

Hotel Booking Analysis

Aishwarya Ashok Kande

Alma Better, Bangalore

Abstract:

Hotels plays a major role in tourism industry. We have hotel booking dataset on which we can perform analysis to get insights which will be useful for customer. This dataset demonstrates the hotel booking pattern of basically two types of hotels, 'City hotel' & 'Resort hotel'. Initially, the dataset had 32 variables describing 119390 observations. Each & every observation in the dataset represents a hotel booking. This dataset comprehends the booking data between 2015 & 2017. Since, this is a real hotel booking data, all the data elements pertaining customer as well as hotel identification were deleted. This dataset can definitely play an important role for research and analysis on revenue management, data mining as well as in other fields. In order to achieve this, we can use data visualization method with several datasets and predict the possibility of the best outcome for the customer to be get satisfied.

1. Problem Statement

Data of different hotels (i.e., excluding the personal information) like booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of

stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things are gathered. 'Hotel industry' is a very volatile industry & the bookings depend on a variety of factors. This makes the analysing patterns available in the dataset which in turn would help the hotels plan better. Different factors affecting the booking pattern, is used to report the trends

This approach builds a customer friendly platform in order to find the best results and solve the multiple problems, The data includes.

1. Hotel : Indicates types of Hotels

- City hotel ● Resort type

2. Is_canceled : Indicates the cancellation of the hotel booking

- Cancellation = 1 ● No Cancellation = 0

3. Lead_time : Time (in days) between booking transaction and actual arrival.

4. Arrival_date_year : Year of arrival

5. Arrival_date_month : Month of arrival

6. Arrival_date_week_number : Week number of arrival date

7. Arrival_date_day_of_month : Day of month of arrival date

8. Stays_in_weekend_nights : No. of weekend nights stayed in a hotel

9. Stays_in_week_nights : No. of weeknights stayed in a hotel

10. Adults : No. of adults in a single booking record.

11. Children : No. of children in a single booking record.

12. Babies : No. of babies in single booking record.

13. Meal : Type of meal chosen

- BB:- bed and breakfast
- HB:-Half board (Breakfast and dinner)
- FB:- Full Board (All meals included)
- SC:- Self catering (No meals Included)

14. Country : Country of origin.

15. Market_segment : Market segment for booking

- Aviation
- Complimentary
- Corporate
- Direct
- Groups
- Online (TA)
- Offline (TA/TO)

16. Distribution_channel : Via which medium booking

- Corporate
- Direct
- GDS: - Global Distribution System
- TA/TO: - Travel Agent/Operator

17. Is_repeated_guest :

- 0 for new customer
- 1 for repeated customer

18. Previous_cancellations : No. of previous canceled bookings.

19. Previous_bookings_not_canceled :
No. of previous non-canceled bookings.

20. Reserved_room_type : Room type
reserved by a customer.

21. Assigned_room_type : Room type
assigned to the customer.

22. Booking_changes : No. of booking
changes done by customers.

23. Deposit_type : Type of deposit at the
time of making a booking

- No deposit
- Refundable
- No refund

24. Agent : Id of agent for booking

25. Company : Id of the company
making a booking

26. Days_in_waiting_list : No. of days in
waiting to book

27. Customer_type : Type of customer

- Contract: - bookings done by the
contract
- Group: - Group booking
- Transient: - Customer staying for
shorter period
- Transient-Party: - Group of customers
staying for a shorter period

28. ADR : Average Daily rate of hotels.

29. Required_car_parking_spaces : No.
of car parking preferred by customers at
the time of booking

30. Total_of_special_requests : Total no. of special request.

31. Reservation_status :

- checked out
- canceled
- not showed

a) Canceled – booking was canceled by the customer

b) Check-Out – customer has checked in but already departed

c) No-Show – customer did not check-in and did inform the hotel of the reason why

32. Reservation_status_date : Date of making reservation status of the hotel

2. Introduction

Hotel Booking Analysis: This is a hotel booking Analysis Project. There are two types of hotel in this data set, which are city hotel and resort hotel. Major details on this data set was reservation of hotel, cancelation, ADR of the hotel, data on the basis of month etc... Shape of the data is 119390 rows and 32 columns.

Some challenges, which faced during the project developing is data was complex, many missing value are missing etc... So, after cleaning the data, it is started to visualize its pattern to understand what is on the dataset.

Libraries which are used in this project is NumPy, pandas, matplotlib, seaborn etc... After completing data cleaning, it is started to visualize the data.

First, it deployed how much number of hotels are there in the given data. Then made analysis on how much hotel are confirmed and how many hotels have been canceled. Then made analysis on which countries people was the most visitor and which are least visitor. Then come across what is the average daily rate with respect to month and years, Project end with having a correlation graph and conclusion.

3. Steps Involved:

1. Framing the questions :

It's all about the points on which the analysis is based upon.

2. Exploratory Data Analysis :

First step is to import libraries such as NumPy, pandas, matplotlib, seaborn. Then load the raw dataset. This data has many unprocessed values which cannot be considered for the study. Here is the workflow of correcting it for our analysis.

3. Data Cleaning :

Handling Null Values

- Company Id and Agent Id: - These columns have null values of 93% and 15% respectively. Hence, these columns are dropped.
- Country: - This has null values less than 5% thus the null values are filled with the mode value.
- Children and

babies: - There are only 4 null values so the null value is filled with median

4. **Data Manipulation:** Creating new columns

- Total night stays =
 $\text{stays_in_weekend_nights} + \text{stays_in_week_nights}$
- Total members =
 $\text{Adults} + \text{childrens} + \text{babies}$

5. **Data Visualization :** Data visualization is the practice of translating information into a visual context, such as a map or graph, to make it easier to understand and gain insights from them.

The graphs used here for study are: -

- Histogram.
- Pie Chart.
- Bar Plot.
- Line Plot.
- Scatter Plot.

6. **Conclusion :** It's about summing up the whole process and stating the final results.

Challenges faced:

- Data cleaning threw a major challenge as data needs to get cleaned for accurate results but in a way such that the data does not lose its meaning.

- Choosing between the type of visualization to be used and also making a choice between matplotlib & seaborn.

Conclusion:

1. City Hotel seems to be more preferred among travellers and it also generates more revenue & profit.
2. Dataset contains booking data of 3 different years(2017,2016,2015), out of which, maximum hotel bookings took place in 2016 & 2015 witnessed the least number of hotel bookings.
3. Out of all months, 'August' witnessed highest number of hotel bookings whereas 'January' witnessed the least.
4. Most number of bookings are made from Portugal & Great Britain.
5. Its observed that 'City hotels' was more cancelled as compared to 'Resort hotels'.
6. As checked BB Meal type has high preference where as FB Meal Type has very low preference
7. We have found that, In 2015 highest ADR is in august month which is 122 and lowest ADR is in November month. For 2016 , highest ADR is in august month which is 148 and lowest ADR is in January month. For 2017 , highest ADR is in august month which is 168 and lowest ADR is in January month.

8. We found that if guest's stay days is getting decreased, ADR is getting high.

9. We have found that, almost 80000 customers dont want parking slot for car and as checked only 2 customers wants 8 parking slots. So we can consider 1 parking slot for each customer is fine.

10. Highest corelation value between axis is 17% positive which is in between adr and special requests so they are directly correspondent to each other.