```
import pandas as pd
import numpy as np
import matplotlib as mp
import seaborn as sns
```

```
df = pd.read_csv('House_Rent_Dataset.csv')
df
```

```
      BHK   Rent  Size    Area Type             Area Locality
City  \
0       2  10000  1100   Super Area                   Bandel
Kolkata
1       2  20000   800   Super Area  Phool Bagan, Kankurgachi
Kolkata
2       2  17000  1000   Super Area   Salt Lake City Sector 2
Kolkata
3       2  10000   800   Super Area               Dumdum Park
Kolkata
4       2   7500   850  Carpet Area             South Dum Dum
Kolkata
...   ...    ...   ...          ...                       ...        .
..
4741    2  15000  1000  Carpet Area              Bandam Kommu
Hyderabad
4742    3  29000  2000   Super Area      Manikonda, Hyderabad
Hyderabad
4743    3  35000  1750  Carpet Area      Himayath Nagar, NH 7
Hyderabad
4744    3  45000  1500  Carpet Area               Gachibowli
Hyderabad
4745    2  15000  1000  Carpet Area           Suchitra Circle
Hyderabad

     Furnishing Status  Bathroom Point of Contact
0          Unfurnished         2    Contact Owner
1       Semi-Furnished         1    Contact Owner
2       Semi-Furnished         1    Contact Owner
3          Unfurnished         1    Contact Owner
4          Unfurnished         1    Contact Owner
...                ...       ...              ...
4741    Semi-Furnished         2    Contact Owner
4742    Semi-Furnished         3    Contact Owner
4743    Semi-Furnished         3    Contact Agent
4744    Semi-Furnished         2    Contact Agent
4745       Unfurnished         2    Contact Owner

[4746 rows x 9 columns]
```

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

```
BHK                 0
Rent                0
Size                0
Area Type           0
Area Locality       0
City                0
Furnishing Status   0
Bathroom            0
Point of Contact    0
dtype: int64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4746 entries, 0 to 4745
Data columns (total 9 columns):
 #   Column             Non-Null Count  Dtype 
---  ------             --------------  ----- 
 0   BHK                4746 non-null   int64 
 1   Rent               4746 non-null   int64 
 2   Size               4746 non-null   int64 
 3   Area Type          4746 non-null   object
 4   Area Locality      4746 non-null   object
 5   City               4746 non-null   object
 6   Furnishing Status  4746 non-null   object
 7   Bathroom           4746 non-null   int64 
 8   Point of Contact   4746 non-null   object
dtypes: int64(4), object(5)
memory usage: 333.8+ KB
```

```
'''
REGPLOT
    - Rent
    - Size

BOXPLOT
    - BHK
    - Area Type
    - City
    - Furnishing Status
    - Bathroom
    - Point of Contact
'''
```

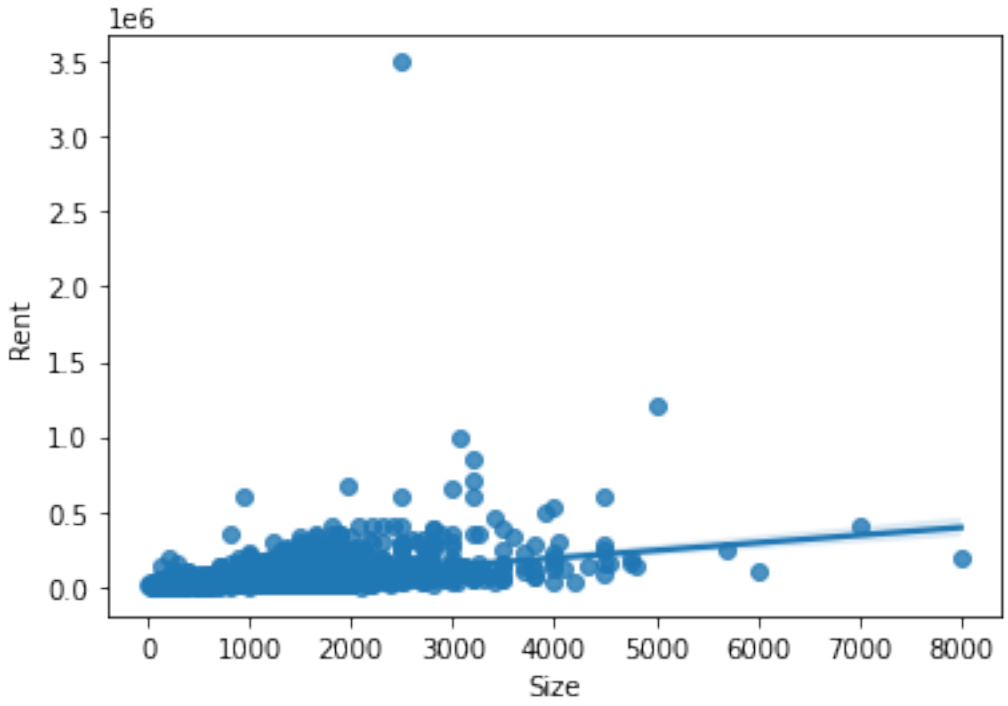
```
'\nREGPLOT\n    - Rent\n    - Size\n\nBOXPLOT\n    - BHK\n    - Area
Type\n    - City\n    - Furnishing Status\n    - Bathroom\n    - Point
of Contact\n'
```

```
# REGPLOT
```

```
sns.regplot(x="Size",y="Rent", data=df)

# From the output of the regplot between 'Size' & 'Rent'.
# It can be clearly seen that Size of the house directly effects the
rent.
# As Size increases Rent also increases gradually.
```

```
<AxesSubplot:xlabel='Size', ylabel='Rent'>
```

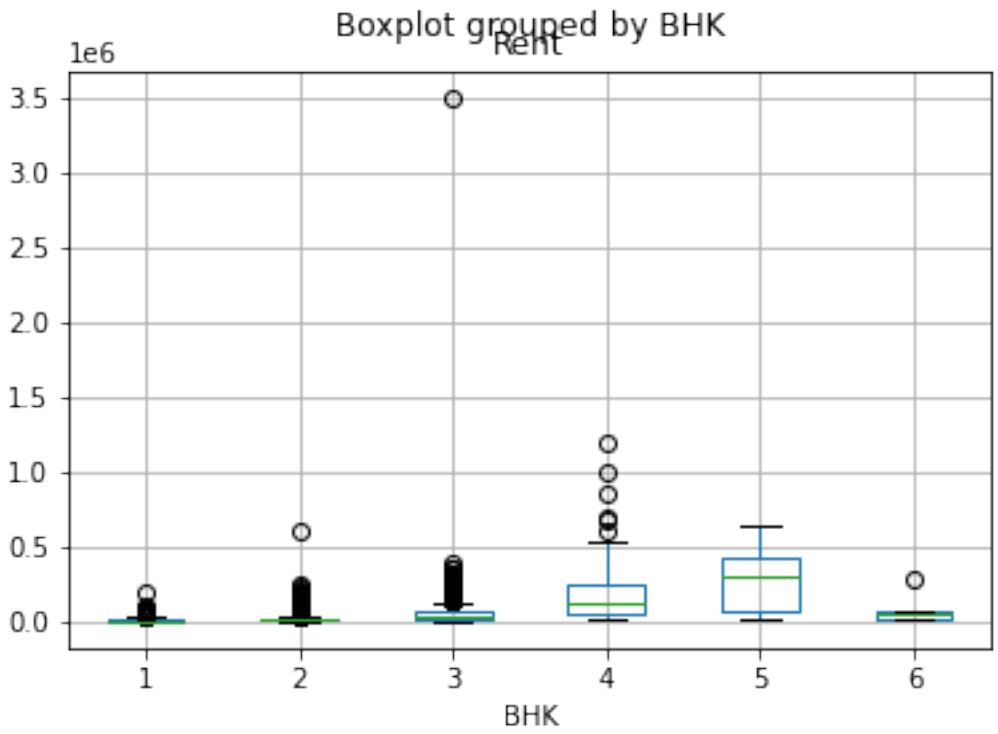



```
# BOXPLOT

df.boxplot(by ='BHK', column =['Rent'])

# 'BHK' is important for the higher values such as 4 & 5 BHK.
```

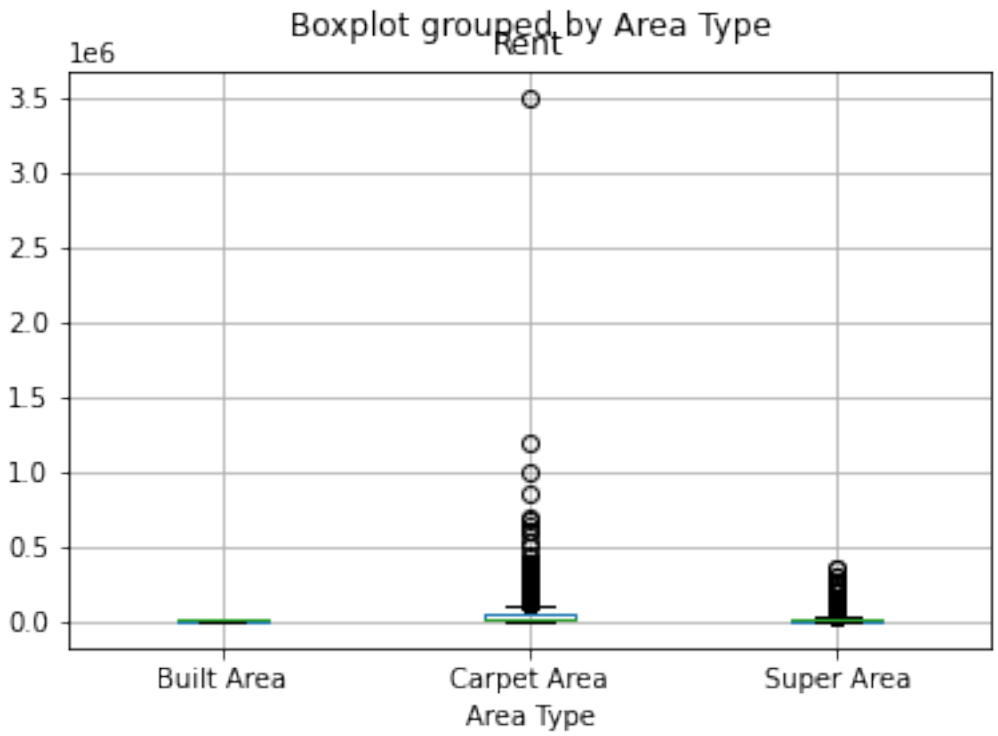
```
<AxesSubplot:title={'center':'Rent'}, xlabel='BHK'>
```




```
df.boxplot(by ='Area Type', column =['Rent'])

# This attribute is important in increase of 'Rent'.
```

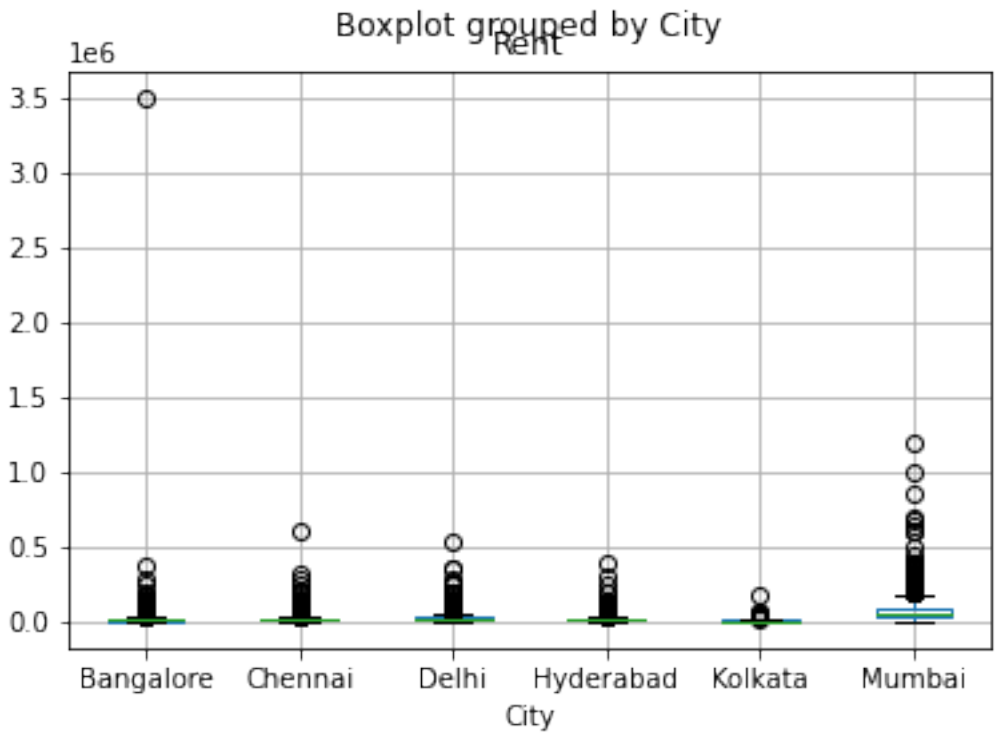
```
<AxesSubplot:title={'center':'Rent'}, xlabel='Area Type'>
```




```
df.boxplot(by ='City', column =['Rent'])

# It can be clearly seen that 'Rent' for house in Mumbai are expensive
as compared to rest all cities.
# Box of 'Mumbai' is not overlapping with rest all cities.
```

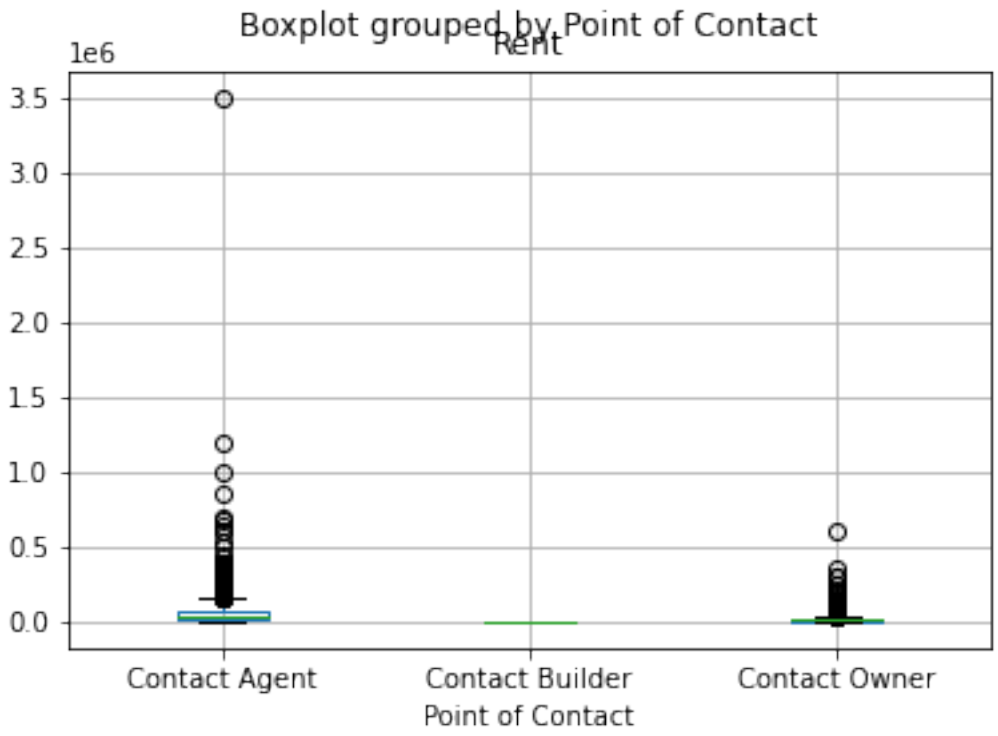
```
<AxesSubplot:title={'center':'Rent'}, xlabel='City'>
```




```
df.boxplot(by ='Point of Contact', column =['Rent'])

# Not important.
```

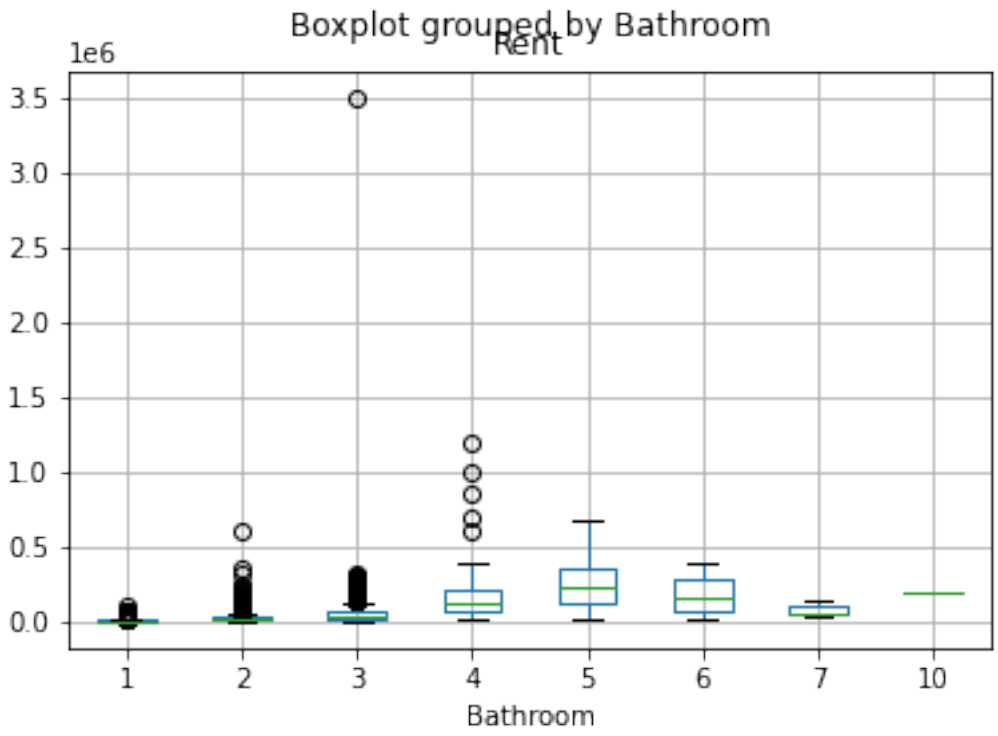
```
<AxesSubplot:title={'center':'Rent'}, xlabel='Point of Contact'>
```




```
df.boxplot(by ='Bathroom', column =['Rent'])

# Not important.
```

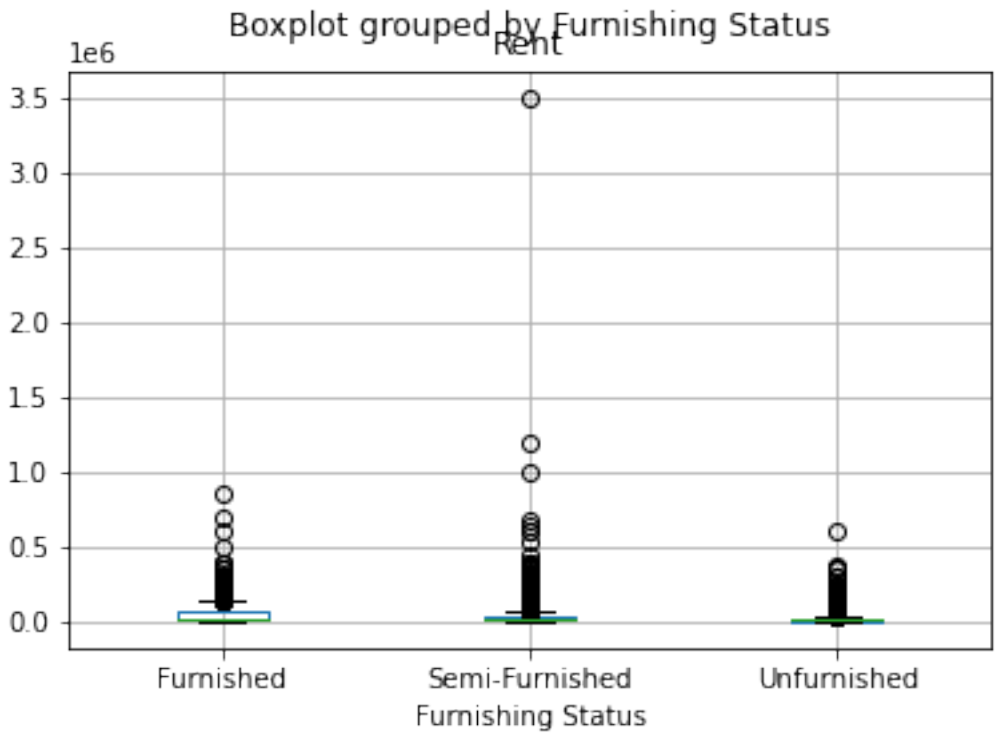
```
<AxesSubplot:title={'center':'Rent'}, xlabel='Bathroom'>
```




```
df.boxplot(by ='Furnishing Status', column =['Rent'])

# Rent of furnished house will be more than rest all others.
```

```
<AxesSubplot:title={'center':'Rent'}, xlabel='Furnishing Status'>
```




```
'''
Conclusion:-

From the whole above anlysis it can be concluded that 'Size' & 'City'
are the most important attribute which directly relates to 'Rent' of
the House.

'''
```

```
"\nConclusion:-\n\nFrom the whole above anlysis it can be concluded
that 'Size' & 'City' are the most important attribute which directly
relates to 'Rent' of the House.\n\n"
```