**PROJECT REPORT**

**ON**

# "ASSESSMENT OF LABOUR FORCE PARTICIPATION, LITERACY RATE, WORKER POPULATION RATIO AND UNEMPLOYMENT RATE OF ALL THE STATES/U.T OF INDIA: A STATISTICAL STUDY USING R, PYTHON AND EXCEL AND DATA SOURCED FROM MoSPI"

**SUBMITTED IN THE PARTIAL FULFILMENT OF THE REQUIREMENT FOR**

**THE AWARD OF THE DEGREE OF**

**BACHELOR OF SCIENCE (HONS,)**

**IN STATISTICS**

**BY**


| | |
|---|---|
| MOHAMMAD ISMAIL | 21STBSA137 |
| MOHD FARHAN KHAN | 21STBSA157 |
| MOHD KAIF | 21STBSA196 |
| MOHD AHSAN | 21STBSA257 |
| MUNEER AHMAD | 21STBSA387 |
| ABHI PRANAY | 21STBSA440 |
| SUSHANT SARASWAT | 21STBSA443 |
| ZAINUL MEHNDI | 21STBSA454 |
| MOHD SUHAIL | 21STBSA459 |
| ADIL KHAN | 21STBSA460 |
| HAFIZ NAYAB AHMAD | 21STBSA464 |
| FARHAN RAZA | 21STBSA921 |


**UNDER THE SUPERVISION OF**


**PROF. AQUIL AHMED & DR. AIJAZ AHMAD DAR**

**DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH**

**ALIGARH MUSLIM UNIVERSITY, ALIGARH, UTTAR PRADESH, INDIA**


**2023-2024**

**DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH**
**ALIGARH MUSLIM UNIVERSITY, ALIGARH, U.P. 202002**

## CERTIFICATE

This is to certify that the project titled "*ASSESSMENT OF LABOUR FORCE PARTICIPATION, LITERACY RATE, WORKER POPULATION RATIO AND UNEMPLOYMENT RATE OF ALL THE STATES/U.T OF INDIA: A STATISTICAL STUDY USING R, PYTHON AND EXCEL AND DATA SOURCED FROM MoSPI*" has been submitted by the following group of students of *Bachelor of Science (Statistics), Final Year, 2023-2024,* under the supervision for the course "**STB-6S1-Project**".

| NAME | ROLL NUMBER |
|---|---|
| MOHAMMAD ISMAIL | 21STBSA137 |
| MOHD FARHAN KHAN | 21STBSA157 |
| MOHD KAIF | 21STBSA196 |
| MOHD AHSAN | 21STBSA257 |
| MUNEER AHMAD | 21STBSA387 |
| ABHI PRANAY | 21STBSA440 |
| SUSHANT SARASWAT | 21STBSA443 |
| ZAINUL MEHNDI | 21STBSA454 |
| MOHD SUHAIL | 21STBSA459 |
| ADIL KHAN | 21STBSA460 |
| HAFIZ NAYAB AHMAD | 21STBSA464 |
| FARHAN RAZA | 21STBSA921 |

**PROF. AQUIL AHMED** **DR. AIJAZ AHMAD DAR**

Dedicated to Late **Sir Syed Ahmad Khan**, the founder of Aligarh Muslim University, and our parents, and all the teachers who have given best of today for our best tomorrow.

# ABSTRACT

This project undertakes a comprehensive analysis of key economic and education indicators in India, focusing on the literacy rate, labour force participation rate, worker population ratio, and unemployment rate. Leveraging data from the Ministry of Statistics and Programme Implementation (MOSPI), we delve into the intricate dynamics shaping India's social and economic landscape.

By employing statistical analysis, we aim to uncover the multifaceted relationship between these indicators and various social and economic factors. Through rigorous examination at both national and regional levels, we seek to identify patterns, trends, and disparities across different states and zones of India.

Our research endeavours to elucidate the underlying drivers influencing these social and economic indicators, considering factors such as education and economic development. By adopting a multi-dimensional approach, we aspire to provide a nuanced understanding of the complex interplay between socio-economic dynamics and trends.

The outcomes of this study hold significant implications for policy formulation and strategic planning. Insights derived from our analysis can inform targeted interventions and policies tailored to address the specific needs and challenges faced by different regions of India. Moreover, our findings contribute to the broader discourse on socio-economic development in the country.

Through this project, we amalgamate quantitative analysis with empirical data sourced from MOSPI, grounding our findings in the rich context of India's diverse social and economic landscape. Ultimately, our research aims to foster informed decision-making and facilitate sustainable development initiatives tailored to promote inclusive growth across the nation.

The dissertation is consisting of three chapters; Chapter I compromises of introduction and various preliminaries and data sources used in the current study; Chapter II is devoted to the methodology used in this project work; Chapter III is consist of various statistical tools that are implemented along with the results that are interpreted in detail; the conclusion of the of the study is summed up and the recommendations are made accordingly.

# Acknowledgement

All praise and thanks are to God-The Almighty, the Merciful and Omniscient whose blessings enabled us to come across the completion of this work in present form. First of all, we would like to thank our supervisors, ***Prof. Aquil Ahmed*** and ***Dr. Aijaz Ahmad Dar (Assistant Professor)***, Department of Statistics & Operations Research, Aligarh Muslim University, Aligarh for their valuable time, constant support, cooperation and strong motivation. We are grateful to them for their generous help in finding out the errors and omissions in the preparation of this project. We owe a deep sense of gratitude to ***Prof. Athar Ali Khan,*** Chairman of the Department for having provided us with all the necessary facilities that were needed for completion of this work. Last but not least, we have no words in our command through which we can express our gratitude towards our parents, our well-wishers who always supported and extended a helping hand.

I. MOHAMMAD ISMAIL
II. MOHD FARHAN KHAN
III. MOHD KAIF
IV. MOHD AHSAN
V. MUNEER AHMAD
VI. ABHI PRANAY
VII. SUSHANT SARASWAT
VIII. ZAINUL MEHNDI
IX. MOHD SUHAIL
X. ADIL KHAN
XI. HAFIZ NAYAB AHMAD
XII. FARHAN RAZA

## Table of Content

# CHAPTER - 1

# INTRODUCTION

## 1.1 India & States/Union Territories of India

India's vast regional expanse encompasses a rich tapestry of topography, natural wonders, cultural traditions, linguistic diversity, and economic landscapes. This multifaceted diversity presents both opportunities and challenges for sustaining the country's unity and prosperity. From the towering Himalayas in the north to the sun-kissed beaches of the south, and from the bustling metropolises of Mumbai and Delhi to the serene backwaters of Kerala, India's geographical and cultural diversity is unparalleled.

States are autonomous administrative units with their own state governments, sharing governance responsibilities with the central government. Union territories, on the other hand, are directly governed by the central government, although some have their own local administrations with limited autonomy.

India comprises twenty-eight states and eight union territories. The states and union territories are further subdivided into districts and smaller administrative divisions.

**List of states:**

i. Andhra Pradesh

ii. Arunachal Pradesh

iii. Assam

iv. Bihar

v. Chhattisgarh

vi. Goa

vii. Gujarat

viii. Haryana

ix. Himachal Pradesh

x. Jharkhand

xi. Karnataka

xii. Kerala

xiii. Madhya Pradesh



*The content of this chapter has been prepared with the help of various resources, whose reference is given in the Reference section.

xiv. Maharashtra

xv. Manipur

xvi. Meghalaya

xvii. Mizoram

xviii. Nagaland

xix. Odisha

xx. Punjab

xxi. Rajasthan

xxii. Sikkim

xxiii. Tamil Nadu

xxiv. Telangana

xxv. Tripura

xxvi. Uttar Pradesh

xxvii. Uttarakhand

xxviii. West Bengal

**List of Union Territories:**

i. Andaman and Nicobar Islands

ii. Chandigarh

iii. Dadra and Nagar Haveli and Daman and Diu

iv. Delhi

v. Jammu and Kashmir

vi. Ladakh

vii. Lakshadweep

viii. Puducherry


## 1.2 Demography of India

India is a big country with lots of people and different cultures. There are over 1.3 billion people in India, and many of them are young. This means there are lots of young people who can work and do things. But having many young people also means there are challenges in finding jobs, getting education.

People in India speak many languages, with Hindi being the main one for the whole country. Each state can also have its own official languages. This diversity in languages adds to India's culture but can sometimes make things better.

According to the 2011 census of India, Hindus make up 79.8% of the population (966.3 million), Muslims 14.2% (172.2 million), Christians 2.3% (28.7 million), Sikhs 1.7% (20.8 million), Buddhists 0.7% (8.4 million), and Jains 0.4% (4.5 million). The Hindu population grew by 16.8% between 2001-2011, while the Muslim population grew by 24.6% during the same period. The share of Hindus in the total population declined from 81.4% in 2001 to 79.8% in 2011, while the share of Muslims increased from 13.4% to 14.2%.

India's population is spread between urban and rural areas, with many living in villages reliant on agriculture, despite a growing urban migration trend.

The government is committed to achieving balanced development across both rural and urban areas. Indian cities have transitioned from ancient agricultural hubs to contemporary urban centres, now grappling with issues such as inadequate planning and poor infrastructure. Government initiatives are focused on tackling these challenges to enhance urban living conditions.

In recent years, significant shifts have occurred in the way Indians live, learn, and work. These transformations are impacting family dynamics and societal structures. India is actively addressing these challenges while capitalizing on its youthful demographic to foster sustainable growth.

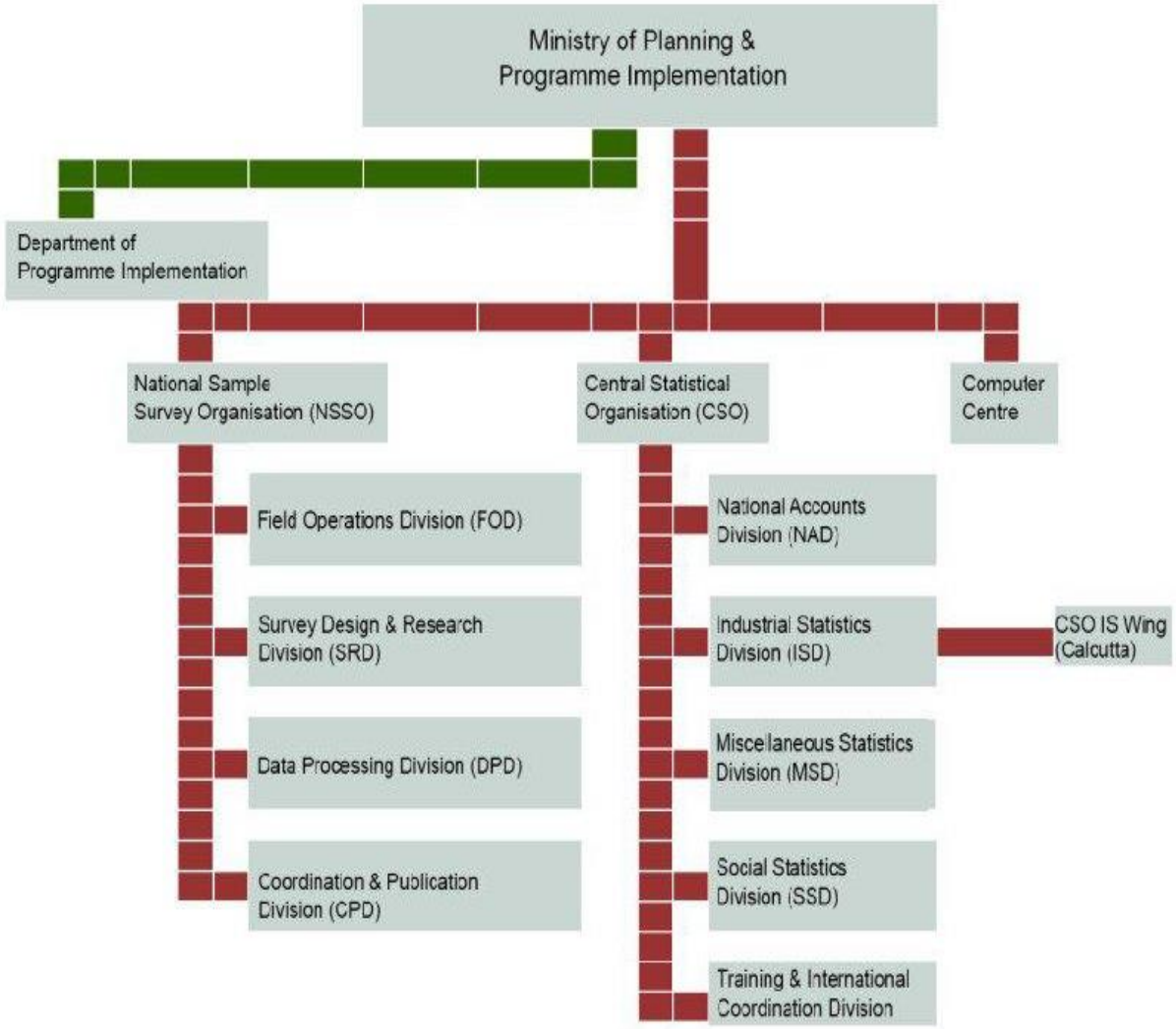
## 1.3 Ministry of Statistics and Programme Implementation (MoSPI)

The Ministry of Statistics and Programme Implementation (MoSPI) is a crucial government department responsible for ensuring the comprehensiveness and quality of statistical data dissemination in India. The Ministry's duties include conducting surveys using sophisticated sampling techniques to provide accurate and reliable data. MoSPI was officially established as an independent ministry on October 15, 1999, signifying its importance in the governmental framework.

### 1.3.1 Structure of MoSPI

MoSPI is divided into two main departments: the Department of Statistics and the Department of Programme Implementation. The Department of Statistics houses the National Statistical Office (NSO), which includes the Central Statistical Office (CSO), the Computer Centre, and the National Sample Survey Office (NSSO). These entities are responsible for collecting, compiling, and analysing statistical data to inform policy decisions.

On the Programme Implementation side, MoSPI is organized into three divisions: the Twenty Point Programme, which monitors key development initiatives; Infrastructure Monitoring and Project Monitoring, which tracks the progress of significant infrastructure projects; and the Member of Parliament Local Area Development Scheme, which oversees development projects funded by MPs. Additionally, MoSPI supervises the National Statistical Commission, an advisory body established through a Government of India resolution to promote statistical excellence, and manages the Indian Statistical Institute, an autonomous institute of national importance established by an Act of Parliament.




## i. National Sample Survey Office(NSSO)

The National Sample Survey Office (NSSO), headed by a Director General, conducts large-scale sample surveys across India in various fields. These surveys include nationwide household surveys on socio-economic topics, the Annual Survey of Industries (ASI), and data collection on rural and urban prices. NSSO also plays a crucial role in improving crop statistics by overseeing area enumeration and crop estimation surveys conducted by state agencies. Additionally, NSSO maintains a framework for urban area units used in urban sample surveys.

NSSO is organized into four divisions:

- Survey Design and Research Division (SDRD): Located in Kolkata, this division handles technical planning, survey concepts, sampling design,

inquiry schedule design, tabulation planning, and survey result analysis and presentation.

- Field Operations Division (FOD): Headquartered in Delhi/Faridabad with a network of offices nationwide, FOD collects primary data for NSS surveys.
- Data Processing Division (DPD): Based in Kolkata with additional centers, DPD manages sample selection, software development, data processing, validation, and tabulation for NSS surveys, including processing rural price and wage data and the Periodic Labour Force Survey (PLFS).
- Survey Coordination Division (SCD): Located in New Delhi, SCD coordinates NSSO's activities, publishes the bi-annual journal "Sarvekshana," and organizes national seminars on socio-economic survey results.

**ii. Central Statistical Office(CSO)**

This ministry serves as the central agency for the systematic development of statistical systems in the country, coordinating statistical activities among various stakeholders including the Government of India, State Governments, and international agencies. The Central Statistical Office (CSO) operates as the statistical arm of the Ministry, performing several key functions:

- Acting as the central agency for the planned development of the statistical system, setting and maintaining standards in statistics covering concepts, definitions, data collection methods, processing, and result dissemination.
- Coordinating statistical efforts across government ministries, State Statistical Bureaus (SSBs), advising on statistical methodology, and analysing data for government departments.
- Publishing the Index of Industrial Production (IIP) monthly as quick estimates, conducting the Annual Survey of Industries (ASI), and providing statistical insights into the organized manufacturing sector's growth and composition.
- Conducting nationwide Economic Censuses and follow-up enterprise surveys, processing data from socio-economic surveys, and assessing survey feasibility and sampling designs.

- Disseminating statistical information through various publications to government, semi-government, and private entities, and providing data to international agencies like the UNSD, ESCAP, ILO, and others upon request.
- Maintaining relationships with international statistical bodies such as UNSD, ESCAP, SIAP, IMF, ADB, FAO, ILO, etc.
- Compiling national accounts, publishing annual estimates of national product, consumption expenditure, capital formation, savings, and preparing comparable State Domestic Product (SDP) estimates.
- Conducting large-scale sample surveys to study socio-economic issues like employment, consumer expenditure, housing, literacy, health, nutrition, and family welfare.
- Providing grants to NGOs and research institutions for special studies, surveys, report printing, and financing seminars related to official statistics.

**iii. Computer centre**

The Computer Centre, under the leadership of an Additional Director General, is responsible for various functions:

The centre oversees the design, development, and upkeep of the Ministry's website, Data Catalogue, and CPI Web Portal. It ensures that these platforms adhere to best practices, including the International Household Survey Network (IHSN) framework and SDMX (Statistical Data and Metadata eXchange) guidelines where relevant.

Additionally, the Computer Centre plays a crucial role in disseminating reports and statistics generated by the Ministry, employing internationally recognized standards and practices.

Furthermore, the centre facilitates the use of NIC Cloud Services for hosting and maintaining the Ministry's applications and portals, ensuring efficient and reliable service delivery.

## 1.4 Periodic Labour Force Survey (PLFS)

One of the significant challenges in obtaining reliable employment and unemployment data in India has been addressed through the evolution of survey methodologies conducted by the National Sample Survey Office (NSSO).

Historically, the NSSO conducted Employment and Unemployment Surveys on a five-year basis, providing crucial statistics on the employment and unemployment situation in the country at five-year intervals. However, starting from April 1, 2017, the NSSO introduced a new survey framework known as the Periodic Labour Force Survey (PLFS), marking a pivotal shift in how employment and unemployment data are collected and analysed.

The PLFS represents a departure from the traditional quinquennial surveys in several key aspects. Firstly, it aims to provide more frequent and timely data by transitioning from a five-year cycle to annual and quarterly intervals. This shift is crucial as it allows policymakers and researchers to have access to more up-to-date information, enabling them to make informed decisions and interventions in the labour market.

Methodologically, the PLFS differs from its predecessor surveys in terms of survey methodology, data collection mechanisms, and sampling design. These changes are aimed at improving the accuracy and granularity of the employment and unemployment data captured through the survey.

The PLFS officially commenced its data collection from July 2017 onwards, with the first report covering the one-year period from July 2017 to June 2018. One of the notable aspects of the PLFS is its focus on providing quarterly employment and unemployment data, a feature that enhances the timeliness and relevance of the information for stakeholders across various sectors.

Overall, the adoption of the PLFS represents a significant advancement in the field of labour market statistics in India, offering a more dynamic and responsive framework for capturing and analysing employment and unemployment trends on a regular basis.

### 1.4.1 Objectives of Periodic labour force survey(PLFS)

The Periodic Labour Force Survey (PLFS) has been structured with two primary objectives:

**i.** To assess the fluctuations in labour force participation and employment status within a short timeframe of three months, specifically focusing on urban areas through the Current Weekly Status (CWS) approach.

**ii.** To provide quarterly updates on key labour force indicators, including the Worker Population Ratio (WPR), Labour Force Participation Rate (LFPR), and

Unemployment Rate (UR) for urban areas, thereby enabling a closer examination of trends and changes in these indicators over time.

Additionally, the PLFS aims to annually present estimates for both rural and urban areas, encompassing all critical parameters under both usual status and CWS methodologies. This comprehensive approach ensures that a wide range of labour market indicators is captured and reported with reasonable accuracy, facilitating a more nuanced understanding of the labour market dynamics at regular intervals throughout the year.

## 1.4.2 Unemployment and its types

The unemployment rate is a measure of the percentage of people in the labour force who are actively seeking work but are unable to find employment. It is calculated by dividing the number of unemployed individuals by the total labour force and multiplying by 100. This rate is a key economic indicator used to assess the health of an economy. High unemployment rates may indicate economic distress, while low rates suggest a thriving job market. The unemployment rate can be influenced by various factors, including economic conditions, technological changes, and government policies.

- **Types of Unemployment**

There are many types of Unemployment with context to PLFS.

**i. Frictional Unemployment:** This form of unemployment occurs when individuals are temporarily without work as they transition between jobs or seek better employment opportunities. It is often voluntary as individuals may be exploring new career paths or awaiting suitable job offers.

**ii. Structural Unemployment:** Structural unemployment arises due to a mismatch between the skills and qualifications of job seekers and the requirements of available job positions. This mismatch can be caused by technological changes, shifts in consumer demand, or changes in industry structures.

**iii. Cyclical Unemployment:** Cyclical unemployment is closely tied to fluctuations in the business cycle. During economic downturns or recessions, businesses may reduce their workforce, leading to increased unemployment rates. Conversely, economic expansions tend to create more job opportunities and lower unemployment rates.

**iv. Seasonal Unemployment:** This type of unemployment occurs in industries or occupations that experience fluctuations in demand based on seasonal factors. For instance, sectors like agriculture, tourism, or retail may see variations in employment levels throughout the year due to seasonal demand patterns.

**v. Underemployment:** Underemployment refers to a situation where individuals are employed but are not fully utilizing their skills, education, or experience in their current jobs. This can result from part-time work, temporary positions, or jobs that do not align with an individual's qualifications.

## 1.4.3 Parameters of PLFS

The major parameters of PLFS are given as:

- Labour Force Participation Rate (LFPR),
- Worker Population Ratio (WPR),
- Unemployment Rate (UR)

**i. Labour Force Participation Rate (LFPR)**

The Labour Force Participation Rate (LFPR) is a crucial measure used in labour economics to gauge the extent of workforce engagement among individuals of working age. It is determined by dividing the labour force, which encompasses both employed individuals and those actively seeking employment, by the total working-age population and then multiplying the result by 100 to express it as a percentage.

Mathematically, the LFPR can be expressed as follows:

$$LFPR = \left(\frac{Labour\ Force}{Working - age\ Population}\right) \times 100$$

In this formula:

- The "Labour Force" denotes the combined number of employed individuals and those actively looking for work.
- The "Working-age Population" represents the total count of individuals within a specified age range, typically between 15 and 64 years, who are deemed suitable for participation in the labour market.

By utilizing this formula, we derive a percentage value that signifies the proportion of the working-age population actively participating in economic activities. A higher LFPR indicates a greater level of workforce engagement, while a lower LFPR may indicate challenges such as unemployment or decreased workforce involvement.

LFPR data plays a pivotal role for policymakers, economists, and researchers as it facilitates the analysis of labour market trends, tracking changes in workforce participation patterns over time, and devising strategies to enhance employment prospects and foster economic growth.

**ii. Worker Population Ratio (WPR)**

The Worker Population Ratio (WPR) is a significant metric used in economics to determine the percentage of the working-age population that is currently employed. It is calculated by dividing the total number of employed individuals by the working-age population and then multiplying the result by 100 to express it as a percentage.

Expressed mathematically, the Worker Population Ratio can be stated as:

$$WPR = \left(\frac{Number\ of\ Employed\ Individuals}{Working - age\ Population}\right) \times 100$$

Here, the "Number of Employed Individuals" refers to the total count of people who are currently working, while the "Working-age Population" represents the total number of individuals within a specific age range (typically 15 to 64 years) who are deemed suitable for participation in the labour market.

The Worker Population Ratio offers insights into the employment level within a given population. A higher WPR indicates a larger proportion of the working-age population is employed, reflecting a healthy labour market and economic activity. Conversely, a lower WPR may indicate challenges such as unemployment or underemployment.

Policymakers, economists, and analysts rely on the Worker Population Ratio to analyse labour market dynamics, monitor employment trends, and develop strategies to promote job creation and foster economic development.

**(iii) Unemployment Rate (UR)**

The Unemployment Rate (UR) is a significant economic measure that indicates the percentage of unemployed individuals actively seeking work within the total labour force. It is computed by dividing the number of unemployed individuals by the total labour force and then multiplying the result by 100 to express it as a percentage.

Expressed mathematically, the Unemployment Rate can be defined as:

$$UR = \left(\frac{Number\ of\ Unemployed\ Individuals}{Total\ Labour\ Force}\right) \times 100$$

Here, the "Number of Unemployed Individuals" refers to individuals who are currently not employed but are actively searching for employment. The "Total Labour Force" encompasses both employed individuals and those seeking work.

The Unemployment Rate serves as a crucial gauge for assessing labour market conditions. A higher UR indicates a larger proportion of the labour force is unemployed, suggesting potential economic challenges such as limited job opportunities or economic downturns. Conversely, a lower UR indicates a smaller percentage of the labour force is unemployed, reflecting a more robust job market.

Policymakers, economists, and analysts rely on the Unemployment Rate to analyse trends in unemployment, evaluate labour market health, and develop strategies to address unemployment issues and promote job creation.

## 1.4.4 Activity Status

The activity status of an individual is determined by analysing their activities within a specified reference period. This assessment is particularly focused on the activities undertaken by the person over the last 365 days before the survey date, known as the usual activity status.

There are distinct types of activity statuses used for classification:

**i. Principal Activity Status (PS):** This status is attributed to the activity in which a person invests a substantial amount of time, meeting the major time criterion, during the 365-day period leading up to the survey date. It serves as the usual principal activity status for the individual.

**ii. Subsidiary Economic Activity Status (SS):** In addition to the usual principal activity status, if an individual engages in some form of economic activity for 30 days or more within the 365-day reference period, it is categorized as their subsidiary economic activity status.

**iii. Current Weekly Status (CWS):** This status is determined based on the person's activities over the last 7 days immediately preceding the survey date. It provides a snapshot of the individual's activity status for that particular week, known as the current weekly status.

## 1.5 Literacy rate

Literacy meaning the ability to read, write, and understand information effectively. It encompasses not only the basic skills of understanding written texts but also the capacity to comprehend, analyse, and communicate ideas. Literacy empowers individuals to access knowledge, participate in society, and express their views, opinions, and ideas. Literacy lays a crucial foundation for education, employment, and personal growth. By employing literacy skills in the younger generation, we can engage them in diverse texts, think critically, and contribute meaningfully to their communities, fostering a more inclusive and informed society.

The literacy rate refers to the percentage of the population of a certain age group who can read and write.

### 1.5.1 Types of Literacy Rate

Various types of literacy rate concerning age groups are elaborated as follows:

**i. Overall Literacy Rate:** This is the general literacy rate calculated across all age groups in a population. It represents the percentage of individuals, regardless of age, who possess basic reading and writing skills.

**ii. Youth Literacy Rate:** This refers to the literacy rate specifically among the youth population, usually defined as individuals aged 15 to 24 years. The youth literacy rate focuses on assessing literacy skills among younger individuals who are typically completing their formal education or transitioning into the workforce.

**iii. Adult Literacy Rate:** The adult literacy rate is calculated for individuals aged 15 years and older, excluding the youth population. It measures literacy skills among adults and is often used to evaluate the effectiveness of adult education programs and initiatives.

**iv. Elderly Literacy Rate:** This literacy rate focuses on individuals in the older age groups, typically those aged 65 years and above. It assesses literacy skills among older adults and considers factors such as lifelong learning opportunities and access to educational resources for seniors.

**v. Child Literacy Rate:** This refers to the literacy rate among children, usually those aged 5 to 14 years, who are in the process of acquiring basic reading and writing skills during their early education years.

**vi. Primary Literacy Rate:** The primary literacy rate assesses literacy skills among individuals in the primary school age group, typically aged 6 to 11 years. It focuses on foundational literacy skills essential for further education and lifelong learning.

**vii. Secondary Literacy Rate:** This literacy rate pertains to individuals in the secondary school age group, usually aged 12 to 17 years. It evaluates literacy skills among adolescents and teenagers as they progress through higher levels of education.

# CHAPTER-2

# RESEARCH METHODOLOGY

## 2.1 Objectives

The primary objective of this project is to assess the labour force using secondary data from the Periodic Labour Force Survey (PLFS) and the National Sample Survey Office (NSSO). The study focuses on four key variables: Literacy Rate (LR), Labour Force Participation Rate (LFPR), Unemployment Rate (UR), and Worker Population Rate (WPR). The analysis is conducted across different States and Union Territories, as well as demographic factors such as gender (male and female) and residence (urban and rural).

By analysing this data, our aim is to identify significant patterns in the labour force, uncover regional and demographic disparities and suggest targeted policies for improvement into the factors influencing labour force dynamics.

## 2.2 Target Population

The targeted population for this project comprises individuals sample surveyed. The dataset includes a comprehensive range of demographic factors, including States/Union Territories, gender (male and female) and residence (urban and rural). This secondary data source provides a detailed analysis across these factors, enabling a thorough examination of labour force participation and employment trends. By using this dataset, we aim to draw meaningful insights into the labour market's dynamics and variations between different population groups. The reliability and scope of this data ensure robust findings that can inform effective policymaking and contribute to the understanding of employment patterns in both urban and rural settings.

## 2.3 Data Source

In this study we are using secondary data. The data for this study was obtained from the Periodic Labour Force Survey (PLFS) and authorized by Ministry of Statistics and Program Implementation (MoSPI). The PLFS is a nationwide survey conducted by the National Sample Survey Office (NSSO) to collect information on India's labour market. The survey provides comprehensive data on various aspects of employment, unemployment, and labour force characteristics. The data is collected through a representative sample of households, making it a reliable and generalizable source. As secondary data, the PLFS offers advantages such as reduced data collection costs, increased efficiency, and access to a large dataset.

### 2.3.1 Significance of Data

This data holds significant importance for our project on labour force assessment. Firstly, these datasets provide a comprehensive and reliable source of information on key labour market indicators such as literacy rates, labour force participation, unemployment rates, and worker population rates. Secondly, the inclusion of demographic factors such as gender and urban/rural residence allows for a analysis of labour force dynamics across different population segments.

Moreover, being secondary data, it offers a cost-effective and efficient means of conducting extensive research without the need for primary data collection. Overall, the significance of these datasets lies in their ability to provide valuable insights into labour force characteristics.

### 2.3.2 Variables Under Study

i) Literacy Rate (LR)

ii) Labour Force Participation Rate (LFPS)

iii) Unemployment Rate (UR)

iv) Worker Population Rate (WPR)

## 2.4 Statistical Methods

### 2.4.1 Descriptive Statistics

### i) Mean

Arithmetic mean of a set of observations is their sum divided by the number of observations. In the case of group or continuous frequency distribution, X is taken as the midpoint of the corresponding class and the value of arithmetic mean is obtained by the following formula:

$$\mathrm{A.M} = \overline{\mathrm{X}} = \frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i}$$

### ii) Median

Median of distribution is the value of the variable which divides it into two equal parts. It is the va1ue which exceeds and is exceeded by the same number of observations, i.e., it is the value such that the number of observations above it is equal to the number of observations below it, the median is thus a positional average. In the case of group or continuous frequency distribution, the class corresponding to the cumulative frequency just greater than N/2 is called the median class and the value of median is obtained by the following formula:

$$Median = l_m + \frac{h_m}{f_m}\left(\frac{N}{2} - C_p\right)$$

Here,

$l_m$ = lower limit of the median class,

$f_m$ = frequency of the median class,

$h$ = class width

$C_p$ = cumulative frequency of the class preceding the median class**,**

N = sum of frequencies.

## iii) Mode

Mode is the value which occurs most frequently in a set of observations. In other words, mode is the value of the variable which is predominant in the series. In the case of discrete frequency distribution mode is the value of x corresponding to maximum frequency. In the case of group or continuous frequency distribution, Identify the modal class corresponding to the maximum frequency and the value of mode is obtained by the following formula:

$$Mode = l + \frac{f_m - f_p}{2f_m - f_p - f_s} \times h$$

Here,

l = lower limit of the modal class

h = class width

$f_m$ = frequency of the modal class

$f_p$ = frequency of the preceding class

$f_s$ = frequency of the successive class

## iv) Standard Deviation

Standard Deviation is a measure which shows how much variation (such as spread, dispersion) from the mean exists. Standard deviation, usually denoted by the Greek letter small sigma (σ), is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean.

The formula of standard deviation is:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Here,

$X_i$ = Each individual data points

$\bar{X}$ = Mean of the data points

N = Total number of data points

## v) Variance

Variance is the measure of how notably a collection of data is spread out. If all the data values are identical, then it indicates the variance is zero. All non-zero variances are considered to be positive. A little variance represents that the data points are close to the mean, and to each other, whereas if the data points are highly spread out from the mean and from one another indicates the high variance. In short, the variance is defined as the average of the squared distance from each point to the mean. The square of standard deviation is called the variance and is given by:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Here,

$X_i$ = Each individual data points

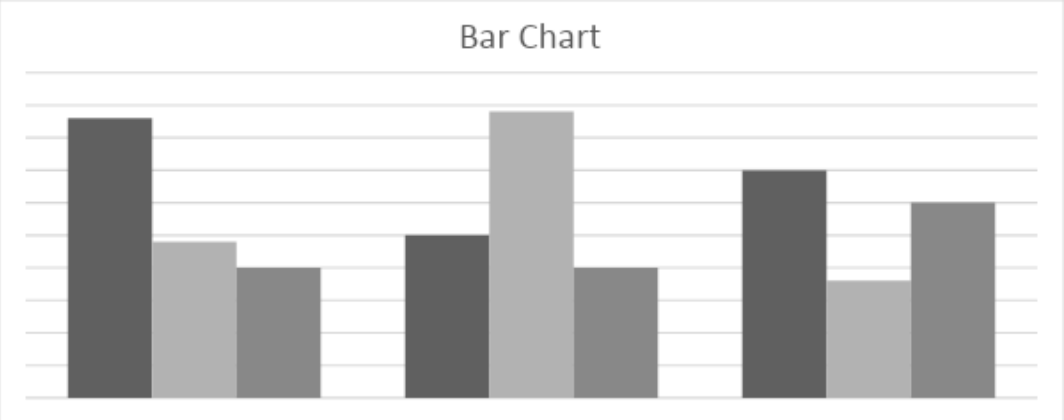
$\bar{X}$ = Mean of the data points

N = Total number of data points

## 2.4.2. Graphical Visualization

### i) Bar Chart

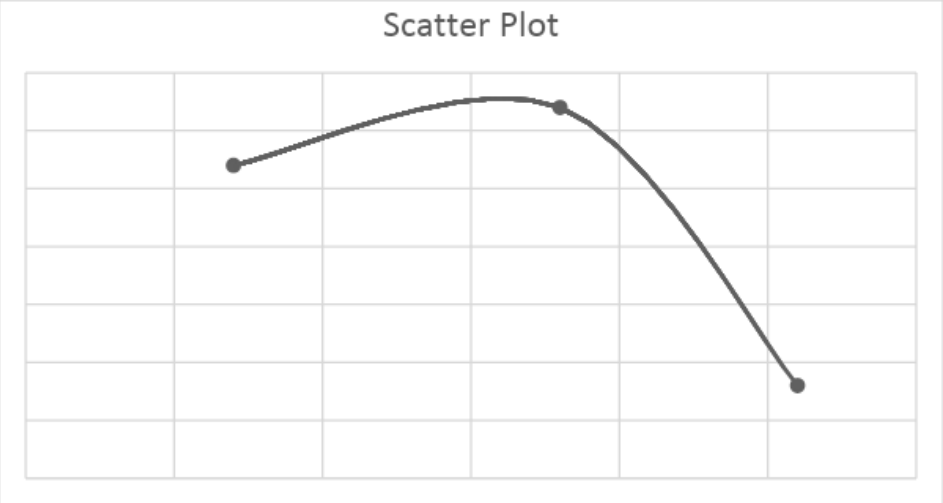
The pictorial representations of grouped data, in the form of vertical or horizontal rectangular bars, where the lengths of the bars are equivalent to the measure of data, are known as bar graphs or bar charts. The bars drawn are of uniform width, and the variable quantity is represented on one of the axes. Also, the measure of the variable is depicted on the other axes. The heights or the lengths of the bars denote the value of the variable, and these graphs are also used to compare certain quantities. The frequency distribution tables can be easily represented using bar charts which simplify the calculations and understanding of data.




### ii) Scatter Plot

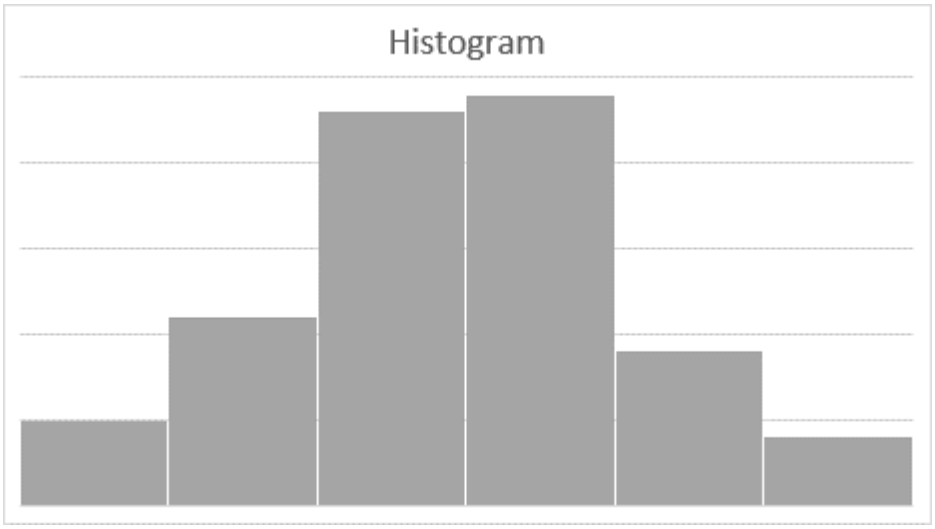
Scatter plots are important in statistics because they can show the extent of correlation, if any, between the values of observed quantities or phenomena (called variables). If no correlation exists between the variables, the points appear randomly scattered on the

coordinate plane. If a large correlation exists, the points concentrate near a straight line. Scatter plots are useful data visualization tools for illustrating a trend.



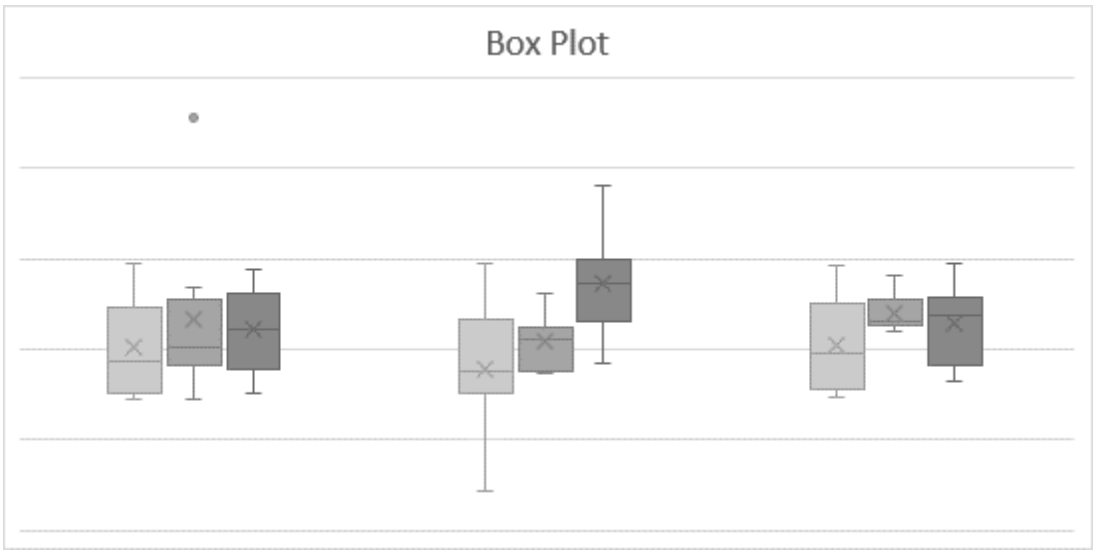

## iii) Histogram

A histogram is an approximate representation of the distribution of numerical data. Histograms give a rough sense of the density of the underlying distribution of the data, and often for density estimation: estimating the probability density function of the underlying variable. The total area of a histogram used for probability density is always normalized to 1. Histograms are sometimes confused with bar charts. A histogram is used for continuous data, where the bins represent ranges of data, while a bar chart is a plot of categorical variables. A histogram is a graph that shows the distribution of numerical data. It is a type of bar chart that shows the frequency or number of observations within different numerical ranges, called bins. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The histogram provides a visual representation of the distribution of the data, showing the number of observations that fall within each bin. This can be useful for identifying patterns and trends in the data, and for making comparisons between different datasets.



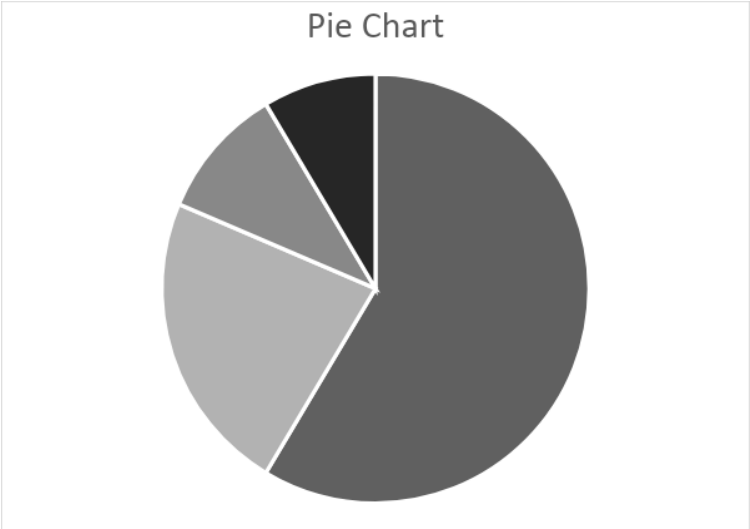

## iv) Box Plot

Boxplot is a method for graphically demonstrating the locality, spread and skewness groups of numerical data through their quartiles. In addition to the box on a box plot, there can be lines (which are called whiskers) extending from the box indicating variability outside the upper and lower quartiles, thus, the plot is also termed as the box-and-whisker plot and the box-and-whisker diagram. Outliers that differ significantly from the rest of the dataset may be plotted as individual points beyond the whiskers on the box-plot. Box plots are nonparametric: they display variation in samples of a statistical population without making any assumptions of the underlying statistical distribution. A boxplot is a standardized way of displaying the dataset based on the five-number summary: the minimum, the maximum, the sample median, and the first and third quartiles.



### v) Pie Chart

A pie chart is a valuable tool for visually representing categorical data. It consists of a circular graph divided into segments (or "slices"), each corresponding to a specific category. The size of each slice reflects the proportion of that category relative to the whole dataset. It is commonly used to illustrate the distribution of discrete variables, such as the percentage of different product types sold, the composition of survey responses, or the allocation of time spent on various activities. When creating a pie chart, we ensure that the sum of all categories equals 100%. Interpretation involves comparing the sizes of slices to understand the relative significance of each group. However, it's essential to recognize the limitations of pie charts. They work best for straightforward data arrangements and are less effective when comparing multiple pie charts or dealing with complex datasets. Pie charts require categorical data and are most effective when emphasizing proportions within a whole.




## 2.4.3. Correlation

Correlation is a statistical measure that expresses the extent to which two variables are linearly related (meaning they change together at a constant rate). It's a common tool for describing simple relationships without making a statement about cause and effect. The sample correlation coefficient, r, quantifies the strength of the relationship. Correlations are also tested for statistical significance. Correlation can't look at the presence or effect of other variables outside of the two being explored. Importantly, correlation doesn't tell us about cause and effect. We describe correlations with a unit-free measure called the correlation coefficient which ranges

from -1 to +1 and is denoted by r. Statistical significance is indicated with a p-value. Therefore, correlations are typically written with two key numbers: r and p.

- The closer r is to zero, the weaker the linear relationship.
- Positive r values indicate a positive correlation, where the values of both variables tend to increase together.
- Negative r values indicate a negative correlation, where the values of one variable tend to increase when the values of the other variable decrease.
- The p-value gives us evidence that we can meaningfully conclude that the population correlation coefficient is likely different from zero, based on what we observe from the sample.
- "Unit-free measure" means that correlations exist on their own scale.
- Once we've obtained a significant correlation, we can also look at its strength. A perfect positive correlation has a value of 1, and a perfect negative correlation has a value of -1. But in the real world, we would never expect to see a perfect correlation unless one variable is a proxy measure for the other. In fact, seeing a perfect correlation number can alert you to an error in your data!
- Another useful piece of information is the N, or number of observations. As with most statistical tests, knowing the size of the sample helps us judge the strength of our sample and how well it represents the population

There are several correlation coefficients measuring the degree of correlation. The most common of these is the Pearson correlation coefficient, which is sensitive only to a linear relationship between two variables (which may be present even when one variable is a nonlinear function of the other). Other correlation coefficients – such as Spearman's rank correlation – have been developed to be more robust than Pearson's, that is, more sensitive to nonlinear relationships.

The most familiar measure of dependence between two quantities is the Pearson product moment correlation coefficient (PPMCC), or "Pearson's correlation coefficient", commonly called simply "the correlation coefficient". It is obtained by taking the ratio of the covariance of the two variables in question of our numerical dataset, normalized to the square root of their variances. Mathematically, one simply divides the covariance of the two variables by the product of their standard deviations

A Pearson product-moment correlation coefficient attempts to establish a line of best fit through a dataset of two variables by essentially laying out the expected values and the resulting Pearson's correlation coefficient indicates how far away the actual dataset is from the expected values. Depending on the sign of our Pearson's correlation coefficient, we can end up with either a negative or positive correlation if there is any sort of relationship between the variables of our data set.

The population correlation coefficient between two random variables X and Y is defined as:

$$\rho(x,y) = corr(X,Y) = \frac{cov(X,Y)}{\sigma X \sigma Y} = \frac{E[(X-\mu x)(Y-\mu y)]}{\sigma X \sigma Y} \qquad \text{if, } \sigma X, \sigma Y > 0$$

where E is the expected value operator, cov means covariance, and corr is a widely used alternative notation for the correlation coefficient. The Pearson correlation is defined only if both standard deviations are finite and positive. An alternative formula purely in terms of moments is:

$$\rho xy = \frac{E(XY) - E(Y)E(Y)}{\sqrt{E(X^2) - E(X)^2}\sqrt{E(Y^2) - E(Y)^2}}$$

If the variables are independent, Pearson's correlation coefficient is 0, but the converse is not true because the correlation coefficient detects only linear dependencies between two variables.

### 2.4.4. Tests of Hypothesis

A statistical hypothesis test is a method of statistical inference used to decide whether the data at hand sufficiently support a particular hypothesis. Hypothesis testing allows us to make probabilistic statements about population parameters.

Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analysed. All analysts use a random population sample to test two different hypotheses: the null hypothesis and the alternative hypothesis.

The null hypothesis is usually a hypothesis of equality between population parameters, H0 is the symbol for it. and the alternative hypothesis is effectively the opposite of a null hypothesis, H1 is the symbol for it. Thus, they are mutually exclusive, and only one can be true. However, one of the two hypotheses will always be true.

Depending on the population distribution, you can classify the statistical hypothesis into two types.

**Null Hypothesis:** A simple hypothesis specifies an exact value for the parameter.

$H_0 = \mu_1 = \mu_2 = \cdots = \mu_n$ n = 1,2,3….

**Alternate Hypothesis:** A composite hypothesis specifies a range of values.

$H_1 = Atleast\ one\ \mu_i\ is\ different.$

## Parametric tests

The basic principle behind the parametric tests is that we have a fixed set of parameters that are used to determine a probabilistic model that may be used in Machine Learning as well. Parametric tests are those tests for which we have prior knowledge of the population distribution (i.e., normal), or if not then we can easily approximate it to a normal distribution which is possible with the help of the Central Limit Theorem.

## i) Proportion Test:

A proportion test, also known as a test of proportions or a test for comparing two proportions, is a statistical method used to determine whether two proportions or percentages are significantly different from each other. This test is used when the data being analysed consists of categorical variables with two possible outcomes. The null hypothesis for a proportion test is that there is no significant difference between the two proportions being compared, while the alternative hypothesis is that there is a significant difference between them. The test involves calculating a test statistic based on the difference between the two proportions and their standard errors. This test statistic is then compared to a critical value from a statistical table or calculated using software, and the result is interpreted based on the level of significance chosen by the researcher

The test statistic is given by

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$\hat{p}$ = sample proportion

$p_0 =$ population proportion under null hypothesis

$n$ = sample size

**ii) Analysis of Variance (ANOVA)**

ANOVA (Analysis of Variance) is a statistical test used to compare the means of two or more groups. It is used to determine whether there is a significant difference between the means of the groups and can be used to compare more than two groups at a time. ANOVA is based on the idea of partitioning the total variance in a dataset into different components, and testing whether the variance between the groups is significantly different from the variance within the groups. There are several types of ANOVA tests, including one-way ANOVA, two-way ANOVA, and repeated measures ANOVA. One-way ANOVA is used to compare the means of two or more groups that have been randomly assigned to different treatment conditions. Two-way ANOVA is used to compare the means of two or more groups that have been assigned to different treatment conditions in two different variables. Repeated measures ANOVA is used to compare the means of two or more groups that have been measured multiple times on the same dependent variable. ANOVA is a powerful tool for comparing group means, but it has some limitations. It assumes that the data are normally distributed and that the variances of the groups are equal. It also assumes that the groups are independent, meaning that the observations in one group are not related to the observations in the other groups. If these assumptions are not met, alternative statistical tests, such as nonparametric tests or generalized linear models, may be more appropriate.

**Non-Parametric Test:** Non-parametric tests are experiments that do not require the underlying population for assumptions. It does not rely on any data referring to any particular parametric group of probability distributions. Non-parametric methods are also called distribution-free tests since they do not have any underlying population.

## i) Chi-Square Test:

The Chi-square test of independence is a statistical test used to determine whether two categorical variables are related or independent of each other. In other words, it is used to evaluate the association between two variables when the variables are categorical. The test is based on the concept of expected and observed frequencies. The observed frequency is the actual number of times a particular category occurs in a sample; while the expected frequency

is the number of times that category is expected to occur if the two variables are independent. The test compares the observed frequencies with the expected frequencies to determine if there is a significant association between the two variables.

The Formula of Chi-Square is:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Here,

$\chi^2$ = chi squared

$O_i$ = observed value

$E_i$ = expected value

In conclusion, the Chi-square test of independence is a useful statistical test for analysing the association between two categorical variables. It is commonly used in social science research, market research, and other fields where categorical data are collected

## ii) Shapiro-Wilk Test:

The Shapiro-Wilk test is a way to tell if a random sample comes from a normal distribution. The test gives you a W value; small values indicate your sample is not normally distributed (you can reject the null hypothesis that your population is normally distributed if your values are under a certain threshold).

The formula for the W value is

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

where, xi are the ordered random sample values and ai are constants generated from the covariance, variances and means of the sample (size n) from a normally distributed sample. The test has limitations, most importantly that the test has a bias by sample size. The larger the sample, the more likely you'll get a statistically significant result.

The null-hypothesis of this test is that the population is normally distributed. Thus, if the p value is less than the chosen alpha level, then the null hypothesis is rejected and there is

evidence that the data tested are not normally distributed. On the other hand, if the p value is greater than the chosen alpha level, then the null hypothesis (that the data came from a normally distributed population) can't be rejected (e.g., for an alpha level of .05, a data set with a p value of less than 0.05 rejects the null hypothesis that the data are from a normally distributed population – consequently, a data set with a p value more than the .05 alpha value fails to reject the null hypothesis that the data is from a normally distributed population).

### 2.4.5. Regression Analysis

In statistical modelling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable (often called the 'outcome' or 'response' variable, or a 'label' in machine learning parlance) and one or more independent variables (often called 'predictors', 'covariates', 'explanatory variables' or 'features'). The most common form of regression analysis is linear regression, in which one finds the line (or a more complex linear combination) that most closely fits the data according to a specific mathematical criterion.

## Linear Models

Linear models are a way of describing a response variable in terms of a linear combination of predictor variables. The response should be a continuous variable and be at least approximately normally distributed. Such models find wide application, but cannot handle clearly discrete or skewed continuous responses.

A linear model specifies a linear relationship between a dependent variable and independent variables.

$$Y = \beta_0 + \beta_1 X_1 + \cdots . + \beta_n X_n + \varepsilon$$

where Y is the dependent variable, $X_i$ are independent variables, $\beta_i$ are parameters of the model.

### i) Simple Linear Regression (SLR)

In statistics, simple linear regression is a linear regression model with a single explanatory variable That is, it concerns two-dimensional sample points with one independent variable and one dependent variable (conventionally, the x and y coordinates in a Cartesian coordinate

system) and finds a linear function (a non-vertical straight line) that, as accurately as possible, predicts the dependent variable values as a function of the independent variable. The adjective simple refers to the fact that the outcome variable is related to a single predictor.

The formula for a Simple linear regression is:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Y is the response variable, X is the explanatory variables, $\beta_0$ is the y-intercept (constant term), $\beta_1$ is the slope coefficients for each explanatory variable, ε is the random error term (also known as the residuals).

## ii) Multiple Linear Regression (MLR)

Multiple linear regression is used to estimate the relationship between two or more independent variables and one dependent variable. We can use multiple linear regression when we want to know:

1. How strong the relationship is between two or more independent variables and one dependent variable (e.g., how rainfall, temperature, and amount of fertilizer added affect crop growth)
2. The value of the dependent variable at a certain value of the independent variables (e.g., the expected yield of a crop at certain levels of rainfall, temperature, and fertilizer addition).

The formula for a multiple linear regression is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \varepsilon$$

where, for I = 1,2,3……n observations

Yi is the response variable, Xi is the explanatory variables, β0 is the y-intercept (constant term) ,βp is the slope coefficients for each explanatory variable ,ε is the random error term (also known as the residuals)

## 2.5. Softwares and Programming Languages Used

### 2.5.1. Microsoft Excel

Excel is a spreadsheet program from Microsoft where one can record data in the form of tables. It is easy to analyse data in an Excel spreadsheet. Microsoft Excel enables users to format, organize and calculate data in a spreadsheet. Excel contains a large number of boxes called cells that are ordered in rows and columns. Data is placed in these cells.

**Benefits of Using MS Excel**

- **Easy To Store Data:** Since there is no limit to the amount of information that can be saved in a spreadsheet, MS Excel is widely used to save data or to analyse data. Filtering information in Excel is easy and convenient.
- **Easy To Recover Data:** If the information is written on a piece of paper, finding it may take longer, however, this is not the case with excel spreadsheets. Finding and recovering data is easy.
- **Application of Mathematical Formulas:** Doing calculations has become easier and less time-taking with the formulas option in MS excel
- **More Secure:** These spreadsheets can be password secured in a laptop or personal computer and the probability of losing them is way lesser in comparison to data written in registers or piece of paper.

### 2.5.2 Python

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form.

There is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping

through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power.

**Python’s features are:**

- Easy-to-understand Python features a small number of keywords, a straightforward structure, and a well-defined syntax. This enables the pupil to swiftly learn the language.
- Easily readable Python code is easier to read and understand since it is more clearly stated.
- Python’s source code is quite straightforward to maintain.
- Python has a large standard library that is particularly portable and cross-platform compatible on UNIX, Windows, and Macintosh.
- Python features an interactive mode that allows interactive testing and debugging of code snippets.
- Python is portable, meaning it can operate on a broad range of hardware systems and has the same user interface across all of them.

**i) NumPy Library**

NumPy is the most popular python library for matrix/vector computations. Due to python’s popularity, it is also one of the leading libraries for numerical operations on large quantities of data. If you have Python and PIP already installed on a system, then installation of NumPy is very easy.

Once NumPy is installed, import it in your applications by adding the import keyword. If you have Python and PIP already installed on a system, then installation of NumPy is very easy. Once NumPy is installed, import it in your applications by adding the import keyword.

**ii) Pandas Library**

Pandas is a very popular library for working with data (its goal is to be the most powerful and flexible open-source tool, and in our opinion, it has reached that goal). Data Frames are at the centre of pandas. A Data Frame is structured like a table or spreadsheet. The rows and the columns both have indexes, and you can perform operations on rows or columns separately.

A pandas Data Frame can be easily changed and manipulated. Pandas has helpful functions for handling missing data, performing operations on columns and rows, and transforming data. If that wasn't enough, a lot of SQL functions have counterparts in pandas, such as join, merge, filter by, and group by. With all of these powerful tools, it should come as no surprise that pandas is very popular among data scientists.

### 2.5.3 R Programming Language

R is a programming language and software environment for analysis, graphics representation and reporting. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently managed by the R Development Core Team. The core of R is an interpreted computer language which allows branching and looping as well as modular programming using functions. R allows integration with the procedures written in the C, C++, .Net, Python or FORTRAN languages for efficiency. R is freely available under the GNU General Public License, and precompiled binary versions are provided for various operating systems like Linux, Windows and Mac.

**Features of R**

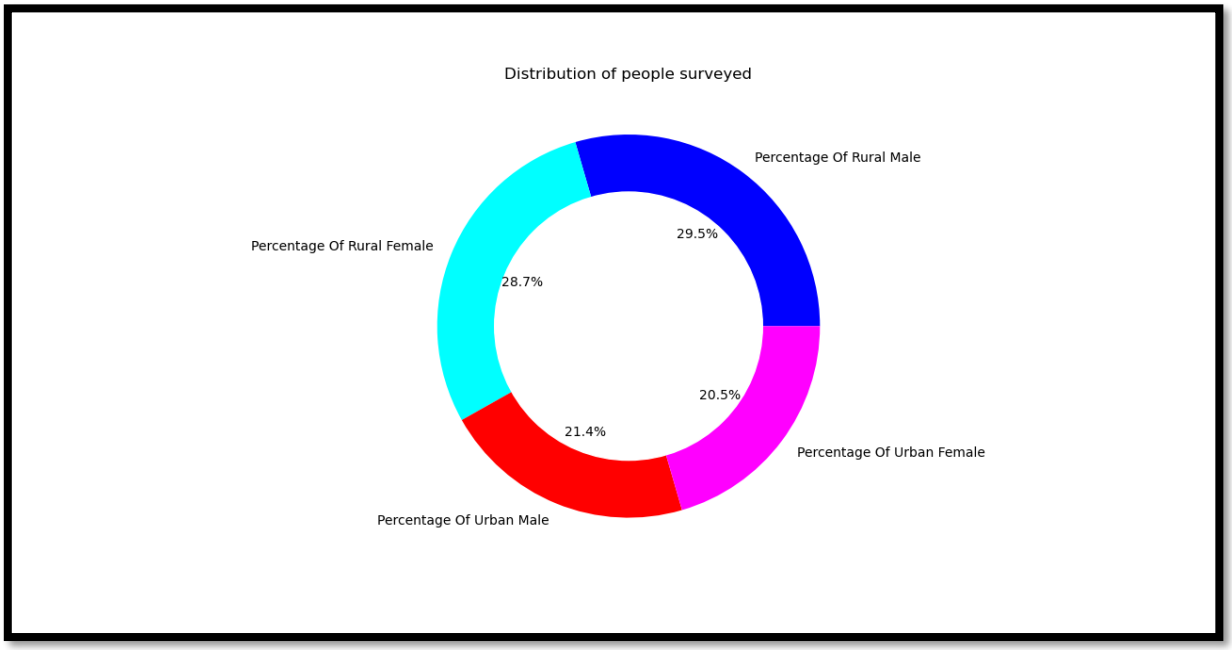
- Ris a well-developed, simple and effective programming language which includes conditionals, loops, user defined recursive functions and input and output facilities.
- R has an effective data handling and storage facility.
- R provides a suite of operators for calculations on arrays, lists, vectors, and matrices
- R provides a large, coherent and integrated collection of tools for data analysis.
- R provides graphical facilities for data analysis and display either directly at the computer or printing at the papers.

# Chapter – 3

# Data Analysis

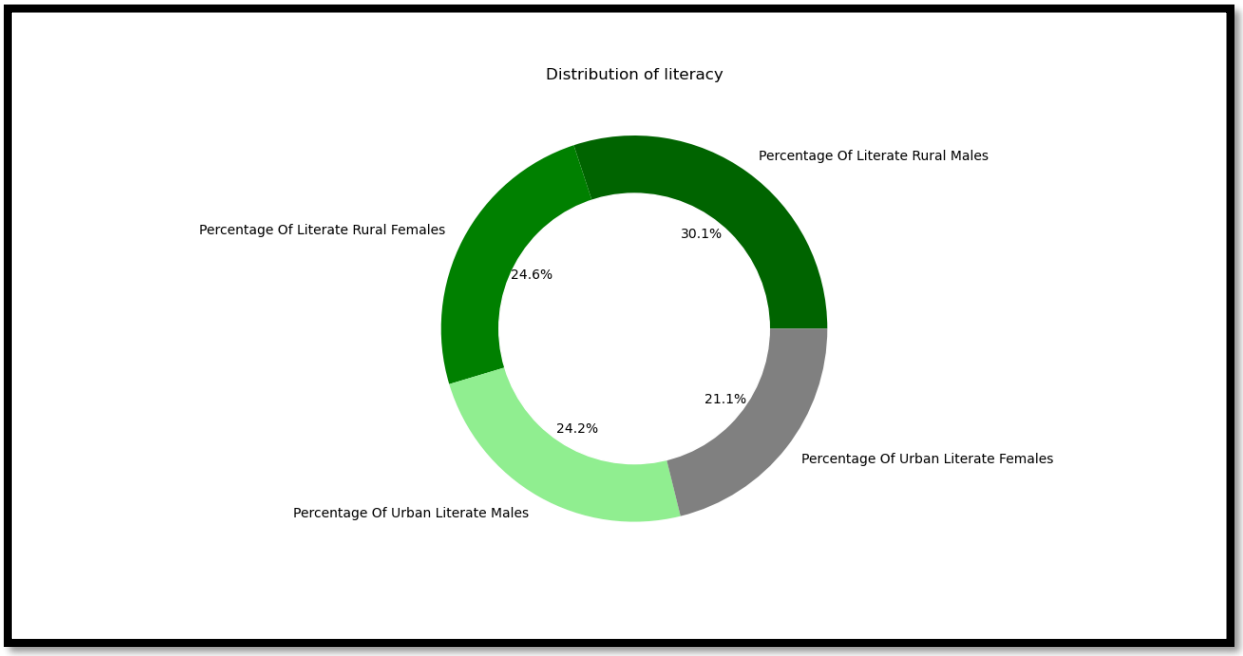
# 3.1 Exploratory Data Analysis

- **Distribution of people surveyed**

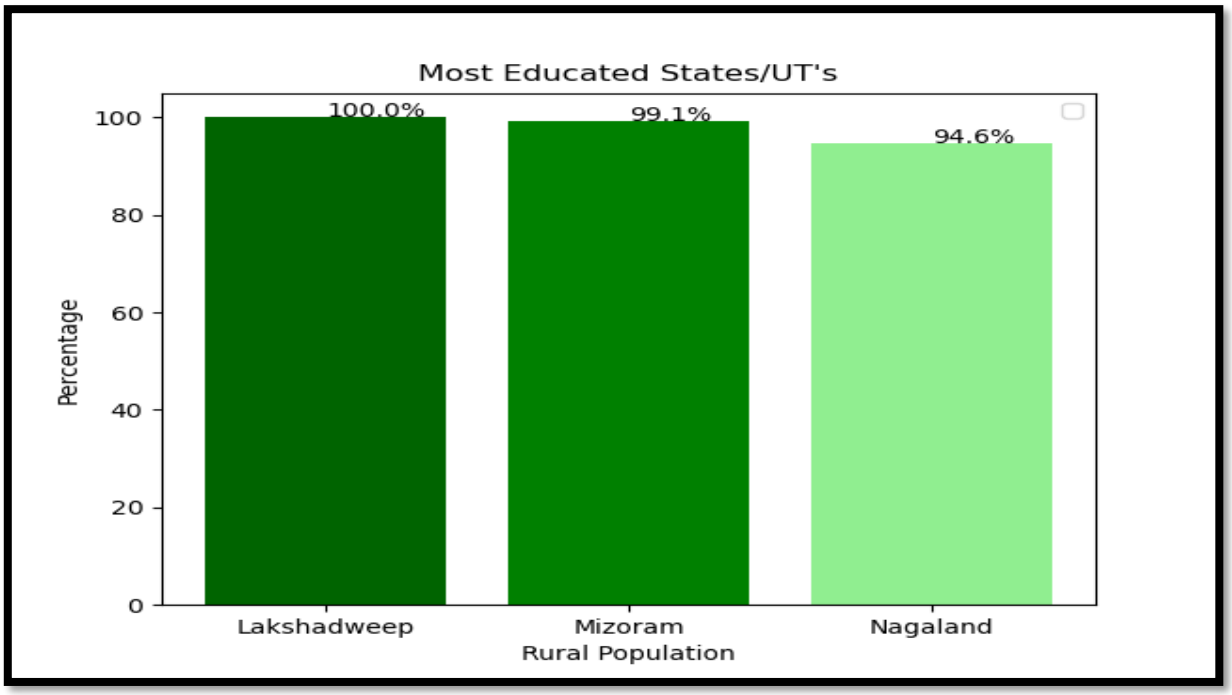


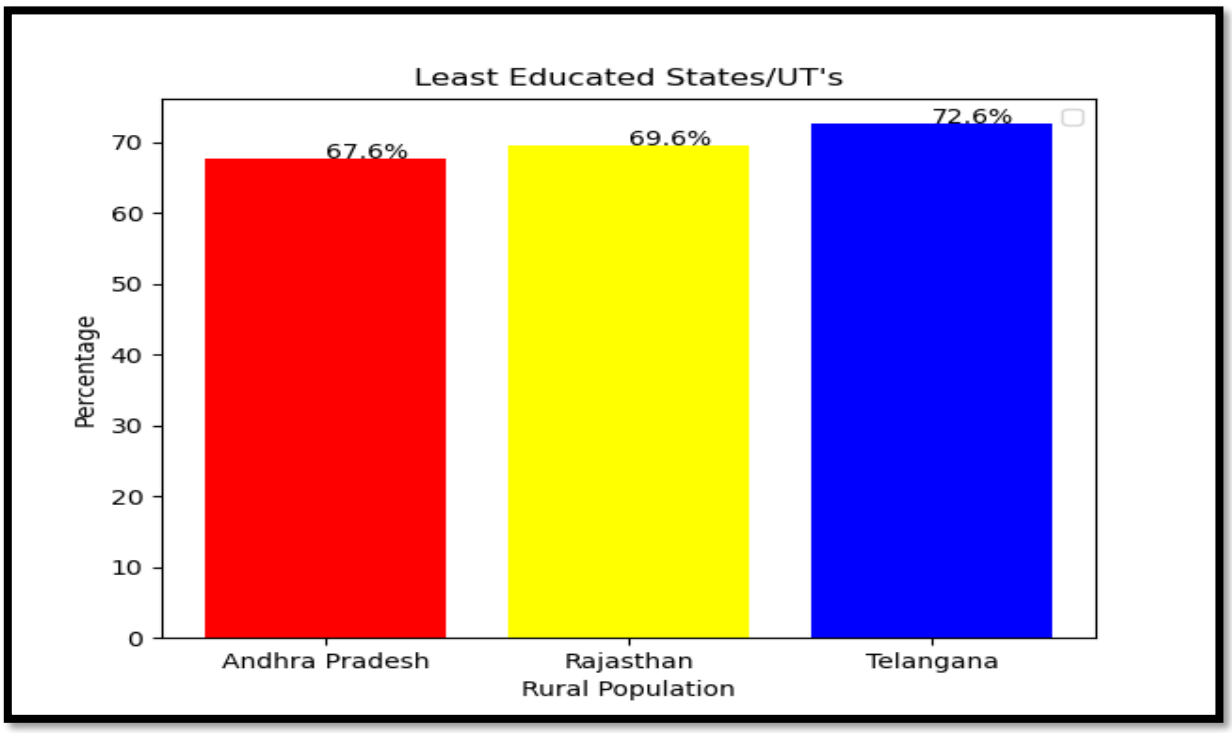

## Literacy Rate Exploratory Data Analysis

- **Distribution based on literacy**

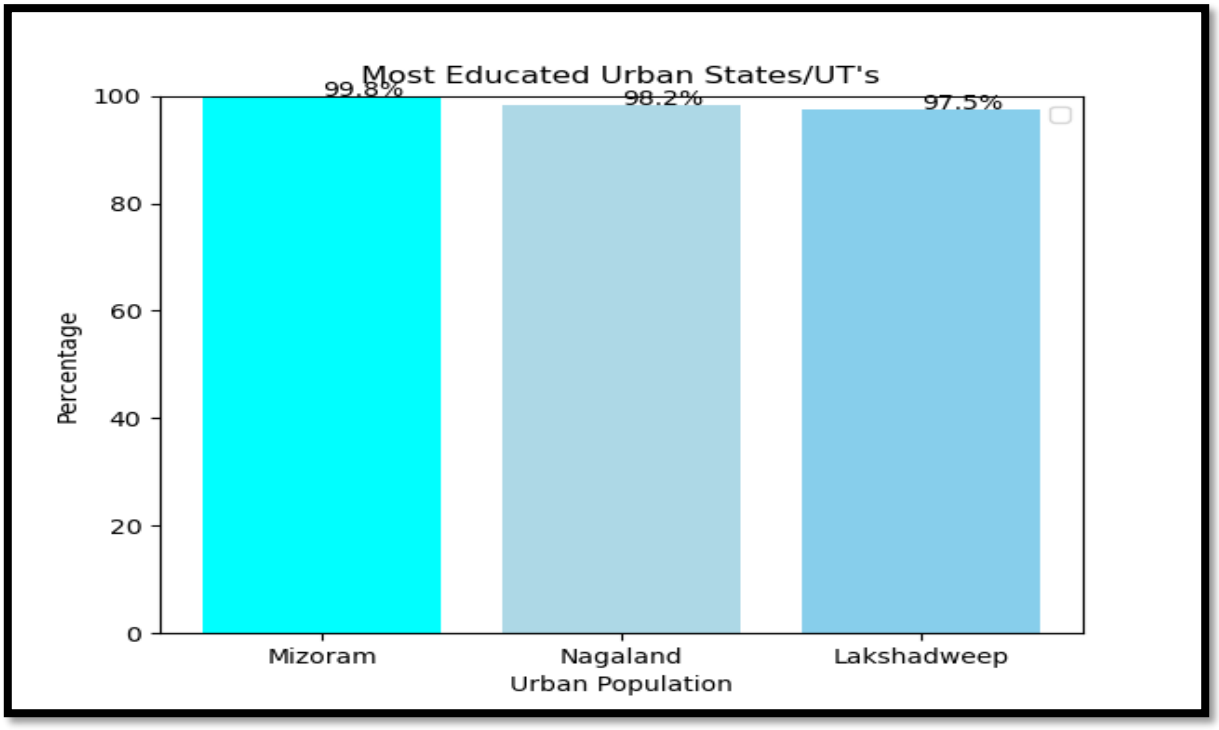


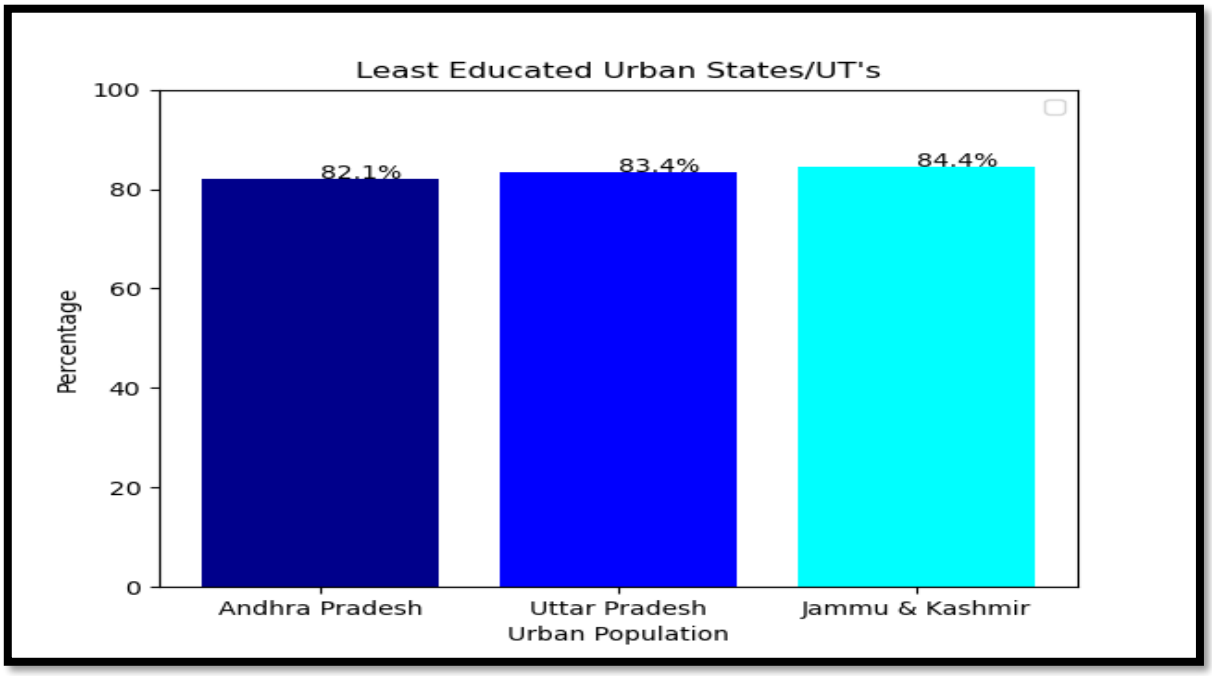
- **Comparison of states/Union territories based on percentage of educated population in rural areas**



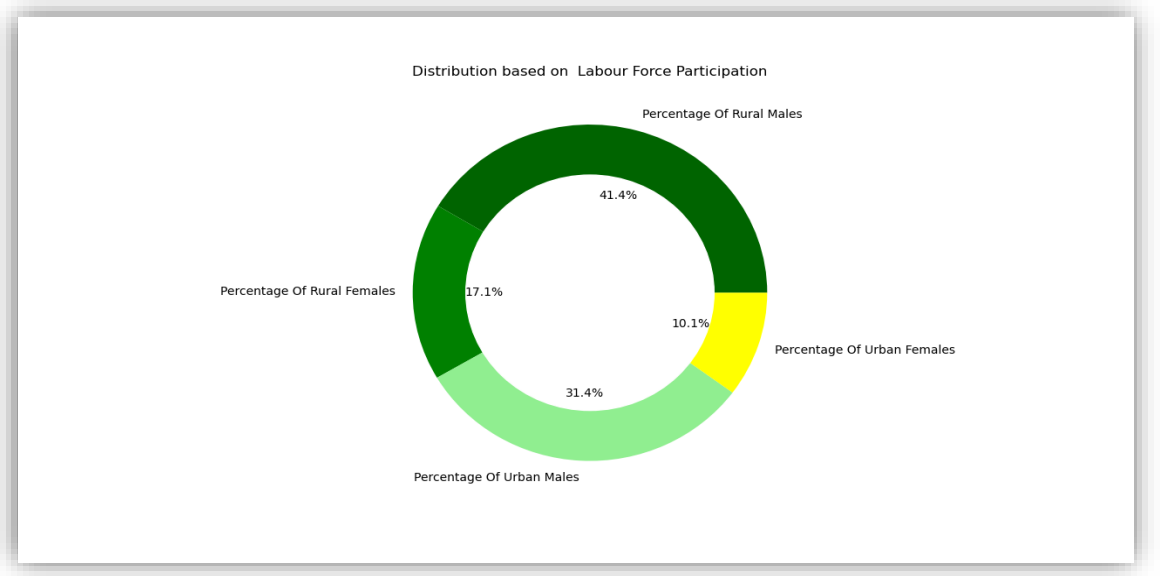
- **Comparison of states/Union territories based on percentage of educated population in urban areas**



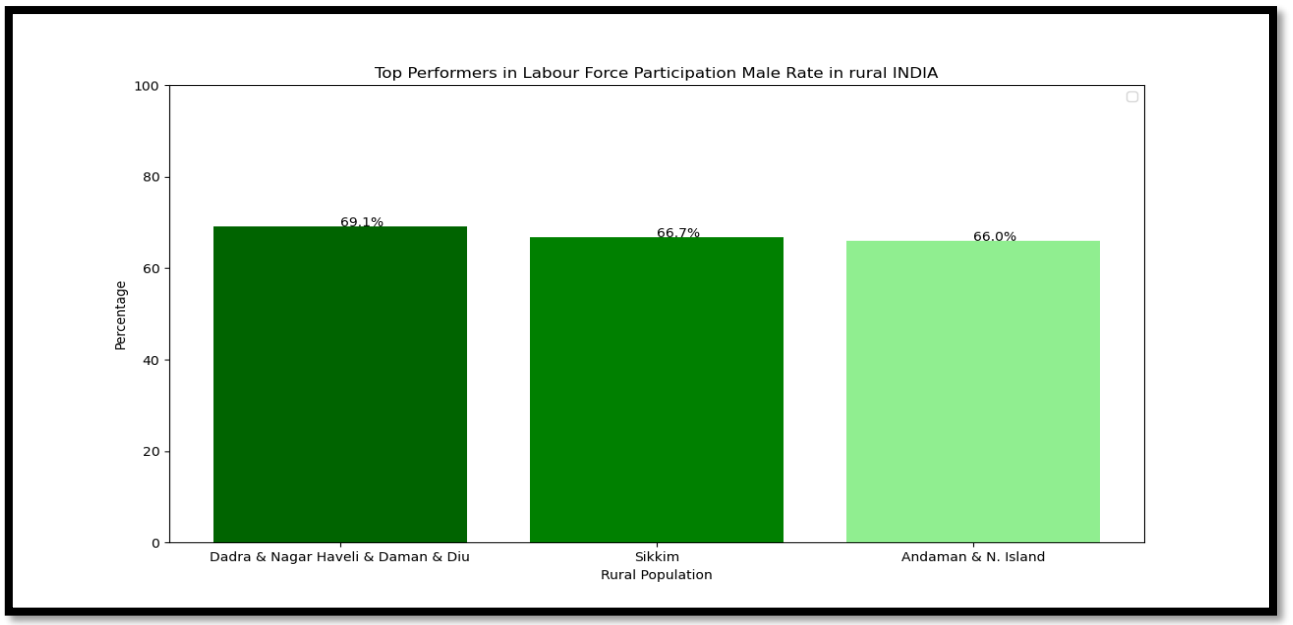



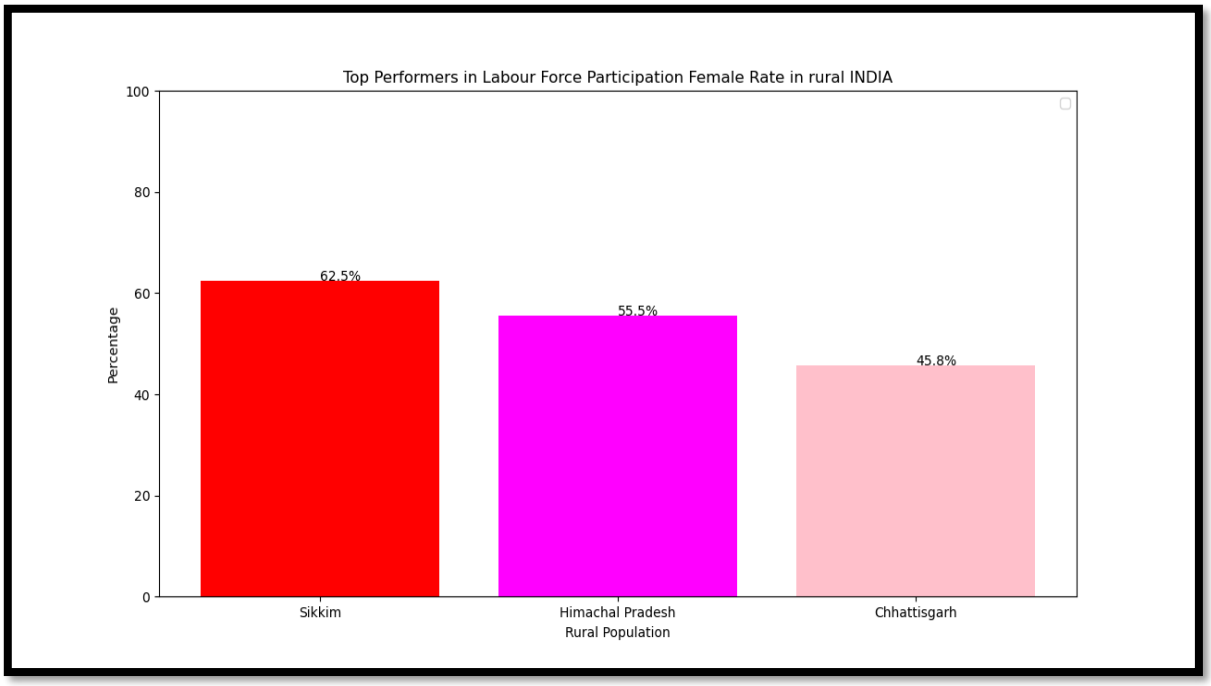
# Labour Force Participation Rate Exploratory Data Analysis

- **Distribution based on labour force**

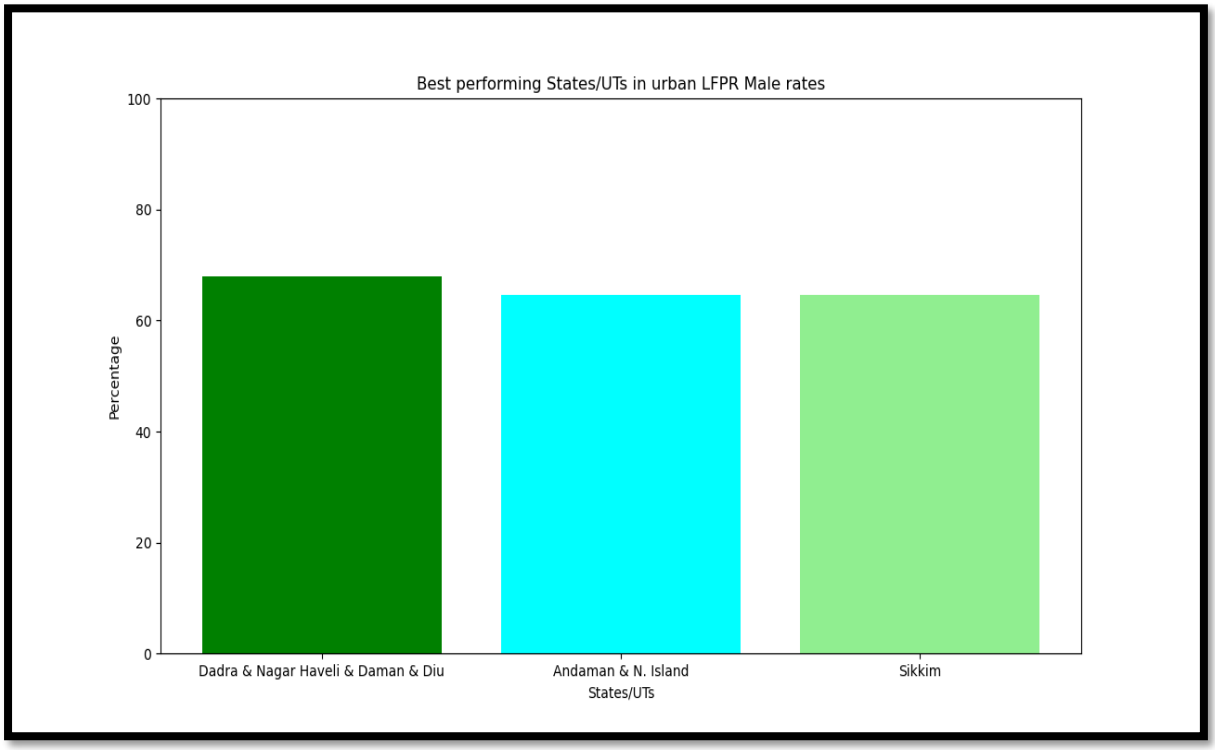


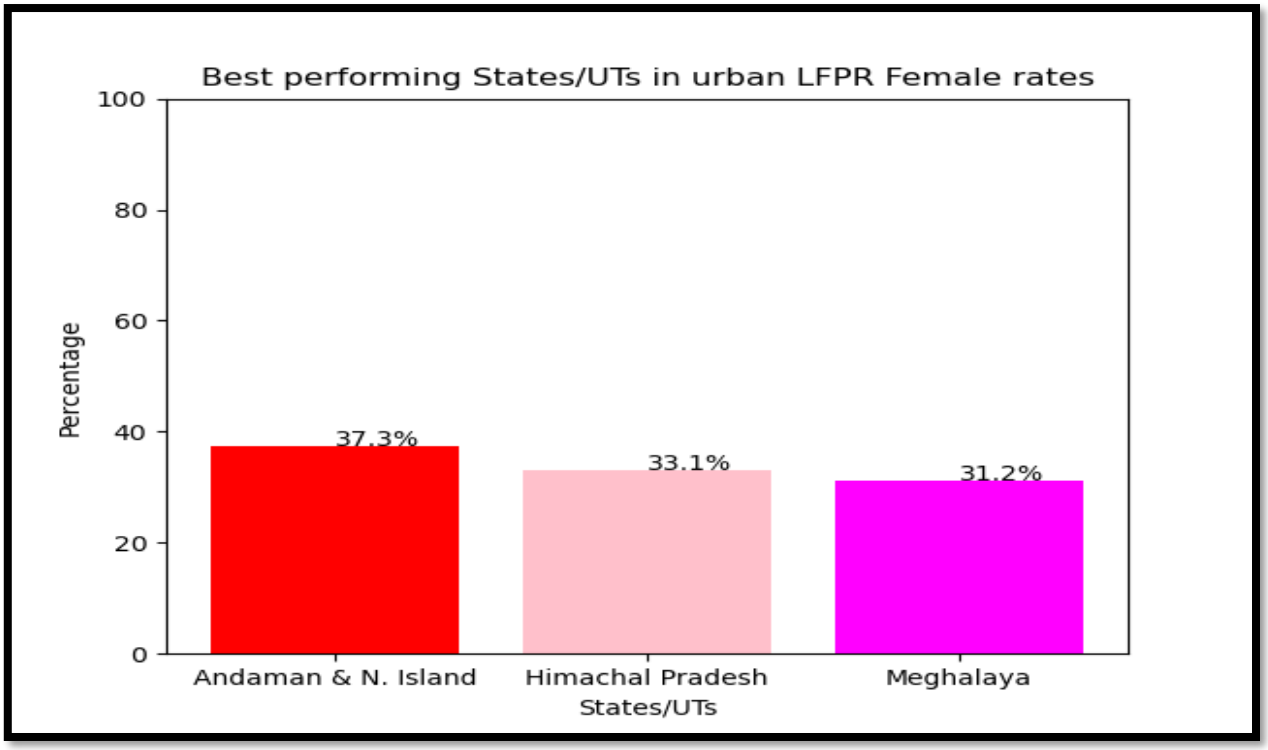

- **Comparison of states/Union territories based on percentage of Labour force participation in rural areas**





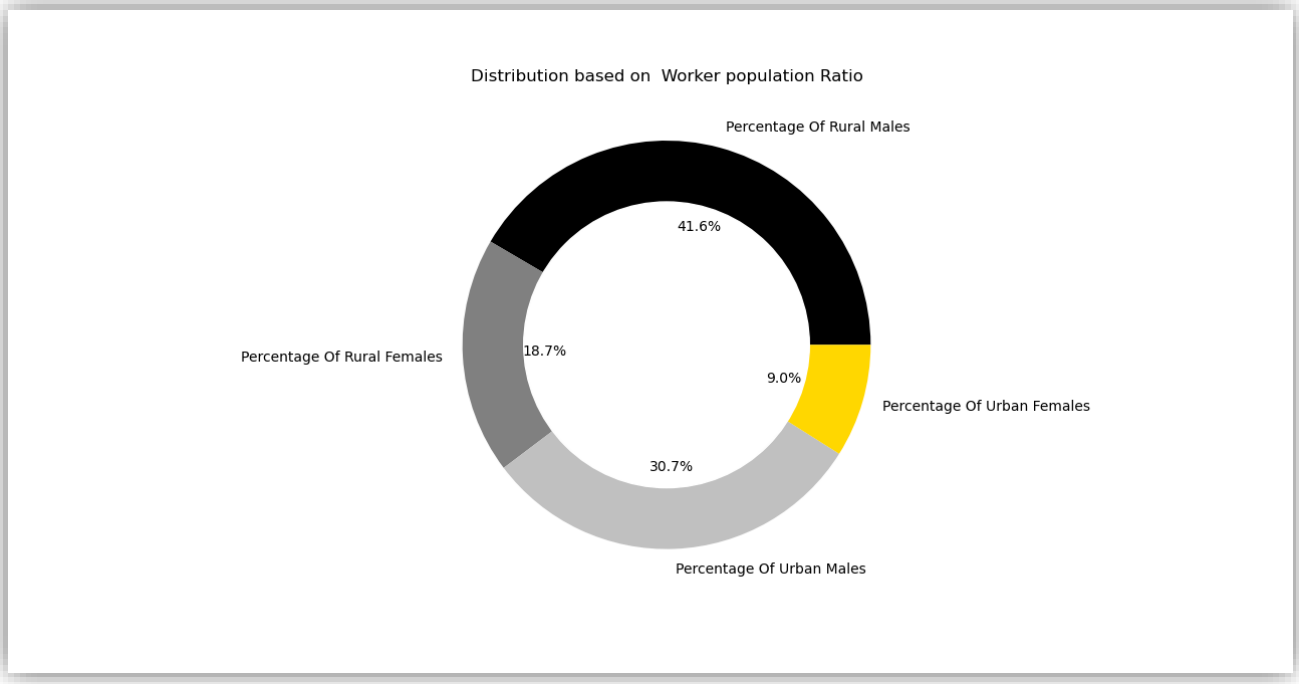

- **Comparison of states/Union territories based on percentage of Labour force participation in urban areas**



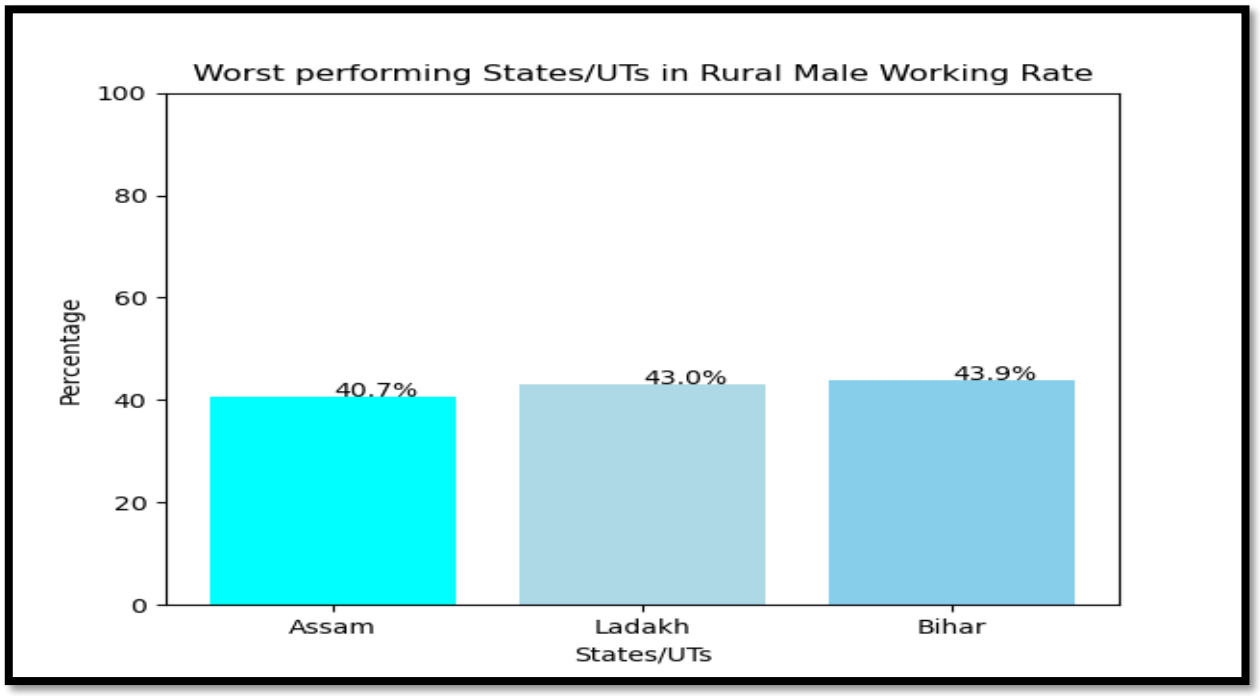


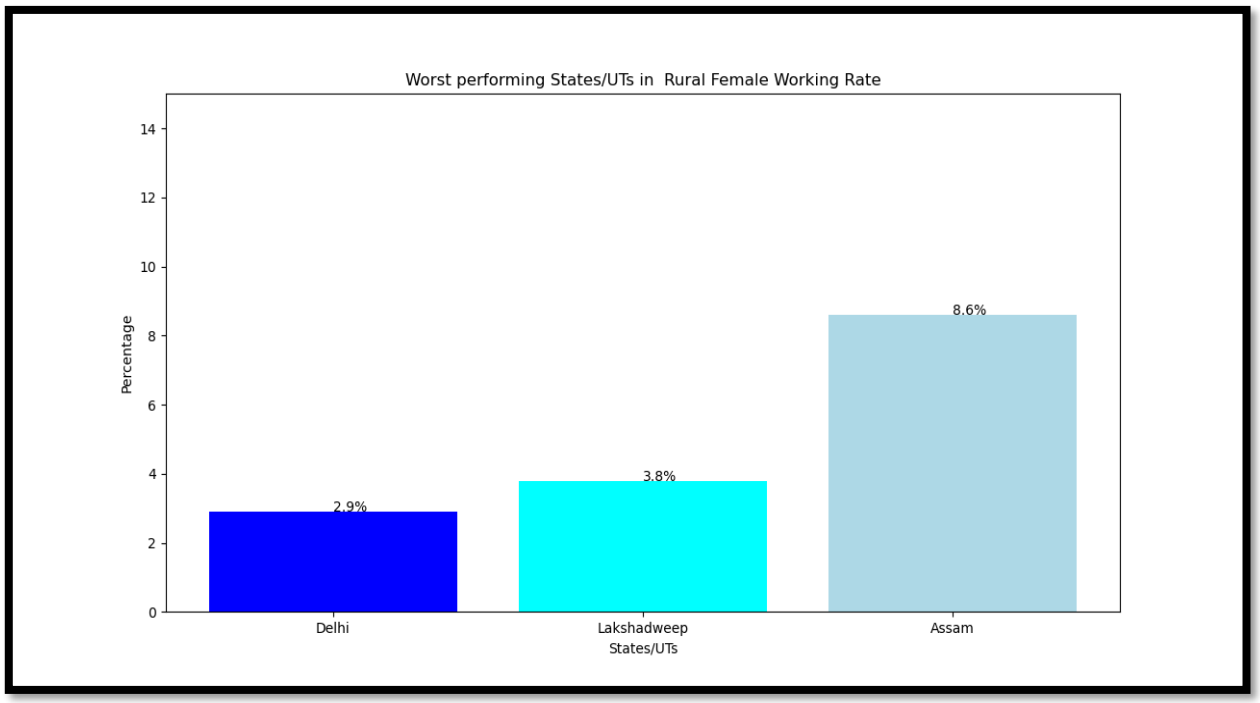

## Worker Population Ratio Exploratory Data Analysis

- **Distribution based on worker population ratio**

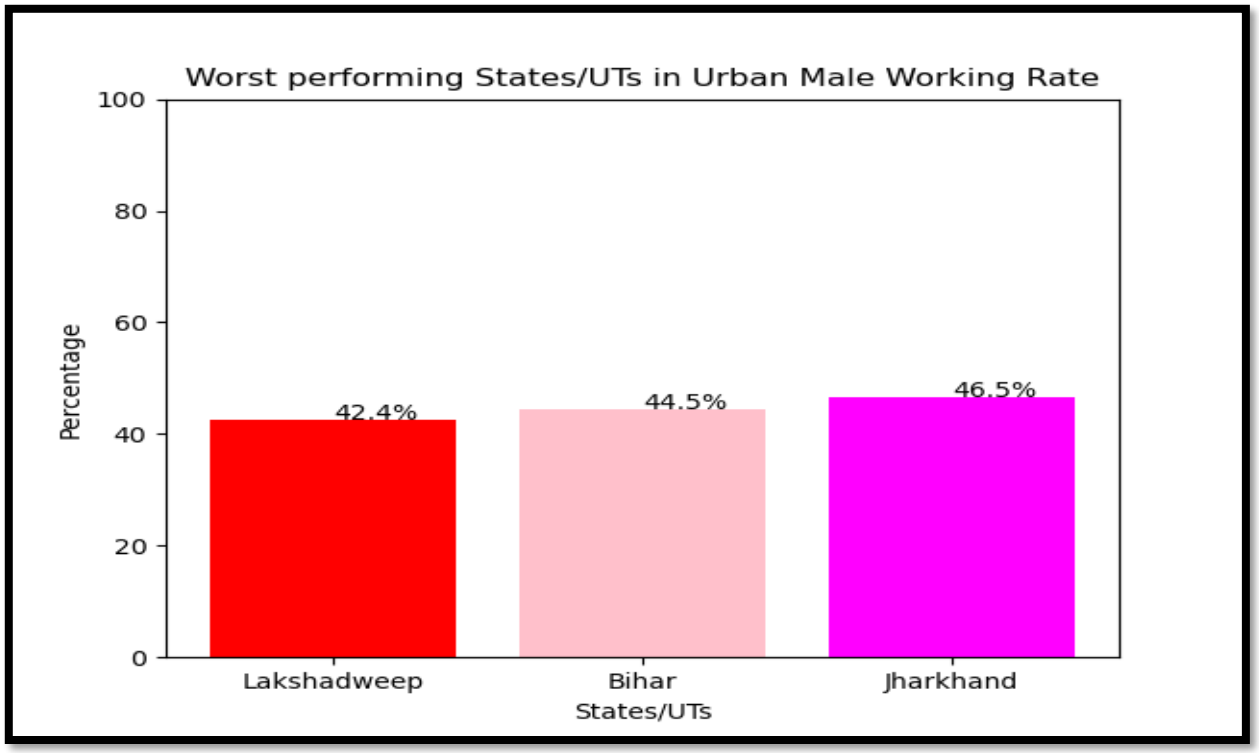


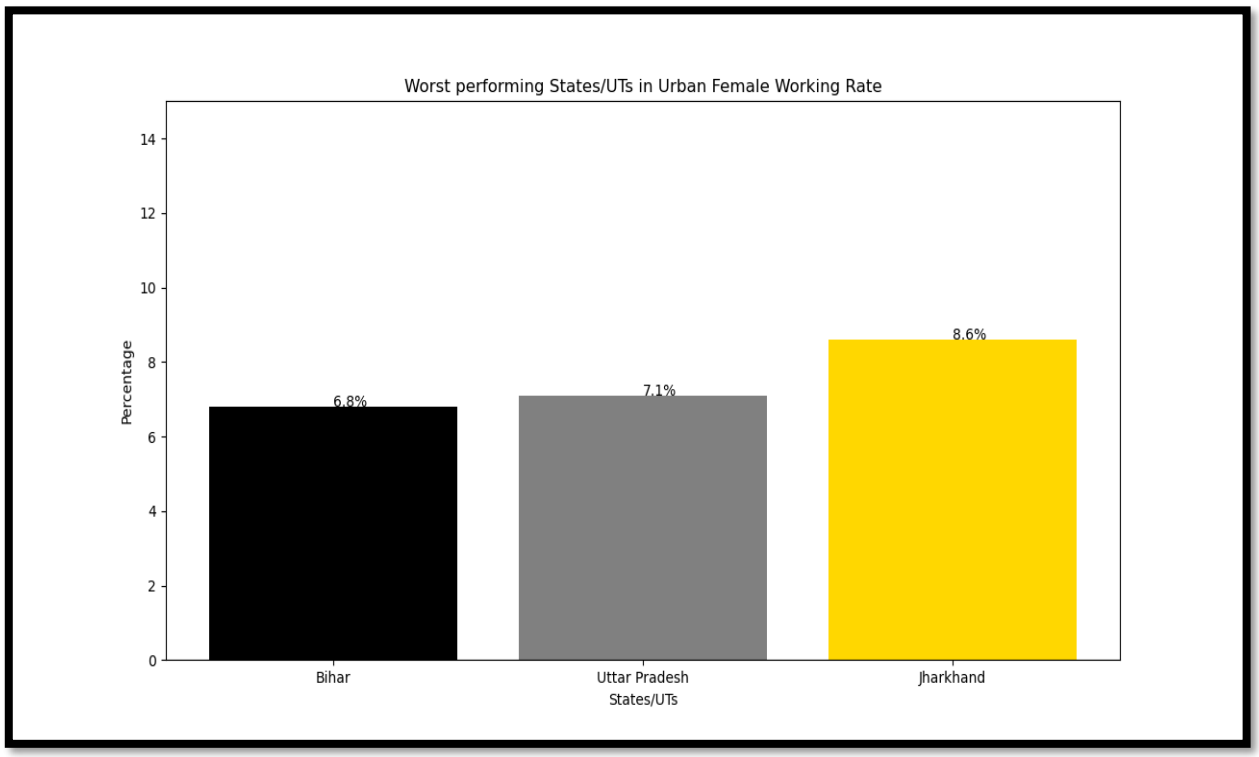
- **Comparison of states/Union territories based on percentage of worker population ratio in rural areas**



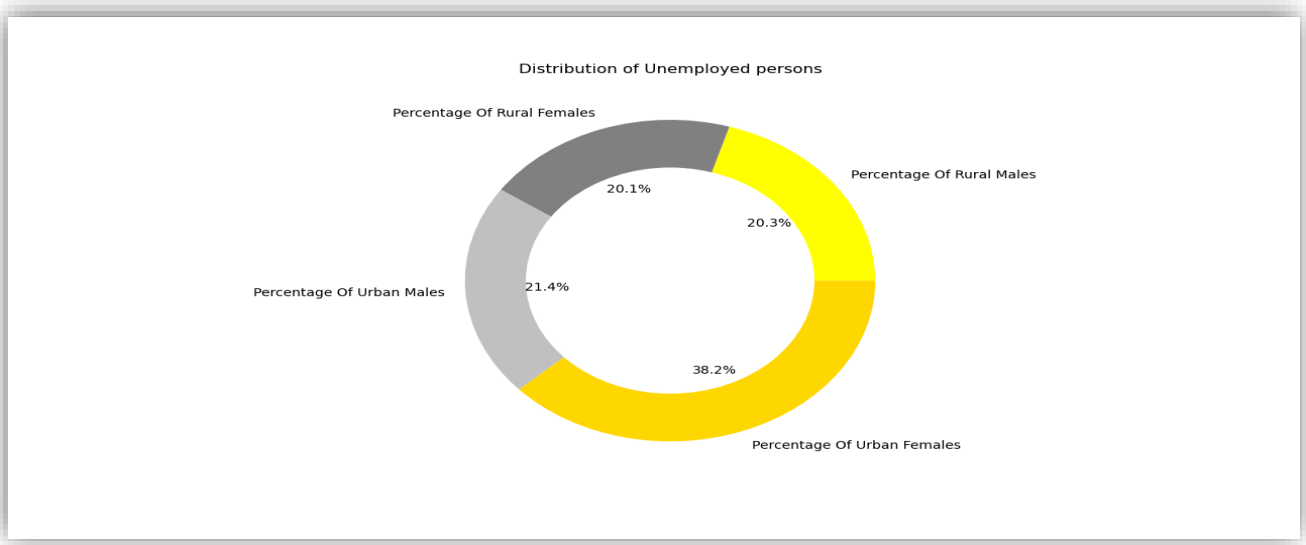
- **Comparison of states/Union territories based on percentage of worker population ratio in urban areas**



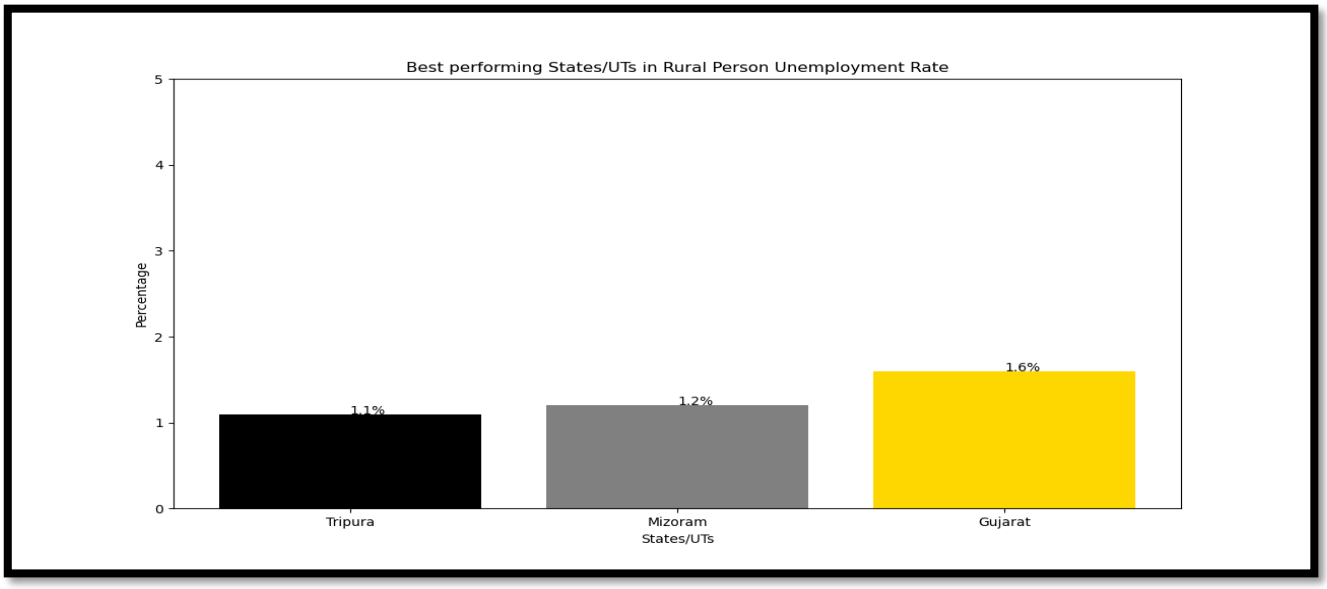



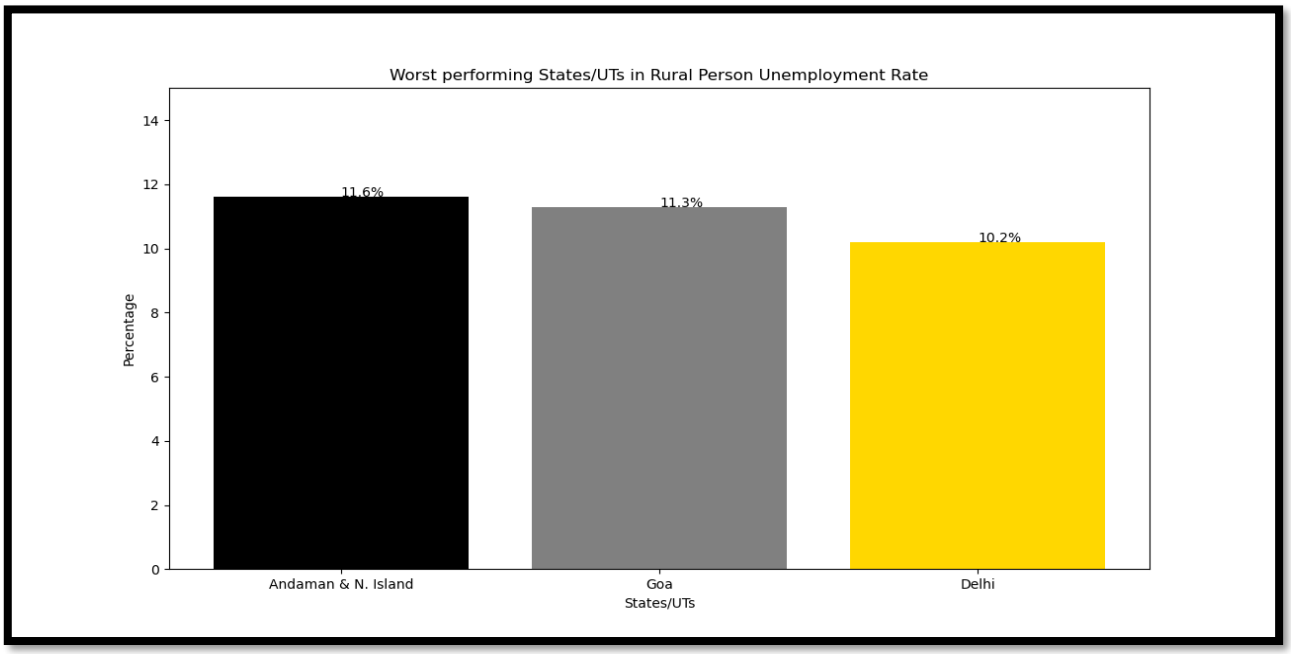
# Unemployment Rate Exploratory Data Analysis

- **Distribution of Unemployed people**

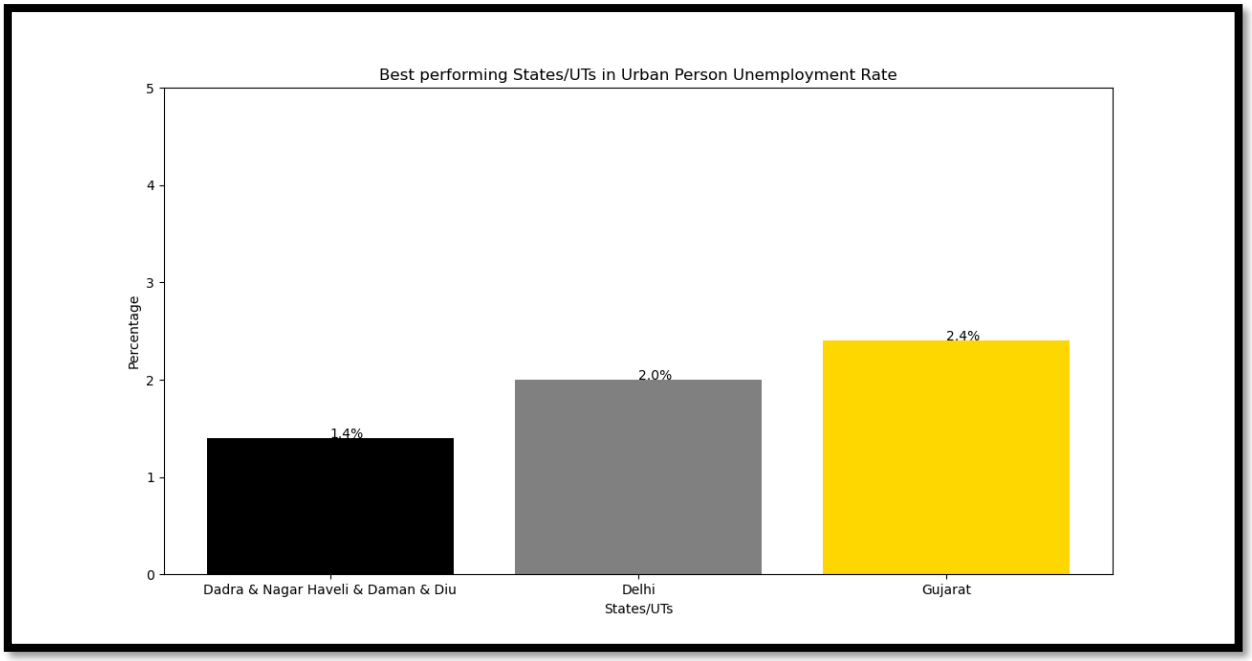


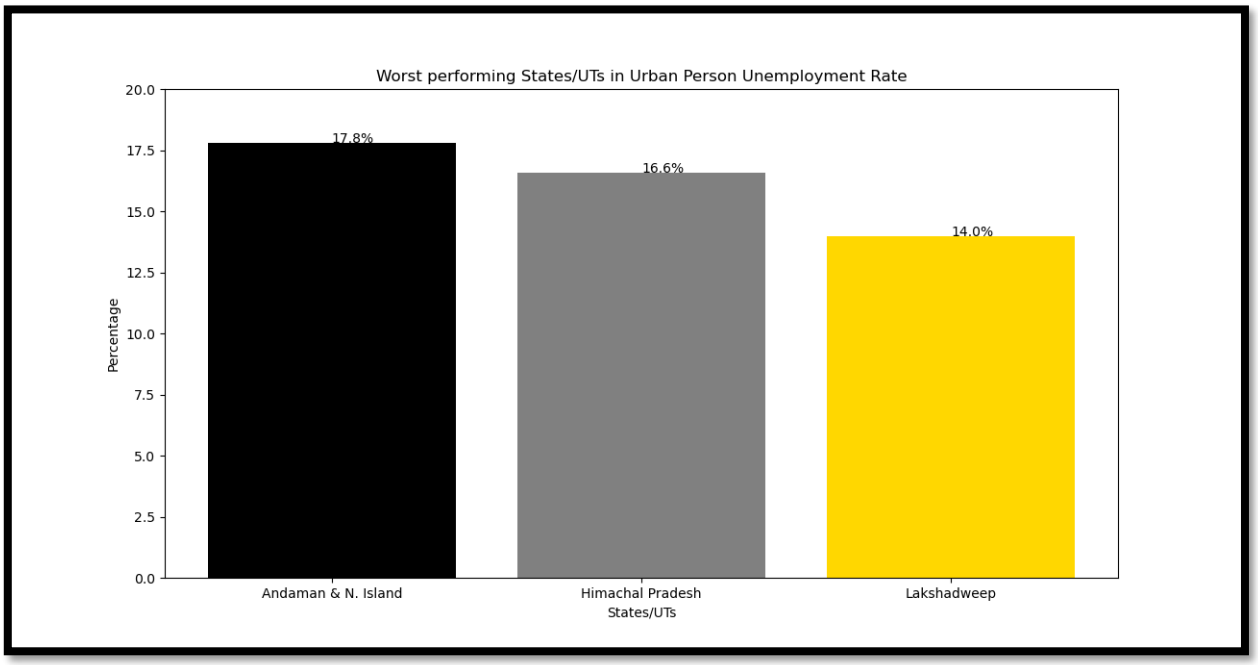

- **Comparison of states/Union territories based on percentage of unemployment rate in rural areas**



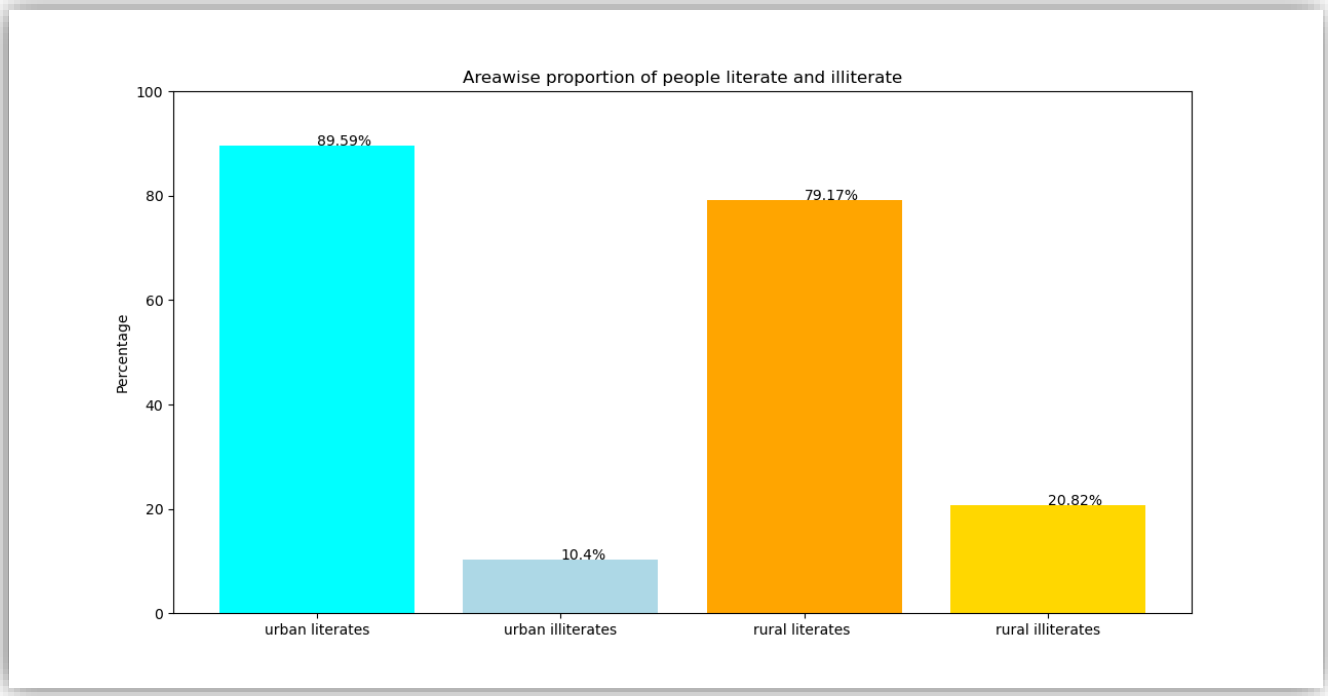



- **Comparison of states/Union territories based on percentage of unemployment rate in urban areas**




Worst performing States/UTs in Urban Person Unemployment Rate
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Percentage
17.8%
16.6%
14.0%
Andaman & N. Island
Himachal Pradesh
Lakshadweep
States/UTs

## 3.2.1 Assessment of Literacy




### 3.2.1.1 Area wise dependency on Literacy Rate

In this particular associativity test, we are going check the association between literate people and illiterate people Area wise (Urban and rural).

$H_0$ : There is no association between literacy and areas of India.

$H_A$ : There exists an association between literacy and areas of India

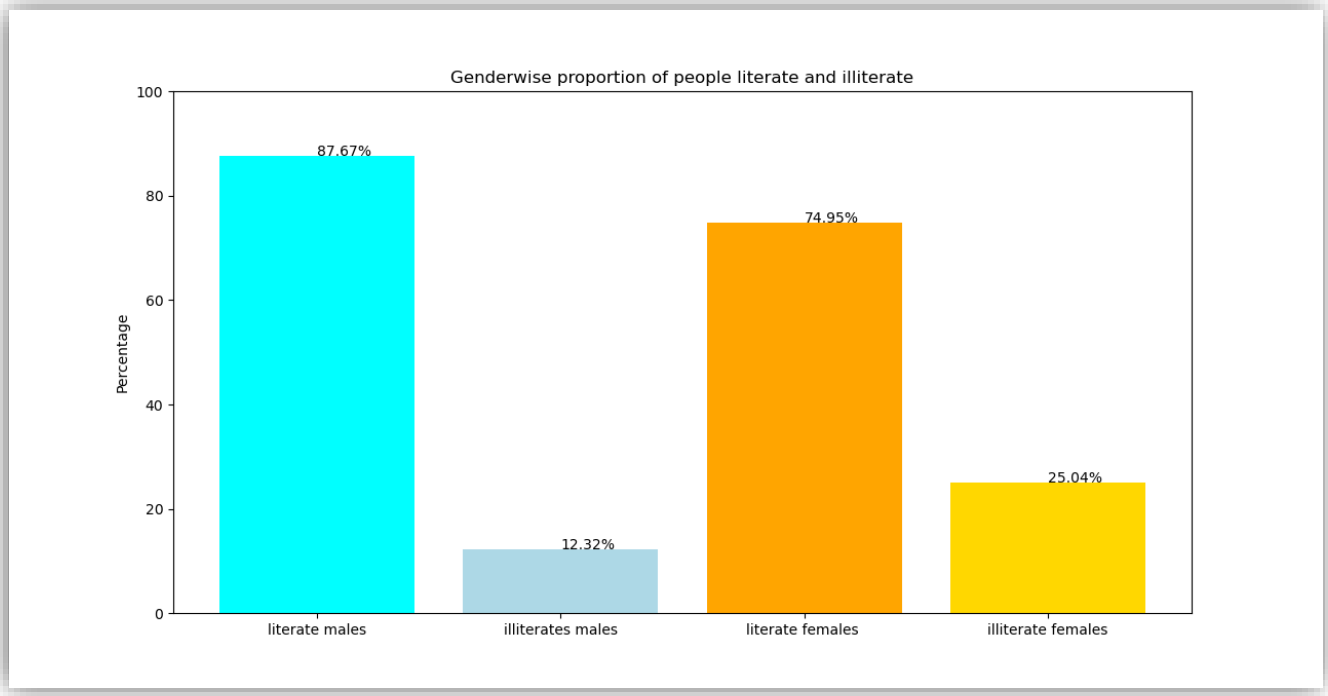
```
data = matrix(c(157273,18268,193170,50801), nrow = 2, byrow = TRUE)
colnames(data) = c("literates", "illiterates")
rownames(data) = c("Urban", "rural")
print(data)
```

```
##       literates illiterates
## Urban    157273       18268
## rural    193170       50801
```

```
test = chisq.test(data)
print(test)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  data
## X-squared = 8052.1, df = 1, p-value < 2.2e-16
```

Since, the P value is smaller than 0.05, so we reject the null hypothesis and accept the alternative hypothesis that there exists a significant association between literacy and areas of India.




### 3.2.1.2 Gender wise dependency on literacy

$H_0$: This associativity test checks for associativity between literacy and gender.

$H_A$ : There exists an association between literacy and gender

```
data = matrix(c(186725,26259,157272,52556), nrow = 2, byrow = TRUE)
colnames(data) = c("literates", "illiterates")
rownames(data) = c("Male", "Female")
print(data)
```

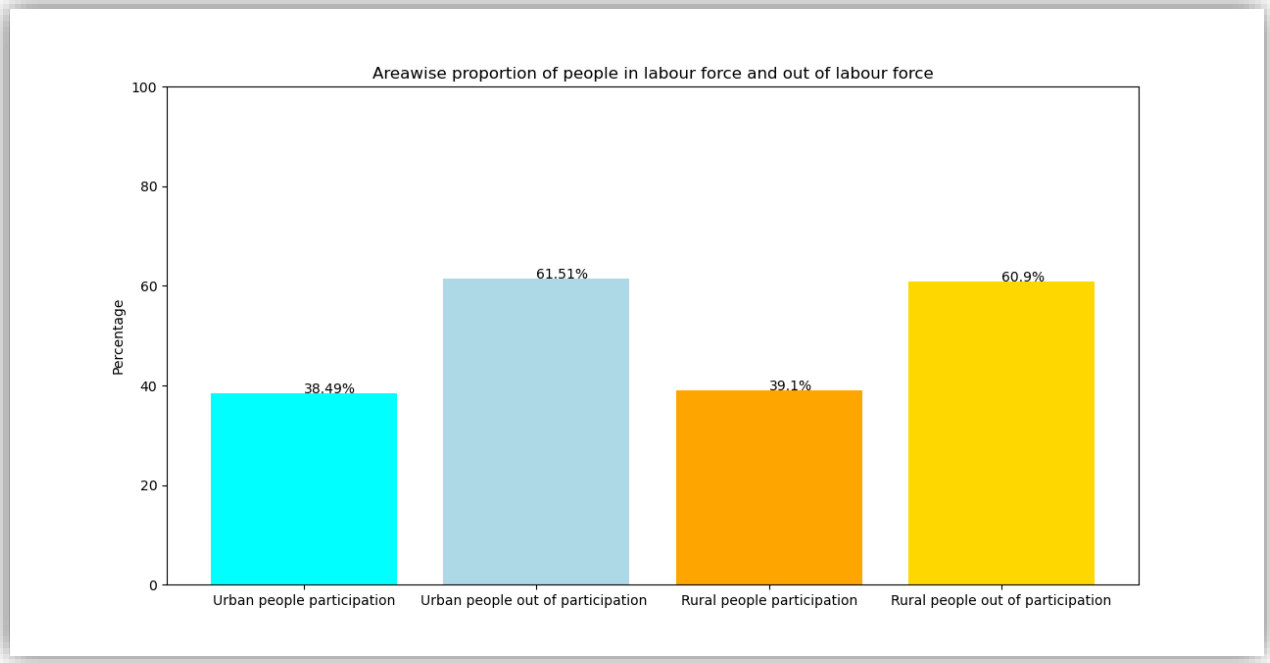
```
##        literates illiterates
## Male      186725       26259
## Female    157272       52556
```

```r
test = chisq.test(data)
print(test)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  data
## X-squared = 11272, df = 1, p-value < 2.2e-16
```

The P value of the above performed associativity test is less than 0.05. So, we reject the null hypothesis and accept the alternative hypothesis that there exists a significant association between literacy and gender.

## 4.2.2 Assessment of Labour Force Participation




### 4.2.2.1 Area wise Dependency on Labour Force Particiption Rate

In this particular associativity test, we are going check the association between Labour force participation Area wise (Urban and rural).

**$H_0$**: There is no association between Labour Force Participation and areas of India.

**$H_A$:** There exists an association between Labour Force participation and areas of India

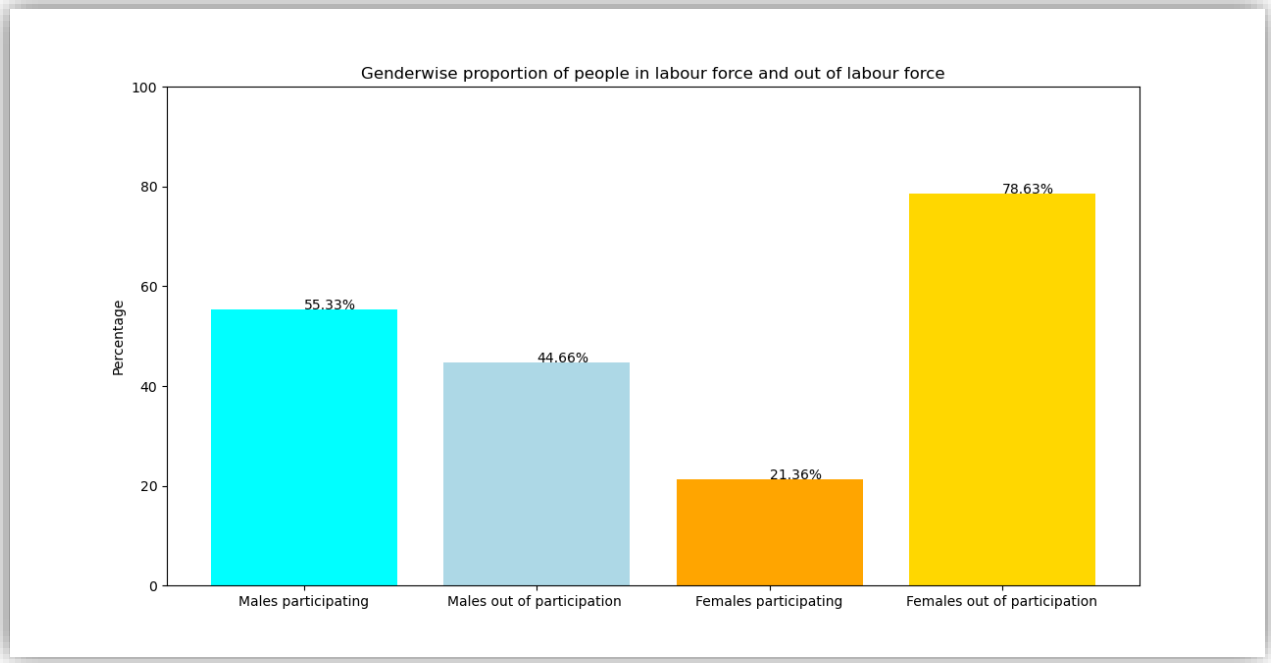
```
data = matrix(c(67583,107958,95393,148578), nrow = 2, byrow = TRUE)
colnames(data) = c("Participation", "No Participation")
rownames(data) = c("Urban", "rural")
print(data)
```

```
##       Participation No Participation
## Urban         67583           107958
## rural         95393           148578
```

```
test = chisq.test(data)
print(test)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  data
## X-squared = 15.46, df = 1, p-value = 8.426e-05
```

Since, the P value is smaller than 0.05, so we reject the null hypothesis and accept the alternative hypothesis that there exists a significant association between labour Force participation and areas of INDIA.




### 4.2.2.2 Gender wise dependency on labour Force Participation

This associativity test checks for associativity between labour Force Participation and gender.

**$H_0$**: There is no association between labour Force Participation and gender.

**$H_A$**: There exists an association between labour Force Participation and gender

```
data = matrix(c(118020,95264,44060,162168), nrow = 2, byrow = TRUE)
colnames(data) = c("Participation", "No Participation")
rownames(data) = c("Male", "Female")
print(data)
```

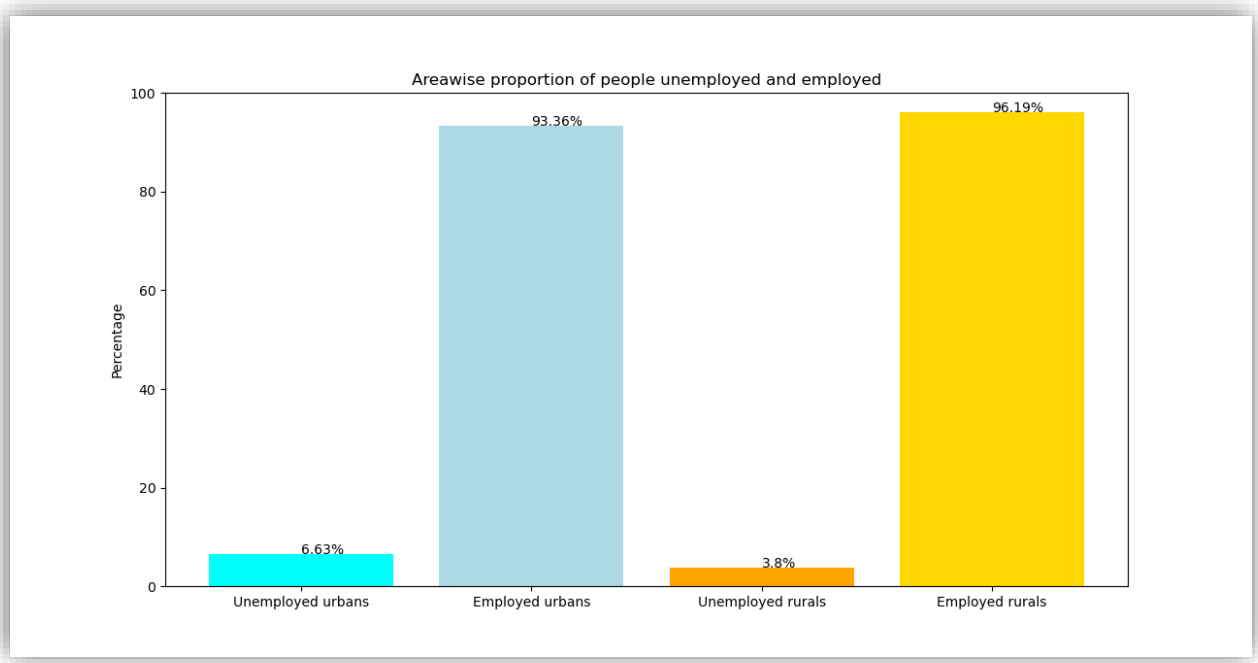
```
##        Participation No Participation
## Male          118020            95264
## Female         44060           162168
```

```
test = chisq.test(data)
print(test)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  data
## X-squared = 51031, df = 1, p-value < 2.2e-16
```

The P value of the above performed associativity test is less than 0.05. So, we reject the null hypothesis and accept the alternative hypothesis that there exists a significant association between labour Force Participation and gender.

## 3.2.3 Assessment of unemployment rate




### 3.2.3.1 Area wise Dependency on Unemployment

In this particular associativity test, we are going check the association between Unemployment Area wise (Urban and rural).

**$H_0$**: There is no association between Unemployment and areas of India.

**$H_A$**: There exists an association between Unemployment and areas of India.

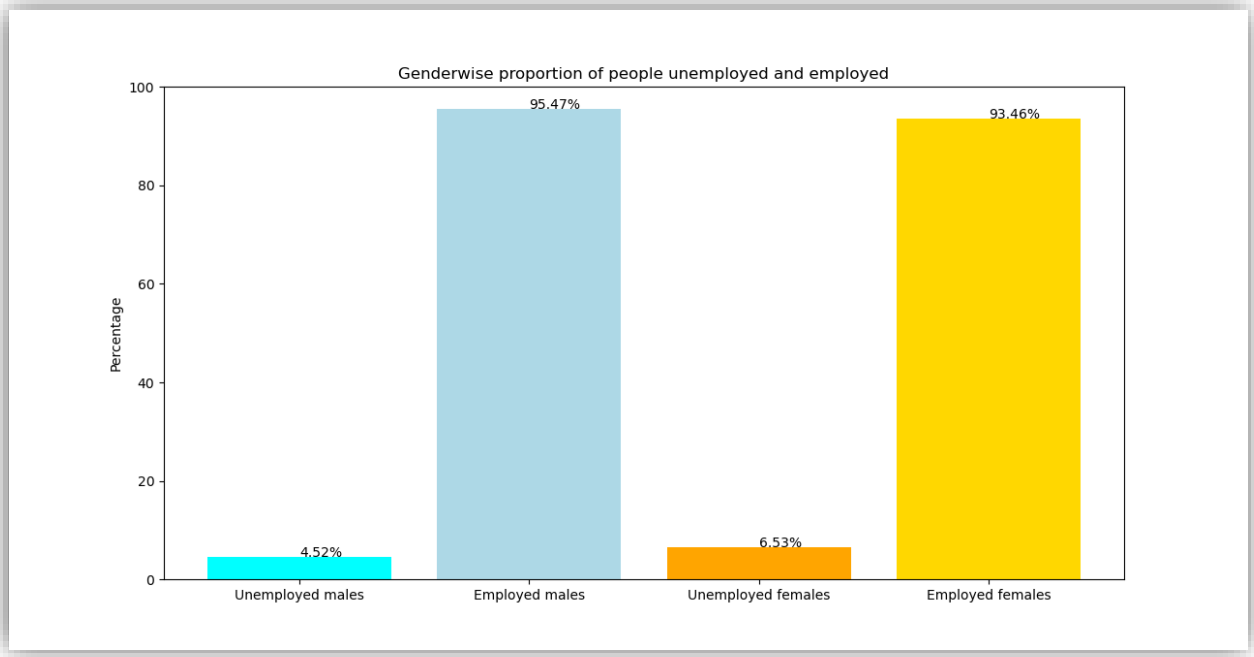
```
data = matrix(c(11652,163889,9278,234693), nrow = 2, byrow = TRUE)
colnames(data) = c("Unemployed", "Not unemployed")
rownames(data) = c("Urban", "rural")
print(data)

##       Unemployed Not unemployed
## Urban      11652         163889
## rural       9278         234693

test = chisq.test(data)
print(test)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  data
## X-squared = 1730.2, df = 1, p-value < 2.2e-16
```

Since, the P value is smaller than 0.05 so we reject the null hypothesis and accept the alternative hypothesis that there exists a significant association between Unemployment and areas of INDIA.




### 3.2.3.2 Gender wise Dependency on Unemployment

This associativity test checks for associativity between unemployment and gender.

**$H_O$**: There is no association between unemployment and gender.

**$H_A$**: There exists an association between unemployment and gender

```
data = matrix(c(9653,203631,13475,192753), nrow = 2, byrow = TRUE)
colnames(data) = c("Unemployed", "Not Unemployed")
rownames(data) = c("Male", "Female")
print(data)
```

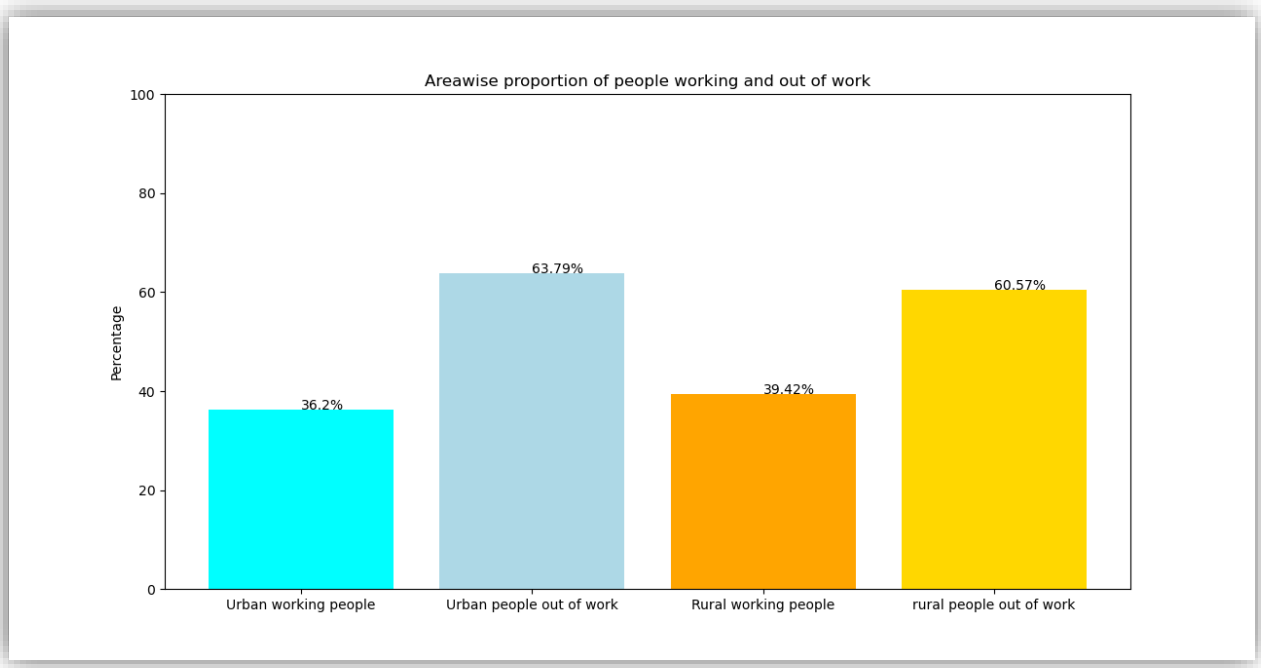
```
##        Unemployed Not Unemployed
## Male         9653         203631
## Female      13475         192753
```

```r
test = chisq.test(data)
print(test)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  data
## X-squared = 811.29, df = 1, p-value < 2.2e-16
```

The P value of the above-performed associativity test is less than 0.05. So, we reject the null hypothesis and accept the alternative hypothesis that there exists a significant association between Unemployment and gender.

## 3.2.4 Assessment of Working Population




### 3.2.4.1 Areawise dependency on Worker Population

In this particular associativity test, we are going check the association between Worker Population Area wise (Urban and rural).

**$H_0$**: There is no association between Worker Population and areas of India.

**$H_A$**: There exists an association between Worker Population and areas of India.

```
data = matrix(c(63546,111995,96197,147774), nrow = 2, byrow = TRUE)
colnames(data) = c("Working", "Not Working")
rownames(data) = c("Urban", "rural")
print(data)
```

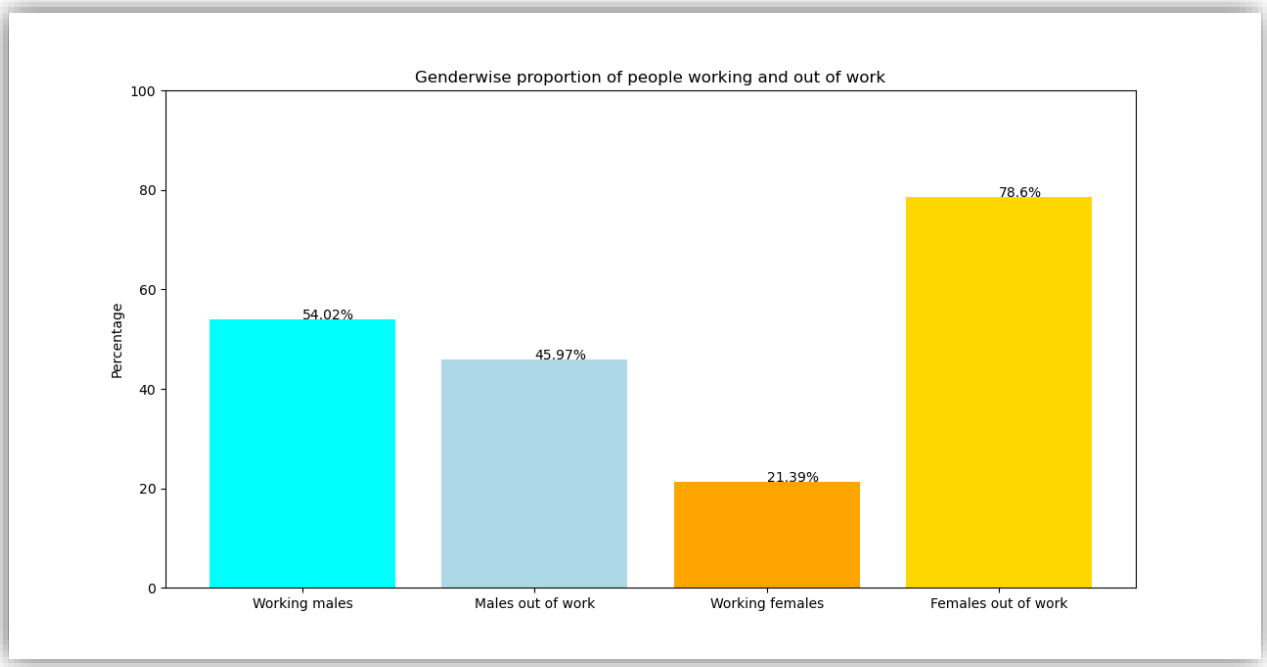
```
##       Working Not Working
## Urban   63546       111995
## rural   96197       147774
```

```
test = chisq.test(data)
print(test)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
```

```
## data:  data
## X-squared = 451.46, df = 1, p-value < 2.2e-16
```

Since, the P value is smaller than 0.05 so we reject the null hypothesis and accept the alternative hypothesis that there exists a significant association between Worker Population and areas of India.




### 3.2.4.2 Gender wise dependency on Worker Population

This associativity test checks for associativity between Worker Population and gender.

**$H_0$**: There is no association between Worker Population and gender.

**$H_A$**: There exists an association between Worker Population and gender

```
data = matrix(c(115222,98063,44114,162114), nrow = 2, byrow = TRUE)
colnames(data) = c("Working", "Not Working")
rownames(data) = c("Male", "Female")
print(data)
```

```
##        Working Not Working
## Male    115222       98063
## Female   44114      162114
```

```
test = chisq.test(data)
print(test)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
```

```
## data:  data
## X-squared = 47395, df = 1, p-value < 2.2e-16
```

The P value of the above-performed associativity test is less than 0.05. So, we reject the null hypothesis and accept the alternative hypothesis that there exists a significant association between Worker Population and gender.

## 3.3.1 Literacy Among States/U.T.

### *3.3.1.1 Area-wise proportion test on States/U.T. of India*

In this section we are going to perform a two sample proportion test on the rural and urban regions of each state. To continue with the test let us look at some of the notations used:

- $H_0$ - Null Hypothesis
- $H_1$ - Alternative Hypothesis
- $P_{L,R}$ - Proportion of literate people in rural area
- $P_{L,U}$ - Proportion of literate people in urban area

#### *Andhra Pradesh*

For Andhra Pradesh we have the following table representing the number of literate people along with number of people surveyed in rural and urban region as follows:

```
##       No. of Literate People Surveyed
## Rural             5189             7676
## Urban             4910             5980
```

$$H_0 : P_{L,R} = P_{L,U}$$

$$H_1 : P_{L,R} < P_{L,U}$$

```
prop.test(c(5189,4910),c(7676,5980),conf.level = 0.95,alternative = "less")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(5189, 4910) out of c(7676, 5980)
## X-squared = 366.47, df = 1, p-value < 2.2e-16
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1.0000000 -0.1329322
## sample estimates:
##    prop 1    prop 2
## 0.6760031 0.8210702
```

From the above proportion test, we may find that the null hypothesis is rejected and it is found that proportion of literate people in rural areas is less as compared to proportion of literate people in urban areas of Andhra Pradesh.

Similarly, the proportion test for the rest of the states is as follows:

| States | P-Value | Null Hypotheses |
|---|---|---|
| Arunachal Pradesh | 2.20E-16 | Rejected |
| Assam | 2.20E-16 | Rejected |
| Bihar | 2.20E-16 | Rejected |
| Chhattisgarh | 2.20E-16 | Rejected |
| Delhi | 5.61E-01 | Accepted |
| Goa | 1.13E-01 | Accepted |
| Gujarat | 2.20E-16 | Rejected |
| Haryana | 2.20E-16 | Rejected |
| Himachal Pradesh | 3.09E-05 | Rejected |
| Jharkhand | 2.20E-16 | Rejected |
| Karnataka | 2.20E-16 | Rejected |
| Kerala | 2.20E-16 | Rejected |
| Madhya Pradesh | 2.20E-16 | Rejected |
| Maharashtra | 2.20E-16 | Rejected |
| Manipur | 1.64E-12 | Rejected |
| Meghalaya | 4.68E-07 | Rejected |
| Mizoram | 3.64E-05 | Rejected |
| Nagaland | 1.79E-09 | Rejected |
| Odisha | 2.20E-16 | Rejected |
| Punjab | 2.20E-16 | Rejected |
| Rajasthan | 2.20E-16 | Rejected |
| Sikkim | 2.20E-16 | Rejected |
| Tamil Nadu | 2.20E-16 | Rejected |
| Telangana | 2.20E-16 | Rejected |
| Tripura | 1.17E-10 | Rejected |
| Uttarakhand | 4.96E-07 | Rejected |
| Uttar Pradesh | 2.20E-16 | Rejected |
| West Bengal | 2.20E-16 | Rejected |
| Andaman & N. Island | 0.002534 | Rejected |
| Chandigarh | 0.4975 | Accepted |
| Dadra & Nagar Haveli & Daman & Diu | 6.41E-07 | Rejected |
| Jammu & Kashmir | 2.20E-16 | Rejected |
| Ladakh | 3.40E-08 | Rejected |
| Lakshadweep | 0.9845 | Accepted |
| Puducherry | 2.15E-13 | Rejected |

### 3.3.1.2 Area-wise proportion test of literacy for all India

For overall India we have the following table representing the number of literate people along with number of people surveyed in rural and urban region as follows:

```
##       No. of Literate People Surveyed
## Rural                 193170                  243971
## Urban                 157273                  175541
```

$$H_0 : P_{L,R} = P_{L,U}$$

$$H_1 : P_{L,R} < P_{L,U}$$

```
prop.test(c(193170,157273),c(243971,175541),conf.level = 0.95,alternative = "
less")

##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(193170, 157273) out of c(243971, 175541)
## X-squared = 8052.1, df = 1, p-value < 2.2e-16
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1.0000000 -0.1023468
## sample estimates:
##    prop 1    prop 2
## 0.7917744 0.8959331
```

From the above proportion test, we may find that the null hypothesis is rejected and it is found that proportion of literate people in rural areas is less as compared to proportion of literate people in urban areas of India.

## 3.3.2 Labour Force Participation Among States/U.T.

### *3.3.2.1 Gender-wise proportion test of Labour Force Participation on States/U.T. of India*

In this section we are going to perform a two sample proportion test on men and women of each state. To continue with the test let us understand some notations used:

- $H_0$ - Null Hypothesis
- $H_1$ - Alternative Hypothesis
- $P_{L,M}$ - Proportion of men in labour
- $P_{L,W}$ - Proportion of women in labour

#### *Andhra Pradesh*

For Andhra Pradesh we have the following table representing the number of people participating in labour along with number of people surveyed among men and women as follows:

```
##       No. of people in Labour People Surveyed
## Men                      4119            6820
## Women                    2400            6837
```

$$H_0 : P_{L,M} = P_{L,W}$$

$$H_1 : P_{L,M} > P_{L,W}$$

```
prop.test(c(4119,2400),c(6820,6837),conf.level = 0.95,alternative = "greater"
)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(4119, 2400) out of c(6820, 6837)
## X-squared = 874.46, df = 1, p-value < 2.2e-16
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.2391785 1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.6039589 0.3510312
```

From the above proportion test, we may find that the null hypothesis is rejected and it is found that proportion of men in labour is greater as compared to proportion of women in labour in Andhra Pradesh.

Similarly, we have done the proportion tests for other states as well:

| States | P-Value | Null Hypotheses |
|---|---|---|
| Arunachal Pradesh | 2.20E-16 | Rejected |
| Assam | 2.20E-16 | Rejected |
| Bihar | 2.20E-16 | Rejected |
| Chhattisgarh | 2.20E-16 | Rejected |
| Delhi | 2.20E-16 | Rejected |
| Goa | 2.20E-16 | Rejected |
| Gujarat | 2.20E-16 | Rejected |
| Haryana | 2.20E-16 | Rejected |
| Himachal Pradesh | 6.84E-13 | Rejected |
| Jharkhand | 2.20E-16 | Rejected |
| Karnataka | 2.20E-16 | Rejected |
| Kerala | 2.20E-16 | Rejected |
| Madhya Pradesh | 2.20E-16 | Rejected |
| Maharashtra | 2.20E-16 | Rejected |
| Manipur | 2.20E-16 | Rejected |
| Meghalaya | 2.20E-16 | Rejected |
| Mizoram | 2.20E-16 | Rejected |
| Nagaland | 2.20E-16 | Rejected |
| Odisha | 2.20E-16 | Rejected |
| Punjab | 2.20E-16 | Rejected |
| Rajasthan | 2.20E-16 | Rejected |
| Sikkim | 2.58E-08 | Rejected |
| Tamil Nadu | 2.20E-16 | Rejected |
| Telangana | 2.20E-16 | Rejected |
| Tripura | 2.20E-16 | Rejected |
| Uttarakhand | 2.20E-16 | Rejected |
| Uttar Pradesh | 2.20E-16 | Rejected |
| West Bengal | 2.20E-16 | Rejected |
| Andaman & N. Island | 2.20E-16 | Rejected |
| Chandigarh | 2.20E-16 | Rejected |
| Dadra & Nagar Haveli & Daman & Diu | 2.20E-16 | Rejected |
| Jammu & Kashmir | 2.20E-16 | Rejected |
| Ladakh | 1.24E-05 | Rejected |
| Lakshadweep | 2.20E-16 | Rejected |
| Puducherry | 2.20E-16 | Rejected |

### 3.3.2.2 Gender-wise proportion test of Labour Force Participation for all India

For overall India we have the following table representing the number of people participating in labour along with number of people surveyed among men and women as follows:

```
##       No. of people in labour People Surveyed
## Men                    119545          213285
## Women                   49403          206228
```

$$H_0 : P_{L,M} = P_{L,W}$$

$$H_1 : P_{L,M} > P_{L,W}$$

```
prop.test(c(119545,49403),c(213285,206228),conf.level = 0.95,alternative = "g
reater")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(119545, 49403) out of c(213285, 206228)
## X-squared = 44896, df = 1, p-value < 2.2e-16
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.3185858 1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.5604942 0.2395552
```

From the above proportion test, we may find that the null hypothesis is rejected and it is found that proportion of men in labour is greater as compared to proportion of women in labour in India.

## 3.3.3 Unemployment Among States/U.T.

### *3.3.3.1 Area-wise proportion test on States/U.T. of India*

In this section we are going to perform a two sample proportion test on the rural and urban regions of each state. To continue with the test let us look at some of the notations used:

- $H_0$ - Null Hypothesis
- $H_1$ - Alternative Hypothesis
- $P_{U,R}$ - Proportion of unemployed people in rural area
- $P_{U,U}$ - Proportion of unemployed people in urban area

#### *Andhra Pradesh*

For Andhra Pradesh we have the following table representing the number of unemployed people along with number of people surveyed in rural and urban region as follows:

```
##       No. of Unemployed People Surveyed
## Rural                284            7676
## Urban                425            5980
```

$$H_0 : P_{U,R} = P_{U,U}$$

$$H_1 : P_{U,R} \neq P_{U,U}$$

```
prop.test(c(284,425),c(7676,5980),conf.level = 0.95,alternative = "two.sided"
)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(284, 425) out of c(7676, 5980)
## X-squared = 78.584, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.04198203 -0.02616157
## sample estimates:
##     prop 1     prop 2
## 0.03699844 0.07107023
```

From the above proportion test, we may find that the null hypothesis is rejected and it is found that proportion of unemployed people in rural areas is not equal to proportion of literate people in urban areas of Andhra Pradesh.

Similarly, we have carried out the proportion test for other states as follows:

| States | P-Value | Null Hypotheses |
|---|---|---|
| Arunachal Pradesh | 2.20E-16 | Rejected |
| Assam | 2.20E-16 | Rejected |
| Bihar | 2.20E-16 | Rejected |
| Chhattisgarh | 2.20E-16 | Rejected |
| Delhi | 0.8781 | Accepted |
| Goa | 0.09639 | Accepted |
| Gujarat | 0.0004211 | Rejected |
| Haryana | 0.6914 | Accepted |
| Himachal Pradesh | 6.84E-13 | Rejected |
| Jharkhand | 2.20E-16 | Rejected |
| Karnataka | 2.20E-16 | Rejected |
| Kerala | 0.001374 | Rejected |
| Madhya Pradesh | 2.20E-16 | Rejected |
| Maharashtra | 2.20E-16 | Rejected |
| Manipur | 0.07507 | Accepted |
| Meghalaya | 2.20E-16 | Rejected |
| Mizoram | 4.50E-08 | Rejected |
| Nagaland | 2.20E-16 | Rejected |
| Odisha | 4.83E-08 | Rejected |
| Punjab | 0.3752 | Accepted |
| Rajasthan | 2.20E-16 | Rejected |
| Sikkim | 0.7868 | Accepted |
| Tamil Nadu | 0.7546 | Accepted |
| Telangana | 2.20E-16 | Rejected |
| Tripura | 1.93E-08 | Rejected |
| Uttarakhand | 0.0008223 | Rejected |
| Uttar Pradesh | 2.20E-16 | Rejected |
| West Bengal | 2.20E-16 | Rejected |
| Andaman & N. Island | 0.0004972 | Rejected |
| Chandigarh | 0.6347 | Accepted |
| Dadra & Nagar Haveli & Daman & Diu | 0.0008858 | Rejected |
| Jammu & Kashmir | 2.20E-16 | Rejected |
| Ladakh | 0.004857 | Rejected |
| Lakshadweep | 0.008935 | Rejected |
| Puducherry | 0.07731 | Accepted |

### 3.3.3.2 Area-wise proportion test of unemployment for all India

For overall India we have the following table representing the number of unemployed people along with number of people surveyed in rural and urban region as follows:

```
##       No. of unemployed People Surveyed
## Rural              9278          243971
## Urban             11652          175541
```

$$H_0 : P_{U,R} = P_{U,U}$$

$$H_1 : P_{U,R} \neq P_{U,U}$$

```
prop.test(c(9278,11652),c(243971,175541),conf.level = 0.95,alternative = "two
.sided")

##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(9278, 11652) out of c(243971, 175541)
## X-squared = 1730.2, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.02974347 -0.02695362
## sample estimates:
##     prop 1     prop 2
## 0.03802911 0.06637766
```

From the above proportion test, we may find that the null hypothesis is rejected and it is found that proportion of unemployed people in rural areas is not equal to proportion of unemployed people in urban areas of India.

## 3.3.4 Working Population Among States/U.T.

### *3.3.4.1 Gender-wise proportion test of Working Population on States/U.T. of India*

In this section we are going to perform a two sample proportion test on men and women of each state. To continue with the test let us understand some notations used:

- $H_0$ - Null Hypothesis
- $H_1$ - Alternative Hypothesis
- $P_{W,M}$ - Proportion of working men
- $P_{W,W}$ - Proportion of working women

#### *Andhra Pradesh*

For Andhra Pradesh we have the following table representing the number of people that are working along with number of people surveyed among men and women as follows:

```
##       No. of Working People People Surveyed
## Men                    3928            6820
## Women                  2297            6837
```

$$H_0 : P_{W,M} = P_{W,W}$$

$$H_1 : P_{W,M} > P_{W,W}$$

```
prop.test(c(3928,2297),c(6820,6837),conf.level = 0.95,alternative = "greater"
)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(3928, 2297) out of c(6820, 6837)
## X-squared = 791.78, df = 1, p-value < 2.2e-16
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.2262328 1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.5759531 0.3359661
```

From the above proportion test, we may find that the null hypothesis is rejected and it is found that proportion of working men is greater as compared to proportion of working women in Andhra Pradesh.

Similarly, we have carried out the proportion test on all the states as follows:

| States | P-Value | Null Hypotheses |
|---|---|---|
| Arunachal Pradesh | 2.20E-16 | Rejected |
| Assam | 2.20E-16 | Rejected |
| Bihar | 2.20E-16 | Rejected |
| Chhattisgarh | 2.20E-16 | Rejected |
| Delhi | 2.20E-16 | Rejected |
| Goa | 2.20E-16 | Rejected |
| Gujarat | 2.20E-16 | Rejected |
| Haryana | 2.20E-16 | Rejected |
| Himachal Pradesh | 2.73E-16 | Rejected |
| Jharkhand | 2.20E-16 | Rejected |
| Karnataka | 2.20E-16 | Rejected |
| Kerala | 2.20E-16 | Rejected |
| Madhya Pradesh | 2.20E-16 | Rejected |
| Maharashtra | 2.20E-16 | Rejected |
| Manipur | 2.20E-16 | Rejected |
| Meghalaya | 2.20E-16 | Rejected |
| Mizoram | 2.20E-16 | Rejected |
| Nagaland | 2.20E-16 | Rejected |
| Odisha | 2.20E-16 | Rejected |
| Punjab | 2.20E-16 | Rejected |
| Rajasthan | 2.20E-16 | Rejected |
| Sikkim | 3.64E-09 | Rejected |
| Tamil Nadu | 2.20E-16 | Rejected |
| Telangana | 2.20E-16 | Rejected |
| Tripura | 2.20E-16 | Rejected |
| Uttarakhand | 2.20E-16 | Rejected |
| Uttar Pradesh | 2.20E-16 | Rejected |
| West Bengal | 2.20E-16 | Rejected |
| Andaman & N. Island | 2.20E-16 | Rejected |
| Chandigarh | 2.20E-16 | Rejected |
| Dadra & Nagar Haveli & Daman & Diu | 2.20E-16 | Rejected |
| Jammu & Kashmir | 2.20E-16 | Rejected |
| Ladakh | 2.20E-16 | Rejected |
| Lakshadweep | 2.20E-16 | Rejected |
| Puducherry | 2.20E-16 | Rejected |

### 3.3.4.2 Gender-wise proportion test of Working People for all India

For overall India we have the following table representing the number of people that are working along with number of people surveyed among men and women as follows:

```
##       No. of Working People People Surveyed
## Men                  113041          213285
## Women                 42895          206228
```

$$H_0 : P_{W,M} = P_{W,W}$$

$$H_1 : P_{W,M} > P_{W,W}$$

```
prop.test(c(113041,42895),c(213285,206228),conf.level = 0.95,alternative = "g
reater")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(113041, 42895) out of c(213285, 206228)
## X-squared = 46548, df = 1, p-value < 2.2e-16
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.3196903 1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.5299998 0.2079979
```

From the above proportion test, we may find that the null hypothesis is rejected and it is found that proportion of working men is greater as compared to proportion of working women in India.

## 3.4 Regression Analysis

This section will examine the correlation between India's labor force participation rate and several current characteristics, such as the country's working population rate, unemployment rate, and literacy rate.

The labor force participation rate is the response variable in our regression analysis, and the dependent variables are the working population rate, the unemployment rate, and the literacy rate. To do this, we will build a basic linear regression model.

### 3.4.1 Normality check for Labour Force Participation Rate

The response variable, or labour force participation, needs to abide by the linear model's normalty assumption. Therefore, we run the Shapiro-Wilk Test to satisfy the requirement.

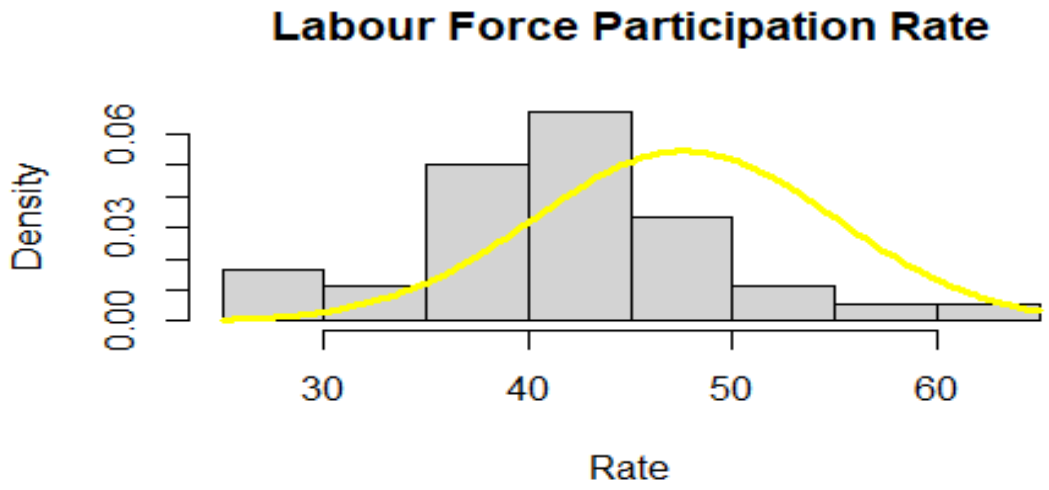
$\boldsymbol{H_0}$ : The labour Force participation rate Follows Normal Distribution

$\boldsymbol{H_1}$ : The labour force participation rate does not follow normal distribution

```
shapiro.test(data$LFPR.Rate)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$LFPR.Rate
## W = 0.97468, p-value = 0.5663
```

From the above test, we get a p-value of 0.5663, which favors null hypothesis. Therefore, null hypothesis is accepted and we can say that labour force participation rate follows normal distribution.



### 3.4.2 Labour Force Participation Rate and Literacy Rate

This model tends to explain the relationship between labour force participation rate and literacy rate.

$$Model: LFPR = \beta_0 + \beta_1 * LR + \varepsilon$$

Where we have the following notations:

- ***LFPR*** : Labour Force Participation Rate
- ***LR*** : Literacy Rate
- $\boldsymbol{\varepsilon}$ : Random Error Term

```
M1<-lm(data$LFPR.Rate~data$Literacy.Rate)
summary(M1)
```

```
##
## Call:
## lm(formula = data$LFPR.Rate ~ data$Literacy.Rate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.457  -4.971  -0.507   4.085  19.199
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         31.7498    14.3493   2.213   0.0337 *
## data$Literacy.Rate   0.1163     0.1675   0.694   0.4924
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.396 on 34 degrees of freedom
## Multiple R-squared:  0.01397,    Adjusted R-squared:  -0.01503
## F-statistic: 0.4817 on 1 and 34 DF,  p-value: 0.4924
```

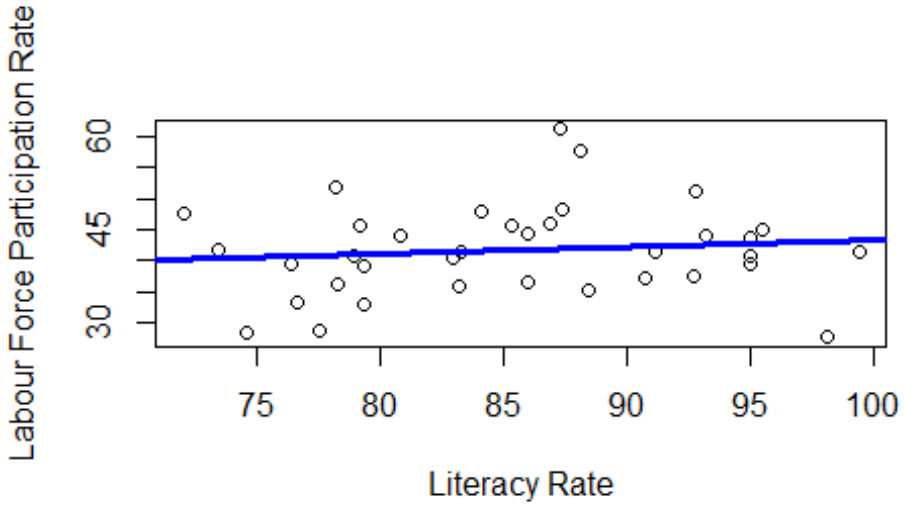
From the above summary of the model we get the values of regression coefficients as follows:

$$\beta_0 = 31.7498$$

$$\beta_1 = 0.1163$$

Therefore, the fitted model is as follows:

$$Labour\ Force\ Participation\ Rate = 31.7498 + 0.1163\ Literacy\ Rate + Error$$




### 3.4.3 Labour Force Participation Rate and Unemployment Rate

This model tends to explain the relationship between labour force participation rate and unemployment rate.

$$\text{Model} : LFPR = \beta_0 + \beta_1 * UR + \varepsilon$$

Where we have the following notations:

- $\boldsymbol{LFPR}$ : Labour Force Participation Rate
- $\boldsymbol{UR}$ : Unemployment Rate
- $\boldsymbol{\varepsilon}$ : Random Error Term

```
M2<-lm(data$LFPR.Rate~data$Unemployment.Rate)
summary(M2)

##
## Call:
## lm(formula = data$LFPR.Rate ~ data$Unemployment.Rate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.9775  -4.6220  -0.3235   4.0547  19.4296
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             4.167e+01  2.542e+00  16.389   <2e-16 ***
## data$Unemployment.Rate 7.392e-04  4.253e-01   0.002    0.999
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 7.448 on 34 degrees of freedom
## Multiple R-squared:  8.883e-08,  Adjusted R-squared:  -0.02941
## F-statistic: 3.02e-06 on 1 and 34 DF,  p-value: 0.9986
```

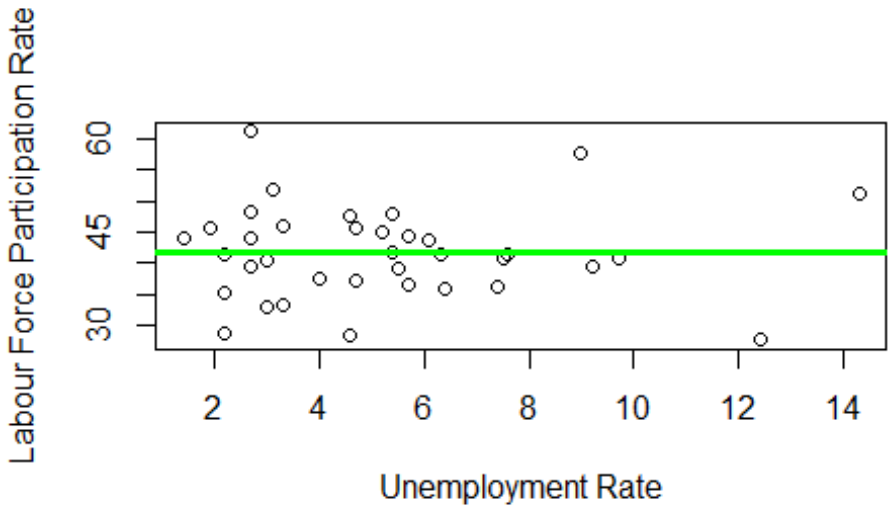
From the above summary of the model we get the values of regression coefficients as follows:

$$\beta_0 = 41.67$$

$$\beta_1 = 0.00074$$

Therefore, the fitted model is as follows:

$$Labour\ Force\ Participation\ Rate = 41.67 + 0.00074\ Unemployment\ Rate + Error$$




### 3.4.4 Labour Force Participation Rate and Working Rate

This model tends to explain the relationship between labour force participation rate and Working Population rate.

$$\text{Model}: LFPR = \beta_0 + \beta_1 * WR + \varepsilon$$

Where we have the following notations:

- ***LFPR*** : Labour Force Participation Rate
- ***WR*** : Working Rate
- $\boldsymbol{\varepsilon}$ : Random Error Term

```
M3<-lm(data$LFPR.Rate~data$Working.Rate)
summary(M3)
```

```
##
## Call:
## lm(formula = data$LFPR.Rate ~ data$Working.Rate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5750 -1.0187 -0.1061  0.4439  5.0083
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        1.24137    1.30991   0.948     0.35
## data$Working.Rate  1.02393    0.03267  31.340   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.362 on 34 degrees of freedom
## Multiple R-squared:  0.9665, Adjusted R-squared:  0.9656
## F-statistic: 982.2 on 1 and 34 DF,  p-value: < 2.2e-16
```

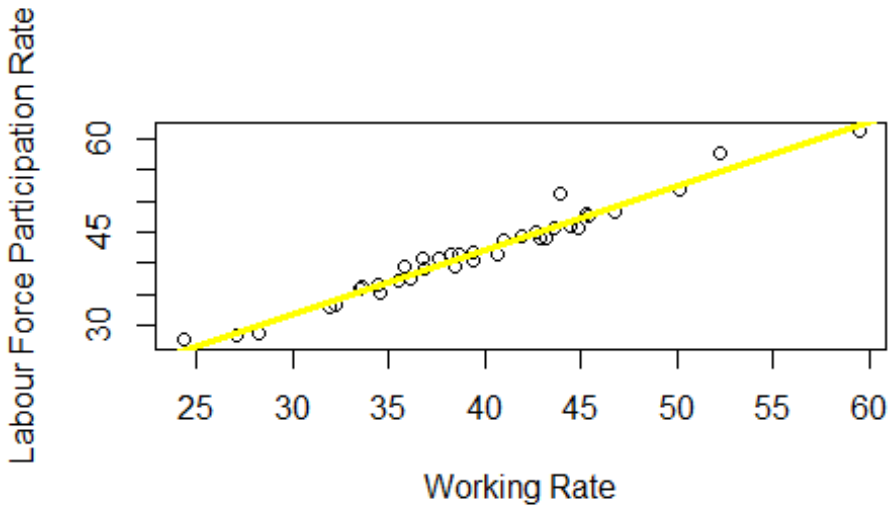
From the above summary of the model we get the values of regression coefficients as follows:

$$\beta_0 = 1.2413$$

$$\beta_1 = 1.0239$$

Therefore, the fitted model is as follows:

$$\textit{Labour Force Participation Rate} = 1.2413 + 1.0239\ \textit{Working Rate} + \textit{Error}$$



# Results & Findings

## Literacy Rate

- The area-wise associativity test resulted in supporting the argument that literacy differs across different areas of India.
- The gender-wise associativity tests reveals that there are links between gender and literacy in India.
- The state-wise proportion test concludes that in every state and union territory of India except Delhi (p-value = 0.561), Goa (p-value = 0.113), Chandigarh (p-value = 0.4975) and Lakshadweep (p-value = 0.9845), the proportion of literates in rural areas are comparatively less than the proportion of literates in urban areas.
- The two sample proportion test on rural and urban areas of India verifies the above claim, that proportion of literates in rural areas are less as compared to proportion of literates in urban areas

## Labour Force Participation Rate

- The area-wise associativity test resulted in supporting the argument that labour force participation differs across different regions of India.
- The gender-wise associativity tests reveals that there are links between gender and labour force participation in India.
- The state-wise proportion test supports the alternative hypothesis that major proportion of labour is composed of men in each state.
- The two sample proportion test on men and women in India were found to support the above claim, that proportion of men in labour force exceeds proportion of women in labour force.

## Unemployment Rate

- The associativity test between unemployment and areas (Rural & Urban) results in favouring the link among the two quantities.
- The gender-wise associativity report suggests that unemployment is somewhat linked to classification in genders.
- The state-wise proportion test supports the argument that in every state proportion of unemployed people in rural and urban is not equal except in the states Delhi (p-value = 0.8781), Goa (p-value =

0.09369), Haryana (p-value = 0.6914), Manipur (p-value = 0.07507), Punjab (p-value = 0.3752), Sikkim (p-value = 0.7868), Tamil Nadu (p-value = 0.7546), Chandigarh (p-value = 0.6347), and Puducherry (p-value = 0.07731).

- The two sample proportion test on rural and urban area of India reveals that proportion of unemployed vary across the two regions.

## Working Population Ratio

- The area-wise associativity test between working and area-wise categorization supports the hypothesis that there is a link between the two categories.
- The gender-wise associativity test summarizes that there is an association between gender and work participation.
- The state-wise proportion test claims that in each state men are the major contributors in working composition.
- The two sample proportion test states that in India, majority of men are working.

## Relationship Between Labour Force Participation Rate and Literacy Rate

*Labour Force Participation Rate = 31.7498 + 0.1163*Literacy Rate*

## Relationship Between Labour Force Participation Rate and Unemployment Rate

*Labour Force Participation Rate = 41.67 + 0.000074*Unemployment Rate*

## Relationship Between Labour Force Participation Rate and Working Rate

*Labour Force Participation Rate = 1.2413 + 1.0239*Working Rate*

# References

- Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkata.

- Miller, Irwin and Miller, Marylees (2006): John E. Freund's Mathematical Statistics with Applications, (7th Edn.), Pearson Education, Asia.

- Data Science from Scratch: First Principles with Python, Author: Joel Grus

- Montgomery, D. C., Peck, E. A. and Vining, G. G. (2012): Introduction to Linear Regression Analysis (Fifth edition), Wiley.

- Braun W J, Murdoch D J (2007): A First Course in Statistical Programming with R. Cambridge University Press. New York

# Appendix

# Literacy Rate Area-Wise

| State/UT | Total Persons Surveyed Rural | No. Of Rural Literate Persons (7) | Urban Person Surveyed Urban | No. of Literate Persons (7) Urban |
|---|---|---|---|---|
| Andhra Pradesh | 7676 | 5189 | 5980 | 4910 |
| Arunachal Pradesh | 5202 | 4307 | 2612 | 2380 |
| Assam | 12176 | 9351 | 3305 | 3123 |
| Bihar | 15808 | 11650 | 6057 | 5136 |
| Chhattisgarh | 6260 | 4770 | 3396 | 2934 |
| Delhi | 218 | 194 | 3320 | 2935 |
| Goa | 705 | 664 | 937 | 896 |
| Gujarat | 8174 | 6629 | 7810 | 7076 |
| Haryana | 6340 | 5078 | 5056 | 4500 |
| Himachal Pradesh | 5860 | 5139 | 1137 | 1045 |
| Jharkhand | 8001 | 6193 | 4377 | 3856 |
| Karnataka | 7464 | 5620 | 8107 | 7345 |
| Kerala | 7219 | 6735 | 8618 | 8359 |
| Madhya Pradesh | 12704 | 9223 | 8221 | 7210 |
| Maharashtra | 14804 | 12287 | 16276 | 15006 |
| Manipur | 4950 | 4430 | 4671 | 4367 |
| Meghalaya | 4408 | 4170 | 1773 | 1729 |
| Mizoram | 2379 | 2358 | 4602 | 4593 |
| Nagaland | 2609 | 2468 | 1733 | 1702 |
| Odisha | 11709 | 9098 | 3819 | 3395 |
| Punjab | 6546 | 5302 | 6363 | 5568 |
| Rajasthan | 12448 | 8664 | 7554 | 6436 |
| Sikkim | 1632 | 1389 | 944 | 903 |
| Tamil Nadu | 11361 | 9396 | 11698 | 10575 |
| Telangana | 4767 | 3461 | 5444 | 4932 |
| Tripura | 4790 | 4436 | 2309 | 2228 |
| Uttarakhand | 3877 | 3292 | 2963 | 2637 |
| Uttar Pradesh | 28242 | 21153 | 14611 | 12186 |
| West Bengal | 13740 | 11088 | 11106 | 9773 |
| Andaman & N. Island | 838 | 764 | 811 | 769 |
| Chandigarh | 224 | 207 | 1081 | 1002 |
| Dadra & Nagar Haveli & Daman & Diu | 593 | 482 | 657 | 597 |
| Jammu & Kashmir | 8865 | 6826 | 5684 | 4797 |
| Ladakh | 693 | 539 | 453 | 409 |
| Lakshadweep | 237 | 237 | 562 | 548 |
| Puducherry | 452 | 381 | 1494 | 1416 |
| all India | 243971 | 193170 | 175541 | 157273 |

# Literacy Rate Gender-Wise

| State/UT | No. of Literate Male (7) | No. of Male Surveyed | No. of Literate Female (7) | No. of Female Surveyed |
|---|---|---|---|---|
| Andhra Pradesh | 5381 | 6820 | 4465 | 6837 |
| Arunachal Pradesh | 3511 | 3994 | 3064 | 3820 |
| Assam | 6345 | 8072 | 5660 | 7409 |
| Bihar | 9455 | 11433 | 6875 | 10432 |
| Chhattisgarh | 4115 | 4864 | 3421 | 4792 |
| Delhi | 1738 | 1873 | 1387 | 1665 |
| Goa | 810 | 823 | 750 | 819 |
| Gujarat | 7554 | 8283 | 6084 | 7701 |
| Haryana | 5473 | 6014 | 4015 | 5382 |
| Himachal Pradesh | 3231 | 3434 | 2929 | 3563 |
| Jharkhand | 5406 | 6271 | 4434 | 6107 |
| Karnataka | 6871 | 7934 | 5712 | 7637 |
| Kerala | 7260 | 7516 | 7797 | 8321 |
| Madhya Pradesh | 9166 | 10860 | 6834 | 10065 |
| Maharashtra | 14673 | 15914 | 12345 | 15166 |
| Manipur | 4483 | 4764 | 4230 | 4857 |
| Meghalaya | 2862 | 2987 | 3009 | 3194 |
| Mizoram | 3476 | 3490 | 3467 | 3491 |
| Nagaland | 2071 | 2153 | 2075 | 2189 |
| Odisha | 6616 | 7783 | 5700 | 7745 |
| Punjab | 5794 | 6637 | 4967 | 6272 |
| Rajasthan | 8636 | 10293 | 6068 | 9709 |
| Sikkim | 1209 | 1330 | 1037 | 1246 |
| Tamil Nadu | 10421 | 11540 | 9411 | 11519 |
| Telangana | 4439 | 5210 | 3661 | 5001 |
| Tripura | 3426 | 3587 | 3196 | 3512 |
| Uttarakhand | 3185 | 3458 | 2702 | 3382 |
| Uttar Pradesh | 18421 | 21852 | 14470 | 21001 |
| West Bengal | 10848 | 12541 | 9746 | 12305 |
| Andaman & N. Island | 804 | 848 | 726 | 801 |
| Chandigarh | 651 | 685 | 560 | 620 |
| Dadra & Nagar Haveli & Daman & Diu | 589 | 647 | 500 | 603 |
| Jammu & Kashmir | 6397 | 7456 | 5001 | 7093 |
| Ladakh | 552 | 628 | 355 | 518 |
| Lakshadweep | 370 | 373 | 414 | 426 |
| Puducherry | 871 | 918 | 901 | 1028 |
| all India | 187110 | 213285 | 157968 | 206228 |

## Labour Force Participation Rate Gender-Wise

| State/UT | NO. of Male in Labour | Total Male Surveyed | NO. of Female in Labour | Female Surveyed |
|---|---|---|---|---|
| Andhra Pradesh | 4119 | 6820 | 2400 | 6837 |
| Arunachal Pradesh | 2213 | 3994 | 1536 | 3820 |
| Assam | 3374 | 8072 | 704 | 7409 |
| Bihar | 5282 | 11433 | 949 | 10432 |
| Chhattisgarh | 2967 | 4864 | 2017 | 4792 |
| Delhi | 1060 | 1873 | 185 | 1665 |
| Goa | 488 | 823 | 187 | 819 |
| Gujarat | 5152 | 8283 | 2172 | 7701 |
| Haryana | 3284 | 6014 | 813 | 5382 |
| Himachal Pradesh | 2119 | 3434 | 1899 | 3563 |
| Jharkhand | 3129 | 6271 | 965 | 6107 |
| Karnataka | 4673 | 7934 | 2192 | 7637 |
| Kerala | 4179 | 7516 | 2055 | 8321 |
| Madhya Pradesh | 6494 | 10860 | 1802 | 10065 |
| Maharashtra | 9612 | 15914 | 4717 | 15166 |
| Manipur | 2492 | 4764 | 1093 | 4857 |
| Meghalaya | 1476 | 2987 | 1220 | 3194 |
| Mizoram | 1766 | 3490 | 1138 | 3491 |
| Nagaland | 1178 | 2153 | 786 | 2189 |
| Odisha | 4444 | 7783 | 1634 | 7745 |
| Punjab | 4102 | 6637 | 1254 | 6272 |
| Rajasthan | 5527 | 10293 | 2825 | 9709 |
| Sikkim | 880 | 1330 | 693 | 1246 |
| Tamil Nadu | 6739 | 11540 | 3525 | 11519 |
| Telangana | 3011 | 5210 | 1685 | 5001 |
| Tripura | 2174 | 3587 | 952 | 3512 |
| Uttarakhand | 1864 | 3458 | 653 | 3382 |
| Uttar Pradesh | 11472 | 21852 | 2835 | 21001 |
| West Bengal | 7826 | 12541 | 2252 | 12305 |
| Andaman & N. Island | 555 | 848 | 289 | 801 |
| Chandigarh | 394 | 685 | 102 | 620 |
| Dadra & Nagar Haveli & Daman & Diu | 442 | 647 | 149 | 603 |
| Jammu & Kashmir | 4041 | 7456 | 1241 | 7093 |
| Ladakh | 293 | 628 | 177 | 518 |
| Lakshadweep | 188 | 373 | 40 | 426 |
| Puducherry | 536 | 918 | 267 | 1028 |
| all India | 119545 | 213285 | 49403 | 206228 |

# Labour Force Participation Rate Area-Wise

| State/UT | No. of Rural Person in Labour | Rural Person Surveyed | No. of Urban Person In Labour | No. Of Urban Person Surveyed |
|---|---|---|---|---|
| Andhra Pradesh | 3846 | 7676 | 2494 | 5980 |
| Arunachal Pradesh | 2559 | 5202 | 1066 | 2612 |
| Assam | 3470 | 12176 | 1305 | 3305 |
| Bihar | 4474 | 15808 | 1781 | 6057 |
| Chhattisgarh | 3355 | 6260 | 1477 | 3396 |
| Delhi | 66 | 218 | 1179 | 3320 |
| Goa | 273 | 705 | 396 | 937 |
| Gujarat | 4087 | 8174 | 3140 | 7810 |
| Haryana | 2219 | 6340 | 1901 | 5056 |
| Himachal Pradesh | 3446 | 5860 | 518 | 1137 |
| Jharkhand | 2664 | 8001 | 1339 | 4377 |
| Karnataka | 3396 | 7464 | 3364 | 8107 |
| Kerala | 2974 | 7219 | 3232 | 8618 |
| Madhya Pradesh | 5196 | 12704 | 2918 | 8221 |
| Maharashtra | 7343 | 14804 | 6706 | 16276 |
| Manipur | 1827 | 4950 | 1780 | 4671 |
| Meghalaya | 1948 | 4408 | 718 | 1773 |
| Mizoram | 1025 | 2379 | 1836 | 4602 |
| Nagaland | 1208 | 2609 | 716 | 1733 |
| Odisha | 4613 | 11709 | 1428 | 3819 |
| Punjab | 2697 | 6546 | 2641 | 6363 |
| Rajasthan | 5378 | 12448 | 2787 | 7554 |
| Sikkim | 1054 | 1632 | 452 | 944 |
| Tamil Nadu | 5317 | 11361 | 4820 | 11698 |
| Telangana | 2374 | 4767 | 2129 | 5444 |
| Tripura | 2122 | 4790 | 965 | 2309 |
| Uttarakhand | 1458 | 3877 | 996 | 2963 |
| Uttar Pradesh | 9461 | 28242 | 4705 | 14611 |
| West Bengal | 5551 | 13740 | 4542 | 11106 |
| Andaman & N. Island | 428 | 838 | 417 | 811 |
| Chandigarh | 95 | 224 | 404 | 1081 |
| Dadra & Nagar Haveli & Daman & Diu | 317 | 593 | 296 | 657 |
| Jammu & Kashmir | 3200 | 8865 | 2126 | 5684 |
| Ladakh | 281 | 693 | 194 | 453 |
| Lakshadweep | 65 | 237 | 156 | 562 |
| Puducherry | 222 | 452 | 551 | 1494 |
| all India | 95393 | 243971 | 67583 | 175541 |

## Unemployment Rate Area-Wise

| State/UT | Rural Unemployed People | Rural Person Surveyed | Urban Unemployed People | Urban Person Surveyed |
|---|---|---|---|---|
| Andhra Pradesh | 284 | 7676 | 425 | 5980 |
| Arunachal Pradesh | 239 | 5202 | 293 | 2612 |
| Assam | 244 | 12176 | 235 | 3305 |
| Bihar | 664 | 15808 | 527 | 6057 |
| Chhattisgarh | 131 | 6260 | 285 | 3396 |
| Delhi | 22 | 218 | 66 | 3320 |
| Goa | 80 | 705 | 82 | 937 |
| Gujarat | 131 | 8174 | 187 | 7810 |
| Haryana | 406 | 6340 | 334 | 5056 |
| Himachal Pradesh | 486 | 5860 | 189 | 1137 |
| Jharkhand | 176 | 8001 | 328 | 4377 |
| Karnataka | 142 | 7464 | 357 | 8107 |
| Kerala | 614 | 7219 | 862 | 8618 |
| Madhya Pradesh | 229 | 12704 | 485 | 8221 |
| Maharashtra | 340 | 14804 | 814 | 16276 |
| Manipur | 223 | 4950 | 248 | 4671 |
| Meghalaya | 225 | 4408 | 218 | 1773 |
| Mizoram | 29 | 2379 | 161 | 4602 |
| Nagaland | 97 | 2609 | 172 | 1733 |
| Odisha | 609 | 11709 | 290 | 3819 |
| Punjab | 425 | 6546 | 388 | 6363 |
| Rajasthan | 548 | 12448 | 672 | 7554 |
| Sikkim | 44 | 1632 | 23 | 944 |
| Tamil Nadu | 636 | 11361 | 667 | 11698 |
| Telangana | 153 | 4767 | 441 | 5444 |
| Tripura | 53 | 4790 | 69 | 2309 |
| Uttarakhand | 202 | 3877 | 213 | 2963 |
| Uttar Pradesh | 706 | 28242 | 1023 | 14611 |
| West Bengal | 316 | 13740 | 533 | 11106 |
| Andaman & N. Island | 97 | 838 | 144 | 811 |
| Chandigarh | 7 | 224 | 44 | 1081 |
| Dadra & Nagar Haveli & Daman & Diu | 28 | 593 | 9 | 657 |
| Jammu & Kashmir | 585 | 8865 | 637 | 5684 |
| Ladakh | 49 | 693 | 55 | 453 |
| Lakshadweep | 17 | 237 | 79 | 562 |
| Puducherry | 41 | 452 | 97 | 1494 |
| all India | 9278 | 243971 | 11652 | 175541 |

## Unemployment Rate Gender-Wise

| State/UT | No. Of Unemployed Male | Male Surveyed | No. Of Unemployed Female | Female Surveyed |
|---|---|---|---|---|
| Andhra Pradesh | 321 | 6820 | 294 | 6837 |
| Arunachal Pradesh | 228 | 3994 | 195 | 3820 |
| Assam | 121 | 8072 | 489 | 7409 |
| Bihar | 572 | 11433 | 271 | 10432 |
| Chhattisgarh | 180 | 4864 | 105 | 4792 |
| Delhi | 41 | 1873 | 37 | 1665 |
| Goa | 63 | 823 | 121 | 819 |
| Gujarat | 174 | 8283 | 123 | 7701 |
| Haryana | 409 | 6014 | 280 | 5382 |
| Himachal Pradesh | 251 | 3434 | 385 | 3563 |
| Jharkhand | 226 | 6271 | 73 | 6107 |
| Karnataka | 230 | 7934 | 191 | 7637 |
| Kerala | 481 | 7516 | 1240 | 8321 |
| Madhya Pradesh | 282 | 10860 | 332 | 10065 |
| Maharashtra | 605 | 15914 | 379 | 15166 |
| Manipur | 229 | 4764 | 219 | 4857 |
| Meghalaya | 128 | 2987 | 265 | 3194 |
| Mizoram | 59 | 3490 | 105 | 3491 |
| Nagaland | 121 | 2153 | 99 | 2189 |
| Odisha | 475 | 7783 | 318 | 7745 |
| Punjab | 358 | 6637 | 602 | 6272 |
| Rajasthan | 648 | 10293 | 350 | 9709 |
| Sikkim | 27 | 1330 | 45 | 1246 |
| Tamil Nadu | 554 | 11540 | 829 | 11519 |
| Telangana | 271 | 5210 | 195 | 5001 |
| Tripura | 54 | 3587 | 39 | 3512 |
| Uttarakhand | 183 | 3458 | 227 | 3382 |
| Uttar Pradesh | 743 | 21852 | 609 | 21001 |
| West Bengal | 351 | 12541 | 468 | 12305 |
| Andaman & N. Island | 73 | 848 | 203 | 801 |
| Chandigarh | 29 | 685 | 19 | 620 |
| Dadra & Nagar Haveli & Daman & Diu | 10 | 647 | 37 | 603 |
| Jammu & Kashmir | 365 | 7456 | 1121 | 7093 |
| Ladakh | 41 | 628 | 47 | 518 |
| Lakshadweep | 39 | 373 | 88 | 426 |
| Puducherry | 65 | 918 | 89 | 1028 |
| all India | 9007 | 213284 | 10489 | 206228 |

## Working Population Rate Gender-Wise

| State/UT | No. of Working Male | No. of Male Surveyed | No. of Working Female | No. of Female Surveyed |
|---|---|---|---|---|
| Andhra Pradesh | 3928 | 6820 | 2297 | 6837 |
| Arunachal Pradesh | 2089 | 3994 | 1455 | 3820 |
| Assam | 3326 | 8072 | 659 | 7409 |
| Bihar | 5019 | 11433 | 928 | 10432 |
| Chhattisgarh | 2855 | 4864 | 1974 | 4792 |
| Delhi | 1038 | 1873 | 181 | 1665 |
| Goa | 450 | 823 | 159 | 819 |
| Gujarat | 5044 | 8283 | 2141 | 7701 |
| Haryana | 3061 | 6014 | 770 | 5382 |
| Himachal Pradesh | 1964 | 3434 | 1692 | 3563 |
| Jharkhand | 3016 | 6271 | 953 | 6107 |
| Karnataka | 4538 | 7934 | 2138 | 7637 |
| Kerala | 3916 | 7516 | 1756 | 8321 |
| Madhya Pradesh | 6321 | 10860 | 1741 | 10065 |
| Maharashtra | 9246 | 15914 | 4610 | 15166 |
| Manipur | 2372 | 4764 | 1044 | 4857 |
| Meghalaya | 1410 | 2987 | 1118 | 3194 |
| Mizoram | 1735 | 3490 | 1107 | 3491 |
| Nagaland | 1111 | 2153 | 751 | 2189 |
| Odisha | 4179 | 7783 | 1572 | 7745 |
| Punjab | 3883 | 6637 | 1135 | 6272 |
| Rajasthan | 5177 | 10293 | 2719 | 9709 |
| Sikkim | 863 | 1330 | 668 | 1246 |
| Tamil Nadu | 6405 | 11540 | 3260 | 11519 |
| Telangana | 2855 | 5210 | 1620 | 5001 |
| Tripura | 2138 | 3587 | 941 | 3512 |
| Uttarakhand | 1767 | 3458 | 609 | 3382 |
| Uttar Pradesh | 11079 | 21852 | 2751 | 21001 |
| West Bengal | 7612 | 12541 | 2178 | 12305 |
| Andaman & N. Island | 507 | 848 | 215 | 801 |
| Chandigarh | 377 | 685 | 99 | 620 |
| Dadra & Nagar Haveli & Daman & Diu | 435 | 647 | 139 | 603 |
| Jammu & Kashmir | 3847 | 7456 | 1043 | 7093 |
| Ladakh | 274 | 628 | 161 | 518 |
| Lakshadweep | 169 | 373 | 32 | 426 |
| Puducherry | 498 | 918 | 244 | 1028 |
| all India | 113041 | 213285 | 42895 | 206228 |

## Working Population Rate Area-Wise

| State/UT | No. of Rural Working Person | No. of Rural Person Surveyed | No. of Urban Working Person | No. of Urban Person Surveyed |
|---|---|---|---|---|
| Andhra Pradesh | 3700 | 7676 | 2320 | 5980 |
| Arunachal Pradesh | 2440 | 5202 | 948 | 2612 |
| Assam | 3397 | 12176 | 1213 | 3305 |
| Bihar | 4284 | 15808 | 1629 | 6057 |
| Chhattisgarh | 3286 | 6260 | 1352 | 3396 |
| Delhi | 60 | 218 | 1155 | 3320 |
| Goa | 242 | 705 | 362 | 937 |
| Gujarat | 4022 | 8174 | 3062 | 7810 |
| Haryana | 2080 | 6340 | 1775 | 5056 |
| Himachal Pradesh | 3159 | 5860 | 433 | 1137 |
| Jharkhand | 2608 | 8001 | 1239 | 4377 |
| Karnataka | 3336 | 7464 | 3218 | 8107 |
| Kerala | 2722 | 7219 | 2904 | 8618 |
| Madhya Pradesh | 5107 | 12704 | 2746 | 8221 |
| Maharashtra | 7180 | 14804 | 6364 | 16276 |
| Manipur | 1747 | 4950 | 1686 | 4671 |
| Meghalaya | 1851 | 4408 | 629 | 1773 |
| Mizoram | 1013 | 2379 | 1772 | 4602 |
| Nagaland | 1164 | 2609 | 645 | 1733 |
| Odisha | 4367 | 11709 | 1318 | 3819 |
| Punjab | 2520 | 6546 | 2482 | 6363 |
| Rajasthan | 5141 | 12448 | 2546 | 7554 |
| Sikkim | 1027 | 1632 | 441 | 944 |
| Tamil Nadu | 5022 | 11361 | 4539 | 11698 |
| Telangana | 2298 | 4767 | 1954 | 5444 |
| Tripura | 2098 | 4790 | 935 | 2309 |
| Uttarakhand | 1380 | 3877 | 924 | 2963 |
| Uttar Pradesh | 9235 | 28242 | 4383 | 14611 |
| West Bengal | 5427 | 13740 | 4331 | 11106 |
| Andaman & N. Island | 378 | 838 | 343 | 811 |
| Chandigarh | 92 | 224 | 388 | 1081 |
| Dadra & Nagar Haveli & Daman & Diu | 302 | 593 | 292 | 657 |
| Jammu & Kashmir | 2988 | 8865 | 1887 | 5684 |
| Ladakh | 261 | 693 | 171 | 453 |
| Lakshadweep | 61 | 237 | 134 | 562 |
| Puducherry | 202 | 452 | 515 | 1494 |
| all India | 96197 | 243971 | 63546 | 175541 |

## Regression Data

| State/UT | LFPR Rate | Literacy Rate | Unemployment Rate | Working Rate |
|---|---|---|---|---|
| Andhra Pradesh | 47.6 | 72 | 4.6 | 45.4 |
| Arunachal Pradesh | 47.9 | 84.1 | 5.4 | 45.3 |
| Assam | 28.9 | 77.5 | 2.2 | 28.2 |
| Bihar | 28.4 | 74.6 | 4.6 | 27.1 |
| Chhattisgarh | 51.7 | 78.2 | 3.1 | 50.1 |
| Delhi | 35.4 | 88.4 | 2.2 | 34.6 |
| Goa | 40.8 | 95 | 9.7 | 36.8 |
| Gujarat | 45.7 | 85.3 | 1.9 | 44.9 |
| Haryana | 35.8 | 83.2 | 6.4 | 33.5 |
| Himachal Pradesh | 57.5 | 88.1 | 9 | 52.3 |
| Jharkhand | 32.9 | 79.3 | 3 | 31.9 |
| Karnataka | 44.1 | 80.8 | 2.7 | 42.9 |
| Kerala | 39.5 | 95 | 9.2 | 35.8 |
| Madhya Pradesh | 39.6 | 76.4 | 2.7 | 38.5 |
| Maharashtra | 46.1 | 86.9 | 3.3 | 44.5 |
| Manipur | 37.2 | 90.7 | 4.7 | 35.5 |
| Meghalaya | 43.7 | 95 | 6.1 | 41 |
| Mizoram | 41.6 | 99.4 | 2.2 | 40.7 |
| Nagaland | 45 | 95.5 | 5.2 | 42.7 |
| Odisha | 39.1 | 79.3 | 5.5 | 36.9 |
| Punjab | 41.3 | 83.3 | 6.3 | 38.7 |
| Rajasthan | 41.7 | 73.4 | 5.4 | 39.4 |
| Sikkim | 61.1 | 87.3 | 2.7 | 59.5 |
| Tamil Nadu | 44.4 | 86 | 5.7 | 41.9 |
| Telangana | 45.8 | 79.2 | 4.7 | 43.6 |
| Tripura | 43.9 | 93.2 | 1.4 | 43.2 |
| Uttarakhand | 36.6 | 86 | 5.7 | 34.5 |
| Uttar Pradesh | 33.3 | 76.6 | 3.3 | 32.2 |
| West Bengal | 40.6 | 82.9 | 3 | 39.4 |
| Andaman & N. Island | 51.2 | 92.8 | 14.3 | 43.9 |
| Chandigarh | 37.6 | 92.7 | 4 | 36.1 |
| Dadra & Nagar Haveli & Daman & Diu | 48.1 | 87.4 | 2.7 | 46.8 |
| Jammu & Kashmir | 36.3 | 78.3 | 7.4 | 33.6 |
| Ladakh | 40.7 | 78.9 | 7.5 | 37.6 |
| Lakshadweep | 27.7 | 98.1 | 12.4 | 24.3 |
| Puducherry | 41.4 | 91.1 | 7.6 | 38.2 |