**Reading the Data**: Use pd.read\_csv() to read the data (5000 rows \* 16 columns)

**Handling Missing Values**: Use df.isnull() or df.info() to detect missing data, then apply strategies like removing rows (df.dropna()), filling with mean/median/mode (df.fillna()).

**Removing Duplicates**: Identify and remove duplicate rows with df.duplicated() and df.drop\_duplicates(), ensuring data consistency.

**Data Type Conversion**: Checked data types and no conversion was needed

**Outlier Detection**: Detect outliers using statistical methods like Z-scores or IQR, and handle them by either removing or capping the values within a range.

**Normalizing/Scaling Data**: Scaled numeric columns using libraries like Pandas techniques to ensure features have similar ranges, improving model performance.

**Clean Data** : df.to\_csv('Cleaned\_House\_dataset.csv', index=True) to save the cleaned data in csv ignoring the indexes (4990 rows \* 16 columns)