In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [3]:
```
data=pd.read_csv("Houseprices.csv")
```

In [4]:
```
data.shape
```

Out[4]:
```
(1460, 81)
```

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

Out[5]:

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

5 rows × 81 columns

In [6]:
```
data.columns
```

Out[6]:
```
Index(['Id', 'MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street',
       'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig',
       'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType',
       'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd',
       'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType',
       'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual',
       'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1',
       'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating',
       'HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF',
       'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath',
       'HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual',
       'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType',
       'GarageYrBlt', 'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual',
       'GarageCond', 'PavedDrive', 'WoodDeckSF', 'OpenPorchSF',
       'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'PoolQC',
       'Fence', 'MiscFeature', 'MiscVal', 'MoSold', 'YrSold', 'SaleType',
       'SaleCondition', 'SalePrice'],
      dtype='object')
```

In [7]:
```
data.dtypes
```

Out[7]:
```
Id                int64
MSSubClass        int64
MSZoning         object
LotFrontage     float64
LotArea           int64
                 ...
MoSold            int64
YrSold            int64
SaleType         object
SaleCondition    object
SalePrice         int64
Length: 81, dtype: object
```

In [8]:
```
data.tail()
```

Out[8]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	PoolArea	PoolQC	Fence	Miscl
1455	1456	60	RL	62.0	7917	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	

1456	1457	20	RL	85.0	13175	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	MnPrv
1457	1458	70	RL	66.0	9042	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	GdPrv
1458	1459	20	RL	68.0	9717	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN
1459	1460	20	RL	75.0	9937	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN

5 rows × 81 columns

In [9]:
```
data.sample(10)
```

Out[9]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscI
364	365	60	RL	NaN	18800	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	
679	680	20	RL	NaN	9945	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	
1440	1441	70	RL	79.0	11526	Pave	NaN	IR1	Bnk	AllPub	...	0	NaN	NaN	
76	77	20	RL	NaN	8475	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	
659	660	20	RL	75.0	9937	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	MnPrv	
975	976	160	FV	NaN	2651	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	
419	420	20	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	
1239	1240	20	RL	64.0	9037	Pave	NaN	IR1	HLS	AllPub	...	0	NaN	NaN	
14	15	20	RL	NaN	10920	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	GdWo	
562	563	30	RL	63.0	13907	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	

10 rows × 81 columns

In [10]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1460 entries, 0 to 1459
Data columns (total 81 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Id             1460 non-null   int64
 1   MSSubClass     1460 non-null   int64
 2   MSZoning       1460 non-null   object
 3   LotFrontage    1201 non-null   float64
 4   LotArea        1460 non-null   int64
 5   Street         1460 non-null   object
 6   Alley          91 non-null     object
 7   LotShape       1460 non-null   object
 8   LandContour    1460 non-null   object
 9   Utilities      1460 non-null   object
 10  LotConfig      1460 non-null   object
 11  LandSlope      1460 non-null   object
 12  Neighborhood   1460 non-null   object
 13  Condition1     1460 non-null   object
 14  Condition2     1460 non-null   object
 15  BldgType       1460 non-null   object
 16  HouseStyle     1460 non-null   object
 17  OverallQual    1460 non-null   int64
 18  OverallCond    1460 non-null   int64
 19  YearBuilt      1460 non-null   int64
 20  YearRemodAdd   1460 non-null   int64
 21  RoofStyle      1460 non-null   object
 22  RoofMatl       1460 non-null   object
 23  Exterior1st    1460 non-null   object
 24  Exterior2nd    1460 non-null   object
 25  MasVnrType     1452 non-null   object
 26  MasVnrArea     1452 non-null   float64
 27  ExterQual      1460 non-null   object
 28  ExterCond      1460 non-null   object
 29  Foundation     1460 non-null   object
 30  BsmtQual       1423 non-null   object
 31  BsmtCond       1423 non-null   object
 32  BsmtExposure   1422 non-null   object
 33  BsmtFinType1   1423 non-null   object
 34  BsmtFinSF1     1460 non-null   int64
 35  BsmtFinType2   1422 non-null   object
 36  BsmtFinSF2     1460 non-null   int64
 37  BsmtUnfSF      1460 non-null   int64
 38  TotalBsmtSF    1460 non-null   int64
 39  Heating        1460 non-null   object
 40  HeatingQC      1460 non-null   object
 41  CentralAir     1460 non-null   object
 42  Electrical     1459 non-null   object
```

```
 43  1stFlrSF       1460 non-null   int64
 44  2ndFlrSF       1460 non-null   int64
 45  LowQualFinSF   1460 non-null   int64
 46  GrLivArea      1460 non-null   int64
 47  BsmtFullBath   1460 non-null   int64
 48  BsmtHalfBath   1460 non-null   int64
 49  FullBath       1460 non-null   int64
 50  HalfBath       1460 non-null   int64
 51  BedroomAbvGr   1460 non-null   int64
 52  KitchenAbvGr   1460 non-null   int64
 53  KitchenQual    1460 non-null   object
 54  TotRmsAbvGrd   1460 non-null   int64
 55  Functional     1460 non-null   object
 56  Fireplaces     1460 non-null   int64
 57  FireplaceQu    770 non-null    object
 58  GarageType     1379 non-null   object
 59  GarageYrBlt    1379 non-null   float64
 60  GarageFinish   1379 non-null   object
 61  GarageCars     1460 non-null   int64
 62  GarageArea     1460 non-null   int64
 63  GarageQual     1379 non-null   object
 64  GarageCond     1379 non-null   object
 65  PavedDrive     1460 non-null   object
 66  WoodDeckSF     1460 non-null   int64
 67  OpenPorchSF    1460 non-null   int64
 68  EnclosedPorch  1460 non-null   int64
 69  3SsnPorch      1460 non-null   int64
 70  ScreenPorch    1460 non-null   int64
 71  PoolArea       1460 non-null   int64
 72  PoolQC         7 non-null      object
 73  Fence          281 non-null    object
 74  MiscFeature    54 non-null     object
 75  MiscVal        1460 non-null   int64
 76  MoSold         1460 non-null   int64
 77  YrSold         1460 non-null   int64
 78  SaleType       1460 non-null   object
 79  SaleCondition  1460 non-null   object
 80  SalePrice      1460 non-null   int64
dtypes: float64(3), int64(35), object(43)
memory usage: 924.0+ KB
```

In [11]:
```
data.describe()
```

Out[11]:

	Id	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF
count	1460.000000	1460.000000	1201.000000	1460.000000	1460.000000	1460.000000	1460.000000	1460.000000	1452.000000	1460.00000
mean	730.500000	56.897260	70.049958	10516.828082	6.099315	5.575342	1971.267808	1984.865753	103.685262	443.63972
std	421.610009	42.300571	24.284752	9981.264932	1.382997	1.112799	30.202904	20.645407	181.066207	456.09809
min	1.000000	20.000000	21.000000	1300.000000	1.000000	1.000000	1872.000000	1950.000000	0.000000	0.00000
25%	365.750000	20.000000	59.000000	7553.500000	5.000000	5.000000	1954.000000	1967.000000	0.000000	0.00000
50%	730.500000	50.000000	69.000000	9478.500000	6.000000	5.000000	1973.000000	1994.000000	0.000000	383.50000
75%	1095.250000	70.000000	80.000000	11601.500000	7.000000	6.000000	2000.000000	2004.000000	166.000000	712.25000
max	1460.000000	190.000000	313.000000	215245.000000	10.000000	9.000000	2010.000000	2010.000000	1600.000000	5644.00000

8 rows × 38 columns

In [12]:
```
data.isnull().sum()
```

Out[12]:
```
Id                0
MSSubClass        0
MSZoning          0
LotFrontage     259
LotArea           0
                 ...
MoSold            0
YrSold            0
SaleType          0
SaleCondition     0
SalePrice         0
Length: 81, dtype: int64
```

In [13]:
```
median=data['LotFrontage'].median()
```

In [14]:

```
data['LotFrontage'].replace(np.nan,median,inplace=True)
```

In [15]:
```
data.isnull().sum()
```

Out[15]:
```
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 [16]:
```
duplicate=data.duplicated()
print(duplicate.sum())
data[duplicate]
```

0

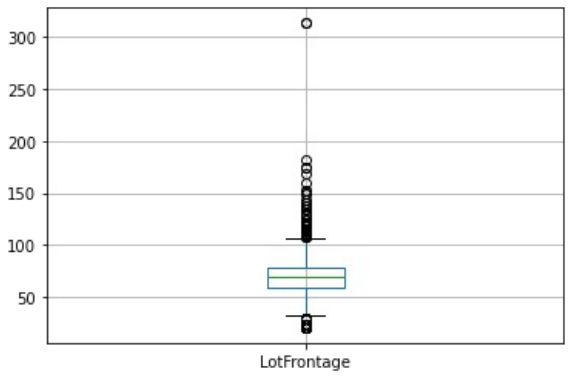
Out[16]:

Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscFeature

0 rows × 81 columns

In [17]:
```
data.boxplot(column=['LotFrontage'])
```

Out[17]:
```
<AxesSubplot:>
```



In [18]:
```
data=data.drop('Street',axis=1)
```

In [19]:
```
data.head()
```

Out[19]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Alley	LotShape	LandContour	Utilities	LotConfig	...	PoolArea	PoolQC	Fence	MiscFea
0	1	60	RL	65.0	8450	NaN	Reg	Lvl	AllPub	Inside	...	0	NaN	NaN	
1	2	20	RL	80.0	9600	NaN	Reg	Lvl	AllPub	FR2	...	0	NaN	NaN	
2	3	60	RL	68.0	11250	NaN	IR1	Lvl	AllPub	Inside	...	0	NaN	NaN	
3	4	70	RL	60.0	9550	NaN	IR1	Lvl	AllPub	Corner	...	0	NaN	NaN	
4	5	60	RL	84.0	14260	NaN	IR1	Lvl	AllPub	FR2	...	0	NaN	NaN	

5 rows × 80 columns

In [20]:
```
data=pd.get_dummies(data,columns=['SaleCondition'])
```

In [21]:
```
data.head()
```

Out[21]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Alley	LotShape	LandContour	Utilities	LotConfig	...	MoSold	YrSold	SaleType	SalePri

0	1	60	RL	65.0	8450	NaN	Reg	Lvl	AllPub	Inside	...	2	2008	WD	2085
1	2	20	RL	80.0	9600	NaN	Reg	Lvl	AllPub	FR2	...	5	2007	WD	1815
2	3	60	RL	68.0	11250	NaN	IR1	Lvl	AllPub	Inside	...	9	2008	WD	2235
3	4	70	RL	60.0	9550	NaN	IR1	Lvl	AllPub	Corner	...	2	2006	WD	1400
4	5	60	RL	84.0	14260	NaN	IR1	Lvl	AllPub	FR2	...	12	2008	WD	2500

5 rows × 85 columns

In [22]:
```
data=data.drop(['Alley','LotShape','LandContour','Utilities','LotConfig','PoolArea','PoolQC','Fence','MiscVal'],a
```

In [23]:
```
data.head()
```

Out[23]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	LandSlope	Neighborhood	Condition1	Condition2	BldgType	...	MoSold	YrSold	SaleTy
0	1	60	RL	65.0	8450	Gtl	CollgCr	Norm	Norm	1Fam	...	2	2008	W
1	2	20	RL	80.0	9600	Gtl	Veenker	Feedr	Norm	1Fam	...	5	2007	W
2	3	60	RL	68.0	11250	Gtl	CollgCr	Norm	Norm	1Fam	...	9	2008	W
3	4	70	RL	60.0	9550	Gtl	Crawfor	Norm	Norm	1Fam	...	2	2006	W
4	5	60	RL	84.0	14260	Gtl	NoRidge	Norm	Norm	1Fam	...	12	2008	W

5 rows × 76 columns

In [24]:
```
data=data.drop(['LandSlope','Neighborhood','Condition1','Condition2','BldgType','OpenPorchSF','EnclosedPorch','3S
```

In [25]:
```
data.head()
```

Out[25]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	HouseStyle	OverallQual	OverallCond	YearBuilt	YearRemodAdd	...	MoSold	YrSold	Sal
0	1	60	RL	65.0	8450	2Story	7	5	2003	2003	...	2	2008	
1	2	20	RL	80.0	9600	1Story	6	8	1976	1976	...	5	2007	
2	3	60	RL	68.0	11250	2Story	7	5	2001	2002	...	9	2008	
3	4	70	RL	60.0	9550	2Story	7	5	1915	1970	...	2	2006	
4	5	60	RL	84.0	14260	2Story	8	5	2000	2000	...	12	2008	

5 rows × 66 columns

In [26]:
```
data=data.drop(['RoofStyle','RoofMatl','GarageType','GarageFinish'],axis=1)
```

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

Out[27]:

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	HouseStyle	OverallQual	OverallCond	YearBuilt	YearRemodAdd	...	MoSold	YrSold	Sal
0	1	60	RL	65.0	8450	2Story	7	5	2003	2003	...	2	2008	
1	2	20	RL	80.0	9600	1Story	6	8	1976	1976	...	5	2007	
2	3	60	RL	68.0	11250	2Story	7	5	2001	2002	...	9	2008	
3	4	70	RL	60.0	9550	2Story	7	5	1915	1970	...	2	2006	
4	5	60	RL	84.0	14260	2Story	8	5	2000	2000	...	12	2008	

5 rows × 62 columns

In [28]:
```
data.dtypes
```

Out[28]:
```
Id                        int64
MSSubClass                int64
MSZoning                 object
LotFrontage             float64
LotArea                   int64
                          ...
SaleCondition_AdjLand     uint8
```

```
SaleCondition_Alloca       uint8
SaleCondition_Family       uint8
SaleCondition_Normal       uint8
SaleCondition_Partial      uint8
Length: 62, dtype: object
```

In [29]:
```
x=data[['YearBuilt','BedroomAbvGr','KitchenAbvGr','GarageArea','MoSold','YrSold','GarageCars','MSSubClass','LotAr
```

In [30]:
```
y=y=data[['SalePrice']]
```

In [31]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.30)
```

In [32]:
```
model=LinearRegression()
model.fit(x_train,y_train)
```

Out[32]:
```
LinearRegression()
```

In [33]:
```
model.score(x_train,y_train)
```

Out[33]:
```
0.5140903398492345
```

In [34]:
```
model.score(x_test,y_test)
```

Out[34]:
```
0.5416013933930424
```

In [37]:
```
from sklearn.preprocessing import PolynomialFeatures
from sklearn import linear_model
poly=PolynomialFeatures(degree=2,interaction_only=True)
x_train1=poly.fit_transform(x_train)
x_test1=poly.fit_transform(x_test)
poly_clf=linear_model.LinearRegression()
poly_clf.fit(x_train1,y_train)
y_pred=poly_clf.predict(x_test1)
print(poly_clf.score(x_train1,y_train))
```

```
0.6159658771807315
```

In [38]:
```
print(poly_clf.score(x_test1,y_test))
```

```
0.6304856704642168
```

the above one is underfitting model

In [ ]:
```

```