 NSS regions are from the same cluster in unweighted mean method. Another NSS region of Haryana that was originally part of this cluster (includes city of Gurgaon) has been shifted in another LIZ to form a separate LIZ with capital Delhi.

The 12th LIZ contains 3 NSS regions of Gujarat and 1 NSS region of Rajasthan. These are contiguous regions. 3 out of these 4 regions are part of the same cluster in Hierarchical method.

The 13th LIZ in 7th cluster has multiple states from North-Eastern India. It includes 3 states having single NSS region each and 1 NSS region from Manipur state and 3 NSS regions of Assam. All these 7 NSS regions are part of the same cluster in Hierarchical method. The metropolitan city of Guwahati can represent this LIZ.

The 14th LIZ includes 4 NSS regions with 1 region each from Jharkhand and Chhattisgarh and 2 NSS regions from Orissa. These are contiguous regions and these are part of the same cluster in Hierarchical method. In the 8th cluster, there are 4 LIZs.

The 15th LIZ has single NSS region from Jammu & Kashmir (Ladakh) that is located in the border region of both Pakistan and China. Leh town can represent this region.

The 16th LIZ includes state of Sikkim with single NSS region and 1 NSS region from West Bengal. Both part of Eastern India and they are contiguous regions. The region from West Bengal has been shifted from 1st cluster and both of them are part of the same cluster in Hierarchical method. Most of the area in this LIZ is mountainous.

The 17th LIZ includes 1 NSS region each from four North-Eastern Indian states of Manipur, Meghalaya, Mizoram and Assam. They do not form a geographically contiguous region but are not located far away. These 4 regions are part of the same cluster in Hierarchical method.

The 18th LIZ has 3 NSS regions from Central India from two states of Chhattisgarh and Maharashtra. The NSS region of Maharashtra has been shifted from 7th cluster. All these regions are contiguous and they are part of the same cluster in Hierarchical method.

The 19th LIZ cluster originally has 3 NSS regions but 2 more NSS regions have been added from other clusters in 20th and 21st LIZs. The 19th LIZ has one urban city of Chandigarh which is joint capital of two states of Punjab and Haryana belonging to Northern India.

In 20th LIZ with Delhi, we have added Eastern region of Haryana so that it covers larger part of industrial areas of Delhi capital region. They are contiguous and they form part of the same cluster in hierarchical method. The capital city of Delhi can represent 20th LIZ.

In 21st LIZ belonging to 9th cluster, one more NSS region from 8th cluster from the same state of Maharashtra is added to completely cover Mumbai-Pune industrial corridor. These two regions are part of the same cluster in Hierarchical method. The city of Mumbai can represent this LIZ.

The 22nd LIZ of 10th cluster has 2 NSS regions from Jammu & Kashmir state. They are part of the same cluster in Hierarchical method.

The 23rd LIZ also has 2 NSS regions but from two Northern Indian states of Uttarakhand and Uttar Pradesh.

The 24th LIZ has 5 NSS regions from four different states of Northern and Eastern India. These NSS regions are contiguous and are part of the same cluster in Hierarchical method.

The 25th LIZ has 5 NSS regions with 3 NSS regions from Western state of Maharashtra and 2 NSS regions of Southern state of Karnataka. These are contiguous regions. 4 of these 5 regions are part of the same cluster in Hierarchical method.

Table 7: Composition of Recommended Living Income Zones (25) for Urban India

Cluster	Living Income Zone	State	NSS Regions
1st Cluster, unweighted value= 0.319	1	Andhra Pradesh	Coastal Northern, Coastal Southern, Inland Southern (1)
		Telangana	Inland North-Western, Inland North-Eastern (1)
	2	Karnataka	Coastal & Ghats (1)
		Kerala	Northern, Southern (1)
		Goa	All (1)
		Tamil Nadu	Coastal, Southern, Inland ((1)
	3	Orissa	Northern (1)
		West Bengal	Eastern Plain, Southern Plain, Central Plain (1)
2nd Cluster, unweighted value= 0.327	4	Punjab	Southern (2)
		Haryana	Western (2)
		Rajasthan	Western, North-Eastern, South-Eastern, Northern (2)
		Gujarat	Dry Areas, Saurastra (2)
	5	Uttar Pradesh	Central, Southern, South Upper Ganga Plain (2)
	6	Madhya Pradesh	Vindya, Malwa, South, South-Western, Northern, Central (2)
3rd Cluster, unweighted value= 0.458	7	Karnataka	Inland Southern (3)
		Tamil Nadu	Coastal Northern (3)
		Puducherry	All (3)
	8	Andaman & Nichobar Island	All (3)

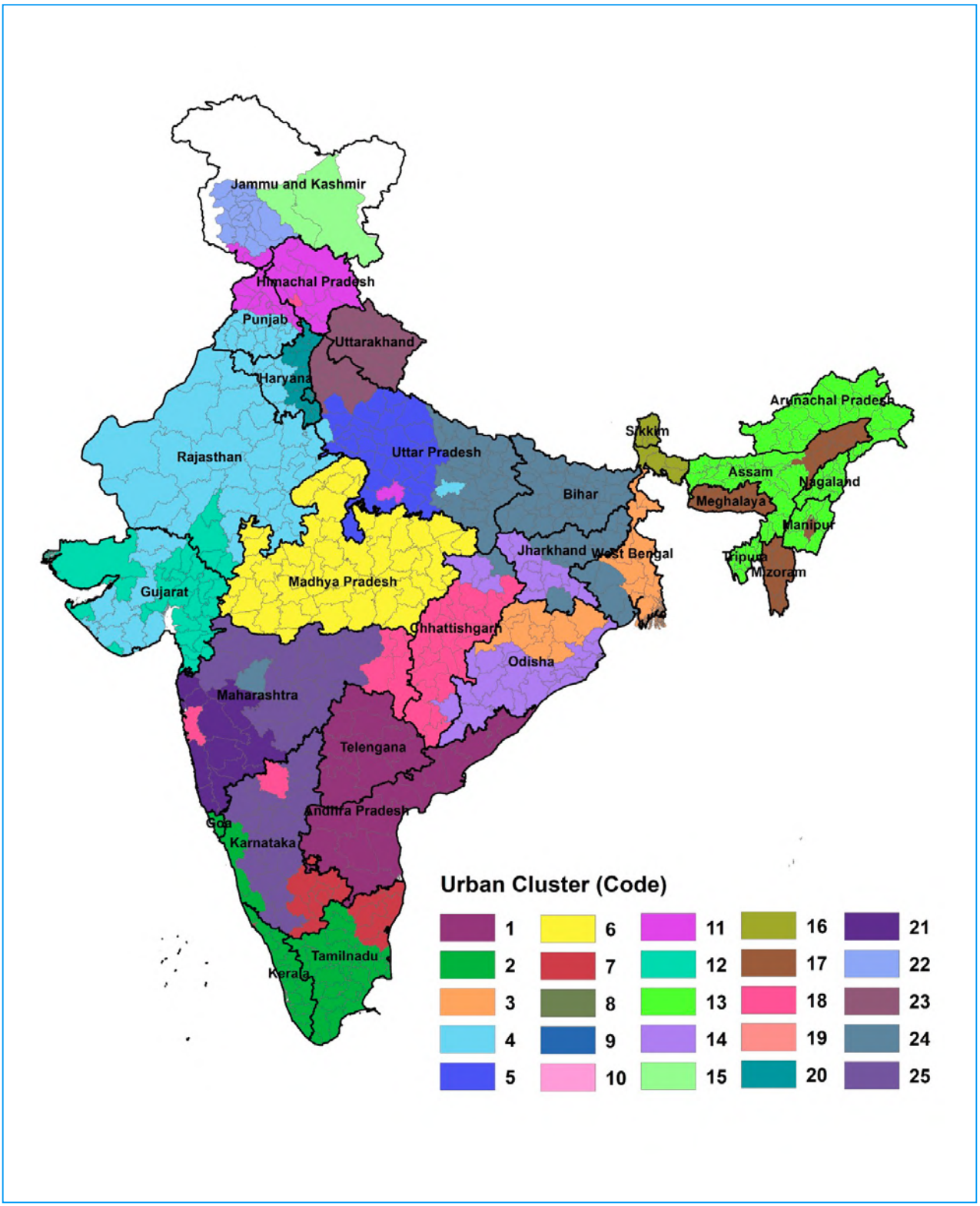
Cluster	Living Income Zone	State	NSS Regions
4th Cluster, unweighted value= 0.585	9	Daman & Diu	All (4)
		Dadra & Nagar Haveli	All (4)
5th Cluster, unweighted value= 0.401	10	Lakshadweep	All (5)
6th Cluster, unweighted value= 0.436	11	Jammu & Kashmir	Mountainous (6)
		Himachal Pradesh	Central Plain, Trans Himalayan & Southern (6)
		Punjab	Northern (6)
	12	Rajasthan	Southern (6)
		Gujarat	South-Eastern, Plain Northern, Kachchh (6)
7th Cluster, unweighted value= 0.302	13	Arunachal Pradesh	All (7)
		Nagaland	All (7)
		Manipur	Hills (7)
		Assam	Plains Western, Cachar Plain, Brahmaputra Plain (7)
		Tripura	All (7)
	14	Orissa	Coastal, Southern (7)
		Jharkhand	Ranchi Plateau (7)
		Chhattisgarh	Northern (7)
8th Cluster, unweighted value= 0.348	15	Jammu & Kashmir	Ladak (8)
	16	West Bengal	Himalayan
		Sikkim	All (8)
	17	Manipur	Plains (8)
		Meghalaya	All (8)
		Mizoram	All (5)
		Assam	Plains Eastern (8)
	18	Chhattisgarh	Mahanadi Basin, Southern (8)
		Maharashtra	Eastern (7)

Cluster	Living Income Zone	State	NSS Regions
9th Cluster, unweighted value= 0.531	19	Chandigarh	All (9)
	20	Delhi	All (9)
		Haryana	Eastern (6)
	21	Maharashtra	Coastal (9)
		Maharashtra	Inland Western (8)
10th Cluster, unweighted value= 0.264	22	Jammu & Kashmir	Outer Hills, Jhelam Valley (10)
	23	Uttarakhand	All (10)
		Uttar Pradesh	North Upper Ganga Plain (10)
	24	Uttar Pradesh	Eastern (10)
		Bihar	Northern Plain, Central Plain (10)
		Jharkhand	Hazaribagh Plateau (10)
		West Bengal	Western Plain (10)
	25	Maharashtra	Inland Northern, Inland Central, Inland Eastern (10)
		Karnataka	Inland Eastern, Inland Northern (10)

Note: In 2nd LIZ, 3 NSS regions – 1 each from states of Kerala and Tamil Nadu and small state of Goa is added. All these three regions originally belong to 3rd cluster and they are part of Western Coastal areas and they substantially rely on tourism. Hilly North-Eastern state of Mizoram is shifted to 17th LIZ that includes other North-Eastern states. One region of Haryana is shifted to 20th LIZ to combine with capital Delhi as it covers large part of industrial areas of capital Delhi. In 21st LIZ, one NSS region from 8th cluster from the same state of Maharashtra is added to completely cover Mumbai-Pune industrial corridor.

Map 4 provides a graphical presentation of our urban living income zones of India.

Map 4: Urban Living Income Zones (LIZs) of India



Note: Composition of districts in NSS regions in each urban living income zone (LIZ) is given in Table 11.

8. SUMMARY AND IMPLICATION OF THE STUDY

In this section, a summary of the study is provided and this is followed by implications from the findings of this study.

8.1 SUMMARY

India is a large and diversified country with 1.34 billion population. Large differences exist in terms of level of living, food habits, gender equity, labour market conditions, energy use, and public provisioning of education and health facilities. This means that there are large differences across India in living incomes and living wages. At the same time, these factors are more similar in neighbouring NSS regions of the same state or neighbouring states. Thus, geographical contiguity or geographical proximity becomes imperative to identifying living income zones.

We have used in this report three alternative aggregating methodologies to help identify living income zones for India. This is done separately for rural India and urban India starting from 88 NSS urban and rural regions.

1. Hierarchical Cluster Analysis
2. K-Means Cluster Analysis
3. Composite Index

Of the above three methods, unweighted composite index is a supervised method based on a composite index. The other two methods (Hierarchical and K-means) are unsupervised methods. In unsupervised methods after providing the standardised values to the models, these methods on their own identify which NSS regions to include in which living income clusters. In that sense, unsupervised methods are generally better, because they are free from assumptions made to construct a composite index.

For all of these three methods, we included 8 indicators for both rural and urban areas covering economic, demographic, social and cultural factors. These indicators were chosen from a list of 50 odd indicators at NSS regional level separately for rural and urban areas.

The unweighted composite index method indicated 53 and 52 geographically contiguous living income zones for rural and urban areas respectively from the original 88 NSS regions for each. The Hierarchical and K-means methods indicated a maximum of 33 geographically contiguous zone for both rural and urban India.

We concluded that the unweighted index method did not provide a good basis for identifying living income zones for either rural or urban areas. In this method, the number of geographical contiguous zones was far larger compared to other two methods. Geographical contiguity is an important consideration as similarity in terms of level of living, food habits, energy use and public provisioning of education and health facilities are more homogeneous in neighbouring NSS regions of the same state or neighbouring states in a highly populous and diversified country such as India. From this perspective, aggregating 88 NSS regions into over 50 contiguous zones makes this method inefficient.

We also concluded that K-means method provides better results for both rural and urban areas compared to Hierarchical method among the two unsupervised methods.

For rural areas, Hierarchical method has two very large clusters which have 29 and 25 NSS regions within them. All of the remaining eight Hierarchical clusters have less than 10 NSS regions within them. In K-means method, the sizes of the clusters are much more balanced with 5 out of 10 clusters containing at least 10

NSS regions within them and so the size of different clusters are less divergent. The largest geographically contiguous zone in K-means method includes 15 NSS regions compared to 28 NSS regions in Hierarchical cluster method. Given these results, we consider that the K-means cluster method provides a much better basis for identifying living income zones for rural India.

Similar results regarding methods are found for urban areas in terms of the distribution of Living Income (LI) clusters. First, the Hierarchical method indicates 3 very large LI clusters covering most NSS regions (77 out of the 88 NSS regions). In contrast, the largest K-means method LI cluster includes 16 NSS regions. Second, 4 clusters in Hierarchical method included only one NSS region compared to none in K-means method. Thus, also for urban areas, K-means cluster method provides the best basis for identifying living income zones.

After identifying living income zones (LIZ) for rural and urban areas from K-means methodology, we checked whether the combinations of NSS regions in each LIZ from the K-means method belong to the same cluster in either the unweighted/composite index method or the Hierarchical method. We feel that when this happens that it helps to provide further confidence in LIZ results from the K-means method. In most of the LIZs in both rural and urban areas, this has been found to be true. It is worth noting that the NSS regions in identified LIZs are mostly contiguous areas geographically, which also adds confidence to results.

In the end, we identified 24 rural living income zones, and 25 urban living income zones for India in this report (details given in the analysis section). Maps 3 and 4 indicate results graphically. However, when few isolated living income zones in distant areas that cover a single NSS region are excluded (Andaman & Nichobar Island in both rural and urban areas and Lakshadweep Island in the urban areas), the number of living income zones where independent follow up studies should be conducted, can be reduced by some margin.

8.2 IMPLICATIONS OF THIS STUDY

There are several implications of this study:

- First, living income and living expenses, of course, vary from one rural area to another rural area and also from one city to another city in the same living income zone. But living income study estimates conducted in multiple areas in a single LIZ are likely to be similar, or to lie within a narrow acceptable band, say plus or minus 10 per cent.

- Second, this study identified 24 rural and 25 urban living income zones with similar economic, social, demographic and cultural factors throughout each zone. Once these 49 living income studies are conducted (including both rural and urban areas), it would be possible to estimate all India level of living income by using total workforce as weights for each LIZ.

- Third, this study also provides some direction on whether living income study results for one LIZ can be considered as a benchmark value for areas belonging to other LIZs. It can be used as benchmark, provided they belong to the same cluster (out of 10 clusters) either in K-means or in unweighted index method.

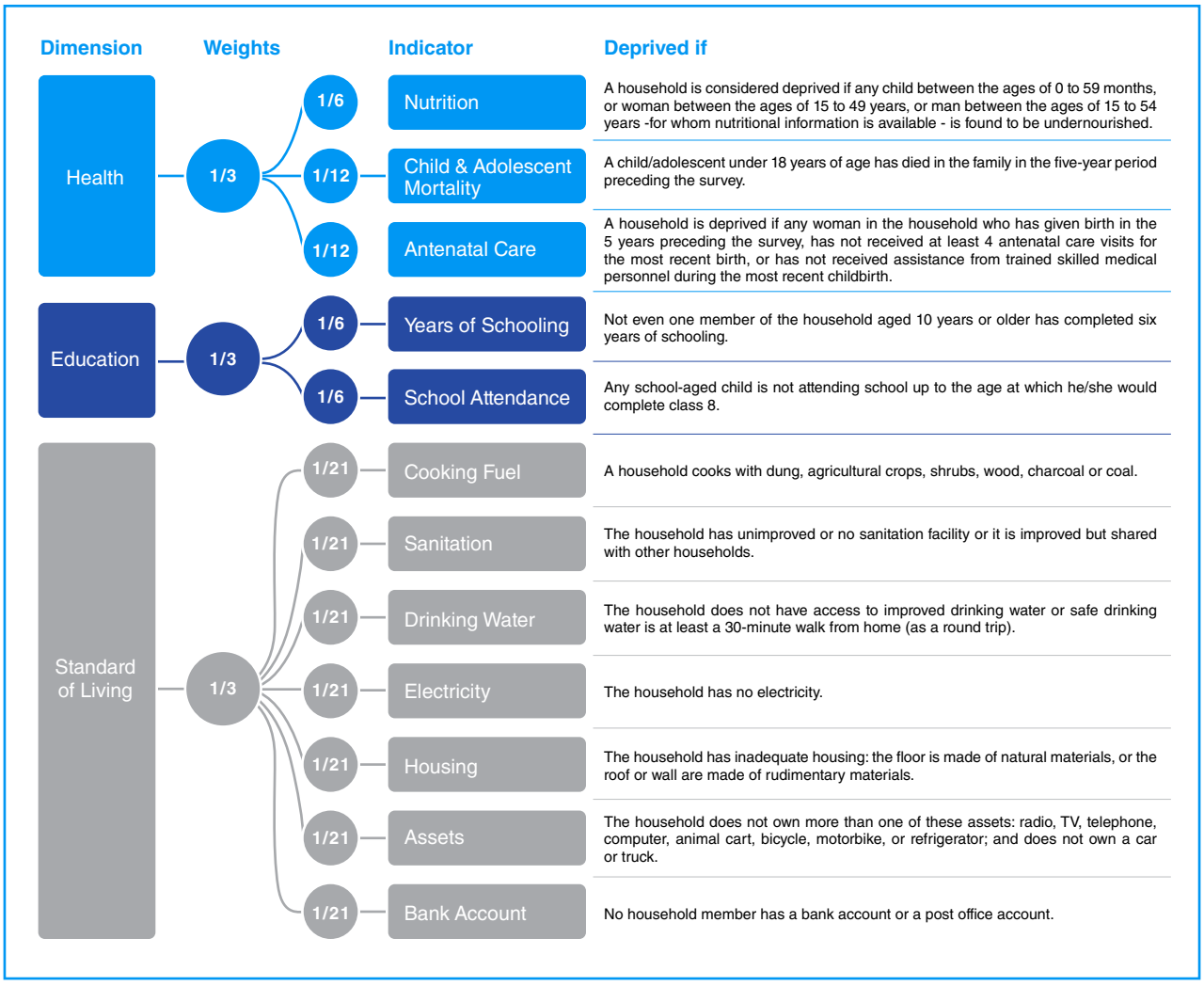
- Lastly, since LIZs have been identified on the basis of economic, social, demographic and cultural factors and not directly adopting Anker Methodology, these LIZs can be used as the basis for calculating representative living incomes not only for the Anker Methodology but also for similar methods used by other organizations (like Indian Government, ILO, etc,).

ANNEXES

ANNEX I: MULTIDIMENSIONAL POVERTY INDEX (MPI)

The Government of India think-tank NITI Aayog released Multidimensional Poverty Index (MPI) in 2021. This is India's first ever national MPI measure. It is based on the National Family Health Survey (NFHS)-4, 2015-16. It uses the globally accepted robust methodology developed by the Oxford Poverty and Human Initiative (OPHI) and the United Nations Development Programme (UNDP). It captures multiple and simultaneous deprivations faced by households across the country. The MPI is calculated by using 12 indicators including nutrition, child and adolescent mortality, antenatal care, years of schooling, school attendance, cooking fuel, sanitation, drinking water, electricity, housing assets, and bank account. These indicators have been grouped into three dimensions namely: health, education and standard of living. The National MPI dimensions, indicators and weights are given below.

Annex I Figure 1: Multi-dimensional Poverty Index (MPI) indicators flow chart



This NITI Aayog report presents an in-depth analysis of the headcount rate and intensity of multidimensional poverty at the national, state and district levels. This shift in focus from income or consumption expenditure as the basis of poverty estimation is founded on the policy narrative at national level that human and capability development along with access to basic infrastructure is at the centre of India's development policy. The MPI shows that Bihar (51.9%) and Jharkhand (42.2%) registered the highest poverty rate among all the states, while Kerala (1%), and Goa (4%) have the lowest poverty rate. The findings of MPI serve as a useful source for measuring the situation at baseline i.e., before the large-scale rollout of government important welfare schemes. Further, the baseline report can be updated upon the release of the NFHS-5 (2019-20) dataset to measure the impact of government flagship programmes and schemes.

ANNEX II: LIST OF NSS REGIONS AND THEIR COMPOSITION OF DISTRICTS

In this annex, the detailed composition of each NSS regions in terms of their constituent districts are presented. It helps in identifying villages or cities where survey needs to be conducted in each living income zone (LIZ). In the identified LIZ, only constituent NSS regions have been presented in the main text. From the detailed district lists of these NSS regions in a specific LIZ, decision can be made in which village or city future living wage and living income studies need to be conducted.

Annex II Table 1: Districts included in NSS regions

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	Andaman & Nicobar Islands (35)	351	Andaman & Nicobar Islands	1.	Nicobars	(01)
				2.	North & Middle Andaman	(02)
				3.	South Andaman	(03)
2.	Andhra Pradesh (28)	281	Coastal Northern	4.	Srikakulam	(01)
				5.	Vizianagaram	(02)
				6.	Visakhapatnam	(03)
				7.	East Godavari	(04)
				8.	West Godavari	(05)
3.		282	Coastal Southern	9.	Krishna	(06)
				10.	Guntur	(07)
				11.	Prakasam	(08)
				12.	Sri Potti Sriramulu Nellore	(09)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
4.		283	Inland Southern	13.	Y.S.R. (Cuddapah)	(10)
				14.	Kurnool	(11)
				15.	Anantapur	(12)
				16.	Chittoor	(13)
5.	Arunachal Pradesh (12)	121	Arunachal Pradesh	17.	Tawang	(01)
				18.	West Kameng	(02)
				19.	East Kameng	(03)
				20.	Papum Pare	(04)
				21.	Upper Subansiri	(05)
				22.	West Siang	(06)
				23.	East Siang	(07)
				24.	Upper Siang	(08)
				25.	Changlang	(09)
				26.	Tirap	(10)
				27.	Lower Subansiri	(11)
				28.	Kurung Kumey	(12)
				29.	Dibang Valley	(13)
				31.	Lohit	(15)
				32.	Anjaw	(16)
6.	Assam (18)	181	Plains Eastern	33.	Lakhimpur	(08)
				34.	Dhemaji	(09)
				35.	Tinsukia	(10)
				36.	Dibrugarh	(11)
				37.	Sivasagar	(12)
				38.	Jorhat	(13)
				39.	Golaghat	(14)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
7.		182	Plains Western	40.	Kokrajhar	(01)
				41.	Dhubri	(02)
				42.	Goalpara	(03)
				43.	Barpeta	(04)
				44.	Bongaigaon	(20)
				45.	Chirang	(21)
				46.	Kamrup	(22)
				47.	Kamrup Metropolitan	(23)
				48.	Nalbari	(24)
				49.	Baksa	(25)
8.		183	Cachar Plain	50.	Karbi Anglong	(15)
				51.	Dima Hasao	(16)
				52.	Cachar	(17)
				53.	Karimganj	(18)
				54.	Hailakandi	(19)
9.		184	Central Brahmaputra Plains	55.	Morigaon	(05)
				56.	Nagaon	(06)
				57.	Sonitpur	(07)
				58.	Darrang	(26)
				59.	Udalguri	(27)
10.	Bihar (10)	101	Northern	60.	Pashchim Champaran	(01)
				61.	Purba Champaran	(02)
				62.	Sheohar	(03)
				63.	Sitamarhi	(04)
				64.	Madhubani	(05)
				65.	Supaul	(06)
				66.	Araria	(07)
				67.	Kishanganj	(08)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
				68.	Purnia	(09)
				69.	Katihar	(10)
				70.	Madhepura	(11)
				71.	Saharsa	(12)
				72.	Darbhanga	(13)
				73.	Muzaffarpur	(14)
				74.	Gopalganj	(15)
				75.	Siwan	(16)
				76.	Saran	(17)
				77.	Vaishali	(18)
				78.	Samastipur	(19)
				79.	Begusarai	(20)
				80.	Khagaria	(21)
				81.	Bhagalpur	(22)
				82.	Banka	(23)
				83.	Munger	(24)
				84.	Lakhisarai	(25)
				85.	Sheikhpura	(26)
				86.	Nalanda	(27)
				87.	Patna	(28)
11.		102	Central	88.	Bhojpur	(29)
				89.	Buxar	(30)
				90.	Kaimur (Bhabua)	(31)
				91.	Rohtas	(32)
				92.	Aurangabad	(33)
				93.	Gaya	(34)
				94.	Nawada	(35)
				95.	Jamui	(36)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
				96.	Jehanabad	(37)
				97.	Arwal	(38)
12.	Chandigarh (04)	041	Chandigarh	98.	Chandigarh	(01)
13.	Chhattisgarh (22)	221	Northern Chhattisgarh	99.	Koriya	(01)
				100.	Surguja	(02)
				101.	Surajpur	(26)
				102.	Balrampur	(27)
14.		222	Mahanadi Basin	103.	Jashpur	(03)
				104.	Raigarh	(04)
				105.	Korba	(05)
				106.	Janjgir-Champa	(06)
				107.	Bilaspur	(07)
				108.	Kabeerdham	(08)
				109.	Rajnandgaon	(09)
				110.	Durg	(10)
				111.	Raipur	(11)
				112.	Mahasamund	(12)
				113.	Dhamtari	(13)
				114.	Balodabazar	(19)
				115.	Gariyaband	(20)
				116.	Bemetara	(23)
				117.	Balod	(24)
				118.	Mungeli	(25)
15.		223	Southern Chhattisgarh	119.	Uttar Bastar Kanker	(14)
				120.	Bastar	(15)
				121.	Narayanpur	(16)
				122.	Dakshin Bastar Dantewada	(17)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
				123.	Bijapur	(18)
				124.	Kondagaon	(21)
				125.	Sukama	(22)
16.	Dadra& Nagar Haveli (26)	261	Dadra & Nagar Haveli	126.	Dadra & Nagar Haveli	(01)
17.	Daman & Diu (25)	251	Daman & Diu	127.	Diu	(01)
				128.	Daman	(02)
18.	Delhi (07)	071	Delhi	129.	North West	(01)
				130.	North	(02)
				131.	North East	(03)
				132.	East	(04)
				133.	New Delhi	(05)
				134.	Central	(06)
				135.	West	(07)
				136.	South West	(08)
				137.	South	(09)
19.	Goa (30)	301	Goa	138.	North Goa	(01)
				139.	South Goa	(02)
20.	Gujarat (24)	241	South Eastern	140.	Panch Mahals	(17)
				141.	Dohad	(18)
				142.	Vadodara	(19)
				143.	Narmada	(20)
				144.	Bharuch	(21)
				145.	The Dangs	(22)
				146.	Navsari	(23)
				147.	Valsad	(24)
				148.	Surat	(25)
				149.	Tapi	(26)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
21.		242	Plains Northern	150.	Mahesana	(04)
				151.	Sabar Kantha	(05)
				152.	Gandhinagar	(06)
				153.	Ahmadabad	(07)
				154.	Anand	(15)
				155.	Kheda	(16)
22.		243	Dry areas	156.	Banas Kantha	(02)
				157.	Patan	(03)
23.		244	Kachchh	158.	Kachchh	(01)
24.		245	Saurashtra	159.	Surendranagar	(08)
				160.	Rajkot	(09)
				161.	Jamnagar	(10)
				162.	Porbandar	(11)
				163.	Junagadh	(12)
				164.	Amreli	(13)
				165.	Bhavnagar	(14)
25.	Haryana (06)	061	Eastern	166.	Panchkula	(01)
				167.	Ambala	(02)
				168.	Yamunanagar	(03)
				169.	Kurukshetra	(04)
				170.	Kaithal	(05)
				171.	Karnal	(06)
				172.	Panipat	(07)
				173.	Sonipat	(08)
				174.	Rohtak	(14)
				175.	Jhajjar	(15)
				176.	Gurgaon	(18)
				177.	Mewat	(19)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
				178.	Faridabad	(20)
				179.	Palwal	(21)
26.		062	Western	180.	Jind	(09)
				181.	Fatehabad	(10)
				182.	Sirsa	(11)
				183.	Hisar	(12)
				184.	Bhiwani	(13)
				185.	Mahendragarh	(16)
				186.	Rewari	(17)
27	Himachal Pradesh (02)	021	Central	187.	Kangra	(02)
				188.	Kullu	(04)
				189.	Mandi	(05)
				190.	Hamirpur	(06)
				191.	Una	(07)
28.		022	Trans Himalayan & Southern	192.	Chamba	(01)
				193.	Lahul & Spiti	(03)
				194.	Bilaspur	(08)
				195.	Solan	(09)
				196.	Sirmaur	(10)
				197.	Shimla	(11)
				198.	Kinnaur	(12)
29.	Jammu & Kashmir (01)	011	Mountainous	199.	Kathua	(07)
				200.	Jammu	(21)
				201.	Samba	(22)
30.		012	Outer Hills	202.	Punch	(05)
				203.	Rajouri	(06)
				204.	Doda	(16)
				205.	Ramban	(17)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
				206.	Kishtwar	(18)
				207.	Udhampur	(19)
				208.	Reasi	(20)
31.		013	Jhelum Valley	209.	Kupwara	(01)
				210.	Badgam	(02)
				211.	Baramula	(08)
				212.	Bandipore	(09)
				213.	Srinagar	(10)
				214.	Ganderbal	(11)
				215.	Pulwama	(12)
				216.	Shupiyan	(13)
				217.	Anantnag	(14)
				218.	Kulgam	(15)
32.		014	Ladakh	219.	Leh	(03)
				220.	Kargil	(04)
33.	Jharkhand (20)	201	Ranchi Plateau	221.	Garhwa	(01)
				222.	Lohardaga	(11)
				223.	Purbi Singhbhum	(12)
				224.	Palamu	(13)
				225.	Latehar	(14)
				226.	Ranchi	(19)
				227.	Khunti	(20)
				228.	Gumla	(21)
				229.	Simdega	(22)
				230.	Pashchimi Singh-bhum	(23)
				231.	Saraikela-Kharsawan	(24)
				232.	Chatra	(02)
				233.	Kodarma	(03)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
34.		202	Hazaribagh Plateau	234.	Giridih	(04)
				235.	Deoghar	(05)
				236.	Godda	(06)
				237.	Sahibganj	(07)
				238.	Pakur	(08)
				239.	Dhanbad	(09)
				240.	Bokaro	(10)
				241.	Hazaribagh	(15)
				242.	Ramgarh	(16)
				243.	Dumka	(17)
				244.	Jamtara	(18)
35.	Karnataka (29)	291	Coastal & Ghats	245.	Uttara Kannada	(09)
				246.	Udupi	(15)
				247.	Dakshina Kannada	(21)
36.		292	Inland Eastern	248.	Shimoga	(14)
				249.	Chikmagalur	(16)
				250.	Hassan	(20)
				251.	Kodagu	(22)
37.		293	Inland Southern	252.	Tumkur	(17)
				253.	Bangalore	(18)
				254.	Mandya	(19)
				255.	Mysore	(23)
				256.	Chamarajanagar	(24)
				257.	Kolar	(27)
				258.	Chikkaballapura	(28)
				259.	Bangalore Rural	(29)
				260.	Ramanagara	(30)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
38.		294	Inland Northern	261.	Belgaum	(01)
				262.	Bagalkot	(02)
				263.	Bijapur	(03)
				264.	Bidar	(04)
				265.	Raichur	(05)
				266.	Koppal	(06)
				267.	Gadag	(07)
				268.	Dharwad	(08)
				269.	Haveri	(10)
				270.	Bellary	(11)
				271.	Chitradurga	(12)
				272.	Davanagere	(13)
				273.	Gulbarga	(25)
				274.	Yadgir	(26)
39.	Kerala (32)	321	Northern	275.	Kasaragod	(01)
				276.	Kannur	(02)
				277.	Wayanad	(03)
				278.	Kozhikode	(04)
				279.	Malappuram	(05)
				280.	Palakkad	(06)
40.		322	Southern	281.	Thrissur	(07)
				282.	Ernakulam	(08)
				283.	Idukki	(09)
				284.	Kottayam	(10)
				285.	Alappuzha	(11)
				286.	Pathanamthitta	(12)
				287.	Kollam	(13)
				288.	Thiruvananthapuram	(14)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
41.	Lakshadweep (31)	311	Lakshadweep	289.	Lakshadweep	(01)
42.	Madhya Pradesh (23)	231	Vindhya	290.	Tikamgarh	(07)
				291.	Chhatarpur	(08)
				292.	Panna	(09)
				293.	Satna	(12)
				294.	Rewa	(13)
				295.	Umaria	(14)
				296.	Shahdol	(43)
				297.	Anuppur	(44)
				298.	Sidhi	(45)
				299.	Singrauli	(46)
43.		232	Central	300.	Sagar	(10)
				301.	Damoh	(11)
				302.	Vidisha	(26)
				303.	Bhopal	(27)
				304.	Sehore	(28)
				305.	Raisen	(29)
44.		233	Malwa	306.	Neemuch	(15)
				307.	Mandsaur	(16)
				308.	Ratlam	(17)
				309.	Ujjain	(18)
				310.	Shajapur	(19)
				311.	Dewas	(20)
				312.	Dhar	(21)
				313.	Indore	(22)
				314.	Rajgarh	(25)
				315.	Jhabua	(47)
				316.	Alirajpur	(48)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
45.		234	South	317.	Katni	(33)
				318.	Jabalpur	(34)
				319.	Narsimhapur	(35)
				320.	Dindori	(36)
				321.	Mandla	(37)
				322.	Chhindwara	(38)
				323.	Seoni	(39)
				324.	Balaghat	(40)
46.		235	South Western	325.	Khargone (West Nimar)	(23)
				326.	Barwani	(24)
				327.	Betul	(30)
				328.	Harda	(31)
				329.	Hoshangabad	(32)
				330.	Khandwa (East Nimar)	(49)
				331.	Burhanpur	(50)
47.		236	Northern	332.	Sheopur	(01)
				333.	Morena	(02)
				334.	Bhind	(03)
				335.	Gwalior	(04)
				336.	Datia	(05)
				337.	Shivpuri	(06)
				338.	Guna	(41)
				339.	Ashoknagar	(42)
48.	Maharashtra (27)	271	Coastal	340.	Thane	(21)
				341.	Mumbai Suburban	(22)
				342.	Mumbai	(23)
				343.	Raigarh	(24)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
				344.	Ratnagiri	(32)
				345.	Sindhudurg	(33)
49.		272	Inland Western	346.	Pune	(25)
				347.	Ahmadnagar	(26)
				348.	Solapur	(30)
				349.	Satara	(31)
				350.	Kolhapur	(34)
				351.	Sangli	(35)
50.		273	Inland Northern	352.	Nandurbar	(01)
				353.	Dhule	(02)
				354.	Jalgaon	(03)
				355.	Nashik	(20)
51.		274	Inland Central	356.	Nanded	(15)
				357.	Hingoli	(16)
				358.	Parbhani	(17)
				359.	Jalna	(18)
				360.	Aurangabad	(19)
				361.	Bid	(27)
				362.	Latur	(28)
				363.	Osmanabad	(29)
52.		275	Inland Eastern	364.	Buldana	(04)
				365.	Akola	(05)
				366.	Washim	(06)
				367.	Amravati	(07)
				368.	Wardha	(08)
				369.	Nagpur	(09)
				370.	Yavatmal	(14)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
53.		276	Eastern	371.	Bhandara	(10)
				372.	Gondiya	(11)
				373.	Gadchiroli	(12)
				374.	Chandrapur	(13)
54.	Manipur (14)	141	Plains	375.	Bishnupur	(04)
				376.	Thoubal	(05)
				377.	Imphal West	(06)
				378.	Imphal East	(07)
55.		142	Hills	379.	Senapati	(01)
				380.	Tamenglong	(02)
				381.	Churachandpur	(03)
				382.	Ukhrul	(08)
				383.	Chandel	(09)
56.	Meghalaya (17)	171	Meghalaya	384.	West Garo Hills	(01)
				385.	East Garo Hills	(02)
				386.	South Garo Hills	(03)
				387.	West Khasi Hills	(04)
				388.	Ribhoi	(05)
				389.	East Khasi Hills	(06)
				390.	Jaintia Hills	(07)
57.	Mizoram (15)	151	Mizoram	391.	Mamit	(01)
				392.	Kolasib	(02)
				393.	Aizwal	(03)
				394.	Champhai	(04)
				395.	Serchhip	(05)
				396.	Lunglei	(06)
				397.	Lawngtlai	(07)
				398.	Saiha	(08)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
58.	Nagaland (13)	131	Nagaland	399.	Mon	(01)
				400.	Mokokchung	(02)
				401.	Zunheboto	(03)
				402.	Wokha	(04)
				403.	Dimapur	(05)
				404.	Phek	(06)
				405.	Tuensang	(07)
				406.	Longleng	(08)
				407.	Kiphire	(09)
				408.	Kohima	(10)
				409.	Peren	(11)
59.	Odisha (21)	211	Coastal	410.	Baleshwar	(08)
				411.	Bhadrak	(09)
				412.	Kendrapara	(10)
				413.	Jagatsinghapur	(11)
				414.	Cuttack	(12)
				415.	Jajapur	(13)
				416.	Nayagarh	(16)
				417.	Khordha	(17)
				418.	Puri	(18)
60.		212	Southern	419.	Ganjam	(19)
				420.	Gajapati	(20)
				421.	Kandhamal	(21)
				422.	Baudh	(22)
				423.	Subarnapur	(23)
				424.	Balangir	(24)
				425.	Nuapada	(25)
				426.	Kalahandi	(26)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
				427.	Rayagada	(27)
				428.	Nabarangapur	(28)
				429.	Koraput	(29)
				430.	Malkangiri	(30)
61.		213	Northern	431.	Bargarh	(01)
				432.	Jharsuguda	(02)
				433.	Sambalpur	(03)
				434.	Debagarh	(04)
				435.	Sundargarh	(05)
				436.	Kendujhar	(06)
				437.	Mayurbhanj	(07)
				438.	Dhenkanal	(14)
				439.	Anugul	(15)
62.	Puducherry (34)	341	Puducherry	440.	Yanam	(01)
				441.	Puducherry	(02)
				442.	Mahe	(03)
				443.	Karaikal	(04)
63.	Punjab (03)	031	Northern	444.	Gurdaspur	(01)
				445.	Kapurthala	(02)
				446.	Jalandhar	(03)
				447.	Hoshiarpur	(04)
				448.	Shahid Bhagat Singh Nagar	(05)
				449.	Amritsar	(15)
				450.	Tarn Taran	(16)
				451.	Rupnagar	(17)
				452.	Sahibzada Ajit Singh Nagar	(18)
				453.	Pathankot	(21)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
64.		032	Southern	454.	Fatehgarh Sahib	(06)
				455.	Ludhiana	(07)
				456.	Moga	(08)
				457.	Firozpur	(09)
				458.	Muktsar	(10)
				459.	Faridkot	(11)
				460.	Bhatinda	(12)
				461.	Mansa	(13)
				462.	Patiala	(14)
				463.	Sangrur	(19)
				464.	Barnala	(20)
				465.	Fazilka	(22)
65.	Rajasthan (08)	081	Western	466.	Bikaner	(03)
				467.	Jodhpur	(15)
				468.	Jaisalmer	(16)
				469.	Barmer	(17)
				470.	Jalor	(18)
				471.	Sirohi	(19)
				472.	Pali	(20)
66.		082	North-Eastern	473.	Alwar	(06)
				474.	Bharatpur	(07)
				475.	Dhaulpur	(08)
				476.	Karauli	(09)
				477.	Sawai Madhopur	(10)
				478.	Dausa	(11)
				479.	Jaipur	(12)
				480.	Ajmer	(21)
				481.	Tonk	(22)
				482.	Bhilwara	(24)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
67.		083	Southern	483.	Rajsamand	(25)
				484.	Dungarpur	(26)
				485.	Banswara	(27)
				486.	Udaipur	(32)
68.		084	South-Eastern	487.	Bundi	(23)
				488.	Chittaurgarh	(28)
				489.	Kota	(29)
				490.	Baran	(30)
				491.	Jhalawar	(31)
				492.	Pratapgarh	(33)
69.		085	Northern	493.	Sri Ganganagar	(01)
				494.	Hanumangarh	(02)
				495.	Churu	(04)
				496.	Jhunjhunun	(05)
				497.	Sikar	(13)
				498.	Nagaur	(14)
70.	Sikkim (11)	111	Sikkim	499.	North District	(01)
				500.	West District	(02)
				501.	South District	(03)
				502.	East District	(04)
71.	Tamil Nadu (33)	331	Coastal Northern	503.	Thiruvallur	(01)
				504.	Chennai	(02)
				505.	Kancheepuram	(03)
				506.	Vellore	(04)
				507.	Tiruvannamalai	(05)
				508.	Viluppuram	(06)
				509.	Cuddalore	(16)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
72.		332	Coastal	510.	Karur	(12)
				511.	Tiruchirappalli	(13)
				512.	Perambalur	(14)
				513.	Ariyalur	(15)
				514.	Nagapattinam	(17)
				515.	Thiruvarur	(18)
				516.	Thanjavur	(19)
				517.	Pudukkottai	(20)
73.		333	Southern	518.	Dindigul	(11)
				519.	Sivaganga	(21)
				520.	Madurai	(22)
				521.	Theni	(23)
				522.	Virudhunagar	(24)
				523.	Ramanathapuram	(25)
				524.	Thoothukkudi	(26)
				525.	Tirunelveli	(27)
				526.	Kanniyakumari	(28)
74.		334	Inland	527.	Salem	(07)
				528.	Namakkal	(08)
				529.	Erode	(09)
				530.	The Nilgiris	(10)
				531.	Dharmapuri	(29)
				532.	Krishnagiri	(30)
				533.	Coimbatore	(31)
				534.	Tiruppur	(32)
				535.	Adilabad	(01)
				536.	Komaram Bheem	(02)
				537.	Mancherial	(03)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
75.	Telangana (36)	361	Inland North Western	538.	Nirmal	(04)
				539.	Nizamabad	(05)
				540.	Kamareddy	(15)
				541.	Sangareddy	(16)
				542.	Medak	(17)
				543.	Siddipet	(18)
				544.	Medchal-Malkajgiri	(21)
				545.	Hyderabad	(22)
				546.	Rangareddy	(23)
				547.	Vikarabad	(24)
				548.	Mahbubnagar	(25)
				549.	Jogulamba	(26)
				550.	Wanaparthy	(27)
				551.	Nagarkurnool	(28)
76.		362	Inland North Eastern	552.	Jagtial	(06)
				553.	Peddapalli	(07)
				554.	Jayashankar	(08)
				555.	Bhadradri	(09)
				556.	Mahabubabad	(10)
				557.	Warangal Rural	(11)
				558.	Warangal Urban	(12)
				559.	Karimnagar	(13)
				560.	Rajanna	(14)
				561.	Jangaon	(19)
				562.	Yadadri	(20)
				563.	Nalgonda	(29)
				564.	Suryapet	(30)
				565.	Khammam	(31)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
77.	Tripura (16)	161	Tripura	566.	West Tripura	(01)
				567.	South Tripura	(02)
				568.	Dhalai	(03)
				569.	North Tripura	(04)
78.	Uttarakhand (05)	051	Uttarakhand	570.	Uttarkashi	(01)
				571.	Chamoli	(02)
				572.	Rudraprayag	(03)
				573.	Tehri Garhwal	(04)
				574.	Dehradun	(05)
				575.	Garhwal	(06)
				576.	Pithoragarh	(07)
				577.	Bageshwar	(08)
				578.	Almora	(09)
				579.	Champawat	(10)
				580.	Nainital	(11)
				581.	Udham Singh Nagar	(12)
				582.	Hardwar	(13)
79	Uttar Pradesh (09)	091	Northern Upper Ganga Plains	583.	Saharanpur	(01)
				584.	Muzaffarnagar	(02)
				585.	Bijnor	(03)
				586.	Moradabad	(04)
				587.	Rampur	(05)
80.		092	Central	588.	Jyotiba Phule Nagar	(06)
				589.	Meerut	(07)
				590.	Baghpat	(08)
				591.	Ghaziabad	(09)
				592.	Gautam Buddha Nagar	(10)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
				593.	Sitapur	(23)
				594.	Hardoi	(24)
				595.	Unnao	(25)
				596.	Lucknow	(26)
				597.	Rae Bareli	(27)
				598.	Kanpur Dehat	(32)
				599.	Kanpur Nagar	(33)
				600.	Fatehpur	(41)
				601.	Bara Banki	(45)
				602.	Pratapgarh	(42)
				603.	Kaushambi	(43)
				604.	Allahabad	(44)
				605.	Faizabad	(46)
				606.	Ambedkar Nagar	(47)
				607.	Sultanpur	(48)
				608.	Bahraich	(49)
				609.	Shrawasti	(50)
				610.	Balrampur	(51)
81.		093	Eastern	611.	Gonda	(52)
				612.	Siddharthnagar	(53)
				613.	Basti	(54)
				614.	Sant Kabir Nagar	(55)
				615.	Maharajganj	(56)
				616.	Gorakhpur	(57)
				617.	Kushinagar	(58)
				618.	Deoria	(59)
				619.	Azamgarh	(60)
				620.	Mau	(61)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
				621.	Ballia	(62)
				622.	Jaunpur	(63)
				623.	Ghazipur	(64)
				624.	Chandauli	(65)
				625.	Varanasi	(66)
				626.	Sant Ravidas Nagar (Bhadohi)	(67)
				627.	Mirzapur	(68)
				628.	Sonbhadra	(69)
82.		904	Southern	629.	Jalaun	(34)
				630.	Jhansi	(35)
				631.	Lalitpur	(36)
				632.	Hamirpur	(37)
				633.	Mahoba	(38)
				634.	Banda	(39)
				635.	Chitrakoot	(40)
83.		095	Southern Upper Ganga Plains	636.	Bulandshahr	(11)
				637.	Aligarh	(12)
				638.	Mahamaya Nagar	(13)
				639.	Mathura	(14)
				640.	Agra	(15)
				641.	Firozabad	(16)
				642.	Mainpuri	(17)
				643.	Budaun	(18)
				644.	Bareilly	(19)
				645.	Pilibhit	(20)
				646.	Shahjahanpur	(21)
				647.	Kheri	(22)
				648.	Farrukhabad	(28)

sl. no.	State/u.t. (code)	NSS region		Detailed composition of region		
		no	description	sl. no.	name of district	code
(1)	(2)	(3)	(4)	(5)	(6)	(7)
				649.	Kannauj	(29)
				650.	Etawah	(30)
				651.	Auraiya	(31)
				652.	Etah	(70)
				653.	Kanshiram Nagar	(71)
84.	West Benga (19)	191	Himalayan	654.	Darjiling	(01)
				655.	Jalpaiguri	(02)
				656.	Koch Bihar	(03)
85.		192	Eastern Plains	657.	Uttar Dinajpur	(04)
				658.	Dakshin Dinajpur	(05)
				659.	Maldah	(06)
				660.	Murshidabad	(07)
				661.	Birbhum	(08)
				662.	Nadia	(10)
86.		193	Southern Plains	663.	North Twenty Four Parganas	(11)
				664.	Kolkata	(16)
				665.	South Twenty Four Parganas	(17)
87.		194	Central Plains	666.	Barddhaman	(09)
				667.	Hugli	(12)
				668.	Haora	(15)
88.		195	Western Plains	669.	Bankura	(13)
				670.	Puruliya	(14)
				671.	Paschim Medinipur	(18)
				672.	Purba Medinipur	(19)

ANNEX III: Map of NSS Regions

ANNEX IV: DEFINITIONS OF VARIABLES GENERATED AND USED IN THIS STUDY

A large number of variables were generated for this study. The main data source is periodic labour force survey of NSSO. Some variables were also generated from Sample Registration System (SRS), National Health Family Survey (NFHS) IV and Census of India, 2011.

The concepts behind generation of these variables are described below one after another.

A. Periodic Labour Force Survey (PLFS) 2018-19 conducted by National Sample Survey Organisation

1. Literacy level: A person who can both read and write a simple message with understanding in at least one language was considered as literate for the purpose of the survey.

Literacy rate is calculated as percentage of 'not literate' to total population.

2. Activity status: It is the activity situation in which a person was found during a specified reference period with regard to the person's participation in economic and non-economic activities. According to this, a person could be in one or a combination of the following three broad activity statuses during the reference period:

i. working or being engaged in economic activity (work),

ii. being not engaged in economic activity (work) but either making tangible efforts to seek 'work' or being available for 'work' if 'work' is available, and

iii. being not engaged in any economic activity (work) and also not available for 'work'.

Broad activity statuses mentioned in (i) and (ii) above are associated with 'being in labour force' and the last with 'not being in the labour force'. Within the labour force, broad activity status (i) and (ii) were associated with 'employment' and 'unemployment', respectively.

Identification of each individual under any one of the three broad activity statuses (viz. employed/ unemployed/not in labour force) was done by adopting either the major time criterion or priority criterion. The former was used for classification of persons according to the usual status approach and the latter for classification of persons according to the current status approach (i.e., current weekly status and current daily status approaches).

If a person categorised as engaged in economic activity was found to be pursuing more than one economic activity during the reference period, the appropriate detailed activity in terms of status and industry was considered as that one corresponding to which relatively long time had been spent.

Code description

Working (or employed) as self-employed

11 worked in household enterprises (self-employed) as own-account worker

12 worked in household enterprises (self-employed) as an employer

21 worked in household enterprises (self-employed) as helper

Regular wage/salaried employee

31 worked as regular wage salaried employee

Casual labour

41 worked as casual labour in public works other than MGNREG public works

42 worked as casual labour in Mahatma Gandhi NREG public works (for current weekly status approach)

51 worked as casual labour in other types of work

61 did not work owing to sickness though there was work in household enterprise (for current weekly status approach)

62 did not work owing to other reasons though there was work in household enterprise (for current weekly status approach)

71 did not work owing to sickness but had regular salaried/wage employment (for current weekly status approach)

72 did not work owing to other reasons but had regular salaried/wage employment (for current weekly status approach)

Not working but seeking/available for work (or unemployed)

81 sought work or did not seek but was available for work (for usual status approach)

81 sought work (for current weekly status approach)

82 did not seek but was available for work (for current weekly status approach)

Neither working nor available for work (or not in labour force)

91 attended educational institutions

92 attended to domestic duties only

93 attended to domestic duties and was also engaged in free collection of goods (vegetables, roots, firewood, cattle feed, etc.), sewing, tailoring, weaving, etc. for household use

94 rentiers, pensionners, remittance récipients, etc.

95 not able to work owing to disability

97 others (including beggars, prostitutes, etc.)

98 did not work owing to sickness (for current weekly status approach)

99 children of age 0-4 years

3. Workers (or employed): Relevant activity status codes 11 to 72 were assigned for workers. Workers were further categorized as *self-employed (*relevant activity status codes: 11, 12, 21, 61, 62)*, regular wage /salaried employee (*relevant activity status codes: 31, 71, 72)*,* and *casual labour (*relevant activity status codes: 41, 42 and 51).

4. Seeking or available for work (or unemployed): Activity status codes 81 or 82 were assigned for unemployed.

5. Labour force: Persons who were either 'working' (or employed) or 'seeking or available for work' (or unemployed) constituted the labour force. Persons with activity status codes 11 to 72 and 81 or 82 constituted the labour force.

6. Not in labour force: Activity status codes 91-95, 97, 98 and 99 were assigned for persons belonging to category 'not in labour force'.

7. Self-employed: Persons who operated their own farm or non-farm enterprises or were engaged independently in a profession or trade on own-account or with one or a few partners were deemed to be self-employed in household enterprises. The remuneration of the self-employed consists of a non-separable combination of two parts: a reward for their labour and profit of their enterprise.

8. Regular wage/salaried employee: These were persons who worked in others' farm or nonfarm enterprises (both household and non-household) and, in return, received salary or wages on a regular basis (i.e. not on the basis of daily or periodic renewal of work contract). This category included not only persons getting time wage but also persons receiving piece wage or salary and paid apprentices, both full time and part-time.

9. Casual labour: A person who was casually engaged in others' farm or non-farm enterprises (both household and non-household) and, in return, received wages according to the terms of the daily or periodic work contract, was considered as a casual labour.

10. Usual activity status: The activity status on which a person spent relatively long time (major time criterion) during the 365 days preceding the date of survey was considered the *usual principal activity status* of the person. To decide the usual principal activity of a person, he/she was first categorised as belonging to the labour force or not, during the reference period on the basis of major time criterion. Persons, thus adjudged as not belonging to the labour force, were assigned the broad activity status 'neither working nor available for work'.

For the persons belonging to the labour force, the broad activity status of either 'working' or 'not working but seeking and/or available for work' was then ascertained again on the basis of the relatively long time spent in the labour force during the 365 days preceding the date of survey.

Usual principal status of a person was determined as the status on which the person spent relatively long time (major time criterion) during the 365 days preceding the date of survey. Such persons might have also pursued, in addition to his/her usual principal status, some economic activity for 30 days or more during the reference period of 365 days preceding the date of survey. The status in which such economic activity was pursued during the reference period of 365 days preceding the date of survey was the *subsidiary economic activity status* of the person.

11. Current weekly activity status (CWS): The current weekly activity status of a person is the activity status obtaining for a person during a reference period of 7 days preceding the date of survey. *A person is considered working (or employed)) if he/she worked for at least one hour on at least one day during the 7 days preceding the date of survey or if he/she had work for at least 1 hour on at least one day during the 7 days preceding the date of the survey but did not do the work.*

12. Earnings of the regular salaried/wage employees and casual labours: For collecting information on earning of 'regular salaried/wage employees' and 'casual labours', following points were taken into consideration:

(a) Earnings of the regular salaried/wage employees and casual labours relates to remuneration in cash and in kind which were paid, as a rule at regular intervals, for time worked (including overtime payments) or work done together with remuneration for time not worked, such as for annual vacation, other paid leave or holidays.

(b) It relates to employees' gross remuneration (i.e. the total before any deductions were made by the employer in respect of taxes, contributions of employees to social security and pension schemes, life insurance premiums and other obligations of employees).

(c) Earnings exclude employers' contributions paid to social security and pension schemes in respect of their employees and also the benefits received by employees under these schemes.

(d) Earnings exclude severance and termination pay.

(e) Wages in kind were evaluated at the respective current retail price.

(f) Bonus (expected or paid) was duly apportioned for the reference period for inclusion in earnings.

13. Gross earnings of self-employed persons: The gross earnings of the self-employed persons during a reference period were derived by the procedure stated below:

(a) Gross earnings of the self-employed persons during a reference period was obtained by deducting total expenses from the gross output of that period.

(b) The gross output corresponds to the sum of the values of all goods and services produced during the reference period, including any part which has been retained for own consumption or given free of charge or at reduced prices to hired labour.

(c) The valuation of output was made at basic price. The basic price is defined as the amount receivable by the producer from the purchaser for a unit of good or service produced as output minus any tax payable on product (like, excise duties, sales tax, non-deductible vat, etc.) plus any subsidy receivable on the product for selling those to the ultimate consumer at lesser price.

(d) Total expenses include (i) current expenses of the enterprise, such as purchase of raw materials, tools and equipment, fuel, electricity, etc., (ii) payments to hired labour, (iii) rent paid for fixed capital items and interest payments on financial assets, (iv) taxes on production (like, recurrent taxes on land & buildings, business or professional licence fees, road tax, registration fee of vehicles, etc.) paid by the enterprise, reduced by subsidies received on production (like, employing physically challenged persons, installing pollution control equipment, etc.). Thus, gross earning = gross output – total expenses.

(f) If the owners of the enterprise are from the same household, earning was judged by considering equal distribution of income among all the owners. If the owners of the enterprise were from the different households, earning from the partnership business was be distributed, according to the agreement (verbal or written), among the partner households.

(g) For helpers in household enterprises, gross earnings was considered as zero (0).

14. For calculating organised sector workers, two criteria were used:

i) All types of enterprises except for Proprietary and Partnership are considered to be part of organised sector.

ii) All enterprises having number of workers 10 or more are considered part of organised sector.

15. Broad sector of employment was defined as:

i) Primary sector contains agriculture and allied sectors like plantation, animal husbandry, etc.

ii) Secondary sector includes mining & quarrying, manufacturing, electricity, gas & water supply and construction.

iii) Tertiary sector consists of trade, hotel & restaurants, transport, storage & communication, finance, business & real estate and public administration, education & health.

16. Location of workplace. Irrespective of usual residence of workers either in rural or urban areas, it is important to know where they actually work. Workers can stay in rural/urban areas but can commute to urban/rural location. Therefore, location of workplace has three broad classifications: i) rural, ii) urban, and iii) no fixed workplace which is termed as mobile.

17. Earnings from employment. In PLFS, information on earnings from employment was collected for all the three categories of workers, viz., *self-employed persons*, *regular wage/ salaried employees,* and *casual labour* as follows

i). For self-employed persons in current weekly status, information on earnings during the last 30 days from the self-employment activity in which the person was working as per current weekly status was collected.

ii). For regular wage/salaried employees in current weekly status, information on earnings during the preceding calendar month from the regular wage/ salaried work in which the person was employed in the current weekly status was collected.

iii). For casual labour, information on earnings was collected for the casual labour work in which the person was engaged for each day of the reference week

18. Household Monthly Income. Household monthly income is obtained by adding monthly earnings of regular salaried/wage employees, casual labours and gross earnings of self-employed persons within the household.

19. Monthly per capita income. This is obtained by dividing household monthly income by corresponding household size.

20. Household type. The household type was decided based on the sources of the household's income during the 365 days preceding the date of survey. For this purpose, only the household's income (net income and not gross income) from economic activities was considered; the incomes of servants and paying guests were not taken into account.

For **rural** areas, a household belonged to any one of the following six household types:

1. Self-employed in agriculture
2. Self-employed in non-agriculture
3. Regular wage/salary earning
4. Casual labour in agriculture
5. Casual labour in non-agriculture
6. Others

For **urban** areas, the household types are:

1. Self-employed
2. Regular wage/salary earning
3. Casual labour
4. Others

B. Sample Registration System (SRS), 2018

21. Crude birth rate, Crude death rate and Infant mortality rate. The Sample Registration System (SRS) in India is carried out by the Office of Registrar General & Census Commissioner, India with an objective of providing reliable annual estimates of birth rate, death rate, infant mortality rate and various other fertility and mortality indicators. SRS is one of the largest demographic surveys in the world covering about 8.1 million population. It serves as the main source of information on fertility and mortality both at the State and National levels.

Apart from the large sample size and geographic spread in most of the districts, the system has a unique feature of dual recording, which involves continuous enumeration and retrospective half yearly surveys. The continuous enumeration and retrospective surveys are followed by the process of matching of the two records and subsequent field verification of unmatched andpartially matched events. The system provides for a cross-check on the correctness and completeness of the events of birth and death listed by the two independent functionaries.

Total fertility rate (TFR) is not available at NSS regional levels.

Definitions

Crude birth rate = (CBR)

$$\frac{\text{Number of live births during the year}}{\text{Mid-year population}} \times 1000$$

Crude death rate = (CDR)

$$\frac{\text{Number of deaths during the year}}{\text{Mid-year population}} \times 1000$$

Infant mortality rate = (IMR)

$$\frac{\text{Number of infant deaths during the year}}{\text{Number of live births during the year}} \times 1000$$

For calculating birth, death and infant mortality rates, we could not calculate these separately for rural and urban areas at NSS regional level of disaggregation because of data paucity even when altogether 8.1 million population was covered. The reliable estimation of demographic variables at a disaggregated level requires very large number of observations that is possible when census data are available. The estimation is available only at combined rural and urban areas at NSS regional level.

For major states, NSS regional level data was generated by aggregating district level data. For smaller states and union territories, the information was available at respective state and union territories level. Since smaller states and union territories do not have more than one NSS region, NSS regional level data represented them.

C. National Family Health Survey IV (2015-16)

The survey was conducted in two phases of full calendar years of 2015 and 2016 by dividing India into two zones. The primary objective of the 2015-16 National Family Health Survey was to provide essential data on health and family welfare, as well as data on emerging issues in these areas.

22. Vegetarianism. *Vegetarian* is defined as complete abstaining from eating chicken, meat, fish and egg. A limitation of calculating share of vegetarian population is that it has much higher weighing of women, as the ratio of women to men in the survey is 6.24:1. The estimation was initially done at the district level and then these were aggregated to arrive at values for the NSS regional level.

D. Census of India, 2011

23. Population density. Census of India district level data were used to calculate population density. Population density is calculated as the ratio of population to geographical area in square kilometres.

For calculating population density at NSS regional levels, the district level population and geographical area were summed up at the corresponding NSS regional levels and then population density was obtained by dividing population by geographical areas. The limitation of this calculation is that population density estimation is not available separately for rural and urban areas. The main problem was getting geographical areas for urban areas. For metropolitan cities and larger towns, geographical area data are separately available. But for smaller towns that are not notified, geographical areas are not separately available. The main problem areas are census towns most of which are not notified and have more than 50 thousand population.

Population density calculations pertain to the year 2011 as population projections at district level are not available after the census of 2011.

ANNEX V Table 1: List of Variables Generated and Used in This Study

Sl. No.	Description of Variables	Rural	Urban	Total
1	Monthly Per capita Consumption Expenditure	Yes	Yes	Yes
2	Household Usual Monthly Consumption Expenditure	Yes	Yes	Yes
3	Size of Households	Yes	Yes	Yes
4	Distribution of Type of Households	Yes	Yes	No
5	Monthly Income of Households	Yes	Yes	Yes
6	Monthly Per Capita Income	Yes	Yes	Yes
7	Literacy Rate (Male)	Yes	Yes	Yes
8	Literacy Rate (Female)	Yes	Yes	Yes
9	Literacy Rate (Persons)	Yes	Yes	Yes
10	Labour Force Participation Rate (Male), UPSS	Yes	Yes	Yes

Sl. No.	Description of Variables	Rural	Urban	Total
11	Labour Force Participation Rate (Female), UPSS	Yes	Yes	Yes
12	Labour Force Participation Rate (Persons), UPSS	Yes	Yes	Yes
13	Labour Force Participation Rate (Male), UPS	Yes	Yes	Yes
14	Labour Force Participation Rate (Female), UPS	Yes	Yes	Yes
15	Labour Force Participation Rate (Persons), UPS	Yes	Yes	Yes
16	Work Force Participation Rate (Male), UPS	Yes	Yes	Yes
17	Work Force Participation Rate (Female), UPS	Yes	Yes	Yes
18	Work Force Participation Rate (Persons), UPS	Yes	Yes	Yes
19	Work Force Participation Rate (Male), CWS	Yes	Yes	Yes
20	Work Force Participation Rate (Female), CWS	Yes	Yes	Yes
21	Work Force Participation Rate (Persons), CWS	Yes	Yes	Yes
22	Unemployment Rate (Male), UPS	Yes	Yes	Yes
23	Unemployment Rate (Female), UPS	Yes	Yes	Yes
24	Unemployment Rate (Person), UPS	Yes	Yes	Yes
25	Unemployment Rate (Male), CWS	Yes	Yes	Yes
26	Unemployment Rate (Female), CWS	Yes	Yes	Yes
27	Unemployment Rate (Person), CWS	Yes	Yes	Yes
28	Monthly Average Earnings of Regular/Salaried Workers, Male	Yes	Yes	Yes
29	Monthly Average Earnings of Regular/Salaried Workers, Female	Yes	Yes	Yes
30	Monthly Average Earnings of Regular/Salaried Workers, Person	Yes	Yes	Yes
31	Daily Wage Rate of Casual Workers, Male	Yes	Yes	Yes
32	Daily Wage Rate of Casual Workers, Female	Yes	Yes	Yes
33	Daily Wage Rate of Casual Workers, Person	Yes	Yes	Yes
34	Broad Sectoral Share of Employment, Male	Yes	Yes	Yes
35	Broad Sectoral Share of Employment, Female	Yes	Yes	Yes
36	Broad Sectoral Share of Employment, Person	Yes	Yes	Yes

Sl. No.	Description of Variables	Rural	Urban	Total
37	Share of Manufacturing Employment, UPS, Male	Yes	Yes	Yes
38	Share of Manufacturing Employment, UPS, Female	Yes	Yes	Yes
39	Share of Manufacturing Employment, UPS, Person	Yes	Yes	Yes
40	Share of Rural/Urban Location of Work, UPS, Male	Yes	Yes	No
41	Share of Rural/Urban Location of Work, UPS, Female	Yes	Yes	No
42	Share of Rural/Urban Location of Work, UPS, Person	Yes	Yes	No
43	Share of Organised Sector Employment, UPS, Male	Yes	Yes	Yes
44	Share of Organised Sector Employment, UPS, Female	Yes	Yes	Yes
45	Share of Organised Sector Employment, UPS, Person	Yes	Yes	Yes
46	Share of Organised Manufacturing Sector Employment, UPS, Male	Yes	Yes	Yes
47	Share of Organised Manufacturing Sector Employment, UPS, Female	Yes	Yes	Yes
48	Share of Organised Manufacturing Sector Employment, UPS, Person	Yes	Yes	Yes
49	Number of Organised Manufacturing Sector Workers, UPS, Male	Yes	Yes	Yes
50	Number of Organised Manufacturing Sector Workers, UPS, Female	Yes	Yes	Yes
51	Number of Organised Manufacturing Sector Workers, UPS, Person	Yes	Yes	Yes
52	Share of Vegetarian Population	Yes	Yes	Yes
53	Crude Birth Rate	No	No	Yes
54	Crude Death Rate	No	No	Yes
55	Infant Mortality Rate	No	No	Yes
56	Population, 2011	Yes	Yes	Yes
57	Area in squared Kilometre, 2011	No	No	Yes
58	Population Density, 2011	No	No	Yes

Notes: UPS indicates usual participation status.

ANNEX VI: DETAILED ANALYSIS OF COMPOSITION OF LIZS IN UNWEIGHTED INDEX METHOD AND HIERARCHICAL CLUSTER METHOD FOR RURAL AREAS

In this Annex, detailed analysis of results from the unweighted index and Hierarchical cluster methods are presented one after another for rural areas. The next Annex provides similar results for urban areas. A detailed analysis of results for K-means cluster method has already been presented in the main text. The reason for presenting these two other analyses in this annex is because although this paper relies mainly on results of the K-means method to identify living income zones (LIZ), results from the two other methods help inform results from the K-means method as well as being interesting in their own right.

Unweighted Index Method in Rural Areas

Annex Table 1 presents the composition of each cluster in terms of specific NSS regions along with their location in states or union territories and specific geographically contiguous zones. Annex Map 1 gives geographical composition of each of 10 clusters. The composition of clusters 1 and 2 and also clusters 9 and 10 have already been discussed. Cluster 3 consists of 12 NSS regions, which are spread over comparatively large number of states stretching from West Bengal of Eastern India to North Indian states of Uttar Pradesh and Jammu & Kashmir and further to Central Indian states of Madhya Pradesh, Western Indian state of Maharashtra and northern part of South Indian states of Karnataka.

Cluster 4 includes an even larger number of states mostly belonging to all regions except the South Indian states. It also contains two geographically contiguous zones. The first one, containing six rural NSS regions, spreads over three states of Orissa, Chhattisgarh and Madhya Pradesh belonging to Eastern and Central part of India. The second geographically contiguous zone includes three NSS regions of Eastern and North-Eastern states of West Bengal and Assam.

Cluster 5 contains 16 NSS regions spreading over different part of India but 9 of these regions are in Southern India. Out of these 9 regions, eight regions cover the whole state of Andhra Pradesh, and Telangana and most of Karnataka state is geographically contiguous. In addition, two regions of North-East India consisting of state of Arunachal Pradesh and one region of Assam is also geographically contiguous.

Cluster 6 also comprises16 NSS regions. These are mostly from Northern, Western and North-Eastern India. There is one contiguous large region spreading over the mountainous state of Uttarakhand in Northern India, agriculturally prosperous states of Haryana and Punjab and further extending to some part of Rajasthan in Western India; this contiguous region contains 7 NSS regions. However, weather and food habits are comparatively more diverse. Apart from this, there are 2 contiguous regions that contain 2 NSS regions each belonging to Central Indian states of Chhattisgarh and Southern Indian state of Tamil Nadu.

Clusters 7 and 8 include 7 and 6 NSS regions respectively, but they are distantly located and only 2 NSS regions in cluster 8 are geographically contiguous. But one of them is less important because it includes the union territory of Chandigarh with negligible rural populace.

In a nutshell, rural India is very diverse according to the unweighted index method. This analysis shows that economic, socio-cultural and demographic development captured by the unweighted index method is spread out across all geographical areas in the country. Still in low and moderately developed clusters (up to cluster 6), the level of development is similar in broader geographically contiguous regions. However, the situation is different in more developed regions (cluster 7 to cluster 10), which are geographically more diverse.

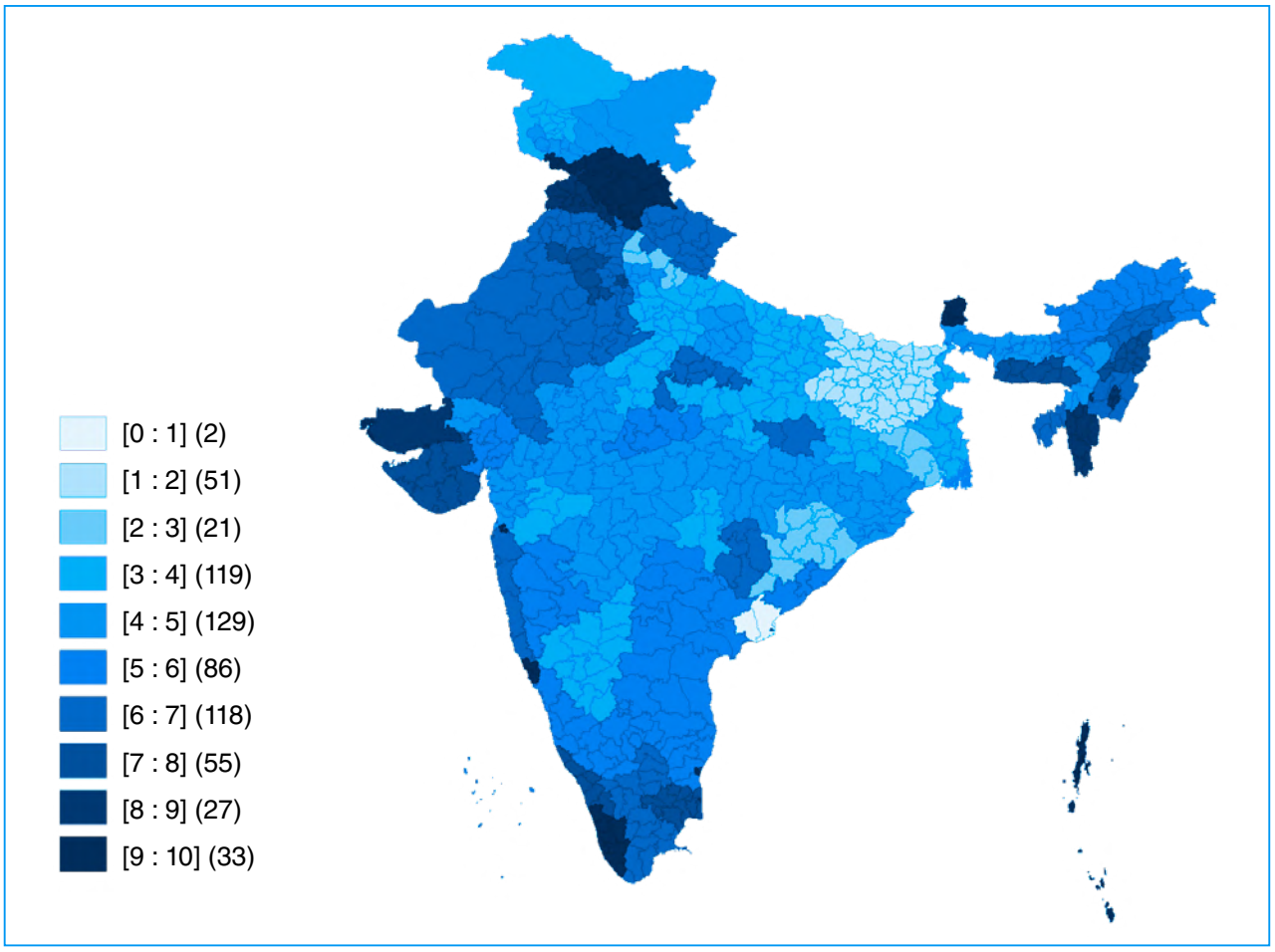
ANNEX VI Table 1: Composition of Clusters in Rural India in Unweighted Index Methodology

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
1	1	Bihar	Northern, Central Plain (1)	0.201
	1	Jharkhand	Hazaribagh Plateau (1)	0.213
2	2	West Bengal	Western Plain (2)	0.255
	2	Orissa	Southern (2)	0.247
	3	Uttar Pradesh	North Upper Ganga Plain (2)	0.253
3	4	Jammu & Kashmir	Jhelum Valley (3)	0.275
	5	Uttar Pradesh	Eastern, South Upper Ganga Plain (3)	0.278
	6	Madhya Pradesh	Vindya, Northern (3)	0.293
	7	West Bengal	Eastern Plains, Southern Plain, Central Plain (3)	0.288
	8	Jharkhand	Ranchi Plateau (3)	0.291
	9	Maharashtra	Inland Northern (3)	0.294
	10	Karnataka	Inland Northern (3)	0.273
4	11	Jammu & Kashmir	Outer Hills, Ladakh (4)	0.319
	12	Rajasthan	South-Eastern (4)	0.325
	13	Uttar Pradesh	Central (4)	0.329
	14	Assam	Plain Western, Cachar Plain (4)	0.311
	14	West Bengal	Himalayan (4)	0.301
	15	Orissa	Coastal, Northern (4)	0.312
	15	Chhattisgarh	Mahanadi Basin (4)	0.326
	15	Madhya Pradesh	Malwa, South, South Western (4)	0.311
	16	Gujarat	Dry Areas (4)	0.326
	17	Gujarat	South-Eastern (4)	0.333
	18	Maharashtra	Inland Central, Inland Eastern (4)	0.321

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
5	19	Arunachal Pradesh	All (5)	0.365
	19	Assam	Central Brahmaputra Plain (5)	0.339
	20	Madhya Pradesh	Central (5)	0.343
	21	Gujarat	Plain Northern (5)	0.359
	22	Daman & Diu	All (5)	0.359
	23	Maharashtra	Inland Western (5)	0.350
	24	Andhra Pradesh	Coastal Northern, Coastal Southern, Inland Southern (5)	0.349
	24	Telengana	Inland North-Western, Inland North-Eastern (5)	0.356
	24	Karnataka	Coastal & Ghats, Inland Eastern, Inland Southern (5)	0.344
	25	Lakshadweep	All (5)	0.348
	26	Tamil Nadu	Coastal Northern (5)	0.368
6	27	Punjab	Southern (6)	0.380
	27	Uttarakhand	All (6)	0.370
	27	Haryana	Eastern (6)	0.382
	27	Rajasthan	Western, North-Eastern, Southern, Northern (6)	0.384
	28	Uttar Pradesh	Southern (6)	0.374
	29	Manipur	Hills (6)	0.376
	30	Tripura	All (6)	0.371
	31	Assam	Plain Eastern (6)	0.386
	32	Chhattisgarh	Northern, Southern (6)	0.384
	33	Maharashtra	Coastal (6)	0.380
	34	Tamil Nadu	Southern, Inland (6)	0.391

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
7	35	Haryana	Western	0.413
	36	Delhi	All (7)	0.432
	37	Nagaland	All (7)	0.418
	38	Meghalaya	All (7)	0.406
	39	Gujarat	Saurastra (7)	0.429
	40	Kerala	Northern (7)	0.431
	41	Tamil Nadu	Coastal (7)	0.407
8	42	Punjab	Northern	0.469
	42	Chandigarh	All	0.468
	43	Manipur	Plains	0.439
	44	Mizoram	All	0.448
	45	Gujarat	Kachchh	0.446
	46	Puduchery	All	0.469
9	47	Jammu & Kashmir	Mountainous	0.506
	47	Himachal Pradesh	Trans Himalayan & Southern	0.485
	48	Sikkim	All	0.492
	49	Dadra & Nagar Haveli	All	0.504
10	50	Himachal Pradesh	Central Plain	0.520
	51	Goa	All	0.509
	52	Kerala	Southern	0.526
	53	Andaman & Nichobar Island	All	0.542

ANNEX VI Map 1: Unweighted Index Method Living Income Clusters, Rural India (number of districts in each cluster in brackets)



Note: The unweighted index method clusters are presented in order of overall achievement in terms of chosen variables representing economic, socio-cultural and demographic factors/characteristics. Therefore, first cluster is ranked lowest and last cluster is ranked highest in that order.

Hierarchical Cluster Method in Rural Areas

In the 1st cluster (that contains 29 NSS regions), one large geographically contiguous zone (2nd) contains 25 NSS regions. This geographical zone contains 5 states (Rajasthan, Gujarat, Punjab, Haryana and Madhya Pradesh), one union territory of Chandigarh, and four out of five NSS regions of Uttar Pradesh state with a huge landmass of northern, central and western India. Another large geographical zone (14th) in the 5th cluster covers three states (Bihar, Jharkhand and Orissa) of eastern India and one region of Chhattisgarh state of central India. The number of NSS regions is eight in this geographically contiguous zone but it has huge mass of rural population of over 200 million.

The 2nd and 4th clusters (original cluster naming by hierarchical method) are mostly hilly and mountainous regions in North-East, Central and North-Western India except for Goa that is in 2nd cluster and is a coastal state. That 8th cluster that comprises three regions of West Bengal of Eastern India and one region of Tamil Nadu. These regions are densely populated and undertake extensive rice cultivation.

In short, unlike in unweighted index method, in this Hierarchical method the number of NSS regions assigned to each cluster can be divided into two groups: 2 clusters with minimum of 25 NSS regions and in the remaining 8 clusters none of them contain even 10 NSS regions. Again, the largest geographically contiguous zone includes 25 NSS regions covering a huge landmass of northern, central and western India. In contrast, 2 clusters possess 2 NSS regions each and 1 cluster has a single NSS region.

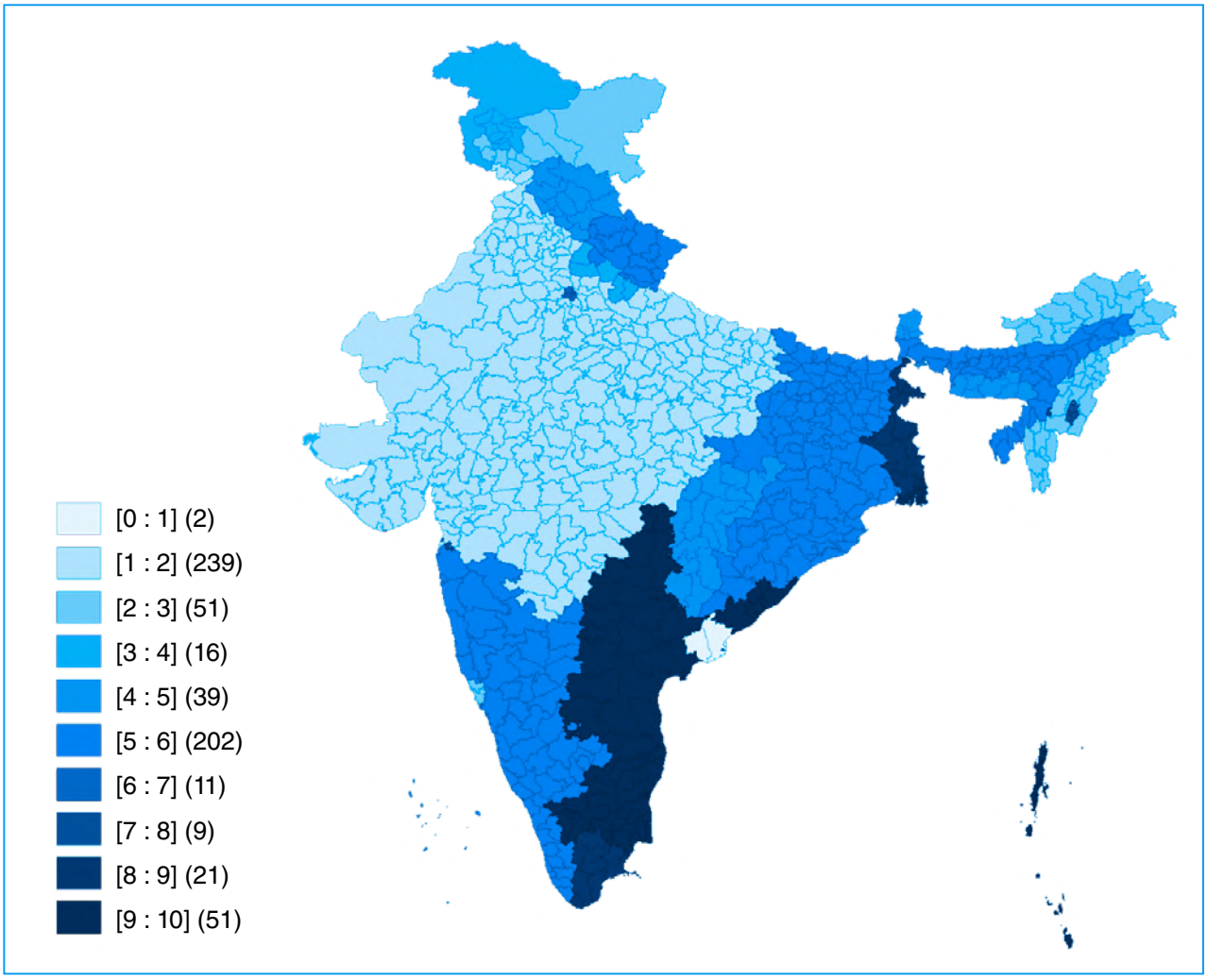
ANNEX VI Table 2: Composition of Living Income Clusters in Rural India in Hierarchical Methodology

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
1st Cluster, unweighted value= 0.399	1	Jammu & Kashmir	Mountainous (1)	0.506
	2	Punjab	Northern, Southern (1)	
	2	Chandigarh	All (1)	
	2	Haryana	Eastern, Western (1)	
	2	Rajasthan	Western, North-Eastern, Southern, South-Eastern, Northern (1)	
	2	Gujarat	South-Eastern, Plain Northern, Dry Areas, Kachchh, Saurash-tra (1)	
	2	Madhya Pradesh	Vindya, Central, Malwa, South, South-Western, Northern (1)	
	2	Uttar Pradesh	Central, Eastern, South Upper Ganga Plain, Southern (1)	0.381
	3	Maharashtra	Inland Northern, Inland Central, Inland Eastern (1)	0.311
2nd cluster, unweighted value = 0.356	4	Jammu & Kashmir	Outer Hills, Ladak (2)	0.319
	5	Arunachal Pradesh	All (2)	
	5	Nagaland	All (2)	
	5	Manipur	Hills (2)	
	5	Mizoram	All (2)	0.402
	6	Goa	All (2)	0.348
3rd cluster, unweighted value = 0.264	7	Jammu & Kashmir	Jhelam Valley (3)	0.275
	8	Uttar Pradesh	North Upper Ganga Plain (3)	0.253

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
4th cluster, unweighted value = 0.435	9	Himachal Pradesh	Central Plain, Southern (4)	0.502
	10	Sikkim	All (4)	0.492
	11	Meghalaya	All (4)	0.406
	12	Chhattisgarh	Mahanadi Basin, Southern (4)	0.355
5th cluster, unweighted value = 0.351	13	Uttarakhand	All (5)	0.370
	14	Bihar	Northern Plain, Southern Plain (5)	
	14	Jharkhand	Ranchi Plateau, Hazaribagh Plateau (5)	
	14	Orissa	Coastal, Southern, Northern (5)	
	14	Chhattisgarh	Northern (5)	0.282
	15	Assam	Plain Eastern, Plain Western, Cachar Plain, Central Brahma-putra Plain (5)	
	15	Tripura	All (5)	
	15	West Bengal	Himalayan (5)	0.319
	16	West Bengal	Western Plain (5)	0.255
	17	Karnataka	Coastal Ghat, Inland Eastern, Inland Southern, Inland North-ern (5)	
	17	Kerala	Northern, Southern (5)	
	17	Maharashtra	Coastal, Inland Western (5)	0.400
	18	Lakshadweep	All (5)	0.479
6th cluster, unweighted value = 0.396	19	Delhi	All (6)	0.432
	20	Daman & Diu	All (6)	0.359
7th cluster, unweighted value = 0.471	21	Manipur	Plain (7)	0.439
	22	Puduchery	All (7)	0.469
	23	Dadra & Nagar Haveli	All (7)	0.504
8th cluster, unweighted value = 0.339	24	West Bengal	Eastern, Plain, Southern Plain, Central Plain (8)	0.288
	25	Tamil Nadu	Southern (8)	0.389

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
9th cluster, unweighted value = 0.352	26	Maharashtra	Eastern (9)	
	26	Andhra Pradesh	Coastal Northern, Coastal Southern, Inland Southern (9)	
	26	Telangana	Inland North-Western, Inland North Eastern (9)	0.342
	27	Tamil Nadu	Coastal Northern, Coastal, Inland (9)	0.389
10th cluster, unweighted value =0.542	28	Andaman & Nichobar	All (10)	0.542

ANNEX VI Map 2: Hierarchical Method Living Income Clusters, Rural (number of districts in each cluster in brackets)



Note: Cluster numbers (from 1 to 10) do not indicate any ranking in Hierarchical method. It shows dissimilarity between different clusters.

ANNEX VII: DETAILED ANALYSIS OF COMPOSITION OF LIVING INCOME CLUSTERS IN UNWEIGHTED INDEX METHOD AND HIERARCHICAL CLUSTER METHOD FOR URBAN AREAS

In this annex, detailed analysis of composition of unweighted index method and Hierarchical cluster method is presented one after another for the urban areas of India. A similar detailed analysis of K-means cluster has already been presented in the main text of this paper.

Unweighted Index Method

Annex Table 3 presents the composition of each cluster in terms of specific NSS regions along with their location in state or union territories and specific geographically contiguous zones. Annex Map 3 gives geographical composition of each of 10 clusters.

The largest number of contiguous NSS regions is in contiguous zone 27 in the 5th cluster that encompasses 11 NSS regions including two southern states of Andhra Pradesh and Telangana that were part of the same state for several decades reflecting uniformity in the development of towns and cities across all these NSS regions. In the contiguous zone, another 5 NSS regions are part of the geographically contiguous states of Orissa and Chhattisgarh but they were never part of the same administrative structure. Additionally, 1 NSS region of Maharashtra state is also part of this contiguous zone. Moreover, 5th cluster contains another two geographically contiguous zones. One lies in North-Eastern India with three NSS regions located in three states of Arunachal Pradesh, Nagaland and Assam. Another zone is in Western part of India having two NSS regions of states of Rajasthan and Gujarat.

The top three clusters (8th to 10th) effectively have one contiguous zone that include Delhi with Eastern Haryana which is basically an extension of Delhi Metropolis where substantial part of linked manufacturing and service activities are located in cities of Gurgaon and Sonipat.

Apart from Delhi metropolis in the 8th cluster, there are other large metropolises and industrial areas. This 8th cluster contains coastal region of Maharashtra that has Mumbai metropolis along with large industrial suburb of Thane. Also in this cluster is Inland Southern region of Karnataka that includes Bangalore metropolis along with its suburb and the city of Mysore. Apart from them, 8th cluster also has South-Eastern region of Gujarat state that has two large industrial cities of Vadodara and Surat. However, another large metropolis of Kolkata in the state of West Bengal in eastern India is in the 5th cluster reflecting lower level of organised manufacturing and service activity in the city and its suburban areas.

The 7th cluster contains two NSS regions of Southern state of Tamil Nadu. Coastal Northern region include city of Chennai along with its suburb and Inland region of Tamil Nadu that includes industrial towns of Coimbatore and Tirupur. This cluster also includes all two NSS regions of the state of Punjab located in Northern India. Punjab is a state whose urban development has been led by small and medium sized towns and its small and medium sized industrial activity is spread out all over the state with concentration of activities in the city of Ludhiana located in Southern NSS region of the state. This cluster also includes the tourism dependent state of Goa and union territory of Puducherry. In this cluster, there are two contiguous zones containing three NSS regions each in Northern and Southern India but spreading over more than one state.

The 4th cluster includes 18 NSS regions, but it has several geographically contiguous zones. Four NSS regions of western state of Maharashtra are in one zone. Another zone has 4 NSS regions including three regions of Central Indian state of Madhya Pradesh and South-Eastern region of Western state of Rajasthan.

Even in the 3rd cluster, there are several contiguous regions but most of these belong to the same state – Assam in North-East India and Uttar Pradesh in Northern India.

In a nutshell, major metropolitan areas and industrial zones of India except for Kolkata are in 7th and 8th clusters. Cluster 10 contains small uniform union territories and cluster 9 contains regions with extensive tourism activities. In contrast, NSS regions with very low level of urbanisation and industrial activities are located in clusters 1 and 2.

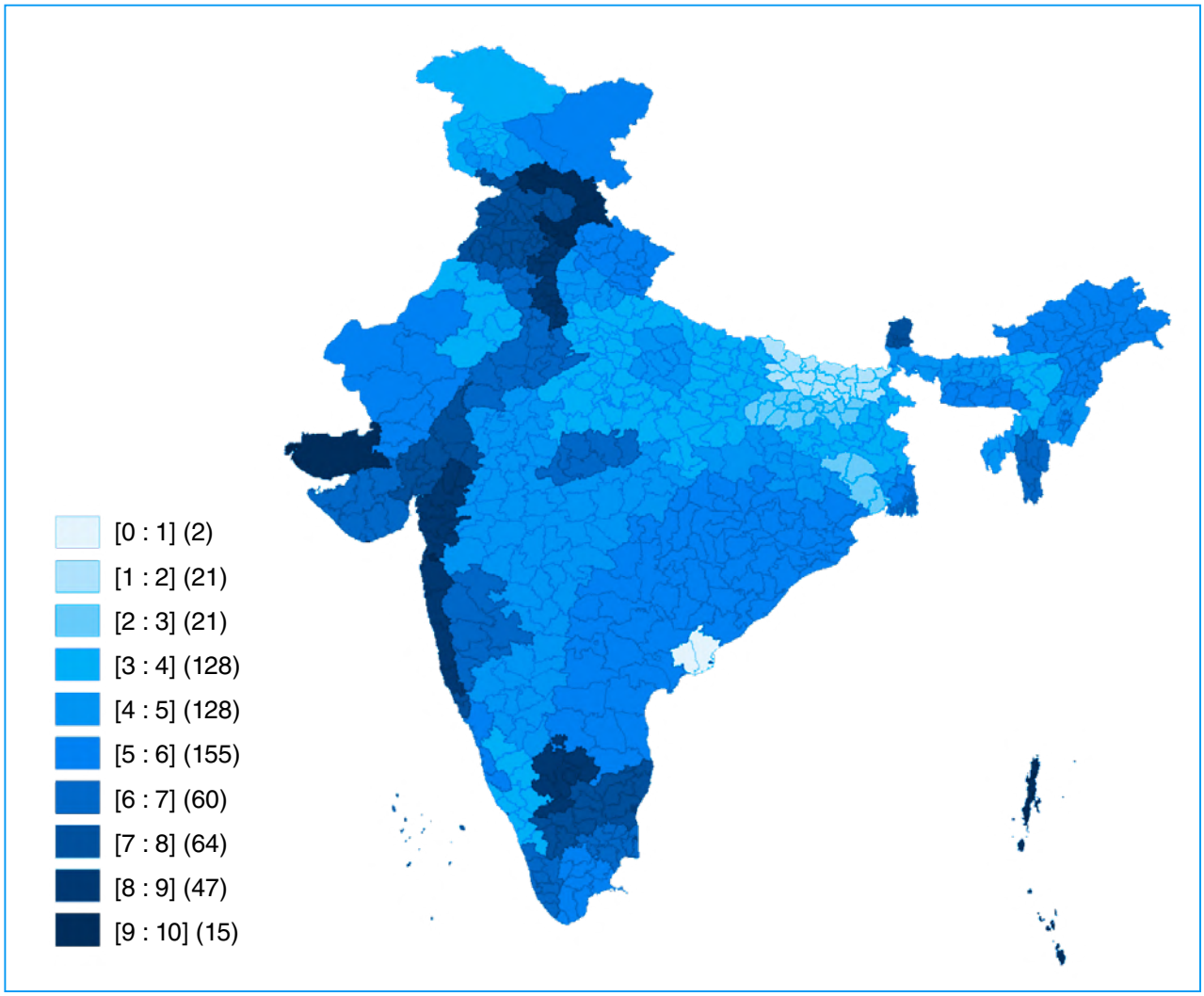
ANNEX VII Table 1: Composition of Living Income Clusters in Urban India in Unweighted Index Methodology

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
1	1	Bihar	Northern (1)	0.105
2	2	Bihar	Central Plain (2)	0.191
	3	West Bengal	Western Plain (2)	0.168
3	4	Rajasthan	Northern (3)	0.235
	5	Uttar Pradesh	Eastern, Southern, Southern Upper Ganga (3)	0.227
	6	Assam	Cachar Plain, Central Brahmaputra Plain (3)	0.229
	7	West Bengal	Eastern Plain (3)	0.327
	8	Jharkhand	Hazaribagh Plateau (3)	0.231
	9	Madhya Pradesh	Vindya, Northern (3)	0.230
	10	Karnataka	Inland Eastern (3)	0.232
	10	Kerala	Northern (3)	0.244
4	11	Jammu & Kashmir	Outer Hills (4)	0.291
	12	Uttar Pradesh	Northern Upper Ganga Plain, Central (4)	0.295
	13	Manipur	Hills (4)	0.258
	14	Tripura	All (4)	0.303
	15	West Bengal	Himalayan, Central Plain (4)	0.270
	15	Assam	Plains Western (4)	0.272
	16	Jharkhand	Ranchi Plateau (4)	0.294
	16	Chhattisgarh	Northern (4)	0.236

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
	17	Rajasthan	South-Eastern	0.260
	17	Madhya Pradesh	Malwa, South, South-Western (4)	0.290
	18	Maharashtra	Inland Northern, Inland Central, Inland Eastern (4)	0.270
	19	Karnataka	Inland Northern (4)	0.267
5	20	Jammu & Kashmir	Jhelum Valley (5)	0.228
	21	Uttarakhand	All (5)	0.350
	22	Rajasthan	Western (5)	0.356
	22	Gujarat	Dry Areas (5)	0.322
	23	Arunachal Pradesh	All (5)	0.339
	23	Nagaland	All (5)	0.308
	23	Assam	Plains Eastern (5)	0.327
	24	Manipur	Plains (5)	0.333
	25	Meghalaya	All (5)	0.320
	26	West Bengal	Southern Plain (5)	0.338
	27	Orissa	Coastal, Northern, Southern (5)	0.321
	27	Chhattisgarh	Mahanadi Basin, Southern (5)	0.331
	27	Maharashtra	Eastern (5)	0.336
	27	Andhra Pradesh	Coastal Northern, Coastal Southern, Inland Southern (5)	0.354
	27	Telangana	Inland North-Western, Inland Eastern (5)	0.328
	28	Karnataka	Coastal & Ghats (5)	0.330
	29	Tamil Nadu	Southern (5)	0.350
6	30	Haryana	Western (6)	0.360
	30	Rajasthan	North-Eastern (6)	0.370
	31	Mizoram	All (6)	0.391
	32	Madhya Pradesh	Central (6)	0.391
	33	Gujarat	Saurastra (6)	0.383

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
	34	Maharashtra	Inland Western (6)	0.386
	35	Kerala	Southern (6)	0.386
	36	Tamil Nadu	Coastal (6)	0.372
7	397	Rajasthan	Southern (7)	0.448
	37	Gujarat	Plain Northern (7)	0.436
	38	Goa	All (7)	0.448
	39	Lakshadweep	All (7)	0.411
	40	Tamil Nadu	Coastal Northern, Inland (7)	0.473
	40	Puducherry	All (7)	0.419
	41	Jammu & Kashmir	Mountainous, Ladakh (7)	0.384
	42	Punjab	Northern, Southern (7)	0.414
	42	Himachal Pradesh	Central Plain (7)	0.425
	43	Sikkim	All (7)	0.419
8	44	Delhi	All (8)	0.488
	44	Haryana	Eastern (8)	0.503
	45	Gujarat	South-Eastern (8)	0.480
	46	Gujarat	Kachchh (9)	0.511
	47	Karnataka	Inland Southern (8)	0.479
	48	Maharashtra	Coastal (8)	0.500
9	49	Himachal Pradesh	Trans Himalayan & Southern (9)	0.523
	50	Andaman & Nichobar Island	All (9)	0.548
10	51	Daman & Diu	All (10)	0.599
	51	Dadra & Nagar Haveli	All (10)	0.571
	52	Chandigarh	All (10)	0.606

ANNEX VII Map 1: Unweighted Index Method Living Income Clusters, Urban India



Note: Only clusters in the unweighted Index method are presented in order of overall achievement in terms of chosen variables representing economic, socio-cultural and demographic factors/characteristics. Therefore, first cluster is ranked lowest and last cluster is ranked higher in that order. Hierarchical and K-means clusters cannot be ranked.

Hierarchical Method Cluster in Urban Areas

Annex Table 4 presents the composition of each cluster in terms of specific NSS regions along with their location in state or union territories and specific geographically contiguous zones. Annex Map 4 gives geographical composition of each of 10 clusters.

We begin by analysing smaller clusters. Cluster 4 has 3 observations. The first two clusters are in mountainous state of Himachal Pradesh in Northern India and the other one is region of Kachchh in the Western state of Gujarat. Cluster 5 has two NSS regions and these are Capital Delhi and Chandigarh, joint capital of states of Punjab and Haryana in Northern India. These are mostly urban regions with high concentration of organised service sector activities.

The 7th, 9th and 10th clusters are single NSS region clusters. The 7th cluster contains coastal region of Western India and the other two clusters (9th and 10th) have one island each located far away from mainland and located in the Arabian sea and Bay of Bengal respectively.

Coming back to large clusters (1st to 3rd), the 1st geographically concentrated zone in cluster 1 contains 4 states (Rajasthan, Punjab, Haryana and Madhya Pradesh), four out of five NSS regions of Uttar Pradesh state and three out of five NSS regions of Gujarat with a huge landmass of northern, central and western India. The 4th geographically concentrated zone in cluster 2 contains two states (Bihar and Jharkhand) of eastern India and one region of Chhattisgarh and West Bengal states and two out of three NSS regions of Orissa state. The 15th geographically contiguous zone in cluster 3 includes four large states of southern India (Andhra Pradesh, Telangana, Tamil Nadu and Kerala), two out of four NSS regions of Karnataka state, two out of six NSS regions of Maharashtra state and two small state and union territory (Goa and Puducherry). The 8th cluster includes union territories of Daman & Diu and Dadra & Nagar Haveli in Western India that are mostly urbanised and have large concentration of organised manufacturing workers.

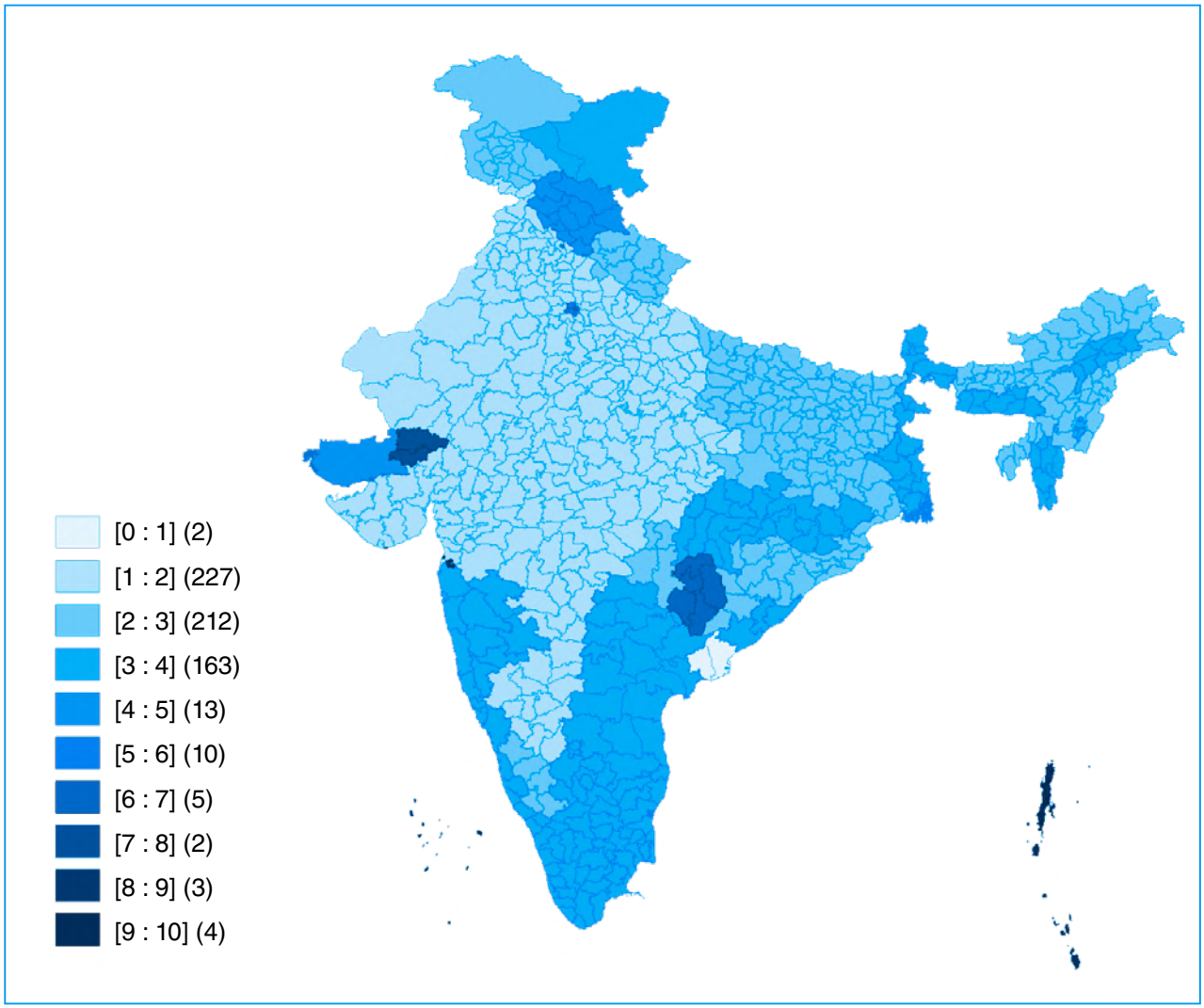
ANNEX VII Table 2: Composition of Living Income Clusters in Urban India in Hierarchical Methodology

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
1st Cluster, unweighted value= 0.319	1	Jammu & Kashmir	Mountainous (1)	
	1	Punjab	Northern, Southern (1)	
	1	Haryana	Eastern, Western (1)	
	1	Rajasthan	Western, North-Eastern, Southern, South-Eastern, Northern (1)	
	1	Gujarat	South-Eastern, Plain Northern, Saurashtra (1)	
	1	Uttar Pradesh	North Upper Ganga Plain, Central, Southern, South Upper Ganga Plain (1)	
	1	Madhya Pradesh	Vindya, Central, Malwa, South, South-Western, Northern (1)	0.369
	2	Karnataka	Inland Northern (1)	
	2	Maharashtra	Inland Northern, Inland Central, Inland Eastern (1)	0.269
2nd Cluster, unweighted value= 0.282	3	Jammu & Kashmir	Outer Hills, Jhelum Valley (2)	0.260
	4	Bihar	Northern Plain, Central Plain (2)	
	4	Uttar Pradesh	Eastern (2)	
	4	West Bengal	Western Plain (2)	

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
	4	Jharkhand	Ranchi Plateau, Hazaribagh Plateau (2)	
	4	Orissa	Coastal, Southern (2)	
	4	Chhattisgarh	Northern (2)	0.232
	5	Uttarakhand	All (2)	0.350
	6	Arunachal Pradesh	All (2)	
	6	Nagaland	All (2)	
	6	Manipur	Hills (2)	
	6	Assam	Plains Western, Cachar Plain, Brahmaputra Plain (2)	
	6	Tripura	All (2)	0.290
	7	Karnataka	Inland Eastern (2)	0.232
	7	Maharashtra	Eastern (2)	0.336
	8	Jammu & Kashmir	Ladakh (3)	
	9	Sikkim	All (3)	
	9	West Bengal	Himalayan, Eastern Plain, Southern Plain, Central Plain (3)	0.349
	10	Manipur	Plains (3)	0.333
	11	Mizoram	All (3)	0.391
3rd Cluster, unweighted value= 0.349	12	Assam	Plains Eastern (3)	0.327
	13	Meghalaya	All (3)	0.32
	14	Orissa	Northern (3)	
	14	Chhattisgarh	Mahanadi Basin (3)	0.325
	15	Maharashtra	Coastal, Inland Western (3)	
	15	Goa	All (3)	
	15	Karnataka	Coastal & Ghat, Inland Southern (3)	
	15	Kerala	Northern, Southern (3)	

Cluster	Geographically Contiguous Zones	State	NSS Regions	Unweighted Index Value
	15	Tamil Nadu	Coastal Northern, Coastal, Inland, Southern (3)	
	15	Puducherry	All (3)	
	15	Andhra Pradesh	Coastal Northern, Coastal Southern, Inland Southern (9)	
	15	Telangana	Inland North-Western, Inland North-Eastern (9)	0.399
4th Cluster, unweighted value= 0.425	16	Himachal Pradesh	Central Plain, Trans Himalayan & Southern (4)	0.474
	17	Gujarat	Kachchh (4)	0.511
5th Cluster, unweighted value= 0.547	18	Delhi	All (5)	0.488
	19	Chandigarh	All (5)	0.606
6th Cluster, unweighted value= 0.337	20	Chhattisgarh	Southern (6)	0.337
7th Cluster, unweighted value= 0.322	21	Gujarat	Dry Areas (7)	0.322
8th Cluster, unweighted value= 0.585	22	Daman & Diu	All (8)	
	22	Dadra & Nagar Haveli	All (8)	0.585
9th Cluster, unweighted value= 0.411	23	Lakshadweep	All (9)	0.411
10th Cluster, unweighted value= 0.548	24	Andaman & Nichobar Island	All (10)	0.548

ANNEX VII Map 2: Hierarhical Method Living Income Clusters, Urban India4



Note: Cluster numbers (from 1 to 10) do not indicate any ranking in Hierarchical method. It shows dissimilarity between different clusters.