hospitality_analysis_pandas

November 5, 2024

[366]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[367]:
```
df_bookings=pd.read_csv("fact_bookings.csv")
```

[368]:
```
import os
```

[369]:
```
current_directory = os.getcwd()
```

[370]:
```
current_directory
```

[370]:
```
'C:\\Users\\adian\\Downloads\\study\\python\\codebasics python\\source-
code\\project_hospitality_analysis\\datasets'
```

[371]:
```
df_bookings.head()
```

[371]:

	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
0	May012216558RT11	16558	27-04-22	1/5/2022	2/5/2022	-3.0	RT1	direct online	1.0	Checked Out	10010	10010
1	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others	NaN	Cancelled	9100	3640
2	May012216558RT13	16558	28-04-22	1/5/2022	4/5/2022	2.0	RT1	logtrip	5.0	Checked Out	9100000	9100
3	May012216558RT14	16558	28-04-22	1/5/2022	2/5/2022	-2.0	RT1	others	NaN	Cancelled	9100	3640
4	May012216558RT15	16558	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online	5.0	Checked Out	10920	10920

[298]:
```
df_bookings.shape
```

[298]:
```
(134590, 12)
```

[372]:
```
x=df_bookings.property_id.unique()
```

[301]:
```
x
```

[301]:
```
array([16558, 16559, 16560, 16561, 16562, 16563, 17558, 17559, 17560,
       17561, 17562, 17563, 18558, 18559, 18560, 18561, 18562, 18563,
       19558, 19559, 19560, 19561, 19562, 19563, 17564], dtype=int64)
```

[373]:
```
len(x)
```

[373]:
```
25
```

[374]:
```
df_bookings.room_category.unique()  # category of rooms
```

[374]:
```
array(['RT1', 'RT2', 'RT3', 'RT4'], dtype=object)
```

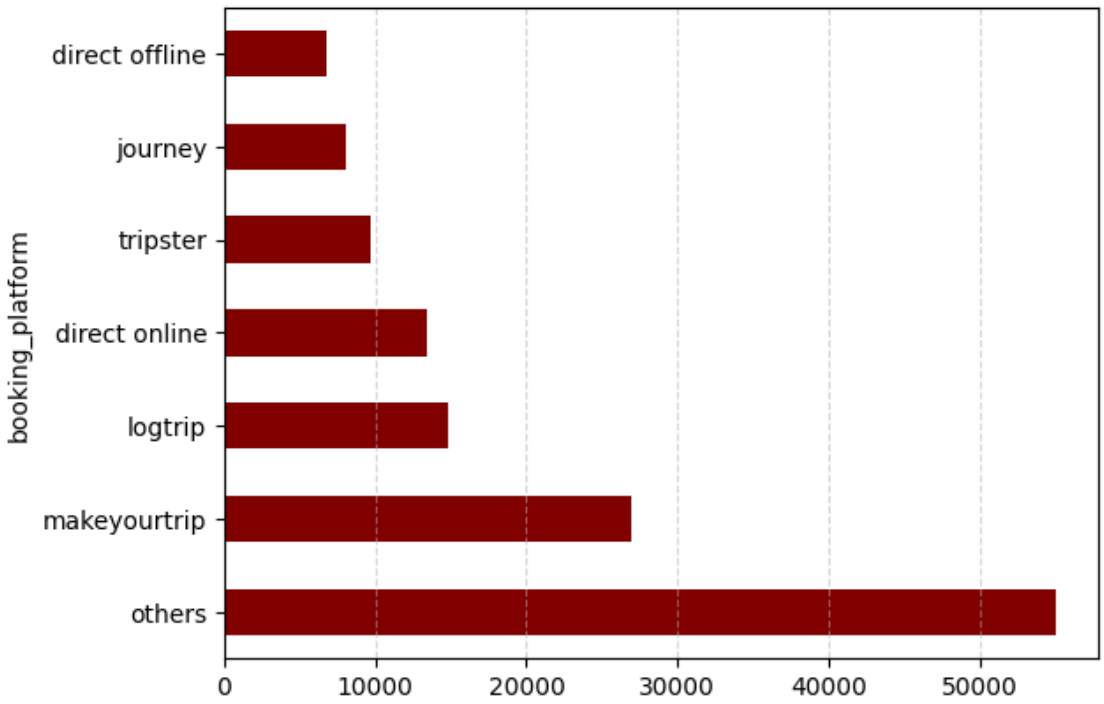
[375]:
```
df_bookings.booking_platform.unique() # booking platforms
```

[375]:
```
array(['direct online', 'others', 'logtrip', 'tripster', 'makeyourtrip',
       'journey', 'direct offline'], dtype=object)
```

[376]:
```
df_bookings.booking_platform.value_counts()  #no of bookings using each platform
```

[376]:
```
booking_platform
others            55066
makeyourtrip      26898
logtrip           14756
direct online     13379
tripster           9630
journey            8106
direct offline     6755
Name: count, dtype: int64
```

[377]:
```
ax=df_bookings.booking_platform.value_counts().plot(kind='barh', color='maroon')
ax.grid(axis='x', linestyle='--', alpha=0.5)
plt.show()
```



[378]:
```
df_bookings.describe()
```

[378]:

	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	7600.000000
50%	17564.000000	2.000000	4.000000	1.350000e+04	11700.000000
75%	18563.000000	2.000000	5.000000	1.800000e+04	15300.000000
max	19563.000000	6.000000	5.000000	2.856000e+07	45220.000000

[379]:
```
df_bookings.revenue_realized.min(),df_bookings.revenue_realized.max()
```

[379]: (2600, 45220)

[380]:
```
df_new_data=pd.read_csv("new_data_august.csv")
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")
```

[381]:
```
df_new_data.head()
```

[381]:

	property_id	property_name	category	city	room_category	room_class	check_in_date	mmm yy	week no	day_type	successful_bookings	capacity	occ%
0	16559	Atliq Exotica	Luxury	Mumbai	RT1	Standard	01-Aug-22	Aug-22	W 32	weekeday	30	30	100.00
1	19562	Atliq Bay	Luxury	Bangalore	RT1	Standard	01-Aug-22	Aug-22	W 32	weekeday	21	30	70.00
2	19563	Atliq Palace	Business	Bangalore	RT1	Standard	01-Aug-22	Aug-22	W 32	weekeday	23	30	76.67
3	19558	Atliq Grands	Luxury	Bangalore	RT1	Standard	01-Aug-22	Aug-22	W 32	weekeday	30	40	75.00
4	19560	Atliq City	Business	Bangalore	RT1	Standard	01-Aug-22	Aug-22	W 32	weekeday	20	26	76.92

[382]:
```
df_date.head()
```

[382]:

	date	mmm yy	week no	day_type
0	1-May-22	22-May	W 19	weekend
1	2-May-22	22-May	W 19	weekeday
2	3-May-22	22-May	W 19	weekeday
3	4-May-22	22-May	W 19	weekeday
4	5-May-22	22-May	W 19	weekeday

[383]:
```
df_hotels.head()
```

[383]:

	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

[312]:
```
df_rooms.head()
```

[312]:
```
  room_id    room_class
0     RT1      Standard
1     RT2         Elite
2     RT3       Premium
3     RT4  Presidential
```

[313]:
```
df_agg_bookings.head()
```

[313]:
```
   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
```

[384]:
```
df_agg_bookings.describe()
```

[384]:
```
         property_id  successful_bookings     capacity
count   9200.000000          9200.000000  9198.000000
mean   18040.640000            14.655761    25.280496
std     1099.818325             7.736170    11.442080
min    16558.000000             1.000000     3.000000
25%    17558.000000             9.000000    18.000000
50%    17564.000000            14.000000    25.000000
75%    18563.000000            19.000000    34.000000
max    19563.000000           123.000000    50.000000
```

[385]:
```
df_agg_bookings.property_id.unique()
```

[385]:
```
array([16559, 19562, 19563, 17558, 16558, 17560, 19558, 19560, 17561,
       16560, 16561, 16562, 16563, 17559, 17562, 17563, 18558, 18559,
       18561, 18562, 18563, 19559, 19561, 17564, 18560], dtype=int64)
```

[386]:
```
df_agg_bookings.groupby('property_id')["successful_bookings"].sum()
```

[386]:
```
property_id
16558    3153
16559    7338
16560    4693
```

```
16561    4418
16562    4820
16563    7211
17558    5053
17559    6142
17560    6013
17561    5183
17562    3424
17563    6337
17564    3982
18558    4475
18559    5256
18560    6638
18561    6458
18562    7333
18563    4737
19558    4400
19559    4729
19560    6079
19561    5736
19562    5812
19563    5413
Name: successful_bookings, dtype: int64
```

[387]:
```
df_agg_bookings[df_agg_bookings.successful_bookings>df_agg_bookings.
↪capacity]['check_in_date']
```

[387]:
```
3         1-May-22
12        1-May-22
4136     11-Jun-22
6209      2-Jul-22
8522     25-Jul-22
9194     31-Jul-22
Name: check_in_date, dtype: object
```

[388]:
```
df_agg_bookings.capacity.max()
```

[388]:
```
50.0
```

[389]:
```
df_agg_bookings[df_agg_bookings.capacity==df_agg_bookings.capacity.
↪max()]['property_id'].unique()
```

[389]:
```
array([17558], dtype=int64)
```

[65]:
```
# data cleaning
```

[390]:
```
df_bookings.describe()
```

[390]:
```
          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
```

[321]:
```
df_bookings.head()
```

[321]:
```
          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
4  May012216558RT15        16558     27-04-22      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
4        4.0           RT1    direct online            5.0    Checked Out

   revenue_generated  revenue_realized
0              10010             10010
1               9100              3640
2            9100000              9100
3               9100              3640
4              10920             10920
```

[391]:
```
df_bookings[df_bookings.no_guests<=0]
```

[391]:
```
              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
```

[49]:
```
# no_guests can't be less than zero. So, these are wrong values. We can't␣
↪ignore these bookings as these have created revenues.
```

[323]:
```
df_bookings.shape
```

[323]:
```
(134590, 12)
```

[392]:
```
df_bookings = df_bookings[df_bookings.no_guests>0]
```

[70]:
```
df_bookings.shape    # we discarded 12 records only
```

[70]:
```
(134578, 12)
```

[393]:
```
df_bookings.count()
```

[393]:
```
booking_id           134578
property_id          134578
booking_date         134578
check_in_date        134578
checkout_date        134578
no_guests            134578
room_category        134578
booking_platform     134578
ratings_given         56679
booking_status       134578
revenue_generated    134578
revenue_realized     134578
dtype: int64
```

[394]:
```
df_bookings.revenue_generated.min(),df_bookings.revenue_generated.max()
```

[394]:
```
(6500, 28560000)
```

[82]:
```
# 28560000 this amount from one booking is not possible(this is an outlier- 3␣
↪standard deviation from mean). We will remove such values.
```

[395]:
```
avg,sd=df_bookings.revenue_generated.mean(),df_bookings.revenue_generated.std()
```

[396]:
```
avg,sd
```

[396]:
```
(15378.036937686695, 93040.1549314641)
```

[397]:
```
higher_lim=avg+3*sd
lower_lim=avg-3*sd
```

[398]:
```
higher_lim,lower_lim
```

[398]:
```
(294498.50173207896, -263742.4278567056)
```

[83]:
```
# So, Any revenue_generated value greater than 294498 is an outlier. We will␣
↪remove these values.
```

[399]:
```
df_bookings=df_bookings[df_bookings.revenue_generated<higher_lim]
```

[85]:
```
df_bookings.shape
```

[85]:
```
(134573, 12)
```

[86]:
```
# 5 outliers got removed
```

[400]:
```
df_bookings.revenue_realized.describe()
```

[400]:
```
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
```

[401]:
```
higher_limit = df_bookings.revenue_realized.mean() + 3*df_bookings.
↪revenue_realized.std()
higher_limit
```

[401]:
```
33479.358661845814
```

[ ]:
```
# max revenue generated(45220)> higher_limit(33479.35). Let's dig a bit more to␣
↪know if these values are outlier or not
```

[402]:
```
df_bookings[df_bookings.revenue_realized>33479]
```

[402]:

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_given
137	May012216559RT41	16559	27-04-22	1/5/2022	7/5/2022	4.0	RT4	others	NaN
139	May012216559RT43	16559	1/5/2022	1/5/2022	2/5/2022	6.0	RT4	tripster	3.0
143	May012216559RT47	16559	28-04-22	1/5/2022	3/5/2022	3.0	RT4	others	5.0
149	May012216559RT413	16559	24-04-22	1/5/2022	7/5/2022	5.0	RT4	logtrip	NaN
222	May012216560RT45	16560	30-04-22	1/5/2022	3/5/2022	5.0	RT4	others	3.0
…	…	…	…	…	…	…	…	…	…
134328	Jul312219560RT49	19560	31-07-22	31-07-22	2/8/2022	6.0	RT4	direct online	5.0
134331	Jul312219560RT412	19560	31-07-22	31-07-22	1/8/2022	6.0	RT4	others	2.0
134467	Jul312219562RT45	19562	28-07-22	31-07-22	1/8/2022	6.0	RT4	makeyourtrip	4.0
134474	Jul312219562RT412	19562	25-07-22	31-07-22	6/8/2022	5.0	RT4	direct offline	5.0
134581	Jul312217564RT42	17564	31-07-22	31-07-22	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]
```

[93]:
```
# We got 1299 such values and the common thing is "Room_category= RT4" , Let's
↪know more about RT4
```

[403]:
```
df_rooms
```

[403]:
```
  room_id    room_class
0     RT1      Standard
1     RT2         Elite
2     RT3       Premium
3     RT4  Presidential
```

[95]:
```
# RT4 is the presidential suite, whose cost can be really high. Let's find
↪outlier for RT4 room_category
```

[404]:
```
df_bookings[df_bookings.room_category=='RT4'].revenue_realized.describe()
```

[404]:
```
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
```

[405]:
```
higher_lim=23439+3*9048
higher_lim
```

[405]:
```
50583
```

[98]:
```
# As max(45220)<50583, no outliers present, no need to discard any values
```

[406]:
```
# null values (no values available)
df_bookings.isnull().sum()
```

[406]:
```
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
```

[ ]:
```
# only ratings given column have NA values which is normal.
```

Data Transformation

[407]:
```
df_agg_bookings.head()
```

[407]:

	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

[408]:
```
df_agg_bookings.isnull().sum()
```

[408]:
```
property_id            0
check_in_date          0
room_category          0
successful_bookings    0
capacity               2
dtype: int64
```

[409]:
```
df_agg_bookings[df_agg_bookings.capacity.isna()]
```

[409]:

	property_id	check_in_date	room_category	successful_bookings	capacity
8	17561	1-May-22	RT1	22	NaN
14	17562	1-May-22	RT1	12	NaN

Average occupancy rate in each category

[165]:
```
#occupancy pct=successful_bookings/capacity
```

[410]:
```
df_agg_bookings["occ_pct"]=df_agg_bookings["successful_bookings"]/
↪df_agg_bookings["capacity"]
```

[343]:
```
df_agg_bookings.head()
```

[343]:

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct
0	16559	1-May-22	RT1	25	30.0	0.833333
1	19562	1-May-22	RT1	28	30.0	0.933333
2	19563	1-May-22	RT1	23	30.0	0.766667
3	17558	1-May-22	RT1	30	19.0	1.578947
4	16558	1-May-22	RT1	18	19.0	0.947368

[411]:
```
df_agg_bookings["occ_pct"]=df_agg_bookings["occ_pct"].apply(lambda x:
↪round(x*100,2))
```

[412]:
```
df_agg_bookings.head()
```

[412]:

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct
0	16559	1-May-22	RT1	25	30.0	83.33
1	19562	1-May-22	RT1	28	30.0	93.33
2	19563	1-May-22	RT1	23	30.0	76.67
3	17558	1-May-22	RT1	30	19.0	157.89
4	16558	1-May-22	RT1	18	19.0	94.74

[413]:
```
df_agg_bookings.groupby('room_category').occ_pct.mean().round(2)
```

[413]:
```
room_category
RT1    58.22
RT2    58.04
RT3    58.03
RT4    59.30
Name: occ_pct, dtype: float64
```

[347]:
```
df_rooms
```

[347]:
```
  room_id    room_class
0     RT1      Standard
1     RT2         Elite
2     RT3       Premium
3     RT4  Presidential
```

[414]:
```
df=pd.merge(df_agg_bookings,df_rooms,left_on="room_category",right_on="room_id")
```

[415]:
```
df.groupby('room_class').occ_pct.mean().round(2)
```

[415]:
```
room_class
Elite           58.04
Premium         58.03
Presidential    59.30
Standard        58.22
Name: occ_pct, dtype: float64
```

[350]:
```
df.shape
```

[350]:
```
(9200, 8)
```

[416]:
```
df
```

[416]:
```
      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
…             …            …            …                     …       …
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
0       83.33     RT1      Standard
1       93.33     RT1      Standard
2       76.67     RT1      Standard
3      157.89     RT1      Standard
4       94.74     RT1      Standard
…         …     …            …
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
```

```
[9200 rows x 8 columns]
```

[417]:
```
df.drop("room_id",axis=1,inplace=True)
```

[418]:
```
df.head()
```

[418]:
```
   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
```

Average occupancy rate per city

[419]:
```
df_hotels.head()
```

[419]:
```
   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
```

[420]:
```
df1=pd.merge(df_hotels,df,on='property_id')
```

[421]:
```
df1.head()
```

[421]:
```
   property_id property_name category   city check_in_date room_category  \
0        16558  Atliq Grands   Luxury  Delhi      1-May-22           RT1
1        16558  Atliq Grands   Luxury  Delhi      2-May-22           RT1
2        16558  Atliq Grands   Luxury  Delhi      3-May-22           RT1
3        16558  Atliq Grands   Luxury  Delhi      4-May-22           RT1
4        16558  Atliq Grands   Luxury  Delhi      5-May-22           RT1

   successful_bookings  capacity  occ_pct room_class
0                   18      19.0    94.74   Standard
1                   12      19.0    63.16   Standard
2                   14      19.0    73.68   Standard
3                   13      19.0    68.42   Standard
```

```
4                   12      19.0    63.16   Standard
```

[422]:
```
df1.drop("category",axis=1,inplace=True)
```

[423]:
```
df1.head()
```

[423]:
```
   property_id property_name   city check_in_date room_category  \
0        16558  Atliq Grands  Delhi      1-May-22           RT1
1        16558  Atliq Grands  Delhi      2-May-22           RT1
2        16558  Atliq Grands  Delhi      3-May-22           RT1
3        16558  Atliq Grands  Delhi      4-May-22           RT1
4        16558  Atliq Grands  Delhi      5-May-22           RT1

   successful_bookings  capacity  occ_pct room_class
0                   18      19.0    94.74   Standard
1                   12      19.0    63.16   Standard
2                   14      19.0    73.68   Standard
3                   13      19.0    68.42   Standard
4                   12      19.0    63.16   Standard
```

[424]:
```
df1.groupby("city").occ_pct.mean().round(2)
```

[424]:
```
city
Bangalore    56.59
Delhi        61.61
Hyderabad    58.14
Mumbai       57.94
Name: occ_pct, dtype: float64
```

when was the occupancy better? weekday or weekend?

[425]:
```
df_date.head()
```

[425]:
```
        date  mmm yy week no  day_type
0  1-May-22  22-May    W 19   weekend
1  2-May-22  22-May    W 19  weekeday
2  3-May-22  22-May    W 19  weekeday
3  4-May-22  22-May    W 19  weekeday
4  5-May-22  22-May    W 19  weekeday
```

[426]:
```
df2=pd.merge(df_date,df1,left_on="date",right_on="check_in_date")
```

[165]:
```
df2.head()
```

[165]:
```
        date  mmm yy week no day_type  property_id  property_name    city  \
0  1-May-22  22-May    W 19  weekend        16558   Atliq Grands   Delhi
1  1-May-22  22-May    W 19  weekend        16558   Atliq Grands   Delhi
2  1-May-22  22-May    W 19  weekend        16558   Atliq Grands   Delhi
```

```
3  1-May-22  22-May    W 19  weekend        16558   Atliq Grands   Delhi
4  1-May-22  22-May    W 19  weekend        16559  Atliq Exotica  Mumbai

  check_in_date room_category  successful_bookings  capacity  occ_pct  \
0      1-May-22           RT1                   18      19.0    94.74
1      1-May-22           RT2                   21      22.0    95.45
2      1-May-22           RT3                    8       8.0   100.00
3      1-May-22           RT4                    3       3.0   100.00
4      1-May-22           RT1                   25      30.0    83.33

     room_class
0      Standard
1         Elite
2       Premium
3  Presidential
4      Standard
```

[427]:
```
df2.groupby("day_type").occ_pct.mean().round(2)
```

[427]:
```
day_type
weekeday    51.82
weekend     74.24
Name: occ_pct, dtype: float64
```

occupancy rate per month for every city

[428]:
```
df2["mmm yy"].unique()
```

[428]:
```
array(['22-May', '22-Jun', '22-Jul'], dtype=object)
```

[429]:
```
df_jun_22=df2[df2["mmm yy"]=='22-Jun'].groupby("city").occ_pct.mean().round(2)
```

[169]:
```
df_jun_22.sort_values(ascending=False)
```

[169]:
```
city
Delhi        61.46
Mumbai       57.79
Hyderabad    57.69
Bangalore    55.95
Name: occ_pct, dtype: float64
```

add August data

[170]:
```
df.head()
```

[170]:
```
   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
```

[171]:
```
df.columns
```

[171]:
```
Index(['property_id', 'check_in_date', 'room_category', 'successful_bookings',
       'capacity', 'occ_pct', 'room_class'],
      dtype='object')
```

[430]:
```
df_new_data.columns
```

[430]:
```
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')
```

[173]:
```
df.shape
```

[173]:
```
(9200, 7)
```

[174]:
```
df_new_data.shape
```

[174]:
```
(7, 13)
```

[431]:
```
df_latest=pd.concat([df,df_new_data],ignore_index=True,axis=0)
```

[432]:
```
df_latest.tail(10)
```

[432]:
```
      property_id check_in_date room_category  successful_bookings  capacity  \
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
9200        16559     01-Aug-22           RT1                   30      30.0
9201        19562     01-Aug-22           RT1                   21      30.0
9202        19563     01-Aug-22           RT1                   23      30.0
9203        19558     01-Aug-22           RT1                   30      40.0
9204        19560     01-Aug-22           RT1                   20      26.0
9205        17561     01-Aug-22           RT1                   18      26.0
9206        17564     01-Aug-22           RT1                   10      16.0

      occ_pct    room_class  property_name  category       city  mmm yy  \
```

```
9197     50.0  Presidential            NaN       NaN        NaN     NaN
9198     50.0  Presidential            NaN       NaN        NaN     NaN
9199     75.0  Presidential            NaN       NaN        NaN     NaN
9200      NaN      Standard  Atliq Exotica    Luxury     Mumbai  Aug-22
9201      NaN      Standard      Atliq Bay    Luxury  Bangalore  Aug-22
9202      NaN      Standard   Atliq Palace  Business  Bangalore  Aug-22
9203      NaN      Standard   Atliq Grands    Luxury  Bangalore  Aug-22
9204      NaN      Standard     Atliq City  Business  Bangalore  Aug-22
9205      NaN      Standard      Atliq Blu    Luxury     Mumbai  Aug-22
9206      NaN      Standard  Atliq Seasons  Business     Mumbai  Aug-22

     week no  day_type    occ%
9197     NaN       NaN     NaN
9198     NaN       NaN     NaN
9199     NaN       NaN     NaN
9200    W 32  weekeday  100.00
9201    W 32  weekeday   70.00
9202    W 32  weekeday   76.67
9203    W 32  weekeday   75.00
9204    W 32  weekeday   76.92
9205    W 32  weekeday   69.23
9206    W 32  weekeday   62.50
```

[177]:
```
df_latest.shape
```

[177]:
```
(9207, 14)
```

revenue_realized per city

[178]:
```
df_hotels.head(2)
```

[178]:
```
   property_id  property_name category    city
0        16558   Atliq Grands   Luxury   Delhi
1        16559  Atliq Exotica   Luxury  Mumbai
```

[179]:
```
df_bookings.head(2)
```

[179]:
```
          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

   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

   revenue_generated  revenue_realized
1               9100              3640
4              10920             10920
```

[433]:
```
df_bookings_all=pd.merge(df_hotels,df_bookings,on="property_id")
```

[188]:
```
df_bookings_all.head(2)
```

[188]:
```
   property_id property_name category    city        booking_id booking_date  \
0        16558  Atliq Grands   Luxury  Delhi  May012216558RT12     30-04-22
1        16558  Atliq Grands   Luxury  Delhi  May012216558RT15     27-04-22

  check_in_date checkout_date  no_guests room_category booking_platform  \
0      1/5/2022      2/5/2022        2.0           RT1           others
1      1/5/2022      2/5/2022        4.0           RT1    direct online

   ratings_given booking_status  revenue_generated  revenue_realized
0            NaN      Cancelled               9100              3640
1            5.0    Checked Out              10920             10920
```

[434]:
```
df_bookings_all.groupby("city").revenue_realized.sum()
```

[434]:
```
city
Bangalore    420383550
Delhi        294404488
Hyderabad    325179310
Mumbai       668569251
Name: revenue_realized, dtype: int64
```

print month by month revenue

[201]:
```
df_date.head(3)
```

[201]:
```
       date  mmm yy week no  day_type
0  1-May-22  22-May     W 19   weekend
1  2-May-22  22-May     W 19  weekeday
2  3-May-22  22-May     W 19  weekeday
```

[184]:
```
# revenue realised on check in date...so join check in date and date column
```

[435]:
```
pd.merge(df_bookings_all, df_date, left_on="check_in_date", right_on="date")
```

[435]:
```
Empty DataFrame
Columns: [property_id, property_name, category, city, booking_id, booking_date,
check_in_date, checkout_date, no_guests, room_category, booking_platform,
ratings_given, booking_status, revenue_generated, revenue_realized, date, mmm
yy, week no, day_type]
Index: []
```

[186]:
```
# not getting any output as date format is different in both data frames
```

[193]:
```
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   property_id        134573 non-null  int64  
 1   property_name      134573 non-null  object 
 2   category           134573 non-null  object 
 3   city               134573 non-null  object 
 4   booking_id         134573 non-null  object 
 5   booking_date       134573 non-null  object 
 6   check_in_date      134573 non-null  object 
 7   checkout_date      134573 non-null  object 
 8   no_guests          134573 non-null  float64
 9   room_category      134573 non-null  object 
 10  booking_platform   134573 non-null  object 
 11  ratings_given      56676 non-null   float64
 12  booking_status     134573 non-null  object 
 13  revenue_generated  134573 non-null  int64  
 14  revenue_realized   134573 non-null  int64  
dtypes: float64(2), int64(3), object(10)
memory usage: 15.4+ MB
```

[194]:
```
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
```

[436]:
```
# we have to change data type of both data frame columns "checkin_date" and␣
  ↪"date" column to datetime
```

[437]:
```
df_date["date"]=pd.to_datetime(df_date["date"], format="%d-%b-%y")
```

[438]:
```
df_date.head()
```

[438]:

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

```
 3 2022-05-04  22-May    W 19  weekeday
 4 2022-05-05  22-May    W 19  weekeday
```

[439]:
```
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
```

[444]:
```
df_bookings_all["check_in_date"]=pd.
 ↪to_datetime(df_bookings_all["check_in_date"],format="mixed", dayfirst=True)
```

[441]:
```
df_bookings_all.head(100)
```

[441]:

	property_id	property_name	category	city	booking_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform
0	16558	Atliq Grands	Luxury	Delhi	May012216558RT12	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others
1	16558	Atliq Grands	Luxury	Delhi	May012216558RT15	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online
2	16558	Atliq Grands	Luxury	Delhi	May012216558RT16	1/5/2022	1/5/2022	3/5/2022	2.0	RT1	others
3	16558	Atliq Grands	Luxury	Delhi	May012216558RT17	28-04-22	1/5/2022	6/5/2022	2.0	RT1	others
4	16558	Atliq Grands	Luxury	Delhi	May012216558RT18	26-04-22	1/5/2022	3/5/2022	2.0	RT1	logtrip
..	…	…	…	…	…	…	…	…	…	…	…
95	16558	Atliq Grands	Luxury	Delhi	May032216558RT25	26-04-22	3/5/2022	4/5/2022	2.0	RT2	tripster
96	16558	Atliq Grands	Luxury	Delhi	May032216558RT26	27-04-22	3/5/2022	5/5/2022	2.0	RT2	logtrip
97	16558	Atliq Grands	Luxury	Delhi	May032216558RT27	29-04-22	3/5/2022	8/5/2022	2.0	RT2	logtrip
98	16558	Atliq Grands	Luxury	Delhi	May032216558RT28	28-04-22	3/5/2022	4/5/2022	2.0	RT2	tripster
99	16558	Atliq Grands	Luxury	Delhi	May032216558RT29	28-04-22	3/5/2022	9/5/2022	1.0	RT2	others

```
    ratings_given booking_status  revenue_generated  revenue_realized
0             NaN      Cancelled               9100              3640
1             5.0    Checked Out              10920             10920
2             4.0    Checked Out               9100              9100
3             NaN      Cancelled               9100              3640
4             NaN        No Show               9100              9100
..            …            …                  …                 …
95            3.0    Checked Out              12600             12600
96            NaN    Checked Out              12600             12600
97            NaN    Checked Out              12600             12600
98            5.0    Checked Out              12600             12600
99            5.0    Checked Out              12600             12600

[100 rows x 15 columns]
```

[445]:
```
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   property_id        134573 non-null  int64
 1   property_name      134573 non-null  object
 2   category           134573 non-null  object
 3   city               134573 non-null  object
 4   booking_id         134573 non-null  object
 5   booking_date       134573 non-null  object
 6   check_in_date      134573 non-null  datetime64[ns]
 7   checkout_date      134573 non-null  object
 8   no_guests          134573 non-null  float64
 9   room_category      134573 non-null  object
 10  booking_platform   134573 non-null  object
 11  ratings_given      56676 non-null   float64
 12  booking_status     134573 non-null  object
 13  revenue_generated  134573 non-null  int64
 14  revenue_realized   134573 non-null  int64
dtypes: datetime64[ns](1), float64(2), int64(3), object(9)
memory usage: 15.4+ MB
```

[446]:
```
df_bookings_all=pd.merge(df_bookings_all,df_date, left_on="check_in_date",␣
↪right_on="date")
df_bookings_all.head()
```

[446]:
```
   property_id property_name category   city        booking_id booking_date  \
0        16558  Atliq Grands   Luxury  Delhi  May012216558RT12     30-04-22
1        16558  Atliq Grands   Luxury  Delhi  May012216558RT15     27-04-22
2        16558  Atliq Grands   Luxury  Delhi  May012216558RT16     1/5/2022
```

```
3        16558  Atliq Grands   Luxury  Delhi  May012216558RT17     28-04-22
4        16558  Atliq Grands   Luxury  Delhi  May012216558RT18     26-04-22

  check_in_date checkout_date  no_guests room_category booking_platform  \
0    2022-05-01      2/5/2022        2.0           RT1           others
1    2022-05-01      2/5/2022        4.0           RT1    direct online
2    2022-05-01      3/5/2022        2.0           RT1           others
3    2022-05-01      6/5/2022        2.0           RT1           others
4    2022-05-01      3/5/2022        2.0           RT1          logtrip

   ratings_given booking_status  revenue_generated  revenue_realized  \
0            NaN      Cancelled               9100              3640
1            5.0    Checked Out              10920             10920
2            4.0    Checked Out               9100              9100
3            NaN      Cancelled               9100              3640
4            NaN        No Show               9100              9100

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

[447]:
```
df_bookings_all.groupby("mmm yy")["revenue_realized"].sum()
```

[447]:
```
mmm yy
22-Jul    572843348
22-Jun    553925855
22-May    581767396
Name: revenue_realized, dtype: int64
```

revenue_realized per hotel type

[267]:
```
df_hotels.head()
```

[267]:
```
   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
```

print revenue realized per hotel

[268]:
```
df_bookings_all.property_name.unique()
```

[268]:
```
array(['Atliq Grands', 'Atliq Exotica', 'Atliq City', 'Atliq Blu',
       'Atliq Bay', 'Atliq Palace', 'Atliq Seasons'], dtype=object)
```

[270]:
```
df_bookings_all.groupby('property_name')["revenue_realized"].sum().sort_values()
```

[270]:
```
property_name
Atliq Seasons     66086735
Atliq Grands     211462134
Atliq Bay        259996918
Atliq Blu        260851922
Atliq City       285798439
Atliq Palace     304081863
Atliq Exotica    320258588
Name: revenue_realized, dtype: int64
```

Print average rating per city

[272]:
```
df_bookings_all.columns
```

[272]:
```
Index(['property_id', 'property_name', 'category', 'city', 'booking_id',
       'booking_date', 'check_in_date', 'checkout_date', 'no_guests',
       'room_category', 'booking_platform', 'ratings_given', 'booking_status',
       'revenue_generated', 'revenue_realized', 'date', 'mmm yy', 'week no',
       'day type'],
      dtype='object')
```

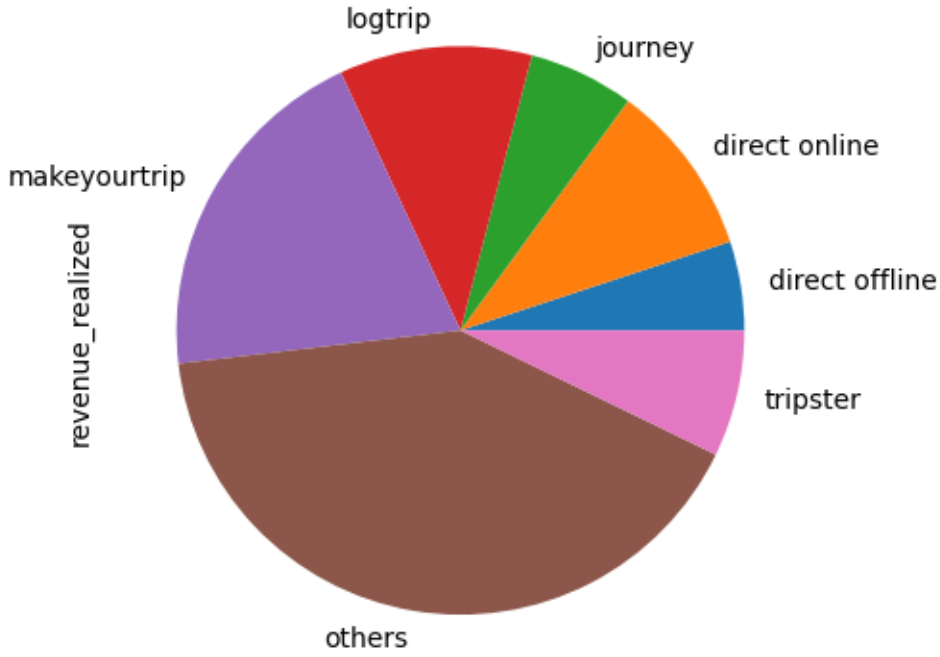
[273]:
```
df_bookings_all.groupby('city')["ratings_given"].mean().round(2).sort_values()
```

[273]:
```
city
Bangalore    3.41
Mumbai       3.65
Hyderabad    3.66
Delhi        3.78
Name: ratings_given, dtype: float64
```

Print a pie chart of revenue realized per booking platform

[277]:
```
df_bookings_all.groupby('booking_platform')["revenue_realized"].sum().
↪plot(kind='pie')
```

[277]:
```
<Axes: ylabel='revenue_realized'>
```


logtrip
journey
direct online
direct offline
tripster
others
revenue_realized
makeyourtrip