Real Estate Price Prediction - using Advanced Linear Regression Techniques

Overview

There are several factors that influence the price a buyer is willing to pay for a house. Some are apparent and obvious and some are not. Nevertheless, a rational approach facilitated by machine learning can be very useful in predicting the house price. A large data set with 79 different features (like living area, number of rooms, location etc) along with their prices are provided for residential homes in Ames, Iowa. The challenge is to learn a relationship between the important features and the price and use it to predict the prices of a new set of houses.

Here's a brief version of what you'll find in the data description file.

- SalePrice - the property's sale price in dollars. This is the target variable that you're trying to predict.
- MSSubClass: The building class
- MSZoning: The general zoning classification
- LotFrontage: Linear feet of street connected to property
- LotArea: Lot size in square feet
- Street: Type of road access
- Alley: Type of alley access
- LotShape: General shape of property
- LandContour: Flatness of the property
- Utilities: Type of utilities available
- LotConfig: Lot configuration
- LandSlope: Slope of property
- Neighborhood: Physical locations within Ames city limits
- Condition1: Proximity to main road or railroad
- Condition2: Proximity to main road or railroad (if a second is present)
- BldgType: Type of dwelling
- HouseStyle: Style of dwelling
- OverallQual: Overall material and finish quality
- OverallCond: Overall condition rating
- YearBuilt: Original construction date
- YearRemodAdd: Remodel date
- RoofStyle: Type of roof
- RoofMatl: Roof material

- Exterior1st: Exterior covering on house
- Exterior2nd: Exterior covering on house (if more than one material)
- MasVnrType: Masonry veneer type
- MasVnrArea: Masonry veneer area in square feet
- ExterQual: Exterior material quality
- ExterCond: Present condition of the material on the exterior
- Foundation: Type of foundation
- BsmtQual: Height of the basement
- BsmtCond: General condition of the basement
- BsmtExposure: Walkout or garden level basement walls
- BsmtFinType1: Quality of basement finished area
- BsmtFinSF1: Type 1 finished square feet
- BsmtFinType2: Quality of second finished area (if present)
- BsmtFinSF2: Type 2 finished square feet
- BsmtUnfSF: Unfinished square feet of basement area
- TotalBsmtSF: Total square feet of basement area
- Heating: Type of heating
- HeatingQC: Heating quality and condition
- CentralAir: Central air conditioning
- Electrical: Electrical system
- 1stFlrSF: First Floor square feet
- 2ndFlrSF: Second floor square feet
- LowQualFinSF: Low quality finished square feet (all floors)
- GrLivArea: Above grade (ground) living area square feet
- BsmtFullBath: Basement full bathrooms
- BsmtHalfBath: Basement half bathrooms
- FullBath: Full bathrooms above grade
- HalfBath: Half baths above grade
- Bedroom: Number of bedrooms above basement level
- Kitchen: Number of kitchens
- KitchenQual: Kitchen quality
- TotRmsAbvGrd: Total rooms above grade (does not include bathrooms)
- Functional: Home functionality rating
- Fireplaces: Number of fireplaces
- FireplaceQu: Fireplace quality
- GarageType: Garage location
- GarageYrBlt: Year garage was built

- GarageFinish: Interior finish of the garage
- GarageCars: Size of garage in car capacity
- GarageArea: Size of garage in square feet
- GarageQual: Garage quality
- GarageCond: Garage condition
- PavedDrive: Paved driveway
- WoodDeckSF: Wood deck area in square feet
- OpenPorchSF: Open porch area in square feet
- EnclosedPorch: Enclosed porch area in square feet
- 3SsnPorch: Three season porch area in square feet
- ScreenPorch: Screen porch area in square feet
- PoolArea: Pool area in square feet
- PoolQC: Pool quality
- Fence: Fence quality
- MiscFeature: Miscellaneous feature not covered in other categories
- MiscVal: Value of miscellaneous feature
- MoSold: Month Sold
- YrSold: Year Sold
- SaleType: Type of sale
- SaleCondition: Condition of sale

Content

1. Import Packages

In [1046]:

```
# Importing packages
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

2. Load data

Read the House Price Dataset using pandas.read_csv function into an object(data)

In [1047]:
```
data = pd.read_csv("D:\WorkPlace_R\DataSets\House Price Dataset.csv")
```

3. Data Preparation

The process of data preparation entails cleansing, structuring and integrating data to make it ready for analysis. Here we first analyze the data statistically and then split the target varibles and normalize, followed by splitting the dataframe into numerical and categorical features.

In [1048]:

```
# checking data types for variables in HousePrice dataframe
data.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 [1049]:
```
# Checking data size
data.shape
```

Out[1049]: (1460, 81)

3.1 Statistical Summary

Here we take a look at the summary of each attribute. This includes the count, mean, the min and max values as well as some percentiles for numeric variables and count, unique, top, freq for categorical variables.

In [1050]:
```
# dataframe with categorical features
data.describe(include=['object'])
```

Out[1050]:

	MSZoning	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	Condition1	...	Gar
count	1460	1460	91	1460	1460	1460	1460	1460	1460	1460	...	137
unique	5	2	2	4	4	2	5	3	25	9	...	6
top	RL	Pave	Grvl	Reg	Lvl	AllPub	Inside	Gtl	NAmes	Norm	...	Attc
freq	1151	1454	50	925	1311	1459	1052	1382	225	1260	...	870

4 rows × 43 columns

In [1051]:
```
# dataframe with numerical features
data.describe(include=['int64'])
```

Out[1051]:

	Id	MSSubClass	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd	BsmtFinSF1	Bsr
count	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	146
mean	730.500000	56.897260	10516.828082	6.099315	5.575342	1971.267808	1984.865753	443.639726	46.5
std	421.610009	42.300571	9981.264932	1.382997	1.112799	30.202904	20.645407	456.098091	161
min	1.000000	20.000000	1300.000000	1.000000	1.000000	1872.000000	1950.000000	0.000000	0.00
25%	365.750000	20.000000	7553.500000	5.000000	5.000000	1954.000000	1967.000000	0.000000	0.00
50%	730.500000	50.000000	9478.500000	6.000000	5.000000	1973.000000	1994.000000	383.500000	0.00
75%	1095.250000	70.000000	11601.500000	7.000000	6.000000	2000.000000	2004.000000	712.250000	0.00
max	1460.000000	190.000000	215245.000000	10.000000	9.000000	2010.000000	2010.000000	5644.000000	147

8 rows × 35 columns

3.2 Spliting Target Variable

Here the Target Variable is separated from data and the distribution is checked.

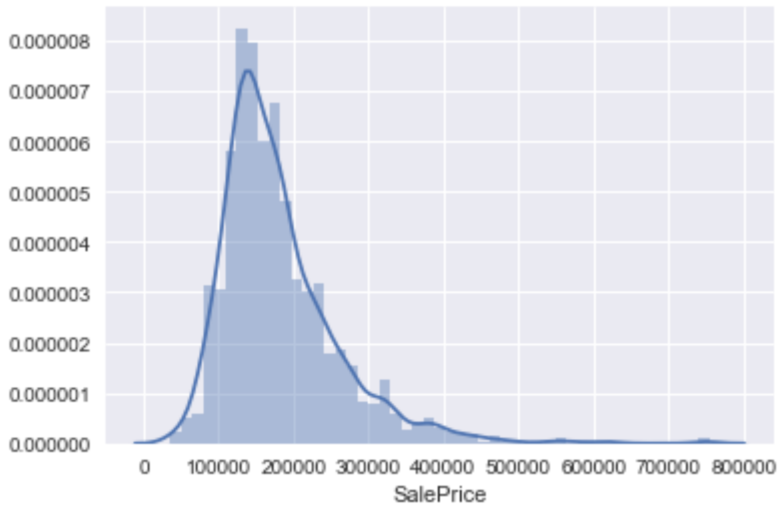
In [1052]:
```
target = data['SalePrice']
target.head()
```

Out[1052]:
```
0    208500
1    181500
2    223500
3    140000
4    250000
Name: SalePrice, dtype: int64
```

In [1053]:
```
# Visualizing the distribution of Salesprice(Dependent) variable
import seaborn as sns
sns.distplot(target,hist=True)
```

```
C:\Users\computer\Anaconda3\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The 'norme
d' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
```

Out[1053]: <matplotlib.axes._subplots.AxesSubplot at 0x15167d00198>



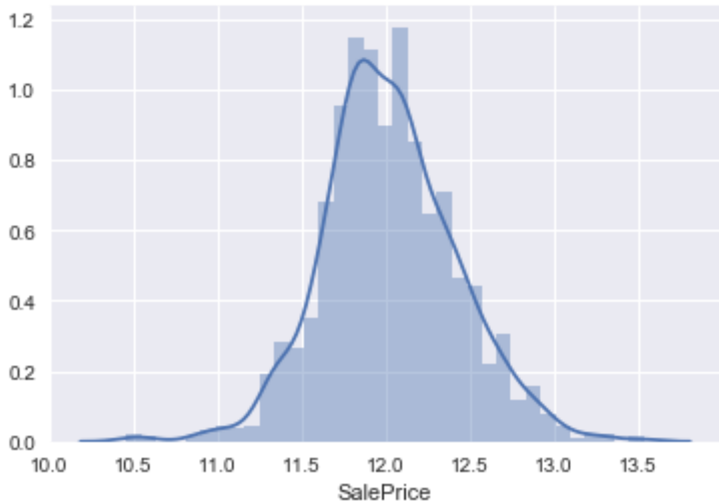
As we can see the distribution is left skewed, so in order to make it normally distributed, we need to use log transformation.

In [1054]:
```
# Log transformation
import numpy as np
target_log  = np.log(target)
```

In [1055]:
```
sns.distplot(target_log,hist=True)
```

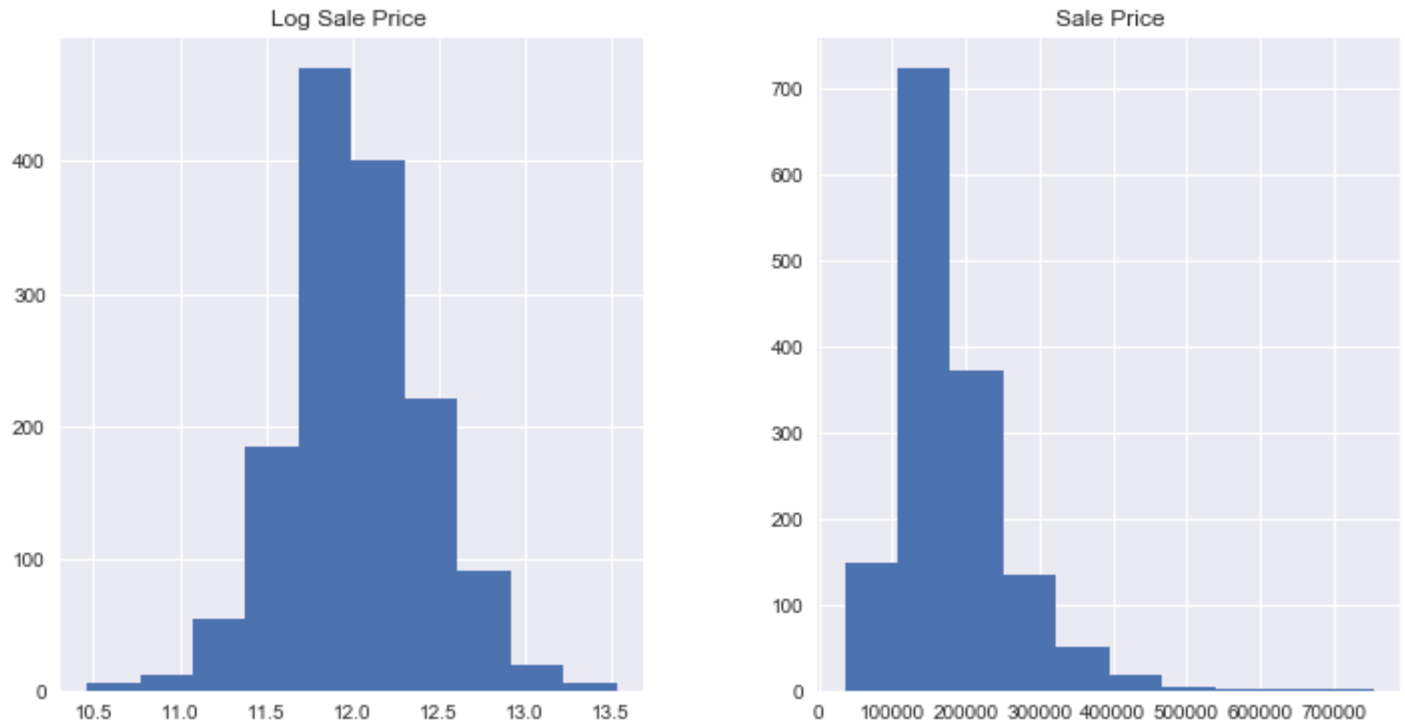
```
C:\Users\computer\Anaconda3\lib\site-packages\matplotlib\axes\_axes.py:6462: UserWarning: The 'norme
d' kwarg is deprecated, and has been replaced by the 'density' kwarg.
  warnings.warn("The 'normed' kwarg is deprecated, and has been "
```

Out[1055]: <matplotlib.axes._subplots.AxesSubplot at 0x151684745c0>



In [1056]:
```
import matplotlib
matplotlib.rcParams['figure.figsize'] = (12.0, 6.0)
prices = pd.DataFrame({"Sale Price":data["SalePrice"],"Log Sale Price ":target_log})
prices.hist()
```

Out[1056]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x00000151689947B8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x00000151689C3320>]], dtype=object)
```



After using log transformation, the Target variable is normally distributed.

In [1057]:
```
# drop target variable from dataset
raw_data = data
data = data.drop(["SalePrice"], axis=1)
data.head()
```

Out[1057]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	ScreenPorch	PoolA
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	...	0	0
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	...	0	0
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub	...	0	0
3	4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub	...	0	0
4	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub	...	0	0

5 rows × 80 columns

3.3 Feature Engineering

In [1058]:
```
#MSSubClass=The building class
data['MSSubClass'] = data['MSSubClass'].apply(str)

#Changing OverallCond into a categorical variable
data['OverallCond'] = data['OverallCond'].astype(str)

#Year and month sold are transformed into categorical features.
data['YrSold'] = data['YrSold'].astype(str)
data['MoSold'] = data['MoSold'].astype(str)
```

In [1059]:

```
# Adding total sqfootage feature
data['TotalSF'] = data['TotalBsmtSF'] + data['1stFlrSF'] + data['2ndFlrSF']
# Removing TotalBsmtSF,1stFlrSF, 2ndFlrSF and Id
data = data.drop(["TotalBsmtSF"], axis=1)
data = data.drop(["1stFlrSF"], axis=1)
data = data.drop(["2ndFlrSF"], axis=1)
data = data.drop(["Id"], axis=1)
data.head()
```

Out[1059]:

	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig	...	PoolArea	P
0	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	Inside	...	0	N
1	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	FR2	...	0	N
2	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub	Inside	...	0	N
3	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub	Corner	...	0	N
4	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub	FR2	...	0	N

5 rows × 77 columns

3.4 Split Dataframe into numeric and categorical

Split dataframe into 2 with:

- categorical features
- numerical features

In [1060]:

```
# save all categorical columns in list
categorical_columns = [col for col in data.columns.values if data[col].dtype == 'object']

# dataframe with categorical features
data_cat = data[categorical_columns]
# dataframe with numerical features
data_num = data.drop(categorical_columns, axis=1)
```

In [1061]:

```
# Using describe function in numeric dataframe
data_num.describe()
```

Out[1061]:

	LotFrontage	LotArea	OverallQual	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	BsmtFinSF2	Bsr
count	1201.000000	1460.000000	1460.000000	1460.000000	1460.000000	1452.000000	1460.000000	1460.000000	1460
mean	70.049958	10516.828082	6.099315	1971.267808	1984.865753	103.685262	443.639726	46.549315	567.
std	24.284752	9981.264932	1.382997	30.202904	20.645407	181.066207	456.098091	161.319273	441.
min	21.000000	1300.000000	1.000000	1872.000000	1950.000000	0.000000	0.000000	0.000000	0.00
25%	59.000000	7553.500000	5.000000	1954.000000	1967.000000	0.000000	0.000000	0.000000	223.
50%	69.000000	9478.500000	6.000000	1973.000000	1994.000000	0.000000	383.500000	0.000000	477.
75%	80.000000	11601.500000	7.000000	2000.000000	2004.000000	166.000000	712.250000	0.000000	808.
max	313.000000	215245.000000	10.000000	2010.000000	2010.000000	1600.000000	5644.000000	1474.000000	2336

8 rows × 30 columns

In [1062]:
```
# Printing 5 head observation in categorical dataframe
data_cat.head()
```

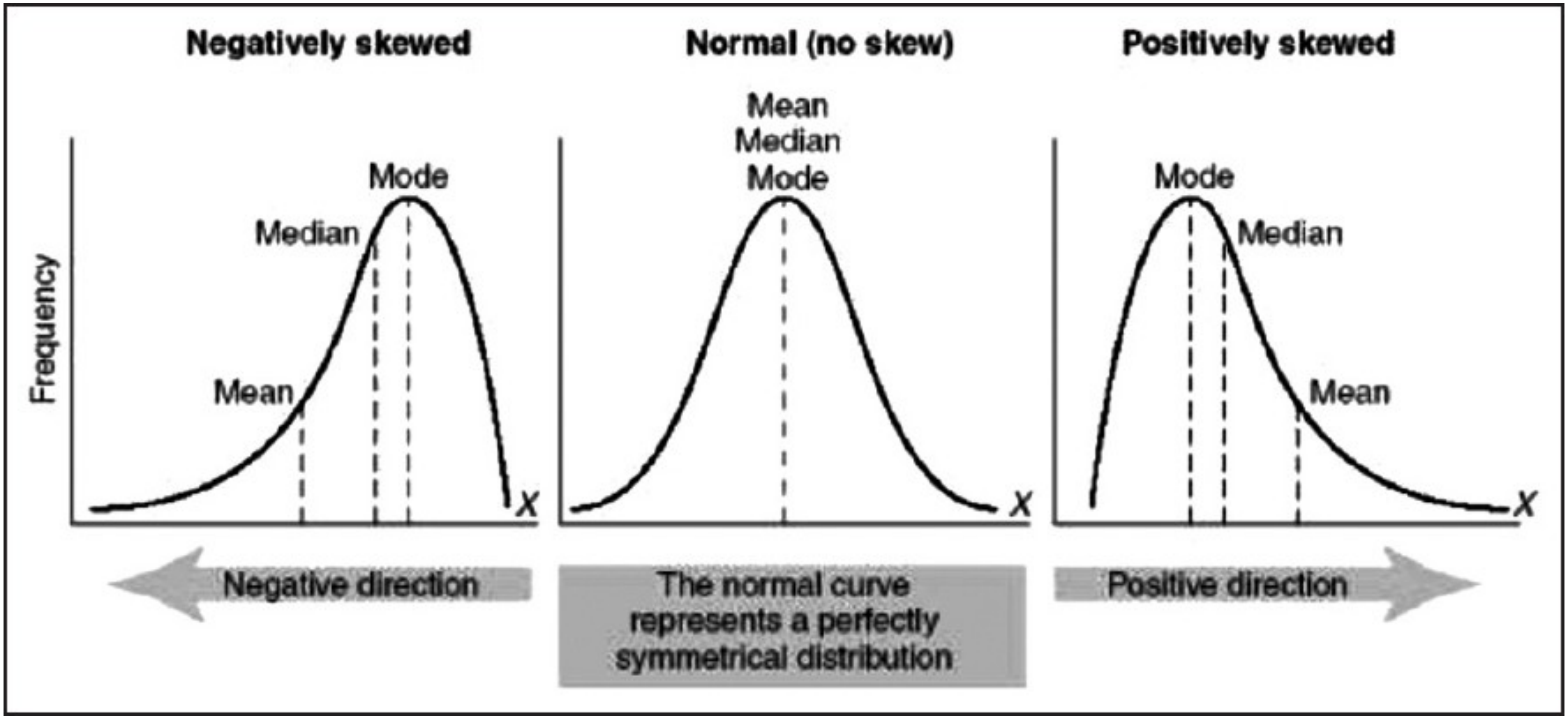
Out[1062]:

	MSSubClass	MSZoning	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	...	Garage
0	60	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	CollgCr	...	TA
1	20	RL	Pave	NaN	Reg	Lvl	AllPub	FR2	Gtl	Veenker	...	TA
2	60	RL	Pave	NaN	IR1	Lvl	AllPub	Inside	Gtl	CollgCr	...	TA
3	70	RL	Pave	NaN	IR1	Lvl	AllPub	Corner	Gtl	Crawfor	...	TA
4	60	RL	Pave	NaN	IR1	Lvl	AllPub	FR2	Gtl	NoRidge	...	TA

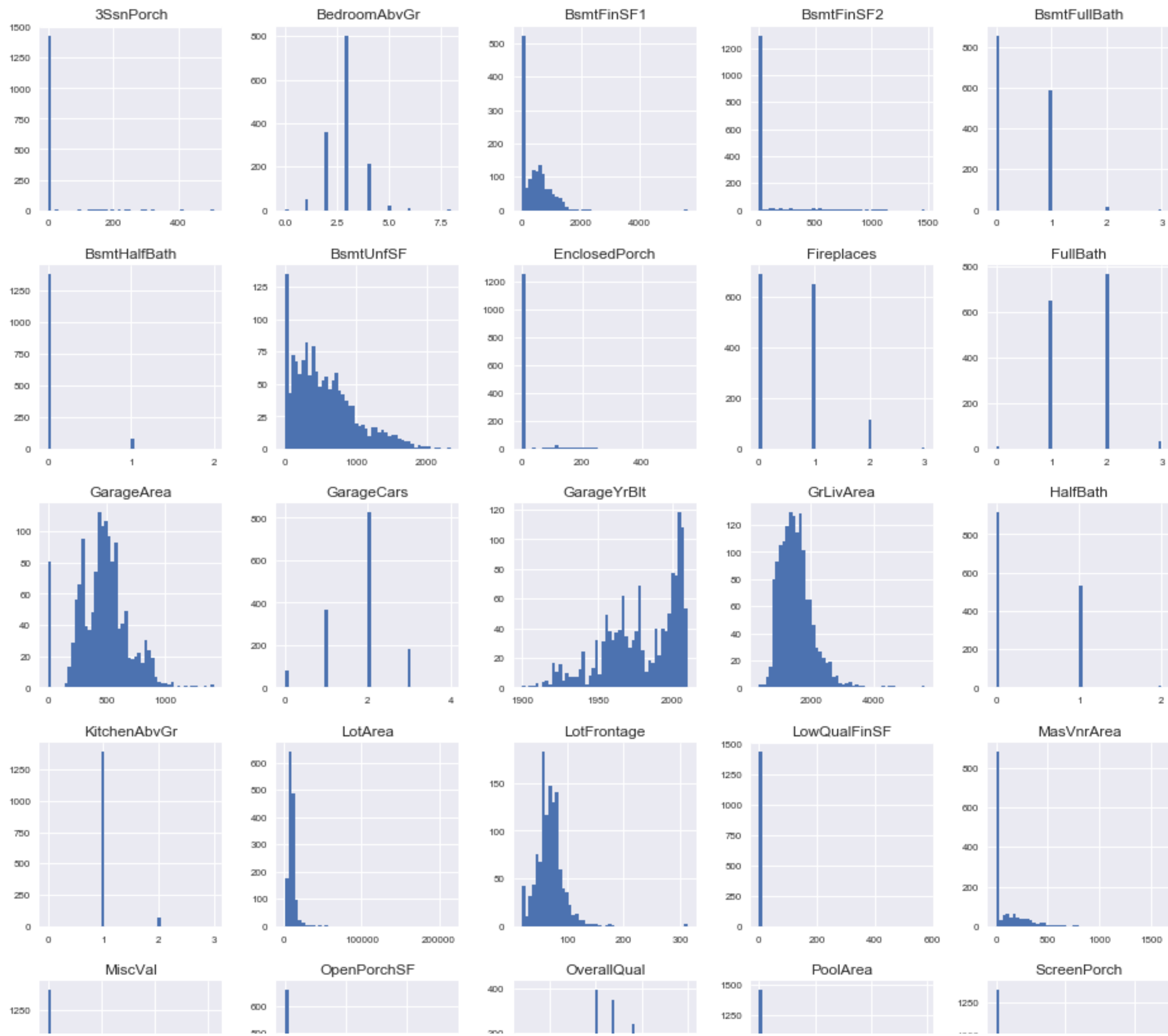
5 rows × 47 columns

3.5 Reduce Skewness for Numeric Features

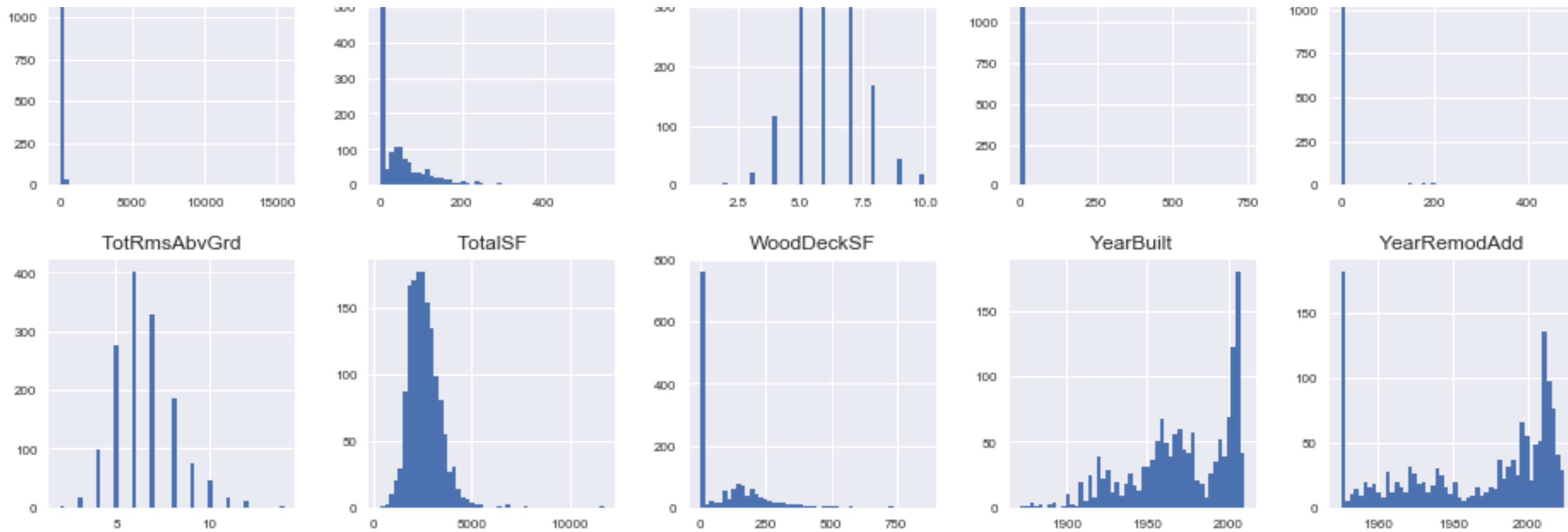
Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. Here we are interested in the variables which have skewness more than 0.75



In [1063]:
```
data_num.hist(figsize=(16, 20), bins=50, xlabelsize=8, ylabelsize=8); # ; avoid having the matplotli
b verbose informations
```



3SsnPorch
BedroomAbvGr
BsmtFinSF1
BsmtFinSF2
BsmtFullBath
BsmtHalfBath
BsmtUnfSF
EnclosedPorch
Fireplaces
FullBath
GarageArea
GarageCars
GarageYrBlt
GrLivArea
HalfBath
KitchenAbvGr
LotArea
LotFrontage
LowQualFinSF
MasVnrArea
MiscVal
OpenPorchSF
OverallQual
PoolArea
ScreenPorch



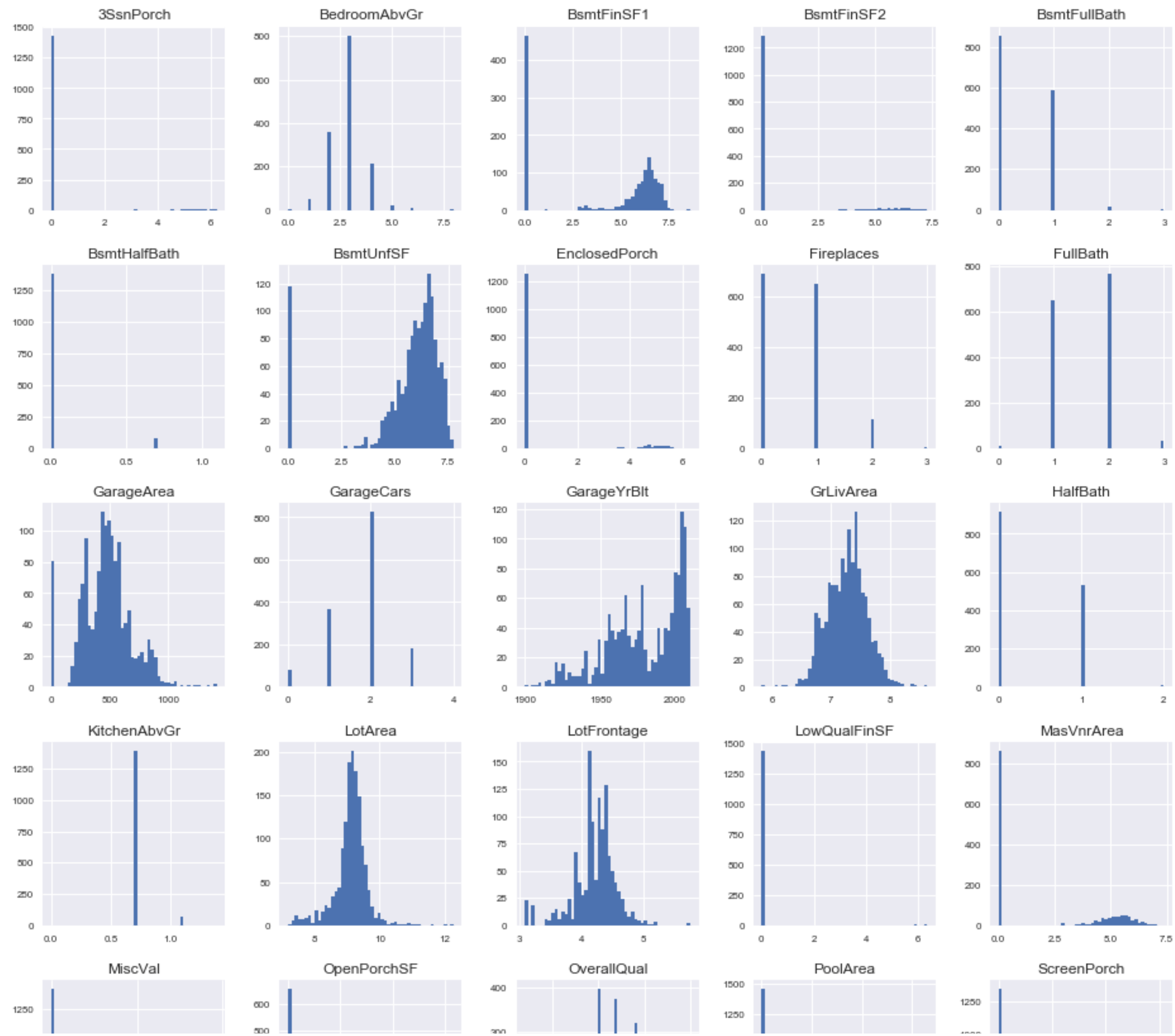
In [1064]:

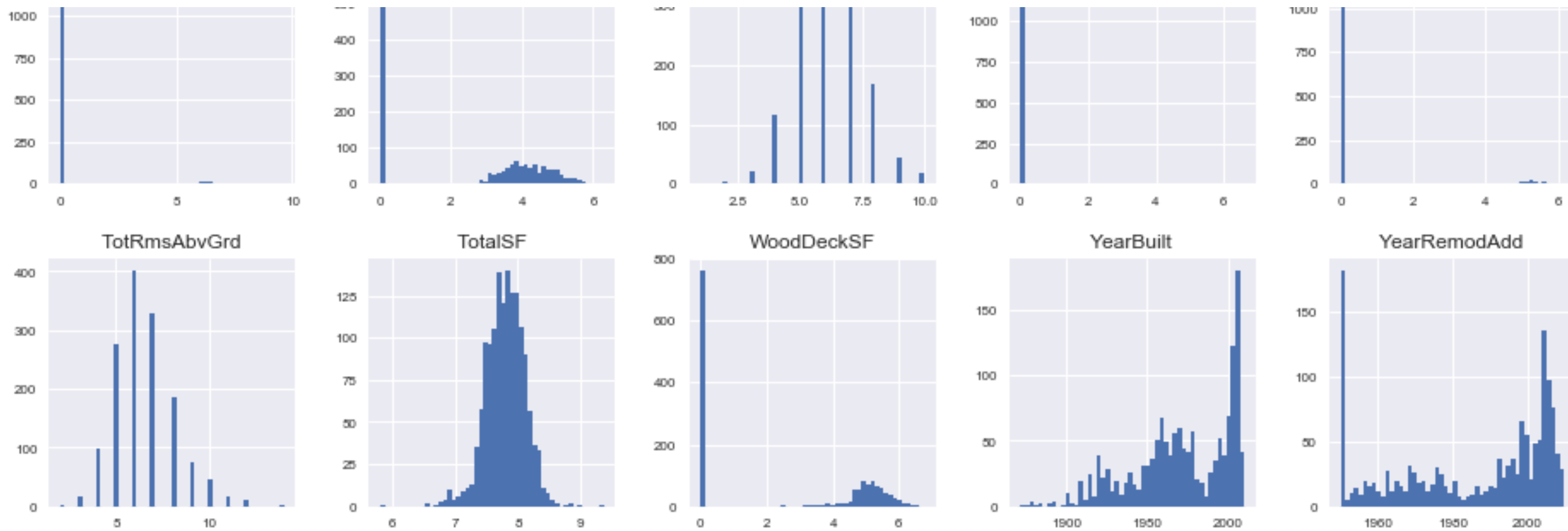
```
from scipy.stats import skew
data_num_skew = data_num.apply(lambda x: skew(x.dropna()))
data_num_skew = data_num_skew[data_num_skew > .75]

# apply log + 1 transformation for all numeric features with skewnes over .75
data_num[data_num_skew.index] = np.log1p(data_num[data_num_skew.index])
```

In [1065]:
```
# List of variables has skewness more than 0.75
data_num_skew
```

Out[1065]:
```
LotFrontage       2.160866
LotArea          12.195142
MasVnrArea        2.666326
BsmtFinSF1        1.683771
BsmtFinSF2        4.250888
BsmtUnfSF         0.919323
LowQualFinSF      9.002080
GrLivArea         1.365156
BsmtHalfBath      4.099186
KitchenAbvGr      4.483784
WoodDeckSF        1.539792
OpenPorchSF       2.361912
EnclosedPorch     3.086696
3SsnPorch        10.293752
ScreenPorch       4.117977
PoolArea         14.813135
MiscVal          24.451640
TotalSF           1.774874
dtype: float64
```

In [1066]:
```
data_num.hist(figsize=(16, 20), bins=50, xlabelsize=8, ylabelsize=8); # ; avoid having the matplotli
b verbose informations
```


3SsnPorch
BedroomAbvGr
BsmtFinSF1
BsmtFinSF2
BsmtFullBath
BsmtHalfBath
BsmtUnfSF
EnclosedPorch
Fireplaces
FullBath
GarageArea
GarageCars
GarageYrBlt
GrLivArea
HalfBath
KitchenAbvGr
LotArea
LotFrontage
LowQualFinSF
MasVnrArea
MiscVal
OpenPorchSF
OverallQual
PoolArea
ScreenPorch



3.6 Mean Normalization

In [1070]:

```
data_num = ((data_num - data_num.mean())/(data_num.max() - data_num.min()))
data_num.describe()
```

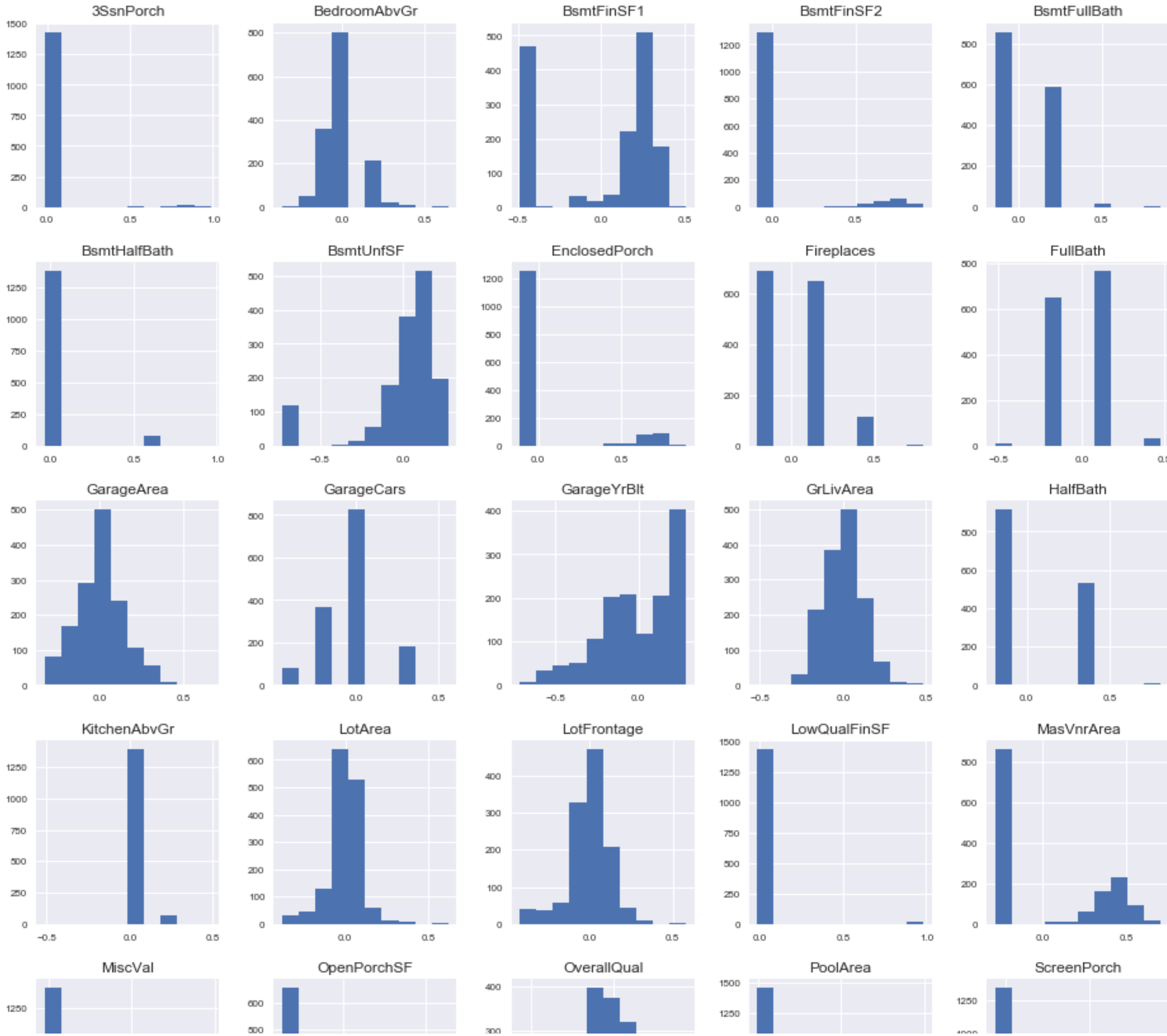
Out[1070]:

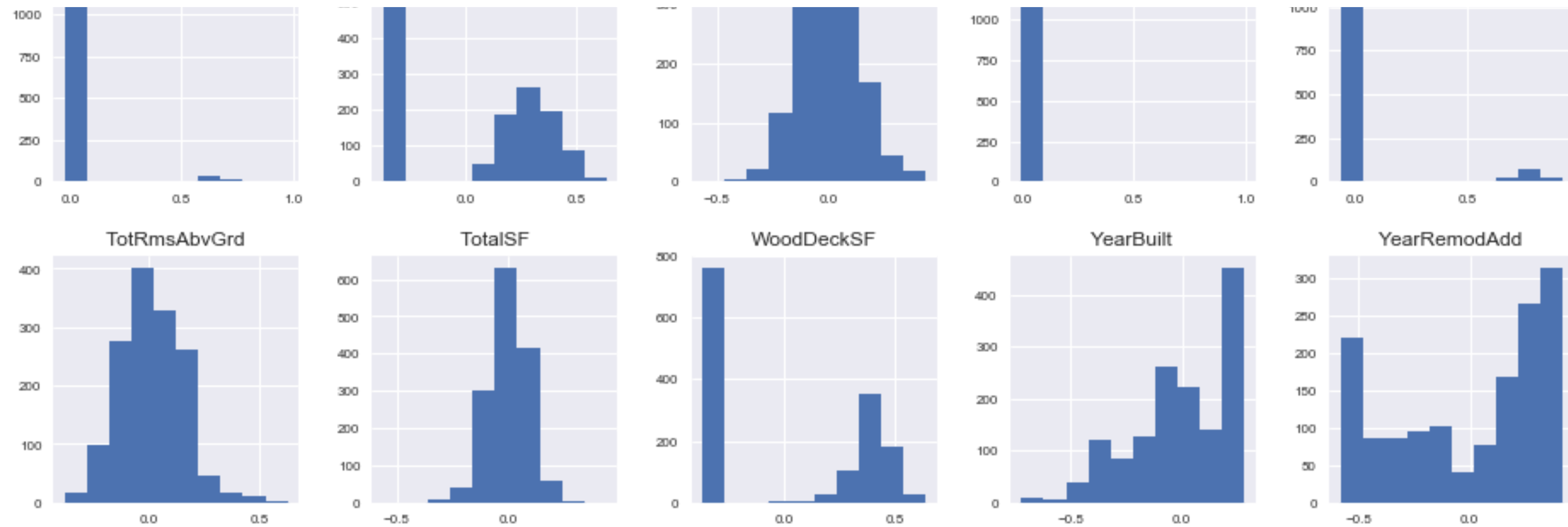
	LotFrontage	LotArea	OverallQual	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	BsmtFinSF2
count	1.201000e+03	1.460000e+03	1.460000e+03	1.460000e+03	1.460000e+03	1.452000e+03	1.460000e+03	1.460000e+03
mean	-2.368815e-18	2.231377e-18	1.296528e-17	1.126382e-17	-4.828710e-18	1.364840e-16	3.726091e-18	9.429291e-18
std	1.302418e-01	1.012732e-01	1.536663e-01	2.188616e-01	3.440901e-01	3.566180e-01	3.463615e-01	2.528701e-01
min	-4.198342e-01	-3.797634e-01	-5.665906e-01	-7.193319e-01	-5.810959e-01	-2.889448e-01	-4.896358e-01	-8.982469e-02
25%	-4.241893e-02	-3.544346e-02	-1.221461e-01	-1.251290e-01	-2.977626e-01	-2.889448e-01	-4.896358e-01	-8.982469e-02
50%	1.556841e-02	8.988780e-03	-1.103501e-02	1.255211e-02	1.522374e-01	-2.889448e-01	1.993641e-01	-8.982469e-02
75%	7.047235e-02	4.854699e-02	1.000761e-01	2.082043e-01	3.189041e-01	4.047022e-01	2.708912e-01	-8.982469e-02
max	5.801658e-01	6.202366e-01	4.334094e-01	2.806681e-01	4.189041e-01	7.110552e-01	5.103642e-01	9.101753e-01

8 rows × 30 columns

In [1073]:

```
data_num.hist(figsize=(16, 20),xlabelsize=8, ylabelsize=8);
```


3SsnPorch
BedroomAbvGr
BsmtFinSF1
BsmtFinSF2
BsmtFullBath
BsmtHalfBath
BsmtUnfSF
EnclosedPorch
Fireplaces
FullBath
GarageArea
GarageCars
GarageYrBlt
GrLivArea
HalfBath
KitchenAbvGr
LotArea
LotFrontage
LowQualFinSF
MasVnrArea
MiscVal
OpenPorchSF
OverallQual
PoolArea
ScreenPorch



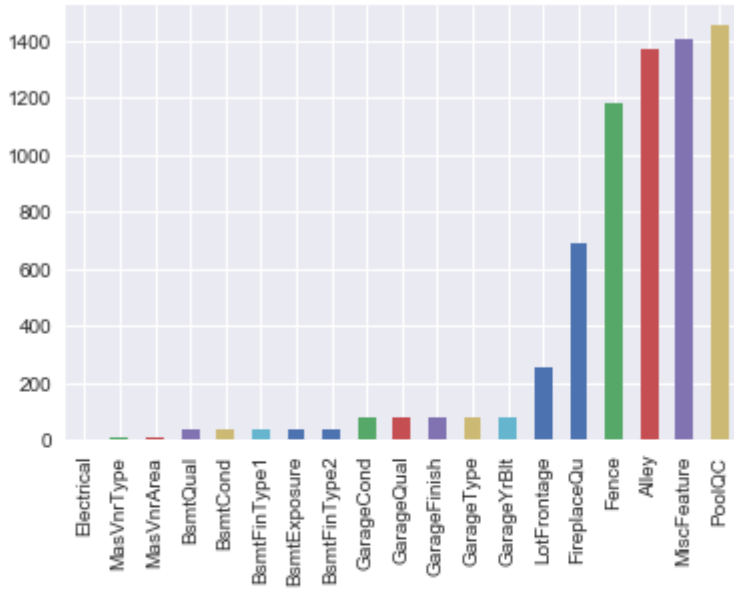
After Mean Normalization the data scale will change and it will not affect original data distribution

4. Missing Data Analysis

If the missing values are not handled properly we may end up drawing an inaccurate inference about the data. Due to improper handling, the result obtained will differ from the ones where the missing values are present.

In [933]:
```
# first we'll visualize null count in overall dataframe
null_in_HousePrice = data.isnull().sum()
null_in_HousePrice = null_in_HousePrice[null_in_HousePrice > 0]
null_in_HousePrice.sort_values(inplace=True)
null_in_HousePrice.plot.bar()
```

Out[933]: <matplotlib.axes._subplots.AxesSubplot at 0x151597dd8d0>



In [934]:

```python
# Printing total numbers and percentage of missing data
total = data.isnull().sum().sort_values(ascending=False)
percent = (data.isnull().sum()/data.isnull().count()).sort_values(ascending=False)
missing_data = pd.concat([total, percent], axis=1, keys=['Total', 'Percent'])
missing_data.head(15)
```

Out[934]:

	Total	Percent
PoolQC	1453	0.995205
MiscFeature	1406	0.963014
Alley	1369	0.937671
Fence	1179	0.807534
FireplaceQu	690	0.472603
LotFrontage	259	0.177397
GarageType	81	0.055479
GarageCond	81	0.055479
GarageYrBlt	81	0.055479
GarageFinish	81	0.055479
GarageQual	81	0.055479
BsmtFinType2	38	0.026027
BsmtExposure	38	0.026027
BsmtCond	37	0.025342
BsmtQual	37	0.025342

5. Missing Data Treatment

We may leave the data as it is or do data imputation to replace them. Suppose the number of cases of missing values is extremely small; then we may drop or omit those values from the analysis. In statistical language, if the number of the cases is less than 5% of the sample, then we can drop them.

If there is a larger number of missing values, then it is better to drop those cases (rather than do imputation) and replace them.

5.1 Handling Missing Values in Numerical Columns

Here we do data imputation. If the number of missing values is more than 260, we drop those values from the analysis.

In [935]:
```
data_len = data_num.shape[0]

# check what is percentage of missing values in categorical dataframe
for col in data_num.columns.values:
    missing_values = data_num[col].isnull().sum()
    #print("{} - missing values: {} ({:0.2f}%)".format(col, missing_values, missing_values/data_len*1
00))

    # drop column if there is more than 50 missing values
    if missing_values > 260:
        #print("droping column: {}".format(col))
        data_num = data_num.drop(col, axis = 1)
    # if there is less than 260 missing values than fill in with median valu of column
    else:
        #print("filling missing values with median in column: {}".format(col))
        data_num = data_num.fillna(data_num[col].median())
```

5.2 Handling Missing Values in Categorical Columns

Here we do data imputation. If the number of missing values is more than 50, we drop the column from the analysis.

In [936]:
```
data_len = data_cat.shape[0]

# check what is percentage of missing values in categorical dataframe
for col in data_cat.columns.values:
    missing_values = data_cat[col].isnull().sum()
    #print("{} - missing values: {} ({:0.2f}%)".format(col, missing_values, missing_values/data_len*1
00))

    # drop column if there is more than 50 missing values
    if missing_values > 50:
        print("droping column: {}".format(col))
        data_cat.drop(col, axis = 1)
    # if there is less than 50 missing values than fill in with median valu of column
    else:
        #print("filling missing values with XXX: {}".format(col))
        #data_cat = data_cat.fillna('XXX')
        pass
```

```
droping column: Alley
droping column: FireplaceQu
droping column: GarageType
droping column: GarageFinish
droping column: GarageQual
droping column: GarageCond
droping column: PoolQC
droping column: Fence
droping column: MiscFeature
```

In [937]:
```
data_cat.describe()
```

Out[937]:

	MSSubClass	MSZoning	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	...	G
count	1460	1460	1460	91	1460	1460	1460	1460	1460	1460	...	1
unique	15	5	2	2	4	4	2	5	3	25	...	5
top	20	RL	Pave	Grvl	Reg	Lvl	AllPub	Inside	Gtl	NAmes	...	T
freq	536	1151	1454	50	925	1311	1459	1052	1382	225	...	1

4 rows × 47 columns

6. Dummy Coding for Categorical Variables

Dummy coding is a way of incorporating nominal variables into regression analysis. It allows us to turn categories into something a regression can treat as having a high (1) and low (0) score. Any binary variable can be thought of as having directionality, because if it is higher, it is category 1, but if it is lower, it is category 0. This allows the regression look at directionality by comparing two sides, rather than expecting each unit to correspond with some kind of increase.

In [938]:
```
data_cat.columns
```

Out[938]:
```
Index(['MSSubClass', 'MSZoning', 'Street', 'Alley', 'LotShape', 'LandContour',
       'Utilities', 'LotConfig', 'LandSlope', 'Neighborhood', 'Condition1',
       'Condition2', 'BldgType', 'HouseStyle', 'OverallCond', 'RoofStyle',
       'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType', 'ExterQual',
       'ExterCond', 'Foundation', 'BsmtQual', 'BsmtCond', 'BsmtExposure',
       'BsmtFinType1', 'BsmtFinType2', 'Heating', 'HeatingQC', 'CentralAir',
       'Electrical', 'KitchenQual', 'Functional', 'FireplaceQu', 'GarageType',
       'GarageFinish', 'GarageQual', 'GarageCond', 'PavedDrive', 'PoolQC',
       'Fence', 'MiscFeature', 'MoSold', 'YrSold', 'SaleType',
       'SaleCondition'],
      dtype='object')
```

In [939]:
```
# Using pandas.get_dummies function to Convert categorical variable into dummy/indicator variables
data_cat_dummies  = pd.get_dummies(data_cat,drop_first=True)
```

In [940]:
```
# Viewing dimensionality of the DataFrame.
data_cat_dummies.head()
```

Out[940]:

	MSSubClass_160	MSSubClass_180	MSSubClass_190	MSSubClass_20	MSSubClass_30	MSSubClass_40	MSSubClass_4
0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0

5 rows × 246 columns

In [941]:
```
print("Numerical features : " + str(len(data_num.columns)))
print("Categorical features : " + str(len(data_cat_dummies.columns)))
```

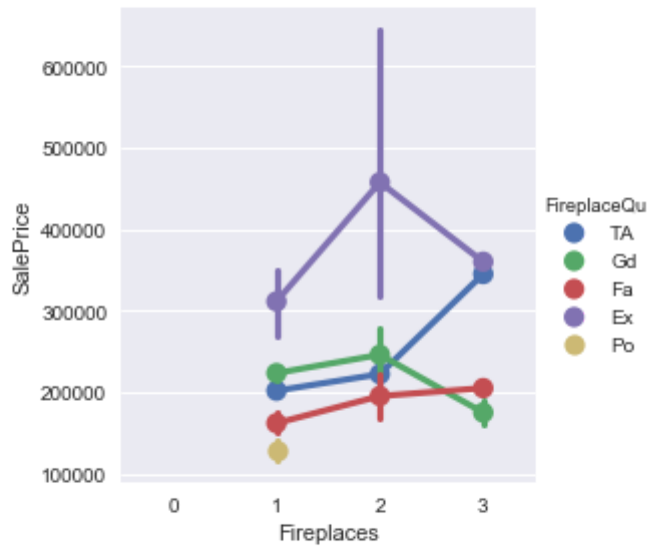
```
Numerical features : 31
Categorical features : 246
```

In [942]:
```
# using concat function we merging two dataframe for furthere analysis
newdata = pd.concat([data_num, data_cat_dummies], axis=1)
```

7. Exploratory Data Analysis

In [1010]:

```
sns.factorplot("Fireplaces","SalePrice",data=raw_data,hue="FireplaceQu");
```



If there are two fireplaces, the Sales Price increases. Also, if there are fireplace of Excellent quality in the house the Sales Price increases.

In [1011]:

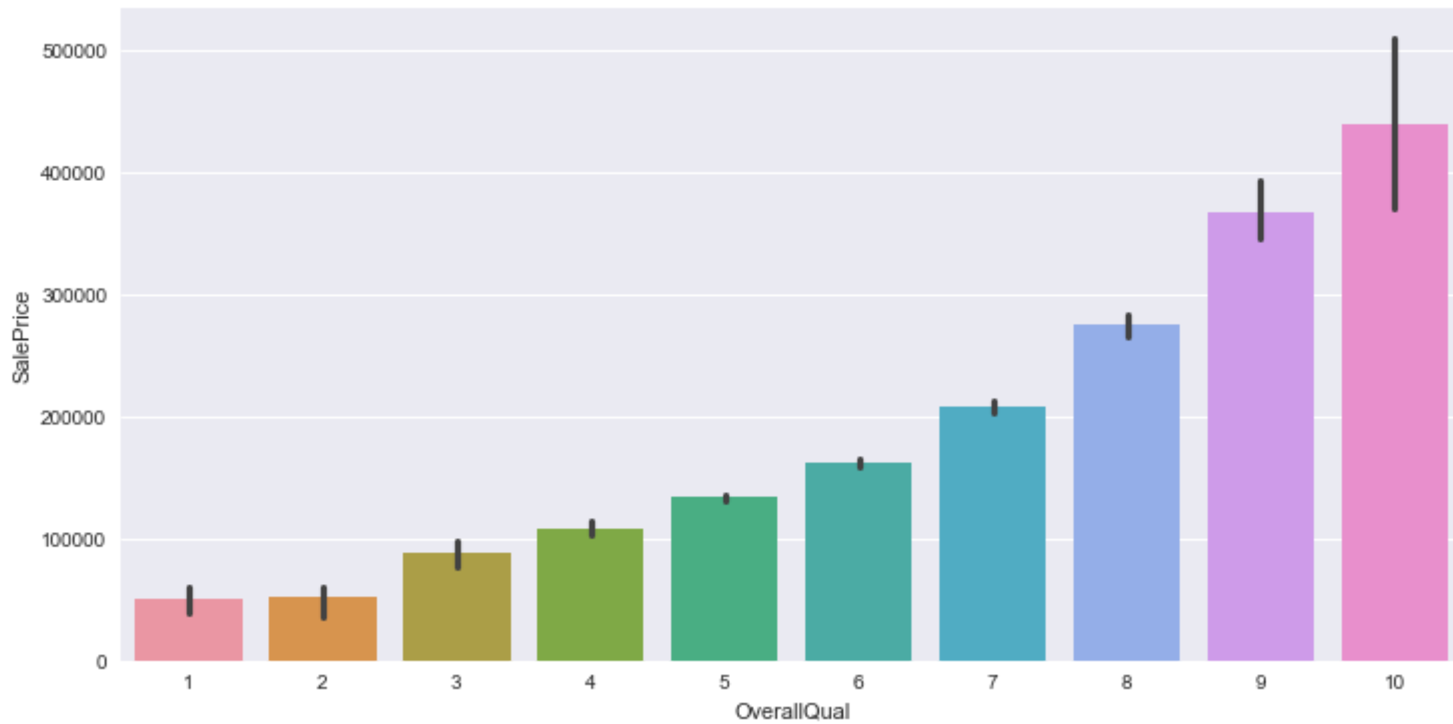
```
# If fireplace is missing that means that house doesn't have a FireplaceQu
FireplaceQu = raw_data["FireplaceQu"].fillna('None')
pd.crosstab(raw_data.Fireplaces, raw_data.FireplaceQu)
```

Out[1011]:

FireplaceQu	Ex	Fa	Gd	None	Po	TA
Fireplaces						
0	0	0	0	690	0	0
1	19	28	324	0	20	259
2	4	4	54	0	0	53
3	1	1	2	0	0	1

In [1012]:
```
sns.barplot(raw_data.OverallQual,raw_data.SalePrice)
```

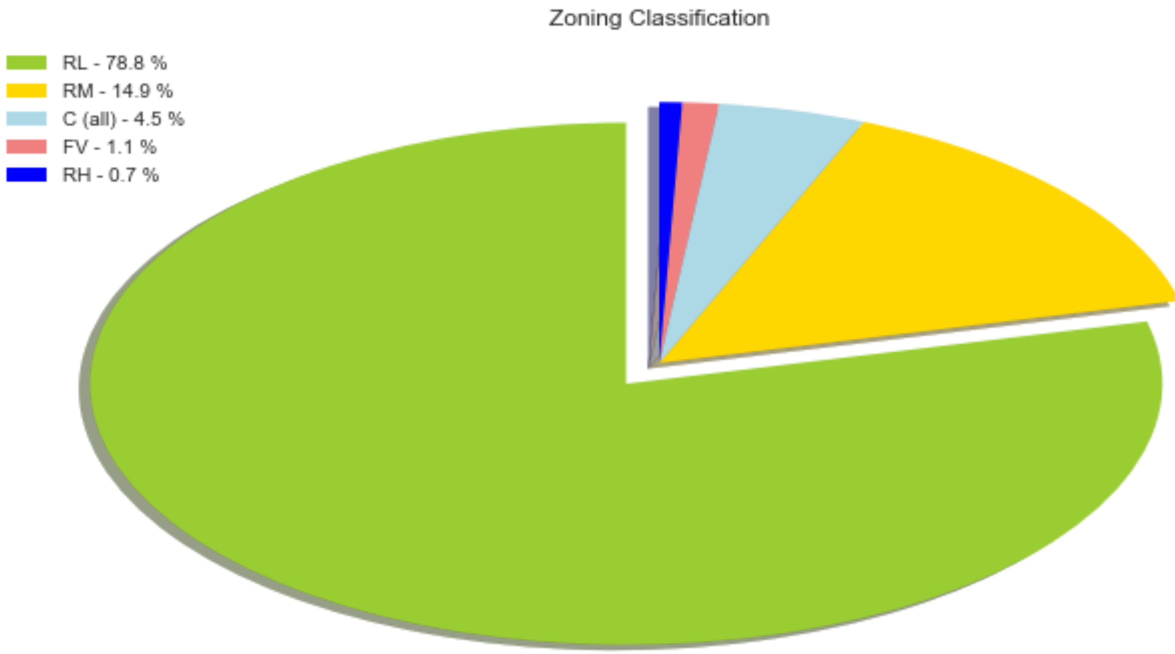
Out[1012]: <matplotlib.axes._subplots.AxesSubplot at 0x1515ee84198>

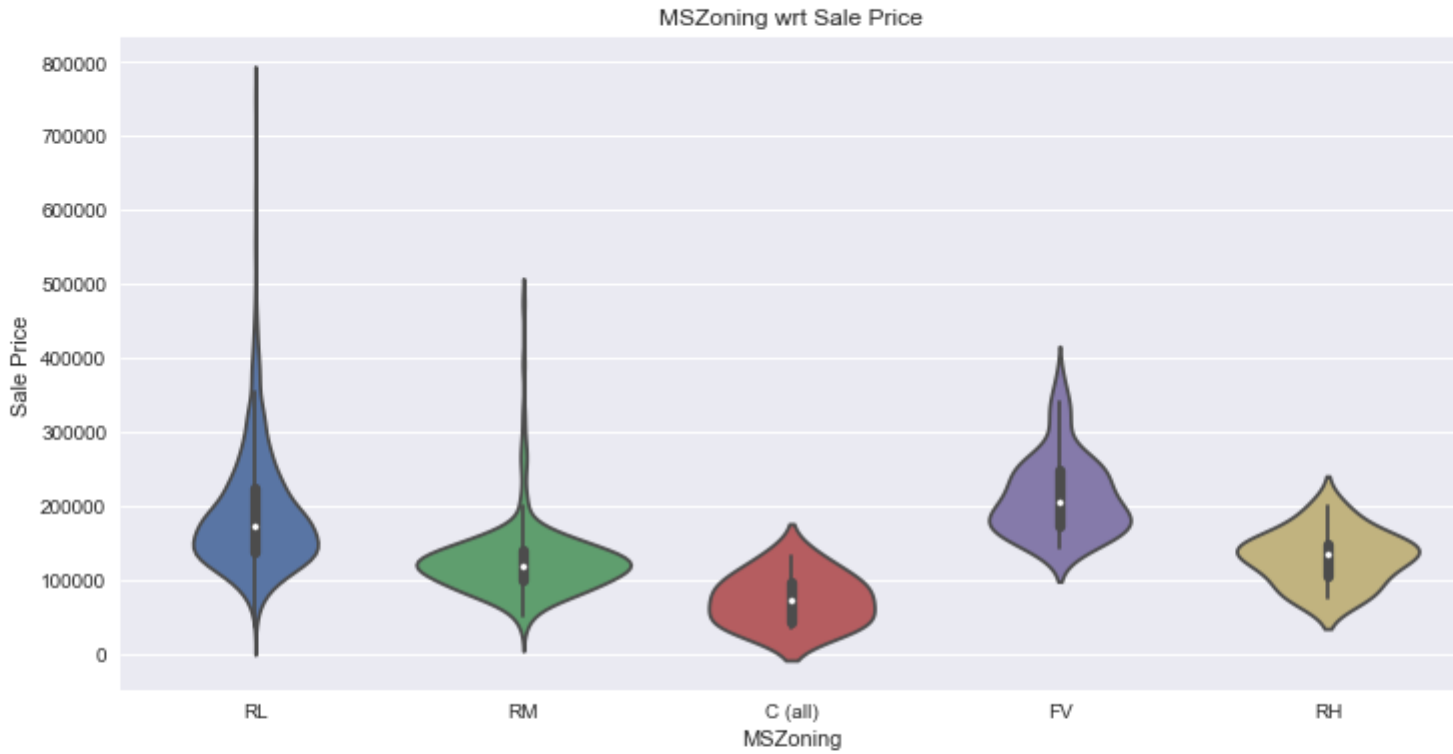


As we can see, the Sales Price increases with the increase in Overall Quality.

In [1014]:

```
# MSZoning
labels = raw_data["MSZoning"].unique()
sizes = raw_data["MSZoning"].value_counts().values
explode=[0.1,0,0,0,0]
parcent = 100.*sizes/sizes.sum()
labels = ['{0} - {1:1.1f} %'.format(i,j) for i,j in zip(labels, parcent)]

colors = ['yellowgreen', 'gold', 'lightblue', 'lightcoral','blue']
patches, texts= plt.pie(sizes, colors=colors,explode=explode,
                        shadow=True,startangle=90)
plt.legend(patches, labels, loc="best")

plt.title("Zoning Classification")
plt.show()

sns.violinplot(raw_data.MSZoning,raw_data["SalePrice"])
plt.title("MSZoning wrt Sale Price")
plt.xlabel("MSZoning")
plt.ylabel("Sale Price");
```

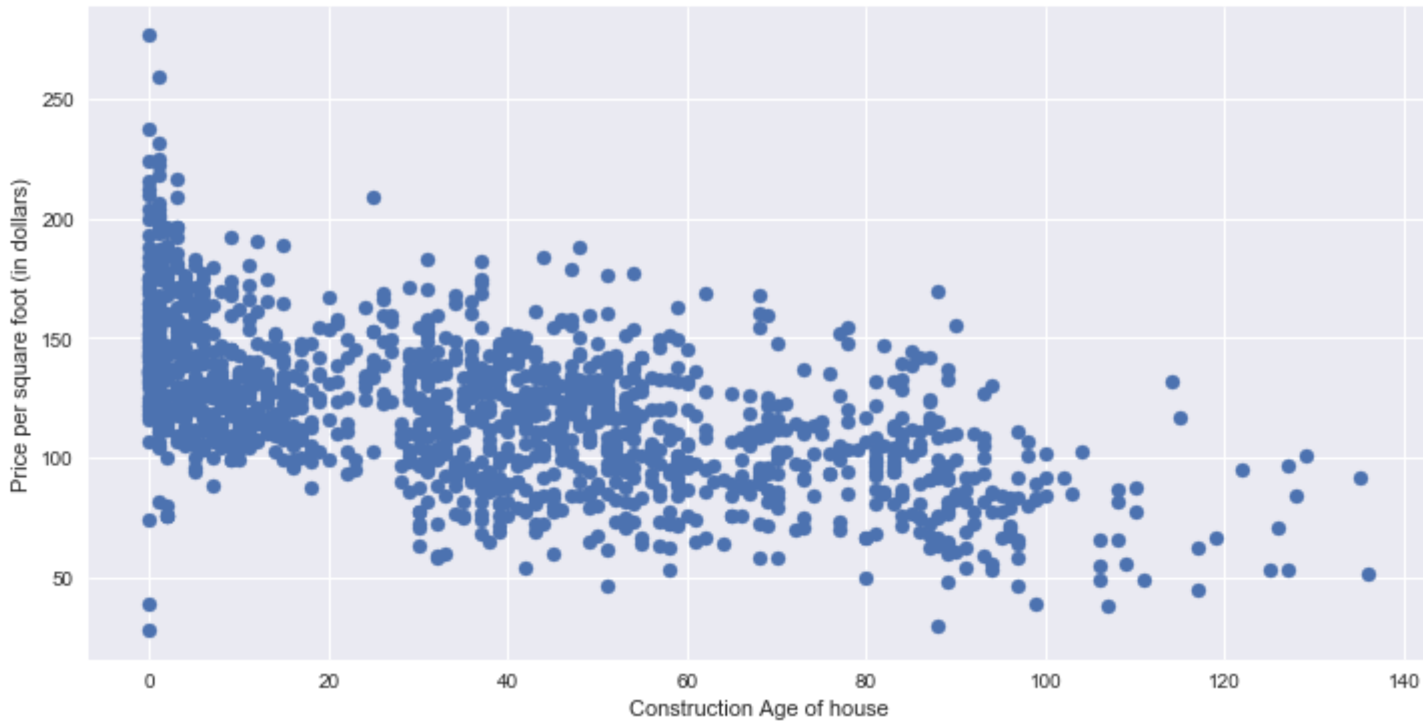

Zoning Classification
RL - 78.8 %
RM - 14.9 %
C (all) - 4.5 %
FV - 1.1 %
RH - 0.7 %


MSZoning wrt Sale Price
Sale Price
800000
700000
600000
500000
400000
300000
200000
100000
0
RL
RM
C (all)
FV
RH
MSZoning

In [1015]:
```
# SalePrice per Square Foot
SalePriceSF = raw_data['SalePrice']/raw_data['GrLivArea']
plt.hist(SalePriceSF, color="green")
plt.title("Sale Price per Square Foot")
plt.ylabel('Number of Sales')
plt.xlabel('Price per square feet');
```

Most of the sales happend in 100 to 150 square feet

In [1016]:
```
ConstructionAge = raw_data['YrSold'] - raw_data['YearBuilt']
plt.scatter(ConstructionAge, SalePriceSF)
plt.ylabel('Price per square foot (in dollars)')
plt.xlabel("Construction Age of house");
```



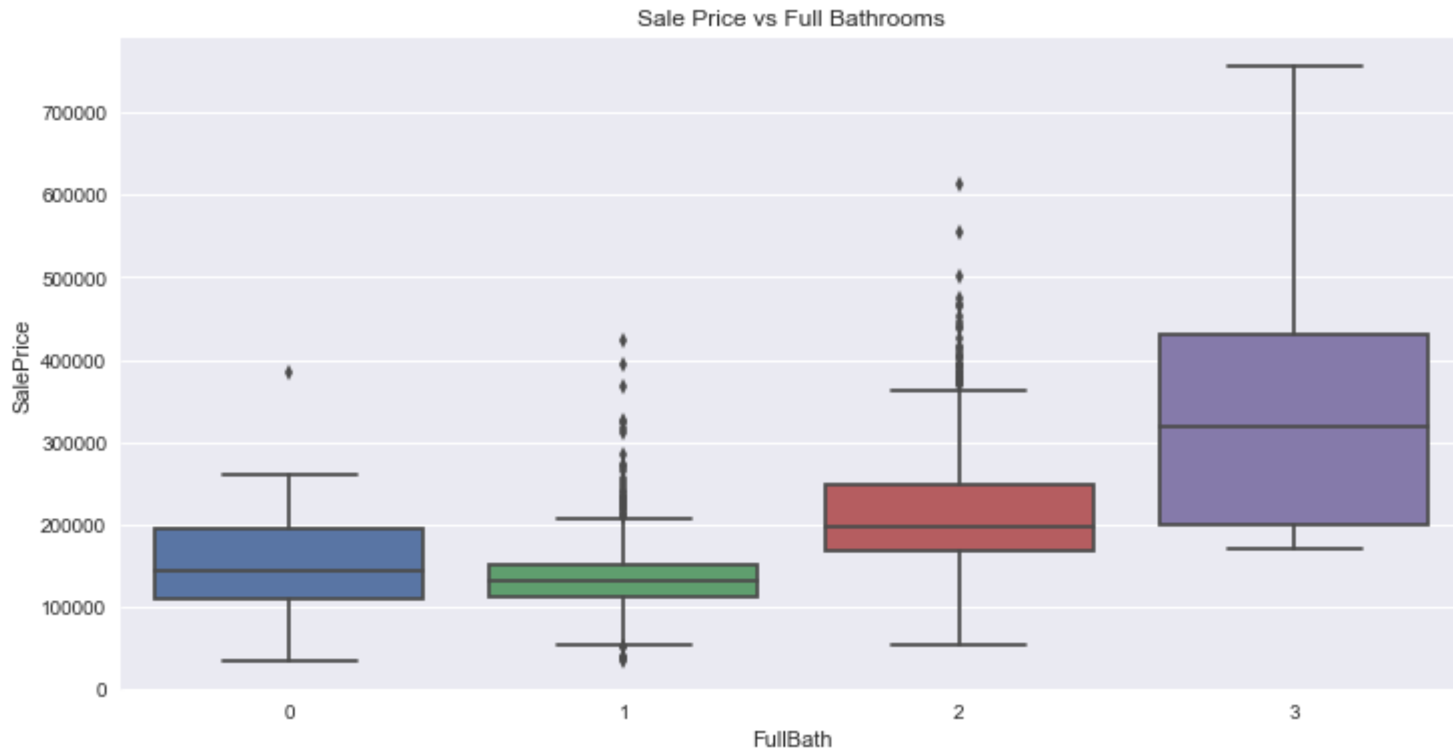
From the above representation, price of house goes down with its age.

In [1018]:
```
# Heating and AC arrangements
sns.stripplot(x="HeatingQC", y="SalePrice",data=raw_data,hue='CentralAir',jitter=True,split=True)
plt.title("Sale Price vs Heating Quality");
```

```
C:\Users\computer\Anaconda3\lib\site-packages\seaborn\categorical.py:2586: UserWarning: The `split` parameter has been renamed to `dodge`.
  warnings.warn(msg, UserWarning)
```

Having AC definitely escalates price of house.

In [1019]:
```
sns.boxplot(raw_data["FullBath"],raw_data["SalePrice"])
plt.title("Sale Price vs Full Bathrooms");
```



In [1020]:
```
# Kitchen Quality
sns.factorplot("KitchenAbvGr","SalePrice",data=raw_data,hue="KitchenQual")
plt.title("Sale Price vs Kitchen");
```

Having one Kitchen of Excellent quality hikes house price.

7.1 Correlation

In [945]:
```
# Check Correlation
data_num.corr()
```

```
Out[945]:
```

	Id	LotFrontage	LotArea	OverallQual	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	Bsmtl
Id	1.000000	-0.018167	-0.017483	-0.028365	-0.012713	-0.021998	-0.030614	-0.012806	-0.005
LotFrontage	-0.018167	1.000000	0.646757	0.207340	0.085230	0.061083	0.114949	0.061908	0.0328
LotArea	-0.017483	0.646757	1.000000	0.178220	0.021943	0.027672	0.070586	0.096966	0.084
OverallQual	-0.028365	0.207340	0.178220	1.000000	0.572323	0.550684	0.413760	0.054199	-0.101
YearBuilt	-0.012713	0.085230	0.021943	0.572323	1.000000	0.592855	0.412414	0.151209	-0.068
YearRemodAdd	-0.021998	0.061083	0.027672	0.550684	0.592855	1.000000	0.224582	0.012105	-0.102
MasVnrArea	-0.030614	0.114949	0.070586	0.413760	0.412414	0.224582	1.000000	0.187073	-0.063
BsmtFinSF1	-0.012806	0.061908	0.096966	0.054199	0.151209	0.012105	0.187073	1.000000	0.185
BsmtFinSF2	-0.005280	0.032854	0.084312	-0.101469	-0.068793	-0.102425	-0.063038	0.185051	1.000
BsmtUnfSF	-0.008539	0.106388	0.047510	0.268446	0.090576	0.167086	0.093676	-0.265488	-0.272
LowQualFinSF	-0.033418	-0.004481	-0.010681	-0.029924	-0.182266	-0.057486	-0.108585	-0.063563	0.006
GrLivArea	0.004185	0.320382	0.385457	0.614207	0.235500	0.311456	0.305297	-0.014440	-0.042
BsmtFullBath	0.002289	0.063432	0.138279	0.111098	0.187599	0.119470	0.116191	0.591148	0.150
BsmtHalfBath	-0.020778	-0.000991	0.047896	-0.041567	-0.040385	-0.013277	0.039029	0.109654	0.117
FullBath	0.005587	0.164084	0.179193	0.550600	0.468271	0.439046	0.280397	-0.085258	-0.098
HalfBath	0.006784	0.014088	0.038728	0.273458	0.242656	0.183331	0.152342	-0.005373	-0.048
BedroomAbvGr	0.037719	0.267436	0.279173	0.101676	-0.070651	-0.040581	0.089910	-0.094666	0.007
KitchenAbvGr	0.003069	0.021505	0.002947	-0.181780	-0.171485	-0.148805	-0.060024	-0.134300	-0.036
TotRmsAbvGrd	0.027239	0.321295	0.360131	0.427452	0.095589	0.191740	0.233967	-0.091730	-0.045
Fireplaces	-0.019772	0.217844	0.327765	0.396765	0.147716	0.112581	0.225613	0.146194	0.039
GarageYrBlt	0.000070	0.041479	-0.019681	0.518018	0.780555	0.618130	0.311101	0.026019	-0.108
GarageCars	0.016570	0.273036	0.272010	0.600671	0.537850	0.420622	0.389981	0.070302	-0.053
GarageArea	0.017634	0.317929	0.322048	0.562022	0.478954	0.371600	0.370968	0.117933	-0.018

	Id	LotFrontage	LotArea	OverallQual	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	Bsmtl
WoodDeckSF	-0.048039	0.057052	0.122915	0.283256	0.307474	0.285773	0.161553	0.082144	0.0591
OpenPorchSF	0.006396	0.132994	0.137754	0.445101	0.389065	0.382743	0.209857	0.015075	-0.077
EnclosedPorch	-0.005415	-0.019167	-0.015186	-0.158648	-0.446694	-0.226996	-0.186104	-0.127733	0.0385
3SsnPorch	-0.039543	0.055061	0.055375	0.027546	0.028977	0.052398	0.047053	0.051775	-0.022
ScreenPorch	0.004489	0.044815	0.087518	0.049983	-0.059629	-0.039983	0.037424	0.073392	0.0638
PoolArea	0.055796	0.120014	0.097257	0.072651	0.005805	0.009025	0.005363	0.042147	0.0680
MiscVal	-0.038611	0.025451	0.080298	-0.085131	-0.077819	-0.074574	-0.054456	0.018808	0.0279
TotalSF	-0.000938	0.372405	0.427126	0.678565	0.366477	0.362181	0.382008	0.167094	0.0193

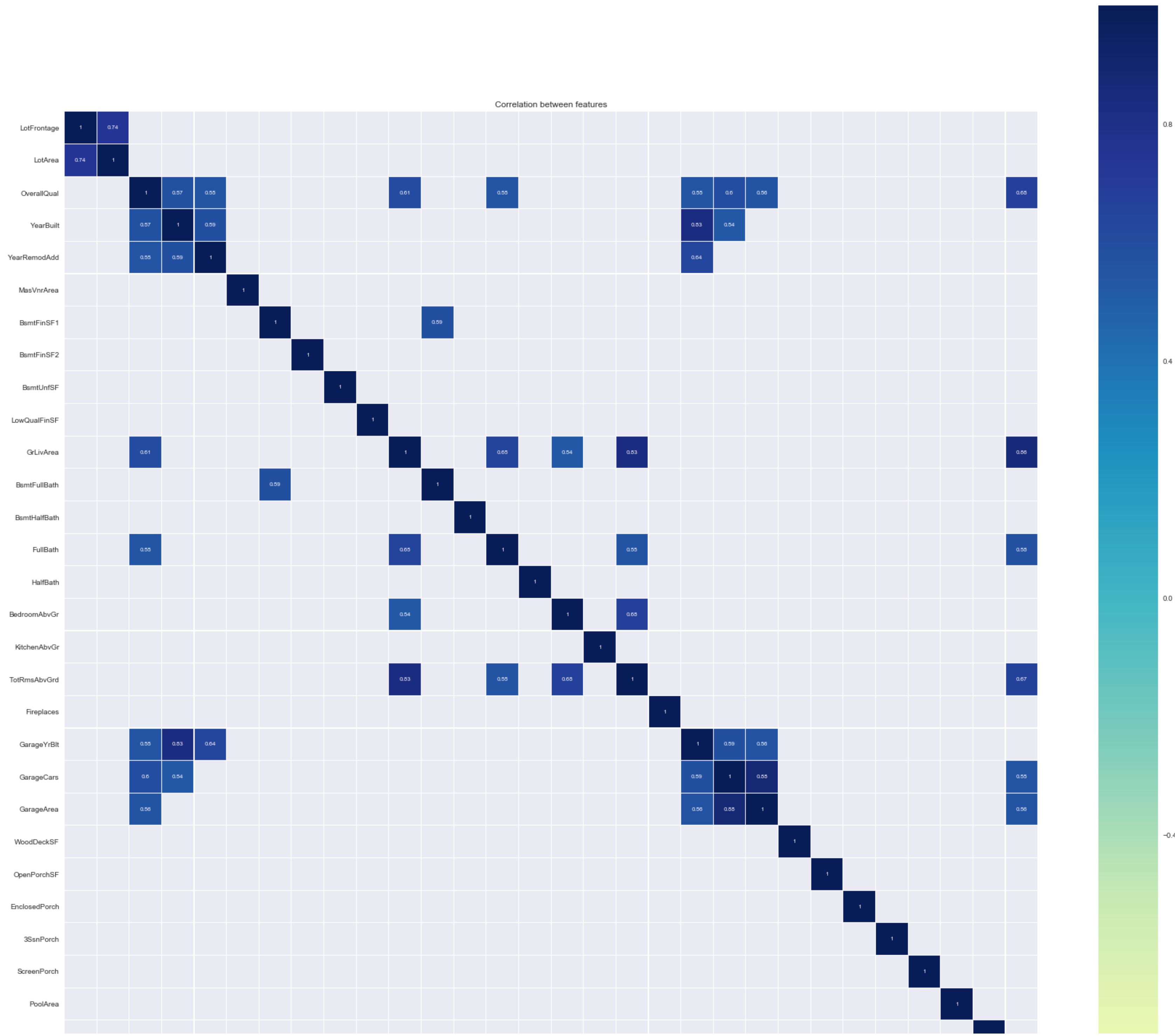
31 rows × 31 columns

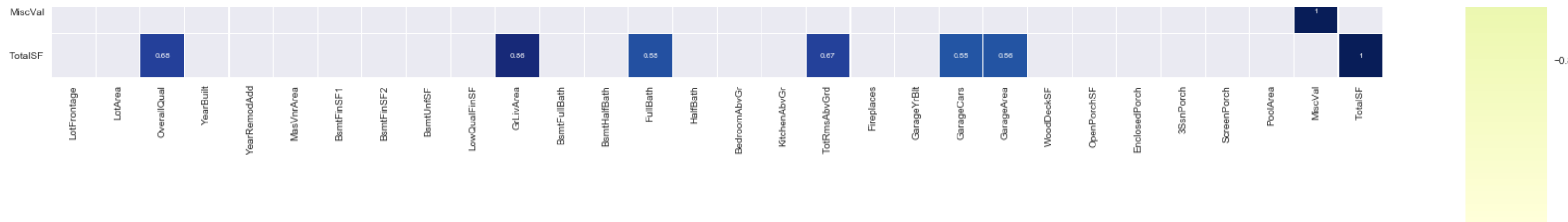
7.2 Correlation Plot

In [1109]:

```
# Ploting correlation plot
import matplotlib.pyplot as plt
corr=data_num.corr()
plt.figure(figsize=(30, 30))

sns.heatmap(corr[(corr >= 0.5) | (corr <= -0.5)],
            cmap='YlGnBu', vmax=1.0, vmin=-1.0, linewidths=0.1,
            annot=True, annot_kws={"size": 8}, square=True);
plt.title('Correlation between features')
```

```
Out[1109]: Text(0.5,1,'Correlation between features')
```


Correlation between features
LotFrontage
LotArea
OverallQual
YearBuilt
YearRemodAdd
MasVnrArea
BsmtFinSF1
BsmtFinSF2
BsmtUnfSF
LowQualFinSF
GrLivArea
BsmtFullBath
BsmtHalfBath
FullBath
HalfBath
BedroomAbvGr
KitchenAbvGr
TotRmsAbvGrd
Fireplaces
GarageYrBlt
GarageCars
GarageArea
WoodDeckSF
OpenPorchSF
EnclosedPorch
3SsnPorch
ScreenPorch
PoolArea
0.8
0.4
0.0
-0.4



As we can see above there are few features which show high multicollinearity from heatmap. Dark Blue squares on diagonal line has high multicollinearity

8. Linear Regression Modeling

8.1 Preparation of Datasets

Split the dataset into Train & Test

In [947]:
```
#Let us now split the dataset into train & test
from sklearn.cross_validation import train_test_split
x_train,x_test, y_train, y_test = train_test_split(newdata, target_log, test_size = 0.30, random_stat
e=0)
print("x_train ",x_train.shape)
print("x_test ",x_test.shape)
print("y_train ",y_train.shape)
print("y_test ",y_test.shape)
```

```
x_train  (1022, 277)
x_test  (438, 277)
y_train  (1022,)
y_test  (438,)
```

8.2 Building a Linear Regression Base Model

In [948]:

```
# Lets build Linear Regression model using statsmodel
import statsmodels.api as sm

# Building Linear Regression model using OLS
model1 = sm.OLS(y_train, x_train).fit()
# Note the Swap of X and Y
```

In [949]:
```python
# Printing Linear Regression Summary
model1.summary()
```

Out[949]:

OLS Regression Results

Dep. Variable:	SalePrice	**R-squared:**	0.959
Model:	OLS	**Adj. R-squared:**	0.944
Method:	Least Squares	**F-statistic:**	65.93
Date:	Wed, 25 Jul 2018	**Prob (F-statistic):**	0.00
Time:	13:25:33	**Log-Likelihood:**	1107.7
No. Observations:	1022	**AIC:**	-1683.
Df Residuals:	756	**BIC:**	-372.1
Df Model:	265		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Id	-0.0241	0.012	-2.019	0.044	-0.048	-0.001
LotFrontage	0.0760	0.051	1.481	0.139	-0.025	0.177
LotArea	0.5298	0.079	6.705	0.000	0.375	0.685
OverallQual	0.3003	0.050	6.007	0.000	0.202	0.398
YearBuilt	0.2447	0.065	3.744	0.000	0.116	0.373
YearRemodAdd	0.0465	0.019	2.476	0.013	0.010	0.083
MasVnrArea	0.0532	0.047	1.131	0.259	-0.039	0.146
BsmtFinSF1	0.0963	0.051	1.878	0.061	-0.004	0.197
BsmtFinSF2	0.0031	0.077	0.040	0.968	-0.147	0.153
BsmtUnfSF	-0.0365	0.026	-1.420	0.156	-0.087	0.014
LowQualFinSF	0.0097	0.040	0.244	0.807	-0.068	0.088
GrLivArea	0.5945	0.179	3.318	0.001	0.243	0.946
BsmtFullBath	0.0877	0.031	2.860	0.004	0.028	0.148

BsmtHalfBath	0.0424	0.027	1.591	0.112	-0.010	0.095
FullBath	0.1111	0.036	3.095	0.002	0.041	0.182
HalfBath	0.0608	0.024	2.528	0.012	0.014	0.108
BedroomAbvGr	-0.0683	0.062	-1.108	0.268	-0.189	0.053
KitchenAbvGr	-0.1893	0.115	-1.642	0.101	-0.416	0.037
TotRmsAbvGrd	0.0414	0.059	0.699	0.485	-0.075	0.158
Fireplaces	0.0570	0.036	1.571	0.117	-0.014	0.128
GarageYrBlt	-0.0136	0.038	-0.363	0.717	-0.087	0.060
GarageCars	0.0141	0.052	0.274	0.784	-0.087	0.115
GarageArea	0.1986	0.063	3.152	0.002	0.075	0.322
WoodDeckSF	0.0375	0.010	3.625	0.000	0.017	0.058
OpenPorchSF	0.0082	0.013	0.651	0.515	-0.017	0.033
EnclosedPorch	0.0099	0.015	0.644	0.520	-0.020	0.040
3SsnPorch	-0.0051	0.034	-0.153	0.879	-0.071	0.061
ScreenPorch	0.0374	0.015	2.413	0.016	0.007	0.068
PoolArea	-0.0234	0.164	-0.143	0.887	-0.345	0.299
MiscVal	0.0331	0.091	0.362	0.718	-0.147	0.213
TotalSF	0.9518	0.214	4.438	0.000	0.531	1.373
MSSubClass_160	-0.0687	0.058	-1.185	0.236	-0.183	0.045
MSSubClass_180	0.0609	0.079	0.774	0.439	-0.094	0.215
MSSubClass_190	-0.0235	0.154	-0.153	0.879	-0.325	0.278
MSSubClass_20	-0.0455	0.068	-0.673	0.501	-0.178	0.087
MSSubClass_30	-0.0968	0.072	-1.348	0.178	-0.238	0.044
MSSubClass_40	0.0152	0.119	0.128	0.898	-0.218	0.248
MSSubClass_45	-0.6041	0.174	-3.476	0.001	-0.945	-0.263

MSSubClass_50	-0.0974	0.087	-1.114	0.265	-0.269	0.074
MSSubClass_60	-0.0696	0.083	-0.837	0.403	-0.233	0.094
MSSubClass_70	-0.0448	0.084	-0.531	0.596	-0.210	0.121
MSSubClass_75	-0.1197	0.113	-1.061	0.289	-0.341	0.102
MSSubClass_80	-0.0783	0.101	-0.774	0.439	-0.277	0.120
MSSubClass_85	-0.0145	0.093	-0.156	0.876	-0.196	0.167
MSSubClass_90	-0.0375	0.040	-0.939	0.348	-0.116	0.041
MSZoning_FV	0.5248	0.065	8.078	0.000	0.397	0.652
MSZoning_RH	0.4490	0.062	7.224	0.000	0.327	0.571
MSZoning_RL	0.4599	0.055	8.337	0.000	0.352	0.568
MSZoning_RM	0.4566	0.051	8.895	0.000	0.356	0.557
Street_Pave	0.1125	0.061	1.829	0.068	-0.008	0.233
Alley_Pave	0.0556	0.027	2.039	0.042	0.002	0.109
LotShape_IR2	0.0042	0.022	0.191	0.849	-0.039	0.048
LotShape_IR3	0.0684	0.048	1.412	0.158	-0.027	0.163
LotShape_Reg	0.0134	0.008	1.580	0.115	-0.003	0.030
LandContour_HLS	-0.0116	0.031	-0.379	0.705	-0.072	0.049
LandContour_Low	-0.0320	0.036	-0.901	0.368	-0.102	0.038
LandContour_Lvl	-0.0066	0.021	-0.306	0.760	-0.049	0.036
Utilities_NoSeWa	-0.2118	0.138	-1.540	0.124	-0.482	0.058
LotConfig_CulDSac	0.0272	0.017	1.560	0.119	-0.007	0.061
LotConfig_FR2	-0.0169	0.021	-0.800	0.424	-0.058	0.024
LotConfig_FR3	-0.1918	0.151	-1.272	0.204	-0.488	0.104
LotConfig_Inside	-0.0140	0.010	-1.460	0.145	-0.033	0.005
LandSlope_Mod	0.0225	0.021	1.065	0.287	-0.019	0.064

LandSlope_Sev	-0.1630	0.053	-3.052	0.002	-0.268	-0.058
Neighborhood_Blueste	0.0024	0.092	0.026	0.979	-0.178	0.183
Neighborhood_BrDale	-0.0958	0.065	-1.472	0.142	-0.224	0.032
Neighborhood_BrkSide	-0.0571	0.054	-1.049	0.294	-0.164	0.050
Neighborhood_ClearCr	-0.0558	0.051	-1.091	0.275	-0.156	0.045
Neighborhood_CollgCr	-0.1002	0.041	-2.421	0.016	-0.182	-0.019
Neighborhood_Crawfor	0.0299	0.050	0.594	0.552	-0.069	0.129
Neighborhood_Edwards	-0.1658	0.046	-3.597	0.000	-0.256	-0.075
Neighborhood_Gilbert	-0.0973	0.043	-2.281	0.023	-0.181	-0.014
Neighborhood_IDOTRR	-0.1833	0.063	-2.903	0.004	-0.307	-0.059
Neighborhood_MeadowV	-0.1905	0.065	-2.935	0.003	-0.318	-0.063
Neighborhood_Mitchel	-0.1735	0.047	-3.661	0.000	-0.267	-0.080
Neighborhood_NAmes	-0.1339	0.045	-2.959	0.003	-0.223	-0.045
Neighborhood_NPkVill	-0.0465	0.067	-0.696	0.487	-0.178	0.085
Neighborhood_NWAmes	-0.1287	0.046	-2.798	0.005	-0.219	-0.038
Neighborhood_NoRidge	0.0117	0.046	0.252	0.801	-0.079	0.103
Neighborhood_NridgHt	-0.0413	0.043	-0.958	0.338	-0.126	0.043
Neighborhood_OldTown	-0.1538	0.056	-2.759	0.006	-0.263	-0.044
Neighborhood_SWISU	-0.0748	0.054	-1.384	0.167	-0.181	0.031
Neighborhood_Sawyer	-0.1217	0.046	-2.631	0.009	-0.212	-0.031
Neighborhood_SawyerW	-0.1013	0.044	-2.291	0.022	-0.188	-0.015
Neighborhood_Somerst	-0.0835	0.049	-1.717	0.086	-0.179	0.012
Neighborhood_StoneBr	0.0169	0.048	0.356	0.722	-0.077	0.111
Neighborhood_Timber	-0.0883	0.046	-1.937	0.053	-0.178	0.001
Neighborhood_Veenker	-0.0656	0.060	-1.087	0.277	-0.184	0.053

Condition1_Feedr	-0.0134	0.031	-0.429	0.668	-0.074	0.048
Condition1_Norm	0.0591	0.026	2.264	0.024	0.008	0.110
Condition1_PosA	0.0274	0.053	0.518	0.605	-0.076	0.131
Condition1_PosN	0.0763	0.041	1.864	0.063	-0.004	0.157
Condition1_RRAe	-0.0767	0.047	-1.650	0.099	-0.168	0.015
Condition1_RRAn	0.0382	0.039	0.971	0.332	-0.039	0.116
Condition1_RRNe	-0.0059	0.079	-0.075	0.940	-0.161	0.149
Condition1_RRNn	0.0611	0.069	0.887	0.375	-0.074	0.196
Condition2_Feedr	0.0727	0.158	0.461	0.645	-0.237	0.382
Condition2_Norm	-0.0028	0.137	-0.020	0.984	-0.272	0.267
Condition2_PosA	-1.792e-14	8.26e-16	-21.700	0.000	-1.95e-14	-1.63e-14
Condition2_PosN	-1.3327	0.176	-7.563	0.000	-1.679	-0.987
Condition2_RRAe	-0.4817	0.262	-1.839	0.066	-0.996	0.032
Condition2_RRAn	5.845e-14	1.77e-15	33.094	0.000	5.5e-14	6.19e-14
Condition2_RRNn	-1.917e-14	9.83e-16	-19.511	0.000	-2.11e-14	-1.72e-14
BldgType_2fmCon	-0.1050	0.130	-0.805	0.421	-0.361	0.151
BldgType_Duplex	-0.0375	0.040	-0.939	0.348	-0.116	0.041
BldgType_Twnhs	-0.0380	0.073	-0.525	0.600	-0.180	0.104
BldgType_TwnhsE	-0.0459	0.068	-0.676	0.500	-0.179	0.088
HouseStyle_1.5Unf	0.5219	0.164	3.175	0.002	0.199	0.845
HouseStyle_1Story	-0.0420	0.053	-0.787	0.432	-0.147	0.063
HouseStyle_2.5Fin	-0.0071	0.090	-0.078	0.937	-0.185	0.170
HouseStyle_2.5Unf	0.0942	0.089	1.053	0.293	-0.081	0.270
HouseStyle_2Story	-0.0444	0.047	-0.936	0.350	-0.138	0.049
HouseStyle_SFoyer	-0.0613	0.074	-0.830	0.407	-0.206	0.084

HouseStyle_SLvl	-0.0347	0.078	-0.445	0.657	-0.188	0.118
OverallCond_2	4.4277	0.119	37.362	0.000	4.195	4.660
OverallCond_3	4.3127	0.101	42.712	0.000	4.114	4.511
OverallCond_4	4.3848	0.098	44.641	0.000	4.192	4.578
OverallCond_5	4.4242	0.098	44.990	0.000	4.231	4.617
OverallCond_6	4.4580	0.098	45.310	0.000	4.265	4.651
OverallCond_7	4.5036	0.099	45.538	0.000	4.309	4.698
OverallCond_8	4.5130	0.100	45.273	0.000	4.317	4.709
OverallCond_9	4.5200	0.100	45.069	0.000	4.323	4.717
RoofStyle_Gable	0.0340	0.083	0.408	0.683	-0.130	0.198
RoofStyle_Gambrel	-0.0215	0.095	-0.228	0.820	-0.207	0.164
RoofStyle_Hip	0.0516	0.084	0.615	0.538	-0.113	0.216
RoofStyle_Mansard	0.0724	0.098	0.735	0.463	-0.121	0.266
RoofStyle_Shed	0.2745	0.181	1.514	0.130	-0.081	0.630
RoofMatl_CompShg	7.3263	0.167	43.934	0.000	6.999	7.654
RoofMatl_Membran	5.014e-15	6.81e-16	7.358	0.000	3.68e-15	6.35e-15
RoofMatl_Metal	7.5603	0.196	38.556	0.000	7.175	7.945
RoofMatl_Roll	7.3087	0.201	36.399	0.000	6.914	7.703
RoofMatl_Tar&Grv	7.3515	0.164	44.942	0.000	7.030	7.673
RoofMatl_WdShake	7.3065	0.179	40.725	0.000	6.954	7.659
RoofMatl_WdShngl	7.3859	0.174	42.533	0.000	7.045	7.727
Exterior1st_AsphShn	-0.0509	0.062	-0.823	0.411	-0.172	0.070
Exterior1st_BrkComm	-0.5255	0.163	-3.228	0.001	-0.845	-0.206
Exterior1st_BrkFace	0.0562	0.076	0.740	0.459	-0.093	0.205
Exterior1st_CBlock	-0.0517	0.066	-0.781	0.435	-0.182	0.078

Exterior1st_CemntBd	-0.0911	0.107	-0.850	0.395	-0.301	0.119
Exterior1st_HdBoard	-0.0013	0.077	-0.017	0.987	-0.152	0.149
Exterior1st_ImStucc	-0.0047	0.131	-0.036	0.971	-0.262	0.253
Exterior1st_MetalSd	-0.0089	0.092	-0.097	0.923	-0.190	0.172
Exterior1st_Plywood	0.0073	0.077	0.095	0.925	-0.143	0.158
Exterior1st_Stone	-0.0266	0.123	-0.216	0.829	-0.268	0.215
Exterior1st_Stucco	0.0184	0.083	0.222	0.824	-0.144	0.181
Exterior1st_VinylSd	0.0045	0.074	0.061	0.951	-0.141	0.150
Exterior1st_Wd Sdng	-0.0573	0.074	-0.780	0.436	-0.202	0.087
Exterior1st_WdShing	-0.0096	0.079	-0.123	0.902	-0.164	0.145
Exterior2nd_AsphShn	-0.0509	0.062	-0.823	0.411	-0.172	0.070
Exterior2nd_Brk Cmn	0.0588	0.108	0.542	0.588	-0.154	0.272
Exterior2nd_BrkFace	-0.0236	0.074	-0.319	0.750	-0.169	0.122
Exterior2nd_CBlock	-0.0517	0.066	-0.781	0.435	-0.182	0.078
Exterior2nd_CmentBd	0.0942	0.103	0.916	0.360	-0.108	0.296
Exterior2nd_HdBoard	-0.0233	0.070	-0.330	0.741	-0.161	0.115
Exterior2nd_ImStucc	-0.0228	0.078	-0.293	0.770	-0.176	0.130
Exterior2nd_MetalSd	-0.0025	0.087	-0.029	0.977	-0.173	0.168
Exterior2nd_Other	-0.0548	0.125	-0.437	0.662	-0.301	0.191
Exterior2nd_Plywood	-0.0339	0.070	-0.487	0.626	-0.171	0.103
Exterior2nd_Stone	0.0260	0.096	0.270	0.787	-0.163	0.215
Exterior2nd_Stucco	0.0244	0.076	0.321	0.748	-0.125	0.174
Exterior2nd_VinylSd	-0.0136	0.069	-0.198	0.843	-0.148	0.121
Exterior2nd_Wd Sdng	0.0311	0.067	0.463	0.644	-0.101	0.163
Exterior2nd_Wd Shng	-0.0206	0.070	-0.295	0.768	-0.158	0.117

MasVnrType_BrkFace	0.0372	0.029	1.268	0.205	-0.020	0.095
MasVnrType_None	0.0669	0.038	1.763	0.078	-0.008	0.141
MasVnrType_Stone	0.0693	0.032	2.201	0.028	0.007	0.131
ExterQual_Fa	0.0209	0.067	0.312	0.755	-0.110	0.152
ExterQual_Gd	-0.0172	0.025	-0.673	0.501	-0.067	0.033
ExterQual_TA	-0.0304	0.028	-1.068	0.286	-0.086	0.025
ExterCond_Fa	-0.1497	0.087	-1.716	0.087	-0.321	0.022
ExterCond_Gd	-0.1222	0.083	-1.473	0.141	-0.285	0.041
ExterCond_Po	-0.2376	0.167	-1.422	0.155	-0.566	0.090
ExterCond_TA	-0.1128	0.083	-1.358	0.175	-0.276	0.050
Foundation_CBlock	0.0246	0.019	1.304	0.193	-0.012	0.062
Foundation_PConc	0.0259	0.020	1.323	0.186	-0.013	0.064
Foundation_Slab	-0.0101	0.069	-0.147	0.883	-0.145	0.124
Foundation_Stone	0.1801	0.058	3.120	0.002	0.067	0.293
Foundation_Wood	0.1163	0.109	1.064	0.288	-0.098	0.331
BsmtQual_Fa	0.0250	0.037	0.673	0.501	-0.048	0.098
BsmtQual_Gd	-0.0452	0.017	-2.594	0.010	-0.079	-0.011
BsmtQual_TA	-0.0345	0.022	-1.602	0.109	-0.077	0.008
BsmtCond_Gd	0.0188	0.029	0.659	0.510	-0.037	0.075
BsmtCond_Po	4.3161	0.158	27.390	0.000	4.007	4.625
BsmtCond_TA	0.0300	0.023	1.279	0.201	-0.016	0.076
BsmtExposure_Gd	0.0520	0.017	3.146	0.002	0.020	0.084
BsmtExposure_Mn	0.0070	0.016	0.435	0.664	-0.025	0.039
BsmtExposure_No	-0.0014	0.011	-0.124	0.901	-0.024	0.021
BsmtFinType1_BLQ	-0.0012	0.015	-0.084	0.933	-0.030	0.027

BsmtFinType1_GLQ	0.0076	0.014	0.564	0.573	-0.019	0.034
BsmtFinType1_LwQ	-0.0441	0.020	-2.176	0.030	-0.084	-0.004
BsmtFinType1_Rec	-0.0209	0.016	-1.315	0.189	-0.052	0.010
BsmtFinType1_Unf	0.0220	0.036	0.607	0.544	-0.049	0.093
BsmtFinType2_BLQ	-0.0865	0.036	-2.381	0.017	-0.158	-0.015
BsmtFinType2_GLQ	-0.0537	0.049	-1.088	0.277	-0.151	0.043
BsmtFinType2_LwQ	-0.0552	0.036	-1.526	0.127	-0.126	0.016
BsmtFinType2_Rec	-0.0464	0.036	-1.292	0.197	-0.117	0.024
BsmtFinType2_Unf	-0.0575	0.063	-0.907	0.365	-0.182	0.067
Heating_GasA	-0.0158	0.124	-0.128	0.898	-0.259	0.227
Heating_GasW	0.0089	0.131	0.068	0.946	-0.248	0.265
Heating_Grav	-0.1993	0.140	-1.428	0.154	-0.473	0.075
Heating_OthW	0.0122	0.173	0.071	0.944	-0.328	0.353
Heating_Wall	0.0378	0.163	0.232	0.817	-0.283	0.358
HeatingQC_Fa	-0.0572	0.025	-2.250	0.025	-0.107	-0.007
HeatingQC_Gd	-0.0190	0.011	-1.781	0.075	-0.040	0.002
HeatingQC_Po	0.0570	0.120	0.474	0.636	-0.179	0.293
HeatingQC_TA	-0.0331	0.011	-3.006	0.003	-0.055	-0.011
CentralAir_Y	0.0507	0.021	2.372	0.018	0.009	0.093
Electrical_FuseF	-0.0291	0.035	-0.841	0.400	-0.097	0.039
Electrical_FuseP	-0.2488	0.105	-2.361	0.018	-0.456	-0.042
Electrical_Mix	-4.3790	0.239	-18.299	0.000	-4.849	-3.909
Electrical_SBrkr	-0.0397	0.016	-2.524	0.012	-0.071	-0.009
KitchenQual_Fa	-0.0768	0.034	-2.270	0.023	-0.143	-0.010
KitchenQual_Gd	-0.0888	0.018	-4.858	0.000	-0.125	-0.053

KitchenQual_TA	-0.0847	0.021	-4.055	0.000	-0.126	-0.044
Functional_Maj2	-0.3974	0.114	-3.484	0.001	-0.621	-0.173
Functional_Min1	-0.0523	0.046	-1.147	0.252	-0.142	0.037
Functional_Min2	-0.0494	0.045	-1.091	0.276	-0.138	0.039
Functional_Mod	-0.0713	0.055	-1.298	0.195	-0.179	0.037
Functional_Sev	1.543e-15	2.08e-16	7.404	0.000	1.13e-15	1.95e-15
Functional_Typ	0.0191	0.040	0.479	0.632	-0.059	0.097
FireplaceQu_Fa	-0.0222	0.026	-0.851	0.395	-0.073	0.029
FireplaceQu_Gd	-0.0048	0.015	-0.317	0.752	-0.034	0.025
FireplaceQu_Po	0.0069	0.030	0.228	0.820	-0.053	0.067
FireplaceQu_TA	-0.0004	0.016	-0.026	0.980	-0.033	0.032
GarageType_Attchd	0.1101	0.048	2.302	0.022	0.016	0.204
GarageType_Basment	0.0812	0.060	1.357	0.175	-0.036	0.199
GarageType_BuiltIn	0.1352	0.050	2.703	0.007	0.037	0.233
GarageType_CarPort	0.0434	0.079	0.553	0.581	-0.111	0.198
GarageType_Detchd	0.1240	0.047	2.615	0.009	0.031	0.217
GarageFinish_RFn	0.0067	0.010	0.652	0.514	-0.013	0.027
GarageFinish_Unf	-0.0016	0.013	-0.127	0.899	-0.026	0.023
GarageQual_Fa	-0.4447	0.142	-3.129	0.002	-0.724	-0.166
GarageQual_Gd	-0.3702	0.146	-2.534	0.011	-0.657	-0.083
GarageQual_Po	-0.1790	0.197	-0.908	0.364	-0.566	0.208
GarageQual_TA	-0.4182	0.140	-2.994	0.003	-0.692	-0.144
GarageCond_Fa	0.2417	0.151	1.602	0.110	-0.055	0.538
GarageCond_Gd	0.2513	0.158	1.590	0.112	-0.059	0.562
GarageCond_Po	0.1645	0.177	0.932	0.352	-0.182	0.511

GarageCond_TA	0.3096	0.149	2.079	0.038	0.017	0.602
PavedDrive_P	-0.0710	0.034	-2.119	0.034	-0.137	-0.005
PavedDrive_Y	-0.0265	0.019	-1.431	0.153	-0.063	0.010
PoolQC_Fa	-1.551e-15	1.54e-16	-10.039	0.000	-1.85e-15	-1.25e-15
PoolQC_Gd	0.2951	0.183	1.611	0.108	-0.065	0.655
Fence_GdWo	-0.0402	0.019	-2.126	0.034	-0.077	-0.003
Fence_MnPrv	0.0016	0.012	0.132	0.895	-0.022	0.026
Fence_MnWw	-0.0612	0.042	-1.451	0.147	-0.144	0.022
MiscFeature_Othr	-0.0999	0.114	-0.880	0.379	-0.323	0.123
MiscFeature_Shed	-0.0070	0.064	-0.110	0.913	-0.132	0.118
MiscFeature_TenC	-1.358e-15	1.14e-16	-11.879	0.000	-1.58e-15	-1.13e-15
MoSold_10	-0.0052	0.023	-0.228	0.820	-0.050	0.040
MoSold_11	0.0008	0.023	0.035	0.972	-0.044	0.045
MoSold_12	0.0126	0.025	0.503	0.615	-0.037	0.062
MoSold_2	0.0015	0.025	0.060	0.952	-0.048	0.051
MoSold_3	0.0131	0.022	0.597	0.551	-0.030	0.056
MoSold_4	0.0071	0.021	0.345	0.730	-0.033	0.048
MoSold_5	0.0249	0.020	1.258	0.209	-0.014	0.064
MoSold_6	0.0268	0.020	1.369	0.171	-0.012	0.065
MoSold_7	0.0161	0.020	0.806	0.421	-0.023	0.055
MoSold_8	0.0182	0.022	0.842	0.400	-0.024	0.061
MoSold_9	0.0014	0.024	0.058	0.954	-0.046	0.049
YrSold_2007	0.0109	0.011	1.022	0.307	-0.010	0.032
YrSold_2008	0.0119	0.011	1.091	0.276	-0.009	0.033
YrSold_2009	-0.0056	0.011	-0.525	0.600	-0.027	0.015

YrSold_2010	-0.0088	0.013	-0.676	0.499	-0.034	0.017
SaleType_CWD	0.2604	0.109	2.392	0.017	0.047	0.474
SaleType_Con	0.0343	0.105	0.326	0.745	-0.172	0.241
SaleType_ConLD	0.2449	0.064	3.808	0.000	0.119	0.371
SaleType_ConLI	-0.0736	0.059	-1.256	0.210	-0.189	0.041
SaleType_ConLw	-0.0168	0.057	-0.296	0.767	-0.128	0.095
SaleType_New	0.3264	0.090	3.639	0.000	0.150	0.503
SaleType_Oth	0.0034	0.083	0.041	0.967	-0.159	0.166
SaleType_WD	-0.0236	0.024	-0.985	0.325	-0.071	0.023
SaleCondition_AdjLand	0.1494	0.075	1.995	0.046	0.002	0.296
SaleCondition_Alloca	0.1037	0.053	1.951	0.051	-0.001	0.208
SaleCondition_Family	0.0847	0.036	2.377	0.018	0.015	0.155
SaleCondition_Normal	0.0496	0.016	3.169	0.002	0.019	0.080
SaleCondition_Partial	-0.2566	0.087	-2.947	0.003	-0.428	-0.086

Omnibus:	234.976	**Durbin-Watson:**	2.028
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1483.744
Skew:	-0.895	**Prob(JB):**	0.00
Kurtosis:	8.625	**Cond. No.**	1.32e+16

In [979]:
```
from IPython.display import Image
```

Model Evaluation Metrics for Regression

Metrics can we used for regression problems are

Mean Absolute Error (MAE) is the mean of the absolute value of the errors:

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n} |y_j - \hat{y}_j|$$

Mean Squared Error (MSE) is the mean of the squared errors:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

Root Mean Squared Error (RMSE) is the square root of the mean of the squared errors:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

In [950]:
```
def rmse(predictions, targets):

    differences = predictions - targets                       # the DIFFERENCEs.

    differences_squared = differences ** 2                    # the SQUAREs of ^

    mean_of_differences_squared = differences_squared.mean()  # the MEAN of ^

    rmse_val = np.sqrt(mean_of_differences_squared)           # ROOT of ^

    return rmse_val
```

In [951]:
```
cols = ['Model', 'R-Squared Value', 'Adj.R-Squared Value', 'RMSE']
models_report = pd.DataFrame(columns = cols)
# Predicting the model on test data
predictions1 = model1.predict(x_test)
```

In [952]:
```
tmp1 = pd.Series({'Model': " Base Linear Regression Model",
                  'R-Squared Value' : model1.rsquared,
                  'Adj.R-Squared Value': model1.rsquared_adj,
                  'RMSE': rmse(predictions1, y_test)})

model1_report = models_report.append(tmp1, ignore_index = True)
model1_report
```

Out[952]:

	Model	R-Squared Value	Adj.R-Squared Value	RMSE
0	Base Linear Regression Model	0.958527	0.943989	0.464362

8.3 Building Model with Constant

In [953]:
```
df_constant = sm.add_constant(newdata)
```

In [954]:
```
x_train1,x_test1, y_train1, y_test1 = train_test_split(df_constant, target_log, test_size = 0.30, ran
dom_state=0)
```

In [955]:
```
# Lets build Linear Regression model using statsmodel
import statsmodels.api as sm

# Building Linear Regression model using OLS
model2 = sm.OLS(y_train1, x_train1).fit()
# Note the Swap of X and Y
```

In [956]:
```
# Printing Linear Regression Summary
model2.summary2()
```

Out[956]:

Model:	OLS	Adj. R-squared:	0.944
Dependent Variable:	SalePrice	AIC:	-1683.3983
Date:	2018-07-25 13:25	BIC:	-372.1468
No. Observations:	1022	Log-Likelihood:	1107.7
Df Model:	265	F-statistic:	65.93
Df Residuals:	756	Prob (F-statistic):	0.00
R-squared:	0.959	Scale:	0.0090580

	Coef.	**Std.Err.**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	9.3135	0.2042	45.6027	0.0000	8.9126	9.7144
Id	-0.0241	0.0119	-2.0193	0.0438	-0.0475	-0.0007
LotFrontage	0.0760	0.0513	1.4807	0.1391	-0.0247	0.1767
LotArea	0.5298	0.0790	6.7045	0.0000	0.3747	0.6849
OverallQual	0.3003	0.0500	6.0067	0.0000	0.2022	0.3985
YearBuilt	0.2447	0.0653	3.7445	0.0002	0.1164	0.3729
YearRemodAdd	0.0465	0.0188	2.4764	0.0135	0.0096	0.0834
MasVnrArea	0.0532	0.0470	1.1307	0.2586	-0.0392	0.1455
BsmtFinSF1	0.0963	0.0513	1.8783	0.0607	-0.0043	0.1969
BsmtFinSF2	0.0031	0.0766	0.0400	0.9681	-0.1472	0.1533
BsmtUnfSF	-0.0365	0.0257	-1.4197	0.1561	-0.0871	0.0140
LowQualFinSF	0.0097	0.0397	0.2438	0.8074	-0.0683	0.0877
GrLivArea	0.5945	0.1792	3.3175	0.0010	0.2427	0.9462
BsmtFullBath	0.0877	0.0307	2.8602	0.0044	0.0275	0.1479
BsmtHalfBath	0.0424	0.0266	1.5907	0.1121	-0.0099	0.0947
FullBath	0.1111	0.0359	3.0949	0.0020	0.0406	0.1815

HalfBath	0.0608	0.0241	2.5283	0.0117	0.0136	0.1080
BedroomAbvGr	-0.0683	0.0616	-1.1083	0.2681	-0.1892	0.0526
KitchenAbvGr	-0.1893	0.1153	-1.6420	0.1010	-0.4156	0.0370
TotRmsAbvGrd	0.0414	0.0592	0.6992	0.4847	-0.0748	0.1576
Fireplaces	0.0570	0.0362	1.5713	0.1165	-0.0142	0.1281
GarageYrBlt	-0.0136	0.0376	-0.3629	0.7168	-0.0875	0.0602
GarageCars	0.0141	0.0515	0.2744	0.7839	-0.0870	0.1153
GarageArea	0.1986	0.0630	3.1518	0.0017	0.0749	0.3224
WoodDeckSF	0.0375	0.0103	3.6249	0.0003	0.0172	0.0578
OpenPorchSF	0.0082	0.0127	0.6509	0.5153	-0.0166	0.0331
EnclosedPorch	0.0099	0.0153	0.6441	0.5197	-0.0202	0.0399
3SsnPorch	-0.0051	0.0337	-0.1529	0.8786	-0.0712	0.0609
ScreenPorch	0.0374	0.0155	2.4127	0.0161	0.0070	0.0678
PoolArea	-0.0234	0.1641	-0.1427	0.8866	-0.3455	0.2987
MiscVal	0.0331	0.0915	0.3618	0.7176	-0.1465	0.2127
TotalSF	0.9518	0.2145	4.4377	0.0000	0.5308	1.3729
MSSubClass_160	-0.0687	0.0580	-1.1852	0.2363	-0.1826	0.0451
MSSubClass_180	0.0609	0.0787	0.7738	0.4393	-0.0935	0.2153
MSSubClass_190	-0.0235	0.1538	-0.1526	0.8788	-0.3253	0.2784
MSSubClass_20	-0.0455	0.0675	-0.6735	0.5008	-0.1781	0.0871
MSSubClass_30	-0.0968	0.0718	-1.3484	0.1779	-0.2376	0.0441
MSSubClass_40	0.0152	0.1188	0.1278	0.8983	-0.2181	0.2485
MSSubClass_45	-0.6041	0.1738	-3.4757	0.0005	-0.9454	-0.2629
MSSubClass_50	-0.0974	0.0874	-1.1145	0.2654	-0.2690	0.0742
MSSubClass_60	-0.0696	0.0832	-0.8366	0.4031	-0.2329	0.0937

MSSubClass_70	-0.0448	0.0843	-0.5307	0.5958	-0.2103	0.1208
MSSubClass_75	-0.1197	0.1128	-1.0609	0.2891	-0.3412	0.1018
MSSubClass_80	-0.0783	0.1013	-0.7735	0.4394	-0.2772	0.1205
MSSubClass_85	-0.0145	0.0927	-0.1563	0.8758	-0.1964	0.1674
MSSubClass_90	-0.0375	0.0399	-0.9394	0.3478	-0.1159	0.0409
MSZoning_FV	0.5248	0.0650	8.0775	0.0000	0.3972	0.6523
MSZoning_RH	0.4490	0.0622	7.2236	0.0000	0.3270	0.5710
MSZoning_RL	0.4599	0.0552	8.3372	0.0000	0.3516	0.5682
MSZoning_RM	0.4566	0.0513	8.8946	0.0000	0.3558	0.5574
Street_Pave	0.1125	0.0615	1.8291	0.0678	-0.0082	0.2332
Alley_Pave	0.0556	0.0273	2.0387	0.0418	0.0021	0.1092
LotShape_IR2	0.0042	0.0222	0.1907	0.8488	-0.0394	0.0479
LotShape_IR3	0.0684	0.0484	1.4124	0.1582	-0.0267	0.1634
LotShape_Reg	0.0134	0.0085	1.5800	0.1145	-0.0033	0.0301
LandContour_HLS	-0.0116	0.0306	-0.3793	0.7045	-0.0717	0.0485
LandContour_Low	-0.0320	0.0355	-0.9009	0.3680	-0.1017	0.0377
LandContour_Lvl	-0.0066	0.0215	-0.3060	0.7597	-0.0488	0.0356
Utilities_NoSeWa	-0.2118	0.1375	-1.5404	0.1239	-0.4818	0.0581
LotConfig_CulDSac	0.0272	0.0175	1.5603	0.1191	-0.0070	0.0615
LotConfig_FR2	-0.0169	0.0211	-0.8003	0.4238	-0.0582	0.0245
LotConfig_FR3	-0.1918	0.1508	-1.2719	0.2038	-0.4877	0.1042
LotConfig_Inside	-0.0140	0.0096	-1.4599	0.1447	-0.0329	0.0048
LandSlope_Mod	0.0225	0.0212	1.0648	0.2873	-0.0190	0.0641
LandSlope_Sev	-0.1630	0.0534	-3.0517	0.0024	-0.2679	-0.0582
Neighborhood_Blueste	0.0024	0.0919	0.0264	0.9790	-0.1781	0.1829

Neighborhood_BrDale	-0.0958	0.0651	-1.4716	0.1416	-0.2236	0.0320
Neighborhood_BrkSide	-0.0571	0.0544	-1.0495	0.2943	-0.1638	0.0497
Neighborhood_ClearCr	-0.0558	0.0511	-1.0914	0.2754	-0.1562	0.0446
Neighborhood_CollgCr	-0.1002	0.0414	-2.4209	0.0157	-0.1815	-0.0190
Neighborhood_Crawfor	0.0299	0.0503	0.5944	0.5524	-0.0688	0.1285
Neighborhood_Edwards	-0.1658	0.0461	-3.5971	0.0003	-0.2563	-0.0753
Neighborhood_Gilbert	-0.0973	0.0427	-2.2808	0.0228	-0.1811	-0.0136
Neighborhood_IDOTRR	-0.1833	0.0631	-2.9032	0.0038	-0.3073	-0.0594
Neighborhood_MeadowV	-0.1905	0.0649	-2.9345	0.0034	-0.3179	-0.0631
Neighborhood_Mitchel	-0.1735	0.0474	-3.6605	0.0003	-0.2665	-0.0804
Neighborhood_NAmes	-0.1339	0.0453	-2.9592	0.0032	-0.2228	-0.0451
Neighborhood_NPkVill	-0.0465	0.0668	-0.6955	0.4869	-0.1776	0.0847
Neighborhood_NWAmes	-0.1287	0.0460	-2.7982	0.0053	-0.2190	-0.0384
Neighborhood_NoRidge	0.0117	0.0465	0.2523	0.8008	-0.0795	0.1029
Neighborhood_NridgHt	-0.0413	0.0431	-0.9578	0.3385	-0.1260	0.0434
Neighborhood_OldTown	-0.1538	0.0557	-2.7594	0.0059	-0.2632	-0.0444
Neighborhood_SWISU	-0.0748	0.0541	-1.3837	0.1669	-0.1810	0.0313
Neighborhood_Sawyer	-0.1217	0.0462	-2.6314	0.0087	-0.2125	-0.0309
Neighborhood_SawyerW	-0.1013	0.0442	-2.2909	0.0222	-0.1882	-0.0145
Neighborhood_Somerst	-0.0835	0.0486	-1.7175	0.0863	-0.1790	0.0119
Neighborhood_StoneBr	0.0169	0.0477	0.3555	0.7223	-0.0766	0.1105
Neighborhood_Timber	-0.0883	0.0456	-1.9369	0.0531	-0.1779	0.0012
Neighborhood_Veenker	-0.0656	0.0603	-1.0871	0.2774	-0.1840	0.0529
Condition1_Feedr	-0.0134	0.0311	-0.4292	0.6679	-0.0744	0.0477
Condition1_Norm	0.0591	0.0261	2.2640	0.0239	0.0078	0.1103

Condition1_PosA	0.0274	0.0528	0.5180	0.6046	-0.0763	0.1310
Condition1_PosN	0.0763	0.0409	1.8644	0.0627	-0.0040	0.1566
Condition1_RRAe	-0.0767	0.0465	-1.6501	0.0993	-0.1680	0.0146
Condition1_RRAn	0.0382	0.0394	0.9713	0.3317	-0.0391	0.1155
Condition1_RRNe	-0.0059	0.0789	-0.0748	0.9404	-0.1608	0.1490
Condition1_RRNn	0.0611	0.0688	0.8873	0.3752	-0.0741	0.1962
Condition2_Feedr	0.0727	0.1577	0.4609	0.6450	-0.2369	0.3822
Condition2_Norm	-0.0028	0.1374	-0.0205	0.9837	-0.2725	0.2668
Condition2_PosA	0.0000	0.0000	0.0294	0.9765	-0.0000	0.0000
Condition2_PosN	-1.3327	0.1762	-7.5630	0.0000	-1.6786	-0.9868
Condition2_RRAe	-0.4817	0.2619	-1.8393	0.0663	-0.9959	0.0324
Condition2_RRAn	-0.0000	0.0000	-3.2980	0.0010	-0.0000	-0.0000
Condition2_RRNn	0.0000	0.0000	1.0685	0.2856	-0.0000	0.0000
BldgType_2fmCon	-0.1050	0.1304	-0.8054	0.4209	-0.3610	0.1510
BldgType_Duplex	-0.0375	0.0399	-0.9394	0.3478	-0.1159	0.0409
BldgType_Twnhs	-0.0380	0.0725	-0.5245	0.6001	-0.1804	0.1043
BldgType_TwnhsE	-0.0459	0.0680	-0.6756	0.4995	-0.1794	0.0876
HouseStyle_1.5Unf	0.5219	0.1644	3.1751	0.0016	0.1992	0.8446
HouseStyle_1Story	-0.0420	0.0534	-0.7867	0.4317	-0.1468	0.0628
HouseStyle_2.5Fin	-0.0071	0.0904	-0.0785	0.9375	-0.1845	0.1704
HouseStyle_2.5Unf	0.0942	0.0895	1.0529	0.2927	-0.0815	0.2699
HouseStyle_2Story	-0.0444	0.0475	-0.9361	0.3495	-0.1376	0.0488
HouseStyle_SFoyer	-0.0613	0.0739	-0.8295	0.4071	-0.2063	0.0837
HouseStyle_SLvl	-0.0347	0.0780	-0.4445	0.6568	-0.1879	0.1185
OverallCond_2	0.9352	0.0715	13.0873	0.0000	0.7949	1.0754

OverallCond_3	0.8202	0.0454	18.0768	0.0000	0.7311	0.9092
OverallCond_4	0.8922	0.0415	21.5101	0.0000	0.8108	0.9736
OverallCond_5	0.9316	0.0397	23.4768	0.0000	0.8537	1.0095
OverallCond_6	0.9655	0.0404	23.9063	0.0000	0.8862	1.0447
OverallCond_7	1.0111	0.0405	24.9477	0.0000	0.9315	1.0906
OverallCond_8	1.0204	0.0424	24.0514	0.0000	0.9371	1.1037
OverallCond_9	1.0274	0.0500	20.5540	0.0000	0.9293	1.1256
RoofStyle_Gable	0.0340	0.0834	0.4081	0.6833	-0.1297	0.1978
RoofStyle_Gambrel	-0.0215	0.0946	-0.2278	0.8198	-0.2072	0.1641
RoofStyle_Hip	0.0516	0.0839	0.6154	0.5385	-0.1130	0.2163
RoofStyle_Mansard	0.0724	0.0985	0.7351	0.4625	-0.1210	0.2658
RoofStyle_Shed	0.2745	0.1813	1.5143	0.1304	-0.0813	0.6303
RoofMatl_CompShg	1.5053	0.0535	28.1131	0.0000	1.4002	1.6104
RoofMatl_Membran	-0.0000	0.0000	-0.1262	0.8996	-0.0000	0.0000
RoofMatl_Metal	1.7394	0.1199	14.5020	0.0000	1.5039	1.9749
RoofMatl_Roll	1.4877	0.1140	13.0528	0.0000	1.2640	1.7115
RoofMatl_Tar&Grv	1.5305	0.0674	22.7170	0.0000	1.3983	1.6628
RoofMatl_WdShake	1.4855	0.0789	18.8377	0.0000	1.3307	1.6403
RoofMatl_WdShngl	1.5650	0.0691	22.6440	0.0000	1.4293	1.7007
Exterior1st_AsphShn	-0.0509	0.0618	-0.8231	0.4107	-0.1722	0.0705
Exterior1st_BrkComm	-0.5255	0.1628	-3.2281	0.0013	-0.8451	-0.2059
Exterior1st_BrkFace	0.0562	0.0759	0.7401	0.4595	-0.0928	0.2051
Exterior1st_CBlock	-0.0517	0.0662	-0.7809	0.4351	-0.1816	0.0783
Exterior1st_CemntBd	-0.0911	0.1071	-0.8505	0.3953	-0.3013	0.1192
Exterior1st_HdBoard	-0.0013	0.0768	-0.0169	0.9865	-0.1520	0.1494

Exterior1st_ImStucc	-0.0047	0.1311	-0.0359	0.9714	-0.2622	0.2528
Exterior1st_MetalSd	-0.0089	0.0920	-0.0970	0.9227	-0.1896	0.1718
Exterior1st_Plywood	0.0073	0.0767	0.0946	0.9247	-0.1433	0.1578
Exterior1st_Stone	-0.0266	0.1232	-0.2155	0.8294	-0.2684	0.2153
Exterior1st_Stucco	0.0184	0.0830	0.2223	0.8242	-0.1444	0.1813
Exterior1st_VinylSd	0.0045	0.0743	0.0612	0.9512	-0.1413	0.1503
Exterior1st_Wd Sdng	-0.0573	0.0735	-0.7799	0.4357	-0.2017	0.0870
Exterior1st_WdShing	-0.0096	0.0786	-0.1227	0.9024	-0.1639	0.1446
Exterior2nd_AsphShn	-0.0509	0.0618	-0.8231	0.4107	-0.1722	0.0705
Exterior2nd_Brk Cmn	0.0588	0.1084	0.5424	0.5877	-0.1541	0.2717
Exterior2nd_BrkFace	-0.0236	0.0742	-0.3185	0.7502	-0.1693	0.1220
Exterior2nd_CBlock	-0.0517	0.0662	-0.7809	0.4351	-0.1816	0.0783
Exterior2nd_CmentBd	0.0942	0.1028	0.9165	0.3597	-0.1076	0.2961
Exterior2nd_HdBoard	-0.0233	0.0704	-0.3305	0.7411	-0.1614	0.1149
Exterior2nd_ImStucc	-0.0228	0.0778	-0.2931	0.7696	-0.1756	0.1300
Exterior2nd_MetalSd	-0.0025	0.0870	-0.0293	0.9767	-0.1733	0.1683
Exterior2nd_Other	-0.0548	0.1254	-0.4367	0.6624	-0.3009	0.1914
Exterior2nd_Plywood	-0.0339	0.0697	-0.4869	0.6265	-0.1707	0.1028
Exterior2nd_Stone	0.0260	0.0962	0.2699	0.7873	-0.1629	0.2148
Exterior2nd_Stucco	0.0244	0.0760	0.3212	0.7482	-0.1247	0.1735
Exterior2nd_VinylSd	-0.0136	0.0686	-0.1978	0.8432	-0.1483	0.1212
Exterior2nd_Wd Sdng	0.0311	0.0673	0.4627	0.6437	-0.1009	0.1632
Exterior2nd_Wd Shng	-0.0206	0.0699	-0.2951	0.7680	-0.1579	0.1166
MasVnrType_BrkFace	0.0372	0.0293	1.2679	0.2052	-0.0204	0.0948
MasVnrType_None	0.0669	0.0380	1.7628	0.0783	-0.0076	0.1414

MasVnrType_Stone	0.0693	0.0315	2.2008	0.0280	0.0075	0.1312
ExterQual_Fa	0.0209	0.0668	0.3123	0.7549	-0.1103	0.1521
ExterQual_Gd	-0.0172	0.0255	-0.6729	0.5012	-0.0672	0.0329
ExterQual_TA	-0.0304	0.0285	-1.0682	0.2858	-0.0863	0.0255
ExterCond_Fa	-0.1497	0.0872	-1.7155	0.0867	-0.3210	0.0216
ExterCond_Gd	-0.1222	0.0829	-1.4731	0.1412	-0.2850	0.0406
ExterCond_Po	-0.2376	0.1671	-1.4224	0.1553	-0.5656	0.0903
ExterCond_TA	-0.1128	0.0831	-1.3578	0.1749	-0.2759	0.0503
Foundation_CBlock	0.0246	0.0189	1.3038	0.1927	-0.0124	0.0616
Foundation_PConc	0.0259	0.0196	1.3234	0.1861	-0.0125	0.0644
Foundation_Slab	-0.0101	0.0685	-0.1470	0.8832	-0.1445	0.1244
Foundation_Stone	0.1801	0.0577	3.1199	0.0019	0.0668	0.2934
Foundation_Wood	0.1163	0.1093	1.0637	0.2878	-0.0983	0.3309
BsmtQual_Fa	0.0250	0.0371	0.6729	0.5012	-0.0479	0.0979
BsmtQual_Gd	-0.0452	0.0174	-2.5943	0.0097	-0.0794	-0.0110
BsmtQual_TA	-0.0345	0.0215	-1.6025	0.1095	-0.0768	0.0078
BsmtCond_Gd	0.0188	0.0285	0.6587	0.5103	-0.0372	0.0748
BsmtCond_Po	0.8235	0.1201	6.8585	0.0000	0.5878	1.0593
BsmtCond_TA	0.0300	0.0235	1.2790	0.2013	-0.0161	0.0761
BsmtExposure_Gd	0.0520	0.0165	3.1463	0.0017	0.0196	0.0844
BsmtExposure_Mn	0.0070	0.0161	0.4348	0.6638	-0.0246	0.0386
BsmtExposure_No	-0.0014	0.0115	-0.1243	0.9011	-0.0240	0.0211
BsmtFinType1_BLQ	-0.0012	0.0146	-0.0837	0.9333	-0.0298	0.0274
BsmtFinType1_GLQ	0.0076	0.0135	0.5638	0.5731	-0.0189	0.0341
BsmtFinType1_LwQ	-0.0441	0.0203	-2.1764	0.0298	-0.0839	-0.0043

BsmtFinType1_Rec	-0.0209	0.0159	-1.3152	0.1889	-0.0522	0.0103
BsmtFinType1_Unf	0.0220	0.0362	0.6068	0.5442	-0.0491	0.0930
BsmtFinType2_BLQ	-0.0865	0.0363	-2.3814	0.0175	-0.1578	-0.0152
BsmtFinType2_GLQ	-0.0537	0.0494	-1.0880	0.2770	-0.1507	0.0432
BsmtFinType2_LwQ	-0.0552	0.0362	-1.5258	0.1275	-0.1262	0.0158
BsmtFinType2_Rec	-0.0464	0.0359	-1.2919	0.1968	-0.1168	0.0241
BsmtFinType2_Unf	-0.0575	0.0634	-0.9071	0.3646	-0.1819	0.0669
Heating_GasA	-0.0158	0.1237	-0.1278	0.8984	-0.2586	0.2270
Heating_GasW	0.0089	0.1307	0.0681	0.9457	-0.2477	0.2655
Heating_Grav	-0.1993	0.1395	-1.4282	0.1536	-0.4732	0.0746
Heating_OthW	0.0122	0.1734	0.0706	0.9438	-0.3281	0.3526
Heating_Wall	0.0378	0.1632	0.2318	0.8168	-0.2826	0.3583
HeatingQC_Fa	-0.0572	0.0254	-2.2495	0.0248	-0.1071	-0.0073
HeatingQC_Gd	-0.0190	0.0107	-1.7809	0.0753	-0.0400	0.0019
HeatingQC_Po	0.0570	0.1203	0.4736	0.6359	-0.1791	0.2931
HeatingQC_TA	-0.0331	0.0110	-3.0059	0.0027	-0.0547	-0.0115
CentralAir_Y	0.0507	0.0214	2.3716	0.0180	0.0087	0.0926
Electrical_FuseF	-0.0291	0.0345	-0.8413	0.4004	-0.0969	0.0388
Electrical_FuseP	-0.2488	0.1054	-2.3611	0.0185	-0.4557	-0.0419
Electrical_Mix	-0.8865	0.2067	-4.2889	0.0000	-1.2922	-0.4807
Electrical_SBrkr	-0.0397	0.0157	-2.5240	0.0118	-0.0705	-0.0088
KitchenQual_Fa	-0.0768	0.0338	-2.2703	0.0235	-0.1431	-0.0104
KitchenQual_Gd	-0.0888	0.0183	-4.8578	0.0000	-0.1247	-0.0529
KitchenQual_TA	-0.0847	0.0209	-4.0550	0.0001	-0.1258	-0.0437
Functional_Maj2	-0.3974	0.1141	-3.4838	0.0005	-0.6213	-0.1735

Functional_Min1	-0.0523	0.0456	-1.1473	0.2516	-0.1419	0.0372
Functional_Min2	-0.0494	0.0453	-1.0912	0.2755	-0.1383	0.0395
Functional_Mod	-0.0713	0.0549	-1.2978	0.1947	-0.1791	0.0365
Functional_Sev	-0.0000	0.0000	-0.8028	0.4223	-0.0000	0.0000
Functional_Typ	0.0191	0.0398	0.4790	0.6321	-0.0590	0.0971
FireplaceQu_Fa	-0.0222	0.0260	-0.8505	0.3953	-0.0733	0.0290
FireplaceQu_Gd	-0.0048	0.0151	-0.3167	0.7516	-0.0344	0.0248
FireplaceQu_Po	0.0069	0.0304	0.2279	0.8198	-0.0527	0.0665
FireplaceQu_TA	-0.0004	0.0164	-0.0256	0.9796	-0.0326	0.0317
GarageType_Attchd	0.1101	0.0478	2.3020	0.0216	0.0162	0.2041
GarageType_Basment	0.0812	0.0598	1.3575	0.1750	-0.0362	0.1986
GarageType_BuiltIn	0.1352	0.0500	2.7033	0.0070	0.0370	0.2334
GarageType_CarPort	0.0434	0.0786	0.5526	0.5807	-0.1109	0.1977
GarageType_Detchd	0.1240	0.0474	2.6146	0.0091	0.0309	0.2171
GarageFinish_RFn	0.0067	0.0103	0.6524	0.5143	-0.0135	0.0269
GarageFinish_Unf	-0.0016	0.0127	-0.1272	0.8988	-0.0265	0.0233
GarageQual_Fa	-0.4447	0.1421	-3.1287	0.0018	-0.7238	-0.1657
GarageQual_Gd	-0.3702	0.1461	-2.5340	0.0115	-0.6571	-0.0834
GarageQual_Po	-0.1790	0.1972	-0.9078	0.3643	-0.5661	0.2081
GarageQual_TA	-0.4182	0.1397	-2.9939	0.0028	-0.6924	-0.1440
GarageCond_Fa	0.2417	0.1509	1.6015	0.1097	-0.0546	0.5380
GarageCond_Gd	0.2513	0.1580	1.5904	0.1122	-0.0589	0.5616
GarageCond_Po	0.1645	0.1766	0.9315	0.3519	-0.1822	0.5111
GarageCond_TA	0.3096	0.1489	2.0790	0.0380	0.0173	0.6020
PavedDrive_P	-0.0710	0.0335	-2.1187	0.0344	-0.1368	-0.0052

PavedDrive_Y	-0.0265	0.0185	-1.4309	0.1529	-0.0628	0.0099
PoolQC_Fa	-0.0000	0.0000	-0.7364	0.4617	-0.0000	0.0000
PoolQC_Gd	0.2951	0.1832	1.6108	0.1076	-0.0645	0.6547
Fence_GdWo	-0.0402	0.0189	-2.1256	0.0339	-0.0772	-0.0031
Fence_MnPrv	0.0016	0.0123	0.1320	0.8950	-0.0225	0.0257
Fence_MnWw	-0.0612	0.0422	-1.4511	0.1472	-0.1440	0.0216
MiscFeature_Othr	-0.0999	0.1136	-0.8796	0.3794	-0.3230	0.1231
MiscFeature_Shed	-0.0070	0.0638	-0.1099	0.9125	-0.1323	0.1182
MiscFeature_TenC	0.0000	0.0000	0.1569	0.8753	-0.0000	0.0000
MoSold_10	-0.0052	0.0229	-0.2279	0.8198	-0.0503	0.0398
MoSold_11	0.0008	0.0228	0.0348	0.9723	-0.0439	0.0455
MoSold_12	0.0126	0.0250	0.5032	0.6150	-0.0366	0.0618
MoSold_2	0.0015	0.0251	0.0598	0.9524	-0.0477	0.0507
MoSold_3	0.0131	0.0220	0.5970	0.5507	-0.0301	0.0564
MoSold_4	0.0071	0.0206	0.3446	0.7305	-0.0334	0.0476
MoSold_5	0.0249	0.0198	1.2576	0.2089	-0.0140	0.0639
MoSold_6	0.0268	0.0196	1.3687	0.1715	-0.0116	0.0652
MoSold_7	0.0161	0.0200	0.8057	0.4207	-0.0232	0.0554
MoSold_8	0.0182	0.0216	0.8422	0.4000	-0.0242	0.0606
MoSold_9	0.0014	0.0243	0.0579	0.9538	-0.0463	0.0491
YrSold_2007	0.0109	0.0107	1.0224	0.3069	-0.0100	0.0318
YrSold_2008	0.0119	0.0109	1.0907	0.2758	-0.0095	0.0332
YrSold_2009	-0.0056	0.0107	-0.5250	0.5997	-0.0267	0.0154
YrSold_2010	-0.0088	0.0130	-0.6759	0.4993	-0.0342	0.0167
SaleType_CWD	0.2604	0.1088	2.3924	0.0170	0.0467	0.4740

SaleType_Con	0.0343	0.1053	0.3258	0.7447	-0.1725	0.2411
SaleType_ConLD	0.2449	0.0643	3.8085	0.0002	0.1187	0.3711
SaleType_ConLI	-0.0736	0.0586	-1.2560	0.2095	-0.1885	0.0414
SaleType_ConLw	-0.0168	0.0567	-0.2960	0.7673	-0.1281	0.0945
SaleType_New	0.3264	0.0897	3.6386	0.0003	0.1503	0.5025
SaleType_Oth	0.0034	0.0828	0.0414	0.9670	-0.1592	0.1660
SaleType_WD	-0.0236	0.0239	-0.9851	0.3249	-0.0705	0.0234
SaleCondition_AdjLand	0.1494	0.0749	1.9954	0.0464	0.0024	0.2963
SaleCondition_Alloca	0.1037	0.0531	1.9512	0.0514	-0.0006	0.2080
SaleCondition_Family	0.0847	0.0356	2.3771	0.0177	0.0148	0.1547
SaleCondition_Normal	0.0496	0.0156	3.1692	0.0016	0.0189	0.0803
SaleCondition_Partial	-0.2566	0.0871	-2.9469	0.0033	-0.4275	-0.0857

Omnibus:	234.976	Durbin-Watson:	2.028
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1483.744
Skew:	-0.895	Prob(JB):	0.000
Kurtosis:	8.625	Condition No.:	10276557703819846

In [957]:
```
# Predicting the model on test data
predictions2 = model2.predict(x_test1)

tmp2 = pd.Series({'Model': " Linear Regression Model with Constant",
                  'R-Squared Value' : model2.rsquared,
                  'Adj.R-Squared Value': model2.rsquared_adj,
                  'RMSE': rmse(predictions2, y_test1)})

model2_report = models_report.append(tmp2, ignore_index = True)
model2_report
```

Out[957]:

	Model	R-Squared Value	Adj.R-Squared Value	RMSE
0	Linear Regression Model with Constant	0.958527	0.943989	0.167912

Multicollinearity

In regression, "multicollinearity" refers to predictors that are correlated with other predictors. Multicollinearity occurs when your model includes multiple factors that are correlated not just to your response variable, but also to each other.

$$VIF = \frac{1}{1 - R^2}$$

A rule of thumb for interpreting the variance inflation factor:

- 1 = not correlated.
- Between 1 and 5 = moderately correlated.
- Greater than 5 = highly correlated.

8.4 Calculating Variance Inflation Factor(VIF)

In [958]:

```
print ("\nVariance Inflation Factor")
cnames = x_train1.columns
for i in np.arange(0,len(cnames)):
    xvars = list(cnames)
    yvar = xvars.pop(i)
    mod = sm.OLS(x_train1[yvar],(x_train1[xvars]))
    res = mod.fit()
    vif = 1/(1-res.rsquared)
    print (yvar,round(vif,3))
```

Variance Inflation Factor

```
C:\Users\computer\Anaconda3\lib\site-packages\statsmodels\regression\linear_model.py:1386: RuntimeWa
rning: divide by zero encountered in double_scalars
  return 1 - self.ssr/self.centered_tss
```

```
const 0.0
Id 1.342
LotFrontage 3.903
LotArea 7.1
OverallQual 6.414
YearBuilt 23.111
YearRemodAdd 4.651
MasVnrArea 31.586
BsmtFinSF1 35.577
BsmtFinSF2 43.956
BsmtUnfSF 4.086
LowQualFinSF 2.479
GrLivArea 50.169
BsmtFullBath 3.124
BsmtHalfBath 1.723
FullBath 4.734
HalfBath 4.046
BedroomAbvGr 4.31
KitchenAbvGr 6.251
TotRmsAbvGrd 7.162
Fireplaces 6.65
GarageYrBlt 7.786
GarageCars 10.078
GarageArea 9.708
WoodDeckSF 1.797
OpenPorchSF 2.119
EnclosedPorch 2.075
3SsnPorch 1.451
ScreenPorch 1.459
PoolArea 8.353
MiscVal 18.066
TotalSF 38.985
MSSubClass_160 16.981
MSSubClass_180 2.722
MSSubClass_190 53.698
MSSubClass_20 119.798
MSSubClass_30 25.49
MSSubClass_40 4.663
MSSubClass_45 36.296
MSSubClass_50 75.425
MSSubClass_60 129.749
MSSubClass_70 33.777
MSSubClass_75 13.917
```

```
MSSubClass_80 39.333
MSSubClass_85 12.166
MSSubClass_90 inf

C:\Users\computer\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: RuntimeWarning: divide by zer
o encountered in double_scalars
```

MSZoning_FV 17.048
MSZoning_RH 5.889
MSZoning_RL 54.661
MSZoning_RM 35.577
Street_Pave 2.077
Alley_Pave 2.547
LotShape_IR2 1.539
LotShape_IR3 1.544
LotShape_Reg 1.903
LandContour_HLS 2.82
LandContour_Low 3.659
LandContour_Lvl 4.557
Utilities_NoSeWa 2.086
LotConfig_CulDSac 2.134
LotConfig_FR2 1.472
LotConfig_FR3 2.507
LotConfig_Inside 2.072
LandSlope_Mod 2.353
LandSlope_Sev 3.121
Neighborhood_Blueste 1.862
Neighborhood_BrDale 4.633
Neighborhood_BrkSide 12.846
Neighborhood_ClearCr 6.766
Neighborhood_CollgCr 17.828
Neighborhood_Crawfor 9.428
Neighborhood_Edwards 15.5
Neighborhood_Gilbert 10.459
Neighborhood_IDOTRR 10.317
Neighborhood_MeadowV 5.516
Neighborhood_Mitchel 8.612
Neighborhood_NAmes 29.1
Neighborhood_NPkVill 3.426
Neighborhood_NWAmes 11.317
Neighborhood_NoRidge 6.939
Neighborhood_NridgHt 9.957
Neighborhood_OldTown 23.253
Neighborhood_SWISU 5.713
Neighborhood_Sawyer 13.541
Neighborhood_SawyerW 9.293
Neighborhood_Somerst 13.828
Neighborhood_StoneBr 3.95
Neighborhood_Timber 6.686
Neighborhood_Veenker 2.397

```
Condition1_Feedr 5.083
Condition1_Norm 8.41
Condition1_PosA 1.836
Condition1_PosN 2.551
Condition1_RRAe 2.364
Condition1_RRAn 2.695
Condition1_RRNe 1.372
Condition1_RRNn 2.085
Condition2_Feedr 10.937
Condition2_Norm 14.482
Condition2_PosA nan
```

```
C:\Users\computer\Anaconda3\lib\site-packages\statsmodels\regression\linear_model.py:1386: RuntimeWa
rning: invalid value encountered in double_scalars
  return 1 - self.ssr/self.centered_tss
```

Condition2_PosN 3.425
Condition2_RRAe 7.566
Condition2_RRAn nan
Condition2_RRNn nan
BldgType_2fmCon 40.404
BldgType_Duplex inf
BldgType_Twnhs 16.356
BldgType_TwnhsE 37.644
HouseStyle_1.5Unf 35.376
HouseStyle_1Story 80.4
HouseStyle_2.5Fin 4.488
HouseStyle_2.5Unf 7.016
HouseStyle_2Story 54.865
HouseStyle_SFoyer 12.395
HouseStyle_SLvl 25.835
OverallCond_2 inf
OverallCond_3 inf
OverallCond_4 inf
OverallCond_5 inf
OverallCond_6 inf
OverallCond_7 inf
OverallCond_8 inf
OverallCond_9 inf
RoofStyle_Gable 134.77
RoofStyle_Gambrel 7.836
RoofStyle_Hip 124.463
RoofStyle_Mansard 6.389
RoofStyle_Shed 7.24
RoofMatl_CompShg inf
RoofMatl_Membran nan
RoofMatl_Metal inf
RoofMatl_Roll inf
RoofMatl_Tar&Grv inf
RoofMatl_WdShake inf
RoofMatl_WdShngl inf
Exterior1st_AsphShn inf
Exterior1st_BrkComm 2.923
Exterior1st_BrkFace 21.489
Exterior1st_CBlock inf
Exterior1st_CemntBd 47.505
Exterior1st_HdBoard 84.65
Exterior1st_ImStucc 1.897
Exterior1st_MetalSd 111.636

```
Exterior1st_Plywood 51.111
Exterior1st_Stone 3.344
Exterior1st_Stucco 12.701
Exterior1st_VinylSd 142.716
Exterior1st_Wd Sdng 75.565
Exterior1st_WdShing 14.027
Exterior2nd_AsphShn inf
Exterior2nd_Brk Cmn 5.173
Exterior2nd_BrkFace 10.75
Exterior2nd_CBlock inf
Exterior2nd_CmentBd 43.785
Exterior2nd_HdBoard 66.434
Exterior2nd_ImStucc 5.308
Exterior2nd_MetalSd 98.531
Exterior2nd_Other 1.733
Exterior2nd_Plywood 53.44
Exterior2nd_Stone 4.072
Exterior2nd_Stucco 10.035
Exterior2nd_VinylSd 120.205
Exterior2nd_Wd Sdng 61.065
Exterior2nd_Wd Shng 15.201
MasVnrType_BrkFace 20.183
MasVnrType_None 39.136
MasVnrType_Stone 9.62
ExterQual_Fa 5.847
ExterQual_Gd 16.286
ExterQual_TA 21.404
ExterCond_Fa 20.496
ExterCond_Gd 70.338
ExterCond_Po 3.079
ExterCond_TA 87.024
Foundation_CBlock 9.836
Foundation_PConc 10.693
Foundation_Slab 9.66
Foundation_Stone 1.83
Foundation_Wood 1.318
BsmtQual_Fa 3.423
BsmtQual_Gd 8.344
BsmtQual_TA 12.98
BsmtCond_Gd 3.95
BsmtCond_Po inf
BsmtCond_TA 5.73
BsmtExposure_Gd 2.474
```

```
BsmtExposure_Mn 2.226
BsmtExposure_No 3.347
BsmtFinType1_BLQ 2.261
BsmtFinType1_GLQ 4.188
BsmtFinType1_LwQ 2.155
BsmtFinType1_Rec 2.413
BsmtFinType1_Unf 30.851
BsmtFinType2_BLQ 3.964
BsmtFinType2_GLQ 2.401
BsmtFinType2_LwQ 5.286
BsmtFinType2_Rec 5.073
BsmtFinType2_Unf 56.432
Heating_GasA 36.344
Heating_GasW 22.362
Heating_Grav 12.82
Heating_OthW 3.316
Heating_Wall 5.872
HeatingQC_Fa 2.277
HeatingQC_Gd 1.837
HeatingQC_Po 1.595
HeatingQC_TA 2.878
CentralAir_Y 3.287
Electrical_FuseF 2.457
Electrical_FuseP 2.448
Electrical_Mix inf
Electrical_SBrkr 2.306
KitchenQual_Fa 3.557
KitchenQual_Gd 8.995
KitchenQual_TA 12.313
Functional_Maj2 2.868
Functional_Min1 4.506
Functional_Min2 6.161
Functional_Mod 3.622
Functional_Sev nan
Functional_Typ 11.688
FireplaceQu_Fa 1.897
FireplaceQu_Gd 4.81
FireplaceQu_Po 1.603
FireplaceQu_TA 5.261
GarageType_Attchd 62.148
GarageType_Basment 5.067
GarageType_BuiltIn 16.807
GarageType_CarPort 3.394
```

```
GarageType_Detchd 49.202
GarageFinish_RFn 2.5
GarageFinish_Unf 4.379
GarageQual_Fa 75.396
GarageQual_Gd 21.025
GarageQual_Po 12.838
GarageQual_TA 199.525
GarageCond_Fa 68.481
GarageCond_Gd 16.445
GarageCond_Po 20.534
GarageCond_TA 210.971
PavedDrive_P 2.192
PavedDrive_Y 3.073
PoolQC_Fa nan
PoolQC_Gd 7.395
Fence_GdWo 1.405
Fence_MnPrv 1.515
Fence_MnWw 1.366
MiscFeature_Othr 2.845
MiscFeature_Shed 17.275
MiscFeature_TenC nan
MoSold_10 3.435
MoSold_11 3.229
MoSold_12 2.597
MoSold_2 2.408
MoSold_3 3.628
MoSold_4 4.501
MoSold_5 5.249
MoSold_6 6.3
MoSold_7 6.082
MoSold_8 3.578
MoSold_9 2.684
YrSold_2007 2.21
YrSold_2008 2.212
YrSold_2009 2.269
YrSold_2010 2.159
SaleType_CWD 1.306
SaleType_Con 1.224
SaleType_ConLD 2.723
SaleType_ConLI 1.509
SaleType_ConLw 1.766
SaleType_New 65.512
SaleType_Oth 1.512
```

```
SaleType_WD 6.931
SaleCondition_AdjLand 1.851
SaleCondition_Alloca 2.167
SaleCondition_Family 1.527
SaleCondition_Normal 3.92
SaleCondition_Partial 63.13
```

Removing variable has threshold value of VIF above 100

Removing variable having VIF above 100

In [959]:
```
vif_100 = ['MSSubClass_20','MSSubClass_60','RoofStyle_Gable','RoofStyle_Hip','RoofMatl_CompShg','Exte
rior1st_MetalSd','Exterior1st_VinylSd','Exterior2nd_VinylSd','GarageQual_TA','GarageCond_TA']
# custom function to remove variables having higer VIF

to_keep = [x for x in x_train1 if x not in vif_100]
# print(to_keep)
x_train2 = x_train1[to_keep]
x_train2.head()
```

Out[959]:

	const	Id	LotFrontage	LotArea	OverallQual	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	BsmtFinS
64	1.0	-0.456134	0.000000	0.006840	0.100076	0.186465	0.218904	0.572034	0.275153	-0.089825
682	1.0	-0.032557	0.000000	-0.223669	-0.011035	0.179219	0.202237	-0.288945	0.310472	-0.089825
960	1.0	0.157985	-0.103554	-0.044634	-0.122146	-0.096144	0.385571	-0.288945	0.268223	-0.089825
1384	1.0	0.448595	-0.036201	0.000150	-0.011035	-0.233825	-0.581096	-0.288945	0.126559	-0.089825
1100	1.0	0.253941	-0.036201	-0.014654	-0.455479	-0.371506	-0.581096	-0.288945	0.167111	-0.089825

5 rows × 268 columns

8.4.1 Building Model after removing VIF above 100

In [960]:
```python
# Lets build Linear Regression model using statsmodel
import statsmodels.api as sm

# Building Linear Regression model using OLS

model3 = sm.OLS(y_train1,x_train2).fit()
# Note the Swap of X and Y
# Printing Linear Regression Summary
model3.summary()
```

Out[960]:

OLS Regression Results

Dep. Variable:	SalePrice	**R-squared:**	0.958
Model:	OLS	**Adj. R-squared:**	0.943
Method:	Least Squares	**F-statistic:**	67.46
Date:	Wed, 25 Jul 2018	**Prob (F-statistic):**	0.00
Time:	13:26:20	**Log-Likelihood:**	1096.2
No. Observations:	1022	**AIC:**	-1678.
Df Residuals:	765	**BIC:**	-411.4
Df Model:	256		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	10.6594	0.212	50.311	0.000	10.243	11.075
Id	-0.0224	0.012	-1.876	0.061	-0.046	0.001
LotFrontage	0.0970	0.051	1.896	0.058	-0.003	0.197
LotArea	0.5045	0.076	6.651	0.000	0.356	0.653
OverallQual	0.3160	0.050	6.380	0.000	0.219	0.413
YearBuilt	0.2287	0.065	3.538	0.000	0.102	0.356
YearRemodAdd	0.0462	0.019	2.490	0.013	0.010	0.083
MasVnrArea	0.0637	0.047	1.357	0.175	-0.028	0.156
BsmtFinSF1	0.1080	0.051	2.111	0.035	0.008	0.209
BsmtFinSF2	-0.0072	0.076	-0.094	0.925	-0.157	0.142
BsmtUnfSF	-0.0387	0.026	-1.503	0.133	-0.089	0.012
LowQualFinSF	0.0269	0.040	0.679	0.498	-0.051	0.105
GrLivArea	0.6034	0.179	3.374	0.001	0.252	0.955

BsmtFullBath	0.0877	0.031	2.866	0.004	0.028	0.148
BsmtHalfBath	0.0396	0.027	1.488	0.137	-0.013	0.092
FullBath	0.1045	0.036	2.910	0.004	0.034	0.175
HalfBath	0.0574	0.024	2.427	0.015	0.011	0.104
BedroomAbvGr	-0.0771	0.061	-1.255	0.210	-0.198	0.044
KitchenAbvGr	-0.1812	0.112	-1.615	0.107	-0.401	0.039
TotRmsAbvGrd	0.0692	0.059	1.172	0.242	-0.047	0.185
Fireplaces	0.0478	0.036	1.322	0.186	-0.023	0.119
GarageYrBlt	0.0065	0.037	0.177	0.859	-0.065	0.078
GarageCars	0.0095	0.051	0.185	0.853	-0.091	0.110
GarageArea	0.1511	0.061	2.460	0.014	0.031	0.272
WoodDeckSF	0.0345	0.010	3.338	0.001	0.014	0.055
OpenPorchSF	0.0080	0.013	0.634	0.526	-0.017	0.033
EnclosedPorch	0.0095	0.015	0.625	0.532	-0.020	0.039
3SsnPorch	-0.0117	0.034	-0.347	0.729	-0.078	0.054
ScreenPorch	0.0445	0.015	2.897	0.004	0.014	0.075
PoolArea	-0.0896	0.161	-0.556	0.579	-0.406	0.227
MiscVal	0.0411	0.092	0.450	0.653	-0.138	0.221
TotalSF	0.9593	0.215	4.461	0.000	0.537	1.381
MSSubClass_160	-0.0469	0.034	-1.365	0.173	-0.114	0.021
MSSubClass_180	0.0532	0.078	0.681	0.496	-0.100	0.206
MSSubClass_190	0.0428	0.136	0.315	0.753	-0.224	0.310
MSSubClass_30	-0.0536	0.026	-2.050	0.041	-0.105	-0.002
MSSubClass_40	0.0507	0.099	0.515	0.607	-0.143	0.244
MSSubClass_45	-0.5522	0.160	-3.460	0.001	-0.866	-0.239

MSSubClass_50	-0.0403	0.048	-0.844	0.399	-0.134	0.053
MSSubClass_70	0.0209	0.030	0.698	0.485	-0.038	0.080
MSSubClass_75	-0.0047	0.083	-0.057	0.955	-0.167	0.158
MSSubClass_80	-0.0300	0.071	-0.421	0.674	-0.170	0.110
MSSubClass_85	0.0225	0.063	0.355	0.723	-0.102	0.147
MSSubClass_90	-0.0175	0.020	-0.865	0.387	-0.057	0.022
MSZoning_FV	0.5095	0.065	7.849	0.000	0.382	0.637
MSZoning_RH	0.4280	0.062	6.922	0.000	0.307	0.549
MSZoning_RL	0.4456	0.055	8.081	0.000	0.337	0.554
MSZoning_RM	0.4428	0.051	8.630	0.000	0.342	0.543
Street_Pave	0.1226	0.061	2.024	0.043	0.004	0.242
Alley_Pave	0.0546	0.027	1.999	0.046	0.001	0.108
LotShape_IR2	0.0174	0.022	0.793	0.428	-0.026	0.061
LotShape_IR3	0.0716	0.049	1.474	0.141	-0.024	0.167
LotShape_Reg	0.0123	0.009	1.450	0.147	-0.004	0.029
LandContour_HLS	-0.0007	0.030	-0.024	0.981	-0.061	0.059
LandContour_Low	-0.0252	0.035	-0.724	0.469	-0.094	0.043
LandContour_Lvl	-0.0022	0.021	-0.102	0.919	-0.044	0.040
Utilities_NoSeWa	-0.2218	0.136	-1.635	0.102	-0.488	0.045
LotConfig_CulDSac	0.0293	0.017	1.683	0.093	-0.005	0.063
LotConfig_FR2	-0.0136	0.021	-0.647	0.518	-0.055	0.028
LotConfig_FR3	-0.1702	0.151	-1.128	0.260	-0.466	0.126
LotConfig_Inside	-0.0126	0.010	-1.330	0.184	-0.031	0.006
LandSlope_Mod	0.0155	0.021	0.733	0.464	-0.026	0.057
LandSlope_Sev	-0.1565	0.053	-2.927	0.004	-0.261	-0.052

Neighborhood_Blueste	0.0123	0.092	0.133	0.894	-0.169	0.193
Neighborhood_BrDale	-0.0905	0.065	-1.386	0.166	-0.219	0.038
Neighborhood_BrkSide	-0.0597	0.054	-1.097	0.273	-0.166	0.047
Neighborhood_ClearCr	-0.0550	0.050	-1.102	0.271	-0.153	0.043
Neighborhood_CollgCr	-0.1006	0.041	-2.426	0.015	-0.182	-0.019
Neighborhood_Crawfor	0.0273	0.050	0.544	0.587	-0.071	0.126
Neighborhood_Edwards	-0.1616	0.046	-3.508	0.000	-0.252	-0.071
Neighborhood_Gilbert	-0.1010	0.043	-2.365	0.018	-0.185	-0.017
Neighborhood_IDOTRR	-0.1844	0.063	-2.925	0.004	-0.308	-0.061
Neighborhood_MeadowV	-0.1705	0.065	-2.626	0.009	-0.298	-0.043
Neighborhood_Mitchel	-0.1725	0.047	-3.642	0.000	-0.265	-0.080
Neighborhood_NAmes	-0.1313	0.045	-2.901	0.004	-0.220	-0.042
Neighborhood_NPkVill	-0.0457	0.067	-0.683	0.495	-0.177	0.086
Neighborhood_NWAmes	-0.1257	0.046	-2.728	0.007	-0.216	-0.035
Neighborhood_NoRidge	0.0111	0.047	0.239	0.811	-0.080	0.103
Neighborhood_NridgHt	-0.0363	0.043	-0.843	0.399	-0.121	0.048
Neighborhood_OldTown	-0.1487	0.056	-2.666	0.008	-0.258	-0.039
Neighborhood_SWISU	-0.0805	0.054	-1.483	0.139	-0.187	0.026
Neighborhood_Sawyer	-0.1176	0.046	-2.538	0.011	-0.209	-0.027
Neighborhood_SawyerW	-0.1011	0.044	-2.284	0.023	-0.188	-0.014
Neighborhood_Somerst	-0.0805	0.049	-1.656	0.098	-0.176	0.015
Neighborhood_StoneBr	0.0257	0.047	0.541	0.589	-0.067	0.119
Neighborhood_Timber	-0.0856	0.046	-1.878	0.061	-0.175	0.004
Neighborhood_Veenker	-0.0643	0.060	-1.065	0.287	-0.183	0.054
Condition1_Feedr	-0.0185	0.030	-0.606	0.544	-0.078	0.041

Condition1_Norm	0.0520	0.026	2.029	0.043	0.002	0.102
Condition1_PosA	0.0063	0.053	0.120	0.904	-0.097	0.110
Condition1_PosN	0.0592	0.041	1.459	0.145	-0.020	0.139
Condition1_RRAe	-0.0854	0.046	-1.853	0.064	-0.176	0.005
Condition1_RRAn	0.0244	0.039	0.624	0.533	-0.052	0.101
Condition1_RRNe	-0.0092	0.079	-0.117	0.907	-0.164	0.146
Condition1_RRNn	0.0479	0.069	0.697	0.486	-0.087	0.183
Condition2_Feedr	0.0424	0.155	0.274	0.784	-0.261	0.346
Condition2_Norm	-0.0263	0.135	-0.194	0.846	-0.292	0.240
Condition2_PosA	9.718e-16	7.27e-16	1.336	0.182	-4.56e-16	2.4e-15
Condition2_PosN	-1.3351	0.174	-7.667	0.000	-1.677	-0.993
Condition2_RRAe	-0.4978	0.262	-1.902	0.058	-1.012	0.016
Condition2_RRAn	-3.375e-16	8.53e-16	-0.396	0.692	-2.01e-15	1.34e-15
Condition2_RRNn	-6.836e-16	5.65e-16	-1.209	0.227	-1.79e-15	4.26e-16
BldgType_2fmCon	-0.1131	0.130	-0.867	0.386	-0.369	0.143
BldgType_Duplex	-0.0175	0.020	-0.865	0.387	-0.057	0.022
BldgType_Twnhs	0.0043	0.036	0.122	0.903	-0.066	0.074
BldgType_TwnhsE	-0.0053	0.022	-0.236	0.814	-0.049	0.039
HouseStyle_1.5Unf	0.5339	0.159	3.364	0.001	0.222	0.845
HouseStyle_1Story	-0.0232	0.045	-0.515	0.607	-0.112	0.065
HouseStyle_2.5Fin	0.0028	0.089	0.031	0.975	-0.172	0.178
HouseStyle_2.5Unf	0.0432	0.087	0.495	0.621	-0.128	0.214
HouseStyle_2Story	-0.0539	0.045	-1.206	0.228	-0.142	0.034
HouseStyle_SFoyer	-0.0361	0.070	-0.516	0.606	-0.173	0.101
HouseStyle_SLvl	-0.0222	0.078	-0.287	0.774	-0.175	0.130

OverallCond_2	1.0515	0.069	15.221	0.000	0.916	1.187
OverallCond_3	0.9616	0.046	20.992	0.000	0.872	1.051
OverallCond_4	1.0332	0.042	24.705	0.000	0.951	1.115
OverallCond_5	1.0767	0.040	26.735	0.000	0.998	1.156
OverallCond_6	1.1115	0.041	27.147	0.000	1.031	1.192
OverallCond_7	1.1536	0.041	27.999	0.000	1.073	1.234
OverallCond_8	1.1624	0.043	27.051	0.000	1.078	1.247
OverallCond_9	1.1891	0.050	23.772	0.000	1.091	1.287
RoofStyle_Gambrel	-0.0575	0.046	-1.246	0.213	-0.148	0.033
RoofStyle_Mansard	0.0436	0.053	0.825	0.410	-0.060	0.148
RoofStyle_Shed	0.2380	0.163	1.459	0.145	-0.082	0.558
RoofMatl_Membran	-2.99e-16	5.31e-16	-0.563	0.574	-1.34e-15	7.43e-16
RoofMatl_Metal	0.1987	0.120	1.660	0.097	-0.036	0.434
RoofMatl_Roll	0.0124	0.110	0.113	0.910	-0.203	0.228
RoofMatl_Tar&Grv	-0.0124	0.044	-0.283	0.777	-0.099	0.074
RoofMatl_WdShake	-0.0203	0.068	-0.299	0.765	-0.153	0.113
RoofMatl_WdShngl	0.1162	0.050	2.310	0.021	0.017	0.215
Exterior1st_AsphShn	-0.0371	0.060	-0.618	0.537	-0.155	0.081
Exterior1st_BrkComm	-0.5460	0.149	-3.660	0.000	-0.839	-0.253
Exterior1st_BrkFace	0.0612	0.041	1.507	0.132	-0.018	0.141
Exterior1st_CBlock	-0.0745	0.063	-1.181	0.238	-0.198	0.049
Exterior1st_CemntBd	-0.0846	0.086	-0.988	0.323	-0.253	0.083
Exterior1st_HdBoard	-0.0056	0.035	-0.160	0.873	-0.074	0.063
Exterior1st_ImStucc	-0.0167	0.113	-0.147	0.883	-0.239	0.206
Exterior1st_Plywood	0.0010	0.035	0.030	0.976	-0.067	0.069

Exterior1st_Stone	-0.0261	0.101	-0.258	0.797	-0.225	0.173
Exterior1st_Stucco	0.0255	0.056	0.458	0.647	-0.084	0.135
Exterior1st_Wd Sdng	-0.0501	0.031	-1.598	0.111	-0.112	0.011
Exterior1st_WdShing	-0.0065	0.039	-0.169	0.866	-0.082	0.069
Exterior2nd_AsphShn	-0.0371	0.060	-0.618	0.537	-0.155	0.081
Exterior2nd_Brk Cmn	0.0854	0.092	0.928	0.354	-0.095	0.266
Exterior2nd_BrkFace	-0.0150	0.048	-0.314	0.753	-0.109	0.079
Exterior2nd_CBlock	-0.0745	0.063	-1.181	0.238	-0.198	0.049
Exterior2nd_CmentBd	0.0896	0.086	1.043	0.297	-0.079	0.258
Exterior2nd_HdBoard	-0.0077	0.035	-0.218	0.827	-0.077	0.062
Exterior2nd_ImStucc	-0.0009	0.051	-0.019	0.985	-0.100	0.098
Exterior2nd_MetalSd	-0.0022	0.014	-0.156	0.876	-0.030	0.026
Exterior2nd_Other	-0.0664	0.105	-0.631	0.528	-0.273	0.140
Exterior2nd_Plywood	-0.0180	0.034	-0.533	0.594	-0.084	0.048
Exterior2nd_Stone	0.0300	0.077	0.389	0.697	-0.121	0.181
Exterior2nd_Stucco	0.0308	0.056	0.549	0.583	-0.079	0.141
Exterior2nd_Wd Sdng	0.0364	0.032	1.139	0.255	-0.026	0.099
Exterior2nd_Wd Shng	-0.0144	0.034	-0.422	0.673	-0.081	0.052
MasVnrType_BrkFace	0.0352	0.029	1.198	0.231	-0.023	0.093
MasVnrType_None	0.0721	0.038	1.901	0.058	-0.002	0.147
MasVnrType_Stone	0.0672	0.032	2.128	0.034	0.005	0.129
ExterQual_Fa	0.0438	0.066	0.666	0.506	-0.085	0.173
ExterQual_Gd	-0.0173	0.025	-0.681	0.496	-0.067	0.032
ExterQual_TA	-0.0317	0.028	-1.116	0.265	-0.088	0.024
ExterCond_Fa	-0.1465	0.087	-1.677	0.094	-0.318	0.025

ExterCond_Gd	-0.1172	0.083	-1.413	0.158	-0.280	0.046
ExterCond_Po	-0.2115	0.166	-1.273	0.203	-0.538	0.115
ExterCond_TA	-0.1080	0.083	-1.297	0.195	-0.271	0.055
Foundation_CBlock	0.0245	0.019	1.305	0.192	-0.012	0.061
Foundation_PConc	0.0250	0.019	1.283	0.200	-0.013	0.063
Foundation_Slab	-0.0036	0.069	-0.053	0.958	-0.138	0.131
Foundation_Stone	0.1746	0.058	3.018	0.003	0.061	0.288
Foundation_Wood	0.1014	0.110	0.924	0.356	-0.114	0.317
BsmtQual_Fa	0.0190	0.037	0.512	0.609	-0.054	0.092
BsmtQual_Gd	-0.0456	0.017	-2.624	0.009	-0.080	-0.011
BsmtQual_TA	-0.0340	0.022	-1.579	0.115	-0.076	0.008
BsmtCond_Gd	0.0183	0.029	0.640	0.523	-0.038	0.075
BsmtCond_Po	0.9398	0.120	7.861	0.000	0.705	1.175
BsmtCond_TA	0.0273	0.024	1.161	0.246	-0.019	0.073
BsmtExposure_Gd	0.0527	0.016	3.205	0.001	0.020	0.085
BsmtExposure_Mn	0.0103	0.016	0.641	0.522	-0.021	0.042
BsmtExposure_No	-0.0028	0.012	-0.247	0.805	-0.025	0.020
BsmtFinType1_BLQ	-0.0020	0.015	-0.138	0.891	-0.030	0.026
BsmtFinType1_GLQ	0.0089	0.014	0.655	0.512	-0.018	0.035
BsmtFinType1_LwQ	-0.0472	0.020	-2.340	0.020	-0.087	-0.008
BsmtFinType1_Rec	-0.0199	0.016	-1.244	0.214	-0.051	0.012
BsmtFinType1_Unf	0.0309	0.036	0.855	0.393	-0.040	0.102
BsmtFinType2_BLQ	-0.0831	0.036	-2.319	0.021	-0.153	-0.013
BsmtFinType2_GLQ	-0.0429	0.049	-0.876	0.381	-0.139	0.053
BsmtFinType2_LwQ	-0.0504	0.036	-1.420	0.156	-0.120	0.019

BsmtFinType2_Rec	-0.0413	0.035	-1.170	0.242	-0.111	0.028
BsmtFinType2_Unf	-0.0593	0.063	-0.935	0.350	-0.184	0.065
Heating_GasA	-0.0351	0.124	-0.283	0.777	-0.278	0.208
Heating_GasW	-0.0144	0.131	-0.110	0.912	-0.271	0.242
Heating_Grav	-0.2203	0.140	-1.576	0.115	-0.495	0.054
Heating_OthW	-0.0391	0.173	-0.226	0.822	-0.379	0.301
Heating_Wall	0.0020	0.163	0.012	0.990	-0.319	0.323
HeatingQC_Fa	-0.0585	0.025	-2.295	0.022	-0.108	-0.008
HeatingQC_Gd	-0.0210	0.011	-1.965	0.050	-0.042	-2.31e-05
HeatingQC_Po	0.0465	0.121	0.385	0.701	-0.191	0.284
HeatingQC_TA	-0.0351	0.011	-3.180	0.002	-0.057	-0.013
CentralAir_Y	0.0458	0.021	2.184	0.029	0.005	0.087
Electrical_FuseF	-0.0243	0.035	-0.704	0.482	-0.092	0.044
Electrical_FuseP	-0.2441	0.104	-2.340	0.020	-0.449	-0.039
Electrical_Mix	-0.9801	0.206	-4.764	0.000	-1.384	-0.576
Electrical_SBrkr	-0.0381	0.016	-2.442	0.015	-0.069	-0.007
KitchenQual_Fa	-0.0766	0.034	-2.283	0.023	-0.143	-0.011
KitchenQual_Gd	-0.0935	0.018	-5.138	0.000	-0.129	-0.058
KitchenQual_TA	-0.0891	0.021	-4.284	0.000	-0.130	-0.048
Functional_Maj2	-0.3801	0.114	-3.329	0.001	-0.604	-0.156
Functional_Min1	-0.0498	0.046	-1.090	0.276	-0.139	0.040
Functional_Min2	-0.0502	0.045	-1.110	0.267	-0.139	0.039
Functional_Mod	-0.0674	0.055	-1.225	0.221	-0.176	0.041
Functional_Sev	4.234e-16	1.84e-16	2.307	0.021	6.32e-17	7.84e-16
Functional_Typ	0.0216	0.040	0.543	0.587	-0.056	0.100

FireplaceQu_Fa	-0.0164	0.026	-0.631	0.528	-0.067	0.035
FireplaceQu_Gd	-0.0028	0.015	-0.189	0.851	-0.033	0.027
FireplaceQu_Po	0.0073	0.030	0.242	0.809	-0.052	0.067
FireplaceQu_TA	0.0052	0.016	0.322	0.748	-0.027	0.037
GarageType_Attchd	0.0291	0.024	1.198	0.231	-0.019	0.077
GarageType_Basment	0.0093	0.042	0.221	0.825	-0.073	0.091
GarageType_BuiltIn	0.0527	0.028	1.850	0.065	-0.003	0.109
GarageType_CarPort	-0.0405	0.064	-0.637	0.524	-0.165	0.084
GarageType_Detchd	0.0480	0.025	1.930	0.054	-0.001	0.097
GarageFinish_RFn	0.0054	0.010	0.528	0.598	-0.015	0.026
GarageFinish_Unf	-0.0046	0.012	-0.366	0.714	-0.029	0.020
GarageQual_Fa	-0.0301	0.028	-1.088	0.277	-0.084	0.024
GarageQual_Gd	0.0449	0.044	1.014	0.311	-0.042	0.132
GarageQual_Po	0.2498	0.135	1.850	0.065	-0.015	0.515
GarageCond_Fa	-0.0711	0.030	-2.394	0.017	-0.129	-0.013
GarageCond_Gd	-0.0506	0.054	-0.942	0.347	-0.156	0.055
GarageCond_Po	-0.1543	0.090	-1.719	0.086	-0.331	0.022
PavedDrive_P	-0.0719	0.033	-2.148	0.032	-0.138	-0.006
PavedDrive_Y	-0.0249	0.018	-1.345	0.179	-0.061	0.011
PoolQC_Fa	1.432e-16	9.06e-17	1.581	0.114	-3.46e-17	3.21e-16
PoolQC_Gd	0.3599	0.180	1.994	0.047	0.006	0.714
Fence_GdWo	-0.0422	0.019	-2.235	0.026	-0.079	-0.005
Fence_MnPrv	0.0019	0.012	0.151	0.880	-0.022	0.026
Fence_MnWw	-0.0702	0.042	-1.683	0.093	-0.152	0.012
MiscFeature_Othr	-0.1048	0.113	-0.926	0.355	-0.327	0.117

MiscFeature_Shed	-0.0103	0.064	-0.161	0.872	-0.135	0.115
MiscFeature_TenC	-1.798e-16	3.75e-17	-4.791	0.000	-2.53e-16	-1.06e-16
MoSold_10	-0.0026	0.023	-0.113	0.910	-0.047	0.042
MoSold_11	0.0040	0.023	0.175	0.861	-0.040	0.048
MoSold_12	0.0175	0.025	0.705	0.481	-0.031	0.066
MoSold_2	0.0037	0.025	0.146	0.884	-0.046	0.053
MoSold_3	0.0140	0.022	0.638	0.524	-0.029	0.057
MoSold_4	0.0084	0.021	0.409	0.683	-0.032	0.049
MoSold_5	0.0246	0.020	1.240	0.215	-0.014	0.064
MoSold_6	0.0289	0.020	1.477	0.140	-0.010	0.067
MoSold_7	0.0154	0.020	0.772	0.440	-0.024	0.055
MoSold_8	0.0202	0.022	0.931	0.352	-0.022	0.063
MoSold_9	0.0010	0.024	0.041	0.968	-0.047	0.049
YrSold_2007	0.0076	0.011	0.716	0.474	-0.013	0.028
YrSold_2008	0.0089	0.011	0.816	0.415	-0.012	0.030
YrSold_2009	-0.0056	0.011	-0.523	0.601	-0.027	0.015
YrSold_2010	-0.0105	0.013	-0.812	0.417	-0.036	0.015
SaleType_CWD	0.2523	0.109	2.310	0.021	0.038	0.467
SaleType_Con	0.0356	0.106	0.337	0.736	-0.172	0.243
SaleType_ConLD	0.2343	0.063	3.710	0.000	0.110	0.358
SaleType_ConLI	-0.0884	0.058	-1.511	0.131	-0.203	0.026
SaleType_ConLw	-0.0228	0.057	-0.400	0.689	-0.134	0.089
SaleType_New	0.3332	0.090	3.717	0.000	0.157	0.509
SaleType_Oth	0.0165	0.083	0.199	0.842	-0.146	0.179
SaleType_WD	-0.0252	0.024	-1.052	0.293	-0.072	0.022

SaleCondition_AdjLand	0.1477	0.075	1.974	0.049	0.001	0.295
SaleCondition_Alloca	0.1030	0.053	1.942	0.052	-0.001	0.207
SaleCondition_Family	0.0757	0.036	2.131	0.033	0.006	0.146
SaleCondition_Normal	0.0522	0.016	3.334	0.001	0.021	0.083
SaleCondition_Partial	-0.2605	0.087	-2.995	0.003	-0.431	-0.090

Omnibus:	221.319	**Durbin-Watson:**	2.040
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1430.171
Skew:	-0.826	**Prob(JB):**	2.77e-311
Kurtosis:	8.555	**Cond. No.**	1.41e+16

In [961]:
```
vif_100 = ['MSSubClass_20','MSSubClass_60','RoofStyle_Gable','RoofStyle_Hip','RoofMatl_CompShg','Exterior1st_MetalSd','Exterior1st_VinylSd','Exterior2nd_VinylSd','GarageQual_TA','GarageCond_TA']
# custom function to remove variables having higer VIF

to_keep = [x for x in x_test1 if x not in vif_100]
# print(to_keep)
x_test2 = x_test1[to_keep]
x_test2.head()
```

Out[961]:

	const	Id	LotFrontage	LotArea	OverallQual	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	BsmtFinSF
529	1.0	-0.137423	0.000000	0.251185	-0.011035	-0.103390	-0.164429	0.000000	0.333029	-0.089825
491	1.0	-0.163468	0.065799	0.009226	-0.011035	-0.219332	-0.581096	-0.288945	0.205091	0.610792
459	1.0	-0.185401	0.000000	-0.049919	-0.122146	-0.154115	-0.581096	0.400582	0.115299	-0.089825
279	1.0	-0.308773	0.084153	0.019570	0.100076	0.041538	-0.131096	0.484095	0.201895	-0.089825
655	1.0	-0.051062	-0.419834	-0.329602	-0.011035	-0.001941	-0.231096	0.516844	-0.489636	-0.089825

5 rows × 268 columns

In [962]:
```
predictions3 = model3.predict(x_test2)

tmp3 = pd.Series({'Model': " LRM after removing VIF above 100",
                  'R-Squared Value' : model3.rsquared,
                  'Adj.R-Squared Value': model3.rsquared_adj,
                  'RMSE': rmse(predictions3, y_test1)})

model3_report = models_report.append(tmp3, ignore_index = True)
model3_report
```

Out[962]:

	Model	R-Squared Value	Adj.R-Squared Value	RMSE
0	LRM after removing VIF above 100	0.957579	0.943383	0.172947

Checking variable having VIF above 10

In [963]:

```python
# Removing variable has threshold value of VIF above 10
print ("\nVariance Inflation Factor")
cnames = x_train2.columns
for i in np.arange(0,len(cnames)):
    xvars = list(cnames)
    yvar = xvars.pop(i)
    mod = sm.OLS(x_train2[yvar],(x_train2[xvars]))
    res = mod.fit()
    vif = 1/(1-res.rsquared)
    print (yvar,round(vif,3))
```

Variance Inflation Factor

```
C:\Users\computer\Anaconda3\lib\site-packages\statsmodels\regression\linear_model.py:1386: RuntimeWa
rning: divide by zero encountered in double_scalars
  return 1 - self.ssr/self.centered_tss
```

```
const 0.0
Id 1.325
LotFrontage 3.834
LotArea 6.474
OverallQual 6.23
YearBuilt 22.381
YearRemodAdd 4.486
MasVnrArea 31.1
BsmtFinSF1 35.07
BsmtFinSF2 42.99
BsmtUnfSF 4.055
LowQualFinSF 2.435
GrLivArea 49.44
BsmtFullBath 3.076
BsmtHalfBath 1.699
FullBath 4.691
HalfBath 3.869
BedroomAbvGr 4.241
KitchenAbvGr 5.859
TotRmsAbvGrd 7.043
Fireplaces 6.545
GarageYrBlt 7.296
GarageCars 9.844
GarageArea 9.122
WoodDeckSF 1.772
OpenPorchSF 2.073
EnclosedPorch 2.027
3SsnPorch 1.437
ScreenPorch 1.417
PoolArea 7.971
MiscVal 17.874
TotalSF 38.758
MSSubClass_160 5.907
MSSubClass_180 2.651
MSSubClass_190 41.496
MSSubClass_30 3.348
MSSubClass_40 3.173
MSSubClass_45 30.276
MSSubClass_50 22.275
MSSubClass_70 4.197
MSSubClass_75 7.433
MSSubClass_80 19.285
```

```
MSSubClass_85 5.626
MSSubClass_90 inf
```

```
C:\Users\computer\Anaconda3\lib\site-packages\ipykernel_launcher.py:9: RuntimeWarning: divide by zer
o encountered in double_scalars
  if __name__ == '__main__':
```

MSZoning_FV 16.836
MSZoning_RH 5.766
MSZoning_RL 54.011
MSZoning_RM 35.153
Street_Pave 1.994
Alley_Pave 2.528
LotShape_IR2 1.485
LotShape_IR3 1.536
LotShape_Reg 1.886
LandContour_HLS 2.759
LandContour_Low 3.481
LandContour_Lvl 4.458
Utilities_NoSeWa 2.008
LotConfig_CulDSac 2.095
LotConfig_FR2 1.458
LotConfig_FR3 2.482
LotConfig_Inside 2.011
LandSlope_Mod 2.322
LandSlope_Sev 3.093
Neighborhood_Blueste 1.856
Neighborhood_BrDale 4.607
Neighborhood_BrkSide 12.711
Neighborhood_ClearCr 6.38
Neighborhood_CollgCr 17.704
Neighborhood_Crawfor 9.305
Neighborhood_Edwards 15.316
Neighborhood_Gilbert 10.374
Neighborhood_IDOTRR 10.173
Neighborhood_MeadowV 5.458
Neighborhood_Mitchel 8.508
Neighborhood_NAmes 28.801
Neighborhood_NPkVill 3.405
Neighborhood_NWAmes 11.226
Neighborhood_NoRidge 6.903
Neighborhood_NridgHt 9.826
Neighborhood_OldTown 23.038
Neighborhood_SWISU 5.689
Neighborhood_Sawyer 13.456
Neighborhood_SawyerW 9.201
Neighborhood_Somerst 13.665
Neighborhood_StoneBr 3.873
Neighborhood_Timber 6.616
Neighborhood_Veenker 2.376

```
Condition1_Feedr 4.819
Condition1_Norm 8.029
Condition1_PosA 1.81
Condition1_PosN 2.48
Condition1_RRAe 2.298
Condition1_RRAn 2.625
Condition1_RRNe 1.363
Condition1_RRNn 2.058
Condition2_Feedr 10.421
Condition2_Norm 13.919
Condition2_PosA nan

C:\Users\computer\Anaconda3\lib\site-packages\statsmodels\regression\linear_model.py:1386: RuntimeWa
rning: invalid value encountered in double_scalars
  return 1 - self.ssr/self.centered_tss
```

Condition2_PosN 3.309
Condition2_RRAe 7.473
Condition2_RRAn nan
Condition2_RRNn nan
BldgType_2fmCon 40.021
BldgType_Duplex inf
BldgType_Twnhs 3.919
BldgType_TwnhsE 4.057
HouseStyle_1.5Unf 32.622
HouseStyle_1Story 56.521
HouseStyle_2.5Fin 4.327
HouseStyle_2.5Unf 6.601
HouseStyle_2Story 48.129
HouseStyle_SFoyer 10.984
HouseStyle_SLvl 25.254
OverallCond_2 inf
OverallCond_3 inf
OverallCond_4 inf
OverallCond_5 inf
OverallCond_6 inf
OverallCond_7 inf
OverallCond_8 inf
OverallCond_9 inf
RoofStyle_Gambrel 1.85
RoofStyle_Mansard 1.824
RoofStyle_Shed 5.804
RoofMatl_Membran nan
RoofMatl_Metal 1.563
RoofMatl_Roll 1.31
RoofMatl_Tar&Grv 1.887
RoofMatl_WdShake 2.499
RoofMatl_WdShngl 1.375
Exterior1st_AsphShn inf
Exterior1st_BrkComm 2.428
Exterior1st_BrkFace 6.085
Exterior1st_CBlock inf
Exterior1st_CemntBd 30.033
Exterior1st_HdBoard 17.486
Exterior1st_ImStucc 1.4
Exterior1st_Plywood 10.46
Exterior1st_Stone 2.24
Exterior1st_Stucco 5.659
Exterior1st_Wd Sdng 13.567

Exterior1st_WdShing 3.356
Exterior2nd_AsphShn inf
Exterior2nd_Brk Cmn 3.687
Exterior2nd_BrkFace 4.413
Exterior2nd_CBlock inf
Exterior2nd_CmentBd 30.277
Exterior2nd_HdBoard 16.663
Exterior2nd_ImStucc 2.223
Exterior2nd_MetalSd 2.668
Exterior2nd_Other 1.209
Exterior2nd_Plywood 12.477
Exterior2nd_Stone 2.583
Exterior2nd_Stucco 5.415
Exterior2nd_Wd Sdng 13.631
Exterior2nd_Wd Shng 3.567
MasVnrType_BrkFace 20.08
MasVnrType_None 38.718
MasVnrType_Stone 9.572
ExterQual_Fa 5.614
ExterQual_Gd 15.913
ExterQual_TA 21.119
ExterCond_Fa 20.344
ExterCond_Gd 69.591
ExterCond_Po 3.011
ExterCond_TA 86.369
Foundation_CBlock 9.624
Foundation_PConc 10.464
Foundation_Slab 9.579
Foundation_Stone 1.818
Foundation_Wood 1.314
BsmtQual_Fa 3.384
BsmtQual_Gd 8.225
BsmtQual_TA 12.847
BsmtCond_Gd 3.932
BsmtCond_Po inf
BsmtCond_TA 5.687
BsmtExposure_Gd 2.428
BsmtExposure_Mn 2.205
BsmtExposure_No 3.322
BsmtFinType1_BLQ 2.222
BsmtFinType1_GLQ 4.165
BsmtFinType1_LwQ 2.118
BsmtFinType1_Rec 2.407

```
BsmtFinType1_Unf 30.41
BsmtFinType2_BLQ 3.82
BsmtFinType2_GLQ 2.334
BsmtFinType2_LwQ 5.044
BsmtFinType2_Rec 4.864
BsmtFinType2_Unf 55.936
Heating_GasA 35.945
Heating_GasW 22.117
Heating_Grav 12.737
Heating_OthW 3.278
Heating_Wall 5.827
HeatingQC_Fa 2.264
HeatingQC_Gd 1.825
HeatingQC_Po 1.591
HeatingQC_TA 2.861
CentralAir_Y 3.131
Electrical_FuseF 2.431
Electrical_FuseP 2.372
Electrical_Mix inf
Electrical_SBrkr 2.244
KitchenQual_Fa 3.468
KitchenQual_Gd 8.804
KitchenQual_TA 12.063
Functional_Maj2 2.843
Functional_Min1 4.462
Functional_Min2 6.084
Functional_Mod 3.601
Functional_Sev nan
Functional_Typ 11.551
FireplaceQu_Fa 1.866
FireplaceQu_Gd 4.771
FireplaceQu_Po 1.584
FireplaceQu_TA 5.131
GarageType_Attchd 15.837
GarageType_Basment 2.457
GarageType_BuiltIn 5.387
GarageType_CarPort 2.2
GarageType_Detchd 13.415
GarageFinish_RFn 2.445
GarageFinish_Unf 4.212
GarageQual_Fa 2.827
GarageQual_Gd 1.909
GarageQual_Po 5.956
```

GarageCond_Fa 2.624
GarageCond_Gd 1.882
GarageCond_Po 5.251
PavedDrive_P 2.162
PavedDrive_Y 3.03
PoolQC_Fa nan
PoolQC_Gd 7.101
Fence_GdWo 1.392
Fence_MnPrv 1.504
Fence_MnWw 1.322
MiscFeature_Othr 2.794
MiscFeature_Shed 17.049
MiscFeature_TenC nan
MoSold_10 3.377
MoSold_11 3.154
MoSold_12 2.52
MoSold_2 2.379
MoSold_3 3.585
MoSold_4 4.435
MoSold_5 5.191
MoSold_6 6.214
MoSold_7 5.992
MoSold_8 3.559
MoSold_9 2.653
YrSold_2007 2.175
YrSold_2008 2.189
YrSold_2009 2.248
YrSold_2010 2.142
SaleType_CWD 1.302
SaleType_Con 1.22
SaleType_ConLD 2.6
SaleType_ConLI 1.488
SaleType_ConLw 1.756
SaleType_New 64.713
SaleType_Oth 1.495
SaleType_WD 6.881
SaleCondition_AdjLand 1.831
SaleCondition_Alloca 2.133
SaleCondition_Family 1.501
SaleCondition_Normal 3.878
SaleCondition_Partial 62.334

Below are the variable having above 10 VIF threshold

In [964]:
```
VIF_10 = ['MSSubClass_20','MSSubClass_60','MSSubClass_90','YearBuilt','MasVnrArea','BsmtFinSF1','BsmtFinSF2','GrLivArea',
          'GarageYrBlt','MiscVal','TotalSF','MSSubClass_190','MSSubClass_45','Neighborhood_Gilbert', 'Neighborhood_IDOTRR',
          'MSSubClass_50','MSSubClass_80', 'MSZoning_FV','MSZoning_RL','MSZoning_RM','Neighborhood_BrkSide',
          'Neighborhood_CollgCr','Neighborhood_Edwards', 'Neighborhood_NAmes','Neighborhood_OldTown', 'Neighborhood_Sawyer',
          'Neighborhood_Somerst','Condition2_Norm','HouseStyle_1.5Unf','HouseStyle_2Story','HouseStyle_SLvl',
          'Neighborhood_NWAmes', 'Condition2_Feedr','BldgType_2fmCon','Foundation_PConc','KitchenQual_TA',
          'HouseStyle_SFoyer','MasVnrType_BrkFace','HouseStyle_1Story','Exterior1st_CemntBd','Exterior1st_HdBoard',
          'Exterior1st_Plywood','Exterior1st_Wd Sdng','Exterior2nd_CmentBd','Exterior2nd_HdBoard','Exterior2nd_Plywood',
          'Exterior2nd_Wd Sdng','MasVnrType_None','MasVnrType_Stone', 'ExterQual_Gd','ExterQual_TA', 'ExterCond_Fa',
          'ExterCond_Gd','ExterCond_TA','BsmtQual_TA','BsmtFinType1_Unf','BsmtFinType2_Unf','Heating_GasA',
          'Heating_GasW','Heating_Grav','GarageType_BuiltIn','SaleType_New','SaleCondition_Partial', 'GarageType_Attchd',
          'GarageType_Detchd','MiscFeature_Shed','Functional_Typ']
to_keep = [x for x in x_train2 if x not in VIF_10]
#print(to_keep)
x_train2 = x_train2[to_keep]
x_train2.head()
```

Out[964]:

	const	Id	LotFrontage	LotArea	OverallQual	YearRemodAdd	BsmtUnfSF	LowQualFinSF	BsmtFullBath	Bsm
64	1.0	-0.456134	0.000000	0.006840	0.100076	0.218904	0.015060	-0.015717	0.191553	-0.0
682	1.0	-0.032557	0.000000	-0.223669	-0.011035	0.202237	0.002327	-0.015717	0.191553	-0.0
960	1.0	0.157985	-0.103554	-0.044634	-0.122146	0.385571	-0.071504	-0.015717	0.191553	-0.0
1384	1.0	0.448595	-0.036201	0.000150	-0.011035	-0.581096	0.029569	-0.015717	-0.141781	-0.0
1100	1.0	0.253941	-0.036201	-0.014654	-0.455479	-0.581096	-0.728201	-0.015717	-0.141781	-0.0

5 rows × 203 columns

8.4.2 Building Model after removing VIF above 10

In [965]:

```python
# Lets build Linear Regression model using statsmodel
import statsmodels.api as sm

# Building Linear Regression model using OLS

model4 = sm.OLS(y_train1,x_train2).fit()
# Note the Swap of X and Y
# Printing Linear Regression Summary
model4.summary()
```

Out[965]:

OLS Regression Results

Dep. Variable:	SalePrice	**R-squared:**	0.921
Model:	OLS	**Adj. R-squared:**	0.903
Method:	Least Squares	**F-statistic:**	50.59
Date:	Wed, 25 Jul 2018	**Prob (F-statistic):**	0.00
Time:	13:26:58	**Log-Likelihood:**	780.75
No. Observations:	1022	**AIC:**	-1175.
Df Residuals:	829	**BIC:**	-224.1
Df Model:	192		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	10.4990	0.097	108.707	0.000	10.309	10.689
Id	-0.0130	0.015	-0.879	0.380	-0.042	0.016
LotFrontage	0.1636	0.063	2.594	0.010	0.040	0.287
LotArea	0.4557	0.087	5.260	0.000	0.286	0.626
OverallQual	0.7137	0.054	13.143	0.000	0.607	0.820
YearRemodAdd	0.0917	0.022	4.252	0.000	0.049	0.134
BsmtUnfSF	0.0193	0.028	0.696	0.487	-0.035	0.074
LowQualFinSF	-0.0603	0.047	-1.295	0.196	-0.152	0.031
BsmtFullBath	0.1867	0.034	5.563	0.000	0.121	0.253
BsmtHalfBath	0.0681	0.033	2.085	0.037	0.004	0.132
FullBath	0.2577	0.040	6.499	0.000	0.180	0.335
HalfBath	0.0815	0.022	3.748	0.000	0.039	0.124
BedroomAbvGr	0.0276	0.074	0.374	0.708	-0.117	0.173

KitchenAbvGr	-0.2277	0.112	-2.032	0.042	-0.448	-0.008
TotRmsAbvGrd	0.3722	0.063	5.934	0.000	0.249	0.495
Fireplaces	0.1581	0.045	3.529	0.000	0.070	0.246
GarageCars	0.0798	0.060	1.334	0.183	-0.038	0.197
GarageArea	0.3148	0.073	4.319	0.000	0.172	0.458
WoodDeckSF	0.0443	0.013	3.400	0.001	0.019	0.070
OpenPorchSF	0.0337	0.015	2.223	0.027	0.004	0.063
EnclosedPorch	-0.0360	0.018	-1.976	0.048	-0.072	-0.000
3SsnPorch	0.0094	0.043	0.222	0.824	-0.074	0.093
ScreenPorch	0.0558	0.019	2.874	0.004	0.018	0.094
PoolArea	-0.0740	0.193	-0.384	0.701	-0.453	0.305
MSSubClass_160	-0.0353	0.039	-0.910	0.363	-0.111	0.041
MSSubClass_180	-0.0501	0.078	-0.644	0.520	-0.203	0.103
MSSubClass_30	-0.1049	0.026	-4.112	0.000	-0.155	-0.055
MSSubClass_40	0.1281	0.115	1.111	0.267	-0.098	0.354
MSSubClass_70	-0.0103	0.027	-0.380	0.704	-0.063	0.043
MSSubClass_75	0.0808	0.093	0.872	0.384	-0.101	0.263
MSSubClass_85	-0.0159	0.041	-0.390	0.697	-0.096	0.064
MSZoning_RH	0.0241	0.040	0.597	0.550	-0.055	0.103
Street_Pave	0.3014	0.069	4.389	0.000	0.167	0.436
Alley_Pave	0.0971	0.030	3.226	0.001	0.038	0.156
LotShape_IR2	0.0191	0.028	0.685	0.494	-0.036	0.074
LotShape_IR3	0.0654	0.060	1.082	0.280	-0.053	0.184
LotShape_Reg	0.0016	0.010	0.152	0.879	-0.019	0.022
LandContour_HLS	-0.0440	0.037	-1.198	0.231	-0.116	0.028

LandContour_Low	-0.0649	0.042	-1.548	0.122	-0.147	0.017
LandContour_Lvl	0.0056	0.026	0.216	0.829	-0.045	0.056
Utilities_NoSeWa	-0.4455	0.148	-3.016	0.003	-0.735	-0.156
LotConfig_CulDSac	0.0564	0.022	2.593	0.010	0.014	0.099
LotConfig_FR2	-0.0137	0.026	-0.518	0.605	-0.066	0.038
LotConfig_FR3	-0.2216	0.161	-1.376	0.169	-0.538	0.094
LotConfig_Inside	-0.0066	0.012	-0.549	0.583	-0.030	0.017
LandSlope_Mod	0.0429	0.026	1.678	0.094	-0.007	0.093
LandSlope_Sev	-0.0509	0.064	-0.791	0.429	-0.177	0.075
Neighborhood_Blueste	-0.0215	0.102	-0.210	0.834	-0.223	0.180
Neighborhood_BrDale	-0.0369	0.056	-0.662	0.508	-0.146	0.072
Neighborhood_ClearCr	0.0851	0.033	2.550	0.011	0.020	0.151
Neighborhood_Crawfor	0.1808	0.029	6.322	0.000	0.125	0.237
Neighborhood_MeadowV	-0.0509	0.049	-1.038	0.300	-0.147	0.045
Neighborhood_Mitchel	-0.0350	0.025	-1.377	0.169	-0.085	0.015
Neighborhood_NPkVill	-0.0340	0.071	-0.477	0.633	-0.174	0.106
Neighborhood_NoRidge	0.1557	0.028	5.596	0.000	0.101	0.210
Neighborhood_NridgHt	0.1041	0.024	4.365	0.000	0.057	0.151
Neighborhood_SWISU	0.0556	0.039	1.442	0.150	-0.020	0.131
Neighborhood_SawyerW	0.0050	0.023	0.216	0.829	-0.041	0.051
Neighborhood_StoneBr	0.2021	0.038	5.312	0.000	0.127	0.277
Neighborhood_Timber	0.0546	0.029	1.911	0.056	-0.001	0.111
Neighborhood_Veenker	0.0512	0.056	0.911	0.362	-0.059	0.161
Condition1_Feedr	0.0267	0.036	0.734	0.463	-0.045	0.098
Condition1_Norm	0.0821	0.030	2.728	0.007	0.023	0.141

Condition1_PosA	0.0435	0.065	0.670	0.503	-0.084	0.171
Condition1_PosN	0.1224	0.049	2.514	0.012	0.027	0.218
Condition1_RRAe	-0.0206	0.056	-0.371	0.710	-0.130	0.088
Condition1_RRAn	0.0450	0.045	0.995	0.320	-0.044	0.134
Condition1_RRNe	0.0326	0.101	0.324	0.746	-0.165	0.230
Condition1_RRNn	0.1671	0.084	1.996	0.046	0.003	0.332
Condition2_PosA	-1.327e-15	5.79e-16	-2.292	0.022	-2.46e-15	-1.9e-16
Condition2_PosN	-1.1232	0.142	-7.932	0.000	-1.401	-0.845
Condition2_RRAe	-0.3835	0.247	-1.552	0.121	-0.869	0.102
Condition2_RRAn	-9.645e-16	6.24e-16	-1.546	0.122	-2.19e-15	2.6e-16
Condition2_RRNn	-1.041e-15	4.43e-16	-2.347	0.019	-1.91e-15	-1.7e-16
BldgType_Duplex	0.0319	0.041	0.782	0.434	-0.048	0.112
BldgType_Twnhs	0.0248	0.044	0.562	0.574	-0.062	0.111
BldgType_TwnhsE	0.0179	0.025	0.705	0.481	-0.032	0.068
HouseStyle_2.5Fin	-0.0319	0.098	-0.327	0.744	-0.223	0.160
HouseStyle_2.5Unf	-0.1033	0.091	-1.135	0.257	-0.282	0.075
OverallCond_2	1.1022	0.079	14.015	0.000	0.948	1.257
OverallCond_3	0.8942	0.051	17.697	0.000	0.795	0.993
OverallCond_4	1.0059	0.043	23.170	0.000	0.921	1.091
OverallCond_5	1.0560	0.041	25.649	0.000	0.975	1.137
OverallCond_6	1.0922	0.042	26.097	0.000	1.010	1.174
OverallCond_7	1.1135	0.042	26.587	0.000	1.031	1.196
OverallCond_8	1.0782	0.044	24.517	0.000	0.992	1.164
OverallCond_9	1.0793	0.054	19.996	0.000	0.973	1.185
RoofStyle_Gambrel	-0.0777	0.053	-1.462	0.144	-0.182	0.027

RoofStyle_Mansard	0.0911	0.065	1.393	0.164	-0.037	0.219
RoofStyle_Shed	0.3171	0.205	1.545	0.123	-0.086	0.720
RoofMatl_Membran	3.616e-15	4.45e-16	8.129	0.000	2.74e-15	4.49e-15
RoofMatl_Metal	-0.0133	0.150	-0.089	0.929	-0.308	0.281
RoofMatl_Roll	0.0048	0.138	0.034	0.973	-0.267	0.276
RoofMatl_Tar&Grv	-0.0554	0.053	-1.036	0.301	-0.160	0.050
RoofMatl_WdShake	-0.1020	0.084	-1.211	0.226	-0.267	0.063
RoofMatl_WdShngl	0.1358	0.063	2.158	0.031	0.012	0.259
Exterior1st_AsphShn	-0.0046	0.068	-0.067	0.947	-0.139	0.130
Exterior1st_BrkComm	-0.6170	0.180	-3.432	0.001	-0.970	-0.264
Exterior1st_BrkFace	0.1401	0.037	3.815	0.000	0.068	0.212
Exterior1st_CBlock	0.0200	0.075	0.266	0.790	-0.127	0.167
Exterior1st_ImStucc	-0.0420	0.143	-0.294	0.768	-0.322	0.238
Exterior1st_Stone	0.2249	0.117	1.918	0.055	-0.005	0.455
Exterior1st_Stucco	0.0635	0.060	1.050	0.294	-0.055	0.182
Exterior1st_WdShing	-0.0310	0.038	-0.825	0.409	-0.105	0.043
Exterior2nd_AsphShn	-0.0046	0.068	-0.067	0.947	-0.139	0.130
Exterior2nd_Brk Cmn	0.1288	0.107	1.202	0.230	-0.081	0.339
Exterior2nd_BrkFace	-0.0720	0.049	-1.464	0.143	-0.168	0.024
Exterior2nd_CBlock	0.0200	0.075	0.266	0.790	-0.127	0.167
Exterior2nd_ImStucc	0.0252	0.053	0.471	0.638	-0.080	0.130
Exterior2nd_MetalSd	0.0026	0.015	0.176	0.861	-0.026	0.031
Exterior2nd_Other	-0.1648	0.135	-1.218	0.223	-0.430	0.101
Exterior2nd_Stone	-0.0865	0.089	-0.973	0.331	-0.261	0.088
Exterior2nd_Stucco	-0.0166	0.062	-0.268	0.788	-0.138	0.105

Exterior2nd_Wd Shng	-0.0372	0.033	-1.134	0.257	-0.102	0.027
ExterQual_Fa	0.0124	0.062	0.200	0.842	-0.109	0.134
ExterCond_Po	-0.0606	0.169	-0.358	0.721	-0.393	0.272
Foundation_CBlock	6.545e-05	0.013	0.005	0.996	-0.026	0.027
Foundation_Slab	-0.0480	0.051	-0.947	0.344	-0.148	0.052
Foundation_Stone	0.0677	0.070	0.971	0.332	-0.069	0.204
Foundation_Wood	0.1572	0.139	1.135	0.257	-0.115	0.429
BsmtQual_Fa	0.0519	0.035	1.469	0.142	-0.017	0.121
BsmtQual_Gd	-0.0144	0.013	-1.121	0.263	-0.040	0.011
BsmtCond_Gd	0.0239	0.035	0.687	0.493	-0.044	0.092
BsmtCond_Po	1.0424	0.137	7.603	0.000	0.773	1.311
BsmtCond_TA	0.0296	0.028	1.051	0.293	-0.026	0.085
BsmtExposure_Gd	0.0659	0.020	3.219	0.001	0.026	0.106
BsmtExposure_Mn	-0.0017	0.019	-0.089	0.929	-0.040	0.036
BsmtExposure_No	-0.0003	0.014	-0.024	0.981	-0.027	0.026
BsmtFinType1_BLQ	0.0160	0.016	1.017	0.310	-0.015	0.047
BsmtFinType1_GLQ	0.0535	0.013	3.993	0.000	0.027	0.080
BsmtFinType1_LwQ	0.0082	0.023	0.363	0.716	-0.036	0.053
BsmtFinType1_Rec	0.0089	0.017	0.524	0.601	-0.025	0.042
BsmtFinType2_BLQ	-0.0174	0.027	-0.654	0.513	-0.070	0.035
BsmtFinType2_GLQ	0.0335	0.051	0.657	0.511	-0.067	0.134
BsmtFinType2_LwQ	0.0181	0.025	0.715	0.475	-0.032	0.068
BsmtFinType2_Rec	0.0181	0.025	0.718	0.473	-0.031	0.068
Heating_OthW	-0.1016	0.149	-0.684	0.494	-0.393	0.190
Heating_Wall	0.1018	0.116	0.878	0.380	-0.126	0.329

HeatingQC_Fa	-0.0855	0.029	-2.910	0.004	-0.143	-0.028
HeatingQC_Gd	-0.0349	0.013	-2.657	0.008	-0.061	-0.009
HeatingQC_Po	-0.1098	0.146	-0.750	0.454	-0.397	0.178
HeatingQC_TA	-0.0519	0.013	-3.914	0.000	-0.078	-0.026
CentralAir_Y	0.1092	0.024	4.574	0.000	0.062	0.156
Electrical_FuseF	0.0211	0.041	0.516	0.606	-0.059	0.101
Electrical_FuseP	-0.1601	0.126	-1.266	0.206	-0.408	0.088
Electrical_Mix	-1.0351	0.250	-4.147	0.000	-1.525	-0.545
Electrical_SBrkr	-0.0250	0.019	-1.333	0.183	-0.062	0.012
KitchenQual_Fa	0.0253	0.032	0.785	0.433	-0.038	0.089
KitchenQual_Gd	-0.0184	0.012	-1.557	0.120	-0.042	0.005
Functional_Maj2	-0.3935	0.133	-2.957	0.003	-0.655	-0.132
Functional_Min1	0.0076	0.032	0.237	0.813	-0.056	0.071
Functional_Min2	-0.0389	0.027	-1.418	0.157	-0.093	0.015
Functional_Mod	-0.0180	0.049	-0.364	0.716	-0.115	0.079
Functional_Sev	1.658e-16	1.61e-16	1.030	0.303	-1.5e-16	4.82e-16
FireplaceQu_Fa	-0.0203	0.033	-0.621	0.535	-0.085	0.044
FireplaceQu_Gd	-0.0107	0.019	-0.567	0.571	-0.048	0.026
FireplaceQu_Po	-0.0311	0.038	-0.817	0.414	-0.106	0.044
FireplaceQu_TA	-0.0173	0.020	-0.860	0.390	-0.057	0.022
GarageType_Basment	-0.0425	0.045	-0.954	0.340	-0.130	0.045
GarageType_CarPort	-0.1036	0.066	-1.580	0.115	-0.232	0.025
GarageFinish_RFn	-0.0005	0.012	-0.042	0.967	-0.024	0.023
GarageFinish_Unf	-0.0106	0.014	-0.775	0.439	-0.037	0.016
GarageQual_Fa	-0.0519	0.031	-1.662	0.097	-0.113	0.009

GarageQual_Gd	-0.0401	0.055	-0.724	0.469	-0.149	0.069
GarageQual_Po	-0.1462	0.145	-1.008	0.314	-0.431	0.138
GarageCond_Fa	-0.0578	0.034	-1.683	0.093	-0.125	0.010
GarageCond_Gd	0.0099	0.068	0.146	0.884	-0.124	0.143
GarageCond_Po	0.1282	0.088	1.452	0.147	-0.045	0.301
PavedDrive_P	0.0115	0.039	0.295	0.768	-0.065	0.088
PavedDrive_Y	0.0401	0.021	1.870	0.062	-0.002	0.082
PoolQC_Fa	-8.706e-17	9.82e-17	-0.886	0.376	-2.8e-16	1.06e-16
PoolQC_Gd	0.3537	0.217	1.627	0.104	-0.073	0.780
Fence_GdWo	-0.0669	0.024	-2.829	0.005	-0.113	-0.020
Fence_MnPrv	-0.0208	0.015	-1.388	0.166	-0.050	0.009
Fence_MnWw	-0.0666	0.052	-1.276	0.202	-0.169	0.036
MiscFeature_Othr	-0.1412	0.124	-1.135	0.257	-0.385	0.103
MiscFeature_TenC	1.628e-16	6.69e-17	2.435	0.015	3.16e-17	2.94e-16
MoSold_10	-0.0094	0.029	-0.328	0.743	-0.066	0.047
MoSold_11	-0.0147	0.028	-0.518	0.605	-0.071	0.041
MoSold_12	0.0232	0.031	0.738	0.461	-0.038	0.085
MoSold_2	-0.0134	0.031	-0.428	0.669	-0.075	0.048
MoSold_3	0.0034	0.028	0.122	0.903	-0.051	0.057
MoSold_4	-0.0021	0.026	-0.082	0.935	-0.053	0.048
MoSold_5	0.0115	0.025	0.462	0.644	-0.037	0.060
MoSold_6	-0.0008	0.025	-0.032	0.975	-0.049	0.047
MoSold_7	0.0021	0.025	0.083	0.934	-0.047	0.051
MoSold_8	0.0122	0.027	0.448	0.654	-0.041	0.066
MoSold_9	-0.0088	0.031	-0.287	0.774	-0.069	0.051

YrSold_2007	0.0007	0.013	0.053	0.958	-0.025	0.027
YrSold_2008	-0.0036	0.013	-0.265	0.791	-0.030	0.023
YrSold_2009	-0.0013	0.013	-0.094	0.925	-0.027	0.025
YrSold_2010	-0.0284	0.016	-1.763	0.078	-0.060	0.003
SaleType_CWD	0.2634	0.134	1.960	0.050	-0.000	0.527
SaleType_Con	-0.0506	0.130	-0.389	0.697	-0.306	0.205
SaleType_ConLD	0.0204	0.062	0.329	0.742	-0.101	0.142
SaleType_ConLI	-0.2000	0.069	-2.900	0.004	-0.335	-0.065
SaleType_ConLw	-0.1070	0.067	-1.590	0.112	-0.239	0.025
SaleType_Oth	0.0374	0.100	0.373	0.709	-0.159	0.234
SaleType_WD	-0.1192	0.021	-5.595	0.000	-0.161	-0.077
SaleCondition_AdjLand	0.0260	0.090	0.291	0.771	-0.150	0.202
SaleCondition_Alloca	0.0720	0.066	1.095	0.274	-0.057	0.201
SaleCondition_Family	0.0660	0.045	1.470	0.142	-0.022	0.154
SaleCondition_Normal	0.0635	0.018	3.500	0.000	0.028	0.099

Omnibus:	152.032	**Durbin-Watson:**	1.948
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	530.233
Skew:	-0.696	**Prob(JB):**	7.27e-116
Kurtosis:	6.243	**Cond. No.**	1.06e+16

In [966]:

```
VIF_10 = ['MSSubClass_20','MSSubClass_60','MSSubClass_90','YearBuilt','MasVnrArea','BsmtFinSF1','BsmtFinSF2','GrLivArea',
          'GarageYrBlt','MiscVal','TotalSF','MSSubClass_190','MSSubClass_45','Neighborhood_Gilbert','Neighborhood_IDOTRR',
          'MSSubClass_50','MSSubClass_80', 'MSZoning_FV','MSZoning_RL','MSZoning_RM','Neighborhood_BrkSide',
          'Neighborhood_CollgCr','Neighborhood_Edwards', 'Neighborhood_NAmes','Neighborhood_OldTown','Neighborhood_Sawyer',
          'Neighborhood_Somerst','Condition2_Norm','HouseStyle_1.5Unf','HouseStyle_2Story','HouseStyle_SLvl',
          'Neighborhood_NWAmes', 'Condition2_Feedr','BldgType_2fmCon','Foundation_PConc','KitchenQual_TA',
          'HouseStyle_SFoyer','MasVnrType_BrkFace','HouseStyle_1Story','Exterior1st_CemntBd','Exterior1st_HdBoard',
          'Exterior1st_Plywood','Exterior1st_Wd Sdng','Exterior2nd_CmentBd','Exterior2nd_HdBoard','Exterior2nd_Plywood',
          'Exterior2nd_Wd Sdng','MasVnrType_None','MasVnrType_Stone', 'ExterQual_Gd','ExterQual_TA','ExterCond_Fa',
          'ExterCond_Gd','ExterCond_TA','BsmtQual_TA','BsmtFinType1_Unf','BsmtFinType2_Unf','Heating_GasA',
          'Heating_GasW','Heating_Grav','GarageType_BuiltIn','SaleType_New','SaleCondition_Partial','GarageType_Attchd',
          'GarageType_Detchd','MiscFeature_Shed','Functional_Typ']
to_keep = [x for x in x_test2 if x not in VIF_10]
#print(to_keep)
x_test2 = x_test2[to_keep]
x_test2.head()
```

Out[966]:

	const	Id	LotFrontage	LotArea	OverallQual	YearRemodAdd	BsmtUnfSF	LowQualFinSF	BsmtFullBath	Bsmt
529	1.0	-0.137423	0.000000	0.251185	-0.011035	-0.164429	0.136304	-0.015717	0.191553	-0.03
491	1.0	-0.163468	0.065799	0.009226	-0.011035	-0.581096	-0.022164	-0.015717	0.191553	-0.03
459	1.0	-0.185401	0.000000	-0.049919	-0.122146	-0.581096	0.079290	-0.015717	0.191553	-0.03
279	1.0	-0.308773	0.084153	0.019570	0.100076	-0.131096	0.128498	-0.015717	-0.141781	-0.03
655	1.0	-0.051062	-0.419834	-0.329602	-0.011035	-0.231096	0.079535	-0.015717	-0.141781	-0.03

5 rows × 203 columns

In [967]:
```
predictions4 = model4.predict(x_test2)
tmp4 = pd.Series({'Model': " LRM after removing VIF above 10",
                  'R-Squared Value' : model4.rsquared,
                  'Adj.R-Squared Value': model4.rsquared_adj,
                  'RMSE': rmse(predictions4, y_test1)})

model4_report = models_report.append(tmp4, ignore_index = True)
model4_report
```

Out[967]:

	Model	R-Squared Value	Adj.R-Squared Value	RMSE
0	LRM after removing VIF above 10	0.921361	0.903147	0.17904

Checking variable having VIF above 5

In [968]:

```
# Removing variable has threshold value of VIF above 5
print ("\nVariance Inflation Factor")
cnames = x_train2.columns
for i in np.arange(0,len(cnames)):
    xvars = list(cnames)
    yvar = xvars.pop(i)
    mod = sm.OLS(x_train2[yvar],(x_train2[xvars]))
    res = mod.fit()
    vif = 1/(1-res.rsquared)
    print (yvar,round(vif,3))
```

```
Variance Inflation Factor
const 0.0

C:\Users\computer\Anaconda3\lib\site-packages\statsmodels\regression\linear_model.py:1386: RuntimeWa
rning: divide by zero encountered in double_scalars
  return 1 - self.ssr/self.centered_tss
```

Id 1.193
LotFrontage 3.41
LotArea 4.937
OverallQual 4.376
YearRemodAdd 3.542
BsmtUnfSF 2.753
LowQualFinSF 1.965
BsmtFullBath 2.164
BsmtHalfBath 1.498
FullBath 3.34
HalfBath 1.914
BedroomAbvGr 3.583
KitchenAbvGr 3.417
TotRmsAbvGrd 4.649
Fireplaces 5.871
GarageCars 7.853
GarageArea 7.513
WoodDeckSF 1.646
OpenPorchSF 1.759
EnclosedPorch 1.696
3SsnPorch 1.342
ScreenPorch 1.324
PoolArea 6.679
MSSubClass_160 4.399
MSSubClass_180 1.543
MSSubClass_30 1.862
MSSubClass_40 2.539
MSSubClass_70 2.017
MSSubClass_75 5.435
MSSubClass_85 1.356
MSZoning_RH 1.429
Street_Pave 1.498
Alley_Pave 1.793
LotShape_IR2 1.396
LotShape_IR3 1.391
LotShape_Reg 1.609
LandContour_HLS 2.347
LandContour_Low 2.95
LandContour_Lvl 3.82
Utilities_NoSeWa 1.392
LotConfig_CulDSac 1.915
LotConfig_FR2 1.346
LotConfig_FR3 1.654

```
LotConfig_Inside 1.852
LandSlope_Mod 1.988
LandSlope_Sev 2.613
Neighborhood_Blueste 1.339
Neighborhood_BrDale 1.963
Neighborhood_ClearCr 1.666
Neighborhood_Crawfor 1.765
Neighborhood_MeadowV 1.824
Neighborhood_Mitchel 1.433
Neighborhood_NPkVill 2.25
Neighborhood_NoRidge 1.439
Neighborhood_NridgHt 1.758
Neighborhood_SWISU 1.677
Neighborhood_SawyerW 1.481
Neighborhood_StoneBr 1.455
Neighborhood_Timber 1.517
Neighborhood_Veenker 1.201
Condition1_Feedr 4.023
Condition1_Norm 6.474
Condition1_PosA 1.607
Condition1_PosN 2.09
Condition1_RRAe 1.952
Condition1_RRAn 2.055
Condition1_RRNe 1.292
Condition1_RRNn 1.784
Condition2_PosA nan
Condition2_PosN

C:\Users\computer\Anaconda3\lib\site-packages\statsmodels\regression\linear_model.py:1386: RuntimeWa
rning: invalid value encountered in double_scalars
  return 1 - self.ssr/self.centered_tss

 1.279
Condition2_RRAe 3.895
Condition2_RRAn nan
Condition2_RRNn nan
BldgType_Duplex 3.399
BldgType_Twnhs 3.494
BldgType_TwnhsE 3.028
HouseStyle_2.5Fin 3.025
HouseStyle_2.5Unf 4.194
OverallCond_2 inf
OverallCond_3 inf
```

```
C:\Users\computer\Anaconda3\lib\site-packages\ipykernel_launcher.py:9: RuntimeWarning: divide by zer
o encountered in double_scalars
  if __name__ == '__main__':
```

OverallCond_4 inf
OverallCond_5 inf
OverallCond_6 inf
OverallCond_7 inf
OverallCond_8 inf
OverallCond_9 inf
RoofStyle_Gambrel 1.432
RoofStyle_Mansard 1.627
RoofStyle_Shed 5.368
RoofMatl_Membran nan
RoofMatl_Metal 1.434
RoofMatl_Roll 1.221
RoofMatl_Tar&Grv 1.629
RoofMatl_WdShake 2.252
RoofMatl_WdShngl 1.258
Exterior1st_AsphShn inf
Exterior1st_BrkComm 2.062
Exterior1st_BrkFace 2.909
Exterior1st_CBlock inf
Exterior1st_ImStucc 1.299
Exterior1st_Stone 1.752
Exterior1st_Stucco 3.902
Exterior1st_WdShing 1.85
Exterior2nd_AsphShn inf
Exterior2nd_Brk Cmn 2.919
Exterior2nd_BrkFace 2.726
Exterior2nd_CBlock inf
Exterior2nd_ImStucc 1.448
Exterior2nd_MetalSd 1.6
Exterior2nd_Other 1.167
Exterior2nd_Stone 2.011
Exterior2nd_Stucco 3.834
Exterior2nd_Wd Shng 1.937
ExterQual_Fa 2.891
ExterCond_Po 1.829
Foundation_CBlock 2.906
Foundation_Slab 3.066
Foundation_Stone 1.541
Foundation_Wood 1.224
BsmtQual_Fa 1.791
BsmtQual_Gd 2.613
BsmtCond_Gd 3.403
BsmtCond_Po inf

BsmtCond_TA 4.763
BsmtExposure_Gd 2.198
BsmtExposure_Mn 1.853
BsmtExposure_No 2.697
BsmtFinType1_BLQ 1.524
BsmtFinType1_GLQ 2.382
BsmtFinType1_LwQ 1.547
BsmtFinType1_Rec 1.598
BsmtFinType2_BLQ 1.229
BsmtFinType2_GLQ 1.481
BsmtFinType2_LwQ 1.503
BsmtFinType2_Rec 1.45
Heating_OthW 1.409
Heating_Wall 1.712
HeatingQC_Fa 1.762
HeatingQC_Gd 1.605
HeatingQC_Po 1.368
HeatingQC_TA 2.416
CentralAir_Y 2.375
Electrical_FuseF 1.991
Electrical_FuseP 2.038
Electrical_Mix inf
Electrical_SBrkr 1.902
KitchenQual_Fa 1.869
KitchenQual_Gd 2.171
Functional_Maj2 2.257
Functional_Min1 1.299
Functional_Min2 1.306
Functional_Mod 1.698
Functional_Sev nan
FireplaceQu_Fa 1.734
FireplaceQu_Gd 4.368
FireplaceQu_Po 1.461
FireplaceQu_TA 4.566
GarageType_Basment 1.628
GarageType_CarPort 1.366
GarageFinish_RFn 1.962
GarageFinish_Unf 2.927
GarageQual_Fa 2.104
GarageQual_Gd 1.753
GarageQual_Po 4.012
GarageCond_Fa 2.048
GarageCond_Gd 1.76

```
GarageCond_Po 2.967
PavedDrive_P 1.71
PavedDrive_Y 2.386
PoolQC_Fa nan
PoolQC_Gd 6.024
Fence_GdWo 1.273
Fence_MnPrv 1.308
Fence_MnWw 1.209
MiscFeature_Othr 1.972
MiscFeature_TenC nan
MoSold_10 3.108
MoSold_11 2.916
MoSold_12 2.359
MoSold_2 2.187
MoSold_3 3.278
MoSold_4 4.064
MoSold_5 4.8
MoSold_6 5.747
MoSold_7 5.517
MoSold_8 3.295
MoSold_9 2.479
YrSold_2007 1.996
YrSold_2008 1.959
YrSold_2009 2.033
YrSold_2010 1.923
SaleType_CWD 1.153
SaleType_Con 1.081
SaleType_ConLD 1.468
SaleType_ConLI 1.21
SaleType_ConLw 1.44
SaleType_Oth 1.28
SaleType_WD 3.18
SaleCondition_AdjLand 1.53
SaleCondition_Alloca 1.919
SaleCondition_Family 1.401
SaleCondition_Normal 3.048
```

Below are the variable having above 5 VIF threshold

In [969]:
```
VIF_5 = ['LotArea','TotRmsAbvGrd','Fireplaces','GarageCars','GarageArea','PoolArea','MSSubClass_75',
'RoofStyle_Shed',
         'BsmtCond_TA','FireplaceQu_TA','PoolQC_Gd' ,'Condition1_Norm','MoSold_6','MoSold_7']
to_keep = [x for x in x_train2 if x not in VIF_5]
#print(to_keep)
x_train2 = x_train2[to_keep]
x_train2.head()
```

Out[969]:

	const	Id	LotFrontage	OverallQual	YearRemodAdd	BsmtUnfSF	LowQualFinSF	BsmtFullBath	BsmtHalfBath
64	1.0	-0.456134	0.000000	0.100076	0.218904	0.015060	-0.015717	0.191553	-0.035941
682	1.0	-0.032557	0.000000	-0.011035	0.202237	0.002327	-0.015717	0.191553	-0.035941
960	1.0	0.157985	-0.103554	-0.122146	0.385571	-0.071504	-0.015717	0.191553	-0.035941
1384	1.0	0.448595	-0.036201	-0.011035	-0.581096	0.029569	-0.015717	-0.141781	-0.035941
1100	1.0	0.253941	-0.036201	-0.455479	-0.581096	-0.728201	-0.015717	-0.141781	-0.035941

5 rows × 189 columns

8.4.3 Building Model after removing VIF above 5

In [970]:

```
# Lets build Linear Regression model using statsmodel
import statsmodels.api as sm

# Building Linear Regression model using OLS

model5 = sm.OLS(y_train1,x_train2).fit()
# Note the Swap of X and Y
# Printing Linear Regression Summary
model5.summary()
```

Out[970]:

OLS Regression Results

Dep. Variable:	SalePrice	**R-squared:**	0.898
Model:	OLS	**Adj. R-squared:**	0.876
Method:	Least Squares	**F-statistic:**	41.53
Date:	Wed, 25 Jul 2018	**Prob (F-statistic):**	1.49e-321
Time:	13:27:17	**Log-Likelihood:**	645.96
No. Observations:	1022	**AIC:**	-933.9
Df Residuals:	843	**BIC:**	-51.54
Df Model:	178		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	10.6788	0.099	107.880	0.000	10.485	10.873
Id	-0.0011	0.017	-0.067	0.947	-0.034	0.031
LotFrontage	0.3744	0.067	5.629	0.000	0.244	0.505
OverallQual	0.9377	0.058	16.307	0.000	0.825	1.051
YearRemodAdd	0.0885	0.024	3.704	0.000	0.042	0.135
BsmtUnfSF	0.0535	0.031	1.729	0.084	-0.007	0.114
LowQualFinSF	-0.0681	0.049	-1.380	0.168	-0.165	0.029
BsmtFullBath	0.2596	0.037	6.961	0.000	0.186	0.333
BsmtHalfBath	0.0577	0.036	1.607	0.108	-0.013	0.128
FullBath	0.4175	0.042	9.991	0.000	0.335	0.499
HalfBath	0.1458	0.023	6.259	0.000	0.100	0.191
BedroomAbvGr	0.2038	0.069	2.934	0.003	0.067	0.340
KitchenAbvGr	-0.0458	0.119	-0.385	0.700	-0.279	0.187

WoodDeckSF	0.0651	0.015	4.462	0.000	0.036	0.094
OpenPorchSF	0.0563	0.017	3.311	0.001	0.023	0.090
EnclosedPorch	-0.0427	0.020	-2.089	0.037	-0.083	-0.003
3SsnPorch	-0.0035	0.048	-0.073	0.941	-0.097	0.090
ScreenPorch	0.0751	0.022	3.478	0.001	0.033	0.118
MSSubClass_160	-0.0791	0.043	-1.846	0.065	-0.163	0.005
MSSubClass_180	-0.1006	0.087	-1.154	0.249	-0.272	0.071
MSSubClass_30	-0.1024	0.029	-3.569	0.000	-0.159	-0.046
MSSubClass_40	0.1524	0.106	1.431	0.153	-0.057	0.361
MSSubClass_70	0.0139	0.030	0.462	0.644	-0.045	0.073
MSSubClass_85	0.0201	0.046	0.441	0.659	-0.069	0.110
MSZoning_RH	0.0085	0.045	0.187	0.852	-0.081	0.097
Street_Pave	0.1576	0.076	2.087	0.037	0.009	0.306
Alley_Pave	0.0952	0.034	2.830	0.005	0.029	0.161
LotShape_IR2	0.0520	0.031	1.694	0.091	-0.008	0.112
LotShape_IR3	0.0922	0.067	1.367	0.172	-0.040	0.225
LotShape_Reg	-0.0058	0.011	-0.502	0.615	-0.028	0.017
LandContour_HLS	-0.0443	0.041	-1.074	0.283	-0.125	0.037
LandContour_Low	-0.0241	0.047	-0.514	0.607	-0.116	0.068
LandContour_Lvl	-0.0019	0.029	-0.065	0.948	-0.059	0.055
Utilities_NoSeWa	-0.3967	0.165	-2.410	0.016	-0.720	-0.074
LotConfig_CulDSac	0.0805	0.024	3.327	0.001	0.033	0.128
LotConfig_FR2	-0.0227	0.030	-0.763	0.446	-0.081	0.036
LotConfig_FR3	-0.2757	0.180	-1.531	0.126	-0.629	0.078
LotConfig_Inside	-0.0068	0.013	-0.506	0.613	-0.033	0.020

LandSlope_Mod	0.0189	0.029	0.660	0.509	-0.037	0.075
LandSlope_Sev	0.1125	0.067	1.668	0.096	-0.020	0.245
Neighborhood_Blueste	0.0037	0.115	0.032	0.974	-0.221	0.228
Neighborhood_BrDale	-0.0657	0.062	-1.057	0.291	-0.188	0.056
Neighborhood_ClearCr	0.1223	0.037	3.316	0.001	0.050	0.195
Neighborhood_Crawfor	0.2149	0.032	6.801	0.000	0.153	0.277
Neighborhood_MeadowV	-0.0438	0.054	-0.808	0.419	-0.150	0.063
Neighborhood_Mitchel	-0.0052	0.028	-0.186	0.852	-0.060	0.050
Neighborhood_NPkVill	-0.0592	0.080	-0.740	0.459	-0.216	0.098
Neighborhood_NoRidge	0.2042	0.031	6.591	0.000	0.143	0.265
Neighborhood_NridgHt	0.1588	0.026	6.011	0.000	0.107	0.211
Neighborhood_SWISU	0.0146	0.043	0.342	0.733	-0.069	0.098
Neighborhood_SawyerW	0.0280	0.026	1.080	0.280	-0.023	0.079
Neighborhood_StoneBr	0.2322	0.043	5.447	0.000	0.149	0.316
Neighborhood_Timber	0.0734	0.032	2.314	0.021	0.011	0.136
Neighborhood_Veenker	0.0980	0.063	1.551	0.121	-0.026	0.222
Condition1_Feedr	-0.0312	0.024	-1.305	0.192	-0.078	0.016
Condition1_PosA	-0.0086	0.065	-0.131	0.896	-0.137	0.119
Condition1_PosN	0.0647	0.045	1.453	0.147	-0.023	0.152
Condition1_RRAe	-0.0762	0.053	-1.452	0.147	-0.179	0.027
Condition1_RRAn	-0.0141	0.039	-0.367	0.714	-0.090	0.061
Condition1_RRNe	-0.1153	0.108	-1.072	0.284	-0.326	0.096
Condition1_RRNn	0.0658	0.089	0.740	0.460	-0.109	0.240
Condition2_PosA	1.384e-16	3.49e-16	0.397	0.692	-5.47e-16	8.23e-16
Condition2_PosN	-1.0770	0.158	-6.831	0.000	-1.386	-0.768

Condition2_RRAe	0.1724	0.157	1.097	0.273	-0.136	0.481
Condition2_RRAn	-7.291e-16	3.31e-16	-2.203	0.028	-1.38e-15	-7.94e-17
Condition2_RRNn	1.481e-15	3.9e-16	3.801	0.000	7.16e-16	2.25e-15
BldgType_Duplex	-0.0137	0.045	-0.304	0.762	-0.102	0.075
BldgType_Twnhs	-0.0495	0.048	-1.022	0.307	-0.144	0.046
BldgType_TwnhsE	-0.0491	0.027	-1.805	0.071	-0.103	0.004
HouseStyle_2.5Fin	0.1092	0.084	1.306	0.192	-0.055	0.273
HouseStyle_2.5Unf	-0.0152	0.060	-0.253	0.801	-0.133	0.103
OverallCond_2	1.2078	0.087	13.849	0.000	1.037	1.379
OverallCond_3	0.9550	0.056	16.929	0.000	0.844	1.066
OverallCond_4	1.0720	0.048	22.540	0.000	0.979	1.165
OverallCond_5	1.0885	0.045	24.236	0.000	1.000	1.177
OverallCond_6	1.1327	0.046	24.759	0.000	1.043	1.222
OverallCond_7	1.1507	0.046	25.126	0.000	1.061	1.241
OverallCond_8	1.0914	0.048	22.527	0.000	0.996	1.186
OverallCond_9	1.1366	0.060	19.031	0.000	1.019	1.254
RoofStyle_Gambrel	-0.0869	0.059	-1.461	0.144	-0.204	0.030
RoofStyle_Mansard	0.0389	0.072	0.538	0.591	-0.103	0.181
RoofMatl_Membran	7.539e-16	2.64e-16	2.858	0.004	2.36e-16	1.27e-15
RoofMatl_Metal	-0.1072	0.167	-0.642	0.521	-0.435	0.221
RoofMatl_Roll	0.0592	0.156	0.380	0.704	-0.246	0.365
RoofMatl_Tar&Grv	-0.0623	0.059	-1.062	0.289	-0.177	0.053
RoofMatl_WdShake	-0.0231	0.086	-0.269	0.788	-0.192	0.146
RoofMatl_WdShngl	0.1669	0.070	2.383	0.017	0.029	0.304
Exterior1st_AsphShn	-0.0367	0.076	-0.485	0.627	-0.185	0.112

Exterior1st_BrkComm	-0.8046	0.201	-4.006	0.000	-1.199	-0.410
Exterior1st_BrkFace	0.1699	0.041	4.118	0.000	0.089	0.251
Exterior1st_CBlock	0.0030	0.082	0.037	0.970	-0.158	0.164
Exterior1st_ImStucc	-0.1014	0.161	-0.631	0.528	-0.417	0.214
Exterior1st_Stone	0.2768	0.130	2.131	0.033	0.022	0.532
Exterior1st_Stucco	0.0695	0.068	1.024	0.306	-0.064	0.203
Exterior1st_WdShing	-0.0282	0.042	-0.673	0.501	-0.110	0.054
Exterior2nd_AsphShn	-0.0367	0.076	-0.485	0.627	-0.185	0.112
Exterior2nd_Brk Cmn	0.2202	0.120	1.829	0.068	-0.016	0.456
Exterior2nd_BrkFace	-0.0997	0.055	-1.802	0.072	-0.208	0.009
Exterior2nd_CBlock	0.0030	0.082	0.037	0.970	-0.158	0.164
Exterior2nd_ImStucc	0.0474	0.060	0.787	0.431	-0.071	0.166
Exterior2nd_MetalSd	0.0040	0.016	0.248	0.804	-0.028	0.036
Exterior2nd_Other	-0.1668	0.153	-1.094	0.274	-0.466	0.133
Exterior2nd_Stone	-0.0617	0.099	-0.621	0.535	-0.257	0.133
Exterior2nd_Stucco	0.0056	0.069	0.081	0.936	-0.130	0.141
Exterior2nd_Wd Shng	-0.0819	0.036	-2.258	0.024	-0.153	-0.011
ExterQual_Fa	-0.0159	0.066	-0.242	0.809	-0.145	0.113
ExterCond_Po	-0.1535	0.189	-0.811	0.418	-0.525	0.218
Foundation_CBlock	-0.0025	0.015	-0.170	0.865	-0.032	0.027
Foundation_Slab	-0.0446	0.051	-0.873	0.383	-0.145	0.056
Foundation_Stone	0.1450	0.077	1.874	0.061	-0.007	0.297
Foundation_Wood	0.1960	0.155	1.262	0.207	-0.109	0.501
BsmtQual_Fa	0.0659	0.039	1.686	0.092	-0.011	0.143
BsmtQual_Gd	-0.0229	0.014	-1.636	0.102	-0.050	0.005

BsmtCond_Gd	0.0015	0.024	0.063	0.950	-0.046	0.049
BsmtCond_Po	0.9291	0.151	6.161	0.000	0.633	1.225
BsmtExposure_Gd	0.0687	0.023	3.025	0.003	0.024	0.113
BsmtExposure_Mn	-0.0029	0.021	-0.137	0.891	-0.045	0.039
BsmtExposure_No	-0.0096	0.015	-0.634	0.526	-0.039	0.020
BsmtFinType1_BLQ	0.0222	0.018	1.253	0.211	-0.013	0.057
BsmtFinType1_GLQ	0.0490	0.015	3.262	0.001	0.019	0.078
BsmtFinType1_LwQ	-0.0022	0.025	-0.089	0.929	-0.052	0.047
BsmtFinType1_Rec	0.0113	0.019	0.589	0.556	-0.026	0.049
BsmtFinType2_BLQ	-0.0023	0.030	-0.076	0.939	-0.061	0.056
BsmtFinType2_GLQ	0.0537	0.057	0.937	0.349	-0.059	0.166
BsmtFinType2_LwQ	0.0207	0.028	0.737	0.461	-0.034	0.076
BsmtFinType2_Rec	0.0043	0.028	0.151	0.880	-0.051	0.060
Heating_OthW	-0.2856	0.163	-1.750	0.081	-0.606	0.035
Heating_Wall	0.1580	0.130	1.214	0.225	-0.097	0.413
HeatingQC_Fa	-0.0797	0.033	-2.422	0.016	-0.144	-0.015
HeatingQC_Gd	-0.0374	0.015	-2.539	0.011	-0.066	-0.008
HeatingQC_Po	-0.0772	0.164	-0.470	0.638	-0.399	0.245
HeatingQC_TA	-0.0509	0.015	-3.430	0.001	-0.080	-0.022
CentralAir_Y	0.1362	0.027	5.096	0.000	0.084	0.189
Electrical_FuseF	-0.0047	0.046	-0.104	0.917	-0.095	0.085
Electrical_FuseP	-0.1933	0.141	-1.373	0.170	-0.470	0.083
Electrical_Mix	-0.9150	0.279	-3.280	0.001	-1.463	-0.367
Electrical_SBrkr	-0.0156	0.021	-0.747	0.456	-0.057	0.025
KitchenQual_Fa	-0.0222	0.036	-0.620	0.536	-0.093	0.048

KitchenQual_Gd	-0.0174	0.013	-1.326	0.185	-0.043	0.008
Functional_Maj2	-0.4493	0.149	-3.015	0.003	-0.742	-0.157
Functional_Min1	0.0722	0.035	2.033	0.042	0.002	0.142
Functional_Min2	-0.0143	0.031	-0.463	0.644	-0.075	0.046
Functional_Mod	-0.0004	0.055	-0.008	0.994	-0.109	0.108
Functional_Sev	-5.086e-16	1.69e-16	-3.012	0.003	-8.4e-16	-1.77e-16
FireplaceQu_Fa	0.0321	0.031	1.035	0.301	-0.029	0.093
FireplaceQu_Gd	0.0320	0.013	2.514	0.012	0.007	0.057
FireplaceQu_Po	0.0558	0.039	1.442	0.150	-0.020	0.132
GarageType_Basment	-0.0157	0.050	-0.317	0.751	-0.113	0.082
GarageType_CarPort	-0.1119	0.073	-1.540	0.124	-0.254	0.031
GarageFinish_RFn	0.0070	0.013	0.532	0.595	-0.019	0.033
GarageFinish_Unf	0.0123	0.015	0.832	0.406	-0.017	0.041
GarageQual_Fa	-0.0784	0.035	-2.256	0.024	-0.147	-0.010
GarageQual_Gd	-0.0003	0.062	-0.005	0.996	-0.122	0.122
GarageQual_Po	-0.1657	0.161	-1.028	0.304	-0.482	0.151
GarageCond_Fa	-0.0458	0.038	-1.196	0.232	-0.121	0.029
GarageCond_Gd	0.0349	0.075	0.464	0.643	-0.112	0.182
GarageCond_Po	0.1587	0.097	1.629	0.104	-0.033	0.350
PavedDrive_P	0.0422	0.043	0.973	0.331	-0.043	0.127
PavedDrive_Y	0.0566	0.024	2.389	0.017	0.010	0.103
PoolQC_Fa	-5.049e-17	1.05e-16	-0.481	0.630	-2.56e-16	1.55e-16
Fence_GdWo	-0.0775	0.027	-2.910	0.004	-0.130	-0.025
Fence_MnPrv	-0.0392	0.017	-2.333	0.020	-0.072	-0.006
Fence_MnWw	-0.0075	0.059	-0.128	0.898	-0.123	0.108

MiscFeature_Othr	-0.1329	0.132	-1.007	0.314	-0.392	0.126
MiscFeature_TenC	9.72e-18	7.62e-17	0.128	0.899	-1.4e-16	1.59e-16
MoSold_10	-0.0041	0.022	-0.186	0.852	-0.047	0.039
MoSold_11	-0.0036	0.022	-0.166	0.868	-0.046	0.039
MoSold_12	0.0128	0.026	0.489	0.625	-0.039	0.064
MoSold_2	-0.0070	0.027	-0.263	0.792	-0.059	0.045
MoSold_3	0.0192	0.020	0.973	0.331	-0.020	0.058
MoSold_4	0.0012	0.017	0.068	0.946	-0.033	0.035
MoSold_5	0.0217	0.015	1.412	0.158	-0.008	0.052
MoSold_8	0.0115	0.020	0.589	0.556	-0.027	0.050
MoSold_9	-0.0224	0.025	-0.889	0.374	-0.072	0.027
YrSold_2007	-0.0024	0.015	-0.162	0.871	-0.032	0.027
YrSold_2008	-0.0120	0.015	-0.796	0.426	-0.042	0.018
YrSold_2009	-0.0074	0.015	-0.488	0.625	-0.037	0.022
YrSold_2010	-0.0415	0.018	-2.324	0.020	-0.077	-0.006
SaleType_CWD	0.1624	0.149	1.089	0.276	-0.130	0.455
SaleType_Con	-0.1282	0.147	-0.873	0.383	-0.416	0.160
SaleType_ConLD	0.0174	0.070	0.250	0.803	-0.119	0.154
SaleType_ConLI	-0.2174	0.077	-2.813	0.005	-0.369	-0.066
SaleType_ConLw	-0.1921	0.075	-2.576	0.010	-0.339	-0.046
SaleType_Oth	-0.0750	0.112	-0.668	0.504	-0.295	0.145
SaleType_WD	-0.1639	0.024	-6.900	0.000	-0.211	-0.117
SaleCondition_AdjLand	0.0095	0.101	0.094	0.925	-0.188	0.207
SaleCondition_Alloca	0.0957	0.071	1.354	0.176	-0.043	0.234
SaleCondition_Family	0.0967	0.050	1.919	0.055	-0.002	0.196

SaleCondition_Normal	0.0848	0.020	4.200	0.000	0.045	0.124

Omnibus:	51.427	**Durbin-Watson:**	1.922
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	110.123
Skew:	-0.307	**Prob(JB):**	1.22e-24
Kurtosis:	4.487	**Cond. No.**	1.06e+16

In [971]:

```
VIF_5 = ['LotArea','TotRmsAbvGrd','Fireplaces','GarageCars','GarageArea','PoolArea','MSSubClass_75',
'RoofStyle_Shed',
         'BsmtCond_TA','FireplaceQu_TA','PoolQC_Gd' ,'Condition1_Norm','MoSold_6','MoSold_7']
to_keep = [x for x in x_test2 if x not in VIF_5]
#print(to_keep)
x_test2 = x_test2[to_keep]
x_test2.head()
```

Out[971]:

	const	Id	LotFrontage	OverallQual	YearRemodAdd	BsmtUnfSF	LowQualFinSF	BsmtFullBath	BsmtHalfBath	
529	1.0	-0.137423	0.000000	-0.011035	-0.164429	0.136304	-0.015717	0.191553	-0.035941	C
491	1.0	-0.163468	0.065799	-0.011035	-0.581096	-0.022164	-0.015717	0.191553	-0.035941	-
459	1.0	-0.185401	0.000000	-0.122146	-0.581096	0.079290	-0.015717	0.191553	-0.035941	-
279	1.0	-0.308773	0.084153	0.100076	-0.131096	0.128498	-0.015717	-0.141781	-0.035941	C
655	1.0	-0.051062	-0.419834	-0.011035	-0.231096	0.079535	-0.015717	-0.141781	-0.035941	-

5 rows × 189 columns

In [972]:
```
predictions5 = model5.predict(x_test2)
tmp5 = pd.Series({'Model': "LRM after removing VIF above 5",
                  'R-Squared Value' : model5.rsquared,
                  'Adj.R-Squared Value': model5.rsquared_adj,
                  'RMSE': rmse(predictions5, y_test1)})

model5_report = models_report.append(tmp5, ignore_index = True)
model5_report
```

Out[972]:

	Model	R-Squared Value	Adj.R-Squared Value	RMSE
0	LRM after removing VIF above 5	0.897625	0.876009	0.187339

8.5 Removing Variable based on Insignificant Variables using P-value

In [973]:
```
X = x_train2
Y = y_train1
```

In [974]:
```python
def feature_selection(X, Y,
                      initial_list=[],
                      threshold_in=0.05,
                      threshold_out = 0.05,
                      verbose=True):
    """ Perform a forward-backward feature selection
    based on p-value from statsmodels.api.OLS
    Arguments:
        X - pandas.DataFrame with candidate features
        y - list-like with the target
        initial_list - list of features to start with (column names of X)
        threshold_in - include a feature if its p-value < threshold_in
        threshold_out - exclude a feature if its p-value > threshold_out
        verbose - whether to print the sequence of inclusions and exclusions
    Returns: list of selected features
    Always set threshold_in < threshold_out to avoid infinite looping.
    See https://en.wikipedia.org/wiki/Stepwise_regression for the details
    """
    included = list(initial_list)
    while True:
        changed=False
        # forward step
        excluded = list(set(X.columns)-set(included))
        new_pval = pd.Series(index=excluded)
        for new_column in excluded:
            model = sm.OLS(Y, sm.add_constant((X[included+[new_column]]))).fit()
            new_pval[new_column] = model.pvalues[new_column]
        best_pval = new_pval.min()
        if best_pval < threshold_in:
            best_feature = new_pval.argmin()
            included.append(best_feature)
            changed=True
            if verbose:
                print('Add  {:30} with p-value {:.6}'.format(best_feature, best_pval))

        # backward step
        model = sm.OLS(Y, sm.add_constant(pd.DataFrame(X[included]))).fit()
        # use all coefs except intercept
        pvalues = model.pvalues.iloc[1:]
        worst_pval = pvalues.max() # null if pvalues is empty
        if worst_pval > threshold_out:
            changed=True
```

```
            worst_feature = pvalues.argmax()
            included.remove(worst_feature)
            if verbose:
                print('Drop {:30} with p-value {:.6}'.format(worst_feature, worst_pval))
        if not changed:
            break
    return included

result = feature_selection(X, Y)

print('resulting features:')
print(result)
```

```
C:\Users\computer\Anaconda3\lib\site-packages\statsmodels\base\model.py:1036: RuntimeWarning: invali
d value encountered in true_divide
  return self.params / self.bse
C:\Users\computer\Anaconda3\lib\site-packages\scipy\stats\_distn_infrastructure.py:879: RuntimeWarni
ng: invalid value encountered in greater
  return (self.a < x) & (x < self.b)
C:\Users\computer\Anaconda3\lib\site-packages\scipy\stats\_distn_infrastructure.py:879: RuntimeWarni
ng: invalid value encountered in less
  return (self.a < x) & (x < self.b)
C:\Users\computer\Anaconda3\lib\site-packages\scipy\stats\_distn_infrastructure.py:1821: RuntimeWarn
ing: invalid value encountered in less_equal
  cond2 = cond0 & (x <= self.a)
C:\Users\computer\Anaconda3\lib\site-packages\ipykernel_launcher.py:30: FutureWarning: 'argmin' is d
eprecated. Use 'idxmin' instead. The behavior of 'argmin' will be corrected to return the positional
minimum in the future. Use 'series.values.argmin' to get the position of the minimum now.
```

```
Add  const                        with p-value 0.0
Add  OverallQual                  with p-value 6.83555e-235
Add  LotFrontage                  with p-value 1.60411e-31
Add  FullBath                     with p-value 3.99905e-25
Add  BsmtFullBath                 with p-value 4.76676e-27
Add  CentralAir_Y                 with p-value 2.60409e-21
Add  HalfBath                     with p-value 5.38676e-14
Add  Condition2_PosN              with p-value 1.24722e-09
Add  Neighborhood_Crawfor         with p-value 2.19761e-09
Add  LotConfig_CulDSac            with p-value 2.05224e-10
Add  WoodDeckSF                   with p-value 1.10802e-08
Add  YearRemodAdd                 with p-value 8.26616e-07
Add  Neighborhood_ClearCr         with p-value 1.41076e-06
Add  Exterior1st_BrkFace          with p-value 5.60284e-06
Add  ScreenPorch                  with p-value 2.14382e-05
Add  Neighborhood_NridgHt         with p-value 2.59856e-05
Add  Neighborhood_NoRidge         with p-value 1.95135e-06
Add  OverallCond_3                with p-value 1.12707e-05
Add  MSSubClass_30                with p-value 9.50052e-06
Add  SaleType_WD                  with p-value 6.73814e-06
Add  OverallCond_5                with p-value 1.03874e-05
Add  BsmtExposure_Gd              with p-value 6.00888e-05
Add  Neighborhood_StoneBr         with p-value 7.3234e-05
Add  Functional_Maj2              with p-value 0.000184661
Add  Exterior2nd_Wd Shng          with p-value 0.000440383
Add  FireplaceQu_Gd               with p-value 0.000449202
Add  Exterior1st_BrkComm          with p-value 0.000874312
Add  MSSubClass_160               with p-value 0.000895898
Add  Alley_Pave                   with p-value 0.000652212
Add  OpenPorchSF                  with p-value 0.00321711
Add  PavedDrive_Y                 with p-value 0.00314738
Add  BedroomAbvGr                 with p-value 0.00395003
Add  OverallCond_4                with p-value 0.00468097
Add  Heating_OthW                 with p-value 0.00891206
Add  Neighborhood_Timber          with p-value 0.00984877
Add  SaleCondition_Normal         with p-value 0.0105163
Add  SaleType_ConLI               with p-value 0.00720382
Add  YrSold_2010                  with p-value 0.0111472
Add  BsmtUnfSF                    with p-value 0.0161466
Add  LotShape_IR2                 with p-value 0.0140732
Add  GarageQual_Fa                with p-value 0.0176694
Add  Utilities_NoSeWa             with p-value 0.021616
Add  BsmtHalfBath                 with p-value 0.0185948
```

```
Add  OverallCond_8                  with p-value 0.0202107
Add  SaleType_ConLw                 with p-value 0.0290347
Add  Fence_MnPrv                    with p-value 0.029409
Add  Fence_GdWo                     with p-value 0.0204549
Add  RoofMatl_WdShngl               with p-value 0.0307887
Add  HeatingQC_TA                   with p-value 0.043081
Add  Exterior2nd_Brk Cmn            with p-value 0.0472503
Add  RoofStyle_Gambrel              with p-value 0.0458707
resulting features:
['const', 'OverallQual', 'LotFrontage', 'FullBath', 'BsmtFullBath', 'CentralAir_Y', 'HalfBath', 'Con
dition2_PosN', 'Neighborhood_Crawfor', 'LotConfig_CulDSac', 'WoodDeckSF', 'YearRemodAdd', 'Neighborh
ood_ClearCr', 'Exterior1st_BrkFace', 'ScreenPorch', 'Neighborhood_NridgHt', 'Neighborhood_NoRidge',
'OverallCond_3', 'MSSubClass_30', 'SaleType_WD', 'OverallCond_5', 'BsmtExposure_Gd', 'Neighborhood_S
toneBr', 'Functional_Maj2', 'Exterior2nd_Wd Shng', 'FireplaceQu_Gd', 'Exterior1st_BrkComm', 'MSSubCl
ass_160', 'Alley_Pave', 'OpenPorchSF', 'PavedDrive_Y', 'BedroomAbvGr', 'OverallCond_4', 'Heating_Oth
W', 'Neighborhood_Timber', 'SaleCondition_Normal', 'SaleType_ConLI', 'YrSold_2010', 'BsmtUnfSF', 'Lo
tShape_IR2', 'GarageQual_Fa', 'Utilities_NoSeWa', 'BsmtHalfBath', 'OverallCond_8', 'SaleType_ConLw',
'Fence_MnPrv', 'Fence_GdWo', 'RoofMatl_WdShngl', 'HeatingQC_TA', 'Exterior2nd_Brk Cmn', 'RoofStyle_G
ambrel']
```

In [1146]:
```
df_train= x_train2.filter(['const', 'OverallQual', 'LotFrontage', 'FullBath', 'BsmtFullBath', 'Centr
alAir_Y', 'HalfBath', 'Condition2_PosN', 'Neighborhood_Crawfor', 'LotConfig_CulDSac', 'WoodDeckSF',
'YearRemodAdd', 'Neighborhood_ClearCr', 'Exterior1st_BrkFace', 'ScreenPorch', 'Neighborhood_NridgHt'
, 'Neighborhood_NoRidge', 'OverallCond_3', 'MSSubClass_30', 'SaleType_WD', 'OverallCond_5', 'BsmtExp
osure_Gd', 'Neighborhood_StoneBr', 'Functional_Maj2', 'Exterior2nd_Wd Shng', 'FireplaceQu_Gd', 'Exte
rior1st_BrkComm', 'MSSubClass_160', 'Alley_Pave', 'OpenPorchSF', 'PavedDrive_Y', 'BedroomAbvGr', 'Ov
erallCond_4', 'Heating_OthW', 'Neighborhood_Timber', 'SaleCondition_Normal', 'SaleType_ConLI', 'YrSo
ld_2010', 'BsmtUnfSF', 'LotShape_IR2', 'GarageQual_Fa', 'Utilities_NoSeWa', 'BsmtHalfBath', 'Overall
Cond_8', 'SaleType_ConLw', 'Fence_MnPrv', 'Fence_GdWo', 'RoofMatl_WdShngl', 'HeatingQC_TA', 'Exterio
r2nd_Brk Cmn', 'RoofStyle_Gambrel'])
df_test= x_test2.filter(['const', 'OverallQual', 'LotFrontage', 'FullBath', 'BsmtFullBath', 'Central
Air_Y', 'HalfBath', 'Condition2_PosN', 'Neighborhood_Crawfor', 'LotConfig_CulDSac', 'WoodDeckSF', 'Y
earRemodAdd', 'Neighborhood_ClearCr', 'Exterior1st_BrkFace', 'ScreenPorch', 'Neighborhood_NridgHt',
'Neighborhood_NoRidge', 'OverallCond_3', 'MSSubClass_30', 'SaleType_WD', 'OverallCond_5', 'BsmtExpos
ure_Gd', 'Neighborhood_StoneBr', 'Functional_Maj2', 'Exterior2nd_Wd Shng', 'FireplaceQu_Gd', 'Exteri
or1st_BrkComm', 'MSSubClass_160', 'Alley_Pave', 'OpenPorchSF', 'PavedDrive_Y', 'BedroomAbvGr', 'Over
allCond_4', 'Heating_OthW', 'Neighborhood_Timber', 'SaleCondition_Normal', 'SaleType_ConLI', 'YrSold
_2010', 'BsmtUnfSF', 'LotShape_IR2', 'GarageQual_Fa', 'Utilities_NoSeWa', 'BsmtHalfBath', 'OverallCo
nd_8', 'SaleType_ConLw', 'Fence_MnPrv', 'Fence_GdWo', 'RoofMatl_WdShngl', 'HeatingQC_TA', 'Exterior2
nd_Brk Cmn', 'RoofStyle_Gambrel'])
df_train.isna().sum().sum(), df_test.isna().sum().sum()
```

Out[1146]: (0, 0)

8.5.1 Building Model after removing insignificant variables using p-value

In [976]:
```
# Building Linear Regression model using OLS

model6 = sm.OLS(y_train1,df_train).fit()
# Note the Swap of X and Y
model6.summary()
```

Out[976]:

OLS Regression Results

Dep. Variable:	SalePrice	**R-squared:**	0.882
Model:	OLS	**Adj. R-squared:**	0.876
Method:	Least Squares	**F-statistic:**	144.9
Date:	Wed, 25 Jul 2018	**Prob (F-statistic):**	0.00
Time:	13:30:22	**Log-Likelihood:**	572.72
No. Observations:	1022	**AIC:**	-1043.
Df Residuals:	971	**BIC:**	-792.0
Df Model:	50		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	11.9214	0.029	413.585	0.000	11.865	11.978
OverallQual	1.0235	0.048	21.113	0.000	0.928	1.119
LotFrontage	0.4657	0.051	9.068	0.000	0.365	0.566
FullBath	0.4014	0.036	11.173	0.000	0.331	0.472
BsmtFullBath	0.3198	0.030	10.516	0.000	0.260	0.379
CentralAir_Y	0.1512	0.021	7.350	0.000	0.111	0.192
HalfBath	0.1486	0.021	7.159	0.000	0.108	0.189
Condition2_PosN	-0.9963	0.145	-6.884	0.000	-1.280	-0.712
Neighborhood_Crawfor	0.2280	0.026	8.803	0.000	0.177	0.279
LotConfig_CulDSac	0.1008	0.019	5.296	0.000	0.063	0.138
WoodDeckSF	0.0730	0.013	5.687	0.000	0.048	0.098
YearRemodAdd	0.0951	0.019	5.026	0.000	0.058	0.132
Neighborhood_ClearCr	0.1312	0.031	4.214	0.000	0.070	0.192

Exterior1st_BrkFace	0.1181	0.026	4.623	0.000	0.068	0.168
ScreenPorch	0.0770	0.020	3.839	0.000	0.038	0.116
Neighborhood_NridgHt	0.1601	0.024	6.737	0.000	0.113	0.207
Neighborhood_NoRidge	0.1966	0.029	6.840	0.000	0.140	0.253
OverallCond_3	-0.1696	0.038	-4.459	0.000	-0.244	-0.095
MSSubClass_30	-0.0888	0.024	-3.669	0.000	-0.136	-0.041
SaleType_WD	-0.1270	0.020	-6.452	0.000	-0.166	-0.088
OverallCond_5	-0.0635	0.011	-5.641	0.000	-0.086	-0.041
BsmtExposure_Gd	0.0653	0.018	3.687	0.000	0.031	0.100
Neighborhood_StoneBr	0.2098	0.038	5.562	0.000	0.136	0.284
Functional_Maj2	-0.4011	0.105	-3.838	0.000	-0.606	-0.196
Exterior2nd_Wd Shng	-0.0972	0.027	-3.534	0.000	-0.151	-0.043
FireplaceQu_Gd	0.0328	0.012	2.782	0.006	0.010	0.056
Exterior1st_BrkComm	-0.7530	0.171	-4.391	0.000	-1.090	-0.416
MSSubClass_160	-0.1273	0.028	-4.547	0.000	-0.182	-0.072
Alley_Pave	0.1186	0.029	4.140	0.000	0.062	0.175
OpenPorchSF	0.0526	0.016	3.390	0.001	0.022	0.083
PavedDrive_Y	0.0535	0.018	2.937	0.003	0.018	0.089
BedroomAbvGr	0.2133	0.056	3.821	0.000	0.104	0.323
OverallCond_4	-0.0828	0.023	-3.528	0.000	-0.129	-0.037
Heating_OthW	-0.4132	0.146	-2.821	0.005	-0.701	-0.126
Neighborhood_Timber	0.0887	0.028	3.156	0.002	0.034	0.144
SaleCondition_Normal	0.0533	0.016	3.332	0.001	0.022	0.085
SaleType_ConLI	-0.2078	0.074	-2.819	0.005	-0.352	-0.063
YrSold_2010	-0.0356	0.013	-2.653	0.008	-0.062	-0.009

BsmtUnfSF	0.0551	0.022	2.556	0.011	0.013	0.097
LotShape_IR2	0.0622	0.028	2.223	0.026	0.007	0.117
GarageQual_Fa	-0.0662	0.026	-2.508	0.012	-0.118	-0.014
Utilities_NoSeWa	-0.4172	0.149	-2.791	0.005	-0.711	-0.124
BsmtHalfBath	0.0695	0.032	2.146	0.032	0.006	0.133
OverallCond_8	-0.0528	0.022	-2.384	0.017	-0.096	-0.009
SaleType_ConLw	-0.1582	0.068	-2.326	0.020	-0.292	-0.025
Fence_MnPrv	-0.0346	0.016	-2.215	0.027	-0.065	-0.004
Fence_GdWo	-0.0569	0.025	-2.320	0.021	-0.105	-0.009
RoofMatl_WdShngl	0.1517	0.066	2.306	0.021	0.023	0.281
HeatingQC_TA	-0.0231	0.011	-2.059	0.040	-0.045	-0.001
Exterior2nd_Brk Cmn	0.1731	0.086	2.015	0.044	0.004	0.342
RoofStyle_Gambrel	-0.1053	0.053	-1.999	0.046	-0.209	-0.002

Omnibus:	51.109	**Durbin-Watson:**	1.930
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	121.015
Skew:	-0.267	**Prob(JB):**	5.27e-27
Kurtosis:	4.599	**Cond. No.**	88.1

In [977]:

```
# prediction on test data
predictions6  = model6.predict(df_test)

tmp6 = pd.Series({'Model': " LRM after removing Insignificant Variables",
                  'R-Squared Value' : model6.rsquared,
                  'Adj.R-Squared Value': model6.rsquared_adj,
                  'RMSE': rmse(predictions6, y_test1)})

model6_report = models_report.append(tmp6, ignore_index = True)
model6_report
```

Out[977]:

	Model	R-Squared Value	Adj.R-Squared Value	RMSE
0	LRM after removing Insignificant Variables	0.881849	0.875765	0.178451

Multiplicative Interactions

“*” will also include the individual columns that were multiplied together

For Example

("y ~ a * b" , data = df) you'll have 3 independent variables which is the results of "a" multiply by "b" + "a" itself + "b" itself

In [1116]:

```
target = pd.DataFrame(y_train1,columns=['SalePrice'])
data = pd.concat([x_train2, target], axis=1)

# Building Linear Regression model using OLS
import statsmodels.formula.api as smf
interaction = smf.ols(formula= 'SalePrice ~ OverallQual * YearRemodAdd *  BsmtFullBath', data = data
).fit()
# Note the Swap of X and Y
interaction.summary()
```

Out[1116]:

OLS Regression Results

Dep. Variable:	SalePrice	**R-squared:**	0.697
Model:	OLS	**Adj. R-squared:**	0.695
Method:	Least Squares	**F-statistic:**	332.8
Date:	Wed, 25 Jul 2018	**Prob (F-statistic):**	1.42e-257
Time:	19:14:30	**Log-Likelihood:**	90.995
No. Observations:	1022	**AIC:**	-166.0
Df Residuals:	1014	**BIC:**	-126.6
Df Model:	7		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	12.0221	0.008	1471.315	0.000	12.006	12.038
OverallQual	1.8759	0.058	32.411	0.000	1.762	1.989
YearRemodAdd	0.1947	0.026	7.388	0.000	0.143	0.246
OverallQual:YearRemodAdd	0.2085	0.151	1.383	0.167	-0.087	0.504
BsmtFullBath	0.3998	0.047	8.487	0.000	0.307	0.492
OverallQual:BsmtFullBath	0.0595	0.333	0.179	0.858	-0.593	0.712
YearRemodAdd:BsmtFullBath	-0.0965	0.158	-0.609	0.542	-0.407	0.214
OverallQual:YearRemodAdd:BsmtFullBath	-1.5047	0.908	-1.657	0.098	-3.286	0.277

Omnibus:	38.984	**Durbin-Watson:**	2.008
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	61.827
Skew:	-0.321	**Prob(JB):**	3.75e-14
Kurtosis:	4.020	**Cond. No.**	132.

Diagnostic Plot

1. Residual plot

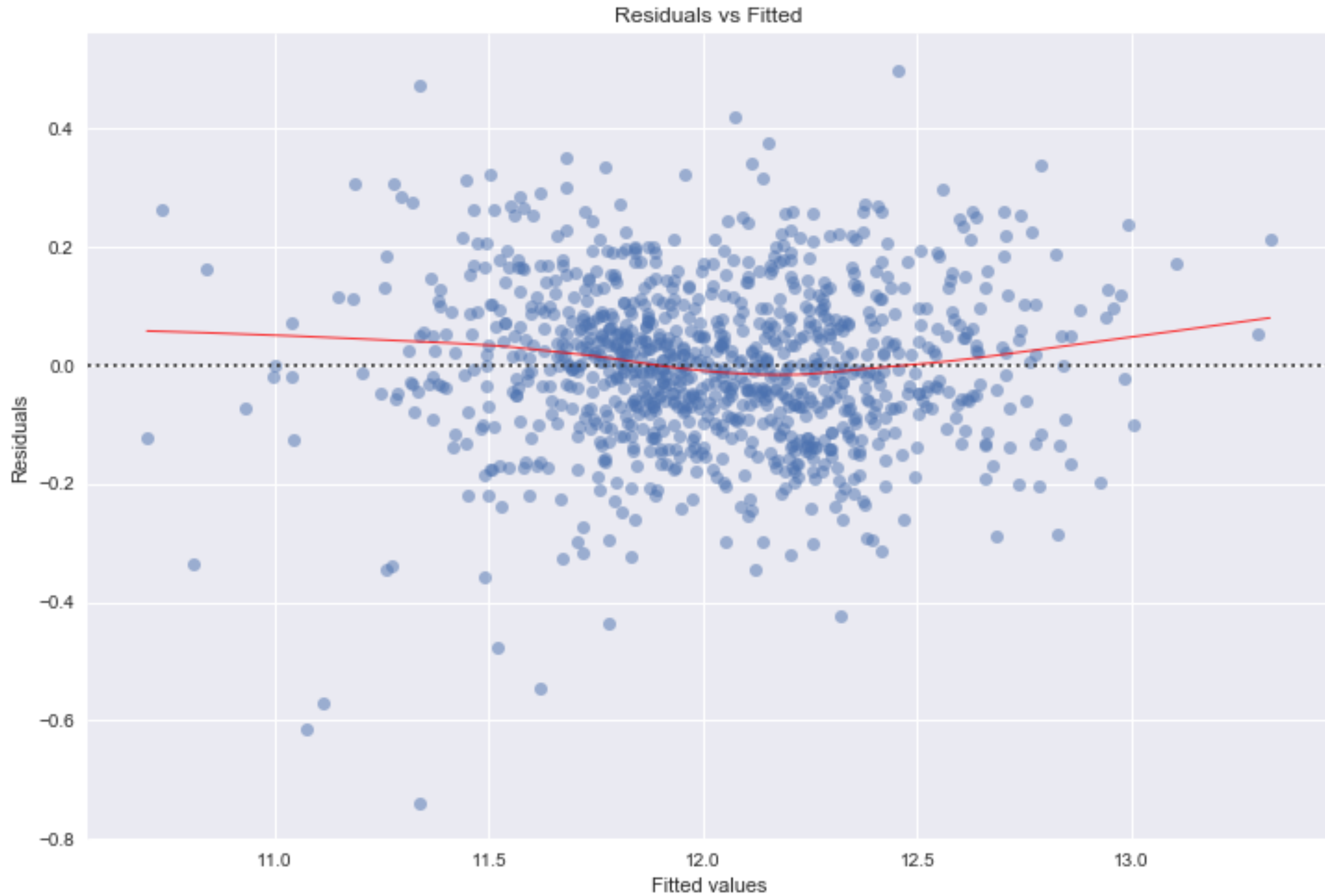
A scatterplot of fitted values against residuals, with a “locally weighted scatterplot smoothing (lowess)” regression line showing any apparent trend.

This one can be easily plotted using seaborn residplot with fitted values as x parameter, and the dependent variable as y. lowess=True makes sure the lowess regression line is drawn. Additional parameters are passed to underlying matplotlib scatter and line functions using scatter_kws and line_kws, also titles and labels are set using matplotlib methods.

In [1119]:

```python
plot_lm_1 = plt.figure(1)
plot_lm_1.set_figheight(8)
plot_lm_1.set_figwidth(12)

# fitted values (need a constant term for intercept)
model_fitted_y = model6.fittedvalues

plot_lm_1.axes[0] = sns.residplot(model_fitted_y, 'SalePrice', data=data,
                          lowess=True,
                          scatter_kws={'alpha': 0.5},
                          line_kws={'color': 'red', 'lw': 1, 'alpha': 0.8})

plot_lm_1.axes[0].set_title('Residuals vs Fitted')
plot_lm_1.axes[0].set_xlabel('Fitted values')
plot_lm_1.axes[0].set_ylabel('Residuals')
```

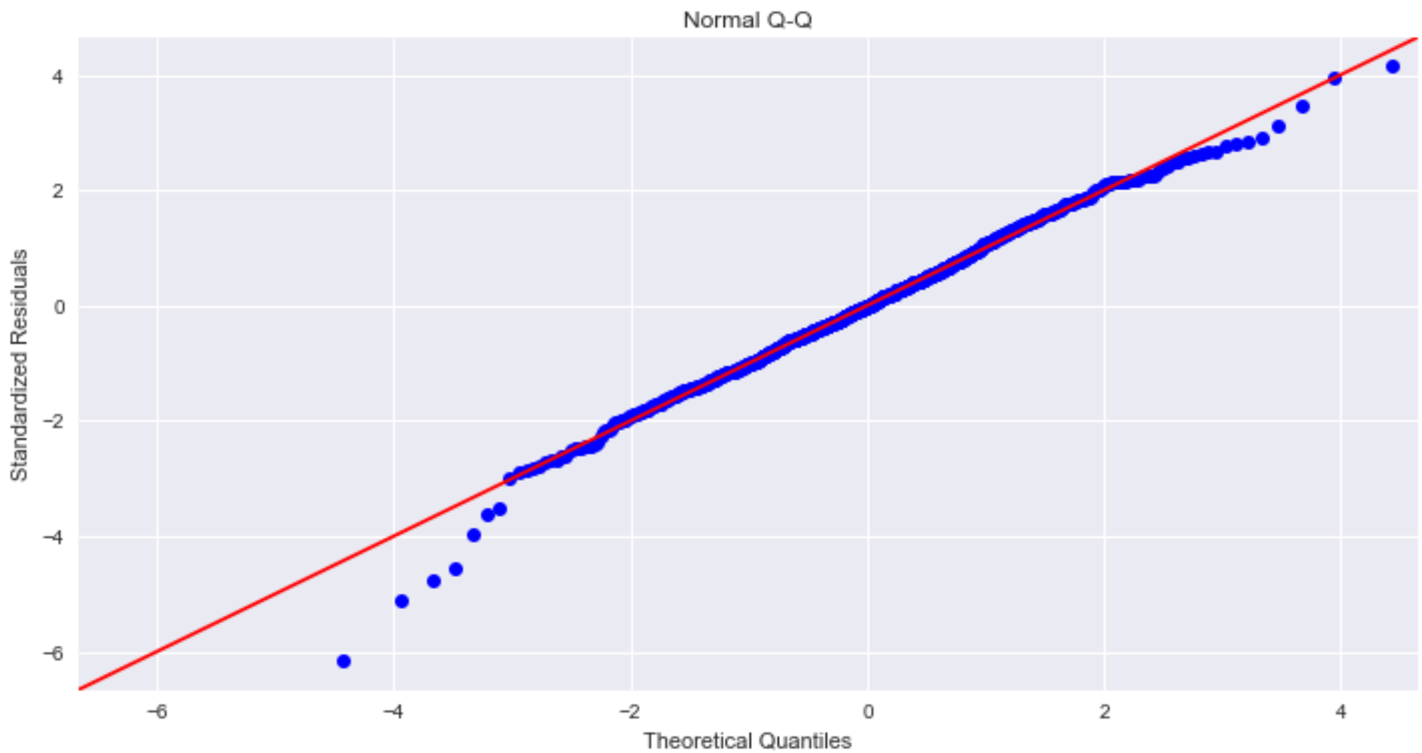
```
Out[1119]: Text(0,0.5,'Residuals')
```



2. QQ Plot

This plot shows if residuals are normally distributed.This plots the standardized (z-score) residuals against the theoretical normal quantiles. Anything quite off the diagonal lines may be a concern for further investigation.

In [1129]:
```
res = model6.resid
import scipy.stats as stats
fig = sm.qqplot(res, stats.t, fit=True, line='45')
plt.title('Normal Q-Q')
plt.xlabel('Theoretical Quantiles')
plt.ylabel('Standardized Residuals');
plt.show()
```



3. Scale-Location Plot

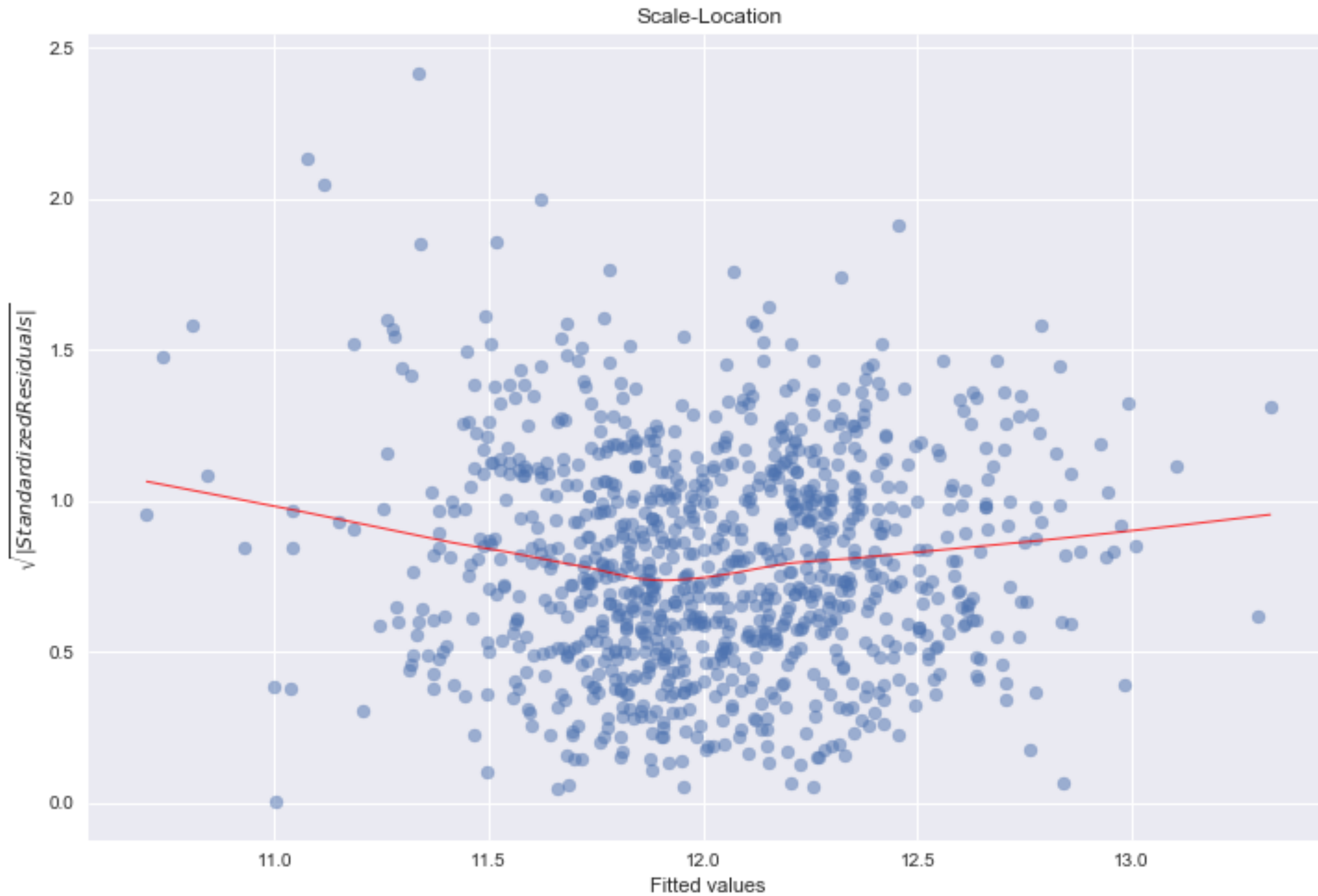
This is another residual plot, showing their spread, which you can use to assess heteroscedasticity.

It’s essentially a scatter plot of absolute square-rooted normalized residuals and fitted values, with a lowess regression line.

In [1133]:

```
# normalized residuals
model_norm_residuals = model6.get_influence().resid_studentized_internal
# absolute squared normalized residuals
model_norm_residuals_abs_sqrt = np.sqrt(np.abs(model_norm_residuals))

plot_lm_3 = plt.figure(3)
plot_lm_3.set_figheight(8)
plot_lm_3.set_figwidth(12)
plt.scatter(model_fitted_y, model_norm_residuals_abs_sqrt, alpha=0.5)
sns.regplot(model_fitted_y, model_norm_residuals_abs_sqrt,
            scatter=False,
            ci=False,
            lowess=True,
            line_kws={'color': 'red', 'lw': 1, 'alpha': 0.8})

plot_lm_3.axes[0].set_title('Scale-Location')
plot_lm_3.axes[0].set_xlabel('Fitted values')
plot_lm_3.axes[0].set_ylabel('$\sqrt{|Standardized Residuals|}$');
```

```
C:\Users\computer\Anaconda3\lib\site-packages\statsmodels\stats\outliers_influence.py:309: RuntimeWa
rning: invalid value encountered in sqrt
  return  self.results.resid / sigma / np.sqrt(1 - hii)
```



4. Leverage plot

This plot shows if any outliers have influence over the regression fit. Anything outside the group and outside “Cook’s Distance” lines, may have an influential effect on model fit.

In [ ]:

```
plot_lm_4 = plt.figure(4)
plot_lm_4.set_figheight(8)
plot_lm_4.set_figwidth(12)

# cook's distance, from statsmodels internals
model_cooks = model6.get_influence().cooks_distance[0]

plt.scatter(model_leverage, model_norm_residuals, alpha=0.5)
sns.regplot(model_leverage, model_norm_residuals,
            scatter=False,
            ci=False,
            lowess=True,
            line_kws={'color': 'red', 'lw': 1, 'alpha': 0.8})

plot_lm_4.axes[0].set_xlim(0, 0.20)
plot_lm_4.axes[0].set_ylim(-3, 5)
plot_lm_4.axes[0].set_title('Residuals vs Leverage')
plot_lm_4.axes[0].set_xlabel('Leverage')
plot_lm_4.axes[0].set_ylabel('Standardized Residuals')

# annotations
leverage_top_3 = np.flip(np.argsort(model_cooks), 0)[:3]

for i in leverage_top_3:
    plot_lm_4.axes[0].annotate(i,
                               xy=(model_leverage[i],
                                   model_norm_residuals[i]))

# shenanigans for cook's distance contours
def graph(formula, x_range, label=None):
    x = x_range
    y = formula(x)
    plt.plot(x, y, label=label, lw=1, ls='--', color='red')

p = len(model_fit.params) # number of model parameters

graph(lambda x: np.sqrt((0.5 * p * (1 - x)) / x),
      np.linspace(0.001, 0.200, 50),
      'Cook\'s distance') # 0.5 line
graph(lambda x: np.sqrt((1 * p * (1 - x)) / x),
      np.linspace(0.001, 0.200, 50)) # 1 line
plt.legend(loc='upper right');
```

In [978]:

```
# Comparison of various model
cols = ['Model', 'R-Squared Value', 'Adj.R-Squared Value', 'RMSE']
clas_model = pd.DataFrame(columns = cols)
clas_model = clas_model.append([model1_report,model2_report,model3_report,model4_report,model5_report
,model6_report])
clas_model
```

Out[978]:

	Model	R-Squared Value	Adj.R-Squared Value	RMSE
0	Base Linear Regression Model	0.958527	0.943989	0.464362
0	Linear Regression Model with Constant	0.958527	0.943989	0.167912
0	LRM after removing VIF above 100	0.957579	0.943383	0.172947
0	LRM after removing VIF above 10	0.921361	0.903147	0.179040
0	LRM after removing VIF above 5	0.897625	0.876009	0.187339
0	LRM after removing Insignificant Variables	0.881849	0.875765	0.178451

Occam’s Razor principles can be stated as “when presented with competing hypothetical answers to a problem, one should select the one that makes the fewest assumptions”. According to Occam's Razor principle we consider Linear Regression Model after removing insignificant variables with 0.88 R-Squared value and RMSE of 0.1784 to make the model perform better with new data as well.