**Hotel Booking Analysis – EDA**

**Biswajit Gochhayat**

**Data science trainee,**

**Alma Better, Bangalore**

**Abstract:**

For any hotel, it is important to know how much booking the hotel is going to get in particular months. Knowing it partially helps the hotel to maintain resources so that the best service can be provided with optimal use of resources.

It is also essential for the hotel to know what type of booking it gets and which type of booking it should look for so that it can make more profit. How much deposit should be taken so that the people will avoid cancelation or even though the canceled hotel booking makes some money?

This dataset will help look into the parameters which make customers cancel their bookings.

**Keywords:** EDA, Seaborn, Hotel Booking, Customer Retention, Hotel profitability

**Problem Statement:**

Look at the data and study the booking pattern by month so that next month's resources can be managed prior.

Look for the cancelation rate and see if there is any pattern and what methods can be used to reduce the cancelation

**About Dataset:**

* **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 hot part ora or contract transient, but is associated with at least another transient booking

* **ADRdr**: 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 was the booking cancelled or when the customer checked out of the hotel.

**Steps Involved in EDA:**

1. **Null Value Treatment:** The country and children column was having a few missing values that were replaced with the mode of the column. The company and agent columns had significant data missing so both columns were dropped from further analysis
2. **Column-wise data analysis**: Each column was analysed andConclusions were made from the EDA

**Observations:**

1. **Hotels:** Out of all bookings 66.4 % of bookings are for the city hotel and 33.6. % of bookings are for the resort hotel. The city hotel also has the highest cancelation compared to the resort hotel. The city hotel has cancelations of 41% on the other hand resort hotel has fewer cancelations.
2. **Repeated Customers:** Overall repeated guests are very few compared to new guests Resort Hotel has 4.5 percent reaped guests while the city has only 2.7 percent repeated guests. The resort also has significantly lower cancellations compared to the city hotel. Cancellation for new customers is almost the same for both Resort and City hotel
3. **Deposit Type**: We can see that most of the bookings are for the No deposit category. The cancellation is significantly higher for the non-refundable category. Refundable has very few bookings so it won't be fair to compare. It is very contrary to whatshould ideally happen but the good thing is the hotel is making a few bucks from these bookings. Probably hotel has asystem that checks the probabilityof cancellation and asks for a deposit.
4. **Previous Cancellation:** Repeated customers with no previous cancellation have a very less percentage of booking cancellations. People who have cancelled once have more cancellations. The ones who have cancelled twice have small cancellations but the numbers are small so it can't be trusted and seems counterintuitive. Compared with one cancellation overall people with more cancellations have a higher chance of cancelling the booking.
5. **Meal:** Most people book their room with Bead Breakfast costlier items can be added to the menu
6. **Distribution Channel:** Travel agents are responsible for bringing most of the bookings, accounting for 82.5 % of the total booking The second and Third highest booking comes from direct bookings and corporate bookings.
7. **Lead Time:** Bookings with more than 150 days of lead time mostly get cancelled
8. **Room Type:** Most of the bookings are for low-cost rooms. Most high-value Customers choose the resort hotel while low-value customers choose the city hotel.

**Summary and Suggestions:**

* City Hotel gets higher booking than Resort hotel and the cancelation rate is also higher of Resort Hotel.
* Resort Hotels get the highest values customers.
* In 2016 the booking ratio is more than 15% from all other years.
* The City hotel has more guests during spring and autumn when the prices are also highest, In July and August, there are fewer visitors, although prices are lower. (Thus, customers can get a good deal on bookings in July and August in a city hotel).
* Guest numbers for the Resort hotel go down slightly from June to September, which is also when the prices are highest. (Thus, these months should be avoided for bookings).
* Broadly, April to August is the peak season for bookings. Both hotels have thefewest guests during the winter.
* Weekended are more crowded than weekday
* Online TA, Groups make the most booking and cancellation of hotels as compared to other market segment categories.
* A and D type room is more booked by user.
* A higher lead time has a higher chance of cancellation.
* A history of previous cancellations increasesthe chances of cancellation

**Suggestion:**

* Resort hotels tend to have fewer bookings in comparison to city hotels so they need to work on their marketing strategy and promote the hotels more, especially on social media.

1. Resort hotels could also reduce prices to increase booking percentages.
2. Set Non-refundable Rates, Collect deposits, and implement more rigid cancellation policies.
3. Encourage Direct bookings by offering special discounts
4. May-August happens to be the busiest month, so the hotels should target more customers and try to do more business.

* Although city hotels have more bookings, they also tend to have more cancellations so to prevent this they could take advance money during vacation. This would ensure most bookings to not cancelled. They could also apply no-refund policies or make the refund policies rather strict so the customers choose not to cancel.

1. The majority of the hotels booked are city hotels. Need to spend the most targeting fund on that hotel.
2. We also realize that the high rate of cancellations can be due to high no deposit policies

* It is quite clear most customers travel in pairs and bringing children or babies along is very rare so the hotels could advertise in ways that attract couples more and also business travellers.
* Most guests do not return but as these customers have already visited once, advertisements should be targeted in such ways so they are bound to return the next time they visit. The customers could also be offered special benefits if they do return to stay.
* Majority of the guests are from Western Europe. We should spend a significant amount of our budget on that area.

**References:**

* Alma batter
* Google
* Blogs