HOUSE PRICES ADVANCED REGRESSION TECHNIQUES

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import preprocessing
from sklearn import metrics
```

In [48]:

```
# Fix random seed for reproducibility
np.random.seed(10)
plt.rcParams["figure.figsize"]=(10,5)
```

IMPORT DATA

In [51]:

```
# Train Data
train=pd.read_csv("C:/Users/Me/Desktop/Data Science/Houses/house-prices-advanced-regression-techni
ques/train.csv")
```

In [3]:

```
#Test Data
test=pd.read_csv("C:/Users/Me/Desktop/Data Science/Houses/house-prices-advanced-regression-techniq
ues/test.csv")
```

In [4]:

```
#DATA SHAPE
```

In [5]:

```
train.shape
```

Out[5]:

```
(1460, 81)
```

In [6]:

```
test.shape
```

Out[6]:

```
(1459, 80)
```

In [7]:

```
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 81 columns):
Id               1460 non-null int64
MSSubClass       1460 non-null int64
MSZoning         1460 non-null object
LotFrontage      1201 non-null float64
LotArea          1460 non-null int64
Street           1460 non-null object
Alley            91 non-null object
```

```
LotShape         1460 non-null object
LandContour      1460 non-null object
Utilities        1460 non-null object
LotConfig        1460 non-null object
LandSlope        1460 non-null object
Neighborhood     1460 non-null object
Condition1       1460 non-null object
Condition2       1460 non-null object
BldgType         1460 non-null object
HouseStyle       1460 non-null object
OverallQual      1460 non-null int64
OverallCond      1460 non-null int64
YearBuilt        1460 non-null int64
YearRemodAdd     1460 non-null int64
RoofStyle        1460 non-null object
RoofMatl         1460 non-null object
Exterior1st      1460 non-null object
Exterior2nd      1460 non-null object
MasVnrType       1452 non-null object
MasVnrArea       1452 non-null float64
ExterQual        1460 non-null object
ExterCond        1460 non-null object
Foundation       1460 non-null object
BsmtQual         1423 non-null object
BsmtCond         1423 non-null object
BsmtExposure     1422 non-null object
BsmtFinType1     1423 non-null object
BsmtFinSF1       1460 non-null int64
BsmtFinType2     1422 non-null object
BsmtFinSF2       1460 non-null int64
BsmtUnfSF        1460 non-null int64
TotalBsmtSF      1460 non-null int64
Heating          1460 non-null object
HeatingQC        1460 non-null object
CentralAir       1460 non-null object
Electrical       1459 non-null object
1stFlrSF         1460 non-null int64
2ndFlrSF         1460 non-null int64
LowQualFinSF     1460 non-null int64
GrLivArea        1460 non-null int64
BsmtFullBath     1460 non-null int64
BsmtHalfBath     1460 non-null int64
FullBath         1460 non-null int64
HalfBath         1460 non-null int64
BedroomAbvGr     1460 non-null int64
KitchenAbvGr     1460 non-null int64
KitchenQual      1460 non-null object
TotRmsAbvGrd     1460 non-null int64
Functional       1460 non-null object
Fireplaces       1460 non-null int64
FireplaceQu      770 non-null object
GarageType       1379 non-null object
GarageYrBlt      1379 non-null float64
GarageFinish     1379 non-null object
GarageCars       1460 non-null int64
GarageArea       1460 non-null int64
GarageQual       1379 non-null object
GarageCond       1379 non-null object
PavedDrive       1460 non-null object
WoodDeckSF       1460 non-null int64
OpenPorchSF      1460 non-null int64
EnclosedPorch    1460 non-null int64
3SsnPorch        1460 non-null int64
ScreenPorch      1460 non-null int64
PoolArea         1460 non-null int64
PoolQC           7 non-null object
Fence            281 non-null object
MiscFeature      54 non-null object
MiscVal          1460 non-null int64
MoSold           1460 non-null int64
YrSold           1460 non-null int64
SaleType         1460 non-null object
SaleCondition    1460 non-null object
SalePrice        1460 non-null int64
dtypes: float64(3), int64(35), object(43)
memory usage: 924.0+ KB
```

In [8]:

```
# MISSING VALUES
```

In [9]:

```
train.isnull().sum()
```

Out[9]:

```
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 [10]:

```
test.isnull().sum()
```

Out[10]:

```
Id                0
MSSubClass        0
MSZoning          4
LotFrontage     227
LotArea           0
                 ...
MiscVal           0
MoSold            0
YrSold            0
SaleType          1
SaleCondition     0
Length: 80, dtype: int64
```

#Fill NA values

In [52]:

```
train.fillna(method="bfill", inplace=True)
test.fillna(method="bfill", inplace=True)
```

In [55]:

```
train.fillna(method="ffill", inplace=True)
test.fillna(method="ffill", inplace=True)
```

In [57]:

```
#Check remaining NA values
train.info()
```

Out[57]:

```
Id              0
MSSubClass      0
MSZoning        0
LotFrontage     0
LotArea         0
                ..
MoSold          0
YrSold          0
SaleType        0
SaleCondition   0
```

```
SalePrice        0
Length: 81, dtype: int64
```

In [ ]:

```
```

TOP 6 OF DATA

In [13]:

```
train.head()
```

Out[13]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	PoolArea	PoolQC	Fence	M
0	1	60	RL	65.0	8450	Pave	Grvl	Reg	Lvl	AllPub	...	0	Ex	MnPrv	
1	2	20	RL	80.0	9600	Pave	Grvl	Reg	Lvl	AllPub	...	0	Ex	MnPrv	
2	3	60	RL	68.0	11250	Pave	Grvl	IR1	Lvl	AllPub	...	0	Ex	MnPrv	
3	4	70	RL	60.0	9550	Pave	Grvl	IR1	Lvl	AllPub	...	0	Ex	MnPrv	
4	5	60	RL	84.0	14260	Pave	Grvl	IR1	Lvl	AllPub	...	0	Ex	MnPrv	

5 rows × 81 columns

In [14]:

```
test.head()
```

Out[14]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	ScreenPorch	PoolArea	P
0	1461	20	RH	80.0	11622	Pave	Pave	Reg	Lvl	AllPub	...	120	0	
1	1462	20	RL	81.0	14267	Pave	Pave	IR1	Lvl	AllPub	...	0	0	
2	1463	60	RL	74.0	13830	Pave	Pave	IR1	Lvl	AllPub	...	0	0	
3	1464	60	RL	78.0	9978	Pave	Pave	IR1	Lvl	AllPub	...	0	0	
4	1465	120	RL	43.0	5005	Pave	Pave	IR1	HLS	AllPub	...	144	0	

5 rows × 80 columns

In [15]:

```
# COLUMNS IN THE DATA
train.columns
```

Out[15]:

```
Index(['Id', 'MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street',
       'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig',
       'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType',
       'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd',
       'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType',
       'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual',
       'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1',
       'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating',
       'HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF',
       'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath',
       'HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual',
       'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType',
       'GarageYrBlt', 'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual',
       'GarageCond', 'PavedDrive', 'WoodDeckSF', 'OpenPorchSF',
       'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'PoolQC',
       'Fence', 'MiscFeature', 'MiscVal', 'MoSold', 'YrSold', 'SaleType',
       'SaleCondition', 'SalePrice'],
      dtype='object')
```

In [ ]:

```
```

Encoding

In [16]:

```
label_encode=preprocessing.LabelEncoder()
```

In [17]:

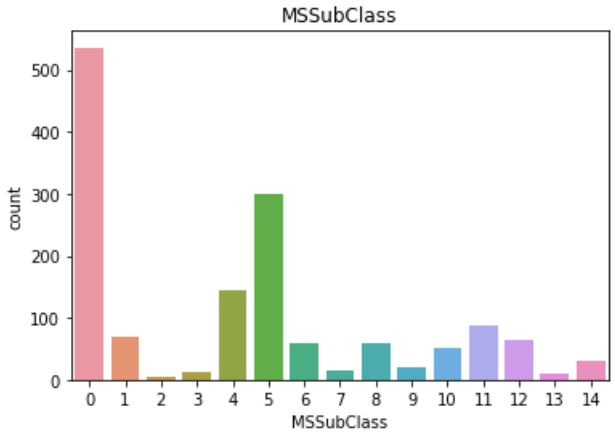
```
columns=['MSSubClass','MSZoning','Street','Alley', 'LotShape', 'LandContour',
'Utilities','LotConfig',
           'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType','HouseStyle','Overal
lQual',
            'OverallCond','RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd','MasVnrType','Exter
Qual',
            'ExterCond', 'Foundation', 'BsmtQual','BsmtCond', 'BsmtExposure', 'BsmtFinType1','BsmtF
inType2',
            'Heating','HeatingQC', 'CentralAir', 'Electrical', 'KitchenQual','Functional','Fireplac
eQu', 'GarageType',
            'GarageFinish','GarageQual','GarageCond','PavedDrive','PoolQC','Fence', 'MiscFeature','
SaleType','SaleCondition']
```

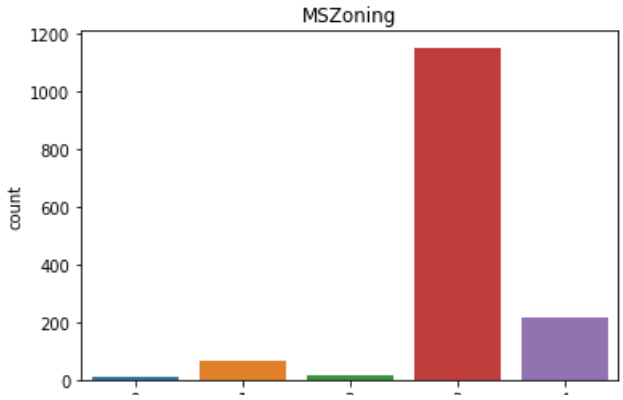
In [24]:

```
# Changing data type to category
for i in columns:
    train[i]=train[i].astype("category")
    test[i]=test[i].astype("category")
```

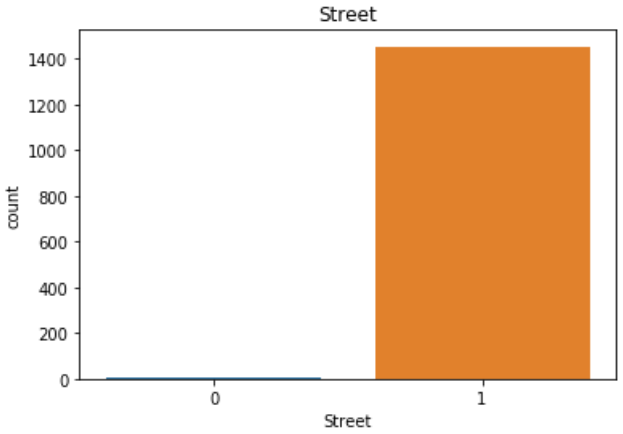
In [31]:

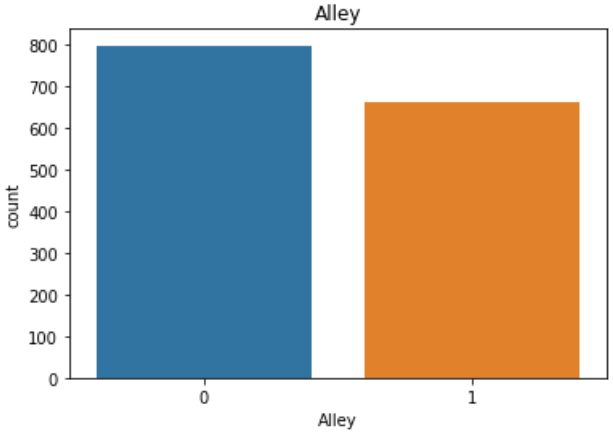
```
# Counting and plotting counts of attributes
for i in columns:
    plt.title(i)
    sns.countplot(i,data=train)
    plt.show()
```

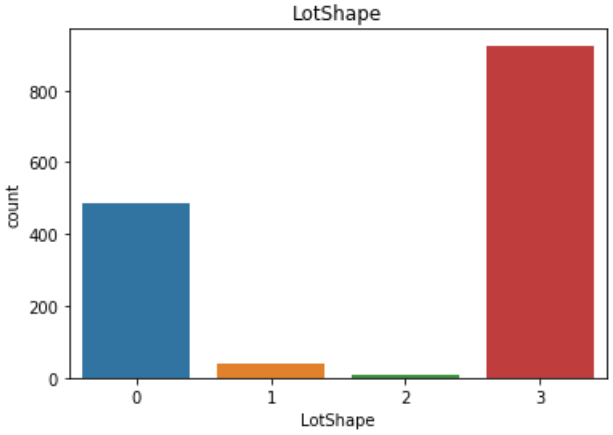


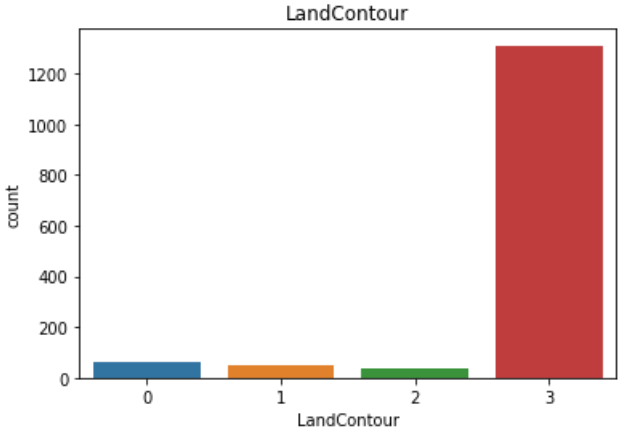


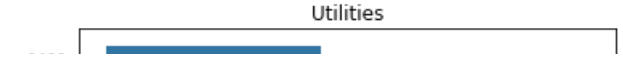

0
1
2
3
4
MSZoning

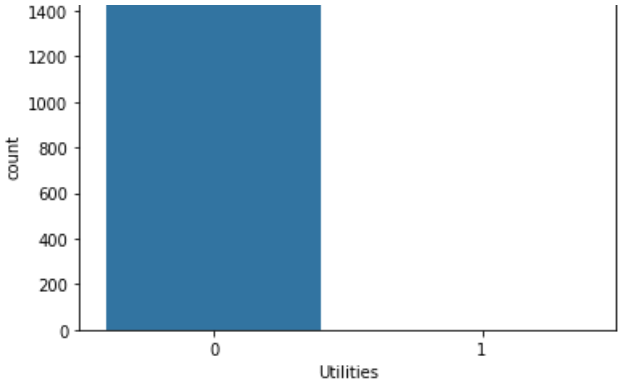

Street
1400
1200
1000
800
600
400
200
0
count
0
1
Street

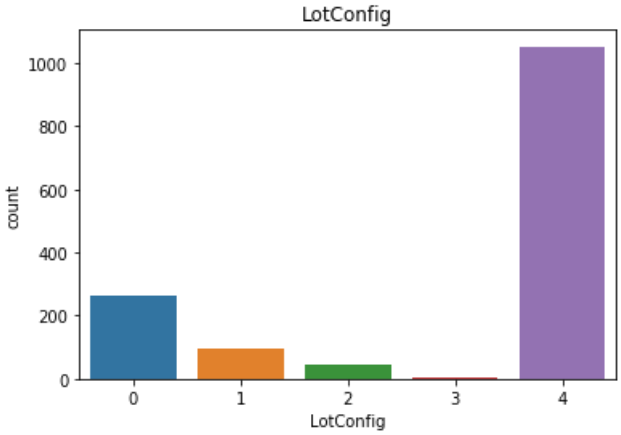

Alley
800
700
600
500
400
300
200
100
0
count
0
1
Alley

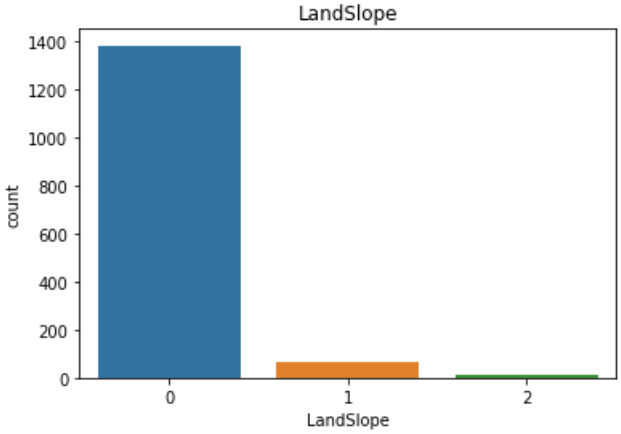

LotShape
800
600
400
200
0
count
0
1
2
3
LotShape

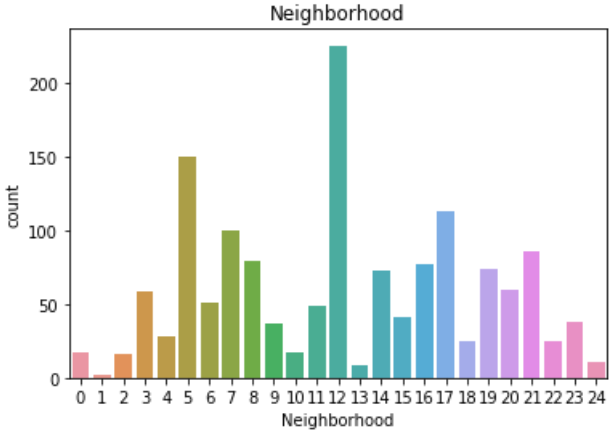

LandContour
1200
1000
800
600
400
200
0
count
0
1
2
3
LandContour


Utilities

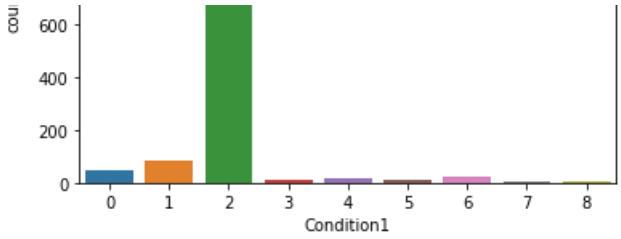

count
Utilities

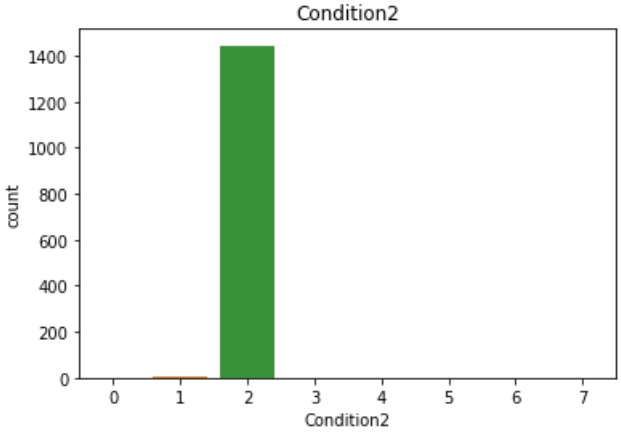

LotConfig
count
LotConfig

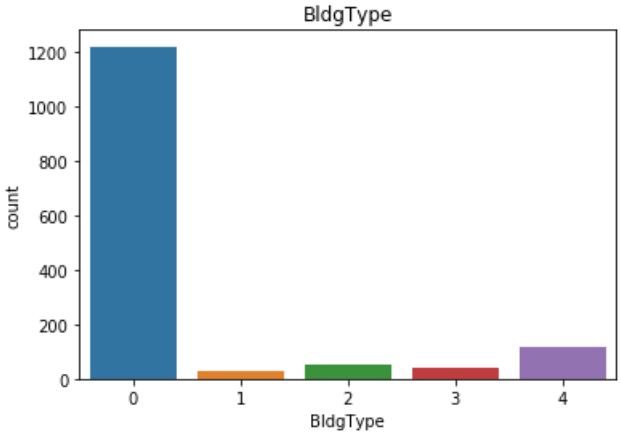

LandSlope
count
LandSlope

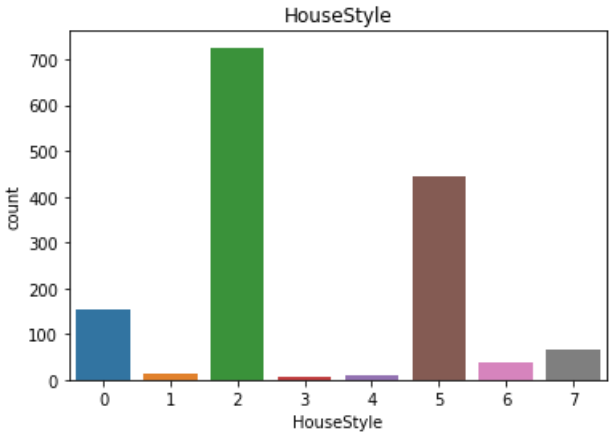

Neighborhood
count
Neighborhood

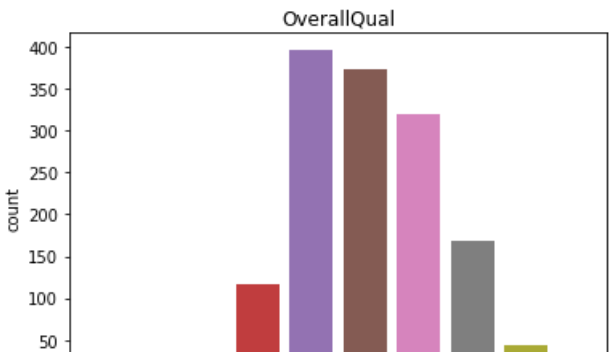

Condition1

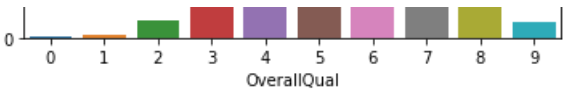

cou
600
400
200
0
0 1 2 3 4 5 6 7 8
Condition1

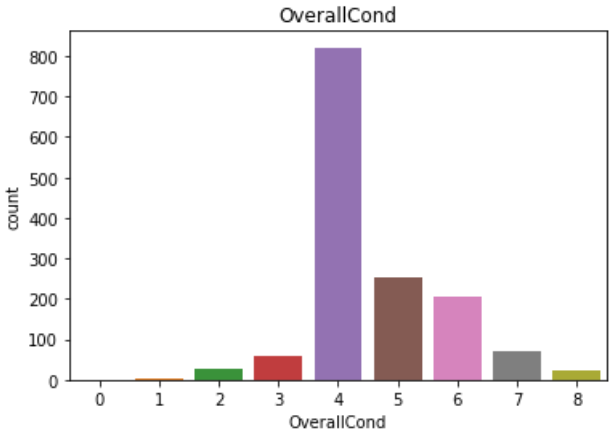

Condition2
count
1400
1200
1000
800
600
400
200
0
0 1 2 3 4 5 6 7
Condition2

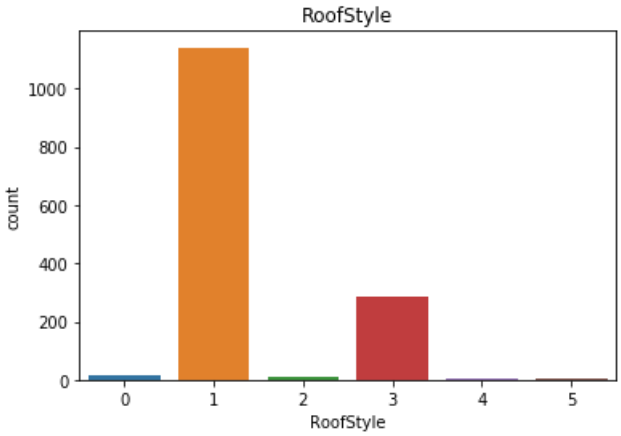

BldgType
count
1200
1000
800
600
400
200
0
0 1 2 3 4
BldgType

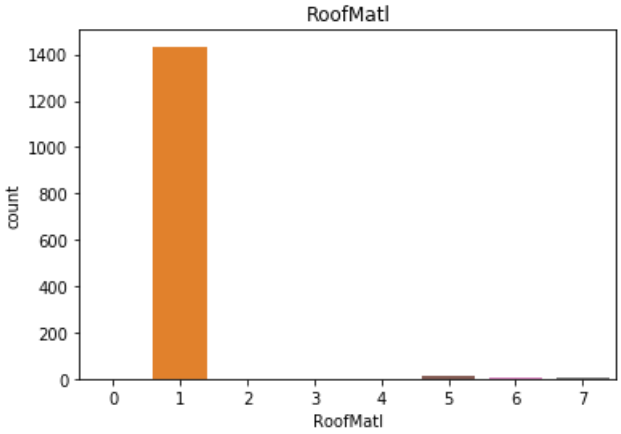

HouseStyle
count
700
600
500
400
300
200
100
0
0 1 2 3 4 5 6 7
HouseStyle

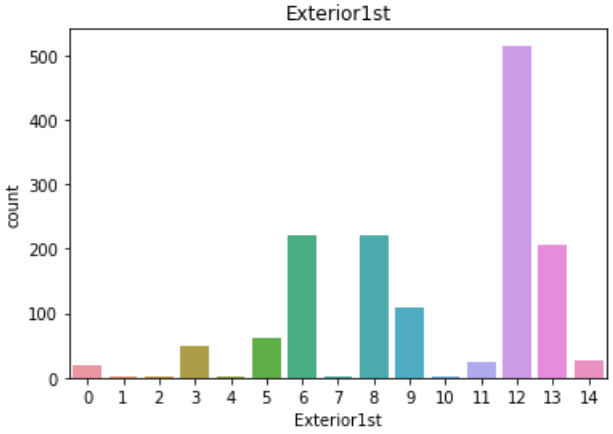

OverallQual
count
400
350
300
250
200
150
100
50

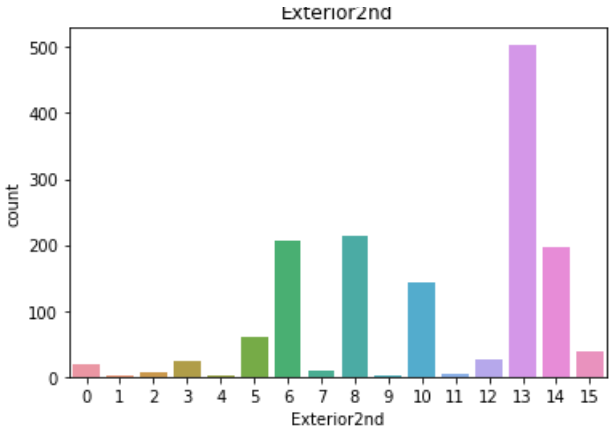

0
0 1 2 3 4 5 6 7 8 9
OverallQual

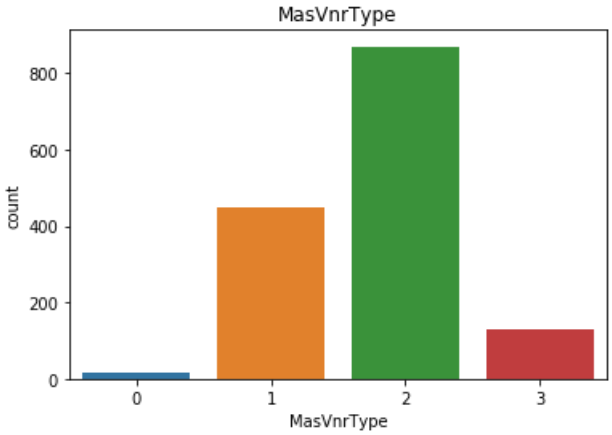

OverallCond
count
800
700
600
500
400
300
200
100
0
0 1 2 3 4 5 6 7 8
OverallCond

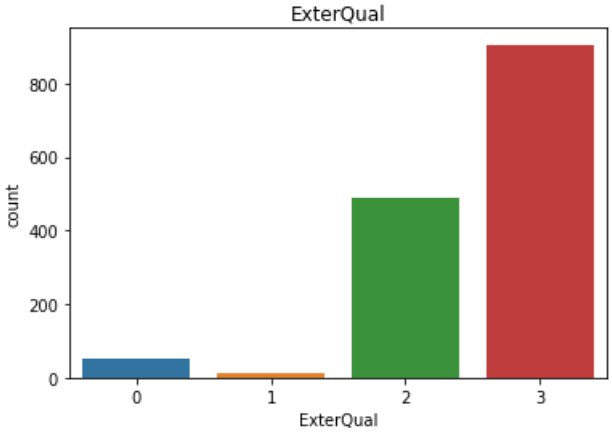

RoofStyle
count
1000
800
600
400
200
0
0 1 2 3 4 5
RoofStyle

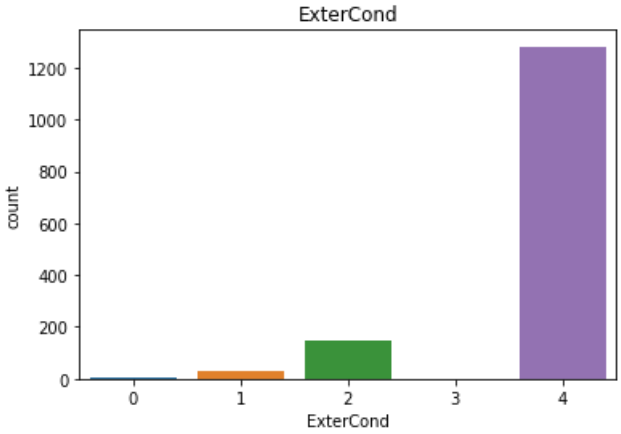

RoofMatl
count
1400
1200
1000
800
600
400
200
0
0 1 2 3 4 5 6 7
RoofMatl


Exterior1st
count
500
400
300
200
100
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Exterior1st

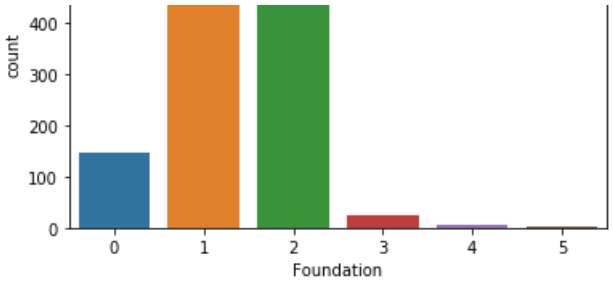

Exterior2nd
count
Exterior2nd

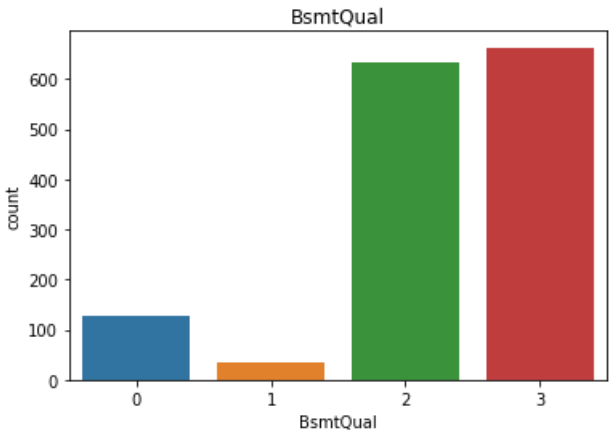

MasVnrType
count
MasVnrType

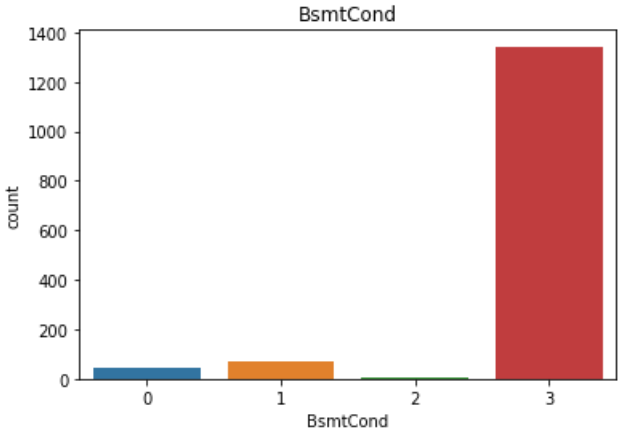

ExterQual
count
ExterQual

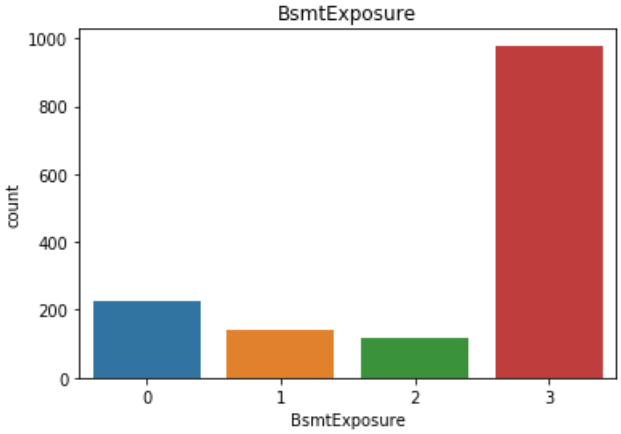

ExterCond
count
ExterCond

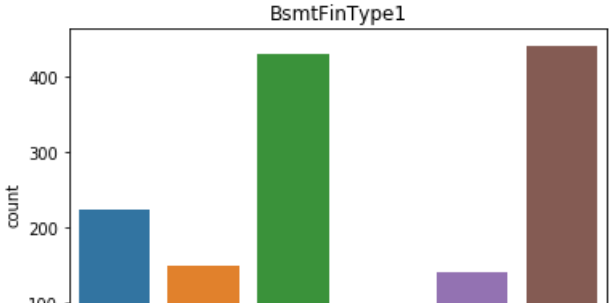

Foundation

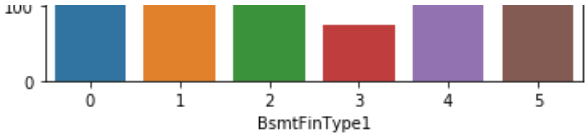

count
400
300
200
100
0
0
1
2
3
4
5
Foundation

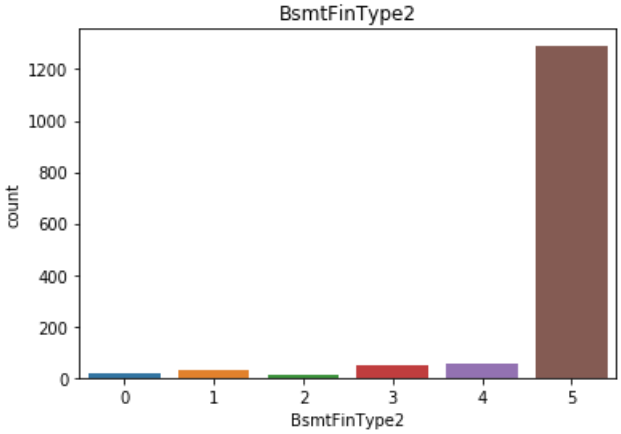

BsmtQual
count
600
500
400
300
200
100
0
0
1
2
3
BsmtQual

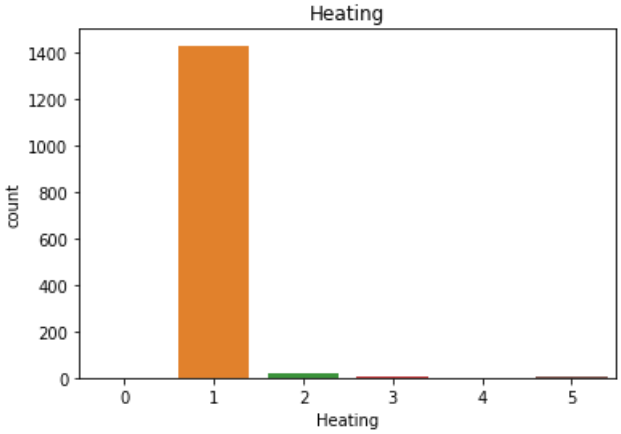

BsmtCond
count
1400
1200
1000
800
600
400
200
0
0
1
2
3
BsmtCond

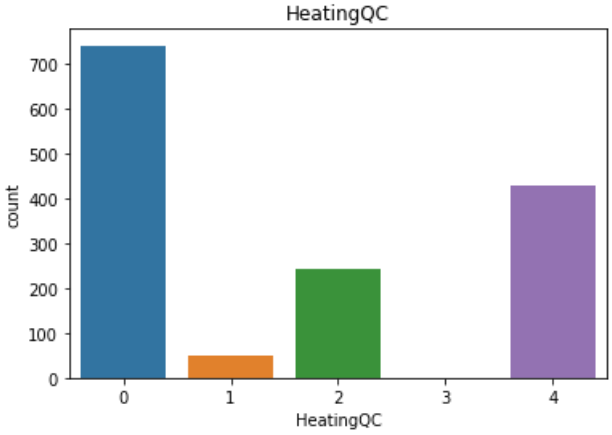

BsmtExposure
count
1000
800
600
400
200
0
0
1
2
3
BsmtExposure

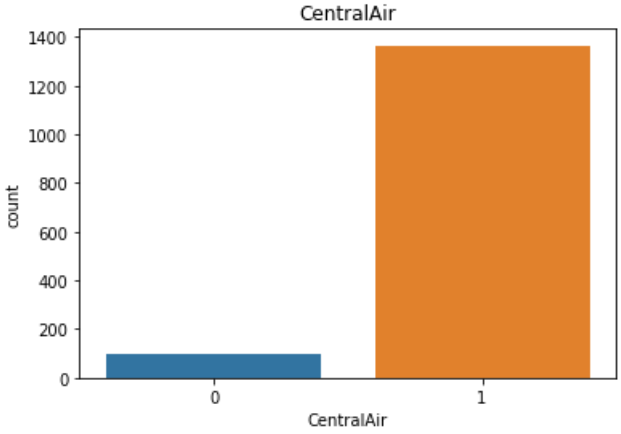

BsmtFinType1
count
400
300
200

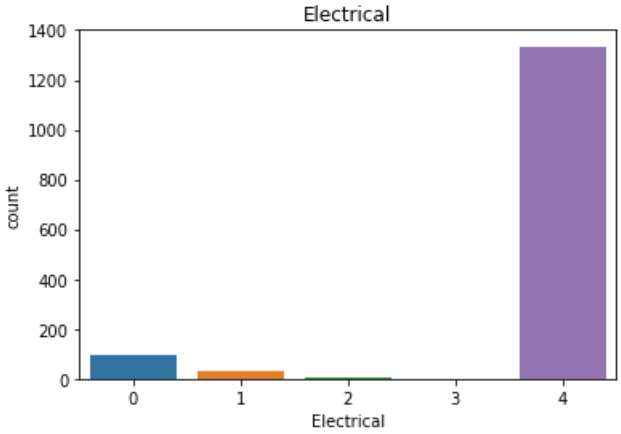

0
1
2
3
4
5
BsmtFinType1

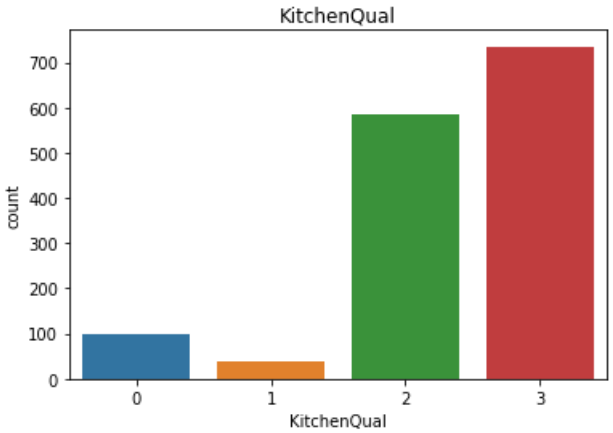

BsmtFinType2
count
0
200
400
600
800
1000
1200
0
1
2
3
4
5
BsmtFinType2

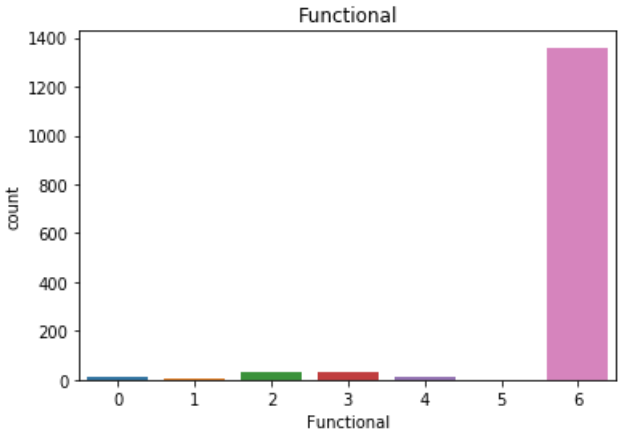

Heating
count
0
200
400
600
800
1000
1200
1400
0
1
2
3
4
5
Heating

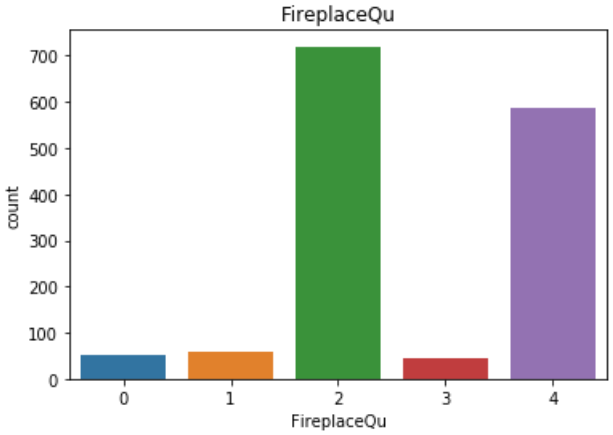

HeatingQC
count
0
100
200
300
400
500
600
700
0
1
2
3
4
HeatingQC

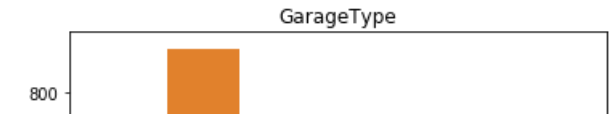

CentralAir
count
0
200
400
600
800
1000
1200
1400
0
1
CentralAir

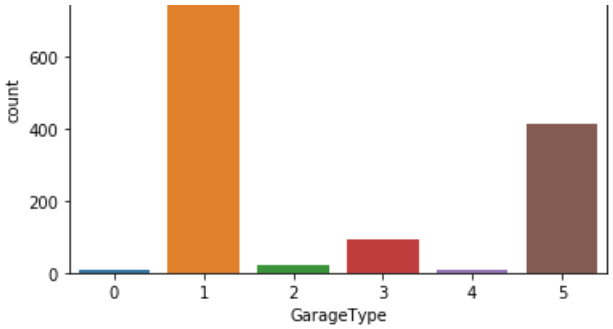

Electrical
count
Electrical

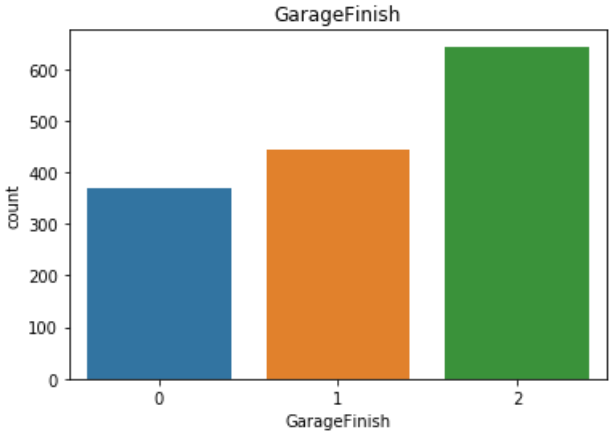

KitchenQual
count
KitchenQual

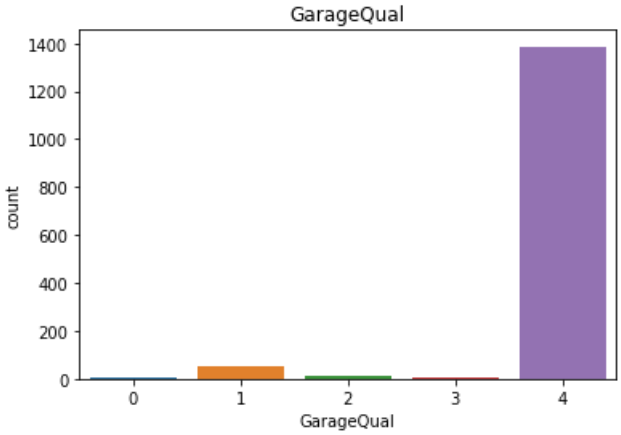

Functional
count
Functional

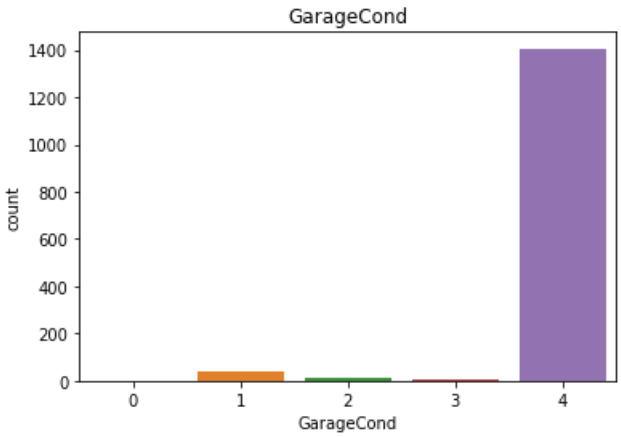

FireplaceQu
count
FireplaceQu

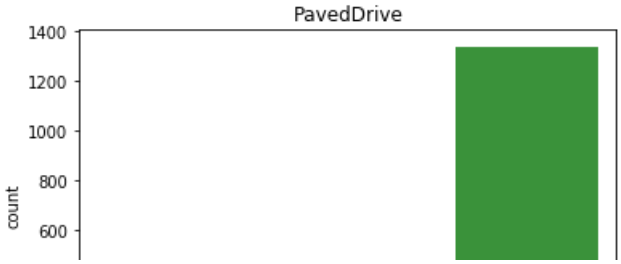

GarageType

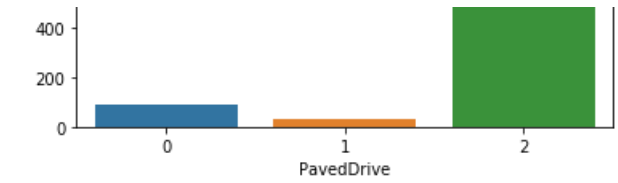

count
600
400
200
0
0
1
2
3
4
5
GarageType

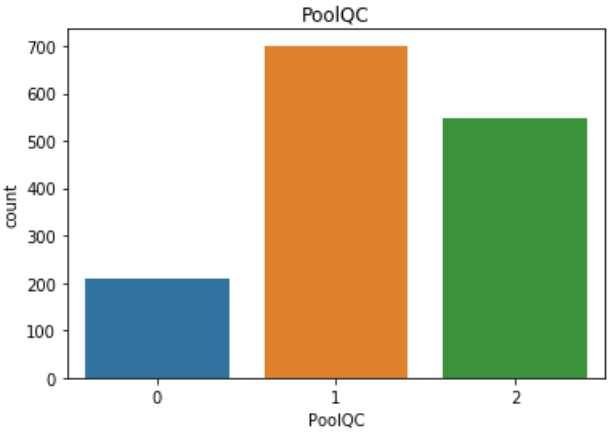

GarageFinish
count
600
500
400
300
200
100
0
0
1
2
GarageFinish

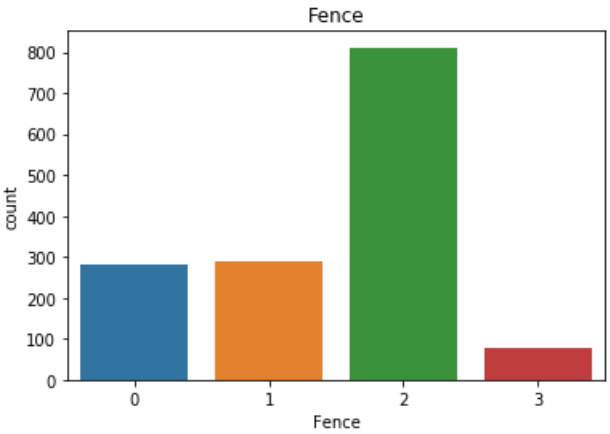

GarageQual
count
1400
1200
1000
800
600
400
200
0
0
1
2
3
4
GarageQual

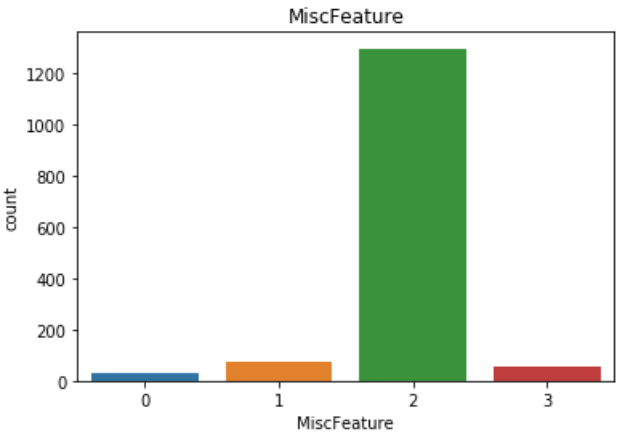

GarageCond
count
1400
1200
1000
800
600
400
200
0
0
1
2
3
4
GarageCond

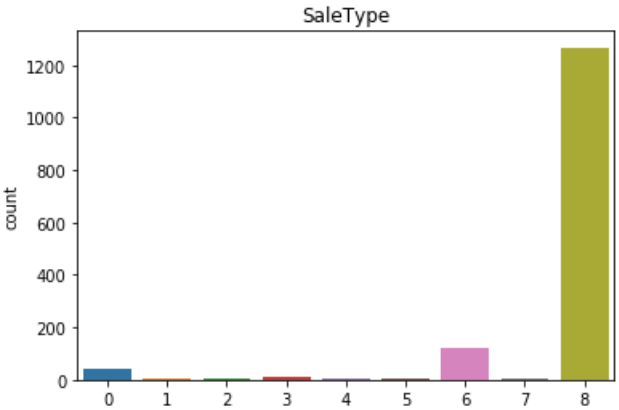

PavedDrive
count
1400
1200
1000
800
600

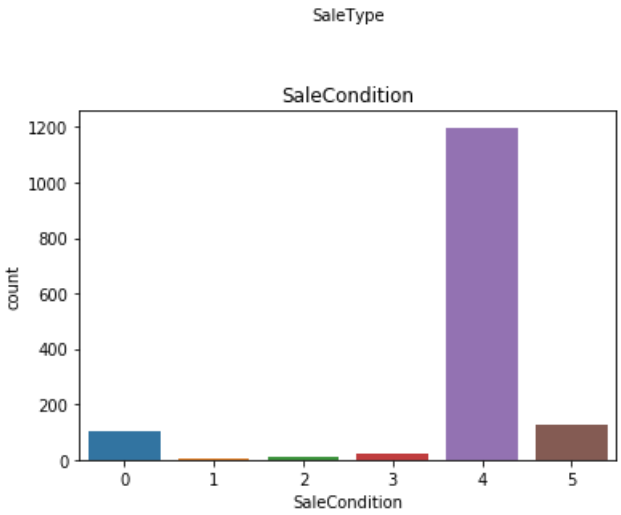

400
200
0
0
1
2
PavedDrive


PoolQC
700
600
500
400
300
200
100
0
count
0
1
2
PoolQC


Fence
800
700
600
500
400
300
200
100
0
count
0
1
2
3
Fence


MiscFeature
1200
1000
800
600
400
200
0
count
0
1
2
3
MiscFeature


SaleType
1200
1000
800
600
400
200
0
count
0
1
2
3
4
5
6
7
8



In [25]:

```
# Label encoding for test and train data
for i in columns:
    train[i]=label_encode.fit_transform(train[i])
    test[i]=label_encode.fit_transform(test[i])
```

In [26]:

```
train
```

Out[26]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	PoolArea	PoolQC	Fe
0	1	5	3	65.0	8450	1	0	3	3	0	...	0	0	
1	2	0	3	80.0	9600	1	0	3	3	0	...	0	0	
2	3	5	3	68.0	11250	1	0	0	3	0	...	0	0	
3	4	6	3	60.0	9550	1	0	0	3	0	...	0	0	
4	5	5	3	84.0	14260	1	0	0	3	0	...	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
1455	1456	5	3	62.0	7917	1	1	3	3	0	...	0	2	
1456	1457	0	3	85.0	13175	1	1	3	3	0	...	0	2	
1457	1458	6	3	66.0	9042	1	1	3	3	0	...	0	2	
1458	1459	0	3	68.0	9717	1	1	3	3	0	...	0	2	
1459	1460	0	3	75.0	9937	1	1	3	3	0	...	0	2	

1460 rows × 81 columns

In [27]:

```
test
```

Out[27]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	ScreenPorch	PoolArea
0	1461	0	2	80.0	11622	1	1	3	3	0	...	120	0
1	1462	0	3	81.0	14267	1	1	0	3	0	...	0	0
2	1463	5	3	74.0	13830	1	1	0	3	0	...	0	0
3	1464	5	3	78.0	9978	1	1	0	3	0	...	0	0
4	1465	11	3	43.0	5005	1	1	0	1	0	...	144	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1454	2915	13	4	21.0	1936	1	0	3	3	0	...	0	0

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	ScreenPorch	PoolArea
1455	2916	13	4	21.0	1894	1	0	3	3	0	...	0	0
1456	2917	0	3	160.0	20000	1	0	3	3	0	...	0	0
1457	2918	9	3	62.0	10441	1	0	3	3	0	...	0	0
1458	2919	5	3	74.0	9627	1	0	3	3	0	...	0	0

1459 rows × 80 columns

Descriptive Statistics

In [28]:

```
train.describe()
```

Out[28]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	
count	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460
mean	730.500000	4.166438	3.028767	70.306164	10516.828082	0.995890	0.453425	1.942466	2.777397	0
std	421.610009	4.161951	0.632017	23.945565	9981.264932	0.063996	0.497997	1.409156	0.707666	0
min	1.000000	0.000000	0.000000	21.000000	1300.000000	0.000000	0.000000	0.000000	0.000000	0
25%	365.750000	0.000000	3.000000	59.000000	7553.500000	1.000000	0.000000	0.000000	3.000000	0
50%	730.500000	4.000000	3.000000	70.000000	9478.500000	1.000000	0.000000	3.000000	3.000000	0
75%	1095.250000	6.000000	3.000000	80.000000	11601.500000	1.000000	1.000000	3.000000	3.000000	0
max	1460.000000	14.000000	4.000000	313.000000	215245.000000	1.000000	1.000000	3.000000	3.000000	1

8 rows × 81 columns

In [29]:

```
#Correlation of Attributes of the data
correlation=train.corr(method="spearman")
correlation
```

Out[29]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	Po
Id	1.000000	0.019226	-0.023256	-0.029631	-0.005364	0.008916	0.022942	0.029786	-0.016388	0.013324	...	0.0
MSSubClass	0.019226	1.000000	0.127902	-0.273617	-0.269570	-0.017276	0.009256	0.070499	-0.003433	-0.029594	...	0.0
MSZoning	-0.023256	0.127902	1.000000	-0.211224	-0.219857	0.036386	-0.082901	0.146903	0.005214	-0.005542	...	0.0
LotFrontage	-0.029631	-0.273617	-0.211224	1.000000	0.523980	-0.041278	0.028328	-0.128586	0.012951	0.036942	...	0.0
LotArea	-0.005364	-0.269570	-0.219857	0.523980	1.000000	-0.053676	-0.011124	-0.310579	-0.080551	0.036246	...	0.0
...	...	...	...	...	...	...	...	...	...	...	...	
MoSold	0.019018	0.018403	-0.035685	0.010191	0.006423	0.002495	0.001799	-0.035617	-0.030269	-0.043864	...	0.0
YrSold	0.001526	-0.027401	-0.016242	0.014633	-0.027473	-0.023701	-0.000329	0.037071	0.022947	0.024530	...	0.0
SaleType	0.026916	0.047649	0.130390	-0.079732	-0.055284	0.008380	0.011871	0.019892	-0.013105	-0.074835	...	0.0
SaleCondition	-0.013279	-0.062102	-0.094488	0.106834	0.080913	0.022390	0.019516	-0.054651	0.021831	-0.063162	...	0.0
SalePrice	-0.018546	0.007192	-0.334909	0.343535	0.456461	0.045814	0.066146	-0.305923	-0.010234	-0.016710	...	0.0

81 rows × 81 columns

In [30]:

```
#Heatmap of correlation between features in train data
correlation.style.background_gradient(cmap='viridis').set_precision(2)
```

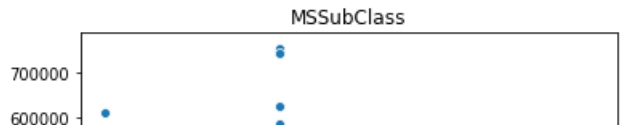
Out[30]:

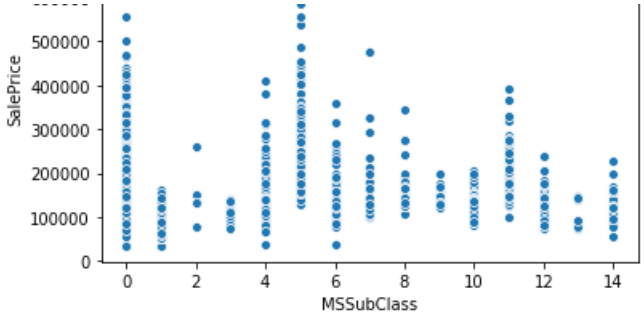
	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig
Id	1	0.019	-0.023	-0.03	-0.0054	0.0089	0.023	0.03	-0.016	0.013	0.047
MSSubClass	0.019	1	0.13	-0.27	-0.27	-0.017	0.0093	0.07	-0.0034	-0.03	0.045
MSZoning	-0.023	0.13	1	-0.21	-0.22	0.036	-0.083	0.15	0.0052	-0.0055	0.018
LotFrontage	-0.03	-0.27	-0.21	1	0.52	-0.041	0.028	-0.13	0.013	0.037	-0.14
LotArea	-0.0054	-0.27	-0.22	0.52	1	-0.054	-0.011	-0.31	-0.081	0.036	-0.2
Street	0.0089	-0.017	0.036	-0.041	-0.054	1	-0.006	-0.011	0.12	0.0017	0.012
Alley	0.023	0.0093	-0.083	0.028	-0.011	-0.006	1	-0.063	-0.018	0.029	-0.024
LotShape	0.03	0.07	0.15	-0.13	-0.31	-0.011	-0.063	1	0.12	-0.036	0.22
LandContour	-0.016	-0.0034	0.0052	0.013	-0.081	0.12	-0.018	0.12	1	0.0088	-0.022
Utilities	0.013	-0.03	-0.0055	0.037	0.036	0.0017	0.029	-0.036	0.0088	1	-0.033
LotConfig	0.047	0.045	0.018	-0.14	-0.2	0.012	-0.024	0.22	-0.022	-0.033	1
LandSlope	0.019	-0.02	-0.028	0.0033	0.12	-0.18	0.017	-0.093	-0.52	-0.0062	0.012
Neighborhood	-0.012	-0.0091	-0.21	0.12	0.093	-0.01	0.036	-0.03	0.038	0.044	-0.034
Condition1	-0.012	-0.013	-0.069	0.0085	0.088	-0.037	0.049	-0.11	0.028	0.003	0.062
Condition2	0.03	-0.03	0.016	0.0069	0.047	-0.00044	0.02	0.0092	-0.00049	0.00018	0.056
BldgType	0.019	0.65	0.1	-0.36	-0.43	-0.049	0.061	0.14	0.029	-0.012	0.1
HouseStyle	0.017	0.53	-0.12	0.082	0.061	0.022	0.01	-0.11	0.07	0.047	-0.035
OverallQual	-0.029	0.11	-0.21	0.21	0.23	0.059	0.055	-0.2	0.017	-0.00032	-0.032
OverallCond	0.0036	-0.072	0.19	-0.052	-0.047	0.049	-0.0037	0.02	-0.033	0.021	-0.039
YearBuilt	-0.0051	0.036	-0.35	0.16	0.1	0.032	0.065	-0.21	0.12	-0.018	0.0084
YearRemodAdd	-0.012	0.0068	-0.2	0.096	0.075	0.063	0.059	-0.13	0.062	-0.03	0.0021
RoofStyle	0.036	-0.14	-0.028	0.16	0.14	-0.02	-0.028	0.0093	0.0079	-0.013	-0.011
RoofMatl	-0.028	-0.019	-0.014	0.064	0.14	0.008	0.011	-0.075	-0.083	-0.0033	-0.07
Exterior1st	-0.037	-0.063	0.013	0.015	0.066	-0.0025	0.01	0.003	-0.014	-0.031	0.02
Exterior2nd	-0.041	-0.086	0.028	0.031	0.049	0.00038	-0.0016	0.0092	-0.036	-0.033	0.0049
MasVnrType	0.051	-0.056	0.019	-0.078	-0.067	0.0092	0.01	0.014	-0.089	-0.035	0.0025
MasVnrArea	-0.039	0.025	-0.088	0.2	0.18	0.013	0.013	-0.11	0.088	0.047	-0.0094
ExterQual	-0.0078	-0.052	0.24	-0.16	-0.13	0.035	-0.035	0.18	-0.018	0.02	0.0061
ExterCond	0.025	0.013	-0.12	0.077	0.033	0.0081	-0.017	-0.019	0.0075	0.0097	0.038
Foundation	0.0043	0.075	-0.26	0.096	0.051	0.045	0.046	-0.14	0.045	-0.018	-0.01
BsmtQual	0.04	-0.11	0.19	-0.12	-0.13	-0.031	-0.062	0.18	-2.3e-05	0.027	0.026
BsmtCond	0.0044	-0.029	-0.014	0.1	0.011	-0.019	0.017	-0.0063	0.018	0.0077	0.031
BsmtExposure	-0.03	-0.066	0.072	-0.11	-0.18	0.07	-0.014	0.15	0.11	0.018	0.026
BsmtFinType1	0.01	0.07	0.089	-0.058	-0.047	-0.015	-0.02	0.077	-0.053	-0.0091	0.029
BsmtFinSF1	-0.013	-0.11	-0.13	0.13	0.17	-0.013	0.03	-0.13	-0.018	-0.013	-0.01
BsmtFinType2	0.014	0.08	0.00079	-0.047	-0.073	0.047	-0.023	0.025	-0.03	-0.066	-0.0067
BsmtFinSF2	-0.0067	-0.084	-0.0055	0.049	0.072	-0.044	0.023	-0.033	0.026	0.073	0.0038
BsmtUnfSF	-0.0095	-0.12	-0.024	0.092	0.078	0.037	0.01	-0.0027	0.034	-0.011	-0.0044
TotalBsmtSF	-0.033	-0.32	-0.2	0.33	0.37	0.0096	0.063	-0.18	0.014	-0.02	-0.035
Heating	0.026	0.062	0.099	-0.047	-0.0035	0.009	-0.0052	0.08	0.017	-0.0037	-0.0042
HeatingQC	0.00045	0.0012	0.15	-0.059	-0.069	-0.052	-0.015	0.1	-0.062	0.012	-0.0042
CentralAir	0.0098	-0.11	-0.12	0.099	0.11	0.07	0.023	-0.11	0.082	0.0069	-0.0092

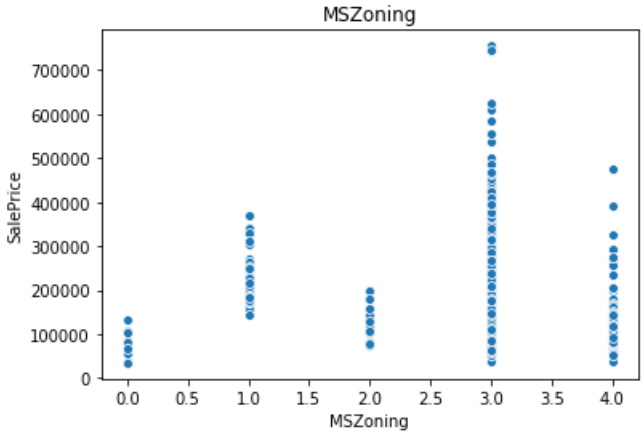
	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig
Electrical	0.009	0.043	-0.12	0.082	0.077	0.019	-0.015	-0.1	0.068	-0.087	-0.012
1stFlrSF	-0.00062	-0.28	-0.17	0.36	0.44	0.0087	0.026	-0.17	-0.036	0.02	-0.044
2ndFlrSF	0.0095	0.49	-0.012	0.028	0.12	0.042	-0.0053	-0.047	-0.022	-0.022	-0.043
LowQualFinSF	-0.028	0.076	0.036	-0.014	-0.02	0.0086	-0.019	0.043	-0.046	-0.0035	-0.0048
GrLivArea	0.0028	0.2	-0.16	0.3	0.45	0.049	0.0094	-0.18	-0.047	-0.0085	-0.069
BsmtFullBath	0.0047	-0.042	-0.071	0.066	0.096	-0.041	0.022	-0.081	-0.018	-0.022	-0.018
BsmtHalfBath	-0.021	0.0025	-0.011	-0.019	0.046	0.016	-0.00071	-0.026	-0.0021	0.11	-0.012
FullBath	0.0073	0.19	-0.24	0.17	0.24	0.048	0.052	-0.16	0.049	-0.028	-0.01
HalfBath	0.0026	0.28	-0.14	0.074	0.14	0.028	-0.0097	-0.12	0.034	-0.02	-0.031
BedroomAbvGr	0.042	0.069	-0.083	0.25	0.34	0.035	-0.014	-0.074	0.0057	0.0059	-0.06
KitchenAbvGr	0.0076	0.28	0.05	0.0021	-0.023	0.014	0.014	0.092	-0.036	-0.0056	0.0028
KitchenQual	-0.011	-0.03	0.18	-0.15	-0.13	-0.034	-0.047	0.15	0.028	-0.021	-0.0042
TotRmsAbvGrd	0.026	0.17	-0.095	0.3	0.41	0.052	-0.031	-0.12	-0.017	0.014	-0.041
Functional	-0.0022	0.031	-0.093	0.028	-0.037	-0.017	0.024	-0.025	0.034	0.0071	-0.019
Fireplaces	-0.014	0.019	-0.09	0.21	0.35	0.0058	0.023	-0.21	-0.063	0.02	-0.054
FireplaceQu	0.031	0.07	-0.008	-0.017	-0.035	0.024	-0.065	-0.039	0.067	-0.018	0.0053
GarageType	-0.004	0.12	0.2	-0.23	-0.22	-0.017	-0.026	0.16	-0.061	0.015	0.00049
GarageYrBlt	0.011	0.067	-0.26	0.088	0.038	0.03	0.082	-0.16	0.074	-0.026	-0.005
GarageFinish	-0.011	-0.025	0.22	-0.2	-0.17	-0.011	-0.069	0.22	-0.05	-0.0091	0.021
GarageCars	0.013	0.024	-0.21	0.29	0.34	-0.028	0.025	-0.2	0.0074	0.0092	-0.076
GarageArea	0.0065	-0.047	-0.25	0.32	0.37	-0.031	0.044	-0.19	0.016	0.01	-0.079
GarageQual	-0.0012	-0.031	-0.18	0.062	0.041	0.033	0.057	-0.1	-0.0015	0.006	0.011
GarageCond	-0.021	-0.04	-0.11	0.053	0.032	-0.013	0.047	-0.071	0.0078	0.0051	0.037
PavedDrive	0.00058	-0.06	-0.18	0.13	0.067	0.021	0.051	-0.12	0.13	0.0078	-0.055
WoodDeckSF	-0.043	0.023	-0.074	0.11	0.18	0.014	0.051	-0.17	-0.017	-0.023	-0.02
OpenPorchSF	-0.0032	0.032	-0.18	0.14	0.18	0.013	0.022	-0.12	0.058	0.036	-0.043
EnclosedPorch	-0.0068	0.011	0.19	-0.1	-0.067	0.026	-0.042	0.11	-0.057	-0.011	-0.045
3SsnPorch	-0.037	-0.036	-0.015	0.056	0.062	0.0083	0.044	-0.036	-0.06	-0.0034	-0.018
ScreenPorch	0.006	-0.022	-0.0096	0.039	0.092	-0.023	0.00034	-0.047	-0.0082	0.094	0.0097
PoolArea	0.056	0.033	-0.015	0.078	0.084	0.0045	-0.023	-0.024	-0.011	-0.0018	-0.048
PoolQC	0.77	0.016	-0.042	-0.05	-0.00037	-0.011	0.012	0.042	0.0055	0.031	0.046
Fence	-0.11	-0.018	0.02	0.0017	-0.044	-0.031	-0.12	-0.0019	0.023	0.048	0.018
MiscFeature	0.16	-0.019	0.03	0.026	0.009	-0.0066	0.044	-0.021	-0.032	0.0027	-0.029
MiscVal	-0.043	-0.033	0.0027	0.017	0.059	-0.16	-0.07	-0.023	0.006	-0.005	-0.022
MoSold	0.019	0.018	-0.036	0.01	0.0064	0.0025	0.0018	-0.036	-0.03	-0.044	0.024
YrSold	0.0015	-0.027	-0.016	0.015	-0.027	-0.024	-0.00033	0.037	0.023	0.025	-0.0045
SaleType	0.027	0.048	0.13	-0.08	-0.055	0.0084	0.012	0.02	-0.013	-0.075	-0.0068
SaleCondition	-0.013	-0.062	-0.094	0.11	0.081	0.022	0.02	-0.055	0.022	-0.063	0.05
SalePrice	-0.019	0.0072	-0.33	0.34	0.46	0.046	0.066	-0.31	-0.01	-0.017	-0.074

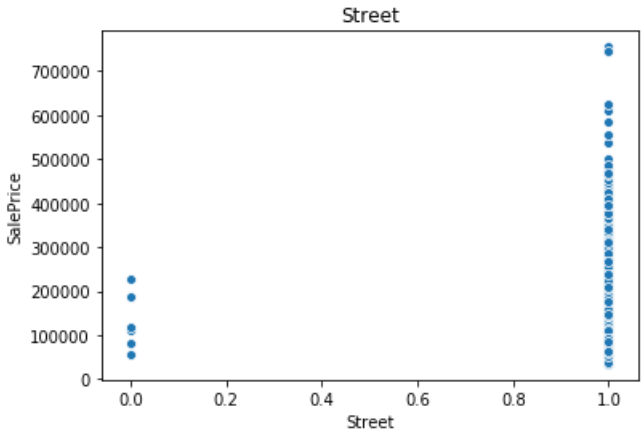
In [32]:

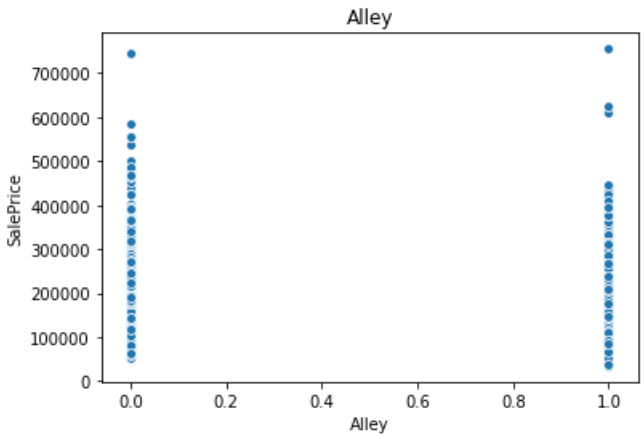
```
for column in columns:
    plt.title(column)
    sns.scatterplot(column,"SalePrice",data=train)
    plt.show()
```

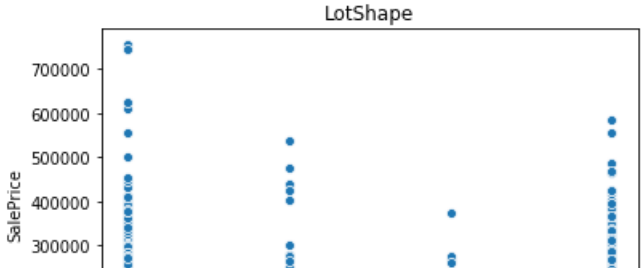


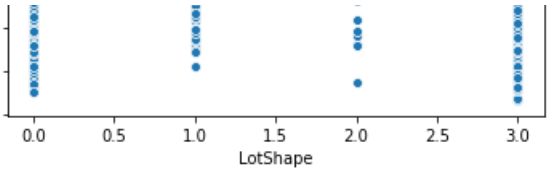

SalePrice
MSSubClass

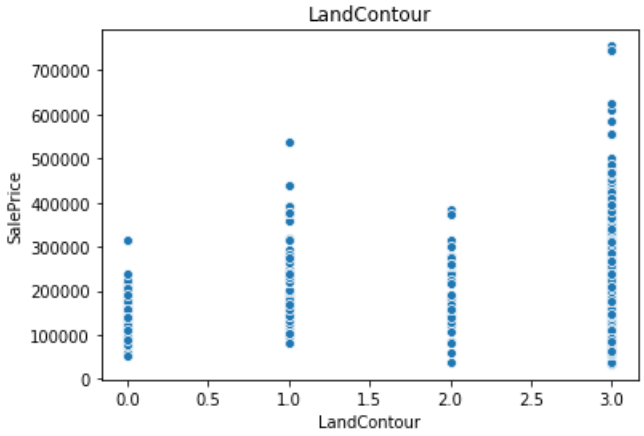

MSZoning
SalePrice
MSZoning

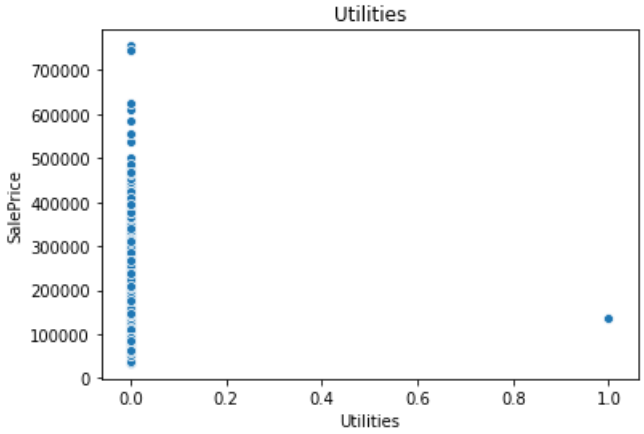

Street
SalePrice
Street

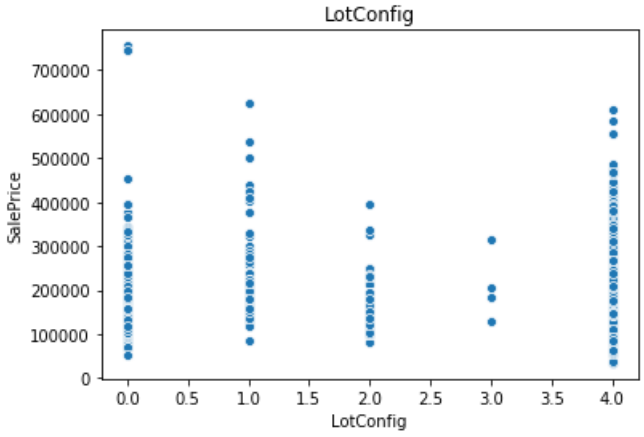

Alley
SalePrice
Alley

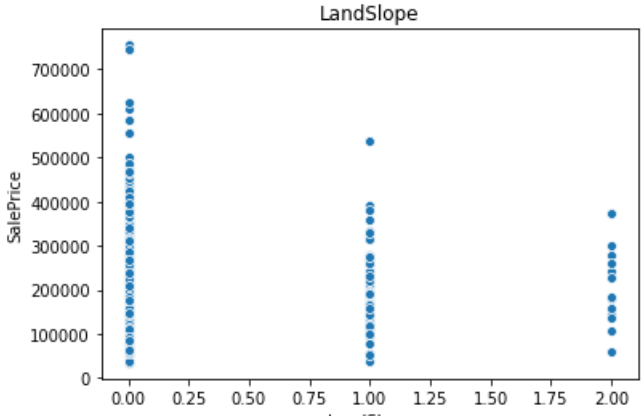

LotShape
SalePrice


200000
100000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
LotShape


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

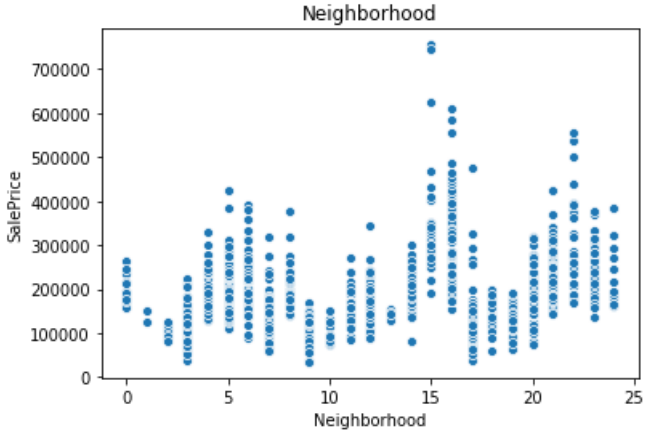

Utilities
700000
600000
500000
400000
300000
200000
100000
0
SalePrice
0.0
0.2
0.4
0.6
0.8
1.0
Utilities


LotConfig
700000
600000
500000
400000
300000
200000
100000
0
SalePrice
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
LotConfig

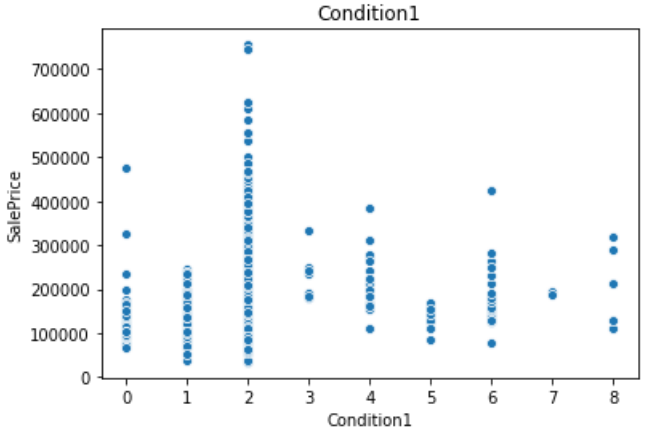

LandSlope
700000
600000
500000
400000
300000
200000
100000
0
SalePrice
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

LandSlope

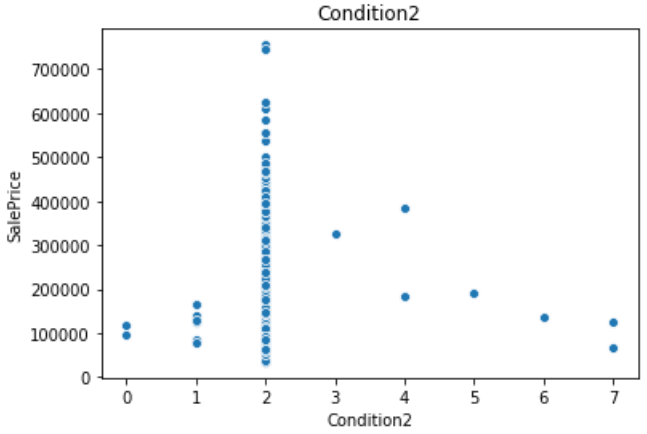
Neighborhood



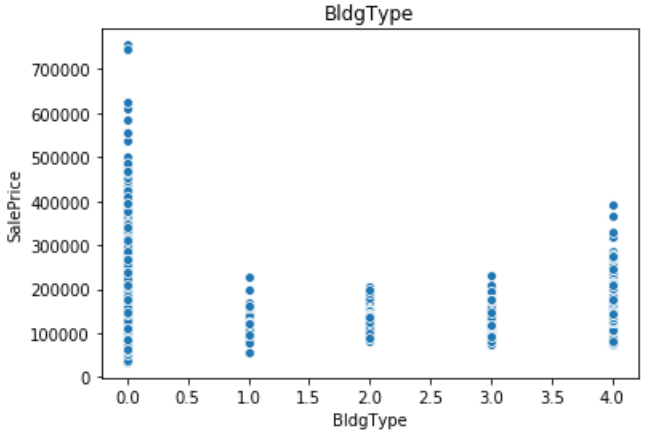
Condition1



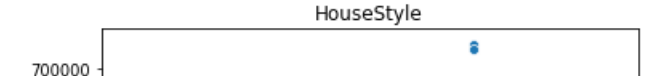
Condition2

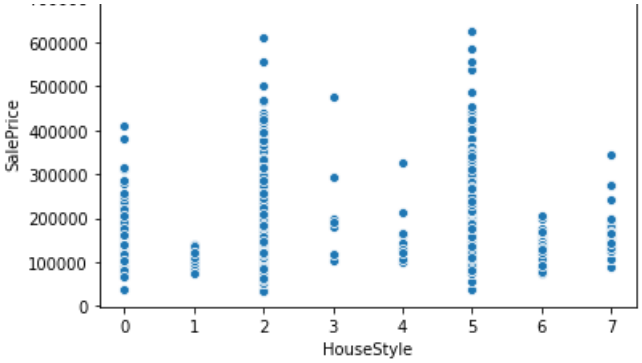


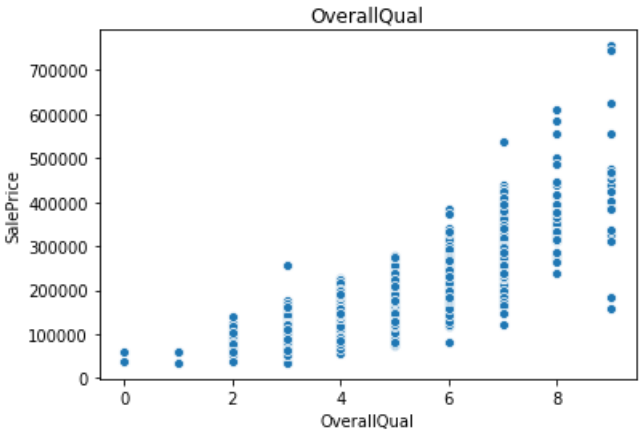
BldgType

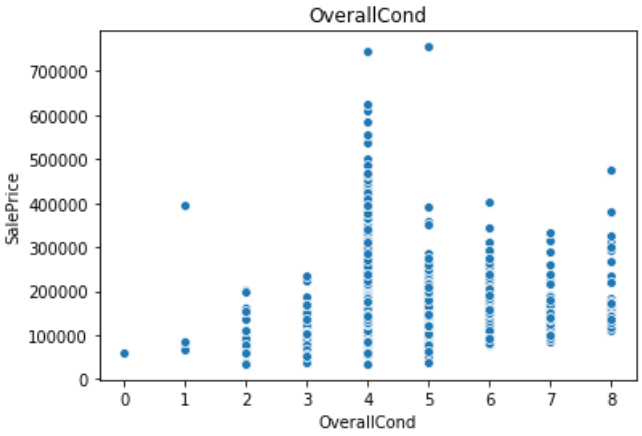


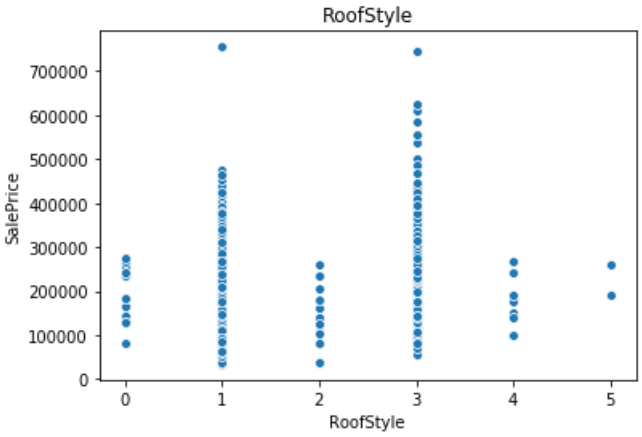
HouseStyle

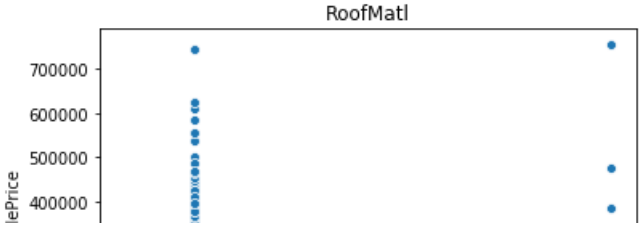


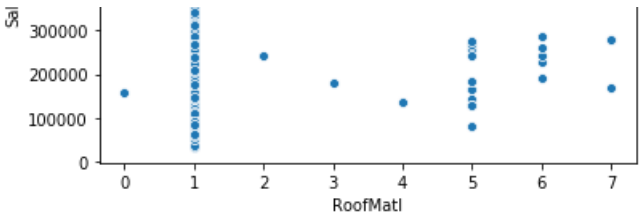

SalePrice
HouseStyle

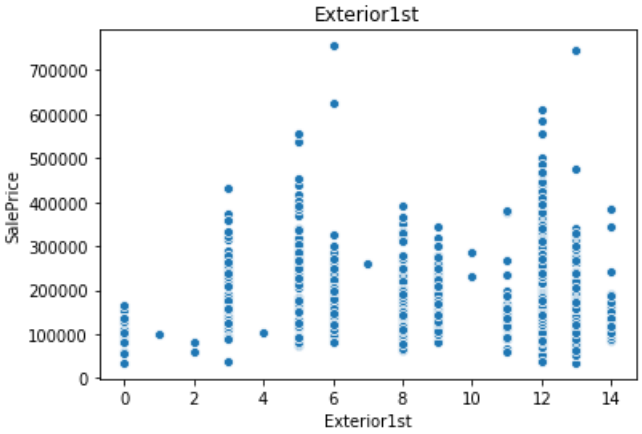

OverallQual
SalePrice
OverallQual

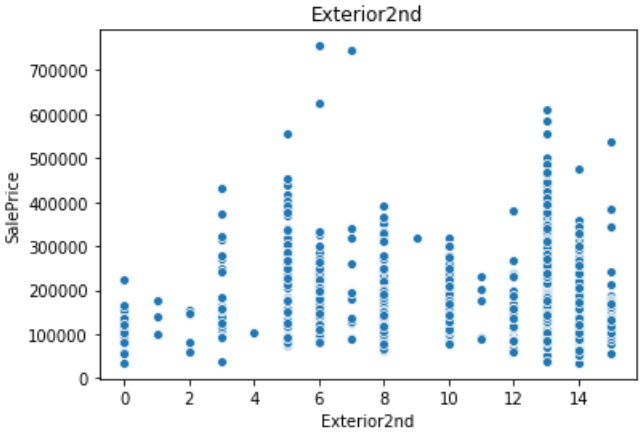

OverallCond
SalePrice
OverallCond

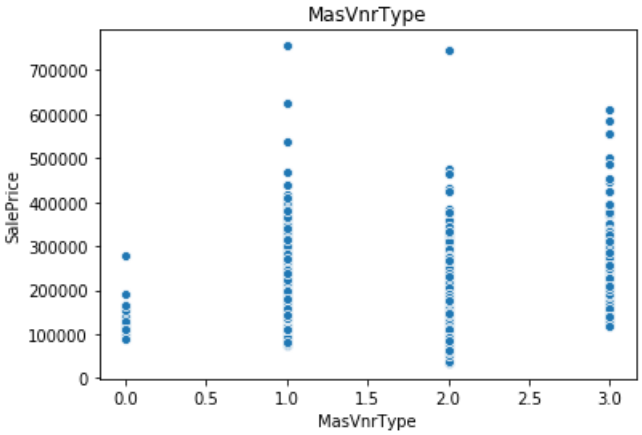

RoofStyle
SalePrice
RoofStyle

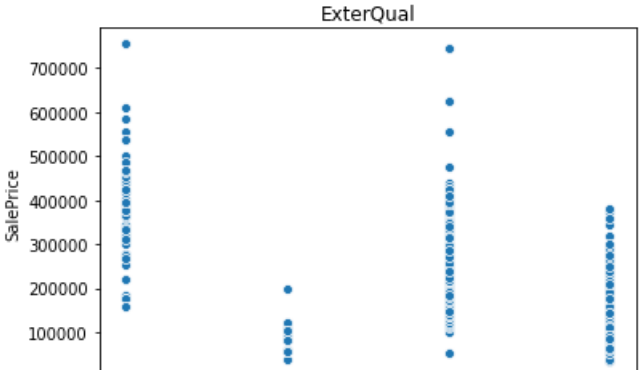

RoofMatl

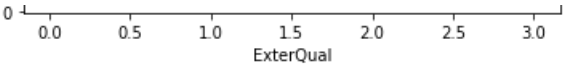

Sal
300000
200000
100000
0
0
1
2
3
4
5
6
7
RoofMatl

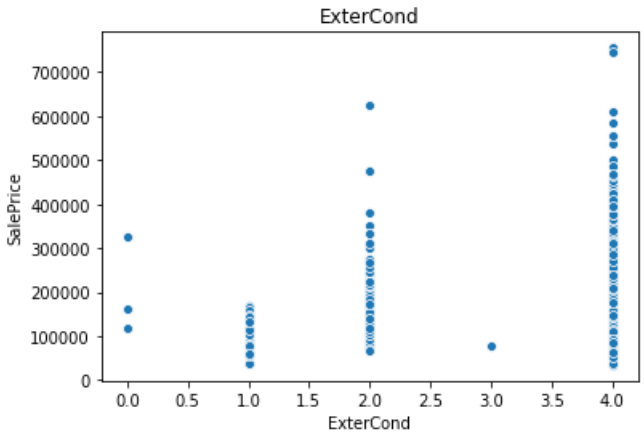

Exterior1st
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
2
4
6
8
10
12
14
Exterior1st

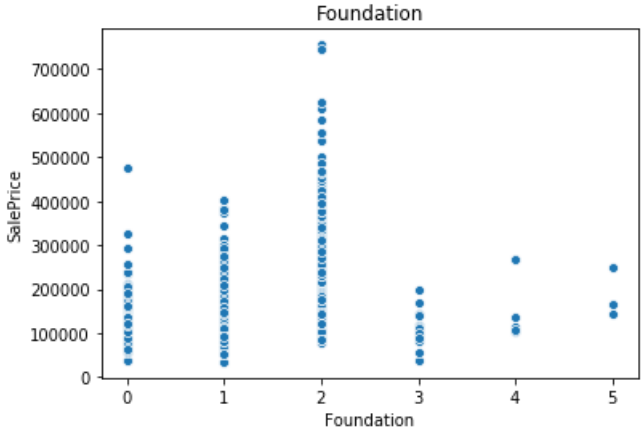

Exterior2nd
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
2
4
6
8
10
12
14
Exterior2nd

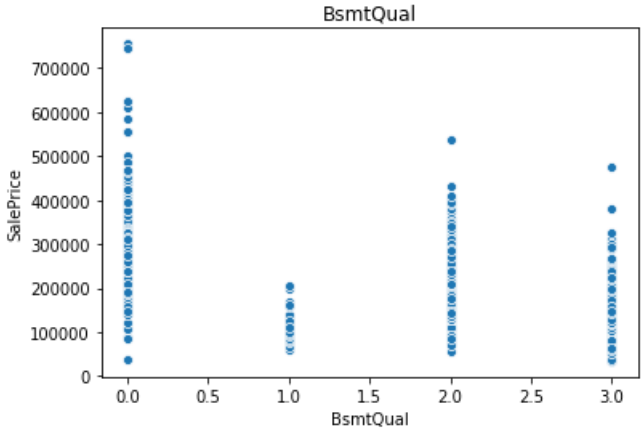

MasVnrType
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
MasVnrType

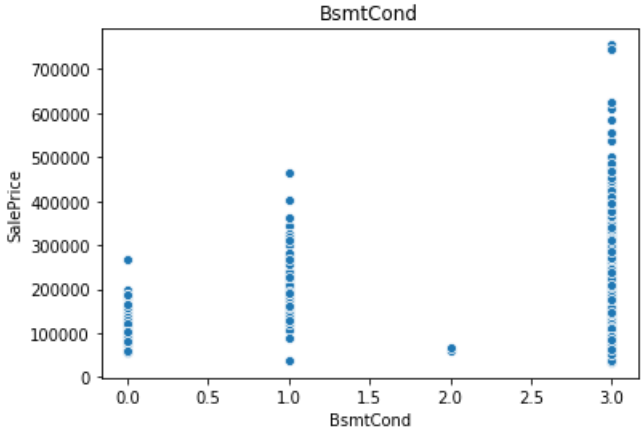

ExterQual
SalePrice
700000
600000
500000
400000
300000
200000
100000

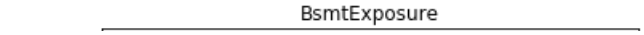

0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
ExterQual

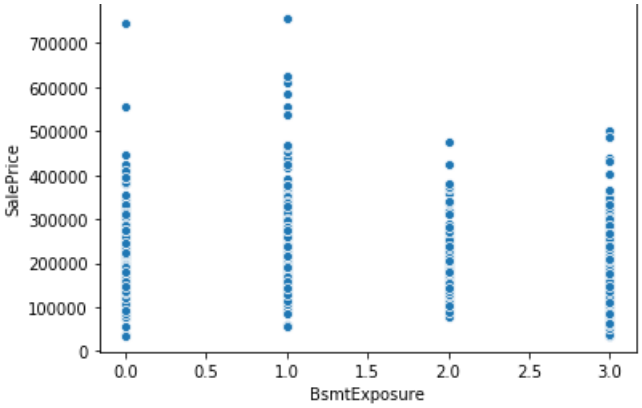

ExterCond
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
ExterCond

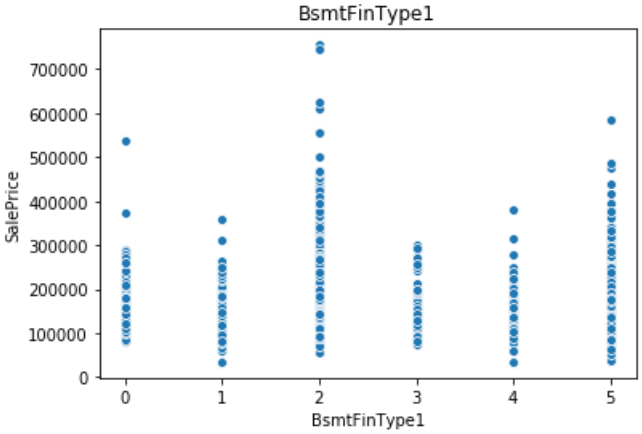

Foundation
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
1
2
3
4
5
Foundation

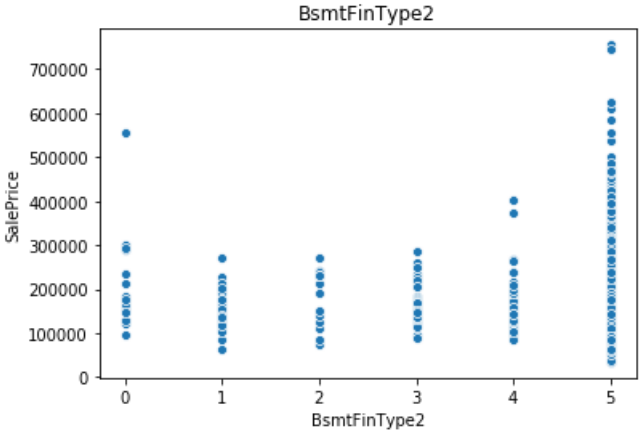

BsmtQual
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
BsmtQual

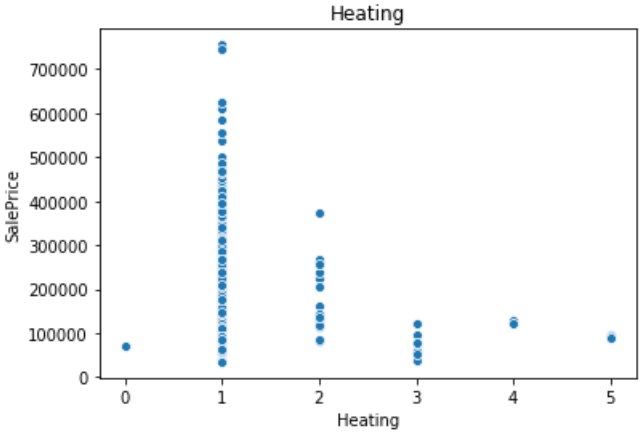

BsmtCond
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
BsmtCond

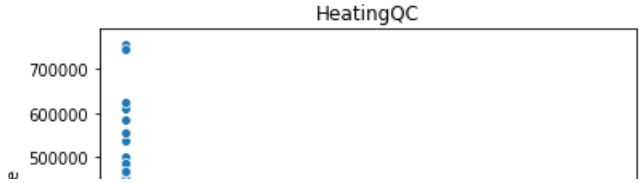

BsmtExposure

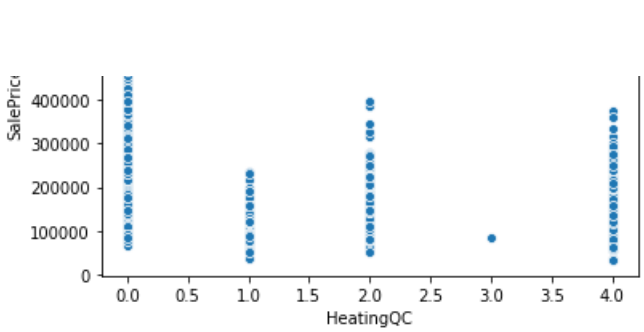

SalePrice
BsmtExposure

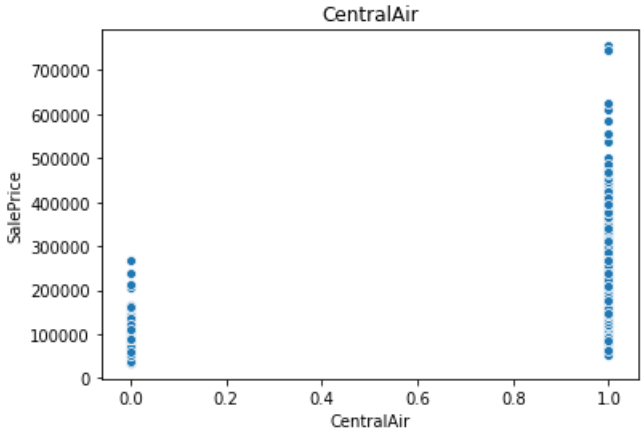

BsmtFinType1
SalePrice
BsmtFinType1

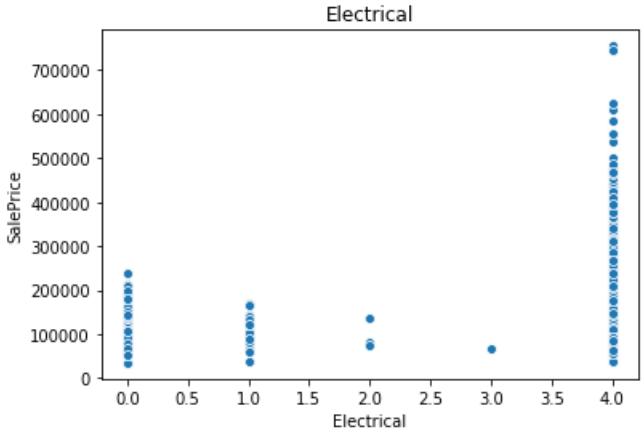

BsmtFinType2
SalePrice
BsmtFinType2

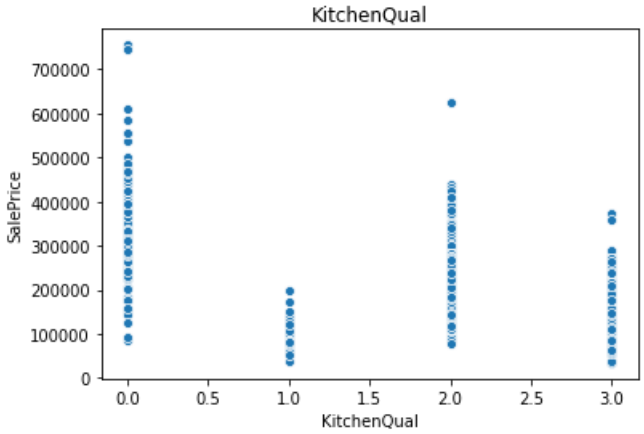

Heating
SalePrice
Heating

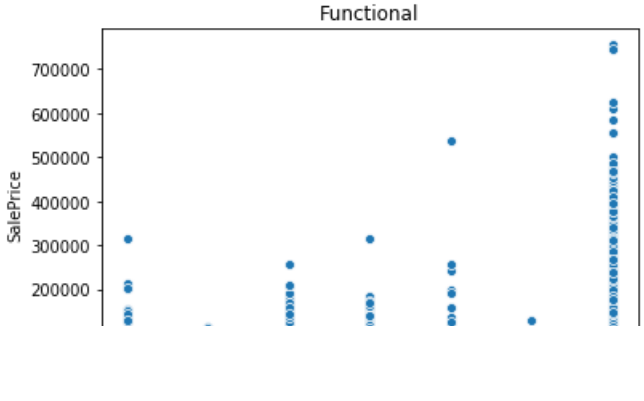

HeatingQC

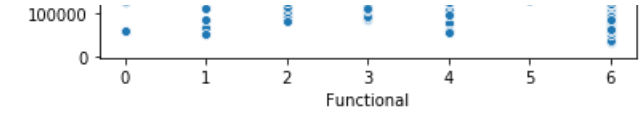

SalePrice
400000
300000
200000
100000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
HeatingQC

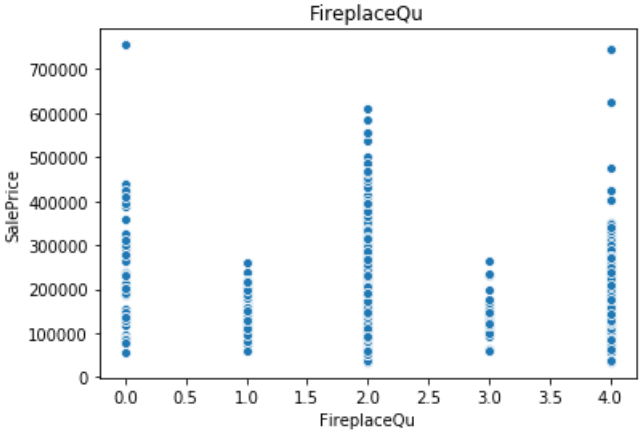

CentralAir
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.0
0.2
0.4
0.6
0.8
1.0
CentralAir

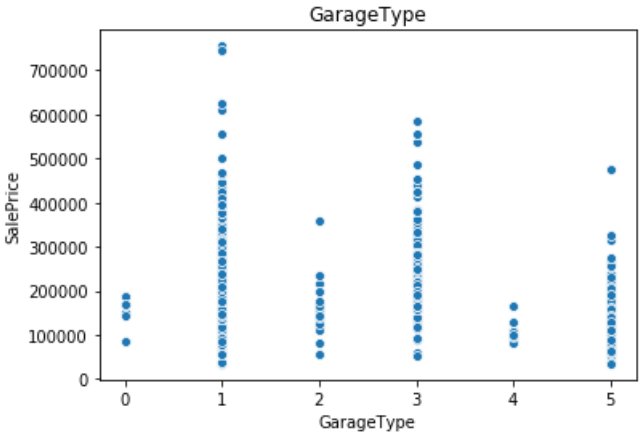

Electrical
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Electrical

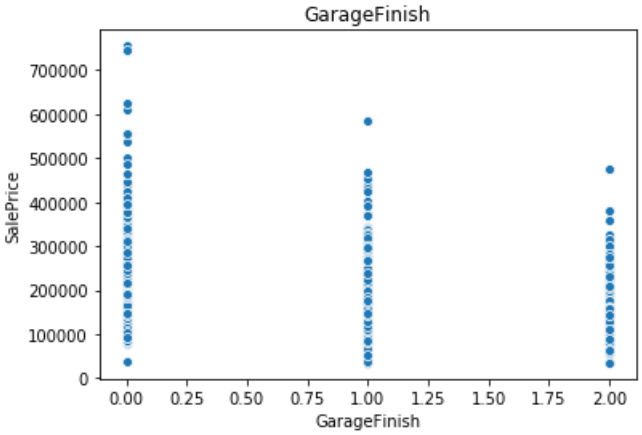

KitchenQual
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
KitchenQual

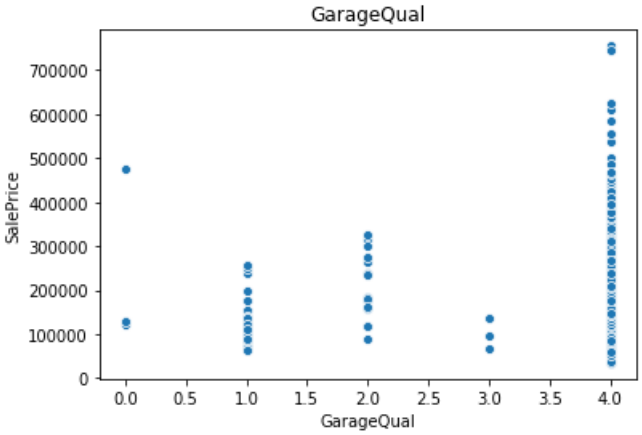

Functional
SalePrice
700000
600000
500000
400000
300000
200000

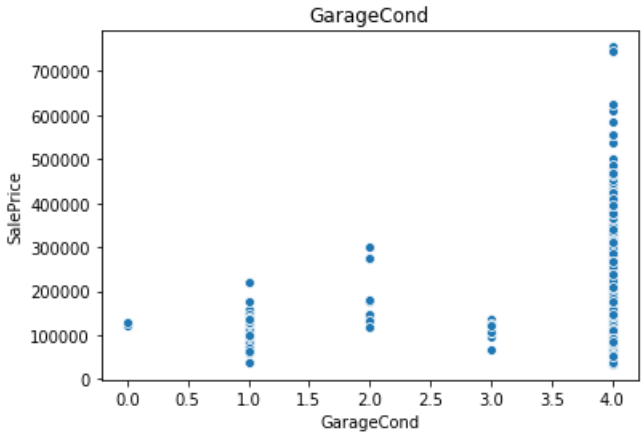

100000
0
0
1
2
3
4
5
6
Functional

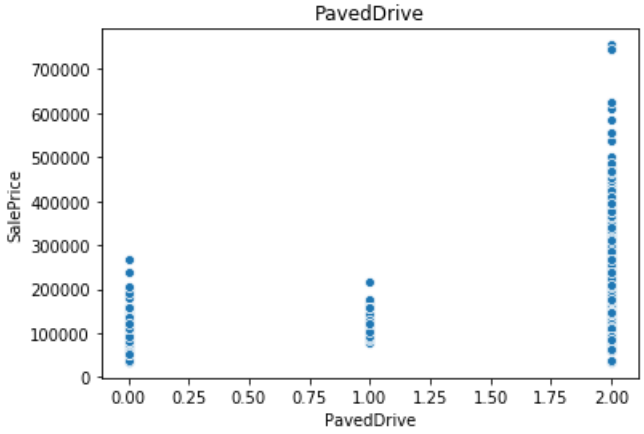

FireplaceQu
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
FireplaceQu

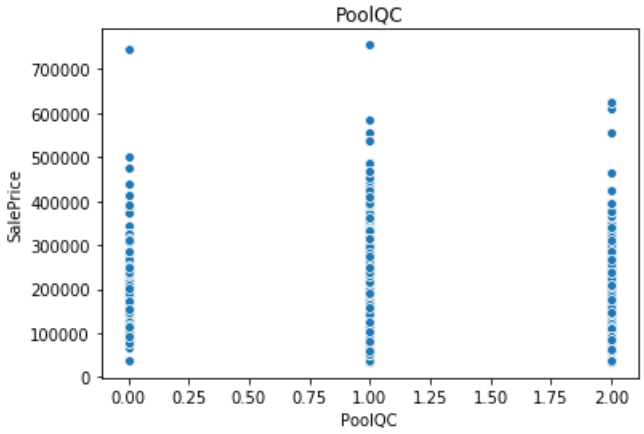

GarageType
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0
1
2
3
4
5
GarageType

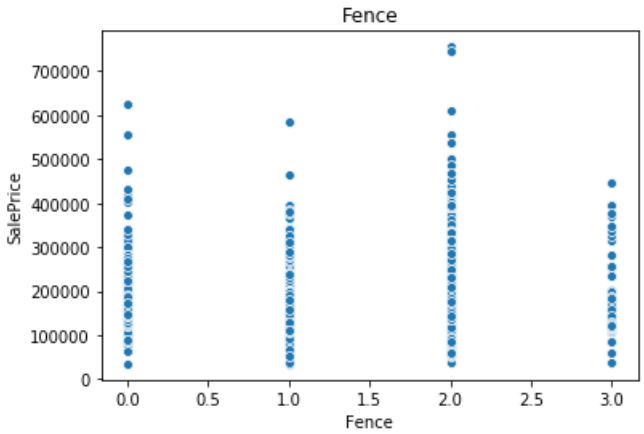

GarageFinish
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
GarageFinish

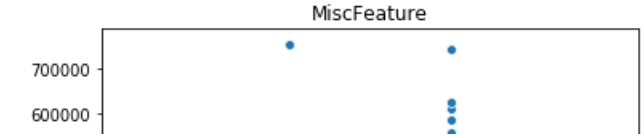

GarageQual
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
GarageQual

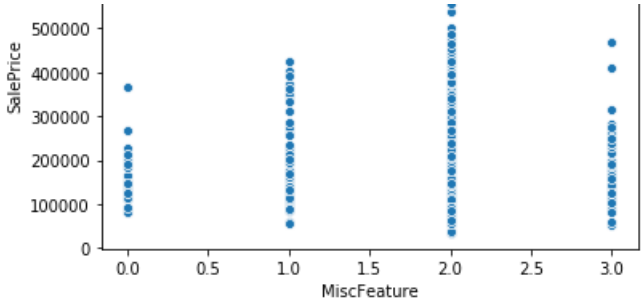

GarageCond
SalePrice
GarageCond

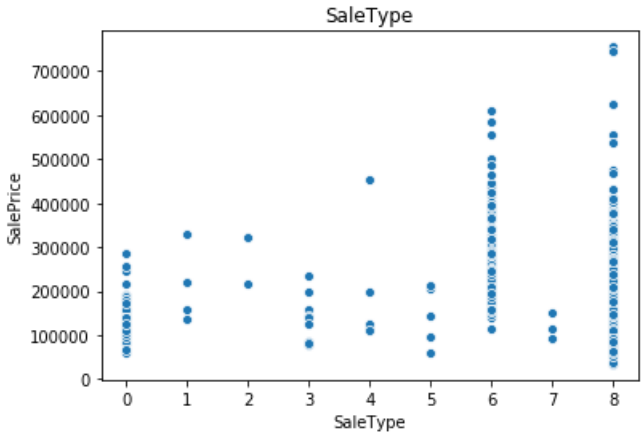

PavedDrive
SalePrice
PavedDrive

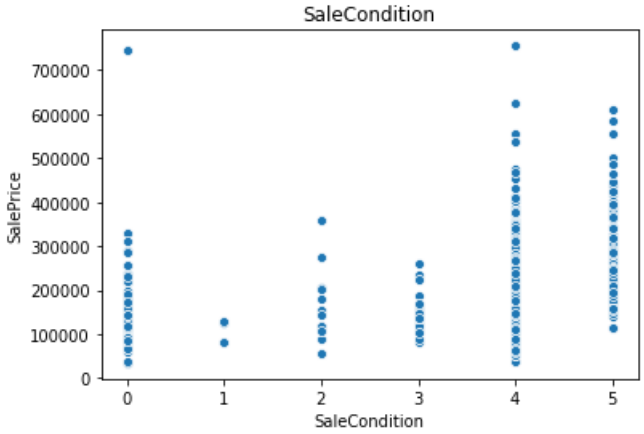

PoolQC
SalePrice
PoolQC


Fence
SalePrice
Fence


MiscFeature

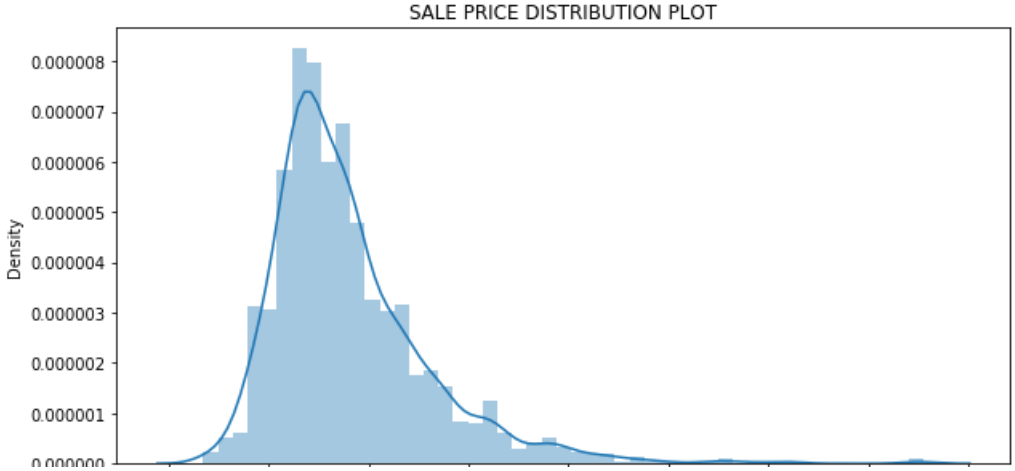






In [49]:

```
# Distribution plot for sales
sns.distplot(train["SalePrice"])
plt.title("SALE PRICE DISTRIBUTION PLOT")
plt.ylabel("Density")
plt.xlabel("Sale Price")
plt.xticks(rotation=90)
plt.show()
```



MODEL

In [34]:

```
x_train=pd.DataFrame(train[['MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street',
       'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig',
       'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType',
       'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd',
       'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType',
       'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual',
       'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1',
       'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating',
       'HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF',
       'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath',
       'HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual',
       'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType',
       'GarageYrBlt', 'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual',
       'GarageCond', 'PavedDrive', 'WoodDeckSF', 'OpenPorchSF',
       'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'PoolQC',
       'Fence', 'MiscFeature', 'MiscVal', 'MoSold', 'YrSold', 'SaleType',
       'SaleCondition']])
y_train=pd.DataFrame(train.iloc[:,-1])
```

In [35]:

```
x_test=pd.DataFrame(test[['MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street',
       'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig',
       'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType',
       'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd',
       'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType',
       'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual',
       'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1',
       'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating',
       'HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF',
       'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath',
       'HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual',
       'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType',
       'GarageYrBlt', 'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual',
       'GarageCond', 'PavedDrive', 'WoodDeckSF', 'OpenPorchSF',
       'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'PoolQC',
       'Fence', 'MiscFeature', 'MiscVal', 'MoSold', 'YrSold', 'SaleType',
       'SaleCondition']])
```

In [36]:

```
#FIT AND PREDICT SALE PRICE
```

In [37]:

```
# Import and use random forest
from sklearn.ensemble import RandomForestRegressor as RF
```

In [38]:

```
# Model Fit and Predict
regressor = RF(n_estimators=100,
               criterion='mse',
               max_features= None,
               max_depth = 14,bootstrap=True)
regressor=regressor.fit(x_train, y_train.values.ravel())
regressor.fit(x_train, y_train.values.ravel())
y_pred = regressor.predict(x_test)
```

In [39]:

```
y_train.values.ravel()
```

Out[39]:

```
array([208500, 181500, 223500, ..., 266500, 142125, 147500], dtype=int64)
```

In [40]:

```
# Make predicted values as DataFrame
y_pred=pd.DataFrame(y_pred)
```

In [41]:

```
y_pred.rename(columns={0:"SalePrice"},inplace=True)
```

In [42]:

```
y_pred["Id"]=test["Id"]
```

In [43]:

```
# values for submission
final=y_pred[["Id","SalePrice"]]
```

In [50]:

```
final.head()
```

Out[50]:

	Id	SalePrice
0	1461	124437.154544
1	1462	152514.897412
2	1463	181226.686719
3	1464	180756.510065
4	1465	200426.302265

In [ ]:

```
#Submission
final.to_csv("PriceSubmission.csv",index=False)
```

In [ ]:

```

```