In [217…
```
import pandas as pd
import numpy as np
from sklearn import preprocessing
from sklearn.preprocessing import OneHotEncoder
from sklearn.linear_model import Ridge
from sklearn import preprocessing
le = preprocessing.LabelEncoder()
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from sklearn.model_selection import KFold
from sklearn.preprocessing import OneHotEncoder
from sklearn.linear_model import LinearRegression


from sklearn.cluster import KMeans
from sklearn.metrics import r2_score, mean_absolute_error
from sklearn.metrics import mean_squared_error

import sklearn as sklearn

from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import accuracy_score
```

In [218…
```
# Setting this option will print all collumns of a dataframe
pd.set_option('display.max_columns', None)
```

In [219…
```
# Setting this option will print all of the data in a feature
pd.set_option('display.max_colwidth', None)
```

In [220…
```
# Setting this option will print all rows di je none, jer mi se nije sve ispisivalo
#pd.set_option('max_rows', 100)
# za nove pande
pd.options.display.max_rows = 100
```

In [221…
```
trainhouse = pd.read_csv('trainhouse.csv')
```

In [222…
```
testhouse = pd.read_csv('testhouse.csv')
#50% uzeli za test
```

In [223…
```
trainhouse.head(15)
```

Out[223]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Uti
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	A
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	A
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	A
3	4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	A
4	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	A
5	6	50	RL	85.0	14115	Pave	NaN	IR1	Lvl	A
6	7	20	RL	75.0	10084	Pave	NaN	Reg	Lvl	A
7	8	60	RL	NaN	10382	Pave	NaN	IR1	Lvl	A
8	9	50	RM	51.0	6120	Pave	NaN	Reg	Lvl	A
9	10	190	RL	50.0	7420	Pave	NaN	Reg	Lvl	A
10	11	20	RL	70.0	11200	Pave	NaN	Reg	Lvl	A
11	12	60	RL	85.0	11924	Pave	NaN	IR1	Lvl	A
12	13	20	RL	NaN	12968	Pave	NaN	IR2	Lvl	A
13	14	20	RL	91.0	10652	Pave	NaN	IR1	Lvl	A
14	15	20	RL	NaN	10920	Pave	NaN	IR1	Lvl	A

In [224…

```
trainvisual = trainhouse.iloc[:,1:81]
trainvisual.head()
```

Out[224]:

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

In [225…

```
trainhouse.isnull().sum()
```

Out[225]:

```
Id                 0
MSSubClass         0
MSZoning           0
LotFrontage      259
LotArea            0
Street             0
Alley           1369
LotShape           0
LandContour        0
Utilities          0
LotConfig          0
LandSlope          0
Neighborhood       0
Condition1         0
Condition2         0
BldgType           0
HouseStyle         0
OverallQual        0
OverallCond        0
YearBuilt          0
YearRemodAdd       0
RoofStyle          0
RoofMatl           0
Exterior1st        0
Exterior2nd        0
MasVnrType         8
MasVnrArea         8
ExterQual          0
ExterCond          0
Foundation         0
BsmtQual          37
BsmtCond          37
BsmtExposure      38
BsmtFinType1      37
BsmtFinSF1         0
BsmtFinType2      38
BsmtFinSF2         0
BsmtUnfSF          0
TotalBsmtSF        0
Heating            0
HeatingQC          0
CentralAir         0
Electrical         1
1stFlrSF           0
2ndFlrSF           0
LowQualFinSF       0
GrLivArea          0
BsmtFullBath       0
BsmtHalfBath       0
FullBath           0
HalfBath           0
BedroomAbvGr       0
KitchenAbvGr       0
KitchenQual        0
TotRmsAbvGrd       0
Functional         0
Fireplaces         0
FireplaceQu      690
GarageType        81
GarageYrBlt       81
GarageFinish      81
GarageCars         0
GarageArea         0
GarageQual        81
```

```
GarageCond         81
PavedDrive          0
WoodDeckSF          0
OpenPorchSF         0
EnclosedPorch       0
3SsnPorch           0
ScreenPorch         0
PoolArea            0
PoolQC           1453
Fence            1179
MiscFeature      1406
MiscVal             0
MoSold              0
YrSold              0
SaleType            0
SaleCondition       0
SalePrice           0
dtype: int64
```

In [226…

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

#sve su tipovi za bins, mogu se grupirat preko encodea u svakoj koloni ili brojevi
```

In [227…

```
#i will check on datatypes of columns
#i will use it do divide categorical from numeric variables and then tobe able to j
#or some selected value i will see
trainhouse.dtypes
```

Out[227]:

```
Id                int64
MSSubClass        int64
MSZoning         object
LotFrontage     float64
LotArea           int64
Street           object
Alley            object
LotShape         object
LandContour      object
Utilities        object
LotConfig        object
LandSlope        object
Neighborhood     object
Condition1       object
Condition2       object
BldgType         object
HouseStyle       object
OverallQual       int64
OverallCond       int64
YearBuilt         int64
YearRemodAdd      int64
RoofStyle        object
RoofMatl         object
Exterior1st      object
Exterior2nd      object
MasVnrType       object
MasVnrArea      float64
ExterQual        object
ExterCond        object
Foundation       object
BsmtQual         object
BsmtCond         object
BsmtExposure     object
BsmtFinType1     object
BsmtFinSF1        int64
BsmtFinType2     object
BsmtFinSF2        int64
BsmtUnfSF         int64
TotalBsmtSF       int64
Heating          object
HeatingQC        object
CentralAir       object
Electrical       object
1stFlrSF          int64
2ndFlrSF          int64
LowQualFinSF      int64
GrLivArea         int64
BsmtFullBath      int64
BsmtHalfBath      int64
FullBath          int64
HalfBath          int64
BedroomAbvGr      int64
KitchenAbvGr      int64
KitchenQual      object
TotRmsAbvGrd      int64
Functional       object
Fireplaces        int64
FireplaceQu      object
GarageType       object
GarageYrBlt     float64
GarageFinish     object
GarageCars        int64
GarageArea        int64
GarageQual       object
```

```
GarageCond        object
PavedDrive        object
WoodDeckSF         int64
OpenPorchSF        int64
EnclosedPorch      int64
3SsnPorch          int64
ScreenPorch        int64
PoolArea           int64
PoolQC            object
Fence             object
MiscFeature       object
MiscVal            int64
MoSold             int64
YrSold             int64
SaleType          object
SaleCondition     object
SalePrice          int64
dtype: object
```

In [228…

```
trainhouse[['LotFrontage','MasVnrArea','GarageYrBlt']] = trainhouse[['LotFrontage'
```

In [229…

```
trainhouse['Alley'] = trainhouse['Alley'].fillna(trainhouse['Alley'].astype('strin
#mode bas ove kolone
```

In [230…

```
trainhouse['BsmtQual'] = trainhouse['BsmtQual'].fillna(trainhouse['BsmtQual'].asty
```

In [231…

```
trainhouse['BsmtCond'] = trainhouse['BsmtCond'].fillna(trainhouse['BsmtCond'].asty
```

In [232…

```
trainhouse['BsmtExposure'] = trainhouse['BsmtExposure'].fillna(trainhouse['BsmtExp
trainhouse['BsmtFinType1'] = trainhouse['BsmtFinType1'].fillna(trainhouse['BsmtFin
trainhouse['BsmtFinType2'] = trainhouse['BsmtFinType2'].fillna(trainhouse['BsmtFin
trainhouse['Electrical'] = trainhouse['Electrical'].fillna(trainhouse['Electrical'
trainhouse['FireplaceQu'] = trainhouse['FireplaceQu'].fillna(trainhouse['Fireplace
trainhouse['GarageType'] = trainhouse['GarageType'].fillna(trainhouse['GarageType'
trainhouse['GarageYrBlt'] = trainhouse['GarageYrBlt'].fillna(trainhouse['GarageYrB
trainhouse['GarageFinish'] = trainhouse['GarageFinish'].fillna(trainhouse['GarageF
trainhouse['GarageQual'] = trainhouse['GarageQual'].fillna(trainhouse['GarageQual'
trainhouse['GarageCond'] = trainhouse['GarageCond'].fillna(trainhouse['GarageCond'
trainhouse['PoolQC'] = trainhouse['PoolQC'].fillna(trainhouse['PoolQC'].astype('st
trainhouse['PoolQC'] = trainhouse['PoolQC'].fillna(trainhouse['PoolQC'].astype('st
trainhouse['Fence'] = trainhouse['Fence'].fillna(trainhouse['Fence'].astype('strin
trainhouse['MiscFeature'] = trainhouse['MiscFeature'].fillna(trainhouse['MiscFeatu
trainhouse['MasVnrType'] = trainhouse['MasVnrType'].fillna(trainhouse['MasVnrType'
```

In [233…

```
trainhouse.isnull().sum().sum()
```

Out[233]:

```
0
```

In [234…

```
trainhouse.head(20)
```

Out[234]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Uti
0	1	60	RL	65.0	8450	Pave	Grvl	Reg	Lvl	A
1	2	20	RL	80.0	9600	Pave	Grvl	Reg	Lvl	A
2	3	60	RL	68.0	11250	Pave	Grvl	IR1	Lvl	A
3	4	70	RL	60.0	9550	Pave	Grvl	IR1	Lvl	A
4	5	60	RL	84.0	14260	Pave	Grvl	IR1	Lvl	A
5	6	50	RL	85.0	14115	Pave	Grvl	IR1	Lvl	A
6	7	20	RL	75.0	10084	Pave	Grvl	Reg	Lvl	A
7	8	60	RL	69.0	10382	Pave	Grvl	IR1	Lvl	A
8	9	50	RM	51.0	6120	Pave	Grvl	Reg	Lvl	A
9	10	190	RL	50.0	7420	Pave	Grvl	Reg	Lvl	A
10	11	20	RL	70.0	11200	Pave	Grvl	Reg	Lvl	A
11	12	60	RL	85.0	11924	Pave	Grvl	IR1	Lvl	A
12	13	20	RL	69.0	12968	Pave	Grvl	IR2	Lvl	A
13	14	20	RL	91.0	10652	Pave	Grvl	IR1	Lvl	A
14	15	20	RL	69.0	10920	Pave	Grvl	IR1	Lvl	A
15	16	45	RM	51.0	6120	Pave	Grvl	Reg	Lvl	A
16	17	20	RL	69.0	11241	Pave	Grvl	IR1	Lvl	A
17	18	90	RL	72.0	10791	Pave	Grvl	Reg	Lvl	A
18	19	20	RL	66.0	13695	Pave	Grvl	Reg	Lvl	A
19	20	20	RL	70.0	7560	Pave	Grvl	Reg	Lvl	A

In [235…

```
testhouse.head()
```

Out[235]:

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

In [236…

```
testhouse.isnull().sum()
```

Out[236]:

Id	0
MSSubClass	0
MSZoning	4
LotFrontage	227
LotArea	0
Street	0
Alley	1352
LotShape	0
LandContour	0
Utilities	2
LotConfig	0
LandSlope	0
Neighborhood	0
Condition1	0
Condition2	0
BldgType	0
HouseStyle	0
OverallQual	0
OverallCond	0
YearBuilt	0
YearRemodAdd	0
RoofStyle	0
RoofMatl	0
Exterior1st	1
Exterior2nd	1
MasVnrType	16
MasVnrArea	15
ExterQual	0
ExterCond	0
Foundation	0
BsmtQual	44
BsmtCond	45
BsmtExposure	44
BsmtFinType1	42
BsmtFinSF1	1
BsmtFinType2	42
BsmtFinSF2	1
BsmtUnfSF	1
TotalBsmtSF	1
Heating	0
HeatingQC	0
CentralAir	0
Electrical	0
1stFlrSF	0
2ndFlrSF	0
LowQualFinSF	0
GrLivArea	0
BsmtFullBath	2
BsmtHalfBath	2
FullBath	0
HalfBath	0
BedroomAbvGr	0
KitchenAbvGr	0
KitchenQual	1
TotRmsAbvGrd	0
Functional	2
Fireplaces	0
FireplaceQu	730
GarageType	76
GarageYrBlt	78
GarageFinish	78
GarageCars	1
GarageArea	1
GarageQual	78

```
GarageCond         78
PavedDrive          0
WoodDeckSF          0
OpenPorchSF         0
EnclosedPorch       0
3SsnPorch           0
ScreenPorch         0
PoolArea            0
PoolQC           1456
Fence            1169
MiscFeature      1408
MiscVal             0
MoSold              0
YrSold              0
SaleType            1
SaleCondition       0
dtype: int64
```

In [237…

```
# MSZoning
# LotFrontage*
# Alley
# Utilities
# Exterior1st
# Exterior2nd
# MasVnrType
# MasVnrArea *
# BsmtQual
# BsmtCond
# BsmtExposure
# BsmtFinType1
# BsmtFinSF1 *
# BsmtFinType2
# BsmtFinSF2 *
# BsmtUnfSF *
# TotalBsmtSF *
# BsmtFullBath*
# BsmtHalfBath *
# KitchenQual
# Functional
# FireplaceQu
# GarageType
# GarageYrBlt *
# GarageFinish
# GarageCars  *
# GarageArea *
# GarageQual
# GarageCond
# PoolQC
# Fence
# MiscFeature
# SaleType
```

In [238…

```
testhouse.dtypes
```

Out[238]:

```
Id                int64
MSSubClass        int64
MSZoning         object
LotFrontage     float64
LotArea           int64
Street           object
Alley            object
LotShape         object
LandContour      object
Utilities        object
LotConfig        object
LandSlope        object
Neighborhood     object
Condition1       object
Condition2       object
BldgType         object
HouseStyle       object
OverallQual       int64
OverallCond       int64
YearBuilt         int64
YearRemodAdd      int64
RoofStyle        object
RoofMatl         object
Exterior1st      object
Exterior2nd      object
MasVnrType       object
MasVnrArea      float64
ExterQual        object
ExterCond        object
Foundation       object
BsmtQual         object
BsmtCond         object
BsmtExposure     object
BsmtFinType1     object
BsmtFinSF1      float64
BsmtFinType2     object
BsmtFinSF2      float64
BsmtUnfSF       float64
TotalBsmtSF     float64
Heating          object
HeatingQC        object
CentralAir       object
Electrical       object
1stFlrSF          int64
2ndFlrSF          int64
LowQualFinSF      int64
GrLivArea         int64
BsmtFullBath    float64
BsmtHalfBath    float64
FullBath          int64
HalfBath          int64
BedroomAbvGr      int64
KitchenAbvGr      int64
KitchenQual      object
TotRmsAbvGrd      int64
Functional       object
Fireplaces        int64
FireplaceQu      object
GarageType       object
GarageYrBlt     float64
GarageFinish     object
GarageCars      float64
GarageArea      float64
GarageQual       object
```

```
GarageCond        object
PavedDrive        object
WoodDeckSF         int64
OpenPorchSF        int64
EnclosedPorch      int64
3SsnPorch          int64
ScreenPorch        int64
PoolArea           int64
PoolQC            object
Fence             object
MiscFeature       object
MiscVal            int64
MoSold             int64
YrSold             int64
SaleType          object
SaleCondition     object
dtype: object
```

In [239…

```
for column in ['MSZoning','Alley','Utilities','Exterior1st','Exterior2nd','MasVnrT
    testhouse[column] = testhouse[column].fillna(testhouse[column].mode()[0])
```

In [240…

```
testhouse.isnull().sum()
```

Out[240]:

```
Id                 0
MSSubClass         0
MSZoning           0
LotFrontage      227
LotArea            0
Street             0
Alley              0
LotShape           0
LandContour        0
Utilities          0
LotConfig          0
LandSlope          0
Neighborhood       0
Condition1         0
Condition2         0
BldgType           0
HouseStyle         0
OverallQual        0
OverallCond        0
YearBuilt          0
YearRemodAdd       0
RoofStyle          0
RoofMatl           0
Exterior1st        0
Exterior2nd        0
MasVnrType         0
MasVnrArea        15
ExterQual          0
ExterCond          0
Foundation         0
BsmtQual           0
BsmtCond           0
BsmtExposure       0
BsmtFinType1       0
BsmtFinSF1         1
BsmtFinType2       0
BsmtFinSF2         1
BsmtUnfSF          1
TotalBsmtSF        1
Heating            0
HeatingQC          0
CentralAir         0
Electrical         0
1stFlrSF           0
2ndFlrSF           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
GarageYrBlt       78
GarageFinish       0
GarageCars         1
GarageArea         1
GarageQual         0
```

```
GarageCond          0
PavedDrive          0
WoodDeckSF          0
OpenPorchSF         0
EnclosedPorch       0
3SsnPorch           0
ScreenPorch         0
PoolArea            0
PoolQC              0
Fence               0
MiscFeature         0
MiscVal             0
MoSold              0
YrSold              0
SaleType            0
SaleCondition       0
dtype: int64
```

In [241…

```
testhouse[['LotFrontage','MasVnrArea','BsmtFinSF1','BsmtFinSF2','BsmtUnfSF','Totall
```

In [242…

```
testhouse.isnull().sum().sum()
```

Out[242]:

```
0
```

In [243…

```
#encoding
```

In [244…

```
trainhouse.head(5)
```

Out[244]:

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

In [245…

```
#another way
#i took categorical variables which had null, addedthose that needed encoding and e
#encoded month year and some numeric variables, so that they just belong to differe
#the result (same as different numbers from 1 to 5 maybe should just mean differen
#bigger value if style is 4th and not 2nd)
encodecolstrain= ['MSSubClass','MSZoning','Street','Alley','LotShape','LandContour

for col in encodecolstrain:
    trainhouse[col]=le.fit_transform(trainhouse[col])
    print(le.classes_)

trainhouse.head(5)
```

```
[ 20  30  40  45  50  60  70  75  80  85  90 120 160 180 190]
['C (all)' 'FV' 'RH' 'RL' 'RM']
['Grvl' 'Pave']
['Grvl' 'Pave']
['IR1' 'IR2' 'IR3' 'Reg']
['Bnk' 'HLS' 'Low' 'Lvl']
['AllPub' 'NoSeWa']
['Corner' 'CulDSac' 'FR2' 'FR3' 'Inside']
['Gtl' 'Mod' 'Sev']
['Blmngtn' 'Blueste' 'BrDale' 'BrkSide' 'ClearCr' 'CollgCr' 'Crawfor'
 'Edwards' 'Gilbert' 'IDOTRR' 'MeadowV' 'Mitchel' 'NAmes' 'NPkVill'
 'NWAmes' 'NoRidge' 'NridgHt' 'OldTown' 'SWISU' 'Sawyer' 'SawyerW'
 'Somerst' 'StoneBr' 'Timber' 'Veenker']
['Artery' 'Feedr' 'Norm' 'PosA' 'PosN' 'RRAe' 'RRAn' 'RRNe' 'RRNn']
['Artery' 'Feedr' 'Norm' 'PosA' 'PosN' 'RRAe' 'RRAn' 'RRNn']
['1Fam' '2fmCon' 'Duplex' 'Twnhs' 'TwnhsE']
['1.5Fin' '1.5Unf' '1Story' '2.5Fin' '2.5Unf' '2Story' 'SFoyer' 'SLvl']
[1872 1875 1880 1882 1885 1890 1892 1893 1898 1900 1904 1905 1906 1908
 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923
 1924 1925 1926 1927 1928 1929 1930 1931 1932 1934 1935 1936 1937 1938
 1939 1940 1941 1942 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954
 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968
 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982
 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996
 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010]
[1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963
 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977
 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991
 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005
 2006 2007 2008 2009 2010]
['Flat' 'Gable' 'Gambrel' 'Hip' 'Mansard' 'Shed']
['ClyTile' 'CompShg' 'Membran' 'Metal' 'Roll' 'Tar&Grv' 'WdShake'
 'WdShngl']
['AsbShng' 'AsphShn' 'BrkComm' 'BrkFace' 'CBlock' 'CemntBd' 'HdBoard'
 'ImStucc' 'MetalSd' 'Plywood' 'Stone' 'Stucco' 'VinylSd' 'Wd Sdng'
 'WdShing']
['AsbShng' 'AsphShn' 'Brk Cmn' 'BrkFace' 'CBlock' 'CmentBd' 'HdBoard'
 'ImStucc' 'MetalSd' 'Other' 'Plywood' 'Stone' 'Stucco' 'VinylSd'
 'Wd Sdng' 'Wd Shng']
['BrkCmn' 'BrkFace' 'None' 'Stone']
['Ex' 'Fa' 'Gd' 'TA']
['Ex' 'Fa' 'Gd' 'Po' 'TA']
['BrkTil' 'CBlock' 'PConc' 'Slab' 'Stone' 'Wood']
['Ex' 'Fa' 'Gd' 'TA']
['Fa' 'Gd' 'Po' 'TA']
['Av' 'Gd' 'Mn' 'No']
['ALQ' 'BLQ' 'GLQ' 'LwQ' 'Rec' 'Unf']
['ALQ' 'BLQ' 'GLQ' 'LwQ' 'Rec' 'Unf']
['Floor' 'GasA' 'GasW' 'Grav' 'OthW' 'Wall']
['Ex' 'Fa' 'Gd' 'Po' 'TA']
['N' 'Y']
['FuseA' 'FuseF' 'FuseP' 'Mix' 'SBrkr']
['Ex' 'Fa' 'Gd' 'TA']
['Maj1' 'Maj2' 'Min1' 'Min2' 'Mod' 'Sev' 'Typ']
['Ex' 'Fa' 'Gd' 'Po' 'TA']
['2Types' 'Attchd' 'Basment' 'BuiltIn' 'CarPort' 'Detchd']
[1900. 1906. 1908. 1910. 1914. 1915. 1916. 1918. 1920. 1921. 1922. 1923.
 1924. 1925. 1926. 1927. 1928. 1929. 1930. 1931. 1932. 1933. 1934. 1935.
 1936. 1937. 1938. 1939. 1940. 1941. 1942. 1945. 1946. 1947. 1948. 1949.
 1950. 1951. 1952. 1953. 1954. 1955. 1956. 1957. 1958. 1959. 1960. 1961.
 1962. 1963. 1964. 1965. 1966. 1967. 1968. 1969. 1970. 1971. 1972. 1973.
 1974. 1975. 1976. 1977. 1978. 1979. 1980. 1981. 1982. 1983. 1984. 1985.
 1986. 1987. 1988. 1989. 1990. 1991. 1992. 1993. 1994. 1995. 1996. 1997.
 1998. 1999. 2000. 2001. 2002. 2003. 2004. 2005. 2006. 2007. 2008. 2009.
```

```
 2010.]
['Fin' 'RFn' 'Unf']
['Ex' 'Fa' 'Gd' 'Po' 'TA']
['Ex' 'Fa' 'Gd' 'Po' 'TA']
['N' 'P' 'Y']
['Ex' 'Fa' 'Gd']
['GdPrv' 'GdWo' 'MnPrv' 'MnWw']
['Gar2' 'Othr' 'Shed' 'TenC']
[ 1  2  3  4  5  6  7  8  9 10 11 12]
[2006 2007 2008 2009 2010]
['COD' 'CWD' 'Con' 'ConLD' 'ConLI' 'ConLw' 'New' 'Oth' 'WD']
['Abnorml' 'AdjLand' 'Alloca' 'Family' 'Normal' 'Partial']
```

Out[245]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utili
0	1	5	3	65.0	8450	1	0	3	3	
1	2	0	3	80.0	9600	1	0	3	3	
2	3	5	3	68.0	11250	1	0	0	3	
3	4	6	3	60.0	9550	1	0	0	3	
4	5	5	3	84.0	14260	1	0	0	3	

In [246…

```
#another way
#i took categorical variables which had null, addedthose that needed encoding and e
#encoded month year and some numeric variables, so that they just belong to differe
#the result (same as different numbers from 1 to 5 maybe should just mean differen
#bigger value if style is 4th and not 2nd)
encodecolstest= ['MSSubClass','MSZoning','Street','Alley','LotShape','LandContour'

for col in encodecolstest:
    testhouse[col]=le.fit_transform(testhouse[col])
    print(le.classes_)

testhouse.head(5)
```

```
[ 20  30  40  45  50  60  70  75  80  85  90 120 150 160 180 190]
['C (all)' 'FV' 'RH' 'RL' 'RM']
['Grvl' 'Pave']
['Grvl' 'Pave']
['IR1' 'IR2' 'IR3' 'Reg']
['Bnk' 'HLS' 'Low' 'Lvl']
['AllPub']
['Corner' 'CulDSac' 'FR2' 'FR3' 'Inside']
['Gtl' 'Mod' 'Sev']
['Blmngtn' 'Blueste' 'BrDale' 'BrkSide' 'ClearCr' 'CollgCr' 'Crawfor'
 'Edwards' 'Gilbert' 'IDOTRR' 'MeadowV' 'Mitchel' 'NAmes' 'NPkVill'
 'NWAmes' 'NoRidge' 'NridgHt' 'OldTown' 'SWISU' 'Sawyer' 'SawyerW'
 'Somerst' 'StoneBr' 'Timber' 'Veenker']
['Artery' 'Feedr' 'Norm' 'PosA' 'PosN' 'RRAe' 'RRAn' 'RRNe' 'RRNn']
['Artery' 'Feedr' 'Norm' 'PosA' 'PosN']
['1Fam' '2fmCon' 'Duplex' 'Twnhs' 'TwnhsE']
['1.5Fin' '1.5Unf' '1Story' '2.5Unf' '2Story' 'SFoyer' 'SLvl']
[1879 1880 1890 1895 1896 1900 1901 1902 1905 1907 1910 1912 1914 1915
 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929
 1930 1931 1932 1934 1935 1936 1937 1938 1939 1940 1941 1942 1945 1946
 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960
 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974
 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988
 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002
 2003 2004 2005 2006 2007 2008 2009 2010]
[1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963
 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977
 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991
 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005
 2006 2007 2008 2009 2010]
['Flat' 'Gable' 'Gambrel' 'Hip' 'Mansard' 'Shed']
['CompShg' 'Tar&Grv' 'WdShake' 'WdShngl']
['AsbShng' 'AsphShn' 'BrkComm' 'BrkFace' 'CBlock' 'CemntBd' 'HdBoard'
 'MetalSd' 'Plywood' 'Stucco' 'VinylSd' 'Wd Sdng' 'WdShing']
['AsbShng' 'AsphShn' 'Brk Cmn' 'BrkFace' 'CBlock' 'CmentBd' 'HdBoard'
 'ImStucc' 'MetalSd' 'Plywood' 'Stone' 'Stucco' 'VinylSd' 'Wd Sdng'
 'Wd Shng']
['BrkCmn' 'BrkFace' 'None' 'Stone']
['Ex' 'Fa' 'Gd' 'TA']
['Ex' 'Fa' 'Gd' 'Po' 'TA']
['BrkTil' 'CBlock' 'PConc' 'Slab' 'Stone' 'Wood']
['Ex' 'Fa' 'Gd' 'TA']
['Fa' 'Gd' 'Po' 'TA']
['Av' 'Gd' 'Mn' 'No']
['ALQ' 'BLQ' 'GLQ' 'LwQ' 'Rec' 'Unf']
['ALQ' 'BLQ' 'GLQ' 'LwQ' 'Rec' 'Unf']
['GasA' 'GasW' 'Grav' 'Wall']
['Ex' 'Fa' 'Gd' 'Po' 'TA']
['N' 'Y']
['FuseA' 'FuseF' 'FuseP' 'SBrkr']
['Ex' 'Fa' 'Gd' 'TA']
['Maj1' 'Maj2' 'Min1' 'Min2' 'Mod' 'Sev' 'Typ']
['Ex' 'Fa' 'Gd' 'Po' 'TA']
['2Types' 'Attchd' 'Basment' 'BuiltIn' 'CarPort' 'Detchd']
[1895. 1896. 1900. 1910. 1915. 1916. 1917. 1918. 1919. 1920. 1921. 1922.
 1923. 1924. 1925. 1926. 1927. 1928. 1930. 1932. 1934. 1935. 1936. 1937.
 1938. 1939. 1940. 1941. 1942. 1943. 1945. 1946. 1947. 1948. 1949. 1950.
 1951. 1952. 1953. 1954. 1955. 1956. 1957. 1958. 1959. 1960. 1961. 1962.
 1963. 1964. 1965. 1966. 1967. 1968. 1969. 1970. 1971. 1972. 1973. 1974.
 1975. 1976. 1977. 1978. 1979. 1980. 1981. 1982. 1983. 1984. 1985. 1986.
 1987. 1988. 1989. 1990. 1991. 1992. 1993. 1994. 1995. 1996. 1997. 1998.
 1999. 2000. 2001. 2002. 2003. 2004. 2005. 2006. 2007. 2008. 2009. 2010.
 2207.]
['Fin' 'RFn' 'Unf']
```

```
['Fa' 'Gd' 'Po' 'TA']
['Ex' 'Fa' 'Gd' 'Po' 'TA']
['N' 'P' 'Y']
['Ex' 'Gd']
['GdPrv' 'GdWo' 'MnPrv' 'MnWw']
['Gar2' 'Othr' 'Shed']
[ 1  2  3  4  5  6  7  8  9 10 11 12]
[2006 2007 2008 2009 2010]
['COD' 'CWD' 'Con' 'ConLD' 'ConLI' 'ConLw' 'New' 'Oth' 'WD']
['Abnorml' 'AdjLand' 'Alloca' 'Family' 'Normal' 'Partial']
```

Out[246]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	U
0	1461	0	2	80.0	11622	1	0	3	3	
1	1462	0	3	81.0	14267	1	0	0	3	
2	1463	5	3	74.0	13830	1	0	0	3	
3	1464	5	3	78.0	9978	1	0	0	3	
4	1465	11	3	43.0	5005	1	0	0	1	

In [247…

```
# from sklearn.preprocessing import StandardScaler
# mms = MinMaxScaler()
# trainhouse[['LotFrontage','LotArea', 'MasVnrArea','BsmtFinSF1','BsmtFinSF2','Bsm
```

In [248…

```
trainhouse.head()
```

Out[248]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utili
0	1	5	3	65.0	8450	1	0	3	3	
1	2	0	3	80.0	9600	1	0	3	3	
2	3	5	3	68.0	11250	1	0	0	3	
3	4	6	3	60.0	9550	1	0	0	3	
4	5	5	3	84.0	14260	1	0	0	3	

In [249…

```
# # def log (x):
# #     return np.log1p(x)
# # #i created a function to normalize data with log, but with logplus1 because od
# # #an error, it's not possible because it is undefined.if base is biggerthen 1 i

# trainhouse1 = trainhouse[['LotFrontage','LotArea', 'MasVnrArea','BsmtFinSF1','Bs
```

In [250…

```
# trainhouse1.head()
```

In [251…

```
# trainhouse.head()
```

In [252…

```
# trainhouse[['LotFrontage','LotArea', 'MasVnrArea','BsmtFinSF2','Bsm
```

In [253…

```
trainhouse.head()
```

Out[253]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utili
0	1	5	3	65.0	8450	1	0	3	3	
1	2	0	3	80.0	9600	1	0	3	3	
2	3	5	3	68.0	11250	1	0	0	3	
3	4	6	3	60.0	9550	1	0	0	3	
4	5	5	3	84.0	14260	1	0	0	3	

In [254…

```
housevisual = trainhouse.iloc[:,1:81]
housevisual.head()
```

Out[254]:

	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities
0	5	3	65.0	8450	1	0	3	3	0
1	0	3	80.0	9600	1	0	3	3	0
2	5	3	68.0	11250	1	0	0	3	0
3	6	3	60.0	9550	1	0	0	3	0
4	5	3	84.0	14260	1	0	0	3	0

In [255…

```
len(trainhouse.columns)
```

Out[255]:

```
81
```

In [256…

```
# # def log (x):
# #     return np.log1p(x)

# testhouse1 = testhouse[['LotFrontage','LotArea', 'MasVnrArea','BsmtFinSF1','Bsmt
```

In [257…

```
# testhouse1.head()
```

In [258…

```
# testhouse[['LotFrontage','LotArea', 'MasVnrArea','BsmtFinSF1','BsmtFinSF2','Bsmt
```

In [259…

```
# testhouse.head()
```

In [260…

```
len(testhouse.columns)
```

Out[260]:

```
80
```

In [261…

```
#iz train zadanog dataseta X će biti Xtrain koji ću podijeliti na X_train i X_val
#datasetu ću primijeniti crossvalidation umejsto traintestsplita, prije toga izdvo
#iz Y zadanog dataseta test data odvojeni će ostati za testiranje kao X_test, nema

Xa=trainhouse.iloc[:,1:80]
print(Xa)
#izbacila sam kolonu indeksa  i kolonu cijena koja je Y
```

```
      MSSubClass  MSZoning  LotFrontage  LotArea  Street  Alley  LotShape  \
0              5         3         65.0     8450       1      0         3
1              0         3         80.0     9600       1      0         3
2              5         3         68.0    11250       1      0         0
3              6         3         60.0     9550       1      0         0
4              5         3         84.0    14260       1      0         0
...          ...       ...          ...      ...     ...    ...       ...
1455           5         3         62.0     7917       1      0         3
1456           0         3         85.0    13175       1      0         3
1457           6         3         66.0     9042       1      0         3
1458           0         3         68.0     9717       1      0         3
1459           0         3         75.0     9937       1      0         3

      LandContour  Utilities  LotConfig  LandSlope  Neighborhood  Condition1  \
0               3          0          4          0             5           2
1               3          0          2          0            24           1
2               3          0          4          0             5           2
3               3          0          0          0             6           2
4               3          0          2          0            15           2
...           ...        ...        ...        ...           ...         ...
1455            3          0          4          0             8           2
1456            3          0          4          0            14           2
1457            3          0          4          0             6           2
1458            3          0          4          0            12           2
1459            3          0          4          0             7           2

      Condition2  BldgType  HouseStyle  OverallQual  OverallCond  YearBuilt  \
0              2         0           5            7            5        104
1              2         0           2            6            8         77
2              2         0           5            7            5        102
3              2         0           5            7            5         19
4              2         0           5            8            5        101
...          ...       ...         ...          ...          ...        ...
1455           2         0           5            6            5        100
1456           2         0           2            6            6         79
1457           2         0           5            7            9         44
1458           2         0           2            5            6         51
1459           2         0           2            5            6         66

      YearRemodAdd  RoofStyle  RoofMatl  Exterior1st  Exterior2nd  MasVnrType  \
0               53          1         1           12           13           1
1               26          1         1            8            8           2
2               52          1         1           12           13           1
3               20          1         1           13           15           2
4               50          1         1           12           13           1
...            ...        ...       ...          ...          ...         ...
1455            50          1         1           12           13           2
1456            38          1         1            9           10           3
1457            56          1         1            5            5           2
1458            46          3         1            8            8           2
1459            15          1         1            6            6           2

      MasVnrArea  ExterQual  ExterCond  Foundation  BsmtQual  BsmtCond  \
0          196.0          2          4           2         2         3
1            0.0          3          4           1         2         3
2          162.0          2          4           2         2         3
3            0.0          3          4           0         3         1
4          350.0          2          4           2         2         3
...          ...        ...        ...         ...       ...       ...
1455         0.0          3          4           2         2         3
1456       119.0          3          4           1         2         3
1457         0.0          0          2           4         3         1
1458         0.0          3          4           1         3         3
1459         0.0          2          4           1         3         3
```

```
      BsmtExposure  BsmtFinType1  BsmtFinSF1  BsmtFinType2  BsmtFinSF2  \
0                3            2         706             5           0
1                1            0         978             5           0
2                2            2         486             5           0
3                3            0         216             5           0
4                0            2         655             5           0
...            ...          ...         ...           ...         ...
1455             3            5           0             5           0
1456             3            0         790             4         163
1457             3            2         275             5           0
1458             2            2          49             4        1029
1459             3            1         830             3         290

      BsmtUnfSF  TotalBsmtSF  Heating  HeatingQC  CentralAir  Electrical  \
0           150          856        1          0           1           4
1           284         1262        1          0           1           4
2           434          920        1          0           1           4
3           540          756        1          2           1           4
4           490         1145        1          0           1           4
...         ...          ...      ...        ...         ...         ...
1455        953          953        1          0           1           4
1456        589         1542        1          4           1           4
1457        877         1152        1          0           1           4
1458          0         1078        1          2           1           0
1459        136         1256        1          2           1           4

      1stFlrSF  2ndFlrSF  LowQualFinSF  GrLivArea  BsmtFullBath  BsmtHalfBath  \
0          856       854             0       1710             1             0
1         1262         0             0       1262             0             1
2          920       866             0       1786             1             0
3          961       756             0       1717             1             0
4         1145      1053             0       2198             1             0
...        ...       ...           ...        ...           ...           ...
1455       953       694             0       1647             0             0
1456      2073         0             0       2073             1             0
1457      1188      1152             0       2340             0             0
1458      1078         0             0       1078             1             0
1459      1256         0             0       1256             1             0

      FullBath  HalfBath  BedroomAbvGr  KitchenAbvGr  KitchenQual  \
0            2         1             3             1            2
1            2         0             3             1            3
2            2         1             3             1            2
3            1         0             3             1            2
4            2         1             4             1            2
...        ...       ...           ...           ...          ...
1455         2         1             3             1            3
1456         2         0             3             1            3
1457         2         0             4             1            2
1458         1         0             2             1            2
1459         1         1             3             1            3

      TotRmsAbvGrd  Functional  Fireplaces  FireplaceQu  GarageType  \
0                8           6           0            2           1
1                6           6           1            4           1
2                6           6           1            4           1
3                7           6           1            2           5
4                9           6           1            4           1
...            ...         ...         ...          ...         ...
1455             7           6           1            4           1
1456             7           2           2            4           1
1457             9           6           2            2           1
1458             5           6           0            2           1
```

```
1459             6            6            0             2            1

      GarageYrBlt  GarageFinish  GarageCars  GarageArea  GarageQual  \
0              89             1           2         548           4
1              62             1           2         460           4
2              87             1           2         608           4
3              84             2           3         642           4
4              86             1           3         836           4
...           ...           ...         ...         ...         ...
1455           85             1           2         460           4
1456           64             2           2         500           4
1457           29             1           1         252           4
1458           36             2           1         240           4
1459           51             0           1         276           4

      GarageCond  PavedDrive  WoodDeckSF  OpenPorchSF  EnclosedPorch  \
0              4           2           0           61              0
1              4           2         298            0              0
2              4           2           0           42              0
3              4           2           0           35            272
4              4           2         192           84              0
...          ...         ...         ...          ...            ...
1455           4           2           0           40              0
1456           4           2         349            0              0
1457           4           2           0           60              0
1458           4           2         366            0            112
1459           4           2         736           68              0

      3SsnPorch  ScreenPorch  PoolArea  PoolQC  Fence  MiscFeature  MiscVal  \
0             0            0         0       2      2            2        0
1             0            0         0       2      2            2        0
2             0            0         0       2      2            2        0
3             0            0         0       2      2            2        0
4             0            0         0       2      2            2        0
...         ...          ...       ...     ...    ...          ...      ...
1455          0            0         0       2      2            2        0
1456          0            0         0       2      2            2        0
1457          0            0         0       2      0            2     2500
1458          0            0         0       2      2            2        0
1459          0            0         0       2      2            2        0

      MoSold  YrSold  SaleType  SaleCondition
0          1       2         8              4
1          4       1         8              4
2          8       2         8              4
3          1       0         8              0
4         11       2         8              4
...      ...     ...       ...            ...
1455       7       1         8              4
1456       1       4         8              4
1457       4       4         8              4
1458       3       4         8              4
1459       5       2         8              4

[1460 rows x 79 columns]
```

In [262…

```
X=trainhouse.iloc[:,1:80]
print(X)
```

```
      MSSubClass  MSZoning  LotFrontage  LotArea  Street  Alley  LotShape  \
0              5         3         65.0     8450       1      0         3
1              0         3         80.0     9600       1      0         3
2              5         3         68.0    11250       1      0         0
3              6         3         60.0     9550       1      0         0
4              5         3         84.0    14260       1      0         0
...          ...       ...          ...      ...     ...    ...       ...
1455           5         3         62.0     7917       1      0         3
1456           0         3         85.0    13175       1      0         3
1457           6         3         66.0     9042       1      0         3
1458           0         3         68.0     9717       1      0         3
1459           0         3         75.0     9937       1      0         3

      LandContour  Utilities  LotConfig  LandSlope  Neighborhood  Condition1  \
0               3          0          4          0             5           2
1               3          0          2          0            24           1
2               3          0          4          0             5           2
3               3          0          0          0             6           2
4               3          0          2          0            15           2
...           ...        ...        ...        ...           ...         ...
1455            3          0          4          0             8           2
1456            3          0          4          0            14           2
1457            3          0          4          0             6           2
1458            3          0          4          0            12           2
1459            3          0          4          0             7           2

      Condition2  BldgType  HouseStyle  OverallQual  OverallCond  YearBuilt  \
0              2         0           5            7            5        104
1              2         0           2            6            8         77
2              2         0           5            7            5        102
3              2         0           5            7            5         19
4              2         0           5            8            5        101
...          ...       ...         ...          ...          ...        ...
1455           2         0           5            6            5        100
1456           2         0           2            6            6         79
1457           2         0           5            7            9         44
1458           2         0           2            5            6         51
1459           2         0           2            5            6         66

      YearRemodAdd  RoofStyle  RoofMatl  Exterior1st  Exterior2nd  MasVnrType  \
0               53          1         1           12           13           1
1               26          1         1            8            8           2
2               52          1         1           12           13           1
3               20          1         1           13           15           2
4               50          1         1           12           13           1
...            ...        ...       ...          ...          ...         ...
1455            50          1         1           12           13           2
1456            38          1         1            9           10           3
1457            56          1         1            5            5           2
1458            46          3         1            8            8           2
1459            15          1         1            6            6           2

      MasVnrArea  ExterQual  ExterCond  Foundation  BsmtQual  BsmtCond  \
0          196.0          2          4           2         2         3
1            0.0          3          4           1         2         3
2          162.0          2          4           2         2         3
3            0.0          3          4           0         3         1
4          350.0          2          4           2         2         3
...          ...        ...        ...         ...       ...       ...
1455         0.0          3          4           2         2         3
1456       119.0          3          4           1         2         3
1457         0.0          0          2           4         3         1
1458         0.0          3          4           1         3         3
1459         0.0          2          4           1         3         3
```

```
      BsmtExposure  BsmtFinType1  BsmtFinSF1  BsmtFinType2  BsmtFinSF2  \
0                3            2         706             5           0
1                1            0         978             5           0
2                2            2         486             5           0
3                3            0         216             5           0
4                0            2         655             5           0
...            ...          ...         ...           ...         ...
1455             3            5           0             5           0
1456             3            0         790             4         163
1457             3            2         275             5           0
1458             2            2          49             4        1029
1459             3            1         830             3         290

      BsmtUnfSF  TotalBsmtSF  Heating  HeatingQC  CentralAir  Electrical  \
0           150          856        1          0           1           4
1           284         1262        1          0           1           4
2           434          920        1          0           1           4
3           540          756        1          2           1           4
4           490         1145        1          0           1           4
...         ...          ...      ...        ...         ...         ...
1455        953          953        1          0           1           4
1456        589         1542        1          4           1           4
1457        877         1152        1          0           1           4
1458          0         1078        1          2           1           0
1459        136         1256        1          2           1           4

      1stFlrSF  2ndFlrSF  LowQualFinSF  GrLivArea  BsmtFullBath  BsmtHalfBath  \
0          856       854             0       1710             1             0
1         1262         0             0       1262             0             1
2          920       866             0       1786             1             0
3          961       756             0       1717             1             0
4         1145      1053             0       2198             1             0
...        ...       ...           ...        ...           ...           ...
1455       953       694             0       1647             0             0
1456      2073         0             0       2073             1             0
1457      1188      1152             0       2340             0             0
1458      1078         0             0       1078             1             0
1459      1256         0             0       1256             1             0

      FullBath  HalfBath  BedroomAbvGr  KitchenAbvGr  KitchenQual  \
0            2         1             3             1            2
1            2         0             3             1            3
2            2         1             3             1            2
3            1         0             3             1            2
4            2         1             4             1            2
...        ...       ...           ...           ...          ...
1455         2         1             3             1            3
1456         2         0             3             1            3
1457         2         0             4             1            2
1458         1         0             2             1            2
1459         1         1             3             1            3

      TotRmsAbvGrd  Functional  Fireplaces  FireplaceQu  GarageType  \
0                8           6           0            2           1
1                6           6           1            4           1
2                6           6           1            4           1
3                7           6           1            2           5
4                9           6           1            4           1
...            ...         ...         ...          ...         ...
1455             7           6           1            4           1
1456             7           2           2            4           1
1457             9           6           2            2           1
1458             5           6           0            2           1
```

```
1459             6            6            0             2            1

      GarageYrBlt  GarageFinish  GarageCars  GarageArea  GarageQual  \
0              89             1           2         548           4
1              62             1           2         460           4
2              87             1           2         608           4
3              84             2           3         642           4
4              86             1           3         836           4
...           ...           ...         ...         ...         ...
1455           85             1           2         460           4
1456           64             2           2         500           4
1457           29             1           1         252           4
1458           36             2           1         240           4
1459           51             0           1         276           4

      GarageCond  PavedDrive  WoodDeckSF  OpenPorchSF  EnclosedPorch  \
0              4           2           0           61              0
1              4           2         298            0              0
2              4           2           0           42              0
3              4           2           0           35            272
4              4           2         192           84              0
...          ...         ...         ...          ...            ...
1455           4           2           0           40              0
1456           4           2         349            0              0
1457           4           2           0           60              0
1458           4           2         366            0            112
1459           4           2         736           68              0

      3SsnPorch  ScreenPorch  PoolArea  PoolQC  Fence  MiscFeature  MiscVal  \
0             0            0         0       2      2            2        0
1             0            0         0       2      2            2        0
2             0            0         0       2      2            2        0
3             0            0         0       2      2            2        0
4             0            0         0       2      2            2        0
...         ...          ...       ...     ...    ...          ...      ...
1455          0            0         0       2      2            2        0
1456          0            0         0       2      2            2        0
1457          0            0         0       2      0            2     2500
1458          0            0         0       2      2            2        0
1459          0            0         0       2      2            2        0

      MoSold  YrSold  SaleType  SaleCondition
0          1       2         8              4
1          4       1         8              4
2          8       2         8              4
3          1       0         8              0
4         11       2         8              4
...      ...     ...       ...            ...
1455       7       1         8              4
1456       1       4         8              4
1457       4       4         8              4
1458       3       4         8              4
1459       5       2         8              4

[1460 rows x 79 columns]
```

In [263…

```
Xa.head()
```

Out[263]:

	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities
0	5	3	65.0	8450	1	0	3	3	0
1	0	3	80.0	9600	1	0	3	3	0
2	5	3	68.0	11250	1	0	0	3	0
3	6	3	60.0	9550	1	0	0	3	0
4	5	3	84.0	14260	1	0	0	3	0

In [264…

```
len(X.columns)
```

Out[264]:

```
79
```

In [265…

```
Ya=trainhouse.iloc[:,80]
print(Ya)
```

```
0       208500
1       181500
2       223500
3       140000
4       250000
         ...
1455    175000
1456    210000
1457    266500
1458    142125
1459    147500
Name: SalePrice, Length: 1460, dtype: int64
```

In [266…

```
Y=trainhouse.iloc[:,80]
print(Y)
#sve iz traindataseta i x i y
```

```
0       208500
1       181500
2       223500
3       140000
4       250000
         ...
1455    175000
1456    210000
1457    266500
1458    142125
1459    147500
Name: SalePrice, Length: 1460, dtype: int64
```

In [267…

```
X_test1 = testhouse.iloc[:,1:80]
#izbacila sam kolonu indeksa samo
```

In [268…

```
X_test1.head()
```

Out[268]:

	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities
0	0	2	80.0	11622	1	0	3	3	0
1	0	3	81.0	14267	1	0	0	3	0
2	5	3	74.0	13830	1	0	0	3	0
3	5	3	78.0	9978	1	0	0	3	0
4	11	3	43.0	5005	1	0	0	1	0

In [269…
```
len(X_test1.columns)
```

Out[269]:
```
79
```

In [270…
```
#X = X.astype(int)
```

In [271…
```
#Y = Y.astype(int)
```

In [272…
```
#X_test1 = X_test1.astype(int)
```

In [273…
```
#I will split X on train i val part using cross validation, not train test split he
```

In [274…
```
#kf = KFold(n_splits=5)
#kf.get_n_splits(trainhouse)
#X i Y su mi dio train data koji zelim podilit na cross val dio, xtest nisam jos d

#SVE OK, PODIJELIS KOLONE NA X I Y I ONDA MEDJU NJIJMA CROSS VALIDATION, TJ NA X-U
#SAMO SAM JA OVDJE XI Y NAPRAVILA OD TRAIN DATASETA DA BUDU ZAPRAVO TRAIN I VALIDA
#ZVATI VALIDATION, CESTO TAKO I IDE
#DA SAM UZELA CIJELI TRAINDATA NEKE TABLICE ZA TRAIN I TESTDATA ZA TEST, DA IMAM Y
#DIJELILA PREKO CROSS VALIDATIONA  PO INDEKSIMA TRAINA A TESTIRALA NA TESTU
# ne znam jesam prvo morala podilit cross val pa izdvojit y onda mozda, ne i u prir
#prije izdvojila iz train dataseta ali ista stvar
```

In [275…
```
# #print(kf)
# k=5
# kf = KFold(n_splits=5, random_state=None, shuffle=False)
# #ovo je sam objekt
# # for i, (train_index, val_index) in enumerate(kf.split(trainhouse)):
# #         print(f"Fold {i}:")
# #         print(f"  Train: index={train_index}")
# #         print(f"  Val:  index={val_index}")

#         #preko x-a dobijas indekse za x i y dio
```

In [276…
```
# kf = KFold(n_splits=5)
# kf.get_n_splits(trainhouse)
# #X i Y su mi dio train data koji zelim podilit na cross val dio, xtest nisam jos
# #ne znam jesam prvo morala podilit cross val pa izdvojit y onda mozda
# #ne dole prikaz bilo bi isto isti indeksi
```

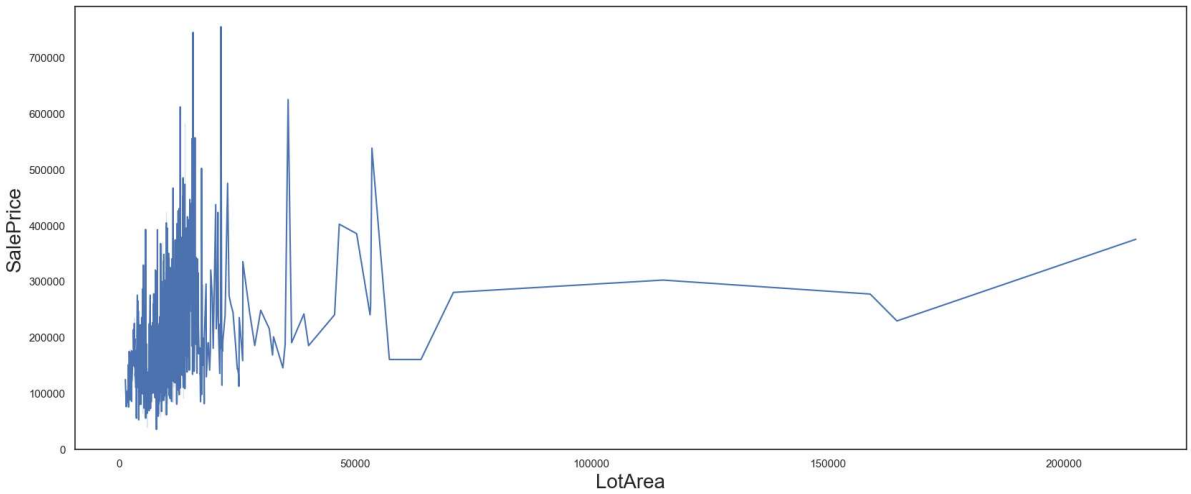
In [277…
```
# print(kf)
# KFold(n_splits=5, random_state=None, shuffle=False)
# for i, (train_index, test_index) in enumerate(kf.split(trainhouse)):
#         print(f"Fold {i}:")
#         print(f"  Train: index={train_index}")
#         print(f"  Test:  index={test_index}")
```

In [278…

```
#LotArea SalePrice
```

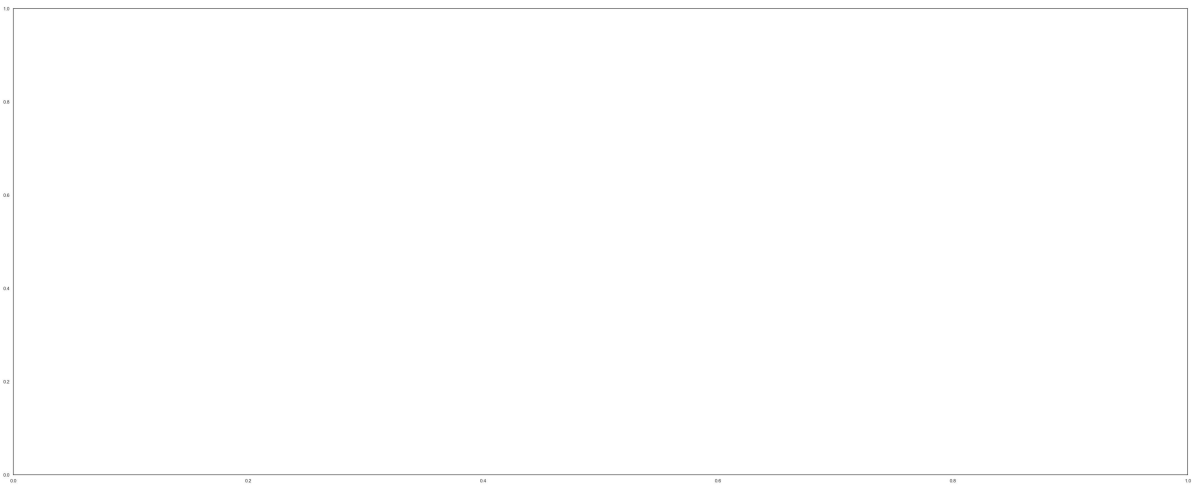
In [279…

```
fig, ax = plt.subplots(figsize = (20, 8))
sns.lineplot(data=housevisual, x='LotArea', y='SalePrice')
plt.xlabel('LotArea', fontsize=20)
plt.ylabel('SalePrice',fontsize=20)
plt.pause(0.001)
plt.show()
```



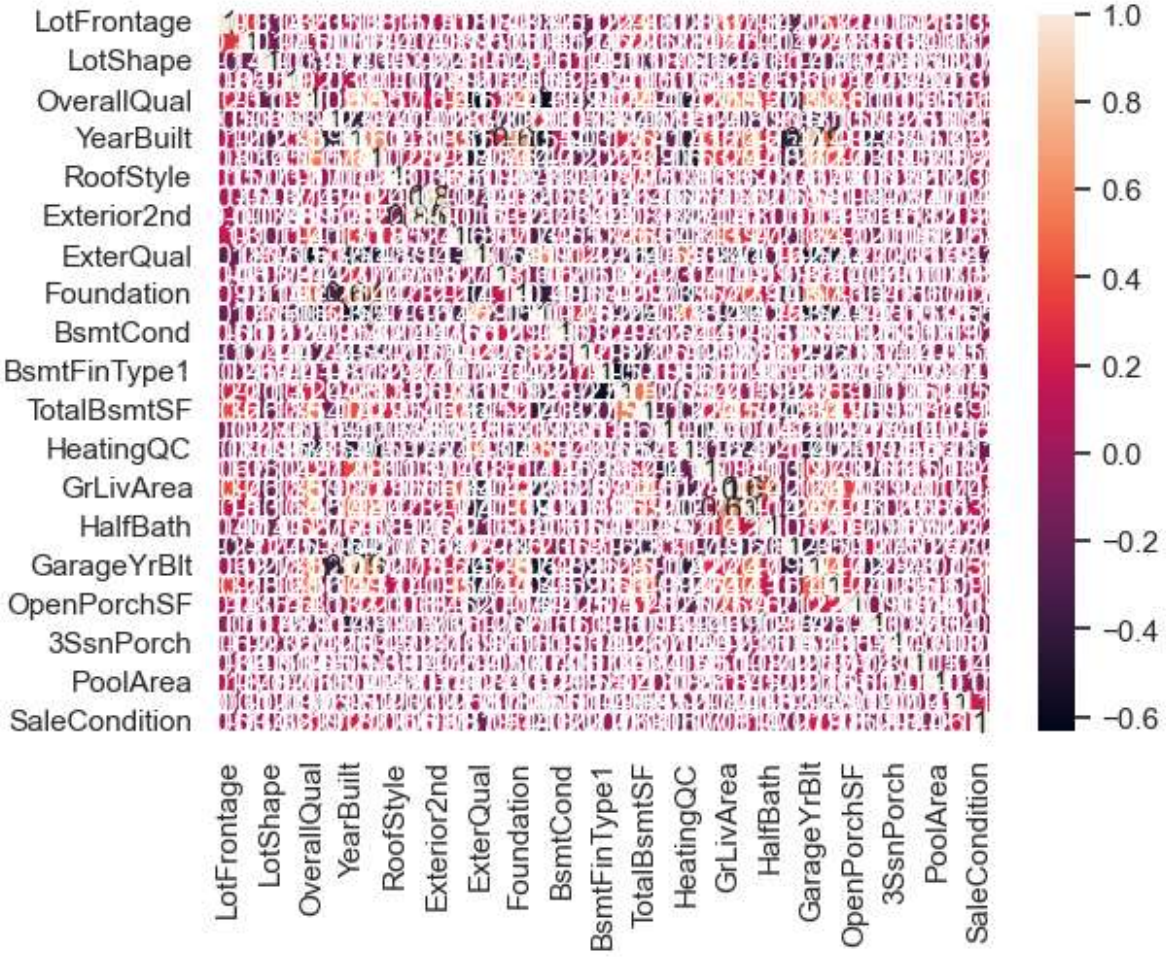
In [280…

```
# df = pd.DataFrame(data=housevisual(size=(100, 26)), columns = ['MSSubClass','MSZ
housevisual2= trainhouse[['LotFrontage','LotArea','LotShape','Neighborhood','Overa

corr=housevisual2.corr()
plt.subplots(figsize=(50,20))
plt.pause(0.001)
sns.heatmap(corr,annot=True)
```



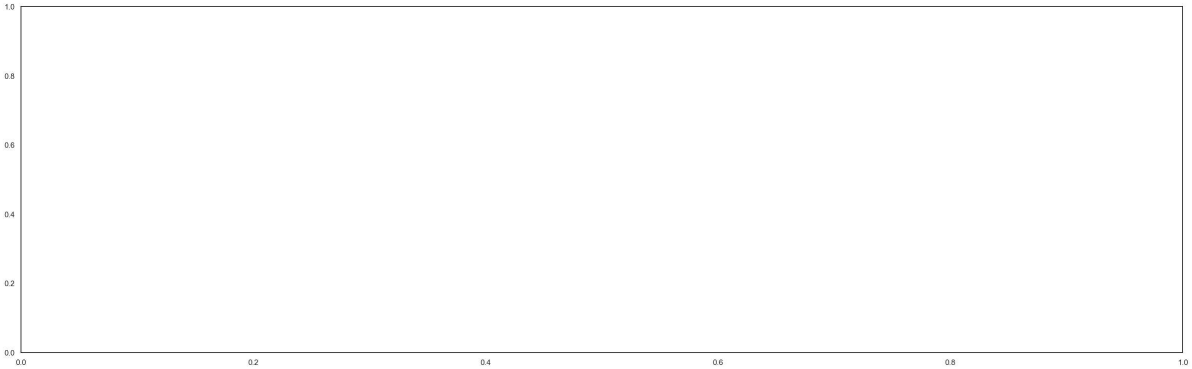
Out[280]:

```
<AxesSubplot:>
```



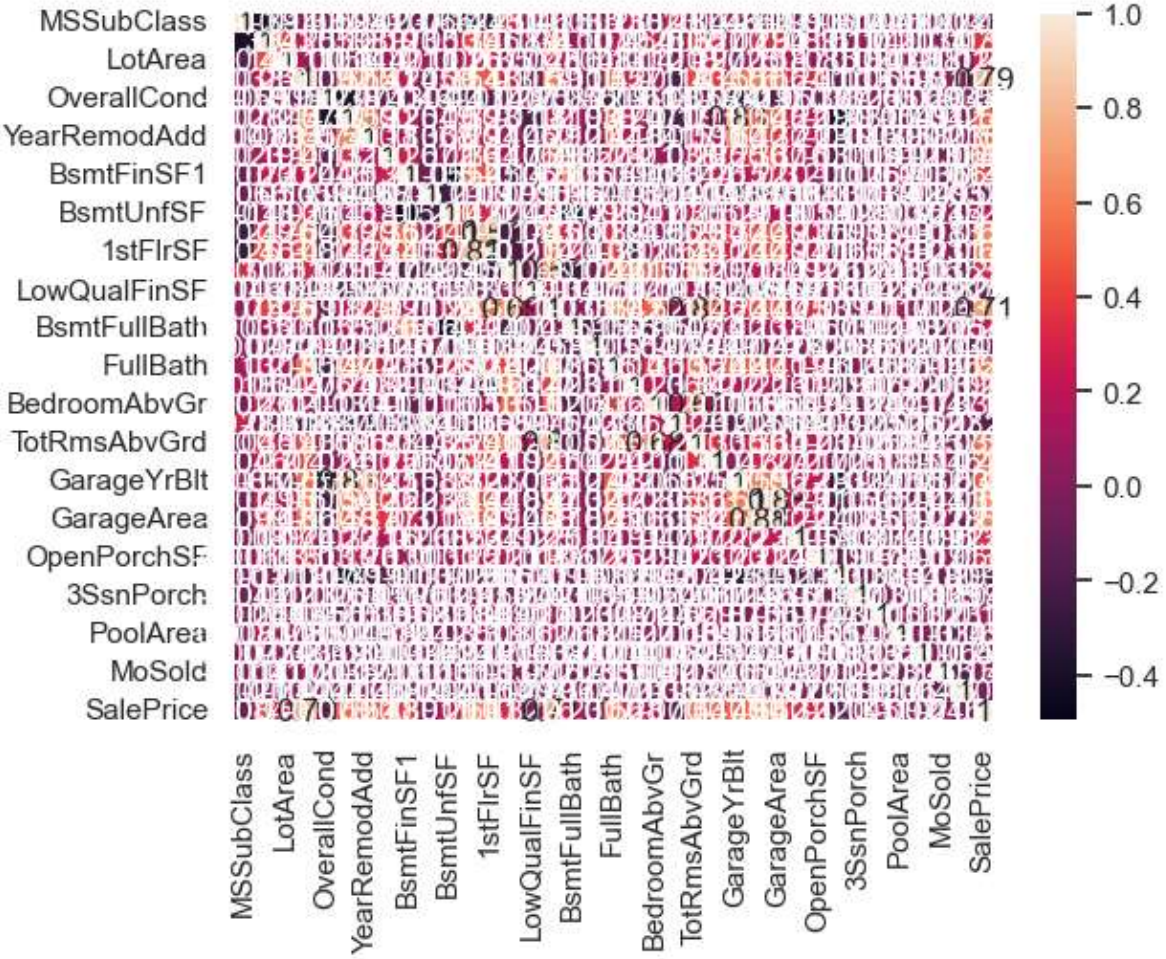
In [281…

```
corr=trainvisual.corr()
plt.subplots(figsize=(30,9))
plt.pause(0.001)
sns.heatmap(corr,annot=True)
#first, not normalized data
```



Out[281]:

```
<AxesSubplot:>
```

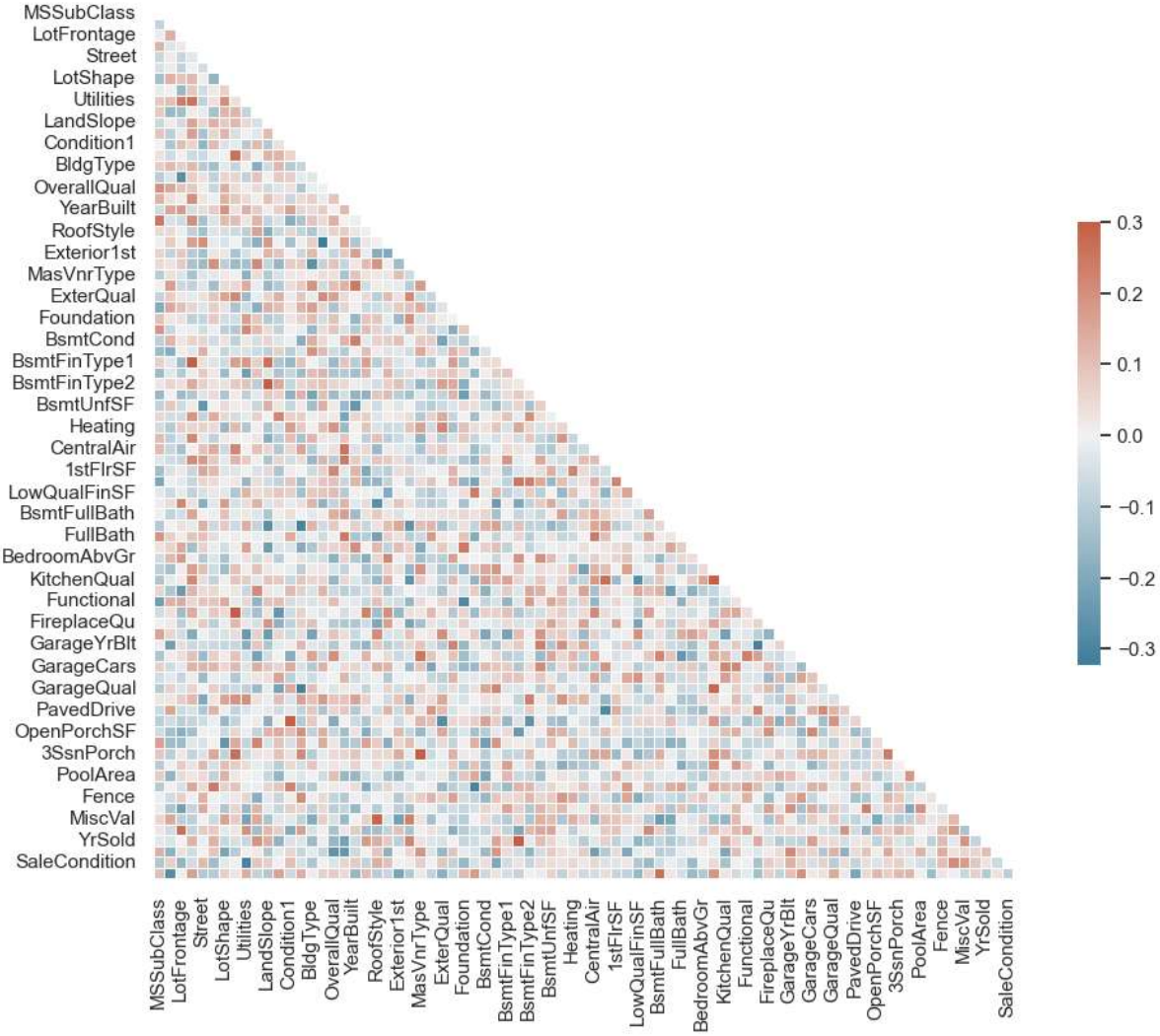


In [282…

```
sns.set_theme(style="white")

# Generate a large random dataset
rs = np.random.RandomState(33)
d = pd.DataFrame(data=rs.normal(size=(100, 80)),
                 columns=trainvisual.columns. values)

#100*80 radi broja kolona ukupnih

# Compute the correlation matrix
corr = d.corr()

# Generate a mask for the upper triangle
mask = np.triu(np.ones_like(corr, dtype=bool))

# Set up the matplotlib figure
f, ax = plt.subplots(figsize=(11, 9))

# Generate a custom diverging colormap
cmap = sns.diverging_palette(230, 20, as_cmap=True)

# Draw the heatmap with the mask and correct aspect ratio
sns.heatmap(corr, mask=mask, cmap=cmap, vmax=.3, center=0,
            square=True, linewidths=.5, cbar_kws={"shrink": .5})

#source:https://seaborn.pydata.org/examples/many_pairwise_correlations.html
```

Out[282]:

```
<AxesSubplot:>
```



In [ ]:

In [ ]:

In [283…

```
#kf.split( )

k = 5
kf = KFold(n_splits=k, random_state=None)

#po indeksima dijelis xcili train i y kolonu
```

In [284…

```
for train_index , test_index in kf.split(X):
    X_train , X_test = X.iloc[train_index,:],X.iloc[test_index,:]
    Y_train , Y_test = Y[train_index] , Y[test_index]

print(f"  Train: index={train_index}")
print(f"  Test:  index={test_index}")
```

```
  Train: index=[   0    1    2 ... 1165 1166 1167]
  Test:  index=[1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1
181
 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195
 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209
 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223
 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237
 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251
 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265
 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279
 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293
 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307
 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321
 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335
 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349
 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363
 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377
 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391
 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405
 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419
 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433
 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447
 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459]
```

In [285…

```
X_train.shape
```

Out[285]:

```
(1168, 79)
```

In [286…

```
Y_train.shape
```

Out[286]:

```
(1168,)
```

In [287…

```
X_test.shape
```

Out[287]:

```
(292, 79)
```

In [288…

```
Y_test.shape
```

Out[288]:

```
(292,)
```

In [289…

```
# for train_index, val_index in kf.split(X,Y)
#     break
```

In [290…

```
#train_index
```

In [291…

```
#val_index
```

In [292…

```
# for train_index, val_index in kf.split(trainhouse):
#     X_train, X_val = X.iloc[train_index,:], X.iloc[train_index,:]
#     Y_train, Y_val = Y[val_index], Y[val_index]

#     ili
# X_tr=trainhouse.loc[train_index]
# Y_tr=trainhouse.loc[train_index]

# X_vl=trainhouse.loc[val_index]
# Y_vl=trainhouse.loc[val_index]
```

In [293…

```
model = Ridge()
#ja mislim da on već u sebi ima cross validation pa ne moze izvodit
```

In [294…
```
model.fit(X_train, Y_train)
```
Out[294]:
```
Ridge()
```

In [295…
```
ytrainpredict = model.predict(X_test)
```

In [296…
```
#Y_test.to_numpy()
```

In [297…
```
#Y_val ranije gore definirala
```

In [ ]:

In [298…
```
r_squared =  r2_score(Y_test, ytrainpredict)
```

In [299…
```
print(r_squared)
```
```
0.6492643697016399
```

In [300…
```
model.score(X_test, Y_test)*100
```
Out[300]:
```
64.92643697016399
```

In [301…
```
from sklearn.ensemble import RandomForestRegressor

model1 = RandomForestRegressor(n_estimators = 1000)
```

In [302…
```
model1.fit(X_train, Y_train)
```
Out[302]:
```
RandomForestRegressor(n_estimators=1000)
```

In [303…
```
model.score(X_test, Y_test)*100
#mean_squared_error(Y_test, ytrainpredict,squared=False)
```
Out[303]:
```
64.92643697016399
```

In [ ]:

In [304…
```
# from sklearn import datasets
# X_train,Y_train = datasets.load_X(return_X_train_Y_train=True)
from xgboost import XGBRegressor
from sklearn.model_selection import cross_val_score
scores = cross_val_score(XGBRegressor(objective='reg:squarederror'), X_train, Y_tra
(-scores)**0.5
#root mean squared error as negative square root of the five scores;
# x**0.5 means x to the 1/2 power; i.e. square root
#resulta along different cross-val iterations, just to try
```
Out[304]:
```
array([27802.77324056, 25537.791917  , 44489.46727796, 28370.67976971,
       21236.06971898])
```

In [ ]:

In [305…
```
from sklearn.ensemble import GradientBoostingRegressor
model2 = GradientBoostingRegressor(n_estimators = 100, max_depth = 5)
```

In [306…
```
model2.fit(X_train, Y_train)
```

Out[306]:
```
GradientBoostingRegressor(max_depth=5)
```

In [307…
```
model2.score(X_test, Y_test)*100
```

Out[307]:
```
82.17775147727038
```

In [308…
```
ytrainpredict2 = model2.predict(X_test)
```

In [309…
```
mean_squared_error(Y_test, ytrainpredict2,squared=False)
```

Out[309]:
```
34569.46080746758
```

In [310…
```
r_squared2 =  r2_score(Y_test, ytrainpredict2)
```

In [311…
```
print(r_squared2)
```

```
0.8217775147727038
```

In [312…
```
#gradient umist oxgboost
#na statquest ucila 0 i 1 umisto nan za xgboost trazi
#ode za nan modei  median pa gradient bolje
```

In [ ]:

In [313…
```
# # Alternative code to silence potential errors
# scores = cross_val_score(XGBRegressor(objective='reg:squarederror'), X_train, Y_t
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: