

```
In [11]: #1) Import the necessary libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [12]: %matplotlib inline
```

```
In [76]: #2) Read the dataset
housing=pd.read_csv("PEP1.csv")
```

```
In [77]: #2.1 Understand the dataset
housing.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 81 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   Id                1460 non-null    int64  
 1   MSSubClass         1460 non-null    int64  
 2   MSZoning          1460 non-null    object  
 3   LotFrontage        1201 non-null    float64 
 4   LotArea            1460 non-null    int64  
 5   Street             1460 non-null    object  
 6   Alley              91  non-null     object  
 7   LotShape            1460 non-null    object  
 8   LandContour         1460 non-null    object  
 9   Utilities           1460 non-null    object  
 10  LotConfig           1460 non-null    object  
 11  LandSlope           1460 non-null    object  
 12  Neighborhood        1460 non-null    object  
 13  Condition1          1460 non-null    object  
 14  Condition2          1460 non-null    object  
 15  BldgType            1460 non-null    object  
 16  HouseStyle          1460 non-null    object  
 17  OverallQual         1460 non-null    int64  
 18  OverallCond         1460 non-null    int64  
 19  YearBuilt           1460 non-null    int64  
 20  YearRemodAdd        1460 non-null    int64  
 21  RoofStyle            1460 non-null    object  
 22  RoofMatl             1460 non-null    object  
 23  Exterior1st          1460 non-null    object  
 24  Exterior2nd          1460 non-null    object  
 25  MasVnrType           1452 non-null    object  
 26  MasVnrArea           1452 non-null    float64 
 27  ExterQual            1460 non-null    object  
 28  ExterCond            1460 non-null    object  
 29  Foundation           1460 non-null    object  
 30  BsmtQual             1423 non-null    object  
 31  BsmtCond             1423 non-null    object  
 32  BsmtExposure         1422 non-null    object  
 33  BsmtFinType1          1423 non-null    object  
 34  BsmtFinSF1            1460 non-null    int64  
 35  BsmtFinType2          1422 non-null    object  
 36  BsmtFinSF2            1460 non-null    int64  
 37  BsmtUnfSF             1460 non-null    int64  
 38  TotalBsmtSF           1460 non-null    int64  
 39  Heating               1460 non-null    object  
 40  HeatingQC              1460 non-null    object  
 41  CentralAir            1460 non-null    object  
 42  Electrical             1459 non-null    object  
 43  1stFlrSF              1460 non-null    int64  
 44  2ndFlrSF              1460 non-null    int64  
 45  LowQualFinSF           1460 non-null    int64  
 46  GrLivArea              1460 non-null    int64  
 47  BsmtFullBath           1460 non-null    int64  
 48  BsmtHalfBath           1460 non-null    int64  
 49  FullBath               1460 non-null    int64  
 50  HalfBath                1460 non-null    int64  
 51  Bedroom                 1460 non-null    int64  
 52  Kitchen                  1460 non-null    int64  
 53  KitchenQual             1460 non-null    object  
 54  TotRmsAbvGrd            1460 non-null    int64  
 55  Functional                1460 non-null    object  
 56  Fireplaces               1460 non-null    int64  
 57  FireplaceQu              770  non-null     object  
 58  GarageType                1379 non-null    object

```

```

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

```

In [78]: #2.2 Print the name of the columns  
`print(housing.columns)`

```

Index(['Id', 'MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street',
       'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig',
       'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType',
       'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAd
d',
       'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType',
       'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual',
       'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1',
       'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating',
       'HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF',
       'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBat
h',
       'HalfBath', 'Bedroom', 'Kitchen', 'KitchenQual', 'TotRmsAbvGrd',
       'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType', 'GarageYrBl
t',
       'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual', 'GarageCon
d',
       'PavedDrive', 'WoodDeckSF', 'OpenPorchSF', 'EnclosedPorch', '3SsnPorc
h',
       'ScreenPorch', 'PoolArea', 'PoolQC', 'Fence', 'MiscFeature', 'MiscVa
l',
       'MoSold', 'YrSold', 'SaleType', 'SaleCondition', 'SalePrice'],
      dtype='object')

```

In [79]: #2.3 Print the shape of the dataframe  
`print(housing.shape)`  
`(1460, 81)`

In [80]: # 2.4 Check for null values  
`housing.isnull().values.any()`

Out[80]: True

In [81]: `print(housing.isnull().sum())`

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

In [82]: *# 2.5 Print the unique values*

```
housing.nunique()
```

Out [82]:

```
Id          1460
MSSubClass   15
MSZoning     5
LotFrontage   110
LotArea       1073
...
MoSold       12
YrSold       5
SaleType      9
SaleCondition 6
SalePrice     663
Length: 81, dtype: int64
```

In [141...]: *# 2.6 Select the numerical and categorical variables*

```
numerical_var=list(housing._get_numeric_data().columns)
numerical_var
```

```
Out[141]: ['Id',
'MSSubClass',
'LotFrontage',
'LotArea',
'OverallQual',
'OverallCond',
'YearBuilt',
'YearRemodAdd',
'MasVnrArea',
'BsmtFinSF1',
'BsmtFinSF2',
'BsmtUnfSF',
'TotalBsmtSF',
'1stFlrSF',
'2ndFlrSF',
'LowQualFinSF',
'GrLivArea',
'BsmtFullBath',
'BsmtHalfBath',
'FullBath',
'HalfBath',
'Bedroom',
'Kitchen',
'TotRmsAbvGrd',
'Fireplaces',
'GarageYrBlt',
'GarageCars',
'GarageArea',
'WoodDeckSF',
'OpenPorchSF',
'EnclosedPorch',
'3SsnPorch',
'ScreenPorch',
'PoolArea',
'MiscVal',
'MoSold',
'YrSold',
'SalePrice']
```

```
In [142]: categorical_var=list(set(housing.columns)- set(numerical_var))
```

```
In [143]: categorical_var
```

```
Out[143]: ['MSZoning',
 'PoolQC',
 'LotShape',
 'Condition2',
 'LandSlope',
 'Condition1',
 'Utilities',
 'LotConfig',
 'BsmtCond',
 'GarageType',
 'Electrical',
 'HouseStyle',
 'Exterior2nd',
 'PavedDrive',
 'ExterCond',
 'MasVnrType',
 'SaleType',
 'RoofStyle',
 'Foundation',
 'Alley',
 'GarageFinish',
 'BldgType',
 'Exterior1st',
 'Neighborhood',
 'HeatingQC',
 'CentralAir',
 'Functional',
 'Street',
 'Heating',
 'BsmtExposure',
 'BsmtQual',
 'BsmtFinType1',
 'ExterQual',
 'GarageCond',
 'KitchenQual',
 'RoofMatl',
 'LandContour',
 'BsmtFinType2',
 'GarageQual',
 'SaleCondition']
```

In [86]: # 3.1 EDA of numerical variables  
housing.describe()

	<b>Id</b>	<b>MSSubClass</b>	<b>LotFrontage</b>	<b>LotArea</b>	<b>OverallQual</b>	<b>OverallCond</b>
<b>count</b>	1460.000000	1460.000000	1201.000000	1460.000000	1460.000000	1460.000000
<b>mean</b>	730.500000	56.897260	70.049958	10516.828082	6.099315	5.575342
<b>std</b>	421.610009	42.300571	24.284752	9981.264932	1.382997	1.112799
<b>min</b>	1.000000	20.000000	21.000000	1300.000000	1.000000	1.000000
<b>25%</b>	365.750000	20.000000	59.000000	7553.500000	5.000000	5.000000
<b>50%</b>	730.500000	50.000000	69.000000	9478.500000	6.000000	5.000000
<b>75%</b>	1095.250000	70.000000	80.000000	11601.500000	7.000000	6.000000
<b>max</b>	1460.000000	190.000000	313.000000	215245.000000	10.000000	9.000000

8 rows x 38 columns

In [87]: housing.corr()

Out[87]:

	<b>Id</b>	<b>MSSubClass</b>	<b>LotFrontage</b>	<b>LotArea</b>	<b>OverallQual</b>	<b>OverallCond</b>
<b>Id</b>	1.000000	0.011156	-0.010601	-0.033226	-0.028365	0.012609
<b>MSSubClass</b>	0.011156	1.000000	-0.386347	-0.139781	0.032628	-0.059316
<b>LotFrontage</b>	-0.010601	-0.386347	1.000000	0.426095	0.251646	-0.059213
<b>LotArea</b>	-0.033226	-0.139781	0.426095	1.000000	0.105806	-0.005636
<b>OverallQual</b>	-0.028365	0.032628	0.251646	0.105806	1.000000	-0.091932
<b>OverallCond</b>	0.012609	-0.059316	-0.059213	-0.005636	-0.091932	1.000000
<b>YearBuilt</b>	-0.012713	0.027850	0.123349	0.014228	0.572323	-0.375983
<b>YearRemodAdd</b>	-0.021998	0.040581	0.088866	0.013788	0.550684	0.073741
<b>MasVnrArea</b>	-0.050298	0.022936	0.193458	0.104160	0.411876	-0.128101
<b>BsmtFinSF1</b>	-0.005024	-0.069836	0.233633	0.214103	0.239666	-0.046231
<b>BsmtFinSF2</b>	-0.005968	-0.065649	0.049900	0.111170	-0.059119	0.040229
<b>BsmtUnfSF</b>	-0.007940	-0.140759	0.132644	-0.002618	0.308159	-0.136841
<b>TotalBsmtSF</b>	-0.015415	-0.238518	0.392075	0.260833	0.537808	-0.171098
<b>1stFlrSF</b>	0.010496	-0.251758	0.457181	0.299475	0.476224	-0.144203
<b>2ndFlrSF</b>	0.005590	0.307886	0.080177	0.050986	0.295493	0.028942
<b>LowQualFinSF</b>	-0.044230	0.046474	0.038469	0.004779	-0.030429	0.025494
<b>GrLivArea</b>	0.008273	0.074853	0.402797	0.263116	0.593007	-0.079686
<b>BsmtFullBath</b>	0.002289	0.003491	0.100949	0.158155	0.111098	-0.054942
<b>BsmtHalfBath</b>	-0.020155	-0.002333	-0.007234	0.048046	-0.040150	0.117821
<b>FullBath</b>	0.005587	0.131608	0.198769	0.126031	0.550600	-0.194149
<b>HalfBath</b>	0.006784	0.177354	0.053532	0.014259	0.273458	-0.060769
<b>Bedroom</b>	0.037719	-0.023438	0.263170	0.119690	0.101676	0.012980
<b>Kitchen</b>	0.002951	0.281721	-0.006069	-0.017784	-0.183882	-0.087001
<b>TotRmsAbvGrd</b>	0.027239	0.040380	0.352096	0.190015	0.427452	-0.057583
<b>Fireplaces</b>	-0.019772	-0.045569	0.266639	0.271364	0.396765	-0.023820
<b>GarageYrBlt</b>	0.000072	0.085072	0.070250	-0.024947	0.547766	-0.324297
<b>GarageCars</b>	0.016570	-0.040110	0.285691	0.154871	0.600671	-0.185758
<b>GarageArea</b>	0.017634	-0.098672	0.344997	0.180403	0.562022	-0.151521
<b>WoodDeckSF</b>	-0.029643	-0.012579	0.088521	0.171698	0.238923	-0.003334
<b>OpenPorchSF</b>	-0.000477	-0.006100	0.151972	0.084774	0.308819	-0.032589
<b>EnclosedPorch</b>	0.002889	-0.012037	0.010700	-0.018340	-0.113937	0.070356
<b>3SsnPorch</b>	-0.046635	-0.043825	0.070029	0.020423	0.030371	0.025504
<b>ScreenPorch</b>	0.001330	-0.026030	0.041383	0.043160	0.064886	0.054811
<b>PoolArea</b>	0.057044	0.008283	0.206167	0.077672	0.065166	-0.001985
<b>MiscVal</b>	-0.006242	-0.007683	0.003368	0.038068	-0.031406	0.068777
<b>MoSold</b>	0.021172	-0.013585	0.011200	0.001205	0.070815	-0.003511
<b>YrSold</b>	0.000712	-0.021407	0.007450	-0.014261	-0.027347	0.043950
<b>SalePrice</b>	-0.021917	-0.084284	0.351799	0.263843	0.790982	-0.077856

Id	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond
----	------------	-------------	---------	-------------	-------------

---

38 rows × 38 columns

In [88]: `housing.info()`

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 81 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   Id                1460 non-null    int64  
 1   MSSubClass         1460 non-null    int64  
 2   MSZoning          1460 non-null    object  
 3   LotFrontage        1201 non-null    float64 
 4   LotArea            1460 non-null    int64  
 5   Street             1460 non-null    object  
 6   Alley              91  non-null     object  
 7   LotShape            1460 non-null    object  
 8   LandContour         1460 non-null    object  
 9   Utilities           1460 non-null    object  
 10  LotConfig           1460 non-null    object  
 11  LandSlope           1460 non-null    object  
 12  Neighborhood        1460 non-null    object  
 13  Condition1          1460 non-null    object  
 14  Condition2          1460 non-null    object  
 15  BldgType            1460 non-null    object  
 16  HouseStyle          1460 non-null    object  
 17  OverallQual         1460 non-null    int64  
 18  OverallCond         1460 non-null    int64  
 19  YearBuilt           1460 non-null    int64  
 20  YearRemodAdd        1460 non-null    int64  
 21  RoofStyle            1460 non-null    object  
 22  RoofMatl             1460 non-null    object  
 23  Exterior1st          1460 non-null    object  
 24  Exterior2nd          1460 non-null    object  
 25  MasVnrType           1452 non-null    object  
 26  MasVnrArea           1452 non-null    float64 
 27  ExterQual            1460 non-null    object  
 28  ExterCond            1460 non-null    object  
 29  Foundation           1460 non-null    object  
 30  BsmtQual             1423 non-null    object  
 31  BsmtCond             1423 non-null    object  
 32  BsmtExposure         1422 non-null    object  
 33  BsmtFinType1          1423 non-null    object  
 34  BsmtFinSF1            1460 non-null    int64  
 35  BsmtFinType2          1422 non-null    object  
 36  BsmtFinSF2            1460 non-null    int64  
 37  BsmtUnfSF             1460 non-null    int64  
 38  TotalBsmtSF           1460 non-null    int64  
 39  Heating               1460 non-null    object  
 40  HeatingQC              1460 non-null    object  
 41  CentralAir            1460 non-null    object  
 42  Electrical             1459 non-null    object  
 43  1stFlrSF              1460 non-null    int64  
 44  2ndFlrSF              1460 non-null    int64  
 45  LowQualFinSF           1460 non-null    int64  
 46  GrLivArea              1460 non-null    int64  
 47  BsmtFullBath           1460 non-null    int64  
 48  BsmtHalfBath           1460 non-null    int64  
 49  FullBath               1460 non-null    int64  
 50  HalfBath                1460 non-null    int64  
 51  Bedroom                 1460 non-null    int64  
 52  Kitchen                  1460 non-null    int64  
 53  KitchenQual             1460 non-null    object  
 54  TotRmsAbvGrd            1460 non-null    int64  
 55  Functional                1460 non-null    object  
 56  Fireplaces               1460 non-null    int64  
 57  FireplaceQu              770  non-null     object  
 58  GarageType                1379 non-null    object

```

```

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

```

```
In [89]: # 3.2 Missing value treatment
# Filling the missing values of Alley column with our assumption that there
housing.loc[housing["Alley"].isnull(),"Alley"]="No alley"

In [90]: housing.loc[housing["PoolQC"].isnull(),"PoolQC"]="No Pool"

In [91]: # Dropping fence and Misc Feature as more than 70% of data is missing for th
housing=housing.drop("MiscFeature",axis=1)
housing=housing.drop("Fence",axis=1)

In [92]: housing=housing.drop("FireplaceQu",axis=1)

In [93]: housing.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 78 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   Id                1460 non-null    int64  
 1   MSSubClass         1460 non-null    int64  
 2   MSZoning          1460 non-null    object  
 3   LotFrontage        1201 non-null    float64 
 4   LotArea            1460 non-null    int64  
 5   Street             1460 non-null    object  
 6   Alley              1460 non-null    object  
 7   LotShape            1460 non-null    object  
 8   LandContour        1460 non-null    object  
 9   Utilities          1460 non-null    object  
 10  LotConfig           1460 non-null    object  
 11  LandSlope           1460 non-null    object  
 12  Neighborhood        1460 non-null    object  
 13  Condition1         1460 non-null    object  
 14  Condition2         1460 non-null    object  
 15  BldgType           1460 non-null    object  
 16  HouseStyle          1460 non-null    object  
 17  OverallQual        1460 non-null    int64  
 18  OverallCond         1460 non-null    int64  
 19  YearBuilt           1460 non-null    int64  
 20  YearRemodAdd       1460 non-null    int64  
 21  RoofStyle           1460 non-null    object  
 22  RoofMatl            1460 non-null    object  
 23  Exterior1st         1460 non-null    object  
 24  Exterior2nd         1460 non-null    object  
 25  MasVnrType          1452 non-null    object  
 26  MasVnrArea          1452 non-null    float64 
 27  ExterQual           1460 non-null    object  
 28  ExterCond           1460 non-null    object  
 29  Foundation          1460 non-null    object  
 30  BsmtQual            1423 non-null    object  
 31  BsmtCond            1423 non-null    object  
 32  BsmtExposure        1422 non-null    object  
 33  BsmtFinType1        1423 non-null    object  
 34  BsmtFinSF1          1460 non-null    int64  
 35  BsmtFinType2        1422 non-null    object  
 36  BsmtFinSF2          1460 non-null    int64  
 37  BsmtUnfSF           1460 non-null    int64  
 38  TotalBsmtSF         1460 non-null    int64  
 39  Heating              1460 non-null    object  
 40  HeatingQC            1460 non-null    object  
 41  CentralAir           1460 non-null    object  
 42  Electrical           1459 non-null    object  
 43  1stFlrSF             1460 non-null    int64  
 44  2ndFlrSF             1460 non-null    int64  
 45  LowQualFinSF         1460 non-null    int64  
 46  GrLivArea            1460 non-null    int64  
 47  BsmtFullBath         1460 non-null    int64  
 48  BsmtHalfBath         1460 non-null    int64  
 49  FullBath             1460 non-null    int64  
 50  HalfBath             1460 non-null    int64  
 51  Bedroom              1460 non-null    int64  
 52  Kitchen               1460 non-null    int64  
 53  KitchenQual          1460 non-null    object  
 54  TotRmsAbvGrd         1460 non-null    int64  
 55  Functional            1460 non-null    object  
 56  Fireplaces            1460 non-null    int64  
 57  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          1460 non-null   object
72 MiscVal          1460 non-null   int64
73 MoSold           1460 non-null   int64
74 YrSold           1460 non-null   int64
75 SaleType         1460 non-null   object
76 SaleCondition    1460 non-null   object
77 SalePrice        1460 non-null   int64
dtypes: float64(3), int64(35), object(40)
memory usage: 889.8+ KB

```

In [95]: # Replacing False with a numerical value Zero and making an assumption that  
housing.loc[housing["LotFrontage"].isnull(),"LotFrontage"] = 0

In [121...]: # 3.3 Identify the skewness and distribution

```

import sweetviz as sv
sweet_report=sv.analyze(housing,pairwise_analysis="Off")
sweet_report.show_html("housing_analysis_report.html")

```

| | [ 0%] 00:00 -> (? 1  
eft)  
Report housing\_analysis\_report.html was generated! NOTEBOOK/COLAB USERS: the  
web browser MAY not pop up, regardless, the report IS saved in your notebook/colab files.

In [118...]: # 3.4 Identify significant variables using a correlation matrix

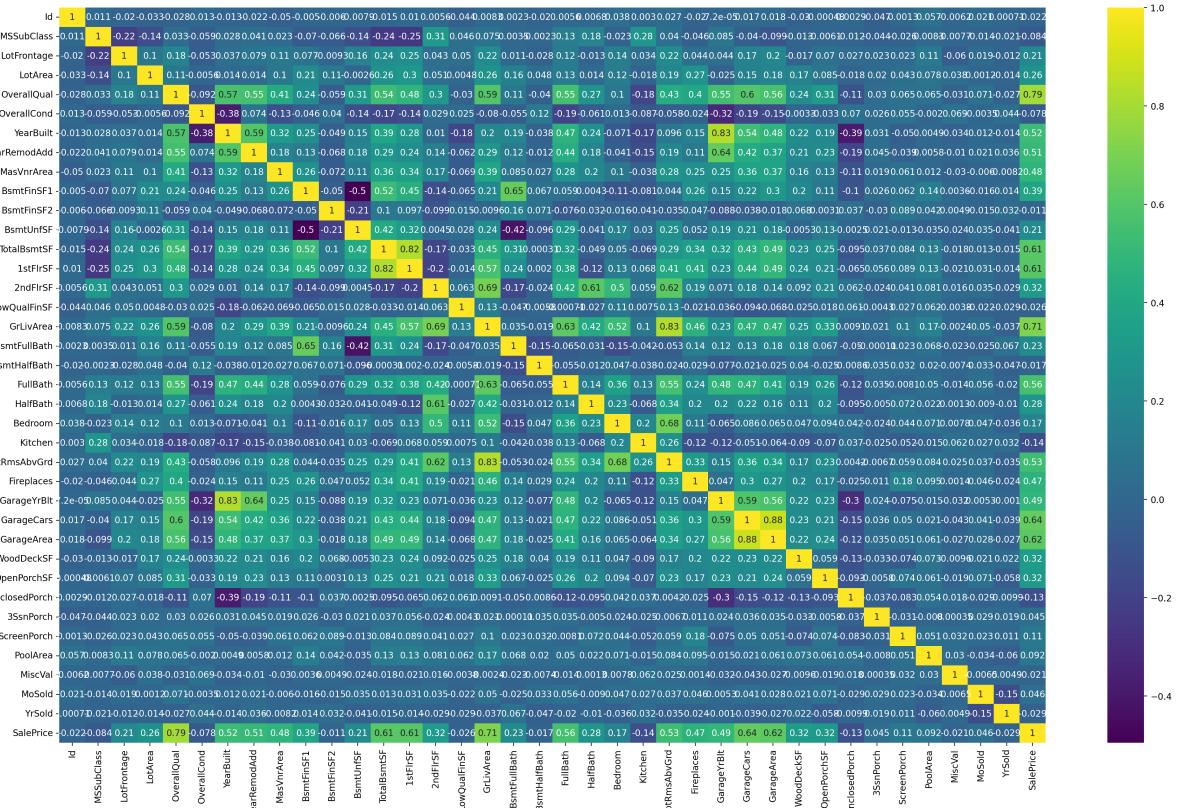
```

plt.figure(figsize=(25,15),dpi=200)
sns.heatmap(housing.corr(),annot=True,cmap="viridis")

```

# We can observe here that Sale Price is highly correlated with OverallQual

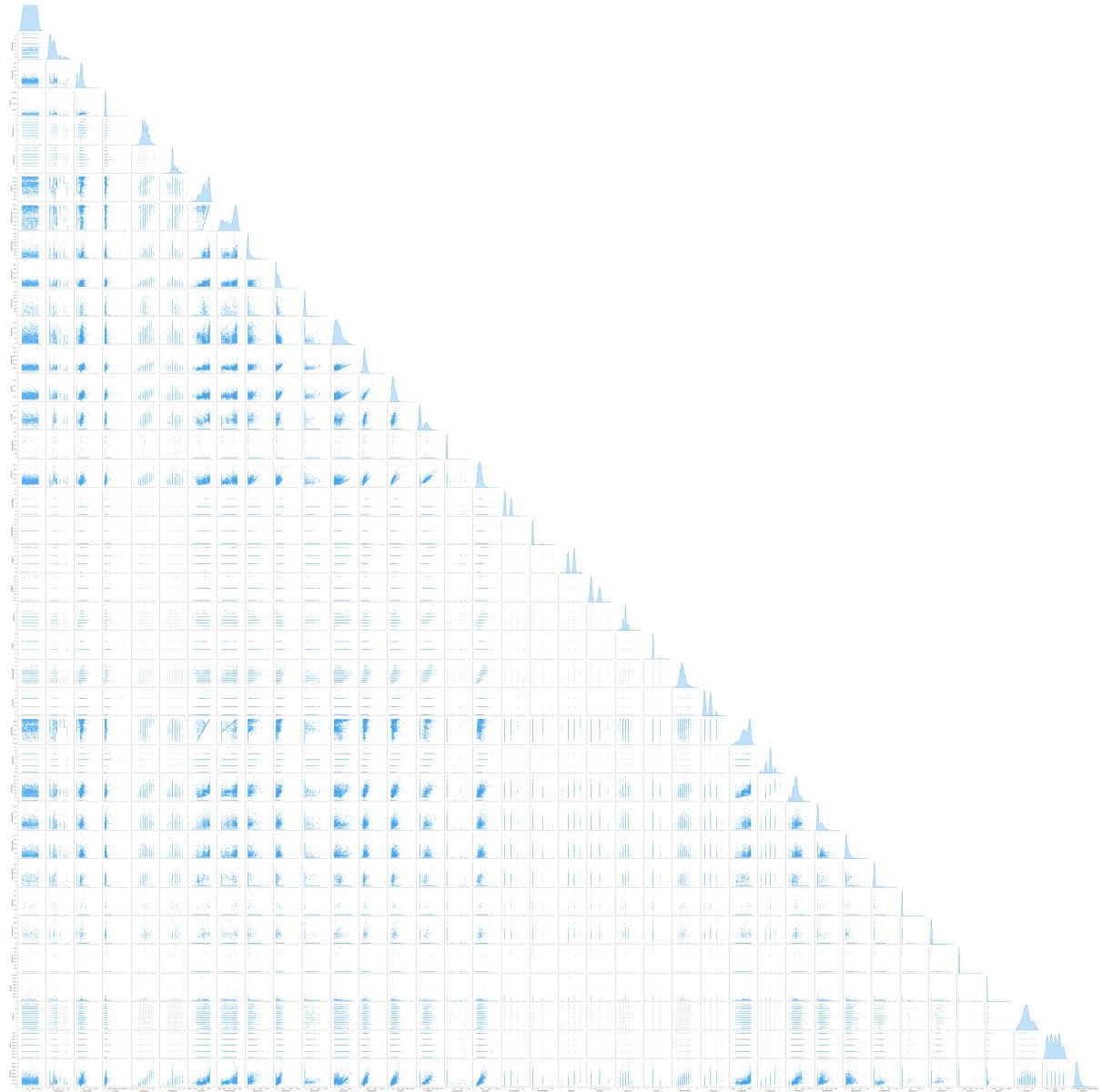
Out[118]: <AxesSubplot:>



In [123]: # 3.5 Pair plot for distribution and density

```
sns.pairplot(housing,corner=True,diag_kind="kde")
```

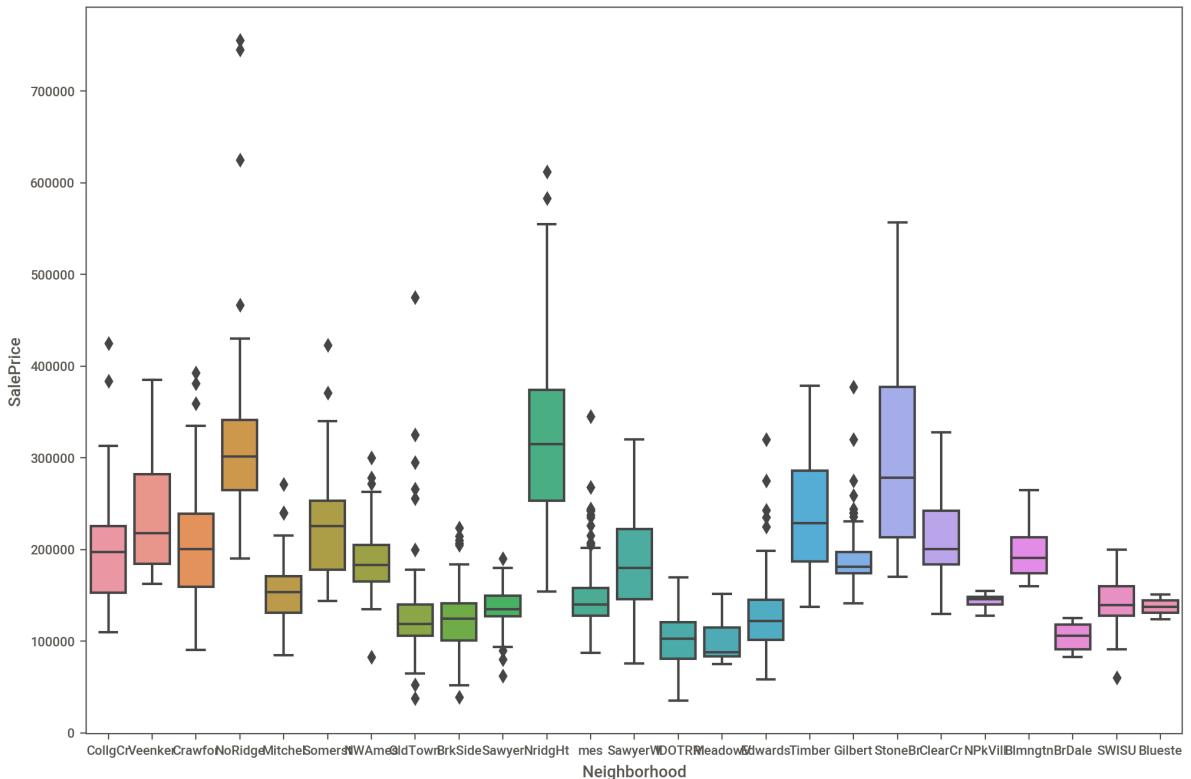
Out[123]: <seaborn.axisgrid.PairGrid at 0x7f8bf40a65b0>



In [128]: # 3.3 Identify the skewness and distribution

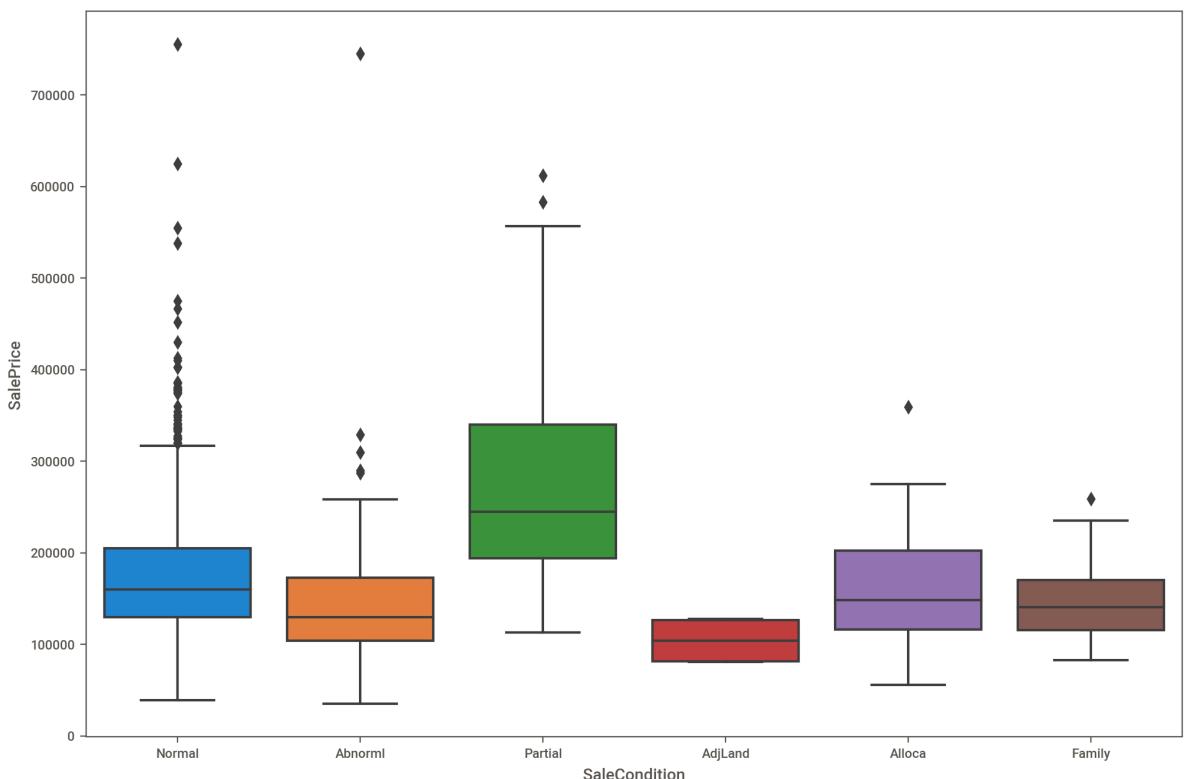
```
plt.figure(figsize=(12,8),dpi=200)
sns.boxplot(x="Neighborhood",y="SalePrice",data=housing)
```

Out[128]: <AxesSubplot:xlabel='Neighborhood', ylabel='SalePrice'>



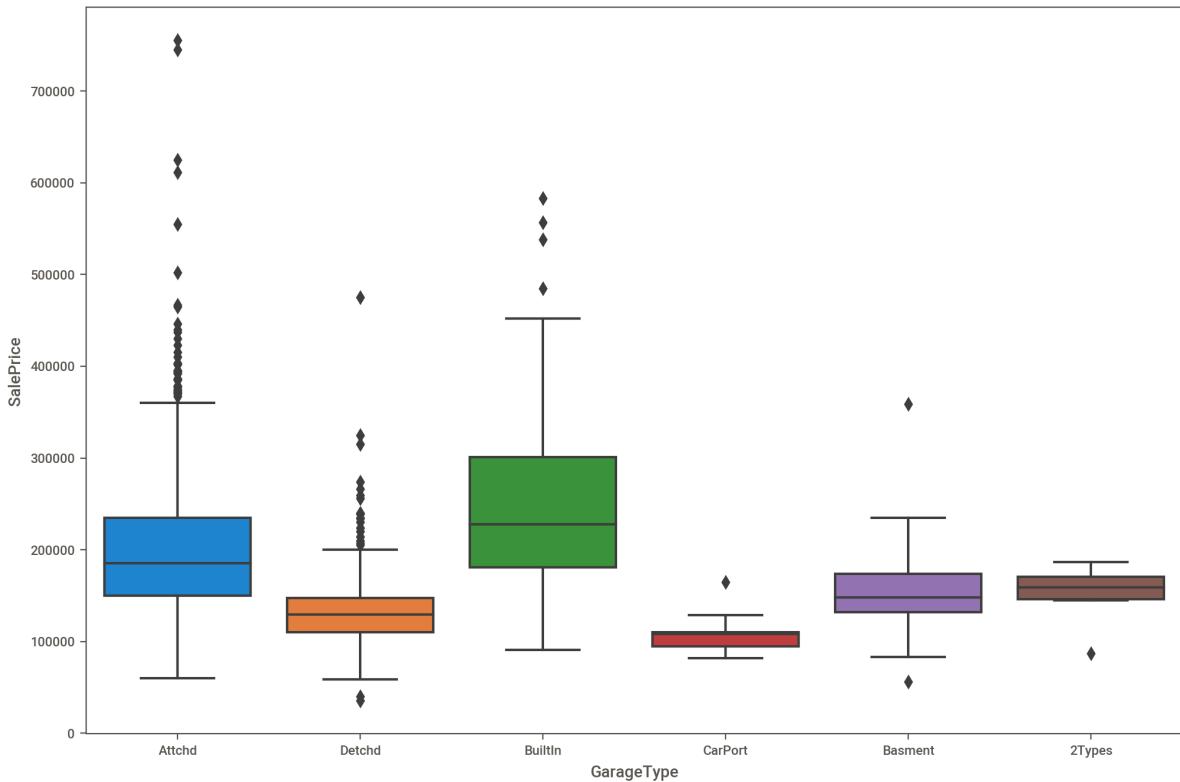
```
In [134]: plt.figure(figsize=(12,8),dpi=200)
sns.boxplot(x="SaleCondition",y="SalePrice",data=housing)
```

```
Out[134]: <AxesSubplot:xlabel='SaleCondition', ylabel='SalePrice'>
```



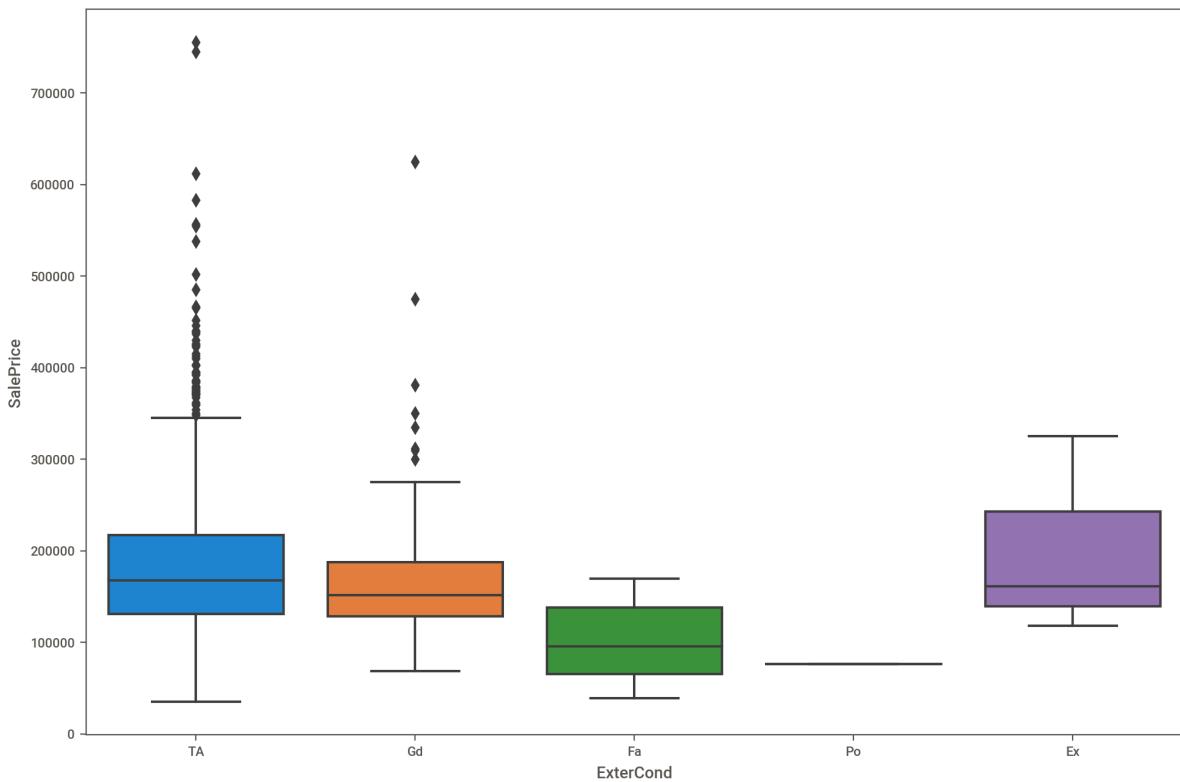
```
In [135]: plt.figure(figsize=(12,8),dpi=200)
sns.boxplot(x="GarageType",y="SalePrice",data=housing)
```

```
Out[135]: <AxesSubplot:xlabel='GarageType', ylabel='SalePrice'>
```



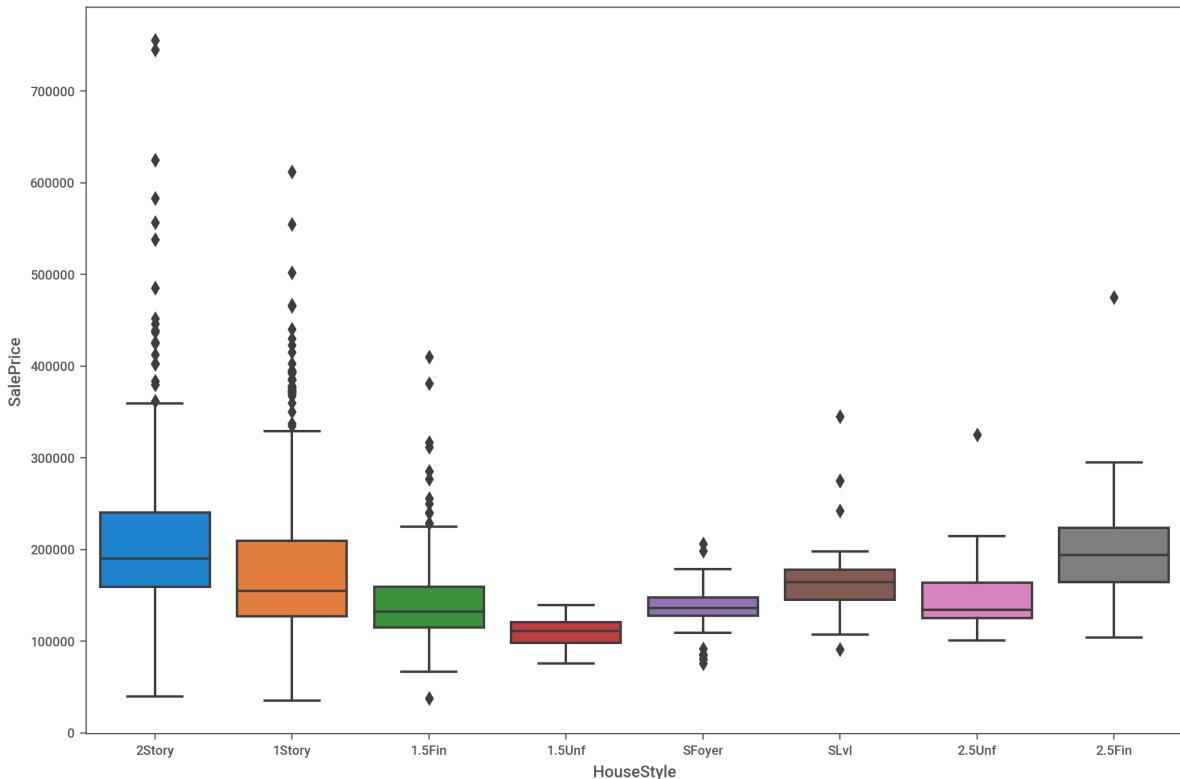
```
In [136]: plt.figure(figsize=(12,8),dpi=200)
sns.boxplot(x="ExterCond",y="SalePrice",data=housing)
```

```
Out[136]: <AxesSubplot:xlabel='ExterCond', ylabel='SalePrice'>
```



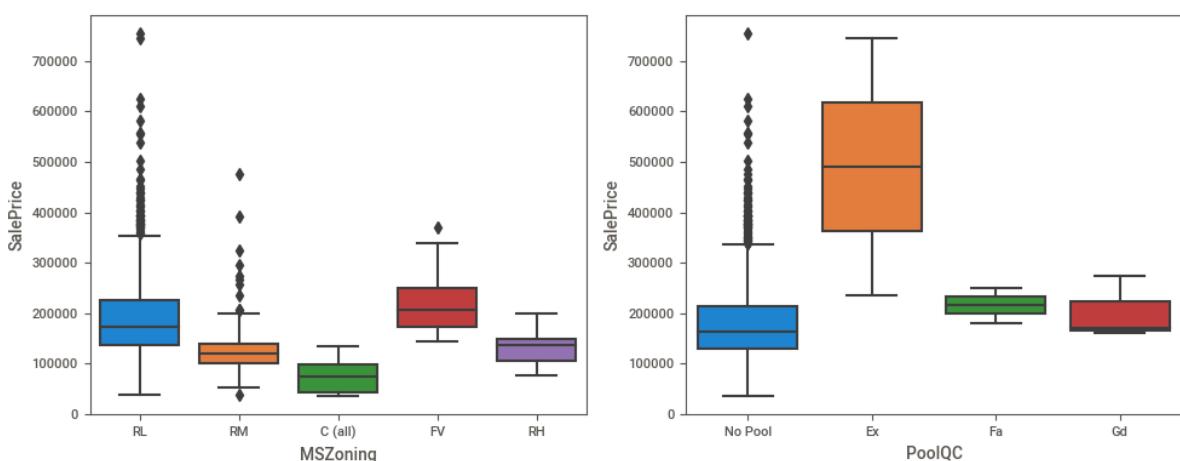
```
In [137]: plt.figure(figsize=(12,8),dpi=200)
sns.boxplot(x="HouseStyle",y="SalePrice",data=housing)
```

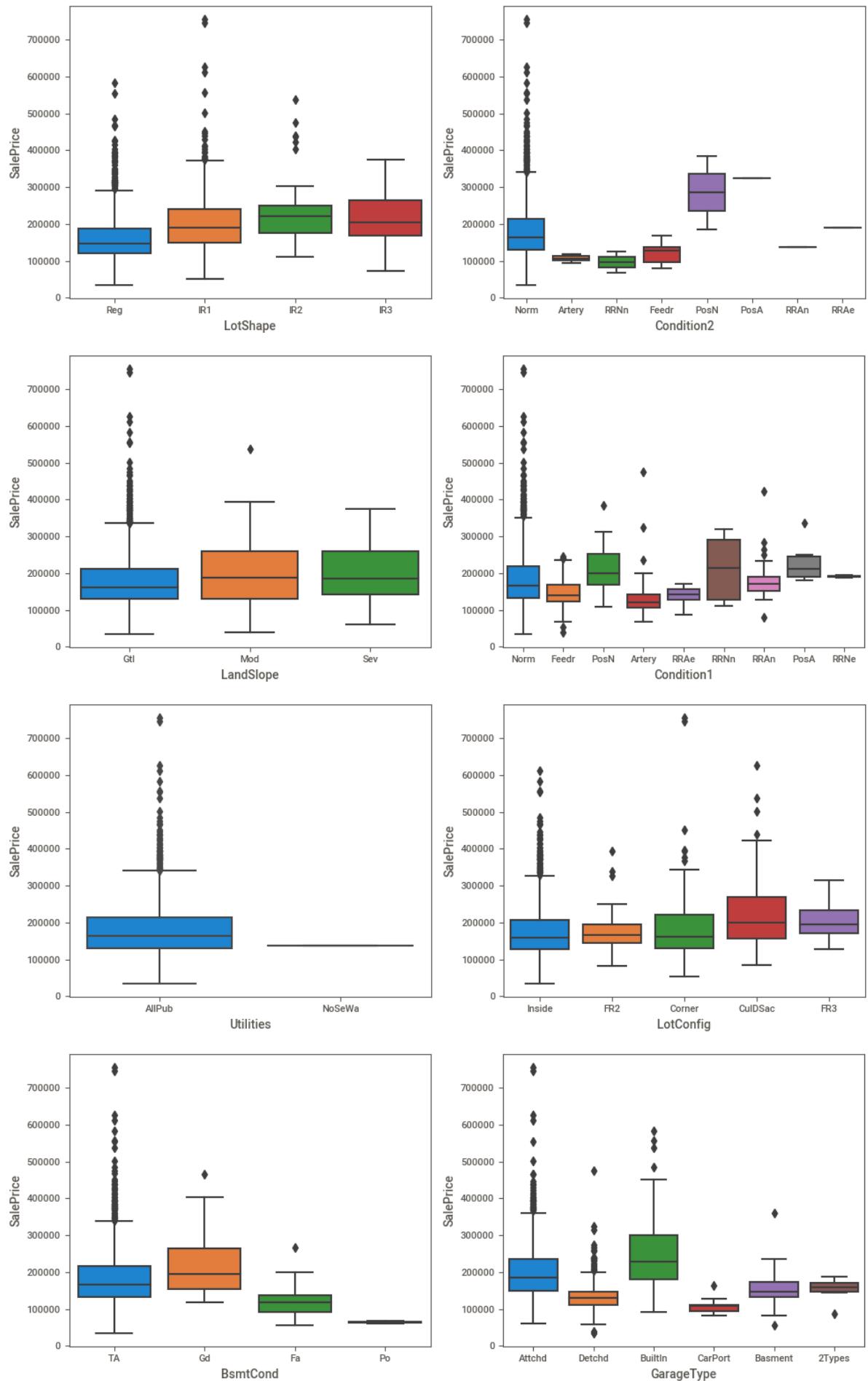
```
Out[137]: <AxesSubplot:xlabel='HouseStyle', ylabel='SalePrice'>
```

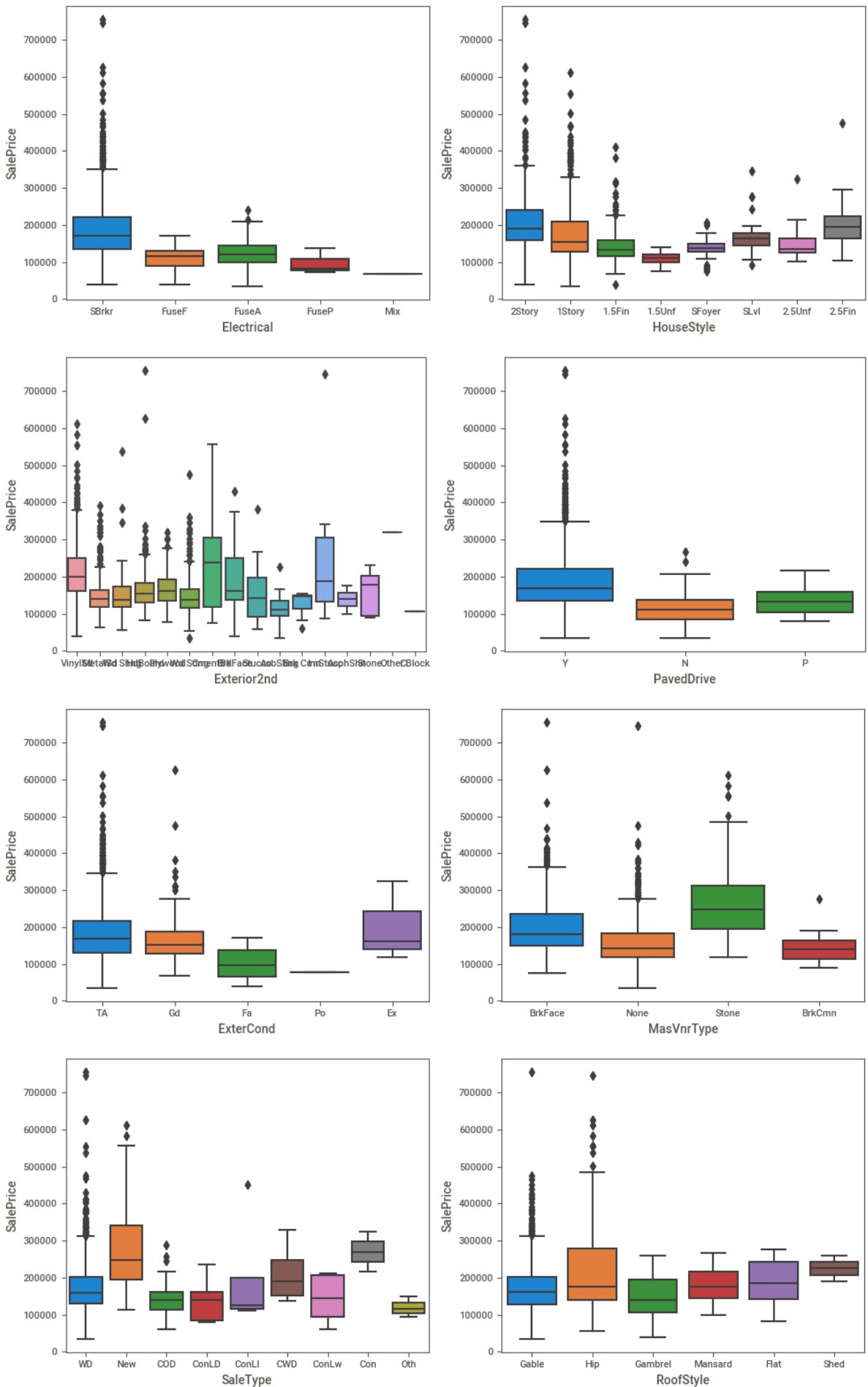


```
In [145]: # Observing different box plots between Sale Price and all Categorical Variables
```

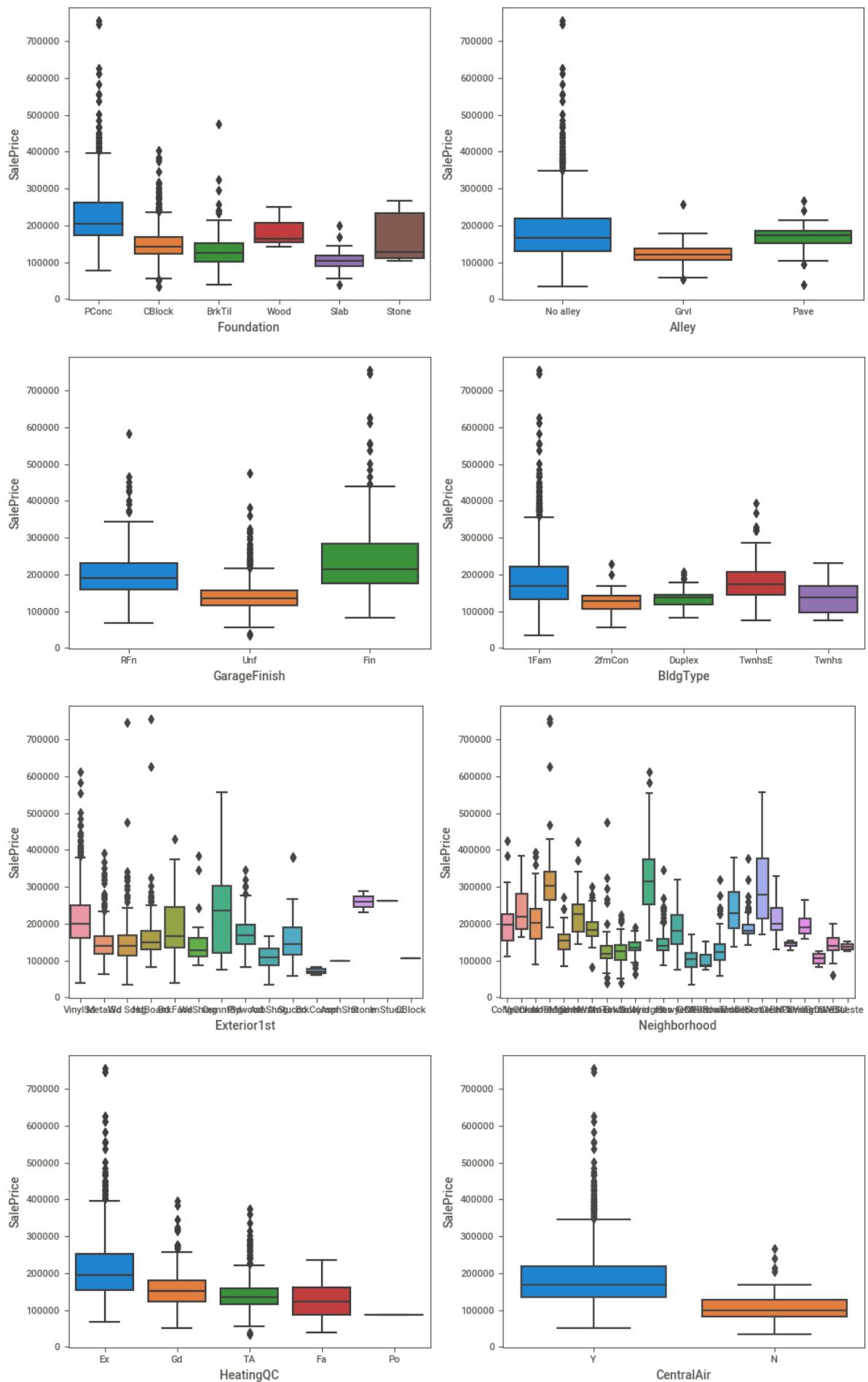
```
In [144]: for i in range(0,len(categorical_var),2):
    plt.figure(figsize=(10,4))
    plt.subplot(121)
    sns.boxplot(x=categorical_var[i],y="SalePrice",data = housing)
    plt.subplot(122)
    sns.boxplot(x=categorical_var[i+1],y="SalePrice",data = housing)
    plt.tight_layout()
    plt.show()
```

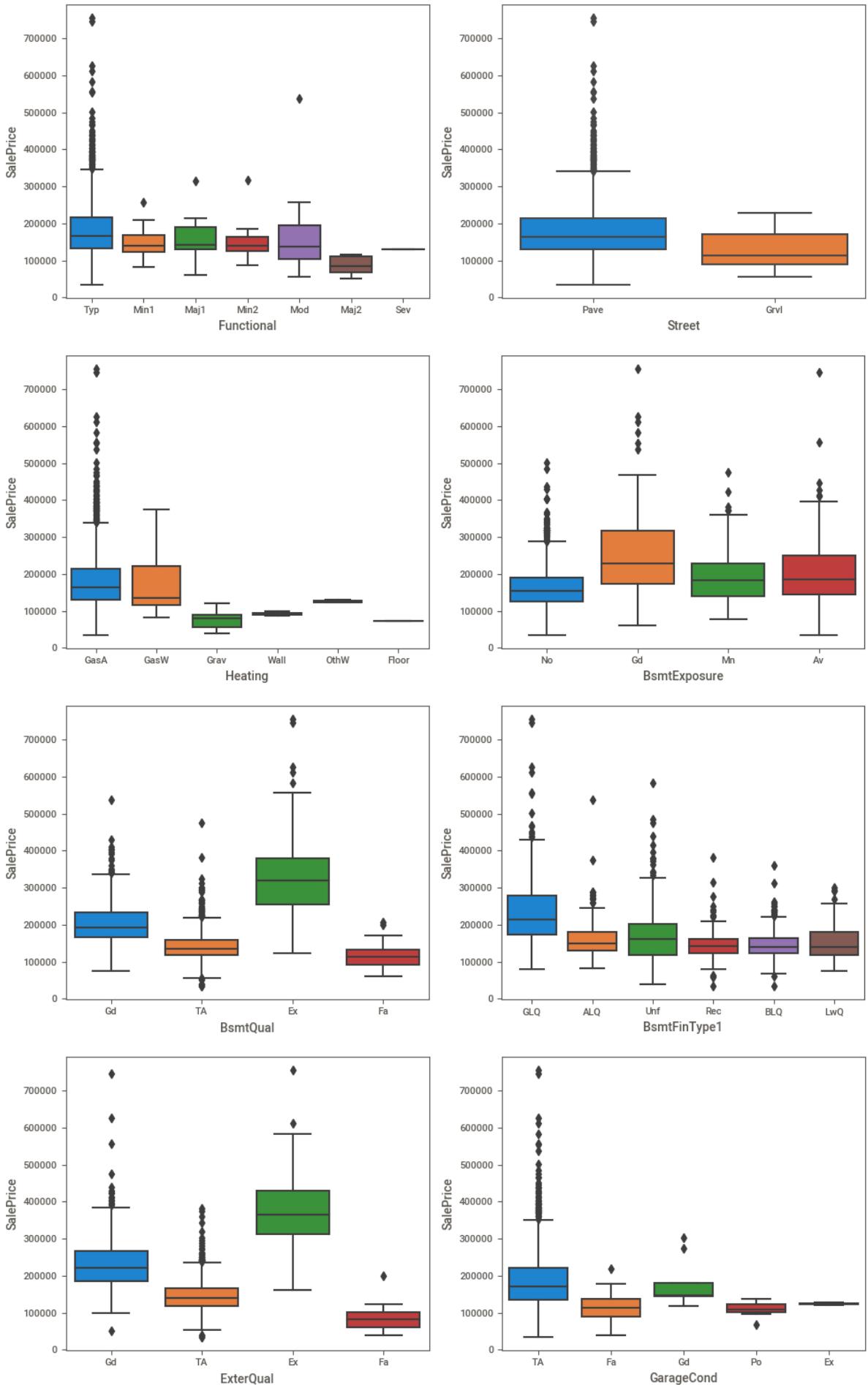


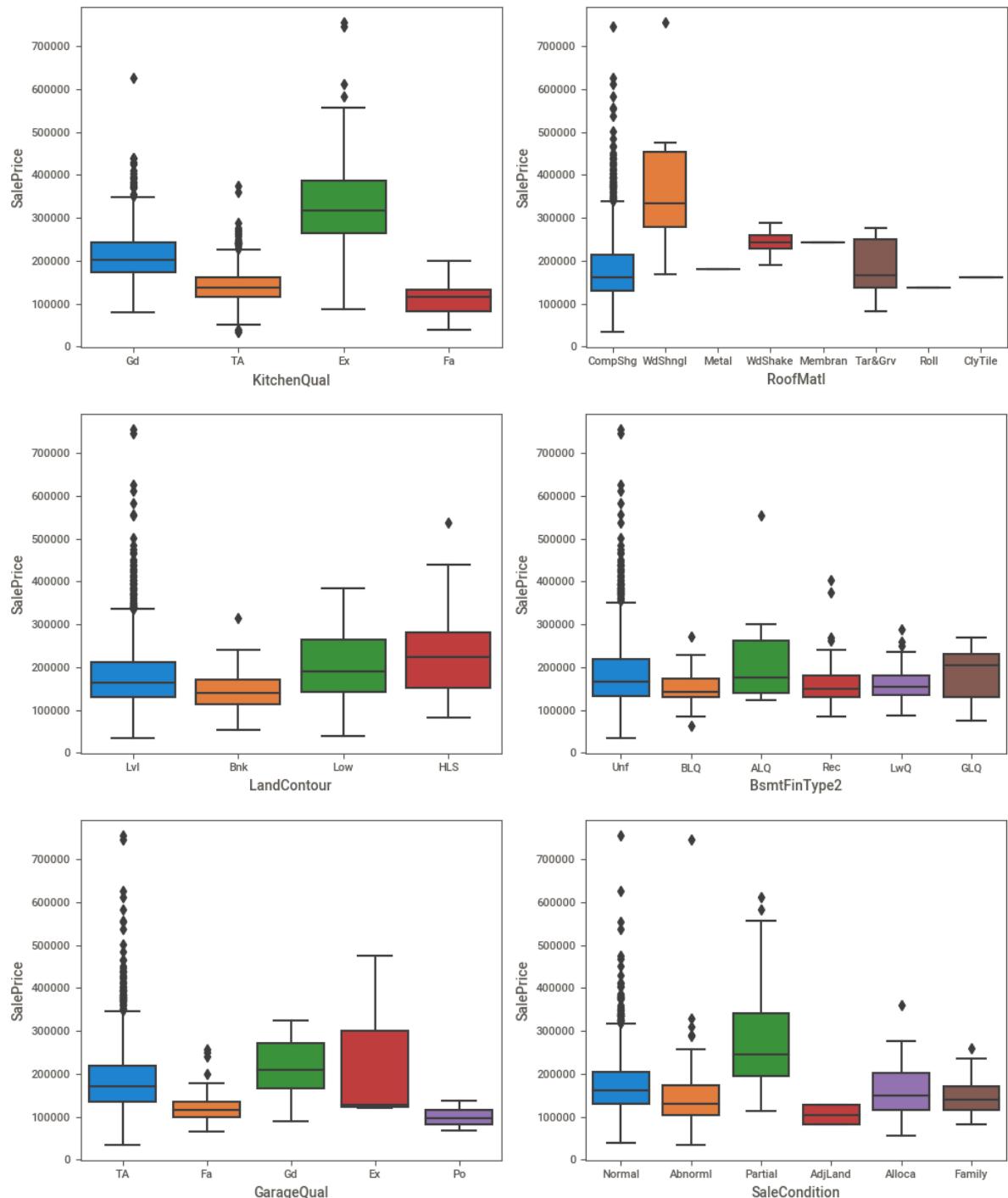




## Proj\_housing







```
In [129]: categorical_var
```

```
Out[129]: ['MSZoning',
'PoolQC',
'MiscFeature',
'LotShape',
'Condition2',
'LandSlope',
'Condition1',
'Utilities',
'LotConfig',
'BsmtCond',
'GarageType',
'Electrical',
'HouseStyle',
'Exterior2nd',
'PavedDrive',
'ExterCond',
'MasVnrType',
'SaleType',
'RoofStyle',
'Foundation',
'Alley',
'GarageFinish',
'BldgType',
'Exterior1st',
'FireplaceQu',
'Neighborhood',
'HeatingQC',
'CentralAir',
'Functional',
'Street',
'Heating',
'BsmtExposure',
'BsmtQual',
'BsmtFinType1',
'ExterQual',
'GarageCond',
'KitchenQual',
'RoofMatl',
'Fence',
'LandContour',
'BsmtFinType2',
'GarageQual',
'SaleCondition']
```

In [ ]: