

# Capstone Project

## Hotel Booking Analysis

By: Vikash Kumar

# Discussion points

To discuss the analysis of the data set on Hotel Booking. We will perform the following analysis on the provided data set:

1. Univariate analysis
2. Hotel-wise analysis
3. Distribution Channel wise analysis
4. Booking cancelation analysis
5. Time-wise analysis
6. Conclusion

By doing this, we'll try to identify the main elements influencing the trends in hotel bookings.

# Data Cleaning

Prior to EDA, cleaning the data is essential since it will get rid of any ambiguous information that can have an impact on the results.

**We will take the following actions while cleaning the data:**

1. Remove duplicate rows.
2. Handling missing values.
3. Convert columns to appropriate data types.
4. Adding important columns.

# Data Summary:

Given data set has different columns of variables crucial for hotel bookings. Some of them are:

1. hotel: The category of hotels, which are two resort hotels and city hotels
2. is\_cancelled: The value of the column shows the cancellation type. If the booking was canceled or not. Values[0,1], where 0 indicates not canceled and 1 indicates canceled.
3. lead\_time: The time between reservation and actual arrival.
4. stayed\_in\_weekend\_nights: The number of weekend nights stay per reservation
5. stayed\_in\_weekday\_nights: The number of weekday nights stays per reservation.
6. meal: Meal preferences per reservation.[BB, FB, HB, SC, Undefined]

7. Country: The origin country of the guest.

8. market\_segment: This column shows how the reservation was made and what is the purpose of the reservation. Eg, corporate means corporate trip.

9. distribution\_channel: The medium through booking was made.

[Direct, Corporate, TA(travel agent)/TO(Tour Operator), undefined, GDS.]

10. Is\_repeated\_guest: Shows if the guest is who has arrived earlier or not. Values[0,1] 0 indicates no and 1 indicated yes person is the repeated guest.

11. days\_in\_waiting\_list: Number of days between actual booking and transaction.

12. customer\_type: Type of customers( Transient, group, etc.)

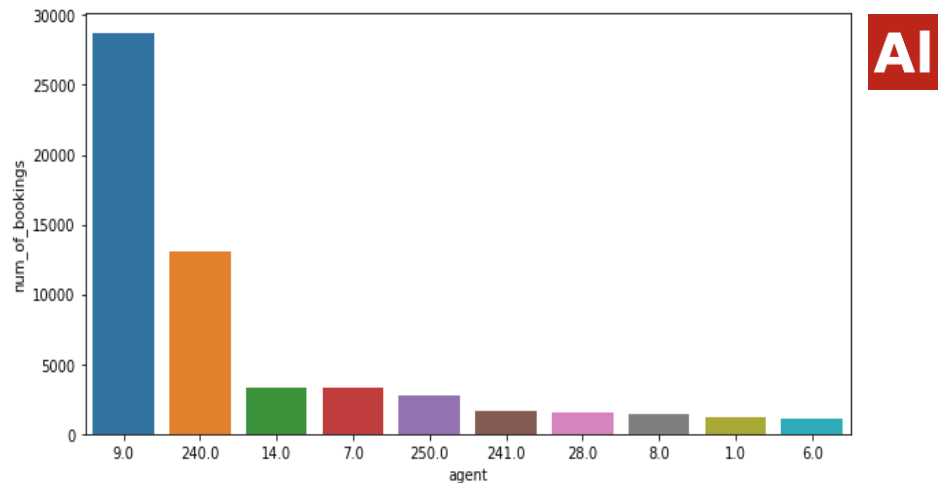
# Univariate Analysis:

We conducted a univariate analysis on a dataset of hotel reservations and answered the following questions:

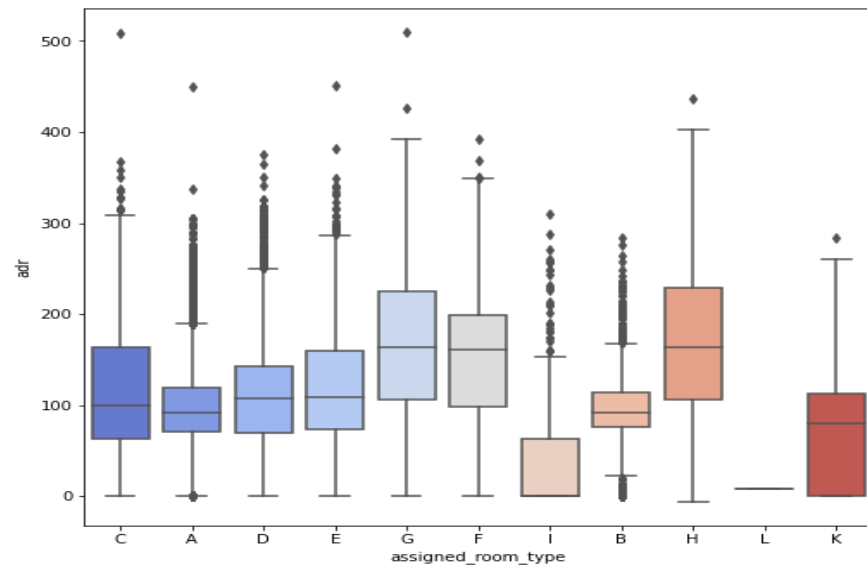
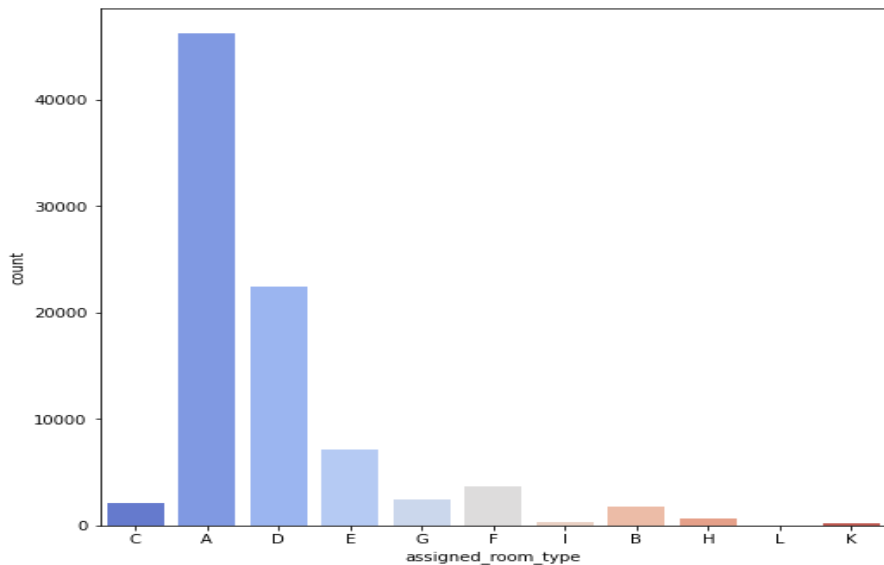
1. Which room type is in most demand and which room type generates the highest adr?
2. What is the most preferred meal by customers?
3. Which agent made most of the bookings?
4. From which country most of the customers are coming?

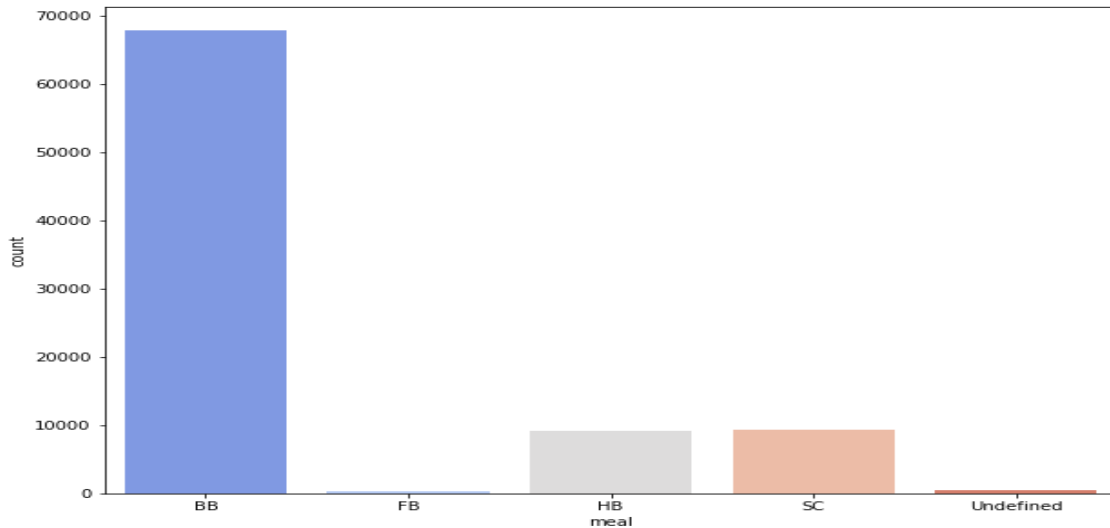
1. Most of the appointments were made by the agent with id number 9.

2. The most popular room type is A, but you may also find better ADR (average daily rate) rooms in kinds H, G, and C. Hotels should provide more A and H room types to increase revenue.



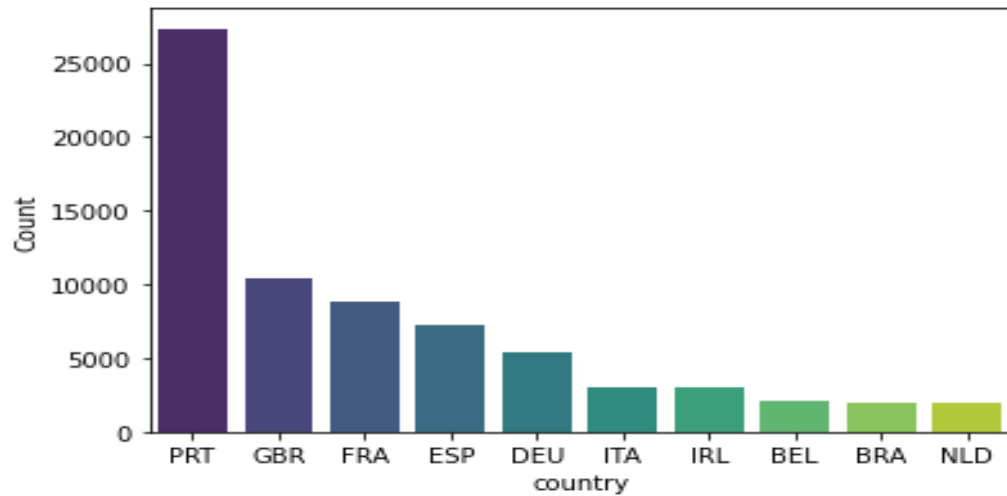
AI



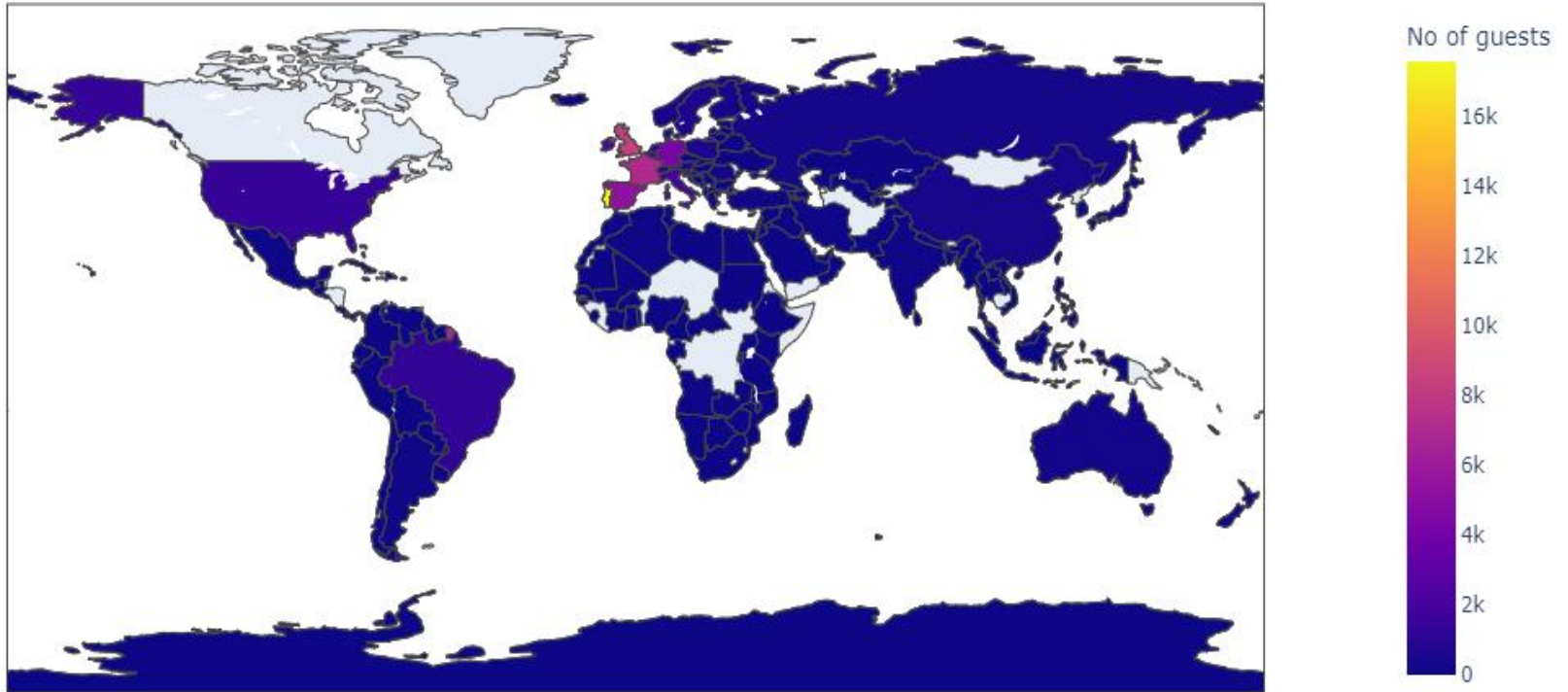


BB is the most popular sort of meal (Bed and breakfast).

The majority of visitors are from Portugal and other European nations.





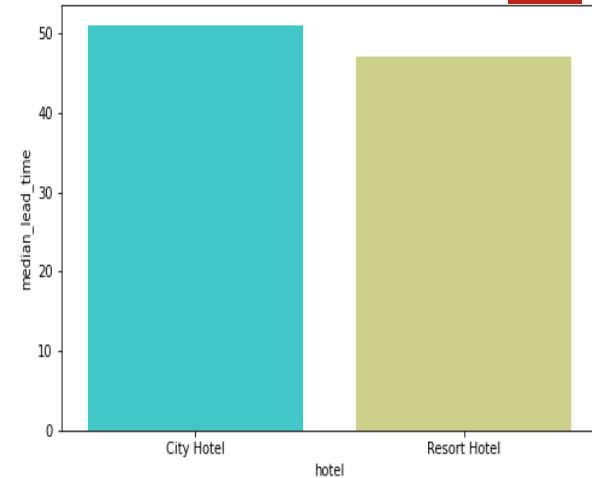
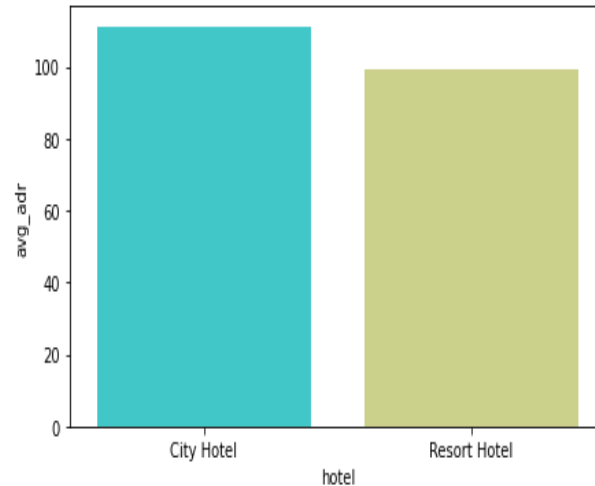
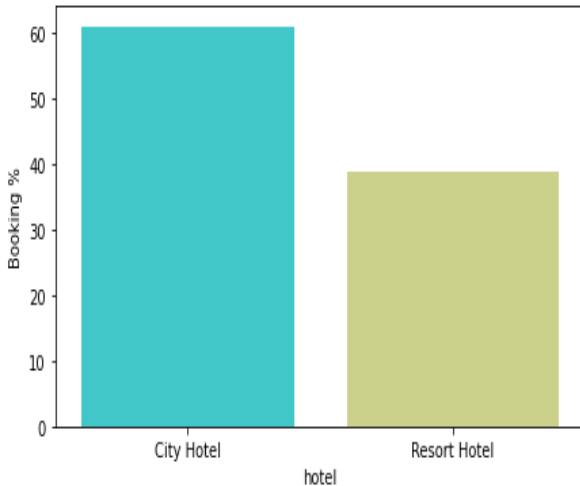


The majority of the consumers are from European nations like Portugal, Spain, France, and the United Kingdom.

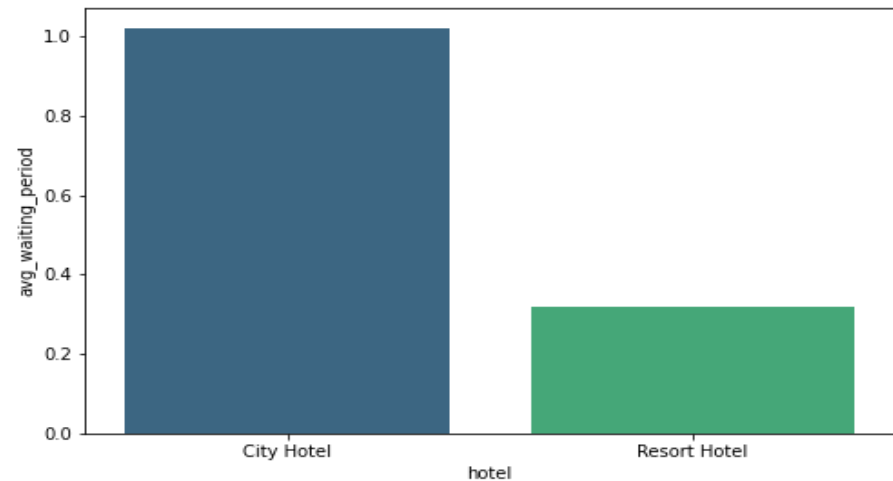
# Hotel wise Analysis

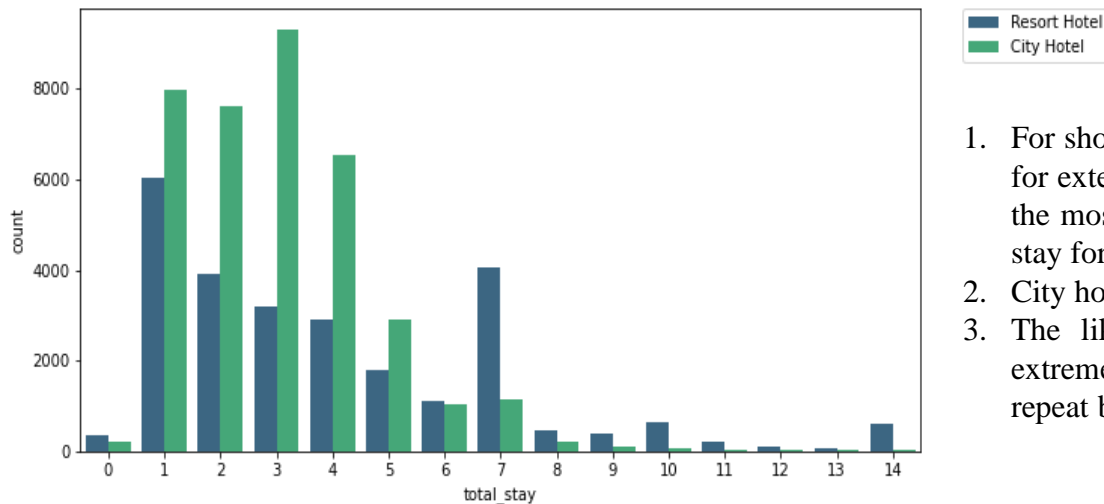
The following questions were addressed through a hotel-wise examination of the provided hotel booking dataset:

1. What is the percentage of bookings in each hotel?
2. Find the higher lead time hotel?
3. Most preferred stay length in each hotel?
4. Which hotel makes more revenue?
5. Which hotel has a higher and how much customer return rate?
6. Find the higher booking cancellations rate?
7. Which hotel, do people have to wait longer to get a booking confirmed?

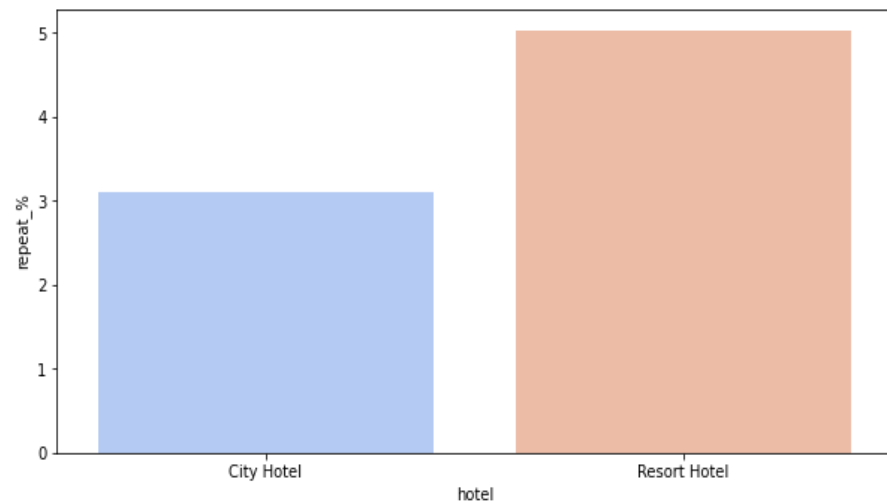
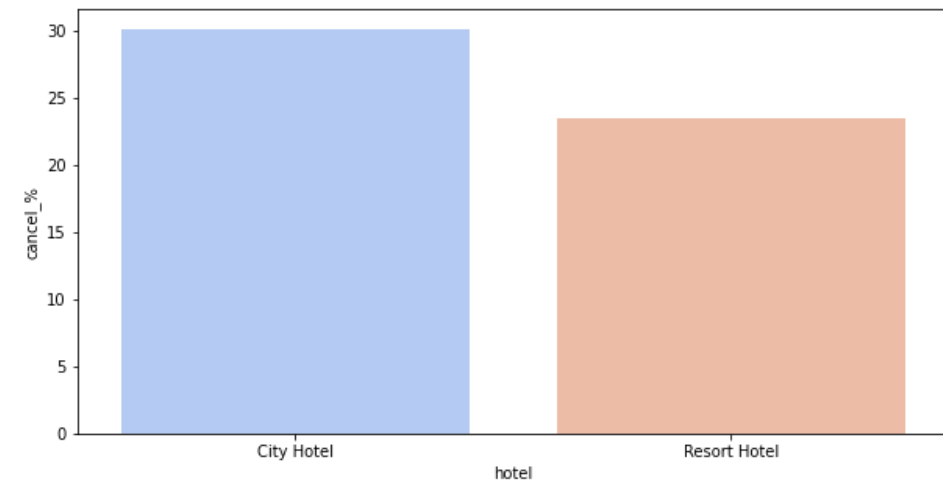


1. 40% of reservations are for resort hotels, while 60% are for city hotels.
2. The average ADR at Resort hotels is a little less than at City hotels. Therefore, it appears that the City hotel is making a little more revenue.
3. A little longer median lead time is found in city hotels. Additionally, the median lead time is much longer in each scenario, indicating that customers typically book their hotel stays far in advance.
4. The wait time at a city hotel is much longer than it is at a resort. As a result, City Hotel is considerably busier than Resort Hotel.





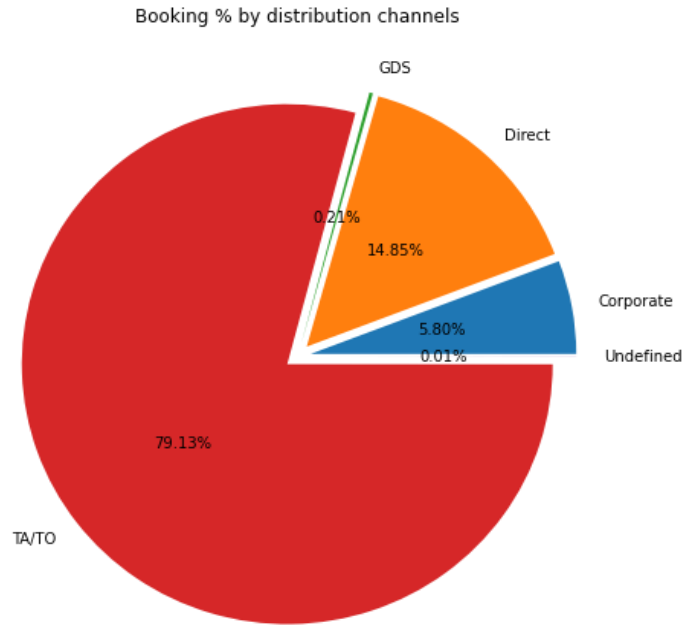
1. For short stays, city hotels are preferred, while resorts are preferred for extended stays. For City Hotels, stays of less than four days are the most typical. Less than 7 days are the most typical duration of stay for resorts.
2. City hotels have a noticeably greater rate of canceled reservations.
3. The likelihood that a customer will return to either hotel is extremely low, although Resort Hotel has a slightly greater rate of repeat business than the City Hotel.



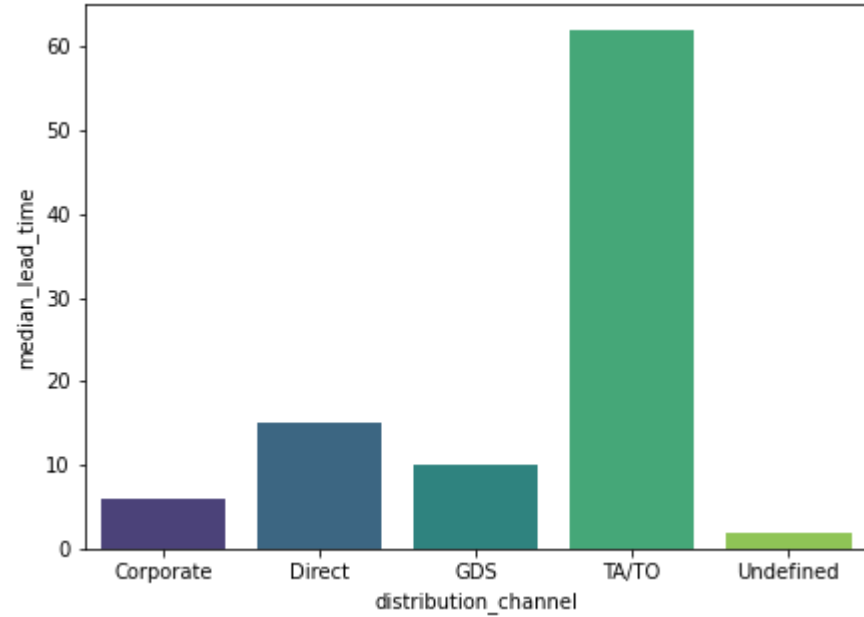
# Distribution channel-wise Analysis

We responded to the following queries while performing a distribution channel-wise analysis of the provided hotel booking dataset:

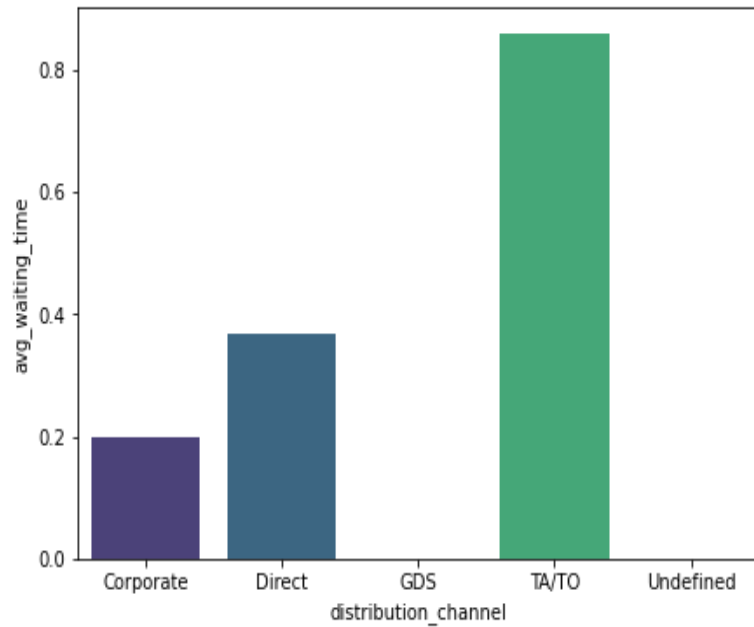
1. What is the most common method of booking hotels?
2. What is the most popular channel for early booking of hotels?
3. Which distribution channel brings better revenue-generating deals for hotels?
4. Which channel has a longer average waiting time?



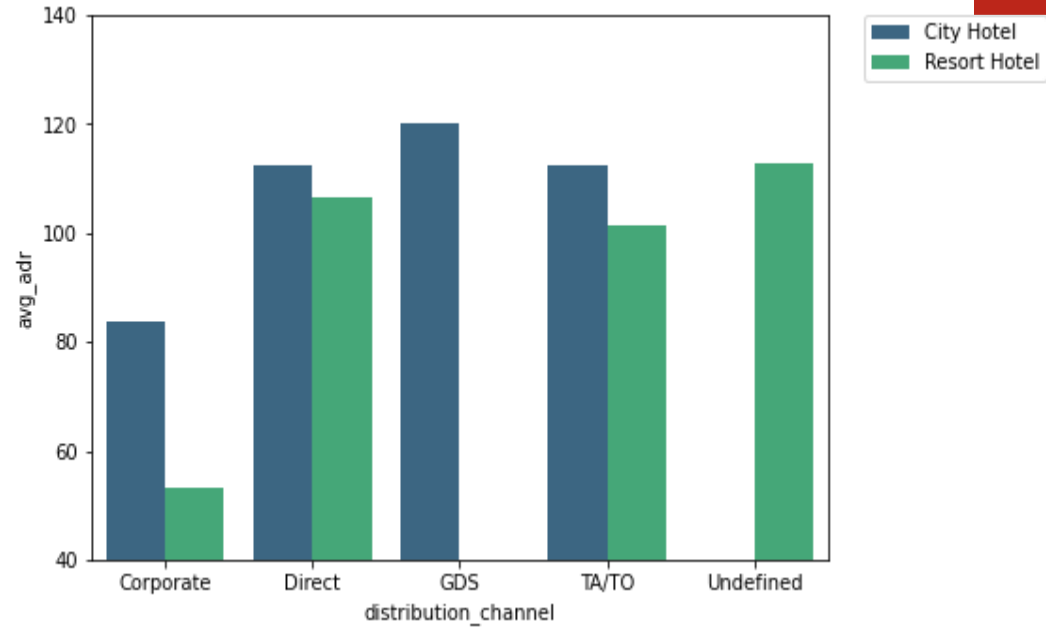
The most popular channel for booking hotels is TA/TO.



TA/TO is most typically used to plan hotel visits in advance. Alternative media, however, are typically appropriate for unannounced guests.



The time it takes to get a reservation confirmation when using TA/TO could be a little longer.



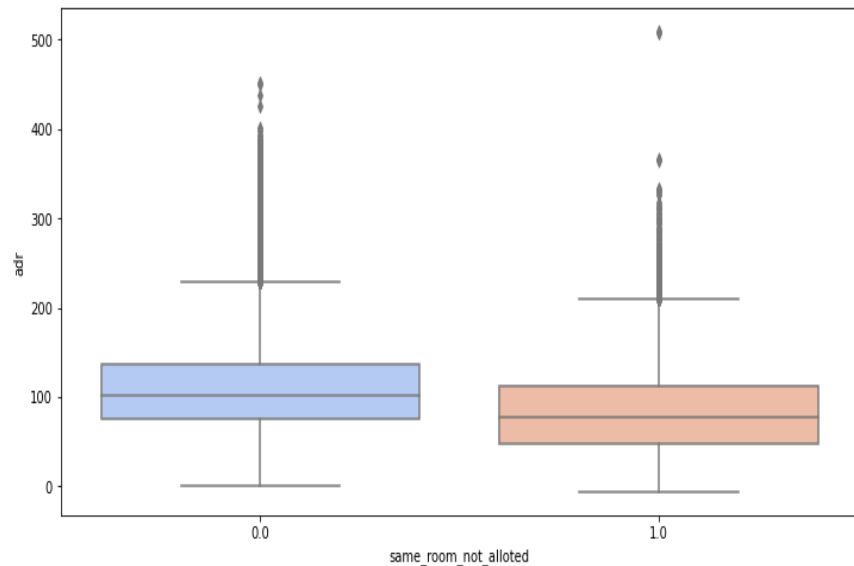
GDS channel enables higher revenue-generating transactions for City hotels in comparison to the majority of bookings, which come via TA/TO. City Hotel may aim to increase its presence on GDS channels in an effort to establish more partnerships that will result in increased revenue.

Direct channel partnerships and TA/TO channel partnerships increase income for resort hotels. Resort hotels must broaden their GDS channel reach in order to increase sales.

# Booking cancellation Analysis

1. Which significant distribution channel has the highest cancellation percentage?
2. effects of lengthier lead times or waiting periods.
3. effects not receiving the requested room type.
4. getting a different room influences the ADR.

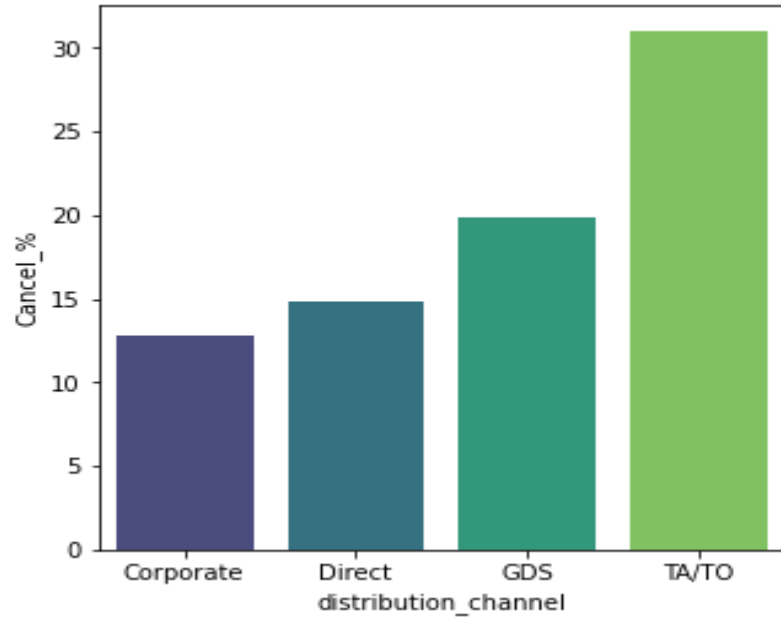
**Getting a different room influences the ADR.**



Not getting the same room does have an impact on the ADR; with a few exceptions, those who didn't get the same room paid a little lower ADR.

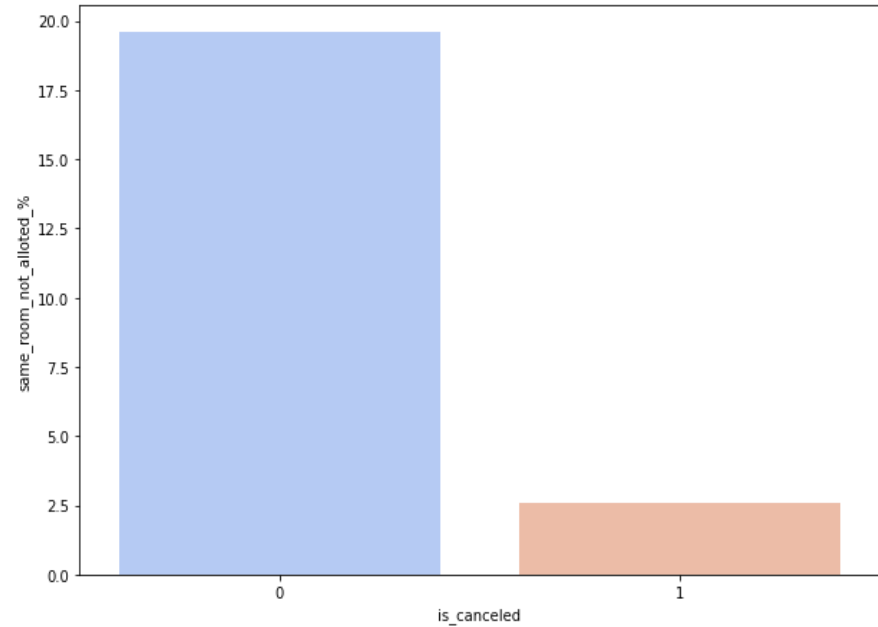


## Distribution channel cancellation percentage



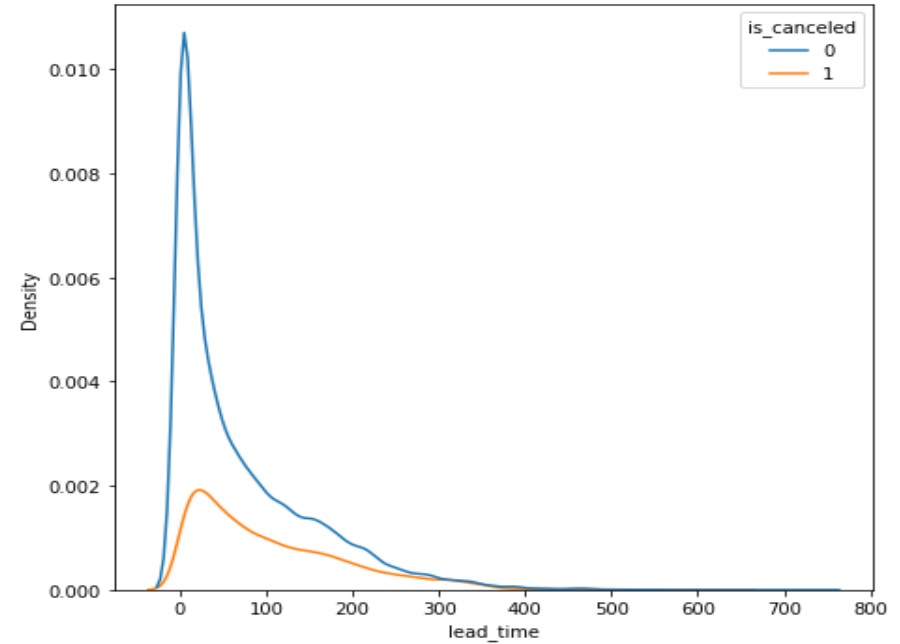
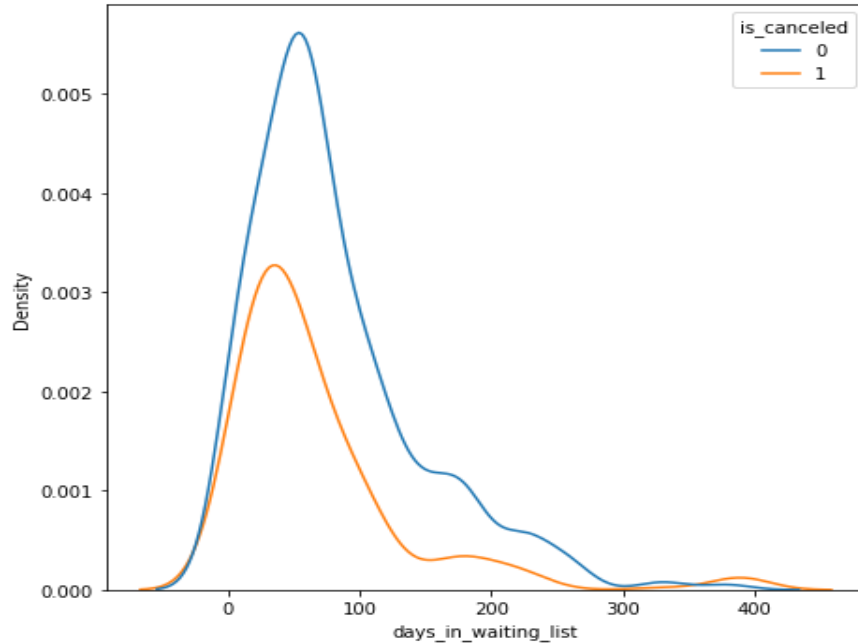
In TA/TO, there are the most canceled reservations. Thus, there is a 30% chance that a reservation booked through TA/TO would be canceled.

## Effects of not receiving the requested room type



As you can see, reservations were not canceled because they did not receive the exact accommodation they requested. Even when requested rooms are not obtained, a sizable portion of reservations is not canceled.

## Effects of lengthier lead times or waiting periods

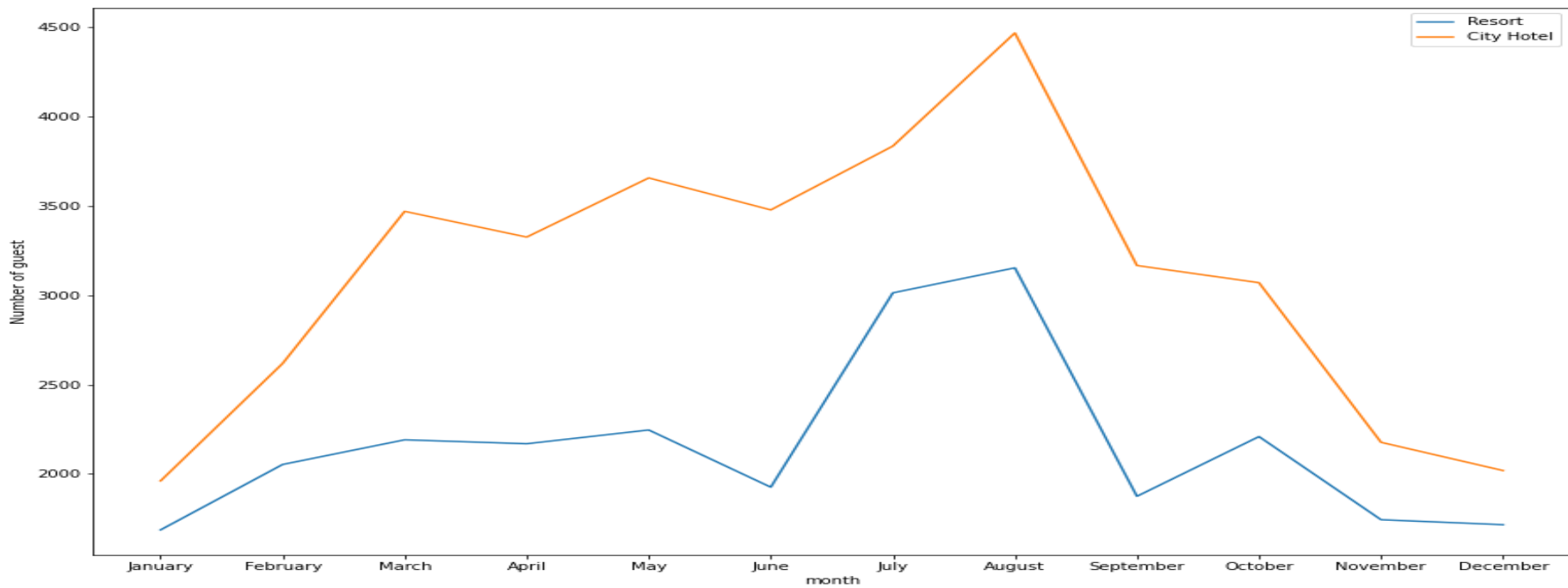


The number of canceled reservations closely relates to the number of reservations. Meaning that as reservations increase, cancellations increase as well. While it is evident that the majority of canceled reservations have waiting periods under 150 days, the majority of confirmed reservations also do. This demonstrates that the waiting period has no influence on reservation cancellations. Additionally, because the curves for cancellation and non-cancellation are comparable for a lead time as well, lead time has no impact on the cancellation of appointments.

# Time-wise Analysis

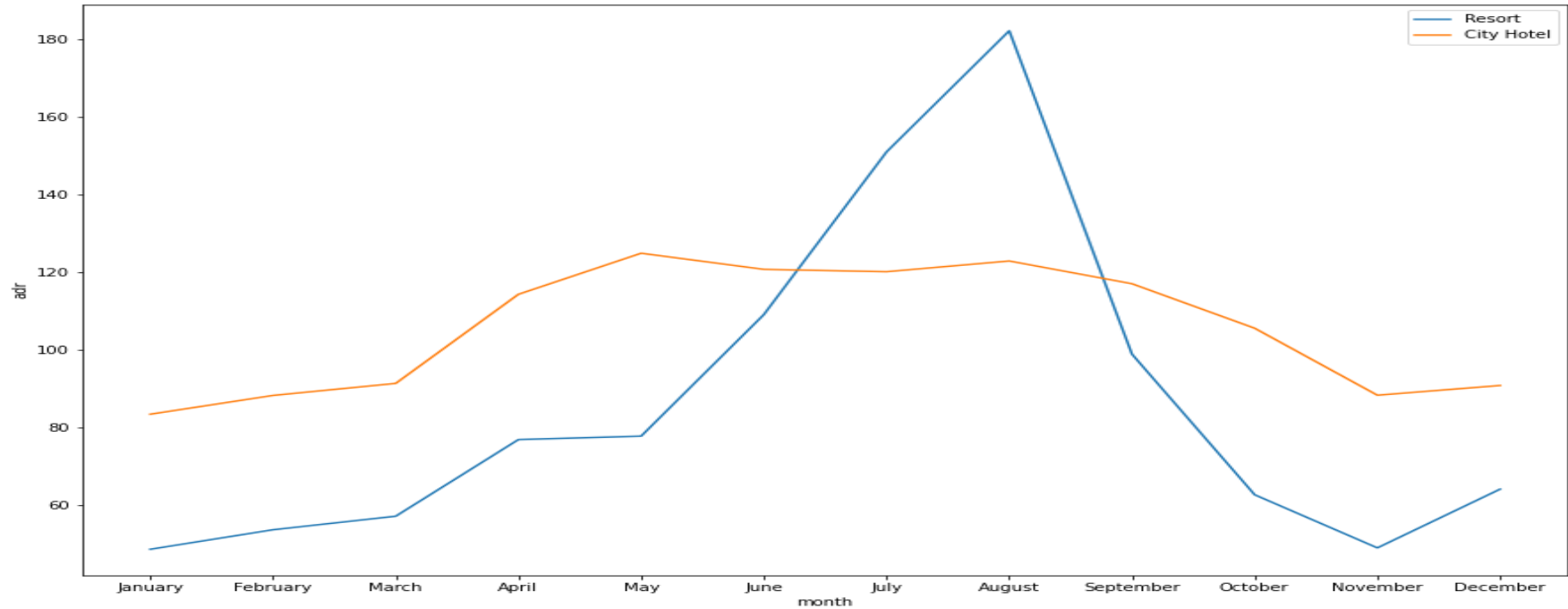
While doing a time-wise analysis of a given hotel booking dataset, we answered the following questions:

1. What are the busiest months for hotels?
2. In which months do hotels charge higher ADR?
3. How do booking numbers and ADR change within a month?

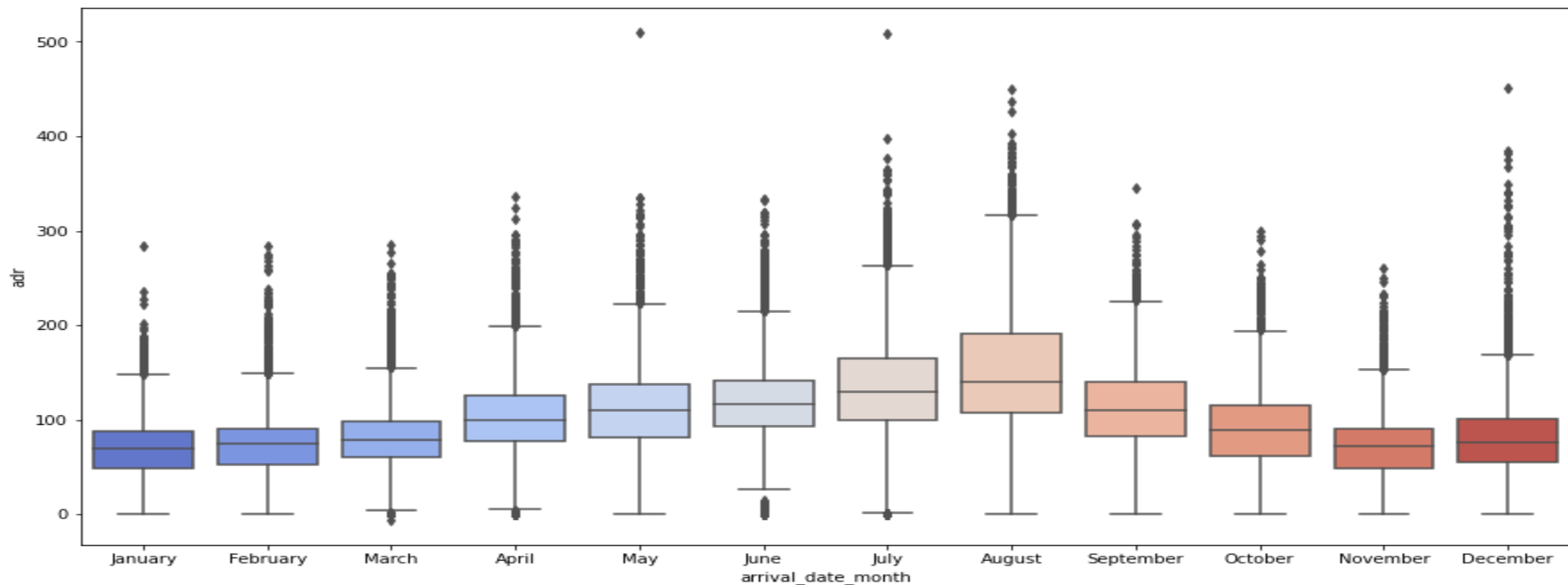


August is the month and has the most visitors.

## Month with the highest revenue

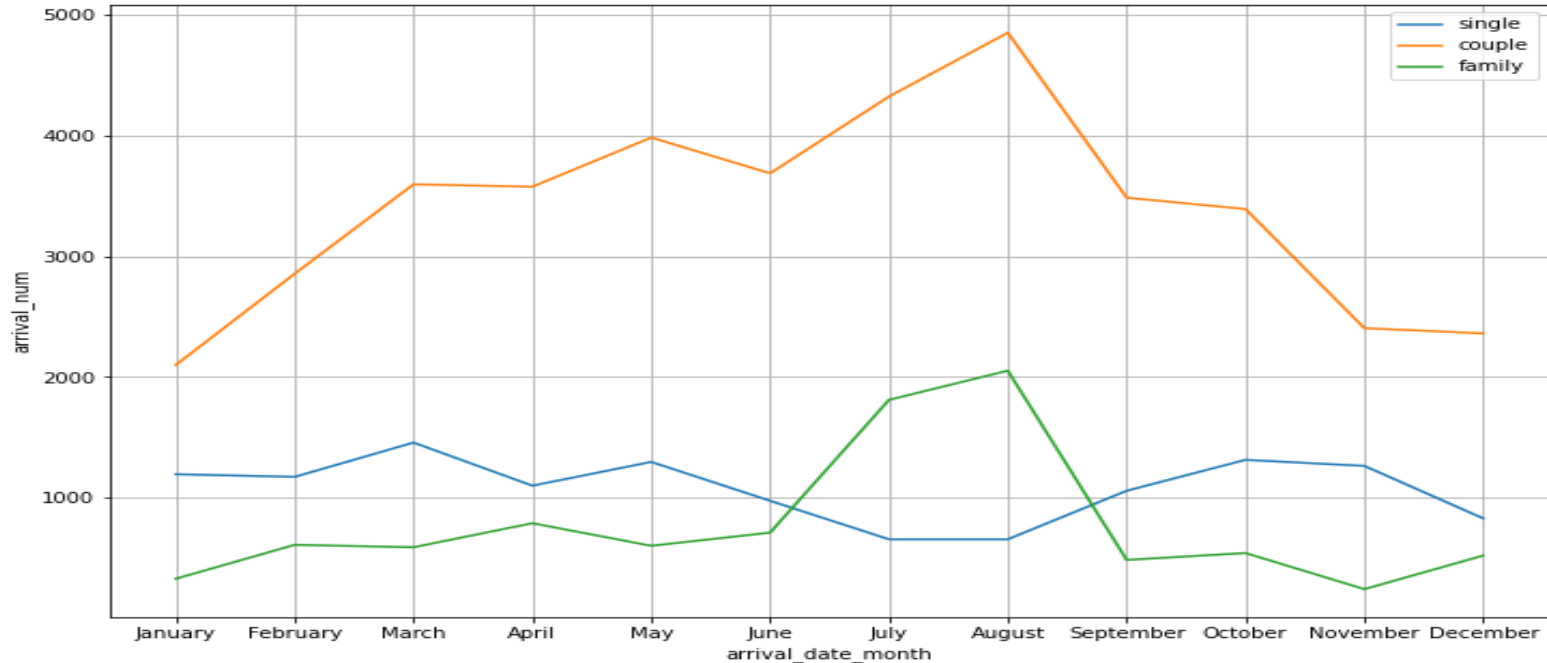


The revenue picture is different; compared to city hotels, resort hotels bring in more money. There was a sharp rise in adr from May to August. The largest number was in August.

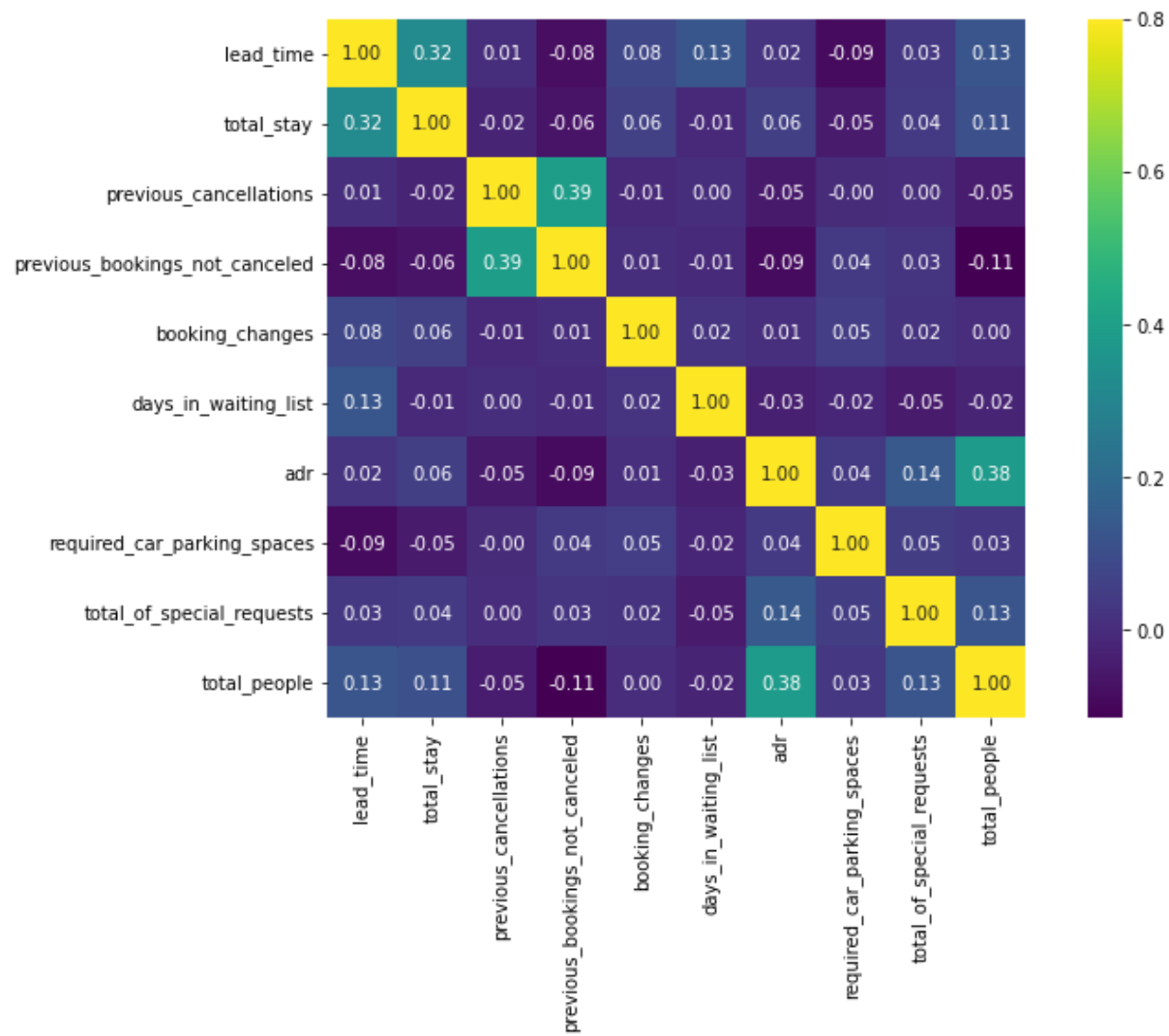


The average ADR climbs from the start of the year through the middle, peaks in August, and then declines through the end of the year. But at the end of the year, hotels do strike some advantageous arrangements with high ADR.

## Reservation percentage according to gender



Most reservations are made by couples. The graph clearly shows that the number of couples and families arriving in the months of July and August suddenly increases. Therefore, better preparations for these types of clients can be made accordingly at that moment.



## Correlation Heatmap

There is a weak link between total stay length and lead time. This may imply that people tend to make few plans before their actual arrival for lengthier hotel stays.

The fact that ADR and total\_people have a slight correlation makes sense given that a higher population number generates higher revenue and, consequently, higher ADR.



# Conclusion:

That's all there is to it! We have reached the end of our exercise. The data has been loaded, null values have been treated, categorical columns encoded, and major reasons that govern hotel bookings have been identified, along with steps to increase them.

1. About 66% of the reservations are for the City Hotel and 34% of the bookings are for the Resort Hotel, so City Hotel is busier than the Resort Hotel. So does the overall ADR.
2. The majority of guests stay less than 7 days at the hotel and for longer stays Resort Hotel is preferred.
3. Both hotels have significantly higher booking cancellation rates and very few guests less than 3% return for any other booking at the City Hotel. 5% of visitors stay at the Resort hotel again.
4. The majority of the visitors were from European nations, with the majority being from Portugal.
5. Customers book through many channels, with TA/TO being the most popular option.
6. Hotel should improve their appeal on the GDS channel because here is where they can find higher adr deals.

7. Nearly 30% of reservations made via TA/TO are canceled.
8. Booking cancellations are unaffected by not receiving the exact accommodation that was reserved, lengthier lead times, and waiting times. Although ADR is reduced by varied room allocation.
9. The busiest and most lucrative months for both hotels are July and August.
10. A month's worth of ADR steadily rises toward the end of the month, with brief spikes on weekends.
11. Since couples are the most frequent hotel customers, hotels can design services to meet their needs in order to boost income.
12. There are more special requests when there are more guests present.
13. Adults and adjacent market segment bookings typically have a significant number of customized requests.
14. Customers can typically find better deals with low ADR for longer stays (more than 15 days).

**Thank You**