

# Assignment No.11

```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
%matplotlib inline
import matplotlib
matplotlib.rcParams["figure.figsize"] = (20,10)
```

```
df1 = pd.read_csv("Bengaluru_House_Data.csv")
df1.head()
```

	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

```
df1.shape
```

```
(13320, 9)
```

```
df1.columns
```

```
Index(['area_type', 'availability', 'location', 'size', 'society',  
      'total_sqft', 'bath', 'balcony', 'price'],  
      dtype='object')
```

```
df1['area_type']
```

```
0      Super built-up Area
1           Plot Area
2      Built-up Area
3      Super built-up Area
4      Super built-up Area
...
13315      Built-up Area
13316      Super built-up Area
13317      Built-up Area
13318      Super built-up Area
13319      Super built-up Area
Name: area_type, Length: 13320, dtype: object
```

```
df1['area_type'].unique()
```

```
array(['Super built-up Area', 'Plot Area', 'Built-up Area',  
      'Carpet Area'], dtype=object)
```

```
df1['area_type'].value_counts()
```

```
area_type
Super built-up Area    8790
Built-up Area          2418
Plot Area              2025
Carpet Area             87
Name: count, dtype: int64
```

```
df2 = df1.drop(['area_type', 'society', 'balcony', 'availability'], axis='columns')
df2.shape
```

```
(13320, 5)
```

```
df2.isnull().sum()
```

```
location      1
size          16
total_sqft     0
bath           73
price          0
dtype: int64
```

```
df2.shape
```

```
(13320, 5)
```

```
df3 = df2.dropna()
df3.isnull().sum()
```

```
location      0
size          0
total_sqft     0
bath          0
price         0
dtype: int64
```

```
df3.shape
```

```
(13246, 5)
```

```
df3['size'].unique()
```

```
array(['2 BHK', '4 Bedroom', '3 BHK', '4 BHK', '6 Bedroom', '3 Bedroom',
      '1 BHK', '1 RK', '1 Bedroom', '8 Bedroom', '2 Bedroom',
      '7 Bedroom', '5 BHK', '7 BHK', '6 BHK', '5 Bedroom', '11 BHK',
      '9 BHK', '9 Bedroom', '27 BHK', '10 Bedroom', '11 Bedroom',
      '10 BHK', '19 BHK', '16 BHK', '43 Bedroom', '14 BHK', '8 BHK',
      '12 Bedroom', '13 BHK', '18 Bedroom'], dtype=object)
```

```
df3['bhk'] = df3['size'].apply(lambda x: int(x.split(' ')[0]))
```

C:\Users\vaishnavi pawar\AppData\Local\Temp\ipykernel\_40352\2222900254.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df3['bhk'] = df3['size'].apply(lambda x: int(x.split(' ')[0]))
```

```
df3['bhk'] = df3['size'].apply(lambda x: int(x.split(' ')[0]))
df3.head()
```

C:\Users\vaishnavi.pawar\AppData\Local\Temp\ipykernel\_40352\2448623301.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df3['bhk'] = df3['size'].apply(lambda x: int(x.split(' ')[0]))
```

	location	size	total_sqft	bath	price	bhk
0	Electronic City Phase II	2 BHK	1056	2.0	39.07	2
1	Chikka Tirupathi	4 Bedroom	2600	5.0	120.00	4
2	Uttarahalli	3 BHK	1440	2.0	62.00	3
3	Lingadheeranahalli	3 BHK	1521	3.0	95.00	3
4	Kothanur	2 BHK	1200	2.0	51.00	2

```
df3.bhk.unique()
```

```
array([ 2,  4,  3,  6,  1,  8,  7,  5, 11,  9, 27, 10, 19, 16, 43, 14, 12,
        13, 18], dtype=int64)
```

```
df3[df3.bhk>20]
```

	location	size	total_sqft	bath	price	bhk
1718	2Electronic City Phase II	27 BHK	8000	27.0	230.0	27
4684	Munnekollal	43 Bedroom	2400	40.0	660.0	43

```
df3.total_sqft.unique()
```

```
array(['1056', '2600', '1440', ..., '1133 - 1384', '774', '4689'],
      dtype=object)
```