Data Cleaning Steps

1. Data Loading and Initial Inspection

The dataset was imported into the Jupyter Notebook using Python libraries. Initial inspection was performed using functions such as head(), info(), and describe() to understand the structure, data types, and summary statistics of the dataset.

---

2. Identifying Missing Values

Missing values were identified using built-in functions to check null counts across all columns. This step helped in understanding which attributes required cleaning or further attention.

---

3. Data Type Review

The dataset's column data types were reviewed to ensure numerical and categorical attributes were correctly interpreted. This step ensured the dataset was suitable for further analysis.

---

4. Handling Inconsistent and Null Values

Based on the initial analysis, missing and inconsistent values were reviewed and handled appropriately to maintain data integrity and avoid errors during analysis.

---

5. Data Consistency Check

Basic consistency checks were performed to ensure the dataset followed logical patterns and did not contain invalid or misleading values that could affect analysis outcomes.

---

6. Prepared Dataset for Analysis

After completing the above steps, the dataset was considered ready for analytical processing and visualization tasks.

---

7. Final Data Validation

After cleaning, the dataset was revalidated to confirm:

- No missing values remained
- Correct data types across all columns
- Logical consistency of numerical values

---

8. Basic Data Exploration

Exploratory Data Analysis (EDA) was carried out using visualizations such as distribution plots, count plots, and box plots. This helped in understanding data distribution, detecting anomalies, and identifying potential outliers.

---

9. Exporting Cleaned Data

The final cleaned dataset was saved and prepared for further analysis, visualization, and reporting in SQL, Power BI, and other analytics tools.