

## Dataset Analysis Report – House Prices Dataset

**Task:** Data Cleaning & Missing Value Handling

**Dataset Used:** House Prices Dataset

### Objective

The objective of this task is to understand and implement data cleaning techniques, especially handling missing values, to improve the quality of a dataset for further analysis or machine learning tasks.

#### 1. Loaded the Dataset

The dataset was loaded using Pandas and basic exploration was done using `.head()`, `.info()`, and `.shape()`.

#### 2. Identified Missing Values

Missing values were identified using:

```
df.isnull().sum()
```

#### 3. Visualized Missing Data

Missing values were visualized using bar charts and heatmaps to understand patterns.

#### 4. Handled Missing Values

- Numerical columns were filled using **mean/median**.
- Categorical columns were filled using **mode**.

#### 5. Removed Columns with High Missing Values

Columns with more than 50% missing values were dropped.

#### 6. Validated the Dataset

After cleaning, the dataset was re-checked for missing values, shape, and data types.

#### 7. Compared Before and After Cleaning

Rows, columns, and overall data quality were compared before and after cleaning.

### Results

- Missing values were successfully handled.
- Dataset quality improved.
- Unnecessary columns were removed.
- The final dataset is ready for analysis or modeling.

### Conclusion

This task helped in understanding the importance of data preprocessing and how proper cleaning improves the reliability of data for future tasks like machine learning and analytics.