

Final_SexyKaggSwagg_Housing

October 2, 2016

```
In [437]: %matplotlib inline
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

import statsmodels.api as sm
from scipy.stats import skew
from sklearn.cross_validation import train_test_split
from sklearn.cross_validation import KFold
```

```
In [438]: #Reading data
data = pd.read_csv('./data/train.csv')
```

1 Data Exploration

1.1 Basic Statistics & Observations

```
In [439]: #Read first 5 lines
data.head()
```

```
Out[439]:
```

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

	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscFeature	Misc
0	Lvl	AllPub	...	0	NaN	NaN		NaN
1	Lvl	AllPub	...	0	NaN	NaN		NaN
2	Lvl	AllPub	...	0	NaN	NaN		NaN
3	Lvl	AllPub	...	0	NaN	NaN		NaN
4	Lvl	AllPub	...	0	NaN	NaN		NaN

	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	2	2008	WD	Normal	208500

1	5	2007	WD	Normal	181500
2	9	2008	WD	Normal	223500
3	2	2006	WD	Abnorml	140000
4	12	2008	WD	Normal	250000

[5 rows x 81 columns]

```
In [440]: #general shape of the data
data.shape
```

```
Out[440]: (1460, 81)
```

```
In [441]: #Basic Statistics about all the *quantitative* columns
data.describe()
```

```
/Users/tiffanychien/anaconda/lib/python3.5/site-packages/numpy/lib/function_base.py:
RuntimeWarning)
```

```
Out[441]:
```

	Id	MSSubClass	LotFrontage	LotArea	OverallQual
count	1460.000000	1460.000000	1201.000000	1460.000000	1460.000000
mean	730.500000	56.897260	70.049958	10516.828082	6.099315
std	421.610009	42.300571	24.284752	9981.264932	1.382997
min	1.000000	20.000000	21.000000	1300.000000	1.000000
25%	365.750000	20.000000	NaN	7553.500000	5.000000
50%	730.500000	50.000000	NaN	9478.500000	6.000000
75%	1095.250000	70.000000	NaN	11601.500000	7.000000
max	1460.000000	190.000000	313.000000	215245.000000	10.000000

	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1
count	1460.000000	1460.000000	1460.000000	1452.000000	1460.000000
mean	5.575342	1971.267808	1984.865753	103.685262	443.639726
std	1.112799	30.202904	20.645407	181.066207	456.098091
min	1.000000	1872.000000	1950.000000	0.000000	0.000000
25%	5.000000	1954.000000	1967.000000	NaN	0.000000
50%	5.000000	1973.000000	1994.000000	NaN	383.500000
75%	6.000000	2000.000000	2004.000000	NaN	712.250000
max	9.000000	2010.000000	2010.000000	1600.000000	5644.000000

	...	WoodDeckSF	OpenPorchSF	EnclosedPorch	3SsnPorch
count	...	1460.000000	1460.000000	1460.000000	1460.000000
mean	...	94.244521	46.660274	21.954110	3.40958
std	...	125.338794	66.256028	61.119149	29.31733
min	...	0.000000	0.000000	0.000000	0.000000
25%	...	0.000000	0.000000	0.000000	0.000000
50%	...	0.000000	25.000000	0.000000	0.000000
75%	...	168.000000	68.000000	0.000000	0.000000
max	...	857.000000	547.000000	552.000000	508.000000

	ScreenPorch	PoolArea	MiscVal	MoSold	YrSold
count	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000
mean	15.060959	2.758904	43.489041	6.321918	2007.815753
std	55.757415	40.177307	496.123024	2.703626	1.328095
min	0.000000	0.000000	0.000000	1.000000	2006.000000
25%	0.000000	0.000000	0.000000	5.000000	2007.000000
50%	0.000000	0.000000	0.000000	6.000000	2008.000000
75%	0.000000	0.000000	0.000000	8.000000	2009.000000
max	480.000000	738.000000	15500.000000	12.000000	2010.000000

	SalePrice
count	1460.000000
mean	180921.195890
std	79442.502883
min	34900.000000
25%	129975.000000
50%	163000.000000
75%	214000.000000
max	755000.000000

[8 rows x 38 columns]

In [442]: *#Basic Statistics about all the *categorical* columns*
data.describe(include = ['O'])

Out [442]:

	MSZoning	Street	Alley	LotShape	LandContour	Utilities	LotConfig	\
count	1460	1460	91	1460	1460	1460	1460	
unique	5	2	2	4	4	2	5	
top	RL	Pave	Grvl	Reg	Lvl	AllPub	Inside	
freq	1151	1454	50	925	1311	1459	1052	

	LandSlope	Neighborhood	Condition1	...	GarageType	\
count	1460	1460	1460	...	1379	
unique	3	25	9	...	6	
top	Gtl	NAmes	Norm	...	Attchd	
freq	1382	225	1260	...	870	

	GarageFinish	GarageQual	GarageCond	PavedDrive	PoolQC	Fence	\
count	1379	1379	1379	1460	7	281	
unique	3	5	5	3	3	4	
top	Unf	TA	TA	Y	Gd	MnPrv	
freq	605	1311	1326	1340	3	157	

	MiscFeature	SaleType	SaleCondition
count	54	1460	1460
unique	4	9	6
top	Shed	WD	Normal
freq	49	1267	1198

[4 rows x 43 columns]

```
In [443]: #Correlations
data.corr()
```

```
Out [443]:
```

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

	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1
Id	0.012609	-0.012713	-0.021998	-0.050298	-0.005024

MSSubClass	-0.059316	0.027850	0.040581	0.022936	-0.0698
LotFrontage	-0.059213	0.123349	0.088866	0.193458	0.2336
LotArea	-0.005636	0.014228	0.013788	0.104160	0.2141
OverallQual	-0.091932	0.572323	0.550684	0.411876	0.2396
OverallCond	1.000000	-0.375983	0.073741	-0.128101	-0.0462
YearBuilt	-0.375983	1.000000	0.592855	0.315707	0.2495
YearRemodAdd	0.073741	0.592855	1.000000	0.179618	0.1284
MasVnrArea	-0.128101	0.315707	0.179618	1.000000	0.2647
BsmtFinSF1	-0.046231	0.249503	0.128451	0.264736	1.0000
BsmtFinSF2	0.040229	-0.049107	-0.067759	-0.072319	-0.0501
BsmtUnfSF	-0.136841	0.149040	0.181133	0.114442	-0.4952
TotalBsmtSF	-0.171098	0.391452	0.291066	0.363936	0.5223
1stFlrSF	-0.144203	0.281986	0.240379	0.344501	0.4458
2ndFlrSF	0.028942	0.010308	0.140024	0.174561	-0.1370
LowQualFinSF	0.025494	-0.183784	-0.062419	-0.069071	-0.0645
GrLivArea	-0.079686	0.199010	0.287389	0.390857	0.2081
BsmtFullBath	-0.054942	0.187599	0.119470	0.085310	0.6492
BsmtHalfBath	0.117821	-0.038162	-0.012337	0.026673	0.0674
FullBath	-0.194149	0.468271	0.439046	0.276833	0.0585
HalfBath	-0.060769	0.242656	0.183331	0.201444	0.0042
BedroomAbvGr	0.012980	-0.070651	-0.040581	0.102821	-0.1073
KitchenAbvGr	-0.087001	-0.174800	-0.149598	-0.037610	-0.0810
TotRmsAbvGrd	-0.057583	0.095589	0.191740	0.280682	0.0443
Fireplaces	-0.023820	0.147716	0.112581	0.249070	0.2600
GarageYrBlt	-0.324297	0.825667	0.642277	0.252691	0.1534
GarageCars	-0.185758	0.537850	0.420622	0.364204	0.2240
GarageArea	-0.151521	0.478954	0.371600	0.373066	0.2969
WoodDeckSF	-0.003334	0.224880	0.205726	0.159718	0.2043
OpenPorchSF	-0.032589	0.188686	0.226298	0.125703	0.1117
EnclosedPorch	0.070356	-0.387268	-0.193919	-0.110204	-0.1023
3SsnPorch	0.025504	0.031355	0.045286	0.018796	0.0264
ScreenPorch	0.054811	-0.050364	-0.038740	0.061466	0.0620
PoolArea	-0.001985	0.004950	0.005829	0.011723	0.1404
MiscVal	0.068777	-0.034383	-0.010286	-0.029815	0.0035
MoSold	-0.003511	0.012398	0.021490	-0.005965	-0.0157
YrSold	0.043950	-0.013618	0.035743	-0.008201	0.0143
SalePrice	-0.077856	0.522897	0.507101	0.477493	0.3864
...	...	WoodDeckSF	OpenPorchSF	EnclosedPorch	3SsnPor
Id	...	-0.029643	-0.000477	0.002889	-0.0466
MSSubClass	...	-0.012579	-0.006100	-0.012037	-0.0438
LotFrontage	...	0.088521	0.151972	0.010700	0.0700
LotArea	...	0.171698	0.084774	-0.018340	0.0204
OverallQual	...	0.238923	0.308819	-0.113937	0.0303
OverallCond	...	-0.003334	-0.032589	0.070356	0.0255
YearBuilt	...	0.224880	0.188686	-0.387268	0.0313
YearRemodAdd	...	0.205726	0.226298	-0.193919	0.0452
MasVnrArea	...	0.159718	0.125703	-0.110204	0.0187

BsmtFinSF1	...	0.204306	0.111761	-0.102303	0.0264
BsmtFinSF2	...	0.067898	0.003093	0.036543	-0.0299
BsmtUnfSF	...	-0.005316	0.129005	-0.002538	0.0207
TotalBsmtSF	...	0.232019	0.247264	-0.095478	0.0373
1stFlrSF	...	0.235459	0.211671	-0.065292	0.0561
2ndFlrSF	...	0.092165	0.208026	0.061989	-0.0243
LowQualFinSF	...	-0.025444	0.018251	0.061081	-0.0042
GrLivArea	...	0.247433	0.330224	0.009113	0.0206
BsmtFullBath	...	0.175315	0.067341	-0.049911	-0.0001
BsmtHalfBath	...	0.040161	-0.025324	-0.008555	0.0351
FullBath	...	0.187703	0.259977	-0.115093	0.0353
HalfBath	...	0.108080	0.199740	-0.095317	-0.0049
BedroomAbvGr	...	0.046854	0.093810	0.041570	-0.0244
KitchenAbvGr	...	-0.090130	-0.070091	0.037312	-0.0246
TotRmsAbvGrd	...	0.165984	0.234192	0.004151	-0.0066
Fireplaces	...	0.200019	0.169405	-0.024822	0.0112
GarageYrBlt	...	0.224577	0.228425	-0.297003	0.0235
GarageCars	...	0.226342	0.213569	-0.151434	0.0357
GarageArea	...	0.224666	0.241435	-0.121777	0.0350
WoodDeckSF	...	1.000000	0.058661	-0.125989	-0.0327
OpenPorchSF	...	0.058661	1.000000	-0.093079	-0.0058
EnclosedPorch	...	-0.125989	-0.093079	1.000000	-0.0373
3SsnPorch	...	-0.032771	-0.005842	-0.037305	1.0000
ScreenPorch	...	-0.074181	0.074304	-0.082864	-0.0314
PoolArea	...	0.073378	0.060762	0.054203	-0.0079
MiscVal	...	-0.009551	-0.018584	0.018361	0.0003
MoSold	...	0.021011	0.071255	-0.028887	0.0294
YrSold	...	0.022270	-0.057619	-0.009916	0.0186
SalePrice	...	0.324413	0.315856	-0.128578	0.0445

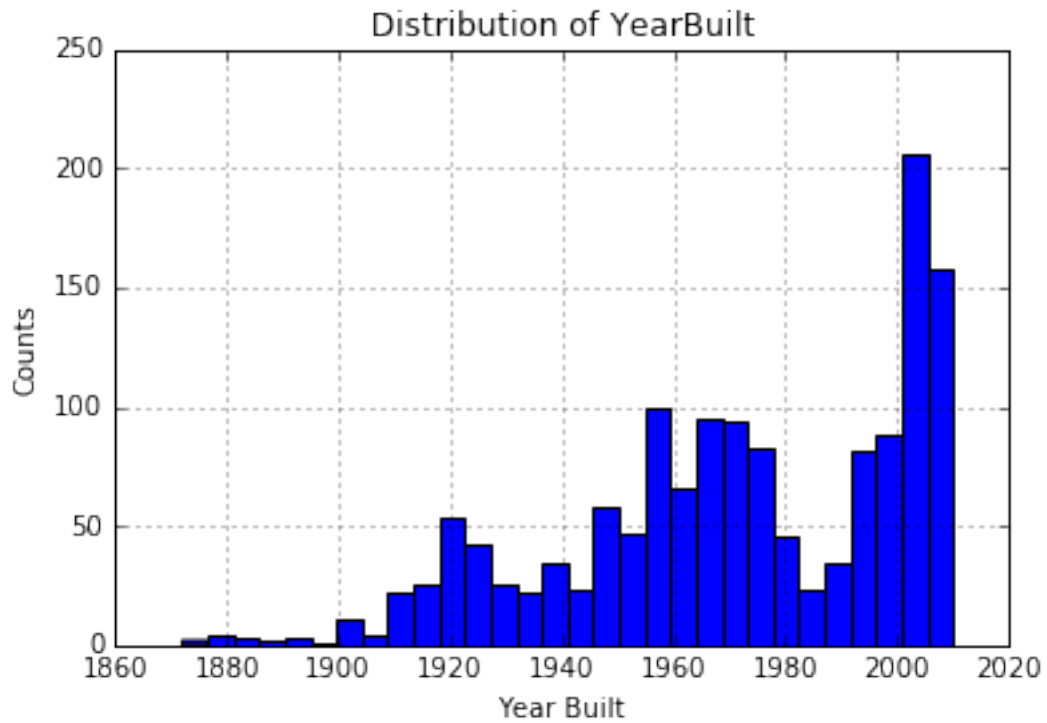
	ScreenPorch	PoolArea	MiscVal	MoSold	YrSold	SaleP
Id	0.001330	0.057044	-0.006242	0.021172	0.000712	-0.02
MSSubClass	-0.026030	0.008283	-0.007683	-0.013585	-0.021407	-0.08
LotFrontage	0.041383	0.206167	0.003368	0.011200	0.007450	0.35
LotArea	0.043160	0.077672	0.038068	0.001205	-0.014261	0.26
OverallQual	0.064886	0.065166	-0.031406	0.070815	-0.027347	0.79
OverallCond	0.054811	-0.001985	0.068777	-0.003511	0.043950	-0.07
YearBuilt	-0.050364	0.004950	-0.034383	0.012398	-0.013618	0.52
YearRemodAdd	-0.038740	0.005829	-0.010286	0.021490	0.035743	0.50
MasVnrArea	0.061466	0.011723	-0.029815	-0.005965	-0.008201	0.47
BsmtFinSF1	0.062021	0.140491	0.003571	-0.015727	0.014359	0.38
BsmtFinSF2	0.088871	0.041709	0.004940	-0.015211	0.031706	-0.01
BsmtUnfSF	-0.012579	-0.035092	-0.023837	0.034888	-0.041258	0.21
TotalBsmtSF	0.084489	0.126053	-0.018479	0.013196	-0.014969	0.61
1stFlrSF	0.088758	0.131525	-0.021096	0.031372	-0.013604	0.60
2ndFlrSF	0.040606	0.081487	0.016197	0.035164	-0.028700	0.31
LowQualFinSF	0.026799	0.062157	-0.003793	-0.022174	-0.028921	-0.02
GrLivArea	0.101510	0.170205	-0.002416	0.050240	-0.036526	0.70

BsmtFullBath	0.023148	0.067616	-0.023047	-0.025361	0.067049	0.22
BsmtHalfBath	0.032121	0.020025	-0.007367	0.032873	-0.046524	-0.01
FullBath	-0.008106	0.049604	-0.014290	0.055872	-0.019669	0.56
HalfBath	0.072426	0.022381	0.001290	-0.009050	-0.010269	0.28
BedroomAbvGr	0.044300	0.070703	0.007767	0.046544	-0.036014	0.16
KitchenAbvGr	-0.051613	-0.014525	0.062341	0.026589	0.031687	-0.13
TotRmsAbvGrd	0.059383	0.083757	0.024763	0.036907	-0.034516	0.53
Fireplaces	0.184530	0.095074	0.001409	0.046357	-0.024096	0.46
GarageYrBlt	-0.075418	-0.014501	-0.032417	0.005337	-0.001014	0.48
GarageCars	0.050494	0.020934	-0.043080	0.040522	-0.039117	0.64
GarageArea	0.051412	0.061047	-0.027400	0.027974	-0.027378	0.62
WoodDeckSF	-0.074181	0.073378	-0.009551	0.021011	0.022270	0.32
OpenPorchSF	0.074304	0.060762	-0.018584	0.071255	-0.057619	0.31
EnclosedPorch	-0.082864	0.054203	0.018361	-0.028887	-0.009916	-0.12
3SsnPorch	-0.031436	-0.007992	0.000354	0.029474	0.018645	0.04
ScreenPorch	1.000000	0.051307	0.031946	0.023217	0.010694	0.11
PoolArea	0.051307	1.000000	0.029669	-0.033737	-0.059689	0.09
MiscVal	0.031946	0.029669	1.000000	-0.006495	0.004906	-0.02
MoSold	0.023217	-0.033737	-0.006495	1.000000	-0.145721	0.04
YrSold	0.010694	-0.059689	0.004906	-0.145721	1.000000	-0.02
SalePrice	0.111447	0.092404	-0.021190	0.046432	-0.028923	1.00

[38 rows x 38 columns]

```
In [466]: # some graphs of different variables
data['YearBuilt'].hist(bins=30)
plt.title('Distribution of YearBuilt')
plt.xlabel('Year Built')
plt.ylabel('Counts')
```

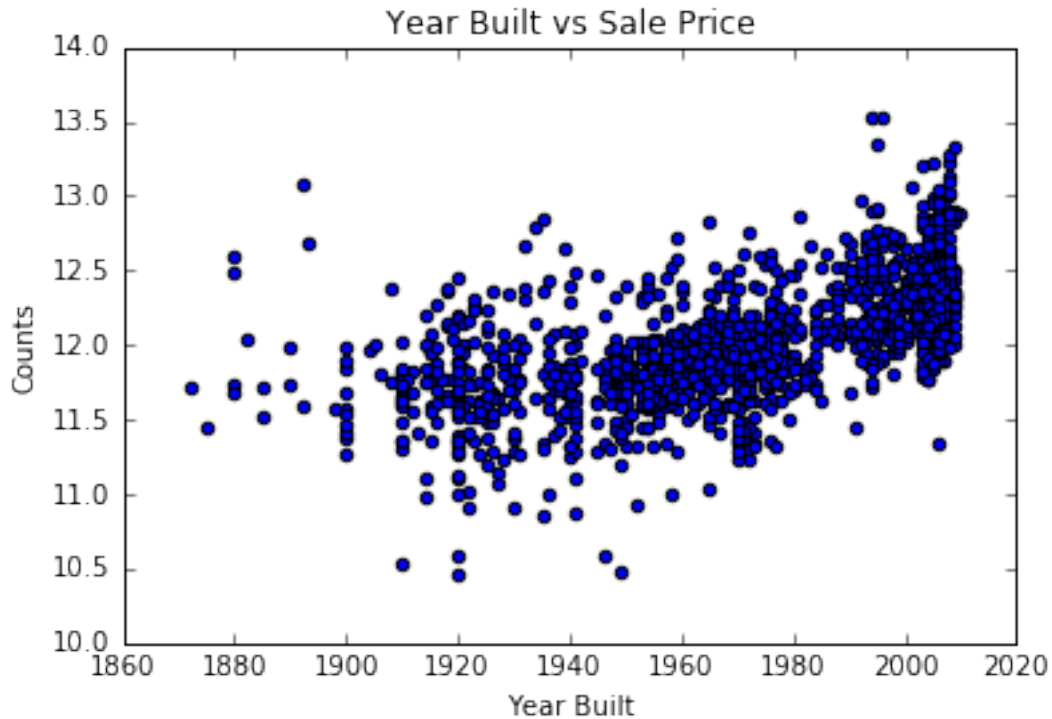
```
Out[466]: <matplotlib.text.Text at 0x11ba49710>
```



There are more new houses in the dataset than old houses, other than a dip around 1990.

```
In [468]: plt.scatter(data['YearBuilt'], data['SalePrice'])  
          plt.title('Year Built vs Sale Price')  
          plt.xlabel('Year Built')  
          plt.ylabel('Sale Price')
```

```
Out[468]: <matplotlib.text.Text at 0x11eb84908>
```

There seems to be a somewhat weak positive correlation between the year the house was built and its price, with quite a few outliers.

2 Data Cleaning & Feature Engineering

```
In [444]: #log-transforming response variables
          data['SalePrice'] = np.log1p(data['SalePrice'])
```

```
In [445]: #Splitting categorical variables into dummies
          data = pd.get_dummies(data, drop_first=True)
```

```
          data.shape
```

```
Out[445]: (1460, 247)
```

```
In [446]: #Filling NAs with column means
          data = data.fillna(data.mean())
```

```
          #And checking if any NAs exist after the code running
          data.isnull().values.any()
```

```
Out[446]: False
```

3 Model Building

```
In [447]: #Splitting datasets into train & test
         local_train, local_test = train_test_split(data, test_size = 0.2, random
```

```
In [448]: #Removing SalesPrices (is the predicting y-variable)

         local_train_y = local_train['SalePrice']
         local_train_x = local_train.drop('SalePrice', axis = 1)

         local_test_y = local_test['SalePrice']
         local_test_x = local_test.drop('SalePrice', axis = 1)
```

```
In [449]: #General statistics to make sure everything is right
         local_train.shape[0] + local_test.shape[0] == data.shape[0]
```

```
Out[449]: True
```

```
In [450]: #General statistics to make sure everything is right
         local_train.shape[1] == local_test.shape[1] == data.shape[1]
```

```
Out[450]: True
```

```
In [451]: #Fitting model
         clf = sm.OLS(local_train_y,local_train_x)
         result = clf.fit()
         result.summary()
```

```
Out[451]: <class 'statsmodels.iolib.summary.Summary'>
         """
```

```

                                OLS Regression Results
=====
Dep. Variable:                  SalePrice      R-squared:
Model:                            OLS         Adj. R-squared:
Method:                 Least Squares         F-statistic:           6.20
Date:                Sun, 02 Oct 2016         Prob (F-statistic):
Time:                  23:10:46         Log-Likelihood:
No. Observations:                1168         AIC:
Df Residuals:                    929         BIC:
Df Model:                        239
Covariance Type:                nonrobust
=====
                                coef      std err          t      P>|t|      [9
-----
Id                        -6.255e-06   8.34e-06   -0.750    0.454    -2.
MSSubClass                -0.0003      0.000   -0.662    0.508
LotFrontage                0.0005      0.000    2.046    0.041    2.
LotArea                   2.351e-06   6.09e-07    3.858    0.000    1.
OverallQual                0.0373      0.006    6.760    0.000
```

OverallCond	0.0368	0.005	7.909	0.000	
YearBuilt	0.0017	0.000	4.227	0.000	
YearRemodAdd	0.0007	0.000	2.502	0.013	
MasVnrArea	1.891e-06	3.21e-05	0.059	0.953	-6.
BsmtFinSF1	5.557e-05	1.5e-05	3.699	0.000	2.
BsmtFinSF2	4.926e-05	2.96e-05	1.662	0.097	-8.
BsmtUnfSF	-9.627e-06	1.45e-05	-0.665	0.506	-3.
TotalBsmtSF	9.52e-05	1.92e-05	4.958	0.000	5.
1stFlrSF	7.979e-05	3.37e-05	2.370	0.018	1.
2ndFlrSF	6.321e-05	2.93e-05	2.158	0.031	5.
LowQualFinSF	3.398e-06	7.5e-05	0.045	0.964	
GrLivArea	0.0001	3.09e-05	4.735	0.000	8.
BsmtFullBath	0.0263	0.010	2.532	0.012	
BsmtHalfBath	0.0029	0.017	0.173	0.862	
FullBath	0.0113	0.012	0.973	0.331	
HalfBath	0.0150	0.011	1.340	0.181	
BedroomAbvGr	0.0053	0.007	0.722	0.471	
KitchenAbvGr	-0.0577	0.037	-1.553	0.121	
TotRmsAbvGrd	0.0088	0.005	1.717	0.086	
Fireplaces	0.0176	0.013	1.383	0.167	
GarageYrBlt	-0.0004	0.000	-1.081	0.280	
GarageCars	0.0231	0.012	1.870	0.062	
GarageArea	0.0001	4.29e-05	2.656	0.008	2.
WoodDeckSF	0.0001	3.06e-05	3.509	0.000	4.
OpenPorchSF	6.613e-05	6.39e-05	1.035	0.301	-5.
EnclosedPorch	2.88e-05	6.85e-05	0.420	0.674	
3SsnPorch	0.0001	0.000	1.285	0.199	-7.
ScreenPorch	0.0003	6.77e-05	3.787	0.000	
PoolArea	0.0001	0.000	0.690	0.490	
MiscVal	3.64e-06	7.13e-06	0.511	0.610	-1.
MoSold	0.0008	0.001	0.623	0.533	
YrSold	0.0015	0.001	3.011	0.003	
MSZoning_FV	0.4914	0.061	8.069	0.000	
MSZoning_RH	0.4877	0.060	8.155	0.000	
MSZoning_RL	0.4702	0.051	9.191	0.000	
MSZoning_RM	0.4413	0.048	9.227	0.000	
Street_Pave	0.1054	0.062	1.695	0.090	
Alley_Pave	0.0472	0.025	1.910	0.056	
LotShape_IR2	0.0302	0.023	1.305	0.192	
LotShape_IR3	0.0128	0.050	0.255	0.798	
LotShape_Reg	0.0036	0.009	0.415	0.678	
LandContour_HLS	0.0439	0.028	1.585	0.113	
LandContour_Low	-0.0321	0.034	-0.947	0.344	
LandContour_Lvl	0.0237	0.020	1.189	0.235	
Utilities_NoSeWa	-0.2010	0.128	-1.571	0.117	
LotConfig_CulDSac	0.0350	0.018	1.923	0.055	
LotConfig_FR2	-0.0297	0.021	-1.385	0.166	
LotConfig_FR3	-0.0636	0.091	-0.702	0.483	

LotConfig_Inside	-0.0097	0.010	-1.005	0.315
LandSlope_Mod	0.0643	0.023	2.844	0.005
LandSlope_Sev	-0.1578	0.066	-2.400	0.017
Neighborhood_Blueste	-0.1314	0.117	-1.121	0.263
Neighborhood_BrDale	-0.0952	0.058	-1.648	0.100
Neighborhood_BrkSide	-0.0300	0.051	-0.585	0.558
Neighborhood_ClearCr	-0.0094	0.050	-0.186	0.852
Neighborhood_CollgCr	-0.0509	0.039	-1.300	0.194
Neighborhood_Crawfor	0.0653	0.046	1.417	0.157
Neighborhood_Edwards	-0.1064	0.043	-2.458	0.014
Neighborhood_Gilbert	-0.0558	0.041	-1.347	0.178
Neighborhood_IDOTRR	-0.0489	0.057	-0.853	0.394
Neighborhood_MeadowV	-0.2280	0.059	-3.895	0.000
Neighborhood_Mitchel	-0.0991	0.044	-2.227	0.026
Neighborhood_NAmes	-0.0668	0.042	-1.574	0.116
Neighborhood_NPkVill	-0.0332	0.076	-0.435	0.664
Neighborhood_NWAmes	-0.0718	0.043	-1.651	0.099
Neighborhood_NoRidge	-0.0088	0.046	-0.193	0.847
Neighborhood_NridgHt	0.0747	0.041	1.826	0.068
Neighborhood_OldTown	-0.0702	0.051	-1.362	0.173
Neighborhood_SWISU	-0.0284	0.051	-0.561	0.575
Neighborhood_Sawyer	-0.0726	0.044	-1.646	0.100
Neighborhood_SawyerW	-0.0481	0.042	-1.148	0.251
Neighborhood_Somerst	-0.0057	0.048	-0.119	0.905
Neighborhood_StoneBr	0.1035	0.044	2.345	0.019
Neighborhood_Timber	-0.0283	0.043	-0.656	0.512
Neighborhood_Veenker	0.0177	0.054	0.330	0.742
Condition1_Feedr	0.0687	0.028	2.457	0.014
Condition1_Norm	0.1108	0.023	4.752	0.000
Condition1_PosA	0.0708	0.052	1.350	0.177
Condition1_PosN	0.1054	0.041	2.579	0.010
Condition1_RRAe	-0.0006	0.047	-0.012	0.991
Condition1_RRAn	0.0839	0.036	2.306	0.021
Condition1_RRNe	0.0476	0.084	0.565	0.572
Condition1_RRNn	0.1036	0.064	1.612	0.107
Condition2_Feedr	0.0887	0.116	0.763	0.446
Condition2_Norm	0.0464	0.101	0.460	0.646
Condition2_PosA	0.2207	0.187	1.178	0.239
Condition2_PosN	-0.8195	0.137	-5.985	0.000
Condition2_RRAe	-0.5692	0.240	-2.374	0.018
Condition2_RRAn	-0.0424	0.153	-0.277	0.782
Condition2_RRNn	-0.0119	0.134	-0.089	0.929
BldgType_2fmCon	0.0275	0.068	0.405	0.686
BldgType_Duplex	0.0042	0.045	0.093	0.926
BldgType_Twnhs	-0.0489	0.054	-0.905	0.366
BldgType_TwnhsE	-0.0283	0.049	-0.578	0.563
HouseStyle_1.5Unf	-0.0026	0.046	-0.056	0.955
HouseStyle_1Story	-0.0395	0.024	-1.653	0.099

HouseStyle_2.5Fin	-0.1062	0.064	-1.668	0.096
HouseStyle_2.5Unf	0.1057	0.061	1.722	0.085
HouseStyle_2Story	-0.0204	0.019	-1.091	0.275
HouseStyle_SFoyer	-0.0237	0.035	-0.686	0.493
HouseStyle_SLvl	-0.0118	0.030	-0.395	0.693
RoofStyle_Gable	-0.0140	0.094	-0.150	0.881
RoofStyle_Gambrel	-0.0089	0.102	-0.088	0.930
RoofStyle_Hip	-0.0112	0.094	-0.120	0.905
RoofStyle_Mansard	0.0620	0.110	0.563	0.573
RoofStyle_Shed	0.4625	0.185	2.496	0.013
RoofMatl_CompShg	2.7998	0.195	14.367	0.000
RoofMatl_Membran	3.1696	0.262	12.103	0.000
RoofMatl_Metal	3.0197	0.258	11.726	0.000
RoofMatl_Roll	2.8164	0.229	12.314	0.000
RoofMatl_Tar&Grv	2.8128	0.215	13.096	0.000
RoofMatl_WdShake	2.6856	0.223	12.032	0.000
RoofMatl_WdShngl	2.8669	0.199	14.399	0.000
Exterior1st_AsphShn	2.016e-15	8.98e-16	2.245	0.025
Exterior1st_BrkComm	-0.1985	0.146	-1.362	0.174
Exterior1st_BrkFace	0.1115	0.080	1.399	0.162
Exterior1st_CBlock	-0.0092	0.068	-0.134	0.894
Exterior1st_CemntBd	-0.1175	0.118	-0.994	0.321
Exterior1st_HdBoard	-0.0005	0.078	-0.006	0.995
Exterior1st_ImStucc	6.596e-15	7.22e-16	9.133	0.000
Exterior1st_MetalSd	0.0762	0.086	0.886	0.376
Exterior1st_Plywood	0.0098	0.077	0.127	0.899
Exterior1st_Stone	0.0983	0.176	0.557	0.578
Exterior1st_Stucco	0.0082	0.089	0.093	0.926
Exterior1st_VinylSd	0.0303	0.079	0.381	0.703
Exterior1st_Wd Sdng	-0.0202	0.076	-0.264	0.792
Exterior1st_WdShing	0.0095	0.081	0.116	0.907
Exterior2nd_AsphShn	-0.0312	0.117	-0.266	0.790
Exterior2nd_Brk Cmn	-0.0165	0.110	-0.150	0.881
Exterior2nd_BrkFace	-0.0890	0.080	-1.118	0.264
Exterior2nd_CBlock	-0.0092	0.068	-0.134	0.894
Exterior2nd_CmentBd	0.1360	0.116	1.172	0.242
Exterior2nd_HdBoard	-0.0311	0.075	-0.414	0.679
Exterior2nd_ImStucc	0.0148	0.085	0.173	0.862
Exterior2nd_MetalSd	-0.0783	0.083	-0.937	0.349
Exterior2nd_Other	-0.1486	0.136	-1.091	0.275
Exterior2nd_Plywood	-0.0398	0.073	-0.543	0.587
Exterior2nd_Stone	-0.0042	0.105	-0.040	0.968
Exterior2nd_Stucco	-0.0236	0.086	-0.275	0.784
Exterior2nd_VinylSd	-0.0373	0.076	-0.494	0.621
Exterior2nd_Wd Sdng	-0.0101	0.073	-0.137	0.891
Exterior2nd_Wd Shng	-0.0250	0.076	-0.327	0.744
MasVnrType_BrkFace	0.0280	0.029	0.961	0.337
MasVnrType_None	0.0251	0.029	0.856	0.392

MasVnrType_Stone	0.0319	0.032	1.010	0.313
ExterQual_Fa	-0.0466	0.063	-0.741	0.459
ExterQual_Gd	0.0165	0.027	0.617	0.537
ExterQual_TA	0.0200	0.030	0.678	0.498
ExterCond_Fa	-0.1134	0.088	-1.287	0.199
ExterCond_Gd	-0.0753	0.083	-0.912	0.362
ExterCond_Po	-0.1086	0.156	-0.695	0.487
ExterCond_TA	-0.0471	0.082	-0.573	0.567
Foundation_CBlock	0.0239	0.017	1.419	0.156
Foundation_PConc	0.0406	0.018	2.236	0.026
Foundation_Slab	-0.0387	0.049	-0.788	0.431
Foundation_Stone	0.0938	0.055	1.721	0.086
Foundation_Wood	-0.1449	0.072	-2.022	0.043
BsmtQual_Fa	-0.0149	0.031	-0.479	0.632
BsmtQual_Gd	-0.0269	0.017	-1.594	0.111
BsmtQual_TA	-0.0394	0.021	-1.909	0.057
BsmtCond_Gd	0.0441	0.028	1.573	0.116
BsmtCond_Po	0.3252	0.151	2.147	0.032
BsmtCond_TA	0.0304	0.022	1.353	0.176
BsmtExposure_Gd	0.0234	0.016	1.437	0.151
BsmtExposure_Mn	-0.0033	0.016	-0.204	0.838
BsmtExposure_No	-0.0133	0.012	-1.109	0.268
BsmtFinType1_BLQ	-0.0097	0.015	-0.657	0.511
BsmtFinType1_GLQ	0.0153	0.013	1.166	0.244
BsmtFinType1_LwQ	-0.0212	0.020	-1.083	0.279
BsmtFinType1_Rec	-0.0093	0.016	-0.574	0.566
BsmtFinType1_Unf	-0.0038	0.015	-0.251	0.802
BsmtFinType2_BLQ	-0.0665	0.038	-1.742	0.082
BsmtFinType2_GLQ	0.0058	0.049	0.118	0.906
BsmtFinType2_LwQ	-0.0301	0.037	-0.806	0.420
BsmtFinType2_Rec	-0.0225	0.034	-0.656	0.512
BsmtFinType2_Unf	-0.0082	0.034	-0.241	0.809
Heating_GasA	0.1424	0.124	1.153	0.249
Heating_GasW	0.2225	0.131	1.693	0.091
Heating_Grav	0.1932	0.160	1.209	0.227
Heating_OthW	0.1105	0.154	0.717	0.474
Heating_Wall	0.1678	0.150	1.119	0.263
HeatingQC_Fa	-0.0165	0.025	-0.666	0.505
HeatingQC_Gd	-0.0200	0.011	-1.813	0.070
HeatingQC_Po	-0.0873	0.131	-0.666	0.505
HeatingQC_TA	-0.0324	0.011	-2.980	0.003
CentralAir_Y	0.0605	0.021	2.879	0.004
Electrical_FuseF	-0.0284	0.033	-0.858	0.391
Electrical_FuseP	-0.0688	0.096	-0.715	0.475
Electrical_Mix	-0.0942	0.226	-0.417	0.677
Electrical_SBrkr	-0.0024	0.016	-0.153	0.878
KitchenQual_Fa	-0.0073	0.036	-0.200	0.841
KitchenQual_Gd	-0.0600	0.019	-3.081	0.002

KitchenQual_TA	-0.0551	0.022	-2.527	0.012
Functional_Maj2	-0.2985	0.082	-3.654	0.000
Functional_Min1	0.0624	0.047	1.326	0.185
Functional_Min2	0.0240	0.047	0.516	0.606
Functional_Mod	-0.0563	0.059	-0.957	0.339
Functional_Sev	6.608e-16	3.09e-16	2.139	0.033
Functional_Typ	0.0756	0.042	1.818	0.069
FireplaceQu_Fa	-0.0046	0.027	-0.173	0.863
FireplaceQu_Gd	0.0048	0.016	0.303	0.762
FireplaceQu_Po	-0.0036	0.035	-0.100	0.920
FireplaceQu_TA	0.0113	0.017	0.650	0.516
GarageType_Attchd	0.1118	0.055	2.023	0.043
GarageType_Basment	0.1019	0.064	1.588	0.113
GarageType_BuiltIn	0.0975	0.058	1.676	0.094
GarageType_CarPort	0.1229	0.072	1.696	0.090
GarageType_Detchd	0.0986	0.055	1.789	0.074
GarageFinish_RFn	-0.0030	0.010	-0.287	0.774
GarageFinish_Unf	-0.0124	0.013	-0.970	0.332
GarageQual_Fa	-0.0681	0.041	-1.663	0.097
GarageQual_Gd	0.0165	0.050	0.331	0.741
GarageQual_Po	-0.1694	0.111	-1.530	0.126
GarageQual_TA	-0.0074	0.040	-0.185	0.853
GarageCond_Fa	-0.1160	0.045	-2.555	0.011
GarageCond_Gd	-0.0605	0.056	-1.078	0.281
GarageCond_Po	0.0170	0.086	0.198	0.843
GarageCond_TA	-0.0690	0.041	-1.696	0.090
PavedDrive_P	-0.0022	0.030	-0.073	0.942
PavedDrive_Y	0.0141	0.018	0.766	0.444
PoolQC_Fa	0.0108	0.166	0.065	0.948
PoolQC_Gd	0.1406	0.148	0.948	0.343
Fence_GdWo	-0.0324	0.019	-1.734	0.083
Fence_MnPrv	-0.0112	0.012	-0.917	0.359
Fence_MnWw	-0.0471	0.041	-1.143	0.254
MiscFeature_Othr	-0.1506	0.149	-1.013	0.312
MiscFeature_Shed	-0.0071	0.021	-0.344	0.731
MiscFeature_TenC	-0.2110	0.169	-1.249	0.212
SaleType_CWD	0.0862	0.062	1.385	0.166
SaleType_Con	0.0979	0.084	1.163	0.245
SaleType_ConLD	0.2808	0.059	4.787	0.000
SaleType_ConLI	-0.0379	0.062	-0.609	0.543
SaleType_ConLw	-0.0452	0.069	-0.656	0.512
SaleType_New	0.1404	0.076	1.845	0.065
SaleType_Oth	0.0559	0.070	0.804	0.422
SaleType_WD	-0.0137	0.021	-0.637	0.524
SaleCondition_AdjLand	0.1236	0.080	1.550	0.121
SaleCondition_Alloca	0.0511	0.047	1.093	0.275
SaleCondition_Family	-0.0018	0.034	-0.054	0.957
SaleCondition_Normal	0.0588	0.016	3.781	0.000

SaleCondition_Partial	-0.0480	0.073	-0.656	0.512
-----------------------	---------	-------	--------	-------

```
=====
Omnibus:                    313.878    Durbin-Watson:
Prob(Omnibus):              0.000    Jarque-Bera (JB):          410
Skew:                      -0.858    Prob(JB):
Kurtosis:                  12.020    Cond. No.                1.3
=====
```

```
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is co
[2] The smallest eigenvalue is 1.26e-21. This might indicate that there a
strong multicollinearity problems or that the design matrix is singular.
"""
```

```
In [452]: local_train_x.shape
```

```
Out[452]: (1168, 246)
```

```
In [453]: local_train_x.drop(['MasVnrArea', 'LowQualFinSF', 'BsmtHalfBath', 'Condition
local_test_x.drop(['MasVnrArea', 'LowQualFinSF', 'BsmtHalfBath', 'Condition1
local_train_x.shape
```

```
Out[453]: (1168, 213)
```

```
In [454]: #Refitting model
clf = sm.OLS(local_train_y, local_train_x)
result = clf.fit()
result.summary()
```

```
Out[454]: <class 'statsmodels.iolib.summary.Summary'>
"""
```

```

                                OLS Regression Results
=====
Dep. Variable:                  SalePrice    R-squared:
Model:                            OLS      Adj. R-squared:
Method:                 Least Squares    F-statistic:          7.31
Date:                 Sun, 02 Oct 2016    Prob (F-statistic):
Time:                  23:10:47    Log-Likelihood:
No. Observations:                1168    AIC:
Df Residuals:                    959    BIC:
Df Model:                        209
Covariance Type:                nonrobust
=====
                                coef    std err          t      P>|t|      [9
-----
Id                -6.633e-06    8.03e-06    -0.826    0.409    -2.
MSSubClass        -0.0003      0.000    -0.753    0.452
LotFrontage        0.0005      0.000     2.079    0.038     2.
LotArea           2.359e-06    5.89e-07     4.005    0.000     1.
```


OverallQual	0.0372	0.005	7.023	0.000
OverallCond	0.0369	0.005	8.195	0.000
YearBuilt	0.0017	0.000	4.495	0.000
YearRemodAdd	0.0007	0.000	2.593	0.010
BsmtFinSF1	5.466e-05	1.14e-05	4.797	0.000
BsmtFinSF2	5.584e-05	2.24e-05	2.497	0.013
BsmtUnfSF	-1.286e-05	1.19e-05	-1.079	0.281
TotalBsmtSF	9.763e-05	1.8e-05	5.428	0.000
1stFlrSF	7.892e-05	9.88e-05	0.799	0.425
2ndFlrSF	6.108e-05	9.4e-05	0.650	0.516
GrLivArea	0.0001	9.75e-05	1.518	0.129
BsmtFullBath	0.0260	0.010	2.710	0.007
FullBath	0.0109	0.011	0.969	0.333
HalfBath	0.0140	0.011	1.296	0.195
BedroomAbvGr	0.0050	0.007	0.716	0.474
KitchenAbvGr	-0.0548	0.027	-2.042	0.041
TotRmsAbvGrd	0.0092	0.005	1.853	0.064
Fireplaces	0.0161	0.010	1.597	0.111
GarageYrBlt	-0.0004	0.000	-1.193	0.233
GarageCars	0.0238	0.012	1.991	0.047
GarageArea	0.0001	4.15e-05	2.745	0.006
WoodDeckSF	0.0001	2.97e-05	3.596	0.000
OpenPorchSF	6.734e-05	6.18e-05	1.089	0.276
EnclosedPorch	3.093e-05	6.58e-05	0.470	0.638
3SsnPorch	0.0001	0.000	1.325	0.185
ScreenPorch	0.0003	6.58e-05	3.917	0.000
PoolArea	0.0002	0.000	1.232	0.218
MiscVal	3.682e-06	6.96e-06	0.529	0.597
MoSold	0.0008	0.001	0.629	0.529
YrSold	0.0015	0.000	3.140	0.002
MSZoning_FV	0.4906	0.059	8.350	0.000
MSZoning_RH	0.4883	0.058	8.463	0.000
MSZoning_RL	0.4700	0.049	9.565	0.000
MSZoning_RM	0.4399	0.046	9.547	0.000
Street_Pave	0.1029	0.059	1.749	0.081
Alley_Pave	0.0469	0.024	1.977	0.048
LotShape_IR2	0.0318	0.022	1.426	0.154
LotShape_IR3	0.0133	0.048	0.274	0.784
LotShape_Reg	0.0039	0.008	0.464	0.643
LandContour_HLS	0.0444	0.027	1.671	0.095
LandContour_Low	-0.0329	0.033	-1.006	0.315
LandContour_Lvl	0.0247	0.019	1.301	0.194
Utilities_NoSeWa	-0.1962	0.121	-1.617	0.106
LotConfig_CulDSac	0.0342	0.018	1.942	0.052
LotConfig_FR2	-0.0295	0.021	-1.408	0.159
LotConfig_FR3	-0.0664	0.088	-0.754	0.451
LotConfig_Inside	-0.0092	0.009	-0.994	0.320
LandSlope_Mod	0.0648	0.022	2.959	0.003

LandSlope_Sev	-0.1556	0.064	-2.445	0.015
Neighborhood_Blueste	-0.1292	0.115	-1.128	0.260
Neighborhood_BrDale	-0.0934	0.055	-1.711	0.087
Neighborhood_BrkSide	-0.0275	0.049	-0.560	0.576
Neighborhood_ClearCr	-0.0042	0.048	-0.087	0.931
Neighborhood_CollgCr	-0.0500	0.038	-1.324	0.186
Neighborhood_Crawfor	0.0670	0.044	1.516	0.130
Neighborhood_Edwards	-0.1052	0.042	-2.524	0.012
Neighborhood_Gilbert	-0.0551	0.040	-1.379	0.168
Neighborhood_IDOTRR	-0.0464	0.055	-0.845	0.399
Neighborhood_MeadowV	-0.2256	0.056	-4.006	0.000
Neighborhood_Mitchel	-0.0984	0.043	-2.301	0.022
Neighborhood_NAmes	-0.0655	0.041	-1.605	0.109
Neighborhood_NPkVill	-0.0356	0.056	-0.640	0.522
Neighborhood_NWAmes	-0.0696	0.042	-1.669	0.095
Neighborhood_NoRidge	-0.0055	0.043	-0.128	0.898
Neighborhood_NridgHt	0.0747	0.039	1.915	0.056
Neighborhood_OldTown	-0.0675	0.049	-1.371	0.171
Neighborhood_SWISU	-0.0278	0.049	-0.568	0.570
Neighborhood_Sawyer	-0.0723	0.042	-1.708	0.088
Neighborhood_SawyerW	-0.0464	0.040	-1.158	0.247
Neighborhood_Somerst	-0.0046	0.046	-0.100	0.920
Neighborhood_StoneBr	0.1035	0.042	2.452	0.014
Neighborhood_Timber	-0.0274	0.042	-0.655	0.513
Neighborhood_Veenker	0.0199	0.052	0.384	0.701
Condition1_Feedr	0.0692	0.024	2.920	0.004
Condition1_Norm	0.1111	0.019	5.824	0.000
Condition1_PosA	0.0722	0.050	1.455	0.146
Condition1_PosN	0.1063	0.038	2.810	0.005
Condition1_RRAn	0.0840	0.033	2.548	0.011
Condition1_RRNe	0.0478	0.081	0.588	0.557
Condition1_RRNn	0.1093	0.060	1.813	0.070
Condition2_Feedr	0.0975	0.083	1.178	0.239
Condition2_Norm	0.0576	0.062	0.925	0.355
Condition2_PosA	0.2246	0.169	1.333	0.183
Condition2_PosN	-0.8095	0.110	-7.386	0.000
Condition2_RRAe	-0.5561	0.219	-2.542	0.011
Condition2_RRAn	-0.0331	0.128	-0.259	0.795
BldgType_2fmCon	0.0262	0.052	0.507	0.612
BldgType_Twnhs	-0.0492	0.048	-1.024	0.306
BldgType_TwnhsE	-0.0278	0.043	-0.643	0.521
HouseStyle_1Story	-0.0392	0.021	-1.864	0.063
HouseStyle_2.5Fin	-0.1049	0.061	-1.718	0.086
HouseStyle_2.5Unf	0.1054	0.059	1.780	0.075
HouseStyle_2Story	-0.0198	0.018	-1.122	0.262
HouseStyle_SFoyer	-0.0227	0.032	-0.701	0.484
HouseStyle_SLvl1	-0.0100	0.027	-0.367	0.714
RoofStyle_Mansard	0.0788	0.056	1.397	0.163

RoofStyle_Shed	0.4802	0.164	2.934	0.003	
RoofMatl_CompShg	2.8039	0.181	15.477	0.000	
RoofMatl_Membran	3.1916	0.229	13.915	0.000	
RoofMatl_Metal	3.0358	0.226	13.455	0.000	
RoofMatl_Roll	2.8252	0.212	13.300	0.000	
RoofMatl_Tar&Grv	2.8268	0.184	15.335	0.000	
RoofMatl_WdShake	2.6833	0.210	12.780	0.000	
RoofMatl_WdShngl	2.8706	0.185	15.488	0.000	
Exterior1st_AsphShn	3.02e-15	4.82e-16	6.263	0.000	2.
Exterior1st_BrkComm	-0.2126	0.093	-2.278	0.023	
Exterior1st_BrkFace	0.1049	0.028	3.781	0.000	
Exterior1st_CemntBd	-0.1236	0.085	-1.453	0.147	
Exterior1st_ImStucc	-4.526e-16	2.97e-16	-1.525	0.128	-1.
Exterior1st_MetalSd	0.0711	0.041	1.738	0.083	
Exterior1st_Stone	0.1054	0.154	0.683	0.495	
Exterior1st_VinylSd	0.0238	0.039	0.611	0.542	
Exterior1st_Wd Sdng	-0.0288	0.020	-1.459	0.145	
Exterior2nd_AsphShn	-0.0230	0.086	-0.266	0.790	
Exterior2nd_BrkFace	-0.0800	0.034	-2.355	0.019	
Exterior2nd_CmentBd	0.1421	0.086	1.651	0.099	
Exterior2nd_HdBoard	-0.0305	0.021	-1.460	0.145	
Exterior2nd_MetalSd	-0.0731	0.042	-1.727	0.085	
Exterior2nd_Other	-0.1430	0.119	-1.204	0.229	
Exterior2nd_Plywood	-0.0322	0.021	-1.558	0.120	
Exterior2nd_Stucco	-0.0130	0.034	-0.389	0.698	
Exterior2nd_VinylSd	-0.0304	0.039	-0.773	0.439	
Exterior2nd_Wd Shng	-0.0164	0.025	-0.649	0.517	
MasVnrType_BrkFace	0.0283	0.029	0.994	0.321	
MasVnrType_None	0.0249	0.028	0.891	0.373	
MasVnrType_Stone	0.0321	0.031	1.037	0.300	
ExterQual_Fa	-0.0533	0.054	-0.997	0.319	
ExterQual_Gd	0.0135	0.025	0.538	0.591	
ExterQual_TA	0.0168	0.028	0.601	0.548	
ExterCond_Fa	-0.1184	0.084	-1.404	0.161	
ExterCond_Gd	-0.0800	0.078	-1.021	0.307	
ExterCond_Po	-0.1198	0.149	-0.802	0.423	
ExterCond_TA	-0.0519	0.078	-0.665	0.506	
Foundation_CBlock	0.0250	0.016	1.557	0.120	
Foundation_PConc	0.0408	0.017	2.349	0.019	
Foundation_Slab	-0.0317	0.043	-0.744	0.457	
Foundation_Stone	0.0917	0.053	1.742	0.082	
Foundation_Wood	-0.1447	0.070	-2.066	0.039	
BsmtQual_Fa	-0.0185	0.029	-0.628	0.530	
BsmtQual_Gd	-0.0292	0.016	-1.807	0.071	
BsmtQual_TA	-0.0418	0.020	-2.141	0.032	
BsmtCond_Gd	0.0432	0.027	1.609	0.108	
BsmtCond_Po	0.3258	0.146	2.236	0.026	
BsmtCond_TA	0.0288	0.021	1.376	0.169	

BsmtExposure_Gd	0.0240	0.015	1.609	0.108	
BsmtExposure_No	-0.0128	0.009	-1.362	0.174	
BsmtFinType1_BLQ	-0.0075	0.013	-0.573	0.567	
BsmtFinType1_GLQ	0.0166	0.011	1.485	0.138	
BsmtFinType1_LwQ	-0.0191	0.018	-1.090	0.276	
BsmtFinType1_Rec	-0.0075	0.014	-0.529	0.597	
BsmtFinType2_BLQ	-0.0620	0.026	-2.396	0.017	
BsmtFinType2_LwQ	-0.0254	0.024	-1.055	0.292	
BsmtFinType2_Rec	-0.0176	0.021	-0.841	0.400	
Heating_GasA	0.1327	0.118	1.122	0.262	
Heating_GasW	0.2099	0.125	1.677	0.094	
Heating_Grav	0.1801	0.146	1.232	0.218	
Heating_OthW	0.1028	0.148	0.696	0.487	
Heating_Wall	0.1601	0.140	1.141	0.254	
HeatingQC_Fa	-0.0175	0.024	-0.741	0.459	
HeatingQC_Gd	-0.0203	0.011	-1.899	0.058	
HeatingQC_Po	-0.0924	0.116	-0.793	0.428	
HeatingQC_TA	-0.0330	0.011	-3.134	0.002	
CentralAir_Y	0.0602	0.020	3.031	0.002	
Electrical_FuseF	-0.0258	0.029	-0.898	0.370	
Electrical_FuseP	-0.0702	0.086	-0.820	0.412	
Electrical_Mix	-0.0907	0.217	-0.418	0.676	
KitchenQual_Gd	-0.0585	0.016	-3.590	0.000	
KitchenQual_TA	-0.0531	0.017	-3.109	0.002	
Functional_Maj2	-0.2975	0.079	-3.787	0.000	
Functional_Min1	0.0631	0.046	1.387	0.166	
Functional_Min2	0.0262	0.045	0.585	0.559	
Functional_Mod	-0.0569	0.057	-0.992	0.321	
Functional_Sev	-9.474e-16	1.67e-16	-5.671	0.000	-1.
Functional_Typ	0.0759	0.040	1.893	0.059	
FireplaceQu_Gd	0.0061	0.013	0.457	0.648	
FireplaceQu_TA	0.0131	0.014	0.916	0.360	
GarageType_Attchd	0.1133	0.048	2.361	0.018	
GarageType_Basment	0.1036	0.058	1.800	0.072	
GarageType_BuiltIn	0.0991	0.051	1.944	0.052	
GarageType_CarPort	0.1280	0.064	1.994	0.046	
GarageType_Detchd	0.1005	0.048	2.107	0.035	
GarageFinish_RFn	-0.0033	0.010	-0.327	0.744	
GarageFinish_Unf	-0.0125	0.012	-1.019	0.308	
GarageQual_Fa	-0.0613	0.023	-2.657	0.008	
GarageQual_Gd	0.0247	0.035	0.705	0.481	
GarageQual_Po	-0.1570	0.111	-1.408	0.159	
GarageCond_Fa	-0.1272	0.054	-2.358	0.019	
GarageCond_Gd	-0.0717	0.066	-1.079	0.281	
GarageCond_TA	-0.0790	0.048	-1.639	0.102	
PavedDrive_Y	0.0149	0.015	0.980	0.327	
PoolQC_Gd	0.1282	0.117	1.095	0.274	
Fence_GdWo	-0.0321	0.018	-1.780	0.075	

Fence_MnPrv	-0.0109	0.012	-0.931	0.352
Fence_MnWw	-0.0467	0.040	-1.170	0.242
MiscFeature_Othr	-0.1512	0.141	-1.069	0.285
MiscFeature_Shed	-0.0073	0.020	-0.362	0.718
MiscFeature_TenC	-0.2123	0.132	-1.614	0.107
SaleType_CWD	0.0863	0.060	1.433	0.152
SaleType_Con	0.0958	0.082	1.162	0.246
SaleType_ConLD	0.2811	0.057	4.970	0.000
SaleType_ConLI	-0.0406	0.060	-0.681	0.496
SaleType_ConLw	-0.0433	0.066	-0.658	0.511
SaleType_New	0.1395	0.074	1.884	0.060
SaleType_Oth	0.0544	0.068	0.799	0.424
SaleType_WD	-0.0137	0.021	-0.667	0.505
SaleCondition_AdjLand	0.1270	0.075	1.690	0.091
SaleCondition_Alloca	0.0509	0.044	1.149	0.251
SaleCondition_Normal	0.0588	0.013	4.387	0.000
SaleCondition_Partial	-0.0477	0.071	-0.670	0.503

```
=====
Omnibus:                314.318    Durbin-Watson:
Prob(Omnibus):           0.000    Jarque-Bera (JB):           413
Skew:                    -0.858    Prob(JB):
Kurtosis:                12.051    Cond. No.                9.2
=====
```

Warnings:

```
[1] Standard Errors assume that the covariance matrix of the errors is co
[2] The smallest eigenvalue is 2.84e-23. This might indicate that there a
strong multicollinearity problems or that the design matrix is singular.
"""
```

```
In [455]: local_train_x.drop(['2ndFlrSF', 'EnclosedPorch', 'MiscVal', 'MoSold', 'LotSha
local_test_x.drop(['2ndFlrSF', 'EnclosedPorch', 'MiscVal', 'MoSold', 'LotShap
local_train_x.shape
```

```
Out[455]: (1168, 176)
```

```
In [456]: #Refitting model
clf = sm.OLS(local_train_y, local_train_x)
result = clf.fit()
result.summary()
```

```
Out[456]: <class 'statsmodels.iolib.summary.Summary'>
"""
```

```

                                OLS Regression Results
=====
Dep. Variable:                SalePrice    R-squared:
Model:                        OLS         Adj. R-squared:
Method:                       Least Squares    F-statistic:           9.12
Date:                         Sun, 02 Oct 2016    Prob (F-statistic):
```

Time: 23:10:47 Log-Likelihood:
 No. Observations: 1168 AIC:
 Df Residuals: 996 BIC:
 Df Model: 172
 Covariance Type: nonrobust

	coef	std err	t	P> t	[95%]
Id	-7.964e-06	7.77e-06	-1.025	0.306	-2.4
MSSubClass	-0.0004	0.000	-2.967	0.003	
LotFrontage	0.0006	0.000	2.591	0.010	
LotArea	2.444e-06	5.27e-07	4.642	0.000	1.4
OverallQual	0.0367	0.005	7.427	0.000	
OverallCond	0.0378	0.004	8.905	0.000	
YearBuilt	0.0017	0.000	5.035	0.000	
YearRemodAdd	0.0008	0.000	2.803	0.005	
BsmtFinSF1	5.359e-05	1.08e-05	4.951	0.000	3.4
BsmtFinSF2	5.41e-05	2.13e-05	2.544	0.011	1.4
BsmtUnfSF	-1.185e-05	1.11e-05	-1.067	0.286	-3.4
TotalBsmtSF	9.583e-05	1.69e-05	5.668	0.000	6.4
1stFlrSF	1.757e-05	3.16e-05	0.556	0.578	-4.4
GrLivArea	0.0002	2.43e-05	8.739	0.000	
BsmtFullBath	0.0272	0.009	2.942	0.003	
FullBath	0.0080	0.011	0.752	0.452	
HalfBath	0.0137	0.010	1.312	0.190	
BedroomAbvGr	0.0067	0.007	0.997	0.319	
KitchenAbvGr	-0.0383	0.023	-1.640	0.101	
TotRmsAbvGrd	0.0085	0.005	1.835	0.067	
Fireplaces	0.0190	0.007	2.596	0.010	
GarageYrBlt	-0.0003	0.000	-1.137	0.256	
GarageCars	0.0245	0.012	2.133	0.033	
GarageArea	0.0001	3.97e-05	2.926	0.004	3.4
WoodDeckSF	0.0001	2.88e-05	3.842	0.000	5.4
OpenPorchSF	6.825e-05	5.92e-05	1.153	0.249	-4.4
3SsnPorch	0.0001	0.000	1.314	0.189	-6.4
ScreenPorch	0.0003	6.34e-05	4.154	0.000	
PoolArea	0.0001	0.000	1.158	0.247	-9.4
YrSold	0.0014	0.000	3.323	0.001	
MSZoning_FV	0.4987	0.051	9.742	0.000	
MSZoning_RH	0.4890	0.056	8.791	0.000	
MSZoning_RL	0.4716	0.048	9.923	0.000	
MSZoning_RM	0.4338	0.045	9.733	0.000	
Street_Pave	0.1060	0.056	1.905	0.057	
Alley_Pave	0.0445	0.023	1.964	0.050	4.4
LotShape_IR2	0.0278	0.021	1.303	0.193	
LandContour_HLS	0.0440	0.025	1.742	0.082	
LandContour_Low	-0.0269	0.031	-0.867	0.386	
LandContour_Lvl	0.0278	0.018	1.557	0.120	

Utilities_NoSeWa	-0.2121	0.114	-1.853	0.064
LotConfig_CulDSac	0.0377	0.017	2.272	0.023
LotConfig_FR2	-0.0255	0.020	-1.284	0.199
LotConfig_FR3	-0.0683	0.086	-0.796	0.426
LotConfig_Inside	-0.0075	0.009	-0.844	0.399
LandSlope_Mod	0.0646	0.021	3.072	0.002
LandSlope_Sev	-0.1609	0.059	-2.749	0.006
Neighborhood_Blueste	-0.1133	0.107	-1.055	0.292
Neighborhood_BrDale	-0.0691	0.039	-1.769	0.077
Neighborhood_CollgCr	-0.0378	0.015	-2.565	0.010
Neighborhood_Crawfor	0.0836	0.021	3.943	0.000
Neighborhood_Edwards	-0.0880	0.017	-5.100	0.000
Neighborhood_Gilbert	-0.0413	0.019	-2.150	0.032
Neighborhood_IDOTRR	-0.0200	0.030	-0.665	0.506
Neighborhood_MeadowV	-0.2116	0.041	-5.181	0.000
Neighborhood_Mitchel	-0.0860	0.022	-3.981	0.000
Neighborhood_NAmes	-0.0508	0.015	-3.325	0.001
Neighborhood_NWAmes	-0.0545	0.019	-2.873	0.004
Neighborhood_NridgHt	0.0823	0.020	4.145	0.000
Neighborhood_OldTown	-0.0395	0.021	-1.908	0.057
Neighborhood_Sawyer	-0.0587	0.019	-3.048	0.002
Neighborhood_SawyerW	-0.0326	0.019	-1.753	0.080
Neighborhood_StoneBr	0.1114	0.027	4.066	0.000
Condition1_Feedr	0.0660	0.023	2.914	0.004
Condition1_Norm	0.1103	0.018	6.148	0.000
Condition1_PosA	0.0706	0.048	1.460	0.145
Condition1_PosN	0.1017	0.036	2.821	0.005
Condition1_RRAn	0.0875	0.031	2.804	0.005
Condition1_RRNn	0.1176	0.057	2.074	0.038
Condition2_Feedr	0.0817	0.075	1.087	0.277
Condition2_Norm	0.0643	0.053	1.207	0.228
Condition2_PosA	0.2672	0.141	1.897	0.058
Condition2_PosN	-0.8036	0.102	-7.865	0.000
Condition2_RRAe	-0.5166	0.199	-2.600	0.009
BldgType_Twnhs	-0.0306	0.024	-1.253	0.210
HouseStyle_1Story	-0.0403	0.015	-2.701	0.007
HouseStyle_2.5Fin	-0.1303	0.050	-2.612	0.009
HouseStyle_2.5Unf	0.1052	0.055	1.906	0.057
HouseStyle_2Story	-0.0187	0.016	-1.205	0.229
HouseStyle_SFoyer	-0.0165	0.026	-0.625	0.532
RoofStyle_Mansard	0.0653	0.052	1.250	0.212
RoofStyle_Shed	0.4748	0.155	3.063	0.002
RoofMatl_CompShg	2.8404	0.170	16.740	0.000
RoofMatl_Membran	3.2300	0.215	15.003	0.000
RoofMatl_Metal	3.0784	0.211	14.602	0.000
RoofMatl_Roll	2.8548	0.200	14.277	0.000
RoofMatl_Tar&Grv	2.8568	0.172	16.601	0.000
RoofMatl_WdShake	2.7366	0.197	13.891	0.000

RoofMatl_WdShngl	2.9120	0.173	16.859	0.000	
Exterior1st_AsphShn	-5.808e-16	4.55e-16	-1.277	0.202	-1.
Exterior1st_BrkComm	-0.1955	0.090	-2.168	0.030	
Exterior1st_BrkFace	0.1078	0.026	4.219	0.000	
Exterior1st_CemntBd	-0.1215	0.082	-1.478	0.140	
Exterior1st_ImStucc	-2.46e-15	3.77e-16	-6.525	0.000	-3
Exterior1st_MetalSd	0.0672	0.039	1.732	0.084	
Exterior1st_Stone	0.0997	0.147	0.676	0.499	
Exterior1st_Wd Sdng	-0.0238	0.015	-1.552	0.121	
Exterior2nd_BrkFace	-0.0744	0.033	-2.276	0.023	
Exterior2nd_CmentBd	0.1439	0.082	1.762	0.078	
Exterior2nd_HdBoard	-0.0242	0.016	-1.475	0.141	
Exterior2nd_MetalSd	-0.0629	0.039	-1.600	0.110	
Exterior2nd_Other	-0.1107	0.111	-0.999	0.318	
Exterior2nd_Plywood	-0.0268	0.017	-1.547	0.122	
Exterior2nd_VinylSd	-0.0017	0.016	-0.104	0.917	
MasVnrType_BrkFace	0.0310	0.027	1.135	0.257	
MasVnrType_None	0.0283	0.027	1.054	0.292	
MasVnrType_Stone	0.0365	0.030	1.229	0.219	
ExterQual_Fa	-0.0668	0.043	-1.555	0.120	
ExterCond_Fa	-0.0699	0.031	-2.275	0.023	
ExterCond_Gd	-0.0285	0.012	-2.434	0.015	
ExterCond_Po	-0.0825	0.121	-0.684	0.494	
Foundation_CBBlock	0.0219	0.016	1.414	0.158	
Foundation_PConc	0.0412	0.017	2.432	0.015	
Foundation_Slab	-0.0325	0.041	-0.801	0.423	
Foundation_Stone	0.0850	0.051	1.677	0.094	
Foundation_Wood	-0.1528	0.068	-2.257	0.024	
BsmtQual_Gd	-0.0230	0.013	-1.734	0.083	
BsmtQual_TA	-0.0347	0.015	-2.349	0.019	
BsmtCond_Gd	0.0388	0.026	1.503	0.133	
BsmtCond_Po	0.2532	0.106	2.381	0.017	
BsmtCond_TA	0.0240	0.020	1.188	0.235	
BsmtExposure_Gd	0.0237	0.014	1.653	0.099	
BsmtExposure_No	-0.0144	0.009	-1.595	0.111	
BsmtFinType1_GLQ	0.0163	0.010	1.558	0.120	
BsmtFinType1_LwQ	-0.0166	0.016	-1.021	0.308	
BsmtFinType2_BLQ	-0.0636	0.025	-2.573	0.010	
BsmtFinType2_LwQ	-0.0278	0.023	-1.195	0.233	
BsmtFinType2_Rec	-0.0166	0.020	-0.826	0.409	
Heating_GasA	0.1447	0.113	1.277	0.202	
Heating_GasW	0.2227	0.120	1.852	0.064	
Heating_Grav	0.1645	0.138	1.194	0.233	
Heating_OthW	0.1085	0.141	0.768	0.443	
Heating_Wall	0.1650	0.134	1.229	0.219	
HeatingQC_Fa	-0.0178	0.023	-0.787	0.432	
HeatingQC_Gd	-0.0203	0.010	-1.960	0.050	
HeatingQC_Po	-0.0869	0.113	-0.769	0.442	

HeatingQC_TA	-0.0334	0.010	-3.283	0.001	
CentralAir_Y	0.0543	0.019	2.861	0.004	
Electrical_FuseF	-0.0282	0.028	-1.010	0.313	
Electrical_FuseP	-0.0673	0.078	-0.863	0.388	
KitchenQual_Gd	-0.0559	0.015	-3.778	0.000	
KitchenQual_TA	-0.0493	0.016	-3.107	0.002	
Functional_Maj2	-0.3334	0.065	-5.140	0.000	
Functional_Min1	0.0419	0.029	1.443	0.149	
Functional_Mod	-0.0781	0.045	-1.745	0.081	
Functional_Sev	1.096e-16	6.77e-17	1.619	0.106	-2.0
Functional_Typ	0.0559	0.019	2.871	0.004	
FireplaceQu_TA	0.0089	0.010	0.857	0.391	
GarageType_Attchd	0.1252	0.046	2.730	0.006	
GarageType_Basment	0.1132	0.055	2.059	0.040	
GarageType_BuiltIn	0.1111	0.049	2.287	0.022	
GarageType_CarPort	0.1350	0.062	2.190	0.029	
GarageType_Detchd	0.1086	0.046	2.380	0.017	
GarageFinish_Unf	-0.0105	0.010	-1.039	0.299	
GarageQual_Fa	-0.0635	0.022	-2.831	0.005	
GarageQual_Gd	0.0277	0.034	0.810	0.418	
GarageQual_Po	-0.1541	0.096	-1.598	0.110	
GarageCond_Fa	-0.1381	0.052	-2.670	0.008	
GarageCond_Gd	-0.1003	0.063	-1.580	0.114	
GarageCond_TA	-0.0933	0.046	-2.026	0.043	
PavedDrive_Y	0.0141	0.015	0.961	0.337	
PoolQC_Gd	0.1410	0.114	1.240	0.215	
Fence_GdWo	-0.0345	0.018	-1.967	0.049	
Fence_MnPrv	-0.0107	0.011	-0.942	0.346	
Fence_MnWw	-0.0504	0.039	-1.290	0.197	
MiscFeature_Othr	-0.0999	0.129	-0.776	0.438	
MiscFeature_TenC	-0.2068	0.128	-1.621	0.105	
SaleType_CWD	0.0925	0.056	1.661	0.097	
SaleType_Con	0.1174	0.077	1.534	0.125	
SaleType_ConLD	0.2813	0.050	5.615	0.000	
SaleType_ConLI	-0.0307	0.055	-0.561	0.575	
SaleType_New	0.1045	0.018	5.706	0.000	
SaleType_Oth	0.0600	0.064	0.934	0.351	
SaleCondition_AdjLand	0.1109	0.072	1.549	0.122	
SaleCondition_Alloca	0.0448	0.043	1.044	0.297	
SaleCondition_Normal	0.0569	0.012	4.600	0.000	
=====					
Omnibus:	300.168	Durbin-Watson:			
Prob(Omnibus) :	0.000	Jarque-Bera (JB) :		385	
Skew:	-0.809	Prob(JB) :			
Kurtosis:	11.752	Cond. No.		1.0	
=====					

Warnings:

```
[1] Standard Errors assume that the covariance matrix of the errors is co
[2] The smallest eigenvalue is 2.34e-23. This might indicate that there a
strong multicollinearity problems or that the design matrix is singular.
"""
```

```
In [457]: #Predictions based off our model
preds = result.predict(local_test_x)
preds[0:5]
```

```
Out[457]: array([ 12.26705006,  11.3662339 ,  11.86414711,  12.45139966,  11.764814
```

```
In [458]: #RMSE
rmse = (sum((local_test_y - preds) ** 2) / preds.shape[0]) ** 0.5
rmse
```

```
Out[458]: 0.11376542632433698
```

4 Actual Kaggle Dataset

```
In [459]: #Import the test dataset (doesn't contain the test SalePrice values)
actual_test = pd.read_csv('./data/test.csv')
actual_test.shape
```

```
Out[459]: (1459, 80)
```

```
In [460]: #Use the whole dataset to re-train
#Removing SalesPrices (is the predicting y-variable)
train_y = data['SalePrice']
train_x = data.drop('SalePrice', axis = 1)
```

```
In [461]: #Feature engineer test dataset predictor columns
actual_test = pd.get_dummies(actual_test, drop_first=True)
actual_test = actual_test.fillna(actual_test.mean())
actual_test.drop(['MasVnrArea', 'LowQualFinSF', 'BsmtHalfBath', 'Condition1_
actual_test.drop(['2ndFlrSF', 'EnclosedPorch', 'MiscVal', 'MoSold', 'LotShape

#Also remove columns not in actual_test from train_x
train_x = train_x.iloc[:, train_x.columns.isin(actual_test.columns)]
train_x.shape
```

```
Out[461]: (1460, 162)
```

```
In [462]: #Fitting model
actual_clf = sm.OLS(train_y, train_x)
actual_result = actual_clf.fit()
actual_result.summary()

actual_preds = actual_result.predict(actual_test)
actual_preds[0:5]
```

```
Out[462]: array([ 11.66476646,  11.9162187 ,  12.08137814,  12.19944302,  12.184636
```

```
In [463]: # Converting predictions back to actual housing values
          actual_preds = np.exp(actual_preds) - 1
          actual_preds
```

```
Out[463]: array([ 116396.51463577,  149673.57267838,  176552.3083296 , ...,
                  161159.65985258,  110287.76978393,  222581.85988514])
```

```
In [464]: #Exporting to CSV
          submission = pd.concat([pd.Series(actual_test['Id']), pd.Series(actual_preds)], axis=1)
          submission.to_csv('submission.csv', index = False)
```

5 Model Interpretation

From the results of the linear regression significance tests, the most important features for predicting house sale price include features related to the size of the house (Lot Area, SF of living area), Overall Quality and Condition, its newness (Year built and remodeled), and location (MSZoning, some neighborhoods).

Those are mostly expected features to be important, but one surprising feature that the model indicates to be very important is roof material (P=0.000 for all dummies of RoofMatl)

```
In [ ]:
```