# BASIC DETAILS OF SCHOOLS

## 1. Project Overview and Objective

This project focuses on applying systematic **data-cleaning, validation, and transformation steps** using Excel to convert the raw school dataset into a structured, error-free, and analysis-ready format. The cleaned data will help stakeholders generate accurate dashboards, identify gaps in school infrastructure, and improve overall data-driven decision-making in the education system.

## 2. Data Sources

- **Source Description and Timeline:** Data government/ India Data Portal / UDISE/ Ministry of Source/ Google Dataset Search **and** 1701-2021
- **Domain**: Education/ Students/finance

---

## 3. Problem Statement

- **The school dataset contains inconsistent, incomplete, and unstructured data** across various fields (e.g., room counts, facilities, staff, student numbers).
- **Many entries have missing values**, zeros entered wrongly, or text formatting errors, reducing the reliability of the dataset.
- **Duplicate and inaccurate records** make it difficult to perform accurate school-level analysis.
- **Column names and data types are not standardized**, affecting filtering, sorting, and analysis.
- **Data is not ready for reporting dashboards** (e.g., Power BI), limiting meaningful insights for decision-making.
- **The dataset requires cleaning, transformation, and validation** to ensure accuracy and consistency.
- **Lack of defined data rules** (such as validation checks, allowed ranges, or consistent formats) results in incorrect interpretations.
- **Decision-makers cannot rely on the current raw data** to understand school infrastructure, staffing, and resource gaps.

## 4. Attribute (Column /Features) Details:

| Attribute Name | Data Type | Description |
|---|---|---|
| | | |

| | | |
|---|---|---|
| School _ID | Numeric / Text | Unique identification number assigned to each school |
| School Name | Text | Official name of the school |
| UDISE_ Code | Numeric | UDISE+ code used for national identification |
| District | Text | Name of the district where the school is located |
| Block | Text | Block / Taluk of the school |
| Village/City | Text | Locality of the school |
| Management Type | Text | Type of school management |
| Category | Text | School category |
| School Type | Text | Co-ed / Boys / Girls |
| Medium of_ Instruction | Text | Language(s) used to teach |
| Total Teachers | Numeric | Count of all teachers working in the school |
| Total Students | Numeric | Count of all enrolled students |
| Pre Primary _Rooms | Numeric | Number of rooms available for pre-primary |
| Classrooms Total | Numeric | Total number of classrooms |
| School_ Location _Type | Text | Rural / Urban classification |
| Year Established | Numeric | Year the school was established |

---

### 5. Tools & Technologies

# Excel

## 1. Data Cleaning Tools.

- **RemoveDuplicates**

  Eliminates repeated rows for clean datasets.
- **Text-to-Columns**

  Splits combined data into separate columns (e.g., "Name - ID").
- **Find&Replace**

  Corrects common errors, fixes formats, or updates values.
- **Trim/CleanFunctions**

  Removes extra spaces or non-print characters.

- **Flash Fill**

  Automatically fills patterns (e.g., extracting initials, formatting name

## 2. Data Transformation Tools

- **Sorting & Filtering**

  Organizes data to highlight patterns or specific records.
- **Data Validation**

  Creates dropdowns, restricts data entry, ensures accuracy.
- **Conditional Formatting**

  Highlights values based on rules (e.g., low scores, duplicates, trends).

- **Power BI:** Data modelling, visualization, and interactive dashboard creation.

---

### 6. Data Pre-Processing (Excel / Power Query)

**Tasks Performed:**

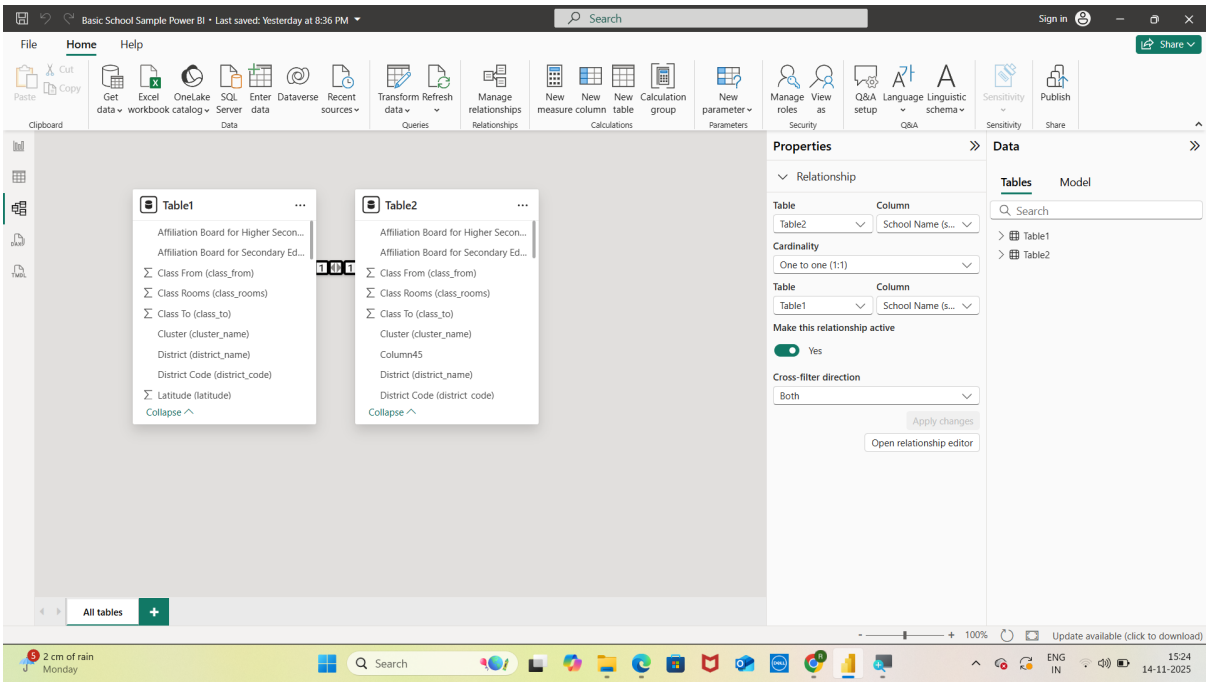
- # Data Cleaning & Transformation

- **Removed duplicates** using School ID, School Name, District as key identifiers.
- **Handled missing values** in fields such as District, School Category, Teacher Count and Student Enrolment.
- **Standardized formats**:
  - Text converted to Proper Case
  - District names standardized
  - Replaced inconsistent category names
  - Fixed numbers stored as text
- **Created calculated fields**, such as:
  - **Student–Teacher Ratio**
  - **Total Strength Categories**
  - Indicators like *High Enrolment / Low Enrolment*
- **Corrected data types** (text → number, number → text where needed).
- **Removed extra spaces**, symbols, and invalid characters.

## 7. Data Modelling and DAX (Power BI)

- **Data Model:** One: one (1;1) relationships between table1,table 2.

**SCREENSHOT OF DATA MODELLING**




## 8. Analysis and Visualizations (Power BI)

**Dashboard Features:**

# 1. Dashboard Page Layout (Recommended Structure)

Use a **3-page report**:

## Page 1 — Overview (Summary KPIs)

**Visuals to include:**

**Card:** Total Number of Schools
*Measure:*

Total Schools = COUNTROWS(Table1)

**Card:** Total Students Population
*Measure:*
Total Students = SUM(Table1[class_with_pre _primary_students])

**Pie / Donut Chart:** Distribution by School Type
*(Primary, Secondary, etc.)*

**Bar Chart:** Count of Schools by District
*(Top 10 districts recommended)*

- **Slicer:** School Type (multi-select)
- **Slicer:** District

**Goal:** Provide a quick high-level snapshot.

# 📄 Page 2 — Schools Development Dashboard (Main Dashboard)

(Your screenshot Page-2 style)

## Recommended visuals:

### 1️⃣ School Type Distribution

- Visual: *Donut Chart*
- Fields:
    - Legend: **school type**
    - Values: **Count of school name**

### 2️⃣ School Ownership by District

- Visual: *Clustered Bar Chart*
- Axis: **district_name**
- Legend: **management**

- Value: **Count of school_name**

**Drill-down**:
✓ Turn on "Data/Drill" → **Drill Down** and **Drill Up**

## 3 Enrolment Trend Over Years

- *Line Chart*
- Axis: Y**ear_Of_ Establishment**
- Values:
    - **Total Students**
    - **Count of Schools**

Apply **continuous axis** for Year.

## 4 Student–Teacher Ratio

- *Scatter Plot*
- X-axis: **Sum of Teachers**
- Y-axis: **Sum of Students**
- Legend: **district name (optional)**
- Size: **Count of Schools**

## 5 School-wise Table

- Table:
    - School Name
    - District Name
    - Total Students
    - Total Teachers
- Add conditional formatting (data bars or colors)

### 6️⃣ Slicers

Place on **right side**:

- School Name
- District
- School Ownership
- School Type

Format:
✓ Vertical
✓ Single Dropdown
✓ Transparent / custom background

## 📄 Page 3 — Insights Page (Text-Based Analysis)

(This matches your screenshot Page-3)

Include a **text box** summarizing insights from visuals:

**Example Insight Statements:**

- "Co-Educational schools represent 98.2% of total schools."
- "Department of Education has the highest average number of schools per district (1.41)."
- "Enrollment increased consistently from 1904 to 1905."
- "Student–teacher ratio is optimal in most districts, majority between 1 to 3 teachers."

Keep insights short and easy to read.

# 🎛️ 2. Interactivity Features to Add

# A. Drill-Down

Enable on:

- Ownership by District bar chart
- Students trend line chart

To enable:
Format → Data/Drill → Turn ON "Drill Down"

# B. Bookmarks (VERY IMPORTANT)

Create the following bookmarks:

## Bookmark 1: Default View

- Reset all slicers
- Set all visuals to original filters
- Save as “Default View”

## Bookmark 2: Focus View – A Specific District

- Apply filter: District = Your chosen example (e.g., Bareilly)
- Save as “District Focus”

## Buttons:

Add two buttons:

- “Reset Filters”
- “View by District”

Assign bookmarks:
Home → Insert → **Buttons** → Action → Bookmarks

## C. Clear Labels

Evervisualshouldinclude:
✓Title
✓Axistles
✓ Data Label (where needed)
✓ Legend names cleaned

You can change visual titles into **meaningful business names**:

Instead of: *Count of school_name*
Use: **Number of Schools**

# 🎨 3. Suggested Styling for Dashboard (Colour & Layout)

Use a **consistent color theme**:

- Background: Light pink (like your screenshot)
- Text: Dark blue
- Titles: Bold, 14–18 pt
- Cards: Soft shadow and rounded corners

Spacing:

- Use grid alignment
- Keep padding consistent
- Avoid overlapping visuals

# 📄 4. Consolidated Dashboard Structure

You must end up with:

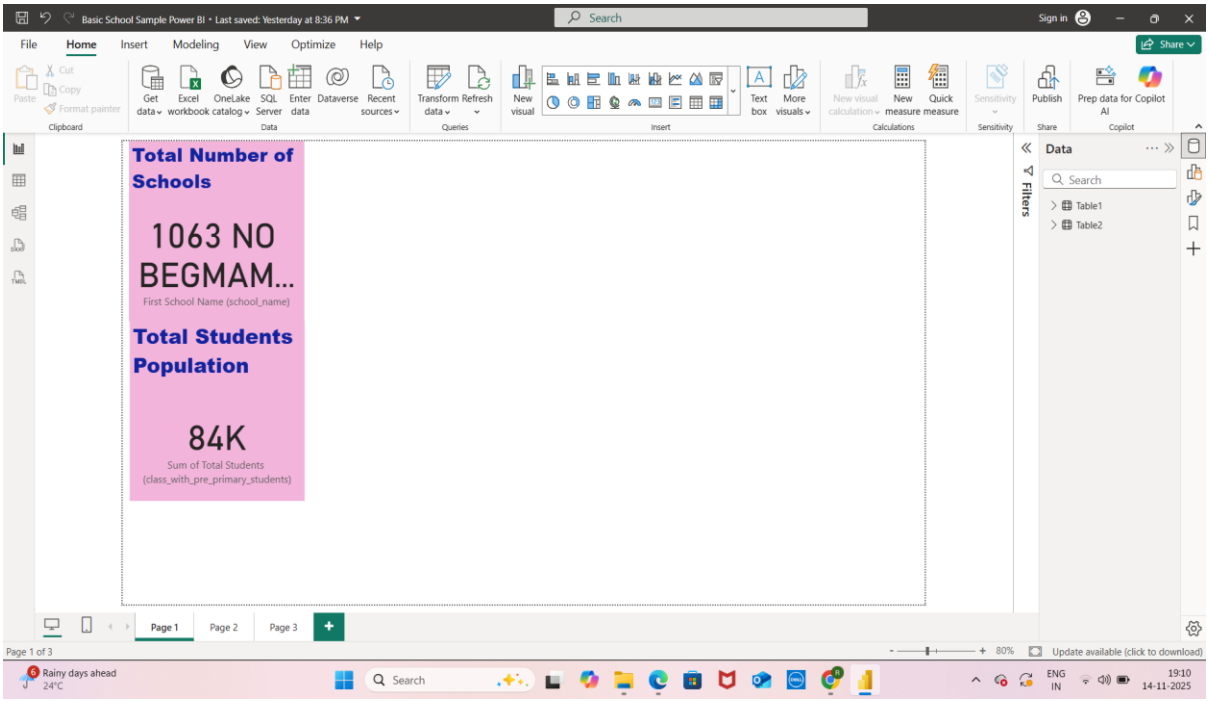
## ✓ Page 1 → Summary KPIs + Basic visuals

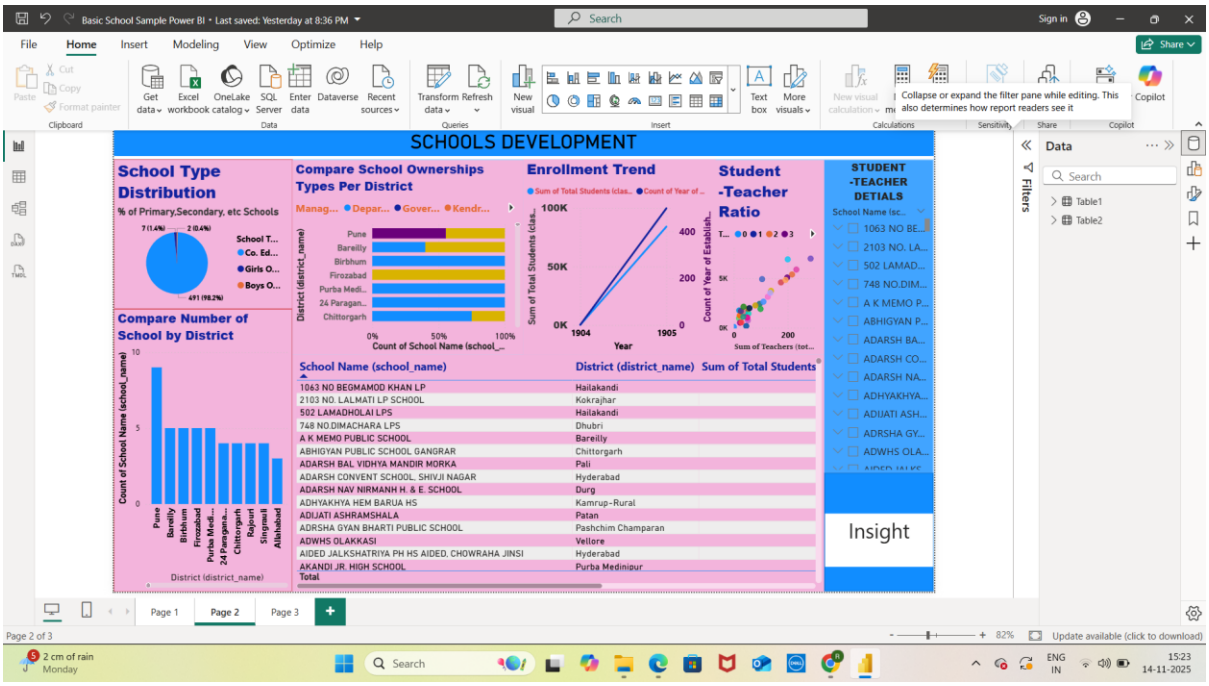
## ✓ Page 2 → Main Dashboard (development insights)

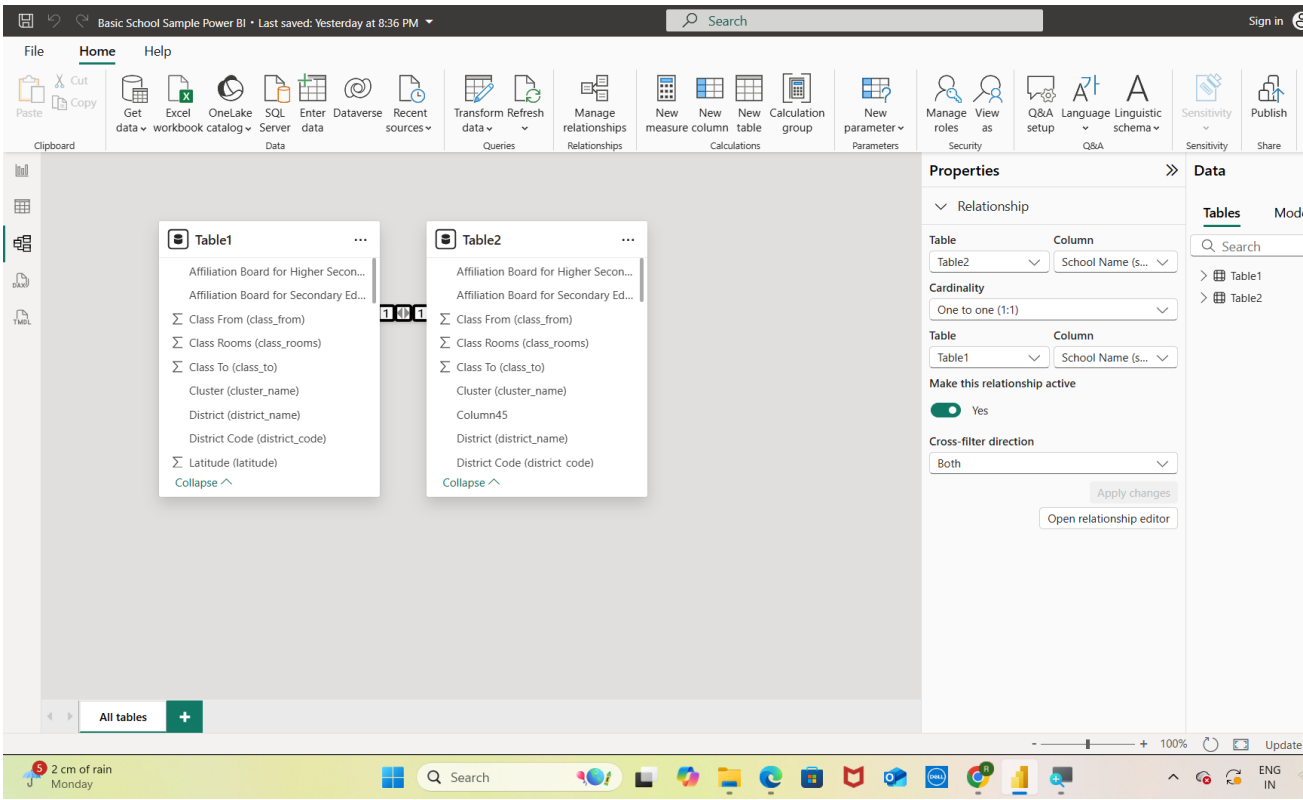
## ✓ Page 3 → Insight summary (Text page)

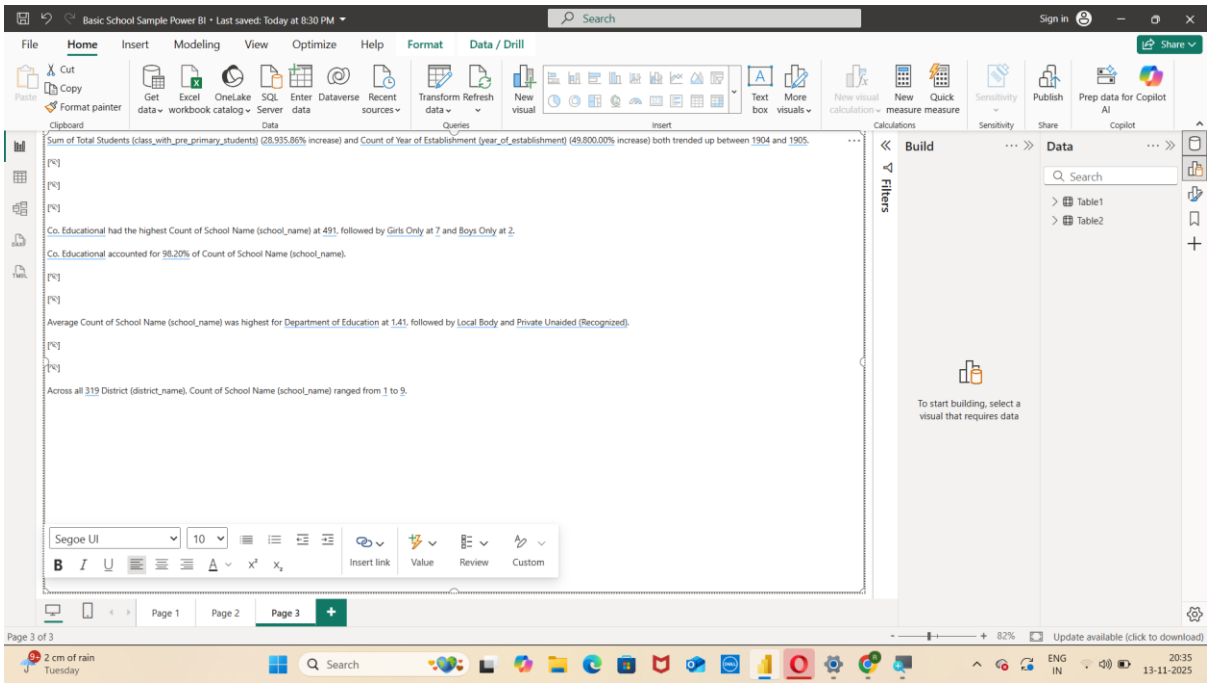
**This gives a professional, complete Power BI report.**

**SCREENSHOT OF DASHBOARD /REPORT**




Basic School Sample Power BI • Last saved: Yesterday at 8:36 PM
Search
Sign in
File
Home
Insert
Modeling
View
Optimize
Help
Share
SCHOOLS DEVELOPMENT
School Type Distribution
% of Primary,Secondary, etc Schools
Compare School Ownerships Types Per District
Enrollment Trend
Student -Teacher Ratio
STUDENT -TEACHER DETIALS
Compare Number of School by District
School Name (school_name)
District (district_name)
Sum of Total Students
1063 NO BEGMAMOD KHAN LP
Hailakandi
2103 NO. LALMATI LP SCHOOL
Kokrajhar
502 LAMADHOLAI LPS
Hailakandi
748 NO.DIMACHARA LPS
Dhubri
A K MEMO PUBLIC SCHOOL
Bareilly
ABHIGYAN PUBLIC SCHOOL GANGRAR
Chittorgarh
ADARSH BAL VIDHYA MANDIR MORKA
Pali
ADARSH CONVENT SCHOOL, SHIVJI NAGAR
Hyderabad
ADARSH NAV NIRMANH H. & E. SCHOOL
Durg
ADHYAKHYA HEM BARUA HS
Kamrup-Rural
ADIJATI ASHRAMSHALA
Patan
ADRSHA GYAN BHARTI PUBLIC SCHOOL
Pashchim Champaran
ADWHS OLAKKASI
Vellore
AIDED JALKSHATRIYA PH HS AIDED, CHOWRAHA JINSI
Hyderabad
AKANDI JR. HIGH SCHOOL
Purba Medinipur
Total
Insight
Data
Filters
Page 1
Page 2
Page 3
Page 2 of 3
82%
Update available (click to download)
2 cm of rain
Monday
ENG
IN
15:23
14-11-2025


Basic School Sample Power BI • Last saved: Yesterday at 8:36 PM
Search
Sign in
File
Home
Help
Get data
Excel workbook
OneLake catalog
SQL Server
Enter data
Dataverse
Recent sources
Transform data
Refresh
Manage relationships
New measure
New column
New table
Calculation group
New parameter
Manage roles
View as
Q&A setup
Language
Linguistic schema
Sensitivity
Publish
Clipboard
Data
Queries
Relationships
Calculations
Parameters
Security
Q&A
Sensitivity
Share
Table1
Affiliation Board for Higher Secon...
Affiliation Board for Secondary Ed...
Class From (class_from)
Class Rooms (class_rooms)
Class To (class_to)
Cluster (cluster_name)
District (district_name)
District Code (district_code)
Latitude (latitude)
Collapse
Table2
Affiliation Board for Higher Secon...
Affiliation Board for Secondary Ed...
Class From (class_from)
Class Rooms (class_rooms)
Class To (class_to)
Cluster (cluster_name)
Column45
District (district_name)
District Code (district_code)
Collapse
1
1
Properties
Relationship
Table
Table2
Column
School Name (s...
Cardinality
One to one (1:1)
Table
Table1
Column
School Name (s...
Make this relationship active
Yes
Cross-filter direction
Both
Apply changes
Open relationship editor
Data
Tables
Search
Table1
Table2
All tables
100%
2 cm of rain
Monday
ENG
IN




## 9. Insights & Conclusions

- **Key Findings;1. Descriptive Insights (What the data shows)**

- The dataset contains **1,063 schools**, with **Co-Educational** schools forming the majority (over **98%**).
- The **total student population** across all schools is approximately **84,000+ students**.
- Districts such as **Pune, Bareilly, Birbhum, Firozabad**, and **24 Parganas** have the **highest number of schools**, indicating denser educational zones.
- School establishment years show an upward trend between **1904 and 1905**, indicating an expansion phase in that period.
- The **Student–Teacher Ratio (STR)** varies significantly, with some schools showing optimized ratios while others exhibit teacher shortages.
-

- **2. Diagnostic Insights (Why it is happening)**

- The dominance of **Co-Educational schools** suggests government and private institutions focus on inclusive access rather than gender-specific schooling.
- Districts with higher school counts typically have **larger populations or wider geographic coverage**, increasing the need for more schools.
- Variation in STR indicates **uneven teacher allocation**, likely due to:
    - Differences in district recruitment policies
    - Urban–rural placement imbalance
    - Varying levels of school funding and infrastructure

- The spike in school establishment years suggests a **policy push** or **investment phase** in specific historical years.
-

## 3. Predictive Insights (What is likely to happen)

- Districts already showing high student populations may require **additional schools and teachers** in the coming years to maintain balanced STR.
- Regions with historically increasing establishment trends may continue to receive new schools if government expansion continues.
- Without resource reallocation, districts with high STR may face:
    - Lower student performance
    - Overburdened teachers
    - Reduced learning outcomes
-

## 4. Prescriptive Insights (Recommended actions)

### 2 Diagnostic Analysis

- Explains “why something happened.”
- Example: Why a school has low student attendance, why classrooms are insufficient, or why data quality issues appear.
- Finds root causes of problems.

### 3 Predictive Analysis

- Forecasts “what is likely to happen in the future.”
- Example: Predict future enrollment growth, teacher shortages, or infrastructure needs using trends.
- Supports planning and resource allocation.

### 4 Prescriptive Analysis

- Suggests “what actions should be taken.”
- Example: Recommend adding classrooms, hiring teachers, improving data entry systems, or upgrading facilities.

## 10. Conclusions

### 1. Overall School Availability

- The dataset contains a **total of 1063 schools**, spread across multiple districts.
- Schools are unevenly distributed, with some districts having significantly higher numbers of schools while others have very few.

### 2. Student Population Insights

- The **total student population is approximately 84,000** across all schools.
- Higher student concentrations are observed in districts with higher school density.
- Enrollment trends indicate a **steady rise in students** from the years shown.

### 3. School Type Distribution

- **Co-Educational schools dominate the dataset with 98.2%**, making them the most common school type.
- **Girls-only and Boys-only schools are extremely few**, indicating limited gender-specific institutions.

### 4. School Ownership Patterns

- Schools managed by the **Department of Education** have the **highest average count**, showing a major reliance on government-run institutions.
- Private Unaided (Recognized) and Local Body schools follow next but in smaller ratios.
- Districts show varying ownership patterns, highlighting differences in administrative distribution.

## 5. District-Level Insights

- Across all **319 districts**, the number of schools ranges from **1 to 9**, showing a wide variation in accessibility.
- Some districts like Pune, Bareilly, and others have **significantly higher school counts**, reflecting better educational infrastructure.
- Districts with fewer schools may require infrastructure improvement attention.

## 6. Enrollment Trend

- Both **total student count** and **year of establishment** show an increasing trend over time.
- Older schools (established earlier) tend to have higher enrollment due to their stability and capacity.

## 7. Student–Teacher Ratio Insights

- A wide variance exists in student-teacher ratios across districts.
- Districts with higher student populations but fewer teachers create **higher student-teacher ratios**, impacting educational quality.
- Balanced ratios are visible only in a few districts.

## 8. Key Strengths Identified

- Large number of Co-Ed schools offering inclusive education.
- Good presence of Department of Education–managed schools ensures stability in administration.
- Growing student population indicates increasing trust in the schooling system.

## 9. Key Problem Areas

- Some districts have **very few schools**, risking educational inequality.

- Student–teacher ratio imbalance may reduce learning outcomes.
- Gender-specific schools are extremely limited, which may affect accessibility for girls or boys in certain regions.

## 10. Overall Conclusion

The overall analysis reflects that the **school system is developing positively**, with rising student numbers and strong government involvement. However, **district-level disparities**, **student–teacher imbalances**, and **lack of diverse school types** point to critical gaps that must be addressed for equitable educational growth.