**HOTEL BOOKING ANALYSIS**

In this document, we embark on a comprehensive analysis of hotel booking patterns, guest preferences, and the factors that influence cancellations. Leveraging the power of SQL, Excel, and Power BI, our aim is to uncover valuable insights that will aid in optimizing hotel operations and enhancing decision-making processes. This analysis is based on a rich dataset encompassing various facets of hotel bookings, guest information, meal preferences, booking sources, room details, and reservation statuses.

The dataset is structured into several tables, each serving a specific purpose in our analysis:

**Room\_Details:** This table provides insights into room reservations and modifications. It sheds light on the type of room initially reserved, the room eventually assigned, and the number of booking changes. These details are instrumental in understanding room allocation dynamics and booking adjustments.

**Reservation\_Status:** This table records the evolving statuses of reservations over time. Linked to the Booking\_Details table, it captures the reservation's last status, such as "Canceled" or "Check-Out," and the date when this status was recorded. This historical data aids in tracking the progression of reservations and their final outcomes.

**Booking\_Details:** The heart of our analysis, this table contains vital information about hotel reservations. It includes a unique booking identifier, hotel type (Resort Hotel or City Hotel), cancellation status, lead time (the number of days between booking and arrival), arrival year, month, week number, and day of the month. Furthermore, it provides details about the number of weekend and weekday nights stayed.

**Guest\_Info:** To understand the composition of guests for each reservation, this table records the number of adults, children, and babies accompanying the booking. These insights help us comprehend the guest demographics and preferences.

**Meal\_And\_Stay\_Details:** This table complements booking information by specifying meal-related and stay-related attributes. It includes the type of meal booked, the Average Daily Rate (ADR) for the stay, the number of required car parking spaces, and the total count of special requests made by guests. These details illuminate guest preferences and service requirements.

**Booking\_Source\_and\_History:** Crucial for understanding booking sources and guest behaviour, this table provides information about market segments, distribution channels, and whether the guest is a repeated visitor. It also records data on previous booking cancellations, deposit types, booking agents, company IDs, waiting list duration, and customer types.

With this diverse and comprehensive dataset, we will embark on a journey to answer critical questions, identify trends, and visualize data through Power BI dashboards. Our objective is to empower hotel management with data-driven insights that enable informed decision-making, enhance guest experiences, and optimize operations. Through structured analysis and the application of suitable MECE (Mutually Exclusive, Collectively Exhaustive) methods, we aim to uncover actionable insights that can shape the future of hotel management and hospitality.

**MECE APPROACH**

The tasks outlined in the analysis can be categorized using the MECE (Mutually Exclusive, Collectively Exhaustive) framework, which ensures that each task falls into a distinct category while collectively covering all aspects of the analysis. Here's how they align with the MECE framework:

**BOOKING PATTERNS and TRENDS:**

1. Analyse booking data to identify trends in the number of bookings over time.

2. Categorize bookings by hotel type (Resort Hotel and City Hotel).

3. Analyse seasonal variations in booking patterns by week number and day of the week.

4. Calculate the average lead time for booking.

**GUEST PREFERENCE:**

1. Analyse the type of meal booked.

2. Examine special requests made by guests and their frequency.

3. Analyse the composition of guests in terms of the number of adults, children, and babies.

4. Analyse the car parking space requirement.

**CANCELLATION ANALYSIS:**

1. Analyse the impact of deposit type on cancellations.

2. Calculate the cancellation rate and identify factors influencing cancellations.

3. Examine the relationship between booking agent IDs and cancellations.

4. Determine if repeated guests have a lower cancellation rate.

**ROOM ALLOCATION and MODIFICATION:**

1. Analyse room allocation dynamics.

2. Calculate the number of changes made to bookings and their impact on revenue.

3. Investigate the impact of room modification on guest satisfaction.

BOOKING SOURCES and REVENUE:

1. Analyse revenues generated from different sources.

2. Explore the historical behaviour of guests.

3. Investigate the influence of customers on revenue.

4. Calculate revenue per available room and revenue per guest.

This MECE framework ensures that all aspects of the analysis are covered, with each task falling into one of the five categories. It helps in organizing and structuring the analysis process, making it easier to derive insights and recommendations from the data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BOOKING PATTERNS and TRENDS | GUEST PREFERENCE | CANCELLATION ANALYSIS | ROOM ALLOCATION and MODIFICATION | BOOKING SOURCES and REVENUE |
| Analyse booking data to identify trends in the no of booking over time | Analyse type of meal booked | Analyse the impact of deposit type | Analyse room allocation dynamics | Analyse revenues generated from different sourness |
| Categorize booking by hotel type (resorts hotel and city hotel) | Examine special requests made by guests and their frequency | Calculating the cancellation rate and identify factors influencing cancellation | Calculate the no of changes made to booking and their impact on revenue | Explore the historical behaviour of guest |
| Analyse seasons variations in booking patterns (by week number and day of week | Analysing the composition of guest (no of adults, children and babies) | Explore the relationship between booking agent Ids and cancellation | Investigate the impact of room modification on guest satisfaction | Investigate the influence of customers on revenue. |
| Calculate the average lead time for booking | Analyse the car parking space requirement | Determine if repeated guests have a lower cancellation rate |  | Calculate revenue per available room and revenue per guest. |

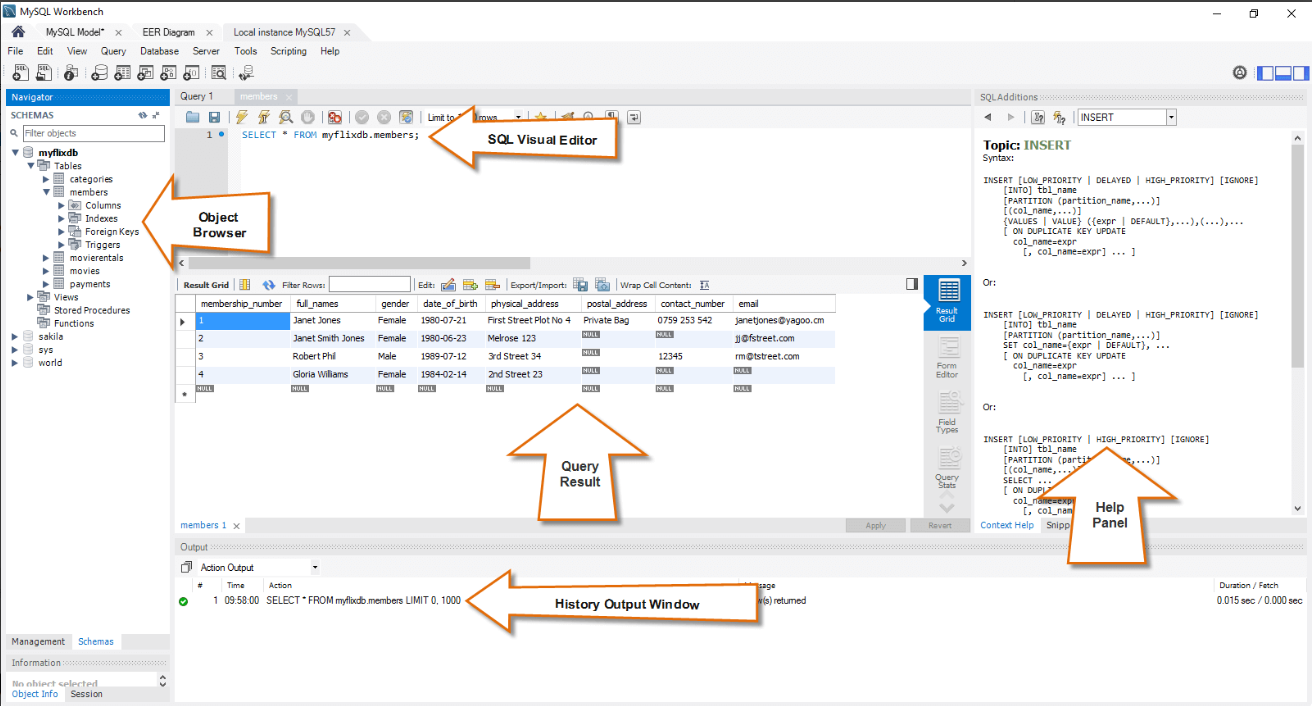
**SETTING UP THE WORKSPACE**

Visit the official MySQL website to download the MySQL installer. You can find it at https://dev.mysql.com/downloads/install. Select the appropriate version for your operating system (Windows, macOS, or Linux) and follow the download instructions.

Run the downloaded installer and follow the installation wizard's instructions. During the installation, you will be prompted to set a root password for MySQL. Make sure to remember this password as it will be needed to access MySQL later.

After a successful installation, you can access MySQL Workbench, which is a graphical user interface for managing MySQL databases.

Launch MySQL Workbench: You can typically find the MySQL Workbench application in your program/application menu. Launch it.

****

**MySQL Workbench Interface:**

**Menu Bar:** The menu bar at the top contains various options for managing and configuring MySQL.

**Object Browser Section:** On the left-hand side, you will find the object browser section. Here, you can see a list of databases, connections, and server management options.

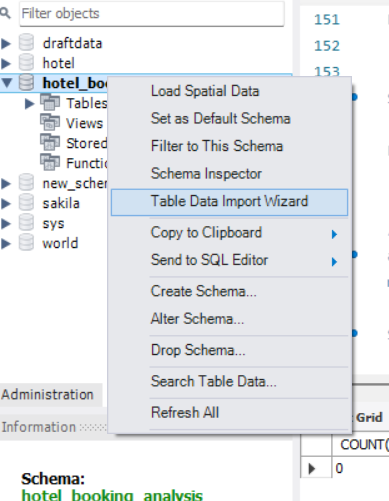
**Query Section:** In the centre, you'll have the Query section. This is where you can write and execute SQL queries against your databases.

**Output Section:** At the bottom, you will find the Output section. This area displays the results and outputs of your SQL queries, including any error messages or query results.

**Connect to a Database:** To get started, you will need to establish a connection to a MySQL server. Click on the "Local Instance 3306" (or similar) in the Navigator section to connect to your local MySQL server.

**Create or Manage Databases:** You can create new databases, tables, and manage your database schema using the Navigator and Query sections.

In the query section, you can execute the following SQL query: "CREATE DATABASE Hotel\_booking\_analysis." This command creates a new database, which you can then locate in the object browser section. By right-clicking on the newly created "Hotel\_booking\_analysis" database, a menu bar will appear, providing access to the "Table Data Import Wizard." This wizard allows you to efficiently import raw data from CSV files one by one, with the duration of the process depending on factors such as the data size and internet speed.



**DATA CLEANING**

Data cleaning is an essential step in preparing the dataset for analysis. In our dataset, which comprises 119,390 rows and 32 columns, we meticulously carried out the following data cleaning operations:

1. Duplicates Detection: We initially checked for duplicate rows within the dataset, and it is noteworthy that no duplicates were found. Each row represents a unique entry.

|  |  |
| --- | --- |
|  |  |

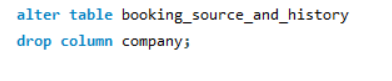
2. Handling Blanks and Replacements: We identified and addressed instances of missing or blank data. Specifically, the "country" column had 488 blank entries, which we replaced with the value "OTHR" to indicate "Other" as an appropriate placeholder. Additionally, the "agent" column had 16,340 missing values, which were replaced with the most frequently occurring values to maintain data integrity.

|  |  |
| --- | --- |
|  | This Gives the count Null data for the country table. You do this with every table to see the blanks and replace them. |
|  | This replaces the blanks with “OTHR” |

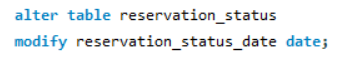
3. Insertion of Missing Values: To ensure completeness and consistency in our dataset, we introduced the "country" values "OTHR" and "Other" into the respective column, thereby providing a comprehensive representation of countries.

|  |  |
| --- | --- |
|  | This adds the new row to the country table |

4. Drop Column: The "company" column presented a unique challenge, as it contained 112,593 blank rows, accounting for more than 75% of the data. Given its extensive lack of data and negligible utility in our analysis, we made the decision to drop this column entirely from the dataset.

this deletes the column company

5. Data Type Adjustment: We observed that certain columns required adjustments in their data types for accurate analysis. The "children" column, originally of float data type, was changed to an integer data type. Similarly, the "adr" (Average Daily Rate) column, also initially of float data type, was converted to an integer data type for consistency. Lastly, the "Reservation\_date" column, which was initially categorized as text data, was transformed into a date data type to facilitate time-based analysis.

this changes the data type of the column reservation statues date to date datatype.

By meticulously performing these data cleaning steps, we ensured that our dataset is now well-prepared for robust analysis, free from duplicates, with appropriately handled missing values, and containing consistent data types throughout. This ensures the reliability and accuracy of the insights and conclusions derived from our subsequent analyses.

**EXPLORATORY DATA ANALYSIS**

Exploratory Data Analysis (EDA) plays a pivotal role in this analysis, aiming to unveil key insights and trends within the dataset related to booking patterns, guest preferences, and factors influencing cancellations. This thorough EDA session has been instrumental in addressing critical questions posed by the dataset.

SQL queries have been our primary analytical tool throughout this process. They have provided direct solutions to various EDA questions, enabling us to extract valuable information promptly. However, in some instances, we've gone a step further by leveraging Excel to refine and delve deeper into the SQL outputs. This combined approach has empowered us to gain a more profound understanding of the data, identify intricate patterns, and highlight critical trends.

The strategic use of filters and pivot tables has significantly enhanced our ability to visualize the dataset, enabling us to identify and address any potential drawbacks efficiently. This comprehensive EDA approach ensures that we harness the full potential of our data to inform decision-making and drive future development strategies effectively.

Here are the few question of EDA,

Understand the distribution of arrival dates, including the most common arrival days and summary statistics for lead times.

The first part of the question involves the identifying the most no of booking made which date and the second part includes the statistics such as the mean, median, and standard deviation.

It is observed for the output that the 17th of the months is having the higher booking or common booking date as 17th.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

Analyse Average Daily Rates (ADR) by meal plan type to identify variations in pricing.

The output shows the average daily rates of different meal plan types, there is no specific pricing data for the meals themselves.

The data indicates the ADR associated with different meal plans, with “HB” having the higher ADR at 120.30, followed by “FB” 109.05 and “BB” at 99.4017.

|  |  |
| --- | --- |
|  |  |

Analyse trends in reservation status dates, including the most common checkout dates and any seasonality patterns.

The output data presents reservation statues dates, including the most common checkout dates and their frequencies and there month and their respective statues count

The list of checkout dates and their frequencies indicates the most common dates when guest tend to check out of the hotel.

|  |  |
| --- | --- |
|  |  |
|  |  |

Compare the distribution of meal plans among different customer types (e.g., Transient, Group) to identify preferences.

The out data shows the distribution of meal plan among the different customer types, along with the total number of special requests made by each customer type.

* Transient customers having the highest total number of special requests, this indicates that transient groups are more likely to make special request related to their meals.
* Transient party customers come next where their count is substantially lower than transient customers.
* The count of the contract customers is significantly lower than the above two, indicating that contract customers is having fewer preferences for specific meal plans.
* Group customers have the lowest count of total special request.

Understanding these preferences can help hotels tailor their meal offerings and improve guest satisfaction for different customer segments.

|  |  |
| --- | --- |
|  |  |

Identify peak booking months and analyse reasons for spikes in bookings, including holidays or events.

The output shows the number of bookings for each month and year, while we don’t have the data holidays or events.

Peak booking month, in the output the month with the highest no. of booking indicates the peak seasons for hotel reservations.

May, October, April and June in 2016 and 2017 consistently appear as the months with high booking number, these months might represent peak tourist seasons, favourable weather or other factors that attract travellers.

|  |  |
| --- | --- |
|  |  |

Explore how reservation statuses vary across different customer types (e.g., Transient, Group) using Excel or SQL. Calculate cancellation rates by customer type.

This output presents the reservation statuses for different customer types, along with the corresponding counts, and the cancellation rates for each customer type are calculated.

* Transient customers have the highest cancellation rate among all customers types, indicating a higher likelihood of cancellation in the group.
* Group customers have the lowest cancellation rate, suggesting that they are more committers to their reservations.
* Contract customers fall in between with a moderate cancellation rate.
* Transient party customers also have a relatively high cancellations rate, through lower than transient customers.

Understanding these cancellation patterns can help hotels tailor their booking and cancellation policies to different customers segments, potentially reducing cancellations and optimizing revenue management.

|  |  |
| --- | --- |
|  |  |
|  |  |

Analyse the distribution of required car parking spaces for each hotel type and determine if one type attracts more guests with cars.

The data showcases the distribution of required car parking spaces for two types of hotels, city hotel and resort hotel.

* City hotel accommodates a significantly larger number od guests who do not require car parking spaces, suggesting that guests at city hotel might be less likely to travel with cars or prefer alternative transportation options.
* Resort hotel on the other hand caters to mix guest, including those who request one car parking space, this could indicate that some guests at resort hotel prefer to travel with cars, potentially due to its location or the nature of the resort experience,
* While there a few cases of guest requesting more than one car parking space, these instances are relatively rare in both types of hotels.

Understanding the parking prefernces of guests at different hotel types can assist in optimizing parking facilities and services as well as tailoring marketing strategies to specific guest needs,

|  |  |
| --- | --- |
|  |  |

Understand the distribution of meal plans (e.g., BB, HB, FB, SC) and identify any patterns or preferences.

|  |  |
| --- | --- |
|  |  |

* Bed & Breakfast is the most popular choice, indicating that many guests prefer to have breakfast included in their reservation while having the flexibility to manage other meals independently.
* Half Board is the second most favoured option, suggesting that guests appreciate having both breakfast and dinner as part of their stay.
* Self Catering appeals to a segment of guests who prefer to cook their meals or explore dining options outside the hotel.
* Full Board is the least selected option, indicating that fewer guests opt for all-inclusive meal plans that cover all three meals.

Understanding meal plan preferences can help hotels tailor their offerings and marketing strategies to meet the diverse dining preferences of their guests, ultimately enhancing guest satisfaction and experience.

**POWER BI**

The Power BI dashboard is a dynamic and visually engaging tool that empowers businesses to make data-driven decisions with ease. It offers a user-friendly interface for data analysis, enabling users to transform raw data into meaningful insights. With customizable reports, interactive visuals, and seamless integration with various data sources, Power BI facilitates the exploration of complex datasets. It allows users to uncover trends, anomalies, and patterns, making it a valuable asset for optimizing operations, understanding customer preferences, and enhancing overall business performance.

Moreover, Power BI's real-time updating capabilities ensure that decision-makers have access to the most current information, while its sharing and collaboration features facilitate effective communication across teams. The ability to create compelling dashboards and reports using Power BI transforms data into actionable knowledge, making it an essential tool for organizations striving to stay competitive in today's data-driven landscape.

To load the data, follow these steps:

1. Launch Power BI.

2. Click on "Get Data."

3. In the dialog box that appears, select "More" to access additional data sources.

4. Search for "MariaDB" and select it.

5. Provide the necessary connection details, including the server information, database name, username, and password.

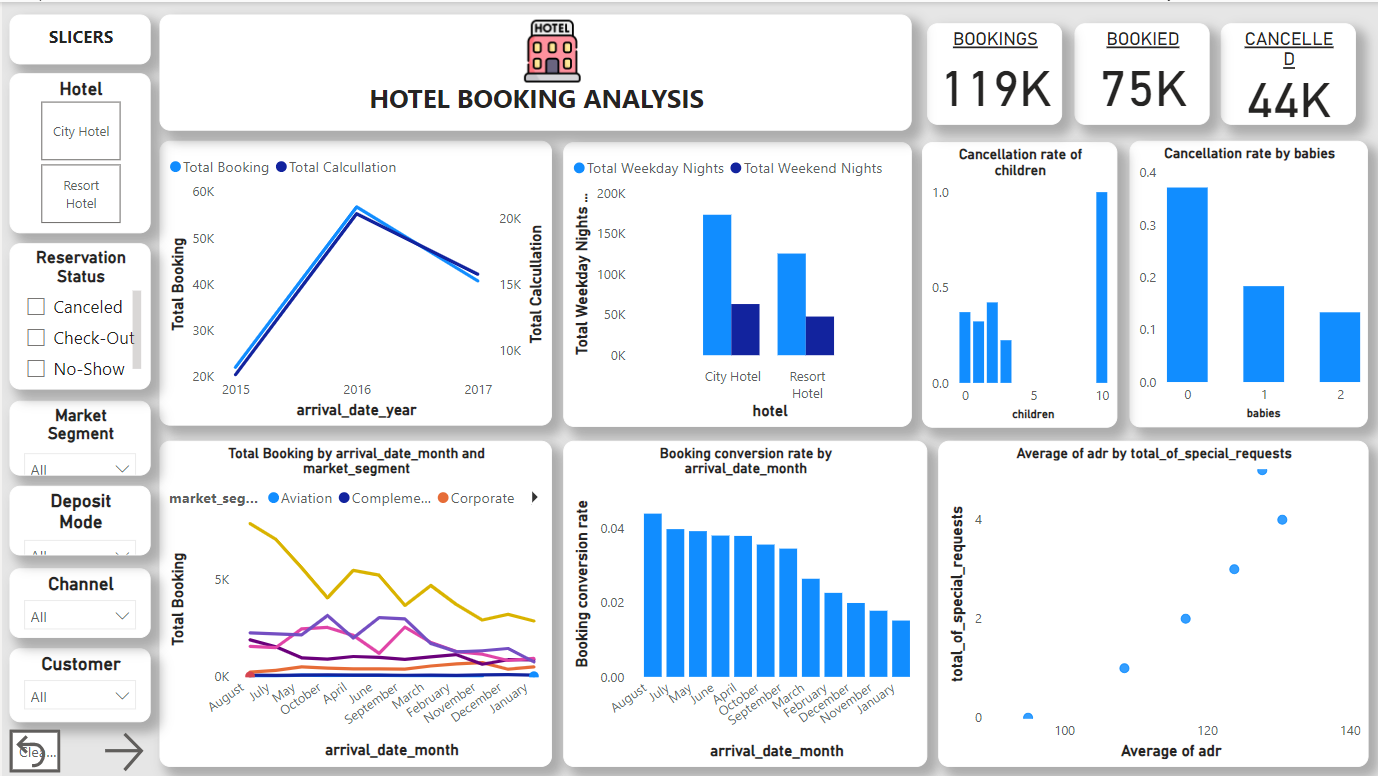
6. After connecting to the MariaDB database, choose the cleaned data stored in CSV files for import.

This process ensures that your data is seamlessly loaded into Power BI for further analysis and visualization. After the loading the data we are ready to go for preparing the dashboard or the report.

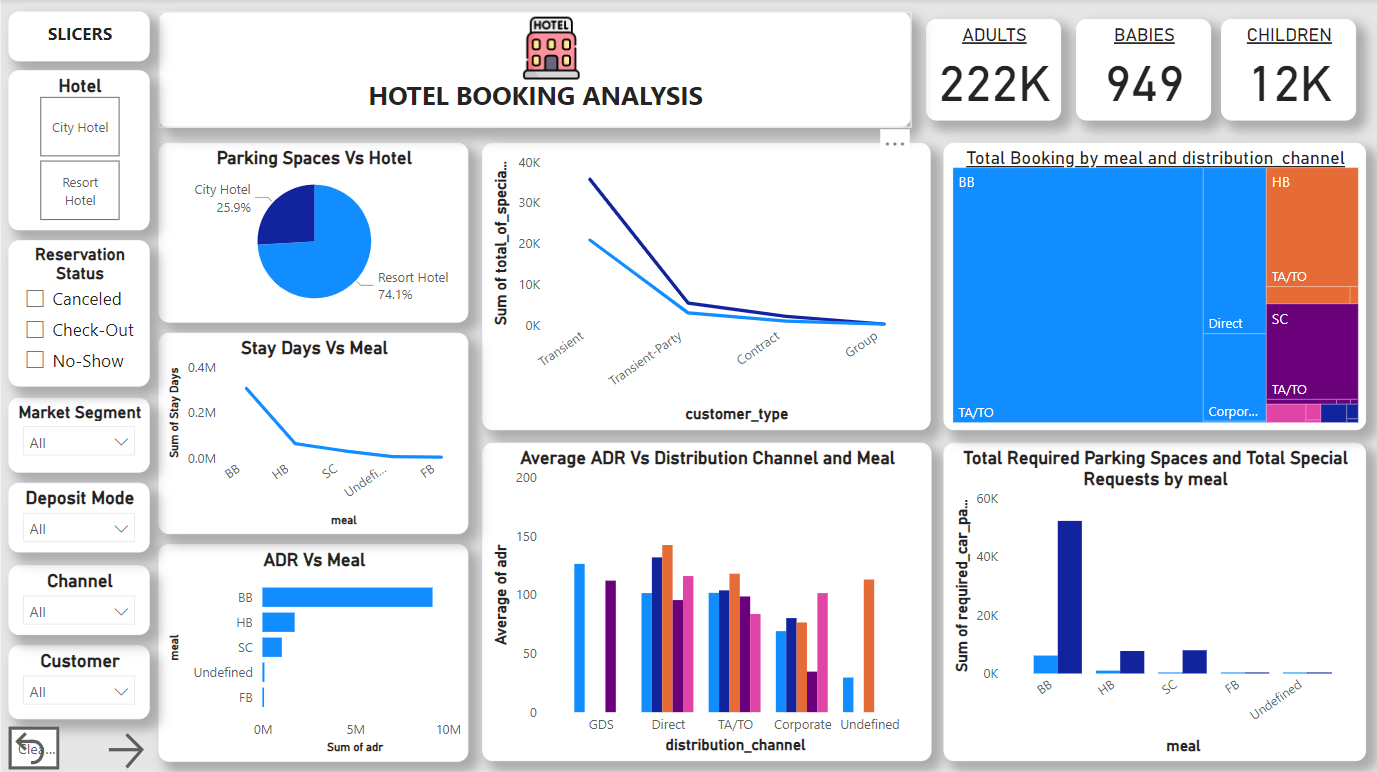
Interactive reports harness the versatility of sliders, offering a powerful tool to empower users in data-driven decision-making. Thoughtfully designed sliders provide users with precision in data filtering and manipulation, resulting in a streamlined decision-making process. These interactive elements not only foster user engagement but also grant users greater control over data exploration, leading to more intuitive and insightful insights.

In addition to sliders, the inclusion of navigation buttons is pivotal in enhancing the interactive report's usability. These buttons greatly improve the user experience by simplifying transitions between various report sections or slides. This seamless and intuitive navigation ensures that users can effortlessly access and digest critical insights, ultimately reinforcing the decision-making process. Interactive reports, with their dynamic sliders and user-friendly navigation, emerge as invaluable tools for those making data-driven decisions.

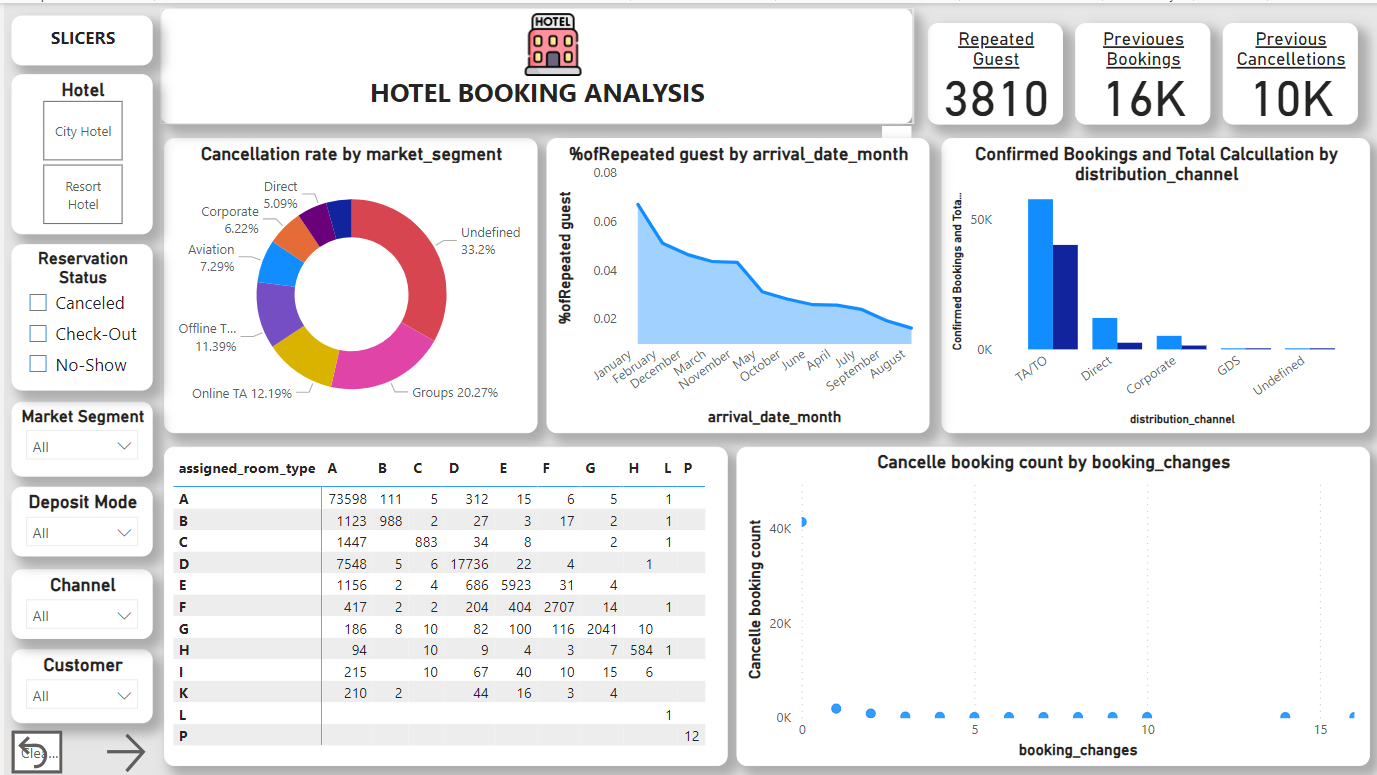
Here I have mentioned the 4 slides of the dashboard



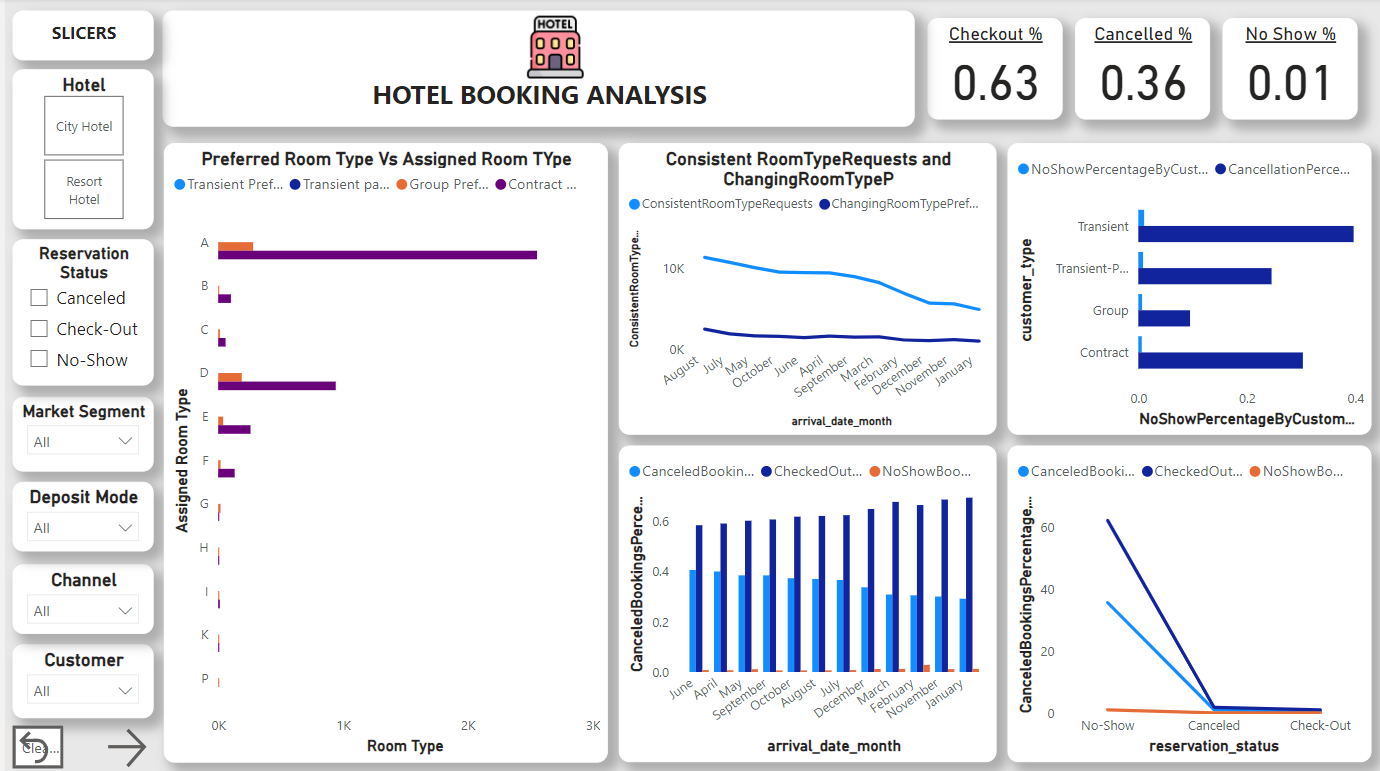
The initial slide of the Power BI dashboard presents a comprehensive overview, featuring total bookings, confirmed reservations, and cancellations. It also incorporates exploratory data analysis (EDA) visuals that offer insights into the dataset. Additionally, slicers are thoughtfully included to provide users with a holistic perspective and interactive control over the data presentation.



This dashboard slide showcases key guest details, including counts of adults, children, and babies. It also addresses critical EDA questions regarding required car parking spaces, total special requests, meal preferences, and Average Daily Rate (ADR). Impressively, it maintains consistency with the slicers introduced in the initial slide, ensuring a seamless and user-friendly experience throughout the presentation.



Slide three of our dashboards offers a comprehensive view of guest information. It sheds light on whether guests are repeat visitors, their historical booking behaviour (including cancellations), and the booking sources they've utilized. Additionally, we delve into customer types and details about the rooms initially reserved and eventually assigned. This cohesive approach allows for a deeper understanding of guest dynamics, historical patterns, and preferences.



Our final slide in the dashboard presents a vital overview of reservation statuses, highlighting the percentages of checkouts, cancellations, and no-shows.

This encapsulates the ultimate outcome of hotel reservations. In essence, our holistic hotel booking analysis dashboard offers data-driven insights that serve as a bedrock for informed decision-making and development initiatives.

These insights enable us to steer our strategies in the right direction and optimize our operations for enhanced guest experiences and overall success.

**INSIGHTS**

* July august are the busier and most profitable months for both hotels.
* City hotels account for approximately 60% of bookings, making them the busier option compared to resort hotels, which represent around 40% of reservations. The city hotel Average daily rate ADR is higher than resort hotel.
* Guests tend to Favor resort hotels for extended stays and opt for city hotels for shorter durations. The most preferred booking channel is TA/TO.
* Interestingly, there are more guests with cars at City hotels (1,926 guests requiring parking spaces) compared to Resort hotels (5,465 guests requiring parking spaces). This suggests that City hotels tend to attract more guests who have vehicles.
* While Transient guests make the most requests, Transient-Party and Contract customers also contribute to the total count, with 8,274 and 2,971 requests, respectively.
* Repeated guests tend to make more special requests, as indicated by higher values in the "total\_of\_special\_requests" column for repeated guest bookings across different distribution channels and market segments.
* "Direct" and "Corporate" channels also have a notable presence of repeated guests, with 3,810 and 1,572 bookings, respectively.
* The "Groups" market segment stands out with a substantial count of repeated guests at 11,580 bookings, indicating its attractiveness to guests who have stayed before.
* In 2017, there was a decrease in both bookings and cancellations compared to 2016, with bookings down by approximately 28.25% and cancellations down by 22.58%. This indicates a reduction in activity from 2016 to 2017.
* Guests with "3 or More Previous Cancellations" have a lower cancellation rate (approximately 59.94%) than those with "1-2 Previous Cancellations," suggesting a shift towards more committed booking behaviour over time.
* Cancellation rates vary notably by customer type, with Transient bookings having the highest rate (~66.96%) and Contract bookings the lowest (~10.62%), indicating a higher likelihood of cancellations for Transient reservations.
* The majority of reservations show consistency between the reserved and assigned room types, as indicated by the "Consistent" label. For example, there are 104,473 instances where both reserved and assigned room types are "A," and 14,917 cases of inconsistency.