

Housing Price Prediction

Submitted by:

Rakesh Shinde

**INTRODUCTION**

**Problem Statement:**

Houses are one of the necessary needs of each and every person around the globe and therefore housing and real estate market is one of the markets which is one of the major contributors in the world’s economy. It is a very large market and there are various companies working in the domain. Data science comes as a very important tool to solve problems in the domain to help the companies increase their overall revenue, profits, improving their marketing strategies and focusing on changing trends in house sales and purchases. Predictive modelling, Market mix modelling, recommendation systems are some of the machine learning techniques used for achieving the business goals for housing companies. Our problem is related to one such housing company.

A US-based housing company named **Surprise Housing** has decided to enter the Australian market. The company uses data analytics to purchase houses at a price below their actual values and flip them at a higher price. For the same purpose, the company has collected a data set from the sale of houses in Australia. The data is provided in the CSV file below.

The company is looking at prospective properties to buy houses to enter the market. You are required to build a model using Machine Learning in order to predict the actual value of the prospective properties and decide whether to invest in them or not. For this company wants to know:

• Which variables are important to predict the price of variable?

• How do these variables describe the price of the house?

**Business Goal:**

You are required to model the price of houses with the available independent variables. This model will then be used by the management to understand how exactly the prices vary with the variables. They can accordingly manipulate the strategy of the firm and concentrate on areas that will yield high returns. Further, the model will be a good way for the management to understand the pricing dynamics of a new market.

**Data Loading to Jupyter**

* Data set was given in two sets train.csv and test.csv
* Loaded to Jupyter Notebook for further analysis

**Uderstanding the Dataset**

* + Train Data set is having 1168 rows and 81 columns including the target column
  + Test Dataset is having 292 row and 80 columns
  + Checked data for null values present in columns are as follows

As we can see that there are multiple columns are having null values. It has to taken care for the either treating it and dropping the column

1. PoolQC 1161

2. MiscFeature 1124

3. Alley 1091

4. Fence 931

5. FireplaceQu 551

6. LotFrontage 214

7. GarageYrBlt 64

8. GarageFinish 64

9. GarageType 64

10. GarageQual 64

11. GarageCond 64

12. BsmtExposure 31

13. BsmtFinType2 31

14. BsmtQual 30

15. BsmtCond 30

16. BsmtFinType1 30

17. MasVnrType 7

18. MasVnrArea 7

For PoolQC, MiscFeature, Alley, Fence, FireplaceQu the Null values are either equal to 50% or more than that so we can think of droping that columns

For rest of the columns we can fill that columns according type of the columns by mode or mean

**Data Type**

The data types are checked and found to be as the mentioned data following are the type of each column

MSSubClass int64

MSZoning int32

LotFrontage float64

LotArea int64

Street int32

LotShape int32

LandContour int32

Utilities int32

LotConfig int32

LandSlope int32

Neighborhood int32

Condition1 int32

Condition2 int32

BldgType int32

HouseStyle int32

OverallQual int64

OverallCond int64

YearBuilt int64

YearRemodAdd int64

RoofStyle int32

RoofMatl int32

Exterior1st int32

Exterior2nd int32

MasVnrType int32

MasVnrArea float64

ExterQual int32

ExterCond int32

Foundation int32

BsmtQual int32

BsmtCond int32

BsmtExposure int32

BsmtFinType1 int32

BsmtFinSF1 int64

BsmtFinType2 int32

BsmtFinSF2 int64

BsmtUnfSF int64

TotalBsmtSF int64

Heating int32

HeatingQC int32

CentralAir int32

Electrical int32

1stFlrSF int64

2ndFlrSF int64

LowQualFinSF int64

GrLivArea int64

BsmtFullBath int64

BsmtHalfBath int64

FullBath int64

HalfBath int64

BedroomAbvGr int64

KitchenAbvGr int64

KitchenQual int32

TotRmsAbvGrd int64

Functional int32

Fireplaces int64

GarageType int32

GarageYrBlt float64

GarageFinish int32

GarageCars int64

GarageArea int64

GarageQual int32

GarageCond int32

PavedDrive int32

WoodDeckSF int64

OpenPorchSF int64

EnclosedPorch int64

3SsnPorch int64

ScreenPorch int64

PoolArea int64

MiscVal int64

MoSold int64

YrSold int64

SaleType int32

SaleCondition int32

SalePrice int64

**Note: - Since the columns and the content in the column is having correct data types so need not to be checked for the void spaces and others**

**Filling the Null Data**

* Null data is filled with the mean and mode of that column according to its type

1. LotFrontage 214
2. GarageYrBlt 64
3. GarageFinish 64
4. GarageType 64
5. GarageQual 64
6. GarageCond 64
7. BsmtExposure 31
8. BsmtFinType2 31
9. BsmtQual 30
10. BsmtCond 30
11. BsmtFinType1 30
12. MasVnrType 7
13. MasVnrArea 7

**Using the Describe Function**

1. MSSubClass, LotArea, OverallQual, OverallCond, BsmtFinSF1, OpenPorchSF, MoSold are positively skewed as mean is more than median
2. LotFrantage is normally distributed
3. Rest is having some complex distribution can be seen from the Histogram

**VISUALIZATION**

**Count Plot for Categorical Data**

Countplot for MSZoning

RL 928

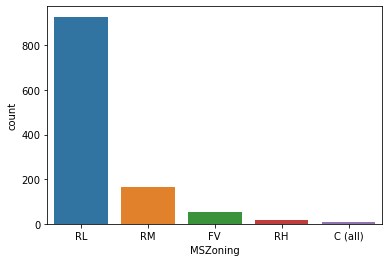
RM 163

FV 52

RH 16

C (all) 9

Name: MSZoning, dtype: int64

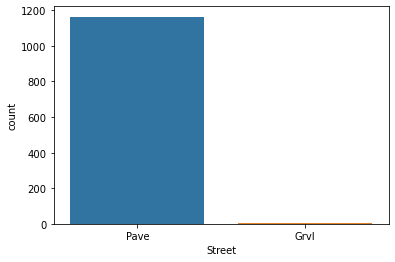
****

Countplot for Street

Pave 1164

Grvl 4

Name: Street, dtype: int64

****

Countplot for LotShape

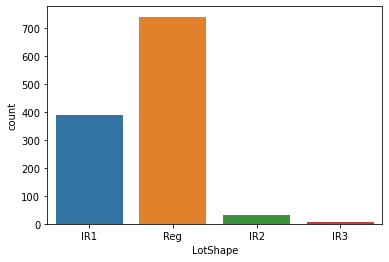
Reg 740

IR1 390

IR2 32

IR3 6

Name: LotShape, dtype: int64

****

Countplot for LandContour

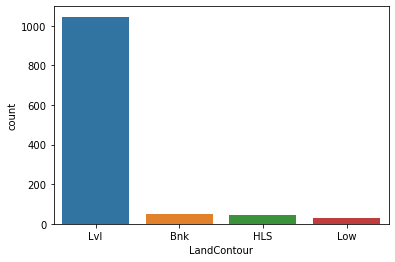
Lvl 1046

Bnk 50

HLS 42

Low 30

Name: LandContour, dtype: int64

****

Countplot for Utilities

AllPub 1168

Name: Utilities, dtype: int64

****

Countplot for LotConfig

Inside 842

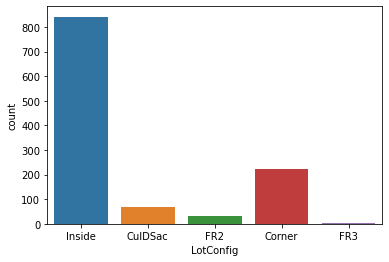
Corner 222

CulDSac 69

FR2 33

FR3 2

Name: LotConfig, dtype: int64

****

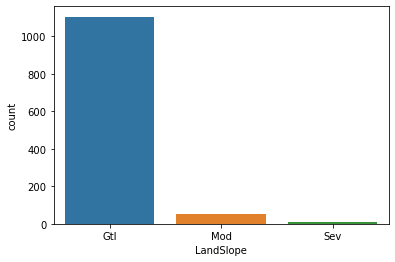
Countplot for LandSlope

Gtl 1105

Mod 51

Sev 12

Name: LandSlope, dtype: int64

****

Countplot for Neighborhood

NAmes 182

CollgCr 118

OldTown 86

Edwards 83

Somerst 68

Gilbert 64

NridgHt 61

Sawyer 60

NWAmes 59

SawyerW 51

BrkSide 50

Crawfor 45

NoRidge 35

Mitchel 34

IDOTRR 30

Timber 24

ClearCr 24

SWISU 21

StoneBr 19

Blmngtn 15

BrDale 11

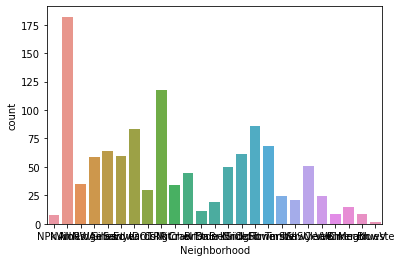
MeadowV 9

Veenker 9

NPkVill 8

Blueste 2

Name: Neighborhood, dtype: int64

****

Countplot for Condition1

Norm 1005

Feedr 67

Artery 38

RRAn 20

PosN 17

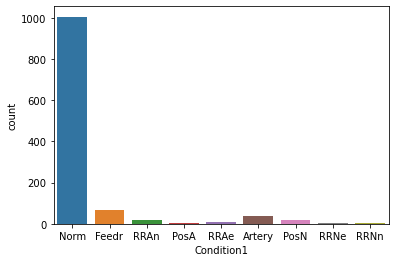
RRAe 9

PosA 6

RRNn 4

RRNe 2

Name: Condition1, dtype: int64

****

Countplot for Condition2

Norm 1154

Feedr 6

PosN 2

Artery 2

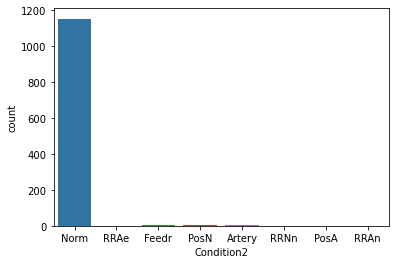
RRAe 1

RRNn 1

PosA 1

RRAn 1

Name: Condition2, dtype: int64

****

Countplot for BldgType

1Fam 981

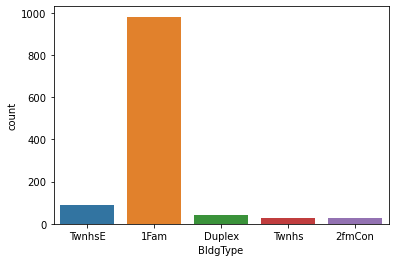
TwnhsE 90

Duplex 41

Twnhs 29

2fmCon 27

Name: BldgType, dtype: int64

****

Countplot for HouseStyle

1Story 578

2Story 361

1.5Fin 121

SLvl 47

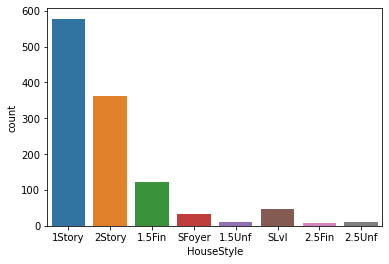
SFoyer 32

1.5Unf 12

2.5Unf 10

2.5Fin 7

Name: HouseStyle, dtype: int64

****

Countplot for RoofStyle

Gable 915

Hip 225

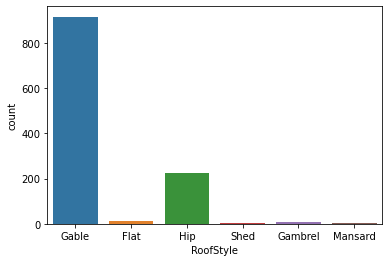
Flat 12

Gambrel 9

Mansard 5

Shed 2

Name: RoofStyle, dtype: int64

****

Countplot for RoofMatl

CompShg 1144

Tar&Grv 10

WdShngl 6

WdShake 4

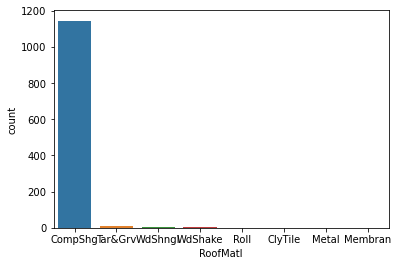
Roll 1

ClyTile 1

Metal 1

Membran 1

Name: RoofMatl, dtype: int64

****

Countplot for Exterior1st

VinylSd 396

HdBoard 179

MetalSd 178

Wd Sdng 174

Plywood 93

CemntBd 42

BrkFace 41

Stucco 22

WdShing 19

AsbShng 19

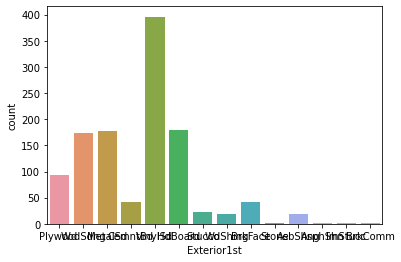
Stone 2

AsphShn 1

ImStucc 1

BrkComm 1

Name: Exterior1st, dtype: int64

****

Countplot for Exterior2nd

VinylSd 387

MetalSd 173

HdBoard 170

Wd Sdng 165

Plywood 118

CmentBd 42

Wd Shng 31

Stucco 23

BrkFace 20

AsbShng 18

ImStucc 8

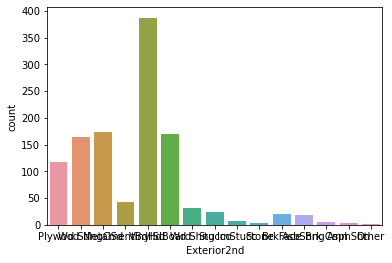
Brk Cmn 5

Stone 4

AsphShn 3

Other 1

Name: Exterior2nd, dtype: int64

****

Countplot for MasVnrType

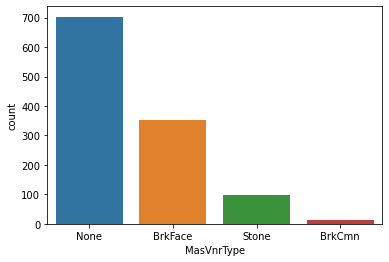
None 703

BrkFace 354

Stone 98

BrkCmn 13

Name: MasVnrType, dtype: int64

****

Countplot for ExterQual

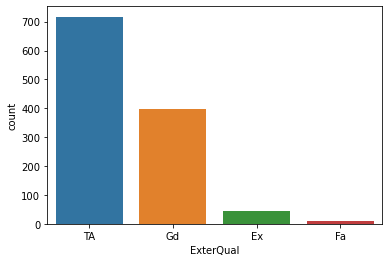
TA 717

Gd 397

Ex 43

Fa 11

Name: ExterQual, dtype: int64

****

Countplot for ExterCond

TA 1022

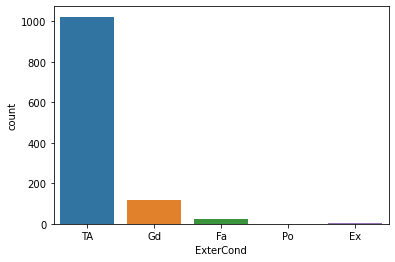
Gd 117

Fa 26

Ex 2

Po 1

Name: ExterCond, dtype: int64

****

Countplot for Foundation

CBlock 516

PConc 513

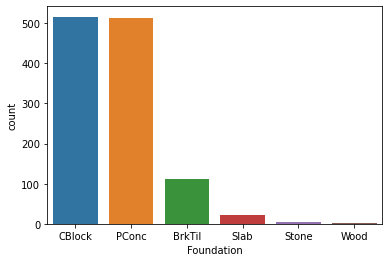
BrkTil 112

Slab 21

Stone 5

Wood 1

Name: Foundation, dtype: int64

****

Countplot for BsmtQual

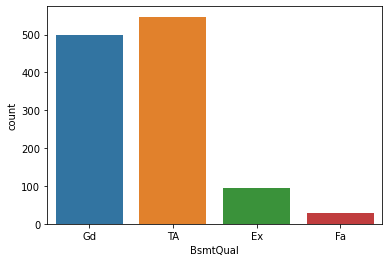
TA 547

Gd 498

Ex 94

Fa 29

Name: BsmtQual, dtype: int64

****

Countplot for BsmtCond

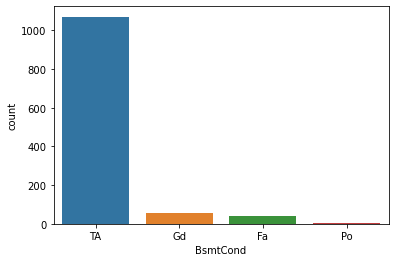
TA 1071

Gd 56

Fa 39

Po 2

Name: BsmtCond, dtype: int64

****

Countplot for BsmtExposure

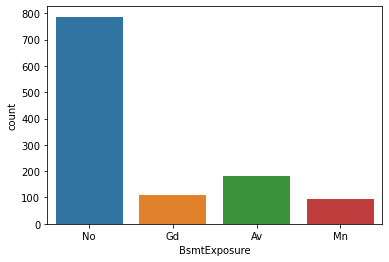
No 787

Av 180

Gd 108

Mn 93

Name: BsmtExposure, dtype: int64

****

Countplot for BsmtFinType1

Unf 375

GLQ 330

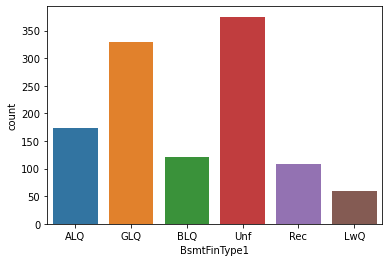
ALQ 174

BLQ 121

Rec 109

LwQ 59

Name: BsmtFinType1, dtype: int64

****

Countplot for BsmtFinType2

Unf 1033

Rec 43

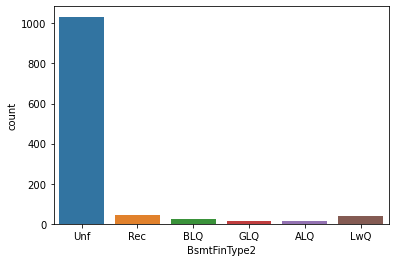
LwQ 40

BLQ 24

ALQ 16

GLQ 12

Name: BsmtFinType2, dtype: int64

****

Countplot for Heating

GasA 1143

GasW 14

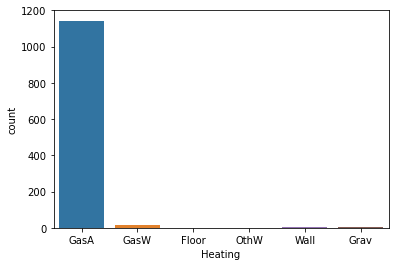
Grav 5

Wall 4

Floor 1

OthW 1

Name: Heating, dtype: int64

****

Countplot for HeatingQC

Ex 585

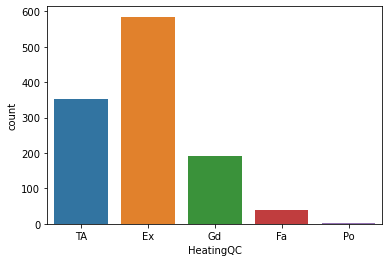
TA 352

Gd 192

Fa 38

Po 1

Name: HeatingQC, dtype: int64

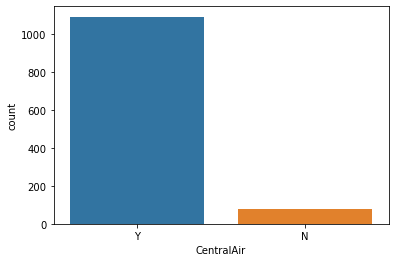
****

Countplot for CentralAir

Y 1090

N 78

Name: CentralAir, dtype: int64

****

Countplot for Electrical

SBrkr 1070

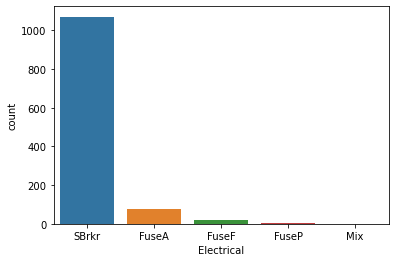
FuseA 74

FuseF 21

FuseP 2

Mix 1

Name: Electrical, dtype: int64

****

Countplot for KitchenQual

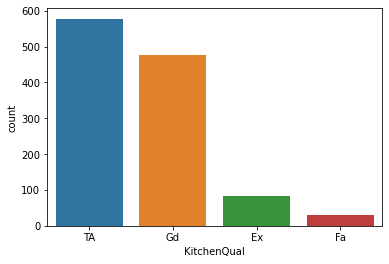
TA 578

Gd 478

Ex 82

Fa 30

Name: KitchenQual, dtype: int64

****

Countplot for Functional

Typ 1085

Min2 30

Min1 25

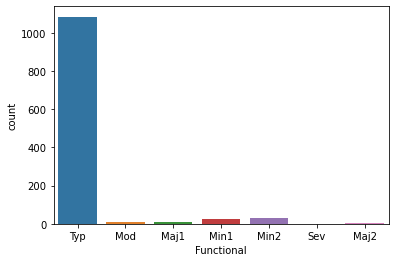
Mod 12

Maj1 11

Maj2 4

Sev 1

Name: Functional, dtype: int64

****

Countplot for GarageType

Attchd 755

Detchd 314

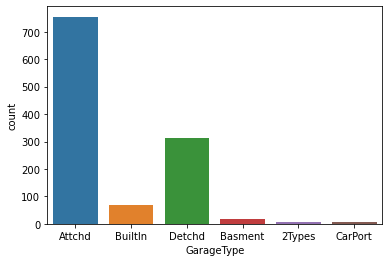
BuiltIn 70

Basment 16

CarPort 8

2Types 5

Name: GarageType, dtype: int64

****

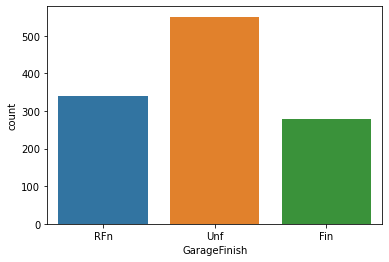
Countplot for GarageFinish

Unf 551

RFn 339

Fin 278

Name: GarageFinish, dtype: int64

****

Countplot for GarageQual

TA 1114

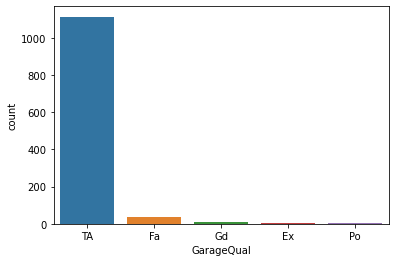
Fa 39

Gd 11

Ex 2

Po 2

Name: GarageQual, dtype: int64

****

Countplot for GarageCond

TA 1125

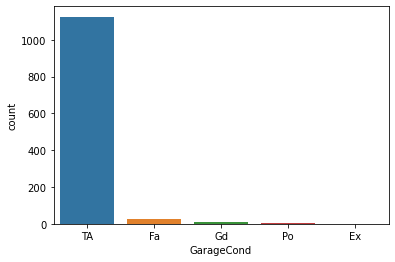
Fa 28

Gd 8

Po 6

Ex 1

Name: GarageCond, dtype: int64

****

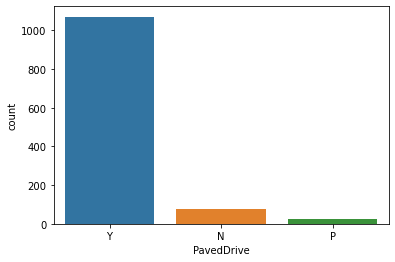
Countplot for PavedDrive

Y 1071

N 74

P 23

Name: PavedDrive, dtype: int64

****

Countplot for SaleType

WD 999

New 106

COD 38

ConLD 8

ConLI 5

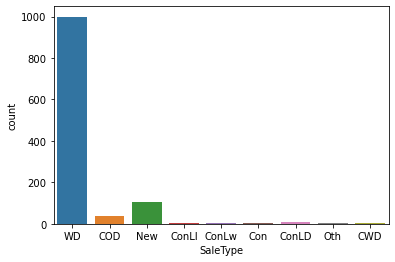
ConLw 4

Oth 3

CWD 3

Con 2

Name: SaleType, dtype: int64

****

Countplot for SaleCondition

Normal 945

Partial 108

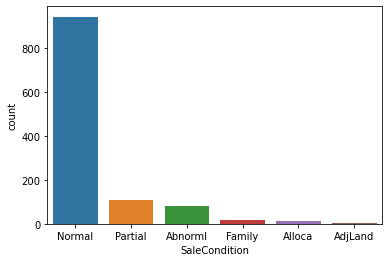
Abnorml 81

Family 18

Alloca 12

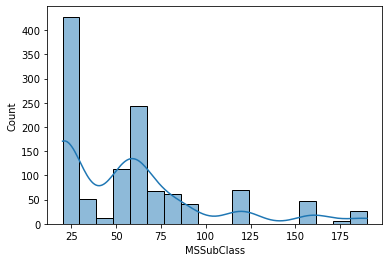
AdjLand 4

Name: SaleCondition, dtype: int64

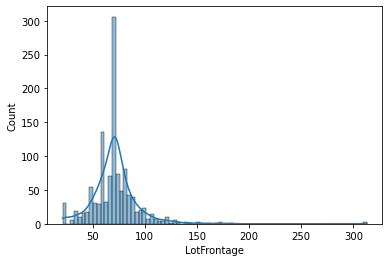
****

**Histogram for Continuous type of columns**

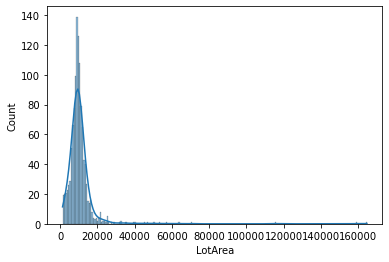
Histplot for MSSubClass

****

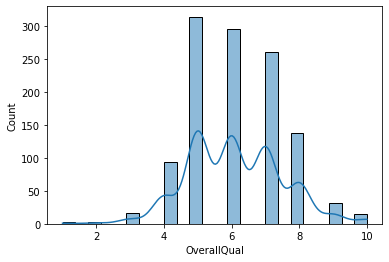
Histplot for LotFrontage

****

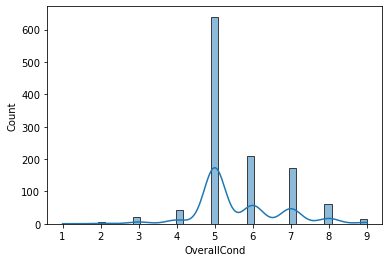
Histplot for LotArea

****

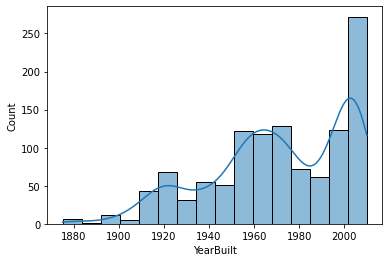
Histplot for OverallQual

****

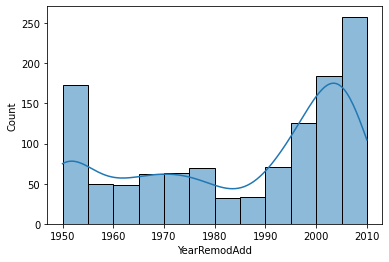
Histplot for OverallCond

****

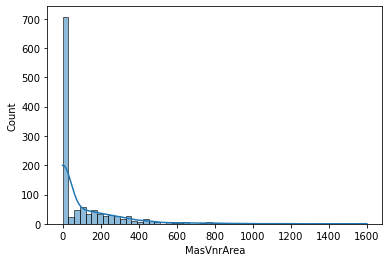
Histplot for YearBuilt

****

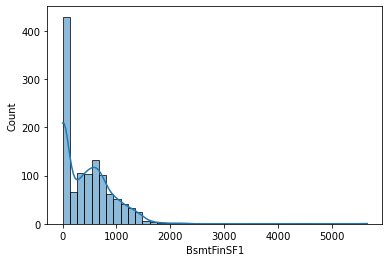
Histplot for YearRemodAdd

****

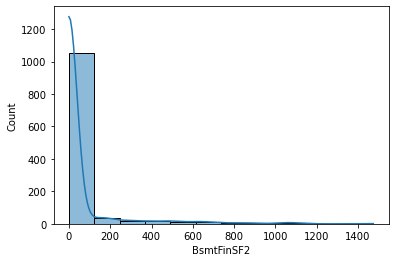
Histplot for MasVnrArea

****

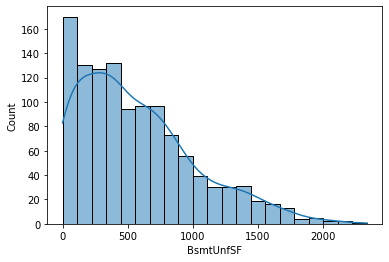
Histplot for BsmtFinSF1

****

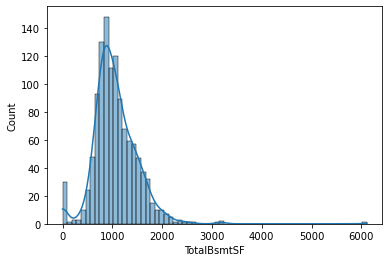
Histplot for BsmtFinSF2

****

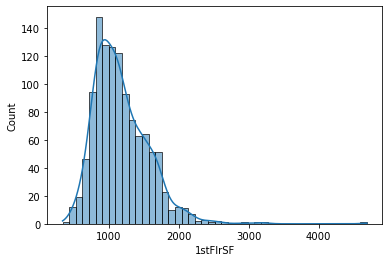
Histplot for BsmtUnfSF

****

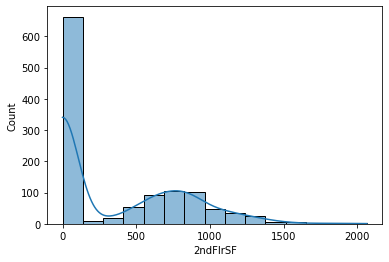
Histplot for TotalBsmtSF

****

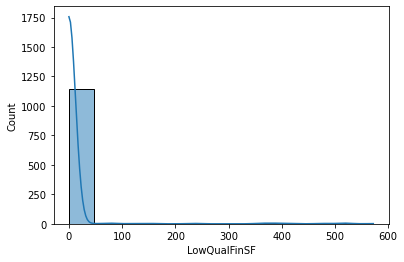
Histplot for 1stFlrSF

****

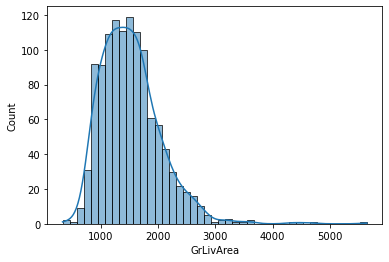
Histplot for 2ndFlrSF

****

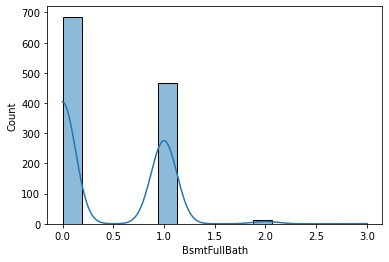
Histplot for LowQualFinSF

****

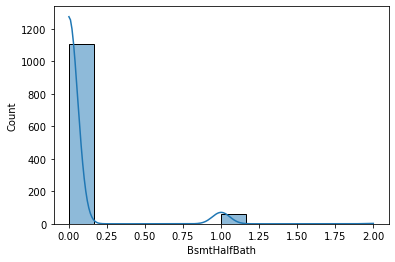
Histplot for GrLivArea

****

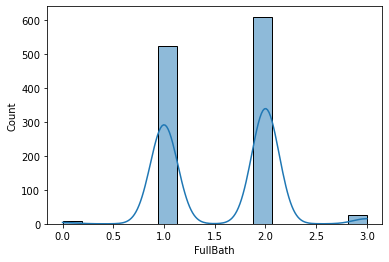
Histplot for BsmtFullBath

****

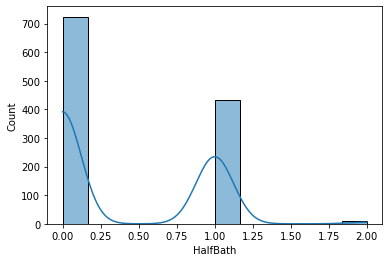
Histplot for BsmtHalfBath

****

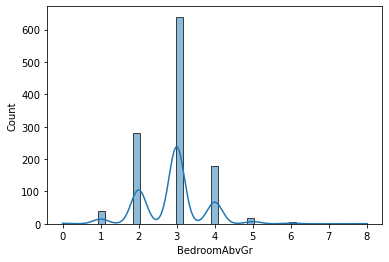
Histplot for FullBath

****

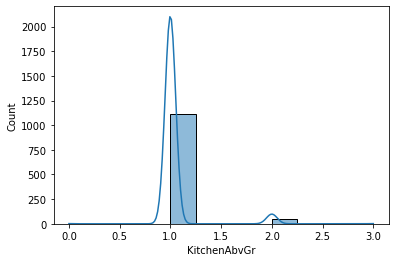
Histplot for HalfBath

****

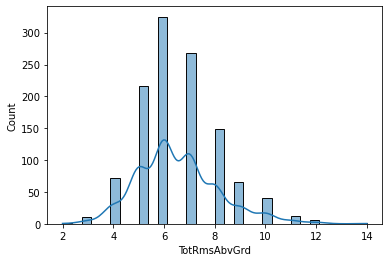
Histplot for BedroomAbvGr

****

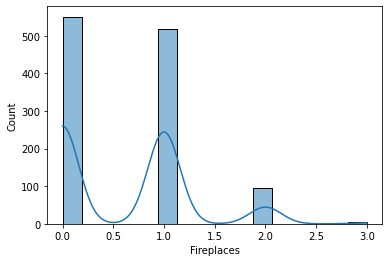
Histplot for KitchenAbvGr

****

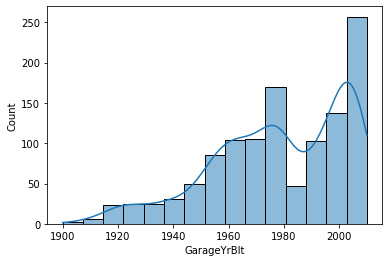
Histplot for TotRmsAbvGrd

****

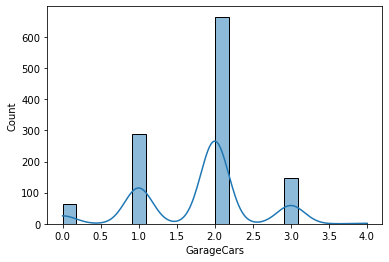
Histplot for Fireplaces

****

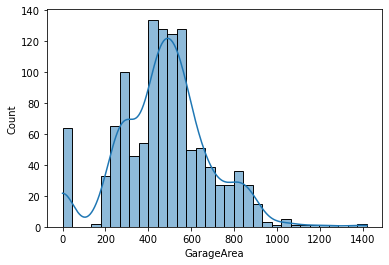
Histplot for GarageYrBlt

****

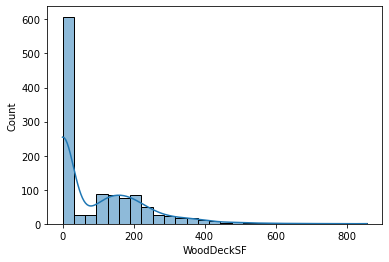
Histplot for GarageCars

****

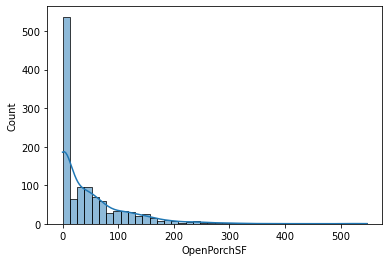
Histplot for GarageArea

****

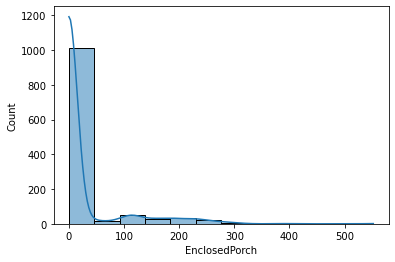
Histplot for WoodDeckSF

****

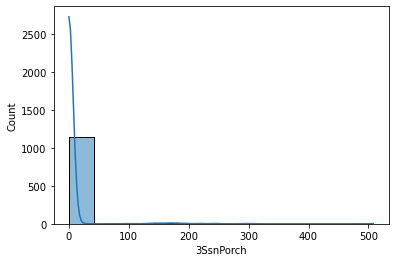
Histplot for OpenPorchSF

****

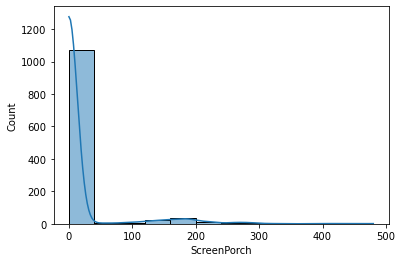
Histplot for EnclosedPorch

****

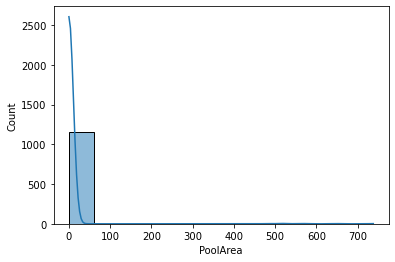
Histplot for 3SsnPorch

****

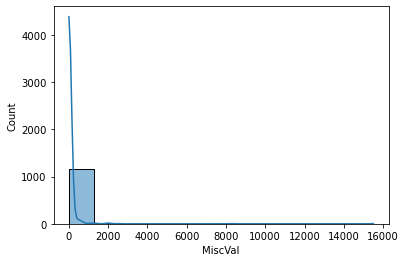
Histplot for ScreenPorch

****

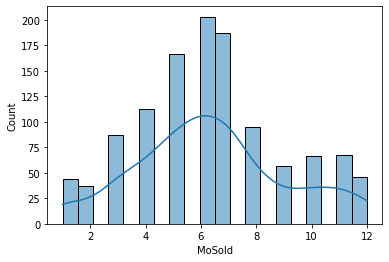
Histplot for PoolArea

****

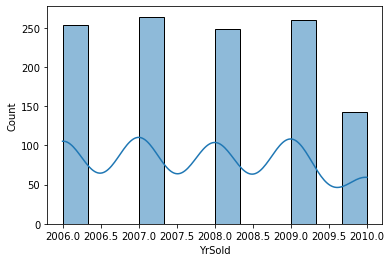
Histplot for MiscVal

****

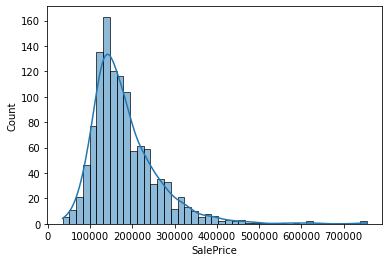
Histplot for MoSold

****

Histplot for YrSold

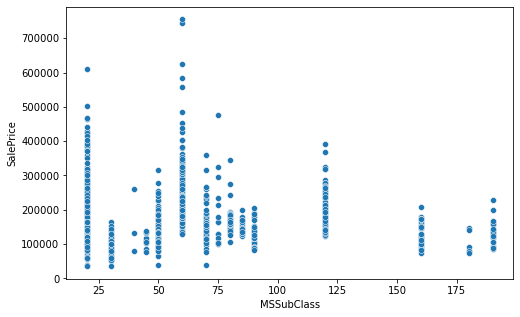
****

Histplot for SalePrice

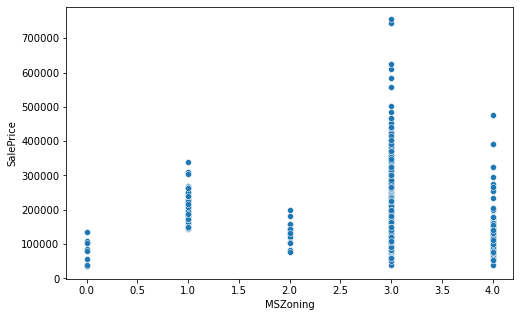
****

**Multivariate Analysis**

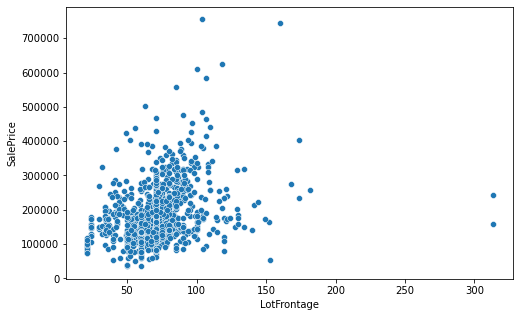
Scatter Plot for MSSubClass

****

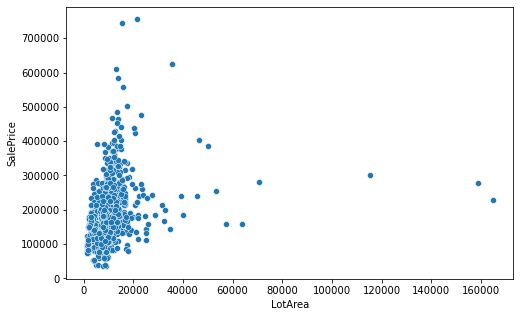
Scatter Plot for MSZoning

****

Scatter Plot for LotFrontage

****

Scatter Plot for LotArea

****

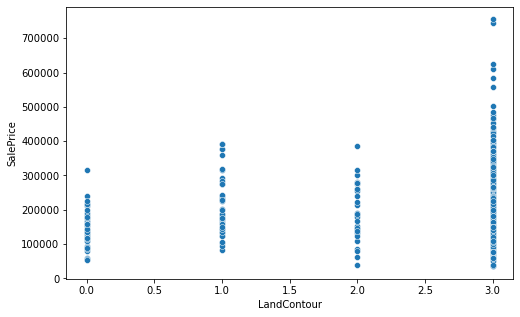
Scatter Plot for Street

****

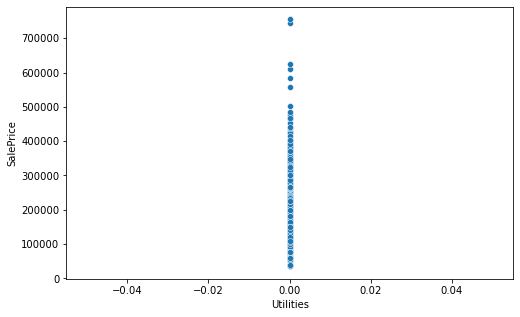
Scatter Plot for LotShape

****

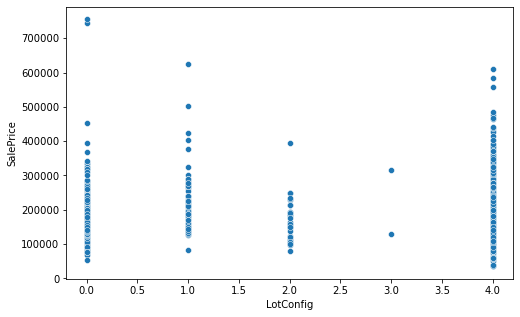
Scatter Plot for LandContour

****

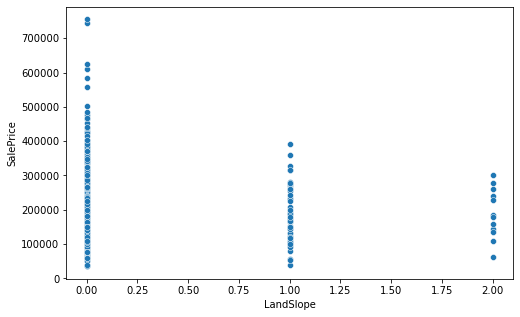
Scatter Plot for Utilities

****

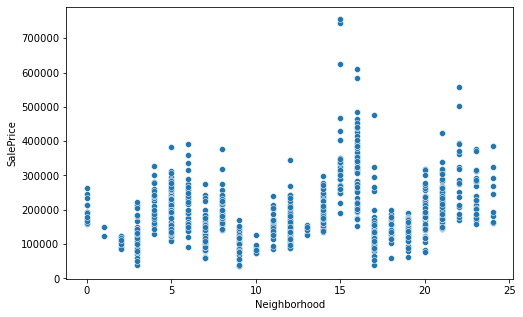
Scatter Plot for LotConfig

****

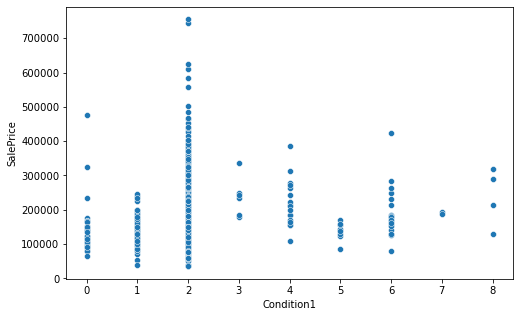
Scatter Plot for LandSlope

****

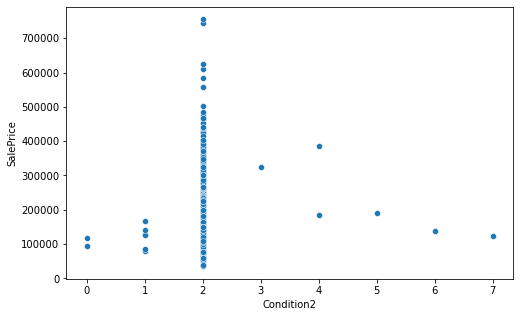
Scatter Plot for Neighborhood

****

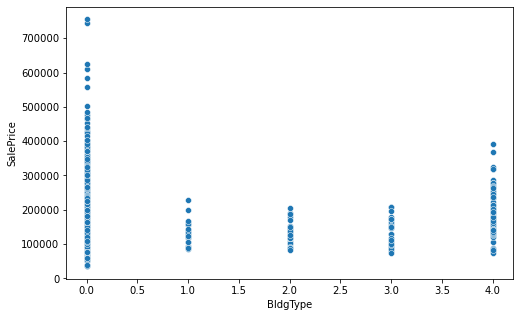
Scatter Plot for Condition1

****

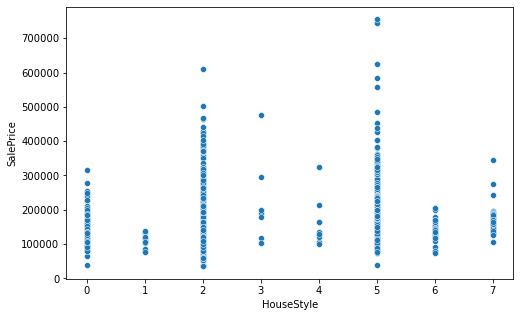
Scatter Plot for Condition2

****

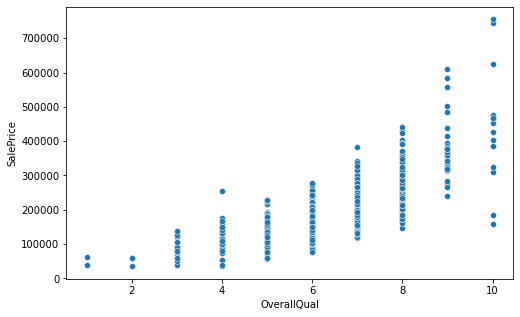
Scatter Plot for BldgType

****

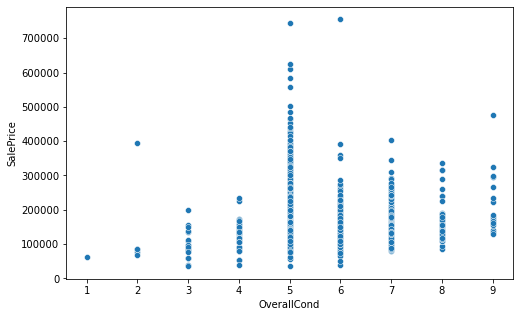
Scatter Plot for HouseStyle

****

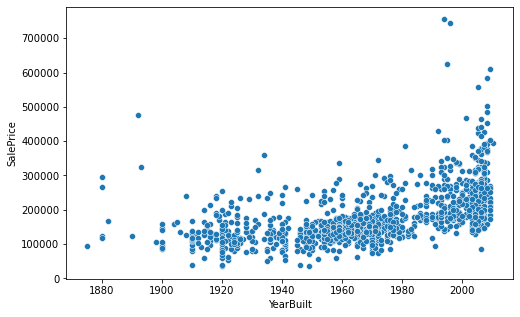
Scatter Plot for OverallQual

****

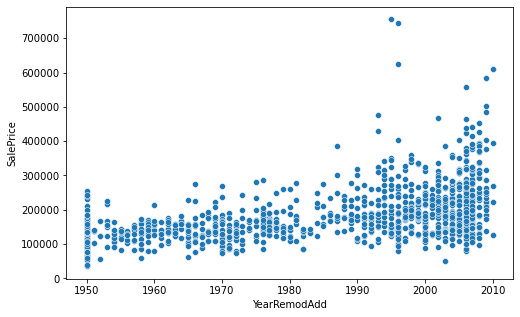
Scatter Plot for OverallCond

****

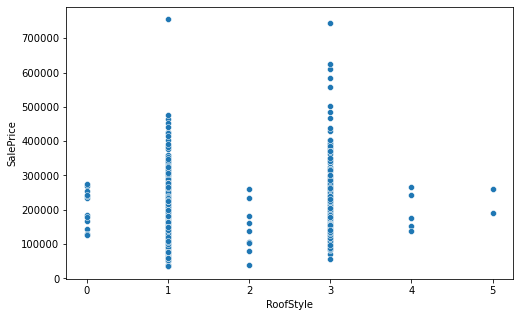
Scatter Plot for YearBuilt

****

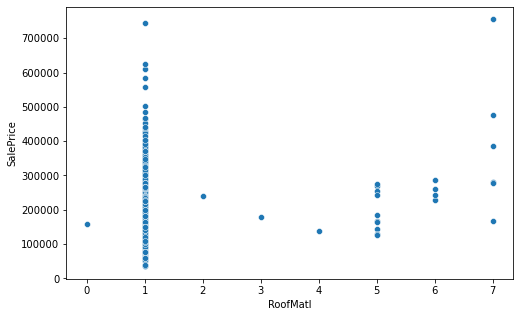
Scatter Plot for YearRemodAdd

****

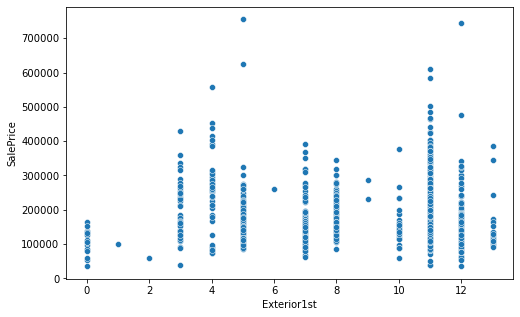
Scatter Plot for RoofStyle

****

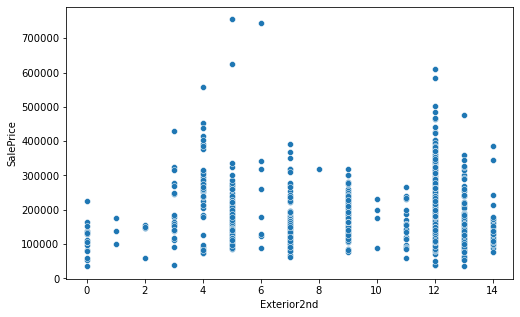
Scatter Plot for RoofMatl

****

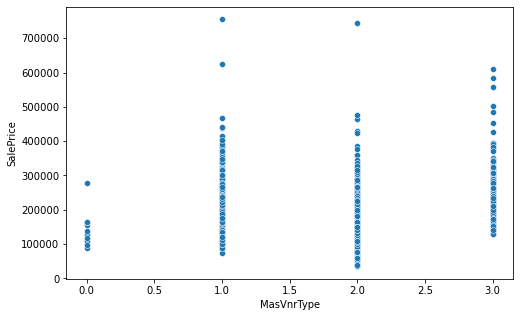
Scatter Plot for Exterior1st

****

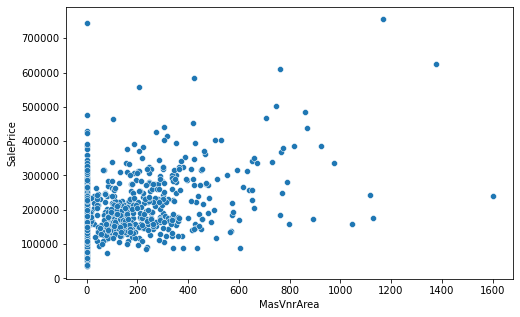
Scatter Plot for Exterior2nd

****

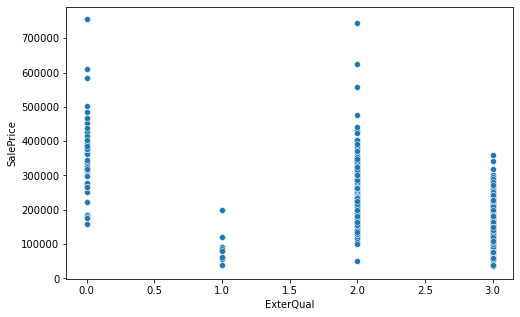
Scatter Plot for MasVnrType

****

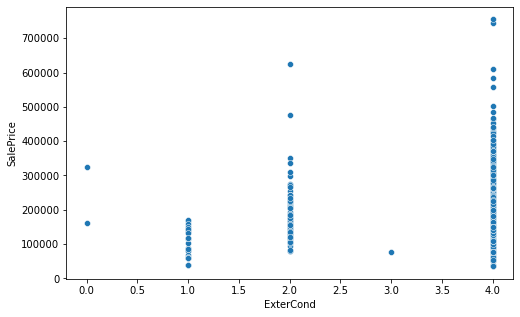
Scatter Plot for MasVnrArea

****

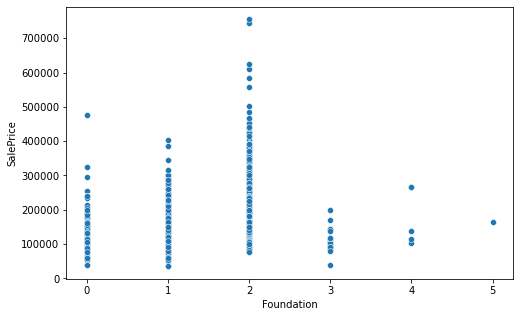
Scatter Plot for ExterQual

****

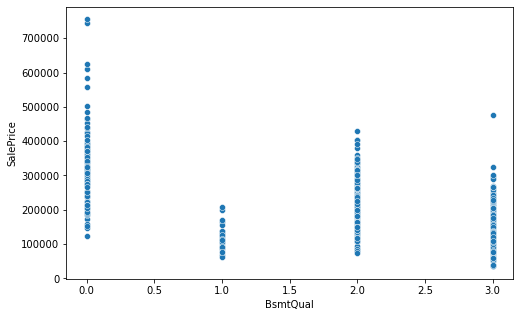
Scatter Plot for ExterCond

****

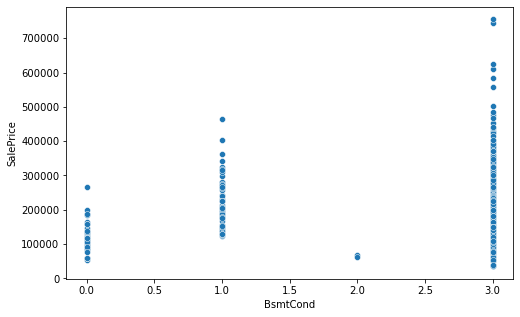
Scatter Plot for Foundation

****

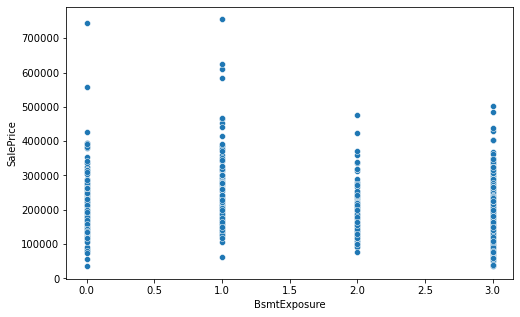
Scatter Plot for BsmtQual

****

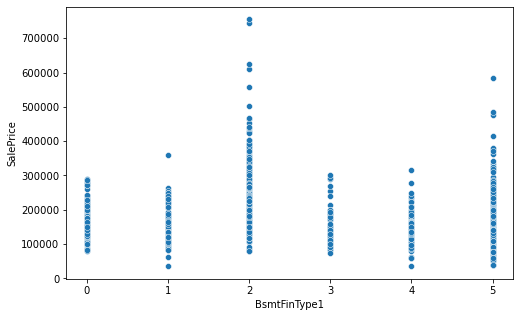
Scatter Plot for BsmtCond

****

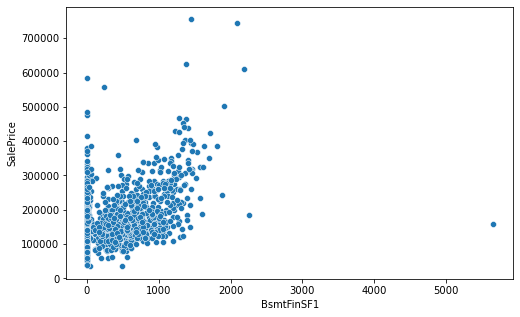
Scatter Plot for BsmtExposure

****

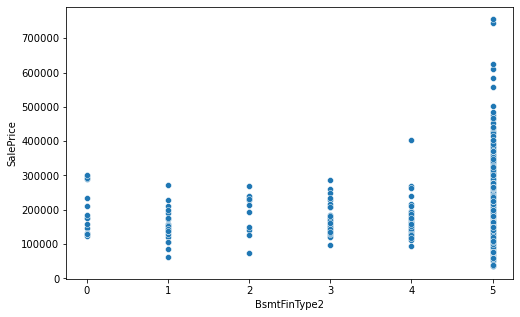
Scatter Plot for BsmtFinType1

****

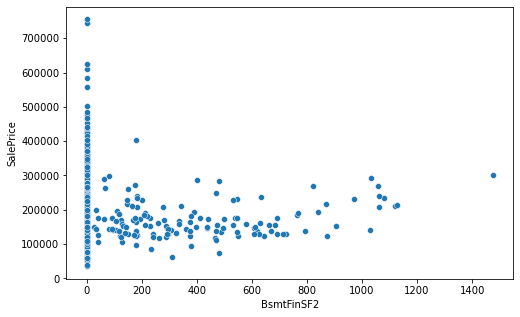
Scatter Plot for BsmtFinSF1

****

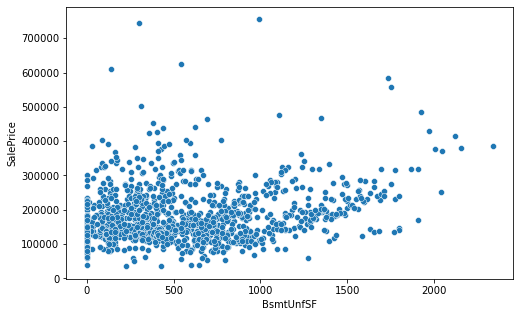
Scatter Plot for BsmtFinType2

****

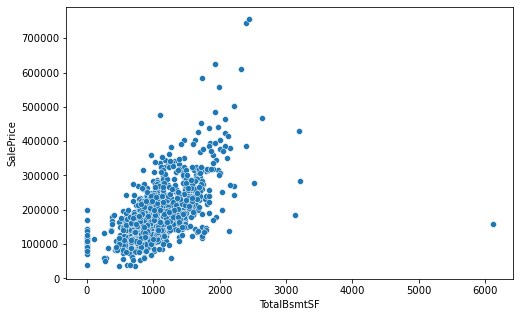
Scatter Plot for BsmtFinSF2

****

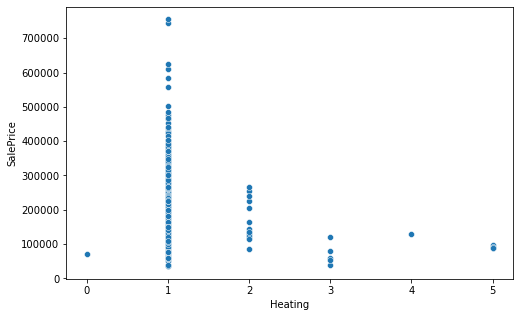
Scatter Plot for BsmtUnfSF

****

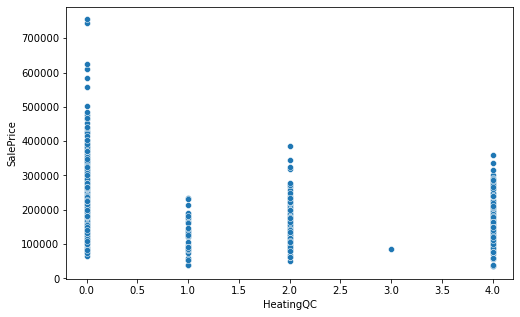
Scatter Plot for TotalBsmtSF

****

Scatter Plot for Heating

****

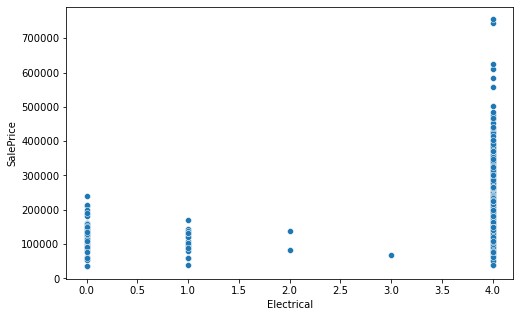
Scatter Plot for HeatingQC

****

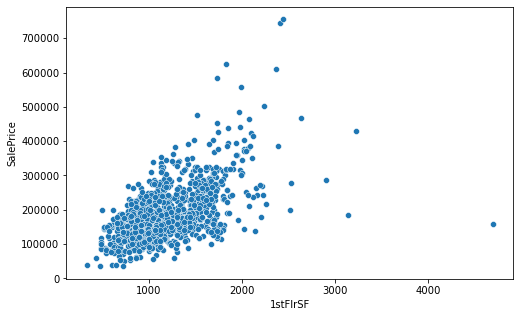
Scatter Plot for CentralAir

****

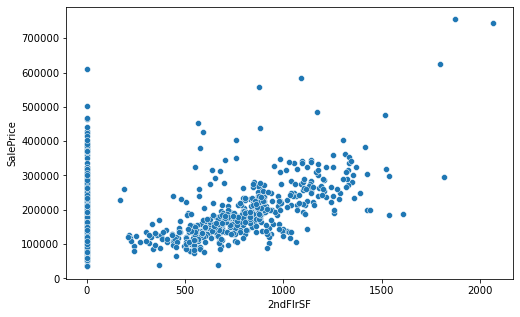
Scatter Plot for Electrical

****

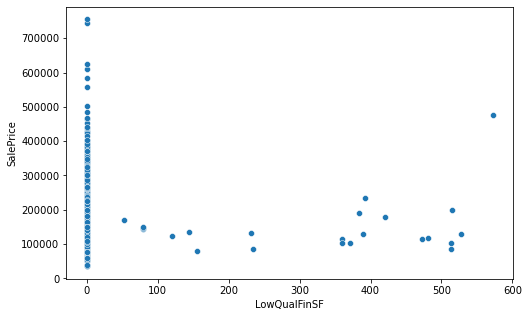
Scatter Plot for 1stFlrSF

****

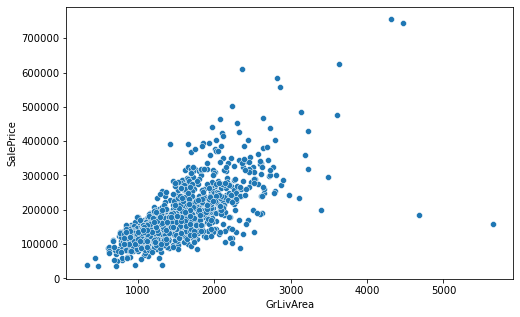
Scatter Plot for 2ndFlrSF

****

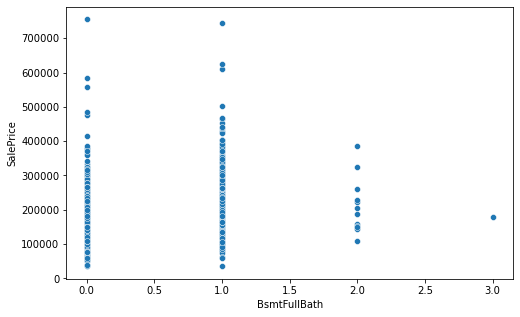
Scatter Plot for LowQualFinSF

****

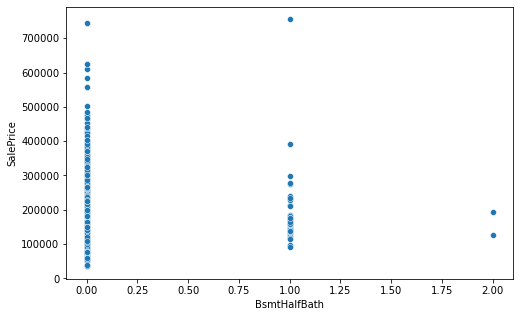
Scatter Plot for GrLivArea

****

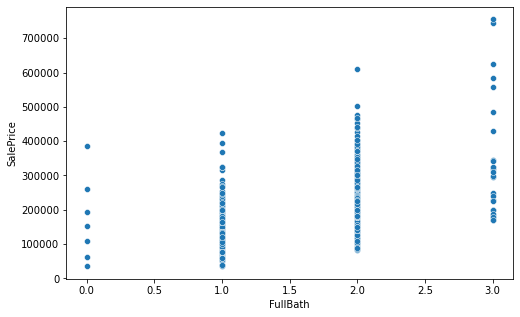
Scatter Plot for BsmtFullBath

****

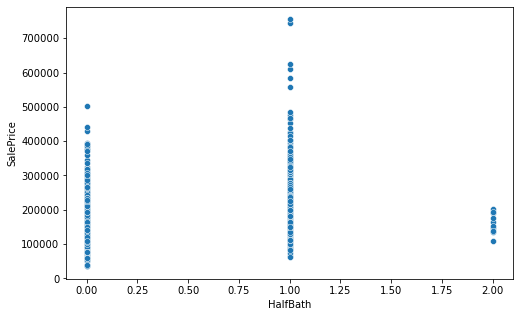
Scatter Plot for BsmtHalfBath

****

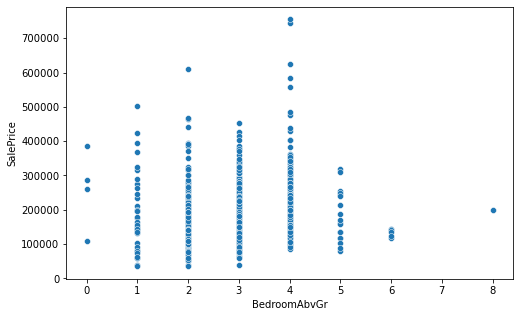
Scatter Plot for FullBath

****

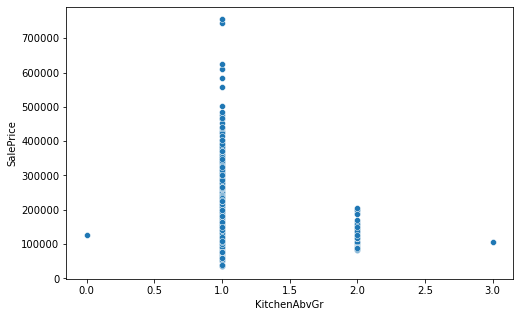
Scatter Plot for HalfBath

****

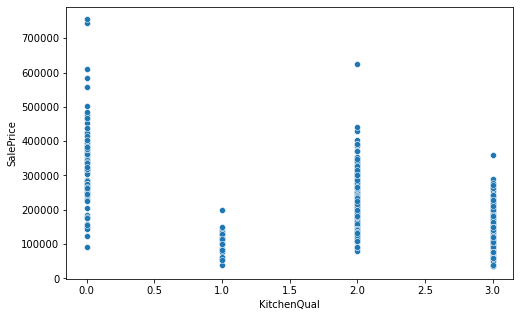
Scatter Plot for BedroomAbvGr

****

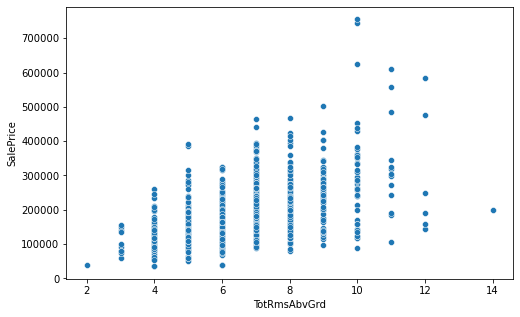
Scatter Plot for KitchenAbvGr

****

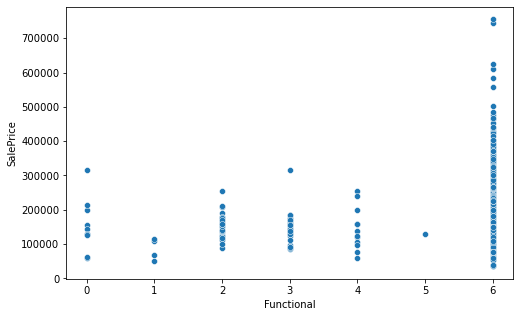
Scatter Plot for KitchenQual

****

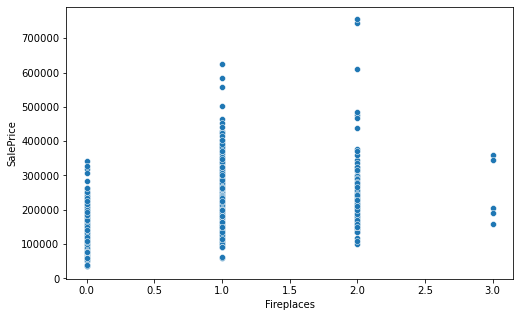
Scatter Plot for TotRmsAbvGrd

****

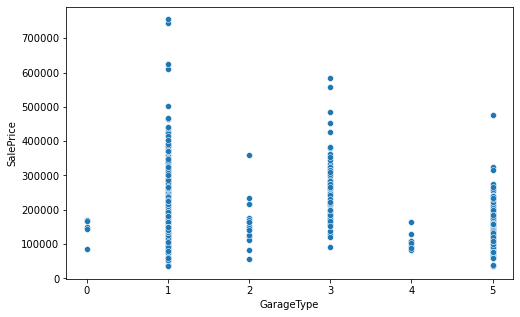
Scatter Plot for Functional

****

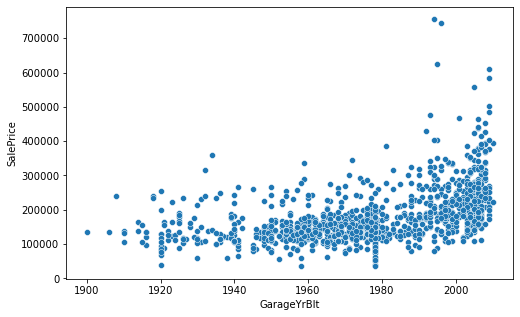
Scatter Plot for Fireplaces

****

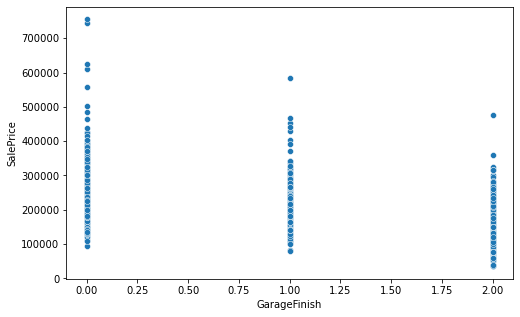
Scatter Plot for GarageType

****

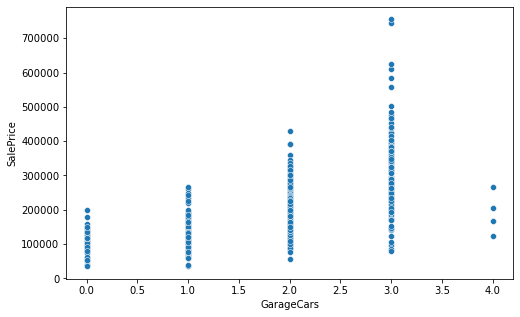
Scatter Plot for GarageYrBlt

****

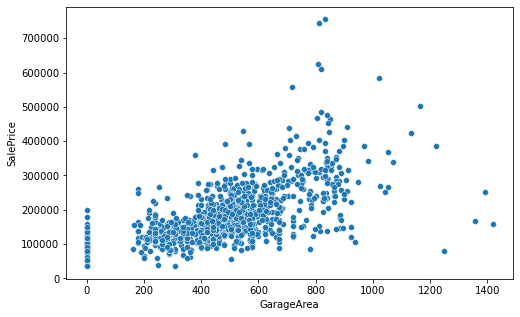
Scatter Plot for GarageFinish

****

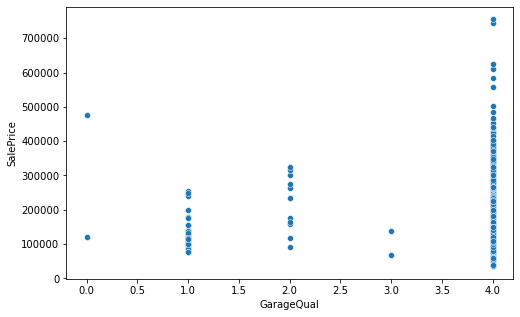
Scatter Plot for GarageCars

****

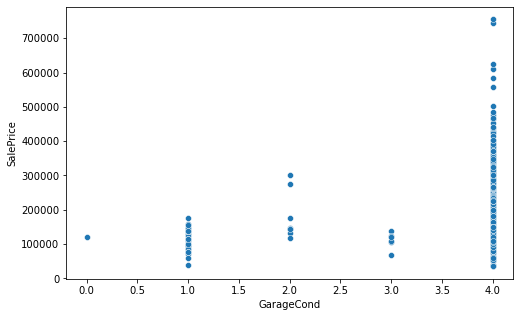
Scatter Plot for GarageArea

****

Scatter Plot for GarageQual

****

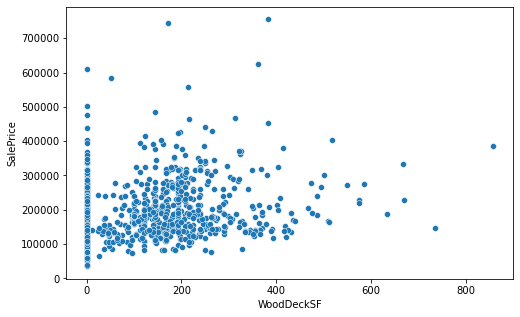
Scatter Plot for GarageCond

****

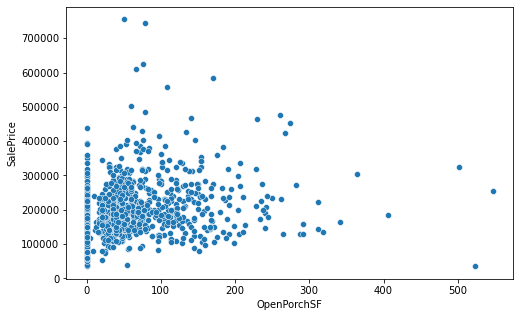
Scatter Plot for PavedDrive

****

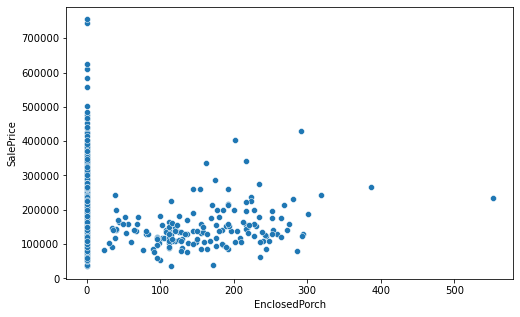
Scatter Plot for WoodDeckSF

****

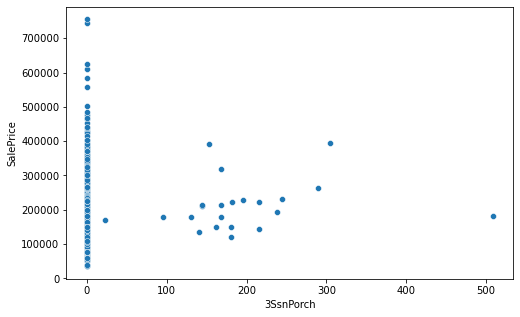
Scatter Plot for OpenPorchSF

****

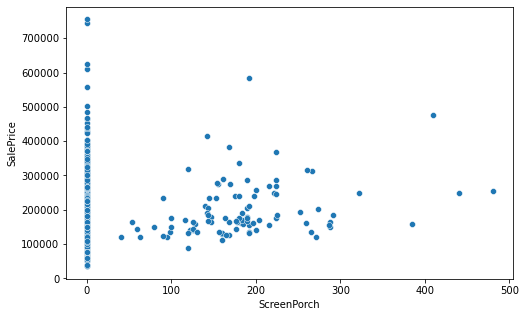
Scatter Plot for EnclosedPorch

****

Scatter Plot for 3SsnPorch

****

Scatter Plot for ScreenPorch

****

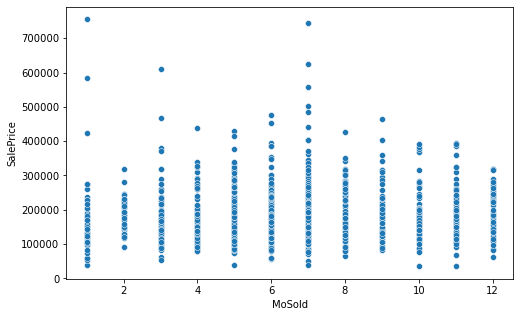
Scatter Plot for PoolArea

****

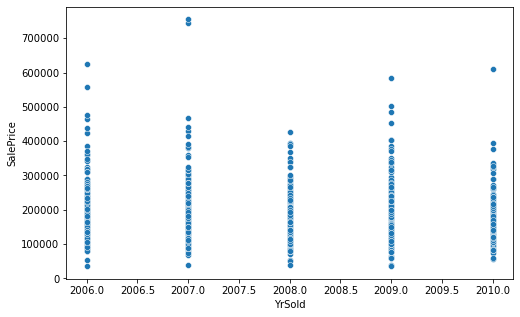
Scatter Plot for MiscVal

****

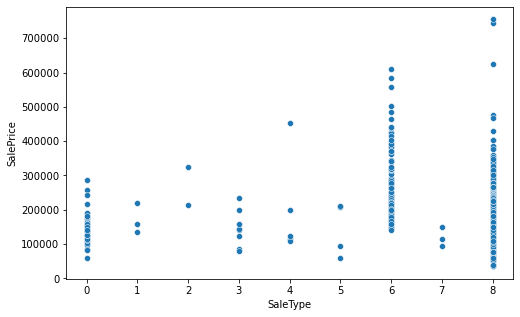
Scatter Plot for MoSold

****

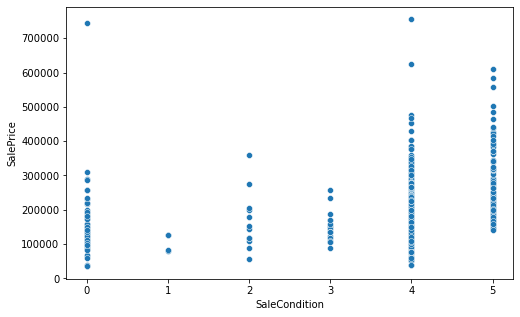
Scatter Plot for YrSold

****

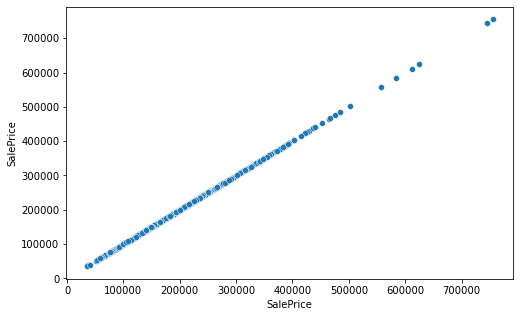
Scatter Plot for SaleType

****

Scatter Plot for SaleCondition

****

Scatter Plot for SalePrice

****

**Observations drawn are from above Data Visualization:**

1. **As LotFrontage area has increased the Sale price is also increased**
2. **For LotArea 0 to 20000 the price might have increased because of other important factors**
3. **Year of built is older, then Sale price is lower and for recent built price has increased**
4. **If year of Remodification is recent one then the price is on higher side**
5. **Masonry veneer area increases the Sale price has increased**
6. **Type 1 finished increases the Sale price has also increased**
7. **As Total Basement Area Increases the Sales price has increased**
8. **As first Floor Area increased the Sale price have increased**
9. **As second Floor Area increased the Sale price have increased**
10. **Increase in trend for ground living area is seen**
11. **As the garage-built year is recent the sale price is higher**
12. **As Garage area increased the sale price has also increased**

**Important Factors Affecting the sale price**

The following factors are more important and positively correlated with the Sale Price (The pearson correlation coefficient is shown against every factor which is important the range is varying if the coefficient is between 0.8 to 0.6 then it is having strong relationship with target variable, and then if it is between 0.6 to 0.4 moderate correlation exits, and less than that up to 0.1 very low correlation is said to be exists.

1. OverallQual 0.789185
2. GrLivArea 0.707300
3. GarageCars 0.628329
4. GarageArea 0.619000
5. TotalBsmtSF 0.595042
6. 1stFlrSF 0.587642
7. FullBath 0.554988
8. TotRmsAbvGrd 0.528363
9. YearBuilt 0.514408
10. YearRemodAdd 0.507831
11. MasVnrArea 0.463626
12. Fireplaces 0.459611
13. GarageYrBlt 0.458007
14. Foundation 0.374169
15. BsmtFinSF1 0.362874
16. OpenPorchSF 0.339500
17. 2ndFlrSF 0.330386
18. LotFrontage 0.323779
19. WoodDeckSF 0.315444
20. HalfBath 0.295592
21. GarageType -0.299470
22. HeatingQC -0.406604
23. GarageFinish -0.537121
24. KitchenQual -0.592468
25. ExterQual -0.624820
26. BsmtQual -0.626850

**Dropping the columns:**

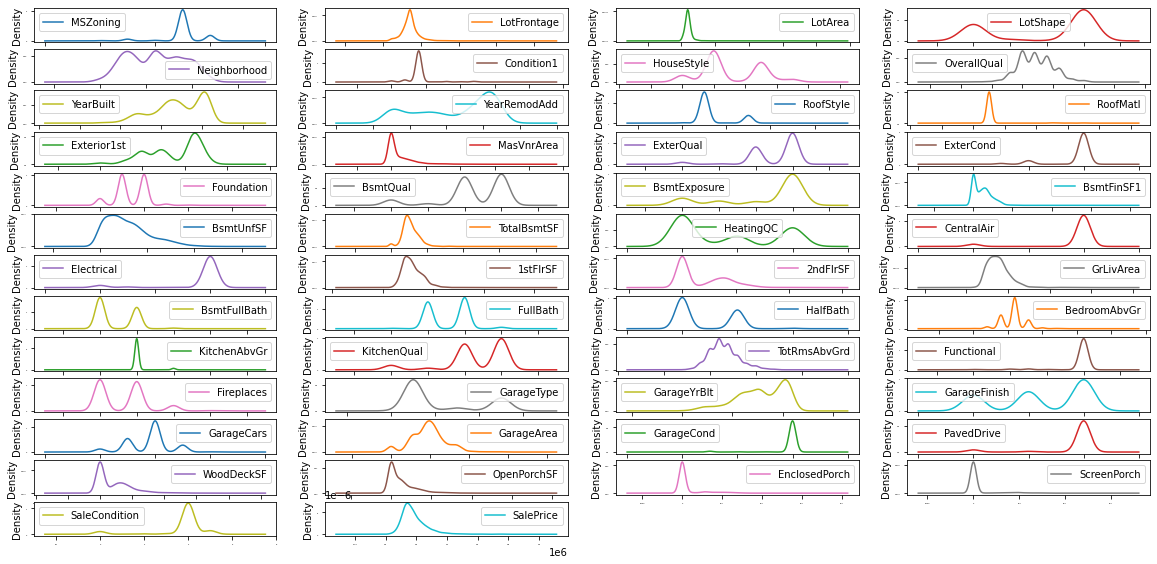
From train database the following columns are dropped because of their very low correlation with the target variable and very high Null values present in the database

[,'Id','PoolQC', 'MiscFeature', 'Alley', 'Fence', 'FireplaceQu', 'MiscVal', 'PoolArea', 'Condition2', 'Utilities', 'Exterior2nd', 'GarageQual', 'MoSold', '3SsnPorch', 'BsmtCond', 'Street', 'Condition2', 'LandContour', 'BsmtFinType2', 'LandSlope', 'MasVnrType', 'BsmtFinSF2', 'BsmtHalfBath', 'MiscVal', 'LowQualFinSF', 'YrSold', 'SaleType', 'LotConfig', 'MSSubClass', 'OverallCond', 'BldgType', 'BsmtFinType1', 'Heating']

**Conversion of Categorical Data**

The categorical data is converted into numerical label by Label Encoder

**Skewness of the data**

****

From plotting the Distplot we can see that there are multiple columns are having the skewness and to be treated

The Highly skewed columns are dropped and then some are treated with Power transform method

**Checking for Outliers**

The outliers are checked with zscore, it is observed that many continuous columns are having skewness and it is to detected

It is observed that 39% data in outliers and cannot be removed it is mainly because of two reasons

1. The data collection strategy is wrong
2. The data is real and will add some important value to the model

Hence, I have decided to keep the outliers as I am not having the domain knowledge of real estate. To check whether outlier is really a outlier or some important data point.

**Training the Train dataset**

As the target variable is continuous this dataset will go with regression type of supervised learning approach

The Best random state with Linear Regression is found out to be 94, following are the Algorithms are used and Cross val Scores are also evaluated.

**Results of various Algorithms**



**I have tabulated the outcomes for all the algorithms**

The Gradient Boosting Regressor is giving a good training and testing accuracy with minimum difference in it, also we can see that the Root mean squared error is also less than the others

The cross val score of for the Gradient Boosting Algorithm also supports the testing score that data is not overfitted.

**Hyperparameter Tuning**

The Hyper parameter tuning is done on Gradient boosting to see the best optimal set of the parameters listed

{'criterion': 'friedman\_mse', 'learning\_rate': 0.1, 'max\_features': 'log2', 'max\_leaf\_nodes': 10 'n\_estimators': 200}

The Testing score is observed to same as that earlier tabulation

**CONCLUSION**

* Following are the factors which are very important in descending order

1. OverallQual 0.789185
2. GrLivArea 0.707300
3. GarageCars 0.628329
4. GarageArea 0.619000
5. TotalBsmtSF 0.595042
6. 1stFlrSF 0.587642
7. FullBath 0.554988
8. TotRmsAbvGrd 0.528363
9. YearBuilt 0.514408
10. YearRemodAdd 0.507831
11. MasVnrArea 0.463626
12. Fireplaces 0.459611
13. GarageYrBlt 0.458007
14. Foundation 0.374169
15. BsmtFinSF1 0.362874
16. HeatingQC -0.406604
17. GarageFinish -0.537121
18. KitchenQual -0.592468
19. ExterQual -0.624820
20. BsmtQual -0.626850

Note: - Negative sign represents the variable is inversely proportional to the target.

Now, Test data has done the preprocessing same as the Train data and tested with saved model.