Hotel booking EDA Technical document

CONTRIBUTION: Lavleen Prabhakar Dekatey (Individual)

ABSTRACT

Hotel is an alternative when people are not resident of that particular location and have no other source of accommodation. So people either reserve hotel online or offline. Here comes the part of hotels to provide accomodation to customer in exchange of money with ease of access to booking through various types of booking. Hotels should make sure that customer are satisfied with their service along with making their side safe to sustain as a business. So hotels should manage their bookings efficiently.

In this project we are trying to have an insight of reasons of cancellation and ways to reduce booking cancellation. We will gain insight of which country is a challenge for sustaining the hotel business, which customers are loyal, whether lead time and waiting time are reason for cancellation of booking, whether reserved room type customer assigned other room type is reason for cancellation and which year had the maximum booking cancellation.

Prime objective of our project is to generate revenue through reduction of booking cancellation, developing efficiency in providing best services to customer and cost of fetching the data through different resources should be reasonable to business.

Problem Statement

There is approximately 37% cancellation during 3 years period whose reason can become reason for falling ratings of the hotels in the long run. So we need to find the reason for cancellation of booking.

We are addressing the following question:

1) Which country has the maximum number of reservation cancellation, in which type of hotel and are which type of customer?
2) Which customer type is mostly a repeated guest and they are from which distribution channel or market segment?

3) Is cancellation of reservation related to lead time that is does longer period of lead-time have more cancellation?
4) The customer who pay for deposit or not needed to pay for deposit do maximum cancellation?
5) Is reserved room type customer provided with assigned room type a reason for reservation cancellation?
6) Whether the customer with maximum booking changes have the cancellation of reservation?
7) Are the customer who wait for reservation are prone to cancel the booking?
8) During which year cancellation were maximum?

Introduction

As a hotel business owner we will want to grab more and more customers but also look after the cancellation of booking which gives a loss if we are not able to fill the vacant room available on time. So In this project we are trying to get the insights of what all factors are responsible for booking cancellation. With the help of given data we will be making few charts to get an idea of factor (variable) name which play an important role in cancellation of bookings.

Types of hotel

There are two types of hotel:

1) City hotel
2) Resort hotel

1) A city hotel is what you probably know best. It provides accommodation and meals to travellers. Often times, people come from all over the world to stay at a hotel so that they can tour around the place that they are staying. They do not spend their day in the hotel and they only come back to sleep and shower, maybe a meal or two. In general, hotels despite having the amenities and luxuries similar to these of resorts are more a place to stay and not a destination.

2) A resort is the destination. Oftentimes, resorts are located in scenic areas but not near a city or anything really. People visit a resort to get away from other people and cities. They stay in the resort and probably don't travel around as they may do when staying at a city hotel.

Hotel cancellation

We found that cancellation of city hotel is maximum in Portugal country which is as follows in different customer type:

Contract   1105.0

Group   29.0

Transient   14446.0

Transient-Party   4501.0

Steps involved

1) EDA:-

 EDA is primarily used to see what data can reveal beyond the formal modelling or hypothesis testing task and provides a provides a better understanding of data set variables and the relationships between them. It can also help determine if the statistical techniques you are considering for data analysis are appropriate.

In our project we tried to have insights about cancellation of booking.

2) Null value treatment:-

It is necessary to fill in missing data values in data sets as most of the machine learning models that you want to use will provide an error if you pass NaN values into it. The easiest way is to just fill them up with 0. This is what we have done while grouping our data.

3) Categorizing numbers:-

We are categorising lead time into a, b and c whose values are 0, from 1 to 104 and above 104. Which helped us having inferences through categories of number.

4) Percentages:-

We calculated percentages in the following way by dividing booking cancelling customer value by total customer in the category:

a)

a_percentage=(hotel_df_a_canceled/lead_time_a)*100

b_percentage=(hotel_df_b_canceled/lead_time_b)*100

c_percentage=(hotel_df_c_canceled/lead_time_c)*100

b)

deposit_type_nodeposit_percentage=(no_deposit_hotel_canceled/no_deposit_hotel)*100

deposit_type_nonrefund_percentage=(nonrefund_hotel_canceled/nonrefund_hotel)*100

deposit_type_refundable_percentage=(refundable_hotel_canceled/refundable_hotel)*100

Conclusion

We should shape our business in such a way that it should be able to sustain losses due to reservation cancellation in coming year such as it was in 2016, have competent and satisfactory rooms to attract more customer by increasing our reviews positively. Manage lead time and booking changes, so as to keep control of the number of cancellation. Losses should not be only seen from business side it should also be seen from customer side like the condition of the room and services we provide so as to bring loyal customers in the long run.