Assessment - 1

Team No. : 4

Meeting Recording Link : https://drive.google.com/drive/folders/1WQ9chdUOzaHrdje6Jc-On5k8PQ9vDpjy?usp=sharing (https://drive.google.com/drive/folders/1WQ9chdUOzaHrdje6Jc-On5k8PQ9vDpjy?usp=sharing)

In [67]:
```
#importing the libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [68]:
```
#loading the crx.data dataset
data=pd.read_csv('data.csv')
```

In [69]:
```
data
```

Out[69]:

	property type(apartment, villa)	property sale(new/resale)	Place	sq ft area	NoOfBedroom	Facing Direction	price	Furnishing Status
0	Apartment	New	Nilgiris	931.0	2	East	6500000	Semi
1	Residential Villa	Resale	Nilgiris	5800.0	4	?	50000000	Semi
2	Residential Villa	New	Nilgiris	600.0	1	South-East	7200000	Unfurnished
3	Residential Villa	New	Pune	2455.0	4	East	29500000	Unfurnished
4	Residential Villa	Resale	Shimla	1100.0	7	East	20000000	Furnished
5	Independent House	Resale	Hubli	7400.0	8	?	25000000	Furnished
6	Apartment	Resale	Bangalore	2400.0	3	East	10000000	Furnished
7	Farm House	Resale	Chennai	6500.0	7	?	70000000	?
8	Villa	Resale	Nilgiris	6969.6	2	North	19000000	Furnished
9	Apartment	Resale	Pune	1311.0	2	West	7500000	Semi
10	Independent House	New	Mysore	13793.0	12	North- west	35000000	Furnished

In [70]:
```
#Task 1 : Replacing ? with NaN
data=data.replace('?',np.nan)
data
```

Out[70]:

	property type(apartment, villa)	property sale(new/resale)	Place	sq ft area	NoOfBedroom	Facing Direction	price	Furnishing Status
0	Apartment	New	Nilgiris	931.0	2	East	6500000	Semi
1	Residential Villa	Resale	Nilgiris	5800.0	4	NaN	50000000	Semi
2	Residential Villa	New	Nilgiris	600.0	1	South-East	7200000	Unfurnished
3	Residential Villa	New	Pune	2455.0	4	East	29500000	Unfurnished
4	Residential Villa	Resale	Shimla	1100.0	7	East	20000000	Furnished
5	Independent House	Resale	Hubli	7400.0	8	NaN	25000000	Furnished
6	Apartment	Resale	Bangalore	2400.0	3	East	10000000	Furnished
7	Farm House	Resale	Chennai	6500.0	7	NaN	70000000	NaN
8	Villa	Resale	Nilgiris	6969.6	2	North	19000000	Furnished
9	Apartment	Resale	Pune	1311.0	2	West	7500000	Semi
10	Independent House	New	Mysore	13793.0	12	North- west	35000000	Furnished
11	Villa	New	Coonoor	1300.0	3	North	9670000	Semi
12	Villa	Resale	Lovedale	600.0	2	East	2500000	Semi
13	Villa	Resale	Yercaude	835.0	2	South	4500000	Unfurnished
14	Villa	Resale	Munnar	5300.0	5	North	80000000	Furnished
15	Villa	New	Kotagiri	1000.0	2	East	8700000	Unfurnished
16	Agricultural Land	New	Vellore	87120.0	NaN	NaN	20000000	NaN
17	Apartment	New	Mumbai	3100.0	4	East	92500000	Unfurnished
18	Apartment	Resale	Ahemdabad	4600.0	3	East	4000000	Unfurnished
19	Villa	Resale	Kolhapur	5000.0	3	NaN	13500000	Furnished

In [71]:
```
#Task 2: Renaming columns to easier notations
data.rename(columns = {'property type(apartment, villa)':'prop_type'}, inplace = True)
data.rename(columns = {'property sale(new/resale)':'prop_sale'}, inplace = True)
data.head()
```

Out[71]:

	prop_type	prop_sale	Place	sq ft area	NoOfBedroom	Facing Direction	price	Furnishing Status
0	Apartment	New	Nilgiris	931.0	2	East	6500000	Semi
1	Residential Villa	Resale	Nilgiris	5800.0	4	NaN	50000000	Semi
2	Residential Villa	New	Nilgiris	600.0	1	South-East	7200000	Unfurnished
3	Residential Villa	New	Pune	2455.0	4	East	29500000	Unfurnished
4	Residential Villa	Resale	Shimla	1100.0	7	East	20000000	Furnished

In [72]:
```
#Task 3: printing unique values of mutiple attributes
print(data['prop_type'].unique())
print('\n')
print(data['prop_sale'].unique())
```

```
['Apartment' 'Residential Villa' 'Independent House' 'Farm House' 'Villa'
 'Apartment ' 'Villa ' 'Agricultural Land']


['New' 'Resale']
```

In [73]:
```
data['NoOfBedroom'] = data['NoOfBedroom'].replace(np.nan, 0)
data
```

Out[73]:

	prop_type	prop_sale	Place	sq ft area	NoOfBedroom	Facing Direction	price	Furnishing Status
0	Apartment	New	Nilgiris	931.0	2	East	6500000	Semi
1	Residential Villa	Resale	Nilgiris	5800.0	4	NaN	50000000	Semi
2	Residential Villa	New	Nilgiris	600.0	1	South-East	7200000	Unfurnished
3	Residential Villa	New	Pune	2455.0	4	East	29500000	Unfurnished
4	Residential Villa	Resale	Shimla	1100.0	7	East	20000000	Furnished
5	Independent House	Resale	Hubli	7400.0	8	NaN	25000000	Furnished
6	Apartment	Resale	Bangalore	2400.0	3	East	10000000	Furnished
7	Farm House	Resale	Chennai	6500.0	7	NaN	70000000	NaN
8	Villa	Resale	Nilgiris	6969.6	2	North	19000000	Furnished
9	Apartment	Resale	Pune	1311.0	2	West	7500000	Semi
10	Independent House	New	Mysore	13793.0	12	North- west	35000000	Furnished
11	Villa	New	Coonoor	1300.0	3	North	9670000	Semi
12	Villa	Resale	Lovedale	600.0	2	East	2500000	Semi
13	Villa	Resale	Yercaude	835.0	2	South	4500000	Unfurnished
14	Villa	Resale	Munnar	5300.0	5	North	80000000	Furnished
15	Villa	New	Kotagiri	1000.0	2	East	8700000	Unfurnished
16	Agricultural Land	New	Vellore	87120.0	0	NaN	20000000	NaN
17	Apartment	New	Mumbai	3100.0	4	East	92500000	Unfurnished
18	Apartment	Resale	Ahemdabad	4600.0	3	East	4000000	Unfurnished
19	Villa	Resale	Kolhapur	5000.0	3	NaN	13500000	Furnished

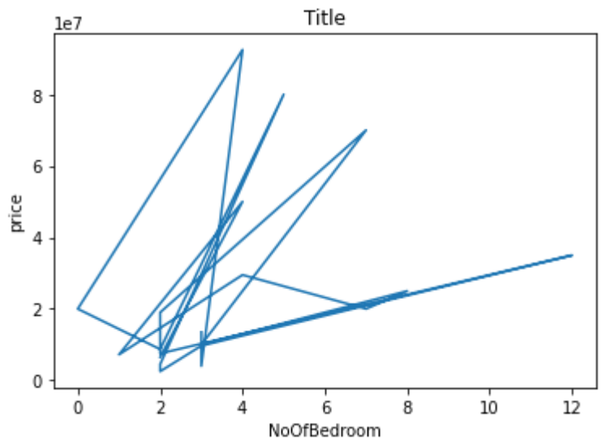
In [74]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   prop_type          20 non-null     object 
 1   prop_sale          20 non-null     object 
 2   Place              20 non-null     object 
 3   sq ft area         20 non-null     float64
 4   NoOfBedroom        20 non-null     object 
 5   Facing Direction   15 non-null     object 
 6   price              20 non-null     int64  
 7   Furnishing Status  18 non-null     object 
dtypes: float64(1), int64(1), object(6)
memory usage: 1.4+ KB
```

In [75]:
```
data['price']=data['price'].astype('double')
data['NoOfBedroom']=data['NoOfBedroom'].astype('double')
```

In [76]:
```
#Tak4: Plotting graph betweeen 'NoOfBedroom' and 'price'
X = (data['NoOfBedroom'])
Y = (data['price'])
plt.xlabel('NoOfBedroom')
plt.ylabel('price')
plt.plot(X,Y)
plt.title('NoOfBedroom vs. Price')
```

Out[76]: Text(0.5, 1.0, 'Title')



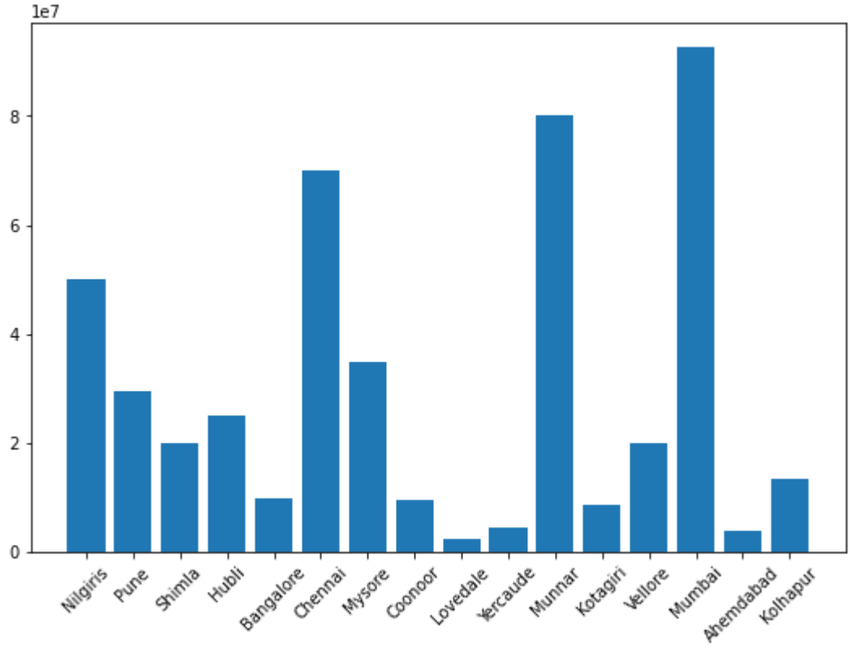
In [ ]:
```

```

In [86]:
```
#Tak 5: Plotting  bar graph betweeen 'Place' and 'price'
groupedby_places = data.groupby('Place')[['price']].mean()
print(groupedby_places.sort_values(by = ['price']))
plt.figure(figsize=(9,6))
plt.bar(x=data['Place'],height=data['price'])
plt.xticks(rotation=45)
```

```
                price
Place
Lovedale    2500000.0
Ahemdabad   4000000.0
Yercaude    4500000.0
Kotagiri    8700000.0
Coonoor     9670000.0
Bangalore  10000000.0
Kolhapur   13500000.0
Pune       18500000.0
Shimla     20000000.0
Vellore    20000000.0
Nilgiris   20675000.0
Hubli      25000000.0
Mysore     35000000.0
Chennai    70000000.0
Munnar     80000000.0
Mumbai     92500000.0
```

Out[86]:
```
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15],
 [Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, '')])
```


1e7
8
6
4
2
0
Nilgiris
Pune
Shimla
Hubli
Bangalore
Chennai
Mysore
Coonoor
Lovedale
Yercaude
Munnar
Kotagiri
Vellore
Mumbai
Ahemdabad
Kolhapur

In [ ]: