**Hotel Booking Analysis**

**Jatin, Abhilasha Manjunath, Krishan Kumar Bafna**

**Data science trainees,**

**Alma Better, Bangalore.**

**Abstract:**

The Data Article describes two datasets with hotel demand data. One of the hotels is City Hotel and the other hotel is Resort Hotel. Both datasets share the same structure, with 31 variables describing the 40060 observations for Resort Hotel and 79330 observations for City Hotel.

Each observation representations of hotel bookings. Both datasets comprehend booking in between the 1st July 2015 and 31st August 2017, including booking that effectively arrived and booking that were cancelled. Since this is hotel real data, all the elements for identification of customer were deleted.

**1.Problem Statement**

Have you ever wondered when the best time of the 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?

**2. Introduction**

### This dataset contains booking information for City Hotel and Resort Hotel, includes when the booking was made, length of the stay, the number of adults, children babies and the number of available parking spaces among other things like reservation status, average daily rates previous cancellation status, deposit type, whether the hotel is providing meals or not. This dataset has 119390 numbers of rows and 31 columns.

## **3. Understanding the variables**

* **is\_canceled** - Whether the booking is canceled or not (0 for no canceled and 1 for canceled)
* **lead\_time** - time (in days) between booking transaction and actual arrival.
* **arrival\_date\_year** - Year of arrival
* **arrival\_date\_month** - month of arrival
* **arrival\_date\_week\_number** - week number of arrival date.
* **arrival\_date\_day\_of\_month** - Day of month of arrival date
* **stays\_in\_weekend\_nights** - No. of weekend nights spent in a hotel
* **stays\_in\_week\_nights** - No. of weeknights spent in a hotel
* **adults** - No. of adults in single booking record.
* **children** - No. of children in single booking record.
* **babies** - No. of babies in single booking record.
* **meal** - Type of meal chosen
* **country** - Country of origin of customers (as mentioned by them)
* **market\_segment** - What segment via booking was made and for what purpose.
* **distribution\_channel** - Via which medium booking was made.
* **is\_repeated\_guest** - Whether the customer has made any booking before(0 for No and 1 for Yes)
* **previous\_cancellations** - No. of previous canceled bookings.
* **previous\_bookings\_not\_canceled** - No. of previous non-canceled bookings.
* **reserved\_room\_type** - Room type reserved by a customer.
* **assigned\_room\_type** - Room type assigned to the customer.
* **booking\_changes** - No. of booking changes done by customers
* **deposit\_type** - Type of deposit at the time of making a booking (No deposit/ Refundable/ No refund)
* **agent** - Id of agent for booking
* **company** - Id of the company making a booking.
* **days\_in\_waiting\_list** - No. of days on waiting list.
* **customer\_type** - Type of customer(Transient, Group, etc.)
* **adr** - Average Daily rate.
* **required\_car\_parking\_spaces** - No. of car parking asked in booking
* **total\_of\_special\_requests** - total no. of special request.
* **reservation\_status** - Whether a customer has checked out or canceled,or not showed.
* **reservation\_status\_date** - Date of making reservation status.

## **4. Python library we used.**

* Numpy
* Pandas
* Matplotlib
* Seaborn
* Klib

# **5. Graphs use for Data Visualization.**

* Bar Plot
* Count Plot
* Scatter Plot
* Heat Map
* Line Plot

**6. Steps involved:**

* **Exploratory Data Analysis**

Exploratory Data Analysis stands for **EDA. ‘EDA’** is applied to **investigate** data and summarize the key **insights**. Here we are understanding the data distributions, some of the values by manipulating data and much more by using **Python functions.**

* **Null values Treatment**

Our dataset contains a large number of null values which might tend to disturb our accuracy hence we filled them at the beginning of our project by putting zero and ‘x’ in order to get a better result.

* **Standardization of features**

Our main motive through this step was to scale our data into a uniform format that would allow us to utilize the data in a better way while performing fitting and applying different algorithms to it.

The basic goal was to enforce a level of consistency or uniformity to certain practices or operations within the selected environment.

**7. Type of Bookings:**

Type of booking assuming one of four categories:

* Contractor - When the booking has an allotment or associated with type of contractor.
* Group – When the booking has associated to any group.
* Transient – When the booking is not a part of any group or contractor and it is not associated to others transient booking.
* Transient-Party – When the booking associated with transient, but it is associated with other transient booking.

.

**8. Approach we used:**

The approach we have used in this

project is defined in the given format-

1) **Loading our data**: In this section we just loaded our dataset in colab

notebook and read the csv file.

2) **Data Cleaning and Processing**: In this section we have tried to remove the null values and for some of the columns we have replaced the null values with the appropriate values with reasonable assumptions.

3) **Analysis and Visualization:** In this section we have tried to explore all variables which can play an important role for the analysis. In the next parts we have tried to explore the effect of one over the other. In the next part we tried to answers our hypothetical questions.

4) **Future scope of further analysis:** There are the room types ‘A’, ‘D’ and ‘E’ which are in top three demanding rooms so, hotel can make rooms like these three types. Also, in the month of August, July and May have maximum number of chances that the guests will definitely come on vacations and also for there other purposes accordingly and many more.

**9. Conclusion:**

That's it! We reached the end of our exercise.

Starting with loading the data so far, we have done EDA, null values treatment, encoding of categorical columns, feature selection and then model building.

* **Most of the customers come from Portugal, Great Britain, France and Spain.**
* **Almost 41% of City Hotel and 27% of Resort Hotel were cancelled. So, we can clearly see that maximum percentage of people cancelled City Hotel.**
* **August Month has the highest bookings in both of the hotels. So over all the years August month is the busiest month for both of the hotels.**
* **Repeated customers is approx. 3.2% for both of the hotels.**
* **City hotel has significantly longer waiting time , hence City Hotel is much busier than Resort Hotel.**
* **There are 46228 Customers who booked City hotel and not cancelled it further also there are 28938 Customers who booked Resort hotel and not cancelled it further.**
* **Maximum waiting day period for Resort Hotel is 185 days and for City Hotel is 379. So, we can easily say that city hotel is busier than Resort Hotel.**
* **The data we can see that August, July and May are the busiest months for Resort hotel and City hotel.**
* **The above data shows the average rent for the Resort hotel for all the 12 months and the average rent is maximum for the month of August and minimum for the month of January.**
* **Also the average rent for the Resort hotel for all the 12 months and the average rent is maximum for the month of August and minimum for the month of January.**
* **For Contract, Transient and Group type customers for Resort Hotel the waiting period is zero whereas for Transient-Party it is 3.12 days So it is clear that if the booking was made as 'Contract' or 'Group' or 'Transient' the rooms will be booked immediately without waiting time.**
* **For Contract, Transient and Group type customers for Resort Hotel the waiting period is zero whereas for Transient-Party it is 7.88 days So it is clear that if the booking was made as 'Contract' or 'Group' or 'Transient' the rooms will be booked immediately without waiting time.**
* **The average rent is maximum for 'Transient' type of customers and minimum for 'Group' type of customers in Resort Hotel.**
* **The maximum three type of rooms that are booked the most. It is seen that room type A has maximum number of bookings and room type E stands in third place**
* **The types of room booked and the rent of the rooms are related. It is observed that the most common rooms booked named room types 'A', 'D' and 'E' have minimal or moderate room rents whereas rarely booked rooms have high room rent.**