AtliQ Hospitality Analysis

Atliq Grands, a renowned player in the hospitality industry with multiple five-star hotels across India, has been experiencing a decline in market share and revenue within the luxury/business hotels category. The objective is to address the declining market share and revenue of Atliq Grands in the luxury/business hotels category by leveraging business and data intelligence. Specifically, the goal is to utilize historical data to gain actionable insights that can inform strategic decision-making and enable Atliq Grands to regain its market position.

In [1]:
```
# importing libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

In [2]:
```
dim_dt = pd.read_csv("dim_date.csv")
```

In [3]:
```
dim_dt.head()
```

Out[3]:

	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

In [9]:
```
dim_dt.shape
```

Out[9]:
```
(92, 4)
```

In [8]:
```
dim_dt.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
```

In [4]:
```
## Change date column type to date
dim_dt['date']=pd.to_datetime(dim_dt['date'])
```

In [5]:
```
dim_dt.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
```

In [6]:
```
##Checking missing values
dim_dt.isnull().sum()
```

Out[6]:
```
date        0
mmm yy      0
week no     0
day_type    0
dtype: int64
```

In [7]:
```
## no missing values found
```

In [8]:
```
dim_ht = pd.read_csv("dim_hotels.csv")
```

In [9]:
```
dim_ht.head()
```

Out[9]:

	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

In [11]:
```
dim_ht.shape
```

Out[11]:
```
(25, 4)
```

In [12]:
```
dim_ht.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 4 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   property_id    25 non-null     int64 
 1   property_name  25 non-null     object
 2   category       25 non-null     object
 3   city           25 non-null     object
dtypes: int64(1), object(3)
memory usage: 932.0+ bytes
```

In [34]:
```
##Checking missing values
dim_ht.isna().sum()
```

Out[34]:
```
property_id      0
property_name    0
category         0
city             0
dtype: int64
```

In [41]:
```
## no missing values found
```

In [10]:
```
dim_rm = pd.read_csv("dim_rooms.csv")
```

In [11]:
```
dim_rm.head()
```

Out[11]:

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

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

Out[12]:
```
(4, 2)
```

In [13]:
```
dim_rm.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4 entries, 0 to 3
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   room_id     4 non-null      object
 1   room_class  4 non-null      object
dtypes: object(2)
memory usage: 196.0+ bytes
```

In [14]:
```
##Checking missing values
dim_rm.isnull().sum()
```

Out[14]:
```
room_id       0
room_class    0
dtype: int64
```

In [15]:
```
## no missing values found
```

In [16]:
```
fact_agg_b = pd.read_csv("fact_aggregated_bookings.csv")
```

In [17]:
```
fact_agg_b.head()
```

Out[17]:

	property_id	check_in_date	room_category	successful_bookings	capacity
0	16559	01-May-22	RT1	25	30
1	19562	01-May-22	RT1	28	30
2	19563	01-May-22	RT1	23	30
3	17558	01-May-22	RT1	13	19
4	16558	01-May-22	RT1	18	19

In [17]:
```
fact_agg_b.shape
```

Out[17]:
```
(9200, 5)
```

In [18]:
```
fact_agg_b.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9200 entries, 0 to 9199
Data columns (total 5 columns):
 #   Column               Non-Null Count  Dtype 
---  ------               --------------  ----- 
 0   property_id          9200 non-null   int64 
 1   check_in_date        9200 non-null   object
 2   room_category        9200 non-null   object
 3   successful_bookings  9200 non-null   int64 
 4   capacity             9200 non-null   int64 
dtypes: int64(3), object(2)
memory usage: 359.5+ KB
```

In [18]:
```
## Change check_in_date data type to date
fact_agg_b['check_in_date']=pd.to_datetime(fact_agg_b['check_in_date'])
```

In [29]:
```
fact_agg_b.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9200 entries, 0 to 9199
Data columns (total 5 columns):
 #   Column               Non-Null Count  Dtype         
---  ------               --------------  -----         
 0   property_id          9200 non-null   int64         
 1   check_in_date        9200 non-null   datetime64[ns]
 2   room_category        9200 non-null   object        
 3   successful_bookings  9200 non-null   int64         
 4   capacity             9200 non-null   int64         
dtypes: datetime64[ns](1), int64(3), object(1)
memory usage: 359.5+ KB
```

In [37]:
```
##Checking missing values
fact_agg_b.isnull().sum()
```

Out[37]:
```
property_id            0
check_in_date          0
room_category          0
successful_bookings    0
capacity               0
dtype: int64
```

In [43]:
```
## no missing values found
```

In [ ]:
```
## calculating ccupancy % ( Ratio of Total Successful Bookings to Total Capacity)
```

In [49]:
```
fact_agg_b["occ_per"] = (fact_agg_b['successful_bookings'] / fact_agg_b['capacity']) *100
```

In [50]:
```
fact_agg_b.head()
```

Out[50]:

	property_id	check_in_date	room_category	successful_bookings	capacity	occ_per
0	16559	2022-05-01	RT1	25	30	83.333333
1	19562	2022-05-01	RT1	28	30	93.333333
2	19563	2022-05-01	RT1	23	30	76.666667
3	17558	2022-05-01	RT1	13	19	68.421053
4	16558	2022-05-01	RT1	18	19	94.736842

In [22]:
```
# reading the data
Atliq_fact_b = pd.read_csv("fact_bookings.csv")
```

In [23]:
```
Atliq_fact_b.head()
```

Out[23]:

	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	2022-04-27	2022-05-01	2022-05-02	3	RT1	direct online	1.0	Checked Out	10010	10010
1	May012216558RT12	16558	2022-04-30	2022-05-01	2022-05-02	2	RT1	others	NaN	Cancelled	9100	3640
2	May012216558RT13	16558	2022-04-28	2022-05-01	2022-05-04	2	RT1	logtrip	5.0	Checked Out	9100	9100
3	May012216558RT14	16558	2022-04-28	2022-05-01	2022-05-02	2	RT1	others	NaN	Cancelled	9100	3640
4	May012216558RT15	16558	2022-04-27	2022-05-01	2022-05-02	4	RT1	direct online	5.0	Checked Out	10920	10920

In [24]:
```
Atliq_fact_b.shape
```

Out[24]:
```
(134590, 12)
```

In [25]:
```
Atliq_fact_b.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134590 entries, 0 to 134589
Data columns (total 12 columns):
 #   Column             Non-Null Count   Dtype  
---  ------             --------------   -----  
 0   booking_id         134590 non-null  object 
 1   property_id        134590 non-null  int64  
 2   booking_date       134590 non-null  object 
 3   check_in_date      134590 non-null  object 
 4   checkout_date      134590 non-null  object 
 5   no_guests          134590 non-null  int64  
 6   room_category      134590 non-null  object 
 7   booking_platform   134590 non-null  object 
 8   ratings_given      56683 non-null   float64
 9   booking_status     134590 non-null  object 
 10  revenue_generated  134590 non-null  int64  
 11  revenue_realized   134590 non-null  int64  
dtypes: float64(1), int64(4), object(7)
memory usage: 12.3+ MB
```

In [30]:
```
##Cahnge booking_date, check_in_date, checkout_date data type to date
Atliq_fact_b['booking_date']=pd.to_datetime(Atliq_fact_b['booking_date'])
Atliq_fact_b['check_in_date']=pd.to_datetime(Atliq_fact_b['check_in_date'])
Atliq_fact_b['checkout_date']=pd.to_datetime(Atliq_fact_b['checkout_date'])
```

In [31]:
```
Atliq_fact_b.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 134590 entries, 0 to 134589
Data columns (total 12 columns):
 #   Column             Non-Null Count   Dtype         
---  ------             --------------   -----         
 0   booking_id         134590 non-null  object        
 1   property_id        134590 non-null  int64         
 2   booking_date       134590 non-null  datetime64[ns]
 3   check_in_date      134590 non-null  datetime64[ns]
 4   checkout_date      134590 non-null  datetime64[ns]
 5   no_guests          134590 non-null  int64         
 6   room_category      134590 non-null  object        
 7   booking_platform   134590 non-null  object        
 8   ratings_given      56683 non-null   float64       
 9   booking_status     134590 non-null  object        
 10  revenue_generated  134590 non-null  int64         
 11  revenue_realized   134590 non-null  int64         
dtypes: datetime64[ns](3), float64(1), int64(4), object(4)
memory usage: 12.3+ MB
```

In [39]:
```
##Checking missing values
Atliq_fact_b.isnull().sum()
```

Out[39]:
```
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        77907
booking_status           0
revenue_generated        0
revenue_realized         0
dtype: int64
```

In [48]:
```
Atliq_fact_b.isnull().mean()*100
```

Out[48]:
```
booking_id            0.000000
property_id           0.000000
booking_date          0.000000
check_in_date         0.000000
checkout_date         0.000000
no_guests             0.000000
room_category         0.000000
booking_platform      0.000000
ratings_given        57.884687
booking_status        0.000000
revenue_generated     0.000000
revenue_realized      0.000000
dtype: float64
```

In [ ]:
```
##57.88% 0f missing values found in 'ratings_given'
```

In [46]:
```
Atliq_fact_b['ratings_given'].value_counts()
```

Out[46]:
```
ratings_given
5.0    19480
3.0    17561
4.0     9653
2.0     6452
1.0     3537
Name: count, dtype: int64
```

In [ ]:
```
## Not removing or replacing na values because, customers may or may not give ratings.
```

In [ ]:
```
## Saving the cleaned and transformed csvs for further analysis and visualization in Power BI
```

In [51]:
```
dim_dt.to_csv('dim_dt.csv')
```

In [52]:
```
dim_ht.to_csv('dim_ht.csv')
```

In [53]:
```
dim_rm.to_csv('dim_rm.csv')
```

In [54]:
```
fact_agg_b.to_csv('fact_agg_b.csv')
```

In [55]:
```
Atliq_fact_b.to_csv('Atliq_fact_b.csv')
```

In [ ]: