

```
In [1]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
```

```
In [2]: data=pd.read_csv(r"C:\Users\DSK_8920444598\Downloads\house_price_train.csv")
```

```
In [3]: pd.set_option("display.max_columns",1000)
pd.set_option("display.max_rows",1000)
```

```
In [4]: data
```

Out[4]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	
0	1	60	RL	65.0	8450	Pave	NaN	Reg		Lvl	AllPub	Inside	Gtl	Co
1	2	20	RL	80.0	9600	Pave	NaN	Reg		Lvl	AllPub	FR2	Gtl	Vee
2	3	60	RL	68.0	11250	Pave	NaN	IR1		Lvl	AllPub	Inside	Gtl	Co
3	4	70	RL	60.0	9550	Pave	NaN	IR1		Lvl	AllPub	Corner	Gtl	Cra
4	5	60	RL	84.0	14260	Pave	NaN	IR1		Lvl	AllPub	FR2	Gtl	NoR
...
1455	1456	60	RL	62.0	7917	Pave	NaN	Reg		Lvl	AllPub	Inside	Gtl	Gi
1456	1457	20	RL	85.0	13175	Pave	NaN	Reg		Lvl	AllPub	Inside	Gtl	NWA
1457	1458	70	RL	66.0	9042	Pave	NaN	Reg		Lvl	AllPub	Inside	Gtl	Cra
1458	1459	20	RL	68.0	9717	Pave	NaN	Reg		Lvl	AllPub	Inside	Gtl	NA
1459	1460	20	RL	75.0	9937	Pave	NaN	Reg		Lvl	AllPub	Inside	Gtl	Edw

1460 rows × 81 columns

Curse of dimensionality above

```
In [5]: data.shape
```

```
Out[5]: (1460, 81)
```

```
In [6]: data.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       1201 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         1452 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
```

INSIGHTS

- Shape of Dataset is 1460x81
- Alley,PoolQC and MiscFeature have majority of null values(>90%)
- 43 columns are of object/35 columns of int/3 columns of float

descriptive stats of data

```
In [7]: data.describe(include='all').T
```

Out[7]:		count	unique	top	freq	mean	std	min	25%	50%	75%	max
	Id	1460.0	NaN	NaN	NaN	730.5	421.610009	1.0	365.75	730.5	1095.25	1460.0
	MSSubClass	1460.0	NaN	NaN	NaN	56.89726	42.300571	20.0	20.0	50.0	70.0	190.0
	MSZoning	1460	5	RL	1151	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	LotFrontage	1201.0	NaN	NaN	NaN	70.049958	24.284752	21.0	59.0	69.0	80.0	313.0
	LotArea	1460.0	NaN	NaN	NaN	10516.828082	9981.264932	1300.0	7553.5	9478.5	11601.5	215245.0
	Street	1460	2	Pave	1454	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Alley	91	2	Grvl	50	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	LotShape	1460	4	Reg	925	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	LandContour	1460	4	Lvl	1311	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Utilities	1460	2	AllPub	1459	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	LotConfig	1460	5	Inside	1052	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	LandSlope	1460	3	Gtl	1382	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Neighborhood	1460	25	NAmes	225	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Condition1	1460	9	Norm	1260	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Condition2	1460	8	Norm	1445	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	BldgType	1460	5	1Fam	1220	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	HouseStyle	1460	8	1Story	726	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	OverallQual	1460.0	NaN	NaN	NaN	6.099315	1.382997	1.0	5.0	6.0	7.0	10.0
	OverallCond	1460.0	NaN	NaN	NaN	5.575342	1.112799	1.0	5.0	5.0	6.0	9.0
	YearBuilt	1460.0	NaN	NaN	NaN	1971.267808	30.202904	1872.0	1954.0	1973.0	2000.0	2010.0
	YearRemodAdd	1460.0	NaN	NaN	NaN	1984.865753	20.645407	1950.0	1967.0	1994.0	2004.0	2010.0
	RoofStyle	1460	6	Gable	1141	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	RoofMatl	1460	8	CompShg	1434	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Exterior1st	1460	15	VinylSd	515	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	Exterior2nd	1460	16	VinylSd	504	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
MasVnrType	1452	4	None	864	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MasVnrArea	1452.0	NaN	NaN	NaN	103.685262	181.066207	0.0	0.0	0.0	166.0	1600.0
ExterQual	1460	4	TA	906	NaN	NaN	NaN	NaN	NaN	NaN	NaN
ExterCond	1460	5	TA	1282	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Foundation	1460	6	PConc	647	NaN	NaN	NaN	NaN	NaN	NaN	NaN
BsmtQual	1423	4	TA	649	NaN	NaN	NaN	NaN	NaN	NaN	NaN
BsmtCond	1423	4	TA	1311	NaN	NaN	NaN	NaN	NaN	NaN	NaN
BsmtExposure	1422	4	No	953	NaN	NaN	NaN	NaN	NaN	NaN	NaN
BsmtFinType1	1423	6	Unf	430	NaN	NaN	NaN	NaN	NaN	NaN	NaN
BsmtFinSF1	1460.0	NaN	NaN	NaN	443.639726	456.098091	0.0	0.0	383.5	712.25	5644.0
BsmtFinType2	1422	6	Unf	1256	NaN	NaN	NaN	NaN	NaN	NaN	NaN
BsmtFinSF2	1460.0	NaN	NaN	NaN	46.549315	161.319273	0.0	0.0	0.0	0.0	1474.0
BsmtUnfSF	1460.0	NaN	NaN	NaN	567.240411	441.866955	0.0	223.0	477.5	808.0	2336.0
TotalBsmtSF	1460.0	NaN	NaN	NaN	1057.429452	438.705324	0.0	795.75	991.5	1298.25	6110.0
Heating	1460	6	GasA	1428	NaN	NaN	NaN	NaN	NaN	NaN	NaN
HeatingQC	1460	5	Ex	741	NaN	NaN	NaN	NaN	NaN	NaN	NaN
CentralAir	1460	2	Y	1365	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Electrical	1459	5	SBrkr	1334	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1stFlrSF	1460.0	NaN	NaN	NaN	1162.626712	386.587738	334.0	882.0	1087.0	1391.25	4692.0
2ndFlrSF	1460.0	NaN	NaN	NaN	346.992466	436.528436	0.0	0.0	0.0	728.0	2065.0
LowQualFinSF	1460.0	NaN	NaN	NaN	5.844521	48.623081	0.0	0.0	0.0	0.0	572.0
GrLivArea	1460.0	NaN	NaN	NaN	1515.463699	525.480383	334.0	1129.5	1464.0	1776.75	5642.0
BsmtFullBath	1460.0	NaN	NaN	NaN	0.425342	0.518911	0.0	0.0	0.0	1.0	3.0
BsmtHalfBath	1460.0	NaN	NaN	NaN	0.057534	0.238753	0.0	0.0	0.0	0.0	2.0
FullBath	1460.0	NaN	NaN	NaN	1.565068	0.550916	0.0	1.0	2.0	2.0	3.0

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
HalfBath	1460.0	NaN	NaN	NaN	0.382877	0.502885	0.0	0.0	0.0	1.0	2.0
BedroomAbvGr	1460.0	NaN	NaN	NaN	2.866438	0.815778	0.0	2.0	3.0	3.0	8.0
KitchenAbvGr	1460.0	NaN	NaN	NaN	1.046575	0.220338	0.0	1.0	1.0	1.0	3.0
KitchenQual	1460	4	TA	735	NaN	NaN	NaN	NaN	NaN	NaN	NaN
TotRmsAbvGrd	1460.0	NaN	NaN	NaN	6.517808	1.625393	2.0	5.0	6.0	7.0	14.0
Functional	1460	7	Typ	1360	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Fireplaces	1460.0	NaN	NaN	NaN	0.613014	0.644666	0.0	0.0	1.0	1.0	3.0
FireplaceQu	770	5	Gd	380	NaN	NaN	NaN	NaN	NaN	NaN	NaN
GarageType	1379	6	Attchd	870	NaN	NaN	NaN	NaN	NaN	NaN	NaN
GarageYrBlt	1379.0	NaN	NaN	NaN	1978.506164	24.689725	1900.0	1961.0	1980.0	2002.0	2010.0
GarageFinish	1379	3	Unf	605	NaN	NaN	NaN	NaN	NaN	NaN	NaN
GarageCars	1460.0	NaN	NaN	NaN	1.767123	0.747315	0.0	1.0	2.0	2.0	4.0
GarageArea	1460.0	NaN	NaN	NaN	472.980137	213.804841	0.0	334.5	480.0	576.0	1418.0
GarageQual	1379	5	TA	1311	NaN	NaN	NaN	NaN	NaN	NaN	NaN
GarageCond	1379	5	TA	1326	NaN	NaN	NaN	NaN	NaN	NaN	NaN
PavedDrive	1460	3	Y	1340	NaN	NaN	NaN	NaN	NaN	NaN	NaN
WoodDeckSF	1460.0	NaN	NaN	NaN	94.244521	125.338794	0.0	0.0	0.0	168.0	857.0
OpenPorchSF	1460.0	NaN	NaN	NaN	46.660274	66.256028	0.0	0.0	25.0	68.0	547.0
EnclosedPorch	1460.0	NaN	NaN	NaN	21.95411	61.119149	0.0	0.0	0.0	0.0	552.0
3SsnPorch	1460.0	NaN	NaN	NaN	3.409589	29.317331	0.0	0.0	0.0	0.0	508.0
ScreenPorch	1460.0	NaN	NaN	NaN	15.060959	55.757415	0.0	0.0	0.0	0.0	480.0
PoolArea	1460.0	NaN	NaN	NaN	2.758904	40.177307	0.0	0.0	0.0	0.0	738.0
PoolQC	7	3	Gd	3	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Fence	281	4	MnPrv	157	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MiscFeature	54	4	Shed	49	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
MiscVal	1460.0	NaN	NaN	NaN	43.489041	496.123024	0.0	0.0	0.0	0.0	15500.0
MoSold	1460.0	NaN	NaN	NaN	6.321918	2.703626	1.0	5.0	6.0	8.0	12.0
YrSold	1460.0	NaN	NaN	NaN	2007.815753	1.328095	2006.0	2007.0	2008.0	2009.0	2010.0
SaleType	1460	9	WD	1267	NaN	NaN	NaN	NaN	NaN	NaN	NaN
SaleCondition	1460	6	Normal	1198	NaN	NaN	NaN	NaN	NaN	NaN	NaN
SalePrice	1460.0	NaN	NaN	NaN	180921.19589	79442.502883	34900.0	129975.0	163000.0	214000.0	755000.0

In [8]: `data.isnull().sum()`

```
Out[8]: Id          0  
        MSSubClass    0  
        MSZoning      0  
        LotFrontage   259  
        LotArea       0  
        Street        0  
        Alley         1369  
        LotShape       0  
        LandContour    0  
        Utilities      0  
        LotConfig      0  
        LandSlope      0  
        Neighborhood   0  
        Condition1    0  
        Condition2    0  
        BldgType       0  
        HouseStyle     0  
        OverallQual    0  
        OverallCond    0  
        YearBuilt      0  
        YearRemodAdd   0  
        RoofStyle      0  
        RoofMatl       0  
        Exterior1st    0  
        Exterior2nd    0  
        MasVnrType     8  
        MasVnrArea     8  
        ExterQual      0  
        ExterCond      0  
        Foundation     0  
        BsmtQual       37  
        BsmtCond       37  
        BsmtExposure   38  
        BsmtFinType1   37  
        BsmtFinSF1     0  
        BsmtFinType2   38  
        BsmtFinSF2     0  
        BsmtUnfSF      0  
        TotalBsmtSF    0  
        Heating        0  
        HeatingQC      0  
        CentralAir     0  
        Electrical     1
```

```
1stFlrSF          0
2ndFlrSF          0
LowQualFinSF      0
GrLivArea         0
BsmtFullBath      0
BsmtHalfBath      0
FullBath          0
HalfBath          0
BedroomAbvGr      0
KitchenAbvGr      0
KitchenQual        0
TotRmsAbvGrd      0
Functional         0
Fireplaces         0
FireplaceQu       690
GarageType         81
GarageYrBlt        81
GarageFinish        81
GarageCars          0
GarageArea          0
GarageQual         81
GarageCond         81
PavedDrive         0
WoodDeckSF         0
OpenPorchSF        0
EnclosedPorch      0
3SsnPorch          0
ScreenPorch         0
PoolArea           0
PoolQC             1453
Fence              1179
MiscFeature        1406
MiscVal            0
MoSold             0
YrSold             0
SaleType            0
SaleCondition       0
SalePrice           0
dtype: int64
```

```
In [9]: for col in data.columns:
    print(f'the missing values of {col} is {round((data[col].isnull().sum()/data.shape[0]*100),2)}%')
```

the missing values of Id is 0.0%
the missing values of MSSubClass is 0.0%
the missing values of MSZoning is 0.0%
the missing values of LotFrontage is 17.74%
the missing values of LotArea is 0.0%
the missing values of Street is 0.0%
the missing values of Alley is 93.77%
the missing values of LotShape is 0.0%
the missing values of LandContour is 0.0%
the missing values of Utilities is 0.0%
the missing values of LotConfig is 0.0%
the missing values of LandSlope is 0.0%
the missing values of Neighborhood is 0.0%
the missing values of Condition1 is 0.0%
the missing values of Condition2 is 0.0%
the missing values of BldgType is 0.0%
the missing values of HouseStyle is 0.0%
the missing values of OverallQual is 0.0%
the missing values of OverallCond is 0.0%
the missing values of YearBuilt is 0.0%
the missing values of YearRemodAdd is 0.0%
the missing values of RoofStyle is 0.0%
the missing values of RoofMatl is 0.0%
the missing values of Exterior1st is 0.0%
the missing values of Exterior2nd is 0.0%
the missing values of MasVnrType is 0.55%
the missing values of MasVnrArea is 0.55%
the missing values of ExterQual is 0.0%
the missing values of ExterCond is 0.0%
the missing values of Foundation is 0.0%
the missing values of BsmtQual is 2.53%
the missing values of BsmtCond is 2.53%
the missing values of BsmtExposure is 2.6%
the missing values of BsmtFinType1 is 2.53%
the missing values of BsmtFinSF1 is 0.0%
the missing values of BsmtFinType2 is 2.6%
the missing values of BsmtFinSF2 is 0.0%
the missing values of BsmtUnfSF is 0.0%
the missing values of TotalBsmtSF is 0.0%
the missing values of Heating is 0.0%
the missing values of HeatingQC is 0.0%
the missing values of CentralAir is 0.0%
the missing values of Electrical is 0.07%

the missing values of 1stFlrSF is 0.0%
the missing values of 2ndFlrSF is 0.0%
the missing values of LowQualFinSF is 0.0%
the missing values of GrLivArea is 0.0%
the missing values of BsmtFullBath is 0.0%
the missing values of BsmtHalfBath is 0.0%
the missing values of FullBath is 0.0%
the missing values of HalfBath is 0.0%
the missing values of BedroomAbvGr is 0.0%
the missing values of KitchenAbvGr is 0.0%
the missing values of KitchenQual is 0.0%
the missing values of TotRmsAbvGrd is 0.0%
the missing values of Functional is 0.0%
the missing values of Fireplaces is 0.0%
the missing values of FireplaceQu is 47.26%
the missing values of GarageType is 5.55%
the missing values of GarageYrBlt is 5.55%
the missing values of GarageFinish is 5.55%
the missing values of GarageCars is 0.0%
the missing values of GarageArea is 0.0%
the missing values of GarageQual is 5.55%
the missing values of GarageCond is 5.55%
the missing values of PavedDrive is 0.0%
the missing values of WoodDeckSF is 0.0%
the missing values of OpenPorchSF is 0.0%
the missing values of EnclosedPorch is 0.0%
the missing values of 3SsnPorch is 0.0%
the missing values of ScreenPorch is 0.0%
the missing values of PoolArea is 0.0%
the missing values of PoolQC is 99.52%
the missing values of Fence is 80.75%
the missing values of MiscFeature is 96.3%
the missing values of MiscVal is 0.0%
the missing values of MoSold is 0.0%
the missing values of YrSold is 0.0%
the missing values of SaleType is 0.0%
the missing values of SaleCondition is 0.0%
the missing values of SalePrice is 0.0%

```
In [10]: null_cols = []
for col in data.columns:
    if data[col].isnull().sum() > 600:
        null_cols.append(col)
```

```
In [11]: null_cols
```

```
Out[11]: ['Alley', 'FireplaceQu', 'PoolQC', 'Fence', 'MiscFeature']
```

```
In [12]: data.drop(null_cols, axis='columns', inplace=True)
```

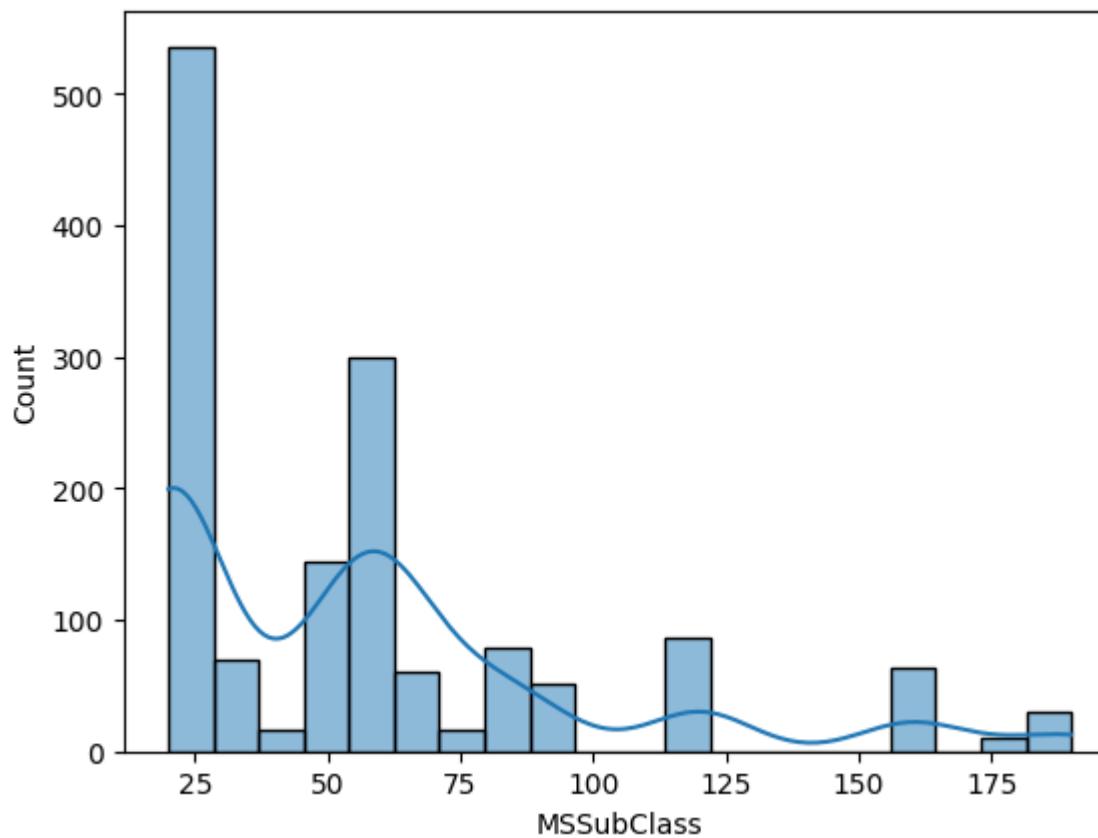
```
In [13]: for col in data.columns:
    print(f'the missing values of {col} is {round((data[col].isnull().sum()/data.shape[0]*100),2)}%')
```

the missing values of Id is 0.0%
the missing values of MSSubClass is 0.0%
the missing values of MSZoning is 0.0%
the missing values of LotFrontage is 17.74%
the missing values of LotArea is 0.0%
the missing values of Street is 0.0%
the missing values of LotShape is 0.0%
the missing values of LandContour is 0.0%
the missing values of Utilities is 0.0%
the missing values of LotConfig is 0.0%
the missing values of LandSlope is 0.0%
the missing values of Neighborhood is 0.0%
the missing values of Condition1 is 0.0%
the missing values of Condition2 is 0.0%
the missing values of BldgType is 0.0%
the missing values of HouseStyle is 0.0%
the missing values of OverallQual is 0.0%
the missing values of OverallCond is 0.0%
the missing values of YearBuilt is 0.0%
the missing values of YearRemodAdd is 0.0%
the missing values of RoofStyle is 0.0%
the missing values of RoofMatl is 0.0%
the missing values of Exterior1st is 0.0%
the missing values of Exterior2nd is 0.0%
the missing values of MasVnrType is 0.55%
the missing values of MasVnrArea is 0.55%
the missing values of ExterQual is 0.0%
the missing values of ExterCond is 0.0%
the missing values of Foundation is 0.0%
the missing values of BsmtQual is 2.53%
the missing values of BsmtCond is 2.53%
the missing values of BsmtExposure is 2.6%
the missing values of BsmtFinType1 is 2.53%
the missing values of BsmtFinSF1 is 0.0%
the missing values of BsmtFinType2 is 2.6%
the missing values of BsmtFinSF2 is 0.0%
the missing values of BsmtUnfSF is 0.0%
the missing values of TotalBsmtSF is 0.0%
the missing values of Heating is 0.0%
the missing values of HeatingQC is 0.0%
the missing values of CentralAir is 0.0%
the missing values of Electrical is 0.07%
the missing values of 1stFlrSF is 0.0%

```
the missing values of 2ndFlrSF is 0.0%
the missing values of LowQualFinSF is 0.0%
the missing values of GrLivArea is 0.0%
the missing values of BsmtFullBath is 0.0%
the missing values of BsmtHalfBath is 0.0%
the missing values of FullBath is 0.0%
the missing values of HalfBath is 0.0%
the missing values of BedroomAbvGr is 0.0%
the missing values of KitchenAbvGr is 0.0%
the missing values of KitchenQual is 0.0%
the missing values of TotRmsAbvGrd is 0.0%
the missing values of Functional is 0.0%
the missing values of Fireplaces is 0.0%
the missing values of GarageType is 5.55%
the missing values of GarageYrBlt is 5.55%
the missing values of GarageFinish is 5.55%
the missing values of GarageCars is 0.0%
the missing values of GarageArea is 0.0%
the missing values of GarageQual is 5.55%
the missing values of GarageCond is 5.55%
the missing values of PavedDrive is 0.0%
the missing values of WoodDeckSF is 0.0%
the missing values of OpenPorchSF is 0.0%
the missing values of EnclosedPorch is 0.0%
the missing values of 3SsnPorch is 0.0%
the missing values of ScreenPorch is 0.0%
the missing values of PoolArea is 0.0%
the missing values of MiscVal is 0.0%
the missing values of MoSold is 0.0%
the missing values of YrSold is 0.0%
the missing values of SaleType is 0.0%
the missing values of SaleCondition is 0.0%
the missing values of SalePrice is 0.0%
```

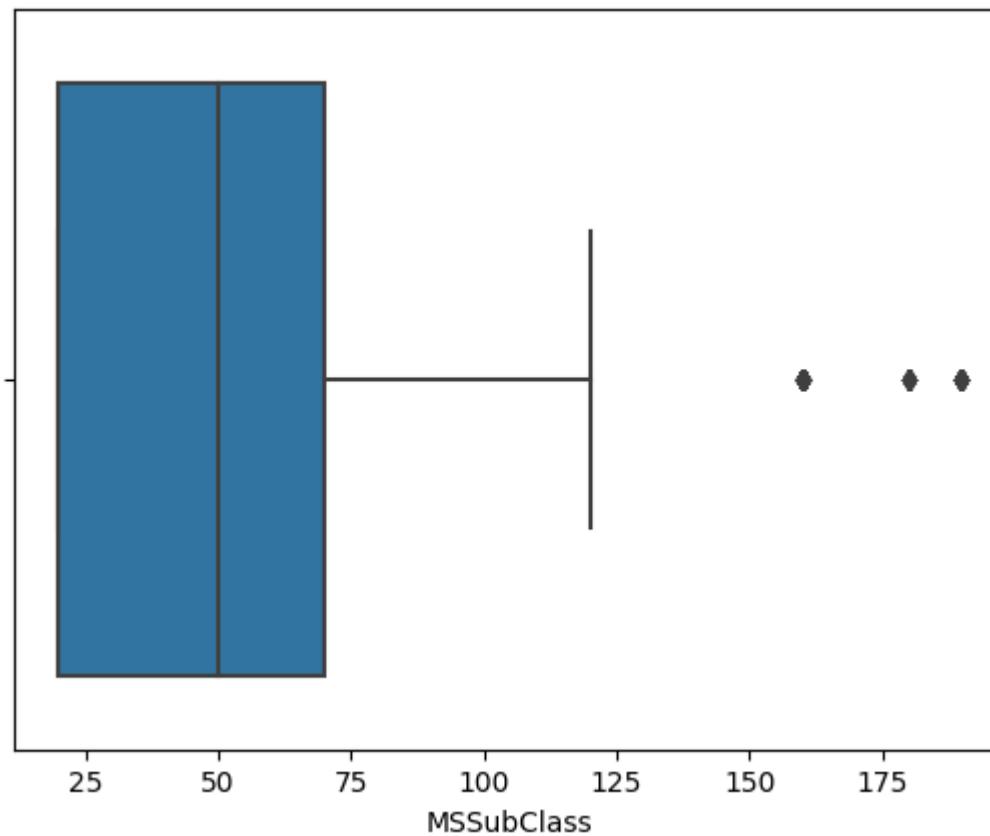
```
In [14]: sns.histplot(x=data['MSSubClass'], bins=20, kde=True)
```

```
Out[14]: <AxesSubplot:xlabel='MSSubClass', ylabel='Count'>
```



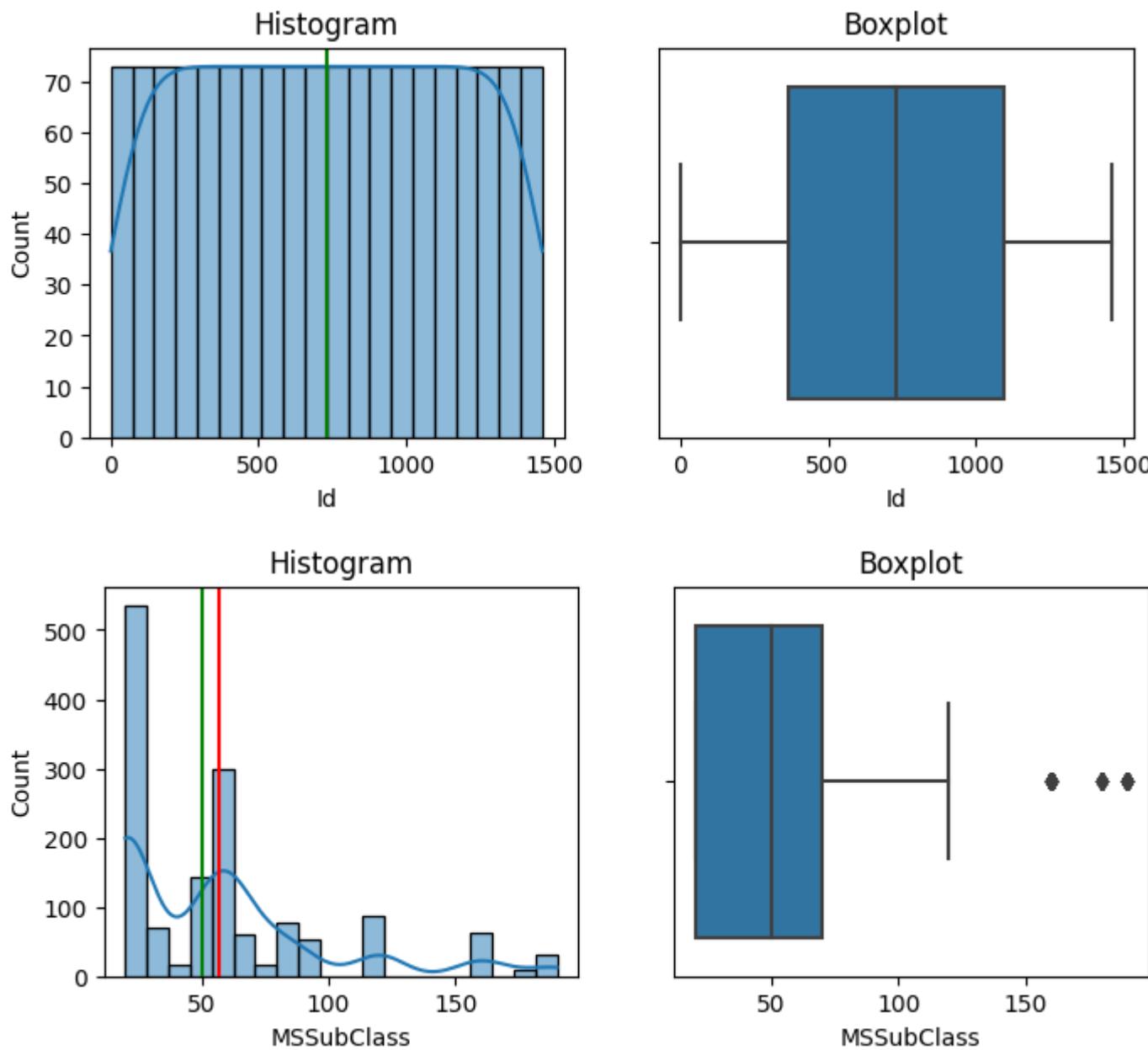
```
In [15]: sns.boxplot(x=data['MSSubClass'])
```

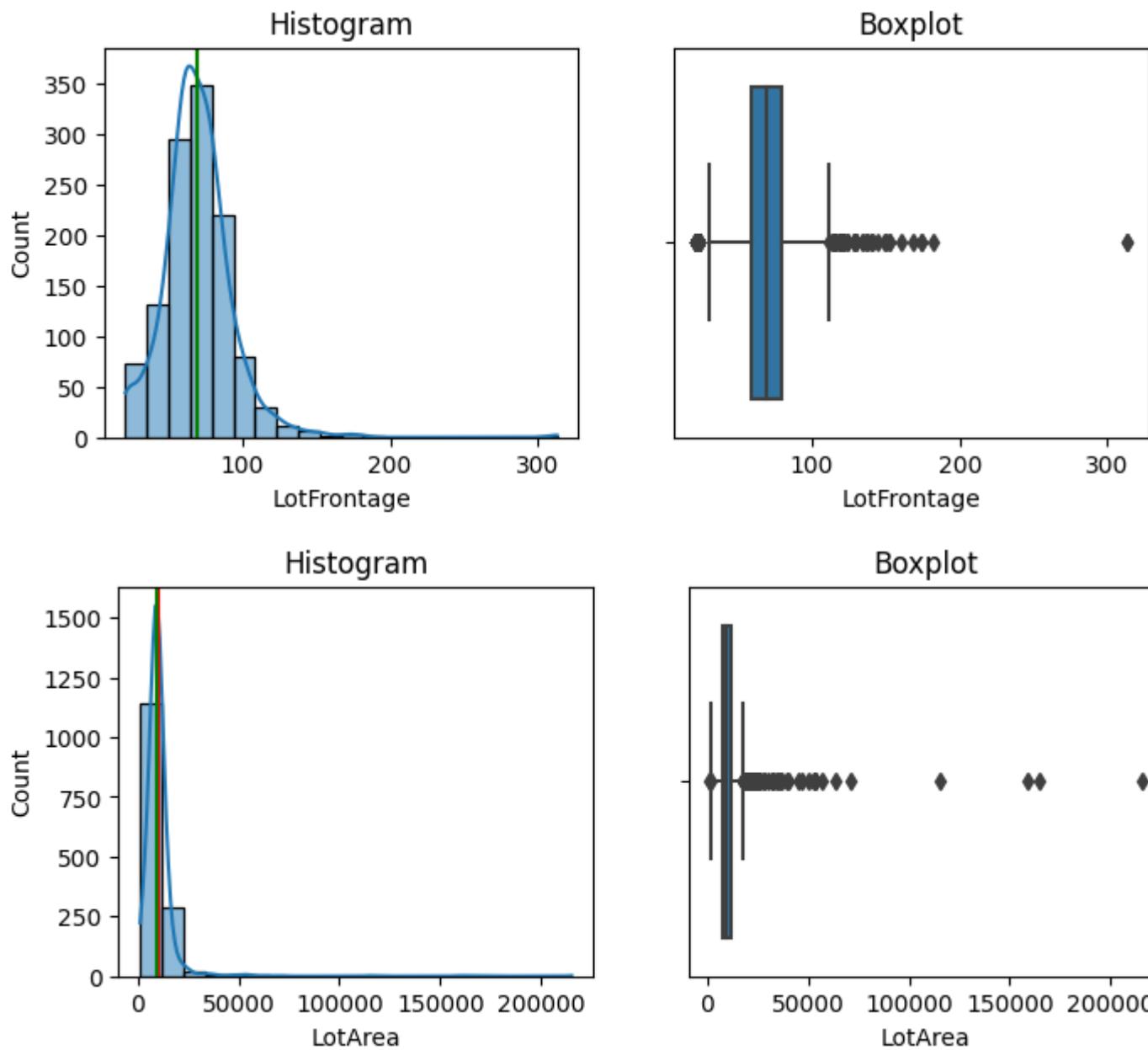
```
Out[15]: <AxesSubplot:xlabel='MSSubClass'>
```

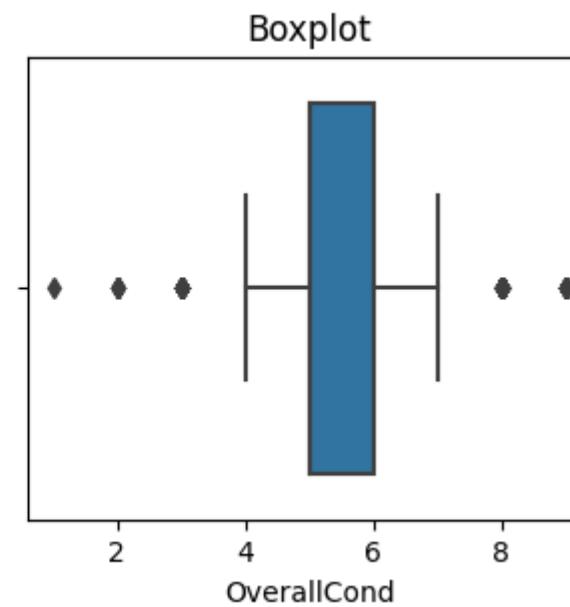
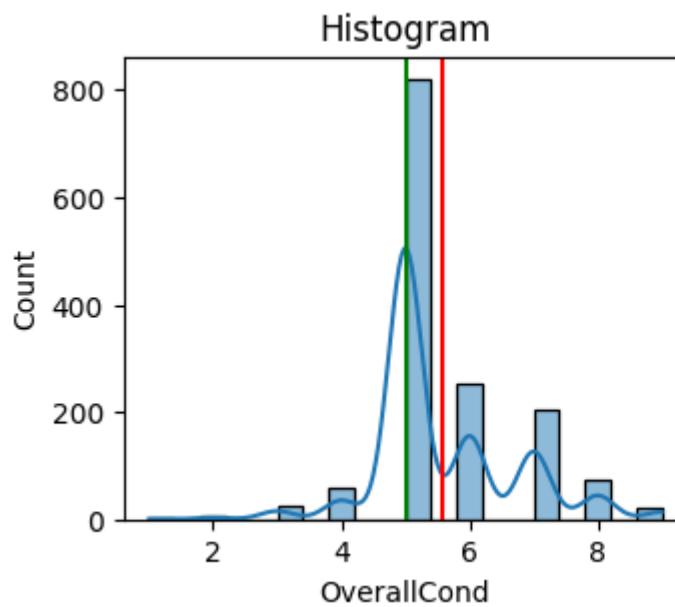
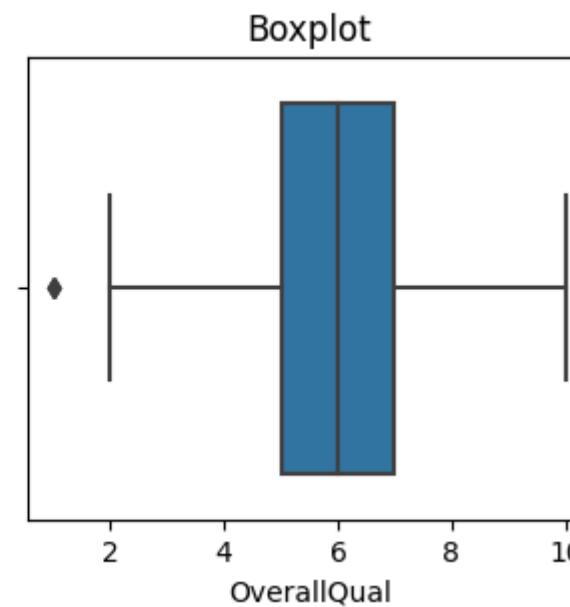
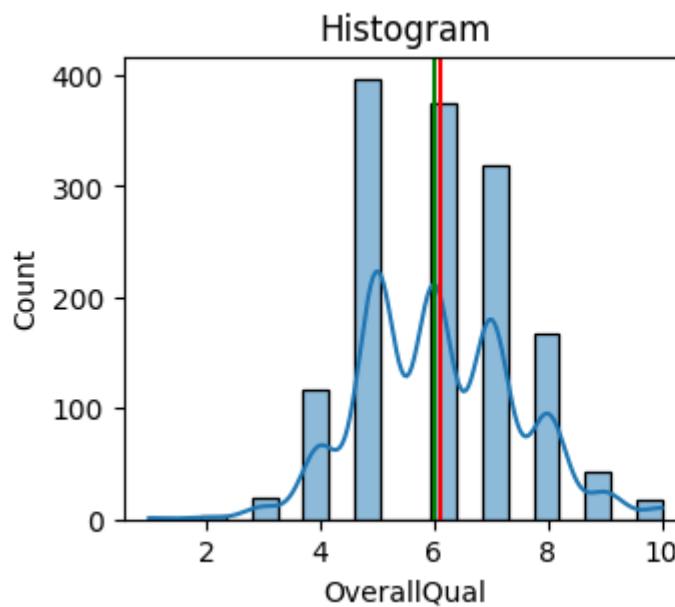


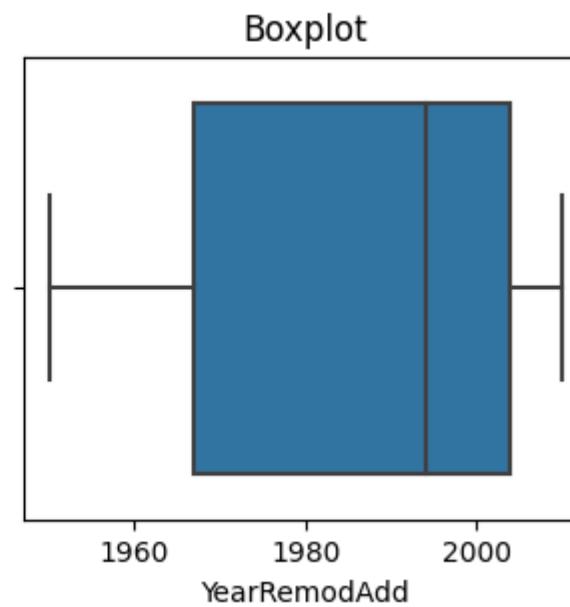
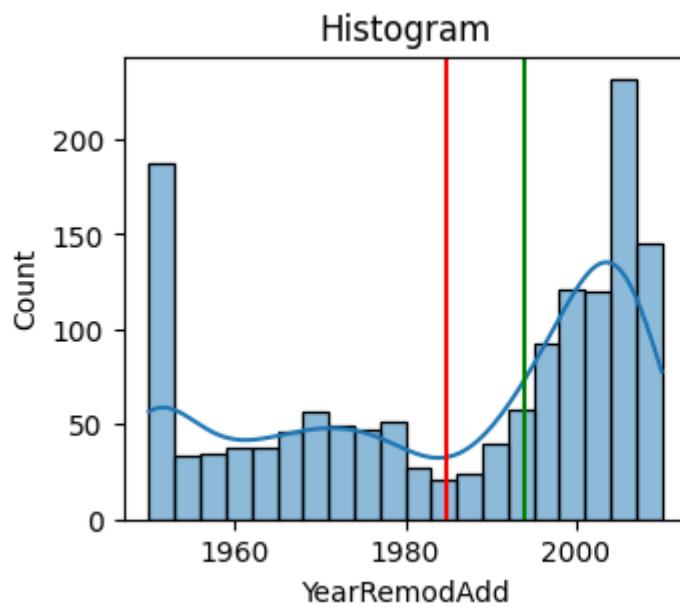
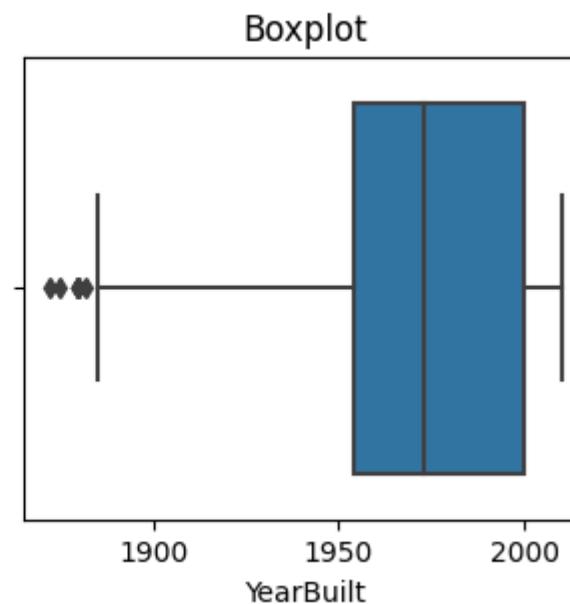
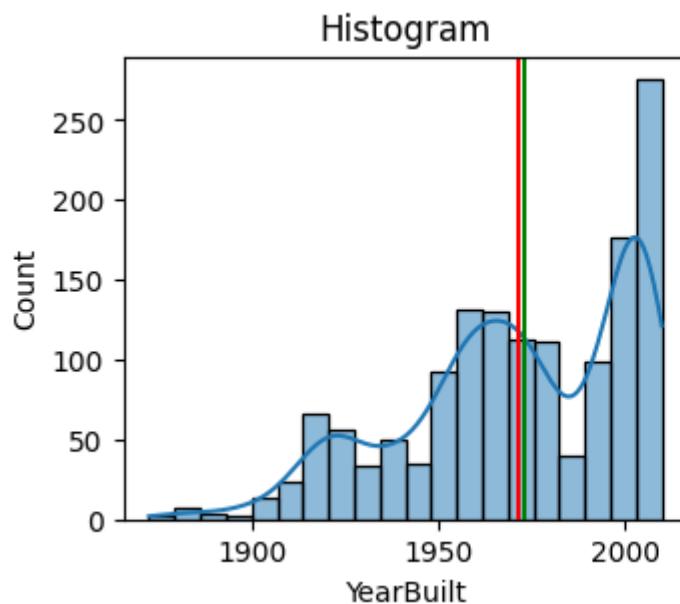
```
In [16]: def uniplot(col):
    plt.figure(figsize=(8,3))
    plt.subplot(1,2,1)
    sns.histplot(x=data[col], bins=20, kde=True)
    plt.axvline(x=data[col].mean(), ymin=0, ymax=1, color="red")
    plt.axvline(x=data[col].median(), ymin=0, ymax=1, color="green")
    plt.title("Histogram")
    plt.subplot(1,2,2)
    sns.boxplot(x=data[col])
    plt.title("Boxplot")
    plt.show()
```

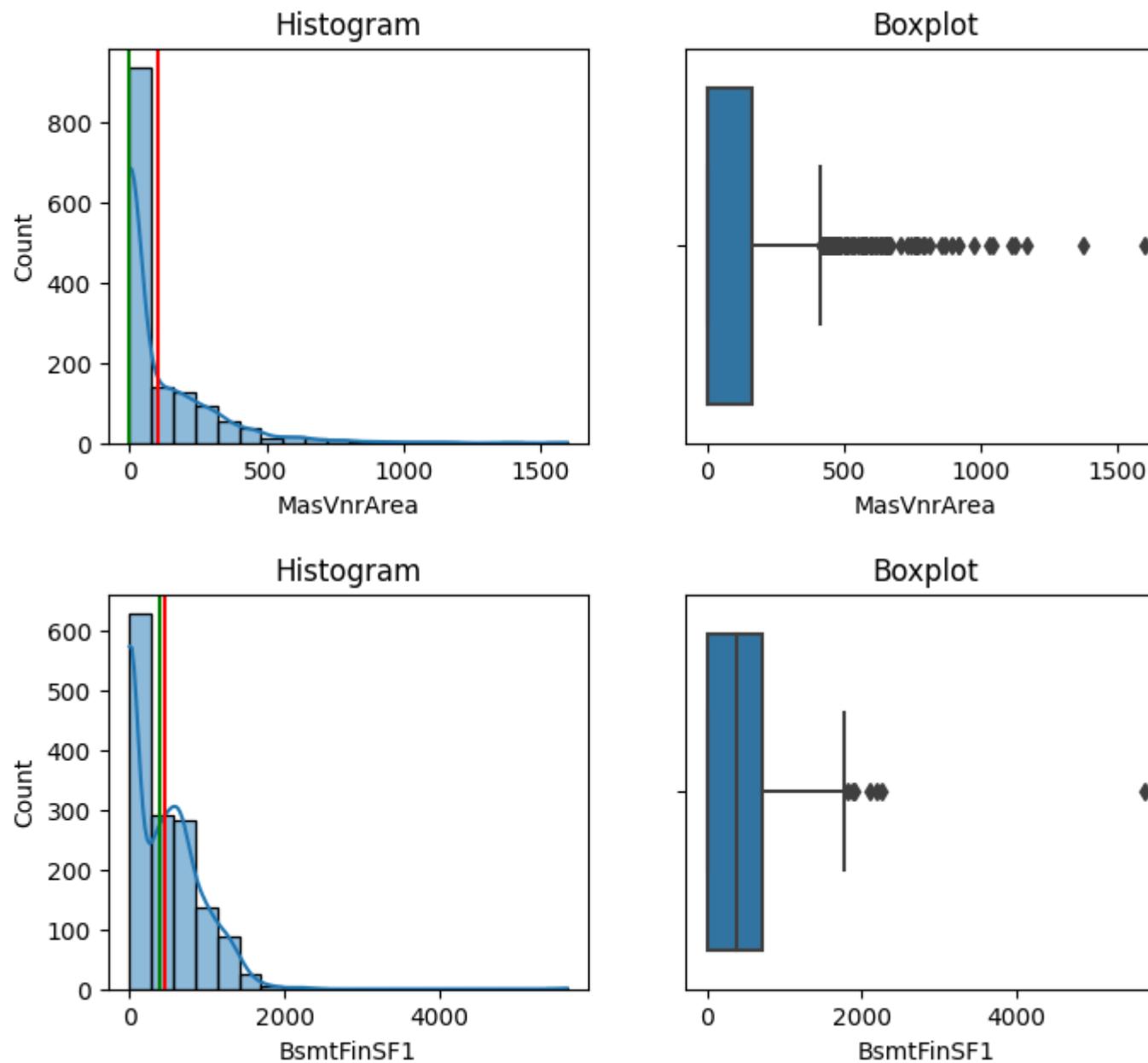
```
In [17]: for col in data.select_dtypes(exclude='object'):
    uniplot(col)
```

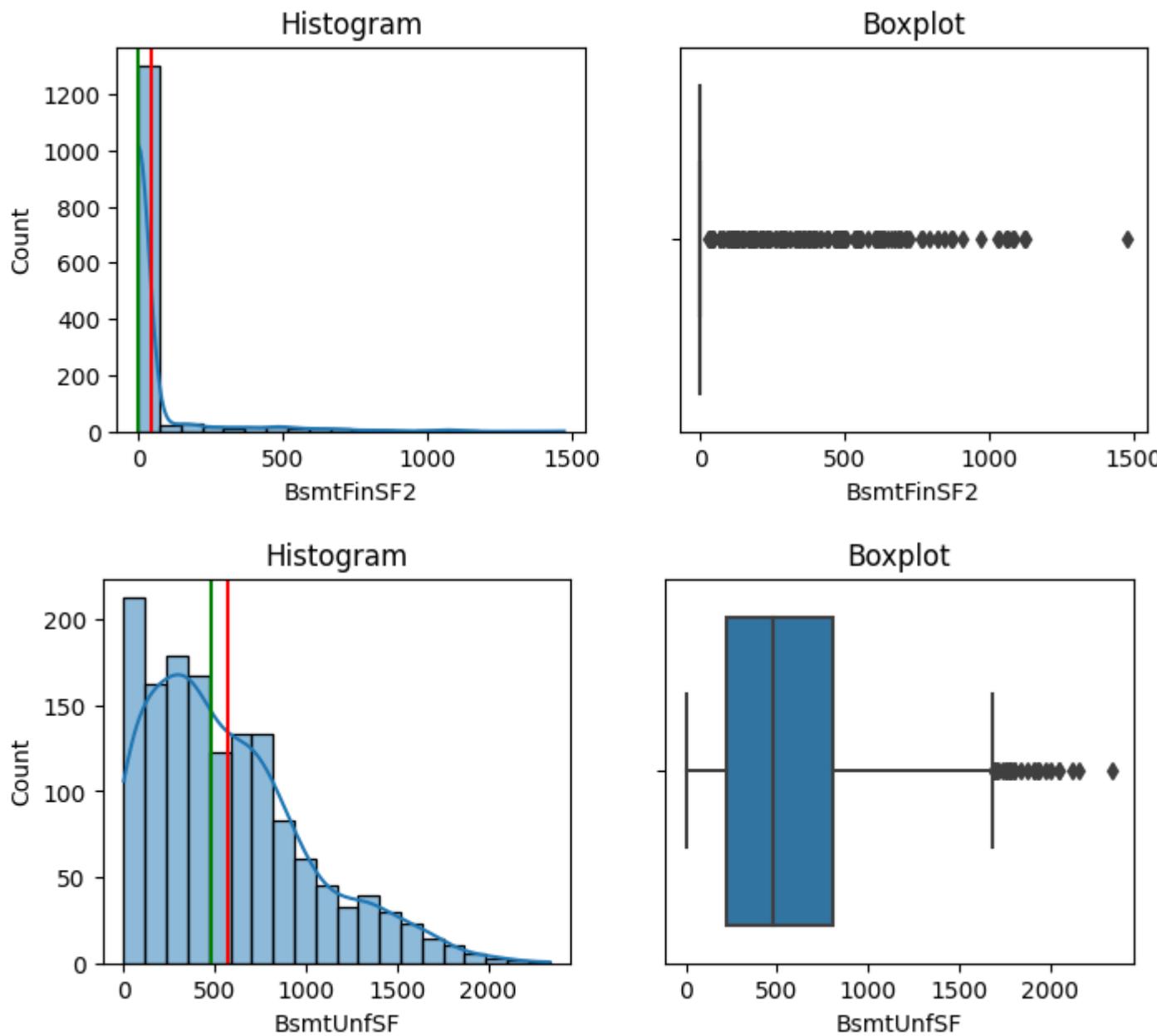


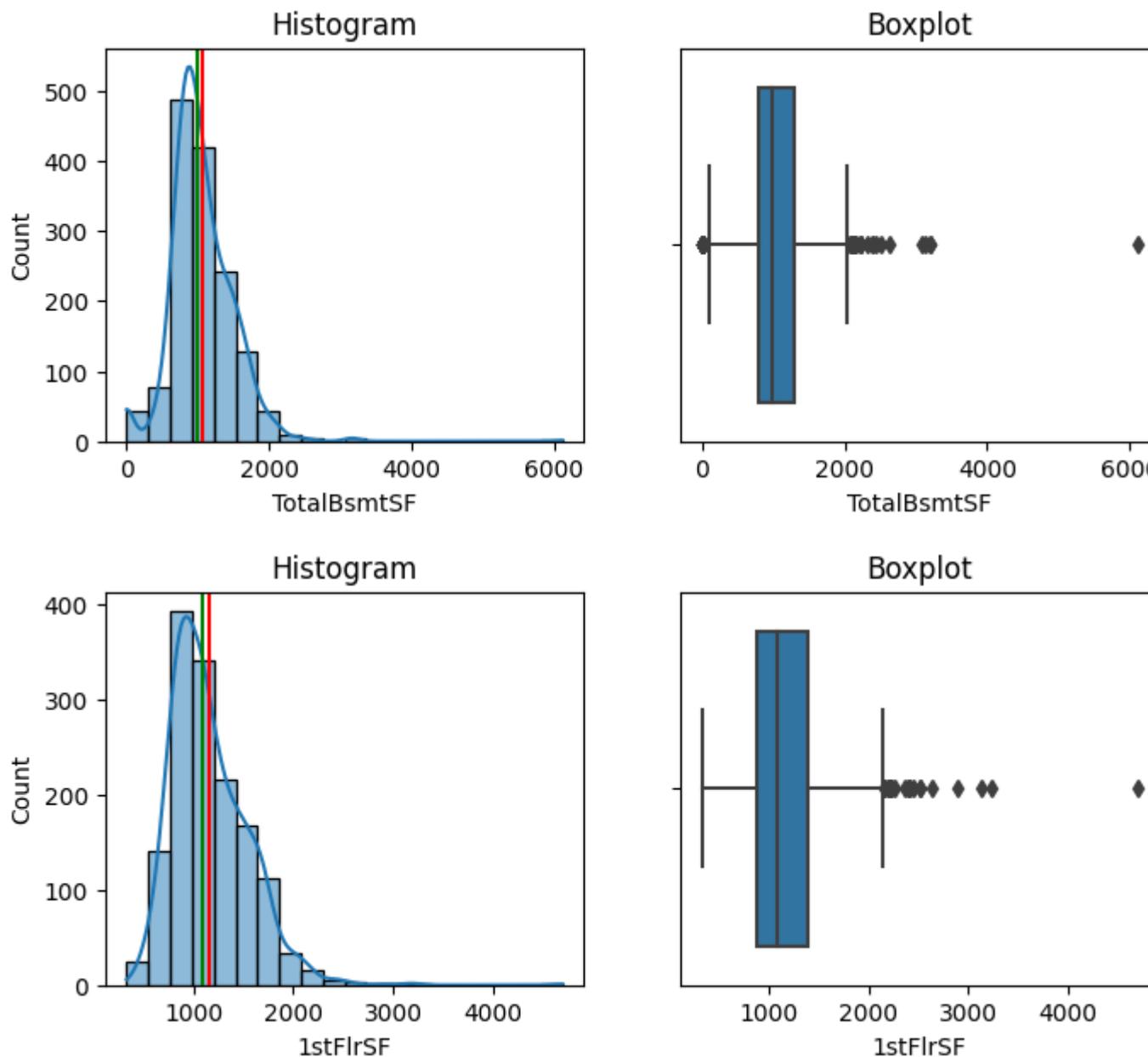


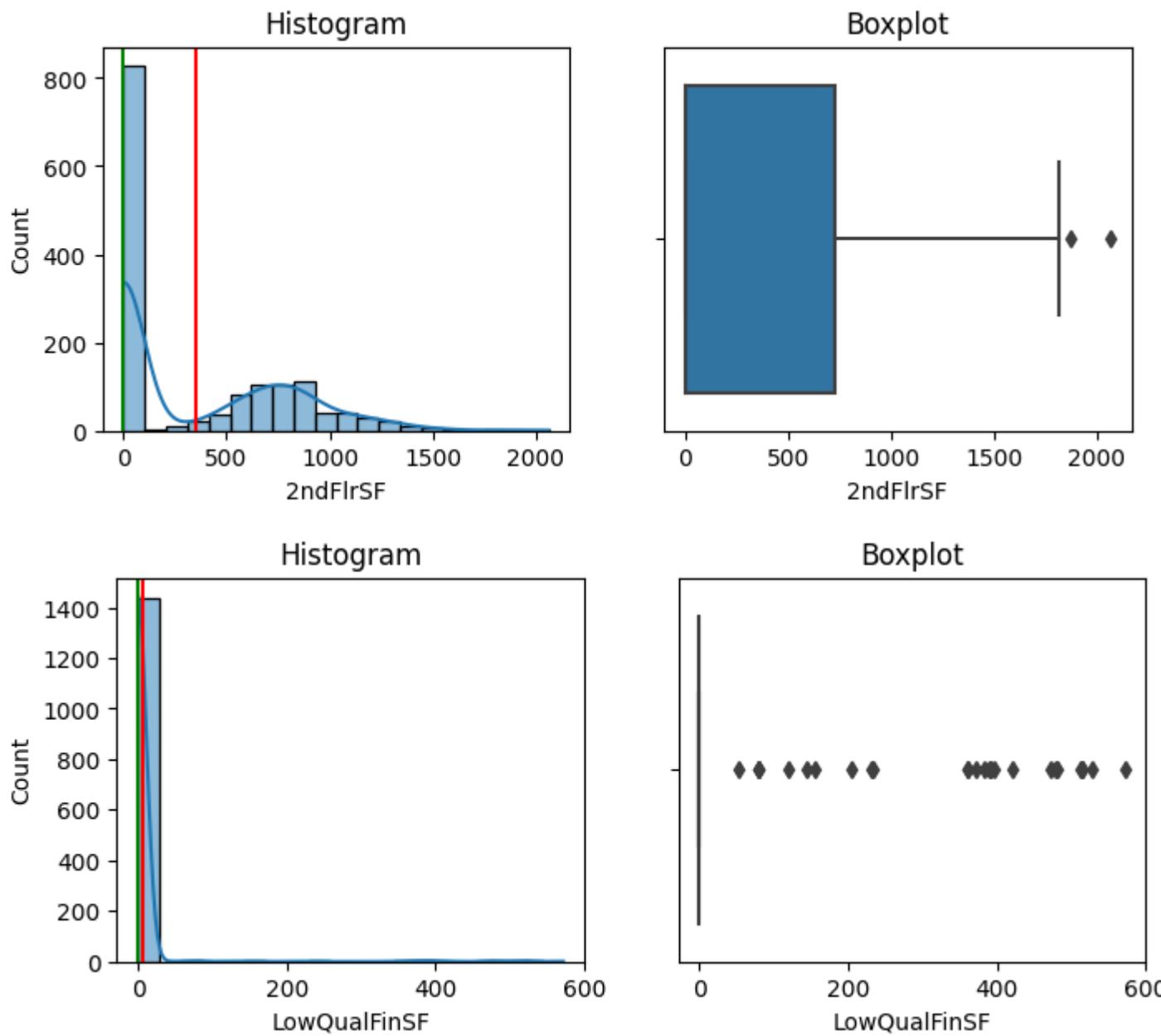


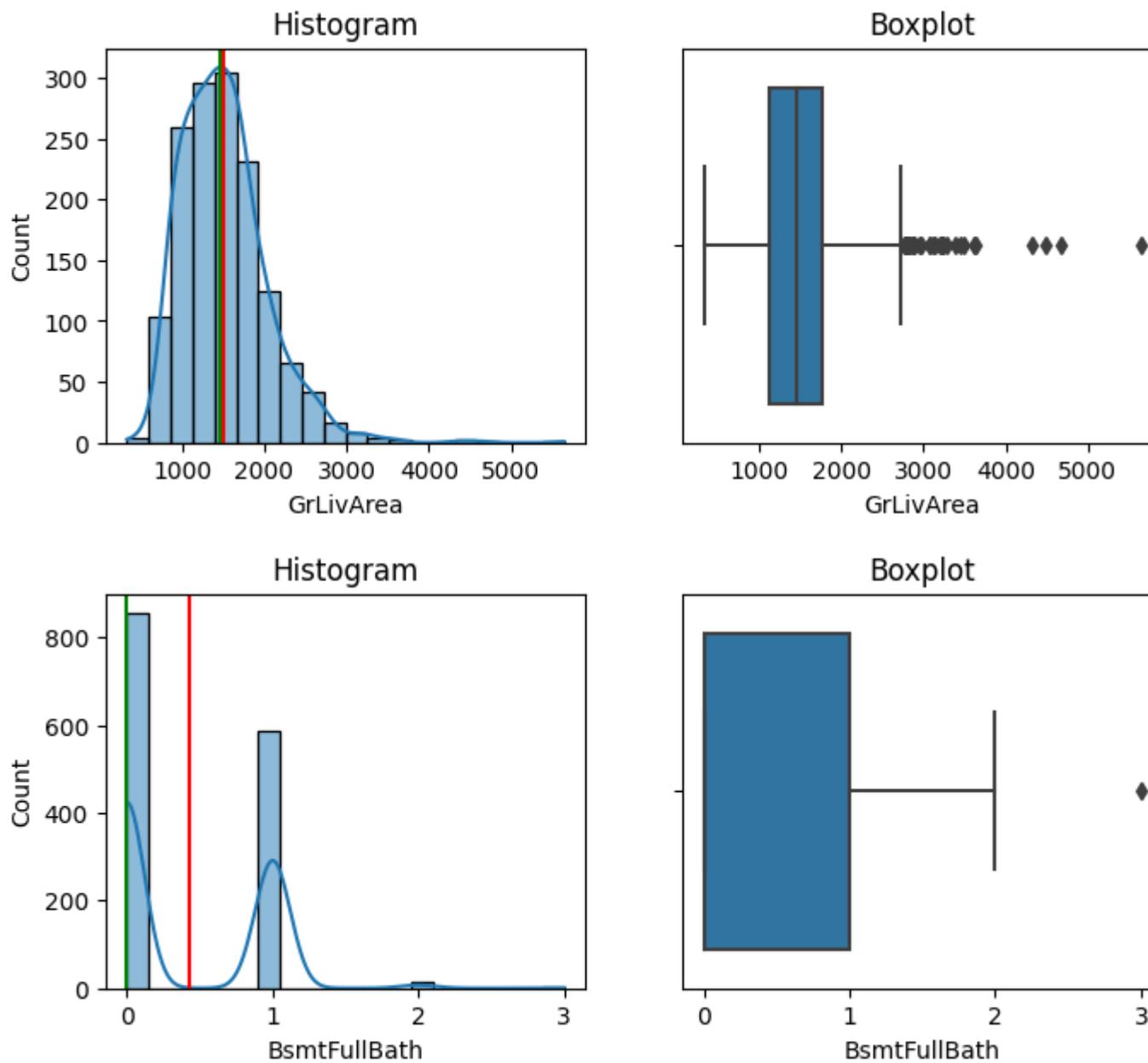


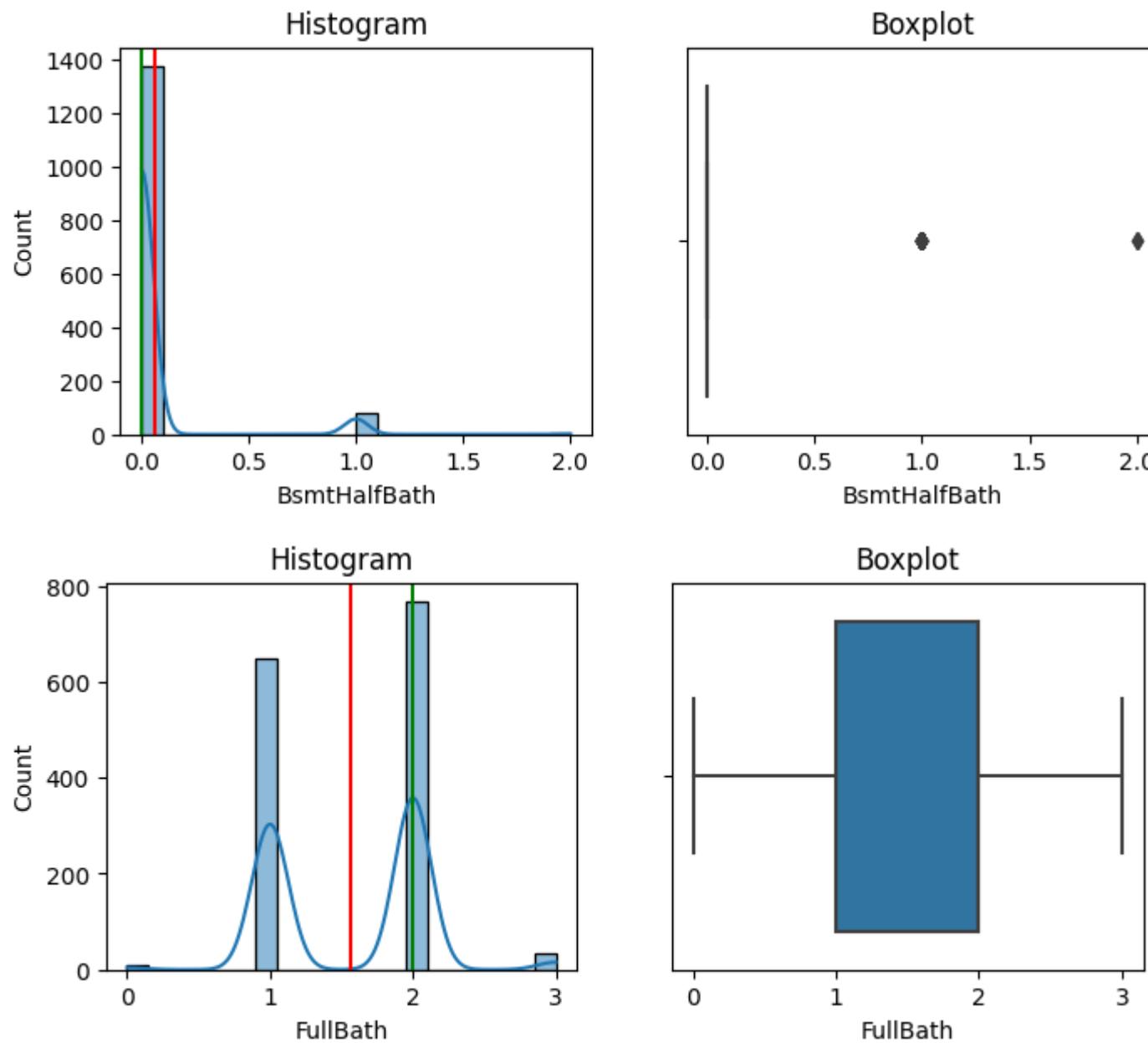


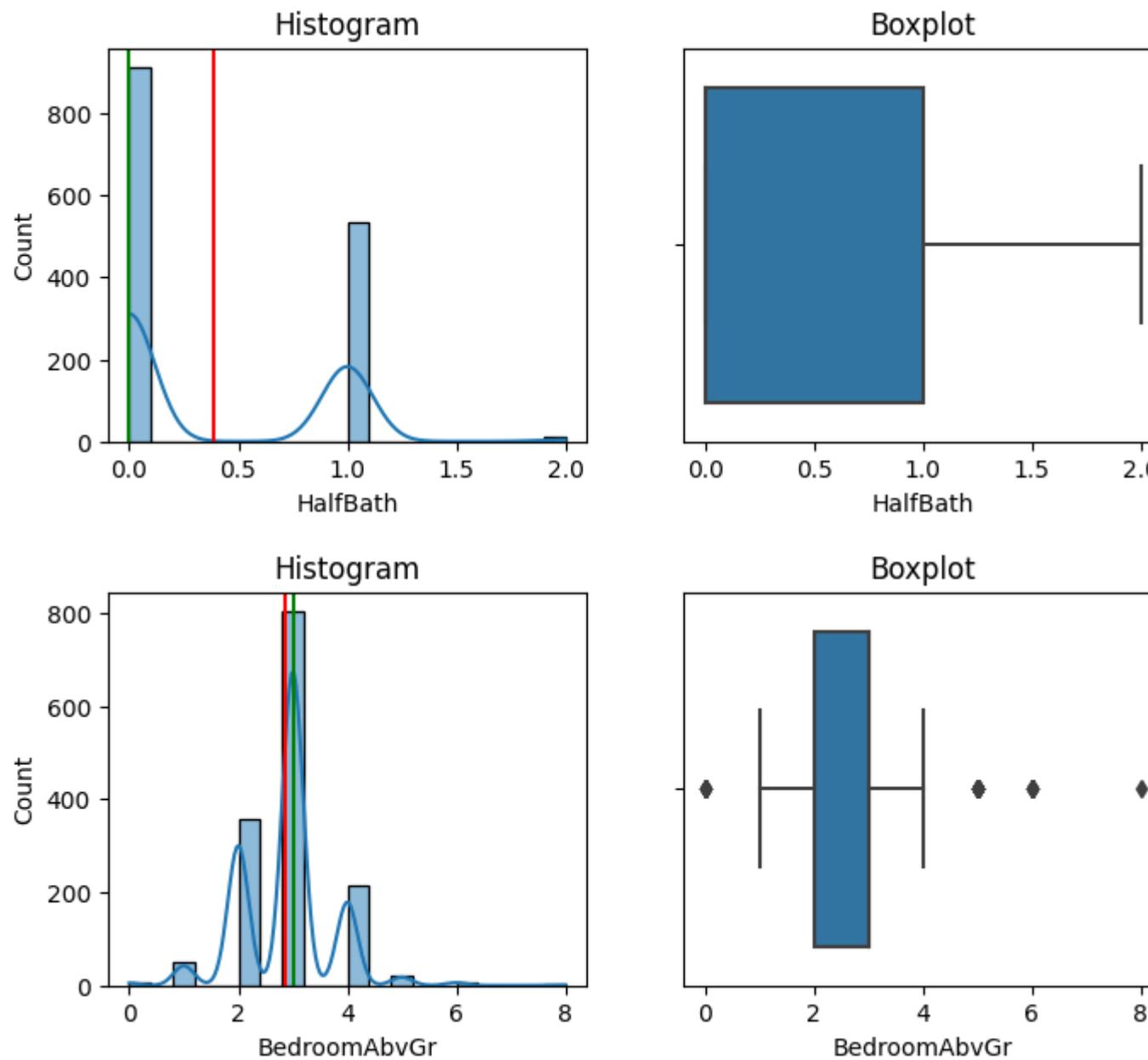


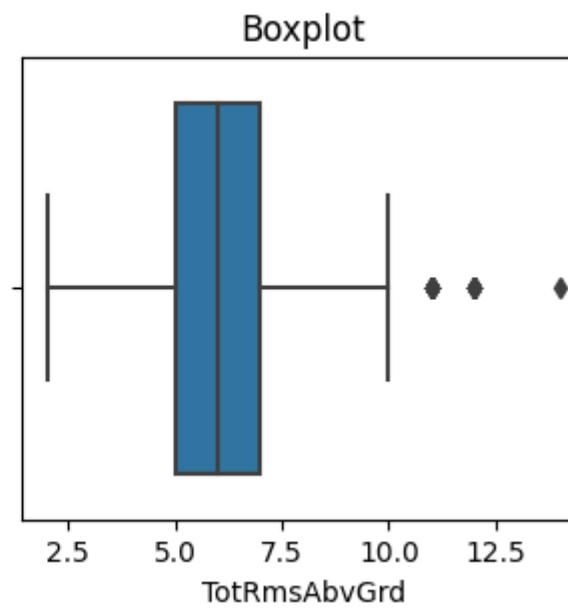
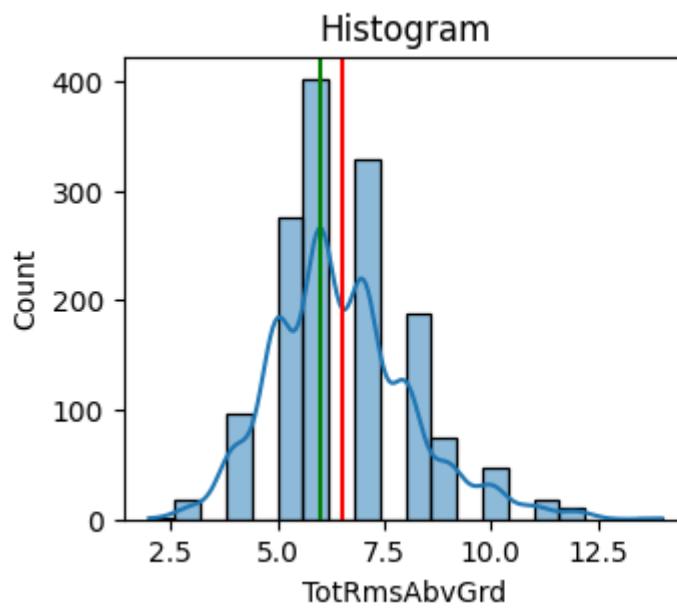
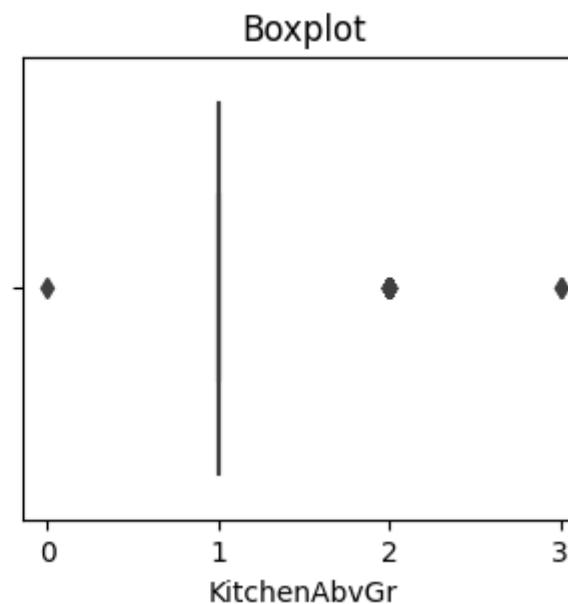
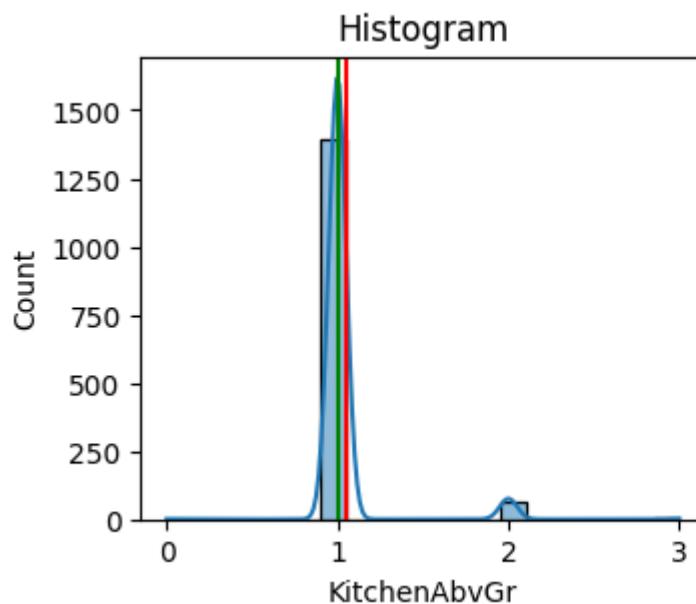


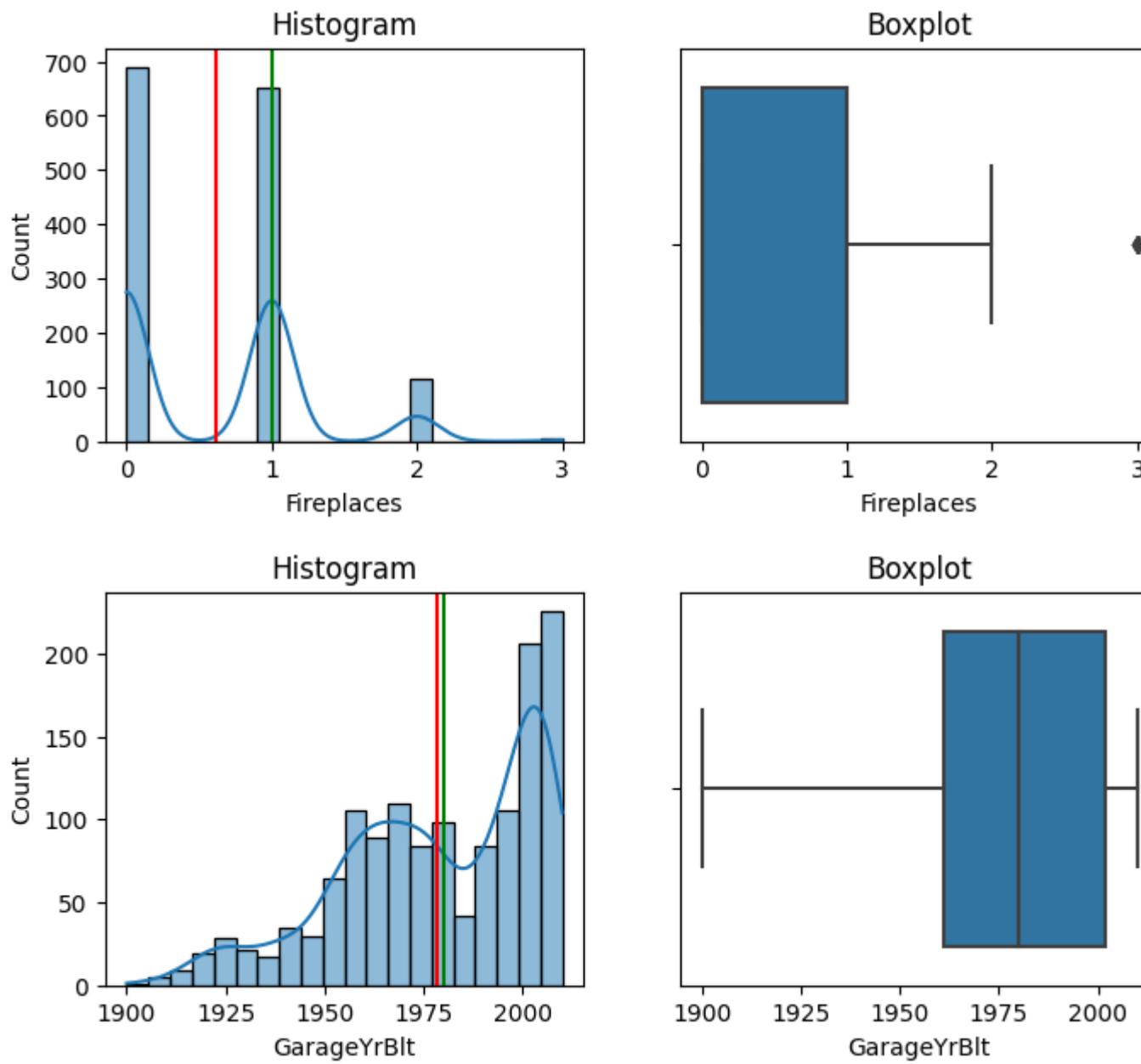


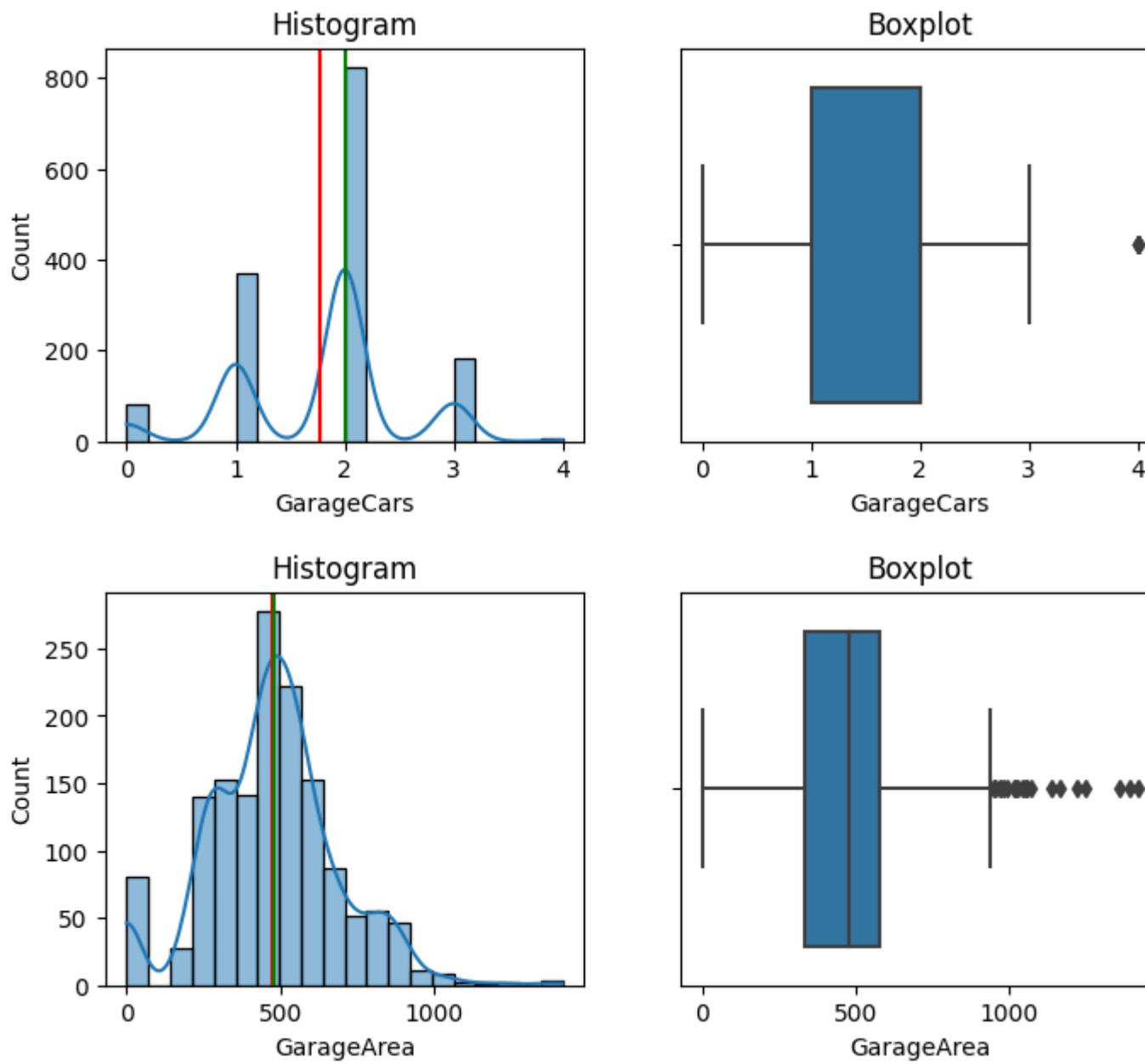


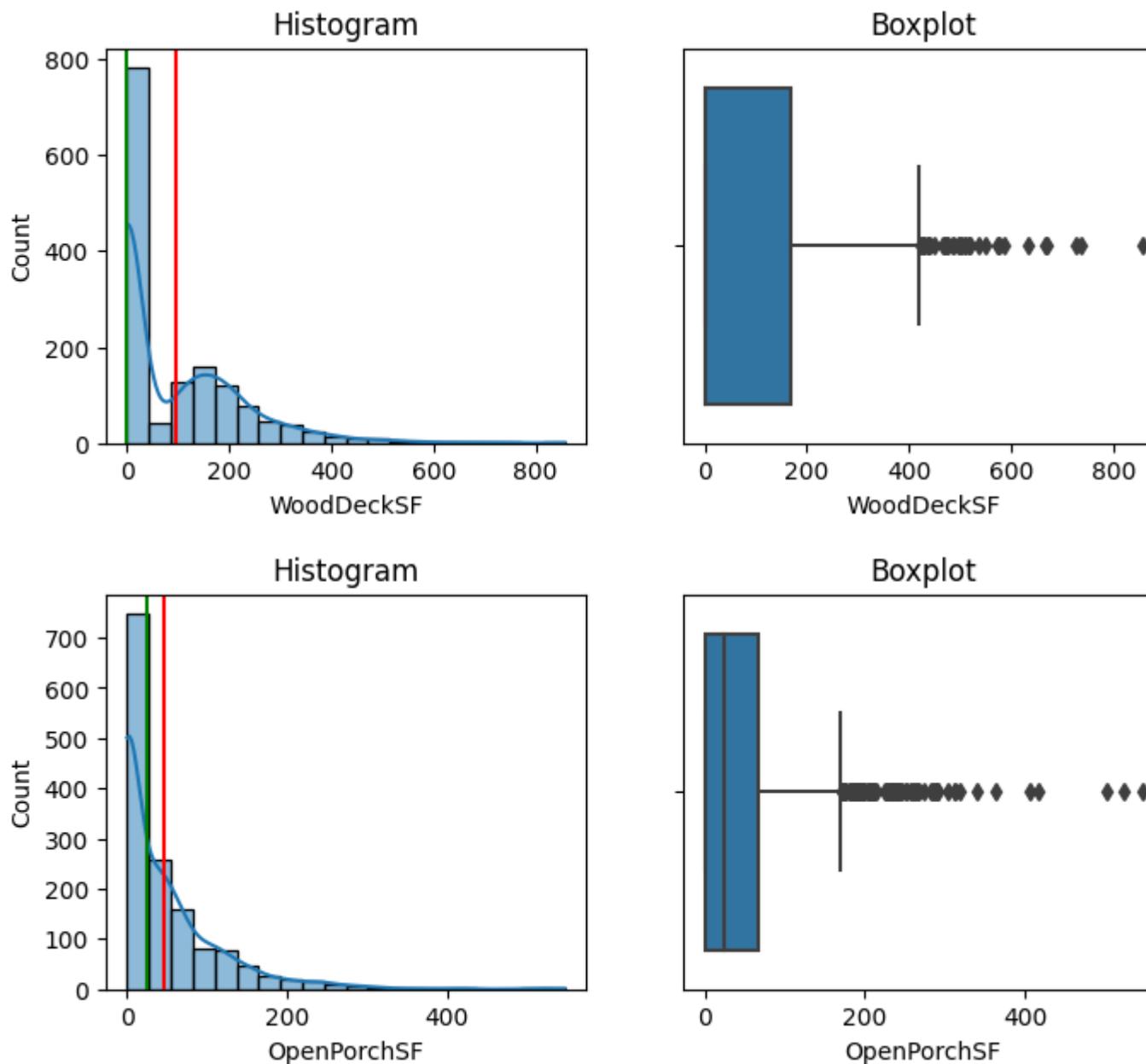


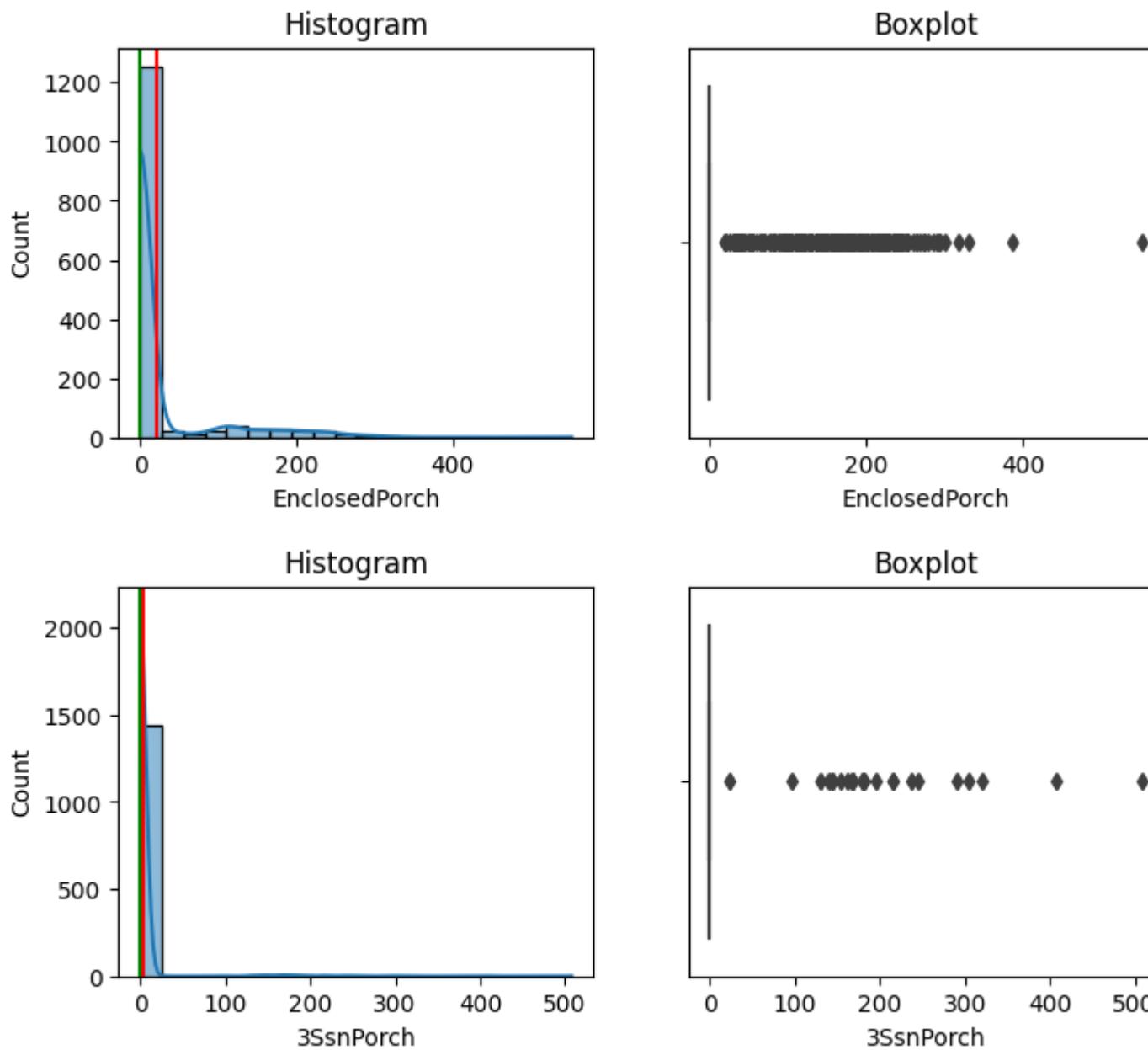


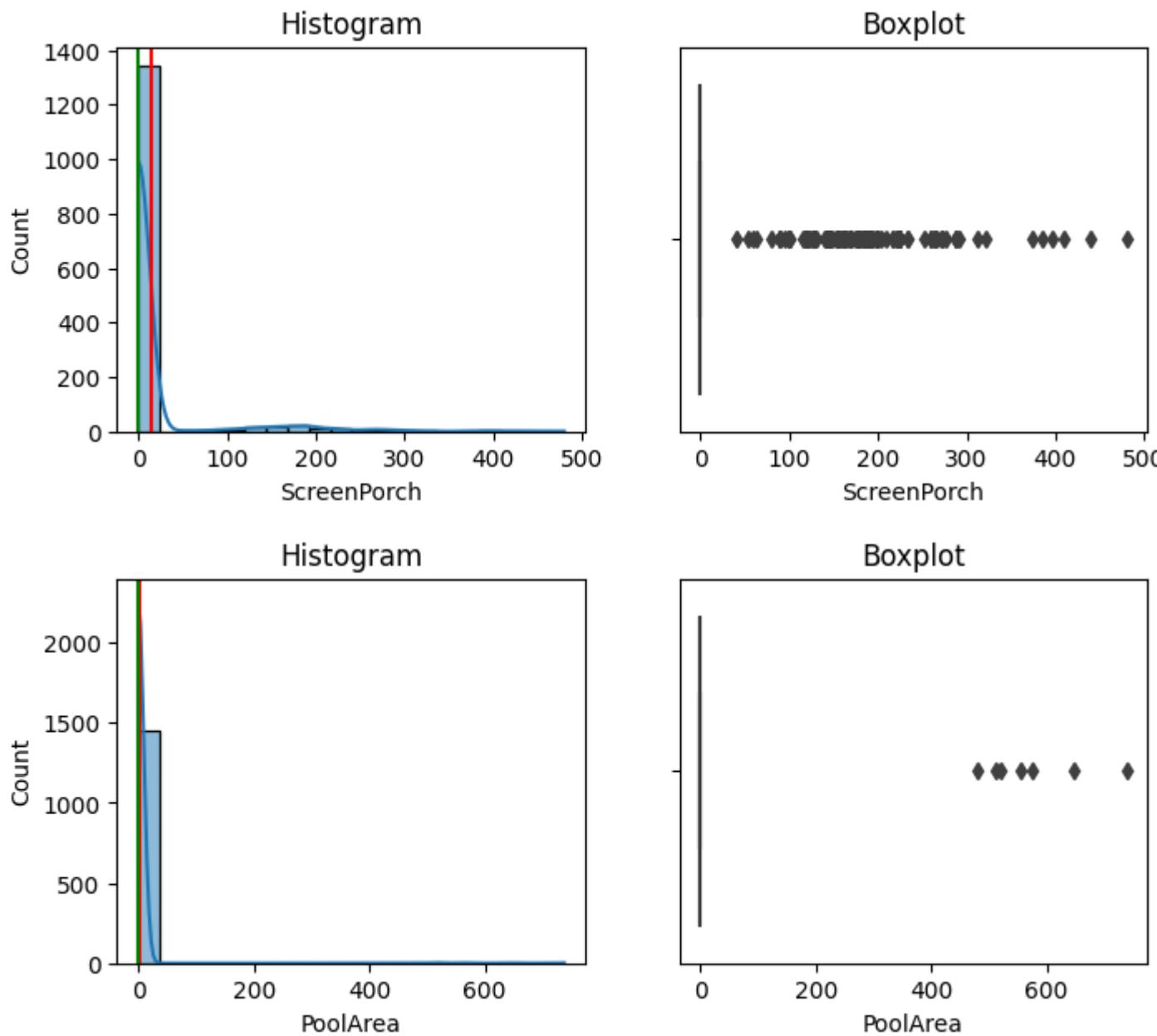


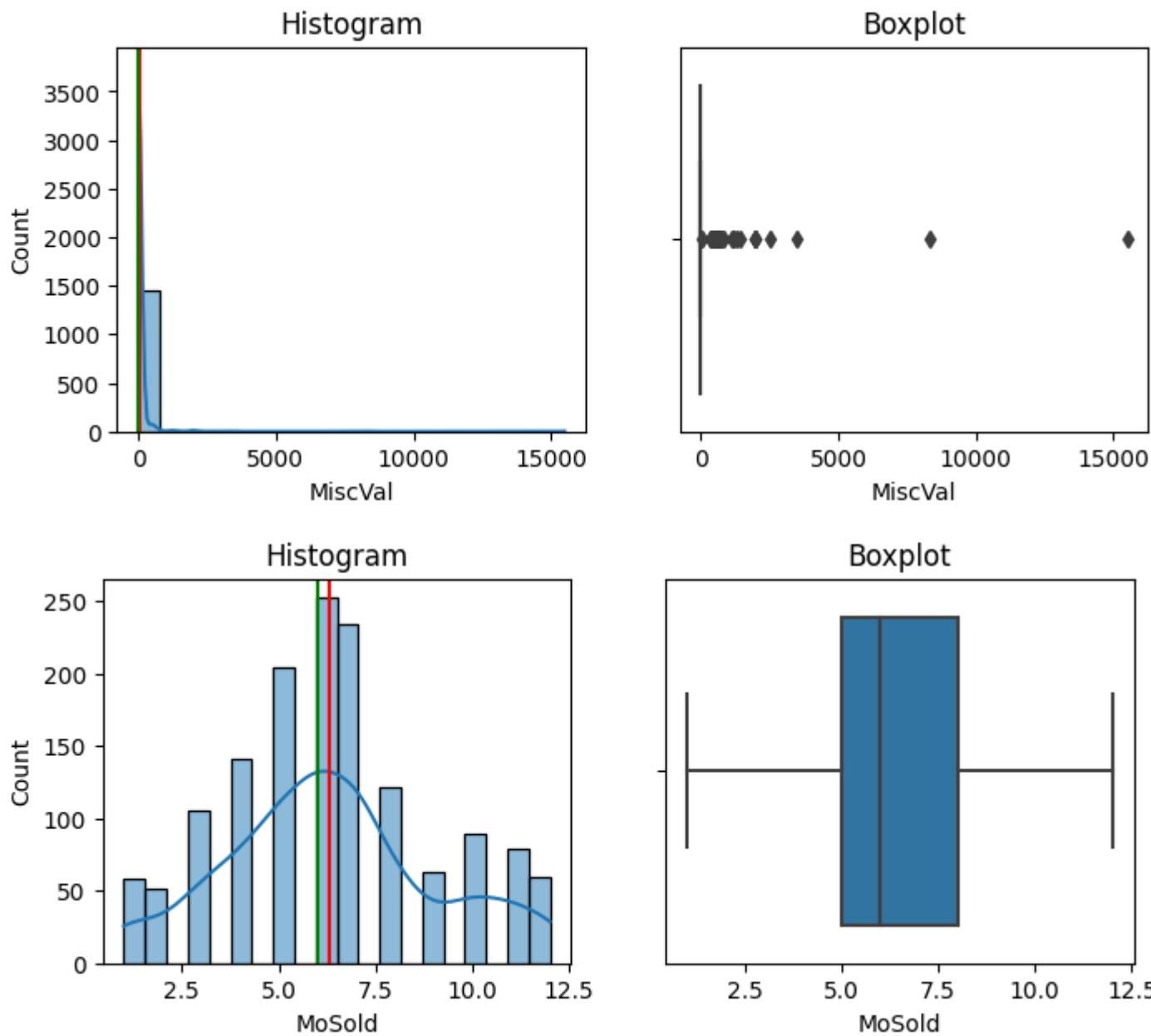


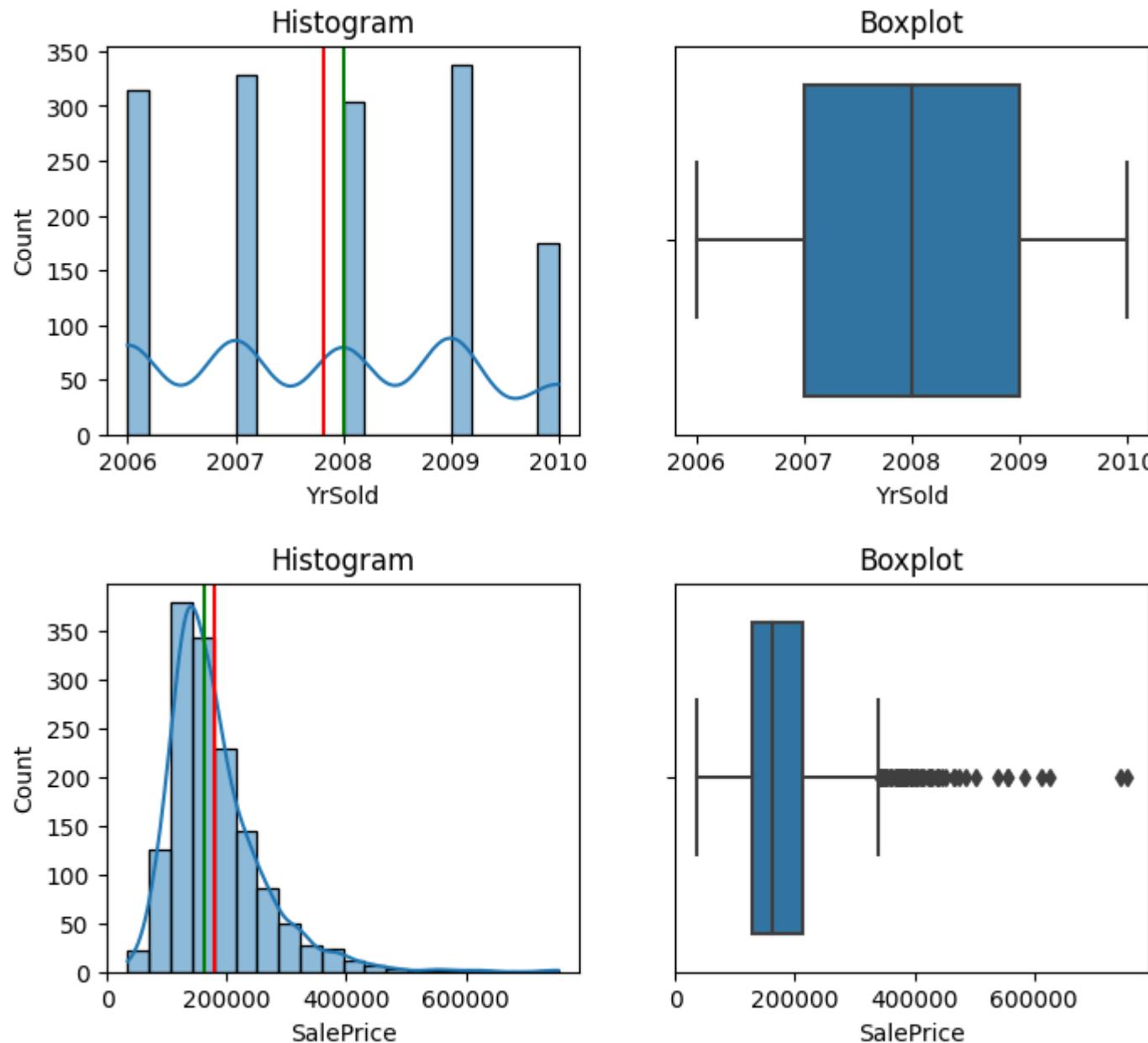








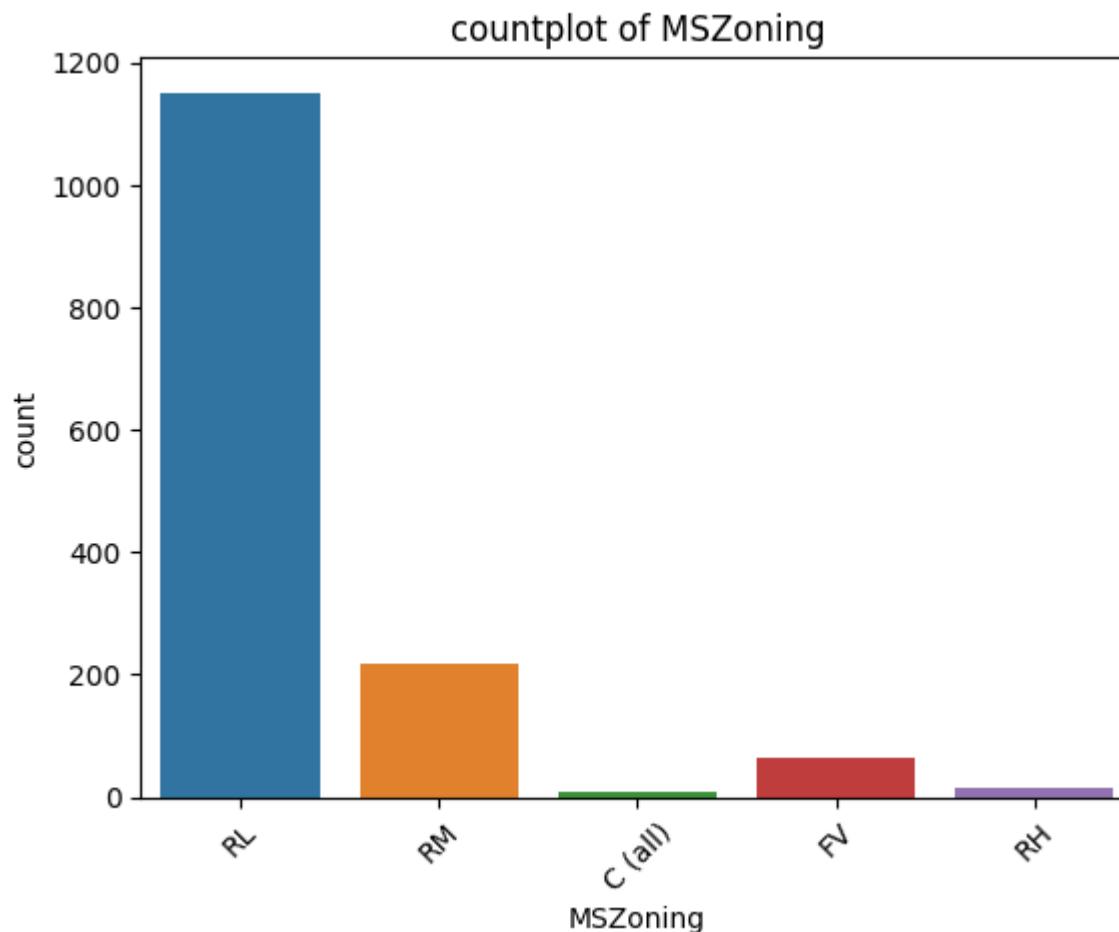


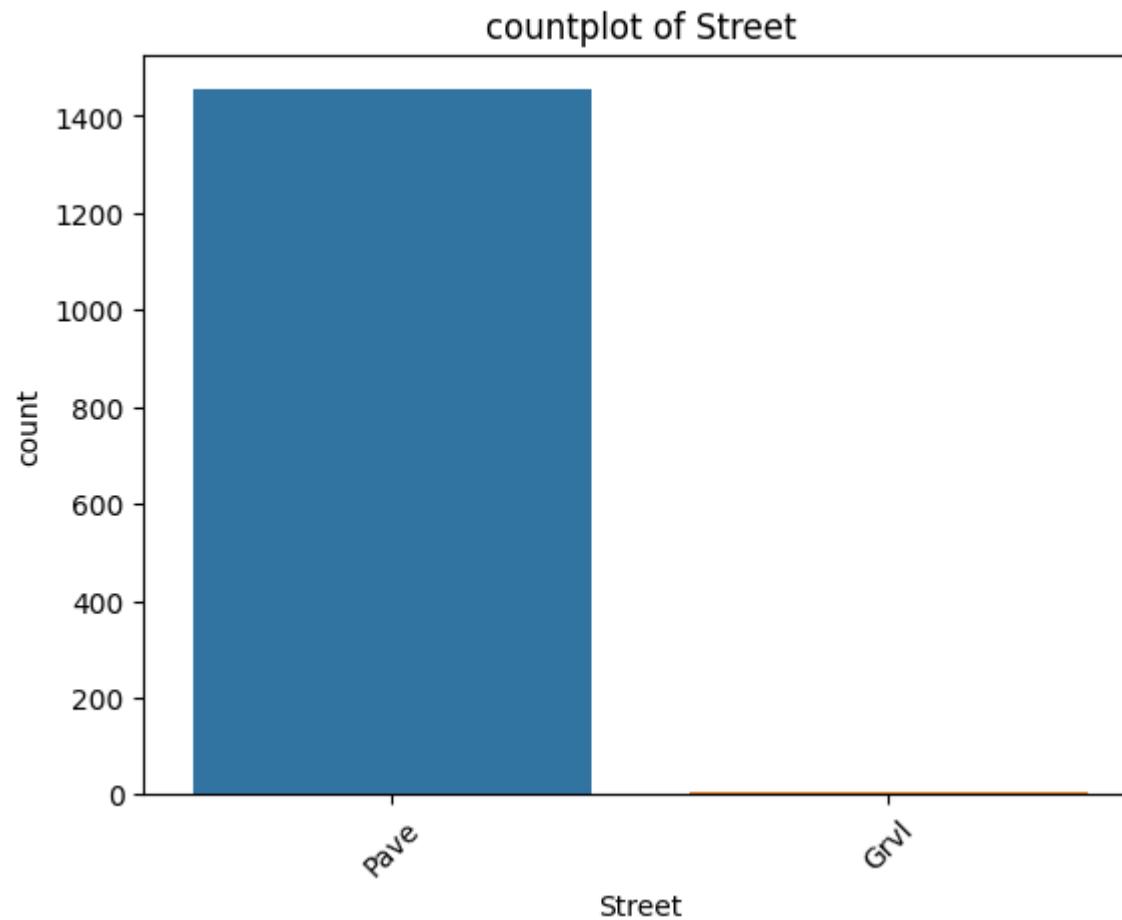


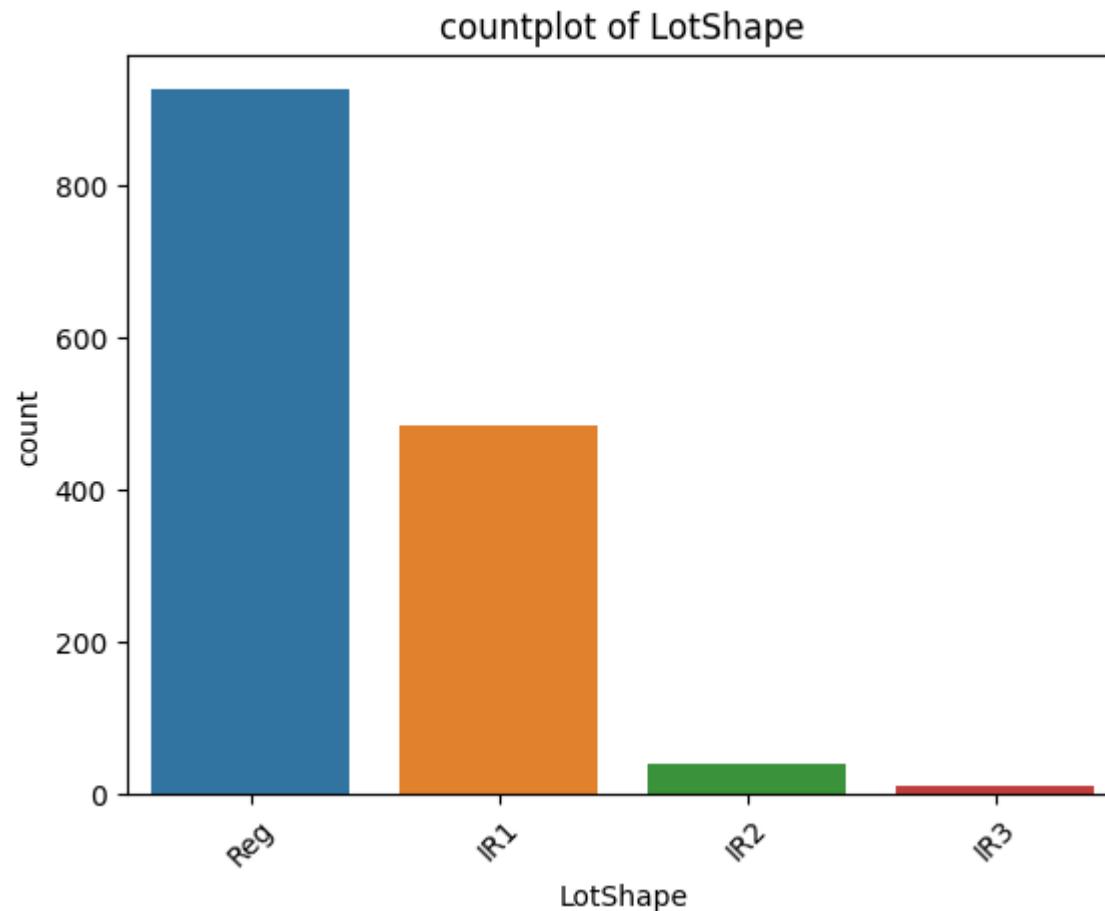
1.mssubclass, the distribution is right skewed, there are outliers(3) present in upper range, the mean value is slightly higher than median value

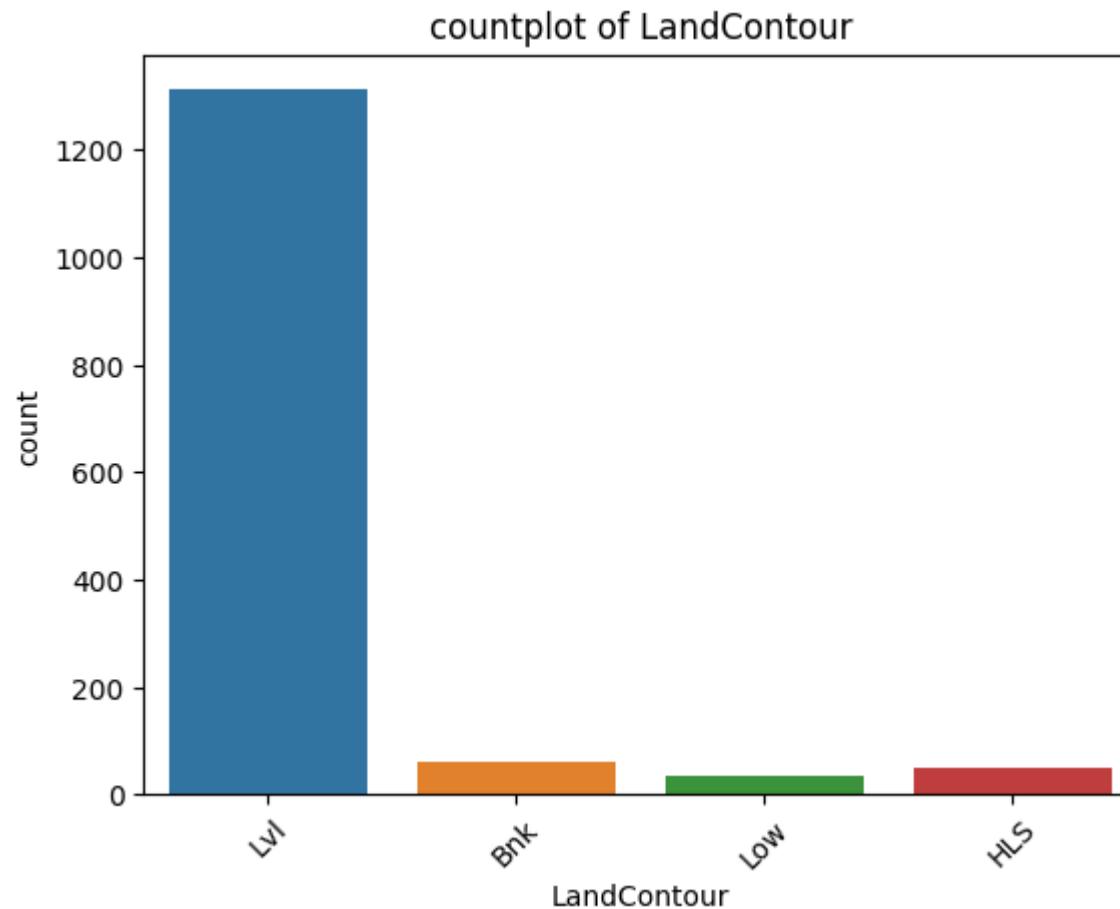
```
In [18]: # do it for all numerical cols and for categorical cols too
```

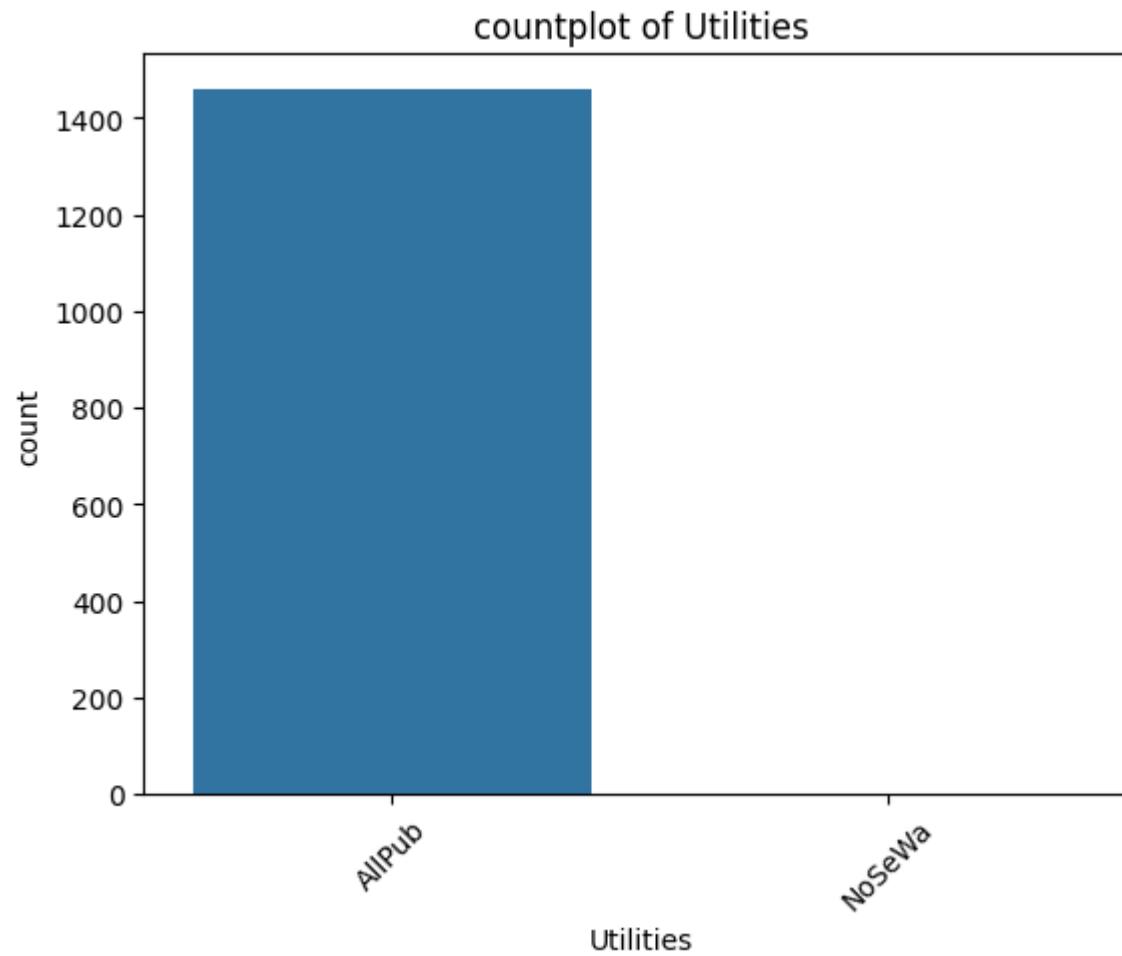
```
In [19]: for col in data.select_dtypes(include='object'):
    sns.countplot(x=data[col])
    plt.title(f'countplot of {col}')
    plt.xticks(rotation=45)
    plt.show()
```

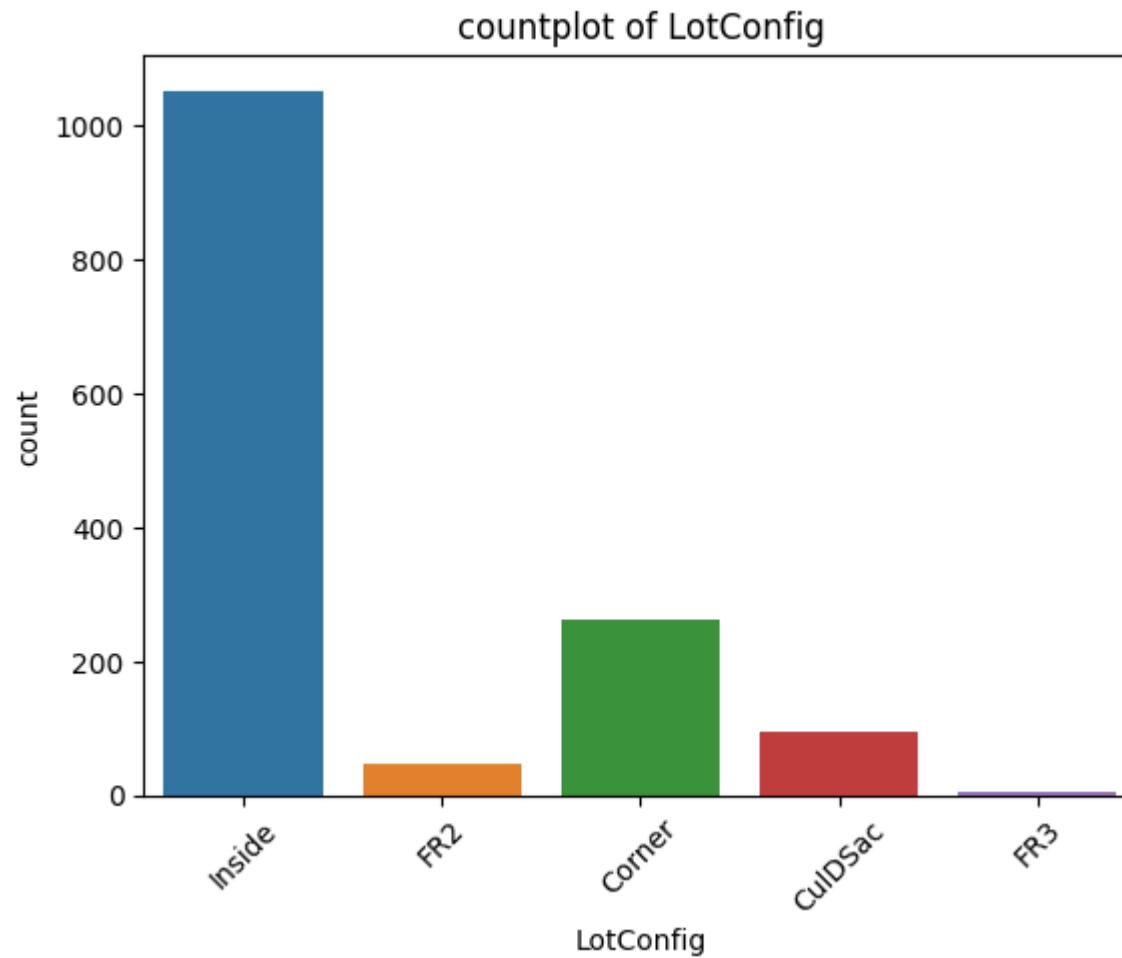


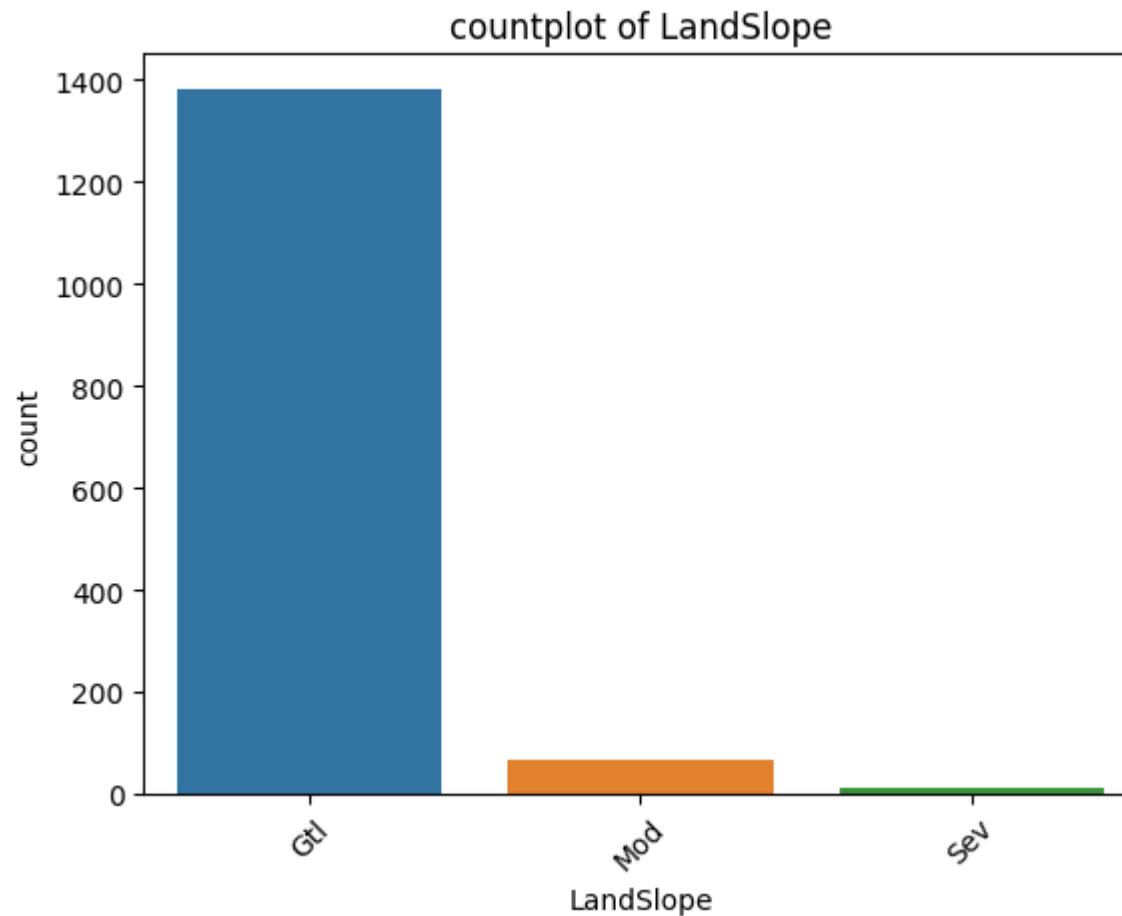




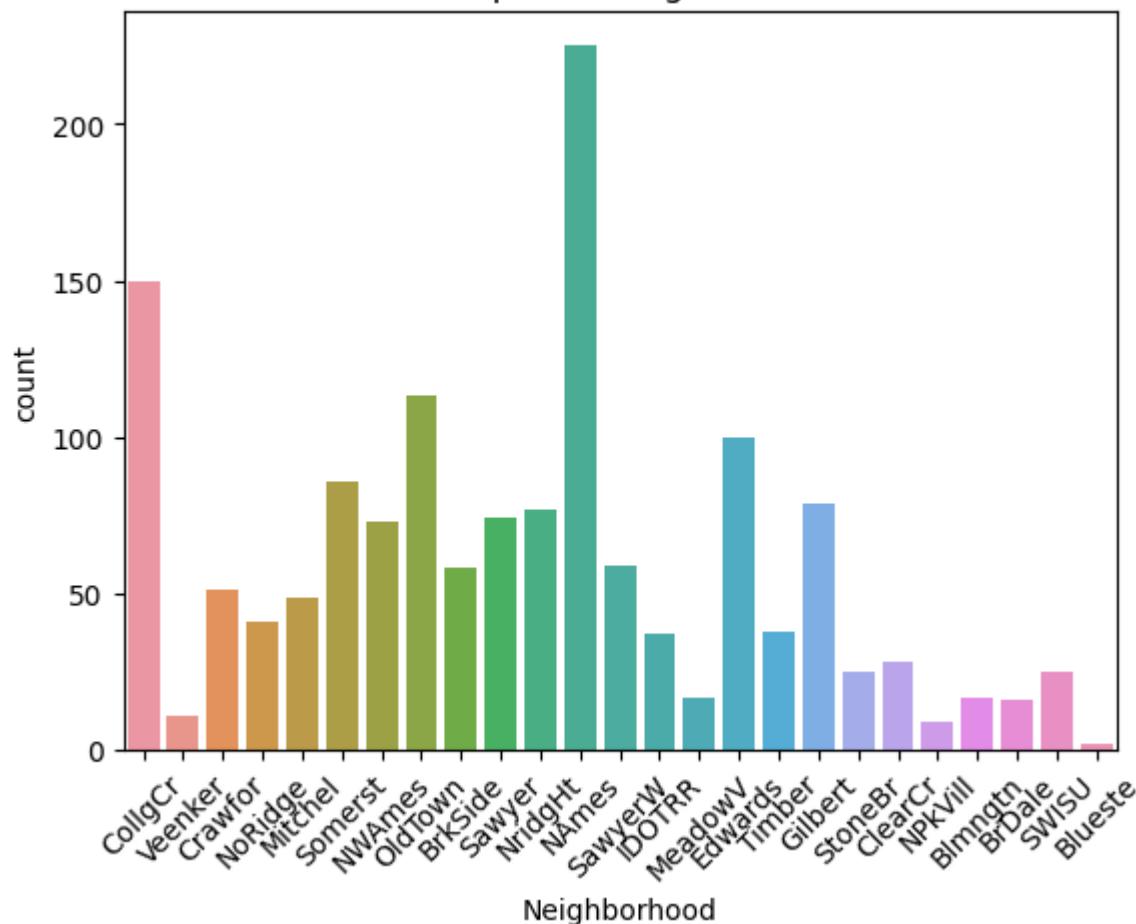


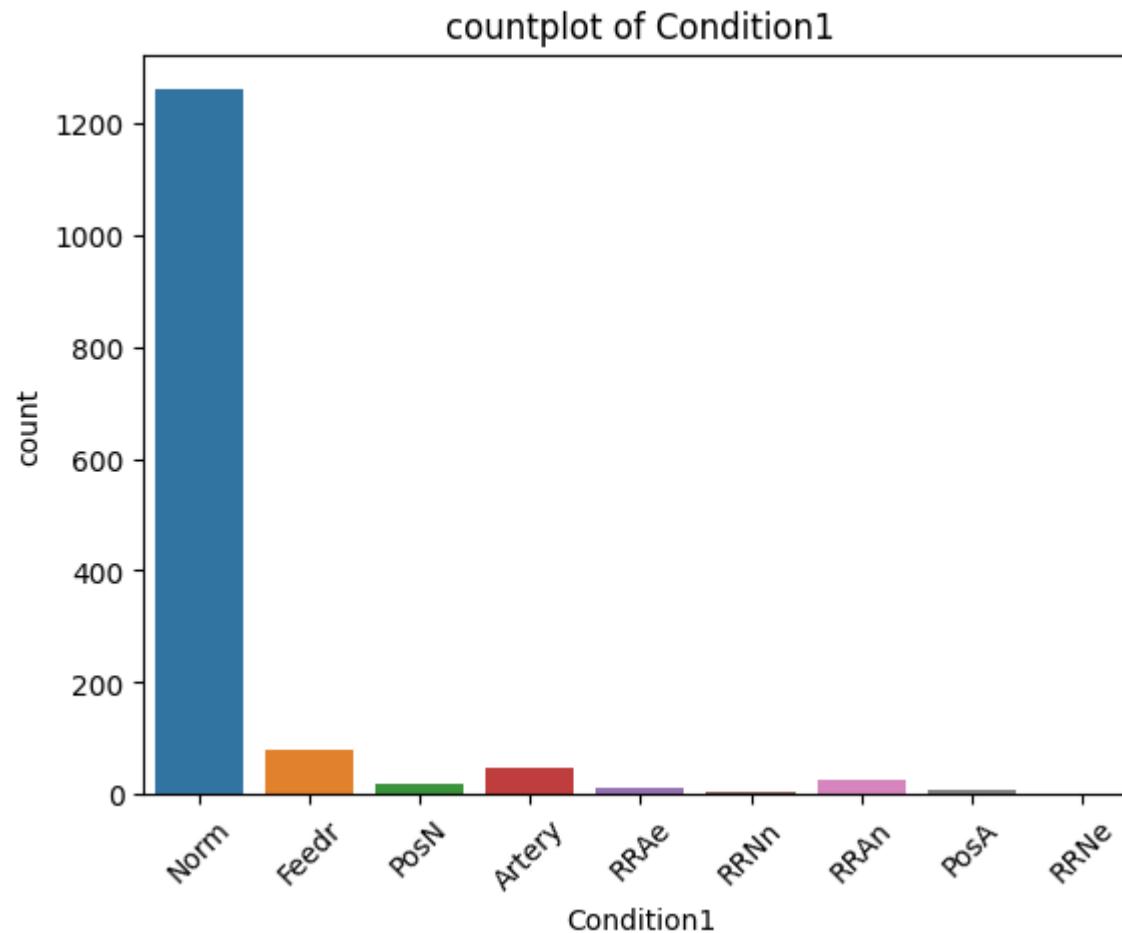


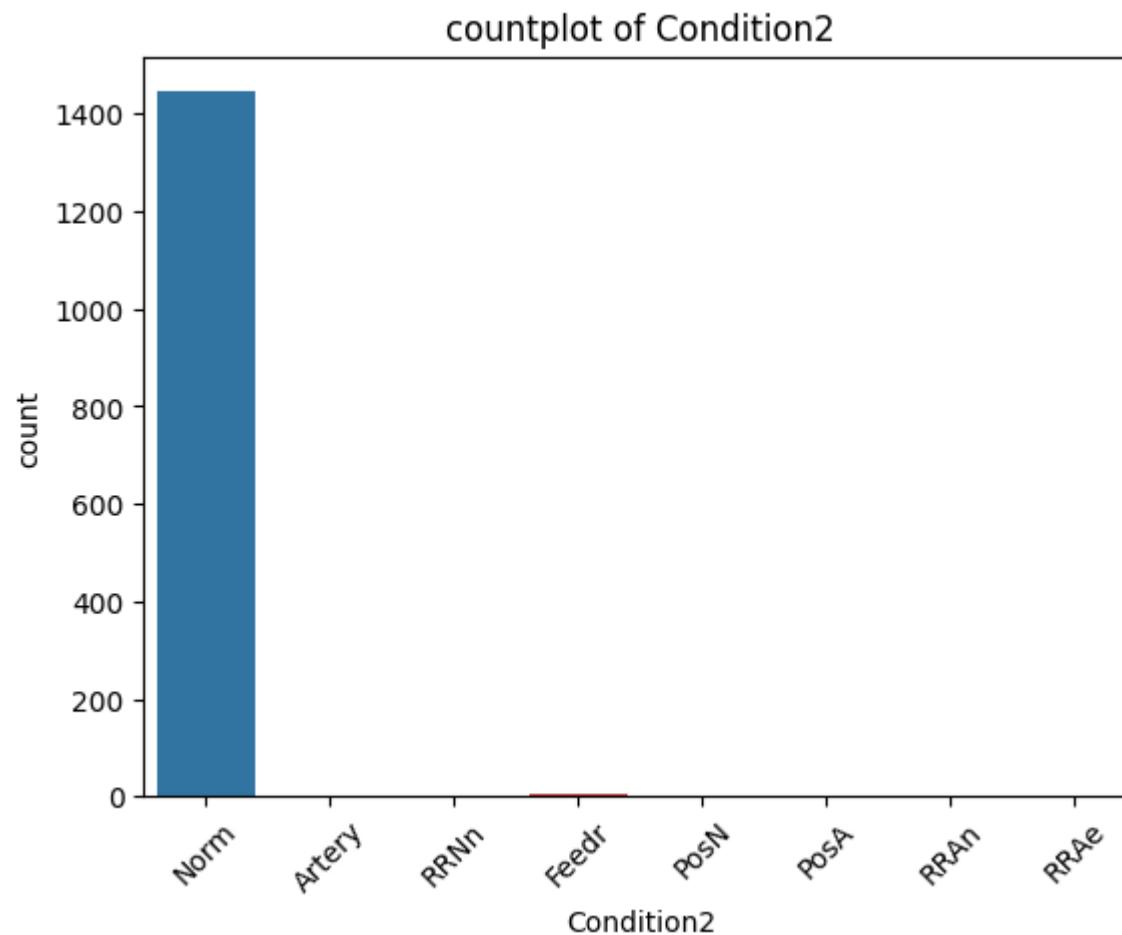


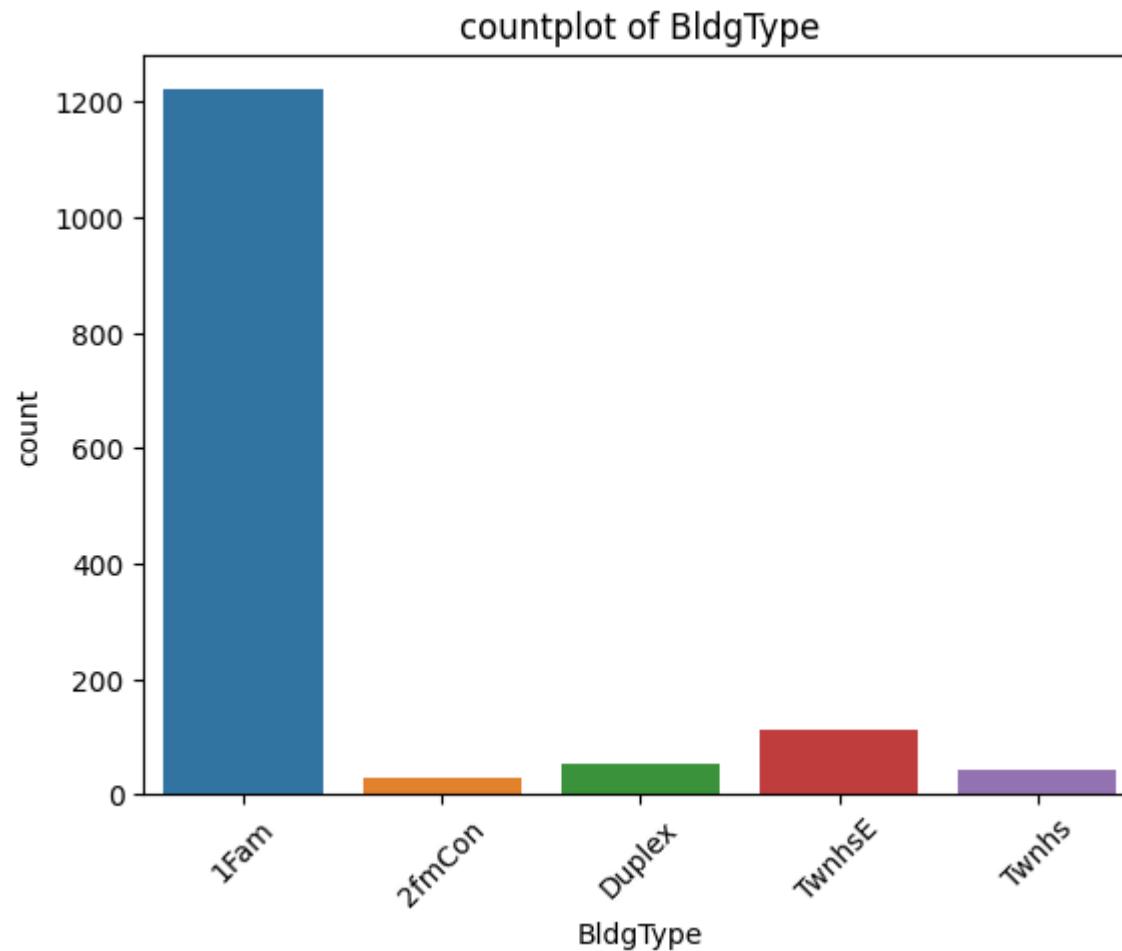


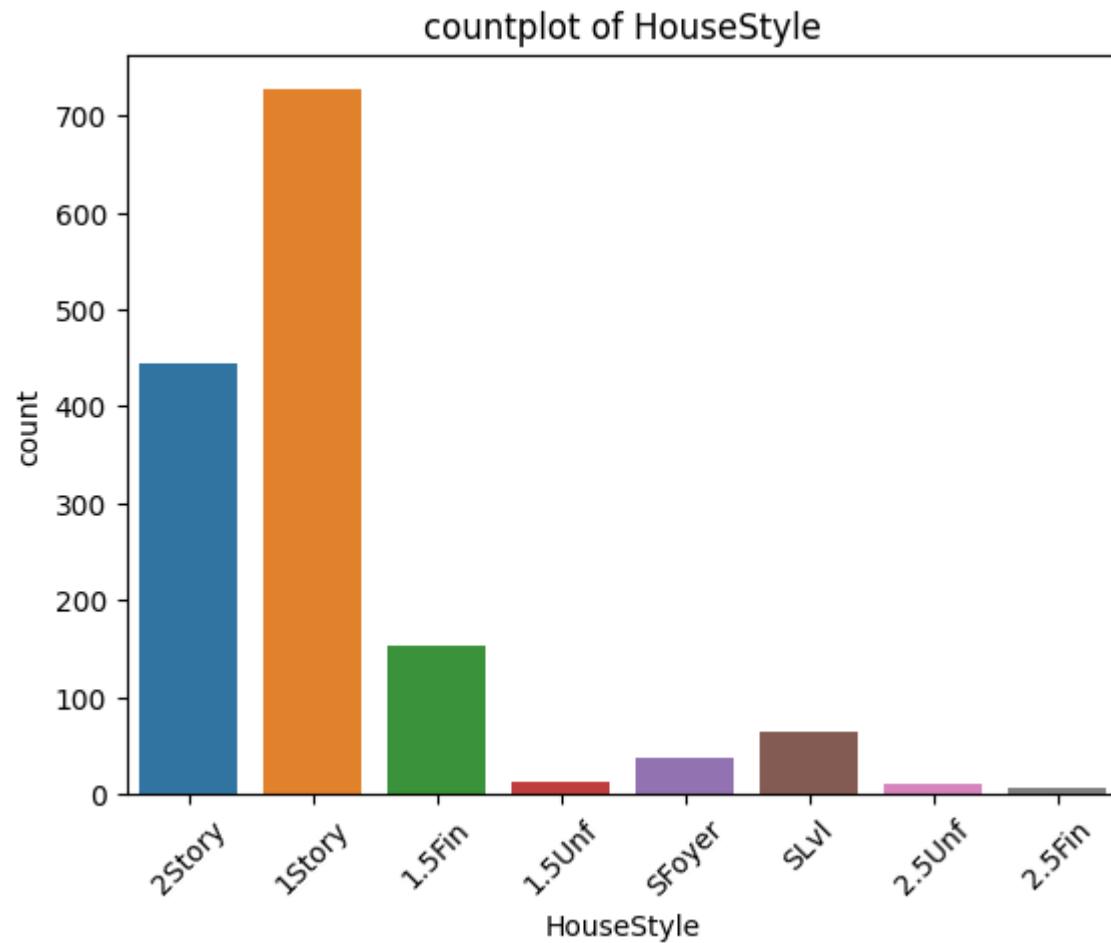
countplot of Neighborhood



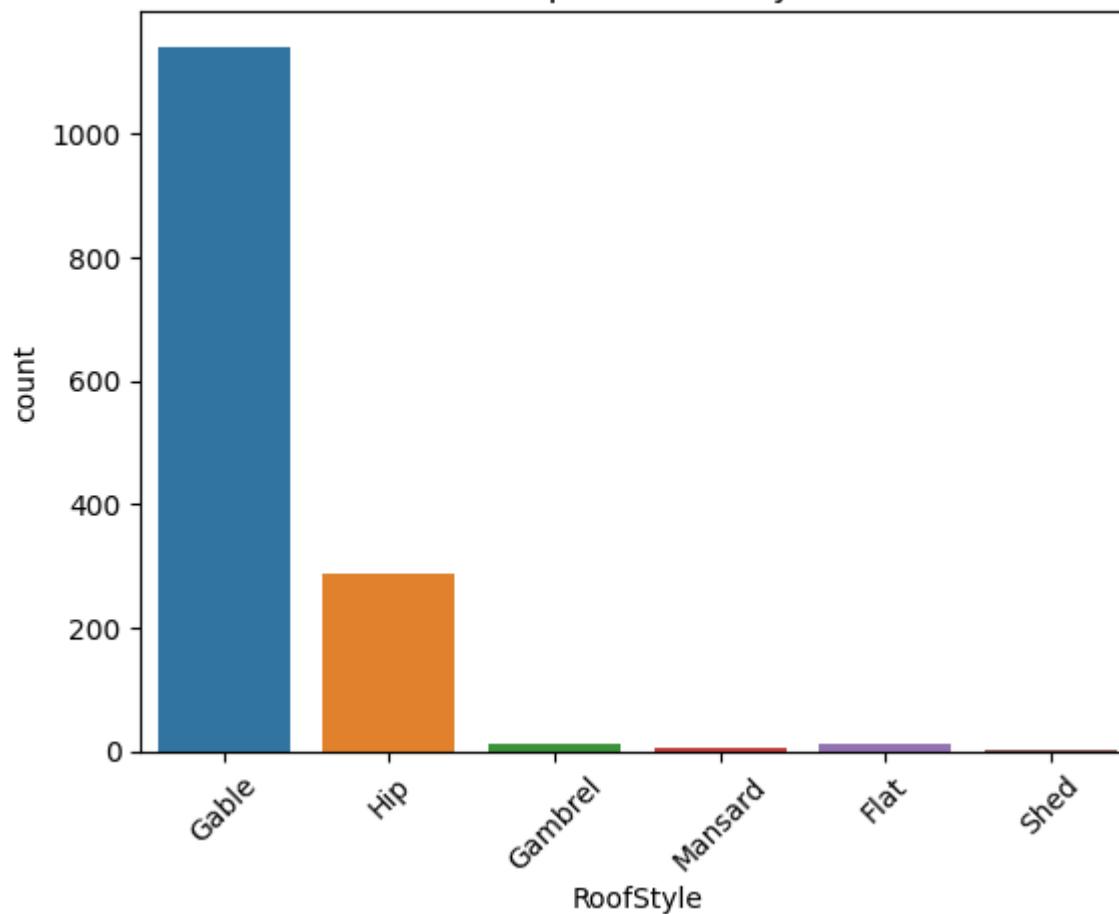


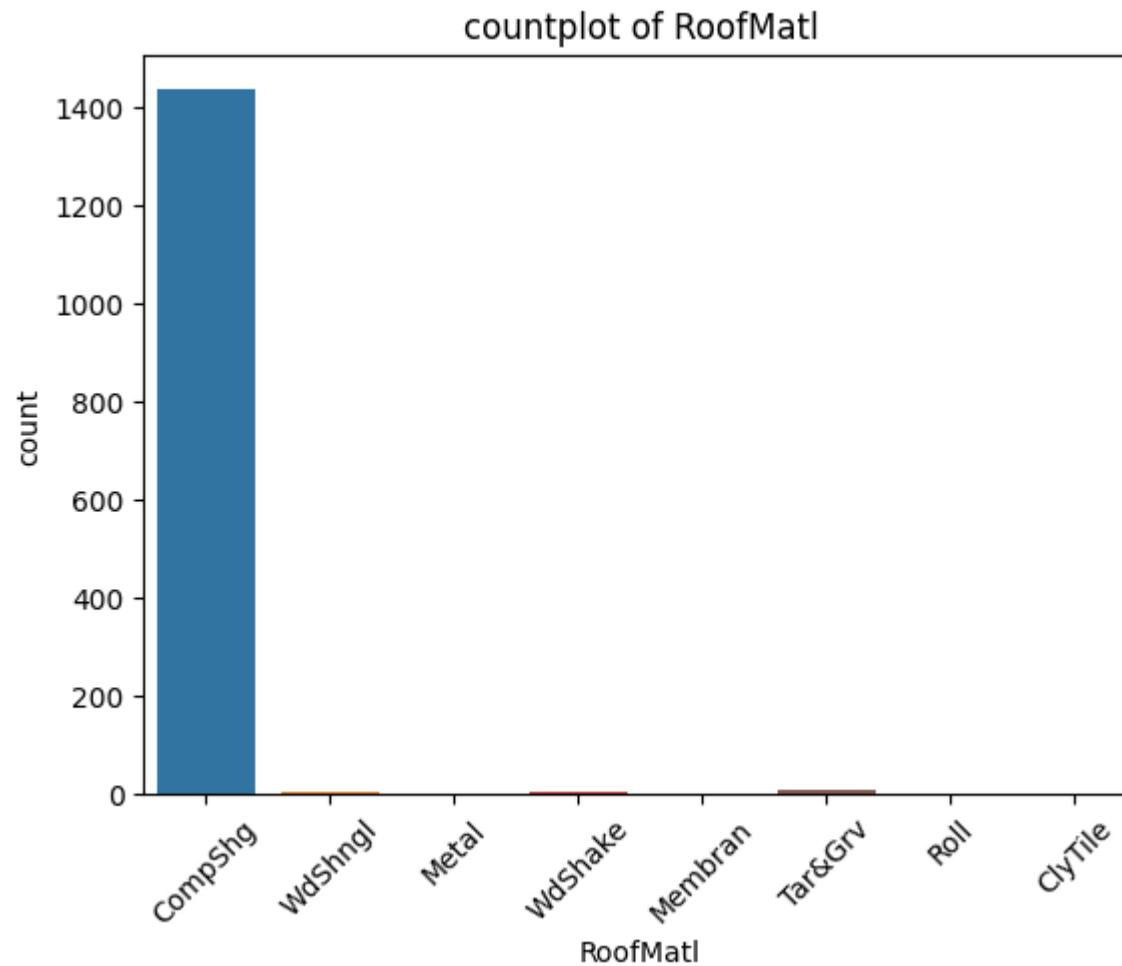


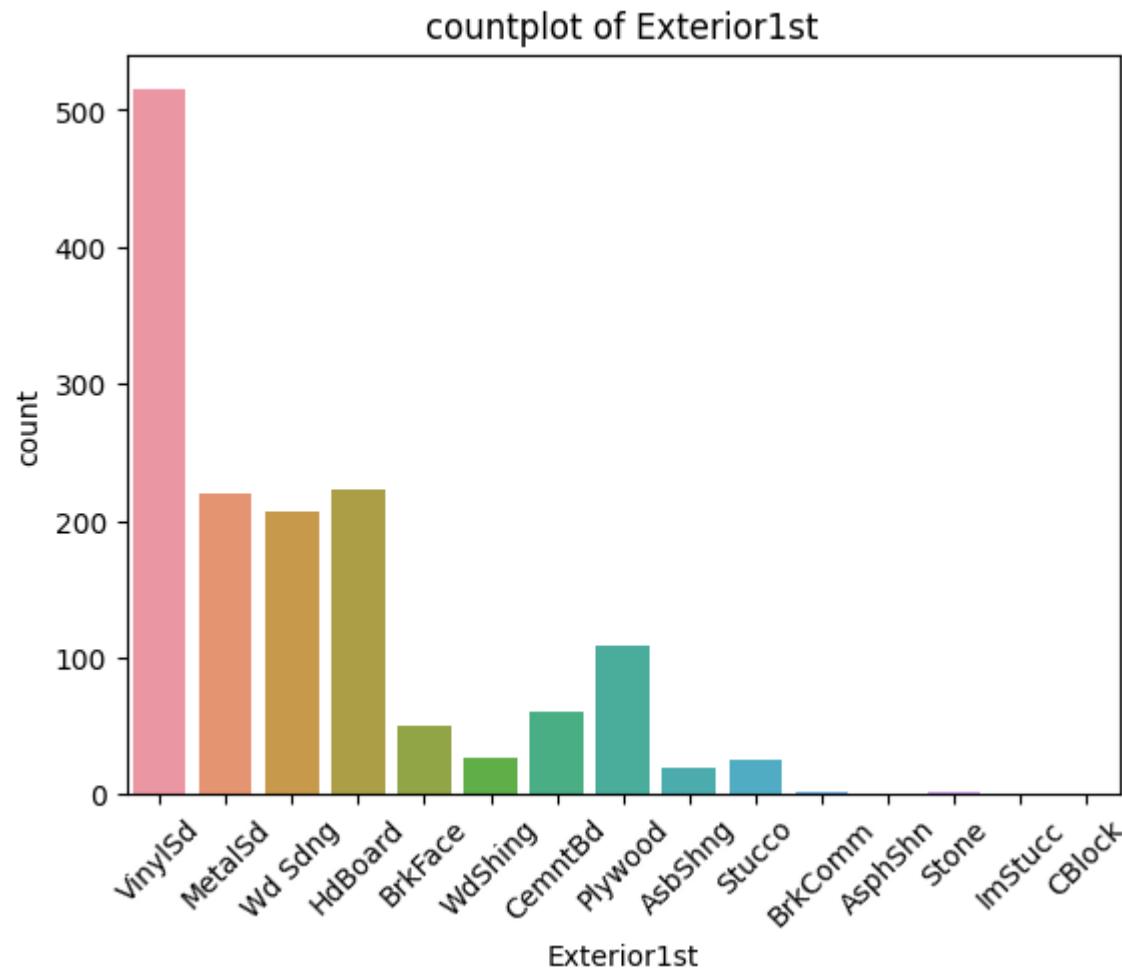


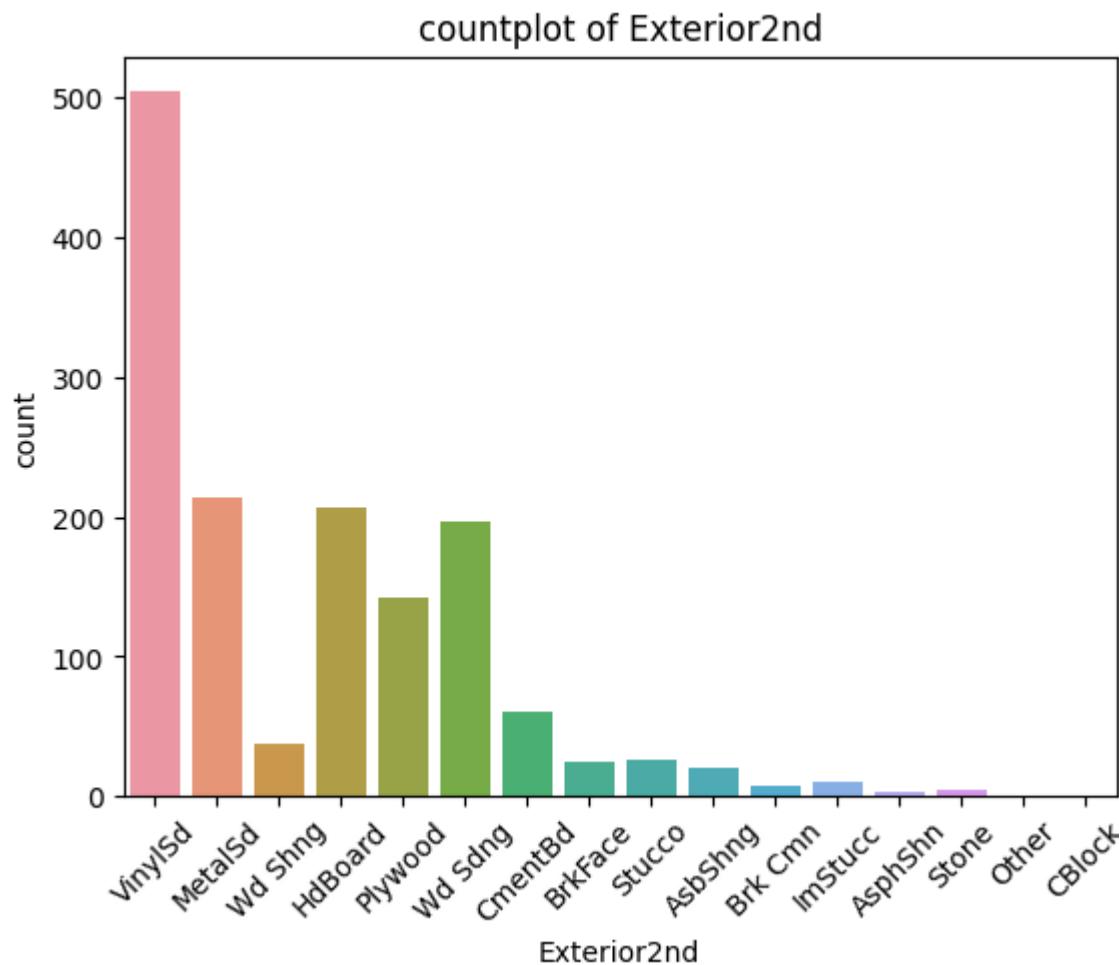


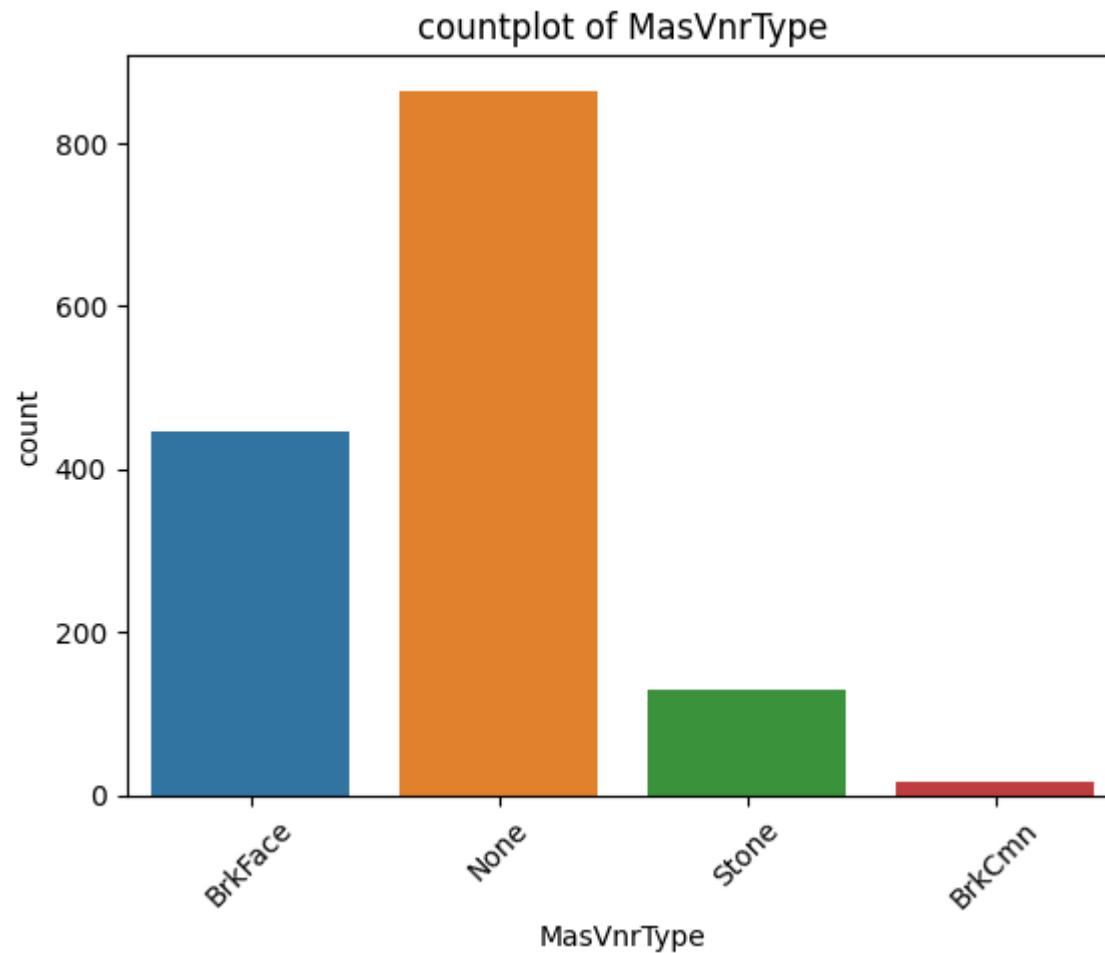
countplot of RoofStyle

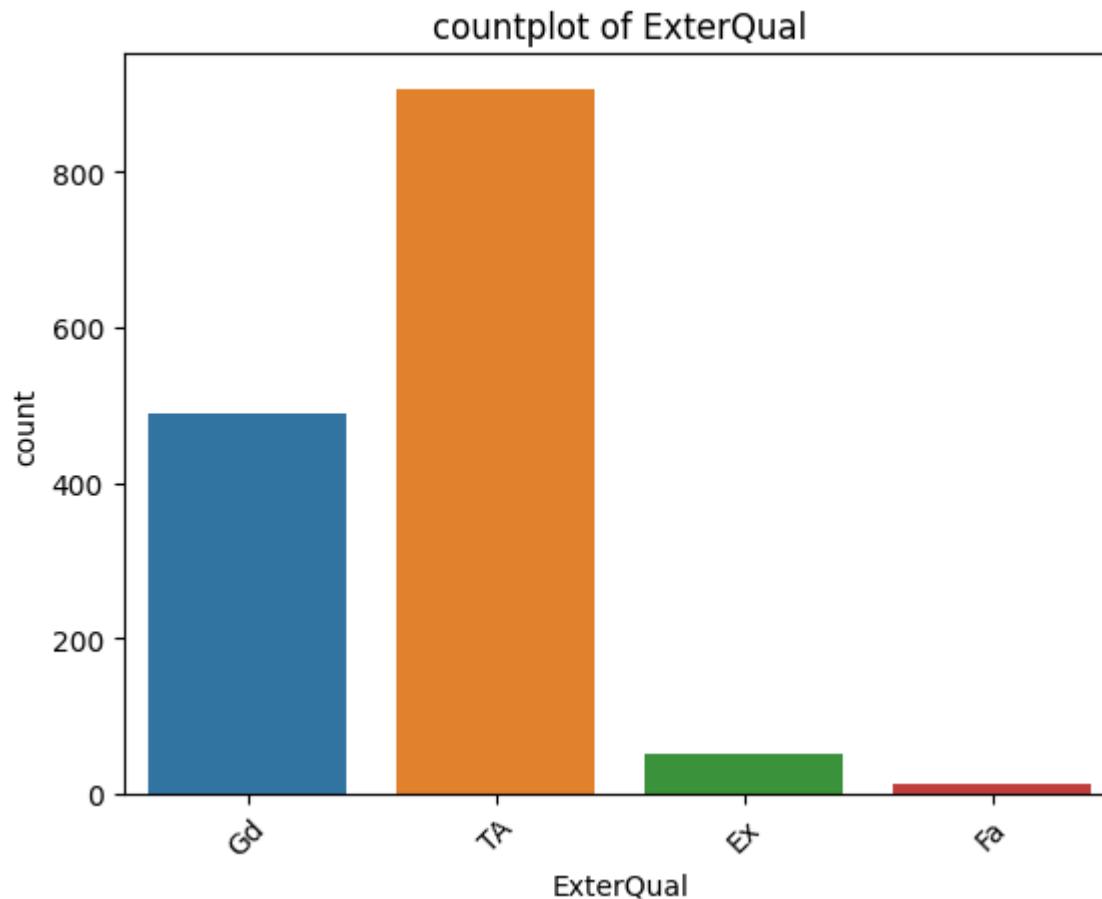


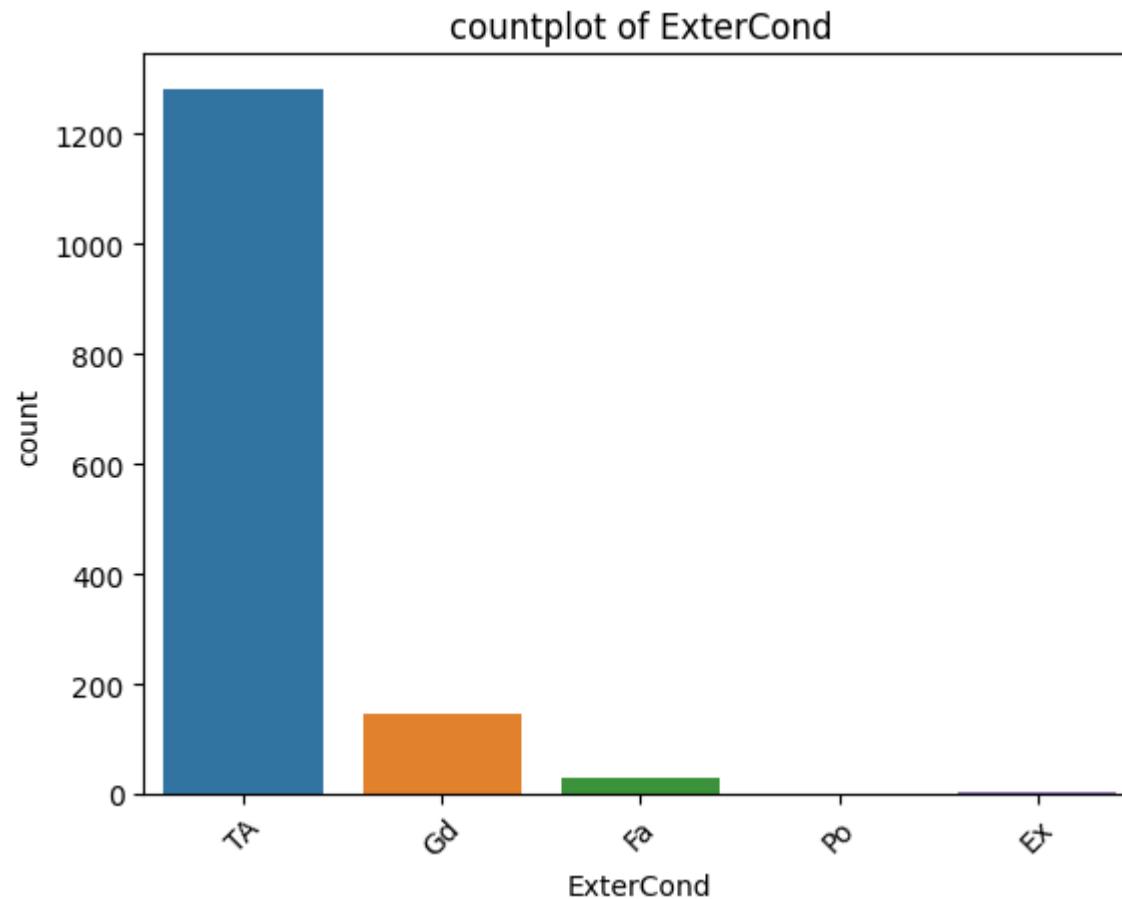


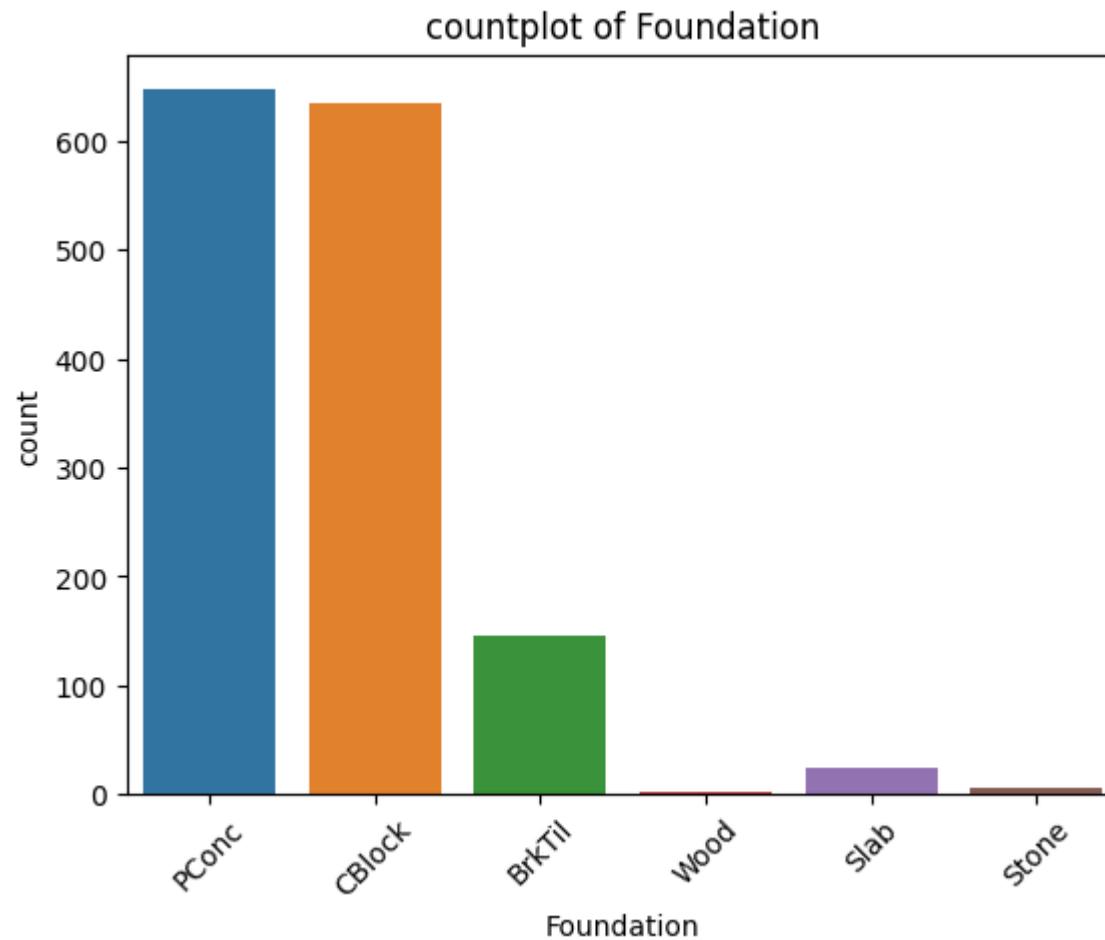


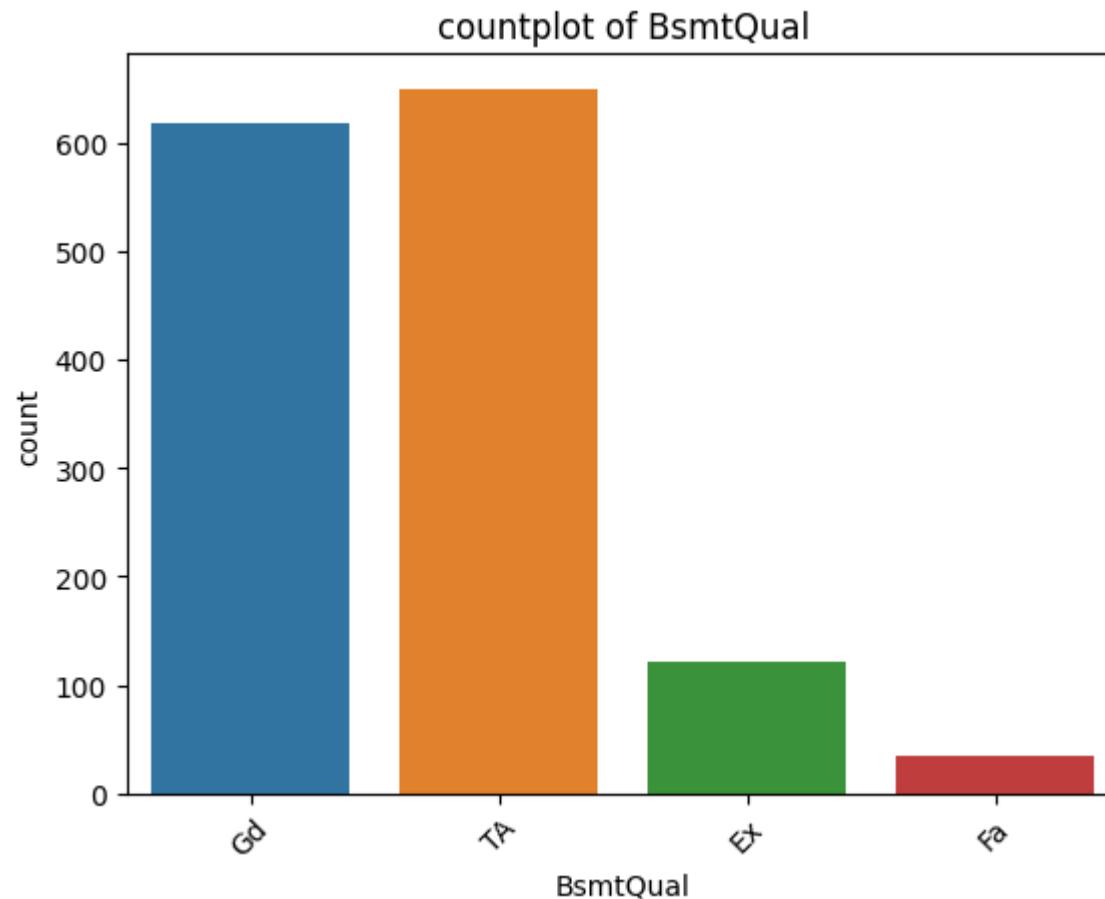


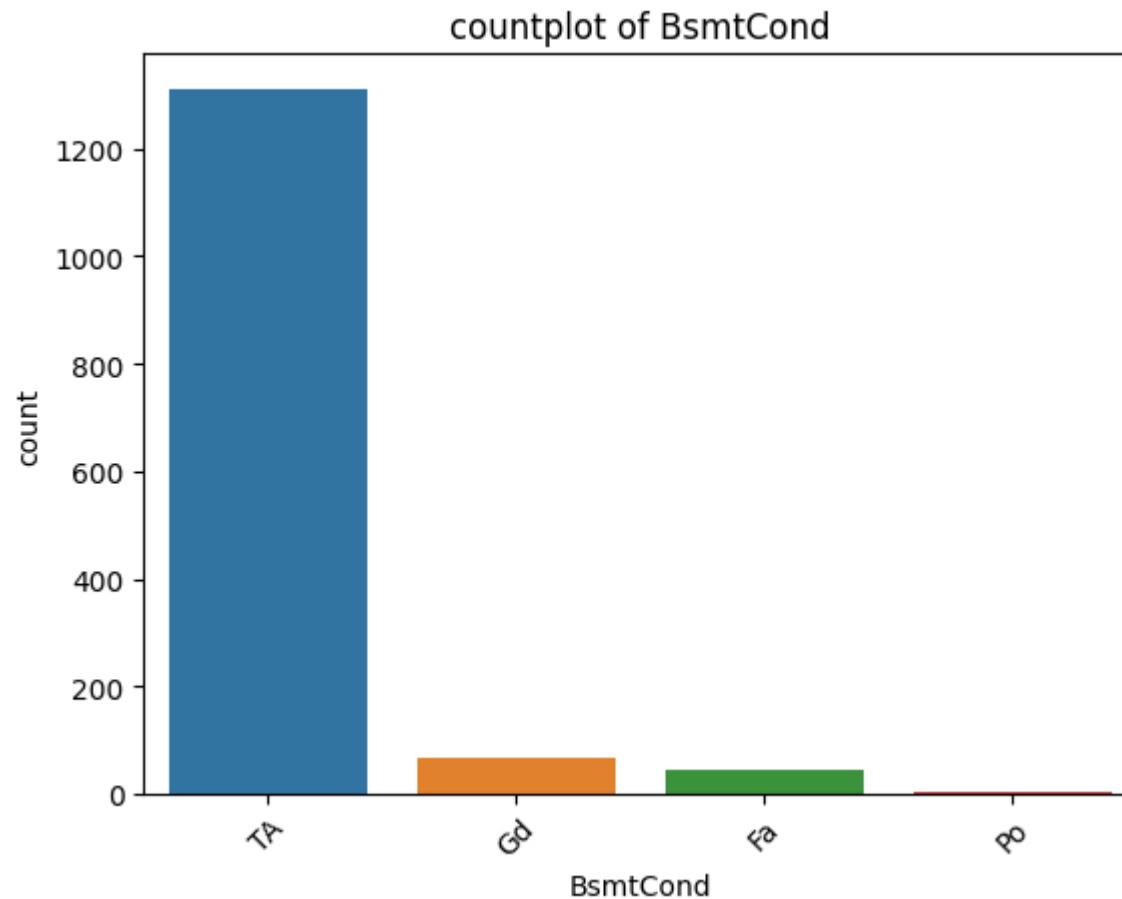


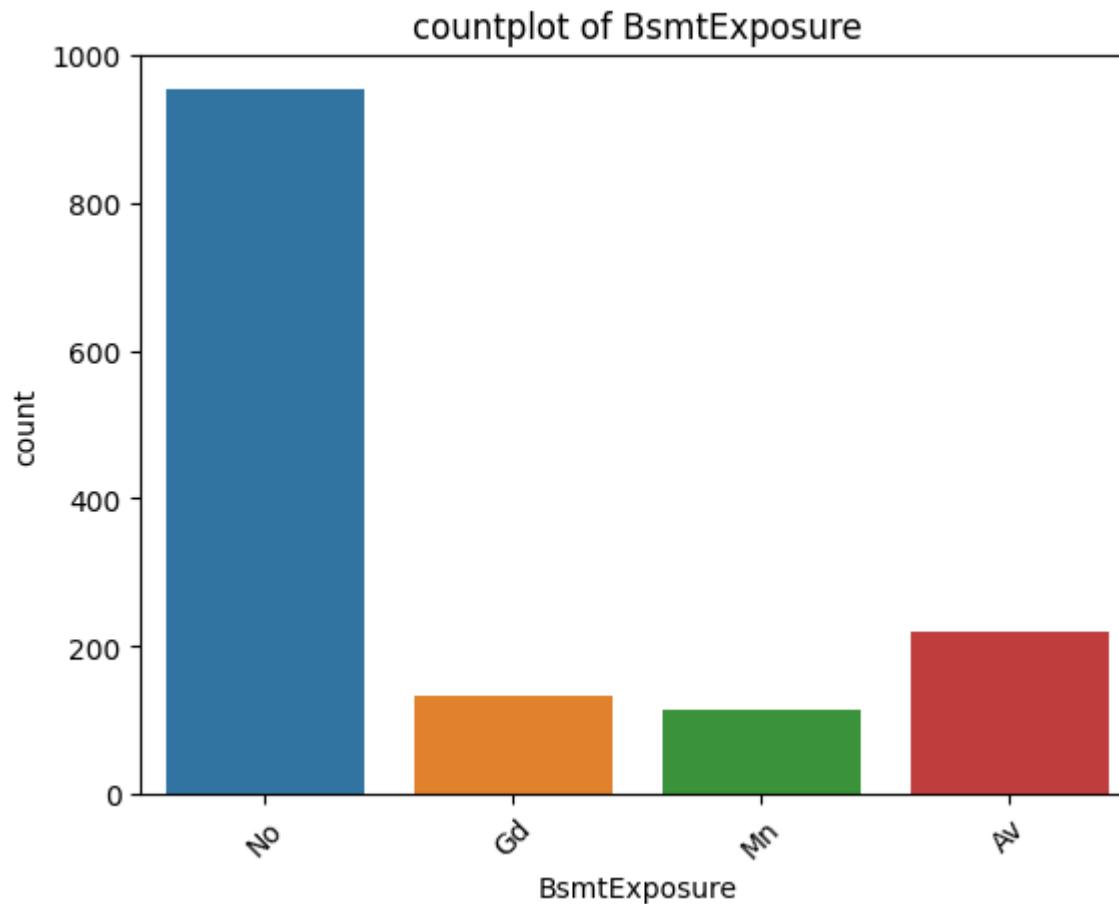


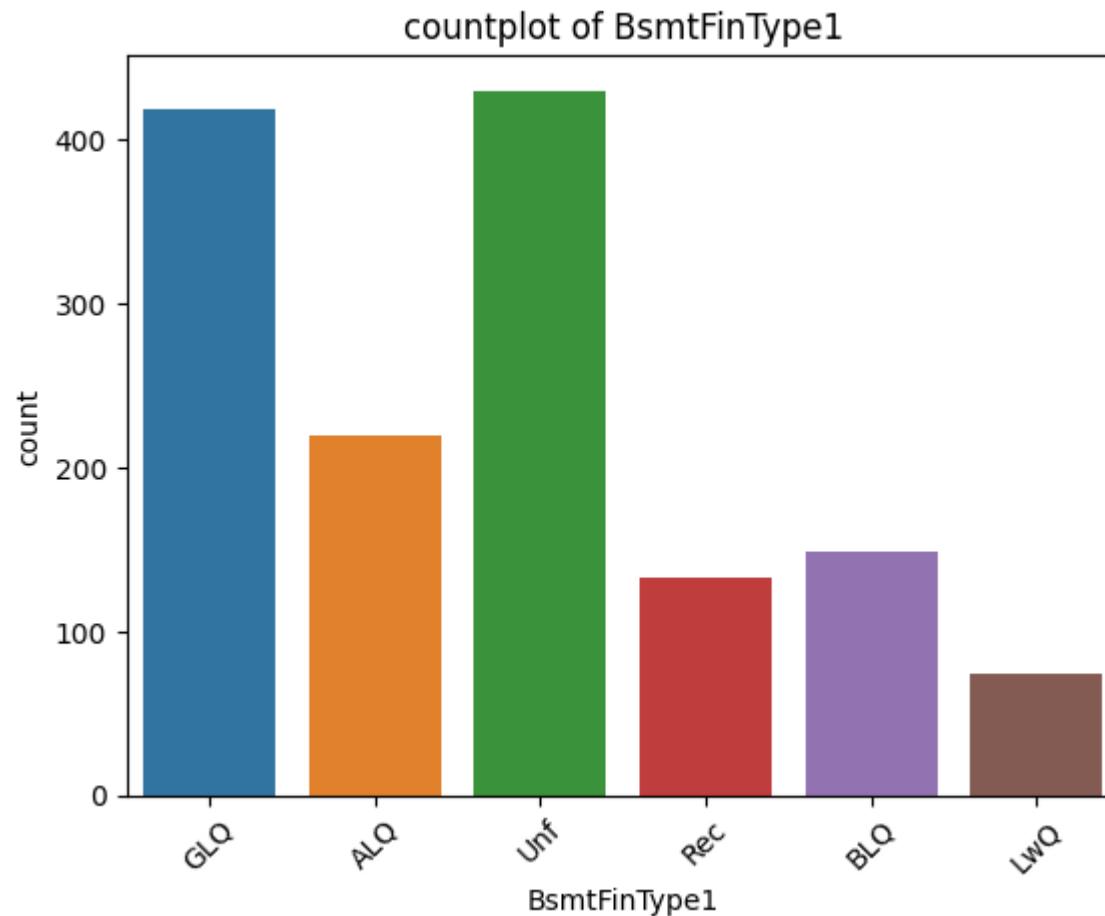


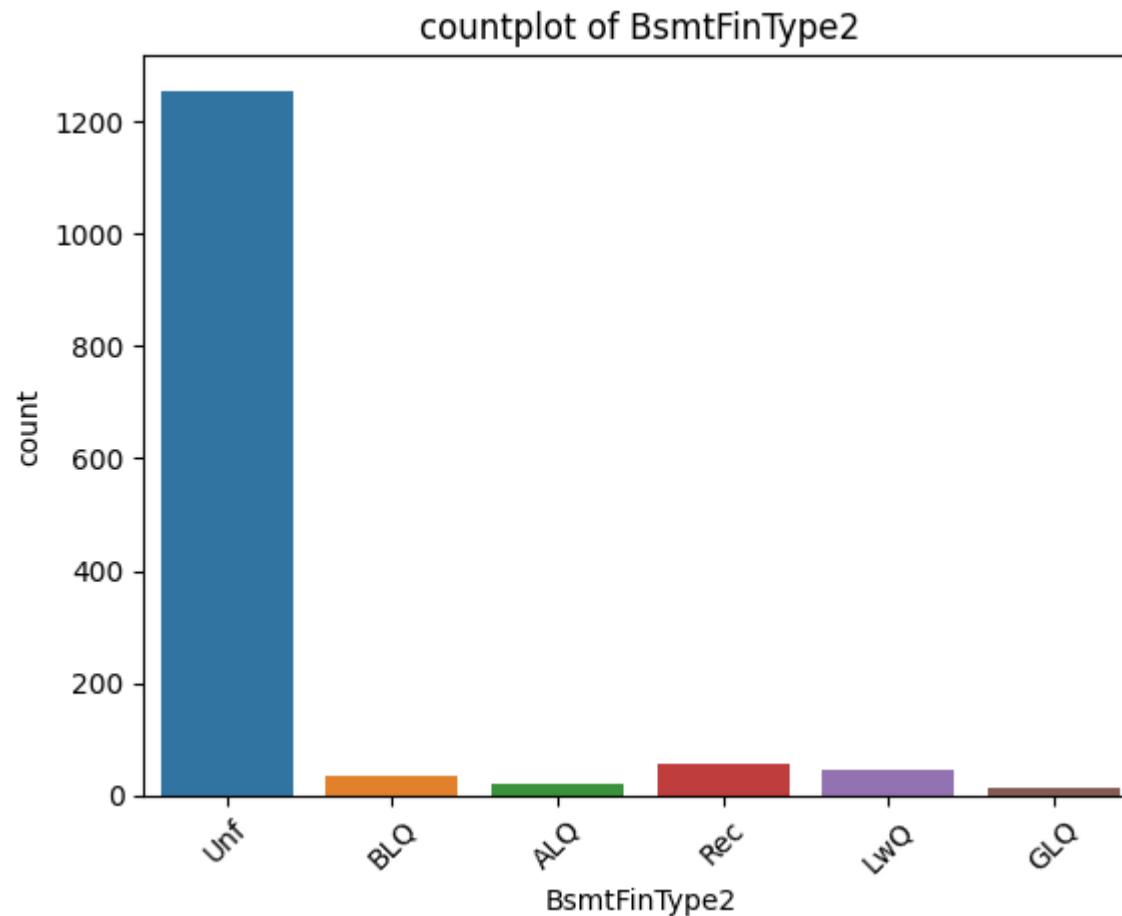


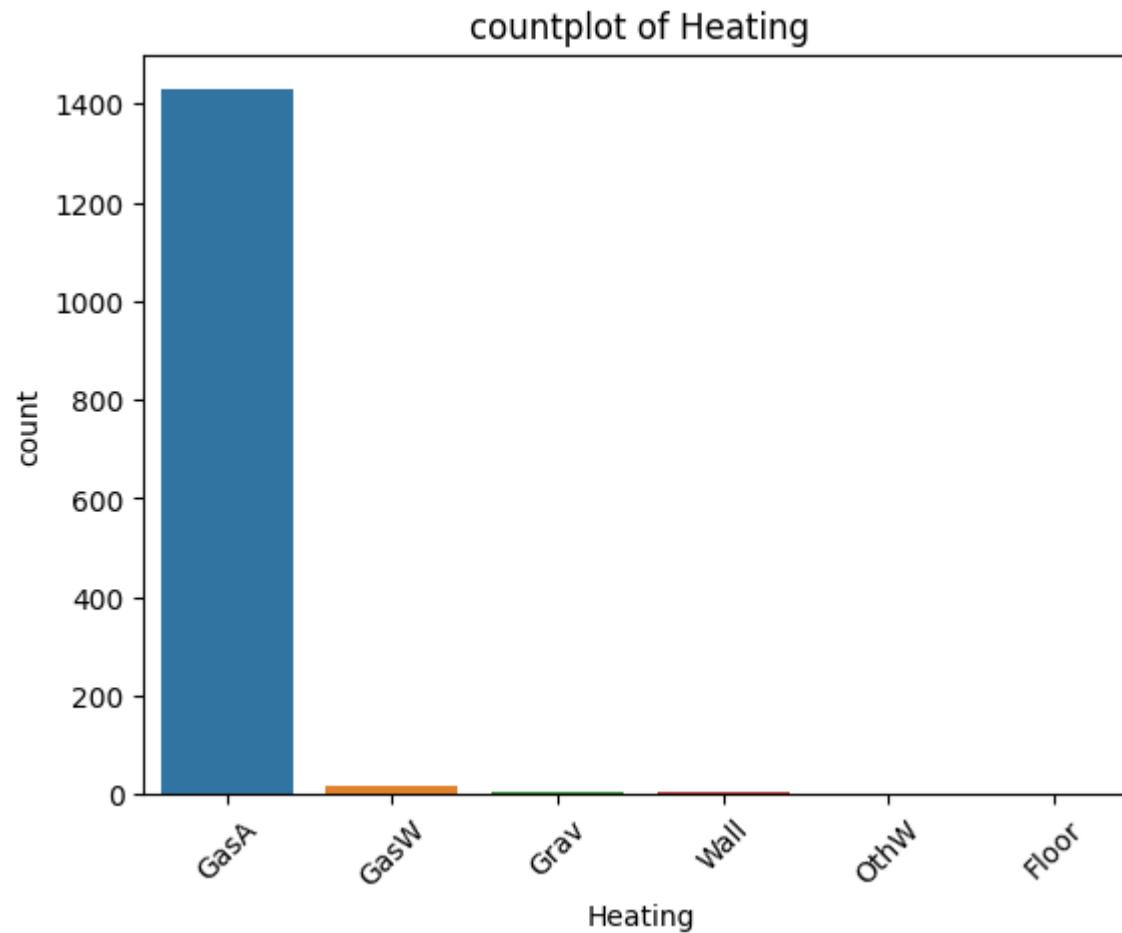


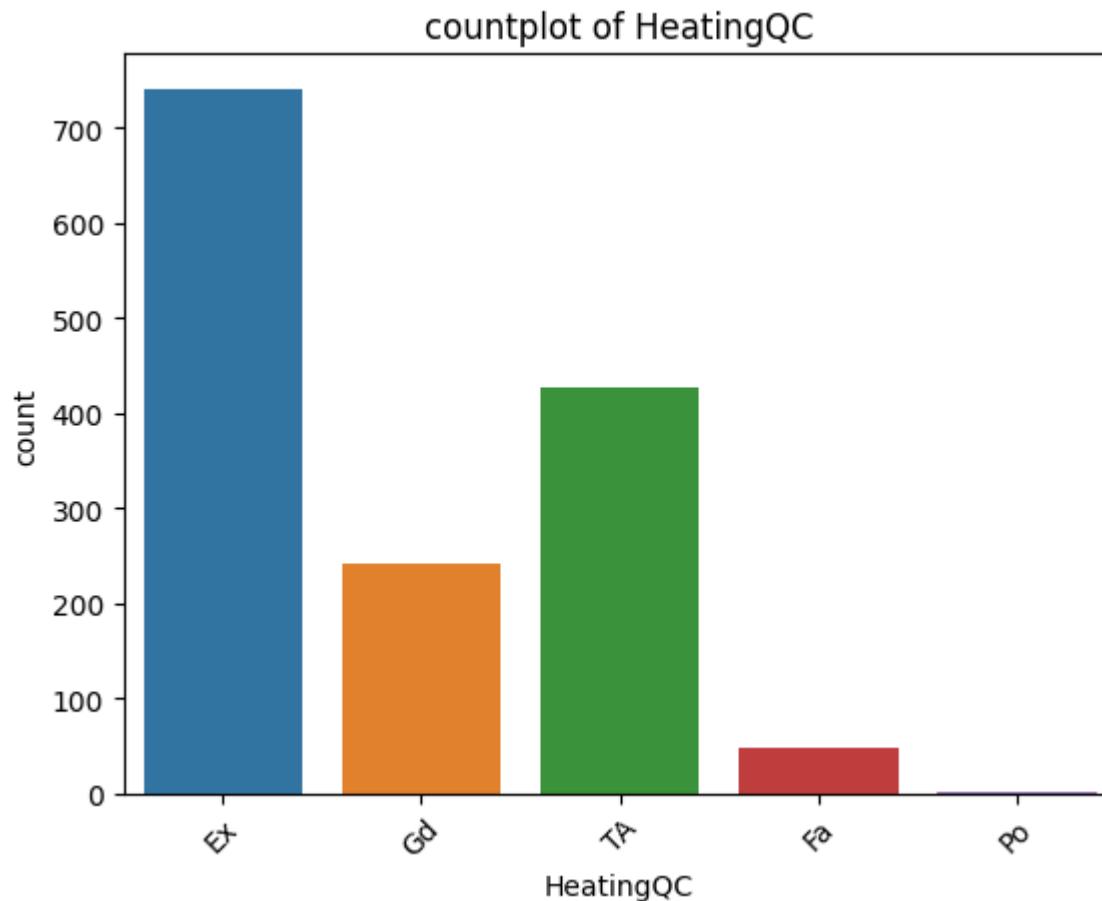


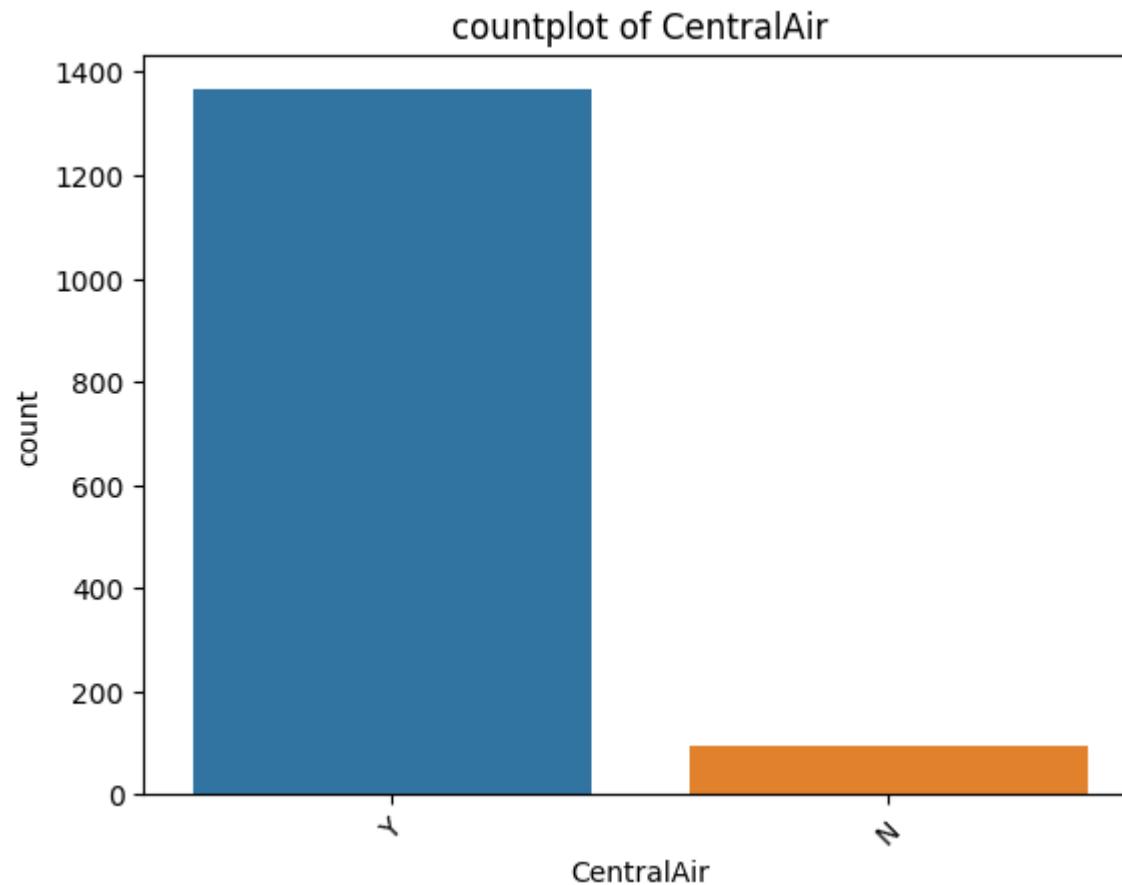


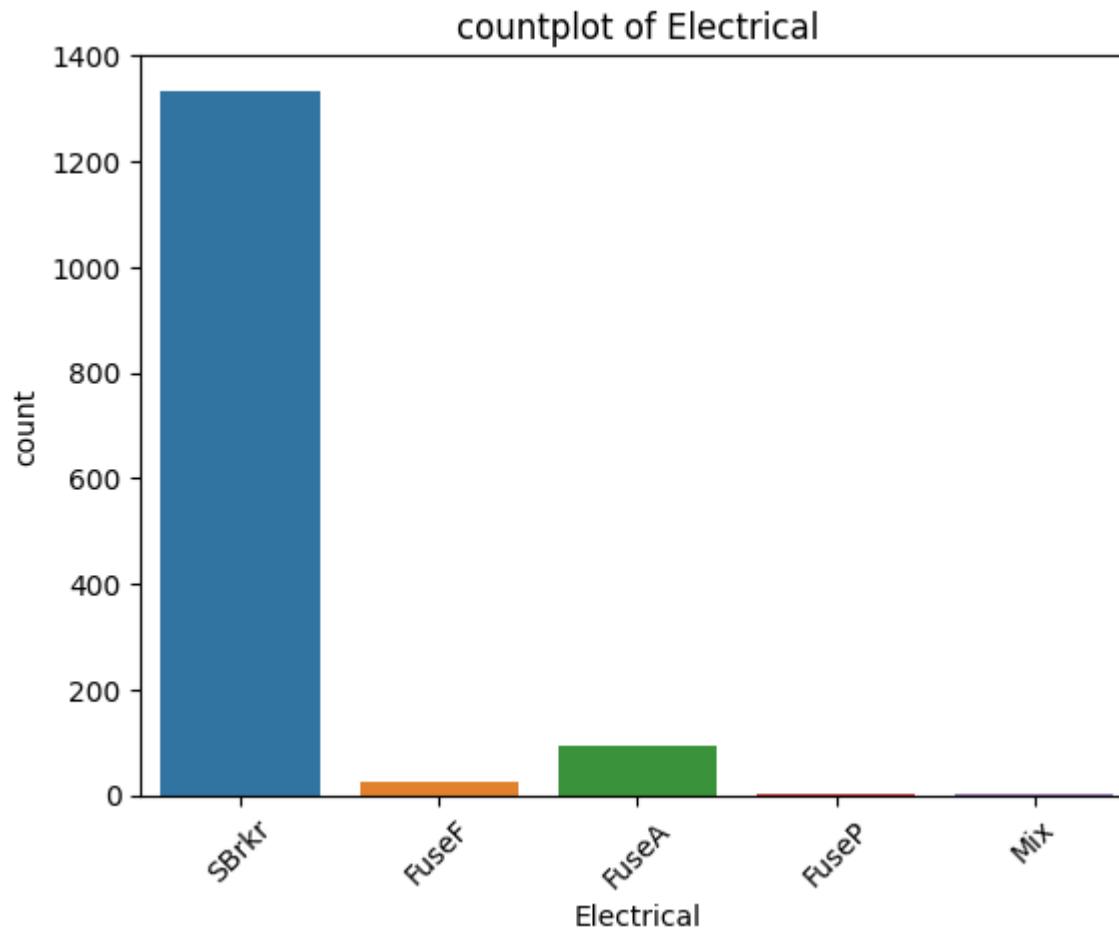


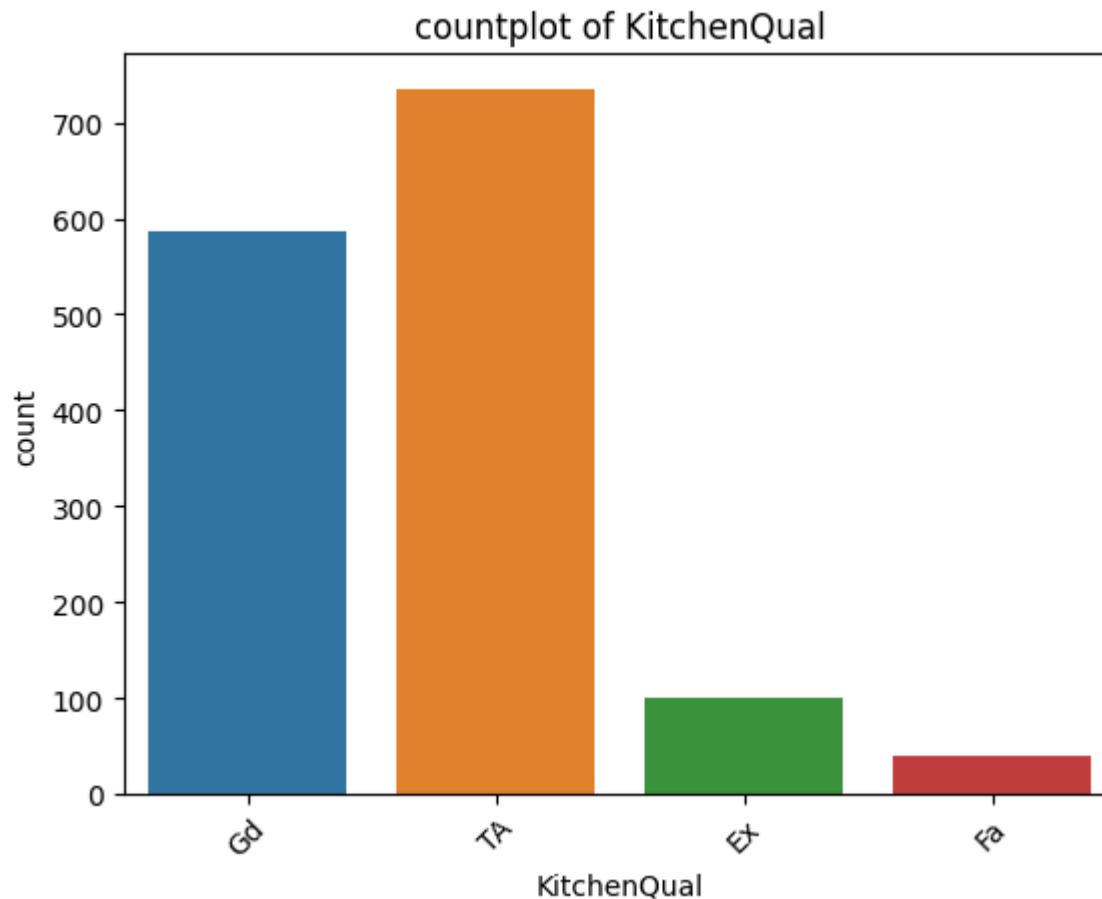


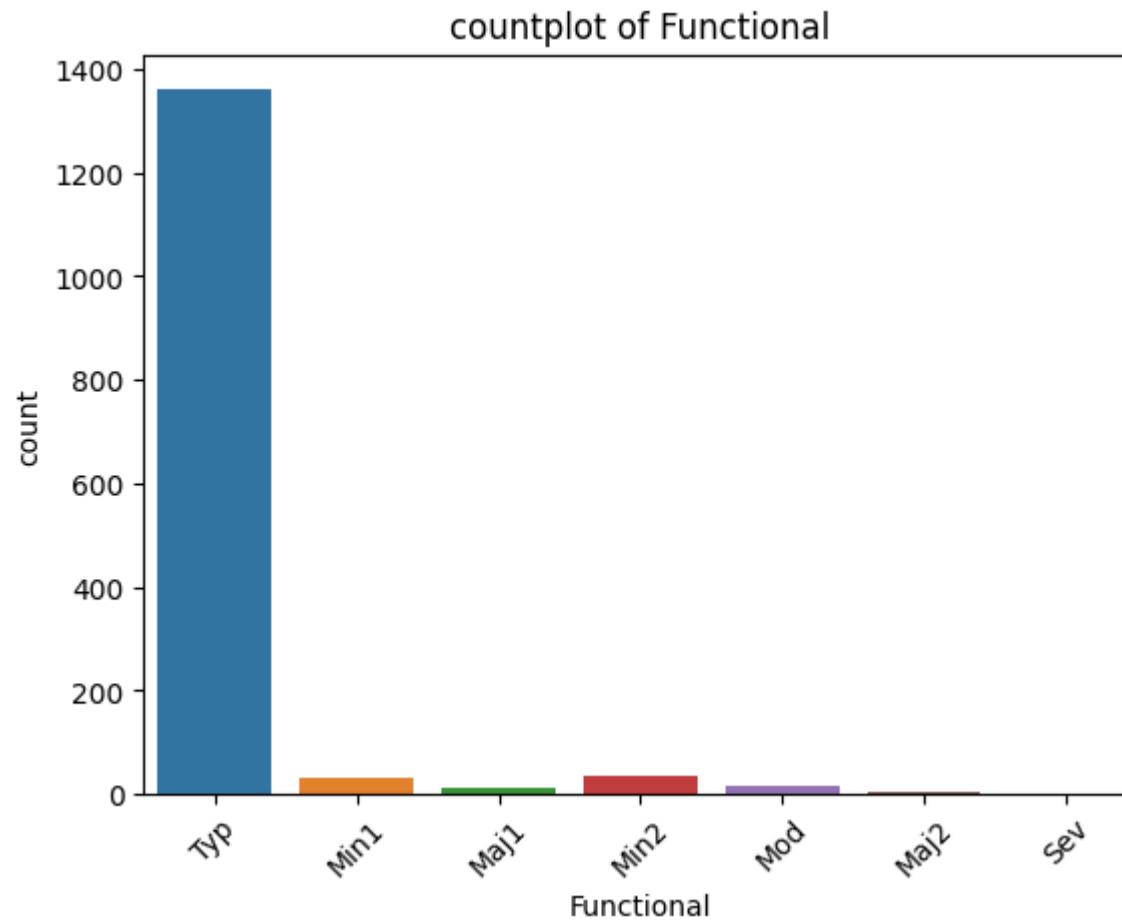


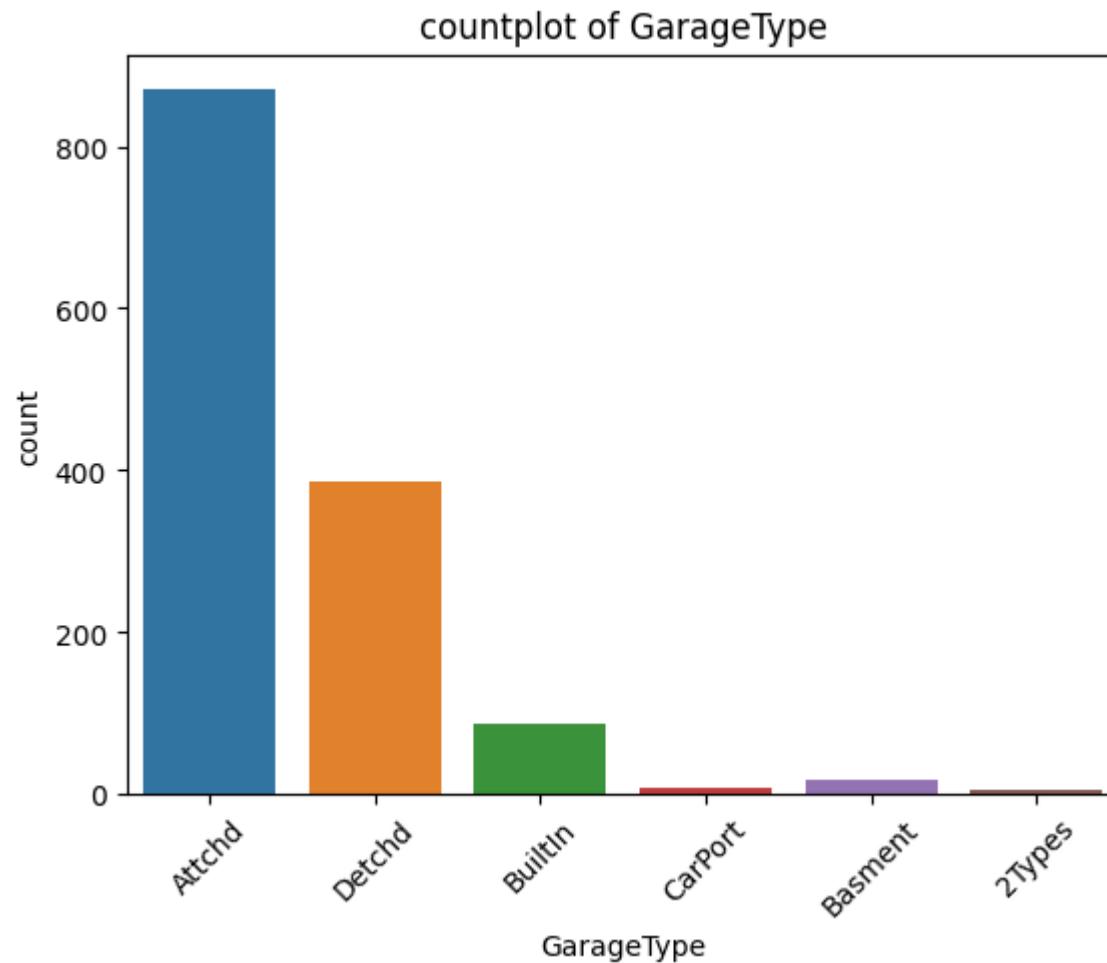


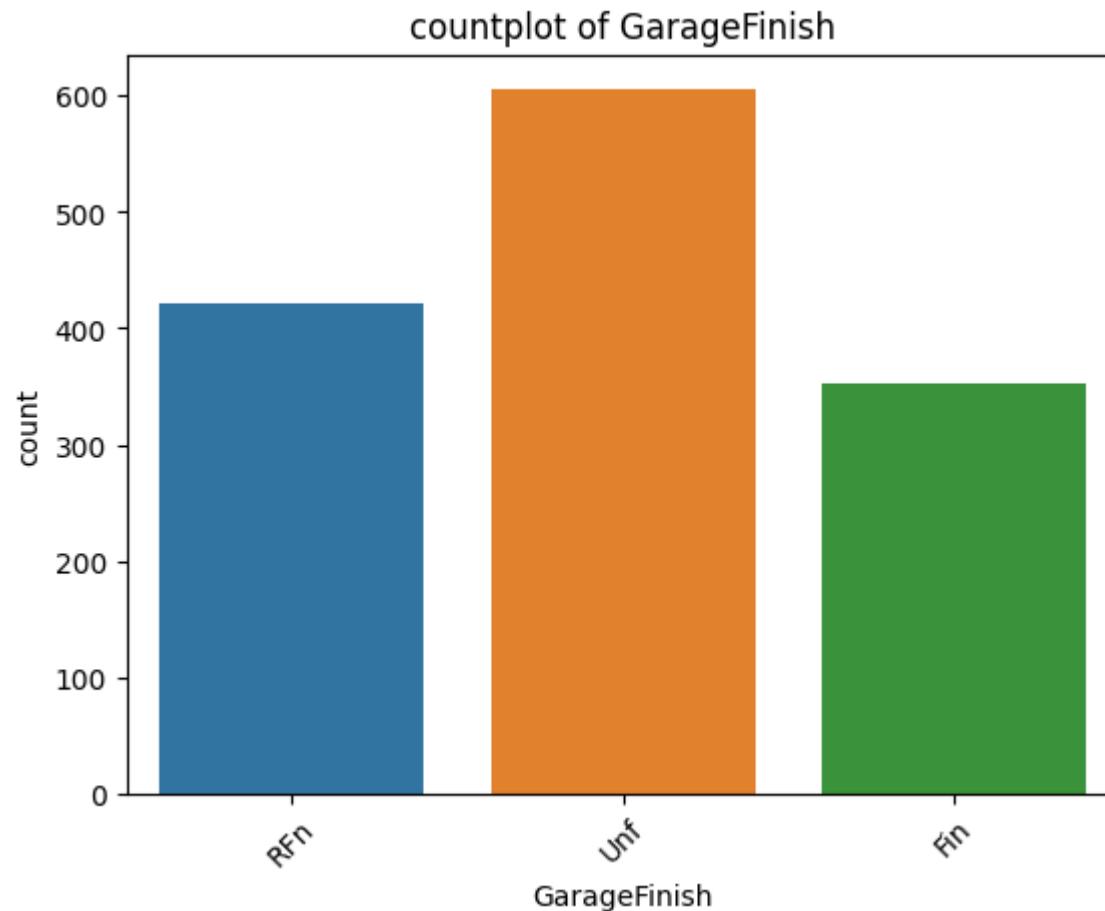


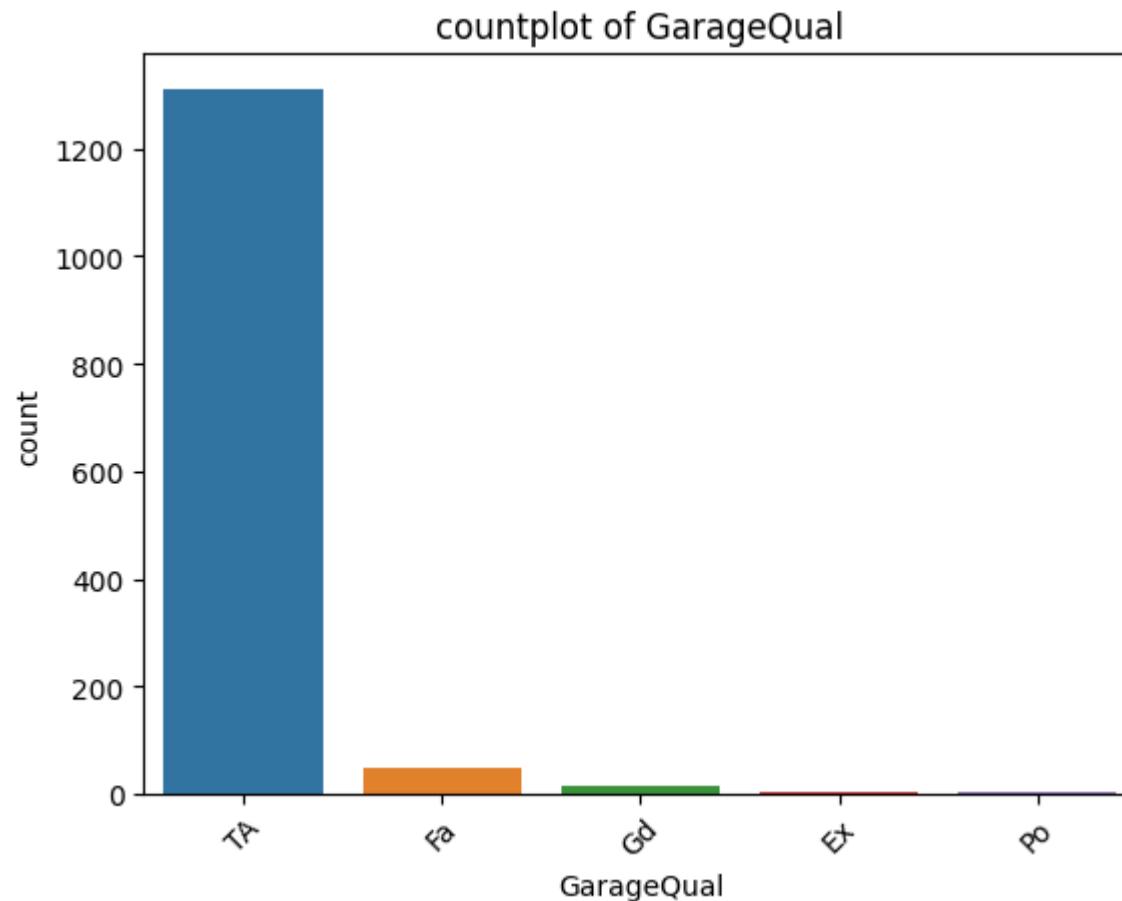


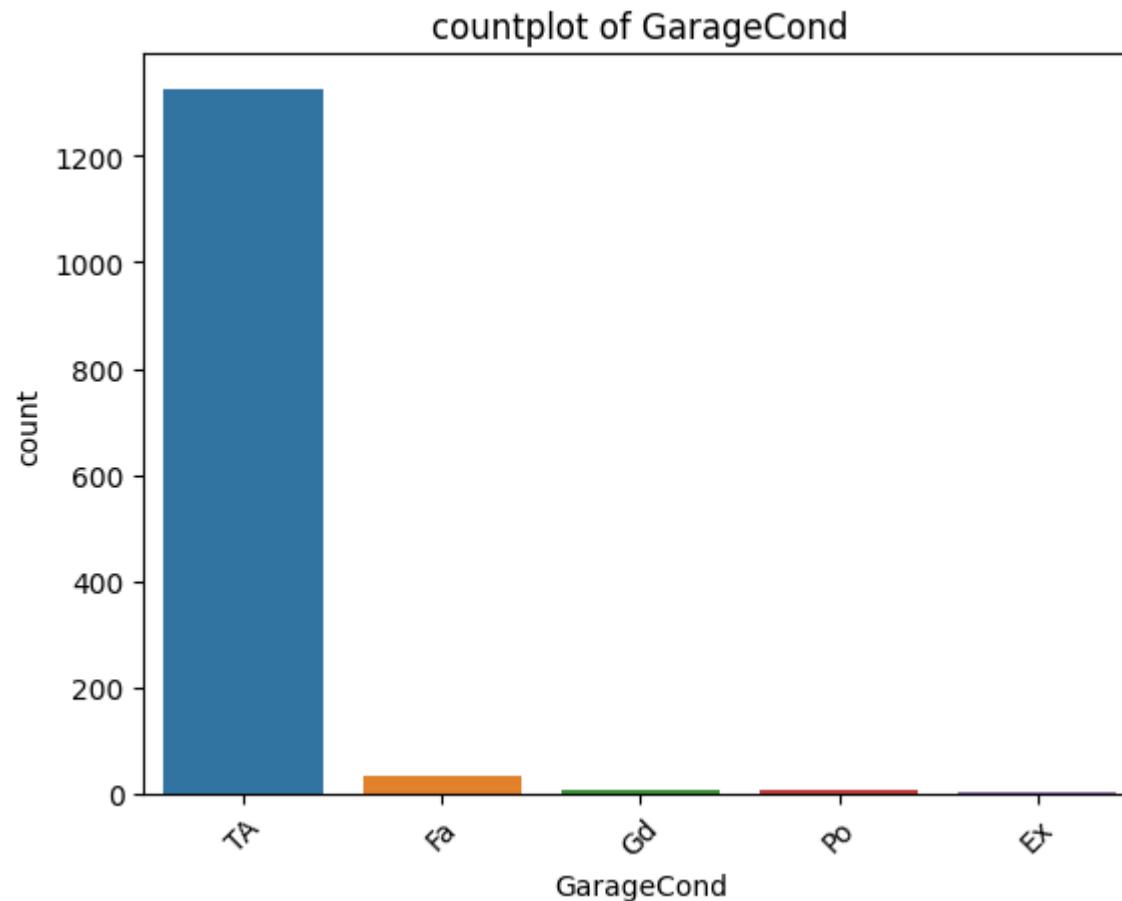


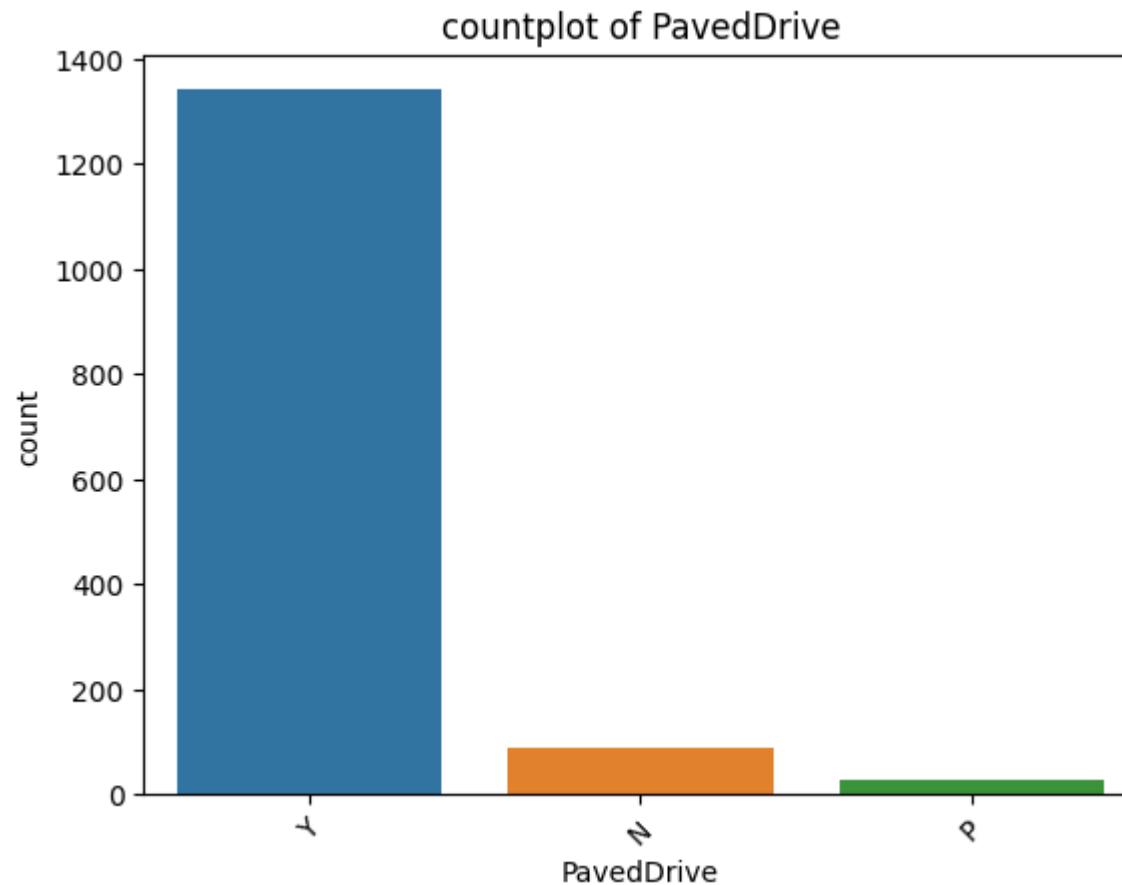


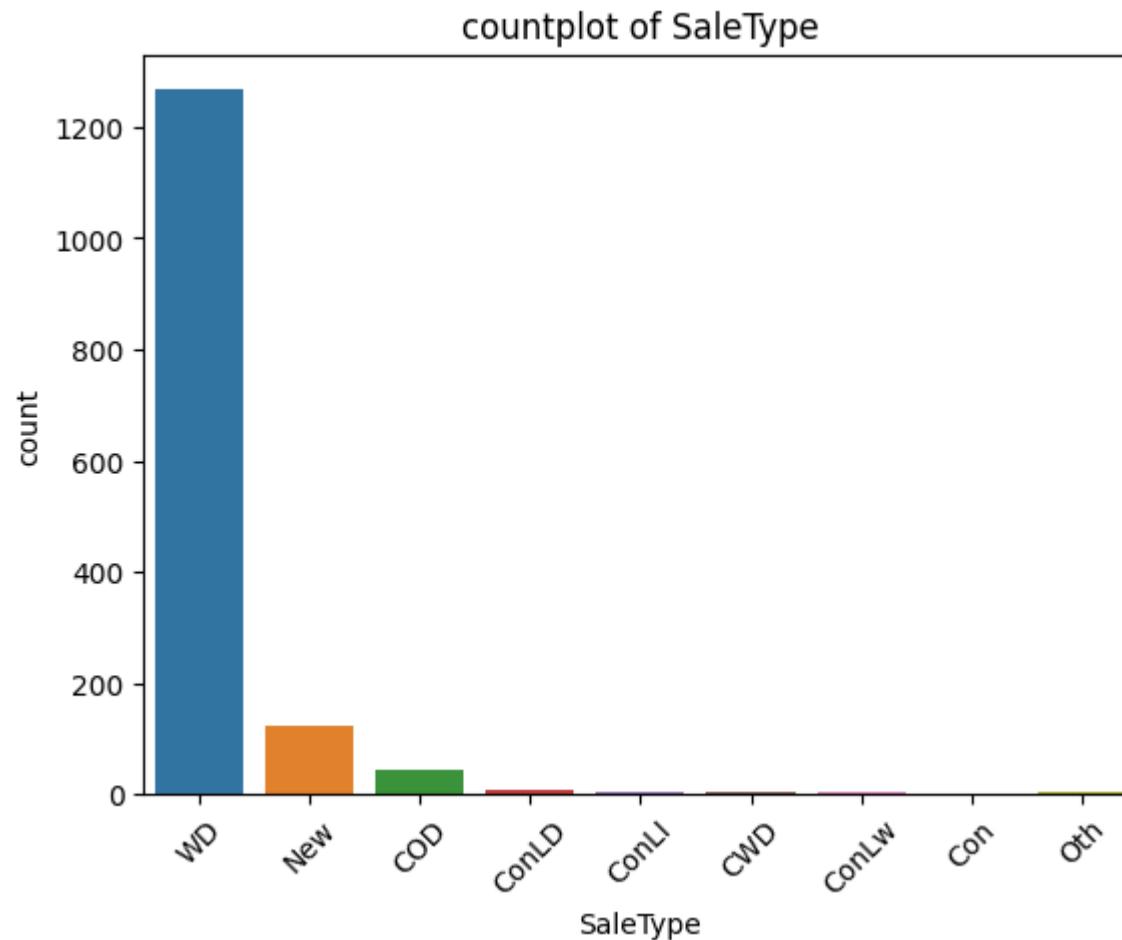


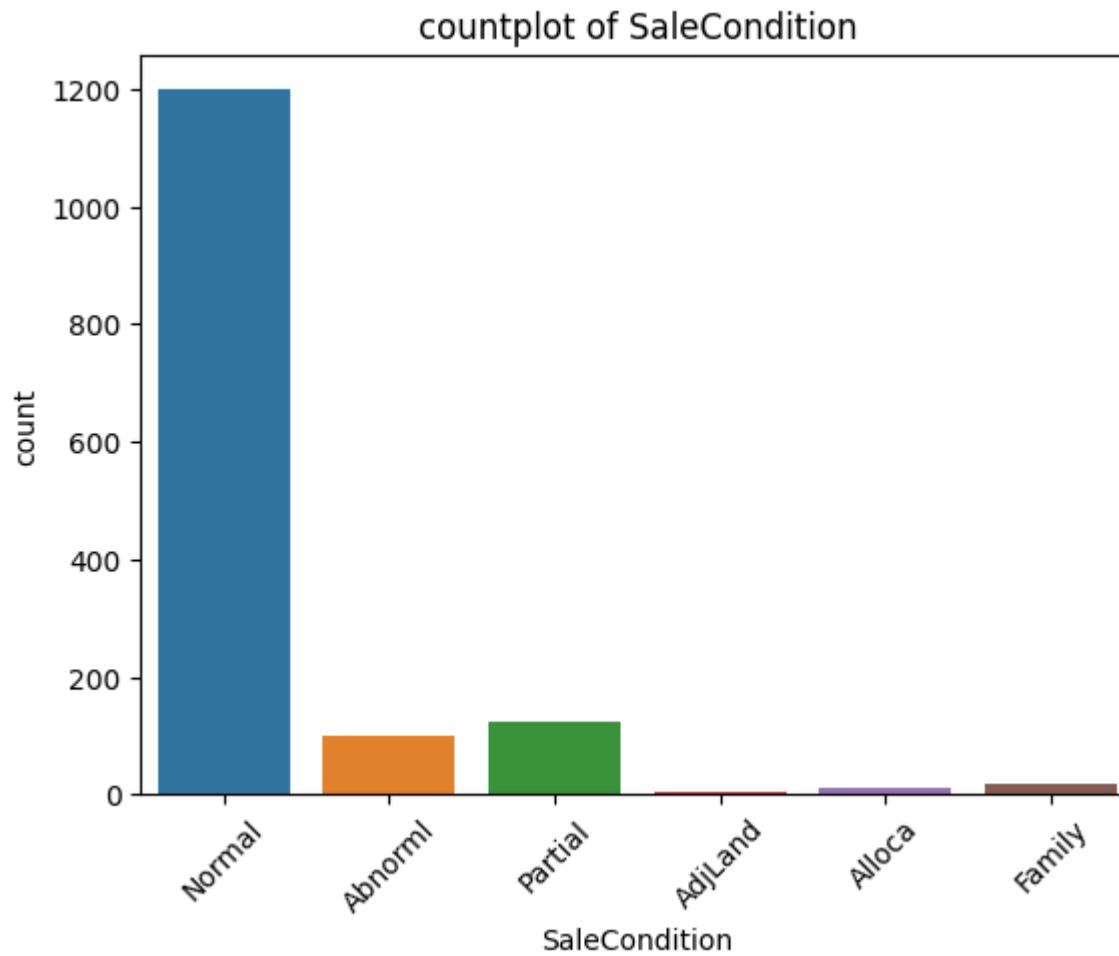






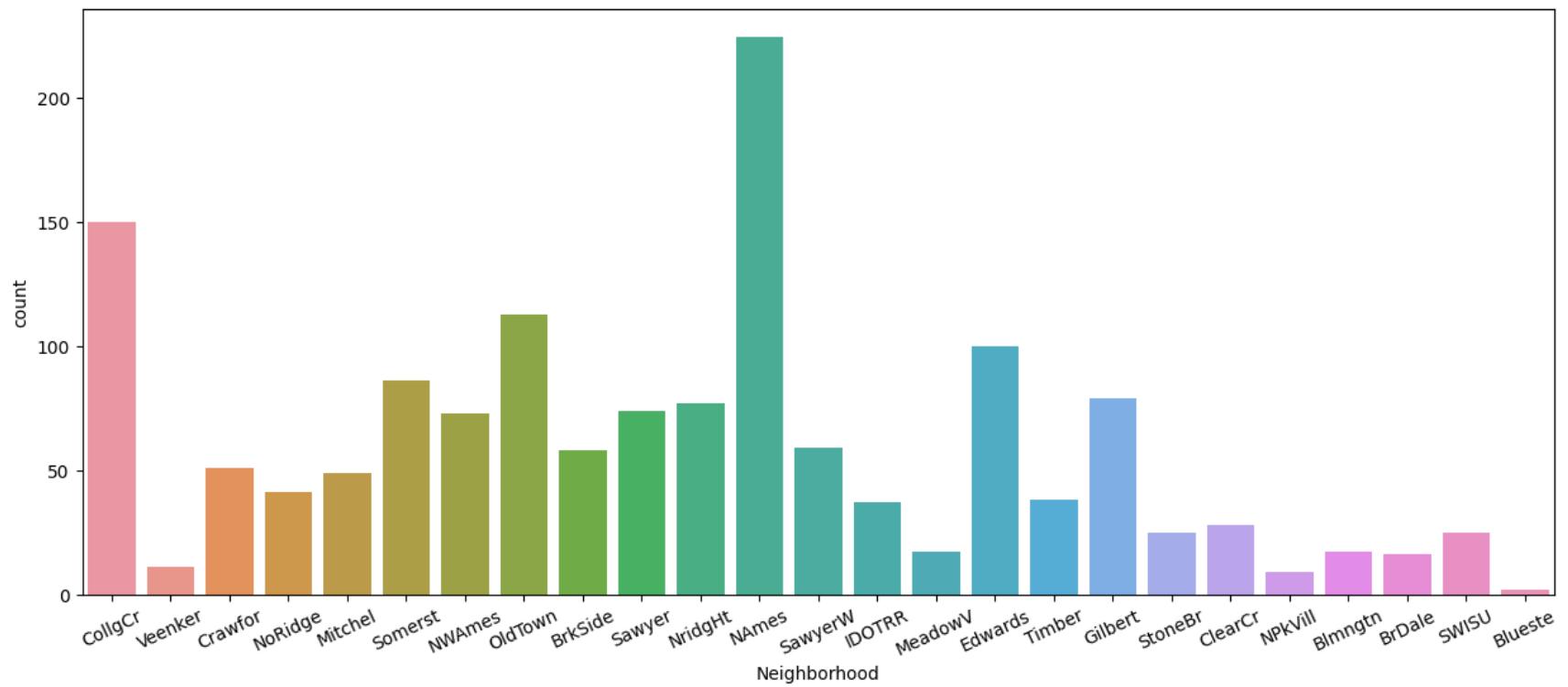






16/08/2023:

```
In [20]: plt.figure(figsize=(15,6))
sns.countplot(x=data['Neighborhood'])
plt.xticks(rotation=25)
plt.show()
```

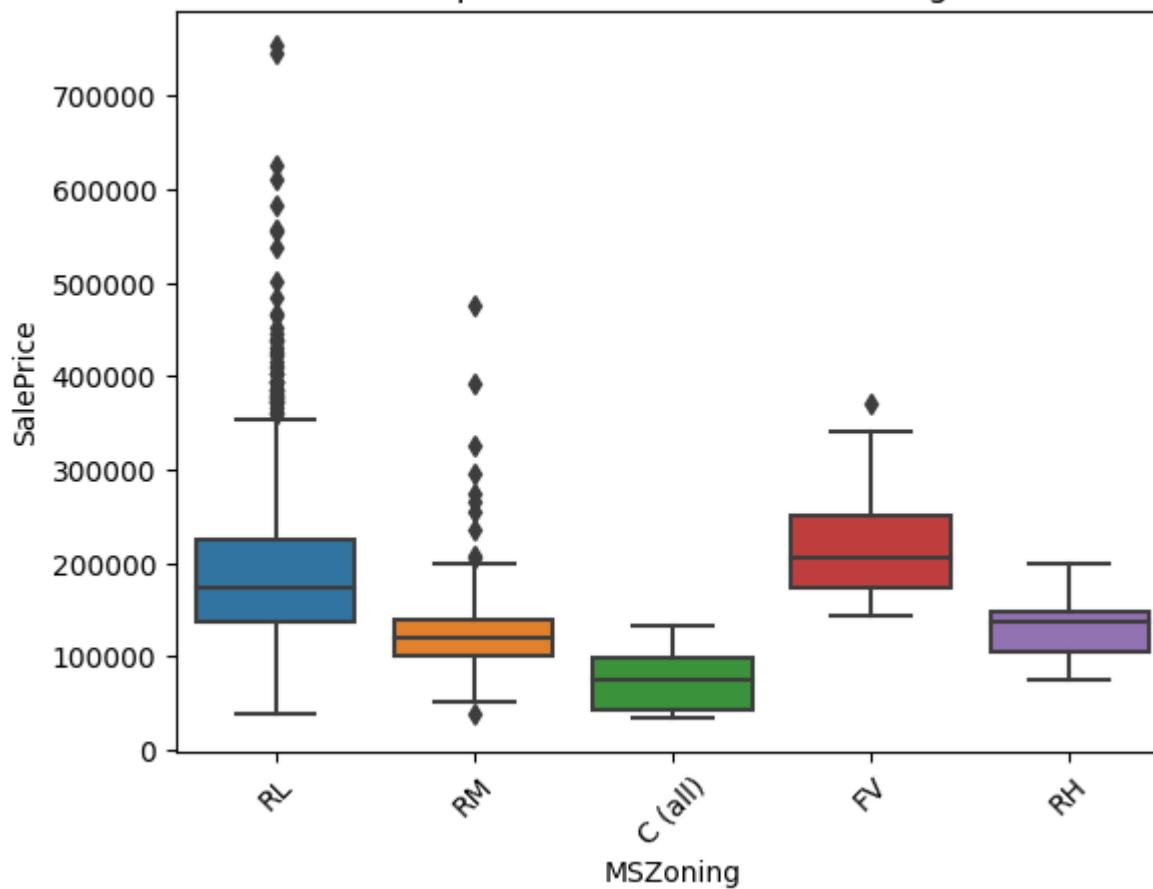


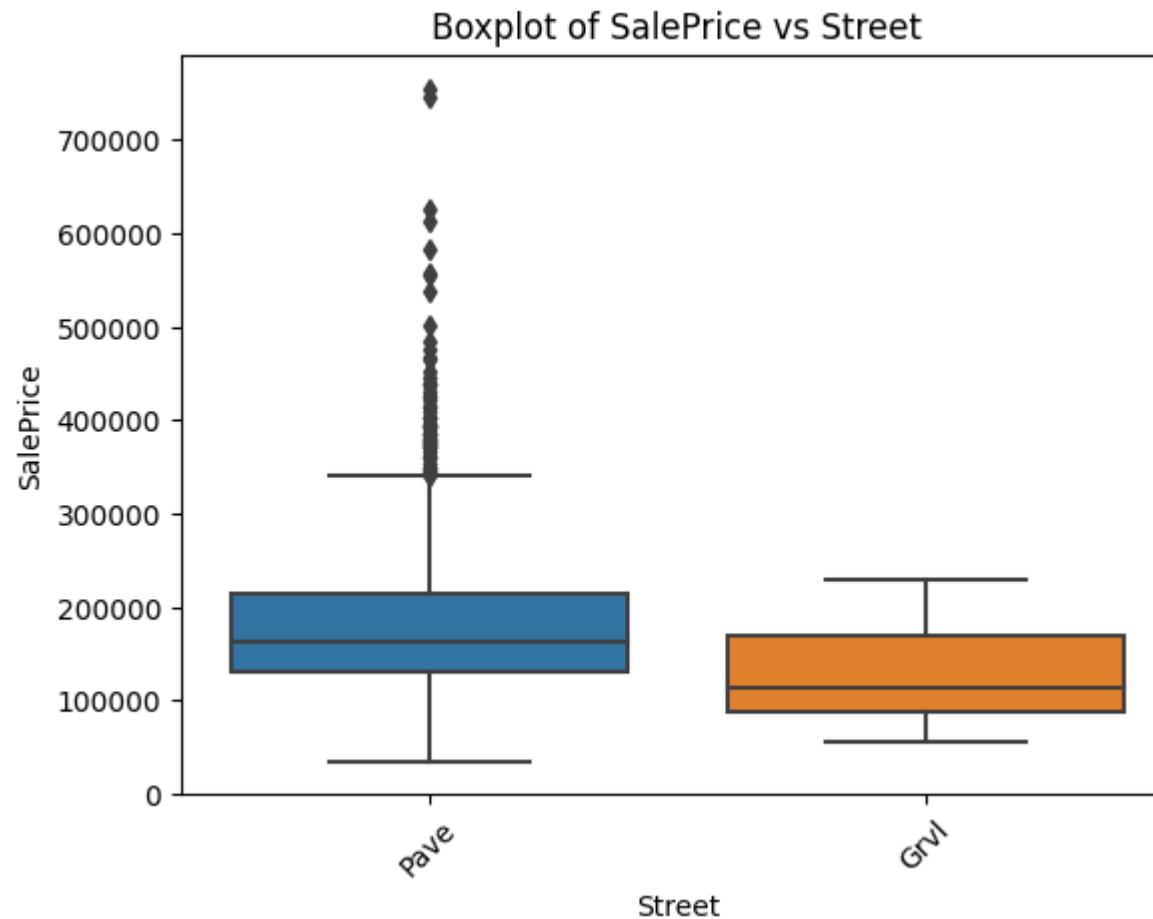
What different plots tell:

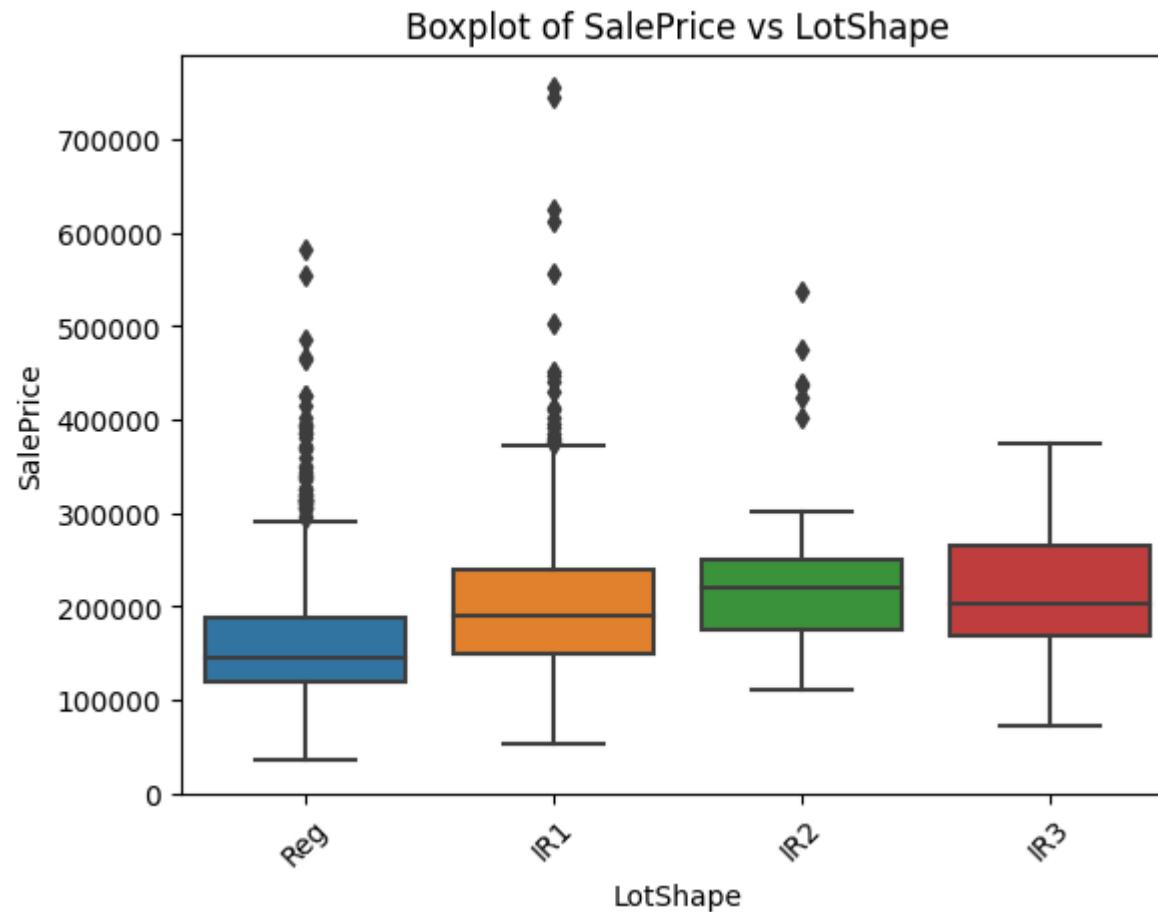
- Histogram/Boxplot(Numerical columns)---> Mean/Median
- Countplot(Categorical columns)-----> Mode

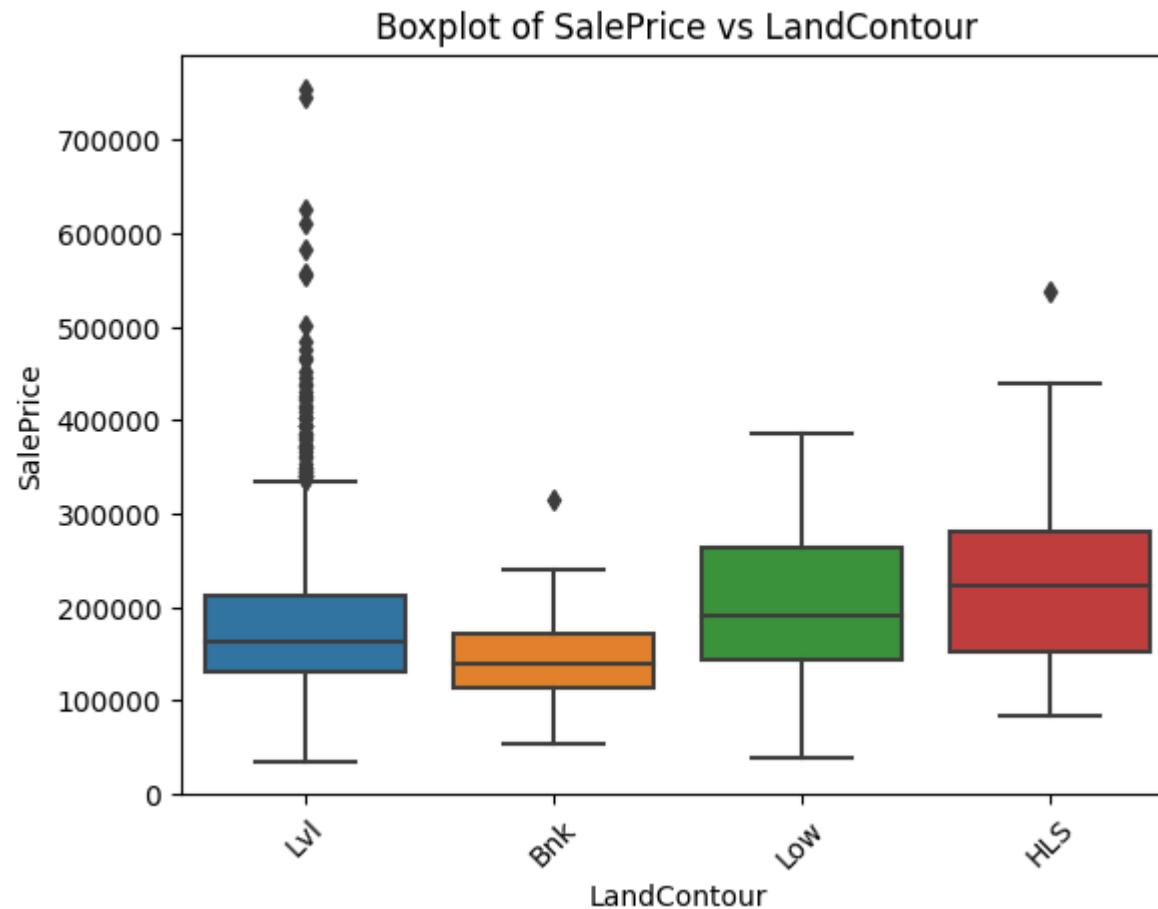
```
In [21]: for col in data.select_dtypes(include='object').columns:
    sns.boxplot(x=data[col],y=data['SalePrice'])
    plt.title(f'Boxplot of SalePrice vs {col}')
    plt.xticks(rotation=45)
    plt.show()
```

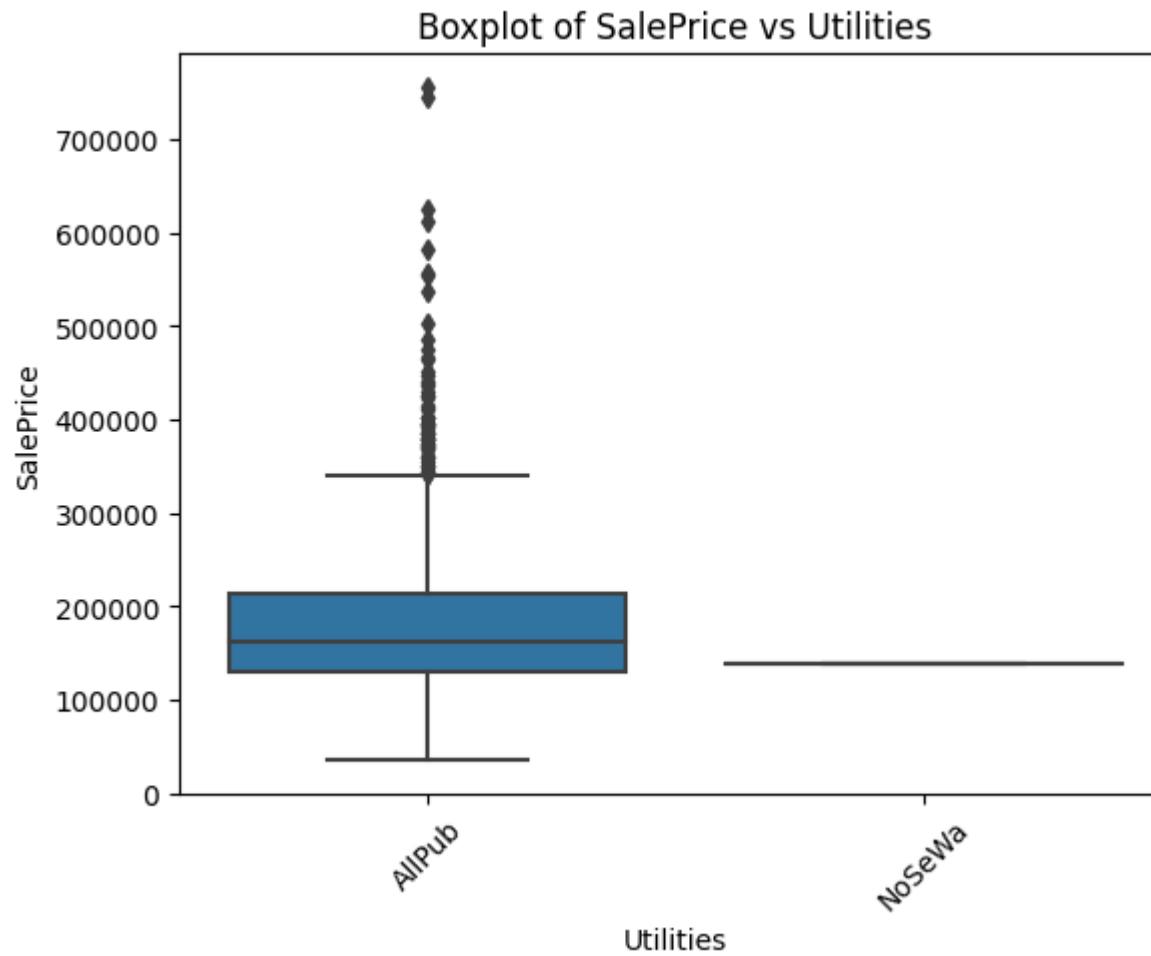
Boxplot of SalePrice vs MSZoning

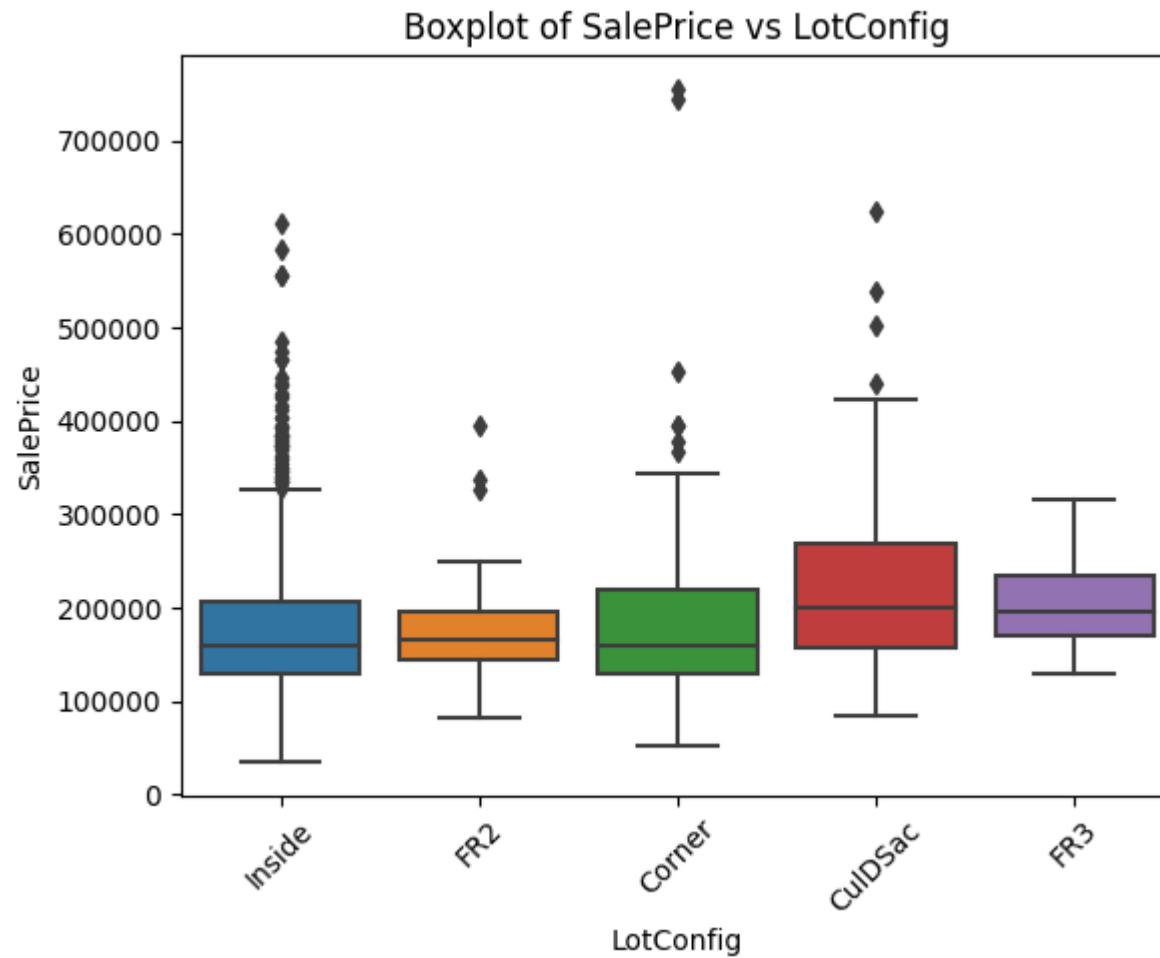


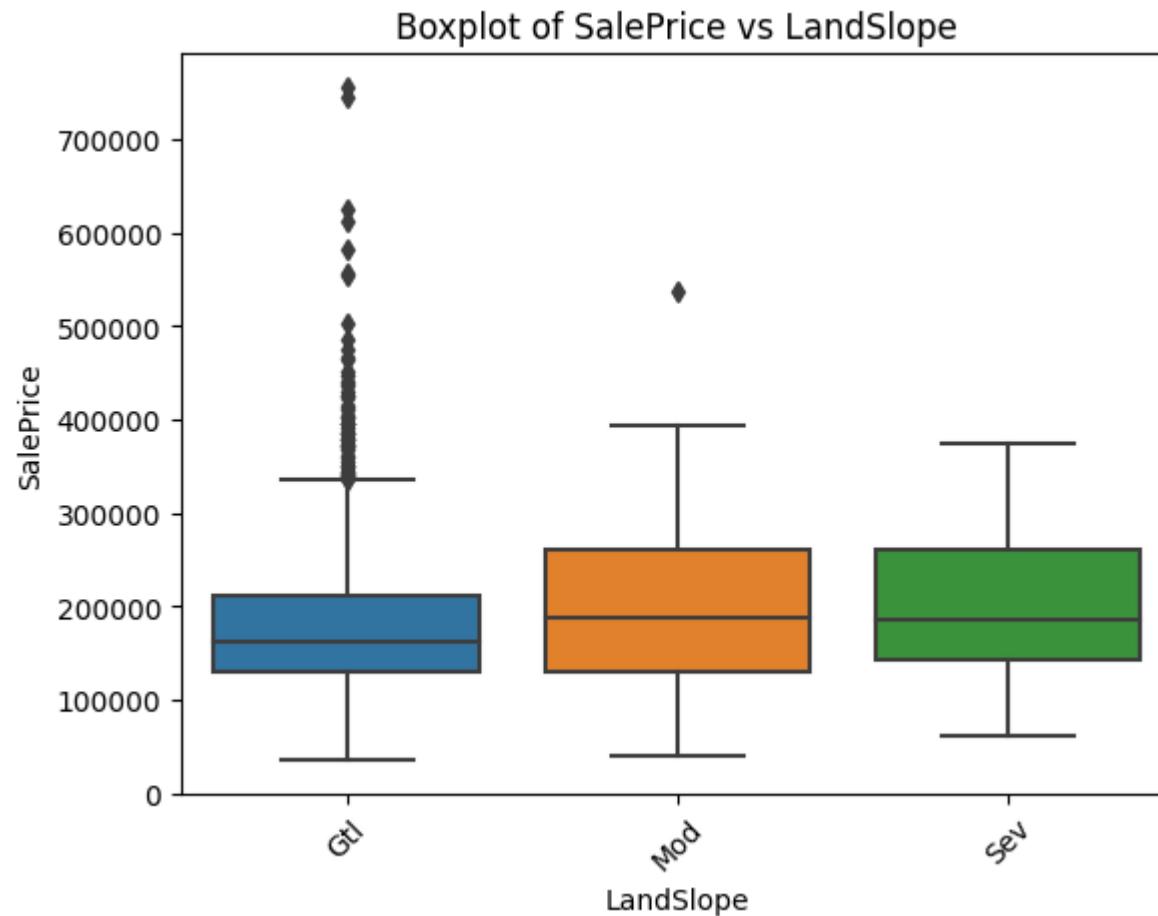


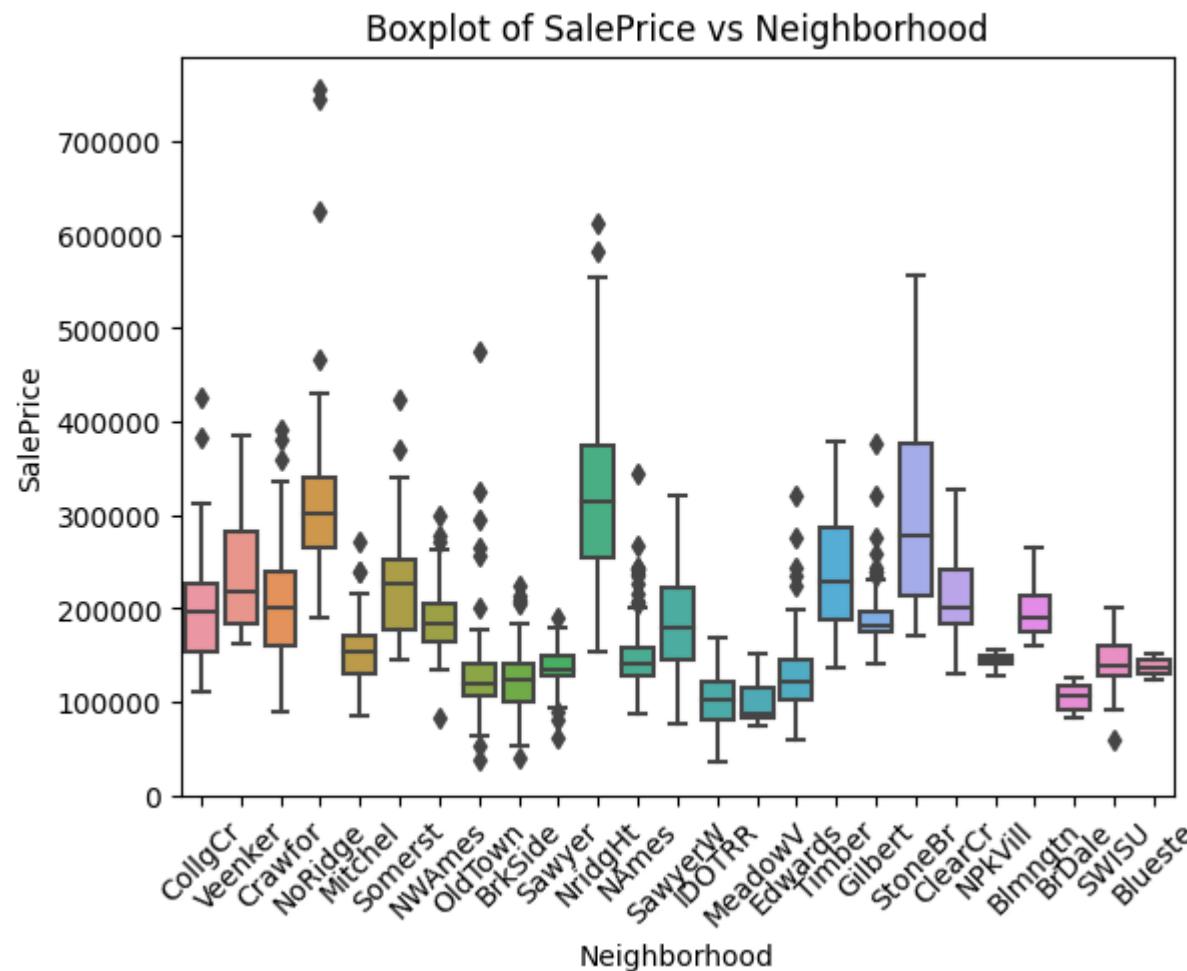


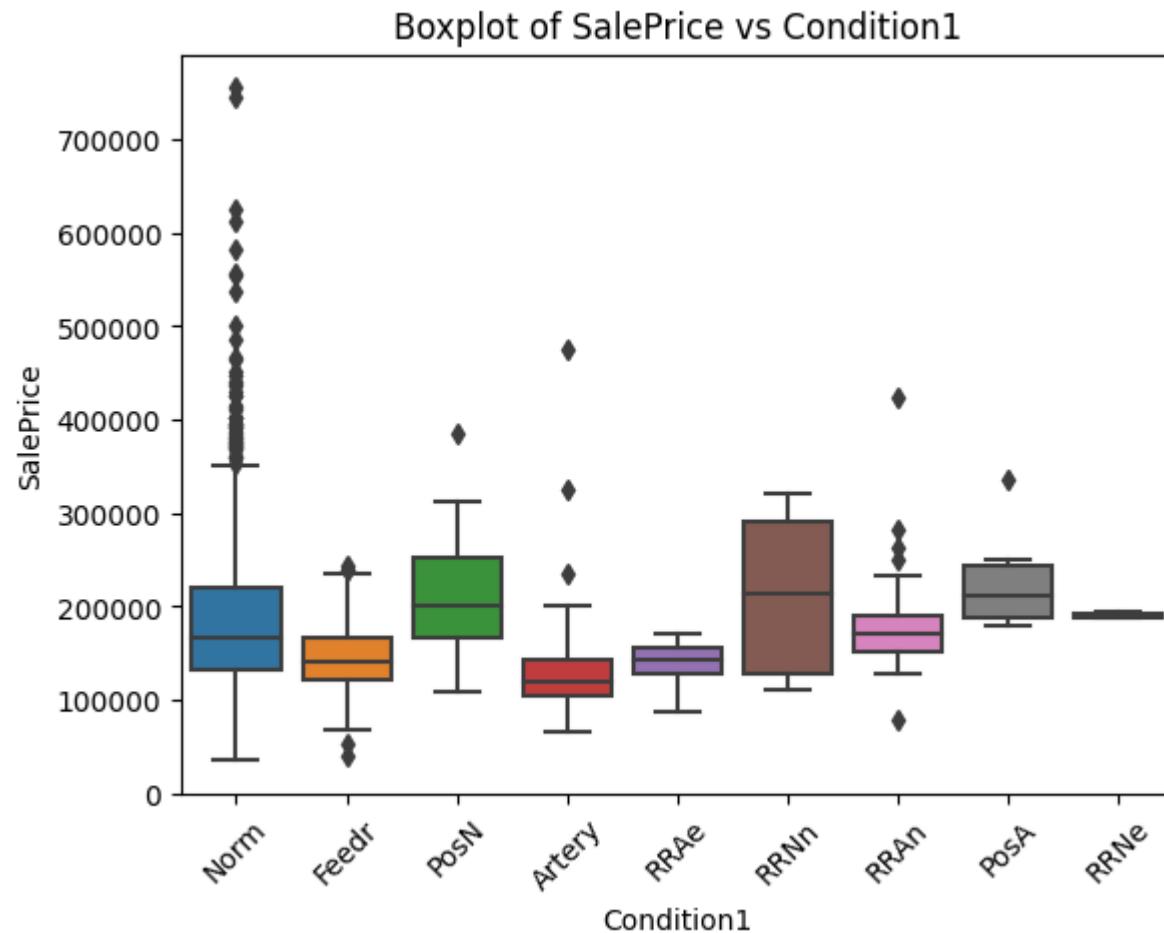


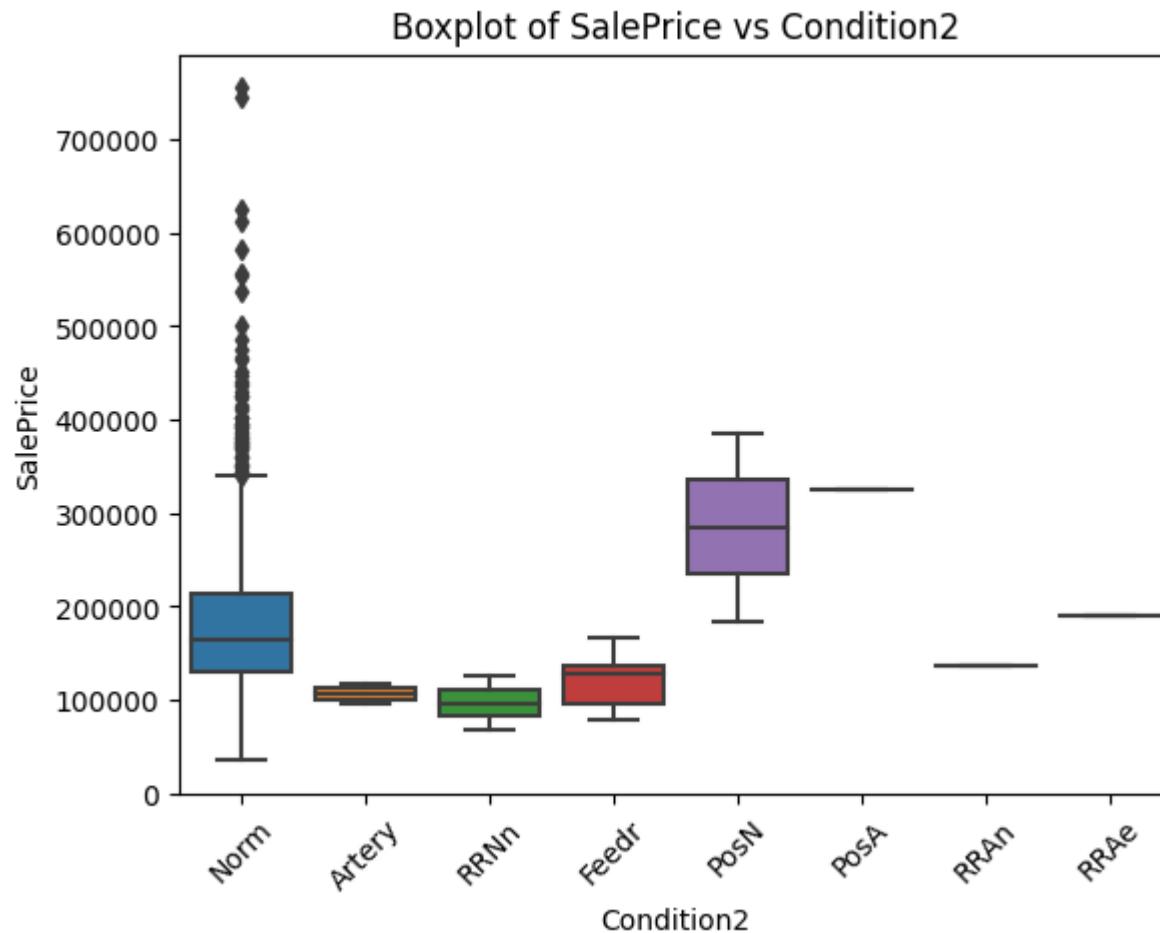


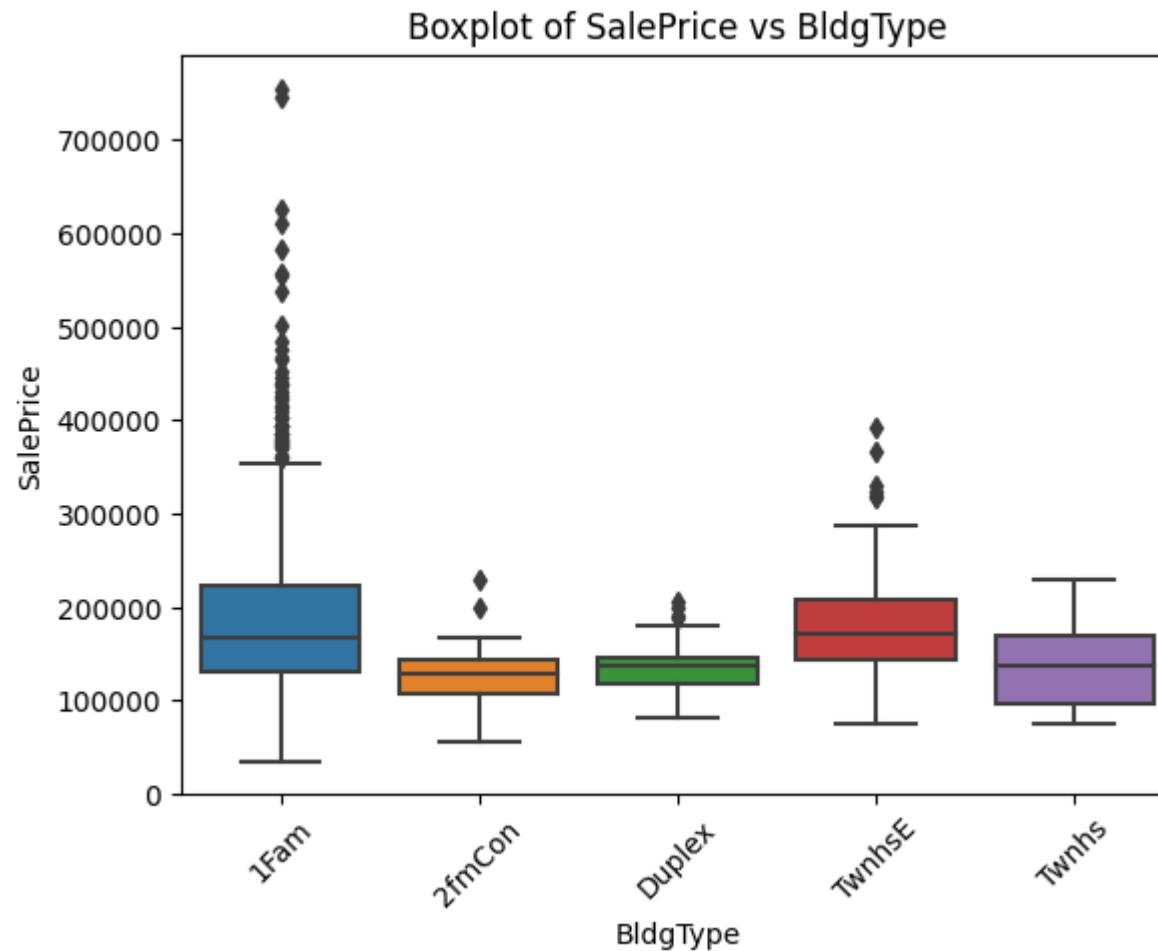


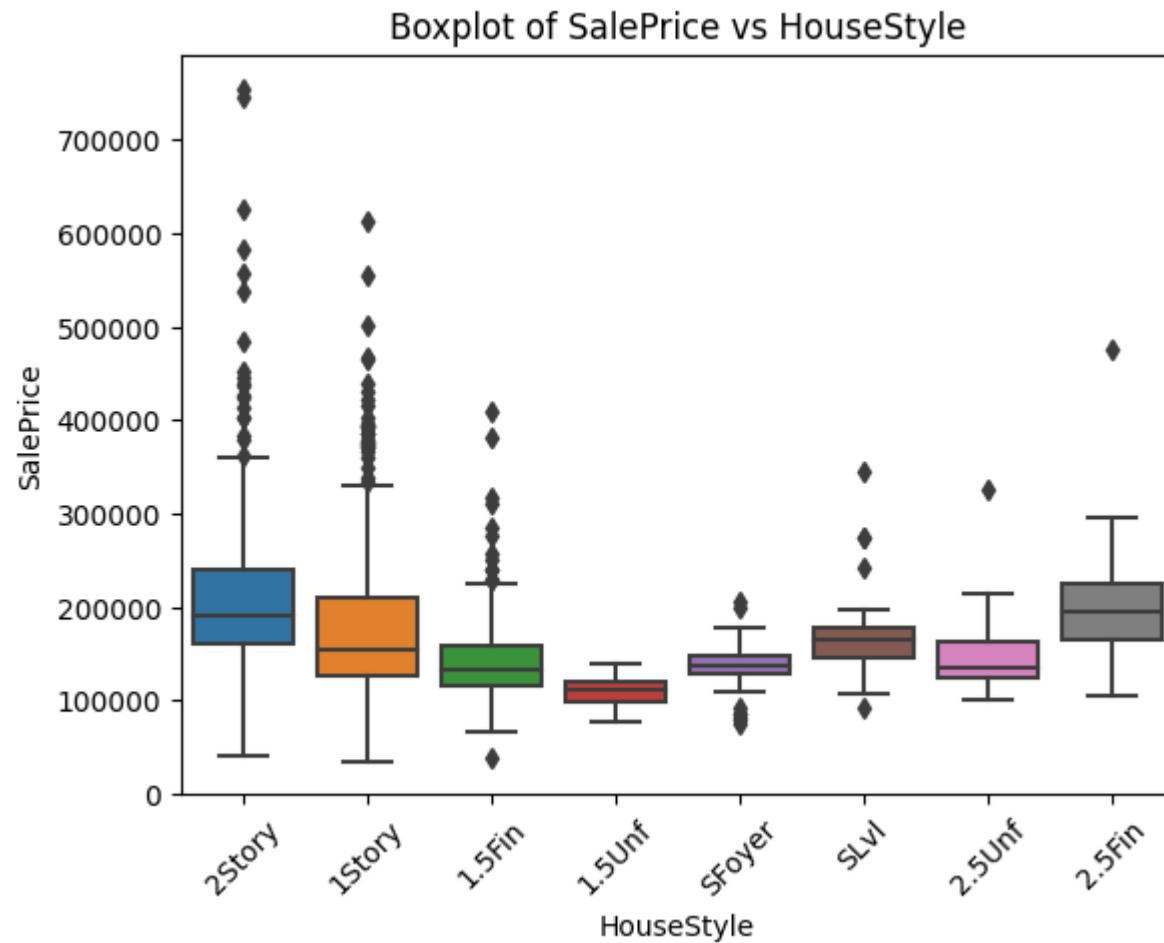


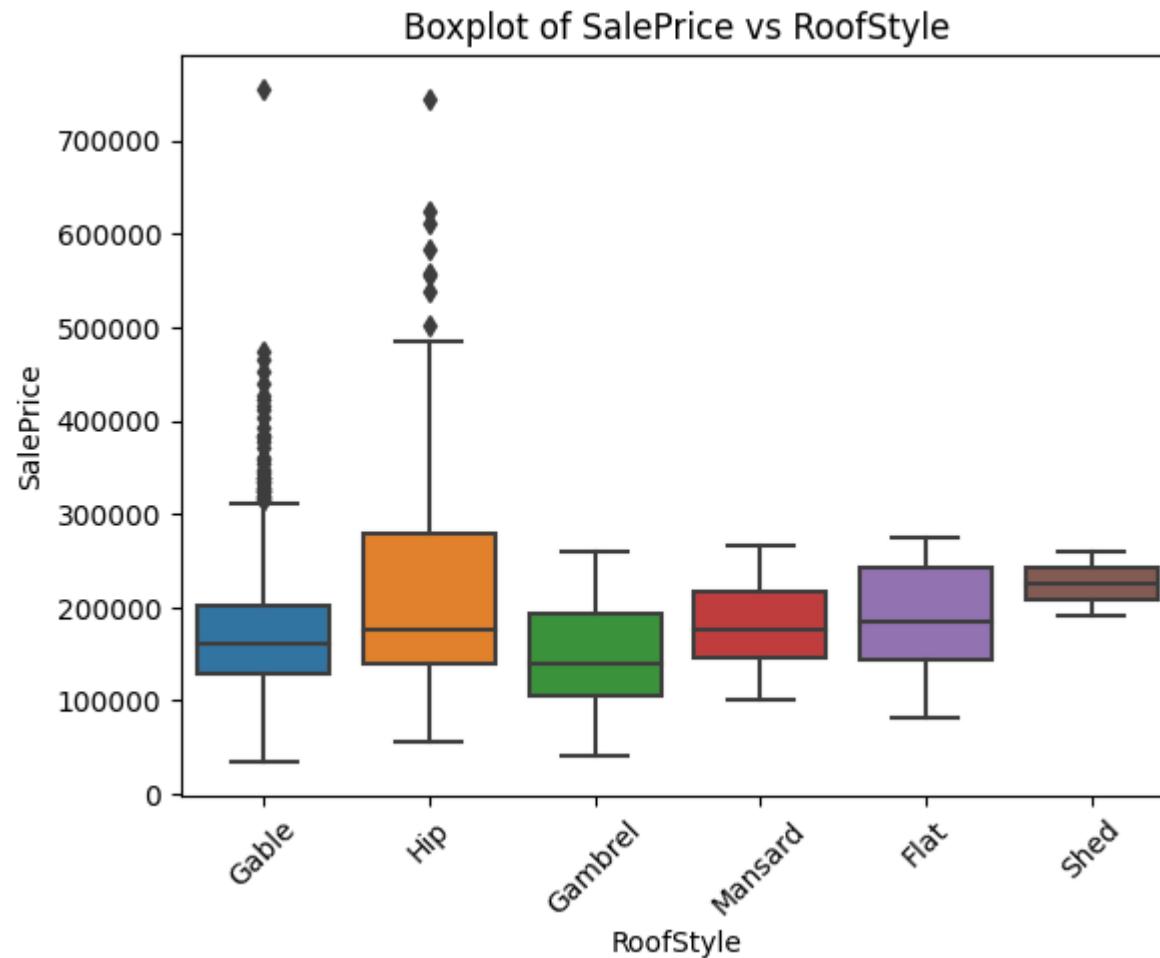


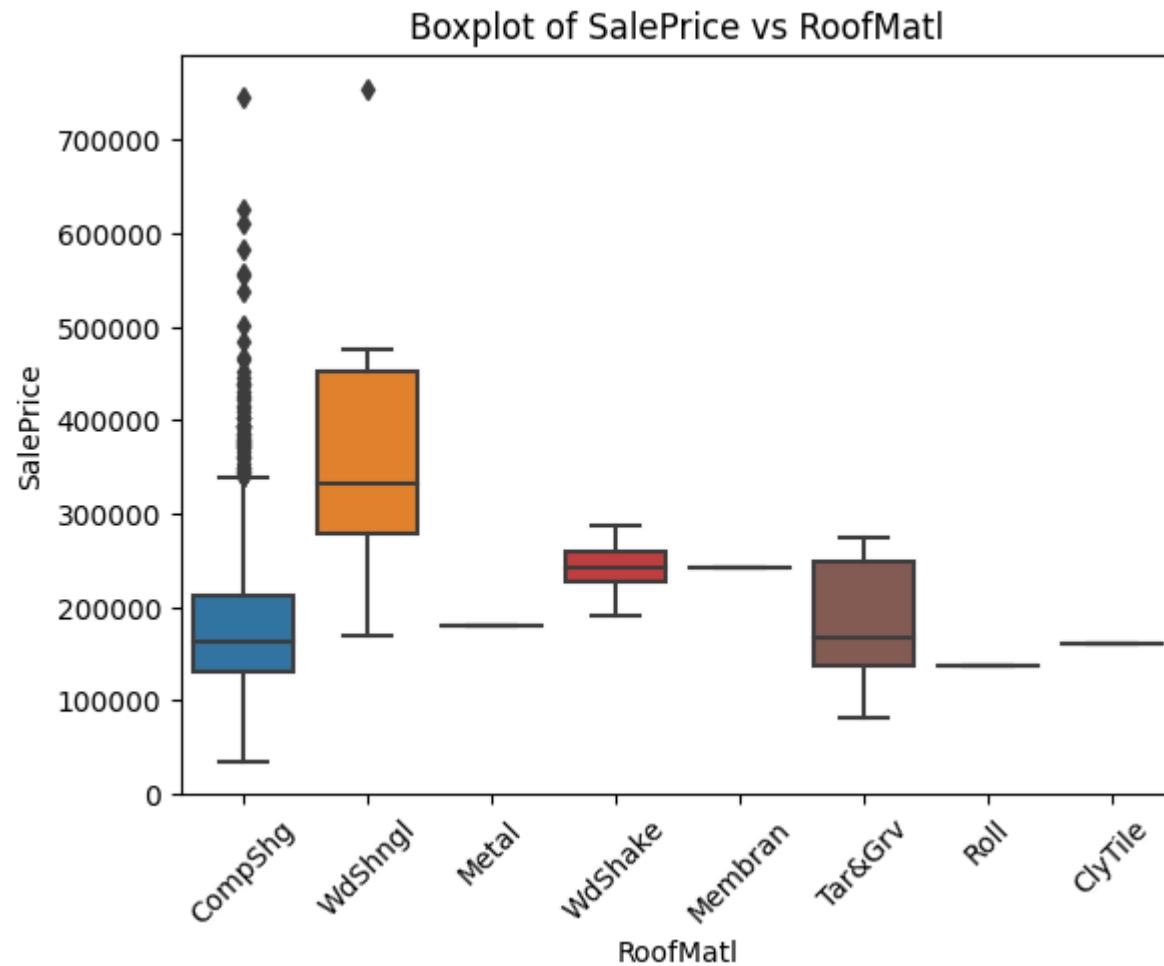


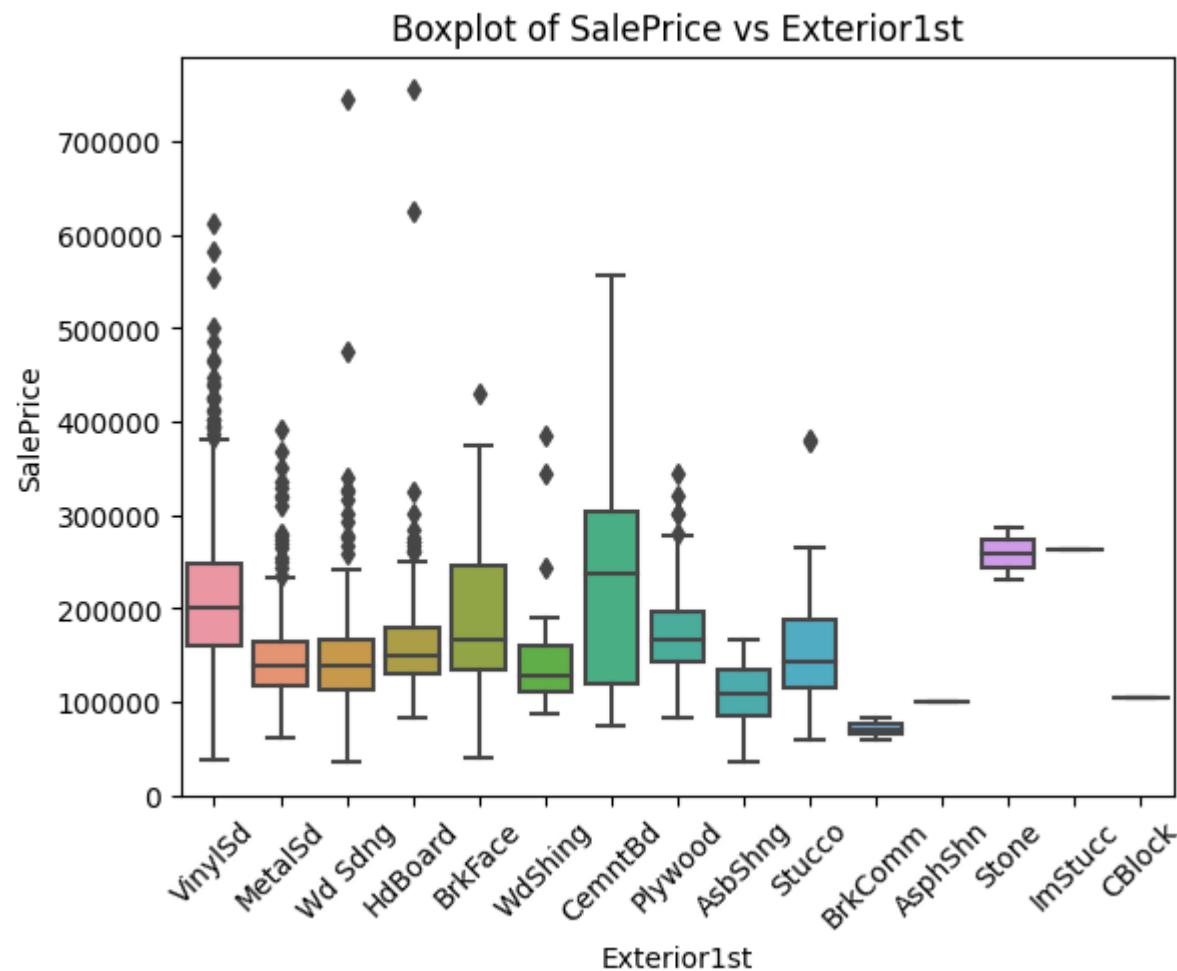


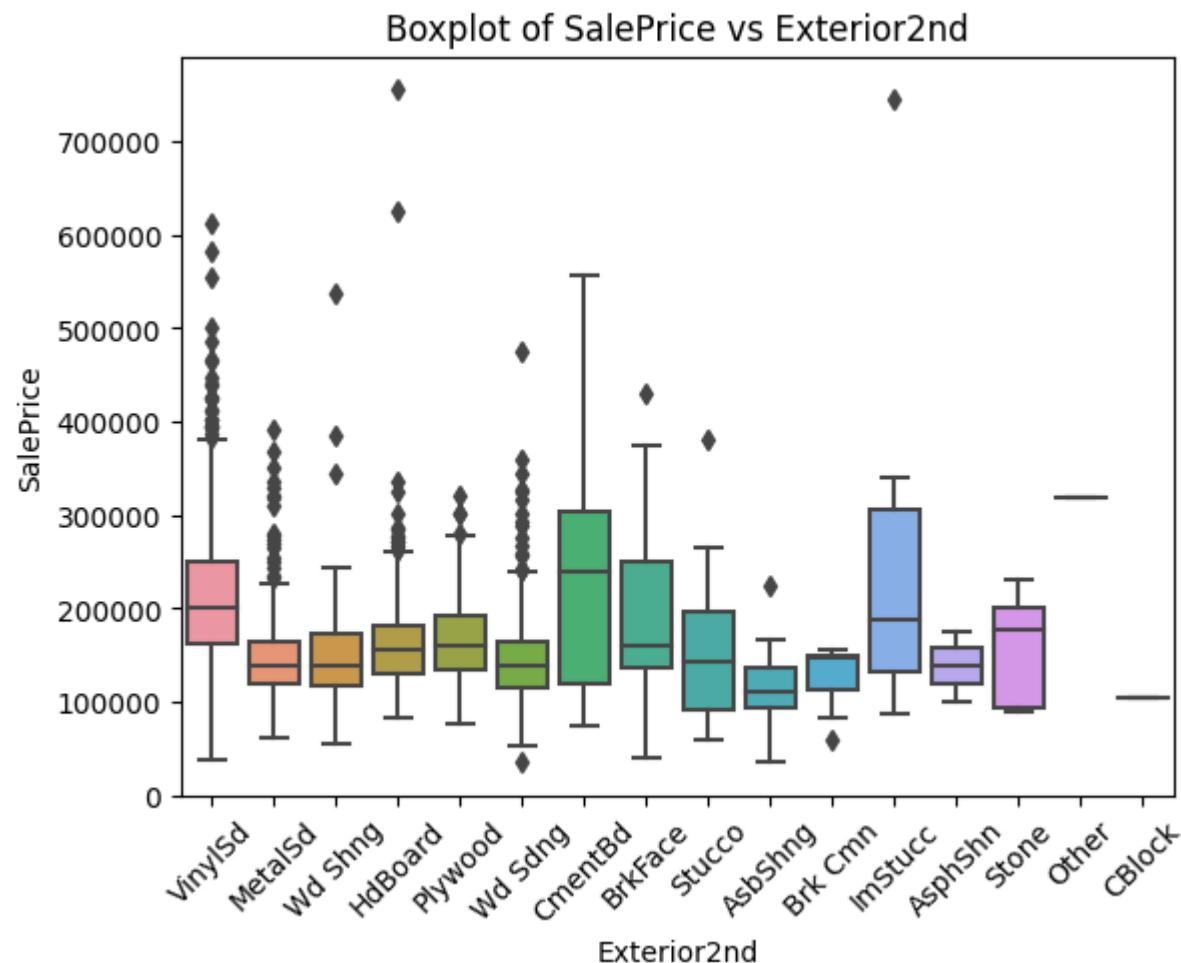


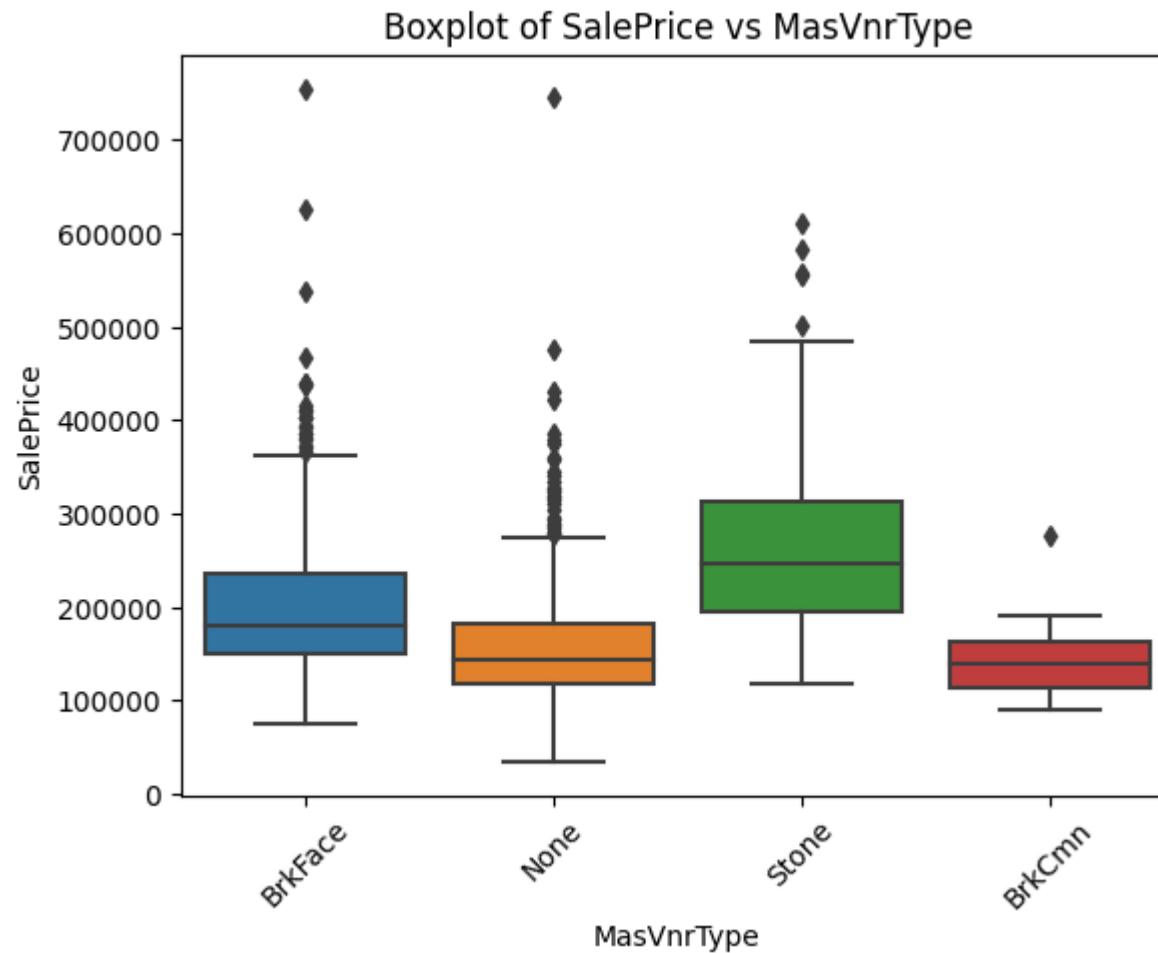


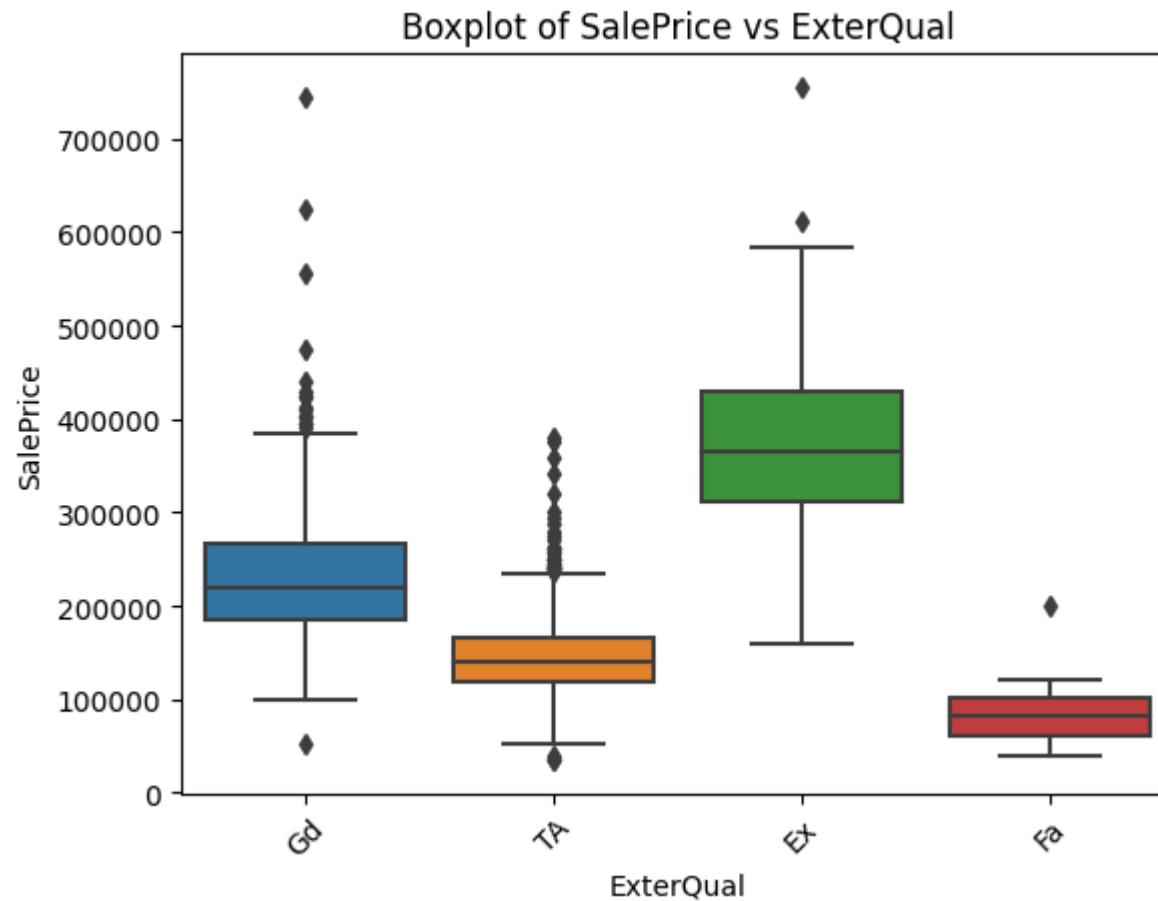


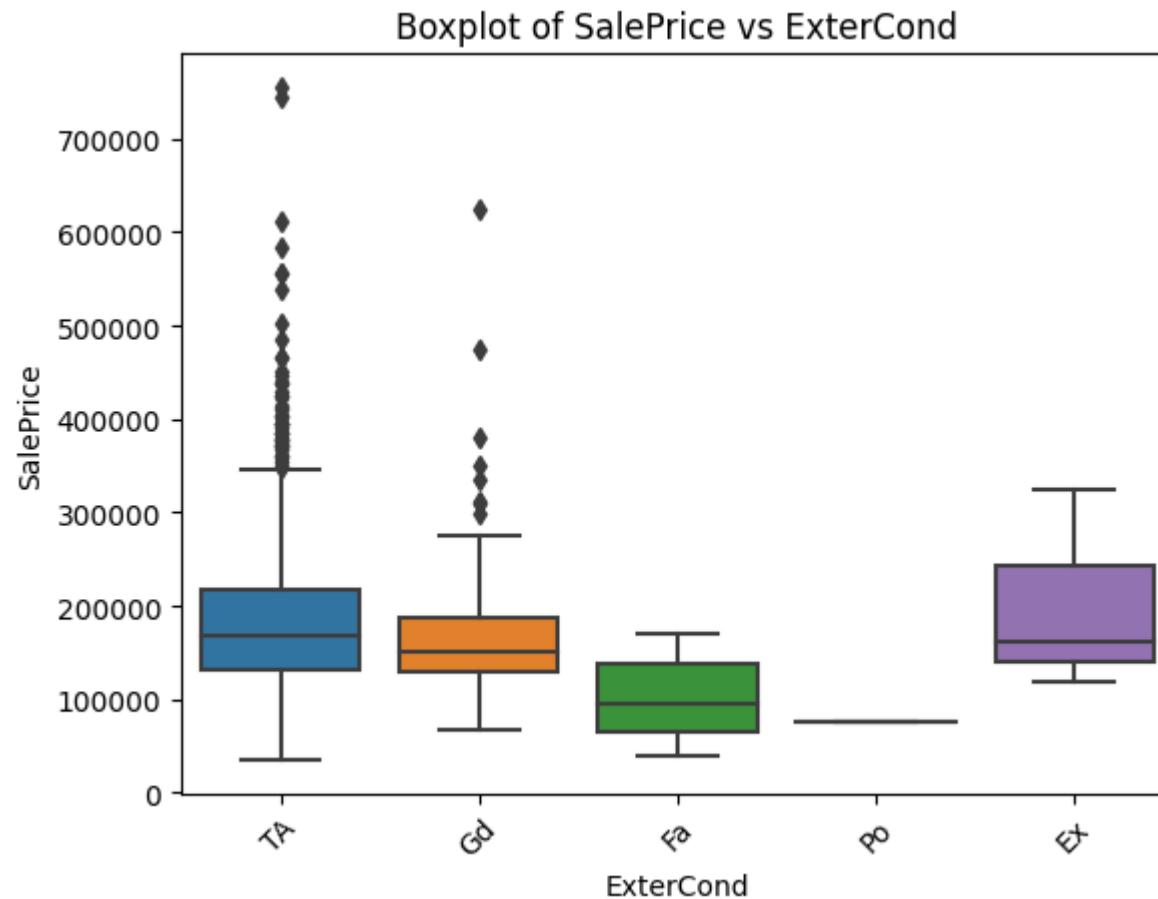


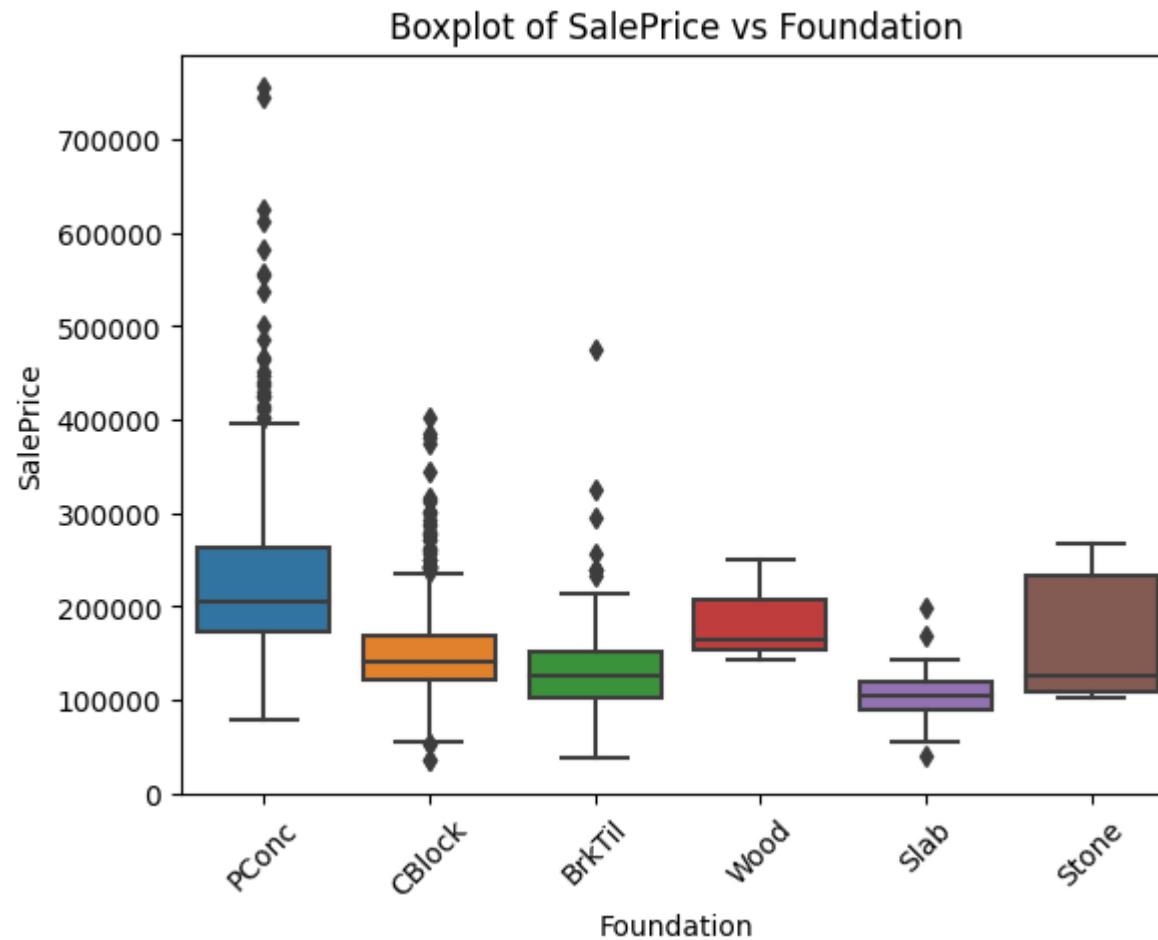


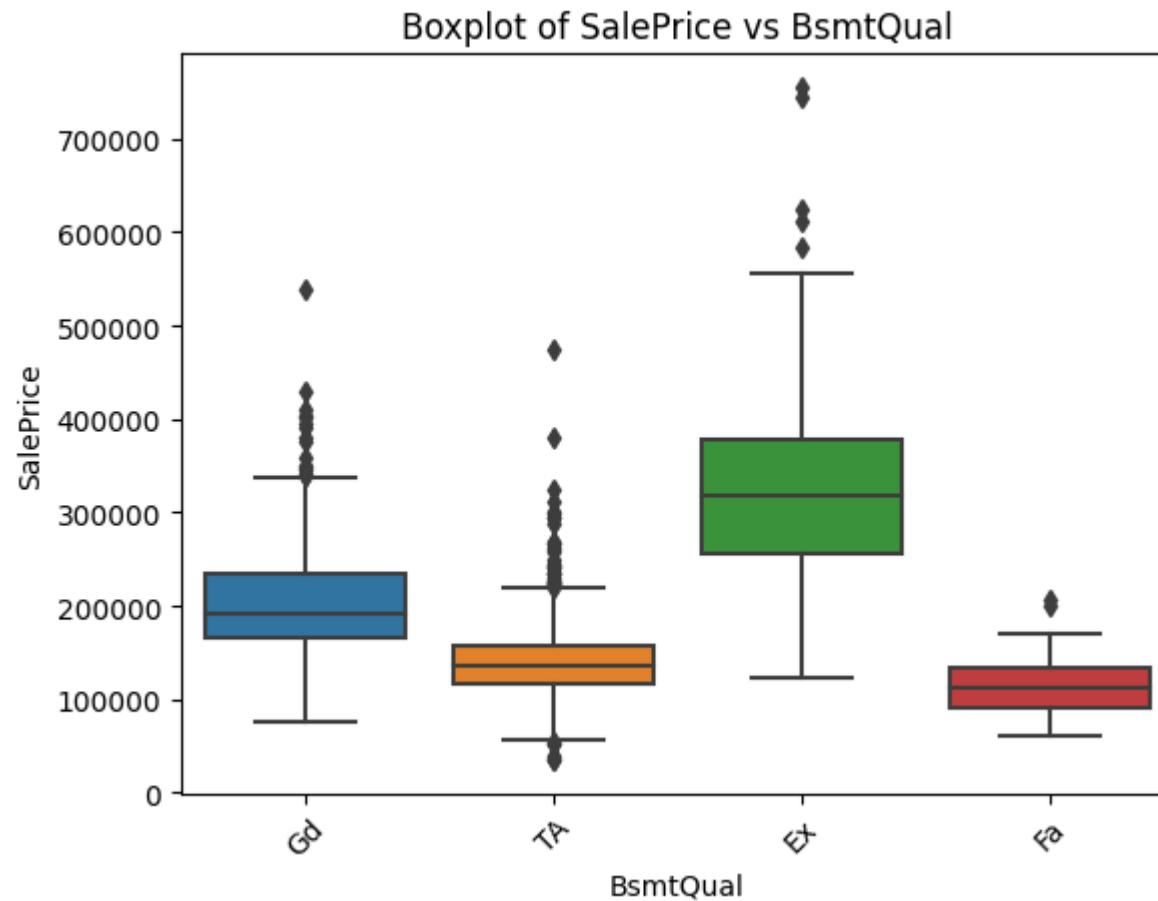


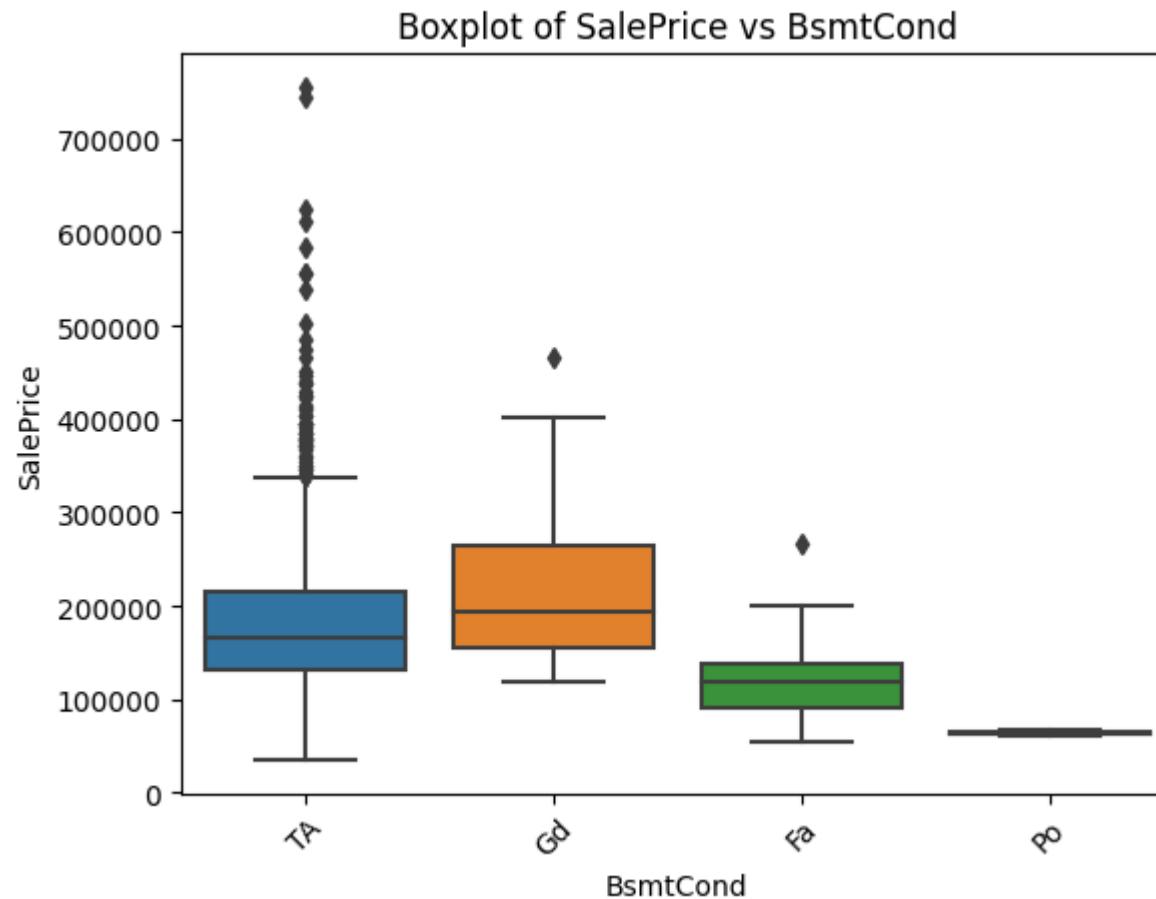


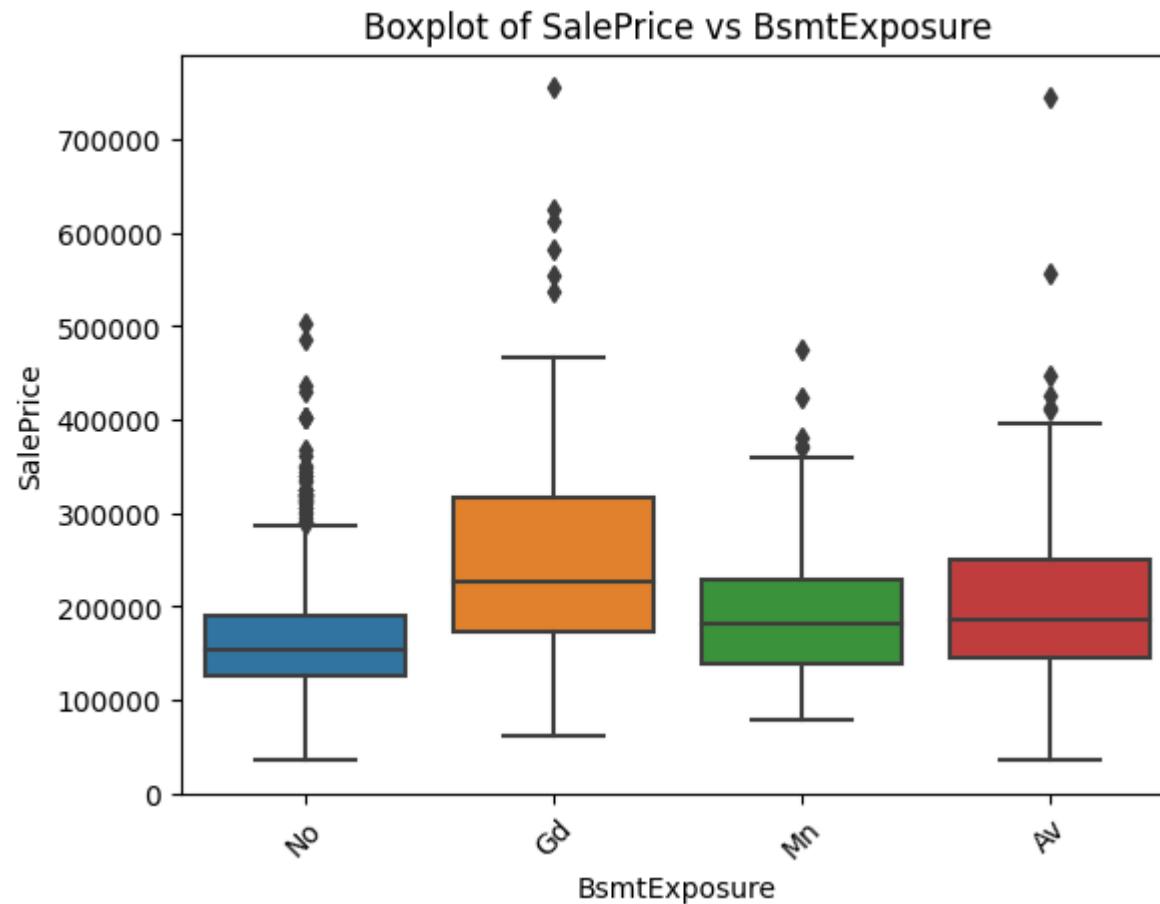


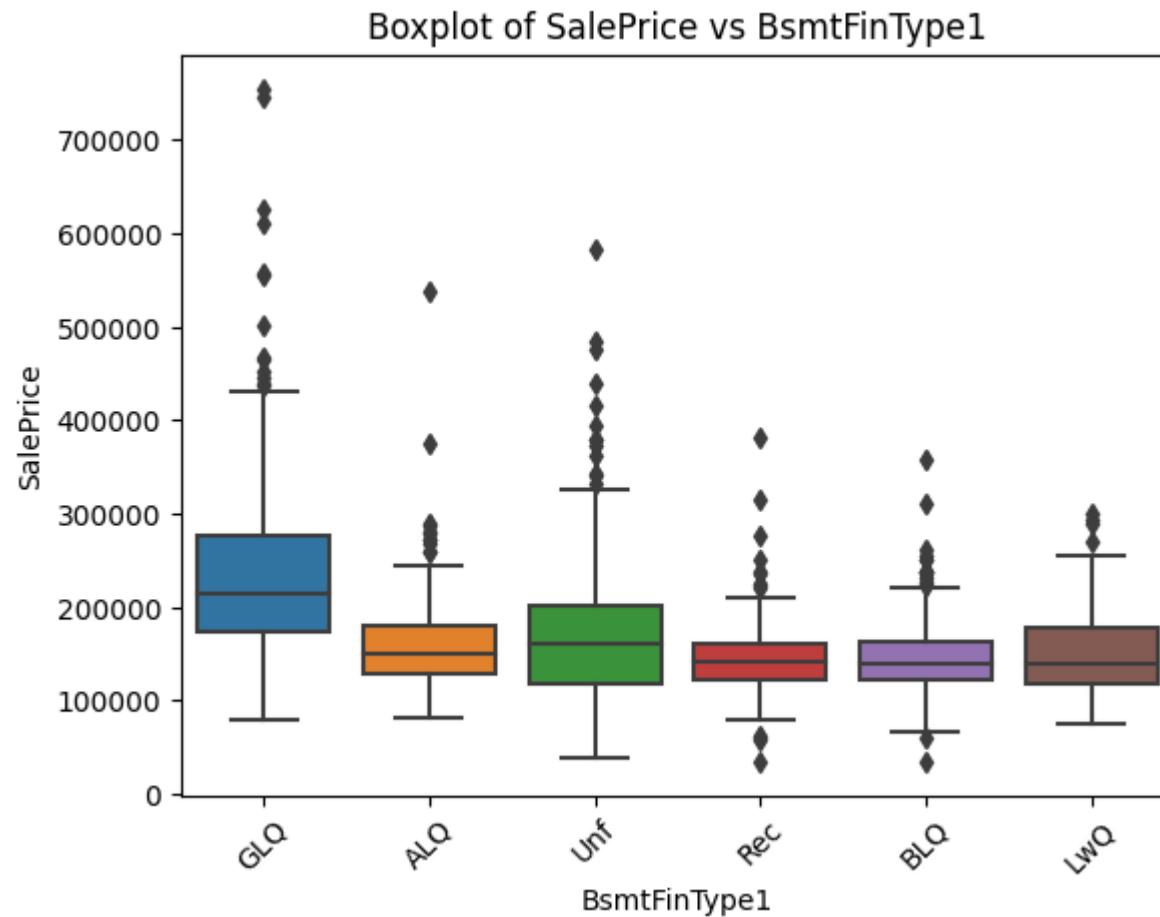


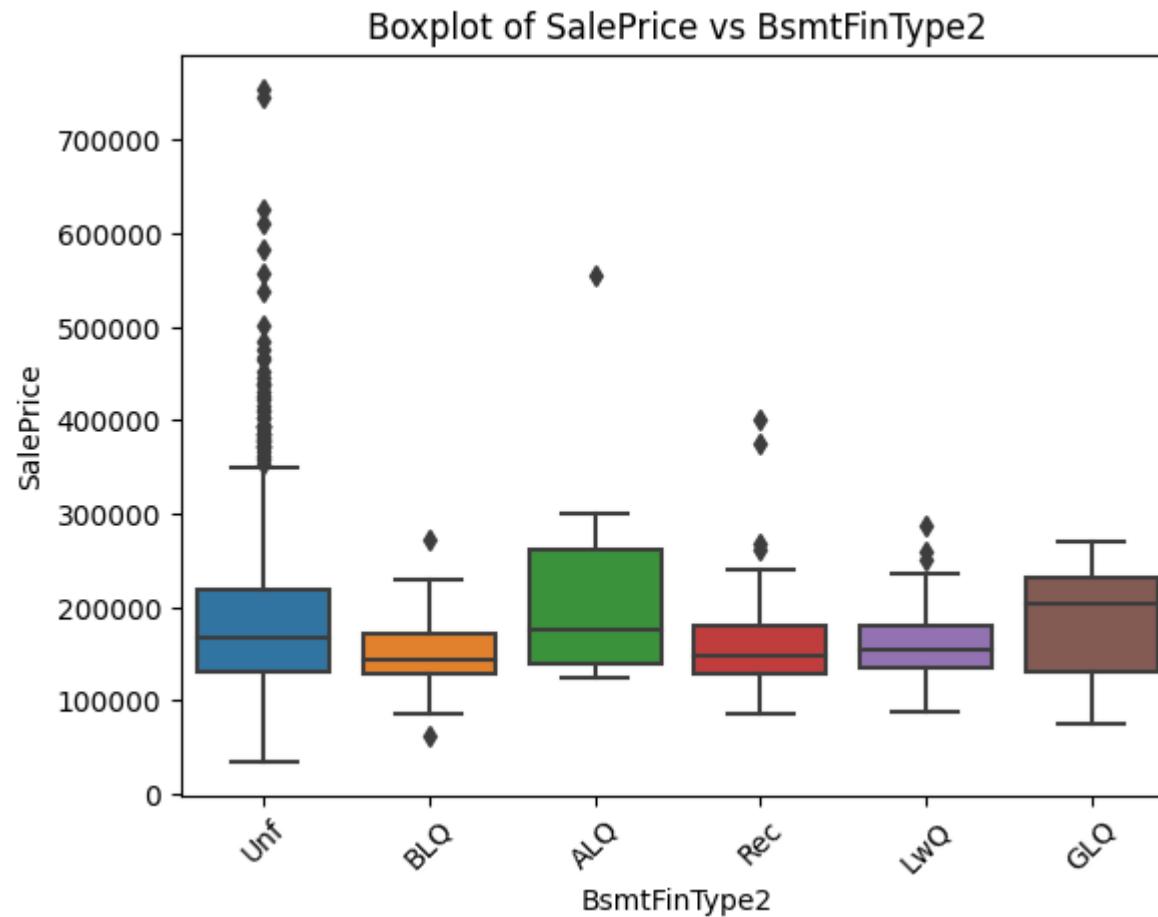


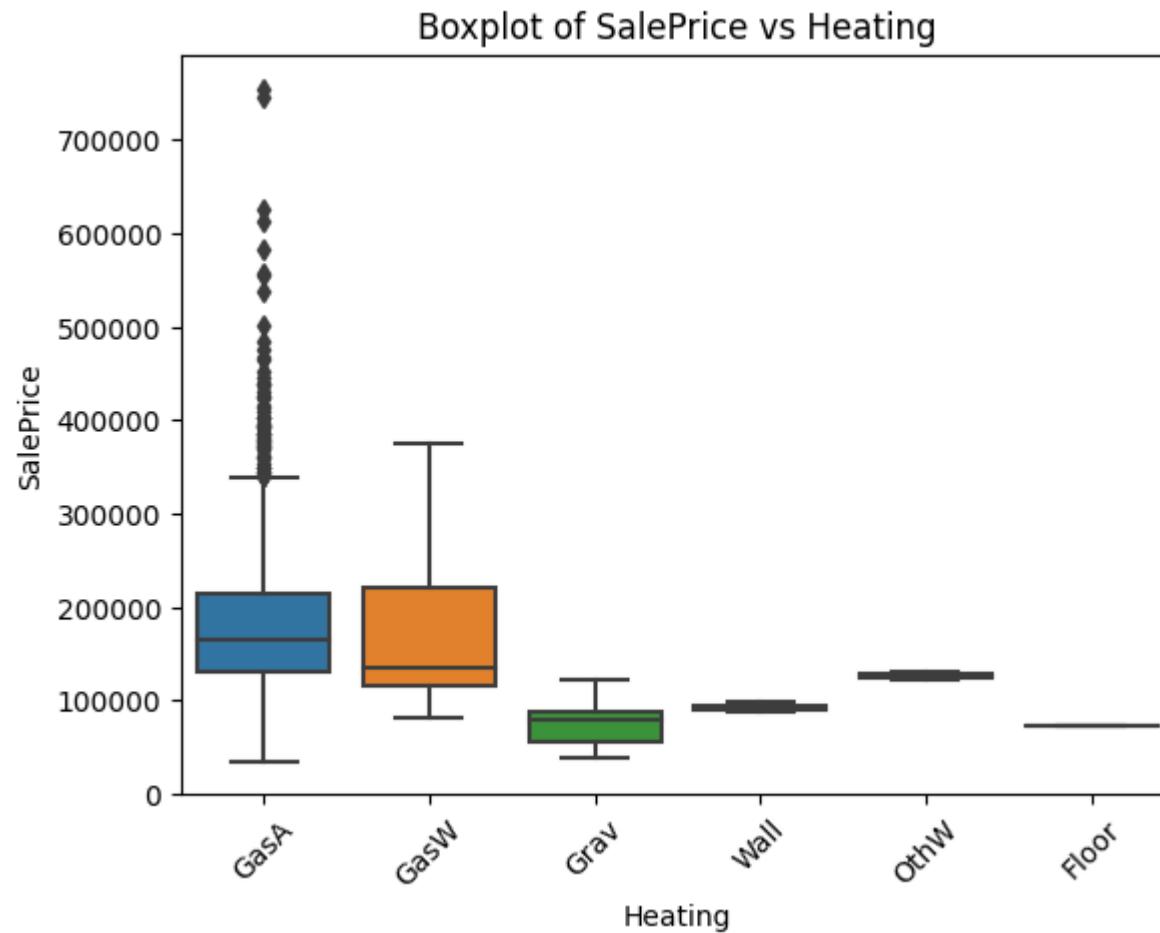


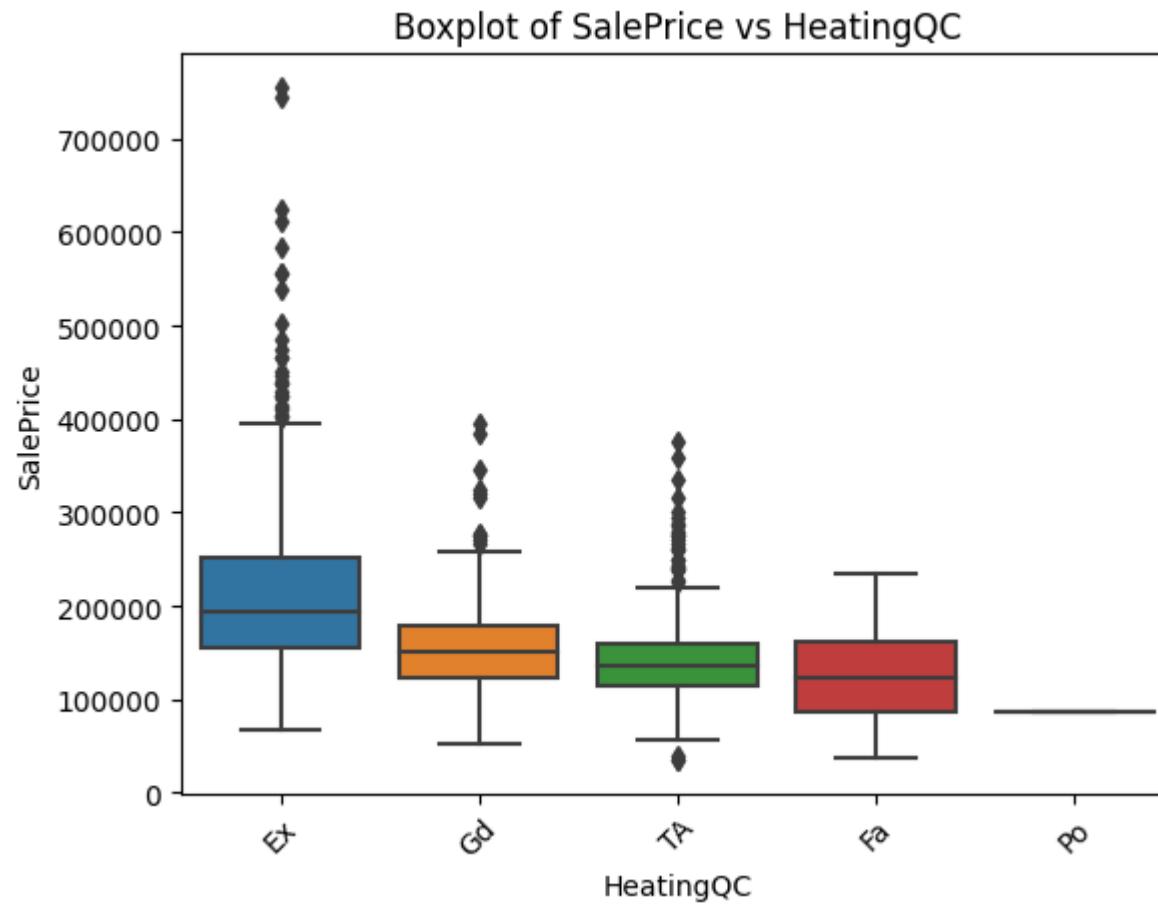


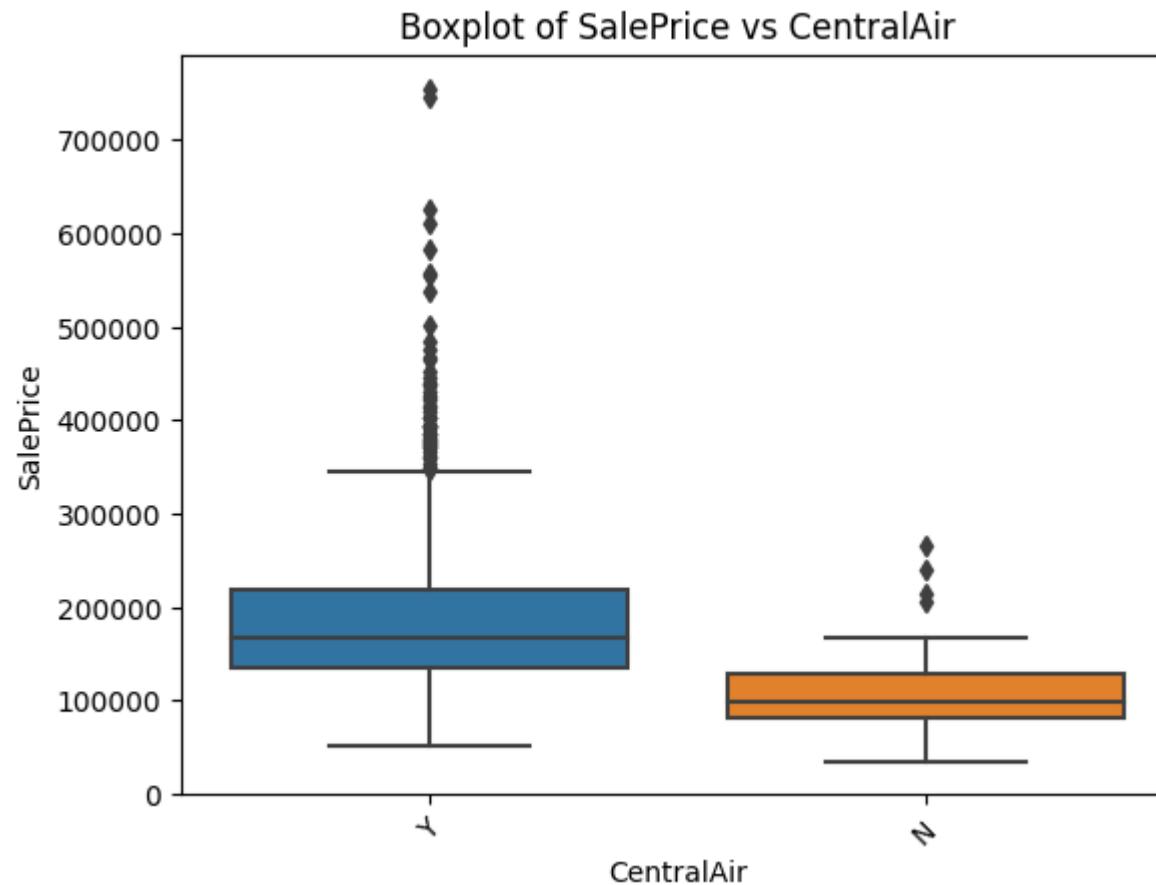


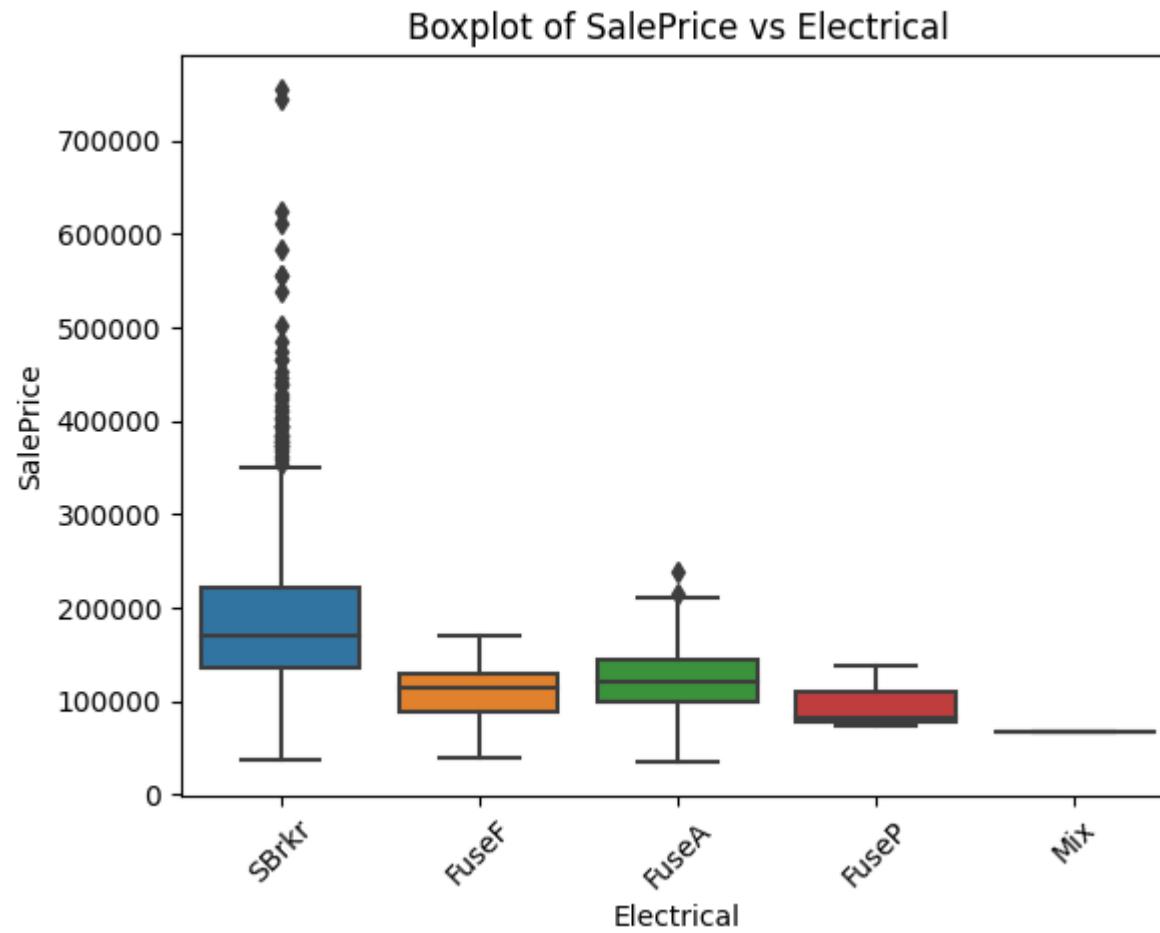


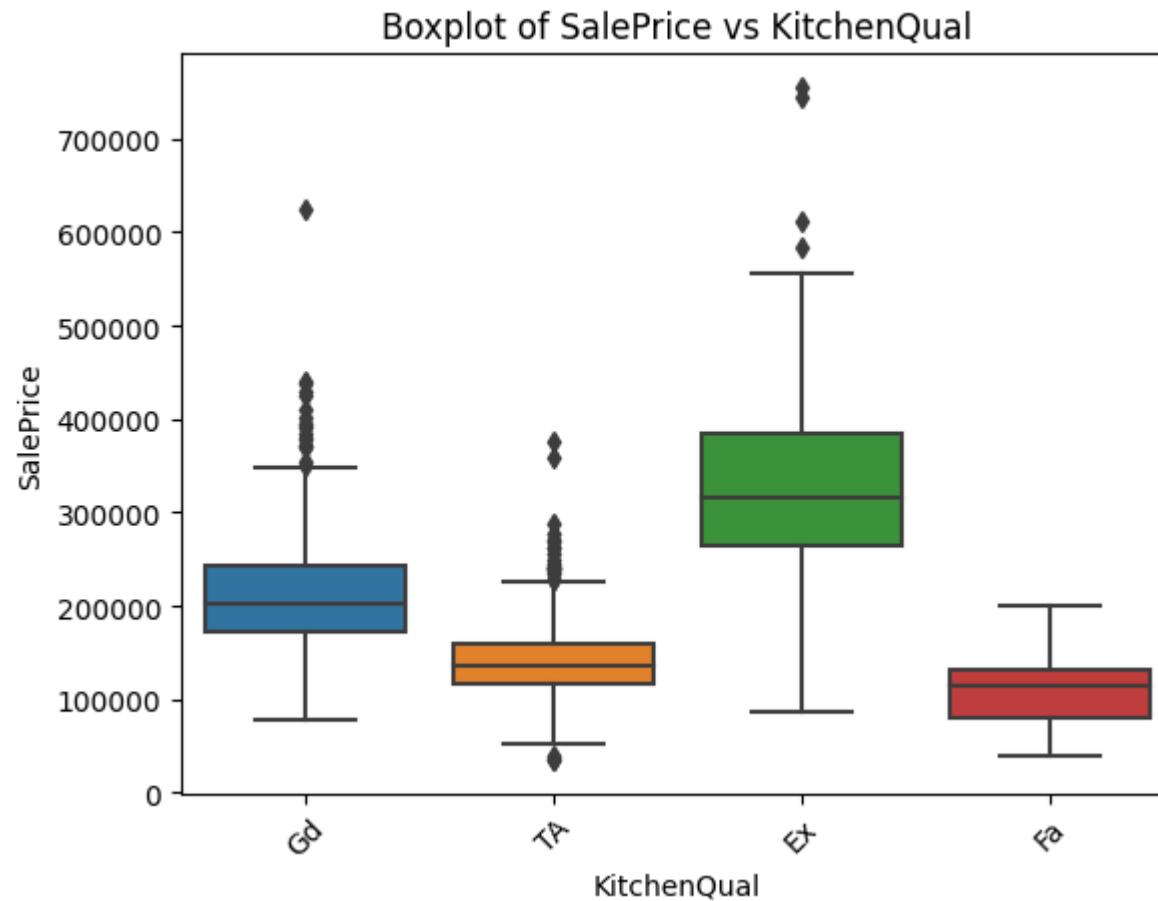




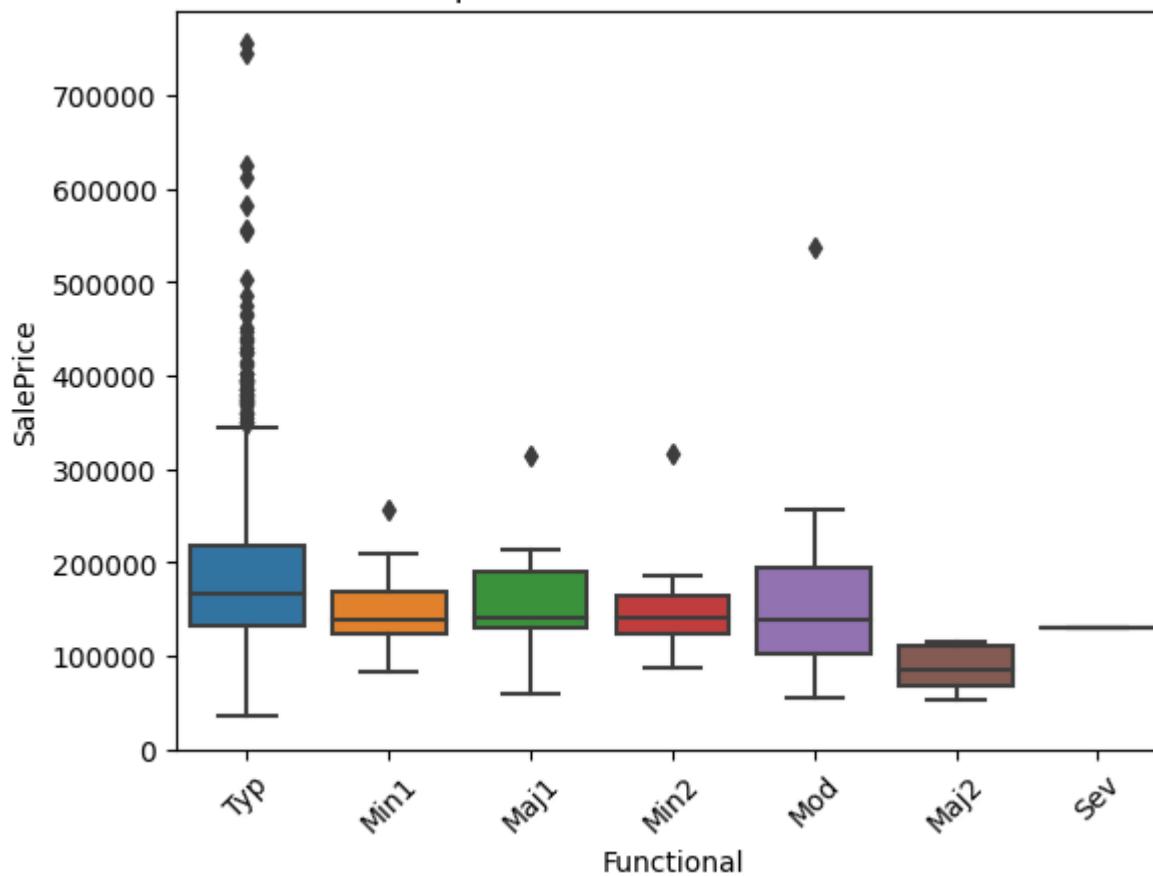


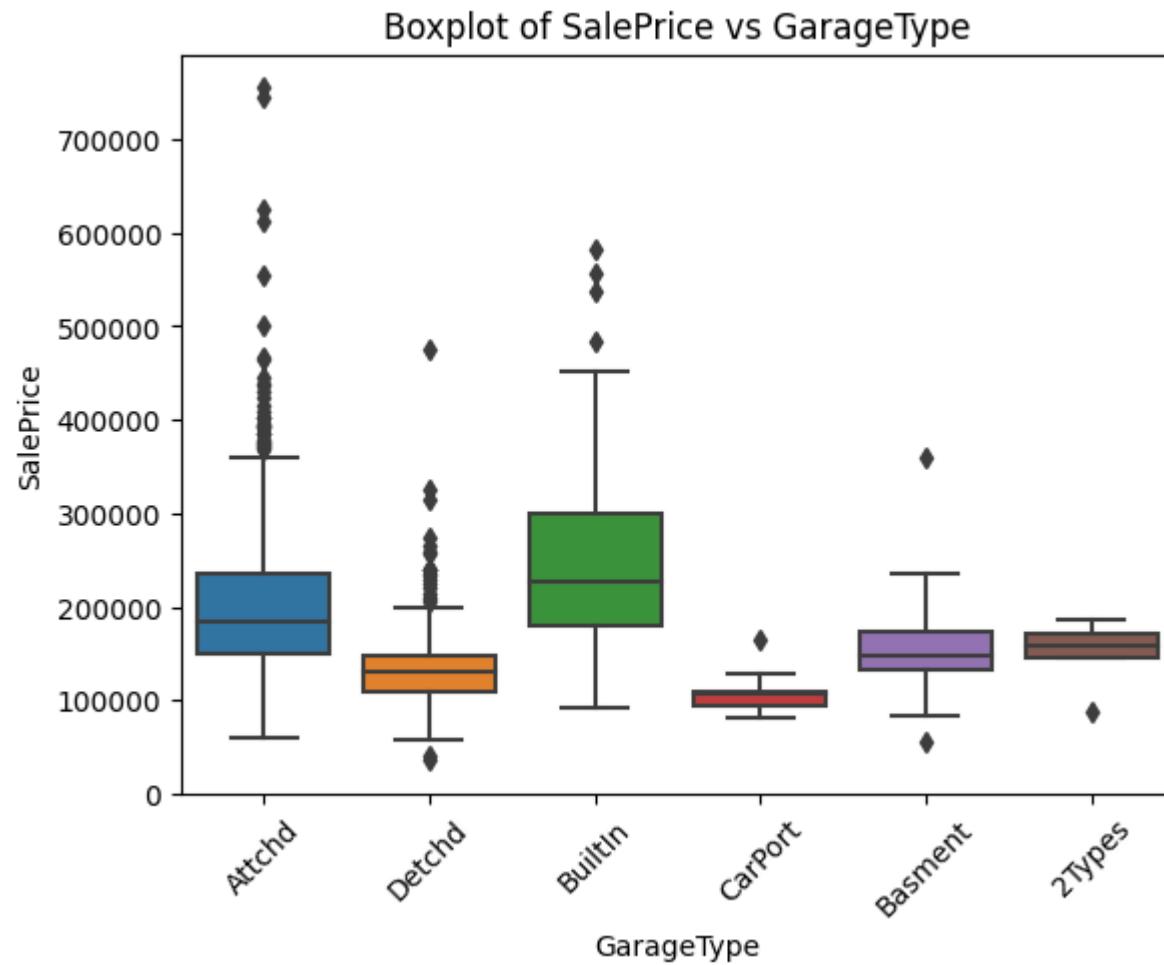


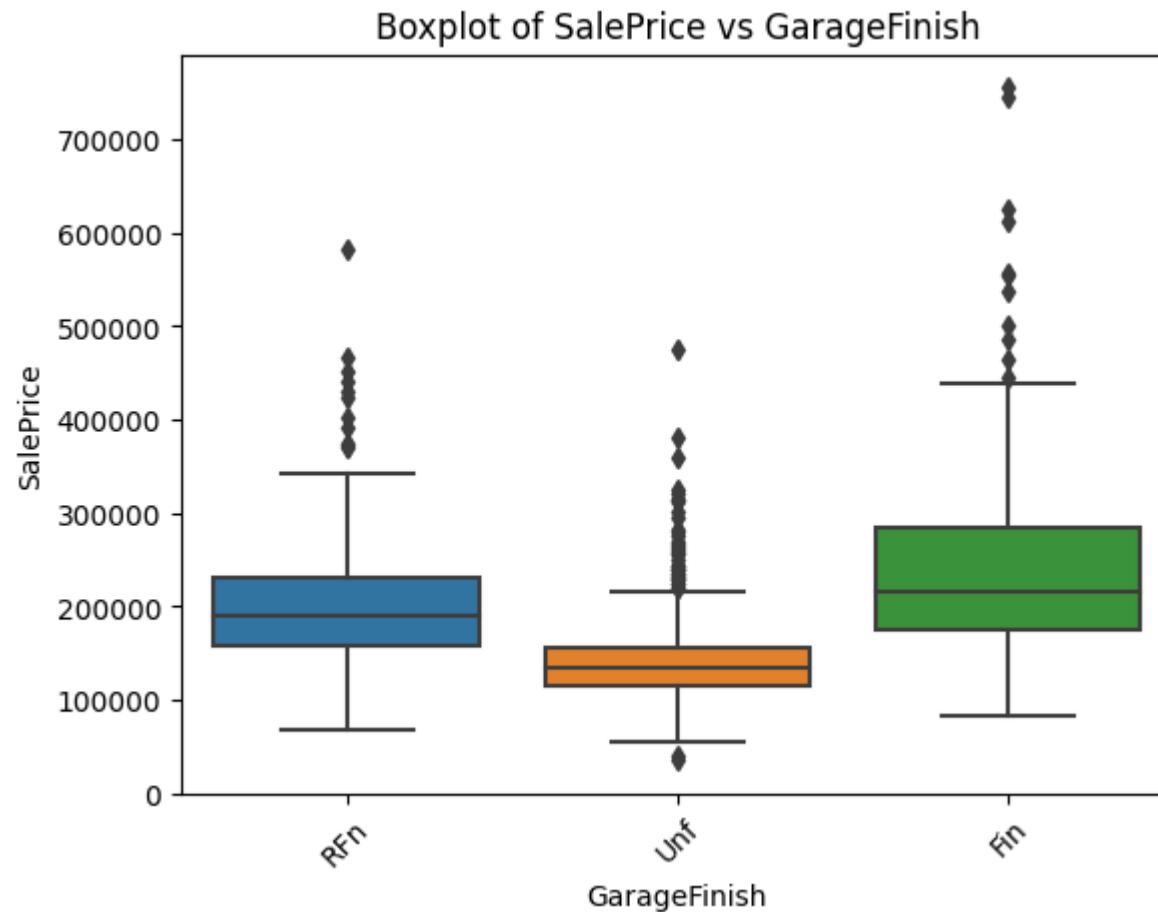


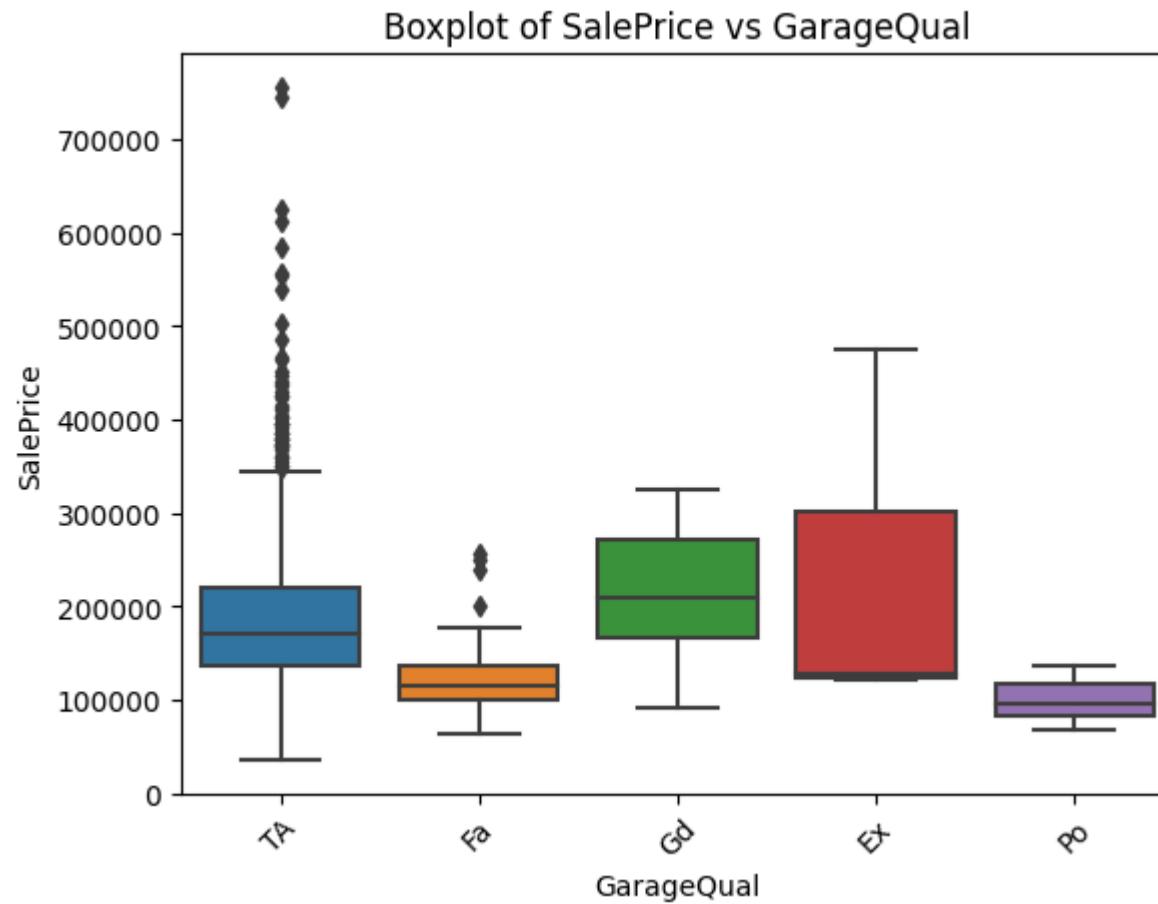


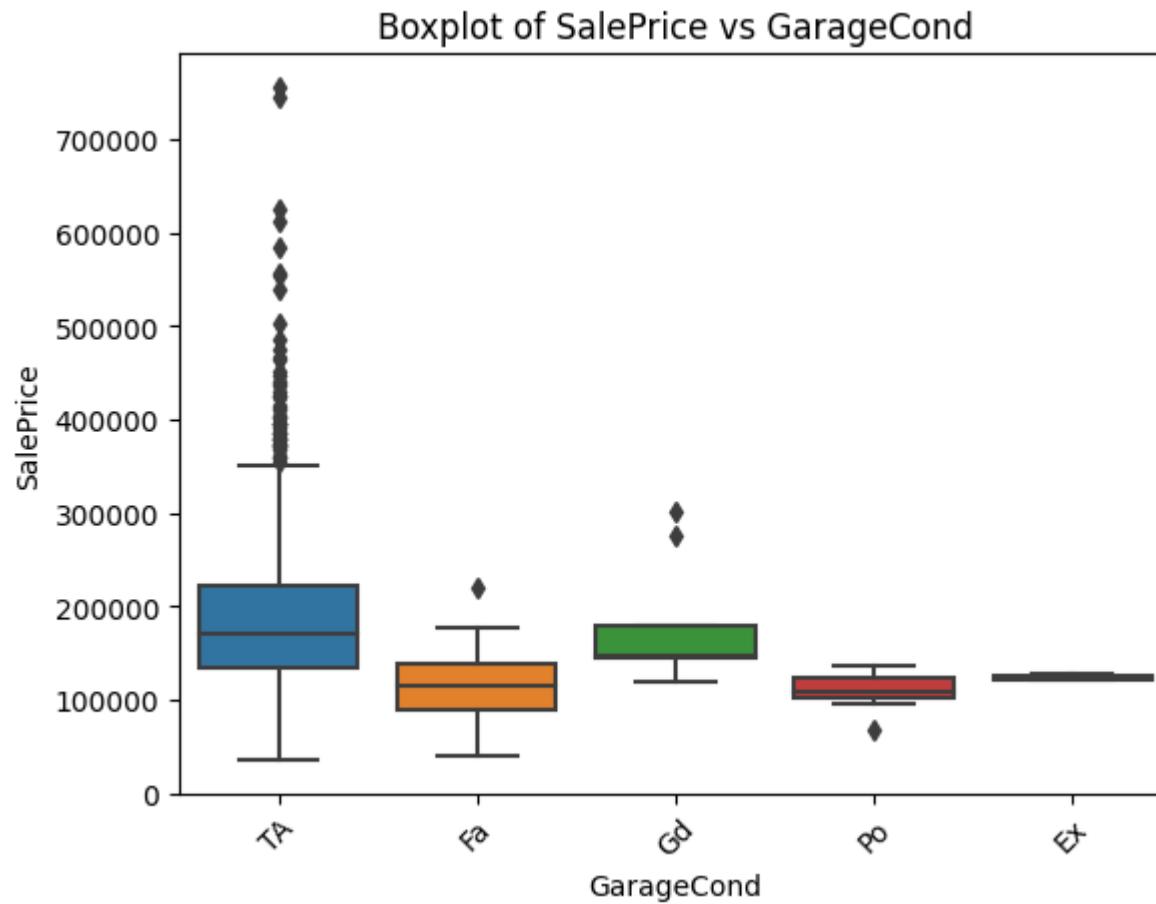
Boxplot of SalePrice vs Functional

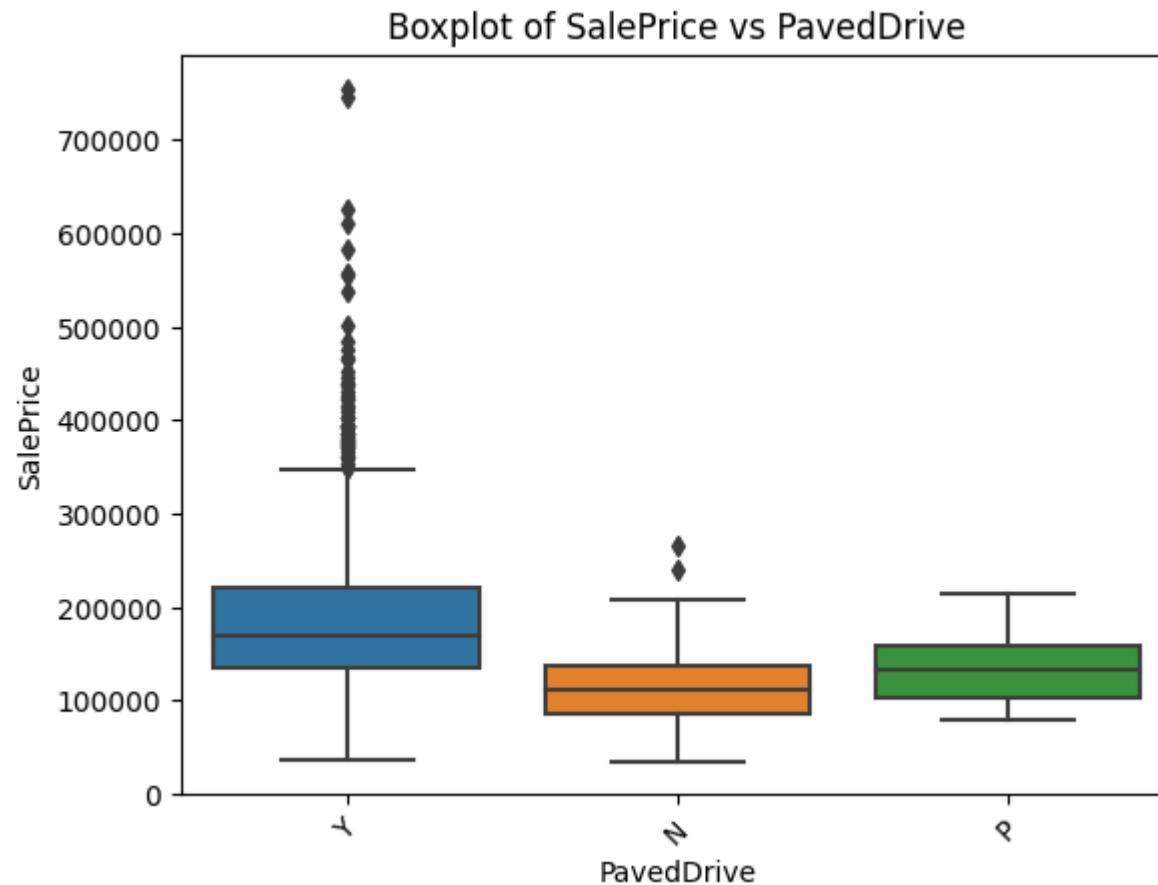


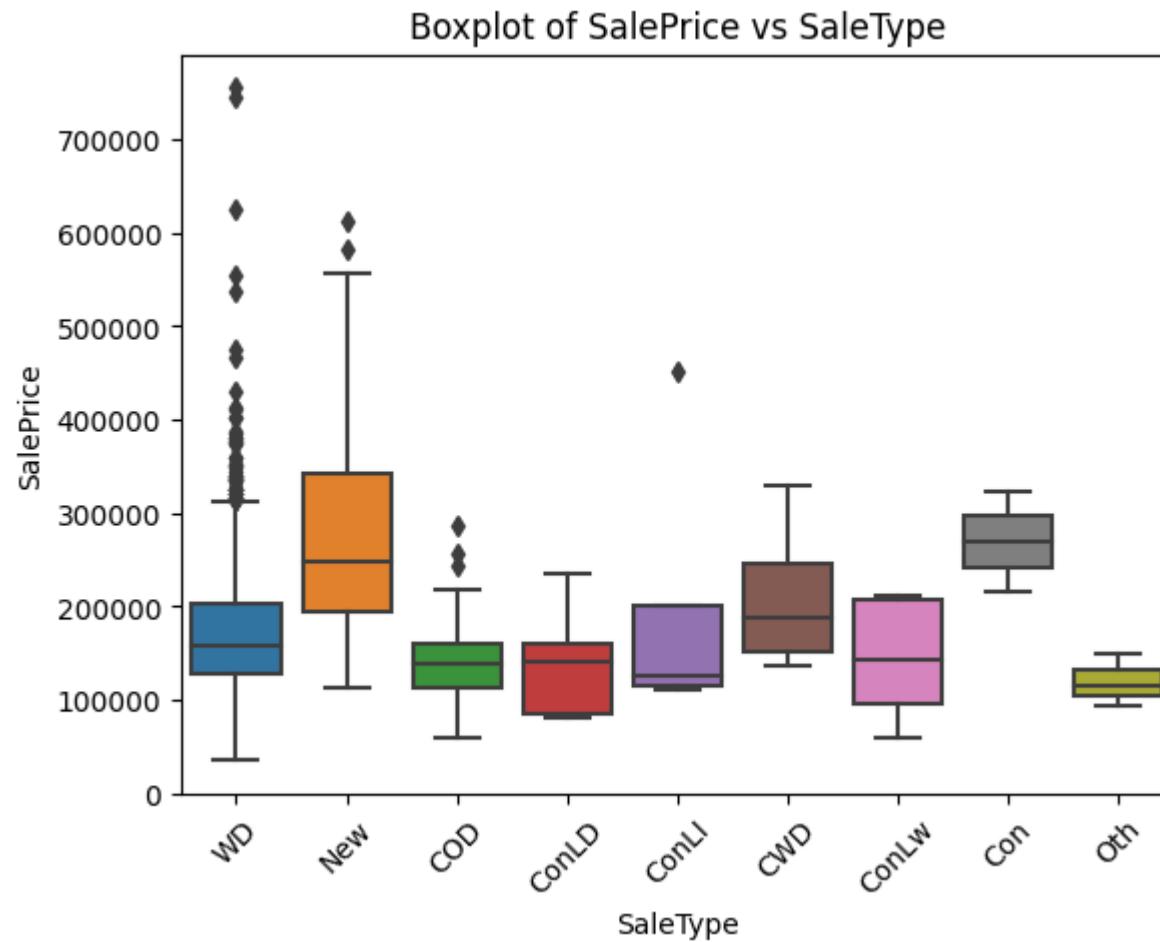


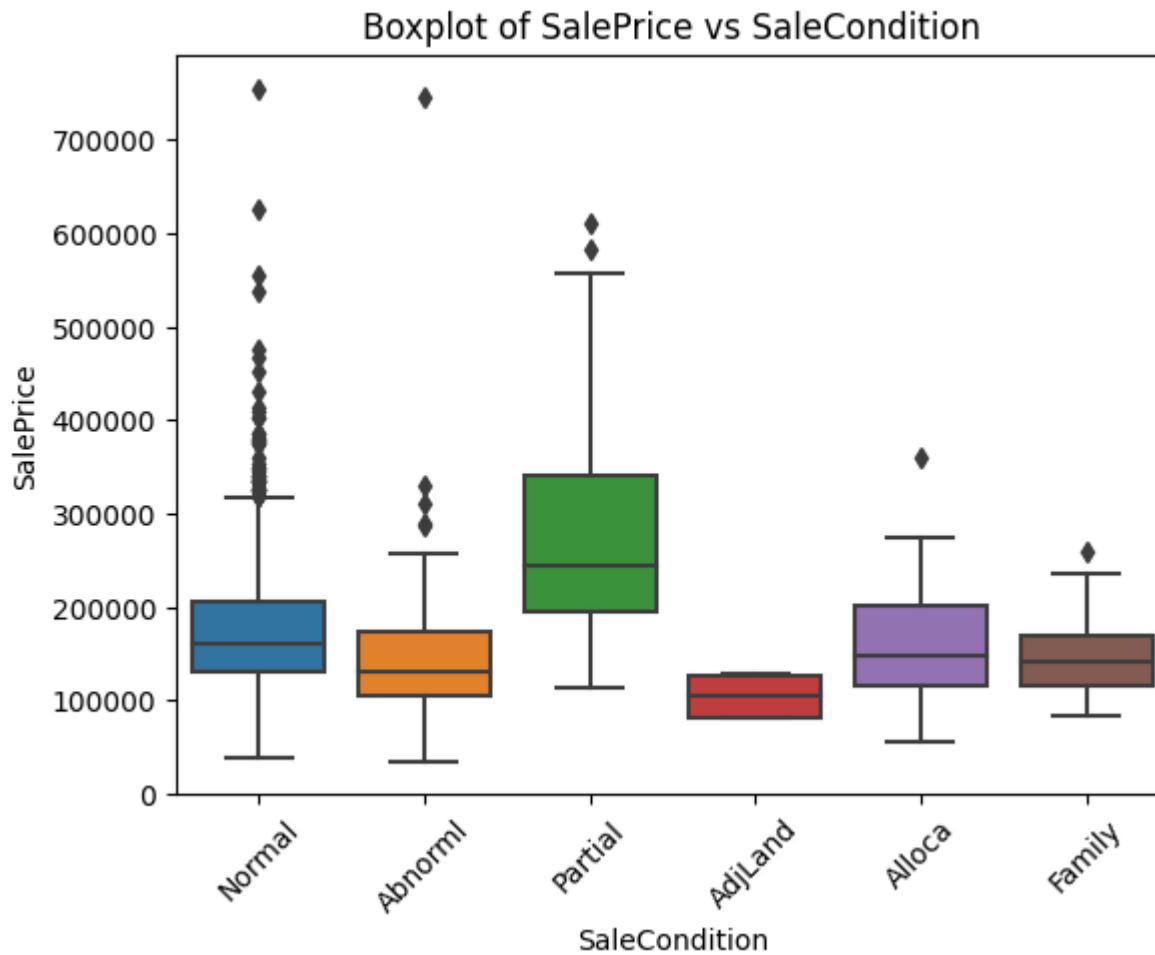












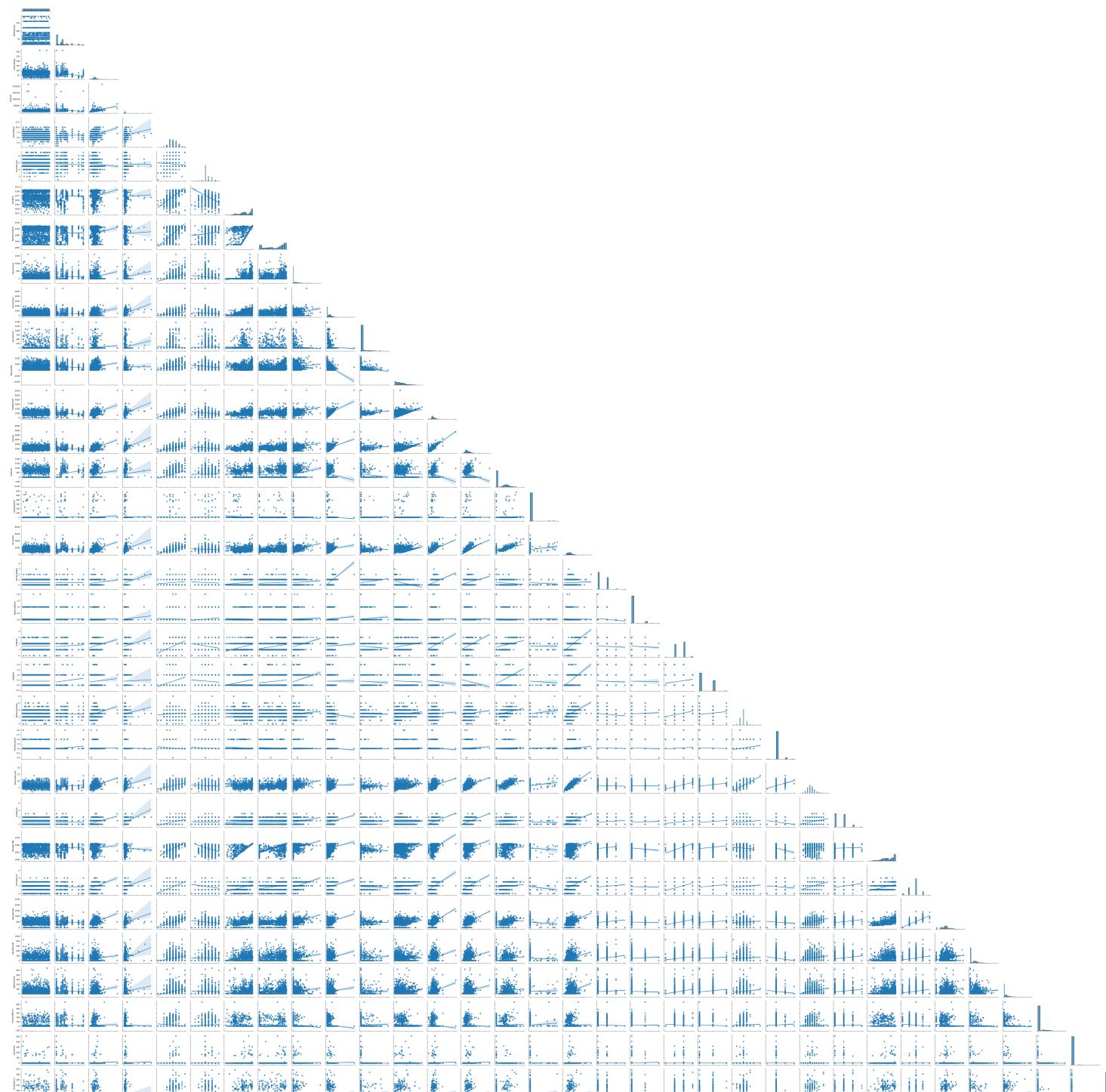
INSIGHTS:

- FV has maxm median value in MSZoning and RL has maxm no of outliers
- Pave has maxm median value in Street and Pave has maxm no of outliers
- IR2 has maxm median value in LotShape and Reg has maxm no of outliers
- HLS has maxm median value in LandContour and Lvl has maxm no of outliers

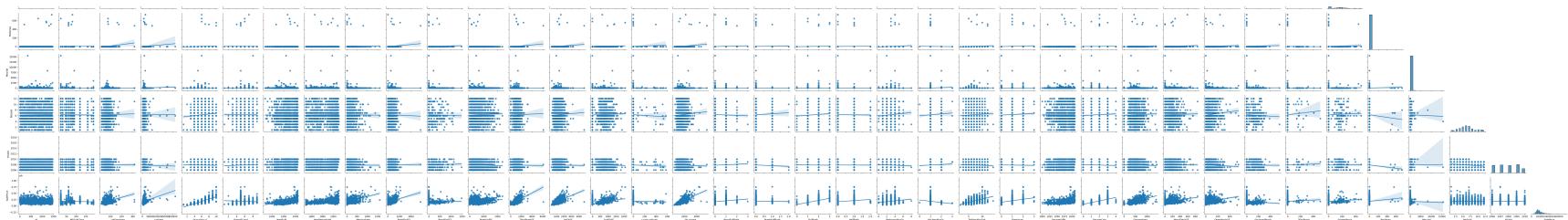
Multivariate Plots:

```
In [22]: sns.pairplot(data,corner=True,kind='reg')
```

```
Out[22]: <seaborn.axisgrid.PairGrid at 0x2130b699ab0>
```



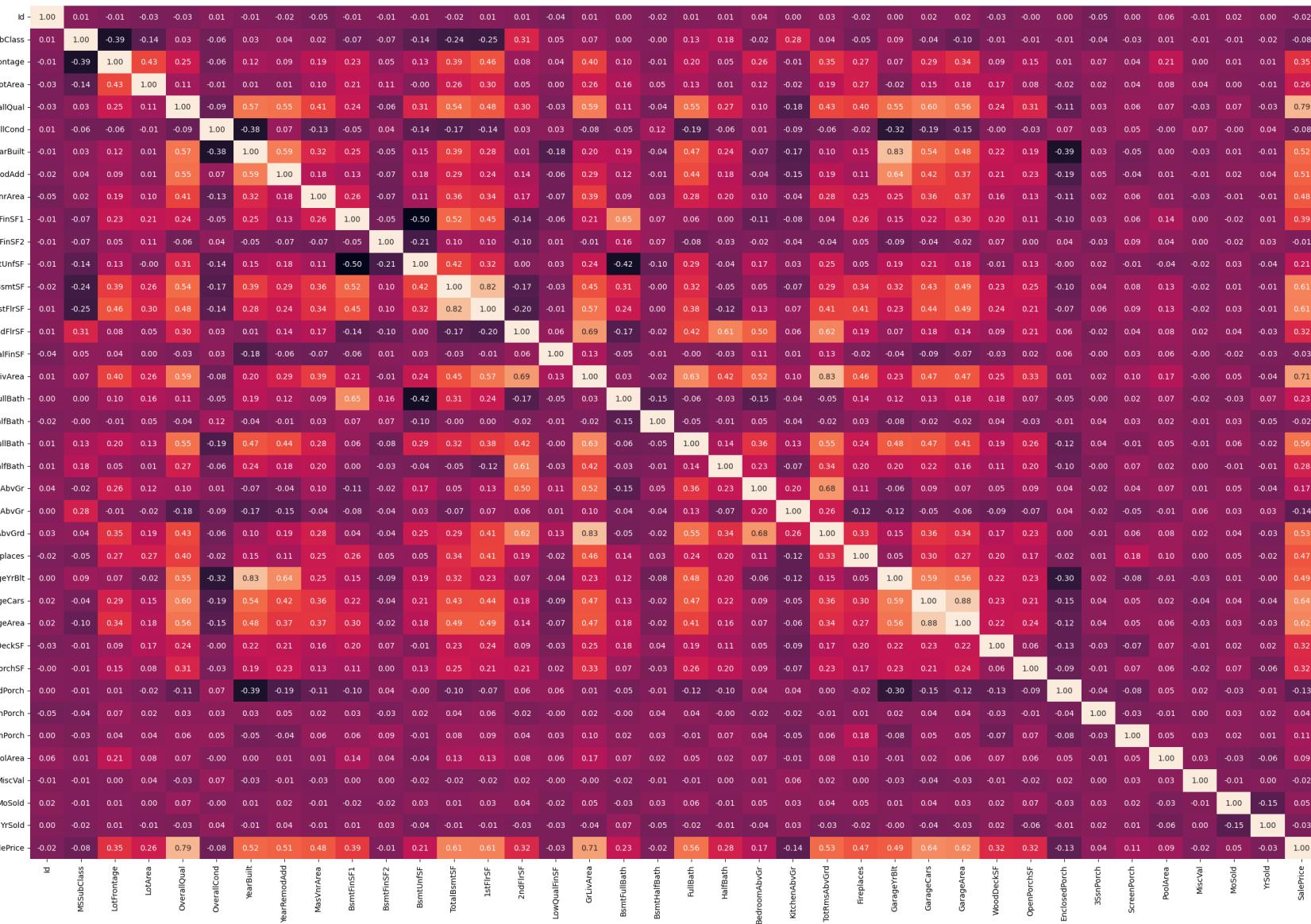
Eda_on_HousePriceDataset-TRAINING(14.08.2023)



```
In [23]: plt.figure(figsize=(30,20))
sns.heatmap(data.corr(), annot=True, cbar=False, fmt='.{2f}')
```

Out[23]: <AxesSubplot:>

Eda_on_HousePriceDataset-TRAINING(14.08.2023)



In [42]: `data.describe().T`

Out[42]:

	count	mean	std	min	25%	50%	75%	max
Id	1460.0	730.500000	421.610009	1.0	365.75	730.5	1095.25	1460.0
MSSubClass	1460.0	56.897260	42.300571	20.0	20.00	50.0	70.00	190.0
LotFrontage	1201.0	70.049958	24.284752	21.0	59.00	69.0	80.00	313.0
LotArea	1460.0	10516.828082	9981.264932	1300.0	7553.50	9478.5	11601.50	215245.0
OverallQual	1460.0	6.099315	1.382997	1.0	5.00	6.0	7.00	10.0
OverallCond	1460.0	5.575342	1.112799	1.0	5.00	5.0	6.00	9.0
YearBuilt	1460.0	1971.267808	30.202904	1872.0	1954.00	1973.0	2000.00	2010.0
YearRemodAdd	1460.0	1984.865753	20.645407	1950.0	1967.00	1994.0	2004.00	2010.0
MasVnrArea	1452.0	103.685262	181.066207	0.0	0.00	0.0	166.00	1600.0
BsmtFinSF1	1460.0	443.639726	456.098091	0.0	0.00	383.5	712.25	5644.0
BsmtFinSF2	1460.0	46.549315	161.319273	0.0	0.00	0.0	0.00	1474.0
BsmtUnfSF	1460.0	567.240411	441.866955	0.0	223.00	477.5	808.00	2336.0
TotalBsmtSF	1460.0	1057.429452	438.705324	0.0	795.75	991.5	1298.25	6110.0
1stFlrSF	1460.0	1162.626712	386.587738	334.0	882.00	1087.0	1391.25	4692.0
2ndFlrSF	1460.0	346.992466	436.528436	0.0	0.00	0.0	728.00	2065.0
LowQualFinSF	1460.0	5.844521	48.623081	0.0	0.00	0.0	0.00	572.0
GrLivArea	1460.0	1515.463699	525.480383	334.0	1129.50	1464.0	1776.75	5642.0
BsmtFullBath	1460.0	0.425342	0.518911	0.0	0.00	0.0	1.00	3.0
BsmtHalfBath	1460.0	0.057534	0.238753	0.0	0.00	0.0	0.00	2.0
FullBath	1460.0	1.565068	0.550916	0.0	1.00	2.0	2.00	3.0
HalfBath	1460.0	0.382877	0.502885	0.0	0.00	0.0	1.00	2.0
BedroomAbvGr	1460.0	2.866438	0.815778	0.0	2.00	3.0	3.00	8.0
KitchenAbvGr	1460.0	1.046575	0.220338	0.0	1.00	1.0	1.00	3.0
TotRmsAbvGrd	1460.0	6.517808	1.625393	2.0	5.00	6.0	7.00	14.0
Fireplaces	1460.0	0.613014	0.644666	0.0	0.00	1.0	1.00	3.0

	count	mean	std	min	25%	50%	75%	max
GarageYrBlt	1379.0	1978.506164	24.689725	1900.0	1961.00	1980.0	2002.00	2010.0
GarageCars	1460.0	1.767123	0.747315	0.0	1.00	2.0	2.00	4.0
GarageArea	1460.0	472.980137	213.804841	0.0	334.50	480.0	576.00	1418.0
WoodDeckSF	1460.0	94.244521	125.338794	0.0	0.00	0.0	168.00	857.0
OpenPorchSF	1460.0	46.660274	66.256028	0.0	0.00	25.0	68.00	547.0
EnclosedPorch	1460.0	21.954110	61.119149	0.0	0.00	0.0	0.00	552.0
3SsnPorch	1460.0	3.409589	29.317331	0.0	0.00	0.0	0.00	508.0
ScreenPorch	1460.0	15.060959	55.757415	0.0	0.00	0.0	0.00	480.0
PoolArea	1460.0	2.758904	40.177307	0.0	0.00	0.0	0.00	738.0
MiscVal	1460.0	43.489041	496.123024	0.0	0.00	0.0	0.00	15500.0
MoSold	1460.0	6.321918	2.703626	1.0	5.00	6.0	8.00	12.0
YrSold	1460.0	2007.815753	1.328095	2006.0	2007.00	2008.0	2009.00	2010.0
SalePrice	1460.0	180921.195890	79442.502883	34900.0	129975.00	163000.0	214000.00	755000.0

SCALING:

```
In [43]: mean=data['SalePrice'].mean()
          std=data['SalePrice'].std()
```

```
In [44]: mean
```

```
Out[44]: 180921.19589041095
```

```
In [45]: std
```

```
Out[45]: 79442.50288288662
```

```
In [46]: scaled=[]
          for ele in data['SalePrice']:
```

```
scale=(ele-mean)/std  
scaled.append(scale)
```

In [47]: scaled

```
Out[47]: [0.3471542701801007,  
 0.007285824194667163,  
 0.5359700735053415,  
 -0.5151045650051659,  
 0.8695446593799337,  
 -0.47734140434011774,  
 1.587044712015849,  
 0.24015864829579756,  
 -0.6422405392441614,  
 -0.7920344098821858,  
 -0.6472756273328345,  
 2.0653780804397925,  
 -0.464753684118435,  
 1.2408824059195742,  
 -0.3011133212365596,  
 -0.6158063267786277,  
 -0.4018150830100214,  
 -1.144490576089302,  
 -0.27593788079319415,  
 -0.5276922852268486,  
 1.817399992072643,  
 -0.5226571971381755,  
 0.6177902549462793,  
 -0.6422405392441614,  
 -0.33887648190160774,  
 0.9488472967765349,  
 -0.580560710157916,  
 1.5744569917941662,  
 0.334566549958418,  
 -1.4151265608554806,  
 -1.7738765871734383,  
 -0.3974093809324324,  
 -0.012854528160025195,  
 -0.19411769935225642,  
 1.2157069654762087,  
 1.6122201524592146,  
 -0.4521659638967523,  
 -0.3514642021232905,  
 -0.9053238918773303,  
 -1.2451923378627638,  
 -0.2633501605715114,  
 -0.13747295835468415,  
 -0.464753684118435,
```

-0.6378348371665724,
-0.5025168447834831,
1.7494263028755561,
0.7397149130134981,
0.8657683433134289,
-0.8549730109905994,
-0.6787449278870413,
-0.04935891680290509,
-0.8360914306580753,
-0.8927361716556476,
2.5688868893071013,
-0.6409817672219931,
-0.00530189602701556,
-0.10600365780047735,
0.19610162751990803,
3.2458544828291984,
-0.705179140352575,
-0.28852560101487684,
-1.006025653650792,
0.27162794885000435,
-0.5151045650051659,
0.4856191926186107,
1.7129219142326764,
-0.011595756137856923,
0.5674393740595484,
-1.2703677783061293,
0.5548516538378656,
0.7940183380498375,
-0.6472756273328345,
0.051342844970556696,
-0.4534247359189205,
-0.9254642442320227,
-1.1319028558676194,
-0.5686023759473174,
-0.6787449278870413,
-0.5591615857810555,
-0.8927361716556476,
0.15833846685485986,
-0.3451703420124491,
0.8066060582715201,
-0.6850387879978826,
-0.15635453868720825,
0.995421861596761,

-0.08712207746795327,
-0.20670541957393915,
-1.2074291771977157,
-0.7215431766407625,
-0.8939949436778158,
-1.0362361821828308,
-0.21929313979562187,
-0.5918896583574305,
0.2999503193487905,
0.051342844970556696,
0.41638673139935567,
-1.0846989050363092,
-1.2326046176410812,
-0.65419887345476,
0.3030972494042112,
-0.03677119658122237,
-0.7798998475884836,
0.22631215605194657,
-0.14376681846552553,
0.8695446593799337,
-1.0186133738724747,
-0.829797570547234,
-0.829797570547234,
0.11428144607897031,
-0.5541264976923823,
-0.011595756137856923,
2.5559215374787683,
0.45414989206440387,
0.9891280014859196,
-0.06194663702458782,
-0.5276922852268486,
-0.32628876167992504,
1.7506850748977245,
-0.21312515688699735,
-0.011595756137856923,
-1.0186133738724747,
-0.5654554458918968,
-0.340135253923776,
0.0009919640838258015,
-1.213723037308557,
-0.6661572076653586,
-1.1822537367543502,
-0.32628876167992504,

-0.38922736278833864,
0.5674393740595484,
0.7940183380498375,
-0.3797865726220766,
0.491913052729452,
-0.011595756137856923,
-0.08712207746795327,
-0.47734140434011774,
-0.12488523813300144,
0.6177902549462793,
0.6366718352788033,
-0.829797570547234,
0.995421861596761,
-0.18782383924141505,
0.2905095291825284,
-0.7039203683304067,
-0.6409817672219931,
-0.9556747727640612,
0.5233823532836589,
-0.5025168447834831,
-0.829797570547234,
-0.7416835289954549,
2.4103067899543427,
0.11428144607897031,
0.6807288560546929,
-0.7039203683304067,
-1.282955498527812,
-0.899030031766489,
1.115005203702747,
0.9312244884661791,
1.7506850748977245,
-0.2318808600173046,
2.9150491954033764,
0.491913052729452,
-0.9783326691630901,
-0.3640519223449732,
-0.6724510677761999,
0.11428144607897031,
1.8214784134244681,
0.03246126463803261,
0.5926148145029139,
-0.6598633475545173,
0.4289744516210384,

0.7310797369414238,
-0.22558699990646322,
0.03875512474887397,
0.7814306178281547,
0.3786235707343075,
-0.10600365780047735,
4.039598356847846,
-1.0186133738724747,
-0.04935891680290509,
0.24141742031796584,
-0.7668589694388204,
0.24015864829579756,
-0.6787449278870413,
3.701781709258547,
-0.09970979768963599,
-0.5780431661135795,
-0.3472221404085834,
1.3227025873605118,
1.6877464737893109,
0.03875512474887397,
0.13945688652233576,
-0.6409817672219931,
-0.6787449278870413,
-0.4081089431208627,
1.6483720849358872,
0.6807288560546929,
-0.968262492985744,
1.1829788928998335,
-0.5151045650051659,
-0.11859137802216008,
-0.8675607312122822,
-0.4018150830100214,
-0.8927361716556476,
-0.00530189602701556,
-0.46601245614060327,
-0.5025168447834831,
1.2094131053653674,
-0.4521659638967523,
-1.0437888143158403,
0.06393056519223943,
0.9032545741336001,
-0.3137010414582423,
-0.24132165018356663,

-0.584966412235505,
0.36603585051262477,
-0.9304993323206958,
1.6436894530134212,
-0.17221506616652849,
0.3018384773820429,
0.24015864829579756,
-0.012854528160025195,
-1.0563765345375231,
2.5846215395842047,
-0.8675607312122822,
1.3730534682472426,
-0.9430870525423785,
-0.7039203683304067,
0.14575074663317714,
-0.4144028032317041,
2.7954658532973906,
-1.0878458350917297,
-0.663639663621022,
0.4478560319535625,
-1.1507844362001434,
0.05763670508139806,
0.17092618707654259,
1.725509634454359,
-0.8549730109905994,
1.0268911621509678,
-0.8864423115448062,
-1.282955498527812,
-0.7668589694388204,
0.3030972494042112,
0.7625490374956306,
-0.552867725670214,
-0.5151045650051659,
-0.011595756137856923,
1.2094131053653674,
-1.3144247990820188,
0.6807288560546929,
-0.09970979768963599,
-0.28852560101487684,
-0.4521659638967523,
0.6177902549462793,
0.334566549958418,
0.491913052729452,

0.6366718352788033,
-1.0563765345375231,
-0.06194663702458782,
1.1968253851436845,
-0.37663964256665594,
-0.6409817672219931,
-1.3584818198579083,
-0.06824049713542918,
0.051342844970556696,
-0.017889616248698285,
-0.760565109327979,
-0.4144028032317041,
1.0709481829268575,
0.7625490374956306,
1.3730534682472426,
-0.5276922852268486,
-0.7102142284412482,
0.3030972494042112,
0.2527463685174803,
-0.5025168447834831,
2.9502696365836445,
0.13945688652233576,
0.5989086746137552,
0.051342844970556696,
0.334566549958418,
0.8015709701828471,
-0.021665932315203102,
-0.2041878755296026,
-0.27593788079319415,
-1.1696660165326676,
-0.7416835289954549,
-0.34422626299582293,
0.6584485912623145,
-0.5667142179140651,
-0.6283940470003104,
0.6807288560546929,
-0.17523611901973232,
-0.48363526445095906,
-0.3640519223449732,
0.7310797369414238,
-0.07453435724627054,
-0.28223174090403547,
-0.3011133212365596,

1.08353590314854,
0.3030972494042112,
-0.3904861348105069,
1.4359920693556563,
1.573198219771998,
0.5548516538378656,
-1.1507844362001434,
-1.2388984777519225,
2.2541938837650335,
-0.19285892733008814,
-0.6158063267786277,
-0.7681177414609887,
2.4430096870902744,
-0.03677119658122237,
0.09539986574644622,
0.995421861596761,
1.1212990638135882,
0.995421861596761,
0.08281214552476351,
2.0357088238772865,
2.178667562434937,
1.5115183906857528,
-0.6891297970699295,
0.768842897606472,
-1.1822537367543502,
1.8010359557844553,
-0.44901903384133157,
0.42268059151019705,
-1.2955432187494949,
-0.7794466896605031,
-0.5276922852268486,
1.2975271469171463,
0.3282726898475766,
0.13945688652233576,
0.6045731487135124,
2.4735474963480764,
0.41638673139935567,
0.27162794885000435,
-0.32628876167992504,
0.27666303693867744,
-1.2451923378627638,
-1.1759598766435089,
1.0709481829268575,

-1.2074291771977157,
-0.5125870209608293,
-0.37034578245581457,
-0.2948194611257182,
-0.33887648190160774,
3.2253868497487423,
1.7262774853878817,
0.11428144607897031,
-1.0815519749808884,
-0.9443458245645467,
-0.5151045650051659,
-0.04306505669206373,
-0.09970979768963599,
-0.5906308863352622,
-0.6409817672219931,
1.2471762660304155,
-0.3137010414582423,
-0.4521659638967523,
0.22127706796327348,
-0.7920344098821858,
0.11428144607897031,
-0.42699052345338684,
-0.27593788079319415,
-0.20041155946309777,
-0.6158063267786277,
-0.23817472012814594,
-0.10726242982264562,
-0.5851929911994953,
-0.7039203683304067,
-0.7290958087737722,
0.4856191926186107,
-1.509534462518101,
-0.4144028032317041,
2.002439479331379,
2.687614266438013,
-0.024183476359539647,
-0.6787449278870413,
0.08595907558018419,
0.41009287128851435,
-1.3207186591928601,
0.7436674571631066,
0.13945688652233576,
-1.2577800580844467,

-0.7039203683304067,
0.12686916630065304,
3.084983418396093,
-0.7794466896605031,
0.4289744516210384,
-0.9367931924315371,
-1.0186133738724747,
-0.9053238918773303,
-0.6535694874436758,
-0.7290958087737722,
-0.14376681846552553,
-1.4340081411880048,
0.7562551773847892,
0.8128999183823615,
-0.20053743666531462,
-0.9179116120990131,
0.9702464211533955,
-0.1626483987980496,
-0.38922736278833864,
-0.829797570547234,
-0.04935891680290509,
1.2471762660304155,
1.9992925492759583,
-1.5221221827397837,
-0.4521659638967523,
0.5170884931728175,
-0.829797570547234,
0.5926148145029139,
0.0026787185935312867,
-0.39552122289918,
0.7310797369414238,
-0.691332648108724,
-0.48992912456180043,
0.31946128569239873,
0.4289744516210384,
-0.8549730109905994,
1.6877464737893109,
-0.5276922852268486,
-0.5780431661135795,
1.1842376649220019,
-0.9052231901155569,
0.18225513527605702,
-0.07453435724627054,

-1.2023940891090426,
-1.2716265503282975,
-0.7353896688846135,
0.0009919640838258015,
-1.2577800580844467,
0.3912112909559902,
-0.8172098503255513,
-0.7794466896605031,
-1.140084874011713,
-0.8927361716556476,
4.708799326993165,
-0.7920344098821858,
-0.2268457719286315,
-0.10600365780047735,
0.36603585051262477,
-0.6724510677761999,
0.11428144607897031,
0.2388998762736293,
-0.7731528295496617,
-0.7668589694388204,
-0.8927361716556476,
1.2471762660304155,
0.2905095291825284,
0.36603585051262477,
0.08910600563560486,
-0.06824049713542918,
-1.0437888143158403,
0.9450709807100302,
-0.2507624403498287,
-0.8927361716556476,
1.0386606805582412,
-0.32628876167992504,
-1.492125645451514,
0.09791740979078277,
-0.7165080885520895,
-0.027456283617177155,
-0.17523611901973232,
-0.43328438356422816,
0.8695446593799337,
0.07651828541392215,
0.3912112909559902,
0.11428144607897031,
-0.4144028032317041,

3.2612115014996514,
0.8821323796016165,
-0.6095124666677864,
0.3521893582687738,
2.505948288198688,
1.4611675097990218,
-1.1511494800865723,
1.8262113962278208,
2.4304219668685914,
-0.32628876167992504,
-0.2129992796847805,
-0.6095124666677864,
-0.42699052345338684,
-0.3137010414582423,
-0.07453435724627054,
-0.2633501605715114,
-1.194841456976033,
-0.829797570547234,
-0.6032186065569449,
-0.10241615753729777,
-0.32628876167992504,
-1.1281265398011144,
-1.8380739603040201,
3.135334299282824,
0.03875512474887397,
-0.6409817672219931,
-0.7668589694388204,
-0.8549730109905994,
0.5762507782147263,
-0.5151045650051659,
1.36046574802556,
-0.42699052345338684,
-0.7102142284412482,
0.4289744516210384,
0.3446367261357642,
-0.2507624403498287,
-0.7102142284412482,
-0.20167033148526606,
0.27370492268658203,
-0.6422405392441614,
-0.5906308863352622,
-1.0626703946483644,
2.7937161601865768,

-0.28852560101487684,
1.0583604627051746,
0.3786235707343075,
0.6681411358330102,
-0.9399401224869578,
-0.38922736278833864,
-0.27593788079319415,
0.048195914915136015,
1.697187263955573,
-0.06194663702458782,
-0.6158063267786277,
3.340023217807607,
-1.194841456976033,
0.24801338571412757,
-0.07453435724627054,
-0.6661572076653586,
-0.9242054722098544,
-1.7826879913286162,
-0.03677119658122237,
-0.9242054722098544,
0.08910600563560486,
-0.8770015213785441,
-0.28852560101487684,
1.1464745042569537,
1.6877464737893109,
0.8443692189365684,
0.40694594123309363,
-0.6032186065569449,
-0.015812642412120635,
0.6052025347245965,
0.36603585051262477,
-0.6472756273328345,
-0.7039203683304067,
1.033185022261809,
-0.5151045650051659,
-0.8612668711014408,
0.9387771205991887,
-0.9179116120990131,
1.2975271469171463,
-0.8549730109905994,
-0.5025168447834831,
-0.9179116120990131,
-0.07453435724627054,

0.6681411358330102,
-0.7479773891062963,
-0.13747295835468415,
-0.9179116120990131,
0.051342844970556696,
1.0961236233702227,
-0.6661572076653586,
1.8136236760061382,
0.41638673139935567,
1.7078868261440032,
-0.5659589547007641,
-0.4823764924287908,
-0.7668589694388204,
0.5485577937270243,
-0.13747295835468415,
-0.5276922852268486,
-0.7857405497713444,
-0.4521659638967523,
-0.20670541957393915,
-0.43957824367506954,
-0.622100186889469,
0.012320912283340254,
0.910996022069935,
-0.7857405497713444,
1.8136236760061382,
-0.6032186065569449,
2.3788123139596924,
-0.6409817672219931,
-0.552867725670214,
-0.47734140434011774,
-1.2766616384169707,
0.06267179317007116,
3.41163475814876,
-0.5402800054485313,
-0.5151045650051659,
-0.8927361716556476,
1.7380973546760417,
-0.8360410797771887,
0.16716245873025945,
0.46044375217524525,
-0.37663964256665594,
1.1842376649220019,
-0.5025168447834831,

0.491913052729452,
-0.37663964256665594,
0.5045007729511347,
0.3030972494042112,
-0.3640519223449732,
0.5548516538378656,
2.242864935565519,
-0.7857405497713444,
1.6625710333459454,
-0.4144028032317041,
1.0143034419292851,
-0.42699052345338684,
-1.3270125193037017,
-0.5465738655593727,
0.028684948571527796,
-0.9493809126532199,
1.685392570107856,
1.5618692715724836,
-1.4340081411880048,
0.7436674571631066,
-0.5780431661135795,
-0.15635453868720825,
-0.19852340142984537,
-0.2633501605715114,
-0.5163633370273342,
-0.3514642021232905,
-0.5780431661135795,
-0.15635453868720825,
-0.7165080885520895,
0.35974199040178345,
-1.2388984777519225,
-0.5226571971381755,
-0.464753684118435,
0.24015864829579756,
-1.5221221827397837,
-1.106727415424254,
-1.2074291771977157,
1.052834453527856,
1.171649944700319,
0.5674393740595484,
2.0653780804397925,
-0.3640519223449732,
2.391123104336498,

-0.474194474284697,
-1.0400124982493355,
-0.32628876167992504,
-0.32628876167992504,
-1.213723037308557,
0.3150555836148098,
-0.9179116120990131,
0.12686916630065304,
-0.5780431661135795,
2.128316681548206,
-1.1696660165326676,
-0.4458721037859109,
-0.4018150830100214,
-1.0500826744266816,
-0.17523611901973232,
0.21372443583026385,
2.782878133075708,
-0.8927361716556476,
-0.5465738655593727,
3.047220257731045,
0.6240841150571207,
-0.6535694874436758,
0.15833846685485986,
-0.1626483987980496,
-0.5465738655593727,
-0.09341593757879463,
-0.973297581074417,
-0.20041155946309777,
0.9639525610425542,
-0.5151045650051659,
-0.4081089431208627,
-1.1822537367543502,
-0.899030031766489,
2.4115403865360676,
-0.6598633475545173,
-0.47734140434011774,
-0.27047481021698383,
-0.09970979768963599,
1.3101148671388292,
0.5045007729511347,
0.334566549958418,
0.5910413494752035,
-0.4043326270543579,

2.6570009308588807,
0.17344373112087913,
-0.5025168447834831,
7.22634337132971,
1.9395008782229652,
-0.9118695063926053,
-0.4962229846726418,
-0.06194663702458782,
-1.157078296310985,
-0.7228019486629308,
-0.53398614533769,
0.18980776740906666,
1.656277173235104,
-0.5151045650051659,
2.2783497188704427,
-0.5151045650051659,
0.403799011177673,
-1.5850607838481974,
1.5241061109074354,
0.9198955402666646,
-0.017386107439830975,
-0.8939949436778158,
-1.6228239445132455,
-0.9836698625370836,
0.10169372585728759,
-0.6535694874436758,
-0.6346879071111518,
-0.20041155946309777,
-0.26964402068235277,
-0.3011133212365596,
2.0150271995530615,
-0.6598633475545173,
1.1842376649220019,
-0.47734140434011774,
-0.7102142284412482,
-0.5780431661135795,
1.7506850748977245,
-0.760565109327979,
0.5170884931728175,
0.17092618707654259,
-0.8927361716556476,
-0.9808502132074267,
0.699610436387217,

0.08281214552476351,
0.5233823532836589,
-0.6233589589116373,
-0.9179116120990131,
-0.22558699990646322,
-1.1004335553134126,
0.7424086851409383,
-0.024183476359539647,
0.11428144607897031,
-0.6158063267786277,
-0.48992912456180043,
-0.024183476359539647,
-0.07453435724627054,
-0.011595756137856923,
1.4964131264197333,
0.6933165762763757,
1.070683840802202,
1.000456949685434,
-1.0437888143158403,
-1.0626703946483644,
-0.23817472012814594,
0.45414989206440387,
1.1905315250328432,
-0.3137010414582423,
-0.10600365780047735,
0.3912112909559902,
-0.2771966528153624,
-0.019148388270866557,
1.3730534682472426,
-0.6724510677761999,
-1.0186133738724747,
0.43149199566537494,
1.9646763186663307,
1.1212990638135882,
1.047434321552754,
0.19610162751990803,
-0.2633501605715114,
0.4520980936682696,
4.494808083224559,
-0.5793019381357478,
-0.9934379334291094,
-0.9304993323206958,
-0.8360914306580753,

2.6947640915239286,
-0.23817472012814594,
0.5107946330619761,
-0.48363526445095906,
-0.464753684118435,
-0.5780431661135795,
-0.06194663702458782,
-0.06320540904675609,
0.07777705743609042,
-0.19411769935225642,
-0.6661572076653586,
-0.24446858023898732,
-0.5276922852268486,
0.6555534156113274,
-0.9191703841211813,
0.08281214552476351,
-0.2608326165271749,
-0.42950806749772336,
1.1186556425670349,
0.5548516538378656,
0.17092618707654259,
-0.12488523813300144,
-0.47104754422927636,
-0.8927361716556476,
3.8276589114753743,
-0.07453435724627054,
0.24015864829579756,
-0.8939949436778158,
0.10169372585728759,
5.060412115945429,
-0.7920344098821858,
0.5885867440319753,
-0.5717493060027381,
0.5359700735053415,
-0.26397954658259554,
-0.9430870525423785,
0.0009919640838258015,
-0.4584598240075936,
-1.5725611776680666,
-0.28978437303704513,
-0.8172098503255513,
0.5535928818156974,
-0.552867725670214,

1.1338867840352709,
-0.32628876167992504,
0.542263933616183,
0.02616740452719125,
-1.106727415424254,
0.5548516538378656,
-0.5213984251160072,
0.6505183275226544,
2.5688868893071013,
-0.899030031766489,
0.10169372585728759,
0.051342844970556696,
-0.42195543536471375,
-0.18782383924141505,
-0.37663964256665594,
0.7059042964980583,
-0.17523611901973232,
-0.51573395101625,
-0.6661572076653586,
-0.3451703420124491,
-1.0186133738724747,
-0.464753684118435,
-0.6346879071111518,
-0.5151045650051659,
-0.2948194611257182,
-0.07579312926843881,
-0.5025168447834831,
-0.340135253923776,
-0.12488523813300144,
0.403799011177673,
-0.5969247464461036,
0.7436674571631066,
0.07651828541392215,
-0.622100186889469,
0.4289744516210384,
-0.2129992796847805,
-0.28852560101487684,
-0.13747295835468415,
-0.6787449278870413,
-0.42699052345338684,
-0.08712207746795327,
-0.3640519223449732,
0.8695446593799337,

0.11365206006788618,
-0.622100186889469,
-0.3640519223449732,
-0.6095124666677864,
0.8768455371085098,
-0.4081089431208627,
0.8556981671360827,
-0.6535694874436758,
-0.15006067857636687,
0.6933165762763757,
-0.899030031766489,
0.24645250840663893,
-0.8172098503255513,
-0.6032186065569449,
-1.4403020012988461,
1.5426981736748608,
-0.612659396723207,
2.128316681548206,
-0.4144028032317041,
-0.5591615857810555,
-0.3011133212365596,
0.08281214552476351,
-0.03677119658122237,
-0.7857405497713444,
-1.0186133738724747,
1.8627157848707008,
-0.4521659638967523,
-0.5717493060027381,
1.0961236233702227,
-0.39552122289918,
-0.7303545807959405,
-0.10600365780047735,
-0.33258262179076636,
-0.20041155946309777,
-0.781234145931982,
-0.5151045650051659,
-0.9367931924315371,
-0.4779330271905368,
5.421981791593043,
-0.5780431661135795,
-0.8927361716556476,
-0.3514642021232905,
-0.011595756137856923,

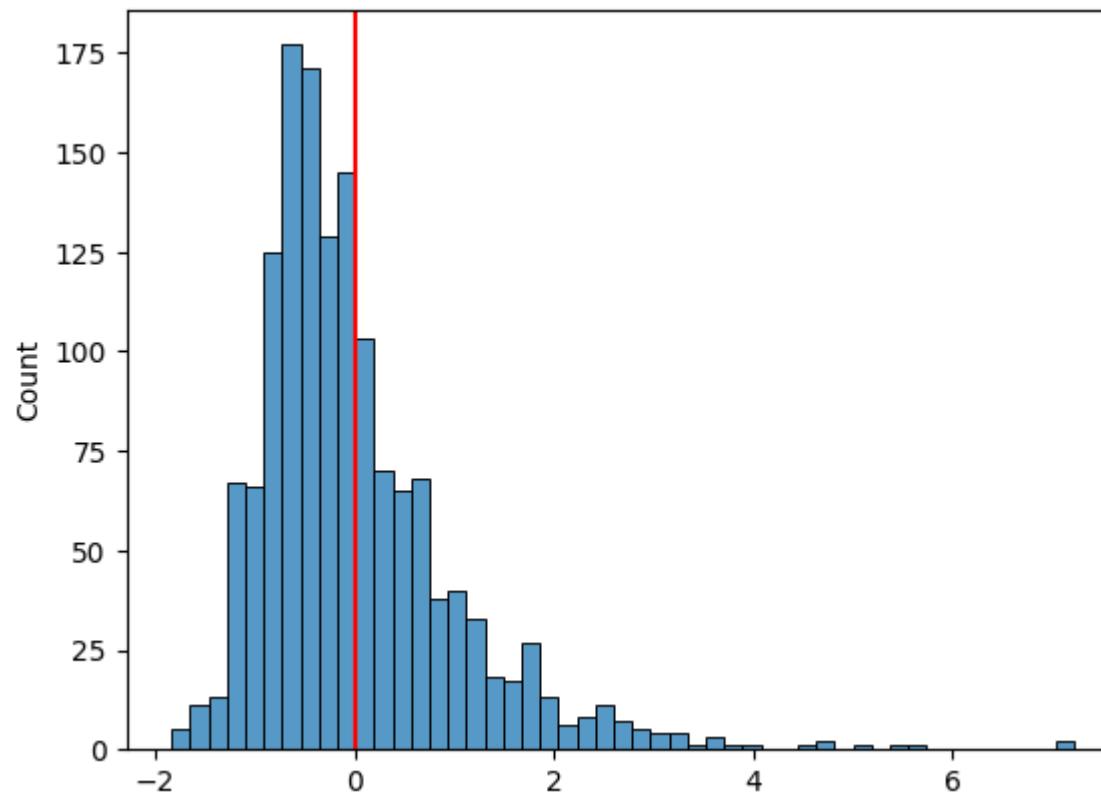
0.7436674571631066,
-0.6976265082195654,
-0.6661572076653586,
0.9324832604883474,
0.8695446593799337,
-0.6283940470003104,
-0.08712207746795327,
-0.33510016583510294,
-0.47104754422927636,
-1.1696660165326676,
-0.4521659638967523,
-0.09048299876714255,
-1.333306379414543,
-1.8329004072929085,
-0.5780431661135795,
0.7184920167197411,
-0.055652776913746455,
0.2527463685174803,
-0.4408370156972378,
-0.137598835556901,
0.15204460674401848,
0.334566549958418,
-0.07453435724627054,
1.3101148671388292,
-0.06194663702458782,
0.699610436387217,
0.5170884931728175,
0.2527463685174803,
-0.7983282699930272,
1.7506850748977245,
0.11428144607897031,
0.768842897606472,
-1.2716265503282975,
0.050084072948388424,
0.9073078200449819,
0.7411373253985483,
0.7990534261385105,
-0.37789841458882417,
0.41638673139935567,
-0.38922736278833864,
-0.47734140434011774,
-0.5465738655593727,
-0.705179140352575,

-0.47734140434011774,
1.1212990638135882,
0.14575074663317714,
0.20868934774159076,
-0.6535694874436758,
-0.7681177414609887,
-0.5918896583574305,
-0.11229751791131871,
-0.6724510677761999,
-0.4521659638967523,
-0.7165080885520895,
-0.6158063267786277,
0.051342844970556696,
-0.32628876167992504,
-0.81091599021471,
1.1464745042569537,
-0.32628876167992504,
0.7310797369414238,
0.42771567959887014,
-0.02544224838170792,
-0.2633501605715114,
-0.5780431661135795,
-1.800310799638972,
-0.5151045650051659,
-0.5780431661135795,
-0.09970979768963599,
-1.0249072339833163,
0.013579684305508526,
-0.16894225890889097,
-0.20041155946309777,
-1.2011353170868744,
0.2388998762736293,
-0.8927361716556476,
-0.5276922852268486,
-0.03173610849254928,
1.952088598444648,
-0.2646718711947881,
0.9438122086878619,
-0.691332648108724,
-0.7039203683304067,
-0.8046221301038685,
2.6971809338064916,
0.17722004718738393,

```
0.20239548763074938,  
2.103141241104841,  
-0.1626483987980496,  
0.07651828541392215,  
-0.08838084949012154,  
1.970970178777172,  
-0.746718617084128,  
-0.5591615857810555,  
0.051342844970556696,  
-1.1319028558676194,  
0.3156849696258939,  
...]
```

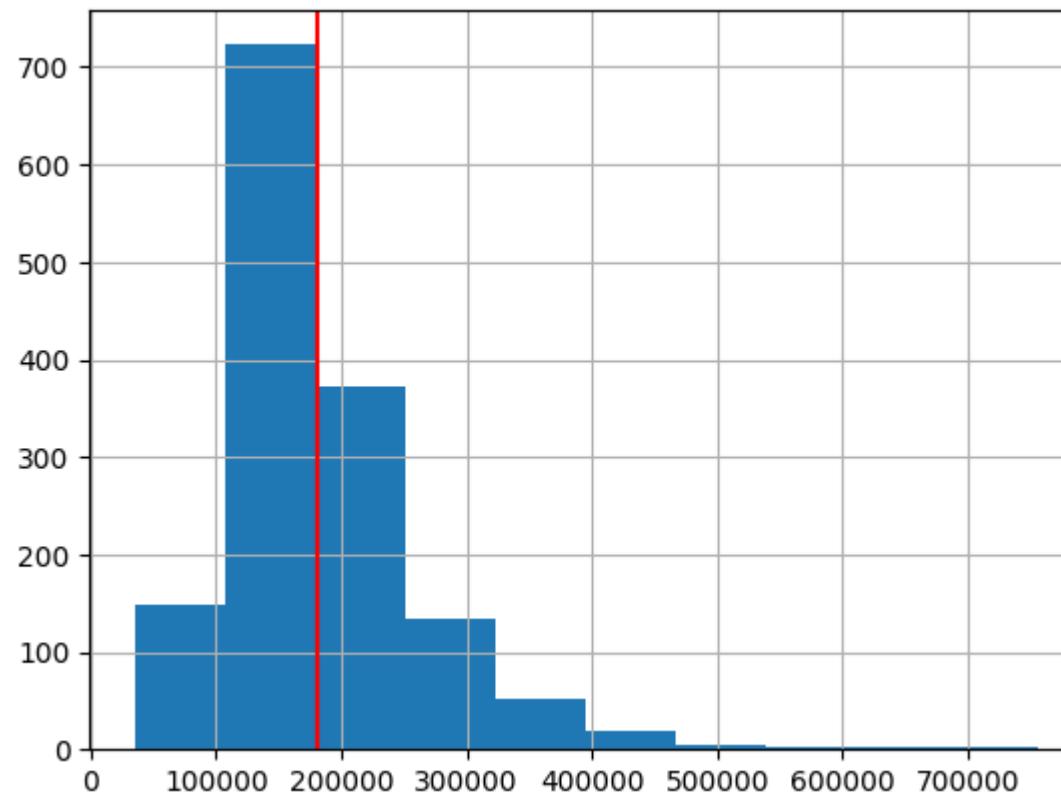
```
In [48]: sns.histplot(x=scaled)  
plt.axvline(x=pd.Series(scaled).mean(), color='red')
```

```
Out[48]: <matplotlib.lines.Line2D at 0x21325aa43d0>
```



```
In [51]: data['SalePrice'].hist()  
plt.axvline(x=data['SalePrice'].mean(), color='red')
```

```
Out[51]: <matplotlib.lines.Line2D at 0x213230f1de0>
```



OUTLIERS:

- Nominal:eg name of species/countries---->One-hot encoding is used
- Ordinal:eg designation,(poor/normal/good)----->Ordinal encoding/Label encoding
- Interval:eg IQ score, Temperature in oC or oF
- Ratio:eg cost of item

```
In [52]: def outliers_per(col):  
    Q1=np.quantile(data[col],.25)  
    Q3=np.quantile(data[col],.75)
```

```
IQR=Q3-Q1

lower_range=Q1-(1.5*IQR)
upper_range=Q3+(1.5*IQR)

upper_count=data[data[col]>upper_range][col].count()
upper_percentage=upper_count/data.shape[0]*100

lower_count=data[data[col]<lower_range][col].count()
lower_percentage=lower_count/data.shape[0]*100

return print('outlier percentage of {col} upper range is {upper}% and lower range is {lower}%')
```

```
In [53]: Q1=np.quantile(data['SalePrice'],.25)
Q3=np.quantile(data['SalePrice'],.75)

IQR=Q3-Q1

lower_range=Q1-(1.5*IQR)
upper_range=Q3+(1.5*IQR)
```

```
In [55]: data[data['SalePrice']>upper_range]['SalePrice']
```

```
Out[55]: 11      345000
         53      385000
         58      438780
        112     383970
       151     372402
       161     412500
       178     501837
       185     475000
       224     386250
       231     403000
       278     415298
       309     360000
       313     375000
       320     342643
       321     354000
       336     377426
       349     437154
       378     394432
       389     426000
       440     555000
       473     440000
       477     380000
       481     374000
       496     430000
       515     402861
       527     446261
       585     369900
       591     451950
       608     359100
       642     345000
       644     370878
       654     350000
       661     402000
       664     423000
       678     372500
       688     392000
       691     755000
       702     361919
       718     341000
       769     538000
       774     395000
       798     485000
      803     582933
```

```
825      385000
877      350000
898      611657
987      395192
990      348000
1046     556581
1142     424870
1169     625000
1181     392500
1182     745000
1228     367294
1243     465000
1267     378500
1268     381000
1353     410000
1373     466500
1388     377500
1437     394617
Name: SalePrice, dtype: int64
```

```
In [57]: upper_range
```

```
Out[57]: 340037.5
```

```
In [56]: len(data[data['SalePrice']>upper_range]['SalePrice'])/data.shape[0]*100
```

```
Out[56]: 4.178082191780822
```

```
In [58]: for i in data.select_dtypes(exclude='object').columns:
    outliers_per(i)
```

In []: