firstly we import the libraries that will be used in our project

```
import numpy as np
import pandas as pd
%matplotlib inline
%matplotlib notebook
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn.preprocessing
```

now we load the csv file in our jupyter notebook and we use the head command to see the upper 5 rows

```
train=pd.read_csv('train.csv')
train.head(5)
```

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscFeature	MiscVal	MoSold	YrSold	SaleType	SaleCondition	SalePrice
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	2	2008	WD	Normal	208500
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	5	2007	WD	Normal	181500
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	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	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	12	2008	WD	Normal	250000

5 rows × 81 columns

```
train.shape
```

```
(1460, 81)
```

so this we have to know the shape of the data set with this we will now how many rows and columns in the data set

```
train.info()
```

```
<class'pd.DataFrame'>
RangeIndex: 1460, 0to1459
Data columns(total 81 columns):
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
dtypes: float64(3), int64(38), object(43)
memory usage: 923.9+ KB
```

```
train.shape
```

```
(1460, 81)
```

so now we find the uniqueness of the columns i.e how many other value or data in one column

```
train.nunique()
```

```
Id               1460
MSSubClass         15
MSZoning            5
LotFrontage       110
LotArea          1073
Street              2
Alley               2
LotShape            4
LandContour         4
Utilities           2
LotConfig           5
LandSlope           3
Neighborhood       25
Condition1          9
Condition2          8
BldgType            5
HouseStyle          8
OverallQual        10
OverallCond         9
YearBuilt          112
YearRemodAdd        61
RoofStyle           6
RoofMatl            8
Exterior1st        15
Exterior2nd        16
MasVnrType          4
MasVnrArea        328
ExterQual           4
ExterCond           5
Foundation          6
BsmtQual            4
BsmtCond            4
BsmtExposure        4
BsmtFinType1        6
BsmtFinSF1        637
BsmtFinType2        6
BsmtFinSF2        144
BsmtUnfSF         780
TotalBsmtSF       721
Heating             6
HeatingQC           5
CentralAir          2
Electrical          5
1stFlrSF          753
2ndFlrSF          417
LowQualFinSF       24
GrLivArea         861
BsmtFullBath        4
BsmtHalfBath        3
FullBath            4
HalfBath            3
BedroomAbvGr        8
KitchenAbvGr        4
KitchenQual         4
TotRmsAbvGrd       12
Functional          7
Fireplaces          4
FireplaceQu         5
GarageType          6
GarageYrBlt        103
GarageFinish        3
GarageCars          5
GarageArea        441
GarageQual          5
GarageCond          5
PavedDrive          3
WoodDeckSF        274
OpenPorchSF       202
EnclosedPorch     120
3SsnPorch          24
ScreenPorch        121
PoolArea             8
PoolQC               3
Fence                4
MiscFeature          4
MiscVal             21
MoSold              12
YrSold               5
SaleType             9
SaleCondition        6
SalePrice          663
dtype: int64
```

NOW WE SEPARATE THE NUMERICAL DATA FROM CATEGORICAL DATA AND FIND THE MISSING VALUE IN NUMERICAL DATA AND FILL THAT VALUE WITH NUMBER.

```
data_num=train.select_dtypes(include=['int64','float64'])
```

```
data_num.head(5)
```

	Id	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	BsmtFinSF2	BsmtUnfSF	TotalBsmtSF	1stFlrSF	2ndFlrSF	LowQualFinSF	GrLivArea	BsmtFullBath	BsmtHalfBath	FullBath	HalfBath	BedroomAbvGr	KitchenAbvGr	TotRmsAbvGrd	Fireplaces	GarageYrBlt	GarageCars	GarageArea	WoodDeckSF	OpenPorchSF	EnclosedPorch	3SsnPorch	ScreenPorch	PoolArea	MiscVal	MoSold	YrSold	SalePrice
0	1	60	65.0	8450	7	5	2003	2003	196.0	706	0	150	856	856	854	0	1710	1	0	2	1	3	1	8	0	2003.0	2	548	0	61	0	0	0	0	0	2	2008	208500
1	2	20	80.0	9600	6	8	1976	1976	0.0	978	0	284	1262	1262	0	0	1262	0	1	2	0	3	1	6	1	1976.0	2	460	298	0	0	0	0	0	0	5	2007	181500
2	3	60	68.0	11250	7	5	2001	2002	162.0	486	0	434	920	920	866	0	1786	1	0	2	1	3	1	6	1	2001.0	2	608	0	42	0	0	0	0	0	9	2008	223500
3	4	70	60.0	9550	7	5	1915	1970	0.0	216	0	540	756	961	756	0	1717	1	0	1	0	3	1	7	1	1998.0	3	642	0	35	272	0	0	0	0	2	2006	140000
4	5	60	84.0	14260	8	5	2000	2000	350.0	655	0	490	1145	1145	1053	0	2198	1	0	2	1	4	1	9	1	2000.0	3	836	192	84	0	0	0	0	0	12	2008	250000

```
data_num.isnull()
```

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

```
data_num.isnull().sum()
```

```
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
```

so firstly we have done the missing value treatment for "LOT FRONTAGE" data and after that for "GARAGEYRBLT"

```
data_num['LotFrontage'].fillna(data_num['LotFrontage'].median(),inplace=True)
```

```
data_num['GarageYrBlt'].fillna(data_num['GarageYrBlt'].median(),inplace=True)
```

```
data_num.isnull().sum()
```

```
Id               0
MSSubClass       0
LotFrontage      0
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      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
```

so now we have just 8 null value in numerical data this is less then 5% so we can drop that value directly

```
data_num.dropna(inplace=True)
```

```
data_num.isnull().sum()
```

```
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
```

with this we have fill all the null value with median method now there is no null value in the numerical data

```
sns.distplot(data_num['SalePrice'])
plt.title('Distribution of SalePrice')
```

Identify significant variables using a correlation matrix

```
data_num.corrwith(data_num['SalePrice'])
```

```
Id               -0.021917
MSSubClass       -0.084284
LotFrontage       0.351799
LotArea           0.263843
OverallQual       0.790982
OverallCond       0.077856
YearBuilt         0.522897
YearRemodAdd      0.507101
MasVnrArea        0.472614
BsmtFinSF1        0.386420
BsmtFinSF2       -0.011378
BsmtUnfSF         0.214479
TotalBsmtSF       0.613581
1stFlrSF          0.605852
2ndFlrSF          0.319334
LowQualFinSF     -0.037963
GrLivArea         0.708624
BsmtFullBath      0.227122
BsmtHalfBath     -0.016844
FullBath          0.560664
HalfBath          0.284108
BedroomAbvGr      0.168213
KitchenAbvGr     -0.135907
TotRmsAbvGrd      0.533723
Fireplaces        0.466929
GarageYrBlt       0.486362
GarageCars        0.640409
GarageArea        0.623431
WoodDeckSF        0.324413
OpenPorchSF       0.315856
EnclosedPorch    -0.128578
3SsnPorch         0.044584
ScreenPorch       0.111447
PoolArea          0.092404
MiscVal          -0.021190
MoSold            0.046432
YrSold           -0.028923
SalePrice         1.000000
dtype: float64
```

```
corr=data_num.corrwith(data_num['SalePrice'])
corr[corr>0.5].sort_values(ascending=False)
```

```
SalePrice    1.000000
OverallQual  0.790982
GrLivArea    0.708624
GarageCars   0.640409
GarageArea   0.623431
TotalBsmtSF  0.613581
1stFlrSF     0.605852
FullBath     0.560664
TotRmsAbvGrd 0.533723
YearBuilt    0.522897
YearRemodAdd 0.507101
GarageYrBlt  0.486362
MasVnrArea   0.472614
Fireplaces   0.466929
BsmtFinSF1   0.386420
LotFrontage  0.351799
WoodDeckSF   0.324413
2ndFlrSF     0.319334
OpenPorchSF  0.315856
HalfBath     0.284108
LotArea      0.263843
BsmtFullBath 0.227122
BsmtUnfSF    0.214479
BedroomAbvGr 0.168213
ScreenPorch  0.111447
PoolArea     0.092404
MoSold       0.046432
3SsnPorch    0.044584
LowQualFinSF -0.037963
MiscVal      -0.021190
BsmtHalfBath -0.016844
BsmtFinSF2   -0.011378
OverallCond  0.077856
MSSubClass   -0.084284
YrSold       -0.028923
EnclosedPorch -0.128578
KitchenAbvGr -0.135907
Id           -0.021917
dtype: float64
```

```
corr=data_num.corrwith(data_num['SalePrice'])
corr[corr>0.5].plot.bar(figsize=(8,6))
```



```
sns.pairplot(data_num,hue='SalePrice')
```

```
<seaborn.axisgrid.PairGrid at 0x1a0d0d0d0>
```


PAIR PLOT

```
%matplotlib inline
sns.pairplot(data_num,hue='SalePrice')
```

```
<seaborn.axisgrid.PairGrid at 0x1a0d0d0d0>
```


EDA of categorical variables

```
data_cat=train.select_dtypes(include=['object'])
```

```
data_cat.head(5)
```

	MSZoning	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	Condition1	Condition2	BldgType	HouseStyle	RoofStyle	RoofMatl	Exterior1st	Exterior2nd	MasVnrType	ExterQual	ExterCond	Foundation	BsmtQual	BsmtCond	BsmtExposure	BsmtFinType1	BsmtFinType2	Heating	HeatingQC	CentralAir	Electrical	KitchenQual	Functional	FireplaceQu	GarageType	GarageFinish	GarageQual	GarageCond	PavedDrive	PoolQC	Fence	MiscFeature	SaleType	SaleCondition
0	RL	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	CollgCr	Norm	Norm	1Fam	2Story	Gable	CompShg	VinylSd	VinylSd	BrkFace	Gd	TA	PConc	Gd	TA	No	GLQ	Unf	GasA	Ex	Y	SBrkr	Gd	Typ	NaN	Attchd	RFn	TA	TA	Y	NaN	NaN	NaN	WD	Normal
1	RL	Pave	NaN	Reg	Lvl	AllPub	FR2	Gtl	Veenker	Feedr	Norm	1Fam	1Story	Gable	CompShg	MetalSd	MetalSd	None	TA	TA	CBlock	Gd	TA	Gd	ALQ	Unf	GasA	Ex	Y	SBrkr	TA	Typ	TA	Attchd	RFn	TA	TA	Y	NaN	NaN	NaN	WD	Abnorml
2	RL	Pave	NaN	IR1	Lvl	AllPub	Inside	Gtl	NoRidge	Norm	Norm	1Fam	2Story	Gable	CompShg	VinylSd	VinylSd	BrkFace	Gd	TA	PConc	Gd	TA	Mn	GLQ	Unf	GasA	Ex	Y	SBrkr	Gd	Typ	TA	Attchd	RFn	TA	TA	Y	NaN	NaN	NaN	WD	Normal
3	RL	Pave	NaN	IR1	Lvl	AllPub	Corner	Gtl	Crawfor	Norm	Norm	1Fam	2Story	Gable	CompShg	Wd Sdng	Wd Shng	None	TA	TA	BrkTil	TA	Gd	No	ALQ	Unf	GasA	Gd	Y	SBrkr	Gd	Typ	Gd	Detchd	Unf	TA	TA	Y	NaN	NaN	NaN	WD	Abnorml
4	RL	Pave	NaN	IR1	Lvl	AllPub	FR2	Gtl	NoRidge	Norm	Norm	1Fam	2Story	Gable	CompShg	VinylSd	VinylSd	BrkFace	Gd	TA	PConc	Gd	TA	Av	GLQ	Unf	GasA	Ex	Y	SBrkr	Gd	Typ	TA	Attchd	RFn	TA	TA	Y	NaN	NaN	NaN	WD	Normal

```
data_cat.isnull()
```

```
MSZoning       False
Street         False
Alley           True
LotShape       False
LandContour    False
Utilities      False
LotConfig      False
LandSlope      False
Neighborhood   False
Condition1     False
Condition2     False
BldgType       False
HouseStyle     False
RoofStyle      False
RoofMatl       False
Exterior1st    False
Exterior2nd    False
MasVnrType     False
ExterQual      False
ExterCond      False
Foundation     False
BsmtQual       False
BsmtCond       False
BsmtExposure   False
BsmtFinType1   False
BsmtFinType2   False
Heating        False
HeatingQC      False
CentralAir     False
Electrical     False
KitchenQual    False
Functional     False
FireplaceQu     True
GarageType     False
GarageFinish   False
GarageQual     False
GarageCond     False
PavedDrive     False
PoolQC          True
Fence           True
MiscFeature     True
SaleType       False
SaleCondition  False
dtype: bool
```

```
data_cat.isnull().sum()
```

```
MSZoning       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
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
FireplaceQu    0
GarageType     0
GarageFinish   0
GarageQual     0
GarageCond     0
PavedDrive     0
PoolQC         0
Fence          0
MiscFeature    0
SaleType       0
SaleCondition  0
dtype: int64
```

NOW WITH THIS WE HAVE SEPARATED THE CATEGORICAL VALUE AND NOW FIND THE MISSING VALUE IN THIS AND REPLACE IT

```
data_cat.isnull()
```

```
MSZoning       False
Street         False
Alley           True
LotShape       False
LandContour    False
Utilities      False
LotConfig      False
LandSlope      False
Neighborhood   False
Condition1     False
Condition2     False
BldgType       False
HouseStyle     False
RoofStyle      False
RoofMatl       False
Exterior1st    False
Exterior2nd    False
MasVnrType     False
ExterQual      False
ExterCond      False
Foundation     False
BsmtQual       False
BsmtCond       False
BsmtExposure   False
BsmtFinType1   False
BsmtFinType2   False
Heating        False
HeatingQC      False
CentralAir     False
Electrical     False
KitchenQual    False
Functional     False
FireplaceQu     True
GarageType     False
GarageFinish   False
GarageQual     False
GarageCond     False
PavedDrive     False
PoolQC          True
Fence           True
MiscFeature     True
SaleType       False
SaleCondition  False
dtype: bool
```

now we have to do the missing value treatment for categorical value with "MODE METHOD"

```
data_cat['Alley'].mode()
```

```
0    Grvl
dtype: object
```

```
data_cat['Alley'].fillna(data_cat['Alley'].mode(),inplace=True)
```

```
data_cat['FireplaceQu'].mode()
```

```
0    Gd
dtype: object
```

```
data_cat['FireplaceQu'].fillna(data_cat['FireplaceQu'].mode(),inplace=True)
```

```
data_cat['PoolQC'].mode()
```

```
0    Gd
dtype: object
```

```
data_cat['PoolQC'].fillna(data_cat['PoolQC'].mode(),inplace=True)
```

```
data_cat['Fence'].mode()
```

```
0    MnPrv
dtype: object
```

```
data_cat['Fence'].fillna(data_cat['Fence'].mode(),inplace=True)
```

```
data_cat['MiscFeature'].mode()
```

```
0    Shed
dtype: object
```

```
data_cat['MiscFeature'].fillna(data_cat['MiscFeature'].mode(),inplace=True)
```

```
data_cat['SaleType'].mode()
```

```
0    WD
dtype: object
```

```
data_cat['SaleType'].fillna(data_cat['SaleType'].mode(),inplace=True)
```

```
data_cat['SaleCondition'].mode()
```

```
0    Normal
dtype: object
```

```
data_cat['SaleCondition'].fillna(data_cat['SaleCondition'].mode(),inplace=True)
```

```
data_cat.isnull()
```

```
MSZoning       False
Street         False
Alley          False
LotShape       False
LandContour    False
Utilities      False
LotConfig      False
LandSlope      False
Neighborhood   False
Condition1     False
Condition2     False
BldgType       False
HouseStyle     False
RoofStyle      False
RoofMatl       False
Exterior1st    False
Exterior2nd    False
MasVnrType     False
ExterQual      False
ExterCond      False
Foundation     False
BsmtQual       False
BsmtCond       False
BsmtExposure   False
BsmtFinType1   False
BsmtFinType2   False
Heating        False
HeatingQC      False
CentralAir     False
Electrical     False
KitchenQual    False
Functional     False
FireplaceQu    False
GarageType     False
GarageFinish   False
GarageQual     False
GarageCond     False
PavedDrive     False
PoolQC         False
Fence          False
MiscFeature    False
SaleType       False
SaleCondition  False
dtype: bool
```

so with this we have fill the categorical null values with mode method now rest of the categorical null value is less then 5% so now we can directly drop that value

```
data_cat.dropna(inplace=True)
```

```
data_cat.isnull()
```

```
MSZoning       False
Street         False
Alley          False
LotShape       False
LandContour    False
Utilities      False
LotConfig      False
LandSlope      False
Neighborhood   False
Condition1     False
Condition2     False
BldgType       False
HouseStyle     False
RoofStyle      False
RoofMatl       False
Exterior1st    False
Exterior2nd    False
MasVnrType     False
ExterQual      False
ExterCond      False
Foundation     False
BsmtQual       False
BsmtCond       False
BsmtExposure   False
BsmtFinType1   False
BsmtFinType2   False
Heating        False
HeatingQC      False
CentralAir     False
Electrical     False
KitchenQual    False
Functional     False
FireplaceQu    False
GarageType     False
GarageFinish   False
GarageQual     False
GarageCond     False
PavedDrive     False
PoolQC         False
Fence          False
MiscFeature    False
SaleType       False
SaleCondition  False
dtype: bool
```

Count plot and box plot for bivarate analysis

```
sns.boxplot(data=data_cat,x='PavementCond',y='SalePrice',hue='LandContour')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a0d0d0d0>
```

```
sns.countplot(data=data_cat,x='LandContour',hue='LotShape')
plt.title('A Tremendous Amount of Lots Are Level')
```


now we have to combine the data_num and data_cat with the help of concat function

```
data=pd.concat([data_num,data_cat],axis=1)
```

```
data.shape
```

```
(1460, 81)
```

```
data.isnull()
```

| | Id | MSSubClass | LotFrontage | LotArea | OverallQual | OverallCond | YearBuilt | YearRemodAdd | MasVnrArea | BsmtFinSF1 | BsmtFinSF2 | BsmtUnfSF | TotalBsmtSF | 1stFlrSF | 2ndFlrSF | LowQualFinSF | GrLivArea | BsmtFullBath | BsmtHalfBath | FullBath | HalfBath | BedroomAbvGr | KitchenAbvGr | TotRmsAbvGrd | Fireplaces | GarageYrBlt | GarageCars | GarageArea | WoodDeckSF | OpenPorchSF | EnclosedPorch | 3SsnPorch | ScreenPorch | PoolArea | MiscVal | MoSold | YrSold | SalePrice | MSZoning | Street | Alley | LotShape | LandContour | Utilities | LotConfig | LandSlope | Neighborhood | Condition1 | Condition2 | BldgType | HouseStyle | RoofStyle | RoofMatl | Exterior1st | Exterior2nd | MasVnrType | ExterQual | ExterCond | Foundation | BsmtQual | BsmtCond | BsmtExposure | BsmtFinType1 | BsmtFinType2 | Heating | HeatingQC | CentralAir | Electrical | KitchenQual | Functional | FireplaceQu | GarageType | GarageFinish | GarageQual | GarageCond | PavedDrive | PoolQC | Fence | MiscFeature | SaleType | SaleCondition |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---