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

Out[5]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	Condition1	C
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	FR2	Gtl	Veenker	Feedr	
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	
3	4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub	Corner	Gtl	Crawfor	Norm	
4	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub	FR2	Gtl	NoRidge	Norm	

In [6]:
```
df.shape
```

Out[6]:
```
(1460, 81)
```

In [10]:
```
df_per=df.isnull().sum()/df.shape[0]*100
df_per
```

Out[10]:
```
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: 81, dtype: float64
```

In [13]:
```
df_drop=df_per[df_per>20].keys()
```

In [14]:
```
df_drop
```

Out[14]:
```
Index(['Alley', 'FireplaceQu', 'PoolQC', 'Fence', 'MiscFeature'], dtype='object')
```

In [15]:
```
df2=df.drop(columns=df_drop)
```

In [16]:
```
df2.shape
```

Out[16]:
```
(1460, 76)
```

In [20]:
```
df2_numeric=df.select_dtypes(include=["int","float"])
```

In [21]:
```
df2_numeric
```

Out[21]:

	Id	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	BsmtFinSF2	BsmtUnfSF	To
0	1	60	65.0	8450	7	5	2003	2003	196.0	706	0	150	
1	2	20	80.0	9600	6	8	1976	1976	0.0	978	0	284	
2	3	60	68.0	11250	7	5	2001	2002	162.0	486	0	434	
3	4	70	60.0	9550	7	5	1915	1970	0.0	216	0	540	
4	5	60	84.0	14260	8	5	2000	2000	350.0	655	0	490	
...	...	...	...	...	...	...	...	...	...	...	...	...	
1455	1456	60	62.0	7917	6	5	1999	2000	0.0	0	0	953	
1456	1457	20	85.0	13175	6	6	1978	1988	119.0	790	163	589	
1457	1458	70	66.0	9042	7	9	1941	2006	0.0	275	0	877	
1458	1459	20	68.0	9717	5	6	1950	1996	0.0	49	1029	0	
1459	1460	20	75.0	9937	5	6	1965	1965	0.0	830	290	136	

460 rows × 38 columns

In [24]:
```
df2_numeric.isnull().sum()
```

Out[24]:
```
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 [27]:
```
num_var=[var for var in df2_numeric if df2_numeric[var].isnull().sum()>0]
```

In [28]:
```
num_var
```

Out[28]:
```
['LotFrontage', 'MasVnrArea', 'GarageYrBlt']
```

In [31]:
```
num_var_miss=df2_numeric[num_var][df2_numeric[num_var].isnull().any(axis=1)]
num_var_miss
```

Out[31]:

	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 [32]:
```
df.head()
```

Out[32]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	Condition1	C
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	FR2	Gtl	Veenker	Feedr	
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	
3	4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub	Corner	Gtl	Crawfor	Norm	
4	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub	FR2	Gtl	NoRidge	Norm	

In [33]:
```
df["LotConfig"].unique()
```

Out[33]:
```
array(['Inside', 'FR2', 'Corner', 'CulDSac', 'FR3'], dtype=object)
```

In [36]:
```
df[df.loc[:,"LotConfig"]=="Inside"]["LotFrontage"]
```

Out[36]:
```
0       65.0
2       68.0
5       85.0
6       75.0
8       51.0
        ...
1455    62.0
1456    85.0
1457    66.0
1458    68.0
1459    75.0
Name: LotFrontage, Length: 1052, dtype: float64
```

In [37]:
```
df[df.loc[:,"LotConfig"]=="Inside"]["LotFrontage"].replace(np.nan,df[df.loc[:,"LotConfig"]=="Inside"]["LotFrontage"].mean())
```

Out[37]:
```
0       65.0
2       68.0
5       85.0
6       75.0
8       51.0
        ...
1455    62.0
1456    85.0
1457    66.0
1458    68.0
1459    75.0
Name: LotFrontage, Length: 1052, dtype: float64
```

In [39]:
```
df_copy=df.copy()
for var in df["LotConfig"].unique():
    df_copy.update(df[df.loc[:,"LotConfig"]==var]["LotFrontage"].replace(np.nan,df[df.loc[:,"LotConfig"]==var]["LotFrontage"
```

In [40]:
```
df_copy.isnull().sum()
```

Out[40]:
```
Id               0
MSSubClass       0
MSZoning         0
LotFrontage      0
LotArea          0
                ..
MoSold           0
YrSold           0
SaleType         0
SaleCondition    0
SalePrice        0
Length: 81, dtype: int64
```

In [42]:
```
num_var
```

Out[42]:
```
['LotFrontage', 'MasVnrArea', 'GarageYrBlt']
```

In [46]:
```
df_copy=df.copy()
num_var=['LotFrontage', 'MasVnrArea', 'GarageYrBlt']
cat_var=["LotConfig","MSZoning","LotShape"]
for cat_var,num_var in zip(cat_var,num_var):
    for var in df[cat_var].unique():
        df_copy.update(df[df.loc[:,cat_var]==var][num_var].replace(np.nan,df[df.loc[:,cat_var]==var][num_var].mean()))
```

In [49]:
```
df_copy[num_var].isnull().sum()
```

Out[49]:
```
0
```

In [ ]: