Capstone Project -EDA
HOTEL BOOKING ANALYSIS

Team Members

Sarath Soman
Hariharapanda Deepak
Shyam Gadekar
Shrikant Kute

Hotel Booking-Exploratory Data Analysis

A hotel is a commercial establishment where bonafide travellers rent the rooms for temporary, overnight lodging with guest facilities.
A resort is a multi Amenity commercial establishment that provide vacationer or a tourist to obtain services like lodging, food, entertainment and shopping.

Hotel Booking-Problem Statement

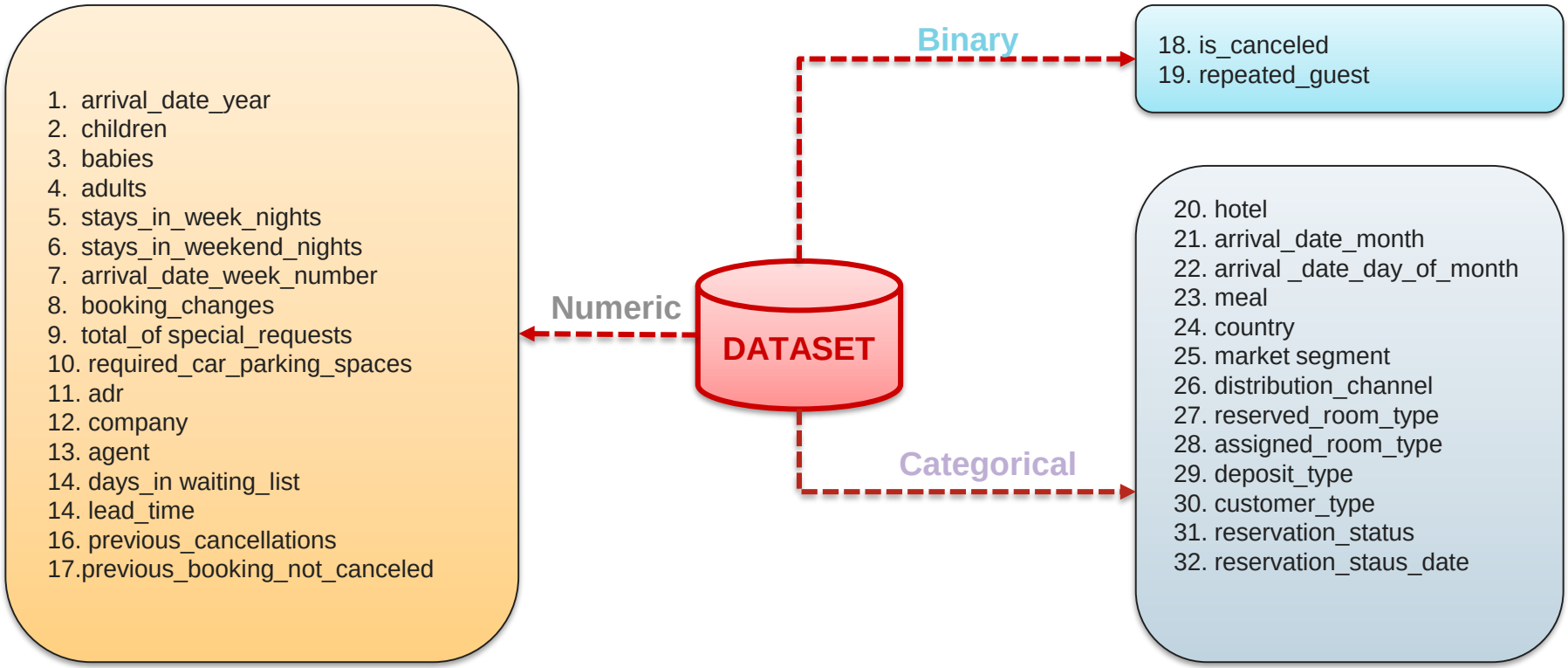
- Understanding the effect of different parameters effecting the hotel performance like when the booking was made, length of stay, number of persons staying, etc. For the period of 2015 July to 2017 August.
- Performing uni-variate, Hotel Wise, Distribution Channel wise, Lead an waiting time analysis and booking cancellation analysis for the data.
- Find out the key factors driving the hotel booking trends and make decisions based on them.

Agenda

To discuss the analysis of given hotel bookings data set from 2015 2017.We'll be doing analysis of given data set in following ways

- Univariate analysis
- Hotel wise analysis
- Time wise analysis
- Room type analysis
- Country Analysis
- Meal Analysis
- Booking cancellation analysis
- Country Analysis
- Market segment analysis
- Lead time analysis
- Customer type analysis
- Average Daily Rate analysis
- By doing this we'll try to find out key factors driving the hotel bookings trends.
- Conclusion

Data Summary



Data Summary (1/2)

Given data set has 32 columns which are:

- hotel : Hotel classification For example, a resort hotel or a city hotel.
- is cancelled : The column value indicates the cancellation type. Whether the reservation has been canceled or not. Values [0,1], where 0 indicates no cancellation.
- lead_time : Time between booking and actual arrival.
- arrival_date_year : year of arrival.
- arrival_date_month : month of arrival.
- arrival_date_week_number : week of arrival.
- arrival_date_day_of_month : date of arrival .
- stayed_in_weekend_nights : Weekend nights per booking.
- stays_in_week_nights : Week day nights per booking.
- adults : Number of adults visiting hotels
- children : Number children's visiting hotels
- babies : Number babies visiting the hotels.
- meal : Meal preferences per booking. [BB, FB, HB, SC, undefined].
- country : Country of origin of the guest.

Data Summary (2/2)

- market_segment : What segment does the customer belong to, e.g. business stands for travel agency, TA for travel agency.
- distribution_channel : How the customer accessed the stay - [direct, corporate, TA TO, undefined, GDS].
- is_repeated_guest : Whether the guest is coming for the first time. Values [0,1] ,1 indicated guest is repeated.
- previous_cancellations : no of bookings cancelled previously
- previous_bookings_not_canceled : no of previous booking which were not cancelled
- reserved_room_type : type of room which is reserved ['C' 'A' 'D' 'E' 'G' 'F' 'H' 'L' 'P' 'B]
- assigned_room_type : type of room which is assigned
- booking_changes : no of changes done in bookings
- deposit_type : type of deposit, ['No Deposit' 'Refundable' 'Non Refund']
- agent : booking is done through agent , then number of agent.
- company : company number
- days_in_waiting_list : Number of days between actual booking and transact.
- customer_type : Type of customers(Transient, group, etc.)
- adr : its average daily rate
- required_car_parking_spaces : no of parking spaces required for car parking
- total_of_special_requests : no special requests done by guests
- reservation_status : status of reservation ['Check-Out' 'Canceled' 'No-Show']
- reservation_status_date : date of reservation status

Data Cleaning & Assumptions taken:

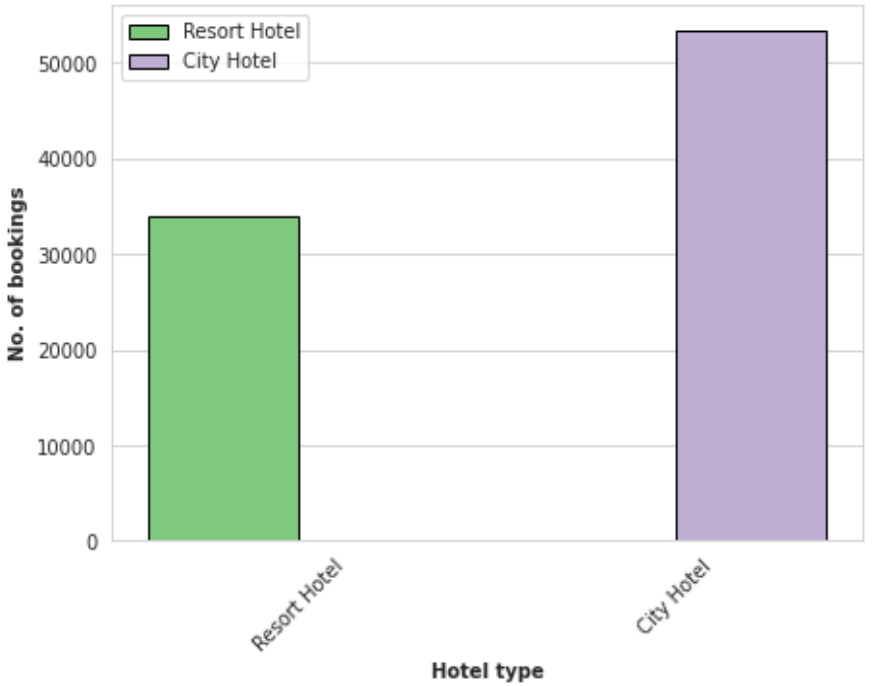
- **Data processing 1-** Removing Duplicates
  - There were 31994 duplicated rows in the dataset.
  - Removing Duplicates.
- **Data processing 2 -** Handling Missing Values

Column Name	Null Values
company	82137
agent	12193
country	452
children	4

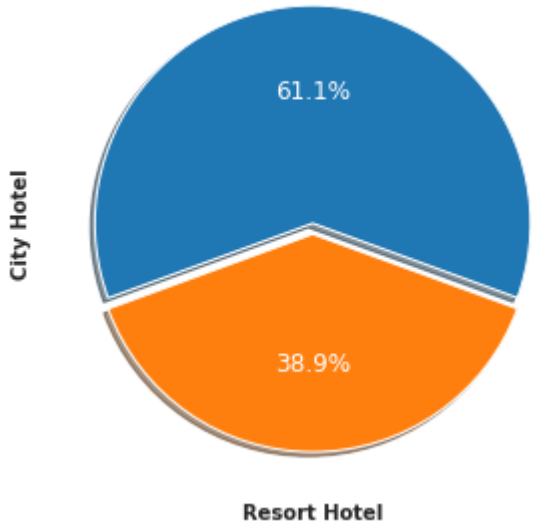
- **Assumption taken -**
  - 452 null values in the country column are considered "other“.
  - The column children has far fewer null values, so deleting it doesn't affect the analysis.
  - The company, agent column has a very high number of null values, assigned 0 to all null values in the column.

Shape of the cleaned data set after data cleaning is (87392,32)

Total number of bookings acc. to Hotel Type



% of bookings acc. to Hotel Type

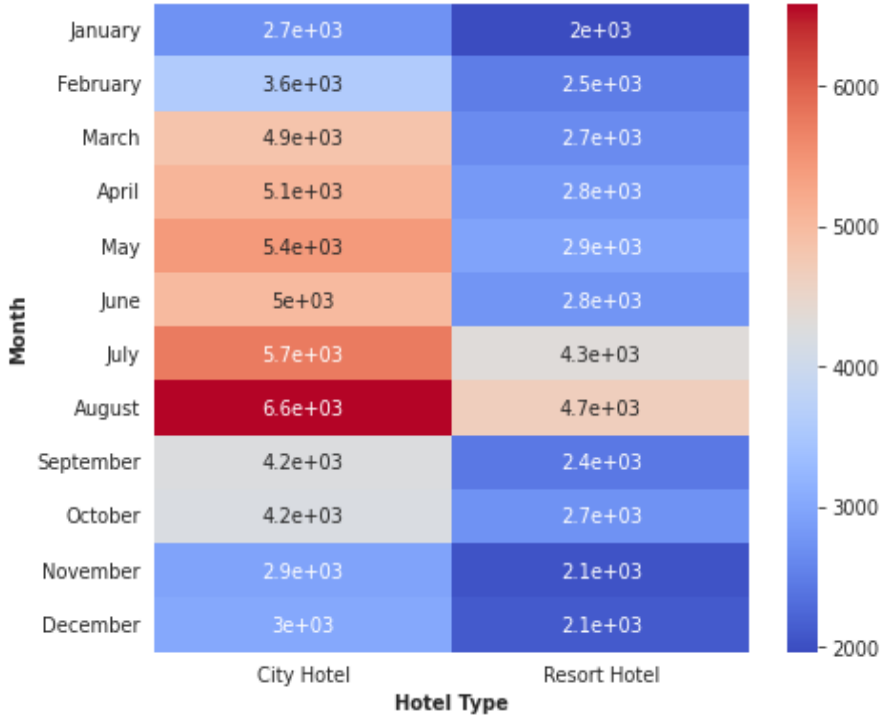


- *It has been clear from both the graph that **City Hotel** is preferred by **most of the customers** and it contributes to **61.1%** of the total booking made.*

Total number of bookings acc . to month

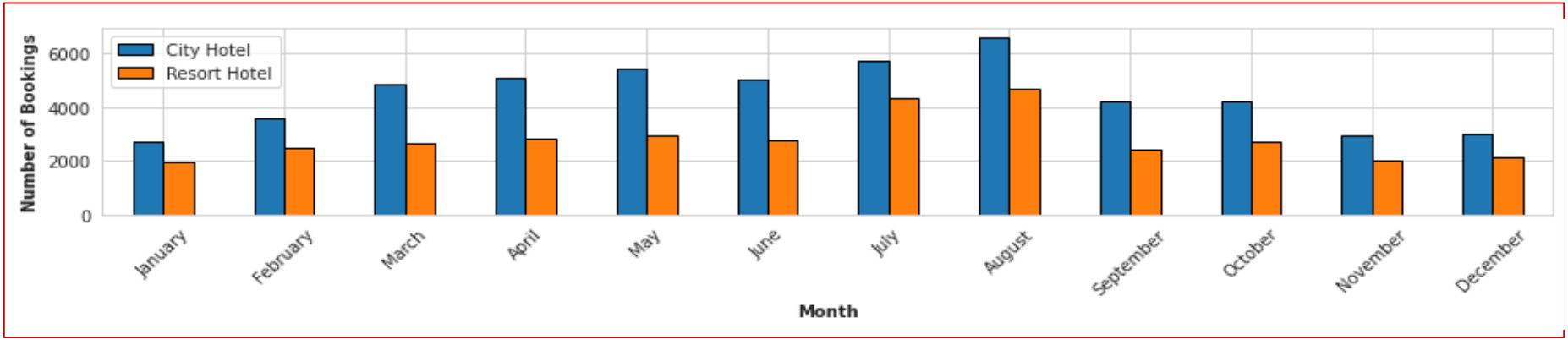
Month	No. of bookings	
	City Hotel	Resort Hotel
January	2730	1963
February	3605	2493
March	4856	2657
April	5080	2828
May	5413	2942
June	5005	2760
July	5744	4313
August	6587	4666
September	4240	2450
October	4208	2726
November	2942	2053
December	3014	2117

Heat Map

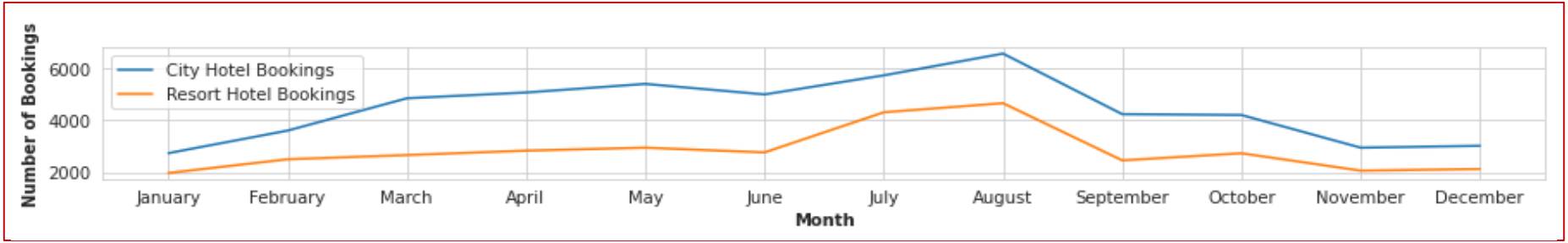


- *From **March to August**, bookings **increased**, and **August** saw the **highest** number of bookings.*
- *From the heat map and bar chart above, It can be concluded that **August** is the **busiest month** for **both the hotels**, followed by July and May.*

Total number of bookings according to months: Bar Plot

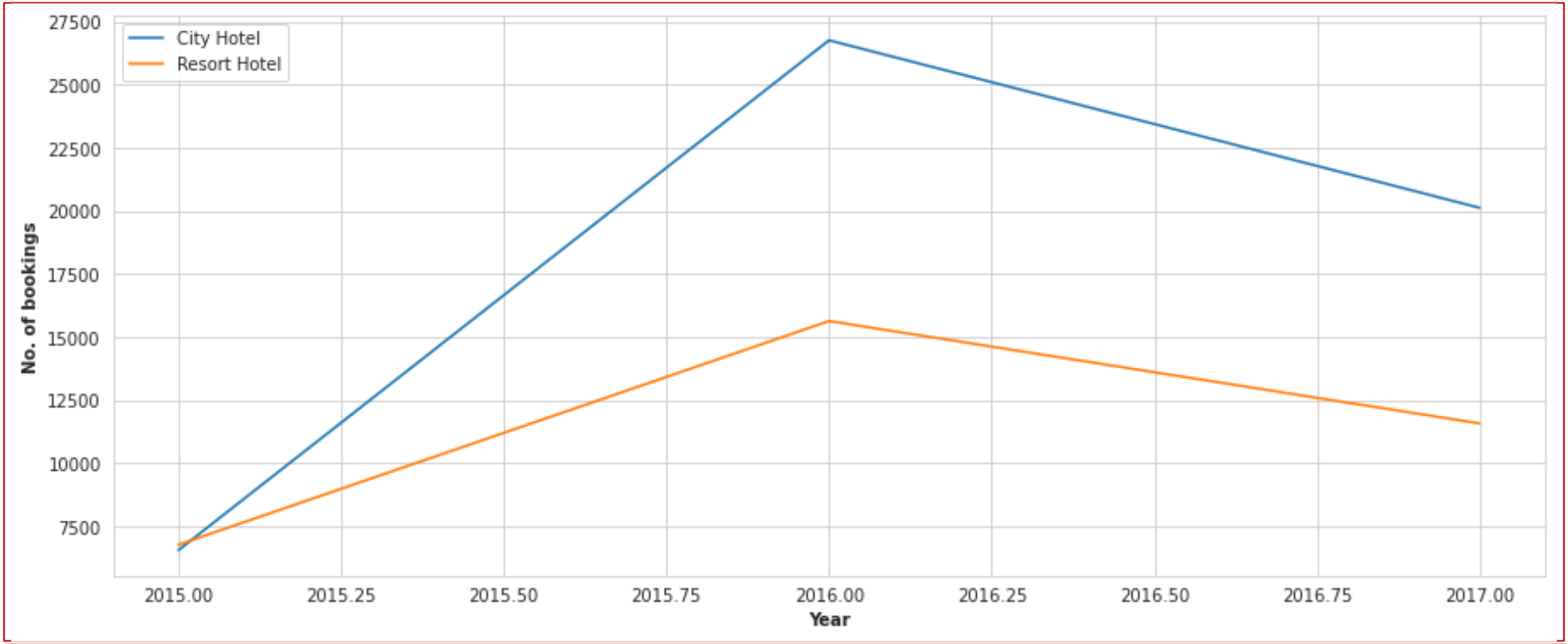


Total number of bookings according to months: Line Plot



- *It is clear from the bar chart that in the City and Resort hotel, the **fewest** bookings occur during the months of **November, December and January.***
- ***Bookings surged** from **February to August, although there was dip** in **June,** and **began** to decline **after September.***

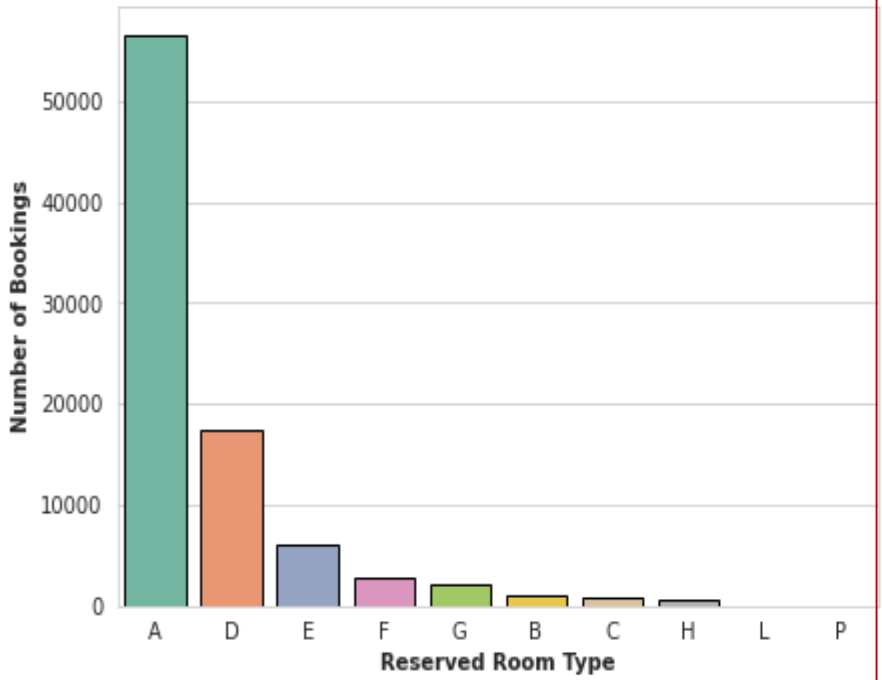
Total number of bookings according to year



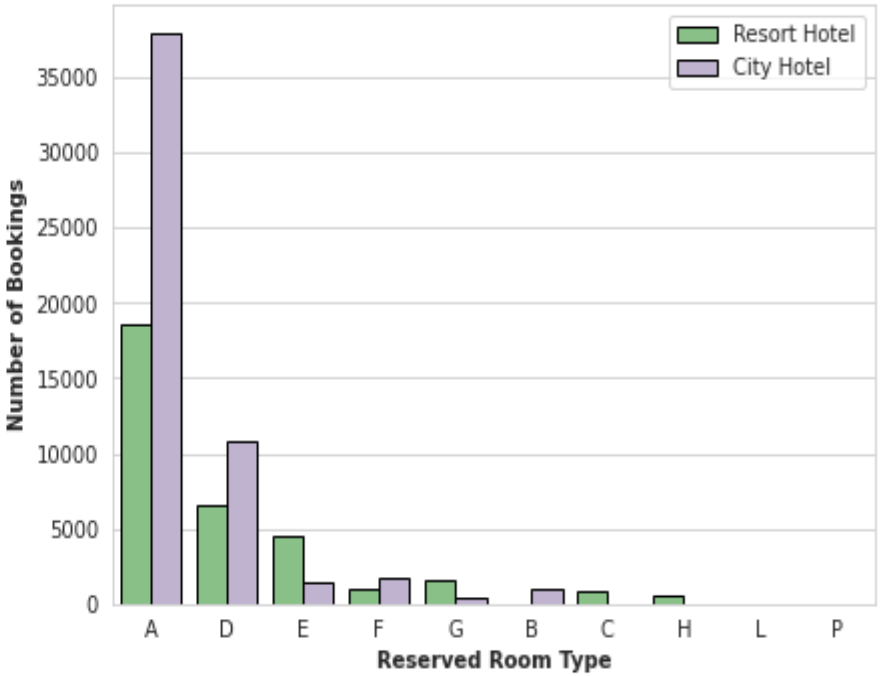
- *From the above line graph, it can be concluded that number of bookings* ***increased from year 2015 to 2016*** *and* ***started declining post year 2016.***

- ***Note: For year 2017 only data up to June is available.***

Room preference

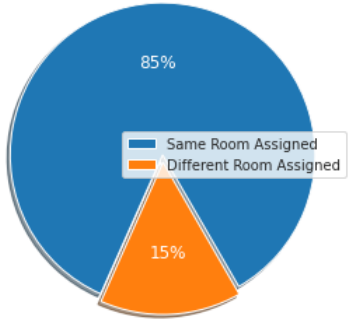


Room preference by Hotel Type

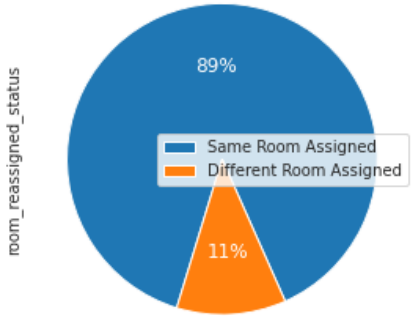


- *Type **A** rooms were **most reserved** by the guest in both the hotels followed by Type D,E,F,G rooms.*

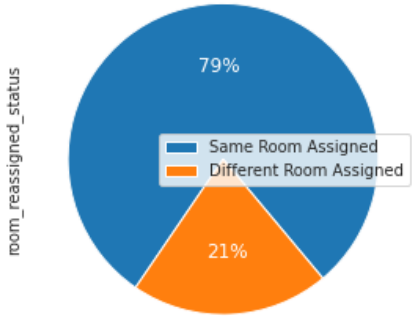
Proportion of rooms re-assigned: Overall



Proportion of rooms re-assigned: City Hotel

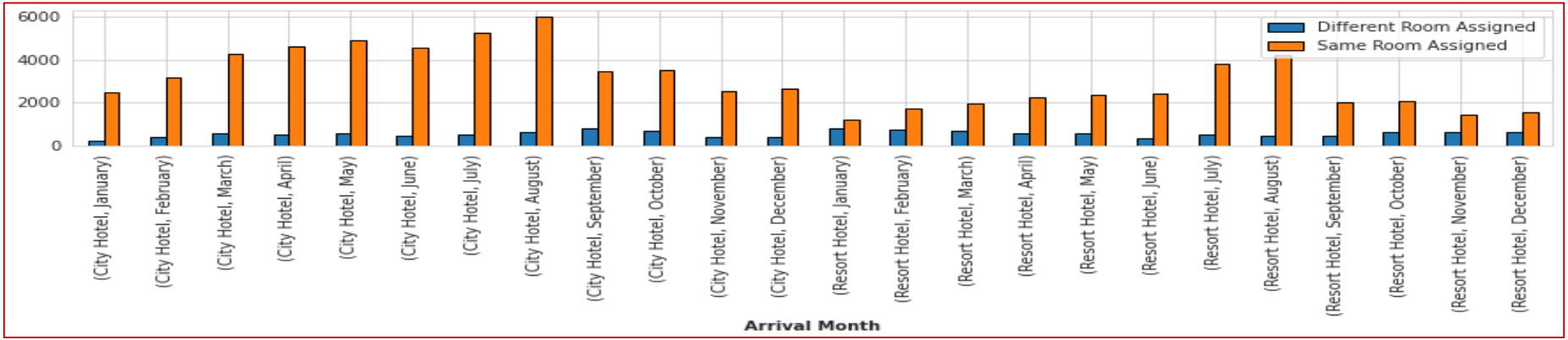


Proportion of rooms re-assigned: Resort Hotel

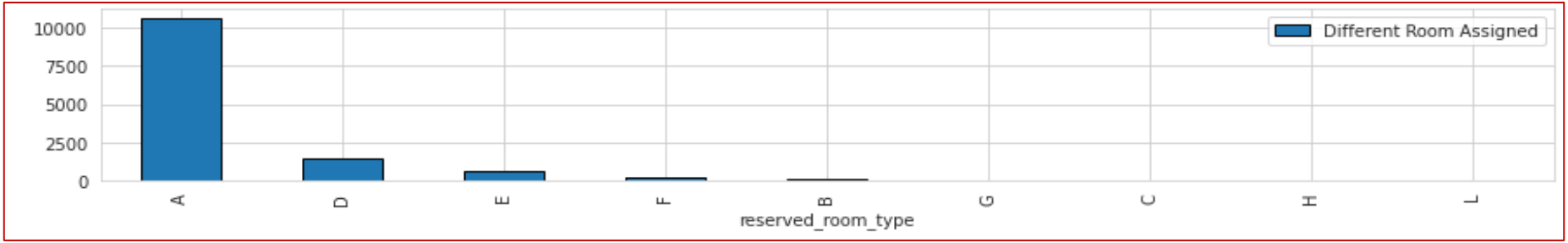


- *Over all **~15% room change** was observed through **out the year.***
- *City Hotel face **11% room change** and Resort Hotel face **21% room change** through out the 3 years.*

Room re-assigned status by Month

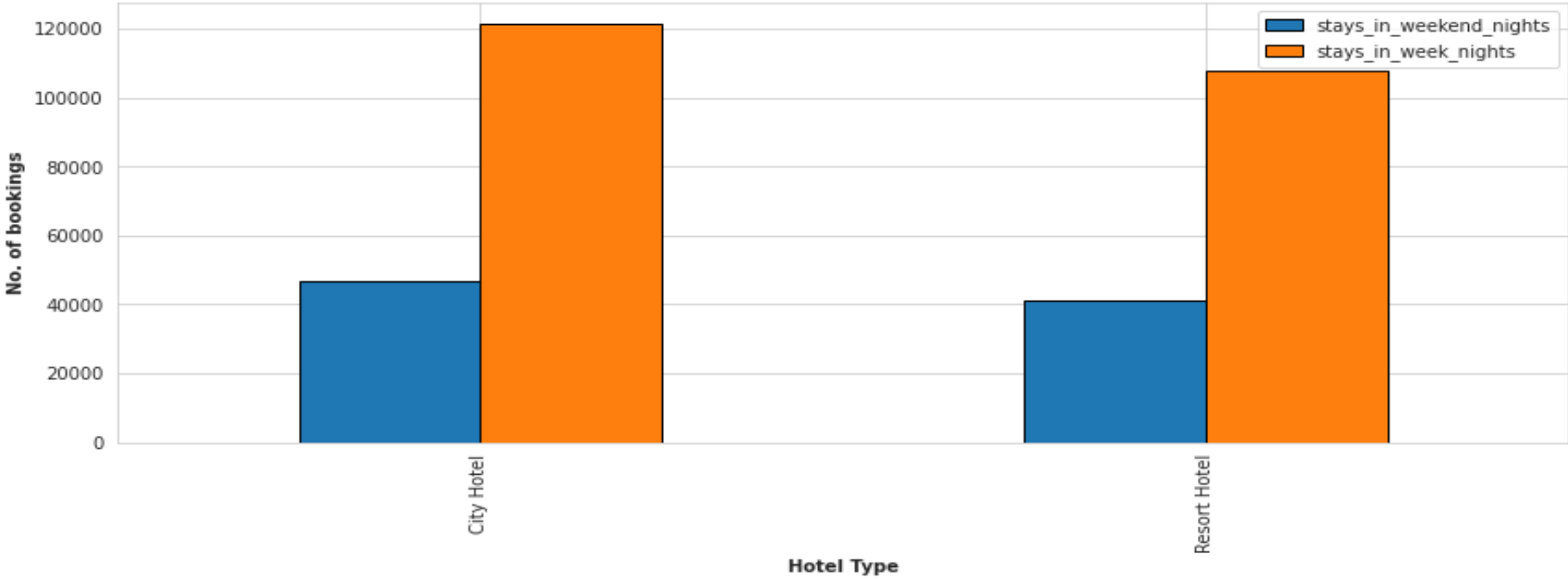


Room re-assigned by room type



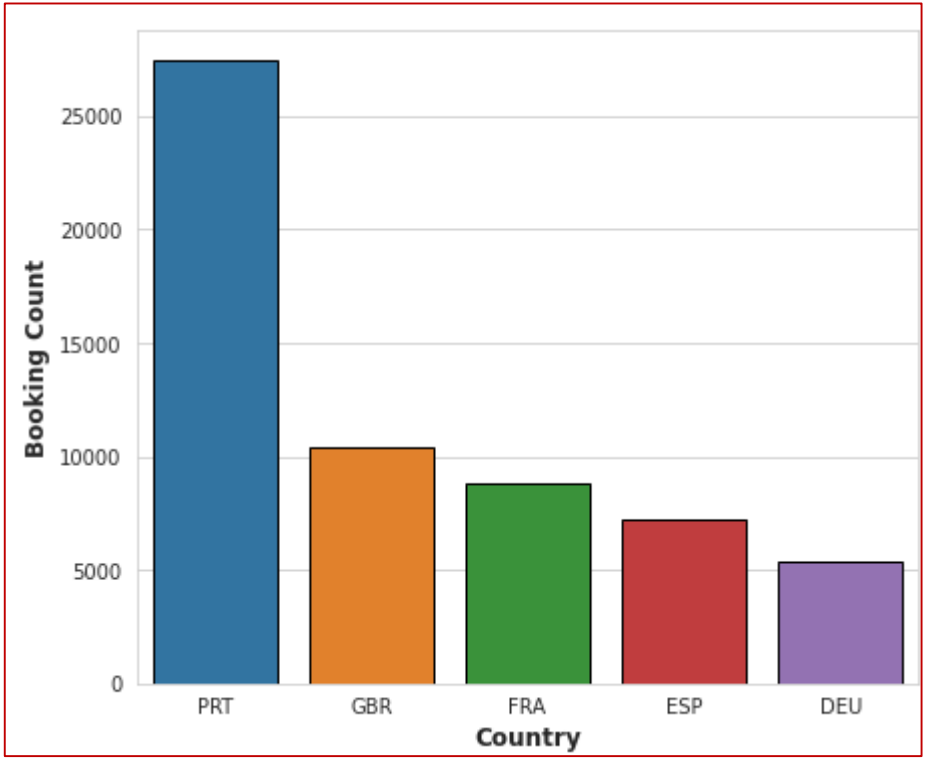
- ***Highest*** *number of changes in room were made in the month of the* ***September for City Hotel*** *and* ***January for Resort Hotel.***
- ***Type A*** *rooms are* ***most susceptible*** *to room shifting, followed by D, E, F, B, and C, and* ***insignificant in H and L.***

Number of stays on weekend nights/week nights

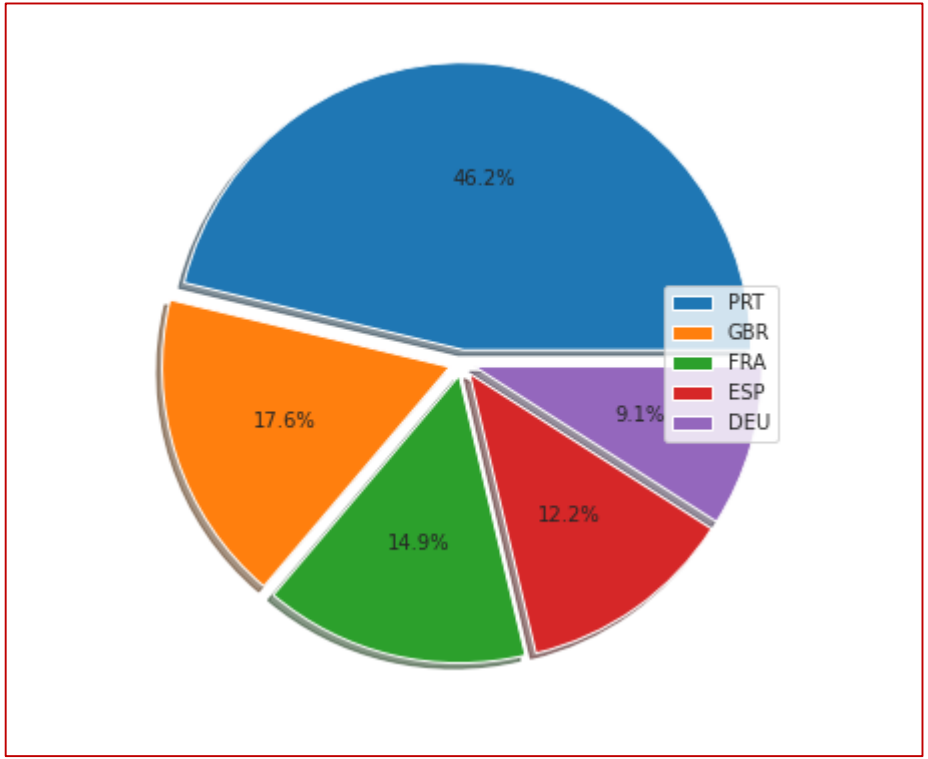


- *As we can see from the bar chart, guests prefer to **stay weeknights the most** at both hotels than weekend nights.*
- *Guests prefer to stay in **City Hotel on weekends the most.***

Top five countries with their booking counts

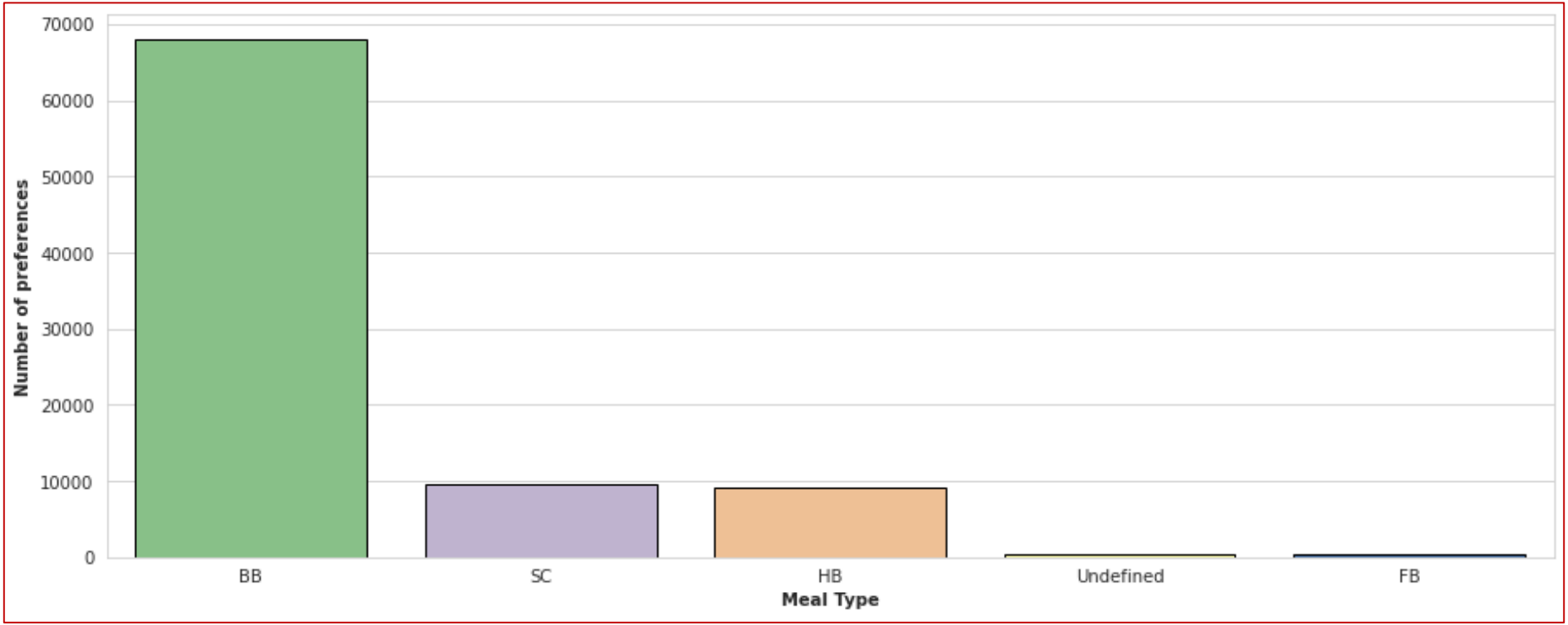


% Top five countries with their booking counts



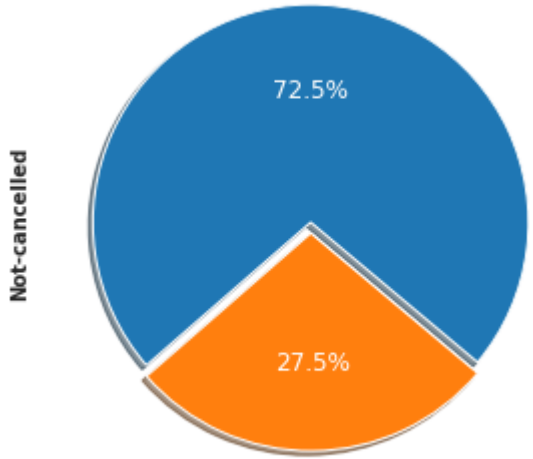
- ***Portugal (PRT)*** *had the* ***highest number of travellers*** *who booked hotels, followed by the* ***United Kingdom (GBR), France (FRA), Spain (ESP) and Germany (DEU)****.*

Meal Consumption Type



- *"**Bed & Breakfast (BB)"** meal type is **mostly preferred** by guests*

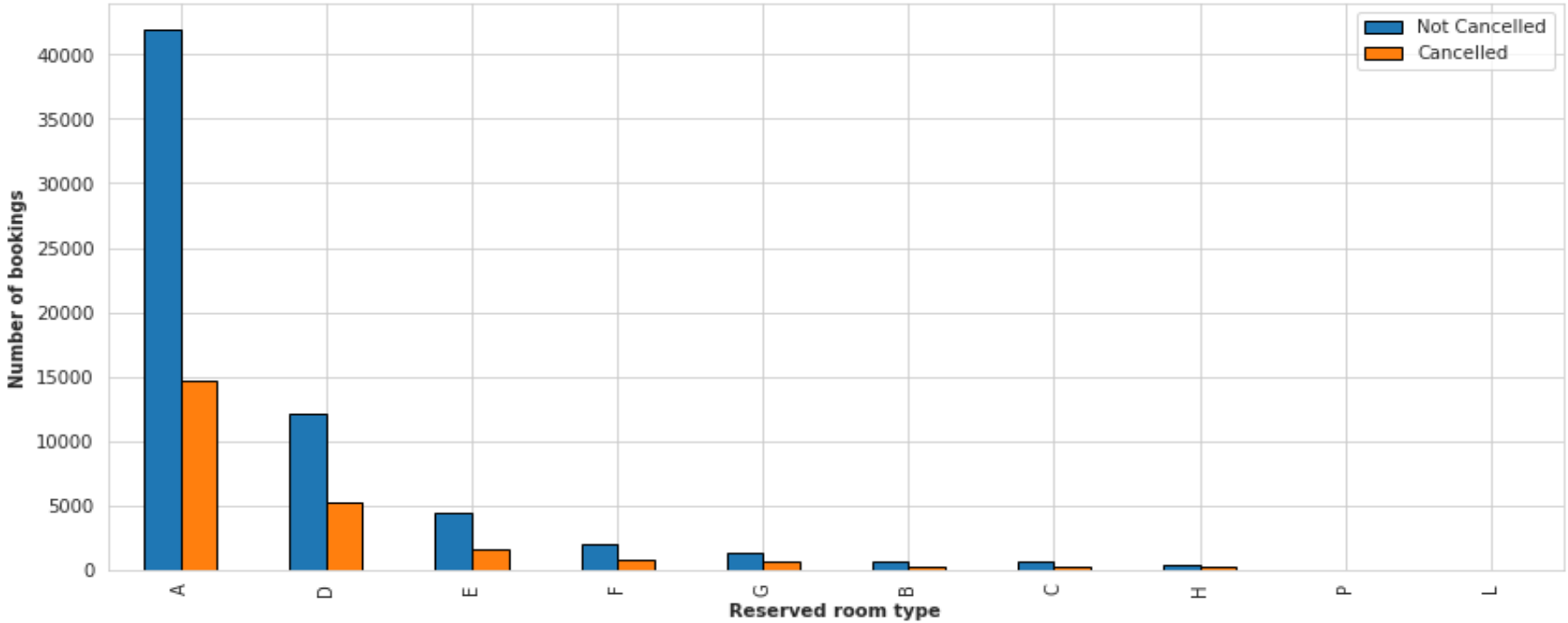
Overall cancellation %



% of cancellation acc. to Hotel Type

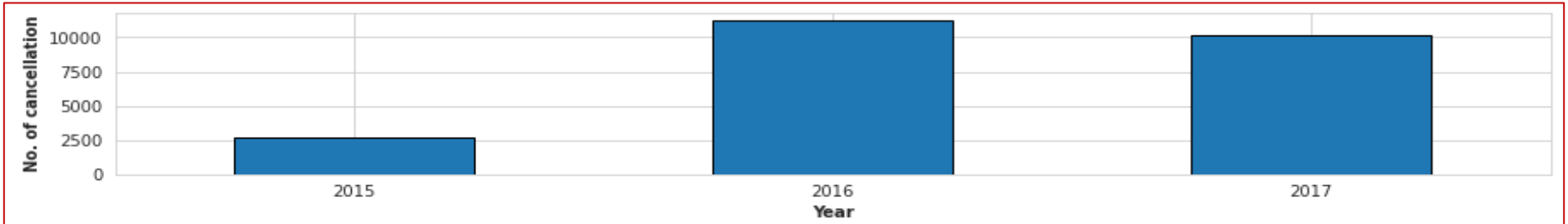
- *It is clear from the pie chart that **overall 27.5%** of cancellations are observed.*
- *From the graph it is observed that **City Hotel has more cancellations** than Resort Hotel.*

Cancellation by room type

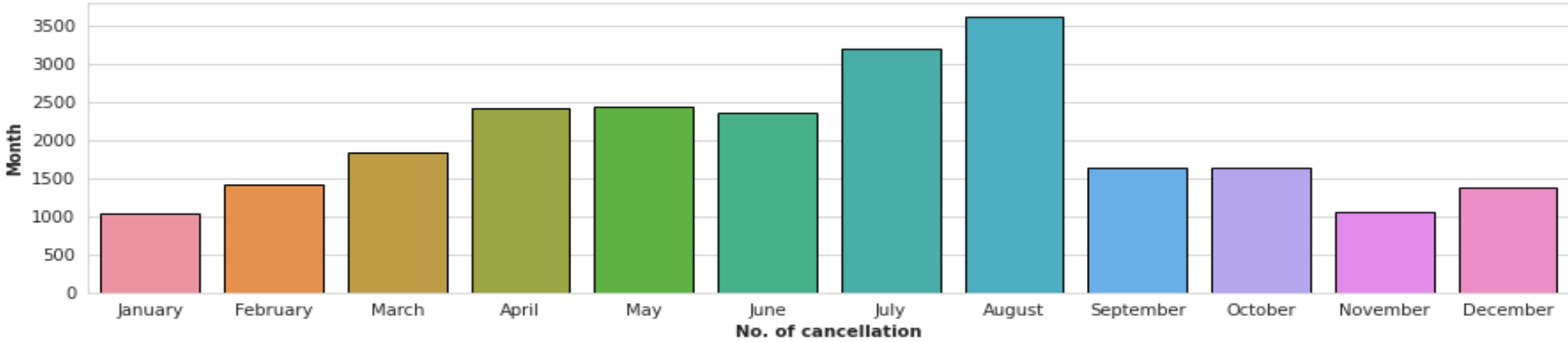


- *Analysis show that **Type A rooms are mostly cancelled** followed by D,E,F,G,B,C,H and negligible in P,L .*

Total number of cancellation according to years

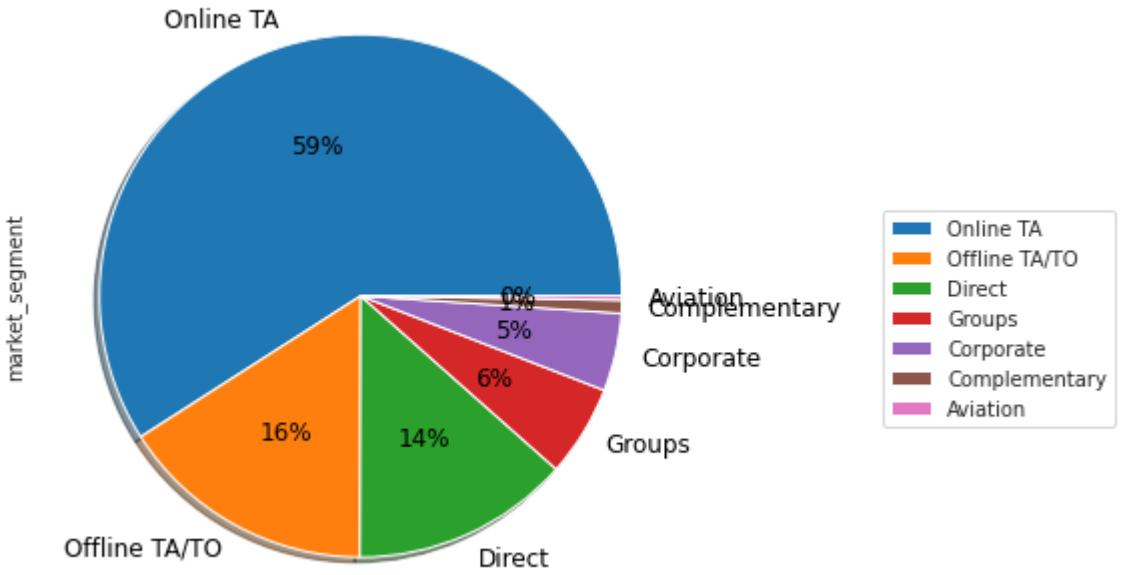


Total number of cancellation according to month



- *Year **2017** have only data up to **June month** and we can see from the bar chart that cancellation of booking is almost equal to year **2016** so it can be concluded that **2017 has the highest cancellation .***
- *August was the month that saw the **highest** number of **cancellations** and the **least in January**.*

Number of bookings by market segment



Market Segment	No. of bookings
Online TA	51617
Offline TA/TO	13889
Direct	11803
Groups	4942
Corporate	4212
Complementary	702
Aviation	227

- *Most bookings are made by **online Travel Agency**, followed by Offline TA/TO. Direct Booking is also lower but almost equal to offline TA/TO .*

Average length of time for reservations prior to arrival(lead time)

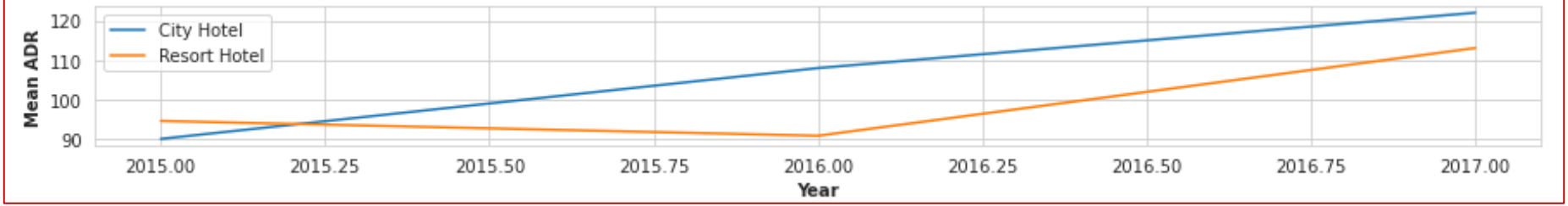
Hotel Type	count	mean	min	25%	50%	75%	max
City Hotel	53424.0	77.684112	0.0	14.0	50.0	**118.0**	629.0
Resort Hotel	33968.0	83.371938	0.0	8.0	47.0	**138.0**	737.0

- *It is clearly visible from the plot that most of the guests prefers to book the hotel* ***2-3 months*** *before arrival.*

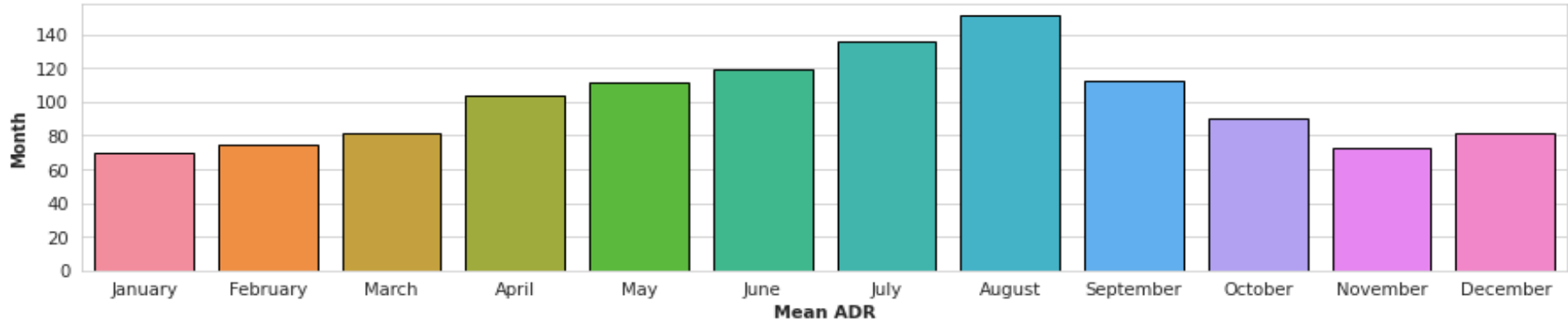
	Lead Time							Waiting Time							
country	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%	75%	max
BEL	2081.000000	94.290245	80.007766	0.000000	28.000000	75.000000	149.000000	396.000000	2081.000000	0.119173	4.283490	0.000000	0.000000	0.000000	0.000000	185.000000
BRA	1995.000000	80.864662	79.065161	0.000000	19.000000	55.000000	127.000000	354.000000	1995.000000	0.204010	4.771928	0.000000	0.000000	0.000000	0.000000	167.000000
CHE	1570.000000	87.754777	76.339889	0.000000	24.000000	68.000000	135.000000	457.000000	1570.000000	0.087261	2.661369	0.000000	0.000000	0.000000	0.000000	98.000000
CN	1093.000000	106.650503	88.937360	0.000000	33.000000	89.000000	166.000000	465.000000	1093.000000	0.143641	3.601288	0.000000	0.000000	0.000000	0.000000	109.000000
DEU	5387.000000	105.089103	89.521446	0.000000	31.000000	83.000000	165.000000	457.000000	5387.000000	1.051977	12.672101	0.000000	0.000000	0.000000	0.000000	224.000000
ESP	7252.000000	52.196773	61.621162	0.000000	8.000000	30.000000	73.000000	367.000000	7252.000000	0.149890	3.556934	0.000000	0.000000	0.000000	0.000000	207.000000
FRA	8837.000000	74.135906	71.760677	0.000000	17.000000	51.000000	111.000000	479.000000	8837.000000	0.846667	12.670392	0.000000	0.000000	0.000000	0.000000	379.000000
GBR	10433.000000	117.419055	100.141790	0.000000	33.000000	93.000000	180.000000	709.000000	10433.000000	0.377264	6.535174	0.000000	0.000000	0.000000	0.000000	150.000000
IRL	3016.000000	114.276857	87.299131	0.000000	41.000000	98.000000	168.000000	465.000000	3016.000000	0.043103	1.573874	0.000000	0.000000	0.000000	0.000000	61.000000
ITA	3066.000000	83.231246	75.470956	0.000000	19.000000	64.000000	128.000000	348.000000	3066.000000	0.961840	9.256337	0.000000	0.000000	0.000000	0.000000	174.000000
NLD	1911.000000	79.920984	75.768942	0.000000	16.000000	56.000000	128.500000	365.000000	1911.000000	0.358974	6.701947	0.000000	0.000000	0.000000	0.000000	185.000000
PRT	27449.000000	65.110277	87.644767	0.000000	3.000000	25.000000	98.000000	737.000000	27449.000000	1.279864	12.941031	0.000000	0.000000	0.000000	0.000000	391.000000
USA	1875.000000	68.748800	74.041411	0.000000	12.000000	44.000000	103.000000	542.000000	1875.000000	0.308800	5.686501	0.000000	0.000000	0.000000	0.000000	147.000000

- ***High lead time & high waiting time*** *suggests* ***the demand is high*** *and the* ***rooms are under-priced*** *and* ***vice-versa.***
- ***Low lead time & high waiting time*** *suggests the* ***rooms are getting cancelled due to unconfirmed booking.***
- ***High lead time & low waiting time is ideal condition for hotel****.*

Trend on ADR according to year

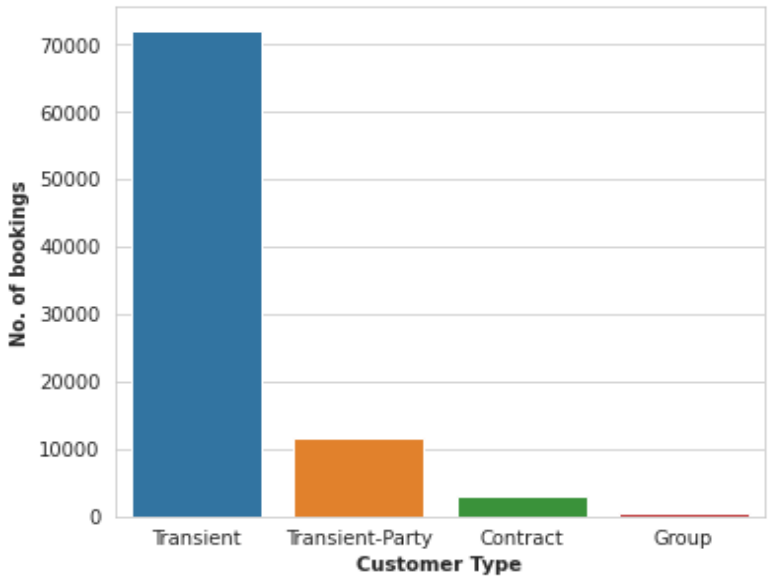


Mean ADR acc. to month

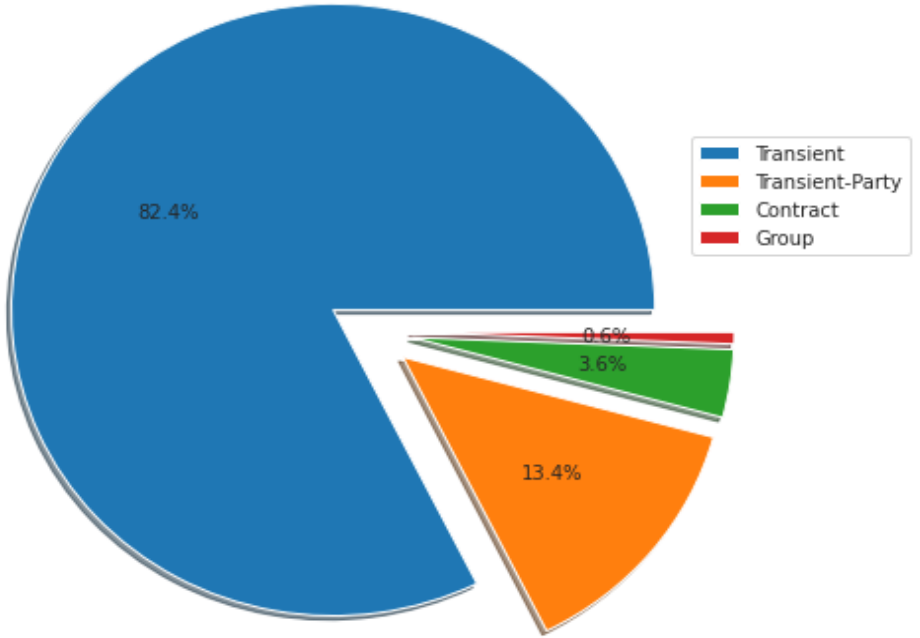


- *From the line plot we can see that **City Hotel had linear increase in ADR through out the year** and but for Resort Hotel **ADR was falling till pre 2016 and post 2016 it started to see a raise**.*
- *The ADR follows an i**ncreasing trend from January to August** and **decreasing trend from September to December.***

Customer Type



% of Customer Type



- *Here we can see the **maximum number** of customers are from **transient category** which is near about **75.1 %**.*

CONCLUSION

- The majority of the hotels reserved are city hotels. City hotel receives approximately 60% of bookings, while Resort hotel receives 40% of bookings; thus, City hotel is busier than Resort hotel. **City hotel will undoubtedly require the most targeted funding.**

- Bookings increased from March to August, with August having the highest number of bookings. The busiest month for both hotels is August, followed by July and May. The months of November, December, and January have the lowest bookings. **Both hotels should be prepared for peak season bookings from March to August, and hotel room preparation (room preparation, maintenance, staffing, etc.) should begin in November, December and January to handle the rush of bookings.**

- In both hotels, guests preferred type A rooms, followed by types D,E,F,G rooms. A 15% change in the room was observed overall. Room changes were most common in type A rooms, followed by D,E,F,B,C, and negligible in H,L.**So, it is clear that Type A are the most preferred followed by Type D rooms and most susceptible to room changes. Therefore, both hotels should focus on increasing the number of Type A rooms by replacing other types of rooms that are being booked by less number of guests. .This will result in increase of ROI for both hotels.**

- Guests prefer to **stay weeknights the most** at both hotels than weekend nights.

- The majority of the guests were from Europe. Guests from Portugal was the highest, followed by the United Kingdom (GBR), France (FRA), Spain (ESP), and Germany (DEU).**This may be due to the ease in visa process for this region.**

CONCLUSION

- Bed and breakfast (BB) is the most ordered meal, followed by SC (no meal), HB (half board), undefined, and FB (full board). **This may be because guests prefer to have breakfast in the morning before leaving the hotel for sightseeing and eat outside the hotel while sightseeing. Hotels should mostly concentrate on break fast quality and quantity for ROI.**

- A total of 27.5% of cancellations have been observed. The most bookings were cancelled in 2017, with the fewest in 2015. August had the highest number of cancellations, while January had the lowest. The City Hotel receives more cancellations than the Resort Hotel. **To minimize cancellations, hotels should track the price of their stay on other channels, including OTAs (online travel agencies) with which the hotel is not a partner. This is to ensure that their customers are not in a hurry to rebook their hotel rooms on other channels at lower prices.**

- The majority of bookings are made through online travel agencies, with offline TA/TO coming in second. Direct booking is also less, but it is nearly equal to offline TA/TO **. This must be because guests want to avoid all preparations/procedures prior to travel. To do this, they hire a travel agent to make all necessary travel arrangements and hotel reservations at best price. Hotel management should promote online TA and give special offer and packages for increasing their hotel bookings.**

- The ADR follows an increasing trend from January to August and decreasing trend from September to December. **ADR is a useful tool for pricing hotel rooms to maximize revenue. ADR values are high from March to August. This means that your maximum income will accrue during this month. An increase in price during this month can be considered in strategy for increasing the income.**

CONCLUSION

- Most of the guests prefers to book the hotel **2-3 months before arrival.**

Lead Time	Waiting Time	Demand	Room Price
High	High	High	Low
Low	Low	Low	High
Low	High	Room getting cancelled due to non-confirmed booking.	
High	**Low**	**Ideal**	**Ideal**

- **High Lead time & Low Waiting time is ideal condition for the hotel.**

- Maximum number of customers are from transient category which is near about 75.1 %.**Transients are usually walk-in or direct booking guests, or groups of guests who only stay 8-10 days, so their visit has a specific purpose. This is the largest market segment, so hotels should offer discounts, special services for this customer and can spend more money on advertising to get this customer segment.**

AI

THE END