



**T.C.
FATİH SULTAN MEHMET VAKIF
ÜNİVERSİTESİ**

MIDTERM

DATA SCIENCE

Meryem Ezber
1521221034

Ayla Gülcü

House Prices

The data consists of 1460 rows and 81 columns. It contains 43 objects, 35 int, 3 float type data.

Depending on other features in the dataset, the sale price is estimated. When the Sale Price column is examined, the following chart is created. (Figure 1)

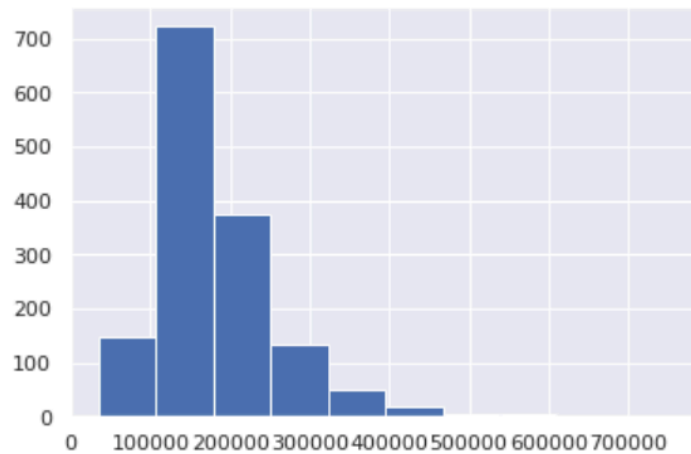


Figure 1

When this value is normalized, a graphic like the following is obtained. (Figure 2)

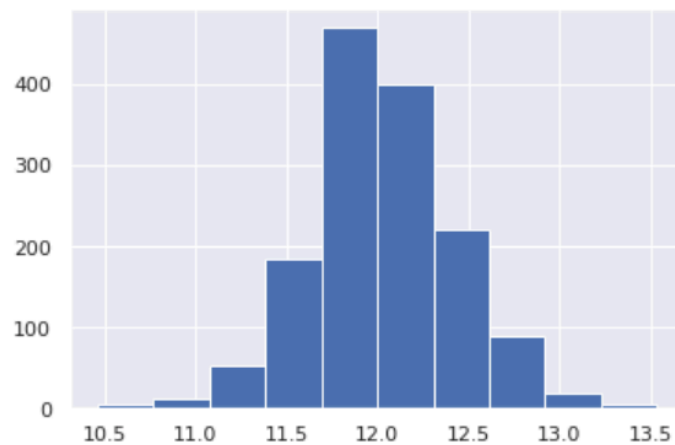


Figure 2

Firstly, missing values in the dataset were cleared. Then outlier values were determined and cleaned.

Feature selection was performed after the dataset was cleaned and prepared. Filter method was used for feature selection.

In this method, you filter and take only the subset of the relevant features. The model is built after selecting the features. The filtering here is done using correlation matrix and it is most commonly done using Pearson correlation.

The correlation coefficient has values between -1 to 1

- A value closer to 0 implies weaker correlation (exact 0 implying no correlation)
- A value closer to 1 implies stronger positive correlation
- A value closer to -1 implies stronger negative correlation

The correlation between Sale Price and other features is as follows. (Figure 3, Figure 4)

```
Skewed_SP      0.948374
OverallQual     0.790982
GrLivArea       0.708624
GarageCars      0.640409
GarageArea      0.623431
TotalBsmtSF     0.613581
1stFlrSF        0.605852
FullBath        0.560664
TotRmsAbvGrd   0.533723
YearBuilt       0.522897
YearRemodAdd    0.507101
Fireplaces      0.466929
BsmtFinSF1      0.386420
WoodDeckSF      0.324413
2ndFlrSF        0.319334
OpenPorchSF     0.315856
HalfBath        0.284108
LotArea         0.263843
BsmtFullBath    0.227122
BsmtUnfSF       0.214479
BedroomAbvGr    0.168213
ScreenPorch     0.111447
PoolArea        0.092404
MoSold          0.046432
3SsnPorch       0.044584
BsmtFinSF2     -0.011378
BsmtHalfBath    -0.016844
MiscVal         -0.021190
Id              -0.021917
LowQualFinSF    -0.025606
YrSold          -0.028923
OverallCond     -0.077856
MSSubClass      -0.084284
EnclosedPorch   -0.128578
KitchenAbvGr    -0.135907
Name: SalePrice, dtype: float64
```

Figure 3

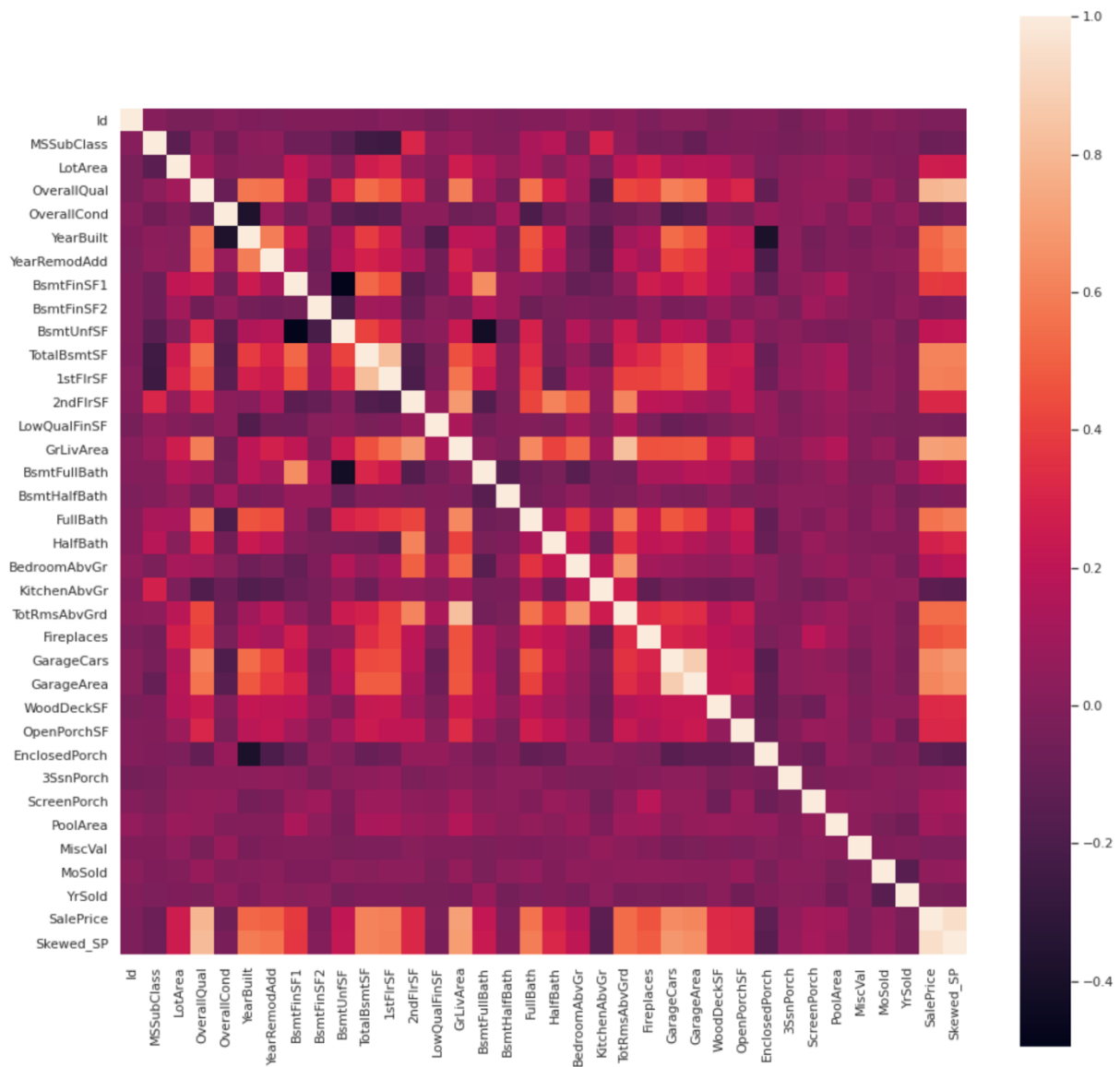


Figure 4

Correlation graphs with the 6 most related features are as follows. (Figure 6)

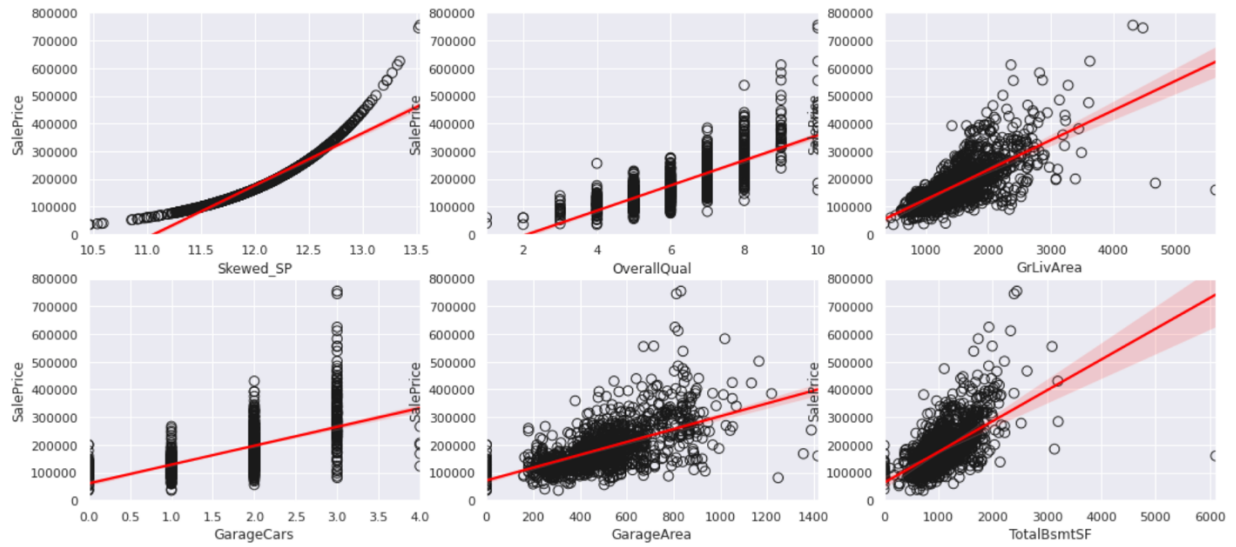


Figure 5

<https://www.kaggle.com/meryemezber/house-prices>