

```
In [1]: import pandas as pd
df = pd.read_csv('C:\\Users\\Asus\\Downloads\\Bengaluru_House_Data.csv')
df
```

Out[1]:

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00
...
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0	231.00
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	NaN	3600	5.0	NaN	400.00
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0	60.00
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCI	4689	4.0	1.0	488.00
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	NaN	550	1.0	1.0	17.00

13320 rows x 9 columns

```
In [2]: len(df)
```

Out[2]: 13320

```
In [3]: x=df[['area_type','availability','location','size','society','total_sqft','bath','balcony']]
y=df['price']
x
```

Out[3]:

	area_type	availability	location	size	society	total_sqft	bath	balcony
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0
...
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	NaN	3600	5.0	NaN
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCI	4689	4.0	1.0
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	NaN	550	1.0	1.0

13320 rows x 8 columns

```
In [4]: y
```

Out[4]:

0	39.07
1	120.00
2	62.00
3	95.00
4	51.00
...	
13315	231.00
13316	400.00
13317	60.00
13318	488.00
13319	17.00

Name: price, Length: 13320, dtype: float64

```
In [5]: from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=10)
```

```
In [6]: x_train
```

Out[6]:

	area_type	availability	location	size	society	total_sqft	bath	balcony
78	Built-up Area	Ready To Move	Kaval Byrasandra	2 BHK	NaN	460	1.0	0.0
6245	Super built-up Area	Ready To Move	Hulimavu	2 BHK	Reiteee	1127	2.0	1.0
3842	Super built-up Area	Ready To Move	Sarjapur Road	2 BHK	NaN	1263	2.0	1.0
703	Super built-up Area	Ready To Move	Kothanur	3 BHK	NaN	1580	3.0	3.0
7891	Super built-up Area	Ready To Move	Singasandra	2 BHK	NaN	1120	2.0	1.0
...
11633	Super built-up Area	Ready To Move	HSR Layout	3 BHK	PuontFa	1590	2.0	2.0
1344	Super built-up Area	Ready To Move	Rukmaiah Layout	3 BHK	NaN	1655	3.0	2.0
12815	Super built-up Area	Ready To Move	Kodichikkanahalli	2 BHK	NaN	900	2.0	1.0
7293	Built-up Area	Ready To Move	Kengeri	2 BHK	Sravema	1230	2.0	2.0
1289	Super built-up Area	Ready To Move	Mysore Road	2 BHK	NaN	1239	2.0	1.0

9324 rows × 8 columns

```
In [7]: x_test.head()
```

Out[7]:

	area_type	availability	location	size	society	total_sqft	bath	balcony
76	Super built-up Area	Ready To Move	Kalena Agrahara	2 BHK	NaN	1150	2.0	3.0
5487	Super built-up Area	Ready To Move	Tumkur Road	2 BHK	Llandee	1250	2.0	2.0
5195	Plot Area	Ready To Move	Varthur Road	4 Bedroom	NaN	1300	3.0	0.0
2411	Super built-up Area	Ready To Move	Dodda Nekkundi	2 BHK	SaisePa	850	2.0	3.0
13217	Plot Area	Ready To Move	T Dasarahalli	6 Bedroom	NaN	1200	3.0	NaN

```
In [8]: y_train.head()
```

Out[8]: 78 22.0
6245 65.0
3842 78.0
703 76.0
7891 60.0
Name: price, dtype: float64

```
In [9]: y_test.head()
```

Out[9]: 76 40.0
5487 33.0
5195 75.0
2411 28.0
13217 125.0
Name: price, dtype: float64