In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot
import seaborn as san
```

In [2]:
```
df=pd.read_csv("Downloads\\train.csv")
```

In [3]:
```
df.head()
```

Out[3]:

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

5 rows × 81 columns

In [4]:
```
pd.set_option('display.max_columns',None)
pd.set_option('display.max_rows',None)
```

In [5]:
```
df.head()
```

Out[5]:

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

In [6]:
```
df.isnull().sum()
```

Out[6]:

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 [7]:
```
df_per=df.isnull().sum()/df.shape[0]*100
```

In [8]:
```
df_per
```

Out[8]:
```
Id               0.000000
MSSubClass       0.000000
MSZoning         0.000000
LotFrontage     17.739726
LotArea          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
OverallQual      0.000000
OverallCond      0.000000
```

In [11]:
```
drop=df_per[df_per>20].keys()
drop
```

Out[11]:
```
Index(['Alley', 'FireplaceQu', 'PoolQC', 'Fence', 'MiscFeature'], dtype='o
bject')
```

In [12]:
```
df2=df.drop(columns=drop)
```

In [13]:
```
df2.shape
```

Out[13]:
```
(1460, 76)
```

In [17]:
```
df2_numeric=df2.select_dtypes(include=["int","float"])
df2_numeric
```

ıAbvGr	TotRmsAbvGrd	Fireplaces	GarageYrBlt	GarageCars	GarageArea	WoodDeckSF	OpenPorchSF	Enc
1	8	0	2003.0	2	548	0	61	
1	6	1	1976.0	2	460	298	0	
1	6	1	2001.0	2	608	0	42	
1	7	1	1998.0	3	642	0	35	
1	9	1	2000.0	3	836	192	84	
1	5	0	1993.0	2	480	40	30	
1	7	1	2004.0	2	636	255	57	
1	7	2	1973.0	2	484	235	204	
2	8	2	1931.0	2	468	90	0	
2	5	2	1939.0	1	205	0	4	

In [18]:
```
df2_numeric.isnull().sum()
```

Out[18]:
```
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
FullBath         0
HalfBath         0
BedroomAbvGr     0
KitchenAbvGr     0
TotRmsAbvGrd     0
Fireplaces       0
GarageYrBlt     81
GarageCars       0
GarageArea       0
WoodDeckSF       0
OpenPorchSF      0
EnclosedPorch    0
3SsnPorch        0
ScreenPorch      0
PoolArea         0
MiscVal          0
MoSold           0
YrSold           0
SalePrice        0
dtype: int64
```

In [19]:
```
df2_numeric_per=df2_numeric.isnull().sum()/df.shape[0]*100
df2_numeric_per
```

Out[19]:
```
Id               0.000000
MSSubClass       0.000000
LotFrontage     17.739726
LotArea          0.000000
OverallQual      0.000000
OverallCond      0.000000
YearBuilt        0.000000
YearRemodAdd     0.000000
MasVnrArea       0.547945
BsmtFinSF1       0.000000
BsmtFinSF2       0.000000
BsmtUnfSF        0.000000
TotalBsmtSF      0.000000
1stFlrSF         0.000000
2ndFlrSF         0.000000
LowQualFinSF     0.000000
GrLivArea        0.000000
BsmtFullBath     0.000000
BsmtHalfBath     0.000000
```

In [26]:
```
null_cols=[var for var in df2_numeric if df2_numeric[var].isnull().sum()>0]
```

In [27]:
```
null_cols
```

Out[27]:
```
['LotFrontage', 'MasVnrArea', 'GarageYrBlt']
```

In [32]:
```
null_var=df2_numeric[null_cols][df[null_cols].isnull().any(axis=1)]
null_rows
```

Out[32]:

	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
31	NaN	0.0	1966.0
39	65.0	0.0	NaN
42	NaN	0.0	1983.0
43	NaN	0.0	1977.0
48	33.0	0.0	NaN
50	NaN	0.0	1997.0

In [35]:
```
df2_numeric.update(df2_numeric.fillna(df2_numeric.mean()))
```

In [37]:
```
df2_numeric.isnull().sum()
```

Out[37]:
```
Id               0
MSSubClass       0
LotFrontage      0
LotArea          0
OverallQual      0
OverallCond      0
YearBuilt        0
YearRemodAdd     0
MasVnrArea       0
BsmtFinSF1       0
BsmtFinSF2       0
BsmtUnfSF        0
TotalBsmtSF      0
1stFlrSF         0
2ndFlrSF         0
LowQualFinSF     0
GrLivArea        0
BsmtFullBath     0
BsmtHalfBath     0
FullBath         0
HalfBath         0
BedroomAbvGr     0
KitchenAbvGr     0
TotRmsAbvGrd     0
Fireplaces       0
GarageYrBlt      0
GarageCars       0
GarageArea       0
WoodDeckSF       0
OpenPorchSF      0
EnclosedPorch    0
3SsnPorch        0
ScreenPorch      0
PoolArea         0
MiscVal          0
MoSold           0
YrSold           0
SalePrice        0
dtype: int64
```

In [ ]:
```

```