In [48]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
import warnings
warnings.filterwarnings('ignore')
```

In [49]:
```
h_price=pd.read_csv("test.csv")
```

In [50]:
```
h_price.head()
```

Out[50]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	ScreenPorch	PoolArea	PoolQC	Fence	MiscFeature	MiscVal	MoSold	YrSold
0	1461	20	RH	80.0	11622	Pave	NaN	Reg	Lvl	AllPub	...	120	0	NaN	MnPrv	NaN	0	6	2010
1	1462	20	RL	81.0	14267	Pave	NaN	IR1	Lvl	AllPub	...	0	0	NaN	NaN	Gar2	12500	6	2010
2	1463	60	RL	74.0	13830	Pave	NaN	IR1	Lvl	AllPub	...	0	0	NaN	MnPrv	NaN	0	3	2010
3	1464	60	RL	78.0	9978	Pave	NaN	IR1	Lvl	AllPub	...	0	0	NaN	NaN	NaN	0	6	2010
4	1465	120	RL	43.0	5005	Pave	NaN	IR1	HLS	AllPub	...	144	0	NaN	NaN	NaN	0	1	2010

5 rows × 80 columns

In [51]:
```
# DATA PRE-PROCESSSING
h_price.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1459 entries, 0 to 1458
Data columns (total 80 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Id             1459 non-null   int64  
 1   MSSubClass     1459 non-null   int64  
 2   MSZoning       1455 non-null   object 
 3   LotFrontage    1232 non-null   float64
 4   LotArea        1459 non-null   int64  
 5   Street         1459 non-null   object 
 6   Alley          107 non-null    object 
 7   LotShape       1459 non-null   object 
 8   LandContour    1459 non-null   object 
 9   Utilities      1457 non-null   object 
 10  LotConfig      1459 non-null   object 
 11  LandSlope      1459 non-null   object 
 12  Neighborhood   1459 non-null   object 
 13  Condition1     1459 non-null   object 
 14  Condition2     1459 non-null   object 
 15  BldgType       1459 non-null   object 
 16  HouseStyle     1459 non-null   object 
 17  OverallQual    1459 non-null   int64  
 18  OverallCond    1459 non-null   int64  
 19  YearBuilt      1459 non-null   int64  
 20  YearRemodAdd   1459 non-null   int64  
 21  RoofStyle      1459 non-null   object 
 22  RoofMatl       1459 non-null   object 
 23  Exterior1st    1458 non-null   object 
 24  Exterior2nd    1458 non-null   object 
 25  MasVnrType     1443 non-null   object 
 26  MasVnrArea     1444 non-null   float64
 27  ExterQual      1459 non-null   object 
 28  ExterCond      1459 non-null   object 
 29  Foundation     1459 non-null   object 
 30  BsmtQual       1415 non-null   object 
 31  BsmtCond       1414 non-null   object 
 32  BsmtExposure   1415 non-null   object 
 33  BsmtFinType1   1417 non-null   object 
 34  BsmtFinSF1     1458 non-null   float64
 35  BsmtFinType2   1417 non-null   object 
 36  BsmtFinSF2     1458 non-null   float64
 37  BsmtUnfSF      1458 non-null   float64
 38  TotalBsmtSF    1458 non-null   float64
 39  Heating        1459 non-null   object 
 40  HeatingQC      1459 non-null   object 
 41  CentralAir     1459 non-null   object 
 42  Electrical     1459 non-null   object 
 43  1stFlrSF       1459 non-null   int64  
 44  2ndFlrSF       1459 non-null   int64  
 45  LowQualFinSF   1459 non-null   int64  
 46  GrLivArea      1459 non-null   int64  
 47  BsmtFullBath   1457 non-null   float64
 48  BsmtHalfBath   1457 non-null   float64
 49  FullBath       1459 non-null   int64  
 50  HalfBath       1459 non-null   int64  
 51  BedroomAbvGr   1459 non-null   int64  
 52  KitchenAbvGr   1459 non-null   int64  
 53  KitchenQual    1458 non-null   object 
 54  TotRmsAbvGrd   1459 non-null   int64  
 55  Functional     1457 non-null   object 
 56  Fireplaces     1459 non-null   int64  
 57  FireplaceQu    729 non-null    object 
 58  GarageType     1383 non-null   object 
 59  GarageYrBlt    1381 non-null   float64
 60  GarageFinish   1381 non-null   object 
 61  GarageCars     1458 non-null   float64
 62  GarageArea     1458 non-null   float64
 63  GarageQual     1381 non-null   object 
 64  GarageCond     1381 non-null   object 
 65  PavedDrive     1459 non-null   object 
 66  WoodDeckSF     1459 non-null   int64  
 67  OpenPorchSF    1459 non-null   int64  
 68  EnclosedPorch  1459 non-null   int64  
 69  3SsnPorch      1459 non-null   int64  
 70  ScreenPorch    1459 non-null   int64  
 71  PoolArea       1459 non-null   int64  
 72  PoolQC         3 non-null      object 
 73  Fence          290 non-null    object 
 74  MiscFeature    51 non-null     object 
 75  MiscVal        1459 non-null   int64  
 76  MoSold         1459 non-null   int64  
 77  YrSold         1459 non-null   int64  
 78  SaleType       1458 non-null   object 
 79  SaleCondition  1459 non-null   object 
dtypes: float64(11), int64(26), object(43)
memory usage: 912.0+ KB
```

In [52]:
```
h_price.describe()
```

Out[52]:

	Id	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	...	GarageArea	WoodDeckSF	OpenPorch
count	1459.000000	1459.000000	1232.000000	1459.000000	1459.000000	1459.000000	1459.000000	1459.000000	1444.000000	1458.000000	...	1458.000000	1459.000000	1459.000000
mean	2190.000000	57.378341	68.580357	9819.161069	6.078821	5.553804	1971.357779	1983.662783	100.709141	439.203704	...	472.768861	93.174777	48.3138
std	421.321334	42.746880	22.376841	4955.517327	1.436812	1.113740	30.390071	21.130467	177.625900	455.268042	...	217.048611	127.744882	68.883
min	1461.000000	20.000000	21.000000	1470.000000	1.000000	1.000000	1879.000000	1950.000000	0.000000	0.000000	...	0.000000	0.000000	0.000
25%	1825.500000	20.000000	58.000000	7391.000000	5.000000	5.000000	1953.000000	1963.000000	0.000000	0.000000	...	318.000000	0.000000	0.000
50%	2190.000000	50.000000	67.000000	9399.000000	6.000000	5.000000	1973.000000	1992.000000	0.000000	350.500000	...	480.000000	0.000000	28.000
75%	2554.500000	70.000000	80.000000	11517.500000	7.000000	6.000000	2001.000000	2004.000000	164.000000	753.500000	...	576.000000	168.000000	72.000
max	2919.000000	190.000000	200.000000	56600.000000	10.000000	9.000000	2010.000000	2010.000000	1290.000000	4010.000000	...	1488.000000	1424.000000	742.000

8 rows × 37 columns

In [53]:
```
# How many null vallues in columns
h_price.shape
```

Out[53]:
```
(1459, 80)
```

In [54]:
```
h_price.isnull().sum()
```

Out[54]:
```
Id                0
MSSubClass        0
MSZoning          4
LotFrontage     227
LotArea           0
                 ..
MiscVal           0
MoSold            0
YrSold            0
SaleType          1
SaleCondition     0
Length: 80, dtype: int64
```

In [55]:
```
sns.heatmap(h_price.isnull(),yticklabels=False,cbar=False)
```

Out[55]:
```
<AxesSubplot:>
```

In [56]:
```
# fill allmissing values
h_price['LotFrontage']=h_price['LotFrontage'].filna(h_price['LotFrontage'].mean())
```

In [57]:
```
h_price['MSZoning']=h_price['MSZoning'].filna(h_price['MSZoning'].mode()[0])
h_price['BsmtQual']=h_price['BsmtQual'].filna(h_price['BsmtQual'].mode()[0])
```

In [58]:
```
h_price['FireplaceQu']=h_price['FireplaceQu'].filna(h_price['FireplaceQu'].mode()[0])
h_price['GarageType']=h_price['GarageType'].filna(h_price['GarageType'].mode()[0])
```

In [59]:
```
h_price['GarageFinish']=h_price['GarageFinish'].filna(h_price['GarageFinish'].mode()[0])
h_price['GarageQual']=h_price['GarageQual'].filna(h_price['GarageQual'].mode()[0])
h_price['GarageCond']=h_price['GarageCond'].filna(h_price['GarageCond'].mode()[0])
```

In [60]:
```
h_price['GarageFinish']=h_price['GarageFinish'].filna(h_price['GarageFinish'].mode()[0])
h_price['GarageQual']=h_price['GarageQual'].filna(h_price['GarageQual'].mode()[0])
h_price['GarageCond']=h_price['GarageCond'].filna(h_price['GarageCond'].mode()[0])
```

In [61]:
```
h_price.drop(['PoolQC','Fence','MiscFeature'],axis=1,inplace=True)
```

In [62]:
```
h_price.shape
```

Out[62]:
```
(1459, 77)
```

In [63]:
```
# remove all null values in data set
h_price.drop(['Id'],axis=1,inplace=True)
```

In [64]:
```
h_price.isnull().sum()
```

Out[64]:
```
MSSubClass      0
MSZoning        0
LotFrontage     0
LotArea         0
Street          0
               ..
MiscVal         0
MoSold          0
YrSold          0
SaleType        1
SaleCondition   0
Length: 76, dtype: int64
```

In [73]:
```
h_price['MasVnrType']=h_price['MasVnrType'].filna(h_price['MasVnrType'].mode()[0])
h_price['MasVnrArea']=h_price['MasVnrArea'].filna(h_price['MasVnrArea'].mode()[0])
```

In [69]:
```
sns.heatmap(h_price.isnull(),yticklabels=False,cbar=False,cmap='coolwarm')
```

Out[69]:
```
<AxesSubplot:>
```

In [67]:
```
h_price['BsmtExposure']=h_price['BsmtExposure'].filna(h_price['BsmtExposure'].mode()[0])
```

In [74]:
```
sns.heatmap(h_price.isnull(),yticklabels=False,cbar=False,cmap='YlGnBu')
```

Out[74]:
```
<AxesSubplot:>
```


In [75]:
```
h_price['BsmtFinType2']=h_price['BsmtFinType2'].filna(h_price['BsmtFinType2'].mode()[0])
```

In [76]:
```
h_price.dropna(inplace=True)
```

In [77]:
```
h_price.shape
```

Out[77]:
```
(93, 76)
```

In [78]:
```
h_price.to_csv('formulatest.csv',index=False)
```

In [ ]:

In [ ]: