

```
import pandas as pd
import numpy as np
```

```
df = pd.read_csv('/content/House_Rent_Dataset.csv')
```

```
df.sample(5)
```

	Posted On	BHK	Rent	Size	Floor	Area Type	Area Locality	City	Furnishing Status	Tenant Preferred	Bathroom	Point of Contact
3136	2022-06-22	2	18000	750	1 out of 3	Super Area	Padur, Old Mahabalipuram Road	Chennai	Semi-Furnished	Bachelors/Family	2	Contact Agent
4366	2022-06-02	2	18000	1000	2 out of 3	Super Area	Kondapur	Hyderabad	Unfurnished	Bachelors/Family	2	Contact Owner
811	2022-06-26	3	40000	1100	21 out of 22	Super Area	Hudco Living Point, Bhandup West	Mumbai	Unfurnished	Family	2	Contact Owner
949	2022-07-06	2	40000	755	12 out of 31	Carpet Area	Omkar Ananta, Goregaon East	Mumbai	Semi-Furnished	Bachelors	2	Contact Agent
4726	2022-05-14	2	16000	1100	2 out of 5	Super Area	Basheer Bagh, NH 7	Hyderabad	Furnished	Family	2	Contact Agent

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

	0
Posted On	0
BHK	0
Rent	0
Size	0
Floor	0
Area Type	0
Area Locality	0
City	0
Furnishing Status	0
Tenant Preferred	0
Bathroom	0
Point of Contact	0

dtype: int64

```
df['Point of Contact'].value_counts() ,df['Area Type'].value_counts() , df['Tenant Preferred'].value_counts() ,df['Furnishing Status'].value_counts(),df['City'].value_
```

```
(Point of Contact
Contact Owner      3216
Contact Agent      1529
Contact Builder         1
Name: count, dtype: int64,
Area Type
Super Area      2446
Carpet Area     2298
Built Area         2
Name: count, dtype: int64,
Tenant Preferred
Bachelors/Family  3444
Bachelors         830
Family            472
Name: count, dtype: int64,
Furnishing Status
Semi-Furnished    2251
Unfurnished       1815
Furnished         680
Name: count, dtype: int64,
City
Mumbai      972
Chennai     891
Bangalore   886
Hyderabad   868
Delhi       605
Kolkata     524
Name: count, dtype: int64)
```


```
df = df.drop(columns=['Posted On' , 'Floor' , 'Area Locality'])
```

```
df.head()
```

	BHK	Rent	Size	Area Type	City	Furnishing Status	Tenant Preferred	Bathroom	Point of Contact
0	2	10000	1100	Super Area	Kolkata	Unfurnished	Bachelors/Family	2	Contact Owner
1	2	20000	800	Super Area	Kolkata	Semi-Furnished	Bachelors/Family	1	Contact Owner
2	2	17000	1000	Super Area	Kolkata	Semi-Furnished	Bachelors/Family	1	Contact Owner
3	2	10000	800	Super Area	Kolkata	Unfurnished	Bachelors/Family	1	Contact Owner
4	2	7500	850	Carpet Area	Kolkata	Unfurnished	Bachelors	1	Contact Owner

```
from sklearn.model_selection import train_test_split
x_train , x_test , y_train , y_test = train_test_split(df.drop(columns=['Rent']), df['Rent'],test_size=0.2,random_state=0)
```

```
x_train.head()
```



McAfee | WebAdvisor

Your download's being scanned.  
We'll let you know if there's an issue.

X

	BHK	Size	Area Type	City	Furnishing Status	Tenant Preferred	Bathroom	Point of Contact
4681	2	700	Super Area	Hyderabad	Unfurnished	Bachelors/Family	2	Contact Owner
630	2	650	Carpet Area	Mumbai	Furnished	Family	2	Contact Agent
1742	3	1200	Super Area	Bangalore	Semi-Furnished	Bachelors/Family	2	Contact Owner
3077	1	560	Super Area	Chennai	Unfurnished	Bachelors/Family	1	Contact Owner
2996	2	600	Super Area	Chennai	Furnished	Bachelors/Family	2	Contact Owner


```
from sklearn.preprocessing import OneHotEncoder , OrdinalEncoder
```

```
from sklearn.compose import ColumnTransformer
transformer = ColumnTransformer(transformers=(
    ('trf1', OrdinalEncoder(categories=[
        ['Built Area' , 'Carpet Area' , 'Super Area'],
        ['Unfurnished', 'Semi-Furnished' , 'Furnished'])),
    ['Area Type', 'Furnishing Status']),
    ('trf2' , OneHotEncoder(sparse_output=False , drop = 'first') ,['City', 'Tenant Preferred' , 'Point of Contact'])
),

remainder ='passthrough')
```

```
transformer.fit_transform(x_train).shape
```

(3796, 14)



McAfee | WebAdvisor

×

Your download's being scanned.

We'll let you know if there's an issue.