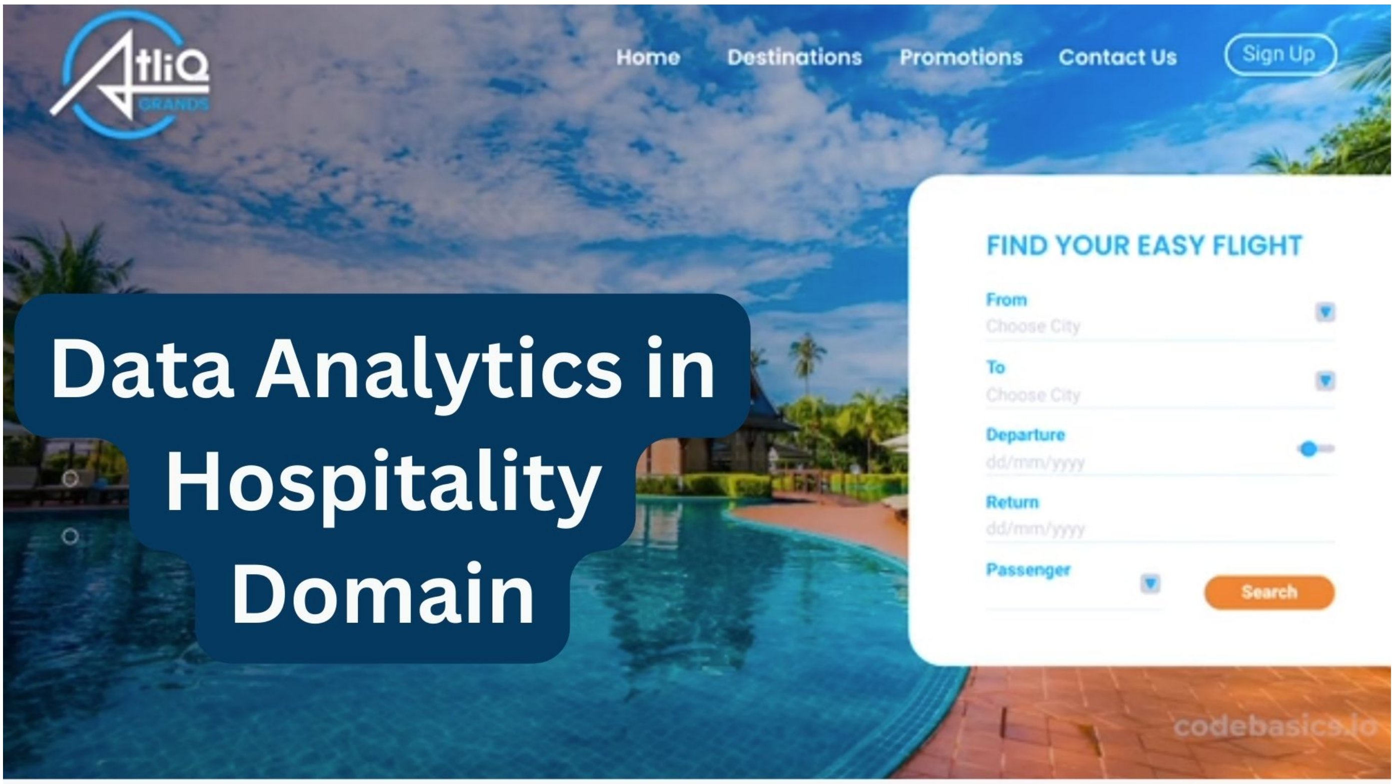


```python
import pandas as pd
```

```python
df_bookings = pd.read_csv('/content/drive/MyDrive/fact_bookings.csv')
df_bookings
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_
0	May012216558RT11	16558	27-04-22	1/5/2022	2/5/2022	-3.0	RT1	direct online	1.0
1	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others	NaN
2	May012216558RT13	16558	28-04-22	1/5/2022	4/5/2022	2.0	RT1	logtrip	5.0
3	May012216558RT14	16558	28-04-22	1/5/2022	2/5/2022	-2.0	RT1	others	NaN
4	May012216558RT15	16558	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online	5.0
...	...	...	...	...	...	...	...	...	...
134585	Jul312217564RT46	17564	29-07-22	31-07-22	3/8/2022	1.0	RT4	makeyourtrip	2.0
134586	Jul312217564RT47	17564	30-07-22	31-07-22	1/8/2022	-4.0	RT4	logtrip	2.0
134587	Jul312217564RT48	17564	30-07-22	31-07-22	2/8/2022	1.0	RT4	tripster	NaN
134588	Jul312217564RT49	17564	29-07-22	31-07-22	1/8/2022	2.0	RT4	logtrip	2.0
134589	Jul312217564RT410	17564	31-07-22	31-07-22	1/8/2022	2.0	RT4	makeyourtrip	NaN

134590 rows × 12 columns

```python
df_bookings.shape
```

```
(134590, 12)
```

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

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

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

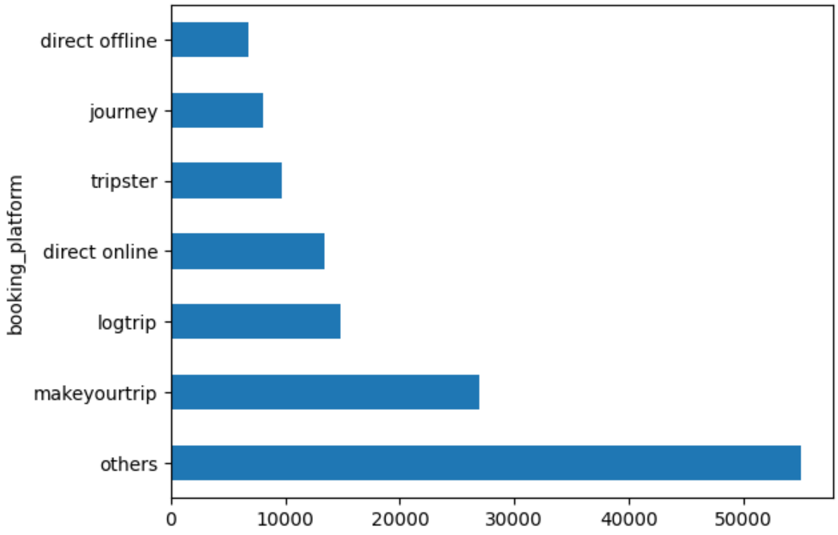
```
array(['direct online', 'others', 'logtrip', 'tripster', 'makeyourtrip',
       'journey', 'direct offline'], dtype=object)
```

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

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

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

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



```
df_bookings.describe()
```

	property_id	no_guests	ratings_given	revenue_generated	revenue_realized
count	134590.000000	134587.000000	56683.000000	1.345900e+05	134590.000000
mean	18061.113493	2.036170	3.619004	1.537805e+04	12696.123256
std	1093.055847	1.034885	1.235009	9.303604e+04	6928.108124
min	16558.000000	-17.000000	1.000000	6.500000e+03	2600.000000
25%	17558.000000	1.000000	3.000000	9.900000e+03	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

	property_id	no_guests	ratings_given	revenue_generated	revenue_realized
max	19563.000000	6.000000	5.000000	2.856000e+07	45220.000000

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

```
(6500, 28560000)
```

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

```
df_hotels.shape
```

```
(25, 4)
```

```
df_hotels.head()
```

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

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

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

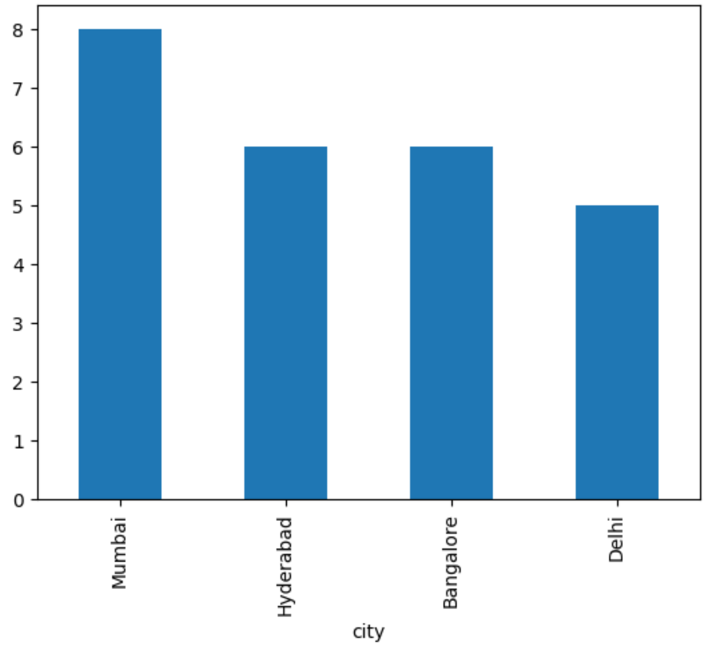
```
category
Luxury      16
Business     9
Name: count, dtype: int64
```

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

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

```
df_hotels.city.value_counts().plot(kind='bar')
```

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



```
df_bookings.head(5)
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_given
0	May012216558RT11	16558	27-04-22	1/5/2022	2/5/2022	-3.0	RT1	direct online	1.0
1	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others	NaN
2	May012216558RT13	16558	28-04-22	1/5/2022	4/5/2022	2.0	RT1	logtrip	5.0
3	May012216558RT14	16558	28-04-22	1/5/2022	2/5/2022	-2.0	RT1	others	NaN
4	May012216558RT15	16558	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online	5.0

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_agg_bookings.head(5)
```

	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

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_agg_bookings.property_id.unique()
```

```
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])
```

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

```
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
```

```
df_agg_bookings
```

	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

9200 rows × 5 columns

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_agg_bookings.capacity.max()
```

```
50.0
```

```
df_agg_bookings[df_agg_bookings.capacity == df_agg_bookings.capacity.max()]
```

	property_id	check_in_date	room_category	successful_bookings	capacity
27	17558	1-May-22	RT2	38	50.0
128	17558	2-May-22	RT2	27	50.0
229	17558	3-May-22	RT2	26	50.0
328	17558	4-May-22	RT2	27	50.0
428	17558	5-May-22	RT2	29	50.0
...	...	...	...	...	...
8728	17558	27-Jul-22	RT2	22	50.0
8828	17558	28-Jul-22	RT2	21	50.0
8928	17558	29-Jul-22	RT2	23	50.0
9028	17558	30-Jul-22	RT2	32	50.0
9128	17558	31-Jul-22	RT2	30	50.0

92 rows × 5 columns

⊞

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_bookings.describe()
```

	property_id	no_guests	ratings_given	revenue_generated	revenue_realized
count	134590.000000	134587.000000	56683.000000	1.345900e+05	134590.000000
mean	18061.113493	2.036170	3.619004	1.537805e+04	12696.123256
std	1093.055847	1.034885	1.235009	9.303604e+04	6928.108124
min	16558.000000	-17.000000	1.000000	6.500000e+03	2600.000000
25%	17558.000000	1.000000	3.000000	9.900000e+03	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

⊞

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

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

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	rating
0	May012216558RT11	16558	27-04-22	1/5/2022	2/5/2022	-3.0	RT1	direct online	1.0

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	rating
3	May012216558RT14	16558	28-04-22	1/5/2022	2/5/2022	-2.0	RT1	others	NaN
17924	May122218559RT44	18559	12/5/2022	12/5/2022	14-05-22	-10.0	RT4	direct online	NaN
18020	May122218561RT22	18561	8/5/2022	12/5/2022	14-05-22	-12.0	RT2	makeyourtrip	NaN
18119	May122218562RT311	18562	5/5/2022	12/5/2022	17-05-22	-6.0	RT3	direct offline	5.0
18121	May122218562RT313	18562	10/5/2022	12/5/2022	17-05-22	-4.0	RT3	direct online	NaN
56715	Jun082218562RT12	18562	5/6/2022	8/6/2022	13-06-22	-17.0	RT1	others	NaN
119765	Jul202219560RT220	19560	19-07-22	20-07-22	22-07-22	-1.0	RT2	others	NaN
134586	Jul312217564RT47	17564	30-07-22	31-07-22	1/8/2022	-4.0	RT4	logtrip	2.0

⊞

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

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

```
df_bookings
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_
1	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others	NaN
2	May012216558RT13	16558	28-04-22	1/5/2022	4/5/2022	2.0	RT1	logtrip	5.0
4	May012216558RT15	16558	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online	5.0
5	May012216558RT16	16558	1/5/2022	1/5/2022	3/5/2022	2.0	RT1	others	4.0
6	May012216558RT17	16558	28-04-22	1/5/2022	6/5/2022	2.0	RT1	others	NaN
...	...	...	...	...	...	...	...	...	...
134584	Jul312217564RT45	17564	30-07-22	31-07-22	1/8/2022	2.0	RT4	others	2.0
134585	Jul312217564RT46	17564	29-07-22	31-07-22	3/8/2022	1.0	RT4	makeyourtrip	2.0
134587	Jul312217564RT48	17564	30-07-22	31-07-22	2/8/2022	1.0	RT4	tripster	NaN
134588	Jul312217564RT49	17564	29-07-22	31-07-22	1/8/2022	2.0	RT4	logtrip	2.0
134589	Jul312217564RT410	17564	31-07-22	31-07-22	1/8/2022	2.0	RT4	makeyourtrip	NaN

134578 rows × 12 columns

⊞

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

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

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_
1	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others	NaN
4	May012216558RT15	16558	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online	5.0
5	May012216558RT16	16558	1/5/2022	1/5/2022	3/5/2022	2.0	RT1	others	4.0
6	May012216558RT17	16558	28-04-22	1/5/2022	6/5/2022	2.0	RT1	others	NaN

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_
7	May012216558RT18	16558	26-04-22	1/5/2022	3/5/2022	2.0	RT1	logtrip	NaN
...	...	...	...	...	...	...	...	...	...
134584	Jul312217564RT45	17564	30-07-22	31-07-22	1/8/2022	2.0	RT4	others	2.0
134585	Jul312217564RT46	17564	29-07-22	31-07-22	3/8/2022	1.0	RT4	makeyourtrip	2.0
134587	Jul312217564RT48	17564	30-07-22	31-07-22	2/8/2022	1.0	RT4	tripster	NaN
134588	Jul312217564RT49	17564	29-07-22	31-07-22	1/8/2022	2.0	RT4	logtrip	2.0
134589	Jul312217564RT410	17564	31-07-22	31-07-22	1/8/2022	2.0	RT4	makeyourtrip	NaN

134573 rows × 12 columns

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

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

```
(134573, 12)
```

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

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

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

```
33479.358661845814
```

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

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	rating
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

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	rating
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

1299 rows × 12 columns

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_bookings[df_bookings.room_category == "RT4"]
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_
47	May012216558RT41	16558	26-04-22	1/5/2022	3/5/2022	2.0	RT4	logtrip	NaN
48	May012216558RT42	16558	27-04-22	1/5/2022	2/5/2022	1.0	RT4	tripster	NaN
49	May012216558RT43	16558	29-04-22	1/5/2022	4/5/2022	2.0	RT4	direct offline	NaN
137	May012216559RT41	16559	27-04-22	1/5/2022	7/5/2022	4.0	RT4	others	NaN
138	May012216559RT42	16559	11/4/2022	1/5/2022	3/5/2022	2.0	RT4	direct offline	NaN
...	...	...	...	...	...	...	...	...	...
134584	Jul312217564RT45	17564	30-07-22	31-07-22	1/8/2022	2.0	RT4	others	2.0
134585	Jul312217564RT46	17564	29-07-22	31-07-22	3/8/2022	1.0	RT4	makeyourtrip	2.0
134587	Jul312217564RT48	17564	30-07-22	31-07-22	2/8/2022	1.0	RT4	tripster	NaN
134588	Jul312217564RT49	17564	29-07-22	31-07-22	1/8/2022	2.0	RT4	logtrip	2.0
134589	Jul312217564RT410	17564	31-07-22	31-07-22	1/8/2022	2.0	RT4	makeyourtrip	NaN

16071 rows × 12 columns

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

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

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

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

```
booking_id              0
property_id             0
booking_date            0
check_in_date           0
```

```
checkout_date            0
no_guests                0
room_category            0
booking_platform         0
ratings_given        77897
booking_status           0
revenue_generated        0
revenue_realized         0
dtype: int64
```

```
df_agg_bookings.isnull().sum()
```

```
property_id            0
check_in_date          0
room_category          0
successful_bookings    0
capacity               2
dtype: int64
```

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

	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

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_agg_bookings.capacity.median()
```

```
25.0
```

```
df_agg_bookings.capacity.fillna(df_agg_bookings.capacity.median(),inplace=True)
```

```
df_agg_bookings.loc[[8,14]]
```

	property_id	check_in_date	room_category	successful_bookings	capacity
8	17561	1-May-22	RT1	22	25.0
14	17562	1-May-22	RT1	12	25.0

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_agg_bookings[df_agg_bookings.successful_bookings > df_agg_bookings.capacity]
```

	property_id	check_in_date	room_category	successful_bookings	capacity

	property_id	check_in_date	room_category	successful_bookings	capacity
3	17558	1-May-22	RT1	30	19.0
12	16563	1-May-22	RT1	100	41.0
4136	19558	11-Jun-22	RT2	50	39.0
6209	19560	2-Jul-22	RT1	123	26.0
8522	19559	25-Jul-22	RT1	35	24.0
9194	18563	31-Jul-22	RT4	20	18.0

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_agg_bookings.shape
```

```
(9200, 5)
```

```
df_agg_bookings = df_agg_bookings[df_agg_bookings.successful_bookings <= df_agg_bookings.capacity]
```

```
df_agg_bookings.shape
```

```
(9194, 5)
```

```
df_agg_bookings.head(5)
```

	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
4	16558	1-May-22	RT1	18	19.0
5	17560	1-May-22	RT1	28	40.0

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_agg_bookings['occ_pct'] = df_agg_bookings['successful_bookings']/df_agg_bookings['capacity']
```

```
df_agg_bookings
```

	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

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct
4	16558	1-May-22	RT1	18	19.0	0.947368
5	17560	1-May-22	RT1	28	40.0	0.700000
...	...	...	...	...	...	...
9195	16563	31-Jul-22	RT4	13	18.0	0.722222
9196	16559	31-Jul-22	RT4	13	18.0	0.722222
9197	17558	31-Jul-22	RT4	3	6.0	0.500000
9198	19563	31-Jul-22	RT4	3	6.0	0.500000
9199	17561	31-Jul-22	RT4	3	4.0	0.750000

9194 rows × 6 columns

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_agg_bookings['occ_pct'] = df_agg_bookings['occ_pct'].round(2)
```

```
df_agg_bookings['occ_pct']=df_agg_bookings['occ_pct']*100
```

```
df_agg_bookings.head(5)
```

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct
0	16559	1-May-22	RT1	25	30.0	83.0
1	19562	1-May-22	RT1	28	30.0	93.0
2	19563	1-May-22	RT1	23	30.0	77.0
4	16558	1-May-22	RT1	18	19.0	95.0
5	17560	1-May-22	RT1	28	40.0	70.0

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_agg_bookings.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 9194 entries, 0 to 9199
Data columns (total 6 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   property_id          9194 non-null   int64  
 1   check_in_date        9194 non-null   object 
 2   room_category        9194 non-null   object 
 3   successful_bookings  9194 non-null   int64  
 4   capacity             9194 non-null   float64
 5   occ_pct              9194 non-null   float64
dtypes: float64(2), int64(2), object(2)
memory usage: 760.8+ KB
```

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

```
df_agg_bookings.groupby('room_category')['occ_pct'].mean()
```

```
room_category
RT1    57.902439
RT2    58.014354
RT3    58.011304
RT4    59.279687
Name: occ_pct, dtype: float64
```

```
df_rooms
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

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

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

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

	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	83.0	RT1	Standard
1	19562	1-May-22	RT1	28	30.0	93.0	RT1	Standard
2	19563	1-May-22	RT1	23	30.0	77.0	RT1	Standard
3	16558	1-May-22	RT1	18	19.0	95.0	RT1	Standard
4	17560	1-May-22	RT1	28	40.0	70.0	RT1	Standard
...	...	...	...	...	...	...	...	...
9189	16563	31-Jul-22	RT4	13	18.0	72.0	RT4	Presidential
9190	16559	31-Jul-22	RT4	13	18.0	72.0	RT4	Presidential
9191	17558	31-Jul-22	RT4	3	6.0	50.0	RT4	Presidential
9192	19563	31-Jul-22	RT4	3	6.0	50.0	RT4	Presidential
9193	17561	31-Jul-22	RT4	3	4.0	75.0	RT4	Presidential

9194 rows × 8 columns

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
df.groupby('room_class')['occ_pct'].mean()
```

```
room_class
Elite          58.014354
Premium        58.011304
Presidential   59.279687
Standard       57.902439
Name: occ_pct, dtype: float64
```

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

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct	room_class
0	16559	1-May-22	RT1	25	30.0	83.0	Standard
1	19562	1-May-22	RT1	28	30.0	93.0	Standard
2	19563	1-May-22	RT1	23	30.0	77.0	Standard
3	16558	1-May-22	RT1	18	19.0	95.0	Standard
4	17560	1-May-22	RT1	28	40.0	70.0	Standard
...	...	...	...	...	...	...	...
9189	16563	31-Jul-22	RT4	13	18.0	72.0	Presidential
9190	16559	31-Jul-22	RT4	13	18.0	72.0	Presidential
9191	17558	31-Jul-22	RT4	3	6.0	50.0	Presidential
9192	19563	31-Jul-22	RT4	3	6.0	50.0	Presidential
9193	17561	31-Jul-22	RT4	3	4.0	75.0	Presidential

9194 rows × 7 columns

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

2. Print average occupancy rate per city

```
df_hotels
```

	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

	property_id	property_name	category	city
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

⊞

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

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

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct	room_class	property_name	category	city
0	16559	1-May-22	RT1	25	30.0	83.0	Standard	Atliq Exotica	Luxury	Mumbai
1	16559	2-May-22	RT1	20	30.0	67.0	Standard	Atliq Exotica	Luxury	Mumbai
2	16559	3-May-22	RT1	17	30.0	57.0	Standard	Atliq Exotica	Luxury	Mumbai
3	16559	4-May-22	RT1	21	30.0	70.0	Standard	Atliq Exotica	Luxury	Mumbai
4	16559	5-May-22	RT1	16	30.0	53.0	Standard	Atliq Exotica	Luxury	Mumbai
...	...	...	...	...	...	...	...	...	...	...
9189	16563	27-Jul-22	RT4	10	18.0	56.0	Presidential	Atliq Palace	Business	Delhi
9190	16563	28-Jul-22	RT4	9	18.0	50.0	Presidential	Atliq Palace	Business	Delhi
9191	16563	29-Jul-22	RT4	9	18.0	50.0	Presidential	Atliq Palace	Business	Delhi
9192	16563	30-Jul-22	RT4	11	18.0	61.0	Presidential	Atliq Palace	Business	Delhi
9193	16563	31-Jul-22	RT4	13	18.0	72.0	Presidential	Atliq Palace	Business	Delhi

9194 rows × 10 columns

⊞

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

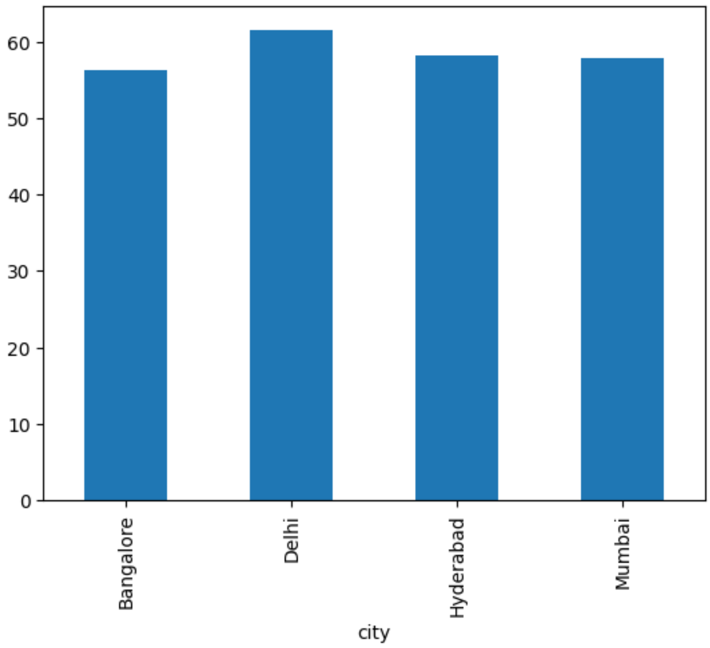
```
</div>
```

```
df.groupby('city')['occ_pct'].mean()
```

```
city
Bangalore    56.332880
Delhi        61.525285
Hyderabad    58.123244
Mumbai       57.897384
Name: occ_pct, dtype: float64
```

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

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



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

```
df.head(3)
```

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct	room_class	property_name	category	city
0	16559	1-May-22	RT1	25	30.0	83.0	Standard	Atliq Exotica	Luxury	Mumbai
1	16559	2-May-22	RT1	20	30.0	67.0	Standard	Atliq Exotica	Luxury	Mumbai
2	16559	3-May-22	RT1	17	30.0	57.0	Standard	Atliq Exotica	Luxury	Mumbai

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_date
```

	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

	date	mmm yy	week no	day_type
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

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df=pd.merge(df,df_date,left_on='check_in_date',right_on ='date')
df.head(3)
```

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct	room_class	property_name	category	city	dat
0	16559	10-May-22	RT1	18	30.0	60.0	Standard	Atliq Exotica	Luxury	Mumbai	10-Ma 22
1	16559	10-May-22	RT2	25	41.0	61.0	Elite	Atliq Exotica	Luxury	Mumbai	10-Ma 22
2	16559	10-May-22	RT3	20	32.0	62.0	Premium	Atliq Exotica	Luxury	Mumbai	10-Ma 22

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df.groupby('day_type')['occ_pct'].mean()
```

```
day_type
weekeday    50.879756
weekend     72.357218
Name: occ_pct, dtype: float64
```

4: In the month of June, what is the occupancy for different cities

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

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

```
df_jun = df[df['mmm yy']=='Jun 22']
```

```
df_jun.groupby('city')['occ_pct'].mean()
```

```
city
Bangalore    56.435388
Delhi        62.490476
Hyderabad    58.474206
Mumbai       58.386905
Name: occ_pct, dtype: float64
```

```
df_aug = pd.read_csv('/content/drive/MyDrive/new_data_august.csv')
df_aug.head(3)
```

	property_id	property_name	category	city	room_category	room_class	check_in_date	mmm yy	week no	day_type	successful_booki
0	16559	Atliq Exotica	Luxury	Mumbai	RT1	Standard	01-Aug-22	Aug-22	W 32	weekeday	30
1	19562	Atliq Bay	Luxury	Bangalore	RT1	Standard	01-Aug-22	Aug-22	W 32	weekeday	21
2	19563	Atliq Palace	Business	Bangalore	RT1	Standard	01-Aug-22	Aug-22	W 32	weekeday	23

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df.columns
```

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

5: We got new data for the month of august. Append that to existing data

```
df_aug.columns
```

```
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')
```

```
df.shape
```

```
(6497, 14)
```

```
df_aug.shape
```

```
(7, 13)
```

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

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_pct	room_class	property_name	category	city
6494	16563	31-Jul-22	RT2	32	38.0	84.0	Elite	Atliq Palace	Business	Delhi
6495	16563	31-Jul-22	RT3	14	20.0	70.0	Premium	Atliq Palace	Business	Delhi
6496	16563	31-Jul-22	RT4	13	18.0	72.0	Presidential	Atliq Palace	Business	Delhi
6497	16559	01-Aug-22	RT1	30	30.0	NaN	Standard	Atliq Exotica	Luxury	Mumbai
6498	19562	01-Aug-22	RT1	21	30.0	NaN	Standard	Atliq Bay	Luxury	Bangalore
6499	19563	01-Aug-22	RT1	23	30.0	NaN	Standard	Atliq Palace	Business	Bangalore
6500	19558	01-Aug-22	RT1	30	40.0	NaN	Standard	Atliq Grands	Luxury	Bangalore
6501	19560	01-Aug-22	RT1	20	26.0	NaN	Standard	Atliq City	Business	Bangalore
6502	17561	01-Aug-22	RT1	18	26.0	NaN	Standard	Atliq Blu	Luxury	Mumbai
6503	17564	01-Aug-22	RT1	10	16.0	NaN	Standard	Atliq Seasons	Business	Mumbai

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
latest_df.shape
```

```
(6504, 15)
```

6. Print revenue realized per city

```
df_bookings.head()
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_given
1	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others	NaN
4	May012216558RT15	16558	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online	5.0
5	May012216558RT16	16558	1/5/2022	1/5/2022	3/5/2022	2.0	RT1	others	4.0
6	May012216558RT17	16558	28-04-22	1/5/2022	6/5/2022	2.0	RT1	others	NaN
7	May012216558RT18	16558	26-04-22	1/5/2022	3/5/2022	2.0	RT1	logtrip	NaN

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_hotels.head()
```

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

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_bookings_all = pd.merge(df_bookings,df_hotels,on='property_id')
df_bookings_all.head(3)
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_given
0	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others	NaN
1	May012216558RT15	16558	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online	5.0
2	May012216558RT16	16558	1/5/2022	1/5/2022	3/5/2022	2.0	RT1	others	4.0

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_bookings_all.groupby('city')['revenue_generated'].sum()
```

```
city
Bangalore    494814675
Delhi        346350550
Hyderabad    381333250
Mumbai       784786980
Name: revenue_generated, dtype: int64
```

7. Print month by month revenue

```
df_bookings_all.head(3)
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_given
0	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others	NaN
1	May012216558RT15	16558	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online	5.0
2	May012216558RT16	16558	1/5/2022	1/5/2022	3/5/2022	2.0	RT1	others	4.0

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

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

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

```
df_bookings_all.info()
```

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

```
df_date.info()
```

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

```
df_date['date'] = pd.to_datetime(df_date['date'])
df_date.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 92 entries, 0 to 91
Data columns (total 4 columns):
 #   Column    Non-Null Count  Dtype         
---  ------    --------------  -----         
 0   date      92 non-null     datetime64[ns]
 1   mmm yy    92 non-null     object        
 2   week no   92 non-null     object        
 3   day_type  92 non-null     object        
dtypes: datetime64[ns](1), object(3)
memory usage: 3.0+ KB
```

```
df_bookings_all.head(3)
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_given
0	May012216558RT12	16558	30-04-22	1/5/2022	2/5/2022	2.0	RT1	others	NaN
1	May012216558RT15	16558	27-04-22	1/5/2022	2/5/2022	4.0	RT1	direct online	5.0
2	May012216558RT16	16558	1/5/2022	1/5/2022	3/5/2022	2.0	RT1	others	4.0

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_bookings_all.info()
```

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

```
df_bookings_all['check_in_date'] = pd.to_datetime(df_bookings_all['check_in_date'],format="mixed")
df_bookings_all.info()
```

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

```
dtypes: datetime64[ns](1), float64(2), int64(3), object(9)
memory usage: 15.4+ MB
```

```
df_bookings_all.head(3)
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_given
0	May012216558RT12	16558	30-04-22	2022-01-05	2/5/2022	2.0	RT1	others	NaN
1	May012216558RT15	16558	27-04-22	2022-01-05	2/5/2022	4.0	RT1	direct online	5.0
2	May012216558RT16	16558	1/5/2022	2022-01-05	3/5/2022	2.0	RT1	others	4.0

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_bookings_all = pd.merge(df_bookings_all,df_date,left_on='check_in_date',right_on ='date')
df_bookings_all.head(3)
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_given
0	May052216558RT11	16558	15-04-22	2022-05-05	7/5/2022	3.0	RT1	tripster	5.0
1	May052216558RT12	16558	30-04-22	2022-05-05	7/5/2022	2.0	RT1	others	NaN
2	May052216558RT13	16558	1/5/2022	2022-05-05	6/5/2022	3.0	RT1	direct offline	5.0

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

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

```
mmm yy
Jul 22    389940912
Jun 22    377191229
May 22    408375641
Name: revenue_realized, dtype: int64
```

8. Print revenue realized per hotel type

```
df_hotels.head(3)
```

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

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_bookings_all.head(3)
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_given
0	May052216558RT11	16558	15-04-22	2022-05-05	7/5/2022	3.0	RT1	tripster	5.0
1	May052216558RT12	16558	30-04-22	2022-05-05	7/5/2022	2.0	RT1	others	NaN
2	May052216558RT13	16558	1/5/2022	2022-05-05	6/5/2022	3.0	RT1	direct offline	5.0

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_revenue=pd.merge(df_bookings_all,df_hotels,on='property_id')
df_revenue.head(3)
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_given
0	May052216558RT11	16558	15-04-22	2022-05-05	7/5/2022	3.0	RT1	tripster	5.0
1	May052216558RT12	16558	30-04-22	2022-05-05	7/5/2022	2.0	RT1	others	NaN
2	May052216558RT13	16558	1/5/2022	2022-05-05	6/5/2022	3.0	RT1	direct offline	5.0

3 rows × 22 columns

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

```
df_revenue.columns
```

```
Index(['booking_id', 'property_id', 'booking_date', 'check_in_date',
       'checkout_date', 'no_guests', 'room_category', 'booking_platform',
       'ratings_given', 'booking_status', 'revenue_generated',
       'revenue_realized', 'property_name_x', 'category_x', 'city_x', 'date',
       'mmm yy', 'week no', 'day_type', 'property_name_y', 'category_y',
       'city_y'],
      dtype='object')
```

```
df_revenue.groupby('property_name_y')['revenue_realized'].sum()
```

```
property_name_y
Atliq Bay        179416721
Atliq Blu        179203544
Atliq City        196555383
Atliq Exotica     219076161
Atliq Grands      145860641
Atliq Palace      209474575
Atliq Seasons      45920757
Name: revenue_realized, dtype: int64
```

Exercise-2 Print average rating per city

```
df_bookings_all.head(3)
```

	booking_id	property_id	booking_date	check_in_date	checkout_date	no_guests	room_category	booking_platform	ratings_given
0	May052216558RT11	16558	15-04-22	2022-05-05	7/5/2022	3.0	RT1	tripster	5.0
1	May052216558RT12	16558	30-04-22	2022-05-05	7/5/2022	2.0	RT1	others	NaN
2	May052216558RT13	16558	1/5/2022	2022-05-05	6/5/2022	3.0	RT1	direct offline	5.0

<svg xmlns="http://www.w3.org/2000/svg" height="24px"viewBox="0 0 24 24" width="24px">

```
</div>
```

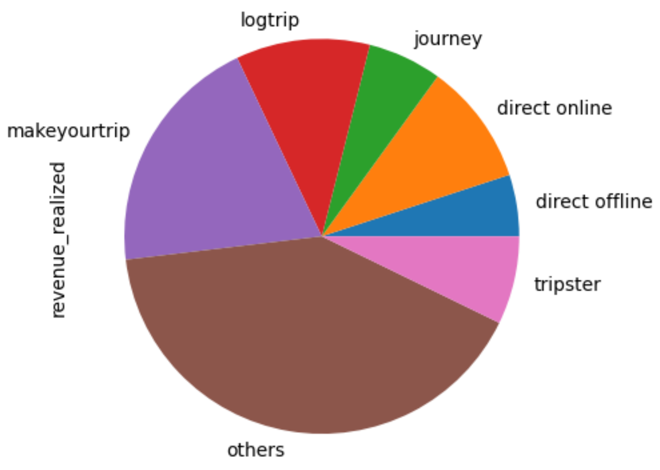
```
df_bookings_all.groupby('city')['ratings_given'].mean()
```

```
city
Bangalore    3.403911
Delhi        3.775088
Hyderabad    3.664286
Mumbai       3.644350
Name: ratings_given, dtype: float64
```

Exercise-3 Print a pie chart of revenue realized per booking platform

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

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


logtrip
journey
direct online
direct offline
tripster
others
makeyourtrip
revenue_realized