

ML/DL Study

f.constant(['enetf.cons

Lookup.KeyValue

Code Review

Code Review

Exploring Data

df.head()

le	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	 PoolArea	PoolQC	Fence	MiscFeature	MiscVal	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	60	RL	65.0	8450	Pave	NaN	Reg	LvI	AllPub	 0	NaN	NaN	NaN	0	2	2008	WD	Normal	208500
1	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	 0	NaN	NaN	NaN	0	5	2007	WD	Normal	181500
2	60	RL	68.0	11250	Pave	NaN	IR1	LvI	AllPub	 0	NaN	NaN	NaN	0	9	2008	WD	Normal	223500
3 4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub	 0	NaN	NaN	NaN	0	2	2006	WD	Abnorml	140000
4	60	RL	84.0	14260	Pave	NaN	IR1	LvI	AllPub	 0	NaN	NaN	NaN	0	12	2008	WD	Normal	250000

5 rows × 81 columns

df = df.drop('Id', axis=1)
df.head(3)

Pythor

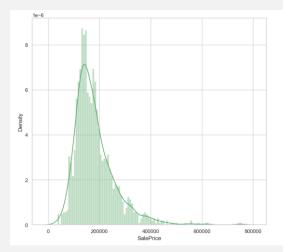
	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	Condition1	Condition2	BldgType	HouseStyle	OverallQual	OverallCond	YearBuilt
0	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm	1Fam	2Story	7	5	2003
1	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	FR2	Gtl	Veenker	Feedr	Norm	1Fam	1Story	6	8	1976
2	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm	1Fam	2Story	7	5	2001

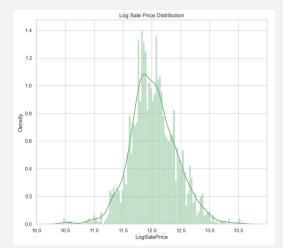
df = df.dropna(axis=1, how='any')

```
MSSubClass
                                MSSubClass
MSZoning
                  0
                                MSZoning
LotFrontage
                 259
                                LotArea
LotArea
                                Street
Street
                                LotShape
MoSold
                                MoSold.
YrSold
                                YrSold
SaleType
                                SaleType
SaleCondition
                                SaleCondition
SalePrice
                                SalePrice
Length: 80, dtype: int64
                                Length: 61, dtype: int64
```

```
def min_max_scaling(data):
       min val = data.min()
       max_val = data.max()
       scaled data = (data - min_val) / (max_val - min_val)
       return scaled data
   df['LogSalePrice'] = min_max_scaling(df['LogSalePrice'])
   print(df['LogSalePrice'].describe())
 ✓ 0.0s
                                                                                                         Python
count
         1460.000000
mean
            0.508683
std
            0.129936
            0.000000
min
25%
            0.427702
50%
            0.501349
75%
            0.589900
            1.000000
```

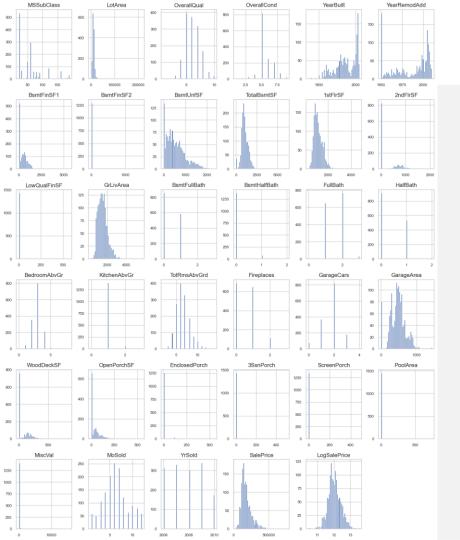
df['LogSalePrice'] = np.log(df['SalePrice'])





Name: LogSalePrice, dtype: float64





MSSubClass	1 01 -01-01004002 02 -002 01 -0.1-0.07.006000-0.1 -0.2 -0.1-0.06 -0.2 -0.04 -0.2 03 02 -0.03 0.2 -0.2 -0.08-0.07 0.1 0.08 0.05 -0.04 0.03 0.05 0.08	1.0		SalePrice		
MSZoning			MSSubClass	0.03		
	-0.1 -0.1 1 03 01 0060 0080 050 005 01 0.04 0.080 0080 09 01 -0.04 02 0.05 02 0.01 0.2 0.1 0.07 0-05 0.1 -0.01 0.1 0.2 0.1 0.1 0.07 0.06 0.2		MSZoning	-0.04		
LotShape	0.004-0.1 0.3 1 0.3 0.07-0.03 0.2 -0.050.006-0.08-0.09-0.1 -0.1 0.1 0.002.0 1 -0.1 0.1 0.06 0.2 0.05 0.2 0.1 0.03 -0.1 0.07 0.2 0.2 0.1 0.1 0.09 0.2		LotArea			
LotConfig	002-008 0.1 0.3 1 -0.020.03 0.04 0.03 0.07-0.05-0.04-0.02-0.03 0.03-0.03 0.05 0.02 0.06 0.03 0.04 0.03 0.02-0.002 0.08 0.04 0.09 0.05 0.08 0.07 0.09	- 0.8	LotShape			
Neighborhood	0.2 -0.1 0.06 0.07-0.02 1 -0.05 0.07-0.04 0.05 0.05 0.06 -0.01 -0.1 0.02-0.010 0.04-0.03 0.1 0.1 0.01 0.04 0.06 -0.08 0.06 0.06 0.2 0.04 0.05 0.07 0.05 0.1		LotConfig			
HouseStyle	-0.02-0.1-0.00-0.03 0 0 3 0 .0 5		Neighborhood			
OverallQua			HouseStyle	-0.1		
OverallCond	01-0009 00=005 003 004 009 02 1 008 02 02 03 005 02 02 01 02-004 02 005 03 0.060 000009 02 007 03 02 0.07 0.2 0.2	- 0.6	OverallQual	0.8		
RoofStyle	-0.07-0.07 0.1 0.0060.07 0.05 0.07 0.06 0.08 1 0.1 0.2 0.1 0.1 0.1 -0.04 0.1 0.04 0.2 -0.1 0.09 0.07-0.05-0.04-0.02 0.1 0.03 0.1 0.06 0.07 0.1 0.04 0.1		OverallCond	-0.2		
Exterior1s	0060 005004 -0.08-0.05 0.05 0.2 -0.2 0.2 0.1 1 0.7 0.3 0.4 0.04 -0.2 -0.1 0.3 0.04 -0.1 -0.09-0.01 -0.2 -0.1 -0.009 0.2 -0.1 0.06 -0.2 -0.2 -0.07 -0.1 -0.2		RoofStyle	0.1		
Exterior2nd	0006001008-009-004006 02 -03 02 02 07 1 04 04 005 -02 -0.1 04 004 -0.1 -0.1 -0.03 -0.3 -0.1 -0.02 0.3 -0.1 0.04 -0.2 -0.2 -0.1 -0.1 -0.2		Exterior1st	-0.2		
ExterQua	-0.1 -0.03 0.03 -0.1 -0.02 -0.01 0.2 -0.5 0.2 0.1 0.3 0.4 1 0.4 -0.01 -0.2 0.2 0.4 -0.2 -0.2 0.3 +0.05 -0.5 0.1 +0.02 0.5 0.2 +0.09 -0.4 +0.3 0.1 +0.2 +0.4 0.1 +0.2 +0.2 0.4 +0.2 +0.2 +0.3 0.1 +0.2 +0.3 0.1 +0.2 +0.3 0.1 +0.2 +0.3 0.1 +0.2 +0.3 0.1 +0.2 +0.3 0.1 +0.2 +0.3 0.1 +0.2 +0.3 0.1 +0.2 +0.3 0.1 +0.2 +0.3 0.1 +0.3 +0.3 0.1 +0.3 +0.3 +0.3 0.1 +0.3 +0.3 +0.3 +0.3 +0.3 +0.3 +0.3 +0.3	- 0.4	Exterior2nd	-0.2		
Foundation	-02 -0.1 009 -0.1 -0.03 -0.1 0.2 -0.6 0.3 0.1 0.4 0.4 0.4 1 0.003 -0.3 0.3 0.5 -0.1 0.2 -0.3 -0.04 -0.5 -0.1 0.1 0.3 -0.2 -0.05 -0.4 -0.3 -0.2 -0.5		ExterQual	-0.4		
BsmtFinSF1	-0.1 -0.1 0.1 0.1 0.06 0.02 0.2 0.2 -0.05 0.1 0.04 0.05 -0.010 0.00 1 -0.6 0.4 -0.2 0.09 0.7 0.01 -0.06 -0.2 -0.02 -0.08 0.2 0.2 0.2 0.2 0.0 0.4		Foundation	-0.5	Co	
BsmtUnfSF	0.06 0.06 0.04 0.00 0.03 0.01 0.04 03 0.2 0.04 0.2 0.2 0.2 0.3 0.6 1 0.4 0.2 0.3 0.1 0.2 0.4 0.3 0.08 0.0 0.1 0.2 0.03 0.2 0.2 0.003 0.1 0.2		BsmtFinSF1	0.4	rrelati	
TotalBsmtSF	-02-008 02 01 0030,004 02 05 -02 01 -0.1 -0.1 -0.2 -0.3 04 04 1 -0.3 08 -0.4 03 03 -0.2 -0.1 -0.1 01 03 04 05 02 02 06	- 0.2	BsmtUnfSF	0.2	ion wi	
HeatingQC	0.04-0.02-0.05 -0.1 -0.03-0.04 0.2 -0.5 0.1 0.04 0.3 0.4 0.4 0.5 -0.06 -0.2 -0.3 1 -0.2 -0.1 -0.3 -0.09 -0.4 -0.1 0.08 0.3 -0.2 -0.1 -0.3 -0.3 -0.1 -0.2 -0.5		TotalBsmtSF	0.6	₿ Sa	
1stFlrSF	-02 -0.1 02 01 005 003 03 04 02 02 004 004 -02 -0.1 04 03 08 -02 1 0.5 04 02 03 -0.3 -0.1 -0.07 02 03 0.4 0.5 02 02 06		HeatingQC	-0.5	lePric	
2ndFlrSF	03 01 001 006 002 01 07 02 0.04 0.1 0.1 0.1 0.2 0.2 0.2 0.1 0.4 0.1 0.5 1 08 0.2 03 0.6 0.4 0.1 0.5 0.1 0.06 0.010.003 0.2 0.1		1stFIrSF	0.6	ď	
GrLivArea	02 001 02 02 006 01 -0.4 06 -0.2 009-0.09-0.1 -0.3 -0.3 009 02 03 -0.3 0.4 06 1 003 06 04 04 -0.2 0.7 0.4 0.5 0.4 02 0.3 0.7	- 0.0	2ndFlrSF	0.1		
smtFullBath	-0.03-0.08 01 0.05 003 001 02 01 -0.05 007 -0.01-0.03-0.05-0.04 07 -0.4 03 -0.09 02 -0.2 003 1 -0.05-0.04 -0.1 -0.07-0.07 0.1 0.1 0.2 0.2 0.09 03		GrLivArea	0.7		
FullBath	02 002 007 02 004 004 -02 06 -0.3 -0.5 -0.2 -0.3 -0.5 -0.5 0.01 03 03 -0.4 03 03 06 -0.05 1 007 02 -0.3 0.5 0.2 0.5 0.4 0.2 0.3 0.5		BsmtFullBath			
HalfBath	02 -0.0376-05 01 003 006 -0.5 02 -0.06-0.04 -0.1 -0.1 -0.1 -0.06-0.08 -0.2 -0.1 -0.3 0.6 0.4 -0.04 0.7 1 0.2 -0.08 0.4 0.2 0.1 007 008 0.2 0.2		FullBath	0.5		
lroomAbvGr	0.080.05 0.1 0.03 0.02 0.08 0.2 0.03 0.00 0.020 0.090 0.2 0.02 0.1 0.2 0.03 0.1 0.08 0.1 0.4 0.4 0.1 0.2 0.2 1 1 0.07 0.5 0.00 0.03 0.04 0.03 0.02 0.00	0.2	HalfBath	0.2 -0.		
	0.07-0.05-0.01-0.1-0.0020.06-0.1-0.3-0.09-0.1-0.2-0.3-0.5-0.3-0.02-0.1-0.1-0.3-0.07-0.1-0.2-0.07-0.3-0.08-0.07-11-0.1-0.04-0.3-0.2-0.1-0.1-0.2		BedroomAbvGr	004		
msAbvGro	01-000501 0070008006 04 04 02 003-01 0.1 02 02 008 02 01 02 02 05 07 007 05 04 05 01 1 02 03 03 01 02 04		KitchenQual	-0.2		
Fireplaces	008 02 02 02 004 02 002 04 007 01 006 004 009 005 02 003 03 0.1 03 01 04 01 02 02 0004 0.04 02 1 03 02 02 02 04		TotRmsAbvGrd	0.4		
SarageCars	005-006 0.1 0.2 009 004-006 06 0.3 006 0.2 0.2 0.4 0.4 0.2 0.2 0.4 0.3 0.4 006 0.5 0.1 0.5 0.1 0.03 0.3 0.3 0.3 1 0.9 0.2 0.3 0.7	0.4	Fireplaces	0.4		
SarageArea	0.04-0.02 0.1 0.1 0.05-0.05-0.02 0.5 -0.2 0.07 -0.2 -0.2 -0.3 -0.3 0.2 0.2 0.5 -0.3 0.5 0.01 0.4 0.2 0.4 0.07 -0.04-0.2 0.3 0.2 0.9 1 0.2 0.2 0.6		GarageCars	0.7		
oodDeckSF	003 -0.1 007 0.1 008 0070007 02 -0.07 0.1 -0.07 -0.1 -0.1 -0.2 02 -0.003 0.2 -0.1 02 0003 0.2 0.2 0.2 008 -0.03 -0.1 0.1 0.2 0.2 0.2 1 0.08 0.3		GarageArea	0.6		
enPorchSF	005 01 006 009 007 005 02 03 02 004 01 01 02 02 009 01 02 02 02 02 03 009 03 02 002 01 02 02 03 02 008 1 04		WoodDeckSF			
SalePrice	003-0.04 02 02 009 01 -0.1 08 -0.2 01 -0.2 -0.2 -0.4 -0.5 0.4 0.2 06 -0.5 06 0.1 0.7 03 05 02-0.04-0.2 0.4 0.4 0.7 06 03 0.4 1	0.6	OpenPorchSF	0.4		
	reas seas seas seas seas seas seas seas					
	MSZor Lots horses photostatic process photostatic process proc			1 1		
	Maria			0.5		

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import mean squared error, r2 score
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X train, y train)
y_pred = model.predict(X_train)
mse = mean_squared_error(y_train, y_pred)
r2 = r2 score(y train, y pred)
print(f"Mean Squared Error: {mse}")
print(f"R-squared: {r2}")
0.0s
                                                                                                      Python
```

Mean Squared Error: 621206121.4955989

R-squared: 0.8460515711954664

Random Forest

```
from sklearn.ensemble import RandomForestRegressor

model_1 = RandomForestRegressor(n_estimators=100, random_state=0)

model_1.fit(X_train, y_train)

y_pred = model_1.predict(X_train)

me = mean_squared_error(y_train, y_pred)

r2 = r2_score(y_train, y_pred)

print(f"Mean Squared Error: {mse}")
print(f"R-squared: {r2}")

✓ 3.5s
Python
```

Mean Squared Error: 89644858.08237426

R-squared: 0.9777840485232812

Gradient Boosting

```
from sklearn.ensemble import GradientBoostingRegressor
model 2 = GradientBoostingRegressor(n estimators=100, learning rate=0.1, random state=0)
model 2.fit(X train, y train)
y_pred = model_2.predict(X_train)
mse = mean squared error(y train, y pred)
r2 = r2 score(y train, y pred)
print(f"Mean Squared Error: {mse}")
print(f"R-squared: {r2}")
0.8s
                                                                                                      Python
```

Mean Squared Error: 250821838.44492674 R-squared: 0.9378408766393095

XGB

```
import xgboost as xgb
  model_4 = xgb.XGBRegressor(n_estimators=100, learning_rate=0.1, random_state=0)
 model 4.fit(X train, y train)
  y_pred = model_4.predict(X_train)
 mse = mean_squared_error(y_train, y_pred)
  r2 = r2 score(y train, y pred)
  print(f"Mean Squared Error: {mse}")
  print(f"R-squared: {r2}")

√ 0.2s

                                                                                                        Python
```

Mean Squared Error: 45084063.98391847

R-squared: 0.9888271854151449

