

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

```
In [2]: df = pd.read_csv("homeprices.csv")
df
```

```
Out[2]:
```

	town	area	price
0	Chennai	2600	5500000
1	Chennai	3000	5650000
2	Chennai	3200	6100000
3	Chennai	3600	6800000
4	Bangalore	2600	5850000
5	Bangalore	2800	6150000
6	Bangalore	3300	6500000
7	Bangalore	3600	7100000
8	Hyderabad	2600	5750000
9	Hyderabad	2900	6000000
10	Hyderabad	3100	6200000
11	Hyderabad	3600	6950000

```
In [3]: df_dum = pd.get_dummies(df,drop_first=True)
df_dum
```

```
Out[3]:
```

	area	price	town_Chennai	town_Hyderabad
0	2600	5500000	1	0
1	3000	5650000	1	0
2	3200	6100000	1	0
3	3600	6800000	1	0
4	2600	5850000	0	0
5	2800	6150000	0	0
6	3300	6500000	0	0
7	3600	7100000	0	0
8	2600	5750000	0	1
9	2900	6000000	0	1
10	3100	6200000	0	1
11	3600	6950000	0	1

```
In [4]: X = df_dum.drop('price',axis='columns') # Independent Variables (X)
```

```
y = df_dum.price # Dependent Variable (y)
```

```
In [5]: X
```

```
Out[5]:
```

	area	town_Chennai	town_Hyderabad
0	2600	1	0
1	3000	1	0
2	3200	1	0
3	3600	1	0
4	2600	0	0
5	2800	0	0
6	3300	0	0
7	3600	0	0
8	2600	0	1
9	2900	0	1
10	3100	0	1
11	3600	0	1

```
In [6]: y
```

```
Out[6]:
```

0	5500000
1	5650000
2	6100000
3	6800000
4	5850000
5	6150000
6	6500000
7	7100000
8	5750000
9	6000000
10	6200000
11	6950000

Name: price, dtype: int64

```
In [7]: from sklearn.model_selection import train_test_split
```

```
In [8]: X_train,X_test = train_test_split(X,test_size=0.3)
print(X_train)
print(X_test)
```

	area	town_Chennai	town_Hyderabad
3	3600	1	0
4	2600	0	0
6	3300	0	0
0	2600	1	0
2	3200	1	0
5	2800	0	0
7	3600	0	0
10	3100	0	1
	area	town_Chennai	town_Hyderabad
9	2900	0	1
11	3600	0	1
1	3000	1	0
8	2600	0	1

```
In [9]: y_train,y_test = train_test_split(y,test_size=0.3)
print(y_train)
print(y_test)
```

```
6    6500000
3    6800000
1    5650000
9    6000000
5    6150000
7    7100000
2    6100000
11   6950000
Name: price, dtype: int64
0    5500000
10   6200000
4    5850000
8    5750000
Name: price, dtype: int64
```

```
In [10]: X_train,X_test,y_train,y_test = train_test_split(X, y,test_size=0.3,random_state=9)
```

```
In [11]: X_train
```

```
Out[11]:
```

	area	town_Chennai	town_Hyderabad
3	3600	1	0
4	2600	0	0
1	3000	1	0
11	3600	0	1
10	3100	0	1
8	2600	0	1
6	3300	0	0
5	2800	0	0

```
In [12]: X_test
```

Out[12]:

	area	town_Chennai	town_Hyderabad
7	3600	0	0
9	2900	0	1
2	3200	1	0
0	2600	1	0

	area	town_Chennai	town_Hyderabad
7	3600	0	0
9	2900	0	1
2	3200	1	0
0	2600	1	0

In [13]: y_train

Out[13]:

3	6800000
4	5850000
1	5650000
11	6950000
10	6200000
8	5750000
6	6500000
5	6150000

Name: price, dtype: int64

In [14]: y_test

Out[14]:

7	7100000
9	6000000
2	6100000
0	5500000

Name: price, dtype: int64