National Rural Drinking Water Analysis

Shreeya Bharat Nelekar
Department of Computer Science
Birla Institute of Technology *and Science, Pilani*
Hyderabad, India

2017A7PS0093H

Saarthak Jain
Department of Computer Science
Birla Institute of Technology *and Science, Pilani*
Hyderabad, India

2017A7PS0083H

Rashi Jain
Department of Computer Science
Birla Institute of Technology *and Science, Pilani*
Hyderabad, India

2017A7PS0082H

I. Introduction

There has been a significant progress in the development of water resources in India, yet problematic management issues remain despite increased funding, resource base and a vast land resource. The biggest challenge before the government is to meet the needs of increasing population. Besides this challenge, other factors affecting water supply in India include political will, environmental sustainability, social dynamics, technological appropriateness and economics. As a part of this project we aim at analysing the extent of dependencies on these parameters and the water supply trends in rural parts of India.

II. Problem Motivation

Water is one of the most important natural resources. The potential of water resources in India is such that it can fulfill the water needs of all the country. Various programmes are implemented for effective allocation of water resources. However none of them satisfy their goals. This has motivated us to understand what are the probable reasons for the partial failure of these projects and schemes, on what parameters is water resource allocation and management based on and so on.

Another reason for taking up this research is to understand the social dynamics in Indian society based on the population distribution of Indian rural population.

III. Objectives

Is The main objective of this research is to recognize the variability in the distribution of potable drinking water across rural India.

- Demographic Analysis:
  - Population distribution in rural areas. (ST, SC and General Category).
  - State-wise population distribution across India.
  - Availability of potable drinking water in rural areas (state-wise comparison).
  - Caste-wise distribution of potable drinking water.
  - Caste-wise analysis of change in the availability of potable drinking water.
  - State-wise comparison of development trend.
- Time-series Analysis:
- Analysis of Jal Jeevan Mission (launched by Central Government).

IV. Background

The Central Government assistance to States for rural water supply started in 1972 with the dispatch of Accelerated Rural Water Supply Program. It was renamed as National Rural Drinking Water Program (NRDWP) in 2009, which is a midway supported plan with finance sharing between the Center and the States. Under NRDWP, one of the goals was to "empower all families to approach and utilize safe and satisfactory drinking water inside premises to the degree conceivable". It was proposed to accomplish the objective by 2030, matching with the United Nation's Sustainable Development Goals. Be that as it may, presently, it is has been wanted to accomplish the objective by 2024 through Jal Jeevan Mission (JJM). As a part of this mission, the data collection of the progress of this mission is being carried out over the years.

V. Methodology

1) Data set Information:

The following data set is dated back to year 2009.

```
In [15]: #getting data type of each attribute
         pop_data.dtypes

Out[15]: State Name                    object
         District Name                 object
         Block Name                    object
         Panchayat Name                object
         Village Name                  object
         Habitation Name               object
         SC Current Population          int64
         ST Current Population          int64
         GENERAL Current Population     int64
         SC Covered Population          int64
         ST Covered Population          int64
         GENERAL Covered Population     int64
         Status                        object
         SC Concentrated               object
         ST Concentrated               object
         Year                          object
         dtype: object
```

Fig. 1. data types of the attributes of the original data set

```
State Name                                    ANDHRA PRADESH
District Name                                 SRIKAKULAM(01)
Block Name                                       VANGARA(02)
Panchayat Name                                V.V.R.PETA(01)
Village Name                                 J.K.GUMADA(001 )
Habitation Name              J.K.GUMADA (0101302001010100)
SC Current Population                                   128
ST Current Population                                    15
GENERAL Current Population                              595
SC Covered Population                                   128
ST Covered Population                                    15
GENERAL Covered Population                              595
Status                                       Fully Covered
SC Concentrated                                         NO
ST Concentrated                                         NO
Year                                            01_04_2009
Name: 0, dtype: object
```

Fig. 2. A sample data point.Data Preprocessing:

1. Data Cleaning:

The raw data contained around 16 lac data points out of which only 15 entries contained missing values. As this ratio is insignificant, we dropped these data points from the data set, as a data cleaning step.

2. Feature Creation:

Newly created attributes :

1. Total Population : It comprises of the SC,ST and GENERAL current populations in a habitation. The main objective was to ease out the calculations for visualizations and further preprocessing.

2.Latitude and Longitude values for every Block were obtained for getting better insights using geographical visualizations. Geopy library was used to obtain these values.

3.Binarization:

Attributes such as

1.SC Concentrated(YES/NO)

2.ST Concentrated(YES/NO)

3.Status (FULLY COVERED/PARTIALLY COVERED)

were binarized into ones and zeros with zeros denoting the latter values.

4.Correlation Analysis:

The dependencies between various attributes amongst one another was captured using correlation matrix.

5.Normalization:

The data points pertaining to population statistics were normalized using Z-Score Normalization wherein the mean and variance calculated State wise.

6.Stratified Sampling:

As the original dataset was enormous containing over 16 lac values. As working on such a huge dataset was computationally difficult,the dataset size had to be reduced to one-fourth of the original size while also minimizing the information loss. Therefore stratified sampling was implemented wherein each bin corresponded to each state and the sampling percentage for each bin was proportional to the ratio of State : Country population.

7.Aggregation:

Another measure of data reduction for ease of visualization. Data was aggregated where the populations of Panchayat,Village and Habitations were merged and only block wise population was considered.

8.Data Slicing:

Dataset was divided into Training(80%) and Testing data(20%) for data analysis using regression.

VI. RESULTS

1)Visualizations

1.Pie-Chart

For getting an idea about the caste wise population across India as well each state. Also distribution of potable water for the same was plotted.

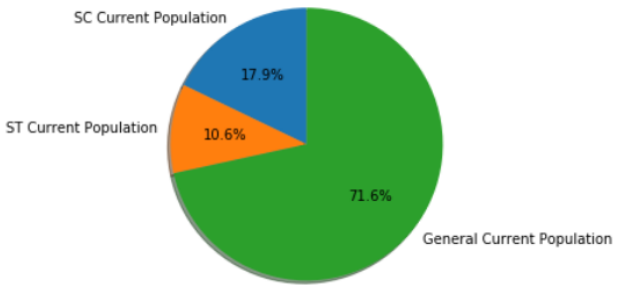


Fig. 3. Caste wise Population Distribution in India

2.JOINT PIE-CHART

To get an idea about the percentage of uncovered population in each caste

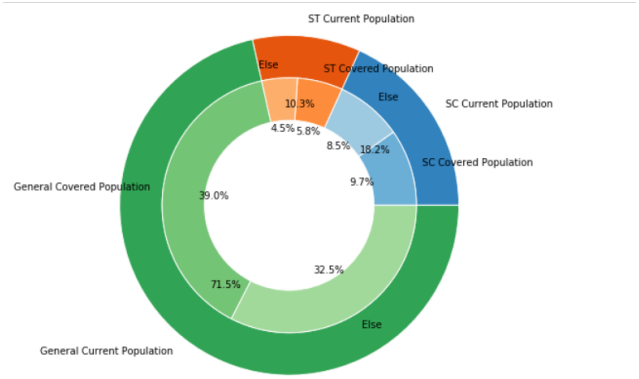
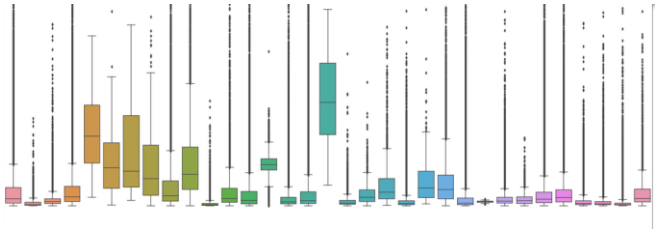


Fig. 4. Current vs Covered Caste Wise Population Distribution

3.BOX PLOT

State wise population statistics were obtained.

Fig. 5. Box Plot for state population



4.JOINT BAR GRAPHS

The percentage of fully covered and partially covered populations in each state drawn on a joint bar graph.

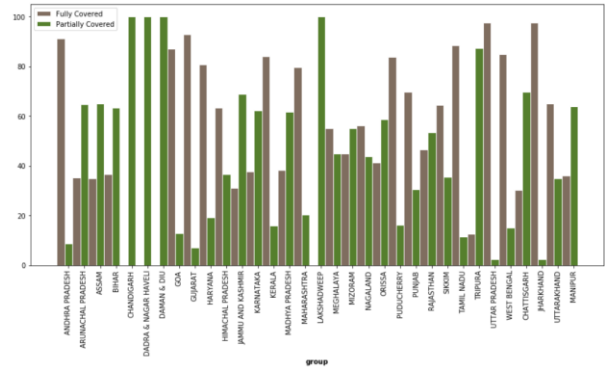


Fig. 6. Statewise Potable Water Coverage in India

5.SCATTER PLOTS

To determine what are the characteristics for classifying a region as SC/ST concentrated as well as there coverage status as being fully or partially covered.

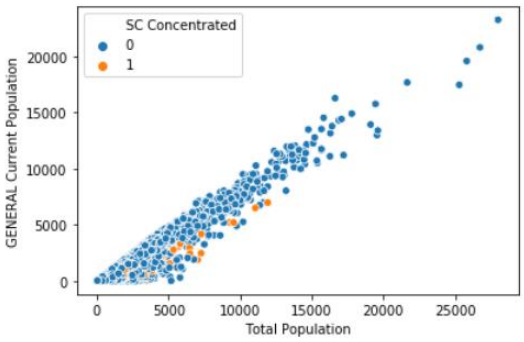
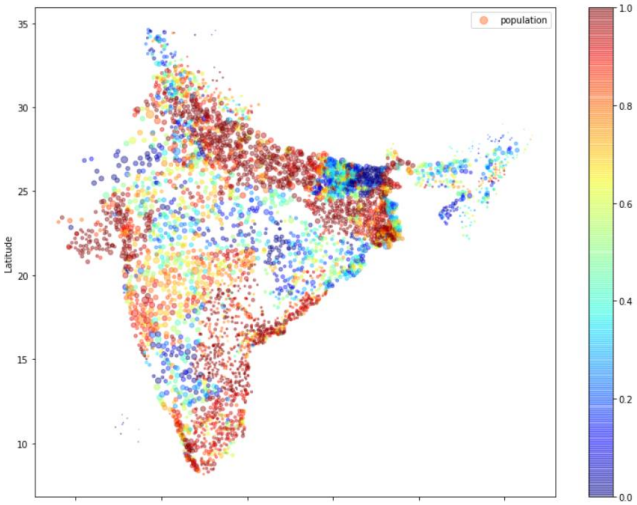


Fig. 7. General Population vs Total Population for Rajasthan

6.GEOGRAPHICAL SCATTER PLOT

Geographical Visualization of the coverage status across India.



8.The ST concentrated areas have total population less than 10,000 irrespective of the state.

CONCLUSIONS/DISCUSSION

1.Almost 46% percent of habitations is only partially covered with potable drinking water supply.

2.From the initial observations,there is no bias amongst the castes as to water supply coverage.

3.More developed states have larger coverage of water supply as is clear from The geographical scatter plot,however Karnataka is an exception.

4.All the Union Territories except Pondicherry are only partially covered with potable water supply.In further data analysis reason behind this can be explored.

5.In some states like Jammu and Kashmir,Karnataka,ST concentrated areas are not fully covered.

6.It is observed that General and ST population are demographically separated in the north-eastern states.

7.Generally ST concentrated areas are observed mostly along the borders of the states while the SC population is evenly spread

BIBLIOGRAPHY

- pandas.pydata.org
- data.gov.in
- ieee.org