**Hotel Booking Analysis**

**Prajakta Dangale**

**Data science trainee,**

**AlmaBetter, Bangalore**

**Abstract:**

Travel and hospitality is one of the most lucrative industries which always have something new to offer to guests. Year-on-year the industry witnesses a rise in travelers globally. Running a hotel business isn’t a cakewalk. Numerous hotels around the world tend to overlook most of the challenges. As a result, they face serious repercussions. Analyzing the data in day today world is very important and nowadays hotel business cannot be run without some sensible and smart use of data.

To overcome these challenges first we analyzed the data and visualized and made proper decisions to help hotel industry. This data article describes the booking information for a city hotel and a resort hotel and includes information such as when the booking was made, length of stay, lead time, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.

So, these information of data gives us idea to analyze and to perform exploratory data analysis using different python packages and libraries. And we came across the solutions for various challenges facing by hoteliers. Data Visualization plays an important role to describe the huge data in a single frame.

***Keywords: packages and libraries, exploratory data analysis, data visualization***

**1.Problem Statement**

A hotel system manages information about rooms, reservations, customers, and customer billing. Hotel industry facing to analyze the problems like change in marketing trends and dynamics, housekeeping issues, customers’ expectations, Data security, mode of bookings etc...

**Context**

Have you ever wondered when the best time of 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 a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions!

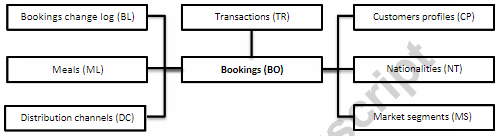
So, the main objective is to build a predictive model, which could make hoteliers to improve their revenue.

* lead \_time:number of days between the time a guest books their room and the time they are scheduled to arrive at the hotel.
* Distribution \_channel: Method or platform by which customer can book the rooms.
* Adr: average daily rate which measures the average revenue hotel receives.

Adr = Room revenue

Number of rooms sold

* matplotlib: Used for creating powerful visualizations
* seaborn: It is a library making statistical graphics.
* hue: Parameter used for coloring the column
* explode: slices and drags the particular position of pie chart.



**2. Introduction**

Travel and Tourism is one of the world‘s most rapidly growing industries. Much of its growth is due to higher disposable incomes, increased leisure time and falling costs of travel. With the invention of rail transport and air transport besides road and sea transport, hotel industry developed from in to present day hotels, it is the part of the hospitality industry.

We can call this industry as a volatile industry and here, bookings depend on various factors like, type of hotels (city, resort), seasonabilty, days and nights stays and many more. This makes analyzing the patterns available in past data more important to help the hotels to improve their business in a better way. Using the historical data, hoteliers can perform campaigns to build up their business. We can use the given data set to predict the future bookings using pandas dataframe techniques.

We will be using the data set available to analyze the factors affecting the hotel bookings. These factors can be used for reporting the trends and predict the future bookings.

### Our goal here is to build a predictive model, which could help hoteliers to improve their business proactively.

## **3. Steps Involved:**

## **3.1 Summary of Data:**

## The dataset has a shape of (119390, 32) which means that it contains 119390 rows and 32 columns. This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, booking countries, number of people who stayed in weekend nights and week nights, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.

**3.2 Data Wrangling:**

First we search for duplicate rows in the dataset and remove it from the dataset. Then find missing value in the dataset. Our dataset contains a large number of null values which might tend to disturb our accuracy This data file consist of null values ‘Nan’ in the columns such as agent, company and children. Replaced those null values with zero. For the missing value in the country column replace it with mode(value that appears most often). Data file consist of different types of data types such as integer, float, object then convert the float data columns into integer data columns (i.e. children, company, agent).

**4.Methodology**

1. Majority of the analysis was EDA which was digging one level deeper and getting the data to answer the above questions
2. To answer few questions, I sliced the data across various cuts.
3. Another methodology included using different processes to view data across segments.
4. I used different plots to visualize the data and found some solutions.

**5.Insights**

* The number of bookings seems to be high in 2016 while the bookings seem to be less in 2015 and 2017. This is majorly due to more data points being available in 2016
* The ratio of bookings cancelled to the confirmed bookings seems high for City Hotels
* We observe that the avg. of bookings in a week is growing year over year
* The growth is higher from 2015 to 2016 as compared to the jump from 2016 to 2017
* The growth is higher from 2015 to 2016 as compared to the jump from 2016 to 2017
* The revenue almost doubled in 2016 but it did not grow at the same rate from 2016 to 2017
* We observe that the number of New Customers and Existing Customers increases drastically in 2016 and the rise is not that high in 2017
* We observe that the New Customers tend to spend significantly lower than the Existing Customers
* Majority of bookings are accompanied with a breakfast or dinner
* There are few Full Board bookings. This gives an opportunity to the hotels to increase revenue from these type of meals by offering some discounts
* The hotels need to have more of Type A room followed by other types
* We can investigate the Type L rooms to conclude about the 4 bookings in these rooms
* We observe that Weekday bookings are higher than the Weekend numbers. That is an interesting finding. We need to investigate our corporate bookings!

**6.Implication to business**

* The above analysis has unraveled some key findings in terms of the hotel type, customer type, customer preferences and time preference
* Based on the results of EDA, hotel can plan on targeting the new customers to increase spends and maintain good relations with the existing customers
* The hotels have an insight on the meal preference and room preference which will help them price the commodities better
* Some plots depict that the growth from 2015 to 2016 was higher as compared to the following year. We can analyze the differences across years to identify the extraordinary performance and match it
* The revenue forecast will help them plan promotions during the low revenue period.

**7.Limitations**

* The data for 2015 and 2017 is for different months. Even though we have converted them to same base line of weekly numbers, there are chances that some weeks perform differently as compared to other weeks
* The definition of new customers is not very well described. A new customer this year will be existing next year, or they can be existing customer from the 2nd booking. A deeper analysis in required based on definition
* The weekday vs weekend analysis can be further drilled for the type of bookings
* The classification model uses few variables. We can tune the model with new variables and adjusting the cost of misclassification.

**8. EDA:**

Analytics in the hotelier world today is important, and nowadays this business cannot be run with some sensible and smart use of data.

Here we demonstrated how to use data to analyze three business important concepts in the fields of revenue management and marketing.

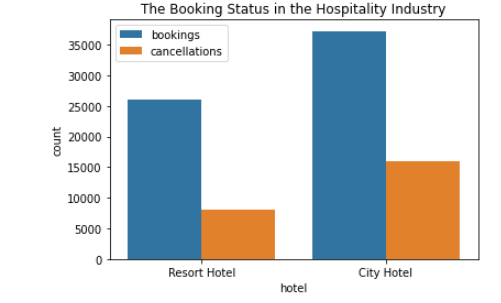
**8.1Countries from which customers visited**

Chart

Description automatically generated

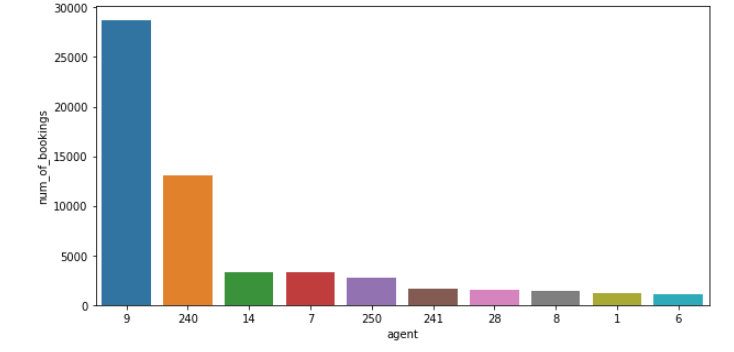
From this analysis, we can observe that country 'PRT' that is Portugal made huge number of bookings as compared to other countries.

# **8.2 Number of booking and cancellation:**



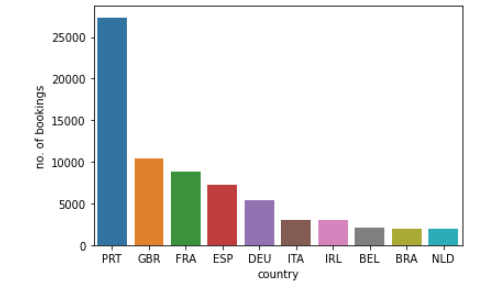
Resort hotel has maximum cancellation than city hotel.

# **8.4 most no. of bookings:**



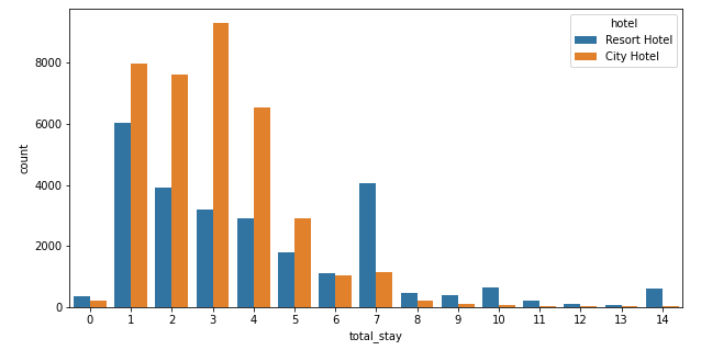
Agent no. 9 has made most no. of bookings

# **8.5 most of the customers visit these hotels**



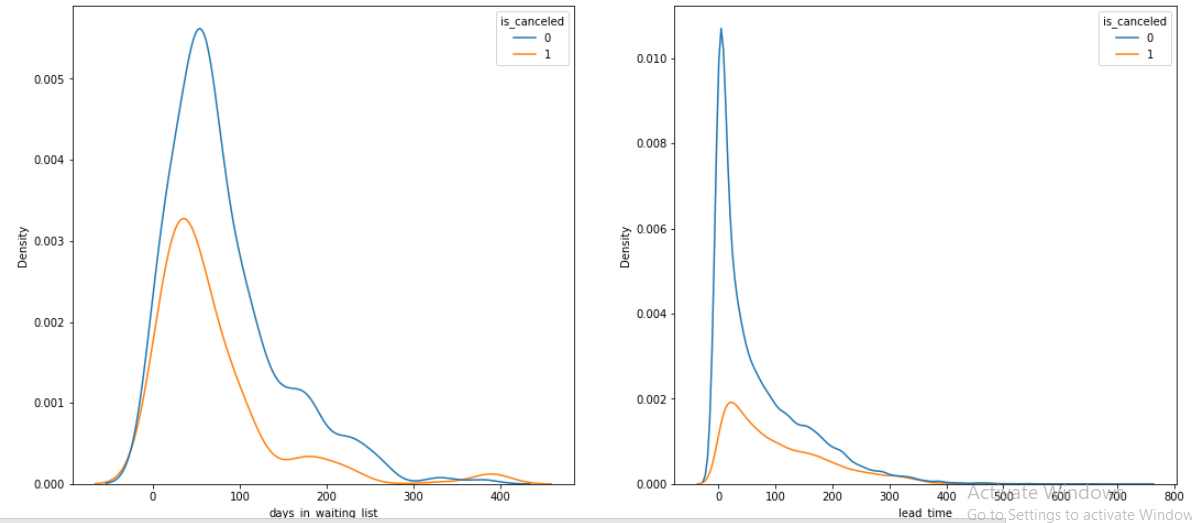
Most of the customers come from Portugal, Great Britain, France and Spain.

# **8.6 preferred stay in each hotel**



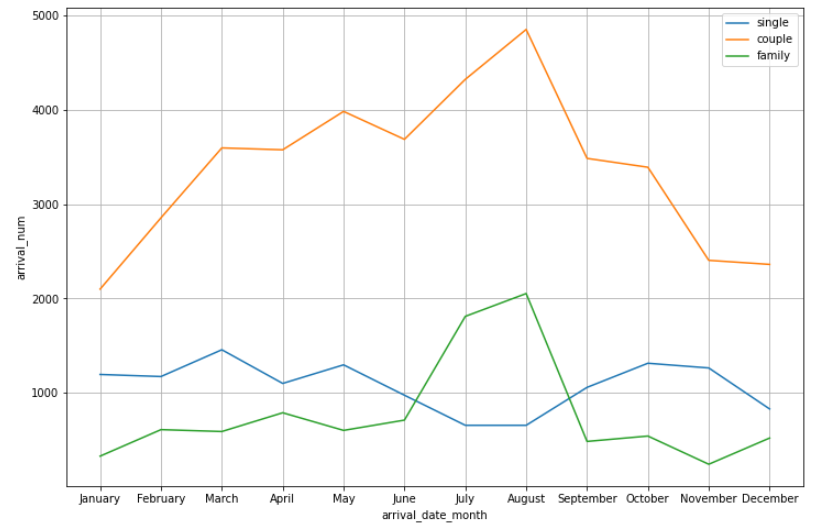
Most common stay length is less than 4 days and generally people prefer City hotel for short stay, but for long stays, Resort Hotel is preferred.

# **8.7 Causes the cancelation of bookings of rooms by customers**



We see that most of the bookings that are cancelled have waiting period of less 150 days but also most of bookings that are not cancelled also have waiting period less than 150 days. Hence this shows that waiting period has no effect on cancellation of bookings. \ \ Also, lead time has no affect on cancellation of bookings, as both curves of cancelation and not cancelation are similar for lead time too.

# **8.7 Divide our customers in three categories of single, couple and family/friends**



Mostly bookings are done by couples(although we are not sure that they are couple as data doesn't tell about that)

It is clear from graph that their is a sudden surge in arrival num of couples and family in months of July and August. So better plans can be planned accordingly at that time for these type of customers.

**9.Conclusion:**

Main aim of the project was to understand and visualize dataset from hotel and customer point of view and give suggestions to reduce these cancellations and increase revenue of hotels. So, from the analysis we can conclude that,

* Out of customers, most of the customers visited hotels from the Portugal country.
* Resort hotel has maximum cancellation than city hotel.
* Agent no. 9 has made most no. of bookings
* Most of the customers come from Portugal, Great Britain, France and Spain.
* Most common stay length is less than 4 days and generally people prefer City hotel for short stay, but for long stays, Resort Hotel is preferred.
* We see that most of the bookings that are cancelled have waiting period of less 150 days but also most of bookings that are not cancelled also have waiting period less than 150 days. Hence this shows that waiting period has no effect on cancellation of bookings. \ \ Also, lead time has no affect on cancellation of bookings, as both curves of cancelation and not cancelation are similar for lead time too.
* Mostly bookings are done by couples(although we are not sure that they are couple as data doesn't tell about that)
* It is clear from graph that their is a sudden surge in arrival num of couples and family in months of July and August. So better plans can be planned accordingly at that time for these type of customers.

From these data analysis, hotel management can work on the areas of improvement to increase the revenue.

**10.References-**

1. Greeksforgeeks
2. European Commission. (2012). Digital Agenda for Europa. Available at: http://ec.europa.eu/digital-agenda/.

3.. Gerbert, P., Schneider, D.