hospitality-analysis-project

January 19, 2024

DATA EXPLORATION

[87]:
```
import pandas as pd
```

[2]:
```
df_bookings = pd.read_csv("datasets/fact_bookings.csv")
df_bookings.head(4)
```

[2]:
```
         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
```

[3]:
```
df_bookings.shape
```

[3]:
```
(134590, 12)
```

[4]:
```
df_bookings.room_category.unique()
```

[4]:
```
array(['RT1', 'RT2', 'RT3', 'RT4'], dtype=object)
```

[5]:
```
df_bookings.booking_platform.unique()
```

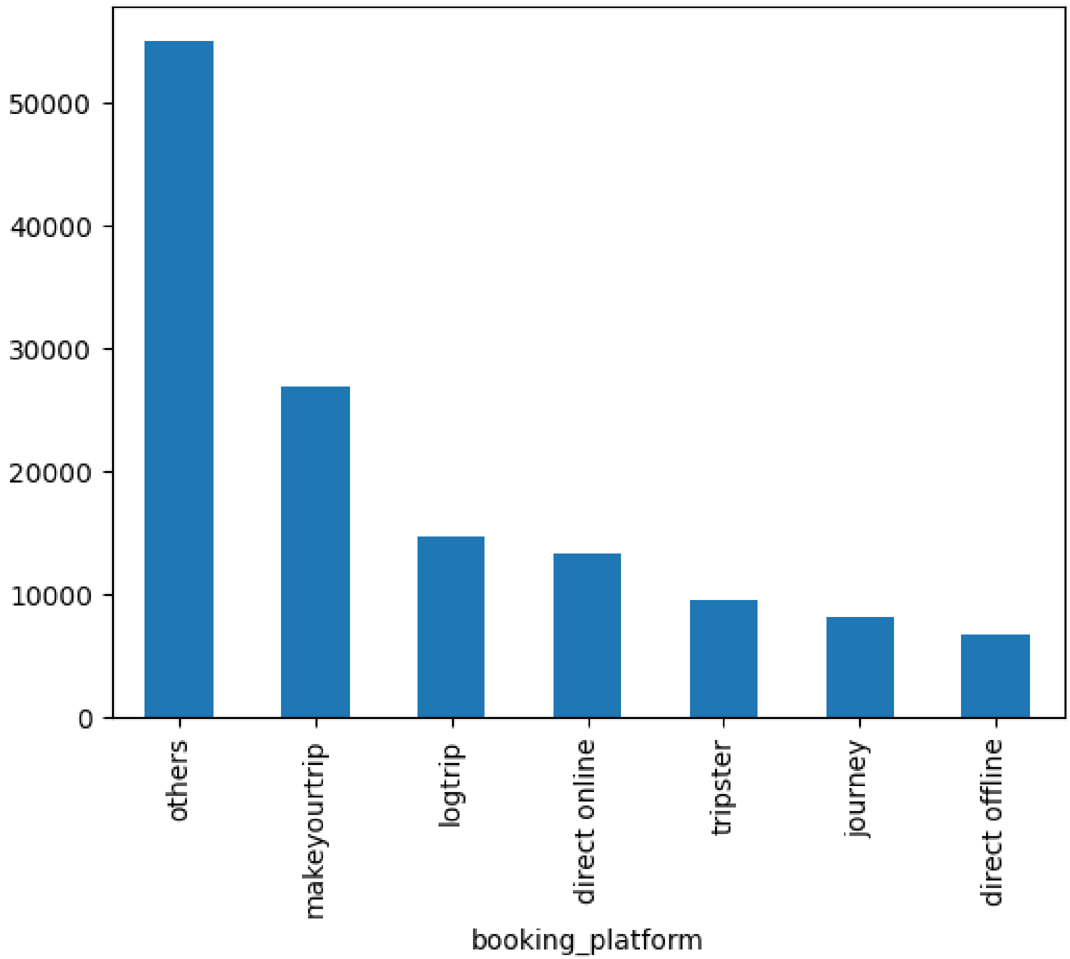
[5]:
```
array(['direct online', 'others', 'logtrip', 'tripster', 'makeyourtrip',
       'journey', 'direct offline'], dtype=object)
```

[6]:
```
df_bookings.booking_platform.value_counts()
```

[6]:
```
booking_platform
others            55066
makeyourtrip      26898
logtrip           14756
direct online     13379
tripster           9630
journey            8106
direct offline     6755
Name: count, dtype: int64
```

[7]:
```
df_bookings.booking_platform.value_counts().plot(kind="bar")
```

[7]:
```
<Axes: xlabel='booking_platform'>
```



[8]:
```
df_bookings.describe()
```

```
[8]:          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
```

```
[9]: df_date = pd.read_csv("datasets/dim_date.csv")
     df_hotels = pd.read_csv("datasets/dim_hotels.csv")
     df_rooms = pd.read_csv("datasets/dim_rooms.csv")
     df_agg_bookings = pd.read_csv("datasets/fact_aggregated_bookings.csv")
```

```
[10]: df_hotels.shape
```

```
[10]: (25, 4)
```

```
[11]: df_hotels.head
```

```
[11]: <bound method NDFrame.head of     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
      5         16563   Atliq Palace  Business      Delhi
      6         17558   Atliq Grands    Luxury     Mumbai
      7         17559  Atliq Exotica    Luxury     Mumbai
      8         17560     Atliq City  Business     Mumbai
      9         17561      Atliq Blu    Luxury     Mumbai
      10        17562      Atliq Bay    Luxury     Mumbai
      11        17563   Atliq Palace  Business     Mumbai
      12        18558   Atliq Grands    Luxury  Hyderabad
      13        18559  Atliq Exotica    Luxury  Hyderabad
```

```
14         18560     Atliq City  Business  Hyderabad
15         18561      Atliq Blu    Luxury  Hyderabad
16         18562      Atliq Bay    Luxury  Hyderabad
17         18563   Atliq Palace  Business  Hyderabad
18         19558   Atliq Grands    Luxury  Bangalore
19         19559  Atliq Exotica    Luxury  Bangalore
20         19560     Atliq City  Business  Bangalore
21         19561      Atliq Blu    Luxury  Bangalore
22         19562      Atliq Bay    Luxury  Bangalore
23         19563   Atliq Palace  Business  Bangalore
24         17564  Atliq Seasons  Business     Mumbai>
```

```
[12]: df_hotels.category.value_counts()
```

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

```
[13]: df_hotels.city.value_counts().sort_values()
```

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

DATA CLEANING

```
[14]: df_bookings.describe()
```

```
[14]:          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
```

[15]:

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

[15]:

```
                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  
```

[16]:

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

[16]:

```
                booking_id  property_id booking_date check_in_date  \
1         May012216558RT12        16558     30-04-22      1/5/2022   
2         May012216558RT13        16558     28-04-22      1/5/2022   
4         May012216558RT15        16558     27-04-22      1/5/2022   
```

```
5        May012216558RT16        16558     1/5/2022      1/5/2022
6        May012216558RT17        16558     28-04-22      1/5/2022
…                   …          …          …           …
134584   Jul312217564RT45        17564     30-07-22      31-07-22
134585   Jul312217564RT46        17564     29-07-22      31-07-22
134587   Jul312217564RT48        17564     30-07-22      31-07-22
134588   Jul312217564RT49        17564     29-07-22      31-07-22
134589  Jul312217564RT410        17564     31-07-22      31-07-22

       checkout_date  no_guests room_category booking_platform  ratings_given  \
1           2/5/2022        2.0           RT1           others            NaN
2           4/5/2022        2.0           RT1          logtrip            5.0
4           2/5/2022        4.0           RT1   direct online             5.0
5           3/5/2022        2.0           RT1           others            4.0
6           6/5/2022        2.0           RT1           others            NaN
…              …        …           …              …             …
134584      1/8/2022        2.0           RT4           others            2.0
134585      3/8/2022        1.0           RT4     makeyourtrip            2.0
134587      2/8/2022        1.0           RT4         tripster            NaN
134588      1/8/2022        2.0           RT4          logtrip            2.0
134589      1/8/2022        2.0           RT4     makeyourtrip            NaN

       booking_status  revenue_generated  revenue_realized
1           Cancelled               9100              3640
2         Checked Out            9100000              9100
4         Checked Out              10920             10920
5         Checked Out               9100              9100
6           Cancelled               9100              3640
…                …                …               …
134584    Checked Out              32300             32300
134585    Checked Out              32300             32300
134587      Cancelled              32300             12920
134588    Checked Out              32300             32300
134589      Cancelled              32300             12920

[134578 rows x 12 columns]
```

[17]:
```
df_bookings.revenue_generated.min(),df_bookings.revenue_generated.max()
```

[17]:
```
(6500, 28560000)
```

[28]:
```
avg, std = df_bookings.revenue_generated.mean(), df_bookings.revenue_generated.
 ↪std()
```

[29]:
```
avg,std
```

[29]:
```
(15378.036937686695, 93040.1549314641)
```

[31]:
```
higher_limit = avg+3*std
higher_limit
```

[31]:
```
294498.50173207896
```

[32]:
```
lower_limit = avg-3*std
lower_limit
```

[32]:
```
-263742.4278567056
```

[33]:
```
df_bookings[df_bookings.revenue_generated<0]
```

[33]:
```
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]
Index: []
```

[34]:
```
df_bookings[df_bookings.revenue_generated>higher_limit]
```

[34]:
```
               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
```

[35]:
```
df_bookings[df_bookings.revenue_generated<higher_limit]
```

[35]:
```
               booking_id  property_id booking_date check_in_date  \
1        May012216558RT12        16558     30-04-22      1/5/2022
4        May012216558RT15        16558     27-04-22      1/5/2022
5        May012216558RT16        16558     1/5/2022      1/5/2022
```

```
6        May012216558RT17       16558     28-04-22      1/5/2022
7        May012216558RT18       16558     26-04-22      1/5/2022
…                     …          …          …            …
134584   Jul312217564RT45       17564     30-07-22      31-07-22
134585   Jul312217564RT46       17564     29-07-22      31-07-22
134587   Jul312217564RT48       17564     30-07-22      31-07-22
134588   Jul312217564RT49       17564     29-07-22      31-07-22
134589  Jul312217564RT410       17564     31-07-22      31-07-22

       checkout_date  no_guests room_category booking_platform  ratings_given  \
1           2/5/2022        2.0           RT1           others            NaN
4           2/5/2022        4.0           RT1    direct online            5.0
5           3/5/2022        2.0           RT1           others            4.0
6           6/5/2022        2.0           RT1           others            NaN
7           3/5/2022        2.0           RT1          logtrip            NaN
…                …          …             …                …              …
134584      1/8/2022        2.0           RT4           others            2.0
134585      3/8/2022        1.0           RT4     makeyourtrip            2.0
134587      2/8/2022        1.0           RT4         tripster            NaN
134588      1/8/2022        2.0           RT4          logtrip            2.0
134589      1/8/2022        2.0           RT4     makeyourtrip            NaN

       booking_status  revenue_generated  revenue_realized
1           Cancelled               9100              3640
4         Checked Out              10920             10920
5         Checked Out               9100              9100
6           Cancelled               9100              3640
7             No Show               9100              9100
…                 …                  …                …
134584    Checked Out              32300             32300
134585    Checked Out              32300             32300
134587      Cancelled              32300             12920
134588    Checked Out              32300             32300
134589      Cancelled              32300             12920

[134573 rows x 12 columns]
```

[36]:

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

[36]:

```
count    134578.000000
mean      12696.011822
std        6927.841641
min        2600.000000
25%        7600.000000
50%       11700.000000
75%       15300.000000
max       45220.000000
```

```
Name: revenue_realized, dtype: float64
```

[37]:

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

[37]:

```
33479.53674501214
```

[38]:

```
df_rooms
```

[38]:

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

[39]:

```
df_bookings
```

[39]:

```
               booking_id  property_id booking_date check_in_date  \
1         May012216558RT12        16558     30-04-22      1/5/2022
2         May012216558RT13        16558     28-04-22      1/5/2022
4         May012216558RT15        16558     27-04-22      1/5/2022
5         May012216558RT16        16558     1/5/2022      1/5/2022
6         May012216558RT17        16558     28-04-22      1/5/2022
…                      …          …          …            …
134584   Jul312217564RT45        17564     30-07-22      31-07-22
134585   Jul312217564RT46        17564     29-07-22      31-07-22
134587   Jul312217564RT48        17564     30-07-22      31-07-22
134588   Jul312217564RT49        17564     29-07-22      31-07-22
134589  Jul312217564RT410        17564     31-07-22      31-07-22

       checkout_date  no_guests room_category booking_platform  ratings_given  \
1           2/5/2022        2.0           RT1           others            NaN
2           4/5/2022        2.0           RT1          logtrip            5.0
4           2/5/2022        4.0           RT1    direct online            5.0
5           3/5/2022        2.0           RT1           others            4.0
6           6/5/2022        2.0           RT1           others            NaN
…                …        …           …                …              …
134584      1/8/2022        2.0           RT4           others            2.0
134585      3/8/2022        1.0           RT4      makeyourtrip            2.0
134587      2/8/2022        1.0           RT4         tripster            NaN
134588      1/8/2022        2.0           RT4          logtrip            2.0
134589      1/8/2022        2.0           RT4      makeyourtrip            NaN

       booking_status  revenue_generated  revenue_realized
1           Cancelled               9100              3640
2         Checked Out            9100000              9100
```

```
4          Checked Out              10920            10920
5          Checked Out               9100             9100
6            Cancelled               9100             3640
…               …                 …                …
134584     Checked Out              32300            32300
134585     Checked Out              32300            32300
134587       Cancelled              32300            12920
134588     Checked Out              32300            32300
134589       Cancelled              32300            12920

[134578 rows x 12 columns]
```

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

```
[40]: 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
```

```
[41]: 23439.308444+ 3*9048.599076
```

```
[41]: 50585.105672000005
```

We can see that max revenue_realized is smaller then the 3std from avg(mean) calculation.So, There wil be no outliers

DATA TRANSFORMATION

```
[42]: df_agg_bookings.head()
```

```
[42]:    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
```

```
[43]: df_agg_bookings["occ_pct"]=df_agg_bookings["successful_bookings"]/
       ↪df_agg_bookings["capacity"]
```

```
[45]: df_agg_bookings.head()
```

[45]:
```
   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
```

[46]:
```
df_agg_bookings["occ_pct"]=df_agg_bookings["occ_pct"].apply(lambda x: round(x*100,2))
```

[47]:
```
df_agg_bookings.head()
```

[47]:
```
   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
```

There is 4 types of Data Transformation 1. Creating new Column 2. Normalization 3. Merge 4. Aggregation

INSIGHT GENERATION

Q1. What is an average occupancy rate in each of the room categories?

[49]:
```
df_agg_bookings.groupby("room_category")["occ_pct"].mean()
```

[49]:
```
room_category
RT1    58.224247
RT2    58.040278
RT3    58.028213
RT4    59.300461
Name: occ_pct, dtype: float64
```

```
[50]: df_rooms
```

```
[50]:   room_id    room_class
      0     RT1      Standard
      1     RT2         Elite
      2     RT3       Premium
      3     RT4  Presidential
```

Now Merge the df_rooms and df_agg_bookings.

```
[52]: df= pd.merge(df_agg_bookings,df_rooms, left_on= "room_category", right_on =␣
       ↪"room_id")
      df.tail(4)
```

```
[52]:       property_id check_in_date room_category  successful_bookings  capacity  \
      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
      9196    72.22     RT4  Presidential
      9197    50.00     RT4  Presidential
      9198    50.00     RT4  Presidential
      9199    75.00     RT4  Presidential
```

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

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

Q2. Print average occupancy Rate per city.

```
[58]: df_hotels.head(4)
```

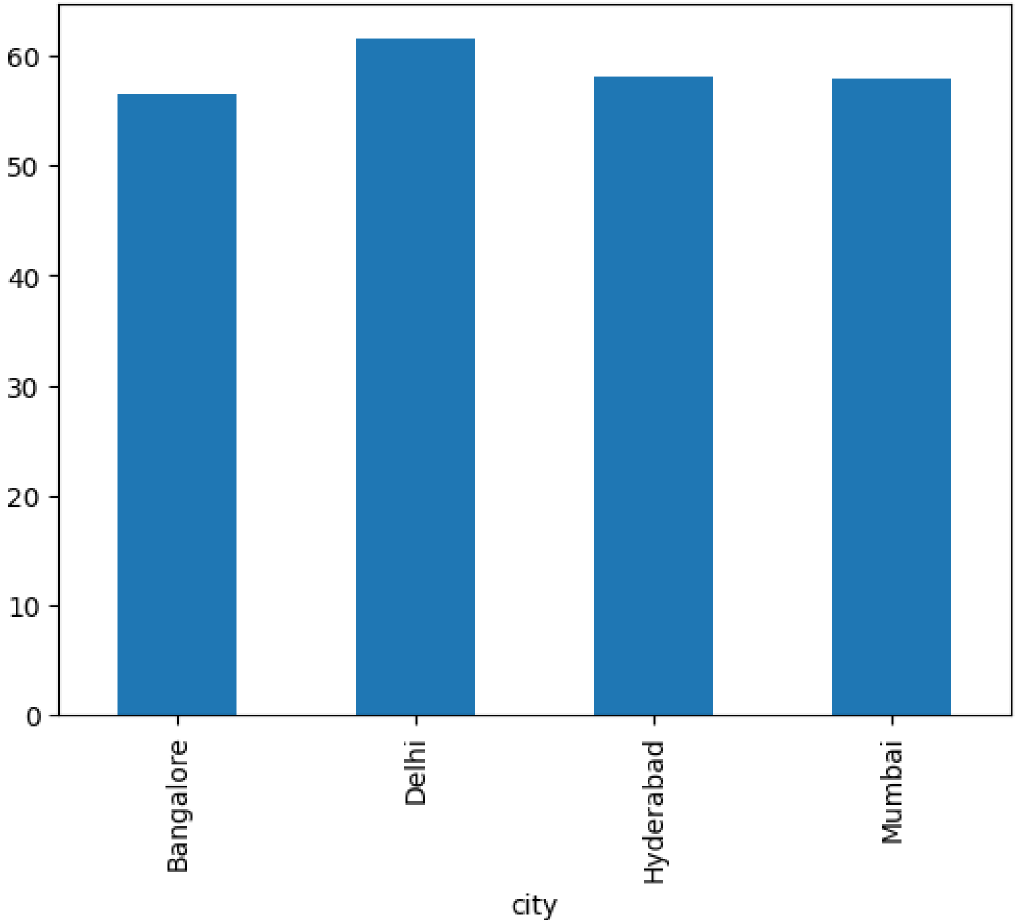
```
[58]:    property_id  property_name  category    city
      0        16558   Atliq Grands    Luxury   Delhi
      1        16559  Atliq Exotica    Luxury  Mumbai
      2        16560     Atliq City  Business   Delhi
      3        16561      Atliq Blu    Luxury   Delhi
```

```
[60]: df= pd.merge(df,df_hotels, on="property_id")
      df.head(4)
```

[60]:
```
   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

   occ_pct room_id room_class  property_name category    city
0    83.33     RT1   Standard  Atliq Exotica   Luxury  Mumbai
1    66.67     RT1   Standard  Atliq Exotica   Luxury  Mumbai
2    56.67     RT1   Standard  Atliq Exotica   Luxury  Mumbai
3    70.00     RT1   Standard  Atliq Exotica   Luxury  Mumbai
```

[61]:
```
df.groupby("city")["occ_pct"].mean().plot(kind="bar")
```

[61]:
```
<Axes: xlabel='city'>
```



[62]:
```
df.groupby("city")["occ_pct"].mean()
```

```
[62]: city
      Bangalore    56.594207
      Delhi        61.606467
      Hyderabad    58.144651
      Mumbai       57.936305
      Name: occ_pct, dtype: float64
```

```
[ ]:
```

Q3.When was the occupancy better? Weekday or Weekend?

```
[64]: df.head(3)
```

```
[64]:    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

         occ_pct room_id room_class  property_name category    city
      0    83.33     RT1   Standard  Atliq Exotica   Luxury  Mumbai
      1    66.67     RT1   Standard  Atliq Exotica   Luxury  Mumbai
      2    56.67     RT1   Standard  Atliq Exotica   Luxury  Mumbai
```

```
[66]: df_date.head(4)
```

```
[66]:         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
```

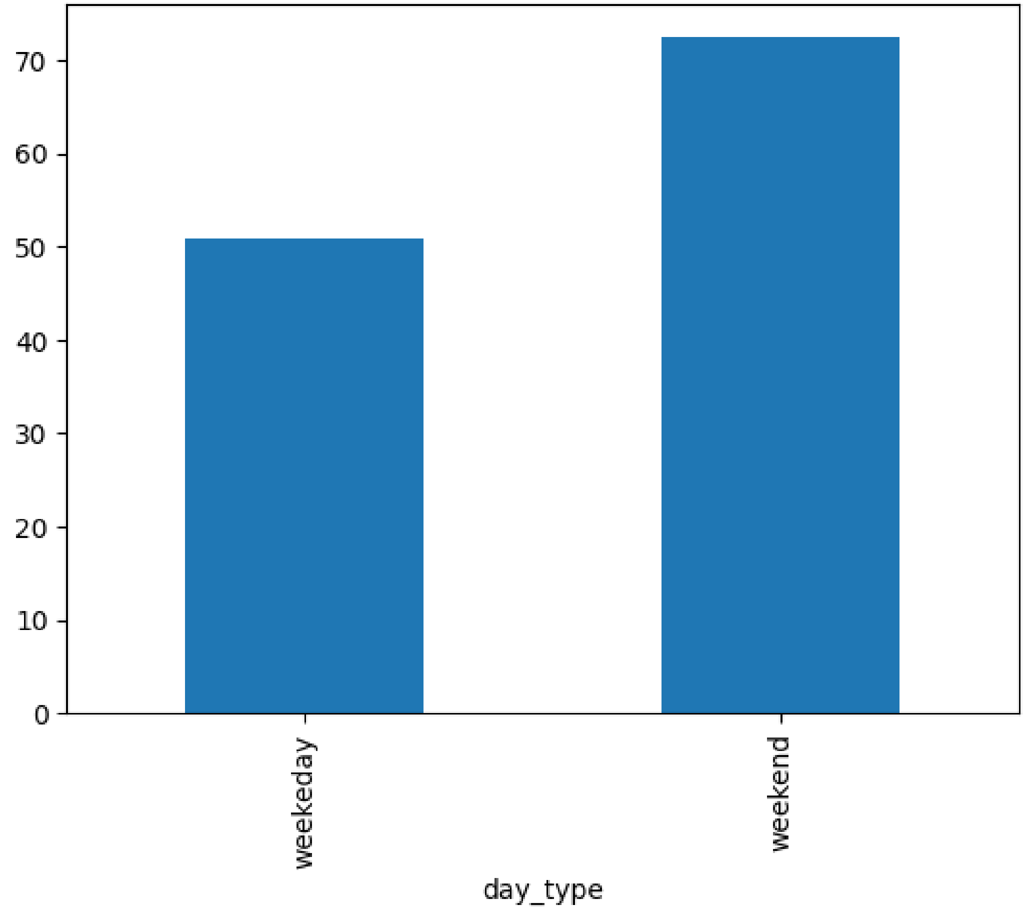
```
[67]: df= pd.merge(df,df_date, left_on= "check_in_date" , right_on = "date")
      df.head(3)
```

```
[67]:    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

         occ_pct room_id room_class  property_name category    city       date  \
      0    60.00     RT1   Standard  Atliq Exotica   Luxury  Mumbai  10-May-22
      1    60.98     RT2      Elite  Atliq Exotica   Luxury  Mumbai  10-May-22
      2    62.50     RT3    Premium  Atliq Exotica   Luxury  Mumbai  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
```

[76]:
```
df.groupby("day_type")["occ_pct"].mean().round(2).plot(kind="bar")
```

[76]:
```
<Axes: xlabel='day_type'>
```



Q4. In the month of june, What is the occupancy for diferent cities

[78]:
```
df["mmm yy"].unique()
```

[78]:
```
array(['May 22', 'Jun 22', 'Jul 22'], dtype=object)
```

[80]:
```
df_jun_22= df[df["mmm yy"]=="Jun 22"]
```

[81]:
```
df_jun_22
```

[81]:
```
      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
…           …           …           …                  …       …
4295        17564     30-Jun-22           RT4                      7      17.0
4296        18560     30-Jun-22           RT1                     18      30.0
4297        18560     30-Jun-22           RT2                     24      40.0
4298        18560     30-Jun-22           RT3                     14      24.0
4299        18560     30-Jun-22           RT4                      8      15.0

      occ_pct room_id    room_class  property_name  category       city  \
2200    66.67     RT1      Standard  Atliq Exotica    Luxury     Mumbai
2201    63.41     RT2         Elite  Atliq Exotica    Luxury     Mumbai
2202    62.50     RT3       Premium  Atliq Exotica    Luxury     Mumbai
2203    61.11     RT4  Presidential  Atliq Exotica    Luxury     Mumbai
2204    63.33     RT1      Standard      Atliq Bay    Luxury  Bangalore
…       …     …           …            …       …        …
4295    41.18     RT4  Presidential  Atliq Seasons  Business     Mumbai
4296    60.00     RT1      Standard     Atliq City  Business  Hyderabad
4297    60.00     RT2         Elite     Atliq City  Business  Hyderabad
4298    58.33     RT3       Premium     Atliq City  Business  Hyderabad
4299    53.33     RT4  Presidential     Atliq City  Business  Hyderabad

           date  mmm yy week no  day_type
2200  10-Jun-22  Jun 22    W 24  weekeday
2201  10-Jun-22  Jun 22    W 24  weekeday
2202  10-Jun-22  Jun 22    W 24  weekeday
2203  10-Jun-22  Jun 22    W 24  weekeday
2204  10-Jun-22  Jun 22    W 24  weekeday
…          …     …     …       …
4295  30-Jun-22  Jun 22    W 27  weekeday
4296  30-Jun-22  Jun 22    W 27  weekeday
4297  30-Jun-22  Jun 22    W 27  weekeday
4298  30-Jun-22  Jun 22    W 27  weekeday
4299  30-Jun-22  Jun 22    W 27  weekeday

[2100 rows x 15 columns]
```

[86]:

```
df_jun_22.groupby("city")["occ_pct"].mean().round(2).
  ↪sort_values(ascending=False)
```

[86]:

```
city
Delhi        62.47
Hyderabad    58.46
Mumbai       58.38
Bangalore    56.58
```

```
Name: occ_pct, dtype: float64
```

[89]:

```
df_data_august = pd.read_csv("datasets/new_data_august.csv")
df_data_august.head(3)
```

[89]:

```
   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

  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

     occ%
0  100.00
1   70.00
2   76.67
```

[90]:

```
df_data_august.columns
```

[90]:

```
Index(['property_id', 'property_name', 'category', 'city', 'room_category',
       'room_class', 'check_in_date', 'mmm yy', 'week no', 'day_type',
       'successful_bookings', 'capacity', 'occ%'],
      dtype='object')
```

[91]:

```
df.columns
```

[91]:

```
Index(['property_id', 'check_in_date', 'room_category', 'successful_bookings',
       'capacity', 'occ_pct', 'room_id', 'room_class', 'property_name',
       'category', 'city', 'date', 'mmm yy', 'week no', 'day_type'],
      dtype='object')
```

[92]:

```
df_data_august.shape
```

[92]:

```
(7, 13)
```

[93]:

```
df.shape
```

[93]:

```
(6500, 15)
```

[95]:

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

[95]:

```
      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_id    room_class property_name  category       city  \
6497    85.00     RT2         Elite    Atliq City  Business  Hyderabad
6498    70.83     RT3       Premium    Atliq City  Business  Hyderabad
6499    80.00     RT4  Presidential    Atliq City  Business  Hyderabad
6500      NaN     NaN      Standard  Atliq Exotica   Luxury     Mumbai
6501      NaN     NaN      Standard     Atliq Bay    Luxury  Bangalore
6502      NaN     NaN      Standard  Atliq Palace  Business  Bangalore
6503      NaN     NaN      Standard  Atliq Grands    Luxury  Bangalore
6504      NaN     NaN      Standard    Atliq City  Business  Bangalore
6505      NaN     NaN      Standard     Atliq Blu    Luxury     Mumbai
6506      NaN     NaN      Standard Atliq Seasons  Business     Mumbai

           date  mmm yy week no  day_type    occ%
6497  31-Jul-22  Jul 22    W 32   weekend     NaN
6498  31-Jul-22  Jul 22    W 32   weekend     NaN
6499  31-Jul-22  Jul 22    W 32   weekend     NaN
6500        NaN  Aug-22    W 32  weekeday  100.00
6501        NaN  Aug-22    W 32  weekeday   70.00
6502        NaN  Aug-22    W 32  weekeday   76.67
6503        NaN  Aug-22    W 32  weekeday   75.00
6504        NaN  Aug-22    W 32  weekeday   76.92
6505        NaN  Aug-22    W 32  weekeday   69.23
6506        NaN  Aug-22    W 32  weekeday   62.50
```

[96]:
```
latest_df.shape
```

[96]:
```
(6507, 16)
```

Q6.Print Revenue Realised per city

[98]:
```
df_bookings.head(4)
```

[98]:
```
          booking_id  property_id booking_date check_in_date checkout_date  \
1  May012216558RT12        16558     30-04-22      1/5/2022      2/5/2022
2  May012216558RT13        16558     28-04-22      1/5/2022      4/5/2022
4  May012216558RT15        16558     27-04-22      1/5/2022      2/5/2022
5  May012216558RT16        16558     1/5/2022      1/5/2022      3/5/2022
```

```
   no_guests room_category booking_platform  ratings_given booking_status  \
1        2.0           RT1           others            NaN      Cancelled
2        2.0           RT1          logtrip            5.0    Checked Out
4        4.0           RT1    direct online            5.0    Checked Out
5        2.0           RT1           others            4.0    Checked Out

   revenue_generated  revenue_realized
1               9100              3640
2            9100000              9100
4              10920             10920
5               9100              9100
```

[100]:
```
df_hotels.head(4)
```

[100]:
```
   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
```

[102]:
```
df_bookings_all=pd.merge(df_bookings,df_hotels, on="property_id")
```

[104]:
```
df_bookings_all.head(3)
```

[104]:
```
         booking_id  property_id booking_date check_in_date checkout_date  \
0  May012216558RT12        16558     30-04-22      1/5/2022      2/5/2022
1  May012216558RT13        16558     28-04-22      1/5/2022      4/5/2022
2  May012216558RT15        16558     27-04-22      1/5/2022      2/5/2022

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

   revenue_generated  revenue_realized property_name category   city
0               9100              3640  Atliq Grands   Luxury  Delhi
1            9100000              9100  Atliq Grands   Luxury  Delhi
2              10920             10920  Atliq Grands   Luxury  Delhi
```

[108]:
```
df_bookings_all.groupby("city")["revenue_realized"].sum()
```

[108]:
```
city
Bangalore    420383550
Delhi        294438788
Hyderabad    325179310
Mumbai       668602231
```

```
Name: revenue_realized, dtype: int64
```

Q7.Print month by month revenue

```
[110]: df_bookings_all.head(3)
```

```
[110]:         booking_id  property_id booking_date check_in_date checkout_date  \
       0  May012216558RT12        16558     30-04-22      1/5/2022      2/5/2022
       1  May012216558RT13        16558     28-04-22      1/5/2022      4/5/2022
       2  May012216558RT15        16558     27-04-22      1/5/2022      2/5/2022

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

          revenue_generated  revenue_realized property_name category   city
       0               9100              3640  Atliq Grands   Luxury  Delhi
       1            9100000              9100  Atliq Grands   Luxury  Delhi
       2              10920             10920  Atliq Grands   Luxury  Delhi
```

```
[111]: df_date.head(3)
```

```
[111]:         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
```

```
[113]: df_bookings_all.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134578 entries, 0 to 134577
Data columns (total 15 columns):
 #   Column             Non-Null Count   Dtype
---  ------             --------------   -----
 0   booking_id         134578 non-null  object
 1   property_id        134578 non-null  int64
 2   booking_date       134578 non-null  object
 3   check_in_date      134578 non-null  object
 4   checkout_date      134578 non-null  object
 5   no_guests          134578 non-null  float64
 6   room_category      134578 non-null  object
 7   booking_platform   134578 non-null  object
 8   ratings_given      56679 non-null   float64
 9   booking_status     134578 non-null  object
 10  revenue_generated  134578 non-null  int64
 11  revenue_realized   134578 non-null  int64
 12  property_name      134578 non-null  object
 13  category           134578 non-null  object
 14  city               134578 non-null  object
dtypes: float64(2), int64(3), object(10)
memory usage: 15.4+ MB
```

[114]:

```python
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
```

[152]:

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

[153]:

```python
df_date.head()
```

[153]:

```
        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
```

[154]:

```python
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
```

[155]:

```python
df_bookings_all["check_in_date"] = pd.
 ↪to_datetime(df_bookings_all["check_in_date"], errors='coerce')
```

[150]:

```python
df_bookings_all.head()
```

```
[150]:         booking_id  property_id booking_date check_in_date checkout_date  \
       0  May012216558RT12        16558     30-04-22    2022-01-05      2/5/2022
       1  May012216558RT13        16558     28-04-22    2022-01-05      4/5/2022
       2  May012216558RT15        16558     27-04-22    2022-01-05      2/5/2022
       3  May012216558RT16        16558     1/5/2022    2022-01-05      3/5/2022
       4  May012216558RT17        16558     28-04-22    2022-01-05      6/5/2022

          no_guests room_category booking_platform  ratings_given booking_status  \
       0        2.0           RT1           others            NaN      Cancelled
       1        2.0           RT1          logtrip            5.0    Checked Out
       2        4.0           RT1   direct online             5.0    Checked Out
       3        2.0           RT1           others            4.0    Checked Out
       4        2.0           RT1           others            NaN      Cancelled

          revenue_generated  revenue_realized property_name category  city
       0               9100              3640  Atliq Grands   Luxury  Delhi
       1            9100000              9100  Atliq Grands   Luxury  Delhi
       2              10920             10920  Atliq Grands   Luxury  Delhi
       3               9100              9100  Atliq Grands   Luxury  Delhi
       4               9100              3640  Atliq Grands   Luxury  Delhi
```

```
[157]: df_bookings_all= pd.
        ↪merge(df_bookings_all,df_date,left_on="check_in_date",right_on="date")
```

```
[158]: df_bookings_all.head(4)
```

```
[158]:         booking_id  property_id booking_date check_in_date checkout_date  \
       0  May052216558RT11        16558     15-04-22    2022-05-05      7/5/2022
       1  May052216558RT12        16558     30-04-22    2022-05-05      7/5/2022
       2  May052216558RT13        16558     1/5/2022    2022-05-05      6/5/2022
       3  May052216558RT14        16558     3/5/2022    2022-05-05      6/5/2022

          no_guests room_category booking_platform  ratings_given booking_status  \
       0        3.0           RT1         tripster            5.0    Checked Out
       1        2.0           RT1           others            NaN      Cancelled
       2        3.0           RT1  direct offline             5.0    Checked Out
       3        2.0           RT1         tripster            3.0    Checked Out

          revenue_generated  revenue_realized property_name category  city  \
       0              10010             10010  Atliq Grands   Luxury  Delhi
       1               9100              3640  Atliq Grands   Luxury  Delhi
       2              10010             10010  Atliq Grands   Luxury  Delhi
       3               9100              9100  Atliq Grands   Luxury  Delhi

                date  mmm yy week no  day_type
       0 2022-05-05  May 22    W 19  weekeday
       1 2022-05-05  May 22    W 19  weekeday
```

```
2 2022-05-05  May 22    W 19  weekeday
3 2022-05-05  May 22    W 19  weekeday
```

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

```
[160]: mmm yy
       Jul 22    60278496
       Jun 22    52903014
       May 22    60961428
       Name: revenue_realized, dtype: int64
```

Q8.Print Revenue Realised per hotel type

```
[162]: df_bookings.head(3)
```

```
[162]:          booking_id  property_id booking_date check_in_date checkout_date  \
       1  May012216558RT12        16558     30-04-22      1/5/2022      2/5/2022
       2  May012216558RT13        16558     28-04-22      1/5/2022      4/5/2022
       4  May012216558RT15        16558     27-04-22      1/5/2022      2/5/2022

          no_guests room_category booking_platform  ratings_given booking_status  \
       1        2.0           RT1           others            NaN      Cancelled
       2        2.0           RT1          logtrip            5.0    Checked Out
       4        4.0           RT1    direct online            5.0    Checked Out

          revenue_generated  revenue_realized
       1               9100              3640
       2            9100000              9100
       4              10920             10920
```

```
[163]: df_hotels.head()
```

```
[163]:    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
```

```
[164]: df_bookings_all = pd.merge(df_bookings,df_hotels, on="property_id")
```

```
[166]: df_bookings_all.head(4)
```

```
[166]:          booking_id  property_id booking_date check_in_date checkout_date  \
       0  May012216558RT12        16558     30-04-22      1/5/2022      2/5/2022
       1  May012216558RT13        16558     28-04-22      1/5/2022      4/5/2022
       2  May012216558RT15        16558     27-04-22      1/5/2022      2/5/2022
```

```
3  May012216558RT16        16558     1/5/2022      1/5/2022      3/5/2022

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

   revenue_generated  revenue_realized property_name category   city
0               9100              3640  Atliq Grands   Luxury  Delhi
1            9100000              9100  Atliq Grands   Luxury  Delhi
2              10920             10920  Atliq Grands   Luxury  Delhi
3               9100              9100  Atliq Grands   Luxury  Delhi
```

[168]:

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

[168]:

```
category
Business     655967037
Luxury      1052636842
Name: revenue_realized, dtype: int64
```

[ ]: