

```
import pandas as pd
```

```
df = pd.read_csv("https://raw.githubusercontent.com/AmenaNajeeb/Data/master/FeatureEngineeringDataset.csv")
```

```
df.head(10)
```

	Order	PID	Sqft	Alley	Utilities	Locality	FamilyAccom	HouseStyle	BHK
0	1	526301100	31770.0	NaN	AllPub	NAmes	1Fam	1Story	1.0
1	2	526350040	11622.0	NaN	AllPub	NAmes	1Fam	1Story	1.0
2	3	526351010	14267.0	NaN	AllPub	NAmes	1Fam	1Story	1.0
3	4	526353030	11160.0	NaN	AllPub	NAmes	1Fam	1Story	2.0
4	5	527105010	13830.0	NaN	AllPub	Gilbert	1Fam	2Story	2.0
5	6	527105030	9978.0	NaN	AllPub	Gilbert	1Fam	2Story	2.0
6	7	527127150	4920.0	NaN	AllPub	StoneBr	1Fam	1Story	2.0
7	8	527145080	5005.0	NaN	AllPub	StoneBr	1Fam	1Story	2.0
8	9	527146030	5389.0	NaN	AllPub	StoneBr	1Fam	1Story	2.0
9	10	527162130	7500.0	NaN	AllPub	Gilbert	1Fam	2Story	2.0

```
df.shape
```

```
(100, 15)
```

```
df.isnull().sum()
```

```
Order      0
PID         0
Sqft        6
Alley      96
Utilities   0
Locality    0
FamilyAccom 0
HouseStyle  0
BHK        21
Baths       0
Fireplaces  0
GarageCars  0
PoolArea   100
SaleCondition 0
SalePrice   0
dtype: int64
```

```
df['Sqft'].fillna(value=df['Sqft'].mean(),inplace=True)
```

```
df.isnull().sum()
```

```
Order      0
PID         0
Sqft        0
Alley      96
Utilities   0
Locality    0
FamilyAccom 0
HouseStyle  0
BHK        21
Baths       0
Fireplaces  0
GarageCars  0
PoolArea   100
SaleCondition 0
SalePrice   0
dtype: int64
```

```
df['BHK'].fillna(value=df['BHK'].mean(),inplace=True)
```

```
df.isnull().sum()
```

```
Order      0
PID         0
Sqft        0
Utilities   0
Locality    0
FamilyAccom 0
```

```
HouseStyle      0
BHK              0
Baths           0
Fireplaces      0
GarageCars      0
SaleCondition    0
SalePrice       0
dtype: int64
```

```
df =df.drop(["PoolArea"],axis=1)
```

```
df.shape
```

```
(100, 14)
```

```
df = df.drop(["Alley"],axis=1)
```

```
df.shape
```

```
(100, 13)
```

```
df.isnull().sum()
```

```
Order          0
PID            0
Sqft           0
Utilities      0
Locality       0
FamilyAccom    0
HouseStyle     0
BHK            0
Baths          0
Fireplaces     0
GarageCars     0
SaleCondition  0
SalePrice      0
dtype: int64
```

```
df['Price_Per_Sqft']=df['SalePrice']/df['Sqft']
```

```
df.columns
```

```
Index(['Order', 'PID', 'Sqft', 'Utilities', 'Locality', 'FamilyAccom',
      'HouseStyle', 'BHK', 'Baths', 'Fireplaces', 'GarageCars',
      'SaleCondition', 'SalePrice', 'Price_Per_Sqft'],
      dtype='object')
```

```
df.head(10)
```

	Order	PID	Sqft	Utilities	Locality	FamilyAccom	HouseStyle	BHK	Baths	Fireplaces	Gara
0	1	526301100	31770.0	AllPub	NAmes	1Fam	1Story	1.0	1	2	
1	2	526350040	11622.0	AllPub	NAmes	1Fam	1Story	1.0	1	0	
2	3	526351010	14267.0	AllPub	NAmes	1Fam	1Story	1.0	1	0	
3	4	526353030	11160.0	AllPub	NAmes	1Fam	1Story	2.0	2	2	
4	5	527105010	13830.0	AllPub	Gilbert	1Fam	2Story	2.0	2	1	
5	6	527105030	9978.0	AllPub	Gilbert	1Fam	2Story	2.0	2	1	
6	7	527127150	4920.0	AllPub	StoneBr	1Fam	1Story	2.0	2	0	
7	8	527145080	5005.0	AllPub	StoneBr	1Fam	1Story	2.0	2	0	
8	9	527146030	5389.0	AllPub	StoneBr	1Fam	1Story	2.0	2	1	
9	10	527162130	7500.0	AllPub	Gilbert	1Fam	2Story	2.0	2	1	

```
df["Bed_Bath_Ratio"]=df["BHK"]/df["Baths"]
```

```
df.columns
```

```
Index(['Order', 'PID', 'Sqft', 'Utilities', 'Locality', 'FamilyAccom',
      'HouseStyle', 'BHK', 'Baths', 'Fireplaces', 'GarageCars',
      'SaleCondition', 'SalePrice', 'Price_Per_Sqft', 'Bed_Bath_Ratio'],
      dtype='object')
```

```
df.head(10)
```

	Order	PID	Sqft	Utilities	Locality	FamilyAccom	HouseStyle	BHK	Baths	Fireplaces	Gara
0	1	526301100	31770.0	AllPub	NAmes	1Fam	1Story	1.0	1	2	
1	2	526350040	11622.0	AllPub	NAmes	1Fam	1Story	1.0	1	0	
2	3	526351010	14267.0	AllPub	NAmes	1Fam	1Story	1.0	1	0	
3	4	526353030	11160.0	AllPub	NAmes	1Fam	1Story	2.0	2	2	
4	5	527105010	13830.0	AllPub	Gilbert	1Fam	2Story	2.0	2	1	
5	6	527105030	9978.0	AllPub	Gilbert	1Fam	2Story	2.0	2	1	
6	7	527127150	4920.0	AllPub	StoneBr	1Fam	1Story	2.0	2	0	
7	8	527145080	5005.0	AllPub	StoneBr	1Fam	1Story	2.0	2	0	
8	9	527146030	5389.0	AllPub	StoneBr	1Fam	1Story	2.0	2	1	
9	10	527162130	7500.0	AllPub	Gilbert	1Fam	2Story	2.0	2	1	

```
df["SaleCondition"].value_counts()
```

```
Normal      89
Partial      8
Family       2
Abnormal     1
Name: SaleCondition, dtype: int64
```

```
from sklearn import preprocessing
```

```
le = preprocessing.LabelEncoder()
df["SaleCondition"]=le.fit_transform(df["SaleCondition"])
```

```
df["SaleCondition"].value_counts()
```

```
2      89
3       8
1       2
0       1
Name: SaleCondition, dtype: int64
```

```
df['FamilyAccom'].value_counts()
```

```
1Fam      97
2Fam       3
Name: FamilyAccom, dtype: int64
```

```
df[["No_Of_Families", "Fam"]]=(df["FamilyAccom"].str.split("F",expand=True))
```

```
df.columns
```

```
Index(['Order', 'PID', 'Sqft', 'Utilities', 'Locality', 'FamilyAccom',
      'HouseStyle', 'BHK', 'Baths', 'Fireplaces', 'GarageCars',
      'SaleCondition', 'SalePrice', 'Price_Per_Sqft', 'Bed_Bath_Ratio',
      'No_Of_Families', 'Fam'],
      dtype='object')
```

```
df["Avg_Price_Per_Locality"]=df.groupby("Locality")["SalePrice"].transform("mean")
```

```
df["Avg_Price_Per_Locality"]
```

```
0      160190.000000
1      160190.000000
2      160190.000000
3      160190.000000
4      185352.533333
...
95      212633.333333
96      212633.333333
97      212633.333333
98      134388.888889
99      134388.888889
Name: Avg_Price_Per_Locality, Length: 100, dtype: float64
```

```
df.dtypes
```

```

Order                int64
PID                  int64
Sqft                 float64
Utilities            object
Locality             object
FamilyAccom         object
HouseStyle          object
BHK                 float64
Baths               int64
Fireplaces          int64
GarageCars          int64
SaleCondition        int64
SalePrice           int64
Price_Per_Sqft      float64
Bed_Bath_Ratio      float64
No_Of_Families      object
Fam                 object
Avg_Price_Per_Locality float64
dtype: object

```

```

df["Locality"].unique()

array(['Names', 'Gilbert', 'StoneBr', 'NWAmes', 'Somerst', 'BrDale',
      'NPKvill', 'NridgHt', 'Blmngtn', 'NoRidge', 'SawyerW', 'Sawyer'],
      dtype=object)

```

```

df.columns

Index(['Order', 'PID', 'Sqft', 'Utilities', 'Locality', 'FamilyAccom',
      'HouseStyle', 'BHK', 'Baths', 'Fireplaces', 'GarageCars',
      'SaleCondition', 'SalePrice', 'Price_Per_Sqft', 'Bed_Bath_Ratio',
      'No_Of_Families', 'Fam', 'Avg_Price_Per_Locality'],
      dtype='object')

```

---

```

df=df.drop(["Order","PID"],axis=1)

```

```

df.columns

Index(['Sqft', 'Utilities', 'Locality', 'FamilyAccom', 'HouseStyle', 'BHK',
      'Baths', 'Fireplaces', 'GarageCars', 'SaleCondition', 'SalePrice',
      'Price_Per_Sqft', 'Bed_Bath_Ratio', 'No_Of_Families', 'Fam',
      'Avg_Price_Per_Locality'],
      dtype='object')

```

```

df["Utilities"].value_counts()

AllPub    100
Name: Utilities, dtype: int64

```

```

df=df.drop(["Utilities"],axis=1)

```