**Day 2 Project report**

**Project name:** Health care

**Group:** Team B

**Team leader:** Achintharya A Patil

**Members:**

|  |  |
| --- | --- |
| **Task/Problem statement** | **Name** |
| 7 | Achintharya |
| 5&6 | Kapil |
| 9 | Aryan Naik |
| 8 | Ardhendu Debnath |

**Objective:** The project involves cleaning and processing census and housing data to extract relevant information for the healthcare department. Initial tasks include identifying and handling missing data, filling in missing values using information from other cells, and visualizing the data to compare missing data before and after processing. Further processing involves extracting information from the housing data, such as the total number of households, livable and dilapidated households, and toilet premises, by correlating it with census data. The processed data is then saved into a CSV file named "census.csv" and "housing.csv" in the "Clean data" folder, respectively. Finally, inconsistencies between redundant data in the housing and census datasets are identified and reported, ensuring data integrity for subsequent analysis.

Problem Statement 5: (Find and process Missing Data)   
Problem Statement 6: (Save Data)

Problem Statement 7: (Process the relevant data)

Problem Statement 8: (Visualize the housing data)

Problem Statement 9: (Inconsistency in different datasets)

**Processes done:**

1. Loaded dataset using pandas.
2. Selected required columns.
3. Renamed columns.
4. Filled missing data of one column using other columns.
5. Merged two datasets.
6. Derived absolute value using percentages and respective data of other dataset.
7. Reported redundant data and compared datasets.
8. Visualized aspects/ columns of dataset.
9. Saved data frames to csv file.

**Outcome:** Cleaned and processed dataset with relevant information extracted for the healthcare department. Missing data is identified and handled, with values filled in using information from other cells. Visualizations are created to compare missing data before and after processing. Information from the housing data, such as total households, livable and dilapidated households, and toilet premises, is extracted and correlated with census data. The processed data is saved into separate CSV files named "census.csv" and "housing.csv" in the "Clean Data" folder. Inconsistencies between redundant data in housing and census datasets are identified, ensuring data integrity for subsequent analysis.