Data Analysis-Hotel Booking

Parashuram N Vaddar
Data Analyst

Abstract:

This data article describes two datasets with hotel demand data. One of the hotels is a resort hotel and the other is a city hotel. Both datasets share the same structure, with 31 variables describing the 119390 observations. Each observation represents a hotel booking. Both datasets comprehend bookings due to arrive between the 1st of July of 2015 and the 31st of August 2017, including bookings that effectively arrived and bookings that were cancelled. Since this is hotel real data, all data elements pertaining hotel or customer identification were deleted. Due to the scarcity of real business data for scientific and educational purposes, these datasets can have an important role for research and education in revenue management, machine learning, or data mining, as well as in other fields.

We will perform exploratory data analysis with python to explore and analyse the data to discover important factors that govern the bookings.

Keywords: Exploratory data analysis, hotel booking, special requests, best time for booking

1. Problem Statement

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.

Data content 31 variables their descriptive information is given below.

1.1 Features Description

1. **hotel**: Hotel (Resort Hotel or City Hotel)

2. **is_canceled**: Value indicating if the booking was cancelled (1) or not (0)

3. **lead_time**: Number of days that elapsed between the entering date of the booking into the PMS and the arrival date

4. **arrival_date_year**: Year of arrival date

5. **arrival_date_month**: Month of arrival date

6. **arrival_date_week_number**: Week number of year for arrival date

7. **arrival_date_day_of_month**: Day of arrival date

8. **stays_in_weekend_nights**: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel

9. **stays_in_week_nights**: Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel

10. **adults**: Number of adults

11. **children**: Number of children

12. **babies**: Number of babies

13. **meal**: Type of meal booked. Categories are presented in standard hospitality meal packages:

14. **country**: Country of origin.

15. **market_segment**: Market segment designation. In categories, the term "TA" means "Travel Agents" and "TO" means "Tour Operators"

16. **distribution_channel**: Booking distribution channel. The term "TA" means "Travel Agents" and "TO" means "Tour Operators"

17. **is_repeated_guest**: Value indicating if the booking name was from a repeated guest (1) or not (0)

18. **previous_cancellations**: Number of previous bookings that were cancelled by the customer prior to the current booking

19. **previous_bookings_not_canceled**: Number of previous bookings not cancelled by the customer prior to the current booking

20. **reserved_room_type**: Code of room type reserved. Code is presented instead of designation for anonymity reasons.

21. **assigned_room_type**: Code for the type of room assigned to the booking.

22. **booking_changes**: Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation

23. **deposit_type**: Indication on if the customer made a deposit to guarantee the booking.

24. **agent**: ID of the travel agency that made the booking

25. **company**: ID of the company/entity that made the booking or responsible for paying the booking.

26. **days_in_waiting_list**: Number of days the booking was in the waiting list before it was confirmed to the customer

27. **customer_type**: Type of booking, assuming one of four categories

28. **adr**: Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights

29. **required_car_parking_spaces**: Number of car parking spaces required by the customer

30. **total_of_special_requests**: Number of special requests made by the customer (e.g., twin bed or high floor)

31. **reservation_status**: Reservation last status, assuming one of three categories

Cancelled – booking was cancelled by the customer
Check-Out – customer has checked in but already departed
No-Show – customer did not check-in and did inform the hotel of the reason why

1. **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 was the booking cancelled or when did the customer checked-out of the hotel

2. Introduction

Customers want to know the best time of year to book a hotel room and the optimal length of stay in order to get the best daily rate. There are two hotels available. One is a city hotel and the other is a resort hotel.

Our goal here is the exploratory data analysis and give the customer the best prediction to book the hotel.

Let's take a look at the hotel bookings of some holidaymakers and try to understand their accommodation requirements for maximum relaxation.

3. Data Cleaning

Data cleaning is the infrastructure and the first block in data science Before we do any analysis. we need to make sure our data is clean and credible.

1. Check for any null values:

There are some null values in the company, agent, country and children's columns.

- if there is no company, the booking would most likely have been made privately, and so these null values can be changed to 0
- if there is no agent, the booking would most likely have been made without an agent, and so these null values can be changed to 0
- if the country is a null value, we can just list it as other
- if there are null values of children, the number of children can be assumed 0

2. Dropping duplicate values:

Looks like there are quite a few duplicate values (31980) present in data which we have to drop.

3. Apparently, there are bookings with 0 adults - seems a little bit irresponsible to book a hotel room just for children and/or babies. Anyway, let's filter out the data where there were no guests (i.e., 0 adults, 0 children and 0 babies)

4. Again, an average daily rate of 5400 seems a little too extreme when the next closest rate is 510. I'm assuming it's just a typo for 540, so let's change this too.

Now we are ready to do some exploratory data analysis.

4. Exploratory Data Analysis (EDA)

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypotheses and to check assumptions with the help of summary statistics and graphical representations.

Following is some analysis which we have done.

I. Which hotels are guests booking?

The database is divided into two types of hotels: "City" hotels and "Resort" hotels.
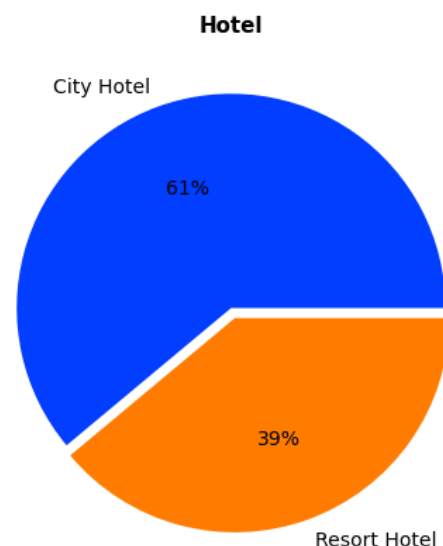
Here City Hotel booking is 53274 out of total booking and Resort booking is 33956.

Almost a third of the bookings are for city hotels, but these include bookings that were both cancelled and not cancelled.
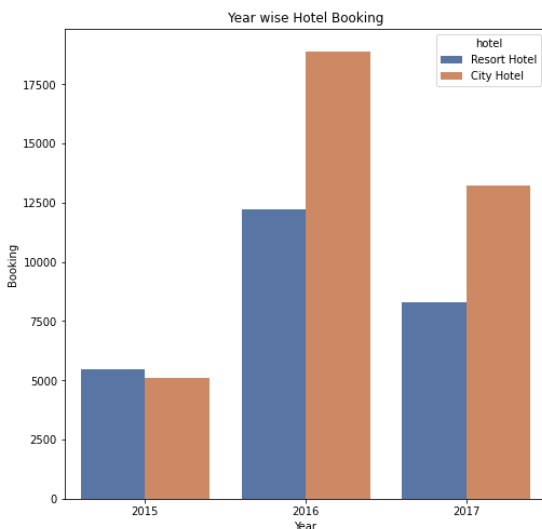
II. Cancelled Booking

Almost 23% of the bookings at the City Hotel were cancelled, while 19% were cancelled at the Resort Hotel. Maybe guests are more prone to cancelling their booking if they don't have to pay a deposit?
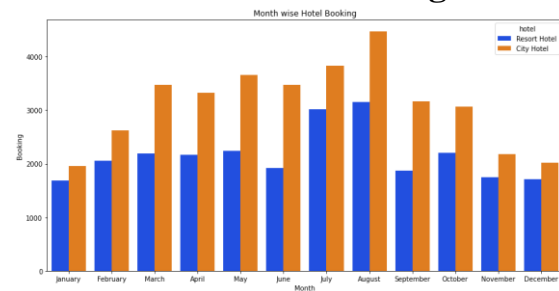
III. Year wise Booking

Both datasets comprehend bookings due to arrive between the 1st of July of 2015 and the 31st of August 2017.

2016 was the best year with the highest number of arrivals for both the Resort Hotel and the City Hotel. 2015 have lest number of bookings.

IV. Month Wise Booking

There was a slight trend with a higher number of arrivals during the summer months (June to mid-September) which could be due to summer holidays for the children and also just the nicer weather!

## V. Month and hotel wise ADR (Average Daily Rate)
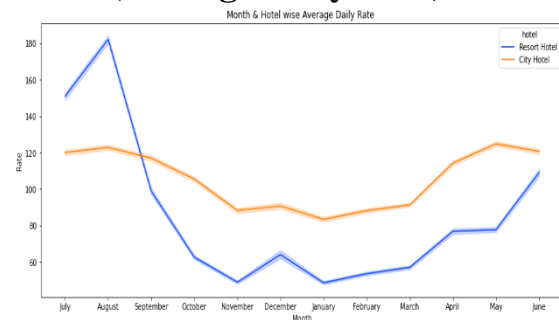
The average daily rate is equal to the average price for each hotel room sold for that day. The average daily rate fluctuates more for resort hotels than city hotels with standard deviations of 60 and 42 respectively. This is especially evident during the summer months (June, July, August and September) with prices ranging from 98 to 182. The winter months (December, January, February and March) generally have lower prices ranging from 57 to 91, except that there is a little peak in December which is probably due to Christmas and New Year's. Even with the seasonal prices, it is the city hotels that maintain a higher price with a higher mean and median throughout the year.

VI. Guest Country of Origin

Country wise Hotel Booking
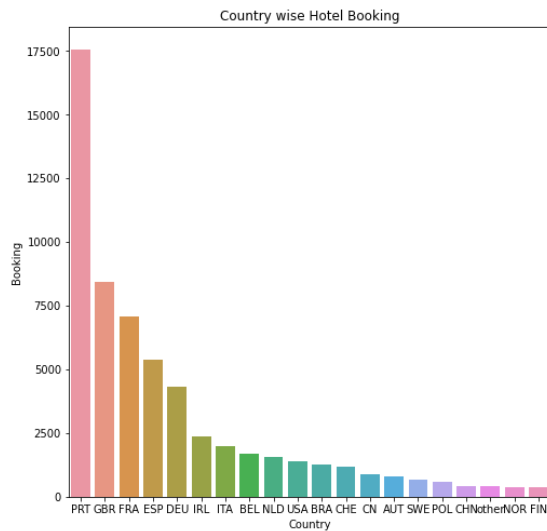
Oh yes, something to be pointed out is that these hotels are based in Portugal! So, most of the guests are local tourists, followed by the British, French and Spanish.

VII. Number of weekend/weekday nights booked

Number of weekend/weekday nights booked
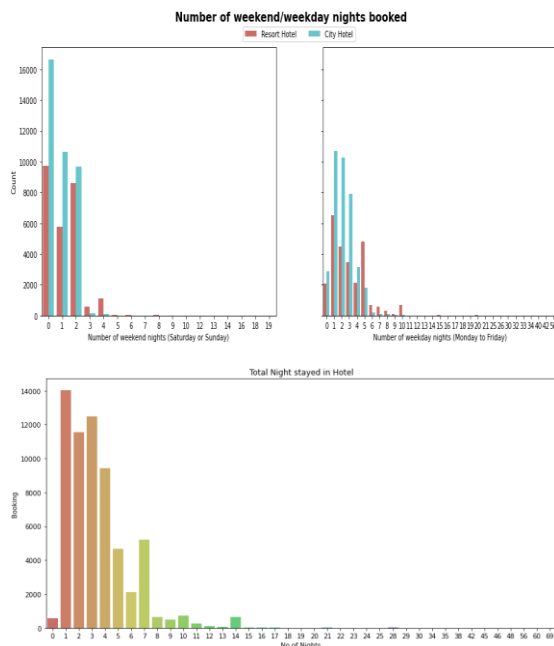
Guests are more likely to book a hotel over the weekday as the majority of the guests don't even stay over for one weekend night, and the number of weekday nights booked even extends to 50! I guess some guests are

having a really long holiday, and probably an expensive one too, right?

- **Total Average Rate**

No of Nights Stayed & Hotel wise Total Average Rate
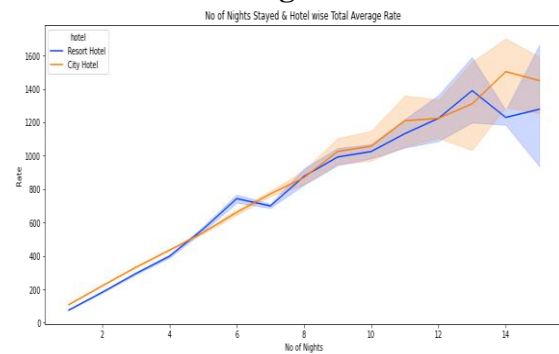
Total Average Rate = adr*($stays\_in\_week\_nights$ + $stays\_in\_weekend\_nights$)
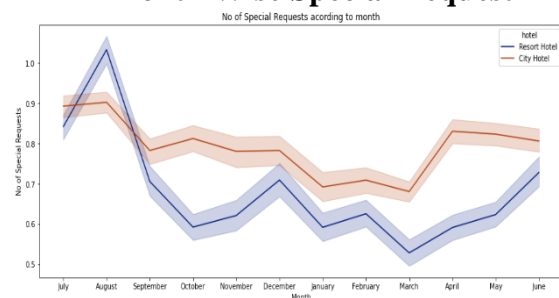
Here we see an increase in total night stayes so the average rate also increases so to get the best deal you have to book when booking rates are low as we see in month wise booking.

VIII. Special Requests
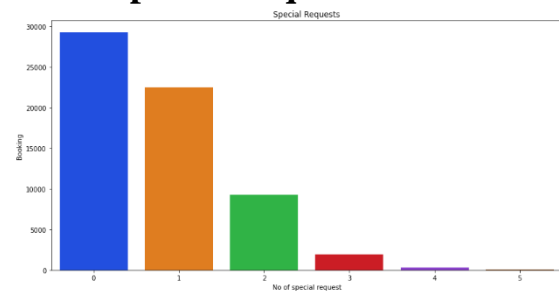
Special Requests

The number of special requests ranges from 0 to 5, with about half being 0.
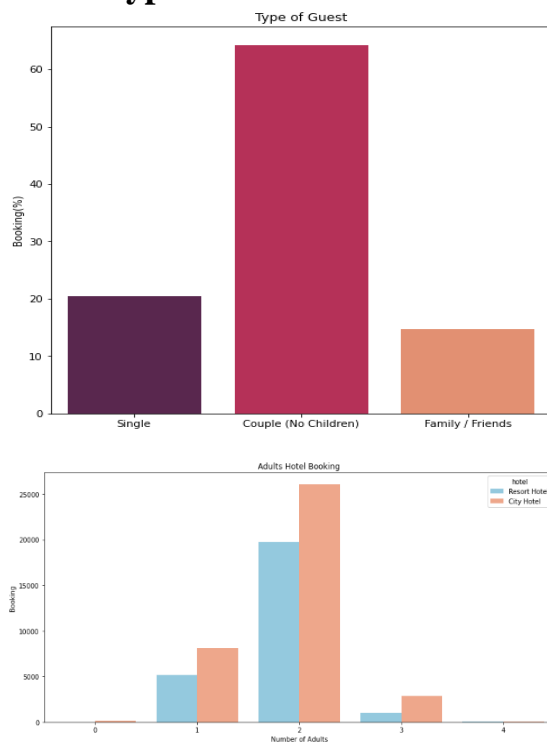
- **Month Wise Special Request**

No of Special Requests according to month

Special requests increase when booking increases. In august month there are more special requests, we can predict that in summer time (June to mid-September) special requests will max when no of
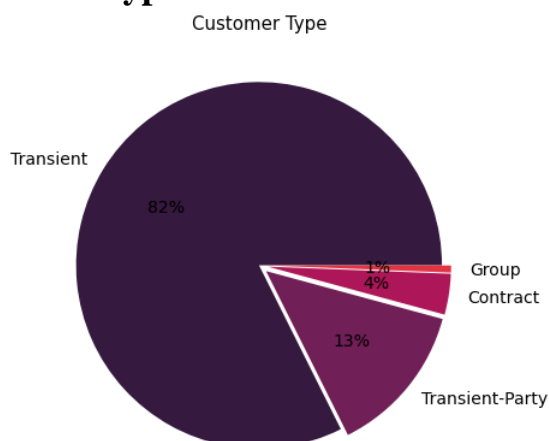
booking are most so we can prepare accordingly.

IX. Type Of Guest

Majority of the adults are pairs. Almost all of the guest's holiday without any children or babies, and those that don't have too huge a preference in terms of a resort or a city hotel, perhaps just a slight lean towards a city hotel.
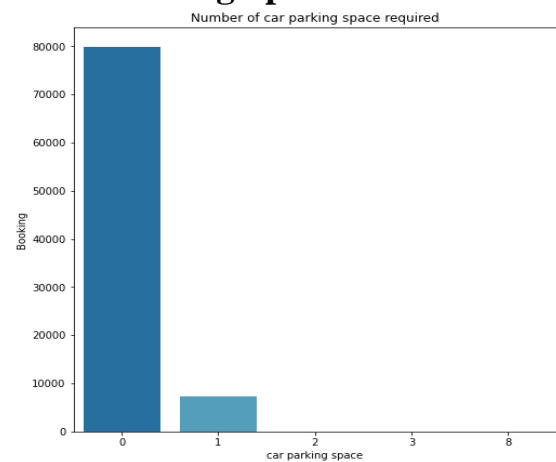
X. Type of Customer

About 82% of them are not part of a party (i.e., Transient) and 13% of them are (i.e., Transient-Party). Contract and Group
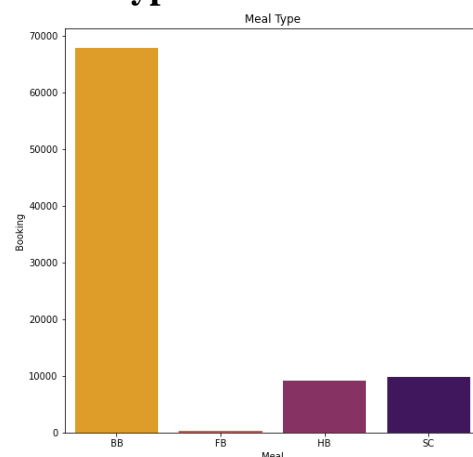
bookings make up the remaining 4% with 3135 and 541 respectively.

XI. Parking Space

The number of car parking spaces requests ranges from 0 to 8, with the majority of them being 0.

XII. Types of meals booked


BB - bed and breakfast

FB - full board

HB - half board

SC - no meal

Approximately 77% of the guests opt for bed and breakfast, followed by 12% for half board which is breakfast and one other meal (usually dinner), 10% for no meal, and the remaining 1% for full board which is breakfast, lunch and dinner.

XIII. Market Segment & Booking Distribution Channel

Market segments and distribution channels
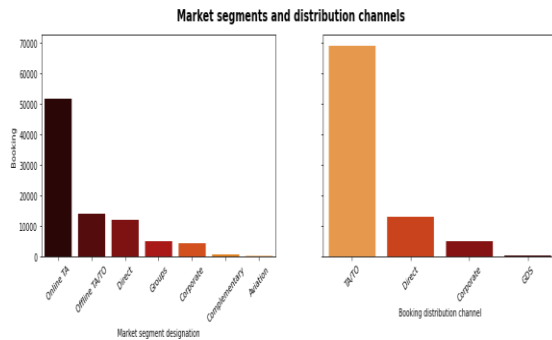
"TA" means "Travel Agents" and "TO" means "Tour Operators"

Majority of the guests booked the hotels through both online and offline travel agents and tour operators (i.e., Online TA, Offline TA/TO, TA/TO) as opposed to Direct bookings themselves and corporate bookings.

XIV. Average daily Rate (Lead Time)
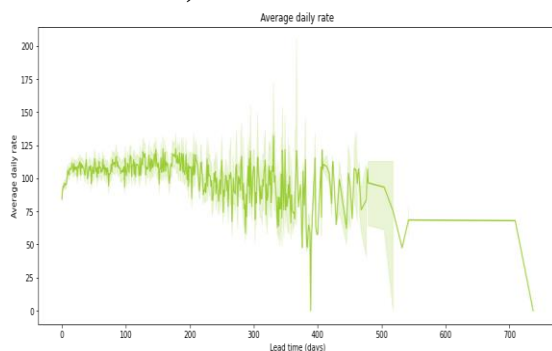
Average daily rate

Making a booking at least 175 days in advance does seem to give a lower price in general, or the next best would be to sweep up any last-minute deals within 10 days of the booking.

XV. Average daily Rate (Repeated Guest)

Average daily rate (Repeated Guest)
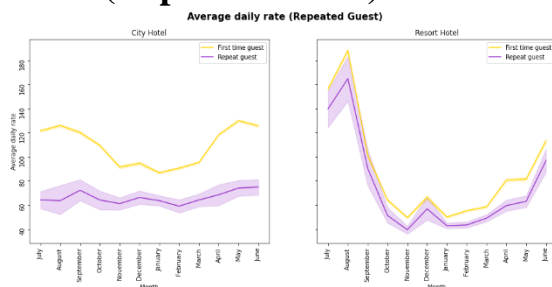
Customers will get better deal if they are repeat guests regardless of whether they book a city or resort hotel! Perhaps it is time to join the hotel loyalty programme.

5. Conclusion

That's it! We reached the end of our exercise. Starting with loading the data so far, we have done null values treatment and EDA.

1) Almost 61% of the total guests opt for a city hotel instead of a resort hotel.
2) In both hotels 19% to 24% bookings were cancelled.
3) Summer months are more expensive for resort hotels while the prices for city hotels don't really fluctuate throughout the year.
4) Special requests increase in summer time when booking is high.
5) Almost all of the guest's holiday without any children or babies.
6) Majority of the guests booked the hotels through both online and offline travel agents and tour operators
7) To get the most bang for your buck you should book at least 175 days in advance, make the booking directly with the hotel and also join any loyalty programme.