**Public Schools Dataset:**

1. **The Problem of the Dataset**

This dataset highlights the distribution and accessibility of public schools,

which can reveal disparities in education access, particularly in

underserved regions.

2. **Reason Behind Selection**

It provides critical insights into how well public schools are distributed

across different regions, making it ideal for spatial and accessibility

analysis.

3. **Problem Being Solved**

The analysis will focus on understanding public schools' geographic

distribution and accessibility to students. It can also help identify areas

needing more educational resources.

**Key Columns**:

**School Name** (Categorical): Identifies each public school.

**Category** (Categorical): School type (high, middle, elementary).

**ZIP Code** (Categorical): Helps analyse distribution by region.

**Longitude/Latitude** (Numerical): Used for mapping and spatial analysis.

**Phone, Address** (Categorical): Useful for contacting or mapping

schools.

4. **Data Cleaning Techniques**

* **Geolocation Cleaning**: Verify and correct longitude and latitude data.
* **Handling Duplicates**: Removing duplicate entries for schools.
* **Standardisation**: Ensuring consistency in categories (e.g., "High
* School" vs "HS").
* Use of uniform text case to ensure uniformity and consistency
* Break down data in smaller segments for better understanding – Location has both latitudes and latitudes
* Keeping track of any updates done to the dataset during data cleaning