**Hotel Booking Analysis**

**SUSOVAN PATRA**

**AlmaBetter, Bangalore**

**Abstract:**

## Exploring, analyzing and finding significance of each feature of given data by showing visual, numerical and tabular forms of different features.

## This can help in understanding varies how hotel booking process is happening in city hotel and resort hotel. We can compare which hotel is best in specific category at specific time.

# ****Importance of Hotel booking analysis:****

Businesses around the world are waking up to the importance of data analytics. Getting the right insights gives you an advantage in a competitive business environment. The hotel industry is no different.

Data analytics in the hotel industry is key to marketing strategy, building customer loyalty, and enhancing productivity. It enables hotels to personalize experiences for their guests, introduce better hotel pricing strategies, and expand their customer base. Here are some ways in which data analytics makes a positive impact on the hotel industry.

* + 1. Customer Data Analysis & Market Segmentation
    2. Real-Time Data and Hotel Pricing Strategies
    3. Inventory Management
    4. Managing Hotel Booking Channels
    5. Demand Forecasting

1. **Problem Statement**

The data two types of hotels one is city and other is resort hotels. The problem statements are

## Best time of year to book a hotel room

## Optimal length of stay in order to get the best daily rate

## Predict whether or not a hotel was likely to receive a disproportionately high number of special requests

**Feature Explanation:**

There are 32 different features are given in dataset. few important features are

1. hotel : there are two types of hotels namely City and Resort Hotel
2. is\_canceled : This column represents

whether the booking is cancelled or not

1. lead\_time: Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
2. arrival\_date\_year: Year of arrival date
3. stays\_in\_weekend\_nights: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
4. stays\_in\_week\_nights: Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
5. meal: Type of meal booked. Categories are presented in standard hospitality meal packages: Undefined/SC – no meal package; BB – Bed & Breakfast; HB – Half board (breakfast and one other meal – usually dinner); FB – Full board (breakfast, lunch and dinner).
6. country: Country of origin. Categories are represented in the ISO 3155–3:2013 format.
7. market\_segment: Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”.
8. distribution\_channel: Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”.
9. is\_repeated\_guest: Value indicating if the booking name was from a repeated guest (1) or not (0)
10. reserved\_room\_type: Code of room type reserved. Code is presented instead of designation for anonymity reasons.

assigned\_room\_type: Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request.

1. Code is presented instead of designation for anonymity reasons.
2. days\_in\_waiting\_list: Number of days the booking was in the waiting list before it was confirmed to the customer.
3. customer\_type: Type of booking, assuming one of four categories: Contract - when the booking has an allotment or other type of contract associated to it; Group – when the booking is associated to a group; Transient – when the booking is not part of a group or contract, and is not associated to other transient booking; Transient-party – when the booking is transient, but is associated to at least other transient booking.
4. adr: Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights.
5. total\_of\_special\_requests: Number of special requests made by the customer (e.g. twin bed or high floor)

**2. Introduction**

Customer wants to book a hotel. There are two hotels are available. One is city hotel and other is resort hotel. He want to know which hotel is best. So with EDA analysis we can conclude the best hotel to book. Each hotel has its own merits and demerits. For example specific hotel has high waiting list then the hotel is very popular and too many people booking the hotel at specific time.

### Our goal here is to compare both hotels with respective each other with some EDA analysis in varies fields of interest and get to some conclusion.

## 

## 3. Data Cleaning

Data cleaning is the most import task in any EDA analysis.

1. Checking null values:

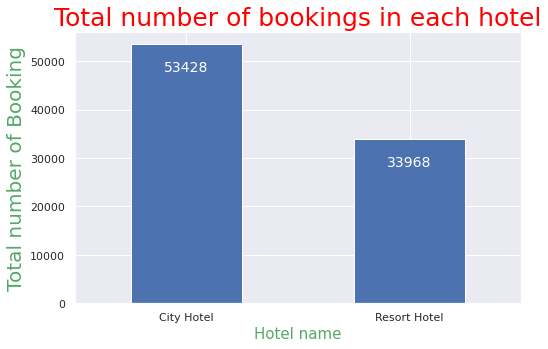
In the given data there are four different columns have null values. Percentage of null values is very high for columns agent, company. So we have removed company column and filled 0 as new id for agent null values. For rest two (county, children) we have removed the rows as there percentage of null values is less.

1. Dropping duplicate values:

Looks like there are quite a few duplicate values present in data which we have dropped.

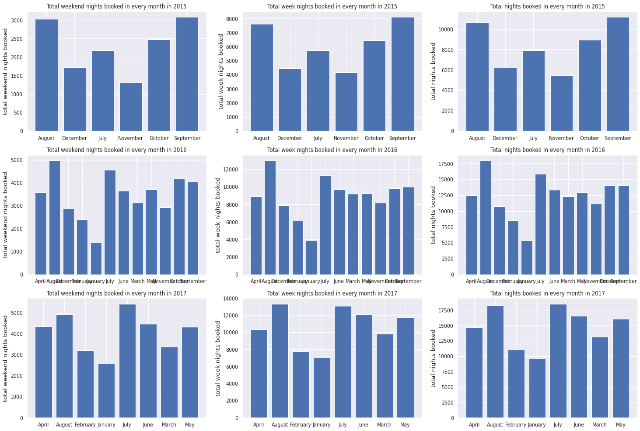
## 4. Exploratory Data Analysis

1. **Total number of booking in each Hotel:**

****

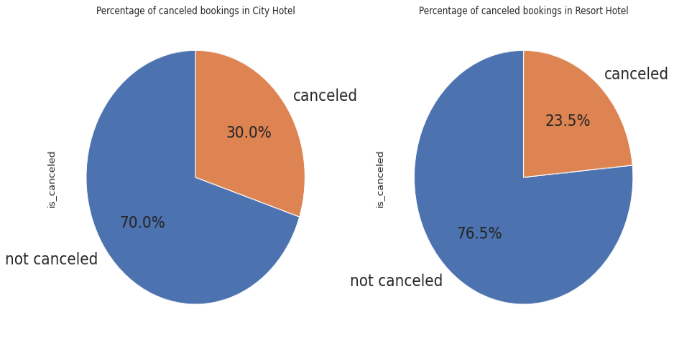
The total number of bookings happening in City Hotel is higher than Resort Hotel

1. **Year wise hotel bookings:**

****

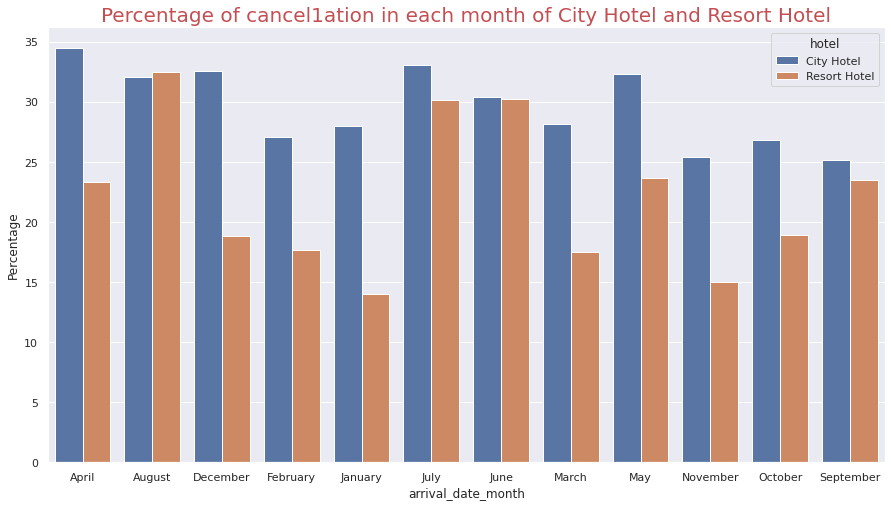
From the above plots we can see the bookings happening in each month of years 2015, 2016 and 2017. Weekend, week and total nights booking is hotels were displayed

1. **Booking Cancelation Percentage for City Hotel and Resort Hotel**

****

Resort hotel has less percentage of cancelation than City Hotel

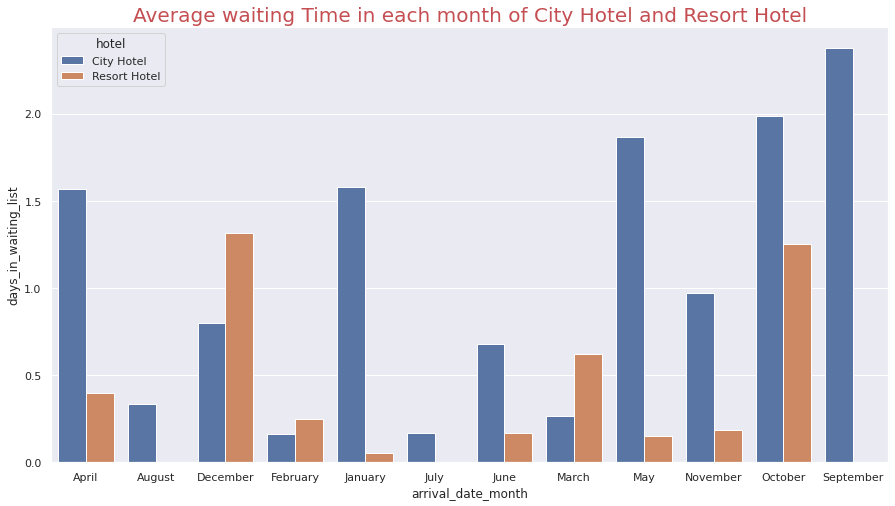
1. **Comparison between cancelation percentage for both Hotel Every month**

****

In the month of January the cancel percentage is very less for Resort Hotel group and in the month of November the cancel percentage is very less for City Hotel.

1. **Average waiting time in Each Month for City Hotel and Resort Hotel:**

The average waiting time in the month of January is very less for Resort Hotel and the average waiting time in the month of February is very less for City Hotel



**So the month January is the best month of year to book a Resort hotel room and the month February is the best time to book City hotel as there is less waiting list**

1. **Number of Special request received by City Hotel and Resort Hotel:**

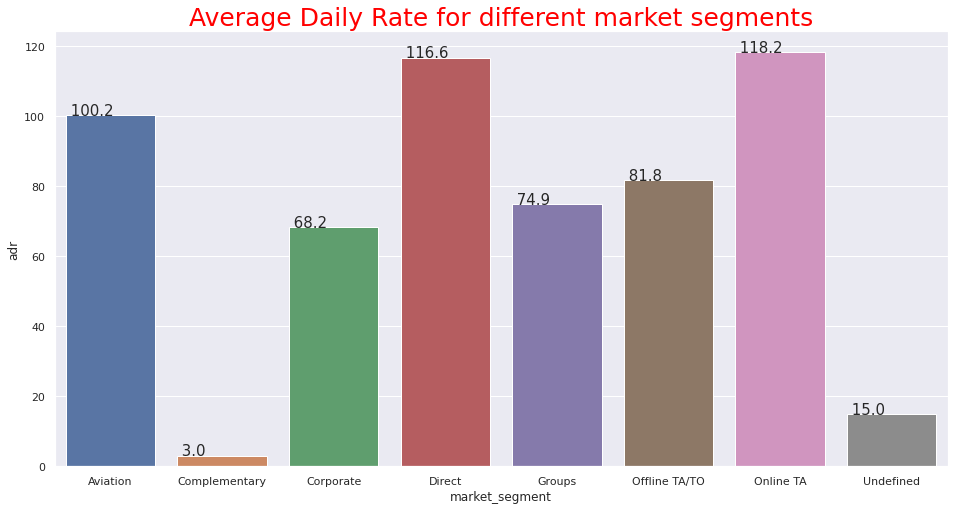
****

Probability of receiving special request is high for both the hotels but the City hotel likely to receive a high number of special requests as its percentage of special requests is 71

1. **Optimal length of stay in order to get the best daily rate:**

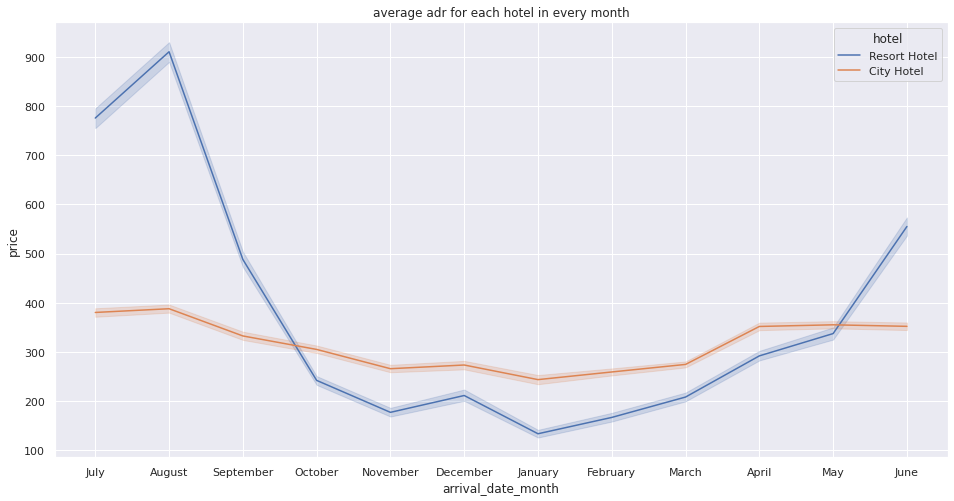
To calculate optimal length of stay we have to consider several factors  
Average daily rate for different market segments  
Price per month of a Hotel  
Price per day for single adults  
Price per day for two adults

1. **Daily rate for different Market segment:**

****

We can see that the average adr is very less in case of Complementary and high for Direct and Online TA market segment.

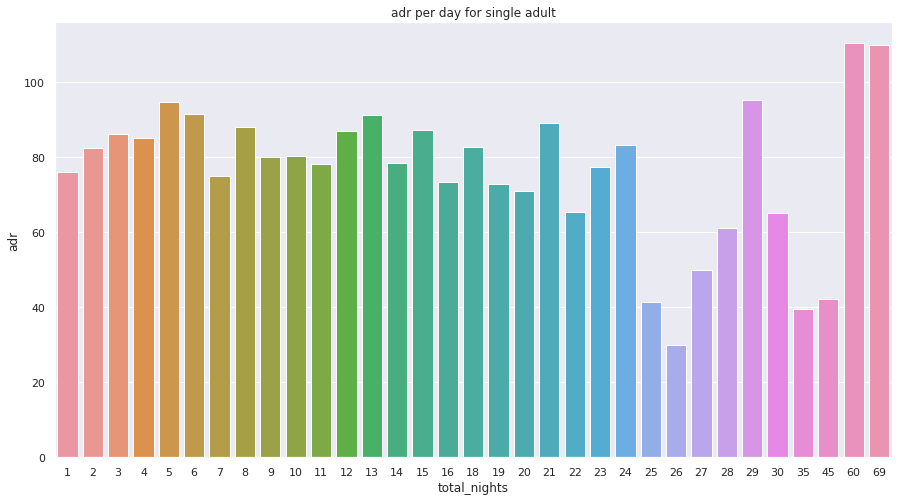
1. **Daily rate for different Months of the year:**

****

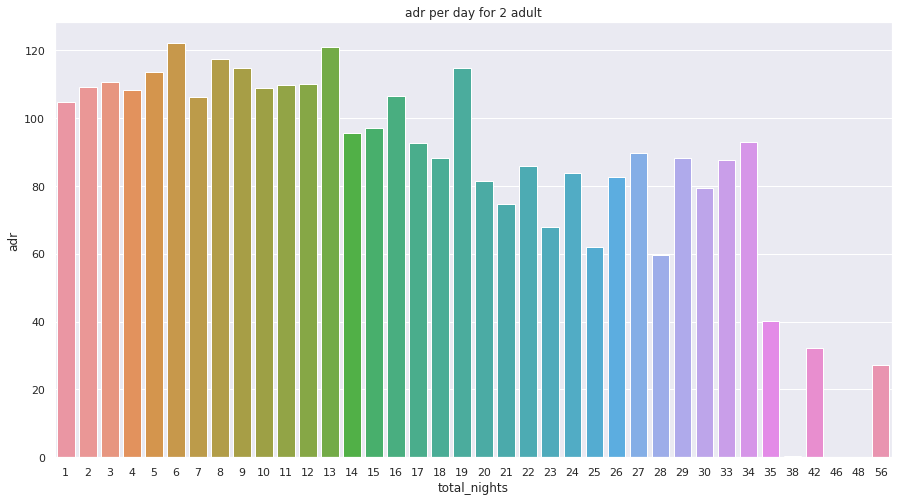
In between November to Jan ADR is very low for both Hotel. From November to January there will be holidays in European countries

Due to Holidays hotel may provide higher discount to attract more guests.

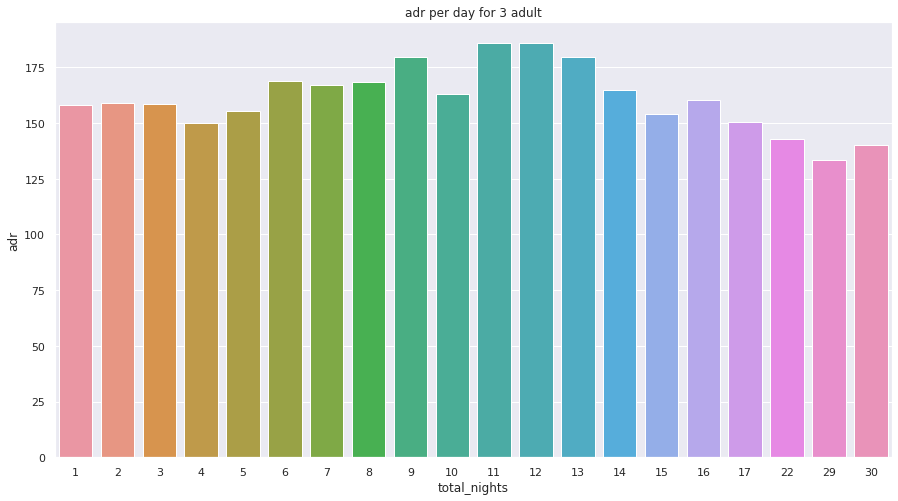
1. **Daily rate As per number of Adults staying :**

****

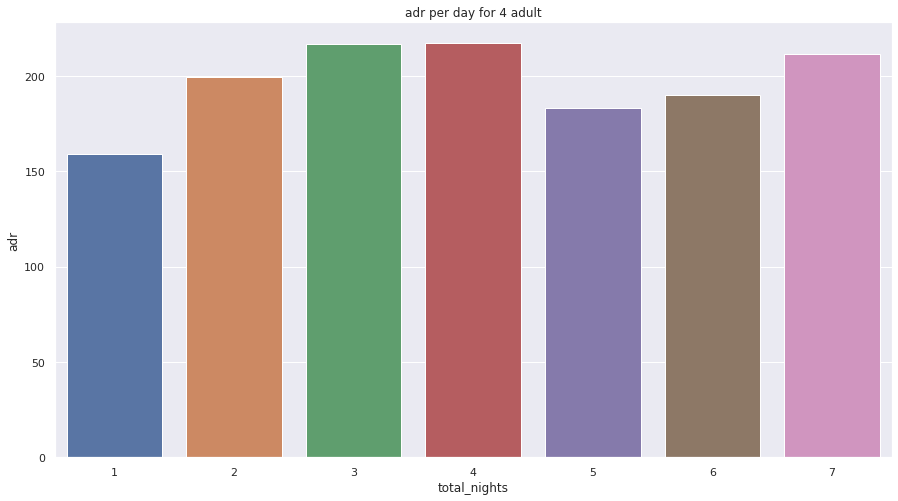
This the Average Daily Rate with only single adult staying



This the Average Daily Rate with two adult staying.



This the Average Daily Rate with three adult staying

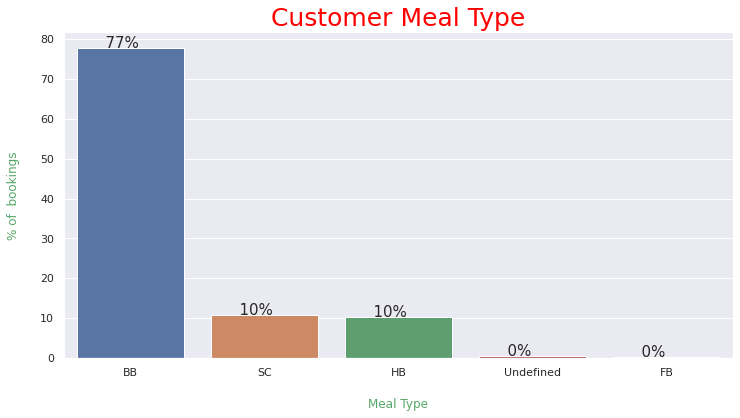


This the Average Daily Rate with four adult staying

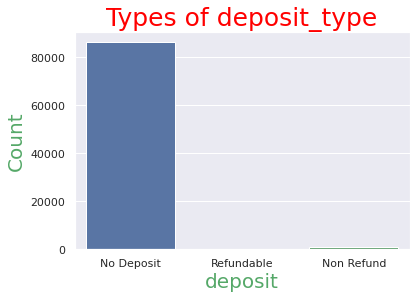
As the number of adults changes the average adr changes

1. **Meals Type:**

* Type of meal booked. Undefined/SC – no meal package
* BB – Bed & Breakfast
* HB – Half board (breakfast and one other meal – usually dinner)
* FB – Full board (breakfast, lunch and dinner)
* The Bread & Breakfast option is the most popular, with a frequency of 77.84%.

****

1. **Deposit Types:**

****There are elements available under deposit column:-

No deposit,

Refundable,

Not refundable

Here in graph we can see that bar

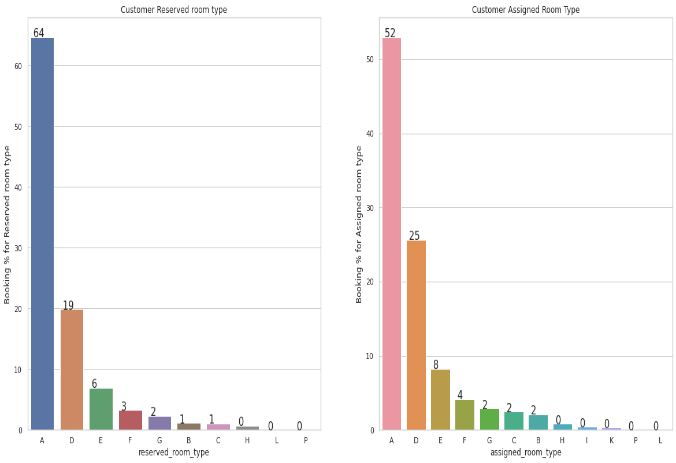
reflecting no deposit is quit high. We can

conclude that majority of the booking

does not require deposit. This could be a

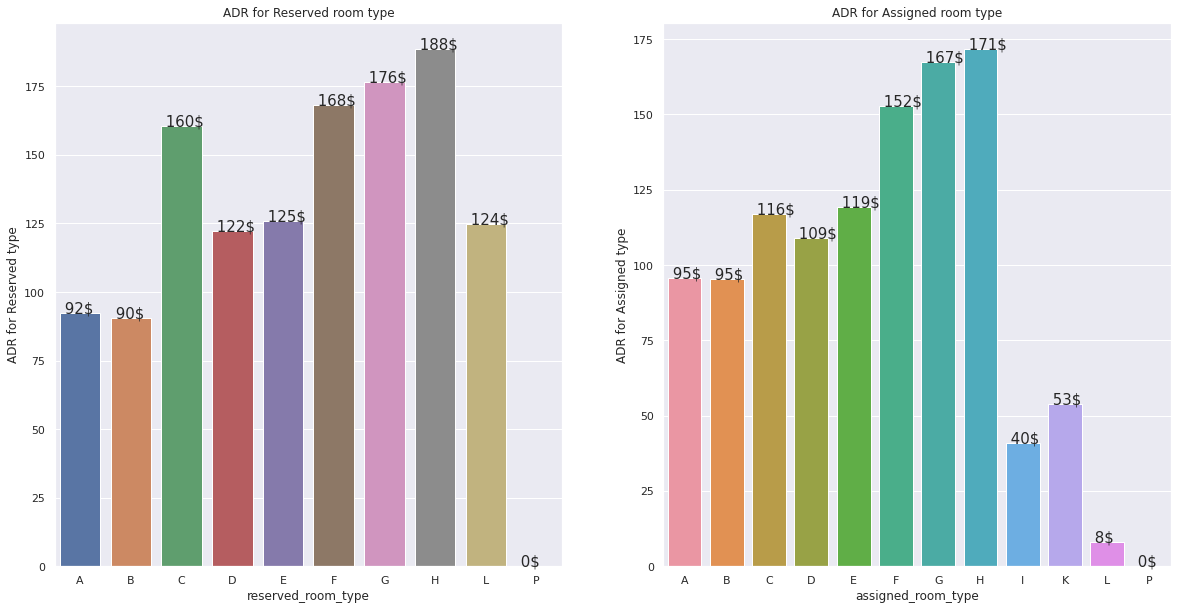
reason why the cancelation rate is high**.**

1. **Reserved room type & Assigned room type bookings:**

****

The Types of Reserved rooms are A, B, C, D, E, F, G, H, L. Room Type A & D seems very popular among the customers. The maximum bookings of around 66% are reserved for the room type A, followed by at 19% reserved for the room type D. The Types of Assigned rooms are A, B, C, D, E, F, G, H, I, K. Maximum bookings are assigned for the room type A i.e around 51% out of 66% of the reserved room type for A.

1. **Reserved room type & Assigned room type bookings with adr:**



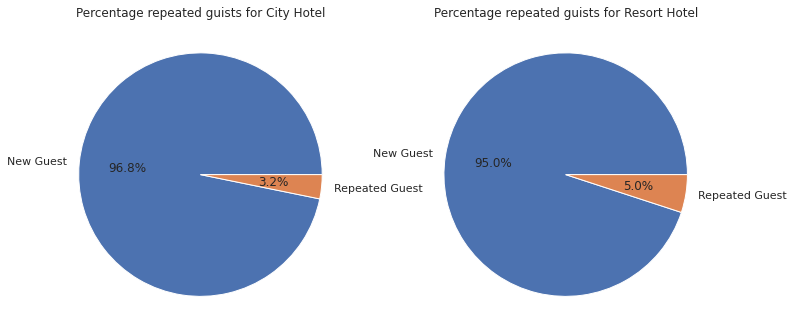
ADR is highest for Reserved room type H followed by Reserved room type G

ADR for room type A & B are the lowest.

As Couple (or 2 adults) is the most popular accommodation type with about 65%, so my guess is Room type A & D can be Single room & Double room because these are most booked room types in hotel industry.

Some room types determined by bed availability as Queen-sized bed, King-sized bed, a room with two twin-sized beds. Also the adr corresponding to the room type A & D are among the lowest.

1. **Overview of repeated guests:**

****

The percentage of repeated guests are low for both the hotels but the percentage of repeated guests are little for Resort Hotel.

Both hotels should work on that to increase the repeated guests percentage.

1. **Looking at types of guests:**

Type of booking, assuming one of four categories:

Contract — when the booking has an allotment or other type of contract associated to it

Group — when the booking is associated to a group

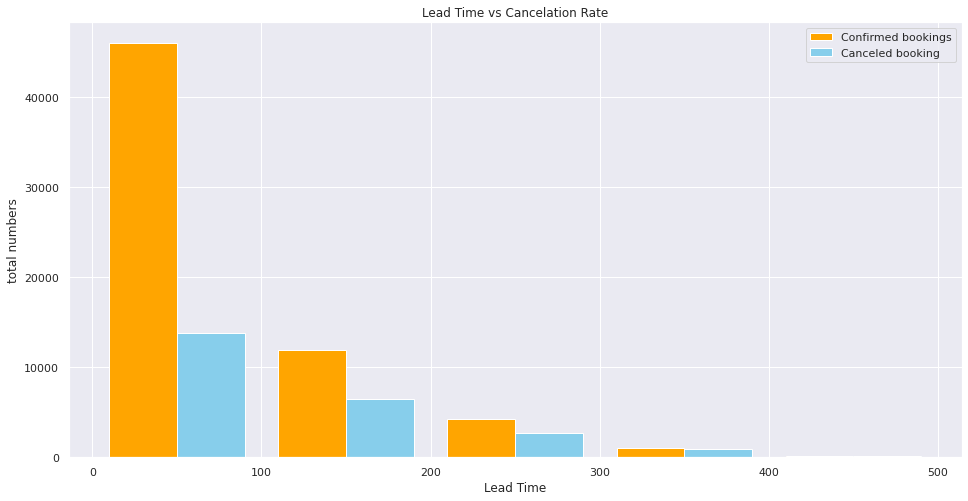
Transient — when the booking is not part of a group or contract, and is not associated to other transient booking

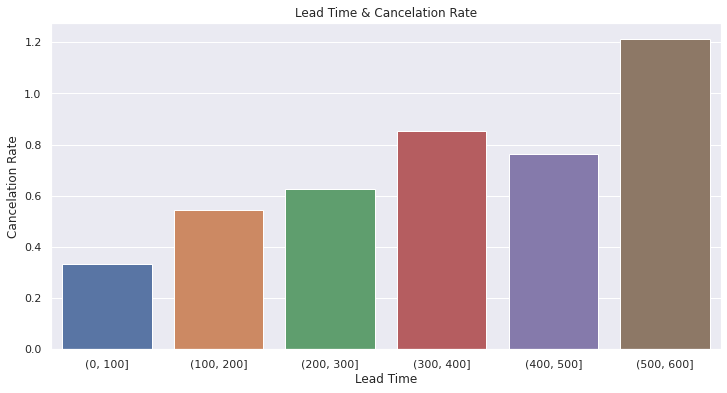
Transient-party — when the booking is transient, but is associated to at least other transient booking

Majority of the bookings are transient. This means that the booking is not part of a group or contract. With the ease of booking directly from the website, most people tend to skip the middleman to ensure quick response from their booking. Transients are the most common customer type, they represent 79.40% of the total customers.

****

1. **Relation between Lead Time & Cancelation Rate:**

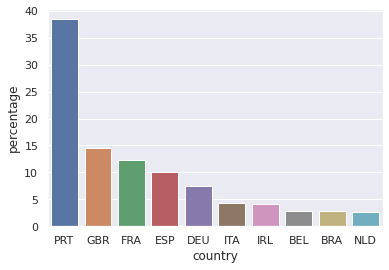
****

****

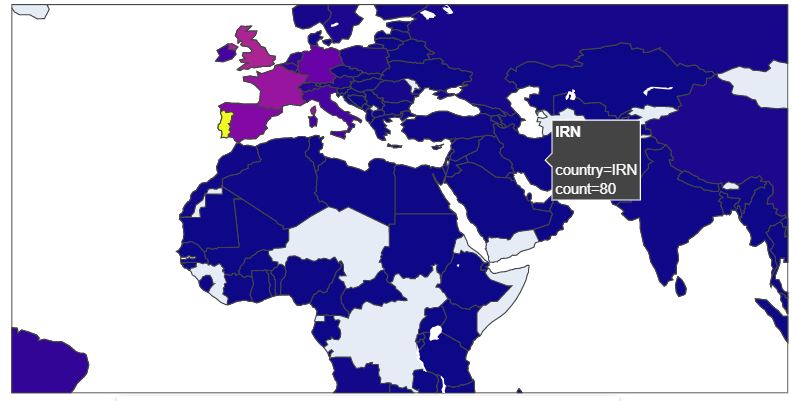
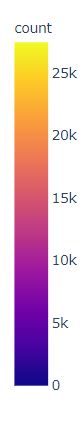
If the lead time in between 0 to 100 the cancellation rate is low and when lead time in between 500 to 600 the cancellation rate is high

We can clearly make the inference that Cancellation rate is directly proportional to the Lead time i.e as the Lead time increases the cancellation rate also increases.

1. **Visitors origin:**

****

The most of the hotel booking where made by country PRT customer.

Most of bookings is done by European countries. In the map we can see the number of booking done by those countries

**8. Conclusion:**

1. The majority of guests come from western European countries.
2. Majority of the hotels booked are city hotel.
3. The number of repeated guests is too low.
4. City hotel has highest percentage of cancellation rate.
5. We also realise that the high rate of cancellations can be due high no deposit policies.
6. The month of January and February has less waiting time compared to other months.
7. Both the hotels has high percentage of special request but the City hotel likely to receive a high number of special requests as its percentage of special requests is 71.
8. The optimal length of stay in order to get the best daily rate varies as the number of adults varies

**References-**

1. AlmaBetter
2. GeeksforGeeks
3. Stack Overflow