

Capstone Project

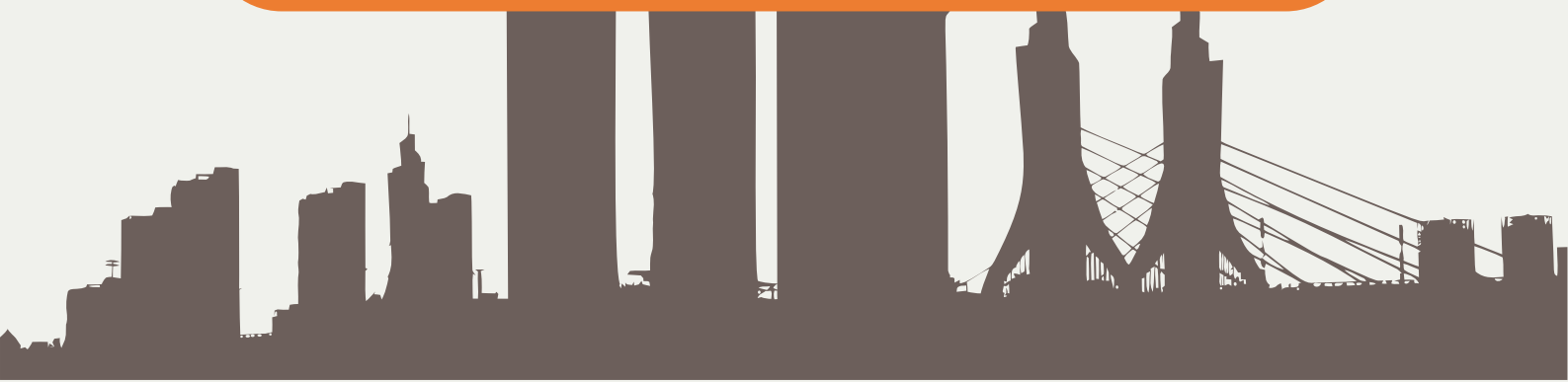
Hotel Booking Analysis

Russel Gabriel Thomas



OVERVIEW

The dataset contains a wealth of information concerning hotels spanning multiple countries. the dataset comprises the information of the booking details, the stay details along with reservation status, Customer history. Additionally it comprises all the sources, segments and market of booking



THE PROCESS

1 : Data Acquisition from GitHub

Obtain the requisite dataset from a designed Github repository, containing essential information on hotel bookings, encompassing various countries and their variations

2 : Data transformation & Enhancement

If necessary, execute data transformation procedures to ensure data quality and consistency. Additionally, consider augmenting the dataset with new problems statements to enrich the analysis potential.

3 : Connecting with Tools

Establish connections between the dataset and various analytical tools. Interface the dataset with Power BI, Excel, and MySQL Workbench, facilitating seamless data integration and processing.

4 : Problem Statement Solution in Power BI

Utilize Power BI to delve into the specified problem statements. Employ its robust features for data visualization exploration & analysis effectively deriving insights & solutions.

5 : Exploratory Data Analysis (EDA)

Perform exploratory data analysis using either Excel or SQL Workbench, depending on the complexity of the analysis. Extract meaningful patterns, relationships, and trends from the data to inform subsequent decision making.

6 : Detailed Documentation

Compile a detailed report that meticulously documents the entire project lifecycle. Include sections on data collection, transformation, problem statement formulation, tools integration, Power BI solutions, EDA insights, and Powerpoint visualizations.

OBJECTIVE

To conduct a comprehensive analysis of hotel bookings data in order to enhance customer satisfaction, improve operational efficiency for the hotels. Analyze booking patterns, guest preferences, and factors influencing cancellations. Using SQL and Excel to identify trends in booking sources and revenue. Developing a Power BI dashboard for tracking booking trends and optimizing hotel operations..

SIGNIFICANCE

The hotel booking analysis plays a significant role in the whole lot of hotels landscape, providing valuable insights for the owners, management and the staffs. By deliving into the factors that influence the lead time, ADR, meal & stay and hotel type can make more informed decisions like the best hotel and room type can be improved and analyze their advantages.

For the hotel owners, the hotel booking analysis serves as a valuable resource for making a profitable revenue, the management can make alot of improvements on their cons/loss. The customers also can analyze & decide according to their interests.

DATA DICTIONARY

ROOM_DETAILS

THE ROOM_DETAILS TABLE PROVIDES INFORMATION RELATED TO ROOM RESERVATIONS AND CHANGES MADE TO THEM. IT IS ASSOCIATED WITH THE BOOKING_DETAILS TABLE VIA THE BOOKING IDENTIFIER. THIS TABLE INCLUDES DETAILS ABOUT THE TYPE OF ROOM INITIALLY RESERVED, THE TYPE OF ROOM EVENTUALLY ASSIGNED, AND THE NUMBER OF CHANGES MADE TO THE BOOKING. IT OFFERS INSIGHTS INTO ROOM ALLOCATION DYNAMICS AND BOOKING MODIFICATIONS.

RESERVATION_STATUS

THE RESERVATION_STATUS TABLE RECORDS THE STATUS OF RESERVATIONS OVER TIME. IT IS CONNECTED TO THE BOOKING_DETAILS TABLE THROUGH THE BOOKING IDENTIFIER. THIS TABLE CAPTURES THE RESERVATION'S LAST STATUS (E.G., CANCELED, CHECK-OUT) AND THE DATE ON WHICH THIS STATUS WAS RECORDED. IT IS VALUABLE FOR TRACKING THE PROGRESSION OF RESERVATIONS AND UNDERSTANDING THEIR FINAL OUTCOMES.

BOOKING_DETAILS

THE BOOKING_DETAILS TABLE CONTAINS ESSENTIAL DETAILS RELATED TO HOTEL RESERVATIONS. IT INCLUDES A UNIQUE BOOKING IDENTIFIER AND INFORMATION ABOUT THE TYPE OF HOTEL (RESORT HOTEL OR CITY HOTEL). ADDITIONALLY, IT RECORDS THE BOOKING'S CANCELLATION STATUS (0 FOR NOT CANCELED, 1 FOR CANCELED), LEAD TIME (NUMBER OF DAYS BETWEEN BOOKING AND ARRIVAL), YEAR, MONTH, WEEK NUMBER, AND DAY OF THE MONTH OF ARRIVAL. THE TABLE ALSO CAPTURES THE NUMBER OF WEEKEND AND WEEKDAY NIGHTS STAYED.

GUEST_INFO

THE GUEST_INFO TABLE PROVIDES INSIGHTS INTO THE GUESTS ASSOCIATED WITH EACH BOOKING. IT IS LINKED TO THE BOOKING_DETAILS TABLE VIA THE BOOKING IDENTIFIER. THIS TABLE RECORDS THE NUMBER OF ADULTS, CHILDREN, AND BABIES ACCOMPANYING THE BOOKING, OFFERING AN UNDERSTANDING OF THE COMPOSITION OF GUESTS FOR EACH RESERVATION.

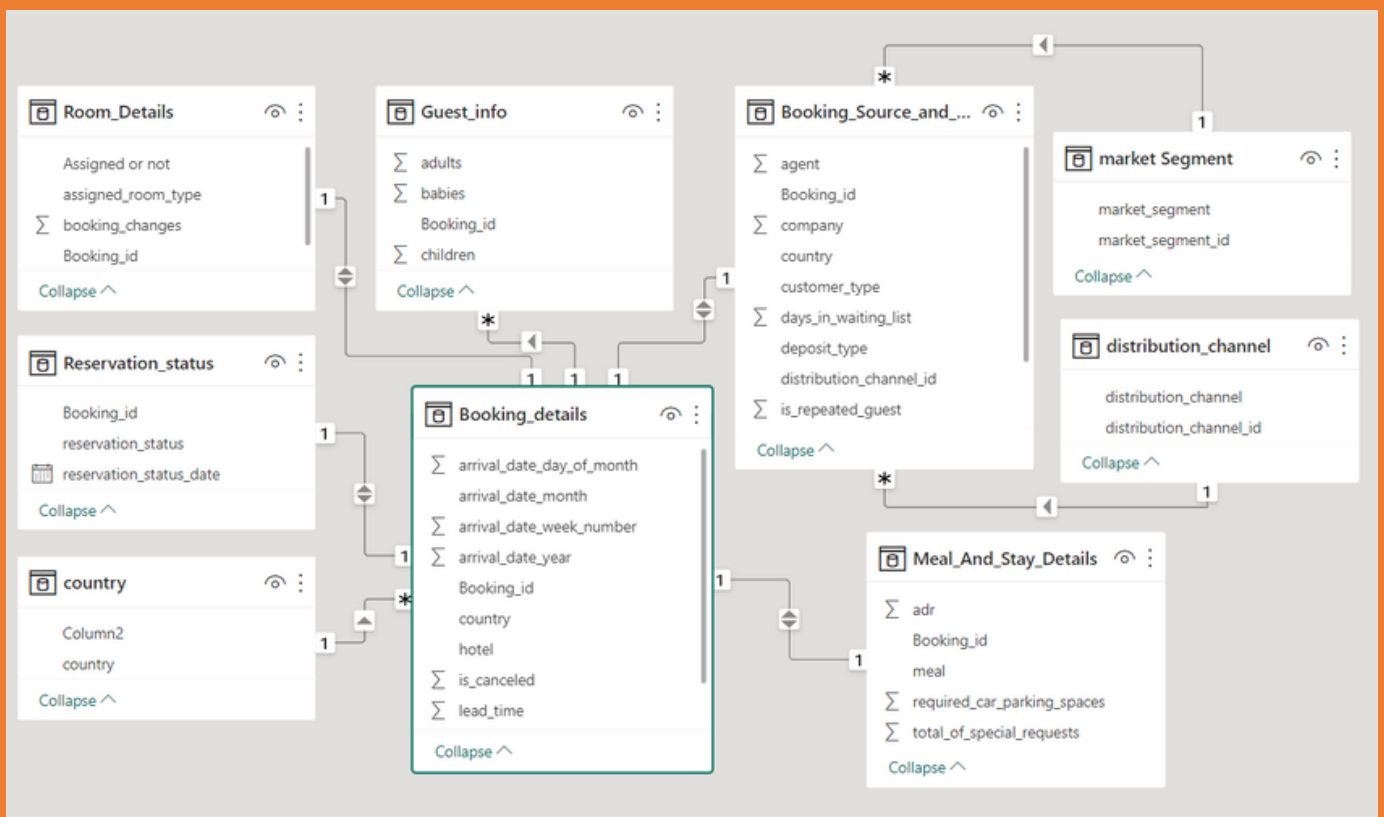
MEAL_AND_STAY_DETAILS

THE MEAL_AND_STAY_DETAILS TABLE COMPLEMENTS THE BOOKING INFORMATION BY SPECIFYING MEAL-RELATED AND STAY-RELATED ATTRIBUTES. IT CONNECTS TO THE BOOKING_DETAILS TABLE VIA THE BOOKING IDENTIFIER. THIS TABLE INCLUDES THE TYPE OF MEAL BOOKED (E.G., BED & BREAKFAST, HALF BOARD), THE AVERAGE DAILY RATE (ADR) FOR THE STAY, THE NUMBER OF REQUIRED CAR PARKING SPACES, AND THE TOTAL COUNT OF SPECIAL REQUESTS MADE BY THE GUEST.

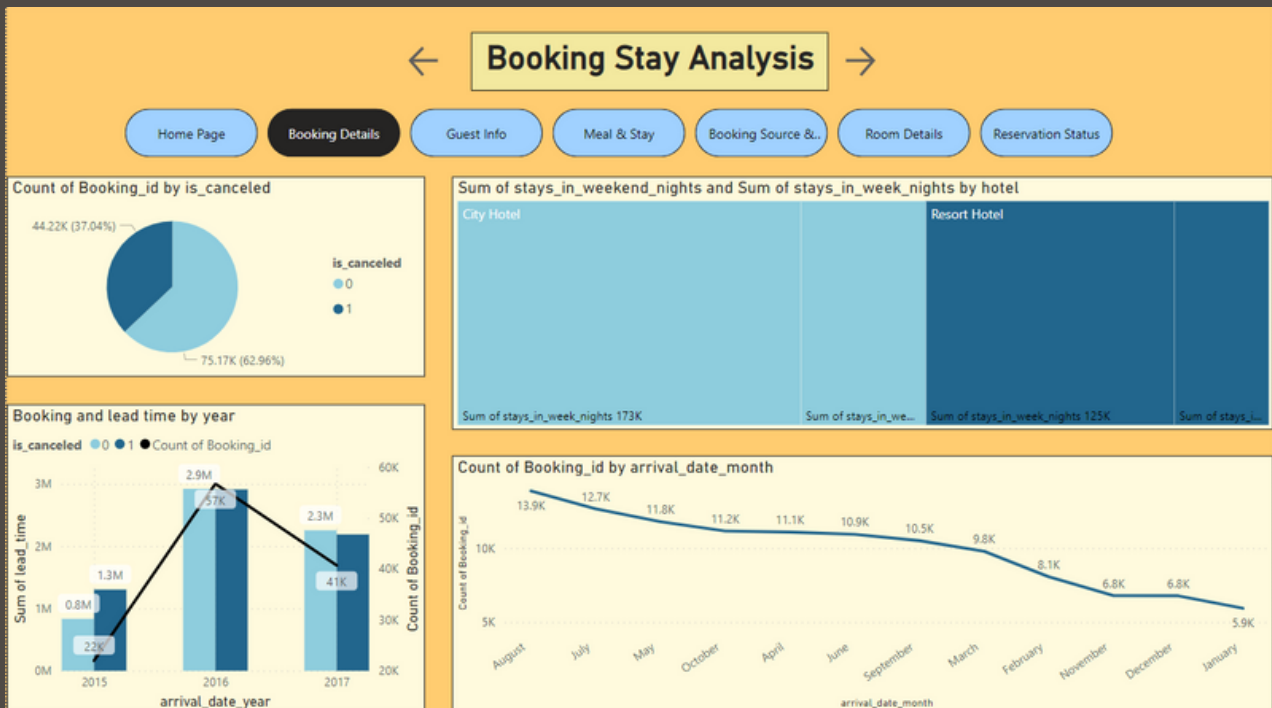
BOOKING_SOURCE_AND_HISTORY

THE `BOOKING_SOURCE_AND_HISTORY` TABLE IS CRUCIAL FOR UNDERSTANDING THE SOURCE OF BOOKINGS AND THE HISTORICAL BEHAVIOR OF GUESTS. IT IS CONNECTED TO THE `BOOKING_DETAILS` TABLE VIA THE BOOKING IDENTIFIER. THIS TABLE ENCOMPASSES INFORMATION SUCH AS THE MARKET SEGMENT (E.G., ONLINE TRAVEL AGENTS, DIRECT BOOKING), DISTRIBUTION CHANNEL (E.G., ONLINE TRAVEL AGENTS, DIRECT BOOKING), AND WHETHER THE GUEST IS A REPEATED VISITOR (0 FOR NOT REPEATED, 1 FOR REPEATED). ADDITIONALLY, IT RECORDS THE NUMBER OF PREVIOUS BOOKING CANCELLATIONS, PREVIOUS BOOKINGS THAT WERE NOT CANCELED, THE DEPOSIT TYPE (E.G., NO DEPOSIT, NON-REFUND), THE BOOKING AGENT'S ID, THE COMPANY'S ID, THE NUMBER OF DAYS A BOOKING SPENT ON THE WAITING LIST, AND THE CUSTOMER TYPE (E.G., TRANSIENT, GROUP).

ER DIAGRAM



Power BI Statements

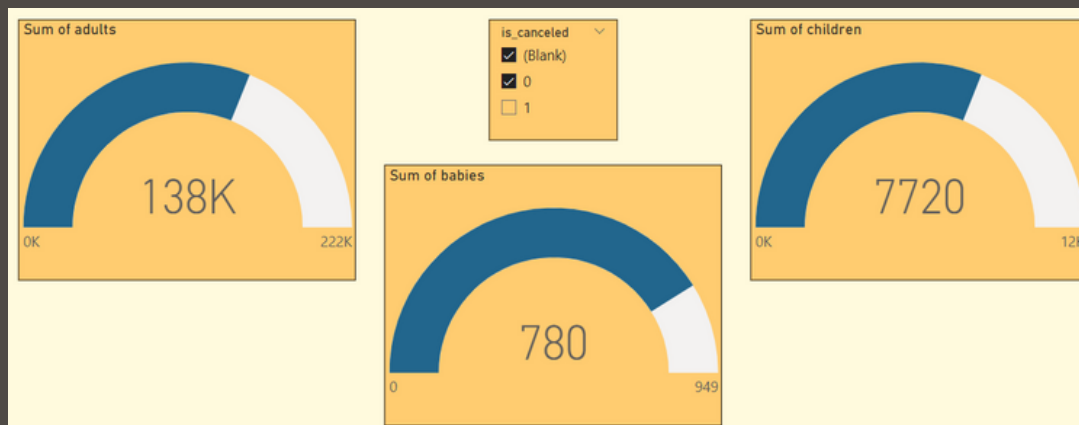


What is the booking trends over the years, including the number of bookings, cancellations. Identify seasonality patterns . Compare the no.of weekends and weekdays for hotel types ?

Analyzing the Booking stay details, the booking trends for the guests' arrival is the August month which is around 13K guests. There is a percentage of 37.04% of the cancellation of the room booking.

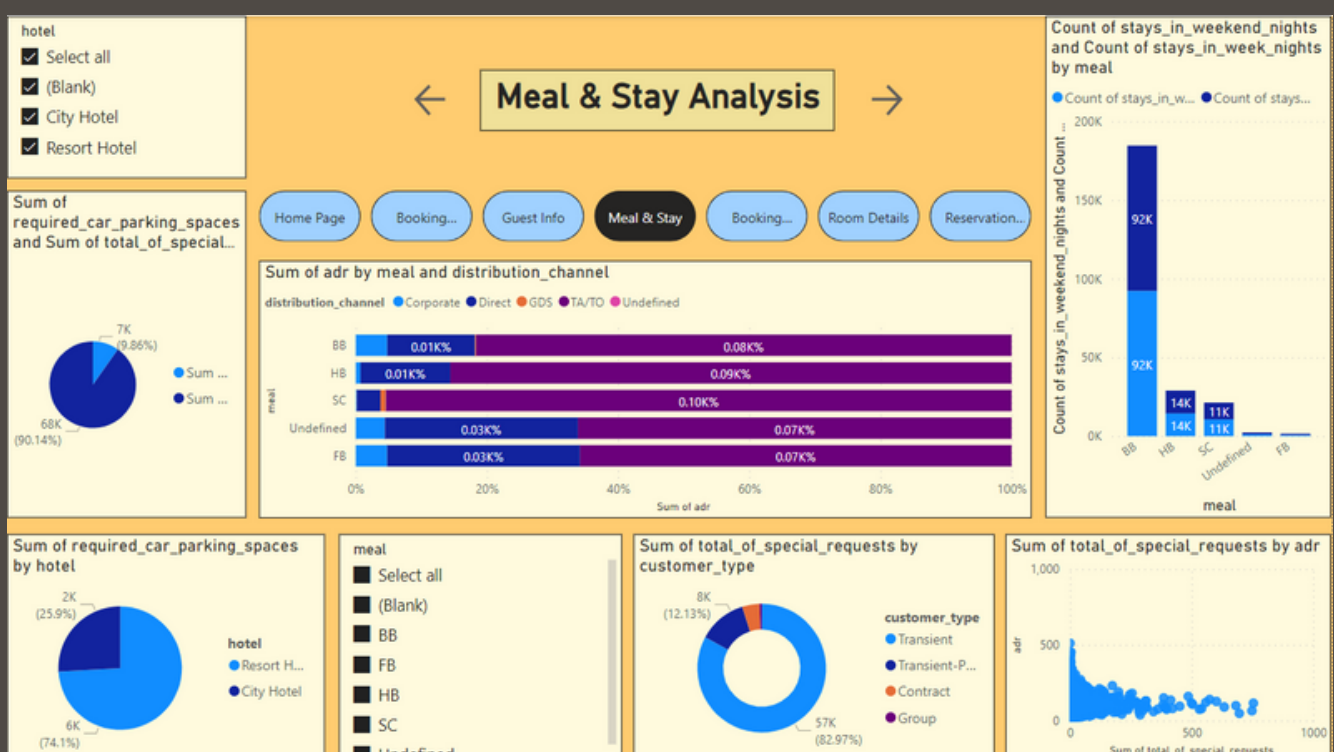
The seasonality pattern of hotels is the best time to visit Europe is during the summer months of June to September. This is the time when the sun is high, the beaches are warm and the weather is sunny.

The majority average of the stay time is from the Resort hotel where the customers have stayed for 3.03 days during weekdays



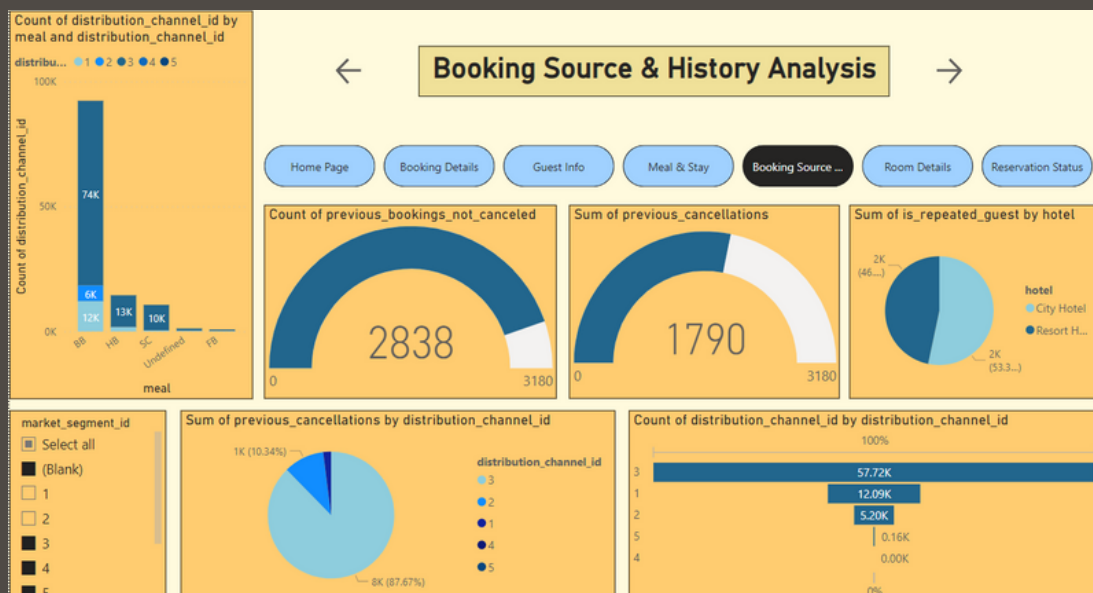
Visualize the distribution of adults, children, and babies in bookings. Explore the impact of children and babies on cancellation rates.

Analyzing the Guest info details, the sum of adults has a 138K customers are checked out whereas, the adults who are cancelled are 81K customers. the sum of children who are checked out are 7720. Whereas the children whose booking are cancelled is around 4K kids. As per in the visualization, the cancellation is around 37.04% of getting cancelled. So the guests mostly who has babies are being checked-out, so this hotel is can consider to be a baby-friendly hotel. The kids are also being checked-out alot, so the hotel is a also being a kid- friendly hotel.



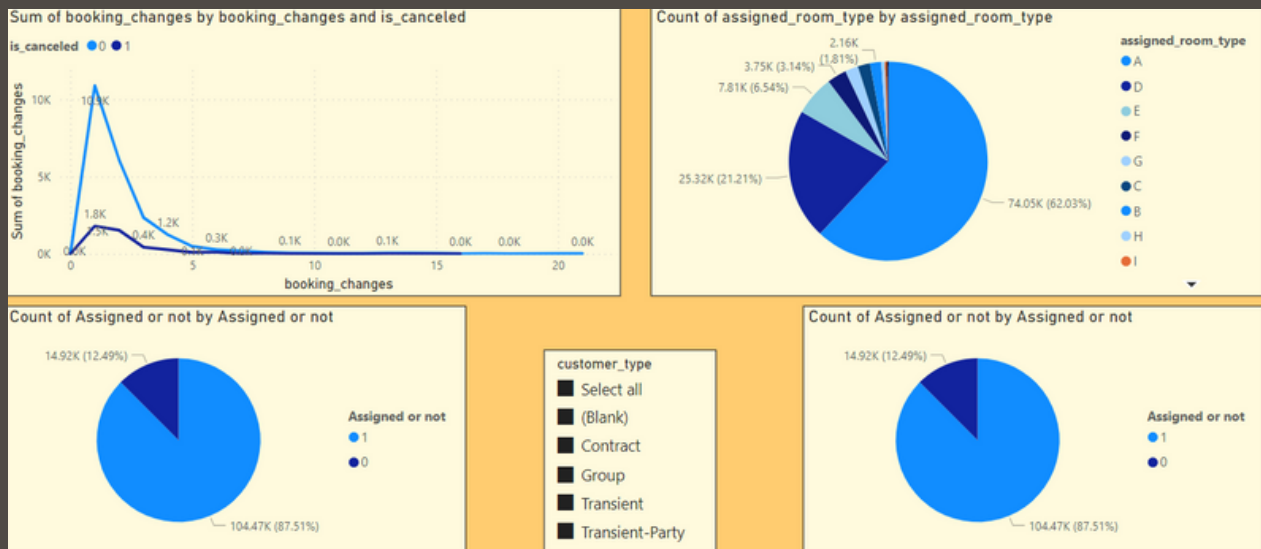
Explore meal plans and their impact on Average Daily Rates (ADR). Analyze meal plan preferences and their association with booking channels

The meal plans for each and every room type and hotel type as per given data, the meal which was preferred most of the customers are the Bed n Breakfast type of meal which around 75% are being opted. The Average Daily Rate (ADR) has been a great impact for the Bed n Breakfast meal because of the price of the adr is being valuable for money. The majority of the distribution channel is the TA/TO channel which has a around 80 % of the customers.



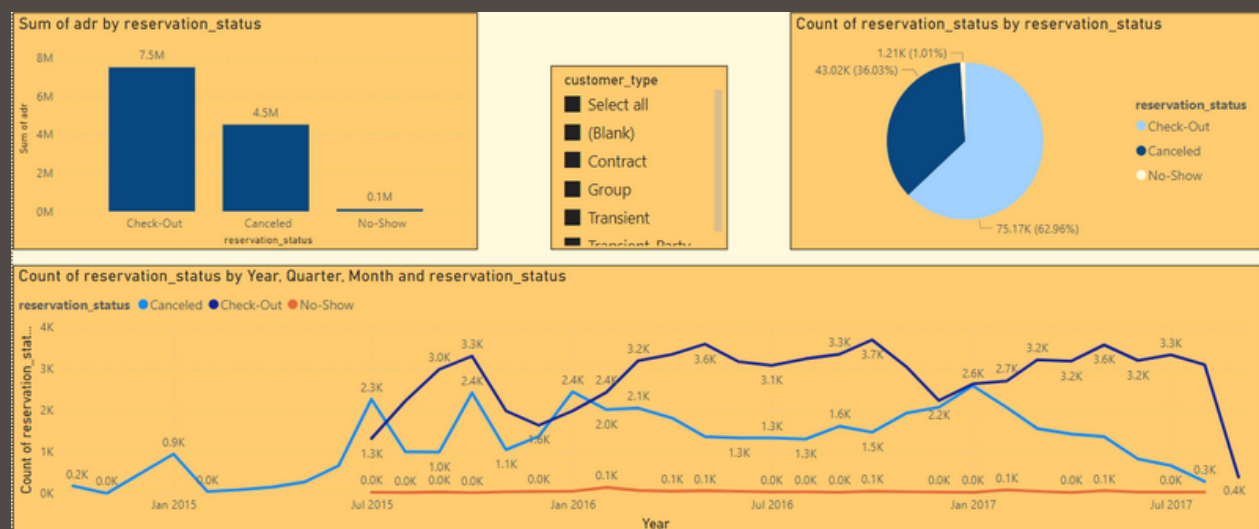
Visualize booking distribution across different market segments and analyze cancellation rates within each segment.

The Booking distribution across different market segments are being distributed mostly for all the markets, the 3rd distribution are being used for all the markets, in which we know that the TA/TO are always the most type distributed customers are arriving. The city type of hotel has the more number of the customers who are retuning, but as per average the resort hotel customers are returning alot.



Investigate the relationship between the number of booking changes made by guests and their likelihood of canceling a booking.

The maximum number of changes which are done by the customers are 1 change done per booking, around 12701 customers have done only a single change. Out of these customers, only 1807 customers have been cancelled their booking, so the cancellation conversion rate is at 14.22 % which tends to be a good sign of the hotel performance.



Visualize how reservation statuses vary across different customer types (e.g., Transient, Group) and identify if certain customer types are more likely to result in cancellations or noshows

The reservation status basically consists of 3 types : Check out, Cancelled and NO-Show . Out of these status the Check Out status has the most number of statuses around 62 % of the whole dataset. As per comparing the customer type , the most common reservation status for every type the Checkout status is still being the maximum out of the 3 statuses.

The transient type of customers are the most common customers in the hotels and around 59 % are Checkout customers , meanwhile the cancelled customers are around 39 % , so the No show customers are in 2 %.

So , by Overall the percentage of cancelled customers are 36.03% and NO-Show of 1.01%.