Import Library

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('/content/train.csv')

df.head()

{"type":"dataframe","variable_name":"df"}

df.shape

(1460, 81)
```

Data **Preprocessing**

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

Id                 0
MSSubClass         0
MSZoning           0
LotFrontage      259
LotArea            0
                ...
MoSold             0
YrSold             0
SaleType           0
SaleCondition      0
SalePrice          0
Length: 81, dtype: int64

df['LotFrontage']=df['LotFrontage'].replace(np.nan,df['LotFrontage'].m
ean())

df.shape

(1460, 81)

df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 81 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Id             1460 non-null   int64
 1   MSSubClass     1460 non-null   int64
 2   MSZoning       1460 non-null   object
```

```
3   LotFrontage    1460 non-null   float64
4   LotArea        1460 non-null   int64
5   Street         1460 non-null   object
6   Alley          91 non-null     object
7   LotShape       1460 non-null   object
8   LandContour    1460 non-null   object
9   Utilities      1460 non-null   object
10  LotConfig      1460 non-null   object
11  LandSlope      1460 non-null   object
12  Neighborhood   1460 non-null   object
13  Condition1     1460 non-null   object
14  Condition2     1460 non-null   object
15  BldgType       1460 non-null   object
16  HouseStyle     1460 non-null   object
17  OverallQual    1460 non-null   int64
18  OverallCond    1460 non-null   int64
19  YearBuilt      1460 non-null   int64
20  YearRemodAdd   1460 non-null   int64
21  RoofStyle      1460 non-null   object
22  RoofMatl       1460 non-null   object
23  Exterior1st    1460 non-null   object
24  Exterior2nd    1460 non-null   object
25  MasVnrType     588 non-null    object
26  MasVnrArea     1452 non-null   float64
27  ExterQual      1460 non-null   object
28  ExterCond      1460 non-null   object
29  Foundation     1460 non-null   object
30  BsmtQual       1423 non-null   object
31  BsmtCond       1423 non-null   object
32  BsmtExposure   1422 non-null   object
33  BsmtFinType1   1423 non-null   object
34  BsmtFinSF1     1460 non-null   int64
35  BsmtFinType2   1422 non-null   object
36  BsmtFinSF2     1460 non-null   int64
37  BsmtUnfSF      1460 non-null   int64
38  TotalBsmtSF    1460 non-null   int64
39  Heating        1460 non-null   object
40  HeatingQC      1460 non-null   object
41  CentralAir     1460 non-null   object
42  Electrical     1459 non-null   object
43  1stFlrSF       1460 non-null   int64
44  2ndFlrSF       1460 non-null   int64
45  LowQualFinSF   1460 non-null   int64
46  GrLivArea      1460 non-null   int64
47  BsmtFullBath   1460 non-null   int64
48  BsmtHalfBath   1460 non-null   int64
49  FullBath       1460 non-null   int64
50  HalfBath       1460 non-null   int64
51  BedroomAbvGr   1460 non-null   int64
```

```
 52  KitchenAbvGr   1460 non-null   int64
 53  KitchenQual    1460 non-null   object
 54  TotRmsAbvGrd   1460 non-null   int64
 55  Functional     1460 non-null   object
 56  Fireplaces     1460 non-null   int64
 57  FireplaceQu    770 non-null    object
 58  GarageType     1379 non-null   object
 59  GarageYrBlt    1379 non-null   float64
 60  GarageFinish   1379 non-null   object
 61  GarageCars     1460 non-null   int64
 62  GarageArea     1460 non-null   int64
 63  GarageQual     1379 non-null   object
 64  GarageCond     1379 non-null   object
 65  PavedDrive     1460 non-null   object
 66  WoodDeckSF     1460 non-null   int64
 67  OpenPorchSF    1460 non-null   int64
 68  EnclosedPorch  1460 non-null   int64
 69  3SsnPorch      1460 non-null   int64
 70  ScreenPorch    1460 non-null   int64
 71  PoolArea       1460 non-null   int64
 72  PoolQC         7 non-null      object
 73  Fence          281 non-null    object
 74  MiscFeature    54 non-null     object
 75  MiscVal        1460 non-null   int64
 76  MoSold         1460 non-null   int64
 77  YrSold         1460 non-null   int64
 78  SaleType       1460 non-null   object
 79  SaleCondition  1460 non-null   object
 80  SalePrice      1460 non-null   int64
dtypes: float64(3), int64(35), object(43)
memory usage: 924.0+ KB
```

```
df['MasVnrType'].value_counts()
```

```
MasVnrType
BrkFace    445
Stone      128
BrkCmn      15
Name: count, dtype: int64
```

```
df['MasVnrType'] = df['MasVnrType'].replace(np.nan ,'BrkFace' )
```

```
df['MasVnrType'].value_counts()
```

```
MasVnrType
BrkFace    1317
Stone       128
BrkCmn       15
Name: count, dtype: int64
```

```
df = df.drop(columns=['Alley' , 'MiscFeature' , 'PoolQC' ,
'Fence'],axis=1)

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

```
Id               0
MSSubClass       0
MSZoning         0
LotFrontage      0
LotArea          0
                ..
MoSold           0
YrSold           0
SaleType         0
SaleCondition    0
SalePrice        0
Length: 77, dtype: int64
```

```
df.shape
```

```
(1460, 77)
```

```
df['FireplaceQu'].value_counts()
```

```
FireplaceQu
Gd    380
TA    313
Fa     33
Ex     24
Po     20
Name: count, dtype: int64
```

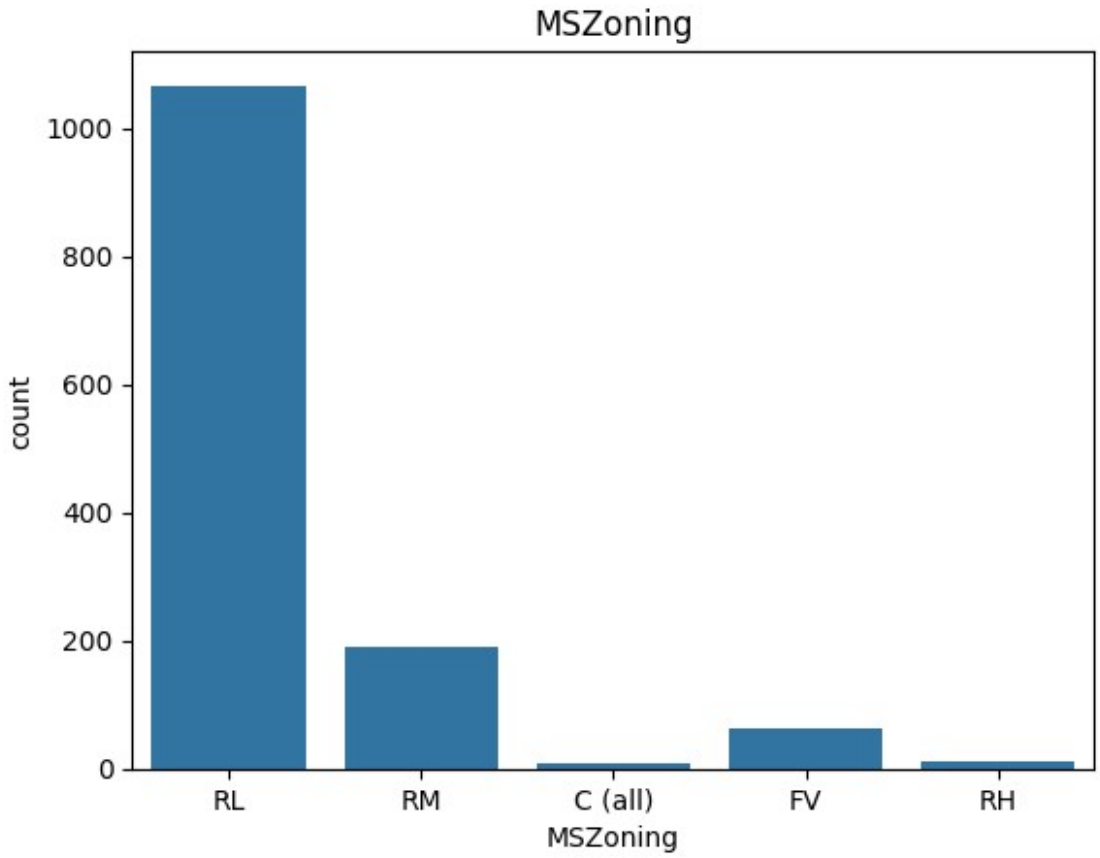
```
df['FireplaceQu'] = df['FireplaceQu'].replace(np.nan,'Gd')

df['FireplaceQu'].value_counts()
```

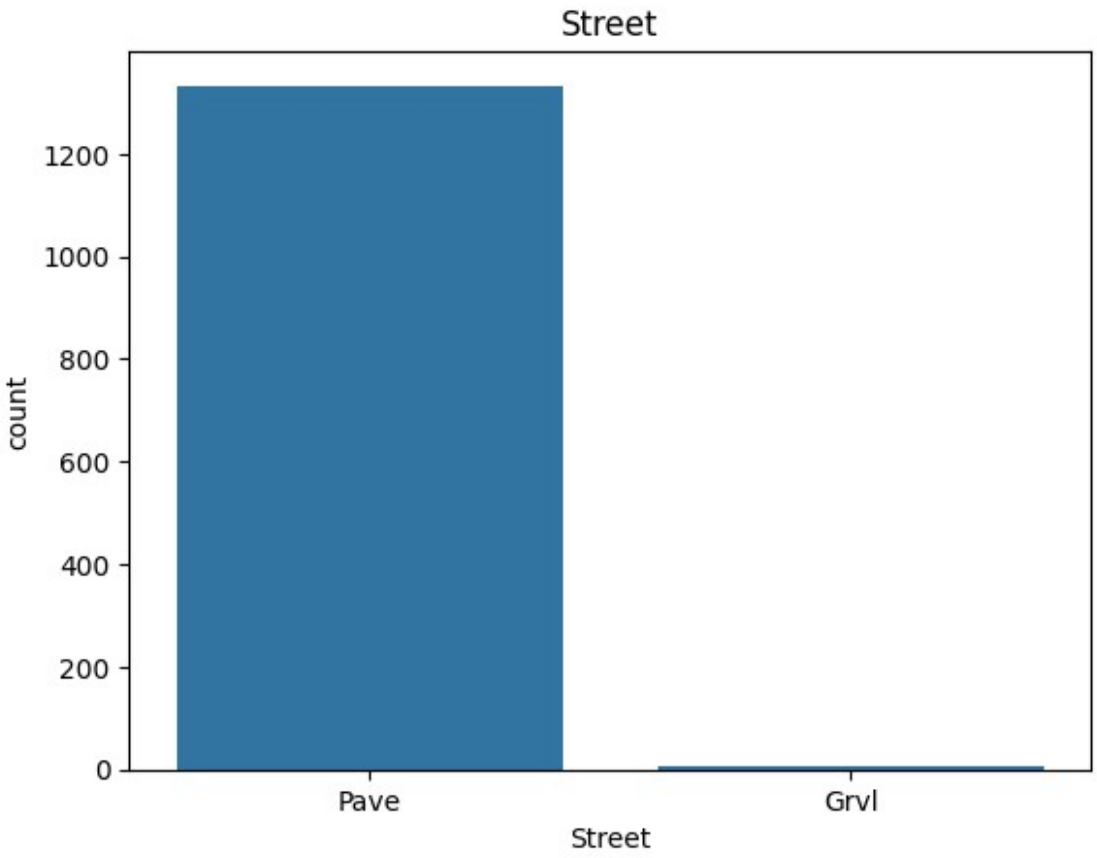
```
FireplaceQu
Gd    1070
TA     313
Fa      33
Ex      24
Po      20
Name: count, dtype: int64
```

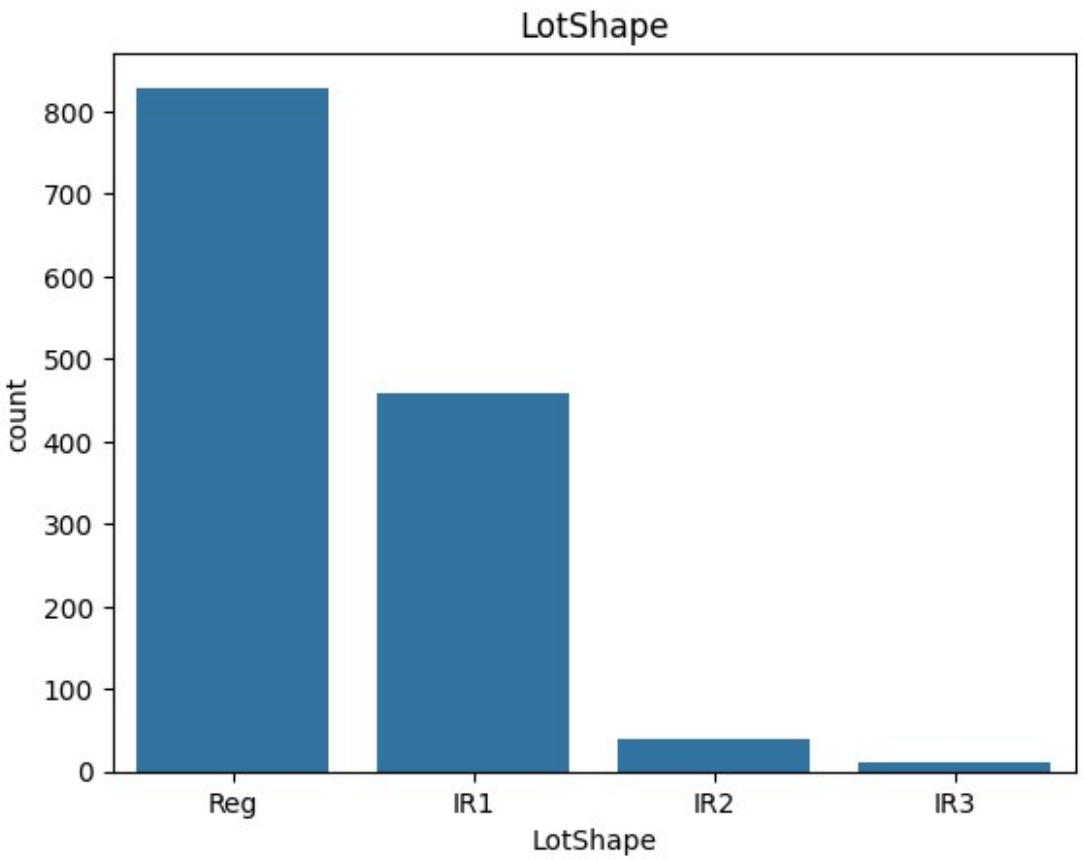
```
df=df.dropna()

df.isnull().values.sum()
df.shape
```

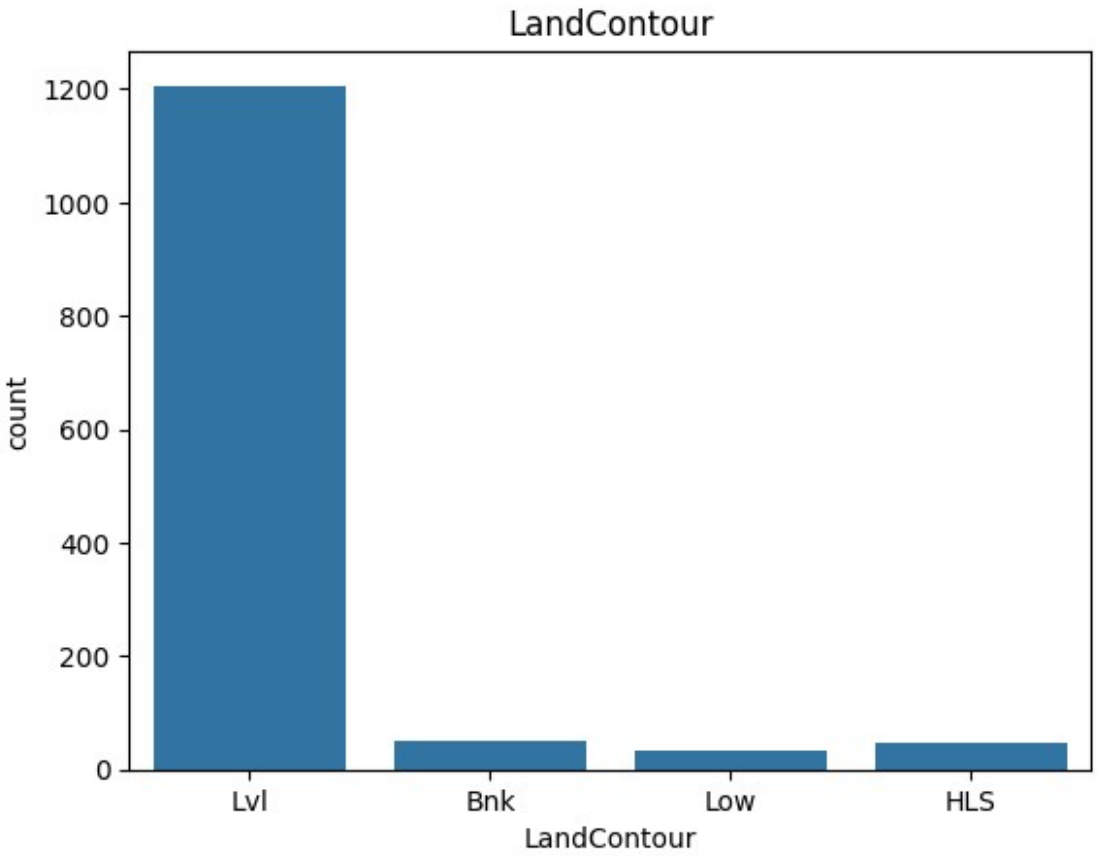
```
(1338, 77)
```

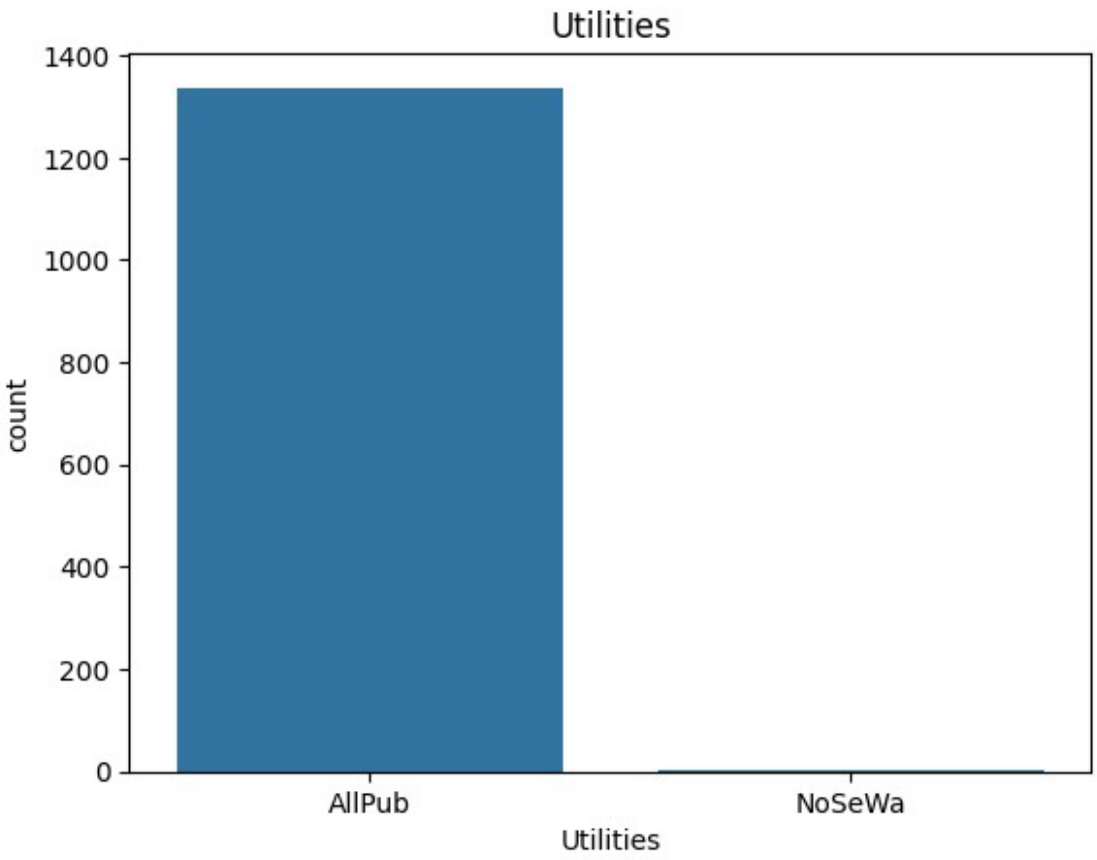
```
cat_data = df.select_dtypes(include='object')
p = cat_data.columns

def fn(p):
  for i in p:
    sns.countplot(x=i,data=df)
    plt.title(i)
    plt.show()
fn(p)
```

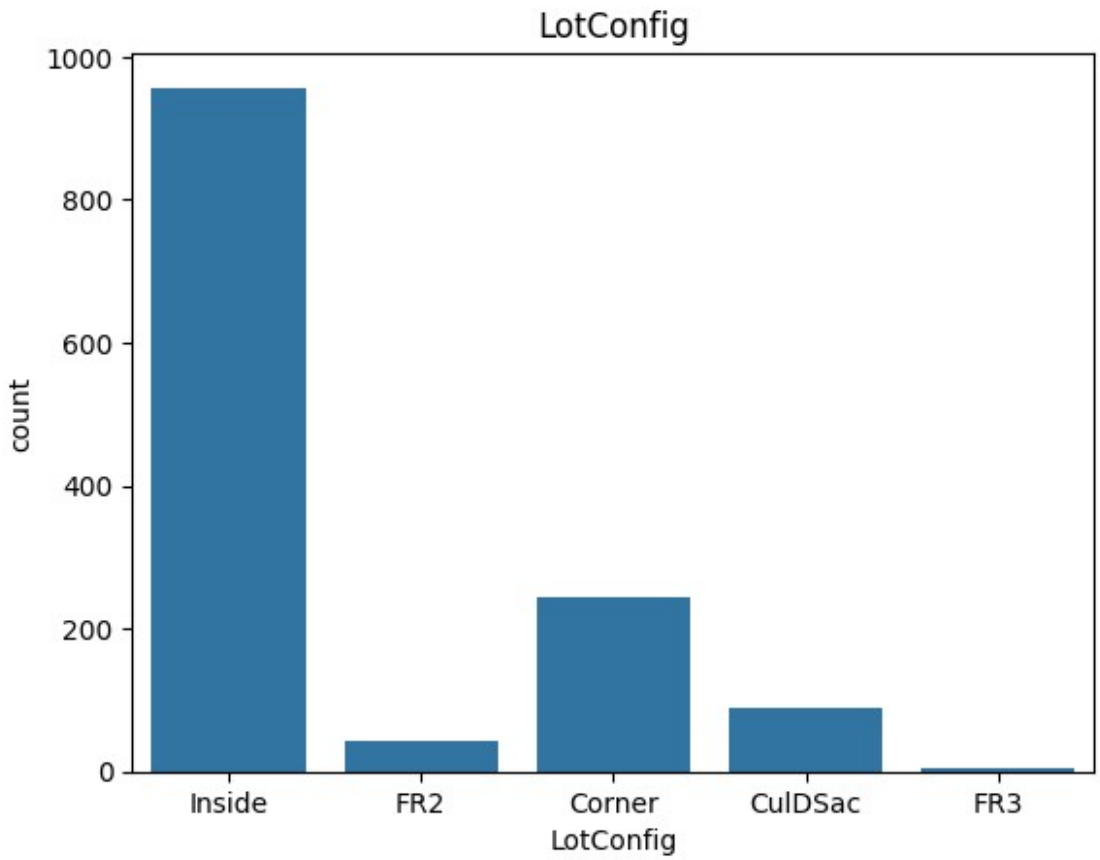


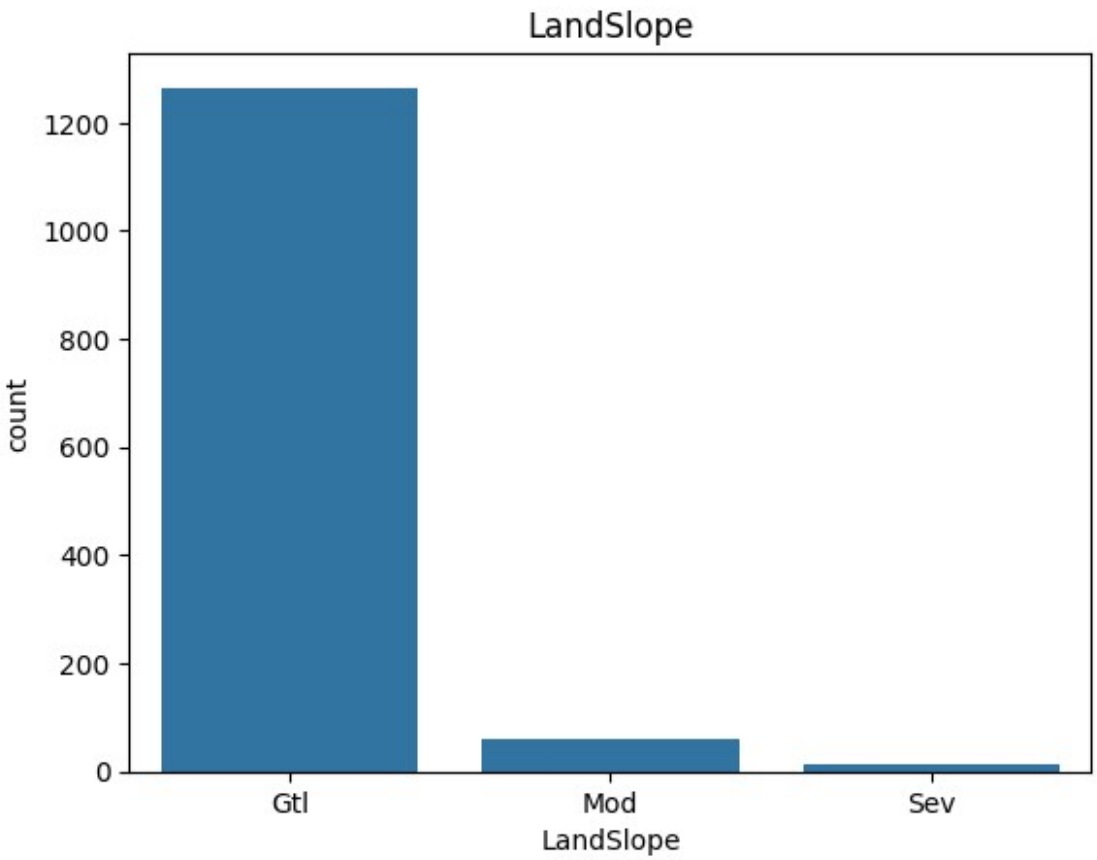

Street
1200
1000
800
600
400
200
0
count
Pave
Grvl
Street

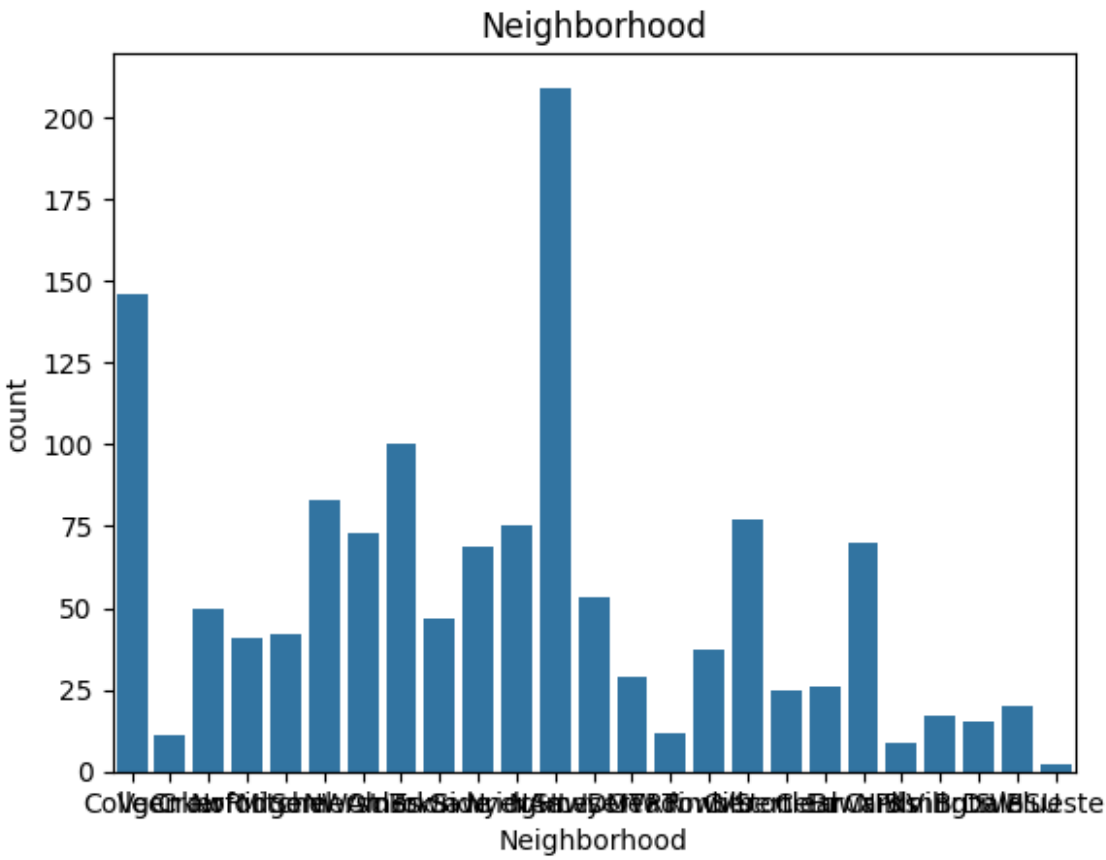

LotShape
800
700
600
500
400
300
200
100
0
count
Reg
IR1
IR2
IR3
LotShape

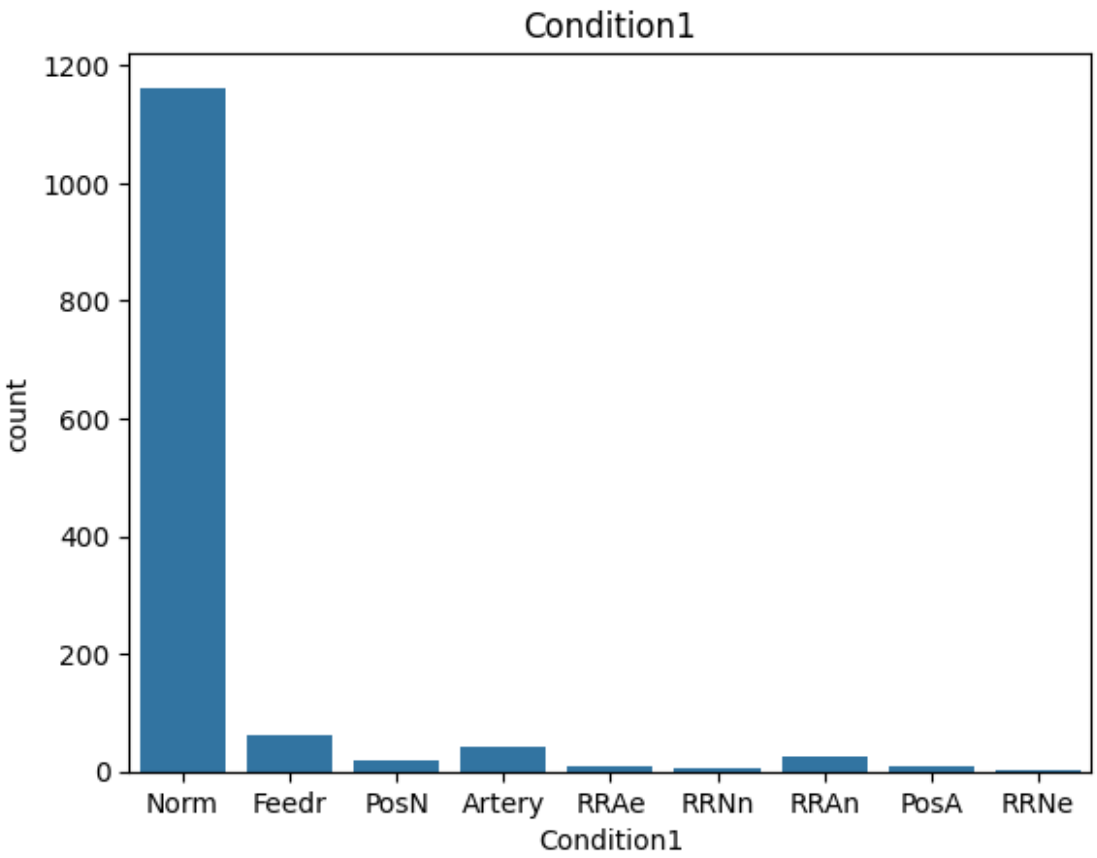

LandContour
1200
1000
800
600
400
200
0
count
Lvl
Bnk
Low
HLS
LandContour

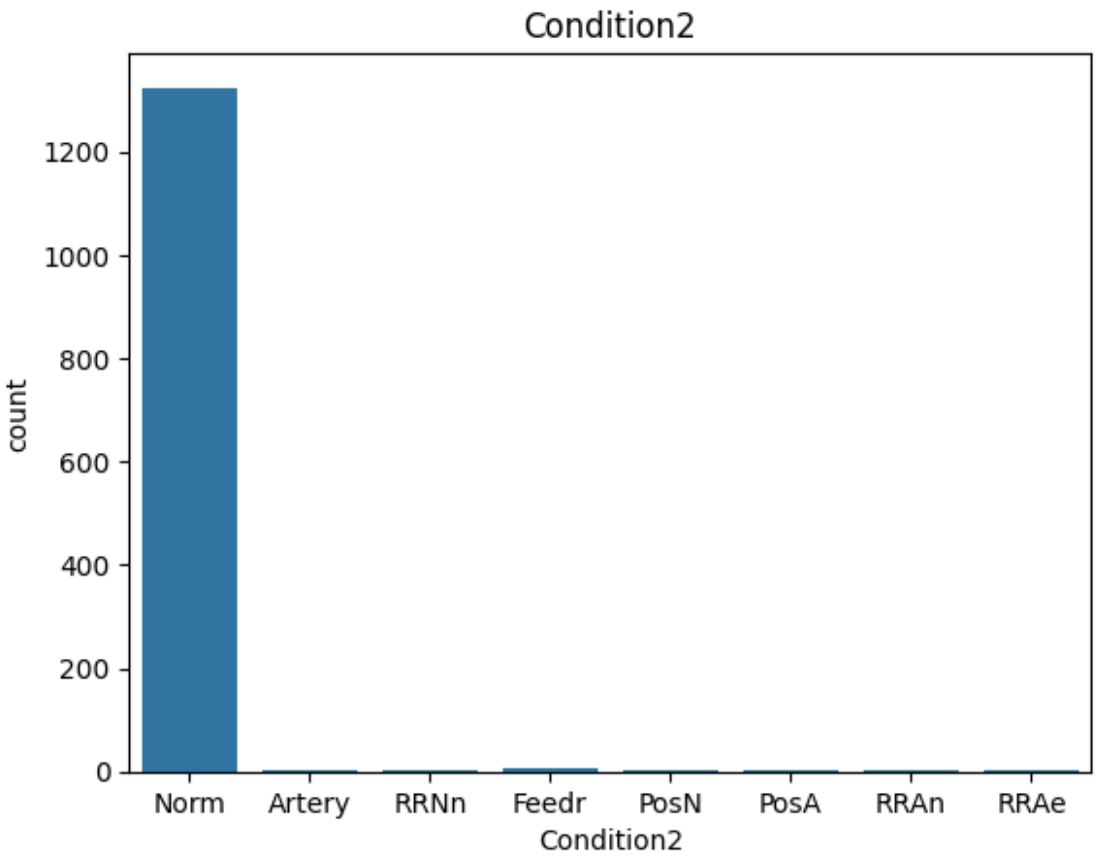

Utilities
1400
1200
1000
800
600
400
200
0
count
AllPub
NoSeWa
Utilities

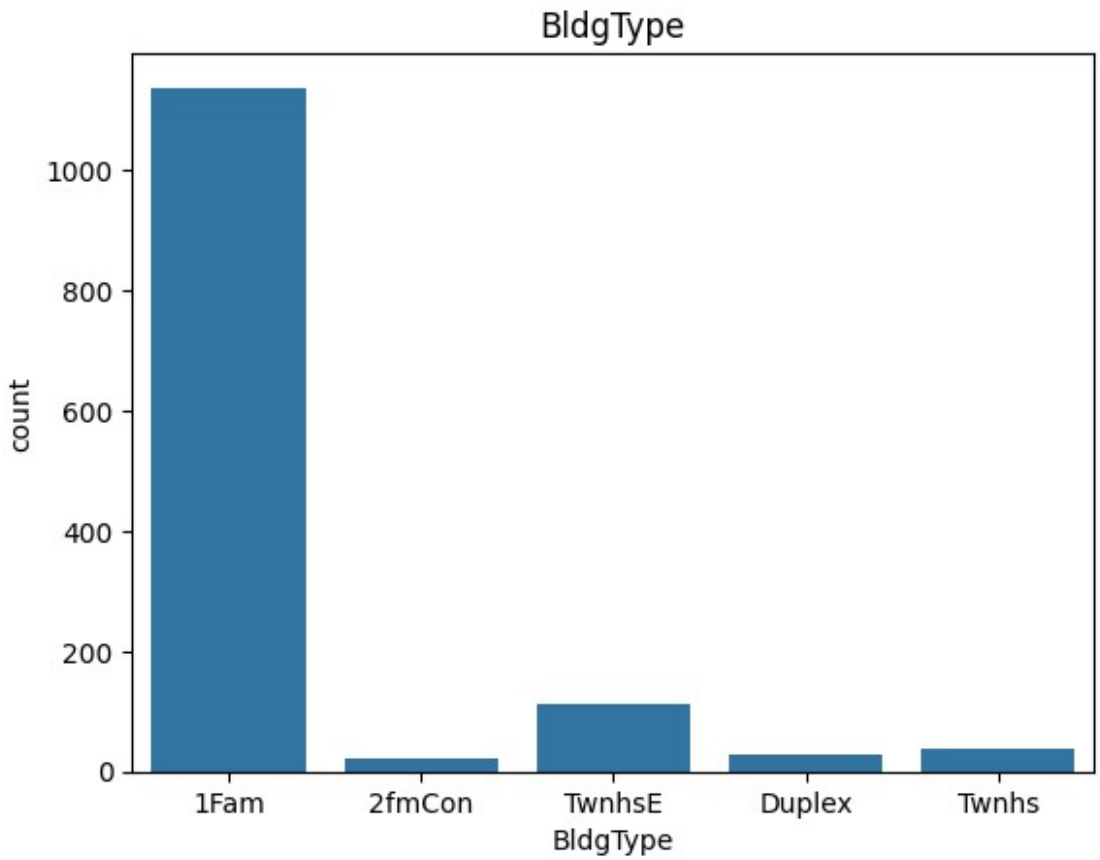

LotConfig
1000
800
600
400
200
0
count
Inside
FR2
Corner
CulDSac
FR3
LotConfig

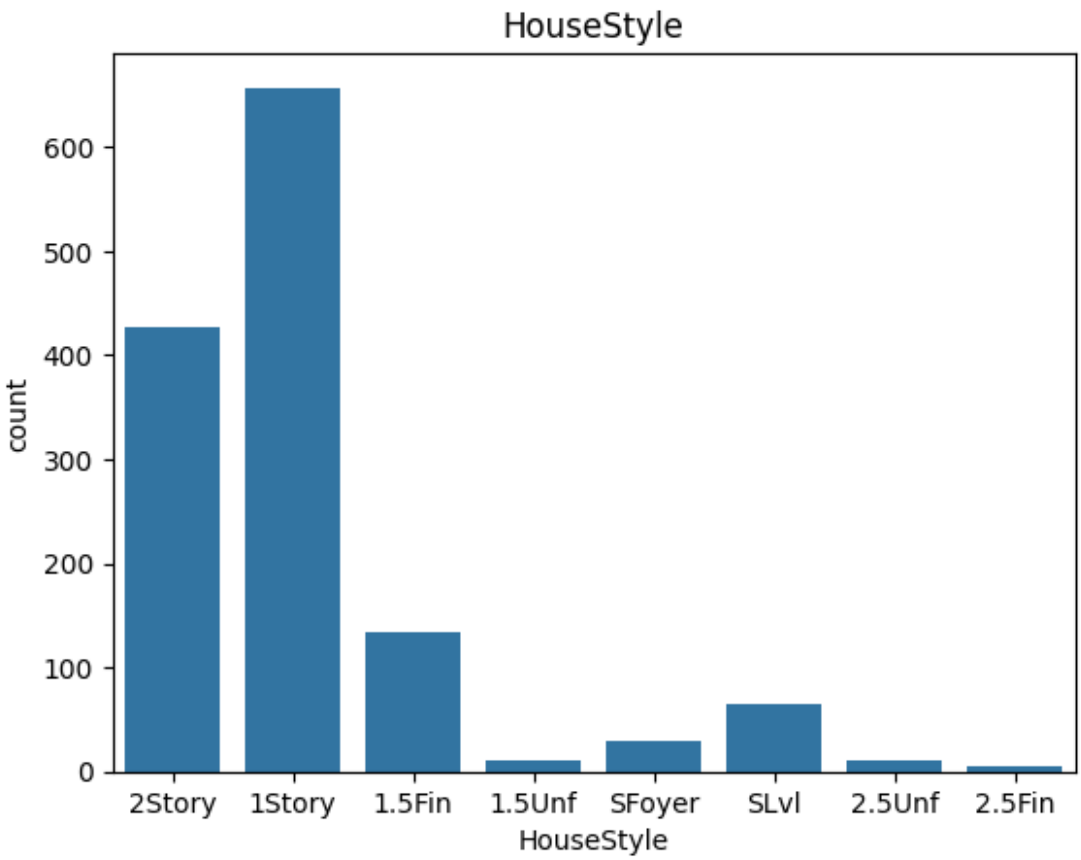

LandSlope
1200
1000
800
600
400
200
0
count
Gtl
Mod
Sev
LandSlope

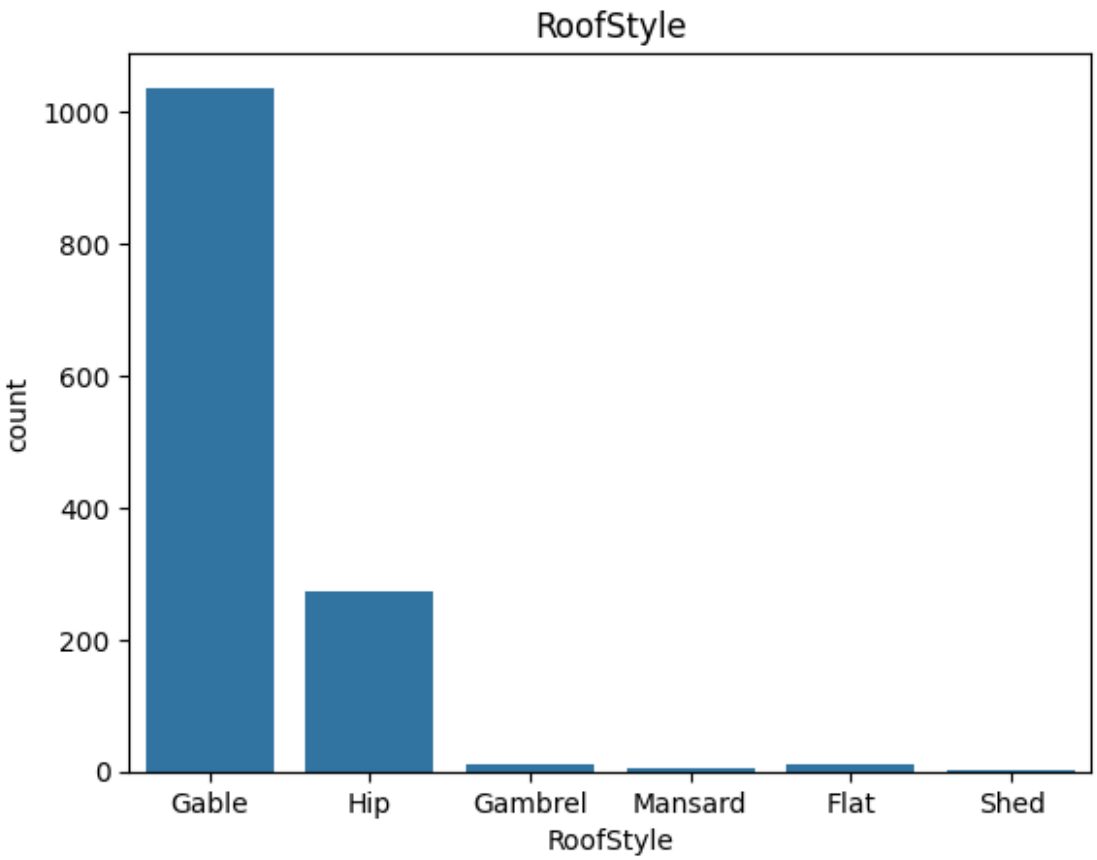

Neighborhood
count
200
175
150
125
100
75
50
25
0
Neighborhood

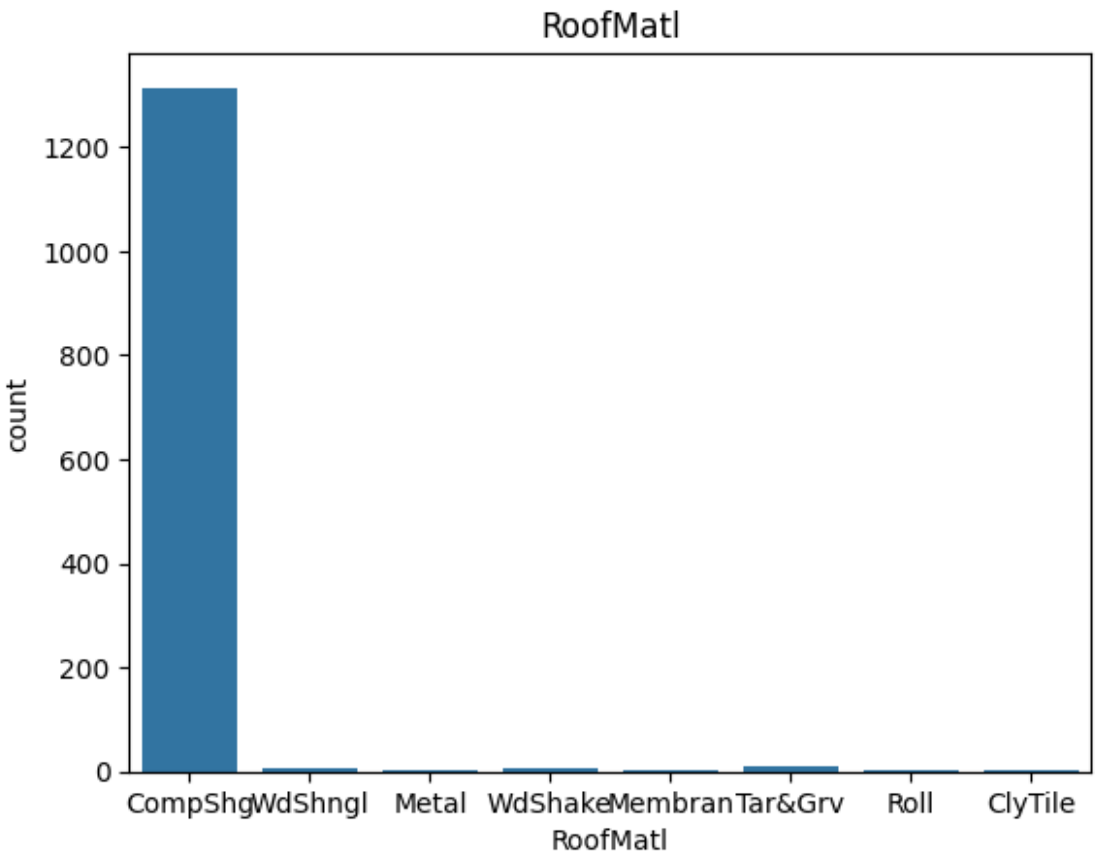

Condition1
1200
1000
800
600
400
200
0
count
Norm
Feedr
PosN
Artery
RRAe
RRNn
RRAn
PosA
RRNe
Condition1

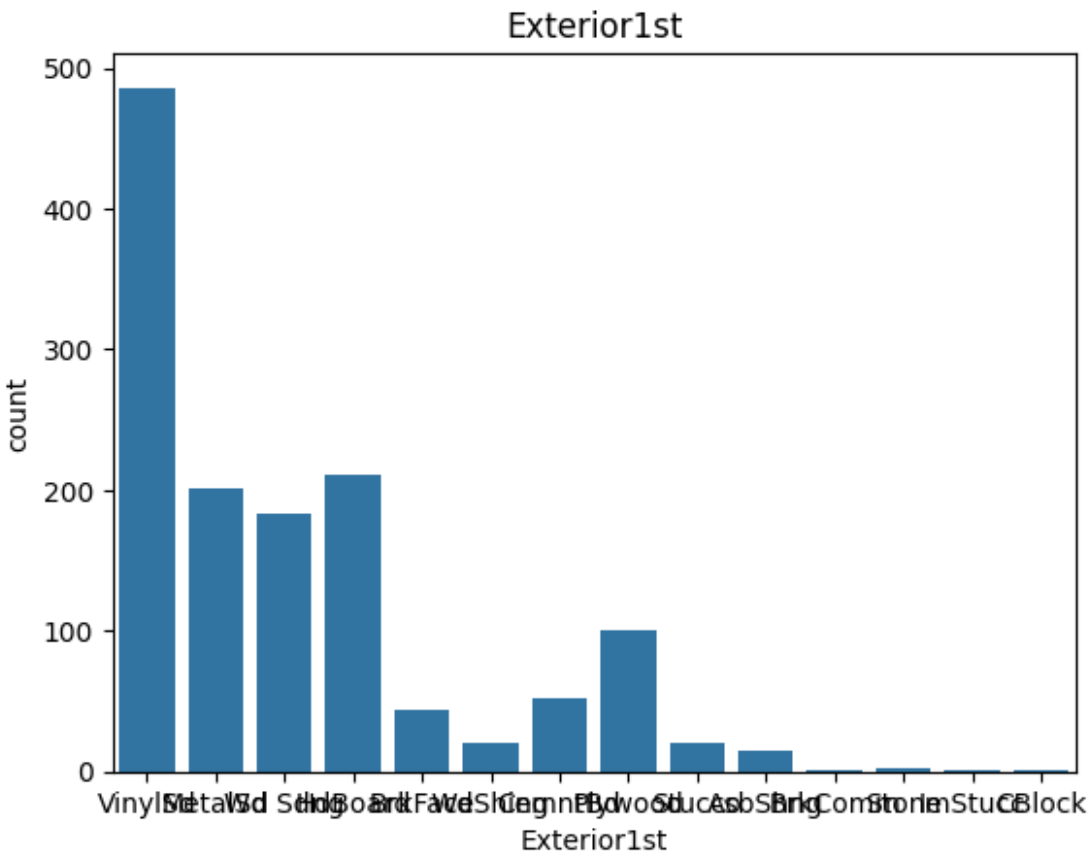

Condition2
count
1200
1000
800
600
400
200
0
Norm
Artery
RRNn
Feedr
PosN
PosA
RRAn
RRAe
Condition2

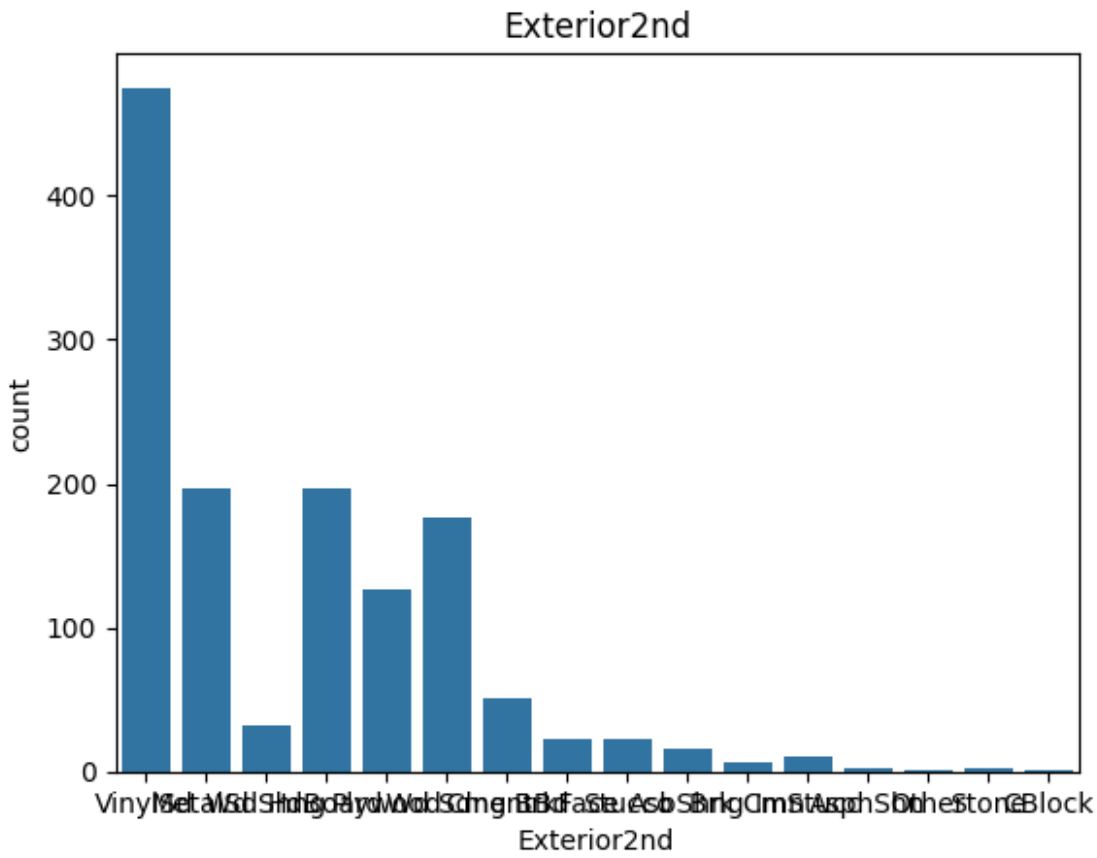

BldgType
count
1000
800
600
400
200
0
1Fam
2fmCon
TwnhsE
Duplex
Twnhs
BldgType

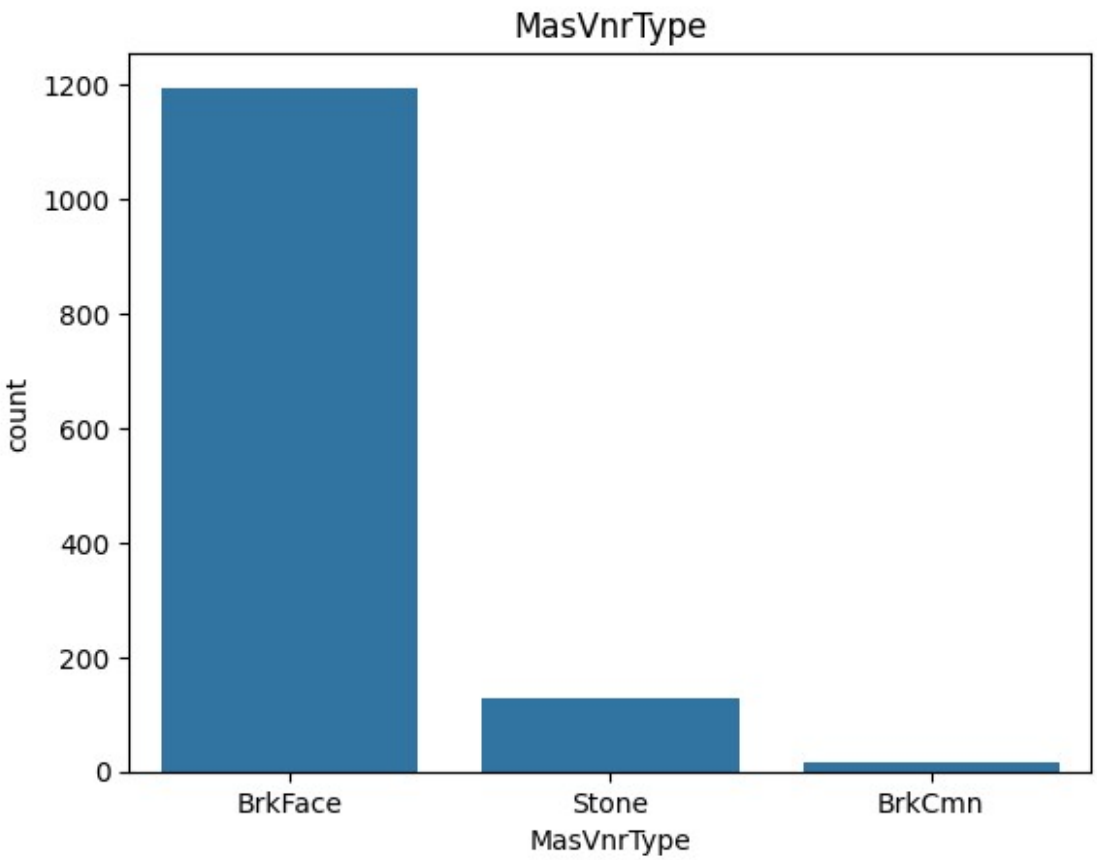

HouseStyle
600
500
400
300
200
100
0
count
2Story
1Story
1.5Fin
1.5Unf
SFoyer
SLvl
2.5Unf
2.5Fin
HouseStyle

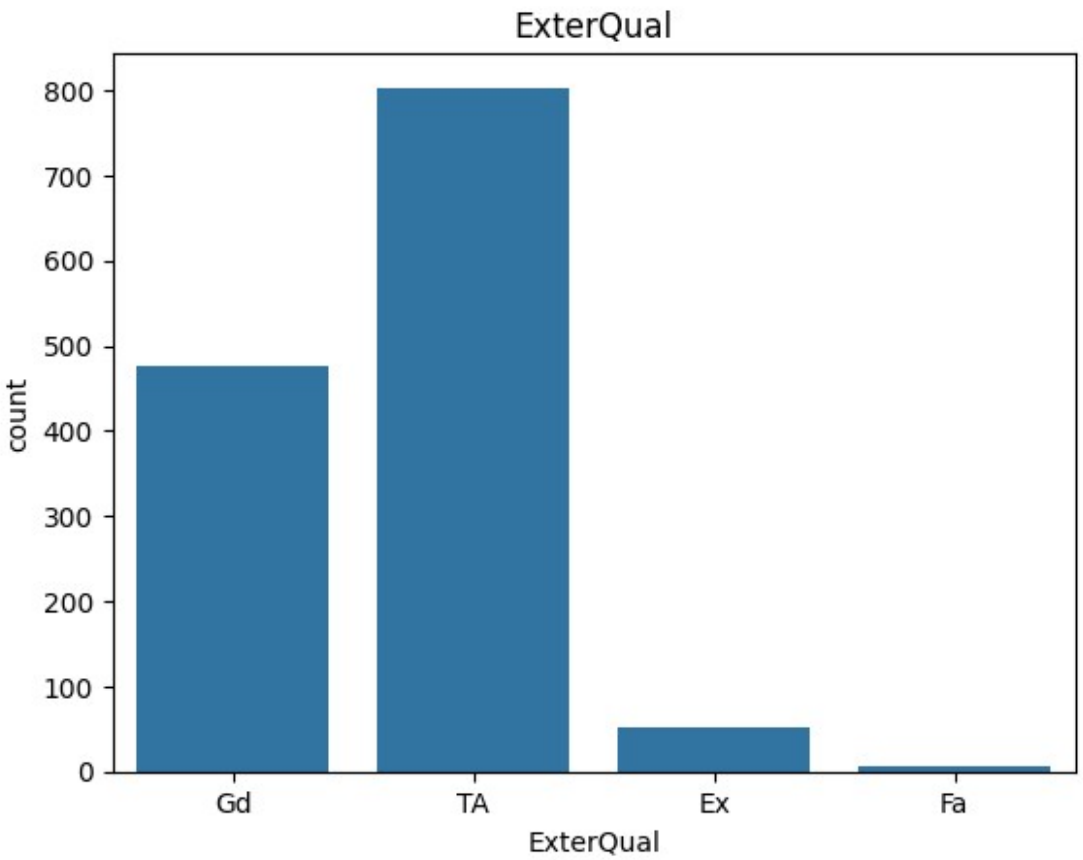

RoofStyle
1000
800
600
400
200
0
count
Gable
Hip
Gambrel
Mansard
Flat
Shed
RoofStyle

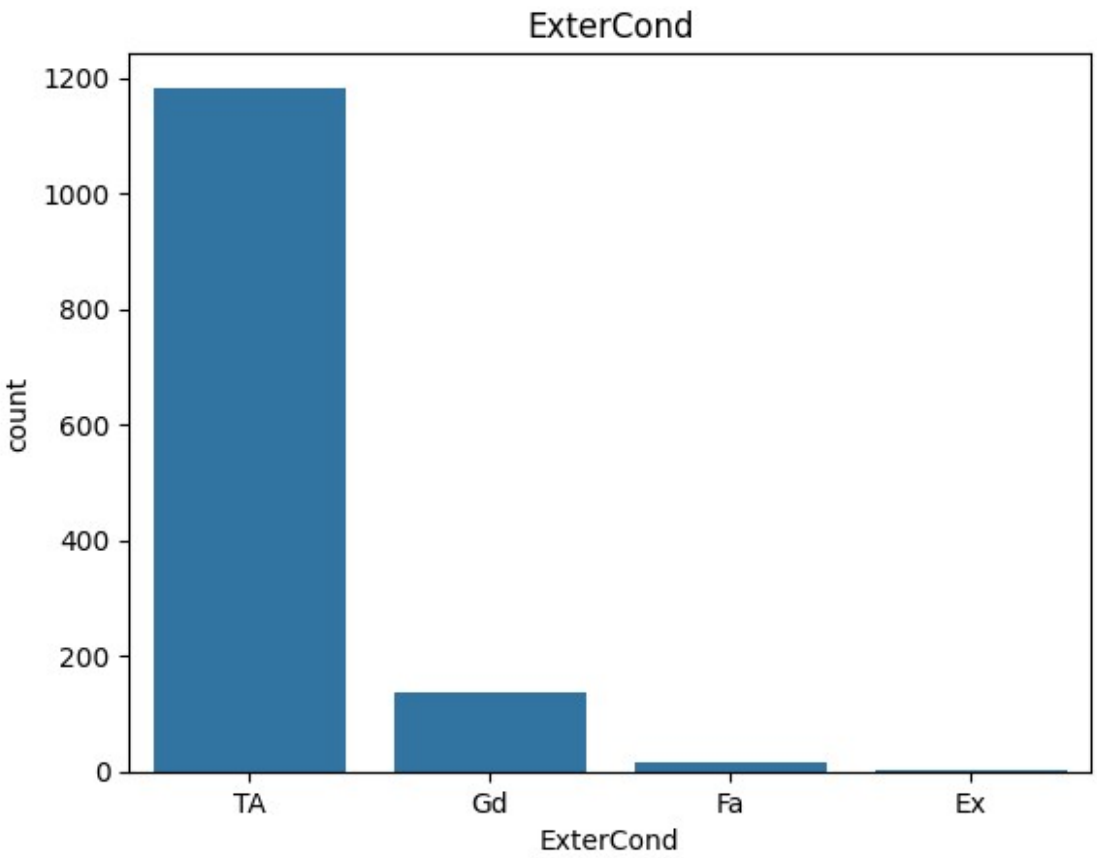

RoofMatl
count
1200
1000
800
600
400
200
0
CompShg
WdShngl
Metal
WdShake
Membran
Tar&Grv
Roll
ClyTile
RoofMatl

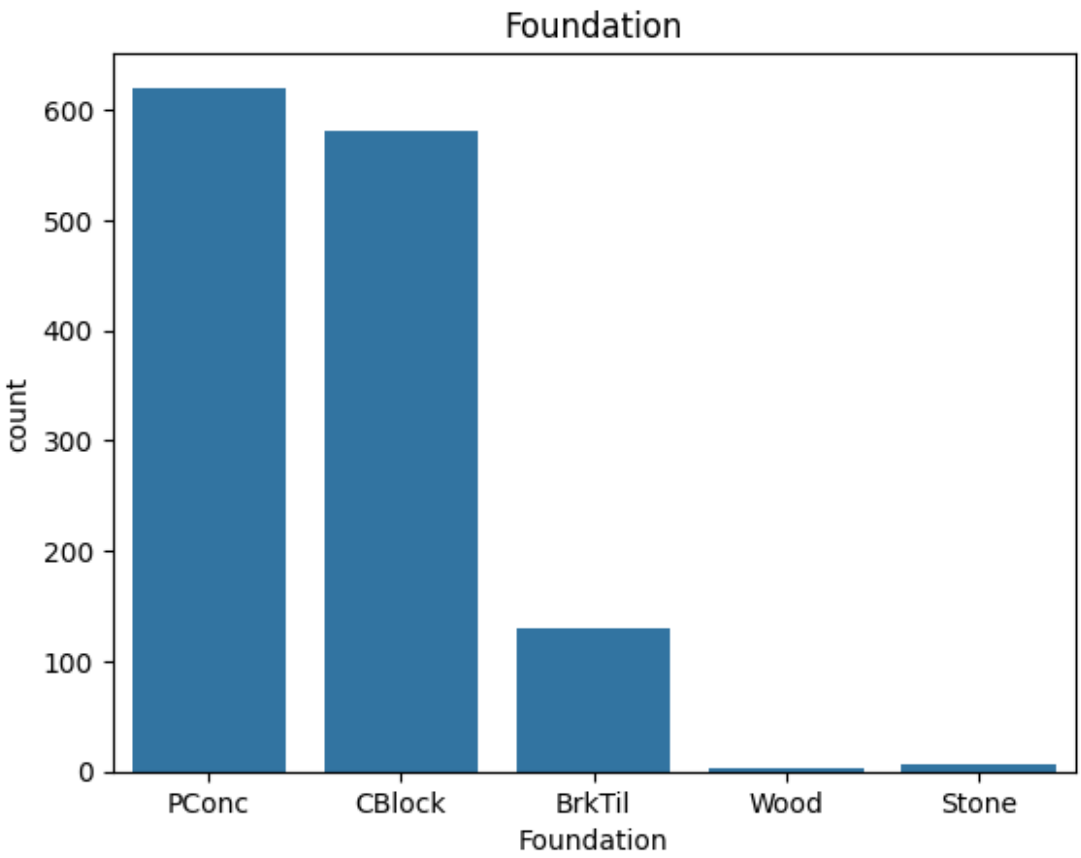

Exterior1st
500
400
300
200
100
0
count
Exterior1st

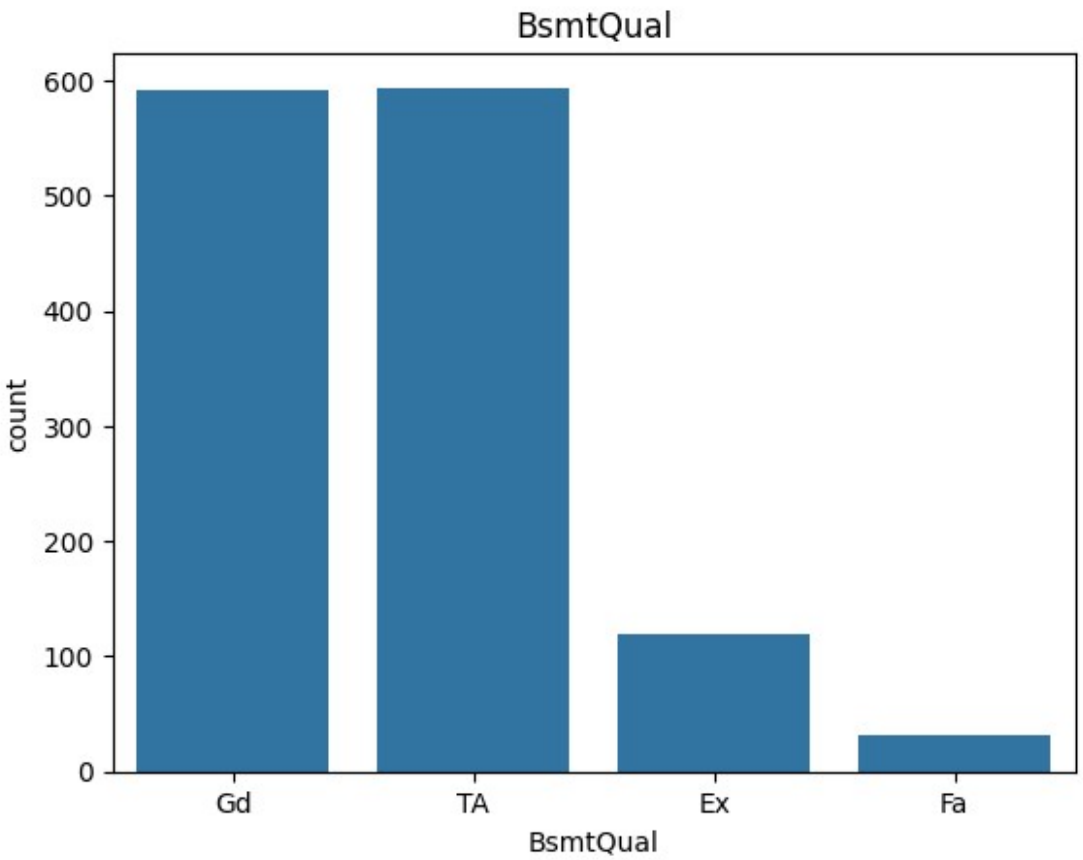

Exterior2nd
count
400
300
200
100
0
Exterior2nd

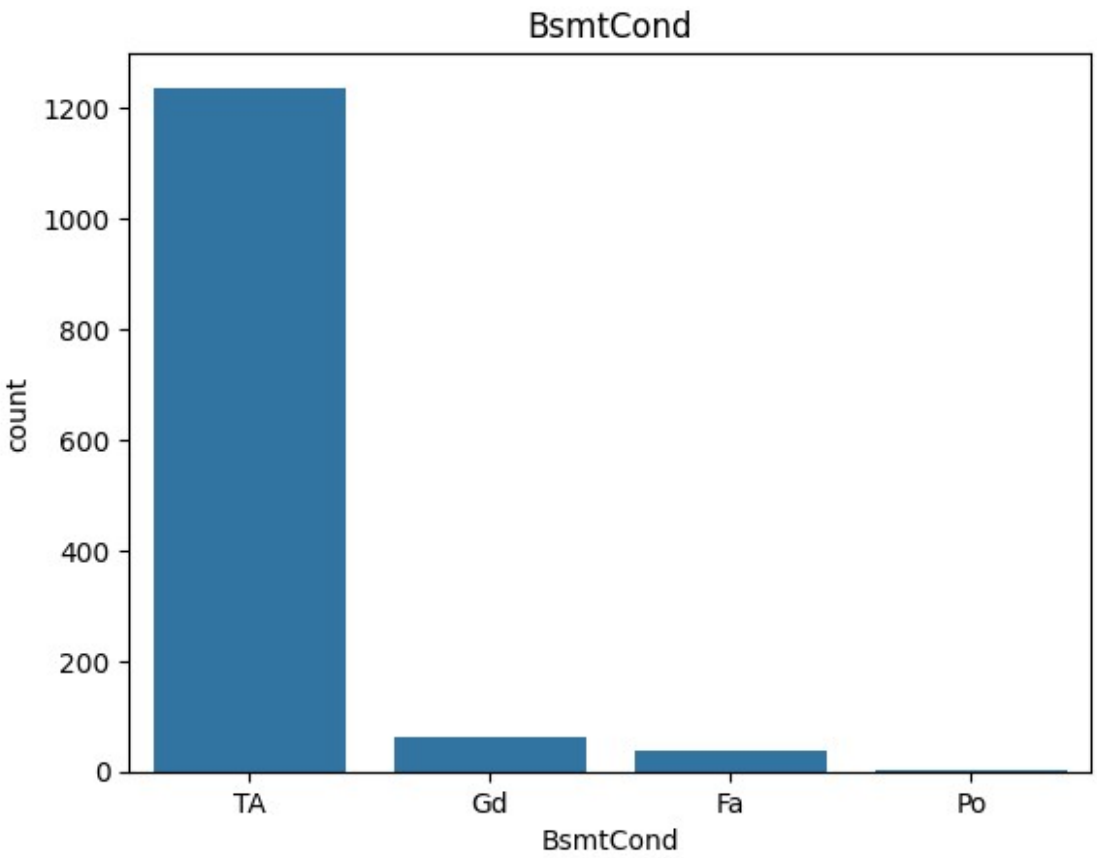

MasVnrType
1200
1000
800
600
400
200
0
count
BrkFace
Stone
BrkCmn
MasVnrType

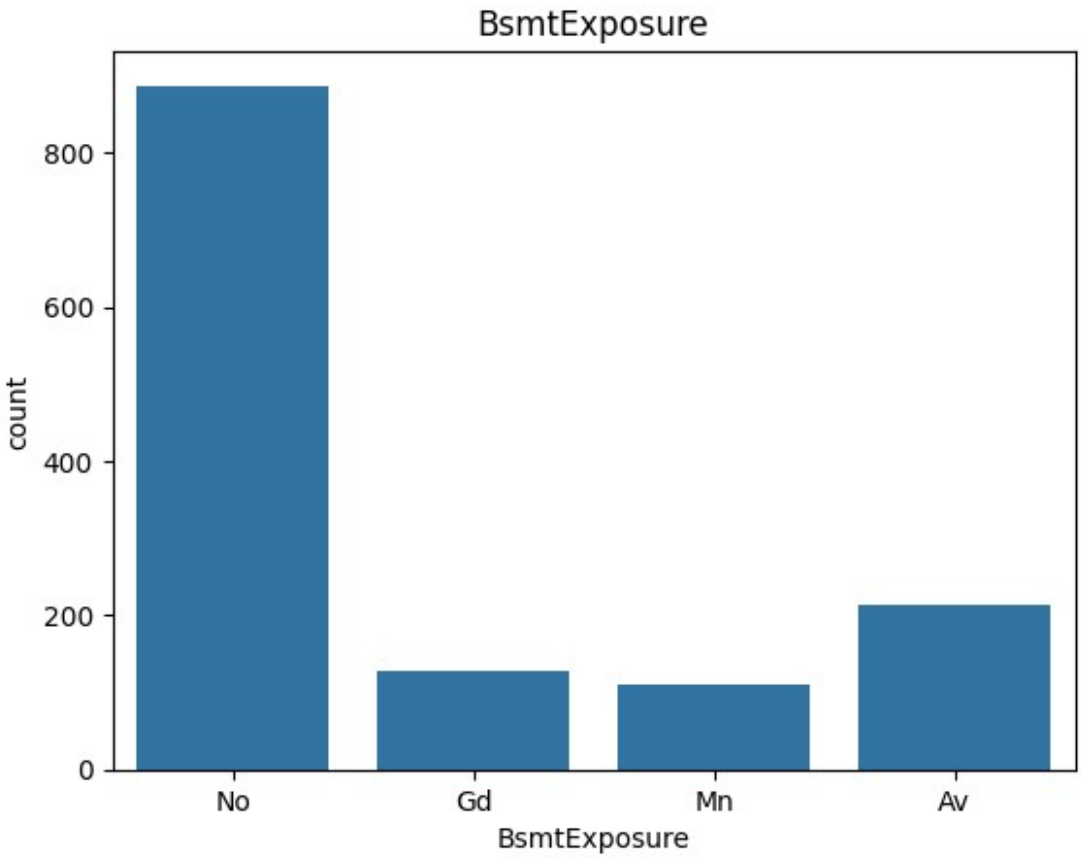

ExterQual
800
700
600
500
400
300
200
100
0
count
Gd
TA
Ex
Fa
ExterQual

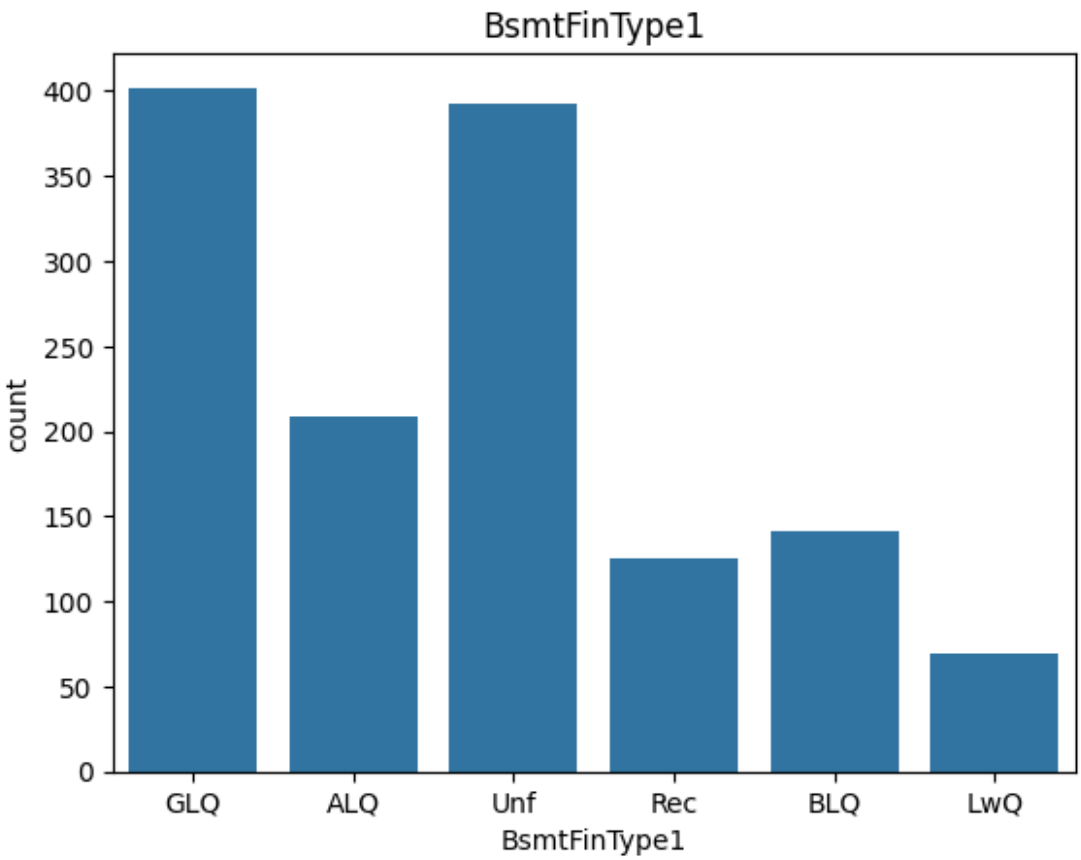

ExterCond
1200
1000
800
600
400
200
0
count
TA
Gd
Fa
Ex
ExterCond

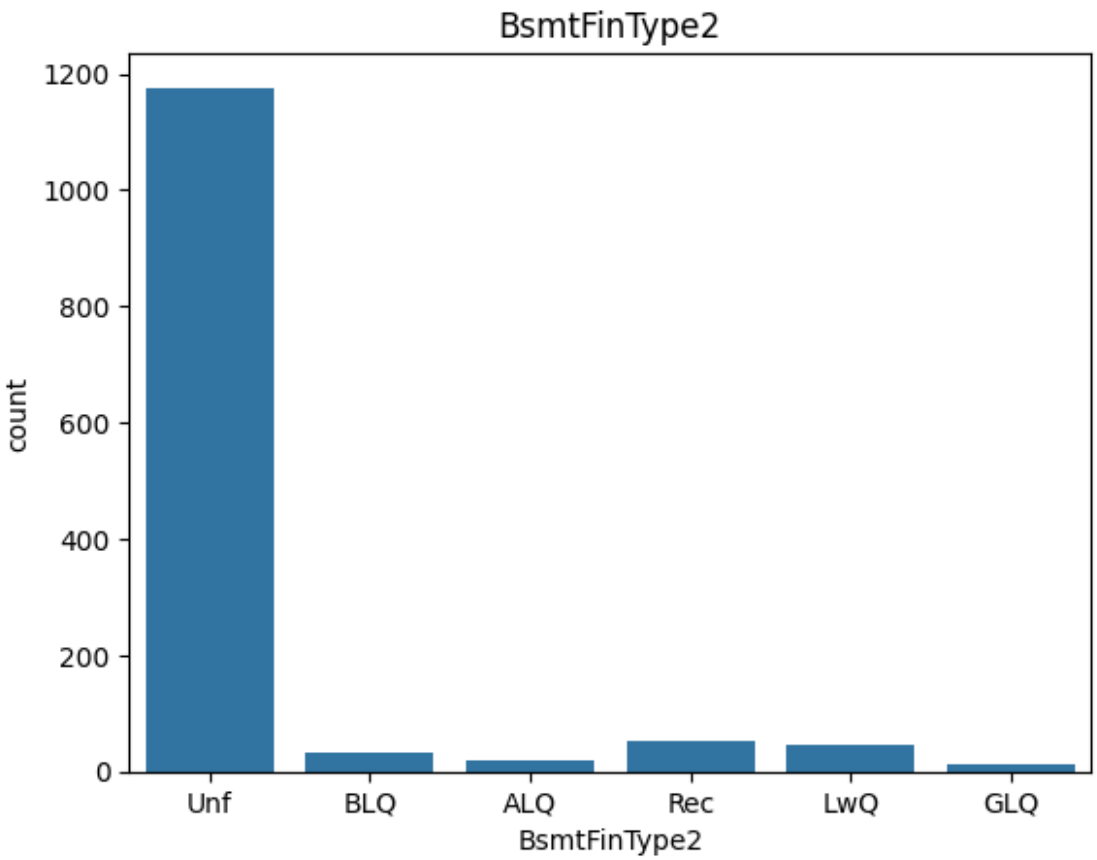

Foundation
600
500
400
300
200
100
0
count
PConc
CBlock
BrkTil
Wood
Stone
Foundation

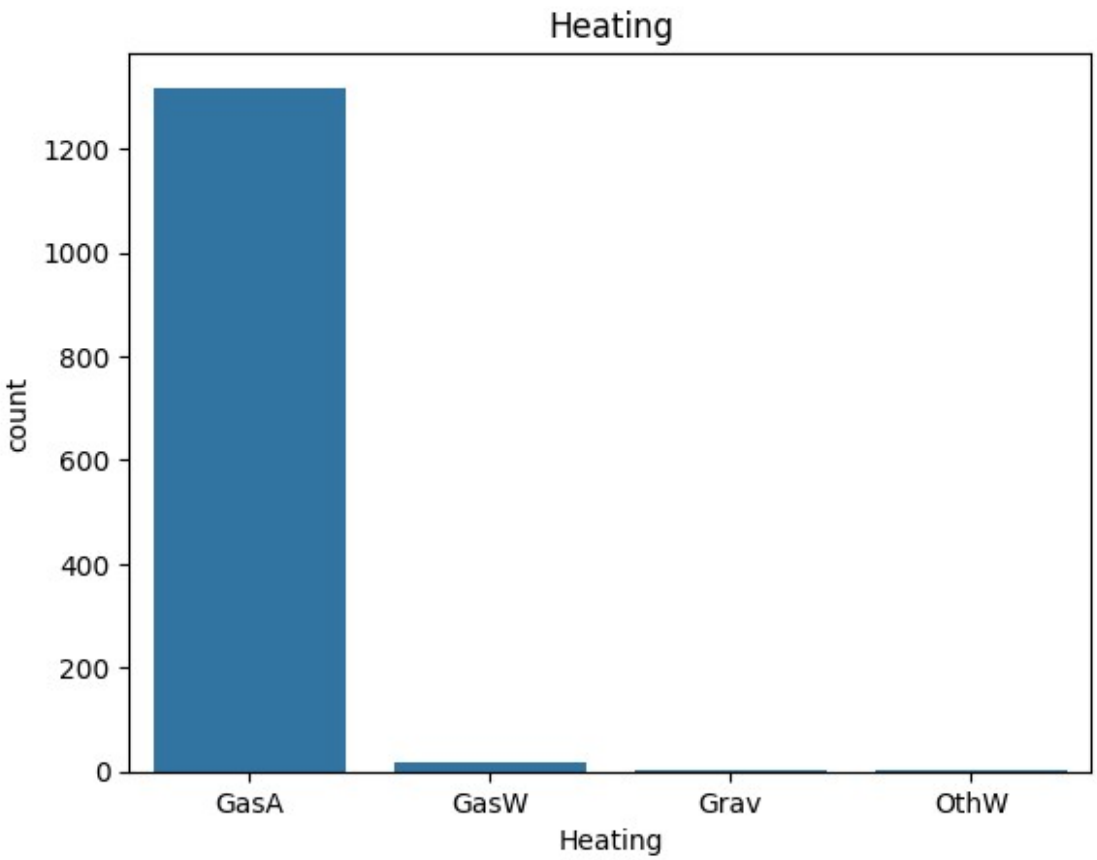

BsmtQual
600
500
400
300
200
100
0
count
Gd
TA
Ex
Fa
BsmtQual

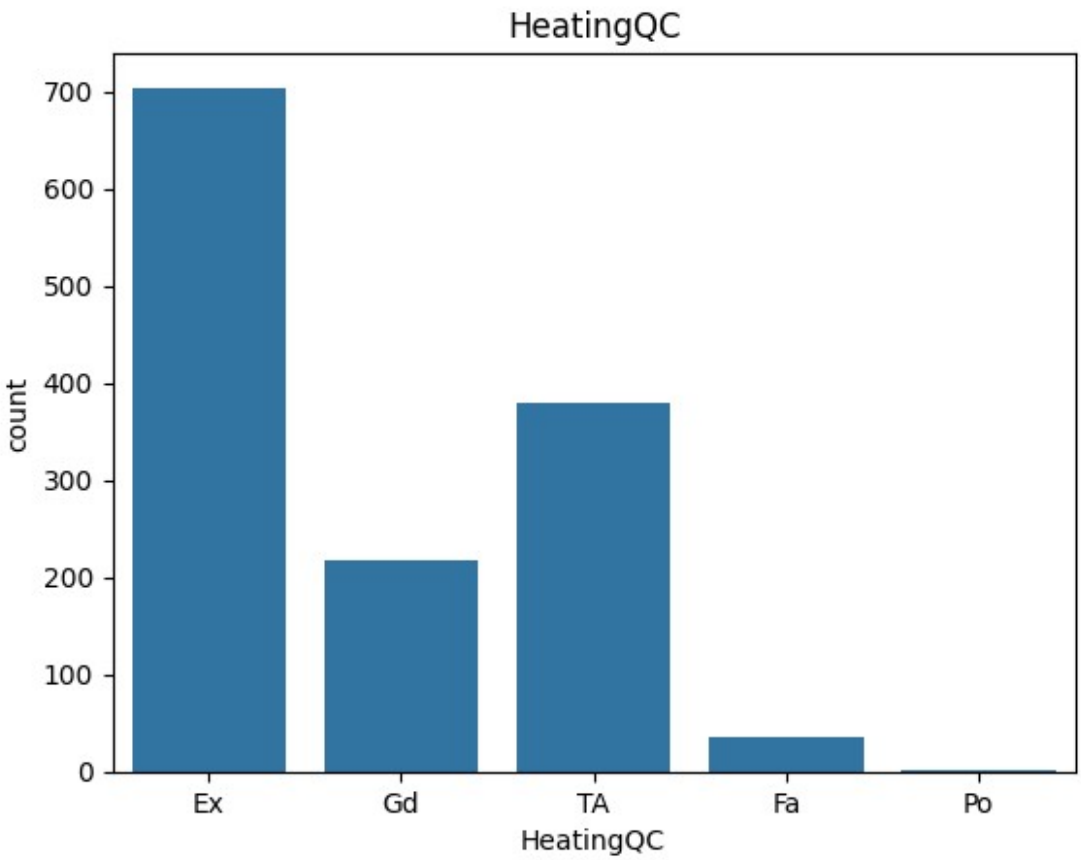

BsmtCond
1200
1000
800
600
400
200
0
count
TA
Gd
Fa
Po
BsmtCond

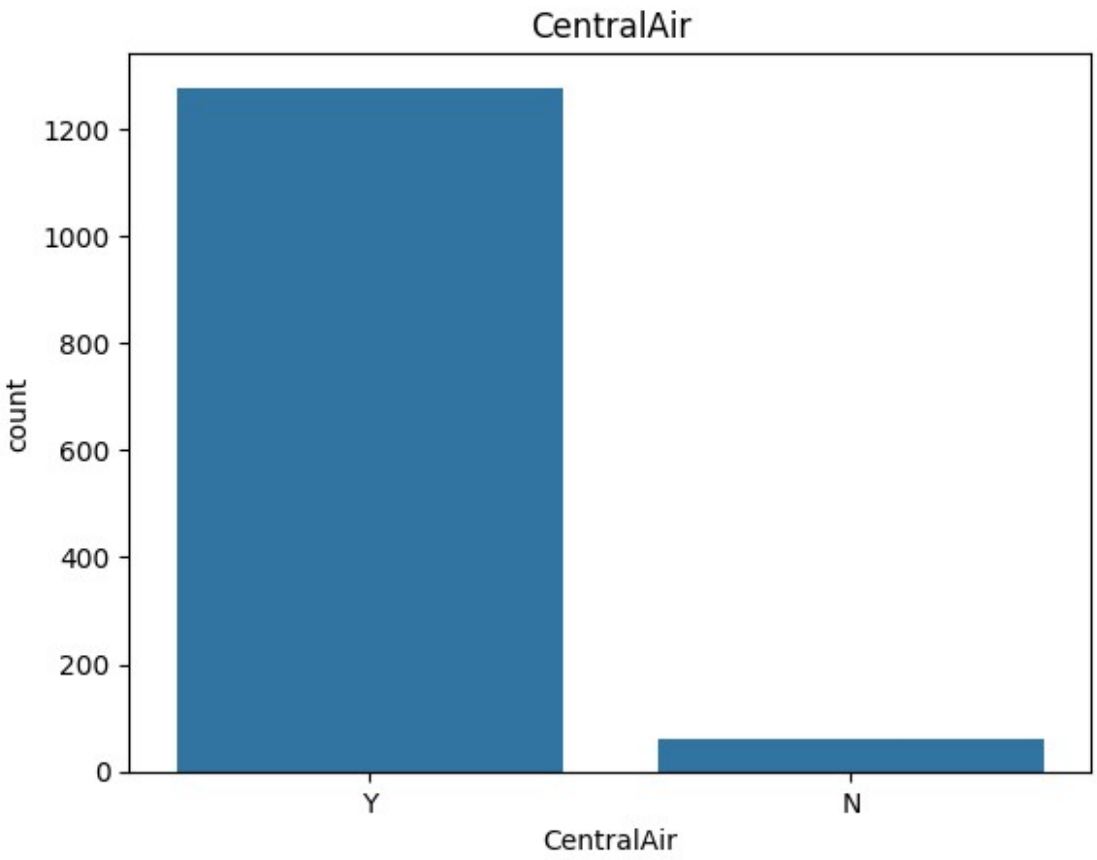

BsmtExposure
800
600
400
200
0
count
No
Gd
Mn
Av
BsmtExposure

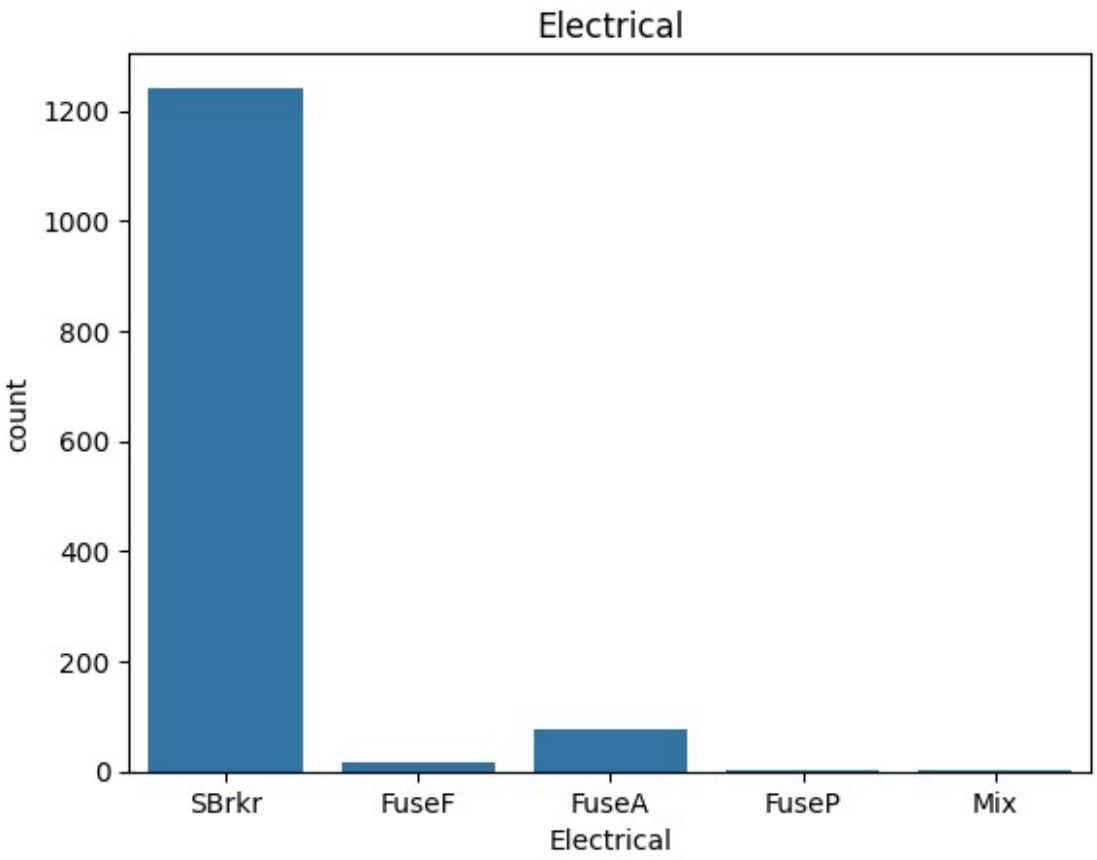

BsmtFinType1
400
350
300
250
200
150
100
50
0
count
GLQ
ALQ
Unf
Rec
BLQ
LwQ
BsmtFinType1

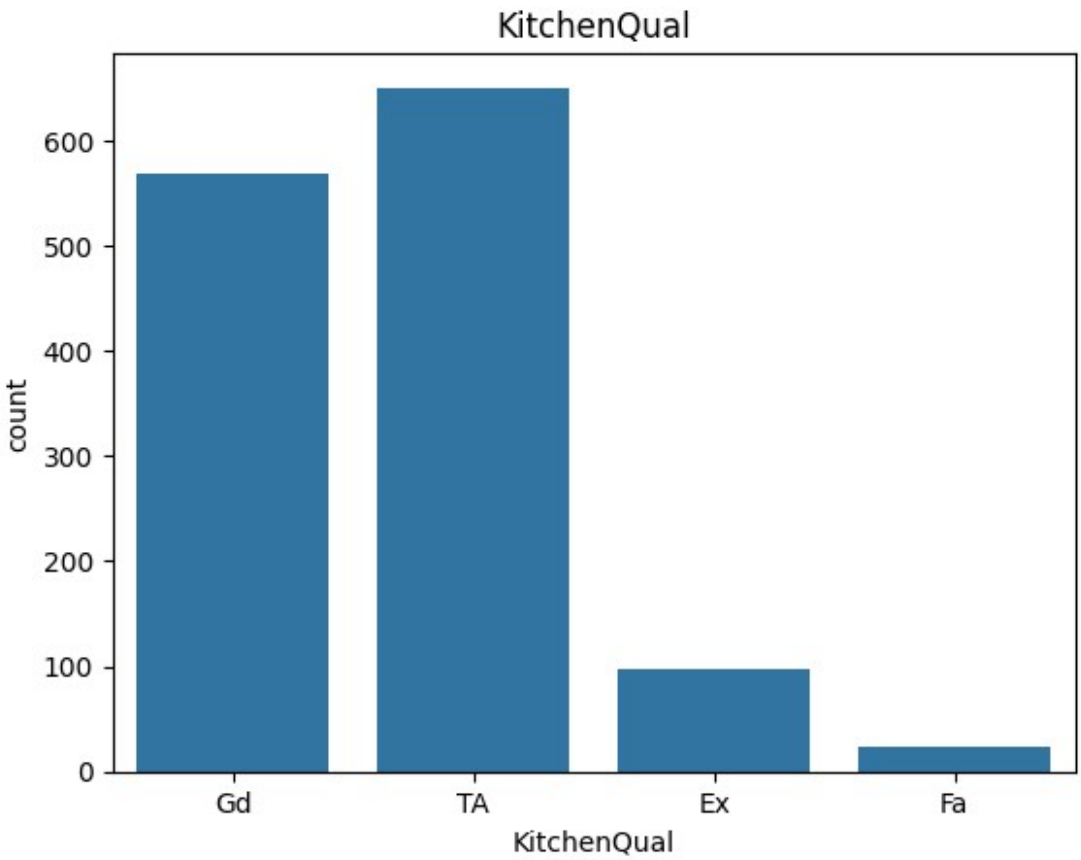

BsmtFinType2
1200
1000
800
600
400
200
0
count
Unf
BLQ
ALQ
Rec
LwQ
GLQ
BsmtFinType2

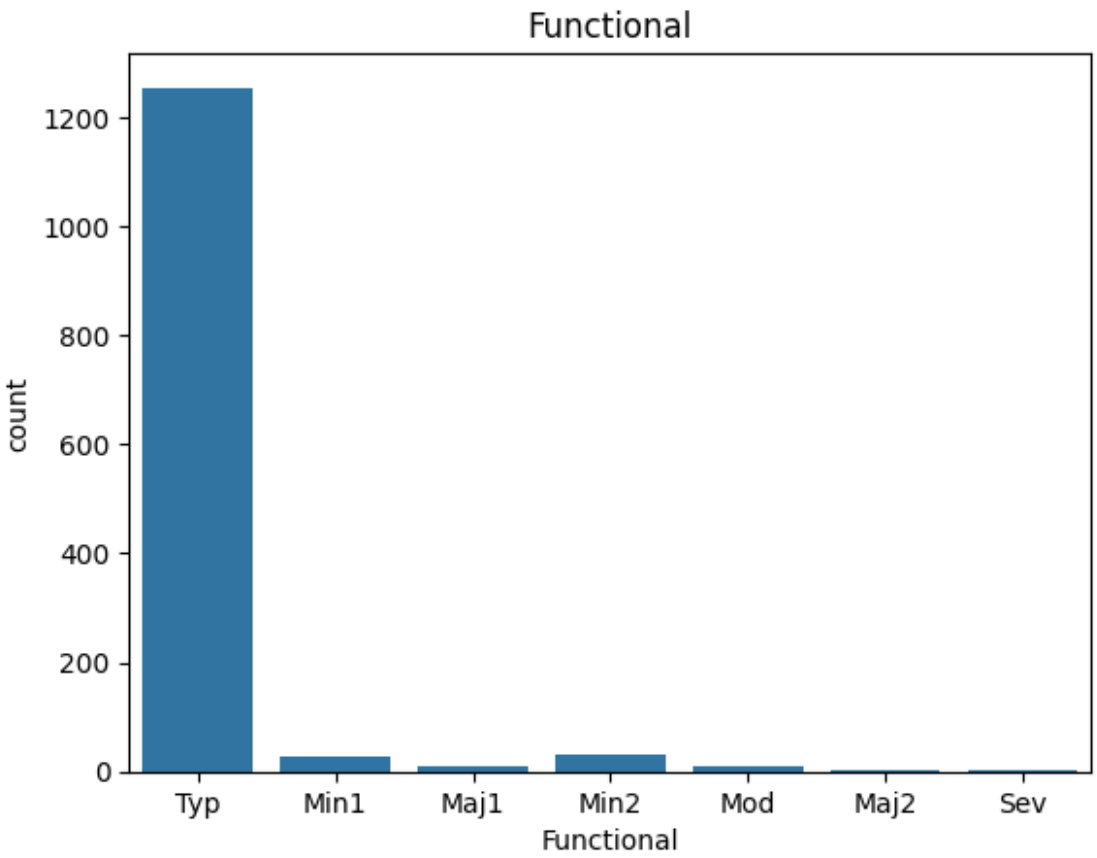

Heating
1200
1000
800
600
400
200
0
count
GasA
GasW
Grav
OthW
Heating

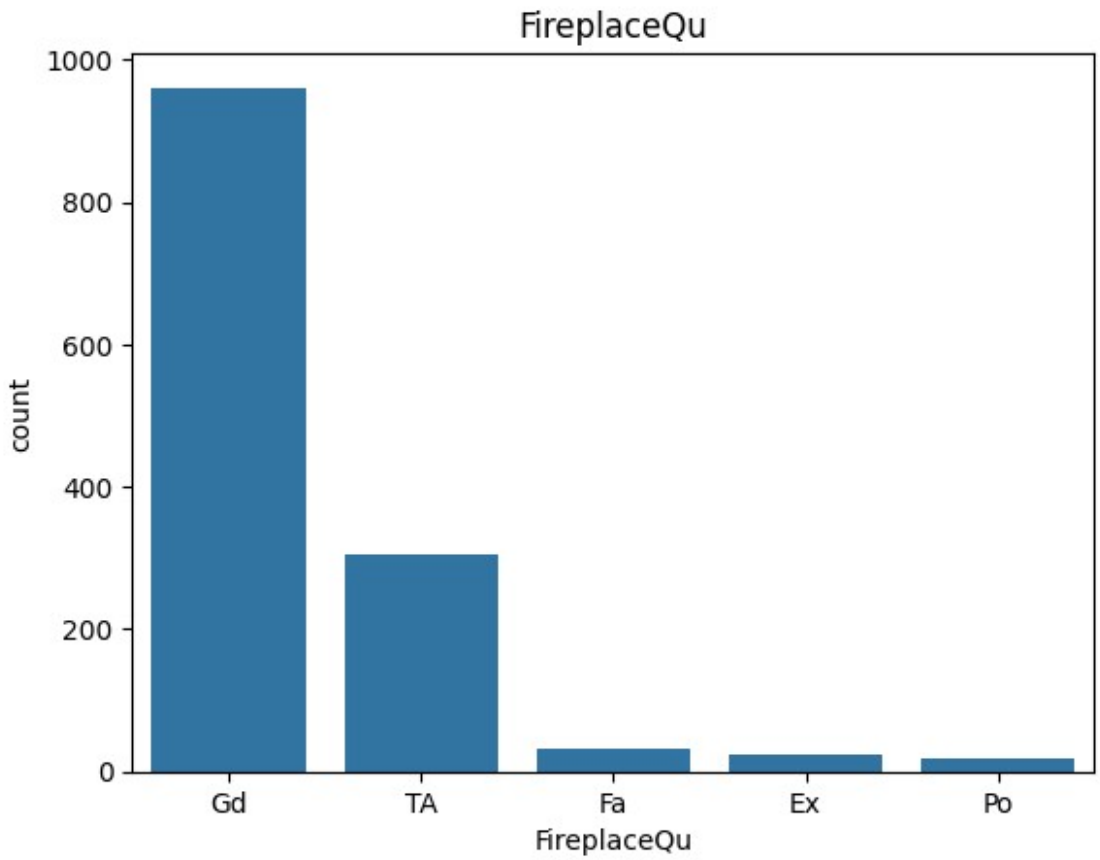

HeatingQC
700
600
500
400
300
200
100
0
count
Ex
Gd
TA
Fa
Po
HeatingQC

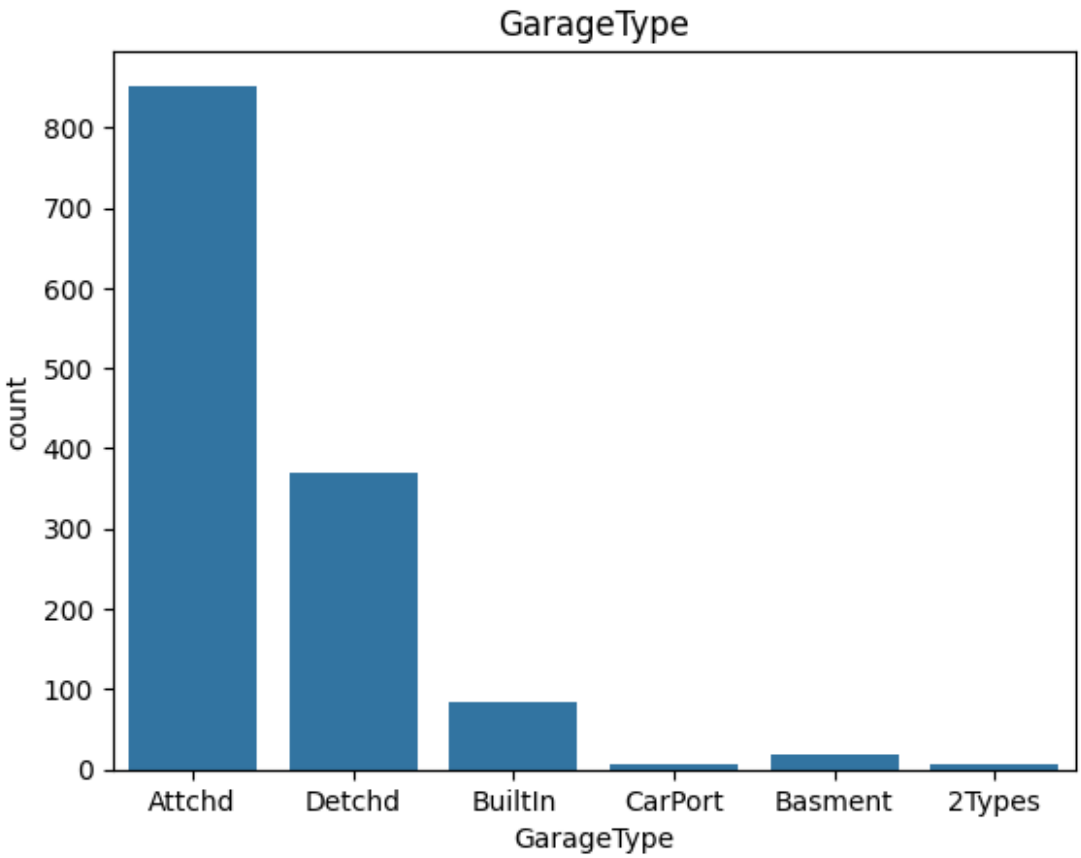

CentralAir
1200
1000
800
600
400
200
0
count
Y
N
CentralAir

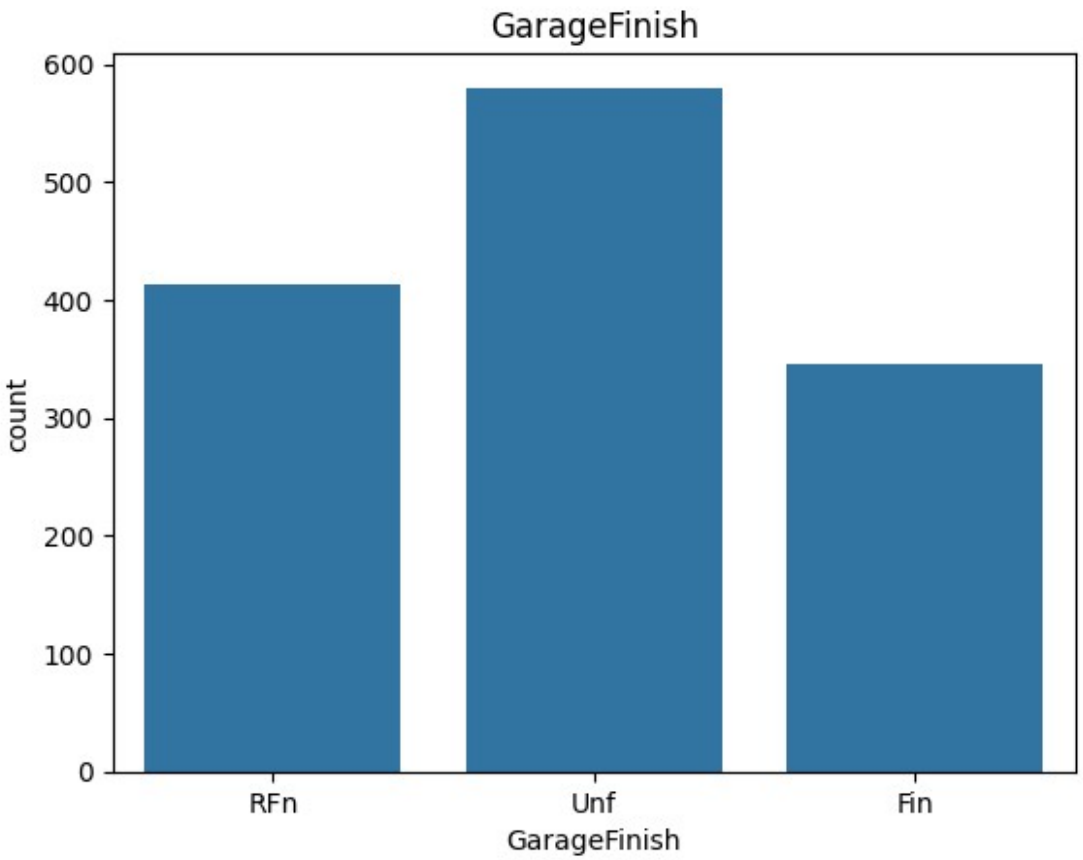

Electrical
1200
1000
800
600
400
200
0
count
SBrkr
FuseF
FuseA
FuseP
Mix
Electrical

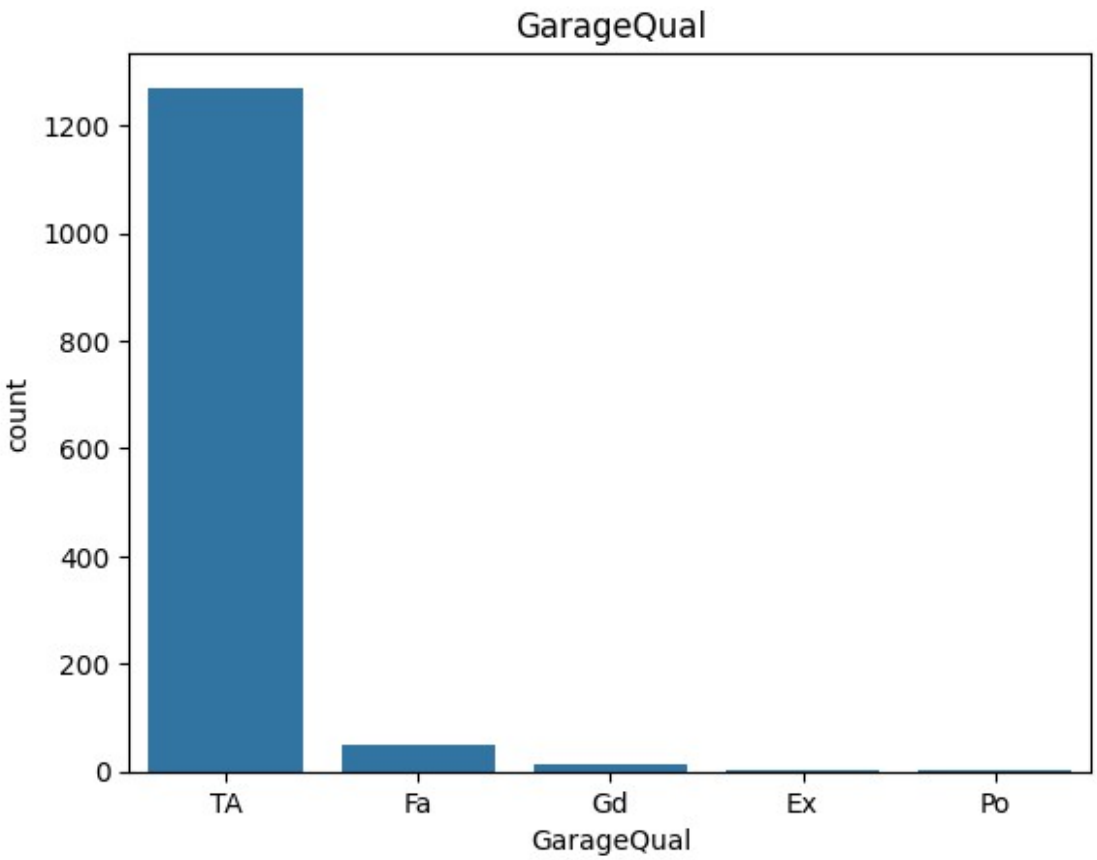

KitchenQual
600
500
400
300
200
100
0
count
Gd
TA
Ex
Fa
KitchenQual

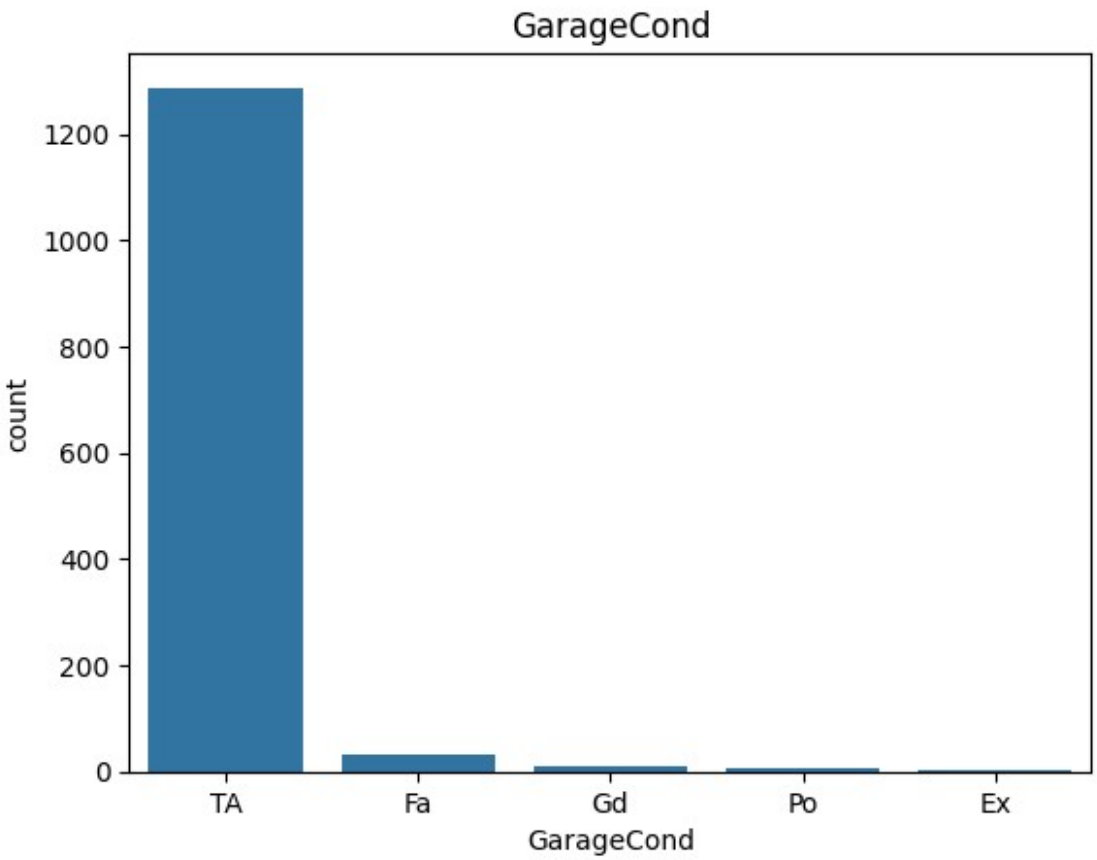

Functional
1200
1000
800
600
400
200
0
count
Typ
Min1
Maj1
Min2
Mod
Maj2
Sev
Functional

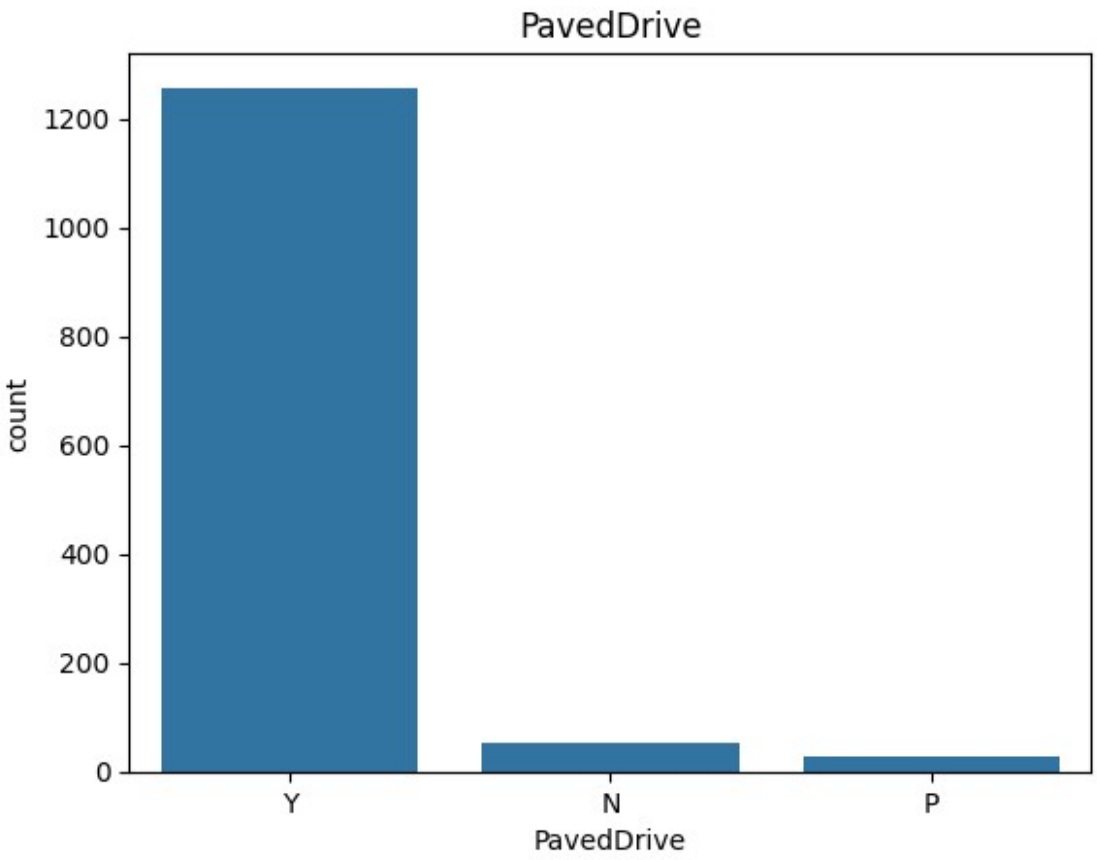

FireplaceQu
1000
800
600
400
200
0
count
Gd
TA
Fa
Ex
Po
FireplaceQu

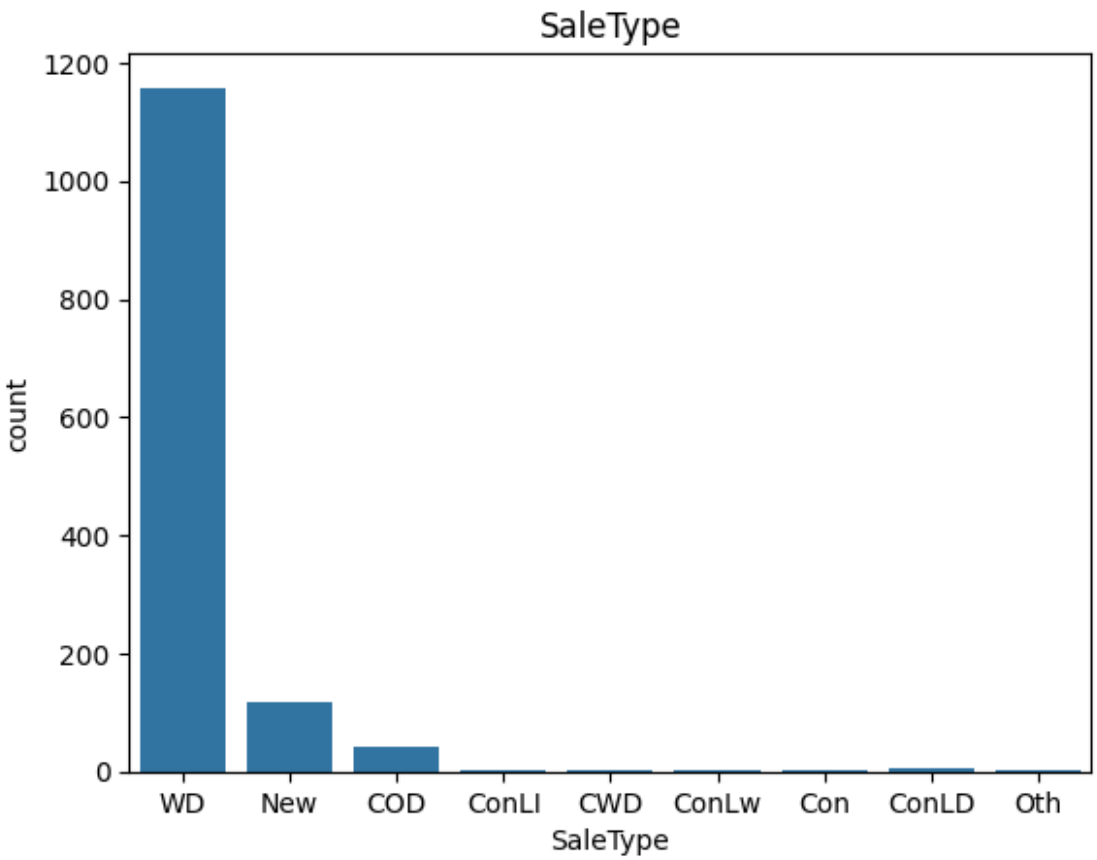

GarageType
count
800
700
600
500
400
300
200
100
0
Attchd
Detchd
BuiltIn
CarPort
Basment
2Types
GarageType

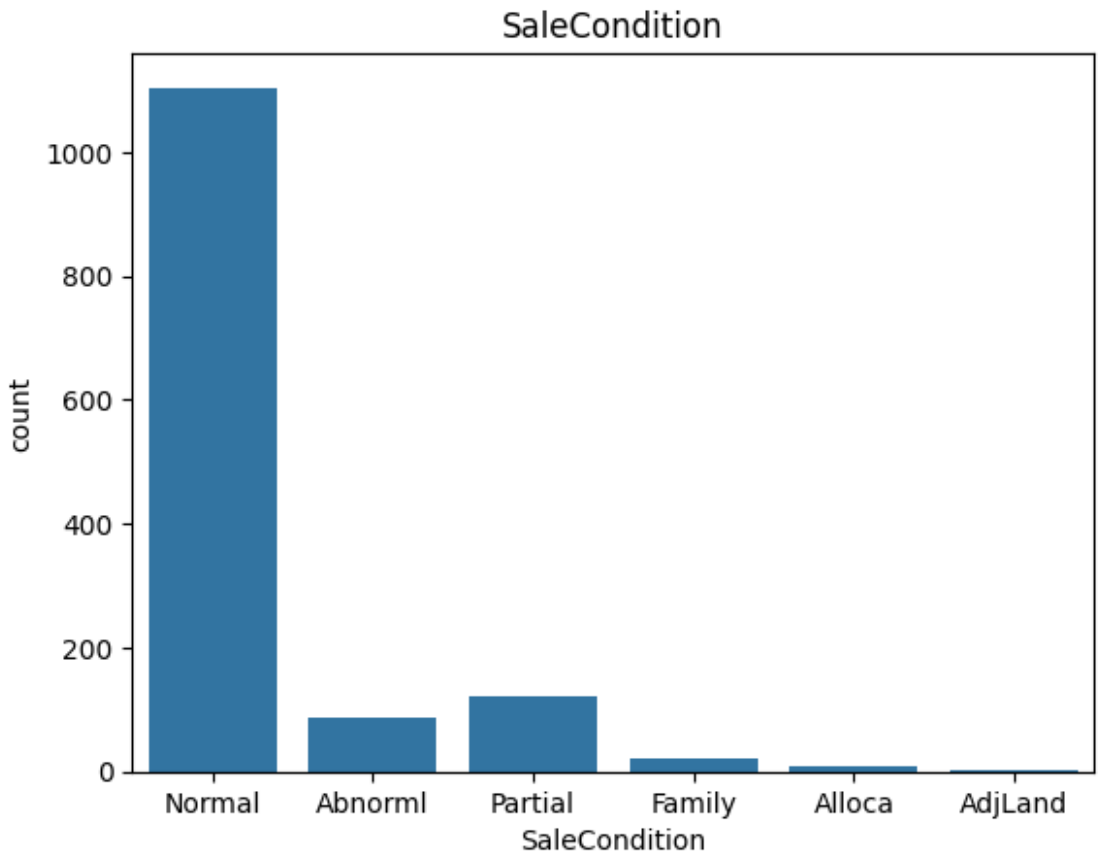

GarageFinish
600
500
400
300
200
100
0
count
RFn
Unf
Fin
GarageFinish


GarageQual
1200
1000
800
600
400
200
0
count
TA
Fa
Gd
Ex
Po
GarageQual


GarageCond
count
1200
1000
800
600
400
200
0
TA
Fa
Gd
Po
Ex
GarageCond


PavedDrive
1200
1000
800
600
400
200
0
count
Y
N
P
PavedDrive


SaleType
1200
1000
800
600
400
200
0
count
WD
New
COD
ConLI
CWD
ConLw
Con
ConLD
Oth
SaleType




```
def fn2(p):
  for i in p:
    s = df[i].value_counts()
    print(i)
    print(s)
fn2(p)
```

```
MSZoning
MSZoning
RL         1066
RM          191
FV           62
RH           11
C (all)       8
Name: count, dtype: int64
Street
Street
Pave    1333
Grvl       5
Name: count, dtype: int64
LotShape
LotShape
Reg    829
```

```
IR1    459
IR2     40
IR3     10
Name: count, dtype: int64
LandContour
LandContour
Lvl    1206
Bnk      52
HLS      48
Low      32
Name: count, dtype: int64
Utilities
Utilities
AllPub    1337
NoSeWa       1
Name: count, dtype: int64
LotConfig
LotConfig
Inside     957
Corner     244
CulDSac     90
FR2         43
FR3          4
Name: count, dtype: int64
LandSlope
LandSlope
Gtl    1265
Mod      61
Sev      12
Name: count, dtype: int64
Neighborhood
Neighborhood
NAmes      209
CollgCr    146
OldTown    100
Somerst     83
Gilbert     77
NridgHt     75
NWAmes      73
Edwards     70
Sawyer      69
SawyerW     53
Crawfor     50
BrkSide     47
Mitchel     42
NoRidge     41
Timber      37
IDOTRR      29
ClearCr     26
```

```
StoneBr     25
SWISU       20
Blmngtn     17
BrDale      15
MeadowV     12
Veenker     11
NPkVill      9
Blueste      2
Name: count, dtype: int64
Condition1
Condition1
Norm      1162
Feedr       63
Artery      43
RRAn        26
PosN        19
RRAe        10
PosA         8
RRNn         5
RRNe         2
Name: count, dtype: int64
Condition2
Condition2
Norm      1324
Feedr        5
Artery       2
RRNn         2
PosN         2
PosA         1
RRAn         1
RRAe         1
Name: count, dtype: int64
BldgType
BldgType
1Fam      1138
TwnhsE     112
Twnhs       38
Duplex      28
2fmCon      22
Name: count, dtype: int64
HouseStyle
HouseStyle
1Story    657
2Story    426
1.5Fin    134
SLvl       64
SFoyer     30
1.5Unf     11
2.5Unf     10
```

```
2.5Fin      6
Name: count, dtype: int64
RoofStyle
RoofStyle
Gable      1037
Hip         272
Flat         11
Gambrel      10
Mansard       6
Shed          2
Name: count, dtype: int64
RoofMatl
RoofMatl
CompShg    1314
Tar&Grv       9
WdShngl       6
WdShake       5
Metal         1
Membran       1
Roll          1
ClyTile       1
Name: count, dtype: int64
Exterior1st
Exterior1st
VinylSd    486
HdBoard    211
MetalSd    201
Wd Sdng    183
Plywood    100
CemntBd     52
BrkFace     44
Stucco      21
WdShing     20
AsbShng     15
Stone        2
BrkComm      1
ImStucc      1
CBlock       1
Name: count, dtype: int64
Exterior2nd
Exterior2nd
VinylSd    475
MetalSd    197
HdBoard    197
Wd Sdng    176
Plywood    127
CmentBd     51
Wd Shng     32
Stucco      23
```

```
BrkFace     22
AsbShng     16
ImStucc     10
Brk Cmn      6
AsphShn      2
Stone        2
Other        1
CBlock       1
Name: count, dtype: int64
MasVnrType
MasVnrType
BrkFace    1195
Stone       128
BrkCmn       15
Name: count, dtype: int64
ExterQual
ExterQual
TA    803
Gd    477
Ex     51
Fa      7
Name: count, dtype: int64
ExterCond
ExterCond
TA    1183
Gd     137
Fa      16
Ex       2
Name: count, dtype: int64
Foundation
Foundation
PConc     620
CBlock    580
BrkTil    129
Stone       6
Wood        3
Name: count, dtype: int64
BsmtQual
BsmtQual
TA    594
Gd    592
Ex    120
Fa     32
Name: count, dtype: int64
BsmtCond
BsmtCond
TA    1237
Gd      62
Fa      38
```

```
Po       1
Name: count, dtype: int64
BsmtExposure
BsmtExposure
No    887
Av    213
Gd    127
Mn    111
Name: count, dtype: int64
BsmtFinType1
BsmtFinType1
GLQ    402
Unf    392
ALQ    209
BLQ    141
Rec    125
LwQ     69
Name: count, dtype: int64
BsmtFinType2
BsmtFinType2
Unf    1176
Rec      53
LwQ      46
BLQ      32
ALQ      19
GLQ      12
Name: count, dtype: int64
Heating
Heating
GasA    1318
GasW      16
Grav       3
OthW       1
Name: count, dtype: int64
HeatingQC
HeatingQC
Ex    704
TA    380
Gd    217
Fa     36
Po      1
Name: count, dtype: int64
CentralAir
CentralAir
Y    1277
N      61
Name: count, dtype: int64
Electrical
Electrical
```

```
SBrkr    1242
FuseA      76
FuseF      17
FuseP       2
Mix         1
Name: count, dtype: int64
KitchenQual
KitchenQual
TA    650
Gd    568
Ex     97
Fa     23
Name: count, dtype: int64
Functional
Functional
Typ     1254
Min2      30
Min1      28
Mod       11
Maj1      10
Maj2       4
Sev        1
Name: count, dtype: int64
FireplaceQu
FireplaceQu
Gd    961
TA    304
Fa     33
Ex     23
Po     17
Name: count, dtype: int64
GarageType
GarageType
Attchd     852
Detchd     369
BuiltIn     85
Basment     19
CarPort      7
2Types       6
Name: count, dtype: int64
GarageFinish
GarageFinish
Unf    580
RFn    413
Fin    345
Name: count, dtype: int64
GarageQual
GarageQual
TA    1270
```

```
Fa      48
Gd      14
Ex       3
Po       3
Name: count, dtype: int64
GarageCond
GarageCond
TA    1287
Fa      33
Gd       9
Po       7
Ex       2
Name: count, dtype: int64
PavedDrive
PavedDrive
Y    1257
N      54
P      27
Name: count, dtype: int64
SaleType
SaleType
WD       1158
New       117
COD        42
ConLD       6
ConLI       4
CWD         4
ConLw       4
Con         2
Oth         1
Name: count, dtype: int64
SaleCondition
SaleCondition
Normal     1104
Partial     120
Abnorml      86
Family       20
Alloca        7
AdjLand       1
Name: count, dtype: int64
```

```
df=df.drop(['Street','Condition2','RoofStyle','Heating','Id'],axis=1)
df.shape
```

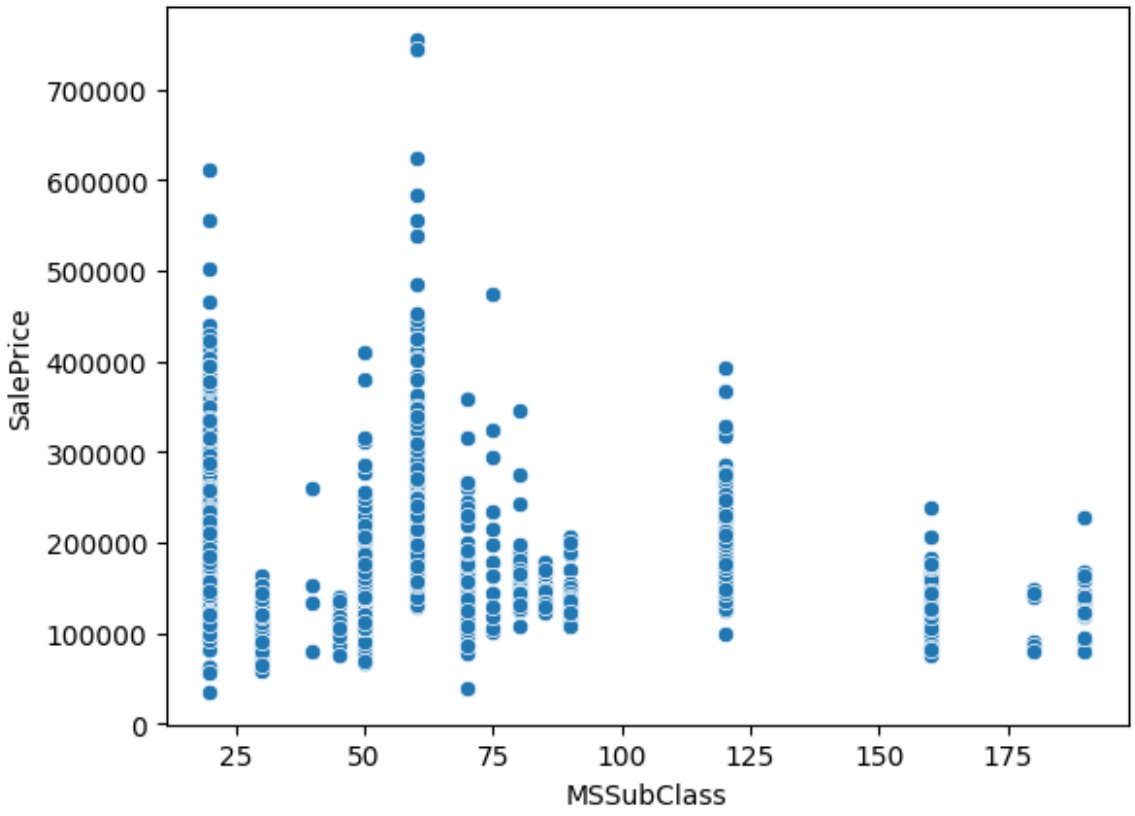
```
(1338, 72)
```

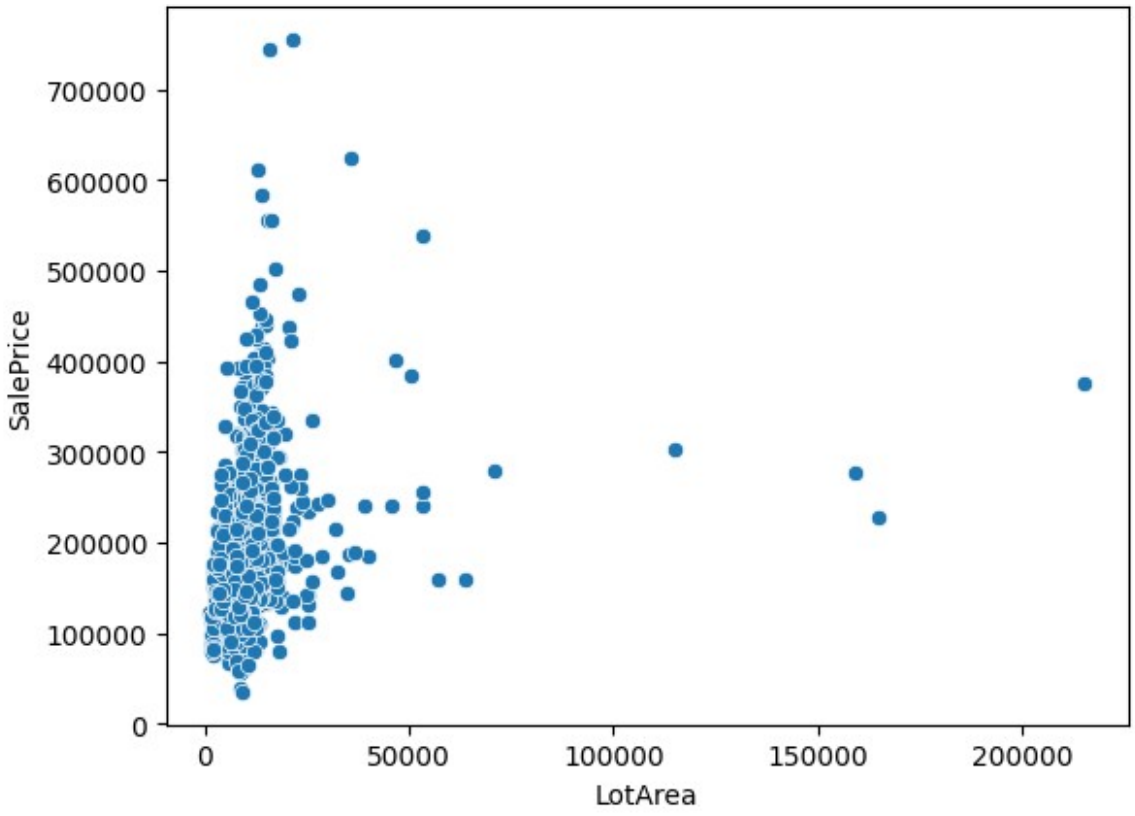
```
df.info()
```

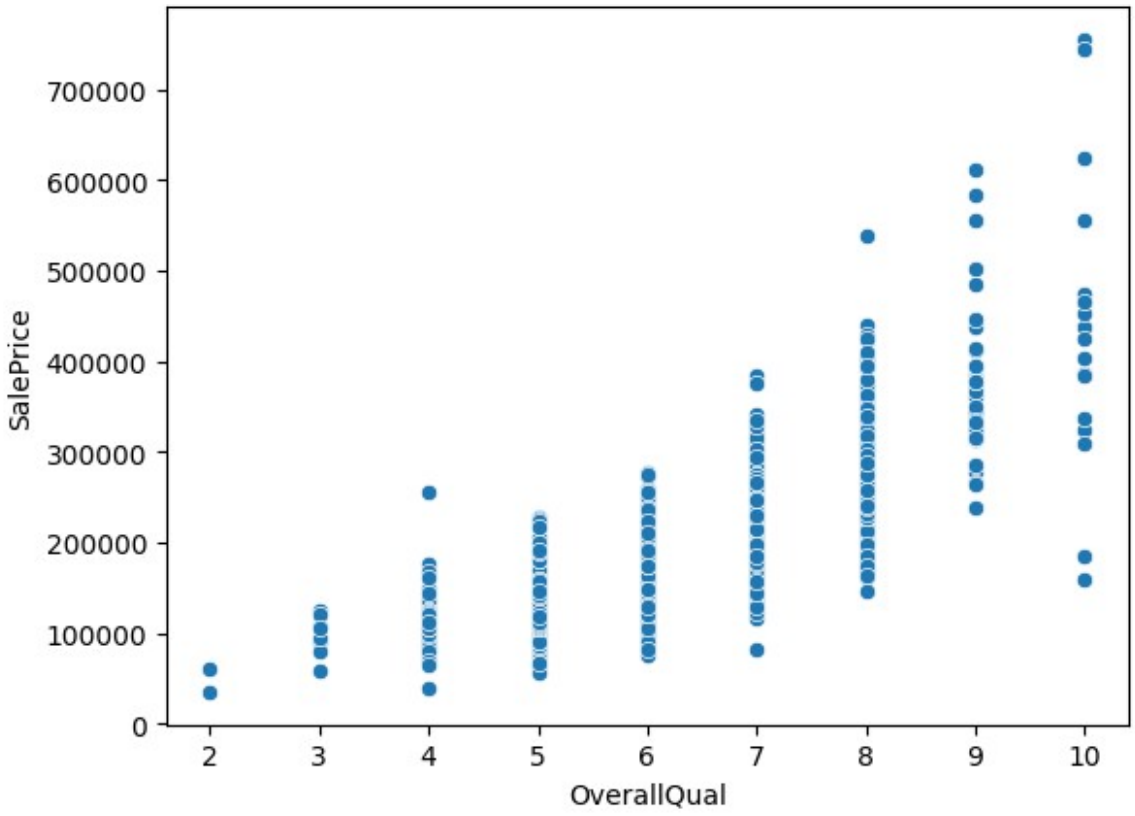
```
<class 'pandas.core.frame.DataFrame'>
Index: 1338 entries, 0 to 1459
```

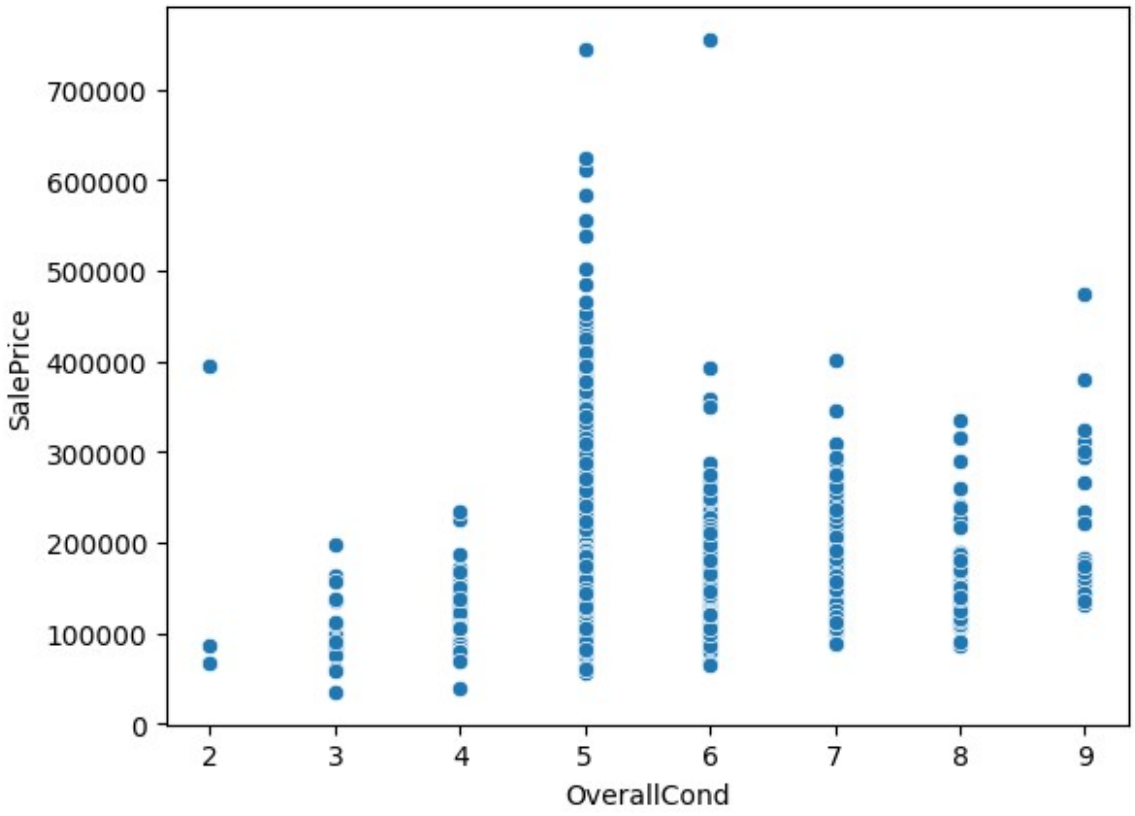
```
Data columns (total 72 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   MSSubClass      1338 non-null   int64
 1   MSZoning        1338 non-null   object
 2   LotFrontage     1338 non-null   float64
 3   LotArea         1338 non-null   int64
 4   LotShape        1338 non-null   object
 5   LandContour     1338 non-null   object
 6   Utilities       1338 non-null   object
 7   LotConfig       1338 non-null   object
 8   LandSlope       1338 non-null   object
 9   Neighborhood    1338 non-null   object
 10  Condition1      1338 non-null   object
 11  BldgType        1338 non-null   object
 12  HouseStyle      1338 non-null   object
 13  OverallQual     1338 non-null   int64
 14  OverallCond     1338 non-null   int64
 15  YearBuilt       1338 non-null   int64
 16  YearRemodAdd    1338 non-null   int64
 17  RoofMatl        1338 non-null   object
 18  Exterior1st     1338 non-null   object
 19  Exterior2nd     1338 non-null   object
 20  MasVnrType      1338 non-null   object
 21  MasVnrArea      1338 non-null   float64
 22  ExterQual       1338 non-null   object
 23  ExterCond       1338 non-null   object
 24  Foundation      1338 non-null   object
 25  BsmtQual        1338 non-null   object
 26  BsmtCond        1338 non-null   object
 27  BsmtExposure    1338 non-null   object
 28  BsmtFinType1    1338 non-null   object
 29  BsmtFinSF1      1338 non-null   int64
 30  BsmtFinType2    1338 non-null   object
 31  BsmtFinSF2      1338 non-null   int64
 32  BsmtUnfSF       1338 non-null   int64
 33  TotalBsmtSF     1338 non-null   int64
 34  HeatingQC       1338 non-null   object
 35  CentralAir      1338 non-null   object
 36  Electrical      1338 non-null   object
 37  1stFlrSF        1338 non-null   int64
 38  2ndFlrSF        1338 non-null   int64
 39  LowQualFinSF    1338 non-null   int64
 40  GrLivArea       1338 non-null   int64
 41  BsmtFullBath    1338 non-null   int64
 42  BsmtHalfBath    1338 non-null   int64
 43  FullBath        1338 non-null   int64
 44  HalfBath        1338 non-null   int64
 45  BedroomAbvGr    1338 non-null   int64
```

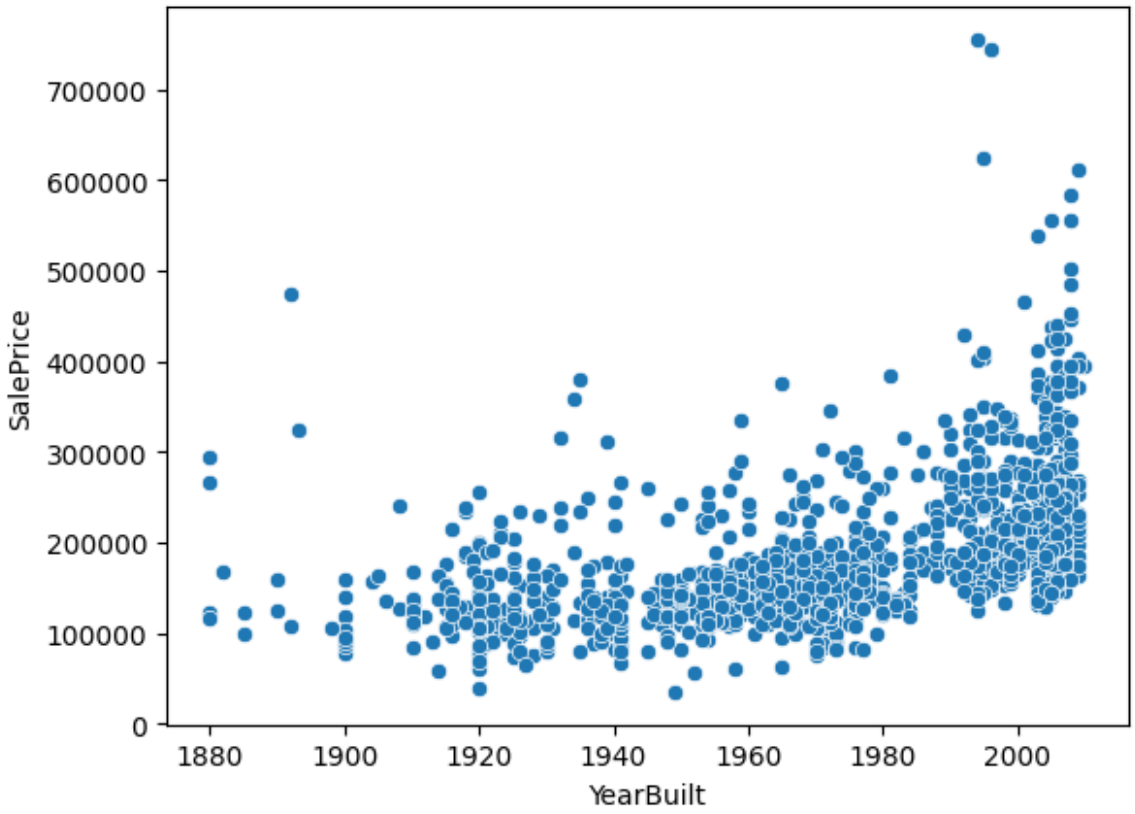
```
 46  KitchenAbvGr   1338 non-null   int64
 47  KitchenQual    1338 non-null   object
 48  TotRmsAbvGrd   1338 non-null   int64
 49  Functional     1338 non-null   object
 50  Fireplaces     1338 non-null   int64
 51  FireplaceQu    1338 non-null   object
 52  GarageType     1338 non-null   object
 53  GarageYrBlt    1338 non-null   float64
 54  GarageFinish   1338 non-null   object
 55  GarageCars     1338 non-null   int64
 56  GarageArea     1338 non-null   int64
 57  GarageQual     1338 non-null   object
 58  GarageCond     1338 non-null   object
 59  PavedDrive     1338 non-null   object
 60  WoodDeckSF     1338 non-null   int64
 61  OpenPorchSF    1338 non-null   int64
 62  EnclosedPorch  1338 non-null   int64
 63  3SsnPorch      1338 non-null   int64
 64  ScreenPorch    1338 non-null   int64
 65  PoolArea       1338 non-null   int64
 66  MiscVal        1338 non-null   int64
 67  MoSold         1338 non-null   int64
 68  YrSold         1338 non-null   int64
 69  SaleType       1338 non-null   object
 70  SaleCondition  1338 non-null   object
 71  SalePrice      1338 non-null   int64
dtypes: float64(3), int64(34), object(35)
memory usage: 763.1+ KB

int_data = df.select_dtypes(include='int64')
s = int_data.columns

def fn3(s):
  for i in s:
    sns.scatterplot(x=df[i],y=df['SalePrice'])
    plt.show()
fn3(s)
```

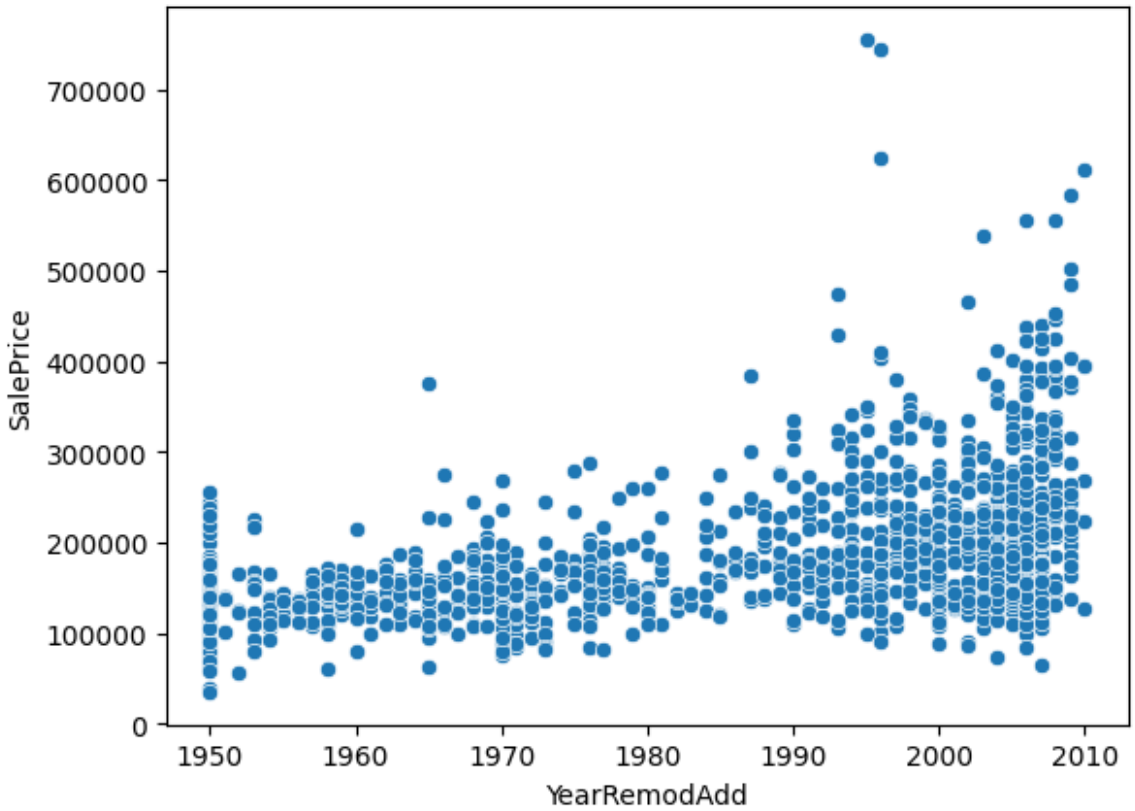

SalePrice
MSSubClass
0
100000
200000
300000
400000
500000
600000
700000
25
50
75
100
125
150
175

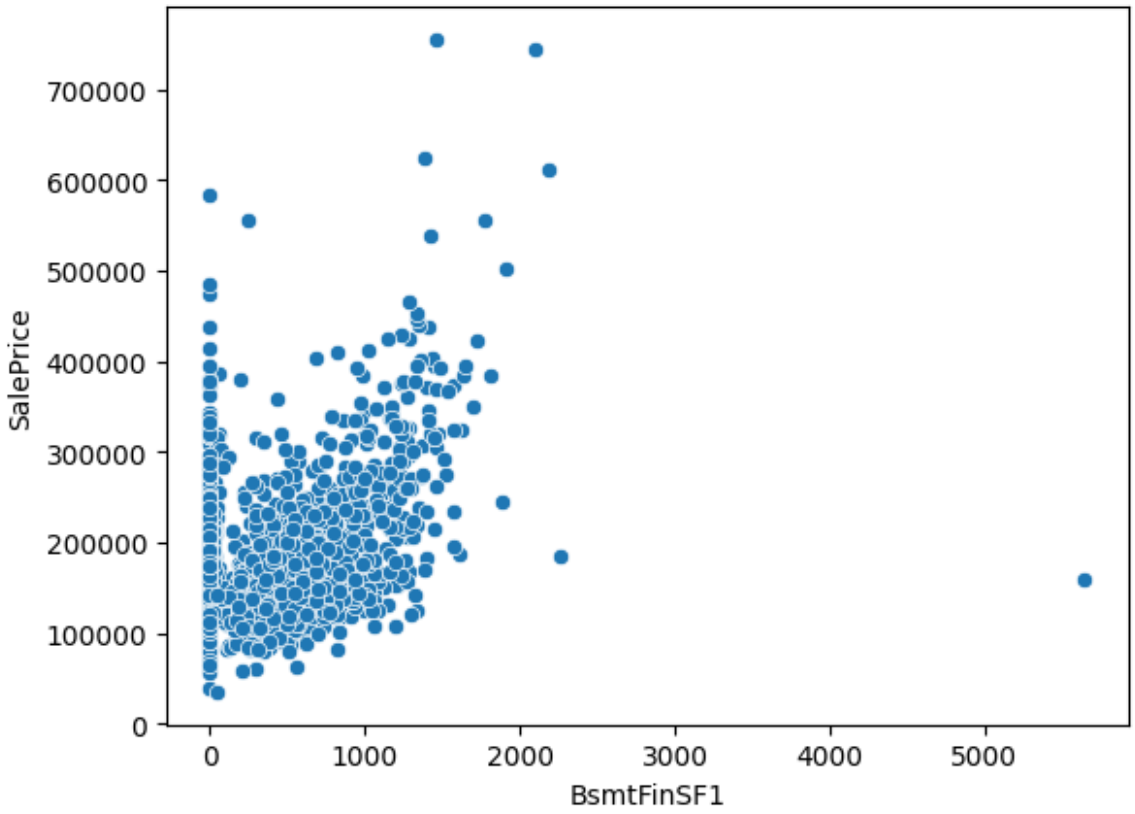

SalePrice
LotArea
0
100000
200000
300000
400000
500000
600000
700000
0
50000
100000
150000
200000

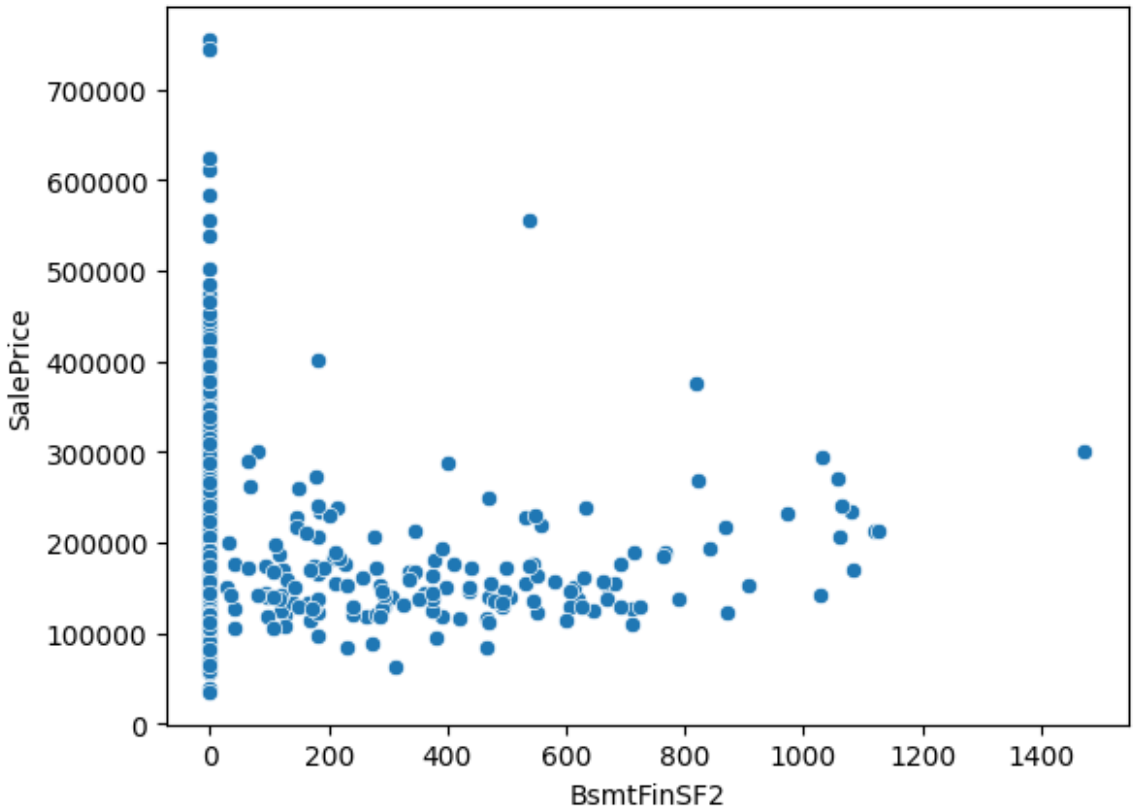

SalePrice
OverallQual
0
100000
200000
300000
400000
500000
600000
700000
2
3
4
5
6
7
8
9
10

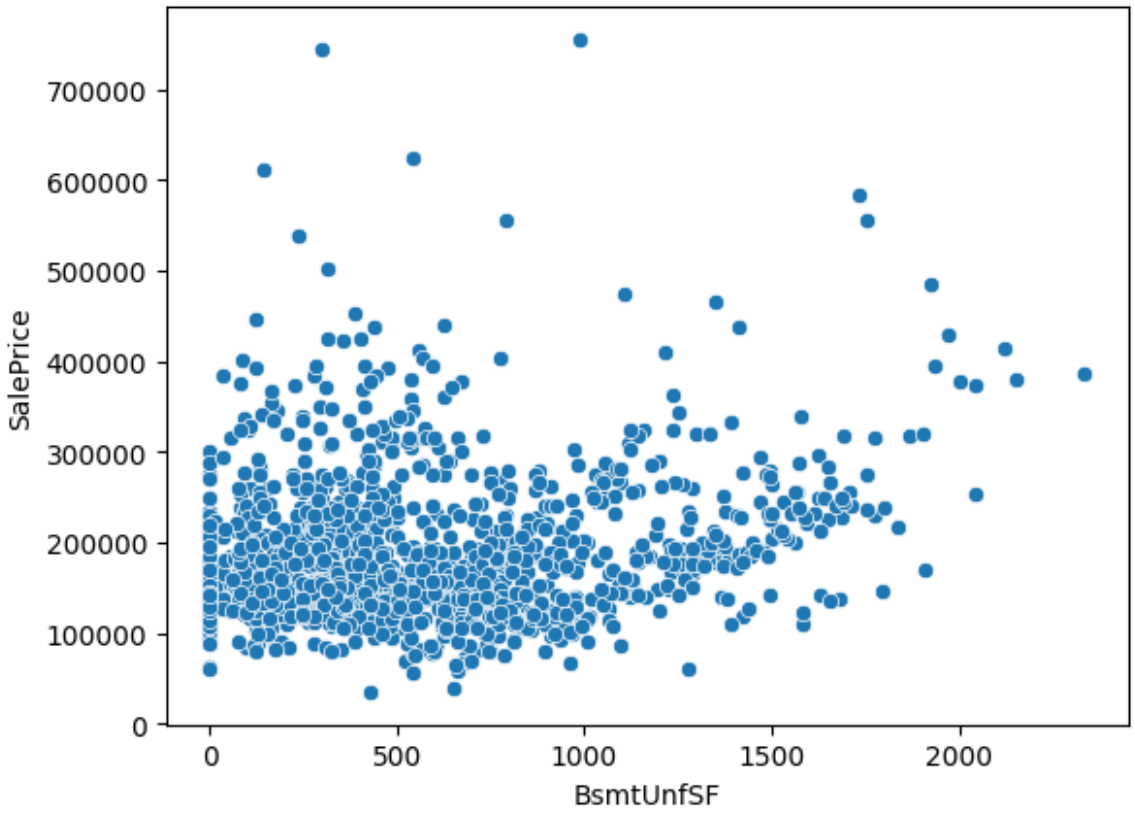

SalePrice
OverallCond
0
100000
200000
300000
400000
500000
600000
700000
2
3
4
5
6
7
8
9

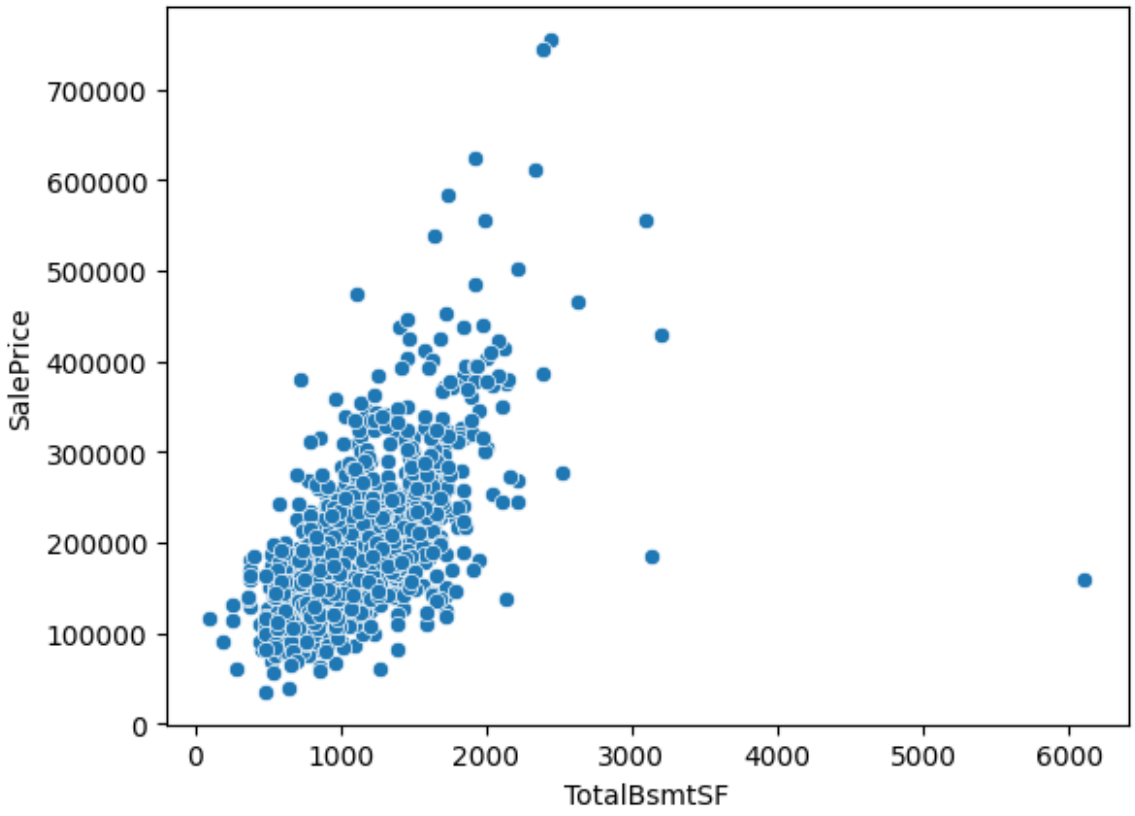

SalePrice
0
100000
200000
300000
400000
500000
600000
700000
1880
1900
1920
1940
1960
1980
2000
YearBuilt

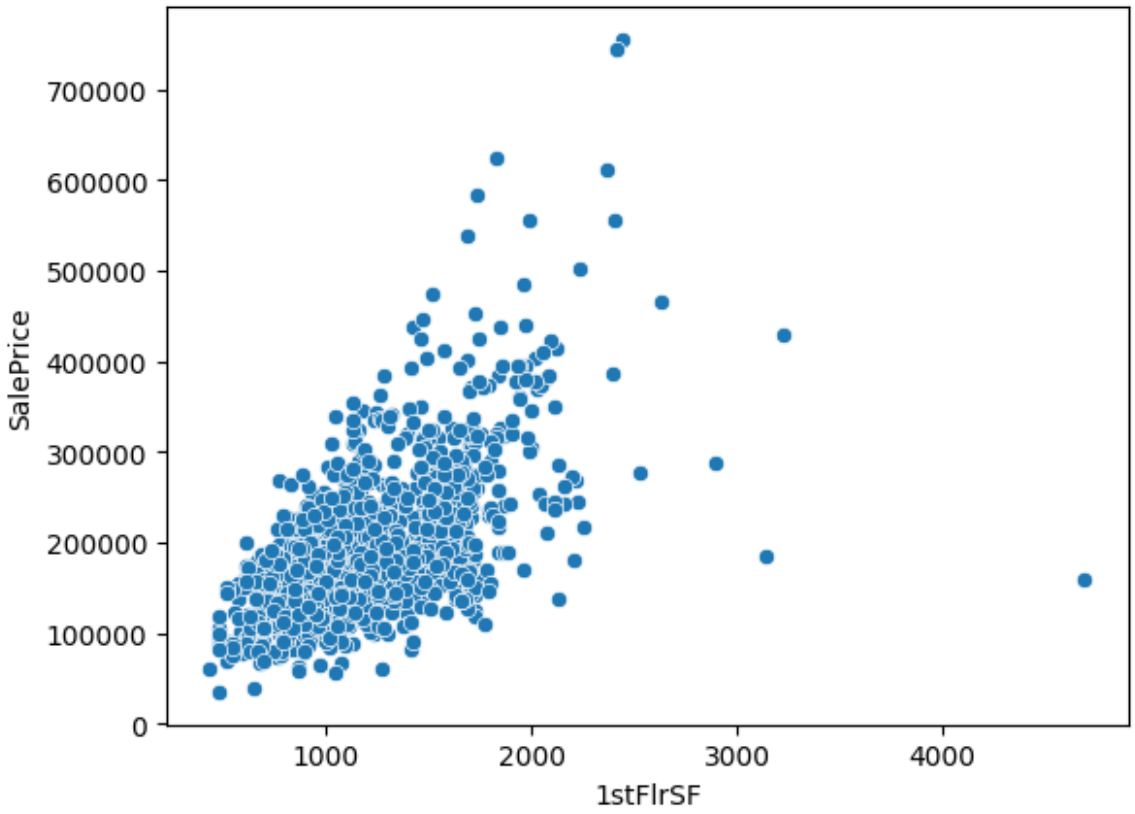

SalePrice
0
100000
200000
300000
400000
500000
600000
700000
1950
1960
1970
1980
1990
2000
2010
YearRemodAdd

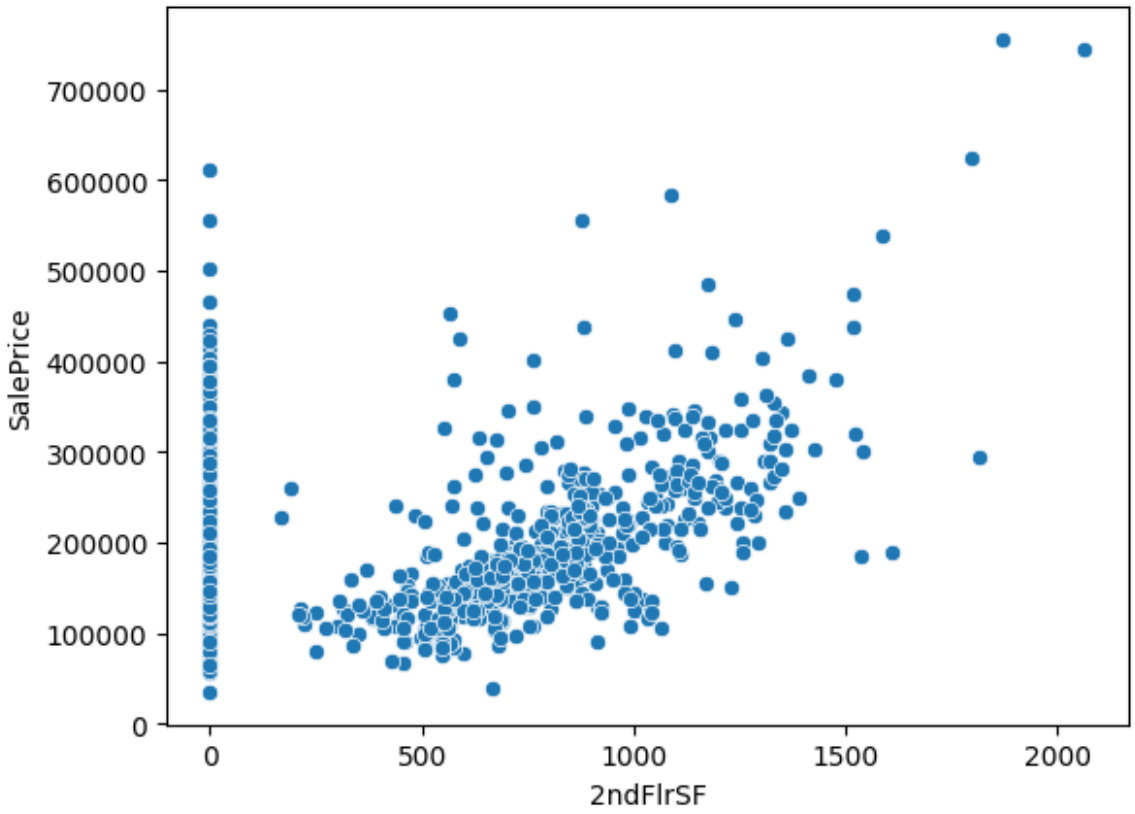

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
1000
2000
3000
4000
5000
BsmtFinSF1

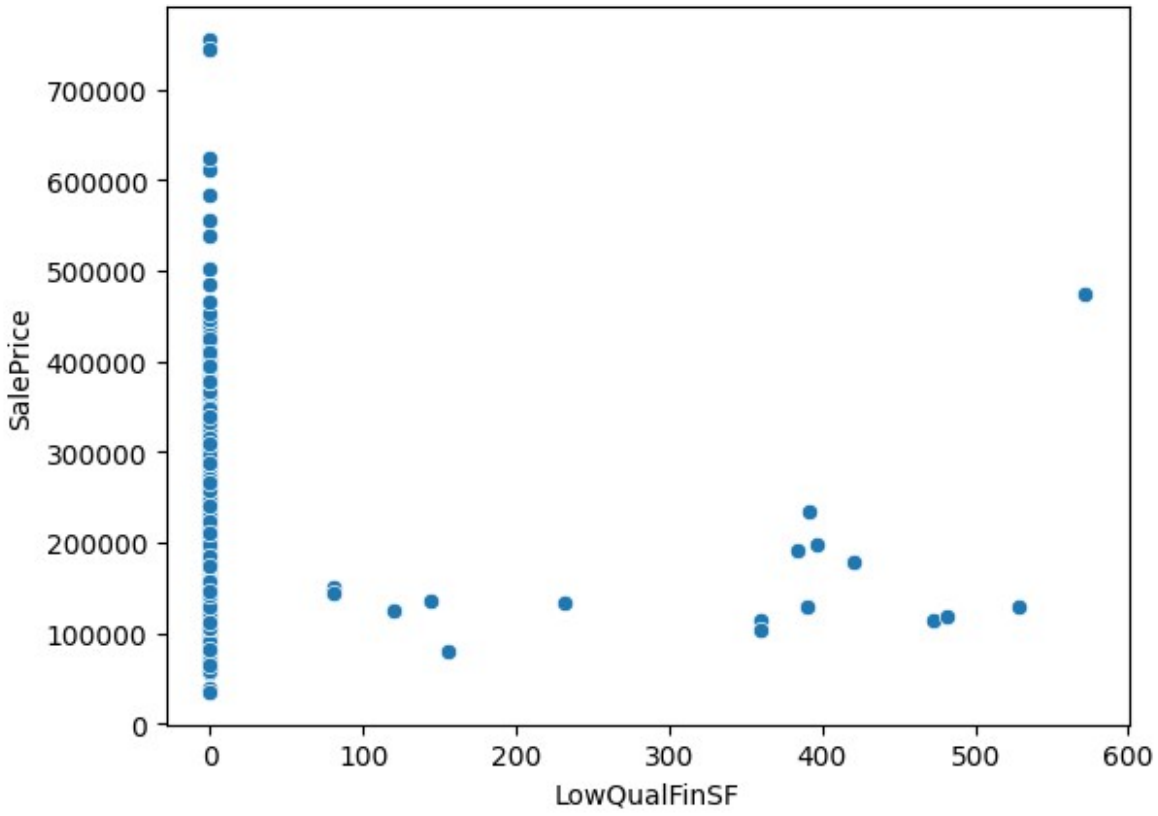

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
200
400
600
800
1000
1200
1400
BsmtFinSF2

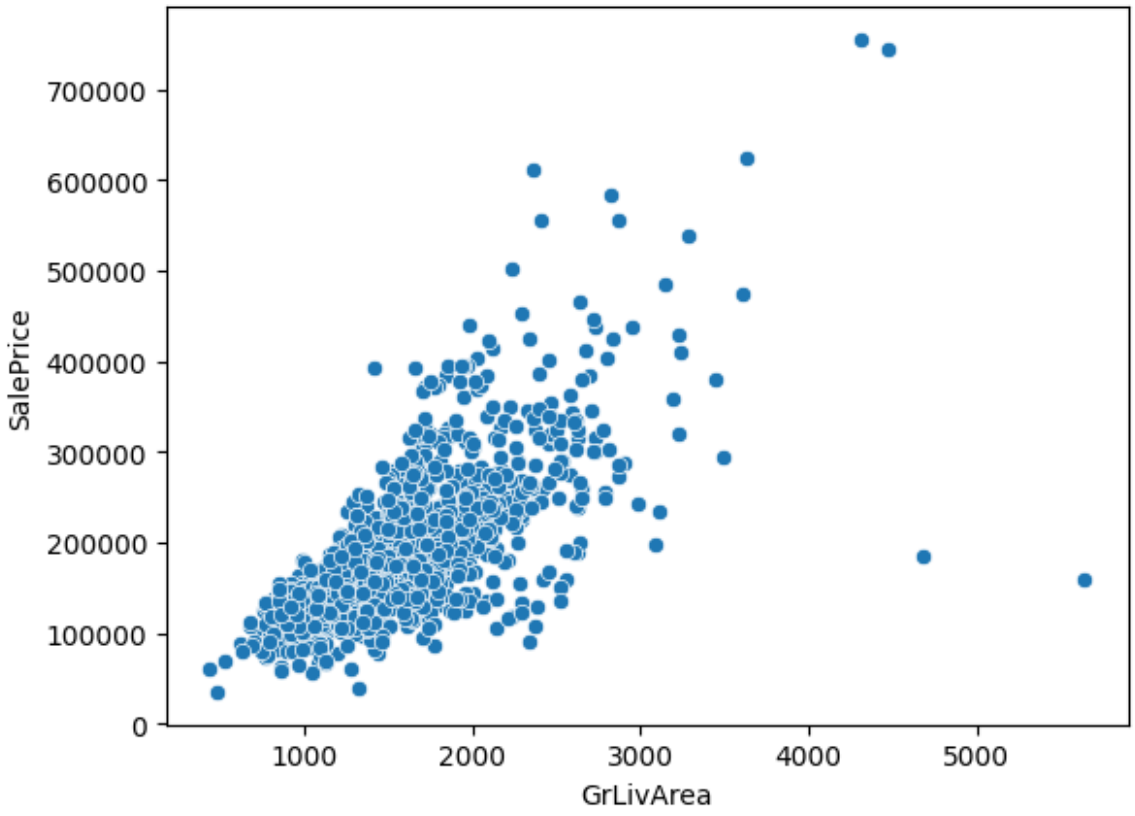

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
500
1000
1500
2000
BsmtUnfSF

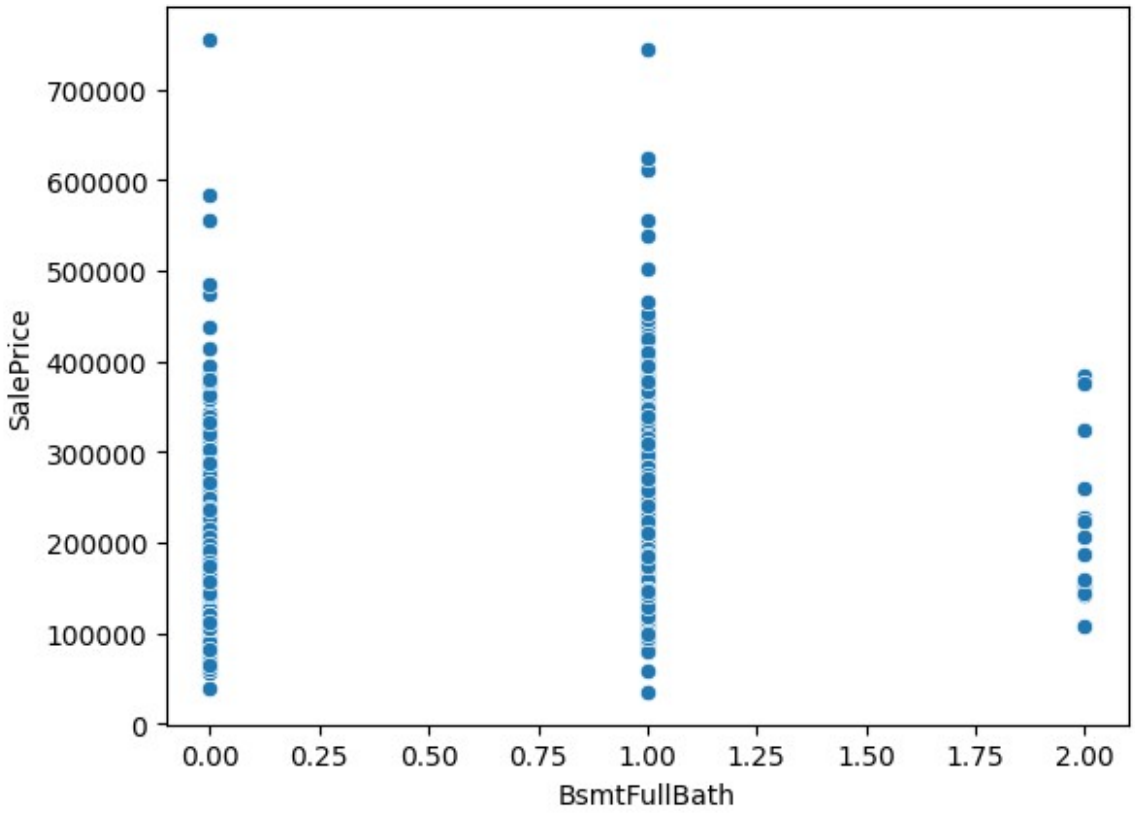

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
1000
2000
3000
4000
5000
6000
TotalBsmtSF

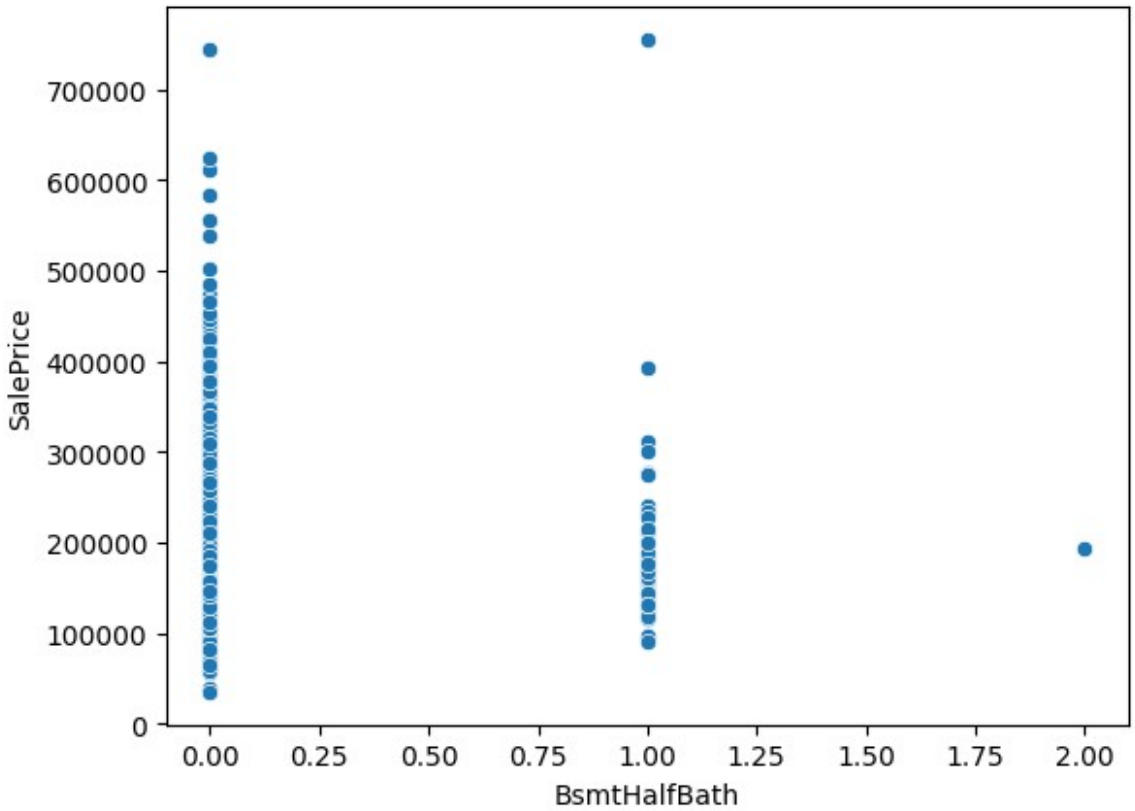

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
1000
2000
3000
4000
1stFlrSF

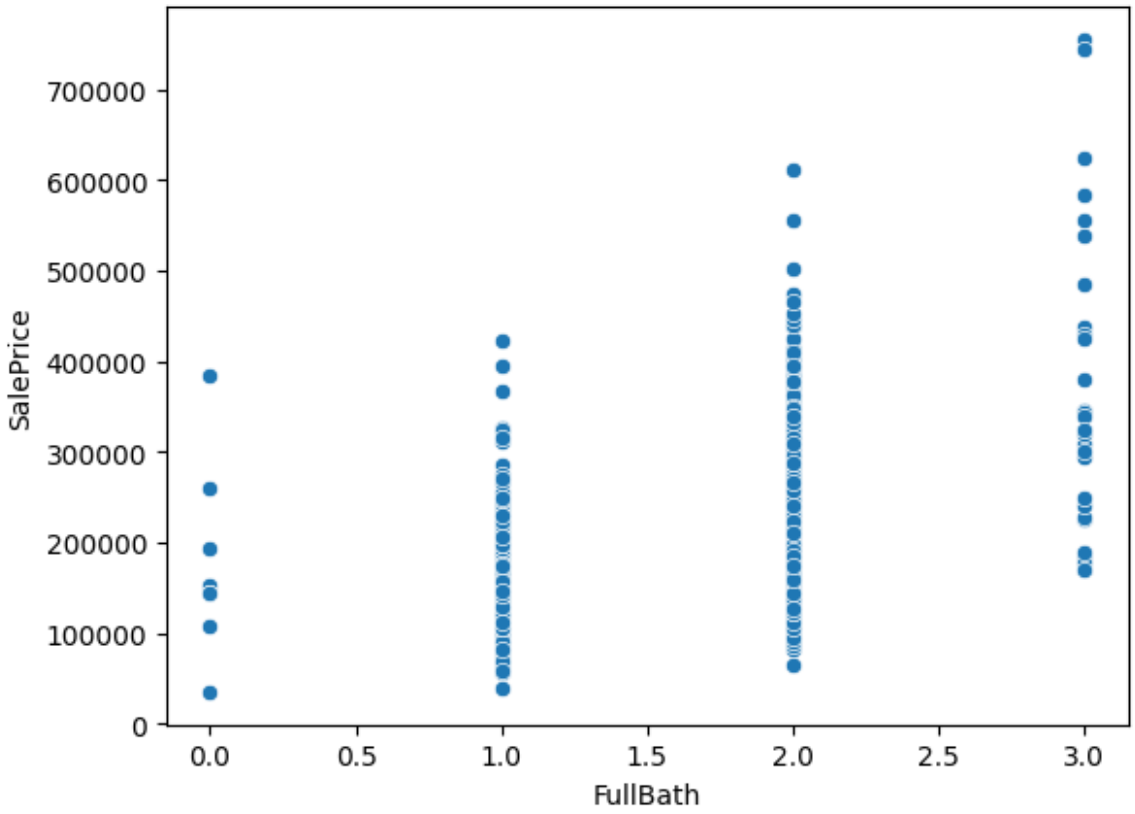

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
500
1000
1500
2000
2ndFlrSF

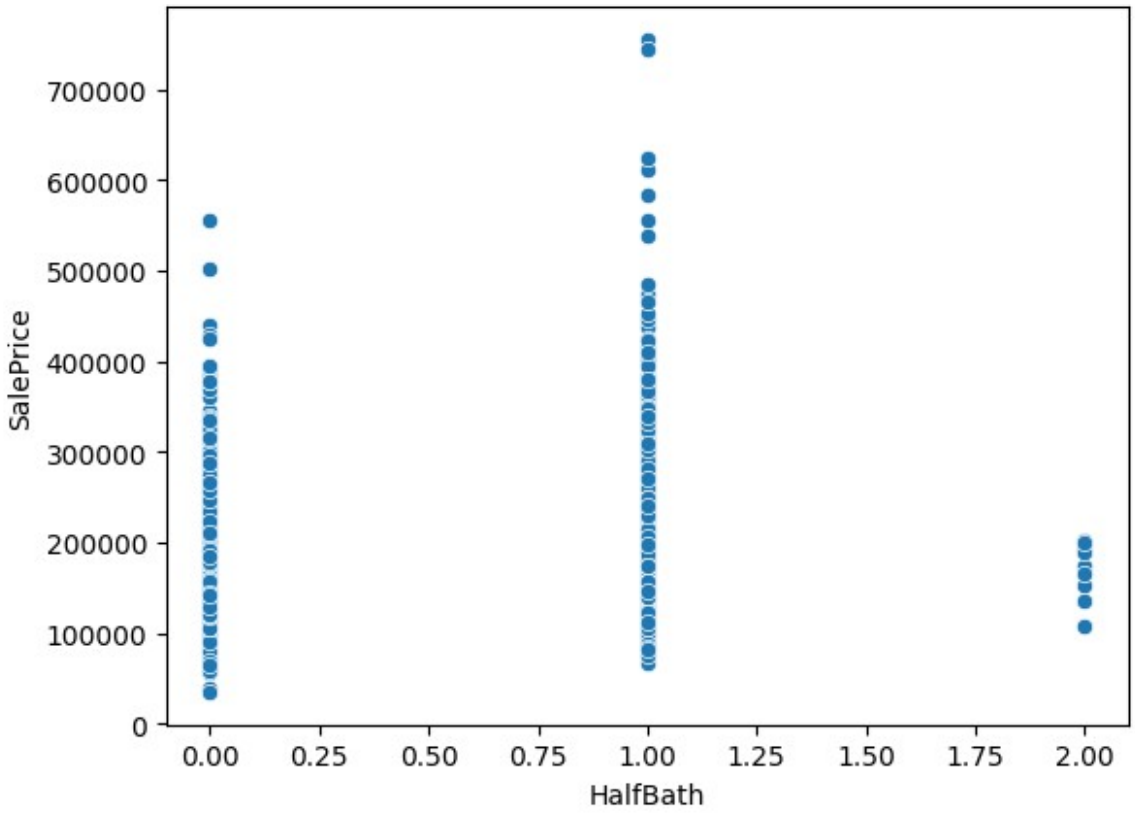

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
100
200
300
400
500
600
LowQualFinSF

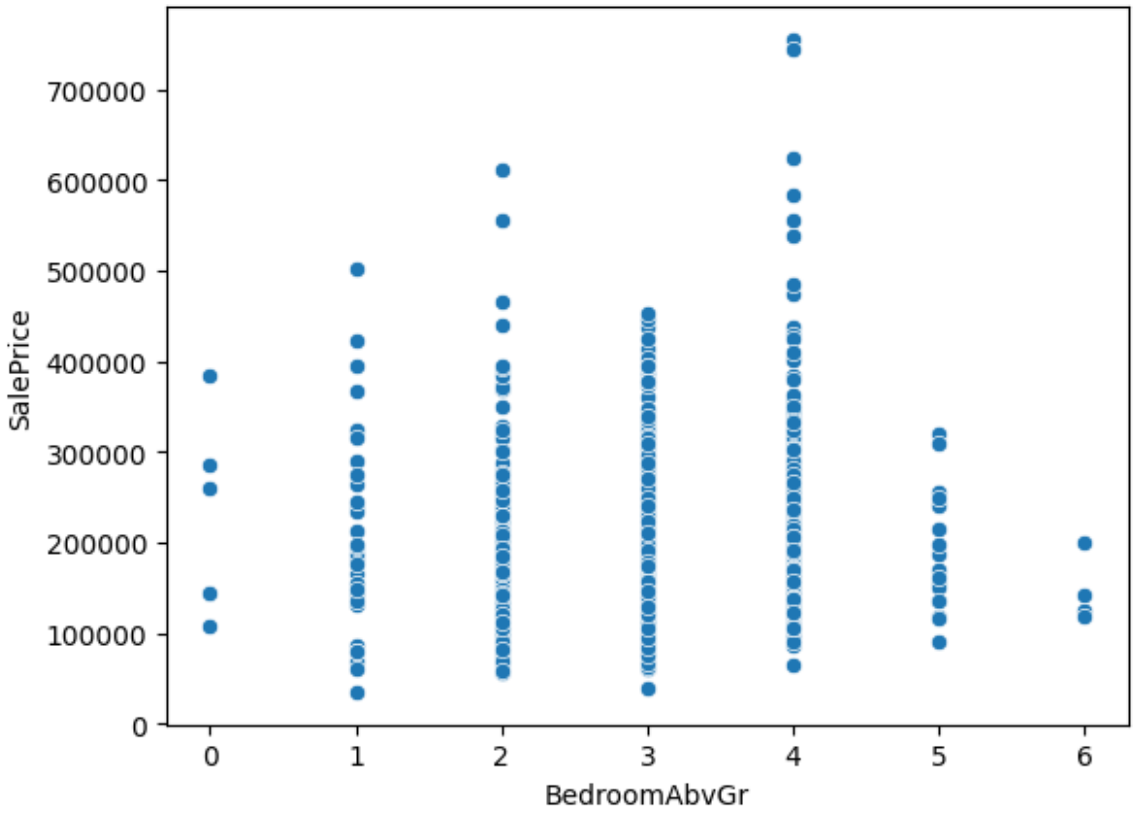

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
1000
2000
3000
4000
5000
GrLivArea

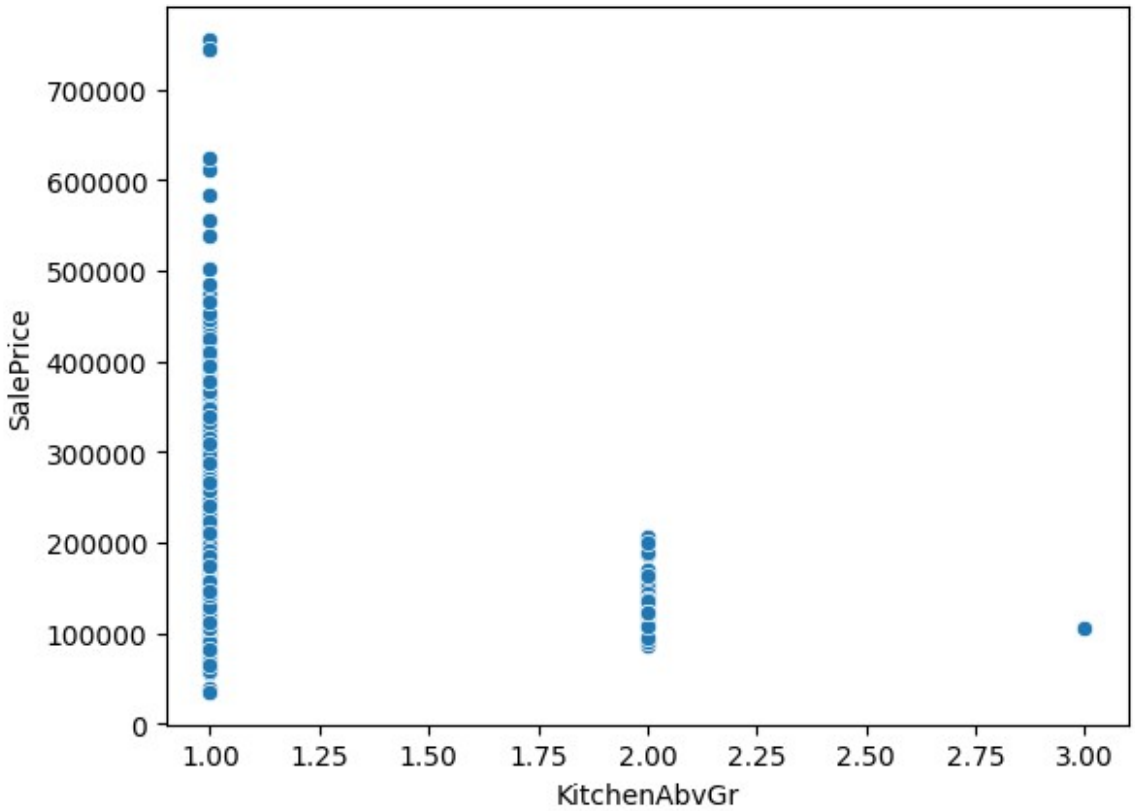

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
BsmtFullBath

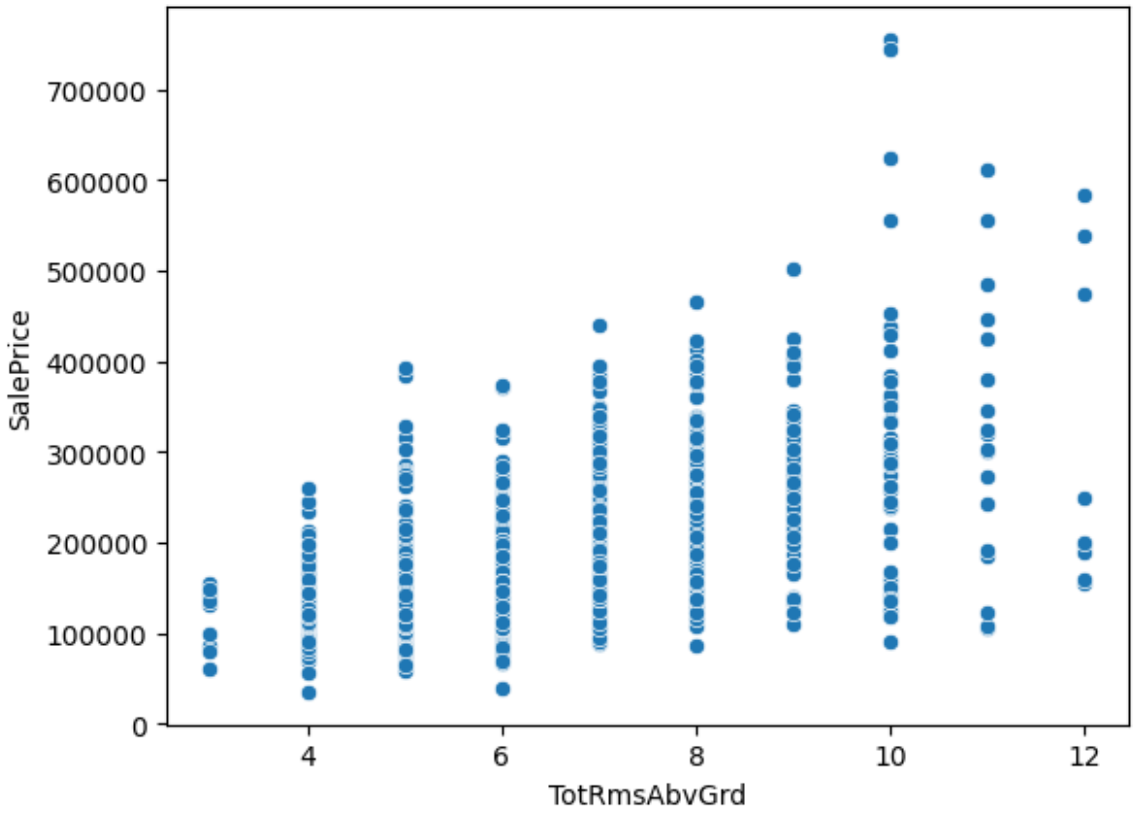

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
BsmtHalfBath

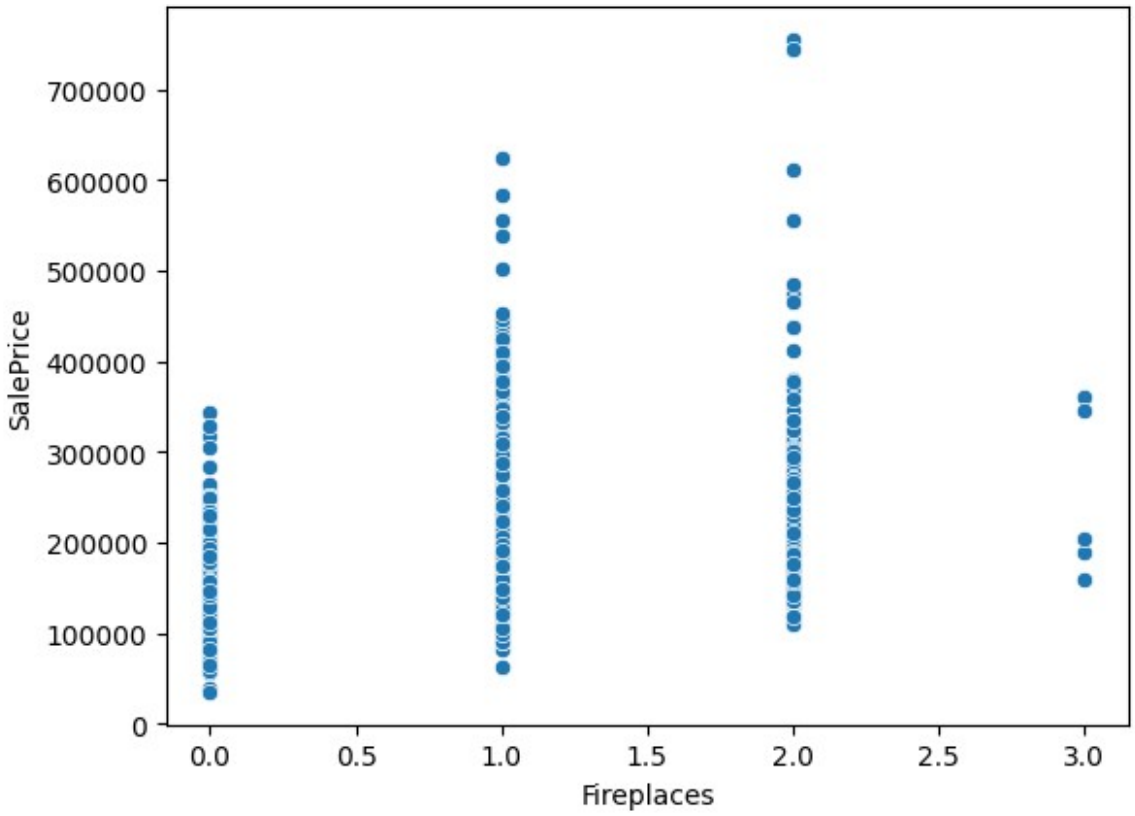

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
FullBath

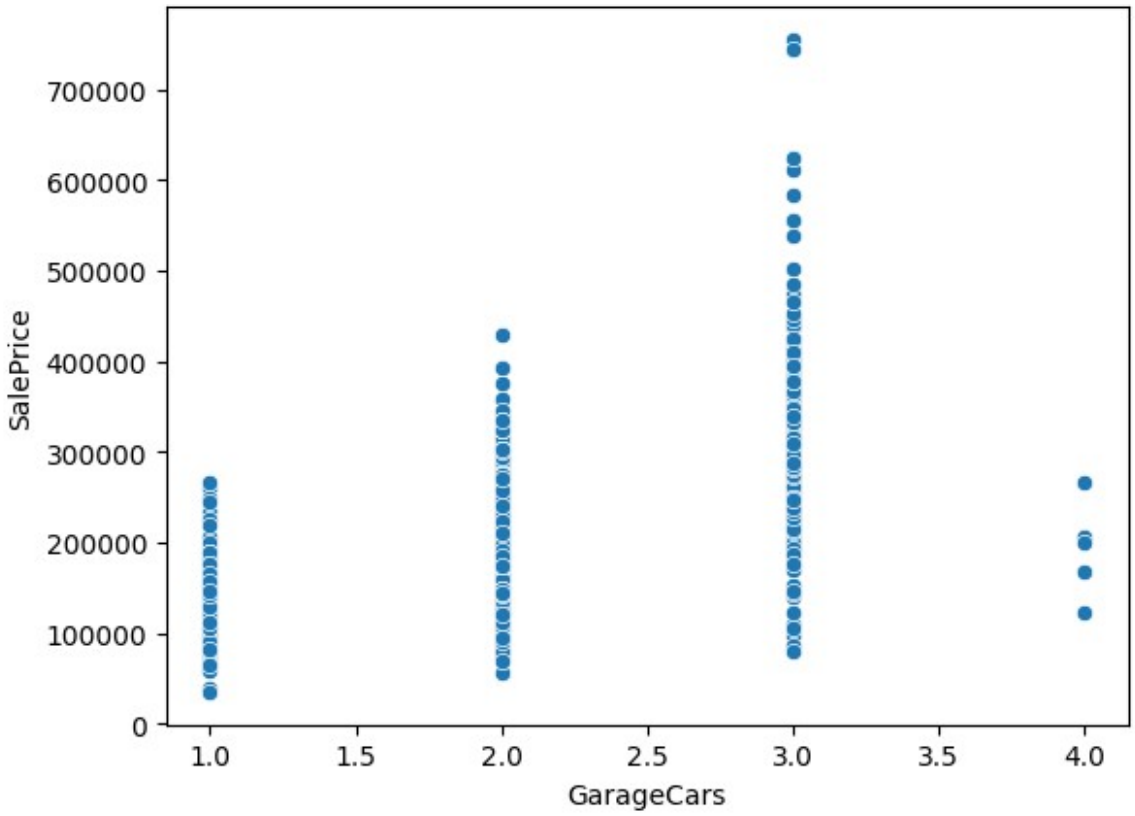

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
HalfBath

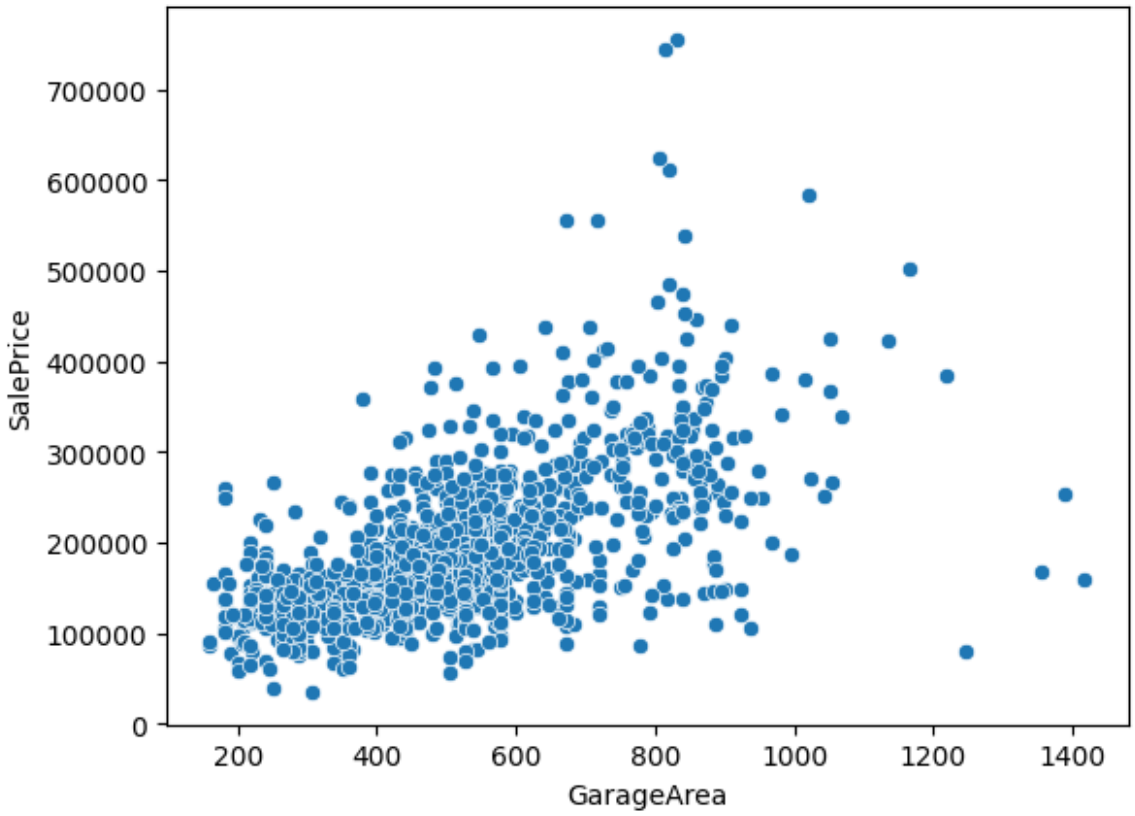

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
1
2
3
4
5
6
BedroomAbvGr

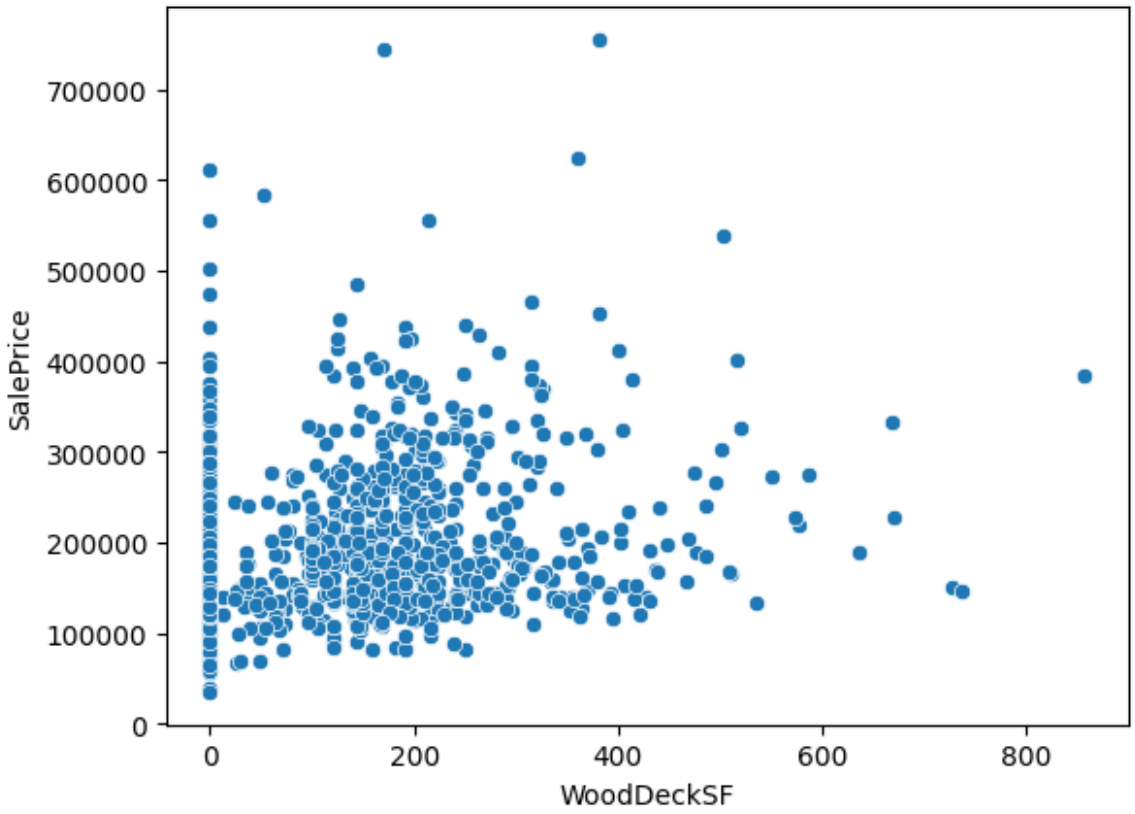

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
KitchenAbvGr

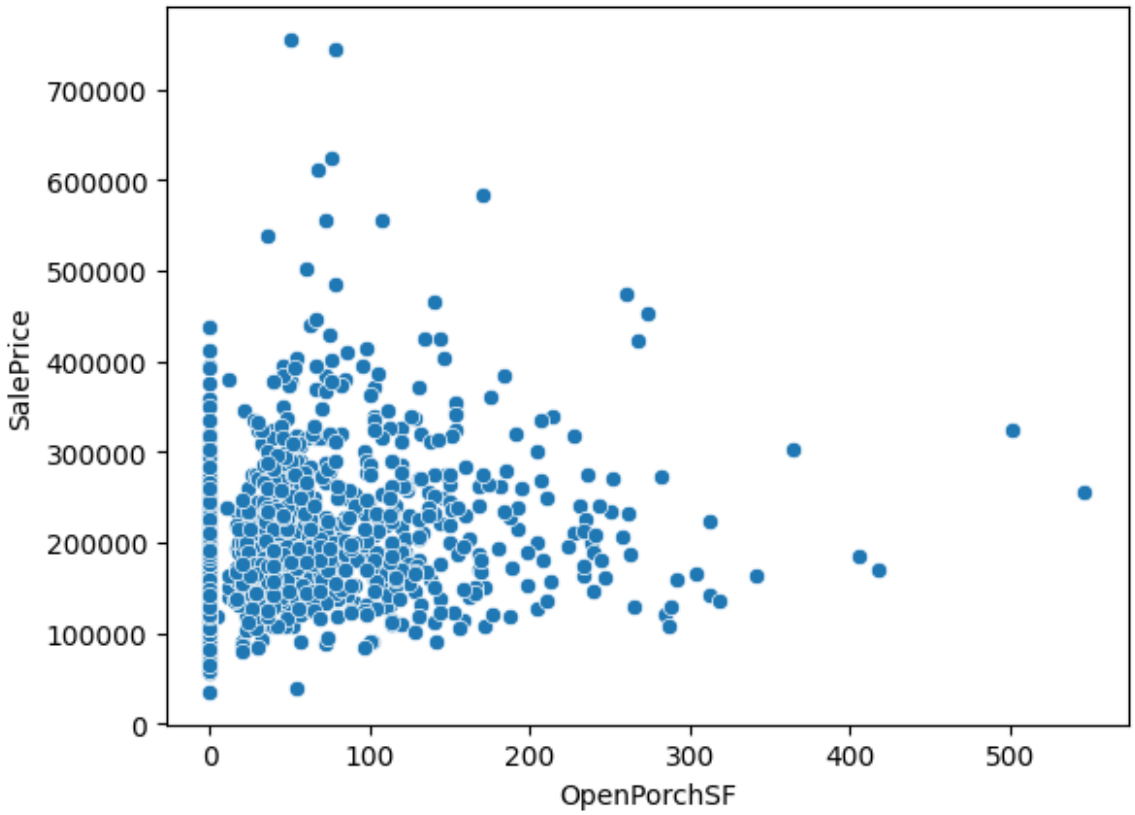

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
4
6
8
10
12
TotRmsAbvGrd

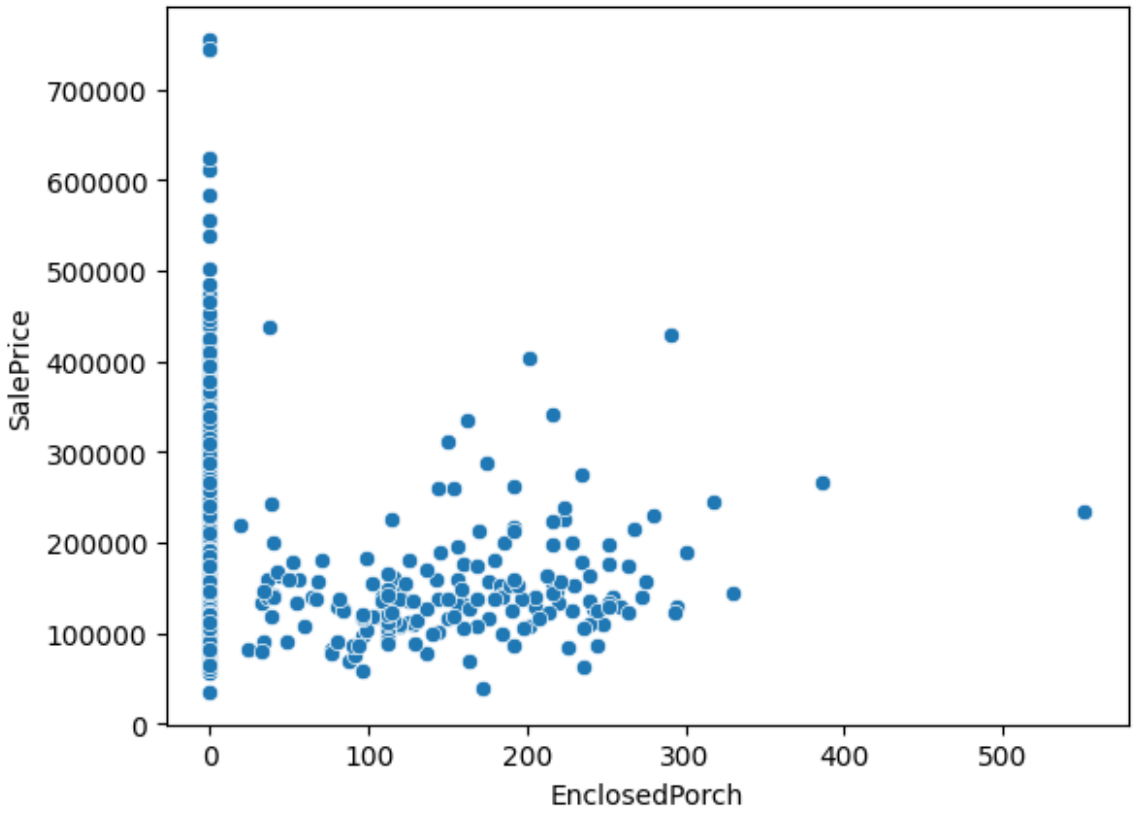

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Fireplaces

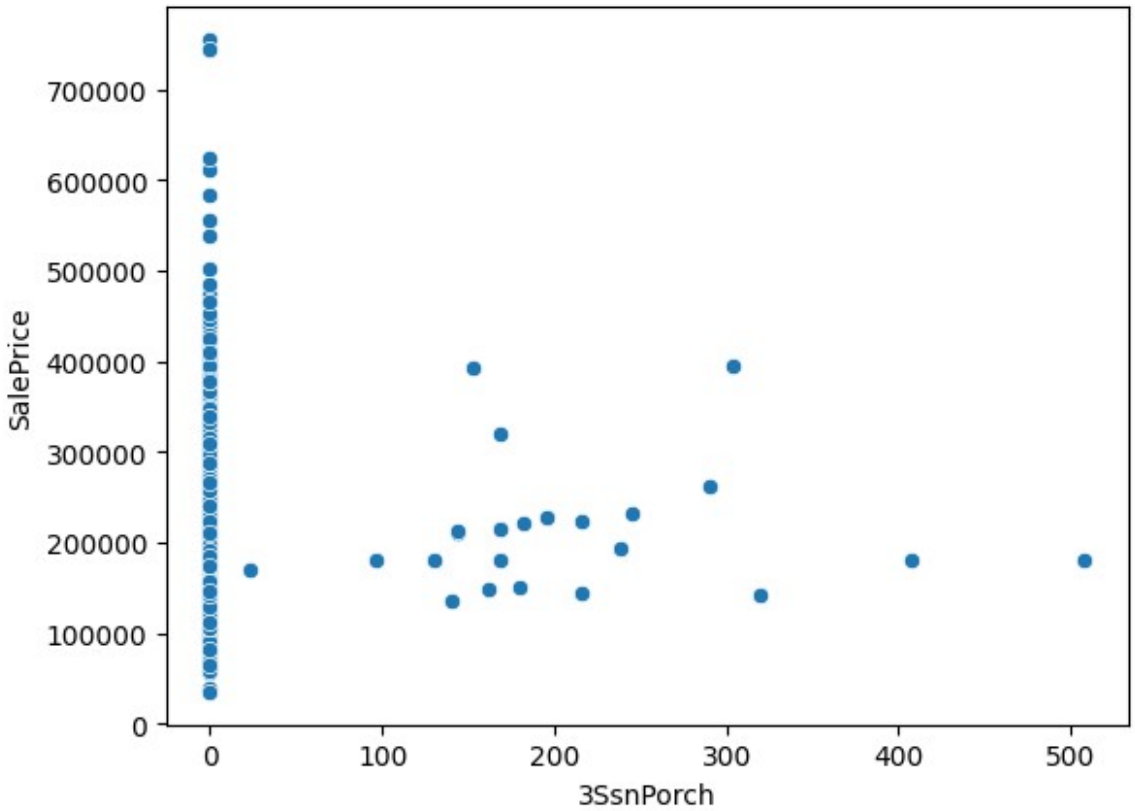

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
GarageCars

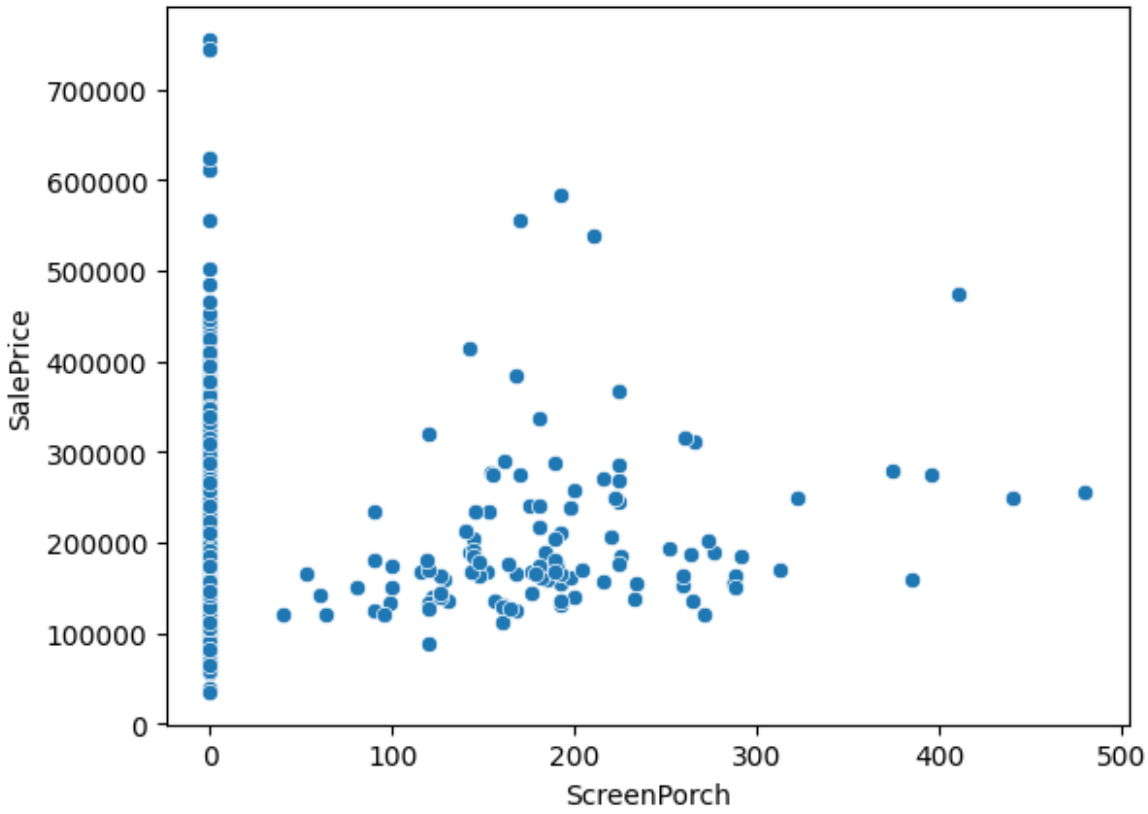

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
200
400
600
800
1000
1200
1400
GarageArea

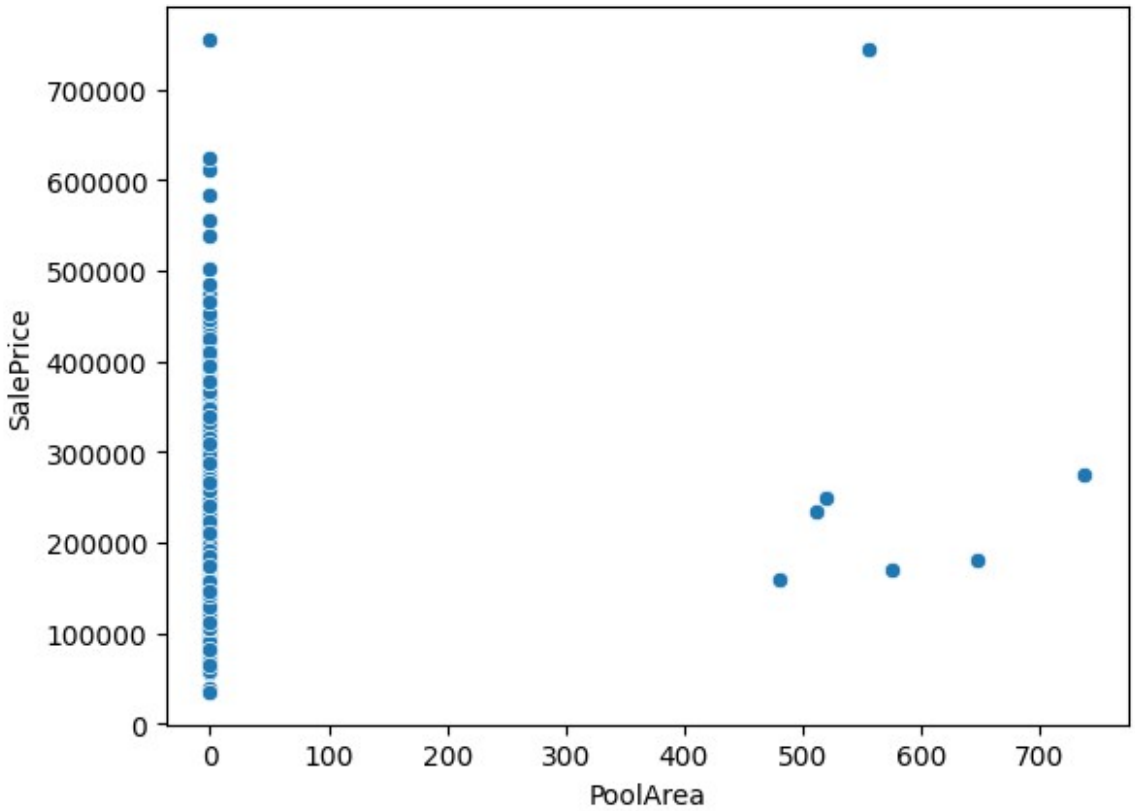

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
200
400
600
800
WoodDeckSF

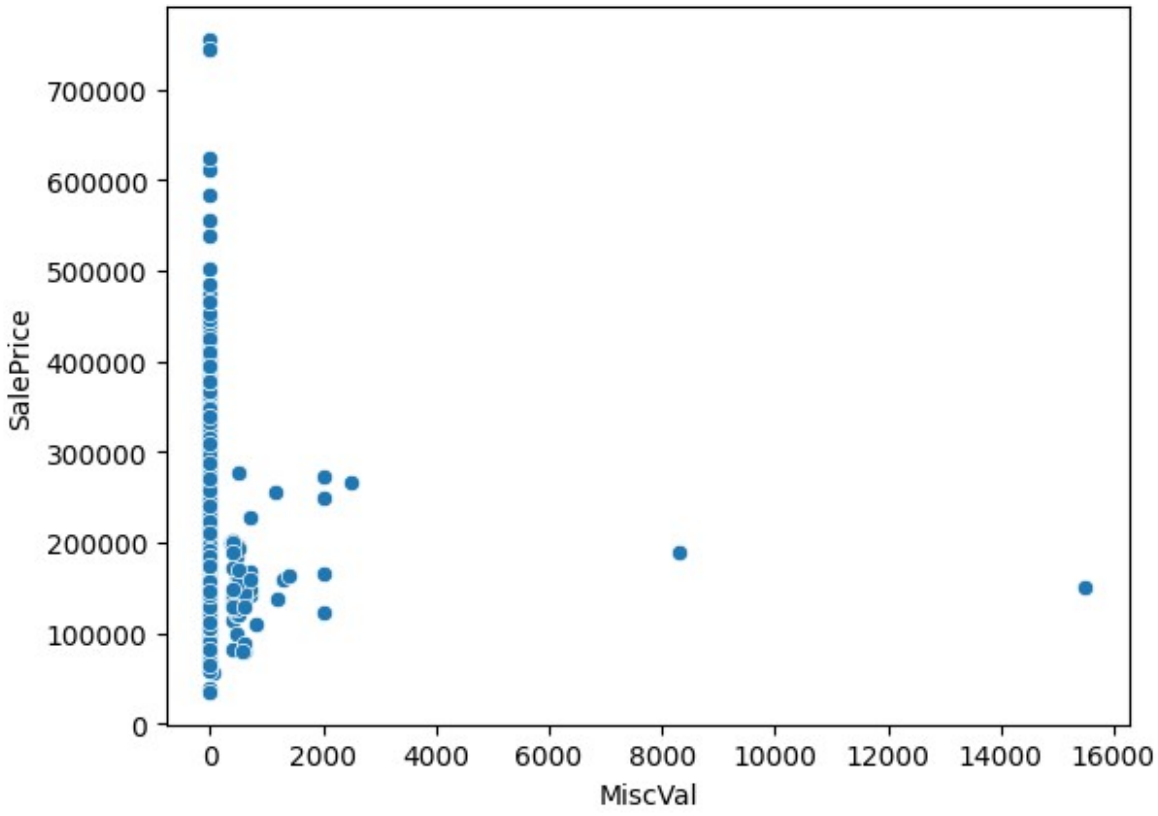

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
100
200
300
400
500
OpenPorchSF

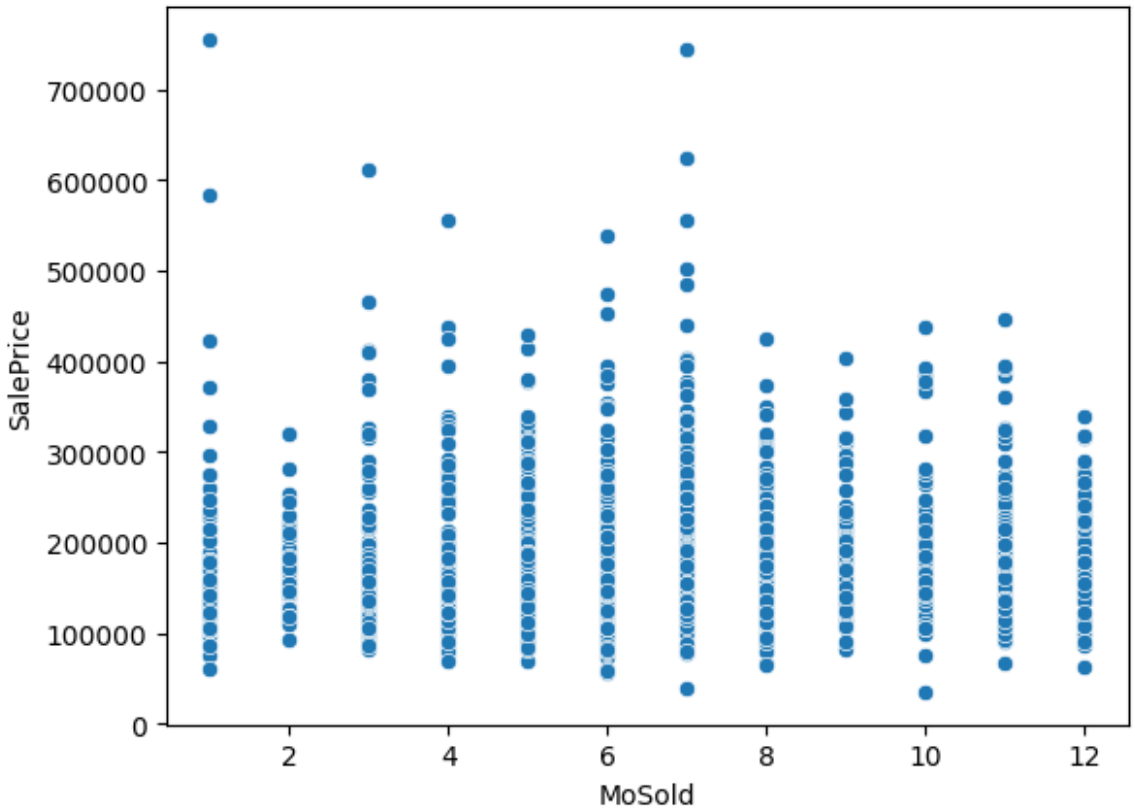

SalePrice
EnclosedPorch
0
100000
200000
300000
400000
500000
600000
700000
0
100
200
300
400
500

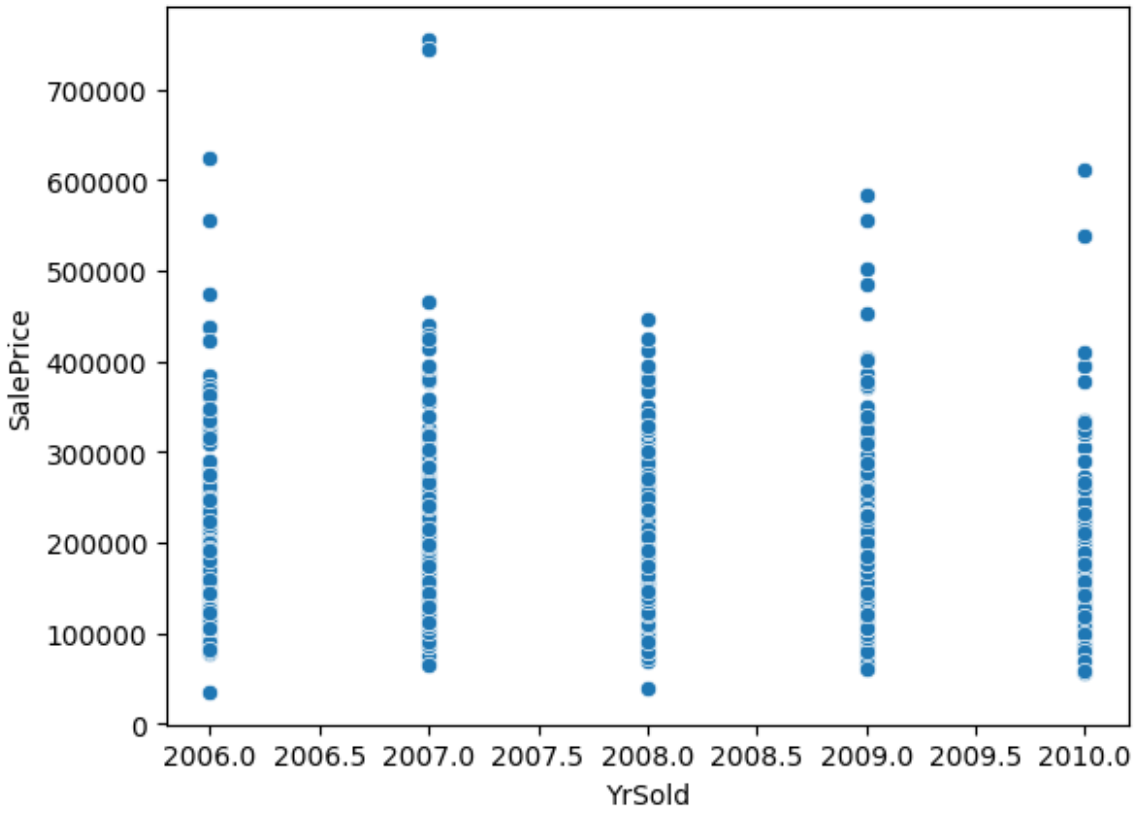

SalePrice
3SsnPorch
0
100000
200000
300000
400000
500000
600000
700000
0
100
200
300
400
500

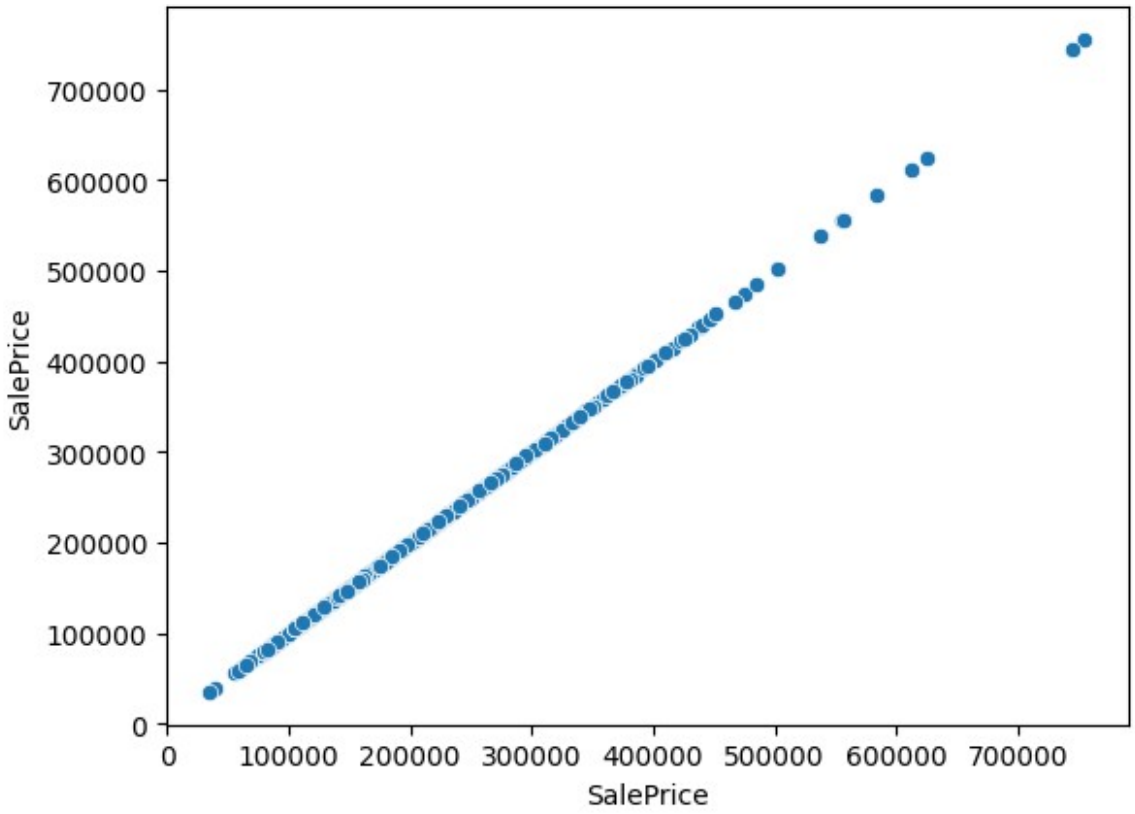

SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
100
200
300
400
500
ScreenPorch


SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
100
200
300
400
500
600
700
PoolArea


SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
2000
4000
6000
8000
10000
12000
14000
16000
MiscVal


SalePrice
700000
600000
500000
400000
300000
200000
100000
0
2
4
6
8
10
12
MoSold


SalePrice
700000
600000
500000
400000
300000
200000
100000
0
2006.0
2006.5
2007.0
2007.5
2008.0
2008.5
2009.0
2009.5
2010.0
YrSold


SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
100000
200000
300000
400000
500000
600000
700000
SalePrice

Outlier Detection

```
df = df[df['LotArea']<150000]
df = df[df['BsmtFinSF1']<3000]
df = df[df['TotalBsmtSF']<4000]
df = df[df['1stFlrSF']<3500]
df = df[df['BsmtHalfBath']<1.25]
df = df[df['KitchenAbvGr']<2.25]
df = df[df['EnclosedPorch']<400]
df = df[df['MiscVal']<4000]

df.shape
```

```
(1329, 72)
```

```
df.head()
```

```
{"type":"dataframe","variable_name":"df"}
```

Seperate Input & output

```
x = df.iloc[:,:-1]
y = df.iloc[:,-1]

x
```

```
{"type":"dataframe","variable_name":"x"}
```

Train_Test_Split

```
from sklearn.model_selection import train_test_split
x_train , x_test , y_train ,y_test =
train_test_split(x,y,test_size=0.2,random_state=42)
```

Seperation on basis of numeric & categorical

```
num_columns = x.select_dtypes(include=['int64', 'float64']).columns
cat_columns = x.select_dtypes(include=['object', 'category']).columns

cat_columns
```

```
Index(['MSZoning', 'LotShape', 'LandContour', 'Utilities',
'LotConfig',
       'LandSlope', 'Neighborhood', 'Condition1', 'BldgType',
'HouseStyle',
       'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType',
'ExterQual',
       'ExterCond', 'Foundation', 'BsmtQual', 'BsmtCond',
'BsmtExposure',
       'BsmtFinType1', 'BsmtFinType2', 'HeatingQC', 'CentralAir',
```

```
'Electrical',
       'KitchenQual', 'Functional', 'FireplaceQu', 'GarageType',
       'GarageFinish', 'GarageQual', 'GarageCond', 'PavedDrive',
'SaleType',
       'SaleCondition'],
      dtype='object')
```

Import Library

```
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
```

Column_Transformer

```
preprocessor = ColumnTransformer(
    transformers=[
        ('cat_data' ,
OneHotEncoder(handle_unknown='ignore',drop='first') , cat_columns),
        ("standardization" , StandardScaler() , num_columns)

    ],remainder='drop'
)
```

Fit transformer to input data

```
x_train_processed = preprocessor.fit_transform(x_train)
x_test_processed = preprocessor.transform(x_test)
```

```
/usr/local/lib/python3.12/dist-packages/sklearn/preprocessing/
_encoders.py:246: UserWarning: Found unknown categories in columns
[10, 11, 15, 34] during transform. These unknown categories will be
encoded as all zeros
  warnings.warn(
```

```
x_test_processed
```

```
array([[ 0.        ,  0.        ,  1.        , ..., -0.15275737,
        -0.52166752,  0.14062912],
       [ 0.        ,  0.        ,  1.        , ..., -0.15275737,
        -1.63295524, -0.61057067],
       [ 0.        ,  0.        ,  1.        , ..., -0.15275737,
        -0.15123828, -0.61057067],
       ...,
       [ 0.        ,  0.        ,  1.        , ..., -0.15275737,
         1.70090792,  0.89182892],
```

```
       [ 0.        ,  0.        ,  1.        , ..., -0.15275737,
        -0.89209676, -0.61057067],
       [ 0.        ,  0.        ,  1.        , ..., -0.15275737,
        -0.15123828, -1.36177046]])
```

Linear Regression

```
model = LinearRegression()
model.fit(X_train_processed,y_train)
```

```
LinearRegression()
```

```
y_pred = model.predict(x_test_processed)
```

R2 Score for Linear Regression(0.8227149376218911)

```
print(r2_score(y_pred,y_test))
```

```
0.8227149376218911
```

Ridge Regression by using GridSearchCV

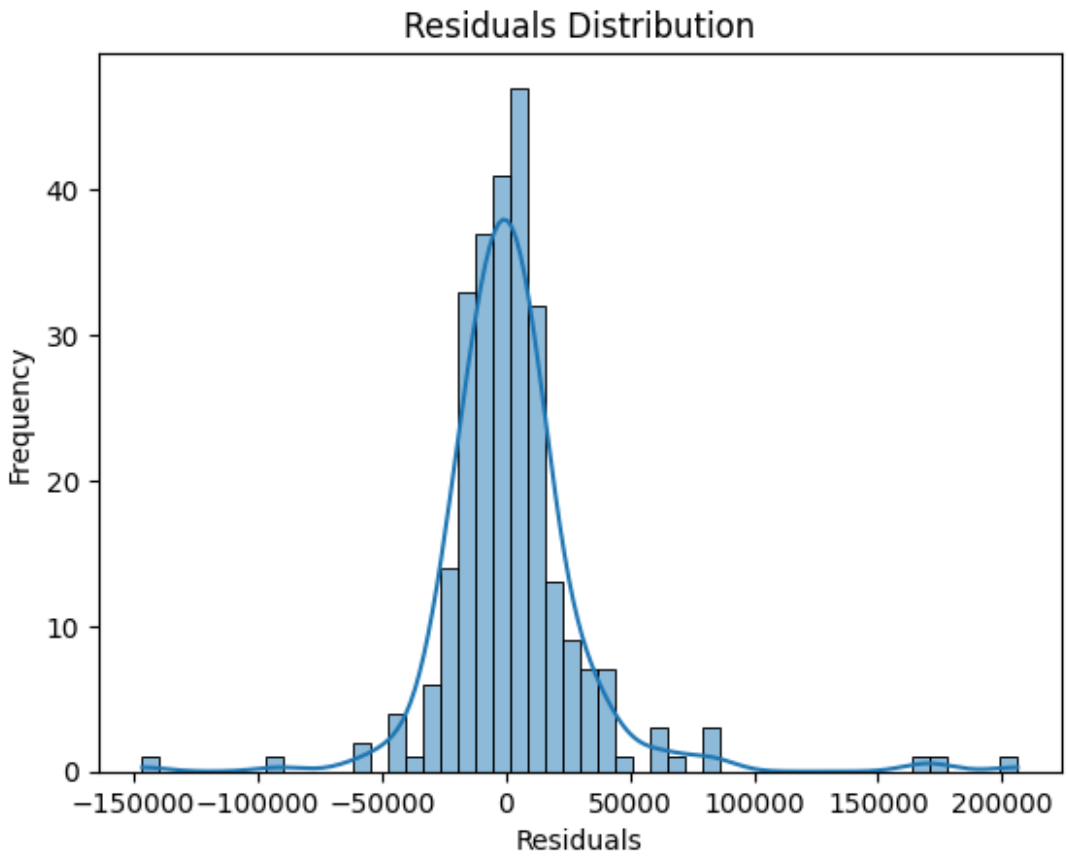
```
from sklearn.linear_model import Ridge
from sklearn.model_selection import GridSearchCV

ridge = Ridge()
parameters = {'alpha':[1e-15,1e-10,1e-8,1e-3,1e-
2,1,5,10,20,30,35,40,45,50,55,100]}
ridge_reg = GridSearchCV(ridge , parameters)
ridge_reg.fit(x_train_processed,y_train)
```

```
/usr/local/lib/python3.12/dist-packages/scipy/_lib/_util.py:1233:
LinAlgWarning: Ill-conditioned matrix (rcond=5.05792e-20): result may
not be accurate.
  return f(*arrays, *other_args, **kwargs)
/usr/local/lib/python3.12/dist-packages/scipy/_lib/_util.py:1233:
LinAlgWarning: Ill-conditioned matrix (rcond=6.32293e-20): result may
not be accurate.
  return f(*arrays, *other_args, **kwargs)

GridSearchCV(estimator=Ridge(),
             param_grid={'alpha': [1e-15, 1e-10, 1e-08, 0.001, 0.01,
```

```
1, 5, 10,
                                      20, 30, 35, 40, 45, 50, 55, 100]})

print(ridge_reg.best_params_)
print(ridge_reg.best_score_)

{'alpha': 10}
0.8628454120885992

print(r2_score(y_test,ridge_reg.predict(x_test_processed)))

0.8800492958111971
```

R2 Score for Ridge Regression(0.8800492958111971)** **

```
residuals = y_test - ridge_reg.predict(x_test_processed)

sns.histplot(residuals, kde=True)  # histogram + KDE
plt.xlabel("Residuals")
plt.ylabel("Frequency")
plt.title("Residuals Distribution")
plt.show()
```



Lasso Regression by using GridSearchCV

```
from sklearn.linear_model import Lasso

parameters = {'alpha':[1e-15,1e-10,1e-8,1e-3,1e-
2,1,5,10,20,30,35,40,45,50,55,100]}
lasso = Lasso()
lasso_reg = GridSearchCV(lasso,parameters)
lasso_reg.fit(x_train_processed,y_train)
```

```
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/
_coordinate_descent.py:695: ConvergenceWarning: Objective did not
converge. You might want to increase the number of iterations, check
the scale of the features or consider increasing regularisation.
Duality gap: 1.889e+11, tolerance: 5.273e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.965e+11, tolerance: 5.155e+08
  model = cd_fast.enet_coordinate_descent(
```

```
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordinate_descent.py:695: ConvergenceWarning: Objective did not converge. You might want to increase the number of iterations, check the scale of the features or consider increasing regularisation. Duality gap: 1.706e+11, tolerance: 4.808e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordinate_descent.py:695: ConvergenceWarning: Objective did not converge. You might want to increase the number of iterations, check the scale of the features or consider increasing regularisation. Duality gap: 1.215e+11, tolerance: 4.884e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordinate_descent.py:695: ConvergenceWarning: Objective did not converge. You might want to increase the number of iterations, check the scale of the features or consider increasing regularisation. Duality gap: 1.893e+11, tolerance: 4.735e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordinate_descent.py:695: ConvergenceWarning: Objective did not converge. You might want to increase the number of iterations, check the scale of the features or consider increasing regularisation. Duality gap: 1.889e+11, tolerance: 5.273e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordinate_descent.py:695: ConvergenceWarning: Objective did not converge. You might want to increase the number of iterations, check the scale of the features or consider increasing regularisation. Duality gap: 1.965e+11, tolerance: 5.155e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordinate_descent.py:695: ConvergenceWarning: Objective did not converge. You might want to increase the number of iterations, check the scale of the features or consider increasing regularisation. Duality gap: 1.706e+11, tolerance: 4.808e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordinate_descent.py:695: ConvergenceWarning: Objective did not converge. You might want to increase the number of iterations, check the scale of the features or consider increasing regularisation. Duality gap: 1.215e+11, tolerance: 4.884e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordinate_descent.py:695: ConvergenceWarning: Objective did not converge. You might want to increase the number of iterations, check the scale of the features or consider increasing regularisation. Duality gap: 1.893e+11, tolerance: 4.735e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
```

```
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.889e+11, tolerance: 5.273e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.965e+11, tolerance: 5.155e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.706e+11, tolerance: 4.808e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.215e+11, tolerance: 4.884e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.893e+11, tolerance: 4.735e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.882e+11, tolerance: 5.273e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.962e+11, tolerance: 5.155e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.700e+11, tolerance: 4.808e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
```

```
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.209e+11, tolerance: 4.884e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.883e+11, tolerance: 4.735e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.820e+11, tolerance: 5.273e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.928e+11, tolerance: 5.155e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.651e+11, tolerance: 4.808e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.156e+11, tolerance: 4.884e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.795e+11, tolerance: 4.735e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
8.998e+09, tolerance: 5.273e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
```

```
the features or consider increasing regularisation. Duality gap:
8.956e+09, tolerance: 5.155e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
6.532e+09, tolerance: 4.808e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
6.574e+09, tolerance: 4.884e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.095e+09, tolerance: 4.735e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
5.552e+08, tolerance: 4.884e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
8.342e+08, tolerance: 4.735e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
9.859e+08, tolerance: 4.735e+08
  model = cd_fast.enet_coordinate_descent(
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_coordina
te_descent.py:695: ConvergenceWarning: Objective did not converge. You
might want to increase the number of iterations, check the scale of
the features or consider increasing regularisation. Duality gap:
1.377e+09, tolerance: 4.735e+08
  model = cd_fast.enet_coordinate_descent(

GridSearchCV(estimator=Lasso(),
             param_grid={'alpha': [1e-15, 1e-10, 1e-08, 0.001, 0.01,
1, 5, 10,
                                   20, 30, 35, 40, 45, 50, 55, 100]})
```

```
print(lasso_reg.best_params_)
print(lasso_reg.best_score_)
```

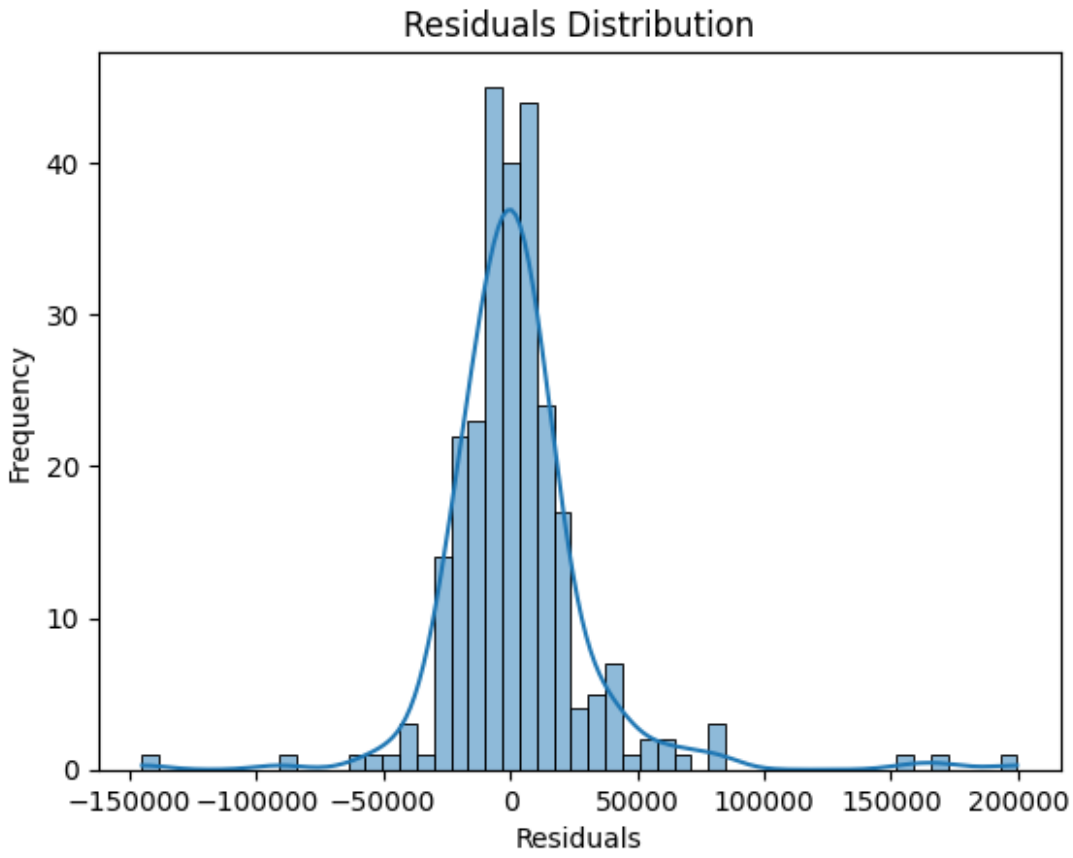
```
{'alpha': 100}
0.8611223388774485
```

```
print(r2_score(y_test , lasso_reg.predict(x_test_processed)))
```

```
0.8839337241343956
```

R2 Score For Lasso Regression is (0.8839337241343956)

```
residuals = y_test - lasso_reg.predict(x_test_processed)

sns.histplot(residuals, kde=True)  # histogram + KDE
plt.xlabel("Residuals")
plt.ylabel("Frequency")
plt.title("Residuals Distribution")
plt.show()
```



We are getting the best R2 score with Lasso Regression