In [1]:
```
import numpy as np
import pandas as pd
```

In [2]:
```
df = pd.read_csv('Housing.csv')
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 545 entries, 0 to 544
Data columns (total 13 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   price             545 non-null    int64 
 1   area              545 non-null    int64 
 2   bedrooms          545 non-null    int64 
 3   bathrooms         545 non-null    int64 
 4   stories           545 non-null    int64 
 5   mainroad          545 non-null    object
 6   guestroom         545 non-null    object
 7   basement          545 non-null    object
 8   hotwaterheating   545 non-null    object
 9   airconditioning   545 non-null    object
 10  parking           545 non-null    int64 
 11  prefarea          545 non-null    object
 12  furnishingstatus  545 non-null    object
dtypes: int64(6), object(7)
memory usage: 55.5+ KB
```

In [4]:
```
df.head()
```

Out[4]:

| | price | area | bedrooms | bathrooms | stories | mainroad | guestroom | basement | hotwaterheating | airconditioning | parking | prefarea |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 13300000 | 7420 | 4 | 2 | 3 | yes | no | no | no | yes | 2 | yes |
| **1** | 12250000 | 8960 | 4 | 4 | 4 | yes | no | no | no | yes | 3 | no |
| **2** | 12250000 | 9960 | 3 | 2 | 2 | yes | no | yes | no | no | 2 | yes |
| **3** | 12215000 | 7500 | 4 | 2 | 2 | yes | no | yes | no | yes | 3 | yes |
| **4** | 11410000 | 7420 | 4 | 1 | 2 | yes | yes | yes | no | yes | 2 | no |

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

Out[5]:

| | price | area | bedrooms | bathrooms | stories | mainroad | guestroom | basement | hotwaterheating | airconditioning | parking | prefarea |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **540** | 1820000 | 3000 | 2 | 1 | 1 | yes | no | yes | no | no | 2 | no |
| **541** | 1767150 | 2400 | 3 | 1 | 1 | no | no | no | no | no | 0 | no |
| **542** | 1750000 | 3620 | 2 | 1 | 1 | yes | no | no | no | no | 0 | no |
| **543** | 1750000 | 2910 | 3 | 1 | 1 | no | no | no | no | no | 0 | no |
| **544** | 1750000 | 3850 | 3 | 1 | 2 | yes | no | no | no | no | 0 | no |

In [6]:
```
df.tail(12)
```

Out[6]:

| | price | area | bedrooms | bathrooms | stories | mainroad | guestroom | basement | hotwaterheating | airconditioning | parking | prefarea |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **533** | 2100000 | 2400 | 3 | 1 | 2 | yes | no | no | no | no | 0 | no |
| **534** | 2100000 | 3000 | 4 | 1 | 2 | yes | no | no | no | no | 0 | no |
| **535** | 2100000 | 3360 | 2 | 1 | 1 | yes | no | no | no | no | 1 | no |
| **536** | 1960000 | 3420 | 5 | 1 | 2 | no | no | no | no | no | 0 | no |
| **537** | 1890000 | 1700 | 3 | 1 | 2 | yes | no | no | no | no | 0 | no |
| **538** | 1890000 | 3649 | 2 | 1 | 1 | yes | no | no | no | no | 0 | no |
| **539** | 1855000 | 2990 | 2 | 1 | 1 | no | no | no | no | no | 1 | no |
| **540** | 1820000 | 3000 | 2 | 1 | 1 | yes | no | yes | no | no | 2 | no |
| **541** | 1767150 | 2400 | 3 | 1 | 1 | no | no | no | no | no | 0 | no |
| **542** | 1750000 | 3620 | 2 | 1 | 1 | yes | no | no | no | no | 0 | no |
| **543** | 1750000 | 2910 | 3 | 1 | 1 | no | no | no | no | no | 0 | no |
| **544** | 1750000 | 3850 | 3 | 1 | 2 | yes | no | no | no | no | 0 | no |

In [19]:
```
df.dropna(inplace=True)
```

In [20]:
```
x = df.iloc[:, 1:]
y = df.iloc[:, 0]
```

In [21]:
```
x.shape
```

Out[21]: (545, 12)

In [22]:
```
y.shape
```

Out[22]: (545,)

In [23]:
```
from sklearn.preprocessing import LabelEncoder
from collections import defaultdict
d = defaultdict(LabelEncoder)
xfit = x.apply(lambda x:d[x.name].fit_transform(x))
l = LabelEncoder()
yfit = l.fit_transform(y)
```

In [24]:
```
xfit
```

Out[24]:

| | area | bedrooms | bathrooms | stories | mainroad | guestroom | basement | hotwaterheating | airconditioning | parking | prefarea | furnishing |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 232 | 3 | 1 | 2 | 1 | 0 | 0 | 0 | 1 | 2 | 1 | |
| **1** | 260 | 3 | 3 | 3 | 1 | 0 | 0 | 0 | 1 | 3 | 0 | |
| **2** | 268 | 2 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 2 | 1 | |
| **3** | 237 | 3 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 3 | 1 | |
| **4** | 232 | 3 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 2 | 0 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **540** | 39 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 2 | 0 | |
| **541** | 15 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **542** | 72 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **543** | 35 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **544** | 90 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |

545 rows × 12 columns

In [37]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(xfit,yfit,test_size=0.3,random_state=0)
```

In [38]:
```
from sklearn.linear_model import LinearRegression
reg = LinearRegression()
```

In [39]:
```
reg.fit(x_train,y_train)
```

Out[39]:
```
LinearRegression()
```

In [40]:
```
print(reg.score(x_test,y_test))
```

```
0.7262896629399543
```

In [ ]:
```

```