

Hotel Booking Analysis

Ajit Kumar Toppo

Data Science trainee,Almabetter

Abstract: Hotel industry generates a lot of data about the booking information which can be analyzed. Patterns available in the past data are important to help the hotels plan better.

Our aim is to perform exploratory data analysis on a data set which contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.

Keywords: Exploratory data analysis (EDA), Python, Pandas, Seaborn, NumPy, Matplotlib

Introduction

In the hotel industry the bookings depend on a variety of factors such as type of hotels, seasonality, days of week and many more. This makes analyzing the patterns available in the past data more important to help the hotels plan better for the future. Using the historical data, hotels

can perform various campaigns to boost the business. We will be using the data available to analyze and try to find answers to some questions.

About the dataset

The dataset has 1,19,390 rows that contains information about the booking information about the city and resort hotel. It contains 32 columns which are

- **Hotel:** hotel type(H1=Resort hotel or H2=City hotel).
- **is_canceled:** Value indicating if the booking was canceled (1) or not (0).
- **lead_time:** Number of days that elapsed between the entering date of the booking and the arrival date.
- **arrival_date_year:** Year of arrival date.
- **arrival_date_month:** Month of arrival date.
- **arrival_date_week_number:** Week number of year for arrival date
- **arrival_date_day_of_month:** Day of arrival date
- **stays_in_weekend_nights:** Number of weekend nights

(Saturday or Sunday) the guest stayed or booked to stay at the hotel.

- **stays_in_week_nights:**

Number of weeknights (Monday to Friday) the guest stayed or booked to stay at the hotel.

- **adults:** Number of adults (an adult is a person older than 19 years of age).

- **children:** Number of children (whereas the word 'child' refers to any child from three months to 12 years of age).

- **babies:** Number of babies (The word 'baby' means any child from birth to age 4).

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

- 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:** Country of origin.

- **market_segment:** Market segment designation. In categories, the term

- 1. “TA” - “Travel Agents” (A travel agent will help you arrange

transportation, find lodgings and suggest an itinerary)

- 2. “TO” - “Tour Operators”. (A tour operator will arrange your transportation, lodgings, food, entertainment, and an itinerary.)

- 3. 'Direct' - "direct book by customer"

- 4. 'Corporate' - "Helping business travelers find the best stays by accessing an unrivaled inventory of competitive hotel deals across any location"

- 5. 'Online TA' - "online travel agent"

- 6. 'Offline TA/too' - 'offline travel agent, offline tour operator'

- 7. 'Complementary' - A Complementary good is a product or service that adds value to another

- 8. 'Groups' - 'huge mass'

- 9. 'Undefined' -

- 10. 'Aviation' - An Aviation Hospitality professional is someone who works at airlines, airports, or other businesses within the aviation or aerospace industry, such as aircraft manufacturers where they take care of

- the needs of the consumer

- **distribution_channel:** 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:** Number of previous bookings that were canceled by the customer before the current booking

- **previous_bookings_not_canceled:** Number of previous bookings not canceled by the customer before the current booking.

- **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. Code is presented instead of designation for anonymity reasons.

- **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.

- **deposit_type:** Indication if the customer deposited to guarantee the booking. This variable can assume three categories:

1. No Deposit – no deposit was made;

2. Non-Refund – a deposit was made in the value of the total stay cost;

3. Refundable – a deposit was made with a value under the total cost of the stay.

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

- **company:** ID of the company/entity that made the booking or is responsible for paying the booking, ID is presented instead of designation for anonymity reasons.

- **days on the waiting list:** Number of days the booking was on the waiting list before it was confirmed to the customer.

- **customer_type:** Type of booking, assuming one of four categories:

1. Contract - when the booking has an allotment or other type of contract associated with it.

2. Group - when the booking is associated with a group.

3. Transient - when the booking is not part of a group or contract

and is not associated with another transient booking.

4. Transient-party -when the booking is transient, but is associated with at least another transient booking.

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

- **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 (e.g. twin bed or high floor).

- **reservation_status:** Reservation is the last status, 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 - the 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 canceled or when the customer checked out of the hotel.

Data Cleaning

There were 4 missing values in the children column, 488 in the country column, 16,340 in the agent column, 1,12,593 in the company column. Even though there are many missing values in columns 'agent' and 'company' we will not delete the entire column and replace the null values with zero because in case we want to find out ID of the travel agency or ID of the company/entity that made the booking or is responsible for paying the booking these columns can give some information about them. Columns like 'children' and 'countries' have few missing values. We can afford to delete a few rows of missing values as we have a large dataset. There are free rows in our dataset which has booking records of rooms where there are no adults, babies or children. Such rows were removed from the dataset. The datatype of columns like 'children', 'company', 'agent' was converted to integer from float. There were also many outliers present which were removed from the dataset.

Exploratory Data Analysis

After the cleaning of data some analysis was performed on the data to understand the scenario and try to find answers to some questions. Which are given here.

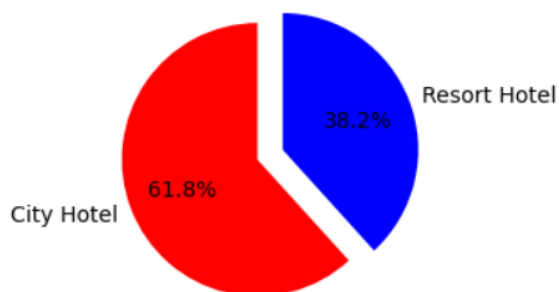
1.How much of the bookings done get cancelled?

After the cleaning of data we were left with 1,18,727 rows which had information about the bookings done out of which 44,137 bookings were cancelled. For further analysis the remaining 74,590 data were used.

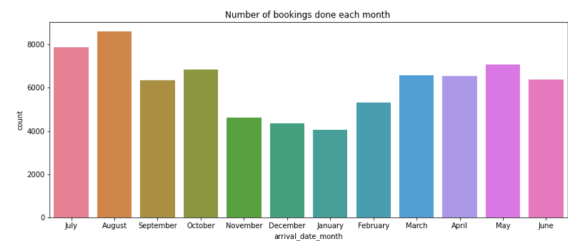
2. Which type of hotel has a higher number of bookings City Hotel or Resort Hotel?

Out of 74,590 bookings data available 46,082 bookings were done for city hotel and 28,508 bookings were done for resort hotel.

Percentage of booking in City Hotel vs Resort Hotel



3. In which month the maximum number of bookings are done?



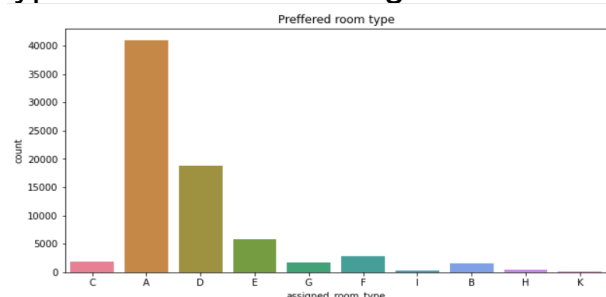
Number of bookings done each month are as follow
(August-8604) (July-7880)
(May-7091) (October-6854)
(March-6566) (April-6528)
(June-6384) (September-6360)
(February-5304) (November-4611)
(December-4347) (January-4061).

4. From which country the maximum number of bookings were done?

Top five countries from which maximum number of bookings were done are as listed below

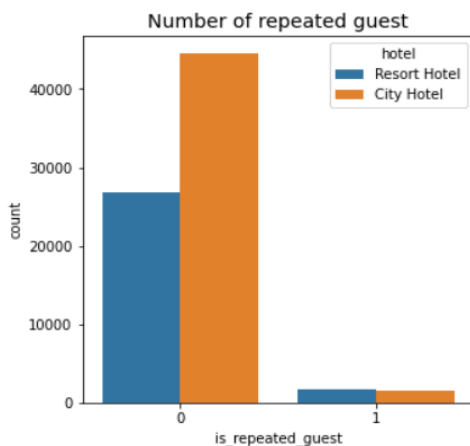
- 1.Portugal-20977
- 2.UnitedKingdom-9668
- 3.France-8468
- 4.Spain-6383
- 5.Germany-6067

5. Which is the most preferred type of room for booking?



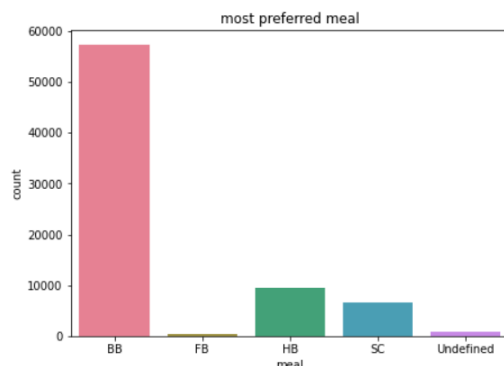
People prefer specific rooms more than others. Data was analyzed to find out which rooms were more preferred for the booking, number of bookings preferred for each room are (A-40930) (D-18796) (E-5774) (F-2806) (C-1905) (G-1755) (B-1643) (H-457) (I-348) (K-176).

6. How many guests are repeated customers?



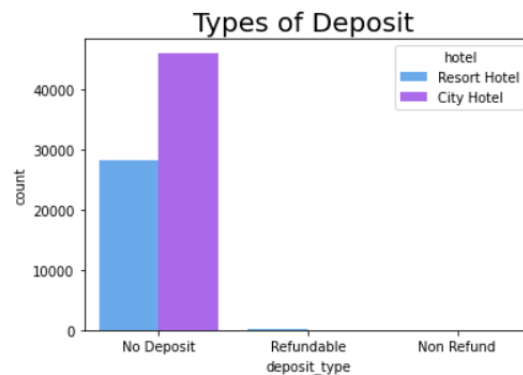
Out of 74,590 bookings data available only 3,201 were repeated customers.

7. Which type of meal is most preferred by the customers?



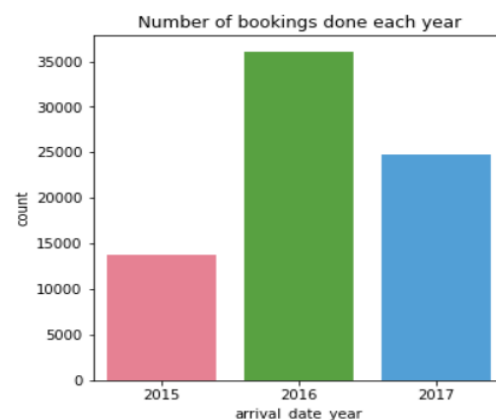
Most people prefer BB which is bed and breakfast, very few prefer FB which is full board (breakfast, lunch & dinner).

8. Which type of deposit customers prefer?



Most people are not giving advanced deposits to ensure booking. There are very few people who give some amount of money for ensuring advanced booking.

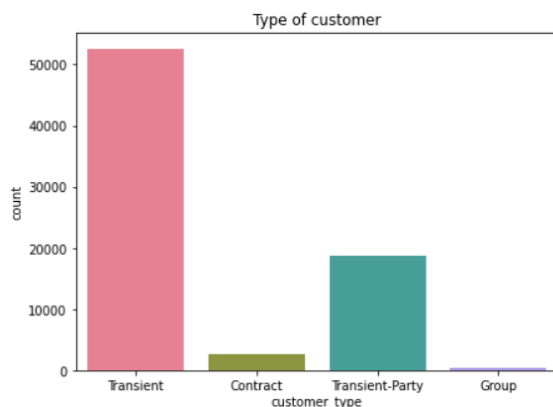
9. In which year most of the bookings were done?



Out of the total booking record available 36,051 booking records

are available from the year 2016. Also 24,814 data are there from 2017 and 13,725 booking data are available to us for the year 2015.

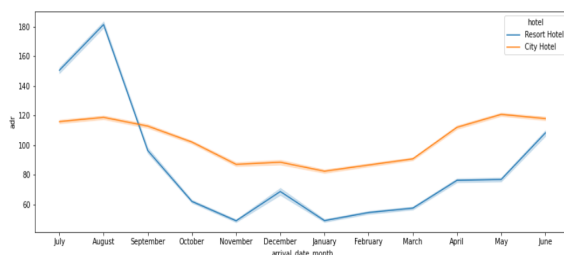
Q10. Which type of customers are most likely to visit?



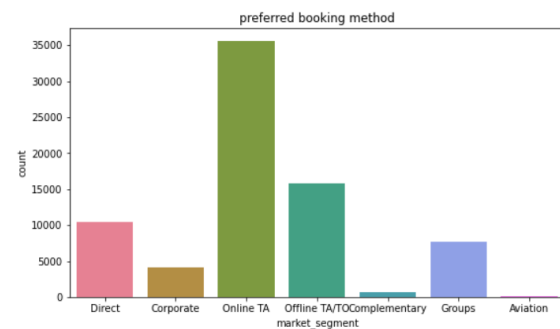
Customers are divided into different types: transient, transient-party, contract and group. From the past booking records available to us 52597 bookings are from transient type, 18673 are from transient-party type, 2810 are from contract type, 510 are from group type.

Q11. How does the ADR vary over the year for both the hotels?

The variation of ADR for both the hotel in different months are shown below

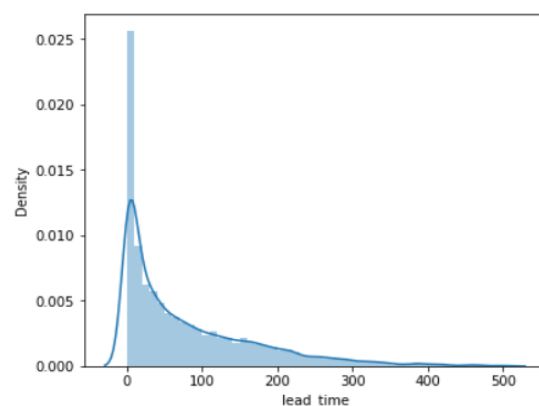


Q12. Which is the most preferred method for the booking?



Out of many methods available for booking 35599 bookings were done by online TA, 15854 bookings were done by offline TA/TO, 10,504 were done by direct method, 7692 were done by group, 4121 were done by corporate, 637 were done by complementary and 183 were done by Aviation.

Q13. Before how many days most of the bookings are done?



We see that there is a positive skewness in the lead time. Most of the bookings done are planned for a close time but still we can see

few percent of people who do planning long before their arrival.

Conclusion

>> We observed that the city hotel has a higher number of booking requests and it also has a higher number of cancellation rates than the resort hotel.

>> 61.8% of bookings are done for the city hotel and 38.2% of bookings are done for the resort hotel.

>>maximum bookings are done in month of July & August while minimum number of bookings are done in the months of January and December.

>>Top countries from where bookings are most done : Portugal,UnitedKingdom,France,S pain,Germany.

>>Most of the bookings done are not by one time customers.

>>People prefer specific rooms more than others, room A&D are the most preferred.

>>Most people prefer BB which is bed and breakfast, very few prefer FB which is full board(breakfast,lunch & dinner).

>>most people do not prefer

giving advanced deposits to ensure booking. There are very few people who give some amount of money for ensuring advanced booking.

>>maximum number of bookings records is of the year 2016.

>>Transient party customers are most likely to visit.

>>Majority of the bookings are done by Online TA and Offline TA/TO.

>>Most of the bookings done are planned for a close time but still we can see few percent of people who do planning long before their arrival.

References

- 1.Rpubs
- 2.Kaggle
- 3.Towardsdatascience
- 4.Medium
- 5.Analytics vidhya