**Practical 2:**

**Aim: Study of any data mining application and find dataset and techniques used in that application, Prepare schema diagram for that application.**

**Application : Education statistics is Indian schools**

**Context**

This dataset contains information about Indian School Education Statistics of the year 2013-2014 to 2015-2016. Many public datasets from Indian Government are scattered and the goal here is to have all those datasets under one umbrella so that it is easy for beginners to find important datasets like this to start their Data Science journey.

**Content**

1. Percentage of Schools with Drinking Water Facility from 2013-14 to 2015-16
2. Gross Enrolment Ratio from 2013-14 to 2015-16
3. Drop-out rate from 2012-13 to 2014-15
4. Percentage of Schools with Computers from 2013-14 to 2015-16
5. Percentage of Schools with Electricity from 2013-14 to 2015-16
6. Schools with Boys Toilet from 2013-14 to 2015-16
7. Schools with Girls Toilet from 2013-14 to 2015-16

**Exploratory Data Analysis (EDA)**

Although many of you might be familiar with what EDA is all about, I would like to take a minute and give a formal definition of EDA and set the tone for this notebook for beginners and experts as well.

In statistics, exploratory data analysis is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task. Exploratory data analysis was promoted by John Tukey to encourage statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiments.

**What do you gain through EDA?**

Below are few among many points where EDA helps us:

1. maximize insight into a data set;
2. uncover underlying structure;
3. extract important variables;
4. detect outliers and anomalies;
5. test underlying assumptions;
6. develop parsimonious models; and
7. determine optimal factor settings.

**EDA Techniques**

Most EDA techniques are graphical in nature with a few quantitative techniques. The reason for the heavy reliance on graphics is that by its very nature the main role of EDA is to open-mindedly explore, and graphics gives the analysts unparalleled power to do so, enticing the data to reveal its structural secrets, and being always ready to gain some new, often unsuspected, insight into the data.

The particular graphical techniques employed in EDA are often quite simple, consisting of various techniques of:

Plotting the raw data (such as data traces, histograms, bihistograms, probability plots, lag plots, block plots, and Youden plots.

Plotting simple statistics such as mean plots, standard deviation plots, box plots, and main effects plots of the raw data.

Positioning such plots so as to maximize our natural pattern-recognition abilities, such as using multiple plots per page.

**Gross Enrollment Ratio**

Gross Enrolment Ratio (GER) or Gross Enrolment Index (GEI) is a statistical measure used in the education sector, to determine the number of students enrolled in school at several different grade levels (like elementary, middle school and high school), and use it to show the ratio of the number of students who live in that country to those who qualify for the particular grade level.

The GER can be over 100% as it includes students who may be older or younger than the official age group.

For instance, in India it improved from 25.8 to 26.3, the GER includes students who are repeating a grade, those who enrolled late and are older than their classmates, or those who have advanced quickly and are younger than their classmates. This allows the total enrolment to exceed the population that corresponds to that level of education.

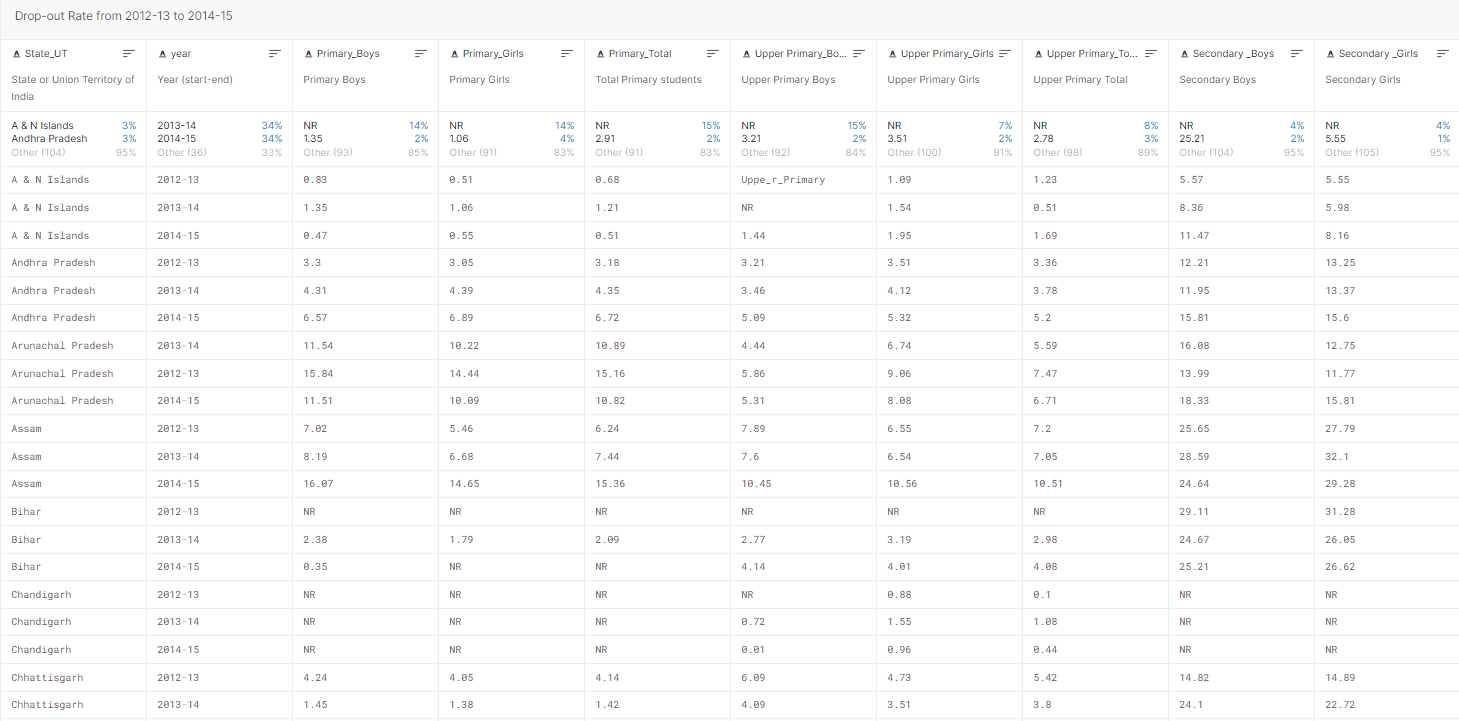
Calculation Method

a = number of students enrolled in a given level

b = population of the age group corresponds to given level of education India

GER=a/b×100

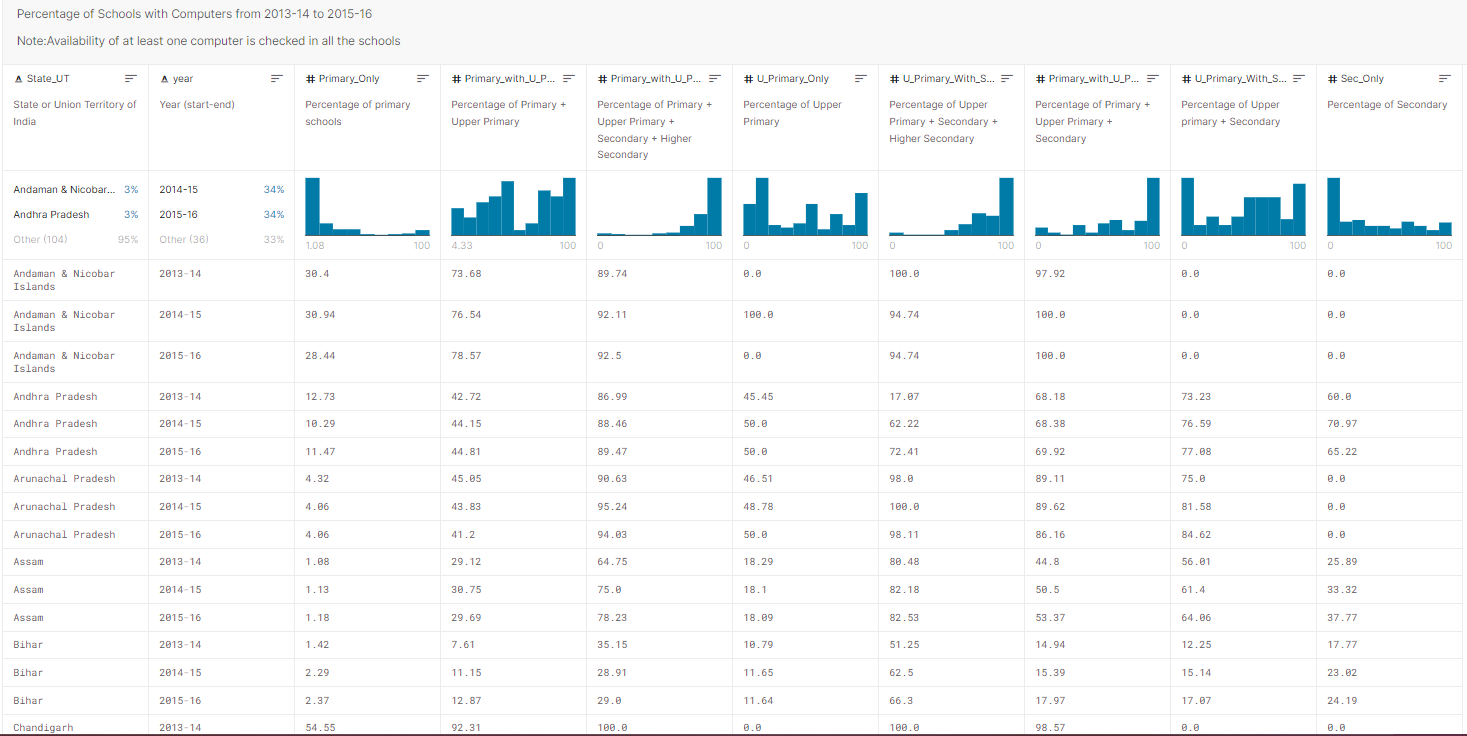
1. Dropout ratio from 2012-13 to 2014-15



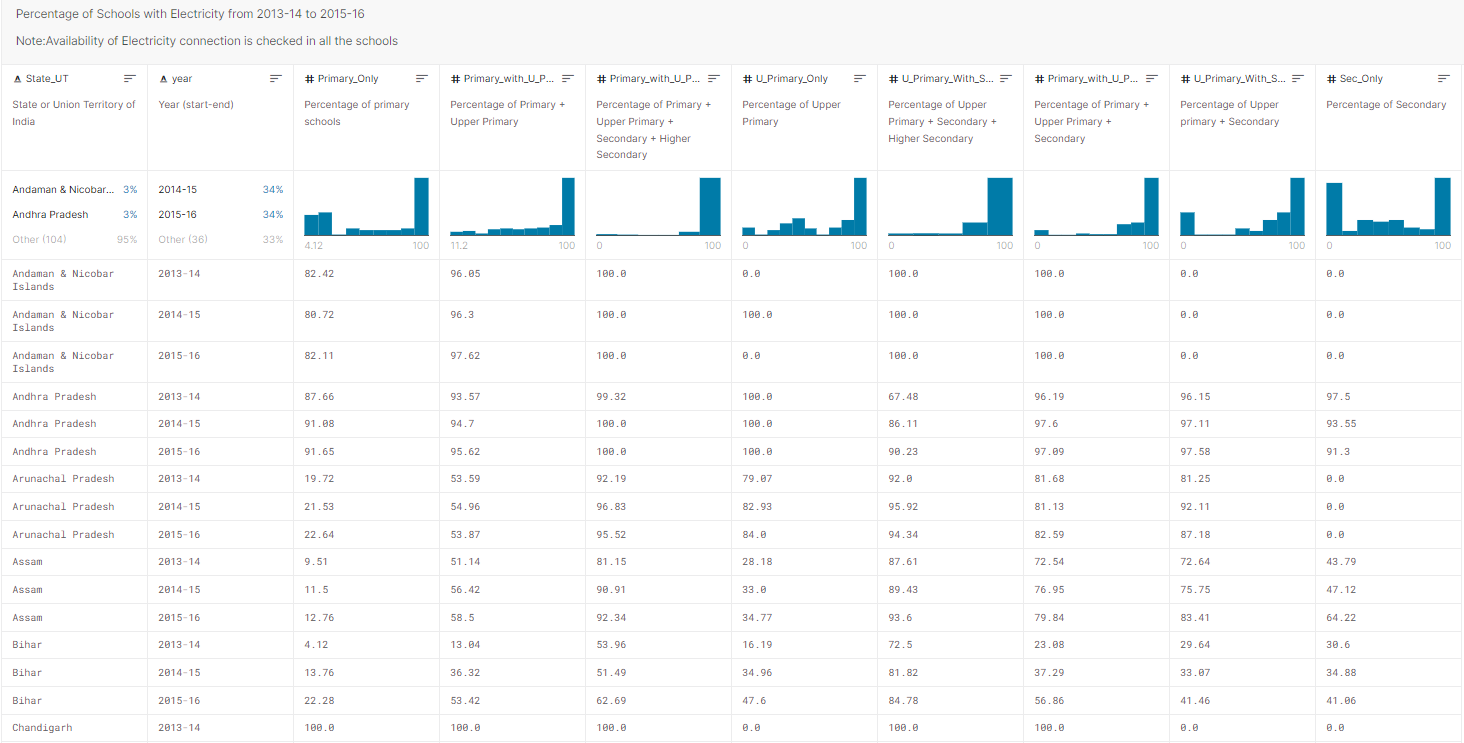
1. Gross Enrolment Ratio from 2013-14 to 2015-16



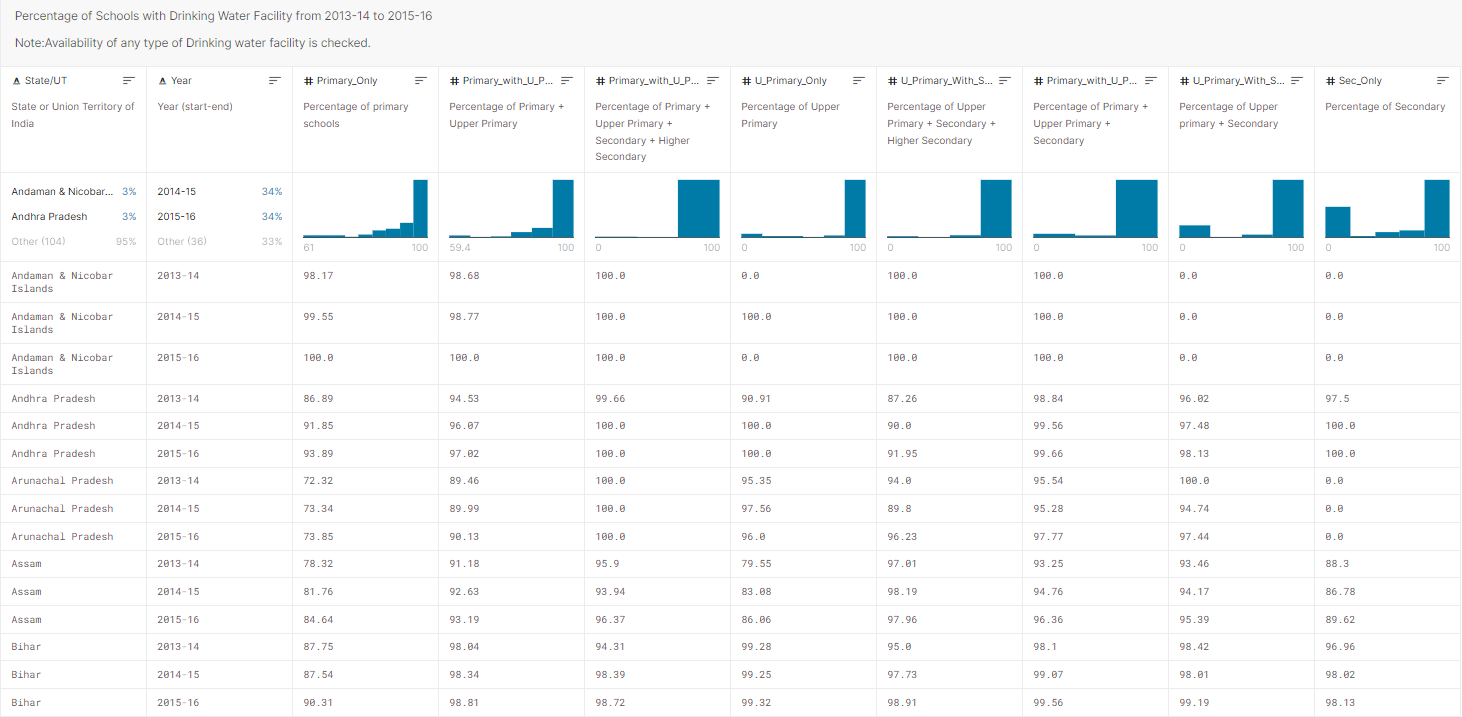
1. Percentage of Schools with Computers from 2013-14 to 2015-16



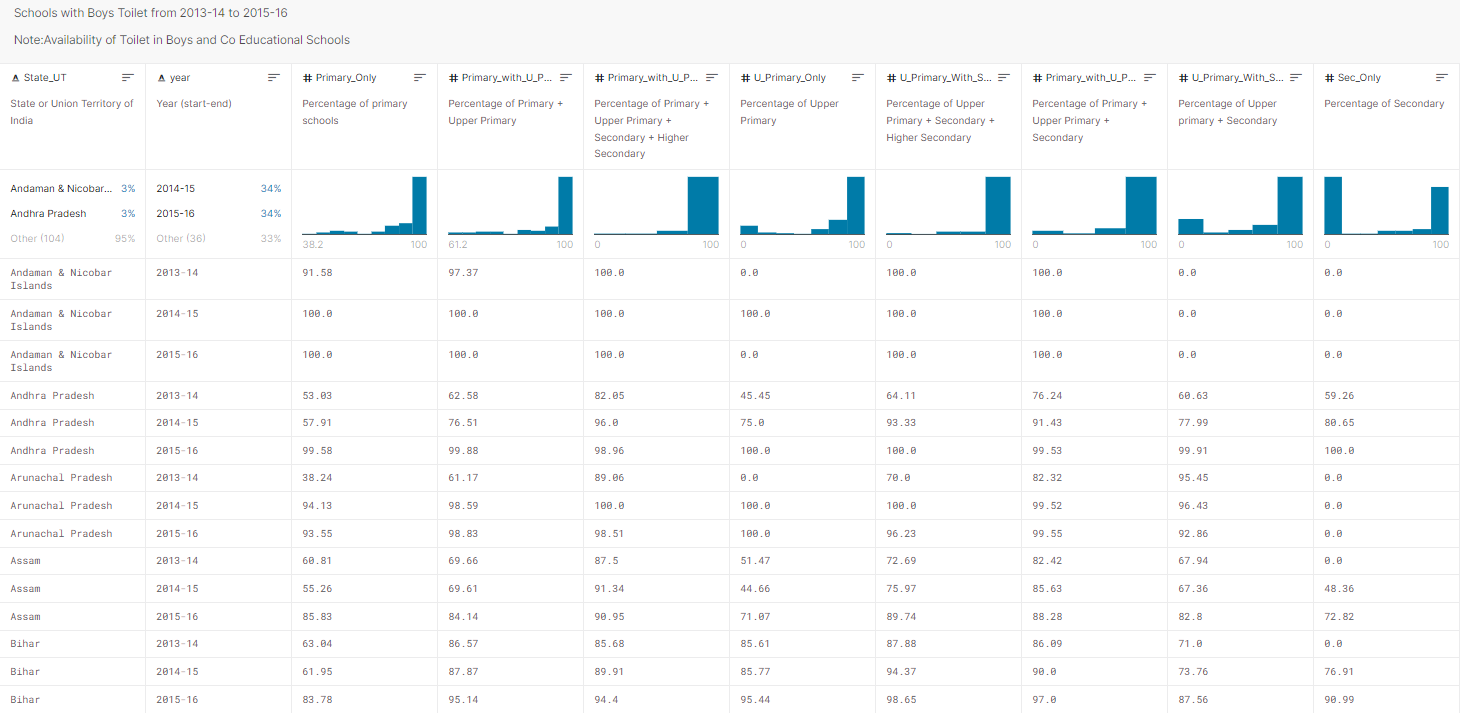
1. Percentage of Schools with Electricity from 2013-14 to 2015-16



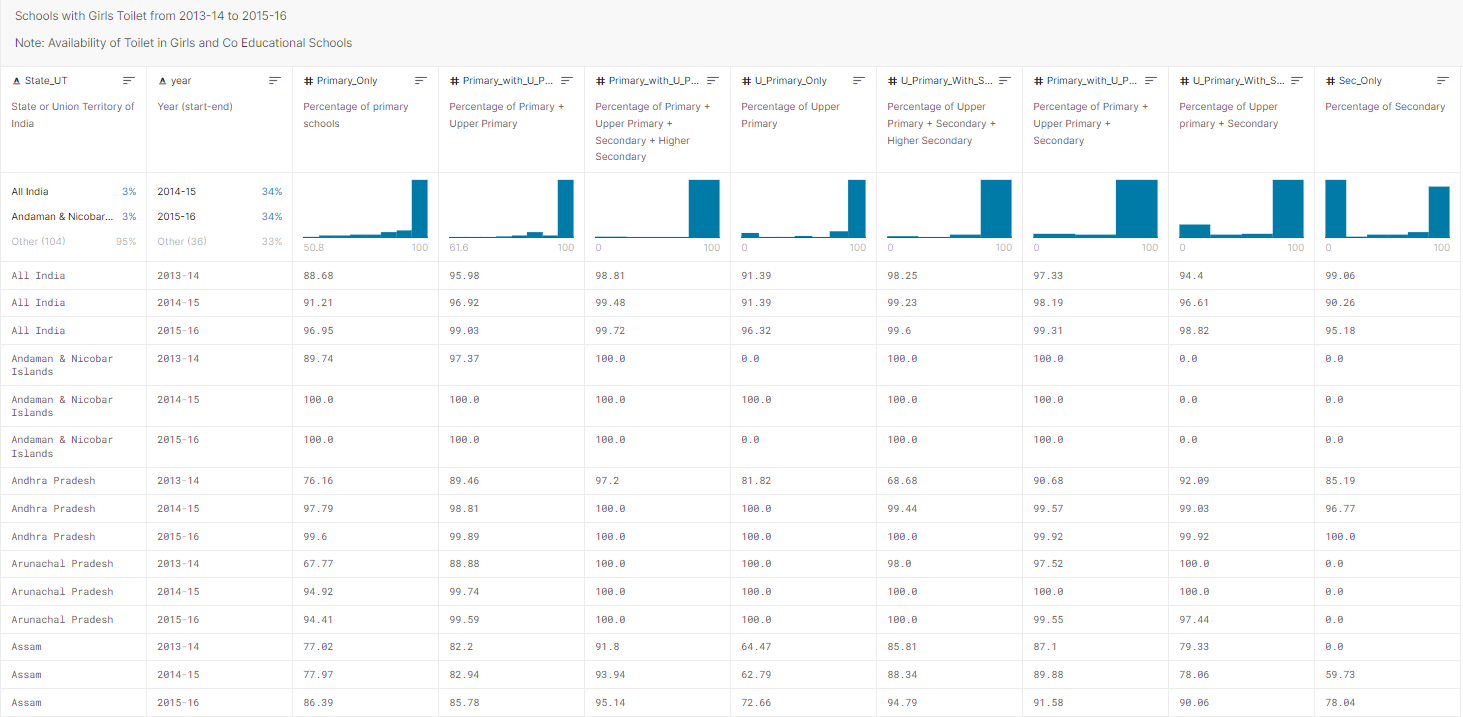
1. Percentage of Schools with Drinking Water Facility from 2013-14 to 2015-16



1. Schools with Boys Toilet from 2013-14 to 2015-16



1. Schools with Girls Toilet from 2013-14 to 2015-16



Schema diagram (Star schema)

