Title: Data Cleaning for Public Health Awareness Campaigns Analysis

Project Description: The Data Cleaning project is an essential phase in our broader effort to analyze public health awareness campaigns. It involves preparing the dataset, which includes information about these campaigns, for subsequent analysis. The project aims to ensure data accuracy, consistency, and reliability to derive meaningful insights. By addressing issues such as missing data, duplicates, and outliers, we will create a clean dataset suitable for robust analysis.

Project Objectives:

Data Assessment: Review the dataset to understand its content, structure, and relevance to public health awareness campaigns.

Missing Data Handling: Identify and address missing values, as missing data can impact the completeness of campaign records.

Duplicate Data Removal: Detect and eliminate duplicate campaign entries to maintain data integrity and avoid double-counting.

Data Type Conversion: Ensure that data types align with the nature of the campaign attributes. Convert data types if needed.

Outlier Detection: Identify and address outliers that may skew campaign performance metrics.

Data Integrity: Verify data integrity, address inconsistencies, and ensure that all entries align with public health campaigns.

Documentation: Maintain detailed documentation of the data cleaning processes for transparency and future reference.

Source:

Data set: https://in.docworkspace.com/d/sIAeL0f_iAaGd5KkG

Code:

Import pandas as pd

Load CSV data into a Pandas DataFrame

Data = pd.read_csv('survey.csv')

Display basic information about the dataset

Print("Data Info:")

```
Print(data.info())

# Check for missing values

Missing_values = data.isnull().sum()

Print("\nMissing Values:")

Print(missing_values)

# Check for duplicate rows

Duplicates = data.duplicated().sum()

Print("\nDuplicate Rows:")

Print(duplicates)

# Remove duplicate rows

Data.drop_duplicates(inplace=True)

# Export the cleaned data to a new CSV file

Data.to_csv('cleaned_data.csv', index=False)
```