In [25]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot
import seaborn as san
```

In [26]:
```
df=pd.read_csv("Downloads\\train.csv")
```

In [27]:
```
df.head()
```

Out[27]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	CollgCr
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	FR2	Gtl	Veenker
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub	Inside	Gtl	CollgCr
3	4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub	Corner	Gtl	Crawfor
4	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub	FR2	Gtl	NoRidge

In [28]:
```
pd.set_option('display.max_columns',None)
pd.set_option('display.max_columns',None)
```

In [29]:
```
df.head()
```

Out[29]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	CollgCr
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	FR2	Gtl	Veenker
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub	Inside	Gtl	CollgCr
3	4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub	Corner	Gtl	Crawfor
4	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub	FR2	Gtl	NoRidge

first we handling categorical missinge variable and fill mode value,we can fill also globel value

In [30]:
```
cat_ver=df.select_dtypes(include=["object"])
```

In [31]:
```
cat_ver
```

Out[31]:

	MSZoning	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	Condition1	Condition2	BldgType
0	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm	1Fam
1	RL	Pave	NaN	Reg	Lvl	AllPub	FR2	Gtl	Veenker	Feedr	Norm	1Fam
2	RL	Pave	NaN	IR1	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm	1Fam
3	RL	Pave	NaN	IR1	Lvl	AllPub	Corner	Gtl	Crawfor	Norm	Norm	1Fam
4	RL	Pave	NaN	IR1	Lvl	AllPub	FR2	Gtl	NoRidge	Norm	Norm	1Fam
...	...	...	...	...	...	...	...	...	...	...	...	...
1455	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	Gilbert	Norm	Norm	1Fam
1456	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	NWAmes	Norm	Norm	1Fam
1457	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	Crawfor	Norm	Norm	1Fam
1458	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	NAmes	Norm	Norm	1Fam
1459	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	Edwards	Norm	Norm	1Fam

1460 rows × 43 columns

In [32]:
```
miss_val_per=cat_ver.isnull().mean()*100
miss_val_per
```

Out[32]:
```
MSZoning         0.000000
Street           0.000000
Alley           93.767123
LotShape         0.000000
LandContour      0.000000
Utilities        0.000000
LotConfig        0.000000
LandSlope        0.000000
Neighborhood     0.000000
Condition1       0.000000
Condition2       0.000000
BldgType         0.000000
HouseStyle       0.000000
RoofStyle        0.000000
RoofMatl         0.000000
Exterior1st      0.000000
Exterior2nd      0.000000
MasVnrType       0.547945
ExterQual        0.000000
ExterCond        0.000000
```

In [33]:
```
drop_col=miss_val_per[miss_val_per>20].keys()
drop_col
```

Out[33]:
```
Index(['Alley', 'FireplaceQu', 'PoolQC', 'Fence', 'MiscFeature'], dtype='object')
```

In [35]:
```
cat_ver.drop(columns=drop_col,axis=1,inplace=True)
cat_ver
```

Out[35]:

	MSZoning	Street	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	Condition1	Condition2	BldgType	Hou
0	RL	Pave	Reg	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm	1Fam	
1	RL	Pave	Reg	Lvl	AllPub	FR2	Gtl	Veenker	Feedr	Norm	1Fam	
2	RL	Pave	IR1	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm	1Fam	
3	RL	Pave	IR1	Lvl	AllPub	Corner	Gtl	Crawfor	Norm	Norm	1Fam	
4	RL	Pave	IR1	Lvl	AllPub	FR2	Gtl	NoRidge	Norm	Norm	1Fam	
...	...	...	...	...	...	...	...	...	...	...	...	
1455	RL	Pave	Reg	Lvl	AllPub	Inside	Gtl	Gilbert	Norm	Norm	1Fam	
1456	RL	Pave	Reg	Lvl	AllPub	Inside	Gtl	NWAmes	Norm	Norm	1Fam	
1457	RL	Pave	Reg	Lvl	AllPub	Inside	Gtl	Crawfor	Norm	Norm	1Fam	
1458	RL	Pave	Reg	Lvl	AllPub	Inside	Gtl	NAmes	Norm	Norm	1Fam	
1459	RL	Pave	Reg	Lvl	AllPub	Inside	Gtl	Edwards	Norm	Norm	1Fam	

In [37]:
```
cat_ver.shape
```

Out[37]:
```
(1460, 38)
```

In [41]:
```
#finding missing collumns
```

In [42]:
```
null_per=cat_ver.isnull().mean()*100
miss_var=null_per[null_per>0].keys()
miss_var
```

Out[42]:
```
Index(['MasVnrType', 'BsmtQual', 'BsmtCond', 'BsmtExposure', 'BsmtFinType1',
       'BsmtFinType2', 'Electrical', 'GarageType', 'GarageFinish',
       'GarageQual', 'GarageCond'],
      dtype='object')
```

In [46]:
```
cat_ver["MasVnrType"].mode()
```

Out[46]:
```
0    None
Name: MasVnrType, dtype: object
```

In [47]:
```
cat_ver["MasVnrType"].value_counts()
```

Out[47]:
```
None       864
BrkFace    445
Stone      128
BrkCmn      15
Name: MasVnrType, dtype: int64
```

In [48]:
```
#we fill mode value:
```

In [49]:
```
cat_ver["MasVnrType"].fillna(cat_ver["MasVnrType"].mode()[0])
```

Out[49]:
```
0       BrkFace
1          None
2       BrkFace
3          None
4       BrkFace
         ...
1455       None
1456      Stone
1457       None
1458       None
1459       None
Name: MasVnrType, Length: 1460, dtype: object
```

In [50]:
```
#we have 11 column of missing value now we can fill all column mode value
```

In [54]:
```
for var in miss_var:
    cat_ver[var].fillna(cat_ver[var].mode()[0],inplace=True)
    print(var,"=",cat_ver[var].mode()[0])
```

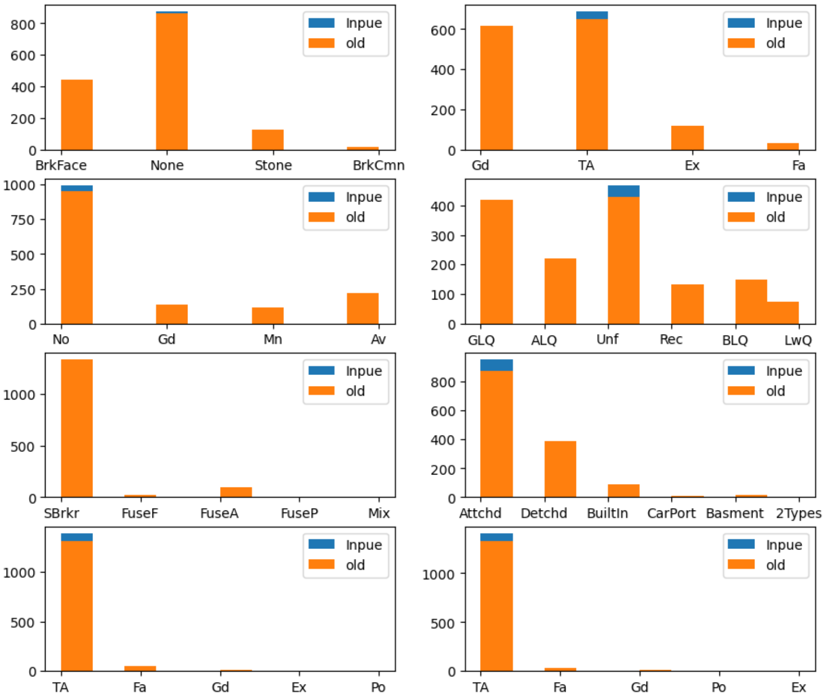
```
MasVnrType = None
BsmtQual = TA
BsmtCond = TA
BsmtExposure = No
BsmtFinType1 = Unf
BsmtFinType2 = Unf
Electrical = SBrkr
GarageType = Attchd
GarageFinish = Unf
GarageQual = TA
GarageCond = TA
```

In [55]:
```
cat_ver.isnull().sum()
```

Out[55]:
```
MSZoning         0
Street           0
LotShape         0
LandContour      0
Utilities        0
LotConfig        0
LandSlope        0
Neighborhood     0
Condition1       0
Condition2       0
BldgType         0
HouseStyle       0
RoofStyle        0
RoofMatl         0
Exterior1st      0
Exterior2nd      0
MasVnrType       0
ExterQual        0
ExterCond        0
```

In [56]:
```
#data dis
```

In [61]:
```
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(16,9))
for i,var in enumerate(miss_var):
    plt.subplot(4,3,i+1)
    plt.hist(cat_ver[var],label="Inpue")
    plt.hist(df[var].dropna(),label="old")
    plt.legend()
```



In [62]:
```
df.update(cat_ver)
df.drop(columns=drop_col,inplace=True)
```

In [63]:
```
df.head()
```

Out[63]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	Cond
0	1	60	RL	65.0	8450	Pave	Reg	Lvl	AllPub	Inside	Gtl	CollgCr	
1	2	20	RL	80.0	9600	Pave	Reg	Lvl	AllPub	FR2	Gtl	Veenker	
2	3	60	RL	68.0	11250	Pave	IR1	Lvl	AllPub	Inside	Gtl	CollgCr	
3	4	70	RL	60.0	9550	Pave	IR1	Lvl	AllPub	Corner	Gtl	Crawfor	
4	5	60	RL	84.0	14260	Pave	IR1	Lvl	AllPub	FR2	Gtl	NoRidge	

In [65]:
```
df.isnull().sum()
```

Out[65]:
```
Id                 0
MSSubClass         0
MSZoning           0
LotFrontage      259
LotArea            0
                ... 
MoSold             0
YrSold             0
SaleType           0
SaleCondition      0
SalePrice          0
Length: 76, dtype: int64
```

In [67]:
```
df.select_dtypes(include=["object"]).isnull().sum()
```

Out[67]:
```
MSZoning         0
Street           0
LotShape         0
LandContour      0
Utilities        0
LotConfig        0
LandSlope        0
Neighborhood     0
Condition1       0
Condition2       0
BldgType         0
HouseStyle       0
RoofStyle        0
RoofMatl         0
Exterior1st      0
Exterior2nd      0
MasVnrType       0
ExterQual        0
ExterCond        0
Foundation       0
BsmtQual         0
BsmtCond         0
BsmtExposure     0
BsmtFinType1     0
BsmtFinType2     0
Heating          0
HeatingQC        0
CentralAir       0
Electrical       0
KitchenQual      0
Functional       0
GarageType       0
GarageFinish     0
GarageQual       0
GarageCond       0
PavedDrive       0
SaleType         0
SaleCondition    0
dtype: int64
```

In [68]:
```
#categorical value missie is complte now we handle numeric value
```

In [69]:
```
df.shape
```

Out[69]:
```
(1460, 76)
```

In [72]:
```
df_numeric=df.select_dtypes(include=["int","float"])
df_numeric
```

Out[72]:

	Id	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	BsmtFinSF
0	1	60	65.0	8450	7	5	2003	2003	196.0	706	
1	2	20	80.0	9600	6	8	1976	1976	0.0	978	
2	3	60	68.0	11250	7	5	2001	2002	162.0	486	
3	4	70	60.0	9550	7	5	1915	1970	0.0	216	
4	5	60	84.0	14260	8	5	2000	2000	350.0	655	
...	...	...	...	...	...	...	...	...	...	...	.
1455	1456	60	62.0	7917	6	5	1999	2000	0.0	0	
1456	1457	20	85.0	13175	6	6	1978	1988	119.0	790	16
1457	1458	70	66.0	9042	7	9	1941	2006	0.0	275	
1458	1459	20	68.0	9717	5	6	1950	1996	0.0	49	102
1459	1460	20	75.0	9937	5	6	1965	1965	0.0	830	29

1460 rows × 38 columns

In [73]:
```
df_numeric.isnull().sum()
```

Out[73]:
```
Id                0
MSSubClass        0
LotFrontage     259
LotArea           0
OverallQual       0
OverallCond       0
YearBuilt         0
YearRemodAdd      0
MasVnrArea        8
BsmtFinSF1        0
BsmtFinSF2        0
BsmtUnfSF         0
TotalBsmtSF       0
1stFlrSF          0
2ndFlrSF          0
LowQualFinSF      0
GrLivArea         0
BsmtFullBath      0
BsmtHalfBath      0
```

In [93]:
```
df_per=df.isnull().sum()/df.shape[0]*100
df_per
```

Out[93]:
```
Id                0.000000
MSSubClass        0.000000
MSZoning          0.000000
LotFrontage      17.739726
LotArea           0.000000
                   ...
MoSold            0.000000
YrSold            0.000000
SaleType          0.000000
SaleCondition     0.000000
SalePrice         0.000000
Length: 76, dtype: float64
```

In [96]:
```
num_cols=[var for var in df_numeric if df_numeric[var].isnull().sum()>0]
num_cols
```

Out[96]:
```
['LotFrontage', 'MasVnrArea', 'GarageYrBlt']
```

In [99]:
```
num_var=df_numeric[num_cols][df_numeric[num_cols].isnull().any(axis=1)]
num_var
```

Out[99]:

	LotFrontage	MasVnrArea	GarageYrBlt
7	NaN	240.0	1973.0
12	NaN	0.0	1962.0
14	NaN	212.0	1960.0
16	NaN	180.0	1970.0
24	NaN	0.0	1968.0
...	...	...	...
1443	NaN	0.0	1916.0
1446	NaN	189.0	1962.0
1449	21.0	0.0	NaN
1450	60.0	0.0	NaN
1453	90.0	0.0	NaN

339 rows × 3 columns

In [100]:
```
#now we tack references colum
```

In [101]:
```
df.head()
```

Out[101]:

RoofStyle	RoofMatl	Exterior1st	Exterior2nd	MasVnrType	MasVnrArea	ExterQual	ExterCond	Foundation	BsmtQual	BsmtCond	BsmtExpos
Gable	CompShg	VinylSd	VinylSd	BrkFace	196.0	Gd	TA	PConc	Gd	TA	
Gable	CompShg	MetalSd	MetalSd	None	0.0	TA	TA	CBlock	Gd	TA	
Gable	CompShg	VinylSd	VinylSd	BrkFace	162.0	Gd	TA	PConc	Gd	TA	
Gable	CompShg	Wd Sdng	Wd Shng	None	0.0	TA	TA	BrkTil	TA	Gd	
Gable	CompShg	VinylSd	VinylSd	BrkFace	350.0	Gd	TA	PConc	Gd	TA	

In [103]:
```
cat_ver=["LotConfig","MSZoning","MasVnrType"]
num_ver=['LotFrontage','MasVnrArea','GarageYrBlt']
```

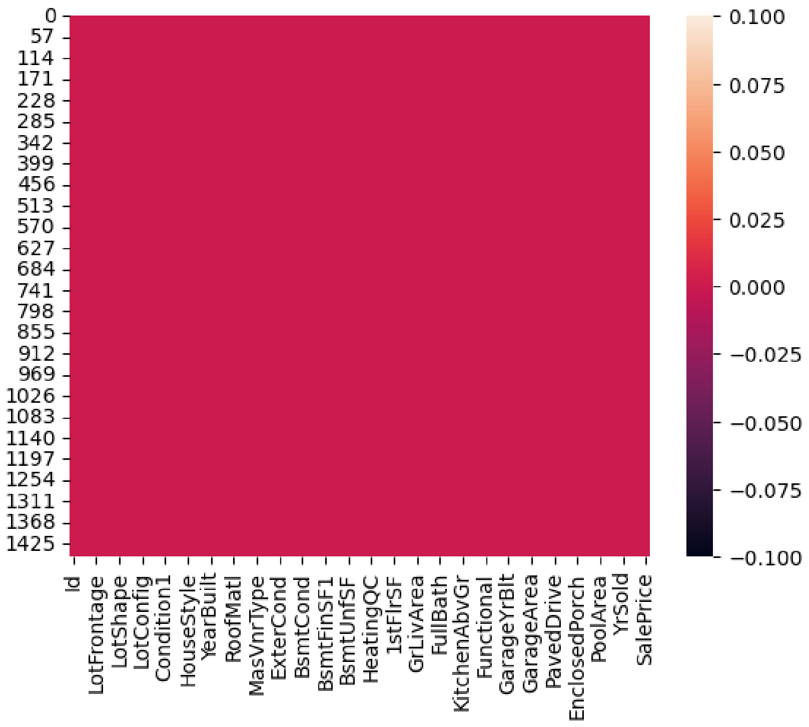
In [104]:
```
for cat_ver,num_ver in zip(cat_ver,num_ver):
    for var in df[cat_ver].unique():
        df.update(df[df.loc[:,cat_ver]==var][num_ver].replace(np.nan,df[df.loc[:,cat_ver]==var][num_ver].mean())
```

In [106]:
```
df.isnull().sum()
```

Out[106]:
```
Id               0
MSSubClass       0
MSZoning         0
LotFrontage      0
LotArea          0
                ..
MoSold           0
YrSold           0
SaleType         0
SaleCondition    0
SalePrice        0
Length: 76, dtype: int64
```

In [107]:
```
sns.heatmap(df.isnull())
```

Out[107]:
```
<Axes: >
```



In [108]:
```
df.shape
```

Out[108]:
```
(1460, 76)
```

```
In [ ]:
```