**Hotel Booking Analysis**

**Snehal D. Ramteke**

**Data science trainee,**

**AlmaBetter, Bangalore**

* **Abstract:**

The hotel industry is one of the most important components of the wider service industry, catering to customers who require overnight accommodation. It is closely associated with the travel industry and the hospitality industry, although there are notable differences in scope.

In this project, we provided the hotel booking dataset for the year 2015-2017. This dataset describes two datasets i.e hotels & resorts. Both datasets comprehend bookings due to arrive between the 1st of July 2015 and the 31st of august 2017, including bookings that effectively arrived and bookings that were canceled. We are provided different columns in this dataset by which we tried to analyze the booking prediction.

Our project can help us understand who is taking the majority of reservations, the number of repeat guests, which country the hotel was booked and so many questions.

**Keywords**- Python for data science.

* **Problem Statement**

We are here to explore a hotel booking dataset to discover important factors that govern the bookings which will help us identify major loopholes & give us insights that will be helpful

To run a profitable hotel business are as follows:

1. Finding the correlation between the numerical data.
2. Which are the months of highest and least occupation?
3. What is the most popular meal package?
4. How many booking changes have been done during the studied period?
5. How many people have been registered in the hotel?
6. What is the most common customer type?

* **Data Description:**
* **hotel** : Hotel (Resort Hotel or 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 into the PMS 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
* **children**: Number of children
* **babies**: Number of babies
* **meal**: Type of meal booked. Categories are presented in standard hospitality meal packages:
  + Undefined/SC – no meal package
  + BB – Bed & Breakfast
  + HB – Half board (breakfast and one other meal – usually dinner)
  + FB – Full board (breakfast, lunch, and dinner)
  + and dinner)
* **country**: Country of origin. Categories are represented in the ISO 3155–3:2013 format
* **market\_segment**: Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”
* **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 prior to the current booking
* **previous\_bookings\_not\_canceled**: Number of previous bookings not canceled by the customer prior to 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 made a deposit to guarantee the booking. This variable can assume three categories:
  + No Deposit – no deposit was made
  + Non Refund \* a deposit was made in the value of the total stay cost
  + 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\_in\_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:
  + Contract - when the booking has an allotment or other type of contract associated with it
  + Group – when the booking is associated with a group
  + Transient – when the booking is not part of a group or contract and is not associated to another transient booking
  + 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 the last status, assuming one of three categories:
  + Canceled – booking was canceled by the customer
  + Check-Out – customer has checked in but already departed
  + No-Show – the customer did not check in and did inform the hotel of the reason why
* **reservation\_status\_date**: The 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 canceled or when the customer checked out of the hotel

.

* **Introduction:**

In this project, we have a dataset of hotels. This dataset contains booking information for a city hotel and a resort hotel and includes information such as when the booking was made. The number of adults, children, length of stay, available parking spaces & other things.

The problem that we faced was that we had large numbers of data from which we had to extract those exact numbers.

**Types of hotels used**:

* Resort Hotel
* City Hotel
* **Most three common and uncommon countries of origin:**

As we can see, Portugal tops the list with 48,586 cases, followed by Great Britain with 12,129 and France with 10,415.

The most uncommon country of origin is Madagascar, Dominica, and Mali

* **The Highest least occupation:**

The month of highest occupation is august with 11.65% of the reservations. The months of least

Occupation is January with 4.94% of the reservations.

* **The most popular meal package:**

The bread and breakfast option is the most popular, with a frequency of 77.26%.

* **The most reserved room type:**

The A room type is the most popular among the clients, with 71.99% of the reservations.

* **How Many Bookings changes done during the studied period:**

There are 25,829 registered changes in bookings during this period.

* **How many people registered in the hotel:**

2,33,934 people registered in the hotel.

* **The most common customer type:**

Transients are the most common customer type, they represent 75% of the total customers.

* **Steps involved in this project:**

**Step 1 -** In the first step, we had different types of datasets in which there were different columns. Firstly we understand data and the collection of data and then we work on data cleaning and manipulation & lastly we work on EDA(Exploratory Data Analysis).

**Step 2** - In this second step, we designed our PowerPoint slides using google presentation.

**Step 3** - In the last step, we prepared the presentation video for the project. In this part of the step, we discussed different results through the videos and audio with PowerPoint slides.

* **Conclusion:**

1. City hotels are the most preferred hotel type by guests. We can say the City hotel is the busiest hotel**.**
2. In this exercise we see if there are any missing values present or not.
3. the majority of reservations are for city hotels.
4. The majority of guests come from western Europe countries.
5. Here adr(average daily rate) is correlated with no\_of\_people. It simply indicates that more revenue is generated when the number of people increases.
6. The number of repeated guests is too low.
7. Waiting time period for City hotels is high as compared to resort hotels. That means city hotels are much busier than Resort hotels.
8. The month of highest occupation is august with 11.65% of the reservations. The months of least
9. The majority of reservations convert into successful transactions.
10. Resort hotels have the most repeated guests.