project1

June 30, 2022

1 Project 1: Feature Engineering

Aurore Prevot

1.1 1- Understanding the dataset

- Importation of the librairies
- Loading of the dataset
- Visualization of the first 5 lines of the dataset

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


data = pd.read_csv("PEP1.csv")
data.head()
```

[1]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	…	PoolArea	PoolQC	Fence	MiscFeature	MiscVal	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	1	60	RL	65.0	8450	Pave	NaN	Reg	…	0	NaN	NaN	NaN	0	2	2008	WD	Normal	208500
1	2	20	RL	80.0	9600	Pave	NaN	Reg	…	0	NaN	NaN	NaN	0	5	2007	WD	Normal	181500
2	3	60	RL	68.0	11250	Pave	NaN	IR1	…	0	NaN	NaN	NaN	0	9	2008	WD	Normal	223500
3	4	70	RL	60.0	9550	Pave	NaN	IR1	…	0	NaN	NaN	NaN	0	2	2006	WD	Abnorml	140000
4	5	60	RL	84.0	14260	Pave	NaN	IR1	…	0	NaN	NaN	NaN	0	12	2008	WD	Normal	250000

```
2   2008        WD         Normal     223500
3   2006        WD        Abnorml     140000
4   2008        WD         Normal     250000

[5 rows x 81 columns]
```

- Visualization of the last 5 lines of the dataset

[2]:
```
data.tail()
```

[2]:
```
         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]
```

- Visualization of the shape of the data (1-a)

[3]:
```
data.shape
```

[3]:
```
(1460, 81)
```

Observation: the column ”Id” only refers to the number of the row.

We can use this column as index for the dataframe.

- Loading the dataset with the ”Id” column as index
- Visualization of the first 5 lines of the dataset

[4]:
```
data = pd.read_csv("PEP1.csv", index_col="Id")
data.head()
```

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

   LandContour Utilities LotConfig  … PoolArea PoolQC Fence MiscFeature  \
Id                                  …
1          Lvl    AllPub    Inside  …        0    NaN   NaN         NaN
2          Lvl    AllPub       FR2  …        0    NaN   NaN         NaN
3          Lvl    AllPub    Inside  …        0    NaN   NaN         NaN
4          Lvl    AllPub    Corner  …        0    NaN   NaN         NaN
5          Lvl    AllPub       FR2  …        0    NaN   NaN         NaN

   MiscVal MoSold  YrSold  SaleType  SaleCondition  SalePrice
Id
1        0      2    2008        WD         Normal     208500
2        0      5    2007        WD         Normal     181500
3        0      9    2008        WD         Normal     223500
4        0      2    2006        WD        Abnorml     140000
5        0     12    2008        WD         Normal     250000

[5 rows x 80 columns]
```

- Visualization of the last 5 lines of the dataset

[5]:
```
data.tail()
```

[5]:
```
      MSSubClass MSZoning  LotFrontage  LotArea Street Alley LotShape  \
Id
1456          60       RL         62.0     7917   Pave   NaN      Reg
1457          20       RL         85.0    13175   Pave   NaN      Reg
1458          70       RL         66.0     9042   Pave   NaN      Reg
1459          20       RL         68.0     9717   Pave   NaN      Reg
1460          20       RL         75.0     9937   Pave   NaN      Reg

     LandContour Utilities LotConfig  … PoolArea PoolQC  Fence MiscFeature  \
Id                                    …
1456         Lvl    AllPub    Inside  …        0    NaN    NaN         NaN
1457         Lvl    AllPub    Inside  …        0    NaN  MnPrv         NaN
1458         Lvl    AllPub    Inside  …        0    NaN  GdPrv        Shed
1459         Lvl    AllPub    Inside  …        0    NaN    NaN         NaN
1460         Lvl    AllPub    Inside  …        0    NaN    NaN         NaN

     MiscVal MoSold  YrSold  SaleType  SaleCondition  SalePrice
Id
```

```
1456       0      8    2007        WD         Normal     175000
1457       0      2    2010        WD         Normal     210000
1458    2500      5    2010        WD         Normal     266500
1459       0      4    2010        WD         Normal     142125
1460       0      6    2008        WD         Normal     147500

[5 rows x 80 columns]
```

- Visualization of the shape of the data (1-a)

[6]:
```
data.shape
```

[6]:
```
(1460, 80)
```

- Getting the info of the dataset

[7]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1460 entries, 1 to 1460
Data columns (total 80 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   MSSubClass     1460 non-null   int64
 1   MSZoning       1460 non-null   object
 2   LotFrontage    1201 non-null   float64
 3   LotArea        1460 non-null   int64
 4   Street         1460 non-null   object
 5   Alley          91 non-null     object
 6   LotShape       1460 non-null   object
 7   LandContour    1460 non-null   object
 8   Utilities      1460 non-null   object
 9   LotConfig      1460 non-null   object
 10  LandSlope      1460 non-null   object
 11  Neighborhood   1460 non-null   object
 12  Condition1     1460 non-null   object
 13  Condition2     1460 non-null   object
 14  BldgType       1460 non-null   object
 15  HouseStyle     1460 non-null   object
 16  OverallQual    1460 non-null   int64
 17  OverallCond    1460 non-null   int64
 18  YearBuilt      1460 non-null   int64
 19  YearRemodAdd   1460 non-null   int64
 20  RoofStyle      1460 non-null   object
 21  RoofMatl       1460 non-null   object
 22  Exterior1st    1460 non-null   object
 23  Exterior2nd    1460 non-null   object
 24  MasVnrType     1452 non-null   object
```

```
25  MasVnrArea     1452 non-null   float64
26  ExterQual      1460 non-null   object
27  ExterCond      1460 non-null   object
28  Foundation     1460 non-null   object
29  BsmtQual       1423 non-null   object
30  BsmtCond       1423 non-null   object
31  BsmtExposure   1422 non-null   object
32  BsmtFinType1   1423 non-null   object
33  BsmtFinSF1     1460 non-null   int64
34  BsmtFinType2   1422 non-null   object
35  BsmtFinSF2     1460 non-null   int64
36  BsmtUnfSF      1460 non-null   int64
37  TotalBsmtSF    1460 non-null   int64
38  Heating        1460 non-null   object
39  HeatingQC      1460 non-null   object
40  CentralAir     1460 non-null   object
41  Electrical     1459 non-null   object
42  1stFlrSF       1460 non-null   int64
43  2ndFlrSF       1460 non-null   int64
44  LowQualFinSF   1460 non-null   int64
45  GrLivArea      1460 non-null   int64
46  BsmtFullBath   1460 non-null   int64
47  BsmtHalfBath   1460 non-null   int64
48  FullBath       1460 non-null   int64
49  HalfBath       1460 non-null   int64
50  BedroomAbvGr   1460 non-null   int64
51  KitchebvGr     1460 non-null   int64
52  KitchenQual    1460 non-null   object
53  TotRmsAbvGrd   1460 non-null   int64
54  Functiol       1460 non-null   object
55  Fireplaces     1460 non-null   int64
56  FireplaceQu    770 non-null    object
57  GarageType     1379 non-null   object
58  GarageYrBlt    1379 non-null   float64
59  GarageFinish   1379 non-null   object
60  GarageCars     1460 non-null   int64
61  GarageArea     1460 non-null   int64
62  GarageQual     1379 non-null   object
63  GarageCond     1379 non-null   object
64  PavedDrive     1460 non-null   object
65  WoodDeckSF     1460 non-null   int64
66  OpenPorchSF    1460 non-null   int64
67  EnclosedPorch  1460 non-null   int64
68  3SsnPorch      1460 non-null   int64
69  ScreenPorch    1460 non-null   int64
70  PoolArea       1460 non-null   int64
71  PoolQC         7 non-null      object
72  Fence          281 non-null    object
```

```
 73  MiscFeature    54 non-null     object
 74  MiscVal        1460 non-null   int64
 75  MoSold         1460 non-null   int64
 76  YrSold         1460 non-null   int64
 77  SaleType       1460 non-null   object
 78  SaleCondition  1460 non-null   object
 79  SalePrice      1460 non-null   int64
dtypes: float64(3), int64(34), object(43)
memory usage: 923.9+ KB
```

Observation: The columns "Alley", "PoolQC", and "MiscFeature" have few values. Some values are numbers and others are strings.

- Visualization of the statistical summary of the dataset

[8]:
```
data.describe()
```

[8]:

	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	\
count	1460.000000	1201.000000	1460.000000	1460.000000	1460.000000	
mean	56.897260	70.049958	10516.828082	6.099315	5.575342	
std	42.300571	24.284752	9981.264932	1.382997	1.112799	
min	20.000000	21.000000	1300.000000	1.000000	1.000000	
25%	20.000000	59.000000	7553.500000	5.000000	5.000000	
50%	50.000000	69.000000	9478.500000	6.000000	5.000000	
75%	70.000000	80.000000	11601.500000	7.000000	6.000000	
max	190.000000	313.000000	215245.000000	10.000000	9.000000	

	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	BsmtFinSF2	…	\
count	1460.000000	1460.000000	1452.000000	1460.000000	1460.000000	…	
mean	1971.267808	1984.865753	103.685262	443.639726	46.549315	…	
std	30.202904	20.645407	181.066207	456.098091	161.319273	…	
min	1872.000000	1950.000000	0.000000	0.000000	0.000000	…	
25%	1954.000000	1967.000000	0.000000	0.000000	0.000000	…	
50%	1973.000000	1994.000000	0.000000	383.500000	0.000000	…	
75%	2000.000000	2004.000000	166.000000	712.250000	0.000000	…	
max	2010.000000	2010.000000	1600.000000	5644.000000	1474.000000	…	

	WoodDeckSF	OpenPorchSF	EnclosedPorch	3SsnPorch	ScreenPorch	\
count	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	
mean	94.244521	46.660274	21.954110	3.409589	15.060959	
std	125.338794	66.256028	61.119149	29.317331	55.757415	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	0.000000	0.000000	
50%	0.000000	25.000000	0.000000	0.000000	0.000000	
75%	168.000000	68.000000	0.000000	0.000000	0.000000	
max	857.000000	547.000000	552.000000	508.000000	480.000000	

	PoolArea	MiscVal	MoSold	YrSold	SalePrice

```
count  1460.000000   1460.000000  1460.000000  1460.000000    1460.000000
mean      2.758904     43.489041     6.321918  2007.815753  180921.195890
std      40.177307    496.123024     2.703626     1.328095   79442.502883
min       0.000000      0.000000     1.000000  2006.000000   34900.000000
25%       0.000000      0.000000     5.000000  2007.000000  129975.000000
50%       0.000000      0.000000     6.000000  2008.000000  163000.000000
75%       0.000000      0.000000     8.000000  2009.000000  214000.000000
max     738.000000  15500.000000    12.000000  2010.000000  755000.000000

[8 rows x 37 columns]
```

- Identification of the null values in the dataset (1-b)

[9]:
```
data.isna().sum(axis=0)
```

[9]:
```
MSSubClass         0
MSZoning           0
LotFrontage      259
LotArea            0
Street             0
                  …
MoSold             0
YrSold             0
SaleType           0
SaleCondition      0
SalePrice          0
Length: 80, dtype: int64
```

Observation : Since the number of columns is too high to see all the columns, we will slice the result

- Visualization of the null values for the first half columns

[10]:
```
data.isna().sum(axis=0)[0:41]
```

[10]:
```
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
dtype: int64
```

- Visualization of the null values for the second half columns

[11]:
```
data.isna().sum(axis=0)[41:81]
```

[11]:
```
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
```

- Visualization of the values in each column (1-c)

[12]:
```
for columm in data:
    values = data[columm].unique()
    print(f"{columm}: {values}")
```

```
MSSubClass: [ 60  20  70  50 190  45  90 120  30  85  80 160  75 180  40]
MSZoning: ['RL' 'RM' 'C (all)' 'FV' 'RH']
LotFrontage: [ 65.  80.  68.  60.  84.  85.  75.  nan  51.  50.  70.  91.  72.
66.
 101.  57.  44. 110.  98.  47. 108. 112.  74. 115.  61.  48.  33.  52.
 100.  24.  89.  63.  76.  81.  95.  69.  21.  32.  78. 121. 122.  40.
 105.  73.  77.  64.  94.  34.  90.  55.  88.  82.  71. 120. 107.  92.
 134.  62.  86. 141.  97.  54.  41.  79. 174.  99.  67.  83.  43. 103.
  93.  30. 129. 140.  35.  37. 118.  87. 116. 150. 111.  49.  96.  59.
  36.  56. 102.  58.  38. 109. 130.  53. 137.  45. 106. 104.  42.  39.
 144. 114. 128. 149. 313. 168. 182. 138. 160. 152. 124. 153.  46.]
LotArea: [ 8450  9600 11250 … 17217 13175  9717]
Street: ['Pave' 'Grvl']
```

```
Alley: [nan 'Grvl' 'Pave']
LotShape: ['Reg' 'IR1' 'IR2' 'IR3']
LandContour: ['Lvl' 'Bnk' 'Low' 'HLS']
Utilities: ['AllPub' 'NoSeWa']
LotConfig: ['Inside' 'FR2' 'Corner' 'CulDSac' 'FR3']
LandSlope: ['Gtl' 'Mod' 'Sev']
Neighborhood: ['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: ['Norm' 'Feedr' 'PosN' 'Artery' 'RRAe' 'RRNn' 'RRAn' 'PosA' 'RRNe']
Condition2: ['Norm' 'Artery' 'RRNn' 'Feedr' 'PosN' 'PosA' 'RRAn' 'RRAe']
BldgType: ['1Fam' '2fmCon' 'Duplex' 'TwnhsE' 'Twnhs']
HouseStyle: ['2Story' '1Story' '1.5Fin' '1.5Unf' 'SFoyer' 'SLvl' '2.5Unf'
'2.5Fin']
OverallQual: [ 7  6  8  5  9  4 10  3  1  2]
OverallCond: [5 8 6 7 4 2 3 9 1]
YearBuilt: [2003 1976 2001 1915 2000 1993 2004 1973 1931 1939 1965 2005 1962
2006
 1960 1929 1970 1967 1958 1930 2002 1968 2007 1951 1957 1927 1920 1966
 1959 1994 1954 1953 1955 1983 1975 1997 1934 1963 1981 1964 1999 1972
 1921 1945 1982 1998 1956 1948 1910 1995 1991 2009 1950 1961 1977 1985
 1979 1885 1919 1990 1969 1935 1988 1971 1952 1936 1923 1924 1984 1926
 1940 1941 1987 1986 2008 1908 1892 1916 1932 1918 1912 1947 1925 1900
 1980 1989 1992 1949 1880 1928 1978 1922 1996 2010 1946 1913 1937 1942
 1938 1974 1893 1914 1906 1890 1898 1904 1882 1875 1911 1917 1872 1905]
YearRemodAdd: [2003 1976 2002 1970 2000 1995 2005 1973 1950 1965 2006 1962 2007
1960
 2001 1967 2004 2008 1997 1959 1990 1955 1983 1980 1966 1963 1987 1964
 1972 1996 1998 1989 1953 1956 1968 1981 1992 2009 1982 1961 1993 1999
 1985 1979 1977 1969 1958 1991 1971 1952 1975 2010 1984 1986 1994 1988
 1954 1957 1951 1978 1974]
RoofStyle: ['Gable' 'Hip' 'Gambrel' 'Mansard' 'Flat' 'Shed']
RoofMatl: ['CompShg' 'WdShngl' 'Metal' 'WdShake' 'Membran' 'Tar&Grv' 'Roll'
 'ClyTile']
Exterior1st: ['VinylSd' 'MetalSd' 'Wd Sdng' 'HdBoard' 'BrkFace' 'WdShing'
'CemntBd'
 'Plywood' 'AsbShng' 'Stucco' 'BrkComm' 'AsphShn' 'Stone' 'ImStucc'
 'CBlock']
Exterior2nd: ['VinylSd' 'MetalSd' 'Wd Shng' 'HdBoard' 'Plywood' 'Wd Sdng'
'CmentBd'
 'BrkFace' 'Stucco' 'AsbShng' 'Brk Cmn' 'ImStucc' 'AsphShn' 'Stone'
 'Other' 'CBlock']
MasVnrType: ['BrkFace' 'None' 'Stone' 'BrkCmn' nan]
MasVnrArea: [1.960e+02 0.000e+00 1.620e+02 3.500e+02 1.860e+02 2.400e+02
2.860e+02
 3.060e+02 2.120e+02 1.800e+02 3.800e+02 2.810e+02 6.400e+02 2.000e+02
```

```
 2.460e+02 1.320e+02 6.500e+02 1.010e+02 4.120e+02 2.720e+02 4.560e+02
 1.031e+03 1.780e+02 5.730e+02 3.440e+02 2.870e+02 1.670e+02 1.115e+03
 4.000e+01 1.040e+02 5.760e+02 4.430e+02 4.680e+02 6.600e+01 2.200e+01
 2.840e+02 7.600e+01 2.030e+02 6.800e+01 1.830e+02 4.800e+01 2.800e+01
 3.360e+02 6.000e+02 7.680e+02 4.800e+02 2.200e+02 1.840e+02 1.129e+03
 1.160e+02 1.350e+02 2.660e+02 8.500e+01 3.090e+02 1.360e+02 2.880e+02
 7.000e+01 3.200e+02 5.000e+01 1.200e+02 4.360e+02 2.520e+02 8.400e+01
 6.640e+02 2.260e+02 3.000e+02 6.530e+02 1.120e+02 4.910e+02 2.680e+02
 7.480e+02 9.800e+01 2.750e+02 1.380e+02 2.050e+02 2.620e+02 1.280e+02
 2.600e+02 1.530e+02 6.400e+01 3.120e+02 1.600e+01 9.220e+02 1.420e+02
 2.900e+02 1.270e+02 5.060e+02 2.970e+02       nan 6.040e+02 2.540e+02
 3.600e+01 1.020e+02 4.720e+02 4.810e+02 1.080e+02 3.020e+02 1.720e+02
 3.990e+02 2.700e+02 4.600e+01 2.100e+02 1.740e+02 3.480e+02 3.150e+02
 2.990e+02 3.400e+02 1.660e+02 7.200e+01 3.100e+01 3.400e+01 2.380e+02
 1.600e+03 3.650e+02 5.600e+01 1.500e+02 2.780e+02 2.560e+02 2.250e+02
 3.700e+02 3.880e+02 1.750e+02 2.960e+02 1.460e+02 1.130e+02 1.760e+02
 6.160e+02 3.000e+01 1.060e+02 8.700e+02 3.620e+02 5.300e+02 5.000e+02
 5.100e+02 2.470e+02 3.050e+02 2.550e+02 1.250e+02 1.000e+02 4.320e+02
 1.260e+02 4.730e+02 7.400e+01 1.450e+02 2.320e+02 3.760e+02 4.200e+01
 1.610e+02 1.100e+02 1.800e+01 2.240e+02 2.480e+02 8.000e+01 3.040e+02
 2.150e+02 7.720e+02 4.350e+02 3.780e+02 5.620e+02 1.680e+02 8.900e+01
 2.850e+02 3.600e+02 9.400e+01 3.330e+02 9.210e+02 7.620e+02 5.940e+02
 2.190e+02 1.880e+02 4.790e+02 5.840e+02 1.820e+02 2.500e+02 2.920e+02
 2.450e+02 2.070e+02 8.200e+01 9.700e+01 3.350e+02 2.080e+02 4.200e+02
 1.700e+02 4.590e+02 2.800e+02 9.900e+01 1.920e+02 2.040e+02 2.330e+02
 1.560e+02 4.520e+02 5.130e+02 2.610e+02 1.640e+02 2.590e+02 2.090e+02
 2.630e+02 2.160e+02 3.510e+02 6.600e+02 3.810e+02 5.400e+01 5.280e+02
 2.580e+02 4.640e+02 5.700e+01 1.470e+02 1.170e+03 2.930e+02 6.300e+02
 4.660e+02 1.090e+02 4.100e+01 1.600e+02 2.890e+02 6.510e+02 1.690e+02
 9.500e+01 4.420e+02 2.020e+02 3.380e+02 8.940e+02 3.280e+02 6.730e+02
 6.030e+02 1.000e+00 3.750e+02 9.000e+01 3.800e+01 1.570e+02 1.100e+01
 1.400e+02 1.300e+02 1.480e+02 8.600e+02 4.240e+02 1.047e+03 2.430e+02
 8.160e+02 3.870e+02 2.230e+02 1.580e+02 1.370e+02 1.150e+02 1.890e+02
 2.740e+02 1.170e+02 6.000e+01 1.220e+02 9.200e+01 4.150e+02 7.600e+02
 2.700e+01 7.500e+01 3.610e+02 1.050e+02 3.420e+02 2.980e+02 5.410e+02
 2.360e+02 1.440e+02 4.230e+02 4.400e+01 1.510e+02 9.750e+02 4.500e+02
 2.300e+02 5.710e+02 2.400e+01 5.300e+01 2.060e+02 1.400e+01 3.240e+02
 2.950e+02 3.960e+02 6.700e+01 1.540e+02 4.250e+02 4.500e+01 1.378e+03
 3.370e+02 1.490e+02 1.430e+02 5.100e+01 1.710e+02 2.340e+02 6.300e+01
 7.660e+02 3.200e+01 8.100e+01 1.630e+02 5.540e+02 2.180e+02 6.320e+02
 1.140e+02 5.670e+02 3.590e+02 4.510e+02 6.210e+02 7.880e+02 8.600e+01
 7.960e+02 3.910e+02 2.280e+02 8.800e+01 1.650e+02 4.280e+02 4.100e+02
 5.640e+02 3.680e+02 3.180e+02 5.790e+02 6.500e+01 7.050e+02 4.080e+02
 2.440e+02 1.230e+02 3.660e+02 7.310e+02 4.480e+02 2.940e+02 3.100e+02
 2.370e+02 4.260e+02 9.600e+01 4.380e+02 1.940e+02 1.190e+02]
ExterQual: ['Gd' 'TA' 'Ex' 'Fa']
ExterCond: ['TA' 'Gd' 'Fa' 'Po' 'Ex']
Foundation: ['PConc' 'CBlock' 'BrkTil' 'Wood' 'Slab' 'Stone']
```

```
BsmtQual: ['Gd' 'TA' 'Ex' nan 'Fa']
BsmtCond: ['TA' 'Gd' nan 'Fa' 'Po']
BsmtExposure: ['No' 'Gd' 'Mn' 'Av' nan]
BsmtFinType1: ['GLQ' 'ALQ' 'Unf' 'Rec' 'BLQ' nan 'LwQ']
BsmtFinSF1: [ 706  978  486  216  655  732 1369  859    0  851  906  998  737
733
  578  646  504  840  188  234 1218 1277 1018 1153 1213  731  643  967
  747  280  179  456 1351   24  763  182  104 1810  384  490  649  632
  941  739  912 1013  603 1880  565  320  462  228  336  448 1201   33
  588  600  713 1046  648  310 1162  520  108  569 1200  224  705  444
  250  984   35  774  419  170 1470  938  570  300  120  116  512  567
  445  695  405 1005  668  821  432 1300  507  679 1332  209  680  716
 1400  416  429  222   57  660 1016  370  351  379 1288  360  639  495
  288 1398  477  831 1904  436  352  611 1086  297  626  560  390  566
 1126 1036 1088  641  617  662  312 1065  787  468   36  822  378  946
  341   16  550  524   56  321  842  689  625  358  402   94 1078  329
  929  697 1573  270  922  503 1334  361  672  506  714  403  751  226
  620  546  392  421  905  904  430  614  450  210  292  795 1285  819
  420  841  281  894 1464  700  262 1274  518 1236  425  692  987  970
   28  256 1619   40  846 1124  720  828 1249  810  213  585  129  498
 1270  573 1410 1082  236  388  334  874  956  773  399  162  712  609
  371  540   72  623  428  350  298 1445  218  985  631 1280  241  690
  266  777  812  786 1116  789 1056   50 1128  775 1309 1246  986  616
 1518  664  387  471  385  365 1767  133  642  247  331  742 1606  916
  185  544  553  326  778  386  426  368  459 1350 1196  630  994  168
 1261 1567  299  897  607  836  515  374 1231  111  356  400  698 1247
  257  380   27  141  991  650  521 1436 2260  719  377 1330  348 1219
  783  969  673 1358 1260  144  584  554 1002  619  180  559  308  866
  895  637  604 1302 1071  290  728    2 1441  943  231  414  349  442
  328  594  816 1460 1324 1338  685 1422 1283   81  454  903  605  990
  206  150  457   48  871   41  674  624  480 1154  738  493 1121  282
  500  131 1696  806 1361  920 1721  187 1138  988  193  551  767 1186
  892  311  827  543 1003 1059  239  945   20 1455  965  980  863  533
 1084 1173  523 1148  191 1234  375  808  724  152 1180  252  832  575
  919  439  381  438  549  612 1163  437  394 1416  422  762  975 1097
  251  686  656  568  539  862  197  516  663  608 1636  784  249 1040
  483  196  572  338  330  156 1390  513  460  659  364  564  306  505
  932  750   64  633 1170  899  902 1238  528 1024 1064  285 2188  465
  322  860  599  354   63  223  301  443  489  284  294  814  165  552
  833  464  936  772 1440  748  982  398  562  484  417  699  696  896
  556 1106  651  867  854 1646 1074  536 1172  915  595 1237  273  684
  324 1165  138 1513  317 1012 1022  509  900 1085 1104  240  383  644
  397  740  837  220  586  535  410   75  824  592 1039  510  423  661
  248  704  412 1032  219  708  415 1004  353  702  369  622  212  645
  852 1150 1258  275  176  296  538 1157  492 1198 1387  522  658 1216
 1480 2096 1159  440 1456  883  547  788  485  340 1220  427  344  756
 1540  666  803 1000  885 1386  319  534  125 1314  602  192  593  804
 1053  532 1158 1014  194  167  776 5644  694 1572  746 1406  925  482
```

```
  189  765   80 1443  259  735  734 1447  548  315 1282  408  309  203
  865  204  790 1320  769 1070  264  759 1373  976  781   25 1110  404
  580  678  958 1336 1079   49  830]
BsmtFinType2: ['Unf' 'BLQ' nan 'ALQ' 'Rec' 'LwQ' 'GLQ']
BsmtFinSF2: [   0   32  668  486   93  491  506  712  362   41  169  869  150
670
   28 1080  181  768  215  374  208  441  184  279  306  180  580  690
  692  228  125 1063  620  175  820 1474  264  479  147  232  380  544
  294  258  121  391  531  344  539  713  210  311 1120  165  532   96
  495  174 1127  139  202  645  123  551  219  606  612  480  182  132
  336  468  287   35  499  723  119   40  117  239   80  472   64 1057
  127  630  128  377  764  345 1085  435  823  500  290  324  634  411
  841 1061  466  396  354  149  193  273  465  400  682  557  230  106
  791  240  547  469  177  108  600  492  211  168 1031  438  375  144
   81  906  608  276  661   68  173  972  105  420  546  334  352  872
  110  627  163 1029]
BsmtUnfSF: [ 150  284  434  540  490   64  317  216  952  140  134  177  175
1494
  520  832  426    0  468  525 1158  637 1777  200  204 1566  180  486
  207  649 1228 1234  380  408 1117 1097   84  326  445  383  167  465
 1296   83 1632  736  192  612  816   32  935  321  860 1410  148  217
  530 1346  576  318 1143 1035  440  747  701  343  280  404  840  724
  295 1768  448   36 1530 1065  384 1288  684 1013  402  635  163  168
  176  370  350  381  410  741 1226 1053  641  516  793 1139  550  905
  104  310  252 1125  203  728  732  510  899 1362   30  958  556  413
  479  297  658  262  891 1304  519 1907  336  107  432  403  811  396
  970  506  884  400  896  253  409   93 1200  572  774  769 1335  340
  882  779  112  470  294 1686  360  441  354  700  725  320  554  312
  968  504 1107  577  660   99  871  474  289  600  755  625 1121  276
  186 1424 1140  375   92  305 1176   78  274  311  710  686  457 1232
 1498 1010  160 2336  630  638  162   70 1357 1194  773  483  235  125
 1390  594 1694  488  357  626  916 1020 1367  798  452  392  975  361
  270  602 1482  680  606   88  342  212 1095   96  628 1560  744 2121
  768  386 1468 1145  244  698 1079  570  476  131  184  143 1092  324
 1541 1470  536  319  599  622  179  292  286   80  712  291  153 1088
 1249  166  906  604  100  818  844  596  210 1603  115  103  673  726
  995  967  721 1656  972  460  208  191  438 1869  371  624  552  322
  598  268  130  484  785  733  953  847  333 1580  411  982  808 1293
  939  784  595  229  114  522  735  405  117  961 1286  672 1141  806
  165 1064 1063  245 1276  892 1008  499 1316  463  242  444  281   35
  356  988  580  651  619  544  387  901  926  135  648   75  788 1307
 1078 1258  273 1436  557  930  780  813  878  122  248  588  524  288
  389  424 1375 1626  406  298 2153  417  739  225  611  237  290  264
  238  363  190 1969  697  414  316  466  420  254  960  397 1191  548
   50  178 1368  169  748  689 1264  467  605 1257  551  678  707  880
  378  223  578  969  379  765  149  912  620 1709  132  993  197 1374
   90  195  706 1163  367 1122 1515   55 1497  450  846   23  390  861
  285 1050  331 2042 1237  113  742  924  512  119  314  308  293  537
```

```
  126  427  309  914  173 1774  823  485 1116  978  636  564  108 1184
  796  366  300  542  645  664  756  247  776  849 1392   38 1406  111
  545  121 2046  161  261  567 1195  874 1342  151  989 1073  927  219
  224  526 1164  761  461  876  859  171  718  138  941  464  250   72
  508 1584  415   82  948  893  864 1349   76  487  652 1240  801  279
 1030  348  234 1198  740   89  586  323 1836  480  456 1935  338 1594
  102  374 1413  491 1129  255 1496  650 1926  154  999 1734  124 1417
   15  834 1649  936  778 1489  442 1434  352  458 1221 1099  416 1800
  227  907  528  189 1273  563  372  702 1090  435  198 1372  174 1638
  894  299  105  676 1120  431  218  110  795 1098 1043  481  666  142
  447  783 1670  277  412  794  239  662 1072  717  546  430  422  188
  266 1181 1753  964 1450 1905 1480  772 1032  220  187   29  495  640
  193  196  720  918 1428   77 1266 1128  692  770  750 1442 1007  501
  691 1550 1680 1330 1710  746  814  515  571  359  355  301  668  920
 1055 1420 1752  304 1302  833  133  549  705  722  799  462  429  810
  155  170  230 1459 1082  758 1290 1074  251  172  868  797  365  418
  730  533  671 1012 1528 1005 1373  500  762  752  399 1042   40   26
  932  278  459  568 1502  543  574  977  449  983  731  120  538  831
  994  341  879  815 1212  866 1630  328  141  364 1380   81  303  940
  764 1048  334 1689  690  792  585  473  246 1045 1405  201   14  841
 1104  241  925 2002   74  661  708 1152  256  804  812 1085  344  425
 1616  976  496  349  971 1393 1622 1352 1795 1017 1588  428  803  693
  858 1284 1203 1652   39  539 1217  257  715  616  240  315 1351 1026
 1571  156   61   95  482 1094   60  862  221  791  398  777  503  734
  709 1252  656 1319 1422  560 1573  589  877  136]
TotalBsmtSF: [ 856 1262  920  756 1145  796 1686 1107  952  991 1040 1175  912
1494
 1253  832 1004    0 1114 1029 1158  637 1777 1060 1566  900 1704 1484
  520  649 1228 1234 1398 1561 1117 1097 1297 1057 1088 1350  840  938
 1150 1752 1434 1656  736  955  794  816 1842  384 1425  970  860 1410
  780  530 1370  576 1143 1947 1453  747 1304 2223  845 1086  462  672
 1768  440  896 1237 1563 1065 1288  684  612 1013  990 1235  876 1214
  824  680 1588  960  458  950 1610  741 1226 1053  641  789  793 1844
  994 1264 1809 1028  729 1092 1125 1673  728  732 1080 1199 1362 1078
  660 1008  924  992 1063 1267 1461 1907  928  864 1734  910 1490 1728
  715  884  969 1710  825 1602 1200  572  774 1392 1232 1572 1541  882
 1149  644 1617 1582  720 1064 1606 1202 1151 1052 2216  968  504 1188
 1593  853  725 1431  855 1726 1360  755 1713 1121 1196  617  848 1424
 1140 1100 1157 1212  689 1070 1436  686  798 1248 1498 1010  713 2392
  630 1203  483 1373 1194 1462  894 1414  996 1694  735  540  626  948
 1845 1020 1367 1444 1573 1302 1314  975 1604  963 1482  506  926 1422
  802  740 1095 1385 1152 1240 1560 2121 1160  807 1468 1575  625  858
  698 1079  768  795 1416 1003  702 1165 1470 2000  700  319  861 1896
  697  972 2136  716 1347 1372 1249 1136 1502 1162  710 1719 1383  844
  596 1056 3206 1358  943 1499 1922 1536 1208 1215  967  721 1684  536
  958 1478  764 1848 1869  616  624  940 1142 1062  888  883 1394 1099
 1268  953  744  608  847  683  870 1580 1856  982 1026 1293  939  784
 1256  658 1041 1682  804  788 1144  961 1260 1310 1141  806 1281 1034
```

```
 1276 1340 1344  988  651 1518  907  901  765  799  648 3094 1440 1258
  915 1517  930  813 1533  872 1242 1364  588  709  560 1375 1277 1626
 1488  808  547 1976 2153 1705 1833 1792 1216  999 1113 1073  954  264
 1269  190 3200  866 1501  777 1218 1368 1084 2006 1244 3138 1379 1257
 1452  528 2035  611  707  880 1051 1581 1838 1650  723  654 1204 1069
 1709  998  993 1374 1389 1163 1122 1496  846  372 1164 1050 2042 1868
 1437  742  770 1722 1814 1430 1058  908  600  965 1032 1299 1120  936
  783 1822 1522  980 1116  978 1156  636 1554 1386  811 1520 1952 1766
  981 1094 2109  525  776 1486 1629 1138 2077 1406 1021 1408  738 1477
 2046  923 1291 1195 1190  874  551 1419 2444 1210  927 1112 1391 1800
  360 1473 1643 1324  270  859  718 1176 1311  971 1742  941 1698 1584
 1595  868 1153  893 1349 1337 1720 1479 1030 1318 1252  983 1860  836
 1935 1614  761 1413  956  712  650  773 1926  731 1417 1024  849 1442
 1649 1568  778 1489 2078 1454 1516 1067 1559 1127 1390 1273  918 1763
 1090 1054 1039 1148 1002 1638  105  676 1184 1109  892 2217 1505 1059
  951 2330 1670 1623 1017 1105 1001  546  480 1134 1104 1272 1316 1126
 1181 1753  964 1466  925 1905 1500  585 1632  819 1616 1161  828  945
  979  561  696 1330  817 1098 1428  673 1241  944 1225 1266 1128  485
 1930 1396  916  822  750 1700 1007 1187  691 1574 1680 1346  985 1657
  602 1022 1082  810 1504 1220 1132 1565 1338 1654 1620 1055  800 1306
 1475 2524 1992 1193  973  854  662 1103 1154  942 1048  727  690 1096
 1459 1251 1247 1074 1271  290  655 1463 1836  803  833  408  533 1012
 1552 1005 1530  974 1567 1006 1042 1298  704  932 1219 1296 1198  959
 1261 1598 1683  818 1600 2396 1624  831 1224  663  879  815 1630 2158
  931 1660  559 1300 1702 1075 1361 1106 1476 1689 2076  792 2110 1405
 1192  746 1986  841 2002 1332  935 1019  661 1309 1328 1085 6110 1246
  771  976 1652 1278 1902 1274 1393 1622 1352  420 1795  544 1510  911
  693 1284 1732 2033  570 1980  814  873  757 1108 2633 1571  984 1205
  714 1746 1525  482 1356  862  839 1286 1485 1594  622  791  708 1223
  913  656 1319 1932  539 1221 1542]
Heating: ['GasA' 'GasW' 'Grav' 'Wall' 'OthW' 'Floor']
HeatingQC: ['Ex' 'Gd' 'TA' 'Fa' 'Po']
CentralAir: ['Y' 'N']
Electrical: ['SBrkr' 'FuseF' 'FuseA' 'FuseP' 'Mix' nan]
1stFlrSF: [ 856 1262  920  961 1145  796 1694 1107 1022 1077 1040 1182  912 1494
 1253  854 1004 1296 1114 1339 1158 1108 1795 1060 1600  900 1704  520
  649 1228 1234 1700 1561 1132 1097 1297 1057 1152 1324 1328  884  938
 1150 1752 1518 1656  736  955  794  816 1842 1360 1425  983  860 1426
  780  581 1370  902 1143 2207 1479  747 1304 2223  845  885 1086  840
  526  952 1072 1768  682 1337 1563 1065  804 1301  684  612 1013  990
 1235  964 1260  905  680 1588  960  835 1225 1610  977 1535 1226 1053
 1047  789  997 1844 1216  774 1282 2259 1436  729 1092 1125 1699  728
  988  772 1080 1199 1586  958  660 1327 1721 1682 1214 1959  928  864
 1734  910 1501 1728  970  875  896  969 1710 1252 1200  572  991 1392
 1232 1572 1541  882 1149  808 1867 1707 1064 1362 1651 2158 1164 2234
  968  769  901 1340  936 1217 1224 1593 1549  725 1431  855 1726  929
 1713 1121 1279  865  848  720 1442 1696 1100 1180 1212  932  689 1236
  810 1137 1248 1498 1010  811 2392  630  483 1555 1194 1490  894 1414
```

```
 1014  798 1566  866  889  626 1222 1872  908 1375 1444 1306 1625 1302
 1314 1005 1604  963 1382 1482  926  764 1422  802 1052  778 1113 1095
 1363 1632 1560 2121 1156 1175 1468 1575  625 1085  858  698 1079 1148
 1644 1003  975 1041 1336 1210 1675 2000 1122 1035  861 1944  697  972
  793 2036  832  716 1153 1088 1372 1472 1249 1136 1553 1163 1898  803
 1719 1383 1445  596 1056 1629 1358  943 1619 1922 1536 1621 1215  993
  841 1684  536 1478 1848 1869 1453  616 1192 1167 1142 1352  495  790
  672 1394 1268 1287  953 1120  752 1319  847  904  914 1580 1856 1007
 1026  939  784 1269  658 1742  788  735 1144  876 1112 1288 1310 1165
  806 1620 1166 1071 1050 1276 1028  756 1344 1602 1470 1196  707  907
 1208 1412  765  827  734  694 2402 1440 1128 1258  933 1689 1888  956
  679  813 1533  888  786 1242  624 1663  833  979  575  849 1277 1634
 1502 1161 1976 1652 1493 2069 1718 1131 1850 1792  916  999 1073 1484
 1766  886 3228 1133  899 1801 1218 1368 2020 1378 1244 3138 1266 1476
  605 2515 1509  751  334  820  880 1159 1601 1838 1680  767  664 1377
  915  768  825 1069 1717 1126 1006 1048  897 1557 1389  996 1134 1496
  846  576  877 1320  703 1429 2042 1521  989 2028  838 1473  779  770
  924 1826 1402 1647 1058  927  600 1186 1940 1029 1032 1299 1054  807
 1828 1548  980 1012 1116 1520 1350 1089 1554 1411  800 1567  981 1094
 1051  822  755  909 2113  525  851 1486 1686 1181 2097 1454 1465 1679
 1437  738 1839  792 2046  923 1291 1668 1195 1190  874  551 1419 2444
 1238 1067 1391 1800 1264  372 1824  859 1576 1178 1325  971 1698 1776
 1616 1146  948 1349 1464 1720 1038  742  757 1506 1836 1690 1220 1117
 1973 1204 1614 1430 1110 1342  966  976 1062 1127 1285  773 1966 1428
 1075 1309 1044  686 1661 1008  944 1489 2084 1434 1160  941 1516 1559
 1099 1701 1307 1456  918 1779  702 1512 1039 1002 1646 1547 1036  676
 1184 1462 1155 1090 1187  954  892 1709 1712  872 2217 1505 1068  951
 2364 1670 1063 1636 1020 1105 1015 1001  546  480 1229 1272 1316 1617
 1098 1788 1466  925 1905 1500 1207 1188 1381  965 1168  561  696 1542
  824  783  673  869 1241 1118 1407  750  691 1574 1504  985 1657 1664
 1082 2898 1687 1654 1055 1803 1532 2524 1733 1992 1771  930 1526 1091
 1523 1364 1130 1096 1338 1103 1154  799  893  829 1240 1459 1251 1247
 1390  438  950  887 1021 1552  812 1530  974  986 1042 1298 1811 1265
 1640 1432  959 1831 1261 1170 2129  818 1124 2411  949 1624  831 1622
  842  663  879  815 1630 1074 2196 1283 1660 1318 1211 2136 1138 1702
 1507 1361 1024 1141 1173 2076 1140 1034 2110 1405  760 1987 1104  713
 2018 1968 1332  935 1357  661 1724 1573 1582 1659 4692 1246  753 1203
 1294 1902 1274 1787 1061  708 1584 1334  693 1284 1172 2156 2053  992
 1078 1980 1281  814 2633 1571  984  754 2117  998 1416 1746 1525 1221
  741 1569 1223  962 1537 1932 1423  913 1578 2073 1256]
2ndFlrSF: [ 854    0  866  756 1053  566  983  752 1142 1218  668 1320  631  716
  676  860 1519  530  808  977 1330  833  765  462  213  548  960  670
 1116  876  612 1031  881  790  755  592  939  520  639  656 1414  884
  729 1523  728  351  688  941 1032  848  836  475  739 1151  448  896
  524 1194  956 1070 1096  467  547  551  880  703  901  720  316 1518
  704 1178  754  601 1360  929  445  564  882  920  518  817 1257  741
  672 1306  504 1304 1100  730  689  591  888 1020  828  700  842 1286
  864  829 1092  709  844 1106  596  807  625  649  698  840  780  568
```

```
  795  648  975  702 1242 1818 1121  371  804  325  809 1200  871 1274
 1347 1332 1177 1080  695  167  915  576  605  862  495  403  838  517
 1427  784  711  468 1081  886  793  665  858  874  526  590  406 1157
  299  936  438 1098  766 1101 1028 1017 1254  378 1160  682  110  600
  678  834  384  512  930  868  224 1103  560  811  878  574  910  620
  687  546  902 1000  846 1067  914  660 1538 1015 1237  611  707  527
 1288  832  806 1182 1040  439  717  511 1129 1370  636  533  745  584
  812  684  595  988  800  677  573 1066  778  661 1440  872  788  843
  713  567  651  762  482  738  586  679  644  900  887 1872 1281  472
 1312  319  978 1093  473  664 1540 1276  441  348 1060  714  744 1203
  783 1097  734  767 1589  742  686 1128 1111 1174  787 1072 1088 1063
  545  966  623  432  581  540  769 1051  761  779  514  455 1426  785
  521  252  813 1120 1037 1169 1001 1215  928 1140 1243  571 1196 1038
  561  979  701  332  368  883 1336 1141  634  912  798  985  826  831
  750  456  602  855  336  408  980  998 1168 1208  797  850  898 1054
  895  954  772 1230  727  454  370  628  304  582 1122 1134  885  640
  580 1112  653  220  240 1362  534  539  650  918  933  712 1796  971
 1175  743  523 1216 2065  272  685  776  630  984  875  913  464 1039
 1259  940  892  725  924  764  925 1479  192  589  992  903  430  748
  587  994  950 1323  732 1357  557 1296  390 1185  873 1611  457  796
  908  550  989  932  358 1392  349  691 1349  768  208  622  857  556
 1044  708  626  904  510 1104  830  981  870  694 1152]
LowQualFinSF: [  0 360 513 234 528 572 144 392 371 390 420 473 156 515  80  53
232 481
 120 514 397 479 205 384]
GrLivArea: [1710 1262 1786 1717 2198 1362 1694 2090 1774 1077 1040 2324  912
1494
 1253  854 1004 1296 1114 1339 2376 1108 1795 1060 1600  900 1704  520
 1317 1228 1234 1700 1561 2452 1097 1297 1057 1152 1324 1328  884  938
 1150 1752 2149 1656 1452  955 1470 1176  816 1842 1360 1425 1739 1720
 2945  780 1158 1111 1370 2034 2473 2207 1479  747 2287 2223  845 1718
 1086 1605  988  952 1285 1768 1230 2142 1337 1563 1065 1474 2417 1560
 1224 1526  990 1235  964 2291 1588  960  835 1225 1610 1732 1535 1226
 1818 1992 1047  789 1517 1844 1855 1430 2696 2259 2320 1458 1092 1125
 3222 1456 1123 1080 1199 1586  754  958  840 1348 1053 2157 2054 1327
 1721 1682 1214 1959 1852 1764  864 1734 1385 1501 1728 1709  875 2035
 1344  969 1993 1252 1200 1096 1968 1947 2462 1232 2668 1541  882 1616
 1355 1867 2161 1707 1382 1767 1651 2158 2060 1920 2234  968 1525 1802
 1340 2082 3608 1217 1593 2727 1431 1726 3112 2229 1713 1121 1279 1310
  848 1284 1442 1696 1100 2062 1212 1392 1236 1436 1954 1248 1498 2267
 1552 2392 1302 2520  987 1555 1194 2794  894 1960 1414 1744 1487 1566
  866 1440 2110 1872 1928 1375 1668 2144 1306 1625 1640 1314 1604 1792
 2574 1316  764 1422 1511 2192  778 1113 1939 1363 2270 1632 1548 2121
 2022 1982 1468 1575 1250  858 1396 1919 1716 2263 1644 1003 1558 1950
 1743 1336 3493 2000 2243 1406  861 1944  972 1118 2036 1641 1432 2353
 2646 1472 2596 2468 2730 1163 2978  803 1719 1383 2134 1192 1056 1629
 1358 1638 1922 1536 1621 1215 1908  841 1684 1112 1577 1478 1626 2728
 1869 1453  720 1595 1167 1142 1352 1924 1505 1574 1394 1268 1287 1664
```

```
  752 1319  904  914 2466 1856 1800 1691 1301 1797  784 1953 1269 1184
 2332 1367 1961  788 1034 1144 1812 1550 1288  672 1572 1620 1639 1680
 2172 2078 1276 1028 2097 1400 2624 1134 1602 2630 1196 1389  907 1208
 1412 1198 1365  630 1661  694 2402 1573 1258 1689 1888 1886 1376 1183
  813 1533 1756 1590 1242 1663 1666 1203 1935 1135 1660 1277 1634 1502
 1969 1072 1976 1652  970 1493 2643 1131 1850 1826 1216  999 1073 1484
 2414 1304 1578  886 3228 1820  899 1218 1801 1322 1911 1378 1041 1368
 2020 2119 2344 1796 2080 1294 1244 4676 2398 1266  928 2713  605 2515
 1509  827  334 1347 1724 1159 1601 1838 2285  767 1496 2183 1635  768
  825 2094 1069 1126 2046 1048 1446 1557  996 1674 2295 1647 2504 2132
  943 1692 1109 1477 1320 1429 2042 2775 2028  838  860 1473  935 1582
 2296  924 1402 1556 1904 1915 1986 2008 3194 1029 2153 1032 1120 1054
  832 1828 2262 2614  980 1512 1790 1116 1520 1350 1750 1554 1411 3395
  800 1387  796 1567 1518 1929 2704 1766  981 1094 1839 1665 1510 1469
 2113 1486 2448 1181 1936 2380 1679 1437 1180 1476 1369 1136 1441  792
  923 1291 1761 1102 1419 4316 2519 1539 1137  616 1148 1391 1164 2576
 1824  729 1178 2554 2418  971 1742 1698 1776 1146 2031  948 1349 1464
 2715 2256 2640 1529 1140 2098 1026 1471 1386 2531 1547 2365 1506 1714
 1836 3279 1220 1117 1973 1204 1614 1603 1110 1342 2084  901 2087 1145
 1062 2013 1895 1564  773 3140 1688 2822 1128 1428 1576 2138 1309 1044
 1008 1052  936 1733 1489 1434 2126 1223 1829 1516 1067 1559 1099 1482
 1165 1416 1701 1775 2358 1646 1445 1779 1481 2654 1426 1039 1372 1002
 1949  910 2610 2224 1155 1090 2230  892 1712 1393 2217 1683 1068  951
 2240 2364 1670  902 1063 1636 2057 2274 1015 2002  480 1229 2127 2200
 1617 1686 2374 1978 1788 2236 1466  925 1905 1500 2069 1971 1962 2403
 1381  965 1958 2872 1894 1308 1098 1095  918 2019  869 1241 2612 2290
 1940 2030 1851 1050  944  691 1504  985 1657 1522 1271 1022 1082 1132
 2898 1264 3082 1654  954 1803 2329 2524 2868 1771  930 1977 1989 1523
 1364 2184 1991 1338 2337 1103 1154 2260 1571 1611 2521  893 1240 1740
 1459 1251 1247 1088  438  950 2622 2021 1690 1658 1964  833 1012  698
 1005 1530 1981  974 2210  986 1020 1868 2828 1006 1298  932 1811 1265
 1580 1876 1671 2108 3627 1261 3086 2345 1343 1124 2514 4476 1130 1221
 1699 1624 1804 1622 1863 1630 1074 2196 1283 1845 1902 1211 1846 2136
 1490 1138 1933 1702 1507 2620 1190 1188 1784 1948 1141 1173 2076 1553
 2058 1405  874 2167 1987 1166 1675 1889 2018 3447 1524 1357 1395 2447
 1659 1970 2372 5642 1246 1983 2526 1708 1122 1274 2810 2599 2112 1787
 1923  708  774 2792 1334  693 1861  872 2169 1913 2156 2634 3238 1865
 1078 1980 2601 1738 1475 1374 2633  790 2117 1762 2784 1746 1584 1912
 2482 1687 1513 1608 2093 1840 1848 1569 2450 2201  804 1537 1932 1725
 2555 2007  913 1346 2073 2340 1256]
BsmtFullBath: [1 0 2 3]
BsmtHalfBath: [0 1 2]
FullBath: [2 1 3 0]
HalfBath: [1 0 2]
BedroomAbvGr: [3 4 1 2 0 5 6 8]
KitchebvGr: [1 2 3 0]
KitchenQual: ['Gd' 'TA' 'Ex' 'Fa']
TotRmsAbvGrd: [ 8  6  7  9  5 11  4 10 12  3  2 14]
```

```
Functiol: ['Typ' 'Min1' 'Maj1' 'Min2' 'Mod' 'Maj2' 'Sev']
Fireplaces: [0 1 2 3]
FireplaceQu: [nan 'TA' 'Gd' 'Fa' 'Ex' 'Po']
GarageType: ['Attchd' 'Detchd' 'BuiltIn' 'CarPort' nan 'Basment' '2Types']
GarageYrBlt: [2003. 1976. 2001. 1998. 2000. 1993. 2004. 1973. 1931. 1939. 1965.
2005.
 1962. 2006. 1960. 1991. 1970. 1967. 1958. 1930. 2002. 1968. 2007. 2008.
 1957. 1920. 1966. 1959. 1995. 1954. 1953.   nan 1983. 1977. 1997. 1985.
 1963. 1981. 1964. 1999. 1935. 1990. 1945. 1987. 1989. 1915. 1956. 1948.
 1974. 2009. 1950. 1961. 1921. 1900. 1979. 1951. 1969. 1936. 1975. 1971.
 1923. 1984. 1926. 1955. 1986. 1988. 1916. 1932. 1972. 1918. 1980. 1924.
 1996. 1940. 1949. 1994. 1910. 1978. 1982. 1992. 1925. 1941. 2010. 1927.
 1947. 1937. 1942. 1938. 1952. 1928. 1922. 1934. 1906. 1914. 1946. 1908.
 1929. 1933.]
GarageFinish: ['RFn' 'Unf' 'Fin' nan]
GarageCars: [2 3 1 0 4]
GarageArea: [ 548  460  608  642  836  480  636  484  468  205  384  736  352
840
  576  516  294  853  280  534  572  270  890  772  319  240  250  271
  447  556  691  672  498  246    0  440  308  504  300  670  826  386
  388  528  894  565  641  288  645  852  558  220  667  360  427  490
  379  297  283  509  405  758  461  400  462  420  432  506  684  472
  366  476  410  740  648  273  546  325  792  450  180  430  594  390
  540  264  530  435  453  750  487  624  471  318  766  660  470  720
  577  380  434  866  495  564  312  625  680  678  726  532  216  303
  789  511  616  521  451 1166  252  497  682  666  786  795  856  473
  398  500  349  454  644  299  210  431  438  675  968  721  336  810
  494  457  818  463  604  389  538  520  309  429  673  884  868  492
  413  924 1053  439  671  338  573  732  505  575  626  898  529  685
  281  539  418  588  282  375  683  843  552  870  888  746  708  513
 1025  656  872  292  441  189  880  676  301  474  706  617  445  200
  592  566  514  296  244  610  834  639  501  846  560  596  600  373
  947  350  396  864  304  784  696  569  628  550  493  578  198  422
  228  526  525  908  499  508  694  874  164  402  515  286  603  900
  583  889  858  502  392  403  527  765  367  426  615  871  570  406
  590  612  650 1390  275  452  842  816  621  544  486  230  261  531
  393  774  749  364  627  260  256  478  442  562  512  839  330  711
 1134  416  779  702  567  832  326  551  606  739  408  475  704  983
  768  632  541  320  800  831  554  878  752  614  481  496  423  841
  895  412  865  630  605  602  618  444  397  455  409  820 1020  598
  857  595  433  776 1220  458  613  456  436  812  686  611  425  343
  479  619  902  574  523  414  738  354  483  327  756  690  284  833
  601  533  522  788  555  689  796  808  510  255  424  305  368  824
  328  160  437  665  290  912  905  542  716  586  467  582 1248 1043
  254  712  719  862  928  782  466  714 1052  225  234  324  306  830
  807  358  186  693  482  813  995  757 1356  459  701  322  315  668
  404  543  954  850  477  276  518 1014  753 1418  213  844  860  748
  248  287  825  647  342  770  663  377  804  936  722  208  662  754
```

```
  622  620  370 1069  372  923  192]
GarageQual: ['TA' 'Fa' 'Gd' nan 'Ex' 'Po']
GarageCond: ['TA' 'Fa' nan 'Gd' 'Po' 'Ex']
PavedDrive: ['Y' 'N' 'P']
WoodDeckSF: [  0 298 192  40 255 235  90 147 140 160  48 240 171 100 406 222 288
49
 203 113 392 145 196 168 112 106 857 115 120  12 576 301 144 300  74 127
 232 158 352 182 180 166 224  80 367  53 188 105  24  98 276 200 409 239
 400 476 178 574 237 210 441 116 280 104  87 132 238 149 355  60 139 108
 351 209 216 248 143 365 370  58 197 263 123 138 333 250 292  95 262  81
 289 124 172 110 208 468 256 302 190 340 233 184 201 142 122 155 670 135
 495 536 306  64 364 353  66 159 146 296 125  44 215 264  88  89  96 414
 519 206 141 260 324 156 220  38 261 126  85 466 270  78 169 320 268  72
 349  42  35 326 382 161 179 103 253 148 335 176 390 328 312 185 269 195
  57 236 517 304 198 426  28 316 322 307 257 219 416 344 380  68 114 327
 165 187 181  92 228 245 503 315 241 303 133 403  36  52 265 207 150 290
 486 278  70 418 234  26 342  97 272 121 243 511 154 164 173 384 202  56
 321  86 194 421 305 117 550 509 153 394 371  63 252 136 186 170 474 214
 199 728 436  55 431 448 361 362 162 229 439 379 356  84 635 325  33 212
 314 242 294  30 128  45 177 227 218 309 404 500 668 402 283 183 175 586
 295  32 366 736]
OpenPorchSF: [ 61   0  42  35  84  30  57 204   4  21  33 213 112 102 154 159
110  90
  56  32  50 258  54  65  38  47  64  52 138 104  82  43 146  75  72  70
  49  11  36 151  29  94 101 199  99 234 162  63  68  46  45 122 184 120
  20  24 130 205 108  80  66  48  25  96 111 106  40 114   8 136 132  62
 228  60 238 260  27  74  16 198  26  83  34  55  22  98 172 119 208 105
 140 168  28  39 148  12  51 150 117 250  10  81  44 144 175 195 128  76
  17  59 214 121  53 231 134 192 123  78 187  85 133 176 113 137 125 523
 100 285  88 406 155  73 182 502 274 158 142 243 235 312 124 267 265  87
 288  23 152 341 116 160 174 247 291  18 170 156 166 129 418 240  77 364
 188 207  67  69 131 191  41 118 252 189 282 135  95 224 169 319  58  93
 244 185 200  92 180 263 304 229 103 211 287 292 241 547  91  86 262 210
 141  15 126 236]
EnclosedPorch: [  0 272 228 205 176  87 172 102  37 144  64 114 202 128 156  44
77 192
 140 180 183  39 184  40 552  30 126  96  60 150 120 112 252  52 224 234
 244 268 137  24 108 294 177 218 242  91 160 130 169 105  34 248 236  32
  80 115 291 116 158 210  36 200  84 148 136 240  54 100 189 293 164 216
 239  67  90  56 129  98 143  70 386 154 185 134 196 264 275 230 254  68
 194 318  48  94 138 226 174  19 170 220 214 280 190 330 208 145 259  81
  42 123 162 286 168  20 301 198 221 212  50  99]
3SsnPorch: [  0 320 407 130 180 168 140 508 238 245 196 144 182 162  23 216  96
153
 290 304]
ScreenPorch: [  0 176 198 291 252  99 184 168 130 142 192 410 224 266 170 154
153 144
 128 259 160 271 234 374 185 182  90 396 140 276 180 161 145 200 122  95
```

```
 120  60 126 189 260 147 385 287 156 100 216 210 197 204 225 152 175 312
 222 265 322 190 233  63  53 143 273 288 263  80 163 116 480 178 440 155
 220 119 165  40]
PoolArea: [  0 512 648 576 555 480 519 738]
PoolQC: [nan 'Ex' 'Fa' 'Gd']
Fence: [nan 'MnPrv' 'GdWo' 'GdPrv' 'MnWw']
MiscFeature: [nan 'Shed' 'Gar2' 'Othr' 'TenC']
MiscVal: [    0   700   350   500   400   480   450 15500  1200   800  2000
600
  3500  1300    54   620   560  1400  8300  1150  2500]
MoSold: [ 2  5  9 12 10  8 11  4  1  7  3  6]
YrSold: [2008 2007 2006 2009 2010]
SaleType: ['WD' 'New' 'COD' 'ConLD' 'ConLI' 'CWD' 'ConLw' 'Con' 'Oth']
SaleCondition: ['Normal' 'Abnorml' 'Partial' 'AdjLand' 'Alloca' 'Family']
SalePrice: [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]
```

Observation: There are no features with only one value.

1.2 2- Generation of dataset for numerical and categorical variables

- Dataset with numerical variables

[13]:
```
numerical_features_df = data.select_dtypes(include=np.number)
numerical_features_df
```

[13]:
```
      MSSubClass  LotFrontage  LotArea  OverallQual  OverallCond  YearBuilt  \
Id
1             60         65.0     8450            7            5       2003
```

```
2             20         80.0     9600            6            8       1976
3             60         68.0    11250            7            5       2001
4             70         60.0     9550            7            5       1915
5             60         84.0    14260            8            5       2000
…            …           …        …               …            …       …
1456          60         62.0     7917            6            5       1999
1457          20         85.0    13175            6            6       1978
1458          70         66.0     9042            7            9       1941
1459          20         68.0     9717            5            6       1950
1460          20         75.0     9937            5            6       1965

      YearRemodAdd  MasVnrArea  BsmtFinSF1  BsmtFinSF2  …  WoodDeckSF  \
Id                                                      …
1             2003       196.0         706           0  …           0
2             1976         0.0         978           0  …         298
3             2002       162.0         486           0  …           0
4             1970         0.0         216           0  …           0
5             2000       350.0         655           0  …         192
…            …           …          …          …   …           …
1456          2000         0.0           0           0  …           0
1457          1988       119.0         790         163  …         349
1458          2006         0.0         275           0  …           0
1459          1996         0.0          49        1029  …         366
1460          1965         0.0         830         290  …         736

      OpenPorchSF  EnclosedPorch  3SsnPorch  ScreenPorch  PoolArea  MiscVal  \
Id
1              61              0          0            0         0        0
2               0              0          0            0         0        0
3              42              0          0            0         0        0
4              35            272          0            0         0        0
5              84              0          0            0         0        0
…             …              …          …            …         …        …
1456           40              0          0            0         0        0
1457            0              0          0            0         0        0
1458           60              0          0            0         0     2500
1459            0            112          0            0         0        0
1460           68              0          0            0         0        0

      MoSold  YrSold  SalePrice
Id
1          2    2008     208500
2          5    2007     181500
3          9    2008     223500
4          2    2006     140000
5         12    2008     250000
…        …     …        …
```

```
1456       8    2007     175000
1457       2    2010     210000
1458       5    2010     266500
1459       4    2010     142125
1460       6    2008     147500

[1460 rows x 37 columns]
```

- List of the names of the columns of the dataset with numerical variables

[14]:
```
numerical_features_list = list(numerical_features_df.columns)
numerical_features_list
```

[14]:
```
['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']
```

- Dataset with categorical variables

[15]:
```
categorical_features_df = data.select_dtypes(exclude=np.number)
categorical_features_df
```

[15]:

Id	MSZoning	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	…	GarageType	GarageFinish	GarageQual	GarageCond	PavedDrive	PoolQC	Fence	MiscFeature	SaleType	SaleCondition
1	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	…	Attchd	RFn	TA	TA	Y	NaN	NaN	NaN	WD	Normal
2	RL	Pave	NaN	Reg	Lvl	AllPub	FR2	Gtl	…	Attchd	RFn	TA	TA	Y	NaN	NaN	NaN	WD	Normal
3	RL	Pave	NaN	IR1	Lvl	AllPub	Inside	Gtl	…	Attchd	RFn	TA	TA	Y	NaN	NaN	NaN	WD	Normal
4	RL	Pave	NaN	IR1	Lvl	AllPub	Corner	Gtl	…	Detchd	Unf	TA	TA	Y	NaN	NaN	NaN	WD	Abnorml
5	RL	Pave	NaN	IR1	Lvl	AllPub	FR2	Gtl	…	Attchd	RFn	TA	TA	Y	NaN	NaN	NaN	WD	Normal
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
1456	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	…	Attchd	RFn	TA	TA	Y	NaN	NaN	NaN	WD	Normal
1457	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	…	Attchd	Unf	TA	TA	Y	NaN	NaN	NaN	WD	Normal
1458	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	…	Attchd	RFn	TA	TA	Y	NaN	NaN	NaN	WD	Normal
1459	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	…	Attchd	Unf	TA	TA	Y	NaN	NaN	NaN	WD	Normal
1460	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	…	Attchd	Fin	TA	TA	Y	NaN	NaN	NaN	WD	Normal

```
1458         TA          Y    NaN  GdPrv        Shed       WD        Normal
1459         TA          Y    NaN    NaN         NaN       WD        Normal
1460         TA          Y    NaN    NaN         NaN       WD        Normal

[1460 rows x 43 columns]
```

- List of the names of the columns of the dataset with numerical variables

[16]:
```
categorical_features_list = list(categorical_features_df.columns)
categorical_features_list
```

[16]:
```
['MSZoning',
 'Street',
 'Alley',
 'LotShape',
 'LandContour',
 'Utilities',
 'LotConfig',
 'LandSlope',
 'Neighborhood',
 'Condition1',
 'Condition2',
 'BldgType',
 'HouseStyle',
 'RoofStyle',
 'RoofMatl',
 'Exterior1st',
 'Exterior2nd',
 'MasVnrType',
 'ExterQual',
 'ExterCond',
 'Foundation',
 'BsmtQual',
 'BsmtCond',
 'BsmtExposure',
 'BsmtFinType1',
 'BsmtFinType2',
 'Heating',
 'HeatingQC',
 'CentralAir',
 'Electrical',
 'KitchenQual',
 'Functiol',
 'FireplaceQu',
 'GarageType',
 'GarageFinish',
 'GarageQual',
```

```
 'GarageCond',
 'PavedDrive',
 'PoolQC',
 'Fence',
 'MiscFeature',
 'SaleType',
 'SaleCondition']
```

1.3 3- EDA of numerical variables

1.3.1 a-Treatment of missing values

- Visualization of the missing values

[17]:
```
numerical_features_df.isna().sum(axis=0)
```

[17]:
```
MSSubClass         0
LotFrontage      259
LotArea            0
OverallQual        0
OverallCond        0
YearBuilt          0
YearRemodAdd       0
MasVnrArea         8
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       81
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
```

Observations: 3 features have missing values: ”LotFrontage”, ”GarageYrBuilt” and ”MasVnrArea”.

- Treatment of the missing values in ”LotFrontage”: The value at -1 will be set for any missing value and not 0 to avoid confusion with the linear footage on the front of the property.

[18]:
```
numerical_features_df["LotFrontage"].fillna(value=-1, inplace=True)
```

```
/usr/local/lib/python3.7/site-packages/pandas/core/series.py:4536:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  downcast=downcast,
```

[19]:
```
numerical_features_df["LotFrontage"].unique()
```

[19]:
```
array([ 65.,  80.,  68.,  60.,  84.,  85.,  75.,  -1.,  51.,  50.,  70.,
        91.,  72.,  66., 101.,  57.,  44., 110.,  98.,  47., 108., 112.,
        74., 115.,  61.,  48.,  33.,  52., 100.,  24.,  89.,  63.,  76.,
        81.,  95.,  69.,  21.,  32.,  78., 121., 122.,  40., 105.,  73.,
        77.,  64.,  94.,  34.,  90.,  55.,  88.,  82.,  71., 120., 107.,
        92., 134.,  62.,  86., 141.,  97.,  54.,  41.,  79., 174.,  99.,
        67.,  83.,  43., 103.,  93.,  30., 129., 140.,  35.,  37., 118.,
        87., 116., 150., 111.,  49.,  96.,  59.,  36.,  56., 102.,  58.,
        38., 109., 130.,  53., 137.,  45., 106., 104.,  42.,  39., 144.,
       114., 128., 149., 313., 168., 182., 138., 160., 152., 124., 153.,
        46.])
```

- Treatment of the missing values in ”GarageYrBuilt”: Only 81 values are missing. We will see if we can set the missing values equal to the year built of the properties.
- Determination of the number of properties which have the garage built in the same year of the house.

[20]:
```
compare_yrbuilt_garageyrbuilt = data[["YearBuilt", "GarageYrBlt"]]
compare_yrbuilt_garageyrbuilt['diff'] = data["YearBuilt"]-data["GarageYrBlt"]
compare_yrbuilt_garageyrbuilt
```

```
/usr/local/lib/python3.7/site-packages/ipykernel_launcher.py:2:
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

[20]:

Id	YearBuilt	GarageYrBlt	diff
1	2003	2003.0	0.0
2	1976	1976.0	0.0
3	2001	2001.0	0.0
4	1915	1998.0	-83.0
5	2000	2000.0	0.0
…	…	…	…
1456	1999	1999.0	0.0
1457	1978	1978.0	0.0
1458	1941	1941.0	0.0
1459	1950	1950.0	0.0
1460	1965	1965.0	0.0

[1460 rows x 3 columns]

[21]:
```
compare_yrbuilt_garageyrbuilt.loc[compare_yrbuilt_garageyrbuilt['diff']==0].
↪shape[0]
```

[21]:
```
1089
```

Observation: 1089 out of 1460 properties have the garage built the same year of the house so it represents 75 % of the properties. So we can set the missing values of "GarageYrBlt" to the corresponding "YearBuilt" values.

[22]:
```
numerical_features_df['GarageYrBlt'].
↪fillna(value=numerical_features_df['YearBuilt'], inplace=True)
```

```
/usr/local/lib/python3.7/site-packages/pandas/core/series.py:4536:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  downcast=downcast,
```

- Treatment of the missing values in "MasVnrArea": Only 8 values are missing out of 1460 which represents about 0.55% of the lines. Thus the lines with missing values will be drop.

[23]:
```
numerical_features_df = numerical_features_df.dropna(axis=0)
numerical_features_df.shape
```

[23]:
```
(1452, 37)
```

[24]:
```
numerical_features_df.isna().sum(axis=0)
```

[24]:
```
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
```

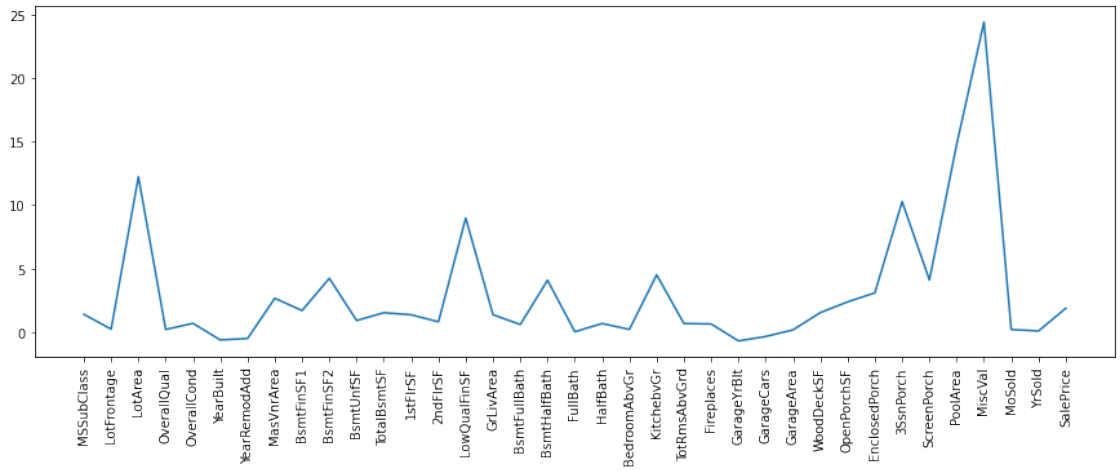
Observation: We don't have any missing values. ### b- Visualization of the skewness of the numerical features

[25]:
```
numerical_features_df.skew(axis=0)
```

[25]:
```
MSSubClass        1.407336
LotFrontage       0.238494
LotArea          12.240033
OverallQual       0.214636
OverallCond       0.694929
YearBuilt        -0.608915
YearRemodAdd     -0.497281
MasVnrArea        2.669084
BsmtFinSF1        1.702885
BsmtFinSF2        4.241902
BsmtUnfSF         0.920938
TotalBsmtSF       1.533040
1stFlrSF          1.373395
2ndFlrSF          0.814485
LowQualFinSF      8.985769
GrLivArea         1.374375
BsmtFullBath      0.605294
BsmtHalfBath      4.090233
FullBath          0.036446
HalfBath          0.683031
BedroomAbvGr      0.217581
KitchebvGr        4.514591
TotRmsAbvGrd      0.680438
Fireplaces        0.648734
GarageYrBlt      -0.688991
GarageCars       -0.338128
GarageArea        0.183300
WoodDeckSF        1.542306
OpenPorchSF       2.385725
EnclosedPorch     3.095363
3SsnPorch        10.275369
ScreenPorch       4.109058
PoolArea         14.787221
MiscVal          24.409889
MoSold            0.210208
YrSold            0.095878
SalePrice         1.884045
dtype: float64
```

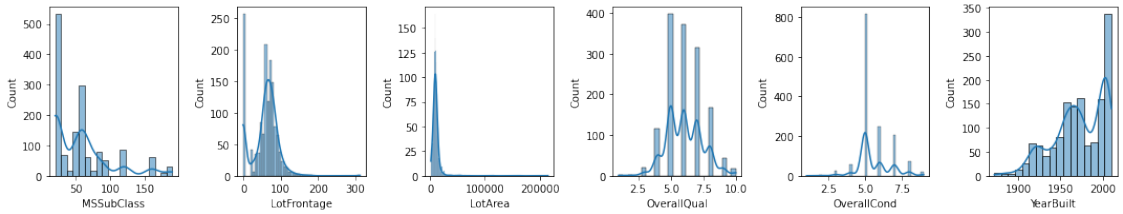
[26]:
```
plt.figure(figsize=(15, 5))
plt.xticks(rotation = 90)
plt.plot(numerical_features_df.skew(axis=0))
```

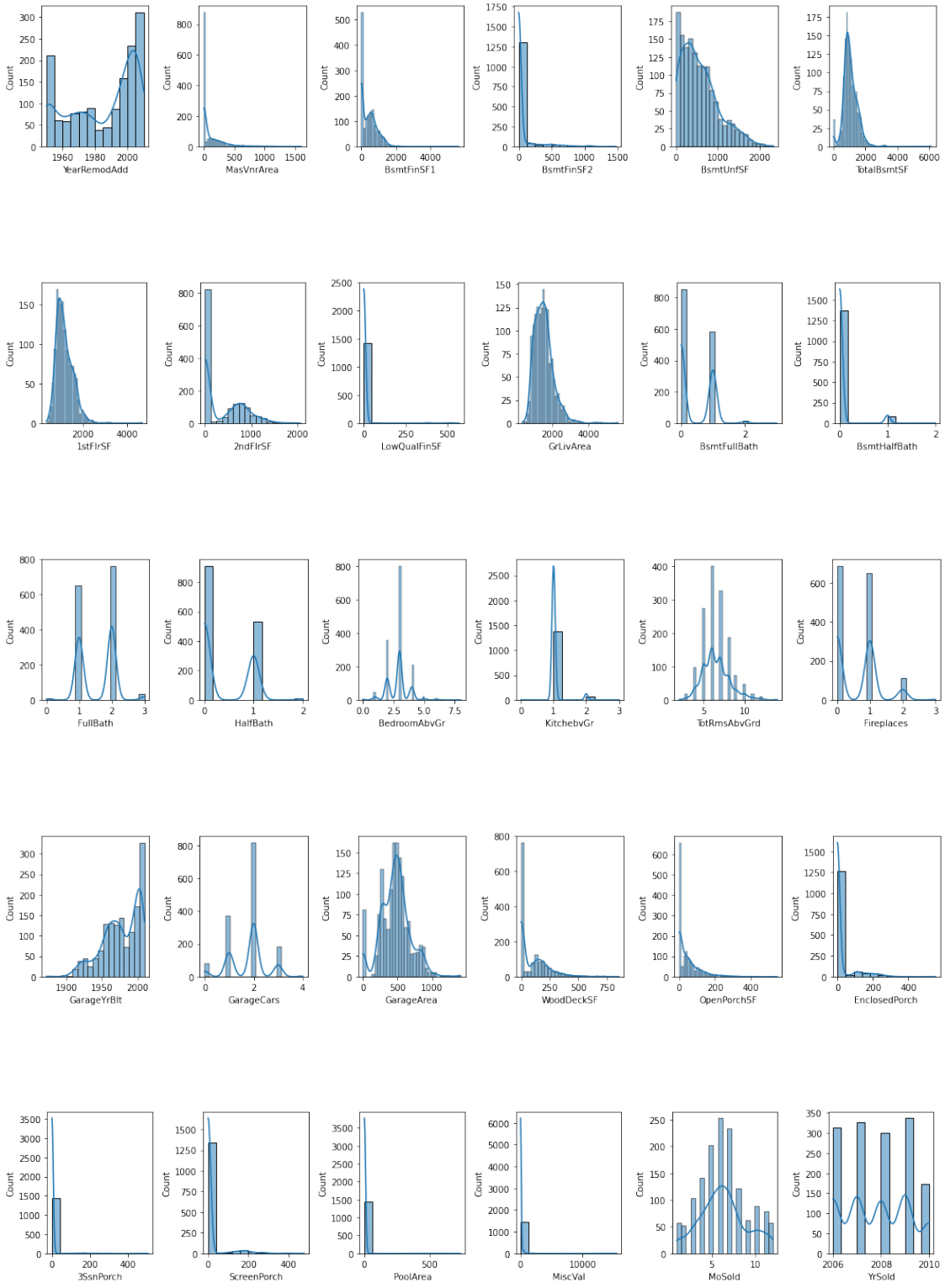
[26]:
```
[<matplotlib.lines.Line2D at 0x7f2d4743d650>]
```



- Visualization of the distribution of the numerical features

[27]:
```
for i in range(0,len(numerical_features_list)-2,6):
        plt.figure(figsize=(15,3))
        plt.subplot(161)
        sns.histplot(numerical_features_df[numerical_features_list[i]],␣
↪kde=True)
        plt.subplot(162)
        sns.histplot(numerical_features_df[numerical_features_list[i+1]],␣
↪kde=True)
        plt.subplot(163)
        sns.histplot(numerical_features_df[numerical_features_list[i+2]],␣
↪kde=True)
        plt.subplot(164)
        sns.histplot(numerical_features_df[numerical_features_list[i+3]],␣
↪kde=True)
        plt.subplot(165)
        sns.histplot(numerical_features_df[numerical_features_list[i+4]],␣
↪kde=True)
        plt.subplot(166)
        sns.histplot(numerical_features_df[numerical_features_list[i+5]],␣
↪kde=True)
        plt.tight_layout()
        plt.show()
```




YearRemodAdd
MasVnrArea
BsmtFinSF1
BsmtFinSF2
BsmtUnfSF
TotalBsmtSF
1stFlrSF
2ndFlrSF
LowQualFinSF
GrLivArea
BsmtFullBath
BsmtHalfBath
FullBath
HalfBath
BedroomAbvGr
KitchenAbvGr
TotRmsAbvGrd
Fireplaces
GarageYrBlt
GarageCars
GarageArea
WoodDeckSF
OpenPorchSF
EnclosedPorch
3SsnPorch
ScreenPorch
PoolArea
MiscVal
MoSold
YrSold
Count

1.3.2 c- Identification of significant variables using a correlation matrix

[28]:
```
corrmatrix = numerical_features_df.corr(method='pearson')
corrmatrix
```

[28]:

	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	\
MSSubClass	1.000000	-0.212759	-0.138054	0.034491	-0.061330	
LotFrontage	-0.212759	1.000000	0.100705	0.172924	-0.055977	
LotArea	-0.138054	0.100705	1.000000	0.106324	-0.002269	
OverallQual	0.034491	0.172924	0.106324	1.000000	-0.090628	
OverallCond	-0.061330	-0.055977	-0.002269	-0.090628	1.000000	
YearBuilt	0.028397	0.034646	0.015639	0.571111	-0.376763	
YearRemodAdd	0.041047	0.077276	0.015126	0.549573	0.075121	
MasVnrArea	0.022936	0.104237	0.104160	0.411876	-0.128101	
BsmtFinSF1	-0.069575	0.077218	0.213063	0.236823	-0.041927	
BsmtFinSF2	-0.066137	-0.009890	0.111686	-0.058039	0.039333	
BsmtUnfSF	-0.138922	0.159377	-0.004227	0.309602	-0.136934	
TotalBsmtSF	-0.236906	0.237426	0.258409	0.537122	-0.167230	
1stFlrSF	-0.250050	0.245865	0.295919	0.476936	-0.138814	
2ndFlrSF	0.308104	0.044036	0.052935	0.298543	0.027473	
LowQualFinSF	0.046413	0.050059	0.004904	-0.029998	0.025140	
GrLivArea	0.076930	0.221295	0.261159	0.594417	-0.076541	
BsmtFullBath	0.003807	0.011913	0.157702	0.108505	-0.051567	
BsmtHalfBath	-0.002633	-0.028136	0.048377	-0.039207	0.117290	
FullBath	0.136306	0.122050	0.122457	0.552266	-0.190396	
HalfBath	0.176165	-0.013566	0.016290	0.271466	-0.061434	
BedroomAbvGr	-0.021651	0.147451	0.117778	0.105900	0.014274	
KitchebvGr	0.286572	0.040195	-0.024697	-0.184642	-0.081254	
TotRmsAbvGrd	0.042406	0.223115	0.187990	0.430549	-0.055964	
Fireplaces	-0.044466	0.047264	0.269643	0.400398	-0.020120	
GarageYrBlt	0.040350	0.042349	0.004387	0.553720	-0.296560	
GarageCars	-0.039043	0.163276	0.154739	0.599734	-0.184866	
GarageArea	-0.098141	0.198667	0.180778	0.560543	-0.151062	
WoodDeckSF	-0.012634	-0.016964	0.173167	0.240652	-0.004530	
OpenPorchSF	-0.005462	0.064812	0.086301	0.303482	-0.031172	
EnclosedPorch	-0.010571	0.031262	-0.023094	-0.112950	0.074731	
3SsnPorch	-0.044049	0.023108	0.020574	0.031029	0.025163	
ScreenPorch	-0.026414	0.022746	0.043511	0.066403	0.054016	
PoolArea	0.008214	0.113478	0.077888	0.065743	-0.002229	
MiscVal	-0.007805	-0.060184	0.038226	-0.031129	0.068642	
MoSold	-0.013840	0.015740	0.003203	0.068760	-0.004034	
YrSold	-0.021529	-0.010378	-0.012977	-0.025186	0.043433	
SalePrice	-0.082813	0.206573	0.264674	0.789997	-0.076294	

	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	BsmtFinSF2	\
MSSubClass	0.028397	0.041047	0.022936	-0.069575	-0.066137	
LotFrontage	0.034646	0.077276	0.104237	0.077218	-0.009890	

```
LotArea          0.015639      0.015126    0.104160    0.213063    0.111686
OverallQual      0.571111      0.549573    0.411876    0.236823   -0.058039
OverallCond     -0.376763      0.075121   -0.128101   -0.041927    0.039333
YearBuilt        1.000000      0.590674    0.315707    0.249239   -0.047816
YearRemodAdd     0.590674      1.000000    0.179618    0.127609   -0.066672
MasVnrArea       0.315707      0.179618    1.000000    0.264736   -0.072319
BsmtFinSF1       0.249239      0.127609    0.264736    1.000000   -0.049287
BsmtFinSF2      -0.047816     -0.066672   -0.072319   -0.049287    1.000000
BsmtUnfSF        0.149810      0.181828    0.114442   -0.496137   -0.209705
TotalBsmtSF      0.392562      0.291492    0.363936    0.520533    0.106309
1stFlrSF         0.284570      0.242488    0.344501    0.443232    0.098824
2ndFlrSF         0.009566      0.140225    0.174561   -0.135715   -0.099560
LowQualFinSF    -0.183749     -0.062045   -0.069071   -0.064345    0.014620
GrLivArea        0.199343      0.288279    0.390857    0.206027   -0.008910
BsmtFullBath     0.186305      0.118169    0.085310    0.647346    0.160189
BsmtHalfBath    -0.037072     -0.011312    0.026673    0.068611    0.070592
FullBath         0.469625      0.440329    0.276833    0.055808   -0.075506
HalfBath         0.240417      0.181063    0.201444    0.001952   -0.031489
BedroomAbvGr    -0.068619     -0.038429    0.102821   -0.105691   -0.016022
KitchebvGr      -0.173951     -0.148527   -0.037610   -0.086473   -0.040459
TotRmsAbvGrd     0.097440      0.193988    0.280682    0.044074   -0.035212
Fireplaces       0.150148      0.114806    0.249070    0.258300    0.047491
GarageYrBlt      0.844290      0.602366    0.271176    0.183446   -0.061006
GarageCars       0.537492      0.419815    0.364204    0.222241   -0.037554
GarageArea       0.478439      0.370674    0.373066    0.295493   -0.017572
WoodDeckSF       0.226891      0.207464    0.159718    0.205350    0.067673
OpenPorchSF      0.185081      0.223491    0.125703    0.107696    0.004294
EnclosedPorch   -0.386839     -0.192367   -0.110204   -0.105608    0.036749
3SsnPorch        0.032037      0.045907    0.018796    0.026995   -0.030186
ScreenPorch     -0.049169     -0.037656    0.061466    0.063299    0.088480
PoolArea         0.005310      0.006145    0.011723    0.141361    0.041610
MiscVal         -0.034048     -0.009927   -0.029815    0.003910    0.004802
MoSold           0.009362      0.018588   -0.005965   -0.016053   -0.014878
YrSold          -0.014441      0.035352   -0.008201    0.016870    0.031851
SalePrice        0.522896      0.507158    0.477493    0.383977   -0.010316

                 …  WoodDeckSF  OpenPorchSF  EnclosedPorch  3SsnPorch  \
MSSubClass       …   -0.012634    -0.005462      -0.010571  -0.044049
LotFrontage      …   -0.016964     0.064812       0.031262   0.023108
LotArea          …    0.173167     0.086301      -0.023094   0.020574
OverallQual      …    0.240652     0.303482      -0.112950   0.031029
OverallCond      …   -0.004530    -0.031172       0.074731   0.025163
YearBuilt        …    0.226891     0.185081      -0.386839   0.032037
YearRemodAdd     …    0.207464     0.223491      -0.192367   0.045907
MasVnrArea       …    0.159718     0.125703      -0.110204   0.018796
BsmtFinSF1       …    0.205350     0.107696      -0.105608   0.026995
BsmtFinSF2       …    0.067673     0.004294       0.036749  -0.030186
```

BsmtUnfSF	…	-0.004192	0.130217	-0.003684	0.020857
TotalBsmtSF	…	0.234182	0.244914	-0.099915	0.037960
1stFlrSF	…	0.238699	0.210625	-0.072610	0.056901
2ndFlrSF	…	0.090962	0.210512	0.064217	-0.024422
LowQualFinSF	…	-0.025669	0.018852	0.061314	-0.004373
GrLivArea	…	0.247981	0.330795	0.005813	0.021000
BsmtFullBath	…	0.175778	0.063937	-0.051483	0.000296
BsmtHalfBath	…	0.039929	-0.024489	-0.008518	0.034966
FullBath	…	0.189982	0.261509	-0.120246	0.036004
HalfBath	…	0.107275	0.196968	-0.093258	-0.004679
BedroomAbvGr	…	0.045614	0.098687	0.038447	-0.024667
KitchebvGr	…	-0.088863	-0.067892	0.028587	-0.024534
TotRmsAbvGrd	…	0.165236	0.237234	0.000861	-0.006657
Fireplaces	…	0.198180	0.170942	-0.029461	0.011447
GarageYrBlt	…	0.239545	0.185259	-0.312712	0.030996
GarageCars	…	0.226669	0.211257	-0.151857	0.036116
GarageArea	…	0.225418	0.238895	-0.121603	0.035410
WoodDeckSF	…	1.000000	0.058911	-0.125486	-0.033008
OpenPorchSF	…	0.058911	1.000000	-0.090870	-0.005401
EnclosedPorch	…	-0.125486	-0.090870	1.000000	-0.037395
3SsnPorch	…	-0.033008	-0.005401	-0.037395	1.000000
ScreenPorch	…	-0.074740	0.075865	-0.083074	-0.031617
PoolArea	…	0.073454	0.061403	0.054397	-0.008036
MiscVal	…	-0.009694	-0.018335	0.018445	0.000298
MoSold	…	0.021789	0.068538	-0.025830	0.029761
YrSold	…	0.021575	-0.055585	-0.008496	0.018714
SalePrice	…	0.324650	0.311268	-0.128778	0.045247

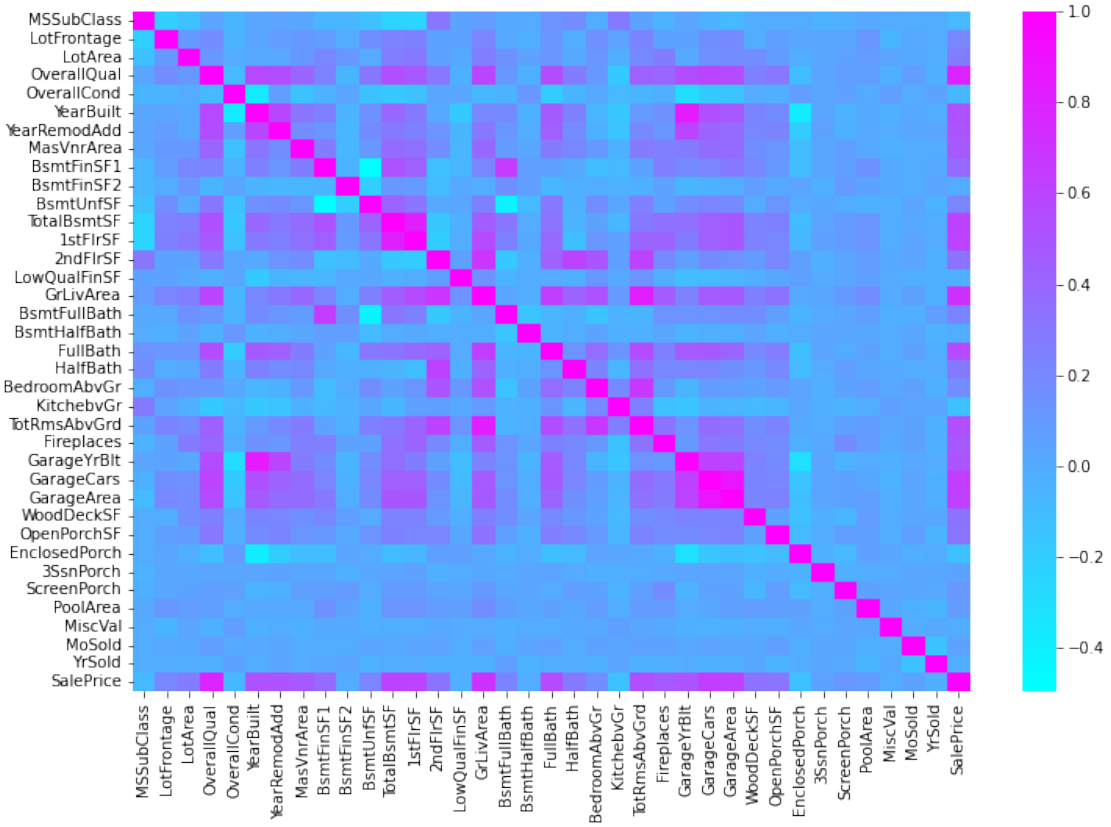
	ScreenPorch	PoolArea	MiscVal	MoSold	YrSold	SalePrice
MSSubClass	-0.026414	0.008214	-0.007805	-0.013840	-0.021529	-0.082813
LotFrontage	0.022746	0.113478	-0.060184	0.015740	-0.010378	0.206573
LotArea	0.043511	0.077888	0.038226	0.003203	-0.012977	0.264674
OverallQual	0.066403	0.065743	-0.031129	0.068760	-0.025186	0.789997
OverallCond	0.054016	-0.002229	0.068642	-0.004034	0.043433	-0.076294
YearBuilt	-0.049169	0.005310	-0.034048	0.009362	-0.014441	0.522896
YearRemodAdd	-0.037656	0.006145	-0.009927	0.018588	0.035352	0.507158
MasVnrArea	0.061466	0.011723	-0.029815	-0.005965	-0.008201	0.477493
BsmtFinSF1	0.063299	0.141361	0.003910	-0.016053	0.016870	0.383977
BsmtFinSF2	0.088480	0.041610	0.004802	-0.014878	0.031851	-0.010316
BsmtUnfSF	-0.012506	-0.035146	-0.023857	0.033432	-0.040377	0.215740
TotalBsmtSF	0.085831	0.126820	-0.018237	0.011558	-0.011451	0.612971
1stFlrSF	0.090338	0.132669	-0.020931	0.031148	-0.009063	0.606849
2ndFlrSF	0.040771	0.081749	0.016257	0.039782	-0.031893	0.322710
LowQualFinSF	0.026627	0.062115	-0.003851	-0.022102	-0.028954	-0.025263
GrLivArea	0.102489	0.170808	-0.002192	0.053792	-0.035801	0.710080
BsmtFullBath	0.024157	0.068057	-0.022813	-0.024940	0.067489	0.225027
BsmtHalfBath	0.031774	0.019937	-0.007484	0.033352	-0.046571	-0.015993

```
FullBath         -0.006959  0.050103 -0.013964  0.058944 -0.019985   0.562491
HalfBath          0.073391  0.022636  0.001528 -0.008772 -0.010056   0.282040
BedroomAbvGr      0.044270  0.070928  0.007728  0.052450 -0.038584   0.171934
KitchebvGr       -0.051430 -0.014485  0.062926  0.031032  0.033943  -0.137419
TotRmsAbvGrd      0.059632  0.083979  0.024853  0.041611 -0.034886   0.536311
Fireplaces        0.185752  0.095602  0.001518  0.052030 -0.024917   0.468930
GarageYrBlt      -0.046853 -0.007671 -0.032264  0.008096 -0.010436   0.507855
GarageCars        0.051277  0.021140 -0.042900  0.039393 -0.038065   0.639686
GarageArea        0.052130  0.061292 -0.027230  0.026719 -0.025754   0.622492
WoodDeckSF       -0.074740  0.073454 -0.009694  0.021789  0.021575   0.324650
OpenPorchSF       0.075865  0.061403 -0.018335  0.068538 -0.055585   0.311268
EnclosedPorch    -0.083074  0.054397  0.018445 -0.025830 -0.008496  -0.128778
3SsnPorch        -0.031617 -0.008036  0.000298  0.029761  0.018714   0.045247
ScreenPorch       1.000000  0.051216  0.031822  0.023695  0.010786   0.113044
PoolArea          0.051216  1.000000  0.029636 -0.033785 -0.059800   0.093109
MiscVal           0.031822  0.029636  1.000000 -0.006400  0.004938  -0.020951
MoSold            0.023695 -0.033785 -0.006400  1.000000 -0.145367   0.045136
YrSold            0.010786 -0.059800  0.004938 -0.145367  1.000000  -0.026180
SalePrice         0.113044  0.093109 -0.020951  0.045136 -0.026180   1.000000

[37 rows x 37 columns]
```

[29]:
```
plt.figure(figsize=(12,8))
sns.heatmap(numerical_features_df.corr(), cmap='cool')
```

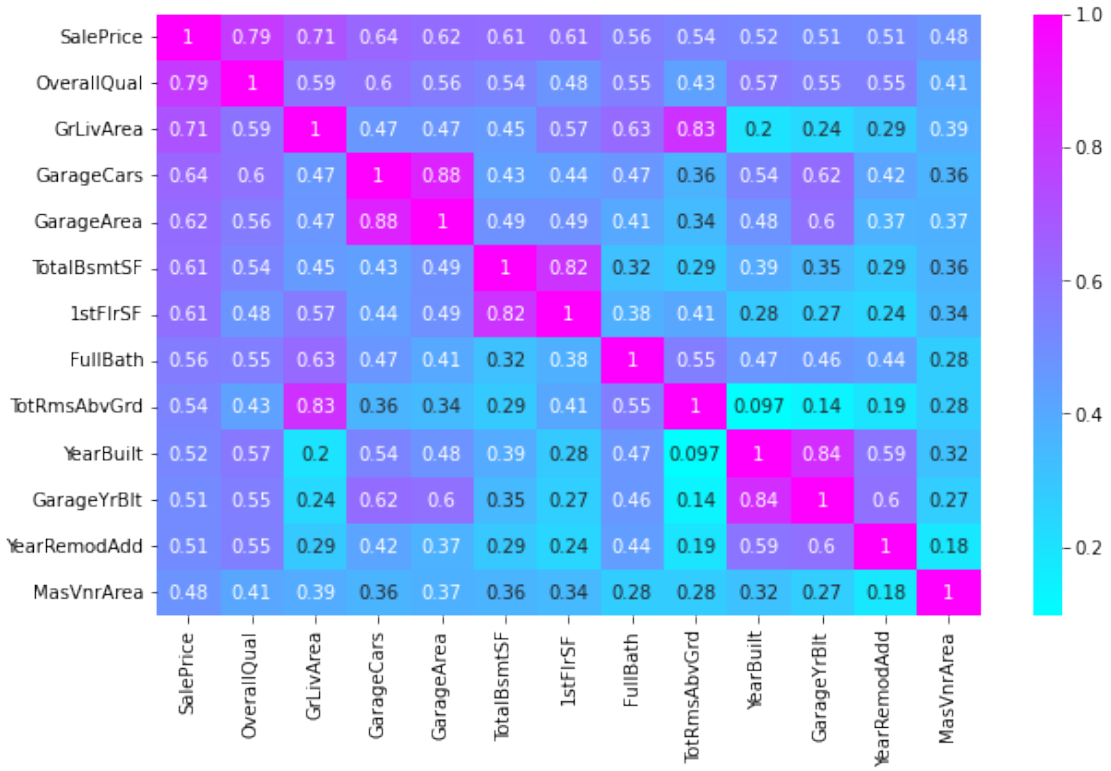
[29]: <AxesSubplot:>



- Visualization of the variables most correlated to the sale price.

[30]:
```
columns = numerical_features_df.corr().nlargest(13, 'SalePrice')['SalePrice'].
↪index
cm = numerical_features_df[columns].corr()
plt.figure(figsize=(10,6))
sns.heatmap(cm, annot=True, cmap = 'cool')
```

[30]: <AxesSubplot:>

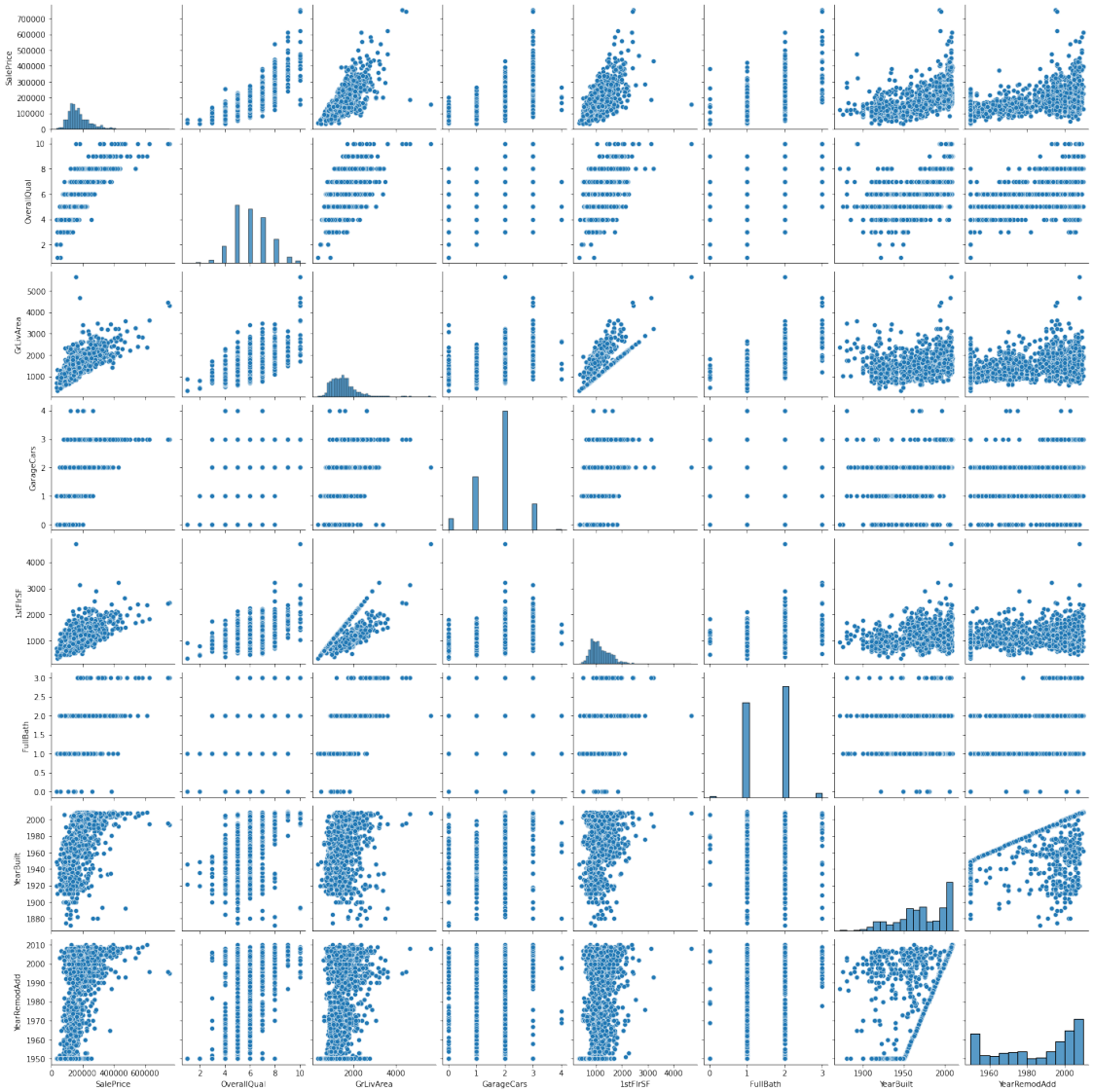


Observations: - The 12 features in this heatmap are the most correlated to "SalePrice". "MasVnArea" value is less than 0.5 so we will discard this feature. We will drop the other 24 features and keep the 11 features above. - "GarageCars" and "GarageArea" are highly correlated so we will keep only "GarageCars". - "YearBuilt" and "GarageYrBuilt" are highly correlated so we will keep only "YearBuilt". - "GrLivArea" and "TotRmsAbvGrd" are highly correlated so we will keep only "GrLivArea". - "TotalBsmtSF" and "1stFlrSF" are highly correlated so we will keep only "1stFlrSF". - We will only keep 7 features : "OverallQual", "GrLivArea", "GarageCars", "1stFlrSF", "FullBath", "YearBuilt", "YearRemodAdd".

1.3.3 d- Pairplot for distribution and density

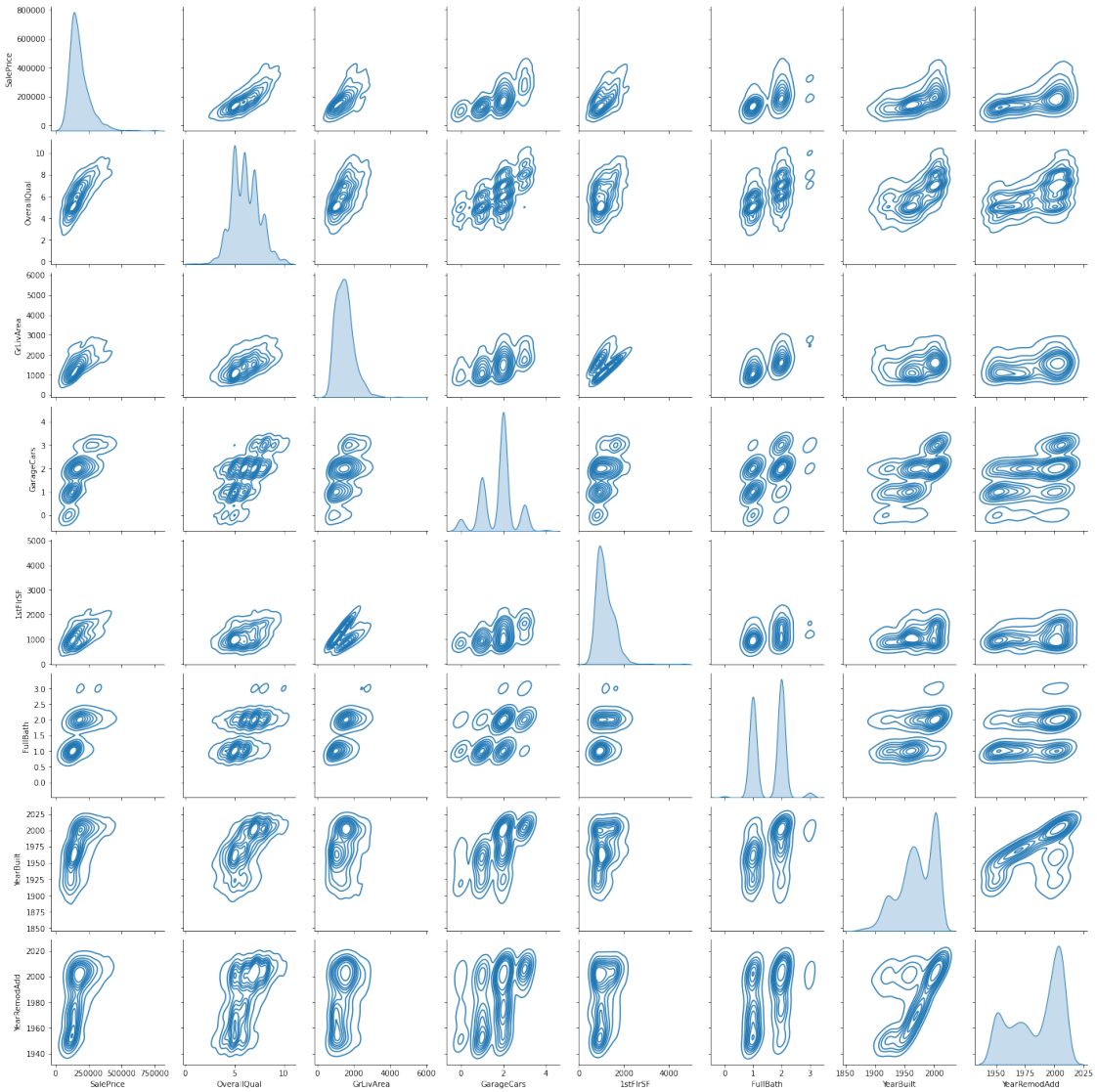
[31]:
```
cols = ["SalePrice", "OverallQual", "GrLivArea", "GarageCars", "1stFlrSF",␣
↪"FullBath", "YearBuilt", "YearRemodAdd"]
sns.pairplot(numerical_features_df[cols])
```

[31]: <seaborn.axisgrid.PairGrid at 0x7f2d44c9d050>



[32]:
```
sns.pairplot(numerical_features_df[cols], kind="kde")
```

[32]: <seaborn.axisgrid.PairGrid at 0x7f2d40de4390>



- Generation of the final dataset for numerical features

[33]:
```
final_numerical_features_df = numerical_features_df[cols]
final_numerical_features_df
```

[33]:

Id	SalePrice	OverallQual	GrLivArea	GarageCars	1stFlrSF	FullBath	\
1	208500	7	1710	2	856	2	
2	181500	6	1262	2	1262	2	
3	223500	7	1786	2	920	2	
4	140000	7	1717	3	961	1	
5	250000	8	2198	3	1145	2	
…	…	…	…	…	…	…	

```
1456     175000            6       1647            2        953          2
1457     210000            6       2073            2       2073          2
1458     266500            7       2340            1       1188          2
1459     142125            5       1078            1       1078          1
1460     147500            5       1256            1       1256          1

      YearBuilt  YearRemodAdd
Id
1          2003          2003
2          1976          1976
3          2001          2002
4          1915          1970
5          2000          2000
…           …             …
1456       1999          2000
1457       1978          1988
1458       1941          2006
1459       1950          1996
1460       1965          1965

[1452 rows x 8 columns]
```

1.4 4- EDA of categorical variables

1.4.1 a-Treatment of missing values

- Visualization of the missing values

[34]:
```
categorical_features_df.isna().sum(axis=0)
```

[34]:
```
MSZoning            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
RoofStyle           0
RoofMatl            0
Exterior1st         0
```

```
Exterior2nd         0
MasVnrType          8
ExterQual           0
ExterCond           0
Foundation          0
BsmtQual           37
BsmtCond           37
BsmtExposure       38
BsmtFinType1       37
BsmtFinType2       38
Heating             0
HeatingQC           0
CentralAir          0
Electrical          1
KitchenQual         0
Functiol            0
FireplaceQu       690
GarageType         81
GarageFinish       81
GarageQual         81
GarageCond         81
PavedDrive          0
PoolQC           1453
Fence            1179
MiscFeature      1406
SaleType            0
SaleCondition       0
dtype: int64
```

- Calculation of the percentage of missing value on each columns

[35]:
```
for name in categorical_features_list:
    percentage_of_missing_values = (categorical_features_df[name].isna().
↪sum(axis=0)/categorical_features_df.shape[0])*100
    print(f"{name} : {percentage_of_missing_values} %")
```

```
MSZoning : 0.0 %
Street : 0.0 %
Alley : 93.76712328767123 %
LotShape : 0.0 %
LandContour : 0.0 %
Utilities : 0.0 %
LotConfig : 0.0 %
LandSlope : 0.0 %
Neighborhood : 0.0 %
Condition1 : 0.0 %
Condition2 : 0.0 %
BldgType : 0.0 %
```

```
HouseStyle : 0.0 %
RoofStyle : 0.0 %
RoofMatl : 0.0 %
Exterior1st : 0.0 %
Exterior2nd : 0.0 %
MasVnrType : 0.547945205479452 %
ExterQual : 0.0 %
ExterCond : 0.0 %
Foundation : 0.0 %
BsmtQual : 2.5342465753424657 %
BsmtCond : 2.5342465753424657 %
BsmtExposure : 2.6027397260273974 %
BsmtFinType1 : 2.5342465753424657 %
BsmtFinType2 : 2.6027397260273974 %
Heating : 0.0 %
HeatingQC : 0.0 %
CentralAir : 0.0 %
Electrical : 0.0684931506849315 %
KitchenQual : 0.0 %
Functiol : 0.0 %
FireplaceQu : 47.26027397260274 %
GarageType : 5.5479452054794525 %
GarageFinish : 5.5479452054794525 %
GarageQual : 5.5479452054794525 %
GarageCond : 5.5479452054794525 %
PavedDrive : 0.0 %
PoolQC : 99.52054794520548 %
Fence : 80.75342465753424 %
MiscFeature : 96.30136986301369 %
SaleType : 0.0 %
SaleCondition : 0.0 %
```

Observations: ”Alley”, ”PoolQC”, ”Fence”, ”MiscFeature” have more than 80% of missing values, so we will drop those features.

- Generation of the new dataset

[36]:
```
categorical_features_df = categorical_features_df.drop(columns=["Alley",␣
↪"PoolQC", "Fence", "MiscFeature"])
categorical_features_df.shape
```

[36]:
```
(1460, 39)
```

- Treatment of the missing value in ”FireplaceQu”

[37]:
```
categorical_features_df["FireplaceQu"].unique()
```

[37]:
```
array([nan, 'TA', 'Gd', 'Fa', 'Ex', 'Po'], dtype=object)
```

[38]:
```
categorical_features_df["FireplaceQu"].value_counts()
```

[38]:
```
Gd    380
TA    313
Fa     33
Ex     24
Po     20
Name: FireplaceQu, dtype: int64
```

Observations: the 2 main values are ”Gd” and ”TA”. Since ”TA” means ”prefabricated fireplace in the main living area or masonry Fireplace in the basement”, we will set the missing values to ”TA”.

- Setting the missing values in ”FireplaceQu” to ”TA”

[39]:
```
categorical_features_df["FireplaceQu"].fillna(value="TA", inplace=True)
categorical_features_df["FireplaceQu"].isna().sum(axis=0)
```

[39]:
```
0
```

Observation: There are no missing values in ”FireplaceQu” feature.

- Dropping the rows with missing values

[40]:
```
categorical_features_df.shape
```

[40]:
```
(1460, 39)
```

[41]:
```
categorical_features_df = categorical_features_df.dropna(axis=0)
categorical_features_df.shape
```

[41]:
```
(1338, 39)
```

Observation: 122 rows were dropped which represents about 8.3% of the rows.

[42]:
```
categorical_features_df.isna().sum(axis=0)
```

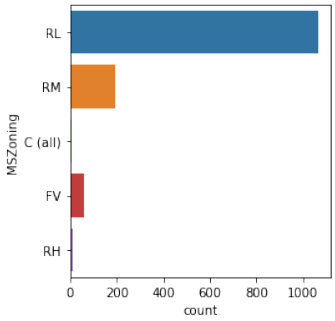
[42]:
```
MSZoning        0
Street          0
LotShape        0
LandContour     0
Utilities       0
LotConfig       0
LandSlope       0
Neighborhood    0
Condition1      0
Condition2      0
BldgType        0
HouseStyle      0
RoofStyle       0
```

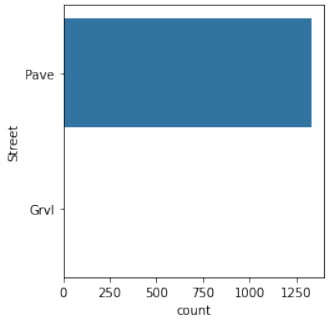
```
RoofMatl         0
Exterior1st      0
Exterior2nd      0
MasVnrType       0
ExterQual        0
ExterCond        0
Foundation       0
BsmtQual         0
BsmtCond         0
BsmtExposure     0
BsmtFinType1     0
BsmtFinType2     0
Heating          0
HeatingQC        0
CentralAir       0
Electrical       0
KitchenQual      0
Functiol         0
FireplaceQu      0
GarageType       0
GarageFinish     0
GarageQual       0
GarageCond       0
PavedDrive       0
SaleType         0
SaleCondition    0
dtype: int64
```

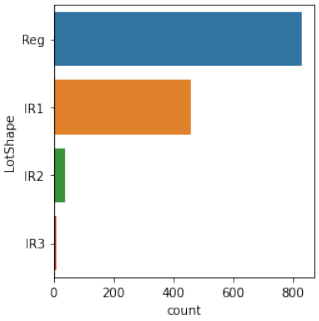
Observation: There are no missing values.

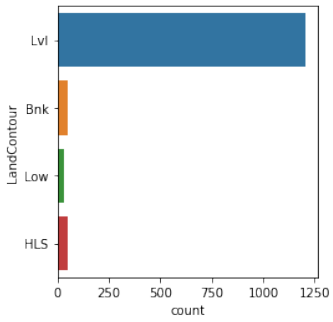
1.4.2 b- Drawing of the count plot and boxplot of the categorical features

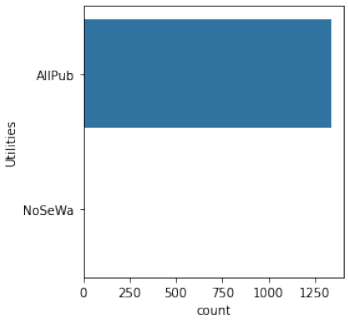
[43]:
```
for i in range(0,len(list(categorical_features_df.columns))-1,3):
    plt.figure(figsize=(15,4))
    plt.subplot(131)
    sns.countplot(y=list(categorical_features_df.columns)[i],␣
 ↪data=categorical_features_df)
    plt.subplot(132)
    sns.countplot(y=list(categorical_features_df.columns)[i+1],␣
 ↪data=categorical_features_df)
    plt.subplot(133)
    sns.countplot(y=list(categorical_features_df.columns)[i+2],␣
 ↪data=categorical_features_df)
    plt.subplots_adjust(wspace=0.5)
```

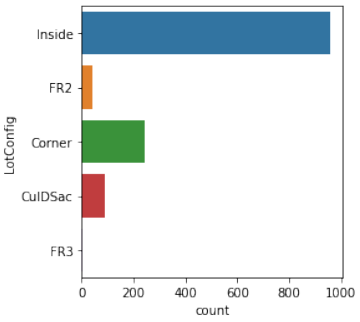

MSZoning
RL
RM
C (all)
FV
RH
0
200
400
600
800
1000
count

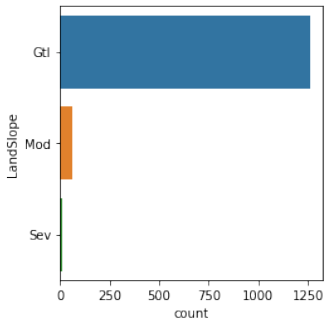

Street
Pave
Grvl
0
250
500
750
1000
1250
count

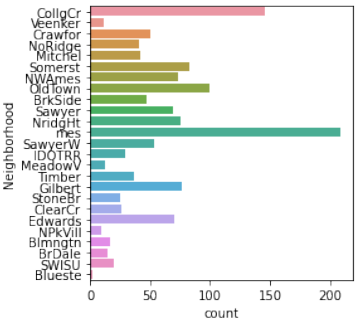

LotShape
Reg
IR1
IR2
IR3
0
200
400
600
800
count

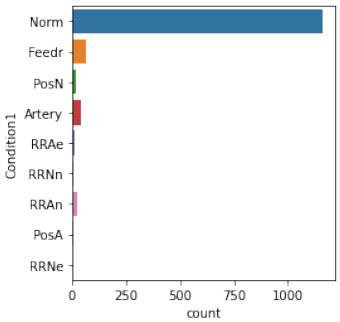

LandContour
Lvl
Bnk
Low
HLS
0
250
500
750
1000
1250
count

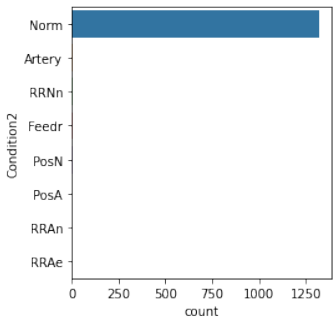

Utilities
AllPub
NoSeWa
0
250
500
750
1000
1250
count

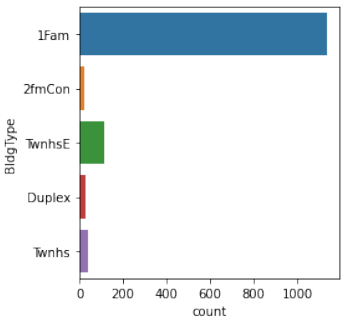

LotConfig
Inside
FR2
Corner
CulDSac
FR3
0
200
400
600
800
1000
count

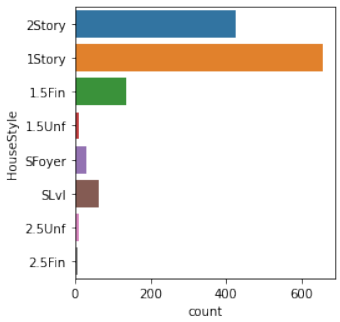

LandSlope
Gtl
Mod
Sev
0
250
500
750
1000
1250
count

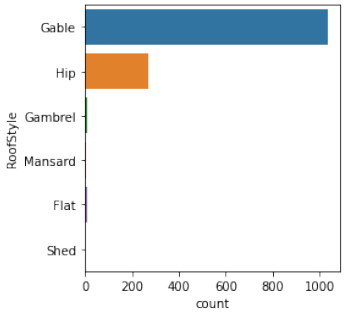

Neighborhood
CollgCr
Veenker
Crawfor
NoRidge
Mitchel
Somerst
NWAmes
OldTown
BrkSide
Sawyer
NridgHt
Names
SawyerW
IDOTRR
MeadowV
Timber
Gilbert
StoneBr
ClearCr
Edwards
NPkVill
Blmngtn
BrDale
SWISU
Blueste
0
50
100
150
200
count

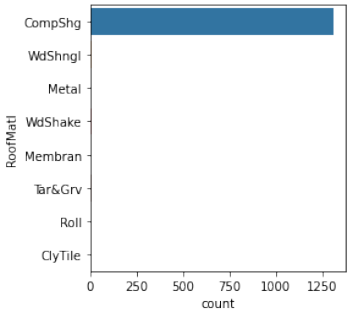

Condition1
Norm
Feedr
PosN
Artery
RRAe
RRNn
RRAn
PosA
RRNe
0
250
500
750
1000
count

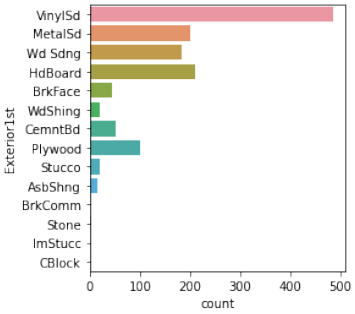

Condition2
Norm
Artery
RRNn
Feedr
PosN
PosA
RRAn
RRAe
0
250
500
750
1000
1250
count

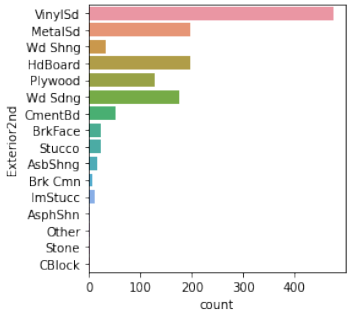

BldgType
1Fam
2fmCon
TwnhsE
Duplex
Twnhs
0
200
400
600
800
1000
count

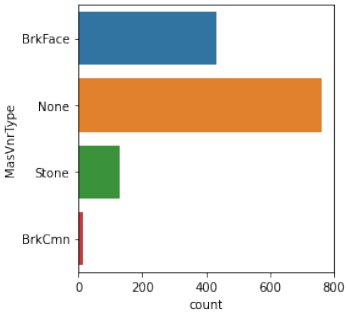

HouseStyle
2Story
1Story
1.5Fin
1.5Unf
SFoyer
SLvl
2.5Unf
2.5Fin
0
200
400
600
count

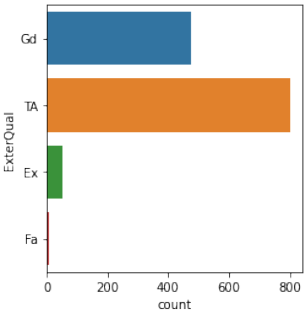

RoofStyle
Gable
Hip
Gambrel
Mansard
Flat
Shed
0
200
400
600
800
1000
count

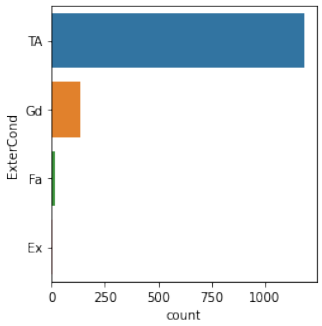

RoofMatl
CompShg
WdShngl
Metal
WdShake
Membran
Tar&Grv
Roll
ClyTile
0
250
500
750
1000
1250
count

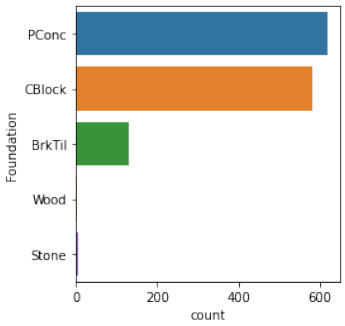

Exterior1st
VinylSd
MetalSd
Wd Sdng
HdBoard
BrkFace
WdShing
CemntBd
Plywood
Stucco
AsbShng
BrkComm
Stone
ImStucc
CBlock
0
100
200
300
400
500
count

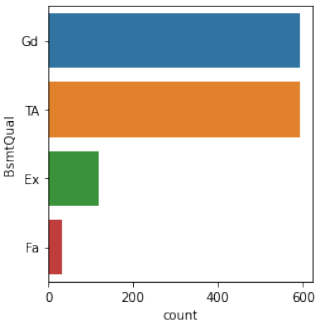

Exterior2nd
VinylSd
MetalSd
Wd Shng
HdBoard
Plywood
Wd Sdng
CmentBd
BrkFace
Stucco
AsbShng
Brk Cmn
ImStucc
AsphShn
Other
Stone
CBlock
0
100
200
300
400
count

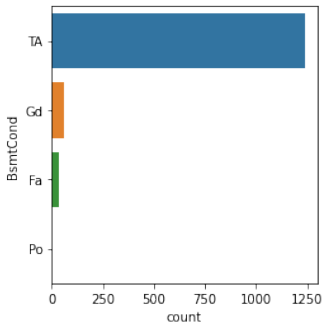

MasVnrType
BrkFace
None
Stone
BrkCmn
0
200
400
600
800
count

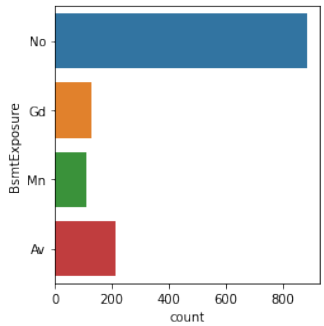

ExterQual
Gd
TA
Ex
Fa
0
200
400
600
800
count

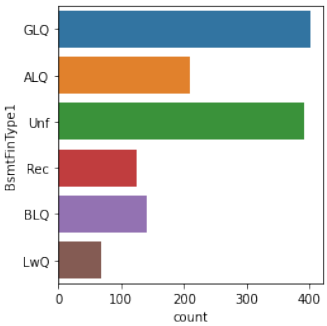

ExterCond
TA
Gd
Fa
Ex
0
250
500
750
1000
count

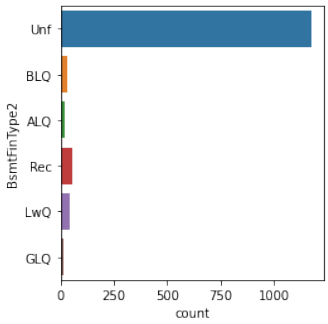

Foundation
PConc
CBlock
BrkTil
Wood
Stone
0
200
400
600
count

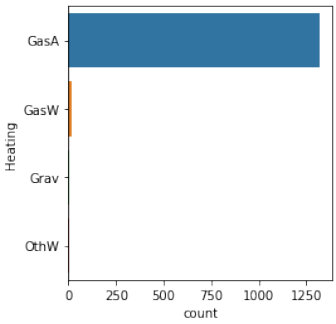

BsmtQual
Gd
TA
Ex
Fa
0
200
400
600
count

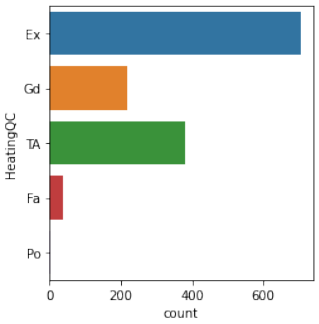

BsmtCond
TA
Gd
Fa
Po
0
250
500
750
1000
1250
count

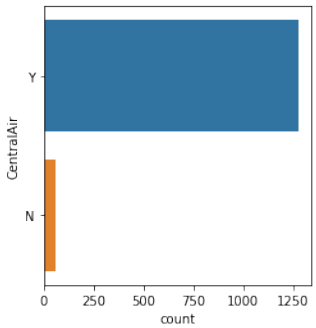

BsmtExposure
No
Gd
Mn
Av
0
200
400
600
800
count

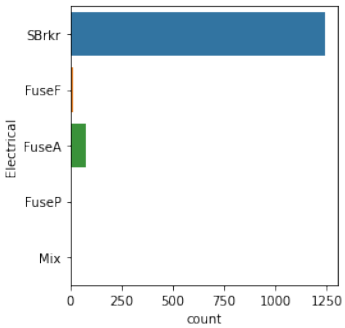

BsmtFinType1
GLQ
ALQ
Unf
Rec
BLQ
LwQ
0
100
200
300
400
count

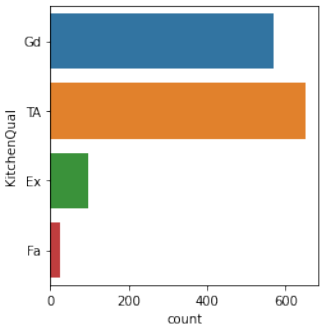

BsmtFinType2
Unf
BLQ
ALQ
Rec
LwQ
GLQ
0
250
500
750
1000
count

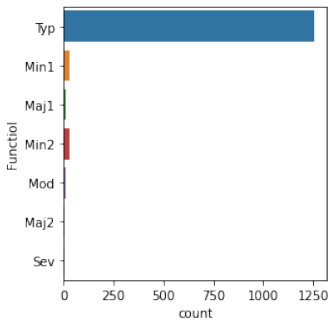

Heating
GasA
GasW
Grav
OthW
0
250
500
750
1000
1250
count

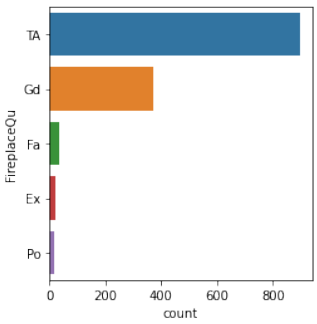

HeatingQC
Ex
Gd
TA
Fa
Po
0
200
400
600
count

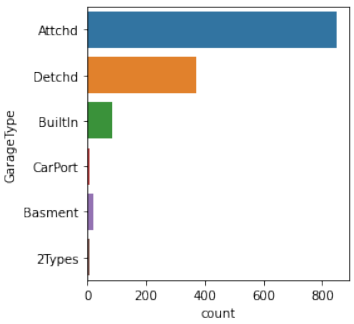

CentralAir
Y
N
0
250
500
750
1000
1250
count

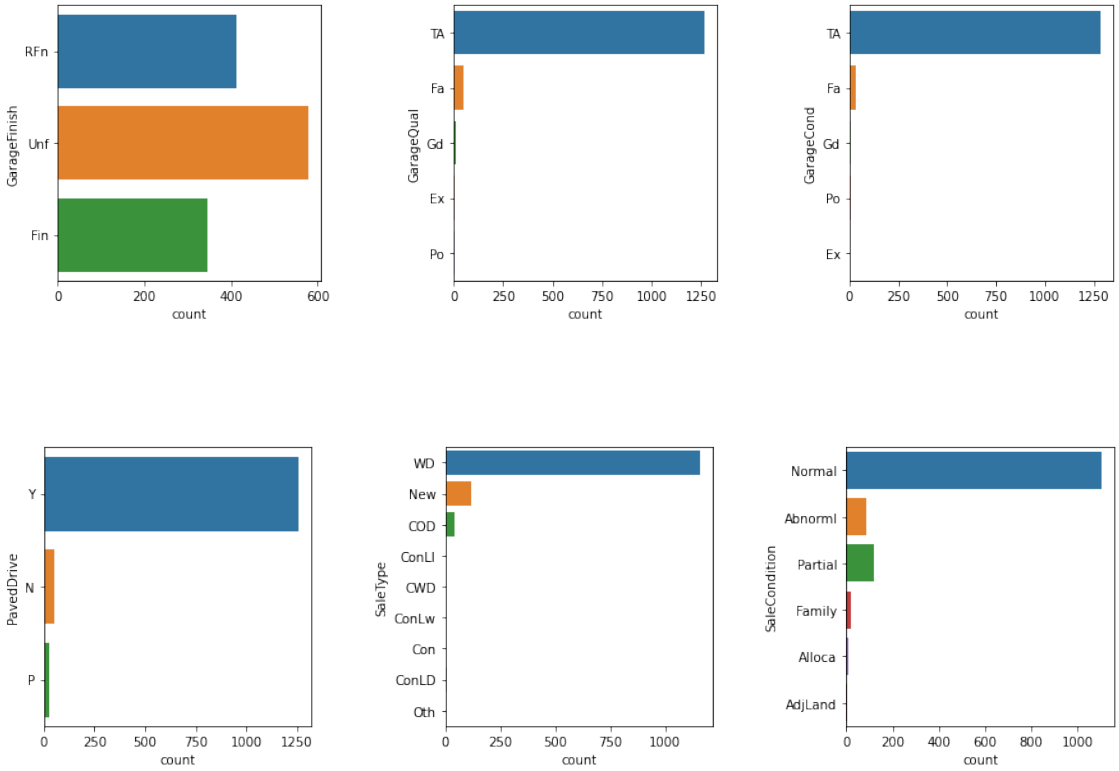

Electrical
SBrkr
FuseF
FuseA
FuseP
Mix
0
250
500
750
1000
1250
count


KitchenQual
Gd
TA
Ex
Fa
0
200
400
600
count

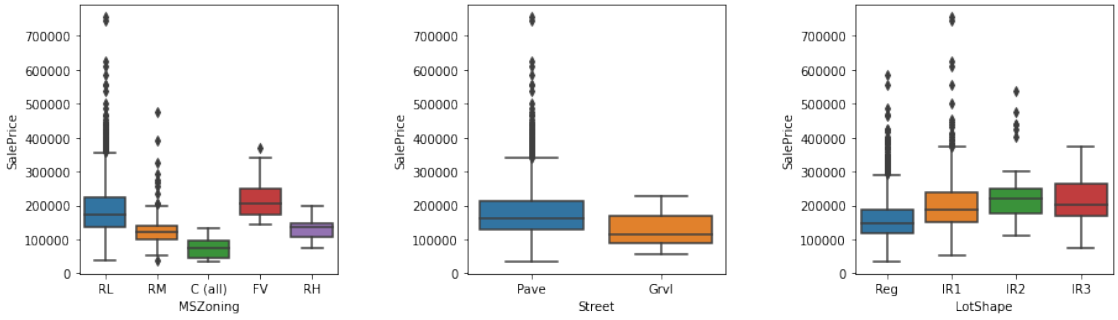

Functiol
Typ
Min1
Maj1
Min2
Mod
Maj2
Sev
0
250
500
750
1000
1250
count

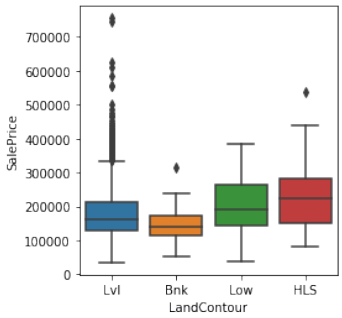

FireplaceQu
TA
Gd
Fa
Ex
Po
0
200
400
600
800
count

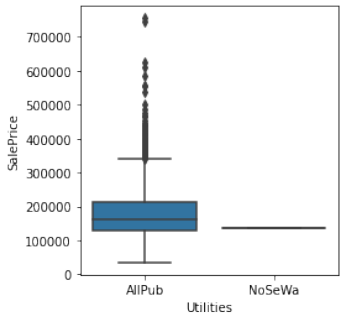

GarageType
Attchd
Detchd
BuiltIn
CarPort
Basment
2Types
0
200
400
600
800
count

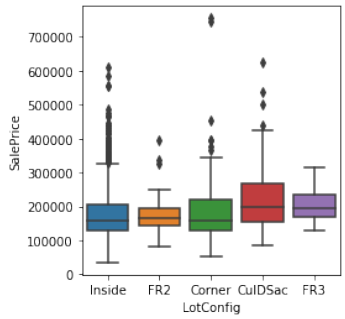


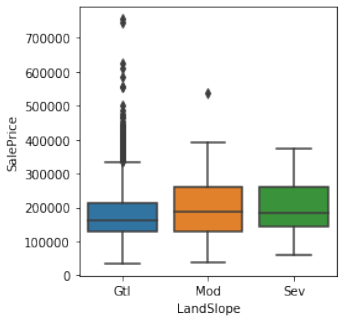
[44]:
```
for i in range(0,len(list(categorical_features_df.columns))-1,3):
    plt.figure(figsize=(15,4))
    plt.subplot(131)
    sns.boxplot(x=list(categorical_features_df.
↪columns)[i],y="SalePrice",data=data)
    plt.subplot(132)
    sns.boxplot(x=list(categorical_features_df.
↪columns)[i+1],y="SalePrice",data=data)
    plt.subplot(133)
    sns.boxplot(x=list(categorical_features_df.
↪columns)[i+2],y="SalePrice",data=data)
    plt.subplots_adjust(wspace=0.5)
```

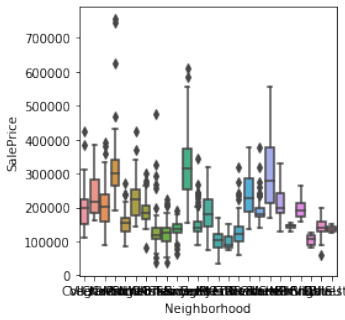


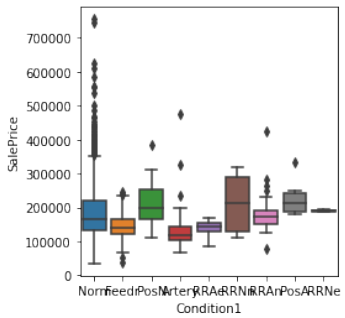

SalePrice
LandContour
Lvl
Bnk
Low
HLS

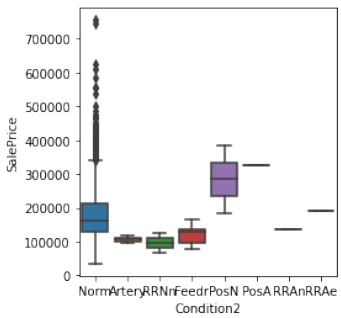

SalePrice
Utilities
AllPub
NoSeWa

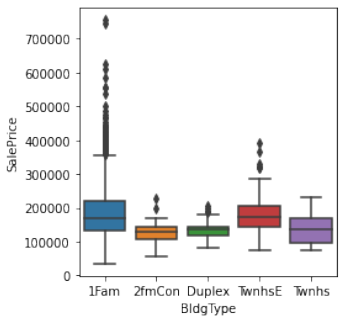

SalePrice
LotConfig
Inside
FR2
Corner
CulDSac
FR3

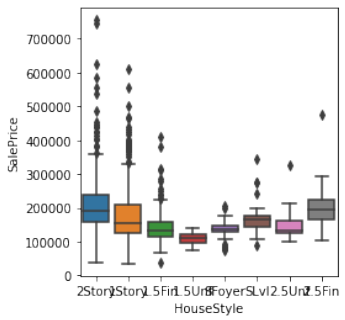

SalePrice
LandSlope
Gtl
Mod
Sev

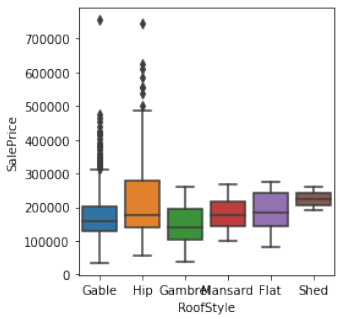

SalePrice
Neighborhood

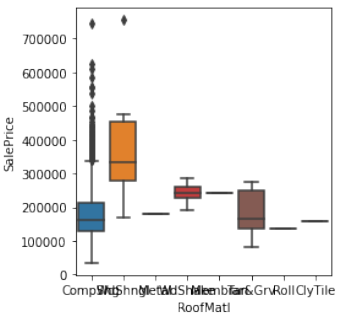

SalePrice
Condition1

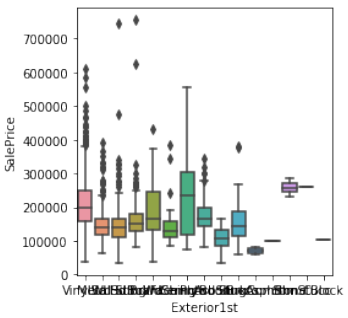

SalePrice
Condition2

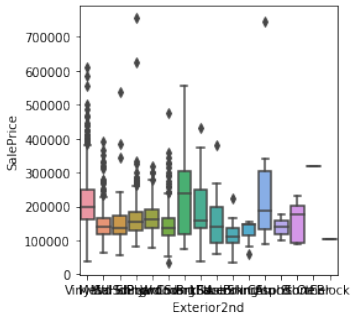

SalePrice
BldgType
1Fam
2fmCon
Duplex
TwnhsE
Twnhs

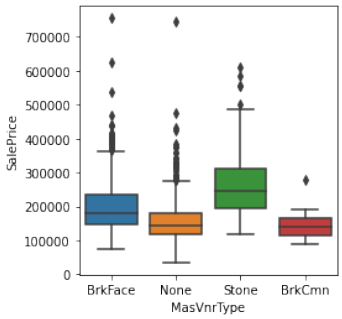

SalePrice
HouseStyle

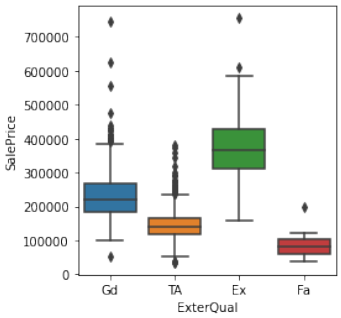

SalePrice
RoofStyle
Gable
Hip
Mansard
Flat
Shed

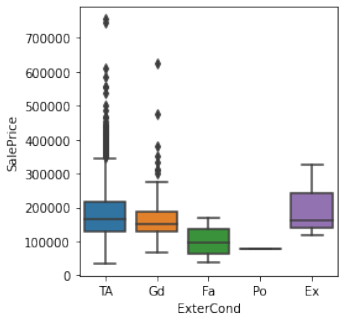

SalePrice
RoofMatl
CompShg
Tar&Grv
Roll
ClyTile

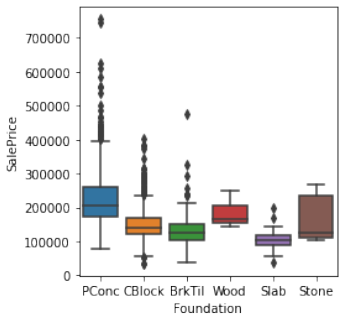

SalePrice
Exterior1st

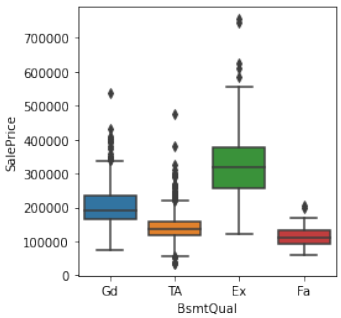

SalePrice
Exterior2nd

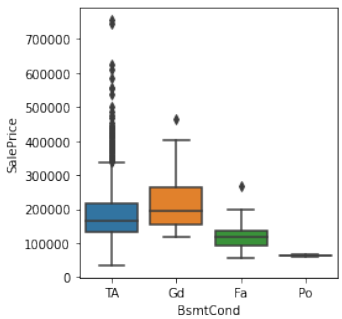

SalePrice
BrkFace
None
Stone
BrkCmn
MasVnrType

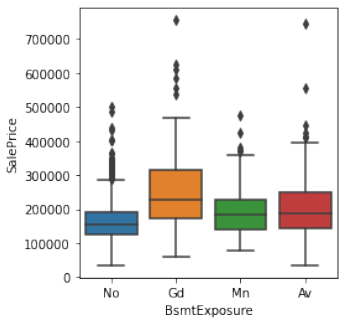

SalePrice
Gd
TA
Ex
Fa
ExterQual

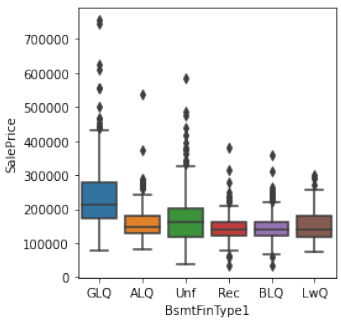

SalePrice
TA
Gd
Fa
Po
Ex
ExterCond

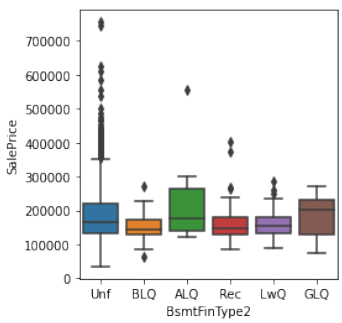

SalePrice
PConc
CBlock
BrkTil
Wood
Slab
Stone
Foundation

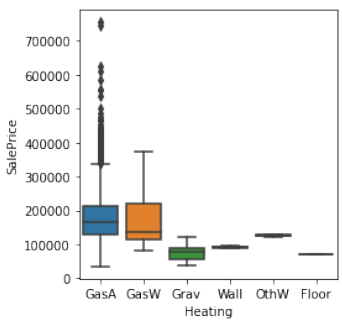

SalePrice
Gd
TA
Ex
Fa
BsmtQual

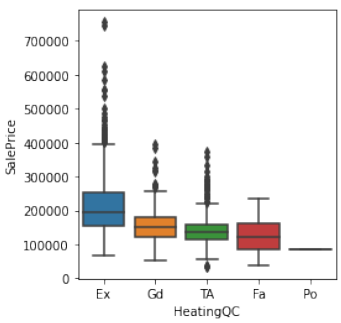

SalePrice
TA
Gd
Fa
Po
BsmtCond

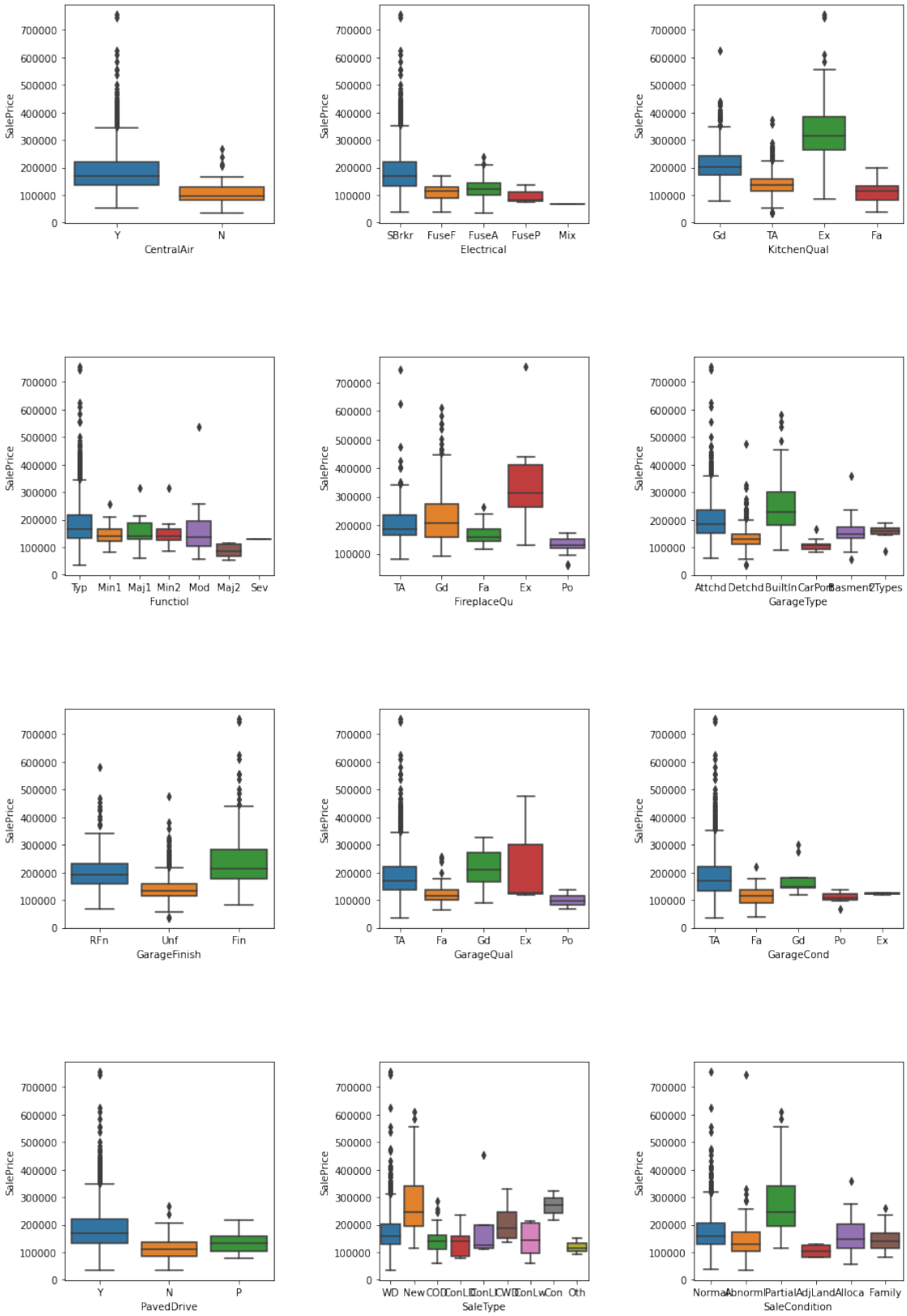

SalePrice
No
Gd
Mn
Av
BsmtExposure


SalePrice
GLQ
ALQ
Unf
Rec
BLQ
LwQ
BsmtFinType1


SalePrice
Unf
BLQ
ALQ
Rec
LwQ
GLQ
BsmtFinType2


SalePrice
GasA
GasW
Grav
Wall
OthW
Floor
Heating


SalePrice
Ex
Gd
TA
Fa
Po
HeatingQC


SalePrice
CentralAir
Y
N
Electrical
SBrkr
FuseF
FuseA
FuseP
Mix
KitchenQual
Gd
TA
Ex
Fa
Functiol
Typ
Min1
Maj1
Min2
Mod
Maj2
Sev
FireplaceQu
TA
Gd
Fa
Ex
Po
GarageType
Attchd
Detchd
BuiltIn
CarPort
Basment
2Types
GarageFinish
RFn
Unf
Fin
GarageQual
TA
Fa
Gd
Ex
Po
GarageCond
TA
Fa
Gd
Po
Ex
PavedDrive
Y
N
P
SaleType
WD
New
COD
Con
SaleCondition
Normal
Abnorml
Partial
Family

1.4.3 c- Identification of the significant features using p-values and Chi-Square values

[45]:
```
from scipy import stats

final_categorical_features_list=[]
for name in list(categorical_features_df.columns):
    crosstab = pd.crosstab(data[name], data["SalePrice"])
    stats.chi2_contingency(crosstab)
    if stats.chi2_contingency(crosstab)[1]<0.05:
        final_categorical_features_list.append(name)
        print(f"For {name} the Chi-Square value= {stats.
 ↪chi2_contingency(crosstab)[0]} and p-value ={stats.
 ↪chi2_contingency(crosstab)[1]}")
```

```
For MSZoning the Chi-Square value= 3147.8911158183737 and p-value
=4.3483250606822396e-11
For Street the Chi-Square value= 888.3129945096931 and p-value
=8.338870380464053e-09
For LotShape the Chi-Square value= 2446.2353573800365 and p-value
=4.724729155980402e-12
For LotConfig the Chi-Square value= 2771.9854545078742 and p-value
=0.045806211958033756
For Neighborhood the Chi-Square value= 16898.75578956907 and p-value
=1.364960102688296e-08
For MasVnrType the Chi-Square value= 2280.7764593012926 and p-value
=9.975416440708533e-07
For ExterQual the Chi-Square value= 2849.766648231622 and p-value
=4.250289171585687e-34
For ExterCond the Chi-Square value= 3192.8402481986723 and p-value
=9.869790306250171e-13
For Foundation the Chi-Square value= 3669.1922311365543 and p-value
=9.66452199704509e-06
For BsmtQual the Chi-Square value= 2592.6160606597737 and p-value
=7.805558340542111e-22
For BsmtCond the Chi-Square value= 2447.2852478512814 and p-value
=1.9050784157897297e-14
For BsmtExposure the Chi-Square value= 2278.0268732438217 and p-value
=1.0980067354518743e-07
For Heating the Chi-Square value= 4201.387994086523 and p-value
=2.477753304101386e-24
For CentralAir the Chi-Square value= 826.856936591472 and p-value
=1.2257126695737677e-05
For KitchenQual the Chi-Square value= 2811.8004076403 and p-value
=1.2820744991685297e-31
```

```
For FireplaceQu the Chi-Square value= 2020.0779576070101 and p-value
=0.00025643727323263203
For GarageFinish the Chi-Square value= 1604.5787796542825 and p-value
=1.0355886930923195e-09
For GarageQual the Chi-Square value= 3107.0556393967954 and p-value
=2.538428577754634e-13
For SaleType the Chi-Square value= 6099.7944527529235 and p-value
=4.560785392696702e-14
For SaleCondition the Chi-Square value= 3950.7393968720835 and p-value
=5.613395680294345e-14
```

[46]:
```
final_categorical_features_list
```

[46]:
```
['MSZoning',
 'Street',
 'LotShape',
 'LotConfig',
 'Neighborhood',
 'MasVnrType',
 'ExterQual',
 'ExterCond',
 'Foundation',
 'BsmtQual',
 'BsmtCond',
 'BsmtExposure',
 'Heating',
 'CentralAir',
 'KitchenQual',
 'FireplaceQu',
 'GarageFinish',
 'GarageQual',
 'SaleType',
 'SaleCondition']
```

[47]:
```
len(final_categorical_features_list)
```

[47]:
```
20
```

Observations: There are 20 significant categorical features.

- Generation of the final dataset for categorical features

[48]:
```
final_categorical_features_df =␣
↪categorical_features_df[final_categorical_features_list]
final_categorical_features_df.shape
```

[48]:
```
(1338, 20)
```

[49]:
```
final_categorical_features_df.head()
```

[49]:

Id	MSZoning	Street	LotShape	LotConfig	Neighborhood	MasVnrType	ExterQual	ExterCond	Foundation	BsmtQual	BsmtCond	BsmtExposure	Heating	CentralAir	KitchenQual	FireplaceQu	GarageFinish	GarageQual	SaleType	SaleCondition
1	RL	Pave	Reg	Inside	CollgCr	BrkFace	Gd	TA	PConc	Gd	TA	No	GasA	Y	Gd	TA	RFn	TA	WD	Normal
2	RL	Pave	Reg	FR2	Veenker	None	TA	TA	CBlock	Gd	TA	Gd	GasA	Y	TA	TA	RFn	TA	WD	Normal
3	RL	Pave	IR1	Inside	CollgCr	BrkFace	Gd	TA	PConc	Gd	TA	Mn	GasA	Y	Gd	TA	RFn	TA	WD	Normal
4	RL	Pave	IR1	Corner	Crawfor	None	TA	TA	BrkTil	TA	Gd	No	GasA	Y	Gd	Gd	Unf	TA	WD	Abnorml
5	RL	Pave	IR1	FR2	NoRidge	BrkFace	Gd	TA	PConc	Gd	TA	Av	GasA	Y	Gd	TA	RFn	TA	WD	Normal

[50]:
```
final_numerical_features_df.shape
```

[50]:
```
(1452, 8)
```

1.5 5- Combining all significant features (numerical and categorical)

[51]:
```
combined_features_df=final_numerical_features_df.
↪join(final_categorical_features_df, how="inner", on="Id")
combined_features_df
```

[51]:

Id	SalePrice	OverallQual	GrLivArea	GarageCars	1stFlrSF	FullBath
1	208500	7	1710	2	856	2
2	181500	6	1262	2	1262	2
3	223500	7	1786	2	920	2
4	140000	7	1717	3	961	1
5	250000	8	2198	3	1145	2
…	…	…	…	…	…	…
1456	175000	6	1647	2	953	2
1457	210000	6	2073	2	2073	2

```
1458     266500            7       2340           1      1188         2
1459     142125            5       1078           1      1078         1
1460     147500            5       1256           1      1256         1

      YearBuilt  YearRemodAdd MSZoning Street  … BsmtCond BsmtExposure  \
Id                                             …
1          2003          2003       RL   Pave  …       TA           No
2          1976          1976       RL   Pave  …       TA           Gd
3          2001          2002       RL   Pave  …       TA           Mn
4          1915          1970       RL   Pave  …       Gd           No
5          2000          2000       RL   Pave  …       TA           Av
…           …             …       …    …  …  …       …           …
1456       1999          2000       RL   Pave  …       TA           No
1457       1978          1988       RL   Pave  …       TA           No
1458       1941          2006       RL   Pave  …       Gd           No
1459       1950          1996       RL   Pave  …       TA           Mn
1460       1965          1965       RL   Pave  …       TA           No

     Heating CentralAir KitchenQual FireplaceQu GarageFinish GarageQual  \
Id
1       GasA          Y          Gd          TA          RFn         TA
2       GasA          Y          TA          TA          RFn         TA
3       GasA          Y          Gd          TA          RFn         TA
4       GasA          Y          Gd          Gd          Unf         TA
5       GasA          Y          Gd          TA          RFn         TA
…       …         …          …          …           …         …
1456    GasA          Y          TA          TA          RFn         TA
1457    GasA          Y          TA          TA          Unf         TA
1458    GasA          Y          Gd          Gd          RFn         TA
1459    GasA          Y          Gd          TA          Unf         TA
1460    GasA          Y          TA          TA          Fin         TA

     SaleType SaleCondition
Id
1          WD        Normal
2          WD        Normal
3          WD        Normal
4          WD       Abnorml
5          WD        Normal
…          …            …
1456       WD        Normal
1457       WD        Normal
1458       WD        Normal
1459       WD        Normal
1460       WD        Normal

[1338 rows x 28 columns]
```

1.6 6- Plotting boxplot for the new dataset to find the features with outliers

[ ]:
```
for i in range(1,len(list(combined_features_df.columns))-1,3):
    plt.figure(figsize=(15,4))
    plt.subplot(131)
    sns.boxplot(x=list(combined_features_df.
 ↪columns)[i],y="SalePrice",data=combined_features_df)
    plt.subplot(132)
    sns.boxplot(x=list(combined_features_df.
 ↪columns)[i+1],y="SalePrice",data=combined_features_df)
    plt.subplot(133)
    sns.boxplot(x=list(combined_features_df.
 ↪columns)[i+2],y="SalePrice",data=combined_features_df)
    plt.subplots_adjust(wspace=0.5)
```

[ ]:
```

```