HOUSE PRICE PREDICTION PROJECT- DEVELOPMENT PART 1

Introduction

House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house. There are three factors that influence the price of a house which include physical conditions, concept and location.

A property's value is important in real estate transactions. Housing price trends are not only the concern of buyers and sellers, but it also indicates the current economic situation. Therefore, it is important to predict housing prices without bias to help both the buyers and sellers make their decisions. This project development may help to predict the house price.

1	Id	To count the records.
2	MSSubClass	Identifies the type of dwelling involved in the sale.
3	MSZoning	Identifies the general zoning classification of the sale.
4	LotArea	Lot size in square feet.
5	LotConfig	Configuration of the lot
6	BldgType	Type of dwelling
7	OverallCond	Rates the overall condition of the house
8	YearBuilt	Original construction year
9	YearRemodAdd	Remodel date (same as construction date if no remodeling or additions).
10	Exterior1st	Exterior covering on house
11	BsmtFinSF2	Type 2 finished square feet.
12	TotalBsmtSF	Total square feet of basement area
13	SalePrice	To be predicted

Code:

Importing Libraries and Dataset

Here we are using

Pandas- To load the Dataframe

Matplotlib- To visualize the data features

Seaborn- To see the correlation between features using heatmap

```
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))
```

Output:

	MSSubClass	MSZoning	LotArea	LotConfig	BldgType	OverallCond	YearBuilt
0	60	RL	8450	Inside	1Fam	5	2003
1	20	RL	9600	FR2	1Fam	8	1976
2	60	RL	11250	Inside	1Fam	5	2001
3	70	RL	9550	Corner	1Fam	5	1915
4	60	RL	14260	FR2	1Fam	5	2000

	YearRemodAdd	Exterior1st	BsmtFinSF2	TotalBsmtSF	SalePrice
0	2003	VinylSd	0.0	856.0	208500.0
1	1976	MetalSd	0.0	1262.0	181500.0
2	2002	VinylSd	0.0	920.0	223500.0
3	1970	Wd Sdng	0.0	756.0	140000.0
4	2000	VinylSd	0.0	1145.0	250000.0

As we have imported the data. So shape method will show us the dimension of the dataset.

Code:

dataset.shape

Output:

(2919,13)

DATA PREPROCESSING:

Code:

obj = (dataset.dtypes == 'object')

object_cols = list(obj[obj].index)

```
print("Categorical variables:",len(object_cols))

int_ = (dataset.dtypes == 'int')
num_cols = list(int_[int_].index)
print("Integer variables:",len(num_cols))

fl = (dataset.dtypes == 'float')
fl_cols = list(fl[fl].index)
print("Float variables:",len(fl_cols))
```

Output:

Categorical Variables : 4

Integer Variables : 6

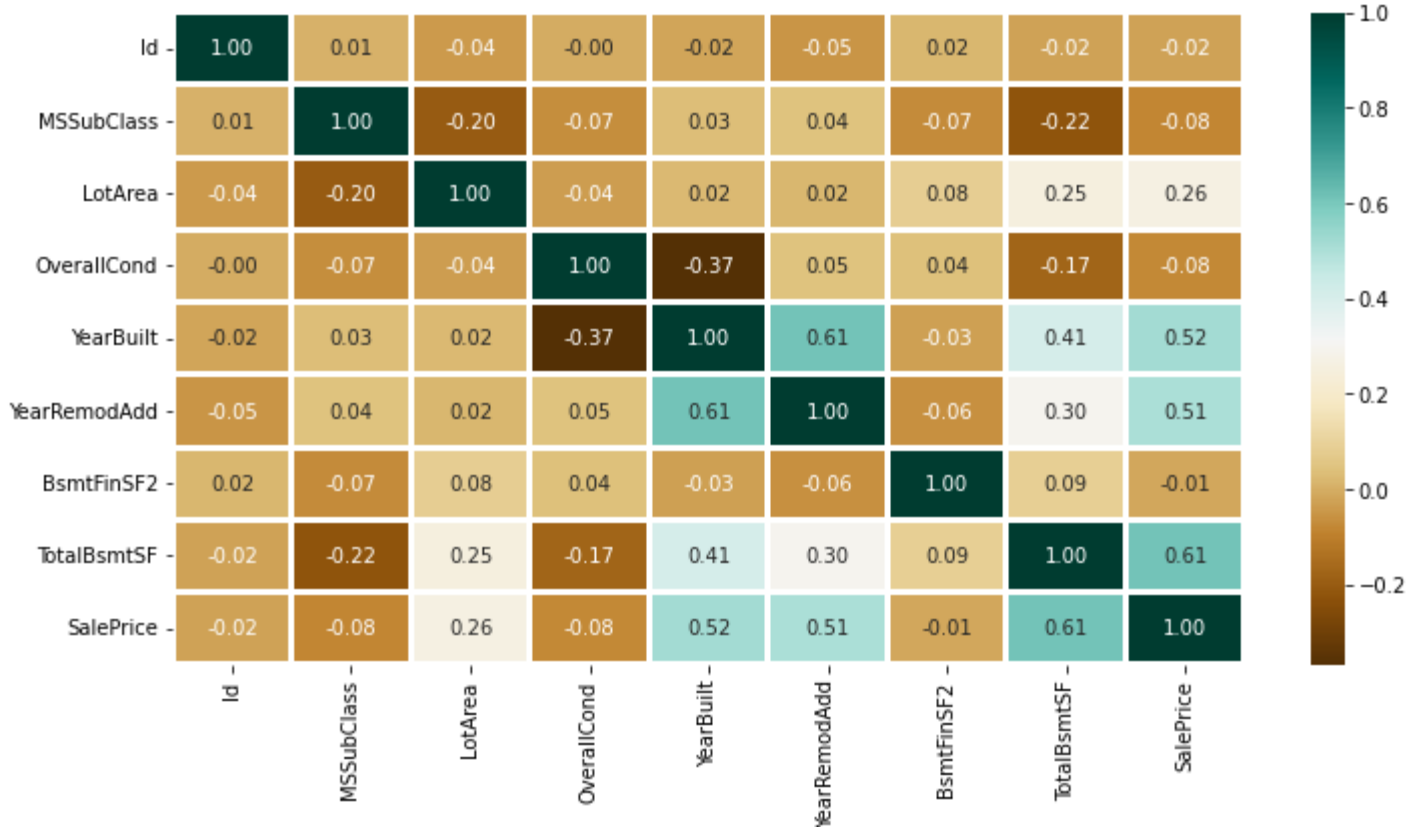
Float Variables : 3

EXPLORATORY DATA ANALYSIS

Code:

```
plt.figure(figsize=(12, 6))
sns.heatmap(dataset.corr(),
                    cmap = 'BrBG',
                    fmt = '.2f',
                   linewidths = 2,
                   annot = True)
```

Output:

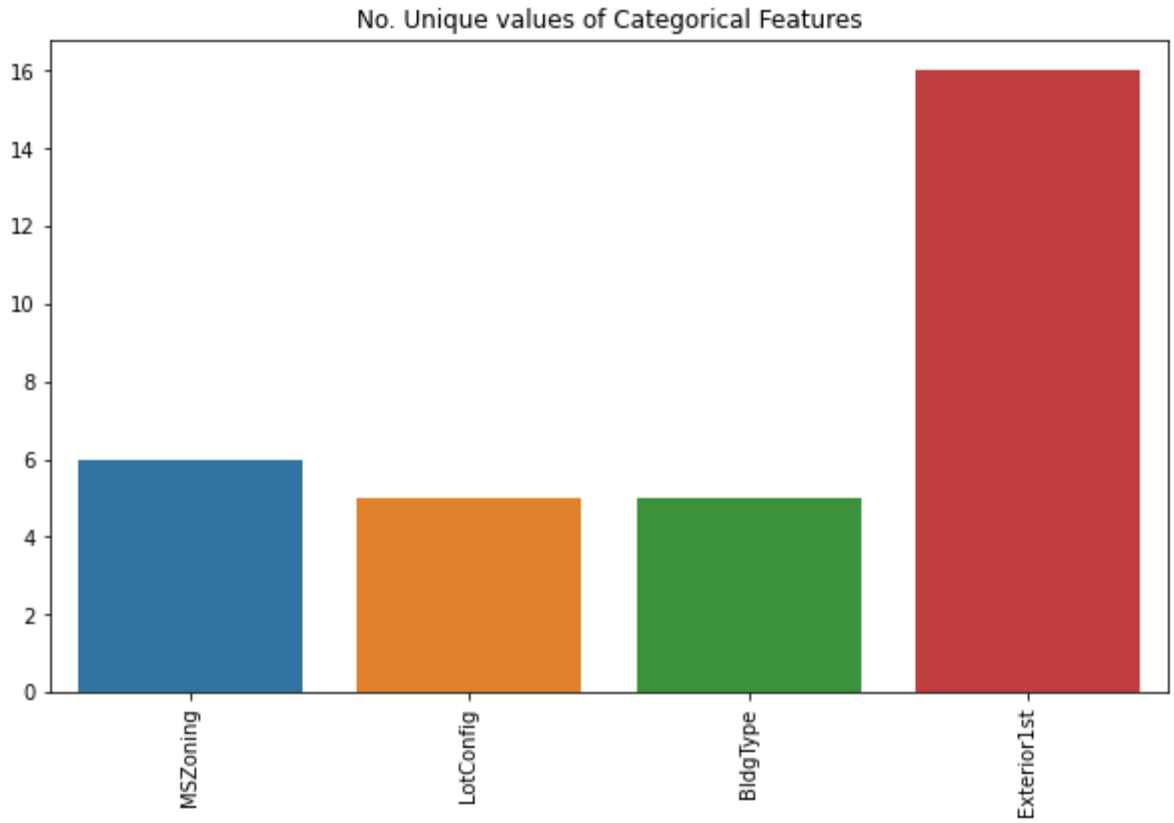

Id
MSSubClass
LotArea
OverallCond
YearBuilt
YearRemodAdd
BsmtFinSF2
TotalBsmtSF
SalePrice
1.00 0.01 -0.04 -0.00 -0.02 -0.05 0.02 -0.02 -0.02
0.01 1.00 -0.20 -0.07 0.03 0.04 -0.07 -0.22 -0.08
-0.04 -0.20 1.00 -0.04 0.02 0.02 0.08 0.25 0.26
-0.00 -0.07 -0.04 1.00 -0.37 0.05 0.04 -0.17 -0.08
-0.02 0.03 0.02 -0.37 1.00 0.61 -0.03 0.41 0.52
-0.05 0.04 0.02 0.05 0.61 1.00 -0.06 0.30 0.51
0.02 -0.07 0.08 0.04 -0.03 -0.06 1.00 0.09 -0.01
-0.02 -0.22 0.25 -0.17 0.41 0.30 0.09 1.00 0.61
-0.02 -0.08 0.26 -0.08 0.52 0.51 -0.01 0.61 1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

To analyze the different categorical features ,let us draw barplots

Code:

```
unique_values = []
for col in object_cols:
unique_values.append(dataset[col].unique().size)
plt.figure(figsize=(10,6))
plt.title('No. Unique values of Categorical Features')
plt.xticks(rotation=90)
sns.barplot(x=object_cols,y=unique_values)
```

Output:

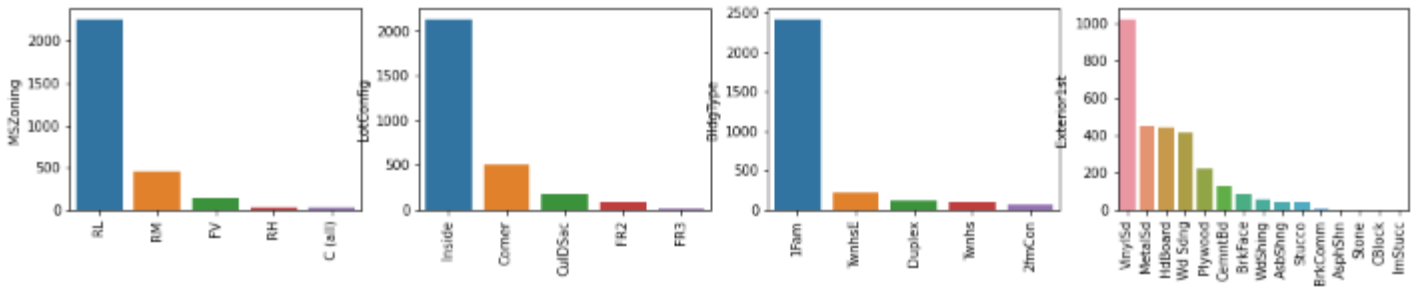


The plot shows that Exterior1st has around 16 unique categories and other features have around 6 unique categories. To findout the actual count of each category we can plot the bargraph of each four features separately.

```
plt.figure(figsize=(18, 36))

plt.title('Categorical Features: Distribution')

plt.xticks(rotation=90)

index = 1
```

```
for col in object_cols:
    y = dataset[col].value_counts()
    plt.subplot(11, 4, index)
    plt.xticks(rotation=90)
    sns.barplot(x=list(y.index), y=y)
    index += 1
```

Output:



Conclusion:

This House price prediction project help us to predict the price of the house and detecting the quality of the house. By including some features we have able to measure the price approximately not be the decimal categorization.