

In [34]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

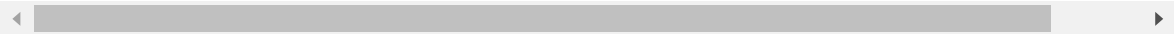
In [2]:

```
df=pd.read_csv("bengaluru_house_prices.csv")
df
```

Out[2]:

	area_type	availability	location	size	society	total_sqft	bath	balcony
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0
...	...	...	...	...	...	...	...	...
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	NaN	3600	5.0	NaN
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	1.0
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	NaN	550	1.0	1.0

13320 rows × 9 columns



In [3]:

```
df.describe
```

Out[3]:

```
<bound method NDFrame.describe of
ty
0      Super built-up Area      19-Dec Electronic City Phase II
1              Plot Area Ready To Move      Chikka Tirupathi
2      Built-up Area Ready To Move      Uttarahalli
3      Super built-up Area Ready To Move      Lingadheeranahalli
4      Super built-up Area Ready To Move      Kothanur
...
13315      Built-up Area Ready To Move      Whitefield
13316 Super built-up Area Ready To Move      Richards Town
13317      Built-up Area Ready To Move      Raja Rajeshwari Nagar
13318 Super built-up Area      18-Jun      Padmanabhanagar
13319 Super built-up Area Ready To Move      Doddathoguru

      size society total_sqft  bath  balcony  price
0      2 BHK Coomee      1056  2.0      1.0  39.07
1      4 Bedroom Theanmp      2600  5.0      3.0 120.00
2      3 BHK      NaN      1440  2.0      3.0  62.00
3      3BHK Soiewre      1521  3.0      1.0  95.00
4      2 BHK      NaN      1200  2.0      1.0  51.00
...
13315  5 Bedroom ArsiaEx      3453  4.0      0.0 231.00
13316  4 BHK      NaN      3600  5.0      NaN 400.00
13317  2 BHK Mahla T      1141  2.0      1.0  60.00
13318  4 BHK SollyCl      4689  4.0      1.0 488.00
13319  1 BHK      NaN      550  1.0      1.0  17.00

[13320 rows x 9 columns]>
```

In [4]:

```
df.rename(columns={'area_type':'Area_Type','availability':'Availability','location': 'L
ocation', 'size': 'Size', 'total_sqft':'Total_sqft','bath':'Bathrooms','balcony':'Balco
ny','price':'Price'},inplace=True)
```

In [5]:

```
df
```

Out[5]:

	Area_Type	Availability	Location	Size	society	Total_sqft	Bathrooms	
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	
...	...	...	...	...	...	...	...	
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	NaN	3600	5.0	
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	NaN	550	1.0	

13320 rows × 9 columns



In [6]:

```
df.Area_Type.isna().sum()
```

Out[6]:

0

In [7]:

```
df.Availability.isna().sum()
```

Out[7]:

0

In [8]:

```
df.Location.isna().sum()
```

Out[8]:

1

In [9]:

```
df.Location.mode()[0]
```

Out[9]:

'Whitefield'

In [10]:

```
df.Location=df.Location.fillna(df.Location.mode()[0])
```

In [11]:

```
df.Location.isna().sum()
```

Out[11]:

0

In [12]:

```
df.Size.isna().sum()
```

Out[12]:

16

In [28]:

```
df.Size=df.Size.fillna(df.Size.mode()[0])
```

In [13]:

```
df.society.isna().sum()
```

Out[13]:

5502

In [25]:

```
df.drop('society',axis=1,inplace=True)
```

In [26]:

```
df.Total_sqft.isna().sum()
```

Out[26]:

0

In [16]:

```
df.Bathrooms.isna().sum()
```

Out[16]:

73

In [17]:

```
df.Bathrooms=df.Bathrooms.fillna(df.Bathrooms.mode()[0])
```

In [18]:

```
df.Bathrooms.isna().sum()
```

Out[18]:

0

In [19]:

```
df.Balcony.isna().sum()
```

Out[19]:

609

In [20]:

```
df.Balcony=df.Balcony.fillna(df.Balcony.mode()[0])
```

In [21]:

```
df.Balcony.isna().sum()
```

Out[21]:

0

In [22]:

```
df.Price.isna().sum()
```

Out[22]:

0

In [29]:

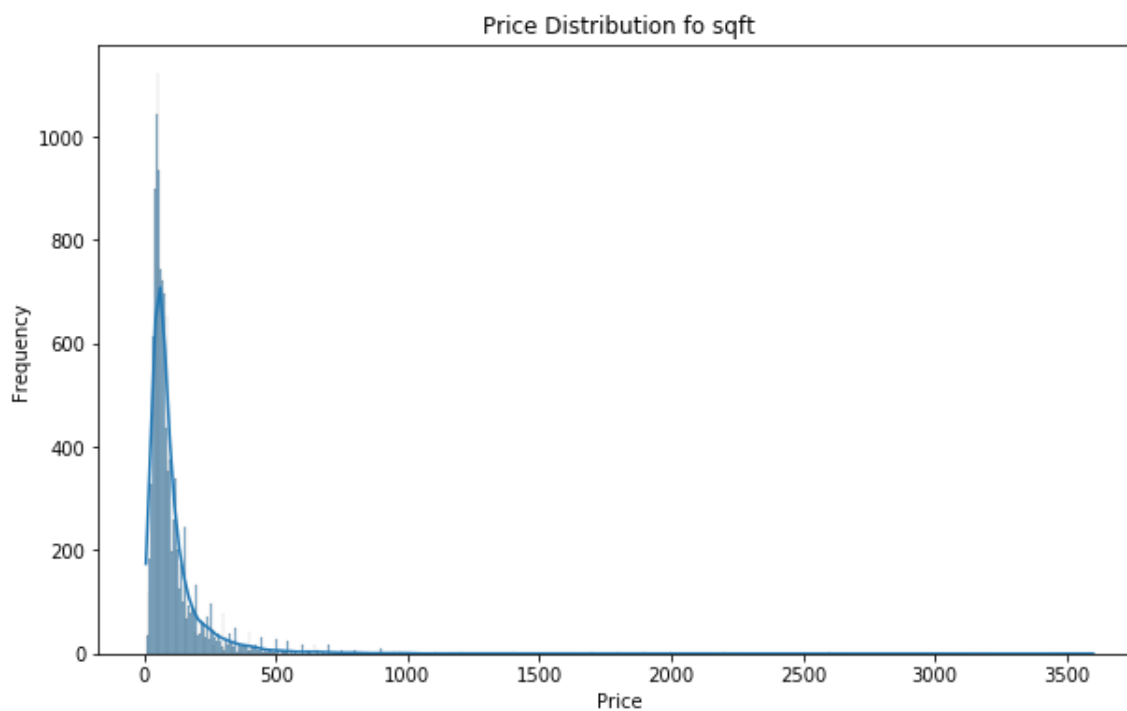
```
df.isna().sum()
```

Out[29]:

```
Area_Type      0
Availability    0
Location       0
Size           0
Total_sqft     0
Bathrooms      0
Balcony        0
Price          0
dtype: int64
```

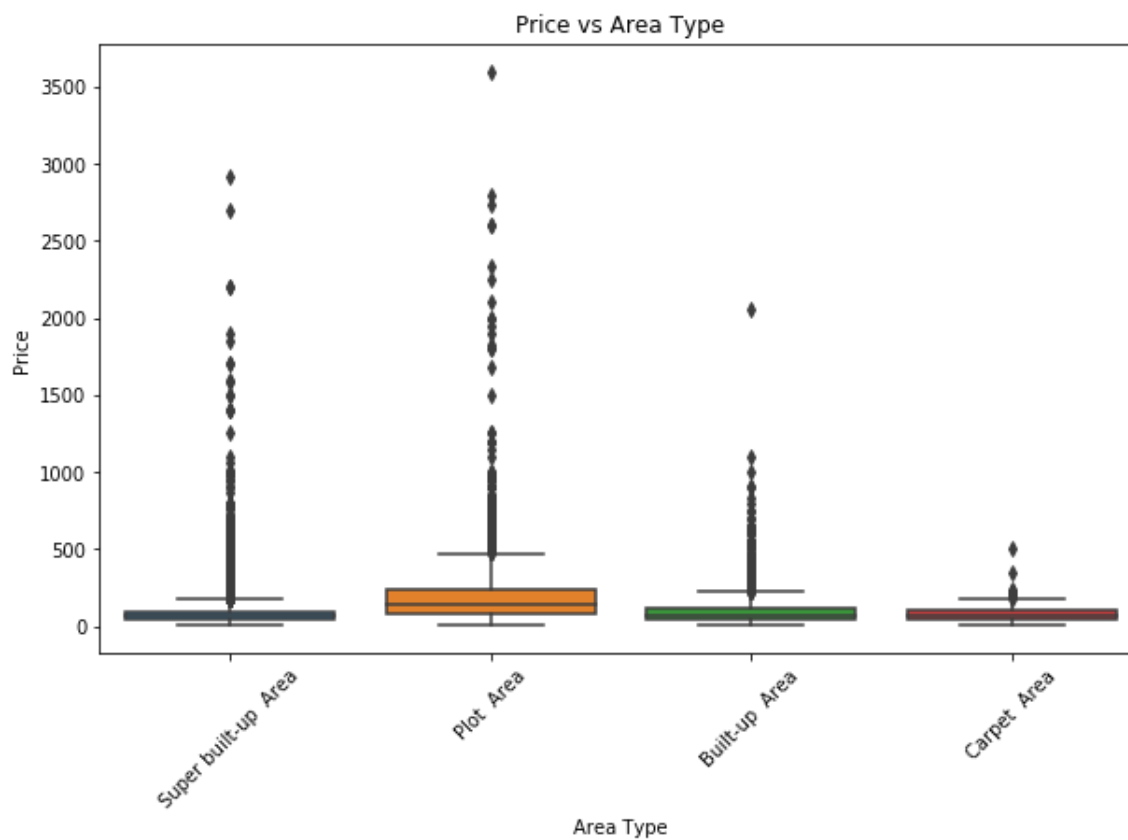
In [46]:

```
plt.figure(figsize=(10, 6))
sns.histplot(df['Price'], kde=True)
plt.xlabel('Price')
plt.ylabel('Frequency')
plt.title('Price Distribution fo sqft')
plt.show()
```



In [41]:

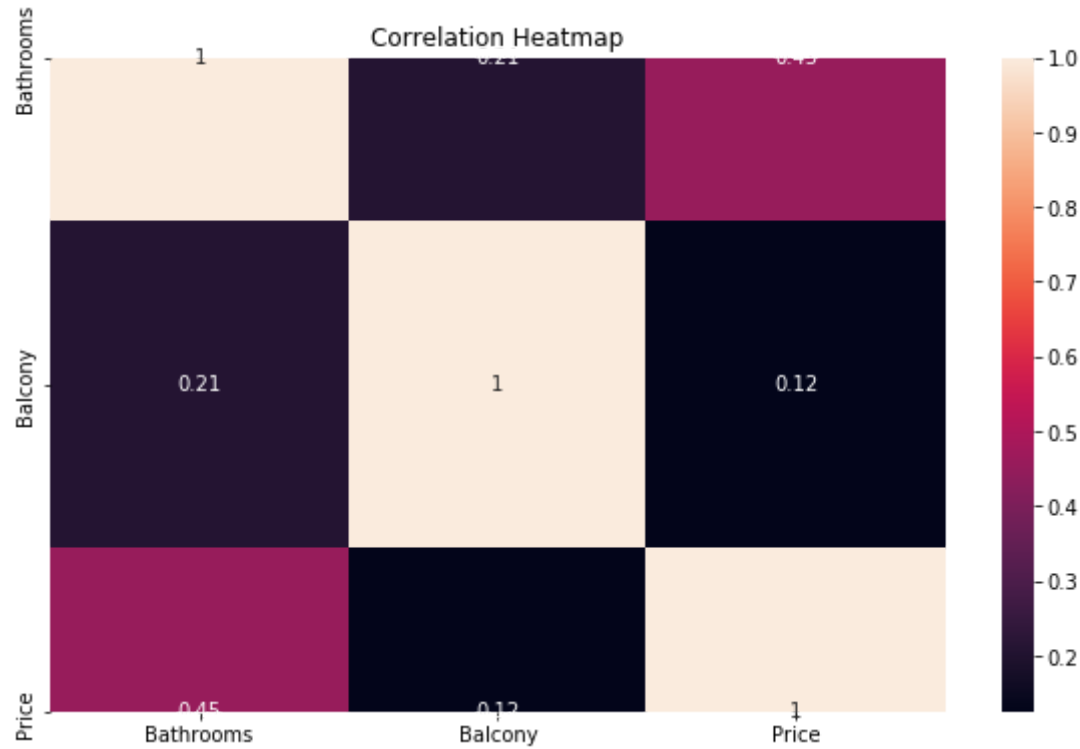
```
plt.figure(figsize=(10, 6))
sns.boxplot(x='Area_Type', y='Price', data=df)
plt.xticks(rotation=45)
plt.xlabel('Area Type')
plt.ylabel('Price')
plt.title('Price vs Area Type')
plt.show()
```



In [43]:

```
corr_matrix = df.corr()

plt.figure(figsize=(10, 6))
sns.heatmap(corr_matrix, annot=True)
plt.title('Correlation Heatmap')
plt.show()
```



In [181]:

```
area_type_grouped = df.groupby('Area_Type').agg({'Price': 'mean'})
x=area_type_grouped.reset_index()
x
```

Out[181]:

	Area_Type	Price
0	Built-up Area	104.285498
1	Carpet Area	89.502356
2	Plot Area	208.495486
3	Super built-up Area	92.971757

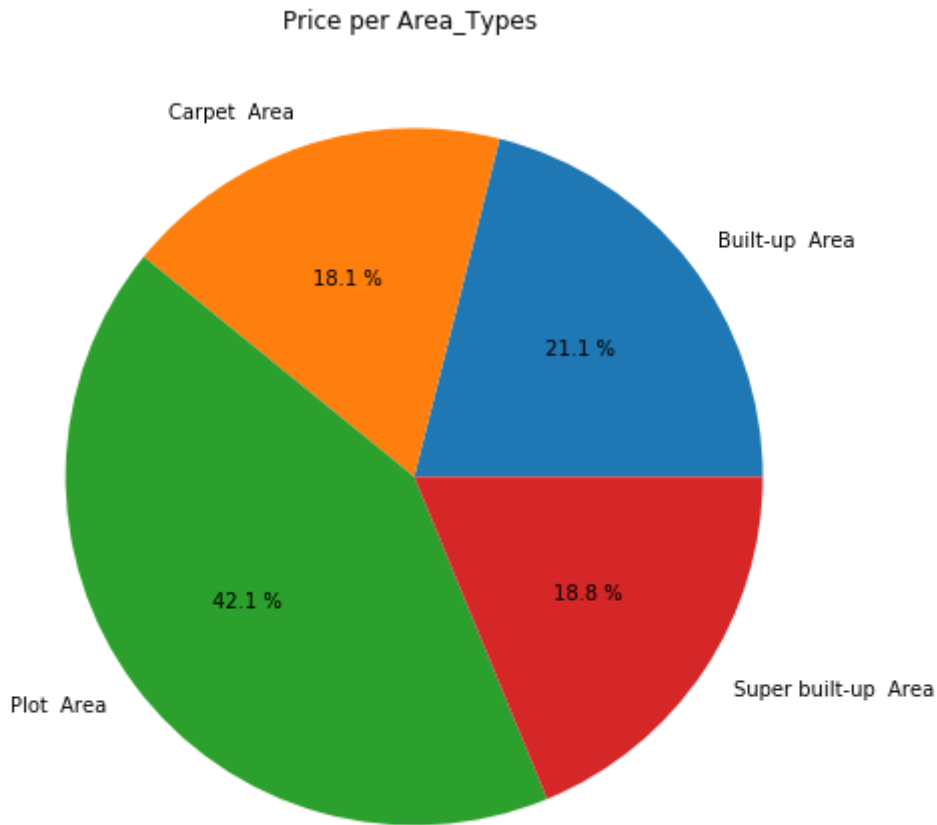


In [187]:

```
plt.figure(figsize=(10, 8))
plt.pie(x.Price, labels=x.Area_Type, autopct='%1f %%')
plt.title('Price per Area_Types ')
```

Out[187]:

Text(0.5, 1.0, 'Price per Area\_Types ')



In [90]:

```
count=df.Location.value_counts()
count
```

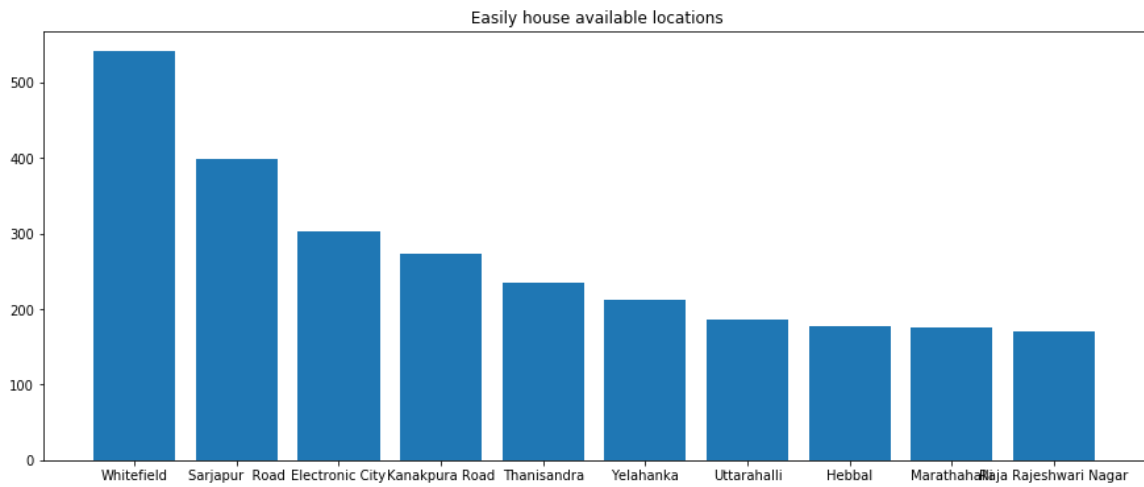
Out[90]:

Whitefield	541
Sarjapur Road	399
Electronic City	302
Kanakpura Road	273
Thanisandra	234
...	
T C Palya main Road	1
Adarsh Nagar	1
Sindhi Colony	1
Anantapuram	1
Jagajyothi layout	1

Name: Location, Length: 1305, dtype: int64

In [156]:

```
plt.figure(figsize=(15, 6))
plt.bar(top_loc.index,top_loc)
plt.title(" Easily house available locations")
plt.show()
```



In [119]:

```
location_counts = df['Location'].value_counts().reset_index()
location_counts.columns = ['Location', 'count']
location_counts
```

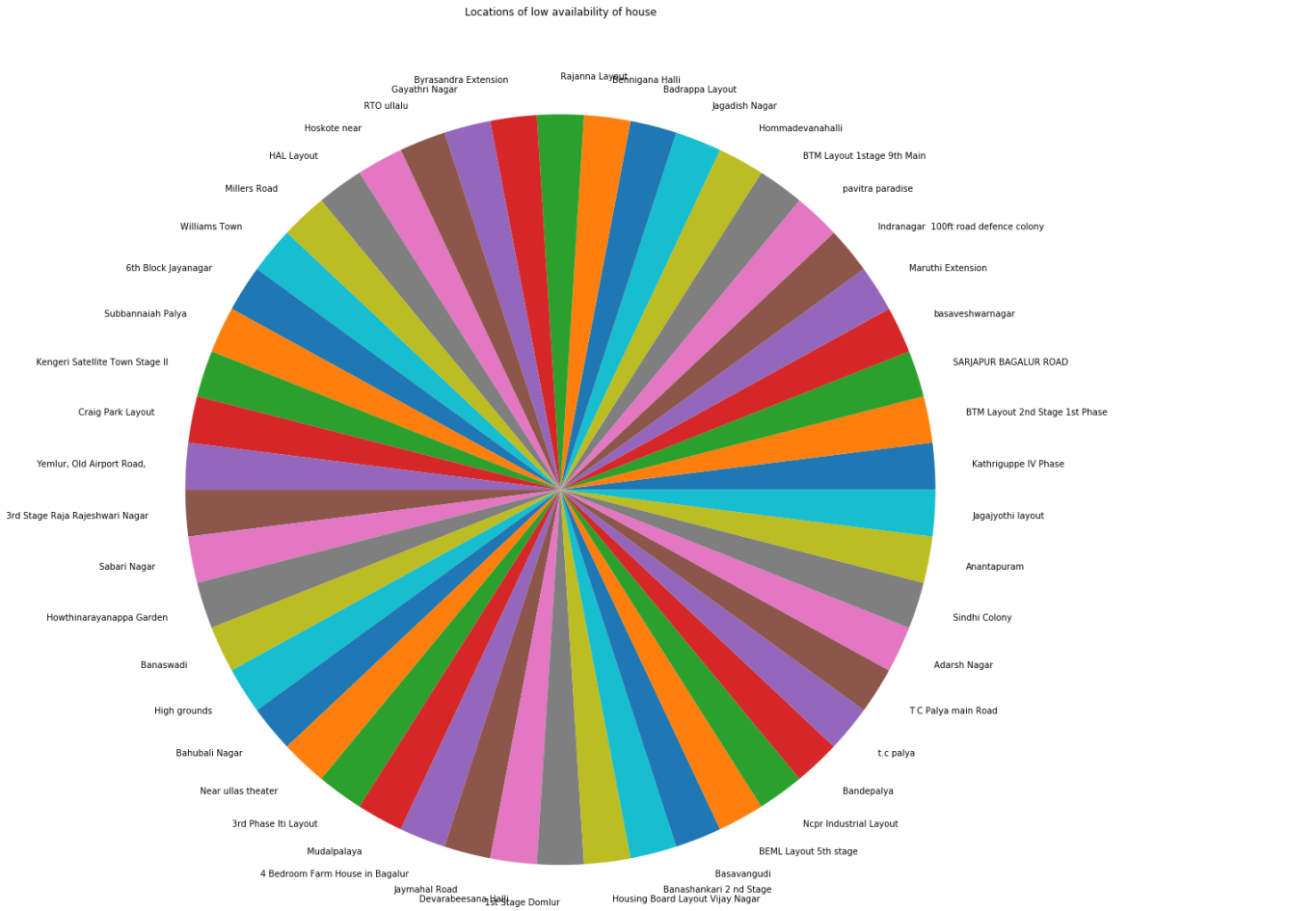
Out[119]:

	Location	count
0	Whitefield	541
1	Sarjapur Road	399
2	Electronic City	302
3	Kanakpura Road	273
4	Thanisandra	234
...	...	...
1300	T C Palya main Road	1
1301	Adarsh Nagar	1
1302	Sindhi Colony	1
1303	Anantapuram	1
1304	Jagajyothi layout	1

1305 rows × 2 columns

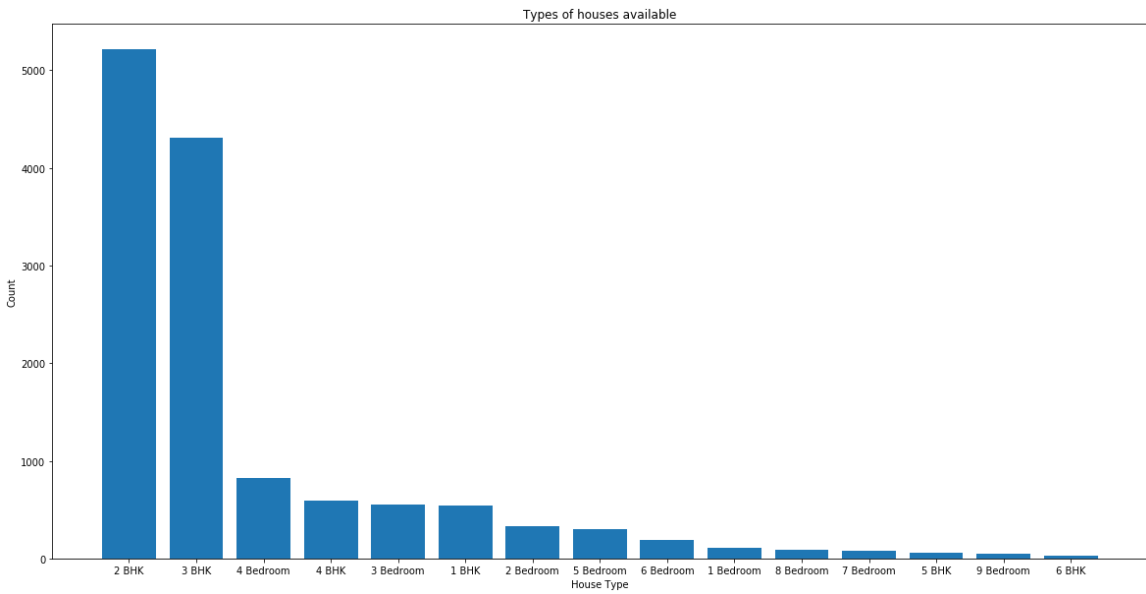
In [157]:

```
plt.figure(figsize=(30, 20))
plt.pie(df.Location.value_counts().tail(50), labels=location_counts.Location.tail(50))
plt.title("Locations of low availability of house")
plt.show()
```



In [153]:

```
plt.figure(figsize=(20, 10))
plt.title('Types of houses available')
plt.xlabel('House Type')
plt.ylabel('Count')
size_count=df.Size.value_counts().reset_index()
size_count.columns=['size','count']
plt.bar(size_count['size'].head(15),size_count['count'].head(15))
plt.show()
```



In [211]:

```
loc_grp=df.groupby('Location').agg({'Price': 'mean'})
x=loc_grp.reset_index().sort_values(by='Price',ascending=False)
x
```

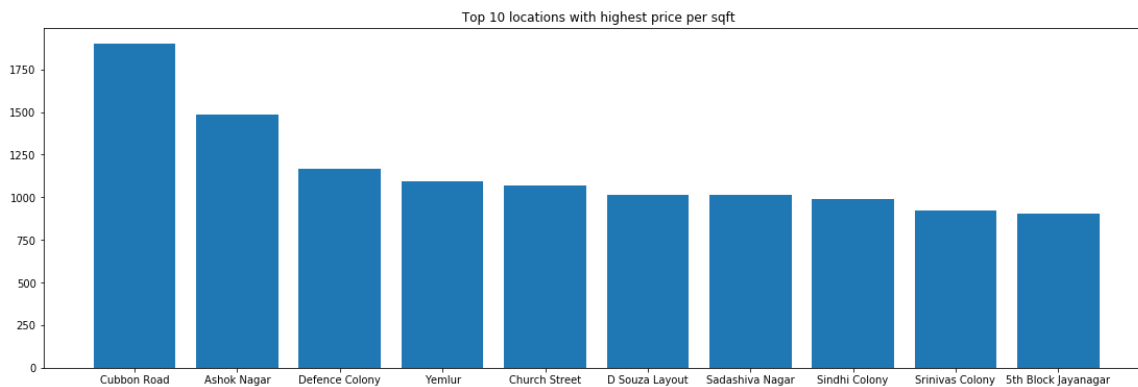
Out[211]:

	Location	Price
351	Cubbon Road	1900.000000
145	Ashok Nagar	1486.000000
362	Defence Colony	1167.714286
1265	Yemlur	1093.388889
339	Church Street	1068.000000
...	...	...
298	Celebrity Paradise Layout	19.245000
162	BAGUR	17.000000
780	Makali	16.000000
0	Anekal	16.000000
105	Alur	15.000000

1305 rows × 2 columns

In [217]:

```
plt.figure(figsize=(19, 6))
high_price = x.head(10)
plt.bar(high_price.Location, high_price.Price)
plt.title(" Top 10 locations with highest price per sqft")
plt.show()
```



In [223]:

```
plt.figure(figsize=(18, 6))
high_price = x.tail(6)
plt.bar(high_price.Location, high_price.Price)
plt.title(" Top 6 locations with lowest price per sqft")
plt.show()
```



In [ ]: