```
import pandas as pd

df_bookings = pd.read_csv("fact_bookings.csv")

df_bookings.head(4)
```

```
         booking_id  property_id booking_date check_in_date
checkout_date  \
0  May012216558RT11        16558     27-04-22      1/5/2022
2/5/2022
1  May012216558RT12        16558     30-04-22      1/5/2022
2/5/2022
2  May012216558RT13        16558     28-04-22      1/5/2022
4/5/2022
3  May012216558RT14        16558     28-04-22      1/5/2022
2/5/2022

   no_guests room_category booking_platform  ratings_given
booking_status  \
0       -3.0           RT1    direct online            1.0    Checked
Out
1        2.0           RT1           others            NaN
Cancelled
2        2.0           RT1          logtrip            5.0    Checked
Out
3       -2.0           RT1           others            NaN
Cancelled

   revenue_generated  revenue_realized
0              10010             10010
1               9100              3640
2            9100000              9100
3               9100              3640
```

```
df_bookings.shape
```

```
(134590, 12)
```

```
df_bookings.room_category.unique()
```

```
array(['RT1', 'RT2', 'RT3', 'RT4'], dtype=object)
```

```
df_bookings.booking_platform.unique()
```

```
array(['direct online', 'others', 'logtrip', 'tripster',
'makeyourtrip',
       'journey', 'direct offline'], dtype=object)
```

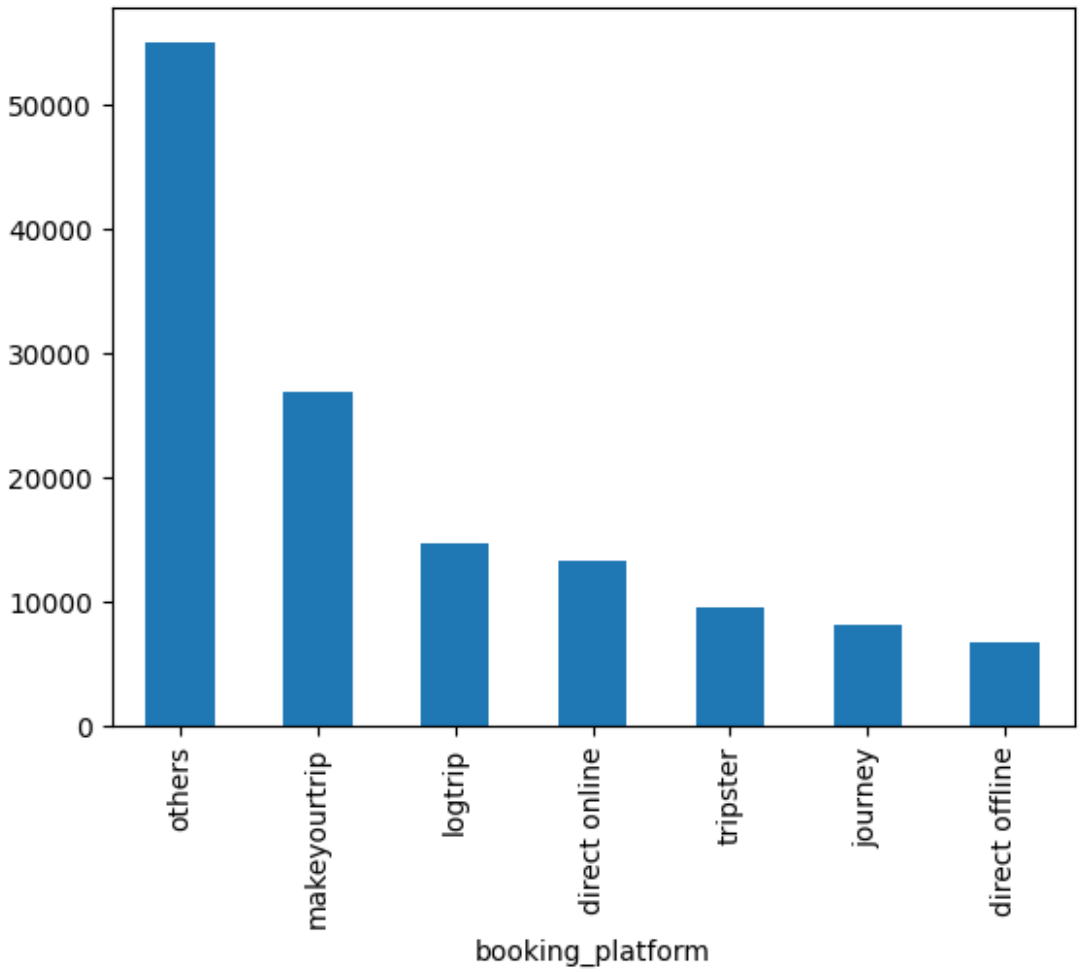
```
df_bookings.booking_platform.value_counts()
```

```
booking_platform
others            55066
```

```
makeyourtrip      26898
logtrip           14756
direct online     13379
tripster           9630
journey            8106
direct offline     6755
Name: count, dtype: int64
```

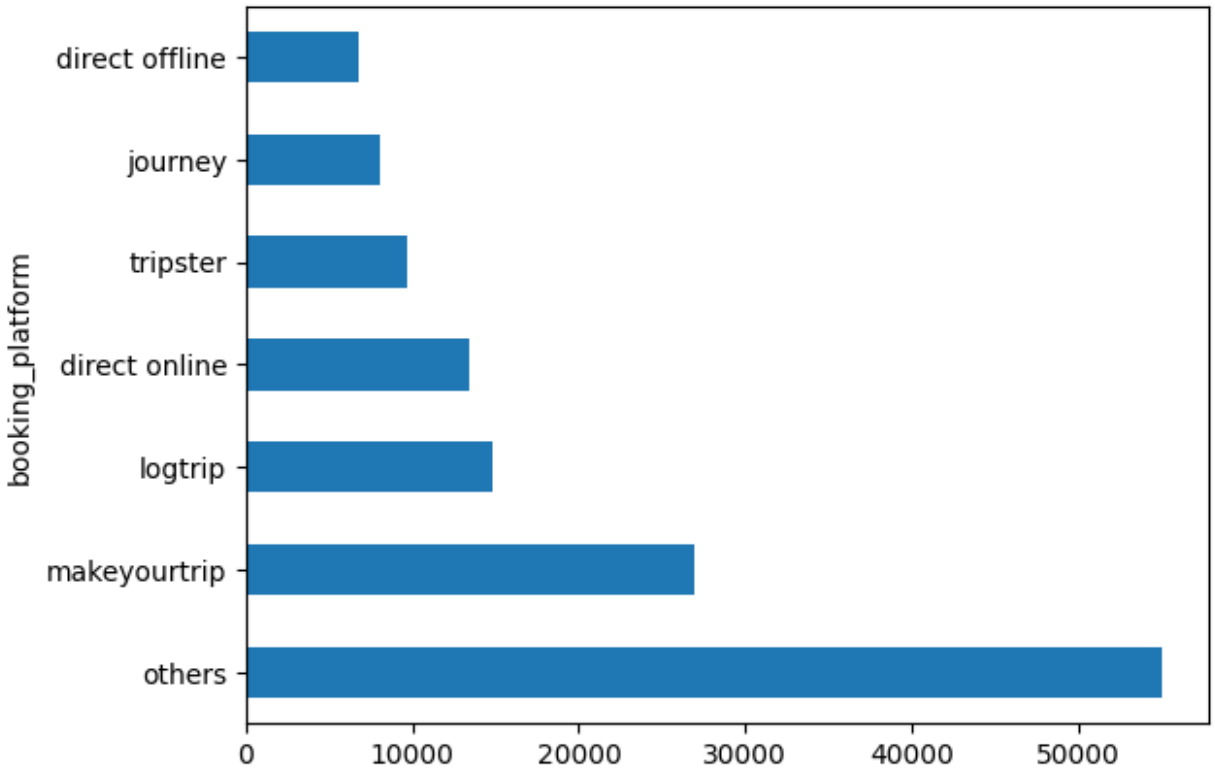
```
import matplotlib.pyplot as plt

df_bookings.booking_platform.value_counts().plot(kind="bar")
```

```
<Axes: xlabel='booking_platform'>
```




```
df_bookings.booking_platform.value_counts().plot(kind="barh")
```

```
<Axes: ylabel='booking_platform'>
```




```
df_bookings.describe()
```

	property_id	no_guests	ratings_given	revenue_generated	revenue_realized
count	134590.000000	134587.000000	56683.000000	1.345900e+05	134590.000000
mean	18061.113493	2.036170	3.619004	1.537805e+04	12696.123256
std	1093.055847	1.034885	1.235009	9.303604e+04	6928.108124
min	16558.000000	-17.000000	1.000000	6.500000e+03	2600.000000
25%	17558.000000	1.000000	3.000000	9.900000e+03	
50%	17564.000000	2.000000	4.000000	1.350000e+04	
75%	18563.000000	2.000000	5.000000	1.800000e+04	
max	19563.000000	6.000000	5.000000	2.856000e+07	

```
25%          7600.000000
50%         11700.000000
75%         15300.000000
max         45220.000000
```

```
df_bookings.revenue_generated.min(),df_bookings.revenue_generated.max(
)
```

```
(6500, 28560000)
```

```
df_date = pd.read_csv('dim_date.csv')
df_hotels = pd.read_csv('dim_hotels.csv')
df_rooms = pd.read_csv('dim_rooms.csv')
df_agg_bookings = pd.read_csv('fact_aggregated_bookings.csv')
```

```
df_hotels.shape
```

```
(25, 4)
```

```
df_hotels.head(4)
```

```
   property_id  property_name  category    city
0        16558   Atliq Grands    Luxury   Delhi
1        16559  Atliq Exotica    Luxury  Mumbai
2        16560     Atliq City  Business   Delhi
3        16561      Atliq Blu    Luxury   Delhi
```

```
df_hotels.category.value_counts()
```

```
category
Luxury      16
Business     9
Name: count, dtype: int64
```

```
df_hotels.city.value_counts()
```

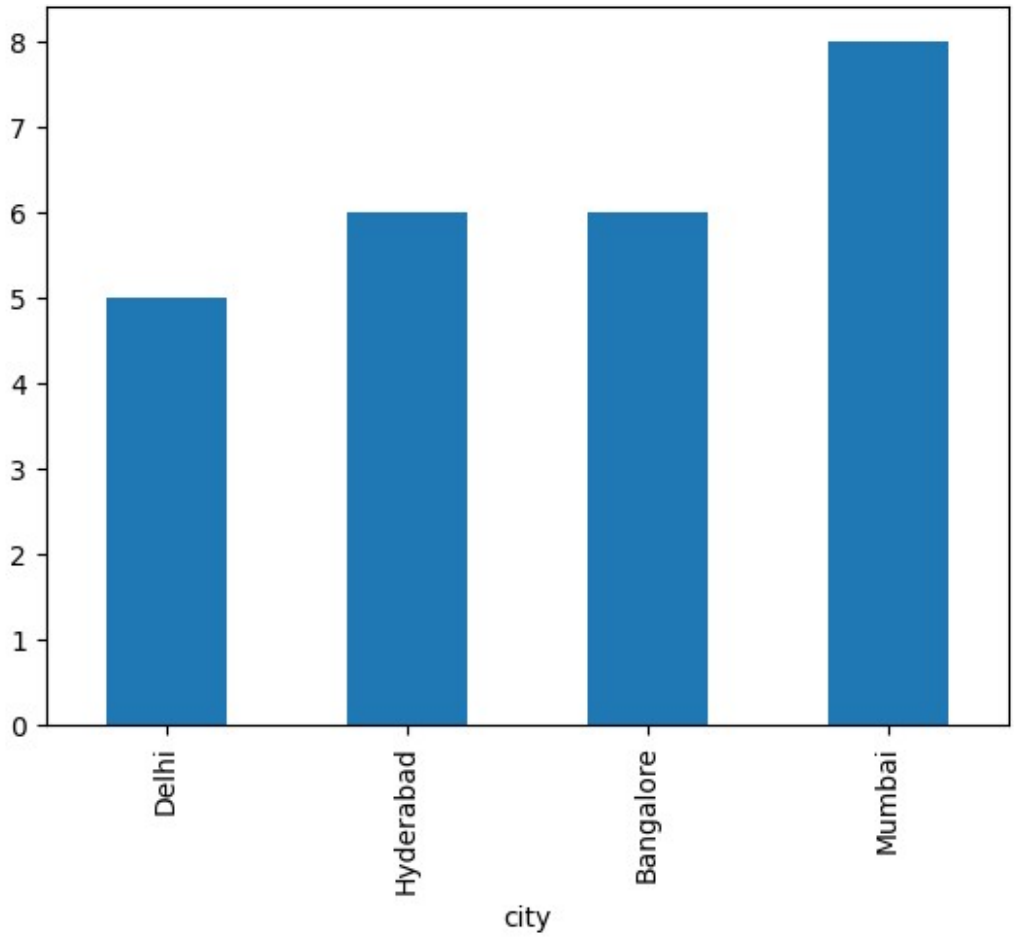
```
city
Mumbai       8
Hyderabad    6
Bangalore    6
Delhi        5
Name: count, dtype: int64
```

```
df_hotels.city.value_counts().sort_values()
```

```
city
Delhi        5
Hyderabad    6
Bangalore    6
Mumbai       8
Name: count, dtype: int64
```

```
df_hotels.city.value_counts().sort_values().plot(kind="bar")
```

```
<Axes: xlabel='city'>
```




```
df_bookings.describe()
```

	property_id	no_guests	ratings_given	revenue_generated
count	134590.000000	134587.000000	56683.000000	1.345900e+05
mean	18061.113493	2.036170	3.619004	1.537805e+04
std	1093.055847	1.034885	1.235009	9.303604e+04
min	16558.000000	-17.000000	1.000000	6.500000e+03
25%	17558.000000	1.000000	3.000000	9.900000e+03
50%	17564.000000	2.000000	4.000000	1.350000e+04
75%	18563.000000	2.000000	5.000000	1.800000e+04

```
max     19563.000000       6.000000       5.000000       2.856000e+07


       revenue_realized
count     134590.000000
mean       12696.123256
std         6928.108124
min         2600.000000
25%         7600.000000
50%        11700.000000
75%        15300.000000
max        45220.000000

df_bookings[df_bookings.no_guests<=0]

               booking_id  property_id booking_date check_in_date  \
0        May012216558RT11        16558     27-04-22      1/5/2022
3        May012216558RT14        16558     28-04-22      1/5/2022
17924    May122218559RT44        18559    12/5/2022     12/5/2022
18020    May122218561RT22        18561     8/5/2022     12/5/2022
18119   May122218562RT311        18562     5/5/2022     12/5/2022
18121   May122218562RT313        18562    10/5/2022     12/5/2022
56715    Jun082218562RT12        18562     5/6/2022      8/6/2022
119765  Jul202219560RT220        19560     19-07-22      20-07-22
134586   Jul312217564RT47        17564     30-07-22      31-07-22

       checkout_date  no_guests room_category booking_platform
ratings_given  \
0           2/5/2022       -3.0           RT1    direct online
1.0
3           2/5/2022       -2.0           RT1           others
NaN
17924       14-05-22      -10.0           RT4    direct online
NaN
18020       14-05-22      -12.0           RT2     makeyourtrip
NaN
18119       17-05-22       -6.0           RT3   direct offline
5.0
18121       17-05-22       -4.0           RT3    direct online
NaN
56715       13-06-22      -17.0           RT1           others
NaN
119765      22-07-22       -1.0           RT2           others
NaN
134586      1/8/2022       -4.0           RT4          logtrip
2.0

       booking_status  revenue_generated  revenue_realized
0         Checked Out              10010             10010
```

```
3           Cancelled                9100               3640
17924         No Show               20900              20900
18020       Cancelled                9000               3600
18119     Checked Out               16800              16800
18121       Cancelled               14400               5760
56715     Checked Out                6500               6500
119765    Checked Out               13500              13500
134586    Checked Out               38760              38760
```

```
df_bookings=df_bookings[df_bookings.no_guests>0]
```

```
df_bookings.shape
```

```
(134578, 12)
```

```
df_bookings.revenue_generated.min(),df_bookings.revenue_generated.max(
)
```

```
(6500, 28560000)
```

```
avg,std =
df_bookings.revenue_generated.mean(),df_bookings.revenue_generated.std
()
```

```
avg,std
```

```
(15378.036937686695, 93040.1549314641)
```

```
higher_limit = avg + 3*std
```

```
higher_limit
```

```
294498.50173207896
```

```
lower_limit = avg - 3*std
```

```
lower_limit
```

```
-263742.4278567056
```

```
df_bookings[df_bookings.revenue_generated>higher_limit]
```

```
               booking_id  property_id booking_date check_in_date  \
2        May012216558RT13        16558     28-04-22      1/5/2022
111      May012216559RT32        16559     29-04-22      1/5/2022
315      May012216562RT22        16562     28-04-22      1/5/2022
562     May012217559RT118        17559     26-04-22      1/5/2022
129176   Jul282216562RT26         16562     21-07-22      28-07-22

       checkout_date  no_guests room_category booking_platform
ratings_given  \
2           4/5/2022        2.0           RT1          logtrip
5.0
```

```
111         2/5/2022        6.0           RT3    direct online
NaN
315         4/5/2022        2.0           RT2   direct offline
3.0
562         2/5/2022        2.0           RT1           others
NaN
129176      29-07-22        2.0           RT2    direct online
3.0

       booking_status  revenue_generated  revenue_realized
2         Checked Out            9100000              9100
111       Checked Out           28560000             28560
315       Checked Out           12600000             12600
562         Cancelled            2000000              4420
129176    Checked Out           10000000             12600
```

```
df_bookings = df_bookings[df_bookings.revenue_generated<higher_limit]
```

```
df_bookings.shape
```

```
(134573, 12)
```

```
df_bookings.revenue_realized.describe()
```

```
count    134573.000000
mean      12695.983585
std        6927.791692
min        2600.000000
25%        7600.000000
50%       11700.000000
75%       15300.000000
max       45220.000000
Name: revenue_realized, dtype: float64
```

```
higher_limit = df_bookings.revenue_realized.mean() +
3*df_bookings.revenue_realized.std()
higher_limit
```

```
33479.358661845814
```

```
df_bookings[df_bookings.revenue_realized>higher_limit]
```

```
               booking_id  property_id booking_date check_in_date  \
137      May012216559RT41        16559     27-04-22      1/5/2022
139      May012216559RT43        16559     1/5/2022      1/5/2022
143      May012216559RT47        16559     28-04-22      1/5/2022
149     May012216559RT413        16559     24-04-22      1/5/2022
222      May012216560RT45        16560     30-04-22      1/5/2022
...                   ...          ...          ...           ...
134328   Jul312219560RT49        19560     31-07-22      31-07-22
134331  Jul312219560RT412        19560     31-07-22      31-07-22
```

```
134467   Jul312219562RT45         19562     28-07-22      31-07-22
134474  Jul312219562RT412         19562     25-07-22      31-07-22
134581   Jul312217564RT42         17564     31-07-22      31-07-22

       checkout_date  no_guests room_category booking_platform
ratings_given  \
137          7/5/2022        4.0           RT4           others
NaN
139          2/5/2022        6.0           RT4         tripster
3.0
143          3/5/2022        3.0           RT4           others
5.0
149          7/5/2022        5.0           RT4          logtrip
NaN
222          3/5/2022        5.0           RT4           others
3.0
...               ...        ...           ...              ...
...
134328       2/8/2022        6.0           RT4    direct online
5.0
134331       1/8/2022        6.0           RT4           others
2.0
134467       1/8/2022        6.0           RT4     makeyourtrip
4.0
134474       6/8/2022        5.0           RT4   direct offline
5.0
134581       1/8/2022        4.0           RT4     makeyourtrip
4.0

       booking_status  revenue_generated  revenue_realized
137       Checked Out              38760             38760
139       Checked Out              45220             45220
143       Checked Out              35530             35530
149       Checked Out              41990             41990
222       Checked Out              34580             34580
...               ...                ...               ...
134328    Checked Out              39900             39900
134331    Checked Out              39900             39900
134467    Checked Out              39900             39900
134474    Checked Out              37050             37050
134581    Checked Out              38760             38760

[1299 rows x 12 columns]

df_rooms

  room_id    room_class
0     RT1      Standard
1     RT2         Elite
```

```
2     RT3       Premium
3     RT4  Presidential
```

```
df_bookings[df_bookings.room_category=="RT4"].revenue_realized.describe()
```

```
count    16071.000000
mean     23439.308444
std       9048.599076
min       7600.000000
25%      19000.000000
50%      26600.000000
75%      32300.000000
max      45220.000000
Name: revenue_realized, dtype: float64
```

```
df_bookings.isnull().sum()
```

```
booking_id                0
property_id               0
booking_date              0
check_in_date             0
checkout_date             0
no_guests                 0
room_category             0
booking_platform          0
ratings_given         77897
booking_status            0
revenue_generated         0
revenue_realized          0
dtype: int64
```

Data Transformation

```
df_agg_bookings.head()
```

```
   property_id check_in_date room_category  successful_bookings
capacity
0        16559      1-May-22           RT1                   25
30.0
1        19562      1-May-22           RT1                   28
30.0
2        19563      1-May-22           RT1                   23
30.0
3        17558      1-May-22           RT1                   30
19.0
```

```
4        16558      1-May-22           RT1                    18
19.0

df_agg_bookings["occ_pct"] =
df_agg_bookings["successful_bookings"]/df_agg_bookings["capacity"]

df_agg_bookings.head()

   property_id check_in_date room_category  successful_bookings
capacity  \
0        16559      1-May-22           RT1                    25
30.0
1        19562      1-May-22           RT1                    28
30.0
2        19563      1-May-22           RT1                    23
30.0
3        17558      1-May-22           RT1                    30
19.0
4        16558      1-May-22           RT1                    18
19.0

    occ_pct
0  0.833333
1  0.933333
2  0.766667
3  1.578947
4  0.947368

df_agg_bookings["occ_pct"] = df_agg_bookings["occ_pct"].apply(lambda x
: round(x*100,2))
df_agg_bookings.head()

   property_id check_in_date room_category  successful_bookings
capacity  \
0        16559      1-May-22           RT1                    25
30.0
1        19562      1-May-22           RT1                    28
30.0
2        19563      1-May-22           RT1                    23
30.0
3        17558      1-May-22           RT1                    30
19.0
4        16558      1-May-22           RT1                    18
19.0

   occ_pct
0    83.33
1    93.33
2    76.67
3   157.89
4    94.74
```

Insights Generation

1. What is an average occupancy rate in each of the room category ?

```
df_agg_bookings.groupby("room_category")["occ_pct"].mean().round(2)
```

```
room_category
RT1    58.22
RT2    58.04
RT3    58.03
RT4    59.30
Name: occ_pct, dtype: float64
```

```
df_rooms
```

```
  room_id    room_class
0     RT1      Standard
1     RT2         Elite
2     RT3       Premium
3     RT4  Presidential
```

```
 df = pd.merge(df_agg_bookings ,df_rooms, left_on="room_category",
right_on="room_id")
df.tail()
```

```
      property_id check_in_date room_category  successful_bookings
capacity  \
9195        16563     31-Jul-22           RT4                   13
18.0
9196        16559     31-Jul-22           RT4                   13
18.0
9197        17558     31-Jul-22           RT4                    3
6.0
9198        19563     31-Jul-22           RT4                    3
6.0
9199        17561     31-Jul-22           RT4                    3
4.0

      occ_pct room_id    room_class
9195    72.22     RT4  Presidential
9196    72.22     RT4  Presidential
9197    50.00     RT4  Presidential
9198    50.00     RT4  Presidential
9199    75.00     RT4  Presidential
```

```
df.groupby("room_class")["occ_pct"].mean().round(2)
```

```
room_class
Elite           58.04
Premium         58.03
Presidential    59.30
Standard        58.22
Name: occ_pct, dtype: float64
```

```
df.drop("room_id", axis=1, inplace=True)
df.head()
```

```
   property_id check_in_date room_category  successful_bookings
capacity  \
0        16559      1-May-22           RT1                   25
30.0
1        19562      1-May-22           RT1                   28
30.0
2        19563      1-May-22           RT1                   23
30.0
3        17558      1-May-22           RT1                   30
19.0
4        16558      1-May-22           RT1                   18
19.0

   occ_pct room_class
0    83.33   Standard
1    93.33   Standard
2    76.67   Standard
3   157.89   Standard
4    94.74   Standard
```

2.Print average occupancy rate per city

```
df_hotels.head()
```

```
   property_id  property_name  category    city
0        16558   Atliq Grands    Luxury   Delhi
1        16559  Atliq Exotica    Luxury  Mumbai
2        16560     Atliq City  Business   Delhi
3        16561      Atliq Blu    Luxury   Delhi
4        16562      Atliq Bay    Luxury   Delhi
```

```
df = pd.merge(df, df_hotels, on="property_id")
df.head()
```

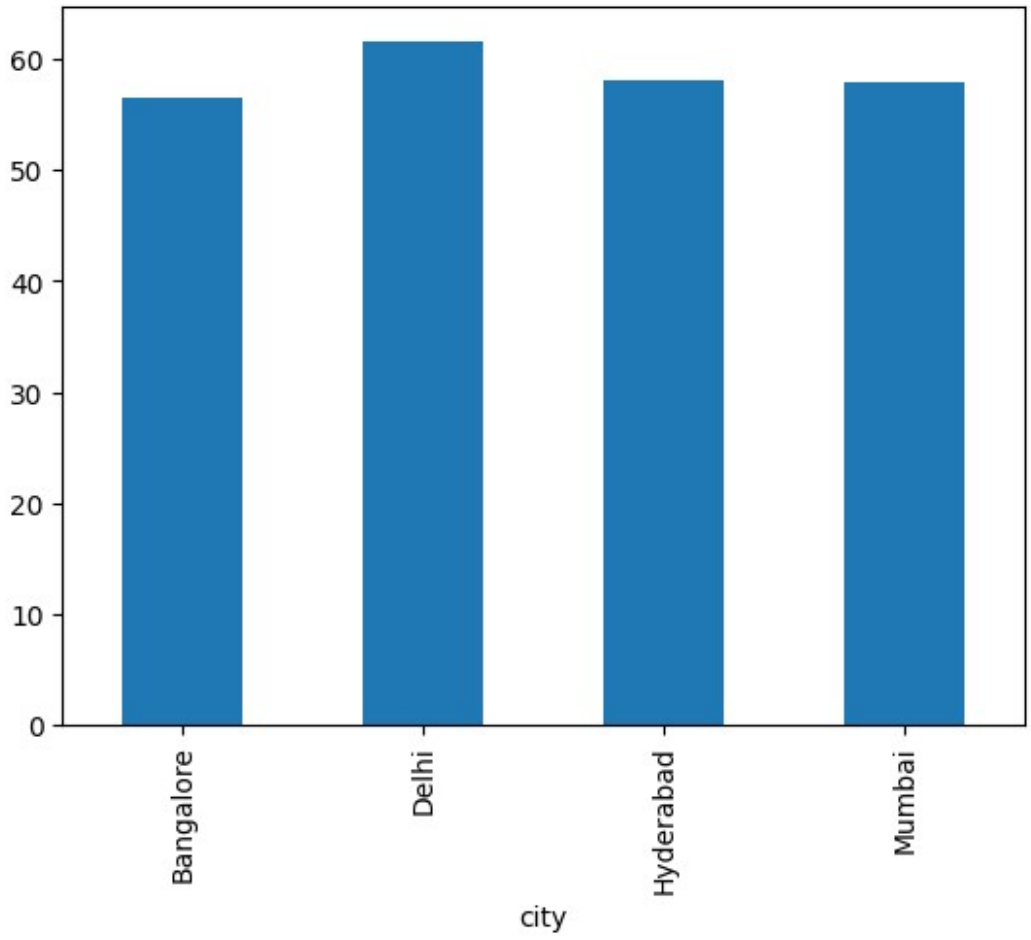
```
   property_id check_in_date room_category  successful_bookings
capacity  \
0        16559      1-May-22           RT1                   25
30.0
1        16559      2-May-22           RT1                   20
30.0
```

```
2         16559      3-May-22            RT1                      17
30.0
3         16559      4-May-22            RT1                      21
30.0
4         16559      5-May-22            RT1                      16
30.0
```

```
   occ_pct room_class  property_name category    city
0    83.33   Standard  Atliq Exotica   Luxury  Mumbai
1    66.67   Standard  Atliq Exotica   Luxury  Mumbai
2    56.67   Standard  Atliq Exotica   Luxury  Mumbai
3    70.00   Standard  Atliq Exotica   Luxury  Mumbai
4    53.33   Standard  Atliq Exotica   Luxury  Mumbai
```

```
df.groupby("city")["occ_pct"].mean().plot(kind="bar")
```

```
<Axes: xlabel='city'>
```



3. When was the occupancy better ? Weekday or Weeked?

```
df.head()
```

```
   property_id check_in_date room_category  successful_bookings
capacity  \
0        16559      1-May-22           RT1                   25
30.0
1        16559      2-May-22           RT1                   20
30.0
2        16559      3-May-22           RT1                   17
30.0
3        16559      4-May-22           RT1                   21
30.0
4        16559      5-May-22           RT1                   16
30.0

   occ_pct room_class  property_name category    city
0    83.33   Standard  Atliq Exotica   Luxury  Mumbai
1    66.67   Standard  Atliq Exotica   Luxury  Mumbai
2    56.67   Standard  Atliq Exotica   Luxury  Mumbai
3    70.00   Standard  Atliq Exotica   Luxury  Mumbai
4    53.33   Standard  Atliq Exotica   Luxury  Mumbai
```

```
df = pd.merge(df, df_date, left_on="check_in_date", right_on="date")
df.head()
```

```
   property_id check_in_date room_category  successful_bookings
capacity  \
0        16559     10-May-22           RT1                   18
30.0
1        16559     10-May-22           RT2                   25
41.0
2        16559     10-May-22           RT3                   20
32.0
3        16559     10-May-22           RT4                   13
18.0
4        19562     10-May-22           RT1                   18
30.0

   occ_pct    room_class  property_name category       city       date
\
0    60.00      Standard  Atliq Exotica   Luxury     Mumbai  10-May-22

1    60.98         Elite  Atliq Exotica   Luxury     Mumbai  10-May-22

2    62.50       Premium  Atliq Exotica   Luxury     Mumbai  10-May-22

3    72.22  Presidential  Atliq Exotica   Luxury     Mumbai  10-May-22

4    60.00      Standard      Atliq Bay   Luxury  Bangalore  10-May-22


   mmm yy week no  day_type
```

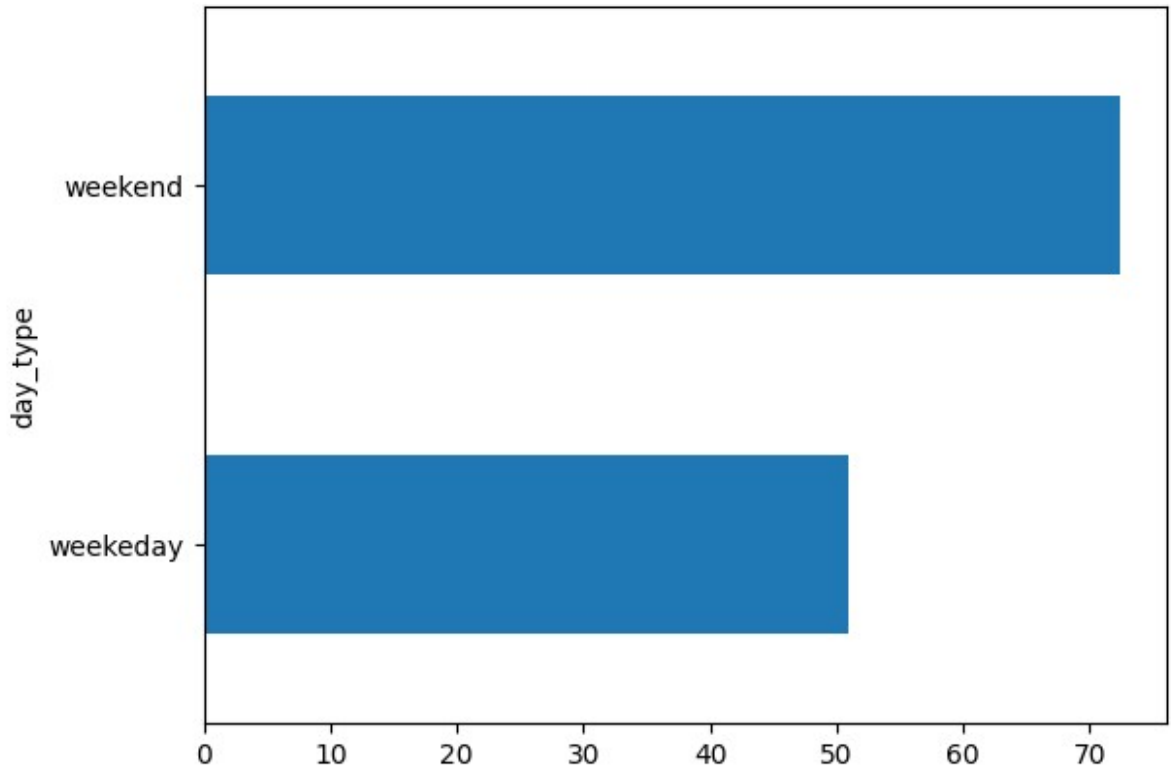
```
0  May 22    W 20  weekeday
1  May 22    W 20  weekeday
2  May 22    W 20  weekeday
3  May 22    W 20  weekeday
4  May 22    W 20  weekeday
```

```
df.groupby("day_type")["occ_pct"].mean().round(2)
```

```
day_type
weekeday    50.90
weekend     72.39
Name: occ_pct, dtype: float64
```

```
df.groupby("day_type")["occ_pct"].mean().round(2).plot(kind="barh")
```

```
<Axes: ylabel='day_type'>
```



4. In the month of june , what is teh occupancy for different cities

```
df["mmm yy"].unique()
```

```
array(['May 22', 'Jun 22', 'Jul 22'], dtype=object)
```

```
df_june_22 = df[df["mmm yy"]=="Jun 22"]
df_june_22.head()
```

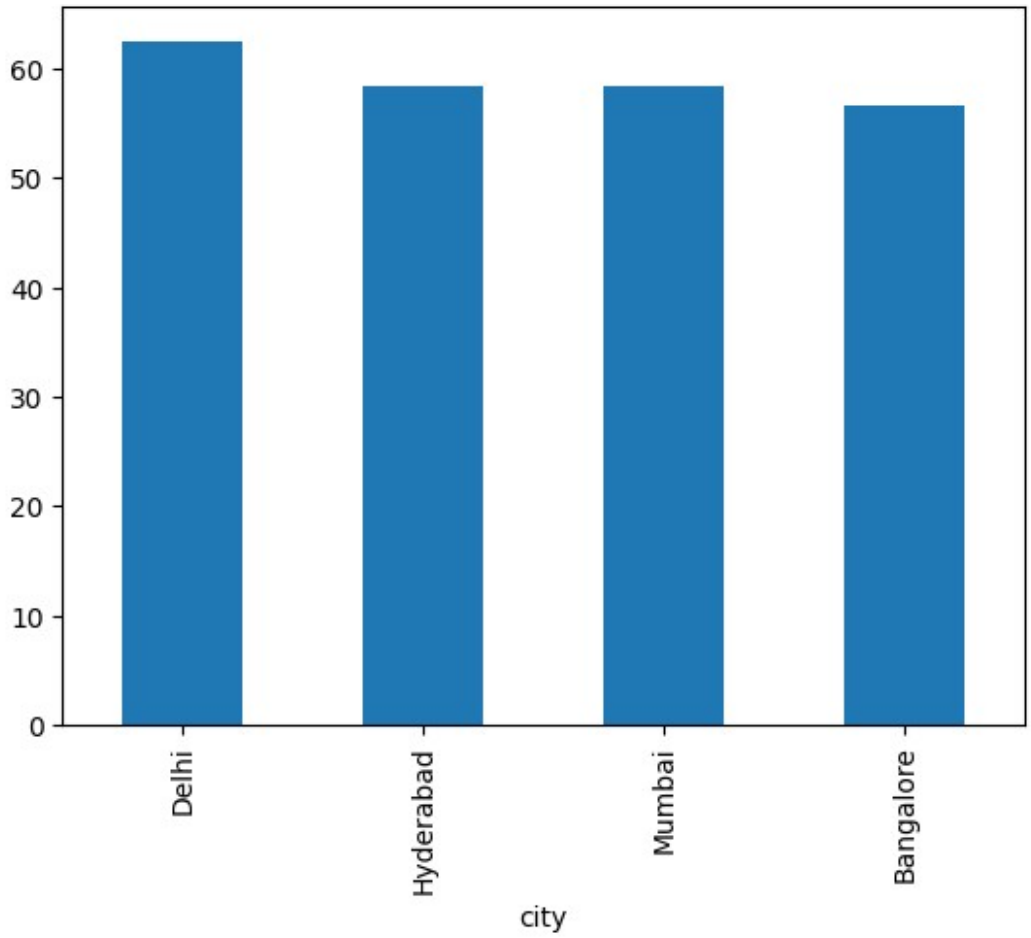
```
      property_id check_in_date room_category  successful_bookings
capacity  \
2200        16559     10-Jun-22           RT1                   20
30.0
2201        16559     10-Jun-22           RT2                   26
41.0
2202        16559     10-Jun-22           RT3                   20
32.0
2203        16559     10-Jun-22           RT4                   11
18.0
2204        19562     10-Jun-22           RT1                   19
30.0

      occ_pct    room_class  property_name category       city
date  \
2200    66.67      Standard  Atliq Exotica   Luxury     Mumbai  10-
Jun-22
2201    63.41         Elite  Atliq Exotica   Luxury     Mumbai  10-
Jun-22
2202    62.50       Premium  Atliq Exotica   Luxury     Mumbai  10-
Jun-22
2203    61.11  Presidential  Atliq Exotica   Luxury     Mumbai  10-
Jun-22
2204    63.33      Standard      Atliq Bay   Luxury  Bangalore  10-
Jun-22

      mmm yy week no  day_type
2200  Jun 22    W 24  weekeday
2201  Jun 22    W 24  weekeday
2202  Jun 22    W 24  weekeday
2203  Jun 22    W 24  weekeday
2204  Jun 22    W 24  weekeday
```

```
df_june_22.groupby("city")
["occ_pct"].mean().round(2).sort_values(ascending=False)
```

```
city
Delhi        62.47
Hyderabad    58.46
Mumbai       58.38
Bangalore    56.58
Name: occ_pct, dtype: float64
```

```
df_june_22.groupby("city")
["occ_pct"].mean().round(2).sort_values(ascending=False).plot(kind="ba
r")
```

```
<Axes: xlabel='city'>
```




```
df_august = pd.read_csv("new_data_august.csv")
df_august
```

```
   property_id  property_name  category       city room_category
room_class  \
0        16559  Atliq Exotica    Luxury     Mumbai           RT1
Standard
1        19562      Atliq Bay    Luxury  Bangalore           RT1
Standard
2        19563   Atliq Palace  Business  Bangalore           RT1
Standard
3        19558   Atliq Grands    Luxury  Bangalore           RT1
Standard
4        19560     Atliq City  Business  Bangalore           RT1
Standard
5        17561      Atliq Blu    Luxury     Mumbai           RT1
Standard
6        17564  Atliq Seasons  Business     Mumbai           RT1
Standard
```

```
  check_in_date  mmm yy week no  day_type  successful_bookings
capacity  \
0     01-Aug-22  Aug-22    W 32  weekeday                   30
30
1     01-Aug-22  Aug-22    W 32  weekeday                   21
30
2     01-Aug-22  Aug-22    W 32  weekeday                   23
30
3     01-Aug-22  Aug-22    W 32  weekeday                   30
40
4     01-Aug-22  Aug-22    W 32  weekeday                   20
26
5     01-Aug-22  Aug-22    W 32  weekeday                   18
26
6     01-Aug-22  Aug-22    W 32  weekeday                   10
16

     occ%
0  100.00
1   70.00
2   76.67
3   75.00
4   76.92
5   69.23
6   62.50
```

```
latest_df = pd.concat([df, df_august], ignore_index=True,axis=0)
latest_df.tail(10)
```

```
      property_id check_in_date room_category  successful_bookings
capacity  \
6497        18560     31-Jul-22           RT2                   34
40.0
6498        18560     31-Jul-22           RT3                   17
24.0
6499        18560     31-Jul-22           RT4                   12
15.0
6500        16559     01-Aug-22           RT1                   30
30.0
6501        19562     01-Aug-22           RT1                   21
30.0
6502        19563     01-Aug-22           RT1                   23
30.0
6503        19558     01-Aug-22           RT1                   30
40.0
6504        19560     01-Aug-22           RT1                   20
26.0
6505        17561     01-Aug-22           RT1                   18
26.0
```

```
6506        17564     01-Aug-22           RT1                    10
16.0

      occ_pct    room_class  property_name  category       city
date  \
6497    85.00         Elite     Atliq City  Business  Hyderabad  31-
Jul-22
6498    70.83       Premium     Atliq City  Business  Hyderabad  31-
Jul-22
6499    80.00  Presidential     Atliq City  Business  Hyderabad  31-
Jul-22
6500      NaN      Standard  Atliq Exotica    Luxury     Mumbai
NaN
6501      NaN      Standard      Atliq Bay    Luxury  Bangalore
NaN
6502      NaN      Standard   Atliq Palace  Business  Bangalore
NaN
6503      NaN      Standard   Atliq Grands    Luxury  Bangalore
NaN
6504      NaN      Standard     Atliq City  Business  Bangalore
NaN
6505      NaN      Standard      Atliq Blu    Luxury     Mumbai
NaN
6506      NaN      Standard  Atliq Seasons  Business     Mumbai
NaN

      mmm yy week no  day_type    occ%
6497  Jul 22    W 32   weekend     NaN
6498  Jul 22    W 32   weekend     NaN
6499  Jul 22    W 32   weekend     NaN
6500  Aug-22    W 32  weekeday  100.00
6501  Aug-22    W 32  weekeday   70.00
6502  Aug-22    W 32  weekeday   76.67
6503  Aug-22    W 32  weekeday   75.00
6504  Aug-22    W 32  weekeday   76.92
6505  Aug-22    W 32  weekeday   69.23
6506  Aug-22    W 32  weekeday   62.50
```

6. Print revenue realized per city

```
df_bookings.head()

         booking_id  property_id booking_date check_in_date
checkout_date  \
1  May012216558RT12        16558     30-04-22      1/5/2022
2/5/2022
4  May012216558RT15        16558     27-04-22      1/5/2022
2/5/2022
5  May012216558RT16        16558     1/5/2022      1/5/2022
```

```
3/5/2022
6  May012216558RT17        16558     28-04-22      1/5/2022
6/5/2022
7  May012216558RT18        16558     26-04-22      1/5/2022
3/5/2022

   no_guests room_category booking_platform  ratings_given
booking_status  \
1        2.0           RT1           others            NaN
Cancelled
4        4.0           RT1    direct online            5.0    Checked
Out
5        2.0           RT1           others            4.0    Checked
Out
6        2.0           RT1           others            NaN
Cancelled
7        2.0           RT1          logtrip            NaN         No
Show

   revenue_generated  revenue_realized
1               9100              3640
4              10920             10920
5               9100              9100
6               9100              3640
7               9100              9100

df_hotels.head()

   property_id  property_name  category    city
0        16558   Atliq Grands    Luxury   Delhi
1        16559  Atliq Exotica    Luxury  Mumbai
2        16560     Atliq City  Business   Delhi
3        16561      Atliq Blu    Luxury   Delhi
4        16562      Atliq Bay    Luxury   Delhi

df_bookings_all = pd.merge(df_bookings, df_hotels, on="property_id")
df_bookings_all.head()

         booking_id  property_id booking_date check_in_date
checkout_date  \
0  May012216558RT12        16558     30-04-22      1/5/2022
2/5/2022
1  May012216558RT15        16558     27-04-22      1/5/2022
2/5/2022
2  May012216558RT16        16558     1/5/2022      1/5/2022
3/5/2022
3  May012216558RT17        16558     28-04-22      1/5/2022
6/5/2022
4  May012216558RT18        16558     26-04-22      1/5/2022
3/5/2022
```

```
   no_guests room_category booking_platform  ratings_given
booking_status  \
0        2.0           RT1           others            NaN
Cancelled
1        4.0           RT1    direct online            5.0    Checked
Out
2        2.0           RT1           others            4.0    Checked
Out
3        2.0           RT1           others            NaN
Cancelled
4        2.0           RT1          logtrip            NaN        No
Show

   revenue_generated  revenue_realized property_name category   city
0               9100              3640  Atliq Grands   Luxury  Delhi
1              10920             10920  Atliq Grands   Luxury  Delhi
2               9100              9100  Atliq Grands   Luxury  Delhi
3               9100              3640  Atliq Grands   Luxury  Delhi
4               9100              9100  Atliq Grands   Luxury  Delhi
```

```
df_bookings_all.groupby("city")["revenue_realized"].sum()
```

```
city
Bangalore    420383550
Delhi        294404488
Hyderabad    325179310
Mumbai       668569251
Name: revenue_realized, dtype: int64
```

7. Print month by month revenue

```
df_bookings_all.head()
```

```
         booking_id  property_id booking_date check_in_date
checkout_date  \
0  May012216558RT12        16558     30-04-22      1/5/2022
2/5/2022
1  May012216558RT15        16558     27-04-22      1/5/2022
2/5/2022
2  May012216558RT16        16558     1/5/2022      1/5/2022
3/5/2022
3  May012216558RT17        16558     28-04-22      1/5/2022
6/5/2022
4  May012216558RT18        16558     26-04-22      1/5/2022
3/5/2022

   no_guests room_category booking_platform  ratings_given
booking_status  \
0        2.0           RT1           others            NaN
```

```
Cancelled
1        4.0           RT1    direct online            5.0    Checked
Out
2        2.0           RT1          others             4.0    Checked
Out
3        2.0           RT1          others             NaN
Cancelled
4        2.0           RT1         logtrip             NaN         No
Show

   revenue_generated  revenue_realized property_name category   city
0               9100              3640  Atliq Grands   Luxury  Delhi
1              10920             10920  Atliq Grands   Luxury  Delhi
2               9100              9100  Atliq Grands   Luxury  Delhi
3               9100              3640  Atliq Grands   Luxury  Delhi
4               9100              9100  Atliq Grands   Luxury  Delhi
```

```
df_date["mmm yy"].unique()
```

```
array(['May 22', 'Jun 22', 'Jul 22'], dtype=object)
```

```
df_date.head()
```

```
        date  mmm yy week no  day_type
0  01-May-22  May 22    W 19   weekend
1  02-May-22  May 22    W 19  weekeday
2  03-May-22  May 22    W 19  weekeday
3  04-May-22  May 22    W 19  weekeday
4  05-May-22  May 22    W 19  weekeday
```

```
pd.merge(df_bookings_all, df_date, left_on="check_in_date",
right_on="date")
```

```
Empty DataFrame
Columns: [booking_id, property_id, booking_date, check_in_date,
checkout_date, no_guests, room_category, booking_platform,
ratings_given, booking_status, revenue_generated, revenue_realized,
property_name, category, city, date, mmm yy, week no, day_type]
Index: []
```

```
df_bookings_all.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134573 entries, 0 to 134572
Data columns (total 15 columns):
 #   Column             Non-Null Count   Dtype
---  ------             --------------   -----
 0   booking_id         134573 non-null  object
 1   property_id        134573 non-null  int64
 2   booking_date       134573 non-null  object
 3   check_in_date      134573 non-null  object
```

```
 4   checkout_date      134573 non-null  object
 5   no_guests          134573 non-null  float64
 6   room_category      134573 non-null  object
 7   booking_platform   134573 non-null  object
 8   ratings_given      56676 non-null   float64
 9   booking_status     134573 non-null  object
 10  revenue_generated  134573 non-null  int64
 11  revenue_realized   134573 non-null  int64
 12  property_name      134573 non-null  object
 13  category           134573 non-null  object
 14  city               134573 non-null  object
dtypes: float64(2), int64(3), object(10)
memory usage: 15.4+ MB
```

```
df_date.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 92 entries, 0 to 91
Data columns (total 4 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   date      92 non-null     object
 1   mmm yy    92 non-null     object
 2   week no   92 non-null     object
 3   day_type  92 non-null     object
dtypes: object(4)
memory usage: 3.0+ KB
```

```
df_date["date"] = pd.to_datetime(df_date["date"])
```

```
C:\Users\souvick majumder\AppData\Local\Temp\
ipykernel_11448\697511496.py:1: UserWarning: Could not infer format,
so each element will be parsed individually, falling back to
`dateutil`. To ensure parsing is consistent and as-expected, please
specify a format.
  df_date["date"] = pd.to_datetime(df_date["date"])
```

```
df_date.head()
```

```
        date  mmm yy week no  day_type
0 2022-05-01  May 22    W 19   weekend
1 2022-05-02  May 22    W 19  weekeday
2 2022-05-03  May 22    W 19  weekeday
3 2022-05-04  May 22    W 19  weekeday
4 2022-05-05  May 22    W 19  weekeday
```

```
df_date.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 92 entries, 0 to 91
Data columns (total 4 columns):
```

```
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   date      92 non-null     datetime64[ns]
 1   mmm yy    92 non-null     object
 2   week no   92 non-null     object
 3   day_type  92 non-null     object
dtypes: datetime64[ns](1), object(3)
memory usage: 3.0+ KB
```

```
df_bookings_all["check_in_date"] =
pd.to_datetime(df_bookings_all["check_in_date"], format='mixed',
errors='coerce', dayfirst=True)
df_bookings_all.head()
```

```
         booking_id  property_id booking_date check_in_date
checkout_date  \
0  May012216558RT12        16558     30-04-22    2022-05-01
2/5/2022
1  May012216558RT15        16558     27-04-22    2022-05-01
2/5/2022
2  May012216558RT16        16558     1/5/2022    2022-05-01
3/5/2022
3  May012216558RT17        16558     28-04-22    2022-05-01
6/5/2022
4  May012216558RT18        16558     26-04-22    2022-05-01
3/5/2022

   no_guests room_category booking_platform  ratings_given
booking_status  \
0        2.0           RT1           others            NaN
Cancelled
1        4.0           RT1    direct online            5.0    Checked
Out
2        2.0           RT1           others            4.0    Checked
Out
3        2.0           RT1           others            NaN
Cancelled
4        2.0           RT1          logtrip            NaN        No
Show

   revenue_generated  revenue_realized property_name category   city
0               9100              3640  Atliq Grands   Luxury  Delhi
1              10920             10920  Atliq Grands   Luxury  Delhi
2               9100              9100  Atliq Grands   Luxury  Delhi
3               9100              3640  Atliq Grands   Luxury  Delhi
4               9100              9100  Atliq Grands   Luxury  Delhi
```

```
df_bookings_all.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134573 entries, 0 to 134572
Data columns (total 15 columns):
 #   Column             Non-Null Count   Dtype
---  ------             --------------   -----
 0   booking_id         134573 non-null  object
 1   property_id        134573 non-null  int64
 2   booking_date       134573 non-null  object
 3   check_in_date      134573 non-null  datetime64[ns]
 4   checkout_date      134573 non-null  object
 5   no_guests          134573 non-null  float64
 6   room_category      134573 non-null  object
 7   booking_platform   134573 non-null  object
 8   ratings_given      56676 non-null   float64
 9   booking_status     134573 non-null  object
 10  revenue_generated  134573 non-null  int64
 11  revenue_realized   134573 non-null  int64
 12  property_name      134573 non-null  object
 13  category           134573 non-null  object
 14  city               134573 non-null  object
dtypes: datetime64[ns](1), float64(2), int64(3), object(9)
memory usage: 15.4+ MB

df_bookings_all= pd.merge(df_bookings_all, df_date,
left_on="check_in_date", right_on="date")
df_bookings_all.head()

         booking_id  property_id booking_date check_in_date
checkout_date  \
0  May012216558RT12        16558     30-04-22    2022-05-01
2/5/2022
1  May012216558RT15        16558     27-04-22    2022-05-01
2/5/2022
2  May012216558RT16        16558     1/5/2022    2022-05-01
3/5/2022
3  May012216558RT17        16558     28-04-22    2022-05-01
6/5/2022
4  May012216558RT18        16558     26-04-22    2022-05-01
3/5/2022

   no_guests room_category booking_platform  ratings_given
booking_status  \
0        2.0           RT1           others            NaN
Cancelled
1        4.0           RT1    direct online            5.0    Checked
Out
2        2.0           RT1           others            4.0    Checked
Out
3        2.0           RT1           others            NaN
Cancelled
```

```
4        2.0           RT1          logtrip             NaN         No
Show

   revenue_generated  revenue_realized property_name category
city  \
0               9100              3640  Atliq Grands   Luxury  Delhi

1              10920             10920  Atliq Grands   Luxury  Delhi

2               9100              9100  Atliq Grands   Luxury  Delhi

3               9100              3640  Atliq Grands   Luxury  Delhi

4               9100              9100  Atliq Grands   Luxury  Delhi


        date  mmm yy week no day_type
0 2022-05-01  May 22    W 19  weekend
1 2022-05-01  May 22    W 19  weekend
2 2022-05-01  May 22    W 19  weekend
3 2022-05-01  May 22    W 19  weekend
4 2022-05-01  May 22    W 19  weekend
```

```
df_bookings_all.groupby("mmm yy")["revenue_realized"].sum()
```

```
mmm yy
Jul 22    572843348
Jun 22    553925855
May 22    581767396
Name: revenue_realized, dtype: int64
```