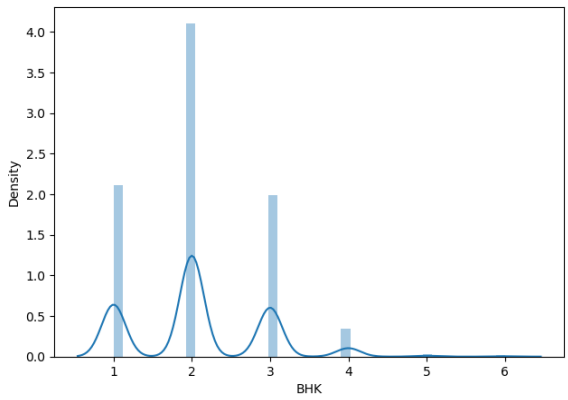
In [63]:
```
#import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [64]:
```
df = pd.read_csv('House_Rent_main2.csv')
df1=df
df1.sample(5)
```

Out[64]:

	Posted On	BHK	Rent	Size	Floor	Area Type	Area Locality	City	Furnishing Status	Te Pref
1968	2022-05-20	1	6500	200	4 out of 4	1	Mahadevapura	3	2	Bach
4387	2022-06-25	2	12000	1000	5 out of 6	1	Upperpally	Hyderabad	2	Bachelors/F
3339	2022-06-23	1	12000	650	1 out of 2	1	Triplicane	5	1	Bachelors/F
1683	2022-05-22	2	40000	1300	2 out of 4	2	Indira Nagar	3	2	Bachelors/F
3461	2022-07-06	2	19000	900	2 out of 3	1	Choolaimedu	5	2	Bachelors/F

In [65]:
```
import warnings
warnings.filterwarnings('ignore')
plt.figure(figsize=(16,5))
plt.subplot(1,2,1)
sns.distplot(df1['BHK'])
plt.subplot(1,2,2)
sns.distplot(df1['Bathroom'])
plt.show()
```

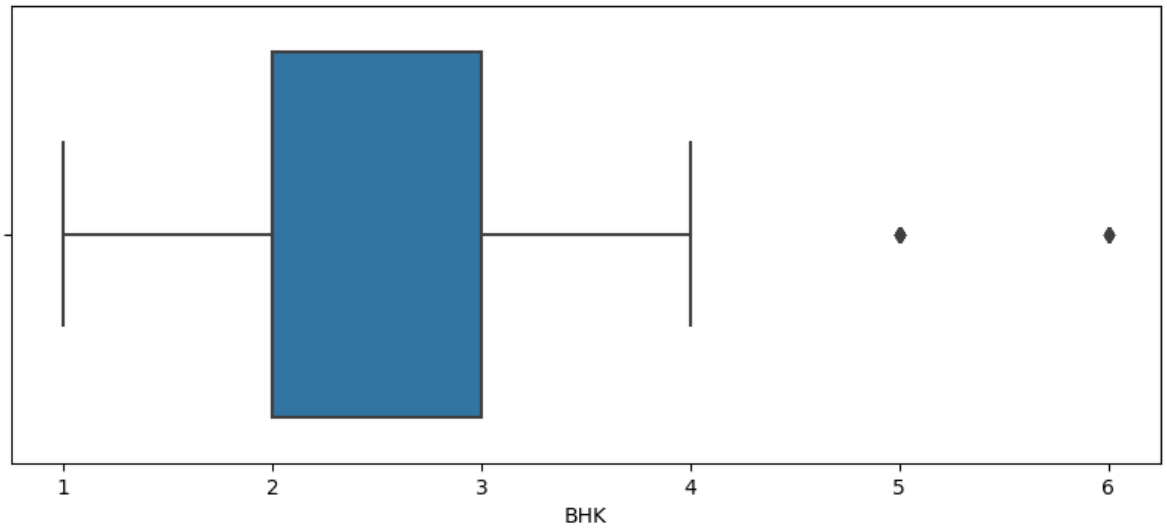


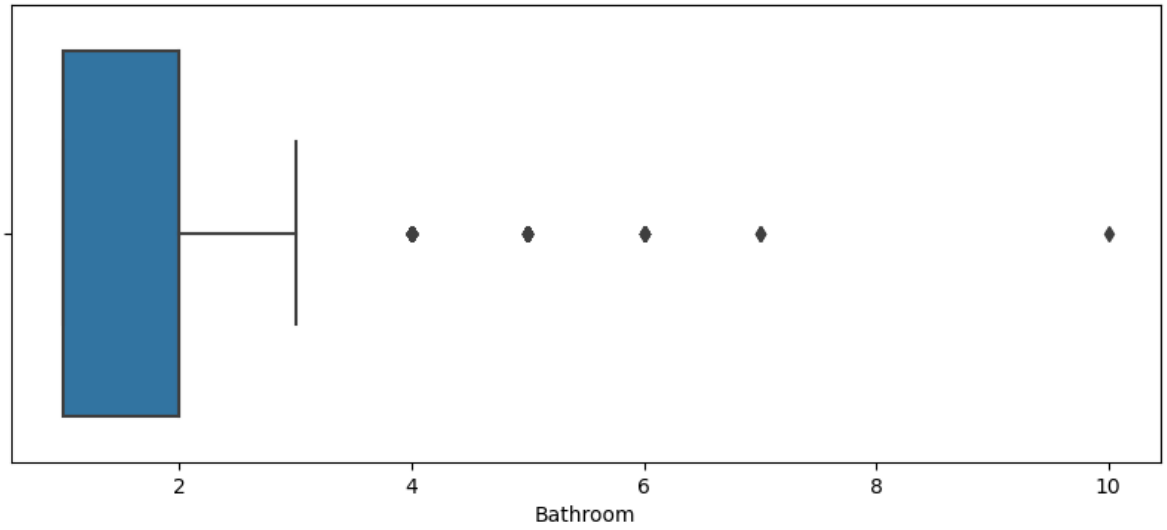
In [66]:
```
#Checking outliers using boxplot
print(plt.figure(figsize = (10, 4), dpi = 100))
sns.boxplot(x = "BHK", data = df1)

print(plt.figure(figsize = (10, 4), dpi = 100))
sns.boxplot(x = "Bathroom", data = df1)
```

```
Figure(1000x400)
Figure(1000x400)
```

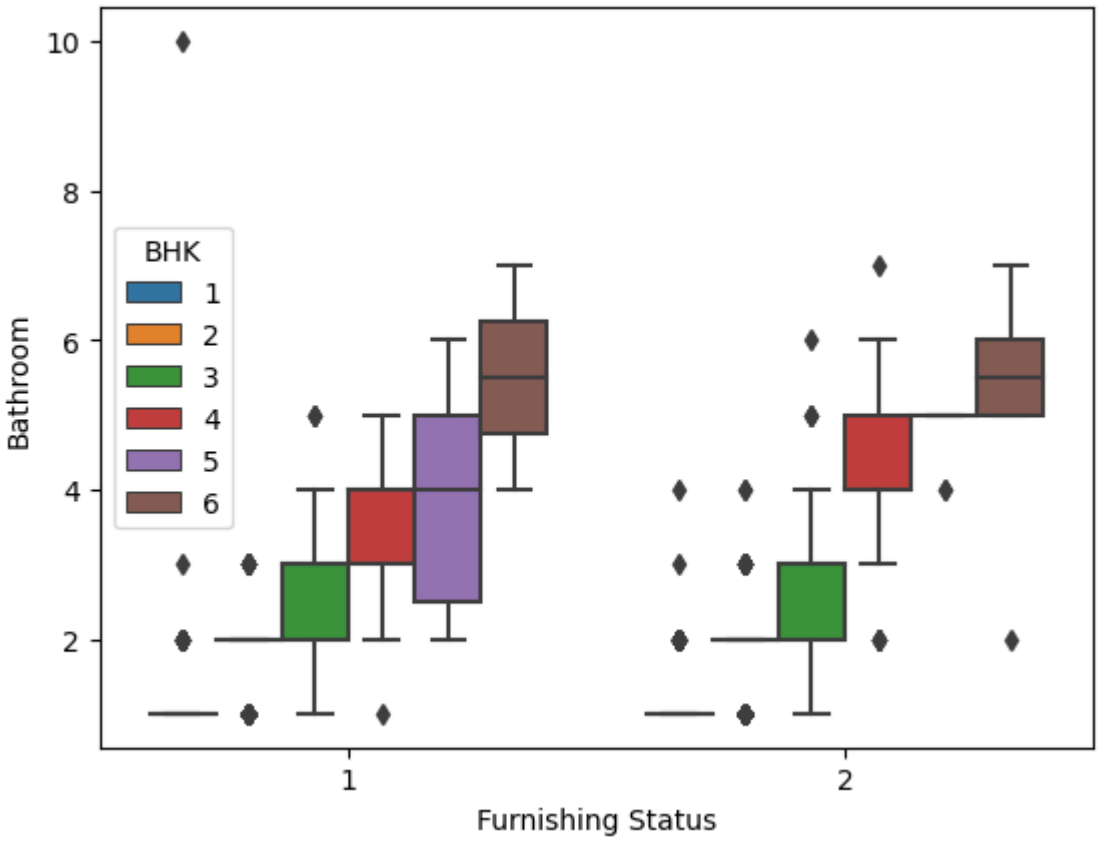
Out[66]: <AxesSubplot:xlabel='Bathroom'>



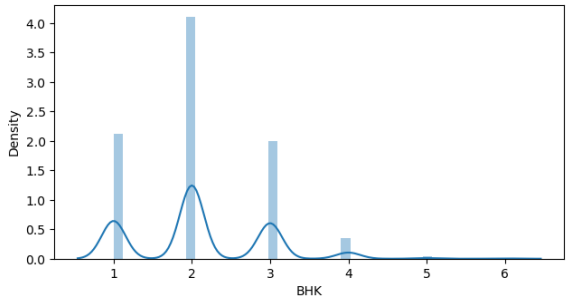


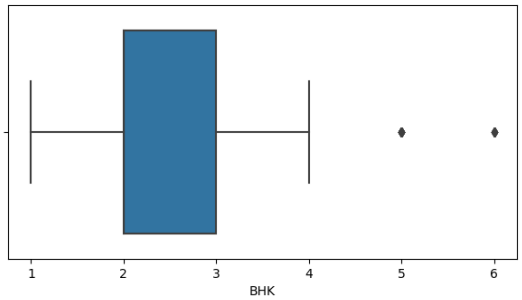
In [67]:
```
sns.boxplot(x ='Furnishing Status', y ='Bathroom', data = df1, hue ='BHK')
```

Out[67]:
```
<AxesSubplot:xlabel='Furnishing Status', ylabel='Bathroom'>
```



In [68]:
```
# compare plots before trimming
plt.figure(figsize=(16,8))
plt.subplot(2,2,1)
sns.distplot(df1['BHK'])
plt.subplot(2,2,2)
sns.boxplot(df1['BHK'])
plt.show()
```





In [69]:
```
# finding boundary values for bmi using z-score
print("Highest allowed",df1['BHK'].mean() + 3*df1['BHK'].std())
print("Lowest allowed",df1['BHK'].mean() - 3*df1['BHK'].std())
```

```
Highest allowed 4.580627788100002
Lowest allowed -0.4129076026807015
```

In [70]:
```
# finding outliers for bmi
df1[(df1['BHK'] > 4.580627788100002) | (df1['BHK'] < -0.4129076026807015)]
```

Out[70]:

	Posted On	BHK	Rent	Size	Floor	Area Type	Area Locality	City	Furnishing Status	
83	2022-06-21	6	20000	1000	1 out of 1	1	Sonarpur	1	2	Bache
460	2022-06-08	5	22500	960	Ground out of 1	1	Kasba	1	1	Bache
521	2022-05-12	5	25000	1880	2 out of 2	2	Dashdrone	1	1	
543	2022-07-04	5	400000	2308	17 out of 31	2	Runwal Elegante, Andheri West	2	2	
666	2022-07-04	5	350000	1880	15 out of 31	2	Runwal Elegante, Andheri West	2	2	
792	2022-06-13	5	600000	3200	10 out of 16	2	Bandra East	2	2	Bache
839	2022-05-19	5	450000	3400	2 out of 5	2	Bandra West	2	2	Bache
848	2022-06-12	5	300000	3000	2 out of 22	2	Chandivali	2	2	Bache
1023	2022-07-09	5	500000	3900	4 out of 6	2	Altamount Road	2	2	Bache
1037	2022-05-31	5	350000	3250	12 out of 18	2	Deonar	2	1	
1287	2022-06-27	5	360000	1880	11 out of 27	2	Runwal Elegante, Andheri West	2	2	
1319	2022-05-31	5	650000	3000	8 out of 10	2	Khar West	2	2	Bache
1384	2022-05-27	5	600000	4500	7 out of 20	2	Bandra West	2	2	
1393	2022-06-29	5	310000	2800	19 out of 33	2	MidCity Shikhar, Andheri West	2	2	Bache
2656	2022-05-22	5	190000	200	2 out of 2	2	Safdarjung Development Area, Hauz Khas	4	1	
3320	2022-06-28	6	280000	4500	Ground out of 1	2	Raja Annamalai Puram	5	2	
3472	2022-05-05	5	50000	3300	Ground out of 2	1	Valasaravakkam, Arcot Road	5	2	Bache
3584	2022-06-02	6	60000	1800	1 out of 3	1	Besant Nagar	5	2	Bache
3622	2022-05-24	5	100000	6000	9 out of 12	1	Navalur	5	2	Bache
3953	2022-05-23	5	45000	2300	3 out of 5	2	Narayanguda	Hyderabad	1	
3989	2022-07-02	5	300000	4050	Ground out of 5	1	Bandlaguda Jagir,, Hyderabad	Hyderabad	1	Bache

	Posted On	BHK	Rent	Size	Floor	Area Type	Area Locality	City	Furnishing Status	
4350	2022-06-18	6	60000	2800	Ground out of 3	1	Boduppal, NH 2 2	Hyderabad	2	Bache
4409	2022-07-02	5	50000	3500	2 out of 3	2	Ramanthapur, NH 2 2	Hyderabad	1	Bache
4479	2022-07-07	6	20000	2400	Ground out of 1	1	Mallikarjuna Nagar, Secunderabad	Hyderabad	1	
4518	2022-05-12	6	35000	3400	Ground out of 2	1	Rampally	Hyderabad	1	Bache
4648	2022-07-06	6	80000	4500	1 out of 2	1	Kakateeya Nagar, Secunderabad	Hyderabad	2	Bache
4696	2022-07-06	6	30000	4200	Ground out of 2	1	Kuntloor	Hyderabad	2	Bache

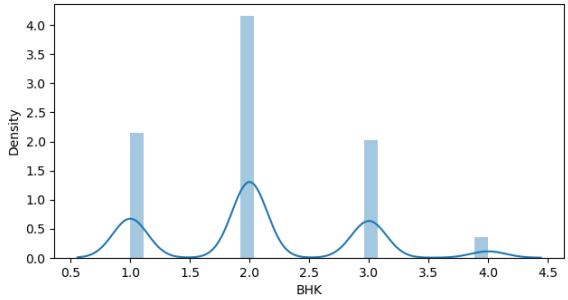
In [71]:
```
# Trimming of outliers for bmi
df2 = df1[(df1['BHK'] < 4.580627788100002) & (df1['BHK'] > -0.4129076026807015
df2
```

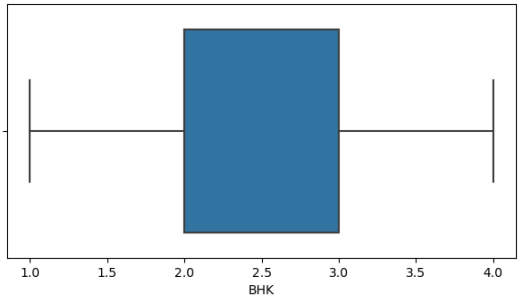
Out[71]:

	Posted On	BHK	Rent	Size	Floor	Area Type	Area Locality	City	Furnishing Status	Te Prefe
0	2022-05-18	2	10000	1100	Ground out of 2	1	Bandel	1	1	Bachelors/F
1	2022-05-13	2	20000	800	1 out of 3	1	Phool Bagan, Kankurgachi	1	2	Bachelors/F
2	2022-05-16	2	17000	1000	1 out of 3	1	Salt Lake City Sector 2	1	2	Bachelors/F
3	2022-07-04	2	10000	800	1 out of 2	1	Dumdum Park	1	1	Bachelors/F
4	2022-05-09	2	7500	850	1 out of 2	2	South Dum Dum	1	1	Bach
...	...	...	...	...	...	...	...	...	...	
4741	2022-05-18	2	15000	1000	3 out of 5	2	Bandam Kommu	Hyderabad	2	Bachelors/F
4742	2022-05-15	3	29000	2000	1 out of 4	1	Manikonda, Hyderabad	Hyderabad	2	Bachelors/F
4743	2022-07-10	3	35000	1750	3 out of 5	2	Himayath Nagar, NH 7	Hyderabad	2	Bachelors/F
4744	2022-07-06	3	45000	1500	23 out of 34	2	Gachibowli	Hyderabad	2	F
4745	2022-05-04	2	15000	1000	4 out of 5	2	Suchitra Circle	Hyderabad	1	Bach

4719 rows × 11 columns

In [72]:
```
# compare plots after trimming
plt.figure(figsize=(16,8))
plt.subplot(2,2,1)
sns.distplot(df2['BHK'])
plt.subplot(2,2,2)
sns.boxplot(df2['BHK'])
plt.show()
```





In [ ]:
```
df2.to_csv("House_Rent")
```

In [ ]:
```

```