DETERMINING THE FACTORS THAT AFFECT THE INFANT MORTALITY AND FERTILITY RATE IN INDIAN STATES USING ANALYTICS

MEENU JOMI

A thesis submitted in fulfilment of the
requirements for the award of the degree of
MASTER OF SCIENCE IN DATA SCIENCE

LIVERPOOL JOHN MOORES UNIVERSITY (LJMU)

JULY 2020

TABLE OF CONTENTS

DEDICATION 1

ACKNOWLEDGEMENTS 1

ABSTRACT 1

LIST OF TABLES 2

LIST OF FIGURES 3

LIST OF EQUATIONS 5

CHAPTER 1: INTRODUCTION 6

1.1 Background of the Study 6

1.2 Problem Statement 6

1.3 Aim and Objectives 7

1.4 Research Questions 8

1.5 Scope of the Study 8

1.6 Significance of the Study 8

1.7 Structure of the Study 8

CHAPTER 2: LITERATURE REVIEW 10

2.1 Introduction 10

2.2 Data Collection 11

2.3 Data Cleaning 12

2.4 Data Analysis and Predictive Modeling 13

2.5 Existing Outcomes 14

2.6 Related Research Publications 15

2.7 Summary 16

CHAPTER 3: RESEARCH METHODOLOGY 17
3.1 Introduction 17
3.2 Data Collection 17
3.3 Data Mining 18
3.4 Data Analysis 19
3.5 Outcomes 23
3.6 Summary 23

CHAPTER 4: ANALYSIS 24
4.1 Introduction 24
4.2 Data Preparation 24
4.2.1 Data Collection 24
4.2.2 Data Understanding and Data Cleaning 25
4.3 Exploratory Data Analysis 25
4.3.1 Analysis on Infant Mortality and Fertility in India (2011) 25
4.3.1.1 Univariate Analysis 25
4.3.1.2 Segmented Analysis 43
4.3.1.3 Bivariate Analysis 55
4.3.2 Comparison of Infant Mortality and Fertility of 2001 with 2011 58
4.3.3 Analysis on Infant Mortality and Fertility among States (2011) 70
4.3.4 Analysis on Infant Mortality and Fertility in each States (2011) 97
4.4 Machine Learning 113
4.4.1 Linear Regression 114
4.4.1.1 Linear Regression with mathematical implications 114
4.4.1.2 Linear Regression in Python 115
4.4.2 Decision tree in Python 117
4.5 Summary 118

CHAPTER 5: RESULTS AND DISCUSSIONS 119
5.1 Introduction 119
5.2 Interpretation of visualizations 119
5.3 Summary 120

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS 121
6.1 Introduction 121
6.2 Discussion and Conclusion 121
6.3 Contribution to knowledge 122
6.4 Future Recommendations 122

REFERENCES 123
APPENDIX A: RESEARCH PROPOSAL 126
APPENDIX B: ETHICS FORMS 129

DEDICATION

This Research paper is dedicated to my family and friends. They have always been supportive and encouraging to get this research done. This paper is also dedicated to all my teachers and professors in the past. I am here because of their hard work and they helped to mold me into the creative and positive person I am today.

ACKNOWLEDGEMENTS

I wish to thank my Thesis supervisor, J V Ramana Reddy for always supporting me through the difficult times and helped me to realize my potential to do the research. I also wish to thank my professors who have enlightened me to understand the basics of research writing. My student mentor has also played a major role in this. Along with that, I wish to thank Liverpool John Moores University for giving me the opportunity and tools to do my research paper.

ABSTRACT

High mortality rates in India have been an issue for a very long time. This has been due to many reasons such as diseases, accidents, self-harms, etc. But one among the top 10 reasons is death during birth or at infancy. This is caused due to many factors. The factors can range from unhealthy conditions in hospitals to financial issues. There are births taken place in houses due to financial shortcomings and the unavailability of proper medical assistance. These are the factors that we know currently, but there could be more factors that cause a similar issue. The data collected from the Indian government census in 2001 and 2011 is being used to find solutions and factors for the high mortality rate in infants along with its connection with fertility rate. Then EDA, Linear regression, and Decision Tree are done to the finalized data. This is further represented in graphs and tables to understand the effect of the factors caused on these major issues.

LIST OF TABLES

Table 3.1: The sum of total number of children ever born against age groups 18
Table 4.1: Highest and Lowest in Andaman and Nicobar islands 98
Table 4.2: Highest and Lowest in Puducherry 98
Table 4.3: Highest and Lowest in Tamil Nadu 99
Table 4.4: Highest and Lowest in Kerala 99
Table 4.5: Highest and Lowest in Lakshadweep 100
Table 4.6: Highest and Lowest in Goa 101
Table 4.7: Highest and Lowest in Karnataka 101
Table 4.8: Highest and Lowest in Andhra Pradesh 102
Table 4.9: Highest and Lowest in Maharashtra 102
Table 4.10: Highest and Lowest in Dadra and Nager Haveli 103
Table 4.11: Highest and Lowest in Daman and Diu 103
Table 4.12: Highest and Lowest in Gujarat 104
Table 4.13: Highest and Lowest in Madhya Pradesh 104
Table 4.14: Highest and Lowest in Chattisgarh 104
Table 4.15: Highest and Lowest in Odisha 105
Table 4.16: Highest and Lowest in Jharkhand 105
Table 4.17: Highest and Lowest in West Bengal 105
Table 4.18: Highest and Lowest in Assam 105
Table 4.19: Highest and Lowest in Meghalaya 106
Table 4.20: Highest and Lowest in Tripura 106
Table 4.21: Highest and Lowest in Mizoram 107
Table 4.22: Highest and Lowest in Manipur 107
Table 4.23: Highest and Lowest in Nagaland 108
Table 4.24: Highest and Lowest in Arunachal Pradesh 108
Table 4.25: Highest and Lowest in Sikkim 109
Table 4.26: Highest and Lowest in Bihar 109
Table 4.27: Highest and Lowest in Uttar Pradesh 110
Table 4.28: Highest and Lowest in Rajasthan 110
Table 4.29: Highest and Lowest in NCT of Delhi 111

Table 4.30: Highest and Lowest in Harayana 111
Table 4.31: Highest and Lowest in Uttarakhand 111
Table 4.32: Highest and Lowest in Chandigarh 112
Table 4.33: Highest and Lowest in Punjab 112
Table 4.34: Highest and Lowest in Himachal Pradesh 113
Table 4.35: Highest and Lowest in Jammu and Kashmir 113

LIST OF FIGURES

Figure 3.1: Religion vs Parity 19
Figure 3.2: Age vs Parity 20
Figure 3.3: Total Children Ever Born vs Age 21
Figure 3.4: Total Children Ever Born vs Religion 22
Figure 3.5: Religion vs Marital Status 22
Figure 3.6: Education vs Marital Status 23
Figure 4.1: Rural/Urban vs Parity 0 26
Figure 4.2: Present Age vs Parity 0 27
Figure 4.3: Present Age vs Parity 3 28
Figure 4.4: Present Age vs Parity 4 29
Figure 4.5: Present Age vs Parity 5 30
Figure 4.6: Present Age vs Parity 7+ 31
Figure 4.7: Present Age vs Total Children Ever Born 32
Figure 4.8: Present Age vs Total Ever Married Women 33
Figure 4.9: All/SC/ST vs Parity 0 34
Figure 4.10: Religious Community vs Parity 0 35
Figure 4.11: Educational Level vs Parity 0 36
Figure 4.12: Educational Level vs Parity 3 37
Figure 4.13: Educational Level vs Number of surviving children 3 37
Figure 4.14: Educational Level vs Parity 4 38
Figure 4.15: Educational Level vs Total Children ever born 39
Figure 4.16: Economic Activity vs Parity 1 40
Figure 4.17: Present Age vs Birth order last year - 1 41
Figure 4.18: Present Age vs Birth order last year - 3 41

Figure 4.19: Present Age vs Birth order last year – 5 42
Figure 4.20: Present Age vs Currently Married Women 43
Figure 4.21: All/SC/ST vs ever Married Women (Rural/Urban) 44
Figure 4.22: Rural/Urban vs Ever Married Women(Present Age) 45
Figure 4.23: Rural/Urban vs Ever Married Women(Rural/Urban) 46
Figure 4.24: Rural/Urban vs Parity 0(Rural /Urban) 47
Figure 4.25: Religious Community vs Parity 0(Rural/Urban) 48
Figure 4.26: Religious Community vs Parity 1(Religious Community) 49
Figure 4.27: Present Age vs Total Ever Married Women (Educational Level) 50
Figure 4.28: Present Age vs Parity 4 (Educational Level) 50
Figure 4.29: Present Age vs Women with 2 Surviving Children (Educational Level) 51
Figure 4.30: Present Age vs Women with 3 Surviving Children (Educational Level) 51
Figure 4.31: Rural/Urban vs Total Ever Married Women (Educational Level) 52
Figure 4.32: Rural/Urban vs Women with 2 Surviving Children (Educational Level) 53
Figure 4.33: Rural/Urban vs Women with 5+ Surviving Children (Educational Level) 54
Figure 4.34: Economic Activity vs Total Ever Married Women (Rural/Urban) 55
Figure 4.35: Heatmap for Last year's birth order, births last year and current marital status .. 56
Figure 4.36: Heatmap for Parity, births till now and ever married women 57
Figure 4.37: Present Age vs Parity 3(all/SC/ST) in 2001 and 2011 59
Figure 4.38: Present Age vs Parity 6(all/SC/ST) in 2001 and 2011 60
Figure 4.39: Religious Community vs Parity 1(rural/urban) in 2001 and 2011 62
Figure 4.40: Present Age vs Number of ever married women (religious community) in 2001 and 2011 63
Figure 4.41: Educational level vs total women in 2001 and 2011 64
Figure 4.42: Educational level vs total ever married women in 2001 and 2011 65
Figure 4.43: Present age vs total ever married women in 2001 and 2011 66
Figure 4.44: Present age vs parity 5 in 2001 and 2011 67
Figure 4.45: Total/rural vs total children ever born in 2001 and 2011 68
Figure 4.46: Educational level vs number of surviving children 0 in 2001 and 2011 69
Figure 4.47: Educational level vs number ever unmarried women in 2001 and 2011 70
Figure 4.48: States vs Parity 0 71
Figure 4.49: States vs Parity 1 72
Figure 4.50: States vs Parity 2 73
Figure 4.51: States vs Parity 3 74

Figure 4.52: States vs Parity 4 75
Figure 4.53: States vs Parity 5 76
Figure 4.54: States vs Parity 6 77
Figure 4.55: States vs Parity 7+ 78
Figure 4.56: States vs Total Women 79
Figure 4.57: States vs Total ever married Women 80
Figure 4.58: States vs Total unmarried Women 81
Figure 4.59: States vs Birth order last year 1 82
Figure 4.60: States vs Birth order last year 2 83
Figure 4.61: States vs Birth order last year 3 84
Figure 4.62: States vs Birth order last year 4 85
Figure 4.63: States vs Birth order last year 5 86
Figure 4.64: States vs Birth order last year 6 87
Figure 4.65: States vs Birth order last year 7+ 88
Figure 4.66: States vs currently married women 89
Figure 4.67: States vs currently unmarried women 90
Figure 4.68: States vs number of female births 91
Figure 4.69: States vs number of surviving children 0 92
Figure 4.70: States vs number of surviving children 1 93
Figure 4.71: States vs number of surviving children 2 94
Figure 4.72: States vs number of surviving children 3 95
Figure 4.73: States vs number of surviving children 4 96
Figure 4.74: States vs number of surviving children 5+ 97
Figure 4.75: Linear model – 1 116
Figure 4.76: Tuning maximum depth 117
Figure 4.77: Decision tree 118

LIST OF EQUATIONS

Equation (1) : Linear Regression line's equation 114
Equation (2): R^2 value 115

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

'State of new-born health in India' (2016) - it focused on factors like rural-urban, poor–rich, facility availability, and gender differentials. The outcome was that there was a requirement for improvement in the health facilities and the introduction of new interventions [5].

'High Neonatal Mortality Rates in Rural India: What Options to Explore?' (2012)- The focus was on the rural areas and the main factor that was looked into was the effects on the neonatal mortality rate depending on birth locations. The conclusion states that there needed to be large scale projects that will help promote home-based new-born care, creating community awareness and community mobilization along with strengthening public-private partnerships [4].

'The association between neonatal death and facility birth in regions of India' (2019) – The objective of this research article was to find out if there is a relation between neonatal deaths and facility births in different states. It turns out that there was a relation in a few states like Uttar Pradesh and Bihar, but was not as robust in other states [3].

'Impact of timing of breastfeeding initiation on neonatal mortality in India' (2018) – The main aim of the study is to examine the timing of initiation of breastfeeding and neonatal deaths. It was found that timely initiation of breastfeeding is beneficial for child survival within the first 28 days of birth in all of the cases of mortality that was looked up [2].

'Mapping Neonatal Mortality in India: A Closer Look' (2017) – This article shows how the mortality rate differs from state to state. As a conclusion, they succeeded to figure out this issue but stated that there is an urgent need for up-to-date data on district-level neonatal mortality in India [1].

In the past, all researches have been done regarding Neonatal deaths and related causes. But in most of them, they have looked at broad factors like residence area, gender, financial background, place of birth, facility, breastfeeding, and so on.

1.2 Problem Statement

Major problems in India include corruption, illiteracy, lack of proper education systems, high rate of mortality and fertility, poverty, and much more. In this research, we focus on one of them which is high mortality rate and fertility rate. Most of the above mentioned issues are

interconnected. The mortality rate has been reducing in the past years but the rate has always been higher than the predicted rate. This issue has been acknowledged by the government in the past but the actions taken have not produced enough results. This shows that more issues need to be acknowledged.

Illiteracy is a major problem in India. This has affected the mortality and fertility rate. There are cases where Female infanticide and the high fertility rate has caused due to the citizens not knowing contraceptives or due to religious acts.
Another major issue is sanitation. Some of the births in India happen at houses where the living conditions are not fit for the same. Sanitation can also reduce the mothers' health and thus affect the children.

In India, the legal timeline for abortions is up to 20 weeks of pregnancy, although exceptional cases can go up to 24 weeks with the court agreement. But there are many instances of illegal abortions happening after this period as well. It is also illegal to reveal the gender of the fetus till birth but despite these rules, some doctors revile the gender. In some cases, this leads to abortion as some families do not prefer female children.
Another major problem is the unevenness of the rates among different states. There are some states in India with a mortality rate as high as 12% in Kerala as compared to 59% in Madhya Pradesh in 2011. This shows that the above stated factors also vary state wise. So this research also looks deeply into those issues as well.

1.3 Aim and Objectives

This research aims to figure out the factors affecting mortality and fertility. Later on, it will be ranked to know which of them are most prominent. This will be done at the national level and state level as well.

The objective will be to find the correlation of mortality with many factors like mothers' age, marital status, religious community, educational level, occupation along with the number of deaths caused; gender, surviving rate, and so on of the child in different years. The relation between fertility rate and infant mortality rate at the state level and national level will also be found. Later on, the results will be compared in the years 2001 and 2011.

1.4 Research Questions

- What are the major factors that affect the mortality rate in infants?

This question is important because infant mortality has been one of the top 10 reasons for death in India. This states that it is one of the major issues. Finding the factors and by how much they affect this issue we can try to solve this problem by coming up with new ideas that can be implemented.

- By how much does the mortality rate in infants differ from state to state?

This is very important to investigate since there has been a huge difference in mortality rate in infants from state to state even when the fertility rate remains almost the same. This shows that just figuring out the factors at the national level is not enough but rather needs to be investigated into the state level as well.

1.5 Scope of the Study

The study is done from 2001 to 2011 to analyze the reasons for infant mortality and high fertility rates. This restricts the chance to explore the situation in the present day due to the unavailability of data. However, the research will help us to realize the issues and factors that have caused infant mortality and high fertility. Although this research can give a base and help in comparison for the coming years when the newly collected data will be analyzed.

1.6 Significance of the Study

Mortality due to causes that can be stopped is a huge disadvantage for a country. Infant mortality especially is one of the issues that usually have solutions. Finding out the factors that cause a high infant mortality rate will help not just to reduce it but also to improve conditions in the country. Most of the time infant mortality rate and the high fertility rate are associated with environmental factors. So this research will provide an insight into the major problems and solving them will improve the conditions of the country while saving lives.

1.7 Structure of the Study

The structure of the thesis is as follows. Chapter 1 presents the background of the research in Infant mortality rate. It also states the problem and the reason why this research is important and significant. It also states the aim and objective of the research. It finally discusses the scope and research questions.

Chapter 2 includes the literature review. This includes the studies and explorations done among the previous research papers and conferences. It has the following steps: data collection, data cleaning, data analysis, and predictive models. This is all in regards to the previous studies. It is written in a way that under each subtitle each consecutive paragraphs are related and from a similar research paper. This is done to create and ease in figuring it out while also following a method. Finally, the existing outcomes are stated.

Chapter 3 includes the research done. It starts with an introduction. The steps followed are Data collection, Data mining, Data analysis, outcome, and a summary.

Chapter 4 includes analysis and machine learning. 4.1 has the introduction, 4.2 has data preparation, 4.3 has exploratory data analysis, 4.4 has machine learning and 4.6 has the summary of the chapter.

Chapter 5 includes results and discussions. 5.1 has an introduction, 5.2 interpreting the visuals, and 5.3 the summary.

Chapter 6 includes conclusions and recommendations. 6.1 has the introduction, 6.2 discusses the issues, and concludes the research, 6.3 has the contributions to knowledge, and 6.4 has the future recommendations and how this research can be taken forward.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In the past, there has been a lot of research on infant mortality rates all over the world. Unfortunately very few researches have taken place specifically for India. Some of these research papers have been about infant mortality rates in a whole country while others were specifically done considering the data found from the hospitals.

When it comes to the country or state level, there have always been many factors that affect the mortality rate. The factors include the educational level of parents, maternal health, place of residence, healthcare availability, age of the mother, financial status of parents and family, health of the newborn, availability of clean drinking water and food, toilet facilities, condition of the houses, and much more.

Diving deeper, it shows us that more factors are determining the infant mortality rate which is at an individual level. This includes the child's health at birth, the place of birth (whether it was at home or a hospital facility), distance between the health center and the location of the house, availability of ease in transportation, premature births, availability of baby incubators, number of check-ups done by the mother before and after birth, pregnancy complications, etc.

The fertility rate has been affected by factors like availability of contraceptives, religious beliefs, educational level of females, knowledge and widespread sex educations, and so on. It is very necessary to reduce the fertility rate in India due to the increasing population and less availability of healthy and comfortable living for the existing population.

Analyzing data and working with machine learning tools have brought out the factors affecting high rates of infant mortality and fertility. This along with data visualization tools has helped to understand the important variables and how much they affect.

2.2 Data Collection

Parental education, income per capita, and health service availability have proven to be the three most important factors affecting infant mortality. Using higher degree polynomial ridge regression on infant mortality 2013 data provided by the UN inter-agency group has shown that female education especially has an impact on reduced infant mortality rates. It also estimated importance in mean years of male schooling, female schooling, and per capita income on reducing infant mortality. The data was collected from Barro and Lee (2013) and data from World Bank (2013) which were publically available [6].

In a study about infant mortality in Russia, they focused on infant mortality and more than 100 factors affecting it. It also focused on data quality and selection of modeling methodology. The data sources were from a variety of national government statistics. This has proven to be a major challenge in the data cleaning process [8].

A research paper on how to reduce infant mortality in Brazil collected data from SIM and SINASC databases available on the DATASUS portal. They collected data on infant deaths and live births in the state of Ceara to study the risk of fetal deaths. In Brazil, an intelligent governance framework for Brazilian healthcare emerged called Stork Network. This helped in increasing the health care needs for pregnant women and newborns. This data was also used for prediction as well [9].

Another study was done in Bangladesh. It is a developing country and the infant mortality rate is a very important factor that affects it. It has an estimate of 160.996 million and ranks 8th in the world. Infant mortality is an important factor since it is usually due to the unavailability of proper health facilities. The data was collected from the 'World Development Indicator, 2015 (WDI-2015)' database for Bangladesh [10].

Maternal mortality and infant mortality rate often decide the development of a region. Research is done on infant mortality rate and maternal mortality rate in East Java used secondary data from the health service profile and Badan Pusat Statistika. The number of samples used was 38. They used the Mixed Geographically Weighted Bivariate Weibull Regression (MGWBWR) technique to solve the issues and find out the factors the caused the same.

A paper on Indonesia's infant mortality rate collected data from the Indonesian health profile, Ministry of Indonesia. The data included infant mortality rates from 1995- 2008 [13].

2.3 Data Cleaning

In a study on the relationship between infant mortality rate and per capita income they analyzed 136 countries and figured out that the per capita income in Qatar was too high. So for this reason they removed Qatar since it was considered as an outlier. In the same analysis, the independent variables that affected infant mortality were socioeconomic conditions, parental education, and health services indicators. The socioeconomic condition includes access to clean water supply, availability of proper sanitation. Health service indicators include physicians, nurses, and hospital beds per 1000 population [6].

In an analysis, the data was collected from various sources and so the cleaning process in that was strenuous. The step were mainly two: Identifying and dealing with the data, and exploratory data analysis. While identifying the data they noticed a lot of missing values. They solved it by inputting values that do not mislead the model. Before that, they removed all the variables with a large amount of missing or corrupted values. Next, they made subsets and used efficient methods to input values by considering the common values in the other variables. The selection criteria made sure that none of the relevant values were dropped from model building [8].

The data collected from the Stork Network had a total of 124,876 births in the state of Ceara in Brazil. Sixteen attributes were selected finally from this data which was related to infant mortality. They include the age of mother, marital status of the mother, mother's schooling, birthplace, number of live births previously, number of stillborn previously, gestation week, type of pregnancy, child's gender, infant's weight at birth, child's skin color at birth, complications at birth and so on. The data collected was fairly clean [9].

In the research paper done on Bangladesh's infant mortality rate, they had around 650 indicators taken from 25 datasets. The database included 55 years of data spanning from 1960 to 2015. Some of the variables were ignored since they were less significant and/or had many missing values. The variables included maternal mortality rate, birth rate, newborns protected against tetanus, female population, rural population, adolescent fertility rate, age of the mother at birth, availability of medications and vitamins, household situations, availability of drinking water, and so on [10].

In the research done in East Java, they used factors like percentage of live-behavioral household clean and healthy, Percentage Antenatal Care (ANC) of Pregnant Women and variables affecting maternal mortality rate in Blitar Regency, the ratio of the number of health facilities on 100,000 population, Percentage of women in East Java who married under 17 years, Percentage of poor, Percentage Antenatal Care (ANC) of Pregnant Women [13].

2.4 Data Analysis and Predictive Modeling

The data on mean female schooling years concerning infant mortality rate was trained. Third-order polynomial ridge regression was used. The third order was used instead of the first order since it decreased the error which increases the coefficient of determination. To not cause overfitting the polynomial order used was not above 3. A similar approach was used for mean years of male schooling data as the independent variable. Using the F test, the null hypothesis was rejected with 99.9% confidence. It was finally found that female education is more influential in reducing infant mortality rates than male education. The next regression analysis was done on Gross National income per capita at Purchasing Power Parity as the independent variable against infant mortality rate. Third-order of the polynomial was chosen again. Using the F test the null hypothesis was rejected yet again. But in this case, the coefficient of determination was lower than the other two done above [6].

The data that was cleaned after being collected from a different source was then put through a lot of machine learning models. The techniques include Linear Regression, Decision Tree, Least-angle Regression, Least Absolute Shrinkage and Selection Operator, Gradient Boosting, Neural Network, Partial Least Square, and Ensemble. The general strategy included data partitioning, model building, and model optimization. The data partitioning step divides the dataset into three: training set for building models, validation set for optimizing the final model, and a test set for evaluating the model-independent from the data set used for modeling. The model building includes selecting the important variables for predicting the target. For the interval target, the common criteria used is the average squared error. This error is reduced when the input is higher. The model optimization step is to select the optimal model from a list of models by assessment criteria like single-board computers (SBC), Bayesian information criterion (BIC), Akaike information criterion (AIC), validation error, and cross-validation error [8].

Many algorithms like Decision Tree, Random Forest, Bayes Net, Neural Networks, and SVM was used to analyze the data collected from Stork Network. The data was first cleaned and then finally put through all the models mentioned. After a thorough process of analyzing and comparing different data, it was clear that Naive Bayes was the right way to go with it. One of the reasons for the success of Naive Bayes is the algorithm's capability to work with incomplete and imprecise information. Out of 124,876 cases, Naïve Bayes predicted 122,689 cases correctly and 2187 cases incorrectly. The main aim was not to identify an unhealthy infant as healthy. This helps to stay on the safe side. Later on, it was used in the application so that once the data of the mother and child was entered, it was applied to the mathematical model which finally showed the result as a percentage in terms of the risk [9].

While doing infant mortality rate analysis in Bangladesh, they started with 650 variables. It was brought down again to 550 variables. ANOVA was applied to the dataset and the variables were brought down to 6 variables by setting the testing with p-value <0.1. Finally, 6 factors were used to create a regression model. The final six factors included maternal mortality ratio, newborns protected against Tetanus Injection, Birth Rate, Female Population, Rural Population Growth, and Adolescent Fertility Rate. According to the model, decreasing values in variables such as maternal mortality ratio, birth rate, crude (per 1,000 people), population, female (% of total), rural population growth (annual %), the adolescent fertility rate (births per 1,000 women ages 15-19) caused the infant mortality reduction [10].

Mixed Geographically Weighted Bivariate Weibull Regression Method was used to analyze infant mortality rate and maternal mortality rate. The mortality rate was done according to a regency basis. The factors were run through the algorithms and results were drawn [13].

2.5 Existing Outcomes

When it comes to female education, it was found that education from first grade to 5th grade has proven to reduce the infant mortality rate. Meanwhile in male education seems to be important from 3rd grade to 7th grade. It was also found that female education was more important than male education to reduce infant mortality. It was also shown that income per capita played an important role as well but education seems to be of more importance. It was concluded that an educated parent is more cautious about the child's health and well-being since they are aware of its importance [6].

By analyzing in Russia regarding infant mortality many factors affecting the same were found. The factors included socioeconomic conditions, health issues of the mother, pregnancy complication, abortion, health infrastructure like hospital beds, number of clinical visits per day and medical staff availability, geographic location, and so on. Usually, traditional surveys and researches are done on this topic but using predictive analysis will help us find more factors that were overlooked to be considered as important [8].

The case study regarding Brazil's infant mortality discussed on a hospital level. They used details from the north-eastern regions of Brazil where this was very sensitive to. This helped in figuring out and making a model that predicted the health status of the mother and newborn. It identified relationships between several factors of infant deaths and premature births along with information on the mother's health during pregnancy [9].

The purpose of the study was to understand the relation of infant mortality rate with the factors in Bangladesh. Using multiple regression and cross-validation techniques it was clear that only 6 variables were important among 16. Although in the study data from 1990-2013 were not included since the data then was not available during that period and had a lot of missing values but the final model created proved a good amount of accuracy and figured out the underlying reasons that caused high infant mortality rate [10].

Finally, the factors influencing infant and maternal mortality included Percentage of live-behavioral household clean and healthy, Percentage Coverage of Pregnant Women K4 and variables affecting MMR in Blitar Regency are Ratio of the number of health facilities on 100,000 population, Percentage of women in East Java who married under 17 years, Percentage of poor, Percentage Antenatal Care (ANC) of Pregnant Women [13].

2.6 Related Research Publications

A lot of researches has been done on these topics. But very few have been done in an Indian context. When it comes to fertility rate there has been almost no researches done regarding the same.

2.7 Summary

Most of these researches have been done in the past. This might be due to less availability of data. Some of the factors that affected infant mortality rate included access to clean water supply, availability of proper sanitation, health service indicators includes physicians, nurses and hospital beds per 1000 population, maternal mortality rate, birth rate, newborns protected against tetanus, female population, rural population, adolescent fertility rate, age of the mother at birth, availability of medications and vitamins, household situations and so on. These are the cases in other developing countries. So, the approach and the factors used there can be very useful in this research.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The mortality trend has always been unstable in India, although it shows a reduction with years. In 1990, the expected mortality was 86.4 deaths per 1000 live births and the actual value came out to be 80.7 deaths per 1000 live births among kids below 1 year age group in India. But later on in 2017, the expected mortality deaths were 34.3 deaths per 1000 live births and the actual value came out to be 36 deaths per 1000 live births. So it is clear that there have been more complications for below one year groups that expected in 2017. The mortality among children can be seen to differ from state to state. In Kerala, the infant mortality rate is 12 per 1000 population as compared to the rate is 35 per 1000 population in Bihar.

The fertility rate in India has constantly reduced. In 1990, the fertility rate in India was 4.04 births per woman as compared to 2017 when it has come to 2.27 births per woman. In Kerala, the fertility rate is 2.35 births per woman as compared to the rate of 3.6 births per woman in Bihar.
So, it is clear that in the two states shown above the fertility rate is almost similar but there is a huge difference in infant mortality rate.

According to Institute for Health Metrics and Evaluation, the ranking of most death caused in India by neonatal disorders have reduced from 4th to 7th from 2007 to 2017 but it is still one among the top ten.

3.2 Data Collection

The Indian census has been conducted 15 times to date. The last one collected was in 2011. In this research, the data used is from the Indian census website. Data from both the years 2001 and 2011 are being used in this research. The data collected are available in state wise order and for the whole of India. The latter includes data for the whole of India and also state wise. The datasets for each state contains values for each district.

There are 10 main categories in the data such as Fertility Tables, Migration Tables, Tables on houses, household amenities and assets, special tables for scheduled tribes, Special tables for

scheduled castes, social and cultural tables, general economic tables, general population data, and household tables. In each of these categories, there are tables for each state and the country as a whole.

The fertility table includes variables like state code, district code, area name, the location being at rural, urban and both, the present age of the mother, the total number of women, total number of women ever married, parity number (the number of time women has given birth either alive or stillborn), the total number of children born till now in terms of male, female and both, religious community, educational level, economic activity, Number of surviving children according to number, male, female and total, number of births last year in terms of gender and total, the order of the Childs' birth, total children ever born, total surviving male and female children, the ratio of male children born to the number of male children surviving, scheduled caste, scheduled tribe, and total.

All the tables were in a raw format. Null values did not exist. If there was a null value it meant that it was not applicable.

3.3 Data Mining

Data mining is the process of examining large data and withdrawing required variables. The main challenge that existed was to combine all of the above-said data. Since they were different topics they could not be brought to one file. In most cases when it was combined a total value was taken. For Example, in the fertility tables when understanding the total number of children ever born while grouping in terms of age group pivot tables was used. They were used to get the sum of the total number of children ever born age-wise as shown in Table 3.1.

Table 3.1: The sum of total number of children ever born against age groups

Present Age	Total children ever born (Females)	Total children ever born (Males)	Total children ever born (Persons)
All ages	986064584	1086632028	2072696612
Less than 15	558416	601706	1160122
15-19	6117124	6415514	12532638
20-24	51812162	54630008	106442170

25-29	107910044	114862248	222772292
30-34	132012644	140804770	272817414
35-39	138957956	150135050	289093006
40-44	111942274	123347924	235290198
45-49	101948506	114821270	216769776
50-54	77266898	87754094	165020992
55-59	68734376	78903458	147637834
60-64	67812744	77071272	144884016
65-69	51486098	58632420	110118518
70-74	34145988	38646592	72792580
75-79	15566870	17687014	33253884
80+	18172242	20542390	38714632

The data for each state was in a different folder. Using a concat code in python package called pandas, the data of each state were piled one on top of another. This ended in making one table with all the states for each table in the categories.

3.4 Data Analysis

The analysis includes understanding data and solving problems. Considering the fertility tables there are a lot of factors. These factors are converted into graphs to analyze better.

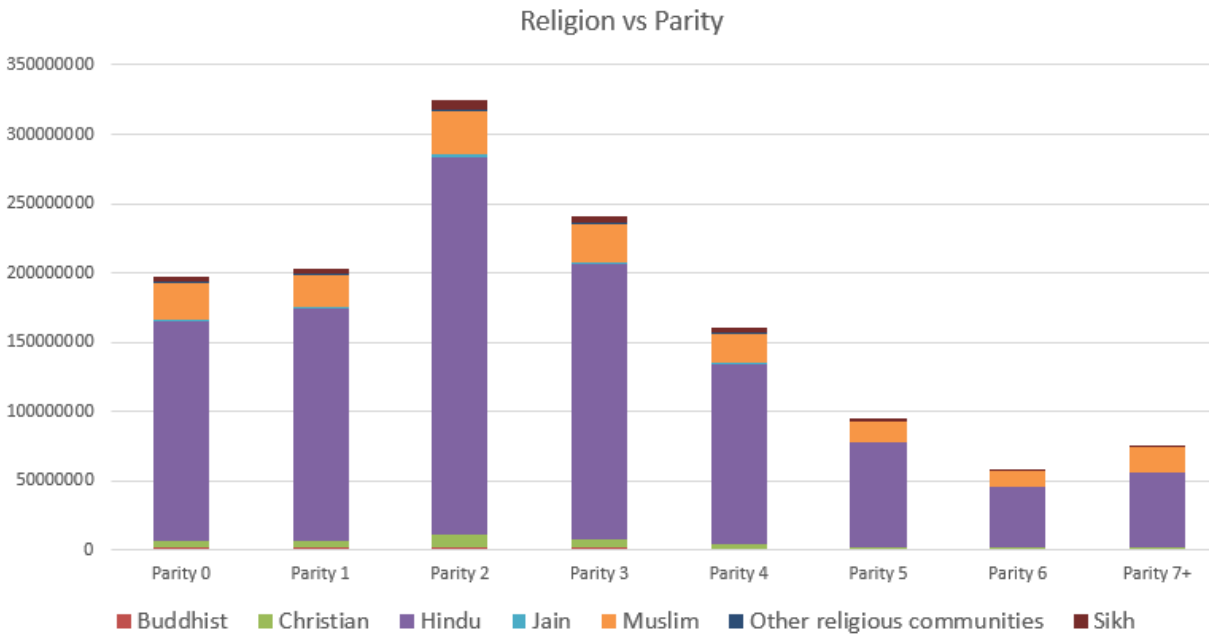

Figure 3.1: Religion vs Parity

It is visible in figure 3.1 that the Hindu population is very high and parity 2 is highest. Although among Muslims and Christians the parity remains constant. The parity 2 is highest in Hindus as compared to others. This graph figure 3.1 also shows a skewed normal curve in terms of parity.

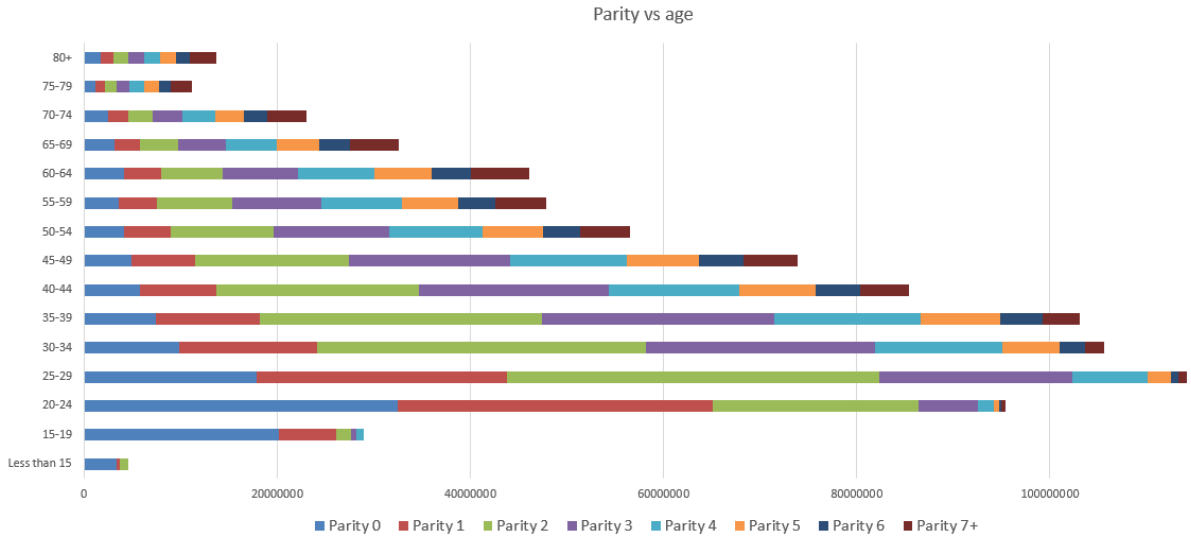

Figure 3.2: Age vs Parity

It is noticed that even women below 19 years have had more than one kid. Also, parity 2 and parity 4 seem to be consistent in all the age groups as seen in figure 3.2.

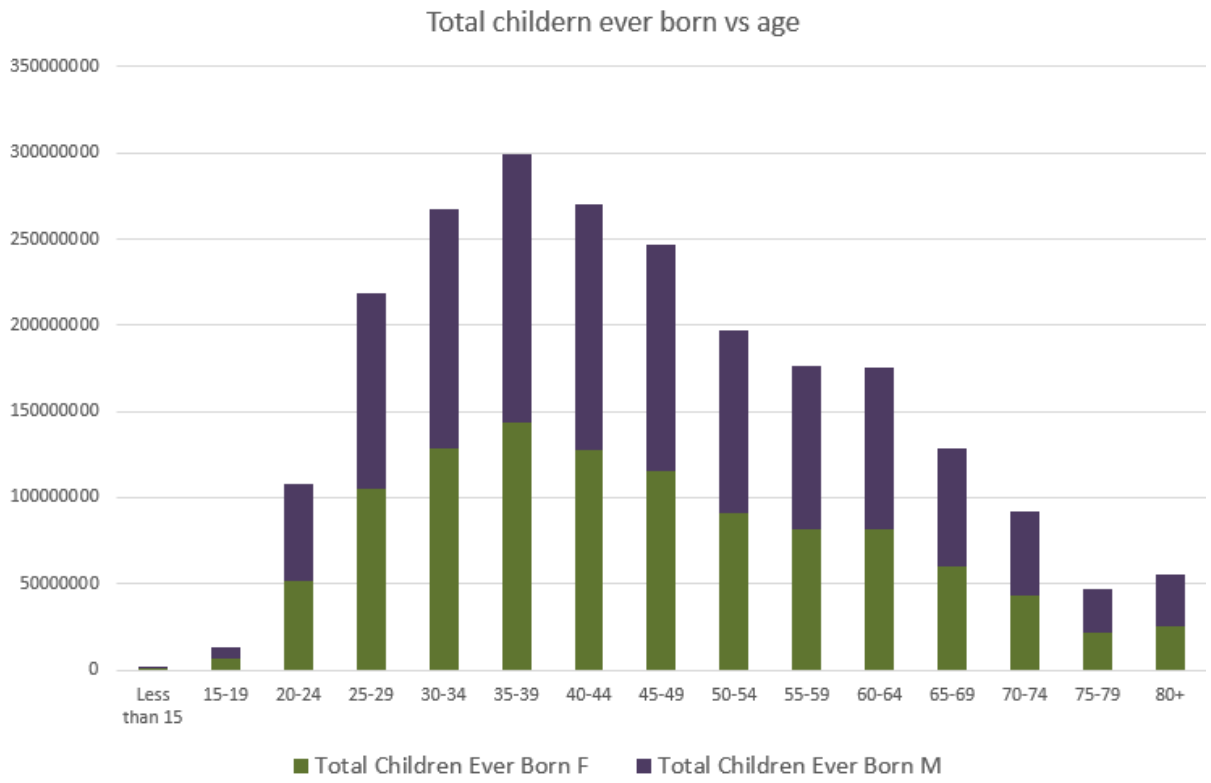

Figure 3.3: Total Children Ever Born vs Age

It is clear from figure 3.3 that there are a lot of kids for women in the age group 35-39. This is a positive sign since the parents at that age will be more settled in life and will be able to take care of the kids' health. It is visible that Male children are more than female in all the age levels.

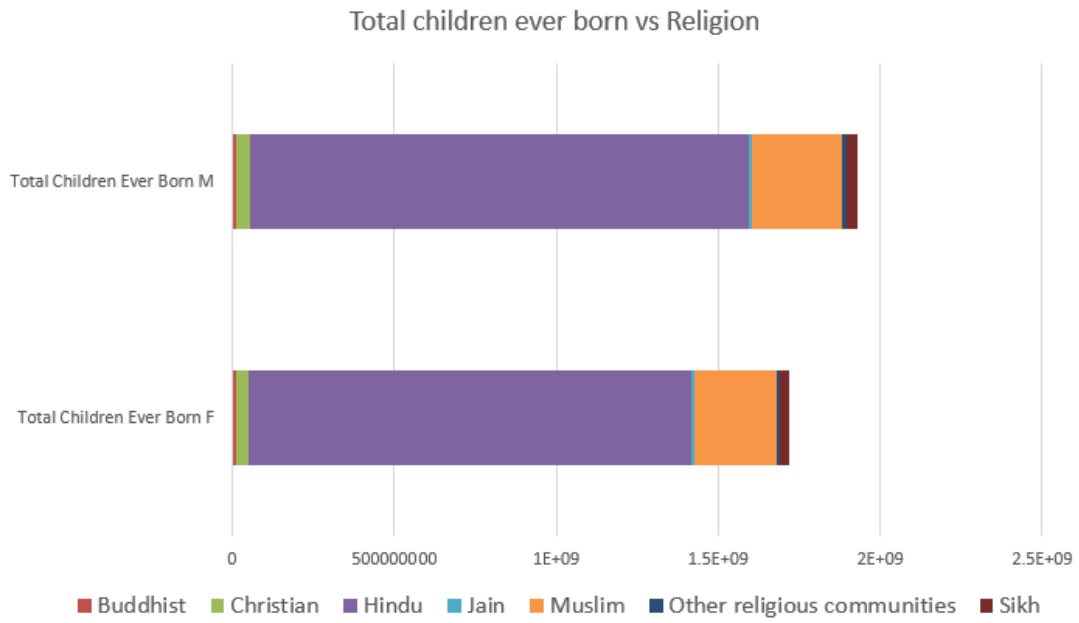

Figure 3.4: Total Children Ever Born vs Religion

It is visible that Hindus have more male children than female. The total number of females born is also less compared to males. It is also visible that Hindus have a huge effect on this difference as seen in figure 3.4

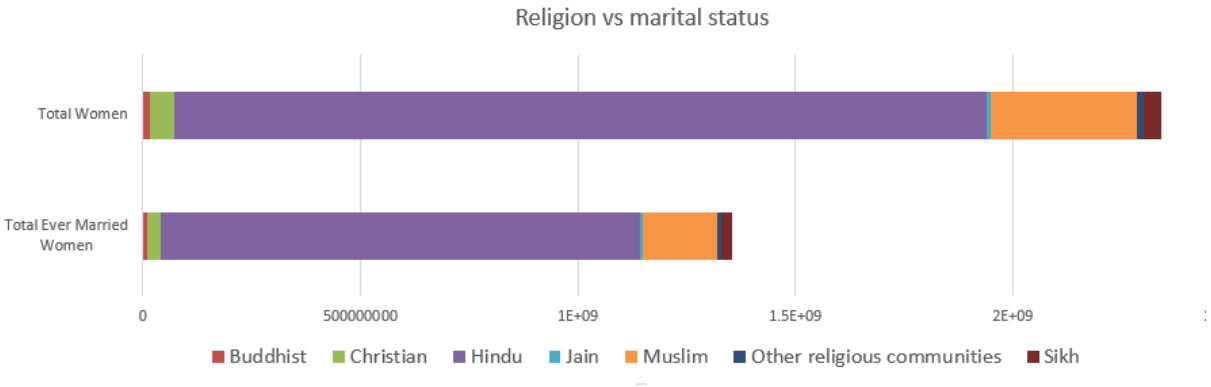

Figure 3.5: Religion vs Marital Status

The number of married women is much higher than unmarried females. This will in turn lead to a higher fertility rate.

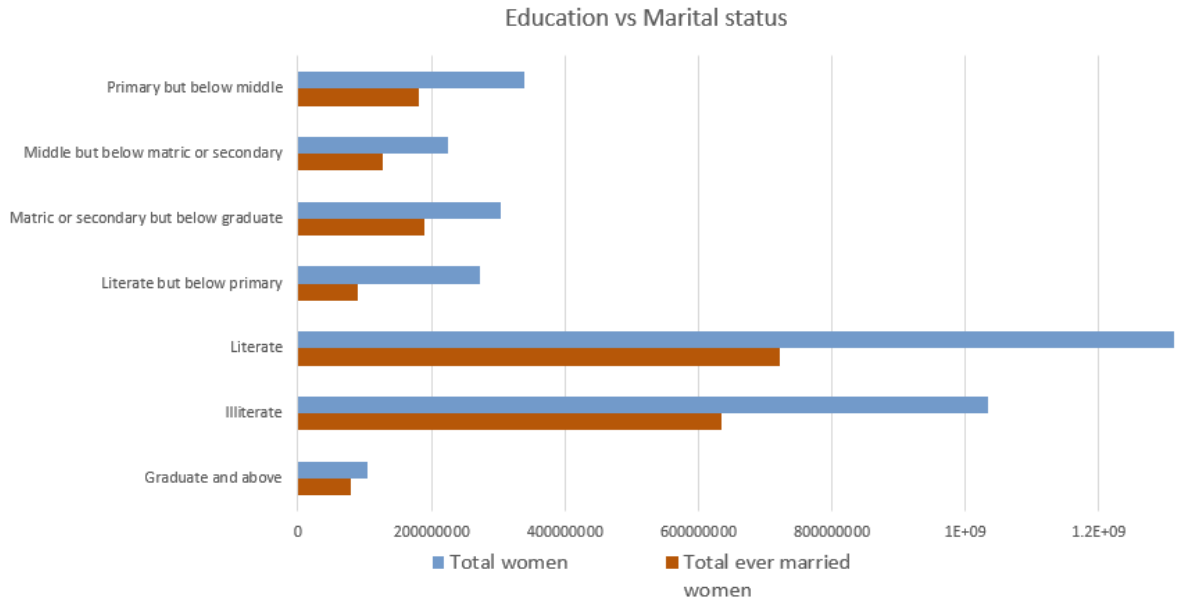

Figure 3.6: Education vs Marital Status

In figure 3.6, half of the total women who are literate are married meanwhile the illiterate ones have more than half married females.

3.5 Outcomes

It is visible that in terms of religion Hindus have the highest population. This shows that they will affect all the factors the most. They do have more male children which are a disadvantage since it will affect the sex-ratio. This along with the current day news shows that it is due to infanticide. Some women are less than 19 years having as many as 4 kids. This is a very unhealthy situation as it will create complications during pregnancy. But the positive side is that people between the ages of 35-39 have most of the kids. This is a good sign because this shows that most of the females have kids at an age when they are matured and financially abled. It has also made clear that having more educated females reduced the number of marriages which in turn reduces fertility rates.

3.6 Summary

The research included data collection from the Indian census. Then it was cleaned and prepared according to the requirements. Later on, an analysis was done which brought out major results as shown above.

CHAPTER 4

ANALYSIS

4.1 Introduction

This analysis is done in two parts. One is at the country level and the other state wise. There existed 12 tables related to fertility and mortality rate. All these were compiled into 4 major tables depending on the factors. An analysis was done separately on these tables. The analysis included univariate analysis, segmented analysis, and bivariate analysis. The same was repeated for both the states and India.

Machine learning tools are used to understand which the most important factors are. Both supervised and unsupervised tools are used for the same. Linear regression is used in supervised learning. A decision tree is used in unsupervised training.

4.2 Data preparation

The data is collected from the Indian census. It is then combined to form 12 major tables. It is then merged into 4 tables. The dataset contains details of India as a whole, each state, and each district. For EDA only India and states data were used.

4.2.1 Data collection

The availability of data for 2001 and 2011 were on two different websites provided by the Indian government. The tables for 2001 were found under the 'Tabulations Plan of Census Year – 2001' and tables for 2011 were found under 'Population Enumeration Data (Final Population)'. Below that there were many series with data for each purpose. This is the same for both 2001 and 2011. The series included general population, general economics, social and cultural, migration, fertility, scheduled caste, and scheduled tribes tables. Out of this, the tables under the fertility table (F-Series) were taken.

The tables ranged from F1 to F14. There were separate tables for scheduled caste and scheduled tribe. Under the series name F1, there are tables for each state and India as a whole. The same existed for each from F1 to F14.

4.2.2 Data Understanding and Data Cleaning

There are many data tables. The columns mainly include survival rates, birth rates, birth order, and marital status. There are different data sets for each attribute. The attributes include religion, present age, rural/urban, SC/ST, educational level, and economic activity. The data of each state and India were in different folders. So using concat they were all brought into one dataset table. The style of each file was different with different column headers. They were all changed to a common header. All the variables and column headers were brought to lower cases to not get repeated.

4.3 Exploratory Data Analysis

EDA includes analysis of data to solve problems and to make a sense of the data. This included three parts: univariate analysis, segmented analysis, and bivariate analysis

4.3.1 Analysis on Infant Mortality and Fertility in India (2011)

This section includes analysis on each topic in India and the state level in the year 2011. Many factors can affect infant mortality. This is a major issue in India. EDA will help analyze the situation and understand the main cause of this problem

4.3.1.1 Univariate Analysis

Univariate analysis means analyzing one variable and understanding the differences among them. This helps in a detailed evaluation of each variable. For example, in the case of the present age, the univariate analysis will help understand which age group most of the population belongs to.

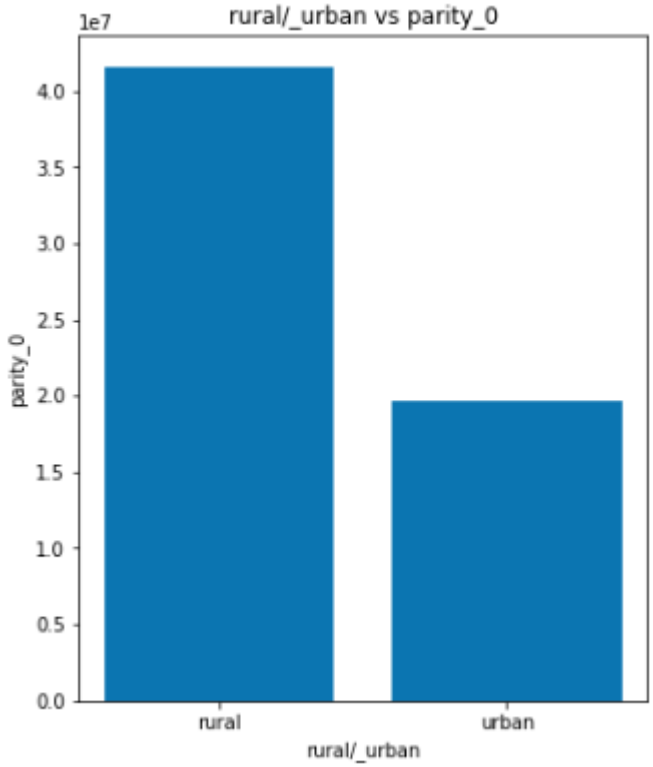

Figure 4.1: Rural/Urban vs Parity 0

In figure 4.1, it is an analysis of the number of women with 0 parity corresponding to rural and urban. It is visible that the rural population is higher than urban. It is the same in case of parity 1, parity 2, parity 3, parity 4, parity 5, parity 6, parity 7+, total children ever born in person, female and male, total women, total married and unmarried women, 1,2 3,4 and 5+ surviving children, female, male,1,2,3,4,5, 6 and 7+ birth order last year, number of women currently married and unmarried, number of births last year total, female and male children, and women with female, male and total surviving children as well.

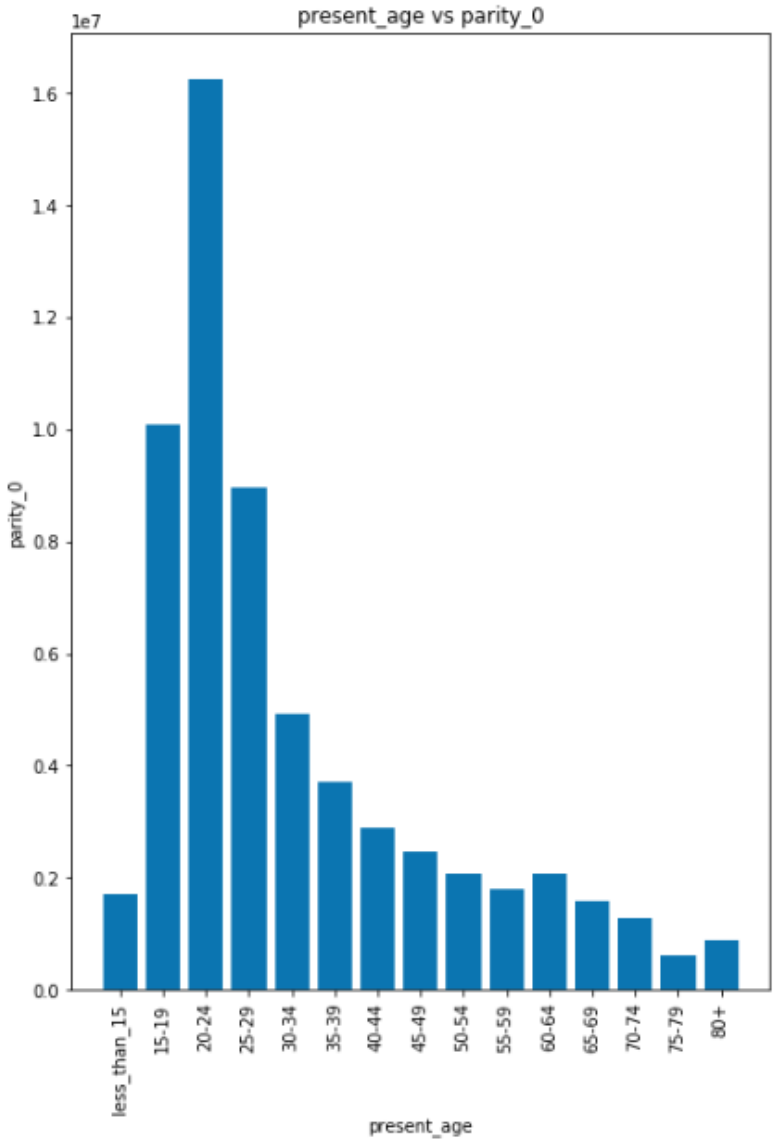

Figure 4.2: Present Age vs Parity 0

In figure 4.2, it is an analysis of the number of women with 0 parity corresponding to the present age of women. In this graph, the age group 20-24 has the highest. Then there is a steep reduction in women with no children. There is a small fluctuation in the age group of 60-64 and 80+. A similar pattern is followed by women with parity 1, parity 2, 0 surviving children, 1 surviving children, and 2 surviving children.

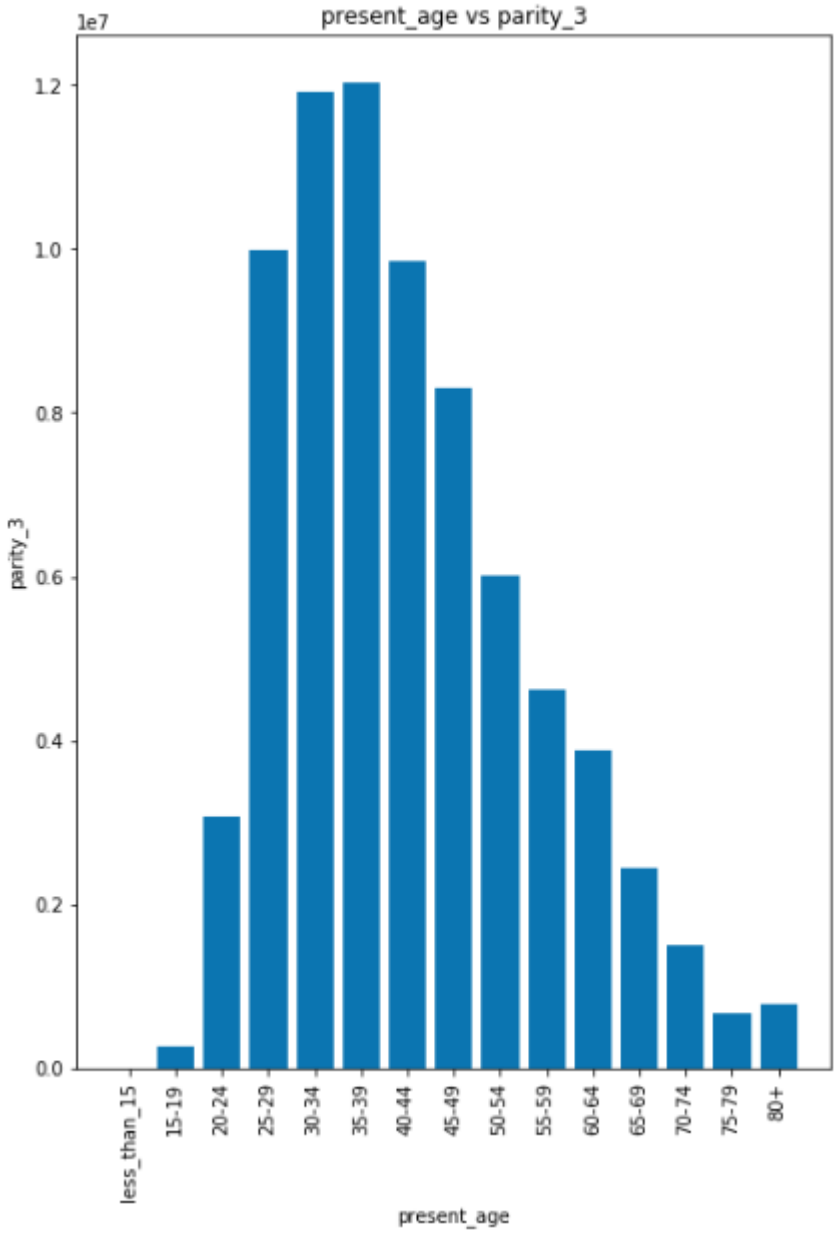

Figure 4.3: Present Age vs Parity 3

In figure 4.3, it is an analysis of the number of women with 3 parity corresponding to the present age of women. Parity 3 is common among the age group 30-39. There are fewer women between the age of 15-24 and no women at the age of less than 15. The slope reduces after the age of 40. A similar pattern is followed by women with 3 surviving children.

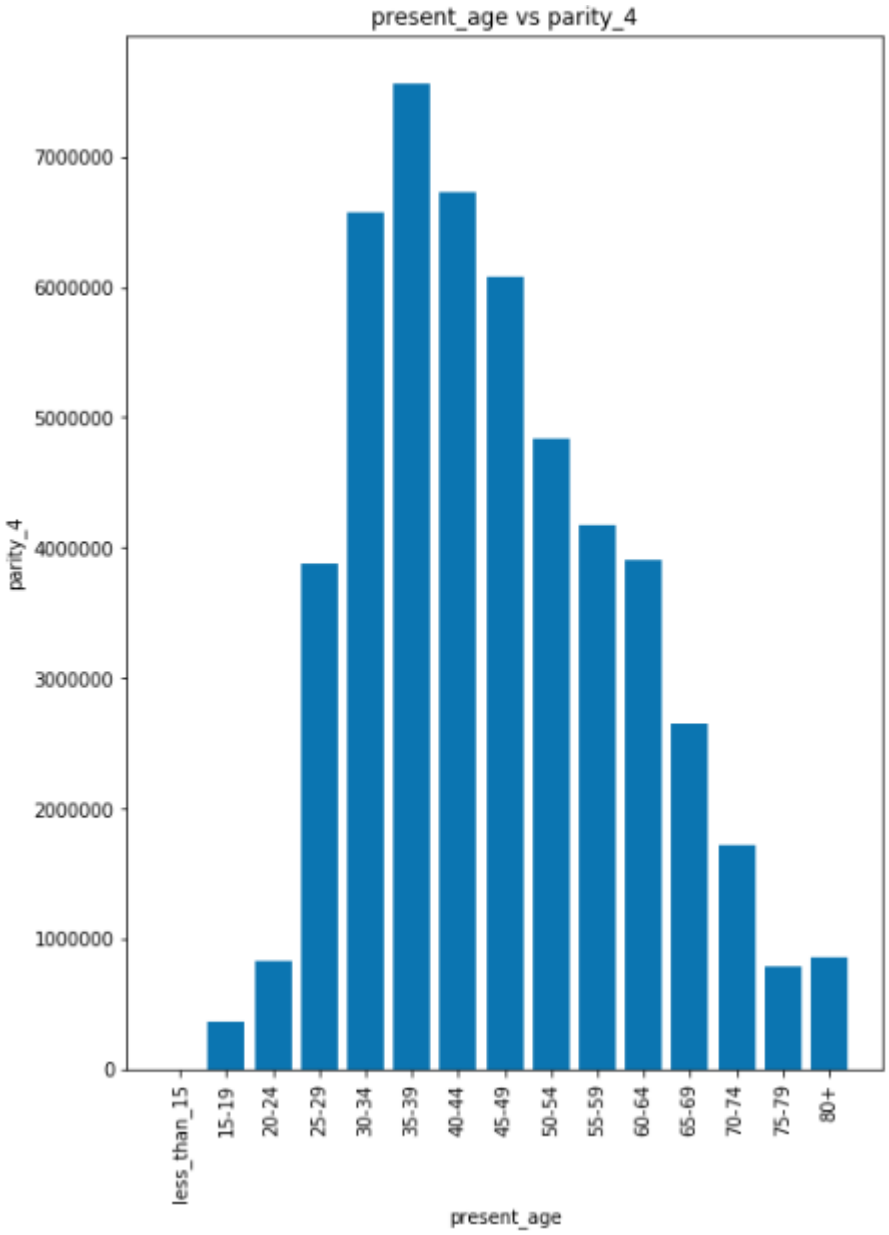

Figure 4.4: Present Age vs Parity 4

In figure 4.4, it is an analysis of the number of women with 4 parity corresponding to the present age of women. Parity 4 is common among the age group 35-39. There are fewer women between the age of 15-24 and no women at the age of less than 15. The age group of 20-24 has a major reduction compared to parity 3. The curve is a normal curve peaking at 35-39 and with an outlier at 80+. A similar pattern is followed by women with 4 surviving children.

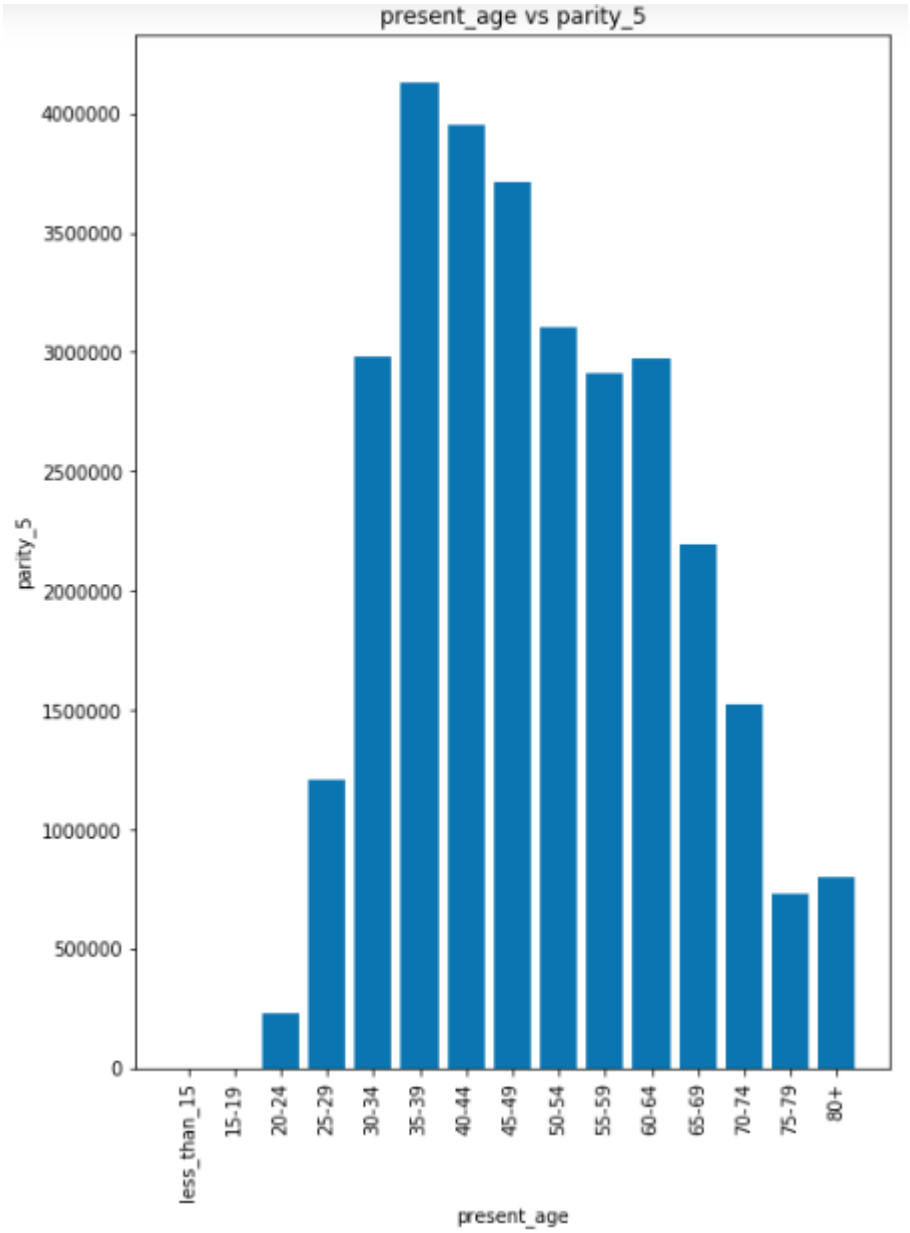

Figure 4.5: Present Age vs Parity 5

In figure 4.5, it is an analysis of the number of women with 5 parity corresponding to the present age of women. Parity 5 is not there among women of age below 19. The slope increases till 35-39 and then decreases gradually. There is a small increase in the age group of 60-64 and then it is gradually decreased with 80+ as an outlier. In parity 6 the peak is at 40-44 with the rest same as party 5. A similar pattern is followed by women with 5+ surviving children.

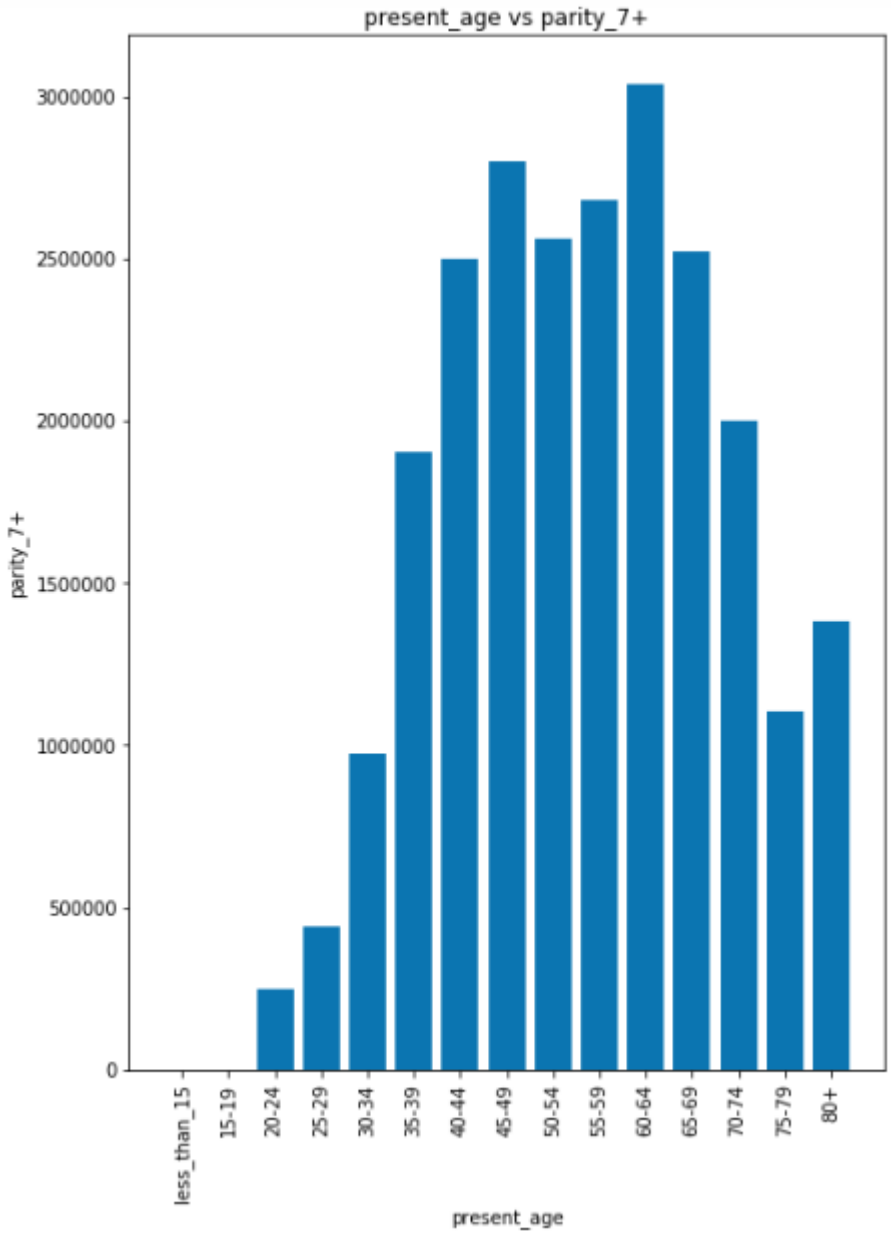

Figure 4.6: Present Age vs Parity 7+

In figure 4.6, it is an analysis of the number of women with 7+ parity corresponding to the present age of women. Parity 7+ is not there among women of age below 19. The number of women increases till 45-49. The peak is at 60-64. There is also an evident increase in women who are 80+.

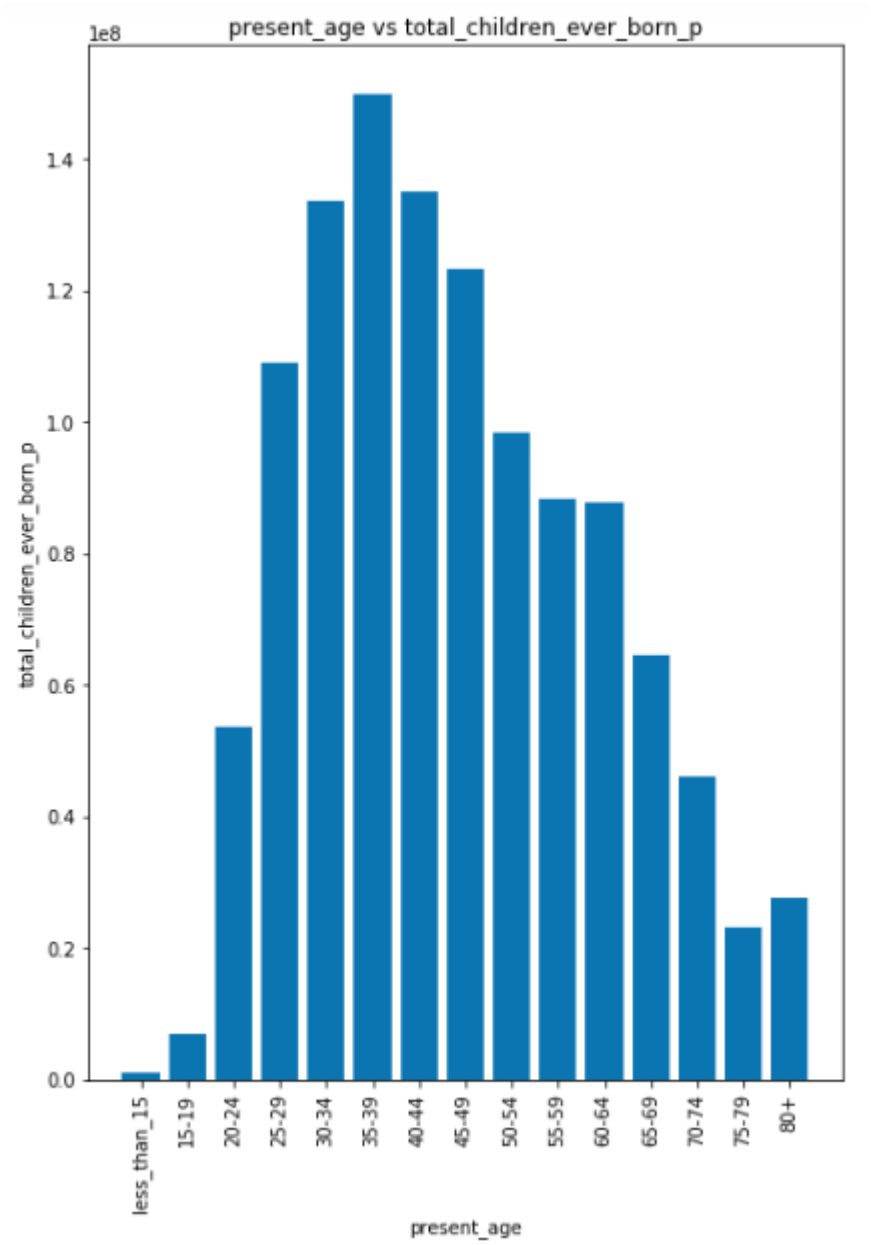

Figure 4.7: Present Age vs Total Children Ever Born

In figure 4.7, it is an analysis of total children ever born corresponding to the present age of women. The graph follows a normal flow with a peak at 35-39. 80+ is an outlier. The graph of total children ever born is similar to the graph of total surviving children in person, males, and females, and total children ever born who are male and female, although the values of the female are lower as compared to males in both the cases.

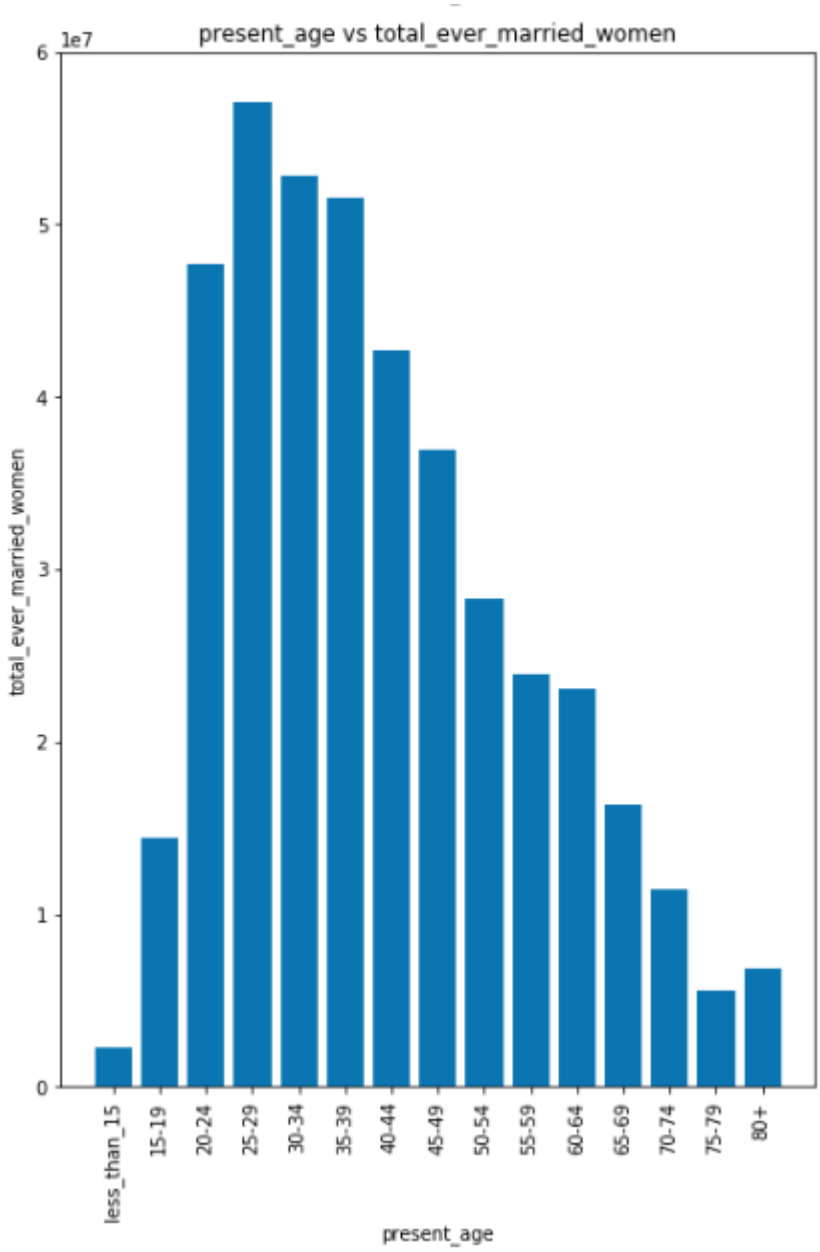

Figure 4.8: Present Age vs Total Ever Married Women

Figure 4.8 is an analysis of total ever married women corresponding to the present age of women. It is visible that there are women who are married below the age of 19. Most of the married women are in the age group of 25-29. The number of married women to date reduces gradually. 80+ is an outlier. The number of unmarried women peaks at ages below 15. The values then reduce until the age of 49. There are very few women who aren't married after that age.

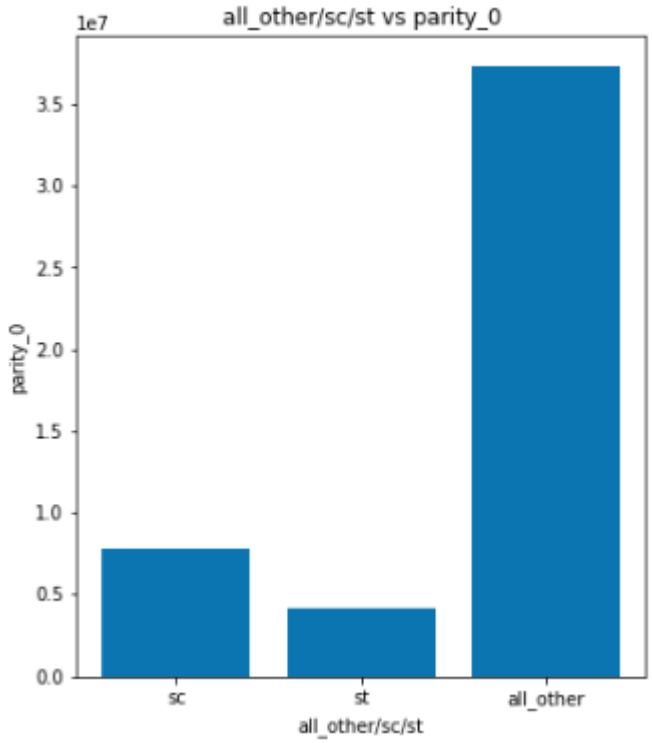

Figure 4.9: All/SC/ST vs Parity 0

Figure 4.9 is an analysis of women belonging to Scheduled caste, Scheduled tribe, or all of them corresponding to parity 0. The values are the highest in All castes in parity 1, parity 2, parity 3, parity 4, parity 5, parity 6, parity 7+, women with 0,1,2,3,4 and 5+ surviving children, total surviving children in total, female and male, total children ever born in total, female and male, Birth order last year – 1,2,3,4 and 5+, the total number of births last year, and total women married and unmarried. The pattern is the same for SC and ST as well. The SC is slightly higher than the ST.

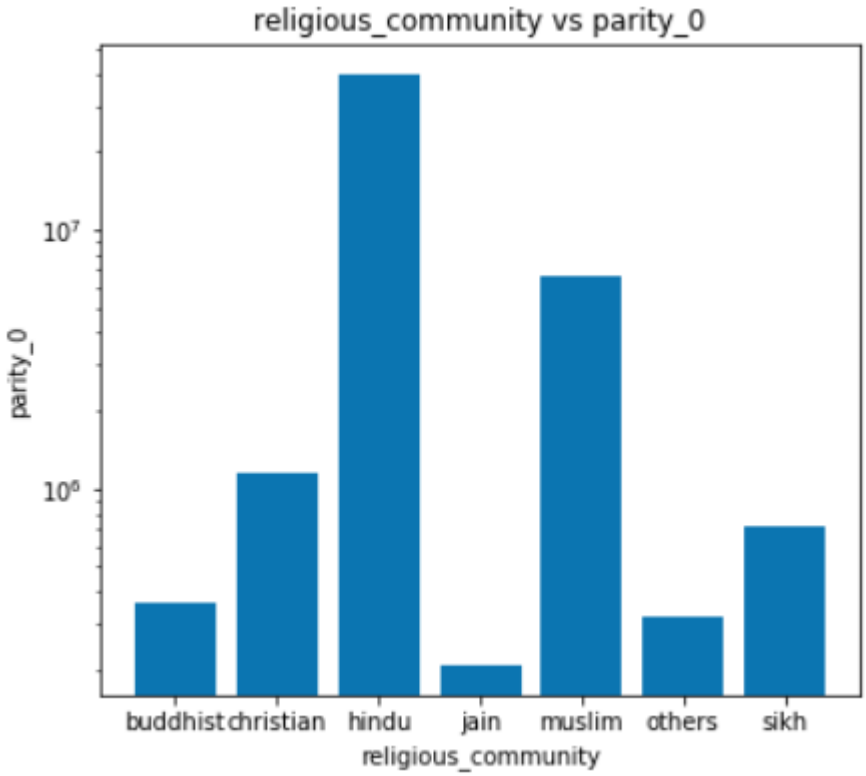

Figure 4.10: Religious Community vs Parity 0

Figure 4.10 is an analysis Religious community of the women corresponding to parity 0. The values are the highest in Hinduism in parity 1, parity 2, parity 3, parity 4, parity 5, parity 6, parity 7+, women with 0,1,2,3,4 and 5+ surviving children total surviving children in total, female and male, total children ever born in total, female and male, Birth order last year – 1,2,3,4 and 5+, the total number of births last year, and total women married and unmarried. The current plot uses log values since Hinduism has a very high value. This made it tough to read the other variables. The rest of the pattern is the same for other values as well.

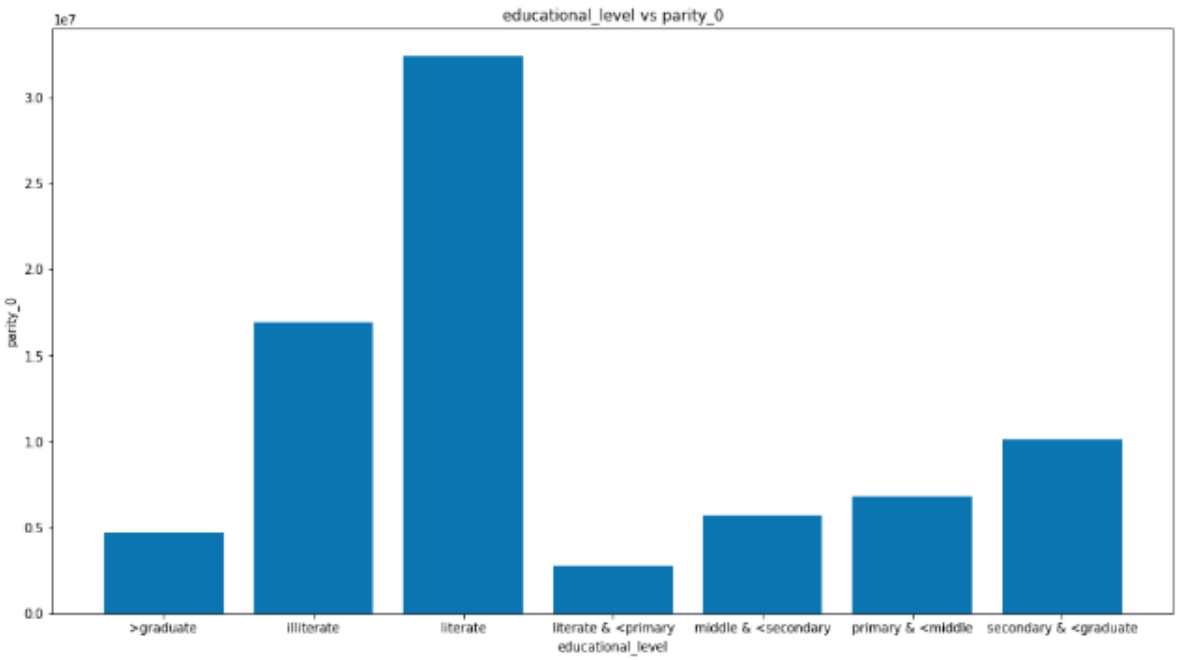

Figure 4.11: Educational Level vs Parity 0

Figure 4.11 is an analysis Educational Level of the women corresponding to parity 0. The literate level is very high for women who have no children. The values are similar in parity 1, parity 2, women with 0,1 and 2 surviving children, total surviving children in total, female and male as well.

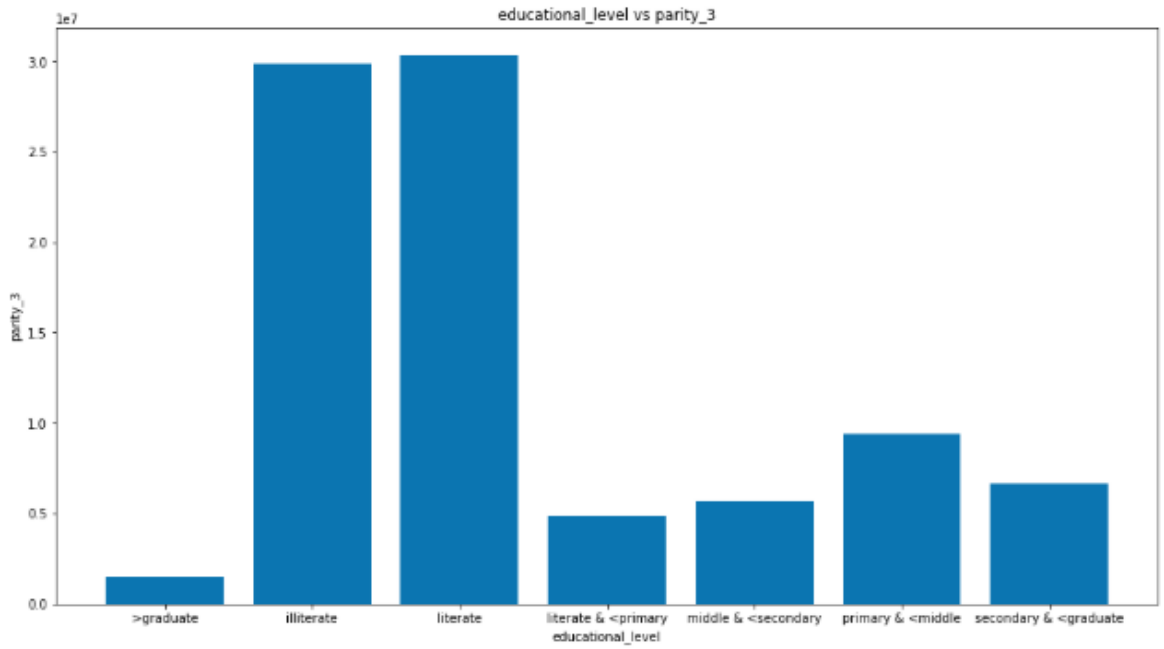

Figure 4.12: Educational Level vs Parity 3

Figure 4.12 is an analysis Educational Level of the women corresponding to parity 3. The literate and illiterate levels are almost similar among this group. There is also a significant reduction in the graduates' level as compared to parity 0 in figure 4.11.

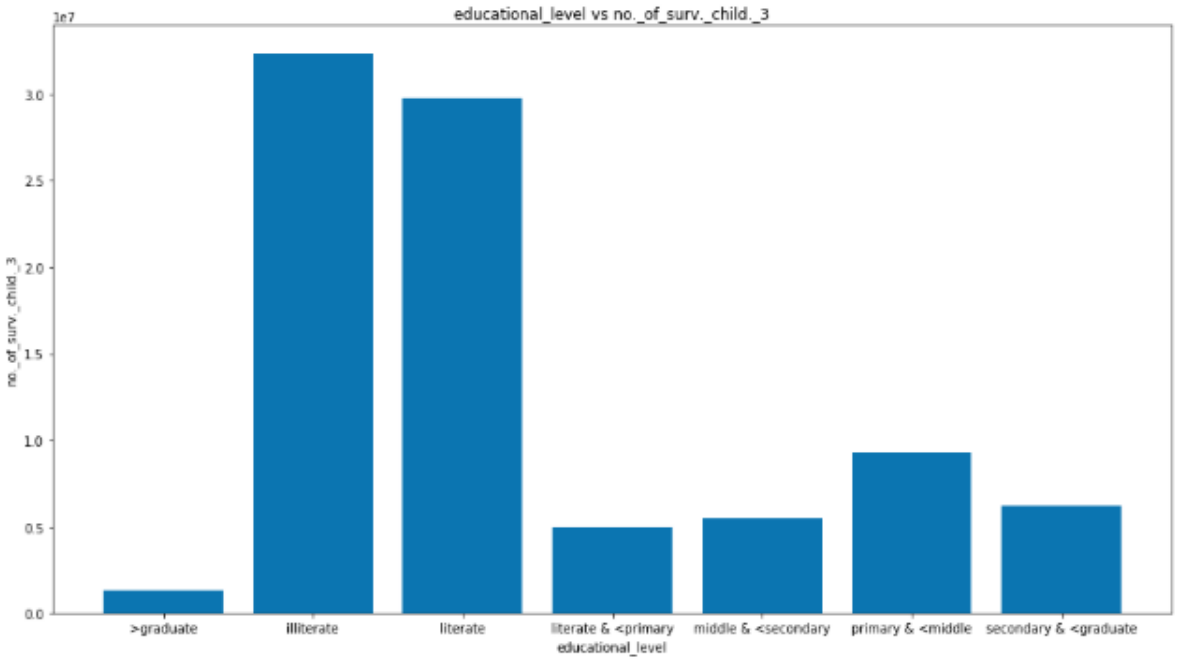

Figure 4.13: Educational Level vs Number of surviving children 3

Figure 4.13 is an analysis Educational Level of the women corresponding to the number of surviving child - 3. The literate level is a little lower than the illiterate level. There are lesser women with secondary and graduate education as compared to women with 0, 1, and 2 surviving children.

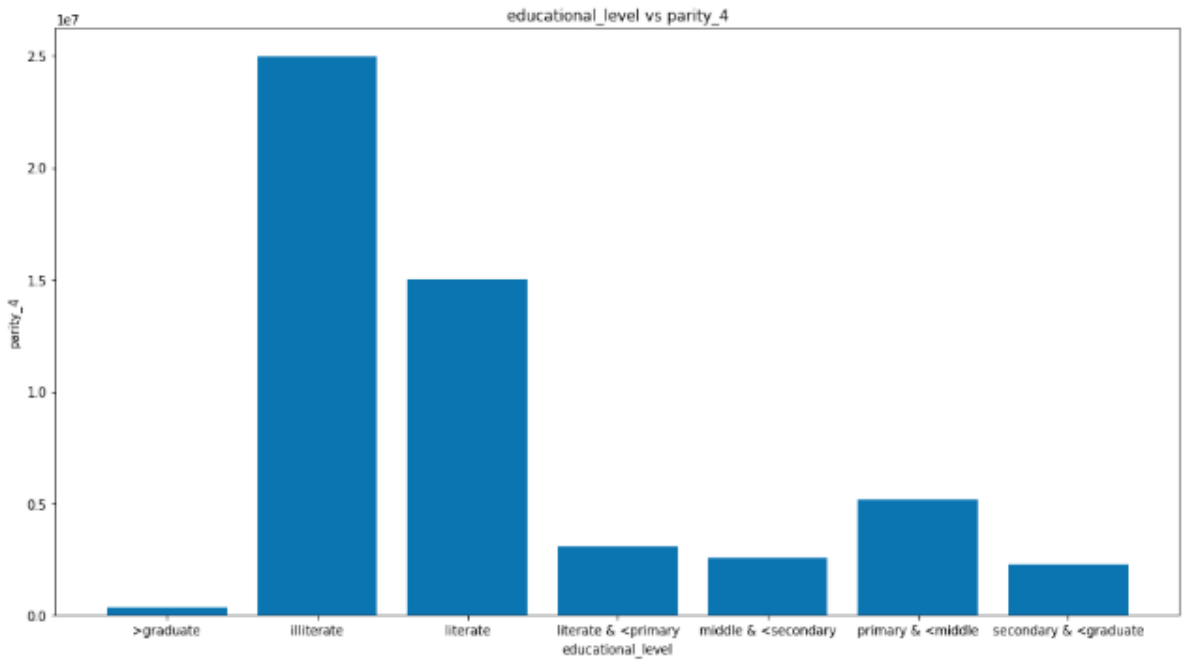

Figure 4.14: Educational Level vs Parity 4

Figure 4.14 is an analysis Educational Level of the women corresponding to parity 4. The illiterate level is very high for women with parity 4. Even though the illiteracy level is lesser than in parity 3 in figure 4.12, but the literate level is very low in parity 4. The values are similar in parity 5, parity 6, parity 7+, and 4 and 5+ surviving children as well.

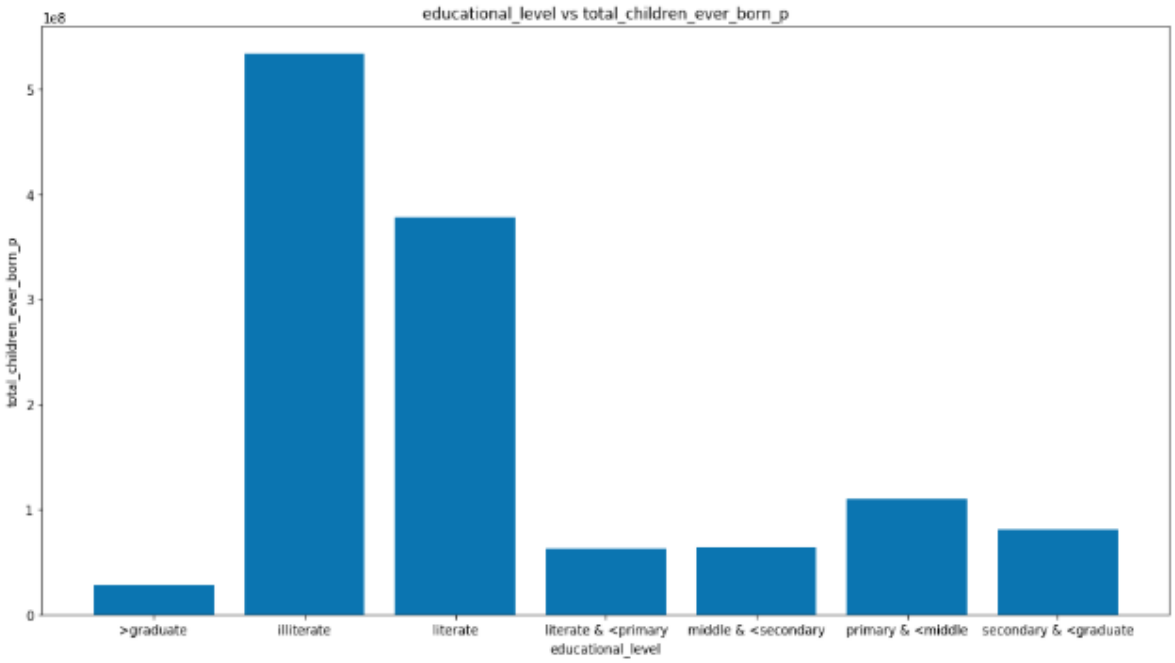

Figure 4.15: Educational Level vs Total Children ever born

Figure 4.15 is an analysis Educational Level of the women corresponding to Total Children ever born. The illiterate level is very among most women. Among the literates, most of the women have only primary education. There are very few that graduate from school. The flow of graph is similar in Total Children ever born female and male, total married and unmarried women, and total, female and male surviving children as well.

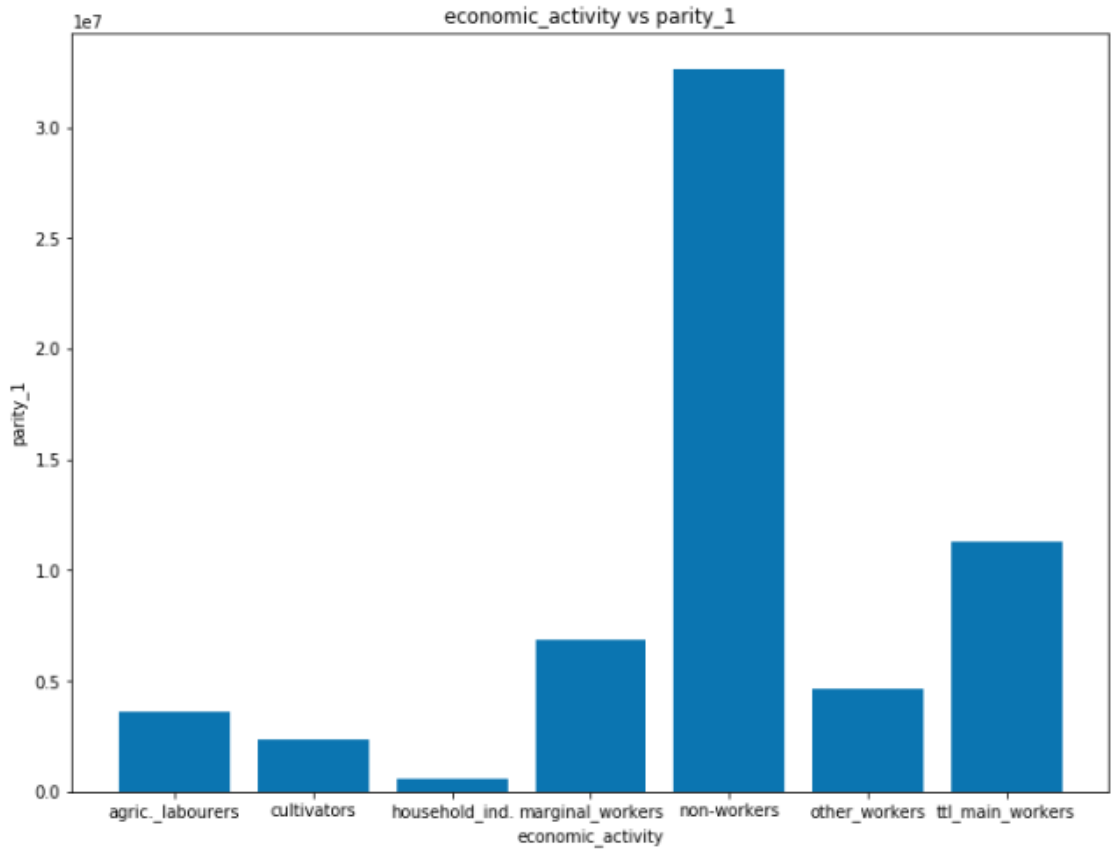

Figure 4.16: Economic Activity vs Parity 1

Figure 4.16 is an analysis Economic Activity of the women corresponding to parity 1. Most women in India are non-worker. This is the same in all the economic instances in India. The values are similar in parity 0, parity 2, parity 3, parity 4, parity 5, parity 6, parity 7+, total children ever born in persons, males and female, total married and unmarried women, women with 0, 1, 2, 3, 4 and 5+ surviving children and total, female and male surviving children as well.

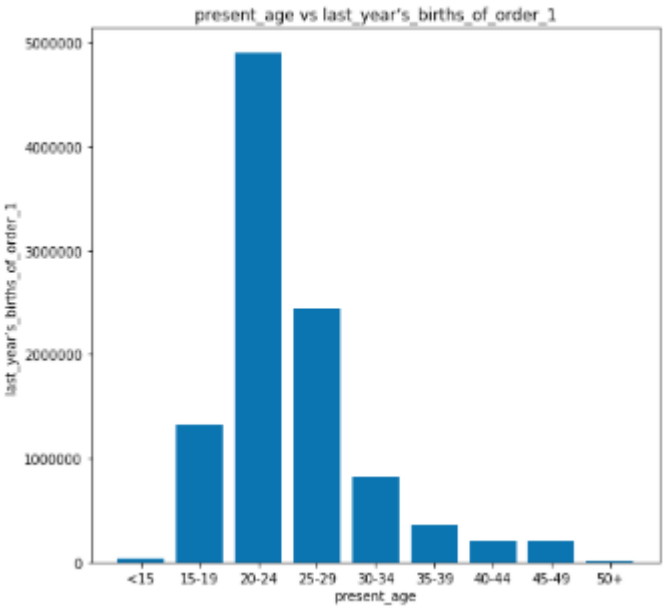

Figure 4.17: Present Age vs Birth order last year - 1

Figure 4.17 is an analysis Present age of the women corresponding to Birth order last year - 1. There are very few women under the age of 15 who had one child last year. Women in the age group 20-24 have the highest values. Then the slope reduces till 50+. The graph of Birth order last year – 2, the number of births last year male and female follows the same pattern.

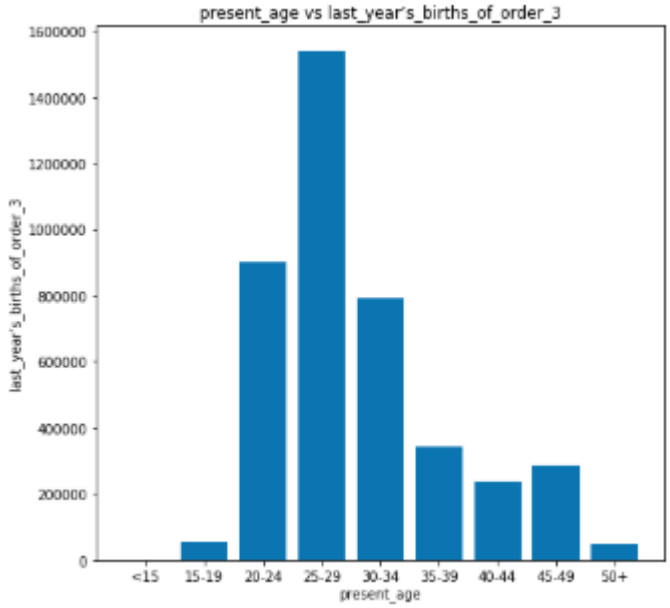

Figure 4.18: Present Age vs Birth order last year - 3

Figure 4.18 is an analysis Present age of the women corresponding to Birth order last year - 3. There are no women in this category below 15. Women in the age group 25-29 have the highest values. Then the slope reduces till 50+. There is a small fluctuation in the age group 45-49. The values are similar in Birth order last year – 4 as well.

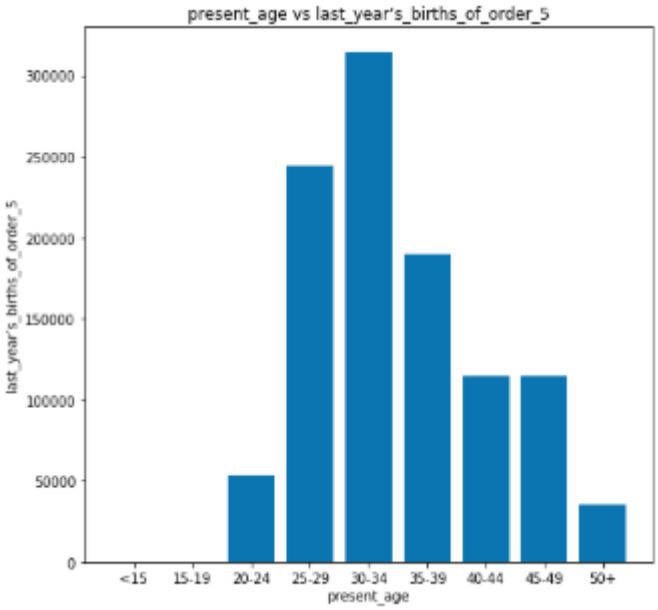

Figure 4.19: Present Age vs Birth order last year – 5

Figure 4.19 is an analysis Present age of the women corresponding to Birth order last year - 5. There are no women in this category below 19. Women in the age group 30-34 have the highest values. Then the slope reduces till 50+. The values are similar in Birth order last year – 6 and 7+ as well.

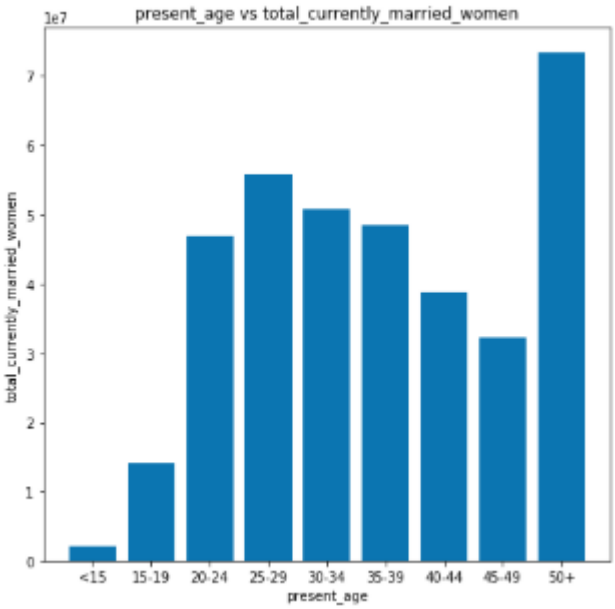

Figure 4.20: Present Age vs Currently Married Women

Figure 4.20 is an analysis Present age of the women corresponding to Currently Married women. There are a few women who are currently married at ages less than 15. The curve is normal and peaking at the 25-24 age group. Women who are 50+ is an outlier and there are 14 times more women who are married than unmarried in the age group of 50+.

4.3.1.2 Segmented Analysis

The segmented analysis involves understanding two variables together. For example, the segmented analysis will help understand the number of women in rural areas who are in the age group 15-19. In this age group and rural area are two variables. This helps in segregate even deeper into the data set to get a clearer idea.

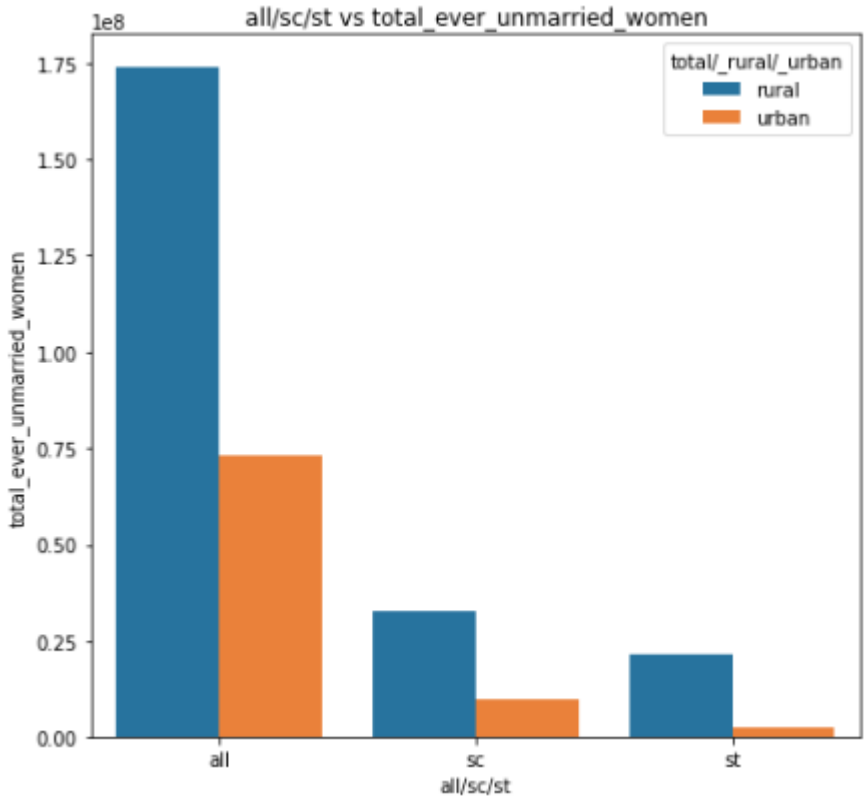

Figure 4.21: All/SC/ST vs Ever Married Women (Rural/Urban)

In figure 4.21, it is an analysis of the number of women ever married corresponding to All/SC/ST in rural/ urban. It is visible that total the unmarried women in rural areas is more than in urban areas in all three cases. It is the same in case of parity 1, parity 2, parity 3, parity 4, parity 5, parity 6, parity 7+, total children ever born in person, female and male,1,2 3,4 and 5+ surviving children, female and male surviving children,1,2,3,4,5, 6 and 7+ birth order last year, number of women currently married and unmarried, number of births last year total, female and male children.

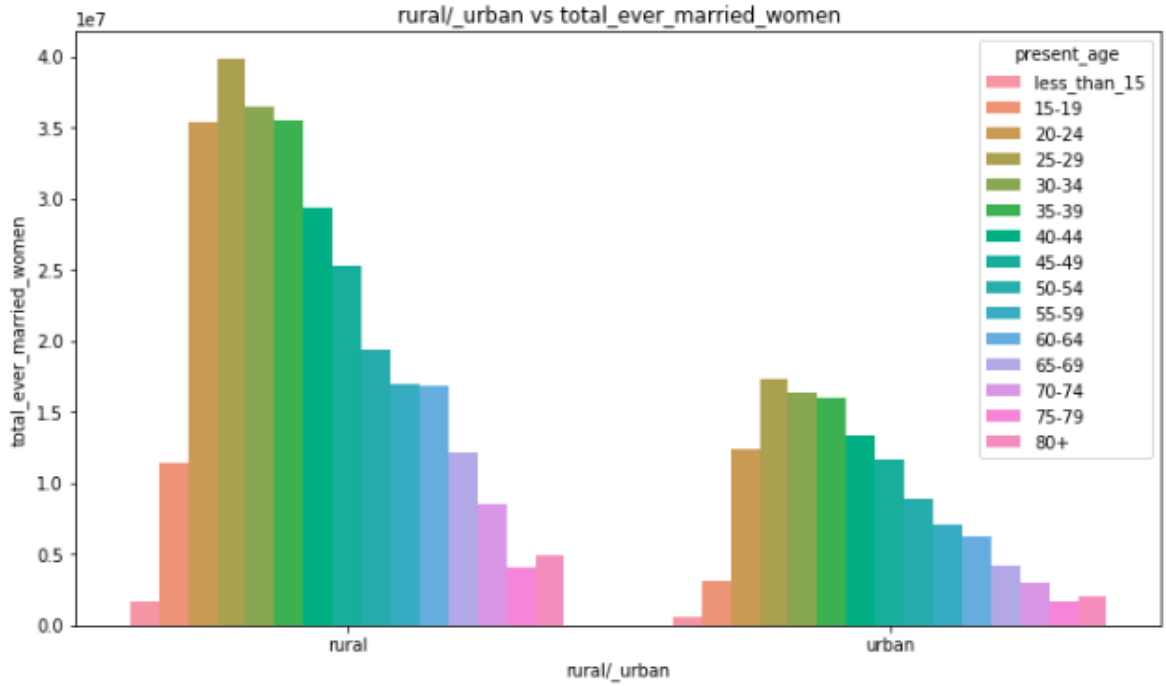

Figure 4.22: Rural/Urban vs Ever Married Women(Present Age)

In figure 4.22, it is an analysis of the number of women ever married corresponding to different age groups in rural or urban. This slope in both rural and urban follows the same as found in Figure 4.8. There is a slight increase in rural areas since the number of married women in rural areas is higher as seen in Figure 4.1. It is the same in case of parity 1, parity 2, parity 3, parity 4, parity 5, parity 6, parity 7+, total children ever born in person, female and male, total women, 1,2 3,4 and 5+ surviving children, female and male surviving children.

In the analysis of the number of births last year (male and female) corresponding to different age groups in rural or urban. It peaks at age 20-24. It is similar in the case of Birth order last year – 1. In another analysis of the Birth order last year – 2 corresponding to different age groups in rural or urban. It peaks at age 20-24 in rural areas but in urban areas, it is highest at 25-29. When the Birth order last year – 3 and 4 were taken corresponding to different age groups in rural or urban. It peaks at age 25-29 in both. It is similar in the case of Birth order last year – 1. Birth order last year – 5 and 6 taken corresponding to different age groups in rural or urban peaks at age 30-34. Birth order last year 7+ taken corresponding to different age groups in rural or urban peaks at age 35-39. The slope is a normal curve in both rural and urban areas in all cases.

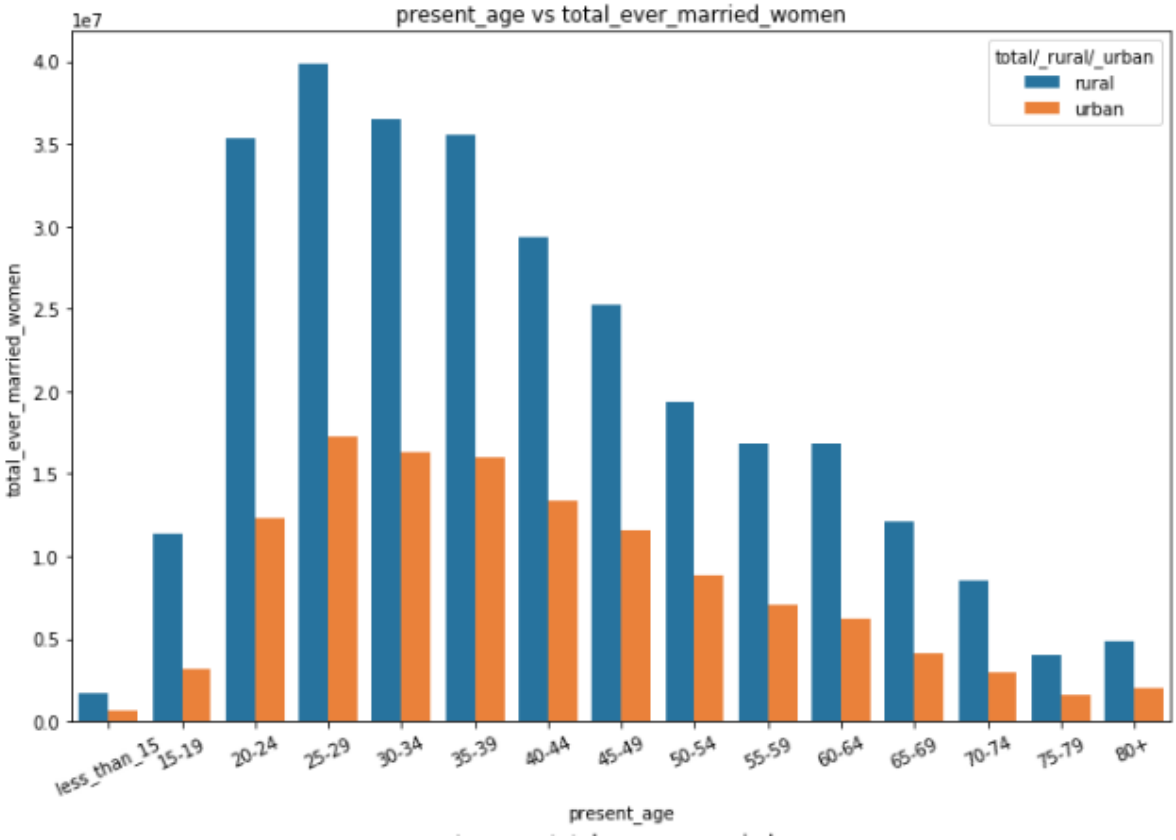

Figure 4.23: Rural/Urban vs Ever Married Women(Rural/Urban)

In figure 4.23, the number of total ever married women parallel to their present age in rural as well as urban areas in different age groups. At the age less than 15 there are very few married women. In the age group 15-19 and 20-24 the graph, the number of women married is higher in rural areas than in urban areas. The graph then falls but the women married in the rural area remain more than urban areas. Most of the unmarried women are in the age group of less than 15 with most women from rural areas. There is a very small part of women who are unmarried after the age of 35.

In the analysis of 1,2,3,4,5, 6 and 7+ birth order last year, the number of women currently married and unmarried, and the number of births last year total, female and male children taken against present age in rural/urban area, the women in the urban area is outnumbered by the women in rural areas.

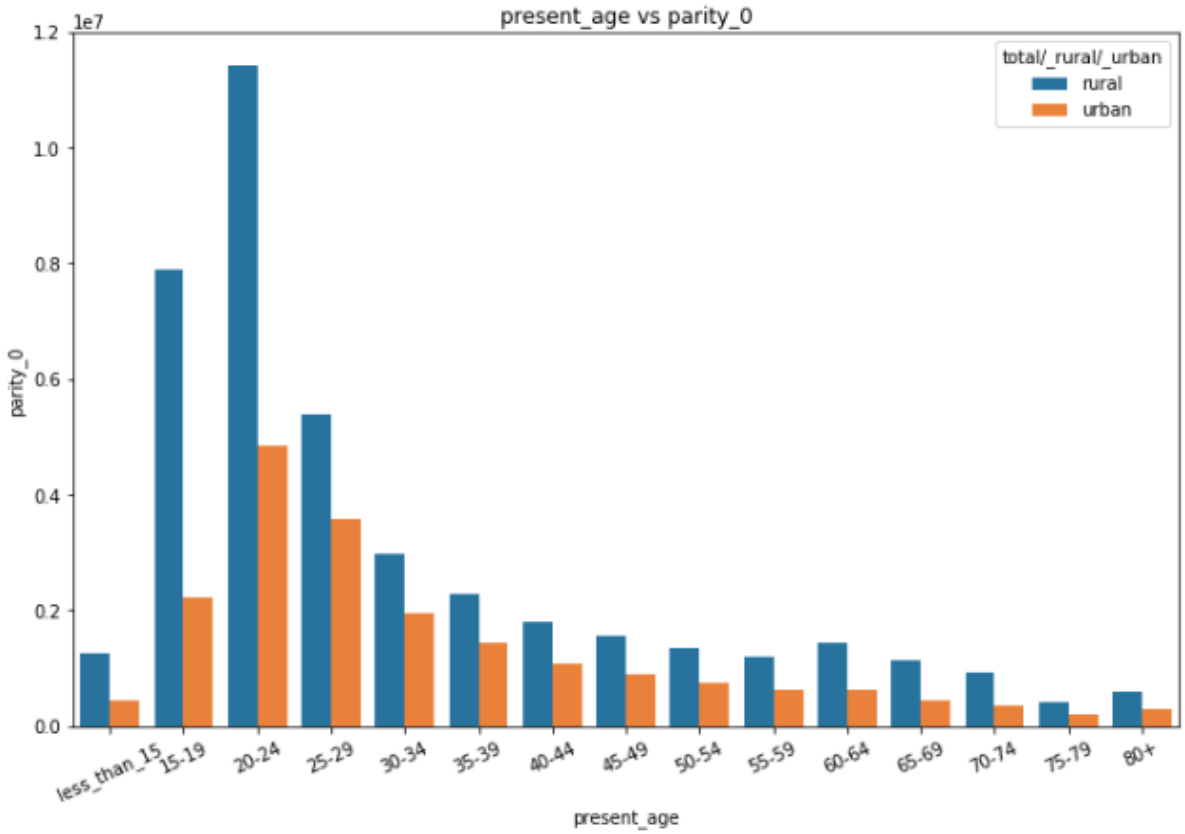

Figure 4.24: Rural/Urban vs Parity 0(Rural /Urban)

In figure 4.24, it is an analysis of the number of women with parity 0 corresponding to rural or urban in different age groups. In all the cases there are more women in rural areas. It is the same in case of parity 1, parity 2, parity 3, parity 4, parity 5, parity 6, parity 7+, total children ever born in person, female and male surviving children, 1,2 3,4 and 5+ surviving children.

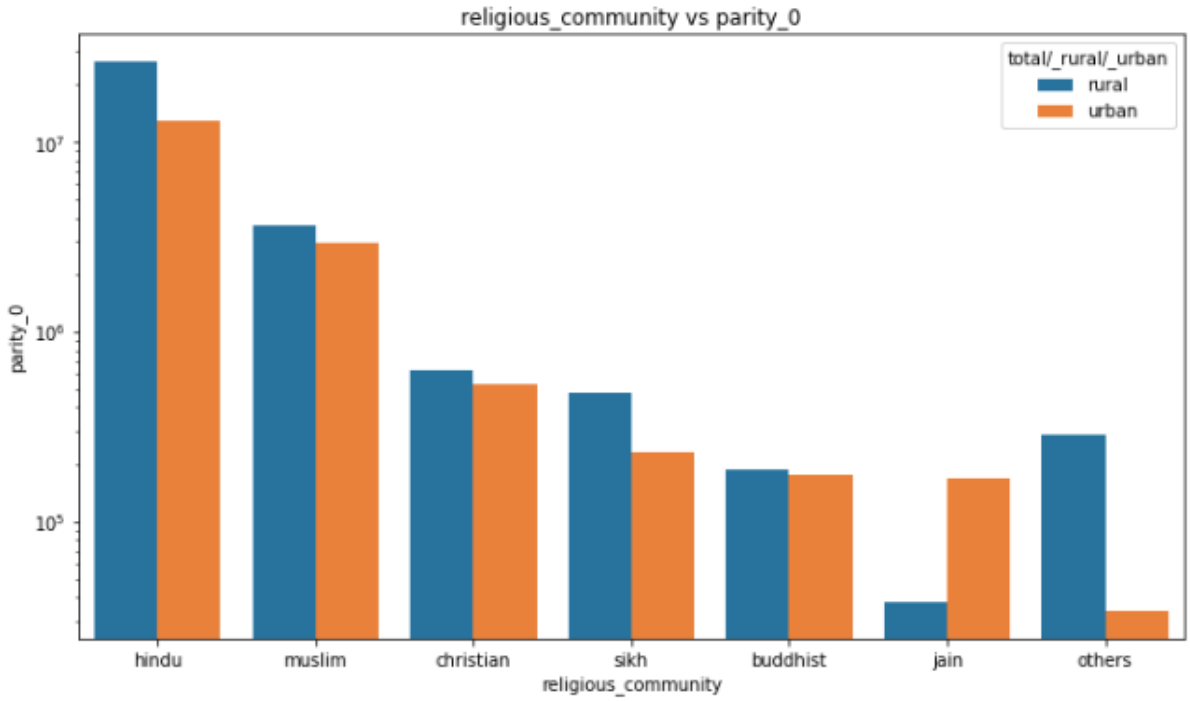

Figure 4.25: Religious Community vs Parity 0(Rural/Urban)

In figure 4.25, it is an analysis of the number of women with parity 0 corresponding to the religious community in rural or urban. There are more women in rural areas except in the Jain community. It is the same in case of parity 1, parity 2, parity 3, parity 4, parity 5, parity 6, parity 7+, total children ever born in person, female and male, total women, total married and unmarried women, 1,2 3,4 and 5+ surviving children, female and male surviving children,1,2,3,4,5, 6 and 7+ birth order last year, number of women currently married and unmarried, and number of births last year total, female and male children

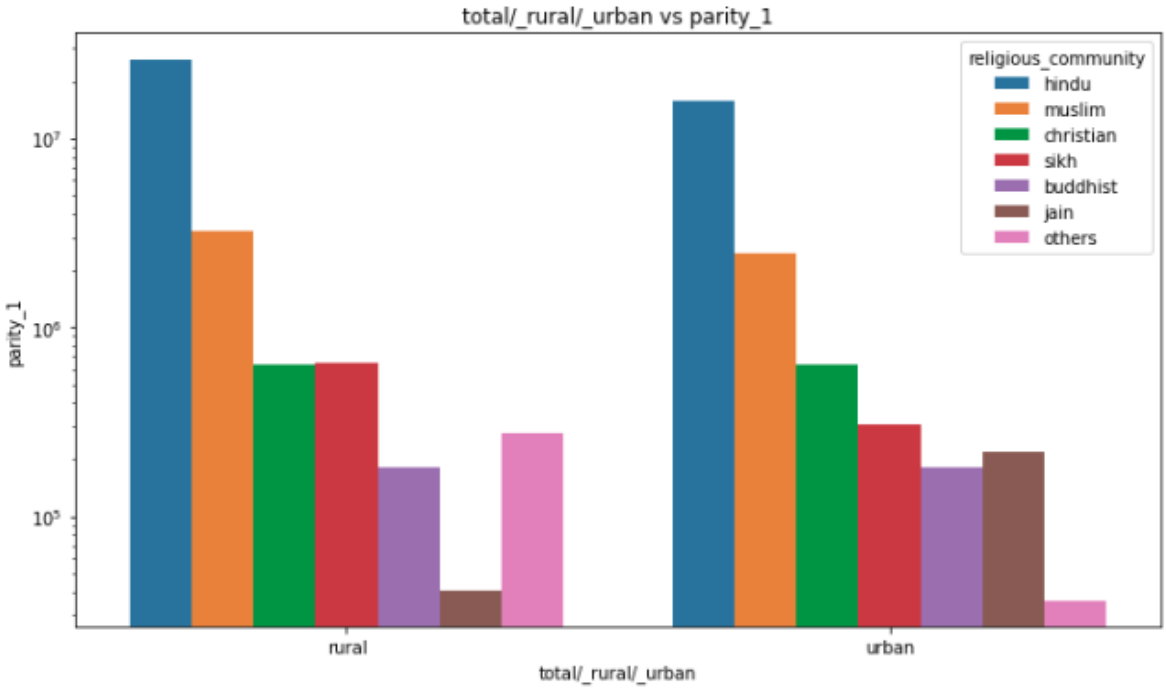

Figure 4.26: Religious Community vs Parity 1(Religious Community)

In figure 4.26, it is an analysis of the number of women with parity 1 corresponding to rural or urban in the religious community. In rural areas, the number of Christians and Sikhs are almost the same. Meanwhile, in urban areas, Christians are more than Sikhs. Hinduism is the most in both. The Jains seem to be lesser in rural as compared to urban. It is the same in case of parity 1, parity 2, parity 3, parity 4, parity 5, parity 6, parity 7+, total children ever born in person, female and male, total women, total married and unmarried women, 1,2 3,4 and 5+ surviving children, female and male surviving children,1,2,3,4,5, 6 and 7+ birth order last year, number of women currently married and unmarried, and number of births last year total, female and male children.

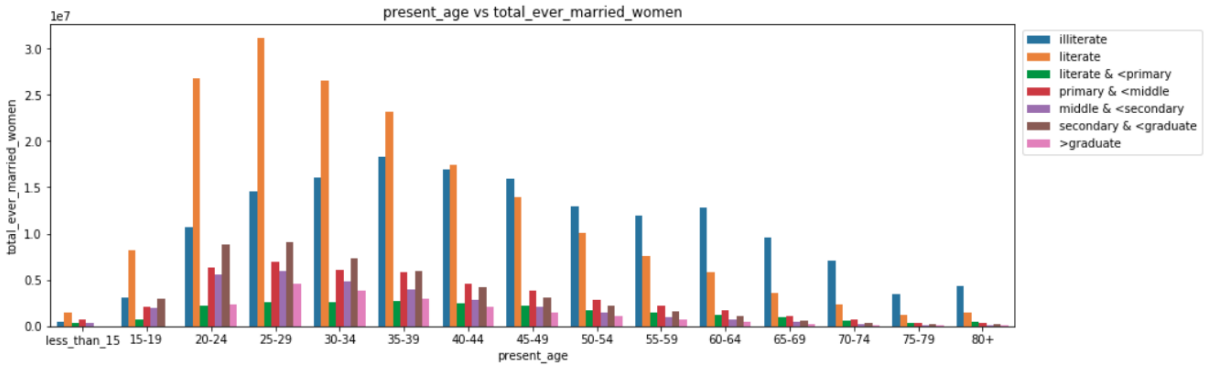

Figure 4.27: Present Age vs Total Ever Married Women (Educational Level)

In figure 4.27, it is an analysis of the number of women present age corresponding to total women ever married in educational level. Literate women are at the peak till age 44 and then illiteracy takes the turn. Most women have secondary education till the age of 34 among the literate women. After the age of 34, most women have only completed primary education. It is the same in case of parity 0, parity 1, parity 2, parity 3, and 0 and 1 surviving children

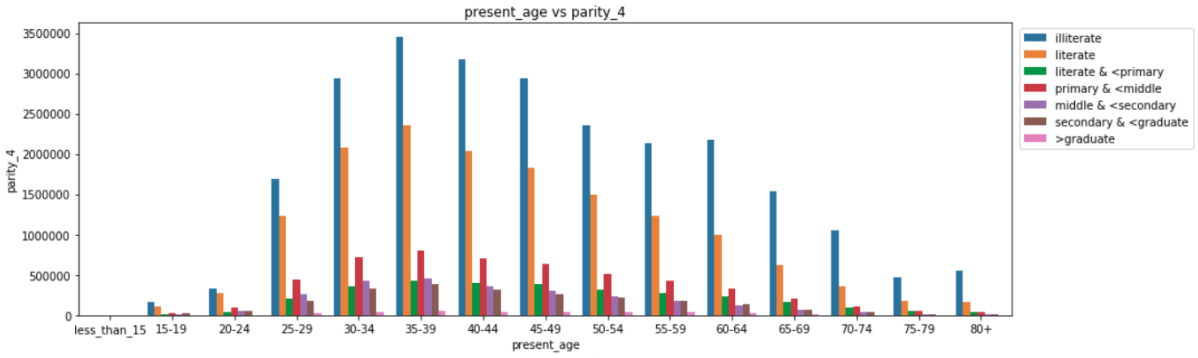

Figure 4.28: Present Age vs Parity 4 (Educational Level)

In figure 4.28, it is an analysis of the number of women present age corresponding to total women ever married in educational level. It is visible that women with 4 children have high rates of illiteracy throughout all ages. And among the literate women, most of them have only primary education. It is the same in case of parity 5, parity 6, parity 7+, and 4 and 5+ surviving children.

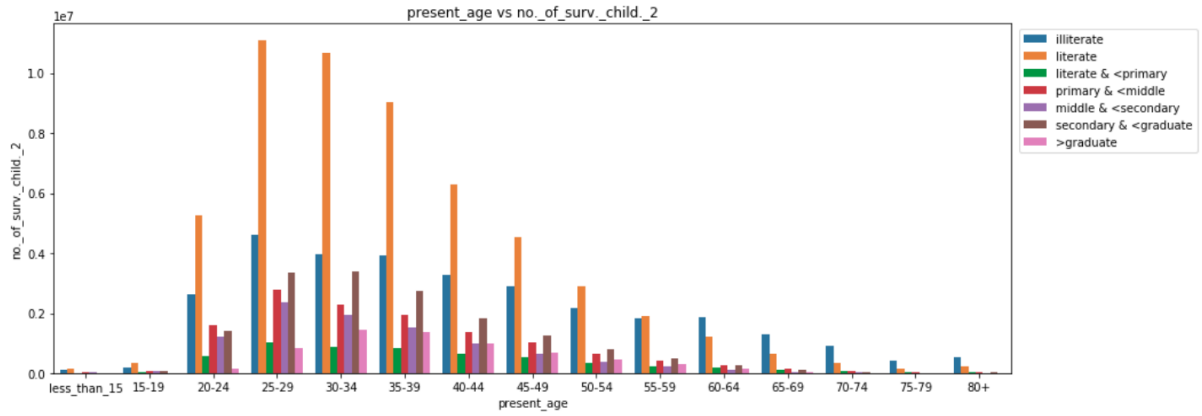

Figure 4.29: Present Age vs Women with 2 Surviving Children (Educational Level)

In figure 4.29, it is an analysis of the number of women present age corresponding to women with 2 surviving children in educational level. Literate women are at the peak till age 59 and then illiteracy takes the turn. Most women have secondary education till the age of 60 among the literate women. After the age of 60, most women have only completed primary education.

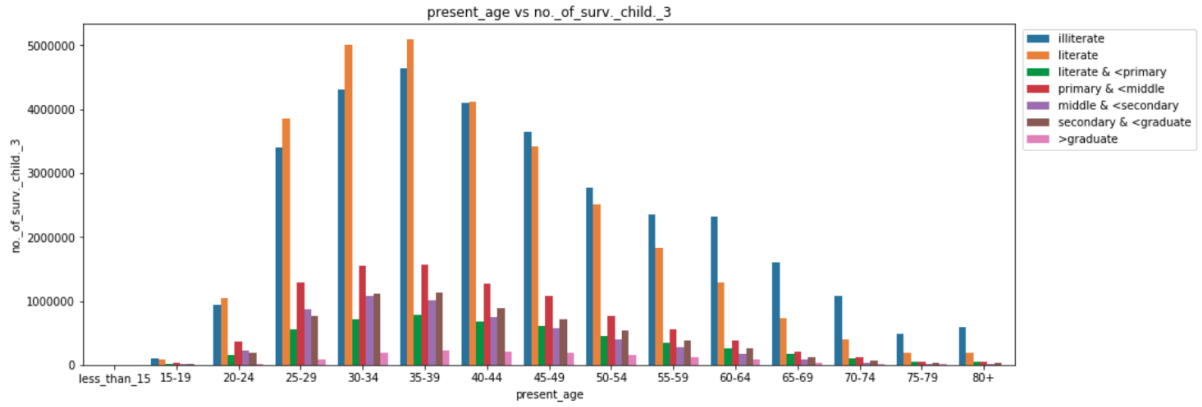

Figure 4.30: Present Age vs Women with 3 Surviving Children (Educational Level)

In figure 4.30, it is an analysis of the number of women present age corresponding to women with 3 surviving children in educational level. Literate women are at the peak till age 44 and then illiteracy takes the turn. Most women have primary education till the age of 74 among the literate women.

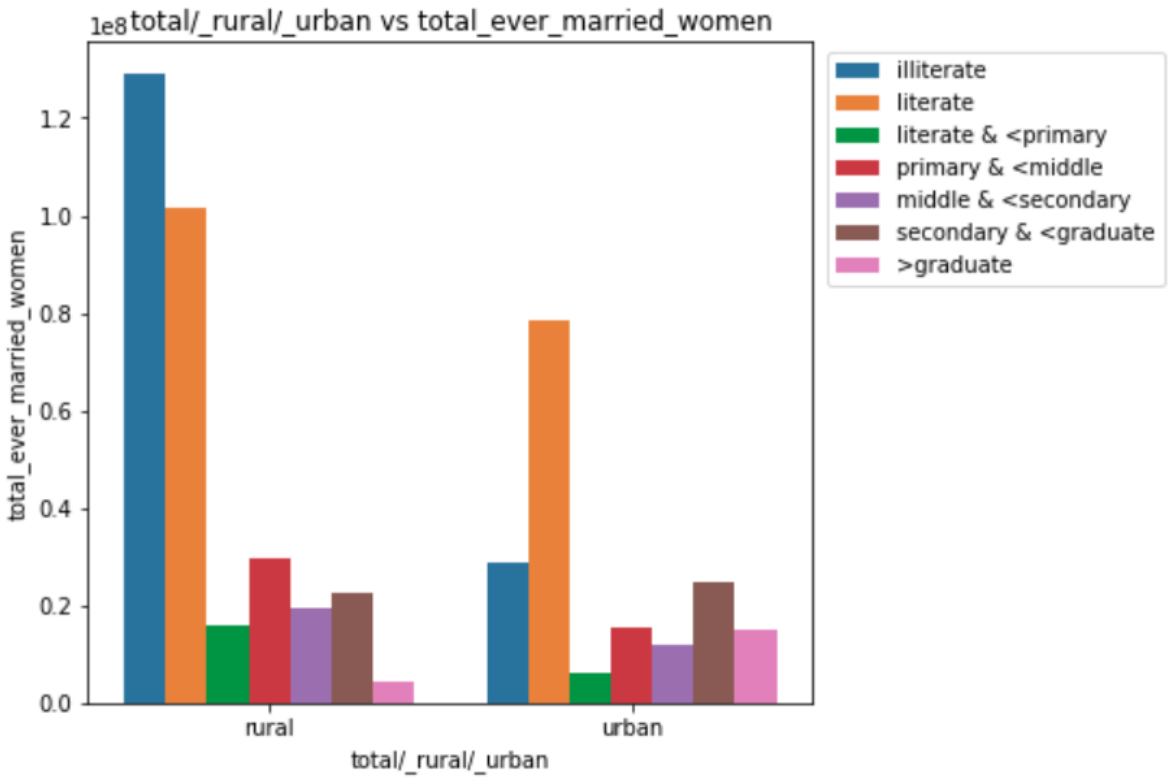

Figure 4.31: Rural/Urban vs Total Ever Married Women (Educational Level)

In figure 4.31, it is an analysis of the number of women in rural/urban areas corresponding to total women ever married in educational level. Illiteracy rates are mostly in rural areas and literacy rates are highest in urban areas . among the literate married women in rural areas, most of them have primary education. Meanwhile in urban regions more married women have finished secondary education. The number of graduates are more in urban as compared to women in rural areas. It is the same in case of 3 and 4 surviving children, female and male surviving children, 3 and 4 birth order last year, and the number of women currently married

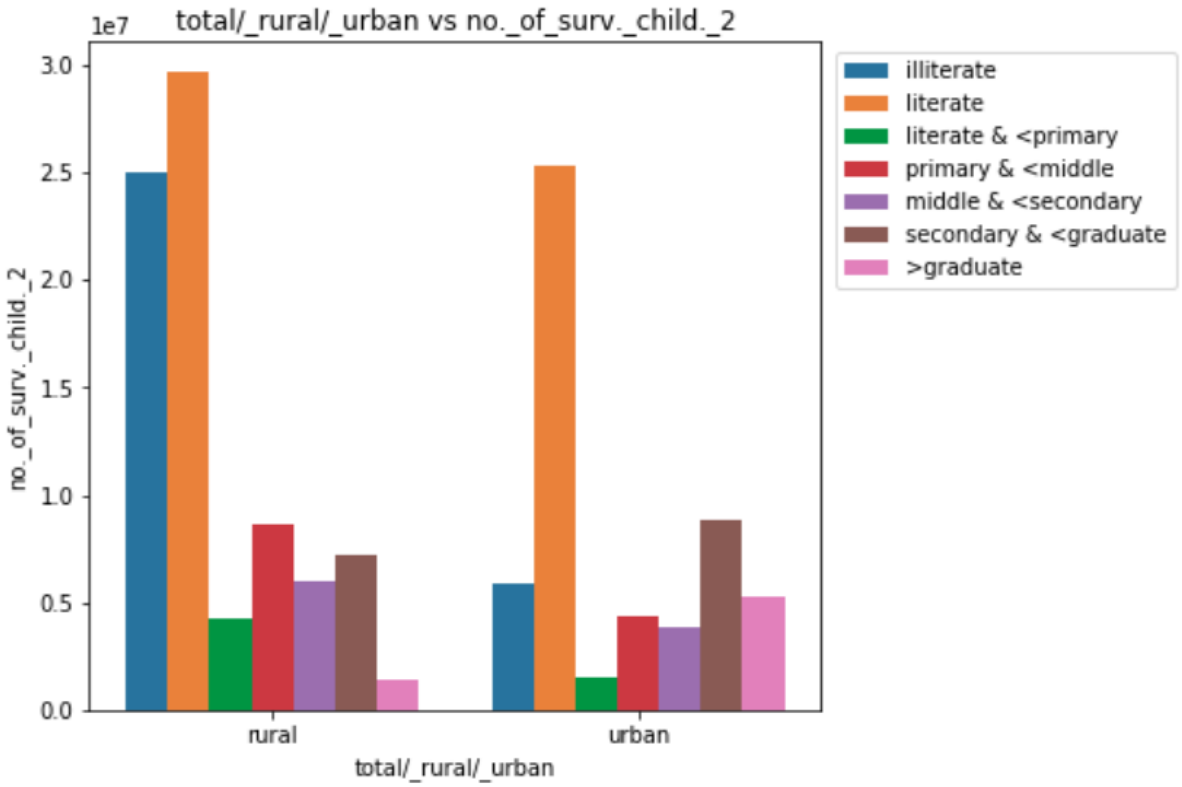

Figure 4.32: Rural/Urban vs Women with 2 Surviving Children (Educational Level)

In figure 4.32, it is an analysis of the number of women in rural/urban areas corresponding to women with 2 surviving children in educational level. The literacy rate is highest in urban and rural areas. Among the literate in rural areas, most of them have primary education. Meanwhile in urban regions more married women have finished secondary education. The number of graduates are more in urban as compared to women in rural areas. It is the same in case of 0,1 and 2 surviving children, 1 and 2 birth order last year, and number of births last year female and male children

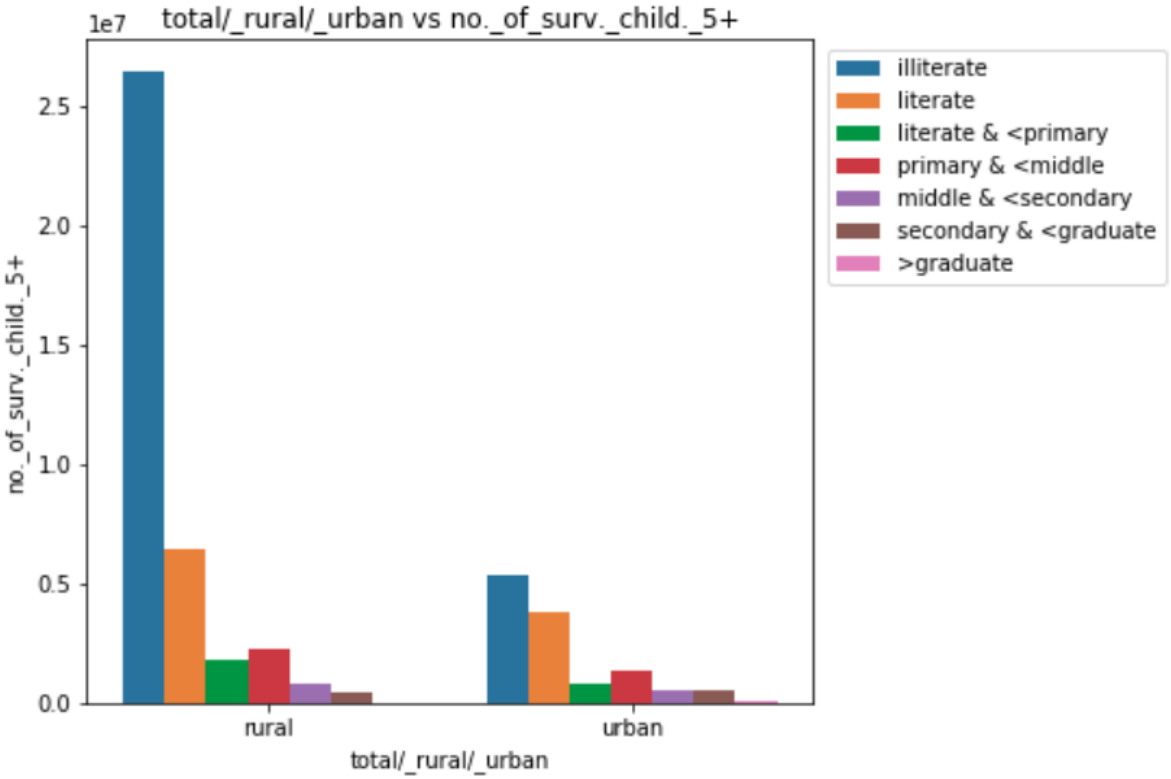

Figure 4.33: Rural/Urban vs Women with 5+ Surviving Children (Educational Level)

In figure 4.33, it is an analysis of the number of women in rural/urban areas corresponding to women with 5 surviving children in educational level. The illiteracy rate is highest in urban and rural areas. Among the literate in rural and urban areas, most of them have primary education. The number of graduates are more in urban as compared to women in rural areas but are not significant. It is the same in the case of 5, 6, and 7+ birth order last year.

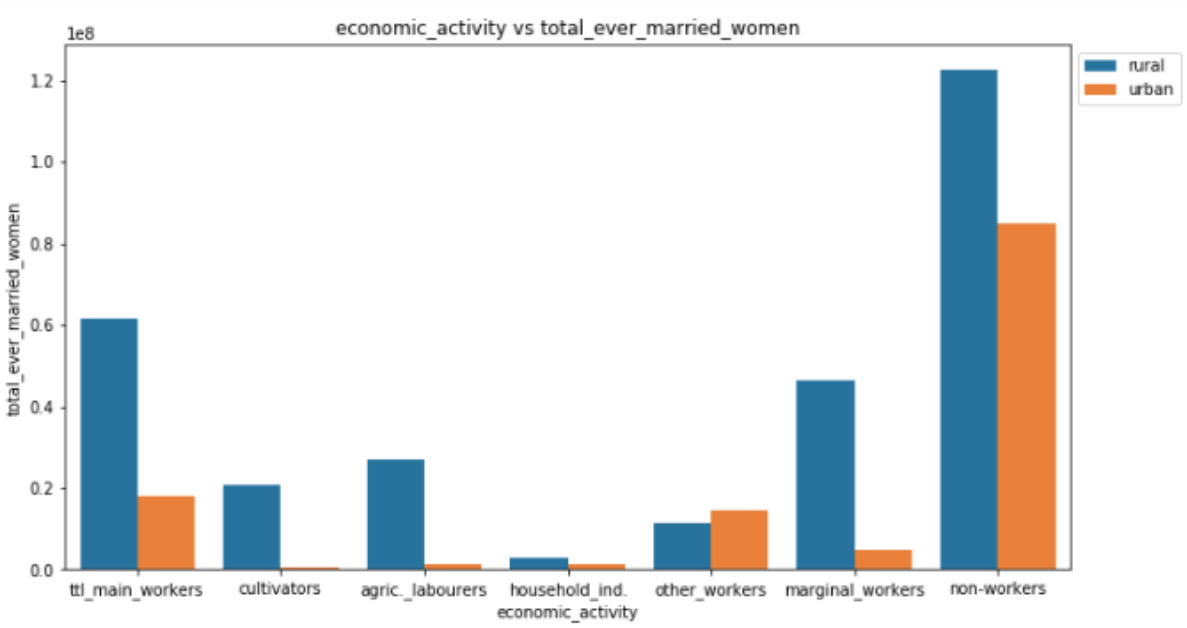

Figure 4.34: Economic Activity vs Total Ever Married Women (Rural/Urban)

In figure 4.34, it is an analysis of the number of women with their occupation corresponding to total women ever married in rural/urban. Most of the married women who work are in rural areas. There is a major difference among the marginal workers when compared in rural and urban It is the same in case of parity 0, parity 1, parity 2, parity 3, parity 4, parity 5, parity 6, parity 7+, total children ever born in person, female and male, total women, total married and unmarried women, 0,1,2 3,4 and 5+ surviving children, female and male surviving children,1,2,3,4,5, 6 and 7+ birth order last year, the number of women currently married and unmarried, and the number of births last year female and male children.

4.3.1.3 Bivariate Analysis

The bivariate analysis helps understand the correlation between the two variables. This helps to understand which factors are the most effective. Heat maps are used to get a better understanding of the variables.

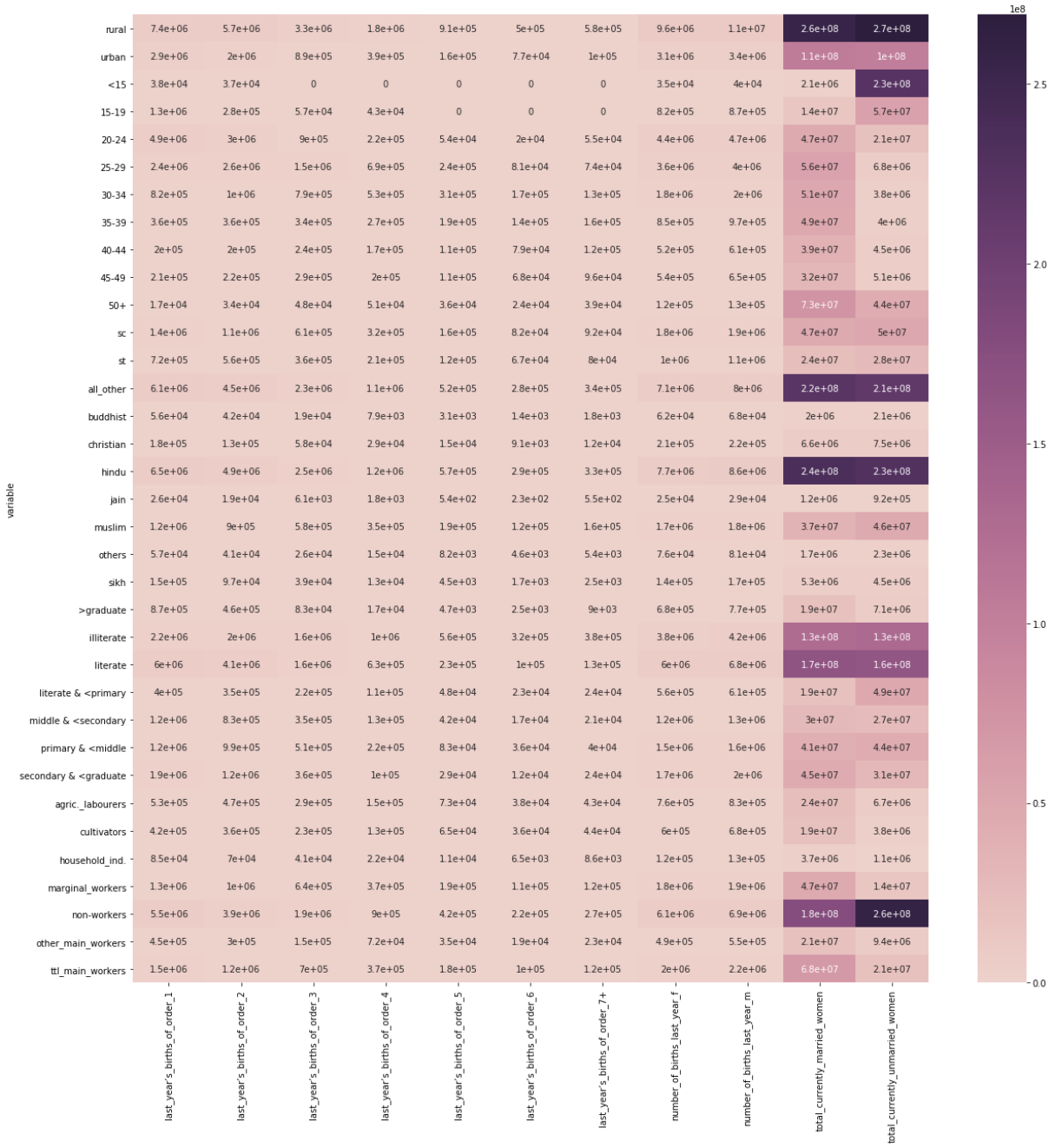

Figure 4.35: Heat map for Last Year's Birth Order, Births Last Year and Current Marital Status

Figure 4.35, shows the correlation between:

- Currently unmarried women with non-workers, Hinduism, Muslim community, below age 15, literate women, illiterate women, women other than SCand ST, urban area, and rural areas.
- Currently married women with non-workers, literate, women other than SCand ST, Hinduism, and rural areas.

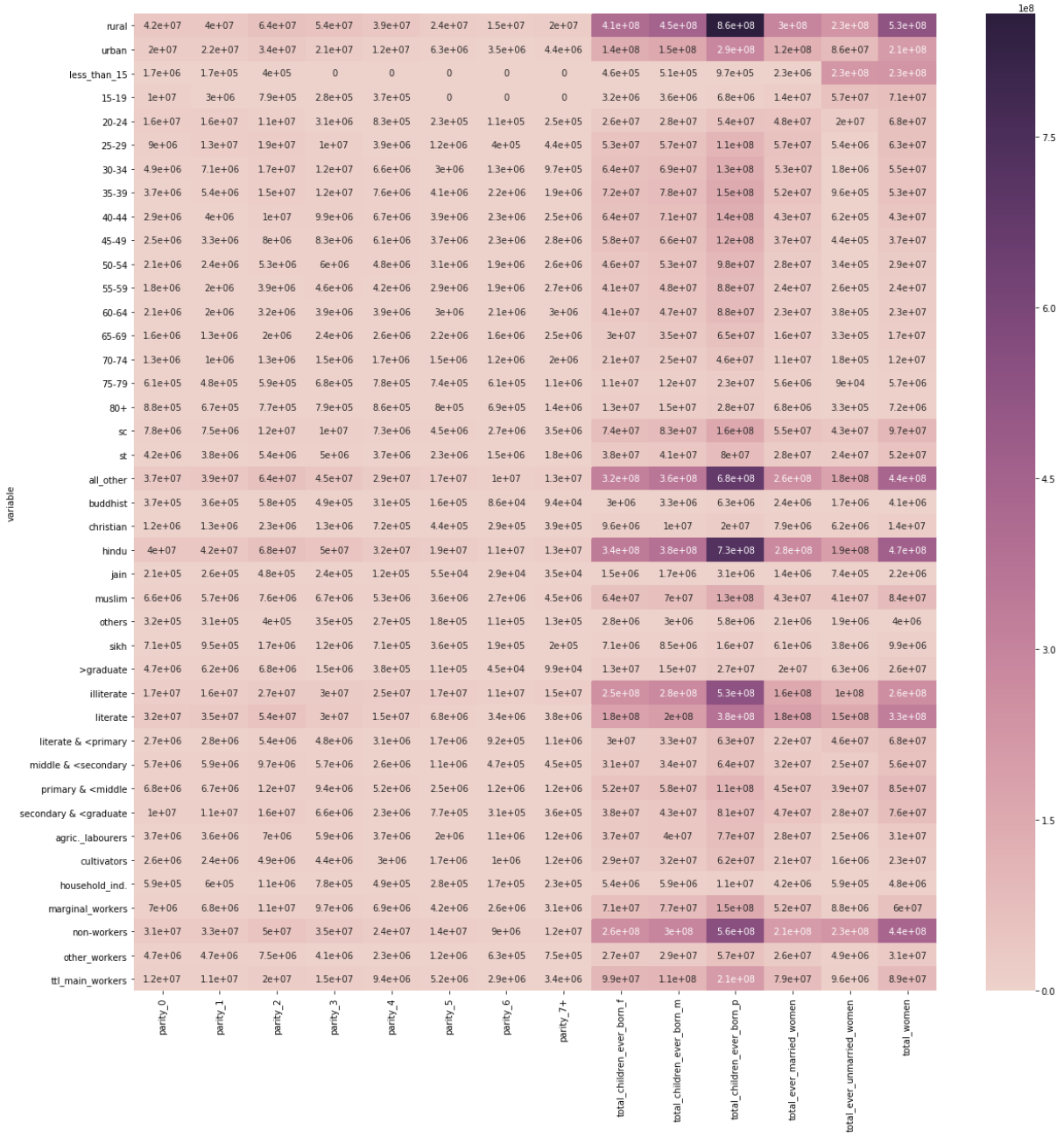

Figure 4.36: Heat Map for Parity, Births Till Now and Ever Married Women

Figure 4.36, shows the correlation between:

- Total women with rural, Hindu community, illiterate, literate, women other than SC and ST and non-workers.
- Currently unmarried women with non-workers, illiterate, literate, women other than SC and ST, Hindu community, less than 15 and rural

- Children ever born with the Hindu community, Muslim community, illiterate, literate, women other than SC and ST, non-workers, total main workers, urban area, and rural areas.
- Ever married women with non-workers, literate, women other than SC and ST, Hindu community, urban and rural areas.

4.3.2 Comparison of Infant Mortality and Fertility of 2001 with 2011

Comparing the values of 2001 and 2011 will help us understand how things have changed over the years. This will also help to understand if the factors that are looked into 2001 have helped and made changes by 2011 or not. The condition right now can be improved by finding what factors have led to a decrease in mortality rate in this time interval.

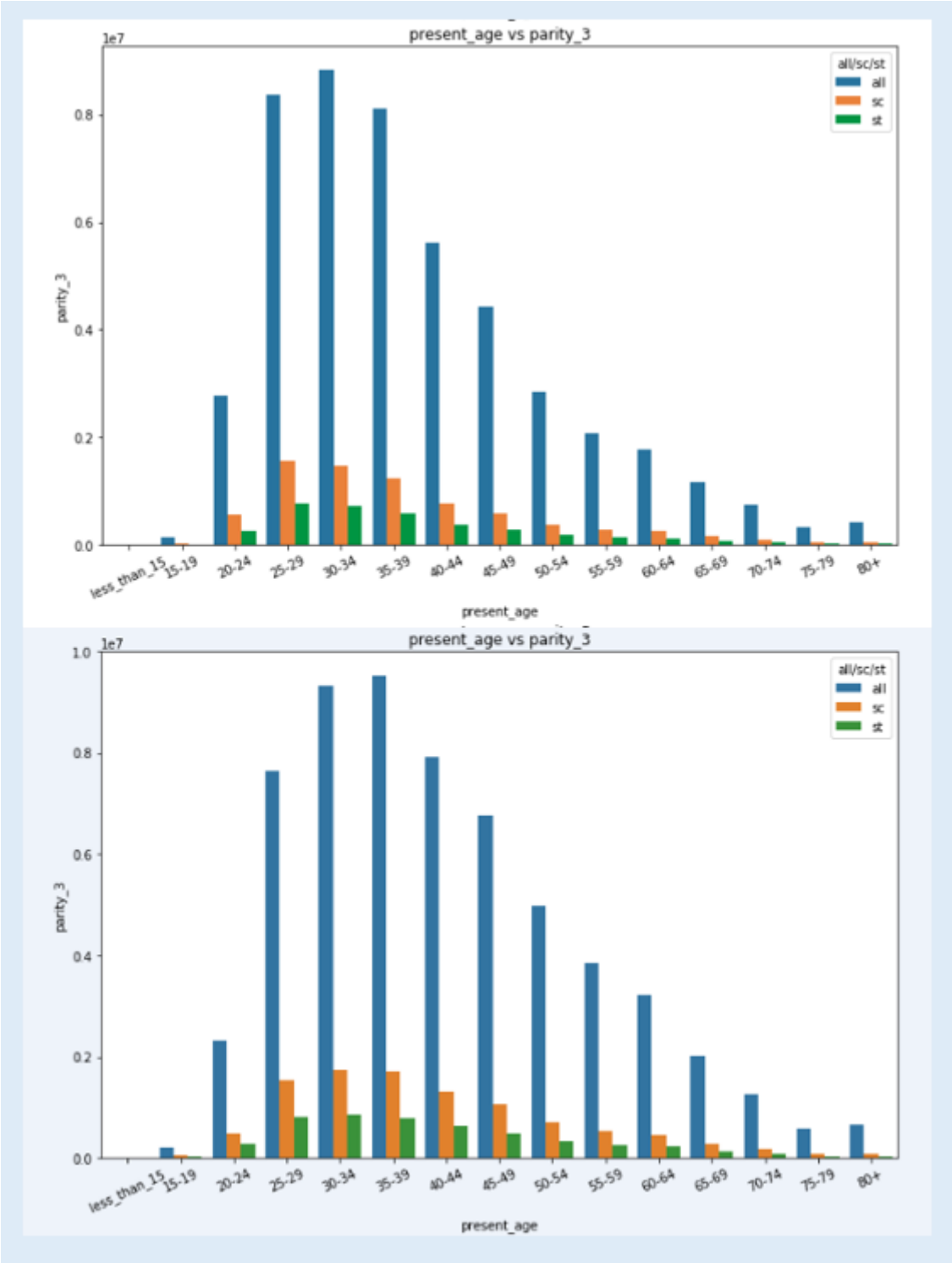

Figure 4.37: Present Age vs Parity 3(all/SC/ST) in 2001 and 2011

Figure 4.37 shows the number of women who had three children corresponding to their present age in 2001 and 2011 respectively. It is visible that in India the parity 3 is highest among age groups 30-34 but in 2011 the peak was at 35-39. This shows that the number of women giving birth to 3 children has reduced among the younger population which is usually

when most women in India have children. This is a good sign because even when parity 0, parity 1, and parity 2 show a similar graph, there are fewer women in the newer generation who are having 3 children. This reduction in fertility rate will in turn reduce the population growth and chances of infant mortality in India.

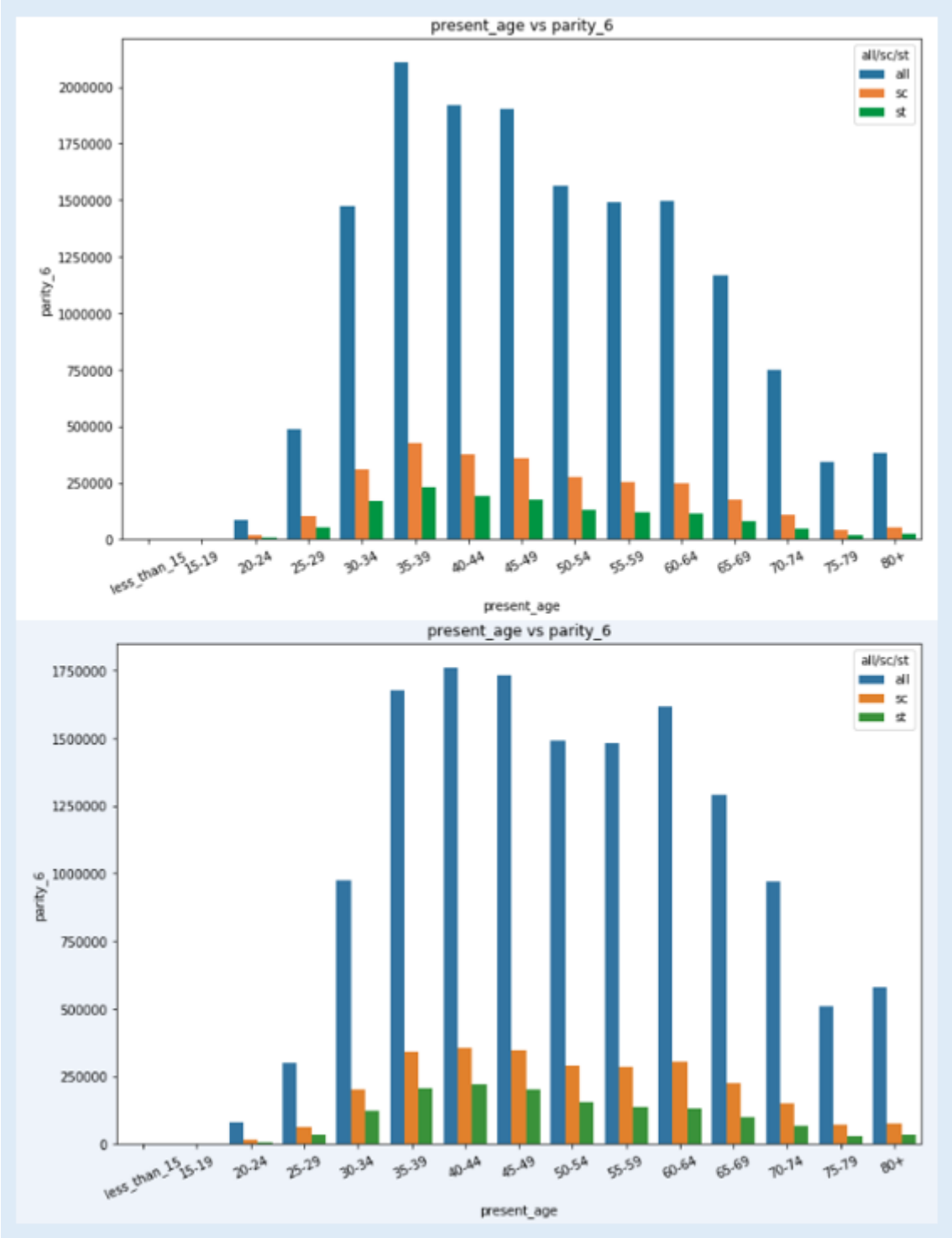

Figure 4.38: Present Age vs Parity 6(all/SC/ST) in 2001 and 2011

Figure 4.38 shows the number of women who had six children corresponding to their present age in 2001 and 2011 respectively. It is visible that the peak value is 2.5 lakh more in 2001 than in 2011. This shows that there is a decrease in the number of women having 6 children. In all of India it is visible that among the age group 35- 39, there is a major decrease in the vale of women. Given that this is a census 10 years apart, the increase in the 45-49 age group in 2011 is because of the peak in the age group 35-39 10 years before. The age group 60-64 is an outlier in 2011.

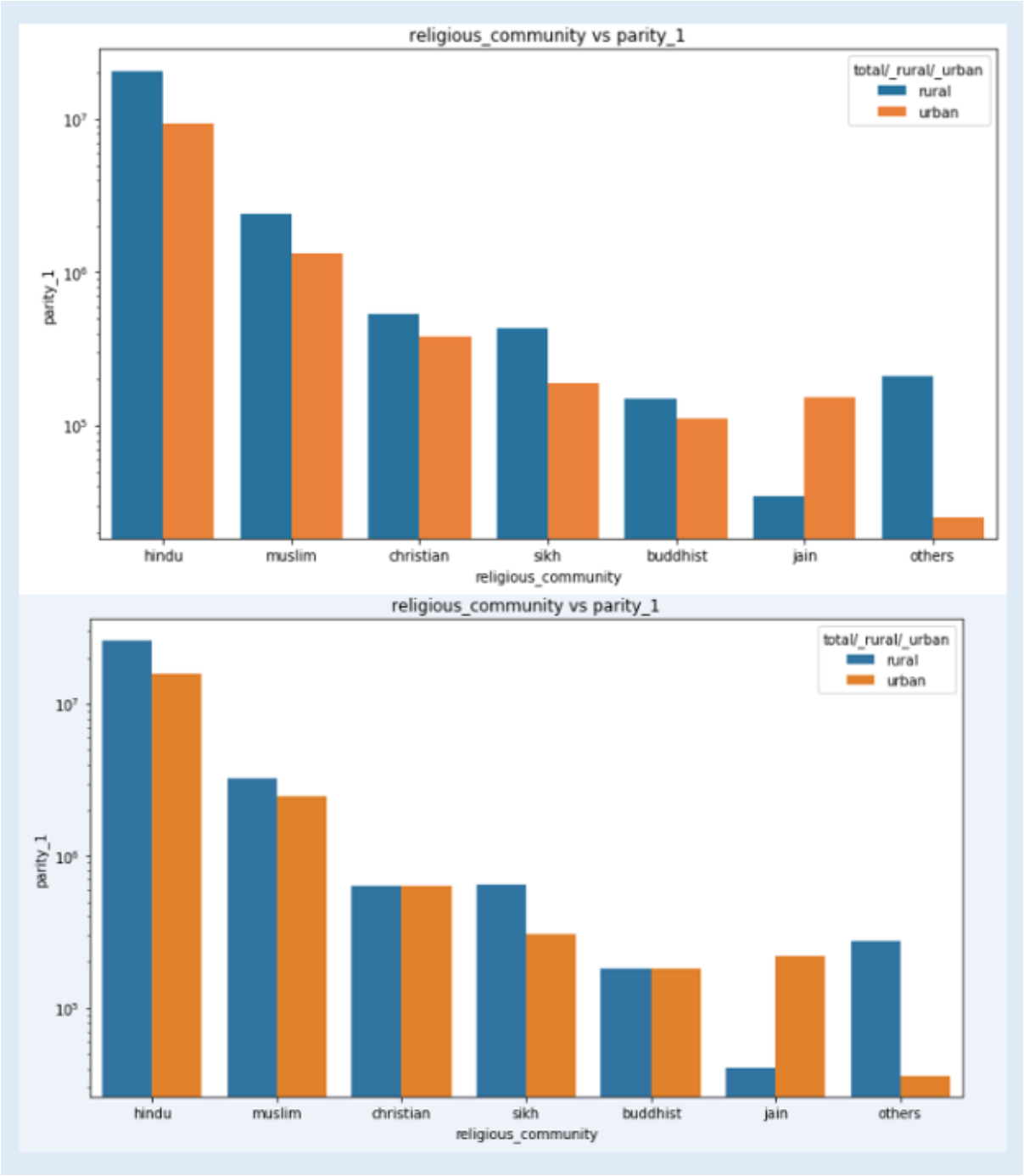

Figure 4.39: Religious Community vs Parity 1(rural/urban) in 2001 and 2011

Figure 4.39 shows the number of women who had 1 child corresponding to their religion (log scale) in 2001 and 2011 respectively. It is visible that in2001 the values of the total women who are Christians and Buddhists have been different in rural and urban, but the same is almost equal in 2011. The rest of the religious communities have followed the same pattern over the years. The same pattern is also found among women having 2 children. This shows

that the major changes have happened among these two religious communities since all the other parity levels are almost the same in both the years.

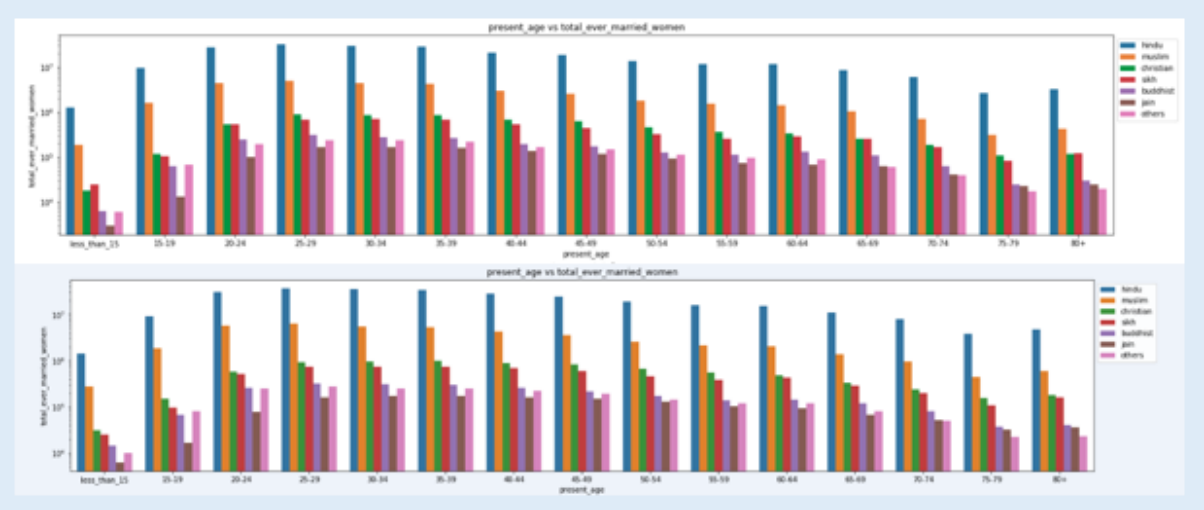

Figure 4.40: Present Age vs Number of ever married women (religious community) in 2001 and 2011

Figure 4.40 shows the number of married women corresponding to their religion (log scale) in 2001 and 2011 respectively. It is visible that there are fewer women among the age group less than 15 getting married in the Sikh community in 2011 as compared to 2001 even when the total number of women stays constant. This is a good sign as women getting married before the age of 19 have more chances of producing premature babies, which in turn leads to high rates of infant mortality.

The same reduction is found in parity 2 among the Sikh community in the age group of less than 15 and, in parity 3 and parity 4 in the age group of 15-19. These results show that the reduction in marriage among women who are less than 15 has led to producing lesser children at a very young age. There is also a major reduction in the total surviving children in this community when the present age is below 19. This clearly shows that there are lesser infant deaths since lesser women below 19 are married.

So the reduction, even though gradual, is a good sign. Sadly the rates have stayed constant among the other communities.

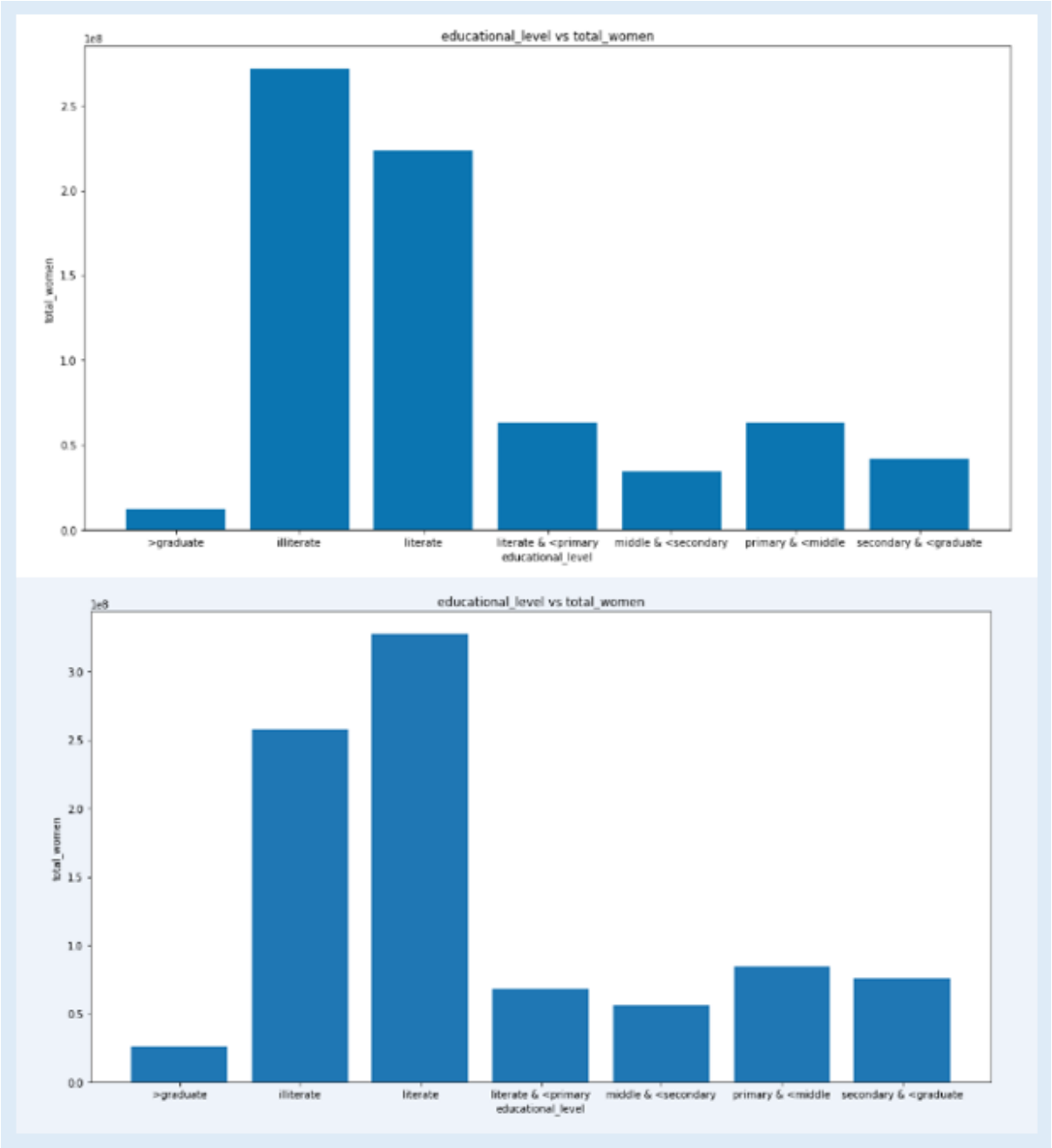

Figure 4.41: Educational level vs total women in 2001 and 2011

Figure 4.41 shows total women corresponding to their educational level in 2001 and 2011 respectively. Illiteracy among women is highest in 2001. But the number of literate women has increased in those 10 years and has exceeded the number of illiterate women. This shows that most females have been educated since.

There is also a gradual increase in the number of females who have graduated from school. This increase in literacy has affected the fertility rate.

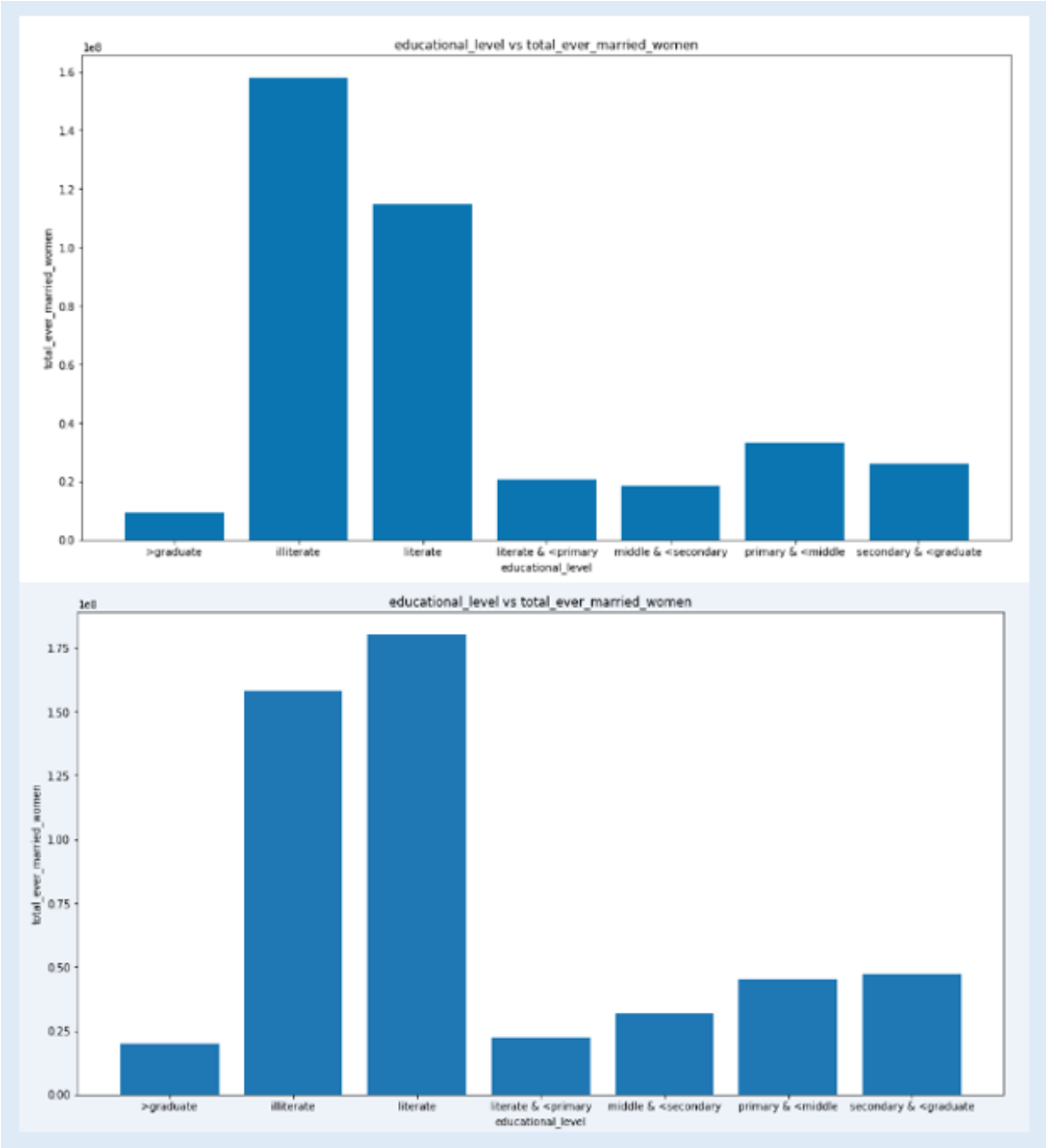

Figure 4.42: Educational level vs total ever married women in 2001 and 2011

Figure 4.42 shows total married women corresponding to their educational level in 2001 and 2011 respectively. Since more women being educated in 2011 than in 2001 has resulted in a major decrease in the number of women who are married. This in turn reduces the fertility rate in India.

Figure 4.43: Present age vs total ever married women in 2001 and 2011

Figure 4.43 shows total married women corresponding to their educational level in 2001 and 2011 respectively. Literacy has increased among the lower age groups. Especially in the age groups 30-34 and 35-39 has shown a major improvement by increasing above the illiteracy rate.

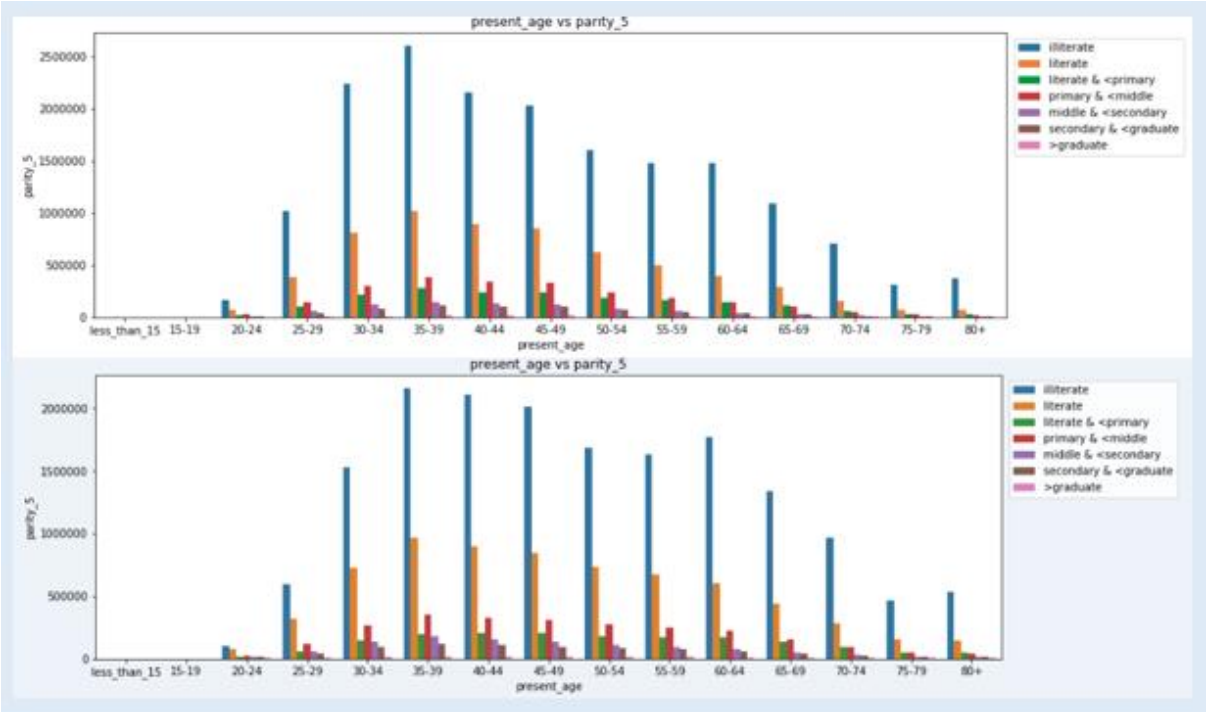

Figure 4.44: Present age vs parity 5 in 2001 and 2011

It is the level of illiteracy that has reduced among the women with parity 0, parity 1, parity 2, and parity 3. But it is not visible among the higher age groups. This shows that the newer generations have more literate women. This means that most women are staying unmarried until they graduate from school. But among women in parity 4, parity 5, parity 6 and parity 7+ the literacy rate has not changed among all ages.

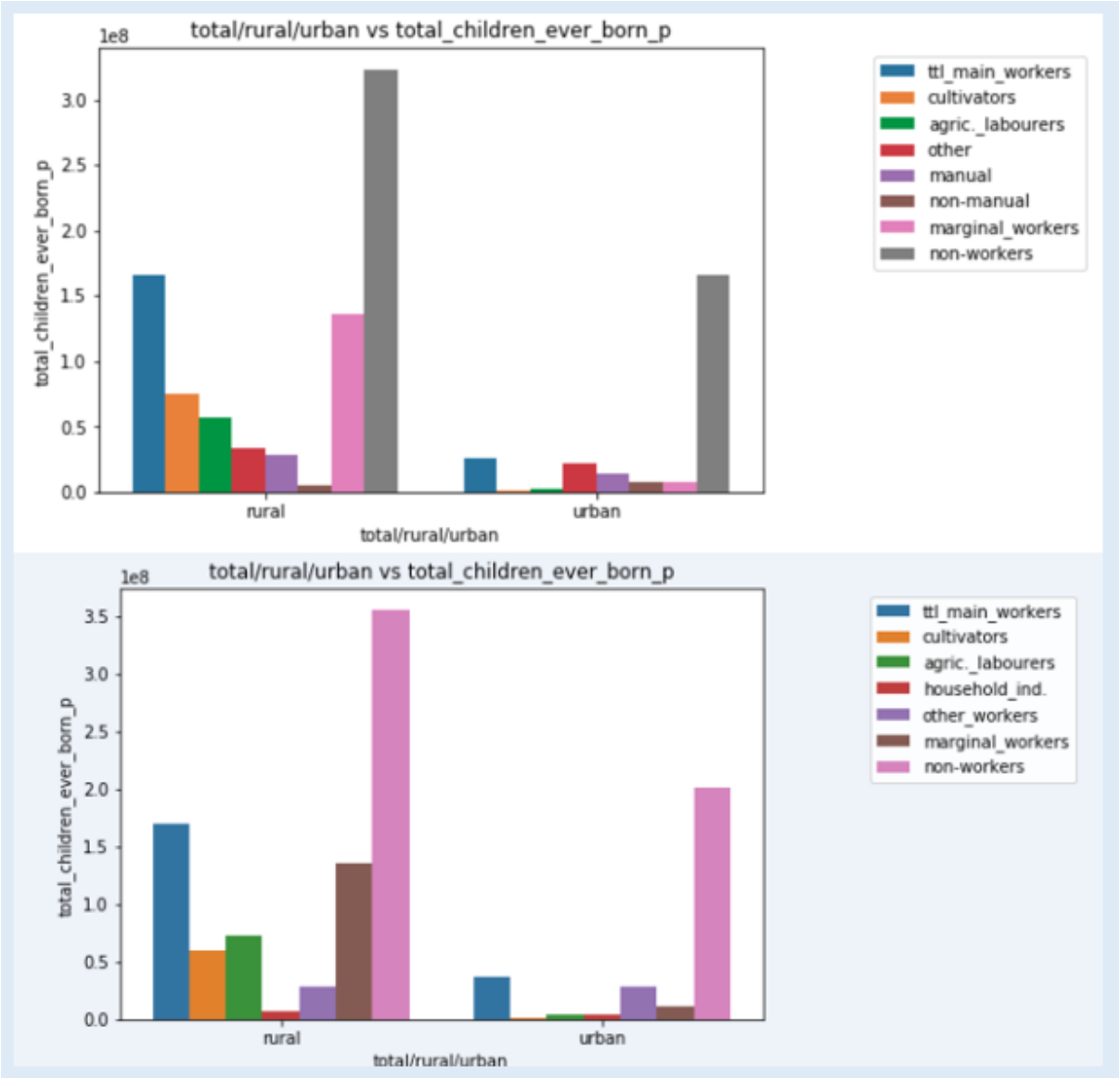

Figure 4.45: Total/rural vs total children ever born in 2001 and 2011

Figure 4.45 shows the occupation of the mothers of children born to date corresponding to the location (rural or urban) in 2001 and 2011 respectively. It is visible that there is an evident reduction in the number of cultivators in rural areas. This is the same in each parity of women. This shows that most women have moved to another profession rather than cultivating. But the level of cultivators remains constant in the urban areas.

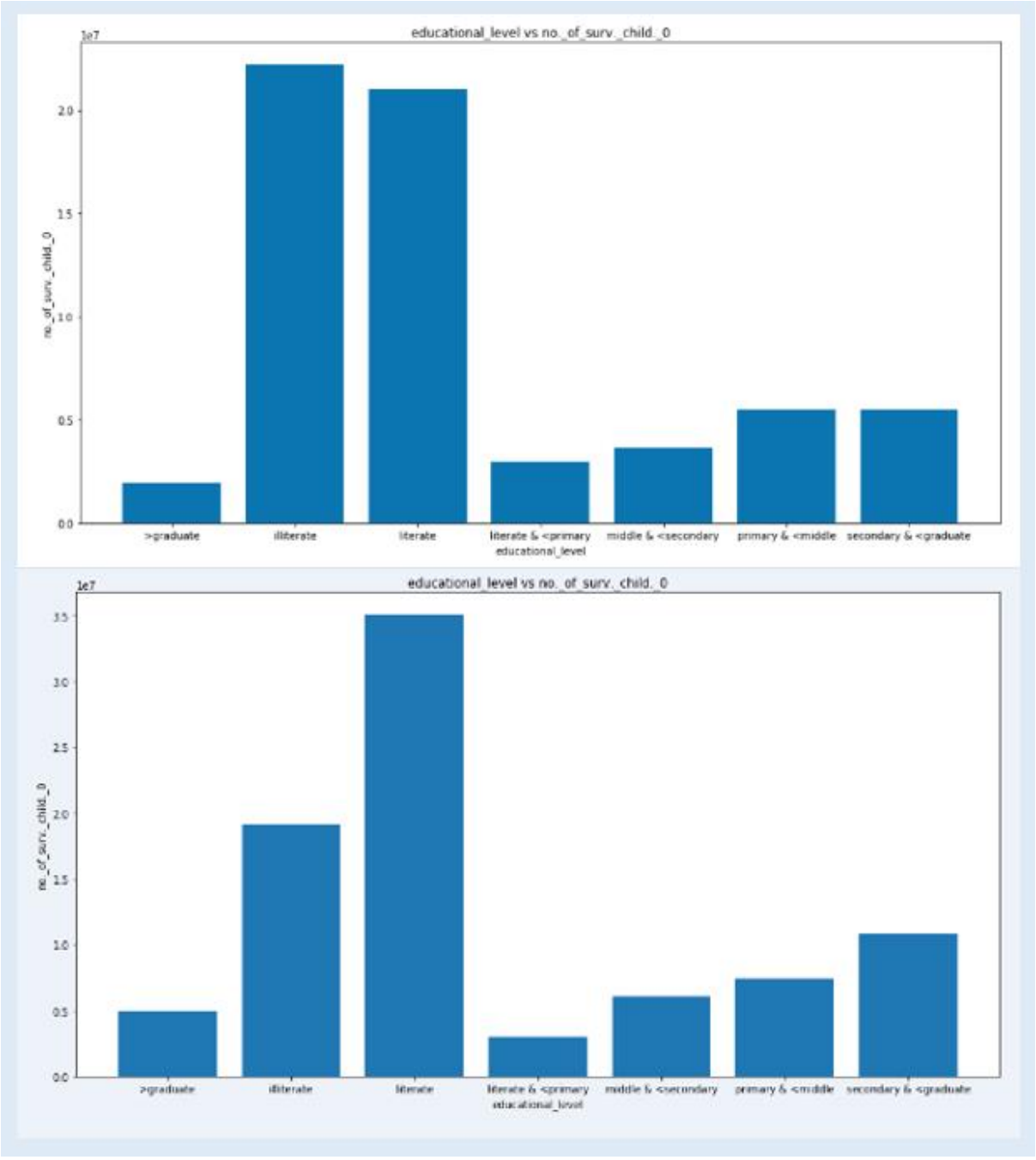

Figure 4.46: Educational level vs number of surviving children 0 in 2001 and 2011

Figure 4.46 shows the total number of women with 0 surviving children corresponding to educational level in 2001 and 2011 respectively. The graph in 2001 shows that the chances of surviving 0 children were highest among illiterate women. But by 2011, the survival of infants among the literates have a very higher chance. It is also seen that literate females have a high value when it comes to parity 0 and parity 1. This means that there are fewer children born to literate women in 2011, thus they have a higher chance of having zero survival.

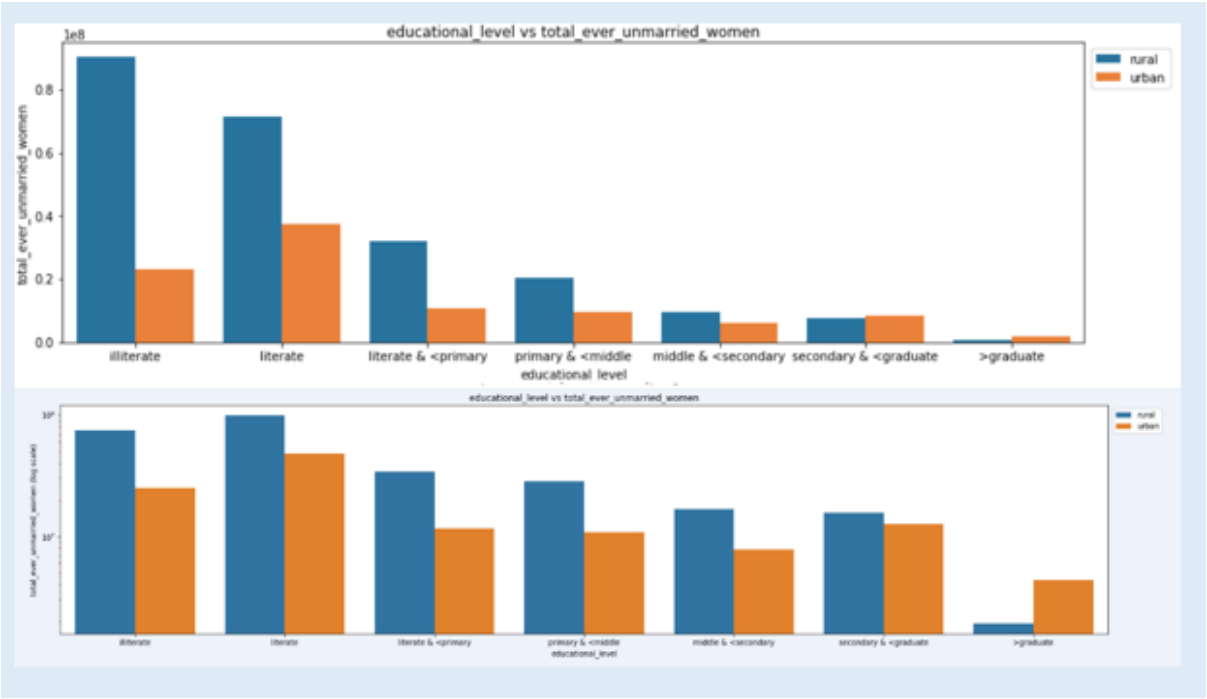

Figure 4.47: Educational level vs number ever unmarried women in 2001 and 2011

Figure 4.47 shows total unmarried women corresponding to educational level in 2001 and 2011 respectively. The graph in 2001 shows that there is an increase in literate women both in urban and rural areas. There is an increase in the number of unmarried women who have attended secondary school in rural areas whereas in urban areas the raises visible among the graduates.

4.3.3 Analysis on Infant Mortality and Fertility among States (2011)

The values of mortality in each state have been very different. This means that just learning about India as a whole will not suffice. So in this section, each state is compared to different variables. This will help figure out which states play a major role in the increase in the mortality rate of India as a whole.

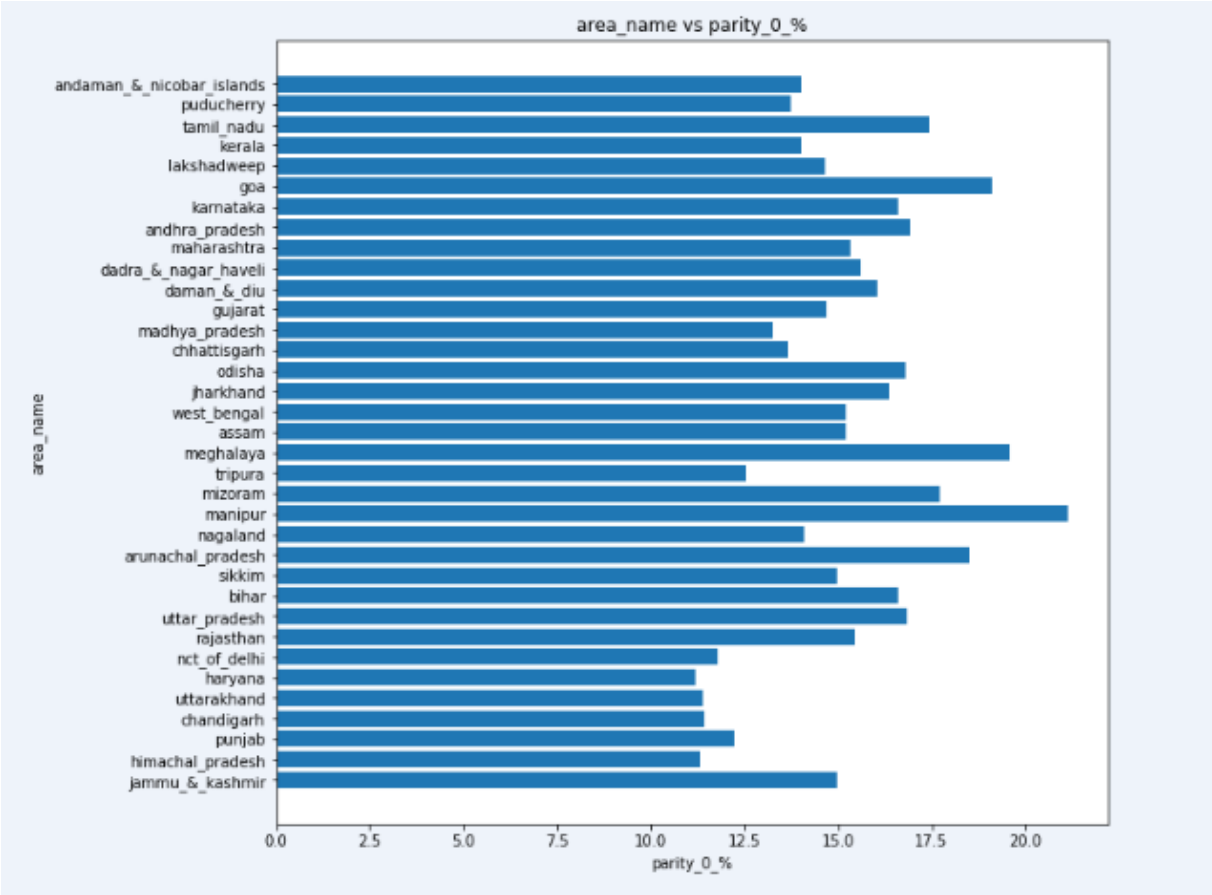

Figure 4.48: States vs Parity 0

In figure 4.48, we see the percentage of women who had no children against the states they live in. The number of women with no children in Manipur is high. It is followed by Meghalaya, Goa, Arunachal Pradesh, Mizoram, and so on. The lowest is in Haryana.

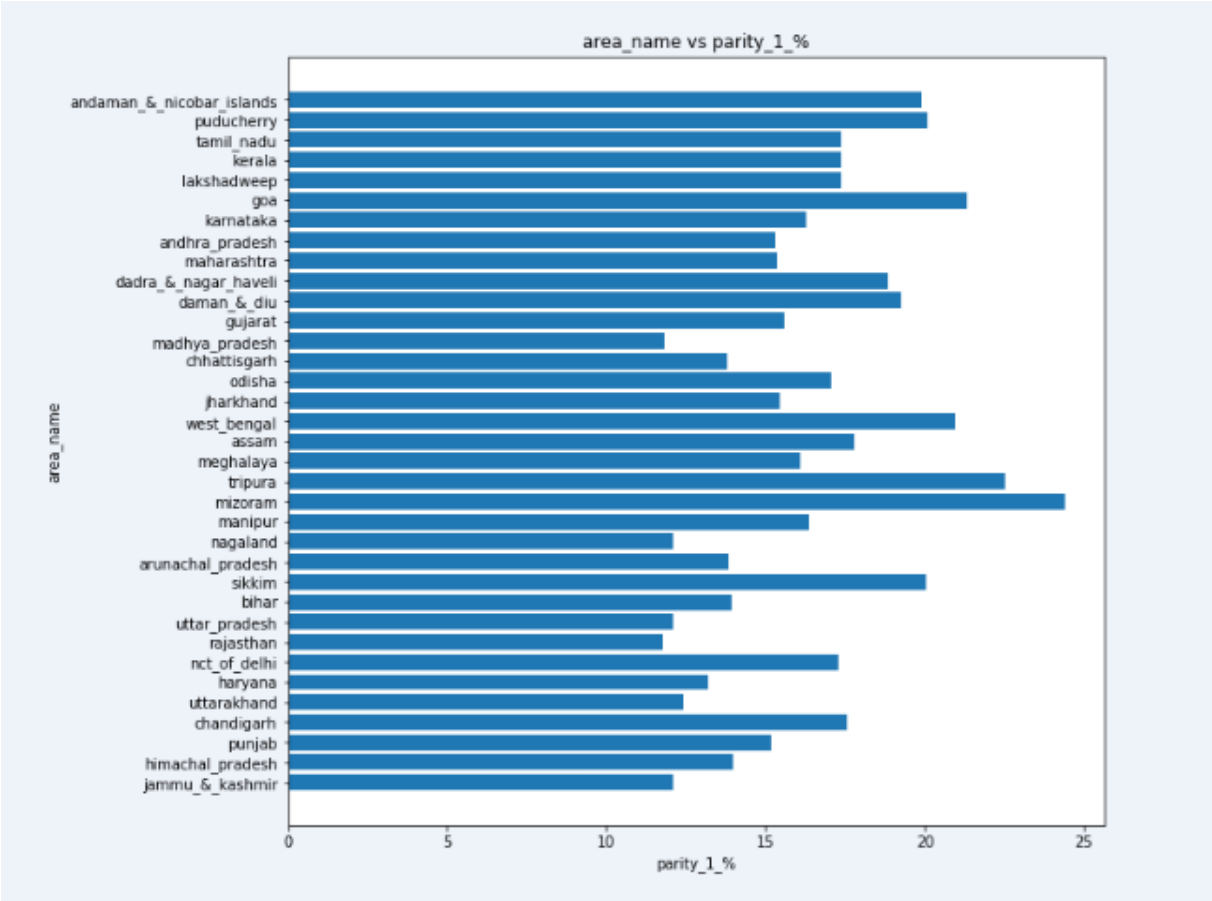

Figure 4.49: States vs Parity 1

In figure 4.49, we see the percentage of women who had one child against the states they live in. The number of women with one child in Mizoram is high. It is followed by Tripura, Goa, West Bengal, Sikkim, and so on. The lowest is in Madhya Pradesh.

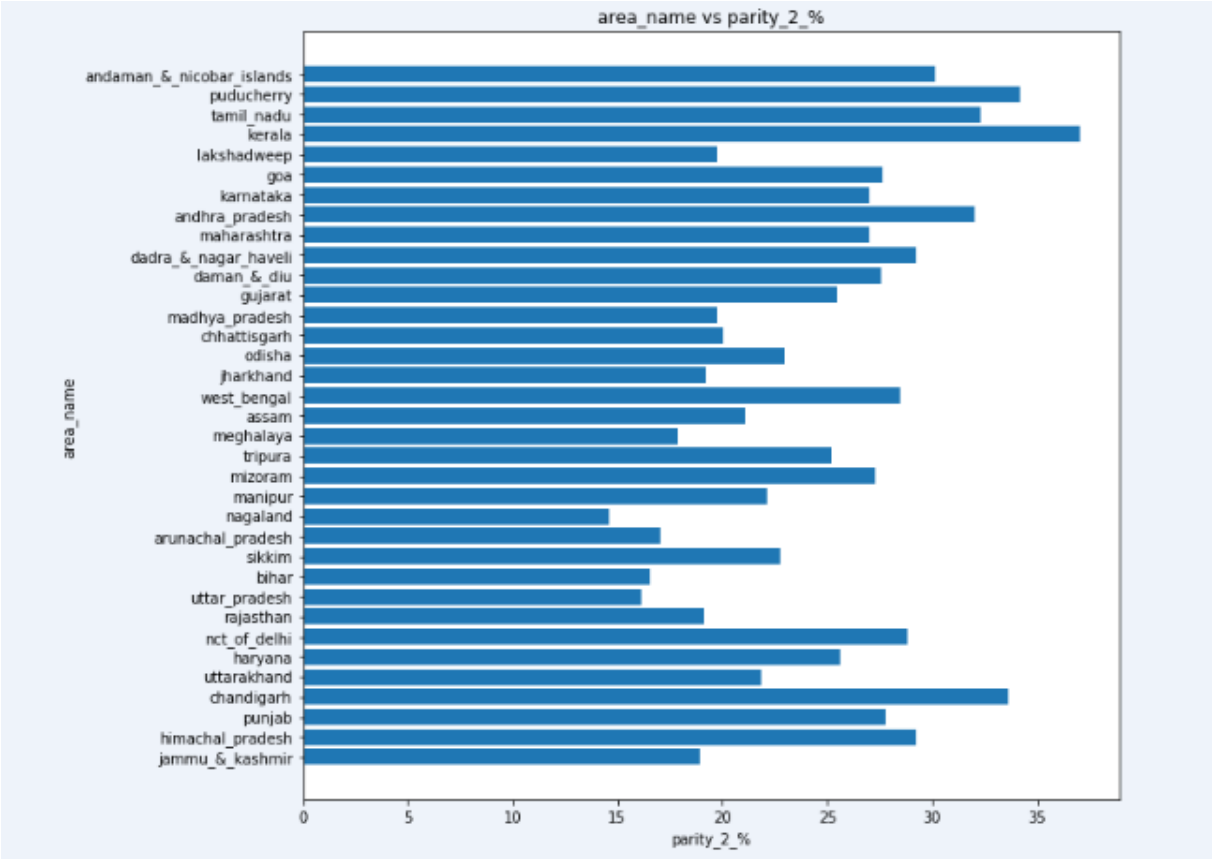

Figure 4.50: States vs Parity 2

In figure 4.50, we see the percentage of women who had 2 children against the states they live in. the number of women with two children in Kerala is high. It is followed by Puducherry, Chandigarh, Tamil Nadu, Andhra Pradesh, and so on. The lowest is in Nagaland.

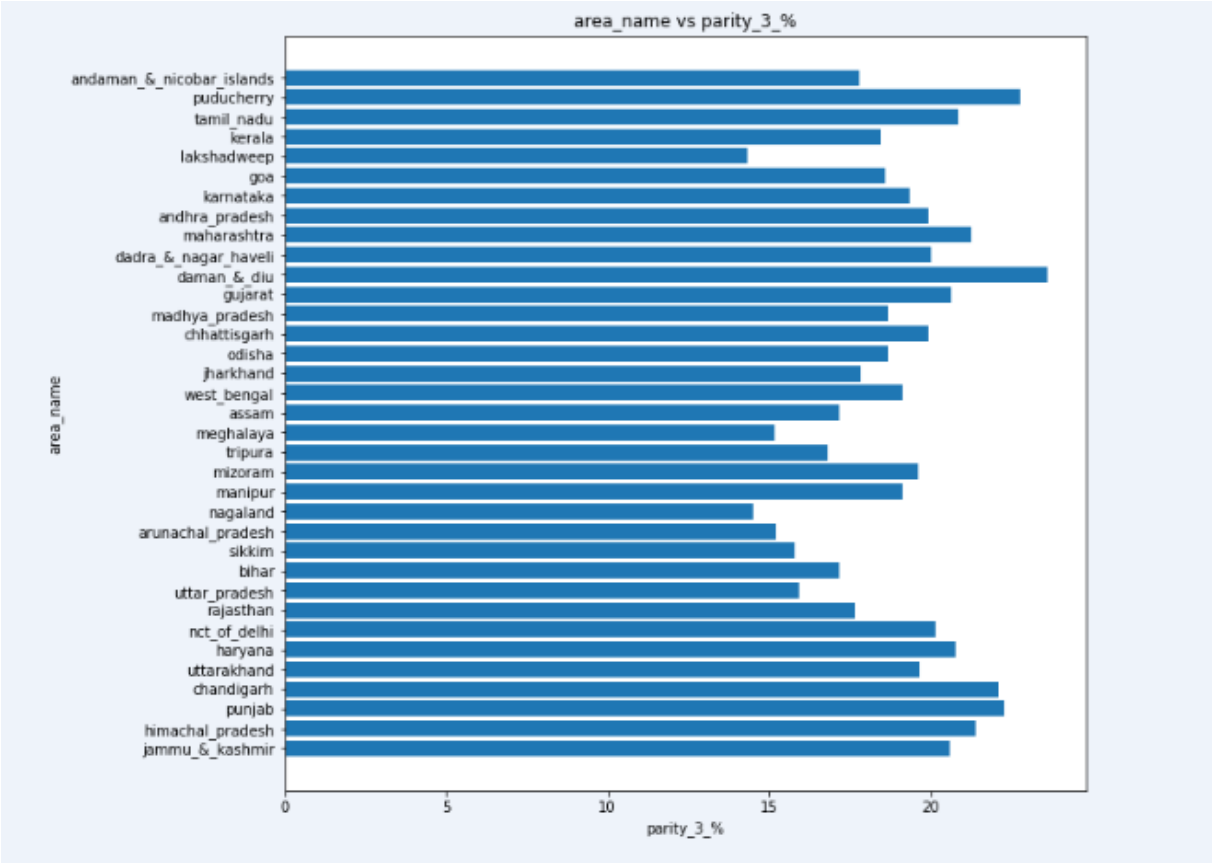

Figure 4.51: States vs Parity 3

In figure 4.51, we see the percentage of women who had 3 children against the states they live in. the number of women with three children in Daman and Diu is high. It is followed by Gujarat, Puducherry, Punjab, Chandigarh so on. The lowest is in Lakshwadeep.

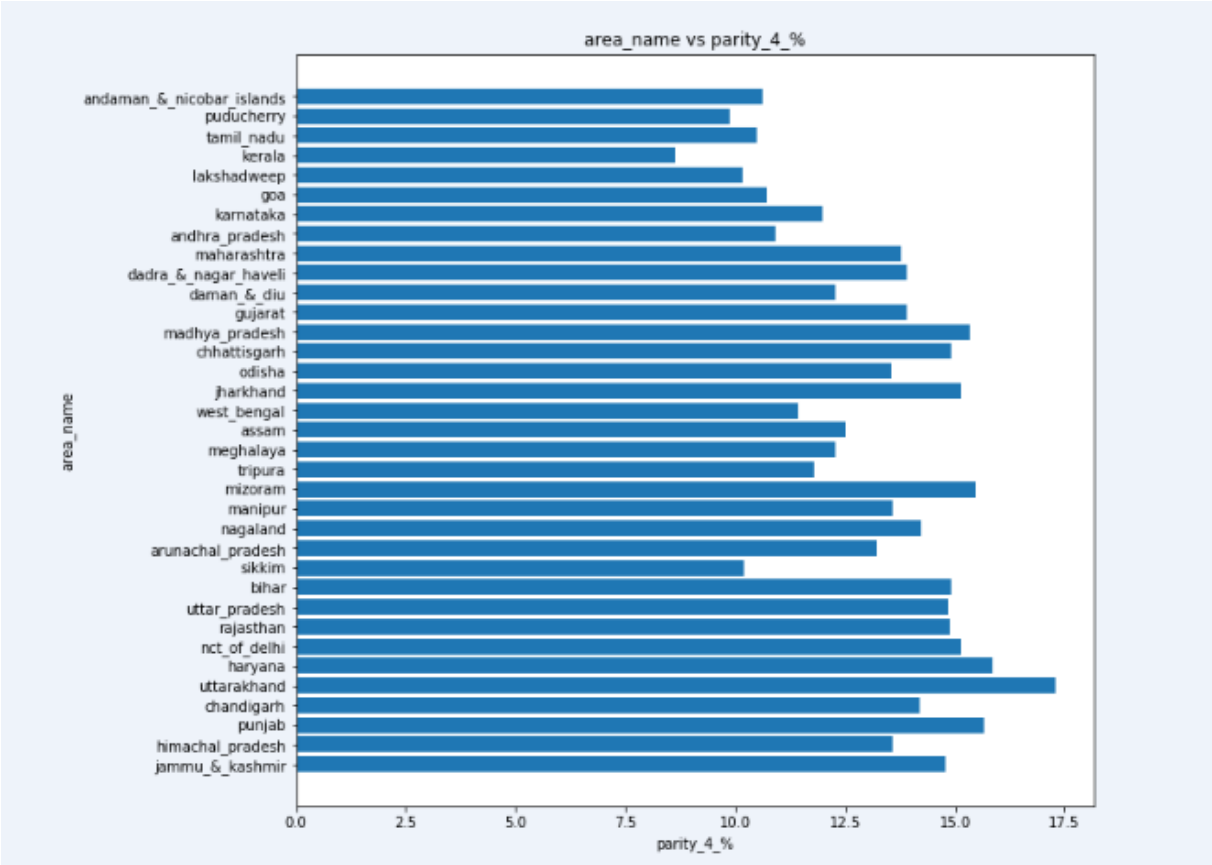

Figure 4.52: States vs Parity 4

In figure 4.52, we see the percentage of women who had 4 children against the states they live in. the number of women with four children in Uttarakhand is high. It is followed by Haryana, Punjab, Madhya Pradesh, Chattisgarh so on. The lowest is in Kerala.

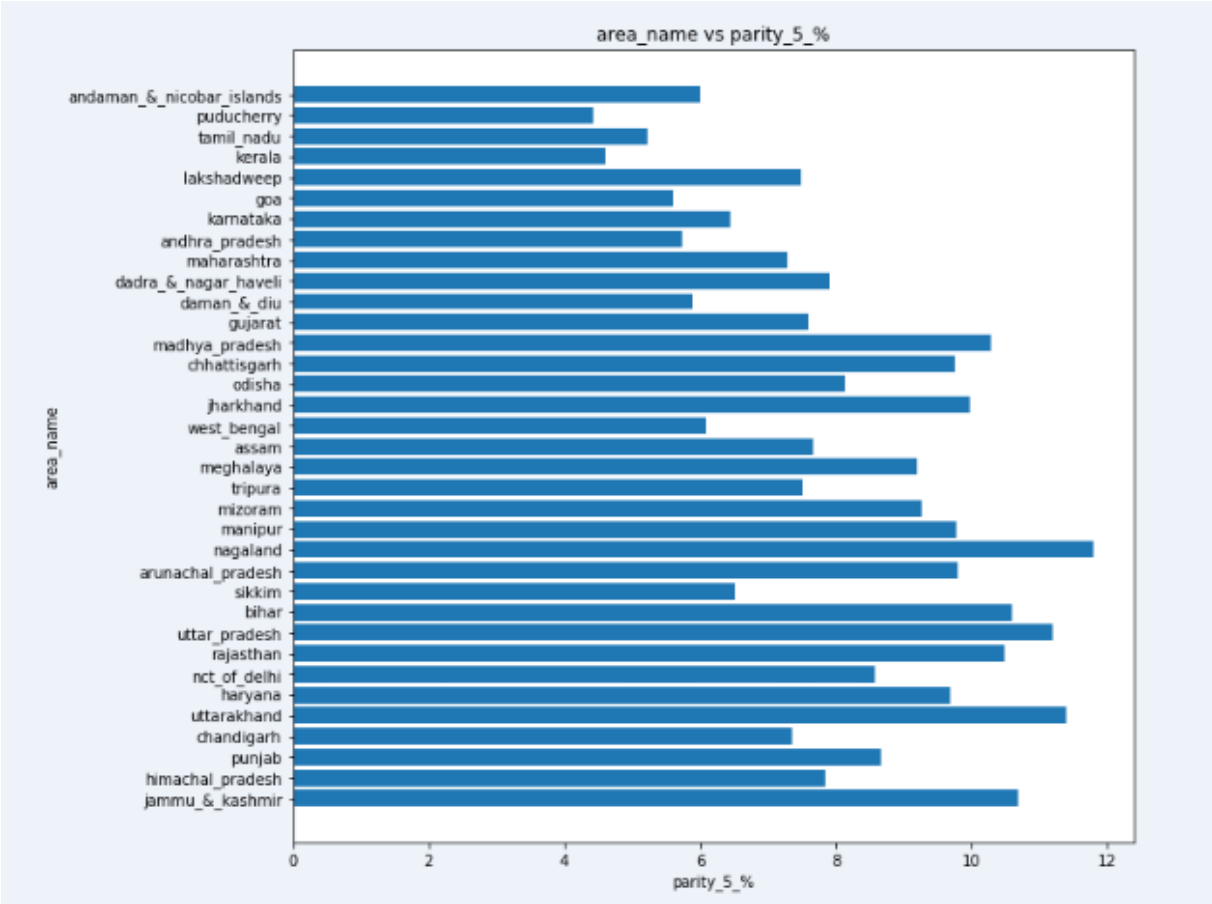

Figure 4.53: States vs Parity 5

In figure 4.53, we see the percentage of women who had 5 children against the states they live in. the number of women with five children in Nagaland is high. It is followed by Bihar, Uttarakhand, Uttar Pradesh, Jammu, and Kashmir so on. The lowest is in Puducherry.

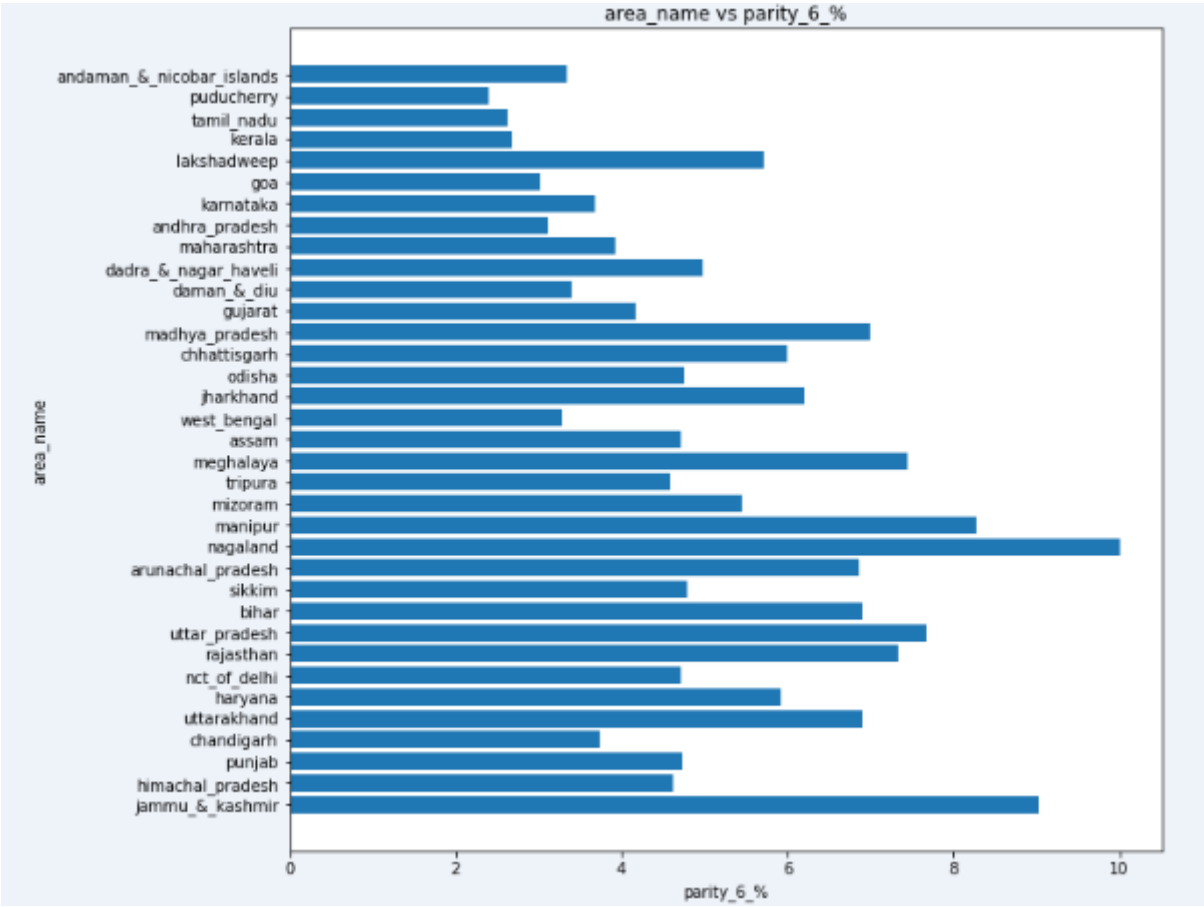

Figure 4.54: States vs Parity 6

In figure 4.54, we see the percentage of women who had 6 children against the states they live in. the number of women with six children in Nagaland is high. It is followed by Manipur, Meghalaya, Uttar Pradesh, Jammu, and Kashmir so on. The lowest is in Puducherry.

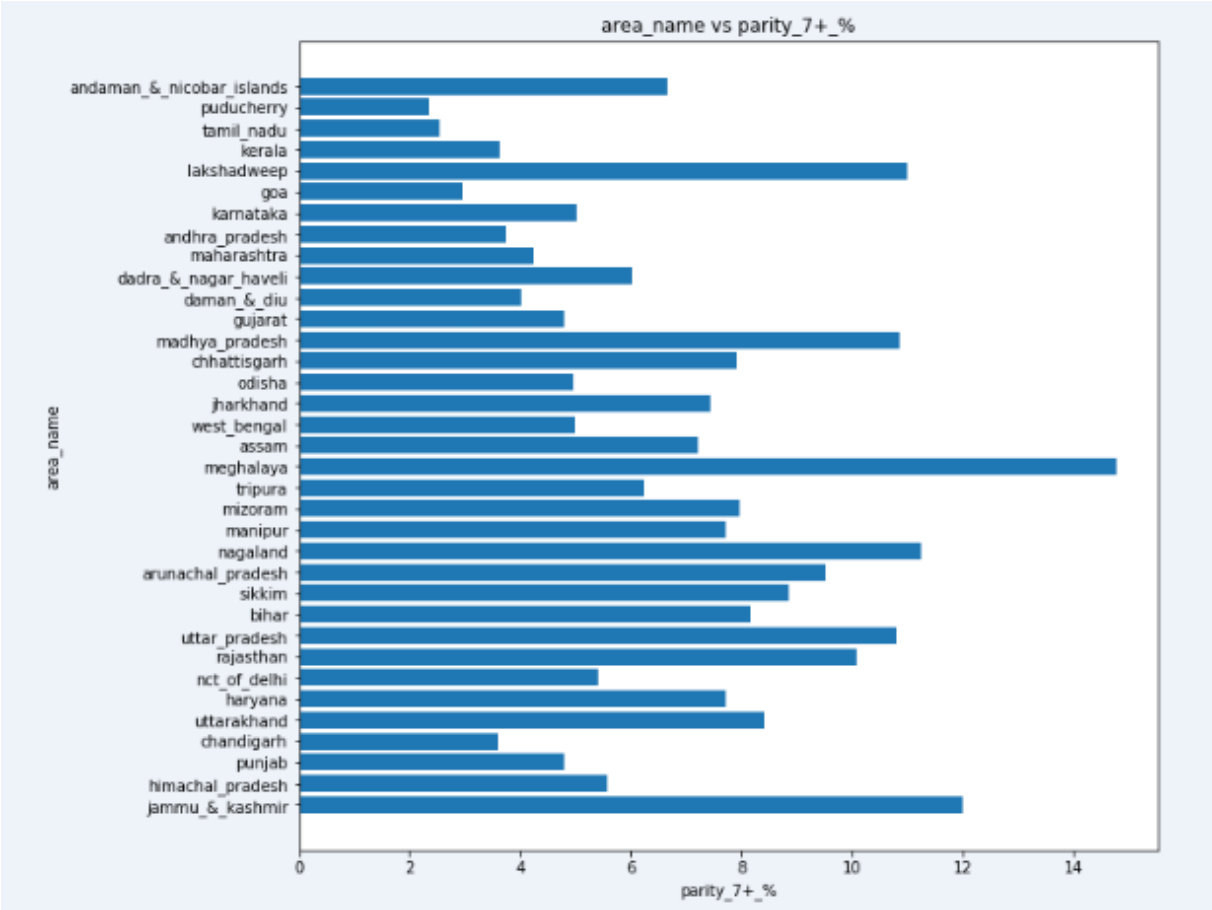

Figure 4.55: States vs Parity 7+

In figure 4.55, we see the percentage of women who had 7+ children against the states they live in. the number of women with seven and more children is the highest in Uttar Pradesh. It is followed by Bihar, West Bengal, Madhya Pradesh, Rajasthan, and so on.

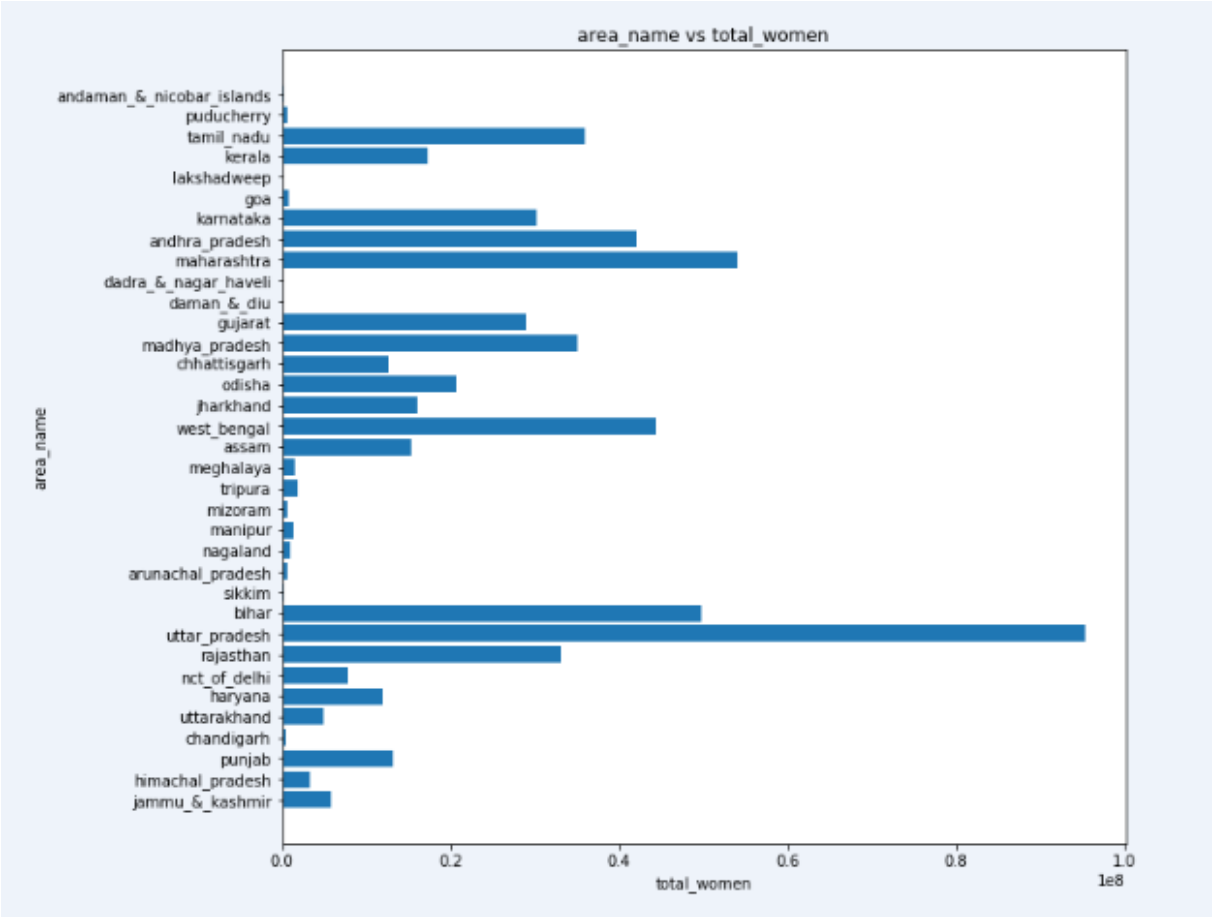

Figure 4.56: States vs Total Women

In figure 4.56, we see the percentage of women against the states they live in. the number of women in India in Uttar Pradesh is high. It is followed by Bihar, Maharashtra, West Bengal, Andhra Pradesh and so on.

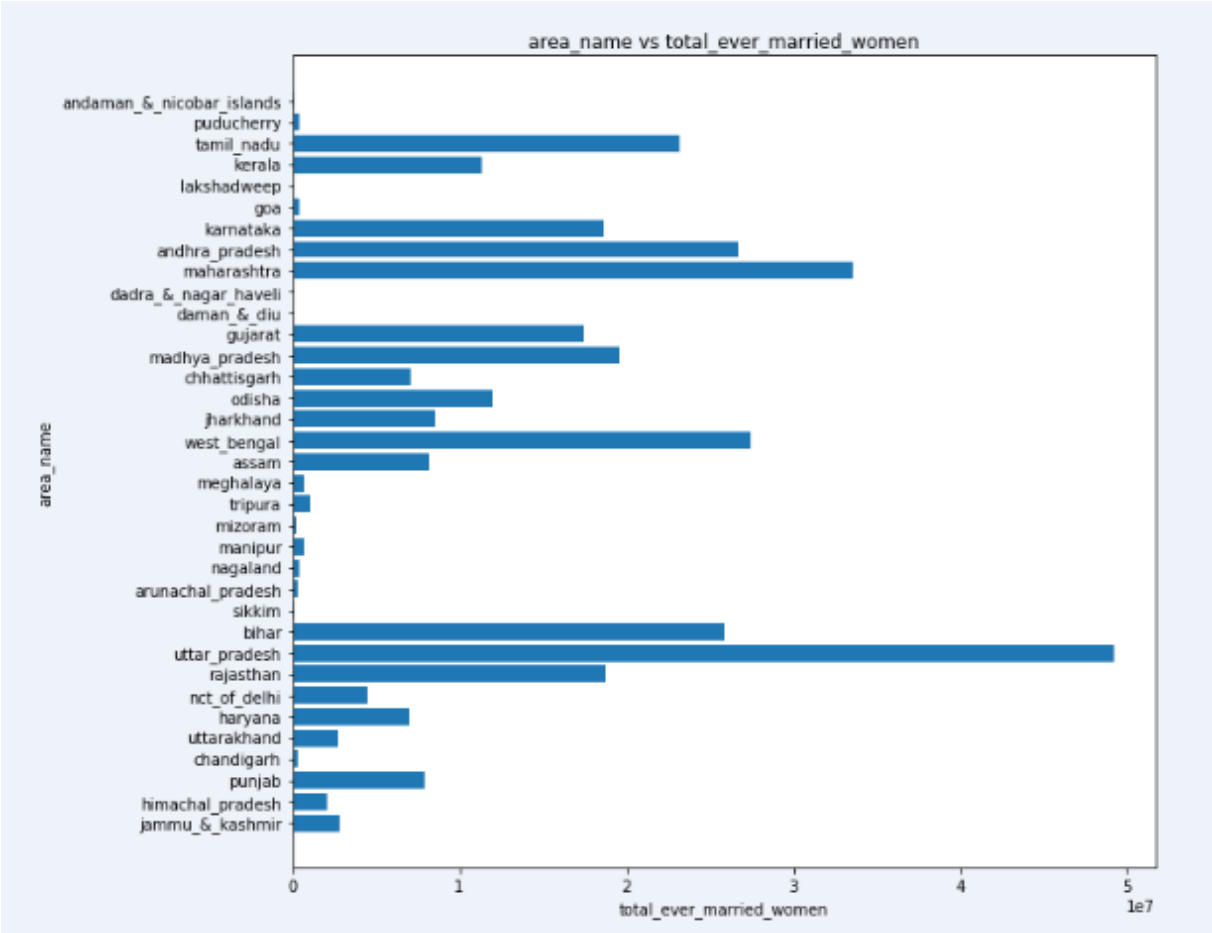

Figure 4.57: States vs Total ever married Women

In figure 4.57, we see the percentage of married women against the states they live in. the number of married women in India is highest in Uttar Pradesh. It is followed by Bihar, Maharashtra, West Bengal, Andhra Pradesh and so on.

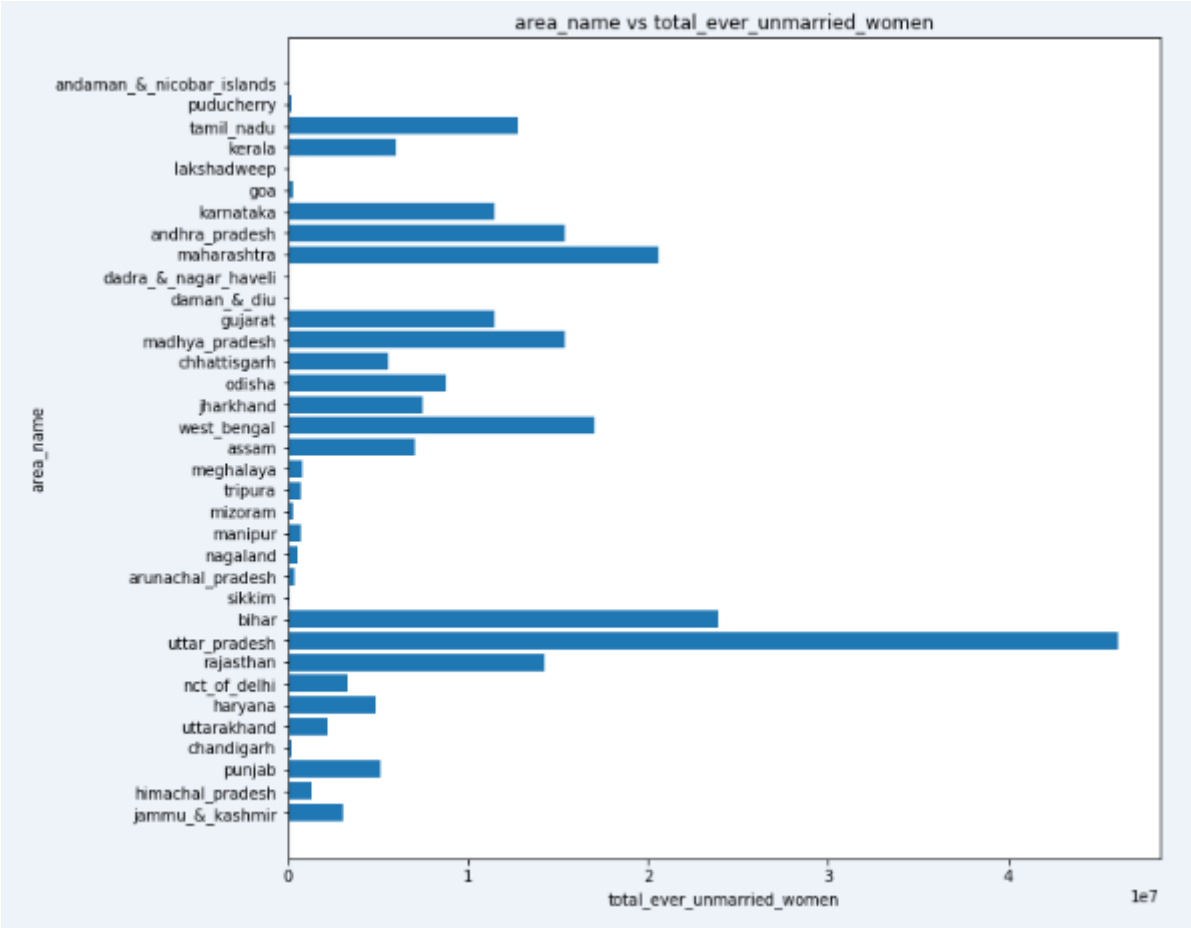

Figure 4.58: States vs Total unmarried Women

In figure 4.58, we see the percentage of unmarried women against the states they live in. the number of unmarried women is highest in Uttar Pradesh. It is followed by Bihar, Maharashtra, West Bengal, Andhra Pradesh and so on.

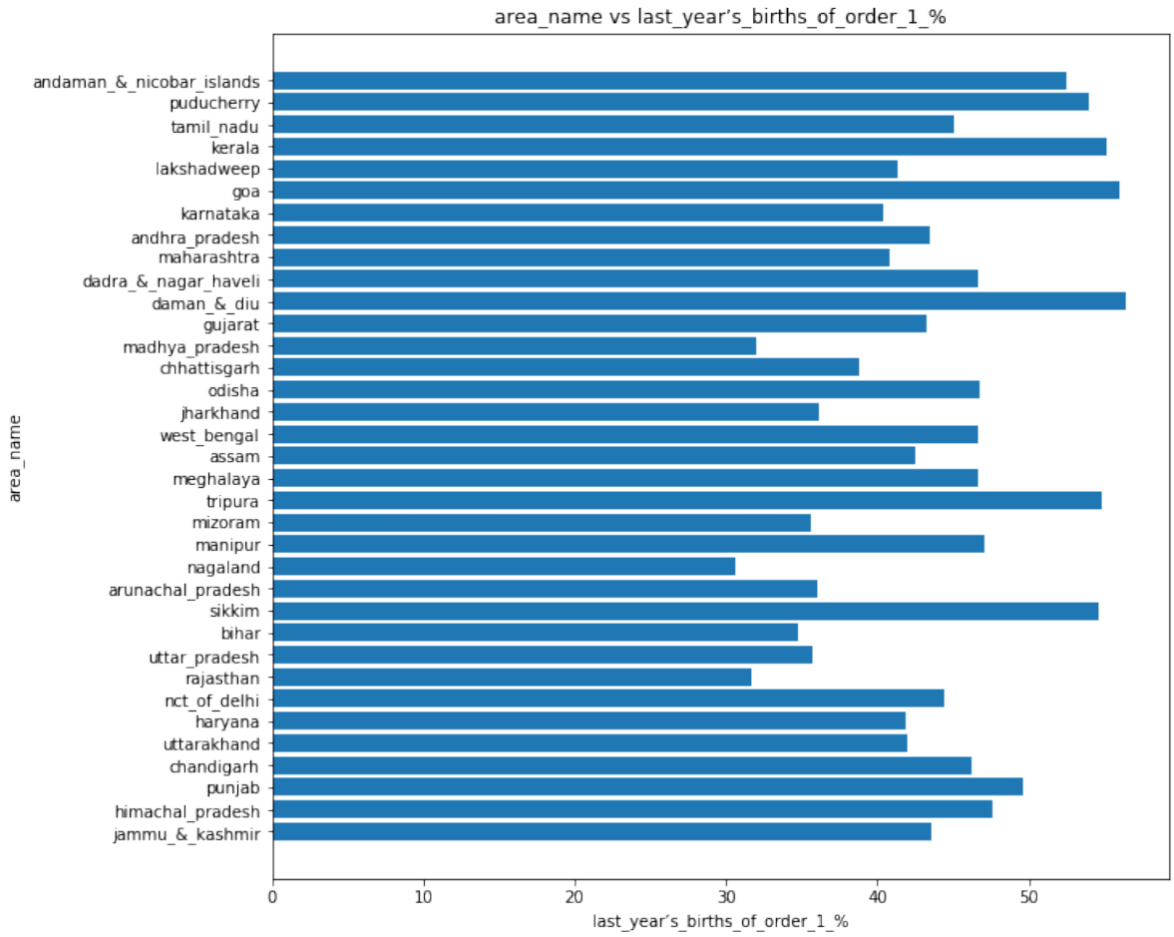

Figure 4.59: States vs Birth order last year 1

In figure 4.59, we see the percentage of women with birth order 1 against the states they live in. The highest is in Daman and Diu. It is followed by Goa, Tripura, Sikkim, Kerala, and so on. The lowest is in Nagaland.

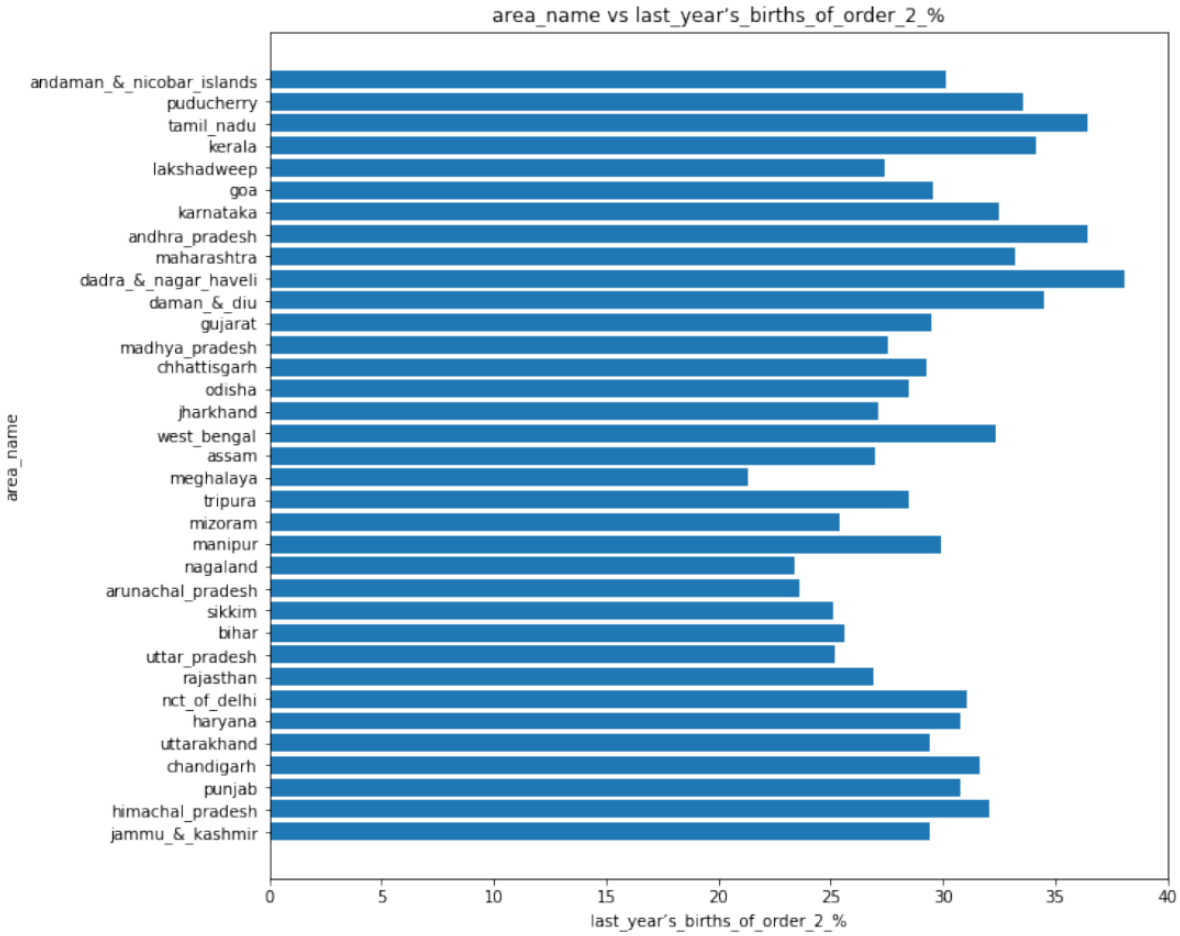

Figure 4.60: States vs Birth order last year 2

In figure 4.60, we see the percentage of women with birth order 2 against the states they live in. highest is in Dadra and Nagar haveli. It is followed by Andhra Pradesh, Tamil Nadu, Daman and Diu, Kerala and so on. The lowest is in Meghalaya.

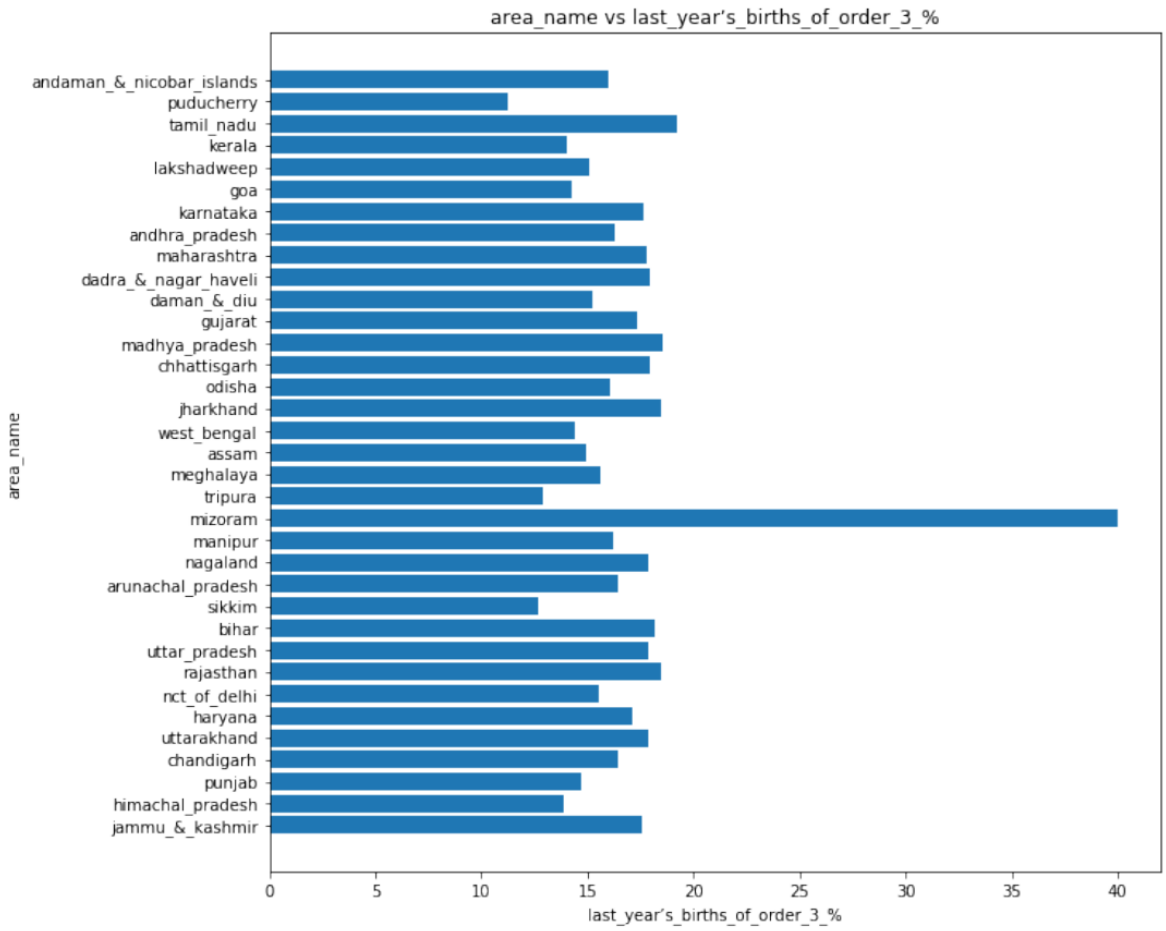

Figure 4.61: States vs Birth order last year 3

In figure 4.61, we see the percentage of women with birth order 3 against the states they live in. Highest is in Mizoram. It is higher than all the other states.

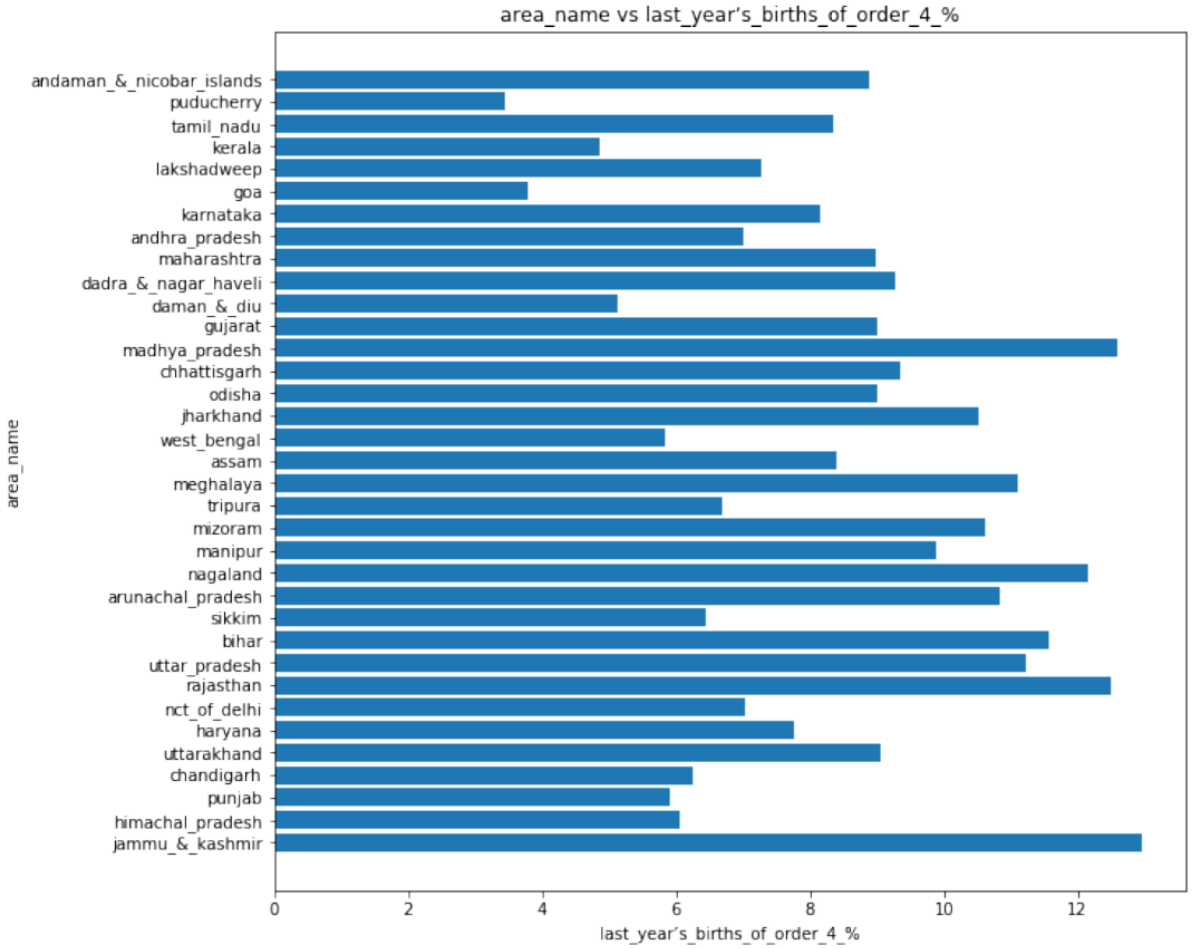

Figure 4.62: States vs Birth order last year 4

In figure 4.62, we see the percentage of women with birth order 4 against the states they live in. highest is in Jammu and Kashmir. It is followed by Madhya Pradesh, Rajasthan, Nagaland, Bihar, and so on. The lowest is in Puducherry.

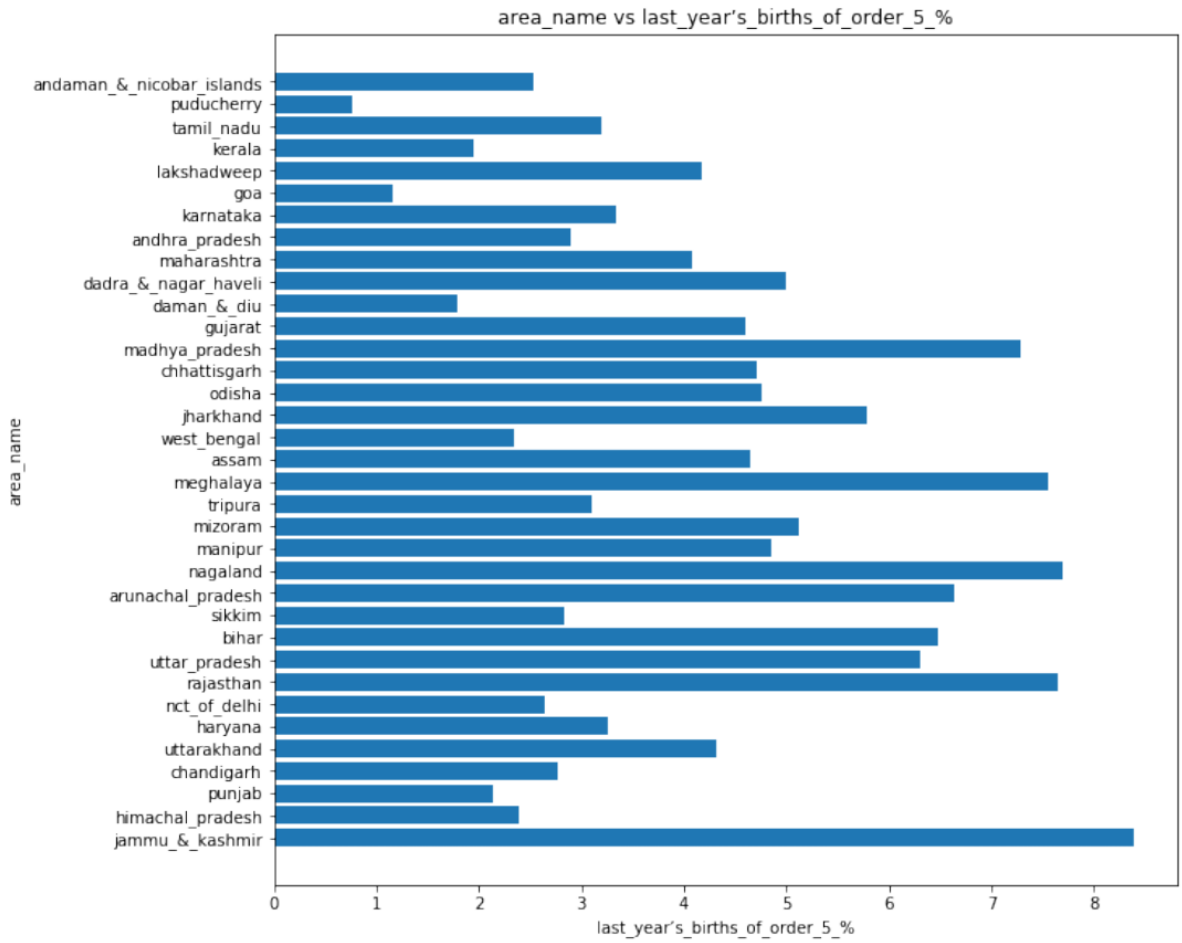

Figure 4.63: States vs Birth order last year 5

In figure 4.63, we see the percentage of women with birth order 5 against the states they live in. highest is in Jammu and Kashmir. It is followed by Meghalaya, Madhya Pradesh, Nagaland, Rajasthan, and so on. The lowest is in Puducherry.

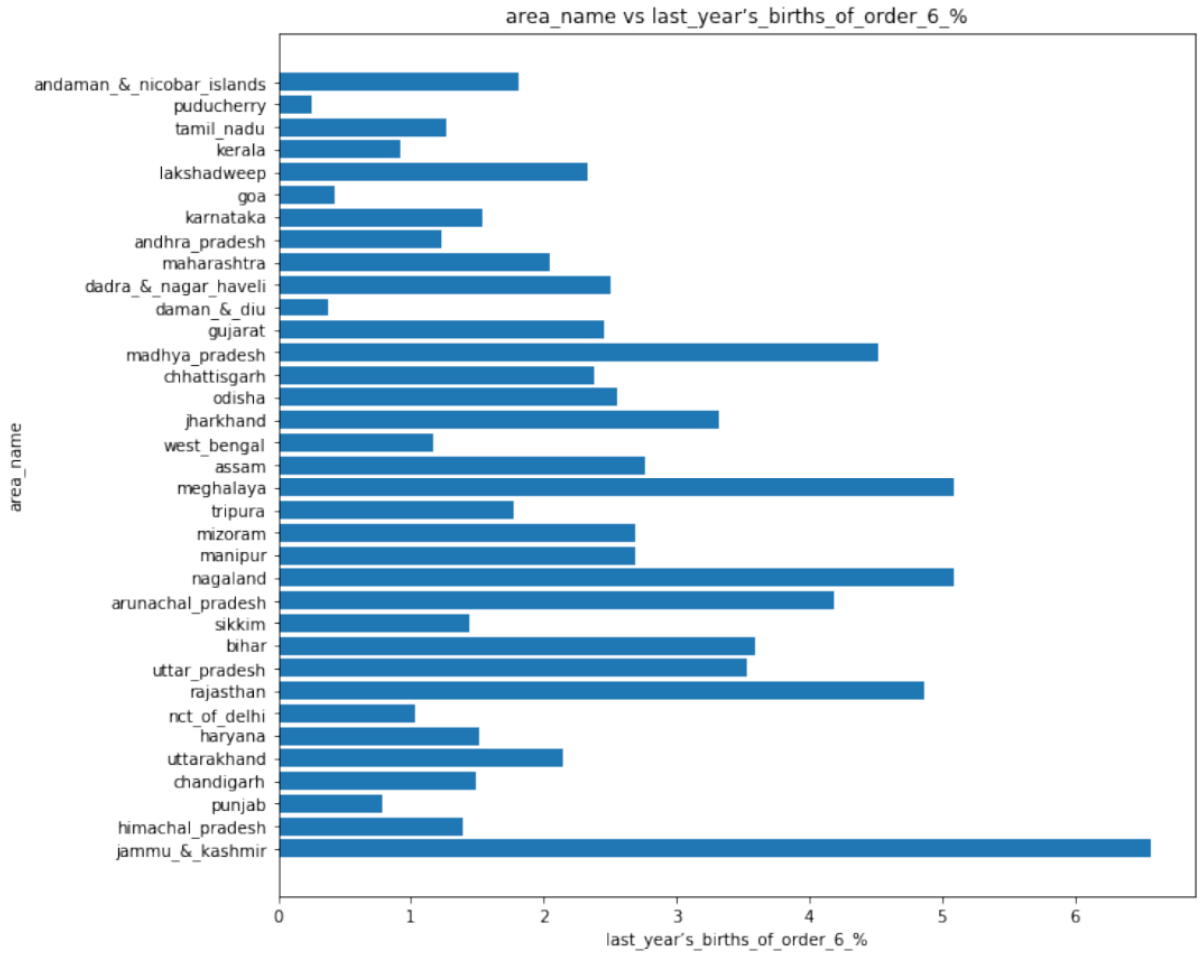

Figure 4.64: States vs Birth order last year 6

In figure 4.64, we see the percentage of women with birth order 6 against the states they live in. highest is in Jammu and Kashmir. It is followed by Meghalaya, Nagaland, Rajasthan, Madhya Pradesh, and so on. The lowest is in Puducherry.

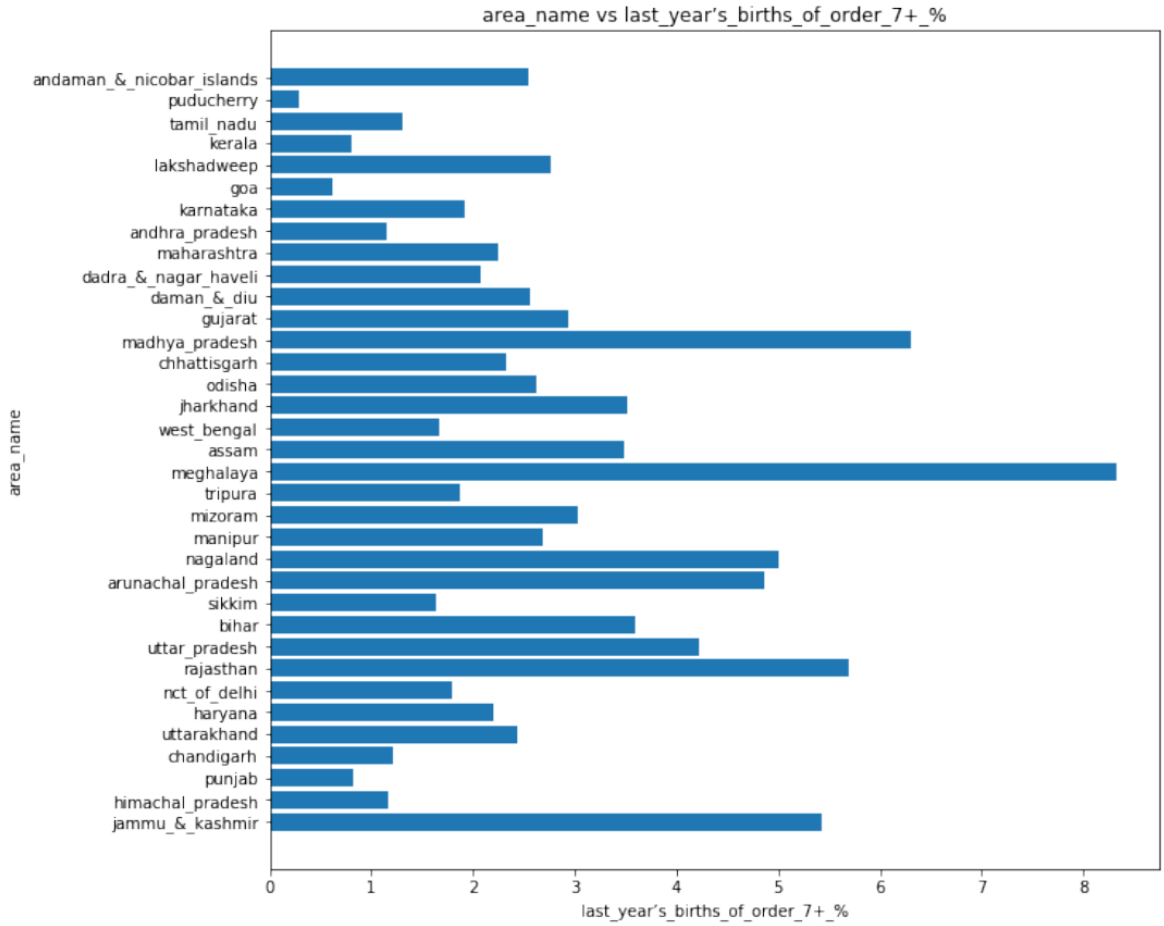

Figure 4.65: States vs Birth order last year 7+

In figure 4.65, we see the percentage of women with birth order 7+ against the states they live in. highest is in Meghalaya. It is followed by Madhya Pradesh, Rajasthan, Jammu and Kashmir, Nagaland, and so on. The lowest is in Puducherry.

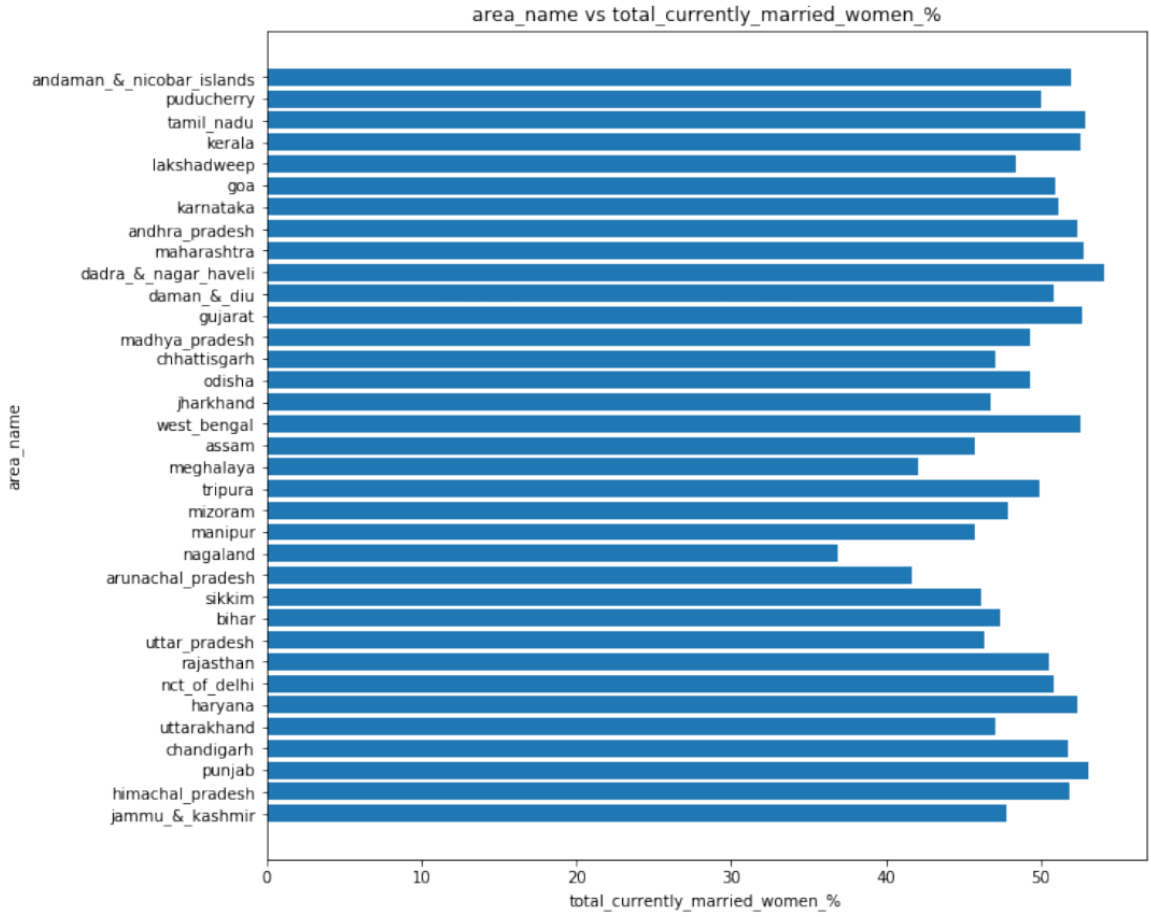

Figure 4.66: States vs currently married women

In figure 4.66, we see the percentage of currently married women against the states they live in. highest is in Dadra and Nagar haveli. It is followed by Punjab, Haryana, West Bengal, Tamil Nadu, and so on. The lowest is in Nagaland.

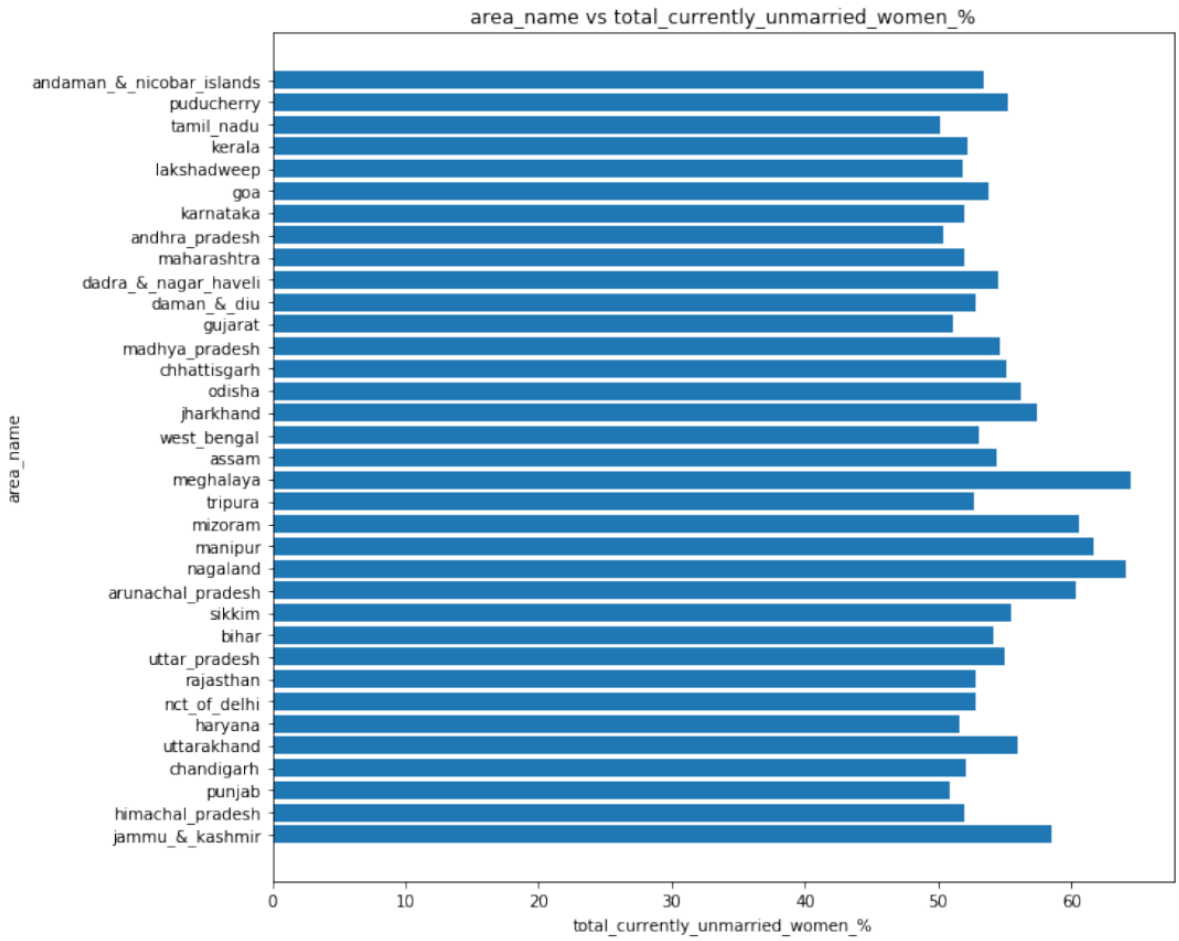

Figure 4.67: States vs currently unmarried women

In figure 4.67, we see the percentage of currently unmarried women against the states they live in. highest is in Meghalaya. It is followed by Nagaland, Manipur, Mizoram, Arunachal Pradesh, and so on.

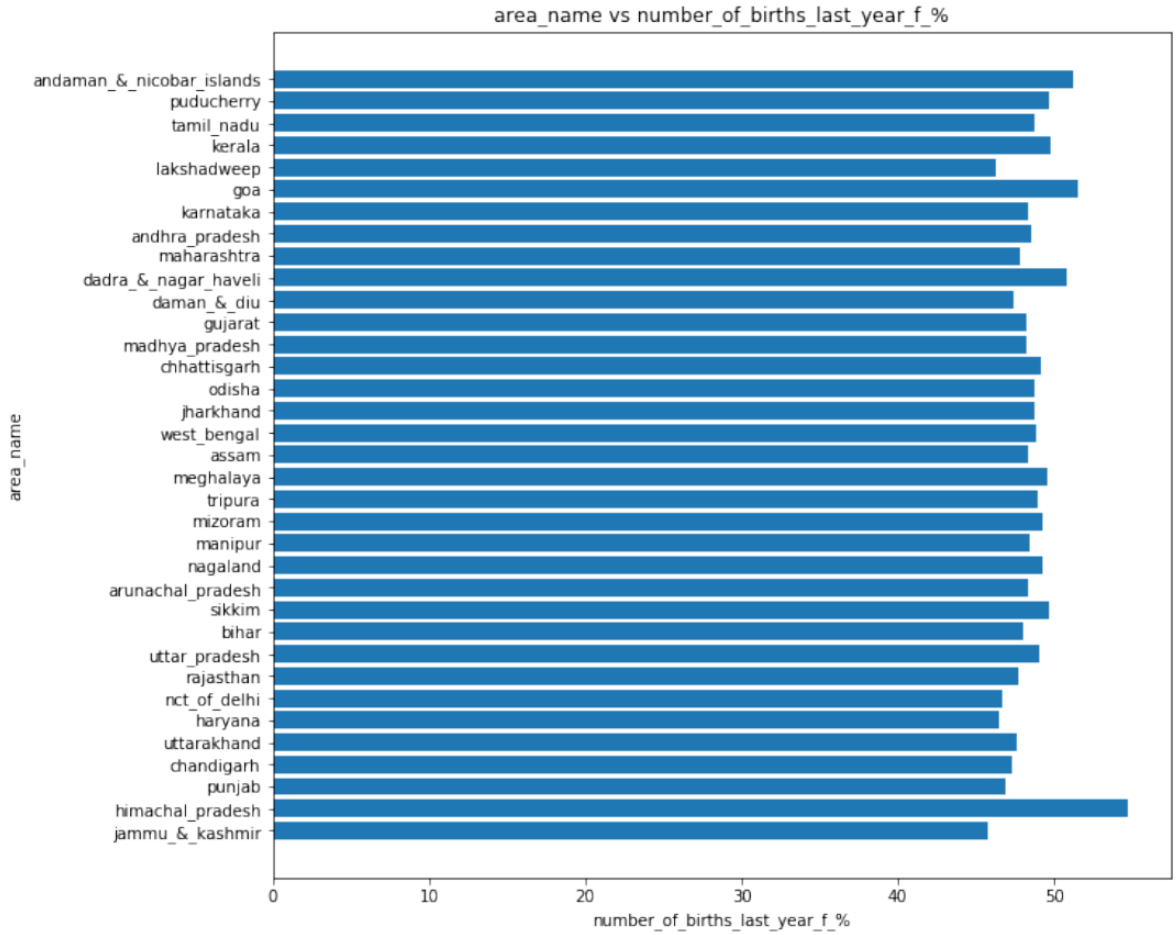

Figure 4.68: States vs number of female births

In figure 4.68, we see the percentage of the number of female births against the states they live in. highest is in Himachal Pradesh. The highest in the percentage of the number of male births against the states they live in is Mizoram.

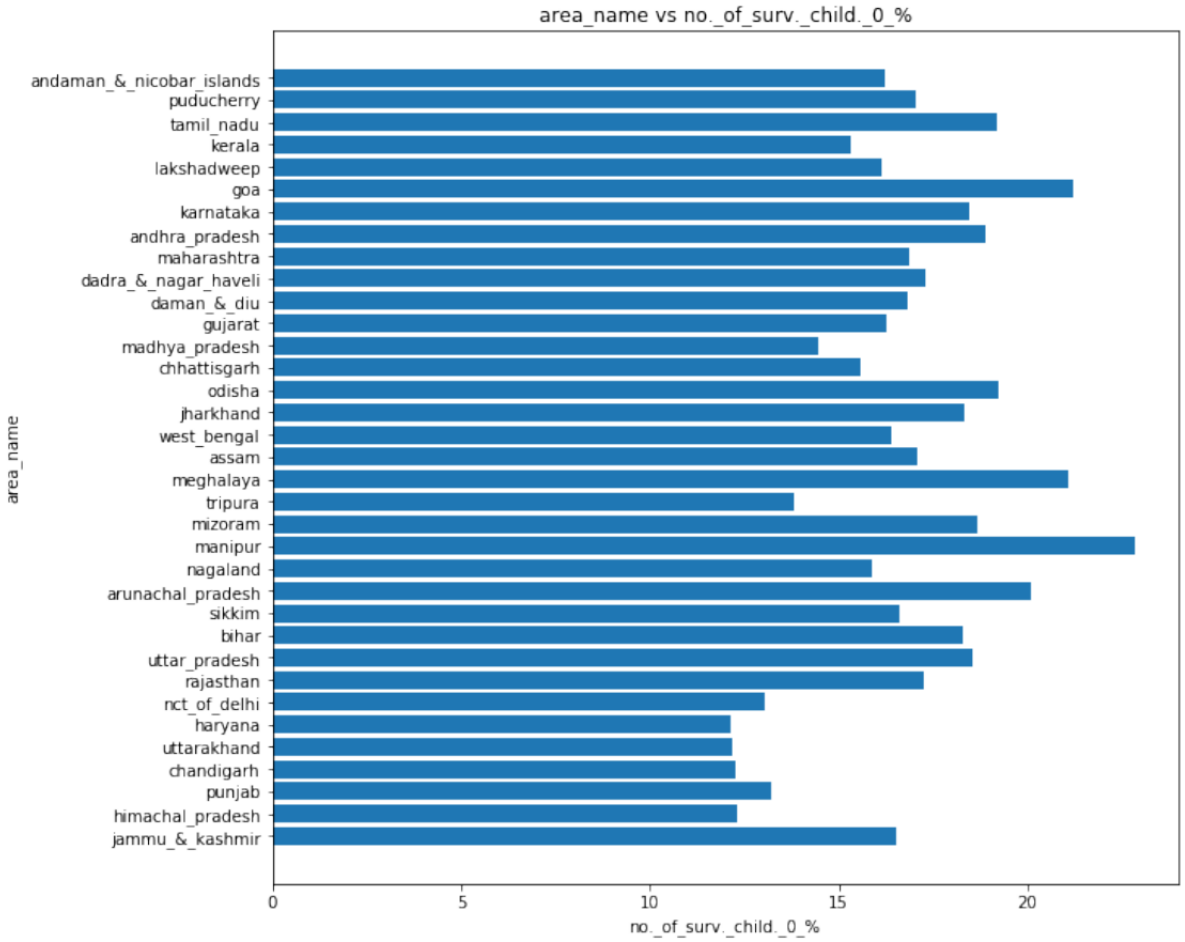

Figure 4.69: States vs number of surviving children 0

In figure 4.69, we see the percentage of the number of surviving children 0 against the states they live in. highest is in Manipur. It is followed by Meghalaya, Goa, Arunachal Pradesh, Odisha, and so on. The lowest are Haryana, Uttarakhand, Himachal Pradesh, and Chandigarh.

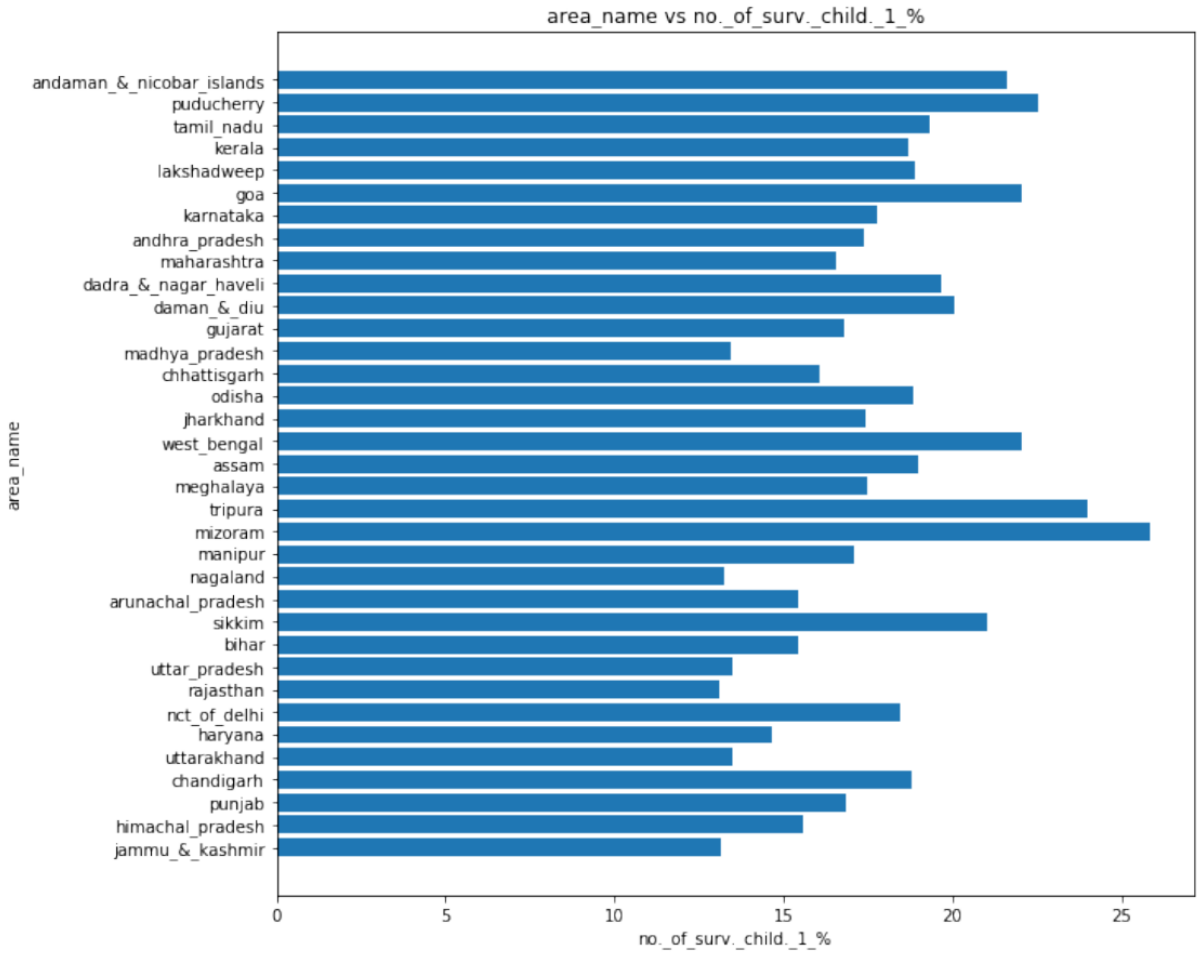

Figure 4.70: States vs number of surviving children 1

In figure 4.70, we see the percentage of the number of surviving children 1 against the states they live in. highest is in Mizoram. It is followed by Tripura, West Bengal, Goa, Puducherry, and so on. The lowest are Rajasthan, Jammu and Kashmir, Uttarakhand, Uttar Pradesh, and Madhya Pradesh.

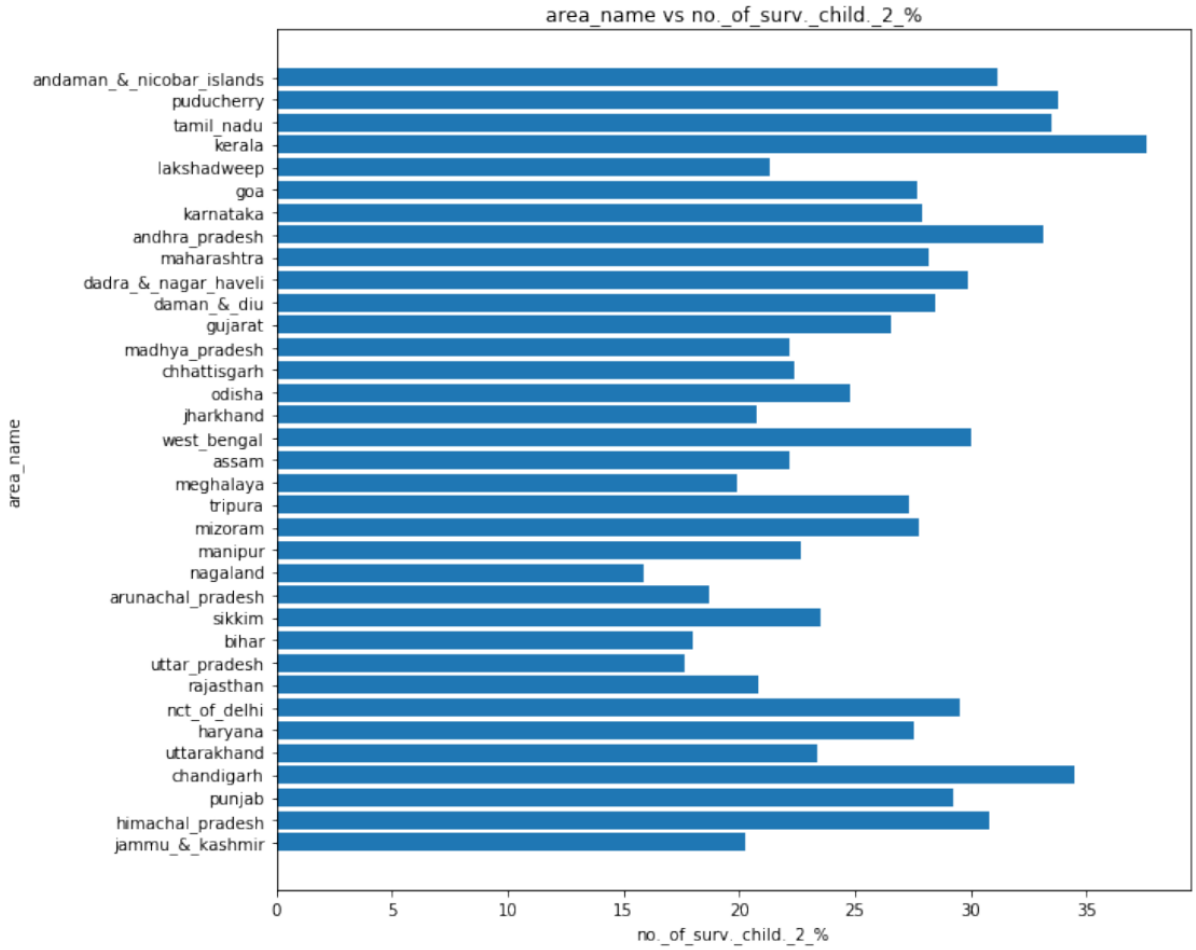

Figure 4.71: States vs number of surviving children 2

In figure 4.71, we see the percentage of the number of surviving children 2 against the states they live in. The highest is in Kerala. It is followed by Chandigarh, Puducherry, Tamil Nadu, Andhra Pradesh and so on. The lowest is Nagaland.

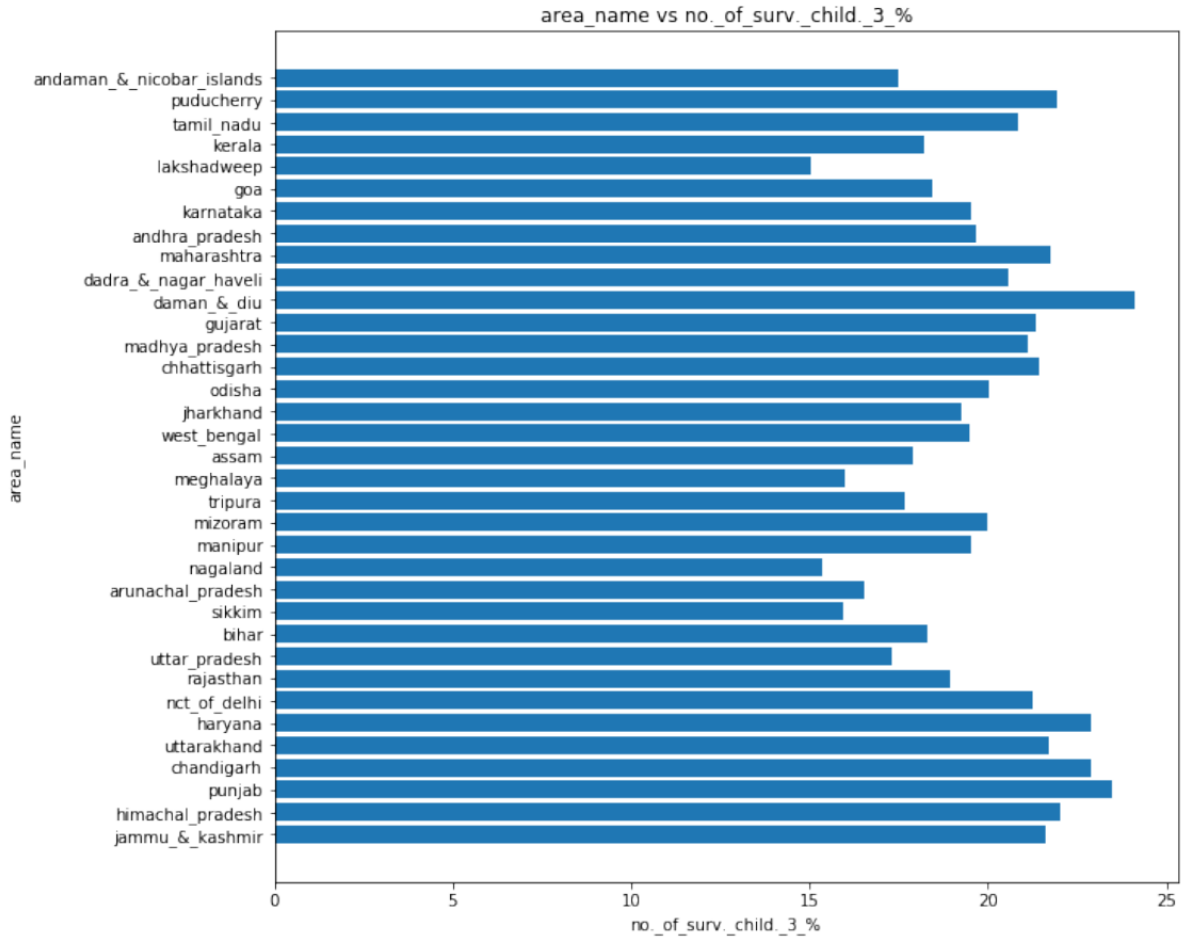

Figure 4.72: States vs number of surviving children 3

In figure 4.72, we see the percentage of the number of surviving children 3 against the states they live in. The highest is in Daman and Diu. It is followed by Punjab, Haryana, Chandigarh, Puducherry, and so on. The lowest are Lakshadweep and Nagaland.

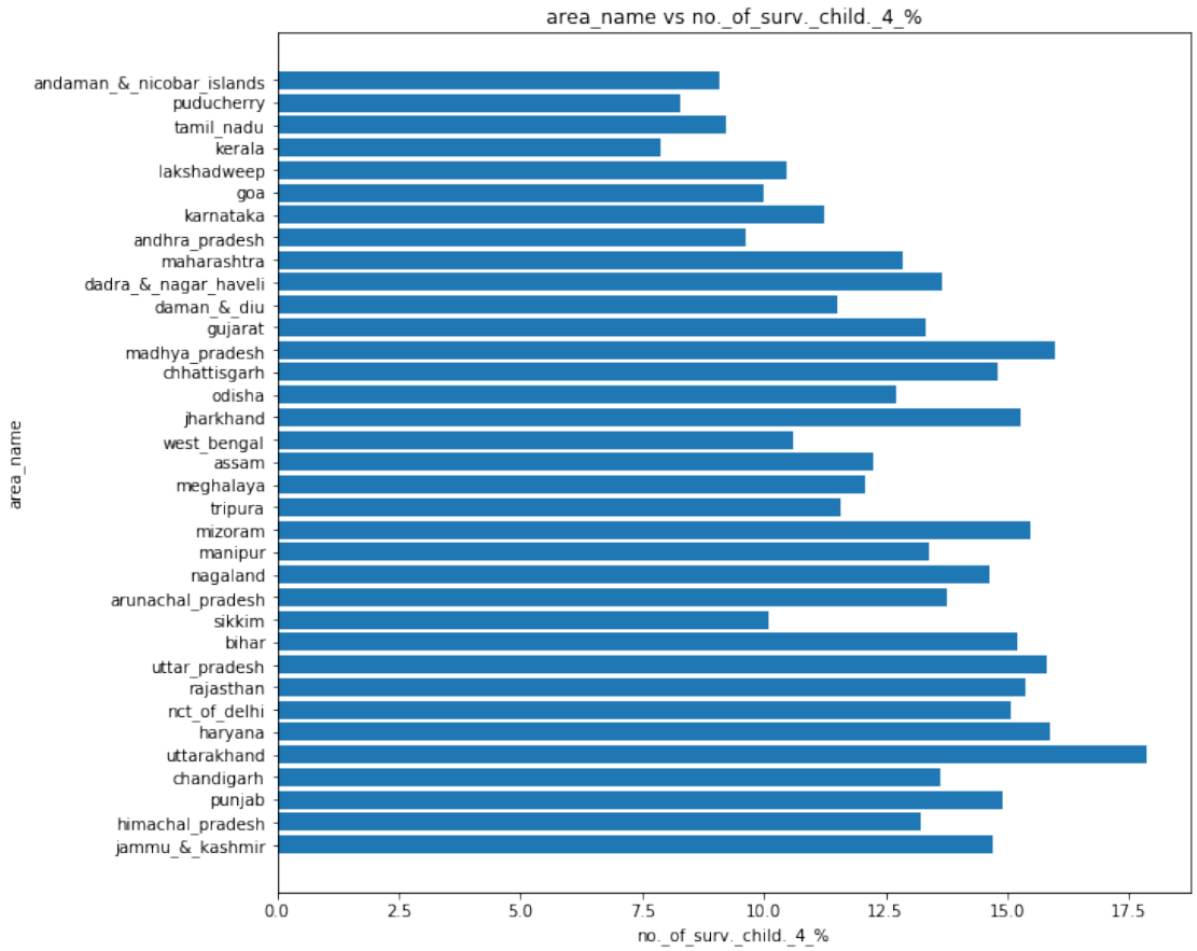

Figure 4.73: States vs number of surviving children 4

In figure 4.73, we see the percentage of the number of surviving children 4 against the states they live in. highest is in Uttarakhand. It is followed by Madhya Pradesh, Uttar Pradesh, Haryana, and so on. The lowest is Kerala.

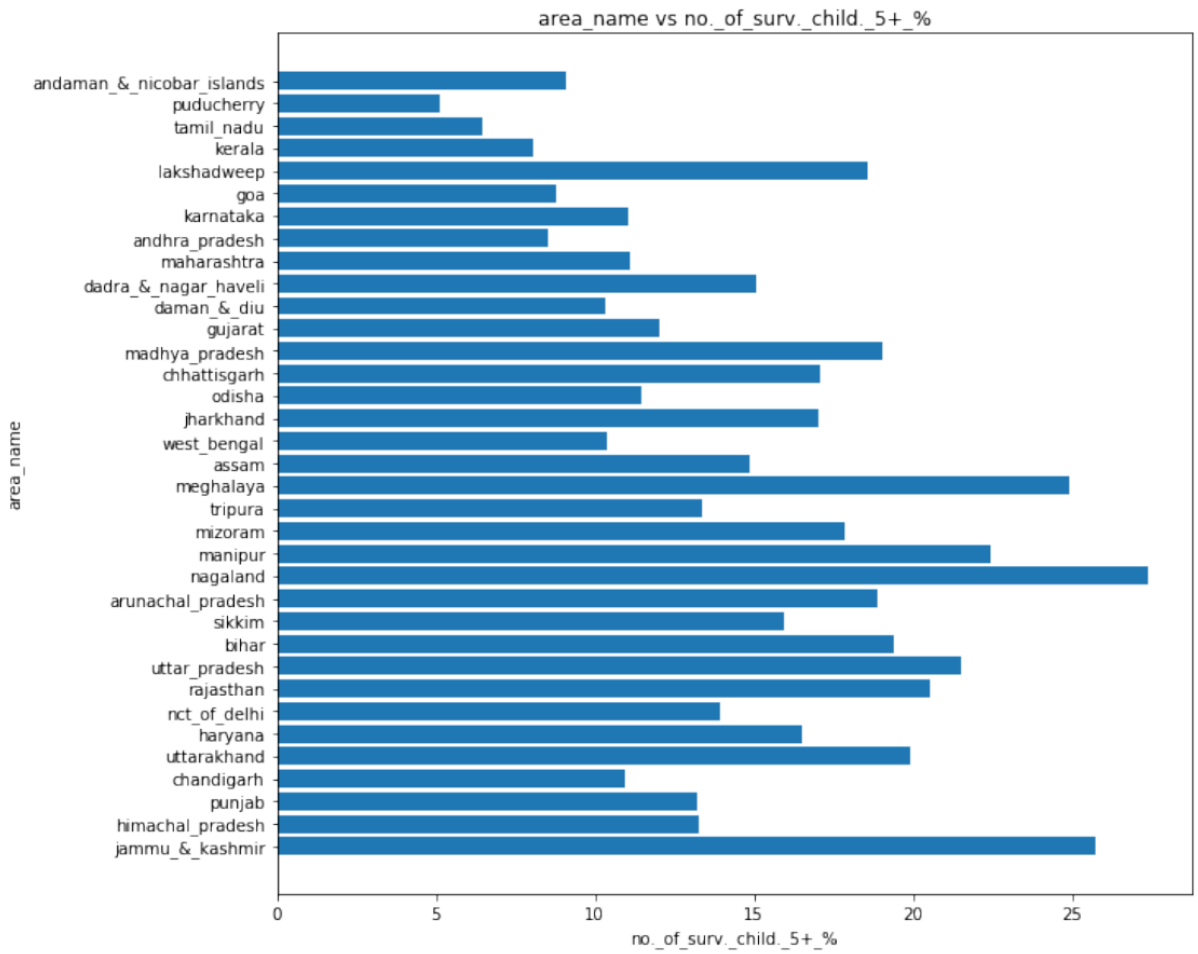

Figure 4.74: States vs number of surviving children 5+

In figure 4.74, we see the percentage of the number of surviving children 5+ against the states they live in. highest is in Nagaland. It is followed by Jammu and Kashmir, Meghalaya, Manipur, and so on. The lowest is Puducherry

The number of surviving female and male children in each state are almost the same.

4.3.4 Analysis on Infant Mortality and Fertility in each state (2011)

In this section, each state is analyzed individually. The top 4 states in each variable were taken and then complied into the tables of each state. This helps understand the main factors affecting infant mortality in each state. Improving on them will help reduce the infant mortality rate in each state which in turn leads to the reduction of infant mortality rate in India.

Table 4.1: Highest and Lowest in Andaman and Nicobar islands

Highest	Lowest
Parity 1 % with less than 15, Hindu and Jain, parity 2 % with less than 15, 15-19, 35-39, urban and ST, parity 3 % with 15-19, total children ever born m % 15-19, total children ever born f % 15-19, parity 5 % 20-24, parity 0% with 80+ and other communities, total ever unmarried women% with urban and Jain, parity 4 % (Jain), last year's births of order 1 % (Jain)	Parity 0, parity 2, parity 3, parity 4, parity 4 %, parity 5, parity 6, parity 7+, total women, total married women, total ever unmarried women, total children ever born p, total children ever born m, total children ever born f, parity 0 % with less than 15, 15-19, 20-24, Hindu, Christian and Jain, parity 7+ % with 20-24,25-29, urban and Jain, parity 3 % with 25-29, 30-34 and Jain, parity 4% with 35-39, rural and urban, parity 5% with 35-39, urban and Jain, parity 6 % with urban and Jain, parity 0% with Hindu, Christian, and Jain, total married women% (Jain), no of surviving child 5+ % (urban)

Table 4.2: Highest and Lowest in Puducherry

Highest	Lowest
Parity 2 %, parity 3 %, total children ever born f % with less than 15, 15-19, parity 2 % with 15-19, 30-34, 35-39, 80+, rural, urban, SC, Hindu and other communities, parity 3 % with 15-19, 30-39, rural and SC, total children ever born m % (20-24), parity 0 % (25-29), parity 1% with 35-39, 80+, other communities and illiterate, total ever married women% with urban and other communities, last year's births of order 4 % (Jain)	Parity 4 %, parity 5, parity 5 %, parity 6, parity 6 %, parity 7+, parity 7+ %, total children ever born m % (less than 15), parity 5 % with 20-24, 25-29, 30-34, 35-39, 80+, rural, urban, SC, Hindu, other communities and non-workers, parity 4 % with 25-29, 30-34, 35-39, rural, SC, Hindu, other communities and non-workers, parity 7+% with 25-29, 80+, rural, urban, SC and Hindu parity 6% with 30-34, 35-39, 80+, rural, urban, SC and Hindu, total ever unmarried women% with urban and other communities, parity 3 % (Jain), parity 0%

	(other communities), no of surviving child 5+ % (rural)

Table 4.3: Highest and Lowest in Tamil Nadu

Highest	Lowest
Parity 0, parity 0 %, parity 1, parity 2, parity 2 %, parity 2% with 30-34, 35-39, rural, urban and ST, parity 0% with 35-39, 80+, rural, ST, Jain and other communities, parity 1% (80+), parity 3 % with rural and ST, total ever married women% with rural and urban, total married women% with ST and Hindu	Parity 4 %, parity 6 %, parity 7+ %, parity 4 % with 15-19, urban, SC and ST, parity 6 % with 20-24, 25-29, 35-39, 80+, rural, urban, SC and ST, parity 7+ % with 20-24, 25-29, 35-39, 80+, rural, urban, ST and Hindu, parity 3% (ST), total children ever born m % (20-24), parity 5% with 80+, rural, urban, ST and Hindu, total ever unmarried women% with rural, urban, ST, Hindu and Christian

Table 4.4: Highest and Lowest in Kerala

Highest	Lowest
Parity 2 %, parity 0 % with less than 15, 25-29 and other communities, total children ever born f % with less than 15, 15-19 parity 1 % with 20-24, 25-29 and Jain parity 2% with 30-34, 35-39, rural, urban, SC, ST, Hindu, Christian, Jain, other communities, literate, total main workers and non-workers, total ever married women % with rural, urban, ST, Hindu, Christian, Jain, literate, non-workers, rural and SC, no of surviving child 2 % with rural, SC and urban	Parity 5 %, parity 7+ %, total children ever born m % with less than 15, 20-24, parity 2 % with 15-19 and 20-24, parity 3 % with 15-19, 20-24, 25-29 and other communities, parity 4 % with 15-19, 20-24, 25-29, 30-34, 35-39, rural, urban, SC, ST, Hindu, Jain, other communities and total main workers, parity 5 % with 20-24, 25-29, 30-34, 35-39, rural, urban, SC, ST, Hindu, other communities and total main workers, parity 6 % with 20-24, 25-29, 30-34, 35-39, rural, urban, SC, ST, Hindu, other communities and total main workers, parity 7+ % with 20-24, 25-29, 30-34, 35-39, 80+, rural, SC,

	ST and total main workers, total ever unmarried women% (rural, urban, ST, SC, Hindu, Christian, Jain, literate and non-workers, parity 0% with Hindu and Christian, no of surviving child 4 % with rural and urban

Table 4.5: Highest and Lowest in Lakshadweep

Highest	Lowest
Parity 7+ %, total children ever born m % (less than 15), parity 0 % with 15-19, 20-24, 25-29, 30-34 and Hindu, total children ever born f % (l15-19), parity 1 % with 25-29, 30-34, ST, Hindu, Christian, Jain and total main workers, parity 7+ % with 80+, rural, urban, ST, illiterate and non- workers, total married women% (Hindu), parity 2% with Hindu and Christian, total married women% with Christian and other communities, total unmarried women% (Jain), parity 4% (other communities	parity 0, parity 2, parity 2 %, parity 3, parity 3 %, parity 4, parity 4 %, parity 5, parity 6, parity 6 %, parity 7+, total women, total married women, total ever unmarried women, total children ever born p, total children ever born m, total children ever born f, parity 1 % with less than 15, 15-19, 80+ and other communities, parity 2 % with less than 15, 15-19, 20-24, 25-29, urban, Jain, other communities and illiterate, total children ever born f % with less than 15, parity 3 % with 15-19, 20-24, 25-29, 30-34, 35-39, rural, urban, ST, Hindu, Christian, Jain, other communities and illiterate total children ever born m % with 15-19 and 20-24, parity 4 % with 20-24, 25-29, 30-34, 35-39, 80+, rural, ST, Hindu, Christian and Jain, parity 5 % with 20-24, 25-29, 30-34, Hindu, Christian, Jain and other communities, parity 6 % with 20-24, 25-29, 30-34, Hindu, Christian, Jain, and other communities), parity 7+ % with 20-24, 30-34, 80+, Hindu, Christian and Jain,

	total ever unmarried women% with urban, Hindu, Christian and other communities, total married women% with Jain, parity 0 % with Jain and other communities

Table 4.6: Highest and Lowest in Goa

Highest	Lowest
Parity 0 %, parity 1 %, parity 0 % with less than 15, 15-19, 20-24, 25-29, 30-34, 35-39, rural, SC, ST, Hindu, other communities and total main workers, total children ever born m % with less than 15, 15-19 and 20-24, parity 1 % with 25-29, 30-34, 35-39, 80+, rural, SC, Hindu, Jain and other communities, parity 2% (80+), parity 3% (80+), total married women% with ST, Christian and other communities, parity 4 % (illiterate), total unmarried women% (total main workers)	Parity 5 %, parity 6, parity 6 %, parity 7+, parity 7+ %, parity 2 % with less than 15, 20-24 and 25-29, parity 1 % (20-24), parity 3 % with 25-29, 30-34, Hindu and other communities, parity 4 % with 25-29, 30-34, 35-39, ST, Hindu, Jain and other communities, parity 5 % with 25-29, 30-34, 35-39, rural, SC, ST, Hindu and Jain, parity 6% with 30-34, 35-39, 80+, rural, ST, Hindu and other communities, parity 7+% with 35-39, rural, SC, ST, Jain and Hindu, total ever unmarried women% with rural, ST and other communities, total married women% (total main workers)

Table 4.7: Highest and Lowest in Karnataka

Highest	Lowest
Total ever married women% with rural and ST, parity 0% with SC and Christian, parity 1% (ST)	Total ever unmarried women% rural and ST

Table 4.8: Highest and Lowest in Andhra Pradesh

Highest	Lowest
Parity 0, parity 1, parity 2, parity 2 %, parity 3, total women, total married women, parity 2 % with 15-19, 20-24, 25-29, 30-34, 80+, rural, SC, ST, illiterate and Hindu, parity 3% (30-34), parity 1% (80+), parity 0 % with urban, Jain and other communities, total married women% with Hindu and illiterate	Parity 5 %, parity 6 %, total children ever born m % (less than 15), parity 4 % with 15-19, urban and literate, parity 1 % with 20-24, 25-29, parity 6 % with 20-24, 80+, rural, urban, SC, ST, Hindu and illiterate, parity 7+ % with 30-34, 35-39, rural, urban and illiterate, parity 5% with 80+, rural, urban, ST and illiterate, total ever unmarried women% with rural, ST, Hindu and illiterate)

Table 4.9: Highest and Lowest in Maharashtra

Highest	Lowest
Parity 0, parity 1, parity 2, parity 3, parity 4, parity 5, parity 6, total women, total married women, total ever unmarried women, total children ever born p, total children ever born male, total children ever born female, parity 2 % (20-24), parity 3 % with rural, SC, ST and total main workers, total ever married women% with rural, Christian and Jain, parity 0% (Christian)	Parity 7+ % with 80+ and Christian, total unmarried women% with Christian and Jain, parity 4% (Christian), parity 5% (Christian), parity 6% (Christian)

Table 4.10: Highest and Lowest in Dadra and Nagar Haveli

Highest	Lowest
parity 1 % with less than 15, 15-19, 80+, urban, SC , Jain and other communities, parity 2 % with less than 15, 25-29, rural and SC, total children ever born f % (less than 15, parity 3 % with 20-24 and 25-29, parity 6 % (20-24), parity 7+ % (20-24), parity 4% with 80+ and other communities, total ever married women% (rural)	Parity 0, parity 2, parity 3, parity 4, parity 5, parity 6, parity 7+, total women, total married women, total ever unmarried women, total children ever born p, total children ever born m, total children ever born f, parity 5% with SC, other communities and literate, parity 6% with SC, Jain and literate, parity 7+% with SC, Jain and literate, parity 0 % with Jain and other communities, parity 4 % (Jain)

Table 4.11: Highest and Lowest in Daman and Diu

Highest	Lowest
Parity 3 %, parity 0 % (less than 15), parity 2 % with less than 15, 15-19, 25-29, 30-34, ST , Christian, Jain and other communities, total children ever born m % with less than 15, 15-19 and 20-24, parity 1 % (20-24) parity 7+ % with 20-24 and Jain, parity 3% with 35-39, 80+, rural, urban and ST, parity 5% (80+), total married women% with ST and Christian, parity 1% (Christian), total married women% (other communities), no of surviving child 3 % with rural and urban, last year's births of order 2 % (Jain)	Parity 0, parity 2, parity 3, parity 4, parity 5, parity 5 %, total women. total married women, total ever unmarried women, total children ever born p, total children ever born m, total children ever born f, total children ever born f % (l5-19), parity 3 % 20-24, Christian and other communities, parity 4 % with 20-24, 25-29, 30-34, Christian, Jain and other communities, parity 5 % with 20-24, 30-34, 35-39, ST, Christian, Jain and other communities, parity 6 % with 20-24, 25-29, 30-34, 35-39, ST, Christian, Jain and other communities, parity 7% with 30-34, ST and Christian, total unmarried women% with Christian and other communities, parity 1% (other communities), parity 0% (ST)

Table 4.12: Highest and Lowest in Gujarat

Highest	Lowest
Parity 2 % with less than 15, 15-19 and other communities, parity 3 % with 15-19 and ST, parity 4 % (15-19), parity 0 % (20-24), total married women% with Jain and other communities	Parity 0 % (less than 15), total children ever born f % with less than 15 and 15-19, total unmarried women% with Jain, other communities and Jain, parity 4% (other communities)

Table 4.13: Highest and Lowest in Madhya Pradesh

Highest	Lowest
Parity 4, parity 4 %, parity 5, parity 6, parity 7+, parity 7+ %, total ever unmarried women, total children ever born p, total children ever born f, parity 2 % (20-24), parity 3 % with 20-24, 25-29, 30-34, Jain and other communities, parity 4 % with 20-24, 25-29, 30-34, 35-39, rural, SC, ST, Hindu and other communities, parity 5 % with 25-29, 30-34, ST and Hindu, parity 6 % with 25-29, 30-34, SC and Hindu, parity 7+% with 25-29, 30-34, 35-39, SC, ST and Hindu	parity 1 %, parity 0 % with 20-24, 25-29, 35-39 and other communities, parity 1 % with 25-29, 30-34, 35-39, rural, urban, SC,ST, Hindu, Jain, other communities and total main workers, parity 2% with Hindu and SC

Table 4.14: Highest and Lowest in Chhattisgarh

Highest	Lowest
Parity 4% with ST, Hindu, Christian, and other communities, parity 5% with Hindu and other communities, parity 3% with Christian, Jain and other communities, parity 7+ % (Jain)	Parity 0% with 30-34 and Christian, parity 1% with 30-34 and Jain, parity 2% (Hindu)

Table 4.15: Highest and Lowest in Odisha

Highest	Lowest
Parity 0% (35-39)	Parity 2 % (15-19), parity 7+ % (80+)

Table 4.16: Highest and Lowest in Jharkhand

Highest	Lowest
Parity 3 % with 20-24 and 25-29, parity 4 % with 25-29, 30-34, 35-39, SC and Hindu, parity 7% (30-34), parity 5% (SC), parity 6% (Christian)	Parity 0 % (15-19), total children ever born f % (15-19), parity 2% with 35-39, SC and Hindu

Table 4.17: Highest and Lowest in West Bengal

Highest	Lowest
Parity 1, parity 1 %, parity 2, parity 3, parity 4, parity 7+, total women, total married women, total ever unmarried women, total children ever born p, total children ever born m, total children ever born f, parity 1 % with 15-19, 20-24, urban, ST, Hindu and literate, parity 2 % with 20-24,25-29 and SC, parity 3% (80+), total ever married women% with urban, Hindu and Jain	Parity 0 % with 15-19, 20-24 and SC, parity 5 % (25-29), parity 6 % (25-29), parity 3 % (urban), parity 7+% (ST), total unmarried women% with Hindu and Jain

Table 4.18: Highest and Lowest in Assam

Highest	Lowest
Parity 1 % (urban)	Parity 3 % (20-24), parity 4% (Hindu)

Table 4.19: Highest and Lowest in Meghalaya

Highest	Lowest
Parity 0 %, parity 6 %, parity 7+ %, parity 1 % (less than 15), parity 2 % (less than 15), total children ever born m % (less than 15), parity 0 % with 15-19, 30-34, 35-39, 80+, rural, urban, ST, SC, Hindu, Jain and illiterate, total children ever born f % (15-19), parity 4 % (20-24), parity 5 % with 20-24, 25-29, 30-34 and Christian, parity 6 % with 20-24, 30-34, 25-29, 35-39, rural, urban, ST, Christian, other communities and other communities, parity 7+ % with 20-24, 25-29, 30-34,35-39, rural, urban, ST, Christian, literate and total main workers, parity 3% (80+), total ever unmarried women with rural, urban, ST, Christian, Jain and other communities, survived child 0 % (SC), last year's births of order 5 % (Jain)	Parity 2 %, parity 3 %, total children ever born m % (15-19), parity 2 % with 25-29, 30-34, 80+, rural, urban, ST, Christian and other communities, parity 3% with 30-34, 35-39, SC, ST, Hindu, Christian, literate, total main workers and non- workers, parity 4% (80+), total ever married women% with rural, urban, ST, Christian, Jain and other communities

Table 4.20: Highest and Lowest in Tripura

Highest	Lowest
Parity 1 %, parity 1 % with less than 15, l5-19, 20-24, 25-29, 30-34, 35-39, rural, urban, SC, ST, Hindu, other communities and non-workers, parity 4% with 80+ and Jain, parity 5% (80+), total ever married women% with urban and total main workers, parity 6 % (Jain), no of surviving child 1 % (urban), last year's births of order 6 % (Jain)	Parity 0 % with less than 15,15-19, 20-24, 25-29, 30-34, urban, SC and ST, parity 3 % with rural, urban, SC and Hindu, parity 1 % (Jain), last year's births of order 1 % (Jain)

Table 4.21: Highest and Lowest in Mizoram

Highest	Lowest
Parity 1 %, parity 4 %, parity 0 % with 15-19, urban and Hindu, parity 1 % with 15-19, 25-29, 30-34, rural, 35-39 and SC, total children ever born m % (15-19), parity 3% with 35-39, urban and Christian, parity 7+ % with 80+, Jain , total ever unmarried women% with rural, Hindu, Christian, Jain and illiterate, parity 4 % with urban, ST, Christian, Jain, other communities, literate and total main workers, parity 5% with Christian and Jain, parity 6 % with Jain and other communities, no of surviving child 1 % (rural), survived child 1% (SC)	Parity 1, parity 1% (80+), parity 4 % with 15-19, 80+ and SC, total children ever born m % (20-24), parity 2% with 80+, rural, Christian, Jain and other communities, parity 3% with 80+ and Jain, parity 0 % with rural, SC, ST, Christian and Jain, parity 5% (SC), parity 7+% (SC), total ever married women% with rural, Hindu, Christian, Jain and illiterate, survived child 4 % (SC), survived child 5+ % (SC)

Table 4.22: Highest and Lowest in Manipur

Highest	Lowest
Parity 0 %, parity 6 %, parity 1 % with less than 15 and 20-24, total children ever born m % (less than 15), parity 4 % (15-19), parity 0 % with 20-24, 25-29, 30-34, 35-39, 80+, rural, ST, Christian and Jain, parity 6 % with 80+, rural, urban, ST, Christian and Jain, total ever unmarried women% with rural, urban, ST and Christian, parity 5 % with urban, Christian and other communities, parity 7+% with urban and Christian, no of surviving child 0 % (rural)	Parity 0 % (less than 15), parity 2 % with 15-19, 20-24, ST, Christian and Jain, parity 3% with 30-34 and ST, total married women% with ST and Christian, parity 1% (Christian)

Table 4.23: Highest and Lowest in Nagaland

Highest	Lowest
Parity 5 %, parity 6 %, parity 7+ %, parity 0 % (20-24), parity 6 % with 20-24, 35-39, 80+, rural, urban, ST, Christian, Jain, other communities, illiterate, literate and total main workers, parity 5 % with rural, urban, ST, Christian, Jain, other communities, literate and total main workers, parity 7+ % with rural, urban, ST, Christian and Jain, total ever unmarried women% with rural, urban, ST, Hindu, Christian, other communities, non- workers and rural, parity 4% (Christian)parity 2 % (total main workers), no of surviving child 5+ % (rural)	Parity 1, parity 1 %, parity 2 %, parity 3 %, parity 1 % with less than 15, 20-24, 80+, rural, ST, Christian, other communities and literate, total children ever born m % with less than 15 and 15-19, parity 2 % with 20-24, 25-29, 30-34, 35-39, 80+, rural, urban, ST, Christian, other communities, literate and total main workers, parity 3 % with 25-29, 35-39, 80+, rural, urban, ST and other communities, parity 4% (80+), total ever married women% with rural, urban, ST, Hindu, Christian, other communities and non- workers, no of surviving child 1 % (rural), no of surviving child 2 % with rural and urban, no of surviving child 3 % (rural)

Table 4.24: Highest and Lowest in Arunachal Pradesh

Highest	Lowest
Parity 0 %, parity 5 % with 20-24, Christian, Jain, and other communities, parity 6 % with 20-24, Christian, Jain, and other communities, parity 7+ % with 20-24, 25-29, Christian and Jain, parity 0% with 80+, rural, urban, ST, Hindu, Jain, literate and non- workers, total ever unmarried women% with urban, ST, Hindu, Christian and Jain, other communities and literate, parity 4% with Christian and Jain, no of surviving child 5+ % (urban), last year's births of	Parity 1, parity 3 %, parity 2 % with less than 15, 25-29, 30-34, 35-39, rural, Christian and other communities, total children ever born m % (15-19), parity 3% with 35-39, rural and Jain, parity 5% (80+), parity 6 % (80+), total ever married women% with rural, urban, ST, Hindu, Christian, Jain, other communities, literate, parity 1% with Christian, Jain and other communities, last year's births of order % (Jain)

order 3 % (Jain), last year's births of order 7+ % (Jain)	

Table 4.25: Highest and Lowest in Sikkim

Highest	Lowest
Parity 1 %, total children ever born f % (less than 15), parity 1 % with 15-19, 30-34, 35-39, rural, urban, SC, ST and Christian, total children ever born m % (15-19), parity 0% (30-34), parity 7+ % with 80, SC and Hindu, total unmarried women% (Hindu), parity 4 % (Jain)	parity 0, parity 1, parity 2, parity 3, parity 3 %, parity 4, parity 4 %, total women, total married women, total ever unmarried women, total children ever born p, total children ever born m, total children ever born f, parity 0 % (15-19), parity 3 % with 15-19, 80+, urban, SC, ST, Hindu and Christian, Parity 4 % with 15-19, 20-24, 80+, rural, urban, SC and ST, parity 5% (80+), total married women% (Hindu), parity 2 % (Jain), parity 6 % (Jain), survived child 3 % (SC)

Table 4.26: Highest and Lowest in Bihar

Highest	Lowest
Parity 0, parity 3, parity 4, parity 5, parity 5 %, parity 6, parity 7+, total women, total married women, total ever unmarried women, total children ever born p, total children ever born m, total children ever born f, parity 3 % (20-24), parity 4 % with 20-24, 25-29, 30-34, 80+, urban, Hindu and non- workers, parity 5 % with 20-24, 25-29, 30-34, 35-39, urban, SC, Hindu and Jain, parity 6 % with 25-29, 30-34, 35-39, urban, SC and Hindu, parity 7% with 30-34, 35-39,	parity 2 %, parity 2% with 30-34, 35-39, urban, SC and Hindu, parity 1% with 35-39, urban, 80+ and Hindu, total ever married women% with rural, urban and Hindu, parity 3% (SC)

SC and Hindu, total unmarried women% (Hindu), parity 0% (Christian)	

Table 4.27: Highest and Lowest in Uttar Pradesh

Highest	Lowest
Parity 0, parity 1, parity 2, parity 3, parity 4, parity 5, parity 5 %, parity 6, parity 6 %, parity 7+, total women, total married women, total ever unmarried women, total children ever born p, total children ever born m, total children ever born f, parity 4 % with 15-19, 20-24, 25-29, 30-34 and ST, parity 5 % with 25-29, 30-34, 35-39, rural, urban, SC, ST and Hindu, parity 6 % with 25-29, 30-34, 35-39, rural, urban, SC and Hindu, parity 7+% with 25-29, 30-34, 35-39, rural, urban, SC, Hindu and Christian, Parity 0 % with urban, SC, Christian and literate, no of surviving child 0 % (urban), total unmarried women % (SC), survived child 5+ % (SC)	Parity 1 %, parity 2 %, parity 0 % (less than 15), parity 1 % with l5-19, 20-24, 25-29, 30-34, 35-39, SC, Hindu and non- workers, parity 2% with 30-34, 35-39, rural, SC, ST and Hindu, parity 3% with 35-39, rural and SC, total ever married women% urban and SC, survived child 2 % (SC)

Table 4.28: Highest and Lowest in Rajasthan

Highest	Lowest
Parity 5, total children ever born m, parity 0 % with less than 15, 15-19 and literate, parity 3 % with 20-24 and 25-29, parity 4 % with 25-29 and 35-39, parity 5 % with 25-29, 30-34, 35-39, rural, SC, ST, Hindu and Jain, parity 6% with SC, ST, Hindu and other communities, parity 7+% with SC and Hindu	Parity 1 %, parity 1 % with less than 15, 15-19, 20-24, 25-29, 30-34, 35-39, 80+, rural, urban, SC, ST and Hindu, parity 2 % with less than 15, 15-19, 80+, urban and SC, parity 3 % with 15-19 and 80+, parity 0 % with 25-29 and 30-34, total unmarried women% (total main workers)

Table 4.29: Highest and Lowest in NCT of Delhi

Highest	Lowest
Parity 6, parity 7+, parity 7+ % (20-24), parity 2% with 80+, Christian and other communities, parity 4% with 80+, rural and urban, parity 1% (Christian)	Parity 0 %, total children ever born f % (less than 15), parity 1 % with 15-19, 25-29, 30-34 and urban, parity 3 % with 20-24 and Christian, total children ever born m % (20-24), parity 0% with 35-39, 80+, rural, urban and Jain, parity 4% (Christian), parity 5% (Christian), parity 6% with Christian and other communities, parity 7+% (Christian), no of surviving child 0 % (rural), last year's births of order 5 % (Jain), last year's births of order 6 % (Jain)

Table 4.30: Highest and Lowest in Haryana

Highest	Lowest
Parity 4 %, total children ever born m % (20-24), parity 3% with 30-34, Hindu and Jain, parity 4% with 35-39, rural, urban and SC, parity 7+ % (80+), total unmarried women% (other communities)	Parity 0 %, parity 1 % with less than 15, 80+ and SC, total children ever born f % (15-19), parity 0 % with 20-24, 25-29, 30-34, 35-39, 80+, rural, urban, SC and Hindu, parity 2% (80+), parity 3% (80+), total married women% (other communities)

Table 4.31: Highest and Lowest in Uttarakhand

Highest	Lowest
Parity 4 %, parity 5 %, parity 3 % with 15-19, 25-29, 30-34, Hindu and literate, parity 4 % with 15-19, 30-34, 35-39, 80+, rural, urban, SC, ST, Hindu and illiterate, parity 2 % (20-24), parity 5% with 35-39, 80+, rural, urban, SC and illiterate, parity 6% with SC	Parity 0 %, parity 0 % with 25-29, 35-39, urban, SC, ST, Hindu and total main workers, parity 1 % with rural, SC, ST and Hindu, total married women% (other communities), survived child 0 % (SC), survived child 1 % (SC)

and Hindu, total unmarried women% (other communities), no of surviving child 4 % with rural, urban and SC	

Table 4.32: Highest and Lowest in Chandigarh

Highest	Lowest
Parity 2 %, parity 3 %, total children ever born f % (less than 15), parity 2% with 80+, Hindu, Christian and Jain, parity 3% with 80+, urban, SC and other communities, parity 6 % (80+), parity 1 % (rural), parity 0 % (non- workers)	Parity 0 %, parity 1, parity 7+ %, total children ever born m % with less than 15 and 15-19, parity 3 % with 15-19 and illiterate, parity 4 % with 20-24 and Christian, parity 5 % with 20-24, Christian and Jain, parity 7+ % with 20-24, Hindu, Christian, Jain and non- workers, parity 0% with 30-34, 80+, urban, Christian and other communities, parity 1% (35-39), parity 6% with Christian, other communities and non-workers), no of surviving child 0 % (urban), last year's births of order 3 % (Jain)

Table 4.33: Highest and Lowest in Punjab

Highest	Lowest
Parity 3 %, parity 4 %, parity 3% with 30-34, rural, SC, Hindu, Christian, Jain, other communities and non- workers, parity 5% (80+), parity 6 % (80+), parity 4 % with rural, SC and Christian, total married women% (Jain), parity 2 % (Jain), survived child 3 % (SC)	Total children ever born f % (less than 15), parity 1 % with 15-19 and Christian, parity 0% with 35-39 and rural, total unmarried women% (Jain), last year's births of order 4 % (Jain), last year's births of order 7+ % (Jain)

Table 4.34: Highest and Lowest in Himachal Pradesh

Highest	Lowest
Parity 3 %, parity 0 % (less than 15), parity 2 % with 25-29, 35-39, urban and Jain, parity 3% with 35-39, urban, SC, ST, Hindu, Christian, other communities, parity 5% (80+)	Parity 0 %, parity 1 % with less than 15, other communities and illiterate, parity 2 % (less than 15), total children ever born f % (less than 15), parity 7+% with 25-29 and urban, parity 0 % with rural, SC, ST, Hindu, illiterate and literate

Table 4.35: Highest and Lowest in Jammu and Kashmir

Highest	Lowest
Parity 5 %, parity 6 %, parity 7+ %, parity 3 % with 15-19, 35-39, urban, Hindu and Christian, parity 4 % with 15-19 and 20-24, parity 5 % with 20-24, 35-39, rural, ST and non- workers, parity 6 % with 20-24, 25-29, 30-34, 35-39, 80+, rural, ST and non-workers, total children ever born m % (20-24), parity 7+% with 25-29, 35-39, 80+, rural and ST, total ever unmarried women% with rural and ST	Parity 1 %, total children ever born f % (l5-19), parity 1 % with rural, urban, ST and Christian, total married women% (ST), parity 2% (ST), no of surviving child 1 % (urban)

4.4 Machine Learning

Machine learning is used on these tables to figure out the importance of each factor that affect the survival rate. The data is provided with many algorithms to attain the expected results. In this case, python and packages like NumPy, pandas, Matplotlib, seaborn, and scikit-learn are used.

The data set is split into train and test set. The train data set is used to train the algorithm and the test is used to make predictions and confirm the accuracy. If the accuracy of the test data set is very low compared to the train data set it means that the model has overfitted. This means

that the model is high variance and low bias. If the model has low variance and high bias, it means that the model is under fitted. This means that the model is not trained well. This will lead to lesser accuracy in the train set itself. So, a bias-variance tradeoff needs to be done. In this, the model is trained without much loss in accuracy but also with a reasonable degree of predictability.

Two methods of machine learning tools are used: supervised and unsupervised. In supervised machine learning, the two options could be logistic regression and linear regression. Linear regression is used since the independent variable is continuous. Under unsupervised machines, learning options include clustering, PCA, Decision Tree, Random Forest, and so on. In this, the decision tree is used since it is not susceptible to outliers. Also, the study expects to understand all the important factors that affect the mortality rate.

4.4.1 Linear Regression

This type of machine learning is used when there is a continuous independent variable. When a model is made there are dependent and independent variables. The independent variables are the ones that affect the dependent variable. The dependent variable is usually related to the issue that needs to be solved. In this research, the dependent variable is the total survival rate of infants.

There are two types of linear regression: simple linear regression and multiple linear regression. Simple linear regression is used when there is only one independent variable. Multiple linear regression is used when there is more than one independent variable. In this case, multiple linear regression is used.

4.4.1.1 Linear Regression with mathematical implications

The first step is to find the best fit line (hyperplane in this case). An equation is formed by the use of the dependent variable(Y values) with coefficients that equates to the independent variable.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \epsilon$$

(1)

In equation (1), the Y is the total survival rate of infants in this research. $B_{1,2,..p}$ values are the coefficient of the independent variables. β_0 is the slope-intercept.

The residual sum of squares is also calculated by adding the square of the distance from the plotted points to the line. Using this the R^2 value is found.

$$R^2 = 1 - (RSS / TSS)$$

(2)

In this RSS is the residual sum of squares, TSS is the sum of errors of the data points from the mean. If the R^2 value is nearer to 1 it means that the deviation of the points from the line is less. But if the R^2 is almost near 0 it means that the data points are very scattered. So having an R^2 value closes to 1 means that the line is a better fit and the results thus obtained are more reliable.

4.4.1.2 Linear Regression in Python

In the first model, the columns include 'state code', 'district code', 'area name', 'total/rural/urban', 'present age', 'total ever married women', 'parity 0', 'parity 1', 'parity 2', 'parity 3', 'parity 4, 'parity 5', 'parity 6', 'parity 7+', 'total children ever born(persons)', 'total children ever born(male)', 'total children ever born(female)', 'all/SC/ST', 'total ever unmarried women ', 'number of surviving children- 0', 'number of surviving children- 2', 'number of surviving children- 3', 'number of surviving children- 4', 'number of surviving children- 5+', 'total surviving children(persons)', 'total surviving children(female)', 'total surviving children(male)', 'total currently married women', 'number of births last year (male)', 'number of births last year (female)', 'last year's births of order 1', 'last year's births of order 2', 'last year's births of order 3', 'last year's births of order 4', 'last year's births of order 5 ', 'last year's births of order 6', 'last year's births of order 7+ and 'total survival rate'.

The table is split into two, categorical variables and numerical variables. A dummy variable is made for each categorical variable. Next, all the independent variable (all except total survival rate) is assigned to X and the dependent variable is assigned to Y.

Using sklearn.model_selection package train_test_split is imported. Using the train_test_split, the data named basic is split into train and test data set in 70-30 proportions respectively. Using sklearn.preprocessing package MinMaxScaler is imported. This is used to standardize the numerical variables.

The feature selection Recursive Feature Elimination (RFE) is used to select the top 15 relevant variables. Next, using Ordinary Least Squares Regression variables are removed one by one by considering the p-value and Variance inflation factor (VIF).

Dep. Variable:	total_survival_rate	**R-squared:**	0.003
Model:	OLS	**Adj. R-squared:**	0.003
Method:	Least Squares	**F-statistic:**	17.38
Date:	Mon, 10 Aug 2020	**Prob (F-statistic):**	6.33e-38
Time:	14:04:16	**Log-Likelihood:**	84649.
No. Observations:	63051	**AIC:**	-1.693e+05
Df Residuals:	63038	**BIC:**	-1.692e+05
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.8838	0.000	3508.816	0.000	0.883	0.884
total_ever_married_women_x	40.2251	18.869	2.132	0.033	3.242	77.208
parity_0	-3.1448	2.833	-1.110	0.267	-8.697	2.407
parity_1	12.8832	6.353	2.028	0.043	0.431	25.335
parity_2	45.0848	25.976	1.736	0.083	-5.828	95.998
parity_3	92.0600	28.873	3.188	0.001	35.469	148.651
parity_4	5.7098	30.854	0.185	0.853	-54.765	66.185
parity_5	101.7483	24.956	4.077	0.000	52.834	150.662
parity_6	-1.2517	13.345	-0.094	0.925	-27.408	24.904
parity_7+	68.7210	30.605	2.245	0.025	8.735	128.707
total_children_ever_born_p	-171.6976	58.089	-2.956	0.003	-285.552	-57.843
total_children_ever_born_m	-288.6808	74.510	-3.874	0.000	-434.720	-142.641
total_children_ever_born_f	-40.3239	52.192	-0.773	0.440	-142.620	61.972
ttl_surv._child._p	49.7232	5.443	9.135	0.000	39.055	60.392
ttl_surv._child._m	128.3401	24.856	5.163	0.000	79.622	177.058

Figure 4.75: Linear model – 1

The R2 value is 0.003, which is very close to 0. This means that the model does not explain the data points very well. After creating many more models, the R2 value remained near 0. Along with that, the VIF values were in infinity. This shows that the data set plots are very scattered. Therefore, linear regression is not a good option for this dataset.

4.4.2 Decision tree in Python

The data for 2011 and 2001 was taken. Then the categorical values are converted into dummies in the basic table. In this case, each categorical values were given a number. Now that all the values have turned numeric, to reduce the confusion while making decision trees.

Next, the data is split into train and test data in 70-30 proportion respectively. All the independent variable (all except total survival rate) is assigned to X and the dependent variable is assigned to Y. A decision tree was generated automatically.

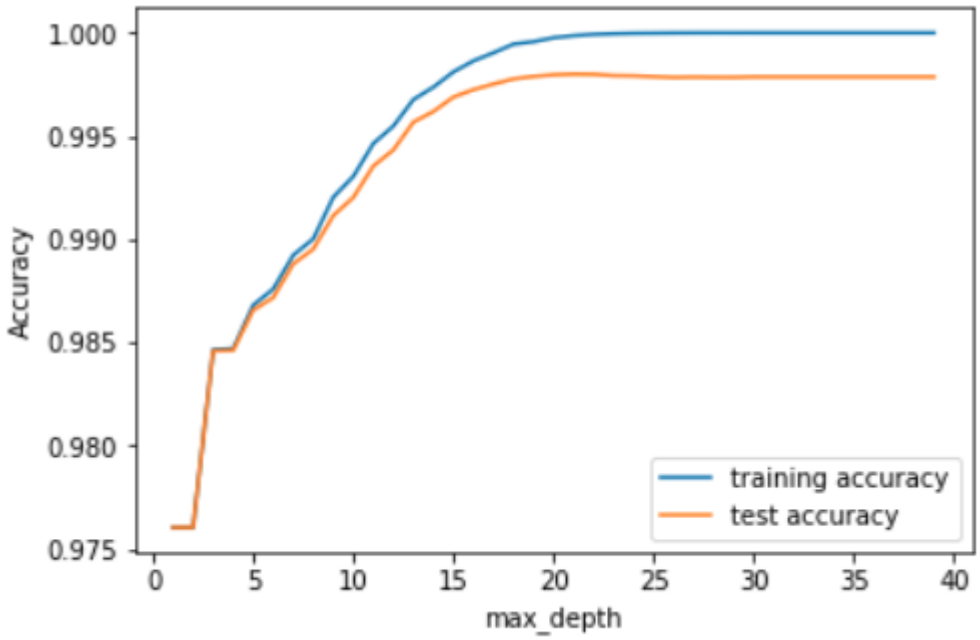

Figure 4.76: Tuning maximum depth

Next for pruning and hyperparameter tuning, a grid search CV is used. First, the maximum depth of the tree was figured with respect to the accuracy of the train and test data set. The depth was taken as 4 since the accuracy of the test and train dataset are almost the same. Similarly, minimum sample leaf and minimum sample splits were found to be 100 each.

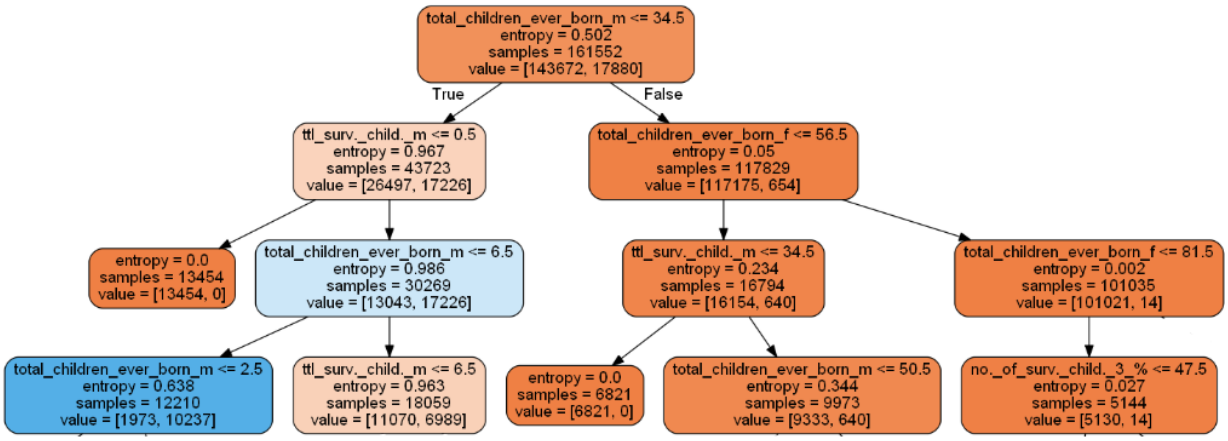

Figure 4.77: Decision tree

A decision tree with those parameters was made, resulting in figure 4.78. The same has been repeated in other models as well. This is the more relevant model. It shows that when total children ever born is more than or equal to 34.5, total surviving children are more prominent to be taken as an important factor, otherwise, total female children ever born. Under total female children ever born, when it is less than or equal to 56.5 total male surviving children is a more prominent factor. So, from this, we can understand the main important factors that affect the mortality rates.

4.5 Summary

This section involved EDA and machine learning. It started with data collection from the census website of India. The data collected was for 2001 and 2011. Then an analysis of the national and state level was conducted. Next, each state was analyzed individually. Next machine learning tools were used to analyze and figure the main important factors affecting infant mortality.

CHAPTER 5

RESULTS AND DISCUSSIONS

5.1 Introduction

This analysis is based on women and the number of children they have. The number of surviving children are also taken into consideration to understand the death rates of infants. India's infant mortality rate has decreased from 2000 to 2017. The annual rate of change from 2010 to 2017 differed among the districts. Inequality between districts increased from 2000 to 2017 in most of the states, especially in Odisha and Assam. If the trend continues like this India would meet the SDG 2030 target (to reach 12 deaths per 1000 live births). Some districts need a higher rate of improvement than they had up to 2017. The major causes of death of children under 5 are infectious diseases, neonatal disorders, malnutrition, etc. India has made a vast improvement to meet the global child survival target. Even though there are wide variations between the districts in India, this shows the importance of conducting the service at the lowest level in districts. Recent studies showed that neonatal deaths are very high in Bihar while compared to other states. Janani Suraksha Yojana program in India provides cash for women to deliver in health facilities which improve the maternal and newborn complications [16].

5.2 Interpretation of visualizations

In India, most of the women belong to Scheduled Caste more than Scheduled tribes. Most women belong to rural areas as compared to urban areas. This means that making changes in rural areas is important. After the 1991 census, the Indian government understood that in gender disparities in child mortality is in a dangerous state and measures like Beti Bachao Beti Padhao have taken to reduce the death rate of girls while compared to boys. But still, the death rate of infant girls is high in northeastern states. So the government must implement new policies for maternal and child health care especially in states like Bihar, Rajasthan, Uttar Pradesh, Haryana, Assam, and Jharkhand. The international program like SDGs is not working in the case of female health in India. To improve this problem government must introduce new programs to teach the importance of a girl child in society the same as a boy child [17].

Most women who have children belong to age groups 20-35. This is a good sign since it is the working class that has most children and they are matured emotionally and financially equipped to take care of themselves and their children. Most women belong to the Hindu community followed by the Muslim community. This means that they will have more influence on mortality and fertility rate.

The education level of women in the age group 20-35 has increased from 2001 top 2011. This is a good sign since educated women know about contraceptives and infant health. This helps with a reduction in fertility rate and infant mortality rates. Education and getting a job for girls is very important nowadays. It will improve the health and social status. It will extend the age of marriage which extends the birth of children also. Since the independence government has raised the age of marriage for girls from 18 to 21 years. Still among rural and uneducated people makes their girl child married before 18 years which leads to the death of mother and child. To bring the change in the child marriage the free education with job guarantees must be given to girls especially in rural areas which will extend the age of marriage [18].

Some women are married and have kids in the age group less than 15. This is illegal in India but still, it is followed among most areas. This practice of child marriage is more common in the north-eastern states of India. This is a great threat to the health of women and newborns. These causes pregnancy complications like premature babies and stillborn.

5.3 Summary

So the most important factors include education of women, age group they belong to, the religion of the mother and location of the mother (rural or urban). Even though there are many initiatives and programs taken by the government, they seem to not reach where it is required or the propaganda is not enough. This could be because most of the women belong to rural areas and the knowledge does not seem to be spreading enough.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

There are new schemes introduced by the government to protect pregnant women and infants under 5. This is an initiative to reduce complications during birth and pregnancy. Last year government launched SUMAN scheme which gives free medicines for pregnant women, mothers up to six months after delivery, and all sick newborns, which will help to bring down the maternal and infant mortality rates in the country. So the women and newborn babies will get good care according to the guidelines of WHO [15].

6.2 Discussion and Conclusion

Since 2001 there was no quality registration system to record the number of death rates of boys to girls. The Civil Registration System in 2011 was more effective to understand the infant death rates compared to the infant birth rates. By this system, we can understand that in northeastern states the death rates of infant girls are very high compared to boys. There is a huge variation within the states during the last two decades. In this article, it shows that there are high deaths especially females as compared to males. The government must review the system for data collection and update the methodology of the registration of deaths below the age of 5 in national state and district levels and update it regularly to understand the infant mortality rate U5MR and female mortality rate[19]

The small state in India, Kerala has achieved the United Nations' sustainable developmental goal, which is to reduce the infant mortality rate to 7 deaths per 1000 live births. Recently a newborn baby with a chronic heart defect survived after operation compared to the old days when it wouldn’t have happened. Earlier there were no modern hospitals and the money was a big issue in the state compared to these days. The state government formed the Hridayam project to save the children born with CHD free of cost. So the child born with severe heart problems survived. The baby with severe heart problems is detected before birth and women are shifted to the hospitals where the treatments are available, so soon after the delivery, the baby will be given full protection including surgery without any delay. These are all the reasons

to introduce Kerala, one among which reduced the infant mortality rate to reach the SDG target [20].

The maternal and infant mortality rates have only mild improvement compared to early surveys. Tamil Nadu, Maharashtra, and Kerala reached the MDG target whereas Kerala is the best among them in both maternal and infant mortality rates. Andhra Pradesh, Gujarat, West Bengal, and Harayana reached near to the MDG targets. The male-female difference had narrowed but the death of female children remained the same. It is narrowed because the number of best surgeons in obstetricians and gynecologists increased. So with more modern hospitals and best doctors we can reach the SDG target within years.

6.3 Contribution to knowledge

This research has helped to understand the relationship between infant mortality rate and many factors. If these factors are looked upon then the infant mortality can be reduced. This research helped realize that most of the women are in rural areas. And concentrating on them will create an improvement.

6.4 Future Recommendations

The restrictions in this research are that the census is taken every 10 years. The last time it was taken was in 2011. This will only give a very small idea of the scenario. The census must at least be taken every 5 years. This will help us understand the condition and give more realistic values. So researching with the next data which should be out in 2021 could lead to understanding the updates in infant mortality rate and fertility rate.

REFERENCES

[1] Myron Anthony Godinho, Shruti Murthy, Theophilus Lakiang, Amitha Puranik & Sreekumaran N. Nair, (2017), Mapping Neonatal Mortality in India: A Closer Look, DOI 10.4103/ijcm.IJCM_327_16, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5682725/

[2] Deepika Phukan, Mukesh Ranjan & L. K. Dwivedi, (2018), Impact of timing of breastfeeding initiation on neonatal mortality in India, DOI https://doi.org/10.1186/s13006-018-0162-0, URL https://internationalbreastfeedingjournal.biomedcentral.com/articles/10.1186/s13006-018-0162-0#rightslink

[3] Diane Coffey, (2019), The association between neonatal death and facility birth in regions of India, DOI 10.4054/DemRes.2019.40.16, URL https://www.demographic-research.org/volumes/vol40/16/default.htm

[4] Ravi Prakash Upadhyay, Palanivel Chinnakali, Oluwakemi Odukoya, Kapil Yadav, Smita Sinha, S. A. Rizwan, Shailaja Daral, Vinoth G. Chellaiyan, and Vijay Silan, (2012), High Neonatal Mortality Rates in Rural India: What Options to Explore?, DOI https://doi.org/10.5402/2012/968921, URL https://www.hindawi.com/journals/isrn/2012/968921/#conclusions

[5] M J Sankar, S B Neogi, J Sharma, M Chauhan, R Srivastava, P K Prabhakar, A Khera, R Kumar, S Zodpey, and V K Paul, (2016), State of newborn health in India, DOI 10.1038/jp.2016.183, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5144119/

[6] George Toscano; Gahangir Hossain, (2018), Predicting the Effect of Parental Education and Income on Infant Mortality Through Statistical Learning, DOI 10.1109/ICDIS.2018.00022, URL https://ieeexplore.ieee.org/document/8367746/authors#authors

[7] Dajie Zhou; Monique Frize, (2006), Predicting Probability of Mortality in the Neonatal Intensive Care Unit, DOI 10.1109/IEMBS.2006.260771, URL https://ieeexplore.ieee.org/document/4462254

[8] Sergey Soshnikov, Carl Lee, Sergey Vladimirov, (2013), A modeling approach to identify factors associated with infant mortality in Russia, DOI 10.1109/ICIS.2013.6607838, URL https://ieeexplore.ieee.org/document/6607838

[9] Ronaldo Ramos; Cristiano Silva; Mario W. L. Moreira; Joel J. P. C. Rodrigues; Mauro Oliveira; Odorico Monteiro (2017), Using predictive classifiers to prevent infant mortality in the Brazilian northeast, DOI 10.1109/HealthCom.2017.8210811, URL https://ieeexplore.ieee.org/document/8210811/references#references

[10] Mst Lucky Sultana Lovely; Mst Irin Sultana; Md Mahmudul Hasan, (2019), Building Predictive Models for Infant Mortality Rate in Bangladesh, DOI 10.1109/ICASERT.2019.8934487, URL https://ieeexplore.ieee.org/document/8934487

[11] Farida Nur Hayati; Purhadi; Bambang Widjanarko Otok, (2018), Predicting the Effect of Parental Education and Income on Infant Mortality Through Statistical Learning, DOI 10.1109/ICDIS.2018.00022, URL https://ieeexplore.ieee.org/document/8367746/authors#authors

[12] Dajie Zhou; Monique Frize, (2006), Predicting Probability of Mortality in the Neonatal Intensive Care Unit, DOI 10.1109/IEMBS.2006.260771, URL https://ieeexplore.ieee.org/document/4462254

[13] Sergey Soshnikov, Carl Lee, Sergey Vladimirov, (2013), Parameter Estimation and Statistical Test of Mixed Geographically Weighted Bivariate Weibull Regression (MGWBWR): On the Cases of Infant Mortality and Maternal Mortality Rate in East Java 2016, DOI 10.1109/SAIN.2018.8673358, URL https://ieeexplore.ieee.org/document/8673358

[14] Shelly Gupta; Shailendra Narayan Singh; Dharminder Kumar (2016), An empirical analysis of maternal health data: A case study of India, DOI 10.1109/NGCT.2016.7877465, URL https://ieeexplore.ieee.org/document/7877465/metrics#metrics

[15] Neetu Chandra Sharma (2019), Government launches SUMAN scheme, assures free medicines for pregnant women, Live Mint[online], 10th October 2019 Available at: https://www.livemint.com/news/india/government-launches-suman-scheme-assures-free-medicines-for-pregnant-women-11570704991082.html [Accessed: 8th August 2020]

[16] Subnational mapping of under-5 and neonatal mortality trends in India: the Global Burden of Disease Study 2000–17(2020), The Lancet [online], v. 395, p.1640-1658
Available at: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30471-2/fulltext#%20
[Accessed: 8th August 2020]

[17] Arjun Mukherjee (2020), UN report on under-5 mortality: India one of the worst in gender disparities, Financial Express [online], January 17, 2020
Available at: https://www.financialexpress.com/opinion/un-report-on-under-5-mortality-india-one-of-the-worst-in-gender-disparities/1825685/
[Accessed: 8th August 2020]
[18] Mary E John(2020), Poverty, not age of marriage, is responsible for women's poor health indicators, The Indian Express [online], 31st July 2020
Available at: https://indianexpress.com/article/opinion/columns/women-health-index-india-marriage-age-6531515/
[Accessed: 8th August 2020]
[19] Ajit Kumar Yadav, F Ram (2019), There Is a Glaring Gender Bias in Death Registrations in India, Vol. 54, Issue No. 51
Available at: https://www.epw.in/engage/article/there-glaring-gender-bias-death-registrations
[Accessed: 8th August 2020]
[20] How Kerala reduced its infant mortality rate to 7 per 1000 (2020), The News Minute [online], 12th May 2020
Available at: https://www.thenewsminute.com/article/how-kerala-reduced-its-infant-mortality-rate-7-1000-124393
[Accessed: 8th August 2020]

APPENDIX A

RESEARCH PROPOSAL

1. Introduction

The mortality trend has always been unstable in India, although it shows a reduction with years. In 1990, the expected mortality was 86.4 deaths per 1000 live births and the actual value came out to be 80.7 deaths per 1000 live births among kids below 1 year age group in India. But later on in 2017, the expected mortality deaths were 34.3 deaths per 1000 live births and the actual value came out to be 36 deaths per 1000 live births. So it is clear that there have been more complications for below one year groups that expected in 2017. The mortality among children can be seen to differ from state to state. In Kerala, the infant mortality rate is 12 per 1000 population as compared to the rate is 35 per 1000 population in Bihar.

The fertility rate in India has constantly reduced. In 1990, the fertility rate in India was 4.04 births per woman as compared to 2017 when it has come to 2.27 births per woman. In Kerala, the fertility rate is 2.35 births per woman as compared to the rate of 3.6 births per woman in Bihar.

So, it is clear that in the two states shown above the fertility rate is almost similar but there is a huge difference in infant mortality rate.

According to Institute for Health Metrics and Evaluation, the ranking of most death caused in India by neonatal disorders have reduced from 4th to 7th from 2007 to 2017 but it is still one among the top ten.

2. Details of the research project

i. Background and related research

'State of new-born health in India' (2016) - it focused on factors like rural-urban, poor–rich, facility availability, and gender differentials. The outcome was that there was a requirement for improvement in the health facilities and the introduction of new interventions.

'High Neonatal Mortality Rates in Rural India: What Options to Explore?' (2012)- The focus was on the rural areas and the main factor that was looked into was the effects on the neonatal mortality rate depending on birth locations. The conclusion states that there needed to be large scale projects that will help promote home-based new-born care, creating community awareness and community mobilization along with strengthening public-private partnerships.

‘The association between neonatal death and facility birth in regions of India’ (2019) – The objective of this research article was to find out if there is a relation between neonatal deaths and facility births in different states. It turns out that there was a relation in few states like Uttar Pradesh and Bihar, but was not as robust in other states.

‘Impact of timing of breastfeeding initiation on neonatal mortality in India’ (2018) – The main aim of the study is to examine the timing of initiation of breastfeeding and neonatal deaths. It was found that timely initiation of breastfeeding is beneficial for child survival within the first 28 days of birth in all of the cases of mortality that was looked up.

‘Mapping Neonatal Mortality in India: A Closer Look’ (2017) – This article shows how the mortality rate differs from state to state. As a conclusion, they succeeded to figure out this issue but stated that there is an urgent need for up-to-date data on district-level neonatal mortality in India.

In the past, all researches have been done regarding Neonatal deaths and related causes. But in most of them, they have looked at broad factors like residence area, gender, financial background, place of birth, facility, breastfeeding, and so on.

ii. Aim

This research aims to figure out the factors affecting mortality and fertility. Later on, it will be ranked to know which of them are most prominent.

iii. Objective

The objective will be to find the correlation of mortality with many factors like mothers’ age, marital status, religious community, educational level, occupation along with the number of deaths caused; gender, surviving rate, and so on of the child in different years. The relation between fertility rate and infant mortality rate at the state level and national level will also be found.

iv. Research Questions

- What are the major factors that affect the mortality rate in infants?

This question is important because infant mortality has been one of the top 10 reasons for death in India. This states that it is one of the major issues. Finding the factors and by how much they

affect this issue we can try to solve this problem by coming up with new ideas that can be implemented.

- By how much does the mortality rate in infants differ from state to state?

This is very important to investigate since there has been a huge difference in mortality rate in infants from state to state even when the fertility rate remains almost the same. This shows that just figuring out the factors at the national level is not enough but rather needs to be investigated into the state level as well.

v. Requirements and resources

The data is collected from the India government census of 2001 and 2011. Further the data 'Annual Health Survivingey: Mortality Schedule' includes the major states in India provided by the 'Ministry of Health and Family Welfare' and 'Department of Health and Family Welfare' (2007 – 2011).

Python and Microsoft Excel for collecting data and doing the analysis. Finally, tableau and Power BI desktop will be used to do data visualization.

vi. Research methodology

The data that will be used for the research will be the census taken by the Indian government in 2001 and 2011. It comes under the category of Fertility Data. Future the data Annual Health Survivingey: Mortality Schedule which includes the major states in India provided by the Ministry of Health and Family Welfare and Department of Health and Family Welfare (2007 – 2011) will also be used. These data are collected from the official website of the Indian Government Census and Open Government Data (OGD) Platform India.

The collected data includes multiple data sets with many variables. It will be merged based on relevant variables that are common to form one final dataset. Treating of missing values will be by replacing with 'not available' since if the data is missing it means that variable is not relevant there.

The next step will be the Exploratory Data Analysis (EDA). This will include multiple steps including univariate analysis, segmented analysis, bivariate analysis, and a few derived metric analysis. These will be done using pivoted tables and graphs. The univariate analysis will help in knowing each factor in detail. The segmented analysis and bivariate will provide the correlation between different variables. And the derived metrics will be used to understand and figure out the data in a more concise manner.

Finally using Linear regression, Random Forest and Clustering can be used to rate the factors to find out which are the most important ones and how to tackle them. For all this, the software used will be Python along with packages provided.

3. Expected outcome

The desired outcome will be to figure out the factors that affect mortality and fertility rates and also to find out the top reasons for the same. This will later be concluded with methods to be implemented for a better chance of survival of young children.

APPENDIX B

ETHICS FORMS

The data used in this research is publically available.