Assignment 10

Data Wrangling on Real Estate Market

In [ ]:
```
! pip install pandas numpy matplotlib
```

```
Requirement already satisfied: pandas in /home/atharv/Downloads/practicals/.main-env/l
ib/python3.12/site-packages (2.2.2)
Requirement already satisfied: numpy in /home/atharv/Downloads/practicals/.main-env/li
b/python3.12/site-packages (2.1.1)
Requirement already satisfied: matplotlib in /home/atharv/Downloads/practicals/.main-e
nv/lib/python3.12/site-packages (3.9.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /home/atharv/Downloads/practi
cals/.main-env/lib/python3.12/site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /home/atharv/Downloads/practicals/.main
-env/lib/python3.12/site-packages (from pandas) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in /home/atharv/Downloads/practicals/.ma
in-env/lib/python3.12/site-packages (from pandas) (2024.1)
Requirement already satisfied: contourpy>=1.0.1 in /home/atharv/Downloads/practicals/.
main-env/lib/python3.12/site-packages (from matplotlib) (1.3.0)
Requirement already satisfied: cycler>=0.10 in /home/atharv/Downloads/practicals/.main
-env/lib/python3.12/site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /home/atharv/Downloads/practical
s/.main-env/lib/python3.12/site-packages (from matplotlib) (4.53.1)
Requirement already satisfied: kiwisolver>=1.3.1 in /home/atharv/Downloads/practical
s/.main-env/lib/python3.12/site-packages (from matplotlib) (1.4.7)
Requirement already satisfied: packaging>=20.0 in /home/atharv/Downloads/practicals/.m
ain-env/lib/python3.12/site-packages (from matplotlib) (24.1)
Requirement already satisfied: pillow>=8 in /home/atharv/Downloads/practicals/.main-en
v/lib/python3.12/site-packages (from matplotlib) (10.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in /home/atharv/Downloads/practicals/.
main-env/lib/python3.12/site-packages (from matplotlib) (3.1.4)
Requirement already satisfied: six>=1.5 in /home/atharv/Downloads/practicals/.main-en
v/lib/python3.12/site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
```

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

In [ ]:
```
df1 = pd.read_csv("Bengaluru_House_Data.csv")
df1.head()
```

Out[ ]:

	area_type	availability	location	size	society	total_sqft	bath	balcony	pric
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.0
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.0
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.0
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.0
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.0

In [ ]:
```
df1.shape
```

Out[ ]:
```
(13320, 9)
```

In [ ]:
```
df1.columns
```

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

In [ ]:
```
df1['area_type']
```

Out[ ]:
```
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
```

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

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

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

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

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

Out[ ]:
```
(13320, 5)
```

In [ ]:
```
df2.isnull().sum()
```

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

In [ ]:
```
df2.shape
```

Out[ ]:
```
(13320, 5)
```

In [ ]:
```
df3 = df2.dropna()
df3.isnull().sum()
```

Out[ ]:
```
location      0
size          0
total_sqft    0
bath          0
price         0
dtype: int64
```

In [ ]:
```
df3.shape
```

Out[ ]:
```
(13246, 5)
```

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

Out[ ]:
```
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)
```

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

```
/tmp/ipykernel_77559/2222900254.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
  df3['bhk'] = df3['size'].apply(lambda x: int(x.split(' ')[0]))
```

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

```
/tmp/ipykernel_77559/2448623301.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
  df3['bhk'] = df3['size'].apply(lambda x: int(x.split(' ')[0]))
```

Out[ ]:

	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

In [ ]:
```
df3.bhk.unique()
```

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

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

Out[ ]:

	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

In [ ]:
```
df3.total_sqft.unique()
```

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

In [ ]:
```

```