Week10

November 3, 2024

1 STOR 320 Introduction to Data Science

1.1 Week 10: Regularization

2 Part 1

2.0.1 Let's first revisit the Ames housing dataset

[ ]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.formula.api as smf
import statsmodels.api as sm
from statsmodels.stats.outliers_influence import variance_inflation_factor
```

[ ]:
```
ames = pd.read_csv('Ames.csv')

pd.set_option('display.max_rows', 200)
pd.set_option('display.max_columns', 200)
ames.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2930 entries, 0 to 2929
Data columns (total 80 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   MSSubClass     2930 non-null   int64
 1   MSZoning       2930 non-null   object
 2   LotFrontage    2440 non-null   float64
 3   LotArea        2930 non-null   float64
 4   Street         2930 non-null   object
 5   Alley          2930 non-null   object
 6   LotShape       2930 non-null   object
 7   LandContour    2930 non-null   object
 8   Utilities      2930 non-null   object
 9   LotConfig      2930 non-null   object
 10  LandSlope      2930 non-null   object
 11  Neighborhood   2930 non-null   object
```

```
12  Condition1    2930 non-null   object
13  Condition2    2930 non-null   object
14  BldgType      2930 non-null   object
15  HouseStyle    2930 non-null   object
16  OverallQual   2930 non-null   int64
17  OverallCond   2930 non-null   int64
18  YearBuilt     2930 non-null   float64
19  YearRemod.Add 2930 non-null   float64
20  RoofStyle     2930 non-null   object
21  RoofMatl      2930 non-null   object
22  Exterior1st   2930 non-null   object
23  Exterior2nd   2930 non-null   object
24  MasVnrType    1155 non-null   object
25  MasVnrArea    2907 non-null   float64
26  ExterQual     2930 non-null   object
27  ExterCond     2930 non-null   object
28  Foundation    2930 non-null   object
29  BsmtQual      2929 non-null   object
30  BsmtCond      2929 non-null   object
31  BsmtExposure  2926 non-null   object
32  BsmtFinType1  2929 non-null   object
33  BsmtFinSF1    2929 non-null   float64
34  BsmtFinType2  2928 non-null   object
35  BsmtFinSF2    2929 non-null   float64
36  BsmtUnfSF     2929 non-null   float64
37  TotalBsmtSF   2929 non-null   float64
38  Heating       2930 non-null   object
39  HeatingQC     2930 non-null   object
40  CentralAir    2930 non-null   object
41  Electrical    2929 non-null   object
42  X1stFlrSF     2930 non-null   float64
43  X2ndFlrSF     2930 non-null   float64
44  LowQualFinSF  2930 non-null   float64
45  GrLivArea     2930 non-null   float64
46  BsmtFullBath  2928 non-null   float64
47  BsmtHalfBath  2928 non-null   float64
48  FullBath      2930 non-null   float64
49  HalfBath      2930 non-null   float64
50  BedroomAbvGr  2930 non-null   float64
51  KitchenAbvGr  2930 non-null   float64
52  KitchenQual   2930 non-null   object
53  TotRmsAbvGrd  2930 non-null   float64
54  Functional    2930 non-null   object
55  Fireplaces    2930 non-null   float64
56  FireplaceQu   2930 non-null   object
57  GarageType    2930 non-null   object
58  GarageYrBlt   2771 non-null   float64
59  GarageFinish  2928 non-null   object
```

```
 60  GarageCars     2929 non-null   float64
 61  GarageArea     2929 non-null   float64
 62  GarageQual     2929 non-null   object
 63  GarageCond     2929 non-null   object
 64  PavedDrive     2930 non-null   object
 65  WoodDeckSF     2930 non-null   float64
 66  OpenPorchSF    2930 non-null   float64
 67  EnclosedPorch  2930 non-null   float64
 68  X3SsnPorch     2930 non-null   float64
 69  ScreenPorch    2930 non-null   float64
 70  PoolArea       2930 non-null   float64
 71  PoolQC         2930 non-null   object
 72  Fence          2930 non-null   object
 73  MiscFeature    106 non-null    object
 74  MiscVal        2930 non-null   float64
 75  MoSold         2930 non-null   float64
 76  YrSold         2930 non-null   float64
 77  SaleType       2930 non-null   object
 78  SaleCondition  2930 non-null   object
 79  SalePrice      2930 non-null   float64
dtypes: float64(34), int64(3), object(43)
memory usage: 1.8+ MB
```

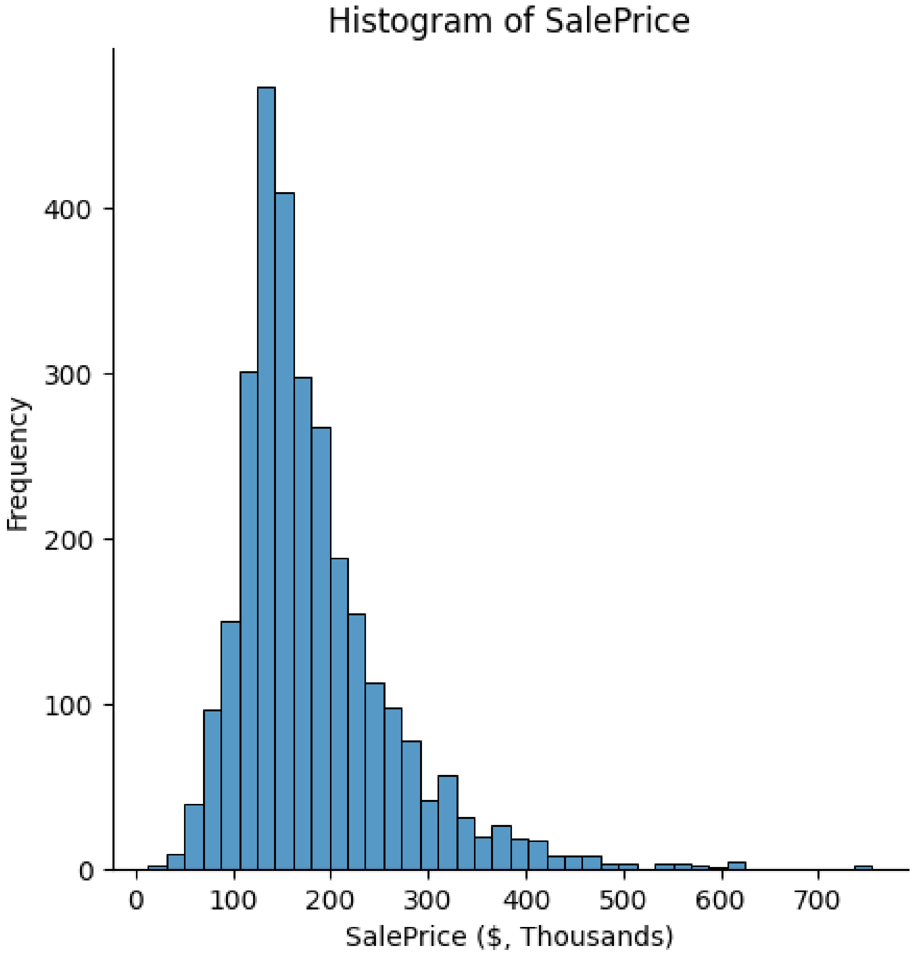
2.0.2 Basic data cleaning and EDA

1.1.1 Dependent Variable

a) A quick visualization of the dependent variable

[ ]:
```python
plt.figure(figsize=(8,6))
ax = sns.displot(ames['SalePrice']/1000, bins=40)

plt.xlabel('SalePrice ($, Thousands)')
plt.ylabel('Frequency')
plt.title('Histogram of SalePrice')

plt.show()
ames['SalePrice'].describe()
```

```
<Figure size 800x600 with 0 Axes>
```



```
[ ]: count      2930.000000
     mean     180796.060068
     std       79886.692357
     min       12789.000000
     25%      129500.000000
     50%      160000.000000
     75%      213500.000000
     max      755000.000000
     Name: SalePrice, dtype: float64
```

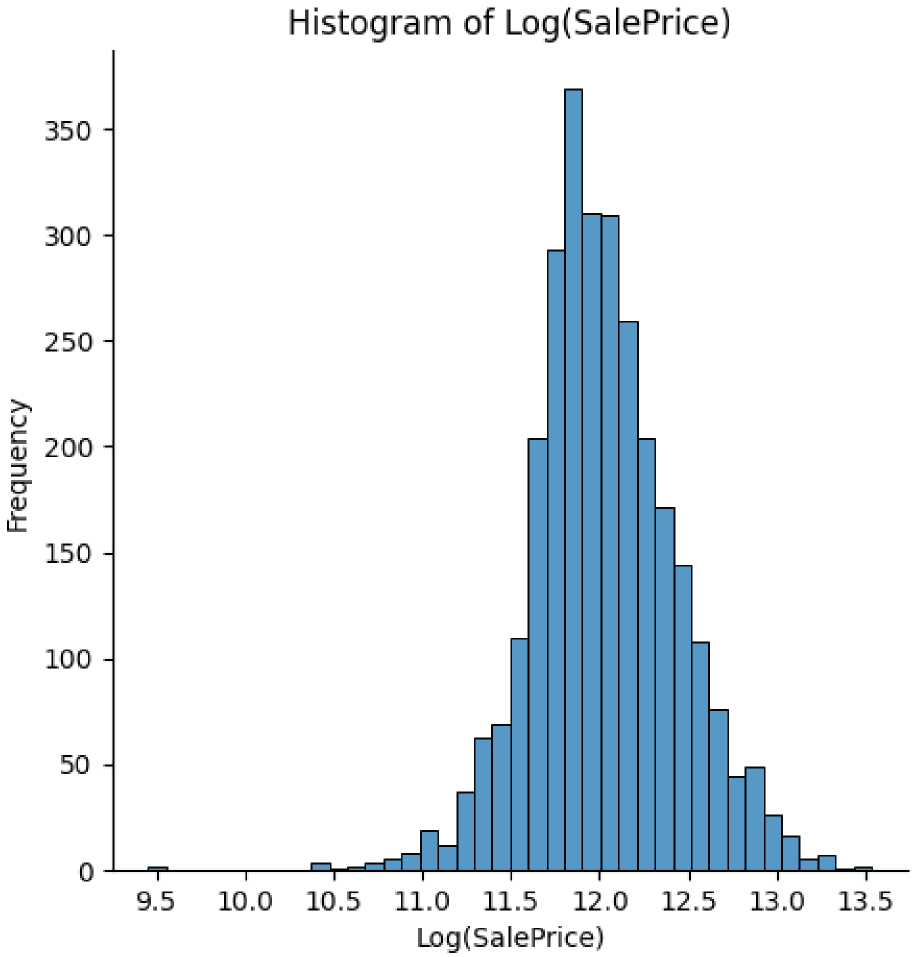
b. Log transformation: Let's take log to be more fair in comparing high vs. low price homes.

```
[ ]: ames['LogSalePrice'] = ames['SalePrice'].apply(np.log)
     ames.drop(columns='SalePrice', inplace=True)
```

```
[ ]: plt.figure(figsize=(8,6))
     ax = sns.displot(ames['LogSalePrice'], bins=40)
```

```
plt.xlabel('Log(SalePrice)')
plt.ylabel('Frequency')
plt.title('Histogram of Log(SalePrice)')

plt.show()
ames['LogSalePrice'].describe()
```

```
<Figure size 800x600 with 0 Axes>
```



```
[ ]: count    2930.000000
     mean       12.020969
     std         0.407587
     min         9.456341
     25%        11.771436
     50%        11.982929
     75%        12.271392
```

```
max         13.534473
Name: LogSalePrice, dtype: float64
```

1.1.2 Independent Variables

a) Column Names: One of the column names contains a dot. We rename it to avoid any problem:

[ ]:
```
ames.rename(columns={'YearRemod.Add':'YearRemodAdd'}, inplace=True)
```

b) Move the new dependent variable (LogSalePrice) to the first column

[ ]:
```
ames = ames[[ames.columns[-1]] + list(ames.columns)[:-1] ]
ames
```

[ ]:
```
      LogSalePrice  MSSubClass MSZoning  LotFrontage  LotArea Street  \
0        12.278393          20       RL        141.0  31770.0   Pave
1        11.561716          20       RH         80.0  11622.0   Pave
2        12.055250          20       RL         81.0  14267.0   Pave
3        12.404924          20       RL         93.0  11160.0   Pave
4        12.154253          60       RL         74.0  13830.0   Pave
…             …          …      …          …      …    …
2925     11.867097          80       RL         37.0   7937.0   Pave
2926     11.782953          20       RL          NaN   8885.0   Pave
2927     11.790557          85       RL         62.0  10441.0   Pave
2928     12.043554          20       RL         77.0  10010.0   Pave
2929     12.144197          60       RL         74.0   9627.0   Pave

         Alley LotShape LandContour Utilities LotConfig LandSlope  \
0     NoAccess      IR1         Lvl    AllPub    Corner       Gtl
1     NoAccess      Reg         Lvl    AllPub    Inside       Gtl
2     NoAccess      IR1         Lvl    AllPub    Corner       Gtl
3     NoAccess      Reg         Lvl    AllPub    Corner       Gtl
4     NoAccess      IR1         Lvl    AllPub    Inside       Gtl
…          …      …          …       …       …        …
2925  NoAccess      IR1         Lvl    AllPub   CulDSac       Gtl
2926  NoAccess      IR1         Low    AllPub    Inside       Mod
2927  NoAccess      Reg         Lvl    AllPub    Inside       Gtl
2928  NoAccess      Reg         Lvl    AllPub    Inside       Mod
2929  NoAccess      Reg         Lvl    AllPub    Inside       Mod

     Neighborhood Condition1 Condition2 BldgType HouseStyle  OverallQual  \
0           NAmes       Norm       Norm     1Fam     1Story            6
1           NAmes      Feedr       Norm     1Fam     1Story            5
2           NAmes       Norm       Norm     1Fam     1Story            6
3           NAmes       Norm       Norm     1Fam     1Story            7
4         Gilbert       Norm       Norm     1Fam     2Story            5
…             …        …        …       …        …          …
2925      Mitchel       Norm       Norm     1Fam       SLvl            6
```

```
2926      Mitchel       Norm       Norm      1Fam     1Story            5
2927      Mitchel       Norm       Norm      1Fam     SFoyer            5
2928      Mitchel       Norm       Norm      1Fam     1Story            5
2929      Mitchel       Norm       Norm      1Fam     2Story            7

      OverallCond  YearBuilt  YearRemodAdd RoofStyle RoofMatl Exterior1st   \
0               5     1960.0        1960.0       Hip  CompShg     BrkFace
1               6     1961.0        1961.0     Gable  CompShg     VinylSd
2               6     1958.0        1958.0       Hip  CompShg     Wd Sdng
3               5     1968.0        1968.0       Hip  CompShg     BrkFace
4               5     1997.0        1998.0     Gable  CompShg     VinylSd
…           …        …           …        …      …         …
2925            6     1984.0        1984.0     Gable  CompShg     HdBoard
2926            5     1983.0        1983.0     Gable  CompShg     HdBoard
2927            5     1992.0        1992.0     Gable  CompShg     HdBoard
2928            5     1974.0        1975.0     Gable  CompShg     HdBoard
2929            5     1993.0        1994.0     Gable  CompShg     HdBoard

     Exterior2nd MasVnrType  MasVnrArea ExterQual ExterCond Foundation   \
0        Plywood      Stone       112.0        TA        TA     CBlock
1        VinylSd        NaN         0.0        TA        TA     CBlock
2        Wd Sdng    BrkFace       108.0        TA        TA     CBlock
3        BrkFace        NaN         0.0        Gd        TA     CBlock
4        VinylSd        NaN         0.0        TA        TA      PConc
…          …        …         …        …       …        …
2925     HdBoard        NaN         0.0        TA        TA     CBlock
2926     HdBoard        NaN         0.0        TA        TA     CBlock
2927     Wd Shng        NaN         0.0        TA        TA      PConc
2928     HdBoard        NaN         0.0        TA        TA     CBlock
2929     HdBoard    BrkFace        94.0        TA        TA      PConc

     BsmtQual BsmtCond BsmtExposure BsmtFinType1  BsmtFinSF1 BsmtFinType2   \
0          TA       Gd           Gd          BLQ       639.0          Unf
1          TA       TA           No          Rec       468.0          LwQ
2          TA       TA           No          ALQ       923.0          Unf
3          TA       TA           No          ALQ      1065.0          Unf
4          Gd       TA           No          GLQ       791.0          Unf
…       …      …          …          …         …           …
2925       TA       TA           Av          GLQ       819.0          Unf
2926       Gd       TA           Av          BLQ       301.0          ALQ
2927       Gd       TA           Av          GLQ       337.0          Unf
2928       Gd       TA           Av          ALQ      1071.0          LwQ
2929       Gd       TA           Av          LwQ       758.0          Unf

      BsmtFinSF2  BsmtUnfSF  TotalBsmtSF Heating HeatingQC CentralAir   \
0            0.0      441.0       1080.0    GasA        Fa          Y
1          144.0      270.0        882.0    GasA        TA          Y
```

```
2            0.0      406.0       1329.0    GasA        TA          Y
3            0.0     1045.0       2110.0    GasA        Ex          Y
4            0.0      137.0        928.0    GasA        Gd          Y
…          …        …          …     …       …        …
2925         0.0      184.0       1003.0    GasA        TA          Y
2926       324.0      239.0        864.0    GasA        TA          Y
2927         0.0      575.0        912.0    GasA        TA          Y
2928       123.0      195.0       1389.0    GasA        Gd          Y
2929         0.0      238.0        996.0    GasA        Ex          Y

     Electrical  X1stFlrSF  X2ndFlrSF  LowQualFinSF  GrLivArea  BsmtFullBath  \
0         SBrkr     1656.0        0.0           0.0     1656.0           1.0
1         SBrkr      896.0        0.0           0.0      896.0           0.0
2         SBrkr     1329.0        0.0           0.0     1329.0           0.0
3         SBrkr     2110.0        0.0           0.0     2110.0           1.0
4         SBrkr      928.0      701.0           0.0     1629.0           0.0
…         …        …        …           …        …           …
2925      SBrkr     1003.0        0.0           0.0     1003.0           1.0
2926      SBrkr      902.0        0.0           0.0      902.0           1.0
2927      SBrkr      970.0        0.0           0.0      970.0           0.0
2928      SBrkr     1389.0        0.0           0.0     1389.0           1.0
2929      SBrkr      996.0     1004.0           0.0     2000.0           0.0

      BsmtHalfBath  FullBath  HalfBath  BedroomAbvGr  KitchenAbvGr  \
0              0.0       1.0       0.0           3.0           1.0
1              0.0       1.0       0.0           2.0           1.0
2              0.0       1.0       1.0           3.0           1.0
3              0.0       2.0       1.0           3.0           1.0
4              0.0       2.0       1.0           3.0           1.0
…            …       …       …           …           …
2925           0.0       1.0       0.0           3.0           1.0
2926           0.0       1.0       0.0           2.0           1.0
2927           1.0       1.0       0.0           3.0           1.0
2928           0.0       1.0       0.0           2.0           1.0
2929           0.0       2.0       1.0           3.0           1.0

     KitchenQual  TotRmsAbvGrd Functional  Fireplaces  FireplaceQu GarageType  \
0             TA           7.0        Typ         2.0           Gd     Attchd
1             TA           5.0        Typ         0.0  NoFirePlace     Attchd
2             Gd           6.0        Typ         0.0  NoFirePlace     Attchd
3             Ex           8.0        Typ         2.0           TA     Attchd
4             TA           6.0        Typ         1.0           TA     Attchd
…          …           …        …         …          …        …
2925          TA           6.0        Typ         0.0  NoFirePlace     Detchd
2926          TA           5.0        Typ         0.0  NoFirePlace     Attchd
2927          TA           6.0        Typ         0.0  NoFirePlace   NoGarage
2928          TA           6.0        Typ         1.0           TA     Attchd
```

```
2929          TA           9.0        Typ         1.0           TA     Attchd

      GarageYrBlt GarageFinish  GarageCars  GarageArea GarageQual GarageCond  \
0          1960.0          Fin         2.0       528.0         TA         TA
1          1961.0          Unf         1.0       730.0         TA         TA
2          1958.0          Unf         1.0       312.0         TA         TA
3          1968.0          Fin         2.0       522.0         TA         TA
4          1997.0          Fin         2.0       482.0         TA         TA
…             …            …         …           …          …          …
2925       1984.0          Unf         2.0       588.0         TA         TA
2926       1983.0          Unf         2.0       484.0         TA         TA
2927          NaN     NoGarage         0.0         0.0   NoGarage   NoGarage
2928       1975.0          RFn         2.0       418.0         TA         TA
2929       1993.0          Fin         3.0       650.0         TA         TA

     PavedDrive  WoodDeckSF  OpenPorchSF  EnclosedPorch  X3SsnPorch  \
0             P       210.0         62.0            0.0         0.0
1             Y       140.0          0.0            0.0         0.0
2             Y       393.0         36.0            0.0         0.0
3             Y         0.0          0.0            0.0         0.0
4             Y       212.0         34.0            0.0         0.0
…          …           …           …             …          …
2925          Y       120.0          0.0            0.0         0.0
2926          Y       164.0          0.0            0.0         0.0
2927          Y        80.0         32.0            0.0         0.0
2928          Y       240.0         38.0            0.0         0.0
2929          Y       190.0         48.0            0.0         0.0

      ScreenPorch  PoolArea  PoolQC    Fence MiscFeature  MiscVal  MoSold  \
0             0.0       0.0  NoPool  NoFence         NaN      0.0     5.0
1           120.0       0.0  NoPool    MnPrv         NaN      0.0     6.0
2             0.0       0.0  NoPool  NoFence        Gar2  12500.0     6.0
3             0.0       0.0  NoPool  NoFence         NaN      0.0     4.0
4             0.0       0.0  NoPool    MnPrv         NaN      0.0     3.0
…           …        …     …      …       …       …     …
2925          0.0       0.0  NoPool    GdPrv         NaN      0.0     3.0
2926          0.0       0.0  NoPool    MnPrv         NaN      0.0     6.0
2927          0.0       0.0  NoPool    MnPrv        Shed    700.0     7.0
2928          0.0       0.0  NoPool  NoFence         NaN      0.0     4.0
2929          0.0       0.0  NoPool  NoFence         NaN      0.0    11.0

      YrSold SaleType SaleCondition
0     2010.0       WD        Normal
1     2010.0       WD        Normal
2     2010.0       WD        Normal
3     2010.0       WD        Normal
4     2010.0       WD        Normal
```

```
…       …       …           …
2925   2006.0      WD         Normal
2926   2006.0      WD         Normal
2927   2006.0      WD         Normal
2928   2006.0      WD         Normal
2929   2006.0      WD         Normal

[2930 rows x 80 columns]
```

2.0.3 1.2 More cleaning: Combine the information from two columns

1.2.1 Fix Condition variables Convert condition variables into dummy variables. We look at both `Condition1` or `Condition2` to decide which conditions each home belongs to.

[ ]:

```
print(np.unique(ames['Condition1']))
print(np.unique(ames['Condition2']))
```

```
['Artery' 'Feedr' 'Norm' 'PosA' 'PosN' 'RRAe' 'RRAn' 'RRNe' 'RRNn']
['Artery' 'Feedr' 'Norm' 'PosA' 'PosN' 'RRAe' 'RRAn' 'RRNn']
```

[ ]:

```
condition_dummy=np.unique(ames['Condition1'])
# In the line above, we implicitly treate 'Norm' as the reference level
ames[condition_dummy]=0

for i in condition_dummy:
    print(i)
    ames.loc[(ames['Condition1']==i) | (ames['Condition2']==i), i]=1

ames.drop(columns=['Condition1', 'Condition2'], inplace=True)
```

```
Artery
Feedr
Norm
PosA
PosN
RRAe
RRAn
RRNe
RRNn
```

[ ]:

```
ames.loc[:,['Artery', 'Feedr', 'Norm']]
```

[ ]:

```
       Artery  Feedr  Norm
0           0      0     1
1           0      1     1
2           0      0     1
3           0      0     1
4           0      0     1
…       …     …    …
```

```
2925        0       0      1
2926        0       0      1
2927        0       0      1
2928        0       0      1
2929        0       0      1

[2930 rows x 3 columns]
```

1.2.2 Encode Exterior Variables

Same treatment as with condition variables

[ ]:
```
print(np.unique(ames['Exterior1st']))
print(np.unique(ames['Exterior2nd']))
```

```
['AsbShng' 'AsphShn' 'BrkComm' 'BrkFace' 'CBlock' 'CemntBd' 'HdBoard'
 'ImStucc' 'MetalSd' 'Plywood' 'PreCast' 'Stone' 'Stucco' 'VinylSd'
 'Wd Sdng' 'WdShing']
['AsbShng' 'AsphShn' 'Brk Cmn' 'BrkFace' 'CBlock' 'CmentBd' 'HdBoard'
 'ImStucc' 'MetalSd' 'Other' 'Plywood' 'PreCast' 'Stone' 'Stucco'
 'VinylSd' 'Wd Sdng' 'Wd Shng']
```

[ ]:
```
# Remove the space in each level to avoid errors in column names later on.
ames['Exterior1st'] = ames['Exterior1st'].str.replace(' ','')
ames['Exterior2nd'] = ames['Exterior2nd'].str.replace(' ','')
print(np.unique(ames['Exterior1st']))
print(np.unique(ames['Exterior2nd']))
```

```
['AsbShng' 'AsphShn' 'BrkComm' 'BrkFace' 'CBlock' 'CemntBd' 'HdBoard'
 'ImStucc' 'MetalSd' 'Plywood' 'PreCast' 'Stone' 'Stucco' 'VinylSd'
 'WdSdng' 'WdShing']
['AsbShng' 'AsphShn' 'BrkCmn' 'BrkFace' 'CBlock' 'CmentBd' 'HdBoard'
 'ImStucc' 'MetalSd' 'Other' 'Plywood' 'PreCast' 'Stone' 'Stucco'
 'VinylSd' 'WdSdng' 'WdShng']
```

[ ]:
```
np.unique(np.concatenate((np.unique(ames['Exterior1st']),np.
 ↪unique(ames['Exterior2nd'])), axis=0 ))
```

[ ]:
```
array(['AsbShng', 'AsphShn', 'BrkCmn', 'BrkComm', 'BrkFace', 'CBlock',
       'CemntBd', 'CmentBd', 'HdBoard', 'ImStucc', 'MetalSd', 'Other',
       'Plywood', 'PreCast', 'Stone', 'Stucco', 'VinylSd', 'WdSdng',
       'WdShing', 'WdShng'], dtype=object)
```

[ ]:
```
exterior_dummy=np.unique(np.concatenate((np.unique(ames['Exterior1st']),np.
 ↪unique(ames['Exterior2nd'])), axis=0 ))
ames[exterior_dummy]=0

for i in exterior_dummy:
    print(i)
    ames.loc[(ames['Exterior1st']==i) | (ames['Exterior2nd']==i), i]=1
```

```
ames.drop(columns=['Exterior1st', 'Exterior2nd'], inplace=True)
```

```
AsbShng
AsphShn
BrkCmn
BrkComm
BrkFace
CBlock
CemntBd
CmentBd
HdBoard
ImStucc
MetalSd
Other
Plywood
PreCast
Stone
Stucco
VinylSd
WdSdng
WdShing
WdShng
```

2.0.4 1.3 Create New Features for Year Built & Year Remodeled

Let's see how sale price is affected by the year built and the year remodeled.

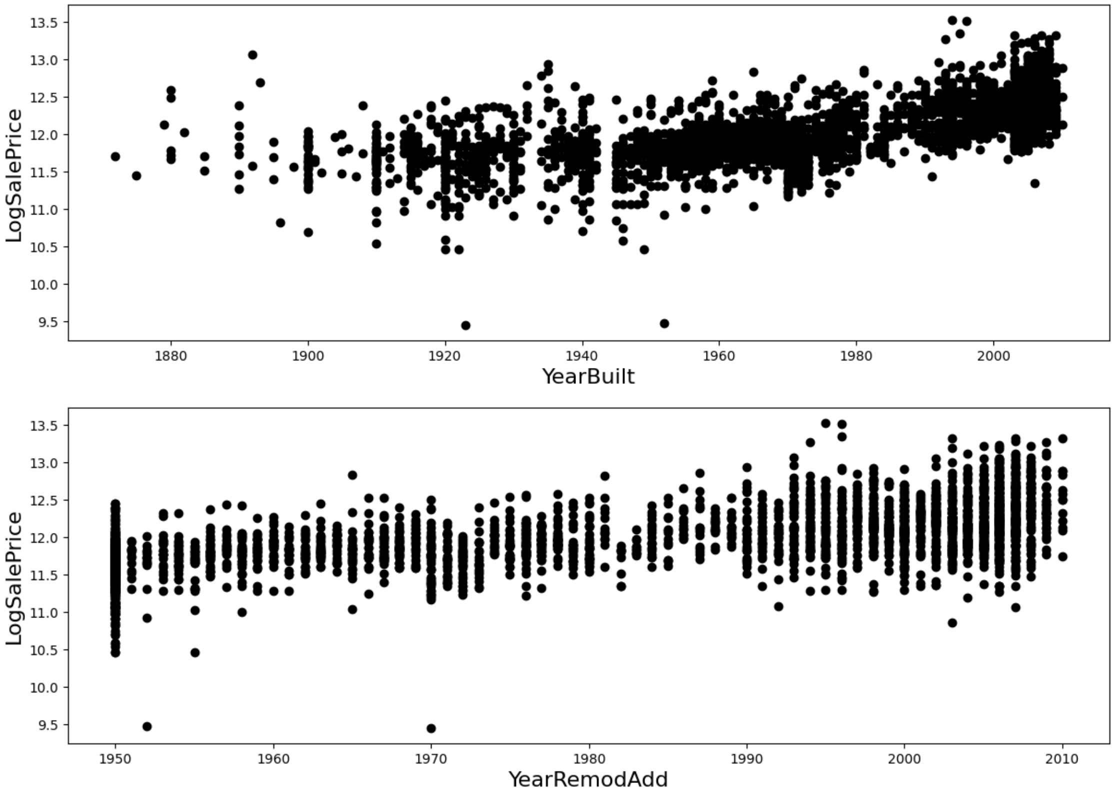
[ ]:
```
print(ames['YearBuilt'].describe(), '\n')
print(ames['YearRemodAdd'].describe())
```

```
count    2930.000000
mean     1971.356314
std        30.245361
min      1872.000000
25%      1954.000000
50%      1973.000000
75%      2001.000000
max      2010.000000
Name: YearBuilt, dtype: float64

count    2930.000000
mean     1984.266553
std        20.860286
min      1950.000000
25%      1965.000000
50%      1993.000000
75%      2004.000000
max      2010.000000
```

```
Name: YearRemodAdd, dtype: float64
```

[ ]:

```
plt.figure(figsize=(14, 10))
plt.subplot(2,1,1)
plt.scatter(ames['YearBuilt'].to_numpy(), ames['LogSalePrice'].to_numpy(),␣
  ↪color='black')
plt.xlabel('YearBuilt', fontsize=16)
plt.ylabel('LogSalePrice', fontsize=16)

plt.subplot(2,1,2)
plt.scatter(ames['YearRemodAdd'].to_numpy(), ames['LogSalePrice'].to_numpy(),␣
  ↪color='black')
plt.xlabel('YearRemodAdd', fontsize=16)
plt.ylabel('LogSalePrice', fontsize=16)
plt.show()
```



Let's add features:

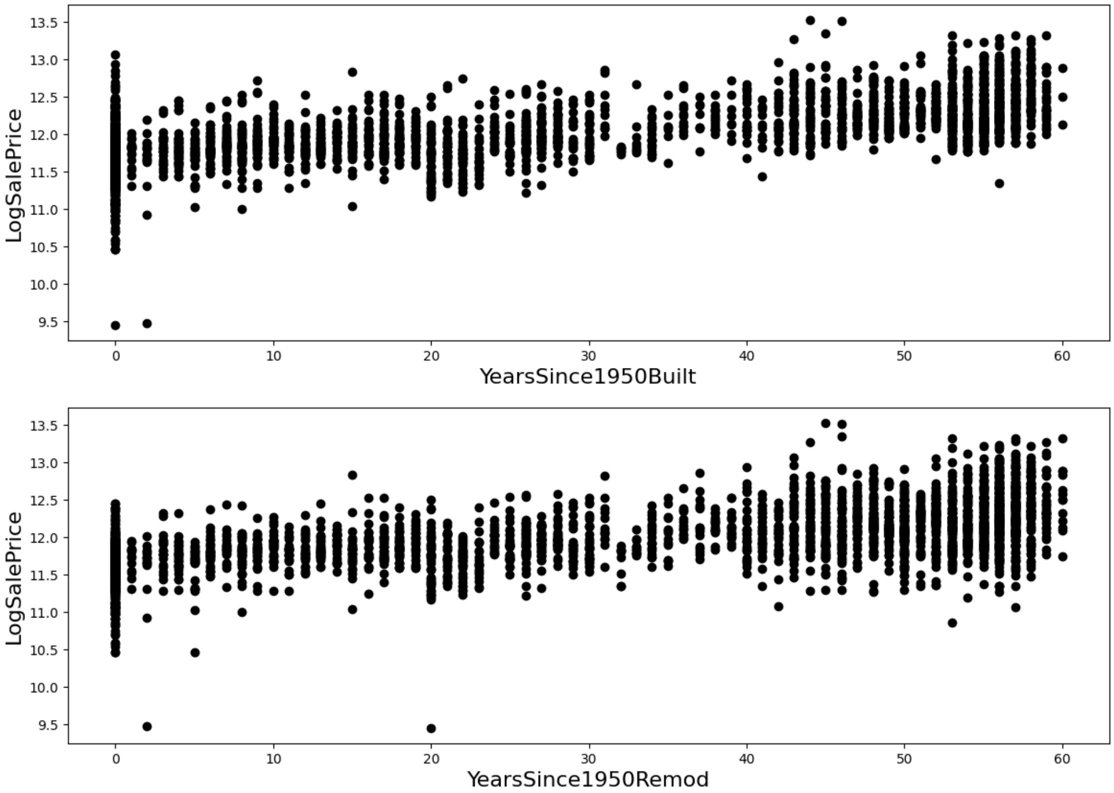
- `YearsSince1950Built` – number of years after 1950 that the home was built (if before 1950, set to 0)
- `YearsSince1950Remod` – number of years after 1950 that the home was remodeled

[ ]:

```
ames['YearsSince1950Built'] = (np.clip(ames['YearBuilt']-1950,0,np.nan)).
 ↪astype('int')
ames['YearsSince1950Remod'] = (np.clip(ames['YearRemodAdd']-1950,0,np.nan)).
 ↪astype('int')

ames.drop(columns=['YearBuilt', 'YearRemodAdd'], inplace=True)
```

[ ]:

```
plt.figure(figsize=(14, 10))
plt.subplot(2,1,1)
plt.scatter(ames['YearsSince1950Built'].to_numpy(), ames['LogSalePrice'].
 ↪to_numpy(), color='black')
plt.xlabel('YearsSince1950Built', fontsize=16)
plt.ylabel('LogSalePrice', fontsize=16)

plt.subplot(2,1,2)
plt.scatter(ames['YearsSince1950Remod'].to_numpy(), ames['LogSalePrice'].
 ↪to_numpy(), color='black')
plt.xlabel('YearsSince1950Remod', fontsize=16)
plt.ylabel('LogSalePrice', fontsize=16)
plt.show()
```



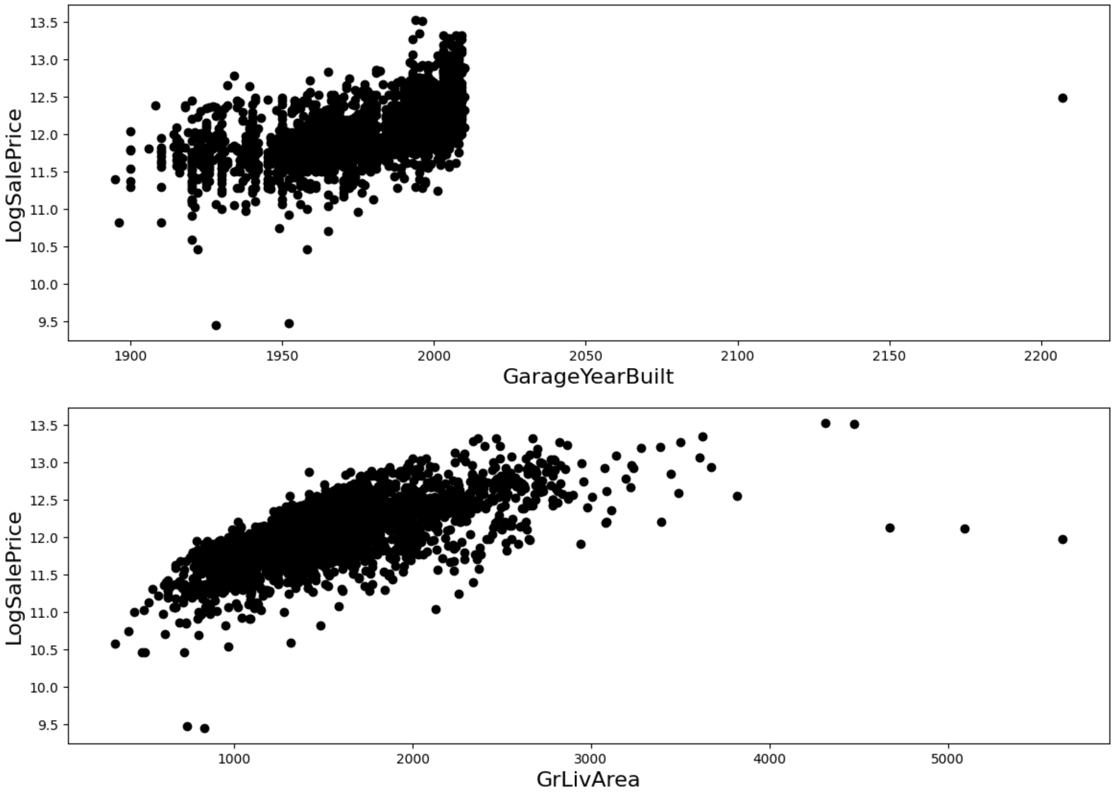
2.0.5 1.4 Remove Outliers and Mistakes

[ ]:

```
plt.figure(figsize=(14, 10))
plt.subplot(2,1,1)
plt.scatter(ames['GarageYrBlt'].to_numpy(), ames['LogSalePrice'].to_numpy(),␣
  ↪color='black')
plt.xlabel('GarageYearBuilt', fontsize=16)
plt.ylabel('LogSalePrice', fontsize=16)

plt.subplot(2,1,2)
plt.scatter(ames['GrLivArea'].to_numpy(), ames['LogSalePrice'].to_numpy(),␣
  ↪color='black')
plt.xlabel('GrLivArea', fontsize=16)
plt.ylabel('LogSalePrice', fontsize=16)
```

[ ]:

```
Text(0, 0.5, 'LogSalePrice')
```



a) Throw away outliers and mistakes:

[ ]:

```
ames
```

```
[ ]:        LogSalePrice  MSSubClass MSZoning  LotFrontage  LotArea Street  \
     0        12.278393          20       RL        141.0  31770.0   Pave
     1        11.561716          20       RH         80.0  11622.0   Pave
     2        12.055250          20       RL         81.0  14267.0   Pave
     3        12.404924          20       RL         93.0  11160.0   Pave
     4        12.154253          60       RL         74.0  13830.0   Pave
     …            …         …      …          …      …    …
     2925     11.867097          80       RL         37.0   7937.0   Pave
     2926     11.782953          20       RL          NaN   8885.0   Pave
     2927     11.790557          85       RL         62.0  10441.0   Pave
     2928     12.043554          20       RL         77.0  10010.0   Pave
     2929     12.144197          60       RL         74.0   9627.0   Pave

             Alley LotShape LandContour Utilities LotConfig LandSlope  \
     0     NoAccess      IR1         Lvl    AllPub    Corner       Gtl
     1     NoAccess      Reg         Lvl    AllPub    Inside       Gtl
     2     NoAccess      IR1         Lvl    AllPub    Corner       Gtl
     3     NoAccess      Reg         Lvl    AllPub    Corner       Gtl
     4     NoAccess      IR1         Lvl    AllPub    Inside       Gtl
     …        …      …         …       …       …       …
     2925  NoAccess      IR1         Lvl    AllPub   CulDSac       Gtl
     2926  NoAccess      IR1         Low    AllPub    Inside       Mod
     2927  NoAccess      Reg         Lvl    AllPub    Inside       Gtl
     2928  NoAccess      Reg         Lvl    AllPub    Inside       Mod
     2929  NoAccess      Reg         Lvl    AllPub    Inside       Mod

          Neighborhood BldgType HouseStyle  OverallQual  OverallCond RoofStyle  \
     0           NAmes     1Fam     1Story            6            5       Hip
     1           NAmes     1Fam     1Story            5            6     Gable
     2           NAmes     1Fam     1Story            6            6       Hip
     3           NAmes     1Fam     1Story            7            5       Hip
     4         Gilbert     1Fam     2Story            5            5     Gable
     …           …      …        …          …          …       …
     2925      Mitchel     1Fam       SLvl            6            6     Gable
     2926      Mitchel     1Fam     1Story            5            5     Gable
     2927      Mitchel     1Fam     SFoyer            5            5     Gable
     2928      Mitchel     1Fam     1Story            5            5     Gable
     2929      Mitchel     1Fam     2Story            7            5     Gable

          RoofMatl MasVnrType  MasVnrArea ExterQual ExterCond Foundation BsmtQual  \
     0     CompShg      Stone       112.0        TA        TA     CBlock       TA
     1     CompShg        NaN         0.0        TA        TA     CBlock       TA
     2     CompShg    BrkFace       108.0        TA        TA     CBlock       TA
     3     CompShg        NaN         0.0        Gd        TA     CBlock       TA
     4     CompShg        NaN         0.0        TA        TA      PConc       Gd
     …       …        …         …       …       …        …      …
     2925  CompShg        NaN         0.0        TA        TA     CBlock       TA
```

```
2926  CompShg        NaN         0.0        TA        TA     CBlock       Gd
2927  CompShg        NaN         0.0        TA        TA      PConc       Gd
2928  CompShg        NaN         0.0        TA        TA     CBlock       Gd
2929  CompShg    BrkFace        94.0        TA        TA      PConc       Gd

     BsmtCond BsmtExposure BsmtFinType1  BsmtFinSF1 BsmtFinType2  BsmtFinSF2  \
0          Gd           Gd          BLQ       639.0          Unf         0.0
1          TA           No          Rec       468.0          LwQ       144.0
2          TA           No          ALQ       923.0          Unf         0.0
3          TA           No          ALQ      1065.0          Unf         0.0
4          TA           No          GLQ       791.0          Unf         0.0
…       …          …          …         …          …         …
2925       TA           Av          GLQ       819.0          Unf         0.0
2926       TA           Av          BLQ       301.0          ALQ       324.0
2927       TA           Av          GLQ       337.0          Unf         0.0
2928       TA           Av          ALQ      1071.0          LwQ       123.0
2929       TA           Av          LwQ       758.0          Unf         0.0

      BsmtUnfSF  TotalBsmtSF Heating HeatingQC CentralAir Electrical  \
0         441.0       1080.0    GasA        Fa          Y      SBrkr
1         270.0        882.0    GasA        TA          Y      SBrkr
2         406.0       1329.0    GasA        TA          Y      SBrkr
3        1045.0       2110.0    GasA        Ex          Y      SBrkr
4         137.0        928.0    GasA        Gd          Y      SBrkr
…         …          …     …       …        …        …
2925      184.0       1003.0    GasA        TA          Y      SBrkr
2926      239.0        864.0    GasA        TA          Y      SBrkr
2927      575.0        912.0    GasA        TA          Y      SBrkr
2928      195.0       1389.0    GasA        Gd          Y      SBrkr
2929      238.0        996.0    GasA        Ex          Y      SBrkr

      X1stFlrSF  X2ndFlrSF  LowQualFinSF  GrLivArea  BsmtFullBath  \
0        1656.0        0.0           0.0     1656.0           1.0
1         896.0        0.0           0.0      896.0           0.0
2        1329.0        0.0           0.0     1329.0           0.0
3        2110.0        0.0           0.0     2110.0           1.0
4         928.0      701.0           0.0     1629.0           0.0
…         …        …           …        …           …
2925     1003.0        0.0           0.0     1003.0           1.0
2926      902.0        0.0           0.0      902.0           1.0
2927      970.0        0.0           0.0      970.0           0.0
2928     1389.0        0.0           0.0     1389.0           1.0
2929      996.0     1004.0           0.0     2000.0           0.0

      BsmtHalfBath  FullBath  HalfBath  BedroomAbvGr  KitchenAbvGr  \
0              0.0       1.0       0.0           3.0           1.0
1              0.0       1.0       0.0           2.0           1.0
```

```
2              0.0       1.0       1.0           3.0           1.0
3              0.0       2.0       1.0           3.0           1.0
4              0.0       2.0       1.0           3.0           1.0
…            …       …       …           …           …
2925           0.0       1.0       0.0           3.0           1.0
2926           0.0       1.0       0.0           2.0           1.0
2927           1.0       1.0       0.0           3.0           1.0
2928           0.0       1.0       0.0           2.0           1.0
2929           0.0       2.0       1.0           3.0           1.0

     KitchenQual  TotRmsAbvGrd Functional  Fireplaces  FireplaceQu GarageType  \
0             TA           7.0        Typ         2.0          Gd     Attchd
1             TA           5.0        Typ         0.0  NoFirePlace     Attchd
2             Gd           6.0        Typ         0.0  NoFirePlace     Attchd
3             Ex           8.0        Typ         2.0          TA     Attchd
4             TA           6.0        Typ         1.0          TA     Attchd
…          …           …        …         …          …        …
2925          TA           6.0        Typ         0.0  NoFirePlace     Detchd
2926          TA           5.0        Typ         0.0  NoFirePlace     Attchd
2927          TA           6.0        Typ         0.0  NoFirePlace   NoGarage
2928          TA           6.0        Typ         1.0          TA     Attchd
2929          TA           9.0        Typ         1.0          TA     Attchd

      GarageYrBlt GarageFinish  GarageCars  GarageArea GarageQual GarageCond  \
0          1960.0          Fin         2.0       528.0         TA         TA
1          1961.0          Unf         1.0       730.0         TA         TA
2          1958.0          Unf         1.0       312.0         TA         TA
3          1968.0          Fin         2.0       522.0         TA         TA
4          1997.0          Fin         2.0       482.0         TA         TA
…           …          …         …         …        …        …
2925       1984.0          Unf         2.0       588.0         TA         TA
2926       1983.0          Unf         2.0       484.0         TA         TA
2927          NaN     NoGarage         0.0         0.0   NoGarage   NoGarage
2928       1975.0          RFn         2.0       418.0         TA         TA
2929       1993.0          Fin         3.0       650.0         TA         TA

     PavedDrive  WoodDeckSF  OpenPorchSF  EnclosedPorch  X3SsnPorch  \
0             P       210.0         62.0            0.0         0.0
1             Y       140.0          0.0            0.0         0.0
2             Y       393.0         36.0            0.0         0.0
3             Y         0.0          0.0            0.0         0.0
4             Y       212.0         34.0            0.0         0.0
…         …         …           …            …         …
2925          Y       120.0          0.0            0.0         0.0
2926          Y       164.0          0.0            0.0         0.0
2927          Y        80.0         32.0            0.0         0.0
2928          Y       240.0         38.0            0.0         0.0
```

```
2929          Y       190.0         48.0            0.0         0.0

      ScreenPorch  PoolArea  PoolQC    Fence MiscFeature  MiscVal  MoSold  \
0             0.0       0.0  NoPool  NoFence         NaN      0.0     5.0
1           120.0       0.0  NoPool    MnPrv         NaN      0.0     6.0
2             0.0       0.0  NoPool  NoFence        Gar2  12500.0     6.0
3             0.0       0.0  NoPool  NoFence         NaN      0.0     4.0
4             0.0       0.0  NoPool    MnPrv         NaN      0.0     3.0
…           …       …     …      …         …      …     …
2925          0.0       0.0  NoPool    GdPrv         NaN      0.0     3.0
2926          0.0       0.0  NoPool    MnPrv         NaN      0.0     6.0
2927          0.0       0.0  NoPool    MnPrv        Shed    700.0     7.0
2928          0.0       0.0  NoPool  NoFence         NaN      0.0     4.0
2929          0.0       0.0  NoPool  NoFence         NaN      0.0    11.0

      YrSold SaleType SaleCondition  Artery  Feedr  Norm  PosA  PosN  RRAe  \
0     2010.0       WD         Normal       0      0     1     0     0     0
1     2010.0       WD         Normal       0      1     1     0     0     0
2     2010.0       WD         Normal       0      0     1     0     0     0
3     2010.0       WD         Normal       0      0     1     0     0     0
4     2010.0       WD         Normal       0      0     1     0     0     0
…      …      …           …     …    …   …   …   …   …
2925  2006.0       WD         Normal       0      0     1     0     0     0
2926  2006.0       WD         Normal       0      0     1     0     0     0
2927  2006.0       WD         Normal       0      0     1     0     0     0
2928  2006.0       WD         Normal       0      0     1     0     0     0
2929  2006.0       WD         Normal       0      0     1     0     0     0

      RRAn  RRNe  RRNn  AsbShng  AsphShn  BrkCmn  BrkComm  BrkFace  CBlock  \
0        0     0     0        0        0       0        0        1       0
1        0     0     0        0        0       0        0        0       0
2        0     0     0        0        0       0        0        0       0
3        0     0     0        0        0       0        0        1       0
4        0     0     0        0        0       0        0        0       0
…    …   …   …      …      …     …      …      …     …
2925     0     0     0        0        0       0        0        0       0
2926     0     0     0        0        0       0        0        0       0
2927     0     0     0        0        0       0        0        0       0
2928     0     0     0        0        0       0        0        0       0
2929     0     0     0        0        0       0        0        0       0

      CemntBd  CmentBd  HdBoard  ImStucc  MetalSd  Other  Plywood  PreCast  \
0           0        0        0        0        0      0        1        0
1           0        0        0        0        0      0        0        0
2           0        0        0        0        0      0        0        0
3           0        0        0        0        0      0        0        0
4           0        0        0        0        0      0        0        0
```

```
…       …      …      …       …      …    …      …      …
2925         0        0        1        0        0      0        0        0
2926         0        0        1        0        0      0        0        0
2927         0        0        1        0        0      0        0        0
2928         0        0        1        0        0      0        0        0
2929         0        0        1        0        0      0        0        0

      Stone  Stucco  VinylSd  WdSdng  WdShing  WdShng  YearsSince1950Built  \
0         0       0        0       0        0       0                   10
1         0       0        1       0        0       0                   11
2         0       0        0       1        0       0                    8
3         0       0        0       0        0       0                   18
4         0       0        1       0        0       0                   47
…     …      …      …     …      …     …                  …
2925      0       0        0       0        0       0                   34
2926      0       0        0       0        0       0                   33
2927      0       0        0       0        0       1                   42
2928      0       0        0       0        0       0                   24
2929      0       0        0       0        0       0                   43

      YearsSince1950Remod
0                      10
1                      11
2                       8
3                      18
4                      48
…                    …
2925                   34
2926                   33
2927                   42
2928                   25
2929                   44

[2930 rows x 105 columns]
```

[ ]:

```
ames.loc[~(ames['GrLivArea']>4000)] # drop 5 obs
```

[ ]:

```
      LogSalePrice  MSSubClass MSZoning  LotFrontage  LotArea Street  \
0        12.278393          20       RL        141.0  31770.0   Pave
1        11.561716          20       RH         80.0  11622.0   Pave
2        12.055250          20       RL         81.0  14267.0   Pave
3        12.404924          20       RL         93.0  11160.0   Pave
4        12.154253          60       RL         74.0  13830.0   Pave
…            …         …      …          …      …    …
2925     11.867097          80       RL         37.0   7937.0   Pave
2926     11.782953          20       RL          NaN   8885.0   Pave
2927     11.790557          85       RL         62.0  10441.0   Pave
```

```
2928     12.043554          20       RL         77.0  10010.0   Pave
2929     12.144197          60       RL         74.0   9627.0   Pave

         Alley LotShape LandContour Utilities LotConfig LandSlope  \
0     NoAccess      IR1         Lvl    AllPub    Corner       Gtl
1     NoAccess      Reg         Lvl    AllPub    Inside       Gtl
2     NoAccess      IR1         Lvl    AllPub    Corner       Gtl
3     NoAccess      Reg         Lvl    AllPub    Corner       Gtl
4     NoAccess      IR1         Lvl    AllPub    Inside       Gtl
…        …      …         …       …       …       …
2925  NoAccess      IR1         Lvl    AllPub   CulDSac       Gtl
2926  NoAccess      IR1         Low    AllPub    Inside       Mod
2927  NoAccess      Reg         Lvl    AllPub    Inside       Gtl
2928  NoAccess      Reg         Lvl    AllPub    Inside       Mod
2929  NoAccess      Reg         Lvl    AllPub    Inside       Mod

     Neighborhood BldgType HouseStyle  OverallQual  OverallCond RoofStyle   \
0           NAmes     1Fam     1Story            6            5       Hip
1           NAmes     1Fam     1Story            5            6     Gable
2           NAmes     1Fam     1Story            6            6       Hip
3           NAmes     1Fam     1Story            7            5       Hip
4         Gilbert     1Fam     2Story            5            5     Gable
…           …      …        …          …          …       …
2925      Mitchel     1Fam       SLvl            6            6     Gable
2926      Mitchel     1Fam     1Story            5            5     Gable
2927      Mitchel     1Fam     SFoyer            5            5     Gable
2928      Mitchel     1Fam     1Story            5            5     Gable
2929      Mitchel     1Fam     2Story            7            5     Gable

     RoofMatl MasVnrType  MasVnrArea ExterQual ExterCond Foundation BsmtQual  \
0     CompShg      Stone       112.0        TA        TA     CBlock       TA
1     CompShg        NaN         0.0        TA        TA     CBlock       TA
2     CompShg    BrkFace       108.0        TA        TA     CBlock       TA
3     CompShg        NaN         0.0        Gd        TA     CBlock       TA
4     CompShg        NaN         0.0        TA        TA      PConc       Gd
…       …        …         …       …       …        …      …
2925  CompShg        NaN         0.0        TA        TA     CBlock       TA
2926  CompShg        NaN         0.0        TA        TA     CBlock       Gd
2927  CompShg        NaN         0.0        TA        TA      PConc       Gd
2928  CompShg        NaN         0.0        TA        TA     CBlock       Gd
2929  CompShg    BrkFace        94.0        TA        TA      PConc       Gd

     BsmtCond BsmtExposure BsmtFinType1  BsmtFinSF1 BsmtFinType2  BsmtFinSF2  \
0          Gd           Gd          BLQ       639.0          Unf         0.0
1          TA           No          Rec       468.0          LwQ       144.0
2          TA           No          ALQ       923.0          Unf         0.0
3          TA           No          ALQ      1065.0          Unf         0.0
```

```
4          TA           No          GLQ       791.0          Unf          0.0
…       …          …          …         …          …         …
2925       TA           Av          GLQ       819.0          Unf          0.0
2926       TA           Av          BLQ       301.0          ALQ        324.0
2927       TA           Av          GLQ       337.0          Unf          0.0
2928       TA           Av          ALQ      1071.0          LwQ        123.0
2929       TA           Av          LwQ       758.0          Unf          0.0

      BsmtUnfSF  TotalBsmtSF Heating HeatingQC CentralAir Electrical  \
0         441.0       1080.0    GasA        Fa          Y      SBrkr
1         270.0        882.0    GasA        TA          Y      SBrkr
2         406.0       1329.0    GasA        TA          Y      SBrkr
3        1045.0       2110.0    GasA        Ex          Y      SBrkr
4         137.0        928.0    GasA        Gd          Y      SBrkr
…         …          …     …       …        …        …
2925      184.0       1003.0    GasA        TA          Y      SBrkr
2926      239.0        864.0    GasA        TA          Y      SBrkr
2927      575.0        912.0    GasA        TA          Y      SBrkr
2928      195.0       1389.0    GasA        Gd          Y      SBrkr
2929      238.0        996.0    GasA        Ex          Y      SBrkr

      X1stFlrSF  X2ndFlrSF  LowQualFinSF  GrLivArea  BsmtFullBath  \
0        1656.0        0.0           0.0     1656.0           1.0
1         896.0        0.0           0.0      896.0           0.0
2        1329.0        0.0           0.0     1329.0           0.0
3        2110.0        0.0           0.0     2110.0           1.0
4         928.0      701.0           0.0     1629.0           0.0
…         …        …           …        …           …
2925     1003.0        0.0           0.0     1003.0           1.0
2926      902.0        0.0           0.0      902.0           1.0
2927      970.0        0.0           0.0      970.0           0.0
2928     1389.0        0.0           0.0     1389.0           1.0
2929      996.0     1004.0           0.0     2000.0           0.0

      BsmtHalfBath  FullBath  HalfBath  BedroomAbvGr  KitchenAbvGr  \
0              0.0       1.0       0.0           3.0           1.0
1              0.0       1.0       0.0           2.0           1.0
2              0.0       1.0       1.0           3.0           1.0
3              0.0       2.0       1.0           3.0           1.0
4              0.0       2.0       1.0           3.0           1.0
…            …       …       …           …           …
2925           0.0       1.0       0.0           3.0           1.0
2926           0.0       1.0       0.0           2.0           1.0
2927           1.0       1.0       0.0           3.0           1.0
2928           0.0       1.0       0.0           2.0           1.0
2929           0.0       2.0       1.0           3.0           1.0
```

```
     KitchenQual  TotRmsAbvGrd Functional  Fireplaces  FireplaceQu GarageType  \
0             TA           7.0        Typ         2.0          Gd     Attchd
1             TA           5.0        Typ         0.0  NoFirePlace     Attchd
2             Gd           6.0        Typ         0.0  NoFirePlace     Attchd
3             Ex           8.0        Typ         2.0          TA     Attchd
4             TA           6.0        Typ         1.0          TA     Attchd
…          …           …        …         …          …        …
2925          TA           6.0        Typ         0.0  NoFirePlace     Detchd
2926          TA           5.0        Typ         0.0  NoFirePlace     Attchd
2927          TA           6.0        Typ         0.0  NoFirePlace   NoGarage
2928          TA           6.0        Typ         1.0          TA     Attchd
2929          TA           9.0        Typ         1.0          TA     Attchd

      GarageYrBlt GarageFinish  GarageCars  GarageArea GarageQual GarageCond  \
0          1960.0          Fin         2.0       528.0         TA         TA
1          1961.0          Unf         1.0       730.0         TA         TA
2          1958.0          Unf         1.0       312.0         TA         TA
3          1968.0          Fin         2.0       522.0         TA         TA
4          1997.0          Fin         2.0       482.0         TA         TA
…           …          …         …         …        …        …
2925       1984.0          Unf         2.0       588.0         TA         TA
2926       1983.0          Unf         2.0       484.0         TA         TA
2927          NaN     NoGarage         0.0         0.0   NoGarage   NoGarage
2928       1975.0          RFn         2.0       418.0         TA         TA
2929       1993.0          Fin         3.0       650.0         TA         TA

     PavedDrive  WoodDeckSF  OpenPorchSF  EnclosedPorch  X3SsnPorch  \
0             P       210.0         62.0            0.0         0.0
1             Y       140.0          0.0            0.0         0.0
2             Y       393.0         36.0            0.0         0.0
3             Y         0.0          0.0            0.0         0.0
4             Y       212.0         34.0            0.0         0.0
…         …          …          …            …         …
2925          Y       120.0          0.0            0.0         0.0
2926          Y       164.0          0.0            0.0         0.0
2927          Y        80.0         32.0            0.0         0.0
2928          Y       240.0         38.0            0.0         0.0
2929          Y       190.0         48.0            0.0         0.0

      ScreenPorch  PoolArea  PoolQC    Fence MiscFeature  MiscVal  MoSold  \
0             0.0       0.0  NoPool  NoFence         NaN      0.0     5.0
1           120.0       0.0  NoPool    MnPrv         NaN      0.0     6.0
2             0.0       0.0  NoPool  NoFence        Gar2  12500.0     6.0
3             0.0       0.0  NoPool  NoFence         NaN      0.0     4.0
4             0.0       0.0  NoPool    MnPrv         NaN      0.0     3.0
…           …        …     …      …         …      …     …
2925          0.0       0.0  NoPool    GdPrv         NaN      0.0     3.0
```

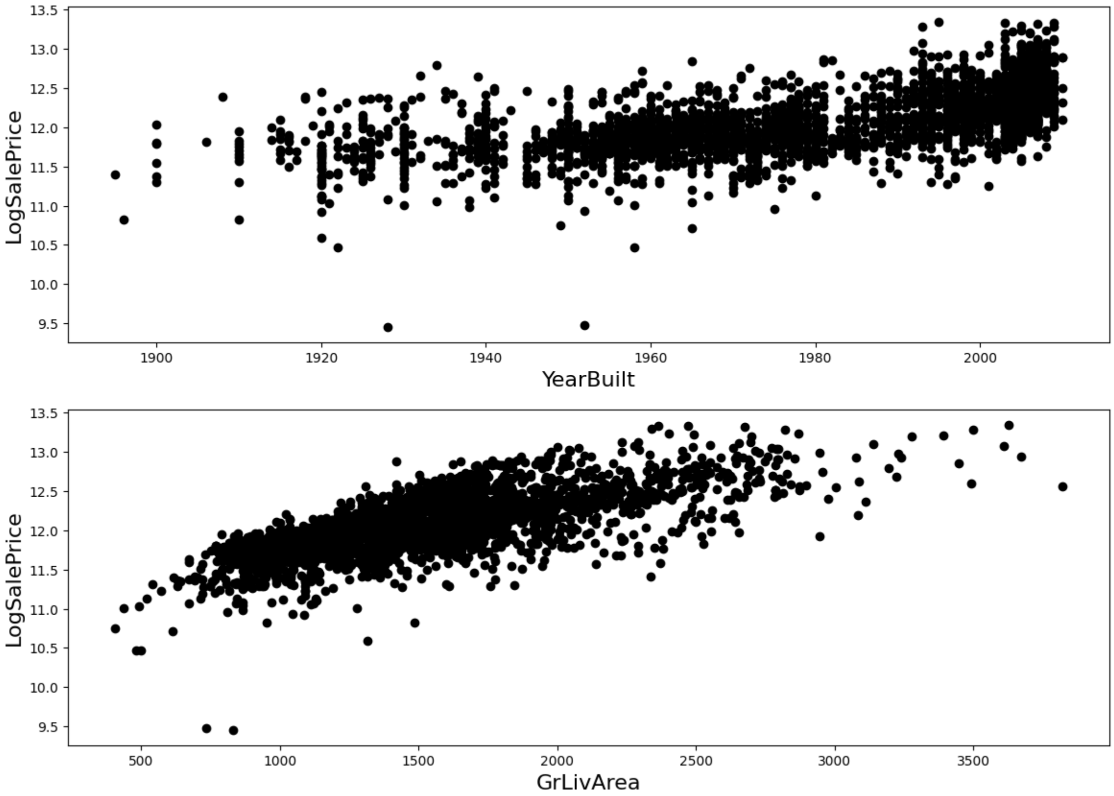
```
2926          0.0       0.0  NoPool    MnPrv         NaN      0.0     6.0
2927          0.0       0.0  NoPool    MnPrv        Shed    700.0     7.0
2928          0.0       0.0  NoPool  NoFence         NaN      0.0     4.0
2929          0.0       0.0  NoPool  NoFence         NaN      0.0    11.0

      YrSold SaleType SaleCondition  Artery  Feedr  Norm  PosA  PosN  RRAe  \
0     2010.0       WD         Normal       0      0     1     0     0     0
1     2010.0       WD         Normal       0      1     1     0     0     0
2     2010.0       WD         Normal       0      0     1     0     0     0
3     2010.0       WD         Normal       0      0     1     0     0     0
4     2010.0       WD         Normal       0      0     1     0     0     0
…      …      …           …     …    …   …   …   …   …
2925  2006.0       WD         Normal       0      0     1     0     0     0
2926  2006.0       WD         Normal       0      0     1     0     0     0
2927  2006.0       WD         Normal       0      0     1     0     0     0
2928  2006.0       WD         Normal       0      0     1     0     0     0
2929  2006.0       WD         Normal       0      0     1     0     0     0

      RRAn  RRNe  RRNn  AsbShng  AsphShn  BrkCmn  BrkComm  BrkFace  CBlock  \
0        0     0     0        0        0       0        0        1       0
1        0     0     0        0        0       0        0        0       0
2        0     0     0        0        0       0        0        0       0
3        0     0     0        0        0       0        0        1       0
4        0     0     0        0        0       0        0        0       0
…    …   …   …      …      …     …      …      …     …
2925     0     0     0        0        0       0        0        0       0
2926     0     0     0        0        0       0        0        0       0
2927     0     0     0        0        0       0        0        0       0
2928     0     0     0        0        0       0        0        0       0
2929     0     0     0        0        0       0        0        0       0

      CemntBd  CmentBd  HdBoard  ImStucc  MetalSd  Other  Plywood  PreCast  \
0           0        0        0        0        0      0        1        0
1           0        0        0        0        0      0        0        0
2           0        0        0        0        0      0        0        0
3           0        0        0        0        0      0        0        0
4           0        0        0        0        0      0        0        0
…       …      …      …      …      …    …      …      …
2925        0        0        1        0        0      0        0        0
2926        0        0        1        0        0      0        0        0
2927        0        0        1        0        0      0        0        0
2928        0        0        1        0        0      0        0        0
2929        0        0        1        0        0      0        0        0

      Stone  Stucco  VinylSd  WdSdng  WdShing  WdShng  YearsSince1950Built  \
0         0       0        0       0        0       0                   10
1         0       0        1       0        0       0                   11
```

```
2       0     0      0     1      0     0              8
3       0     0      0     0      0     0             18
4       0     0      1     0      0     0             47
…     …    …      …     …     …     …          …
2925    0     0      0     0      0     0             34
2926    0     0      0     0      0     0             33
2927    0     0      0     0      0     1             42
2928    0     0      0     0      0     0             24
2929    0     0      0     0      0     0             43

      YearsSince1950Remod
0                      10
1                      11
2                       8
3                      18
4                      48
…                     …
2925                   34
2926                   33
2927                   42
2928                   25
2929                   44

[2925 rows x 105 columns]
```

```
[ ]: ames = ames.loc[(ames['GarageYrBlt']<=2015)] # drop 1 obs
     ames = ames.loc[~(ames['GrLivArea']>4000)] # drop 4 additional obs
```

```
[ ]: plt.figure(figsize=(14, 10))
     plt.subplot(2,1,1)
     plt.scatter(ames['GarageYrBlt'].to_numpy(), ames['LogSalePrice'].to_numpy(),␣
      ↪color='black')
     plt.xlabel('YearBuilt', fontsize=16)
     plt.ylabel('LogSalePrice', fontsize=16)

     plt.subplot(2,1,2)
     plt.scatter(ames['GrLivArea'].to_numpy(), ames['LogSalePrice'].to_numpy(),␣
      ↪color='black')
     plt.xlabel('GrLivArea', fontsize=16)
     plt.ylabel('LogSalePrice', fontsize=16)
```

```
[ ]: Text(0, 0.5, 'LogSalePrice')
```



b) Let’s apply the same 1950 transformation to `GarageYrBlt`.

[ ]:
```
ames['YearsSince1950GarageBuilt'] = (np.clip(ames['GarageYrBlt']-1950,0,np.nan))
ames.drop(columns=['GarageYrBlt'], inplace=True)
```

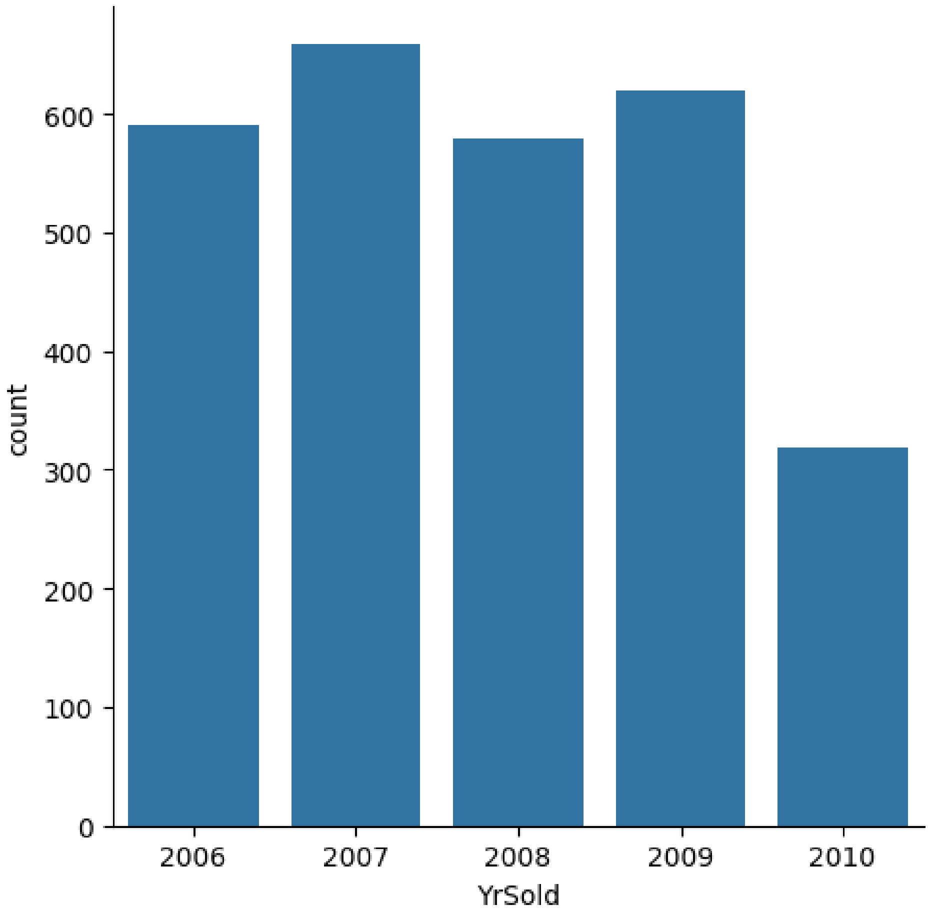
2.0.6 1.5 Numerical variables to categorical variables (Time Sold and Quality Score)

1.5.1 Time Sold: Investigate further the time when the house was sold:

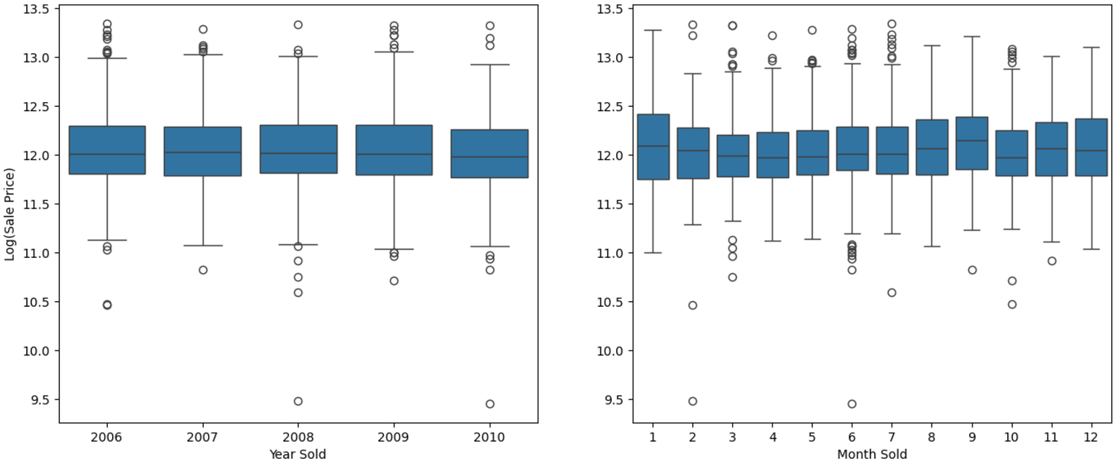
[ ]:
```
ames[['YrSold','MoSold']]=ames[['YrSold','MoSold']].astype('int')
```

[ ]:
```
plt.figure(figsize=(6,4))
sns.catplot(x="YrSold", kind="count",data=ames);
```

```
<Figure size 600x400 with 0 Axes>
```



[ ]:
```
fig, axs=plt.subplots(1,2, figsize=(15,6))
sns.boxplot(x="YrSold", y="LogSalePrice", data=ames, ax= axs[0])
sns.boxplot(x="MoSold", y="LogSalePrice", data=ames, ax= axs[1])
axs[0].set_xlabel('Year Sold')
axs[1].set_xlabel('Month Sold')
axs[0].set_ylabel('Log(Sale Price)')
axs[1].set_ylabel('')
plt.show()
```

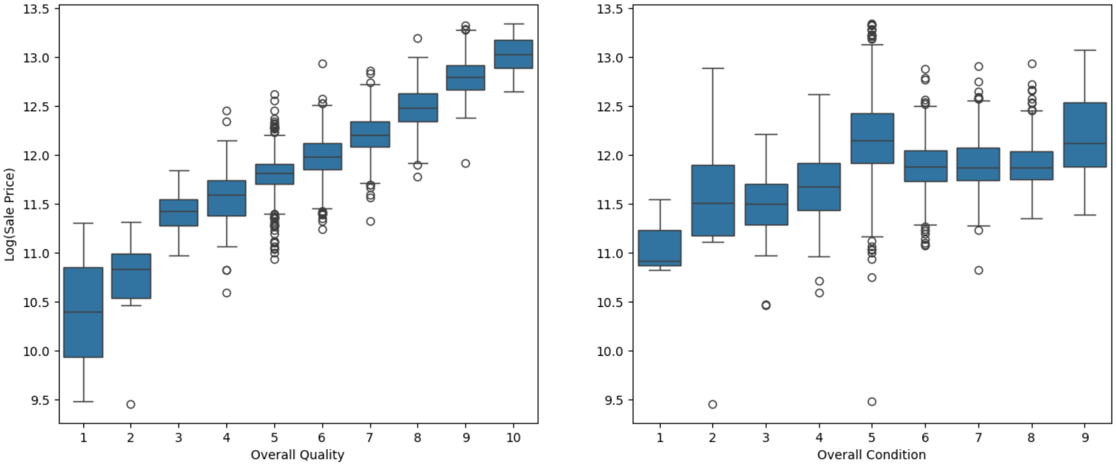


Convert `YrSold` and `MoSold` to categorical variables:

```
[ ]: ames[['YrSold','MoSold']]=ames[['YrSold','MoSold']].astype('category')
```

1.5.2 Quality Variables Let's look at the overall quality/condition variables.

```
[ ]: fig, axs=plt.subplots(1,2, figsize=(15,6))
     sns.boxplot(x="OverallQual", y="LogSalePrice", data=ames, ax= axs[0])
     sns.boxplot(x="OverallCond", y="LogSalePrice", data=ames, ax= axs[1])
     axs[0].set_xlabel('Overall Quality')
     axs[1].set_xlabel('Overall Condition')
     axs[0].set_ylabel('Log(Sale Price)')
     axs[1].set_ylabel('')
     plt.show()
```



Let’s encode them to categorical:

```
[ ]: ames[['OverallQual','OverallCond']]=ames[['OverallQual','OverallCond']].
       ↪astype('category')
```

2.0.7 1.6 Handling NaN Values

```
[ ]: np.sum(ames.isnull(),axis=0)
```

```
[ ]: LogSalePrice          0
     MSSubClass            0
     MSZoning              0
     LotFrontage         479
     LotArea               0
     Street                0
     Alley                 0
     LotShape              0
     LandContour           0
     Utilities             0
     LotConfig             0
     LandSlope             0
     Neighborhood          0
     BldgType              0
     HouseStyle            0
     OverallQual           0
     OverallCond           0
     RoofStyle             0
     RoofMatl              0
     MasVnrType         1629
     MasVnrArea           21
     ExterQual             0
     ExterCond             0
     Foundation            0
     BsmtQual              1
     BsmtCond              1
     BsmtExposure          4
     BsmtFinType1          1
     BsmtFinSF1            1
     BsmtFinType2          2
     BsmtFinSF2            1
     BsmtUnfSF             1
     TotalBsmtSF           1
     Heating               0
     HeatingQC             0
     CentralAir            0
     Electrical            1
     X1stFlrSF             0
```

```
X2ndFlrSF                       0
LowQualFinSF                    0
GrLivArea                       0
BsmtFullBath                    2
BsmtHalfBath                    2
FullBath                        0
HalfBath                        0
BedroomAbvGr                    0
KitchenAbvGr                    0
KitchenQual                     0
TotRmsAbvGrd                    0
Functional                      0
Fireplaces                      0
FireplaceQu                     0
GarageType                      0
GarageFinish                    0
GarageCars                      0
GarageArea                      0
GarageQual                      0
GarageCond                      0
PavedDrive                      0
WoodDeckSF                      0
OpenPorchSF                     0
EnclosedPorch                   0
X3SsnPorch                      0
ScreenPorch                     0
PoolArea                        0
PoolQC                          0
Fence                           0
MiscFeature                  2668
MiscVal                         0
MoSold                          0
YrSold                          0
SaleType                        0
SaleCondition                   0
Artery                          0
Feedr                           0
Norm                            0
PosA                            0
PosN                            0
RRAe                            0
RRAn                            0
RRNe                            0
RRNn                            0
AsbShng                         0
AsphShn                         0
BrkCmn                          0
```

```
BrkComm                          0
BrkFace                          0
CBlock                           0
CemntBd                          0
CmentBd                          0
HdBoard                          0
ImStucc                          0
MetalSd                          0
Other                            0
Plywood                          0
PreCast                          0
Stone                            0
Stucco                           0
VinylSd                          0
WdSdng                           0
WdShing                          0
WdShng                           0
YearsSince1950Built              0
YearsSince1950Remod              0
YearsSince1950GarageBuilt        0
dtype: int64
```

3 In-class activity 1: Only display the column names with missing values

[ ]:
```
missing_columns = ames.columns[ames.isnull().any()]
print(missing_columns)
```

```
Index(['LotFrontage', 'MasVnrType', 'MasVnrArea', 'BsmtQual', 'BsmtCond',
       'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1', 'BsmtFinType2',
       'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Electrical', 'BsmtFullBath',
       'BsmtHalfBath', 'MiscFeature'],
      dtype='object')
```

Dealing with NA columns:

- For numerical variables, NAs arise because some factor has already been set to 0 (e.g., there is no basement), so we convert NAs to 0.
- For categorical variables, we add a new level corresponding to whether the variable is missing from that observation. We'll do this via dummy encoding as usual.

[ ]:
```
# numerical ones:
numerical_cols = ['LotFrontage', 'MasVnrArea','BsmtFinSF1',␣
 ↪'BsmtFinSF2','BsmtUnfSF','TotalBsmtSF','BsmtFullBath',
                  'BsmtHalfBath', 'GarageCars', 'GarageArea',␣
 ↪'GrLivArea','YearsSince1950GarageBuilt']
for var in numerical_cols:
```

```
    print(var, ames[var].dtype)
    ames.loc[np.isnan(ames[var]), [var]] = 0
```

```
LotFrontage float64
MasVnrArea float64
BsmtFinSF1 float64
BsmtFinSF2 float64
BsmtUnfSF float64
TotalBsmtSF float64
BsmtFullBath float64
BsmtHalfBath float64
GarageCars float64
GarageArea float64
GrLivArea float64
YearsSince1950GarageBuilt float64
```

[ ]:
```
ames['MasVnrType'].dtype
```

[ ]:
```
dtype('O')
```

[ ]:
```
# categorical ones:
categorical_cols = ['MasVnrType', 'BsmtQual', 'BsmtCond', 'BsmtExposure',␣
 ↪'BsmtFinType1',
                    'BsmtFinType2', 'Electrical', 'GarageFinish', 'GarageQual',␣
 ↪'GarageCond', 'MiscFeature']

for var in categorical_cols:
    print(var, ames[var].dtype)
    ames.loc[(ames[var]).isna(), [var]] = 'NaN'

ames[categorical_cols]=ames[categorical_cols].astype('category')
```

```
MasVnrType object
BsmtQual object
BsmtCond object
BsmtExposure object
BsmtFinType1 object
BsmtFinType2 object
Electrical object
GarageFinish object
GarageQual object
GarageCond object
MiscFeature object
```

[ ]:
```
# check again:
print(np.sum(ames.isnull().any()))
```

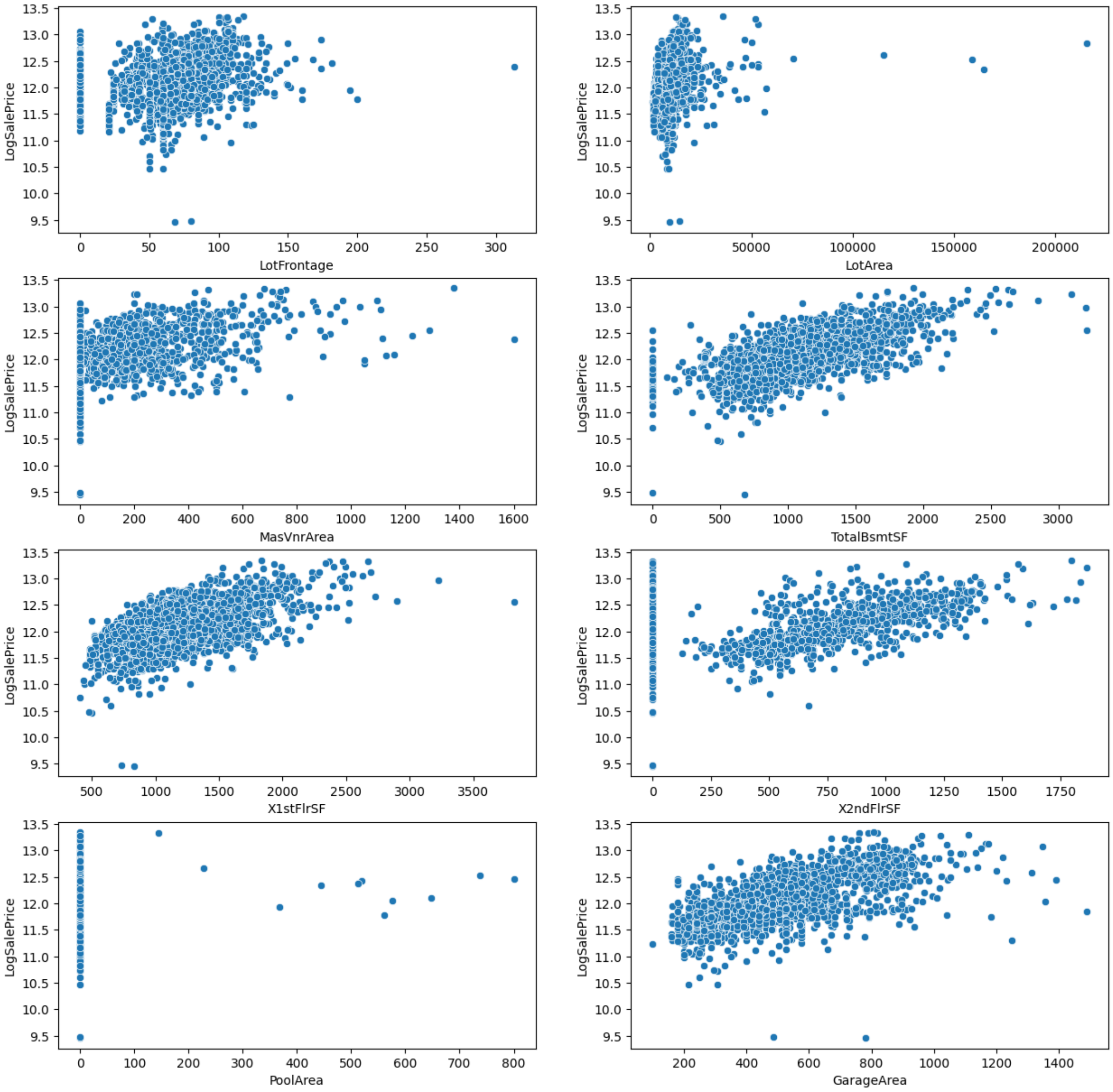
```
0
```

3.0.1 1.7 Some More EDA

3.0.2 1.7.1 Continuous variables

```
[ ]: # a) Scattered plots
fig, axs=plt.subplots(4,2, figsize=(15,15))
sns.scatterplot(data=ames, x="LotFrontage", y="LogSalePrice", ax= axs[0,0])
sns.scatterplot(data=ames, x="LotArea", y="LogSalePrice", ax= axs[0,1])
sns.scatterplot(data=ames, x="MasVnrArea", y="LogSalePrice", ax= axs[1,0])
sns.scatterplot(data=ames, x="TotalBsmtSF", y="LogSalePrice", ax= axs[1,1])
sns.scatterplot(data=ames, x="X1stFlrSF", y="LogSalePrice", ax= axs[2,0])
sns.scatterplot(data=ames, x="X2ndFlrSF", y="LogSalePrice", ax= axs[2,1])
sns.scatterplot(data=ames, x="PoolArea", y="LogSalePrice", ax= axs[3,0])
sns.scatterplot(data=ames, x="GarageArea", y="LogSalePrice", ax= axs[3,1])
```

```
[ ]: <Axes: xlabel='GarageArea', ylabel='LogSalePrice'>
```

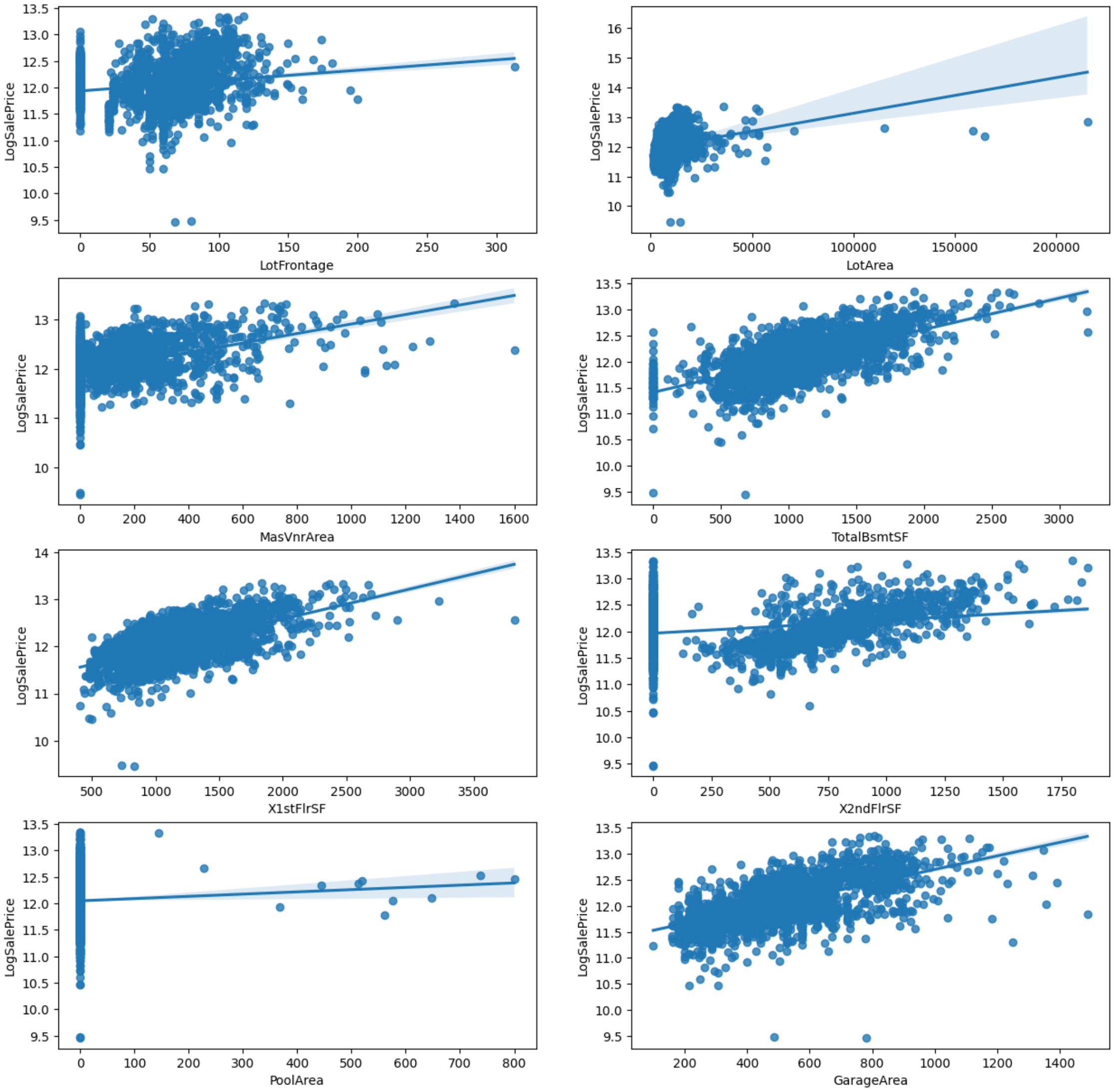


[ ]:

```python
# b) Smoothed versions of those plots (with a trend line).
fig, axs=plt.subplots(4,2, figsize=(15,15))
sns.regplot(data=ames, x="LotFrontage", y="LogSalePrice", ax= axs[0,0])
sns.regplot(data=ames, x="LotArea", y="LogSalePrice", ax= axs[0,1])
sns.regplot(data=ames, x="MasVnrArea", y="LogSalePrice", ax= axs[1,0])
sns.regplot(data=ames, x="TotalBsmtSF", y="LogSalePrice", ax= axs[1,1])
sns.regplot(data=ames, x="X1stFlrSF", y="LogSalePrice", ax= axs[2,0])
sns.regplot(data=ames, x="X2ndFlrSF", y="LogSalePrice", ax= axs[2,1])
sns.regplot(data=ames, x="PoolArea", y="LogSalePrice", ax= axs[3,0])
sns.regplot(data=ames, x="GarageArea", y="LogSalePrice", ax= axs[3,1])
```

[ ]:

```
<Axes: xlabel='GarageArea', ylabel='LogSalePrice'>
```



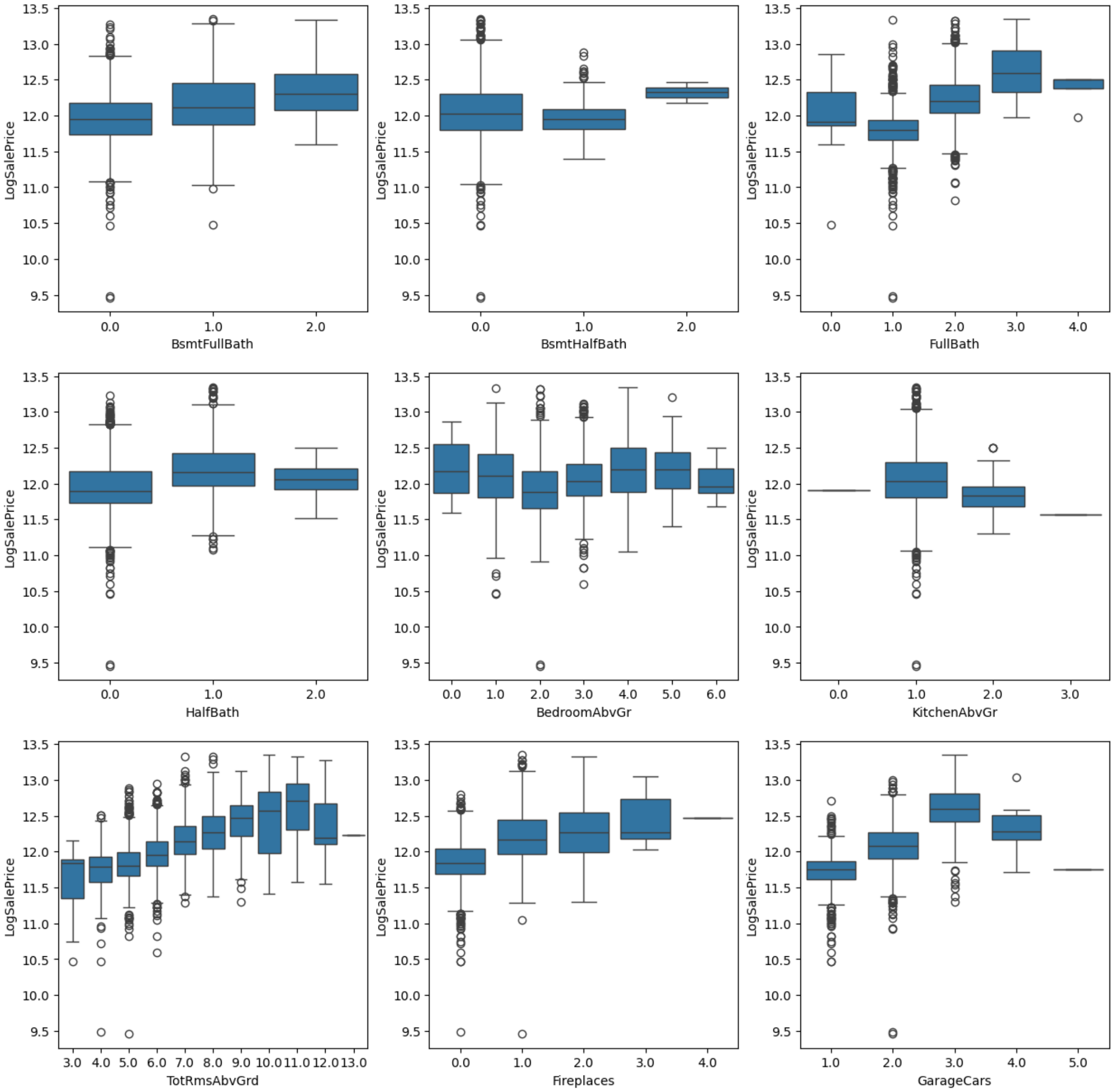
Note: It appears that there are nonlinear relationships. Later, we will use polynomials to enhance our models.

3.0.3 1.7.2 Discrete Variables

Let's look at discrete variables now.

```
[ ]: fig, axs=plt.subplots(3,3, figsize=(15,15))
     sns.boxplot(data=ames, x="BsmtFullBath", y="LogSalePrice", ax= axs[0,0])
     sns.boxplot(data=ames, x="BsmtHalfBath", y="LogSalePrice", ax= axs[0,1])
     sns.boxplot(data=ames, x="FullBath", y="LogSalePrice", ax= axs[0,2])
     sns.boxplot(data=ames, x="HalfBath", y="LogSalePrice", ax= axs[1,0])
     sns.boxplot(data=ames, x="BedroomAbvGr", y="LogSalePrice", ax= axs[1,1])
     sns.boxplot(data=ames, x="KitchenAbvGr", y="LogSalePrice", ax= axs[1,2])
     sns.boxplot(data=ames, x="TotRmsAbvGrd", y="LogSalePrice", ax= axs[2,0])
     sns.boxplot(data=ames, x="Fireplaces", y="LogSalePrice", ax= axs[2,1])
     sns.boxplot(data=ames, x="GarageCars", y="LogSalePrice", ax= axs[2,2])
```

```
[ ]: <Axes: xlabel='GarageCars', ylabel='LogSalePrice'>
```



Let's convert those all to factors.

```
[ ]: discrete_cols = ['BsmtFullBath', 'BsmtHalfBath', 'FullBath', 'HalfBath',␣
      ↪'BedroomAbvGr',
                      'KitchenAbvGr', 'TotRmsAbvGrd', 'Fireplaces', 'GarageCars',␣
      ↪'MSSubClass']
     ames[discrete_cols] = ames[discrete_cols].astype('int').astype('category')
```

3.0.4 1.7.3 Other Categorical Variables

We still have many columns that are categorical. We must encode them before passing them to our models

```
[ ]: still_categorical = ames.columns[ames.dtypes == 'object']
     for col in still_categorical:
         print(col, ':\n', ames[col].unique())
```

```
MSZoning :
 ['RL' 'RH' 'FV' 'RM' 'C' 'I' 'A']
Street :
 ['Pave' 'Grvl']
Alley :
 ['NoAccess' 'Pave' 'Grvl']
LotShape :
 ['IR1' 'Reg' 'IR2' 'IR3']
LandContour :
 ['Lvl' 'HLS' 'Bnk' 'Low']
Utilities :
 ['AllPub' 'NoSewr' 'NoSeWa']
LotConfig :
 ['Corner' 'Inside' 'CulDSac' 'FR2' 'FR3']
LandSlope :
 ['Gtl' 'Mod' 'Sev']
Neighborhood :
 ['NAmes' 'Gilbert' 'StoneBr' 'NWAmes' 'Somerst' 'BrDale' 'NPkVill'
 'NridgHt' 'Blmngtn' 'NoRidge' 'SawyerW' 'Sawyer' 'Greens' 'OldTown'
 'BrkSide' 'IDOTRR' 'ClearCr' 'SWISU' 'Edwards' 'CollgCr' 'Crawfor'
 'Blueste' 'Mitchel' 'Timber' 'MeadowV' 'Veenker' 'GrnHill' 'Landmrk']
BldgType :
 ['1Fam' 'TwnhsE' 'Twnhs' 'Duplex' '2fmCon']
HouseStyle :
 ['1Story' '2Story' '1.5Fin' 'SFoyer' 'SLvl' '2.5Unf' '1.5Unf' '2.5Fin']
RoofStyle :
 ['Hip' 'Gable' 'Mansard' 'Gambrel' 'Shed' 'Flat']
RoofMatl :
 ['CompShg' 'WdShake' 'WdShngl' 'Tar&Grv' 'Membran' 'Roll' 'Metal']
ExterQual :
 ['TA' 'Gd' 'Ex' 'Fa']
ExterCond :
 ['TA' 'Gd' 'Fa' 'Po' 'Ex']
Foundation :
 ['CBlock' 'PConc' 'Wood' 'BrkTil' 'Slab' 'Stone']
Heating :
 ['GasA' 'GasW' 'Grav' 'Wall' 'Floor' 'OthW']
HeatingQC :
 ['Fa' 'TA' 'Ex' 'Gd' 'Po']
CentralAir :
 ['Y' 'N']
KitchenQual :
 ['TA' 'Gd' 'Ex' 'Fa' 'Po']
Functional :
```

```
 ['Typ' 'Mod' 'Min1' 'Min2' 'Maj1' 'Maj2' 'Sal' 'Sev']
FireplaceQu :
 ['Gd' 'NoFirePlace' 'TA' 'Po' 'Ex' 'Fa']
GarageType :
 ['Attchd' 'BuiltIn' 'Basment' 'Detchd' 'CarPort' '2Types']
PavedDrive :
 ['P' 'Y' 'N']
PoolQC :
 ['NoPool' 'Ex' 'Gd' 'TA' 'Fa']
Fence :
 ['NoFence' 'MnPrv' 'GdPrv' 'GdWo' 'MnWw']
SaleType :
 ['WD ' 'New' 'COD' 'Con' 'ConLD' 'Oth' 'ConLw' 'ConLI' 'CWD' 'VWD']
SaleCondition :
 ['Normal' 'Partial' 'Family' 'Abnorml' 'Alloca' 'AdjLand']
```

[ ]:

```
ames[still_categorical] = ames[still_categorical].astype('category')
```

[ ]:

```
ames.head()
```

[ ]:

```
   LogSalePrice MSSubClass MSZoning  LotFrontage  LotArea Street     Alley  \
0     12.278393         20       RL        141.0  31770.0   Pave  NoAccess
1     11.561716         20       RH         80.0  11622.0   Pave  NoAccess
2     12.055250         20       RL         81.0  14267.0   Pave  NoAccess
3     12.404924         20       RL         93.0  11160.0   Pave  NoAccess
4     12.154253         60       RL         74.0  13830.0   Pave  NoAccess

  LotShape LandContour Utilities LotConfig LandSlope Neighborhood BldgType  \
0      IR1         Lvl    AllPub    Corner       Gtl        NAmes     1Fam
1      Reg         Lvl    AllPub    Inside       Gtl        NAmes     1Fam
2      IR1         Lvl    AllPub    Corner       Gtl        NAmes     1Fam
3      Reg         Lvl    AllPub    Corner       Gtl        NAmes     1Fam
4      IR1         Lvl    AllPub    Inside       Gtl      Gilbert     1Fam

  HouseStyle OverallQual OverallCond RoofStyle RoofMatl MasVnrType  \
0     1Story           6           5       Hip  CompShg      Stone
1     1Story           5           6     Gable  CompShg        NaN
2     1Story           6           6       Hip  CompShg    BrkFace
3     1Story           7           5       Hip  CompShg        NaN
4     2Story           5           5     Gable  CompShg        NaN

   MasVnrArea ExterQual ExterCond Foundation BsmtQual BsmtCond BsmtExposure  \
0       112.0        TA        TA     CBlock       TA       Gd           Gd
1         0.0        TA        TA     CBlock       TA       TA           No
2       108.0        TA        TA     CBlock       TA       TA           No
3         0.0        Gd        TA     CBlock       TA       TA           No
4         0.0        TA        TA      PConc       Gd       TA           No
```

```
  BsmtFinType1  BsmtFinSF1 BsmtFinType2  BsmtFinSF2  BsmtUnfSF  TotalBsmtSF  \
0          BLQ       639.0          Unf         0.0      441.0       1080.0
1          Rec       468.0          LwQ       144.0      270.0        882.0
2          ALQ       923.0          Unf         0.0      406.0       1329.0
3          ALQ      1065.0          Unf         0.0     1045.0       2110.0
4          GLQ       791.0          Unf         0.0      137.0        928.0

  Heating HeatingQC CentralAir Electrical  X1stFlrSF  X2ndFlrSF  LowQualFinSF  \
0    GasA        Fa          Y      SBrkr     1656.0        0.0          0.0
1    GasA        TA          Y      SBrkr      896.0        0.0          0.0
2    GasA        TA          Y      SBrkr     1329.0        0.0          0.0
3    GasA        Ex          Y      SBrkr     2110.0        0.0          0.0
4    GasA        Gd          Y      SBrkr      928.0      701.0          0.0

   GrLivArea BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr  \
0     1656.0            1            0        1        0            3
1      896.0            0            0        1        0            2
2     1329.0            0            0        1        1            3
3     2110.0            1            0        2        1            3
4     1629.0            0            0        2        1            3

  KitchenAbvGr KitchenQual TotRmsAbvGrd Functional Fireplaces  FireplaceQu  \
0            1          TA            7        Typ          2           Gd
1            1          TA            5        Typ          0  NoFirePlace
2            1          Gd            6        Typ          0  NoFirePlace
3            1          Ex            8        Typ          2           TA
4            1          TA            6        Typ          1           TA

  GarageType GarageFinish GarageCars  GarageArea GarageQual GarageCond  \
0     Attchd          Fin          2       528.0         TA         TA
1     Attchd          Unf          1       730.0         TA         TA
2     Attchd          Unf          1       312.0         TA         TA
3     Attchd          Fin          2       522.0         TA         TA
4     Attchd          Fin          2       482.0         TA         TA

  PavedDrive  WoodDeckSF  OpenPorchSF  EnclosedPorch  X3SsnPorch  ScreenPorch  \
0          P       210.0         62.0            0.0         0.0          0.0
1          Y       140.0          0.0            0.0         0.0        120.0
2          Y       393.0         36.0            0.0         0.0          0.0
3          Y         0.0          0.0            0.0         0.0          0.0
4          Y       212.0         34.0            0.0         0.0          0.0

   PoolArea  PoolQC    Fence MiscFeature  MiscVal MoSold YrSold SaleType  \
0        0.0  NoPool  NoFence          NaN      0.0      5   2010       WD
1        0.0  NoPool    MnPrv          NaN      0.0      6   2010       WD
2        0.0  NoPool  NoFence         Gar2  12500.0      6   2010       WD
```

```
3        0.0  NoPool  NoFence          NaN      0.0      4   2010      WD
4        0.0  NoPool    MnPrv          NaN      0.0      3   2010      WD

  SaleCondition  Artery  Feedr  Norm  PosA  PosN  RRAe  RRAn  RRNe  RRNn  \
0        Normal       0      0     1     0     0     0     0     0     0
1        Normal       0      1     1     0     0     0     0     0     0
2        Normal       0      0     1     0     0     0     0     0     0
3        Normal       0      0     1     0     0     0     0     0     0
4        Normal       0      0     1     0     0     0     0     0     0

   AsbShng  AsphShn  BrkCmn  BrkComm  BrkFace  CBlock  CemntBd  CmentBd  \
0        0        0       0        0        1       0        0        0
1        0        0       0        0        0       0        0        0
2        0        0       0        0        0       0        0        0
3        0        0       0        0        1       0        0        0
4        0        0       0        0        0       0        0        0

   HdBoard  ImStucc  MetalSd  Other  Plywood  PreCast  Stone  Stucco  VinylSd  \
0        0        0        0      0        1        0      0       0        0
1        0        0        0      0        0        0      0       0        1
2        0        0        0      0        0        0      0       0        0
3        0        0        0      0        0        0      0       0        0
4        0        0        0      0        0        0      0       0        1

   WdSdng  WdShing  WdShng  YearsSince1950Built  YearsSince1950Remod  \
0       0        0       0                   10                   10
1       0        0       0                   11                   11
2       1        0       0                    8                    8
3       0        0       0                   18                   18
4       0        0       0                   47                   48

   YearsSince1950GarageBuilt
0                       10.0
1                       11.0
2                        8.0
3                       18.0
4                       47.0
```

[ ]:

```

```

3.1 2. Regression Analysis and Regularization Methods

3.1.1 Helper Functions

[ ]:

```
def OSR2(y_train, y_test, y_pred):

    SSE = np.sum((y_test - y_pred)**2)
```

```
    SST = np.sum((y_test - np.mean(y_train))**2)

    return (1 - SSE/SST)
```

```
[ ]: def MAE(y_test, y_pred):

    return (np.mean(abs(y_test - y_pred)))
```

```
[ ]: def RMSE(y_test, y_pred):

    return np.sqrt(np.mean((y_test - y_pred)**2))
```

```
[ ]: def print_metrics(model, X_train, y_train, X_test, y_test,␣
  ↪flag_log_sale_price=False):

    if (flag_log_sale_price == True):

        y_pred_train = pd.Series(model.predict(X_train)).reset_index(drop=True)
        y_pred_test = pd.Series(model.predict(X_test)).reset_index(drop=True)
        y_train = y_train.copy().reset_index(drop=True)
        y_test = y_test.copy().reset_index(drop=True)

        print("\nMetrics for Log(Sale Price):\n")

    elif (flag_log_sale_price == False):

        y_pred_train = pd.Series(model.predict(X_train)).apply(np.exp).
  ↪reset_index(drop=True)
        y_pred_test = pd.Series(model.predict(X_test)).apply(np.exp).
  ↪reset_index(drop=True)
        y_train = y_train.copy().apply(np.exp).reset_index(drop=True)
        y_test = y_test.copy().apply(np.exp).reset_index(drop=True)

        print("\nMetrics for Sale Price:\n")

    print('Training R2', OSR2(y_train, y_train, y_pred_train))
    print('Training MAE', MAE(y_train, y_pred_train))
    print('Training RMSE', RMSE(y_train, y_pred_train))

    print('Out-of-sample R2', OSR2(y_train, y_test, y_pred_test))
    print('Out-of-sample MAE', MAE(y_test, y_pred_test))
    print('Out-of-sample RMSE', RMSE(y_test, y_pred_test))

    return None
```

4 In-class activity 2: Randomly select 30% rows as the testing set and the other 70% as the training set

```
[ ]: test_df = ames.sample(n=int(len(ames) * 0.3), random_state=42)
     train_df = ames.drop(test_df.index)
     test_df.shape
```

```
[ ]: (829, 105)
```

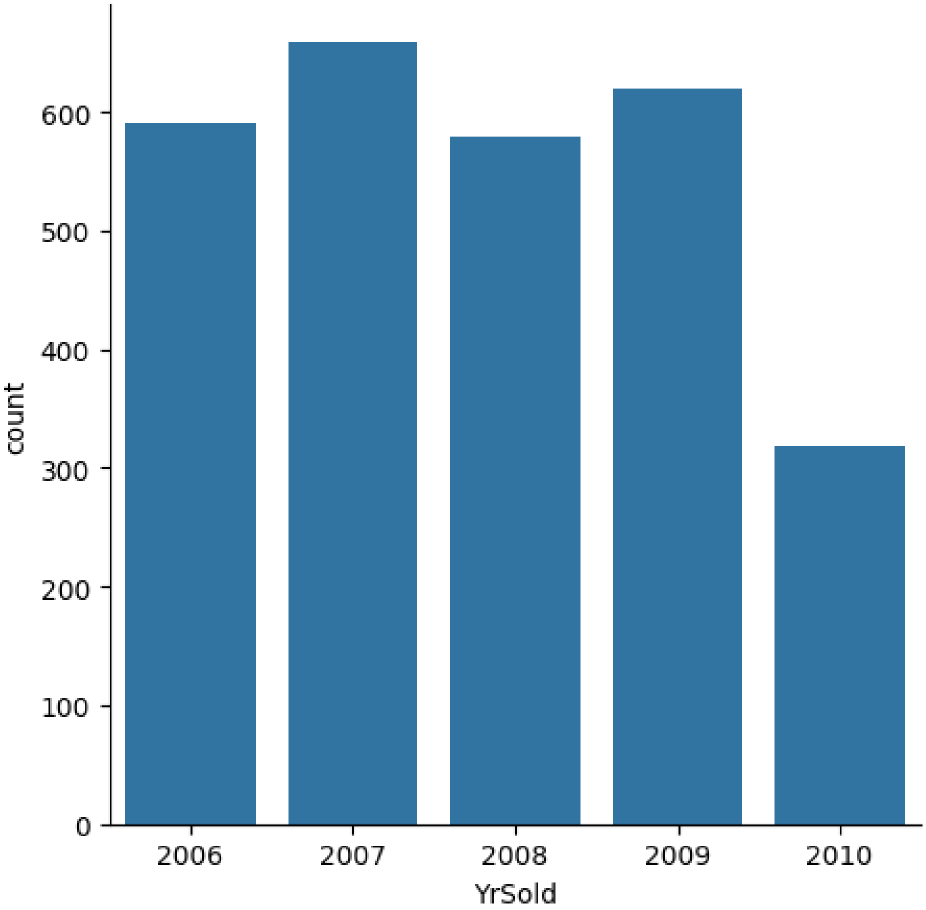
```
[ ]: train_df.shape
```

```
[ ]: (1936, 105)
```

4.0.1 Sequential Split

```
[ ]: sns.catplot(x="YrSold", kind="count",data=ames)
```

```
[ ]: <seaborn.axisgrid.FacetGrid at 0x16a7be240>
```



```
[ ]: ames.loc[ames['YrSold'].isin([2006, 2007, 2008])]
```

```
[ ]:      LogSalePrice MSSubClass MSZoning  LotFrontage  LotArea Street    Alley  \
     989     12.081076         20       RL         87.0  11029.0   Pave  NoAccess
     990     12.377923         20       RL          0.0  12925.0   Pave  NoAccess
     991     12.239991         60       RL         85.0  11075.0   Pave  NoAccess
     992     12.141534         60       RL         72.0   8702.0   Pave  NoAccess
     993     12.013701         60       RL         65.0   8139.0   Pave  NoAccess
     …             …          …        …            …        …      …        …
     2924    11.782953         20       RL        160.0  20000.0   Pave  NoAccess
     2925    11.867097         80       RL         37.0   7937.0   Pave  NoAccess
     2926    11.782953         20       RL          0.0   8885.0   Pave  NoAccess
     2928    12.043554         20       RL         77.0  10010.0   Pave  NoAccess
     2929    12.144197         60       RL         74.0   9627.0   Pave  NoAccess

          LotShape LandContour Utilities LotConfig LandSlope Neighborhood BldgType  \
     989       IR1         Lvl    AllPub    Corner       Gtl        NAmes     1Fam
     990       IR1         Lvl    AllPub    Corner       Gtl        NAmes     1Fam
     991       Reg         Lvl    AllPub    Inside       Gtl        NAmes     1Fam
     992       IR1         Lvl    AllPub    Inside       Gtl      Gilbert     1Fam
     993       Reg         Lvl    AllPub    Inside       Gtl      Gilbert     1Fam
     …         …          …         …         …         …        …
     2924      Reg         Lvl    AllPub    Inside       Gtl      Mitchel     1Fam
     2925      IR1         Lvl    AllPub   CulDSac       Gtl      Mitchel     1Fam
     2926      IR1         Low    AllPub    Inside       Mod      Mitchel     1Fam
     2928      Reg         Lvl    AllPub    Inside       Mod      Mitchel     1Fam
     2929      Reg         Lvl    AllPub    Inside       Mod      Mitchel     1Fam

          HouseStyle OverallQual OverallCond RoofStyle RoofMatl MasVnrType  \
     989     1Story           6           8       Hip  CompShg        NaN
     990     1Story           6           7     Gable  CompShg        NaN
     991     2Story           6           5     Gable  CompShg        NaN
     992     2Story           6           5     Gable  CompShg        NaN
     993     2Story           6           5     Gable  CompShg    BrkFace
     …          …          …           …         …        …         …
     2924    1Story           5           7     Gable  CompShg        NaN
     2925      SLvl           6           6     Gable  CompShg        NaN
     2926    1Story           5           5     Gable  CompShg        NaN
     2928    1Story           5           5     Gable  CompShg        NaN
     2929    2Story           7           5     Gable  CompShg    BrkFace

           MasVnrArea ExterQual ExterCond Foundation BsmtQual BsmtCond  \
     989          0.0        Ex        TA     CBlock       Gd       TA
     990          0.0        TA        TA     CBlock       TA       TA
     991          0.0        TA        TA     CBlock       Fa       TA
     992          0.0        TA        TA      PConc       TA       TA
     993        119.0        TA        TA      PConc       Gd       TA
```

```
…          …       …       …         …      …      …
2924          0.0        TA        TA     CBlock       TA       TA
2925          0.0        TA        TA     CBlock       TA       TA
2926          0.0        TA        TA     CBlock       Gd       TA
2928          0.0        TA        TA     CBlock       Gd       TA
2929         94.0        TA        TA      PConc       Gd       TA

     BsmtExposure BsmtFinType1  BsmtFinSF1 BsmtFinType2  BsmtFinSF2  \
989            No          ALQ       528.0          BLQ       411.0
990            Mn          BLQ       865.0          Unf         0.0
991            Mn          ALQ       500.0          LwQ       276.0
992            No          BLQ       706.0          Unf         0.0
993            No          ALQ       476.0          Unf         0.0
…           …          …         …          …          …
2924           No          ALQ      1224.0          Unf         0.0
2925           Av          GLQ       819.0          Unf         0.0
2926           Av          BLQ       301.0          ALQ       324.0
2928           Av          ALQ      1071.0          LwQ       123.0
2929           Av          LwQ       758.0          Unf         0.0

      BsmtUnfSF  TotalBsmtSF Heating HeatingQC CentralAir Electrical  \
989       245.0       1184.0    GasA        Ex          Y      SBrkr
990       340.0       1205.0    GasA        Ex          Y      SBrkr
991       176.0        952.0    GasA        TA          Y      SBrkr
992       220.0        926.0    GasA        Ex          Y      SBrkr
993       204.0        680.0    GasA        Gd          Y      SBrkr
…          …          …      …       …        …        …
2924        0.0       1224.0    GasA        Ex          Y      SBrkr
2925      184.0       1003.0    GasA        TA          Y      SBrkr
2926      239.0        864.0    GasA        TA          Y      SBrkr
2928      195.0       1389.0    GasA        Gd          Y      SBrkr
2929      238.0        996.0    GasA        Ex          Y      SBrkr

      X1stFlrSF  X2ndFlrSF  LowQualFinSF  GrLivArea BsmtFullBath BsmtHalfBath  \
989      1414.0        0.0           0.0     1414.0            1            0
990      2117.0        0.0           0.0     2117.0            0            0
991      1092.0     1020.0           0.0     2112.0            0            0
992       926.0      678.0           0.0     1604.0            0            0
993       680.0      790.0           0.0     1470.0            0            0
…          …        …           …        …          …          …
2924     1224.0        0.0           0.0     1224.0            1            0
2925     1003.0        0.0           0.0     1003.0            1            0
2926      902.0        0.0           0.0      902.0            1            0
2928     1389.0        0.0           0.0     1389.0            1            0
2929      996.0     1004.0           0.0     2000.0            0            0

     FullBath HalfBath BedroomAbvGr KitchenAbvGr KitchenQual TotRmsAbvGrd  \
```

```
989         1        0            3            1          TA            6
990         2        1            4            1          TA            7
991         2        1            4            1          TA            9
992         2        1            3            1          TA            7
993         2        1            3            1          TA            7
…       …      …          …          …          …          …
2924        1        0            4            1          TA            7
2925        1        0            3            1          TA            6
2926        1        0            2            1          TA            5
2928        1        0            2            1          TA            6
2929        2        1            3            1          TA            9

     Functional Fireplaces  FireplaceQu GarageType GarageFinish GarageCars  \
989        Min1          1           TA     Attchd          Unf          2
990         Typ          2           Gd     Attchd          Fin          2
991         Typ          2           Gd     Attchd          Unf          2
992         Typ          1           TA     Attchd          Fin          2
993         Typ          1           TA    BuiltIn          Fin          2
…         …        …          …        …          …        …
2924        Typ          1           TA     Detchd          Unf          2
2925        Typ          0  NoFirePlace     Detchd          Unf          2
2926        Typ          0  NoFirePlace     Attchd          Unf          2
2928        Typ          1           TA     Attchd          RFn          2
2929        Typ          1           TA     Attchd          Fin          3

      GarageArea GarageQual GarageCond PavedDrive  WoodDeckSF  OpenPorchSF  \
989        601.0         TA         TA          Y         0.0         51.0
990        550.0         TA         TA          Y         0.0         42.0
991        576.0         TA         TA          Y       280.0          0.0
992        470.0         TA         TA          Y         0.0         36.0
993        420.0         TA         TA          Y       192.0         49.0
…          …      …        …        …        …        …
2924       576.0         TA         TA          Y       474.0          0.0
2925       588.0         TA         TA          Y       120.0          0.0
2926       484.0         TA         TA          Y       164.0          0.0
2928       418.0         TA         TA          Y       240.0         38.0
2929       650.0         TA         TA          Y       190.0         48.0

      EnclosedPorch  X3SsnPorch  ScreenPorch  PoolArea  PoolQC    Fence  \
989             0.0         0.0        190.0       0.0  NoPool  NoFence
990             0.0         0.0          0.0       0.0  NoPool  NoFence
991             0.0         0.0          0.0       0.0  NoPool  NoFence
992             0.0         0.0          0.0       0.0  NoPool  NoFence
993             0.0         0.0          0.0       0.0  NoPool  NoFence
…            …         …          …        …       …       …
2924            0.0         0.0          0.0       0.0  NoPool  NoFence
2925            0.0         0.0          0.0       0.0  NoPool    GdPrv
```

```
2926            0.0         0.0          0.0       0.0  NoPool    MnPrv
2928            0.0         0.0          0.0       0.0  NoPool  NoFence
2929            0.0         0.0          0.0       0.0  NoPool  NoFence

     MiscFeature  MiscVal MoSold YrSold SaleType SaleCondition  Artery  Feedr  \
989          NaN      0.0      5   2008       WD        Normal       0      0
990          NaN      0.0      5   2008       WD        Normal       0      0
991          NaN      0.0      6   2008       WD        Normal       0      0
992          NaN      0.0      4   2008       WD        Normal       0      0
993          NaN      0.0     10   2008       WD        Normal       0      0
…          …      …    …    …      …           …     …    …
2924         NaN      0.0      9   2006       WD       Abnorml       0      0
2925         NaN      0.0      3   2006       WD        Normal       0      0
2926         NaN      0.0      6   2006       WD        Normal       0      0
2928         NaN      0.0      4   2006       WD        Normal       0      0
2929         NaN      0.0     11   2006       WD        Normal       0      0

      Norm  PosA  PosN  RRAe  RRAn  RRNe  RRNn  AsbShng  AsphShn  BrkCmn  \
989      1     0     0     0     0     0     0        0        0       0
990      1     0     0     0     0     0     0        0        0       0
991      1     0     0     0     0     0     0        0        0       0
992      1     0     0     0     0     0     0        0        0       0
993      1     0     0     0     0     0     0        0        0       0
…    …   …   …   …   …   …   …      …      …     …
2924     1     0     0     0     0     0     0        0        0       0
2925     1     0     0     0     0     0     0        0        0       0
2926     1     0     0     0     0     0     0        0        0       0
2928     1     0     0     0     0     0     0        0        0       0
2929     1     0     0     0     0     0     0        0        0       0

      BrkComm  BrkFace  CBlock  CemntBd  CmentBd  HdBoard  ImStucc  MetalSd  \
989         0        0       0        0        0        0        0        1
990         0        1       0        0        0        0        0        0
991         0        0       0        0        0        1        0        0
992         0        0       0        0        0        0        0        0
993         0        0       0        0        0        0        0        0
…       …      …     …      …      …      …      …      …
2924        0        0       0        0        0        0        0        0
2925        0        0       0        0        0        1        0        0
2926        0        0       0        0        0        1        0        0
2928        0        0       0        0        0        1        0        0
2929        0        0       0        0        0        1        0        0

      Other  Plywood  PreCast  Stone  Stucco  VinylSd  WdSdng  WdShing  \
989       0        0        0      0       0        0       0        0
990       0        1        0      0       0        0       0        0
991       0        0        0      0       0        0       0        0
```

```
992       0        0        0      0       0        1       0        0
993       0        0        0      0       0        1       0        0
…     …      …      …    …     …      …     …       …
2924      0        0        0      0       0        1       0        0
2925      0        0        0      0       0        0       0        0
2926      0        0        0      0       0        0       0        0
2928      0        0        0      0       0        0       0        0
2929      0        0        0      0       0        0       0        0

      WdShng  YearsSince1950Built  YearsSince1950Remod  \
989        0                    8                   52
990        0                   20                   20
991        0                   19                   19
992        0                   47                   48
993        0                   45                   46
…      …                  …                  …
2924       0                   10                   46
2925       0                   34                   34
2926       0                   33                   33
2928       0                   24                   25
2929       0                   43                   44

      YearsSince1950GarageBuilt
989                        40.0
990                        20.0
991                        19.0
992                        47.0
993                        45.0
…                         …
2924                       10.0
2925                       34.0
2926                       33.0
2928                       25.0
2929                       43.0

[1828 rows x 105 columns]
```

5 In-class activity 3: If we use a categorical variable (for example, `YrSold`), to split the training set and testing set, can we use this categorical variable as a predictor? Why?

We could not bec

[ ]:

```python
ames_train = ames.loc[ames['YrSold'].isin([2006, 2007, 2008])]
ames_test = ames.loc[ames['YrSold'].isin([2009, 2010])]
```

```python
ames = ames.drop(columns = ['YrSold'])
ames_train = ames_train.drop(columns = ['YrSold'])
ames_test = ames_test.drop(columns = ['YrSold'])

print(ames.shape, ames_train.shape, ames_test.shape)
```

```
(2765, 104) (1828, 104) (937, 104)
```

5.1 2.1 Naive Linear Regression

a) with only original features (without polynomial features)

[ ]:

```python
print(ames_train.shape)
all_columns = "+".join(ames_train.columns.difference(["LogSalePrice"]))
my_formula = "LogSalePrice~" + all_columns +'-1'
print(my_formula)

mod_naive = smf.ols(my_formula, data=ames_train)
nlr = mod_naive.fit()

print(nlr.summary())
```

```
(1828, 104)
LogSalePrice~Alley+Artery+AsbShng+AsphShn+BedroomAbvGr+BldgType+BrkCmn+BrkComm+B
rkFace+BsmtCond+BsmtExposure+BsmtFinSF1+BsmtFinSF2+BsmtFinType1+BsmtFinType2+Bsm
tFullBath+BsmtHalfBath+BsmtQual+BsmtUnfSF+CBlock+CemntBd+CentralAir+CmentBd+Elec
trical+EnclosedPorch+ExterCond+ExterQual+Feedr+Fence+FireplaceQu+Fireplaces+Foun
dation+FullBath+Functional+GarageArea+GarageCars+GarageCond+GarageFinish+GarageQ
ual+GarageType+GrLivArea+HalfBath+HdBoard+Heating+HeatingQC+HouseStyle+ImStucc+K
itchenAbvGr+KitchenQual+LandContour+LandSlope+LotArea+LotConfig+LotFrontage+LotS
hape+LowQualFinSF+MSSubClass+MSZoning+MasVnrArea+MasVnrType+MetalSd+MiscFeature+
MiscVal+MoSold+Neighborhood+Norm+OpenPorchSF+Other+OverallCond+OverallQual+Paved
Drive+Plywood+PoolArea+PoolQC+PosA+PosN+PreCast+RRAe+RRAn+RRNe+RRNn+RoofMatl+Roo
fStyle+SaleCondition+SaleType+ScreenPorch+Stone+Street+Stucco+TotRmsAbvGrd+Total
BsmtSF+Utilities+VinylSd+WdSdng+WdShing+WdShng+WoodDeckSF+X1stFlrSF+X2ndFlrSF+X3
SsnPorch+YearsSince1950Built+YearsSince1950GarageBuilt+YearsSince1950Remod-1
                            OLS Regression Results
==============================================================================
Dep. Variable:           LogSalePrice   R-squared:                       0.960
Model:                            OLS   Adj. R-squared:                  0.952
Method:                 Least Squares   F-statistic:                     121.8
Date:                Fri, 01 Nov 2024   Prob (F-statistic):               0.00
Time:                        18:18:55   Log-Likelihood:                 2091.0
No. Observations:                1828   AIC:                            -3574.
Df Residuals:                    1524   BIC:                            -1899.
Df Model:                         303
Covariance Type:            nonrobust
================================================================================
==============
```

```
                                 coef    std err          t      P>|t|
[0.025      0.975]
---------------------------------------------------------------------------------
--------------
Alley[Grvl]                    4.9744      0.139     35.709      0.000
4.701       5.248
Alley[NoAccess]                4.9924      0.139     36.001      0.000
4.720       5.264
Alley[Pave]                    5.0001      0.140     35.839      0.000
4.726       5.274
BedroomAbvGr[T.1]             -0.0148      0.075     -0.198      0.843
-0.162       0.132
BedroomAbvGr[T.2]             -0.0141      0.075     -0.187      0.851
-0.161       0.133
BedroomAbvGr[T.3]             -0.0177      0.075     -0.235      0.814
-0.166       0.130
BedroomAbvGr[T.4]             -0.0183      0.076     -0.241      0.809
-0.167       0.131
BedroomAbvGr[T.5]             -0.0721      0.078     -0.923      0.356
-0.225       0.081
BedroomAbvGr[T.6]              0.0336      0.089      0.376      0.707
-0.141       0.208
BldgType[T.2fmCon]            -0.0799      0.107     -0.744      0.457
-0.290       0.131
BldgType[T.Duplex]            -0.0301      0.017     -1.810      0.070
-0.063       0.003
BldgType[T.Twnhs]             -0.0176      0.043     -0.412      0.680
-0.101       0.066
BldgType[T.TwnhsE]             0.0228      0.039      0.580      0.562
-0.054       0.100
BsmtCond[T.Fa]                 0.0222      0.055      0.402      0.688
-0.086       0.131
BsmtCond[T.Gd]                 0.0333      0.054      0.611      0.541
-0.074       0.140
BsmtCond[T.NaN]               -0.7664      0.043    -17.786      0.000
-0.851      -0.682
BsmtCond[T.NoBasement]        -0.0102      0.014     -0.746      0.456
-0.037       0.017
BsmtCond[T.Po]                -0.0901      0.090     -0.998      0.318
-0.267       0.087
BsmtCond[T.TA]                 0.0265      0.054      0.494      0.621
-0.079       0.132
BsmtExposure[T.Gd]             0.0226      0.010      2.366      0.018
0.004       0.041
BsmtExposure[T.Mn]            -0.0276      0.010     -2.822      0.005
-0.047      -0.008
BsmtExposure[T.NaN]            0.0241      0.062      0.389      0.697
-0.098       0.146
```

```
BsmtExposure[T.No]            -0.0255      0.007     -3.460      0.001
-0.040      -0.011
BsmtExposure[T.NoBasement]    -0.0102      0.014     -0.746      0.456
-0.037       0.017
BsmtFinType1[T.BLQ]           -0.0109      0.010     -1.144      0.253
-0.030       0.008
BsmtFinType1[T.GLQ]            0.0034      0.008      0.405      0.686
-0.013       0.020
BsmtFinType1[T.LwQ]           -0.0366      0.012     -3.063      0.002
-0.060      -0.013
BsmtFinType1[T.NaN]           -0.7664      0.043    -17.786      0.000
-0.851      -0.682
BsmtFinType1[T.NoBasement]    -0.0102      0.014     -0.746      0.456
-0.037       0.017
BsmtFinType1[T.Rec]           -0.0323      0.009     -3.440      0.001
-0.051      -0.014
BsmtFinType1[T.Unf]           -0.0240      0.010     -2.465      0.014
-0.043      -0.005
BsmtFinType2[T.BLQ]           -0.0292      0.024     -1.240      0.215
-0.075       0.017
BsmtFinType2[T.GLQ]            0.0084      0.027      0.307      0.759
-0.045       0.062
BsmtFinType2[T.LwQ]           -0.0238      0.022     -1.070      0.285
-0.067       0.020
BsmtFinType2[T.NaN]           -0.7664      0.043    -17.786      0.000
-0.851      -0.682
BsmtFinType2[T.NoBasement]    -0.0102      0.014     -0.746      0.456
-0.037       0.017
BsmtFinType2[T.Rec]           -0.0297      0.021     -1.392      0.164
-0.071       0.012
BsmtFinType2[T.Unf]           -0.0142      0.023     -0.623      0.533
-0.059       0.030
BsmtFullBath[T.1]              0.0172      0.007      2.569      0.010
0.004       0.030
BsmtFullBath[T.2]              0.0504      0.031      1.613      0.107
-0.011       0.112
BsmtHalfBath[T.1]              0.0138      0.010      1.363      0.173
-0.006       0.034
BsmtHalfBath[T.2]              0.1116      0.122      0.918      0.359
-0.127       0.350
BsmtQual[T.Fa]                -0.0526      0.020     -2.586      0.010
-0.093      -0.013
BsmtQual[T.Gd]                -0.0187      0.011     -1.736      0.083
-0.040       0.002
BsmtQual[T.NaN]               -0.7664      0.043    -17.786      0.000
-0.851      -0.682
BsmtQual[T.NoBasement]        -0.0102      0.014     -0.746      0.456
-0.037       0.017
```

```
BsmtQual[T.Po]                 0.0953      0.090      1.061      0.289
-0.081       0.271
BsmtQual[T.TA]                -0.0110      0.014     -0.795      0.427
-0.038       0.016
CentralAir[T.Y]                0.0243      0.013      1.810      0.070
-0.002       0.051
Electrical[T.FuseF]           -0.0061      0.022     -0.273      0.785
-0.050       0.038
Electrical[T.FuseP]            0.0010      0.057      0.017      0.987
-0.111       0.113
Electrical[T.Mix]             -0.0901      0.090     -0.998      0.318
-0.267       0.087
Electrical[T.NaN]              0.0457      0.089      0.514      0.608
-0.129       0.220
Electrical[T.SBrkr]           -0.0067      0.010     -0.650      0.516
-0.027       0.014
ExterCond[T.Fa]               -0.0101      0.042     -0.241      0.809
-0.092       0.072
ExterCond[T.Gd]                0.0033      0.036      0.091      0.928
-0.068       0.075
ExterCond[T.Po]                1.3374      0.234      5.710      0.000
0.878       1.797
ExterCond[T.TA]                0.0161      0.036      0.445      0.656
-0.055       0.087
ExterQual[T.Fa]               -0.0161      0.040     -0.406      0.685
-0.094       0.062
ExterQual[T.Gd]               -0.0087      0.017     -0.500      0.617
-0.043       0.026
ExterQual[T.TA]               -0.0180      0.019     -0.933      0.351
-0.056       0.020
Fence[T.GdWo]                  0.0092      0.017      0.554      0.580
-0.023       0.042
Fence[T.MnPrv]                -0.0056      0.013     -0.422      0.673
-0.032       0.020
Fence[T.MnWw]                 -0.0097      0.031     -0.310      0.756
-0.071       0.051
Fence[T.NoFence]               0.0035      0.012      0.290      0.772
-0.020       0.027
FireplaceQu[T.Fa]              0.0037      0.023      0.161      0.872
-0.042       0.049
FireplaceQu[T.Gd]              0.0167      0.019      0.897      0.370
-0.020       0.053
FireplaceQu[T.NoFirePlace]     3.0216      0.081     37.211      0.000
2.862       3.181
FireplaceQu[T.Po]             -0.0041      0.026     -0.159      0.874
-0.054       0.046
FireplaceQu[T.TA]              0.0043      0.019      0.225      0.822
-0.033       0.042
```

```
Fireplaces[T.1]                3.0427      0.081     37.341      0.000
2.883       3.203
Fireplaces[T.2]                3.0773      0.082     37.468      0.000
2.916       3.238
Fireplaces[T.3]                3.0624      0.087     35.371      0.000
2.893       3.232
Fireplaces[T.4]                2.7628      0.186     14.833      0.000
2.397       3.128
Foundation[T.CBlock]           0.0217      0.011      1.940      0.053
-0.000       0.044
Foundation[T.PConc]            0.0283      0.012      2.389      0.017
0.005       0.052
Foundation[T.Slab]             0.0086      0.036      0.239      0.811
-0.062       0.079
Foundation[T.Stone]           -0.0132      0.047     -0.281      0.779
-0.106       0.079
Foundation[T.Wood]             0.0412      0.065      0.631      0.528
-0.087       0.169
FullBath[T.1]                  0.0299      0.077      0.390      0.696
-0.120       0.180
FullBath[T.2]                  0.0518      0.077      0.669      0.504
-0.100       0.204
FullBath[T.3]                  0.0972      0.079      1.228      0.220
-0.058       0.252
FullBath[T.4]              -2.134e-14   2.17e-15     -9.849      0.000
-2.56e-14   -1.71e-14
Functional[T.Maj2]             0.0490      0.067      0.728      0.467
-0.083       0.181
Functional[T.Min1]             0.0595      0.033      1.803      0.072
-0.005       0.124
Functional[T.Min2]             0.0302      0.034      0.895      0.371
-0.036       0.097
Functional[T.Mod]              0.0383      0.037      1.045      0.296
-0.034       0.110
Functional[T.Sal]             -0.3892      0.151     -2.584      0.010
-0.685      -0.094
Functional[T.Sev]             -0.2305      0.110     -2.092      0.037
-0.447      -0.014
Functional[T.Typ]              0.0765      0.029      2.619      0.009
0.019       0.134
GarageCars[T.2]                0.0191      0.009      2.184      0.029
0.002       0.036
GarageCars[T.3]                0.0485      0.016      3.051      0.002
0.017       0.080
GarageCars[T.4]                0.0719      0.039      1.846      0.065
-0.005       0.148
GarageCars[T.5]            -4.547e-15   7.02e-16     -6.478      0.000
-5.92e-15   -3.17e-15
```

```
GarageCond[T.Fa]               0.0644      0.083      0.777      0.437
-0.098       0.227
GarageCond[T.Gd]               0.0810      0.084      0.964      0.335
-0.084       0.246
GarageCond[T.Po]               0.2043      0.092      2.216      0.027
0.023       0.385
GarageCond[T.TA]               0.0797      0.081      0.981      0.327
-0.080       0.239
GarageFinish[T.RFn]          -0.0012      0.006     -0.193      0.847
-0.014       0.011
GarageFinish[T.Unf]          -0.0023      0.008     -0.297      0.767
-0.017       0.013
GarageQual[T.Fa]             -0.1674      0.097     -1.723      0.085
-0.358       0.023
GarageQual[T.Gd]             -0.1324      0.095     -1.391      0.165
-0.319       0.054
GarageQual[T.Po]             -0.1549      0.129     -1.199      0.231
-0.408       0.099
GarageQual[T.TA]             -0.1370      0.096     -1.426      0.154
-0.325       0.051
GarageType[T.Attchd]           0.0362      0.024      1.517      0.130
-0.011       0.083
GarageType[T.Basment]          0.0238      0.033      0.712      0.477
-0.042       0.089
GarageType[T.BuiltIn]          0.0293      0.026      1.125      0.261
-0.022       0.080
GarageType[T.CarPort]         -0.0188      0.040     -0.469      0.639
-0.097       0.060
GarageType[T.Detchd]           0.0412      0.024      1.734      0.083
-0.005       0.088
HalfBath[T.1]                  0.0253      0.007      3.429      0.001
0.011       0.040
HalfBath[T.2]                 -0.0831      0.033     -2.519      0.012
-0.148      -0.018
Heating[T.GasA]                0.1503      0.098      1.528      0.127
-0.043       0.343
Heating[T.GasW]                0.1526      0.101      1.505      0.133
-0.046       0.352
Heating[T.Grav]               -0.0518      0.161     -0.323      0.747
-0.367       0.263
Heating[T.OthW]                0.0894      0.137      0.651      0.515
-0.180       0.358
Heating[T.Wall]                0.2007      0.121      1.665      0.096
-0.036       0.437
HeatingQC[T.Fa]               -0.0535      0.016     -3.360      0.001
-0.085      -0.022
HeatingQC[T.Gd]               -0.0112      0.007     -1.649      0.099
-0.024       0.002
```

```
HeatingQC[T.Po]                -0.1209      0.071     -1.713      0.087
-0.259       0.018
HeatingQC[T.TA]                -0.0324      0.007     -4.686      0.000
-0.046      -0.019
HouseStyle[T.1.5Unf]            0.0142      0.074      0.191      0.849
-0.132       0.160
HouseStyle[T.1Story]            0.0423      0.031      1.361      0.174
-0.019       0.103
HouseStyle[T.2.5Fin]           -0.0686      0.068     -1.012      0.312
-0.202       0.064
HouseStyle[T.2.5Unf]            0.0433      0.047      0.930      0.352
-0.048       0.134
HouseStyle[T.2Story]            0.0134      0.032      0.421      0.674
-0.049       0.076
HouseStyle[T.SFoyer]            0.0520      0.043      1.196      0.232
-0.033       0.137
HouseStyle[T.SLvl]              0.0719      0.047      1.536      0.125
-0.020       0.164
KitchenAbvGr[T.1]              -0.1077      0.092     -1.165      0.244
-0.289       0.074
KitchenAbvGr[T.2]              -0.1499      0.098     -1.536      0.125
-0.341       0.041
KitchenAbvGr[T.3]           -2.938e-15   1.15e-15     -2.562      0.011
-5.19e-15   -6.88e-16
KitchenQual[T.Fa]              -0.0536      0.023     -2.374      0.018
-0.098      -0.009
KitchenQual[T.Gd]              -0.0429      0.012     -3.497      0.000
-0.067      -0.019
KitchenQual[T.Po]            -1.77e-15   1.11e-15     -1.602      0.109
-3.94e-15    3.97e-16
KitchenQual[T.TA]              -0.0546      0.014     -3.984      0.000
-0.081      -0.028
LandContour[T.HLS]              0.0102      0.017      0.600      0.549
-0.023       0.044
LandContour[T.Low]              0.0204      0.024      0.852      0.395
-0.027       0.068
LandContour[T.Lvl]              0.0004      0.013      0.030      0.976
-0.025       0.026
LandSlope[T.Mod]                0.0243      0.014      1.753      0.080
-0.003       0.051
LandSlope[T.Sev]               -0.0545      0.041     -1.345      0.179
-0.134       0.025
LotConfig[T.CulDSac]            0.0148      0.011      1.345      0.179
-0.007       0.036
LotConfig[T.FR2]               -0.0202      0.014     -1.406      0.160
-0.048       0.008
LotConfig[T.FR3]               -0.0075      0.028     -0.264      0.792
-0.063       0.048
```

```
LotConfig[T.Inside]            -0.0019      0.006     -0.311      0.756
-0.014       0.010
LotShape[T.IR2]                 0.0056      0.013      0.418      0.676
-0.021       0.032
LotShape[T.IR3]                 0.0151      0.028      0.541      0.588
-0.040       0.070
LotShape[T.Reg]                -0.0023      0.005     -0.413      0.679
-0.013       0.008
MSSubClass[T.30]               -0.0601      0.017     -3.623      0.000
-0.093      -0.028
MSSubClass[T.40]               -0.0257      0.049     -0.521      0.603
-0.123       0.071
MSSubClass[T.45]                0.0233      0.077      0.301      0.764
-0.128       0.175
MSSubClass[T.50]                0.0297      0.032      0.941      0.347
-0.032       0.092
MSSubClass[T.60]               -0.0211      0.032     -0.671      0.502
-0.083       0.041
MSSubClass[T.70]                0.0259      0.034      0.771      0.441
-0.040       0.092
MSSubClass[T.75]                0.0056      0.050      0.113      0.910
-0.092       0.103
MSSubClass[T.80]               -0.0516      0.044     -1.185      0.236
-0.137       0.034
MSSubClass[T.85]                0.0059      0.039      0.152      0.879
-0.070       0.082
MSSubClass[T.90]               -0.0301      0.017     -1.810      0.070
-0.063       0.003
MSSubClass[T.120]              -0.0634      0.040     -1.599      0.110
-0.141       0.014
MSSubClass[T.150]              -0.1852      0.113     -1.637      0.102
-0.407       0.037
MSSubClass[T.160]              -0.1437      0.053     -2.732      0.006
-0.247      -0.041
MSSubClass[T.180]              -0.0974      0.062     -1.559      0.119
-0.220       0.025
MSSubClass[T.190]               0.0401      0.105      0.383      0.702
-0.165       0.246
MSZoning[T.C]                   1.4675      0.059     24.738      0.000
1.351        1.584
MSZoning[T.FV]                  1.7801      0.051     34.822      0.000
1.680        1.880
MSZoning[T.I]                   1.5664      0.143     10.990      0.000
1.287        1.846
MSZoning[T.RH]                  1.7555      0.053     32.836      0.000
1.651        1.860
MSZoning[T.RL]                  1.7681      0.048     36.754      0.000
1.674        1.862
```

```
MSZoning[T.RM]               1.7220      0.049     35.389      0.000
1.627       1.817
MasVnrType[T.BrkFace]        0.0338      0.023      1.468      0.142
-0.011       0.079
MasVnrType[T.CBlock]        -0.2792      0.119     -2.354      0.019
-0.512      -0.047
MasVnrType[T.NaN]            0.0329      0.023      1.416      0.157
-0.013       0.078
MasVnrType[T.Stone]          0.0311      0.024      1.278      0.202
-0.017       0.079
MiscFeature[T.NaN]          -0.0020      0.079     -0.025      0.980
-0.156       0.152
MiscFeature[T.Othr]          0.0943      0.087      1.087      0.277
-0.076       0.265
MiscFeature[T.Shed]         -0.0210      0.075     -0.280      0.779
-0.168       0.126
MiscFeature[T.TenC]         -0.3765      0.160     -2.349      0.019
-0.691      -0.062
MoSold[T.2]                 -0.0380      0.016     -2.444      0.015
-0.069      -0.008
MoSold[T.3]                 -0.0227      0.014     -1.610      0.108
-0.050       0.005
MoSold[T.4]                 -0.0055      0.014     -0.397      0.691
-0.033       0.022
MoSold[T.5]                  0.0027      0.013      0.208      0.835
-0.023       0.028
MoSold[T.6]                 -0.0031      0.013     -0.245      0.806
-0.028       0.022
MoSold[T.7]                  0.0042      0.013      0.332      0.740
-0.020       0.029
MoSold[T.8]                 -0.0159      0.013     -1.188      0.235
-0.042       0.010
MoSold[T.9]                  0.0052      0.015      0.359      0.720
-0.023       0.034
MoSold[T.10]                -0.0245      0.014     -1.710      0.088
-0.053       0.004
MoSold[T.11]                -0.0103      0.015     -0.709      0.479
-0.039       0.018
MoSold[T.12]                -0.0091      0.016     -0.586      0.558
-0.040       0.021
Neighborhood[T.Blueste]      0.1140      0.055      2.073      0.038
0.006       0.222
Neighborhood[T.BrDale]      -0.0058      0.039     -0.149      0.882
-0.082       0.070
Neighborhood[T.BrkSide]      0.0175      0.032      0.546      0.585
-0.045       0.080
Neighborhood[T.ClearCr]      0.0283      0.033      0.857      0.392
-0.037       0.093
```

```
Neighborhood[T.CollgCr]      -0.0253      0.025     -1.007      0.314
-0.075       0.024
Neighborhood[T.Crawfor]       0.0907      0.029      3.156      0.002
0.034       0.147
Neighborhood[T.Edwards]      -0.0473      0.028     -1.720      0.086
-0.101       0.007
Neighborhood[T.Gilbert]      -0.0187      0.026     -0.717      0.473
-0.070       0.032
Neighborhood[T.Greens]        0.0641      0.052      1.238      0.216
-0.037       0.166
Neighborhood[T.GrnHill]       0.5099      0.069      7.345      0.000
0.374       0.646
Neighborhood[T.IDOTRR]        0.0109      0.035      0.313      0.754
-0.057       0.079
Neighborhood[T.Landmrk]       0.0180      0.096      0.188      0.851
-0.170       0.206
Neighborhood[T.MeadowV]      -0.0757      0.041     -1.864      0.063
-0.155       0.004
Neighborhood[T.Mitchel]      -0.0382      0.028     -1.366      0.172
-0.093       0.017
Neighborhood[T.NAmes]        -0.0173      0.027     -0.644      0.520
-0.070       0.035
Neighborhood[T.NPkVill]      -0.0036      0.058     -0.062      0.951
-0.117       0.110
Neighborhood[T.NWAmes]       -0.0222      0.028     -0.805      0.421
-0.076       0.032
Neighborhood[T.NoRidge]       0.0219      0.029      0.745      0.457
-0.036       0.079
Neighborhood[T.NridgHt]       0.0425      0.026      1.635      0.102
-0.008       0.094
Neighborhood[T.OldTown]      -0.0168      0.032     -0.525      0.600
-0.080       0.046
Neighborhood[T.SWISU]        -0.0345      0.035     -0.994      0.320
-0.102       0.034
Neighborhood[T.Sawyer]        0.0097      0.028      0.353      0.724
-0.044       0.064
Neighborhood[T.SawyerW]      -0.0302      0.027     -1.106      0.269
-0.084       0.023
Neighborhood[T.Somerst]       0.0297      0.030      1.003      0.316
-0.028       0.088
Neighborhood[T.StoneBr]       0.0852      0.029      2.892      0.004
0.027       0.143
Neighborhood[T.Timber]       -0.0036      0.027     -0.131      0.896
-0.057       0.050
Neighborhood[T.Veenker]      -0.0013      0.033     -0.038      0.970
-0.066       0.064
OverallCond[T.2]             -0.0162      0.161     -0.101      0.920
-0.332       0.299
```

```
OverallCond[T.3]            -0.0028      0.124     -0.022      0.982
-0.246       0.240
OverallCond[T.4]             0.1160      0.124      0.937      0.349
-0.127       0.359
OverallCond[T.5]             0.1937      0.124      1.565      0.118
-0.049       0.436
OverallCond[T.6]             0.2207      0.124      1.777      0.076
-0.023       0.464
OverallCond[T.7]             0.2577      0.124      2.072      0.038
0.014       0.502
OverallCond[T.8]             0.2761      0.124      2.221      0.026
0.032       0.520
OverallCond[T.9]             0.3119      0.127      2.465      0.014
0.064       0.560
OverallQual[T.2]             0.7215      0.058     12.501      0.000
0.608       0.835
OverallQual[T.3]             1.0120      0.037     27.387      0.000
0.939       1.084
OverallQual[T.4]             1.0524      0.031     34.025      0.000
0.992       1.113
OverallQual[T.5]             1.1148      0.030     36.559      0.000
1.055       1.175
OverallQual[T.6]             1.1438      0.031     37.296      0.000
1.084       1.204
OverallQual[T.7]             1.1785      0.031     38.189      0.000
1.118       1.239
OverallQual[T.8]             1.2299      0.032     39.034      0.000
1.168       1.292
OverallQual[T.9]             1.2885      0.034     38.340      0.000
1.223       1.354
OverallQual[T.10]            1.3182      0.038     34.393      0.000
1.243       1.393
PavedDrive[T.P]              0.0081      0.018      0.448      0.654
-0.027       0.044
PavedDrive[T.Y]              0.0247      0.013      1.957      0.051
-5.44e-05       0.049
PoolQC[T.Fa]                 0.2286      0.162      1.412      0.158
-0.089       0.546
PoolQC[T.Gd]                 0.3563      0.151      2.365      0.018
0.061       0.652
PoolQC[T.NoPool]            -0.1255      0.107     -1.176      0.240
-0.335       0.084
PoolQC[T.TA]                 0.1469      0.099      1.486      0.137
-0.047       0.341
RoofMatl[T.Membran]          0.1658      0.114      1.450      0.147
-0.059       0.390
RoofMatl[T.Metal]            0.0680      0.111      0.615      0.539
-0.149       0.285
```

```
RoofMatl[T.Roll]               0.0973      0.097      1.001      0.317
-0.093       0.288
RoofMatl[T.Tar&Grv]            0.0010      0.046      0.021      0.983
-0.088       0.090
RoofMatl[T.WdShake]           -0.0359      0.039     -0.931      0.352
-0.111       0.040
RoofMatl[T.WdShngl]            0.0876      0.052      1.694      0.090
-0.014       0.189
RoofStyle[T.Gable]            -0.0135      0.052     -0.261      0.794
-0.115       0.088
RoofStyle[T.Gambrel]          -0.0645      0.058     -1.116      0.265
-0.178       0.049
RoofStyle[T.Hip]              -0.0050      0.052     -0.096      0.924
-0.107       0.097
RoofStyle[T.Mansard]          -0.1087      0.065     -1.682      0.093
-0.236       0.018
RoofStyle[T.Shed]             -0.0190      0.081     -0.234      0.815
-0.179       0.141
SaleCondition[T.AdjLand]       0.1867      0.044      4.285      0.000
0.101       0.272
SaleCondition[T.Alloca]        0.0750      0.034      2.174      0.030
0.007       0.143
SaleCondition[T.Family]        0.0209      0.019      1.119      0.263
-0.016       0.058
SaleCondition[T.Normal]        0.0367      0.010      3.607      0.000
0.017       0.057
SaleCondition[T.Partial]       0.0842      0.047      1.791      0.073
-0.008       0.176
SaleType[T.CWD]                0.0224      0.029      0.763      0.445
-0.035       0.080
SaleType[T.Con]                0.0588      0.055      1.073      0.284
-0.049       0.166
SaleType[T.ConLD]              0.0498      0.031      1.597      0.110
-0.011       0.111
SaleType[T.ConLI]             -0.1998      0.058     -3.430      0.001
-0.314      -0.086
SaleType[T.ConLw]             -0.0918      0.049     -1.880      0.060
-0.188       0.004
SaleType[T.New]               -0.0053      0.049     -0.108      0.914
-0.101       0.091
SaleType[T.Oth]                0.1280      0.089      1.446      0.149
-0.046       0.302
SaleType[T.VWD]               -0.0075      0.092     -0.082      0.935
-0.187       0.172
SaleType[T.WD ]                0.0094      0.014      0.661      0.509
-0.018       0.037
Street[T.Pave]                 0.0534      0.045      1.200      0.230
-0.034       0.141
```

```
TotRmsAbvGrd[T.4]              0.0256      0.032      0.810      0.418
-0.036       0.088
TotRmsAbvGrd[T.5]              0.0200      0.032      0.635      0.526
-0.042       0.082
TotRmsAbvGrd[T.6]              0.0315      0.032      0.983      0.326
-0.031       0.094
TotRmsAbvGrd[T.7]              0.0267      0.033      0.807      0.420
-0.038       0.091
TotRmsAbvGrd[T.8]              0.0291      0.034      0.851      0.395
-0.038       0.096
TotRmsAbvGrd[T.9]              0.0087      0.036      0.241      0.810
-0.062       0.079
TotRmsAbvGrd[T.10]            -0.0115      0.039     -0.297      0.766
-0.088       0.064
TotRmsAbvGrd[T.11]            -0.0404      0.043     -0.947      0.344
-0.124       0.043
TotRmsAbvGrd[T.12]             0.0872      0.077      1.139      0.255
-0.063       0.237
TotRmsAbvGrd[T.13]          4.001e-16   9.21e-17      4.346      0.000
2.2e-16    5.81e-16
Utilities[T.NoSeWa]         1.151e-16   6.26e-17      1.838      0.066
-7.74e-18    2.38e-16
Utilities[T.NoSewr]           -0.0517      0.139     -0.372      0.710
-0.324       0.221
Artery                        -0.0877      0.015     -5.813      0.000
-0.117      -0.058
AsbShng                       -0.0202      0.025     -0.823      0.411
-0.068       0.028
AsphShn                        0.1446      0.068      2.120      0.034
0.011       0.278
BrkCmn                         0.0401      0.063      0.637      0.524
-0.083       0.163
BrkComm                        0.2073      0.067      3.112      0.002
0.077       0.338
BrkFace                        0.0524      0.015      3.577      0.000
0.024       0.081
BsmtFinSF1                  5.092e-05   9.19e-06      5.540      0.000
3.29e-05     6.9e-05
BsmtFinSF2                  4.314e-05   1.86e-05      2.317      0.021
6.62e-06    7.97e-05
BsmtUnfSF                  -1.353e-05    8.5e-06     -1.592      0.112
-3.02e-05    3.14e-06
CBlock                         2.8036      0.111     25.216      0.000
2.585       3.022
CemntBd                       -0.0711      0.066     -1.077      0.282
-0.201       0.058
CmentBd                        0.0901      0.066      1.361      0.174
-0.040       0.220
```

```
EnclosedPorch               7.145e-05   4.04e-05      1.767      0.077
-7.86e-06       0.000
Feedr                         -0.0541      0.011     -5.062      0.000
-0.075      -0.033
GarageArea                  8.381e-05   2.65e-05      3.158      0.002
3.18e-05       0.000
GrLivArea                      0.0002   2.25e-05      8.326      0.000
0.000       0.000
HdBoard                       -0.0063      0.012     -0.528      0.597
-0.030       0.017
ImStucc                       -0.0304      0.030     -1.019      0.308
-0.089       0.028
LotArea                     2.117e-06   4.41e-07      4.798      0.000
1.25e-06    2.98e-06
LotFrontage                 9.516e-05    7.7e-05      1.236      0.217
-5.58e-05       0.000
LowQualFinSF               -3.663e-06   5.66e-05     -0.065      0.948
-0.000       0.000
MasVnrArea                  3.583e-05   1.91e-05      1.878      0.061
-1.59e-06    7.33e-05
MetalSd                        0.0177      0.012      1.420      0.156
-0.007       0.042
MiscVal                     7.798e-06   7.97e-06      0.978      0.328
-7.84e-06    2.34e-05
Norm                           0.0444      0.026      1.684      0.092
-0.007       0.096
OpenPorchSF                    0.0001   3.74e-05      3.211      0.001
4.67e-05       0.000
Other                         -0.0695      0.088     -0.786      0.432
-0.243       0.104
Plywood                       -0.0009      0.012     -0.082      0.934
-0.024       0.022
PoolArea                      -0.0004      0.000     -1.284      0.199
-0.001       0.000
PosA                           0.0615      0.027      2.317      0.021
0.009       0.114
PosN                           0.0239      0.021      1.160      0.246
-0.017       0.064
PreCast                             0          0        nan        nan
0           0
RRAe                          -0.0580      0.023     -2.543      0.011
-0.103      -0.013
RRAn                          -0.0188      0.016     -1.144      0.253
-0.051       0.013
RRNe                          -0.0223      0.047     -0.474      0.635
-0.114       0.070
RRNn                           0.0497      0.042      1.193      0.233
-0.032       0.131
```

```
ScreenPorch                      0.0002    4.06e-05      4.759      0.000
0.000       0.000
Stone                           -0.0193       0.048     -0.405      0.686
-0.113       0.074
Stucco                           0.0211       0.019      1.131      0.258
-0.015       0.058
TotalBsmtSF                   8.053e-05    1.24e-05      6.511      0.000
5.63e-05       0.000
VinylSd                          0.0061       0.013      0.478      0.633
-0.019       0.031
WdSdng                           0.0120       0.012      0.981      0.327
-0.012       0.036
WdShing                          0.0059       0.022      0.266      0.790
-0.037       0.049
WdShng                          -0.0026       0.018     -0.143      0.887
-0.038       0.033
WoodDeckSF                    4.619e-05    1.96e-05      2.352      0.019
7.67e-06    8.47e-05
X1stFlrSF                     8.927e-05    2.26e-05      3.950      0.000
4.49e-05       0.000
X2ndFlrSF                        0.0001    2.12e-05      4.812      0.000
6.04e-05       0.000
X3SsnPorch                       0.0002    9.51e-05      2.141      0.032
1.7e-05       0.000
YearsSince1950Built              0.0031       0.001      6.064      0.000
0.002       0.004
YearsSince1950GarageBuilt       -0.0003       0.000     -1.110      0.267
-0.001       0.000
YearsSince1950Remod              0.0005       0.000      2.464      0.014
9.94e-05       0.001
==============================================================================
Omnibus:                      266.117   Durbin-Watson:                   1.943
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1925.093
Skew:                          -0.463   Prob(JB):                         0.00
Kurtosis:                       7.941   Cond. No.                     3.80e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 2.19e-22. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

[ ]:

```
y_train = ames_train['LogSalePrice']
y_test = ames_test['LogSalePrice']
```

[ ]:

```
print_metrics(nlr, ames_train, y_train, ames_test, y_test, flag_log_sale_price␣
  ↪= True)
print_metrics(nlr, ames_train, y_train, ames_test, y_test, flag_log_sale_price␣
  ↪= False)
```

```
Metrics for Log(Sale Price):

Training R2 0.9603376376378353
Training MAE 0.056296645864186146
Training RMSE 0.07708753756375215
Out-of-sample R2 0.84160608337568
Out-of-sample MAE 0.08310808994525071
Out-of-sample RMSE 0.15762880434009238

Metrics for Sale Price:

Training R2 0.9617308456634752
Training MAE 10352.437245152909
Training RMSE 15282.418966032086
Out-of-sample R2 0.3685202291723365
Out-of-sample MAE 15719.441358572112
Out-of-sample RMSE 62283.05295821487
```

6 In-class activity 4: Remove some predictors of the linear regression model based on the output. Does the out-of-sample performance get better?

[ ]:

```

```

b) with polynomial features

[ ]:

```

```

6.0.1 Higher-order Variables

We can construct new features using a polynomial transformation. This is necessary because the regression plots we generated with ‘LogSalePrice ~ [single independent variable]’ exhibits some non-linear relationship. In the function below, you can choose the highest degree of the polynomial features. A higher degree polynomial might cause overfitting concern, but we will later use ‘regularization’ to mitigate this issue.

[ ]:

```
def create_polynomial_features(df, n_degree):

    new_df = None

    for i in range(2, n_degree+1):
```

```
        tmp = df.pow(i)

        affix = '_p'+str(i)
        tmp.columns = list(map(lambda x: x + affix, df.columns))

        if new_df is not None:
            new_df = pd.concat([new_df, tmp], axis=1)
        else:
            new_df = tmp

    return new_df
```

NOTE: An important consideration when creating higher-order variables is that the resulting features will tend to have some degree of linear dependence amongst themselves. This is normal as several new features are based on their zero-th power peer. Such correlation can also yield a high degree of multicollinearity in the regression models. The `sklearn` implementations that we will be using do not automatically account for this phenomenon, therefore we must be careful in selection the `n_degree`, and analyzing the model fit.

```
[ ]: # We only choose a select list of variables to do polynomial transformation.
poly_cols = ['LotFrontage', 'LotArea', 'MasVnrArea', 'BsmtFinSF1',␣
 ↪'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF',
             'X1stFlrSF', 'X2ndFlrSF', 'LowQualFinSF', 'GrLivArea',␣
 ↪'GarageArea', 'WoodDeckSF', 'OpenPorchSF',
             'EnclosedPorch', 'X3SsnPorch', 'ScreenPorch', 'MiscVal',␣
 ↪'YearsSince1950Built',
             'YearsSince1950Remod', 'YearsSince1950GarageBuilt']
```

```
[ ]: n_degree = 2

train_poly_temp = create_polynomial_features(ames_train[poly_cols], n_degree)
test_poly_temp = create_polynomial_features(ames_test[poly_cols], n_degree)

ames_train_poly = pd.concat([ames_train, train_poly_temp], axis=1)
ames_test_poly = pd.concat([ames_test, test_poly_temp], axis=1)

print(ames_train.shape, ames_test.shape)
print(train_poly_temp.shape, test_poly_temp.shape)
print(ames_train_poly.shape, ames_test_poly.shape)
```

```
(1828, 104) (937, 104)
(1828, 21) (937, 21)
(1828, 125) (937, 125)
```

```
[ ]: print(ames_train_poly.shape)
all_columns = "+".join(ames_train_poly.columns.difference(["LogSalePrice"]))
my_formula = "LogSalePrice~" + all_columns +'-1'
```

```python
print(my_formula)

mod_naive_poly = smf.ols(my_formula, data=ames_train_poly)
nlr_poly = mod_naive_poly.fit()

print(nlr_poly.summary())
```

```
(1828, 125)
LogSalePrice~Alley+Artery+AsbShng+AsphShn+BedroomAbvGr+BldgType+BrkCmn+BrkComm+B
rkFace+BsmtCond+BsmtExposure+BsmtFinSF1+BsmtFinSF1_p2+BsmtFinSF2+BsmtFinSF2_p2+B
smtFinType1+BsmtFinType2+BsmtFullBath+BsmtHalfBath+BsmtQual+BsmtUnfSF+BsmtUnfSF_
p2+CBlock+CemntBd+CentralAir+CmentBd+Electrical+EnclosedPorch+EnclosedPorch_p2+E
xterCond+ExterQual+Feedr+Fence+FireplaceQu+Fireplaces+Foundation+FullBath+Functi
onal+GarageArea+GarageArea_p2+GarageCars+GarageCond+GarageFinish+GarageQual+Gara
geType+GrLivArea+GrLivArea_p2+HalfBath+HdBoard+Heating+HeatingQC+HouseStyle+ImSt
ucc+KitchenAbvGr+KitchenQual+LandContour+LandSlope+LotArea+LotArea_p2+LotConfig+
LotFrontage+LotFrontage_p2+LotShape+LowQualFinSF+LowQualFinSF_p2+MSSubClass+MSZo
ning+MasVnrArea+MasVnrArea_p2+MasVnrType+MetalSd+MiscFeature+MiscVal+MiscVal_p2+
MoSold+Neighborhood+Norm+OpenPorchSF+OpenPorchSF_p2+Other+OverallCond+OverallQua
l+PavedDrive+Plywood+PoolArea+PoolQC+PosA+PosN+PreCast+RRAe+RRAn+RRNe+RRNn+RoofM
atl+RoofStyle+SaleCondition+SaleType+ScreenPorch+ScreenPorch_p2+Stone+Street+Stu
cco+TotRmsAbvGrd+TotalBsmtSF+TotalBsmtSF_p2+Utilities+VinylSd+WdSdng+WdShing+WdS
hng+WoodDeckSF+WoodDeckSF_p2+X1stFlrSF+X1stFlrSF_p2+X2ndFlrSF+X2ndFlrSF_p2+X3Ssn
Porch+X3SsnPorch_p2+YearsSince1950Built+YearsSince1950Built_p2+YearsSince1950Gar
ageBuilt+YearsSince1950GarageBuilt_p2+YearsSince1950Remod+YearsSince1950Remod_p2
-1
                            OLS Regression Results
==============================================================================
Dep. Variable:           LogSalePrice   R-squared:                       0.962
Model:                            OLS   Adj. R-squared:                  0.954
Method:                 Least Squares   F-statistic:                     118.8
Date:                Fri, 01 Nov 2024   Prob (F-statistic):               0.00
Time:                        18:18:57   Log-Likelihood:                 2140.1
No. Observations:                1828   AIC:                            -3630.
Df Residuals:                    1503   BIC:                            -1839.
Df Model:                         324
Covariance Type:            nonrobust
================================================================================
================
                                   coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
----------------
Alley[Grvl]                      4.9635      0.140     35.451      0.000
4.689       5.238
Alley[NoAccess]                  4.9859      0.139     35.754      0.000
4.712       5.259
```

```
Alley[Pave]                      4.9939      0.140     35.609      0.000
4.719       5.269
BedroomAbvGr[T.1]               -0.0107      0.074     -0.145      0.885
-0.155       0.134
BedroomAbvGr[T.2]               -0.0093      0.074     -0.126      0.900
-0.154       0.136
BedroomAbvGr[T.3]               -0.0123      0.074     -0.165      0.869
-0.158       0.133
BedroomAbvGr[T.4]               -0.0144      0.075     -0.193      0.847
-0.161       0.132
BedroomAbvGr[T.5]               -0.0677      0.077     -0.880      0.379
-0.218       0.083
BedroomAbvGr[T.6]                0.0342      0.088      0.389      0.697
-0.138       0.206
BldgType[T.2fmCon]              -0.1128      0.106     -1.061      0.289
-0.321       0.096
BldgType[T.Duplex]              -0.0287      0.017     -1.726      0.085
-0.061       0.004
BldgType[T.Twnhs]               -0.0321      0.042     -0.764      0.445
-0.115       0.050
BldgType[T.TwnhsE]               0.0022      0.039      0.057      0.954
-0.074       0.079
BsmtCond[T.Fa]                   0.0069      0.055      0.127      0.899
-0.100       0.114
BsmtCond[T.Gd]                   0.0116      0.054      0.215      0.830
-0.094       0.117
BsmtCond[T.NaN]                 -0.7545      0.043    -17.473      0.000
-0.839      -0.670
BsmtCond[T.NoBasement]          -0.0166      0.014     -1.153      0.249
-0.045       0.012
BsmtCond[T.Po]                  -0.1085      0.089     -1.216      0.224
-0.284       0.067
BsmtCond[T.TA]                   0.0084      0.053      0.158      0.875
-0.096       0.112
BsmtExposure[T.Gd]               0.0269      0.010      2.826      0.005
0.008       0.046
BsmtExposure[T.Mn]              -0.0260      0.010     -2.672      0.008
-0.045      -0.007
BsmtExposure[T.NaN]              0.0131      0.061      0.215      0.830
-0.107       0.133
BsmtExposure[T.No]              -0.0217      0.007     -2.959      0.003
-0.036      -0.007
BsmtExposure[T.NoBasement]      -0.0166      0.014     -1.153      0.249
-0.045       0.012
BsmtFinType1[T.BLQ]             -0.0129      0.009     -1.366      0.172
-0.031       0.006
BsmtFinType1[T.GLQ]              0.0008      0.008      0.095      0.924
-0.016       0.017
```

```
BsmtFinType1[T.LwQ]            -0.0370      0.012     -3.138      0.002
-0.060      -0.014
BsmtFinType1[T.NaN]            -0.7545      0.043    -17.473      0.000
-0.839      -0.670
BsmtFinType1[T.NoBasement]     -0.0166      0.014     -1.153      0.249
-0.045       0.012
BsmtFinType1[T.Rec]            -0.0333      0.009     -3.594      0.000
-0.052      -0.015
BsmtFinType1[T.Unf]            -0.0155      0.011     -1.378      0.168
-0.037       0.007
BsmtFinType2[T.BLQ]            -0.0291      0.023     -1.248      0.212
-0.075       0.017
BsmtFinType2[T.GLQ]            -0.0021      0.027     -0.078      0.938
-0.056       0.051
BsmtFinType2[T.LwQ]            -0.0309      0.022     -1.406      0.160
-0.074       0.012
BsmtFinType2[T.NaN]            -0.7545      0.043    -17.473      0.000
-0.839      -0.670
BsmtFinType2[T.NoBasement]     -0.0166      0.014     -1.153      0.249
-0.045       0.012
BsmtFinType2[T.Rec]            -0.0371      0.021     -1.762      0.078
-0.078       0.004
BsmtFinType2[T.Unf]            -0.0284      0.025     -1.138      0.255
-0.077       0.021
BsmtFullBath[T.1]               0.0120      0.007      1.798      0.072
-0.001       0.025
BsmtFullBath[T.2]               0.0812      0.033      2.465      0.014
0.017       0.146
BsmtHalfBath[T.1]               0.0145      0.010      1.454      0.146
-0.005       0.034
BsmtHalfBath[T.2]               0.1422      0.120      1.187      0.235
-0.093       0.377
BsmtQual[T.Fa]                 -0.0543      0.020     -2.696      0.007
-0.094      -0.015
BsmtQual[T.Gd]                 -0.0198      0.011     -1.860      0.063
-0.041       0.001
BsmtQual[T.NaN]                -0.7545      0.043    -17.473      0.000
-0.839      -0.670
BsmtQual[T.NoBasement]         -0.0166      0.014     -1.153      0.249
-0.045       0.012
BsmtQual[T.Po]                  0.0744      0.089      0.836      0.403
-0.100       0.249
BsmtQual[T.TA]                 -0.0123      0.014     -0.900      0.368
-0.039       0.015
CentralAir[T.Y]                 0.0222      0.013      1.661      0.097
-0.004       0.048
Electrical[T.FuseF]            -0.0063      0.022     -0.288      0.774
-0.049       0.037
```

```
Electrical[T.FuseP]              0.0237      0.056      0.420      0.674
-0.087       0.134
Electrical[T.Mix]               -0.1085      0.089     -1.216      0.224
-0.284       0.067
Electrical[T.NaN]                0.0382      0.088      0.434      0.664
-0.135       0.211
Electrical[T.SBrkr]             -0.0080      0.010     -0.786      0.432
-0.028       0.012
ExterCond[T.Fa]                 -0.0025      0.041     -0.060      0.952
-0.084       0.079
ExterCond[T.Gd]                  0.0187      0.036      0.516      0.606
-0.052       0.090
ExterCond[T.Po]                  1.4030      0.234      6.000      0.000
0.944       1.862
ExterCond[T.TA]                  0.0310      0.036      0.861      0.390
-0.040       0.102
ExterQual[T.Fa]                 -0.0100      0.039     -0.256      0.798
-0.087       0.067
ExterQual[T.Gd]                 -0.0005      0.018     -0.030      0.976
-0.035       0.034
ExterQual[T.TA]                 -0.0108      0.019     -0.554      0.580
-0.049       0.027
Fence[T.GdWo]                    0.0020      0.016      0.119      0.905
-0.030       0.034
Fence[T.MnPrv]                  -0.0081      0.013     -0.613      0.540
-0.034       0.018
Fence[T.MnWw]                   -0.0169      0.031     -0.549      0.583
-0.077       0.043
Fence[T.NoFence]                -0.0032      0.012     -0.273      0.785
-0.026       0.020
FireplaceQu[T.Fa]               -0.0012      0.023     -0.052      0.958
-0.046       0.044
FireplaceQu[T.Gd]                0.0051      0.018      0.278      0.781
-0.031       0.041
FireplaceQu[T.NoFirePlace]       2.9952      0.082     36.637      0.000
2.835       3.156
FireplaceQu[T.Po]               -0.0172      0.025     -0.677      0.499
-0.067       0.033
FireplaceQu[T.TA]               -0.0043      0.019     -0.223      0.824
-0.042       0.033
Fireplaces[T.1]                  3.0216      0.082     36.907      0.000
2.861       3.182
Fireplaces[T.2]                  3.0565      0.082     37.086      0.000
2.895       3.218
Fireplaces[T.3]                  3.0135      0.087     34.716      0.000
2.843       3.184
Fireplaces[T.4]                  2.8565      0.186     15.394      0.000
2.493       3.220
```

```
Foundation[T.CBlock]          0.0216      0.011      1.943      0.052
-0.000       0.043
Foundation[T.PConc]           0.0242      0.012      2.056      0.040
0.001       0.047
Foundation[T.Slab]            0.0018      0.036      0.049      0.961
-0.068       0.072
Foundation[T.Stone]          -0.0180      0.047     -0.387      0.698
-0.109       0.073
Foundation[T.Wood]            0.0413      0.064      0.641      0.521
-0.085       0.168
FullBath[T.1]                 0.0427      0.076      0.565      0.572
-0.105       0.191
FullBath[T.2]                 0.0605      0.076      0.792      0.429
-0.089       0.210
FullBath[T.3]                 0.1131      0.078      1.448      0.148
-0.040       0.266
FullBath[T.4]               2.24e-14   1.46e-15     15.327      0.000
1.95e-14    2.53e-14
Functional[T.Maj2]            0.0230      0.069      0.334      0.739
-0.112       0.158
Functional[T.Min1]            0.0628      0.033      1.920      0.055
-0.001       0.127
Functional[T.Min2]            0.0276      0.034      0.821      0.412
-0.038       0.094
Functional[T.Mod]             0.0274      0.037      0.749      0.454
-0.044       0.099
Functional[T.Sal]            -0.3687      0.152     -2.433      0.015
-0.666      -0.071
Functional[T.Sev]            -0.2319      0.109     -2.126      0.034
-0.446      -0.018
Functional[T.Typ]             0.0875      0.029      3.020      0.003
0.031       0.144
GarageCars[T.2]               0.0186      0.010      1.856      0.064
-0.001       0.038
GarageCars[T.3]               0.0462      0.016      2.849      0.004
0.014       0.078
GarageCars[T.4]               0.0584      0.040      1.466      0.143
-0.020       0.137
GarageCars[T.5]            2.698e-15   1.03e-15      2.622      0.009
6.79e-16    4.72e-15
GarageCond[T.Fa]              0.0675      0.082      0.822      0.411
-0.094       0.229
GarageCond[T.Gd]              0.0868      0.083      1.043      0.297
-0.076       0.250
GarageCond[T.Po]              0.1992      0.091      2.185      0.029
0.020       0.378
GarageCond[T.TA]              0.0833      0.081      1.033      0.302
-0.075       0.241
```

```
GarageFinish[T.RFn]              -0.0010      0.006     -0.156      0.876
-0.013       0.011
GarageFinish[T.Unf]              -0.0060      0.008     -0.781      0.435
-0.021       0.009
GarageQual[T.Fa]                 -0.1885      0.098     -1.923      0.055
-0.381       0.004
GarageQual[T.Gd]                 -0.1530      0.096     -1.597      0.110
-0.341       0.035
GarageQual[T.Po]                 -0.1660      0.129     -1.284      0.199
-0.420       0.088
GarageQual[T.TA]                 -0.1662      0.097     -1.715      0.086
-0.356       0.024
GarageType[T.Attchd]              0.0432      0.024      1.821      0.069
-0.003       0.090
GarageType[T.Basment]             0.0214      0.033      0.643      0.520
-0.044       0.087
GarageType[T.BuiltIn]             0.0358      0.026      1.378      0.168
-0.015       0.087
GarageType[T.CarPort]            -0.0107      0.040     -0.271      0.786
-0.088       0.067
GarageType[T.Detchd]              0.0555      0.024      2.343      0.019
0.009       0.102
HalfBath[T.1]                     0.0209      0.007      2.830      0.005
0.006       0.035
HalfBath[T.2]                    -0.0971      0.033     -2.970      0.003
-0.161      -0.033
Heating[T.GasA]                   0.1664      0.097      1.716      0.086
-0.024       0.357
Heating[T.GasW]                   0.1591      0.100      1.593      0.111
-0.037       0.355
Heating[T.Grav]                  -0.0336      0.158     -0.213      0.831
-0.343       0.276
Heating[T.OthW]                   0.1002      0.135      0.742      0.458
-0.165       0.365
Heating[T.Wall]                   0.1925      0.119      1.624      0.105
-0.040       0.425
HeatingQC[T.Fa]                  -0.0532      0.016     -3.393      0.001
-0.084      -0.022
HeatingQC[T.Gd]                  -0.0075      0.007     -1.114      0.265
-0.021       0.006
HeatingQC[T.Po]                  -0.1181      0.069     -1.702      0.089
-0.254       0.018
HeatingQC[T.TA]                  -0.0306      0.007     -4.461      0.000
-0.044      -0.017
HouseStyle[T.1.5Unf]             -0.0012      0.074     -0.016      0.987
-0.146       0.144
HouseStyle[T.1Story]              0.0209      0.033      0.636      0.525
-0.044       0.085
```

```
HouseStyle[T.2.5Fin]          -0.0739      0.070     -1.055      0.292
-0.211       0.063
HouseStyle[T.2.5Unf]           0.0440      0.046      0.957      0.339
-0.046       0.134
HouseStyle[T.2Story]           0.0046      0.031      0.146      0.884
-0.057       0.066
HouseStyle[T.SFoyer]           0.0292      0.045      0.652      0.515
-0.059       0.117
HouseStyle[T.SLvl]             0.0588      0.047      1.262      0.207
-0.033       0.150
KitchenAbvGr[T.1]             -0.1250      0.091     -1.367      0.172
-0.304       0.054
KitchenAbvGr[T.2]             -0.1718      0.097     -1.778      0.076
-0.361       0.018
KitchenAbvGr[T.3]          -4.375e-15   1.12e-15     -3.897      0.000
-6.58e-15   -2.17e-15
KitchenQual[T.Fa]             -0.0534      0.022     -2.404      0.016
-0.097      -0.010
KitchenQual[T.Gd]             -0.0446      0.012     -3.674      0.000
-0.068      -0.021
KitchenQual[T.Po]           -5.42e-16   1.14e-15     -0.477      0.634
-2.77e-15    1.69e-15
KitchenQual[T.TA]             -0.0542      0.014     -3.990      0.000
-0.081      -0.028
LandContour[T.HLS]             0.0213      0.017      1.256      0.209
-0.012       0.055
LandContour[T.Low]             0.0109      0.024      0.454      0.650
-0.036       0.058
LandContour[T.Lvl]             0.0052      0.013      0.402      0.688
-0.020       0.030
LandSlope[T.Mod]               0.0163      0.014      1.191      0.234
-0.011       0.043
LandSlope[T.Sev]               0.0025      0.042      0.060      0.952
-0.081       0.086
LotConfig[T.CulDSac]           0.0096      0.011      0.874      0.382
-0.012       0.031
LotConfig[T.FR2]              -0.0247      0.014     -1.741      0.082
-0.053       0.003
LotConfig[T.FR3]              -0.0092      0.028     -0.329      0.742
-0.064       0.046
LotConfig[T.Inside]           -0.0009      0.006     -0.150      0.880
-0.013       0.011
LotShape[T.IR2]               -0.0044      0.013     -0.329      0.742
-0.031       0.022
LotShape[T.IR3]                0.0037      0.028      0.132      0.895
-0.051       0.059
LotShape[T.Reg]                0.0005      0.005      0.093      0.926
-0.010       0.011
```

```
MSSubClass[T.30]                -0.0582      0.017     -3.491      0.000
-0.091      -0.026
MSSubClass[T.40]                -0.0310      0.049     -0.626      0.531
-0.128       0.066
MSSubClass[T.45]                 0.0251      0.076      0.330      0.742
-0.124       0.175
MSSubClass[T.50]                 0.0112      0.031      0.357      0.721
-0.050       0.073
MSSubClass[T.60]                -0.0209      0.031     -0.665      0.506
-0.082       0.041
MSSubClass[T.70]                 0.0243      0.033      0.733      0.464
-0.041       0.090
MSSubClass[T.75]                 0.0014      0.049      0.028      0.978
-0.095       0.098
MSSubClass[T.80]                -0.0559      0.043     -1.305      0.192
-0.140       0.028
MSSubClass[T.85]                 0.0159      0.039      0.412      0.680
-0.060       0.092
MSSubClass[T.90]                -0.0287      0.017     -1.726      0.085
-0.061       0.004
MSSubClass[T.120]               -0.0302      0.039     -0.766      0.444
-0.108       0.047
MSSubClass[T.150]               -0.1521      0.118     -1.284      0.199
-0.384       0.080
MSSubClass[T.160]               -0.1089      0.053     -2.065      0.039
-0.212      -0.005
MSSubClass[T.180]               -0.0694      0.062     -1.116      0.265
-0.192       0.053
MSSubClass[T.190]                0.0679      0.104      0.652      0.514
-0.136       0.272
MSZoning[T.C]                    1.4713      0.059     24.956      0.000
1.356       1.587
MSZoning[T.FV]                   1.7984      0.051     35.292      0.000
1.698       1.898
MSZoning[T.I]                    1.4924      0.141     10.608      0.000
1.216       1.768
MSZoning[T.RH]                   1.7684      0.053     33.281      0.000
1.664       1.873
MSZoning[T.RL]                   1.7799      0.048     37.131      0.000
1.686       1.874
MSZoning[T.RM]                   1.7332      0.049     35.721      0.000
1.638       1.828
MasVnrType[T.BrkFace]            0.0196      0.023      0.862      0.389
-0.025       0.064
MasVnrType[T.CBlock]            -0.3480      0.117     -2.965      0.003
-0.578      -0.118
MasVnrType[T.NaN]                0.0277      0.024      1.180      0.238
-0.018       0.074
```

```
MasVnrType[T.Stone]               0.0157      0.024      0.651      0.515
-0.032       0.063
MiscFeature[T.NaN]                0.0236      0.087      0.272      0.786
-0.147       0.194
MiscFeature[T.Othr]               0.0872      0.087      0.999      0.318
-0.084       0.258
MiscFeature[T.Shed]              -0.0027      0.079     -0.034      0.973
-0.157       0.151
MiscFeature[T.TenC]              -0.3784      0.162     -2.341      0.019
-0.695      -0.061
MoSold[T.2]                      -0.0384      0.015     -2.507      0.012
-0.068      -0.008
MoSold[T.3]                      -0.0219      0.014     -1.577      0.115
-0.049       0.005
MoSold[T.4]                      -0.0083      0.014     -0.610      0.542
-0.035       0.018
MoSold[T.5]                       0.0011      0.013      0.086      0.932
-0.024       0.026
MoSold[T.6]                      -0.0017      0.012     -0.139      0.890
-0.026       0.023
MoSold[T.7]                       0.0033      0.012      0.266      0.790
-0.021       0.028
MoSold[T.8]                      -0.0147      0.013     -1.112      0.266
-0.041       0.011
MoSold[T.9]                       0.0037      0.014      0.257      0.798
-0.024       0.032
MoSold[T.10]                     -0.0216      0.014     -1.533      0.126
-0.049       0.006
MoSold[T.11]                     -0.0103      0.014     -0.714      0.475
-0.038       0.018
MoSold[T.12]                     -0.0104      0.015     -0.677      0.499
-0.040       0.020
Neighborhood[T.Blueste]           0.1267      0.054      2.326      0.020
0.020       0.234
Neighborhood[T.BrDale]            0.0042      0.039      0.107      0.915
-0.073       0.081
Neighborhood[T.BrkSide]           0.0211      0.032      0.657      0.511
-0.042       0.084
Neighborhood[T.ClearCr]           0.0209      0.034      0.621      0.534
-0.045       0.087
Neighborhood[T.CollgCr]          -0.0314      0.025     -1.238      0.216
-0.081       0.018
Neighborhood[T.Crawfor]           0.0772      0.029      2.663      0.008
0.020       0.134
Neighborhood[T.Edwards]          -0.0495      0.028     -1.789      0.074
-0.104       0.005
Neighborhood[T.Gilbert]          -0.0302      0.026     -1.154      0.249
-0.082       0.021
```

```
Neighborhood[T.Greens]          0.0771      0.052      1.483      0.138
-0.025      0.179
Neighborhood[T.GrnHill]         0.4596      0.069      6.657      0.000
0.324      0.595
Neighborhood[T.IDOTRR]          0.0129      0.035      0.367      0.714
-0.056      0.082
Neighborhood[T.Landmrk]         0.0131      0.094      0.139      0.890
-0.172      0.198
Neighborhood[T.MeadowV]        -0.0561      0.041     -1.361      0.174
-0.137      0.025
Neighborhood[T.Mitchel]        -0.0392      0.028     -1.384      0.166
-0.095      0.016
Neighborhood[T.NAmes]          -0.0209      0.027     -0.773      0.440
-0.074      0.032
Neighborhood[T.NPkVill]         0.0059      0.057      0.103      0.918
-0.106      0.118
Neighborhood[T.NWAmes]         -0.0254      0.028     -0.901      0.368
-0.081      0.030
Neighborhood[T.NoRidge]         0.0299      0.030      1.001      0.317
-0.029      0.089
Neighborhood[T.NridgHt]         0.0288      0.026      1.104      0.270
-0.022      0.080
Neighborhood[T.OldTown]        -0.0178      0.032     -0.551      0.581
-0.081      0.046
Neighborhood[T.SWISU]          -0.0364      0.035     -1.046      0.296
-0.105      0.032
Neighborhood[T.Sawyer]          0.0084      0.028      0.302      0.762
-0.046      0.063
Neighborhood[T.SawyerW]        -0.0412      0.028     -1.487      0.137
-0.096      0.013
Neighborhood[T.Somerst]         0.0144      0.030      0.488      0.626
-0.044      0.073
Neighborhood[T.StoneBr]         0.0710      0.030      2.367      0.018
0.012      0.130
Neighborhood[T.Timber]         -0.0154      0.027     -0.563      0.574
-0.069      0.038
Neighborhood[T.Veenker]        -0.0114      0.034     -0.341      0.733
-0.077      0.054
OverallCond[T.2]                0.0450      0.158      0.284      0.776
-0.266      0.356
OverallCond[T.3]                0.0153      0.122      0.126      0.900
-0.224      0.255
OverallCond[T.4]                0.1183      0.122      0.971      0.332
-0.121      0.357
OverallCond[T.5]                0.2038      0.122      1.673      0.094
-0.035      0.443
OverallCond[T.6]                0.2354      0.122      1.926      0.054
-0.004      0.475
```

```
OverallCond[T.7]                  0.2765      0.122      2.259      0.024
0.036       0.517
OverallCond[T.8]                  0.2983      0.122      2.438      0.015
0.058       0.538
OverallCond[T.9]                  0.3423      0.125      2.748      0.006
0.098       0.587
OverallQual[T.2]                  0.7610      0.058     13.213      0.000
0.648       0.874
OverallQual[T.3]                  1.0193      0.037     27.679      0.000
0.947       1.092
OverallQual[T.4]                  1.0455      0.031     33.832      0.000
0.985       1.106
OverallQual[T.5]                  1.1034      0.030     36.215      0.000
1.044       1.163
OverallQual[T.6]                  1.1296      0.031     36.839      0.000
1.069       1.190
OverallQual[T.7]                  1.1648      0.031     37.683      0.000
1.104       1.225
OverallQual[T.8]                  1.2162      0.032     38.586      0.000
1.154       1.278
OverallQual[T.9]                  1.2776      0.034     38.117      0.000
1.212       1.343
OverallQual[T.10]                 1.3260      0.038     34.507      0.000
1.251       1.401
PavedDrive[T.P]                   0.0032      0.018      0.177      0.859
-0.032       0.039
PavedDrive[T.Y]                   0.0293      0.013      2.332      0.020
0.005       0.054
PoolQC[T.Fa]                      0.3093      0.167      1.852      0.064
-0.018       0.637
PoolQC[T.Gd]                      0.4395      0.156      2.824      0.005
0.134       0.745
PoolQC[T.NoPool]                 -0.2178      0.108     -2.012      0.044
-0.430      -0.005
PoolQC[T.TA]                      0.0348      0.104      0.333      0.739
-0.170       0.239
RoofMatl[T.Membran]               0.0795      0.115      0.689      0.491
-0.147       0.306
RoofMatl[T.Metal]                 0.0230      0.112      0.206      0.837
-0.197       0.243
RoofMatl[T.Roll]                  0.0743      0.096      0.774      0.439
-0.114       0.262
RoofMatl[T.Tar&Grv]              -0.0012      0.045     -0.026      0.979
-0.090       0.087
RoofMatl[T.WdShake]              -0.0181      0.038     -0.472      0.637
-0.093       0.057
RoofMatl[T.WdShngl]               0.0631      0.051      1.227      0.220
-0.038       0.164
```

```
RoofStyle[T.Gable]             -0.0082      0.052     -0.158      0.874
-0.110       0.093
RoofStyle[T.Gambrel]           -0.0521      0.058     -0.904      0.366
-0.165       0.061
RoofStyle[T.Hip]               -0.0008      0.052     -0.015      0.988
-0.103       0.101
RoofStyle[T.Mansard]           -0.1099      0.064     -1.711      0.087
-0.236       0.016
RoofStyle[T.Shed]              -0.0457      0.081     -0.563      0.574
-0.205       0.113
SaleCondition[T.AdjLand]        0.2002      0.043      4.658      0.000
0.116       0.285
SaleCondition[T.Alloca]         0.0725      0.034      2.127      0.034
0.006       0.139
SaleCondition[T.Family]         0.0227      0.018      1.233      0.218
-0.013       0.059
SaleCondition[T.Normal]         0.0390      0.010      3.881      0.000
0.019       0.059
SaleCondition[T.Partial]        0.0961      0.046      2.077      0.038
0.005       0.187
SaleType[T.CWD]                 0.0156      0.029      0.539      0.590
-0.041       0.072
SaleType[T.Con]                 0.0544      0.054      1.007      0.314
-0.052       0.160
SaleType[T.ConLD]               0.0277      0.031      0.898      0.369
-0.033       0.088
SaleType[T.ConLI]              -0.2024      0.057     -3.531      0.000
-0.315      -0.090
SaleType[T.ConLw]              -0.1059      0.048     -2.196      0.028
-0.201      -0.011
SaleType[T.New]                -0.0274      0.048     -0.567      0.571
-0.122       0.067
SaleType[T.Oth]                 0.1478      0.088      1.678      0.094
-0.025       0.321
SaleType[T.VWD]                -0.0313      0.090     -0.348      0.728
-0.208       0.145
SaleType[T.WD ]                 0.0034      0.014      0.245      0.806
-0.024       0.031
Street[T.Pave]                  0.0204      0.044      0.460      0.645
-0.067       0.107
TotRmsAbvGrd[T.4]               0.0211      0.032      0.670      0.503
-0.041       0.083
TotRmsAbvGrd[T.5]               0.0158      0.032      0.497      0.619
-0.046       0.078
TotRmsAbvGrd[T.6]               0.0185      0.033      0.568      0.570
-0.045       0.082
TotRmsAbvGrd[T.7]               0.0113      0.034      0.337      0.736
-0.055       0.077
```

```
TotRmsAbvGrd[T.8]              0.0164      0.035      0.474      0.636
-0.052       0.084
TotRmsAbvGrd[T.9]             -0.0057      0.036     -0.158      0.874
-0.077       0.065
TotRmsAbvGrd[T.10]            -0.0178      0.039     -0.462      0.644
-0.093       0.058
TotRmsAbvGrd[T.11]            -0.0326      0.042     -0.768      0.443
-0.116       0.051
TotRmsAbvGrd[T.12]             0.0897      0.076      1.180      0.238
-0.059       0.239
TotRmsAbvGrd[T.13]         -3.155e-16   9.89e-17     -3.189      0.001
-5.1e-16   -1.21e-16
Utilities[T.NoSeWa]         2.269e-16   1.48e-16      1.532      0.126
-6.37e-17    5.17e-16
Utilities[T.NoSewr]           -0.1244      0.137     -0.907      0.364
-0.394       0.145
Artery                        -0.0918      0.015     -6.155      0.000
-0.121      -0.063
AsbShng                       -0.0204      0.024     -0.844      0.399
-0.068       0.027
AsphShn                        0.1508      0.068      2.231      0.026
0.018       0.283
BrkCmn                         0.0462      0.062      0.747      0.455
-0.075       0.167
BrkComm                        0.1891      0.066      2.877      0.004
0.060       0.318
BrkFace                        0.0648      0.015      4.445      0.000
0.036       0.093
BsmtFinSF1                   9.94e-05   2.54e-05      3.915      0.000
4.96e-05       0.000
BsmtFinSF1_p2              -3.784e-08   1.73e-08     -2.187      0.029
-7.18e-08   -3.89e-09
BsmtFinSF2                 -1.364e-06   5.17e-05     -0.026      0.979
-0.000       0.000
BsmtFinSF2_p2               4.068e-08   5.78e-08      0.704      0.481
-7.26e-08    1.54e-07
BsmtUnfSF                  -3.462e-05    2.4e-05     -1.444      0.149
-8.17e-05    1.24e-05
BsmtUnfSF_p2                8.152e-09   1.12e-08      0.729      0.466
-1.38e-08    3.01e-08
CBlock                         2.7422      0.110     24.878      0.000
2.526       2.958
CemntBd                       -0.0743      0.065     -1.142      0.254
-0.202       0.053
CmentBd                        0.0959      0.065      1.470      0.142
-0.032       0.224
EnclosedPorch              -9.118e-05   6.25e-05     -1.458      0.145
-0.000    3.15e-05
```

```
EnclosedPorch_p2              4.909e-07   1.46e-07      3.369      0.001
2.05e-07    7.77e-07
Feedr                           -0.0555      0.011     -5.251      0.000
-0.076      -0.035
GarageArea                   -4.593e-06   8.08e-05     -0.057      0.955
-0.000       0.000
GarageArea_p2                 5.741e-08   5.91e-08      0.971      0.332
-5.86e-08    1.73e-07
GrLivArea                        0.0003   7.98e-05      3.686      0.000
0.000       0.000
GrLivArea_p2                 -3.041e-08   1.26e-08     -2.421      0.016
-5.5e-08   -5.77e-09
HdBoard                          0.0004      0.012      0.034      0.973
-0.023       0.024
ImStucc                         -0.0246      0.030     -0.834      0.404
-0.083       0.033
LotArea                       6.268e-06   8.91e-07      7.035      0.000
4.52e-06    8.02e-06
LotArea_p2                   -3.289e-11   6.34e-12     -5.186      0.000
-4.53e-11   -2.05e-11
LotFrontage                   -7.06e-06      0.000     -0.047      0.962
-0.000       0.000
LotFrontage_p2                5.364e-07   1.16e-06      0.464      0.643
-1.73e-06    2.81e-06
LowQualFinSF                  2.652e-05      0.000      0.119      0.905
-0.000       0.000
LowQualFinSF_p2              -7.349e-08   5.25e-07     -0.140      0.889
-1.1e-06    9.57e-07
MasVnrArea                    9.059e-05   4.38e-05      2.069      0.039
4.73e-06       0.000
MasVnrArea_p2                -4.991e-08      4e-08     -1.249      0.212
-1.28e-07    2.85e-08
MetalSd                          0.0174      0.012      1.416      0.157
-0.007       0.042
MiscVal                       2.305e-05   2.05e-05      1.126      0.260
-1.71e-05    6.32e-05
MiscVal_p2                   -9.851e-10   1.21e-09     -0.814      0.416
-3.36e-09    1.39e-09
Norm                             0.0571      0.026      2.196      0.028
0.006       0.108
OpenPorchSF                   8.498e-05   7.67e-05      1.108      0.268
-6.55e-05       0.000
OpenPorchSF_p2                1.204e-07   2.54e-07      0.474      0.636
-3.78e-07    6.19e-07
Other                           -0.0595      0.087     -0.683      0.494
-0.230       0.111
Plywood                          0.0006      0.011      0.052      0.959
-0.022       0.023
```

```
PoolArea                        -0.0007      0.000     -2.091      0.037
-0.001   -4.32e-05
PosA                             0.0471      0.027      1.773      0.076
-0.005       0.099
PosN                             0.0237      0.020      1.164      0.245
-0.016       0.064
PreCast                      -3.194e-19   1.39e-17     -0.023      0.982
-2.75e-17    2.69e-17
RRAe                            -0.0554      0.023     -2.460      0.014
-0.100      -0.011
RRAn                            -0.0160      0.016     -0.989      0.323
-0.048       0.016
RRNe                            -0.0211      0.046     -0.456      0.649
-0.112       0.070
RRNn                             0.0639      0.041      1.559      0.119
-0.016       0.144
ScreenPorch                      0.0003   9.24e-05      3.149      0.002
0.000       0.000
ScreenPorch_p2               -4.331e-07   3.31e-07     -1.310      0.190
-1.08e-06    2.16e-07
Stone                           -0.0021      0.047     -0.044      0.965
-0.095       0.091
Stucco                           0.0205      0.019      1.108      0.268
-0.016       0.057
TotalBsmtSF                   6.341e-05   3.41e-05      1.859      0.063
-3.49e-06       0.000
TotalBsmtSF_p2                3.152e-09   1.53e-08      0.206      0.837
-2.69e-08    3.32e-08
VinylSd                          0.0059      0.013      0.457      0.648
-0.019       0.031
WdSdng                           0.0088      0.012      0.727      0.467
-0.015       0.033
WdShing                          0.0061      0.022      0.278      0.781
-0.037       0.049
WdShng                           0.0041      0.018      0.231      0.818
-0.031       0.039
WoodDeckSF                    8.018e-05   3.14e-05      2.550      0.011
1.85e-05       0.000
WoodDeckSF_p2                -7.926e-08   5.42e-08     -1.462      0.144
-1.86e-07    2.71e-08
X1stFlrSF                        0.0002   8.13e-05      2.313      0.021
2.86e-05       0.000
X1stFlrSF_p2                 -2.878e-08    1.4e-08     -2.059      0.040
-5.62e-08   -1.37e-09
X2ndFlrSF                     7.968e-05   7.94e-05      1.004      0.315
-7.6e-05       0.000
X2ndFlrSF_p2                  1.574e-08   2.92e-08      0.540      0.589
-4.15e-08    7.29e-08
```

```
X3SsnPorch                     8.546e-05      0.000      0.292      0.770
-0.000       0.001
X3SsnPorch_p2                  4.591e-07   1.04e-06      0.443      0.658
-1.57e-06    2.49e-06
YearsSince1950Built            5.089e-05      0.001      0.036      0.972
-0.003       0.003
YearsSince1950Built_p2         6.621e-05   2.77e-05      2.393      0.017
1.19e-05       0.000
YearsSince1950GarageBuilt         0.0006      0.001      0.528      0.597
-0.002       0.003
YearsSince1950GarageBuilt_p2 -1.493e-05   2.13e-05     -0.700      0.484
-5.68e-05    2.69e-05
YearsSince1950Remod               0.0015      0.001      1.889      0.059
-5.77e-05       0.003
YearsSince1950Remod_p2        -2.076e-05   1.44e-05     -1.444      0.149
-4.9e-05    7.45e-06
==============================================================================
Omnibus:                      303.264   Durbin-Watson:                   1.912
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2249.910
Skew:                          -0.564   Prob(JB):                         0.00
Kurtosis:                       8.317   Cond. No.                     4.48e+17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 8.45e-15. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

[ ]:

```
print_metrics(nlr_poly, ames_train_poly, y_train, ames_test_poly, y_test,␣
 ↪flag_log_sale_price = True)
print_metrics(nlr_poly, ames_train_poly, y_train, ames_test_poly, y_test,␣
 ↪flag_log_sale_price = False)
```

```
Metrics for Log(Sale Price):

Training R2 0.9624105774575185
Training MAE 0.05428713187831344
Training RMSE 0.07504602800253764
Out-of-sample R2 0.8389618470389943
Out-of-sample MAE 0.08377525015006256
Out-of-sample RMSE 0.15893909023513655

Metrics for Sale Price:

Training R2 0.9640398351503799
Training MAE 9959.9415261308
```

```
Training RMSE 14814.210314147367
Out-of-sample R2 0.42824990275266905
Out-of-sample MAE 15888.470959526549
Out-of-sample RMSE 59264.318499926754
```

Pay attention to the test set performance.

6.1 2.2 "Common Sense" Linear Regression

We want to perform linear regression with only some of the variables which could be chosen using common sense

[ ]:

```
ames_train[['MSSubClass',␣
 ↪'OverallQual','YearsSince1950Built','GrLivArea','TotRmsAbvGrd','FullBath','SaleCondition']]
```

[ ]:

```
     MSSubClass OverallQual  YearsSince1950Built  GrLivArea TotRmsAbvGrd  \
989          20           6                    8     1414.0            6
990          20           6                   20     2117.0            7
991          60           6                   19     2112.0            9
992          60           6                   47     1604.0            7
993          60           6                   45     1470.0            7
…         …         …                  …        …          …
2924         20           5                   10     1224.0            7
2925         80           6                   34     1003.0            6
2926         20           5                   33      902.0            5
2928         20           5                   24     1389.0            6
2929         60           7                   43     2000.0            9

     FullBath SaleCondition
989         1        Normal
990         2        Normal
991         2        Normal
992         2        Normal
993         2        Normal
…       …           …
2924        1       Abnorml
2925        1        Normal
2926        1        Normal
2928        1        Normal
2929        2        Normal

[1828 rows x 7 columns]
```

[ ]:

```
ames_train_cs = ames_train[['LogSalePrice','MSSubClass',␣
 ↪'OverallQual','YearsSince1950Built','GrLivArea','TotRmsAbvGrd','FullBath','SaleCondition']]
ames_train_cs = ames_train_cs.copy()
ames_train_cs['OverallQual'] = ames_train_cs['OverallQual'].astype('float')
ames_train_cs['TotRmsAbvGrd'] = ames_train_cs['TotRmsAbvGrd'].astype('int')
```

```python
ames_train_cs['FullBath'] = ames_train_cs['FullBath'].astype('int')


ames_test_cs = ames_test[['LogSalePrice','MSSubClass',␣
 ↪'OverallQual','YearsSince1950Built','GrLivArea','TotRmsAbvGrd','FullBath','SaleCondition']]
ames_test_cs = ames_test_cs.copy()
ames_test_cs['OverallQual'] = ames_test_cs['OverallQual'].astype('float')
ames_test_cs['TotRmsAbvGrd'] = ames_test_cs['TotRmsAbvGrd'].astype('int')
ames_test_cs['FullBath'] = ames_test_cs['FullBath'].astype('int')


print(ames_train_cs.shape)
```

```
(1828, 8)
```

[ ]:

```python
all_columns = "+".join(ames_train_cs.columns.difference(["LogSalePrice"]))
my_formula = "LogSalePrice~" + all_columns +'-1'
print(my_formula)

mod_commonsense = smf.ols(my_formula, data=ames_train_cs)
lr_cs = mod_commonsense.fit()

print(lr_cs.summary())
```

```
LogSalePrice~FullBath+GrLivArea+MSSubClass+OverallQual+SaleCondition+TotRmsAbvGr
d+YearsSince1950Built-1
                            OLS Regression Results
==============================================================================
Dep. Variable:           LogSalePrice   R-squared:                       0.859
Model:                            OLS   Adj. R-squared:                  0.857
Method:                 Least Squares   F-statistic:                     437.4
Date:                Fri, 01 Nov 2024   Prob (F-statistic):               0.00
Time:                        18:18:57   Log-Likelihood:                 928.68
No. Observations:                1828   AIC:                            -1805.
Df Residuals:                    1802   BIC:                            -1662.
Df Model:                          25
Covariance Type:            nonrobust
================================================================================
============
                               coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
------------
MSSubClass[20]              10.7050      0.027    400.469      0.000      10.653
10.757
MSSubClass[30]              10.5387      0.031    342.113      0.000      10.478
10.599
MSSubClass[40]              10.5311      0.078    134.816      0.000      10.378
```

```
10.684
MSSubClass[45]               10.6019      0.051    207.974      0.000      10.502
10.702
MSSubClass[50]               10.5715      0.030    352.890      0.000      10.513
10.630
MSSubClass[60]               10.5596      0.031    342.320      0.000      10.499
10.620
MSSubClass[70]               10.5422      0.034    309.101      0.000      10.475
10.609
MSSubClass[75]               10.4234      0.048    216.125      0.000      10.329
10.518
MSSubClass[80]               10.6874      0.032    338.392      0.000      10.625
10.749
MSSubClass[85]               10.7808      0.038    285.581      0.000      10.707
10.855
MSSubClass[90]               10.5422      0.037    286.747      0.000      10.470
10.614
MSSubClass[120]              10.6443      0.030    355.652      0.000      10.586
10.703
MSSubClass[150]              10.3006      0.150     68.787      0.000      10.007
10.594
MSSubClass[160]              10.3885      0.031    338.192      0.000      10.328
10.449
MSSubClass[180]              10.5393      0.057    186.112      0.000      10.428
10.650
MSSubClass[190]              10.5839      0.037    282.665      0.000      10.510
10.657
SaleCondition[T.AdjLand]      0.0346      0.063      0.553      0.581      -0.088
0.157
SaleCondition[T.Alloca]       0.1598      0.047      3.403      0.001       0.068
0.252
SaleCondition[T.Family]       0.0338      0.029      1.153      0.249      -0.024
0.091
SaleCondition[T.Normal]       0.0990      0.015      6.711      0.000       0.070
0.128
SaleCondition[T.Partial]      0.1320      0.019      6.869      0.000       0.094
0.170
FullBath                     -0.0231      0.010     -2.403      0.016      -0.042
-0.004
GrLivArea                     0.0004   1.53e-05     28.971      0.000       0.000
0.000
OverallQual                   0.1159      0.004     27.329      0.000       0.108
0.124
TotRmsAbvGrd                 -0.0163      0.004     -3.792      0.000      -0.025
-0.008
YearsSince1950Built           0.0034      0.000     10.515      0.000       0.003
0.004
===============================================================================
```

```
Omnibus:                      633.662   Durbin-Watson:                   1.749
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             8959.171
Skew:                          -1.225   Prob(JB):                         0.00
Kurtosis:                      13.565   Cond. No.                     7.09e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 7.09e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
[ ]: print_metrics(lr_cs, ames_train_cs, y_train, ames_test_cs, y_test,␣
      ↪flag_log_sale_price = True)
     print_metrics(lr_cs, ames_train_cs, y_train, ames_test_cs, y_test,␣
      ↪flag_log_sale_price = False)
```

```
Metrics for Log(Sale Price):

Training R2 0.8585288469181491
Training MAE 0.10592319141717338
Training RMSE 0.14558908973890214
Out-of-sample R2 0.8347117625903195
Out-of-sample MAE 0.11370443708303558
Out-of-sample RMSE 0.1610227749435649

Metrics for Sale Price:

Training R2 0.8752645912586368
Training MAE 19392.02270765631
Training RMSE 27590.689019810863
Out-of-sample R2 0.8672838314855108
Out-of-sample MAE 19917.235925224228
Out-of-sample RMSE 28553.022575248742
```

```
[ ]:
```

```
[ ]:
```

7 Part 2: Model refinement

7.1 1. Principal Components Regression

We first remove columns that are nearly constant, i.e., have small standard deviation. Then we use sklearn grid-search for cross validation and finally retrain the final model.

[ ]:
```
X_train_poly = ames_train_poly.drop(columns='LogSalePrice')
X_test_poly = ames_test_poly.drop(columns='LogSalePrice')

X_train_poly_wide = pd.get_dummies(X_train_poly)
X_test_poly_wide = pd.get_dummies(X_test_poly)
```

[ ]:
```
y_train = ames_train['LogSalePrice']
y_test = ames_test['LogSalePrice']

X_train_pcr = X_train_poly_wide.loc[:, X_train_poly_wide.std() > 0.1]
X_test_pcr = X_test_poly_wide[X_train_pcr.columns]

print(X_train_poly_wide.shape, X_train_pcr.shape)
print(X_test_poly_wide.shape, X_test_pcr.shape)
```

```
(1828, 397) (1828, 275)
(937, 397) (937, 275)
```

We also standardize the data before feeding it to the PCA step, as recommended by good practice.

[ ]:
```
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.linear_model import LinearRegression
from sklearn.pipeline import Pipeline
from sklearn.model_selection import GridSearchCV

scaler = StandardScaler()
pca = PCA(n_components=5, random_state=88)
lr = LinearRegression()
pipe = Pipeline(steps=[('scaler', scaler), ('pca', pca), ('lr', lr)])

pipe.fit(X_train_pcr, y_train)
```

[ ]:
```
Pipeline(steps=[('scaler', StandardScaler()),
                ('pca', PCA(n_components=5, random_state=88)),
                ('lr', LinearRegression())])
```

[ ]:
```
print_metrics(pipe, X_train_pcr, y_train, X_test_pcr, y_test,␣
  ↪flag_log_sale_price = True)
print_metrics(pipe, X_train_pcr, y_train, X_test_pcr, y_test,␣
  ↪flag_log_sale_price = False)
```

```

Metrics for Log(Sale Price):

Training R2 0.8567244094773699
Training MAE 0.10108699422106086
Training RMSE 0.14651462829012946
Out-of-sample R2 0.8355568543382658
```

```
Out-of-sample MAE 0.10530967892420574
Out-of-sample RMSE 0.16061060586244827

Metrics for Sale Price:

Training R2 0.8927975509686596
Training MAE 17757.429598137704
Training RMSE 25578.202564553925
Out-of-sample R2 0.8987227159041019
Out-of-sample MAE 17636.54874599874
Out-of-sample RMSE 24942.856856990882
```

[ ]:

```

```

7.2 2. Ridge Regression

We can choose `alpha_max` so as the value that makes all coefficientes zero, and then construct a log sequence of `alpha` values trending smaller, decreasing the degree of regularization.

For the case of `Ridge` Regression, alpha value that would make all coefficients zero would be `Inf`, however we can be satisfied with sufficiently small numbers, and work from there.

[ ]:

```python
X_train_rr = X_train_poly_wide
X_test_rr = X_test_poly_wide

print(X_train_rr.shape, X_test_rr.shape)
```

```
(1828, 397) (937, 397)
```

7.2.1 Determine 'alpha_max'

[ ]:

```python
from sklearn.linear_model import Ridge

alpha_max = 10**5
rr = Ridge(alpha=alpha_max, random_state=88)
rr.fit(X_train_rr, y_train)
```

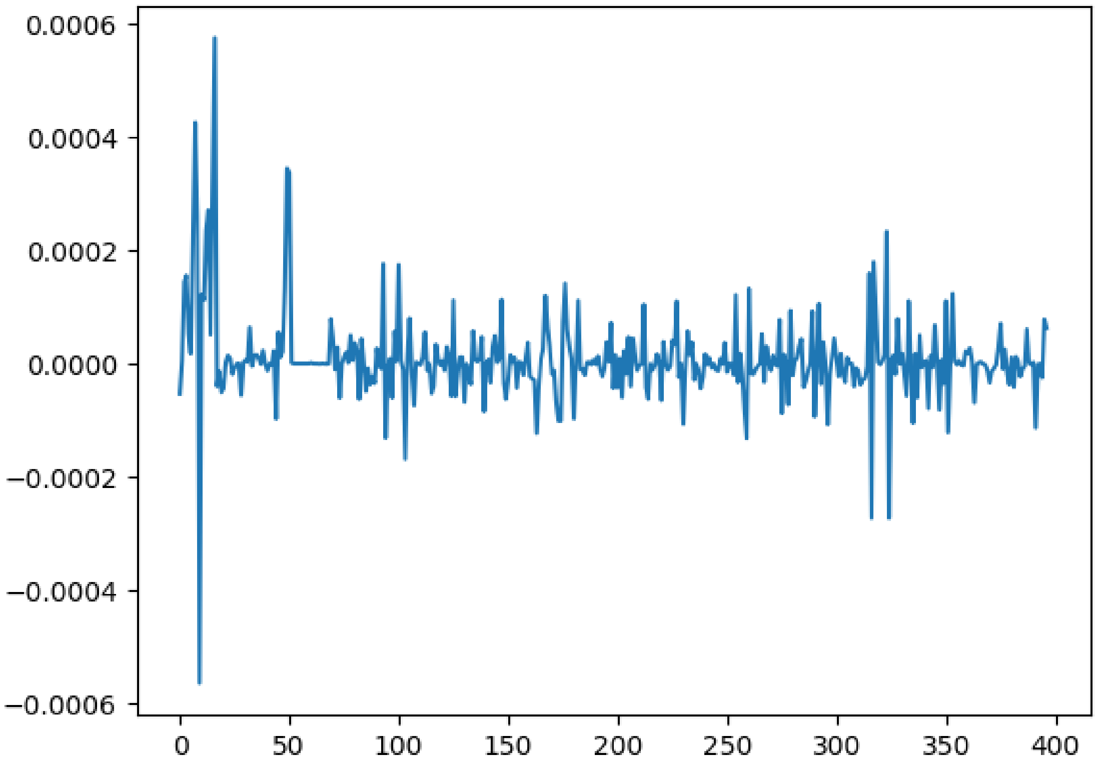
```
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_ridge.py:216: LinAlgWarning: Ill-conditioned
matrix (rcond=5.87405e-17): result may not be accurate.
  return linalg.solve(A, Xy, assume_a="pos", overwrite_a=True).T
```

[ ]:

```
Ridge(alpha=100000, random_state=88)
```

[ ]:

```python
plt.plot(rr.coef_)
plt.show()
print(max(abs(rr.coef_)))
```



```
0.0005755966628195446
```

[ ]:

```python
print_metrics(rr, X_train_rr, y_train, X_test_rr, y_test, flag_log_sale_price = True)
print_metrics(rr, X_train_rr, y_train, X_test_rr, y_test, flag_log_sale_price = False)
```

```
Metrics for Log(Sale Price):

Training R2 0.8452590849089253
Training MAE 0.10467969125067748
Training RMSE 0.1522640805932089
Out-of-sample R2 0.8229399441245389
Out-of-sample MAE 0.11151022706827869
Out-of-sample RMSE 0.16665817850890513

Metrics for Sale Price:

Training R2 0.878237976891266
Training MAE 18648.76164916895
Training RMSE 27259.85846771404
Out-of-sample R2 0.8695106549025007
```

```
Out-of-sample MAE 19389.607748338716
Out-of-sample RMSE 28312.465936142868
```

[ ]:

```

```

7.3 3. Lasso Regression

[ ]:

```python
from sklearn.linear_model import Lasso
```

[ ]:

```python
X_train_lasso = X_train_poly_wide
X_test_lasso = X_test_poly_wide

print(X_train_lasso.shape, X_test_lasso.shape)
```

```
(1828, 397) (937, 397)
```

7.3.1 Lasso Coefficients vs. Degree of Regularization

[ ]:

```python
# Question: What is the best value of lambda in Lasso?
```

7.3.2 Lasso Hyper-parameter Tuning

[ ]:

```python
alphas = np.logspace(-5, 1 , num=50, base=10)
coefs = []
```

[ ]:

```python
from sklearn.linear_model import Lasso

for a in alphas:
    lasso = Lasso(alpha=a, random_state=88)
    lasso.fit(X_train_lasso, y_train)
    coefs.append(lasso.coef_)

plt.figure(figsize=(12, 10))
ax = plt.gca()
ax.plot(alphas, coefs)
ax.set_xscale('log')
plt.xlabel('Alpha')
plt.ylabel('Coefficients')
plt.title('Lasso coefficients as a function of the regularization')
plt.show()
```

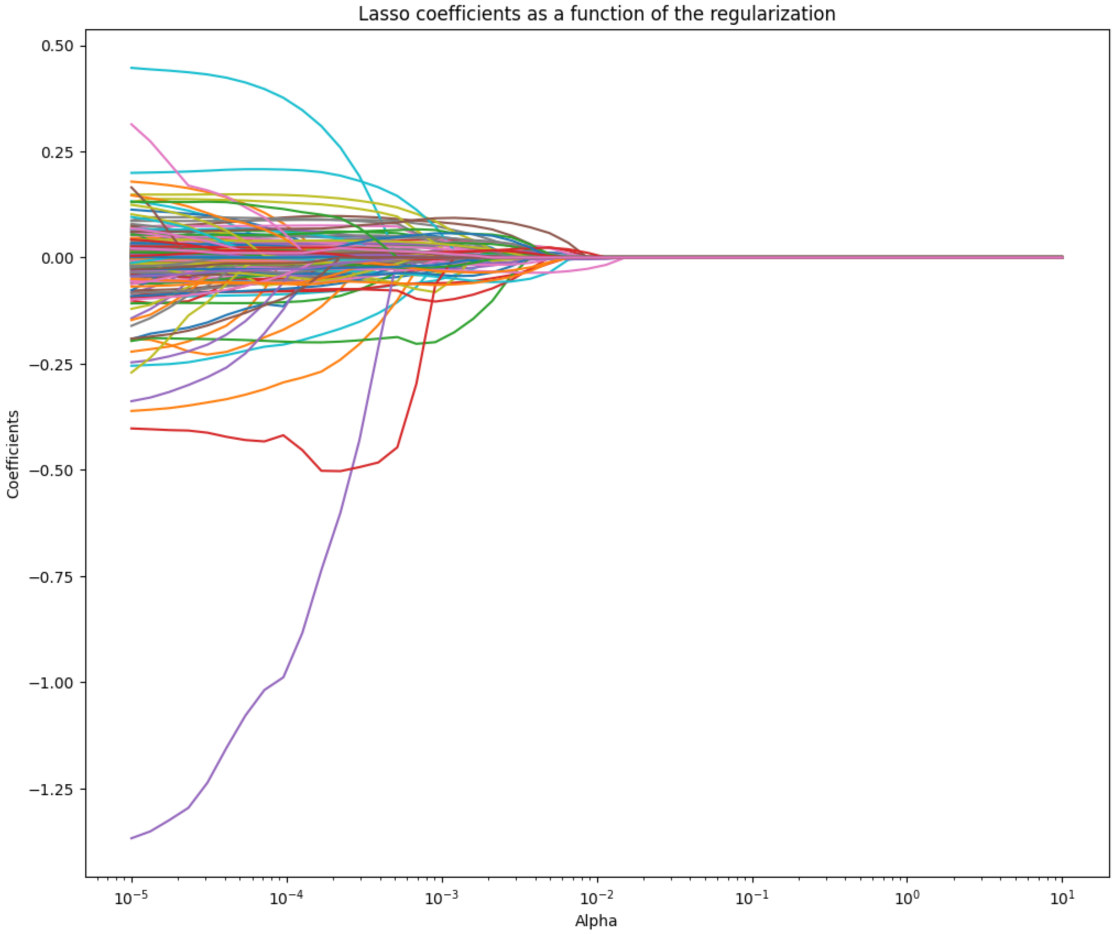
```
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 5.418e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
```

```
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 5.499e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 5.600e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 5.727e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 5.885e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 6.079e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 6.317e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 6.609e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 6.965e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
```

```
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 7.397e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 7.914e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 8.532e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 9.266e+00, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 1.012e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 1.106e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 1.209e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 1.317e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
```

```
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 1.431e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 1.548e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 1.491e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 1.354e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 1.889e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 2.145e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 2.127e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 3.411e-01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
```

```
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 2.100e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 2.168e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 2.190e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 2.220e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
```



7.3.3 Selected Variables

Let's look at some of the variables selected by lasso

[ ]:
```
lasso = Lasso(alpha=0.1, random_state=88)
lasso.fit(X_train_lasso, y_train)
```

```
/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-
packages/sklearn/linear_model/_coordinate_descent.py:697: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations,
check the scale of the features or consider increasing regularisation. Duality
gap: 2.157e+01, tolerance: 2.739e-02
  model = cd_fast.enet_coordinate_descent(
```

[ ]:
```
Lasso(alpha=0.1, random_state=88)
```

[ ]:
```
cols = X_train_lasso.columns
coefs = lasso.coef_
sorted(zip(abs(coefs), cols))
```

```
[ ]: [(0.0, 'Alley_Grvl'),
      (0.0, 'Alley_NoAccess'),
      (0.0, 'Alley_Pave'),
      (0.0, 'Artery'),
      (0.0, 'AsbShng'),
      (0.0, 'AsphShn'),
      (0.0, 'BedroomAbvGr_0'),
      (0.0, 'BedroomAbvGr_1'),
      (0.0, 'BedroomAbvGr_2'),
      (0.0, 'BedroomAbvGr_3'),
      (0.0, 'BedroomAbvGr_4'),
      (0.0, 'BedroomAbvGr_5'),
      (0.0, 'BedroomAbvGr_6'),
      (0.0, 'BldgType_1Fam'),
      (0.0, 'BldgType_2fmCon'),
      (0.0, 'BldgType_Duplex'),
      (0.0, 'BldgType_Twnhs'),
      (0.0, 'BldgType_TwnhsE'),
      (0.0, 'BrkCmn'),
      (0.0, 'BrkComm'),
      (0.0, 'BrkFace'),
      (0.0, 'BsmtCond_Ex'),
      (0.0, 'BsmtCond_Fa'),
      (0.0, 'BsmtCond_Gd'),
      (0.0, 'BsmtCond_NaN'),
      (0.0, 'BsmtCond_NoBasement'),
      (0.0, 'BsmtCond_Po'),
      (0.0, 'BsmtCond_TA'),
      (0.0, 'BsmtExposure_Av'),
      (0.0, 'BsmtExposure_Gd'),
      (0.0, 'BsmtExposure_Mn'),
      (0.0, 'BsmtExposure_NaN'),
      (0.0, 'BsmtExposure_No'),
      (0.0, 'BsmtExposure_NoBasement'),
      (0.0, 'BsmtFinType1_ALQ'),
      (0.0, 'BsmtFinType1_BLQ'),
      (0.0, 'BsmtFinType1_GLQ'),
      (0.0, 'BsmtFinType1_LwQ'),
      (0.0, 'BsmtFinType1_NaN'),
      (0.0, 'BsmtFinType1_NoBasement'),
      (0.0, 'BsmtFinType1_Rec'),
      (0.0, 'BsmtFinType1_Unf'),
      (0.0, 'BsmtFinType2_ALQ'),
      (0.0, 'BsmtFinType2_BLQ'),
      (0.0, 'BsmtFinType2_GLQ'),
      (0.0, 'BsmtFinType2_LwQ'),
      (0.0, 'BsmtFinType2_NaN'),
```

```
(0.0, 'BsmtFinType2_NoBasement'),
(0.0, 'BsmtFinType2_Rec'),
(0.0, 'BsmtFinType2_Unf'),
(0.0, 'BsmtFullBath_0'),
(0.0, 'BsmtFullBath_1'),
(0.0, 'BsmtFullBath_2'),
(0.0, 'BsmtHalfBath_0'),
(0.0, 'BsmtHalfBath_1'),
(0.0, 'BsmtHalfBath_2'),
(0.0, 'BsmtQual_Ex'),
(0.0, 'BsmtQual_Fa'),
(0.0, 'BsmtQual_Gd'),
(0.0, 'BsmtQual_NaN'),
(0.0, 'BsmtQual_NoBasement'),
(0.0, 'BsmtQual_Po'),
(0.0, 'BsmtQual_TA'),
(0.0, 'BsmtUnfSF'),
(0.0, 'CBlock'),
(0.0, 'CemntBd'),
(0.0, 'CentralAir_N'),
(0.0, 'CentralAir_Y'),
(0.0, 'CmentBd'),
(0.0, 'Electrical_FuseA'),
(0.0, 'Electrical_FuseF'),
(0.0, 'Electrical_FuseP'),
(0.0, 'Electrical_Mix'),
(0.0, 'Electrical_NaN'),
(0.0, 'Electrical_SBrkr'),
(0.0, 'EnclosedPorch'),
(0.0, 'ExterCond_Ex'),
(0.0, 'ExterCond_Fa'),
(0.0, 'ExterCond_Gd'),
(0.0, 'ExterCond_Po'),
(0.0, 'ExterCond_TA'),
(0.0, 'ExterQual_Ex'),
(0.0, 'ExterQual_Fa'),
(0.0, 'ExterQual_Gd'),
(0.0, 'ExterQual_TA'),
(0.0, 'Feedr'),
(0.0, 'Fence_GdPrv'),
(0.0, 'Fence_GdWo'),
(0.0, 'Fence_MnPrv'),
(0.0, 'Fence_MnWw'),
(0.0, 'Fence_NoFence'),
(0.0, 'FireplaceQu_Ex'),
(0.0, 'FireplaceQu_Fa'),
(0.0, 'FireplaceQu_Gd'),
```

```
(0.0, 'FireplaceQu_NoFirePlace'),
(0.0, 'FireplaceQu_Po'),
(0.0, 'FireplaceQu_TA'),
(0.0, 'Fireplaces_0'),
(0.0, 'Fireplaces_1'),
(0.0, 'Fireplaces_2'),
(0.0, 'Fireplaces_3'),
(0.0, 'Fireplaces_4'),
(0.0, 'Foundation_BrkTil'),
(0.0, 'Foundation_CBlock'),
(0.0, 'Foundation_PConc'),
(0.0, 'Foundation_Slab'),
(0.0, 'Foundation_Stone'),
(0.0, 'Foundation_Wood'),
(0.0, 'FullBath_0'),
(0.0, 'FullBath_1'),
(0.0, 'FullBath_2'),
(0.0, 'FullBath_3'),
(0.0, 'FullBath_4'),
(0.0, 'Functional_Maj1'),
(0.0, 'Functional_Maj2'),
(0.0, 'Functional_Min1'),
(0.0, 'Functional_Min2'),
(0.0, 'Functional_Mod'),
(0.0, 'Functional_Sal'),
(0.0, 'Functional_Sev'),
(0.0, 'Functional_Typ'),
(0.0, 'GarageCars_1'),
(0.0, 'GarageCars_2'),
(0.0, 'GarageCars_3'),
(0.0, 'GarageCars_4'),
(0.0, 'GarageCars_5'),
(0.0, 'GarageCond_Ex'),
(0.0, 'GarageCond_Fa'),
(0.0, 'GarageCond_Gd'),
(0.0, 'GarageCond_Po'),
(0.0, 'GarageCond_TA'),
(0.0, 'GarageFinish_Fin'),
(0.0, 'GarageFinish_RFn'),
(0.0, 'GarageFinish_Unf'),
(0.0, 'GarageQual_Ex'),
(0.0, 'GarageQual_Fa'),
(0.0, 'GarageQual_Gd'),
(0.0, 'GarageQual_Po'),
(0.0, 'GarageQual_TA'),
(0.0, 'GarageType_2Types'),
(0.0, 'GarageType_Attchd'),
```

```
(0.0, 'GarageType_Basment'),
(0.0, 'GarageType_BuiltIn'),
(0.0, 'GarageType_CarPort'),
(0.0, 'GarageType_Detchd'),
(0.0, 'HalfBath_0'),
(0.0, 'HalfBath_1'),
(0.0, 'HalfBath_2'),
(0.0, 'HdBoard'),
(0.0, 'HeatingQC_Ex'),
(0.0, 'HeatingQC_Fa'),
(0.0, 'HeatingQC_Gd'),
(0.0, 'HeatingQC_Po'),
(0.0, 'HeatingQC_TA'),
(0.0, 'Heating_Floor'),
(0.0, 'Heating_GasA'),
(0.0, 'Heating_GasW'),
(0.0, 'Heating_Grav'),
(0.0, 'Heating_OthW'),
(0.0, 'Heating_Wall'),
(0.0, 'HouseStyle_1.5Fin'),
(0.0, 'HouseStyle_1.5Unf'),
(0.0, 'HouseStyle_1Story'),
(0.0, 'HouseStyle_2.5Fin'),
(0.0, 'HouseStyle_2.5Unf'),
(0.0, 'HouseStyle_2Story'),
(0.0, 'HouseStyle_SFoyer'),
(0.0, 'HouseStyle_SLvl'),
(0.0, 'ImStucc'),
(0.0, 'KitchenAbvGr_0'),
(0.0, 'KitchenAbvGr_1'),
(0.0, 'KitchenAbvGr_2'),
(0.0, 'KitchenAbvGr_3'),
(0.0, 'KitchenQual_Ex'),
(0.0, 'KitchenQual_Fa'),
(0.0, 'KitchenQual_Gd'),
(0.0, 'KitchenQual_Po'),
(0.0, 'KitchenQual_TA'),
(0.0, 'LandContour_Bnk'),
(0.0, 'LandContour_HLS'),
(0.0, 'LandContour_Low'),
(0.0, 'LandContour_Lvl'),
(0.0, 'LandSlope_Gtl'),
(0.0, 'LandSlope_Mod'),
(0.0, 'LandSlope_Sev'),
(0.0, 'LotConfig_Corner'),
(0.0, 'LotConfig_CulDSac'),
(0.0, 'LotConfig_FR2'),
```

```
(0.0, 'LotConfig_FR3'),
(0.0, 'LotConfig_Inside'),
(0.0, 'LotFrontage'),
(0.0, 'LotShape_IR1'),
(0.0, 'LotShape_IR2'),
(0.0, 'LotShape_IR3'),
(0.0, 'LotShape_Reg'),
(0.0, 'LowQualFinSF'),
(0.0, 'MSSubClass_120'),
(0.0, 'MSSubClass_150'),
(0.0, 'MSSubClass_160'),
(0.0, 'MSSubClass_180'),
(0.0, 'MSSubClass_190'),
(0.0, 'MSSubClass_20'),
(0.0, 'MSSubClass_30'),
(0.0, 'MSSubClass_40'),
(0.0, 'MSSubClass_45'),
(0.0, 'MSSubClass_50'),
(0.0, 'MSSubClass_60'),
(0.0, 'MSSubClass_70'),
(0.0, 'MSSubClass_75'),
(0.0, 'MSSubClass_80'),
(0.0, 'MSSubClass_85'),
(0.0, 'MSSubClass_90'),
(0.0, 'MSZoning_A'),
(0.0, 'MSZoning_C'),
(0.0, 'MSZoning_FV'),
(0.0, 'MSZoning_I'),
(0.0, 'MSZoning_RH'),
(0.0, 'MSZoning_RL'),
(0.0, 'MSZoning_RM'),
(0.0, 'MasVnrType_BrkCmn'),
(0.0, 'MasVnrType_BrkFace'),
(0.0, 'MasVnrType_CBlock'),
(0.0, 'MasVnrType_NaN'),
(0.0, 'MasVnrType_Stone'),
(0.0, 'MetalSd'),
(0.0, 'MiscFeature_Gar2'),
(0.0, 'MiscFeature_NaN'),
(0.0, 'MiscFeature_Othr'),
(0.0, 'MiscFeature_Shed'),
(0.0, 'MiscFeature_TenC'),
(0.0, 'MoSold_1'),
(0.0, 'MoSold_10'),
(0.0, 'MoSold_11'),
(0.0, 'MoSold_12'),
(0.0, 'MoSold_2'),
```

```
(0.0, 'MoSold_3'),
(0.0, 'MoSold_4'),
(0.0, 'MoSold_5'),
(0.0, 'MoSold_6'),
(0.0, 'MoSold_7'),
(0.0, 'MoSold_8'),
(0.0, 'MoSold_9'),
(0.0, 'Neighborhood_Blmngtn'),
(0.0, 'Neighborhood_Blueste'),
(0.0, 'Neighborhood_BrDale'),
(0.0, 'Neighborhood_BrkSide'),
(0.0, 'Neighborhood_ClearCr'),
(0.0, 'Neighborhood_CollgCr'),
(0.0, 'Neighborhood_Crawfor'),
(0.0, 'Neighborhood_Edwards'),
(0.0, 'Neighborhood_Gilbert'),
(0.0, 'Neighborhood_Greens'),
(0.0, 'Neighborhood_GrnHill'),
(0.0, 'Neighborhood_IDOTRR'),
(0.0, 'Neighborhood_Landmrk'),
(0.0, 'Neighborhood_MeadowV'),
(0.0, 'Neighborhood_Mitchel'),
(0.0, 'Neighborhood_NAmes'),
(0.0, 'Neighborhood_NPkVill'),
(0.0, 'Neighborhood_NWAmes'),
(0.0, 'Neighborhood_NoRidge'),
(0.0, 'Neighborhood_NridgHt'),
(0.0, 'Neighborhood_OldTown'),
(0.0, 'Neighborhood_SWISU'),
(0.0, 'Neighborhood_Sawyer'),
(0.0, 'Neighborhood_SawyerW'),
(0.0, 'Neighborhood_Somerst'),
(0.0, 'Neighborhood_StoneBr'),
(0.0, 'Neighborhood_Timber'),
(0.0, 'Neighborhood_Veenker'),
(0.0, 'Norm'),
(0.0, 'Other'),
(0.0, 'OverallCond_1'),
(0.0, 'OverallCond_2'),
(0.0, 'OverallCond_3'),
(0.0, 'OverallCond_4'),
(0.0, 'OverallCond_5'),
(0.0, 'OverallCond_6'),
(0.0, 'OverallCond_7'),
(0.0, 'OverallCond_8'),
(0.0, 'OverallCond_9'),
(0.0, 'OverallQual_1'),
```

```
(0.0, 'OverallQual_10'),
(0.0, 'OverallQual_2'),
(0.0, 'OverallQual_3'),
(0.0, 'OverallQual_4'),
(0.0, 'OverallQual_5'),
(0.0, 'OverallQual_6'),
(0.0, 'OverallQual_7'),
(0.0, 'OverallQual_8'),
(0.0, 'OverallQual_9'),
(0.0, 'PavedDrive_N'),
(0.0, 'PavedDrive_P'),
(0.0, 'PavedDrive_Y'),
(0.0, 'Plywood'),
(0.0, 'PoolArea'),
(0.0, 'PoolQC_Ex'),
(0.0, 'PoolQC_Fa'),
(0.0, 'PoolQC_Gd'),
(0.0, 'PoolQC_NoPool'),
(0.0, 'PoolQC_TA'),
(0.0, 'PosA'),
(0.0, 'PosN'),
(0.0, 'PreCast'),
(0.0, 'RRAe'),
(0.0, 'RRAn'),
(0.0, 'RRNe'),
(0.0, 'RRNn'),
(0.0, 'RoofMatl_CompShg'),
(0.0, 'RoofMatl_Membran'),
(0.0, 'RoofMatl_Metal'),
(0.0, 'RoofMatl_Roll'),
(0.0, 'RoofMatl_Tar&Grv'),
(0.0, 'RoofMatl_WdShake'),
(0.0, 'RoofMatl_WdShngl'),
(0.0, 'RoofStyle_Flat'),
(0.0, 'RoofStyle_Gable'),
(0.0, 'RoofStyle_Gambrel'),
(0.0, 'RoofStyle_Hip'),
(0.0, 'RoofStyle_Mansard'),
(0.0, 'RoofStyle_Shed'),
(0.0, 'SaleCondition_Abnorml'),
(0.0, 'SaleCondition_AdjLand'),
(0.0, 'SaleCondition_Alloca'),
(0.0, 'SaleCondition_Family'),
(0.0, 'SaleCondition_Normal'),
(0.0, 'SaleCondition_Partial'),
(0.0, 'SaleType_COD'),
(0.0, 'SaleType_CWD'),
```

```
(0.0, 'SaleType_Con'),
(0.0, 'SaleType_ConLD'),
(0.0, 'SaleType_ConLI'),
(0.0, 'SaleType_ConLw'),
(0.0, 'SaleType_New'),
(0.0, 'SaleType_Oth'),
(0.0, 'SaleType_VWD'),
(0.0, 'SaleType_WD '),
(0.0, 'Stone'),
(0.0, 'Street_Grvl'),
(0.0, 'Street_Pave'),
(0.0, 'Stucco'),
(0.0, 'TotRmsAbvGrd_10'),
(0.0, 'TotRmsAbvGrd_11'),
(0.0, 'TotRmsAbvGrd_12'),
(0.0, 'TotRmsAbvGrd_13'),
(0.0, 'TotRmsAbvGrd_3'),
(0.0, 'TotRmsAbvGrd_4'),
(0.0, 'TotRmsAbvGrd_5'),
(0.0, 'TotRmsAbvGrd_6'),
(0.0, 'TotRmsAbvGrd_7'),
(0.0, 'TotRmsAbvGrd_8'),
(0.0, 'TotRmsAbvGrd_9'),
(0.0, 'Utilities_AllPub'),
(0.0, 'Utilities_NoSeWa'),
(0.0, 'Utilities_NoSewr'),
(0.0, 'VinylSd'),
(0.0, 'WdSdng'),
(0.0, 'WdShing'),
(0.0, 'WdShng'),
(0.0, 'X3SsnPorch'),
(0.0, 'YearsSince1950Built'),
(0.0, 'YearsSince1950GarageBuilt'),
(0.0, 'YearsSince1950Remod'),
(2.5331010634587378e-11, 'LotArea_p2'),
(1.3993291221418433e-09, 'MiscVal_p2'),
(1.9822649889359722e-08, 'X1stFlrSF_p2'),
(2.3802701564232124e-08, 'MasVnrArea_p2'),
(2.7992921253098592e-08, 'TotalBsmtSF_p2'),
(3.339096979427708e-08, 'GrLivArea_p2'),
(3.528953608339322e-08, 'GarageArea_p2'),
(4.1452120239835054e-08, 'BsmtUnfSF_p2'),
(4.293104032165341e-08, 'X2ndFlrSF_p2'),
(5.0982227892613884e-08, 'BsmtFinSF2_p2'),
(6.610737913177183e-08, 'BsmtFinSF1_p2'),
(1.4465097086805764e-07, 'WoodDeckSF_p2'),
(2.3501229806679708e-07, 'EnclosedPorch_p2'),
```

```
 (3.014718222992432e-07, 'ScreenPorch_p2'),
 (3.2004683666256624e-07, 'LowQualFinSF_p2'),
 (4.7339943613552793e-07, 'OpenPorchSF_p2'),
 (8.760323298690916e-07, 'X3SsnPorch_p2'),
 (1.3082321119650769e-06, 'LotFrontage_p2'),
 (4.058185623667801e-06, 'YearsSince1950GarageBuilt_p2'),
 (5.12167435997977e-06, 'LotArea'),
 (1.2272984513062219e-05, 'BsmtFinSF2'),
 (1.330405978797038e-05, 'MiscVal'),
 (4.989002825892379e-05, 'YearsSince1950Remod_p2'),
 (8.040984709135124e-05, 'YearsSince1950Built_p2'),
 (8.647420011977135e-05, 'GarageArea'),
 (0.00013939339924155087, 'GrLivArea'),
 (0.0001403572099503, 'MasVnrArea'),
 (0.00014593688498538664, 'BsmtFinSF1'),
 (0.0001940248434984298, 'OpenPorchSF'),
 (0.00022011306725641135, 'WoodDeckSF'),
 (0.0002444063470411281, 'TotalBsmtSF'),
 (0.00024603323913392506, 'X2ndFlrSF'),
 (0.0004106890632050995, 'X1stFlrSF'),
 (0.0004546671719953639, 'ScreenPorch')]
```

[ ]:

```
print_metrics(rr, X_train_lasso, y_train, X_test_lasso, y_test,␣
 ↪flag_log_sale_price = True)
print_metrics(rr, X_train_lasso, y_train, X_test_lasso, y_test,␣
 ↪flag_log_sale_price = False)
```

```
Metrics for Log(Sale Price):

Training R2 0.8452590849089253
Training MAE 0.10467969125067748
Training RMSE 0.1522640805932089
Out-of-sample R2 0.8229399441245389
Out-of-sample MAE 0.11151022706827869
Out-of-sample RMSE 0.16665817850890513

Metrics for Sale Price:

Training R2 0.878237976891266
Training MAE 18648.76164916895
Training RMSE 27259.85846771404
Out-of-sample R2 0.8695106549025007
Out-of-sample MAE 19389.607748338716
Out-of-sample RMSE 28312.465936142868
```

[ ]:

```

```

```
[ ]:
```