Datasets-223039659-KA

Mental Health Dataset:

1. **The Problem of the Dataset**

   The dataset focuses on mental health issues among university students, which are often overlooked and lead to lower academic performance and well-being.

2. **Reason Behind Selection**

   Mental health in educational settings is crucial, yet not discussed enough. This dataset allows for an in-depth analysis of how various academic, lifestyle and social factors contribute to students' mental health.

3. **Problem Being Solved**

   The analysis will help identify correlations between mental health issues (like depression, and anxiety) and academic factors (such as workload, and CGPA), offering insights into gaps in student support.

   Key Columns:

   - **Gender** (Categorical): Analyses mental health trends across genders.
   - **Age** (Numerical): Looks at the correlation between age and mental health.
   - **CGPA** (Categorical): Reflects academic performance, grouped into ranges.
   - **Sleep Hours, Sports Engagement** (Numerical/Categorical): Tracks students’ lifestyle balance.
   - **Depression, Anxiety, Isolation** (Categorical): Quantifies students' mental health status.
   - **Stress Relief Activities** (Categorical): Activities students use to cope with stress, offering insights into well-being.

4. **Data Cleaning Techniques**

   - **Handling Missing Values**: Imputing missing demographic or lifestyle data.
   - **Outlier Detection**: Identifying extreme values in CGPA or sleep hours.

- **Categorical Standardisation**: Ensuring consistency in gender, degree level, and stress-relief activity labels.

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.

   It also highlights that accessibility of public schools in rural areas are not that likely.

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.

   With helping areas that need more educational resources, it could encourage the investment in public schools in rural areas to ensure fair access to education.

   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").

     Identify values that are very different from the rest of the data and decide whether to remove them or chnage them completely.