

Feature Engineering

July 20, 2023

```
[4]: #import the important libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import scipy as sc
from scipy import stats
import scipy.stats as stats
from scipy.stats import ttest_1samp
import os
import warnings
warnings.filterwarnings("ignore")
%matplotlib inline
```

```
[5]: #download the PEP1.csv file and import the dataset
Pep1 = pd.read_csv('PEP1.csv')
Pep1.head()
```

```
[5]:   Id MSSubClass MSZoning LotFrontage LotArea Street Alley LotShape \
0    1          60      RL       65.0     8450    Pave  NaN    Reg
1    2          20      RL       80.0     9600    Pave  NaN    Reg
2    3          60      RL       68.0    11250    Pave  NaN    IR1
3    4          70      RL       60.0     9550    Pave  NaN    IR1
4    5          60      RL       84.0    14260    Pave  NaN    IR1

  LandContour Utilities ... PoolArea PoolQC Fence MiscFeature MiscVal MoSold \
0        Lvl    AllPub ...        0    NaN    NaN        NaN    0     2
1        Lvl    AllPub ...        0    NaN    NaN        NaN    0     5
2        Lvl    AllPub ...        0    NaN    NaN        NaN    0     9
3        Lvl    AllPub ...        0    NaN    NaN        NaN    0     2
4        Lvl    AllPub ...        0    NaN    NaN        NaN    0    12

  YrSold SaleType SaleCondition SalePrice
0  2008      WD      Normal  208500
1  2007      WD      Normal  181500
2  2008      WD      Normal  223500
3  2006      WD  Abnorml  140000
```

```
4    2008      WD      Normal    250000
```

[5 rows x 81 columns]

```
[6]: Pep1.tail()
```

```
[6]:      Id  MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape \
1455  1456       60      RL       62.0     7917   Pave  NaN  Reg
1456  1457       20      RL       85.0    13175   Pave  NaN  Reg
1457  1458       70      RL       66.0     9042   Pave  NaN  Reg
1458  1459       20      RL       68.0     9717   Pave  NaN  Reg
1459  1460       20      RL       75.0    9937   Pave  NaN  Reg

      LandContour Utilities ... PoolArea PoolQC  Fence MiscFeature MiscVal \
1455      Lvl    AllPub ...        0    NaN    NaN      NaN      0
1456      Lvl    AllPub ...        0    NaN  MnPrv      NaN      0
1457      Lvl    AllPub ...        0    NaN  GdPrv    Shed    2500
1458      Lvl    AllPub ...        0    NaN    NaN      NaN      0
1459      Lvl    AllPub ...        0    NaN    NaN      NaN      0

      MoSold YrSold SaleType SaleCondition SalePrice
1455      8    2007      WD      Normal    175000
1456      2    2010      WD      Normal    210000
1457      5    2010      WD      Normal    266500
1458      4    2010      WD      Normal    142125
1459      6    2008      WD      Normal    147500
```

[5 rows x 81 columns]

```
[7]: #a) Indetify the shape and columns of the dataset
Pep1.shape
```

```
[7]: (1460, 81)
```

```
[8]: Pep1.columns
```

```
[8]: Index(['Id', 'MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street',
       'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig',
       'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType',
       'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd',
       'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType',
       'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual',
       'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1',
       'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating',
       'HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF',
       'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath',
       'HalfBath', 'BedroomAbvGr', 'KitchebvGr', 'KitchenQual', 'TotRmsAbvGrd',
```

```
'Functiol', 'Fireplaces', 'FireplaceQu', 'GarageType', 'GarageYrBlt',
'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual', 'GarageCond',
'PavedDrive', 'WoodDeckSF', 'OpenPorchSF', 'EnclosedPorch', '3SsnPorch',
'ScreenPorch', 'PoolArea', 'PoolQC', 'Fence', 'MiscFeature', 'MiscVal',
'MoSold', 'YrSold', 'SaleType', 'SaleCondition', 'SalePrice'],
dtype='object')
```

```
[9]: #b) indentify variable with null values in columns
pd.set_option('display.max_rows',None)
column_null_value =Pep1.isna().sum()
column_null_value =column_null_value[column_null_value >0]
column_null_value
```

```
[9]: LotFrontage      259
Alley            1369
MasVnrType       8
MasVnrArea       8
BsmtQual         37
BsmtCond         37
BsmtExposure     38
BsmtFinType1     37
BsmtFinType2     38
Electrical        1
FireplaceQu      690
GarageType        81
GarageYrBlt      81
GarageFinish      81
GarageQual        81
GarageCond        81
PoolQC           1453
Fence             1179
MiscFeature       1406
dtype: int64
```

```
[10]: #Alternative for all values and NAN values
Pep1.isna().sum()
```

```
[10]: 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
KitchebvGr	0
KitchenQual	0
TotRmsAbvGrd	0
Functiol	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
```

```
[11]: #identify the unique value
unique_counts = Pep1.nunique()
print("Variables with unique values:")
print(unique_counts)
```

Variables with unique values:

```
Id              1460
MSSubClass      15
MSZoning        5
LotFrontage     110
LotArea         1073
Street          2
Alley           2
LotShape         4
LandContour     4
Utilities        2
LotConfig        5
LandSlope        3
Neighborhood    25
Condition1      9
Condition2      8
```

BldgType	5
HouseStyle	8
OverallQual	10
OverallCond	9
YearBuilt	112
YearRemodAdd	61
RoofStyle	6
RoofMatl	8
Exterior1st	15
Exterior2nd	16
MasVnrType	4
MasVnrArea	327
ExterQual	4
ExterCond	5
Foundation	6
BsmtQual	4
BsmtCond	4
BsmtExposure	4
BsmtFinType1	6
BsmtFinSF1	637
BsmtFinType2	6
BsmtFinSF2	144
BsmtUnfSF	780
TotalBsmtSF	721
Heating	6
HeatingQC	5
CentralAir	2
Electrical	5
1stFlrSF	753
2ndFlrSF	417
LowQualFinSF	24
GrLivArea	861
BsmtFullBath	4
BsmtHalfBath	3
FullBath	4
HalfBath	3
BedroomAbvGr	8
KitchebvGr	4
KitchenQual	4
TotRmsAbvGrd	12
Functiol	7
Fireplaces	4
FireplaceQu	5
GarageType	6
GarageYrBlt	97
GarageFinish	3
GarageCars	5
GarageArea	441

```

GarageQual      5
GarageCond      5
PavedDrive      3
WoodDeckSF     274
OpenPorchSF    202
EnclosedPorch   120
3SsnPorch       20
ScreenPorch     76
PoolArea        8
PoolQC          3
Fence           4
MiscFeature     4
MiscVal         21
MoSold          12
YrSold          5
SaleType         9
SaleCondition    6
SalePrice       663
dtype: int64

```

```
[12]: #split the datasets into numerical variables(num_variable) and categorical
      ↪variables(cat_variable):
num_variable=Pep1.select_dtypes(exclude='object')
cat_variable=Pep1.select_dtypes(include='object')
```

```
[13]: num_variable.head()
```

	Id	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	\
0	1	60	65.0	8450	7	5	2003	
1	2	20	80.0	9600	6	8	1976	
2	3	60	68.0	11250	7	5	2001	
3	4	70	60.0	9550	7	5	1915	
4	5	60	84.0	14260	8	5	2000	

	YearRemodAdd	MasVnrArea	BsmtFinSF1	...	WoodDeckSF	OpenPorchSF	\
0	2003	196.0	706	...	0	61	
1	1976	0.0	978	...	298	0	
2	2002	162.0	486	...	0	42	
3	1970	0.0	216	...	0	35	
4	2000	350.0	655	...	192	84	

	EnclosedPorch	3SsnPorch	ScreenPorch	PoolArea	MiscVal	MoSold	YrSold	\
0	0	0	0	0	0	2	2008	
1	0	0	0	0	0	5	2007	
2	0	0	0	0	0	9	2008	
3	272	0	0	0	0	2	2006	
4	0	0	0	0	0	12	2008	

```

SalePrice
0      208500
1      181500
2      223500
3      140000
4      250000

[5 rows x 38 columns]

```

```
[14]: cat_variable.head()
```

```

[14]:   MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope \
0       RL    Pave   NaN     Reg        Lvl    AllPub    Inside    Gtl
1       RL    Pave   NaN     Reg        Lvl    AllPub    FR2      Gtl
2       RL    Pave   NaN    IR1        Lvl    AllPub    Inside    Gtl
3       RL    Pave   NaN    IR1        Lvl    AllPub    Corner   Gtl
4       RL    Pave   NaN    IR1        Lvl    AllPub    FR2      Gtl

Neighborhood Condition1 ... GarageType GarageFinish GarageQual GarageCond \
0   CollgCr        Norm ... Attchd        RFn       TA       TA
1   Veenker        Feedr ... Attchd        RFn       TA       TA
2   CollgCr        Norm ... Attchd        RFn       TA       TA
3   Crawfor         Norm ... Detchd        Unf       TA       TA
4   NoRidge        Norm ... Attchd        RFn       TA       TA

PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition
0       Y     NaN   NaN       NaN      WD  Normal
1       Y     NaN   NaN       NaN      WD  Normal
2       Y     NaN   NaN       NaN      WD  Normal
3       Y     NaN   NaN       NaN      WD Abnorml
4       Y     NaN   NaN       NaN      WD  Normal

```

```
[5 rows x 43 columns]
```

```

[15]: #EDA(exploratory Data Anaylsis) of Numerical Variable
      #a) Missing Value Treatment
      #Lets find the Nan values in Numerical Variable in PEP1 datsets

```

```

pd.set_option('display.max_rows',None)
column_null_value =Pep1.isna().sum()
column_null_value =column_null_value[column_null_value >0]
column_null_value

```

```

[15]: LotFrontage      259
      Alley          1369
      MasVnrType      8

```

```
MasVnrArea      8
BsmtQual       37
BsmtCond       37
BsmtExposure   38
BsmtFinType1   37
BsmtFinType2   38
Electrical     1
FireplaceQu    690
GarageType     81
GarageYrBlt    81
GarageFinish   81
GarageQual     81
GarageCond     81
PoolQC         1453
Fence          1179
MiscFeature    1406
dtype: int64
```

```
[16]: # Lets now treat the Nan values in the columns
column_null_value = num_variable.isna().sum()
column_null_value=column_null_value[column_null_value>0]
column_null_value
```

```
[16]: LotFrontage    259
MasVnrArea      8
GarageYrBlt    81
dtype: int64
```

```
[17]: # Replace the NAN values by the mean of the values in the same columns
lst= column_null_value.index[0: ].tolist()
print(lst)
```

```
['LotFrontage', 'MasVnrArea', 'GarageYrBlt']
```

```
[18]: def replaceNAN(lst,Pep1):
        for i in lst:
            print("Updating values:"+str(i))
            Pep1[i].fillna(value=Pep1[i].mean(),inplace=True)
        return Pep1
num_variable =replaceNAN(lst,num_variable)
num_variable.head()
```

```
Updating values:LotFrontage
Updating values:MasVnrArea
Updating values:GarageYrBlt
```

```
[18]:   Id MSSubClass LotFrontage LotArea OverallQual OverallCond YearBuilt \
0    1          60     65.0      8450         7           5       2003
1    2          20     80.0      9600         6           8       1976
2    3          60     68.0     11250         7           5       2001
3    4          70     60.0      9550         7           5       1915
4    5          60     84.0     14260         8           5       2000

   YearRemodAdd MasVnrArea BsmtFinSF1 ... WoodDeckSF OpenPorchSF \
0        2003     196.0      706   ...        0        61
1        1976      0.0      978   ...      298        0
2        2002     162.0      486   ...        0        42
3        1970      0.0      216   ...        0        35
4        2000     350.0      655   ...      192        84

   EnclosedPorch 3SsnPorch ScreenPorch PoolArea MiscVal MoSold YrSold \
0            0        0          0        0        0        2    2008
1            0        0          0        0        0        5    2007
2            0        0          0        0        0        9    2008
3          272        0          0        0        0        2    2006
4            0        0          0        0        0       12    2008

   SalePrice
0    208500
1    181500
2    223500
3    140000
4    250000
```

[5 rows x 38 columns]

```
[19]: # check the NAN values in numerical variables
num_variable.isna().sum()
```

```
[19]: Id          0
MSSubClass      0
LotFrontage     0
LotArea         0
OverallQual     0
OverallCond     0
YearBuilt        0
YearRemodAdd    0
MasVnrArea      0
BsmtFinSF1      0
BsmtFinSF2      0
BsmtUnfSF       0
TotalBsmtSF     0
1stFlrSF        0
```

```
2ndFlrSF          0  
LowQualFinSF     0  
GrLivArea         0  
BsmtFullBath     0  
BsmtHalfBath     0  
FullBath          0  
HalfBath          0  
BedroomAbvGr      0  
KitchebvGr        0  
TotRmsAbvGrd     0  
Fireplaces         0  
GarageYrBlt       0  
GarageCars         0  
GarageArea         0  
WoodDeckSF         0  
OpenPorchSF        0  
EnclosedPorch     0  
3SsnPorch          0  
ScreenPorch        0  
PoolArea           0  
MiscVal            0  
MoSold              0  
YrSold              0  
SalePrice           0  
dtype: int64
```

```
[20]: num_variable.dtypes
```

```
[20]: Id                  int64  
MSSubClass          int64  
LotFrontage         float64  
LotArea              int64  
OverallQual         int64  
OverallCond          int64  
YearBuilt            int64  
YearRemodAdd        int64  
MasVnrArea          float64  
BsmtFinSF1          int64  
BsmtFinSF2          int64  
BsmtUnfSF           int64  
TotalBsmtSF         int64  
1stFlrSF             int64  
2ndFlrSF             int64  
LowQualFinSF        int64  
GrLivArea            int64  
BsmtFullBath        int64  
BsmtHalfBath        int64
```

```
FullBath           int64
HalfBath          int64
BedroomAbvGr      int64
KitchebvGr        int64
TotRmsAbvGrd     int64
Fireplaces        int64
GarageYrBlt      float64
GarageCars        int64
GarageArea        int64
WoodDeckSF        int64
OpenPorchSF       int64
EnclosedPorch    int64
3SsnPorch         int64
ScreenPorch       int64
PoolArea          int64
MiscVal           int64
MoSold            int64
YrSold            int64
SalePrice          int64
dtype: object
```

```
[21]: # lets convert all dtypes in int64 for better visualization of graphs
num_variable['LotFrontage']=num_variable['LotFrontage'].astype('int64')
num_variable['MasVnrArea']=num_variable['MasVnrArea'].astype('int64')
num_variable['GarageYrBlt']=num_variable['GarageYrBlt'].astype('int64')
```

```
[22]: num_variable.dtypes
```

```
[22]: Id           int64
MSSubClass      int64
LotFrontage     int64
LotArea         int64
OverallQual     int64
OverallCond     int64
YearBuilt       int64
YearRemodAdd   int64
MasVnrArea     int64
BsmtFinSF1    int64
BsmtFinSF2    int64
BsmtUnfSF      int64
TotalBsmtSF    int64
1stFlrSF        int64
2ndFlrSF        int64
LowQualFinSF   int64
GrLivArea       int64
BsmtFullBath   int64
BsmtHalfBath   int64
```

```

FullBath          int64
HalfBath          int64
BedroomAbvGr     int64
KitchebvGr       int64
TotRmsAbvGrd    int64
Fireplaces        int64
GarageYrBlt      int64
GarageCars        int64
GarageArea        int64
WoodDeckSF       int64
OpenPorchSF      int64
EnclosedPorch    int64
3SsnPorch        int64
ScreenPorch       int64
PoolArea          int64
MiscVal           int64
MoSold            int64
YrSold            int64
SalePrice          int64
dtype: object

```

[23]: num_variable.columns

[23]: Index(['Id', 'MSSubClass', 'LotFrontage', 'LotArea', 'OverallQual',
 'OverallCond', 'YearBuilt', 'YearRemodAdd', 'MasVnrArea', 'BsmtFinSF1',
 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', '1stFlrSF', '2ndFlrSF',
 'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath',
 'HalfBath', 'BedroomAbvGr', 'KitchebvGr', 'TotRmsAbvGrd', 'Fireplaces',
 'GarageYrBlt', 'GarageCars', 'GarageArea', 'WoodDeckSF', 'OpenPorchSF',
 'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'MiscVal',
 'MoSold', 'YrSold', 'SalePrice'],
 dtype='object')

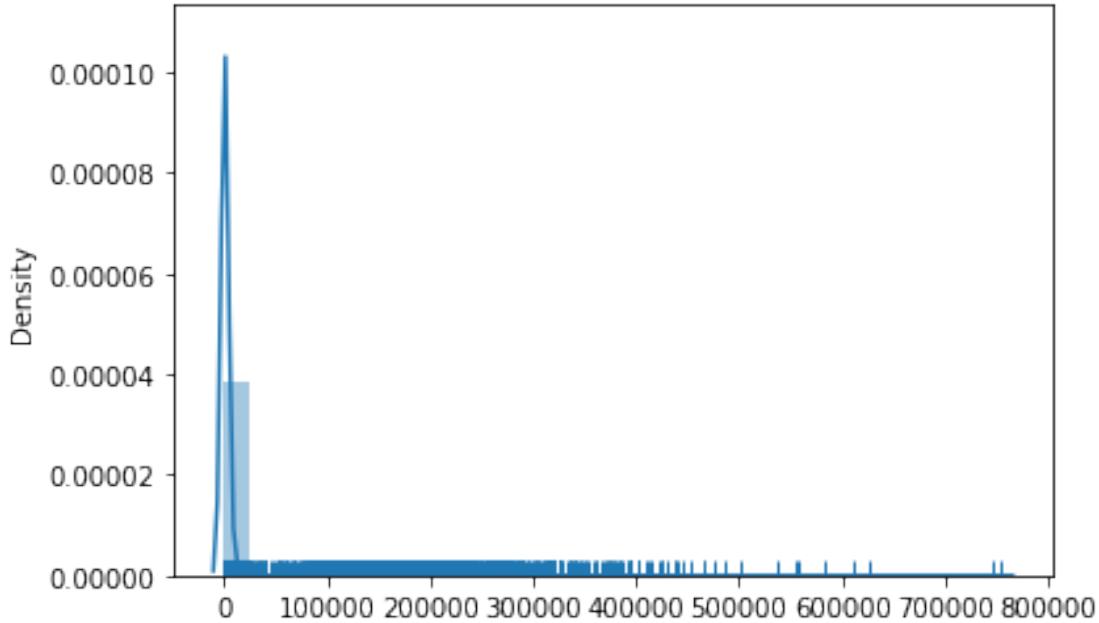
[24]: #b) identify the Skewness and the distribution, import the libraries like
 ↪stats, matplotlib, seaborn
 from scipy.stats import skew
 import seaborn as sns
 import matplotlib.pyplot as plt
 %matplotlib inline

[25]: skewvalue= num_variable.skew(axis=0, numeric_only=True)
 print(skewvalue)

Id	0.000000
MSSubClass	1.407657
LotFrontage	2.386156

```
LotArea          12.207688
OverallQual      0.216944
OverallCond      0.693067
YearBuilt        -0.613461
YearRemodAdd    -0.503562
MasVnrArea       2.676474
BsmtFinSF1       1.685503
BsmtFinSF2       4.255261
BsmtUnfSF        0.920268
TotalBsmtSF      1.524255
1stFlrSF         1.376757
2ndFlrSF         0.813030
LowQualFinSF     9.011341
GrLivArea        1.366560
BsmtFullBath     0.596067
BsmtHalfBath     4.103403
FullBath          0.036562
HalfBath          0.675897
BedroomAbvGr     0.211790
KitchebvGr       4.488397
TotRmsAbvGrd     0.676341
Fireplaces        0.649565
GarageYrBlt      -0.664636
GarageCars        -0.342549
GarageArea        0.179981
WoodDeckSF        1.541376
OpenPorchSF       2.364342
EnclosedPorch     3.089872
3SsnPorch         10.304342
ScreenPorch       4.122214
PoolArea          14.828374
MiscVal           24.476794
MoSold            0.212053
YrSold             0.096269
SalePrice          1.882876
dtype: float64
```

```
[26]: # identify the distribution and density of skewness of numerical variable
sns.distplot(num_variable, bins=30,kde=True,rug=True)
plt.show()
```



```
[44]: import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats

plt.figure(figsize=(12, 7))

for i, column in enumerate(num_variable.columns):
    plt.subplot(3,4,i+1) # Create 4 subplots in the first row
    sns.histplot(data=num_variable[column], kde=True) # Histogram with KDE line
    plt.title(column) # Set the title as the column name

plt.tight_layout()
plt.show()
```

□
→-----

```
ValueError
last)

<ipython-input-44-ae4a13a713c9> in <module>
      6
      7 for i, column in enumerate(num_variable.columns):
----> 8     plt.subplot(3,4,i+1) # Create 4 subplots in the first row
```

```

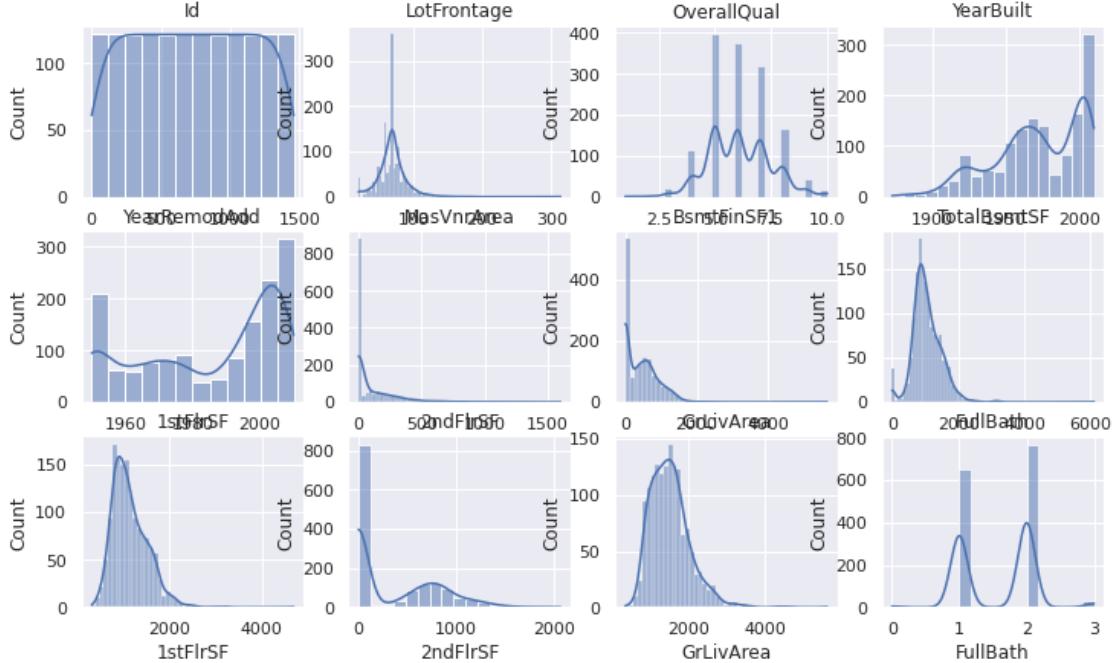
9      sns.histplot(data=num_variable[column], kde=True) # Histogram with KDE line
10     plt.title(column) # Set the title as the column name

/usr/local/lib/python3.7/site-packages/matplotlib/pyplot.py in
1266
1267     # First, search for an existing subplot with a matching spec.
-> 1268     key = SubplotSpec._from_subplot_args(fig, args)
1269
1270     for ax in fig.axes:

/usr/local/lib/python3.7/site-packages/matplotlib/gridspec.py in
1266
1267     if not isinstance(num, Integral) or num < 1 or num >
-> 1268         raise ValueError(
-> 1269             f"num must be 1 <= num <= {rows*cols}, not {num!
1270             r}")
1271
1272     i = j = num
1273
1274     return gs[i-1:j]

```

ValueError: num must be 1 <= num <= 12, not 13



```
[28]: #c) Identify significant variable using a correlation matrix
correlation_matrix= num_variable.corr()
correlation_matrix
```

	Id	MSSubClass	LotFrontage	LotArea	OverallQual	\
Id	1.000000	0.011156	-0.009616	-0.033226	-0.028365	
MSSubClass	0.011156	1.000000	-0.357042	-0.139781	0.032628	
LotFrontage	-0.009616	-0.357042	1.000000	0.306689	0.234227	
LotArea	-0.033226	-0.139781	0.306689	1.000000	0.105806	
OverallQual	-0.028365	0.032628	0.234227	0.105806	1.000000	
OverallCond	0.012609	-0.059316	-0.052842	-0.005636	-0.091932	
YearBuilt	-0.012713	0.027850	0.117555	0.014228	0.572323	
YearRemodAdd	-0.021998	0.040581	0.082775	0.013788	0.550684	
MasVnrArea	-0.050205	0.022899	0.179270	0.103957	0.410220	
BsmtFinSF1	-0.005024	-0.069836	0.215760	0.214103	0.239666	
BsmtFinSF2	-0.005968	-0.065649	0.043298	0.111170	-0.059119	
BsmtUnfSF	-0.007940	-0.140759	0.122249	-0.002618	0.308159	
TotalBsmtSF	-0.015415	-0.238518	0.363366	0.260833	0.537808	
1stFlrSF	0.010496	-0.251758	0.414246	0.299475	0.476224	
2ndFlrSF	0.005590	0.307886	0.072479	0.050986	0.295493	
LowQualFinSF	-0.044230	0.046474	0.036879	0.004779	-0.030429	
GrLivArea	0.008273	0.074853	0.368376	0.263116	0.593007	
BsmtFullBath	0.002289	0.003491	0.091428	0.158155	0.111098	
BsmtHalfBath	-0.020155	-0.002333	-0.006446	0.048046	-0.040150	
FullBath	0.005587	0.131608	0.180431	0.126031	0.550600	
HalfBath	0.006784	0.177354	0.048209	0.014259	0.273458	
BedroomAbvGr	0.037719	-0.023438	0.237016	0.119690	0.101676	
KitchebvGr	0.002951	0.281721	-0.005762	-0.017784	-0.183882	
TotRmsAbvGrd	0.027239	0.040380	0.320166	0.190015	0.427452	
Fireplaces	-0.019772	-0.045569	0.235636	0.271364	0.396765	
GarageYrBlt	0.000136	0.079774	0.064794	-0.024454	0.519278	
GarageCars	0.016570	-0.040110	0.269722	0.154871	0.600671	
GarageArea	0.017634	-0.098672	0.323658	0.180403	0.562022	
WoodDeckSF	-0.029643	-0.012579	0.077032	0.171698	0.238923	
OpenPorchSF	-0.000477	-0.006100	0.137435	0.084774	0.308819	
EnclosedPorch	0.002889	-0.012037	0.009814	-0.018340	-0.113937	
3SsnPorch	-0.046635	-0.043825	0.062317	0.020423	0.030371	
ScreenPorch	0.001330	-0.026030	0.037683	0.043160	0.064886	
PoolArea	0.057044	0.008283	0.180867	0.077672	0.065166	
MiscVal	-0.006242	-0.007683	0.001101	0.038068	-0.031406	
MoSold	0.021172	-0.013585	0.010172	0.001205	0.070815	
YrSold	0.000712	-0.021407	0.006750	-0.014261	-0.027347	
SalePrice	-0.021917	-0.084284	0.334897	0.263843	0.790982	
	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	\

Id	0.012609	-0.012713	-0.021998	-0.050205	-0.005024
MSSubClass	-0.059316	0.027850	0.040581	0.022899	-0.069836
LotFrontage	-0.052842	0.117555	0.082775	0.179270	0.215760
LotArea	-0.005636	0.014228	0.013788	0.103957	0.214103
OverallQual	-0.091932	0.572323	0.550684	0.410220	0.239666
OverallCond	1.000000	-0.375983	0.073741	-0.127775	-0.046231
YearBuilt	-0.375983	1.000000	0.592855	0.314726	0.249503
YearRemodAdd	0.073741	0.592855	1.000000	0.179170	0.128451
MasVnrArea	-0.127775	0.314726	0.179170	1.000000	0.263569
BsmtFinSF1	-0.046231	0.249503	0.128451	0.263569	1.000000
BsmtFinSF2	0.040229	-0.049107	-0.067759	-0.072296	-0.050117
BsmtUnfSF	-0.136841	0.149040	0.181133	0.114182	-0.495251
TotalBsmtSF	-0.171098	0.391452	0.291066	0.362438	0.522396
1stFlrSF	-0.144203	0.281986	0.240379	0.342147	0.445863
2ndFlrSF	0.028942	0.010308	0.140024	0.174019	-0.137079
LowQualFinSF	0.025494	-0.183784	-0.062419	-0.069066	-0.064503
GrLivArea	-0.079686	0.199010	0.287389	0.389883	0.208171
BsmtFullBath	-0.054942	0.187599	0.119470	0.085042	0.649212
BsmtHalfBath	0.117821	-0.038162	-0.012337	0.026674	0.067418
FullBath	-0.194149	0.468271	0.439046	0.275713	0.058543
HalfBath	-0.060769	0.242656	0.183331	0.200792	0.004262
BedroomAbvGr	0.012980	-0.070651	-0.040581	0.102420	-0.107355
KitchebvGr	-0.087001	-0.174800	-0.149598	-0.037371	-0.081007
TotRmsAbvGrd	-0.057583	0.095589	0.191740	0.280025	0.044316
Fireplaces	-0.023820	0.147716	0.112581	0.247902	0.260011
GarageYrBlt	-0.306118	0.781662	0.618672	0.249927	0.150858
GarageCars	-0.185758	0.537850	0.420622	0.363768	0.224054
GarageArea	-0.151521	0.478954	0.371600	0.372558	0.296970
WoodDeckSF	-0.003334	0.224880	0.205726	0.159355	0.204306
OpenPorchSF	-0.032589	0.188686	0.226298	0.124950	0.111761
EnclosedPorch	0.070356	-0.387268	-0.193919	-0.109850	-0.102303
3SsnPorch	0.025504	0.031355	0.045286	0.018797	0.026451
ScreenPorch	0.054811	-0.050364	-0.038740	0.061459	0.062021
PoolArea	-0.001985	0.004950	0.005829	0.011724	0.140491
MiscVal	0.068777	-0.034383	-0.010286	-0.029813	0.003571
MoSold	-0.003511	0.012398	0.021490	-0.005945	-0.015727
YrSold	0.043950	-0.013618	0.035743	-0.008185	0.014359
SalePrice	-0.077856	0.522897	0.507101	0.475227	0.386420

...	WoodDeckSF	OpenPorchSF	EnclosedPorch	3SsnPorch	\
Id	-0.029643	-0.000477	0.002889	-0.046635	
MSSubClass	-0.012579	-0.006100	-0.012037	-0.043825	
LotFrontage	0.077032	0.137435	0.009814	0.062317	
LotArea	0.171698	0.084774	-0.018340	0.020423	
OverallQual	0.238923	0.308819	-0.113937	0.030371	
OverallCond	-0.003334	-0.032589	0.070356	0.025504	
YearBuilt	0.224880	0.188686	-0.387268	0.031355	

YearRemodAdd	...	0.205726	0.226298	-0.193919	0.045286
MasVnrArea	...	0.159355	0.124950	-0.109850	0.018797
BsmtFinSF1	...	0.204306	0.111761	-0.102303	0.026451
BsmtFinSF2	...	0.067898	0.003093	0.036543	-0.029993
BsmtUnfSF	...	-0.005316	0.129005	-0.002538	0.020764
TotalBsmtSF	...	0.232019	0.247264	-0.095478	0.037384
1stFlrSF	...	0.235459	0.211671	-0.065292	0.056104
2ndFlrSF	...	0.092165	0.208026	0.061989	-0.024358
LowQualFinSF	...	-0.025444	0.018251	0.061081	-0.004296
GrLivArea	...	0.247433	0.330224	0.009113	0.020643
BsmtFullBath	...	0.175315	0.067341	-0.049911	-0.000106
BsmtHalfBath	...	0.040161	-0.025324	-0.008555	0.035114
FullBath	...	0.187703	0.259977	-0.115093	0.035353
HalfBath	...	0.108080	0.199740	-0.095317	-0.004972
BedroomAbvGr	...	0.046854	0.093810	0.041570	-0.024478
KitchebvGr	...	-0.090130	-0.070091	0.037312	-0.024600
TotRmsAbvGrd	...	0.165984	0.234192	0.004151	-0.006683
Fireplaces	...	0.200019	0.169405	-0.024822	0.011257
GarageYrBlt	...	0.221131	0.218673	-0.286177	0.023670
GarageCars	...	0.226342	0.213569	-0.151434	0.035765
GarageArea	...	0.224666	0.241435	-0.121777	0.035087
WoodDeckSF	...	1.000000	0.058661	-0.125989	-0.032771
OpenPorchSF	...	0.058661	1.000000	-0.093079	-0.005842
EnclosedPorch	...	-0.125989	-0.093079	1.000000	-0.037305
3SsnPorch	...	-0.032771	-0.005842	-0.037305	1.000000
ScreenPorch	...	-0.074181	0.074304	-0.082864	-0.031436
PoolArea	...	0.073378	0.060762	0.054203	-0.007992
MiscVal	...	-0.009551	-0.018584	0.018361	0.000354
MoSold	...	0.021011	0.071255	-0.028887	0.029474
YrSold	...	0.022270	-0.057619	-0.009916	0.018645
SalePrice	...	0.324413	0.315856	-0.128578	0.044584

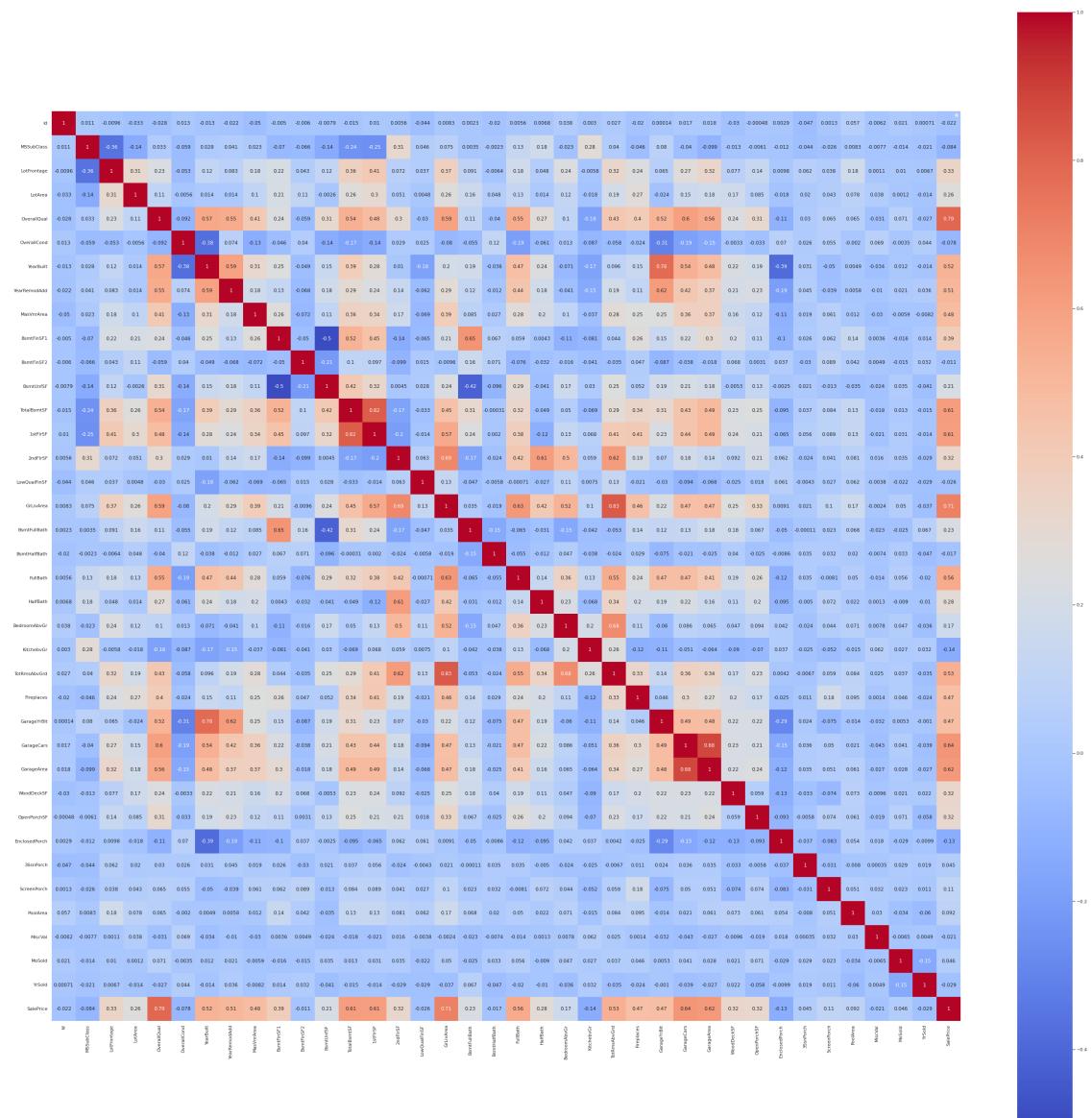
	ScreenPorch	PoolArea	MiscVal	MoSold	YrSold	SalePrice
Id	0.001330	0.057044	-0.006242	0.021172	0.000712	-0.021917
MSSubClass	-0.026030	0.008283	-0.007683	-0.013585	-0.021407	-0.084284
LotFrontage	0.037683	0.180867	0.001101	0.010172	0.006750	0.334897
LotArea	0.043160	0.077672	0.038068	0.001205	-0.014261	0.263843
OverallQual	0.064886	0.065166	-0.031406	0.070815	-0.027347	0.790982
OverallCond	0.054811	-0.001985	0.068777	-0.003511	0.043950	-0.077856
YearBuilt	-0.050364	0.004950	-0.034383	0.012398	-0.013618	0.522897
YearRemodAdd	-0.038740	0.005829	-0.010286	0.021490	0.035743	0.507101
MasVnrArea	0.061459	0.011724	-0.029813	-0.005945	-0.008185	0.475227
BsmtFinSF1	0.062021	0.140491	0.003571	-0.015727	0.014359	0.386420
BsmtFinSF2	0.088871	0.041709	0.004940	-0.015211	0.031706	-0.011378
BsmtUnfSF	-0.012579	-0.035092	-0.023837	0.034888	-0.041258	0.214479
TotalBsmtSF	0.084489	0.126053	-0.018479	0.013196	-0.014969	0.613581
1stFlrSF	0.088758	0.131525	-0.021096	0.031372	-0.013604	0.605852

2ndFlrSF	0.040606	0.081487	0.016197	0.035164	-0.028700	0.319334
LowQualFinSF	0.026799	0.062157	-0.003793	-0.022174	-0.028921	-0.025606
GrLivArea	0.101510	0.170205	-0.002416	0.050240	-0.036526	0.708624
BsmtFullBath	0.023148	0.067616	-0.023047	-0.025361	0.067049	0.227122
BsmtHalfBath	0.032121	0.020025	-0.007367	0.032873	-0.046524	-0.016844
FullBath	-0.008106	0.049604	-0.014290	0.055872	-0.019669	0.560664
HalfBath	0.072426	0.022381	0.001290	-0.009050	-0.010269	0.284108
BedroomAbvGr	0.044300	0.070703	0.007767	0.046544	-0.036014	0.168213
KitchebvGr	-0.051613	-0.014525	0.062341	0.026589	0.031687	-0.135907
TotRmsAbvGrd	0.059383	0.083757	0.024763	0.036907	-0.034516	0.533723
Fireplaces	0.184530	0.095074	0.001409	0.046357	-0.024096	0.466929
GarageYrBlt	-0.074938	-0.014418	-0.031877	0.005265	-0.001041	0.471316
GarageCars	0.050494	0.020934	-0.043080	0.040522	-0.039117	0.640409
GarageArea	0.051412	0.061047	-0.027400	0.027974	-0.027378	0.623431
WoodDeckSF	-0.074181	0.073378	-0.009551	0.021011	0.022270	0.324413
OpenPorchSF	0.074304	0.060762	-0.018584	0.071255	-0.057619	0.315856
EnclosedPorch	-0.082864	0.054203	0.018361	-0.028887	-0.009916	-0.128578
3SsnPorch	-0.031436	-0.007992	0.000354	0.029474	0.018645	0.044584
ScreenPorch	1.000000	0.051307	0.031946	0.023217	0.010694	0.111447
PoolArea	0.051307	1.000000	0.029669	-0.033737	-0.059689	0.092404
MiscVal	0.031946	0.029669	1.000000	-0.006495	0.004906	-0.021190
MoSold	0.023217	-0.033737	-0.006495	1.000000	-0.145721	0.046432
YrSold	0.010694	-0.059689	0.004906	-0.145721	1.000000	-0.028923
SalePrice	0.111447	0.092404	-0.021190	0.046432	-0.028923	1.000000

[38 rows x 38 columns]

```
[29]: sns.set(rc={'figure.figsize':(50,50)})
sns.heatmap(data=correlation_matrix, square=True, annot=True, fmt=' .2g', cmap="coolwarm")
plt.legend()
plt.show()
```

No artists with labels found to put in legend. Note that artists whose label start with an underscore are ignored when legend() is called with no argument.



```
[30]: # lets find the correlation and heat map of value less than 0.3
dfNumHigh = num_variable.loc[:, num_variable.corr().abs()['SalePrice'] > 0.3]
```

```
dfNumHigh['Id'] = num_variable['Id']
```

```
first_column = dfNumHigh.pop('Id')
```

```
dfNumHigh.insert(0, 'Id', first_column)
```

```
[31]: num_variable= dfNumHigh
```

```
[32]: num_variable.head()
```

```
[32]:    Id LotFrontage OverallQual YearBuilt YearRemodAdd MasVnrArea \
0     1          65            7      2003      2003        196
1     2          80            6      1976      1976         0
2     3          68            7      2001      2002       162
3     4          60            7      1915      1970         0
4     5          84            8      2000      2000       350

    BsmtFinSF1 TotalBsmtSF 1stFlrSF 2ndFlrSF GrLivArea FullBath \
0       706        856      856      854      1710        2
1       978       1262     1262        0      1262        2
2       486        920      920      866      1786        2
3       216        756      961      756      1717        1
4       655       1145     1145     1053      2198        2

    TotRmsAbvGrd Fireplaces GarageYrBlt GarageCars GarageArea WoodDeckSF \
0           8          0      2003        2        548        0
1           6          1      1976        2        460      298
2           6          1      2001        2        608        0
3           7          1      1998        3        642        0
4           9          1      2000        3        836      192

    OpenPorchSF SalePrice
0          61   208500
1          0    181500
2          42   223500
3          35   140000
4          84   250000
```

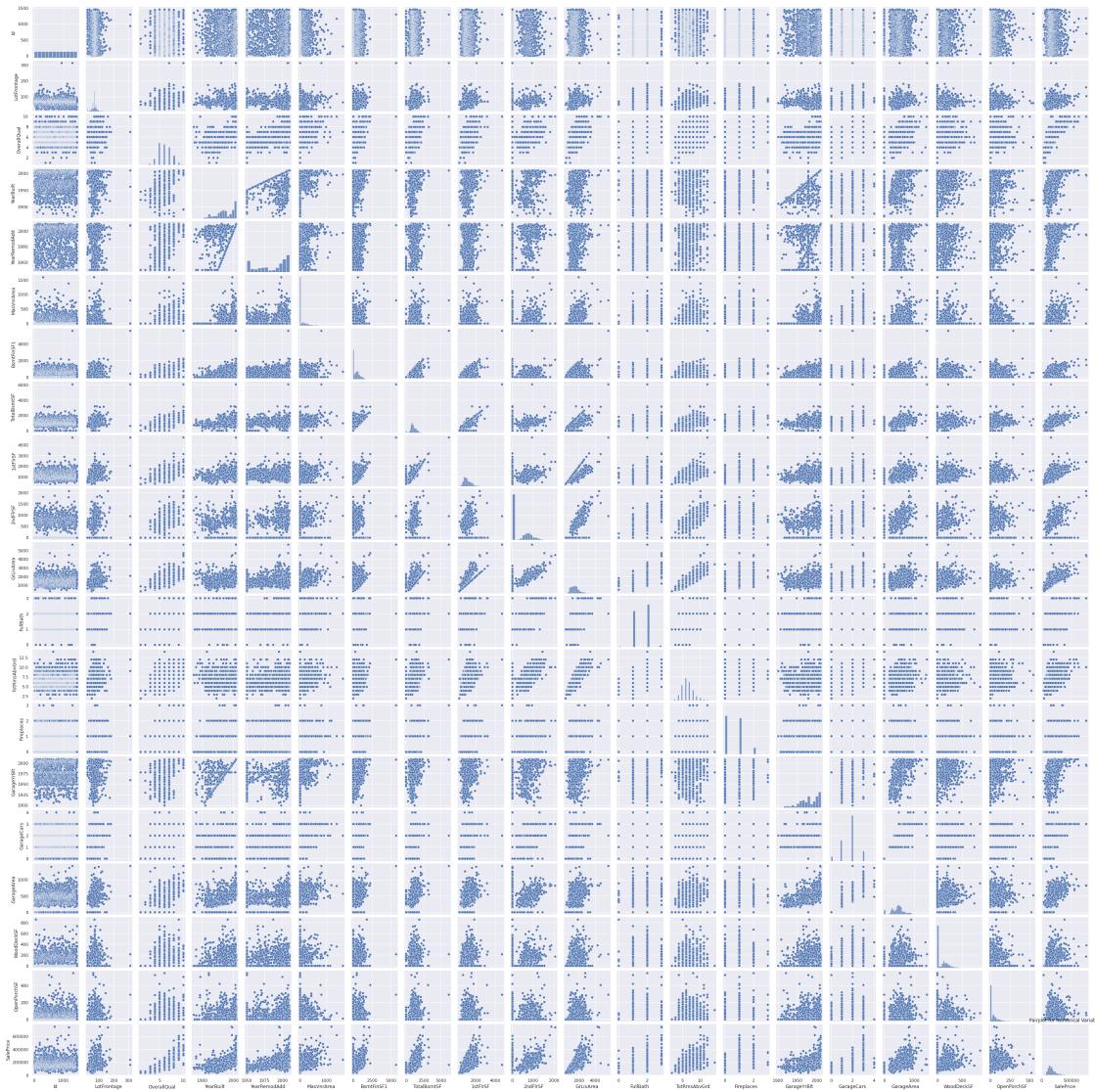
```
[33]: num_variable.dtypes
```

```
[33]: Id          int64
LotFrontage    int64
OverallQual    int64
YearBuilt      int64
YearRemodAdd   int64
MasVnrArea     int64
BsmtFinSF1    int64
TotalBsmtSF   int64
1stFlrSF       int64
2ndFlrSF       int64
GrLivArea      int64
FullBath        int64
TotRmsAbvGrd   int64
Fireplaces     int64
GarageYrBlt    int64
GarageCars      int64
GarageArea      int64
```

```
WoodDeckSF      int64
OpenPorchSF     int64
SalePrice       int64
dtype: object
```

```
[42]: #pairplot for numerical data
skewValue = num_variable.skew(axis=0, numeric_only=True)
print(skewValue)
sns.pairplot(num_variable,height=2)
plt.title("Pairplot for Numerical Variable")
plt.show()
```

```
Id              0.000000
LotFrontage     2.386156
OverallQual     0.216944
YearBuilt       -0.613461
YearRemodAdd   -0.503562
MasVnrArea      2.676474
BsmtFinSF1     1.685503
TotalBsmtSF    1.524255
1stFlrSF        1.376757
2ndFlrSF        0.813030
GrLivArea       1.366560
FullBath         0.036562
TotRmsAbvGrd   0.676341
Fireplaces      0.649565
GarageYrBlt    -0.664636
GarageCars       -0.342549
GarageArea       0.179981
WoodDeckSF      1.541376
OpenPorchSF     2.364342
SalePrice        1.882876
dtype: float64
```



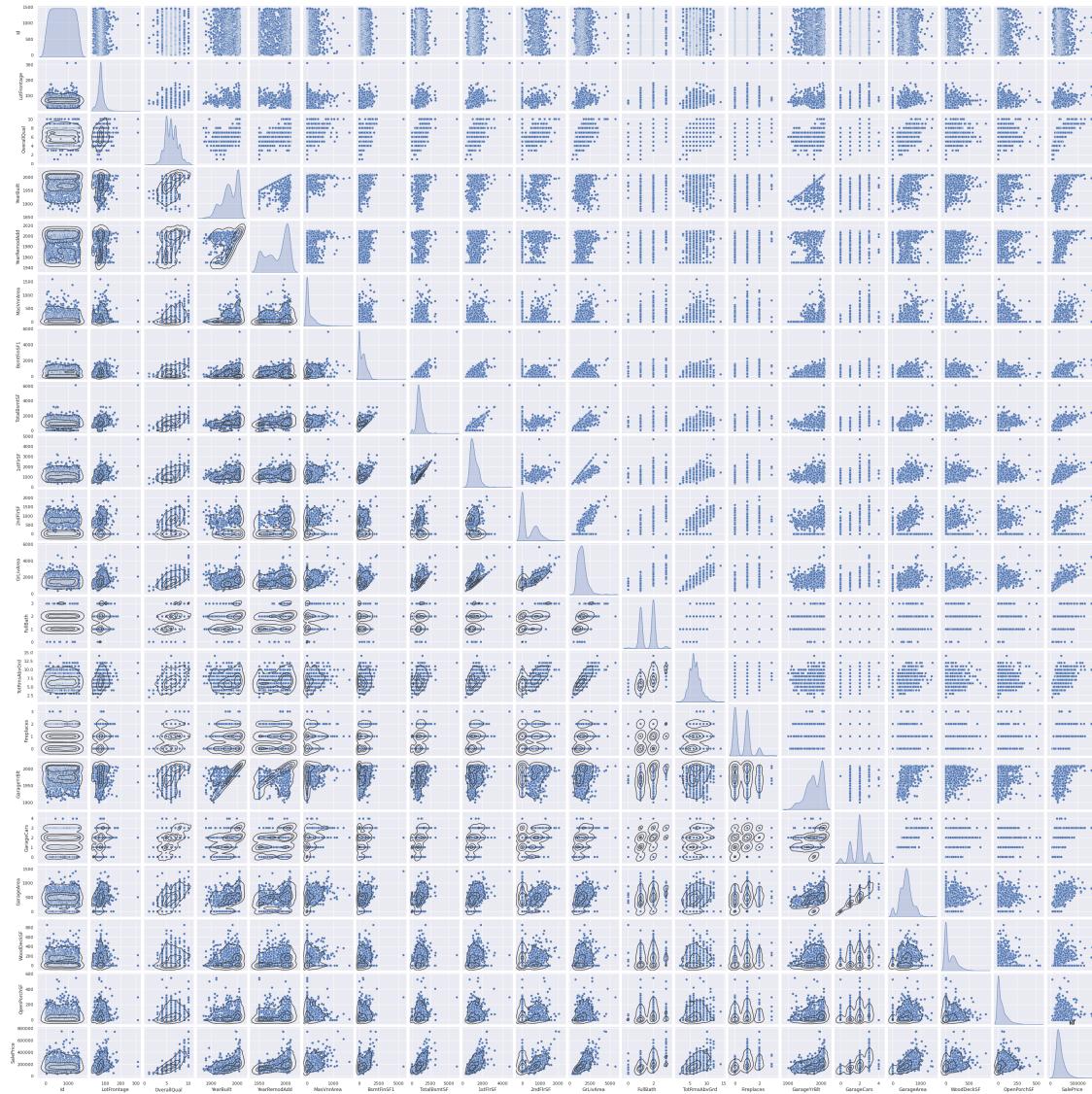
```
[45]: #pairplot for numerical data
def PairPlotLoop(Pep1, columns):
    for col in columns:
        if (Pep1[col].dtype == 'int64') :
            i = num_variable[col].unique()

            plt.figure(figsize=(16,6))
            g = sns.pairplot(num_variable, diag_kind="kde",height=2)
            g.map_lower(sns.kdeplot, levels=4, color=".2")

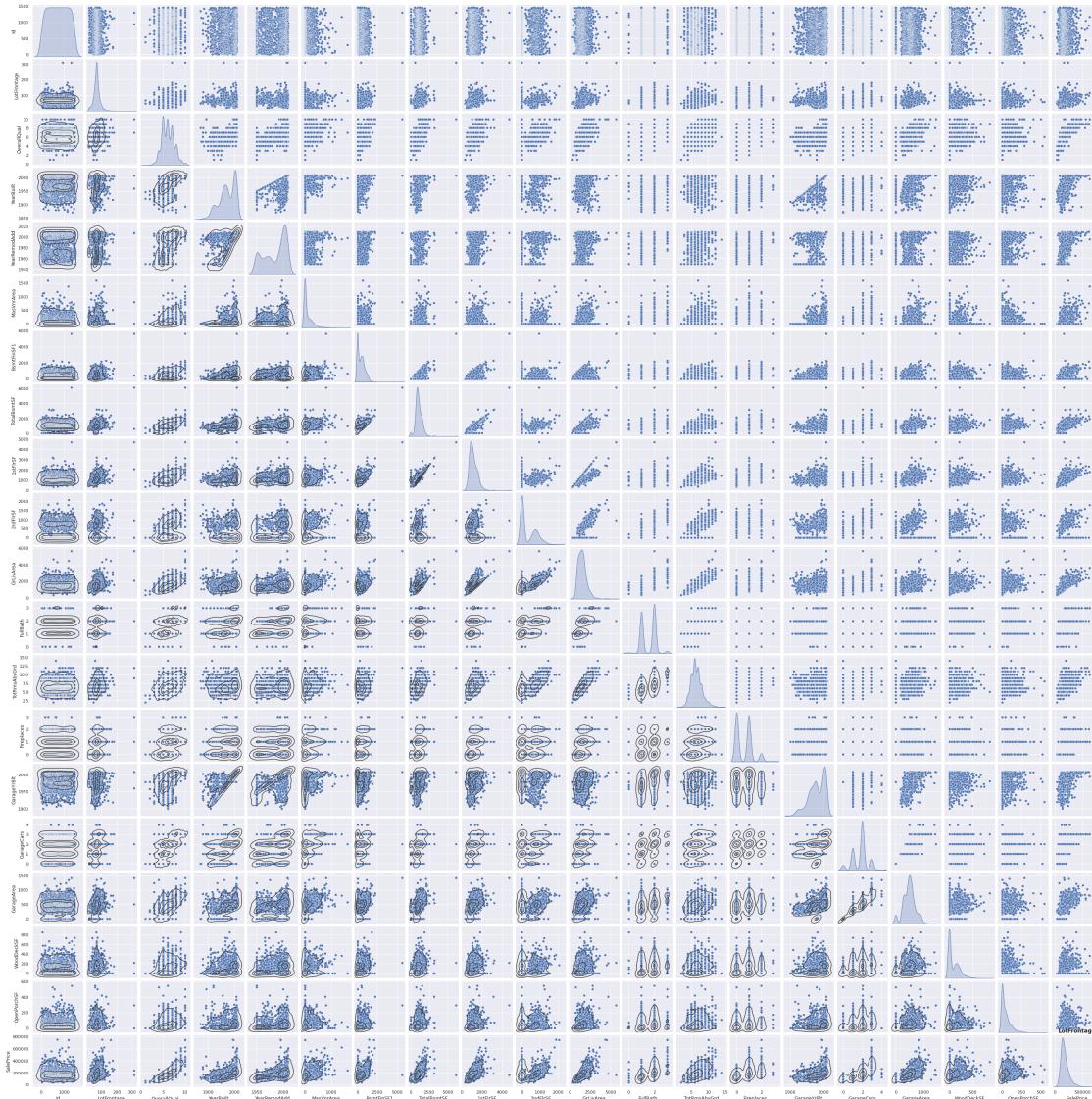
            plt.title(col, fontsize=14, fontweight='bold')
            plt.show()
```

```
[ ]: PairPlotLoop(num_variable,num_variable.columns)
```

<Figure size 1152x432 with 0 Axes>



<Figure size 1152x432 with 0 Axes>



```
[ ]: #Q4.EDA (Exploratory Data Analysis) of categorical variable
#a) Missing value treatment
cat_variable.head()
```

```
[29]: cat_variable['SalePrice']=Pep1['SalePrice']
cat_variable['Id']=Pep1['Id']
```

```
[30]: cat_variable.head()
```

```
[30]: MSZoning Street Alley LotShape LandContour Utilities LotConfig LandSlope \
0      RL    Pave   NaN     Reg       Lvl    AllPub    Inside    Gtl
1      RL    Pave   NaN     Reg       Lvl    AllPub        FR2    Gtl
2      RL    Pave   NaN    IR1       Lvl    AllPub    Inside    Gtl
```

```

3      RL    Pave   NaN     IR1          Lvl    AllPub   Corner    Gtl
4      RL    Pave   NaN     IR1          Lvl    AllPub   FR2       Gtl

Neighborhood Condition1 ... GarageQual GarageCond PavedDrive PoolQC Fence \
0      CollgCr      Norm   ...        TA        TA       Y     NaN   NaN
1      Veenker      Feedr   ...        TA        TA       Y     NaN   NaN
2      CollgCr      Norm   ...        TA        TA       Y     NaN   NaN
3      Crawfor      Norm   ...        TA        TA       Y     NaN   NaN
4      NoRidge      Norm   ...        TA        TA       Y     NaN   NaN

MiscFeature SaleType SaleCondition SalePrice Id
0      NaN        WD      Normal    208500  1
1      NaN        WD      Normal    181500  2
2      NaN        WD      Normal    223500  3
3      NaN        WD      Abnorml  140000  4
4      NaN        WD      Normal    250000  5

[5 rows x 45 columns]

```

```
[31]: #Shift Id in front
first_column = cat_variable.pop('Id')

# insert column using insert(position,column_name,
cat_variable.insert(0, 'Id', first_column)
cat_variable.head()
```

```

[31]:   Id MSZoning Street Alley LotShape LandContour Utilities LotConfig  \
0   1      RL    Pave   NaN     Reg          Lvl    AllPub   Inside
1   2      RL    Pave   NaN     Reg          Lvl    AllPub   FR2
2   3      RL    Pave   NaN     IR1          Lvl    AllPub   Inside
3   4      RL    Pave   NaN     IR1          Lvl    AllPub   Corner
4   5      RL    Pave   NaN     IR1          Lvl    AllPub   FR2

LandSlope Neighborhood ... GarageFinish GarageQual GarageCond PavedDrive \
0      Gtl      CollgCr  ...        RFn        TA        TA       Y
1      Gtl      Veenker   ...        RFn        TA        TA       Y
2      Gtl      CollgCr   ...        RFn        TA        TA       Y
3      Gtl      Crawfor   ...        Unf        TA        TA       Y
4      Gtl      NoRidge   ...        RFn        TA        TA       Y

PoolQC Fence MiscFeature SaleType SaleCondition SalePrice
0      NaN    NaN      NaN        WD      Normal    208500
1      NaN    NaN      NaN        WD      Normal    181500
2      NaN    NaN      NaN        WD      Normal    223500
3      NaN    NaN      NaN        WD      Abnorml  140000
4      NaN    NaN      NaN        WD      Normal    250000

```

[5 rows x 45 columns]

[32]: #NAN Values

```
cat_null_column= cat_variable.isna().sum()
cat_null_column=cat_null_column[cat_null_column>0]
cat_null_column
```

[32]: Alley 1369
MasVnrType 8
BsmtQual 37
BsmtCond 37
BsmtExposure 38
BsmtFinType1 37
BsmtFinType2 38
Electrical 1
FireplaceQu 690
GarageType 81
GarageFinish 81
GarageQual 81
GarageCond 81
PoolQC 1453
Fence 1179
MiscFeature 1406
dtype: int64

[33]: #Replace the NAN Values by using count function of column

```
Pep1_clean = cat_variable.apply(lambda x:x.fillna(x.value_counts().index[0]))
Pep1_clean.head()
```

[33]:

	Id	MSZoning	Street	Alley	LotShape	LandContour	Utilities	LotConfig	\
0	1	RL	Pave	Grvl	Reg	Lvl	AllPub	Inside	
1	2	RL	Pave	Grvl	Reg	Lvl	AllPub	FR2	
2	3	RL	Pave	Grvl	IR1	Lvl	AllPub	Inside	
3	4	RL	Pave	Grvl	IR1	Lvl	AllPub	Corner	
4	5	RL	Pave	Grvl	IR1	Lvl	AllPub	FR2	

	LandSlope	Neighborhood	...	GarageFinish	GarageQual	GarageCond	PavedDrive	\
0	Gtl	CollgCr	...	RFn	TA	TA	Y	
1	Gtl	Veenker	...	RFn	TA	TA	Y	
2	Gtl	CollgCr	...	RFn	TA	TA	Y	
3	Gtl	Crawfor	...	Unf	TA	TA	Y	
4	Gtl	NoRidge	...	RFn	TA	TA	Y	

	PoolQC	Fence	MiscFeature	SaleType	SaleCondition	SalePrice		
0	Gd	MnPrv		Shed	WD	Normal	208500	
1	Gd	MnPrv		Shed	WD	Normal	181500	
2	Gd	MnPrv		Shed	WD	Normal	223500	

```

3      Gd MnPrv        Shed       WD      Abnормл    140000
4      Gd MnPrv        Shed       WD      Normal     250000

```

[5 rows x 45 columns]

```
[34]: # in order to arrange the string to get the count plot correctly
for col in cat_variable:
    print('***** ' + col + ' count= ' + str(len(cat_variable[col].unique())))
    ↪+ ' *****')
    print(cat_variable[col].unique())

```

```

***** Id count= 1460 *****
[ 1 2 3 ... 1458 1459 1460]
***** MSZoning count= 5 *****
['RL' 'RM' 'C (all)' 'FV' 'RH']
***** Street count= 2 *****
['Pave' 'Grvl']
***** Alley count= 3 *****
[nan 'Grvl' 'Pave']
***** LotShape count= 4 *****
['Reg' 'IR1' 'IR2' 'IR3']
***** LandContour count= 4 *****
['Lvl' 'Bnk' 'Low' 'HLS']
***** Utilities count= 2 *****
['AllPub' 'NoSeWa']
***** LotConfig count= 5 *****
['Inside' 'FR2' 'Corner' 'CulDSac' 'FR3']
***** LandSlope count= 3 *****
['Gtl' 'Mod' 'Sev']
***** Neighborhood count= 25 *****
['CollgCr' 'Veenker' 'Crawfor' 'NoRidge' 'Mitchel' 'Somerst' 'NWAmes'
 'OldTown' 'BrkSide' 'Sawyer' 'NridgHt' 'mes' 'SawyerW' 'IDOTRR' 'MeadowV'
 'Edwards' 'Timber' 'Gilbert' 'StoneBr' 'ClearCr' 'NPkVill' 'Blmngtn'
 'BrDale' 'SWISU' 'Blueste']
***** Condition1 count= 9 *****
['Norm' 'Feedr' 'PosN' 'Artery' 'RRAe' 'RRNn' 'RRAn' 'PosA' 'RRNe']
***** Condition2 count= 8 *****
['Norm' 'Artery' 'RRNn' 'Feedr' 'PosN' 'PosA' 'RRAn' 'RR Ae']
***** BldgType count= 5 *****
['1Fam' '2fmCon' 'Duplex' 'TwnhsE' 'Twnhs']
***** HouseStyle count= 8 *****
['2Story' '1Story' '1.5Fin' '1.5Unf' 'SFoyer' 'SLvl' '2.5Unf' '2.5Fin']
***** RoofStyle count= 6 *****
['Gable' 'Hip' 'Gambrel' 'Mansard' 'Flat' 'Shed']
***** RoofMatl count= 8 *****
['CompShg' 'WdShngl' 'Metal' 'WdShake' 'Membran' 'Tar&Grv' 'Roll'

```

```

'ClyTile']
***** Exterior1st count= 15 *****
['VinylSd' 'MetalSd' 'Wd Sdng' 'HdBoard' 'BrkFace' 'WdShing' 'CemntBd'
 'Plywood' 'AsbShng' 'Stucco' 'BrkComm' 'AsphShn' 'Stone' 'ImStucc'
 'CBlock']

***** Exterior2nd count= 16 *****
['VinylSd' 'MetalSd' 'Wd Shng' 'HdBoard' 'Plywood' 'Wd Sdng' 'CmentBd'
 'BrkFace' 'Stucco' 'AsbShng' 'Brk Cmn' 'ImStucc' 'AsphShn' 'Stone'
 'Other' 'CBlock']

***** MasVnrType count= 5 *****
['BrkFace' 'None' 'Stone' 'BrkCmn' nan]

***** ExterQual count= 4 *****
['Gd' 'TA' 'Ex' 'Fa']

***** ExterCond count= 5 *****
['TA' 'Gd' 'Fa' 'Po' 'Ex']

***** Foundation count= 6 *****
['PConc' 'CBlock' 'BrkTil' 'Wood' 'Slab' 'Stone']

***** BsmtQual count= 5 *****
['Gd' 'TA' 'Ex' nan 'Fa']

***** BsmtCond count= 5 *****
['TA' 'Gd' nan 'Fa' 'Po']

***** BsmtExposure count= 5 *****
['No' 'Gd' 'Mn' 'Av' nan]

***** BsmtFinType1 count= 7 *****
['GLQ' 'ALQ' 'Unf' 'Rec' 'BLQ' nan 'LwQ']

***** BsmtFinType2 count= 7 *****
['Unf' 'BLQ' nan 'ALQ' 'Rec' 'LwQ' 'GLQ']

***** Heating count= 6 *****
['GasA' 'GasW' 'Grav' 'Wall' 'OthW' 'Floor']

***** HeatingQC count= 5 *****
['Ex' 'Gd' 'TA' 'Fa' 'Po']

***** CentralAir count= 2 *****
['Y' 'N']

***** Electrical count= 6 *****
['SBrkr' 'FuseF' 'FuseA' 'FuseP' 'Mix' nan]

***** KitchenQual count= 4 *****
['Gd' 'TA' 'Ex' 'Fa']

***** Functiol count= 7 *****
['Typ' 'Min1' 'Maj1' 'Min2' 'Mod' 'Maj2' 'Sev']

***** FireplaceQu count= 6 *****
[nan 'TA' 'Gd' 'Fa' 'Ex' 'Po']

***** GarageType count= 7 *****
['Attchd' 'Detchd' 'BuiltIn' 'CarPort' nan 'Basment' '2Types']

***** GarageFinish count= 4 *****
['RFn' 'Unf' 'Fin' nan]

***** GarageQual count= 6 *****
['TA' 'Fa' 'Gd' nan 'Ex' 'Po']

***** GarageCond count= 6 *****

```

```

['TA' 'Fa' nan 'Gd' 'Po' 'Ex']
***** PavedDrive count= 3 *****
['Y' 'N' 'P']
***** PoolQC count= 4 *****
[nan 'Ex' 'Fa' 'Gd']
***** Fence count= 5 *****
[nan 'MnPrv' 'GdWo' 'GdPrv' 'MnWw']
***** MiscFeature count= 5 *****
[nan 'Shed' 'Gar2' 'Othr' 'TenC']
***** SaleType count= 9 *****
['WD' 'New' 'COD' 'ConLD' 'ConLI' 'CWD' 'ConLw' 'Con' 'Oth']
***** SaleCondition count= 6 *****
['Normal' 'Abnorml' 'Partial' 'AdjLand' 'Alloca' 'Family']
***** SalePrice count= 663 *****
[208500 181500 223500 140000 250000 143000 307000 200000 129900 118000
 129500 345000 144000 279500 157000 132000 149000 90000 159000 139000
 325300 139400 230000 154000 256300 134800 306000 207500 68500 40000
 149350 179900 165500 277500 309000 145000 153000 109000 82000 160000
 170000 130250 141000 319900 239686 249700 113000 127000 177000 114500
 110000 385000 130000 180500 172500 196500 438780 124900 158000 101000
 202500 219500 317000 180000 226000 80000 225000 244000 185000 144900
 107400 91000 135750 136500 193500 153500 245000 126500 168500 260000
 174000 164500 85000 123600 109900 98600 163500 133900 204750 214000
 94750 83000 128950 205000 178000 118964 198900 169500 100000 115000
 190000 136900 383970 217000 259500 176000 155000 320000 163990 136000
 153900 181000 84500 128000 87000 150000 150750 220000 171000 231500
 166000 204000 125000 105000 222500 122000 372402 235000 79000 109500
 269500 254900 162500 412500 103200 152000 127500 325624 183500 228000
 128500 215000 239000 163000 184000 243000 211000 501837 200100 120000
 475000 173000 135000 153337 286000 315000 192000 148500 311872 104000
 274900 171500 112000 143900 277000 98000 186000 252678 156000 161750
 134450 210000 107000 311500 167240 204900 97000 386250 290000 106000
 192500 148000 403000 94500 128200 216500 89500 185500 194500 318000
 262500 110500 241500 137000 76500 276000 151000 73000 175500 179500
 120500 266000 124500 201000 415298 228500 244600 179200 164700 88000
 153575 233230 135900 131000 167000 142500 175000 158500 267000 149900
 295000 305900 82500 360000 165600 119900 375000 188500 270000 187500
 342643 354000 301000 126175 242000 324000 145250 214500 78000 119000
 284000 207000 228950 377426 202900 87500 140200 151500 157500 437154
 318061 95000 105900 177500 134000 280000 198500 147000 165000 162000
 172400 134432 123000 61000 340000 394432 179000 187750 213500 76000
 240000 81000 191000 426000 106500 129000 67000 241000 245500 164990
 108000 258000 168000 339750 60000 222000 181134 149500 126000 142000
 206300 275000 109008 195400 85400 79900 122500 212000 116000 90350
 555000 162900 199900 119500 188000 256000 161000 263435 62383 188700
 124000 178740 146500 187000 440000 251000 132500 208900 380000 297000
 89471 326000 374000 164000 86000 133000 172785 91300 34900 430000
 226700 289000 208300 164900 202665 96500 402861 265000 234000 106250

```

```

184750 315750 446261 200624 107500 39300 111250 272000 248000 213250
179665 229000 263000 112500 255500 121500 268000 325000 316600 135960
142600 224500 118500 146000 131500 181900 253293 369900 79500 185900
451950 138000 319000 114504 194201 217500 221000 359100 313000 261500
75500 137500 183200 105500 314813 305000 165150 139900 209500 93000
264561 274000 370878 143250 98300 205950 350000 145500 97500 197900
402000 423000 230500 173500 103600 257500 372500 159434 285000 227875
148800 392000 194700 755000 335000 108480 141500 89000 123500 138500
196000 312500 361919 213000 55000 302000 254000 179540 52000 102776
189000 130500 159500 341000 103000 236500 131400 93500 239900 299800
236000 265979 260400 275500 158900 179400 215200 337000 264132 216837
538000 134900 102000 395000 221500 175900 187100 161500 233000 107900
160200 146800 269790 143500 485000 582933 227680 135500 159950 144500
55993 157900 224900 271000 224000 183000 139500 232600 147400 237000
139950 174900 133500 189950 250580 248900 169000 200500 66500 303477
132250 328900 122900 154500 118858 142953 611657 125500 255000 154300
173733 75000 35311 238000 176500 145900 169990 193000 117500 184900
253000 239799 244400 150900 197500 172000 116500 214900 178900 37900
99500 182000 167500 85500 178400 336000 159895 255900 117000 395192
195000 197000 348000 173900 337500 121600 206000 232000 136905 119200
227000 203000 213490 194000 287000 293077 310000 119750 84000 315500
262280 278000 139600 556581 84900 176485 200141 185850 328000 167900
151400 91500 138800 155900 83500 252000 92900 176432 274725 134500
184100 133700 118400 212900 163900 259000 239500 94000 424870 174500
116900 201800 218000 235128 108959 233170 245350 625000 171900 154900
392500 745000 186700 104900 262000 219210 116050 271900 229456 80500
137900 367294 101800 138887 265900 248328 465000 186500 169900 171750
294000 165400 301500 99900 128900 183900 378500 381000 185750 68400
150500 281000 333168 206900 295493 111000 156500 72500 52500 155835
108500 283463 410000 156932 144152 216000 274300 466500 58500 237500
377500 246578 281213 137450 193879 282922 257000 223000 274970 182900
192140 143750 64500 394617 149700 149300 121000 179600 92000 287090
266500 142125 147500]

```

[35]: #b) Count Plot for bivariate Analysis

```

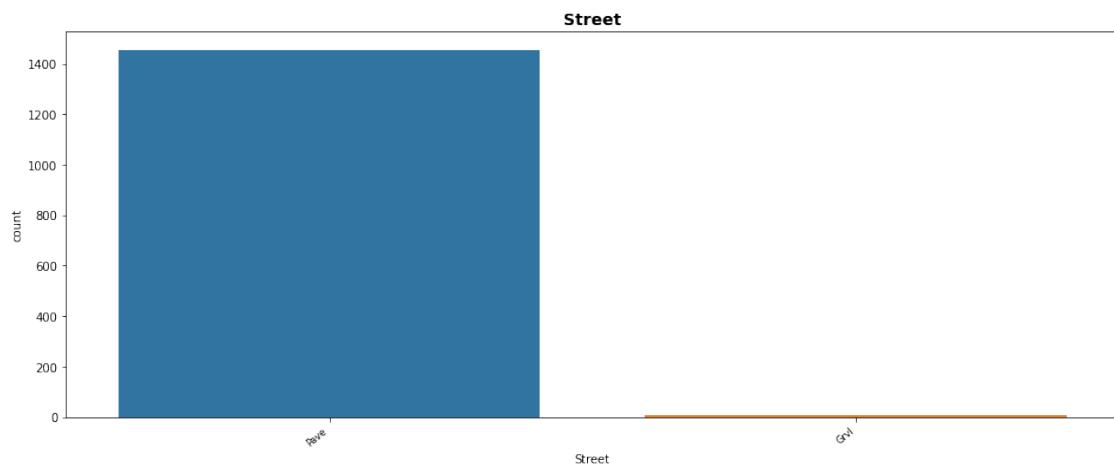
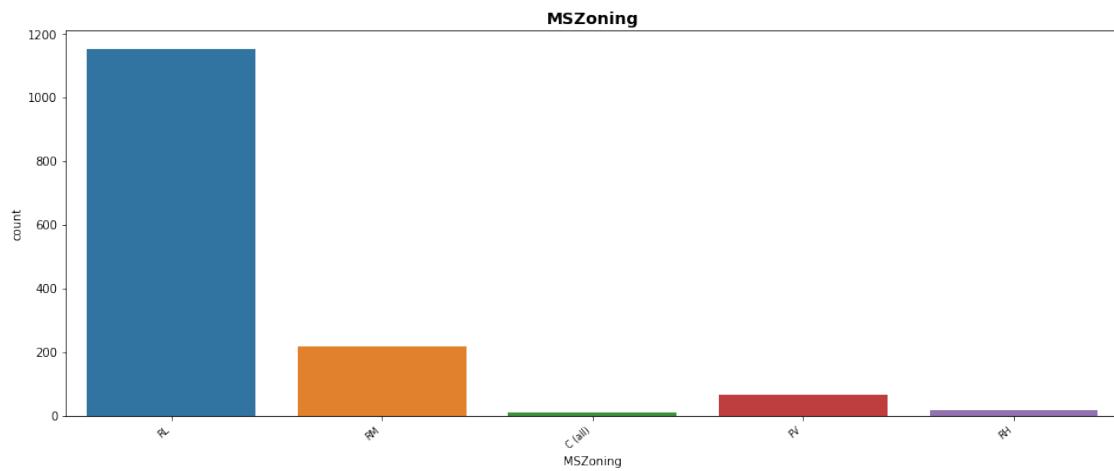
def histPlotLoop(Pep1, columns):
    for col in columns:
        if (Pep1[col].dtype == object) :
            i = cat_variable[col].unique()

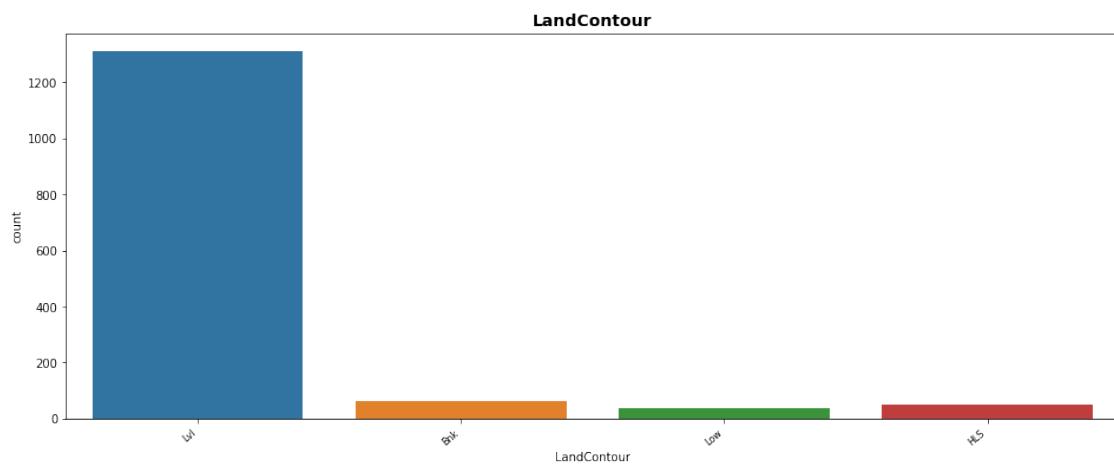
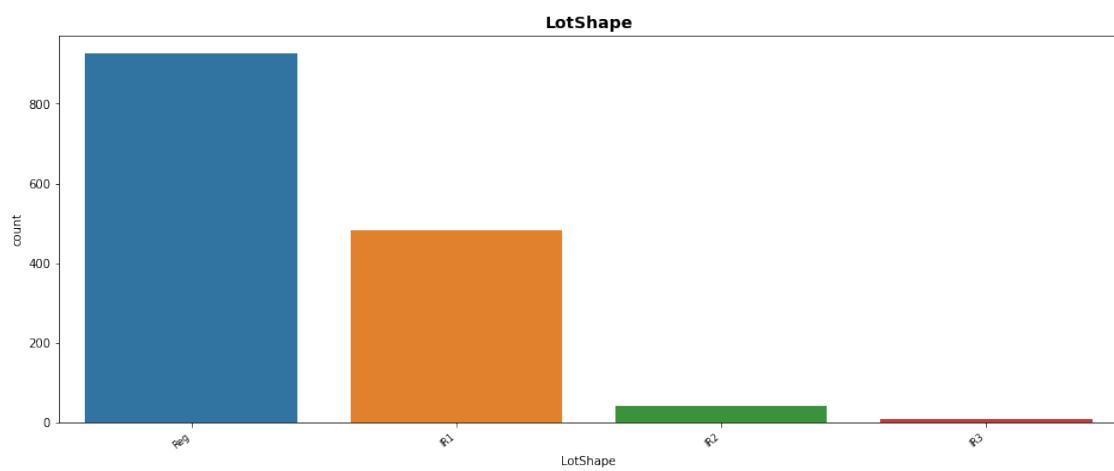
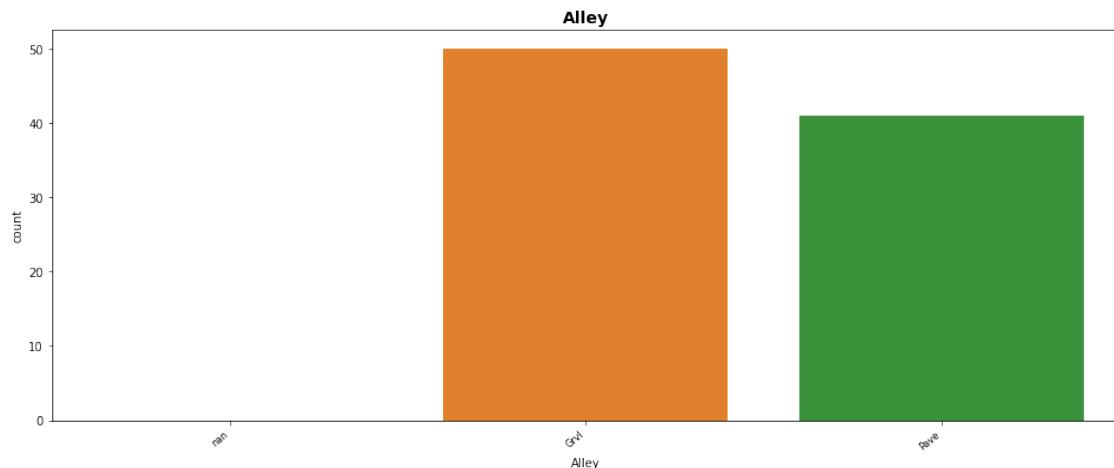
            ## Plotting a histogram with the Categorical Data
            plt.figure(figsize=(16,6))
            ax = sns.countplot(x=col, data=Pep1, order=i)
            ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right", u
            ↪fontsize=8)
            plt.title(col, fontsize=14, fontweight='bold')

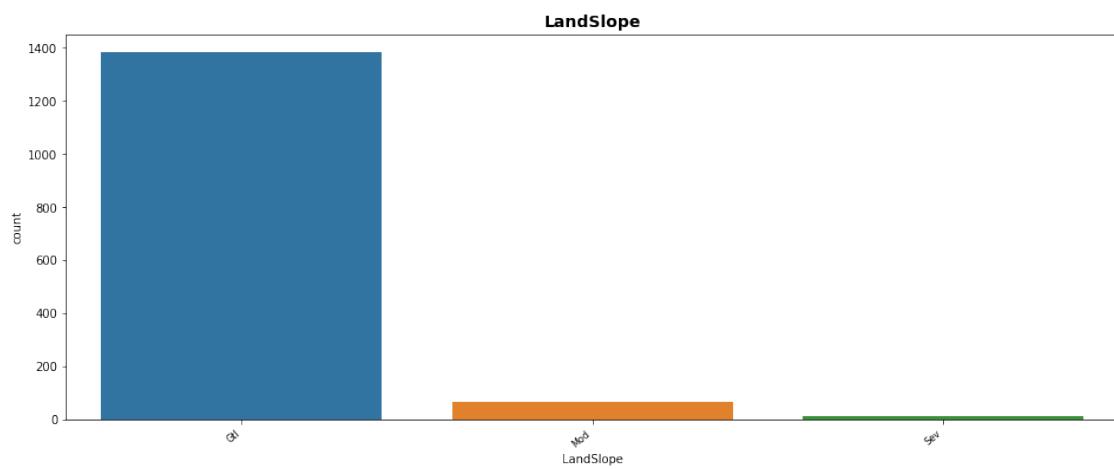
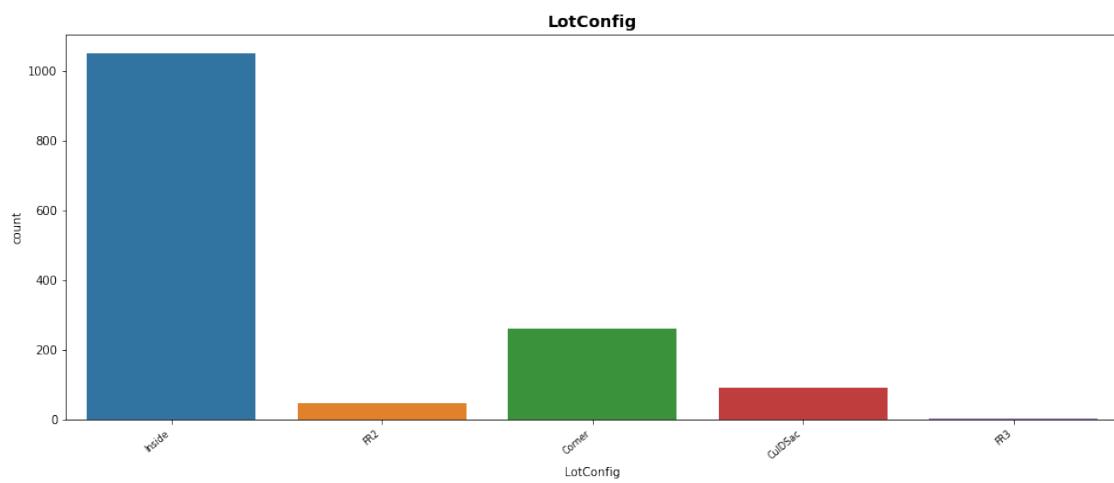
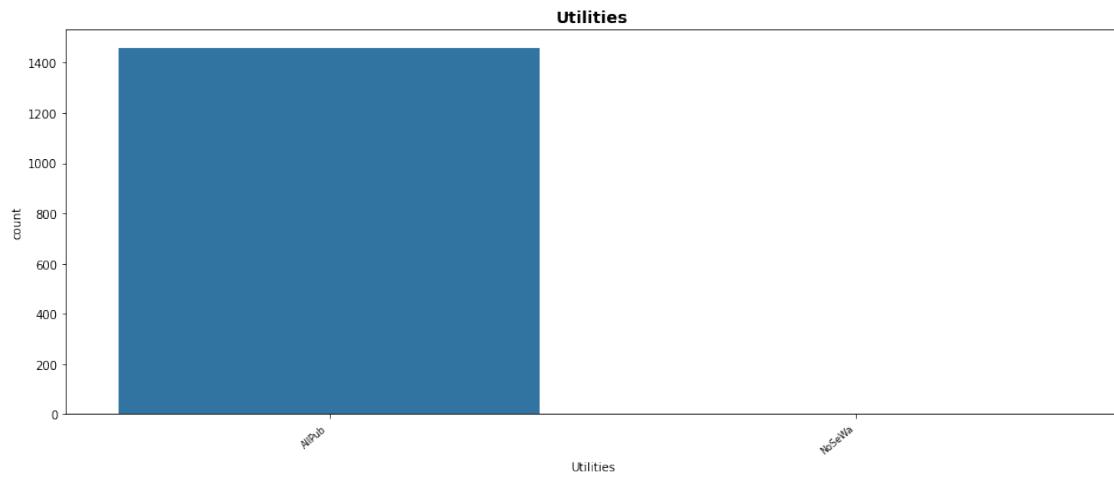
```

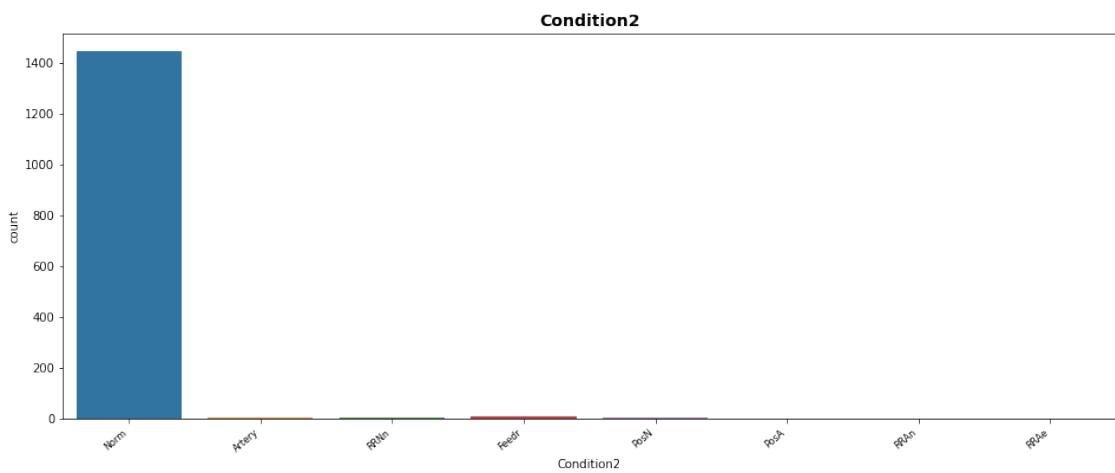
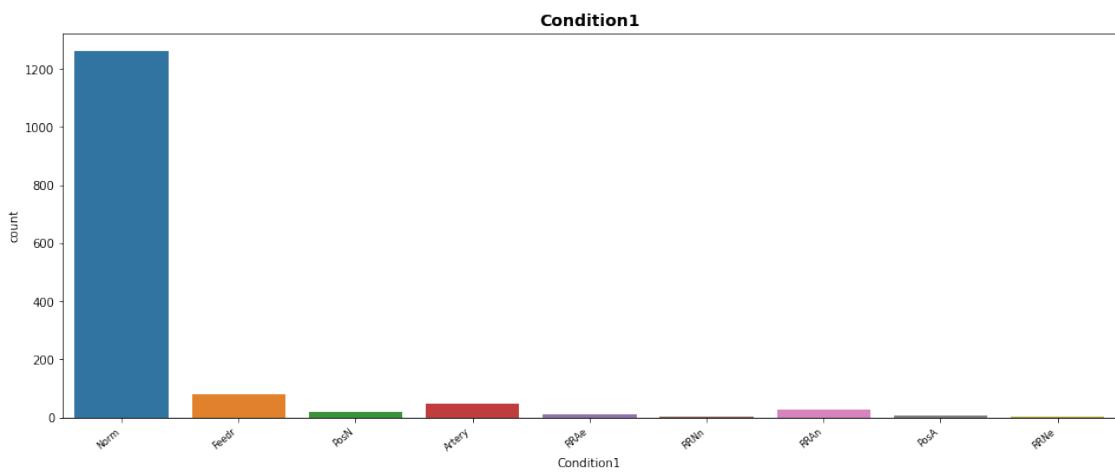
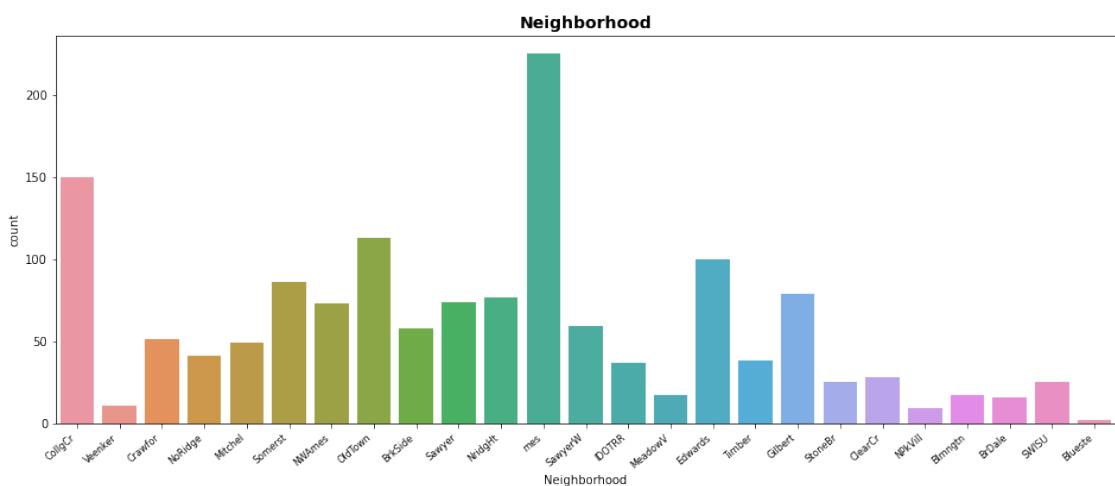
```
plt.show()
```

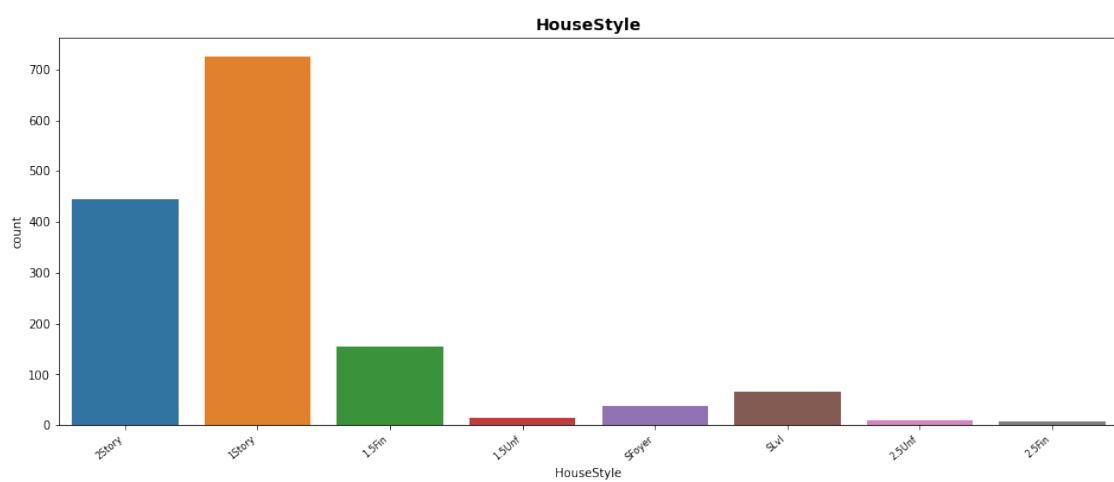
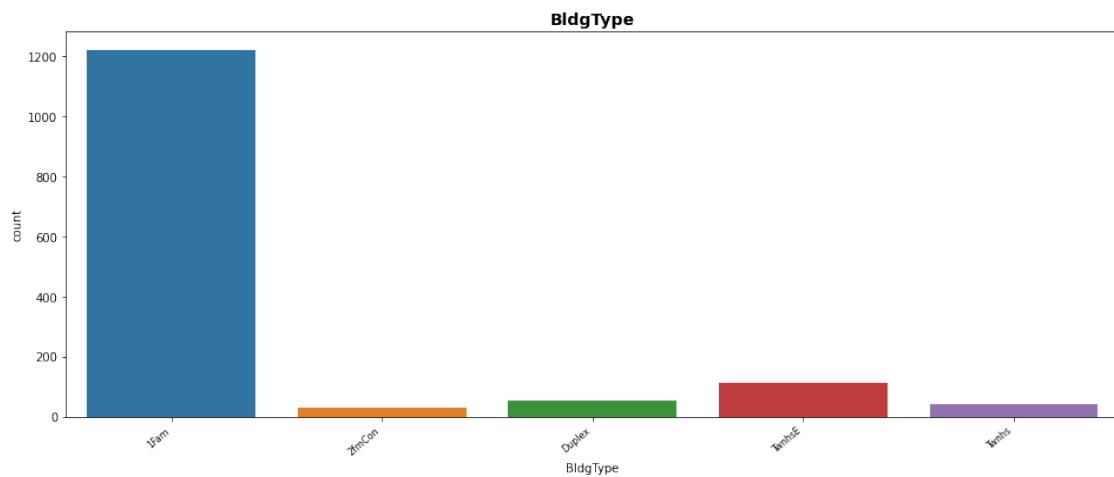
```
[36]: histPlotLoop(cat_variable,cat_variable.columns)
```

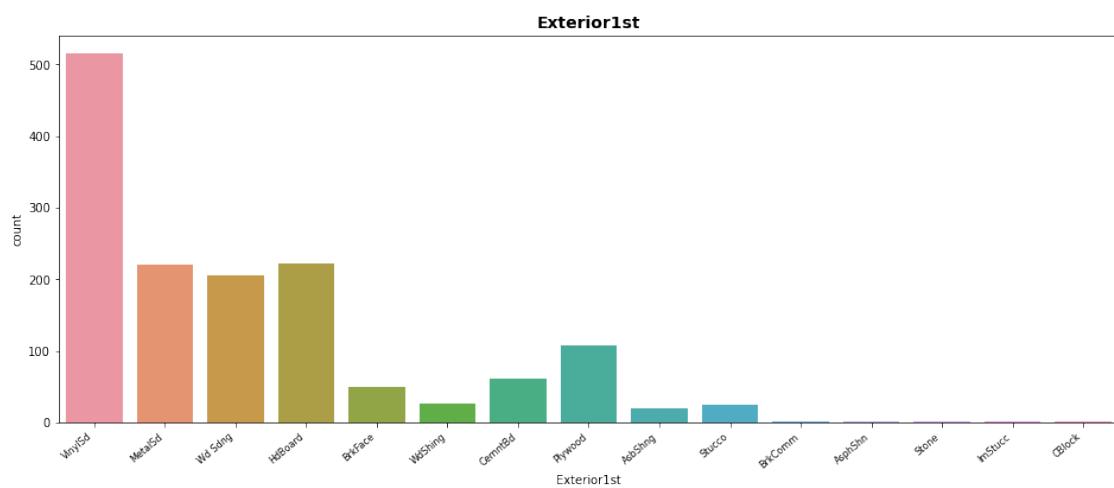
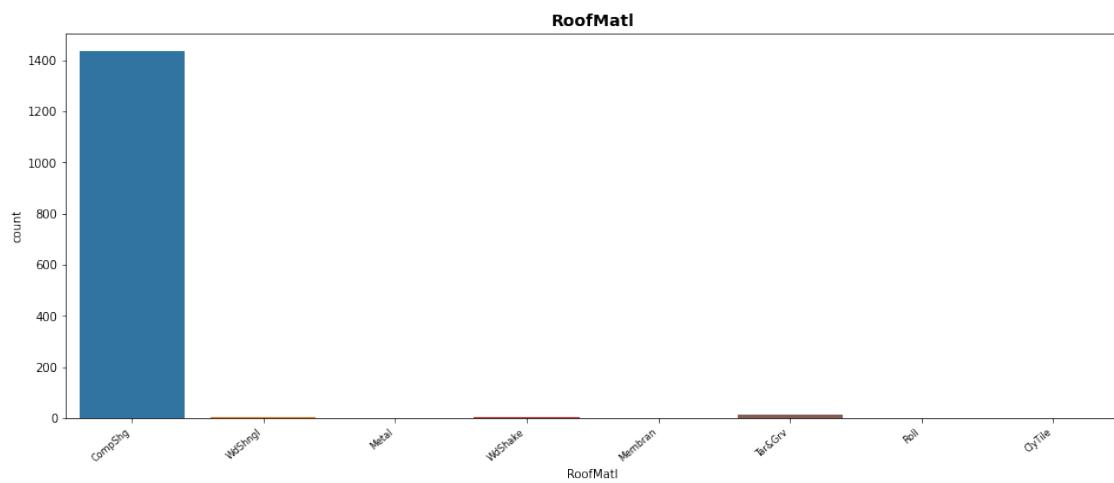
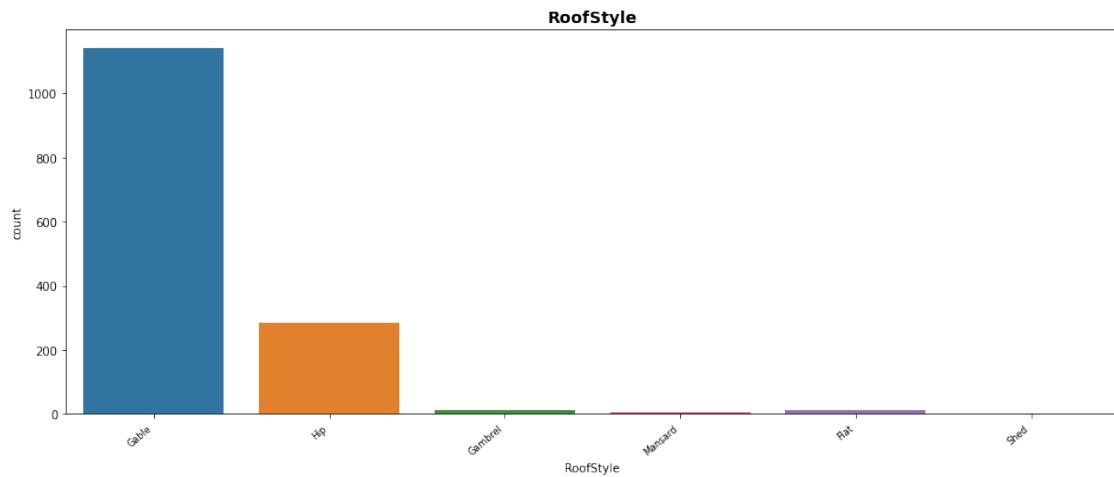


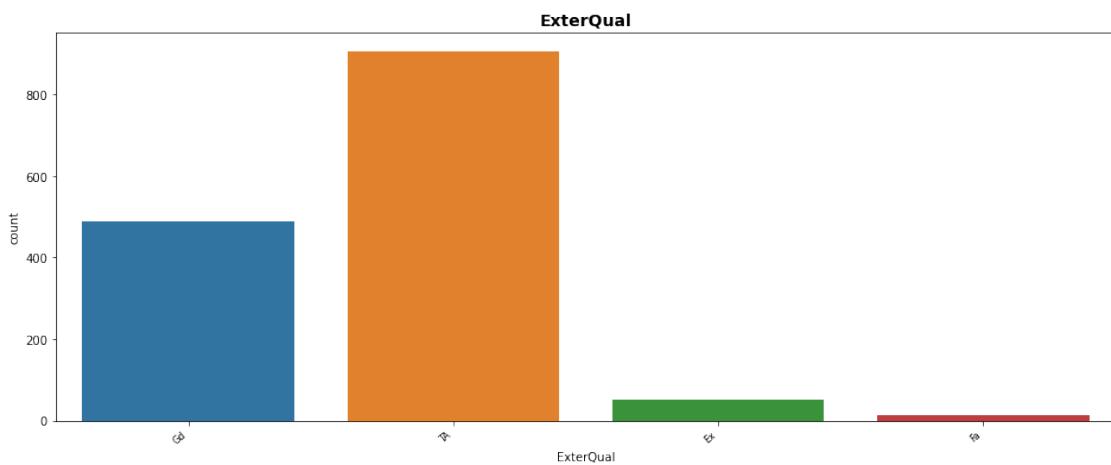
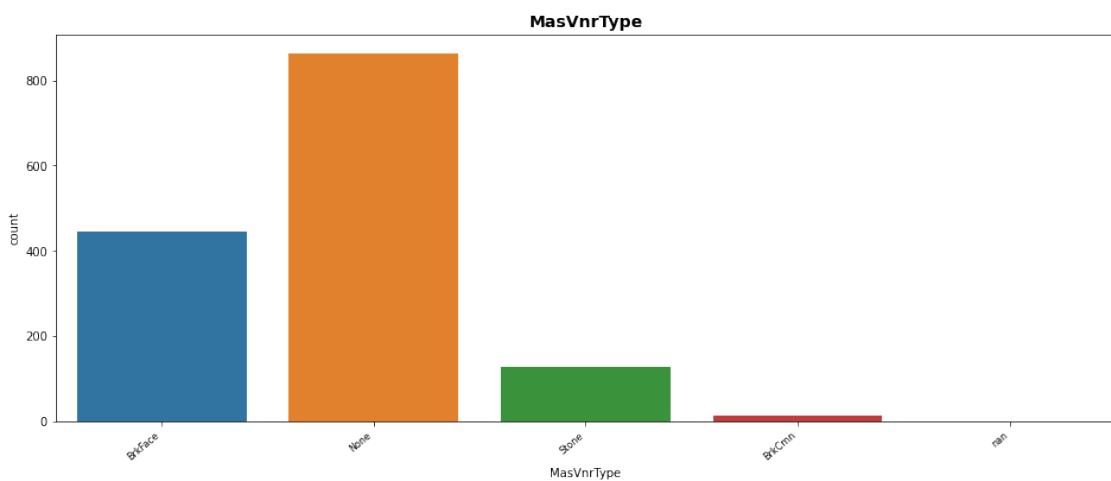
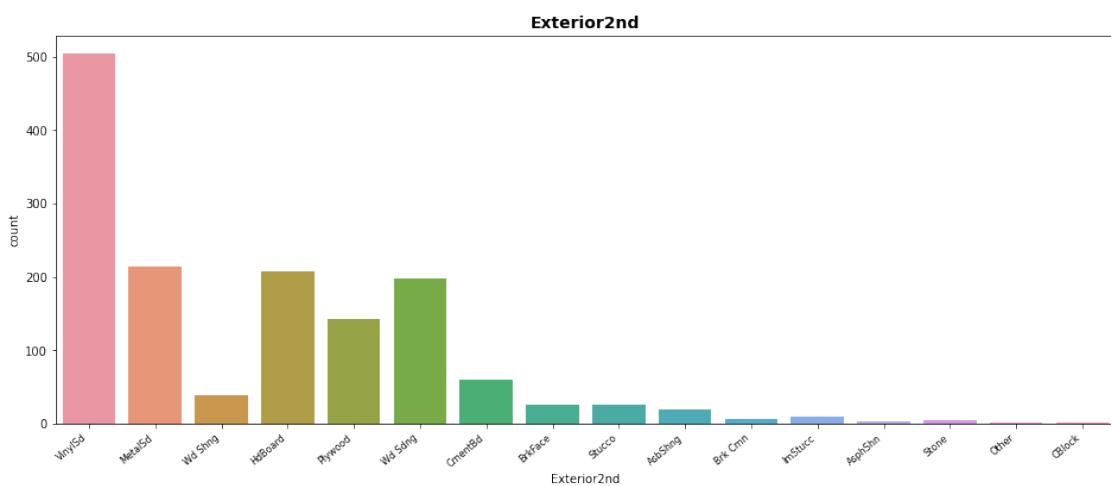


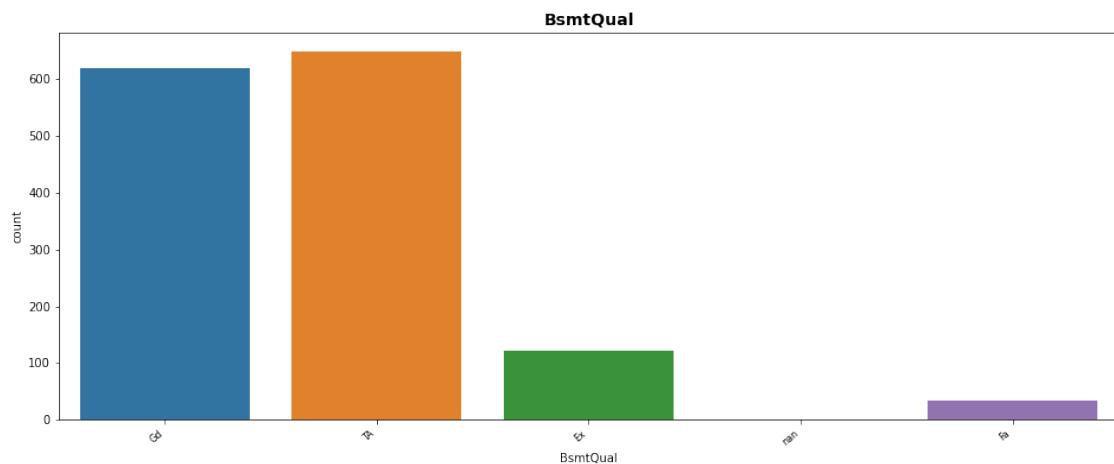
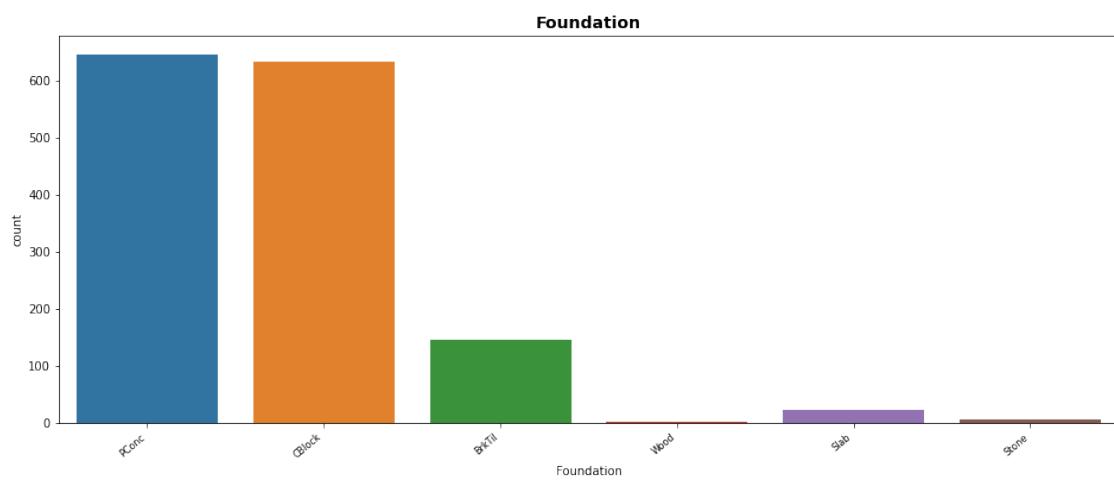
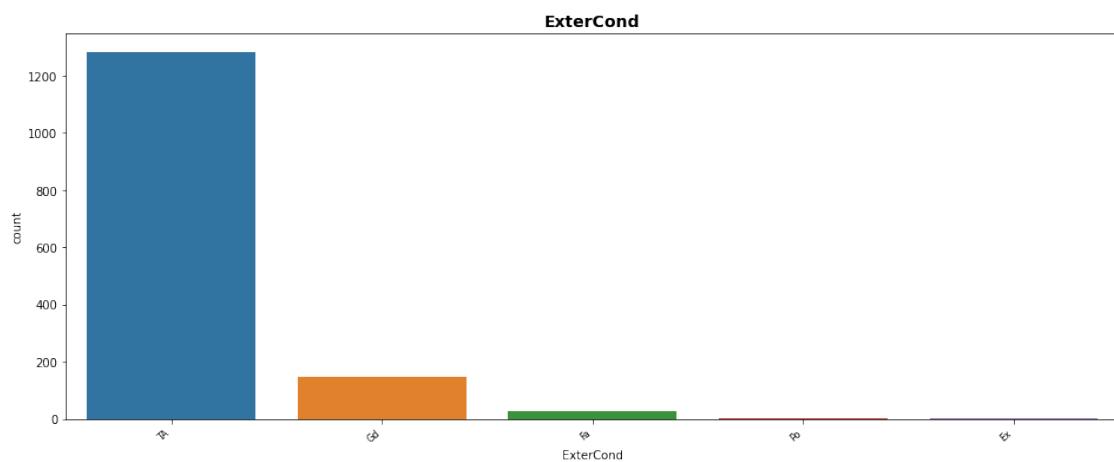


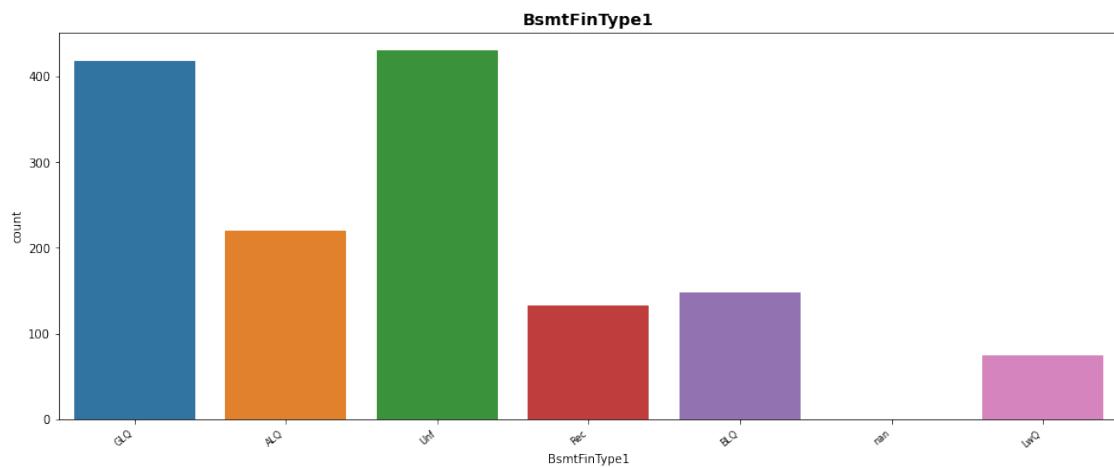
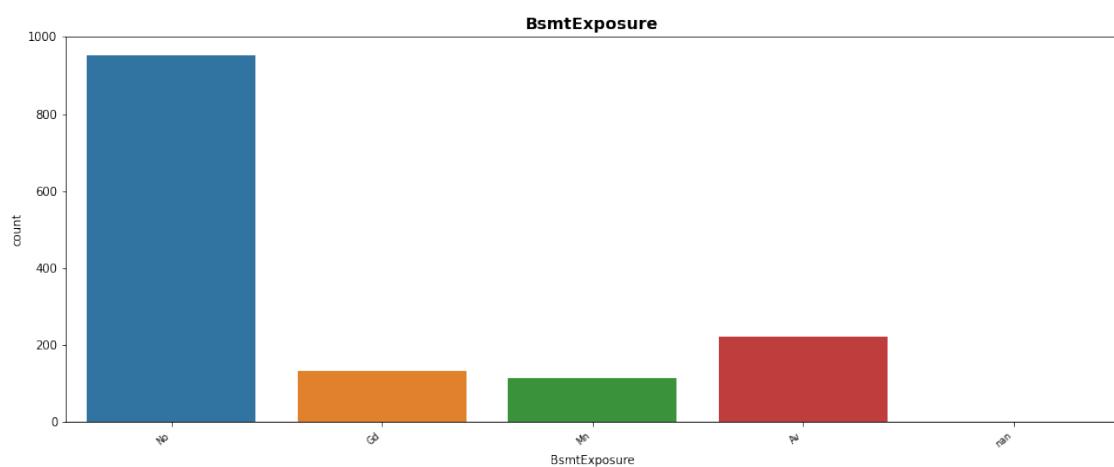
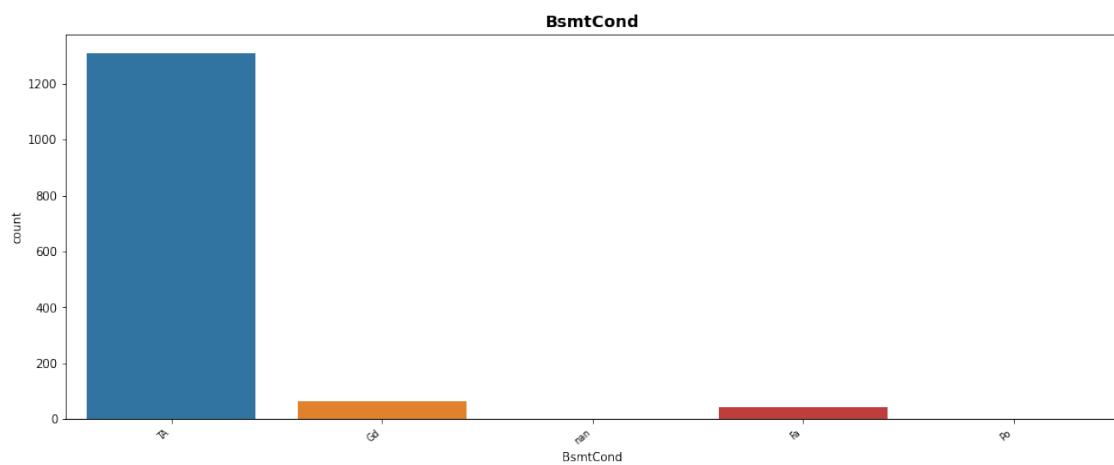


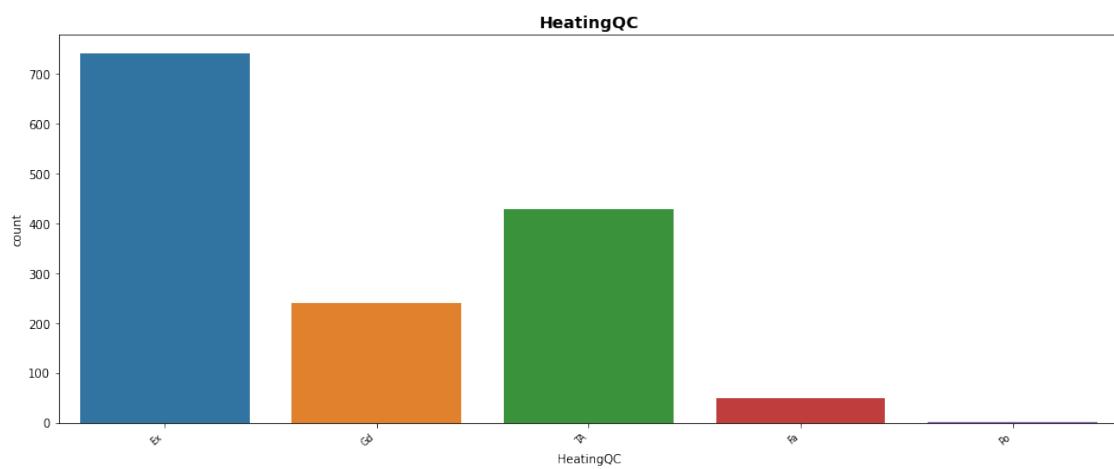
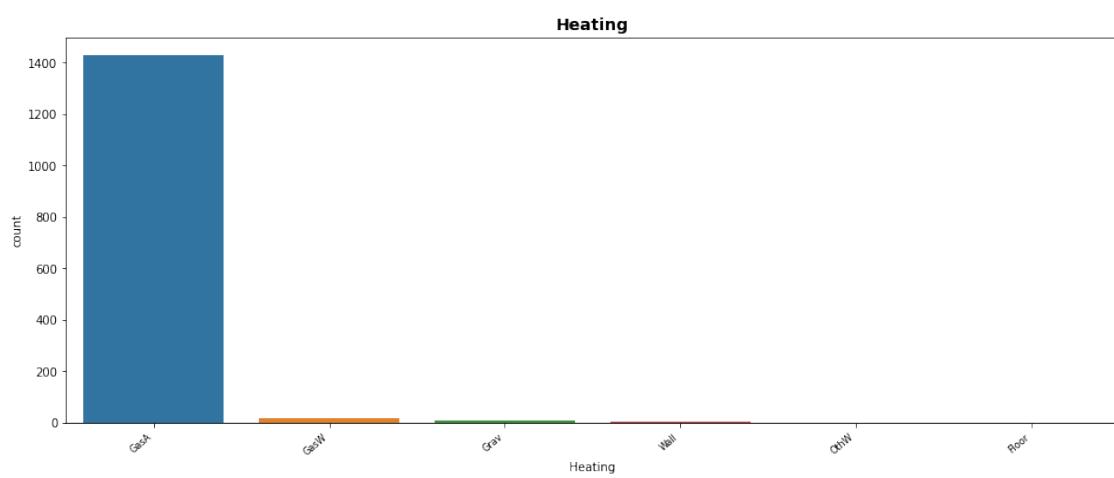
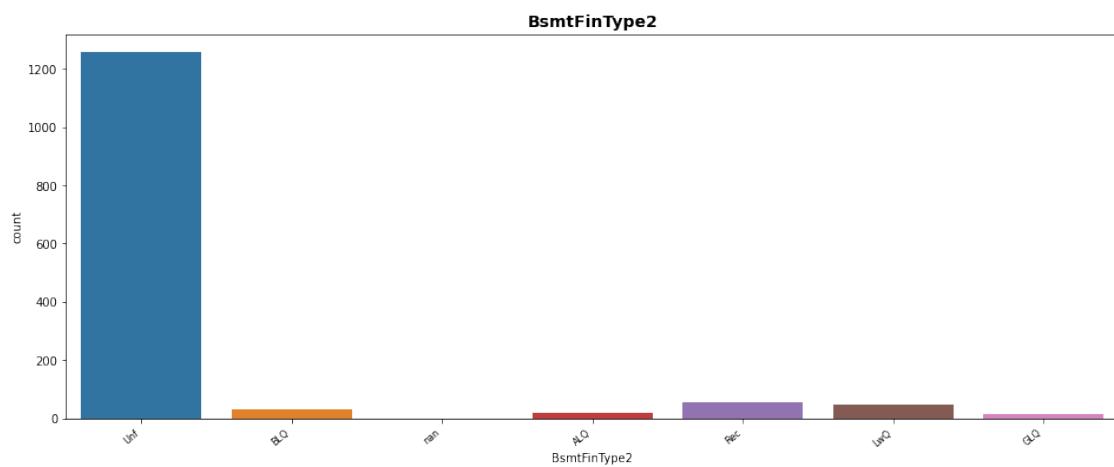


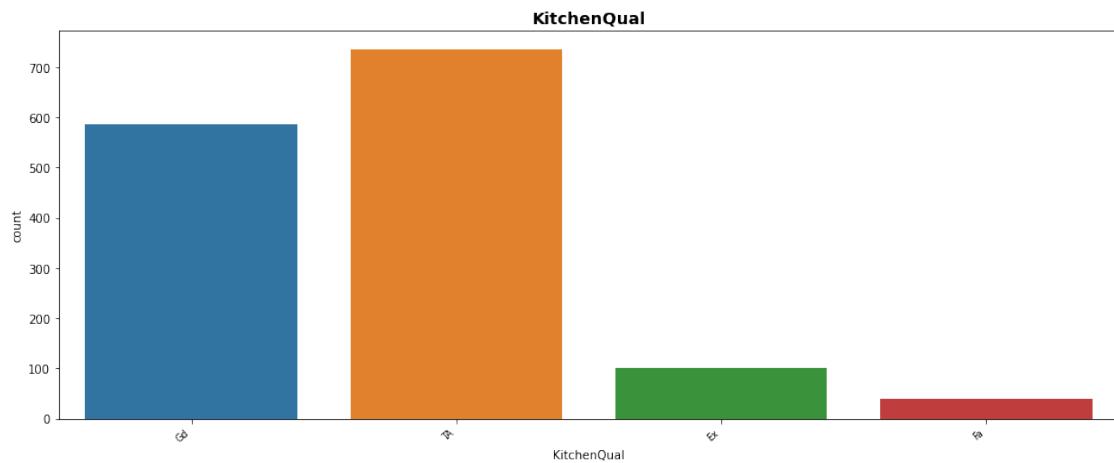
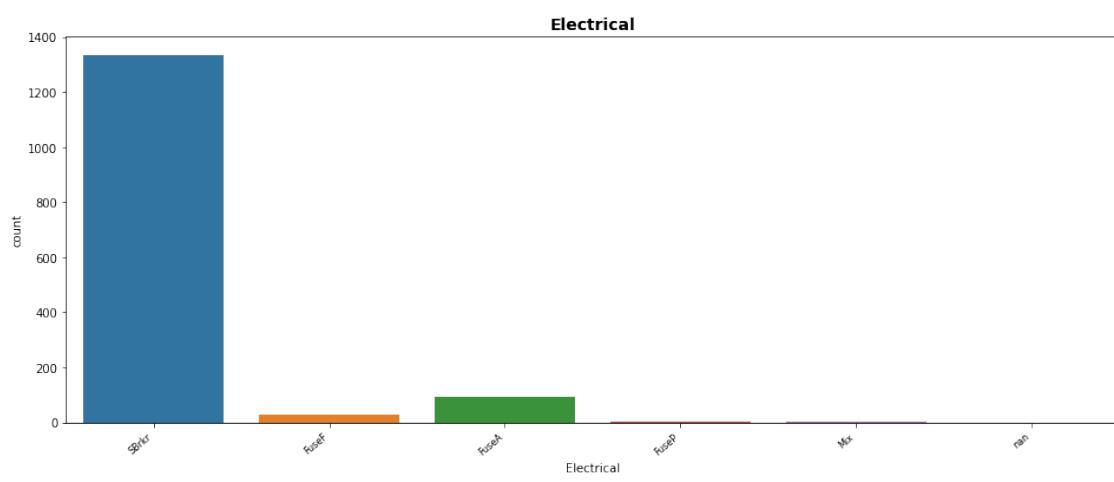
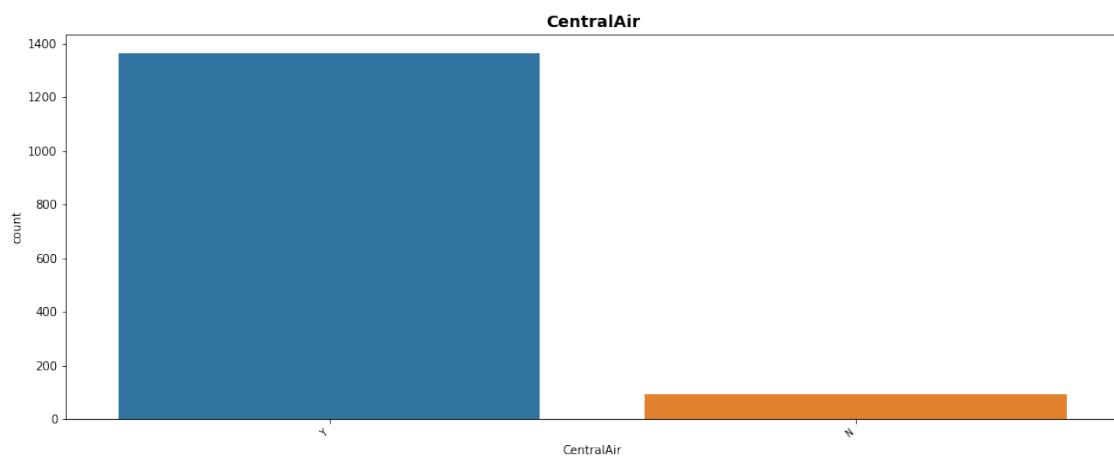


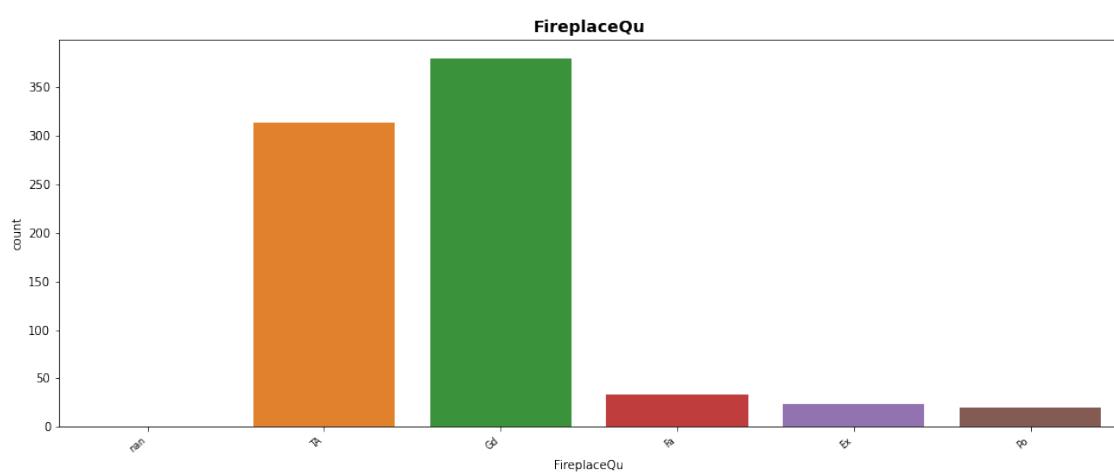
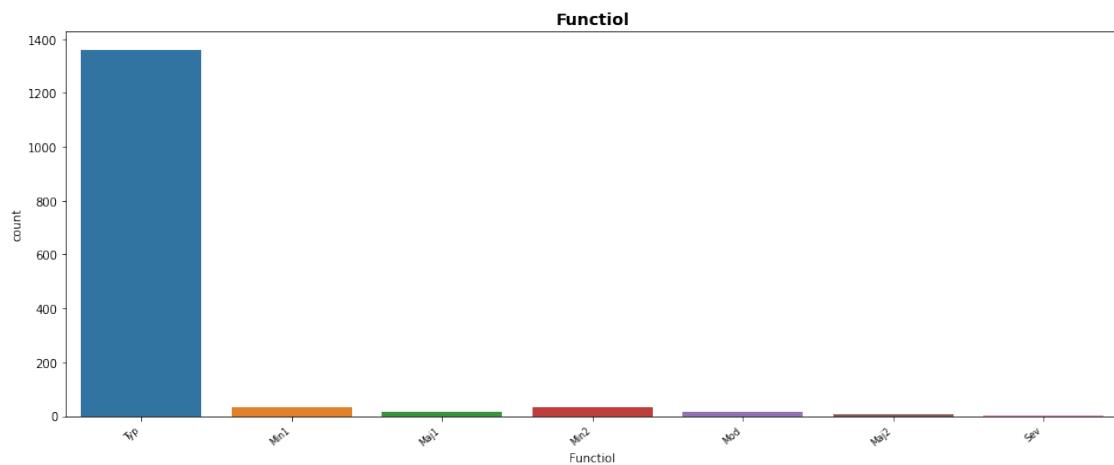


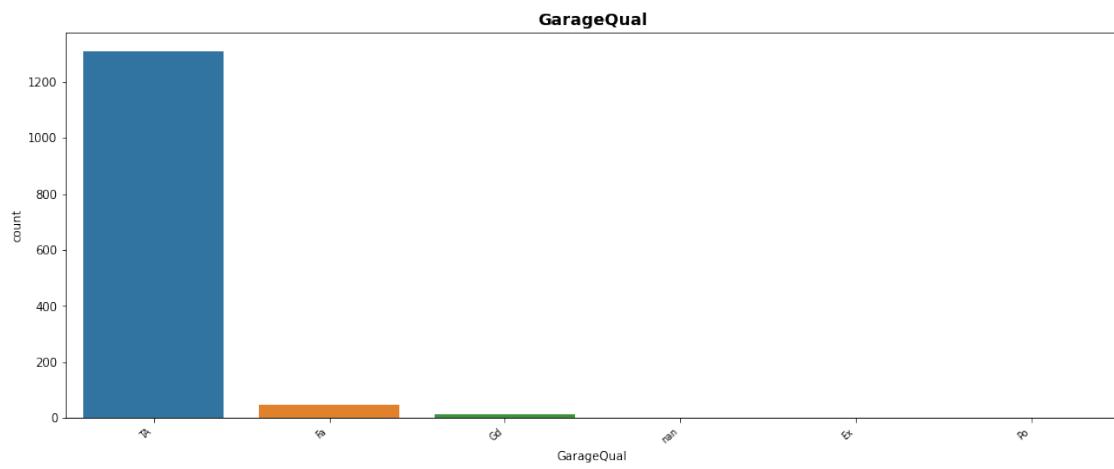
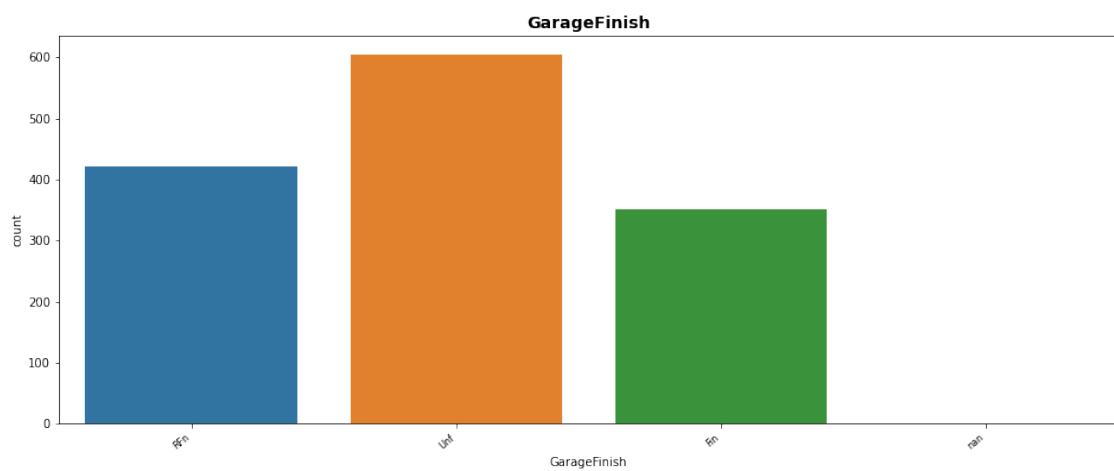
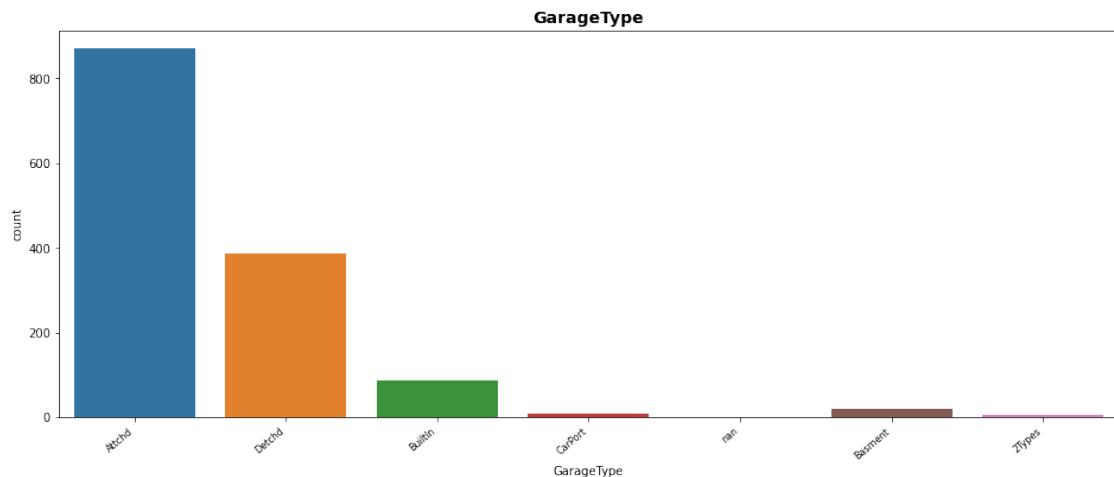


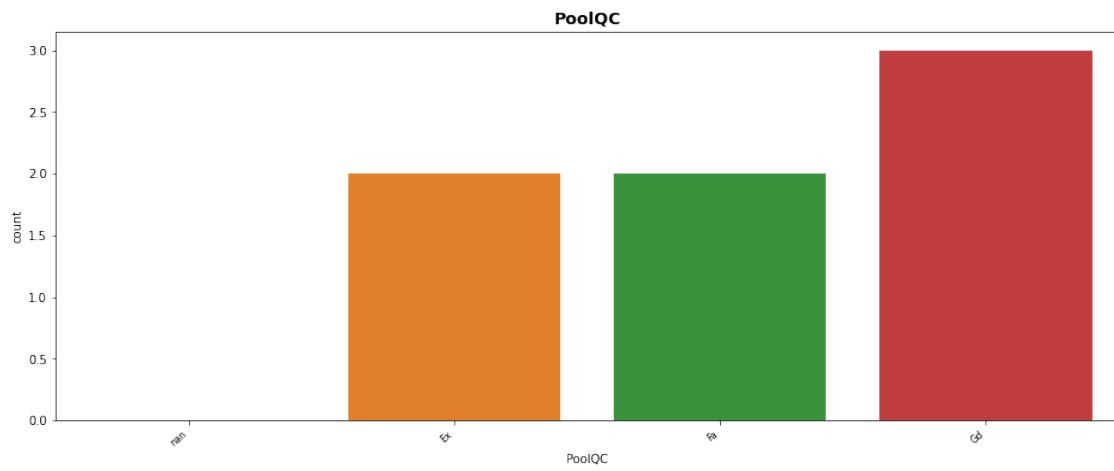
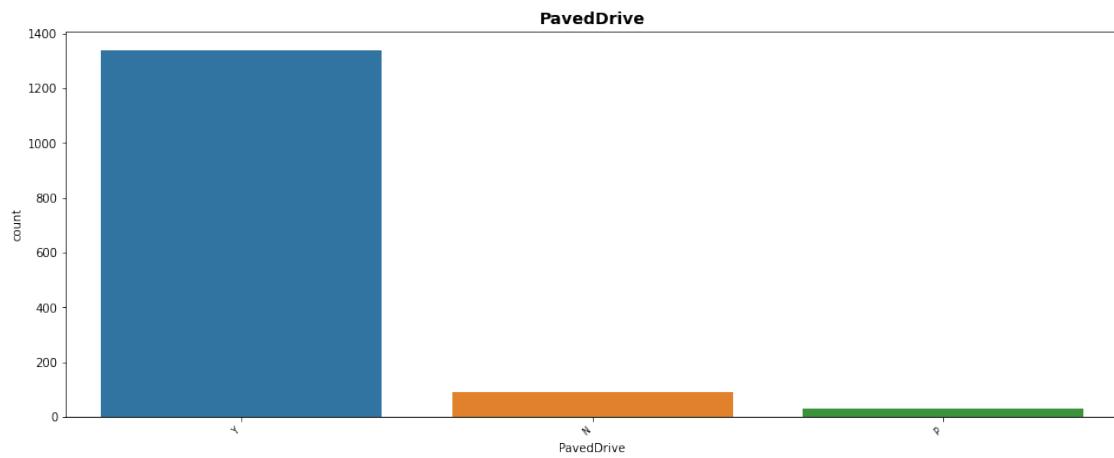
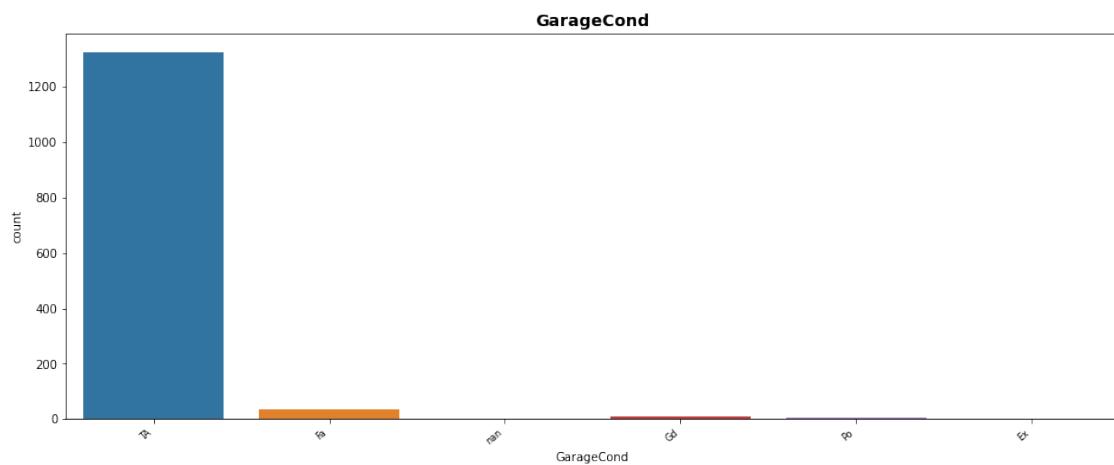


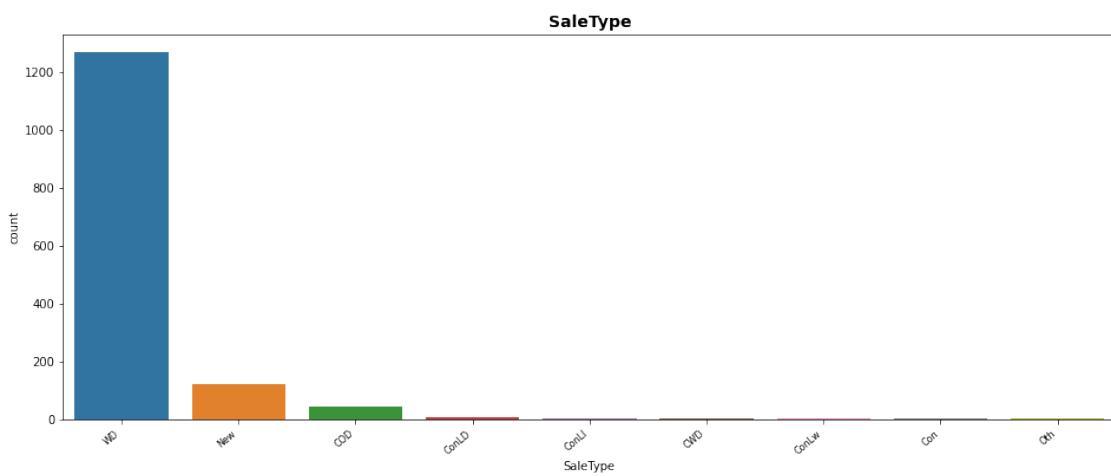
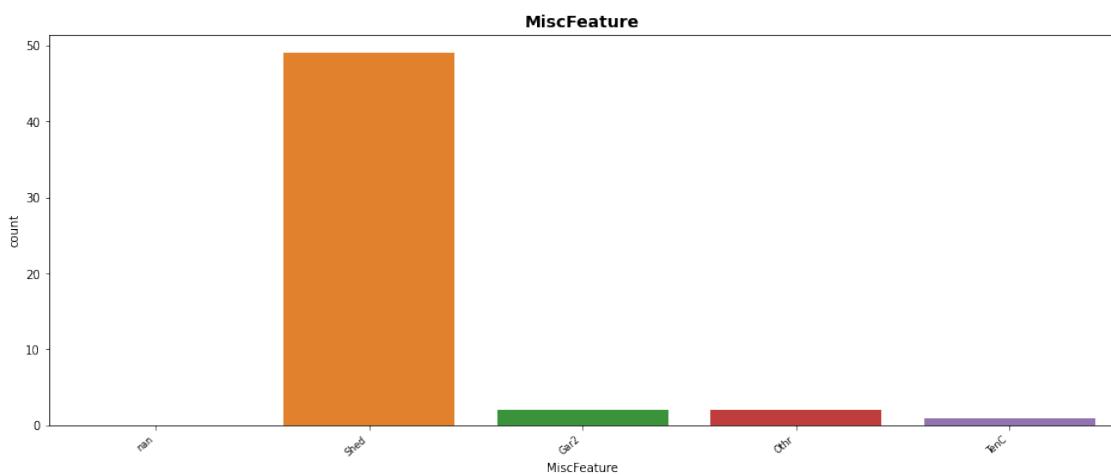
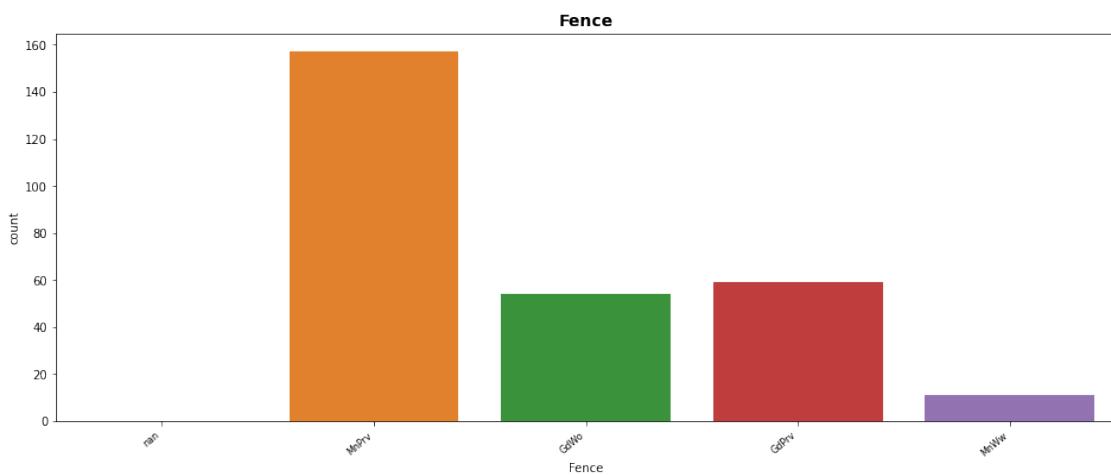


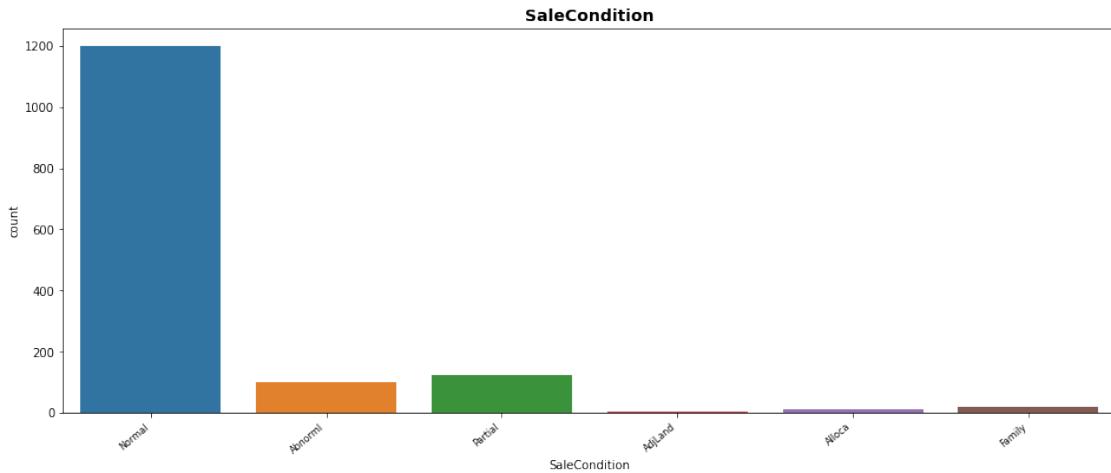












```
[38]: #convert the categorical variable into binary to find the chi_square
! pip install category_encoders
```

Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: category_encoders in ./local/lib/python3.7/site-packages (2.6.1)
Requirement already satisfied: pandas>=1.0.5 in /usr/local/lib/python3.7/site-packages (from category_encoders) (1.1.5)
Requirement already satisfied: scikit-learn>=0.20.0 in /usr/local/lib/python3.7/site-packages (from category_encoders) (1.0.2)
Requirement already satisfied: patsy>=0.5.1 in /usr/local/lib/python3.7/site-packages (from category_encoders) (0.5.1)
Requirement already satisfied: statsmodels>=0.9.0 in /usr/local/lib/python3.7/site-packages (from category_encoders) (0.11.1)
Requirement already satisfied: scipy>=1.0.0 in /usr/local/lib/python3.7/site-packages (from category_encoders) (1.4.1)
Requirement already satisfied: numpy>=1.14.0 in /usr/local/lib/python3.7/site-packages (from category_encoders) (1.21.5)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/site-packages (from pandas>=1.0.5->category_encoders) (2019.3)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/site-packages (from pandas>=1.0.5->category_encoders) (2.8.1)
Requirement already satisfied: six in /usr/local/lib/python3.7/site-packages (from patsy>=0.5.1->category_encoders) (1.14.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/site-packages (from scikit-learn>=0.20.0->category_encoders) (2.2.0)
Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.7/site-

```
packages (from scikit-learn>=0.20.0->category_encoders) (0.14.1)
WARNING: You are using pip version 22.0.3; however, version 23.2 is
available.

You should consider upgrading via the '/usr/local/bin/python3 -m pip install
--upgrade pip' command.
```

```
[40]: import category_encoders as ce
```

```
[41]: encoder = ce.ordinal.OrdinalEncoder(return_df = True)
df_train = encoder.fit_transform(cat_variable)
df_train.head()
```

```
[41]:   MSZoning  Street  Alley  LotShape  LandContour  Utilities  LotConfig \
0          1       1     3         1            1           1           1
1          1       1     3         1            1           1           2
2          1       1     3         2            1           1           1
3          1       1     3         2            1           1           3
4          1       1     3         2            1           1           2

   LandSlope  Neighborhood  Condition1 ... GarageType  GarageFinish \
0          1             1          1 ...           1             1
1          1             2          2 ...           1             1
2          1             1          1 ...           1             1
3          1             3          1 ...           2             2
4          1             4          1 ...           1             1

   GarageQual  GarageCond  PavedDrive  PoolQC  Fence  MiscFeature  SaleType \
0          1           1          1        4       5           5           1
1          1           1          1        4       5           5           1
2          1           1          1        4       5           5           1
3          1           1          1        4       5           5           1
4          1           1          1        4       5           5           1

   SaleCondition
0              1
1              1
2              1
3              2
4              1

[5 rows x 43 columns]
```

```
[40]: cat_variable = df_train
```

```
[41]: cat_variable.head()
```

```
[41]:   Id  MSZoning  Street  Alley  LotShape  LandContour  Utilities  LotConfig \
0    1        1       1     3         1            1           1           1
1    2        1       1     3         1            1           1           2
2    3        1       1     3         2            1           1           1
3    4        1       1     3         2            1           1           3
4    5        1       1     3         2            1           1           2

      LandSlope  Neighborhood ... GarageFinish  GarageQual  GarageCond \
0          1             1   ...           1           1           1
1          1             2   ...           1           1           1
2          1             1   ...           1           1           1
3          1             3   ...           2           1           1
4          1             4   ...           1           1           1

  PavedDrive  PoolQC  Fence  MiscFeature  SaleType  SaleCondition  SalePrice
0          1       4      5          5         1            1      208500
1          1       4      5          5         1            1      181500
2          1       4      5          5         1            1      223500
3          1       4      5          5         1            2      140000
4          1       4      5          5         1            1      250000
```

[5 rows x 45 columns]

```
[42]: cat_variable.corr()
```

```
[42]:          Id  MSZoning  Street  Alley  LotShape  LandContour \
Id  1.000000 -0.011949 -0.008916 -0.001530 -0.024071  0.014769
MSZoning -0.011949  1.000000  0.039678 -0.250108 -0.167918 -0.013396
Street  -0.008916  0.039678  1.000000  0.015720  0.010129  0.097236
Alley   -0.001530 -0.250108  0.015720  1.000000  0.097078 -0.017513
LotShape -0.024071 -0.167918  0.010129  0.097078  1.000000  0.201047
LandContour  0.014769 -0.013396  0.097236 -0.017513  0.201047  1.000000
Utilities  0.013324 -0.011167 -0.001682  0.006407  0.026616 -0.007963
LotConfig  -0.038192 -0.094747  0.004458  0.065500  0.308073  0.021107
LandSlope  0.005847 -0.043444  0.179360  0.030096  0.144248  0.507203
Neighborhood -0.013078 -0.067889  0.053445  0.043543 -0.013295  0.113336
Condition1 -0.010735 -0.026021  0.023914 -0.025739  0.066612 -0.021478
Condition2 -0.016961  0.016014 -0.005881  0.022405  0.052183  0.012096
BldgType   0.019230  0.271750  0.023392 -0.029783 -0.134373 -0.042667
HouseStyle  0.026166 -0.047199  0.025328 -0.008679 -0.067647 -0.004987
RoofStyle   0.015873 -0.088537  0.006881  0.030532  0.043063  0.100625
RoofMatl   0.013375 -0.049864 -0.007749  0.029520  0.118271  0.109717
Exterior1st 0.003624 -0.063732  0.011874 -0.040385 -0.030630  0.034502
Exterior2nd 0.015449 -0.025637  0.015314 -0.073012 -0.068168  0.014566
MasVnrType  0.030431  0.088915 -0.014663 -0.047581  0.025179  0.073687
```

ExterQual	-0.021844	-0.018093	0.143047	-0.052568	-0.111502	0.075493
ExterCond	-0.037865	0.017808	0.002633	-0.063272	-0.047015	0.001553
Foundation	0.020023	0.011181	0.021128	-0.155687	-0.158779	-0.000145
BsmtQual	-0.018676	-0.025631	-0.008485	-0.067690	-0.131625	0.011679
BsmtCond	-0.011355	0.011452	-0.018487	-0.042854	-0.064224	-0.001913
BsmtExposure	0.027692	-0.076712	0.035998	0.111678	0.065124	0.066580
BsmtFinType1	0.009911	0.003967	0.013592	-0.082321	-0.106861	0.006041
BsmtFinType2	0.002847	-0.081051	0.024109	0.057510	-0.008548	-0.024782
Heating	0.061902	0.012889	-0.008194	-0.110547	-0.043320	-0.024849
HeatingQC	0.015001	-0.035643	0.043211	-0.040666	-0.119556	0.028243
CentralAir	-0.009821	0.144970	0.069869	-0.202835	-0.099138	0.023605
Electrical	-0.037032	0.079803	0.020475	-0.119392	-0.088752	0.018633
KitchenQual	-0.020927	-0.030523	0.055746	-0.059979	-0.089061	0.037246
Functiol	0.006018	-0.029668	-0.015465	-0.045178	-0.017841	0.016088
FireplaceQu	-0.000390	0.197296	0.021227	-0.113586	-0.206050	-0.066514
GarageType	-0.007207	0.113987	0.101750	-0.110622	-0.100828	0.039636
GarageFinish	-0.009336	0.039377	0.007698	-0.030331	0.012848	0.059278
GarageQual	-0.017357	0.075028	0.026658	-0.094932	-0.093695	0.014482
GarageCond	-0.012836	0.071517	0.027236	-0.098604	-0.098629	0.017144
PavedDrive	0.008921	0.117402	0.011248	-0.235390	-0.100666	0.081031
PoolQC	-0.033928	0.027003	0.004067	-0.015492	-0.030895	0.011667
Fence	0.012682	0.060229	0.029920	0.015728	0.078494	0.070593
MiscFeature	0.044733	0.033389	-0.163140	0.028485	-0.010451	-0.005240
SaleType	-0.016932	0.058948	0.021171	-0.004045	0.001435	-0.014537
SaleCondition	-0.017712	0.068705	0.022919	-0.016915	-0.004773	0.062826
SalePrice	-0.021917	-0.116047	-0.041036	0.139868	0.267759	0.092009

	Utilities	LotConfig	LandSlope	Neighborhood	...	\
Id	0.013324	-0.038192	0.005847	-0.013078	...	
MSZoning	-0.011167	-0.094747	-0.043444	-0.067889	...	
Street	-0.001682	0.004458	0.179360	0.053445	...	
Alley	0.006407	0.065500	0.030096	0.043543	...	
LotShape	0.026616	0.308073	0.144248	-0.013295	...	
LandContour	-0.007963	0.021107	0.507203	0.113336	...	
Utilities	1.000000	0.062298	-0.005909	0.029258	...	
LotConfig	0.062298	1.000000	0.033673	-0.055593	...	
LandSlope	-0.005909	0.033673	1.000000	0.072891	...	
Neighborhood	0.029258	-0.055593	0.072891	1.000000	...	
Condition1	-0.008311	0.061695	-0.020333	0.002747	...	
Condition2	-0.002397	0.018440	-0.014034	-0.012207	...	
BldgType	-0.010899	-0.106812	-0.056801	0.184198	...	
HouseStyle	0.076219	0.005696	0.018446	0.060100	...	
RoofStyle	-0.011462	0.067337	0.189081	0.061029	...	
RoofMatl	-0.003158	0.112297	0.221080	0.091511	...	
Exterior1st	0.008321	0.022409	0.110339	0.102043	...	
Exterior2nd	0.004602	0.015406	0.093730	0.121032	...	
MasVnrType	-0.031982	-0.003133	0.056888	0.013245	...	

ExterQual	0.012733	-0.037293	0.092514	0.176707	...
ExterCond	-0.008842	0.008542	0.002675	-0.018067	...
Foundation	0.008611	0.000844	0.044388	0.104070	...
BsmtQual	0.006402	0.000397	-0.009260	0.107559	...
BsmtCond	-0.007534	0.007984	-0.007621	0.096439	...
BsmtExposure	-0.017046	0.051940	0.065387	-0.045852	...
BsmtFinType1	-0.029484	-0.005728	0.009217	0.149344	...
BsmtFinType2	0.048337	0.005710	0.048227	0.058629	...
Heating	-0.003339	0.004421	0.003625	0.000671	...
HeatingQC	0.003963	-0.018754	0.057444	0.130972	...
CentralAir	-0.006907	-0.026600	0.010849	0.070703	...
Electrical	0.088392	-0.041057	0.011273	0.075490	...
KitchenQual	-0.026715	-0.046671	0.029612	0.153756	...
Functiol	-0.006303	0.008194	0.097882	0.077679	...
FireplaceQu	-0.021082	-0.073301	-0.069081	-0.037018	...
GarageType	0.056129	-0.048375	0.043883	0.133600	...
GarageFinish	-0.032176	-0.000694	0.020837	0.196493	...
GarageQual	-0.007631	-0.045119	0.005701	0.081193	...
GarageCond	-0.007409	-0.037126	0.007672	0.092220	...
PavedDrive	-0.007368	-0.065917	0.018009	0.065998	...
PoolQC	0.001657	-0.044571	0.014289	0.019372	...
Fence	0.012194	-0.015429	0.035386	0.031296	...
MiscFeature	0.005073	-0.012520	-0.051181	-0.008989	...
SaleType	0.057455	0.012120	-0.056680	0.008466	...
SaleCondition	0.019173	-0.013638	0.017187	0.007101	...
SalePrice	-0.014314	0.109106	0.051152	-0.143621	...

	GarageFinish	GarageQual	GarageCond	PavedDrive	PoolQC	\
Id	-0.009336	-0.017357	-0.012836	0.008921	-0.033928	
MSZoning	0.039377	0.075028	0.071517	0.117402	0.027003	
Street	0.007698	0.026658	0.027236	0.011248	0.004067	
Alley	-0.030331	-0.094932	-0.098604	-0.235390	-0.015492	
LotShape	0.012848	-0.093695	-0.098629	-0.100666	-0.030895	
LandContour	0.059278	0.014482	0.017144	0.081031	0.011667	
Utilities	-0.032176	-0.007631	-0.007409	-0.007368	0.001657	
LotConfig	-0.000694	-0.045119	-0.037126	-0.065917	-0.044571	
LandSlope	0.020837	0.005701	0.007672	0.018009	0.014289	
Neighborhood	0.196493	0.081193	0.092220	0.065998	0.019372	
Condition1	0.012920	-0.015162	-0.016253	0.034886	-0.068573	
Condition2	0.008149	0.019882	0.003607	0.054484	0.005796	
BldgType	0.043835	0.040162	0.038377	-0.064447	0.026355	
HouseStyle	0.013577	0.092276	0.089343	0.038707	0.010586	
RoofStyle	0.045779	-0.028896	-0.012751	0.011743	-0.051893	
RoofMatl	0.023509	0.005744	0.015417	-0.033947	-0.100694	
Exterior1st	0.044642	0.072491	0.074176	0.058754	-0.092011	
Exterior2nd	0.050297	0.100327	0.094935	0.119255	-0.098952	
MasVnrType	0.053773	0.044496	0.044785	0.069944	0.004776	

ExterQual	0.151209	0.189273	0.195280	0.166263	0.020062
ExterCond	0.085163	0.178642	0.168506	0.139432	-0.022881
Foundation	0.024916	0.182532	0.170701	0.243846	0.020203
BsmtQual	0.123399	0.119180	0.122385	0.168455	-0.020983
BsmtCond	0.079177	0.121402	0.143749	0.148613	0.018219
BsmtExposure	0.009487	-0.020492	-0.012015	-0.015039	-0.021710
BsmtFinType1	0.008407	0.092989	0.096412	0.141792	0.032206
BsmtFinType2	-0.000614	0.027401	0.039962	0.009026	-0.077483
Heating	0.057977	0.118384	0.130970	0.141864	0.008075
HeatingQC	0.020111	0.139173	0.148856	0.193215	-0.029891
CentralAir	0.147735	0.288008	0.295220	0.275660	0.016701
Electrical	0.065836	0.151144	0.156999	0.168861	0.018534
KitchenQual	0.225090	0.228018	0.233780	0.164511	0.002532
Functiol	0.036582	0.065887	0.056685	0.047388	0.015240
FireplaceQu	0.022276	0.198937	0.214806	0.108336	0.037704
GarageType	0.505294	0.824347	0.823027	0.242775	0.024623
GarageFinish	1.000000	0.518531	0.521545	0.133507	-0.001018
GarageQual	0.518531	1.000000	0.968264	0.255161	0.010196
GarageCond	0.521545	0.968264	1.000000	0.268934	0.009652
PavedDrive	0.133507	0.255161	0.268934	1.000000	0.017815
PoolQC	-0.001018	0.010196	0.009652	0.017815	1.000000
Fence	0.056200	-0.020630	-0.035884	-0.034126	0.093026
MiscFeature	0.034681	-0.005417	-0.003933	0.000685	0.001044
SaleType	0.026965	0.041921	0.048109	-0.021553	0.012437
SaleCondition	0.051570	0.006738	0.004943	-0.049674	-0.040874
SalePrice	-0.006799	-0.245912	-0.261527	-0.208954	-0.126070

	Fence	MiscFeature	SaleType	SaleCondition	SalePrice
Id	0.012682	0.044733	-0.016932	-0.017712	-0.021917
MSZoning	0.060229	0.033389	0.058948	0.068705	-0.116047
Street	0.029920	-0.163140	0.021171	0.022919	-0.041036
Alley	0.015728	0.028485	-0.004045	-0.016915	0.139868
LotShape	0.078494	-0.010451	0.001435	-0.004773	0.267759
LandContour	0.070593	-0.005240	-0.014537	0.062826	0.092009
Utilities	0.012194	0.005073	0.057455	0.019173	-0.014314
LotConfig	-0.015429	-0.012520	0.012120	-0.013638	0.109106
LandSlope	0.035386	-0.051181	-0.056680	0.017187	0.051152
Neighborhood	0.031296	-0.008989	0.008466	0.007101	-0.143621
Condition1	-0.031674	-0.001242	0.003128	-0.019800	-0.044820
Condition2	-0.016973	-0.064842	0.011960	0.005158	-0.004833
BldgType	0.145490	0.056123	0.044289	0.040365	-0.112611
HouseStyle	-0.163869	-0.021265	-0.006154	-0.029049	-0.188688
RoofStyle	0.000924	-0.013105	-0.031824	0.032990	0.159332
RoofMatl	0.025178	0.023374	-0.014712	0.055436	0.035820
Exterior1st	-0.140353	0.004149	-0.015554	0.000631	-0.120586
Exterior2nd	-0.115956	-0.010142	-0.033452	-0.039999	-0.164716
MasVnrType	0.016956	-0.022719	0.070788	0.124239	0.007123

ExterQual	-0.171845	-0.090986	0.022013	-0.009670	-0.265015
ExterCond	-0.085694	-0.017514	0.003909	-0.038491	-0.121706
Foundation	-0.153194	-0.115557	-0.050491	-0.093201	-0.429678
BsmtQual	-0.038887	-0.036810	0.013701	0.054583	-0.084708
BsmtCond	0.035492	-0.015516	0.016683	0.010336	-0.176352
BsmtExposure	0.062164	0.006711	0.028535	0.092899	0.132027
BsmtFinType1	-0.129792	-0.023372	-0.023989	-0.055771	-0.387251
BsmtFinType2	-0.068567	-0.032791	-0.021281	-0.035900	-0.137066
Heating	0.006423	0.012483	0.018388	-0.032650	-0.106673
HeatingQC	-0.204193	-0.042021	-0.009051	-0.049721	-0.427649
CentralAir	0.005749	-0.002015	0.026522	-0.010168	-0.251328
Electrical	-0.004963	0.029254	-0.016778	-0.024845	-0.226145
KitchenQual	-0.071642	-0.054149	-0.000178	0.033660	-0.114746
Functiol	-0.034905	-0.078691	-0.014494	-0.055108	-0.108367
FireplaceQu	-0.080348	-0.017466	0.021277	0.010237	-0.442834
GarageType	-0.016003	-0.016616	0.051111	0.031789	-0.273539
GarageFinish	0.056200	0.034681	0.026965	0.051570	-0.006799
GarageQual	-0.020630	-0.005417	0.041921	0.006738	-0.245912
GarageCond	-0.035884	-0.003933	0.048109	0.004943	-0.261527
PavedDrive	-0.034126	0.000685	-0.021553	-0.049674	-0.208954
PoolQC	0.093026	0.001044	0.012437	-0.040874	-0.126070
Fence	1.000000	0.120317	0.021994	0.050621	0.175621
MiscFeature	0.120317	1.000000	-0.031855	0.046880	0.072697
SaleType	0.021994	-0.031855	1.000000	0.232149	0.072896
SaleCondition	0.050621	0.046880	0.232149	1.000000	0.142503
SalePrice	0.175621	0.072697	0.072896	0.142503	1.000000

[45 rows x 45 columns]

[]:

```
[43]: #c) Identify signoficant variable(sales and the foundation) using p-values and
      ↪chi-square value
import scipy.stats as stats
from scipy.stats import chi2_contingency
contingency_table=pd.
      ↪crosstab(cat_variable['Foundation'],cat_variable['SalePrice'])
print('contingency_table :-\n',contingency_table)
# Observed Values
Observed_Values = contingency_table.values
print("Observed Values :-\n",Observed_Values)
Chi2=stats.chi2_contingency(contingency_table)
Expected_Values = Chi2[3]
print("Expected Values :-\n",Expected_Values)
no_of_rows=len(contingency_table.iloc[0:2,0])
no_of_columns=len(contingency_table.iloc[0,0:2])
ddof=(no_of_rows-1)*(no_of_columns-1)
```

```

print("Degree of Freedom:-",ddof)
alpha = 0.05
from scipy.stats import chi2
chi_square=sum([(o-e)**2./e for o,e in zip(Observed_Values,Expected_Values)])
chi_square_statistic=chi_square[0]+chi_square[1]
print("chi-square statistic:-",chi_square_statistic)
critical_value=chi2.ppf(q=1-alpha,df=ddof)
print('critical_value:',critical_value)
# P-value
p_value=1-chi2.cdf(x=chi_square_statistic,df=ddof)
print('p-value:',p_value)
print('Significance level: ',alpha)
print('Degree of Freedom: ',ddof)
print('chi-square statistic:',chi_square_statistic)
print('critical_value:',critical_value)
print('p-value:',p_value)
if chi_square_statistic>=critical_value:
    print("Reject H0,There is a relationship between 2 categorical variables")
else:
    print("Retain H0,There is no relationship between 2 categorical variables")

if p_value<=alpha:
    print("Reject H0,There is a relationship between 2 categorical variables")
else:
    print("Retain H0,There is no relationship between 2 categorical variables")

```

contingency_table :-

SalePrice	34900	35311	37900	39300	40000	52000	52500	55000	\
Foundation									
1	0	0	0	0	0	0	0	0	
2	1	1	0	0	0	1	1	0	
3	0	0	1	0	1	0	0	1	
4	0	0	0	0	0	0	0	0	
5	0	0	0	1	0	0	0	1	
6	0	0	0	0	0	0	0	0	

SalePrice	55993	58500	...	485000	501837	538000	555000	556581	\
Foundation			...						
1	0	0	...	1	1	1	1	1	
2	1	0	...	0	0	0	0	0	
3	0	1	...	0	0	0	0	0	
4	0	0	...	0	0	0	0	0	
5	0	0	...	0	0	0	0	0	
6	0	0	...	0	0	0	0	0	

SalePrice	582933	611657	625000	745000	755000				
Foundation									

```

1      1      1      1      1
2      0      0      0      0      0
3      0      0      0      0      0
4      0      0      0      0      0
5      0      0      0      0      0
6      0      0      0      0      0

```

[6 rows x 663 columns]

Observed Values :-

```

[[0 0 0 ... 1 1 1]
 [1 1 0 ... 0 0 0]
 [0 0 1 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]

```

Expected Values :-

```

[[0.44315068 0.44315068 0.44315068 ... 0.44315068 0.44315068 0.44315068]
 [0.43424658 0.43424658 0.43424658 ... 0.43424658 0.43424658 0.43424658]
 [0.1 0.1 0.1 ... 0.1 0.1 0.1 ]
 [0.00205479 0.00205479 0.00205479 ... 0.00205479 0.00205479 0.00205479]
 [0.01643836 0.01643836 0.01643836 ... 0.01643836 0.01643836 0.01643836]
 [0.00410959 0.00410959 0.00410959 ... 0.00410959 0.00410959 0.00410959]]

```

Degree of Freedom:- 1

chi-square statistic:- 2.6056782334384856

critical_value: 3.841458820694124

p-value: 0.10648159423064485

Significance level: 0.05

Degree of Freedom: 1

chi-square statistic: 2.6056782334384856

critical_value: 3.841458820694124

p-value: 0.10648159423064485

Retain H0, There is no relationship between 2 categorical variables

Retain H0, There is no relationship between 2 categorical variables

[44]: cat_variable.dtypes

```

[44]: Id          int64
MSZoning     int64
Street        int64
Alley         int64
LotShape       int64
LandContour    int64
Utilities      int64
LotConfig       int64
LandSlope       int64
Neighborhood   int64
Condition1     int64

```

```

Condition2          int64
BldgType           int64
HouseStyle          int64
RoofStyle           int64
RoofMatl            int64
Exterior1st         int64
Exterior2nd         int64
MasVnrType          int64
ExterQual            int64
ExterCond            int64
Foundation          int64
BsmtQual             int64
BsmtCond             int64
BsmtExposure         int64
BsmtFinType1         int64
BsmtFinType2         int64
Heating              int64
HeatingQC             int64
CentralAir            int64
Electrical            int64
KitchenQual           int64
Functiol              int64
FireplaceQu           int64
GarageType             int64
GarageFinish           int64
GarageQual             int64
GarageCond             int64
PavedDrive            int64
PoolQC                int64
Fence                  int64
MiscFeature            int64
SaleType                int64
SaleCondition           int64
SalePrice                int64
dtype: object

```

[45]: cat_variable.head()

```

[45]:   Id  MSZoning  Street  Alley  LotShape  LandContour  Utilities  LotConfig  \
0    1        1       1      3        1            1           1           1           1
1    2        1       1      3        1            1           1           1           2
2    3        1       1      3        2            1           1           1           1
3    4        1       1      3        2            1           1           1           3
4    5        1       1      3        2            1           1           1           2

      LandSlope  Neighborhood  ...  GarageFinish  GarageQual  GarageCond  \
0            1           ...           1           1           1           1

```

```

1      1      2 ...      1      1      1
2      1      1 ...      1      1      1
3      1      3 ...      2      1      1
4      1      4 ...      1      1      1

   PavedDrive PoolQC Fence MiscFeature SaleType SaleCondition SalePrice
0            1      4      5          5        1            1    208500
1            1      4      5          5        1            1    181500
2            1      4      5          5        1            1    223500
3            1      4      5          5        1            2    140000
4            1      4      5          5        1            1    250000

```

[5 rows x 45 columns]

[46]: *#Q5. Cobime all the significant categorical and numerical variable*

```
Pep1_merge = pd.merge(num_variable, cat_variable, how='outer', on = 'Id')
Pep1_merge.head()
```

decode the cat_variable

```

[46]:   Id MSSubClass LotFrontage LotArea OverallQual OverallCond YearBuilt \
0   1       60        65     8450        7           5      2003
1   2       20        80     9600        6           8      1976
2   3       60        68    11250        7           5      2001
3   4       70        60     9550        7           5      1915
4   5       60        84    14260        8           5      2000

   YearRemodAdd MasVnrArea BsmtFinSF1 ... GarageFinish GarageQual \
0       2003        196        706 ...          1           1
1       1976         0        978 ...          1           1
2       2002        162        486 ...          1           1
3       1970         0        216 ...          2           1
4       2000        350        655 ...          1           1

   GarageCond PavedDrive PoolQC Fence MiscFeature SaleType \
0           1         1      4      5          5        1
1           1         1      4      5          5        1
2           1         1      4      5          5        1
3           1         1      4      5          5        1
4           1         1      4      5          5        1

   SaleCondition SalePrice_y
0             1    208500
1             1    181500
2             1    223500
3             2    140000
4             1    250000

```

[5 rows x 82 columns]

[47]: Pep1_merge.shape

[47]: (1460, 82)

[48]: Pep1_merge.corr()

```
[48]:          Id  MSSubClass  LotFrontage  LotArea  OverallQual \
Id  1.000000  0.011156 -0.009616 -0.033226 -0.028365
MSSubClass  0.011156  1.000000 -0.357042 -0.139781  0.032628
LotFrontage -0.009616 -0.357042  1.000000  0.306689  0.234227
LotArea    -0.033226 -0.139781  0.306689  1.000000  0.105806
OverallQual -0.028365  0.032628  0.234227  0.105806  1.000000
OverallCond  0.012609 -0.059316 -0.052842 -0.005636 -0.091932
YearBuilt   -0.012713  0.027850  0.117555  0.014228  0.572323
YearRemodAdd -0.021998  0.040581  0.082775  0.013788  0.550684
MasVnrArea  -0.050205  0.022899  0.179270  0.103957  0.410220
BsmtFinSF1  -0.005024 -0.069836  0.215760  0.214103  0.239666
BsmtFinSF2  -0.005968 -0.065649  0.043298  0.111170 -0.059119
BsmtUnfSF   -0.007940 -0.140759  0.122249 -0.002618  0.308159
TotalBsmtSF -0.015415 -0.238518  0.363366  0.260833  0.537808
1stFlrSF    0.010496 -0.251758  0.414246  0.299475  0.476224
2ndFlrSF    0.005590  0.307886  0.072479  0.050986  0.295493
LowQualFinSF -0.044230  0.046474  0.036879  0.004779 -0.030429
GrLivArea   0.008273  0.074853  0.368376  0.263116  0.593007
BsmtFullBath 0.002289  0.003491  0.091428  0.158155  0.111098
BsmtHalfBath -0.020155 -0.002333 -0.006446  0.048046 -0.040150
FullBath     0.005587  0.131608  0.180431  0.126031  0.550600
HalfBath     0.006784  0.177354  0.048209  0.014259  0.273458
BedroomAbvGr 0.037719 -0.023438  0.237016  0.119690  0.101676
KitchebvGr   0.002951  0.281721 -0.005762 -0.017784 -0.183882
TotRmsAbvGrd 0.027239  0.040380  0.320166  0.190015  0.427452
Fireplaces   -0.019772 -0.045569  0.235636  0.271364  0.396765
GarageYrBlt  0.000136  0.079774  0.064794 -0.024454  0.519278
GarageCars   0.016570 -0.040110  0.269722  0.154871  0.600671
GarageArea   0.017634 -0.098672  0.323658  0.180403  0.562022
WoodDeckSF   -0.029643 -0.012579  0.077032  0.171698  0.238923
OpenPorchSF  -0.000477 -0.006100  0.137435  0.084774  0.308819
EnclosedPorch 0.002889 -0.012037  0.009814 -0.018340 -0.113937
3SsnPorch   -0.046635 -0.043825  0.062317  0.020423  0.030371
ScreenPorch  0.001330 -0.026030  0.037683  0.043160  0.064886
PoolArea    0.057044  0.008283  0.180867  0.077672  0.065166
MiscVal     -0.006242 -0.007683  0.001101  0.038068 -0.031406
MoSold      0.021172 -0.013585  0.010172  0.001205  0.070815
YrSold      0.000712 -0.021407  0.006750 -0.014261 -0.027347
```

	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	\
SalePrice_x	-0.021917	-0.084284	0.334897	0.263843	0.790982	
MSZoning	-0.011949	0.286937	-0.230329	-0.161640	-0.015122	
Street	-0.008916	0.024969	0.037324	0.197131	-0.058823	
Alley	-0.001530	-0.105995	0.133366	0.060105	0.099179	
LotShape	-0.024071	-0.115361	0.196810	0.315484	0.198994	
LandContour	0.014769	-0.018947	0.087751	0.258792	0.039133	
Utilities	0.013324	-0.022844	-0.000049	0.010123	-0.001881	
LotConfig	-0.038192	-0.069160	0.124568	0.160976	0.060576	
LandSlope	0.005847	-0.025672	0.067513	0.436868	-0.066450	
Neighborhood	-0.013078	0.181163	-0.111131	0.043191	-0.138913	
Condition1	-0.010735	-0.034756	0.074363	0.033795	-0.019525	
Condition2	-0.016961	0.023822	0.032015	0.030878	0.019912	
BldgType	0.019230	0.770840	-0.435955	-0.211111	0.022988	
HouseStyle	0.026166	0.080782	-0.033021	-0.008187	-0.178803	
RoofStyle	0.015873	-0.099083	0.159235	0.137369	0.092467	
RoofMatl	0.013375	-0.024630	0.199497	0.162496	0.033583	
Exterior1st	0.003624	0.061310	-0.014166	0.085558	-0.169506	
Exterior2nd	0.015449	0.045042	-0.015674	0.061748	-0.222293	
MasVnrType	0.030431	-0.067842	-0.002419	0.066471	-0.012461	
ExterQual	-0.021844	-0.039352	-0.013603	0.014713	-0.381530	
ExterCond	-0.037865	0.011163	-0.051035	-0.018784	-0.139091	
Foundation	0.020023	-0.019650	-0.086625	-0.006544	-0.478417	
BsmtQual	-0.018676	-0.102955	0.047313	-0.016246	-0.206902	
BsmtCond	-0.011355	0.018474	-0.055489	-0.025991	-0.227490	
BsmtExposure	0.027692	0.063997	0.071626	0.056750	0.080070	
BsmtFinType1	0.009911	-0.087532	-0.070887	-0.023228	-0.422580	
BsmtFinType2	0.002847	-0.021939	-0.008727	0.039689	-0.207340	
Heating	0.061902	0.031321	-0.019042	0.022623	-0.119230	
HeatingQC	0.015001	0.019458	-0.089009	-0.003581	-0.457083	
CentralAir	-0.009821	0.101774	-0.068603	-0.049755	-0.272038	
Electrical	-0.037032	-0.040325	-0.060872	-0.042676	-0.242794	
KitchenQual	-0.020927	-0.019633	0.028891	0.007505	-0.219938	
Functiol	0.006018	-0.012157	-0.012555	0.034551	-0.157566	
FireplaceQu	-0.000390	0.021743	-0.206344	-0.180300	-0.417722	
GarageType	-0.007207	0.148495	-0.121935	-0.083231	-0.313078	
GarageFinish	-0.009336	0.074824	-0.052183	0.003996	-0.057931	
GarageQual	-0.017357	0.091627	-0.105460	-0.073981	-0.266223	
GarageCond	-0.012836	0.084181	-0.103644	-0.078780	-0.285116	
PavedDrive	0.008921	0.025973	-0.067845	-0.002964	-0.200702	
PoolQC	-0.033928	-0.005221	-0.191645	-0.065167	-0.079987	
Fence	0.012682	0.104436	-0.024406	0.049814	0.200623	
MiscFeature	0.044733	0.048430	-0.007206	-0.110001	0.095170	
SaleType	-0.016932	0.012874	0.006185	-0.013882	0.063406	
SaleCondition	-0.017712	-0.015047	0.069015	0.000199	0.131870	
SalePrice_y	-0.021917	-0.084284	0.334897	0.263843	0.790982	

Id	0.012609	-0.012713	-0.021998	-0.050205	-0.005024
MSSubClass	-0.059316	0.027850	0.040581	0.022899	-0.069836
LotFrontage	-0.052842	0.117555	0.082775	0.179270	0.215760
LotArea	-0.005636	0.014228	0.013788	0.103957	0.214103
OverallQual	-0.091932	0.572323	0.550684	0.410220	0.239666
OverallCond	1.000000	-0.375983	0.073741	-0.127775	-0.046231
YearBuilt	-0.375983	1.000000	0.592855	0.314726	0.249503
YearRemodAdd	0.073741	0.592855	1.000000	0.179170	0.128451
MasVnrArea	-0.127775	0.314726	0.179170	1.000000	0.263569
BsmtFinSF1	-0.046231	0.249503	0.128451	0.263569	1.000000
BsmtFinSF2	0.040229	-0.049107	-0.067759	-0.072296	-0.050117
BsmtUnfSF	-0.136841	0.149040	0.181133	0.114182	-0.495251
TotalBsmtSF	-0.171098	0.391452	0.291066	0.362438	0.522396
1stFlrSF	-0.144203	0.281986	0.240379	0.342147	0.445863
2ndFlrSF	0.028942	0.010308	0.140024	0.174019	-0.137079
LowQualFinSF	0.025494	-0.183784	-0.062419	-0.069066	-0.064503
GrLivArea	-0.079686	0.199010	0.287389	0.389883	0.208171
BsmtFullBath	-0.054942	0.187599	0.119470	0.085042	0.649212
BsmtHalfBath	0.117821	-0.038162	-0.012337	0.026674	0.067418
FullBath	-0.194149	0.468271	0.439046	0.275713	0.058543
HalfBath	-0.060769	0.242656	0.183331	0.200792	0.004262
BedroomAbvGr	0.012980	-0.070651	-0.040581	0.102420	-0.107355
KitchebvGr	-0.087001	-0.174800	-0.149598	-0.037371	-0.081007
TotRmsAbvGrd	-0.057583	0.095589	0.191740	0.280025	0.044316
Fireplaces	-0.023820	0.147716	0.112581	0.247902	0.260011
GarageYrBlt	-0.306118	0.781662	0.618672	0.249927	0.150858
GarageCars	-0.185758	0.537850	0.420622	0.363768	0.224054
GarageArea	-0.151521	0.478954	0.371600	0.372558	0.296970
WoodDeckSF	-0.003334	0.224880	0.205726	0.159355	0.204306
OpenPorchSF	-0.032589	0.188686	0.226298	0.124950	0.111761
EnclosedPorch	0.070356	-0.387268	-0.193919	-0.109850	-0.102303
3SsnPorch	0.025504	0.031355	0.045286	0.018797	0.026451
ScreenPorch	0.054811	-0.050364	-0.038740	0.061459	0.062021
PoolArea	-0.001985	0.004950	0.005829	0.011724	0.140491
MiscVal	0.068777	-0.034383	-0.010286	-0.029813	0.003571
MoSold	-0.003511	0.012398	0.021490	-0.005945	-0.015727
YrSold	0.043950	-0.013618	0.035743	-0.008185	0.014359
SalePrice_x	-0.077856	0.522897	0.507101	0.475227	0.386420
MSZoning	-0.026962	-0.062684	0.033050	-0.040262	-0.150204
Street	-0.042848	-0.021137	-0.065465	-0.017325	0.015643
Alley	-0.098099	0.303155	0.073085	0.071527	0.146783
LotShape	-0.033747	0.229365	0.175488	0.089030	0.157718
LandContour	0.000986	-0.029314	-0.007847	-0.019583	0.078115
Utilities	0.009994	-0.011505	-0.034080	0.063427	-0.019100
LotConfig	0.031625	0.033011	0.042451	0.044738	0.056046
LandSlope	0.010355	-0.073639	-0.059105	-0.021808	0.113834
Neighborhood	-0.059568	-0.115600	-0.155964	-0.079836	-0.057656

Condition1	0.073287	-0.106443	-0.048202	-0.043368	-0.028312
Condition2	0.046472	-0.074740	-0.004373	0.021494	-0.005867
BldgType	-0.150158	0.197934	0.082713	0.048657	-0.007100
HouseStyle	0.132787	-0.251479	-0.193432	-0.121885	0.009929
RoofStyle	0.032252	-0.017865	-0.003283	0.130309	0.129374
RoofMatl	0.012947	-0.012635	-0.028809	0.047498	0.120056
Exterior1st	0.061448	-0.255993	-0.295733	-0.055326	0.093651
Exterior2nd	0.083044	-0.396766	-0.368477	-0.116399	0.032897
MasVnrType	-0.016224	-0.056501	0.050918	-0.268316	0.004900
ExterQual	0.097238	-0.438761	-0.417886	-0.027512	-0.007096
ExterCond	0.183496	-0.270435	-0.081738	-0.074799	-0.058488
Foundation	0.227912	-0.647865	-0.477886	-0.211172	-0.201189
BsmtQual	0.031173	-0.341687	-0.265016	0.022973	-0.066487
BsmtCond	-0.094425	-0.170332	-0.144478	-0.091448	-0.178049
BsmtExposure	-0.090596	0.198262	0.105582	0.041229	0.092050
BsmtFinType1	0.056212	-0.438540	-0.416797	-0.181172	-0.353186
BsmtFinType2	-0.008319	-0.089341	-0.164374	-0.084720	-0.097782
Heating	-0.072006	-0.180570	-0.162898	-0.056406	-0.072048
HeatingQC	0.014105	-0.448855	-0.550017	-0.159243	-0.085672
CentralAir	-0.118969	-0.381831	-0.298878	-0.127347	-0.166468
Electrical	-0.063316	-0.317138	-0.312435	-0.114856	-0.152332
KitchenQual	-0.038593	-0.328549	-0.379318	-0.009715	0.019621
Functiol	-0.095312	-0.169388	-0.082170	-0.082454	-0.058127
FireplaceQu	0.054827	-0.224994	-0.160649	-0.246107	-0.169235
GarageType	0.003701	-0.323697	-0.170951	-0.138710	-0.186534
GarageFinish	-0.023443	-0.092524	-0.001445	-0.027914	0.012877
GarageQual	0.017318	-0.295760	-0.143198	-0.142980	-0.133160
GarageCond	-0.001667	-0.285189	-0.153350	-0.134726	-0.128748
PavedDrive	0.082132	-0.396023	-0.193616	-0.141731	-0.175379
PoolQC	0.023986	0.004755	-0.014097	-0.000381	-0.135647
Fence	-0.155022	0.221814	0.168992	0.108640	0.018913
MiscFeature	-0.075653	0.077926	0.083103	0.051024	0.010375
SaleType	-0.111267	0.077768	0.031197	0.030285	0.013870
SaleCondition	-0.144985	0.149704	0.117656	0.079047	0.027785
SalePrice_y	-0.077856	0.522897	0.507101	0.475227	0.386420

...	GarageFinish	GarageQual	GarageCond	PavedDrive	\
Id	...	-0.009336	-0.017357	-0.012836	0.008921
MSSubClass	...	0.074824	0.091627	0.084181	0.025973
LotFrontage	...	-0.052183	-0.105460	-0.103644	-0.067845
LotArea	...	0.003996	-0.073981	-0.078780	-0.002964
OverallQual	...	-0.057931	-0.266223	-0.285116	-0.200702
OverallCond	...	-0.023443	0.017318	-0.001667	0.082132
YearBuilt	...	-0.092524	-0.295760	-0.285189	-0.396023
YearRemodAdd	...	-0.001445	-0.143198	-0.153350	-0.193616
MasVnrArea	...	-0.027914	-0.142980	-0.134726	-0.141731
BsmtFinSF1	...	0.012877	-0.133160	-0.128748	-0.175379

BsmtFinSF2	..	-0.038860	-0.047197	-0.040690	-0.067842
BsmtUnfSF	..	-0.045533	-0.017723	-0.030091	0.021160
TotalBsmtSF	..	-0.046763	-0.173645	-0.179122	-0.185966
1stFlrSF	..	-0.031200	-0.166320	-0.173799	-0.147598
2ndFlrSF	..	0.039600	-0.044184	-0.055085	0.034872
LowQualFinSF	..	0.092300	0.156270	0.148413	0.076245
GrLivArea	..	0.018484	-0.144605	-0.159889	-0.072562
BsmtFullBath	..	0.004379	-0.065509	-0.056168	-0.115018
BsmtHalfBath	..	-0.000973	-0.031341	-0.029262	-0.012814
FullBath	..	-0.031615	-0.125115	-0.136307	-0.138856
HalfBath	..	-0.011381	-0.113432	-0.125367	-0.087447
BedroomAbvGr	..	0.005135	0.000034	0.002882	0.032287
KitchebvGr	..	0.085262	0.172988	0.172314	0.110847
TotRmsAbvGrd	..	0.051838	-0.081101	-0.085890	-0.008858
Fireplaces	..	-0.002946	-0.182669	-0.199752	-0.084286
GarageYrBlt	..	0.049687	-0.070154	-0.067341	-0.247374
GarageCars	..	-0.244416	-0.577586	-0.579929	-0.253820
GarageArea	..	-0.276140	-0.534628	-0.539459	-0.242255
WoodDeckSF	..	-0.014047	-0.113309	-0.114698	-0.102714
OpenPorchSF	..	-0.009777	-0.030995	-0.044590	-0.034247
EnclosedPorch	..	-0.001902	0.082142	0.082288	0.183532
3SsnPorch	..	-0.070948	-0.019647	-0.030146	-0.023778
ScreenPorch	..	-0.014941	-0.053480	-0.073337	-0.037535
PoolArea	..	0.003750	0.001316	0.001916	-0.019331
MiscVal	..	-0.046945	0.004295	0.003774	0.026407
MoSold	..	-0.014245	-0.020755	-0.013496	-0.009224
YrSold	..	0.006532	-0.006783	0.001632	-0.016057
SalePrice_x	..	-0.006799	-0.245912	-0.261527	-0.208954
MSZoning	..	0.039377	0.075028	0.071517	0.117402
Street	..	0.007698	0.026658	0.027236	0.011248
Alley	..	-0.030331	-0.094932	-0.098604	-0.235390
LotShape	..	0.012848	-0.093695	-0.098629	-0.100666
LandContour	..	0.059278	0.014482	0.017144	0.081031
Utilities	..	-0.032176	-0.007631	-0.007409	-0.007368
LotConfig	..	-0.000694	-0.045119	-0.037126	-0.065917
LandSlope	..	0.020837	0.005701	0.007672	0.018009
Neighborhood	..	0.196493	0.081193	0.092220	0.065998
Condition1	..	0.012920	-0.015162	-0.016253	0.034886
Condition2	..	0.008149	0.019882	0.003607	0.054484
BldgType	..	0.043835	0.040162	0.038377	-0.064447
HouseStyle	..	0.013577	0.092276	0.089343	0.038707
RoofStyle	..	0.045779	-0.028896	-0.012751	0.011743
RoofMatl	..	0.023509	0.005744	0.015417	-0.033947
Exterior1st	..	0.044642	0.072491	0.074176	0.058754
Exterior2nd	..	0.050297	0.100327	0.094935	0.119255
MasVnrType	..	0.053773	0.044496	0.044785	0.069944
ExterQual	..	0.151209	0.189273	0.195280	0.166263

ExterCond	...	0.085163	0.178642	0.168506	0.139432
Foundation	...	0.024916	0.182532	0.170701	0.243846
BsmtQual	...	0.123399	0.119180	0.122385	0.168455
BsmtCond	...	0.079177	0.121402	0.143749	0.148613
BsmtExposure	...	0.009487	-0.020492	-0.012015	-0.015039
BsmtFinType1	...	0.008407	0.092989	0.096412	0.141792
BsmtFinType2	...	-0.000614	0.027401	0.039962	0.009026
Heating	...	0.057977	0.118384	0.130970	0.141864
HeatingQC	...	0.020111	0.139173	0.148856	0.193215
CentralAir	...	0.147735	0.288008	0.295220	0.275660
Electrical	...	0.065836	0.151144	0.156999	0.168861
KitchenQual	...	0.225090	0.228018	0.233780	0.164511
Functiol	...	0.036582	0.065887	0.056685	0.047388
FireplaceQu	...	0.022276	0.198937	0.214806	0.108336
GarageType	...	0.505294	0.824347	0.823027	0.242775
GarageFinish	...	1.000000	0.518531	0.521545	0.133507
GarageQual	...	0.518531	1.000000	0.968264	0.255161
GarageCond	...	0.521545	0.968264	1.000000	0.268934
PavedDrive	...	0.133507	0.255161	0.268934	1.000000
PoolQC	...	-0.001018	0.010196	0.009652	0.017815
Fence	...	0.056200	-0.020630	-0.035884	-0.034126
MiscFeature	...	0.034681	-0.005417	-0.003933	0.000685
SaleType	...	0.026965	0.041921	0.048109	-0.021553
SaleCondition	...	0.051570	0.006738	0.004943	-0.049674
SalePrice_y	...	-0.006799	-0.245912	-0.261527	-0.208954

	PoolQC	Fence	MiscFeature	SaleType	SaleCondition	\
Id	-0.033928	0.012682	0.044733	-0.016932	-0.017712	
MSSubClass	-0.005221	0.104436	0.048430	0.012874	-0.015047	
LotFrontage	-0.191645	-0.024406	-0.007206	0.006185	0.069015	
LotArea	-0.065167	0.049814	-0.110001	-0.013882	0.000199	
OverallQual	-0.079987	0.200623	0.095170	0.063406	0.131870	
OverallCond	0.023986	-0.155022	-0.075653	-0.111267	-0.144985	
YearBuilt	0.004755	0.221814	0.077926	0.077768	0.149704	
YearRemodAdd	-0.014097	0.168992	0.083103	0.031197	0.117656	
MasVnrArea	-0.000381	0.108640	0.051024	0.030285	0.079047	
BsmtFinSF1	-0.135647	0.018913	0.010375	0.013870	0.027785	
BsmtFinSF2	-0.049488	-0.115984	0.022462	0.013232	-0.058805	
BsmtUnfSF	0.044661	0.135130	0.049216	0.062772	0.134302	
TotalBsmtSF	-0.114240	0.113118	0.068617	0.082509	0.142533	
1stFlrSF	-0.113228	0.078714	0.048006	0.083203	0.166194	
2ndFlrSF	-0.112976	0.079138	0.022083	-0.031918	-0.034329	
LowQualFinSF	-0.110205	-0.043218	-0.025386	-0.008159	-0.017711	
GrLivArea	-0.187349	0.119651	0.051313	0.033941	0.092110	
BsmtFullBath	-0.070130	-0.025510	0.005497	-0.004749	-0.012422	
BsmtHalfBath	-0.045949	-0.017256	-0.031731	0.000948	0.021340	
FullBath	-0.067679	0.211935	0.064537	0.043578	0.123659	

HalfBath	-0.029279	0.086564	0.037717	-0.029352	-0.010208
BedroomAbvGr	-0.076052	-0.027185	-0.001891	-0.020038	0.017270
KitchebvGr	0.013386	0.052468	-0.014275	0.027160	0.102303
TotRmsAbvGrd	-0.087717	0.080433	0.032370	0.040563	0.118449
Fireplaces	-0.090908	0.042750	0.005656	-0.001670	0.004490
GarageYrBlt	0.025014	0.228096	0.052995	0.073742	0.148013
GarageCars	-0.032770	0.183841	0.056129	0.019239	0.132221
GarageArea	-0.067879	0.148804	0.036404	0.037023	0.127422
WoodDeckSF	-0.031901	-0.021168	-0.035864	-0.023842	-0.017134
OpenPorchSF	-0.031424	0.070687	0.005352	0.011643	0.064042
EnclosedPorch	-0.109238	-0.062464	-0.035399	-0.009259	-0.048306
3SsnPorch	0.007365	-0.005648	-0.018341	-0.002600	-0.001324
ScreenPorch	-0.059776	-0.009775	-0.025970	-0.029731	-0.028190
PoolArea	-0.884250	-0.084654	0.001214	-0.009887	0.068578
MiscVal	-0.033724	-0.028760	-0.392918	-0.014143	-0.032655
MoSold	0.036368	0.002408	-0.011902	0.040201	0.054679
YrSold	0.060904	-0.051277	-0.059325	-0.021683	-0.123720
SalePrice_x	-0.126070	0.175621	0.072697	0.072896	0.142503
MSZoning	0.027003	0.060229	0.033389	0.058948	0.068705
Street	0.004067	0.029920	-0.163140	0.021171	0.022919
Alley	-0.015492	0.015728	0.028485	-0.004045	-0.016915
LotShape	-0.030895	0.078494	-0.010451	0.001435	-0.004773
LandContour	0.011667	0.070593	-0.005240	-0.014537	0.062826
Utilities	0.001657	0.012194	0.005073	0.057455	0.019173
LotConfig	-0.044571	-0.015429	-0.012520	0.012120	-0.013638
LandSlope	0.014289	0.035386	-0.051181	-0.056680	0.017187
Neighborhood	0.019372	0.031296	-0.008989	0.008466	0.007101
Condition1	-0.068573	-0.031674	-0.001242	0.003128	-0.019800
Condition2	0.005796	-0.016973	-0.064842	0.011960	0.005158
BldgType	0.026355	0.145490	0.056123	0.044289	0.040365
HouseStyle	0.010586	-0.163869	-0.021265	-0.006154	-0.029049
RoofStyle	-0.051893	0.000924	-0.013105	-0.031824	0.032990
RoofMatl	-0.100694	0.025178	0.023374	-0.014712	0.055436
Exterior1st	-0.092011	-0.140353	0.004149	-0.015554	0.000631
Exterior2nd	-0.098952	-0.115956	-0.010142	-0.033452	-0.039999
MasVnrType	0.004776	0.016956	-0.022719	0.070788	0.124239
ExterQual	0.020062	-0.171845	-0.090986	0.022013	-0.009670
ExterCond	-0.022881	-0.085694	-0.017514	0.003909	-0.038491
Foundation	0.020203	-0.153194	-0.115557	-0.050491	-0.093201
BsmtQual	-0.020983	-0.038887	-0.036810	0.013701	0.054583
BsmtCond	0.018219	0.035492	-0.015516	0.016683	0.010336
BsmtExposure	-0.021710	0.062164	0.006711	0.028535	0.092899
BsmtFinType1	0.032206	-0.129792	-0.023372	-0.023989	-0.055771
BsmtFinType2	-0.077483	-0.068567	-0.032791	-0.021281	-0.035900
Heating	0.008075	0.006423	0.012483	0.018388	-0.032650
HeatingQC	-0.029891	-0.204193	-0.042021	-0.009051	-0.049721
CentralAir	0.016701	0.005749	-0.002015	0.026522	-0.010168

Electrical	0.018534	-0.004963	0.029254	-0.016778	-0.024845
KitchenQual	0.002532	-0.071642	-0.054149	-0.000178	0.033660
Functiol	0.015240	-0.034905	-0.078691	-0.014494	-0.055108
FireplaceQu	0.037704	-0.080348	-0.017466	0.021277	0.010237
GarageType	0.024623	-0.016003	-0.016616	0.051111	0.031789
GarageFinish	-0.001018	0.056200	0.034681	0.026965	0.051570
GarageQual	0.010196	-0.020630	-0.005417	0.041921	0.006738
GarageCond	0.009652	-0.035884	-0.003933	0.048109	0.004943
PavedDrive	0.017815	-0.034126	0.000685	-0.021553	-0.049674
PoolQC	1.000000	0.093026	0.001044	0.012437	-0.040874
Fence	0.093026	1.000000	0.120317	0.021994	0.050621
MiscFeature	0.001044	0.120317	1.000000	-0.031855	0.046880
SaleType	0.012437	0.021994	-0.031855	1.000000	0.232149
SaleCondition	-0.040874	0.050621	0.046880	0.232149	1.000000
SalePrice_y	-0.126070	0.175621	0.072697	0.072896	0.142503

	SalePrice_y				
Id		-0.021917			
MSSubClass		-0.084284			
LotFrontage		0.334897			
LotArea		0.263843			
OverallQual		0.790982			
OverallCond		-0.077856			
YearBuilt		0.522897			
YearRemodAdd		0.507101			
MasVnrArea		0.475227			
BsmtFinSF1		0.386420			
BsmtFinSF2		-0.011378			
BsmtUnfSF		0.214479			
TotalBsmtSF		0.613581			
1stFlrSF		0.605852			
2ndFlrSF		0.319334			
LowQualFinSF		-0.025606			
GrLivArea		0.708624			
BsmtFullBath		0.227122			
BsmtHalfBath		-0.016844			
FullBath		0.560664			
HalfBath		0.284108			
BedroomAbvGr		0.168213			
KitchebvGr		-0.135907			
TotRmsAbvGrd		0.533723			
Fireplaces		0.466929			
GarageYrBlt		0.471316			
GarageCars		0.640409			
GarageArea		0.623431			
WoodDeckSF		0.324413			
OpenPorchSF		0.315856			

EnclosedPorch	-0.128578
3SsnPorch	0.044584
ScreenPorch	0.111447
PoolArea	0.092404
MiscVal	-0.021190
MoSold	0.046432
YrSold	-0.028923
SalePrice_x	1.000000
MSZoning	-0.116047
Street	-0.041036
Alley	0.139868
LotShape	0.267759
LandContour	0.092009
Utilities	-0.014314
LotConfig	0.109106
LandSlope	0.051152
Neighborhood	-0.143621
Condition1	-0.044820
Condition2	-0.004833
BldgType	-0.112611
HouseStyle	-0.188688
RoofStyle	0.159332
RoofMatl	0.035820
Exterior1st	-0.120586
Exterior2nd	-0.164716
MasVnrType	0.007123
ExterQual	-0.265015
ExterCond	-0.121706
Foundation	-0.429678
BsmtQual	-0.084708
BsmtCond	-0.176352
BsmtExposure	0.132027
BsmtFinType1	-0.387251
BsmtFinType2	-0.137066
Heating	-0.106673
HeatingQC	-0.427649
CentralAir	-0.251328
Electrical	-0.226145
KitchenQual	-0.114746
Functiol	-0.108367
FireplaceQu	-0.442834
GarageType	-0.273539
GarageFinish	-0.006799
GarageQual	-0.245912
GarageCond	-0.261527
PavedDrive	-0.208954
PoolQC	-0.126070

```
Fence          0.175621
MiscFeature    0.072697
SaleType        0.072896
SaleCondition   0.142503
SalePrice_y     1.000000
```

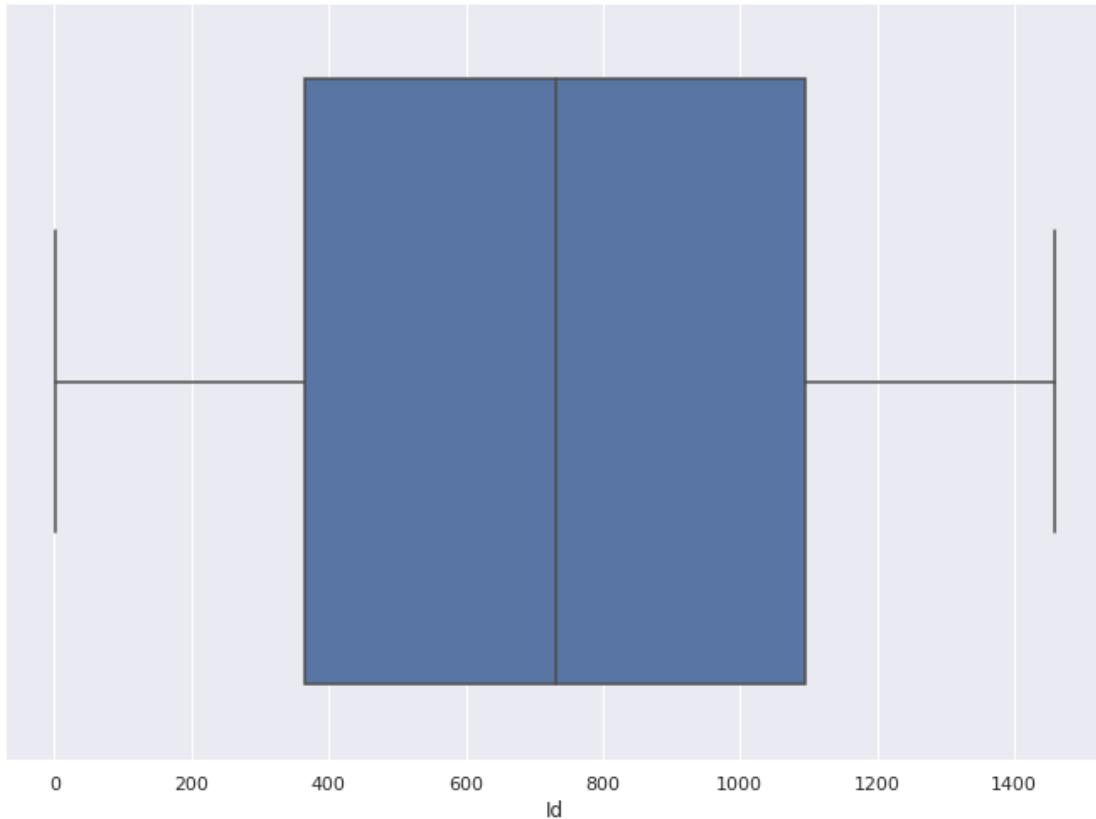
[82 rows x 82 columns]

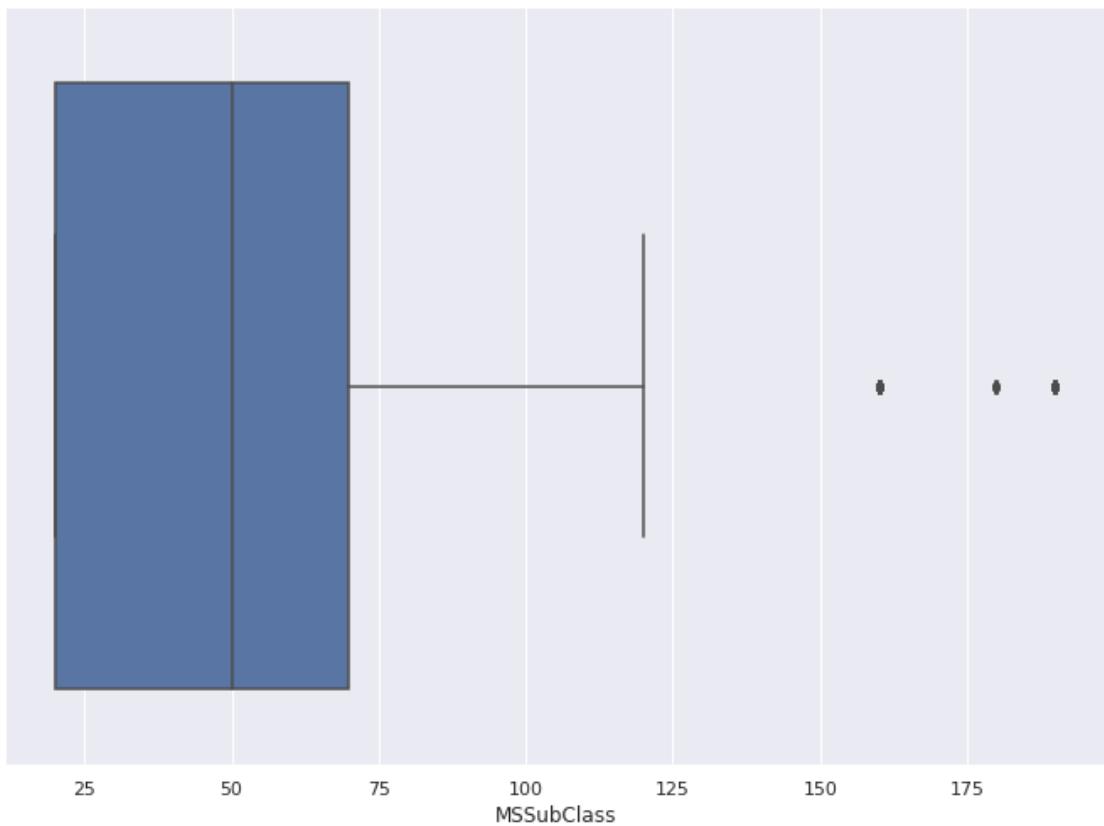
```
[52]: #Q6.Plot box plot for the new dataset to find the variable with outliers
```

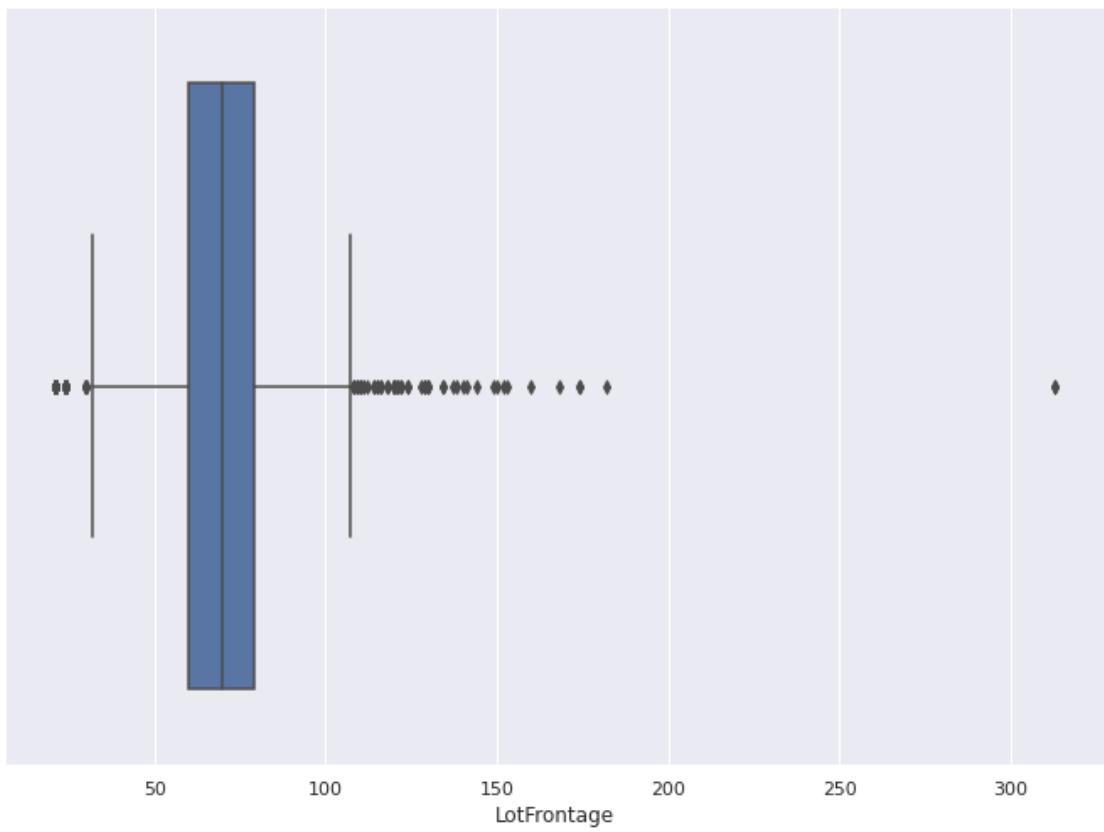
```
def boxplotloop(Pep1, columns):
    for col in columns:
        if Pep1[col].dtype != object:
            sns.set(rc={'figure.figsize':(11.7,8.27)})
            sns.boxplot(Pep1[col])
            plt.show()
```

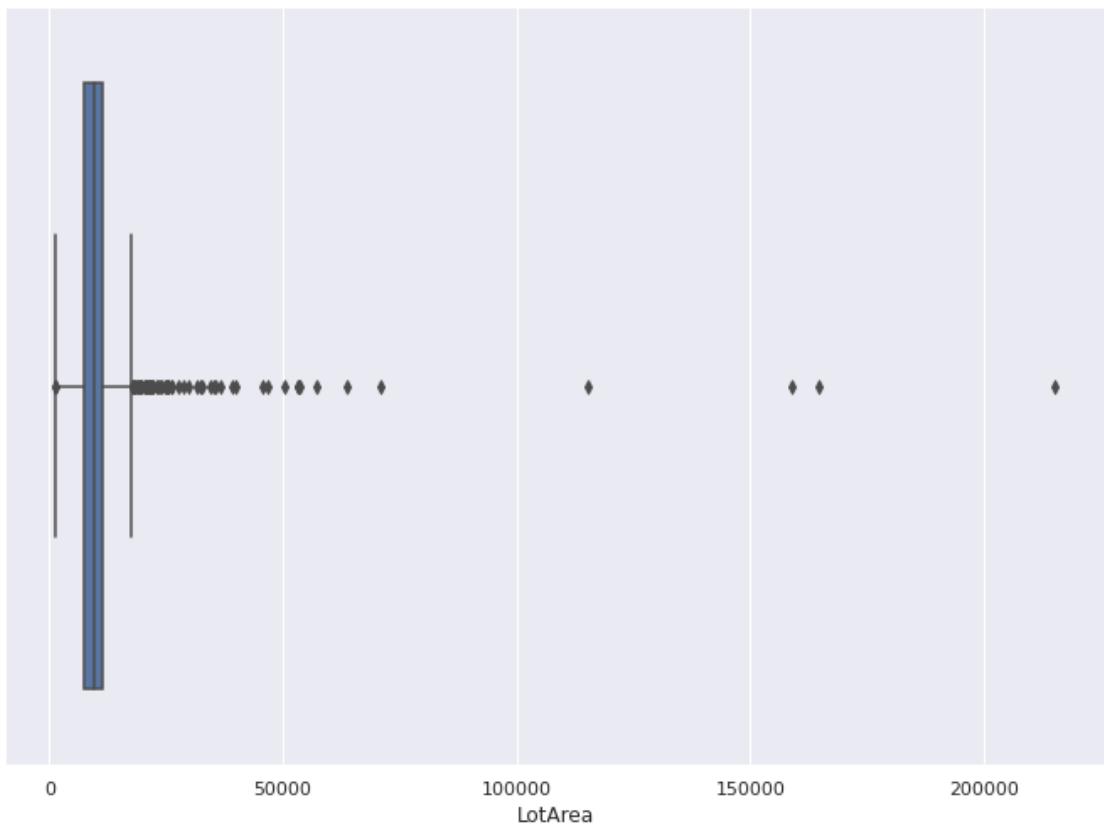
```
[55]: Pep1_final=Pep1_merge
```

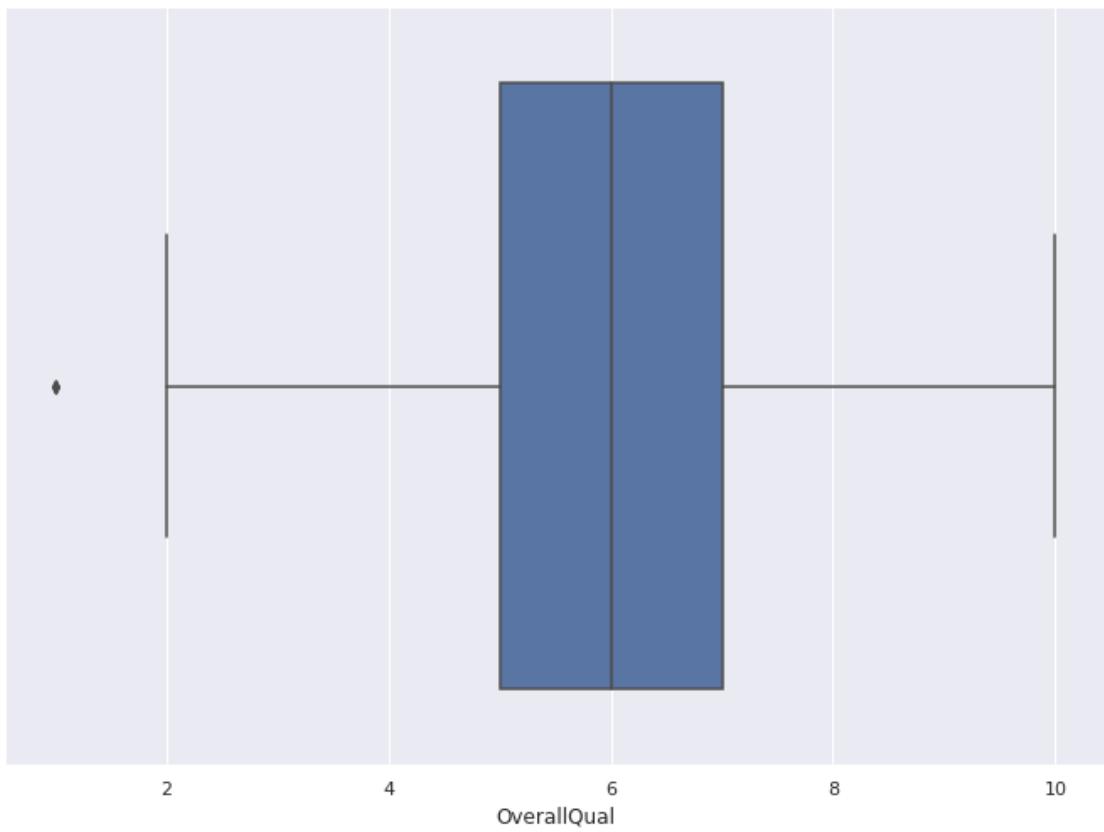
```
[56]: boxplotloop(Pep1_final, Pep1_final.columns)
```

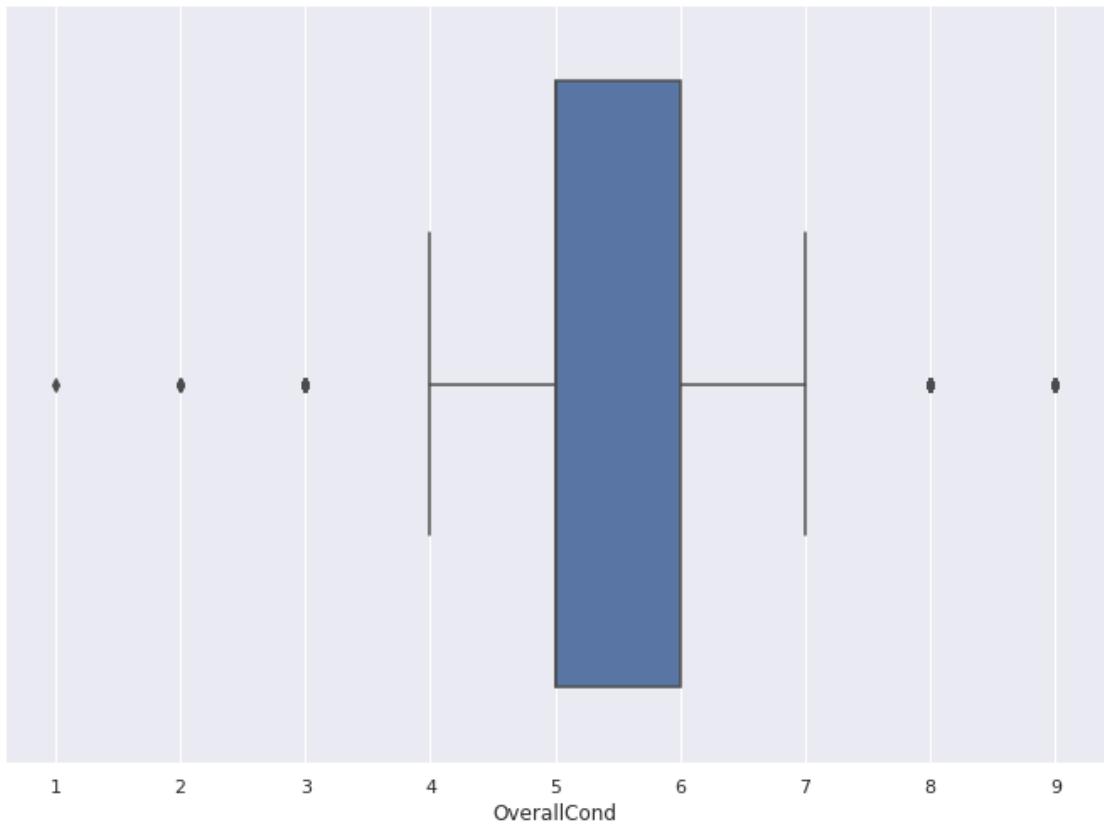


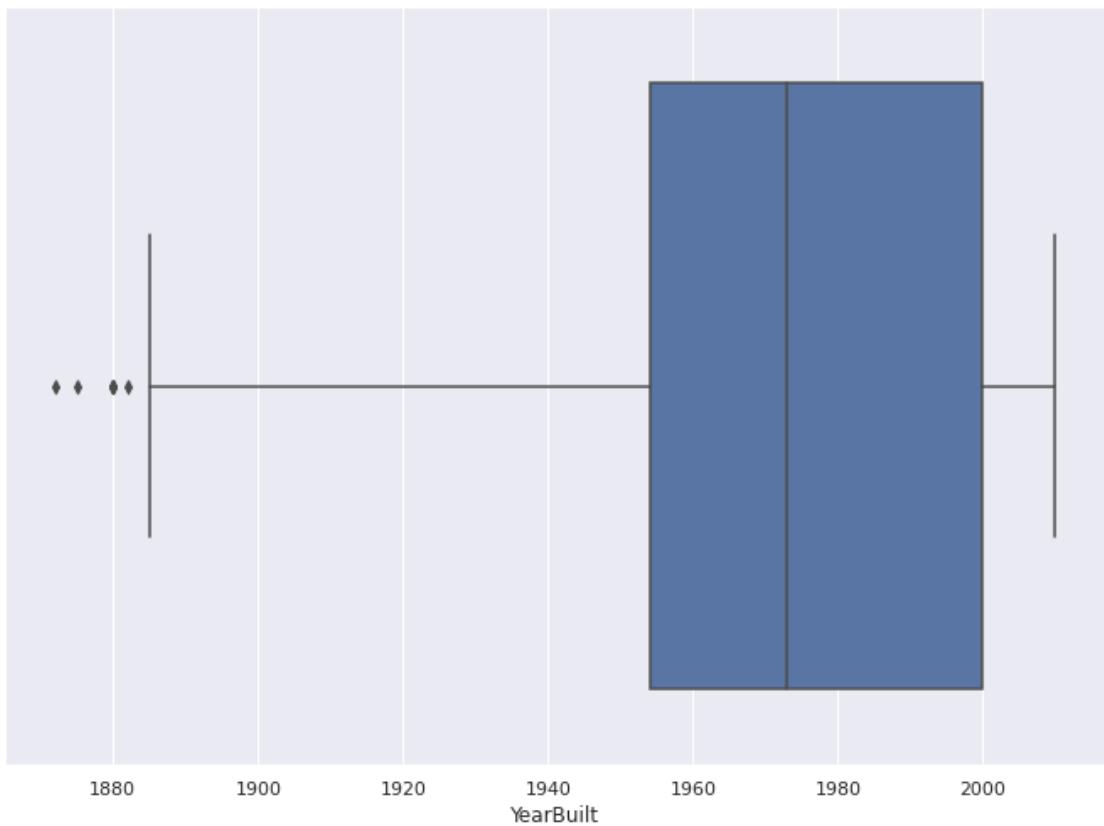


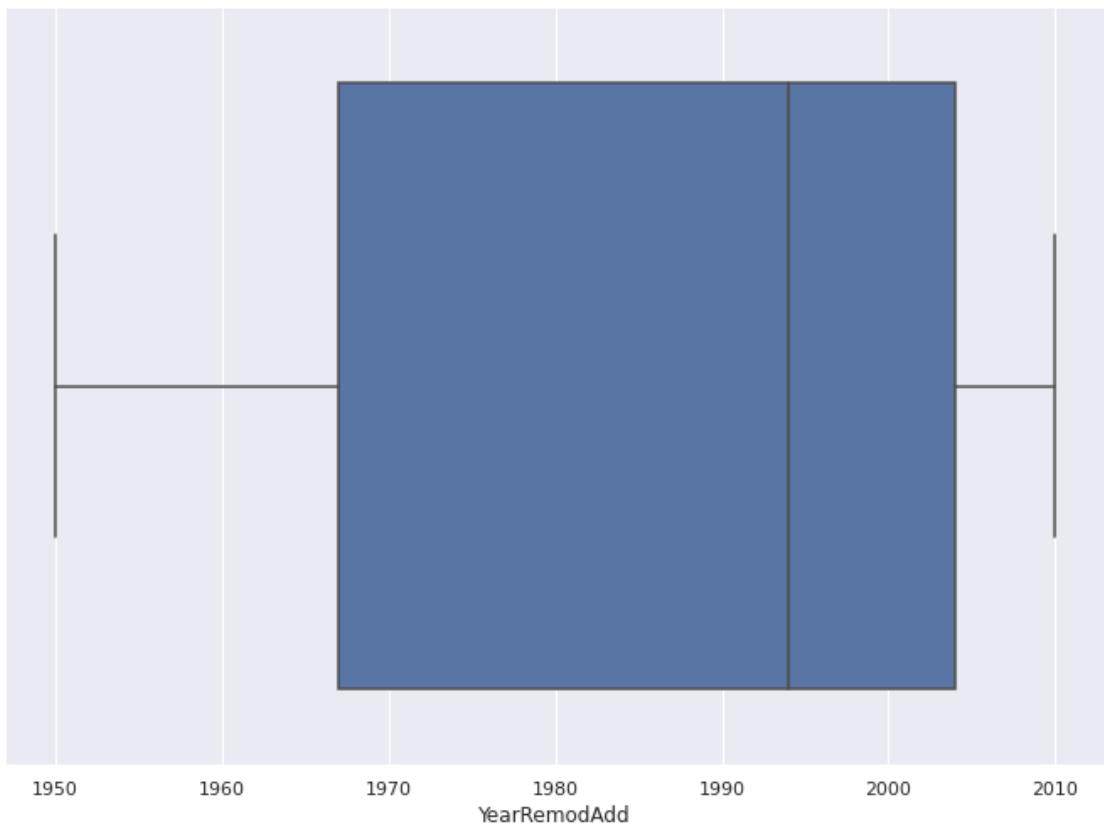


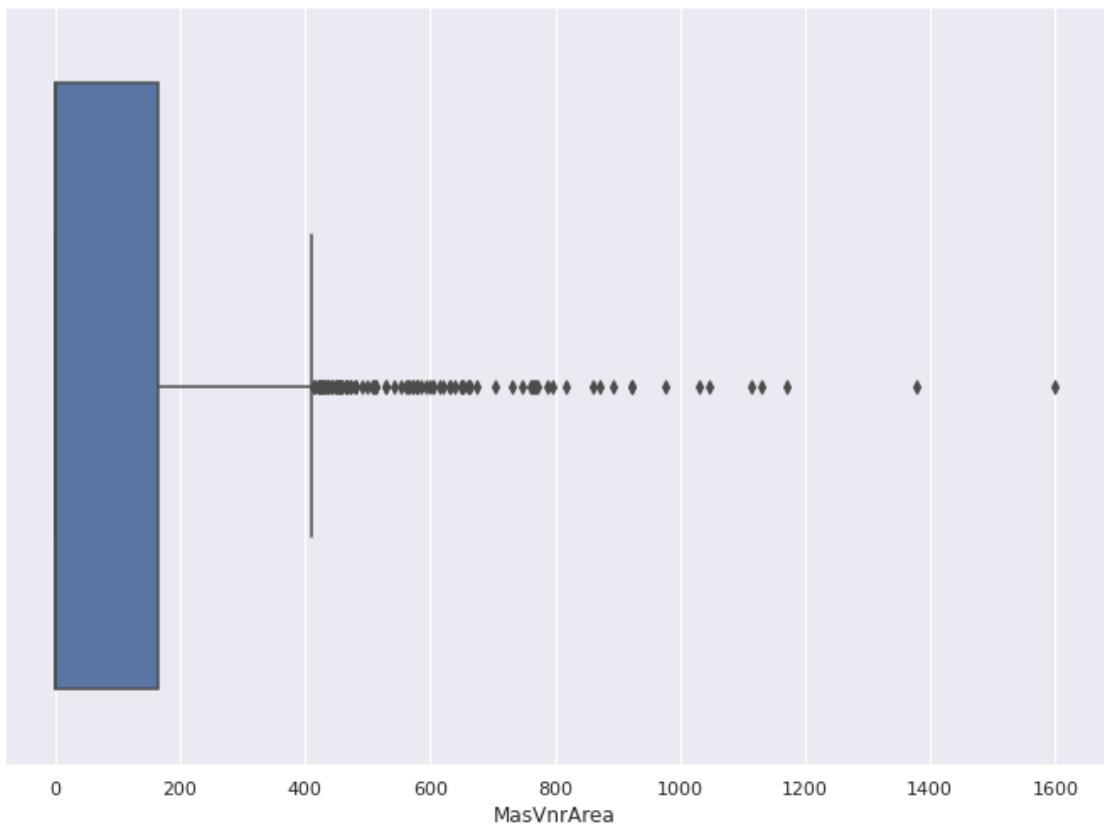


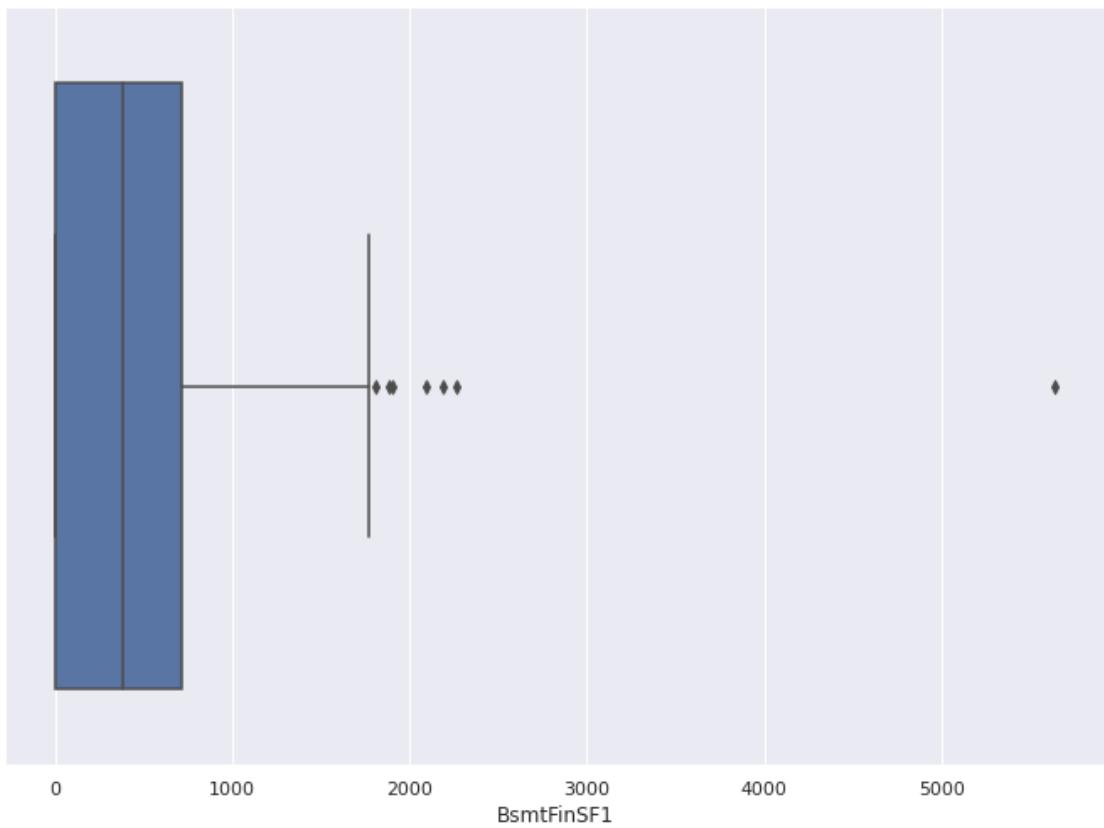


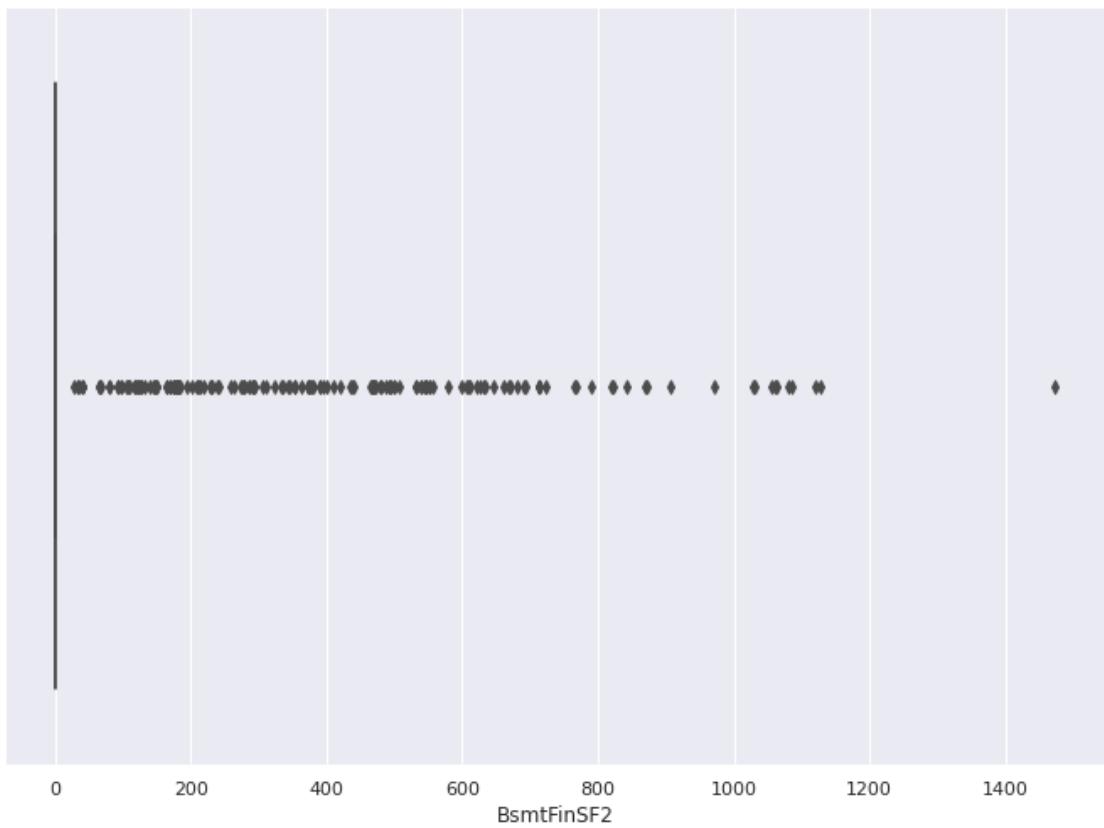


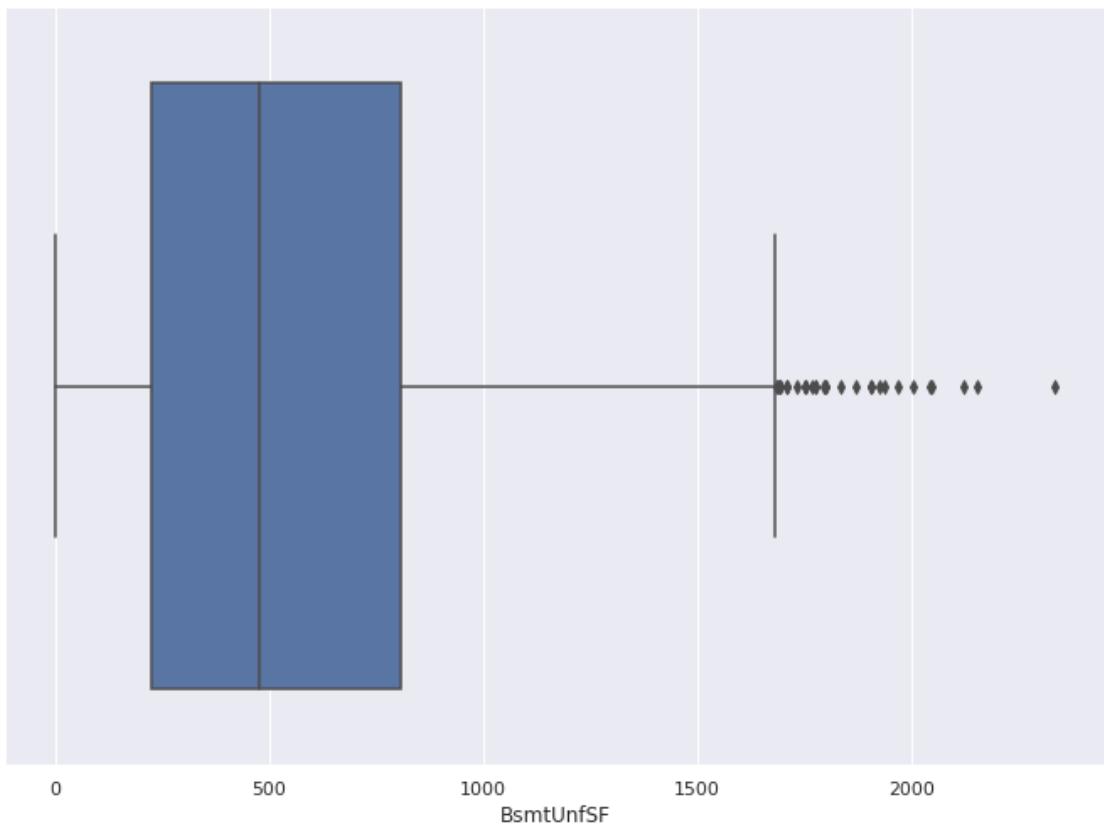


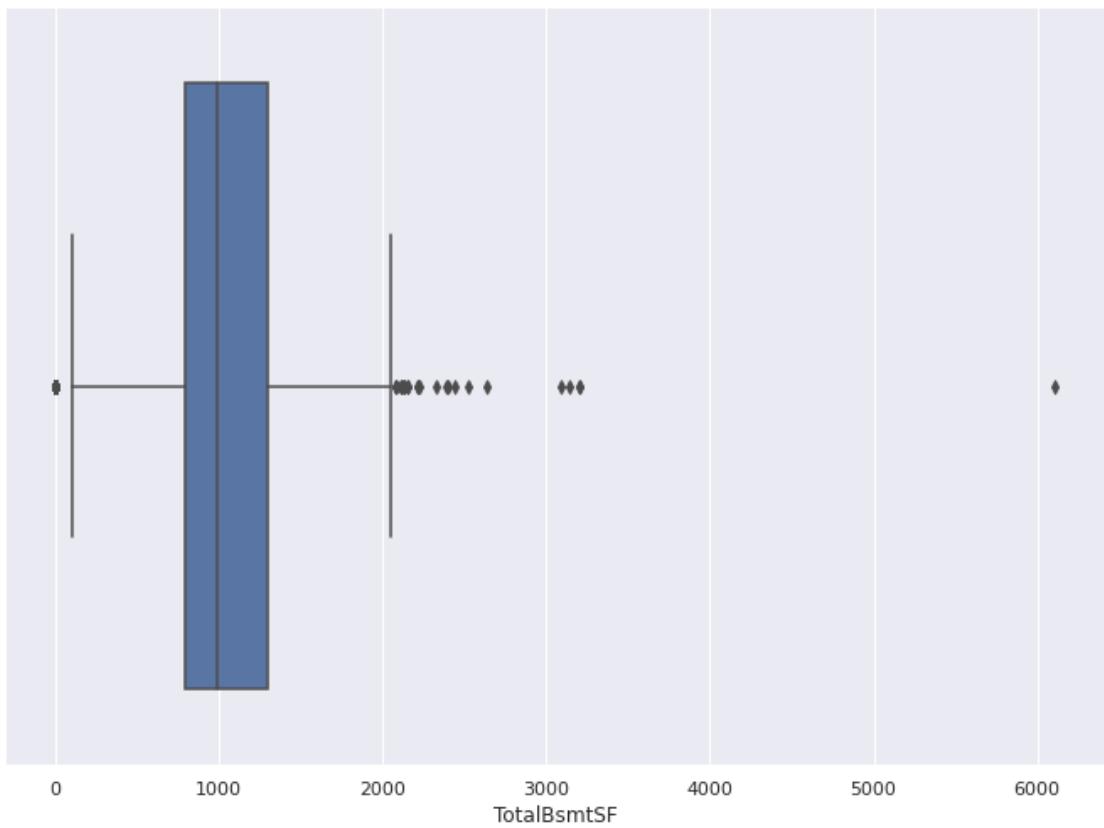


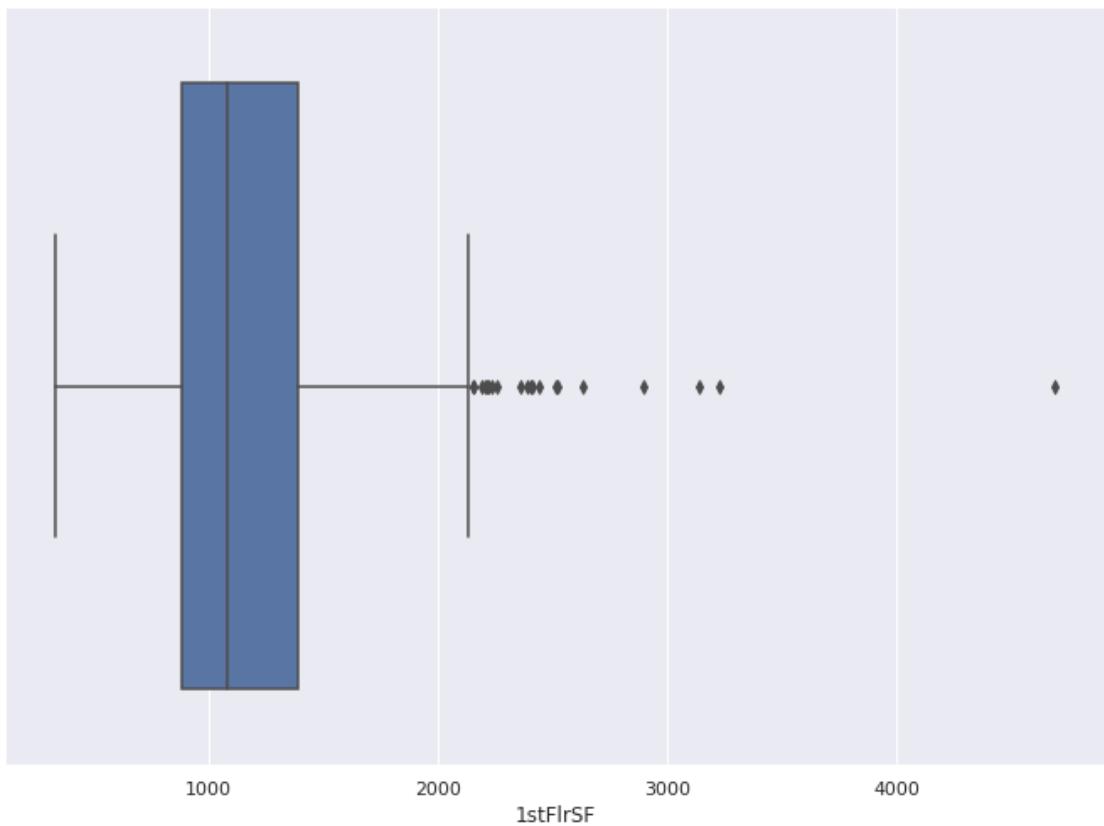


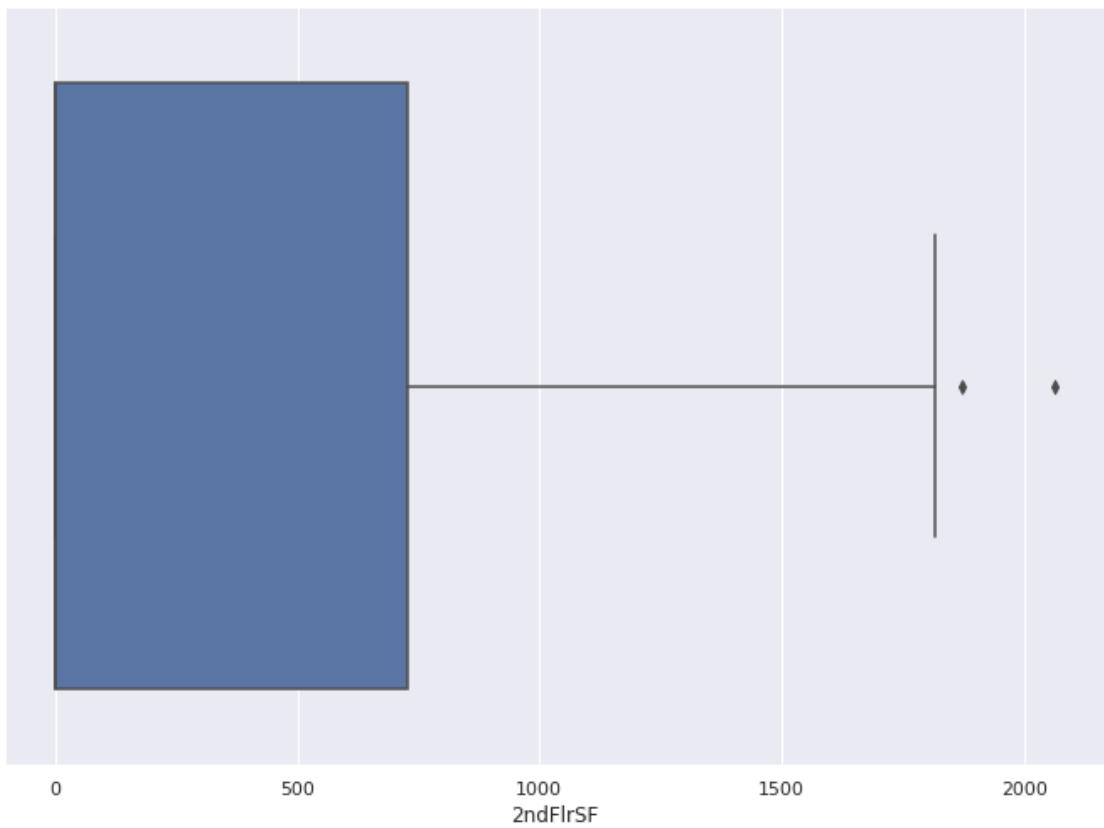


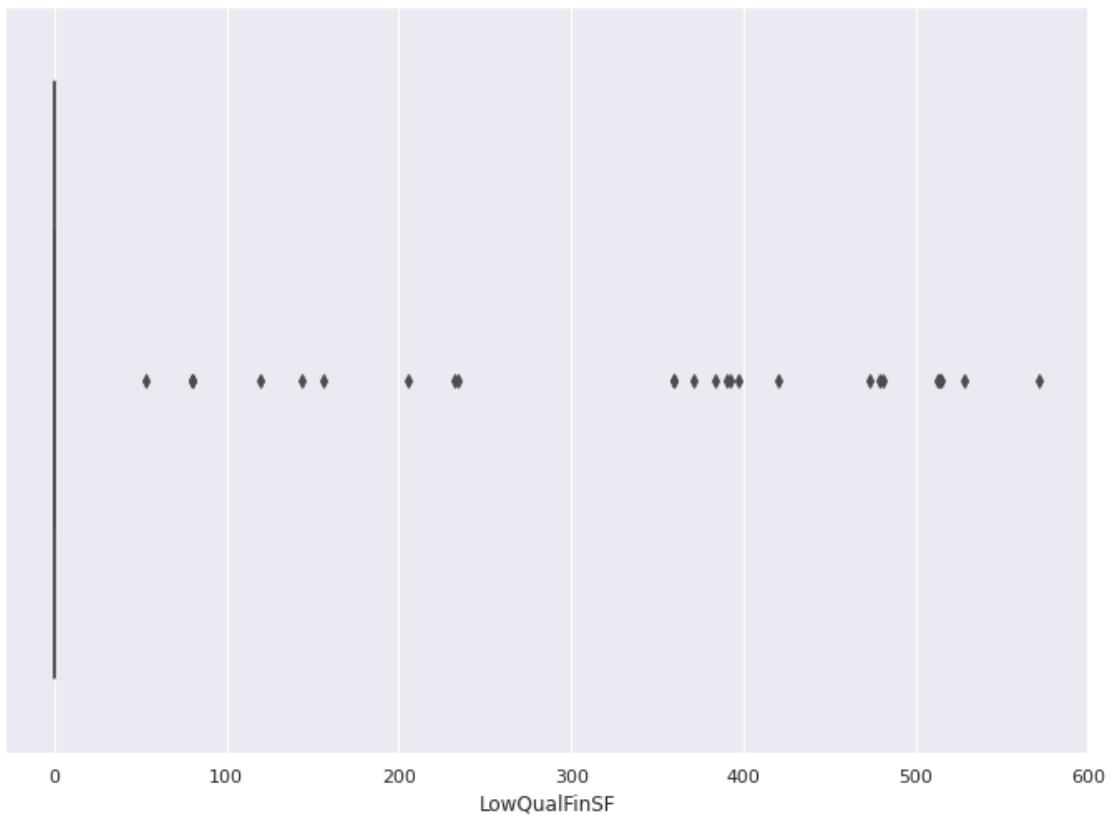


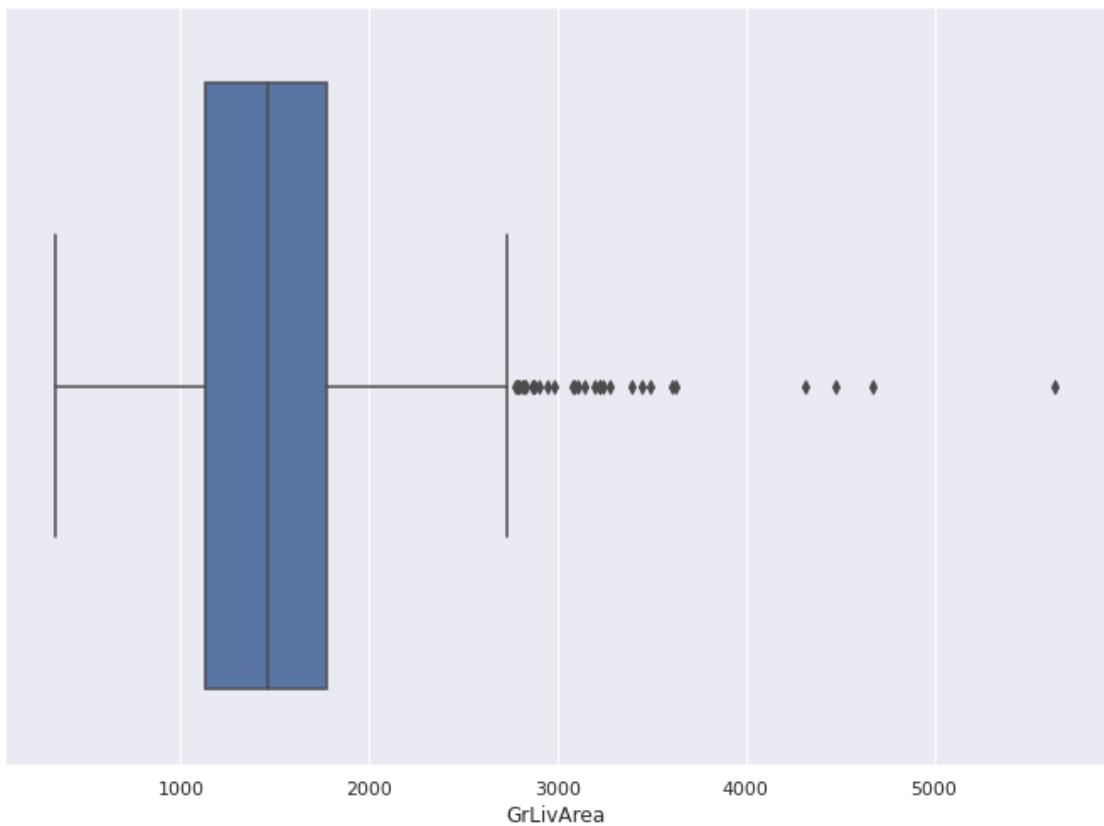


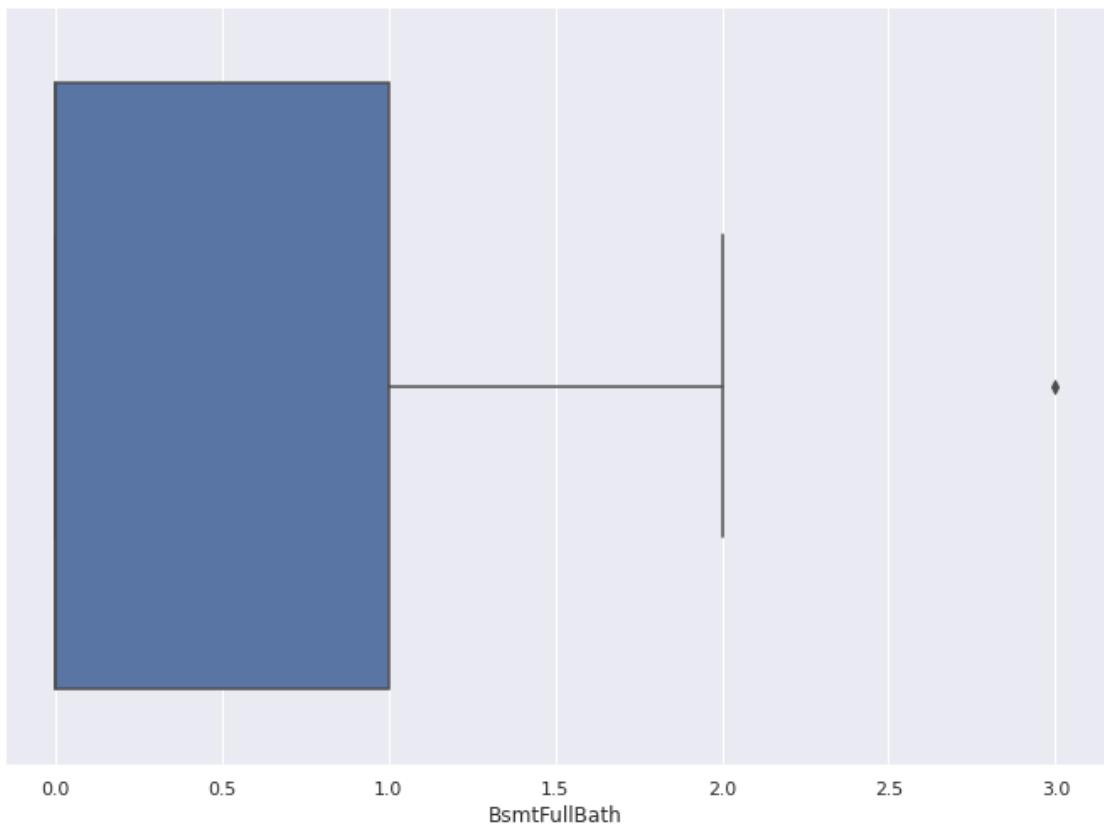


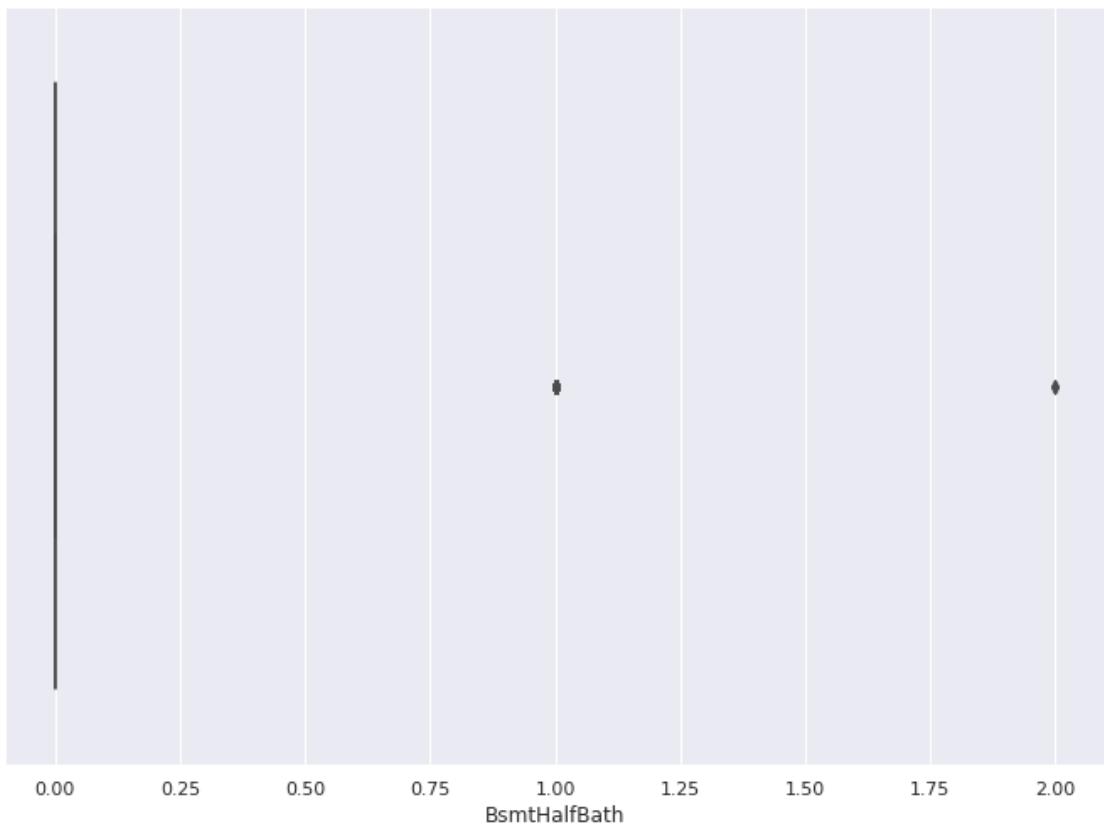


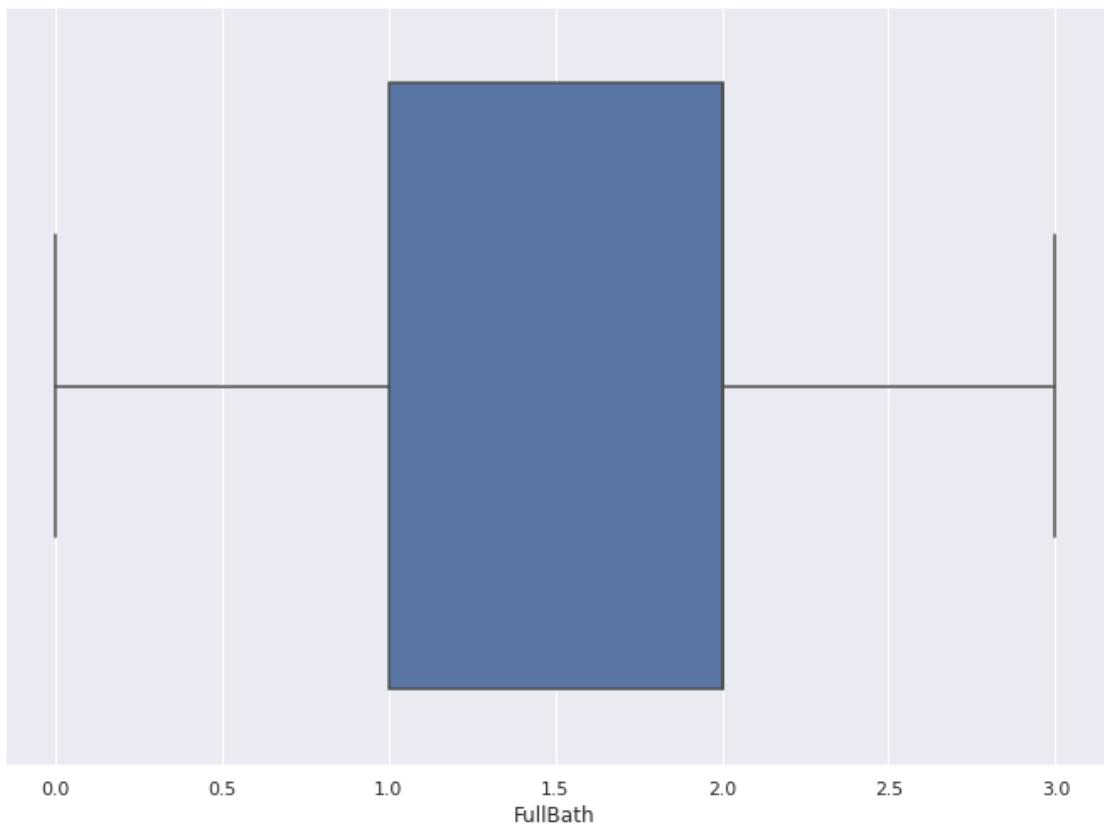


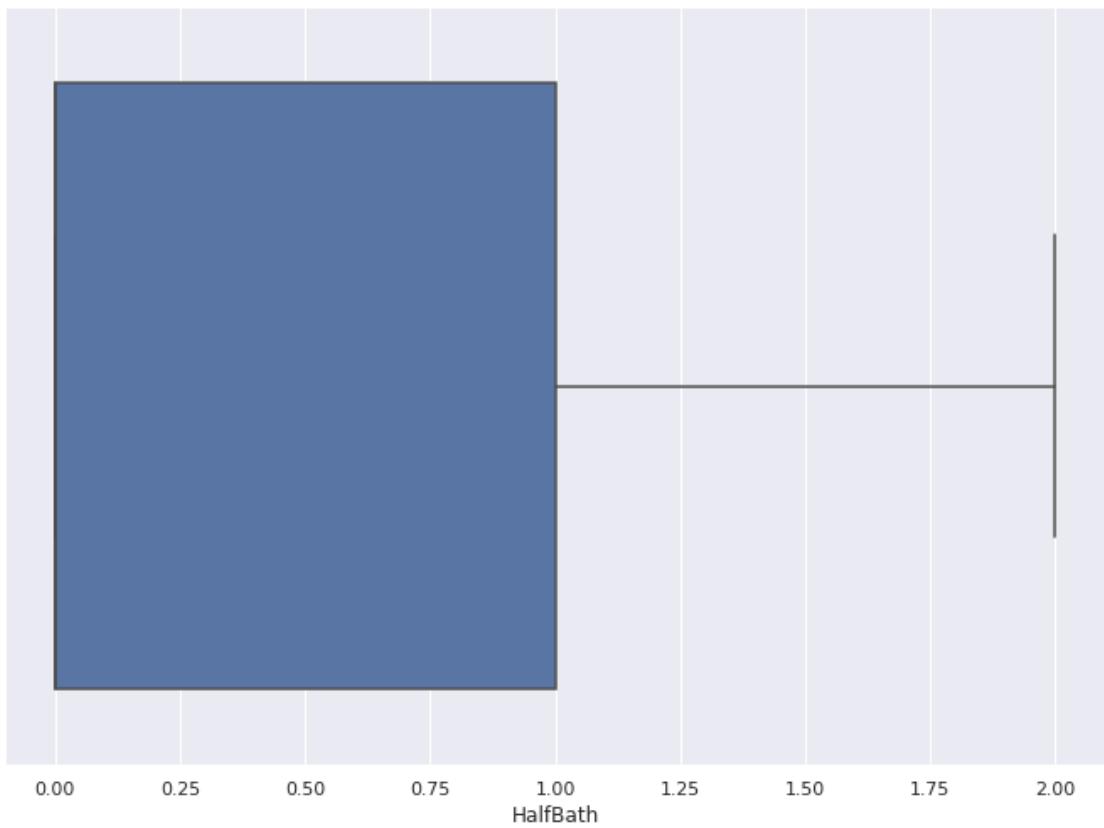


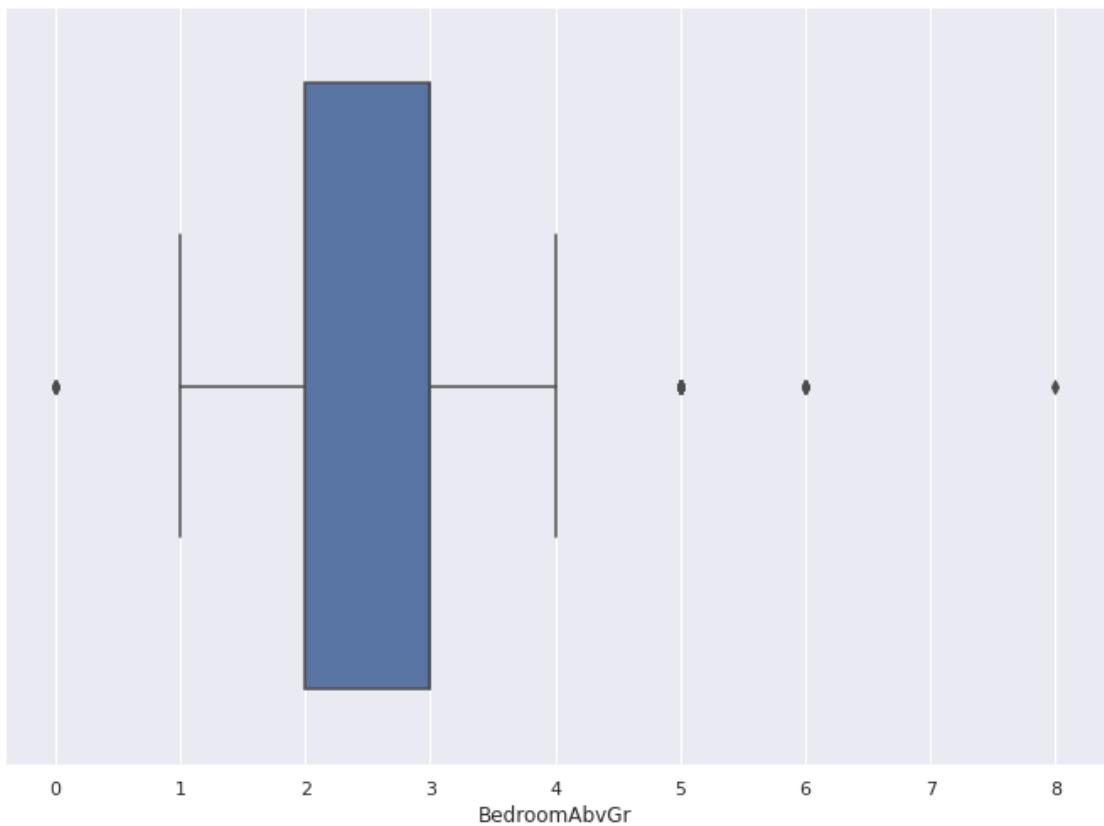


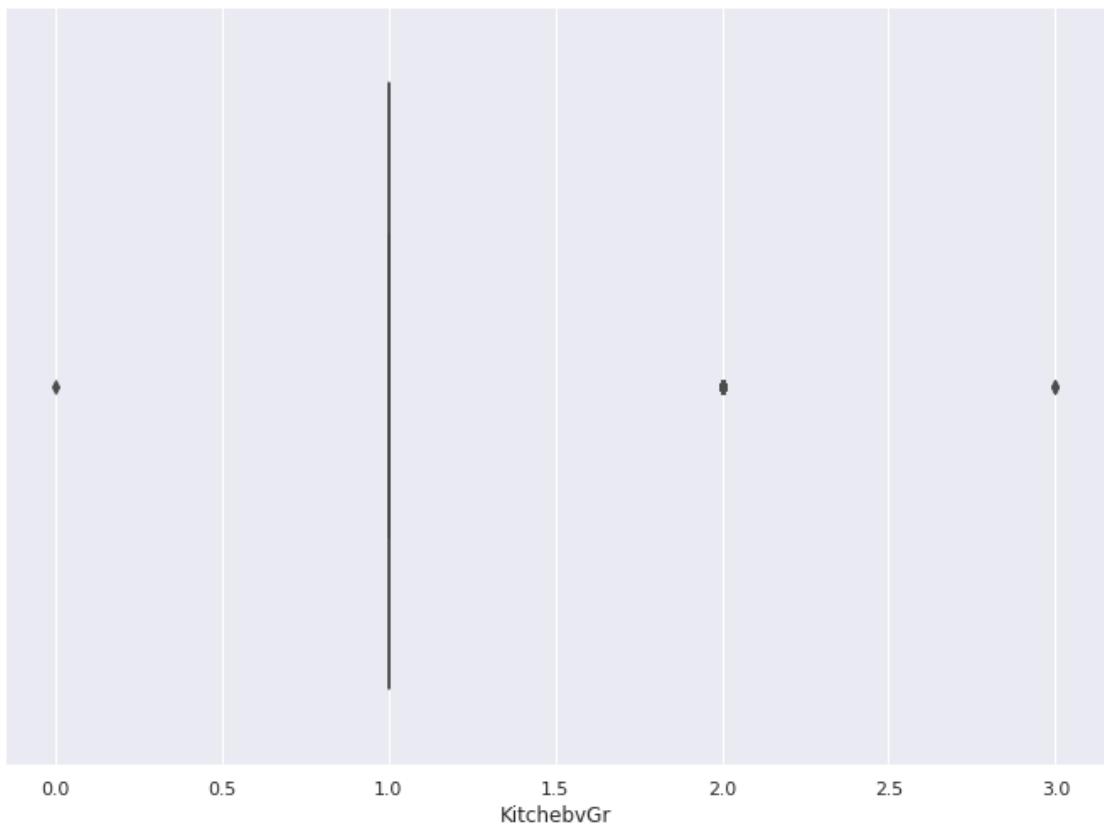


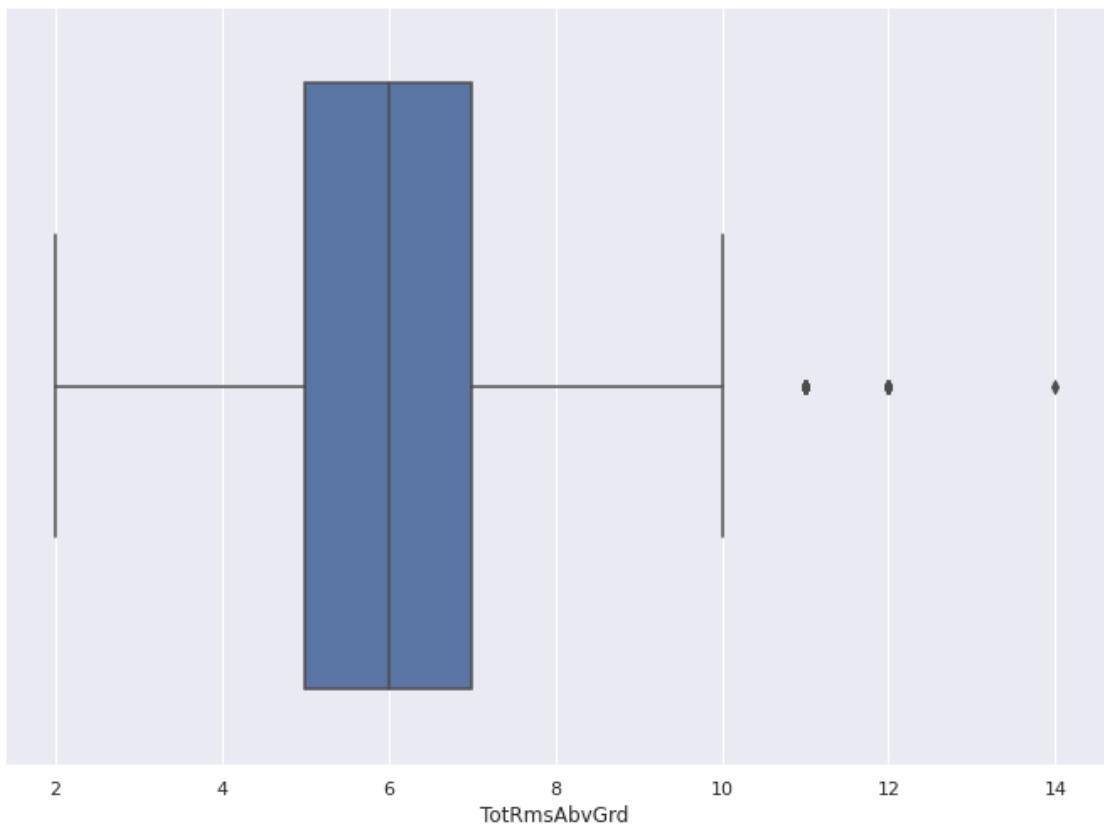


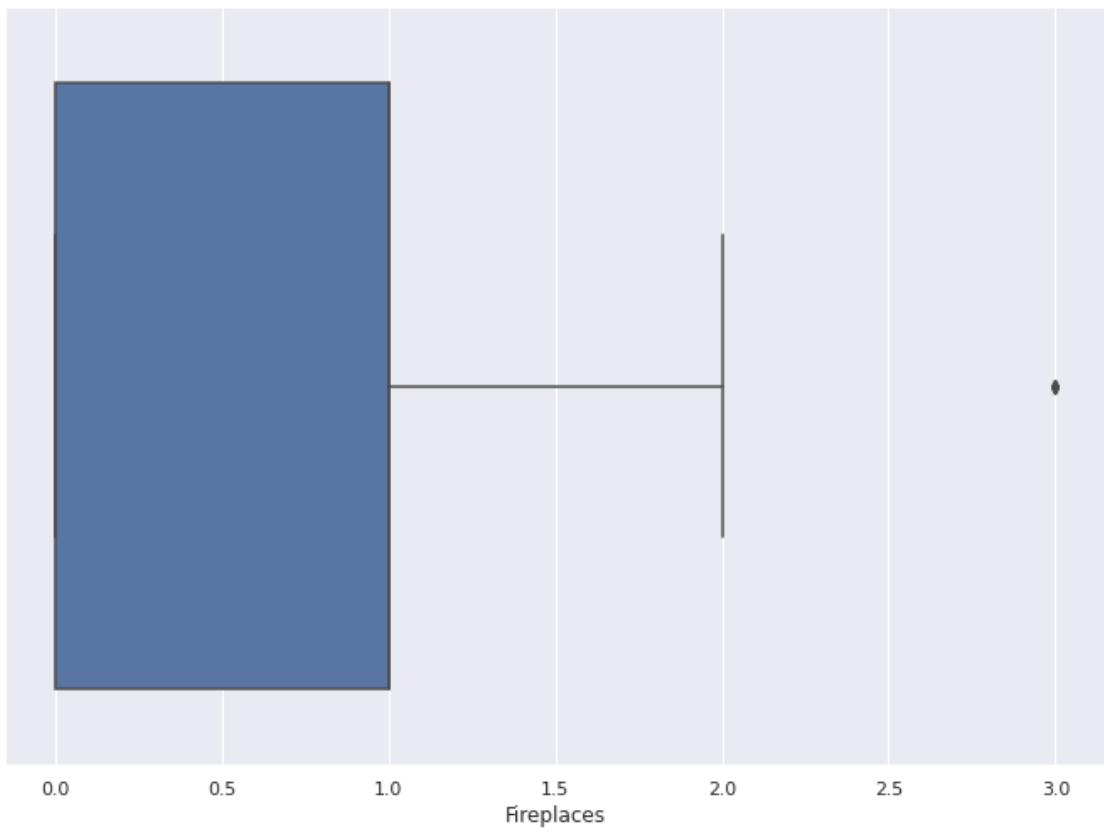


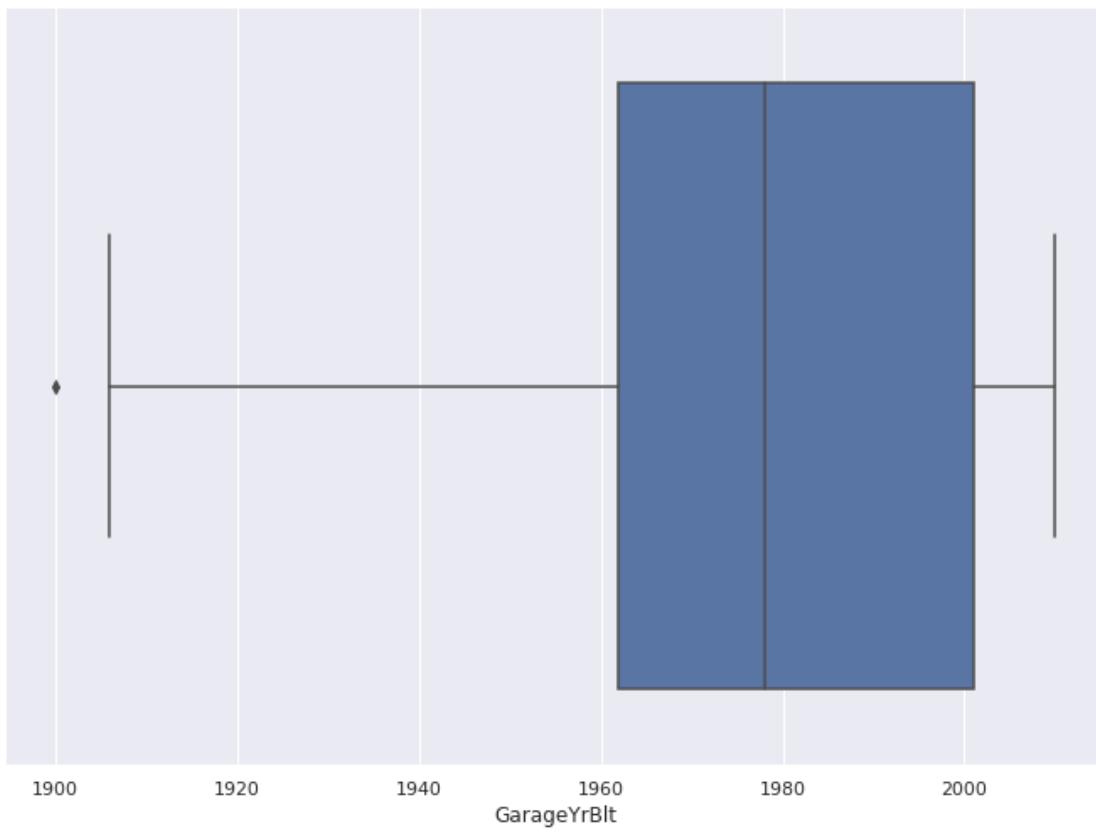


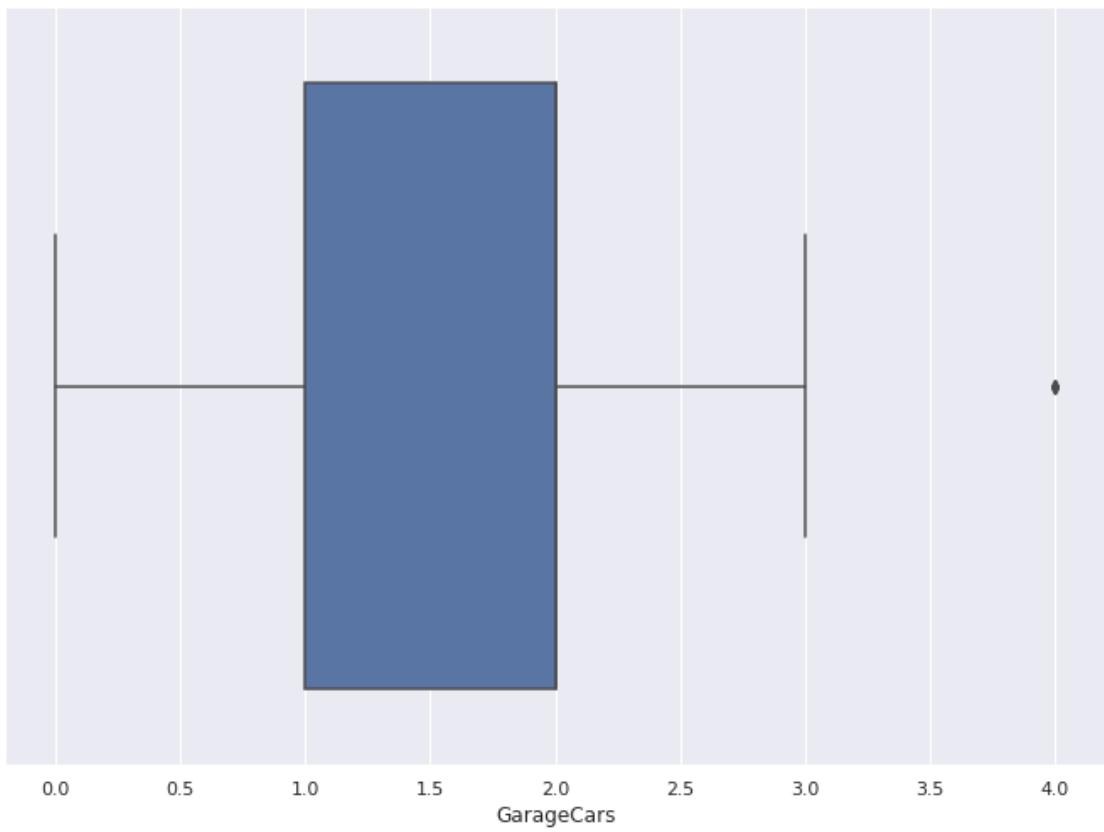


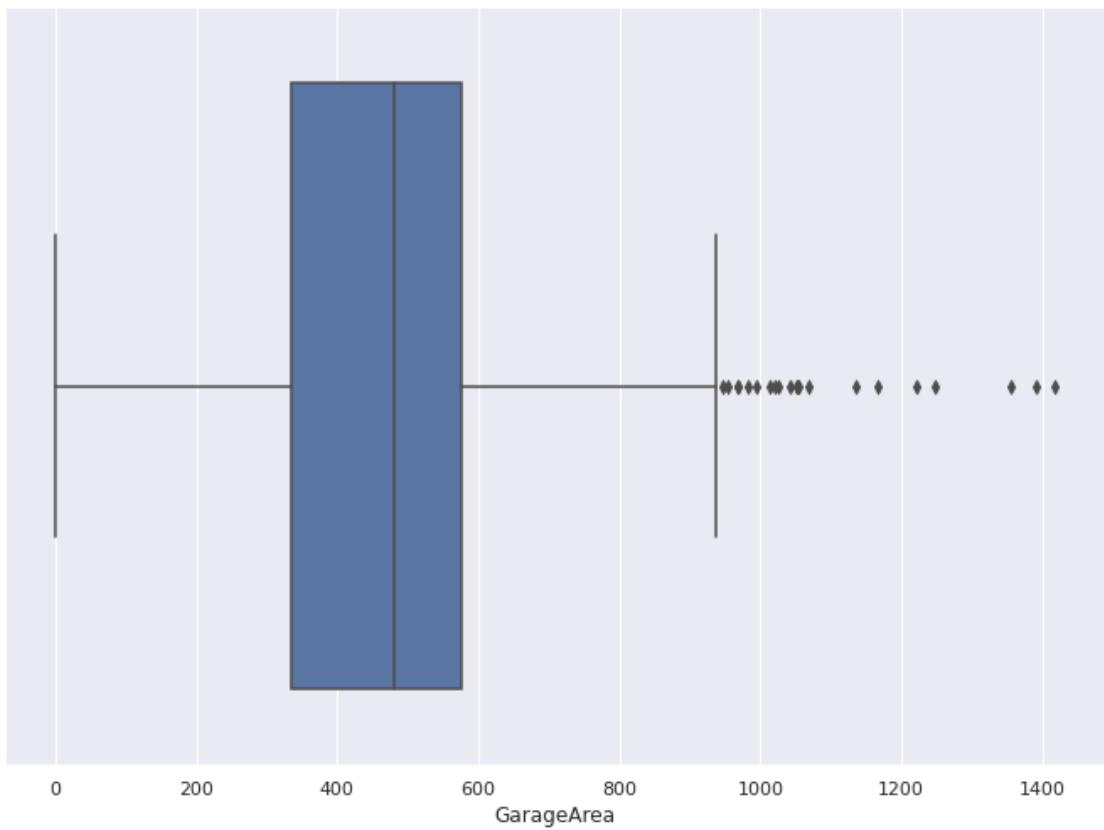


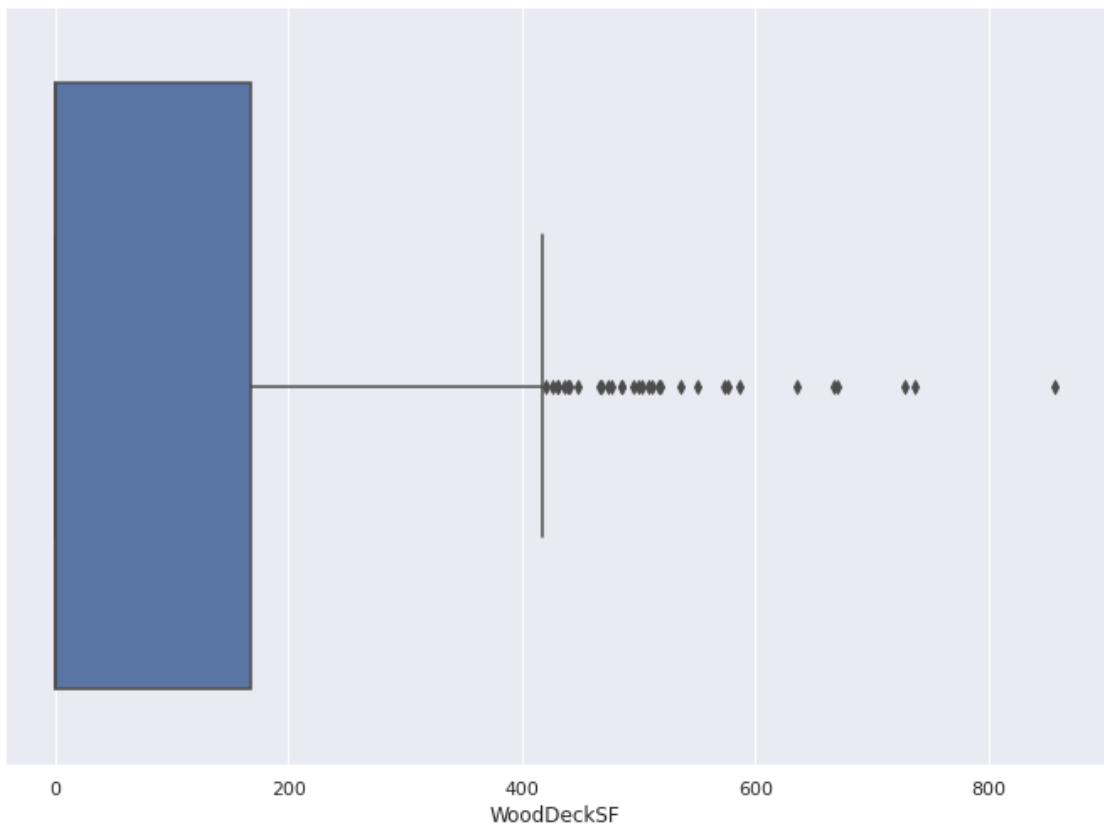


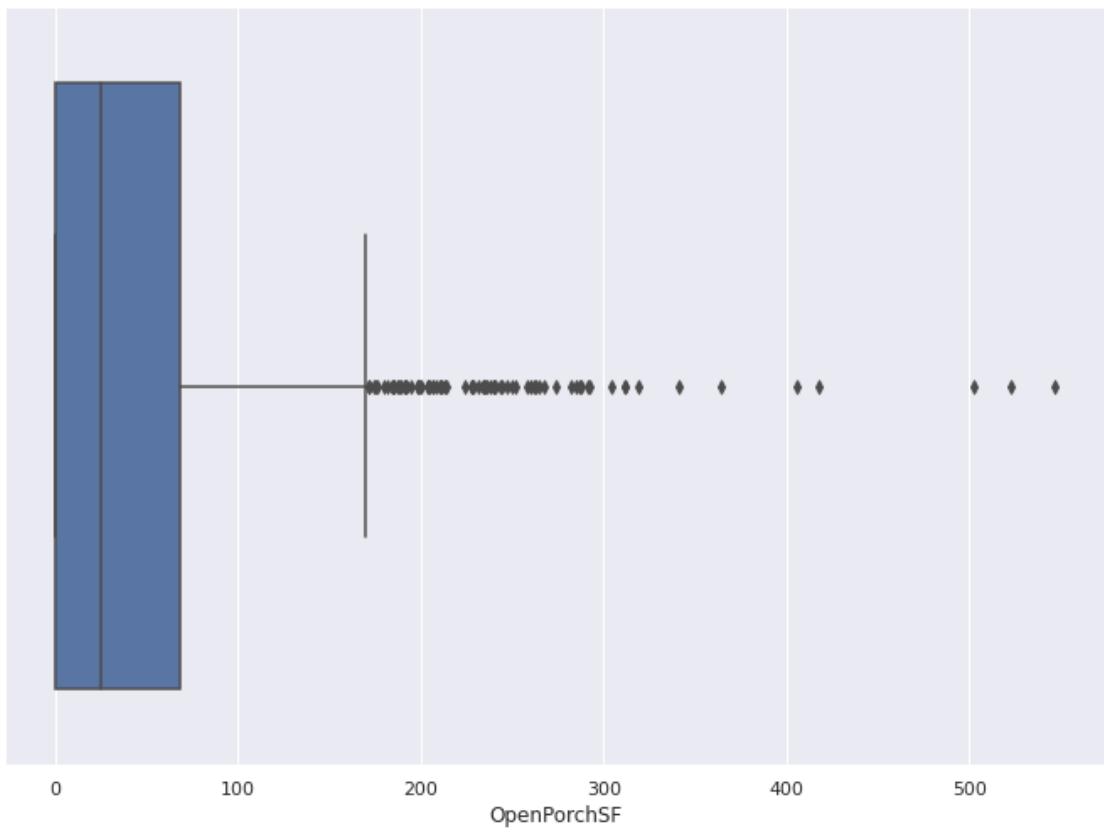


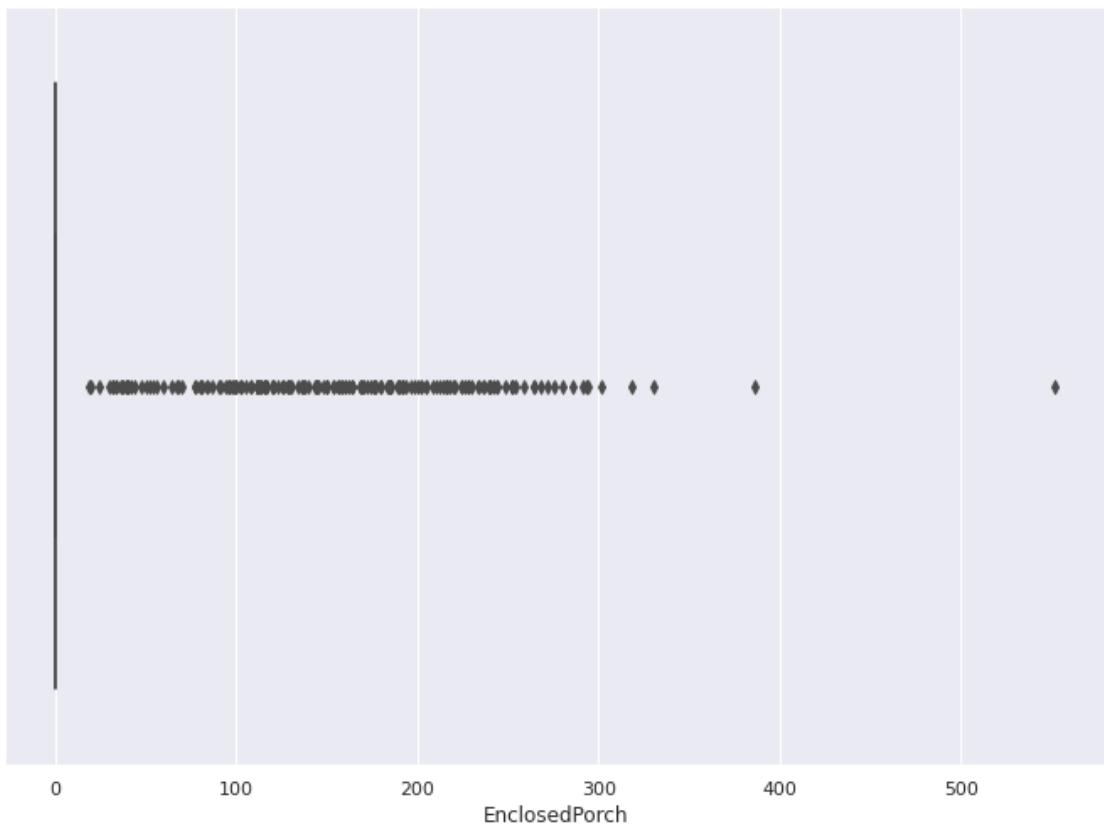


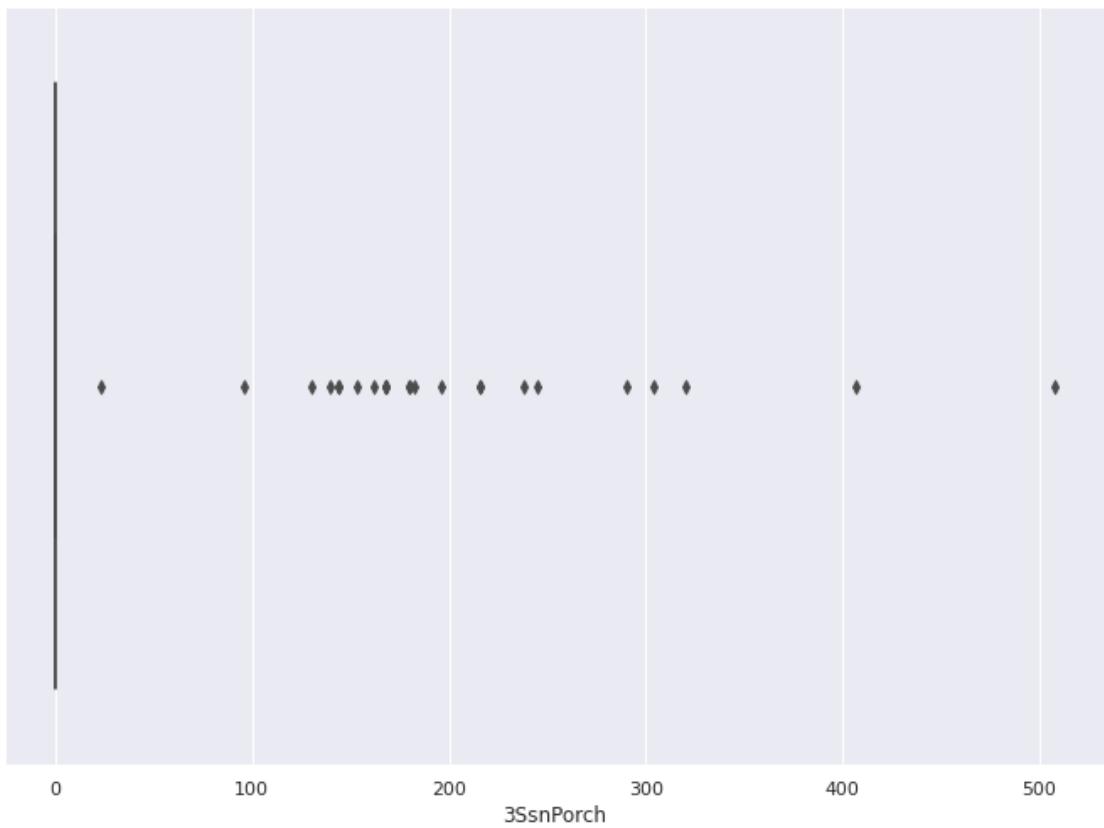


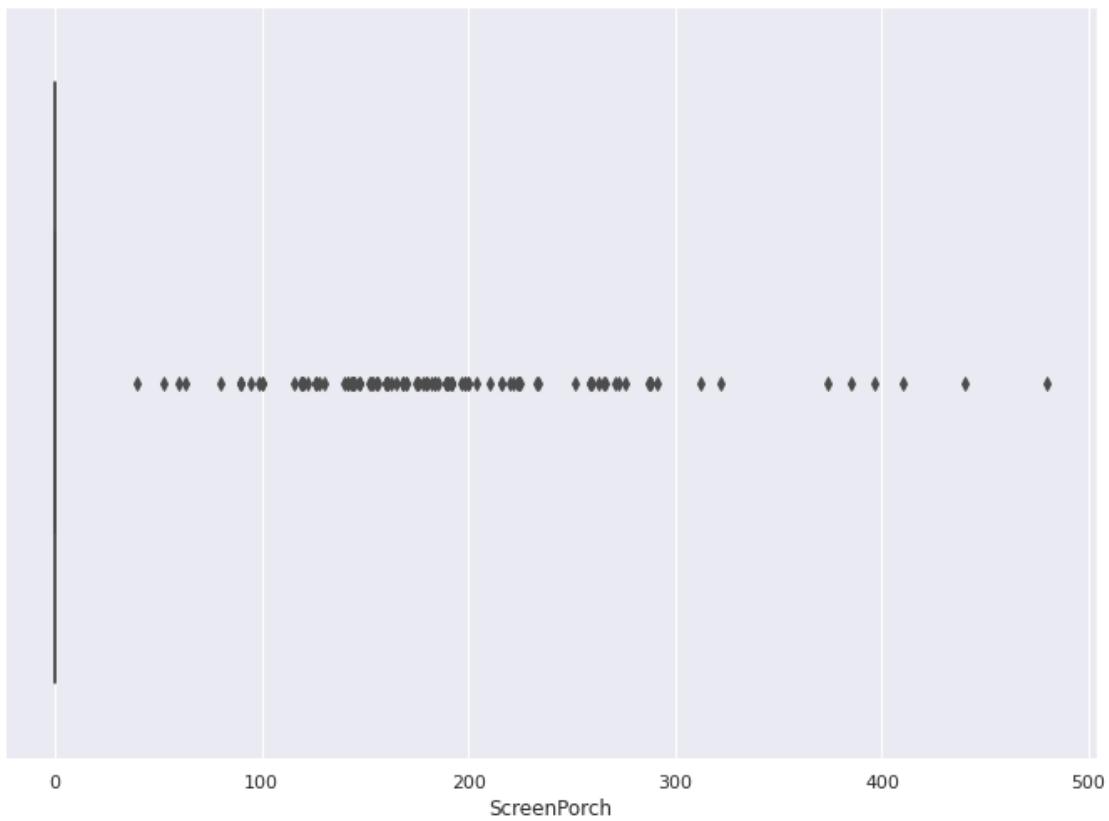


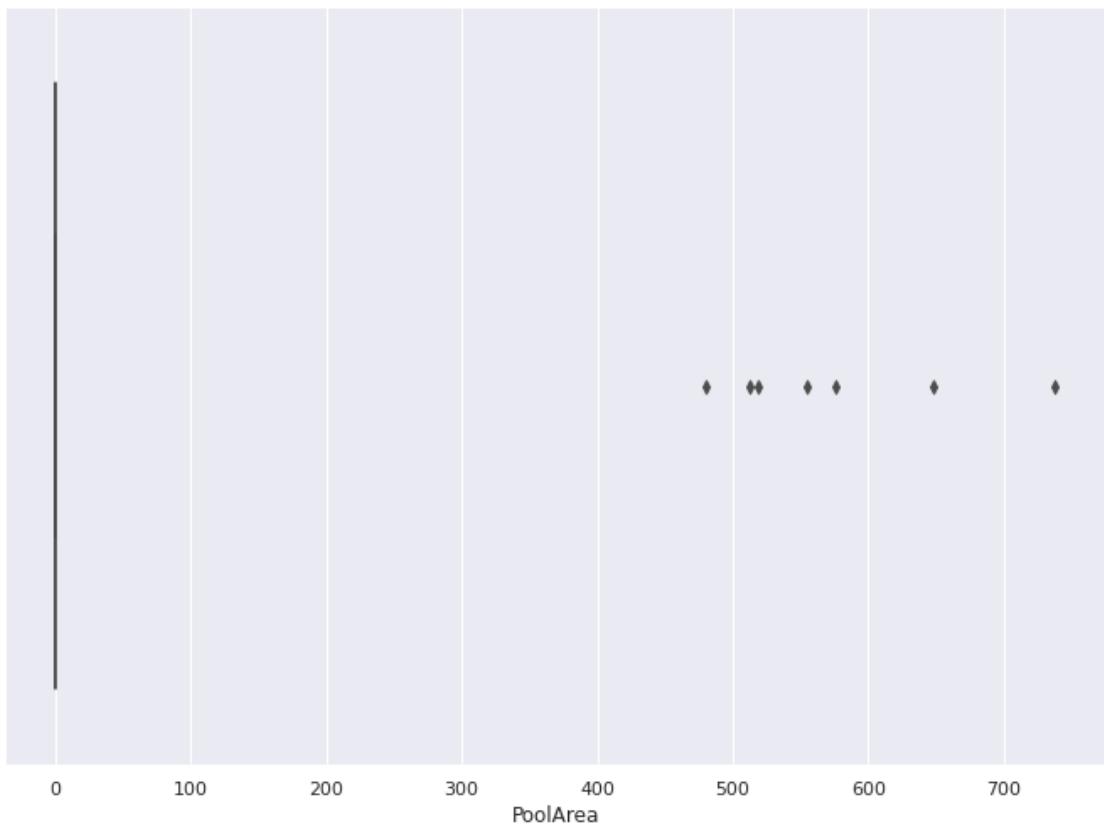


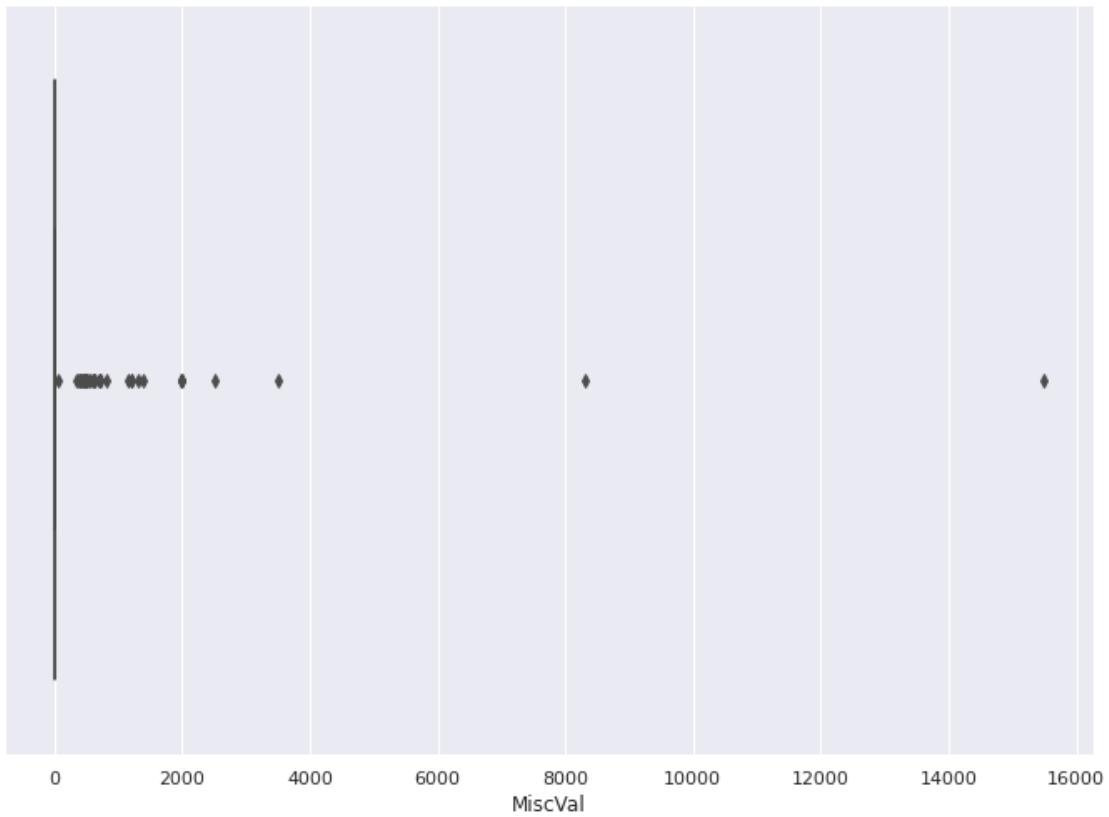


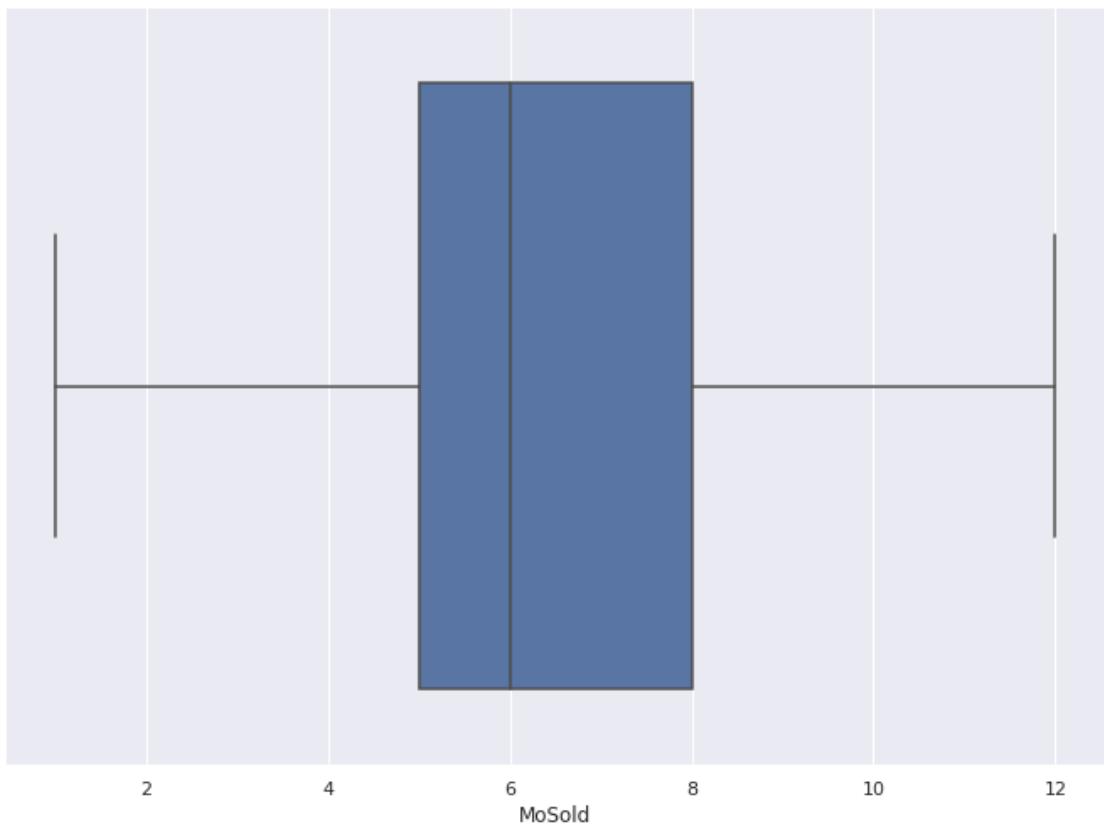


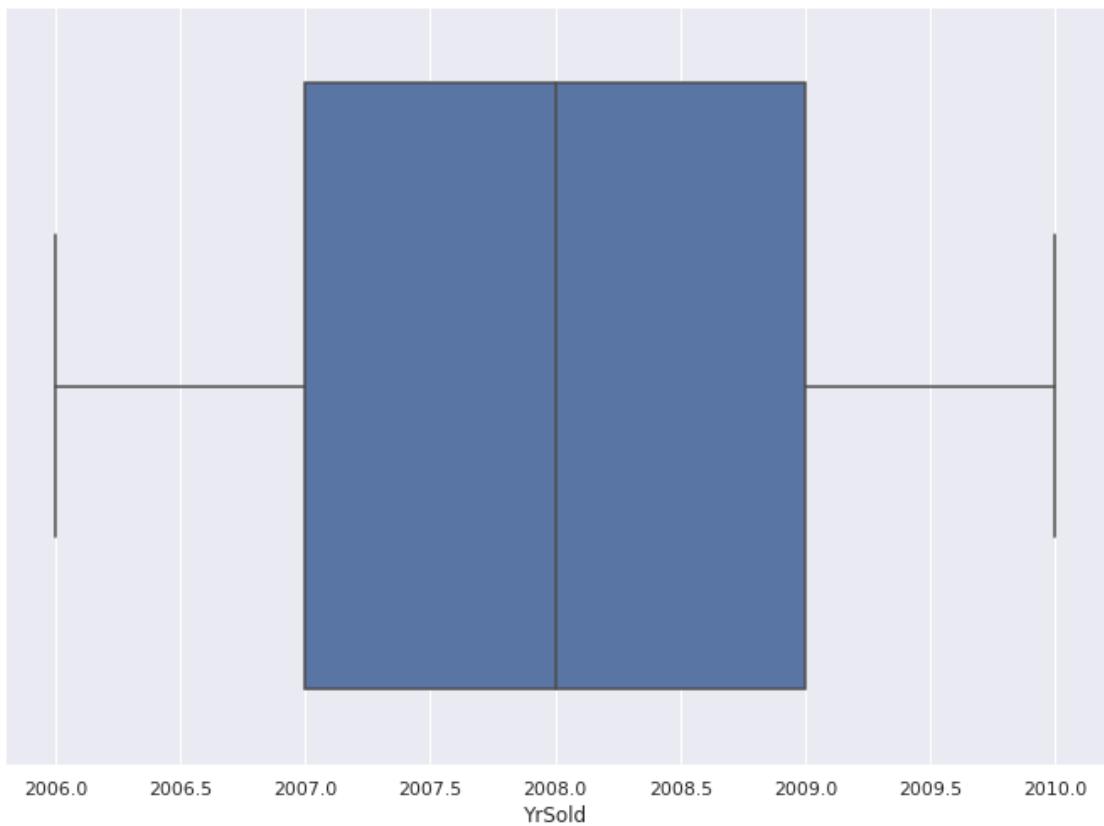


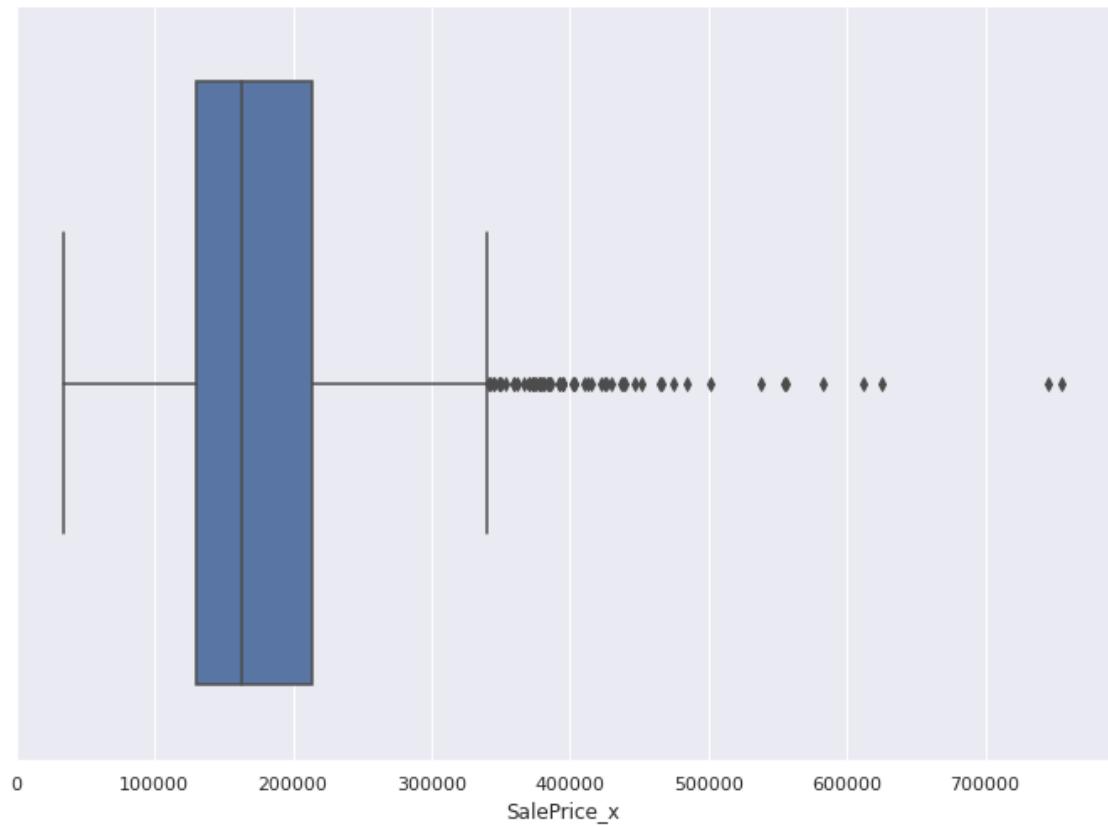


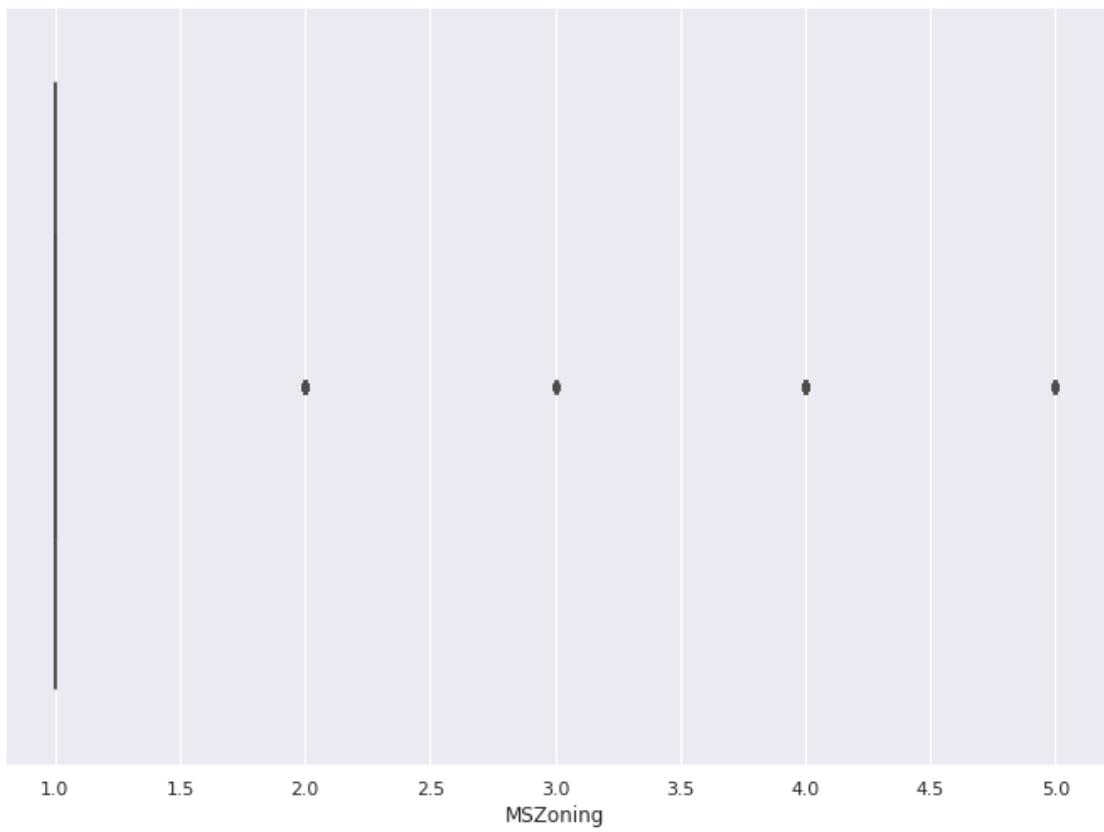


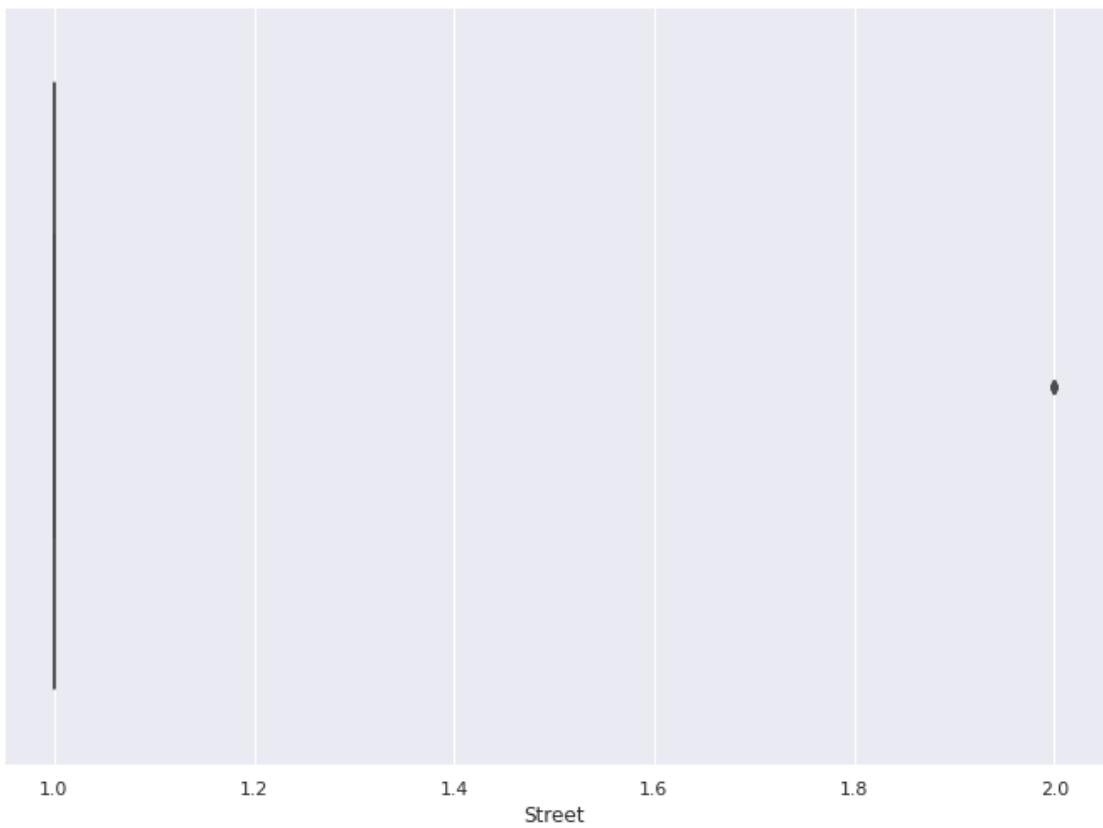


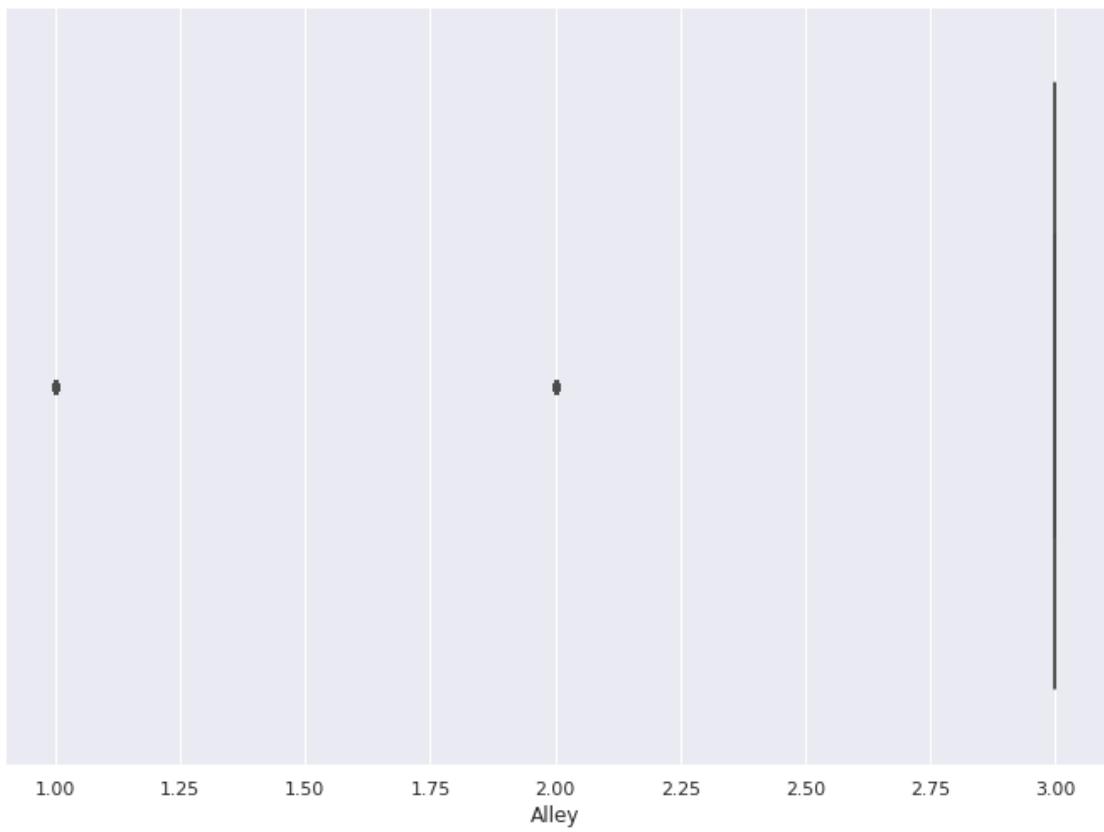


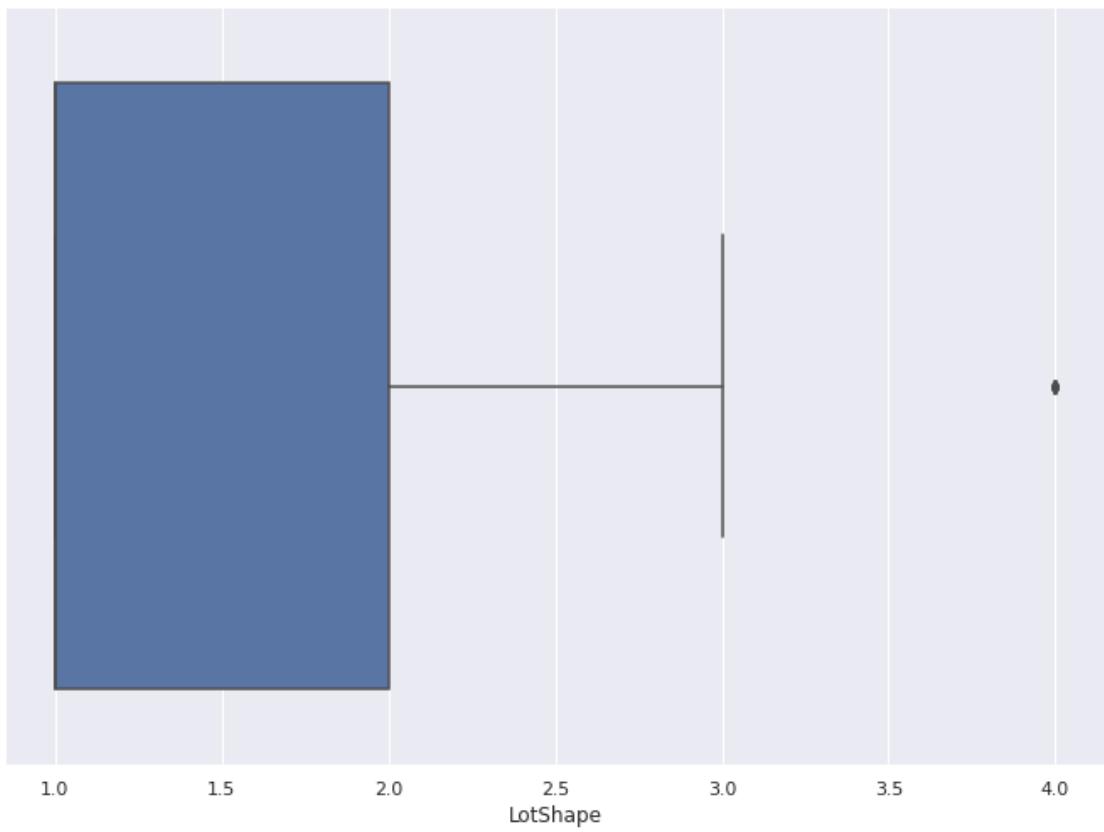


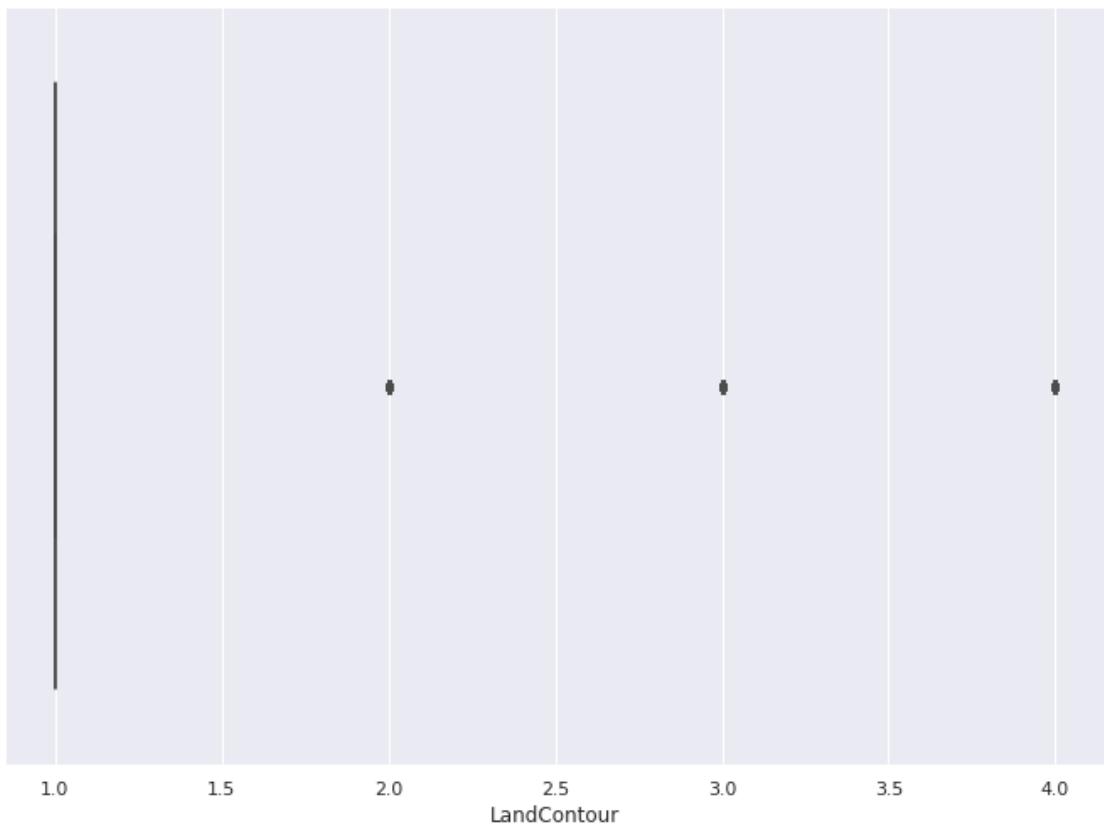


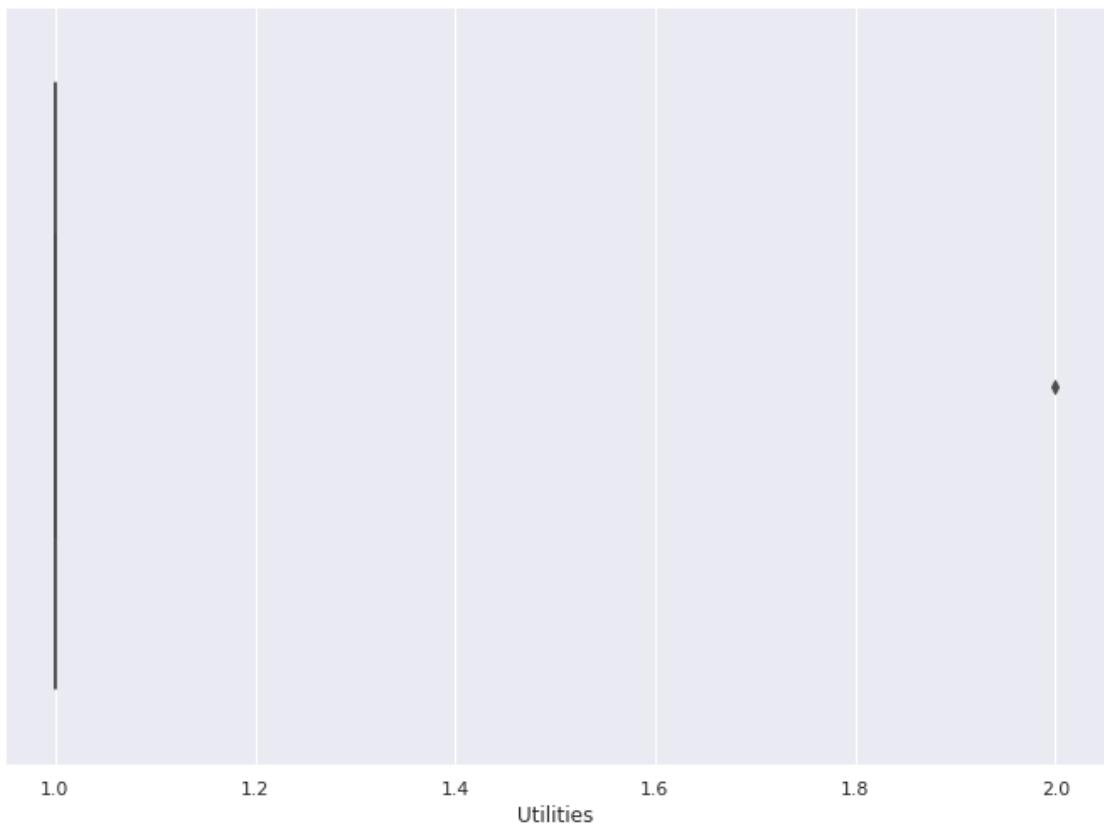


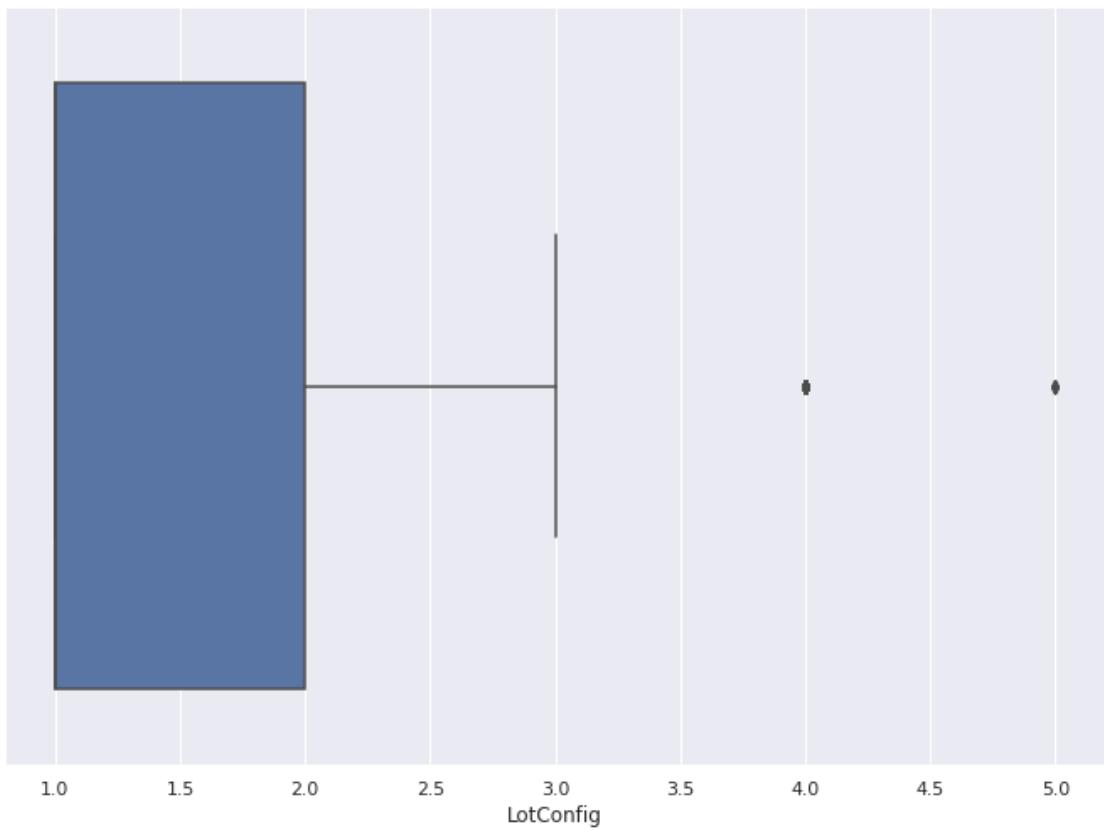


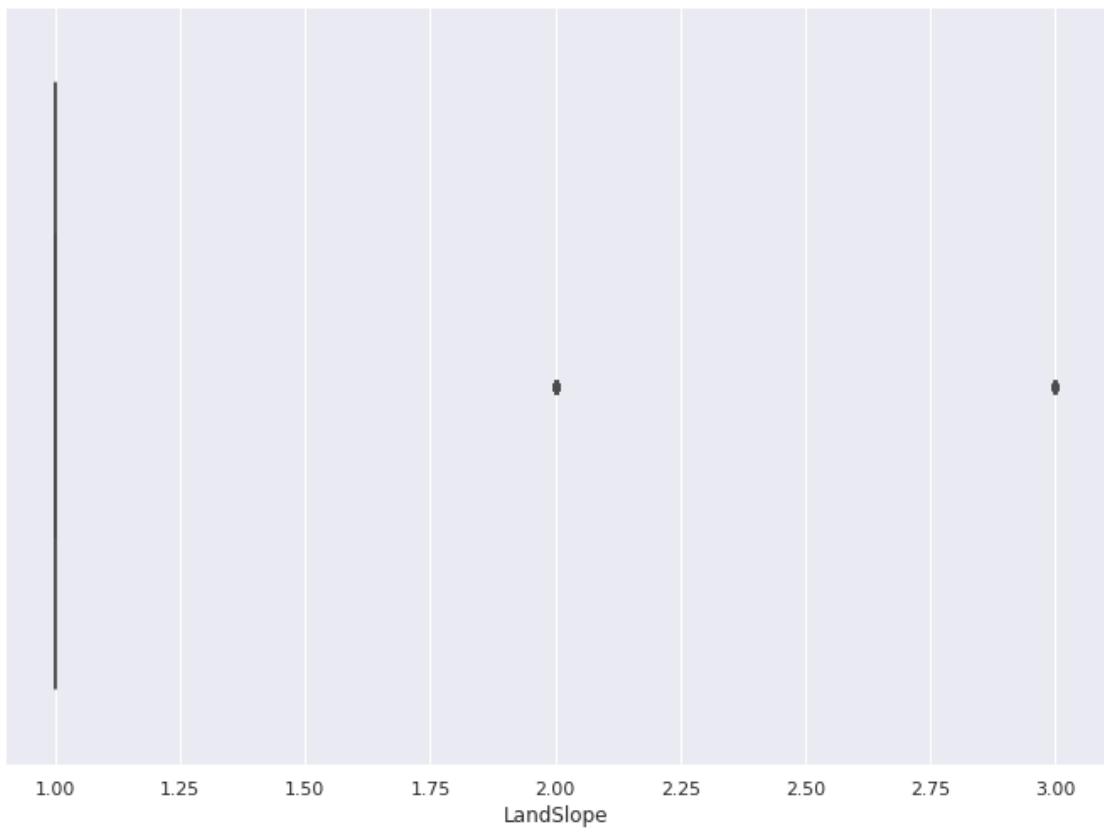


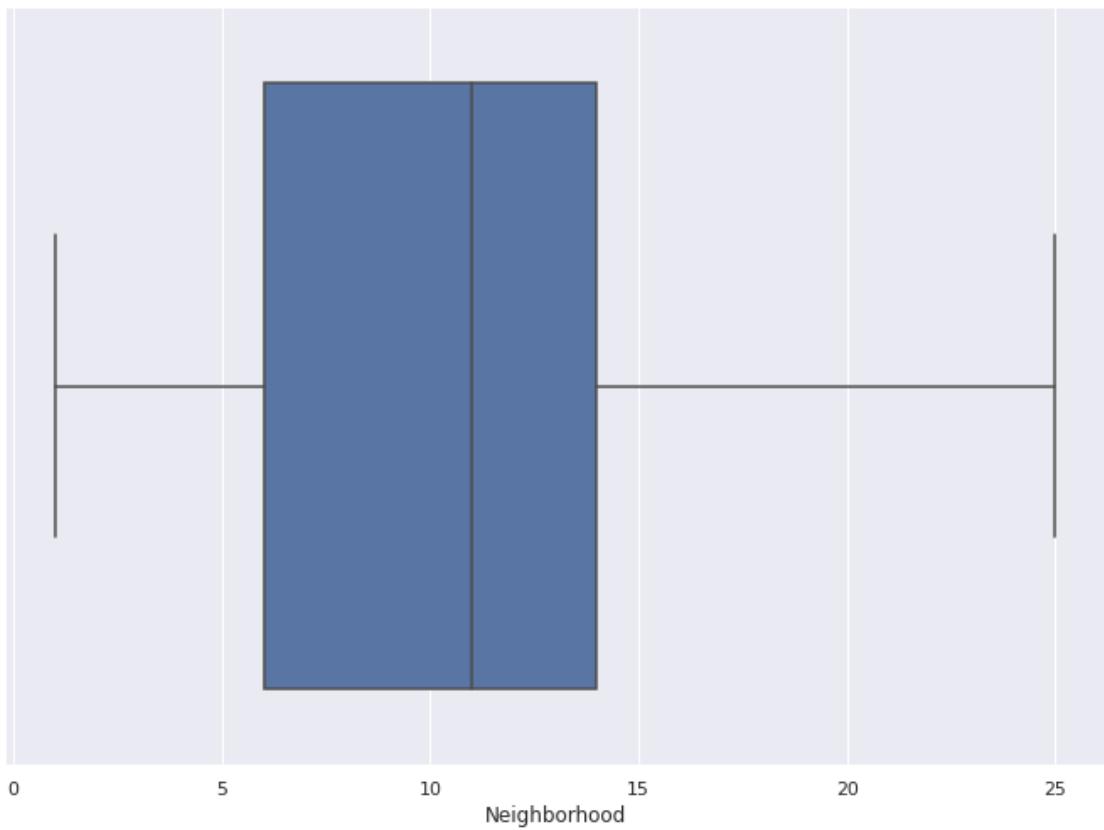


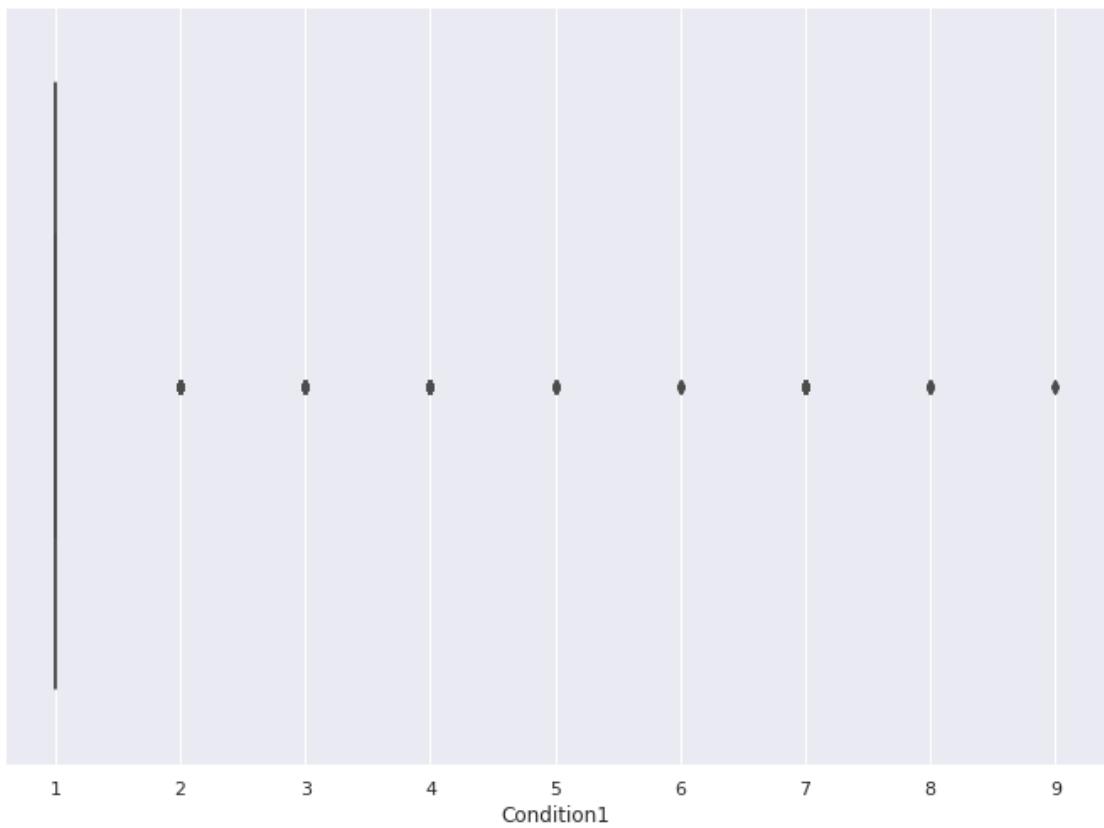


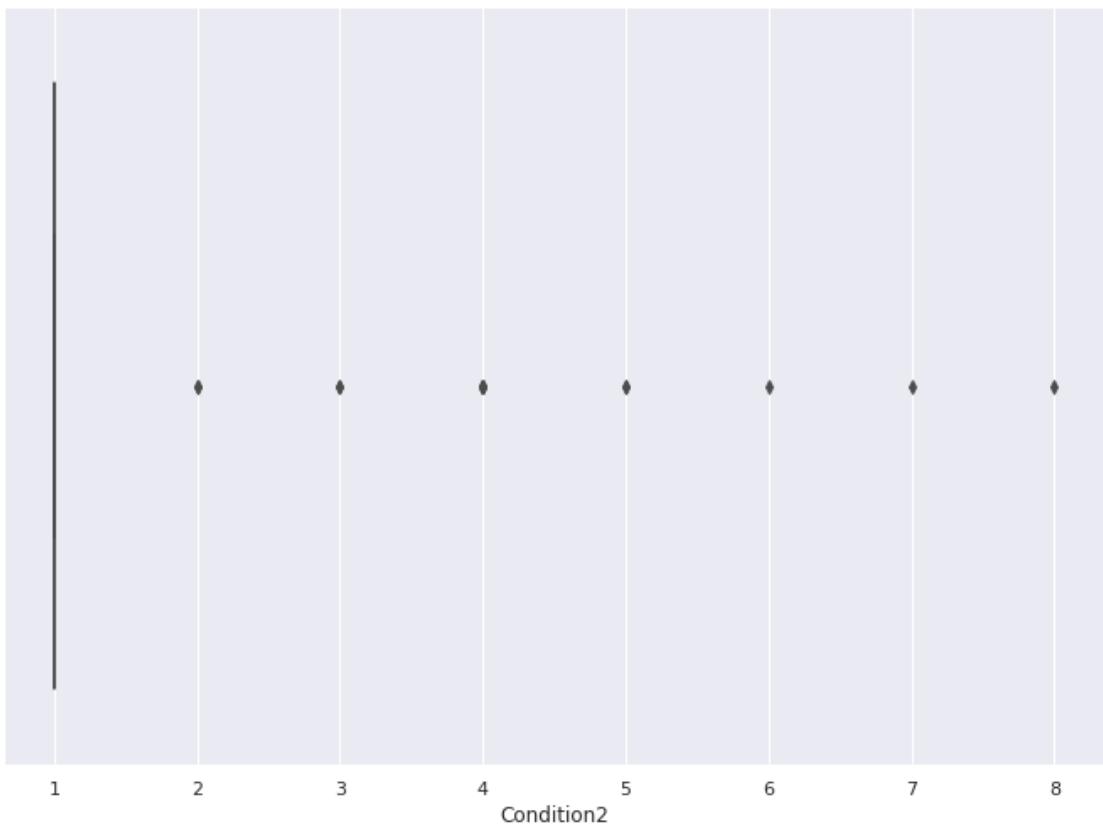


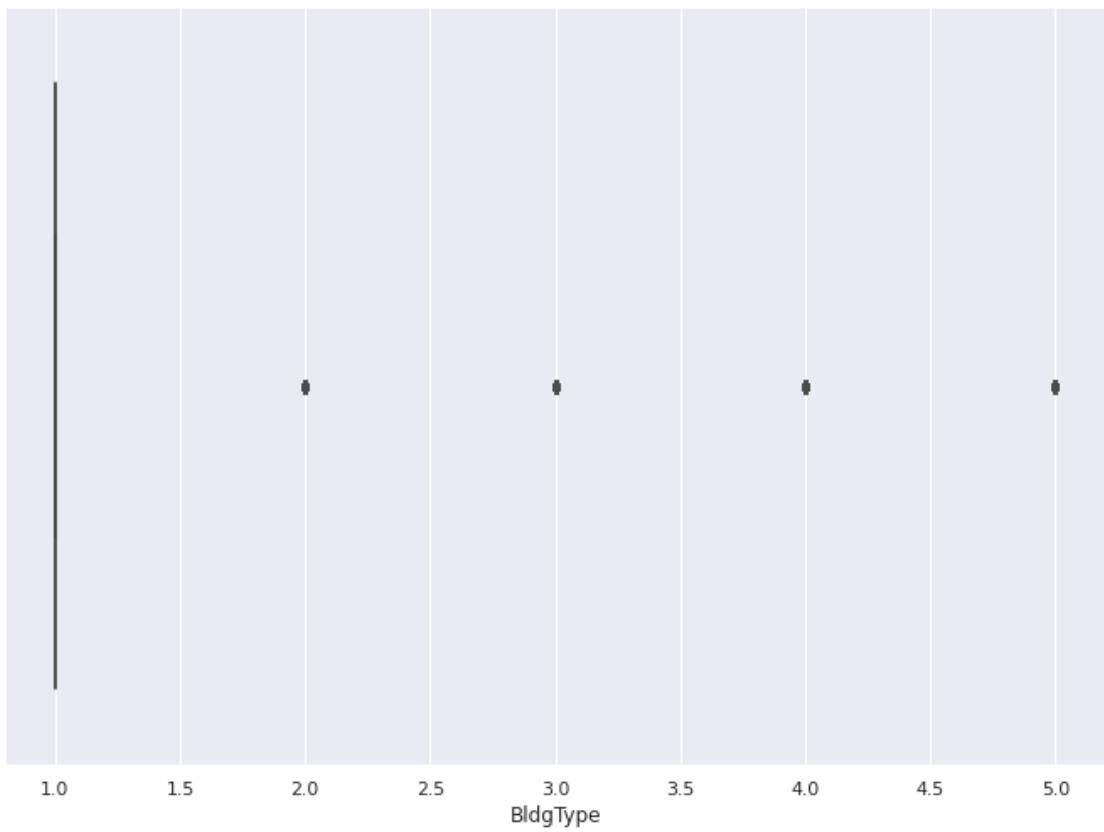


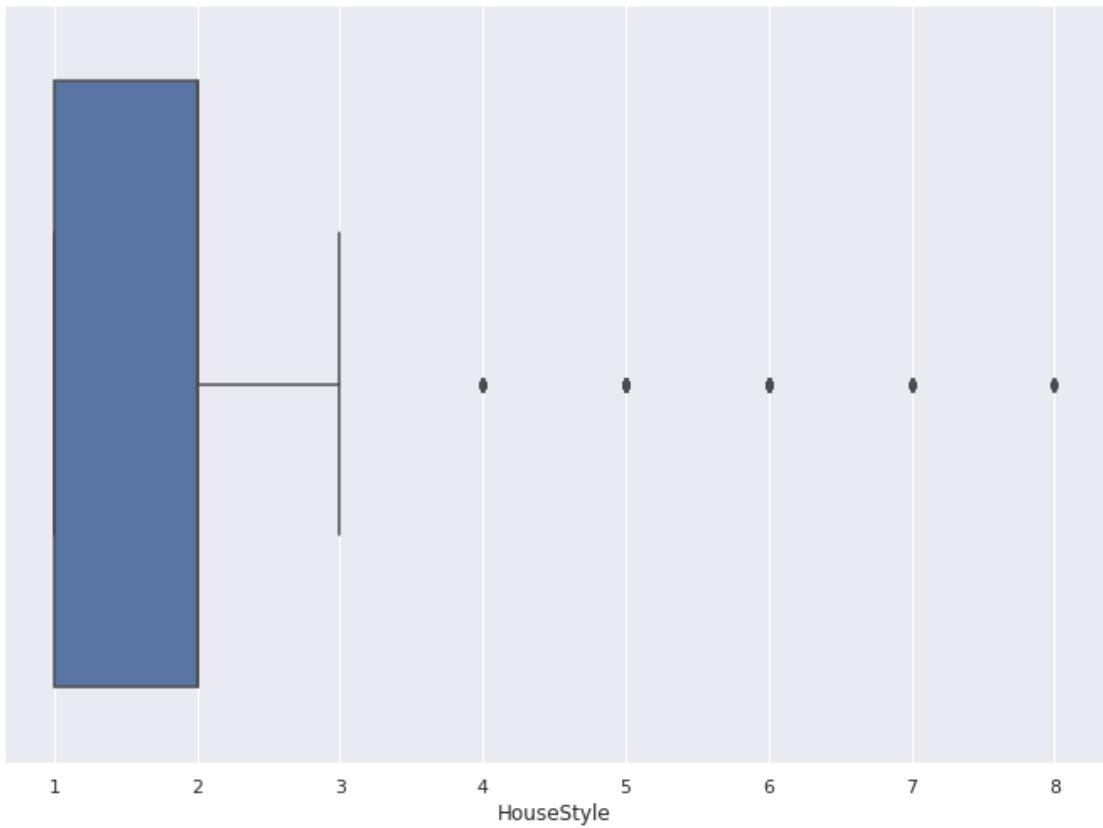


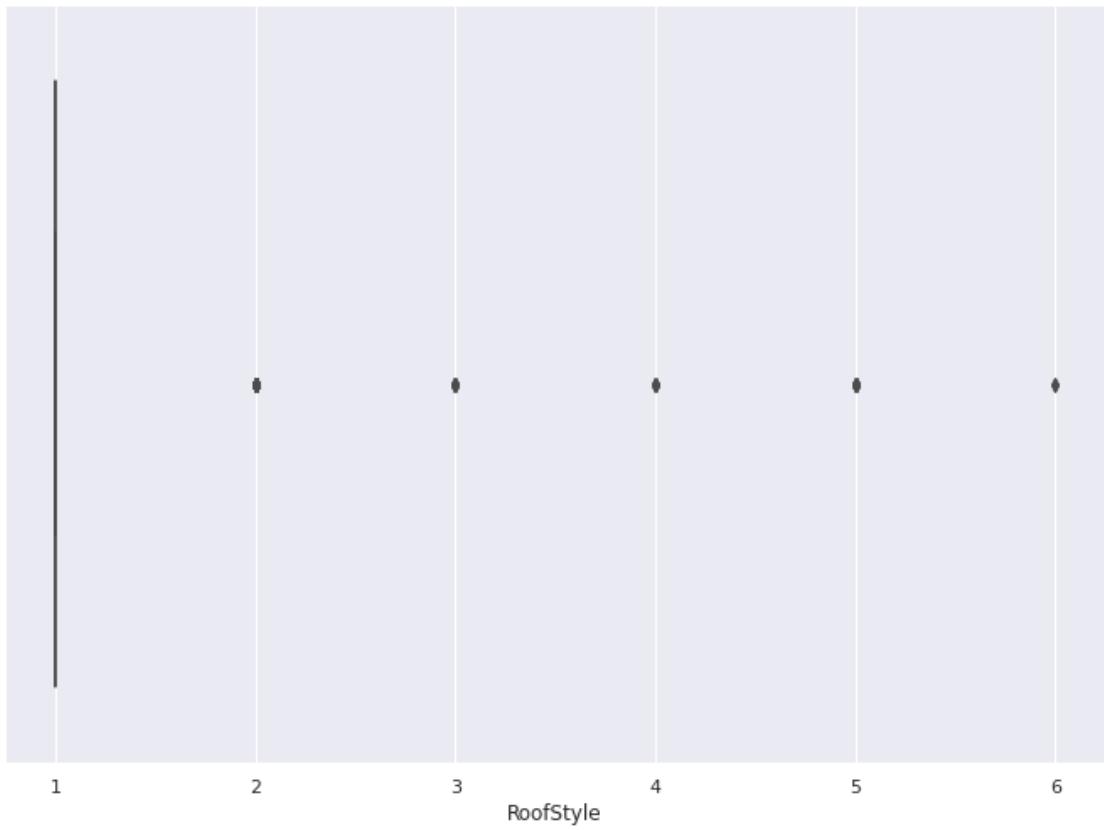


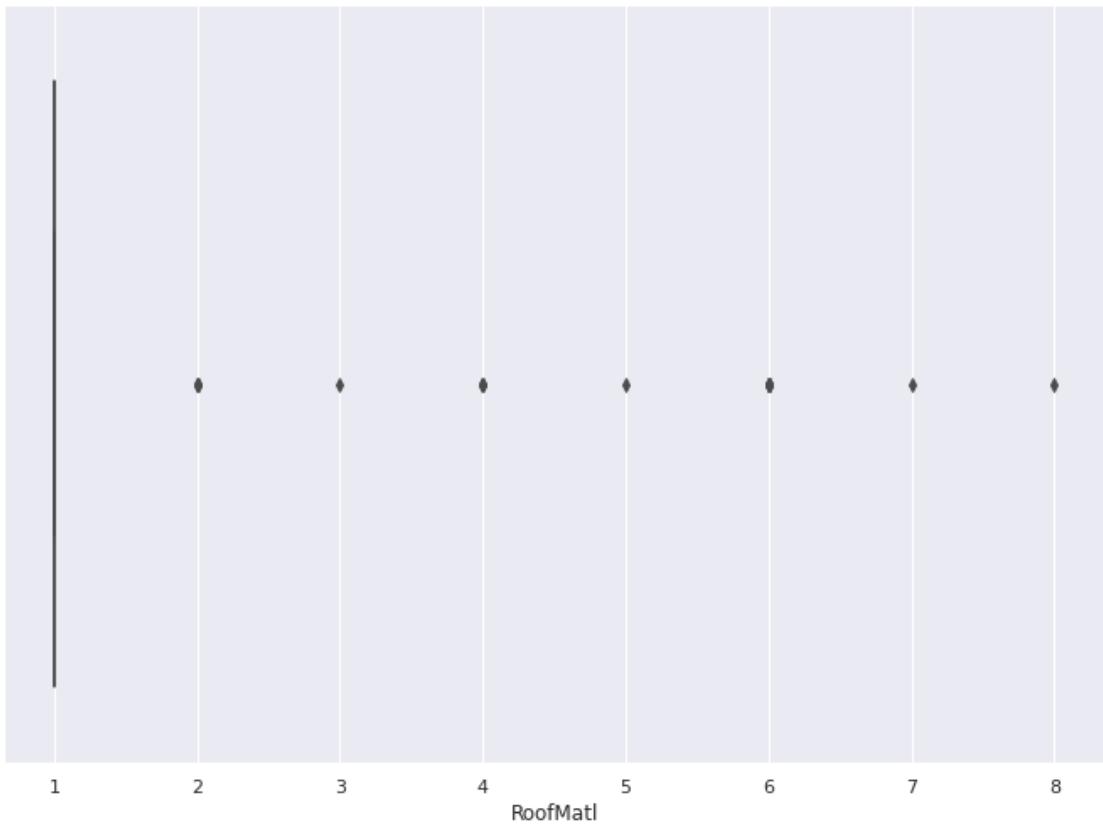


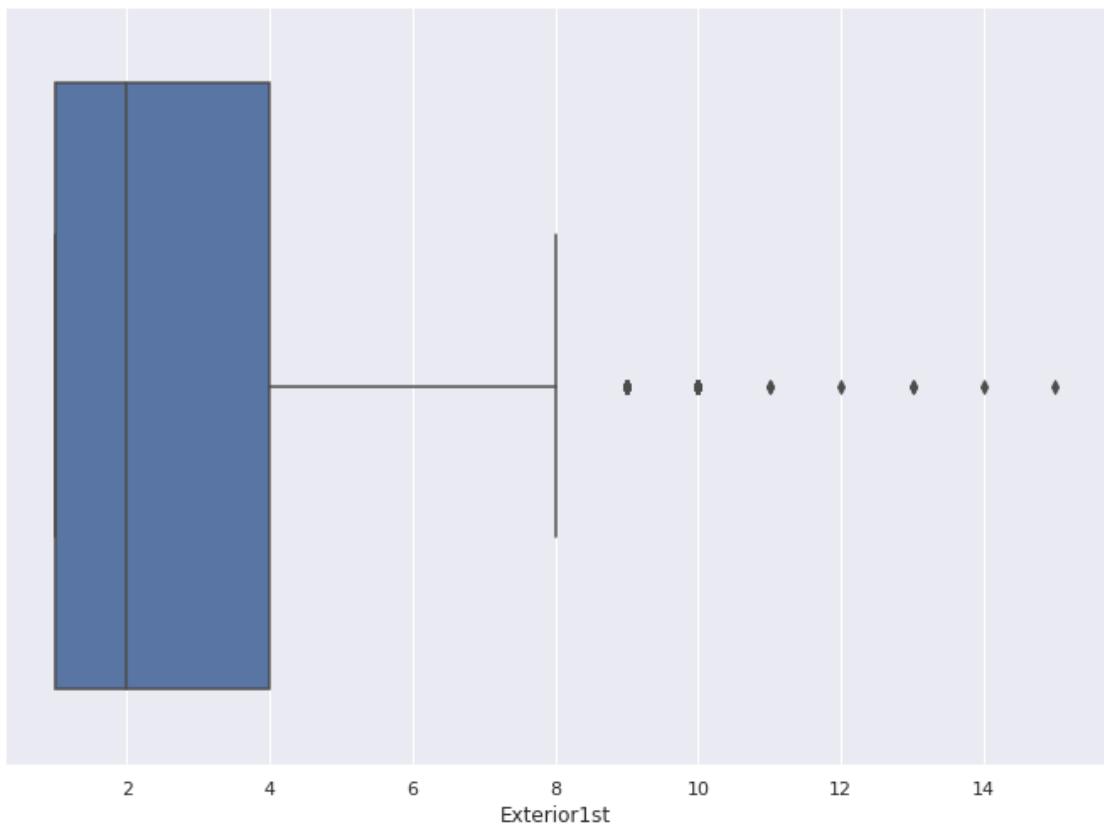


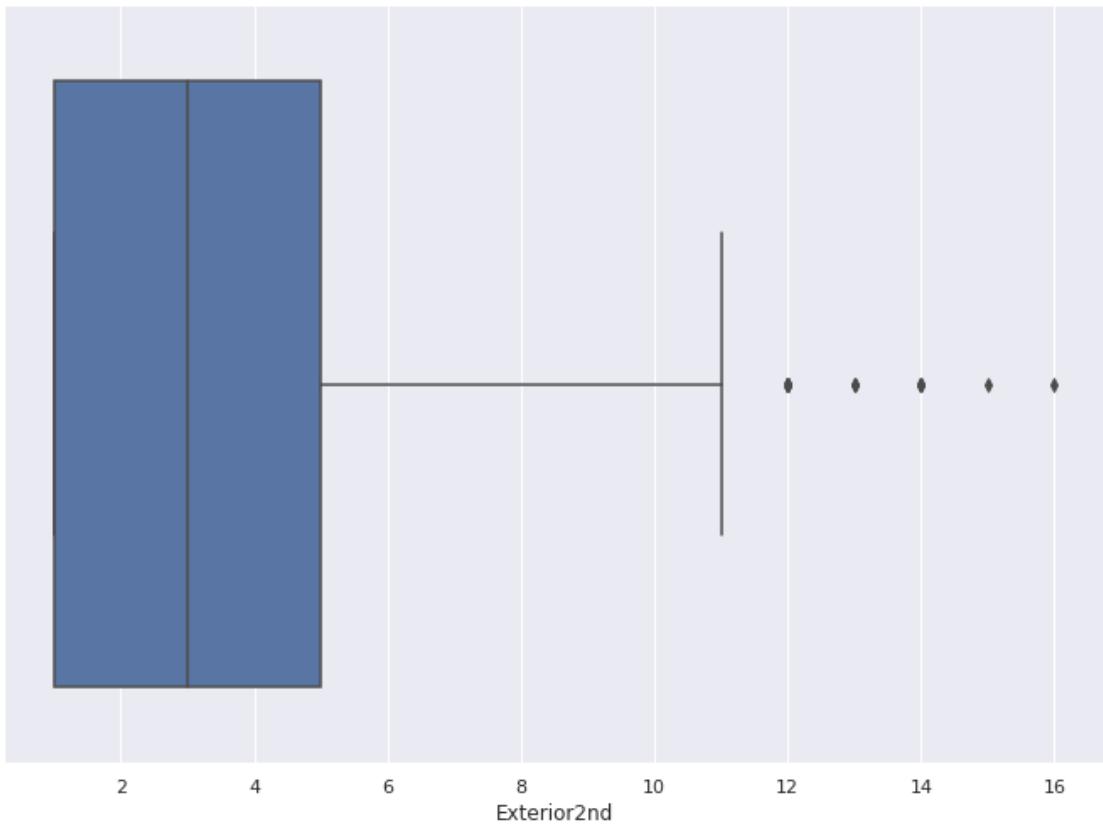


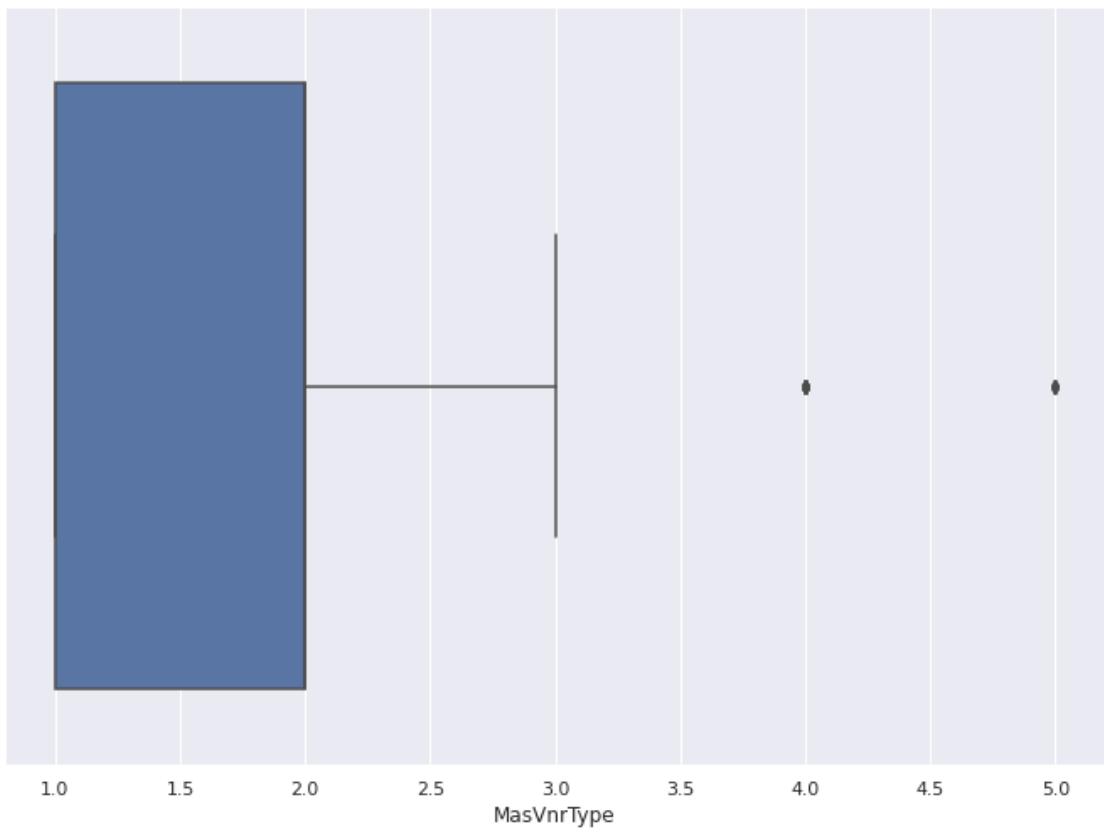


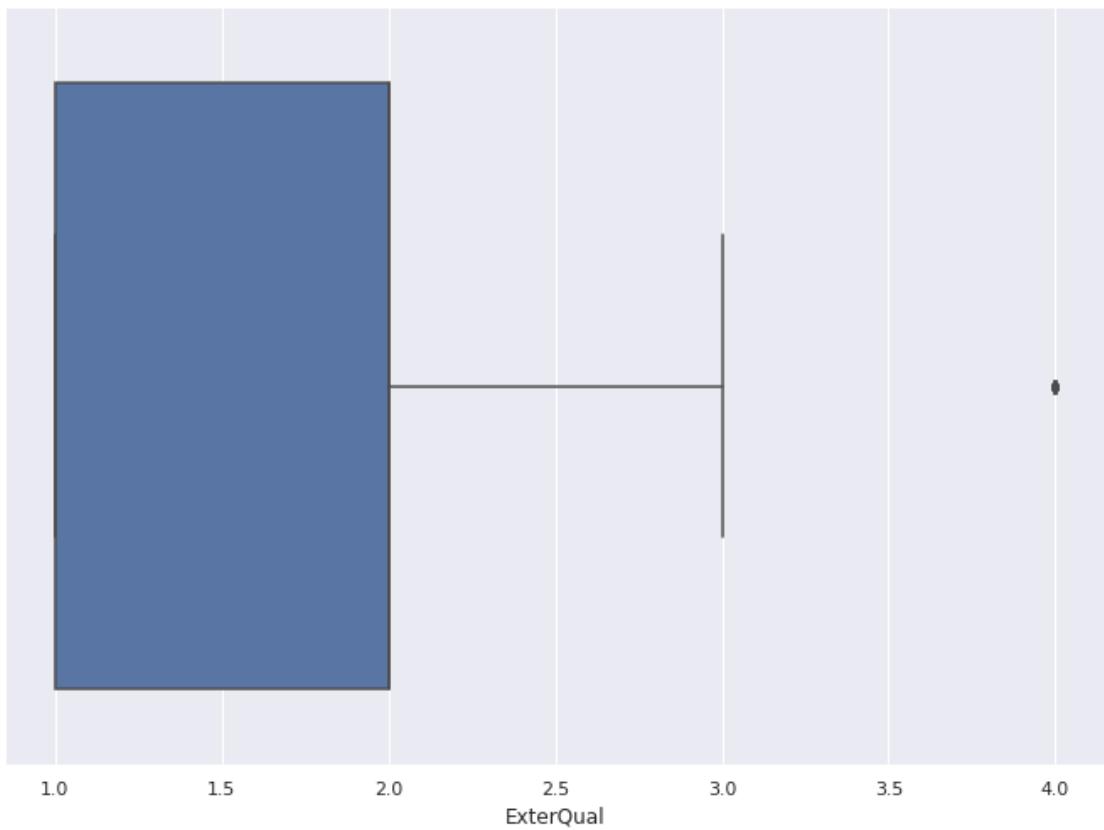


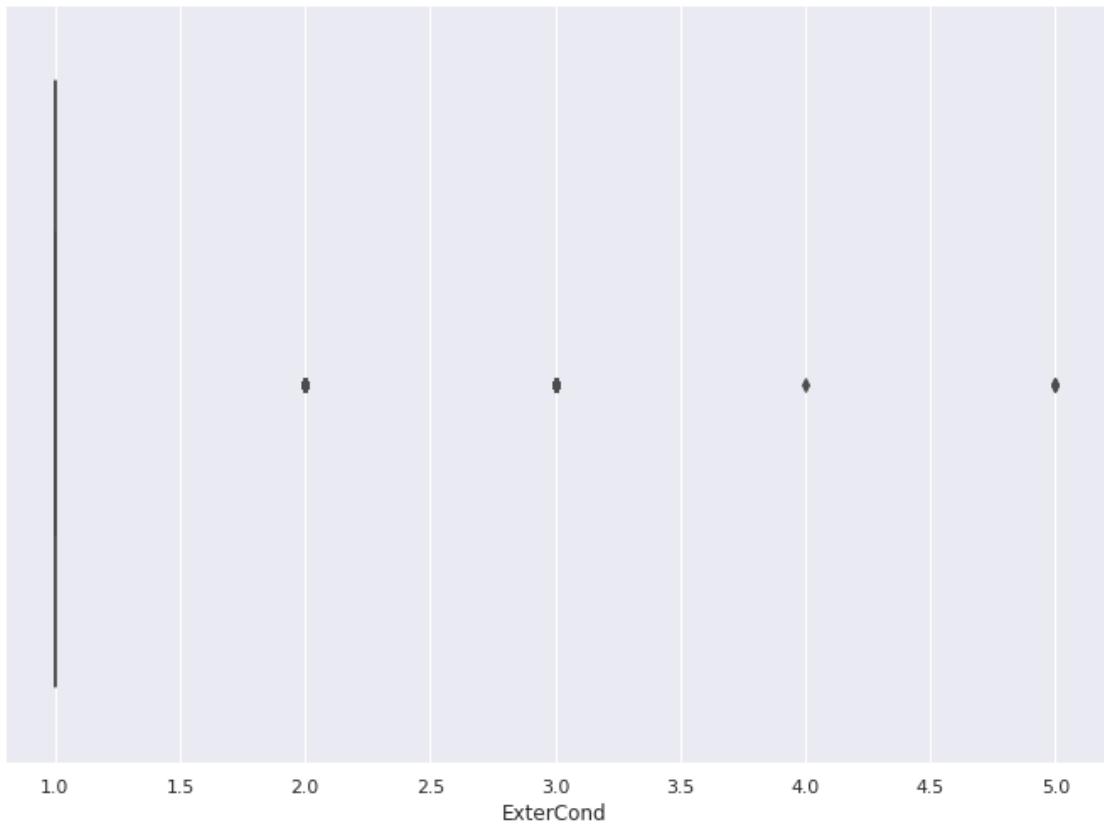


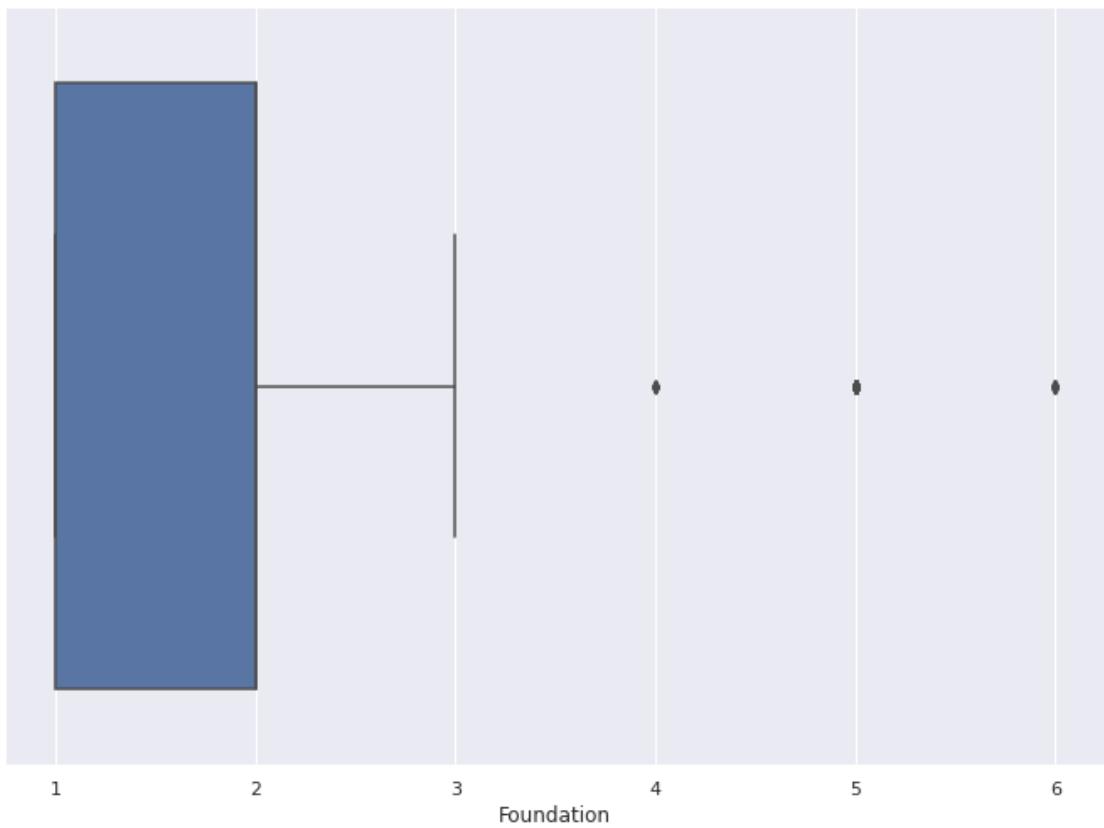


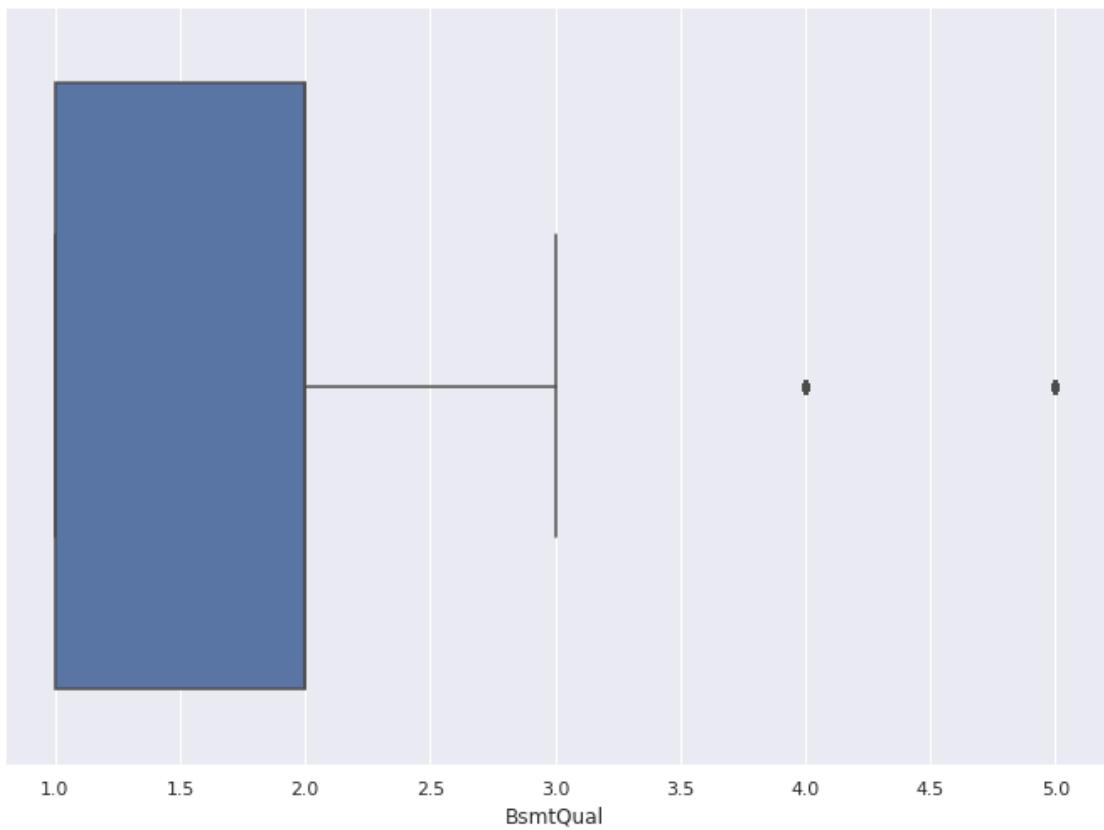


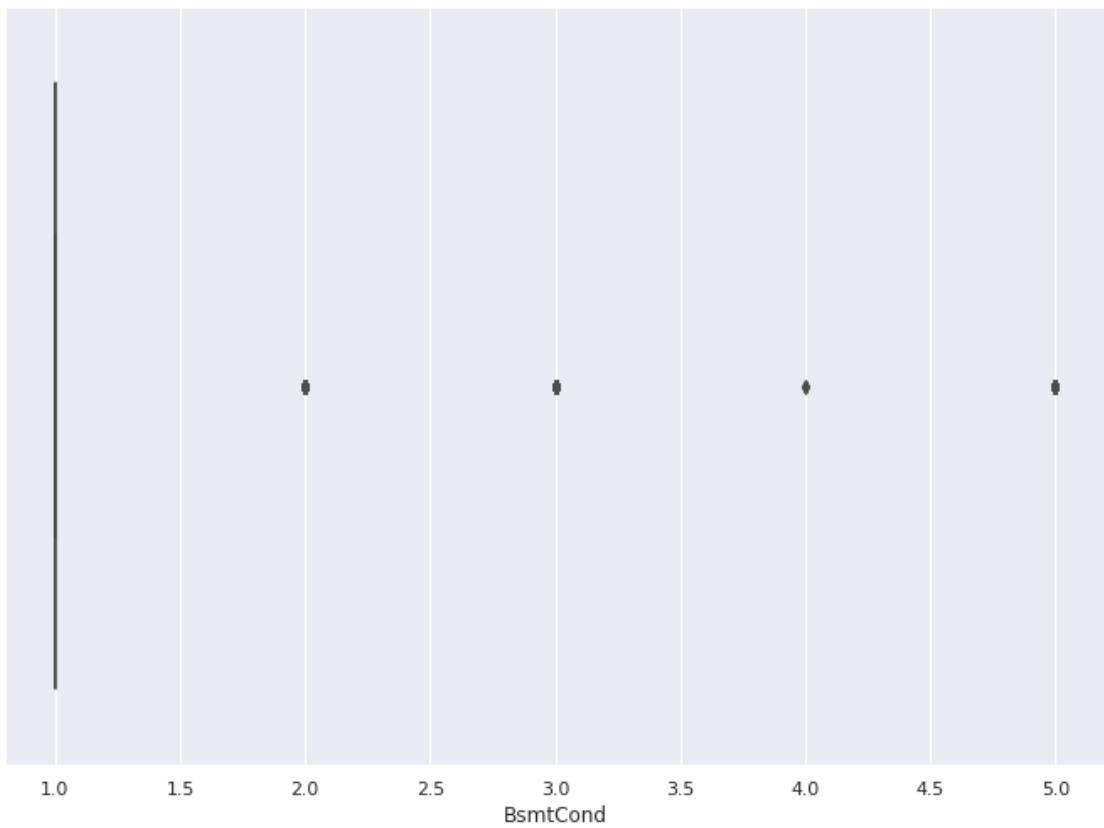


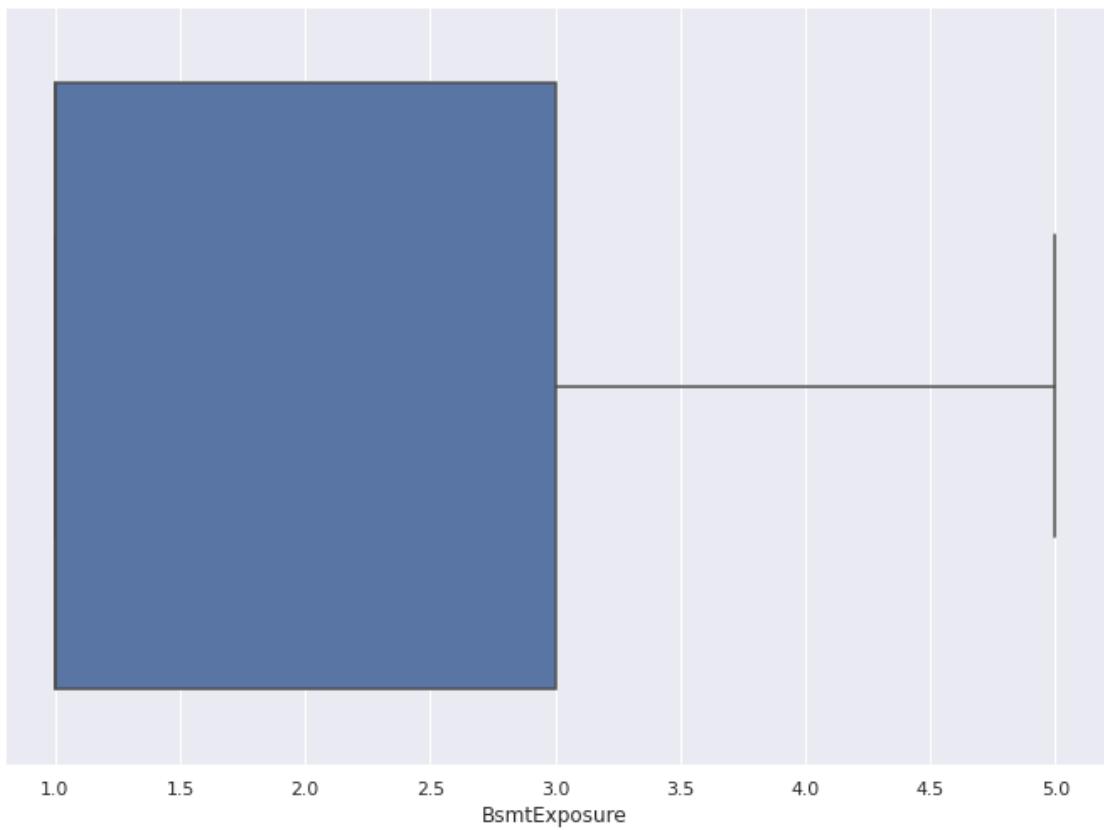


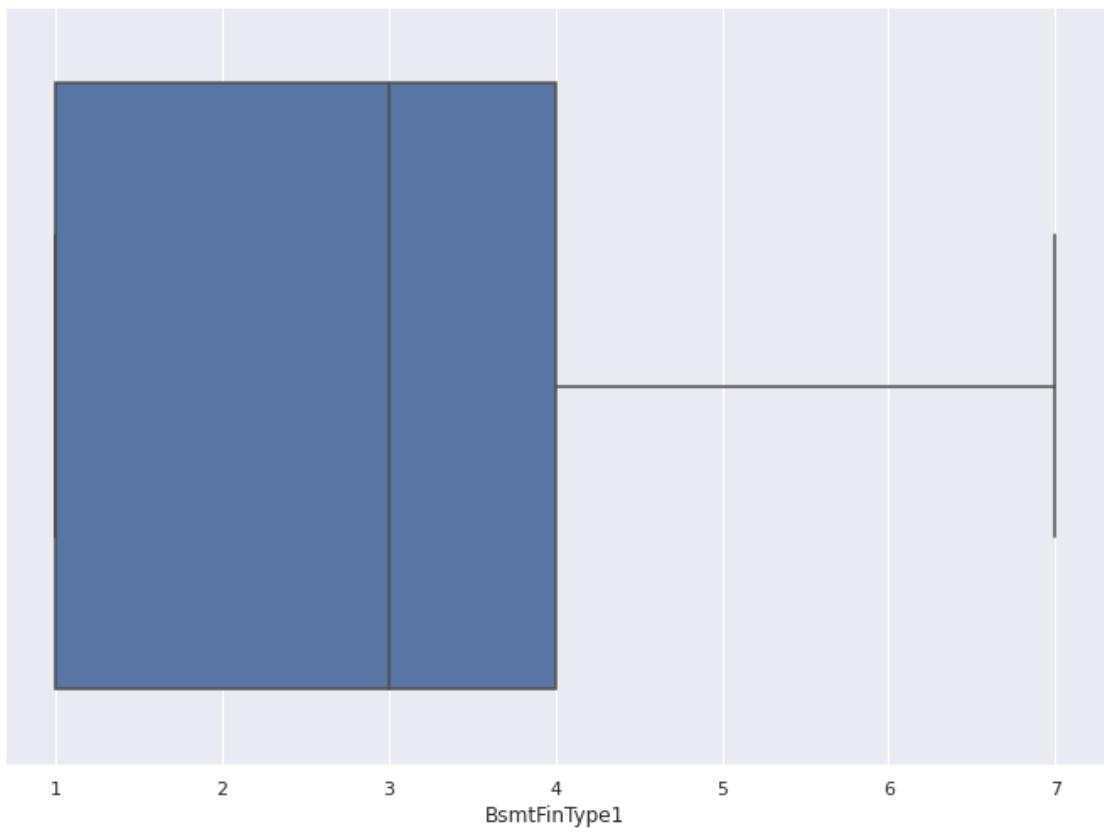


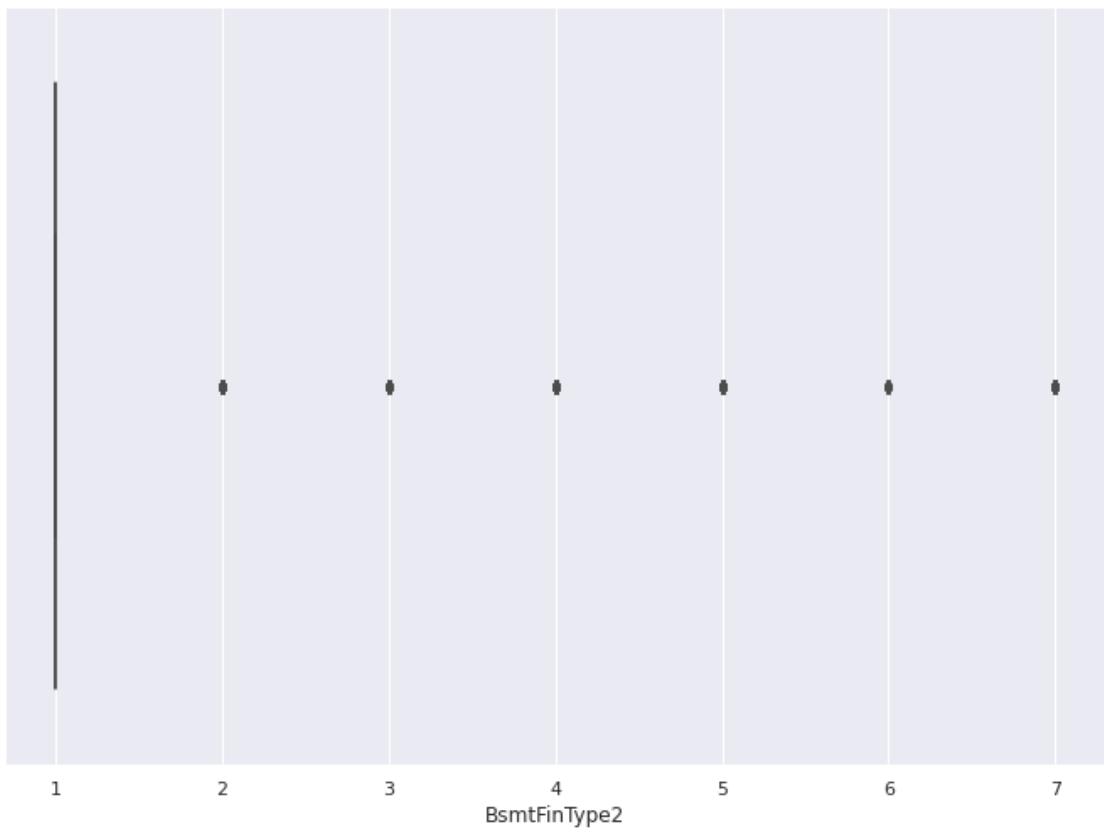


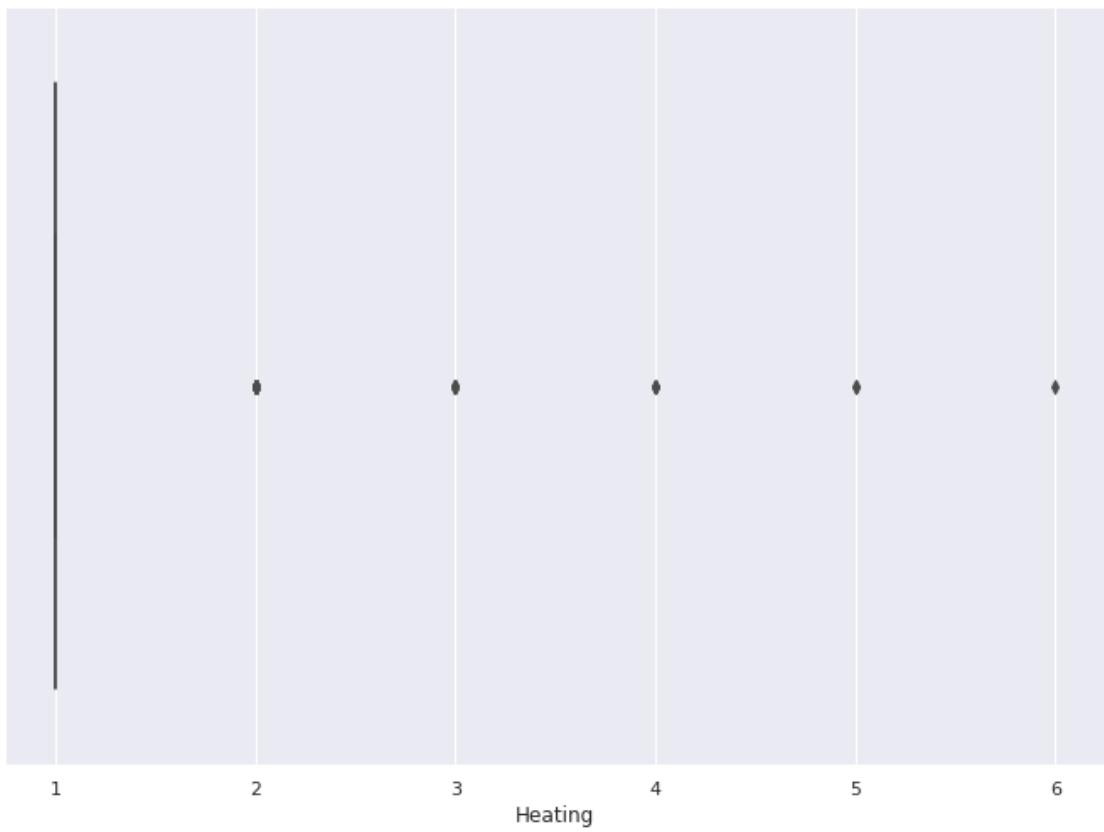


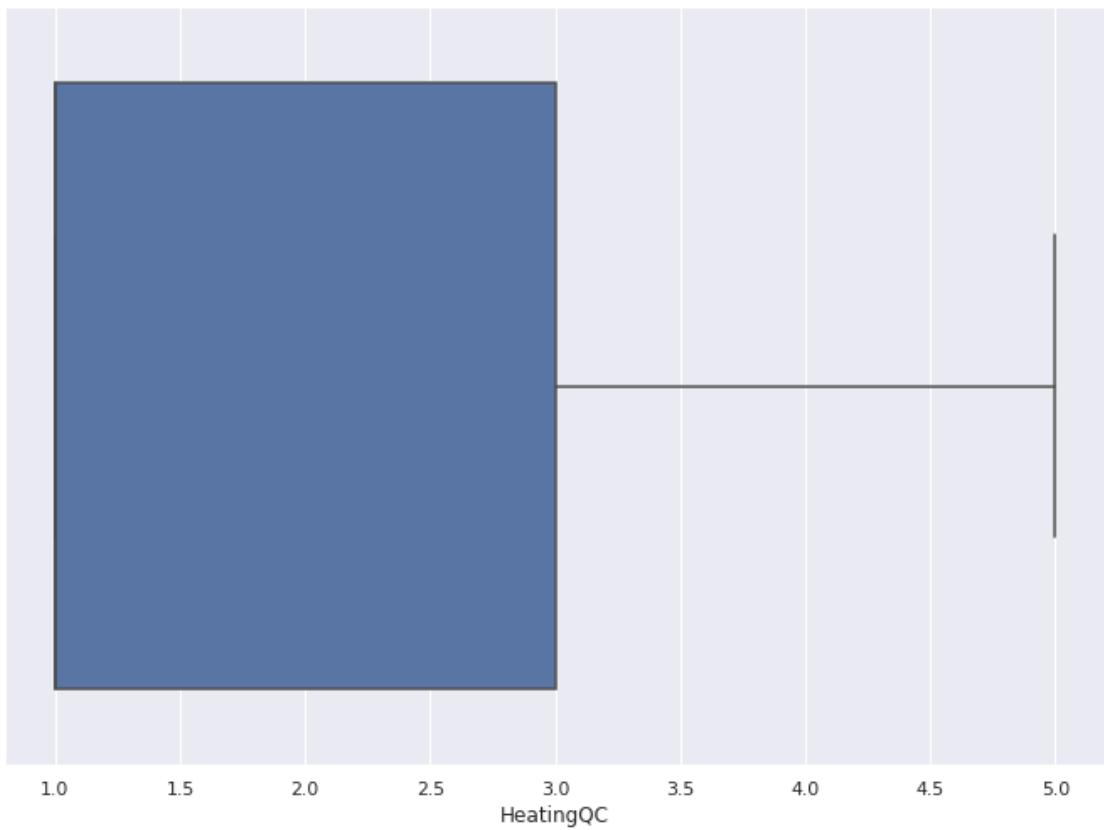


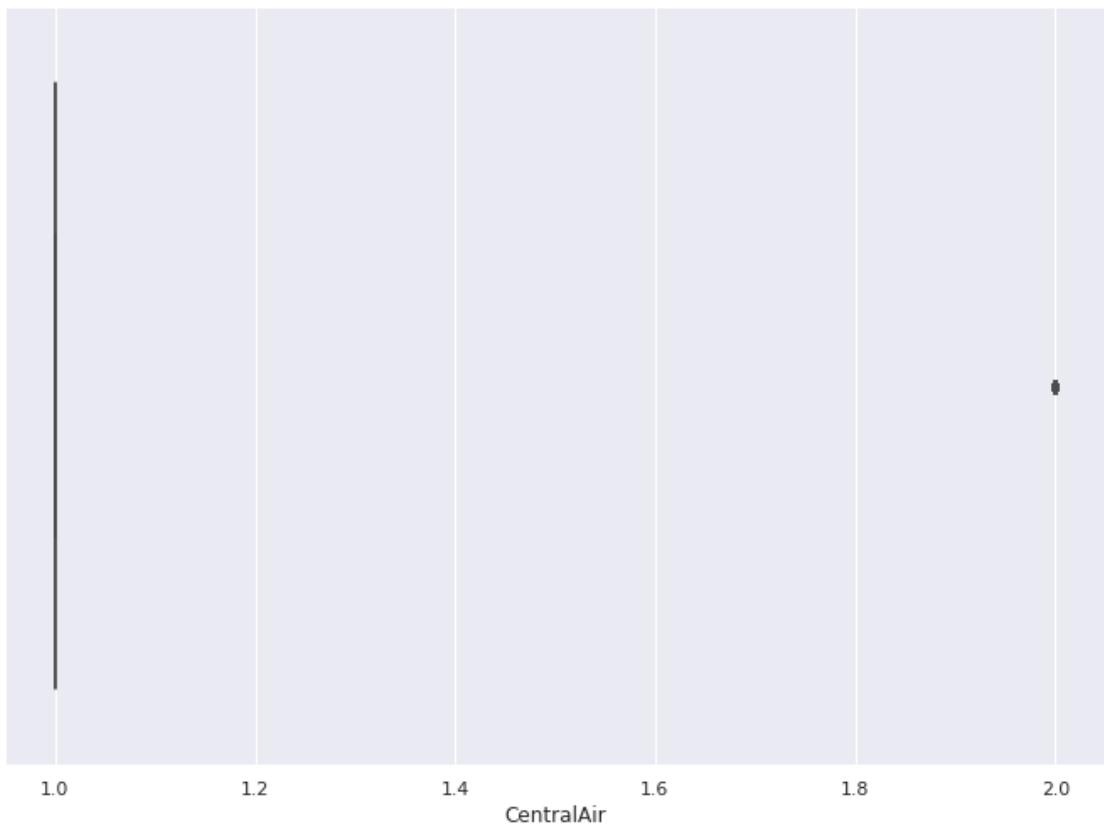


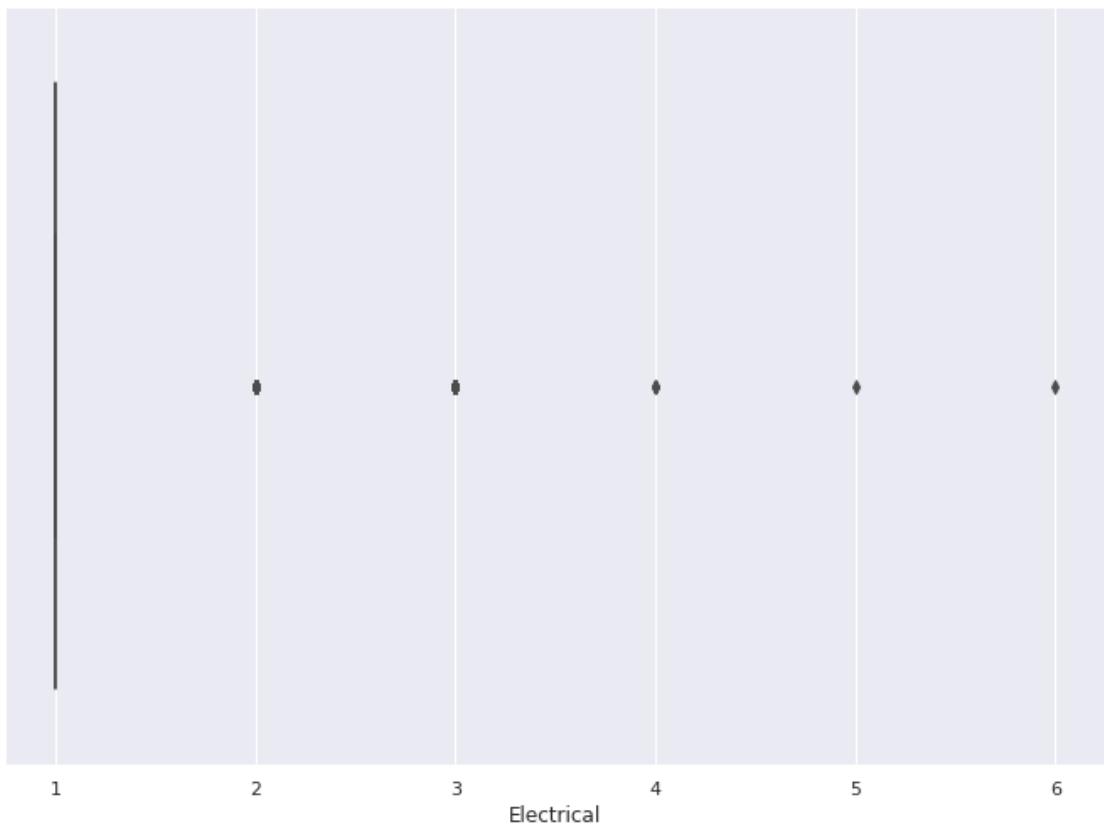


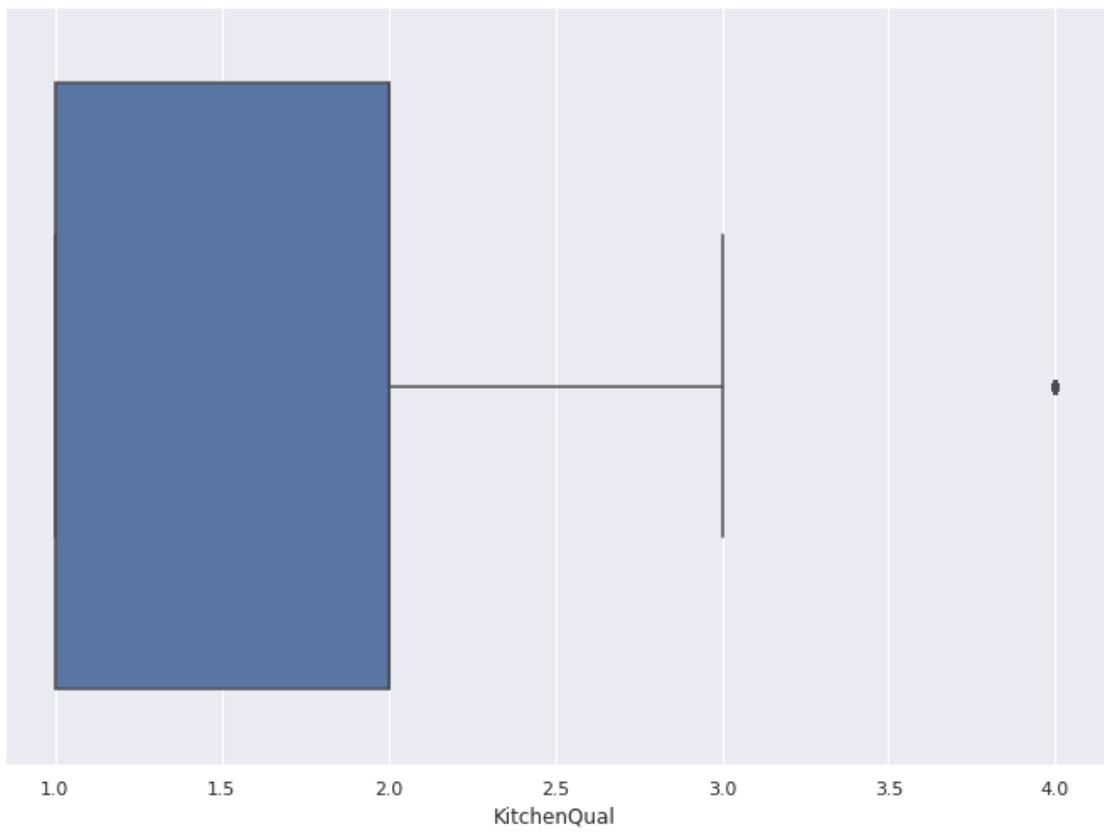


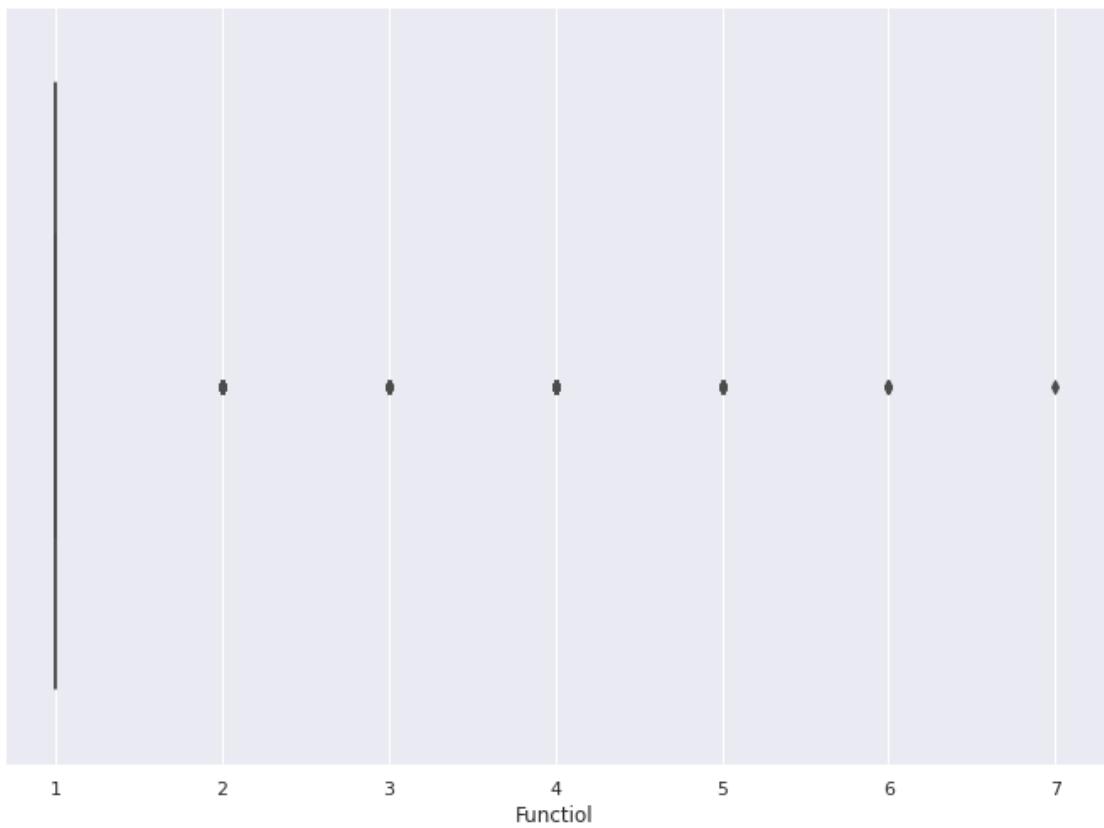


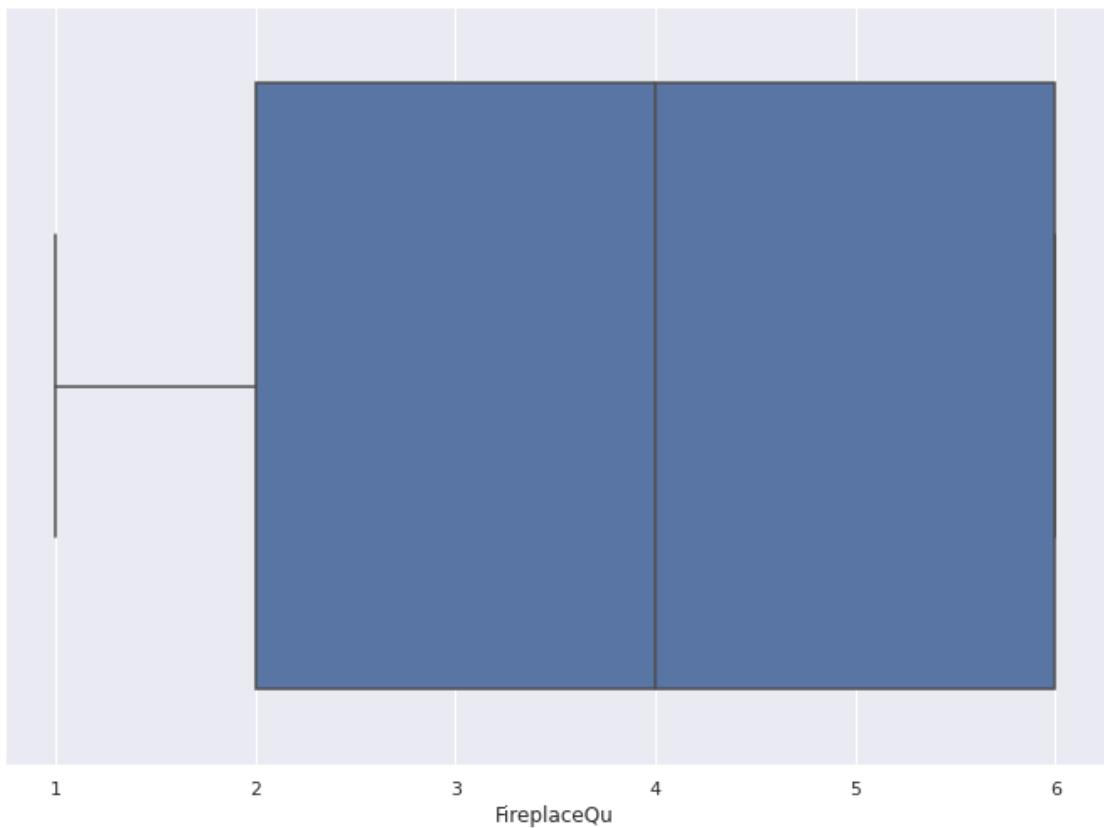


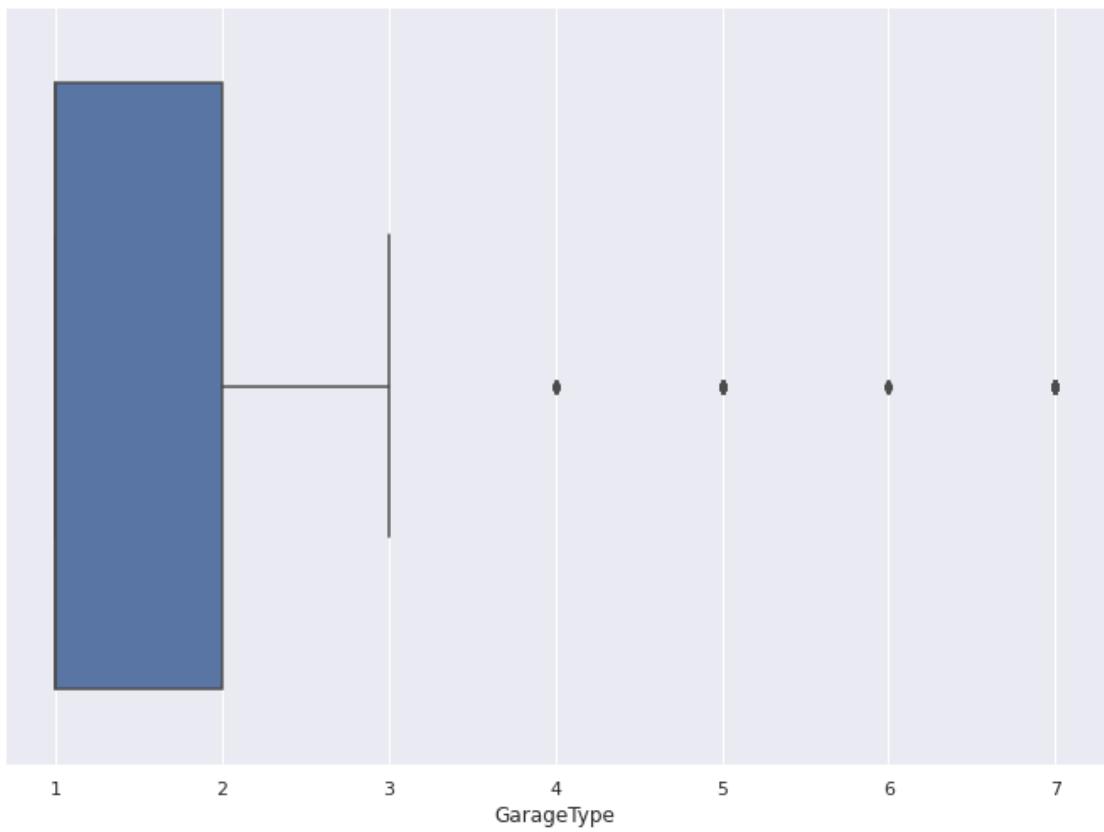


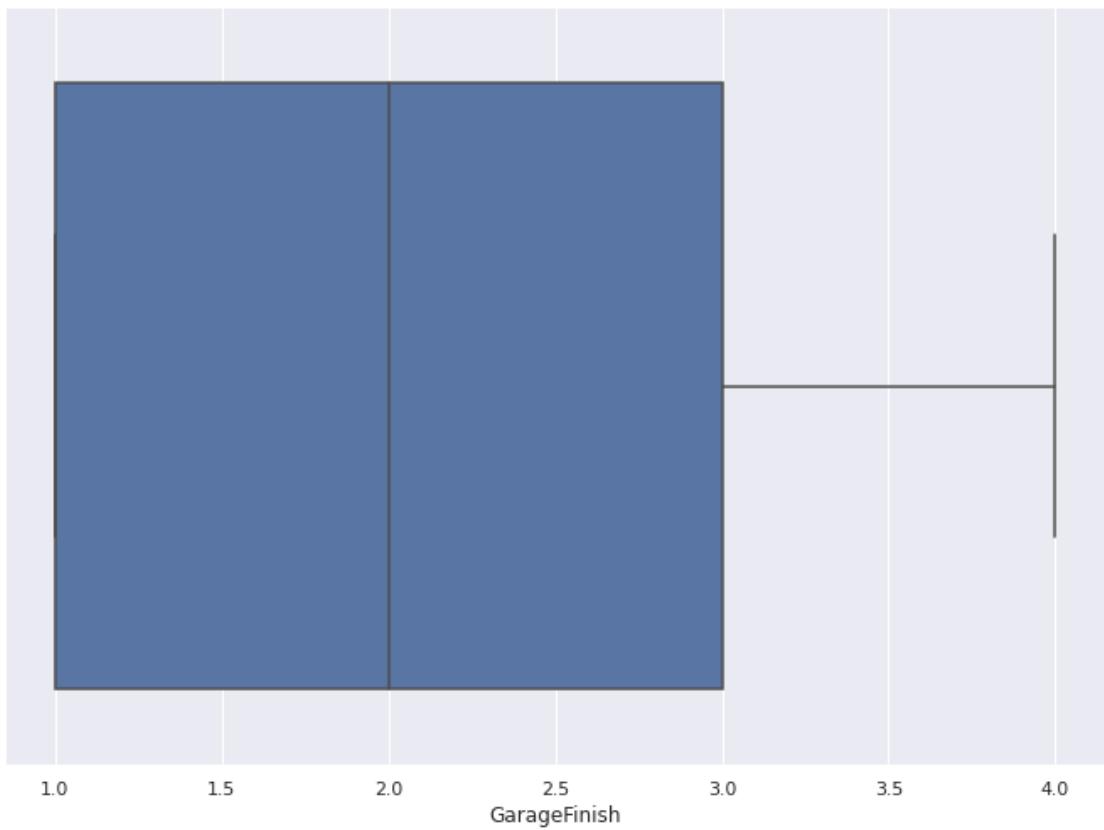


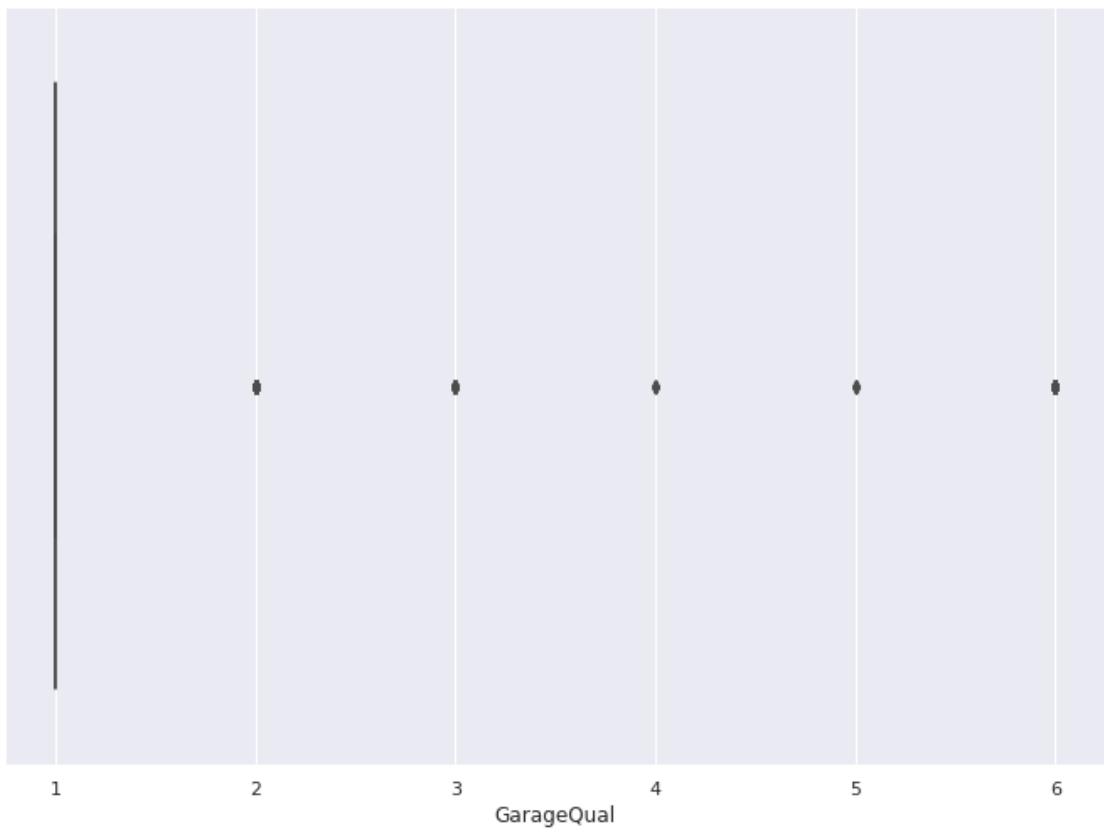


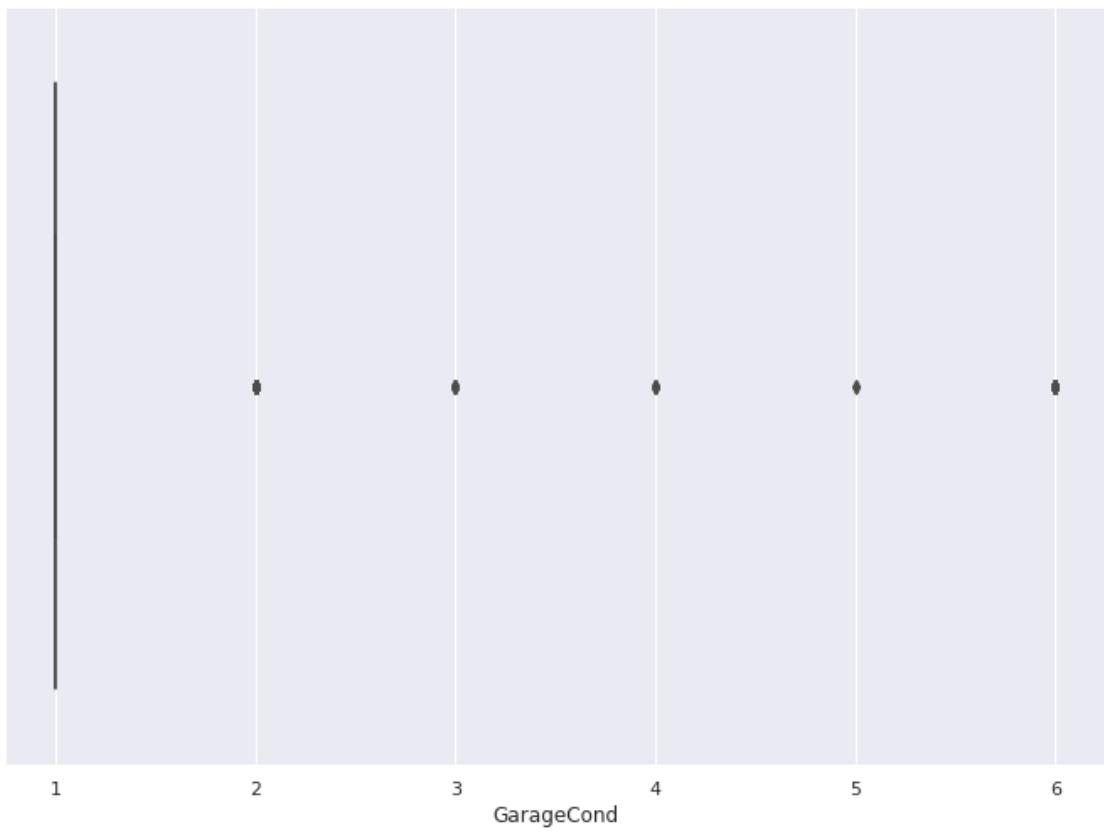


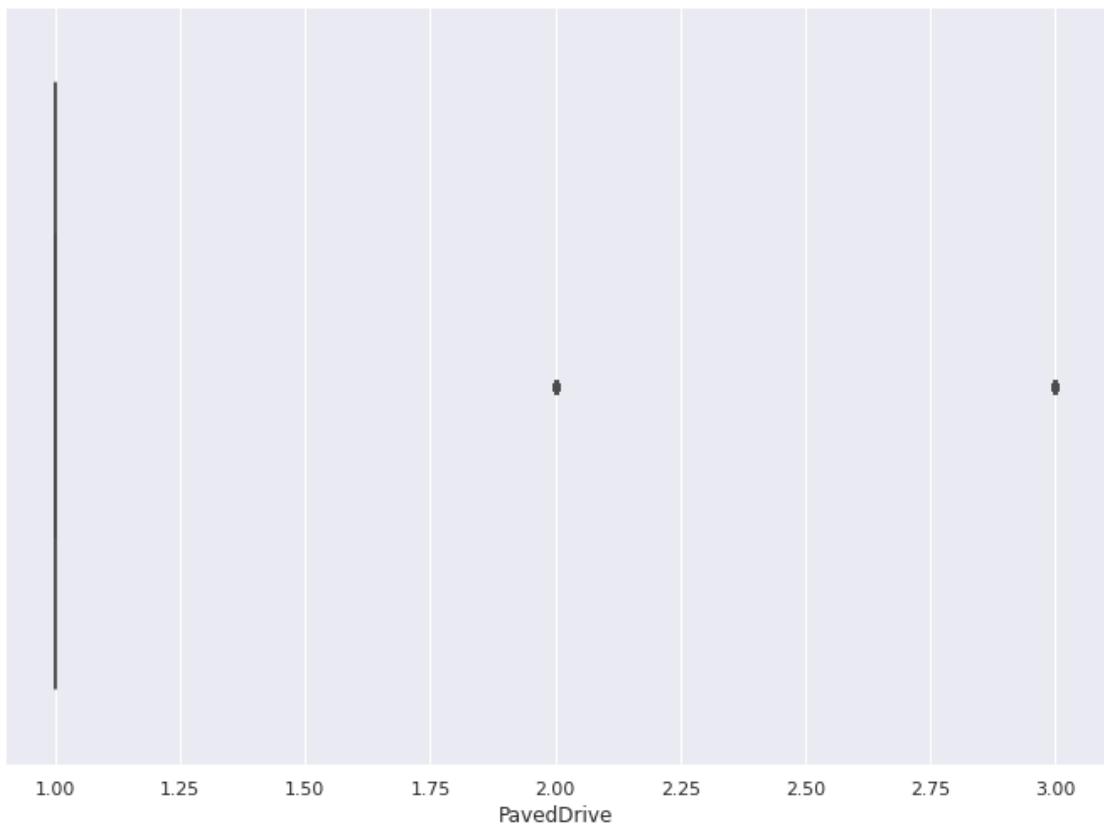


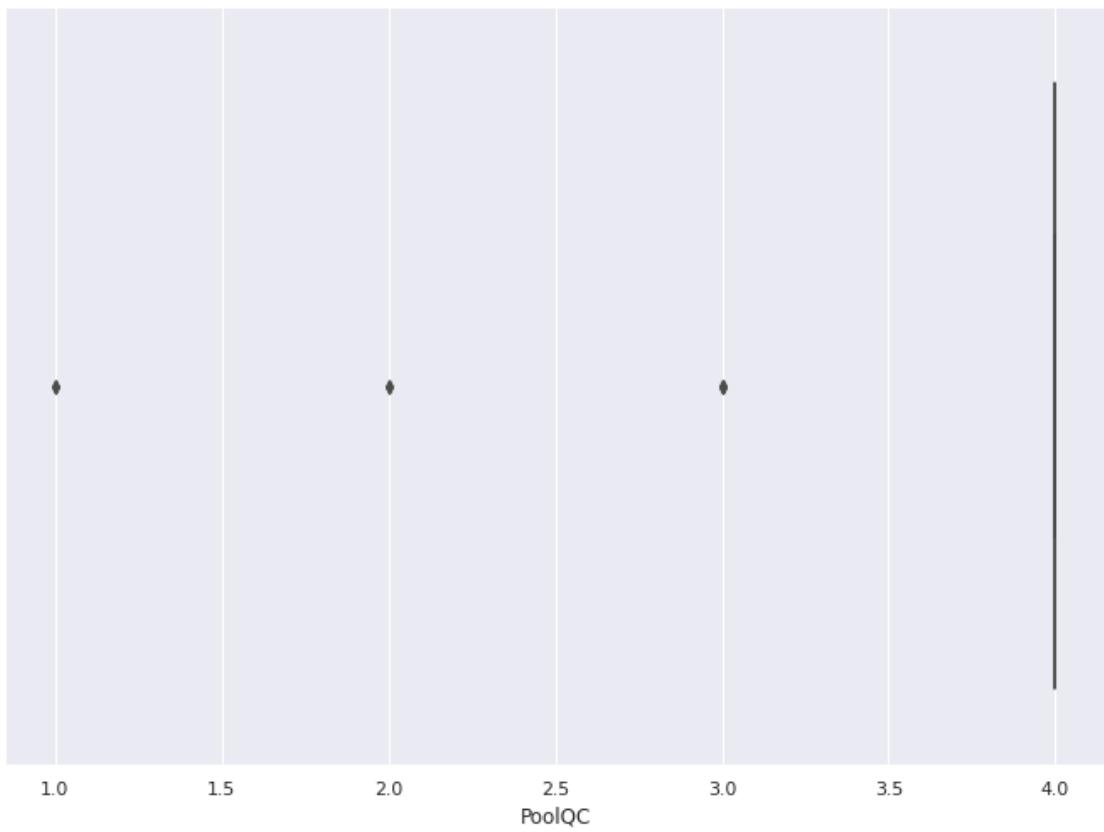


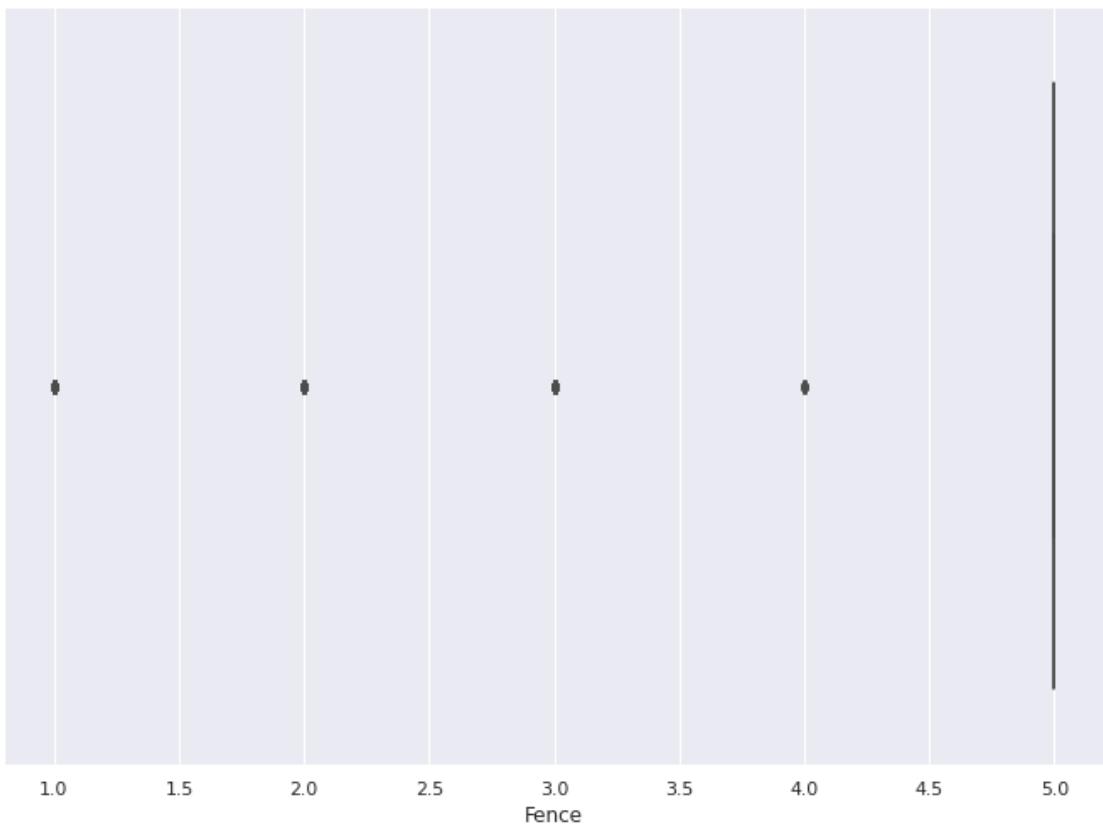


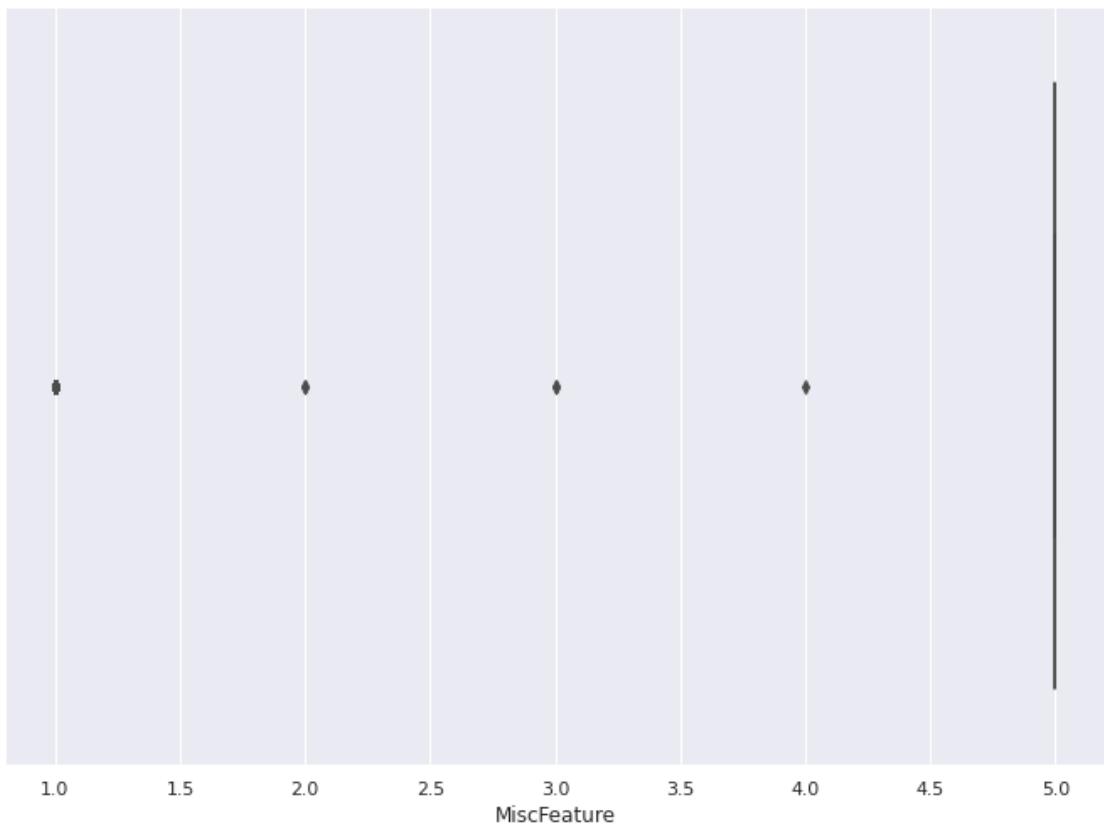


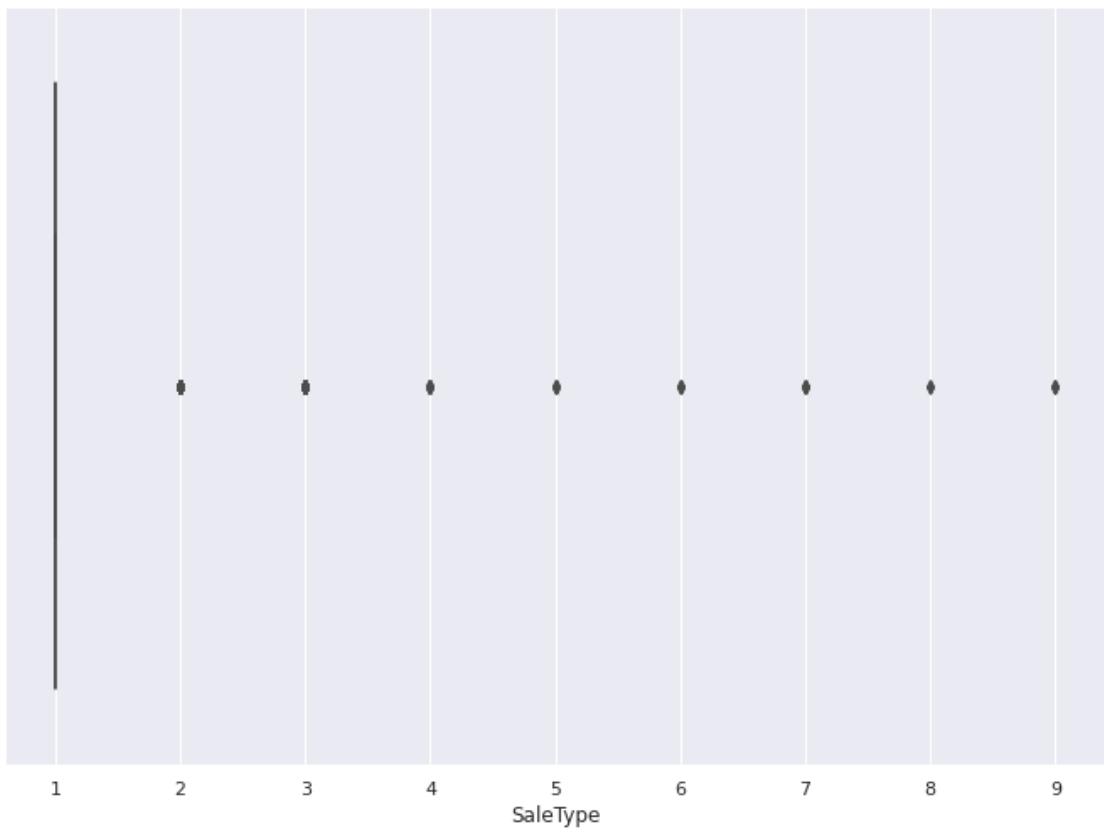


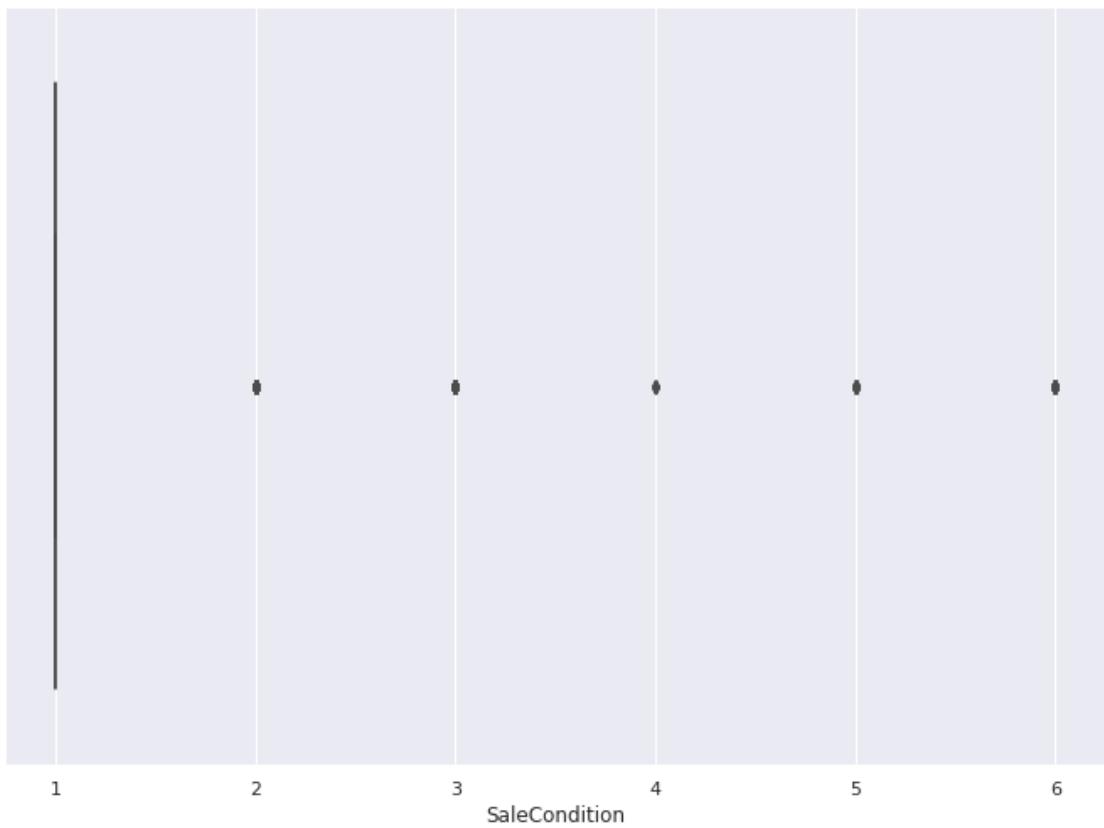


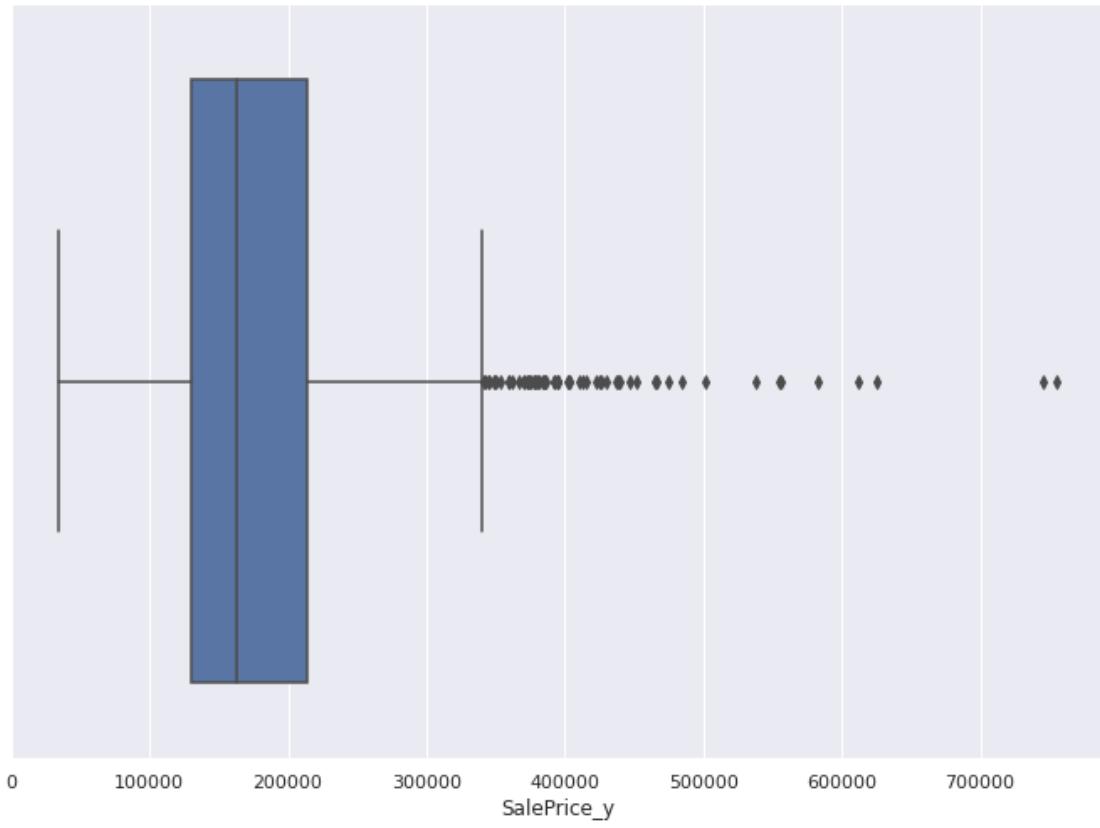












```
[1]: !pip install nbconvert
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: nbconvert in /usr/local/lib/python3.7/site-
packages (5.6.1)
Requirement already satisfied: jupyter-core in /usr/local/lib/python3.7/site-
packages (from nbconvert) (4.6.3)
Requirement already satisfied: defusedxml in /usr/local/lib/python3.7/site-
packages (from nbconvert) (0.6.0)
Requirement already satisfied: pandocfilters>=1.4.1 in
/usr/local/lib/python3.7/site-packages (from nbconvert) (1.4.2)
Requirement already satisfied: nbformat>=4.4 in /usr/local/lib/python3.7/site-
packages (from nbconvert) (5.0.5)
Requirement already satisfied: entrypoints>=0.2.2 in
/usr/local/lib/python3.7/site-packages (from nbconvert) (0.3)
Requirement already satisfied: testpath in /usr/local/lib/python3.7/site-
packages (from nbconvert) (0.4.4)
Requirement already satisfied: mistune<2,>=0.8.1 in
/usr/local/lib/python3.7/site-packages (from nbconvert) (0.8.4)
Requirement already satisfied: bleach in /usr/local/lib/python3.7/site-packages
(from nbconvert) (3.1.4)
```

```

Requirement already satisfied: traitlets>=4.2 in /usr/local/lib/python3.7/site-
packages (from nbconvert) (5.1.1)
Requirement already satisfied: pygments in /usr/local/lib/python3.7/site-
packages (from nbconvert) (2.6.1)
Requirement already satisfied: jinja2>=2.4 in /usr/local/lib/python3.7/site-
packages (from nbconvert) (2.11.1)
Requirement already satisfied: MarkupSafe>=0.23 in
/usr/local/lib/python3.7/site-packages (from jinja2>=2.4->nbconvert) (1.1.1)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in
/usr/local/lib/python3.7/site-packages (from nbformat>=4.4->nbconvert) (4.4.0)
Requirement already satisfied: ipython-genutils in
/usr/local/lib/python3.7/site-packages (from nbformat>=4.4->nbconvert) (0.2.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.7/site-
packages (from bleach->nbconvert) (0.5.1)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.7/site-
packages (from bleach->nbconvert) (1.14.0)
Requirement already satisfied: importlib-resources>=1.4.0 in
/usr/local/lib/python3.7/site-packages (from
jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (5.4.0)
Requirement already satisfied: pyrsistent!=0.17.0,!0.17.1,!0.17.2,>=0.14.0 in
/usr/local/lib/python3.7/site-packages (from
jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (0.16.0)
Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.7/site-
packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (19.3.0)
Requirement already satisfied: importlib-metadata in
/usr/local/lib/python3.7/site-packages (from
jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (1.6.0)
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.7/site-packages (from
jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (4.0.1)
Requirement already satisfied: zipp>=3.1.0 in /usr/local/lib/python3.7/site-
packages (from importlib-
resources>=1.4.0->jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (3.1.0)
WARNING: You are using pip version 22.0.3; however, version 23.2 is
available.

You should consider upgrading via the '/usr/local/bin/python3 -m pip install
--upgrade pip' command.

```

[2]: jupyter nbconvert --to docx Feature_Engineering.ipynb

```

File "<ipython-input-2-d4a150c485ec>", line 1
jupyter nbconvert --to docx Feature_Engineering.ipynb
^
SyntaxError: invalid syntax

```

[]: