In [2]:
```
import pandas as pd
```

In [11]:
```
df_bookings = pd.read_csv("datasets/fact_bookings.csv")
df_bookings.head(2)
```

Out[11]:

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	rati
0	May012216558RT11	16558	27-04-22	1/5/2022	2/5/2022	-3.0	RT1	direct online	
1	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others	

In [12]:
```
df_bookings.shape
```

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

In [14]:
```
df_bookings.room_category.value_counts()
```

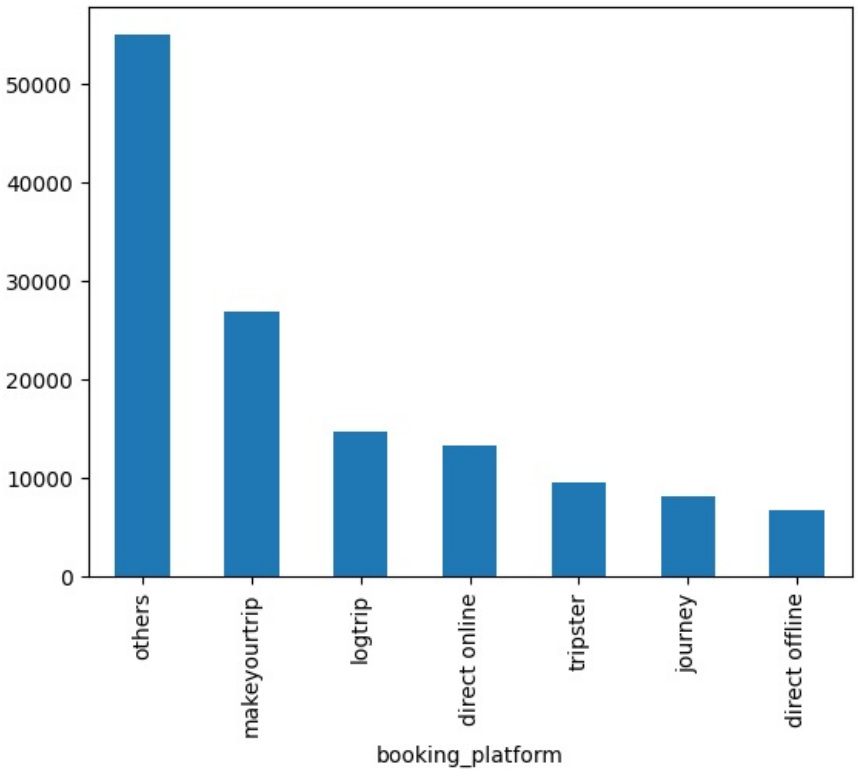
Out[14]:
```
room_category
RT2    49505
RT1    38446
RT3    30566
RT4    16073
Name: count, dtype: int64
```

In [15]:
```
df_bookings.booking_platform.value_counts()
```

Out[15]:
```
booking_platform
others            55066
makeyourtrip      26898
logtrip           14756
direct online     13379
tripster           9630
journey            8106
direct offline     6755
Name: count, dtype: int64
```

In [23]:
```
df_bookings.booking_platform.value_counts().plot(kind="bar")
```

Out[23]:
```
<Axes: xlabel='booking_platform'>
```



In [24]:
```
df_bookings.describe()
```

Out[24]:

	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

In [25]:
```
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")
```

In [26]:
```
df_hotels.shape
```

Out[26]:
```
(25, 4)
```

In [27]:
```
df_hotels.describe()
```

Out[27]:

	property_id
count	25.000000
mean	18040.640000
std	1122.436371
min	16558.000000
25%	17558.000000
50%	17564.000000
75%	18563.000000
max	19563.000000

In [28]:
```
df_rooms.describe()
```

Out[28]:

	room_id	room_class
count	4	4
unique	4	4
top	RT1	Standard
freq	1	1

In [29]:
```
df_hotels.head(4)
```

Out[29]:

	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

In [31]:
```
df_hotels.category.value_counts()
```

Out[31]:
```
category
Luxury      16
Business     9
Name: count, dtype: int64
```

In [32]:
```
df_bookings.describe()
```

Out[32]:

	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

In [36]:
```
df_bookings=df_bookings[df_bookings.no_guests>0]
df_bookings
```

Out[36]:

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform
1	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others
2	May012216558RT13	16558	28-04-22	1/5/2022	4/5/2022	2.0	RT1	logtrip
4	May012216558RT15	16558	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online
5	May012216558RT16	16558	1/5/2022	1/5/2022	3/5/2022	2.0	RT1	others
6	May012216558RT17	16558	28-04-22	1/5/2022	6/5/2022	2.0	RT1	others
...	...	...	...	...	...	...	...	...
134584	Jul312217564RT45	17564	30-07-22	31-07-22	1/8/2022	2.0	RT4	others
134585	Jul312217564RT46	17564	29-07-22	31-07-22	3/8/2022	1.0	RT4	makeyourtrip
134587	Jul312217564RT48	17564	30-07-22	31-07-22	2/8/2022	1.0	RT4	tripste
134588	Jul312217564RT49	17564	29-07-22	31-07-22	1/8/2022	2.0	RT4	logtrip
134589	Jul312217564RT410	17564	31-07-22	31-07-22	1/8/2022	2.0	RT4	makeyourtrip

134578 rows × 12 columns

In [37]:
```
df_bookings.shape
```

Out[37]: (134578, 12)

In [39]:
```
df_bookings.revenue_generated.min(),df_bookings.revenue_generated.max()
```

Out[39]: (6500, 28560000)

In [42]:
```
avg, std = df_bookings.revenue_generated.mean(),df_bookings.revenue_generated.std()
```

In [43]:
```
avg, std
```

Out[43]: (15378.036937686695, 93040.1549314641)

In [50]:
```
higher_limit = avg + 3*std
```

In [51]:
```
higher_limit
```

Out[51]: 294498.50173207896

In [53]:
```
lower_limit = avg - 3*std
lower_limit
```

Out[53]: -263742.4278567056

In [57]:
```
df_bookings[df_bookings.revenue_generated>higher_limit]
```

Out[57]:

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platfor
2	May012216558RT13	16558	28-04-22	1/5/2022	4/5/2022	2.0	RT1	logtr
111	May012216559RT32	16559	29-04-22	1/5/2022	2/5/2022	6.0	RT3	direct onlir
315	May012216562RT22	16562	28-04-22	1/5/2022	4/5/2022	2.0	RT2	direct offlir
562	May012217559RT118	17559	26-04-22	1/5/2022	2/5/2022	2.0	RT1	othe
129176	Jul282216562RT26	16562	21-07-22	28-07-22	29-07-22	2.0	RT2	direct onlir

In [63]:
```
df_bookings =df_bookings[df_bookings.revenue_generated<higher_limit]
df_bookings.shape
```

Out[63]:
```
(134573, 12)
```

In [65]:
```
df_bookings.revenue_realized.describe()
```

Out[65]:
```
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
```

In [67]:
```
higher_limit = df_bookings.revenue_realized.mean()+3* df_bookings.revenue_realized.std()
```

In [68]:
```
higher_limit
```

Out[68]:
```
33479.358661845814
```

In [74]:
```
df_rooms
```

Out[74]:

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

In [76]:
```
df_bookings[df_bookings.room_category=="RT4"].revenue_realized.describe()
```

Out[76]:
```
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
```

Data Transformation

In [79]:
```
df_agg_bookings.head()
```

Out[79]:

	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

In [104…
```
df_agg_bookings["occ_pct"]=df_agg_bookings["successful_bookings"]/ df_agg_bookings["capacity"]
```

In [105…
```
df_agg_bookings.head(5)
```

Out[105…

	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

In [106…
```
df_agg_bookings["occ_pct"] = df_agg_bookings["occ_pct"].apply(lambda x: round(x*100, 2))
df_agg_bookings.head(4)
```

Out[106…

	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

Report Generation

1.what is the average occupancy rate in each of the room categories

In [107…
```
df_agg_bookings.groupby("room_category")["occ_pct"].mean().round(2)
```

Out[107…
```
room_category
RT1    58.22
RT2    58.04
RT3    58.03
RT4    59.30
Name: occ_pct, dtype: float64
```

In [108…
```
df_rooms
```

Out[108…

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

In [109…
```
df= pd.merge(df_agg_bookings,df_rooms, left_on =("room_category"), right_on = ("room_id"))
df.head(5)
```

Out[109…

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

In [112…
```
df.tail(5)
```

Out[112…

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct	room_id	room_class
9195	16563	31-Jul-22	RT4	13	18.0	72.22	RT4	Presidential
9196	16559	31-Jul-22	RT4	13	18.0	72.22	RT4	Presidential
9197	17558	31-Jul-22	RT4	3	6.0	50.00	RT4	Presidential
9198	19563	31-Jul-22	RT4	3	6.0	50.00	RT4	Presidential
9199	17561	31-Jul-22	RT4	3	4.0	75.00	RT4	Presidential

In [113…
```
df.groupby("room_class")["occ_pct"].mean().round(2)
```

Out[113…
```
room_class
Elite           58.04
Premium         58.03
Presidential    59.30
Standard        58.22
Name: occ_pct, dtype: float64
```

In [115…
```
df.drop("room_id", axis=1, inplace=True)
df.head(4)
```

Out[115…

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

2.Print average occupancy rate per city

In [116…
```
df_hotels
```

Out[116…

	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

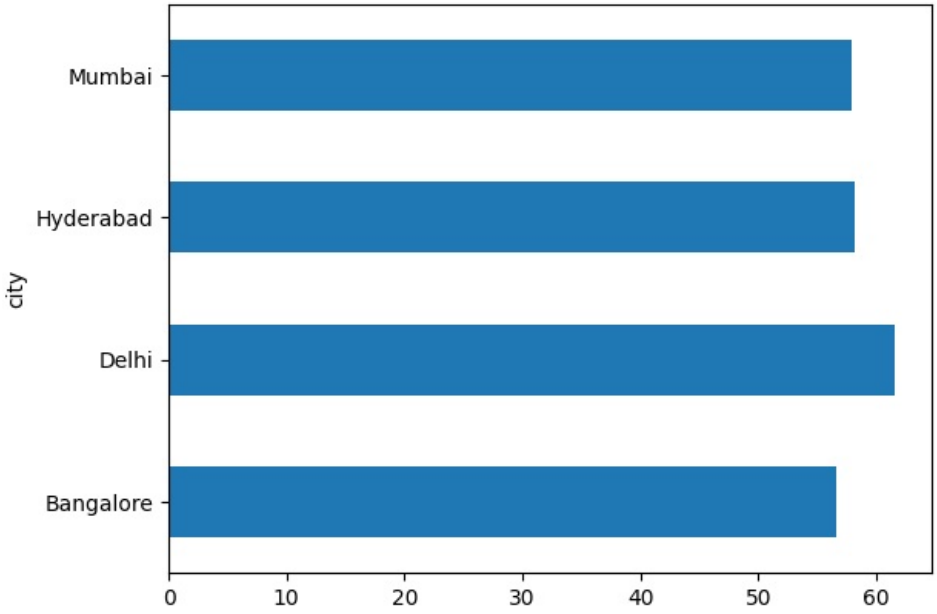
In [117…
```
df = pd.merge(df,df_hotels, on="property_id")
df.head(5)
```

Out[117…

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct	room_class	property_name	category	
0	16559	1-May-22	RT1	25	30.0	83.33	Standard	Atliq Exotica	Luxury	Mun
1	16559	2-May-22	RT1	20	30.0	66.67	Standard	Atliq Exotica	Luxury	Mun
2	16559	3-May-22	RT1	17	30.0	56.67	Standard	Atliq Exotica	Luxury	Mun
3	16559	4-May-22	RT1	21	30.0	70.00	Standard	Atliq Exotica	Luxury	Mun
4	16559	5-May-22	RT1	16	30.0	53.33	Standard	Atliq Exotica	Luxury	Mun

In [119…
```
df.groupby("city")["occ_pct"].mean().plot(kind="barh")
```

Out[119…
```
<Axes: ylabel='city'>
```



3.when is occupancy better weekend or weekdays

In [120…
```
df_date
```

Out[120…

	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
...	...	...	...	...
87	27-Jul-22	Jul 22	W 31	weekeday
88	28-Jul-22	Jul 22	W 31	weekeday
89	29-Jul-22	Jul 22	W 31	weekeday
90	30-Jul-22	Jul 22	W 31	weekend
91	31-Jul-22	Jul 22	W 32	weekend

92 rows × 4 columns

In [121…
```
df = pd.merge(df,df_date, left_on="check_in_date", right_on="date")
df.head(5)
```

Out[121…

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct	room_class	property_name	category	
0	16559	10-May-22	RT1	18	30.0	60.00	Standard	Atliq Exotica	Luxury	M
1	16559	10-May-22	RT2	25	41.0	60.98	Elite	Atliq Exotica	Luxury	M
2	16559	10-May-22	RT3	20	32.0	62.50	Premium	Atliq Exotica	Luxury	M
3	16559	10-May-22	RT4	13	18.0	72.22	Presidential	Atliq Exotica	Luxury	M
4	19562	10-May-22	RT1	18	30.0	60.00	Standard	Atliq Bay	Luxury	Ban

In [122…
```
df.groupby("day_type")["occ_pct"].mean().round(2)
```

Out[122…
```
day_type
weekeday    50.90
weekend     72.39
Name: occ_pct, dtype: float64
```

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

In [124…
```
df["mmm yy"].unique()
```

Out[124…
```
array(['May 22', 'Jun 22', 'Jul 22'], dtype=object)
```

In [128…
```
df_june_22 = df[df["mmm yy"]=="Jun 22"]
```

In [130…
```
df_june_22.head(4
           )
```

Out[130…

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct	room_class	property_name	category	
2200	16559	10-Jun-22	RT1	20	30.0	66.67	Standard	Atliq Exotica	Luxury	I
2201	16559	10-Jun-22	RT2	26	41.0	63.41	Elite	Atliq Exotica	Luxury	I
2202	16559	10-Jun-22	RT3	20	32.0	62.50	Premium	Atliq Exotica	Luxury	I
2203	16559	10-Jun-22	RT4	11	18.0	61.11	Presidential	Atliq Exotica	Luxury	I

In [131…
```
df_june_22.groupby("city")[     "occ_pct"].mean().round(2).sort_values(ascending = False)
```

Out[131…
```
city
Delhi        62.47
Hyderabad    58.46
Mumbai       58.38
Bangalore    56.58
Name: occ_pct, dtype: float64
```

In [135…
```
df_august=pd.read_csv("datasets/new_data_august.csv")
```

In [138…
```
df_august.columns
```

Out[138…
```
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')
```

In [139…
```
df_august.shape
```

Out[139…
```
(7, 13)
```

In [141…
```
df.shape
```

Out[141…
```
(6500, 14)
```

In [145…
```
latest_df = pd.concat([df, df_august], ignore_index = True, axis = 0)
latest_df.tail(5)
```

Out[145…

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct	room_class	property_name	category	
6502	19563	01-Aug-22	RT1	23	30.0	NaN	Standard	Atliq Palace	Business	I
6503	19558	01-Aug-22	RT1	30	40.0	NaN	Standard	Atliq Grands	Luxury	I
6504	19560	01-Aug-22	RT1	20	26.0	NaN	Standard	Atliq City	Business	I
6505	17561	01-Aug-22	RT1	18	26.0	NaN	Standard	Atliq Blu	Luxury	
6506	17564	01-Aug-22	RT1	10	16.0	NaN	Standard	Atliq Seasons	Business	

In [146…
```
latest_df.shape
```

Out[146… (6507, 15)

6. print revenue realized by city

In [147…
```
df_hotels.head(3)
```

Out[147…

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

In [149…
```
df_bookings_all= pd.merge(df_bookings, df_hotels, on ="property_id")
df_bookings_all.head(3)
```

Out[149…

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	rati
0	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others	
1	May012216558RT15	16558	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online	
2	May012216558RT16	16558	1/5/2022	1/5/2022	3/5/2022	2.0	RT1	others	

In [150…
```
df_bookings_all.groupby("city")["revenue_realized"].sum()
```

Out[150…
```
city
Bangalore    420383550
Delhi        294404488
Hyderabad    325179310
Mumbai       668569251
Name: revenue_realized, dtype: int64
```

7.Print by monthly revenue

In [151…
```
df_bookings_all.head(3)
```

Out[151…

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	rati
0	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others	
1	May012216558RT15	16558	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online	
2	May012216558RT16	16558	1/5/2022	1/5/2022	3/5/2022	2.0	RT1	others	

In [ ]: