In [1]:
```python
import pandas as pd
import math
import statsmodels.api as sm
import numpy as np
import scipy
import seaborn as sns
from matplotlib import pyplot as plt
from scipy.stats import pearsonr
from sklearn.decomposition import PCA
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error as MSE
from sklearn.metrics import mean_absolute_error as MAE
from sklearn.model_selection import train_test_split
```

Note that the first 1460 samples in cleaned_data.csv are training and cleaned_data.csv does not have sale price in it.

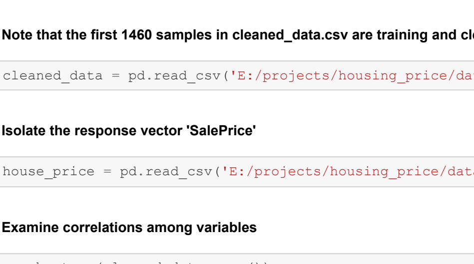
In [2]:
```python
cleaned_data = pd.read_csv('E:/projects/housing_price/data/cleaned_data.csv',index_col=0)
```

Isolate the response vector 'SalePrice'

In [3]:
```python
house_price = pd.read_csv('E:/projects/housing_price/data/train.csv',index_col=0)['SalePrice']
```

Examine correlations among variables

In [4]:
```python
sns.heatmap(cleaned_data.corr())
plt.show()
```



It can be seen that there are highly correlated variables

In [5]:
```python
correlation = cleaned_data.corr()
```

In [6]:
```python
significantly_correlated = set()
for var1 in set(correlation):
    for var2 in set(correlation):
        if correlation.loc[var1][var2]>0.8 and var1!=var2:
            significantly_correlated.add(tuple(sorted([var1,var2])))
```

In [7]:
```python
print("The highly correlated couples are: {}".format(significantly_correlated))
```

```
The highly correlated couples are: {('OverallCond*OverallQual*GrLivArea', 'OverallQual*GrLivArea'),
('GrLivArea', 'OverallCond*GrLivArea'), ('GrLivArea', 'OverallQual*GrLivArea'), ('OverallQual', 'Over
allQual*GrLivArea'), ('OverallCond*GrLivArea', 'OverallCond*OverallQual*GrLivArea'), ('FireplaceQu',
'Fireplaces'), ('SaleCondition_Partial', 'SaleType_New'), ('GrLivArea', 'OverallCond*OverallQual*GrLi
vArea')}
```

As is said in readme.txt file, SaleCondition_Partial means when the house is sold, the house is only partially completed which also means that the house is a new house! So we can safely delete one of these categorical variables.

In [8]:
```python
cleaned_data = cleaned_data.drop(columns=['SaleCondition_Partial'])
```

Take a look at the other pair of high correlation

In [9]:
```python
correlation.loc['FireplaceQu']['Fireplaces']
```

Out[9]:
```
0.8617353520998596
```

We should remove one of the two, lets remove 'Fireplaces'

In [10]:
```python
cleaned_data = cleaned_data.drop(columns=['Fireplaces'])
```

Check again for highly correlated pairs

In [11]:
```python
significantly_correlated = set()
for var1 in set(cleaned_data.corr()):
    for var2 in set(cleaned_data.corr()):
        if correlation.loc[var1][var2]>0.8 and var1!=var2:
            significantly_correlated.add(tuple(sorted([var1,var2])))
if len(significantly_correlated)==0:
    print("There is no more highly correlated variables")
else:
    print("The highly correlated couples are: {}".format(significantly_correlated))
```

```
The highly correlated couples are: {('OverallCond*OverallQual*GrLivArea', 'OverallQual*GrLivArea'),
('GrLivArea', 'OverallCond*GrLivArea'), ('GrLivArea', 'OverallQual*GrLivArea'), ('OverallQual', 'Over
allQual*GrLivArea'), ('OverallCond*GrLivArea', 'OverallCond*OverallQual*GrLivArea'), ('GrLivArea', 'O
verallCond*OverallQual*GrLivArea')}
```

Now there is no more highly correlated pairs (other than interaction terms).

Splitting Train and Test Data

In [12]:
```python
train_data = cleaned_data.head(1460)
test_data = cleaned_data.tail(1459)
```

Outliers

It is a common sense that the lot area and the living area of the house are usually the most important factors for housing price, so we can plot the area against sale price to see if there is anything unusual.

Lot Area vs Sale Price

In [13]:
```python
lot_area = train_data['LotArea']

plt.scatter(lot_area,house_price)
plt.xlabel('lot area')
plt.ylabel('house price')
```

Out[13]:
```
Text(0, 0.5, 'house price')
```


There are four outliers with high leverage. We need to remove them.

In [14]:
```python
# First, to identify the ID's of the houses with the 4 largest lot area, we sort the data frame by lot area.
print(train_data.sort_values('LotArea',ascending=False)['LotArea'][:6])
```

```
Id
314     215245
336     164660
250     159000
707     115149
452      70761
1299     63887
Name: LotArea, dtype: int64
```

In [15]:
```python
# Now, drop them
train_data = train_data.drop(index=[314,336,250,707,452,1299])
```

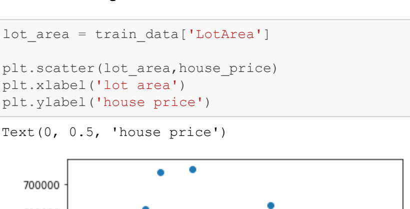
In [16]:
```python
house_price = house_price.drop(index=[314,336,250,707,452,1299])
```

Look at the lot again

In [17]:
```python
lot_area = train_data['LotArea']

plt.scatter(lot_area,house_price)
plt.xlabel('lot area')
plt.ylabel('house price')
```

Out[17]:
```
Text(0, 0.5, 'house price')
```

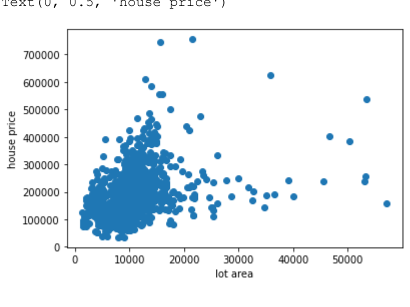


Living Area vs Sale Price

In [18]:
```python
living_area = train_data['GrLivArea']

plt.scatter(living_area,house_price)
plt.xlabel('living area')
plt.ylabel('house price')
```

Out[18]:
```
Text(0, 0.5, 'house price')
```



It is very clear that there are three outliers. Identify and remove them.

In [19]:
```python
# the one with large area but low sale price
print(living_area.sort_values(ascending=False)[:1])
```

```
Id
524    4676
Name: GrLivArea, dtype: int64
```

In [20]:
```python
# the two with very high sale price
print(house_price.sort_values(ascending=False)[:2])
```

```
Id
692     755000
1183    745000
Name: SalePrice, dtype: int64
```

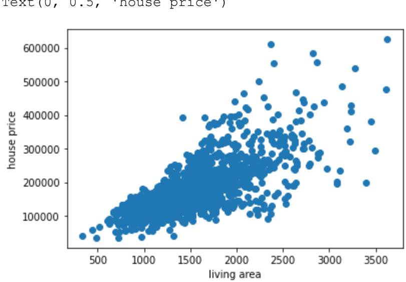
In [21]:
```python
train_data = train_data.drop(index=[524,692,1183])
house_price = house_price.drop(index=[524,692,1183])
```

Plot again

In [22]:
```python
living_area = train_data['GrLivArea']

plt.scatter(living_area,house_price)
plt.xlabel('living area')
plt.ylabel('house price')
```

Out[22]:
```
Text(0, 0.5, 'house price')
```

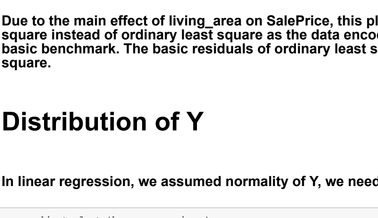


Due to the main effect of living_area on SalePrice, this plot here shows heterscedesticity. We should be using weighted least square instead of ordinary least square as the data encodes homoscedestic. But anyhow, lets still get the baseline model out as a basic benchmark. The basic residuals of ordinary least square will also help us determine the weights of weighted least square.

Distribution of Y

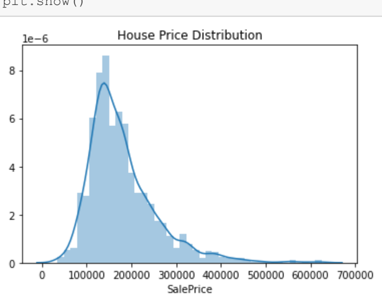
In linear regression, we assumed normality of Y, we need to verify this.

In [23]:
```python
sns.distplot(house_price)
plt.title('House Price Distribution')
plt.show()
```

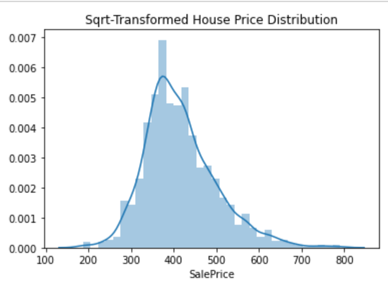


The distribution looks skewed, we consider either take log or take square-root.

In [24]:
```python
# Taking log
log_house_price = np.log(house_price)
sns.distplot(log_house_price)
plt.title('Log-Transformed House Price Distribution')
plt.show()
```



In [25]:
```python
sqrt_house_price = np.sqrt(house_price)
sns.distplot(sqrt_house_price)
plt.title('Sqrt-Transformed House Price Distribution')
plt.show()
```



Based on what we see, the log-transformed house price is closer to normal distribution. So we will use log-transformed data, however it seems like there are three exceptionally low sale prices which introduced a little bit of skewness into the distribution. We need to remove them before use.

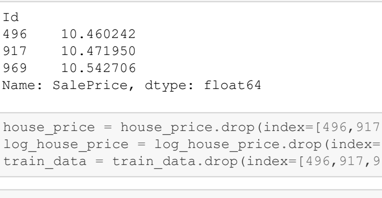
In [26]:
```python
print(log_house_price.sort_values()[:3])
```

```
Id
496    10.460242
917    10.471950
969    10.542706
Name: SalePrice, dtype: float64
```

In [27]:
```python
house_price = house_price.drop(index=[496,917,969])
log_house_price = log_house_price.drop(index=[496,917,969])
train_data = train_data.drop(index=[496,917,969])
```

In [28]:
```python
sns.distplot(log_house_price)
plt.title('log-transformed house price after removing outliers')
```

Out[28]:
```
Text(0.5, 1.0, 'log-transformed house price after removing outliers')
```



Linear Regression Fit

In [29]:
```python
data = sm.add_constant(train_data)
OLS = sm.OLS(log_house_price,train_data)
lin_regression = OLS.fit()
print(lin_regression.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              SalePrice   R-squared:                       0.936
Model:                            OLS   Adj. R-squared:                  0.929
Method:                 Least Squares   F-statistic:                     143.2
Date:                Thu, 16 Jul 2020   Prob (F-statistic):               0.00
Time:                        11:44:27   Log-Likelihood:                 1300.6
No. Observations:                1448   AIC:                            -2331.
Df Residuals:                    1313   BIC:                            -1619.
Df Model:                         134
Covariance Type:            nonrobust
=====================================================================================================
                                        coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------------
LotFrontage                        4.384e-05   9.92e-05      0.442      0.659      -0.000       0.000
LotArea                            4.574e-06   8.01e-07      5.713      0.000       3e-06    6.14e-06
OverallQual                           0.0804      0.026      3.071      0.002       0.029       0.132
OverallCond                           0.0632      0.027      2.345      0.019       0.010       0.116
MasVnrArea                         3.381e-05   2.54e-05      1.334      0.183   -1.59e-05    8.36e-05
BsmtQual                              0.0131      0.008      1.664      0.096      -0.002       0.028
BsmtCond                              0.0011      0.011      0.102      0.919      -0.020       0.022
BsmtExposure                          0.0119      0.004      3.249      0.001       0.005       0.019
HeatingQC                             0.0119      0.004      2.889      0.004       0.004       0.020
CentralAir                            0.0634      0.015      4.170      0.000       0.034       0.093
Electrical                           -0.0175      0.012     -1.515      0.130      -0.040       0.005
LowQualFinSF                      -5.239e-05    7.71e-05     -0.679      0.497      -0.000    9.89e-05
GrLivArea                             0.0004      0.000      3.494      0.000       0.000       0.001
BedroomAbvGr                         -0.0023      0.005     -0.423      0.672      -0.013       0.008
KitchenAbvGr                         -0.0523      0.023     -2.272      0.023      -0.098      -0.007
KitchenQual                           0.0215      0.007      3.096      0.002       0.008       0.035
FireplaceQu                           0.0071      0.002      3.386      0.001       0.003       0.011
GarageFinish                          0.0035      0.005      0.686      0.493      -0.007       0.014
GarageArea                            0.0001   2.28e-05      4.558      0.000    5.92e-05       0.000
PavedDrive                            2.6746      1.341      1.994      0.046       0.043       5.306
WoodDeckSF                         7.726e-05   2.58e-05      2.991      0.003    2.66e-05       0.000
OpenPorchSF                           0.0002   5.12e-05      2.975      0.003    5.18e-05       0.000
EnclosedPorch                      9.222e-05   5.34e-05      1.728      0.084   -1.25e-05       0.000
3SsnPorch                             0.0002   9.85e-05      1.760      0.079   -1.98e-05       0.000
ScreenPorch                           0.0002   5.29e-05      4.351      0.000       0.000       0.000
PoolArea                              0.0001   8.36e-05      1.481      0.139   -4.02e-05       0.000
MiscVal                            1.492e-06   6.29e-06      0.237      0.813   -1.08e-05    1.38e-05
MoSold                               -0.0007      0.001     -0.635      0.526      -0.003       0.001
YrSold                                0.0006      0.002      0.247      0.805      -0.004       0.005
LandContour_Lvl                       0.0045      0.012      0.372      0.710      -0.019       0.028
has_fence                            -0.0020      0.008     -0.249      0.804      -0.018       0.014
gabble_roof                          -0.0046      0.008     -0.602      0.547      -0.020       0.010
LandSlope_Gtl                         0.0116      0.016      0.707      0.480      -0.021       0.044
Bath                                  0.0282      0.006      4.343      0.000       0.015       0.041
GarageScore                           0.0079      0.003      2.279      0.023       0.001       0.015
OverallCond*OverallQual*GrLivArea  3.888e-06   2.73e-06      1.424      0.155   -1.47e-06    9.24e-06
OverallQual*GrLivArea             -2.588e-05   1.59e-05     -1.626      0.104   -5.71e-05    5.34e-06
OverallCond*GrLivArea              -1.63e-05   1.87e-05     -0.872      0.384    -5.3e-05    2.04e-05
OverallCond*OverallQual              -0.0052      0.005     -1.135      0.257      -0.014       0.004
ExterScore                           -0.0036      0.006     -0.560      0.575      -0.016       0.009
BsmtLowQArea                          0.0001   1.45e-05      6.962      0.000    7.25e-05       0.000
BsmtAvgQArea                          0.0001   1.68e-05      8.087      0.000       0.000       0.000
BsmtHiQArea                           0.0002   1.68e-05     10.676      0.000       0.000       0.000
RecentRemod                           0.0033      0.009      0.346      0.729      -0.015       0.022
HouseAge                             -0.0020      0.000     -6.702      0.000      -0.003      -0.001
MSZoning_Others                       0.4217      0.224      1.880      0.060      -0.018       0.862
MSZoning_RL                           0.4632      0.223      2.074      0.038       0.025       0.901
MSZoning_RM                           0.4524      0.224      2.022      0.043       0.014       0.891
LotShape_IR1                          0.3294      0.168      1.960      0.050      -0.000       0.659
LotShape_IR2                          0.3436      0.169      2.034      0.042       0.012       0.675
LotShape_IR3                          0.3362      0.169      1.984      0.047       0.004       0.669
LotShape_Reg                          0.3282      0.168      1.950      0.051      -0.002       0.658
LotConfig_Corner                      0.4514      0.224      2.019      0.044       0.013       0.890
LotConfig_Inside                      0.4379      0.224      1.959      0.050      -0.001       0.876
LotConfig_Others                      0.4480      0.224      2.002      0.045       0.009       0.887
Neighborhood_Blmngtn                  0.0792      0.041      1.929      0.054      -0.001       0.160
Neighborhood_Blueste                  0.0093      0.079      0.117      0.907      -0.146       0.165
Neighborhood_BrDale                  -0.0007      0.043     -0.015      0.988      -0.085       0.083
Neighborhood_BrkSide                  0.0674      0.033      2.066      0.039       0.003       0.131
Neighborhood_ClearCr                  0.0735      0.036      2.052      0.040       0.003       0.144
Neighborhood_CollgCr                  0.0346      0.030      1.159      0.247      -0.024       0.093
Neighborhood_Crawfor                  0.1706      0.031      5.473      0.000       0.109       0.232
Neighborhood_Edwards                  0.0002      0.030      0.006      0.995      -0.059       0.059
Neighborhood_Gilbert                  0.0409      0.031      1.302      0.193      -0.021       0.103
Neighborhood_IDOTRR                  -0.0402      0.036     -1.128      0.260      -0.110       0.030
Neighborhood_MeadowV                 -0.1121      0.043     -2.638      0.008      -0.196      -0.029
Neighborhood_Mitchel                  0.0198      0.032      0.628      0.530      -0.042       0.082
Neighborhood_NAmes                    0.0465      0.029      1.613      0.107      -0.010       0.103
Neighborhood_NPkVill                  0.0932      0.048      1.949      0.052      -0.001       0.187
Neighborhood_NWAmes                   0.0236      0.031      0.770      0.441      -0.037       0.084
Neighborhood_NoRidge                  0.0713      0.035      2.061      0.039       0.003       0.139
Neighborhood_NridgHt                  0.1330      0.033      4.065      0.000       0.069       0.197
Neighborhood_OldTown                  0.0053      0.033      0.162      0.872      -0.059       0.069
Neighborhood_SWISU                    0.0662      0.037      1.777      0.076      -0.007       0.139
Neighborhood_Sawyer                   0.0523      0.030      1.721      0.085      -0.007       0.112
Neighborhood_SawyerW                  0.0439      0.032      1.381      0.168      -0.018       0.106
Neighborhood_Somerst                  0.1257      0.034      3.672      0.000       0.059       0.193
Neighborhood_StoneBr                  0.1797      0.036      4.922      0.000       0.108       0.251
Neighborhood_Timber                   0.0551      0.033      1.650      0.099      -0.010       0.121
Neighborhood_Veenker                  0.0990      0.041      2.394      0.017       0.018       0.180
Condition1_Artery                     0.1105      0.077      1.430      0.153      -0.041       0.262
Condition1_Feedr                      0.1397      0.076      1.842      0.066      -0.009       0.289
Condition1_Norm                       0.1944      0.075      2.585      0.010       0.047       0.342
Condition1_PosA                       0.1354      0.083      1.627      0.104      -0.028       0.299
Condition1_PosN                       0.1953      0.080      2.428      0.015       0.037       0.353
Condition1_RRAe                       0.0760      0.081      0.938      0.348      -0.083       0.235
Condition1_RRAn                       0.1648      0.078      2.122      0.034       0.012       0.317
Condition1_RRNe                       0.1438      0.102      1.411      0.159      -0.056       0.344
Condition1_RRNn                       0.1773      0.087      2.045      0.041       0.007       0.347
Condition2_Artery                     0.2091      0.117      1.795      0.073      -0.019       0.438
Condition2_Feedr                      0.1932      0.100      1.928      0.054      -0.003       0.390
Condition2_Norm                       0.2151      0.090      2.400      0.017       0.039       0.391
Condition2_PosA                       0.2860      0.143      2.000      0.046       0.005       0.567
Condition2_PosN                       0.0397      0.131      0.302      0.762      -0.218       0.297
Condition2_RRAe                       0.0576      0.138      0.418      0.676      -0.213       0.328
Condition2_RRAn                       0.1394      0.129      1.083      0.279      -0.113       0.392
Condition2_RRNn                       0.1971      0.112      1.758      0.079      -0.023       0.417
BldgType_1Fam                         0.3000      0.135      2.215      0.027       0.034       0.566
BldgType_2fmCon                       0.3056      0.135      2.264      0.024       0.041       0.570
BldgType_Duplex                       0.2820      0.136      2.079      0.038       0.016       0.548
BldgType_Twnhs                        0.2071      0.135      1.529      0.127      -0.059       0.473
BldgType_TwnhsE                       0.2426      0.135      1.797      0.073      -0.022       0.507
HouseStyle_1.5Fin                     0.1735      0.085      2.049      0.041       0.007       0.340
HouseStyle_1.5Unf                     0.1892      0.088      2.154      0.031       0.017       0.361
HouseStyle_1Story                     0.1767      0.085      2.080      0.038       0.010       0.343
HouseStyle_2.5Fin                     0.0964      0.096      0.999      0.318      -0.093       0.286
HouseStyle_2.5Unf                     0.1821      0.090      2.031      0.042       0.006       0.358
HouseStyle_2Story                     0.1699      0.085      2.007      0.045       0.004       0.336
HouseStyle_SFoyer                     0.1716      0.086      1.985      0.047       0.002       0.341
HouseStyle_SLvl                       0.1780      0.086      2.075      0.038       0.010       0.346
MasVnrType_BrkFace                    0.4398      0.223      1.971      0.049       0.002       0.878
MasVnrType_None                       0.4467      0.223      2.001      0.046       0.009       0.885
MasVnrType_Others                     0.4508      0.225      2.007      0.045       0.010       0.891
Foundation_BrkTil                     0.2177      0.112      1.944      0.052      -0.002       0.437
Foundation_CBlock                     0.2412      0.112      2.146      0.032       0.021       0.462
Foundation_PConc                      0.2539      0.113      2.257      0.024       0.033       0.475
Foundation_Slab                       0.2414      0.114      2.109      0.035       0.017       0.466
Foundation_Stone                      0.2520      0.120      2.105      0.035       0.017       0.487
Foundation_Wood                       0.1311      0.125      1.050      0.294      -0.114       0.376
Functional_Maj1                       0.2382      0.099      2.416      0.016       0.045       0.432
Functional_Maj2                       0.0784      0.107      0.730      0.465      -0.132       0.289
Functional_Min1                       0.3017      0.099      3.061      0.002       0.108       0.495
Functional_Min2                       0.2905      0.100      2.911      0.004       0.095       0.486
Functional_Mod                        0.1986      0.102      1.940      0.053      -0.002       0.399
Functional_Sev                       -0.1027      0.133     -0.773      0.439      -0.363       0.158
Functional_Typ                        0.3326      0.097      3.414      0.001       0.141       0.524
GarageType_Attchd                     0.4487      0.223      2.009      0.045       0.011       0.887
GarageType_Detchd                     0.4488      0.224      2.008      0.045       0.010       0.887
GarageType_Others                     0.4398      0.224      1.963      0.050       0.000       0.879
SaleType_COD                          0.0737      0.079      0.935      0.350      -0.081       0.229
SaleType_CWD                          0.1839      0.089      2.074      0.038       0.010       0.358
SaleType_Con                          0.2234      0.105      2.137      0.033       0.018       0.429
SaleType_ConLD                        0.1676      0.083      2.025      0.043       0.005       0.330
SaleType_ConLI                        0.0772      0.089      0.868      0.385      -0.097       0.252
SaleType_ConLw                        0.0807      0.088      0.921      0.357      -0.091       0.253
SaleType_New                          0.2776      0.092      3.002      0.003       0.096       0.459
SaleType_Oth                          0.1585      0.093      1.704      0.089      -0.024       0.341
SaleType_WD                           0.0947      0.076      1.241      0.215      -0.055       0.244
SaleCondition_Abnorml                 0.0666      0.064      1.040      0.298      -0.059       0.192
SaleCondition_AdjLand                 0.1458      0.085      1.714      0.087      -0.021       0.313
SaleCondition_Alloca                  0.0975      0.071      1.363      0.173      -0.043       0.238
SaleCondition_Family                  0.0671      0.068      0.994      0.321      -0.065       0.200
SaleCondition_Normal                  0.1321      0.063      2.096      0.036       0.008       0.256
Exterior_HdBoard                      0.2082      0.112      1.859      0.063      -0.012       0.428
Exterior_MetalSd                      0.2353      0.112      2.105      0.036       0.016       0.455
Exterior_Others                       0.2439      0.112      2.182      0.029       0.025       0.463
Exterior_Plywood                      0.2157      0.113      1.908      0.057      -0.006       0.437
Exterior_VinylSd                      0.2291      0.112      2.042      0.041       0.009       0.449
Exterior_Wd Sdng                      0.2051      0.112      1.837      0.066      -0.014       0.424
==============================================================================
Omnibus:                      394.020   Durbin-Watson:                   1.886
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2932.107
Skew:                          -1.060   Prob(JB):                         0.00
Kurtosis:                       9.641   Cond. No.                     1.25e+16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 3.61e-20. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

```
D:\anaconda\lib\site-packages\numpy\core\fromnumeric.py:2542: FutureWarning: Method .ptp is deprecate
d and will be removed in a future version. Use numpy.ptp instead.
  return ptp(axis=axis, out=out, **kwargs)
```

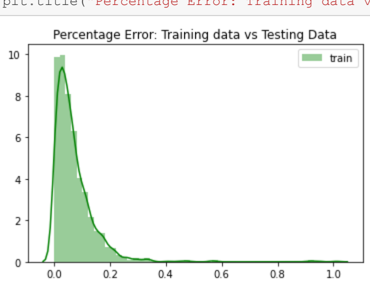
If we were doing a train-validation split on the training data, we can use the second half of the code below to compare the results. But since we have known that the regression model (with our carefully chose features) is very robust, we go ahead without validating.

In [40]:
```python
train_fitted_log_price = lin_regression.predict(train_data)
train_fitted_price = np.exp(train_fitted_log_price)
train_percentage_error = np.abs(house_price-train_fitted_price)/house_price
print("Average percentage error for training data: {}".format(sum(train_percentage_error)/len(train_per
centage_error)))

#test_fitted_log_price = lin_regression.predict(test_data)
#test_fitted_price = np.exp(test_fitted_log_price)
#test_percentage_error = np.abs(np.exp(test_house_price)-test_fitted_price)/np.exp(test_house_price)
#print("Average percentage error for testing data: {}".format(sum(test_percentage_error)/len(test_perce
ntage_error)))
```

```
Average percentage error for training data: 0.07152024776178298
```

In [31]:
```python
sns.distplot(train_percentage_error,color='g',label='train')
#sns.distplot(test_percentage_error,color='r',label='test')
plt.legend()
plt.title("Percentage Error: Training data vs Testing Data");
```



Prediction of OLS

In [32]:
```python
OLS_prediction = pd.DataFrame(np.exp(lin_regression.predict(test_data)))
OLS_prediction = OLS_prediction.rename(columns={0:'SalePrice'})
OLS_prediction.to_csv('E:/projects/housing_price/OLS_prediction.csv',columns=OLS_prediction.columns)
```

Removing Insignificant Predictors

Based on the summary, there are a few predictors with small t-scores together with large p-values. It means that these predictors are weak predictors and should be taken out.

In [33]:
```python
weak_predictors = ['LotFrontage','CentralAir','FireplaceQu','GarageFinish','RecentRemod','YrSold','MoSo
ld','BsmtQual','BsmtCond',
                   'gabble_roof','LandContour_Lvl','Electrical','MasVnrArea','BedroomAbvGr','MiscVal',
'PoolArea','has_fence',
                   'LandSlope_Gtl','ExterScore','SaleCondition_Abnorml','SaleCondition_AdjLand','SaleCo
ndition_Alloca',
                   'SaleCondition_Family','SaleCondition_Normal','3SsnPorch','LowQualFinSF']
```

In [34]:
```python
weak_removed_train_data = train_data.drop(columns=weak_predictors)
```

In [35]:
```python
weak_removed_test_data = test_data.drop(columns=weak_predictors)
```

In [36]:
```python
weak_removed_train_data = sm.add_constant(weak_removed_train_data)
weak_removed_OLS = sm.OLS(log_house_price,weak_removed_train_data)
weak_removed_lin_regression = weak_removed_OLS.fit()
print(weak_removed_lin_regression.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              SalePrice   R-squared:                       0.932
Model:                            OLS   Adj. R-squared:                  0.926
Method:                 Least Squares   F-statistic:                     168.9
Date:                Thu, 16 Jul 2020   Prob (F-statistic):               0.00
Time:                        11:44:27   Log-Likelihood:                 1253.0
No. Observations:                1448   AIC:                            -2288.
Df Residuals:                    1339   BIC:                            -1713.
Df Model:                         108
Covariance Type:            nonrobust
=====================================================================================================
                                        coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------------
LotArea                            4.909e-06   7.98e-07      6.148      0.000    3.34e-06    6.48e-06
OverallQual                           0.0824      0.026      3.183      0.001       0.032       0.133
OverallCond                           0.0621      0.027      2.341      0.019       0.010       0.114
BsmtExposure                          0.0107      0.004      2.955      0.003       0.004       0.018
HeatingQC                             0.0151      0.004      3.726      0.000       0.007       0.023
GrLivArea                             0.0004      0.000      3.297      0.001       0.000       0.001
KitchenAbvGr                         -0.0622      0.023     -2.723      0.007      -0.107      -0.017
KitchenQual                           0.0189      0.007      2.806      0.005       0.006       0.032
GarageArea                            0.0001    2.3e-05      4.646      0.000    6.16e-05       0.000
PavedDrive                            3.0384      0.048     63.860      0.000       2.945       3.132
WoodDeckSF                         8.696e-05   2.59e-05      3.362      0.001    3.62e-05       0.000
OpenPorchSF                           0.0001   5.14e-05      2.845      0.005    4.54e-05       0.000
EnclosedPorch                      8.648e-05    5.4e-05      1.603      0.109   -1.94e-05       0.000
ScreenPorch                           0.0003   5.33e-05      4.749      0.000       0.000       0.000
Bath                                  0.0286      0.007      4.399      0.000       0.016       0.041
GarageScore                           0.0105      0.003      3.208      0.001       0.004       0.017
OverallCond*OverallQual*GrLivArea  2.722e-06   2.71e-06      1.003      0.316    -2.6e-06    8.04e-06
OverallQual*GrLivArea             -1.939e-05   1.57e-05     -1.235      0.217   -5.02e-05    1.14e-05
OverallCond*GrLivArea             -8.965e-06   1.86e-05     -0.483      0.629   -4.54e-05    2.74e-05
OverallCond*OverallQual              -0.0048      0.005     -1.045      0.296      -0.014       0.004
BsmtLowQArea                          0.0001   1.46e-05      6.960      0.000    7.32e-05       0.000
BsmtAvgQArea                          0.0001   1.69e-05      7.980      0.000       0.000       0.000
BsmtHiQArea                           0.0002   1.67e-05     10.837      0.000       0.000       0.000
HouseAge                             -0.0023      0.000     -7.745      0.000      -0.003      -0.002
MSZoning_Others                       0.4720      0.015     31.085      0.000       0.442       0.502
MSZoning_RL                           0.5280      0.012     44.243      0.000       0.505       0.551
MSZoning_RM                           0.5192      0.014     37.702      0.000       0.492       0.546
LotShape_IR1                          0.3786      0.013     30.238      0.000       0.354       0.403
LotShape_IR2                          0.3890      0.018     21.791      0.000       0.354       0.424
LotShape_IR3                          0.3744      0.031     12.205      0.000       0.314       0.435
LotShape_Reg                          0.3772      0.013     29.480      0.000       0.352       0.402
LotConfig_Corner                      0.5096      0.010     52.493      0.000       0.491       0.529
LotConfig_Inside                      0.4991      0.009     53.851      0.000       0.481       0.517
LotConfig_Others                      0.5105      0.011     47.712      0.000       0.489       0.531
Neighborhood_Blmngtn                  0.0922      0.031      3.022      0.003       0.032       0.152
Neighborhood_Blueste                  0.0511      0.075      0.680      0.496      -0.096       0.199
Neighborhood_BrDale                  -0.0063      0.033     -0.190      0.849      -0.072       0.059
Neighborhood_BrkSide                  0.0763      0.020      3.824      0.000       0.037       0.115
Neighborhood_ClearCr                  0.0651      0.024      2.770      0.006       0.019       0.111
Neighborhood_CollgCr                  0.0379      0.013      2.976      0.003       0.013       0.063
Neighborhood_Crawfor                  0.1807      0.018     10.257      0.000       0.146       0.215
Neighborhood_Edwards                  0.0077      0.013      0.573      0.567      -0.019       0.034
Neighborhood_Gilbert                  0.0561      0.017      3.375      0.001       0.023       0.089
Neighborhood_IDOTRR                  -0.0409      0.025     -1.646      0.100      -0.090       0.008
Neighborhood_MeadowV                 -0.0994      0.032     -3.095      0.002      -0.162      -0.036
Neighborhood_Mitchel                  0.0244      0.017      1.448      0.148      -0.009       0.058
Neighborhood_NAmes                    0.0523      0.011      4.764      0.000       0.031       0.074
Neighborhood_NPkVill                  0.1038      0.039      2.666      0.008       0.027       0.180
Neighborhood_NWAmes                   0.0287      0.015      1.852      0.064      -0.002       0.059
Neighborhood_NoRidge                  0.0731      0.021      3.474      0.001       0.032       0.114
Neighborhood_NridgHt                  0.1545      0.018      8.728      0.000       0.120       0.189
Neighborhood_OldTown                  0.0038      0.020      0.191      0.849      -0.036       0.043
Neighborhood_SWISU                    0.0625      0.026      2.433      0.015       0.012       0.113
Neighborhood_Sawyer                   0.0541      0.015      3.597      0.000       0.025       0.084
Neighborhood_SawyerW                  0.0447      0.016      2.763      0.006       0.013       0.076
Neighborhood_Somerst                  0.1398      0.021      6.685      0.000       0.099       0.181
Neighborhood_StoneBr                  0.1858      0.024      7.653      0.000       0.138       0.233
Neighborhood_Timber                   0.0669      0.020      3.362      0.001       0.028       0.106
Neighborhood_Veenker                  0.1042      0.034      3.089      0.002       0.038       0.170
Condition1_Artery                     0.1297      0.021      6.301      0.000       0.089       0.170
Condition1_Feedr                      0.1599      0.017      9.372      0.000       0.126       0.193
Condition1_Norm                       0.2111      0.013     15.846      0.000       0.185       0.237
Condition1_PosA                       0.1593      0.037      4.314      0.000       0.087       0.232
Condition1_PosN                       0.2177      0.027      8.057      0.000       0.165       0.271
Condition1_RRAe                       0.0825      0.033      2.502      0.012       0.018       0.147
Condition1_RRAn                       0.1844      0.024      7.616      0.000       0.137       0.232
Condition1_RRNe                       0.1700      0.069      2.446      0.015       0.034       0.306
Condition1_RRNn                       0.2046      0.046      4.430      0.000       0.114       0.295
Condition2_Artery                     0.2224      0.080      2.777      0.006       0.065       0.380
Condition2_Feedr                      0.1870      0.054      3.494      0.000       0.082       0.292
Condition2_Norm                       0.2372      0.032      7.339      0.000       0.174       0.301
Condition2_PosA                       0.3365      0.116      2.906      0.004       0.109       0.564
Condition2_PosN                       0.0569      0.103      0.554      0.580      -0.145       0.259
Condition2_RRAe                       0.1118      0.102      1.099      0.272      -0.088       0.311
Condition2_RRAn                       0.1686      0.101      1.672      0.095      -0.029       0.366
Condition2_RRNn                       0.1987      0.075      2.659      0.008       0.052       0.345
BldgType_1Fam                         0.3402      0.012     29.456      0.000       0.318       0.363
BldgType_2fmCon                       0.3392      0.021     16.405      0.000       0.299       0.380
BldgType_Duplex                       0.3083      0.021     14.549      0.000       0.267       0.350
BldgType_Twnhs                        0.2427      0.020     12.041      0.000       0.203       0.282
BldgType_TwnhsE                       0.2887      0.015     19.550      0.000       0.260       0.318
HouseStyle_1.5Fin                     0.2019      0.012     16.677      0.000       0.178       0.226
HouseStyle_1.5Unf                     0.2267      0.028      7.978      0.000       0.171       0.282
HouseStyle_1Story                     0.2031      0.012     17.406      0.000       0.180       0.226
HouseStyle_2.5Fin                     0.0864      0.040      2.146      0.032       0.007       0.165
HouseStyle_2.5Unf                     0.1983      0.033      5.994      0.000       0.133       0.263
HouseStyle_2Story                     0.1926      0.011     17.466      0.000       0.171       0.214
HouseStyle_SFoyer                     0.2056      0.021      9.833      0.000       0.165       0.247
HouseStyle_SLvl                       0.2047      0.016     12.687      0.000       0.173       0.236
MasVnrType_BrkFace                    0.5042      0.009     53.264      0.000       0.486       0.523
MasVnrType_None                       0.5037      0.009     53.886      0.000       0.485       0.522
MasVnrType_Others                     0.5113      0.011     45.475      0.000       0.489       0.533
Foundation_BrkTil                     0.2504      0.017     14.516      0.000       0.217       0.284
Foundation_CBlock                     0.2793      0.015     18.172      0.000       0.249       0.309
Foundation_PConc                      0.2895      0.016     18.024      0.000       0.258       0.321
Foundation_Slab                       0.2796      0.026     10.619      0.000       0.228       0.331
Foundation_Stone                      0.2601      0.040      6.475      0.000       0.181       0.339
Foundation_Wood                       0.1604      0.054      2.977      0.003       0.055       0.266
Functional_Maj1                       0.2746      0.032      8.695      0.000       0.213       0.337
Functional_Maj2                       0.1128      0.046      2.465      0.014       0.023       0.203
Functional_Min1                       0.3363      0.025     13.578      0.000       0.288       0.385
Functional_Min2                       0.3292      0.025     13.265      0.000       0.281       0.378
Functional_Mod                        0.2346      0.031      7.666      0.000       0.175       0.295
Functional_Sev                       -0.1384      0.094     -1.467      0.143      -0.323       0.047
Functional_Typ                        0.3701      0.019     19.415      0.000       0.333       0.408
GarageType_Attchd                     0.5121      0.010     50.636      0.000       0.492       0.532
GarageType_Detchd                     0.5069      0.011     47.646      0.000       0.486       0.528
GarageType_Others                     0.5002      0.011     47.111      0.000       0.479       0.521
SaleType_COD                          0.0942      0.022      4.313      0.000       0.051       0.137
SaleType_CWD                          0.1901      0.050      3.770      0.000       0.091       0.289
SaleType_Con                          0.2734      0.070      3.889      0.000       0.135       0.411
SaleType_ConLD                        0.1919      0.037      5.179      0.000       0.119       0.265
SaleType_ConLI                        0.1217      0.046      2.653      0.008       0.032       0.212
SaleType_ConLw                        0.1341      0.046      2.893      0.004       0.043       0.225
SaleType_New                          0.2035      0.019     10.883      0.000       0.167       0.240
SaleType_Oth                          0.1651      0.058      2.854      0.004       0.052       0.279
SaleType_WD                           0.1452      0.016      9.139      0.000       0.114       0.176
Exterior_HdBoard                      0.2390      0.009     27.489      0.000       0.222       0.256
Exterior_MetalSd                      0.2639      0.009     30.847      0.000       0.247       0.281
Exterior_Others                       0.2769      0.009     30.622      0.000       0.259       0.295
Exterior_Plywood                      0.2490      0.012     21.530      0.000       0.226       0.272
Exterior_VinylSd                      0.2562      0.008     30.822      0.000       0.240       0.273
Exterior_Wd Sdng                      0.2343      0.009     25.997      0.000       0.217       0.252
==============================================================================
Omnibus:                      439.132   Durbin-Watson:                   1.904
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3045.522
Skew:                          -1.228   Prob(JB):                         0.00
Kurtosis:                       9.667   Cond. No.                     1.90e+16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 1.56e-20. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

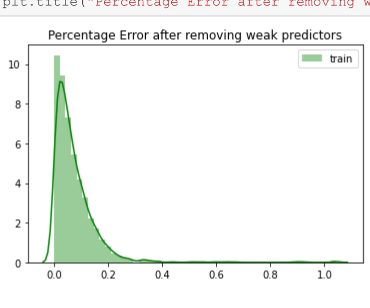
```
D:\anaconda\lib\site-packages\numpy\core\fromnumeric.py:2542: FutureWarning: Method .ptp is deprecate
d and will be removed in a future version. Use numpy.ptp instead.
  return ptp(axis=axis, out=out, **kwargs)
```

In [37]:
```python
weak_removed_train_fitted_log_price = weak_removed_lin_regression.predict(weak_removed_train_data)
weak_removed_train_fitted_price = np.exp(weak_removed_train_fitted_log_price)
weak_removed_train_percentage_error = np.abs(house_price-weak_removed_train_fitted_price)/house_price
print("Average percentage error for training data: {}".format(sum(weak_removed_train_percentage_error)/
len(weak_removed_train_percentage_error)))

#weak_removed_test_fitted_log_price = weak_removed_lin_regression.predict(weak_removed_test_data)
#weak_removed_test_fitted_price = np.exp(weak_removed_test_fitted_log_price)
#weak_removed_test_percentage_error = np.abs(np.exp(weak_removed_test_house_price)-weak_removed_test_fi
tted_price)/np.exp(weak_removed_test_house_price)
#print("Average percentage error for testing data: {}".format(sum(weak_removed_test_percentage_error)/l
en(weak_removed_test_percentage_error)))
```

```
Average percentage error for training data: 0.07372782175076624
```

In [38]:
```python
sns.distplot(weak_removed_train_percentage_error,color='g',label='train')
#sns.distplot(weak_removed_test_percentage_error,color='r',label='test')
plt.legend()
plt.title("Percentage Error after removing weak predictors");
```



Weighted Linear Regression Fit

Get the weights from OLS model --- there are a few common and possible ways to determine weights, all of which aims at decoupling variance from the dependence on mean

In [41]:
```python
# weights = 1/(house_price-train_fitted_price)**2
weights = house_price/ np.abs(house_price-train_fitted_price)
# weights = 1/np.log(train_fitted_price)
# weights = 1/np.log(train_fitted_price)**2
```

In [43]:
```python
WLS = sm.WLS(log_house_price,train_data,weights=weights)
```

In [44]:
```python
weighted_lin_regression = WLS.fit()
```

In [45]:
```python
WLS_train_fitted_log_price = weighted_lin_regression.predict(train_data)
WLS_train_fitted_price = np.exp(WLS_train_fitted_log_price)
WLS_train_percentage_error = np.abs(house_price-WLS_train_fitted_price)/house_price
print("Average percentage error for training data: {}".format(sum(WLS_train_percentage_error)/len(train
_percentage_error)))
```

```
Average percentage error for training data: 0.07016364517176095
```

In [46]:
```python
print(weighted_lin_regression.summary())
```

```
                            WLS Regression Results
==============================================================================
Dep. Variable:              SalePrice   R-squared:                       1.000
Model:                            WLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 7.092e+10
Date:                Thu, 16 Jul 2020   Prob (F-statistic):               0.00
Time:                        11:46:03   Log-Likelihood:                 2225.2
No. Observations:                1448   AIC:                            -4180.
Df Residuals:                    1313   BIC:                            -3468.
Df Model:                         134
Covariance Type:            nonrobust
======================================================================================================
                                         coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------------
LotFrontage                         4.717e-05   4.47e-05      1.056      0.291   -4.05e-05       0.000
LotArea                             4.693e-06   3.91e-07     11.994      0.000    3.93e-06    5.46e-06
OverallQual                            0.0624      0.015      4.146      0.000       0.033       0.092
OverallCond                            0.0406      0.015      2.670      0.008       0.011       0.070
MasVnrArea                          3.796e-05   9.95e-06      3.814      0.000    1.84e-05    5.75e-05
BsmtQual                               0.0126      0.003      3.677      0.000       0.006       0.019
BsmtCond                               0.0022      0.005      0.450      0.653      -0.007       0.012
BsmtExposure                           0.0112      0.001      7.834      0.000       0.008       0.014
HeatingQC                              0.0103      0.002      5.493      0.000       0.007       0.014
CentralAir                             0.0588      0.008      7.472      0.000       0.043       0.074
Electrical                            -0.0148      0.005     -2.691      0.007      -0.026      -0.004
LowQualFinSF                       -4.107e-05   4.82e-05     -0.851      0.395      -0.000    5.36e-05
GrLivArea                              0.0003   6.19e-05      4.472      0.000       0.000       0.000
BedroomAbvGr                          -0.0023      0.002     -0.998      0.318      -0.007       0.002
KitchenAbvGr                          -0.0494      0.015     -3.250      0.001      -0.079      -0.020
KitchenQual                            0.0213      0.003      7.364      0.000       0.016       0.027
FireplaceQu                            0.0075      0.001      9.124      0.000       0.006       0.009
GarageFinish                           0.0041      0.002      1.910      0.056      -0.000       0.008
GarageArea                             0.0001    9.4e-06     11.530      0.000       9e-05       0.000
PavedDrive                             2.2546      0.537      4.199      0.000       1.201       3.308
WoodDeckSF                          6.268e-05   1.13e-05      5.538      0.000    4.05e-05    8.49e-05
OpenPorchSF                            0.0001   2.12e-05      6.879      0.000       0.000       0.000
EnclosedPorch                          0.0001   2.58e-05      4.076      0.000    5.46e-05       0.000
3SsnPorch                              0.0002   3.88e-05      4.356      0.000    9.29e-05       0.000
ScreenPorch                            0.0002   2.42e-05      8.916      0.000       0.000       0.000
PoolArea                               0.0002   6.67e-05      2.545      0.011    3.89e-05       0.000
MiscVal                             3.734e-06   3.42e-06      1.092      0.275   -2.97e-06    1.04e-05
MoSold                                -0.0007      0.000     -1.616      0.106      -0.002       0.000
YrSold                                 0.0014      0.001      1.520      0.129      -0.000       0.003
LandContour_Lvl                        0.0041      0.006      0.724      0.469      -0.007       0.015
has_fence                             -0.0014      0.003     -0.434      0.665      -0.008       0.005
gabble_roof                           -0.0065      0.003     -2.087      0.037      -0.013      -0.000
LandSlope_Gtl                          0.0078      0.008      1.007      0.314      -0.007       0.023
Bath                                   0.0278      0.003     10.173      0.000       0.022       0.033
GarageScore                            0.0082      0.001      5.607      0.000       0.005       0.011
OverallCond*OverallQual*GrLivArea   1.243e-06   1.78e-06      0.700      0.484   -2.24e-06    4.72e-06
OverallQual*GrLivArea              -1.09e-05   9.68e-06     -1.126      0.260   -2.99e-05    8.09e-06
OverallCond*GrLivArea               3.836e-06   1.11e-05      0.344      0.731    -1.8e-05    2.57e-05
OverallCond*OverallQual               -0.0024      0.003     -0.908      0.364      -0.008       0.003
ExterScore                            -0.0040      0.003     -1.540      0.124      -0.009       0.001
BsmtLowQArea                           0.0001   6.51e-06     15.369      0.000    8.73e-05       0.000
BsmtAvgQArea                           0.0001   7.25e-06     18.696      0.000       0.000       0.000
BsmtHiQArea                            0.0002   6.96e-06     26.219      0.000       0.000       0.000
RecentRemod                            0.0042      0.004      1.185      0.236      -0.003       0.011
HouseAge                              -0.0021      0.000    -14.771      0.000      -0.002      -0.002
MSZoning_Others                        0.3518      0.090      3.910      0.000       0.175       0.528
MSZoning_RL                            0.3973      0.090      4.438      0.000       0.222       0.573
MSZoning_RM                            0.3782      0.089      4.234      0.000       0.203       0.553
LotShape_IR1                           0.2816      0.067      4.178      0.000       0.149       0.414
LotShape_IR2                           0.2855      0.068      4.229      0.000       0.153       0.418
LotShape_IR3                           0.2816      0.068      4.140      0.000       0.148       0.415
LotShape_Reg                           0.2786      0.067      4.136      0.000       0.146       0.411
LotConfig_Corner                       0.3818      0.089      4.275      0.000       0.207       0.557
LotConfig_Inside                       0.3688      0.090      4.116      0.000       0.193       0.545
LotConfig_Others                       0.3766      0.090      4.202      0.000       0.201       0.552
Neighborhood_Blmngtn                   0.0712      0.018      4.038      0.000       0.037       0.106
Neighborhood_Blueste                  -0.0025      0.035     -0.072      0.943      -0.071       0.066
Neighborhood_BrDale                   -0.0087      0.018     -0.487      0.626      -0.044       0.026
Neighborhood_BrkSide                   0.0616      0.015      4.172      0.000       0.033       0.091
Neighborhood_ClearCr                   0.0623      0.016      3.832      0.000       0.030       0.094
Neighborhood_CollgCr                   0.0252      0.012      2.085      0.037       0.001       0.049
Neighborhood_Crawfor                   0.1567      0.015     10.803      0.000       0.128       0.185
Neighborhood_Edwards                  -0.0071      0.013     -0.547      0.585      -0.033       0.018
Neighborhood_Gilbert                   0.0266      0.013      2.094      0.036       0.002       0.052
Neighborhood_IDOTRR                   -0.0232      0.016     -1.431      0.153      -0.055       0.009
Neighborhood_MeadowV                  -0.1155      0.017     -6.620      0.000      -0.150      -0.081
Neighborhood_Mitchel                   0.0096      0.012      0.781      0.435      -0.015       0.034
Neighborhood_NAmes                     0.0357      0.012      3.087      0.002       0.013       0.058
Neighborhood_NPkVill                   0.0826      0.021      3.951      0.000       0.042       0.124
Neighborhood_NWAmes                    0.0126      0.013      0.963      0.336      -0.013       0.038
Neighborhood_NoRidge                   0.0607      0.013      4.559      0.000       0.035       0.087
Neighborhood_NridgHt                   0.1202      0.012      9.778      0.000       0.096       0.144
Neighborhood_OldTown                   0.0092      0.014      0.651      0.515      -0.018       0.037
Neighborhood_SWISU                     0.0582      0.018      3.182      0.001       0.022       0.094
Neighborhood_Sawyer                    0.0488      0.012      3.919      0.000       0.024       0.073
Neighborhood_SawyerW                   0.0338      0.012      2.768      0.006       0.010       0.058
Neighborhood_Somerst                   0.1190      0.013      8.948      0.000       0.093       0.145
Neighborhood_StoneBr                   0.1638      0.014     11.826      0.000       0.137       0.191
Neighborhood_Timber                    0.0423      0.013      3.165      0.002       0.016       0.069
Neighborhood_Veenker                   0.0841      0.015      5.557      0.000       0.054       0.114
Condition1_Artery                      0.0845      0.031      2.685      0.007       0.023       0.146
Condition1_Feedr                       0.1157      0.031      3.746      0.000       0.055       0.176
Condition1_Norm                        0.1672      0.031      5.482      0.000       0.107       0.227
Condition1_PosA                        0.1237      0.035      3.489      0.001       0.054       0.193
Condition1_PosN                        0.1718      0.033      5.270      0.000       0.108       0.236
Condition1_RRAe                        0.0543      0.034      1.592      0.112      -0.013       0.121
Condition1_RRAn                        0.1466      0.031      4.682      0.000       0.085       0.208
Condition1_RRNe                        0.1239      0.055      2.249      0.025       0.016       0.232
Condition1_RRNn                        0.1395      0.038      3.638      0.000       0.064       0.215
Condition2_Artery                      0.1665      0.051      3.271      0.001       0.067       0.266
Condition2_Feedr                       0.1638      0.039      4.160      0.000       0.087       0.241
Condition2_Norm                        0.1916      0.035      5.415      0.000       0.122       0.261
Condition2_PosA                        0.3065      0.056      5.439      0.000       0.196       0.417
Condition2_PosN                        0.0086      0.038      0.226      0.821      -0.066       0.083
Condition2_RRAe                        0.0096      0.046      0.210      0.834      -0.080       0.100
Condition2_RRAn                        0.1096      0.036      3.005      0.003       0.038       0.181
Condition2_RRNn                        0.1711      0.061      2.814      0.005       0.052       0.290
BldgType_1Fam                          0.2577      0.054      4.768      0.000       0.152       0.364
BldgType_2fmCon                        0.2630      0.054      4.847      0.000       0.157       0.369
BldgType_Duplex                        0.2297      0.055      4.159      0.000       0.121       0.338
BldgType_Twnhs                         0.1720      0.055      3.116      0.002       0.064       0.280
BldgType_TwnhsE                        0.2048      0.054      3.796      0.000       0.099       0.311
HouseStyle_1.5Fin                      0.1459      0.034      4.319      0.000       0.080       0.212
HouseStyle_1.5Unf                      0.1703      0.035      4.875      0.000       0.102       0.239
HouseStyle_1Story                      0.1536      0.034      4.500      0.000       0.087       0.221
HouseStyle_2.5Fin                      0.0504      0.044      1.156      0.248      -0.035       0.136
HouseStyle_2.5Unf                      0.1533      0.038      4.008      0.000       0.078       0.228
HouseStyle_2Story                      0.1469      0.034      4.296      0.000       0.080       0.214
HouseStyle_SFoyer                      0.1506      0.034      4.414      0.000       0.084       0.218
HouseStyle_SLvl                        0.1562      0.034      4.574      0.000       0.089       0.223
MasVnrType_BrkFace                     0.3713      0.089      4.153      0.000       0.196       0.547
MasVnrType_None                        0.3771      0.090      4.209      0.000       0.201       0.553
MasVnrType_Others                      0.3789      0.090      4.231      0.000       0.203       0.555
Foundation_BrkTil                      0.1931      0.045      4.262      0.000       0.104       0.282
Foundation_CBlock                      0.2149      0.045      4.732      0.000       0.126       0.304
Foundation_PConc                       0.2264      0.046      4.944      0.000       0.137       0.316
Foundation_Slab                        0.2130      0.046      4.605      0.000       0.122       0.304
Foundation_Stone                       0.1947      0.054      3.612      0.000       0.089       0.300
Foundation_Wood                        0.0852      0.058      1.457      0.145      -0.029       0.200
Functional_Maj1                        0.1959      0.041      4.792      0.000       0.116       0.276
Functional_Maj2                        0.0569      0.049      1.162      0.246      -0.039       0.153
Functional_Min1                        0.2785      0.040      6.970      0.000       0.200       0.357
Functional_Min2                        0.2620      0.040      6.533      0.000       0.183       0.341
Functional_Mod                         0.1716      0.040      4.276      0.000       0.093       0.250
Functional_Sev                        -0.1396      0.041     -3.416      0.001      -0.220      -0.059
Functional_Typ                         0.3019      0.038      7.847      0.000       0.226       0.377
GarageType_Attchd                      0.3771      0.090      4.210      0.000       0.201       0.553
GarageType_Detchd                      0.3779      0.089      4.240      0.000       0.203       0.553
GarageType_Others                      0.3722      0.090      4.142      0.000       0.196       0.548
SaleType_COD                           0.0598      0.034      1.778      0.076      -0.006       0.126
SaleType_CWD                           0.1718      0.058      2.951      0.003       0.058       0.286
SaleType_Con                           0.2078      0.046      4.491      0.000       0.117       0.299
SaleType_ConLD                         0.1472      0.036      4.105      0.000       0.077       0.218
SaleType_ConLI                         0.0495      0.042      1.190      0.234      -0.032       0.131
SaleType_ConLw                         0.0591      0.035      1.681      0.093      -0.010       0.128
SaleType_New                           0.1794      0.067      2.690      0.007       0.049       0.310
SaleType_Oth                           0.1772      0.054      3.264      0.001       0.071       0.284
SaleType_WD                            0.0755      0.032      2.328      0.020       0.012       0.139
SaleCondition_Abnorml                 -0.0077      0.068     -0.114      0.909      -0.141       0.126
SaleCondition_AdjLand                  0.0564      0.073      0.775      0.439      -0.086       0.199
SaleCondition_Alloca                   0.0638      0.073      0.873      0.383      -0.080       0.207
SaleCondition_Family                   0.0009      0.070      0.013      0.990      -0.136       0.137
SaleCondition_Normal                   0.0559      0.068      0.824      0.410      -0.077       0.189
Exterior_HdBoard                       0.1724      0.045      3.862      0.000       0.085       0.260
Exterior_MetalSd                       0.1997      0.045      4.465      0.000       0.112       0.287
Exterior_Others                        0.2063      0.045      4.577      0.000       0.118       0.295
Exterior_Plywood                       0.1789      0.045      3.946      0.000       0.090       0.268
Exterior_VinylSd                       0.1950      0.045      4.338      0.000       0.107       0.283
Exterior_Wd Sdng                       0.1750      0.045      3.921      0.000       0.087       0.263
==============================================================================
Omnibus:                      249.570   Durbin-Watson:                   1.895
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               55.479
Skew:                          -0.052   Prob(JB):                     8.97e-13
Kurtosis:                       2.047   Cond. No.                     1.16e+16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.16e+16. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In [47]:
```python
WLS_prediction = pd.DataFrame(np.exp(weighted_lin_regression.predict(test_data)))
```

In [48]:
```python
WLS_prediction = WLS_prediction.rename(columns={0:'SalePrice'})
```

In [51]:
```python
WLS_prediction.to_csv('E:/projects/housing_price/WLS_prdiction.csv',columns=WLS_prediction.columns)
```

In [ ]: