

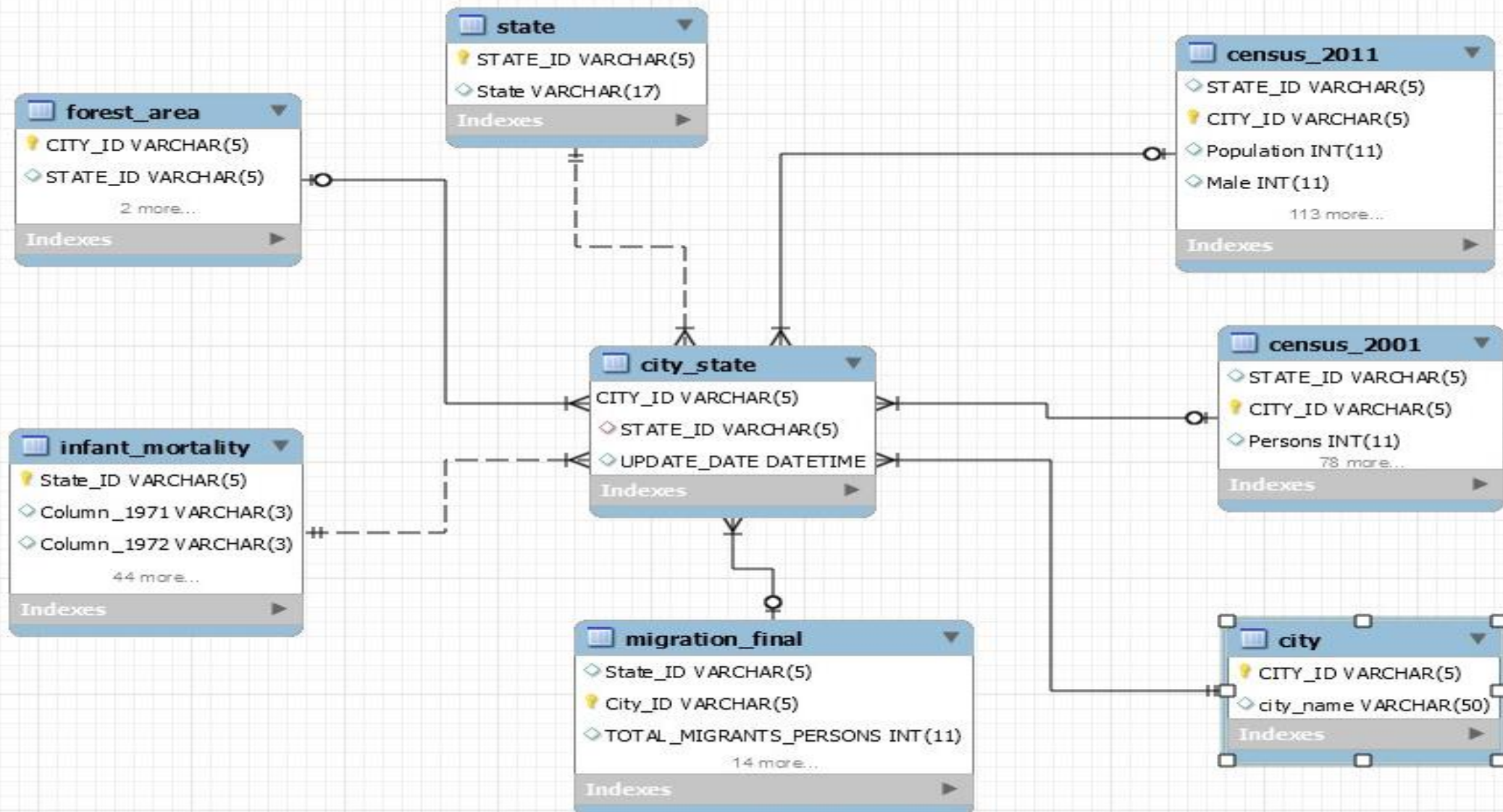
# Demographic Analysis of India

Tech Ninjas -  
Abhishek, Arpit,  
Bhupender and Shishir

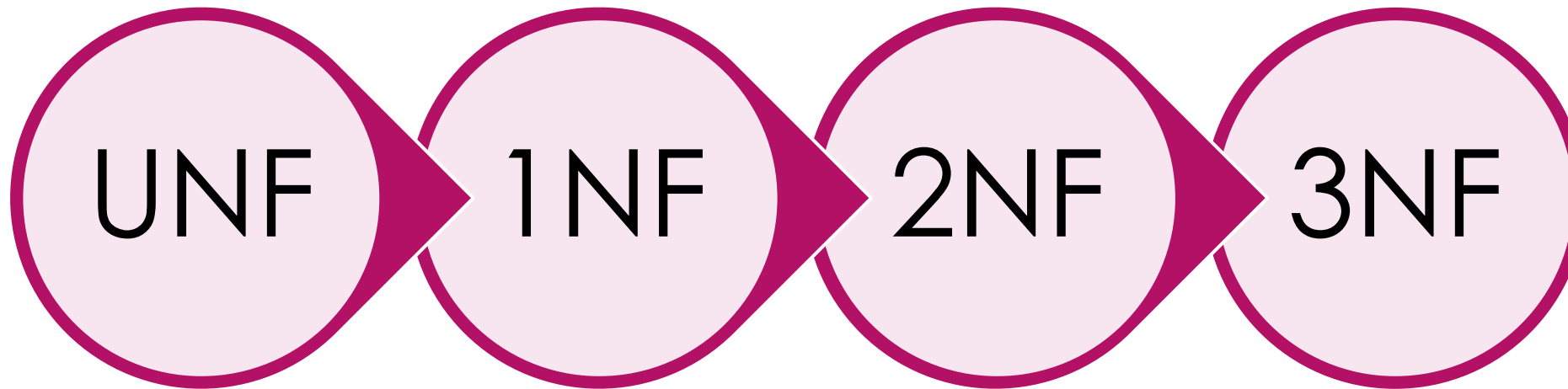
THE OXFORD DICTIONARY OF ECONOMICS DEFINES DEMOGRAPHY AS **“THE STUDY OF THE CHARACTERISTICS OF HUMAN POPULATIONS.”**

FOR OUR PROJECT ,WE ARE SHOWCASING AND ANALYZING SOME OF THE PRESSING DEMOGRAPHIC ASPECTS LIKE POPULATION DENSITY AND RELATED TOPICS LIKE BIRTH RATE AND DEATH RATE, HIGH DENSITY DUE TO INDUSTRIALIZATION AND DEFORESTATION. WE WILL ALSO ADDRESS SOME OF THE SOCIO ECONOMIC ISSUE LIKE ILLITERACY RATE, ILLEGAL IMMIGRATION AND MORTALITY RATE.

# Demographic: Entity Relationship Diagram



# Normalization Process



- Base data was downloaded in UNF

- No Repeating groups
- Data is Atomic
- Each Field is unique
- Primary Key

- 1<sup>st</sup> Normal Form
- All Non key attributes are dependent on all parts of primary key
- Eliminated Partial Dependencies

- 1<sup>st</sup> Normal Form
- 2<sup>nd</sup> Normal Form
- Eliminated transitive dependencies



Normalization

# Introduction – Business Need

The Purpose of this database is to act as a datamart for the metrics, dimensions and facts related to Demographics of India.

We chose this dataset for few different reasons:

- 1) To show population growth over last 20 years and how it has affected forest cover, mortality and gender diversity.
- 2) To find out the reason for Population Migration in India
- 3) To find out the most populated city for every state and tried to infer which cities can be our next metropolitan city as per development.
- 4) To find out Female Literacy growth and how women empowerment and awareness has increased over the years.

# Extract – Transform - Load

## Data Sources

- Census 2001, Census 2011, Forest Cover, Migration, Mortality

## Preparation

- Removed State names and City names from the base tables
- Added primary Keys as City\_id and state\_id to all the base tables using vlookup and fuzzy lookup algorithm within excel for normalization.

## Clean-up

- Used OpenRefine, Excel and MySQL to remove spaces and special characters from state and city names
- Used Fuzzy lookup to do the Primary key mappings to all the tables.

## Transformation

- Large Numbers were converted to "M"(millions) and "T" (Thousands)
- Percentages were calculated and stored in MySQL views

## Loading

- We have used MySQL insert statements and SQLizer tool to load data to the base tables.

## Lessons

- Update statements could only be executed with SQL\_Safe\_Updates = 0

# Create Table & Views Queries



Create SQL  
Queries



CreateView\_Popul  
ationDensity



CreateView\_Femal  
e Literacy



CreateView\_Forest  
Cover



CreateView\_infant  
\_mortality



CreateView\_Popul  
ationGrowth



CreateView\_topcit  
ybypop\_eachstate

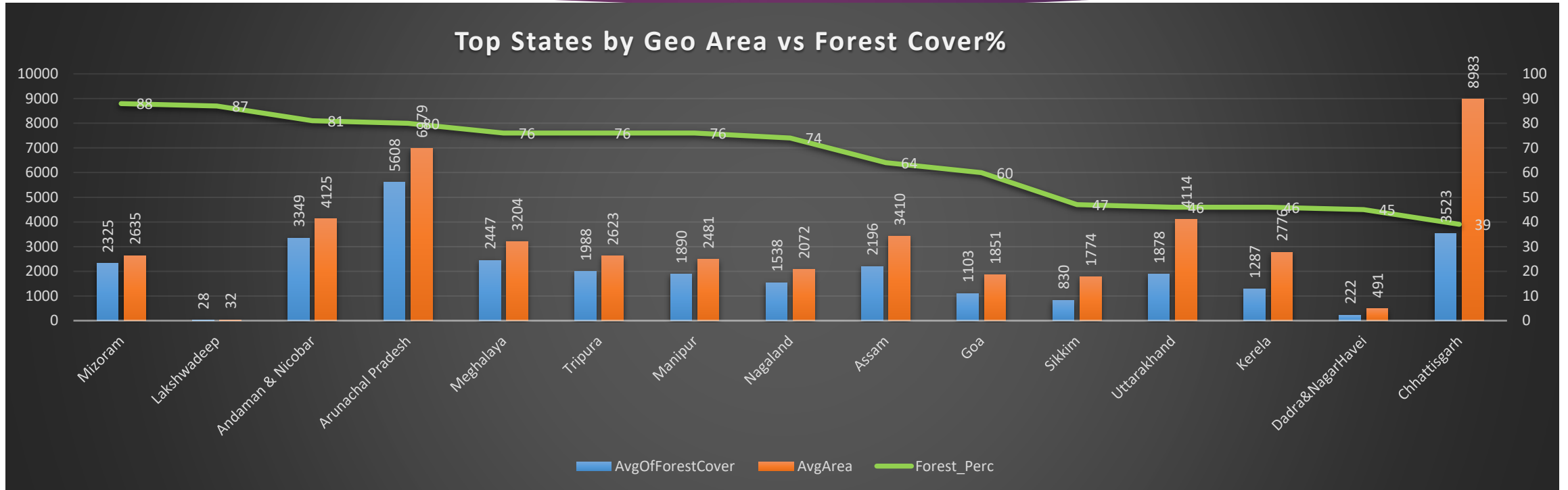


CreateView\_topst  
byareaandforestcc



CreateView\_Migra  
tion

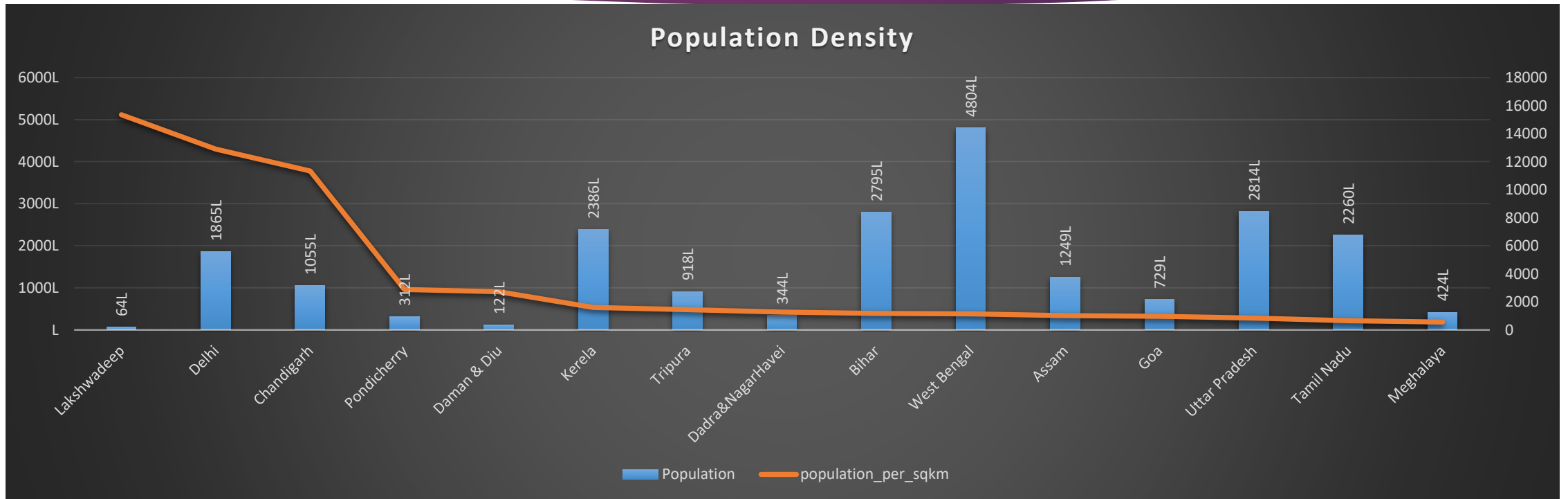
# Query1: Top States by Geographical area and Forest Area % #GoGreen



This chart shows area vs forest cover distribution of the top 15 states in India. Northeast states (7 sisters) and UT's clearly have highest forest covers in India.

# Query2: States by Population Density

## #PopulationExplosion

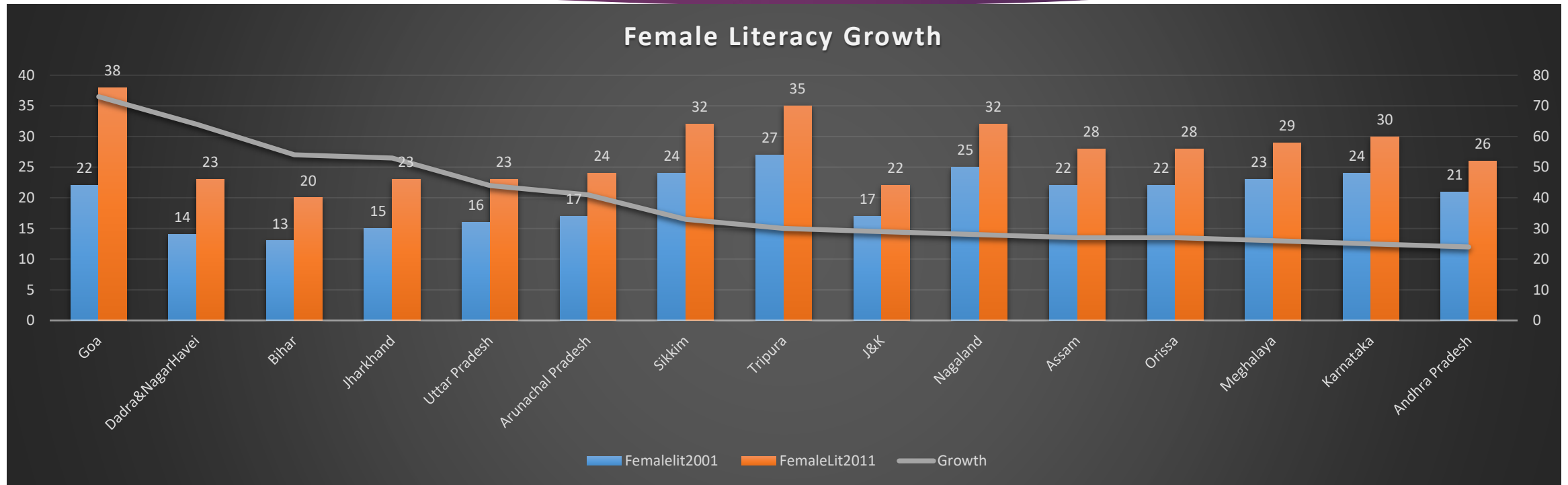


This chart shows population density of states in lakhs and their population per sq kms. Density is shown removing the forest area from the overall land. This shows that UT's have high population density.



# Query3: Female Literacy growth

## #Beti Bachao, Beti Padhao

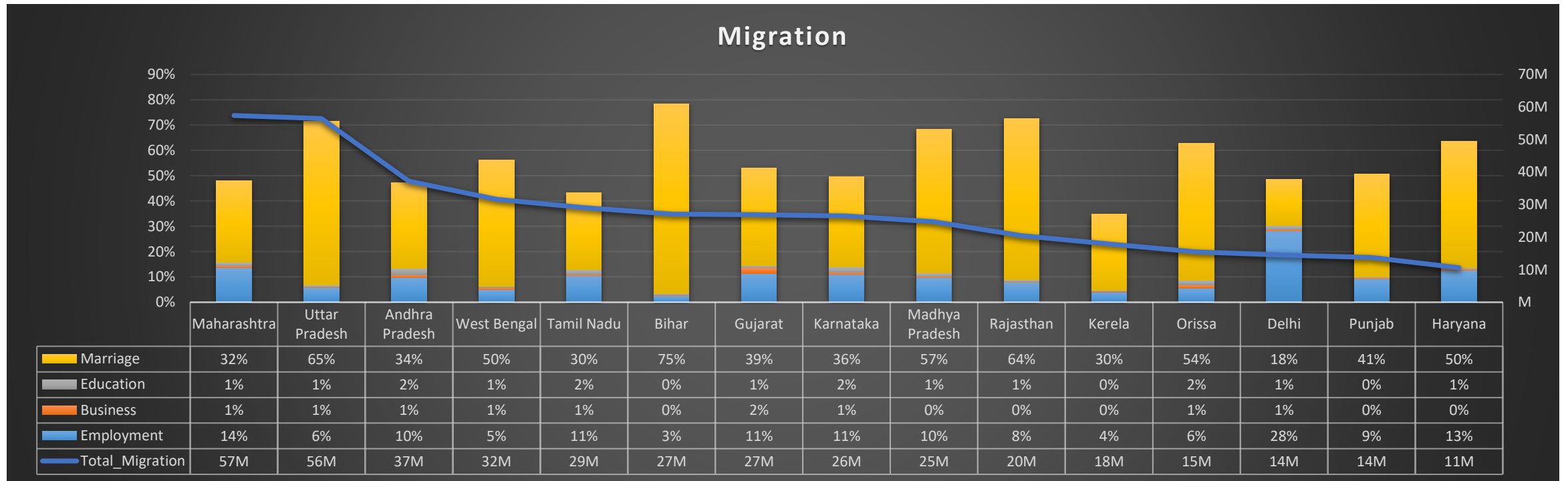


We have compared 2 census data from 2001 and 2011 and could show that how female literacy has grown in those 20 years.

Goa, Dadra&NagarHaveli, Bihar and Jharkhand have increasing trends in terms of female education.

# Query4: Migration in India

## #NRC & CAA



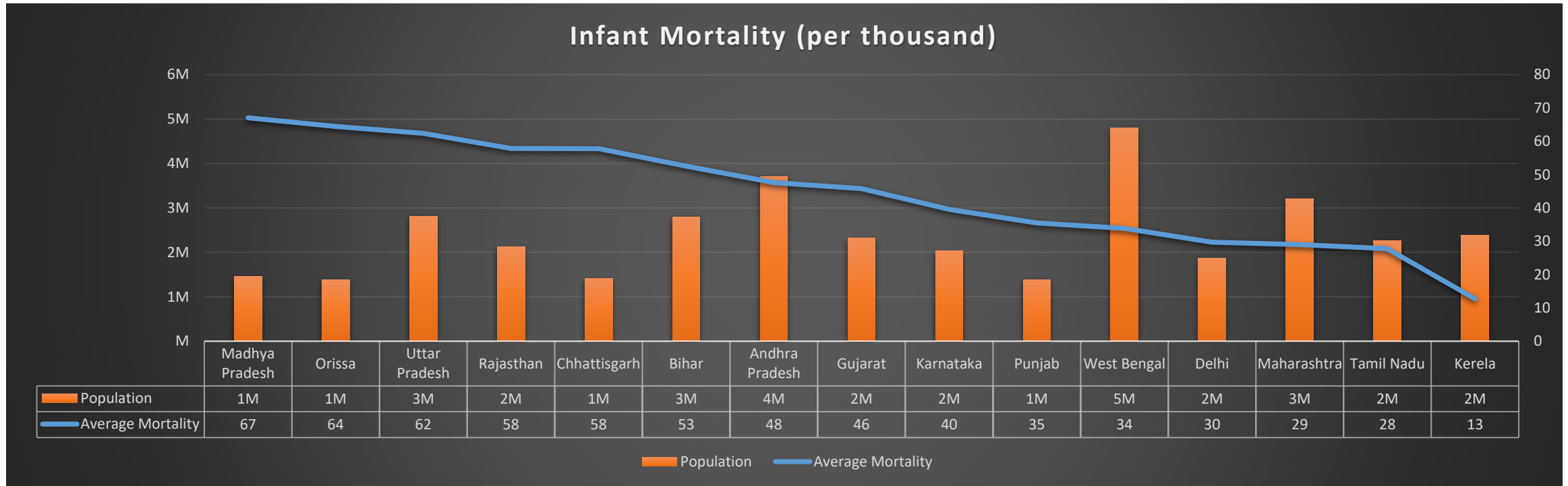
This graph shows states with highest migration counts with reason breakup.

We could infer from this graph that Marriage and Employment are the major reasons people migrate to India.

Maharashtra and UP have registered the highest migrants amongst all the states.

# Query5: Infant Mortality

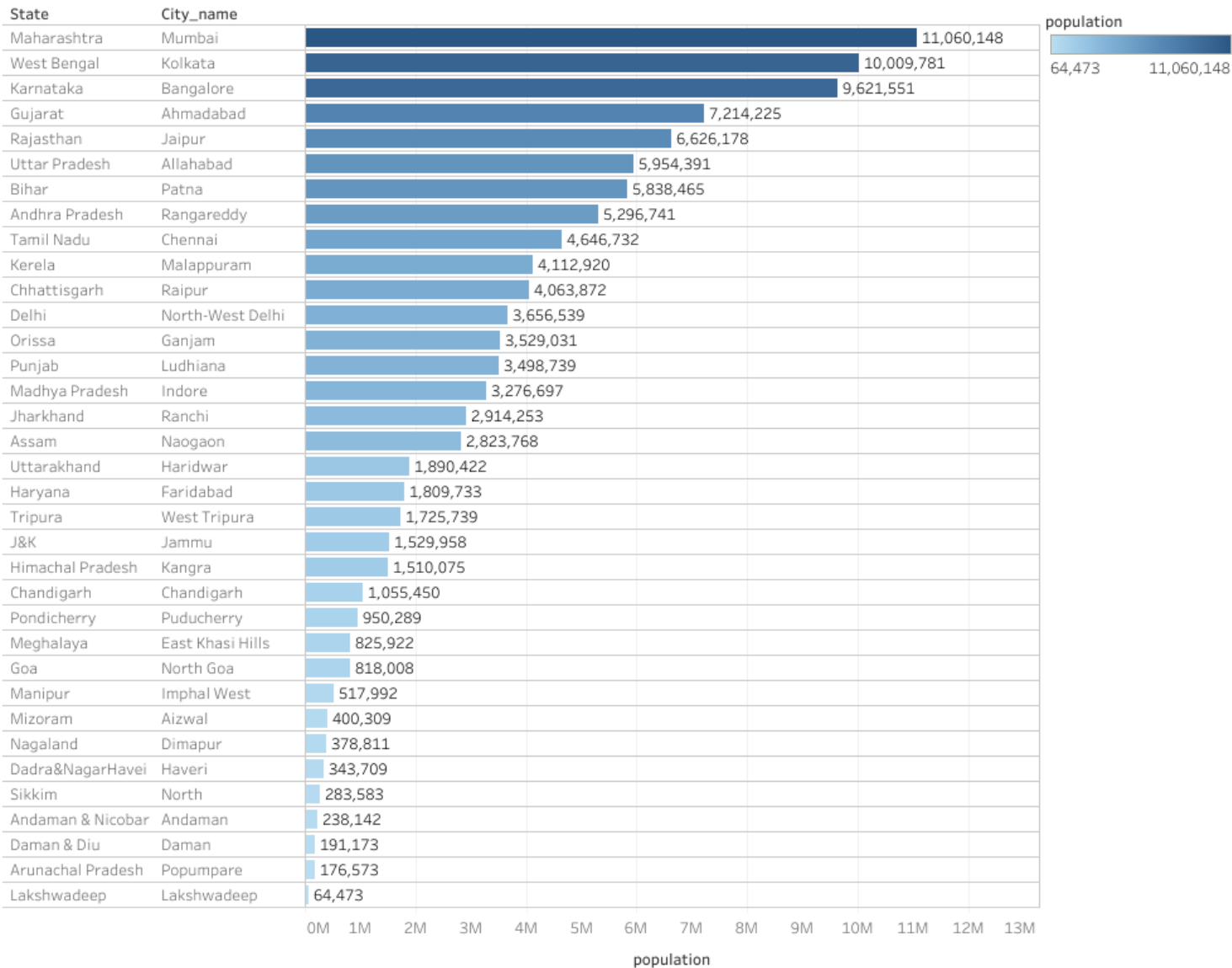
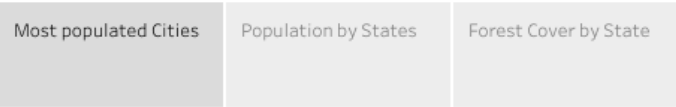
#Today's children are tomorrow's helmsman



This graph shows states with highest population counts with infant mortality rate per thousand. We could infer from this graph that population and Infant mortality have no direct correlation. MP, Orissa have relatively less population than TN and kerala but still have a high infant mortality rate.

Tableau Integration –  
States with their most  
populated cities

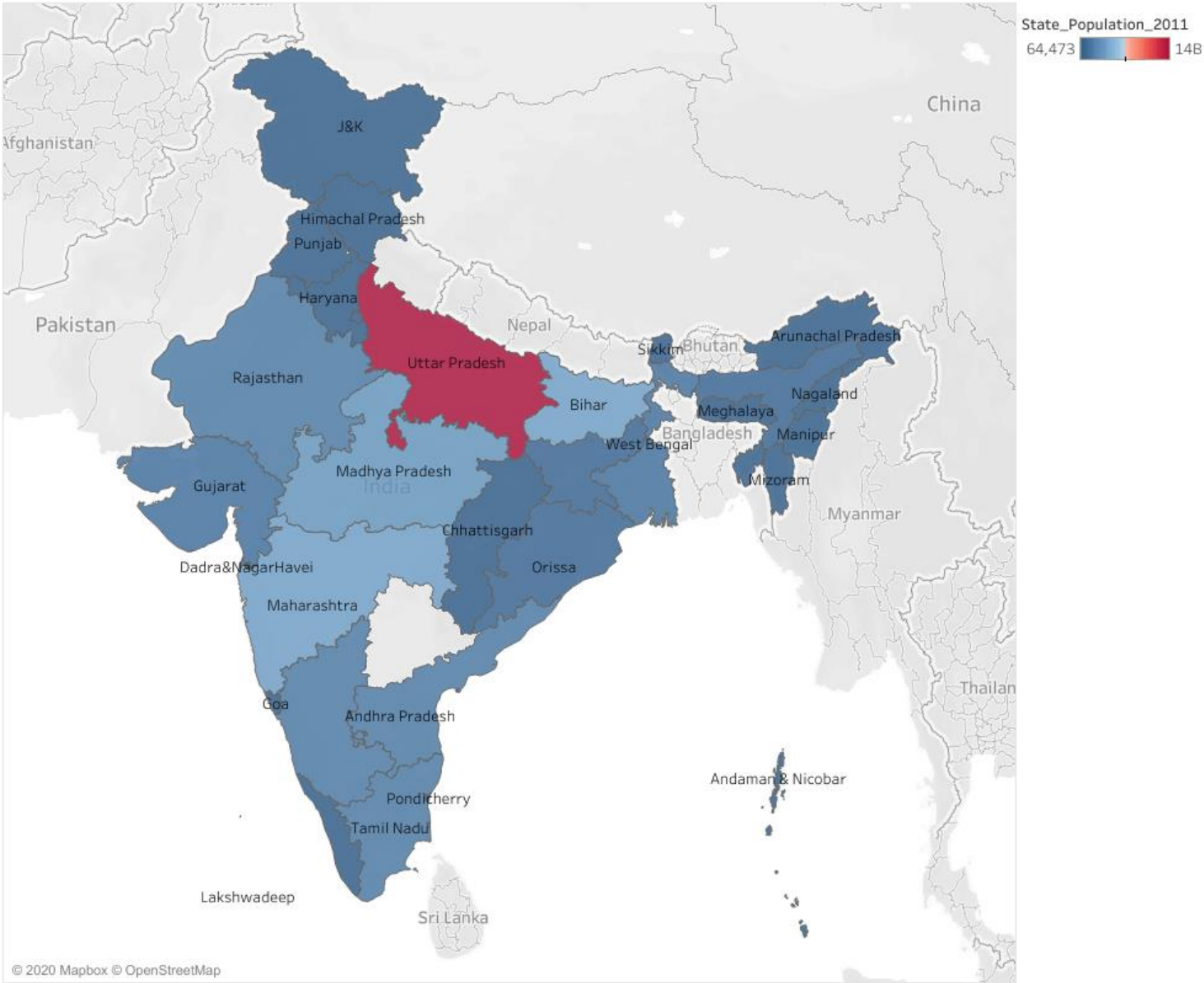
Demographic Analysis



# Tableau Integration – Population by State

## Demographic Analysis

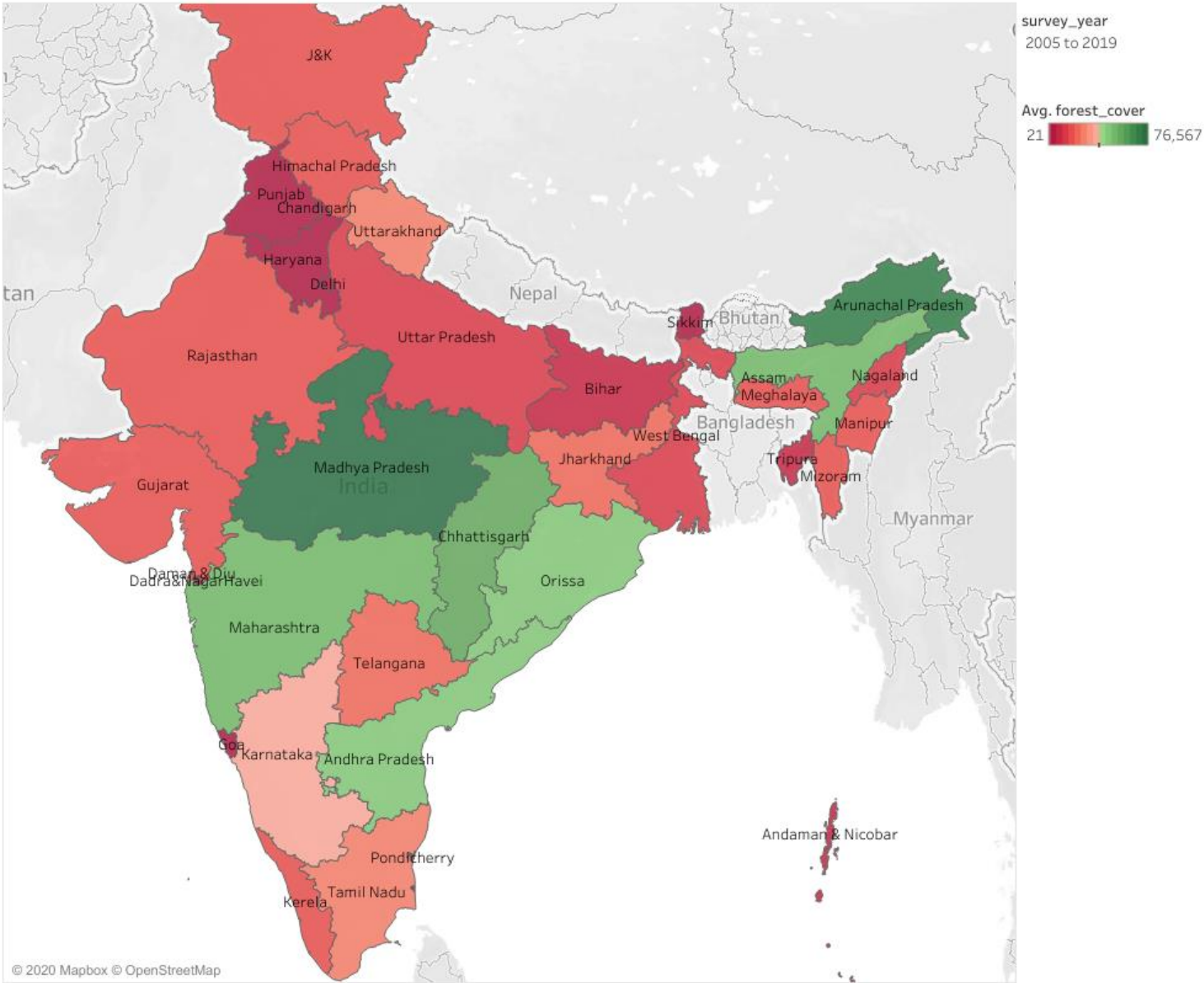
Most populated Cities	Population by States	Forest Cover by State
-----------------------	----------------------	-----------------------



# Tableau Integration - Forest Cover by State

## Demographic Analysis

Most populated Cities	Population by States	Forest Cover by State
-----------------------	----------------------	-----------------------



# Insights, Corrective measures and recommendations from Data Analysis

- ▶ North-East states have highest forest covers in India and needs to be preserved with the help of various green initiatives to help reduce carbon footprints.
- ▶ Delhi, Haryana and Punjab has highest populations and lowest forest covers. More and more tree plantation drives in these states would help reduce pollution as well as improve life expectancy.
- ▶ Union Territories have higher population densities than other states. Measures can be taken to migrate people to other states by means of creating employment opportunities and livelihood to help reduce population concentration.
- ▶ Some states like Goa & Bihar has shown steep growth in the female literacy rates. Best practices from these states can be learnt and implemented across other states to help improve this metric across all the states in India
- ▶ Marriage and Employment are the major reasons people migrate to India. Maharashtra and UP have registered the highest migrants amongst all the states. Measures can be taken to distribute migrants across other states to help reduce population concentration.
- ▶ Population and Infant mortality have no direct correlation. MP, Orissa have relatively less population than TN and Kerala but still have a high infant mortality rate. Better facilities like vaccinations, healthy and vitamin enrich food can be provided at subsidized rates for infants and expectant mothers in MP & Orissa to help decrease infant mortality rates.
- ▶ Jaipur, Ahmedabad, Patna, Prayagraj and Bengaluru could be the next metropolitan cities looking at the population density distribution.



# Lessons Learned

- ▶ Data Wrangling – City name, State name standardization across population, migration, forest cover data using open refine and MYSQL
- ▶ Learnt and used fuzzy lookup method for indexing.
- ▶ Learnt and applied database normalization
- ▶ Learnt to create views, tables using complex MYSQL queries with functions like group by, joins, in, where etc
- ▶ Connected MySQL views to excel reports/dashboards to enable auto-refresh of data.
- ▶ Learnt and used excel as well as Tableau to visualize data in an intuitive manner.
- ▶ Learnt to connect excel as a source for tableau and used geo spatial maps in tableau to visualize data for states and cities.



# Report Files



Demographics -  
Report

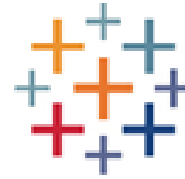


Tableau Dashboard