Project – Hotel Data Analytics

About Project

Atliq Grands, a prestigious hotel chain in India, has been serving guests for over 20 years. Their hotels are located in major cities including Delhi, Mumbai, Bangalore, and Hyderabad. They offer a variety of hotel experiences under different brands such as Atliq Seasons, Atliq Exotica, Atliq Bay, and Atliq Palace and so on.

Guests can choose from a range of room types to suit their needs, including Standard, Elite, Premium, and Presidential. Bookings can be made through their own website, as well as popular platforms like MakeYourTrip, LogTrip, Tripster, or via offline methods. All bookings, regardless of the method, are consolidated into the Atliq Grands booking database, which is utilized for comprehensive data analysis and service optimization.

Problem Statement

Atliq Grands, a leading hotel chain in India, is currently facing significant challenges from competitors, leading to a decline in both revenue and market share. Despite their long-standing presence in cities such as Delhi, Mumbai, Bangalore, and Hyderabad, increased competition has impacted their financial performance. The company is actively seeking strategies to regain its market position and enhance its profitability.

Steps to Solve the Problem

Step 1: Understand Business Problem

Step 2: Data collection and data understanding

Step 3: Data Cleaning and Exploration

Step 4: Data Transformation

Step 5: Insight Generation

Data Collection and Data Understanding

Downloading Dataset from Kaggle -

https://www.kaggle.com/datasets/ad043santhoshs/hospitality-domain

There are total 6 excel files –

1. **Dim_date** – date metadata
2. **Dim_hotels** – all hotels that atliq grands company has
3. **Dim_rooms** – all room types
4. **Fact_aggregated_bookings** – contains successful bookings on a particular date of a particular property id and room type and capacity of that property
5. **Fact_bookings** – transactions data
6. **New_data_august** – august month data (additional)

Fact Table: Stores numbers related to business activities, like sales or bookings. (in fact table the information changes continuously)

Dimension Table: Holds descriptive details about the facts, such as customer information, product names, or dates.

(in dimension table the information doesn't changes that often)

Data Exploration

Here, we summarize statistics, detect anomalies, and visualize distributions to understand patterns and relationships within the dataset. Basically trying to make sense of out all the available data

Data Cleaning

1. **Negative Values -** By using describe function I found out that there was guest count in negatives (no_guest column) – to find all the negative values, I just used the condition to show all the rows of no_guest columns which has values less than 0
   Solution – there was 12 rows or 12 negative values so I thought we should get rid of it, as we have total 1,34,590 records, so if we ignore 12, it doesn't really matters
   How? Just created a new variable and stored the condition to show only those records which contains value greater than 0 (for no_guest column)

2. **Outlier in revenue generated column –**
   The revenue generated maximum count was 28560000 and its not a realistic amount that anyone can pay for a single night stay

   Solution – used 3 standard deviation technique to remove this outlier, by using this I got a value ie. 294498.501 and apply a condition to display those rows which contains value greater than the higher_limit in revenue_generated. And now show those rows which has values less than the higher limit so that I can do further analysis. I prefer this method to ignore the outliers

3. **Got Null values in ratings_given column –** but its not that efficient or important to handle those missing values as it's a purely customer behavior based data and cannot be solve technically as customers can be lazy for giving reviews even though they enjoyed their stay, sometimes customers may forget to give ratings or there could be any other reasons. So the best practice could be just to avoid the ratings_given column as it will not impact our analysis process

Data Exploration

By considering 2 columns (successful_bookings and capacity) we will find out the occupancy rate and will add to newly created column (occ_pct) [occupancy percentage]

Occ_pct = Successful_bookings / capacity

Values will be in percentage form

Insights Generation

Found out the answer to the following questions:

1. What is the average occupancy rate in each of the room categories?
2. When was the occupancy better? Weekday or weekend?
3. In the month of June, what is the occupancy for different cities
4. What was the revenue realized per city?
5. What was the revenue of each month?