Capstone Project

Hotel Booking Analysis

By Ayush Kumar

Points for Discussion

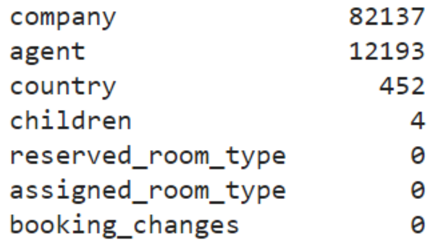
- Data Summary
- Data Preparation & Cleaning
- Hotel Types
- Cancellations
- Cancellations by Deposit Type
- Lead Time & Cancellations
- Bookings & Cancellations by Months
- Avg. Daily Rate over Months
- Bookings by Country
- Bookings & Cancellations by Market Segments
- Correlation of features
- Distribution of ADR
- Car Parking Space
- Room Types
- Top Agents
- Special Request by Country
- Lead Time
- Bookings by Distribution Channel
- Meals
- Previous cancellations & Cancellations
- Repeated Guests
- Conclusions

Data Summary

The **Hotel Booking Dataset** contains booking information for city and resort type hotels mainly in the years 2015, 2016 and 2017. It includes information like when the booking was made, length of stay, the number of adults, children and babies, and the number of available parking spaces, meals etc. All personally identifying information has been removed from the data. This dataset contains more than 119,000 rows of data.

Data Preparation & Cleaning

- There were 31994 duplicate data, so deleted those rows.
- 4 columns (*company*, *agent*, *country* & *children*) had missing values.
- There may be some cases when customer didn't booked hotel via any agent or via any company. Hence we will replace null values by 0 in these columns.



```
company                82137
agent                  12193
country                  452
children                   4
reserved_room_type         0
assigned_room_type         0
booking_changes            0
```

- children column is numeric and skewed, hence choose median for imputing missing values.
- country column is categorical column. We use mode of *‘country’* column, But it can lead to bias towards a specific country that occurs most frequently in the data. Hence I created a new category *‘others’* for missing values.

Data Preparation & Cleaning

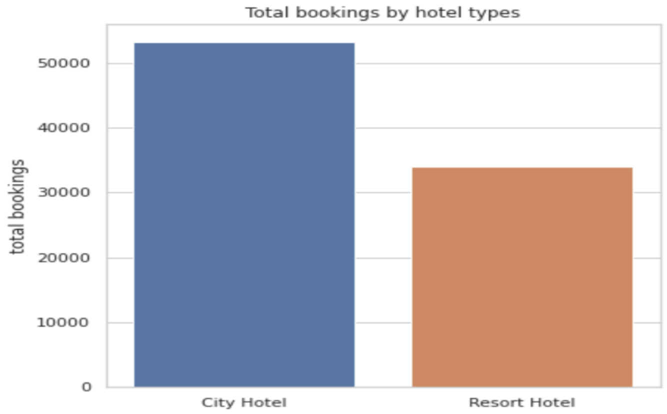
- There were 166 rows with 0 adults, children and babies which seems unlikely hence dropped them.
- Created new column for *total_guest* = *(adults* + *children)*. Ignored babies because generally they are not charged.
- Some values in ‘adr’ was negative, which must be an error so imputed them with median since it was a skewed data.
- Created new column for *total_stays_night* = *(weekend_nights* + *week_nights)* to analyze average length of stay.
- Assigned appropriate data types for some columns.

Hotel Types

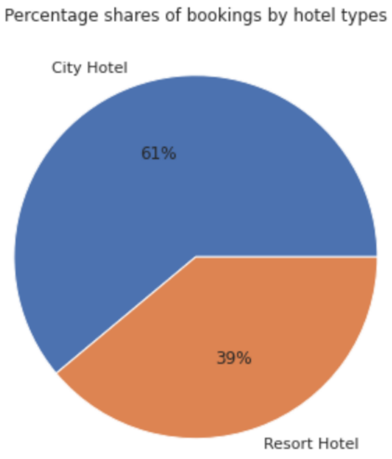
There are two types of hotels:

- City Hotel
- Resort Hotel

There are around 52k bookings in City Hotels which is 61% of total bookings in the dataset, similarly around 34k bookings (39%) in Resort Hotels.



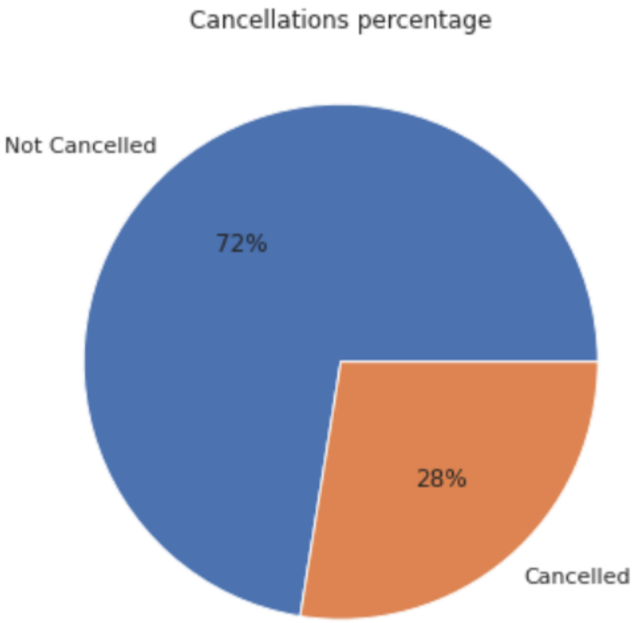
Total bookings by hotel types



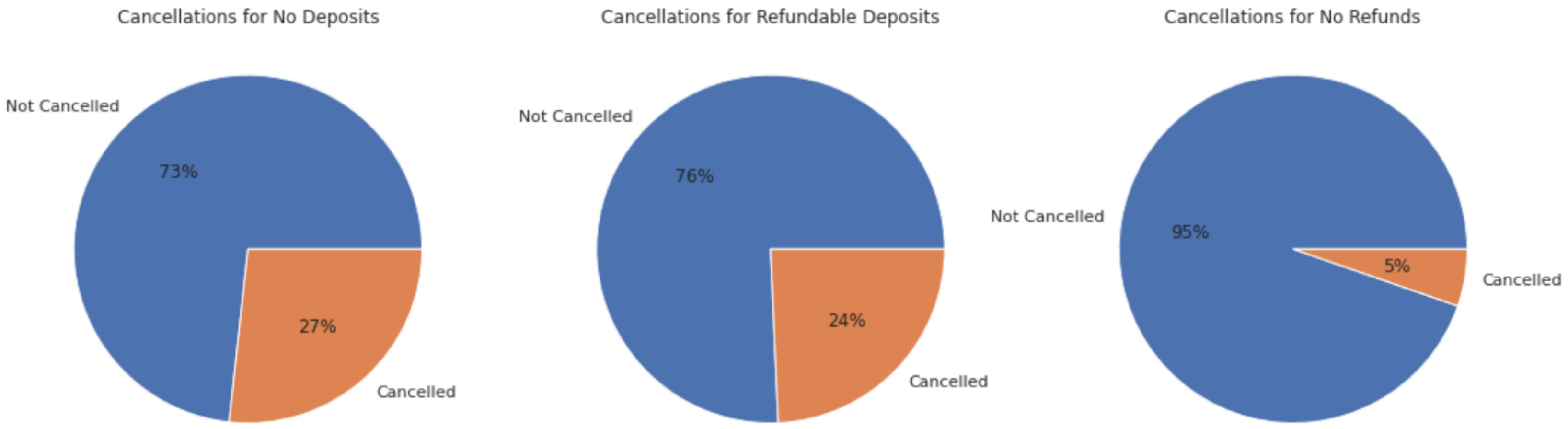
Percentage shares of bookings by hotel types

Cancellations

Out of all the bookings, 28% of them was cancelled and 72% was not cancelled.



Cancellations by Deposit Type

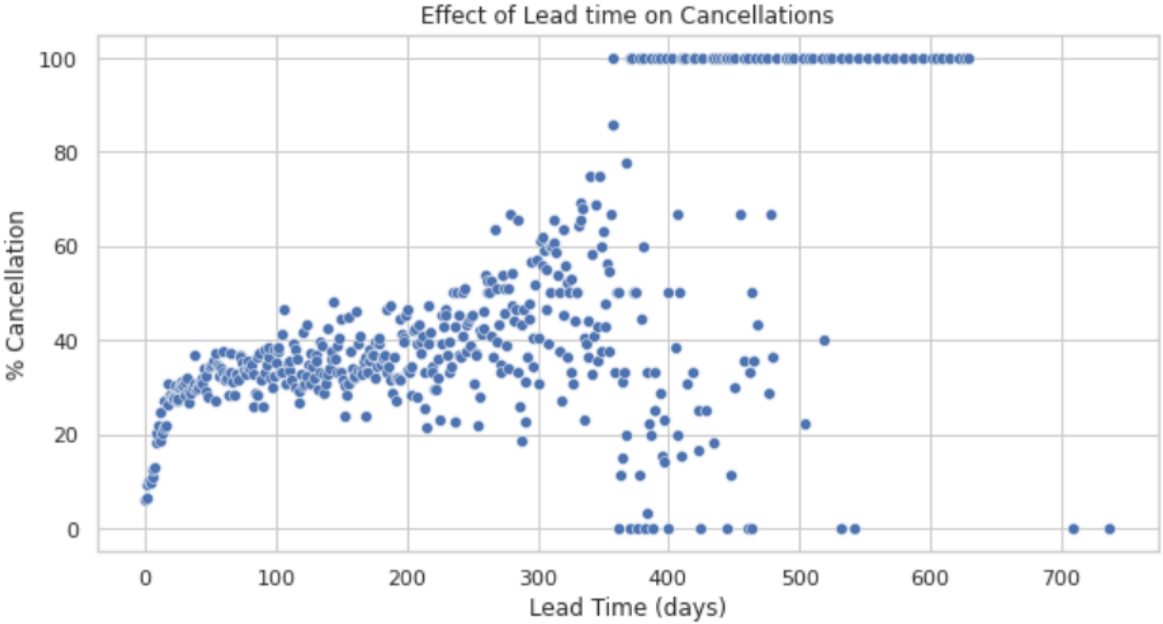


```
Total cancellation percentage : 28%
```

Around 1/4th of bookings were cancelled where customer will get the money back. Only 5% cancels their bookings when they can't get their money back.

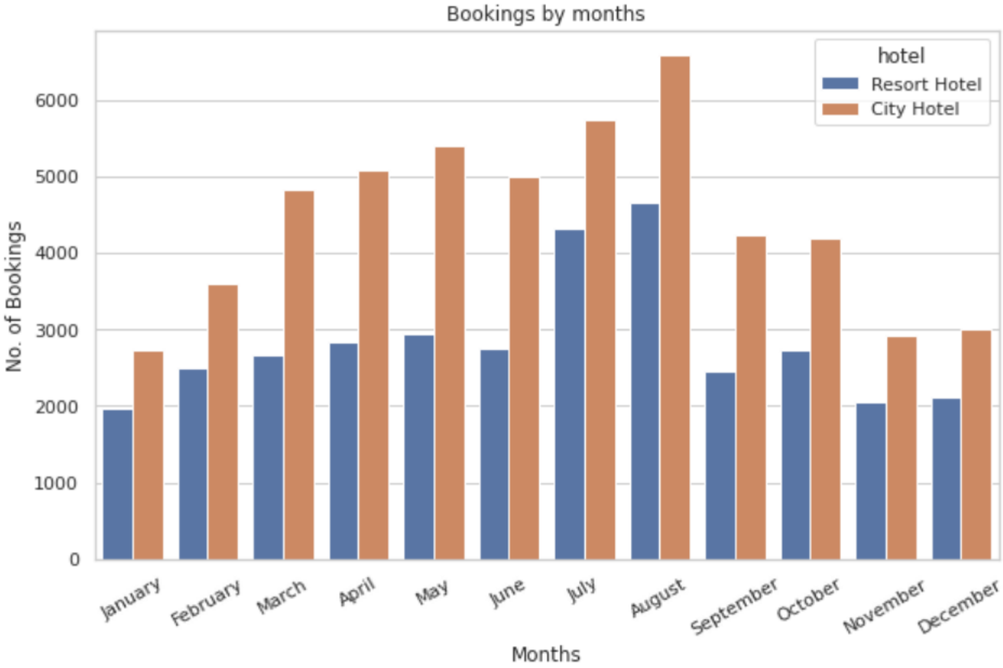
Lead Time vs Cancellations

Bookings having less lead time are less often cancelled whereas high lead time bookings are more often cancelled.



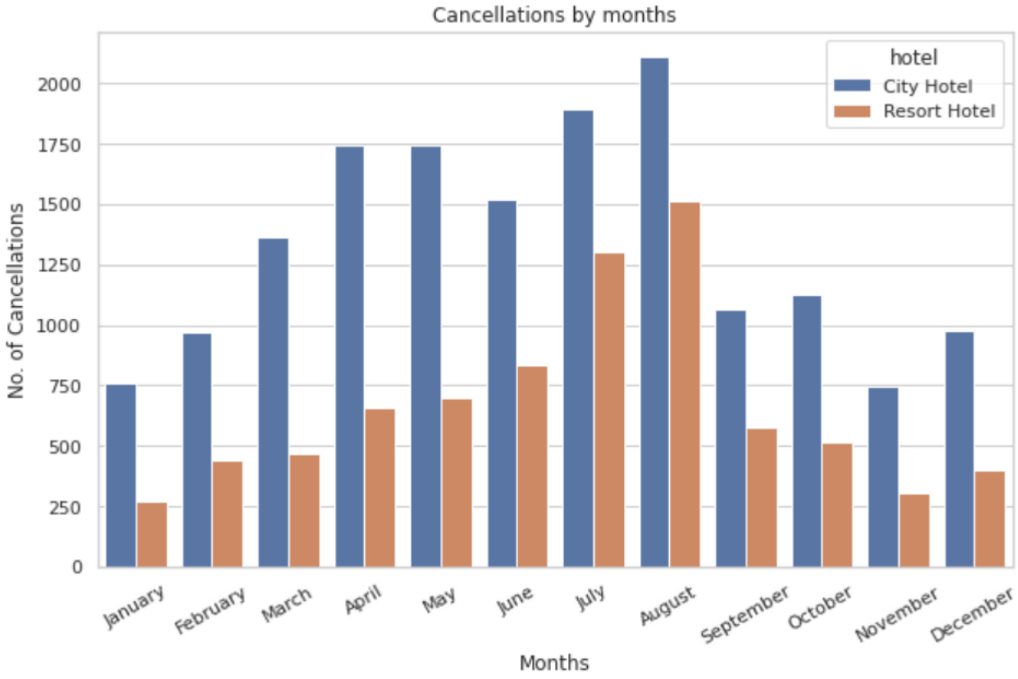
Monthly Bookings

For both hotel types, winter months have less bookings compared to mid-year (or summer) months.



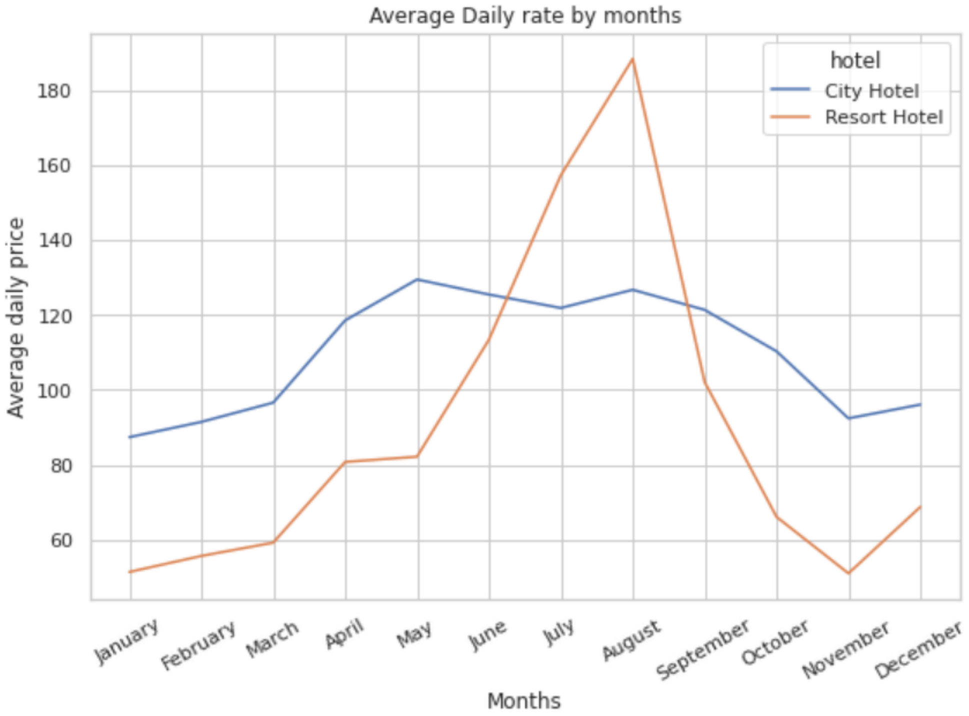
Monthly Cancellations

Graphs for total bookings and total cancellations for both hotel types are similar which means with more number of bookings, more are the cancellations.



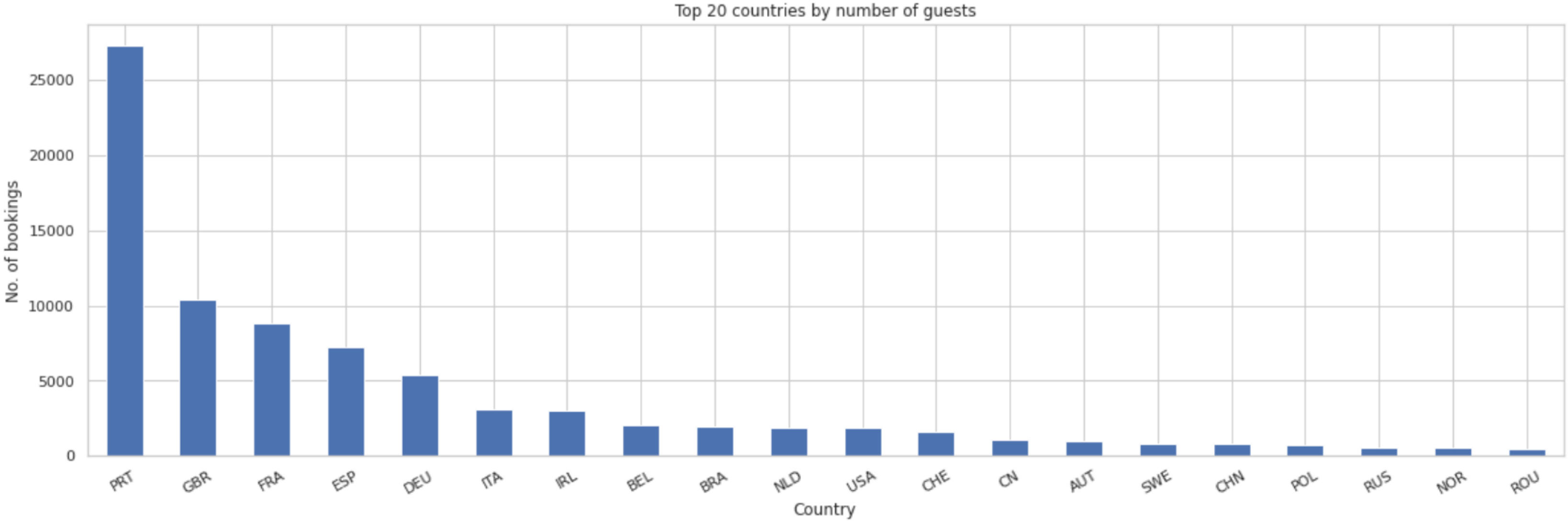
Daily Rates by Months

adr is higher in summer months and lower in winter months, and we saw the similar pattern in total bookings over months. Simple market principle i.e., more demand more price and less demand less price.



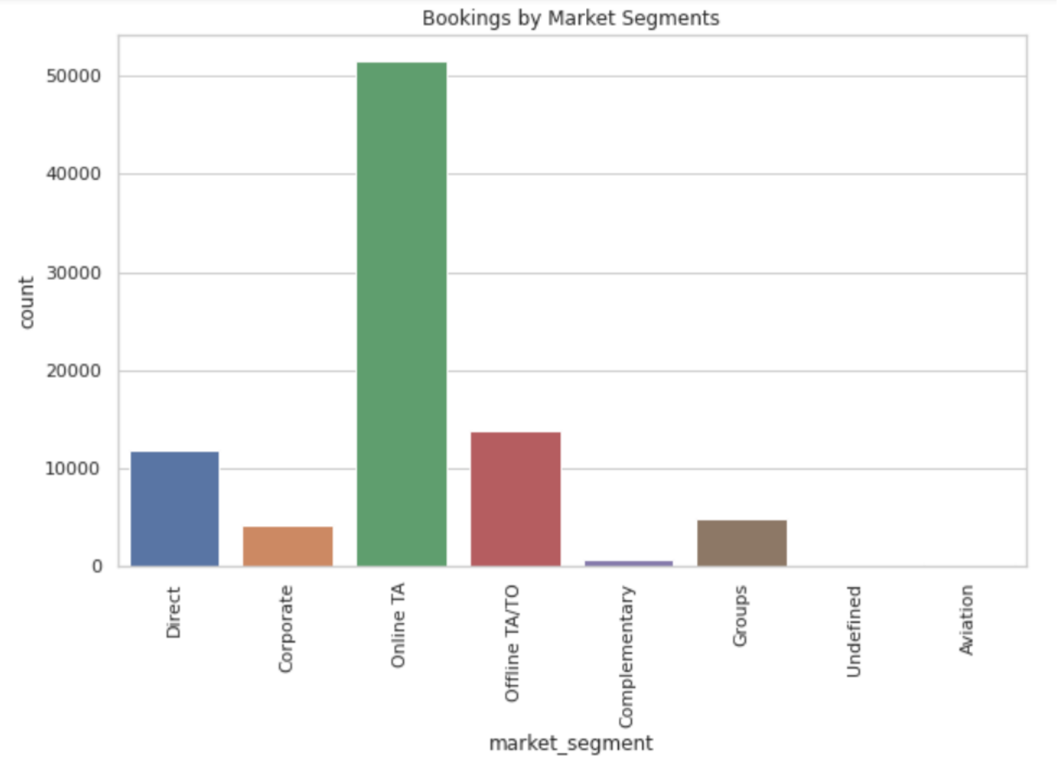
Bookings by Country

Portugal has significantly higher number of bookings compared to other countries. Next major countries are GBR, FRA, ESP etc. Most of the top countries are European countries.



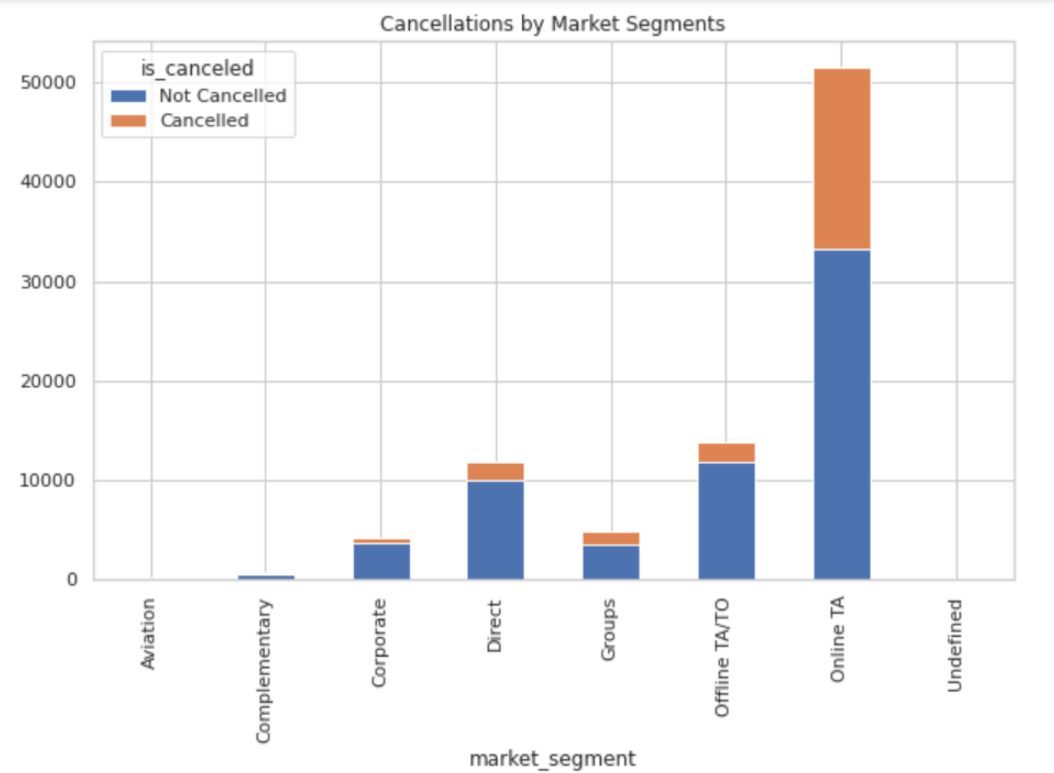
Bookings by Market Segments

Indirect bookings through travel agents (specially Online) are higher compared to other bookings like direct, corporate, group etc.



Cancellations by Market Segments

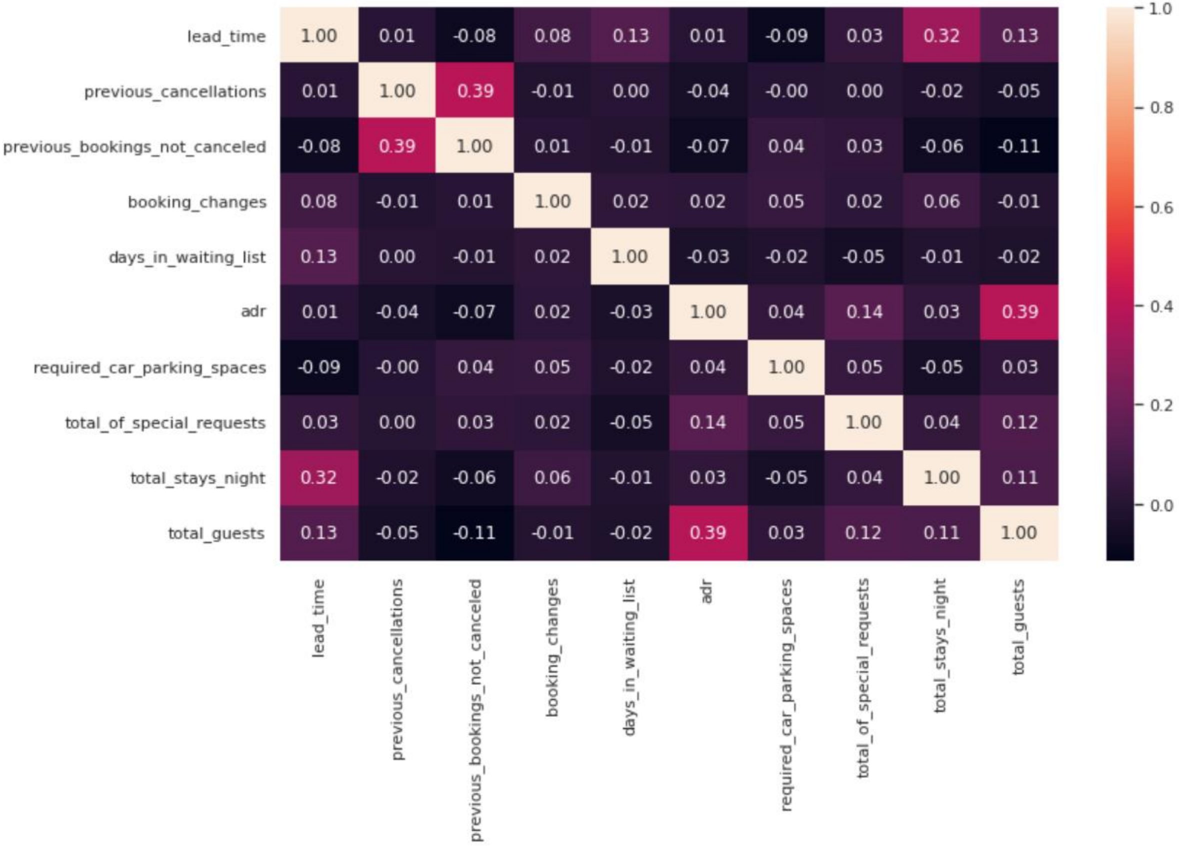
We see that Online TA has highest proportion of cancellations (around 35%) but at the same time number of non cancelled bookings are also the highest for Online TA. Cancellation proportions for other segments are less.



Correlation of features

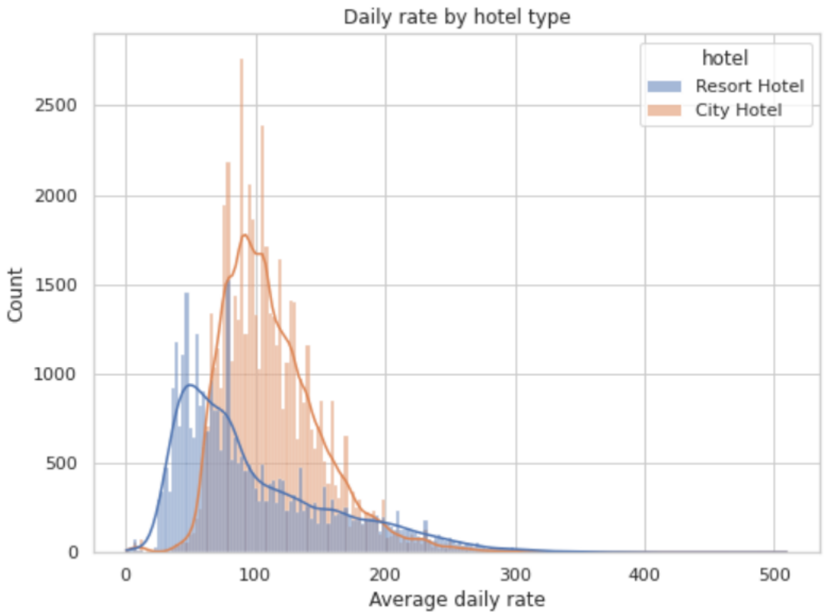
Lead time and total stays night have little correlation. This may indicates that people who wants to stay longer generally plans early hence longer lead time.

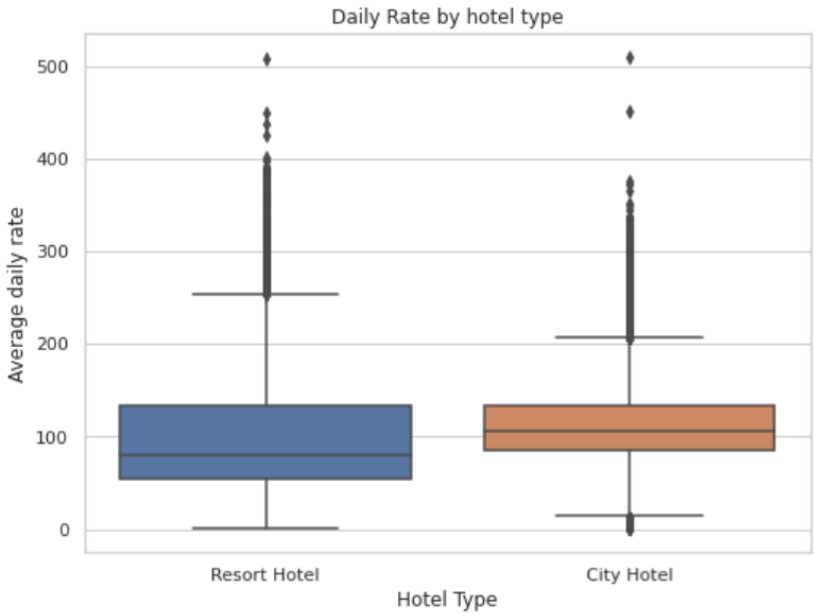
adr and total guests are also slightly correlated, which makes sense as more number of people means more cost/charges.



	lead_time	previous_cancellations	previous_bookings_not_canceled	booking_changes	days_in_waiting_list	adr	required_car_parking_spaces	total_of_special_requests	total_stays_night	total_guests
lead_time	1.00	0.01	-0.08	0.08	0.13	0.01	-0.09	0.03	0.32	0.13
previous_cancellations	0.01	1.00	0.39	-0.01	0.00	-0.04	-0.00	0.00	-0.02	-0.05
previous_bookings_not_canceled	-0.08	0.39	1.00	0.01	-0.01	-0.07	0.04	0.03	-0.06	-0.11
booking_changes	0.08	-0.01	0.01	1.00	0.02	0.02	0.05	0.02	0.06	-0.01
days_in_waiting_list	0.13	0.00	-0.01	0.02	1.00	-0.03	-0.02	-0.05	-0.01	-0.02
adr	0.01	-0.04	-0.07	0.02	-0.03	1.00	0.04	0.14	0.03	0.39
required_car_parking_spaces	-0.09	-0.00	0.04	0.05	-0.02	0.04	1.00	0.05	-0.05	0.03
total_of_special_requests	0.03	0.00	0.03	0.02	-0.05	0.14	0.05	1.00	0.04	0.12
total_stays_night	0.32	-0.02	-0.06	0.06	-0.01	0.03	-0.05	0.04	1.00	0.11
total_guests	0.13	-0.05	-0.11	-0.01	-0.02	0.39	0.03	0.12	0.11	1.00

Daily Rate Distribution

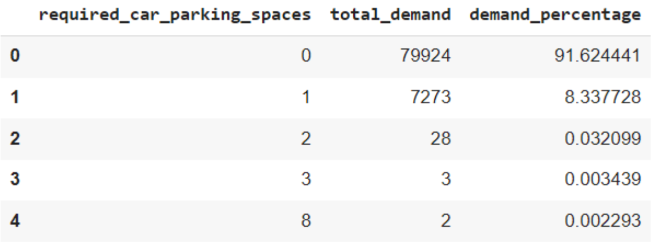




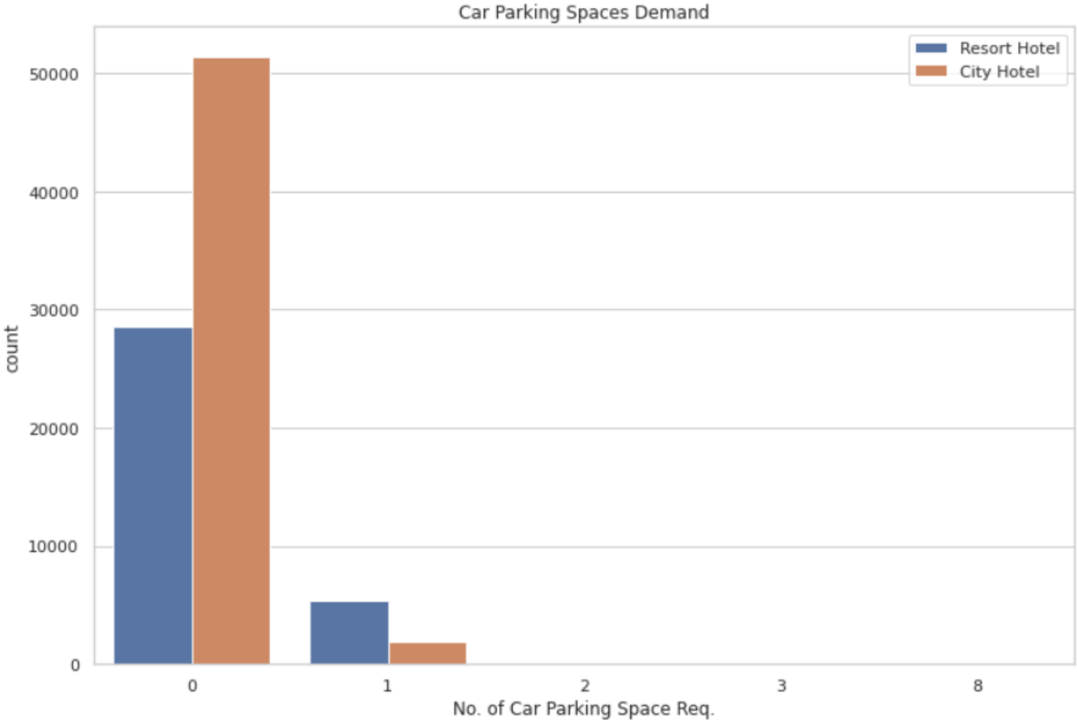
Car Parking Space

More than 91% guests don't need parking space.

Approx. 8% guests need parking space for 1 car.

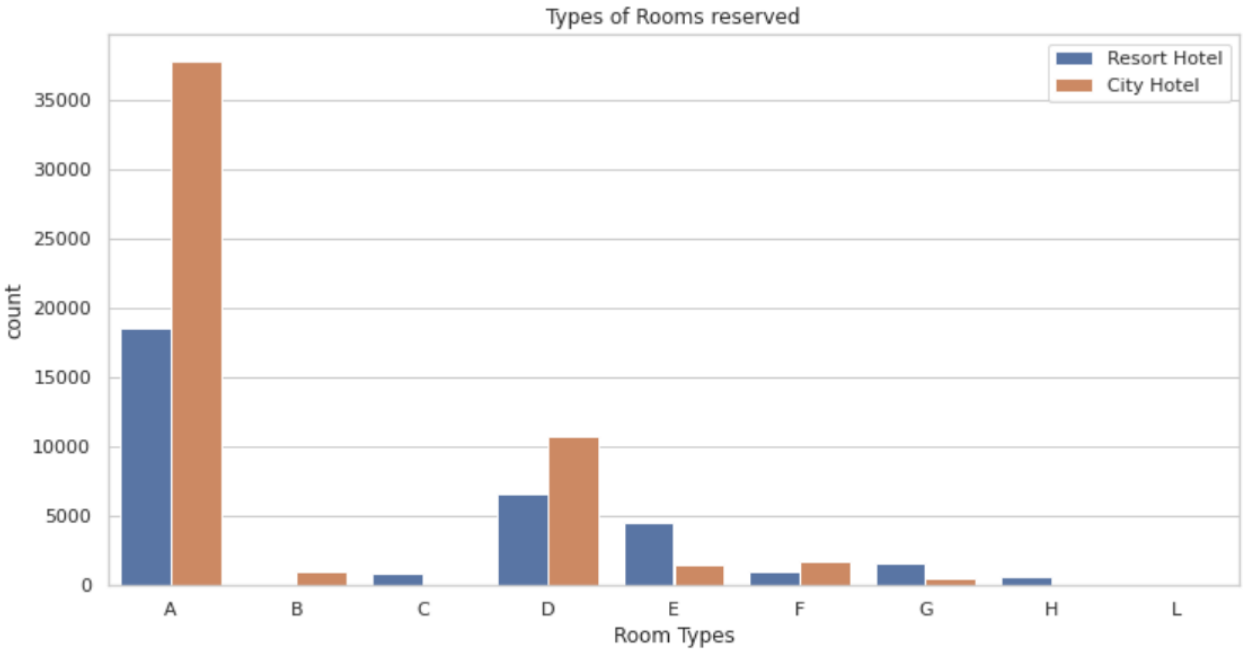


	required_car_parking_spaces	total_demand	demand_percentage
0	0	79924	91.624441
1	1	7273	8.337728
2	2	28	0.032099
3	3	3	0.003439
4	8	2	0.002293



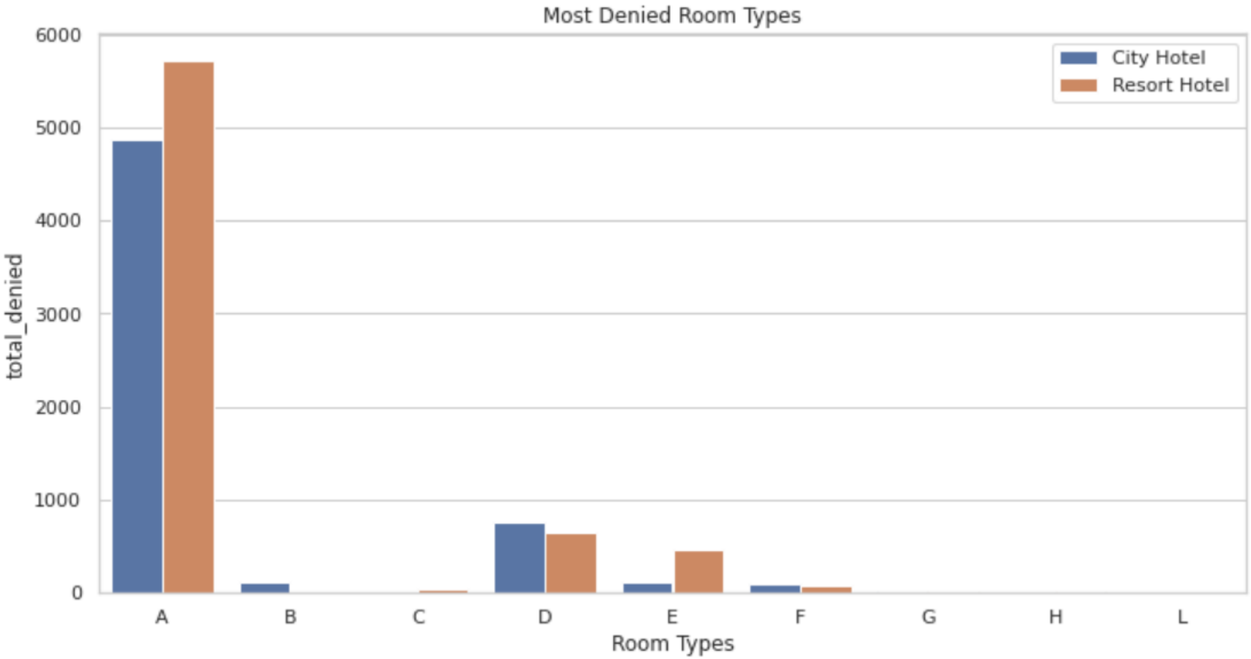
Room Types

Most of guests reserves rooms of type 'A', also significant number of guests also reserves room type 'D' & 'E'. Demand for rest types are very minimal.



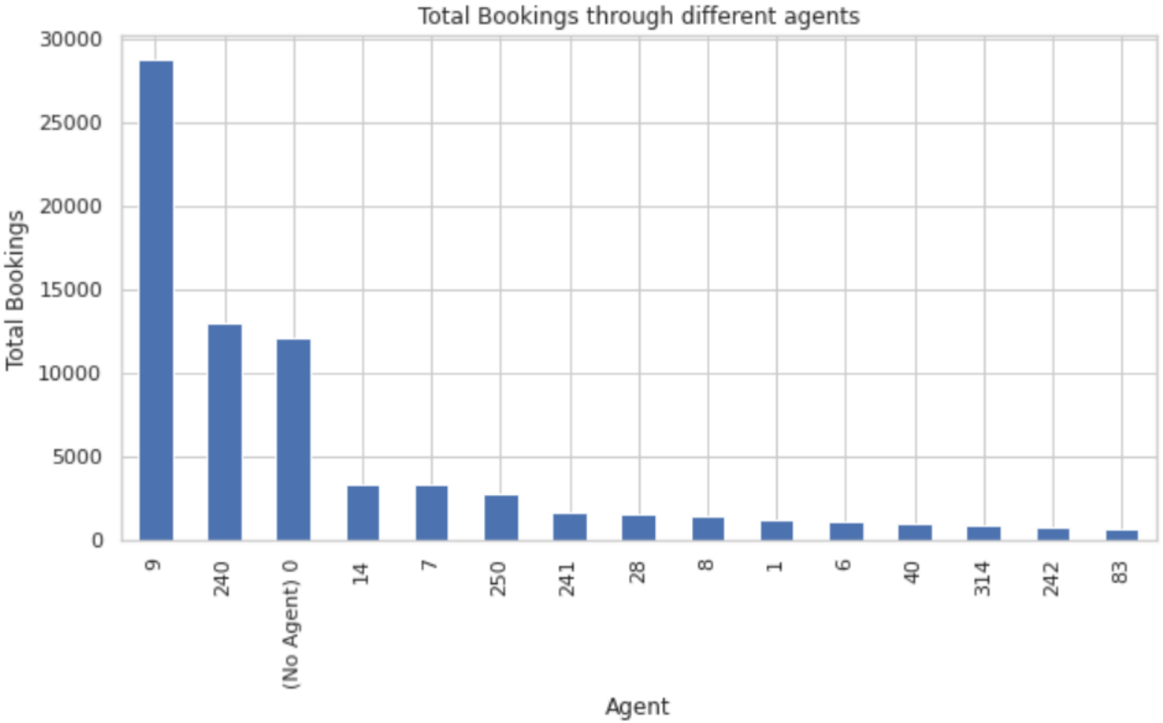
Room Types

Similar to number of bookings, room type ‘A’ has most number of denials. Followed by ‘D’ and ‘E’ etc.



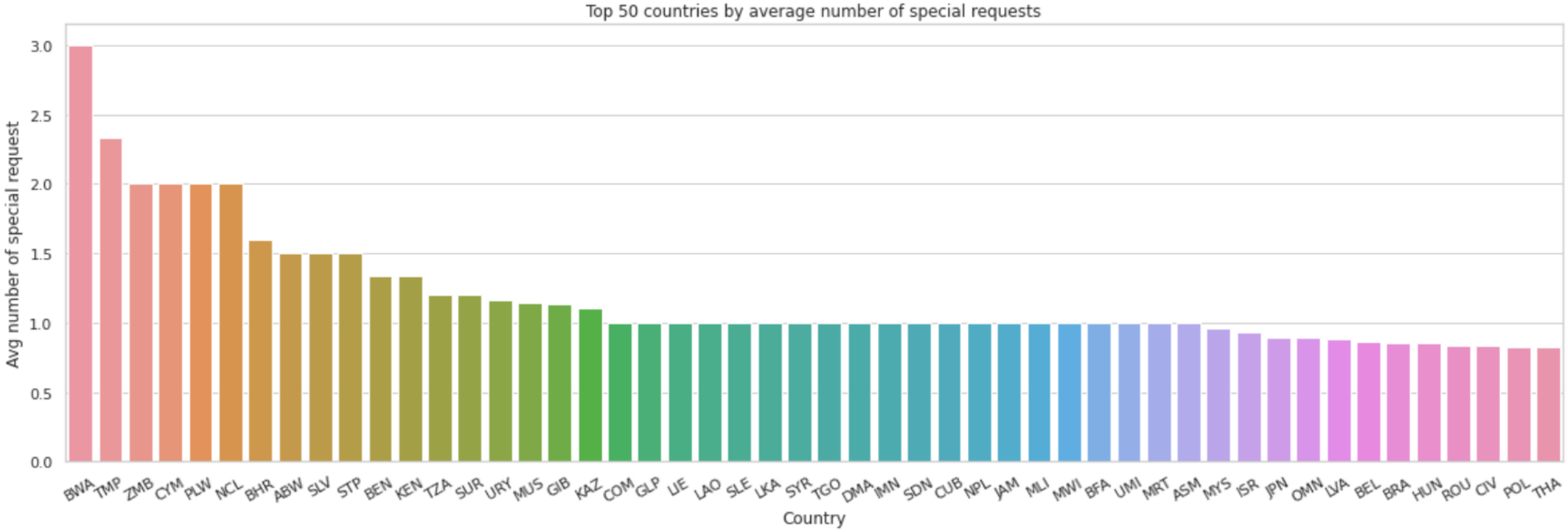
Top Agents

Agent no. 9 has made most no. of bookings followed by 240, 14 etc.



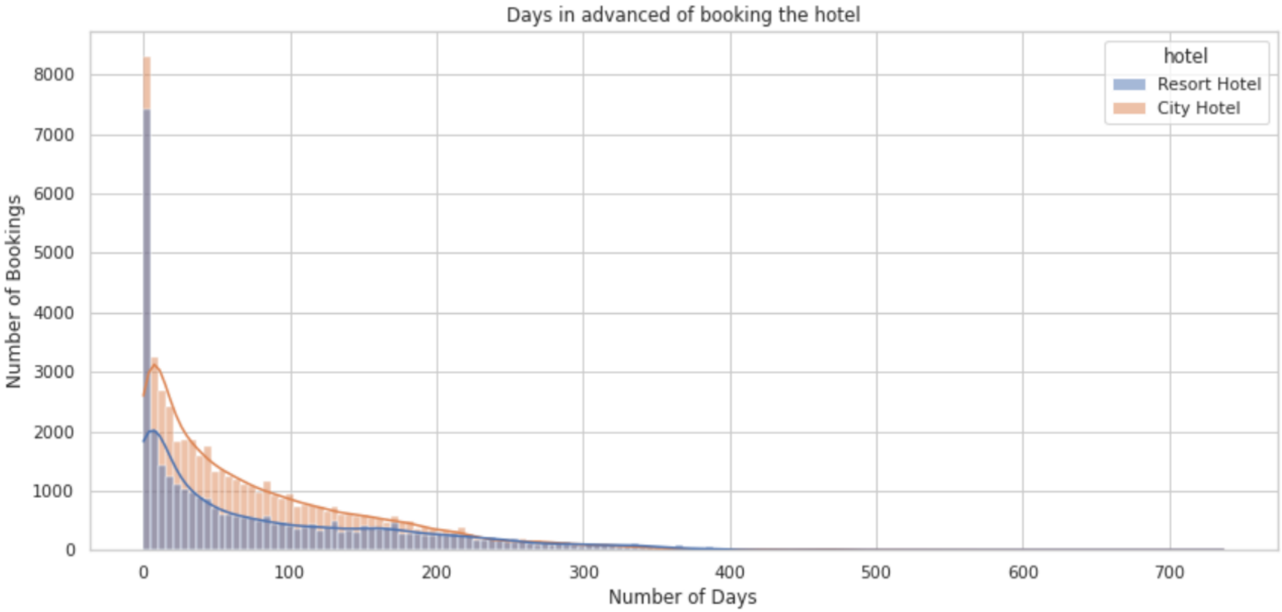
Special Requests by Country

Botswana, East Timor, Zambia are some countries with highest number of special requests.



Lead Time

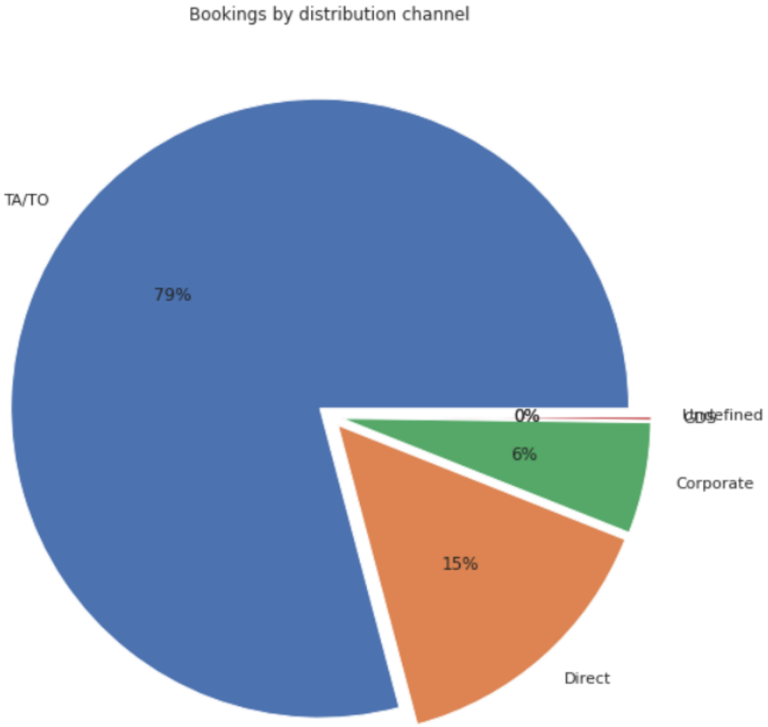
Majority of the bookings are done within 100 days of check-in date to the hotel.



Bookings by Distribution Channel

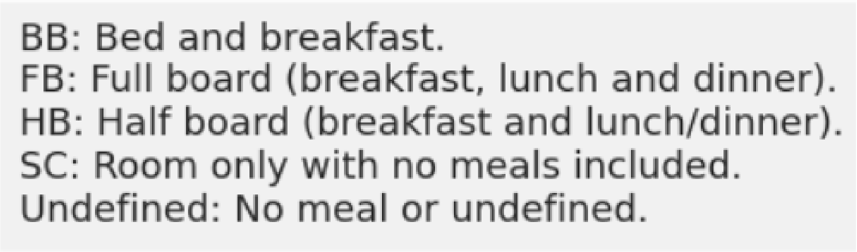
Around 79% of bookings are done through TA/TO distribution channel.

Other major distribution channels are Direct and Corporate.

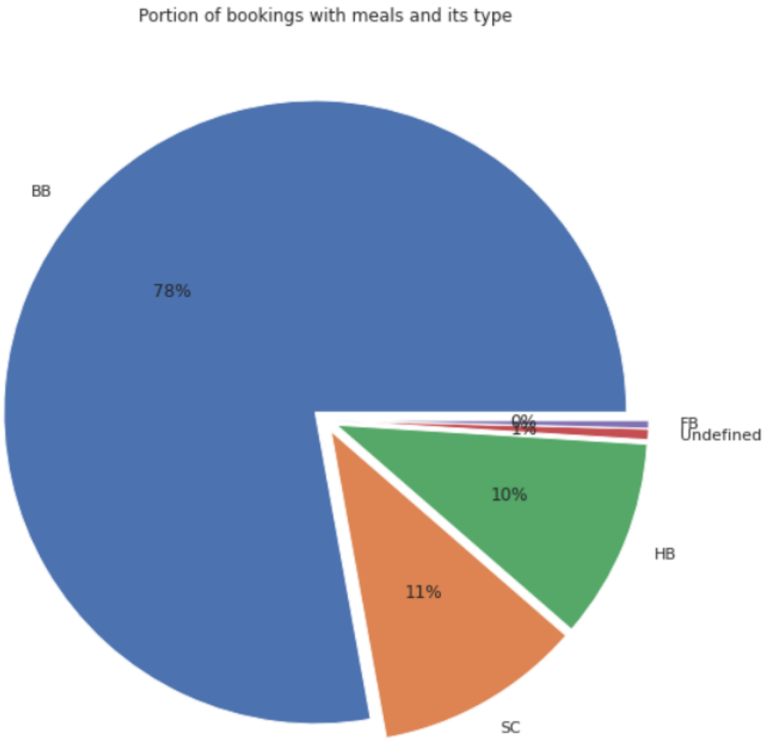


Meals

Almost 90% of total guests books meal in their bookings, among which majority books meal of type BB (78%) followed by SC (11%), HB (10%) etc.



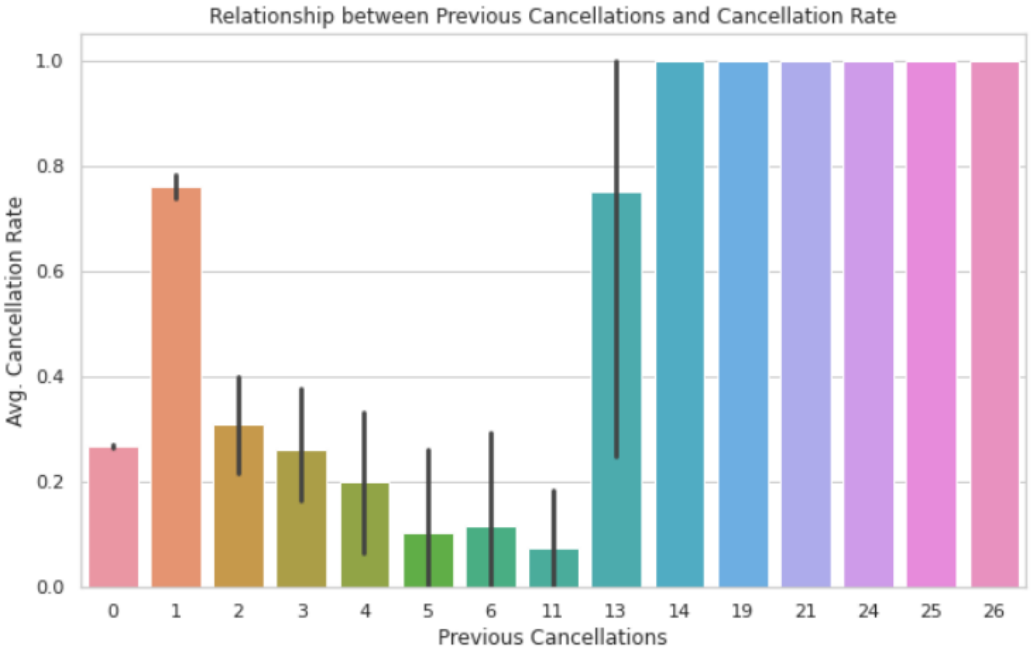
BB: Bed and breakfast.
FB: Full board (breakfast, lunch and dinner).
HB: Half board (breakfast and lunch/dinner).
SC: Room only with no meals included.
Undefined: No meal or undefined.



Portion of bookings with meals and its type

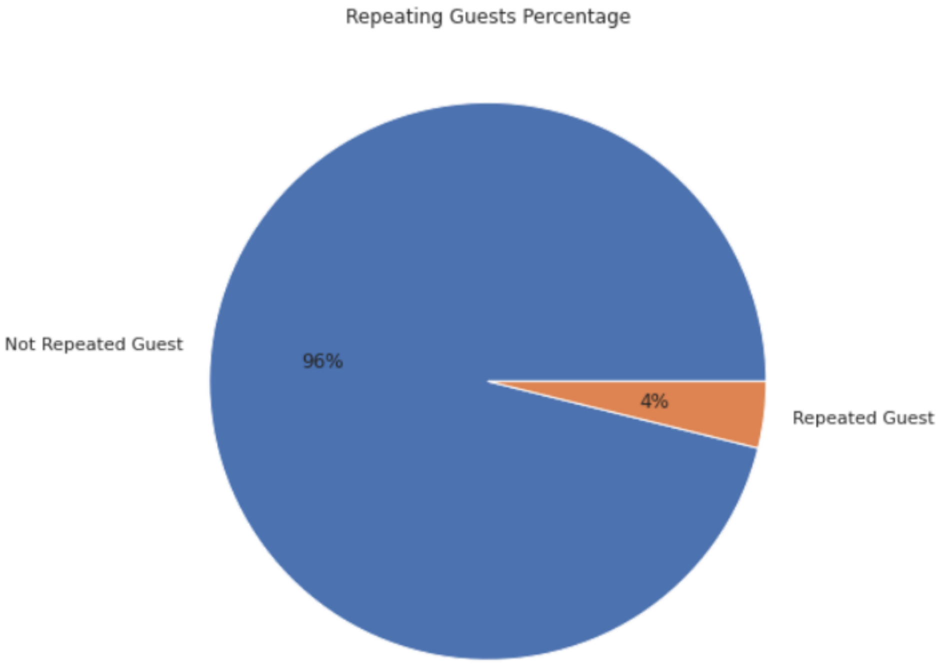
Previous and Current cancellations

Those who have cancelled previously once have high chance of cancellation compared to 0 previous cancellations. For higher values of previous cancellations, numbers are very less and can be ignored.



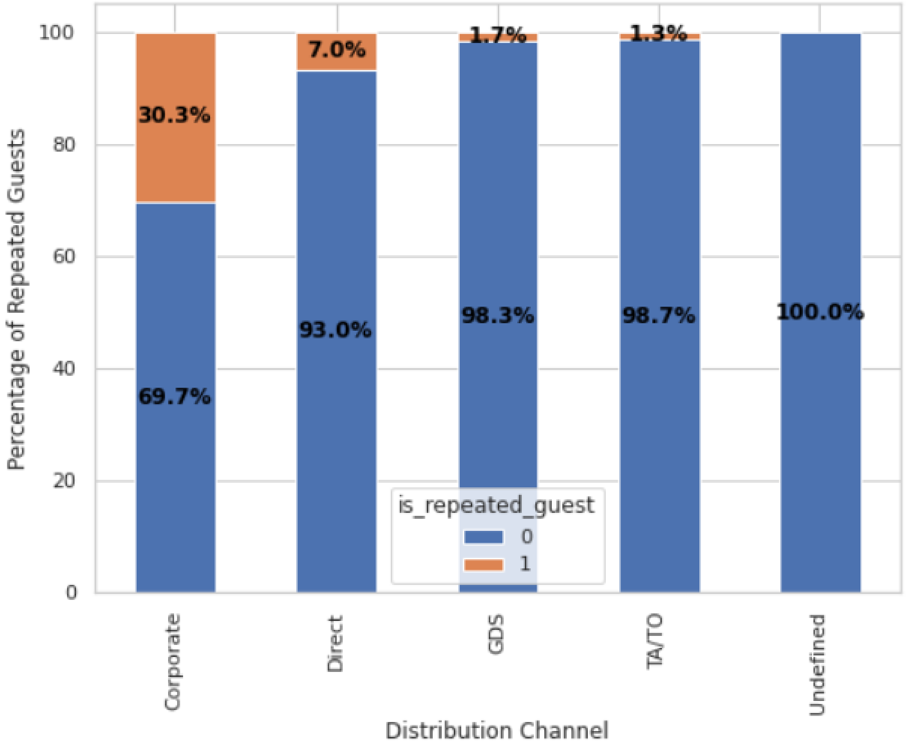
Repeated Guests

Around 4% of total bookings are made by repeated guests. So majority bookings are from new customers.



Repeated Guests by Distribution Channel

Corporate has most number of repeated guests followed by Direct, GDS etc.



Conclusions

- City Hotels have more bookings (61%) compared to Resort Hotels (39%).
- Non-refundable deposits decreases significantly the chances of cancellation.
- Longer lead time increases the chances of cancellations.
- Summer months (April to August) have more bookings compared to winter months.
- Portugal has the highest number of customers followed by Britain, France etc.
- More bookings comes through TA/TO (Online/Offline) compared to Direct, Corporates etc.
- Most customers don't need parking space around 91%.
- Room Type A is the most preferred room type among customers.
- Agent no. 9 has made most no. of bookings followed by 240, 14 etc.
- Almost 90% of total guests books meal in their bookings, among which majority books meal of type BB (78%) followed by SC (11%), HB (10%) etc.
- Corporate has the most percentage of repeated guests while TA/TO has the least whereas in the case of cancelled bookings TA/TO has the most percentage while Corporate has the least.

Thank You!