House Price Prediction

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn.linear_model import LinearRegression, Ridge, Lasso
from sklearn.model_selection import train_test_split, cross_validate,GridSearchCV
from sklearn.svm import SVR
from sklearn.ensemble import RandomForestRegressor
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import OrdinalEncoder,StandardScaler
from sklearn.ensemble import BaggingRegressor
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.model_selection import cross_val_score
from sklearn.tree import DecisionTreeRegressor
from lazy_impute import *
from sklearn.preprocessing import MinMaxScaler,Normalizer
from sklearn.impute import SimpleImputer
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import r2_score
```

```
train = pd.read_csv('train.csv')
test = pd.read_csv('test.csv')
```

```
pd.pretty_head(train)
```

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

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

  MoSold  YrSold  SaleType  SaleCondition  SalePrice
0      2    2008        WD         Normal     208500
1      5    2007        WD         Normal     181500
2      9    2008        WD         Normal     223500
3      2    2006        WD        Abnorml     140000
4     12    2008        WD         Normal     250000

[5 rows x 81 columns]
```

```
pd.pretty_describe(train)
```

```
                Id  MSSubClass  LotFrontage       LotArea  OverallQual  \
count  1460.000000 1460.000000  1201.000000   1460.000000  1460.000000
mean    730.500000   56.897260    70.049958  10516.828082     6.099315
std     421.610009   42.300571    24.284752   9981.264932     1.382997
min       1.000000   20.000000    21.000000   1300.000000     1.000000
25%     365.750000   20.000000    59.000000   7554.500000     5.000000
50%     730.500000   50.000000    69.000000   9478.500000     6.000000
75%    1095.250000   70.000000    80.000000  11601.500000     7.000000
max    1460.000000  190.000000   313.000000  215245.000000    10.000000

       OverallCond    YearBuilt  YearRemodAdd  MasVnrArea  BsmtFinSF1  \
count  1460.000000  1460.000000   1460.000000    1452.000    1460.000
mean      5.575342  1971.267808   1984.865753     103.685     443.639
std       1.112799    30.202904     20.645407     181.066     456.981
min       1.000000  1872.000000   1950.000000       0.000       0.000
25%       5.000000  1954.000000   1967.000000       0.000     383.500
50%       5.000000  1973.000000   1994.000000       0.000     477.500
75%       6.000000  2000.000000   2004.000000     166.000     808.000
max       9.000000  2010.000000   2010.000000    1600.000    5644.000

       BsmtFinSF2  BsmtUnfSF  TotalBsmtSF     1stFlrSF     2ndFlrSF  \
count    1460.000    1460.000    1460.000  1460.000000  1460.000000
mean       46.549     567.240    1057.429   1162.626712   346.992466
std       161.319     441.369     438.705    386.587738   436.528436
min         0.000       0.000       0.000    334.000000     0.000000
25%         0.000     223.000     795.750    882.000000     0.000000
50%         0.000     477.500    991.500   1087.000000     0.000000
75%         0.000     808.000    1298.250   1391.250000   728.000000
max      1474.000    2336.000    6110.000   4692.000000  2065.000000

       LowQualFinSF  GrLivArea  BsmtFullBath  BsmtHalfBath     FullBath  \
count   1460.000000  1460.000    1460.000000   1460.000000  1460.000000
mean       5.844521  1515.463       0.425342      0.057534     1.565068
std       48.623081   525.480       0.518911      0.238753     0.550916
min        0.000000   334.000       0.000000      0.000000     0.000000
25%        0.000000  1129.500       0.000000      0.000000     1.000000
50%        0.000000  1464.000       0.000000      0.000000     2.000000
75%        0.000000  1776.750       1.000000      0.000000     2.000000
max      572.000000  5642.000       3.000000      2.000000     3.000000

       HalfBath  BedroomAbvGr  KitchenAbvGr  TotRmsAbvGrd  Fireplaces  \
count  1460.000   1460.000000   1460.000000   1460.000000  1460.000000
mean      0.382      2.866438      1.046575      6.517808     0.613014
std       0.502      0.815778      0.220338      1.625393     0.644666
min       0.000      0.000000      0.000000      2.000000     0.000000
25%       0.000      2.000000      1.000000      5.000000     0.000000
50%       0.000      3.000000      1.000000      6.000000     1.000000
75%       1.000      3.000000      1.000000      7.000000     1.000000
max       2.000      8.000000      3.000000     14.000000     3.000000

       GarageYrBlt  GarageCars  GarageArea  WoodDeckSF  OpenPorchSF  \
count  1379.000000  1460.000000  1460.000    1460.000    1460.000
mean   1978.506164     1.767123   472.980      94.244      46.660
std      24.689725     0.747315   213.804     125.338      66.256
min    1900.000000     0.000000     0.000       0.000       0.000
25%    1961.000000     1.000000   334.500       0.000      27.000
50%    1980.000000     2.000000   480.000      0.000      28.000
75%    2000.000000     2.000000   568.500     168.000      68.000
max    2010.000000     4.000000  1418.000     857.000     547.000

       EnclosedPorch  3SsnPorch  ScreenPorch      PoolArea       MiscVal  \
count    1460.000000  1460.000    1460.000    1460.000000    1460.000000
mean       21.954110     3.409     15.060       2.758904      43.489041
std        61.119149     29.317     55.757      40.177307     496.123024
min         0.000000     0.000      0.000       0.000000       0.000000
25%         0.000000     0.000      0.000       0.000000       0.000000
50%         0.000000     0.000      0.000       0.000000       0.000000
75%         0.000000     0.000      0.000       0.000000       0.000000
max       552.000000   508.000    480.000     738.000000   15500.000000

            MoSold       YrSold      SalePrice
count  1460.000000  1460.000000    1460.000000
mean      6.321918  2007.815753  180921.195890
std       2.703626     1.328095   79442.502883
min       1.000000  2006.000000   34900.000000
25%       5.000000  2007.000000  129975.000000
50%       6.000000  2008.000000  163000.000000
75%       8.000000  2009.000000  214000.000000
max      12.000000  2010.000000  755000.000000

[8 rows x 38 columns]
```

```
pd.pretty_head(test)
```

```
   Id  MSSubClass  MSZoning  LotFrontage  LotArea Street Alley LotShape  \
0  1461          20        RH         80.0    11622   Pavt   NaN      Reg
1  1462          20        RL         81.0    14267   Pavt   NaN      IR1
2  1463          60        RL         74.0    13830   Pavt   NaN      IR1
3  1464          60        RL         78.0     9978   Pavt   NaN      IR1
4  1465         120        RL         43.0     5005   Pavt   NaN      IR1

  LandContour Utilities  ...  PoolArea  PoolQC  Fence  MiscFeature  MiscVal  \
0         Lvl    AllPub  ...         0     NaN  MnPrv          NaN        0
1         Lvl    AllPub  ...         0     NaN    NaN         Shed     400
2         Lvl    AllPub  ...         0     NaN  MnPrv          NaN        0
3         Lvl    AllPub  ...         0     NaN    NaN          NaN        0
4         HLS    AllPub  ...         0     NaN    NaN          NaN        0

  MoSold  YrSold  SaleType  SaleCondition
0      6    2010        WD         Normal
1      6    2010        WD         Normal
2      3    2010        WD         Normal
3      6    2010        WD         Normal
4      1    2010        WD         Normal

[5 rows x 80 columns]
```

```
pd.pretty_describe(test)
```

```
                Id  MSSubClass  LotFrontage       LotArea  OverallQual  \
count  1459.000000 1459.000000  1232.000000   1459.000000  1459.000000
mean    2190.000000   57.378341    68.580357   9819.161069     6.078821
std      421.321334   42.746491    22.376841   4955.517327     1.436812
min     1461.000000   20.000000    21.000000   1470.000000     1.000000
25%     1825.500000   20.000000    58.000000   7391.000000     5.000000
50%     2190.000000   50.000000    67.000000   9399.000000     6.000000
75%     2554.500000   70.000000    80.000000  11517.500000     7.000000
max     2919.000000  190.000000   200.000000  56581.000000    10.000000

       OverallCond    YearBuilt  YearRemodAdd  MasVnrArea  BsmtFinSF1  \
count  1459.000000  1459.000000   1459.000000    1444.000    1458.000
mean      5.538671  1971.357779   1983.662783     100.709     439.203
std       1.116883    30.390071     21.130467     177.626     455.268
min       1.000000  1879.000000   1950.000000       0.000       0.000
25%       5.000000  1953.000000   1963.000000       0.000     219.250
50%       5.000000  1973.000000   1992.000000       0.000     465.000
75%       6.000000  2000.000000   2004.000000     164.000     805.500
max       9.000000  2010.000000   2010.000000    1290.000    4010.000

       BsmtFinSF2  BsmtUnfSF  TotalBsmtSF     1stFlrSF     2ndFlrSF  \
count    1458.000    1458.000    1458.000  1459.000000  1459.000000
mean       52.619     570.837    1047.075   1156.534939   325.967788
std       159.351     439.544     396.856    398.123751   420.610568
min         0.000       0.000       0.000    407.000000     0.000000
25%         0.000     219.250     784.750    873.500000     0.000000
50%         0.000     460.500    988.500   1079.000000     0.000000
75%         0.000     797.750    1305.000   1382.500000   676.000000
max      1526.000    2140.000    5095.000   5095.000000  2053.000000

       LowQualFinSF  GrLivArea  BsmtFullBath  BsmtHalfBath     FullBath  \
count   1459.000000  1459.000    1457.000000   1457.000000  1459.000000
mean       6.078821  1486.045       0.435642      0.062081     1.568539
std       48.740932   485.566       0.518314      0.241732     0.553405
min        0.000000   407.000       0.000000      0.000000     0.000000
25%        0.000000  1117.500       0.000000      0.000000     1.000000
50%        0.000000  1432.000       0.000000      0.000000     2.000000
75%        0.000000  1744.750       1.000000      0.000000     2.000000
max      360.000000  5095.000       3.000000      2.000000     3.000000

       HalfBath  BedroomAbvGr  KitchenAbvGr  TotRmsAbvGrd  Fireplaces  \
count  1459.000   1459.000000   1459.000000   1459.000000  1459.000000
mean      0.380      2.854676      1.042466      6.385195     0.581220
std       0.503      0.828577      0.205404      1.508895     0.645011
min       0.000      0.000000      0.000000      3.000000     0.000000
25%       0.000      2.000000      1.000000      5.000000     0.000000
50%       0.000      3.000000      1.000000      6.000000     1.000000
75%       1.000      3.000000      1.000000      7.000000     1.000000
max       2.000      6.000000      2.000000     15.000000     4.000000

       GarageYrBlt  GarageCars  GarageArea  WoodDeckSF  OpenPorchSF  \
count  1381.000000  1458.000000  1458.000    1459.000    1459.000
mean   1977.721899     1.730453   472.768      93.174      48.314
std      26.431599     0.747814   215.394     128.183      67.227
min    1895.000000     0.000000     0.000       0.000       0.000
25%    1959.000000     1.000000   320.000       0.000      30.000
50%    1979.000000     2.000000   480.000      0.000      26.000
75%    2000.000000     2.000000   568.000     168.000      70.000
max    2207.000000     5.000000  1488.000     736.000     406.000

       EnclosedPorch  3SsnPorch  ScreenPorch      PoolArea       MiscVal  \
count    1459.000000  1459.000    1459.000    1459.000000    1459.000000
mean       23.689811     2.251     16.062       1.744345      58.167866
std        64.244246     25.141     56.609      30.492036     630.806671
min         0.000000     0.000      0.000       0.000000       0.000000
25%         0.000000     0.000      0.000       0.000000       0.000000
50%         0.000000     0.000      0.000       0.000000       0.000000
75%         0.000000     0.000      0.000       0.000000       0.000000
max       552.000000   390.000    576.000     800.000000   17000.000000

            MoSold       YrSold
count  1459.000000  1459.000000
mean      6.104181  2008.021948
std       2.721602     1.301808
min       1.000000  2006.000000
25%       4.000000  2007.000000
50%       6.000000  2008.000000
75%       8.000000  2009.000000
max      12.000000  2010.000000

[8 rows x 36 columns]
```

```
missing_values = train.isna().sum().sort_values(ascending=False)
print(missing_values[missing_values > 0])
```

```
PoolQC       1453
MiscFeature  1406
Alley        1369
Fence        1179
MasVnrType    872
FireplaceQu   690
LotFrontage   259
GarageType     81
GarageYrBlt    81
GarageFinish   81
GarageQual     81
GarageCond     81
BsmtExposure   38
BsmtFinType2   38
BsmtCond       37
BsmtQual       37
BsmtFinType1   37
MasVnrArea      8
Electrical      1
dtype: int64
```

```
pd.pretty_describe(y)
```

```
count     1460.000000
mean    180921.195890
std      79442.502883
min      34900.000000
25%     129975.000000
50%     163000.000000
75%     214000.000000
max     755000.000000
Name: SalePrice, dtype: float64
```

```
pd.pretty_describe(X)
```

```
                Id  MSSubClass  LotFrontage       LotArea  OverallQual  \
count  1460.000000 1460.000000  1201.000000   1460.000000  1460.000000
mean    730.500000   56.897260    70.049958  10516.828082     6.099315
std     421.610009   42.300571    24.284752   9981.264932     1.382997
min       1.000000   20.000000    21.000000   1300.000000     1.000000
25%     365.750000   20.000000    59.000000   7554.500000     5.000000
50%     730.500000   50.000000    69.000000   9478.500000     6.000000
75%    1095.250000   70.000000    80.000000  11601.500000     7.000000
max    1460.000000  190.000000   313.000000  215245.000000    10.000000

       OverallCond    YearBuilt  YearRemodAdd  MasVnrArea  BsmtFinSF1  \
count  1460.000000  1460.000000   1460.000000    1452.000    1460.000
mean      5.575342  1971.267808   1984.865753     103.685     443.639
std       1.112799    30.202904     20.645407     181.066     456.981
min       1.000000  1872.000000   1950.000000       0.000       0.000
25%       5.000000  1954.000000   1967.000000       0.000     383.500
50%       5.000000  1973.000000   1994.000000       0.000     477.500
75%       6.000000  2000.000000   2004.000000     166.000     808.000
max       9.000000  2010.000000   2010.000000    1600.000    5644.000

       BsmtFinSF2  BsmtUnfSF  TotalBsmtSF     1stFlrSF     2ndFlrSF  \
count    1460.000    1460.000    1460.000  1460.000000  1460.000000
mean       46.549     567.240    1057.429   1162.626712   346.992466
std       161.319     441.369     438.705    386.587738   436.528436
min         0.000       0.000       0.000    334.000000     0.000000
25%         0.000     223.000     795.750    882.000000     0.000000
50%         0.000     477.500    991.500   1087.000000     0.000000
75%         0.000     808.000    1298.250   1391.250000   728.000000
max      1474.000    2336.000    6110.000   4692.000000  2065.000000

       LowQualFinSF  GrLivArea  BsmtFullBath  BsmtHalfBath     FullBath  \
count   1460.000000  1460.000    1460.000000   1460.000000  1460.000000
mean       5.844521  1515.463       0.425342      0.057534     1.565068
std       48.623081   525.480       0.518911      0.238753     0.550916
min        0.000000   334.000       0.000000      0.000000     0.000000
25%        0.000000  1129.500       0.000000      0.000000     1.000000
50%        0.000000  1464.000       0.000000      0.000000     2.000000
75%        0.000000  1776.750       1.000000      0.000000     2.000000
max      572.000000  5642.000       3.000000      2.000000     3.000000

       HalfBath  BedroomAbvGr  KitchenAbvGr  TotRmsAbvGrd  Fireplaces  \
count  1460.000   1460.000000   1460.000000   1460.000000  1460.000000
mean      0.382      2.866438      1.046575      6.517808     0.613014
std       0.502      0.815778      0.220338      1.625393     0.644666
min       0.000      0.000000      0.000000      2.000000     0.000000
25%       0.000      2.000000      1.000000      5.000000     0.000000
50%       0.000      3.000000      1.000000      6.000000     1.000000
75%       1.000      3.000000      1.000000      7.000000     1.000000
max       2.000      8.000000      3.000000     14.000000     3.000000

       GarageYrBlt  GarageCars  GarageArea  WoodDeckSF  OpenPorchSF  \
count  1379.000000  1460.000000  1460.000    1460.000    1460.000
mean   1978.506164     1.767123   472.980      94.244      46.660
std      24.689725     0.747315   213.804     125.338      66.256
min    1900.000000     0.000000     0.000       0.000       0.000
25%    1961.000000     1.000000   334.500       0.000      27.000
50%    1980.000000     2.000000   480.000      0.000      28.000
75%    2000.000000     2.000000   568.500     168.000      68.000
max    2010.000000     4.000000  1418.000     857.000     547.000

       EnclosedPorch  3SsnPorch  ScreenPorch      PoolArea       MiscVal  \
count    1460.000000  1460.000    1460.000    1460.000000    1460.000000
mean       21.954110     3.409     15.060       2.758904      43.489041
std        61.119149     29.317     55.757      40.177307     496.123024
min         0.000000     0.000      0.000       0.000000       0.000000
25%         0.000000     0.000      0.000       0.000000       0.000000
50%         0.000000     0.000      0.000       0.000000       0.000000
75%         0.000000     0.000      0.000       0.000000       0.000000
max       552.000000   508.000    480.000     738.000000   15500.000000

            MoSold       YrSold
count  1460.000000  1460.000000
mean      6.321918  2007.815753
std       2.703626     1.328095
min       1.000000  2006.000000
25%       5.000000  2007.000000
50%       6.000000  2008.000000
75%       8.000000  2009.000000
max      12.000000  2010.000000

[8 rows x 36 columns]
```

```
pd.pretty_describe(X_test)
```

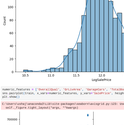
```
                Id  MSSubClass  LotFrontage       LotArea  OverallQual  \
count  1459.000000 1459.000000  1232.000000   1459.000000  1459.000000
mean    2190.000000   57.378341    68.580357   9819.161069     6.078821
std      421.321334   42.746491    22.376841   4955.517327     1.436812
min     1461.000000   20.000000    21.000000   1470.000000     1.000000
25%     1825.500000   20.000000    58.000000   7391.000000     5.000000
50%     2190.000000   50.000000    67.000000   9399.000000     6.000000
75%     2554.500000   70.000000    80.000000  11517.500000     7.000000
max     2919.000000  190.000000   200.000000  56581.000000    10.000000

       OverallCond    YearBuilt  YearRemodAdd  MasVnrArea  BsmtFinSF1  \
count  1459.000000  1459.000000   1459.000000    1444.000    1458.000
mean      5.538671  1971.357779   1983.662783     100.709     439.203
std       1.116883    30.390071     21.130467     177.626     455.268
min       1.000000  1879.000000   1950.000000       0.000       0.000
25%       5.000000  1953.000000   1963.000000       0.000     219.250
50%       5.000000  1973.000000   1992.000000       0.000     465.000
75%       6.000000  2000.000000   2004.000000     164.000     805.500
max       9.000000  2010.000000   2010.000000    1290.000    4010.000

       BsmtFinSF2  BsmtUnfSF  TotalBsmtSF     1stFlrSF     2ndFlrSF  \
count    1458.000    1458.000    1458.000  1459.000000  1459.000000
mean       52.619     570.837    1047.075   1156.534939   325.967788
std       159.351     439.544     396.856    398.123751   420.610568
min         0.000       0.000       0.000    407.000000     0.000000
25%         0.000     219.250     784.750    873.500000     0.000000
50%         0.000     460.500    988.500   1079.000000     0.000000
75%         0.000     797.750    1305.000   1382.500000   676.000000
max      1526.000    2140.000    5095.000   5095.000000  2053.000000

       LowQualFinSF  GrLivArea  BsmtFullBath  BsmtHalfBath     FullBath  \
count   1459.000000  1459.000    1457.000000   1457.000000  1459.000000
mean       6.078821  1486.045       0.435642      0.062081     1.568539
std       48.740932   485.566       0.518314      0.241732     0.553405
min        0.000000   407.000       0.000000      0.000000     0.000000
25%        0.000000  1117.500       0.000000      0.000000     1.000000
50%        0.000000  1432.000       0.000000      0.000000     2.000000
75%        0.000000  1744.750       1.000000      0.000000     2.000000
max      360.000000  5095.000       3.000000      2.000000     3.000000

       HalfBath  BedroomAbvGr  KitchenAbvGr  TotRmsAbvGrd  Fireplaces  \
count  1459.000   1459.000000   1459.000000   1459.000000  1459.000000
mean      0.380      2.854676      1.042466      6.385195     0.581220
std       0.503      0.828577      0.205404      1.508895     0.645011
min       0.000      0.000000      0.000000      3.000000     0.000000
25%       0.000      2.000000      1.000000      5.000000     0.000000
50%       0.000      3.000000      1.000000      6.000000     1.000000
75%       1.000      3.000000      1.000000      7.000000     1.000000
max       2.000      6.000000      2.000000     15.000000     4.000000

       GarageYrBlt  GarageCars  GarageArea  WoodDeckSF  OpenPorchSF  \
count  1381.000000  1458.000000  1458.000    1459.000    1459.000
mean   1977.721899     1.730453   472.768      93.174      48.314
std      26.431599     0.747814   215.394     128.183      67.227
min    1895.000000     0.000000     0.000       0.000       0.000
25%    1959.000000     1.000000   320.000       0.000      30.000
50%    1979.000000     2.000000   480.000      0.000      26.000
75%    2000.000000     2.000000   568.000     168.000      70.000
max    2207.000000     5.000000  1488.000     736.000     406.000

       EnclosedPorch  3SsnPorch  ScreenPorch      PoolArea       MiscVal  \
count    1459.000000  1459.000    1459.000    1459.000000    1459.000000
mean       23.689811     2.251     16.062       1.744345      58.167866
std        64.244246     29.317     56.609      30.492036     630.806671
min         0.000000     0.000      0.000       0.000000       0.000000
25%         0.000000     0.000      0.000       0.000000       0.000000
50%         0.000000     0.000      0.000       0.000000       0.000000
75%         0.000000     0.000      0.000       0.000000       0.000000
max       552.000000   390.000    576.000     800.000000   17000.000000

            MoSold       YrSold
count  1459.000000  1459.000000
mean      6.104181  2008.021948
std       2.721602     1.301808
min       1.000000  2006.000000
25%       4.000000  2007.000000
50%       6.000000  2008.000000
75%       8.000000  2009.000000
max      12.000000  2010.000000

[8 rows x 36 columns]
```

```
pd.pretty_describe(y)
```

```
count     1460.000000
mean    180921.195890
std      79442.502883
min      34900.000000
25%     129975.000000
50%     163000.000000
75%     214000.000000
max     755000.000000
Name: SalePrice, dtype: float64
```

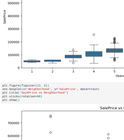
```
plt.figure(figsize=(8, 6))
sns.histplot(y, bins=30, kde=True)
plt.title('Distribution of SalePrice')
plt.show()
```

```
train['LogSalePrice'] = np.log(train['SalePrice'])
plt.figure(figsize=(8, 6))
sns.histplot(train['LogSalePrice'], kde=True)
plt.title('Distribution of LogSalePrice')
plt.show()
```

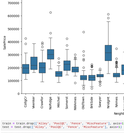



```
numeric_features = ['GrLivArea', 'OverallQual', 'TotalBsmtSF', 'YearBuilt']
sns.pairplot(train, x_vars=numeric_features, y_vars='SalePrice', height=4, aspect=1, kind='scatter')
plt.show()
```

```
/opt/anaconda3/anaconda3/lib/python3/dist-packages/seaborn/distrib.py:425: UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```




```
plt.figure(figsize=(8, 6))
sns.boxplot(x='OverallQual', y='SalePrice', data=train)
plt.title('SalePrice vs OverallQual')
plt.show()
```




```
plt.figure(figsize=(8, 6))
sns.boxplot(x='Neighborhood', y='SalePrice', data=train)
plt.title('SalePrice vs Neighborhood')
plt.xticks(rotation=45)
plt.show()
```

```
train = train.drop(['Alley', 'PoolQC', 'Fence', 'MiscFeature'], axis=1)
test = test.drop(['Alley', 'PoolQC', 'Fence', 'MiscFeature'], axis=1)
```

```
for col in train.columns:
    if train[col].dtype == 'object':
        train[col] = train[col].fillna(train[col].mode()[0])
    else:
        train[col] = train[col].fillna(train[col].median())

for col in test.columns:
    if test[col].dtype == 'object':
        test[col] = test[col].fillna(test[col].mode()[0])
    else:
        test[col] = test[col].fillna(test[col].median())
```

```
label_encoders = {}
for col in train.columns:
    if train[col].dtype == 'object':
        le = LabelEncoder()
        train[col] = le.fit_transform(train[col])
        label_encoders[col] = le

for col in test.columns:
    if test[col].dtype == 'object':
        le = label_encoders[col]
        test[col] = le.transform(test[col])
except KeyError:
        test[col] = -1
        test[col] = le.fit_transform(test[col])
label_encoders[col] = le
```

```
X = train.drop(['Id', 'SalePrice'], axis=1)
y = train['SalePrice']
X_test = test.drop(['Id'], axis=1)
```

```
y = train['SalePrice']
```

```
sns.heatmap(X.corr(), cmap='coolwarm')
plt.title('Feature Correlation Heatmap')
```

```
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
dtrain = xgb.DMatrix(X_train, label=y_train)
dval = xgb.DMatrix(X_val, label=y_val)
dtest = xgb.DMatrix(X_test)
```

```
params = {
    'objective': 'reg:squarederror',
    'eta': 0.01,
    'max_depth': 6,
    'subsample': 0.8,
    'colsample_bytree': 0.8,
    'seed': 42
}

num_rounds = 100
```

```
bst = xgb.train(params, dtrain, num_rounds, evals=[(dval, 'eval')], early_stopping_rounds=10)
```

```
[0]	eval-rmse:79866.71875
[1]	eval-rmse:79282.51953
[2]	eval-rmse:78705.21094
[3]	eval-rmse:78134.51563
[4]	eval-rmse:77570.32031
[5]	eval-rmse:76012.64844
[6]	eval-rmse:74509.09766
[7]	eval-rmse:73059.07031
[8]	eval-rmse:71660.64844
[9]	eval-rmse:70312.64844
[10]	eval-rmse:69013.94688
[11]	eval-rmse:67762.69688
[12]	eval-rmse:66557.00625
[13]	eval-rmse:65394.74453
[14]	eval-rmse:64274.59375
[15]	eval-rmse:63195.01641
[16]	eval-rmse:62154.64453
[17]	eval-rmse:61152.01953
[18]	eval-rmse:60185.79453
[19]	eval-rmse:59254.79453
[20]	eval-rmse:58358.01953
[21]	eval-rmse:57494.79453
[22]	eval-rmse:56663.01953
[23]	eval-rmse:55862.79453
[24]	eval-rmse:55092.01953
[25]	eval-rmse:54350.79453
[26]	eval-rmse:53637.01953
[27]	eval-rmse:52950.79453
[28]	eval-rmse:52290.01953
[29]	eval-rmse:51654.79453
[30]	eval-rmse:51043.01953
[31]	eval-rmse:50454.79453
[32]	eval-rmse:49888.01953
[33]	eval-rmse:49342.79453
[34]	eval-rmse:48817.01953
[35]	eval-rmse:48311.79453
[36]	eval-rmse:47825.01953
[37]	eval-rmse:47356.79453
[38]	eval-rmse:46905.01953
[39]	eval-rmse:46469.79453
[40]	eval-rmse:46049.01953
[41]	eval-rmse:45643.79453
[42]	eval-rmse:45251.01953
[43]	eval-rmse:44872.79453
[44]	eval-rmse:44506.01953
[45]	eval-rmse:44151.79453
[46]	eval-rmse:43809.01953
[47]	eval-rmse:43477.79453
[48]	eval-rmse:43157.01953
[49]	eval-rmse:42846.79453
[50]	eval-rmse:42546.01953
[51]	eval-rmse:42255.79453
[52]	eval-rmse:41974.01953
[53]	eval-rmse:41701.79453
[54]	eval-rmse:41437.01953
[55]	eval-rmse:41181.79453
[56]	eval-rmse:40933.01953
[57]	eval-rmse:40692.79453
[58]	eval-rmse:40459.01953
[59]	eval-rmse:40232.79453
[60]	eval-rmse:40013.01953
[61]	eval-rmse:39799.79453
[62]	eval-rmse:39592.01953
[63]	eval-rmse:39390.79453
[64]	eval-rmse:39195.01953
[65]	eval-rmse:39004.79453
[66]	eval-rmse:38819.01953
[67]	eval-rmse:38638.79453
[68]	eval-rmse:38463.01953
[69]	eval-rmse:38291.79453
[70]	eval-rmse:38125.01953
[71]	eval-rmse:37962.79453
[72]	eval-rmse:37804.01953
[73]	eval-rmse:37649.79453
[74]	eval-rmse:37498.01953
[75]	eval-rmse:37350.79453
[76]	eval-rmse:37206.01953
[77]	eval-rmse:37065.79453
[78]	eval-rmse:36927.01953
[79]	eval-rmse:36792.79453
[80]	eval-rmse:36660.01953
[81]	eval-rmse:36531.79453
[82]	eval-rmse:36405.01953
[83]	eval-rmse:36281.79453
[84]	eval-rmse:36160.01953
[85]	eval-rmse:36041.79453
[86]	eval-rmse:35925.01953
[87]	eval-rmse:35811.79453
[88]	eval-rmse:35699.01953
[89]	eval-rmse:35589.79453
[90]	eval-rmse:35482.01953
[91]	eval-rmse:35376.79453
[92]	eval-rmse:35273.01953
[93]	eval-rmse:35171.79453
[94]	eval-rmse:35072.01953
[95]	eval-rmse:34974.79453
[96]	eval-rmse:34879.01953
[97]	eval-rmse:34785.79453
[98]	eval-rmse:34693.01953
[99]	eval-rmse:34602.79453
```

```
y_pred = bst.predict(dtest)
```

```
y_test_pred = bst.predict(dval)
```

```
y_test_pred
```

```
array([180546.12, 161237.40, 114594.125, 230273.42, 339410.06,
       191024.  , 178854.  , 178854.  , 178854.  , 178854.  ,
       178854
```