## INTRODUCTION TO DATA MANAGEMENT PROJECT REPORT

(Project Semester August-December 2019)



**Health Analysis on Empowered Action Group (EAG) states of India**

Submitted by

Harshit Anand

Registration No. 11704351 Section: KM064

Course Code: INT217

Under the Guidance of

## Assistant Professor Maneet Kaur UID: 15709

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

**INDEX**

**Contents Page No**

1. Certificate 3
2. Declaration 4
3. Acknowledgement 5
4. Introduction 6
5. The dataset contains the following features 8
6. Objectives/Scope of the Analysis 9
7. Source of dataset 9
8. ETL Process 10
9. Analysis on Dataset 16

* State wise population analysis 16
* District wise population data analysis 17
* District wise sex ratio data analysis 18
* District wise marriage below legal age data analysis 19
* District wise literacy rate data analysis 21
* District wise child suffering from disease data analysis 23
* District wise Institutional delivery data analysis 25
* District wise crude death and birth rate data analysis 27
* District wise birth registration data analysis 29
* District wise people chronic disease data analysis 31
* District wise schooling status data analysis 32
* Dashboard 34

1. Conclusion 35
2. Bibliography 35

**CERTIFICATE**

This is to certify that Harshit Anandbearing Registration no. **11704351** has completed **INT217** project titled, **“Health Analysis on Empowered Action Group (EAG) states of India”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

## Miss Maneet Kaur Assistant Professor

**School of Computer Science and Engineering**

Lovely Professional University Phagwara, Punjab.

Date: 16-11-2019

## DECLARATION

I, Harshit Anand, student of Computer Science and Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 16-11-2019

Registration: 11704351

Harshit Anand

## Acknowledgement

I would like to express my special thanks of gratitude to my teacher Miss Maneet Kaur who gave me the golden opportunity to do this wonderful project on the topic Health Analysis on Empowered Action Group (EAG) states of India which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to them.

Secondly I would also like to thank friends who helped me a lot in finalizing this project within the limited time frame.

Introduction

The survey was conducted in Empowered Action Group (EAG) states Uttarakhand, Rajasthan, Uttar Pradesh, Bihar, Jharkhand, Odisha, Chhattisgarh & Madhya Pradesh and Assam. District Level Factsheet–2010-11 was released by Shri P. K. Pradhan, Secretary, Health at New Delhi today.  
  
Evidence based planning requires disaggregated data at the District Level. In the absence of vital data at the district level, State level estimates are being used for formulating district plans and setting the milestones. In the process, the hotspots (districts requiring special attention) very often get masked by the State average. This statistical fallacy compounds the problems of the districts acutely, more so in the health sector.  
  
To overcome this problem, the Annual Health Survey (AHS) was conceived during a meeting of the National Commission of Population held in 2005 under the Chairmanship of the Prime Minister wherein it was decided that “there should be an Annual Health Survey of all districts which could be published/monitored and compared against benchmarks”. The objective of the AHS is to yield a comprehensive, representative and reliable dataset on core vital indicators including composite ones like Infant Mortality Rate, Maternal Mortality Ratio and Total Fertility Rate along with their co-variates (process and outcome indicators) at the district level and map the changes therein on an annual basis. These benchmarks would help in better and holistic understanding and timely monitoring of various determinants on well-being and health of population particularly Reproductive and Child Health.  
  
Realizing the need for preparing a comprehensive district health profile on key parameters based on a community set up, the AHS has been designed to yield benchmarks of core vital and health indicators at the district level on fertility and mortality; prevalence of disabilities, injuries, acute and chronic illness and access to health care for these morbidities; and access to maternal, child health and family planning services.  
  
Annual Health Survey (AHS) is implemented by the Office of Registrar General, India in all the 284 districts (as per 2001 Census) in 8 Empowered Action Group (EAG) States (Bihar, Jharkhand, Uttar Pradesh, Uttarakhand, Madhya Pradesh, Chhattisgarh, Orissa and Rajasthan) and Assam for a three year period (i.e., a Base-line Survey followed by two updation surveys) spread over 2010-11 to 2012-13. These nine States, which account for about 48 percent of the total population, 59 percent of Births, 70 percent of Infant Deaths, 75 percent of Under 5 Deaths and 62 percent of Maternal Deaths in the country, are the high focus States in view of their relatively higher fertility and mortality. A representative sample of about 20.1 million population and 4.1 million households were covered in 20,694 statistically selected PSUs (Census Enumeration Blocks in case of urban areas and villages or a segment thereof in rural areas) in these 9 AHS States during baseline and to be followed every year. With the present coverage, the AHS is the largest demographic survey in the world and is two and half times that of the Sample Registration System.

In India, the eight socioeconomically backward states of Bihar, Chhattisgarh, Jharkhand, Madhya Pradesh, Orissa, Rajasthan, Uttaranchal and Uttar Pradesh, referred to as the Empowered Action Group (EAG) states, lag behind in the demographic transition and have the highest infant mortality rates in the country. Neonatal mortality constitutes about 60% of the total infant mortality in India and is highest in the EAG states. This study assesses the levels and trends in neonatal mortality in the EAG states and examines the impact of bio-demographic compared with health care determinants on neonatal mortality. Data from India’s Sample Registration System (SRS) and National Family and Health Survey (NFHS-2, 1998–99) are used. Cox proportional hazard models are applied to estimate adjusted neonatal mortality rates by health care, bio-demographic and socioeconomic determinants. Variations in neonatal mortality by these determinants suggest that universal coverage of all pregnant women with full antenatal care, providing assistance at delivery and postnatal care including emergency care are critical inputs for achieving a reduction in neonatal mortality. Health interventions are also required that focus on curtailing the high risk of neonatal deaths arising from the mothers’ younger age at childbirth, low birth weight of children and higher order births with short birth intervals.  
  
The fieldwork for Baseline Survey was carried out during July, 2010 to March, 2011. In all, 161 indicators are available from AHS Baseline. In the first phase of dissemination, 9 Core Vital Indicators, viz., crude birth rate, crude death rate, natural growth rate, infant mortality rate, neo-natal and post neo-natal mortality rate, under 5 mortality rate, sex ratio at birth, sex ratio (0-4 years) and overall sex ratio have already been released.  
  
In this phase of dissemination, data on remaining 152 indicators pertaining to total fertility rate, abortion, family planning practices, ante-natal care, delivery care, postnatal care, immunization, childhood diseases, breastfeeding and supplementation, birth registration, disability, injury, morbidity, personal habits, etc., are being released.  
  
Total fertility rate, current usage of family planning, unmet need for family planning, mean ate at marriage for female, marriage among females below legal age (18 years), ante-natal care, delivery care, janani suraksha yojana, post-natal & new born care and immunization.  
  
Earlier Dr. C. Chandramouli, Registrar General & CCI made a brief presentation on the Annual Health Survey highlighting Total Fertility Rate, Current Usage of Family Planning, Unmet Need for Family Planning, Mean age at Marriage for Female and Marriage among Females below legal age (18 years).

# **The dataset contains the following features**:

1. **State**:
2. **District**:
3. **Population**: A population is the number of living people that live together in the same place.
4. **Sex Ratio**: The sex ratio is the ratio of males to females in a population.
5. **Literacy Rate**: Total number of literate persons in a given age group, expressed as a percentage of the total population in that age group.
6. **Crude Birth Rate**: The crude birth rate is the number of live birthsoccurring among the population of a given geographical area during a given year, per 1,000 mid-year total population of the given geographical area during the same year..
7. **Crude Death Rate**: The crude death rate is the number of deaths occurring among the population of a given geographical area during a given year, per 1,000 mid-year total population of the given geographical area during the same year
8. **Chronic Illness:** A chronic illness is a long-term health condition that may not have a cure.
9. **Diarrhea**: It refers to unusually loose or watery stools. Usually caused by bacteria, viruses, or parasites,
10. **Acute respiratory disease**: A sudden condition in which breathing is difficult and the oxygen levels in the blood abruptly drop lower than normal.
11. **Fever**: any body temperature above the normal oral measurement of 98.6 degrees
12. **Birth registration:** The registration of births is carried out under the provisions of the Registration of Birth and Death (RBD) Act, 1969 and the Registrar General.
13. **Schooling Status:** What percentage of student of age 6 to 17 are attending school or left the school.

## Objectives/Scope of the Analysis

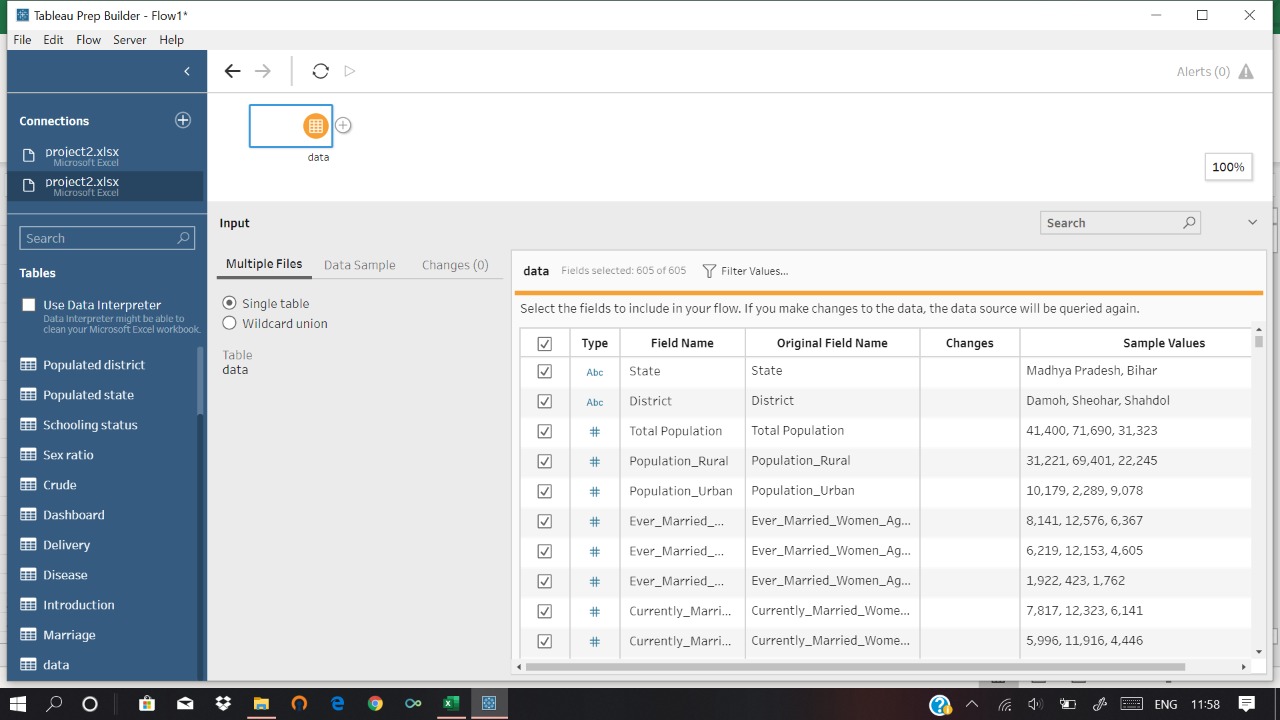
* State wise population analysis
* District wise population data analysis
* District wise sex ratio data analysis
* District wise marriage below legal age data analysis
* District wise literacy rate data analysis
* District wise child suffering from disease data analysis
* District wise Institutional delivery data analysis
* District wise crude death and birth rate data analysis
* District wise birth registration data analysis
* District wise people suffering from chronic disease data analysis
* District wise schooling status data analysis

## Source of dataset

All the dataset used in this is collected from very well-known online website [www.kaggle.com](http://www.kaggle.com/)

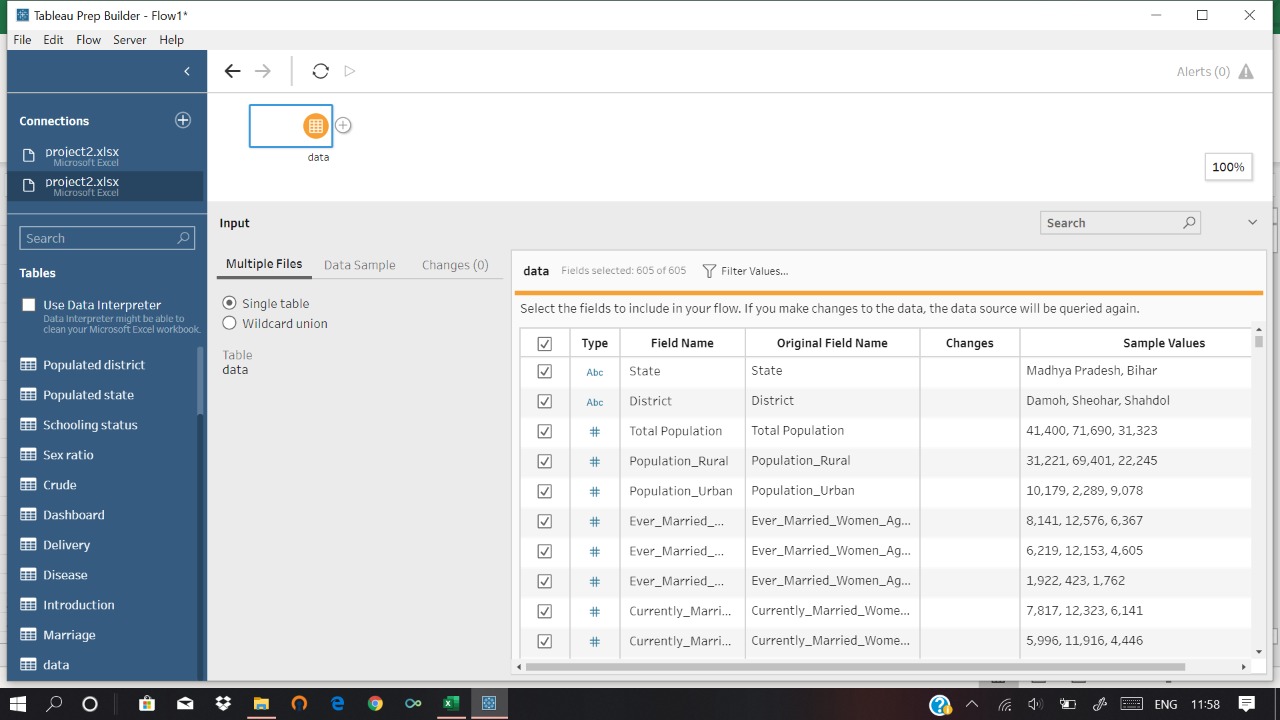
# ETL Process

**Step1**

****

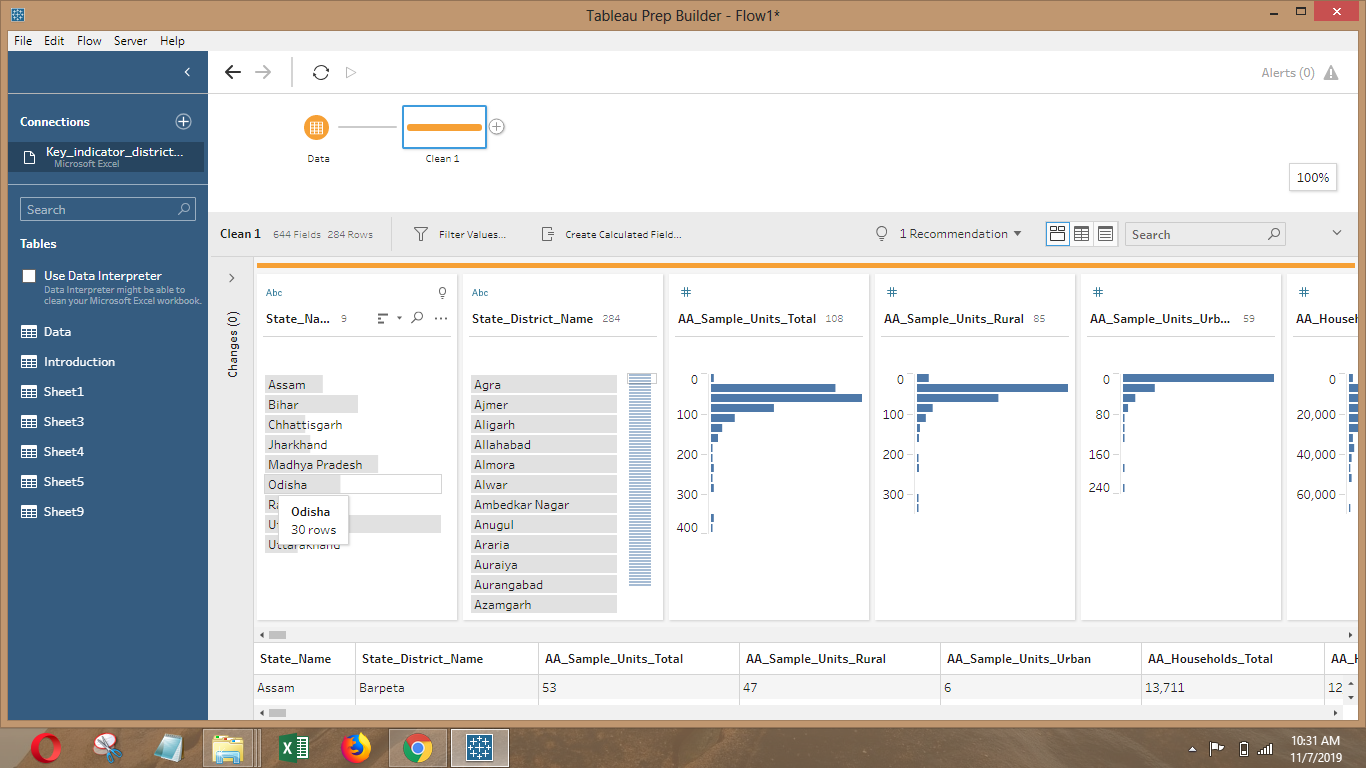
In this Step File Containing Data is connected to Tableau Prep

# Step2

****

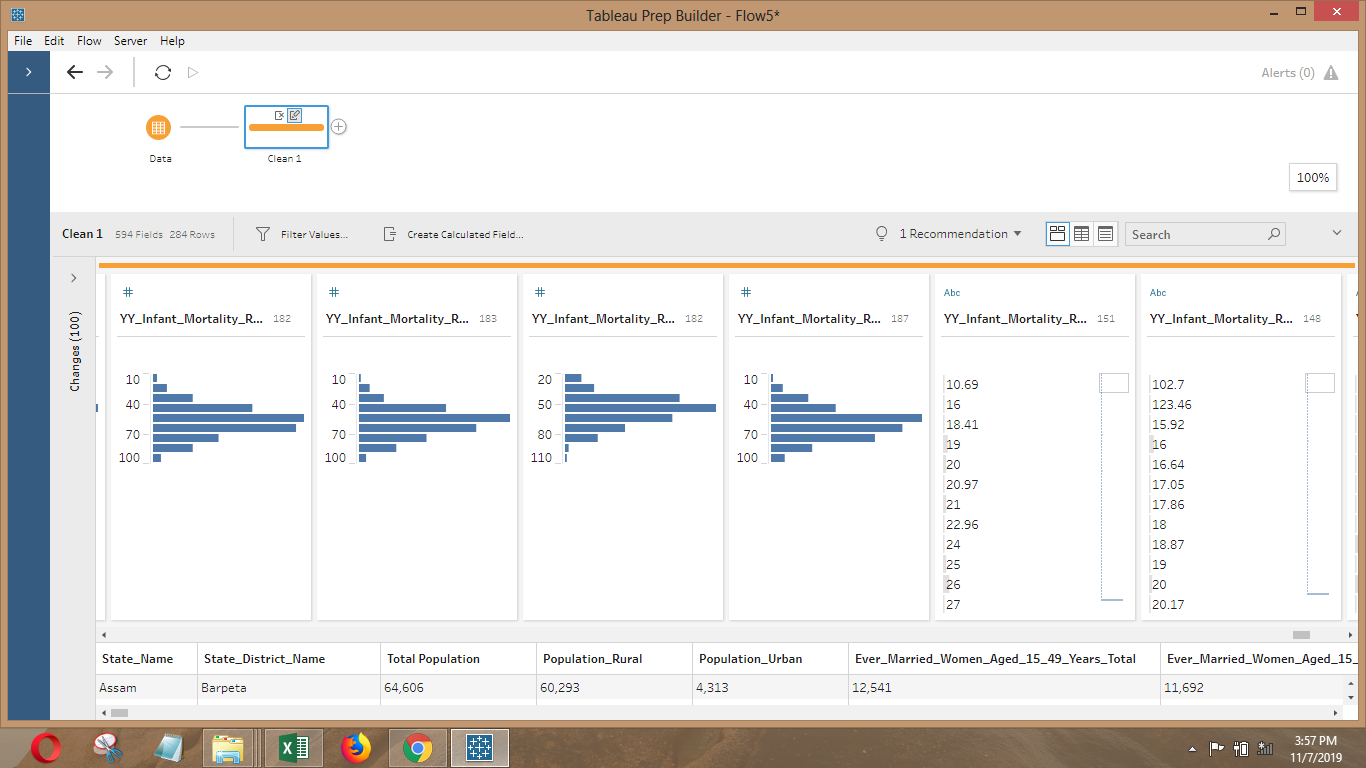
In this process I have Selected the Data from the column for further cleaning process that I am going to use in my project.

# Step3

****

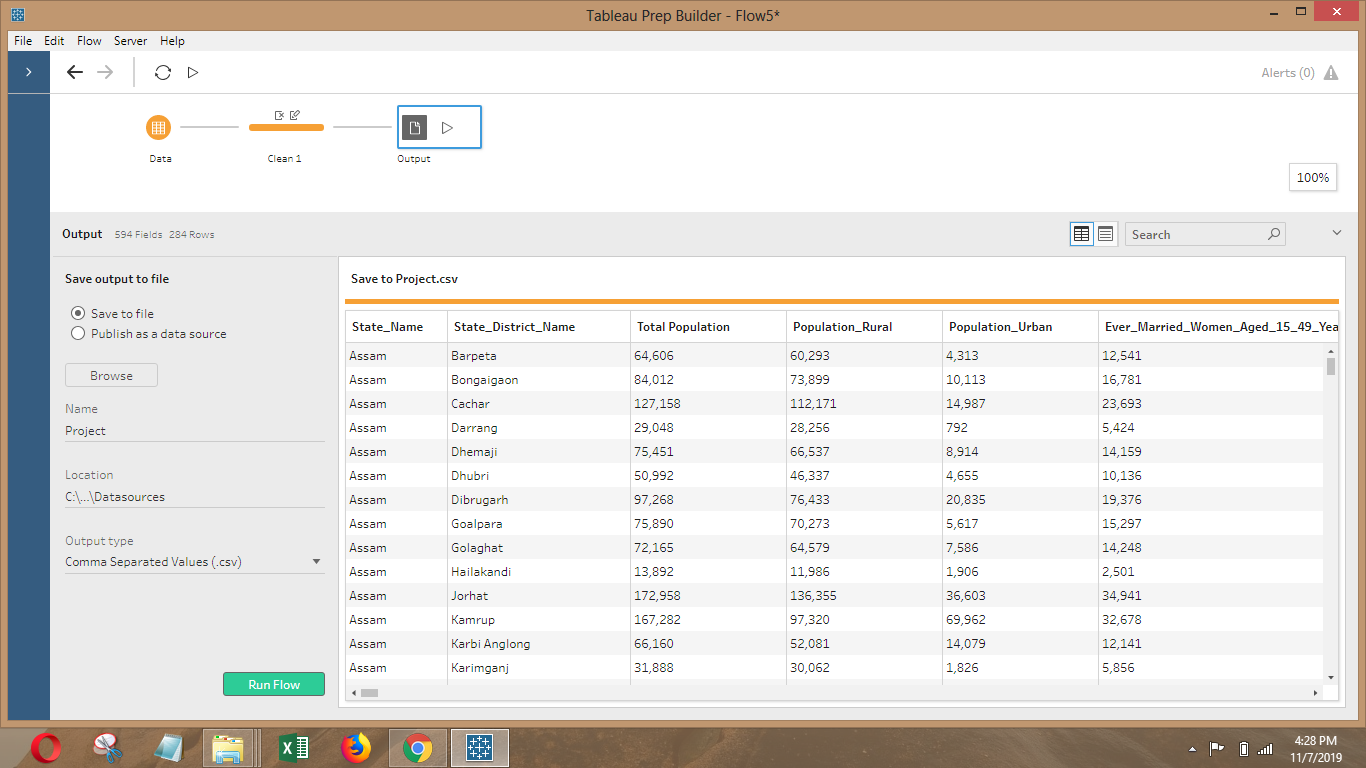
In this process I have add steps to clean data

# Step4

****

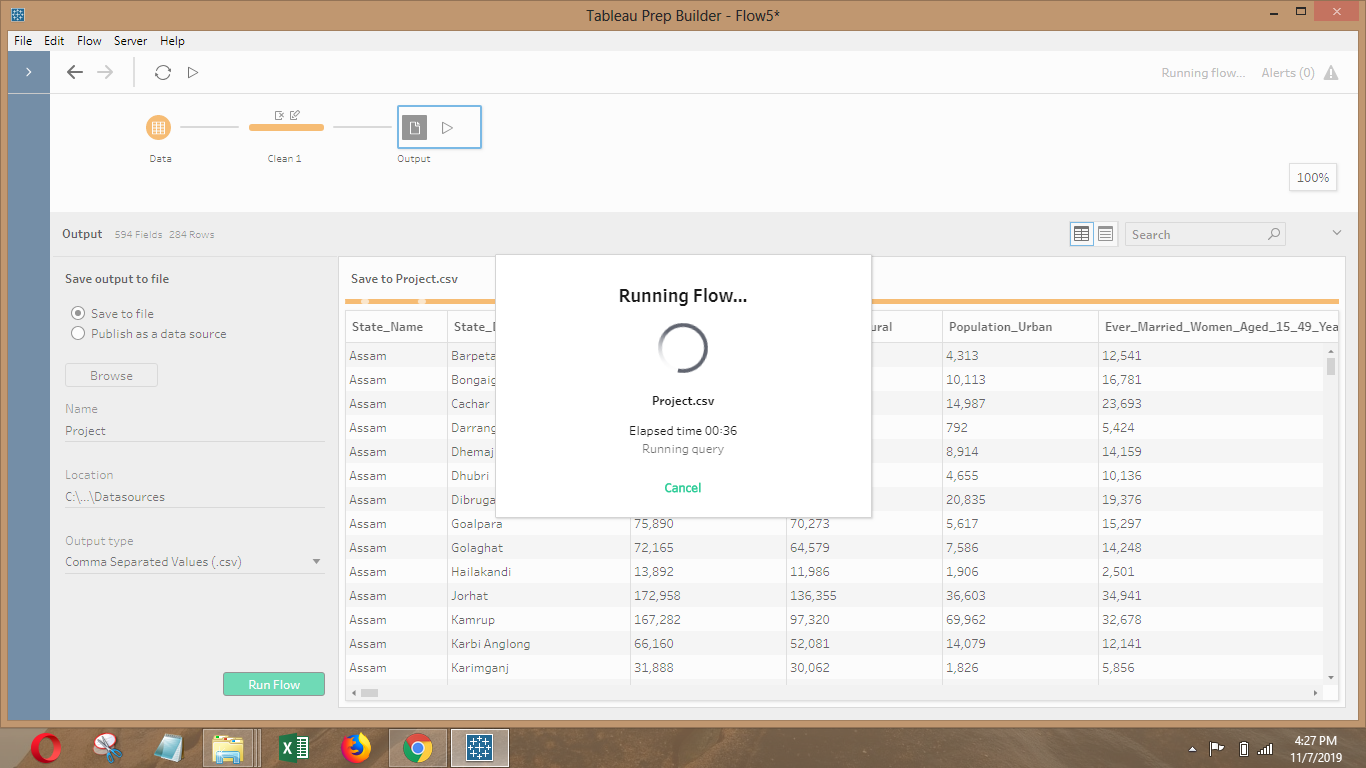
In this process I have deleted the column and that are not useful after that I change the column name into relevant name after that transform the value where the Null are given in place of the data

# Step5

****

In this process I have add Output process after all the cleaning process

# Step6

****

Finally, in this process I have save the Data in(.csv) file and Run Flow to Save the processed data.

**Analysis on Dataset**

1. **State wise Total Population data analysis:-**

## Description:

In this we are analysis the data of population of the 9 state. For that we have made the pivot table from the original dataset. Pivot table contains District in row and total population of state as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Uttar Pradesh are most populated among 9 state.

1. **District wise Total Population data analysis:-**

## Description:

In this we are analysis the data of population of the 285 district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and total population of district as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Dehradun is more populated among 285 district.

1. **District wise sex ratio data analysis:-**

## Description:

In this we are analysis the sex ratio data of the 285 district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and sex ratio as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Tehri Garhwal has highest sex rate among 285 district.

1. **District wise Literacy rate data analysis:-**
2. **Male**

## Description:

In this we are analysis the data of Literacy rate of male of district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and percentage of male as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* chamoli has highest percentage literate male among 285 district.

1. **Female**

## Description:

In this we are analysis the data of literacy rate of female of district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and percentage of literate female as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* North cachar hills has highest percentage of literate female infection among 285 district.

1. **District wise marriage below legal age data analysis:-**
2. Female

## Description:

In this we are analysis the data of marriage of female below legal age of district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and percentage of female as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Bhilwara has percentage of marriage of female below legal age among 285 district.

1. **Male**

## Description:

In this we are analysis the data of marriage of male below legal age of district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and percentage of male as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Jhabua has percentage of marriage of male below legal age among 285 district.

1. **District wise children suffering from disease data analysis:-**
2. **Acute Respiratory Infection**

## Description:

In this we are analysis the data of children suffering from acute respiratory infection of district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and percentage of children suffering acute respiratory infection as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Rampur has percentage of children suffering from acute respiratory infection among 285 district.

1. **Fever**

## Description:

In this we are analysis the data of children suffering from fever of district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and percentage of children suffering fever as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Raisen has percentage of children suffering from fever among 285

1. **District wise Institutional delivery data analysis:-**
2. Private Institution

## Description:

In this we are analysis the data of delivery in private institution of district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and private institution as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Jhunjhunum has percentage of delivery in government institution among 285

1. Government Institution

## Description:

In this we are analysis the data of delivery in government institution of district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and government institution as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Guna has percentage of delivery in government institution among 285 district.

1. **District wise Crude birth and Crude death rate data analysis:-**
2. **Crude Birth Rate**

## Description:

In this we are analysis the data of crude birth rate of district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and crude birth rate as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Shrawasti has highest crude Birth rate among 285 district.

1. **Crude Death Rate**

## Description:

In this we are analysis the data of crude death rate of district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and crude death rate as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Shrawasti has highest crude death rate among 285 district.

1. **District wise Birth registration data analysis:-**
2. **Best Birth Registered**

## Description:

In this we are analysis the data of birth registration of district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and birth registration percentage as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Barmn has highest percentage of birth registration among 285 district

1. **Worst Birth Registered**

## Description:

In this we are analysis the data of birth registration of district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and birth registration percentage as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Auraiya has worst rate of birth registration among 285 district.

1. **District wise Population suffering from chronic disease data analysis:-**

## Description:

In this we are analysis the data of people suffering from chronic illness of the 285 district. For that we have made the pivot table from the original dataset. Pivot table contains District in row and value of dropout as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Jorhat has highest number of people suffering from chronic illness

Among 285 district..

1. **District wise Schooling status data analysis:-**
2. **Attending School**

## Description:

In this we are analysis the data of children of age 6 to 17 currently attending school. For that we have made the pivot table from the original dataset. Pivot table contains District in row and value of currently attending school as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

According to this pivot chart:

* Samastipur has highest rate of attending school among 285 district.

1. **Dropout**

## 

## Description:

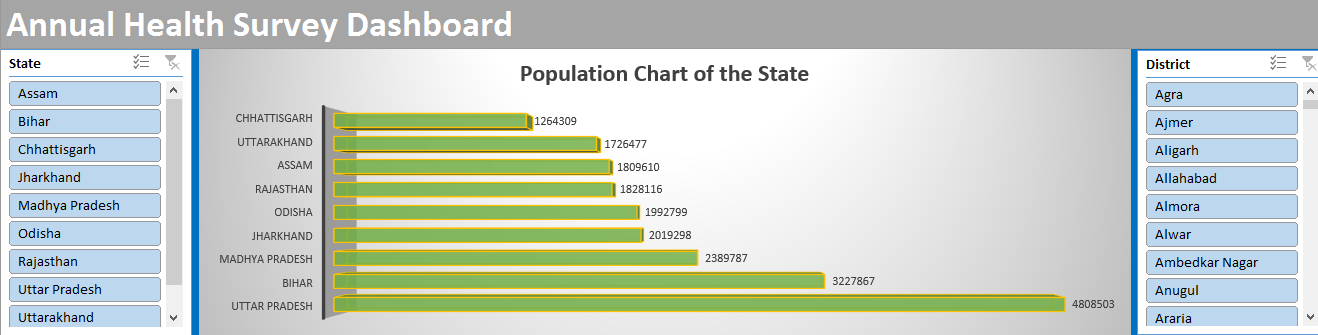
In this we are analysis the data of children of 6 to 17 dropout from school. For that we have made the pivot table from the original dataset. Pivot table contains District in row and value of dropout as columns. In chart area we’ve added data labels to each value which makes easy for the viewer to see.

## Conclusion:

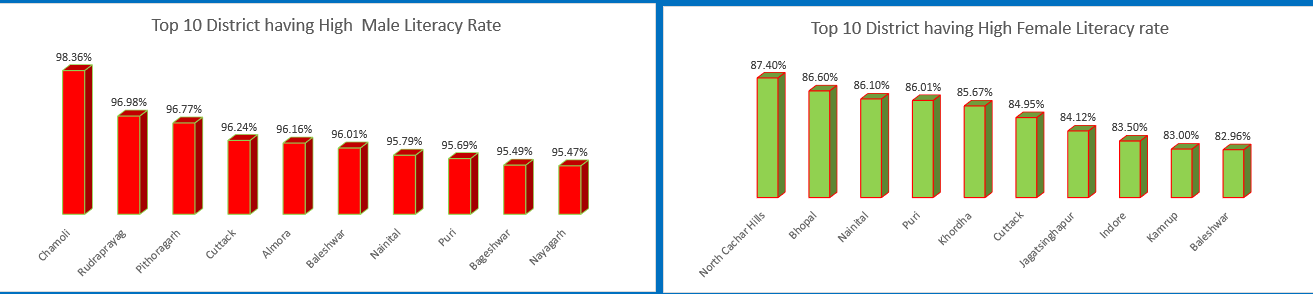
According to this pivot chart:

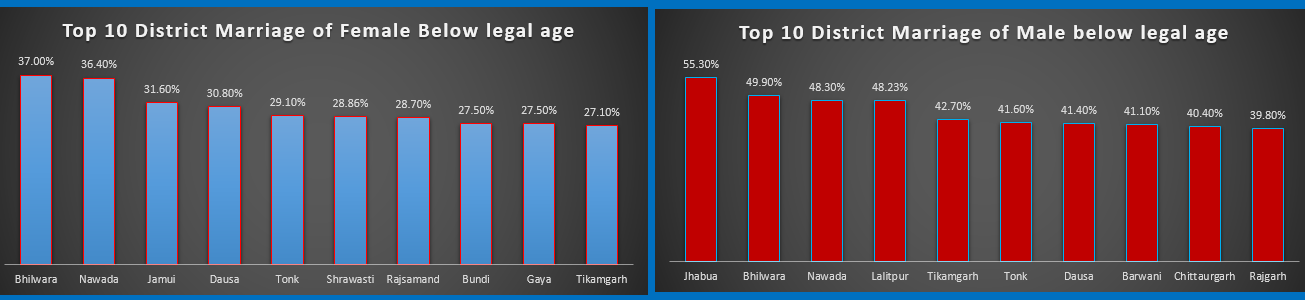
* Budaun has highest rate of dropout among 285 district.

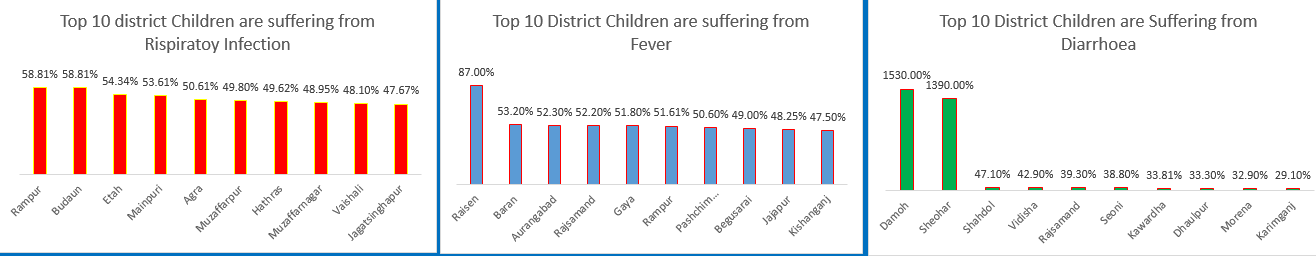
DASHBOARD











# Conclusion

To make public health services more accessible, there is a need to improve the quality of services, enlarge infrastructure to reduce waiting time, and enhance the physical reach to inpatients in the EAG states of India. Public health services will then be able to compete with those in the private sector.

## Bibliography

* + [Kaggle.com](http://www.kaggle.com/)
  + Wikipedia
  + Google
  + Youtube