```
GROUP 6
```

In [16]: import numpy as np import pandas as pd import matplotlib.pyplot as plt import seaborn as sb

no

no

0

no

unfurnished

In [15]: df=pd.read_csv("Housing.csv")

In [4]: **df**

Out[4]:

price area bedrooms bathrooms stories mainroad guestroom basement hotwaterheating airconditioning parking prefarea furnishingstatus **0** 13300000 7420 3 2 furnished yes no yes **1** 12250000 8960 4 furnished yes no no no yes 3 no **2** 12250000 9960 3 2 2 2 semi-furnished yes no yes no yes **3** 12215000 7500 2 3 furnished yes yes no no yes yes 4 2 2 **4** 11410000 7420 1 furnished yes yes yes no yes no 1820000 3000 2 540 1 1 yes yes no no 2 no unfurnished **541** 1767150 2400 3 1 1 no no no no no 0 no semi-furnished 542 1750000 3620 2 1 1 0 unfurnished yes no no no 543 1750000 2910 furnished 1 1 no 0 no no no no no 3 **544** 1750000 3850 1 2 0 unfurnished yes no no no no no

545 rows × 13 columns

df.head()

price area bedrooms bathrooms stories mainroad guestroom basement hotwaterheating airconditioning parking prefarea furnishingstatus Out[5]: **0** 13300000 7420 3 yes no no 2 yes furnished no yes **1** 12250000 8960 yes no no no no furnished yes **2** 12250000 9960 3 2 2 yes 2 semi-furnished yes no no no yes **3** 12215000 7500 furnished 2 3 no yes no yes yes **4** 11410000 7420 2 2 furnished yes no yes no yes yes

In [6]: df.tail()

Out[6]: price area bedrooms bathrooms stories mainroad guestroom basement hotwaterheating airconditioning parking prefarea furnishingstatus **540** 1820000 3000 2 unfurnished 1 1 2 yes no yes no no no **541** 1767150 2400 no no no no semi-furnished **542** 1750000 3620 2 1 1 0 yes no no no no no unfurnished **543** 1750000 2910 no no no no no furnished

no

no

544 1750000 3850

3

1

2

yes

In [7]: df.shape

(545, 13)

In [8]: df.isnull()

ut[8]:		price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus	
	0	False	False	False	False	False	False	False	False	False	False	False	False	False	
	1	False	False	False	False	False	False	False	False	False	False	False	False	False	
	2	False	False	False	False	False	False	False	False	False	False	False	False	False	
	3	False	False	False	False	False	False	False	False	False	False	False	False	False	
	4	False	False	False	False	False	False	False	False	False	False	False	False	False	
!	540	False	False	False	False	False	False	False	False	False	False	False	False	False	
į	541	False	False	False	False	False	False	False	False	False	False	False	False	False	
į	542	False	False	False	False	False	False	False	False	False	False	False	False	False	
į	543	False	False	False	False	False	False	False	False	False	False	False	False	False	
	544		False	False	False	False	False	False	False	False	False	False	False	False	

545 rows × 13 columns

df.duplicated()

False Out[9]: False False False False

540 False 541 False 542 False 543 False False 544

Length: 545, dtype: bool

df.drop duplicates()

In [11]:	<pre>ar.arop_duplicates()</pre>													
Out[11]:		price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
,	0	13300000	7420	4	2	3	yes	no	no	no	yes	2	yes	furnished
	1	12250000	8960	4	4	4	yes	no	no	no	yes	3	no	furnished
	2	12250000	9960	3	2	2	yes	no	yes	no	no	2	yes	semi-furnished
	3	12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	furnished
	4	11410000	7420	4	1	2	yes	yes	yes	no	yes	2	no	furnished
	540	1820000	3000	2	1	1	yes	no	yes	no	no	2	no	unfurnished
	541	1767150	2400	3	1	1	no	no	no	no	no	0	no	semi-furnished
	542	1750000	3620	2	1	1	yes	no	no	no	no	0	no	unfurnished
	543	1750000	2910	3	1	1	no	no	no	no	no	0	no	furnished
	544	1750000	3850	3	1	2	yes	no	no	no	no	0	no	unfurnished

545 rows × 13 columns

df.hist() In [12]:

array([[<AxesSubplot:title={'center':'price'}>, Out[12]: <AxesSubplot:title={'center':'area'}>], [<AxesSubplot:title={'center':'bedrooms'}>, <AxesSubplot:title={'center':'bathrooms'}>], [<AxesSubplot:title={'center':'stories'}>,

<AxesSubplot:title={'center':'parking'}>]], dtype=object) price area 200 -100 100 500 bath # 6000 15000 0.**b**edroom**s**.0 200 200 storie⁄s 2parking3 200 -200 -100

In [13]: df.hist("price")

array([[<AxesSubplot:title={'center':'price'}>]], dtype=object)

