Capstone Project
Hotel Booking EDA

Team Members

Saurabh Hase

Aakash Sharma

Ever wondered when the best time to book a Hotel Room

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests?
This hotel booking dataset can help you explore those questions!

Variables description

- **hotel** : It has information about the type of Hotel booked. (H1 = Resort Hotel or H2 = City Hotel).
- **is_canceled** : It indicates if the booking was cancelled (1) or not (0).
- **lead_time** : It tells about the number of days that elapsed between the entering date of the booking into the PMS and the arrival date.
- **arrival_date_year** : Year of arrival date.
- **arrival_date_month** : Month of arrival date.
- **arrival_date_week** : Week number of year for arrival date.
- **arrival_date_day** : Day of arrival date.
- **stays_in_week_nights** : Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel.
- **adults** : Number of adults.
- **children** : Number of children.
- **babies** : Number of babies.

- **meal** : It indicates type of meal booked. It includes the following categories. 1. Undefined/SC – no meal package; 2. BB – Bed & Breakfast; 3. HB – Half board (breakfast and one other meal – usually dinner); 4. FB – Full board (breakfast, lunch and dinner)
- **country** : It tells about country of origin. Categories are represented in the ISO 3155–3:2013 format
- **market_segment** : It is Market segment designation. In categories, the term TA means “Travel Agents” and TO means “Tour Operators”.
- **distribution_channel** : It includes Booking distribution channel. The term TA means “Travel Agents” and TO means “Tour Operators”.
- **is_repeated_guest** : Value indicating if the booking name was from a repeated guest (1) or not (0).
- **previous_cancellations** : It indicates number of previous bookings that were cancelled by the guest prior to the current booking.
- **previous_bookings_not_canceled** : It indicates number of previous bookings not cancelled by the guest prior to the current booking.

- **reserved_room_type** : It contains code of room type reserved.
- **assigned_room_type** : It contains code for the type of room assigned to the booking which may differ from the reserved room type.
- **booking_changes** : It indicates number of changes made to the booking.
- **deposit_type** : It indicates that if the guest has made a deposit to guarantee the booking. This feature can assume three categories: No Deposit – no deposit was made; Non Refund – a deposit was made in the value of the total stay cost; Refundable – a deposit was made with a value under the total cost of stay.
- **agent** : ID of the travel agency that made the booking.
- **company** : ID of the company that made the booking.
- **days_in_waiting_list** : It indicates number of days the booking was in the waiting list.
- **customer_type** : Type of booking, assuming one of four categories: 1. Contract - when the booking has an allotment or other type of contract associated to it; 2. Group – when the booking is associated to a group; 3. Transient – when the booking isn't part of a group/contract, and isn't associated to other transient booking. 4. Transient-party – when the booking is transient but associated with at least other transient booking

- **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** : The last status of reservation, assuming one of three categories: 1. Canceled – booking was canceled by the customer; 2. Check-Out – customer has checked in but already departed; 3. No-Show – customer did not check-in and did inform the hotel of the reason why
- **reservation_status_date** : Date at which the last status was set. This variable can be used in conjunction with the Reservation Status to understand when the booking was cancelled or when the checked checked-out of the hotel.

Data Cleaning

Null value count :

- country 488
- agent 16340
- company 112593
- children 4

Dealing with null values:

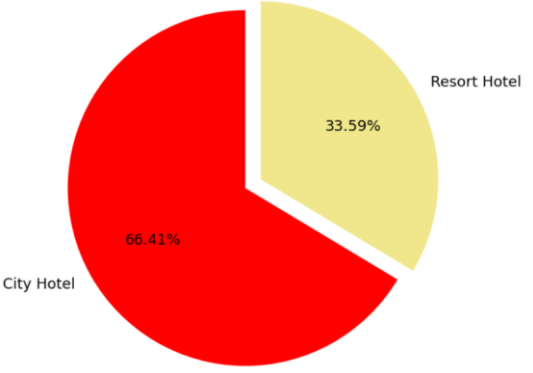
1. Company : Replaced all NAN with zeros
2. Agent : Replaced all NAN with zeros
3. Country : Replaced all NAN with ‘unknown’
4. Children : Replaced all NAN with mean

Cleaning unwanted rows :

Dropped the rows which is having zero guests

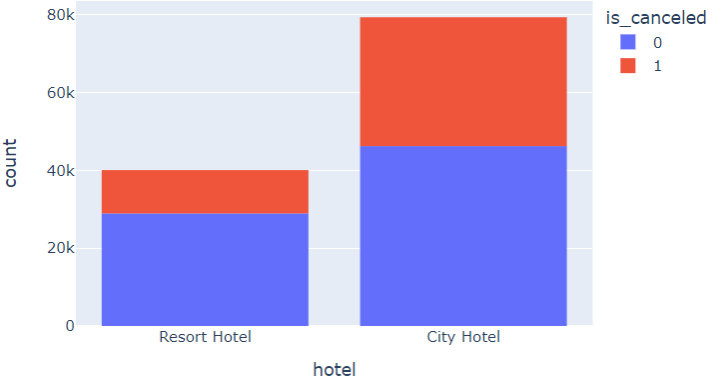
Exploratory Data Analysis

Overview of the type of Hotels



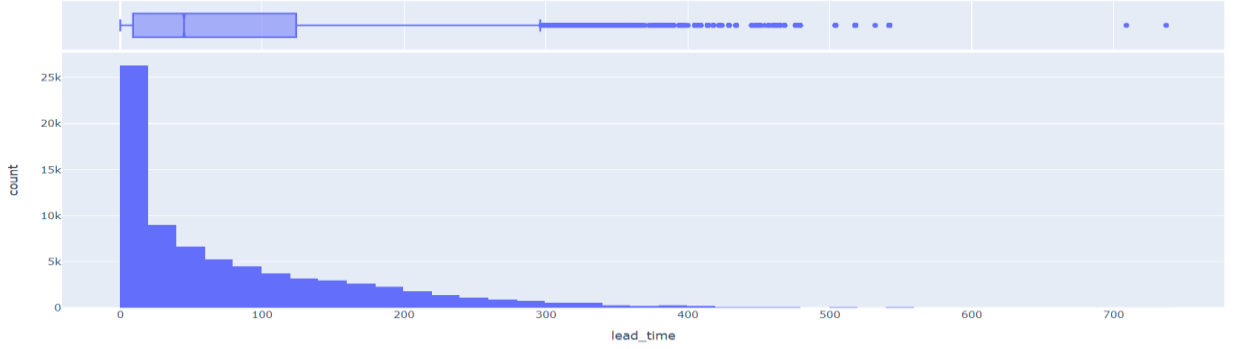
- Resort booking 33.59%
- City booking 66.41%

Cancelled bookings

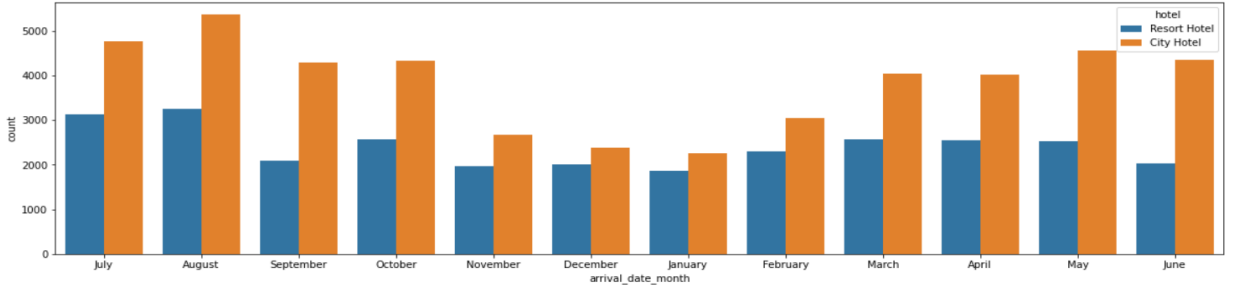


- The total Booking Cancellation percentage is 37.04 %
- The cancellation percentage in resort hotels is 27.76%
- The cancellation percentage in city hotels is 41.73%

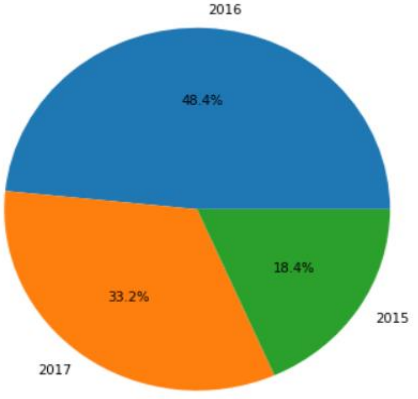
Lead time of booking



Month-wise booking

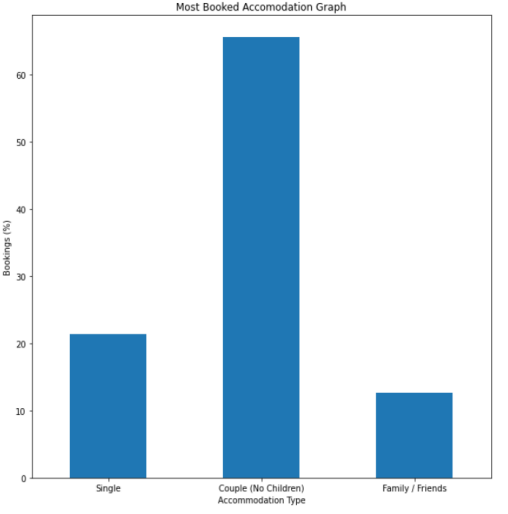


Year wise booking



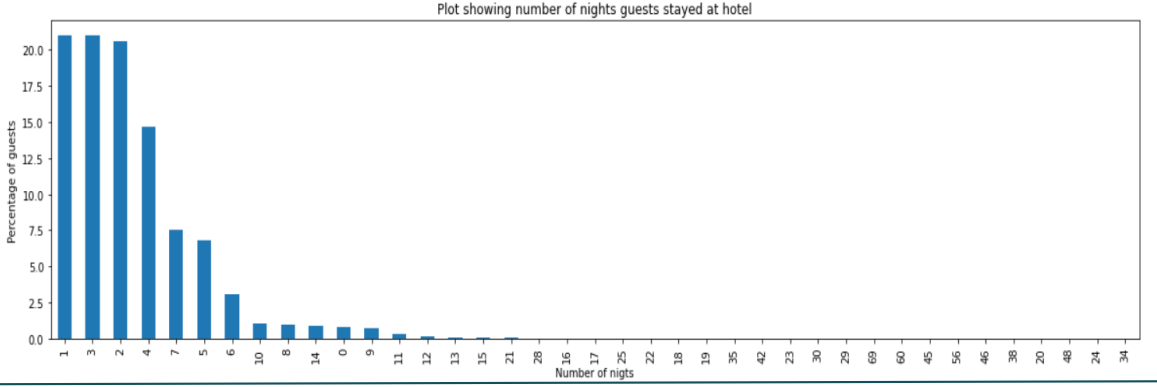
- Year 2015 lowest booking
- Year 2016 it increases
- Year 2017 it decrease again

Accommodation type



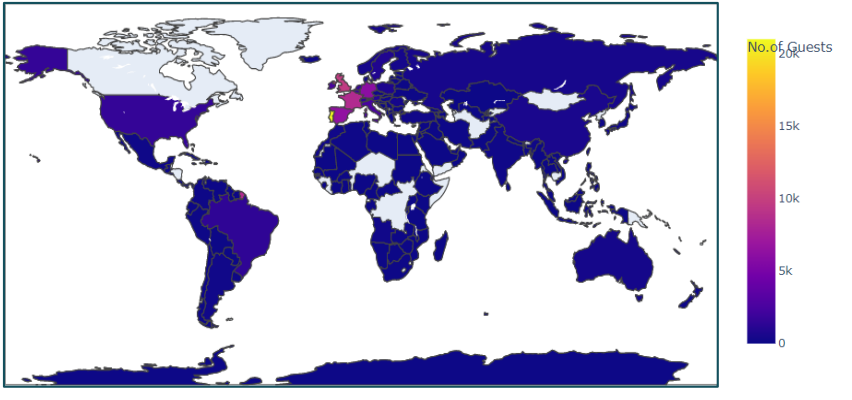
- Couples without children are the most frequent guests
- Family/friends are least

Duration of guest staying

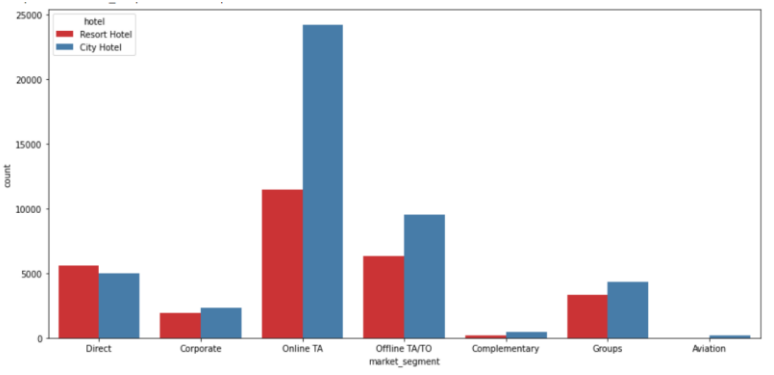


- Most of the guests stays for 1-4 nights
- Most of guest come from Portugal

Locality wise booking

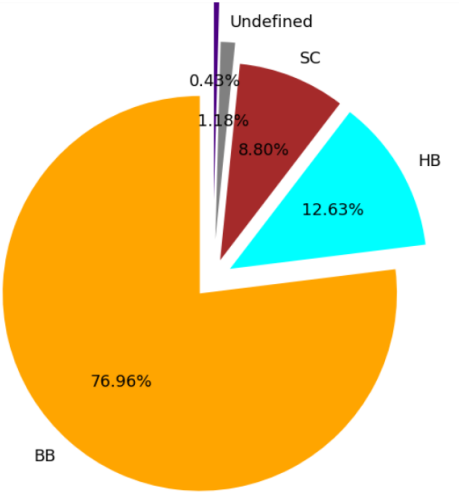


Booking methods



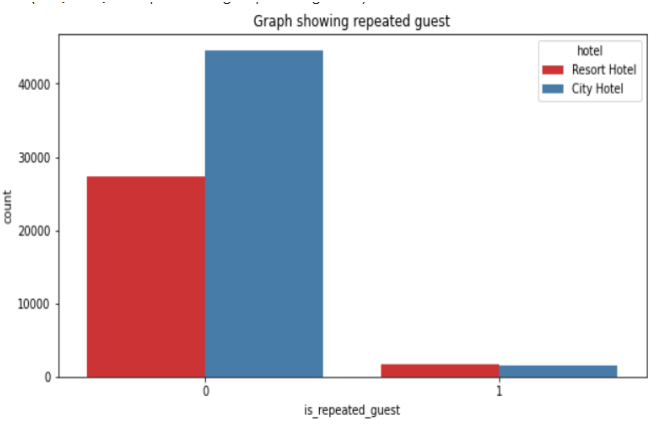
- Most of the Guests preferred online booking in city hotel as well as resort hotel
- Resort hotel does not preferred Aviation method

Preferred meal by guests



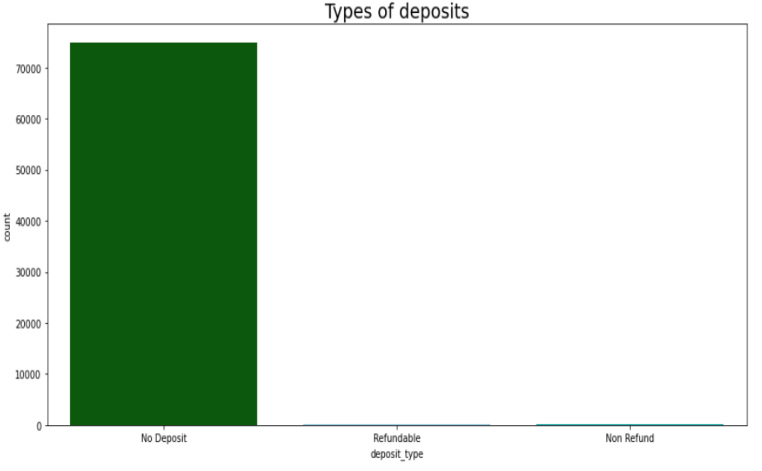
- 77% Guests prefers Bed & Breakfast meal
- 12.63% Guests prefers Half board

Repeated guest



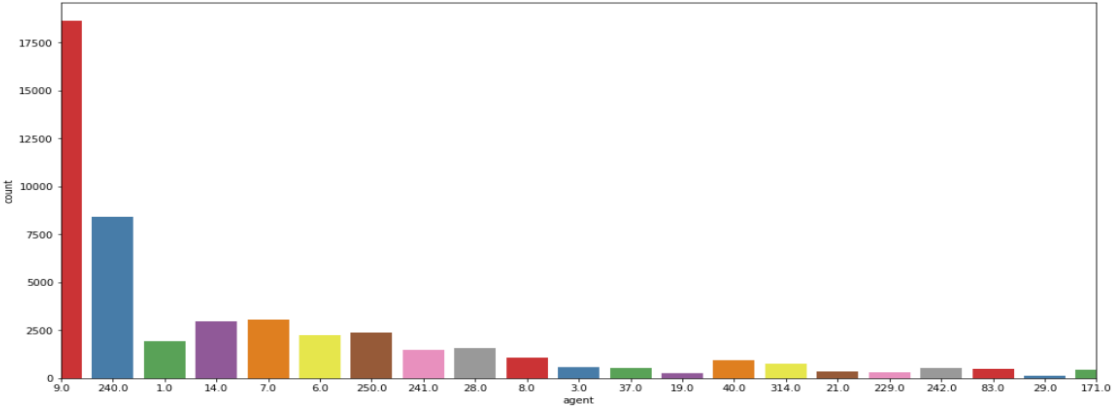
- Guests doesn’t repeat often

Distribution of deposit types



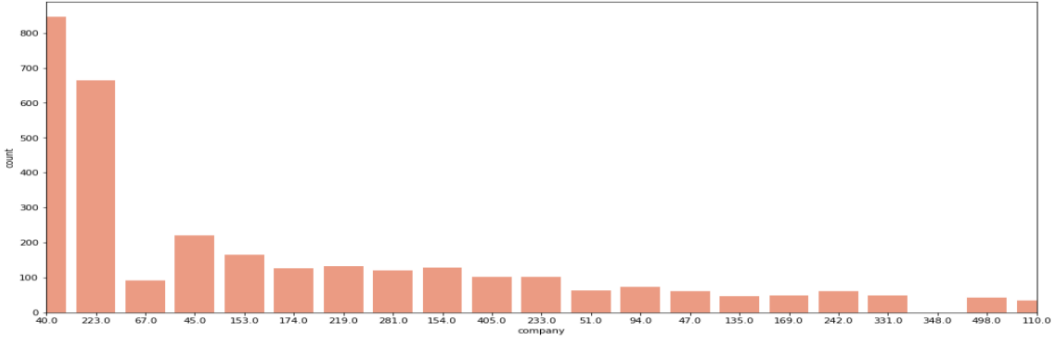
- Most of the guests doesn’t paid deposit

Agent wise booking



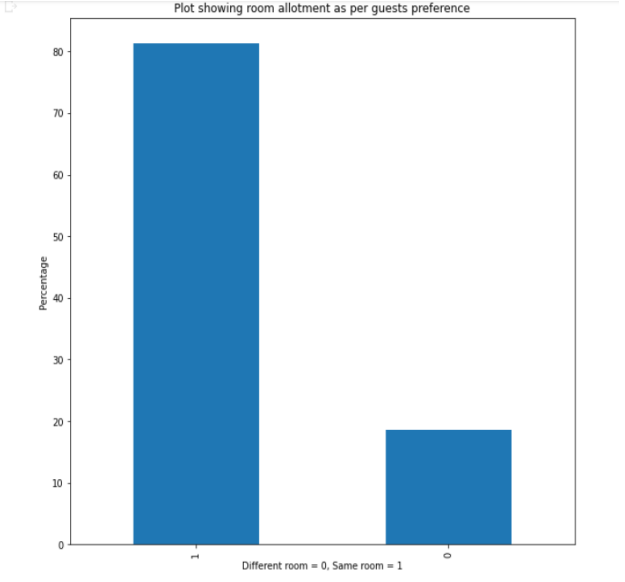
- Most of the boking get from Agent number 9

Company wise booking



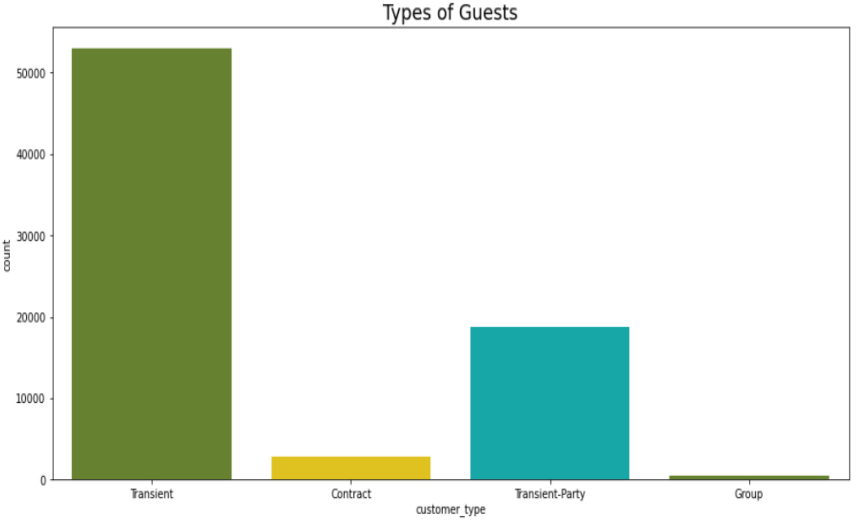
- Most of the boking get from Company number 40

Preference rooms



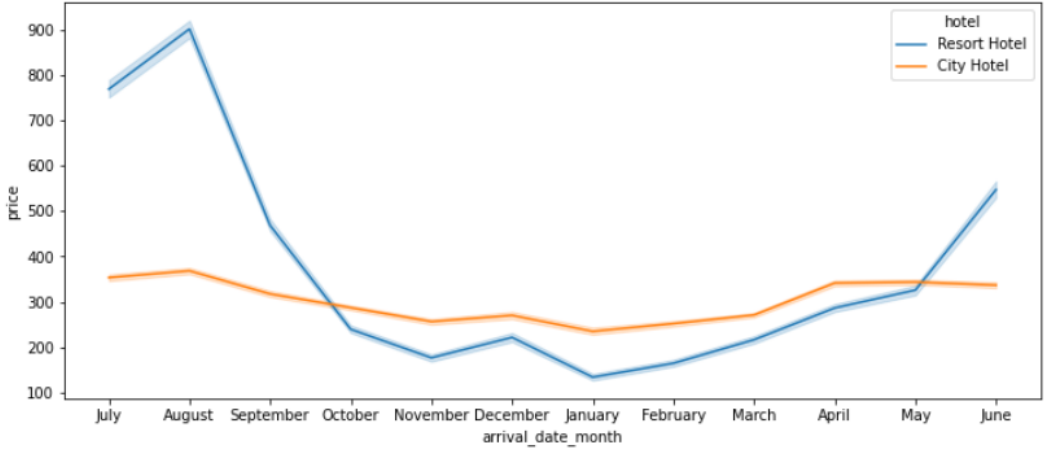
- Around 80% guests got the room they preferred while booking

Types of guests



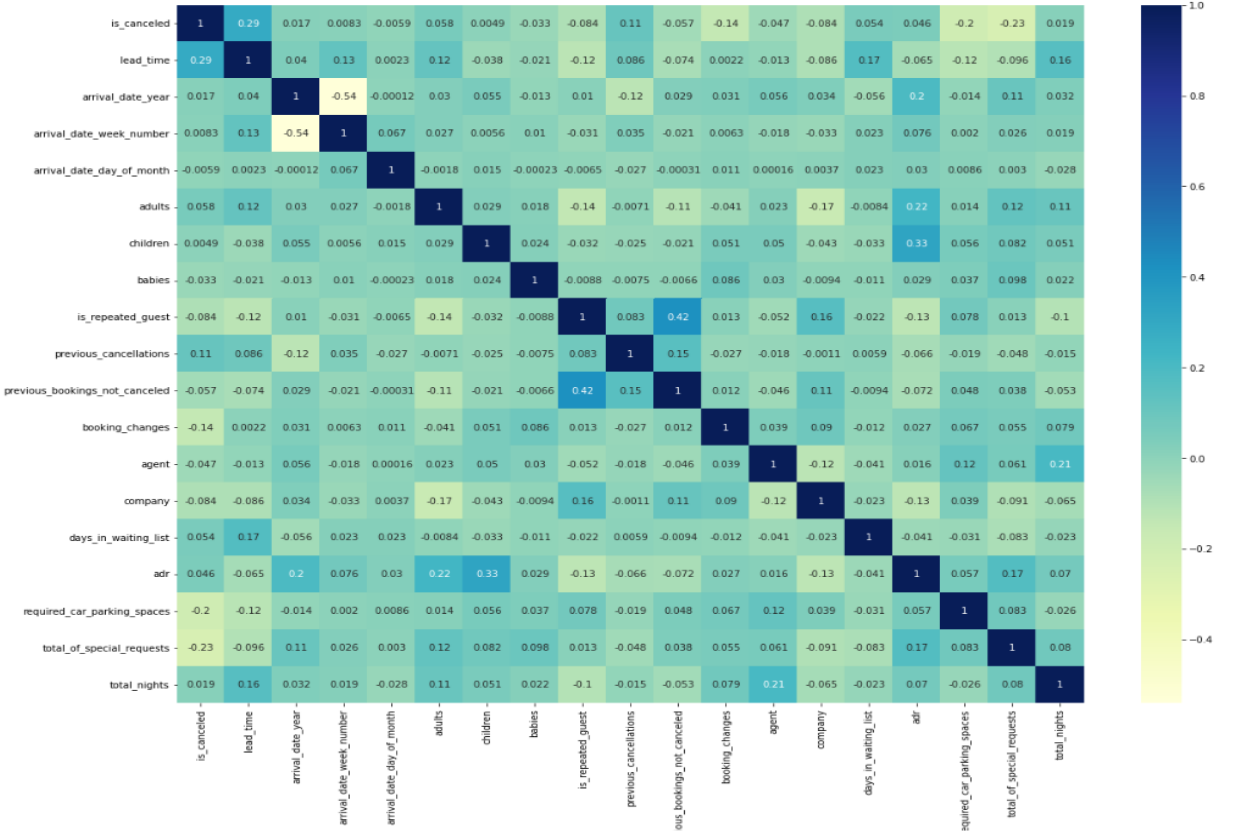
- Around 50k guests are Transient type
- Very few people do booking with group

Month wise pricing



- August month have most booking and so costing for august month is highest
- Autumn beginning months resort hotels have higher pricing than city hotels
- City hotel prices constant throughout the year

Correlation between variables



Conclusion

- Most of the booking is for city hotels compare to resort hotels.
- 63% of time the guests checked-in whereas 37% of time the booking got cancelled.
- Booking Cancellation percentage in Resort hotel is : 27.76 % while Booking cancellation percentage in City hotel is : 41.73 %
- Most frequent lead time for booking confirmation is 0 to 10 days
- Booking rate for year 2015, 2016 & 2017 is 18.4%, 48.4% & 33.2% respectively
- Summer end and Autumn beginning months have more bookings
- Only in 2016 we have booking for whole year otherwise in 2015 we have booking for only July to December and in 2017 we have booking from January to August
- Most of the guests stayed at the hotel for 2 nights
- Majority of the stays are over the weekday's night.
- Majority of bookings is for couple type accommodation rising up to 65%, followed by 21% for single type accommodation and family/friends type up to 12%

- Guests preferred to book meal BB Bed & Breakfast
- Mostly guests come from PRT country
- The online travel agents are preferred by the guests for reservation
- Number of repeated guests is low
- 87% of the guests got their preferred room.
- Majority of the booking does not require deposit
- Agent number 9 getting maximum booking
- Company no. 40 getting maximum booking followed by company no. 223
- Majority of the bookings are transient
- prices of resort hotels are much higher during summer end and Autumn start because booking rate is high for this months
- prices of city hotels are almost same throughout the year
- A mild correlation is found between repetition of the guest and not cancelling the booking which means the guest who are coming to the hotel again are not cancelling their bookings frequently.

Thank
you