



HOTEL BOOKING

EDA Report



- Anjali Solanki
- Rutvi Dave
- Melvin Mathai



Abhishek Wavhal



⇨ Summary of the Project:

- The main Objective of this project is to analyze and understand the business generation of one the most reputed hotel chain in the country.
- The hotel chain consists of two sub units namely: Resort Hotel and City Hotel.
- We have been provided with a data set (for 3 years:2015, 2016, 2017) which consists of data for both these sub units.
- So we will be using this data after a thorough process of cleaning and editing, use them to compare the business outputs from both these sub units which will give us an overview of the business done the Hotel chain.





EFFECT OF PROJECT :

- As mentioned in the previous slide, the main objective of this project is to analyze the data for both City Hotel and Resort Hotel and compare them to get an overview of the total business done by the hotel and the business contribution from both of the units.
- Which will help us getting insights on how the overall business can be developed and where all there is scope of improvement.

Data Variable:

- ❖ **Hotel:** Type of hotel(City or Resort)
- ❖ **is_cancelled:** If the booking was cancelled(1) or not(0)
- ❖ **lead_time:** Number of days before the actual arrival of the guests
- ❖ **arrival_date_year:** Year of arrival date
- ❖ **arrival_date_month:** Month of arrival date
- ❖ **arrival_date_week_number:** Week number of year for arrival date
- ❖ **arrival_date_day_of_month:** Day of arrival date
- ❖ **stays_in_weekend_nights:** Number of weekend nights(Saturday or Sunday) spent at the hotel by the guests.
- ❖ **stays_in_week_nights:** Number of weeknights(Monday to Friday) spent at the hotel by the guests.
- ❖ **adults:** Number of adults among the guests
- ❖ **children:** Number of children
- ❖ **babies:** Number of babies
- ❖ **meal:** Type of meal booked
- ❖ **country:** country of the guests
- ❖ **market_segment:** Designation of market segment
- ❖ **is_repeated_guest:** If the booking was from a repeated guest(1) or not(0)
- ❖ **previous_cancellation:** Number of previous bookings that were cancelled by the customer prior to the current booking
- ❖ **previous_bookings_not_cancelled:** Number of previous bookings not cancelled by the customer prior to the current booking
- ❖ **reserved_room_type:** Code from room type reserved
- ❖ **assigned_room_type:** Code of room type assigned
- ❖ **booking_changes:** Number of changes made to the booking
- ❖ **deposit_type:** Type of deposit made by the guest
- ❖ **agent:** ID of travel agent who made the booking
- ❖ **company:** ID of the company that made the booking
- ❖ **days_in_waiting_list:** Number of the days the booking was in the waiting list
- ❖ **customer_type:** Type of customer, assuming one of four categories
- ❖ **adr:** Average daily rate
- ❖ **required_car_parking_spaces:** Number of car parking spaces required by the customer
- ❖ **total_of_special_requests:** Number of special requests made by the customer
- ❖ **reservation_status:** Reservation status(Canceled, check-out or no-show)

Data Cleaning and Wrangling Process:

- First of all we extracted and loaded our data set in power query editor.
- After loading the first thing we did was we removed errors from the entire data set.
- We found out that the total number of rows in the data set was **1,19,386**.
- As there are chances that there might be duplicate records in this data set, so we selected the entire table and removed duplicate data from the entire table.
- After removing the duplicate data we came to know that there were **31994** duplicate records in the entire data set which means only **87392** were unique data after removing the duplicates.
- We then started looking at all the columns to check if any more changes can be made or not.
- Removed the “**arrival_date_week_number**” column as that was not quite useful for our analysis.
- Then we made a new column named as **Date** by merging the Day, Month and Year column and changed the data type of this new column into date from any.
- We added a new column “**Total stay days**” by adding the data of weekdays and weekend stays column and the change the data type to whole numbers.



Continued . . .

- We added one more column “Total number of People” by adding the data of the columns Adults, Children and Babies.
- We noticed that the above new column we made had some rows with **zero** as values in it, **172** rows to be specific. So we replaced these values with an **error** and removed the errors.
- So now we after removing the 172 records we have total of **87220** unique records.
- We added a new column using DAX query named as **“status of room assigned”** which shows whether the person has got the same room type as per his/her choice.
- Then in the column named **“agent”** we found out that there were **12134** records having **NULL** value, so as the number of records having NULL value was high we replaced the same with **OTHERS**.
- We removed the “company” column as that was not useful for the analysis.
- Then after all these steps we added the column named **“Order ID”** which we used as our **primary key** in this data.

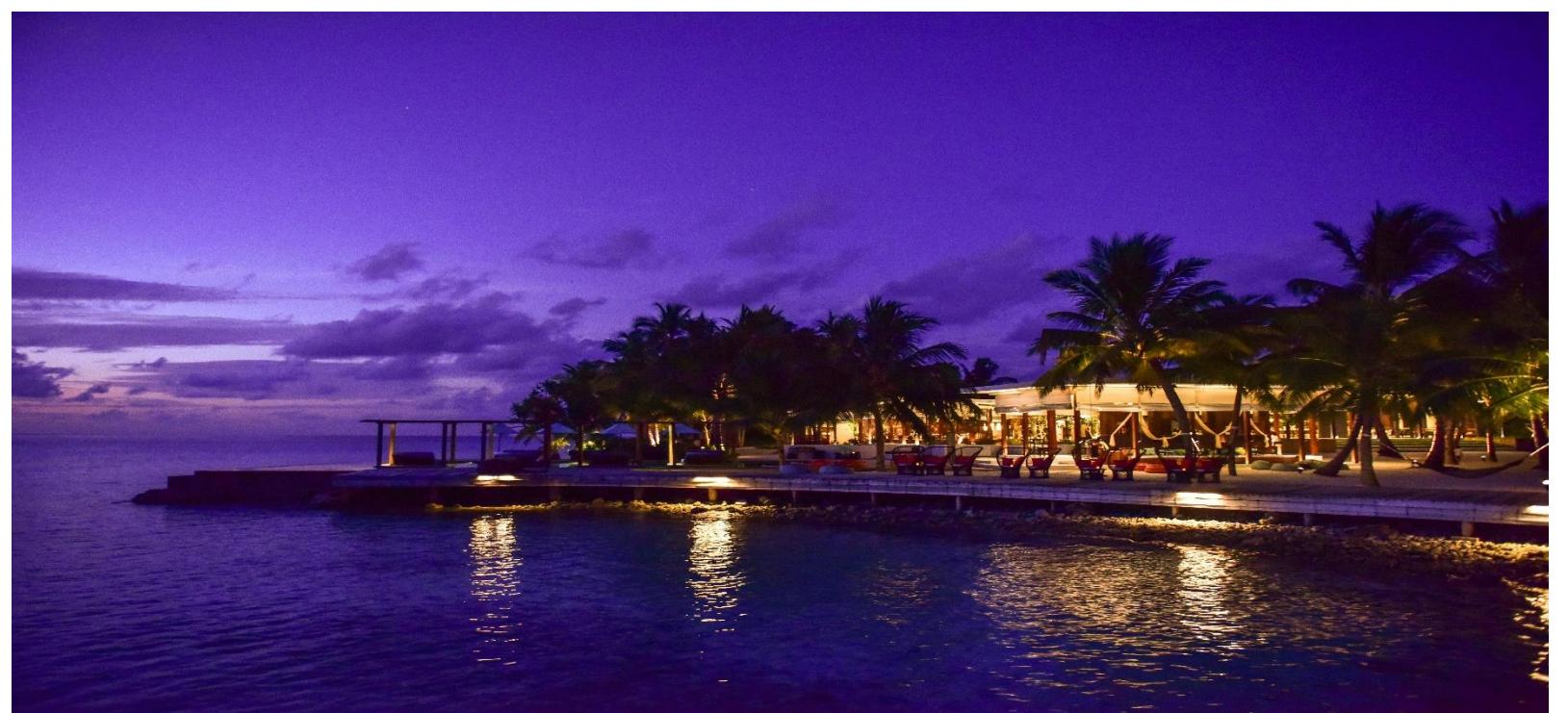


Data Manipulation is the modification of information to make it easier to read or more structured.



Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools.

Primary Data Visual Comparison Using Matplotlib
We already know that we are dealing with a 2-Hotel data comparison but, before moving ahead, we would have to see the shares of each hotel in the analysis. Let's use the Data Visualization tool i.e. matplotlib (alternatively, seaborn can also be used), for plotting the shares of each hotel data.



TYPES OF HOTEL PLOT

Types of hotel

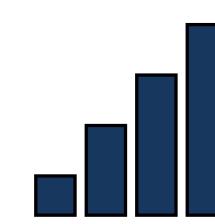
● City Hotel ● Resort Hotel

33.97K
(38.87%)

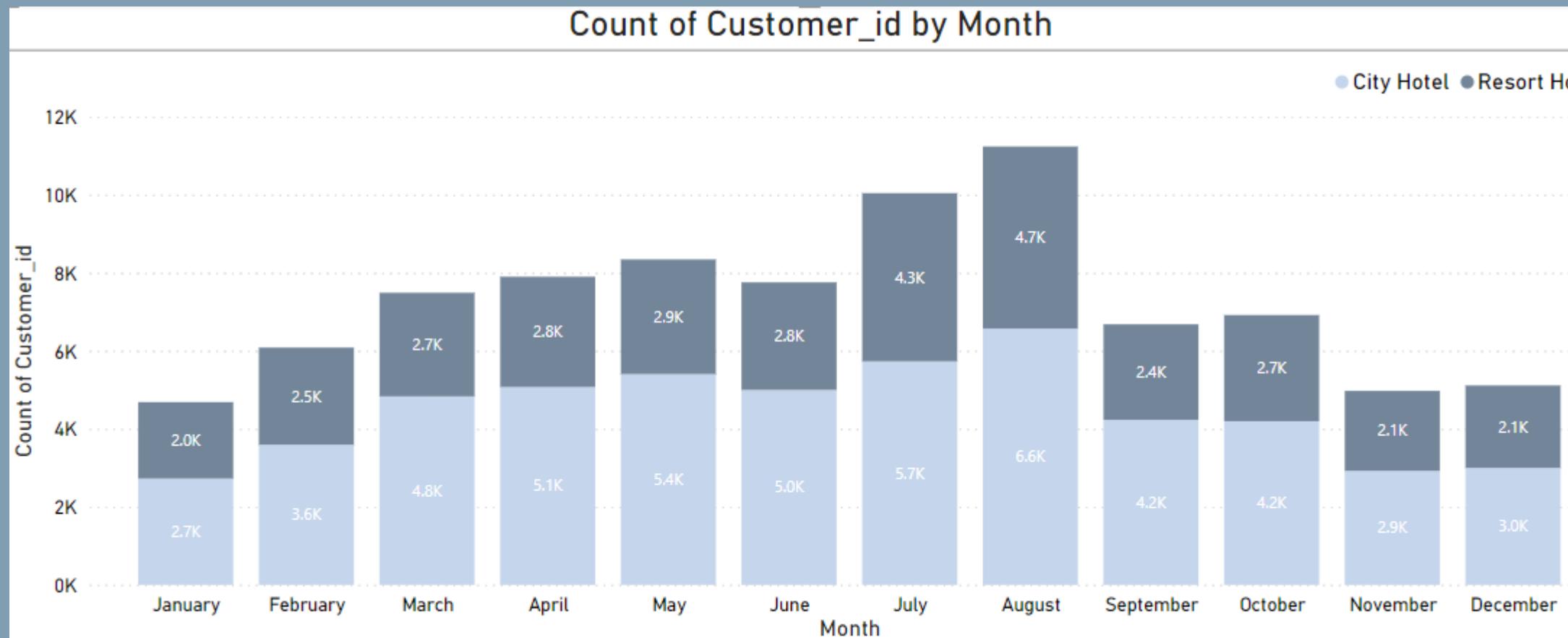
53.42K
(61.13%)

i.e. the data we will be dealing with has 61.13 percent share of City Hotel and the rest is of the Resort Hotel.





FABRICATING INSIGHTS



Analysing Top Arrival Month

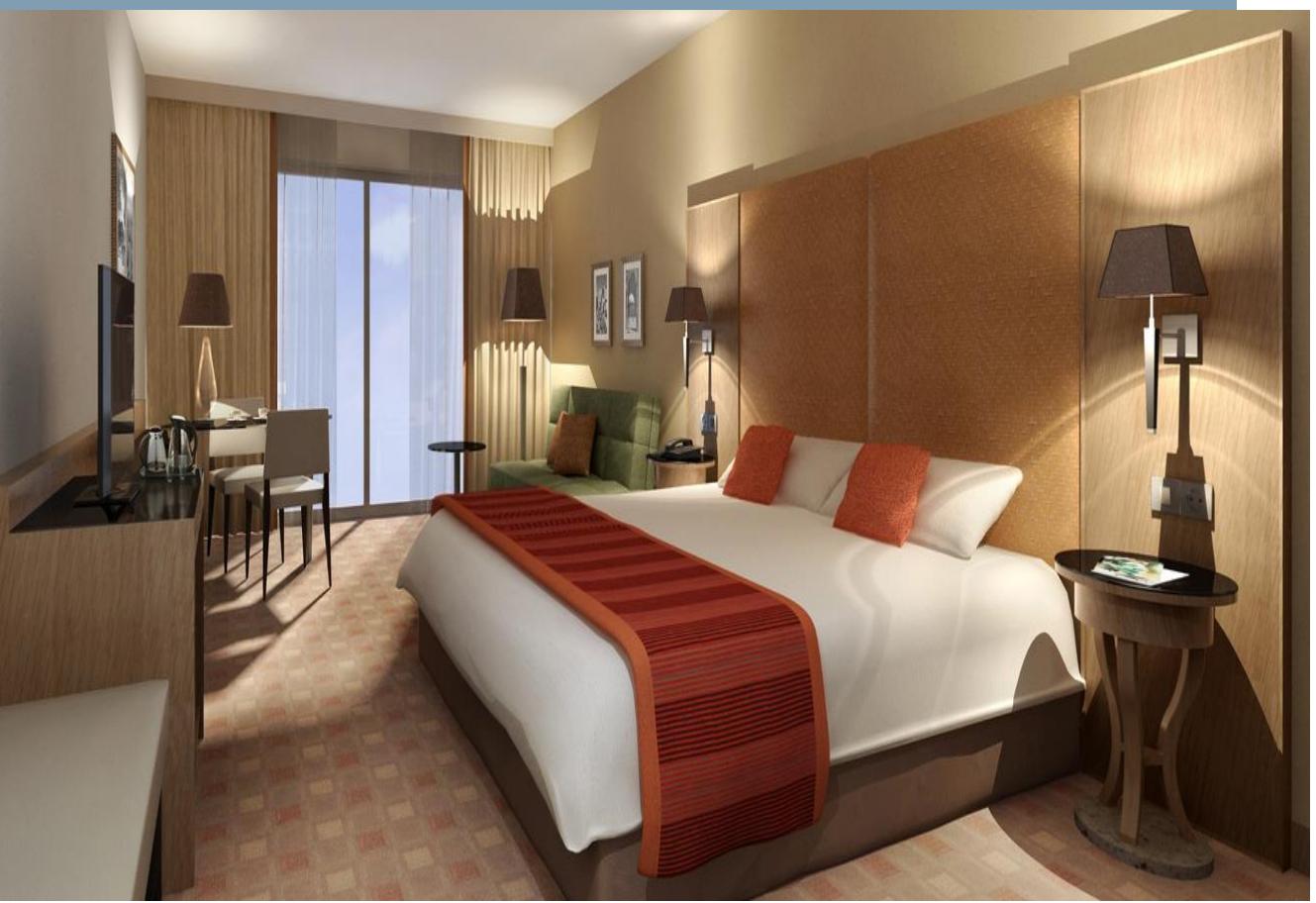
- This bar chart represents a bifurcated monthly booking details between city hotel and resort hotel.
- which gives us a fair idea that both the hotels have received the highest booking in the month of August and the lowest booking in the month of January.

Continued...

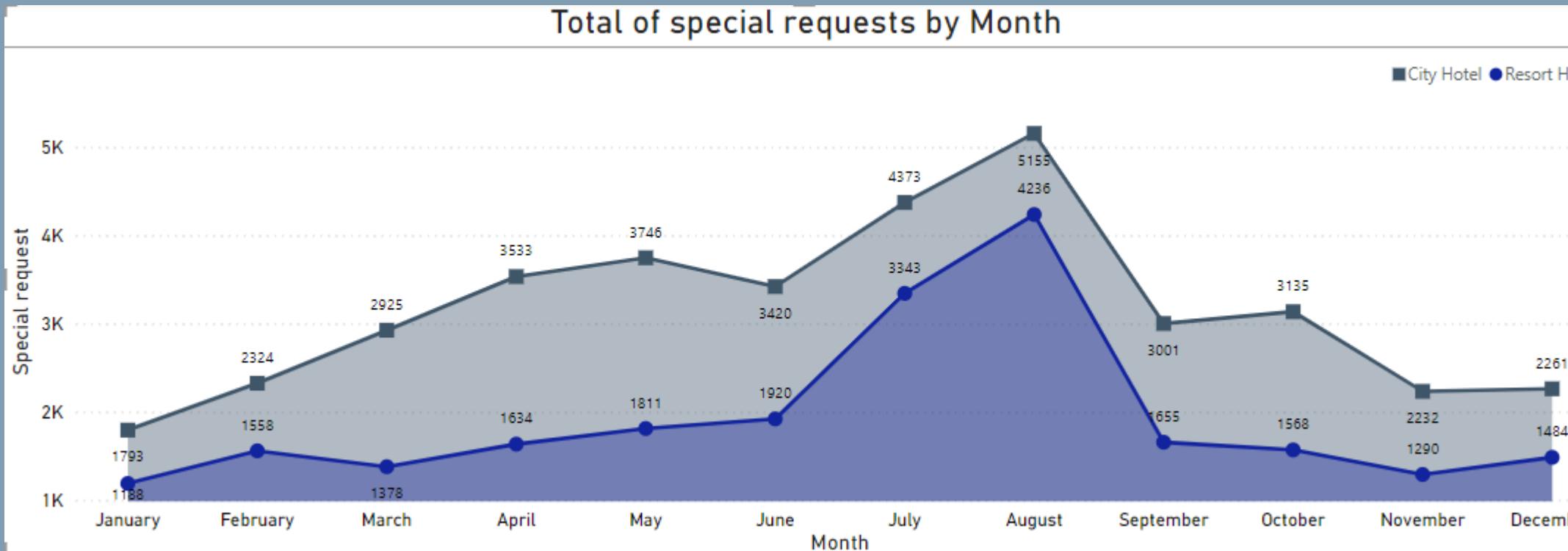
MOST POPULAR ROOM TYPE & MEAL CHOICES

Sum of total days by Roomtype and Meal						
Room types	BB	FB	HB	SC	Undefined	Total
A	134522	894	21951	27012	1777	186156
B	3224		87	60		3371
C	2665	151	1292	10	146	4264
D	59241	489	11868	266	337	72201
E	20609	249	7558	125	200	28741
F	8768	67	1526	105	63	10529
G	7269	18	1451	5	60	8803
H	1838	16	549			2403
L	7					7
Total	238143	1884	46282	27583	2583	316475

- This Matrix enlightens us about a very important aspect that is customer preferences.
- In this Matrix we can find out the room type preferences and the meal preferences and the combination of both of them .
- we get an idea that majority of them go for a combination of Room type A and BB meals. (Which shows that people want greater room experience and can adjust on their meal choices).



Continued...

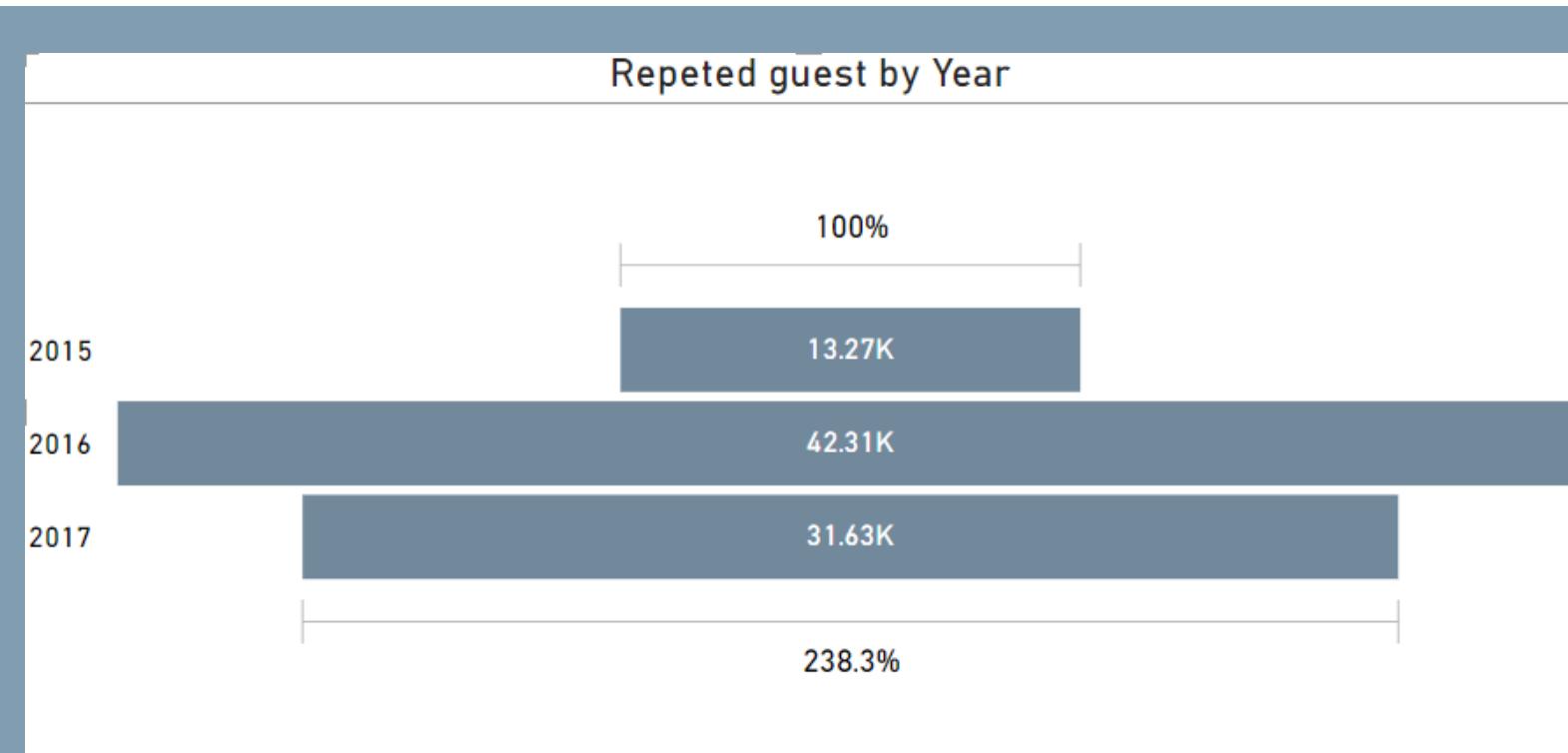


ANALYSE SPECIAL REQUESTS BY MONTH

- This data gives us a simple bifurcation on special request received by both the hotel units.
- Which gives us a clear picture that city hotel receives maximum order, that is why they have received more special requests and still they are able to manage the guests.
- Maximum requests have been seen in the month of July, August only for both the hotels.

Continued...

REPETED GUEST AND DEPOSIT TYPE

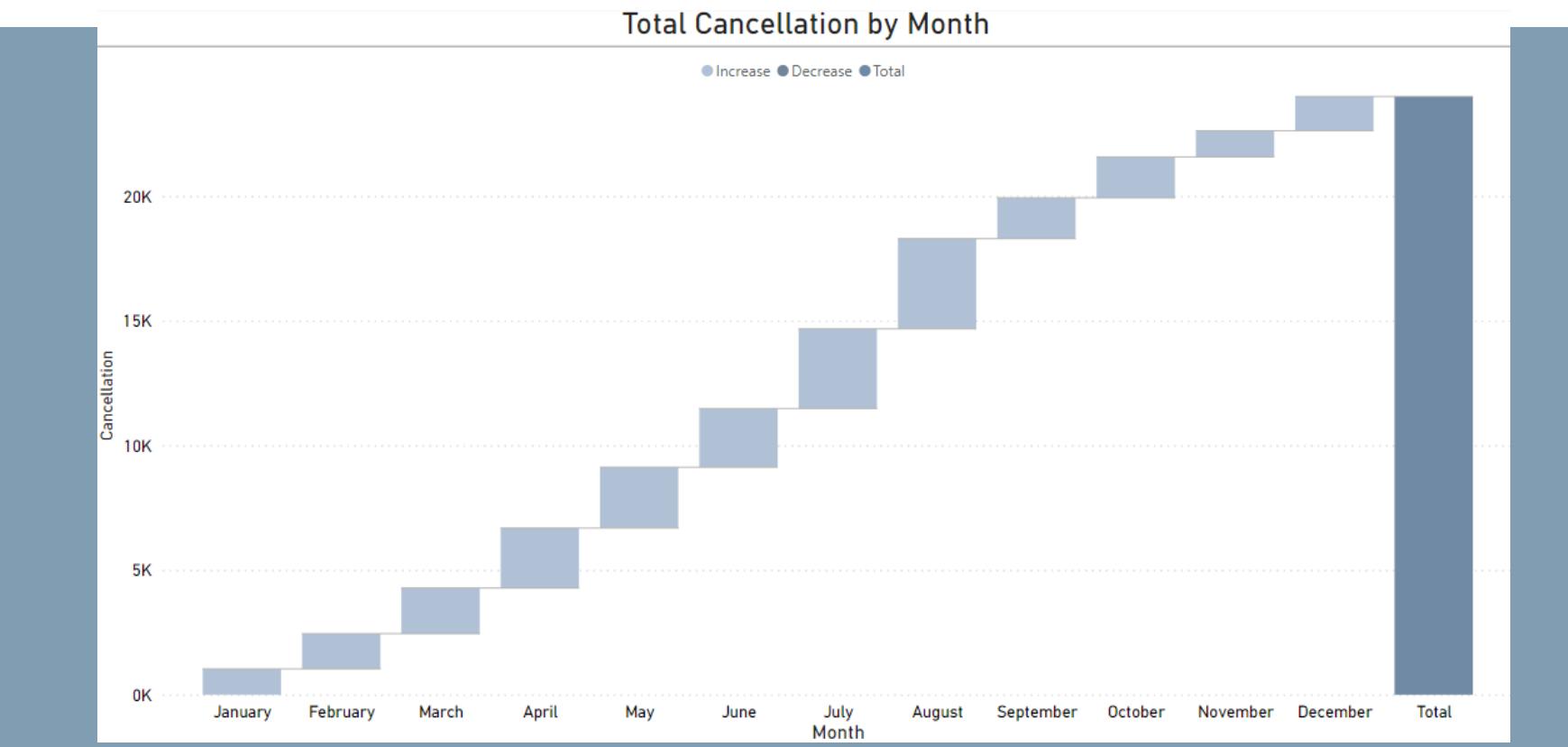
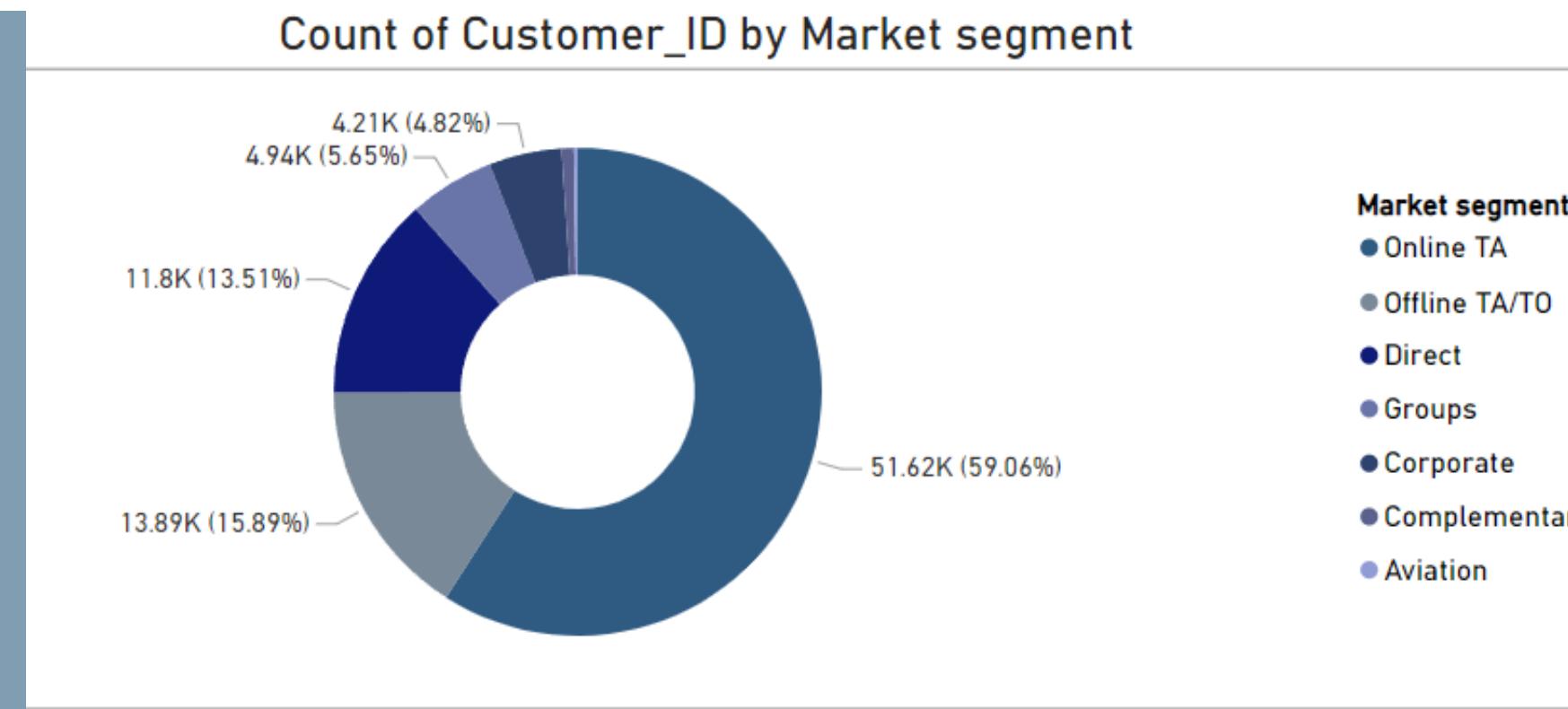


Deposit type	Count of customer
No Deposit	86075
Non Refund	1038
Refundable	107
Total	87220

- This funnel chart represents the repeated booking data both the hotels received within our timeframe of 3 years.
- The chart shows that 2016 was the year where both these hotels received the maximum repeated bookings.
- The reason for this exact specific hike in that particular year could be because the country was hosting the T20 world cup.
- This data shows us what deposit type is selected by the people while booking their stay with the hotel.
- This shows that many people to proceed with their bookings without paying any deposit.
- The reason could be as we saw in one of our slides that the maximum bookings coming through online segment or sense of doubt in mind while making any advance payment.

Continued...

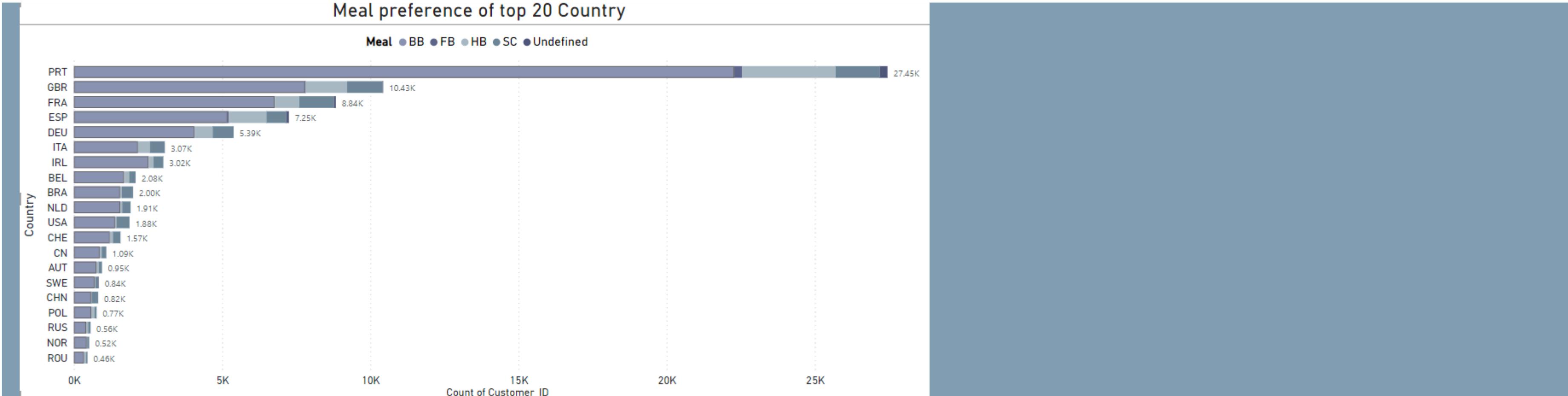
Market segments and cancellation



- The donut chart here represents the total bookings received from various segments as mentioned.
- From this we can clearly get a picture that the maximum bookings come from the Online TA segment
- This means there might be maximum benefit/discounts available for people while booking through online segment.

- This chart shows us total cancellation on a monthly basis.
- Where August Month has seen the maximum cancellation.

Continued...



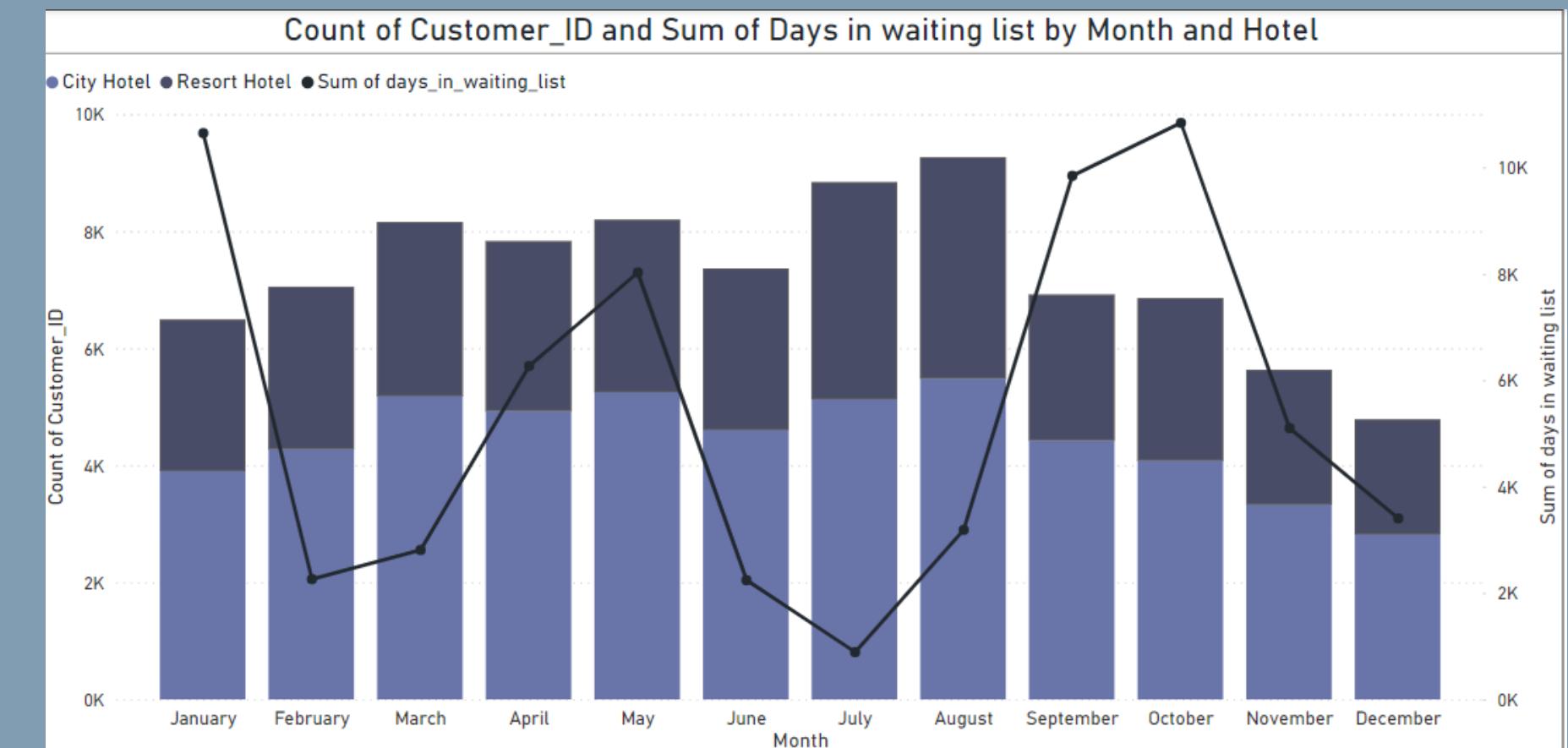
- This chart gives us dual information.
- It shows us from which countries we have received the maximum orders from and gives us their meal preferences too.

Continued...

Analyse the days spent on the waiting list

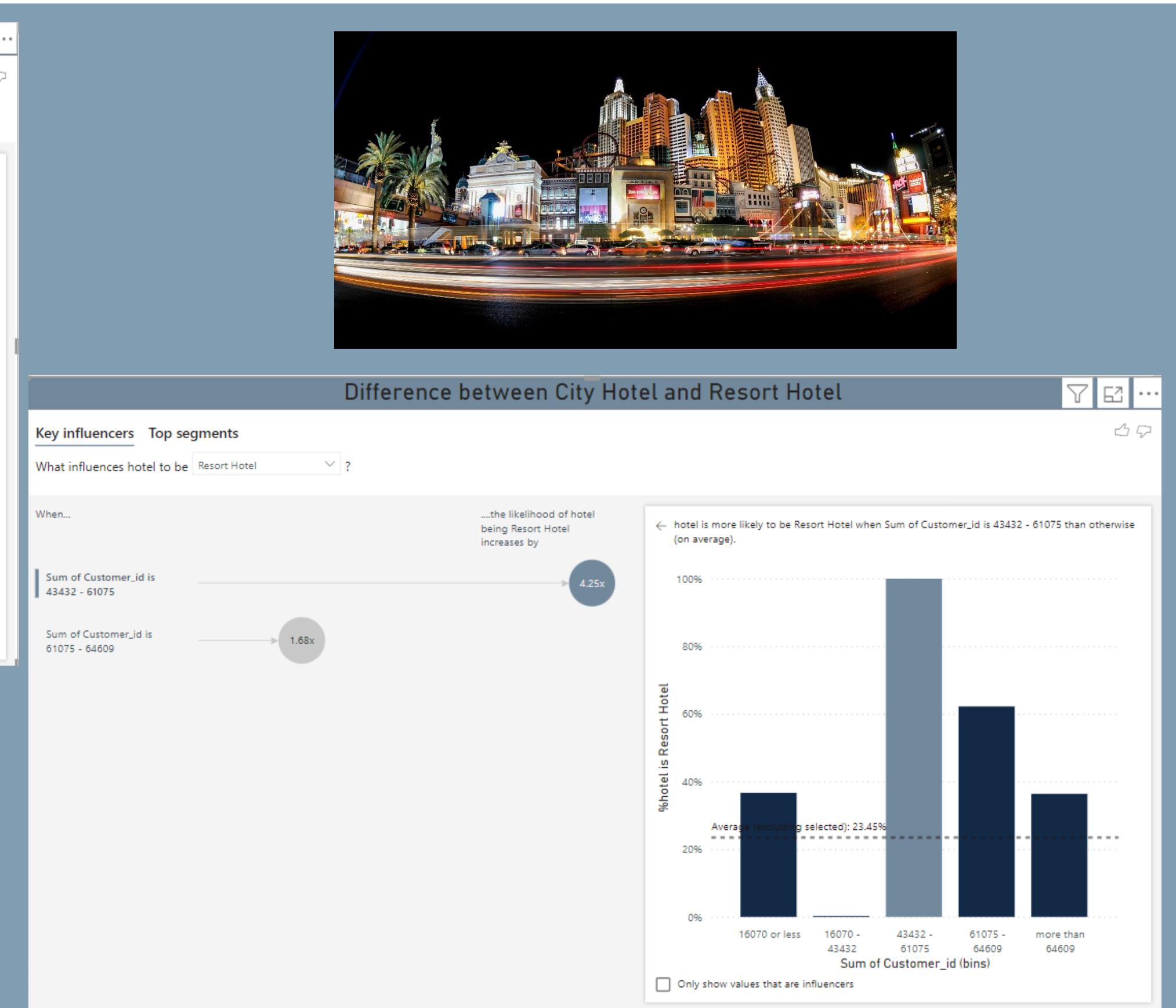
Hotel	City Hotel		Resort Hotel		Total		
	Month	Sum of Days in waiting list	Count of Customer_ID	Sum of Days in waiting list	Count of Customer_ID	Sum of Days in waiting list	Count of Customer_ID
January		7834	3902	2812	2588	10646	6490
February		1613	4277	647	2777	2260	7054
March		1354	5184	1455	2974	2809	8158
April		5243	4930	1026	2901	6269	7831
May		7493	5248	532	2949	8025	8197
June		1791	4606	450	2760	2241	7366
July		876	5131	12	3709	888	8840
August		3186	5485	0	3779	3186	9264
September		9743	4426	101	2491	9844	6917
October		7524	4082	3314	2776	10838	6858
November		4688	3331	412	2303	5100	5634
December		3164	2822	239	1961	3403	4783
Total		54509	53424	11000	33968	65509	87392

- With the help of this chart and table we get a brief idea about the monthly waiting list in both the hotels.
- Where City hotel experiences high waiting list in the month of September and Resort hotel experiences the highest waiting list in the month of October.



Continued...

DIFFERENCE BETWEEN HOTELS



HOTEL BOOKING

Count of Customer by Customer Type and Market Segment

customer_type	Aviation	Complementary	Corporate	Direct	Groups	Offline TA/TO	Online TA	Total
Contract		2	18	14	63	1714	1328	3139
Group	2	6	28	130	13	192	173	544
Transient	209	665	3134	10803	918	8904	47353	71986
Transient-Party	16	29	1032	856	3948	3079	2763	11723
Total	227	702	4212	11803	4942	13889	51617	87392

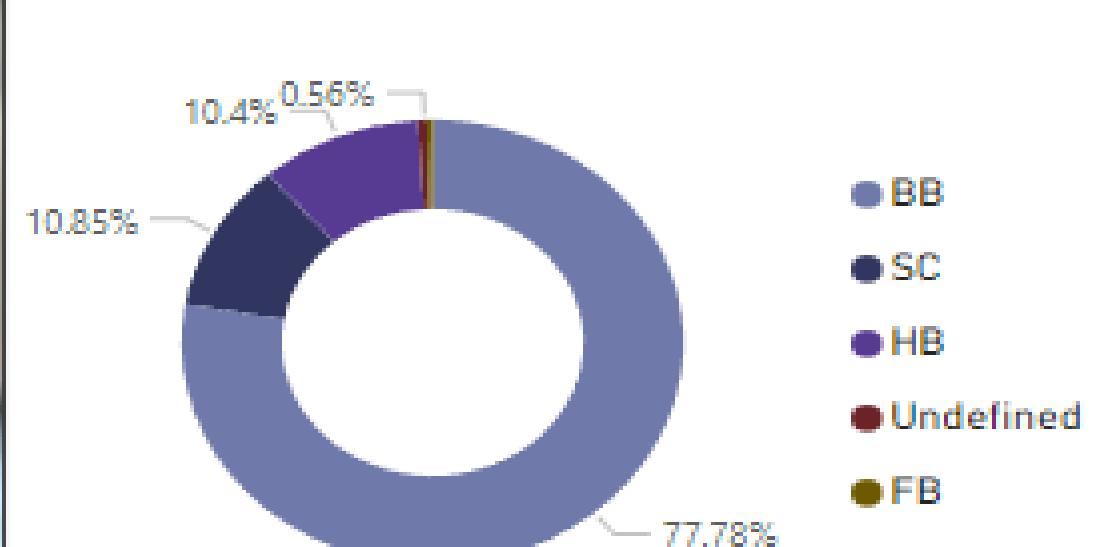
Repeated guest and Required Car parking by Month



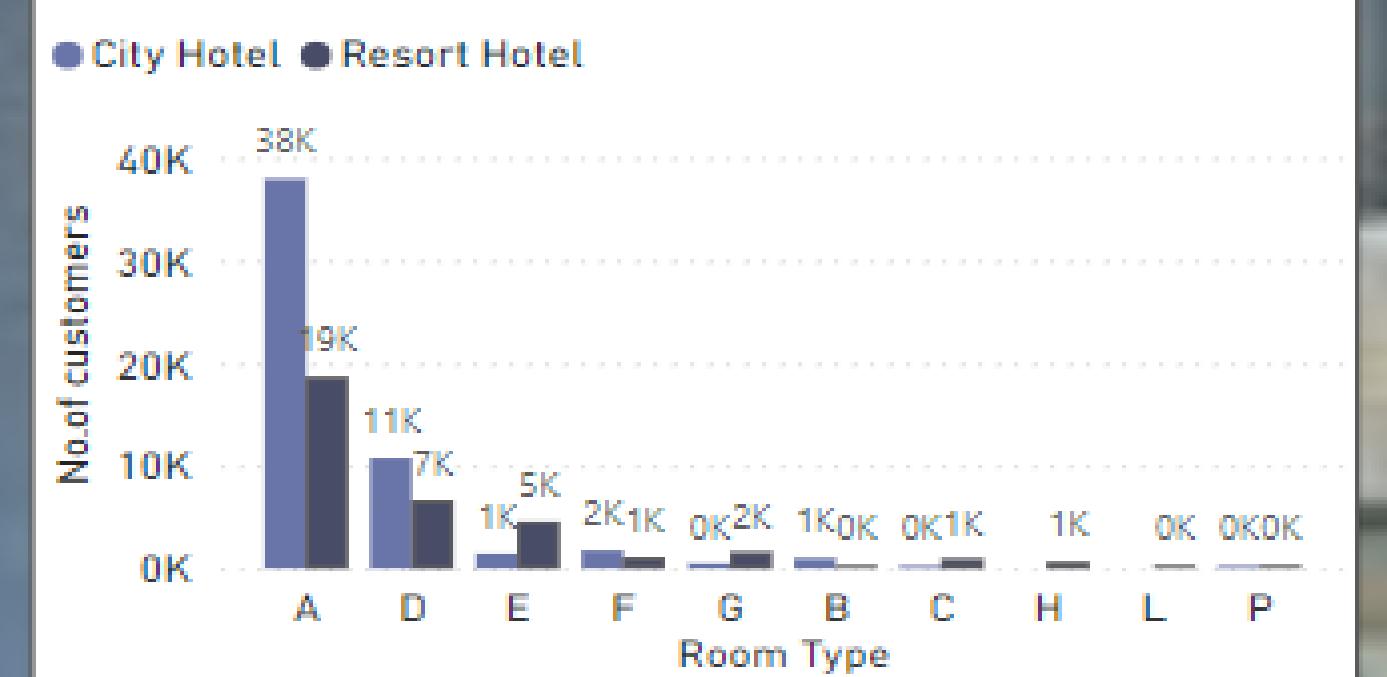
Adr By Year



Meal preference by Customer



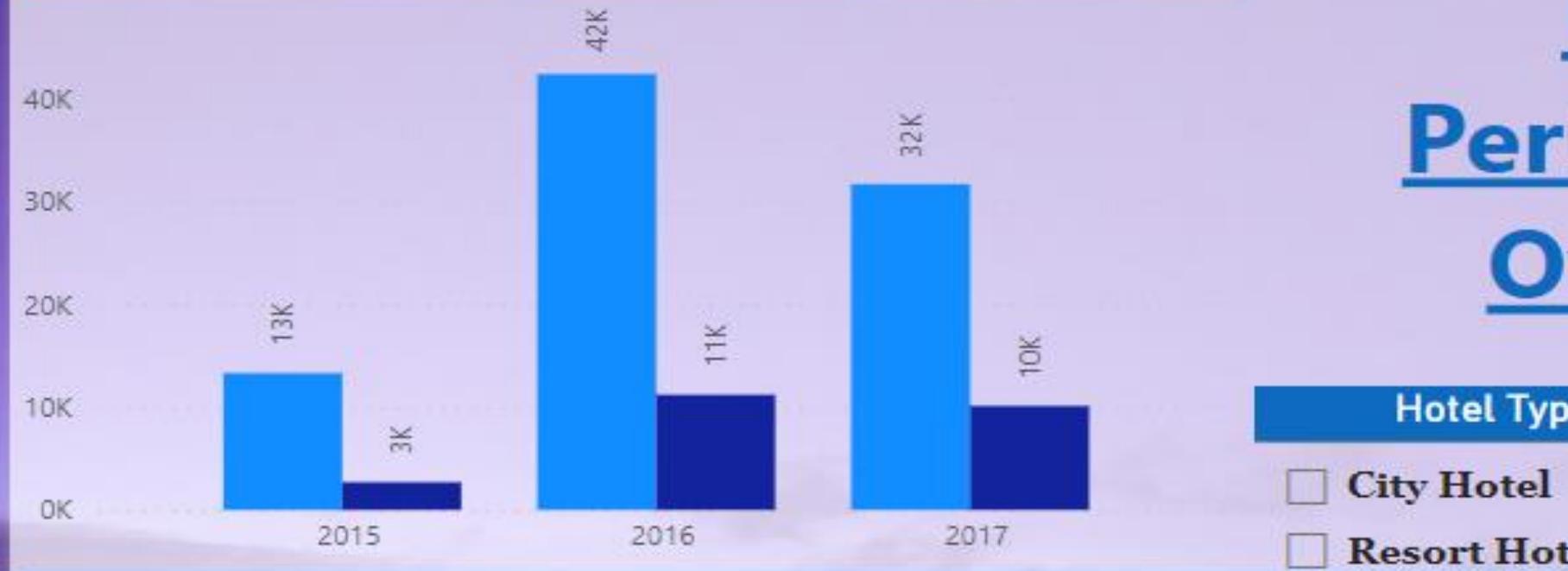
Room preference by Customer



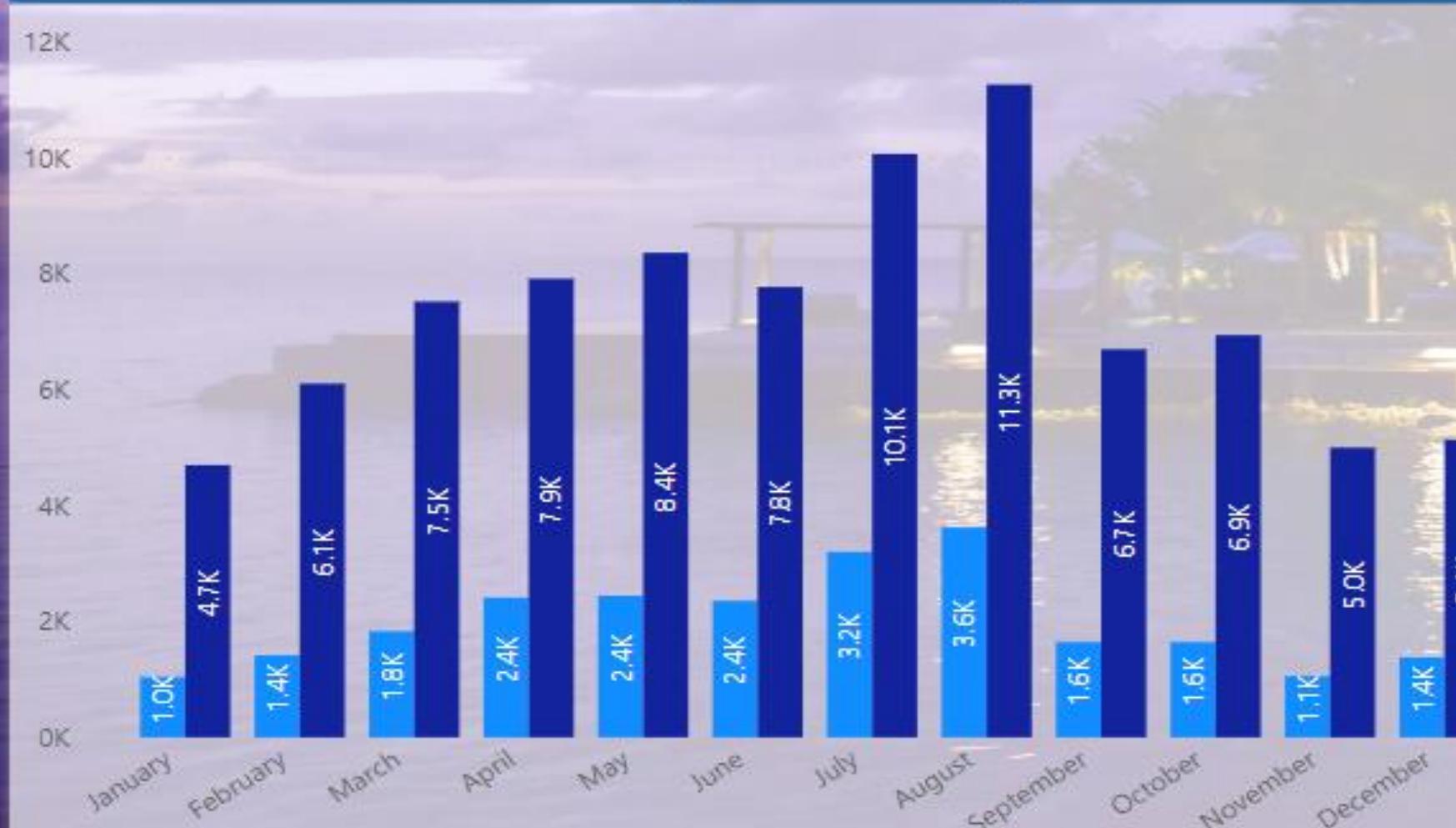


DASHBOARD

Year wise booking and Hotel booking cancelled



Month wise booking and Hotel booking cancelled



Hotels Performance Overview

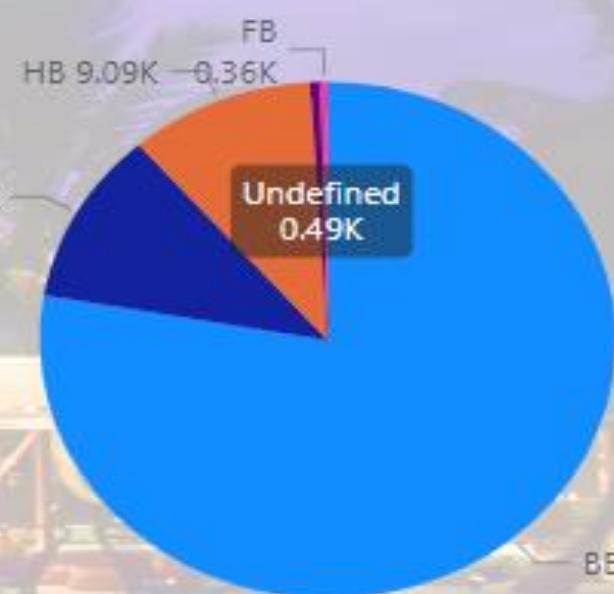
Hotel Types

- City Hotel
- Resort Hotel

Total Revenues

9.29M

Prefer meal by guest



Total Booking

87.39K

distribution_channel	Contract	Group	Transient	Transient-Party	Total
Corporate	948.00	2,389.20	2,36,491.58	1,08,299.40	3,48,128.18
Direct	1,075.20	10,380.70	12,73,027.90	1,32,943.45	14,17,427.25
GDS			21,777.53		21,777.53
TA/TO	2,89,128.58	33,123.00	63,96,830.22	7,86,804.03	75,05,885.83
Undefined			112.70		112.70
Total	2,91,151.78	45,892.90	79,28,239.93	10,28,046.88	92,93,331.49

Average of lead_time

79.89

Repeated Guest

3415



...

Business generation



Booking data bifurcation on the basis of year and preferences

Total bookings

87.22K

Total no. of people

177K

Room Preferences

reserved room type	City Hotel	Resort Hotel	Total
A	37834	18599	56433
B	989	3	992
C	13	901	914
D	10745	6629	17374
E	1457	4578	6035
F	1754	1066	2820
G	477	1573	2050
H		596	596
I		6	6
Total	53269	33951	87220

Year wise bookings received

hotel ● City Hotel ● Resort Hotel



Month Wise Bookings received

hotel ● City Hotel ● Resort Hotel



Preferences by Meal Choice

hotel ● City Hotel ● Resort Hotel



DERIVING CONCLUSIONS

- After doing this analysis we could conclude the following things:
- Maximum business is generated from the City hotel as compared to Resort Hotel.
- The Maximum number of bookings is generated in the year 2016.
- If we take an average of all the three years the maximum bookings have been received in the month of August.
- We could get an insight on the meal and room type preferences from these bookings which shows BB type of meals is preferred the most and Room type A and D are preferred maximum in the overall bookings.
- We got an idea on an average how many repeated guests the hotels can expect. That is nearly about 20 to 30 %.
- We could see that many people tend to book their stay without paying any advance deposit.
- The maximum bookings have been generated from the Online TA agent segment.
- We could see that there few months where booking orders exceeds so much that people need to wait for their bookings. (January and May month experience the highest waiting list.)
- Both these hotels receive bookings from many other countries also from which Portugal, UK and France are the top three countries.



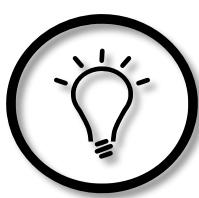
SUGGESTIONS :

Resort Hotel

- As we can see **Resort hotel not contributing towards the business as compared to that of City hotel, we can make some changes or implement the operational strategies used by City Hotel.**
- **We Can work on the advertising and branding part to attract more customers towards Resort hotel.**
- **We can add or enhance more niche activities and infrastructure and at Resort Hotel which creates an amazing experience for the customers and which at a whole will increase the revenue/business for them.**

City Hotel

- **City hotel is contributing quiet well to the overall business so to sustain its business it should continue with its operational strategy and can take few calculated risks to make any changes.**



SUGGESTIONS :



Common Suggestions

- Both the hotels should work on increasing Room type A and D in their infrastructure as that two room types are preferred the most.
- While they should focus on giving some good benefits or offers for other room types which makes them almost fully occupied even during non-seasonal time.
- Both of them should focus on advertising through various promotional means to showcase their infrastructural facility, services, cultural belonging and hospitality which in general will increase their visitors from all the corners of the world and increase their business more and more.

q u e s t i o n s

**THANK
YOU**

