PREPROCESSING THE DATASET FOR HOUSE PRICE PREDICTION

INTRODUCTION:

This document outlines the preprocessing steps undertaken to prepare a comprehensive dataset for the task of house price prediction using machine learning. The dataset may be derived from various sources, and this document provides a detailed description of the data integration and preprocessing procedures.

DATA SOURCES:

The dataset comprises information from various sources, which may include real estate listings, property records, or other relevant data sources.

DATA INTEGRATION:

The initial step involves importing data from each source using the Pandas library in Python. The data files are read and stored in separate DataFrames. These DataFrames are then merged horizontally (column-wise) to create a consolidated dataset, ensuring that duplicate columns are removed to avoid redundancy.

DATA PREPROCESSING:

The preprocessing of the dataset involves several essential tasks, similar to those outlined in your previous document for energy consumption:

1. HANDLING MISSING VALUES:

- Address missing values using techniques such as interpolation, forward-fill, or backward-fill to ensure a complete dataset.

2. FEATURE ENGINEERING:

- Create additional features to enhance the dataset's predictive power. This may include transformations, scaling, or the creation of new derived features.

3. DATA TYPE CONVERSION:

- Check and modify data types to ensure consistency. In particular, non-numeric data types should be converted to numerical types to make them compatible with machine learning algorithms.

PROGRAM:

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

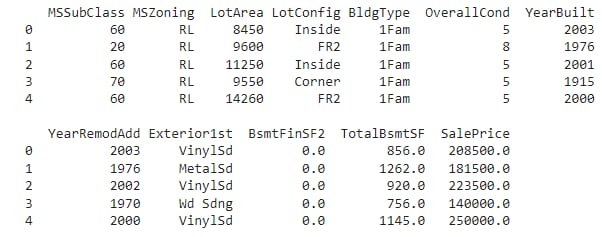
dataset = pd.read\_excel("HousePricePrediction.xlsx")

# Printing first 5 records of the dataset

print(dataset.head(5))

dataset.shape

RESULT:



CONCLUSION:

The successful preprocessing of the dataset is a critical step in any machine learning project, including house price prediction. By combining data from multiple sources and ensuring data quality, you have created a solid foundation for future research and modeling in the domain of predicting house prices.