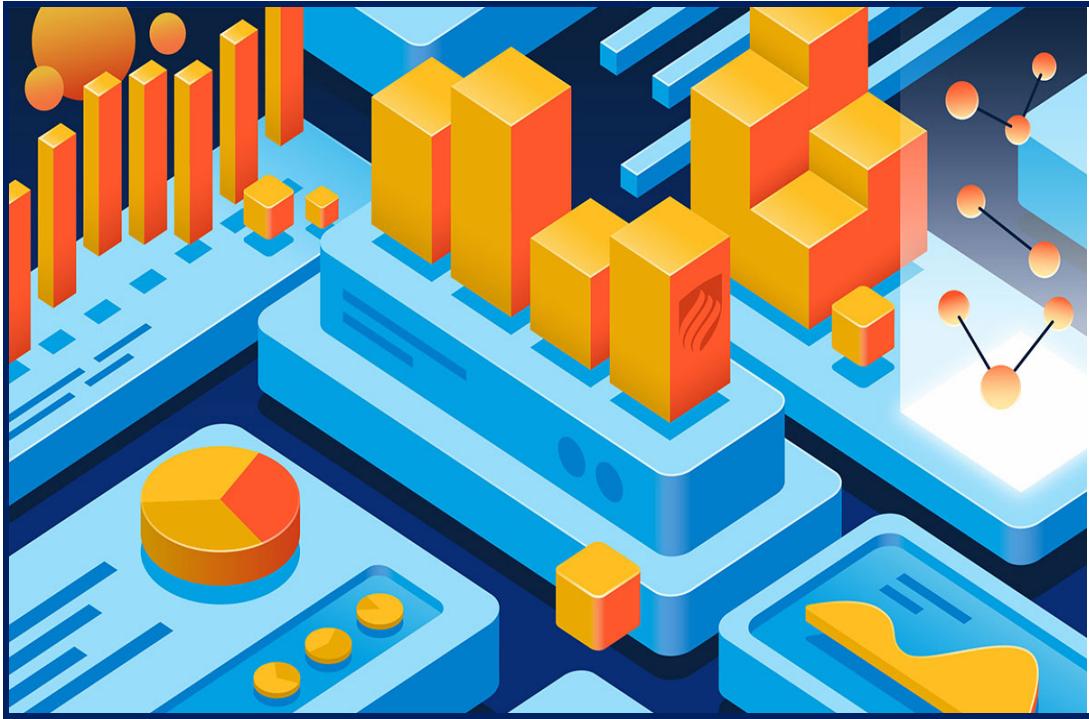


## PROJECT REPORT



☞ **Project Theme:** Water and Sanitization

☞ **Project Title:** Visual Analytics and Insights into Water Quality Affected Habitations

### Presented By:

Damarla Venkata Sai Chandana,  
Student at G.Narayananamma Institute of  
Technology and Science for women, Hyderabad.

## **TABLE OF CONTENTS**

S.no.	TITLE	Page no.
1	Introduction 1.1 Overview 1.2 Purpose	3 4
2	Literature Survey 2.1 Existing problem 2.2 Proposed solution	4 4
3	Software/Hardware Requirements	4
4	Experimental Investigations	5
5	Result	5-16
6	Advantages & Disadvantages	17
7	Applications	17
8	Conclusion	17
9	Future Scope	17
10	Bibliography	17

## **INTRODUCTION**

### **1.1 Overview**

This project is a data visualization developed from a dataset taken from [data.gov.in](http://data.gov.in) using **Tableau** and finally, the worksheets, dashboards and story are published on Tableau Public. The theme chosen for project is "Water and Sanitization" and the dataset chosen is "Water Quality affected Habitations in India as of 2012".

The dataset contains the following columns:

#### **Dimensions**

- ☞ State Name
- ☞ District Name
- ☞ Panchayat Name
- ☞ Block Name
- ☞ Habitation Name
- ☞ Village Name
- ☞ Date

#### **Measures**

- ☞ Quality Parameter

The **Quality parameters** which are affecting are:

- ☞ Arsenic
- ☞ Iron
- ☞ Flouride
- ☞ Nitrate
- ☞ Salinity

**The visualizations developed out of plethora of tools provided by Tableau on the dataset are:**

- The Quality parameter Analysis
- No. of affected habitations in each State
- Arsenic affected States
- Flouride affected States
- Iron affected States
- Nitrate affected States
- Salinity affected States
- Quality Parameter prominence based on State
- % affect of Quality parameters in each State
- Top 10 affected Districts in each State
- Top 10 affected Panchayats/Villages/Blocks/Habitations in each State
- Visual Analytics and Insights into Water Quality affected Habitations in India (Dashboard)
- Story of Water Quality affected Habitations (Story)

## 1.2 Purpose

The purpose of visualizing data is to analyze various factors in a dataset. This task of visualizing data comes under Exploratory Data Analytics (EDA) part as we go on exploring various factors and ultimately come to a meaningful conclusion in order to solve an existing problem. The dataset being chosen gives great insights into water quality affected habitations in India, this helps government to understand which areas are affected most and by what parameter in what percentage, finally it helps the officials to take necessary action in the relevant area in order to improve the water quality in that area.

## LITERATURE SURVEY

### 2.1 Existing problem

As per the dataset, we could observe that there are states like Rajasthan, West Bengal, Karnataka etc. which are affected too much by the lack of water quality. And each state is affected by various parameters. These things are going unnoticed. Because no one can understand a raw data. Government suffers in getting proper insights into the data which should help them in decision-making. Due to improper visualizations, officials could not come to an appropriate and valid decision.

### 2.2 Proposed Solution

In order to overcome the existing problem, **Tableau software** is used to achieve good insights into data by various visualizations like pie chart, scatter plot, packaged bubble chart, maps view, etc. An **interactive dashboard** is being built from the visualizations and a **story** is made out of each story point (Story gives us conclusions from each story point). This visualizations help officials take a necessary action in the affected areas. Thus, problem gets solved.

## SOFTWARE/ HARDWARE REQUIREMENTS

### Software

- ☞ Tableau
- ☞ Tableau Public
- ☞ Dataset

### Hardware

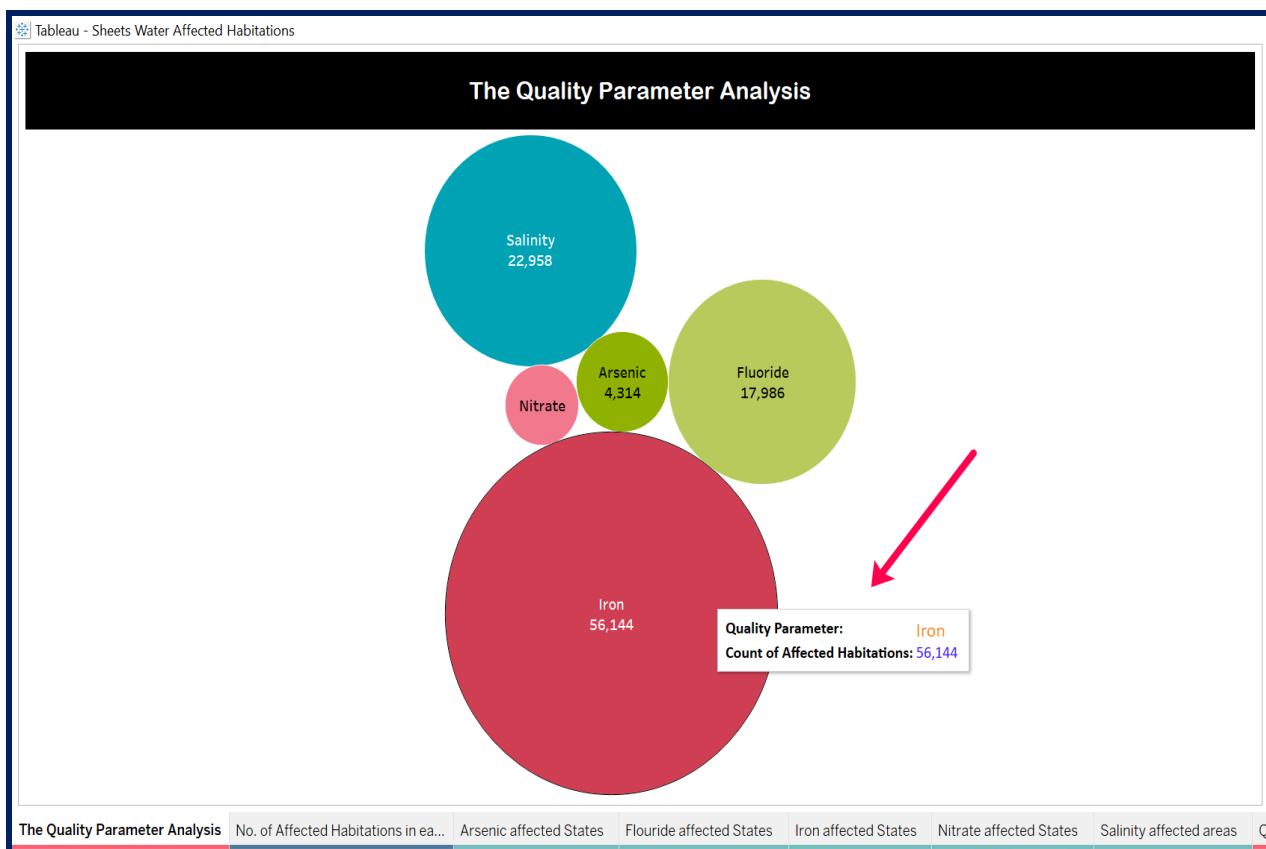
- ☞ A desktop/laptop with atleast 4 GB RAM
- ☞ Any browser

## EXPERIMENTAL INVESTIGATIONS

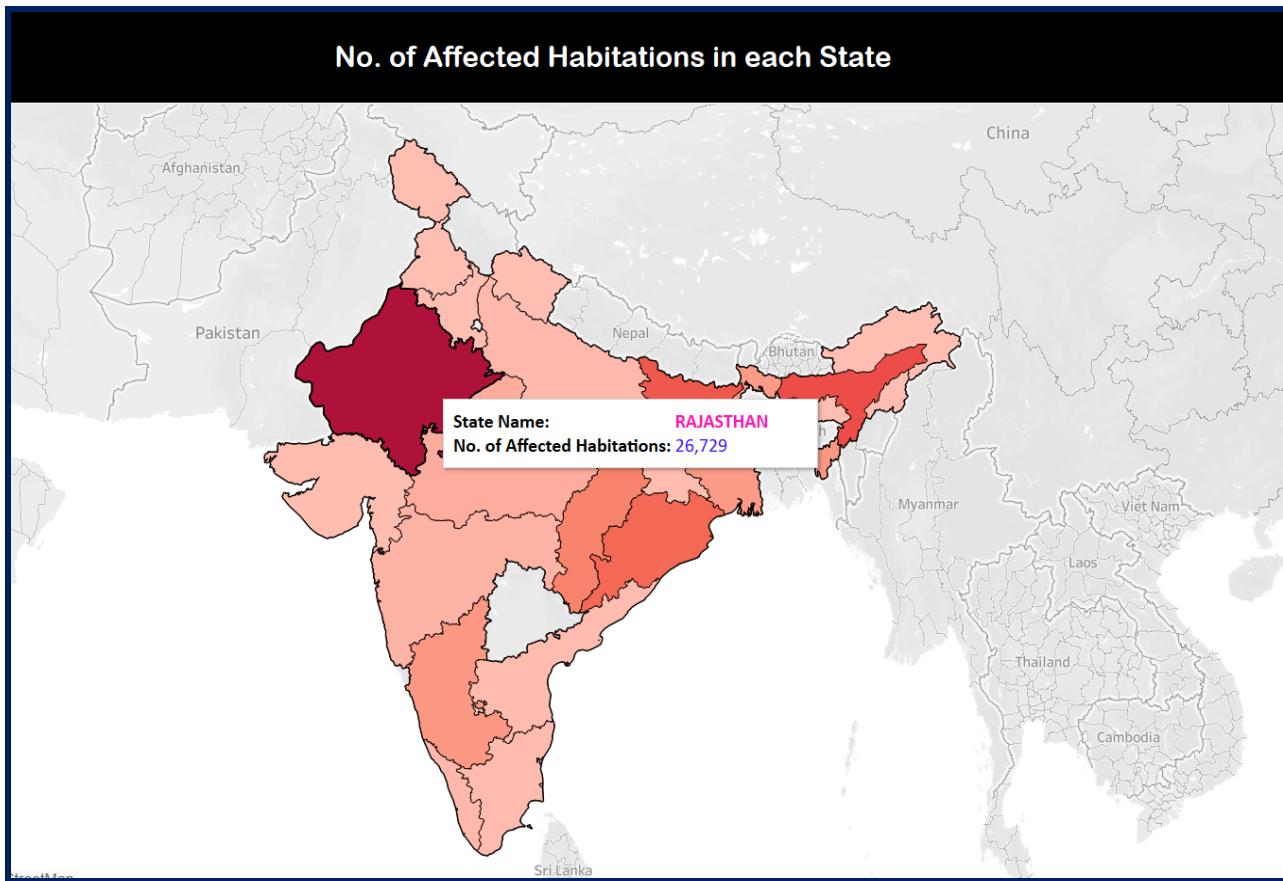
- 1) Which is the quality parameter which is affecting most of the habitations in India?
- 2) Which state is most affected by lack of water quality?
- 3) Explore which quality parameter is dominant in which state?
- 4) Which parameters are affecting each state and by what percentage of total count?
- 5) What are the most affected Districts in each state due to poor water quality?
- 6) What are the most affected Villages in each state due to poor water quality?
- 7) What are the most affected Panchayats in each state due to poor water quality?
- 8) What are the most affected Blocks in each state due to poor water quality?
- 9) What are the most affected Habitations in each state due to poor water quality?
- 10) Dashboard connecting all the investigations.
- 11) Story connecting all the meaningful conclusions from story points.

## RESULT

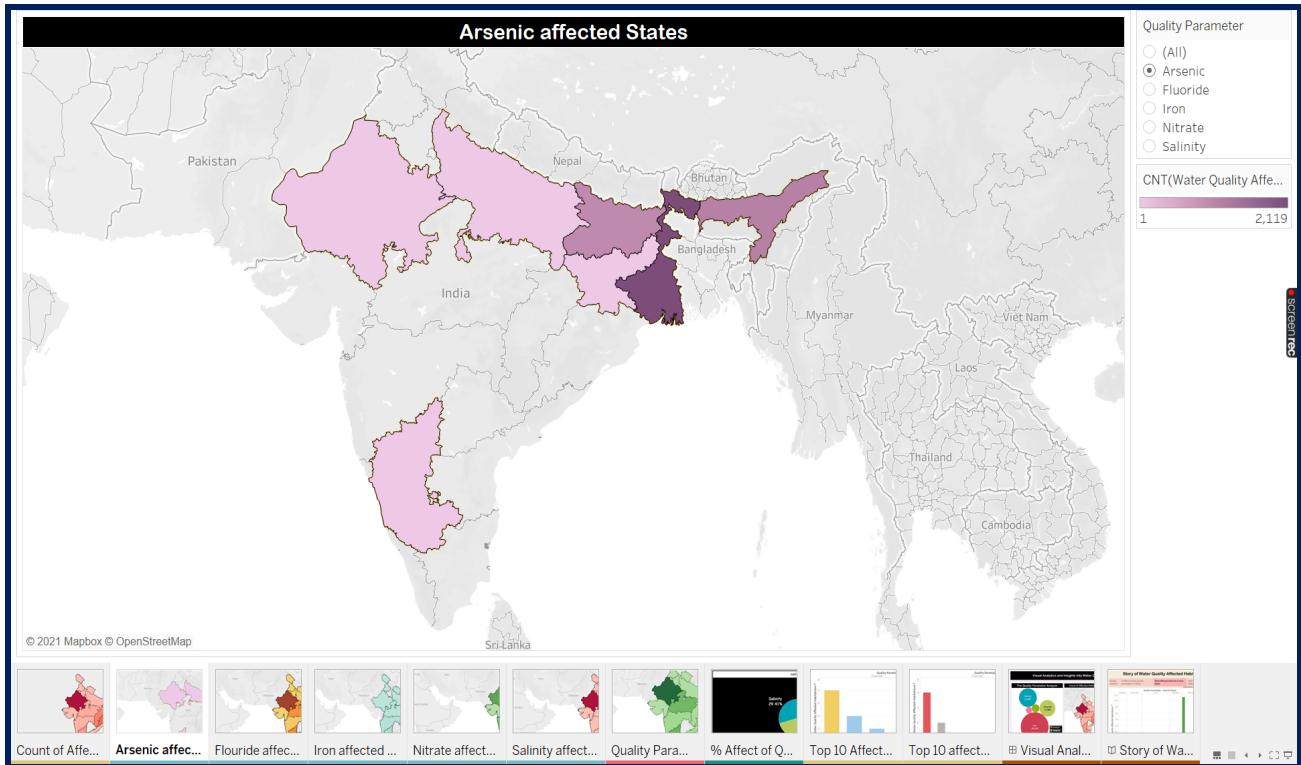
- 1) The Quality parameter Analysis



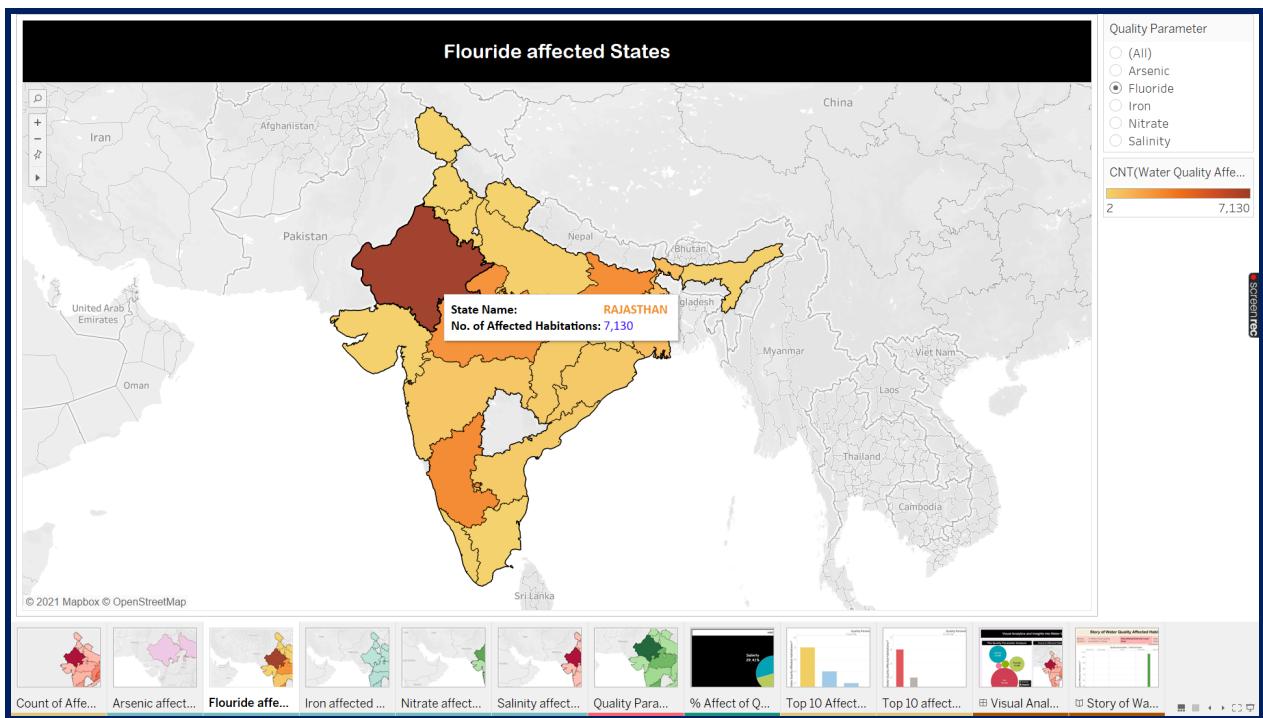
2) No. of affected habitations in each State



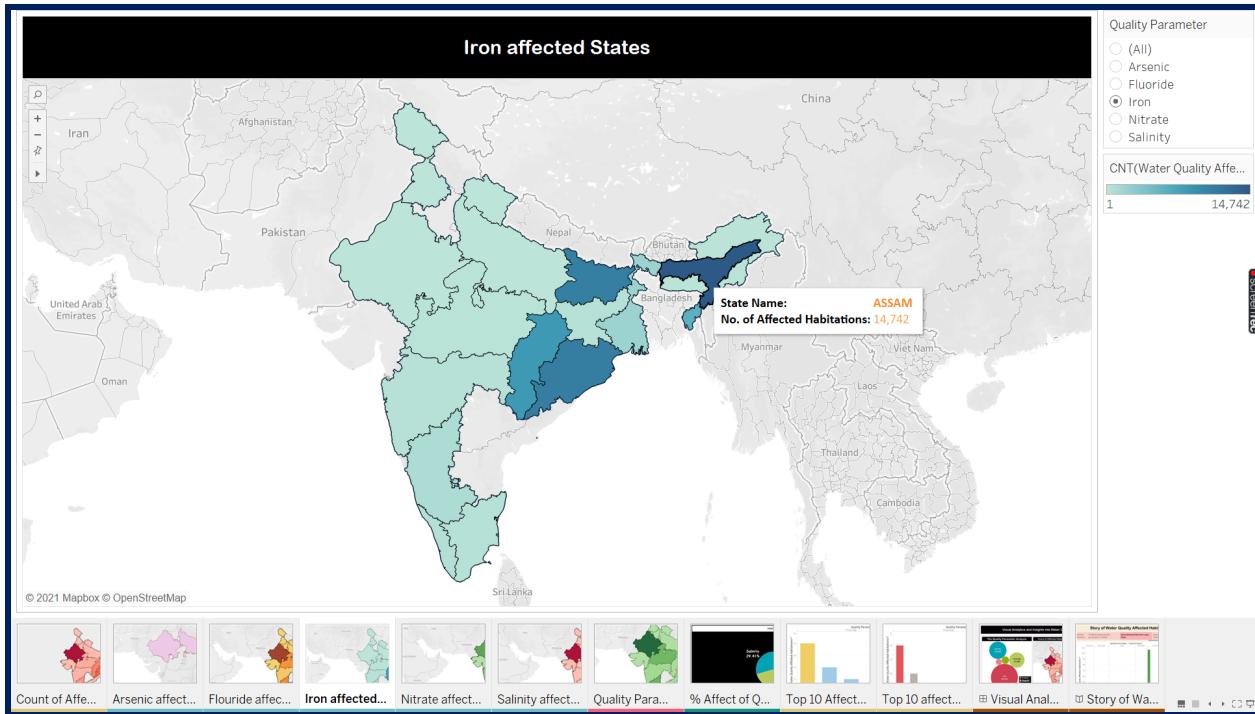
### 3) Arsenic affected States



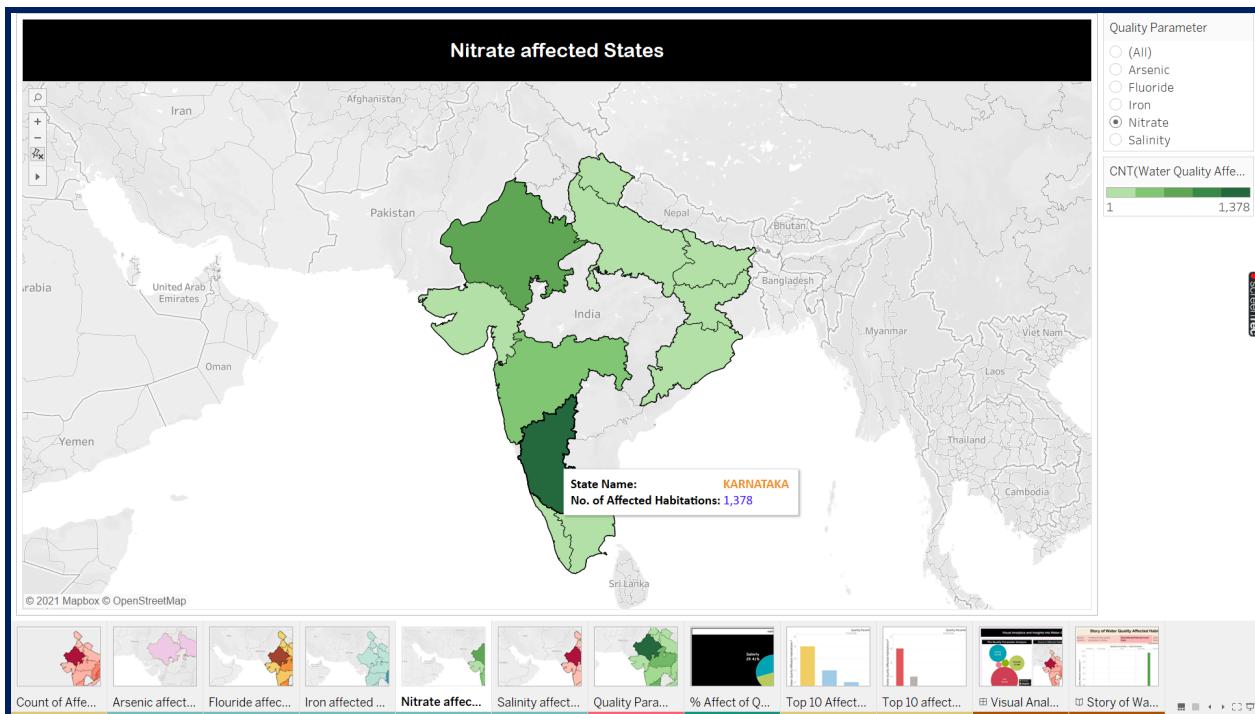
### 4) Flouride affected States



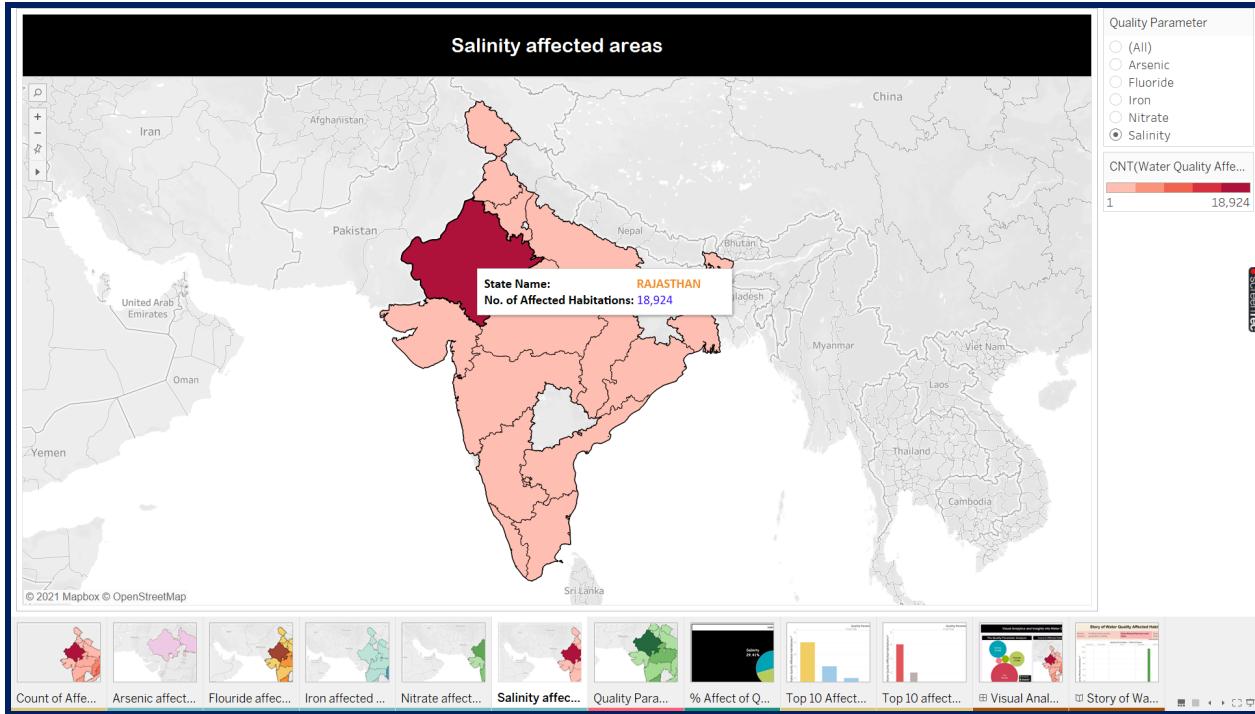
## 5) Iron affected States



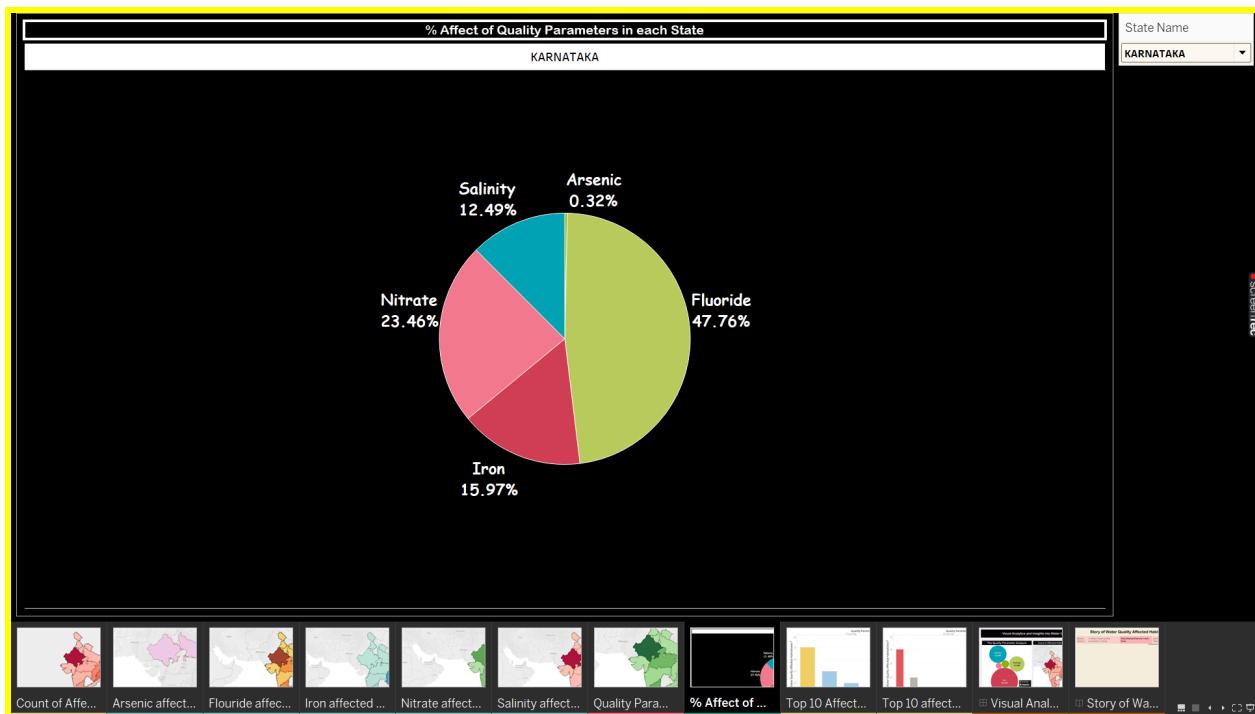
## 6) Nitrate affected States



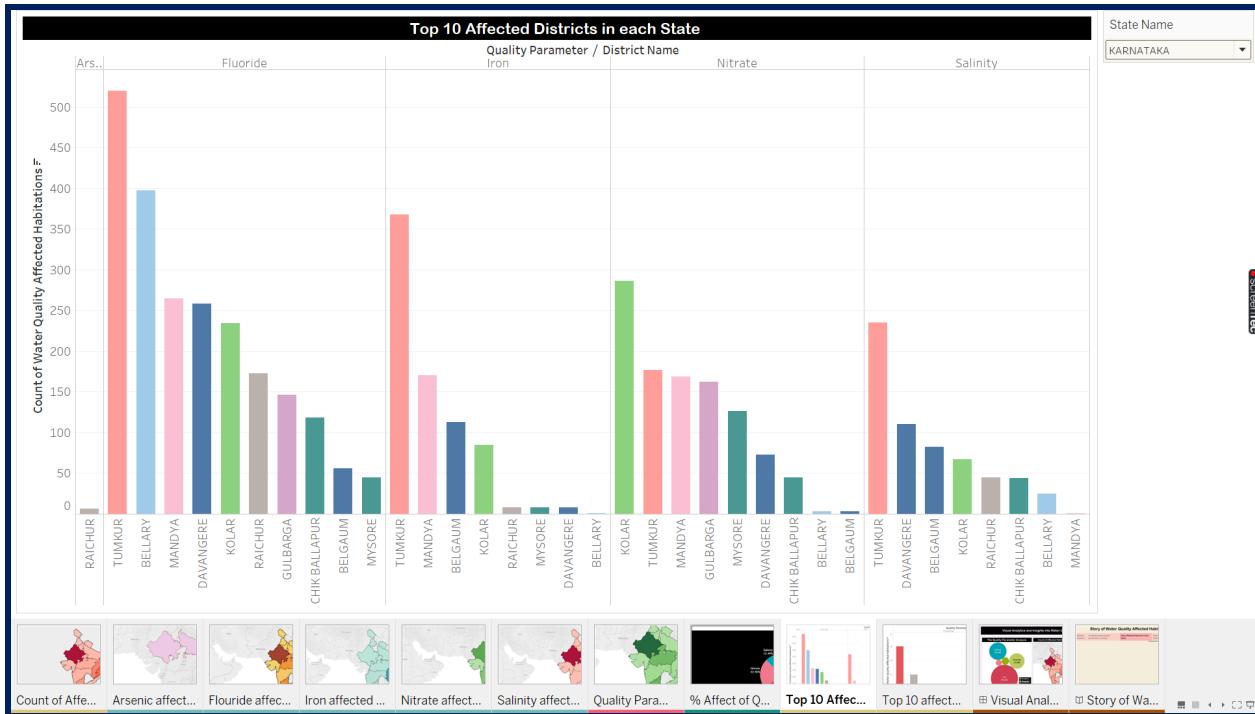
## 7) Salinity affected States



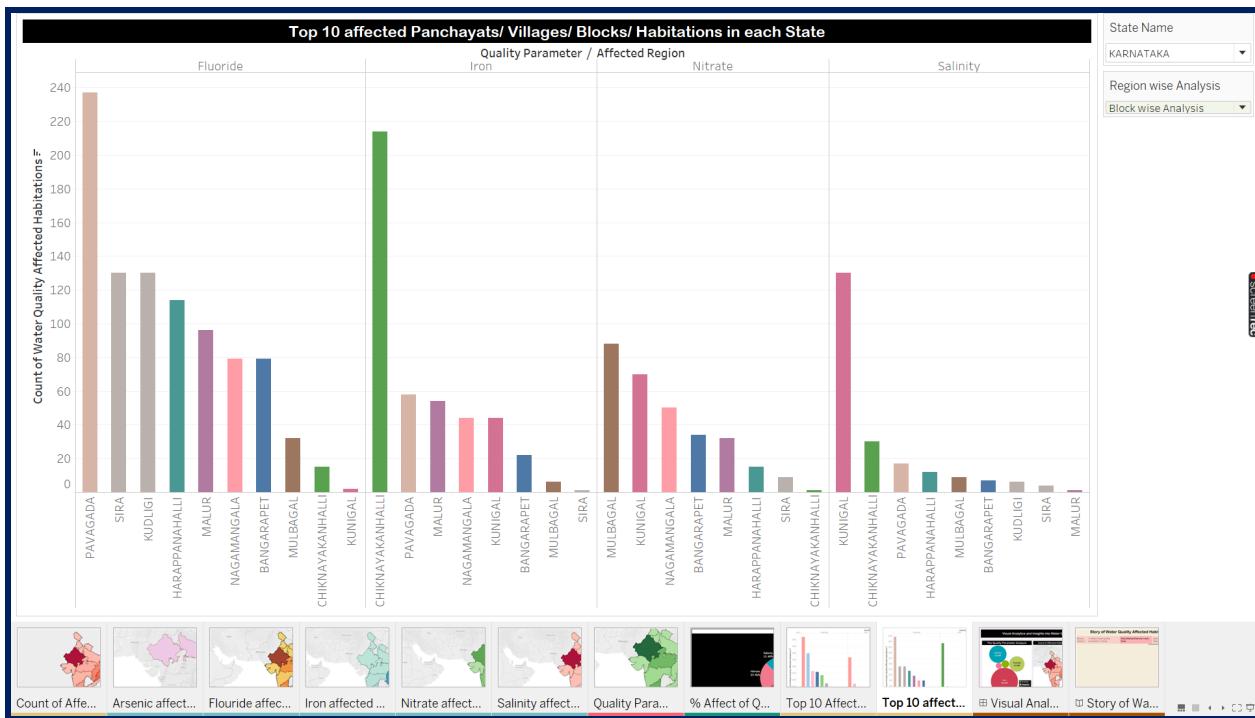
## 8) % affect of Quality parameters in each State



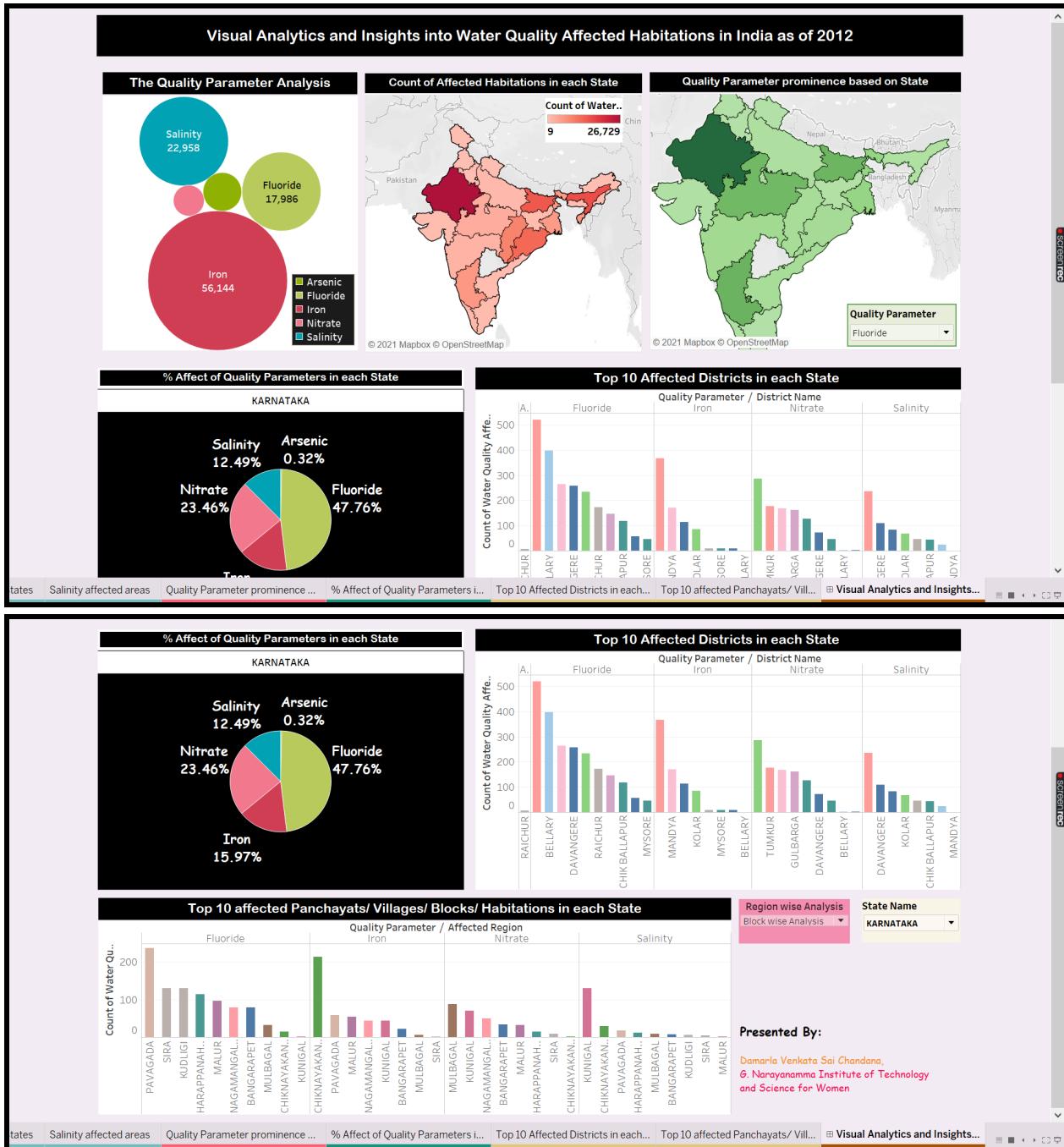
## 9) Top 10 affected Districts in each State



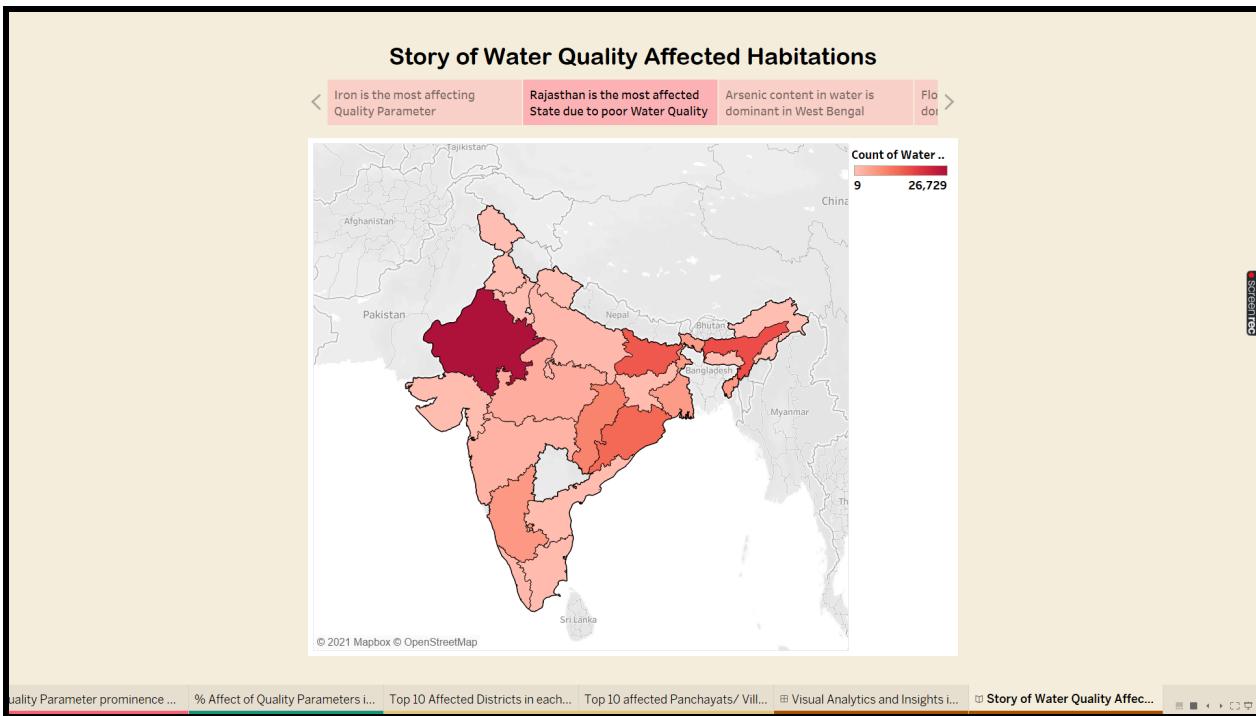
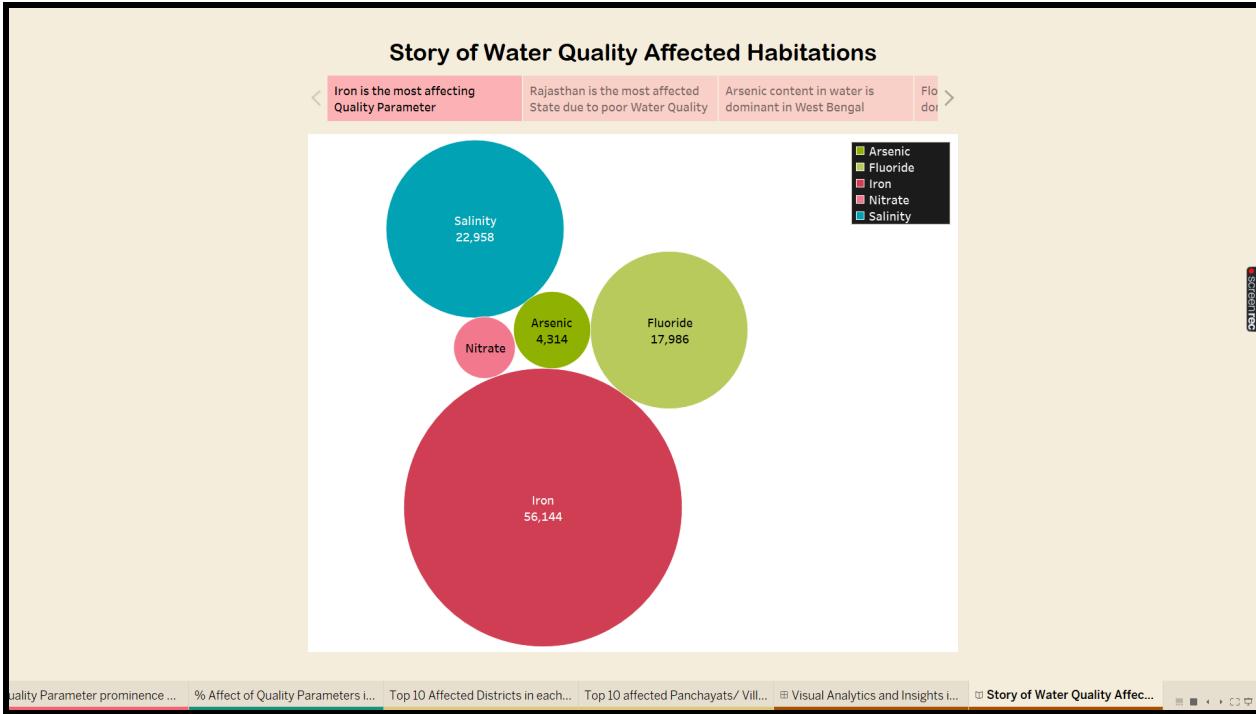
## 10) Top 10 affected Panchayats/Villages/Blocks/Habitations in each State

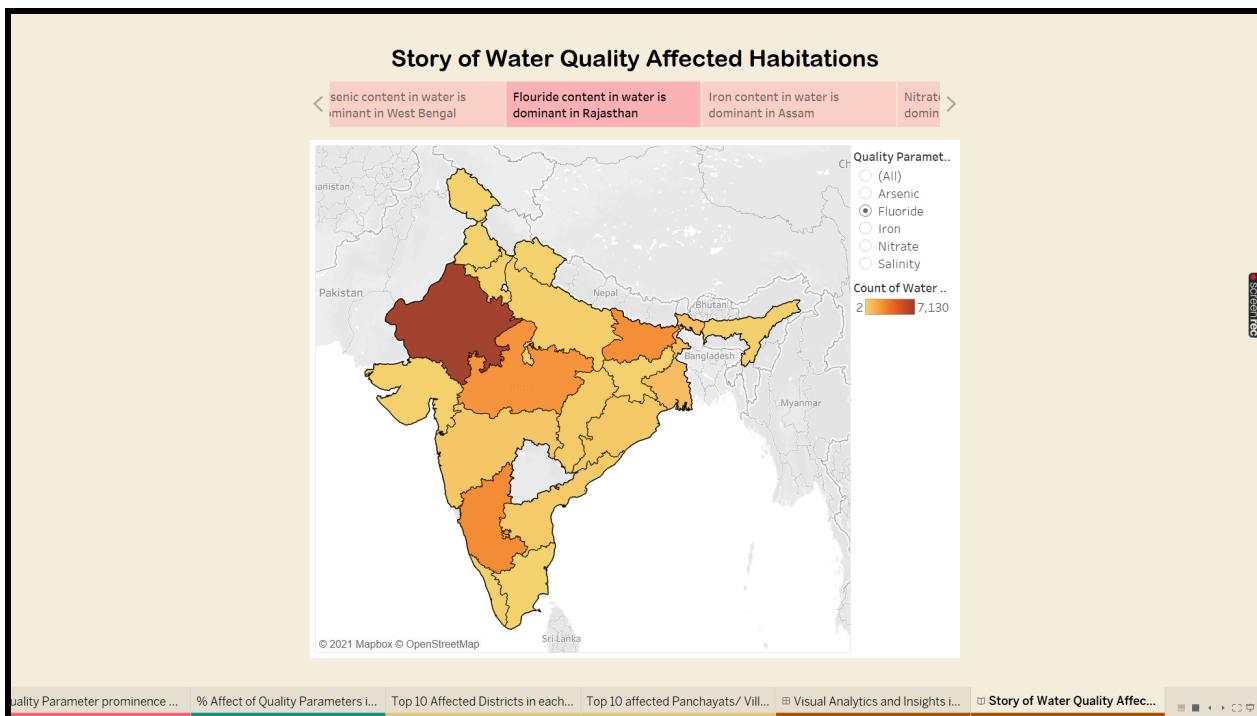
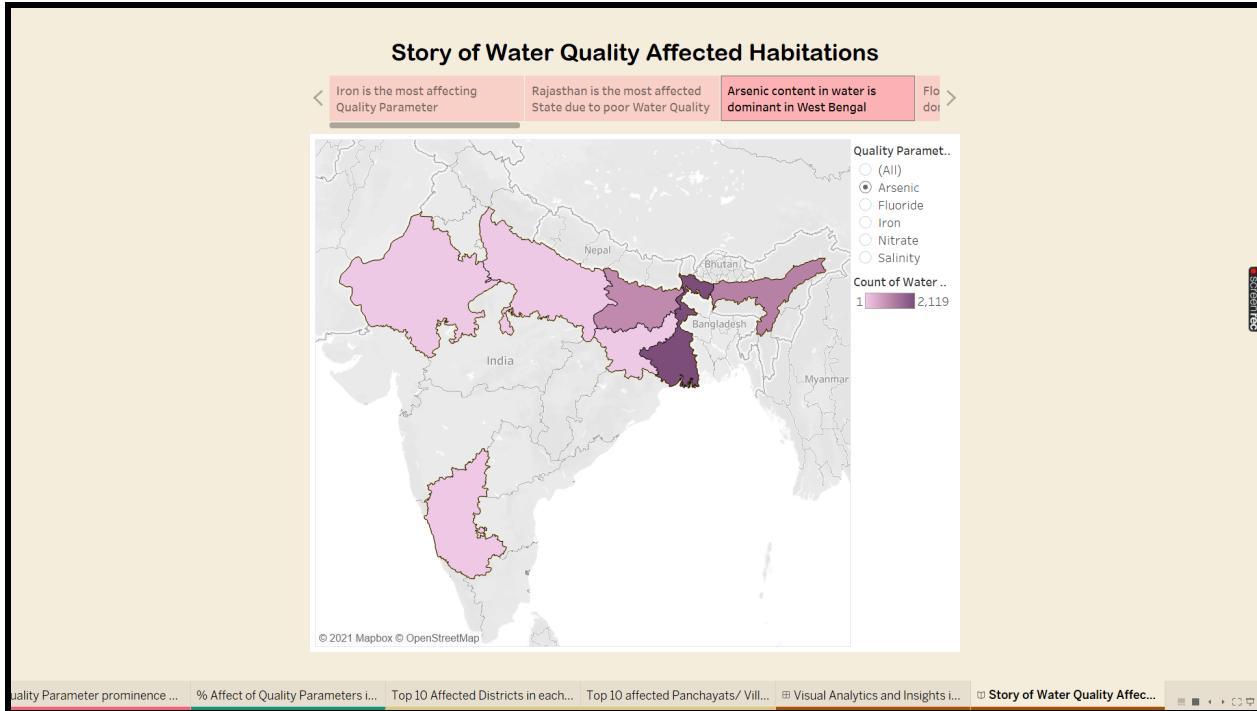


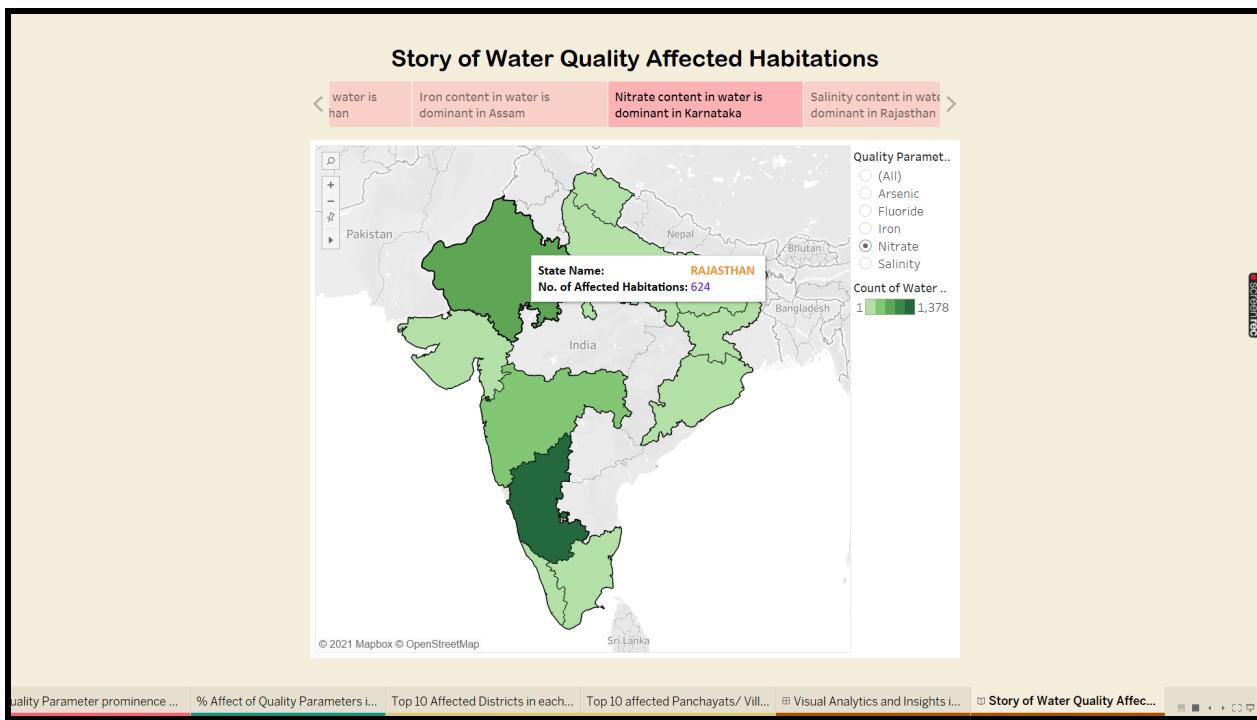
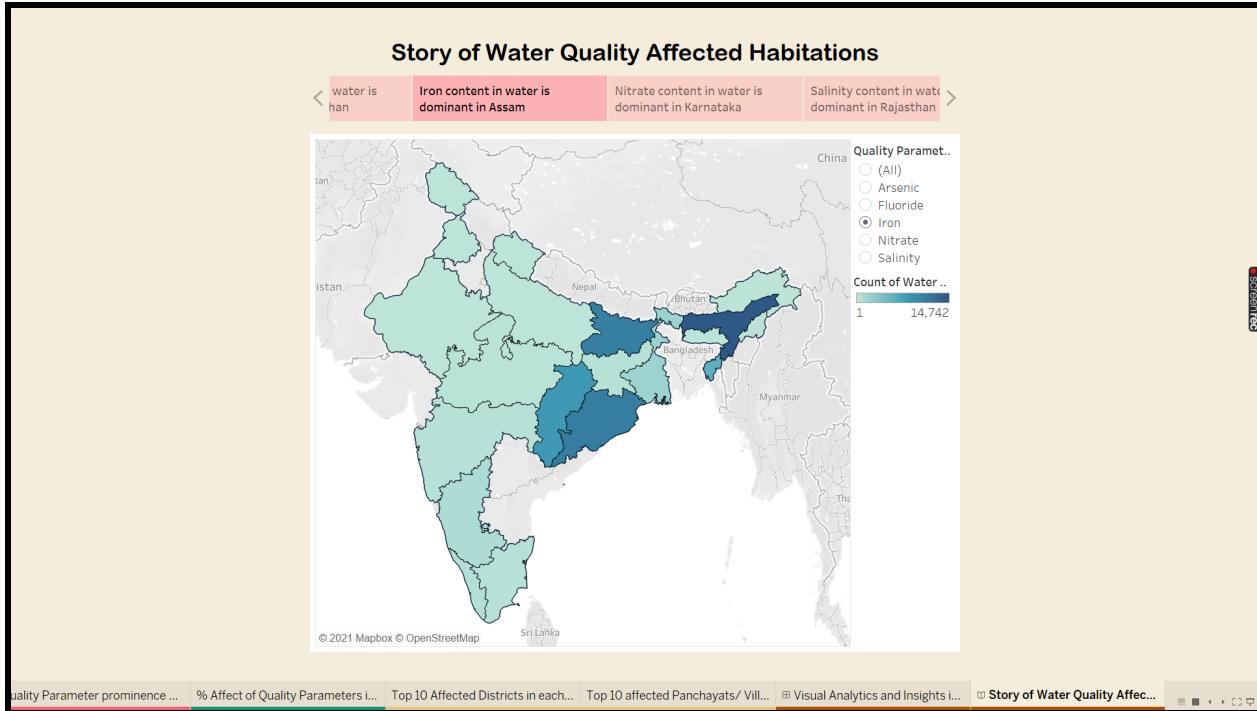
## 11) Dashboard

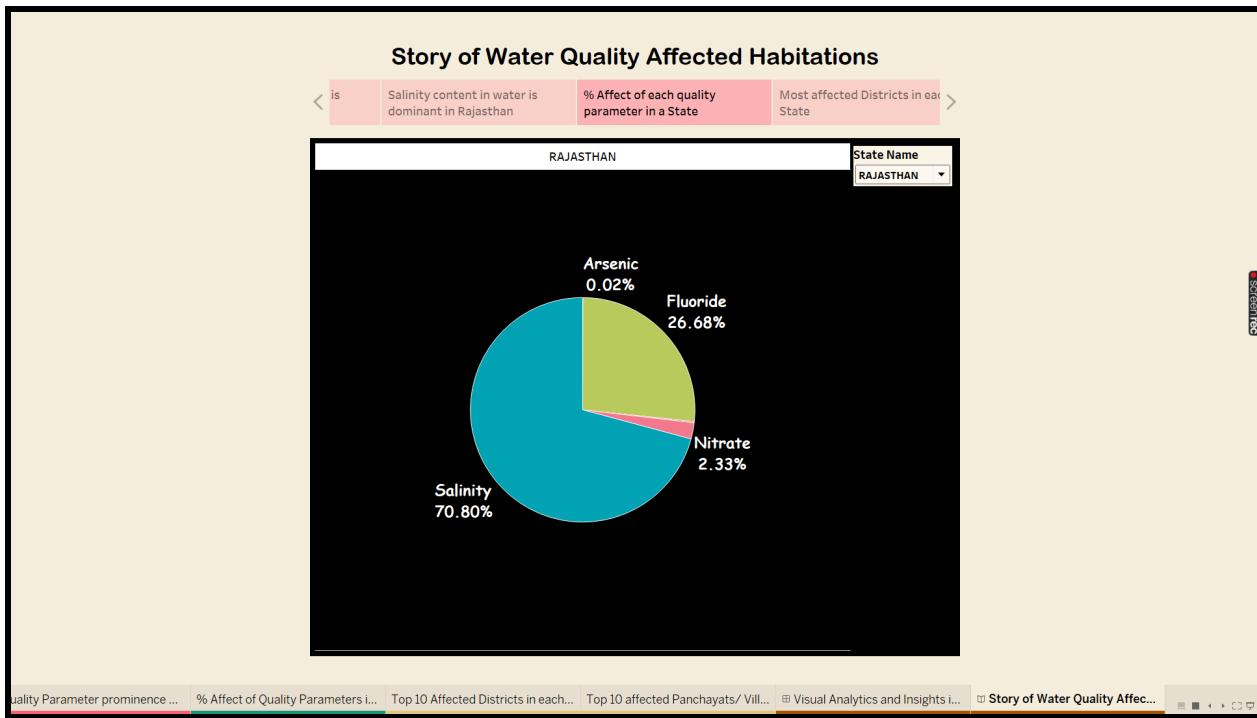
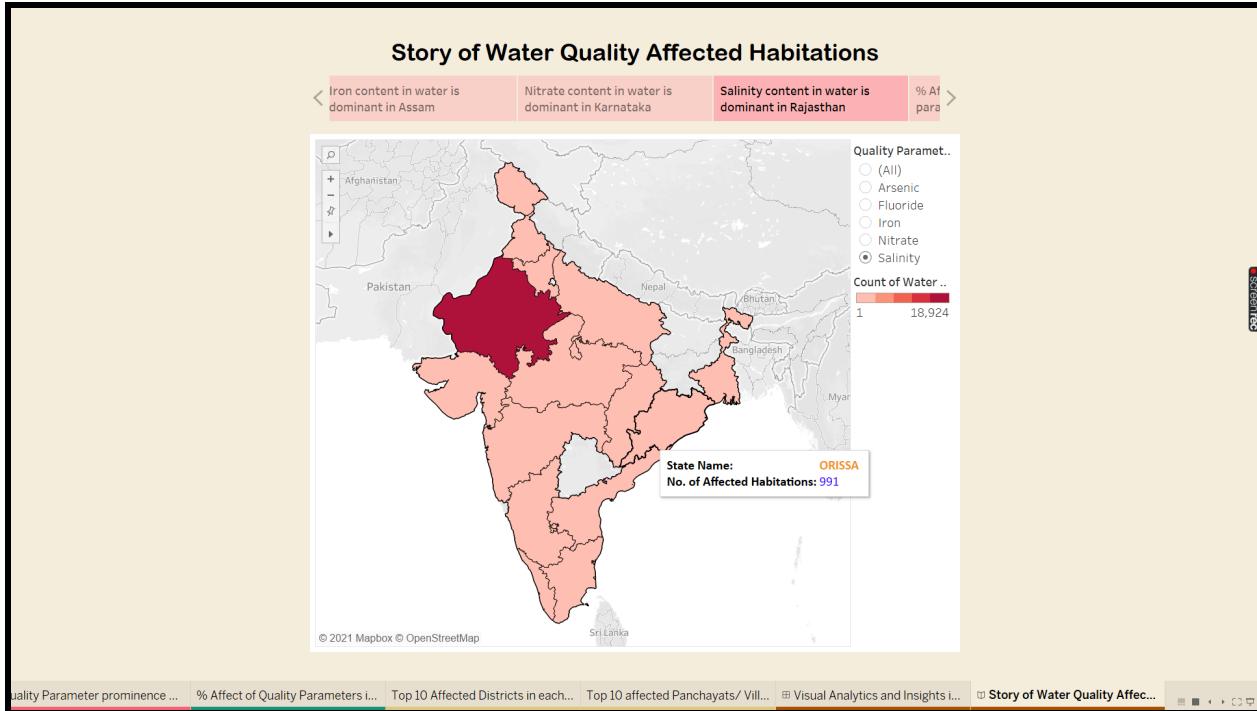


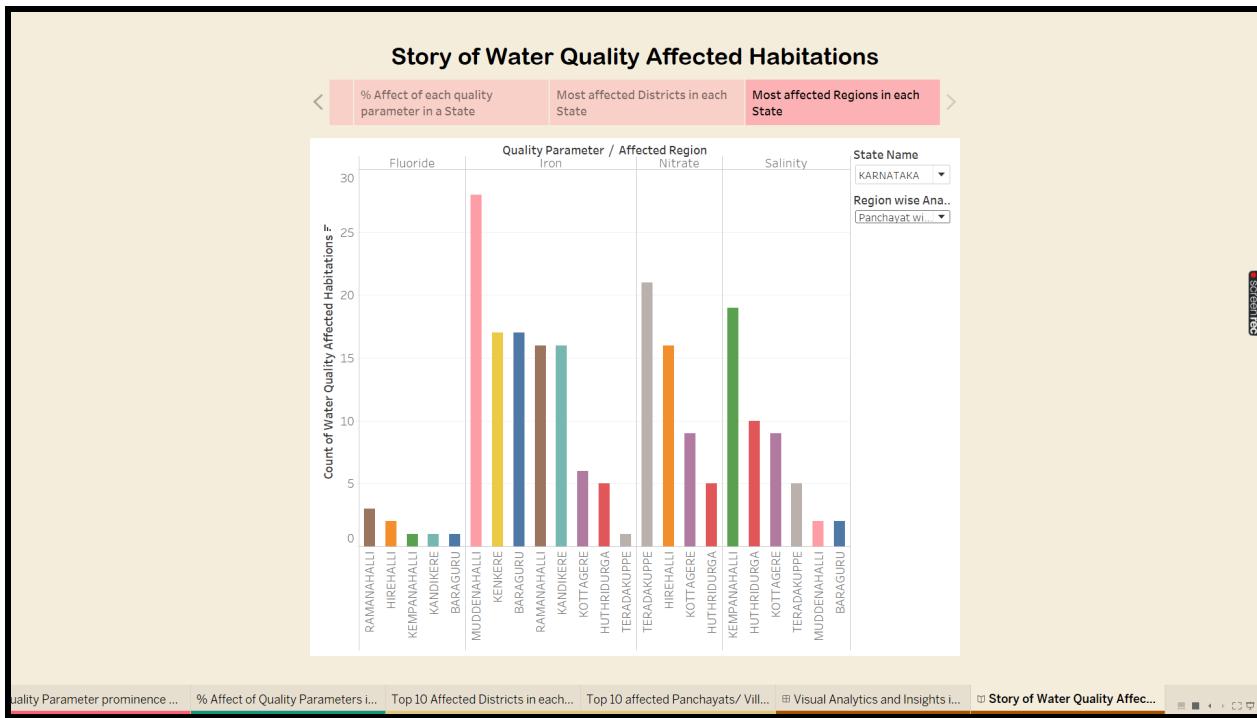
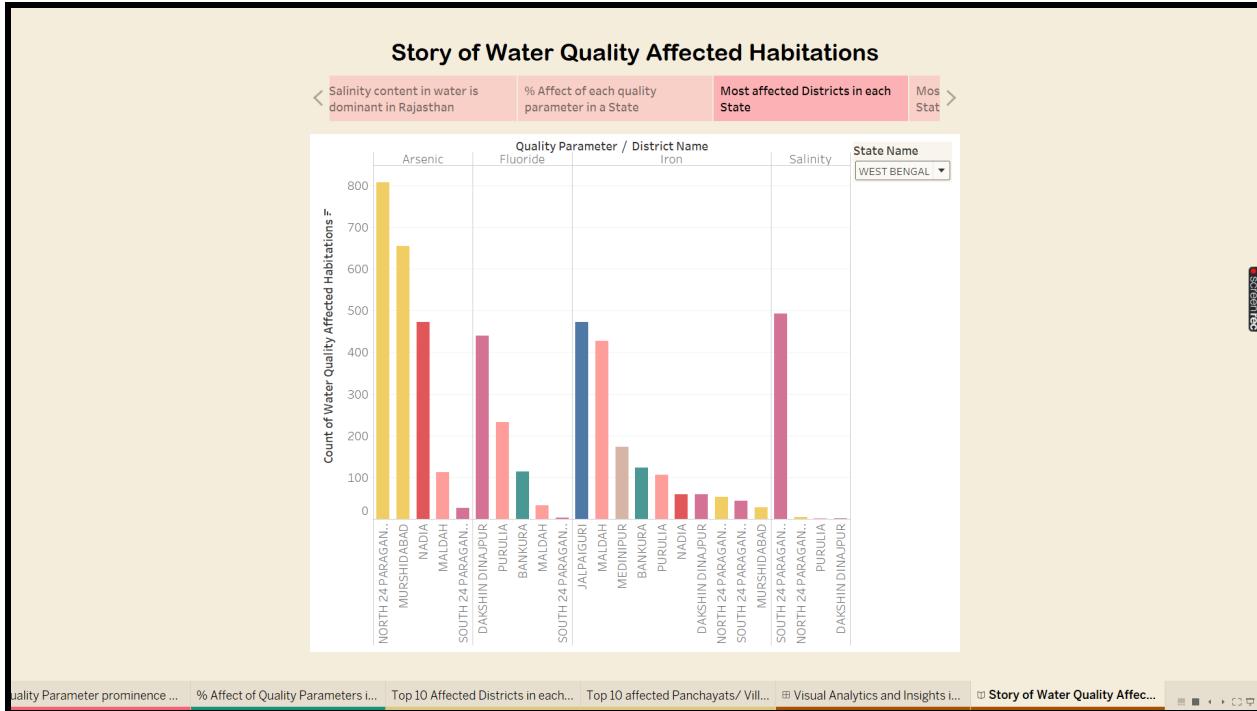
## 12) Story











## **ADVANTAGES**

- ☞ The visualization gives more insights into data.
- ☞ These insights help the respective authorities to take a proper decision.
- ☞ If a necessary action is taken to develop the affected habitations, it would solve the problem of lack of water quality.

## **DISADVANTAGES**

If the dataset provided is not that accurate, then wrong visualizations would be produced, which in turn affects the decision-making i.e, it may lead to wrong decisions/wrong actions. Thus it is necessary that the dataset and the visualizations should be accurate.

## **APPLICATIONS**

- ☞ Data Analyzing
- ☞ Data Visualization
- ☞ Data Cleaning
- ☞ Exploratory Data Analysis
- ☞ Prediction/ forecasting
- ☞ Decision Making

## **CONCLUSION**

- Iron is the most affecting quality parameter
- Rajasthan is the most affected state due to lack of water quality
- Arsenic content in water is dominant in West Bengal
- Flouride content in water is dominant in Rajasthan
- Iron content in water is dominant in Assam
- Nitrate content in water is dominant in Karnataka
- Salinity content in water is dominant in Rajasthan
- % Affect of each quality parameter in a State
- Most affected Districts in each State
- Most affected Regions in each State

## **FUTURE SCOPE**

To take necessary actions in-order to improve water quality in India, these analytics help a lot.

## **BIBLIOGRAPHY**

[https://help.tableau.com/current/pro/desktop/en-us/gettingstarted\\_overview.htm](https://help.tableau.com/current/pro/desktop/en-us/gettingstarted_overview.htm)

The END

Thank You