# Makaan Dataset Cleaning Assignment 1

First we will import all the necessary libraries required to work on the datasets In [1]:

```
# Importing all the necessary libraries
import pandas as pd
import numpy as np
from pandas profiling import ProfileReport
from sklearn.impute import SimpleImputer
import missingno as msno
# for plotting
import matplotlib.pyplot as plt
from feature engine.discretisation import EqualWidthDiscretiser
from feature engine.transformation import BoxCoxTransformer
# for plotting
import matplotlib.pyplot as plt
# for Q-Q plots
import scipy.stats as stats
from sklearn.model selection import train test split, cross val score
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import StandardScaler
from feature engine.encoding import OneHotEncoder
from feature engine.imputation import CategoricalImputer
from sklearn.linear model import LogisticRegression
from sklearn.naive bayes import GaussianNB
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy score, precision score, recall score, f1 score
C:\Users\ADMIN\AppData\Local\Temp\ipykernel 4472\337837679.py:5: DeprecationWarning:
`import pandas profiling` is going to be deprecated by April 1st. Please use `import
ydata profiling` instead.
  from pandas profiling import ProfileReport
Reading all the datasets and combining into one \( \)
In [2]:
makaan2 = pd.read csv("D:\SIBMB\SEM 3\ML\Data Cleaning\makaan new2.csv")
```

makaan3 = pd.read csv("D:\SIBMB\SEM 3\ML\Data Cleaning\makaan new3.csv")

```
In [4]:
```

makaan5 = pd.read\_csv("D:\SIBMB\SEM 3\ML\Data Cleaning\makaan\_new5.csv")

#### In [5]:

makaan6 = pd.read csv("D:\SIBMB\SEM 3\ML\Data Cleaning\makaan new6.csv")

#### In [6]:

# Combining all the four datasets into one

dataframes = (makaandata for makaandata in [makaan2, makaan3, makaan5, makaan6])
makaandata = pd.concat(dataframes, ignore index=True)

makaandata.head(n=10)

#### Out[6]:

	item	proj_na me	address	city	price	rate	price_rang	status
0		Royal Embassy	Nagarbhavi	Bangalore	63L	4,500/ sq ft the price/ sq ft of this propert	<i class="icon-&lt;br&gt;down-round"></i>	Ready to M move
1		Royal Embassy	Nagarbhavi	Bangalore	49.5L	4,500/ sq ft the price/ sq ft of this propert	<i class="icon-&lt;br&gt;down-round"></i>	Ready to M move
2		Royal Embassy	Nagarbhavi	Bangalore	49.5L	4,500/ sq ft the price/ sq ft of this propert	<i class="icon-&lt;br&gt;down-round"></i>	Ready to M move
	2 BHK Independent	Varanasi	Varanasi	Bangalore	72L	7,200/ sq ft the	<i class="icon-&lt;br&gt;down-round"></i>	Ready 3

	item	proj_na me	address	city	price	rate	price_rang	status
	House					price/ sq ft of this propert.		move
4	4 BHK Independent House	Varanasi	Varanasi	Bangalore	1.05Cr	4,038/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>	Ready to move
5	1 BHK Independent House	Urban Clusters Layout	Urban Clusters Layout	Bangalore	65L	5,416/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>	Ready to move
6	1 BHK Independent House	Urban Clusters Layout	Urban Clusters Layout	Bangalore	70L	5,833/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>	Ready to move
7	2 BHK Independent House	Varanasi	Varanasi	Bangalore	58L	6,444/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>	Ready 3 to y move

	item	proj_na me	address	city	price	rate	price_rang	status
8	2 BHK Independent House	Varanasi	Varanasi	Bangalore	69.5L	6,950/ sq ft the price/ sq ft of this propert	<i class="icon-&lt;br&gt;down-round"></i>	Ready to move
9	2 BHK Independent House	Varanasi	Varanasi	Bangalore	70L	7,000/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>	Ready to move

# In [7]:

makaandata.sample(n=10)

#Gives randomly any 10 values from the dataset after every execution

# Out[7]:

Out[7]	•				i		
	item	proj_name	address	city	price	rate	price_rang
8087	2 BHK Apartment	kengeri satellite town	kengeri satellite town	Bangalore	43.7L	3,800/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>
6200	Residential Plot	telecom smart city near somathahalli	Nallurahalli Road	Bangalore	14.38L	1,199/ sq ft the price/ sq ft of	<i class="icon-&lt;br&gt;down-round"></i>

	item	proj_name	address	city	price	rate	price_rang
7037	1 BHK Apartment	Subha 9 Sky Vue	Anekal City	Bangalore	22.88L	this propert 4,401/sq ft the price/sq ft of this propert	NaN
7927	2 BHK Apartment	Bellandur	Bellandur	Bangalore	61.9L	5,430/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>
6240	3 BHK Villa	Jakkuru	Jakkuru	Bangalore	72L	4,780/ sq ft the price/ sq ft of this propert.	NaN
3601	3 BHK Apartment	Sobha City Mykonos	Kannur on Thanisandra Main Road	Bangalore	1.2Cr	7,648/ sq ft the price/ sq ft of this propert.	<i class="icon-up-&lt;br&gt;round"></i>
1369	3 BHK	Banashankari	Banashankari	Bangalore	47.5L	3,800/	

	item	proj_name	address	city	price	rate	price_rang
	Apartment					sq ft the price/ sq ft of this propert.	down-round">
5205	2 BHK Villa	Vaishnavi 2bhk villas	Whitefield	Bangalore	45.25L	5,478/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>
5616	Residential Plot	ABD Downtown	Bommenahalli	Bangalore	27.54L	2,700/ sq ft the price/ sq ft of this propert	<i class="icon-&lt;br&gt;down-round"></i>
3590	4 BHK Apartment	Century Renata	Richmond Town	Bangalore	13Cr	26,162/ sq ft the price/ sq ft of this proper	<i class="icon-up-&lt;br&gt;round"></i>

# Checking for null or duplicate and unique values in the DataFrame¶

In [8]:

print(makaandata.shape)
(8158, 20)

In [9]:

print(makaandata.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8158 entries, 0 to 8157

```
Data columns (total 20 columns):
                      Non-Null Count Dtype
    Column
    ----
                      -----
 0
    item
                      8158 non-null
                                      object
 1
    proj name
                      8158 non-null
                                      object
                      8158 non-null object
 2
    address
                      8158 non-null object
 3
    city
 4
    price
                      8158 non-null
                                      object
                      8158 non-null object
 5
    rate
 6
    price rang
                      6869 non-null object
 7
    status
                      8158 non-null
                                      object
 8
                      8158 non-null
                                      object
    age
 9
                      8158 non-null
    nature
                                      object
                      8158 non-null
 10 description
                                      object
 11 price Negotiable 8158 non-null
                                      object
12 deposit
                      8158 non-null
                                      object
13 furniture
                      8158 non-null
                                      object
 14 score
                      8158 non-null
                                      object
15 Amenities
                      8158 non-null object
16 proj launch
                      8158 non-null
                                      object
17 builder
                      8158 non-null
                                      object
18 builder exp
                      8158 non-null
                                      object
19 past projects
                      8158 non-null
                                      object
dtypes: object(20)
memory usage: 1.2+ MB
None
In [10]:
makaandata.nunique()
print("Unique Values:\n", makaandata)
Unique Values:
                                                proj name \
                          item
0
             3 BHK Apartment
                                           Royal Embassy
1
             2 BHK Apartment
                                            Royal Embassy
2
             2 BHK Apartment
                                           Royal Embassy
3
     2 BHK Independent House
                                                Varanasi
     4 BHK Independent House
                                                Varanasi
                 5 BHK Villa State Bank Of India Colony
8153
8154
            Residential Plot
                                             GREEN WOODS
8155
             1 BHK Apartment
                                         Brigade Panorama
            Residential Plot
                                             abd downtown
8156
8157 2 BHK Independent House
                                           Artha reveria
                        address
                                      city
                                            price \
                                               63L
0
                     Nagarbhavi Bangalore
1
                     Nagarbhavi Bangalore
                                             49.5L
2
                     Nagarbhavi Bangalore
                                             49.5L
3
                       Varanasi Bangalore
                                              72L
                       Varanasi Bangalore 1.05Cr
4
                                       . . .
                                              . . .
. . .
8153 State Bank Of India Colony Bangalore
                                               90L
```

```
KR Puram Bhattarahalli Bangalore
8154
                                              19.8L
8155
                      Kumbalgodu Bangalore
                                                 50L
                                              16.2L
8156
                       Bandapura Bangalore
                          Marsur Bangalore
8157
                                              49.1L
                                                    rate \
       4,500/ sq ft the price/ sq ft of this propert...
0
1
       4,500/ sq ft the price/ sq ft of this propert...
2
       4,500/ sq ft the price/ sq ft of this propert...
3
       7,200/ sq ft the price/ sq ft of this propert...
4
       4,038/ sq ft the price/ sq ft of this propert...
. . .
       3,461/ sq ft the price/ sq ft of this propert...
8153
       1,650/ sq ft the price/ sq ft of this propert...
8154
8155
       5,630/ sq ft the price/ sq ft of this propert...
8156
       2,700/ sq ft the price/ sq ft of this propert...
8157
       4,091/ sq ft the price/ sq ft of this propert...
                                                              age nature \
                           price rang
                                               status
0
      <i class="icon-down-round"></i> Ready to move
                                                              Null
                                                                   Resale
1
      <i class="icon-down-round"></i></i>
                                       Ready to move
                                                              Null
                                                                    Resale
2
      <i class="icon-down-round"></i></i>
                                       Ready to move
                                                              Null
                                                                    Resale
3
      <i class="icon-down-round"></i> Ready to move 3 - 4 years
                                                                    Resale
      <i class="icon-down-round"></i> Ready to move
                                                      3 - 4 years
                                                                    Resale
8153 <i class="icon-down-round"></i> Ready to move
                                                      1 - 2 years
                                                                    Resale
8154 <i class="icon-down-round"></i>
                                        Unfurnished
                                                              Null
                                                                       New
8155 <i class="icon-down-round"></i>
                                      Ready to move 1 - 2 years
                                                                       New
                                        Unfurnished
8156 <i class="icon-down-round"></i>
                                                              Null
                                                                       New
8157 <i class="icon-down-round"></i> Ready to move 1 - 2 years
                                                                       New
                                             description price Negotiable
0
      3 BHK Apartment available for sale in Malagala...
                                                                       No
1
      2 BHK Apartment available for sale in Malagala...
                                                                       No
2
      2 BHK Apartment available for sale in Malagala...
                                                                       No
      2 BHK Independent House available for sale in ...
3
                                                                       No
4
      4 BHK Independent House available for sale in ...
                                                                       No
                                                                      . . .
     5 BHK Villa available for sale in , State Bank...
8153
                                                                       No
8154 A plot is available for sale in KR Puram Bhatt...
                                                                      No
8155
     It's a 1 bhk multistorey apartment situated in...
                                                                      No
8156 A plot is available at a prime location in Ban...
                                                                      Yes
8157 A 2 bhk property is available for sale in Mars...
                                                                      Yes
                       furniture score
         deposit
0
      No Deposit
                     Unfurnished
                                   8.2
                     Unfurnished 8.2
1
     No Deposit
2
     No Deposit
                     Unfurnished
                                  8.2
3
     No Deposit
                     Unfurnished Null
4
      No Deposit
                     Unfurnished Null
                                   . . .
                              . . .
. . .
8153 No Deposit Semi-Furnished Null
```

```
8154 No Deposit
                      Unfurnished Null
8155 No Deposit
                      Unfurnished
                                     8.8
8156 No Deposit
                      Unfurnished Null
8157 No Deposit
                      Unfurnished Null
                                                Amenities proj launch \
0
      ['Gymnasium', 'Landscaped Gardens', 'Lift Avai...
                                                                  Null
      ['Gymnasium', 'Landscaped Gardens', 'Lift Avai...
1
                                                                  Null
2
      ['Gymnasium', 'Landscaped Gardens', 'Lift Avai...
                                                                  Null
3
                                     ['Vaastu Compliant']
                                                                  Null
4
                                     ['Vaastu Compliant']
                                                                  Null
                                                       . . .
8153
                                                        []
                                                                  Null
8154
                                                        [1
                                                                  Null
     ['Gymnasium', 'Landscaped Gardens', 'Swimming ...
8155
                                                                    216
8156 ['Indoor Games', 'ATM', 'Jogging Track', 'Powe...
                                                                  Null
8157
     ['Landscaped Gardens', 'Lift Available', 'Swim...
                                                                  Null
                            builder builder exp past projects
0
      About builder: YGAM Builders
                                            Null
                                                              1
1
      About builder: YGAM Builders
                                            Null
                                                              1
2
      About builder: YGAM Builders
                                            Null
                                                              1
3
                               Null
                                            Null
                                                           Null
4
                                Null
                                            Null
                                                           Null
                                 . . .
                                             . . .
. . .
8153
                               Null
                                            Null
                                                           Null
8154
                               Null
                                            Null
                                                           Null
8155 About builder: Brigade Group
                                        34 years
                                                             42
8156
                               Null
                                            Null
                                                           Null
8157
                               Null
                                            Null
                                                           Null
[8158 rows x 20 columns]
In [11]:
null values = makaandata.isnull().sum()
print("Null Values:\n", null values)
Null Values:
 item
                         0
proj name
                        0
address
                        0
                        0
city
                        0
price
rate
                        0
                     1289
price rang
status
                        0
                        0
age
nature
                        0
                        0
description
price Negotiable
                        0
                        0
deposit
furniture
                        0
score
                        0
```

```
Amenities
                      0
proj launch
builder
                      0
                      0
builder exp
past projects
                      0
dtype: int64
In [12]:
(makaandata == "Null").sum()
Out[12]:
item
                      0
proj name
                      0
address
                      0
                      0
city
                      0
price
rate
price rang
                      0
                      0
status
                   4768
age
nature
                      0
                      0
description
price Negotiable
                      0
deposit
                      0
                      0
furniture
score
                   4410
Amenities
                      0
proj launch
                   5838
builder
                   5458
builder exp
                   6400
past projects
                   5634
dtype: int64
In [13]:
null values=makaandata.isnull()
# Count the number of null values in each column
null_counts=makaandata.isnull().sum()
# Display the DataFrame with null values (True where null, False where not null)
print(null values)
# Display the count of null values in each column
print(null counts)
      item proj name address city price rate price rang status \
0
     False
               False False False False
                                                         False
                                                                 False
1
     False
               False
                        False False False
                                                         False False
2
               False False False False
                                                         False False
    False
```

False False False False

... ... ...

False False

False False

False False

False False

False

False

3

False

False

. . .

8153 False

8155 False

8154 False

. . .

8156	False	False	e False	False	False	False	e 1	False	False	Э	
			e False								
	age	nature o	description	price	_Negoti	able	deposit	furn	iture	score	\
0	False	False	False		E	Talse	False		False	False	
1	False	False	False	:	E	Talse	False		False	False	
2	False	False	False	<b>:</b>	E	Talse	False		False	False	
3	False	False	False	<b>:</b>	E	Talse	False		False	False	
4	False	False	False	:	E	Talse	False		False	False	
		• • •									
			False				False				
			False				False				
			False				False				
			False								
8157	False	False	False	:	E	alse	False		False	False	
•		_	_launch bu				_				
0			False								
1			False								
			False								
3			False					False			
4			False								
0152			···								
			False					False			
8154 8155			False False								
	Fa.		False			alse		False			
	Fa		False								
0137	ra	136	raise	raise	L	aise		raise			
[8158	rows x	20 column	nsl								
item	200 11	20 0020	0								
proj	name		0								
addre	-		0								
city			0								
price	<u> </u>		0								
rate			0								
price	rang	12	289								
statu	_		0								
age			0								
natur	·e		0								
descr	iption		0								
price	_Negoti	able	0								
depos	it		0								
furni	ture		0								
score	!		0								
Ameni	ties		0								
proj_	launch		0								
build	ler		0								
build	ler_exp		0								
past_	project	S	0								
dtype	: int64										

# Handling the Null Values¶

# In [14]:

# Replace specific representations of null values with pandas NaN
makaandata.replace(['Null', '[]'], pd.NA, inplace=True)
makaandata

# Out[14]:

	item	proj_nam e	address	city	price	rate	price_rang	
0	3 BHK Apartment	Royal Embassy	Nagarbhavi	Bangalore	63L	4,500/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>	R
1	2 BHK Apartment	Royal Embassy	Nagarbhavi	Bangalore	49.5L	4,500/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>	Rem
2	2 BHK Apartment	Royal Embassy	Nagarbhavi	Bangalore	49.5L	4,500/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>	R(m
3	2 BHK Independent House	Varanasi	Varanasi	Bangalore	72L	7,200/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>	Rom

	item	proj_nam e	address	city	price	rate	price_rang	
4	4 BHK Independent House	Varanasi	Varanasi	Bangalore	1.05Cr	4,038/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>	Rem
•••								
8153	5 BHK Villa	State Bank Of India Colony	State Bank Of India Colony	Bangalore	90L	3,461/ sq ft the price/ sq ft of this propert	<i class="icon-&lt;br&gt;down-round"></i>	Rem
8154	Residential Plot	GREEN WOODS	KR Puram Bhattarahalli	Bangalore	19.8L	1,650/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>	Ui
8155	1 BHK Apartment	Brigade Panorama	Kumbalgodu	Bangalore	50L	5,630/ sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i>	Rem
8156	Residential Plot	abd downtown	Bandapura	Bangalore	16.2L	2,700/ sq ft the price/	<i class="icon-&lt;br&gt;down-round"></i>	U

item	proj_nam e	address	city	price	rate	price_rang
					sq ft of this propert.	
2 BHK Independent House	Artha reveria	Marsur	Bangalore	49.1L	4,091 / sq ft the price/ sq ft of this propert.	<i class="icon-&lt;br&gt;down-round"></i> n

#### $8158 \text{ rows} \times 20 \text{ columns}$

null values = makaandata.isnull()

#### In [15]:

print(null values) item proj name address city price rate price rang status \ 0 False False False False False False False 1 False False False False False False False 2 False False False False False False False 3 False False False False False False False

False False False False False False False . . . . . . ... ... ... . . . . . . . . . 8153 False False False False False False False

8154FalseFalseFalseFalseFalseFalse8155FalseFalseFalseFalseFalseFalse8156FalseFalseFalseFalseFalseFalse

8157 False False False False False False False False

nature description price Negotiable deposit furniture 0 False False False False False False True 1 True False False False False False False 2 True False False False False False False 3 False False False False False False True False False False False False False True . 8153 False False False False False False True 8154 True False False False False False True 8155 False False False False False False False True 8156 False False False False False True 8157 False False False False False False True

0	False	True	False	True	False
1	False	True	False	True	False
2	False	True	False	True	False
3	False	True	True	True	True
4	False	True	True	True	True
8153	True	True	True	True	True
8154	True	True	True	True	True
8155	False	False	False	False	False
8156	False	True	True	True	True
8157	False	True	True	True	True

[8158 rows x 20 columns]

#### **Dropping the unnecessary columns**

Removing the columns "city," "price range," and "deposit" from the dataset.¶

The reason for this is that these columns contain a limited number of unique values (1, 2, and 1, respectively).¶

In particular, the city is fixed as Bengaluru, and there are no deposits required for the houses. Hence, retaining these columns may not add significant value to the algorithms. ¶

Additionally, the "price range" column has only two distinct entries, which do not offer relevant information, further justifying its exclusion from the analysis.¶

#### In [17]:

makaandata.drop(columns=["city","price\_rang", "deposit"], inplace=True)
makaandata.head(n=10)

#### Out[17]:

	item	proj_na me	address	price	rate	status	age	nature	description	pric
0	3 BHK Apartment	Royal Embassy	Nagarbhavi	63L	4,500/ sq ft the price/ sq ft of this propert	Ready to move	< <b>NA</b> >	Resale	3 BHK Apartment available for sale in Malagala	No
1	2 BHK Apartment	Royal Embassy	Nagarbhavi	49.5L	4,500/ sq ft the price/ sq ft of this propert.	Ready to move	<na &gt;</na 	Resale	2 BHK Apartment available for sale in Malagala	No
2	2 BHK	Royal	Nagarbhavi	49.5L	4,500/	Ready	<na< th=""><th>Resale</th><th>2 BHK</th><th>No</th></na<>	Resale	2 BHK	No

	item	proj_na me	address	price	rate	status	age	nature	description	pric
	Apartment	Embassy			sq ft the price/ sq ft of this propert.	to move	>		Apartment available for sale in Malagala	
3	2 BHK Independent House	Varanasi	Varanasi	72L	7,200/ sq ft the price/ sq ft of this propert.	Ready to move	3 – 4 years	Resale	2 BHK Independent House available for sale in	No
4	4 BHK Independent House	Varanasi	Varanasi	1.05Cr	4,038/ sq ft the price/ sq ft of this propert.	Ready to move	3 – 4 years	Resale	4 BHK Independent House available for sale in	No
5	1 BHK Independent House	Urban Clusters Layout	Urban Clusters Layout	65L	5,416/ sq ft the price/ sq ft of this propert.	Ready to move	1 – 2 years	New	A spacious 1 bhk independent house is availabl	No
6	1 BHK Independent House	Urban Clusters Layout	Urban Clusters Layout	70L	5,833/ sq ft the price/ sq ft of this propert.	Ready to move	1 - 2 years	New	It's a 1 bhk independent house situated in Urb	No

	item	proj_na me	address	price	rate	status	age	nature	description	pric
7	2 BHK Independent House	Varanasi	Varanasi	58L	6,444/ sq ft the price/ sq ft of this propert.	Ready to move	3 – 4 years	Resale	2 BHK Independent House available for sale in	No
8	2 BHK Independent House	Varanasi	Varanasi	69.5L	6,950/ sq ft the price/ sq ft of this propert.	Ready to move	3 – 4 years	Resale	2 BHK Independent House available for sale in	No
9	2 BHK Independent House	Varanasi	Varanasi	70L	7,000/ sq ft the price/ sq ft of this propert.	Ready to move	3 - 4 years	Resale	2 BHK Independent House available for sale in	No

# Percentage of missing data in the entire dataset

In [18]:

```
print("age = ", (makaandata['age'] == "Null").sum()/len(makaandata))
print("score = ", (makaandata['score'] == "Null").sum()/len(makaandata))
print("proj_launch = ", (makaandata['proj_launch'] == "Null").sum()/len(makaandata))
print("builder = ", (makaandata['builder'] == "Null").sum()/len(makaandata))
print("builder_exp = ", (makaandata['builder_exp'] == "Null").sum()/len(makaandata))
print("past_projects = ", (makaandata['past_projects'] == "Null").sum()/len(makaandata))
age = 0.0
score = 0.0
proj_launch = 0.0
builder = 0.0
builder_exp = 0.0
past projects = 0.0
```

Here, We can clearly see that the columns proj\_launch and builder\_exp column has almost 70% of missing data  $\P$ 

Also the data in the proj\_launch column does not have a standard unit thus the values are ranging from numbers to date¶

So not much useful information could be derived from this column¶

In [19]:

makaandata.drop(columns=["proj\_launch","builder\_exp"], inplace=True)
makaandata.head(n=10)

#### Out[19]:

	item	proj_na me	address	price	rate	status	age	nature	description	prie
0	3 BHK Apartment	Royal Embassy	Nagarbhavi	63L	4,500/ sq ft the price/ sq ft of this propert.	Ready to move	< <b>NA</b> >	Resale	3 BHK Apartment available for sale in Malagala	No
1	2 BHK Apartment	Royal Embassy	Nagarbhavi	49.5L	4,500/ sq ft the price/ sq ft of this propert.	Ready to move	< <b>NA</b> >	Resale	2 BHK Apartment available for sale in Malagala	No
2	2 BHK Apartment	Royal Embassy	Nagarbhavi	49.5L	4,500/ sq ft the price/ sq ft of this propert.	Ready to move	< <b>NA</b> >	Resale	2 BHK Apartment available for sale in Malagala	No
3	2 BHK Independent House	Varanasi	Varanasi	72L	7,200/ sq ft the price/ sq ft of this	Ready to move	3 - 4 years	Resale	2 BHK Independent House available for sale in	No

	item	proj_na me	address	price	rate	status	age	nature	description	pric
					propert.					
4	4 BHK Independent House	Varanasi	Varanasi	1.05Cr	4,038/ sq ft the price/ sq ft of this propert.	Ready to move	3 – 4 years	Resale	4 BHK Independent House available for sale in	No
5	1 BHK Independent House	Urban Clusters Layout	Urban Clusters Layout	65L	5,416/ sq ft the price/ sq ft of this propert.	Ready to move	1 – 2 years	New	A spacious 1 bhk independent house is availabl	No
6	1 BHK Independent House	Urban Clusters Layout	Urban Clusters Layout	70L	5,833/ sq ft the price/ sq ft of this propert	Ready to move	1 – 2 years	New	It's a 1 bhk independent house situated in Urb	No
7	2 BHK Independent House	Varanasi	Varanasi	58L	6,444/ sq ft the price/ sq ft of this propert	Ready to move	3 – 4 years	Resale	2 BHK Independent House available for sale in	No
8	2 BHK Independent	Varanasi	Varanasi	69.5L	6,950/ sq ft the	Ready to	3 – 4	Resale	2 BHK Independent	No

	item	proj_na me	address	price	rate	status	age	nature	description	pric t
	House				price/ sq ft of this propert.	move	years		House available for sale in	
9	2 BHK Independent House	Varanasi	Varanasi	70L	7,000/ sq ft the price/ sq ft of this propert.	Ready to move	3 – 4 years	Resale	2 BHK Independent House available for sale in	No

### In [20]:

print(makaandata.shape)
(8158, 15)

# Handling the Rate column for numeric values ¶

```
In [21]:
```

import re

```
# Define a function to extract numeric part from 'rate' column
def extract_numeric_rate(rate_str):
    rate match = re.search(r'[\d,.]+', rate str)
    if rate_match:
        return float(rate match.group().replace(',', '').replace('L',
'00000').replace('Cr', '000000'))
    else:
        return None
In [22]:
# Define a function to extract string characters from 'rate' column
def extract string rate(rate str):
    rate match = re.search(r'[a-zA-Z]+', rate str)
    if rate match:
        return rate match.group()
    else:
        return None
In [23]:
# Convert 'rate' column to string to handle non-string entries
makaandata['rate'] = makaandata['rate'].astype(str)
In [24]:
# Apply the functions to create two separate columns for 'rate'
```

```
makaandata['rate value'] = makaandata['rate'].apply(extract numeric rate)
makaandata['rate description'] = makaandata['rate'].apply(lambda x: x.split('the price/
sq ft of ')[-1])
makaandata['rate value']
makaandata['rate description']
Out[24]:
0
        this property is lower than the avg price/ sq ...
1
        this property is lower than the avg price/ sq ...
        this property is lower than the avg price/ sq ...
2
3
        this property is lower than the avg price/ sq ...
        this property is lower than the avg price/ sq ...
8153
       this property is lower than the avg price/ sq ...
       this property is lower than the avg price/ sq ...
8154
8155
       this property is lower than the avg price/ sq ...
8156
       this property is lower than the avg price/ sq ...
8157
        this property is lower than the avg price/ sq ...
Name: rate description, Length: 8158, dtype: object
In [25]:
# Convert 'rate value' column back to float (since it was converted to string during
extraction)
makaandata['rate value'] = makaandata['rate value'].astype(float)
makaandata['rate value']
Out[25]:
0
        4500.0
1
        4500.0
2
        4500.0
3
        7200.0
        4038.0
        . . .
8153
        3461.0
8154
       1650.0
8155
       5630.0
8156 2700.0
8157
        4091.0
Name: rate value, Length: 8158, dtype: float64
In [26]:
```

#### Out[26]:

makaandata.head()

	item	proj_na me	address	price	rate	status	age	nature	description	prio t
	3 BHK Apartment	Royal Embassy	Nagarbhavi	63L	4,500/ sq ft the price/ sq ft of this	Ready to move	< <b>NA</b> >		3 BHK Apartment available for sale in Malagala	No

	item	proj_na me	address	price	rate	status	age	nature	description	pric
					propert.					
1		Royal Embassy	Nagarbhavi	49.5L	4,500/ sq ft the price/ sq ft of this propert.	Ready to move	<na &gt;</na 	Resale	2 BHK Apartment available for sale in Malagala	No
2	2 BHK Apartment	Royal Embassy	Nagarbhavi	49.5L	4,500/ sq ft the price/ sq ft of this propert.	Ready to move	< <b>NA</b> >	Resale	2 BHK Apartment available for sale in Malagala	No
3	2 BHK Independent House	Varanasi	Varanasi	72L	7,200/ sq ft the price/ sq ft of this propert.	Ready to move	3 - 4 years	Resale	2 BHK Independent House available for sale in	No
4	4 BHK Independent House	Varanasi	Varanasi	1.05Cr	4,038/ sq ft the price/ sq ft of this propert	Ready to move	3 – 4 years	Resale	4 BHK Independent House available for sale in	No

Handling the Price column for numeric values¶ In [27]:

```
import re
# Define a function to convert price from string to numeric
def convert_price_to_numeric(price_str):
    if 'L' in price str:
        return float(re.sub(r'[^\d.]', '', price str)) * 100000 # Convert Lakhs to
numeric value
    elif 'Cr' in price str:
       return float(re.sub(r'[^\d.]', '', price str)) * 10000000 # Convert Crores to
numeric value
   else:
        return None
# Convert 'price' column to numeric format
makaandata['price value'] = makaandata['price'].apply(convert_price_to_numeric)
# Display the updated DataFrame with numeric 'price value'
print(makaandata[['price', 'price_value']].head())
   price price value
     63L 6300000.0
0
   49.5L 4950000.0
   49.5L 4950000.0
2
    72L 7200000.0
3
4 1.05Cr 10500000.0
Dropping the Extra Rate & Price column for numeric values ¶
```

```
columns to drop = ['price','rate']
makaandata = makaandata.drop(columns=columns to drop)
```

# In [29]:

makaandata.sample(5)

#### Out[29]:

	item	proj_nam e	address	status	age	nature	description	price_N tiab
969	3 BHK Apartment	Reputed Amity Ramapriya Apartment	Banashankari	Ready to move	<na &gt;</na 	Resale	3 BHK Apartment available for sale in Promont	Yes
2557	Residential Plot	Hoskote	Hoskote	Ready to move	<na &gt;</na 	New	Residential Plot available for sale in Swamy V	No
7138	2 BHK	balaji elite	Hennagara	Ready to	1 – 2	New	It's a 2 bhk	Yes

	item	proj_nam e	address	status	age	nature	description	price_N tiabl
	Apartment	qq		move	years		multistorey apartment situated in	
3658	3 BHK Villa	Abhee Prakruthi Villa	Chandapura	Under Construction	<na &gt;</na 	New	Well designed 3 bhk villa is available at a pr	No
4912	3 BHK Villa	Yelahanka	Yelahanka	Under Construction	<na &gt;</na 	New	A khatha, \ nBIAAPA Approved \ n100% loan facilit	Yes

# In [30]:

makaandata.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8158 entries, 0 to 8157
Data columns (total 16 columns):

#	Column	Non-Null Count	Dtype
0	item	8158 non-null	object
1	proj_name	8158 non-null	object
2	address	8158 non-null	object
3	status	8158 non-null	object
4	age	3390 non-null	object
5	nature	8158 non-null	object
6	description	8158 non-null	object
7	price_Negotiable	8158 non-null	object
8	furniture	8158 non-null	object
9	score	3748 non-null	object
10	Amenities	6519 non-null	object
11	builder	2700 non-null	object
12	past_projects	2524 non-null	object
13	rate_value	8158 non-null	float64
14	rate_description	8158 non-null	object
15	price_value	8158 non-null	float64

dtypes: float64(2), object(14)

memory usage: 1019.9+ KB

# **Splitting the Amenities Column**¶ In [31]:

```
# Perform one-hot encoding on the 'Amenities' column
amenities dummies = makaandata['Amenities'].str.strip('[]').str.get dummies(', ')
# Concatenate the one-hot encoded DataFrame with the original DataFrame
makaandata = pd.concat([makaandata, amenities dummies], axis=1)
# Drop the original 'Amenities' column
makaandata.drop('Amenities', axis=1, inplace=True)
# Print the modified DataFrame
print(makaandata)
                         item
                                               proj name \
0
             3 BHK Apartment
                                          Royal Embassy
1
             2 BHK Apartment
                                          Royal Embassy
2
             2 BHK Apartment
                                           Royal Embassy
  2 BHK Independent House
3
                                                Varanasi
    4 BHK Independent House
4
                                                Varanasi
. . .
8153
                 5 BHK Villa State Bank Of India Colony
8154
           Residential Plot
                                            GREEN WOODS
            1 BHK Apartment
                                        Brigade Panorama
8155
8156
            Residential Plot
                                           abd downtown
8157 2 BHK Independent House
                                           Artha reveria
                        address
                                       status
                                                     age nature \
                     Nagarbhavi Ready to move
0
                                                      <NA> Resale
1
                     Nagarbhavi Ready to move
                                                     <NA> Resale
2
                     Nagarbhavi Ready to move
                                                      <NA> Resale
3
                       Varanasi Ready to move 3 - 4 years Resale
4
                       Varanasi Ready to move 3 - 4 years Resale
                                          . . .
8153 State Bank Of India Colony Ready to move 1 - 2 years Resale
8154
       KR Puram Bhattarahalli Unfurnished
                                                      <NA>
                                                             New
8155
                   Kumbalgodu Ready to move 1 - 2 years
                                                             New
8156
                      Bandapura Unfurnished
                                                     <NA>
                                                              New
8157
                         Marsur Ready to move 1 - 2 years
                                                              New
                                          description price Negotiable
     3 BHK Apartment available for sale in Malagala...
0
                                                                    No
1
     2 BHK Apartment available for sale in Malagala...
                                                                   No
     2 BHK Apartment available for sale in Malagala...
2
                                                                   No
3
     2 BHK Independent House available for sale in ...
                                                                   No
     4 BHK Independent House available for sale in ...
                                                                   No
                                                                   . . .
8153 5 BHK Villa available for sale in , State Bank...
                                                                   No
8154 A plot is available for sale in KR Puram Bhatt...
                                                                   No
8155 It's a 1 bhk multistorey apartment situated in...
                                                                   No
8156 A plot is available at a prime location in Ban...
                                                                  Yes
8157 A 2 bhk property is available for sale in Mars...
                                                                  Yes
```

	furniture	score		'Shopp	oing Ma	11'	'Sofa'	'Sports	Facility'	\
0	Unfurnished	8.2				0	C	)	1	
1	Unfurnished	8.2				0	C	)	1	
2	Unfurnished	8.2				0	C	)	1	
3	Unfurnished	<na></na>				0	C	)	0	
4	Unfurnished	<na></na>				0	C	)	0	
8153	Semi-Furnished	<na></na>				0	C	)	0	
8154	Unfurnished	<na></na>				0	C	)	0	
8155	Unfurnished	8.8				0	C	)	1	
8156	Unfurnished	<na></na>				1	C	)	1	
8157	Unfurnished	<na></na>				0	C	)	0	
	'Staff Quarter'	'Swin	nming	Pool'	'TV'	'Va	aastu C	Compliant'	'Wardrobe	' \
0	0			0	0			1		0
1	0			0	0			1		0
2	0			0	0			1		0
3	0			0	0			1		0
4	0			0	0			1		0
										•
8153	0			0	0			0		0
8154	0			0	0			0		0
8155	0			1	0			0		0
8156	1			1	0			1		0
8157	0			1	0			0		0
	'Washing Machin	ne' 'V	√ifi'							
0		0	0							
1		0	0							
2		0	0							
3		0	0							
4		0	0							
		• • •								
8153		0	0							

[8158 rows x 50 columns]

# In [32]:

amenities\_dummies

# Out[32]:

	"Children's play area"			'ATM	'BED '		'Car Parking'	'Club House '	'Dining Table'	'Gas connecti
0	1	1	0	0	0	0	1	0	0	0

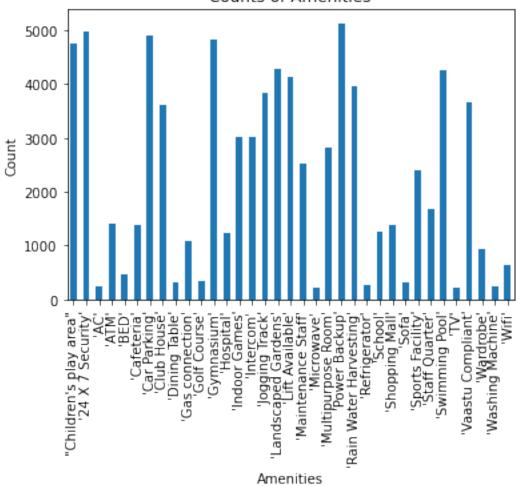
	"Children's play area"			'ATM	'BED	'Cafeteria '	'Car Parking'	'Club House	Dining	'Gas connect
1	1	1	0	0	0	0	1	0	0	0
2	1	1	0	0	0	0	1	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
8153	0	0	0	0	0	0	0	0	0	0
8154	0	0	0	0	0	0	0	0	0	0
8155	1	1	0	0	0	0	1	1	0	0
8156	1	1	0	1	0	0	1	0	0	0
8157	0	1	0	0	0	0	1	0	0	0

# 8158 rows × 35 columns

# In [33]:

```
subset_columns = amenities_dummies.columns
amenities_dummies[subset_columns].sum().plot(kind='bar')
plt.title('Counts of Amenities')
plt.xlabel('Amenities')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.show()
```

#### Counts of Amenities



#### In [34]:

amenities counts = makaandata.sum()

# Print the counts for each amenity
print(amenities\_counts)

3 BHK Apartment 2 BHK Apartment 2 BHK Apartmen... item Royal Embassy Royal Embassy Royal Embassy Vara... proj name address NagarbhaviNagarbhaviNagarbhaviVaranasiVaranasi... status Ready to moveReady to moveReady to moveReady t... ResaleResaleResaleResaleNewNewResaleResa... nature description 3 BHK Apartment available for sale in Malagala... price Negotiable furniture UnfurnishedUnfurnishedUnfurnishedUnfurnishedUn... 44049109.0 rate value rate description this property is lower than the avg price/ sq ... price\_value 77435486000.0 "Children's play area" 4758 '24 X 7 Security' 4965 'AC' 254 'ATM' 1410 'BED' 455 'Cafeteria' 1374 4903 'Car Parking' 'Club House' 3601

'Dining Table'	324
'Gas connection'	1086
'Golf Course'	335
'Gymnasium'	4836
'Hospital'	1246
'Indoor Games'	3023
'Intercom'	3011
'Jogging Track'	3830
'Landscaped Gardens'	4289
'Lift Available'	4131
'Maintenance Staff'	2527
'Microwave'	225
'Multipurpose Room'	2823
'Power Backup'	5130
'Rain Water Harvesting'	3966
'Refrigerator'	256
'School'	1264
'Shopping Mall'	1380
'Sofa'	305
'Sports Facility'	2396
'Staff Quarter'	1668
'Swimming Pool'	4256
'TV'	226
'Vaastu Compliant'	3664
'Wardrobe'	937
'Washing Machine'	251
'Wifi'	645
dtype: object	

dtype: object

C:\Users\ADMIN\AppData\Local\Temp\ipykernel\_4472\407195955.py:1: FutureWarning:
Dropping of nuisance columns in DataFrame reductions (with 'numeric\_only=None') is
deprecated; in a future version this will raise TypeError. Select only valid columns
before calling the reduction.

amenities\_counts = makaandata.sum()

#### In [35]:

makaandata.sample(5)

### Out[35]:

	item	proj_name	address	status	age	nature	description	price_ tial
5678	3 BHK Apartment	Prestige Song Of The South	Begur	Under Construction	< <b>NA</b> >	New	This spacious 3 bhk multistorey apartment is a	No

	item	proj_name	address	status	age	nature	description	price_ tial
4390	3 BHK Villa	Doddajala	Doddajala	Under Construction	< <b>NA</b> >	New	A 3 bhk property is available for sale in Dodd	Yes
8065	3 BHK Apartment	SWASTHIK GLOBAL NAVEEN	Ideal Homes Co Operative Housing Society	Ready to move	1 – 2 years	New	It's a 3 bhk multistorey apartment situated in	Yes
7140	3 BHK Independent House	Immadihalli Whitefield	Immadihalli Whitefield	Under Construction	< <b>NA</b> >	New	Well designed 3 bhk independent house is avail	No
1779	2 BHK Apartment	RMV Extension Stage 2	RMV Extension Stage 2	Ready to move	2 – 3 years	Resale	2 BHK Apartment available for sale in 10th Cro	No

# 5 rows $\times$ 50 columns

In [36]:

makaandata.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8158 entries, 0 to 8157

-----

Data columns (total 50 columns):

#	Column	Non-Null Count	Dtype
0	item	8158 non-null	object
1	proj_name	8158 non-null	object
2	address	8158 non-null	object
3	status	8158 non-null	object
4	age	3390 non-null	object
5	nature	8158 non-null	object

```
6
   description
                          8158 non-null object
7
   price Negotiable
                          8158 non-null object
8
  furniture
                          8158 non-null object
9
                          3748 non-null object
  score
10 builder
                         2700 non-null object
                         2524 non-null object
11 past projects
12 rate value
                         8158 non-null float64
                         8158 non-null object
13 rate description
14 price value
                         8158 non-null float64
15 "Children's play area" 8158 non-null int64
16 '24 X 7 Security'
                          8158 non-null int64
                          8158 non-null int64
17 'AC'
18 'ATM'
                          8158 non-null int64
19 'BED'
                          8158 non-null int64
                         8158 non-null int64
20 'Cafeteria'
21 'Car Parking'
                         8158 non-null int64
22 'Club House'
                         8158 non-null int64
23 'Dining Table'
                         8158 non-null int64
24 'Gas connection'
                         8158 non-null int64
25 'Golf Course'
                         8158 non-null int64
26 'Gymnasium'
                         8158 non-null int64
27 'Hospital'
                         8158 non-null int64
28 'Indoor Games'
                         8158 non-null int64
29 'Intercom'
                          8158 non-null int64
30 'Jogging Track'
                         8158 non-null int64
31 'Landscaped Gardens'
                         8158 non-null int64
32 'Lift Available'
                          8158 non-null int64
33 'Maintenance Staff'
                         8158 non-null int64
34 'Microwave'
                         8158 non-null int64
35 'Multipurpose Room'
                         8158 non-null int64
                          8158 non-null int64
36 'Power Backup'
37 'Rain Water Harvesting' 8158 non-null int64
38 'Refrigerator'
                          8158 non-null int64
                          8158 non-null int64
39 'School'
40 'Shopping Mall'
                         8158 non-null int64
41 'Sofa'
                          8158 non-null int64
42 'Sports Facility'
                         8158 non-null int64
43 'Staff Quarter'
                         8158 non-null int64
44 'Swimming Pool'
                         8158 non-null int64
45 'TV'
                          8158 non-null int64
46 'Vaastu Compliant'
                         8158 non-null int64
47 'Wardrobe'
                          8158 non-null int64
48 'Washing Machine'
                         8158 non-null int64
49 'Wifi'
                         8158 non-null int64
```

dtypes: float64(2), int64(35), object(13)

memory usage: 3.1+ MB

# Group by function is used for analysis \( \) In [37]:

# Group the data by 'item' and 'nature', and then count the occurrences of Nature of the Makaan.

```
nature_counts = makaandata.groupby(['item', 'nature'])['nature'].count()
```

- # Reset the index of the resulting Series and convert it back to a DataFrame
  nature counts = nature counts.reset index(name='count')
- # Pivot the DataFrame to get 'item' and counts of 'New' and 'Resale' as separate columns

nature\_counts\_pivot = nature\_counts.pivot\_table(index=['item'], columns='nature',
values='count', fill value=0)

- # Rename the columns for better clarity
  nature\_counts\_pivot.rename(columns={'New': 'Count\_New', 'Resale': 'Count\_Resale'},
  inplace=True)
- # Reset the index to get 'item' and 'proj\_name' as separate columns
  nature\_counts\_pivot.reset\_index(inplace=True)
- # Print the resulting DataFrame with counts of 'New' and 'Resale'
  nature\_counts\_pivot

#### Out[37]:

nature	item	Count_New	Count_Resale
0	Residential Plot	1197	253
1	1 BHK Apartment	186	33
2	1 BHK Independent Floor	3	0
3	1 BHK Independent House	188	14
4	1 BHK Villa	6	0
5	1 RK Studio Apartment	1	2
6	10 BHK Independent House	11	0
7	11 BHK Independent House	1	0
8	11 BHK Villa	1	0
9	14 BHK Independent House	0	1
10	15 BHK Apartment	0	1
11	2 BHK Apartment	1669	467

nature	item	Count_New	Count_Resale
12	2 BHK Independent Floor	11	20
13	2 BHK Independent House	308	202
14	2 BHK Villa	279	8
15	3 BHK Apartment	1361	395
16	3 BHK Independent Floor	14	14
17	3 BHK Independent House	185	87
18	3 BHK Villa	492	51
19	3 BHK penthouse	4	0
20	4 BHK Apartment	235	33
21	4 BHK Independent Floor	1	8
22	4 BHK Independent House	61	67
23	4 BHK Villa	98	30
24	5 BHK Apartment	5	2
25	5 BHK Independent Floor	2	1
26	5 BHK Independent House	19	36
27	5 BHK Villa	10	18
28	6 BHK Apartment	3	0
29	6 BHK Independent Floor	0	1
30	6 BHK Independent House	11	19
31	6 BHK Villa	2	1

nature	item	Count_New	Count_Resale
32	7 BHK Apartment	2	0
33	7 BHK Independent Floor	2	0
34	7 BHK Independent House	1	6
35	7 BHK Villa	3	0
36	8 BHK Independent House	4	4
37	9 BHK Independent House	1	7

# In [38]:

#Counting wheather it is Negotiable or not with respect ot item
negotiable\_counts = makaandata.groupby('item')
['price\_Negotiable'].value\_counts().unstack(fill\_value=0)

# Rename the columns to have more meaningful names
negotiable\_counts.rename(columns={'No': 'Count\_Negotiable\_No', 'Yes':
'Count\_Negotiable\_Yes'}, inplace=True)

negotiable\_counts

#### Out[38]:

price_Negotiable	Count_Negotiable_No	Count_Negotiable_Yes
item		
Residential Plot	968	482
1 BHK Apartment	105	114
1 BHK Independent Floor	3	0
1 BHK Independent House	202	0
1 BHK Villa	3	3
1 RK Studio Apartment	3	0

price_Negotiable	Count_Negotiable_No	Count_Negotiable_Yes
item		
10 BHK Independent House	11	0
11 BHK Independent House	0	1
11 BHK Villa	0	1
14 BHK Independent House	0	1
15 BHK Apartment	0	1
2 BHK Apartment	830	1306
2 BHK Independent Floor	12	19
2 BHK Independent House	319	191
2 BHK Villa	11	276
3 BHK Apartment	887	869
3 BHK Independent Floor	21	7
3 BHK Independent House	143	129
3 BHK Villa	265	278
3 BHK penthouse	0	4
4 BHK Apartment	193	75
4 BHK Independent Floor	8	1
4 BHK Independent House	85	43
4 BHK Villa	61	67
5 BHK Apartment	3	4

price_Negotiable	Count_Negotiable_No	Count_Negotiable_Yes
item		
5 BHK Independent Floor	3	0
5 BHK Independent House	34	21
5 BHK Villa	10	18
6 BHK Apartment	2	1
6 BHK Independent Floor	1	0
6 BHK Independent House	20	10
6 BHK Villa	1	2
7 BHK Apartment	2	0
7 BHK Independent Floor	0	2
7 BHK Independent House	4	3
7 BHK Villa	3	0
8 BHK Independent House	6	2
9 BHK Independent House	4	4

# Treating the Status and Furniture Column¶ In [39]:

```
# Apply one-hot encoding to the 'status' column
status dummies = pd.get dummies(makaandata['status'])
```

<sup>#</sup> Concatenate the dummy columns with the original DataFrame
makaandata = pd.concat([makaandata, status\_dummies], axis=1)

<sup>#</sup> Drop the original 'status' column
makaandata.drop(columns=['status'], inplace=True)

```
# Now 'mrg df cleaned' will have the dummy variables for 'status' as separate columns
print(makaandata.head())
                                  proj name
                                                address
                                                                 age nature
                       item
           3 BHK Apartment Royal Embassy Nagarbhavi
                                                                <NA>
                                                                      Resale
1
          2 BHK Apartment
                            Royal Embassy Nagarbhavi
                                                                <NA>
                                                                      Resale
2
          2 BHK Apartment Royal Embassy Nagarbhavi
                                                                <NA>
                                                                      Resale
3 2 BHK Independent House
                                  Varanasi
                                             Varanasi 3 - 4 years
                                                                      Resale
                                  Varanasi
  4 BHK Independent House
                                               Varanasi 3 - 4 years Resale
                                         description price Negotiable
  3 BHK Apartment available for sale in Malagala...
  2 BHK Apartment available for sale in Malagala...
                                                                   No
  2 BHK Apartment available for sale in Malagala...
                                                                   No
  2 BHK Independent House available for sale in ...
                                                                   No
4 4 BHK Independent House available for sale in ...
                                                                   No
     furniture score
                                           builder ... 'Staff Quarter'
O Unfurnished 8.2 About builder: YGAM Builders
1 Unfurnished 8.2 About builder: YGAM Builders ...
                                                                      0
2 Unfurnished 8.2 About builder: YGAM Builders ...
                                                                      0
3 Unfurnished <NA>
                                                                      0
                                              <NA>
4 Unfurnished <NA>
                                              <NA> ...
                                                                      0
   'Swimming Pool' 'TV'
                         'Vaastu Compliant'
                                             'Wardrobe'
                                                         'Washing Machine'
0
                 0
                                          1
                                                      0
                                                                         0
                      0
                                          1
1
                 0
                                                      0
                                                                         0
                                          1
                                                      0
2
                      0
                                                                         0
3
                 0
                                          1
                                                      0
                                                                         0
                      0
                      0
                                          1
4
   'Wifi'
          Ready to move
                         Under Construction Unfurnished
0
        0
                       1
1
        0
                       1
                                           0
                                                        0
2
                                           0
                                                        0
        0
                       1
3
        0
                       1
4
        0
                                           0
                       1
                                                        0
[5 rows x 52 columns]
In [40]:
# Apply one-hot encoding to the 'furniture' column
furniture_dummies = pd.get_dummies(makaandata['furniture'])
# Concatenate the dummy columns with the original DataFrame
makaandata = pd.concat([makaandata, status dummies], axis=1)
# Drop the original 'furniture' column
makaandata.drop(columns=['furniture'], inplace=True)
# Now 'mrg df cleaned' will have the dummy variables for 'status' as separate columns
print(makaandata.head())
                       item
                                  proj name
                                                                 age nature \
                                                address
```

```
0
          3 BHK Apartment
                            Royal Embassy Nagarbhavi
                                                               <NA> Resale
1
          2 BHK Apartment Royal Embassy Nagarbhavi
                                                               <NA> Resale
2
          2 BHK Apartment Royal Embassy Nagarbhavi
                                                                <NA> Resale
3 2 BHK Independent House
                                  Varanasi
                                            Varanasi 3 - 4 years Resale
4 4 BHK Independent House
                                             Varanasi 3 - 4 years Resale
                                  Varanasi
                                        description price Negotiable score \
  3 BHK Apartment available for sale in Malagala...
                                                                        8.2
  2 BHK Apartment available for sale in Malagala...
                                                                   No
                                                                        8.2
  2 BHK Apartment available for sale in Malagala...
                                                                   No
                                                                        8.2
  2 BHK Independent House available for sale in ...
                                                                   No <NA>
4 4 BHK Independent House available for sale in ...
                                                                   No <NA>
                       builder past projects ... 'Vaastu Compliant'
O About builder: YGAM Builders
                                            1
                                           1 ...
                                                                     1
1 About builder: YGAM Builders
2 About builder: YGAM Builders
                                            1 ...
                                                                     1
3
                          <NA>
                                                                     1
                                        <NA>
                                               . . .
4
                          <NA>
                                        <NA> ...
                                                                     1
  'Wardrobe'
              'Washing Machine'
                                 'Wifi' Ready to move Under Construction
                                                     1
0
          0
                              0
                                     0
1
          0
                              0
                                      0
                                                     1
                                                                         0
2
                              0
                                      0
                                                                         0
          0
                                                     1
3
          0
                              0
                                      0
                                                     1
                                                                         0
4
          0
                              0
                                      0
                                                     1
  Unfurnished Ready to move Under Construction Unfurnished
            0
0
                           1
                                                0
1
             0
                           1
                                                0
                                                             0
2
             0
                                                0
                                                             0
                           1
3
             0
                           1
                                                0
                                                             0
4
                                                0
                                                             0
                           1
```

[5 rows x 54 columns]

#### In [41]:

makaandata.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8158 entries, 0 to 8157
Data columns (total 54 columns):

#	Column	Non-Null Count	Dtype
0	item	8158 non-null	object
1	proj_name	8158 non-null	object
2	address	8158 non-null	object
3	age	3390 non-null	object
4	nature	8158 non-null	object
5	description	8158 non-null	object
6	price_Negotiable	8158 non-null	object
7	score	3748 non-null	object
8	builder	2700 non-null	object

```
9
   past projects
                           2524 non-null object
 10 rate value
                           8158 non-null float64
                          8158 non-null object
 11 rate description
12 price value
                           8158 non-null float64
13 "Children's play area" 8158 non-null int64
14 '24 X 7 Security'
                          8158 non-null int64
15 'AC'
                           8158 non-null int64
16 'ATM'
                           8158 non-null int64
17 'BED'
                           8158 non-null int64
18 'Cafeteria'
                           8158 non-null int64
19 'Car Parking'
                          8158 non-null int64
 20 'Club House'
                          8158 non-null int64
 21 'Dining Table'
                          8158 non-null int64
                          8158 non-null int64
 22 'Gas connection'
 23 'Golf Course'
                          8158 non-null int64
 24 'Gymnasium'
                          8158 non-null int64
 25 'Hospital'
                          8158 non-null int64
 26 'Indoor Games'
                          8158 non-null int64
 27 'Intercom'
                          8158 non-null int64
 28 'Jogging Track'
                          8158 non-null int64
 29 'Landscaped Gardens'
                          8158 non-null int64
 30 'Lift Available'
                          8158 non-null int64
 31 'Maintenance Staff'
                          8158 non-null int64
 32 'Microwave'
                           8158 non-null int64
 33 'Multipurpose Room'
                          8158 non-null int64
 34 'Power Backup'
                          8158 non-null int64
   'Rain Water Harvesting' 8158 non-null int64
                          8158 non-null int64
 36 'Refrigerator'
 37 'School'
                          8158 non-null int64
 38 'Shopping Mall'
                           8158 non-null int64
                          8158 non-null int64
 39 'Sofa'
 40 'Sports Facility'
                          8158 non-null int64
 41 'Staff Quarter'
                           8158 non-null int64
 42 'Swimming Pool'
                          8158 non-null int64
 43 'TV'
                          8158 non-null int64
 44 'Vaastu Compliant'
                          8158 non-null int64
 45 'Wardrobe'
                           8158 non-null int64
 46 'Washing Machine'
                          8158 non-null int64
 47 'Wifi'
                           8158 non-null int64
                          8158 non-null uint8
 48 Ready to move
 49 Under Construction
                          8158 non-null uint8
 50 Unfurnished
                           8158 non-null uint8
 51 Ready to move
                           8158 non-null uint8
 52 Under Construction
                          8158 non-null uint8
 53 Unfurnished
                           8158 non-null uint8
dtypes: float64(2), int64(35), object(11), uint8(6)
memory usage: 3.0+ MB
```

#### In [42]:

# Count unique values of 'Ready to move' and 'Under Construction' with respect to
'item' and 'price\_value'
status\_cols = ['Ready to move', 'Under Construction', 'Unfurnished']

```
status_count = makaandata.groupby(['item', 'price_value'])
[status_cols].sum().reset_index()
```

status\_count

# Out[42]:

Out[42	-1-	<u> </u>		I			
	item	price_valu e	Ready to move	Ready to move	Under	Under Construction	Unfurnished
0	Residential Plot	190000.0	0	0	0	0	1
1	Residential Plot	300000.0	0	0	0	0	1
2	Residential Plot	360000.0	0	0	0	0	2
3	Residential Plot	504000.0	1	1	0	0	0
4	Residential Plot	599000.0	5	5	0	0	0
2625	9 BHK Independent House	15000000.0	1	1	0	0	0
2626	9 BHK Independent House	16000000.0	2	2	0	0	0
2627	9 BHK Independent House	16500000.0	1	1	0	0	0
2628	9 BHK Independent	25000000.0	1	1	0	0	0

	item	price_valu e	Ready to move	Ready to move	Under	Under Construction	Unfurnished
	House						
2629	9 BHK Independent House	50000000.0	1	1	0	0	0

 $2630 \text{ rows} \times 8 \text{ columns}$