

```
In [1]: import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
```

```
In [2]: df_raw = pd.read_csv(r"D:\Datasets\Mumbai House Prices.csv")
```

```
In [3]: df_raw.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 76038 entries, 0 to 76037
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype  
---  -
 0   bhk         76038 non-null  int64  
 1   type        76038 non-null  object  
 2   locality    76038 non-null  object  
 3   area        76038 non-null  int64  
 4   price       76038 non-null  float64 
 5   price_unit  76038 non-null  object  
 6   region      76038 non-null  object  
 7   status      76038 non-null  object  
 8   age         76038 non-null  object  
dtypes: float64(1), int64(2), object(6)
memory usage: 5.2+ MB
```

```
In [4]: df_raw.describe()
```

```
Out[4]:
```

| | bhk | area | price |
|-------|--------------|--------------|--------------|
| count | 76038.000000 | 76038.000000 | 76038.000000 |
| mean | 2.015111 | 1024.536850 | 29.38227 |
| std | 0.922754 | 670.276165 | 32.90345 |
| min | 1.000000 | 127.000000 | 1.00000 |
| 25% | 1.000000 | 640.000000 | 1.75000 |
| 50% | 2.000000 | 872.000000 | 5.50000 |
| 75% | 3.000000 | 1179.000000 | 59.00000 |
| max | 10.000000 | 16000.000000 | 99.99000 |

```
In [5]: df_raw.head()
```

| Out[5]: | bhk | type | locality | area | price | price_unit | region | status | age |
|---------|-----|-----------|---|------|-------|------------|----------------------|-----------------------|-----|
| 0 | 3 | Apartment | Lak And Hanware The Residency Tower | 685 | 2.50 | Cr | Andheri West | Ready to move | New |
| 1 | 2 | Apartment | Radheya Sai Enclave Building No 2 | 640 | 52.51 | L | Naigaon East | Under Construction | New |
| 2 | 2 | Apartment | Romell Serene | 610 | 1.73 | Cr | Borivali West | Under Construction | New |
| 3 | 2 | Apartment | Soundlines Codename Urban Rainforest | 876 | 59.98 | L | Panvel | Under Construction | New |
| 4 | 2 | Apartment | Origin Oriana | 659 | 94.11 | L | Mira Road East | Under Construction | New |

Multivariate Analysis

```
In [7]: df_bhk_info = df_raw["bhk"].value_counts().head(3)
display(df_bhk_info)
```

```
bhk
2    31429
1    24613
3    15273
Name: count, dtype: int64
```

```
In [8]: x = df_bhk_info[::].to_list()
print(x)
```

```
[31429, 24613, 15273]
```

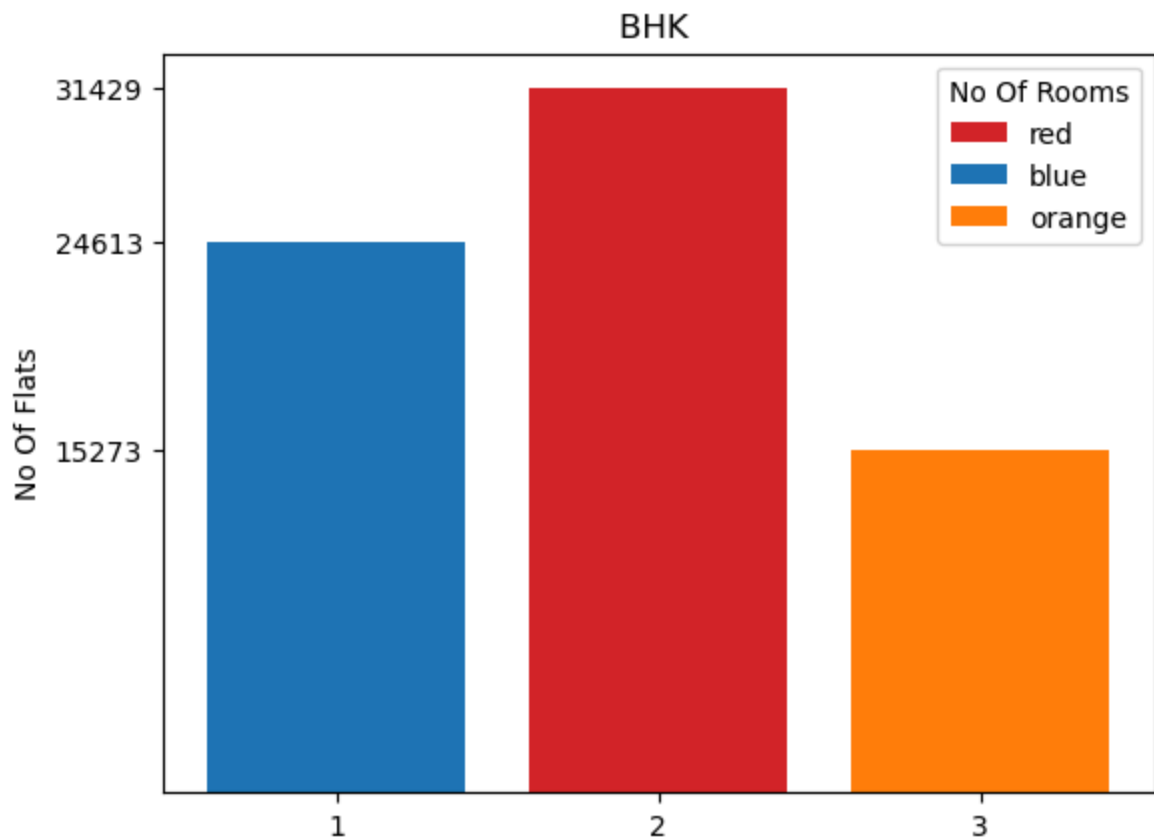
```
In [9]: fig, ax = plt.subplots()

bar_labels = ['red', 'blue', 'orange']
bar_colors = ['tab:red', 'tab:blue', 'tab:orange']

ax.bar(df_bhk_info.index, df_bhk_info.values, label=bar_labels, color=bar_colors)
ax.set_yticks(df_bhk_info.values)
ax.set_xticks(df_bhk_info.index)

ax.set_ylabel('No Of Flats')
ax.set_title('BHK ')
ax.legend(title='No Of Rooms')

plt.show()
```



The above graph represents the no of available flats for top 3 No of BHK Flats

Count of No of Residences According to their type

```
In [12]: df_raw["type"].value_counts()
```

```
Out[12]: type
Apartment          74854
Studio Apartment    882
Villa              226
Independent House   73
Penthouse           3
Name: count, dtype: int64
```

Top three Residences Accordong to Sqft area of house

```
In [14]: df_raw[df_raw["area"] == df_raw["area"].max()]
```

Out[14]:

| | bhk | type | locality | area | price | price_unit | region | status | age |
|--------------|-----|-----------|--------------------------|-------|-------|------------|----------------------|---------------|---------|
| 60864 | 10 | Apartment | JP Chalet Amar | 16000 | 50.0 | Cr | Juhu | Ready to move | Unknown |
| 68538 | 10 | Apartment | Indiabulls Blu Tower D | 16000 | 40.0 | Cr | Worli | Ready to move | Unknown |
| 68539 | 10 | Apartment | Sunteck Signature Island | 16000 | 50.0 | Cr | Bandra Kurla Complex | Ready to move | Unknown |

Details of the Apartment with least Sqft Area

In [16]: `df_raw[df_raw["area"] == df_raw["area"].min()]`

Out[16]:

| | bhk | type | locality | area | price | price_unit | region | status | age |
|--------------|-----|-----------|--------------------|------|-------|------------|----------------|--------------------|-----|
| 20784 | 1 | Apartment | JP North Barcelona | 127 | 59.0 | L | Mira Road East | Under Construction | New |
| 23438 | 1 | Apartment | JP North Barcelona | 127 | 59.0 | L | Mira Road East | Under Construction | New |
| 24503 | 1 | Apartment | JP North Barcelona | 127 | 59.0 | L | Mira Road East | Under Construction | New |
| 29816 | 1 | Apartment | JP North Barcelona | 127 | 72.0 | L | Mira Road East | Under Construction | New |
| 42185 | 1 | Apartment | JP North Barcelona | 127 | 59.0 | L | Mira Road East | Under Construction | New |

We can see here that Apartments has the least sqft area according to house and area of house is 127sqft.

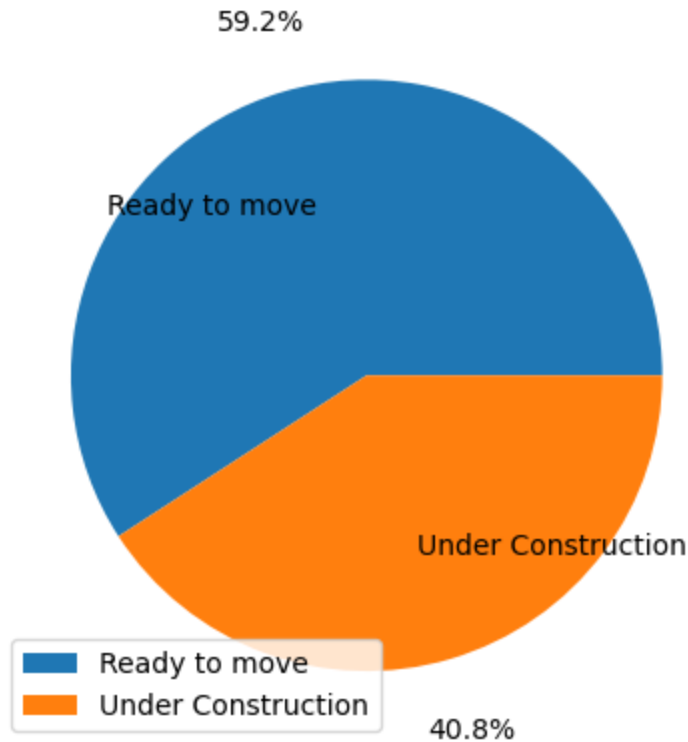
In [18]: `df_info_status = df_raw['status'].value_counts()`

In [19]: `fig, ax = plt.subplots()`

```
# bar_labels = ['red', 'blue', 'orange']
# bar_colors = ['tab:red', 'tab:blue', 'tab:orange']
```

```
ax.pie(df_info_status.values, labels = df_info_status.index, autopct='%1.1f%%', pctdis
# ax.set_yticks(df_bhk_info.values)
# ax.set_xticks(df_bhk_info.index)

# ax.set_ylabel('No Of Flats')
# ax.set_title('BHK ')
ax.legend()
plt.show()
```



```
In [20]: df_raw[df_raw["price_unit"] == "Cr" ].max()
```

```
Out[20]: bhk          10
type          Villa
locality      Zueb Aayesha Palace
area          16000
price          60.0
price_unit     Cr
region        vile parle west
status        Under Construction
age           Unknown
dtype: object
```

Details of Flat with max price when price unit is in Crores

```
In [22]: df_raw[df_raw["price_unit"] == "L"].max()
```

```
Out[22]: bhk                    5
         type                  Villa
         locality      Zinal Angel Residency
         area                    5000
         price                99.99
         price_unit          L
         region                mumbai
         status      Under Construction
         age                  Unknown
         dtype: object
```

Details of flat with maximum price when price unit is in lakhs

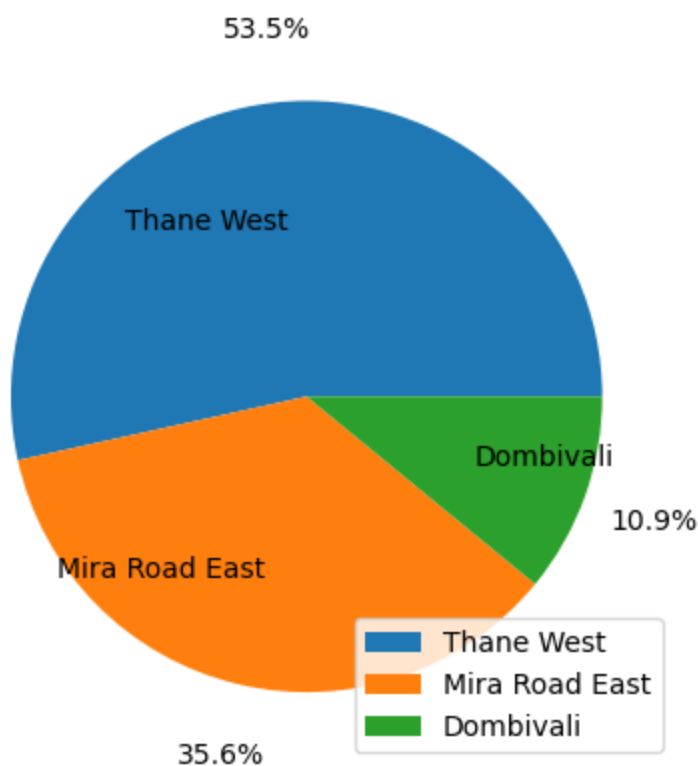
```
In [24]: df_region_info = df_raw["region"].value_counts().head(3)
```

```
In [25]: fig, ax = plt.subplots()

# bar_labels = ['red', 'blue', 'orange']
# bar_colors = ['tab:red', 'tab:blue', 'tab:orange']

ax.pie(df_region_info.values, labels = df_region_info.index, autopct='%1.1f%%', pctdis
# ax.set_yticks(df_bhk_info.values)
# ax.set_xticks(df_bhk_info.index)

# ax.set_ylabel('No Of Flats')
# ax.set_title('BHK ')
ax.legend()
plt.show()
```



Maximum no of Flats are in Thane west

```
In [27]: df_raw.head()
```

```
Out[27]:
```

| | bhk | type | locality | area | price | price_unit | region | status | age |
|---|-----|-----------|---|------|-------|------------|----------------------|-----------------------|-----|
| 0 | 3 | Apartment | Lak And Hanware The Residency Tower | 685 | 2.50 | Cr | Andheri West | Ready to move | New |
| 1 | 2 | Apartment | Radheya Sai Enclave Building No 2 | 640 | 52.51 | L | Naigaon East | Under Construction | New |
| 2 | 2 | Apartment | Romell Serene | 610 | 1.73 | Cr | Borivali West | Under Construction | New |
| 3 | 2 | Apartment | Soundlines Codename Urban Rainforest | 876 | 59.98 | L | Panvel | Under Construction | New |
| 4 | 2 | Apartment | Origin Oriana | 659 | 94.11 | L | Mira Road East | Under Construction | New |

```
In [28]: df_raw[(df_raw["region"] == "Thane West") & (df_raw["price_unit"] == "Cr")].max()
```

```
Out[28]: bhk                6
type                Villa
locality      Yash The Vivek Flat Owners CHS Ltd
area                8000
price              19.0
price_unit         Cr
region            Thane West
status            Under Construction
age                Unknown
dtype: object
```

Details of residence and the maximum price of residence is 19.0 Cr in Thane West which has the largest no of Residences

```
In [30]: df_raw[(df_raw["region"] == "Thane West") & (df_raw["price_unit"] == "L")].min()
```

```
Out[30]: bhk          1
         type          Apartment
         locality      Aakar Manas Residency
         area          183
         price          10.99
         price_unit      L
         region          Thane West
         status          Ready to move
         age            New
         dtype: object
```

Details of residence and the minimum price of residence is 10.99 Lakhs in Thane West which has the largest no of Residences

```
In [32]: list_of_regions = set(df_raw["region"].to_list())
         final_list = list(set([i.upper() for i in list_of_regions]))
         len(final_list)
```

```
Out[32]: 228
```

There are a total of 228 regions in Mumbai Locations with available residences

```
In [34]: # for i in list_of_regions:
         #     display(df_raw[(df_raw["region"] == i)])
```

Details of the residences according to all 228 regions can be found after executing the above cell

```
In [36]: df_info_age = df_raw["age"].value_counts()
```

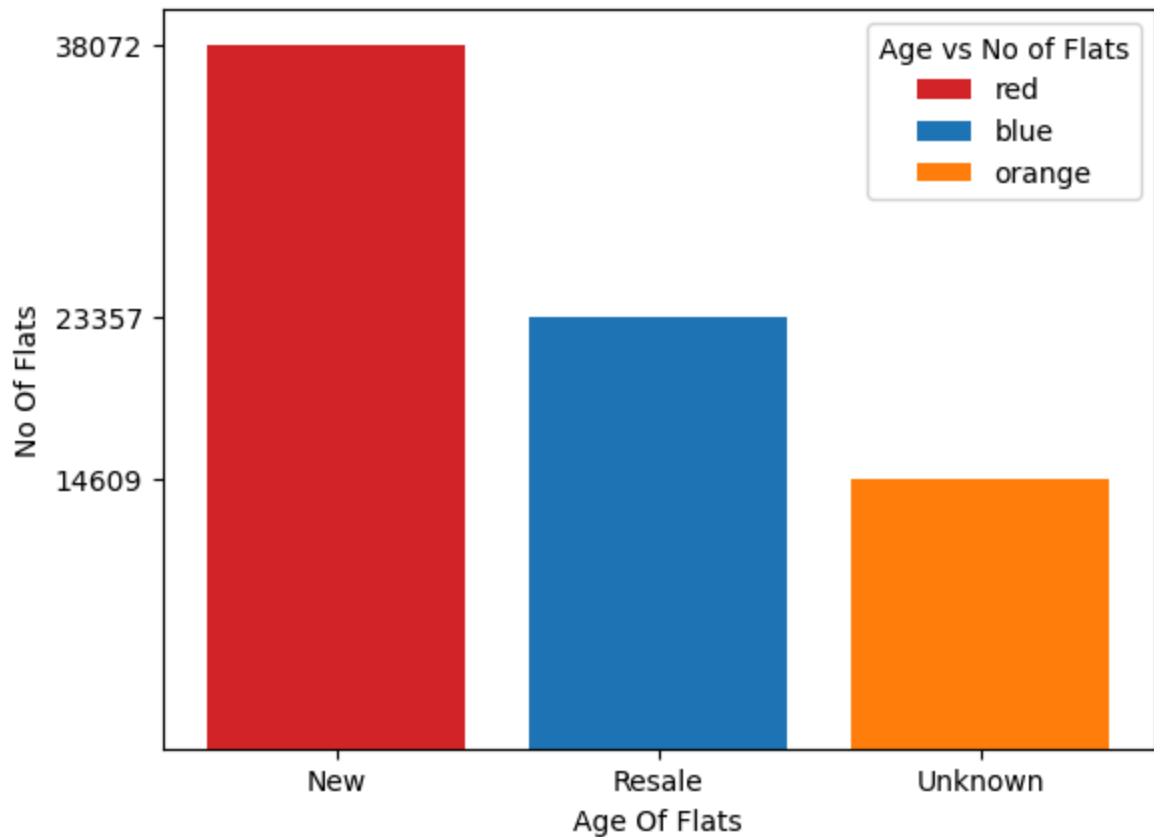
```
In [37]: fig, ax = plt.subplots()

         bar_labels = ['red', 'blue', 'orange']
         bar_colors = ['tab:red', 'tab:blue', 'tab:orange']

         ax.bar(df_info_age.index, df_info_age.values, label=bar_labels, color=bar_colors)
         ax.set_yticks(df_info_age.values)
         ax.set_xticks(df_info_age.index)

         ax.set_ylabel('No Of Flats')
         ax.set_xlabel('Age Of Flats')
         # ax.set_title('Age of Flats ')
         ax.legend(title='Age vs No of Flats')

         plt.show()
```

```
In [38]: df_raw["price_unit"].value_counts()
```

```
Out[38]: price_unit
Cr      40981
L       35057
Name: count, dtype: int64
```

We can also say that there are 35,057 flats with maximum price upto 1 Crore and 40981 flats with price more than 1 Crore

```
In [40]: df_raw.head()
```

| Out[40]: | bhk | type | locality | area | price | price_unit | region | status | age |
|----------|-----|-----------|---|------|-------|------------|----------------------|-----------------------|-----|
| 0 | 3 | Apartment | Lak And Hanware The Residency Tower | 685 | 2.50 | Cr | Andheri West | Ready to move | New |
| 1 | 2 | Apartment | Radheya Sai Enclave Building No 2 | 640 | 52.51 | L | Naigaon East | Under Construction | New |
| 2 | 2 | Apartment | Romell Serene | 610 | 1.73 | Cr | Borivali West | Under Construction | New |
| 3 | 2 | Apartment | Soundlines Codename Urban Rainforest | 876 | 59.98 | L | Panvel | Under Construction | New |
| 4 | 2 | Apartment | Origin Oriana | 659 | 94.11 | L | Mira Road East | Under Construction | New |

```
In [41]: len(df_raw["locality"].to_list())
```

Out[41]: 76038

```
In [42]: list_locality = [i.upper() for i in df_raw["locality"].to_list()]
```

```
In [43]: list_locality_final = list(set(list_locality))
len(list_locality_final)
```

Out[43]: 9778

After careful calibration it has been found that there are in total 9778 localities with available Residences

```
In [45]: df_locality = pd.DataFrame(list_locality_final).to_csv("Locality_List.csv")
```

Since the list is too big and cant be viewed and analysed in form of DataFrame very easily we can also convert it to excel and csv format for less technical analysis but more precise or can also be used for presentation to clients

```
In [47]: list_res_type = list(set(df_raw["type"].to_list()))
print(list_res_type)
```

```
['Studio Apartment', 'Apartment', 'Penthouse', 'Villa', 'Independent House']
```

```
In [48]: df_raw[[df_raw["type"] == "Apartment"] and df_raw["price_unit"] == "Cr"].max()
```

```
Out[48]: bhk                10  
         type                Villa  
         locality      Zueb Aayesha Palace  
         area                16000  
         price                60.0  
         price_unit          Cr  
         region      vile parle west  
         status      Under Construction  
         age                Unknown  
         dtype: object
```

```
In [49]: for i in list_res_type:  
         print(df_raw[df_raw["type"] == i].max())  
         print("\n\n")
```

| | |
|------------|--------------------|
| bhk | 2 |
| type | Studio Apartment |
| locality | Yashvi Pride |
| area | 2637 |
| price | 99.0 |
| price_unit | L |
| region | Worli |
| status | Under Construction |
| age | Unknown |

dtype: object

| | |
|------------|---------------------|
| bhk | 10 |
| type | Apartment |
| locality | Zoeb Aayesha Palace |
| area | 16000 |
| price | 99.99 |
| price_unit | L |
| region | vile parle west |
| status | Under Construction |
| age | Unknown |

dtype: object

| | |
|------------|----------------------------|
| bhk | 4 |
| type | Penthouse |
| locality | Runwal Forest Tower 5 To 8 |
| area | 2500 |
| price | 3.9 |
| price_unit | Cr |
| region | Koper Khairane |
| status | Ready to move |
| age | Unknown |

dtype: object

| | |
|------------|---|
| bhk | 7 |
| type | Villa |
| locality | Wadhwa Wise City South Block Phase I Plot RZ8 ... |
| area | 12400 |
| price | 90.0 |
| price_unit | L |
| region | Worli |
| status | Under Construction |
| age | Resale |

dtype: object

| | |
|----------|---------------------------|
| bhk | 6 |
| type | Independent House |
| locality | Wadhwa Magnolia Cluster 3 |
| area | 7500 |

```
price                62.0
price_unit           L
region              sector
status              Under Construction
age                 Resale
dtype: object
```

Above find the details of the most expensive residences according to their type

```
In [51]: list_price_unit = list(set(df_raw["price_unit"].to_list()))
```

```
In [52]: for i in list_price_unit:
          print(i)
          display(df_raw[df_raw["price_unit"]== i].min())
```

```
Cr
bhk                1
type              Apartment
locality          Arcade
area              288
price             1.0
price_unit        Cr
region            Agripada
status            Ready to move
age               New
dtype: object
```

```
L
bhk                1
type              Apartment
locality          AJ Shri Jay
area              127
price             4.49
price_unit        L
region            Adaigaon
status            Ready to move
age               New
dtype: object
```

```
In [53]: for i in list_price_unit:
          print(i)
          display(df_raw[df_raw["price_unit"]== i].max())
          print("\n")
```

```
Cr
```

```

bhk          10
type          Villa
locality      Zueb Aayesha Palace
area          16000
price         60.0
price_unit      Cr
region        vile parle west
status        Under Construction
age           Unknown
dtype: object

```

```

L
bhk          5
type          Villa
locality      Zinal Angel Residency
area          5000
price        99.99
price_unit      L
region        mumbai
status        Under Construction
age           Unknown
dtype: object

```

```
In [54]: df_raw.head()
```

```
Out[54]:
```

| | bhk | type | locality | area | price | price_unit | region | status | age |
|---|-----|-----------|---|------|-------|------------|----------------------|-----------------------|-----|
| 0 | 3 | Apartment | Lak And Hanware The Residency Tower | 685 | 2.50 | Cr | Andheri West | Ready to move | New |
| 1 | 2 | Apartment | Radheya Sai Enclave Building No 2 | 640 | 52.51 | L | Naigaon East | Under Construction | New |
| 2 | 2 | Apartment | Romell Serene | 610 | 1.73 | Cr | Borivali West | Under Construction | New |
| 3 | 2 | Apartment | Soundlines Codename Urban Rainforest | 876 | 59.98 | L | Panvel | Under Construction | New |
| 4 | 2 | Apartment | Origin Oriana | 659 | 94.11 | L | Mira Road East | Under Construction | New |

```
In [55]: df_ready_to_move = df_raw.where(df_raw["status"] == "Ready to move")
```

```
In [56]: df_rtm_final = df_ready_to_move.dropna()
```

Details of the cheapest ready to move Apartment

```
In [130... df_rtm_lakhs = df_rtm_final[df_rtm_final["price_unit"] == "L"].copy()
```

```
In [148... for i in list_res_type:  
    print(df_rtm_lakhs[df_rtm_lakhs["type"] == i].min())  
    print("\n")
```

| | |
|------------|------------------|
| bhk | 1.0 |
| type | Studio Apartment |
| locality | Akruti |
| area | 149.0 |
| price | 5.5 |
| price_unit | L |
| region | Airoli |
| status | Ready to move |
| age | New |

dtype: object

| | |
|------------|---------------|
| bhk | 1.0 |
| type | Apartment |
| locality | AJ Shri Jay |
| area | 150.0 |
| price | 4.49 |
| price_unit | L |
| region | Adaigaon |
| status | Ready to move |
| age | New |

dtype: object

| | |
|------------|-----|
| bhk | NaN |
| type | NaN |
| locality | NaN |
| area | NaN |
| price | NaN |
| price_unit | NaN |
| region | NaN |
| status | NaN |
| age | NaN |

dtype: object

| | |
|------------|---------------------------|
| bhk | 1.0 |
| type | Villa |
| locality | Assets The Chalets Avenue |
| area | 560.0 |
| price | 30.0 |
| price_unit | L |
| region | Badlapur West |
| status | Ready to move |
| age | New |

dtype: object

| | |
|------------|-------------------------|
| bhk | 1.0 |
| type | Independent House |
| locality | Bhoomi Homes Naksashtra |
| area | 400.0 |
| price | 20.0 |
| price_unit | L |
| region | Bhandup East |
| status | Ready to move |

age
dtype: object

New

Details of the cheapest residences according to their type which are in Ready to Move Status

```
In [161... df_rtm_cr = df_rtm_final[df_rtm_final["price_unit"] == "Cr"].copy()
```

```
In [163... for i in list_res_type:  
    print(i)  
    print(df_rtm_cr[df_rtm_cr["type"] == i].max())  
    print("\n")
```

Studio Apartment
bhk 1.0
type Studio Apartment
locality Swaraj Homes Kalpita Enclave
area 2637.0
price 6.46
price_unit Cr
region Worli
status Ready to move
age Resale
dtype: object

Apartment
bhk 10.0
type Apartment
locality Zueb Aayesha Palace
area 16000.0
price 60.0
price_unit Cr
region vile parle west
status Ready to move
age Unknown
dtype: object

Penthouse
bhk 4.0
type Penthouse
locality Runwal Forest Tower 5 To 8
area 2500.0
price 3.9
price_unit Cr
region Koper Khairane
status Ready to move
age Unknown
dtype: object

Villa
bhk 7.0
type Villa
locality Vasant Vasant Vihar
area 12400.0
price 51.0
price_unit Cr
region Worli
status Ready to move
age Resale
dtype: object

Independent House
bhk 6.0
type Independent House
locality Wadhwa Juhu Anmol Chs

```
area                7500.0
price               32.0
price_unit          Cr
region              sector
status              Ready to move
age                 Resale
dtype: object
```

Details of the most expensive residences according to their type which are in Ready to Move Status

```
In [173...] df_under_con = df_raw.where(df_raw["status"] == "Under Construction")
```

```
In [177...] df_under_con_final = df_under_con.dropna()
```

```
In [187...] df_uc_lakhs = df_under_con_final[df_under_con_final["price_unit"] == "L"].copy()
```

```
In [189...] for i in list_res_type:
    print(df_uc_lakhs[df_uc_lakhs["type"] == i].min())
    print("\n")
```

| | |
|------------|----------------------|
| bhk | 1.0 |
| type | Studio Apartment |
| locality | A Plus Golden Corner |
| area | 139.0 |
| price | 9.67 |
| price_unit | L |
| region | Ambernath East |
| status | Under Construction |
| age | New |

dtype: object

| | |
|------------|-----------------------------|
| bhk | 1.0 |
| type | Apartment |
| locality | Green Woods Building 2 To 6 |
| area | 127.0 |
| price | 8.5 |
| price_unit | L |
| region | Agripada |
| status | Under Construction |
| age | New |

dtype: object

| | |
|------------|-----|
| bhk | NaN |
| type | NaN |
| locality | NaN |
| area | NaN |
| price | NaN |
| price_unit | NaN |
| region | NaN |
| status | NaN |
| age | NaN |

dtype: object

| | |
|------------|--|
| bhk | 1.0 |
| type | Villa |
| locality | Aditya Creations The Ocean Homes The Pearl |
| area | 800.0 |
| price | 37.0 |
| price_unit | L |
| region | Naigaon East |
| status | Under Construction |
| age | New |

dtype: object

| | |
|------------|--------------------|
| bhk | 1.0 |
| type | Independent House |
| locality | Dosti Nest Phase 1 |
| area | 400.0 |
| price | 27.0 |
| price_unit | L |
| region | Ambernath West |
| status | Under Construction |

age
dtype: object

New

Details of the cheapest residences according to their type which are in Under Construction Status.

Please note that there are no Penthouses available in this category

```
In [185... df_uc_cr = df_under_con_final[df_under_con_final["price_unit"] == "Cr"].copy()
```

```
In [193... for i in list_res_type:  
    print(df_uc_cr[df_uc_cr["type"] == i].max())  
    print("\n")
```

| | |
|------------|-------------------------|
| bhk | 1.0 |
| type | Studio Apartment |
| locality | Piramal Revanta Tower 1 |
| area | 733.0 |
| price | 1.3 |
| price_unit | Cr |
| region | Mulund West |
| status | Under Construction |
| age | New |

dtype: object

| | |
|------------|--------------------|
| bhk | 8.0 |
| type | Apartment |
| locality | Zee Sahyadri |
| area | 10893.0 |
| price | 55.55 |
| price_unit | Cr |
| region | Worli |
| status | Under Construction |
| age | Unknown |

dtype: object

| | |
|------------|-----|
| bhk | NaN |
| type | NaN |
| locality | NaN |
| area | NaN |
| price | NaN |
| price_unit | NaN |
| region | NaN |
| status | NaN |
| age | NaN |

dtype: object

| | |
|------------|---|
| bhk | 6.0 |
| type | Villa |
| locality | Wadhwa Wise City South Block Phase I Plot RZ8 ... |
| area | 8000.0 |
| price | 15.0 |
| price_unit | Cr |
| region | Virar |
| status | Under Construction |
| age | Resale |

dtype: object

| | |
|------------|---------------------------|
| bhk | 5.0 |
| type | Independent House |
| locality | Wadhwa Magnolia Cluster 3 |
| area | 1756.0 |
| price | 2.93 |
| price_unit | Cr |
| region | Panvel |
| status | Under Construction |

age
dtype: object

New

Details of the most expensive residences according to their type which are in Under Construction Status.

Please note that there are no Penthouses available in this category

In []: