



**Project Title:** **Analyse the hotel booking data**

Dataset: https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand

Description of individual fields

1. Hotel: Name of the hotel (Resort Hotel or City Hotel)
2. Is \_cancelled: If the booking was cancelled (1) or not (0)
3. Lead\_time: Number of days before the actual arrival of the guests.
4. Arrival\_date\_year: Year of arrival date
5. Arrival\_date\_month: Month of month arrival date
6. Arrival\_date\_week\_number: week number of year for arrival date
7. Arrival\_date\_day\_of\_month: day of arrival date
8. Stays\_in\_weekend\_nights: Number of weekend nights (Saturday or Sunday) spent
9. Stays\_in\_week\_nights: Number of weeknights (Monday to Friday) spent at the hotel.
10. Adults: Number of adults among guests
11. Children: Number of children among guests
12. Babies: Number of babies among guests
13. Meal: Type of meal booked
14. Country: Country of guests
15. Market\_segment: Designation of market segment
16. Distribution\_channel: Name of booking distribution channel
17. Is\_repeated\_guest: if the booking was from a repeated guest (1) or not (0)
18. Previous\_cancellations: Number of previous bookings that were cancelled by the customer to the current booking
19. Previous\_booking\_not\_cancelled : Number of previous bookings not cancelled by the customer prior to the current booking
20. Reserved\_room\_type: Code of room type reserved
21. Assigned\_room\_type: Code of room type assigned
22. Booking\_changes: Number of changes made to the booking
23. Deposit\_type: Type of the deposit made by the guest
24. Agent: ID of travel agent who made the booking
25. Company: ID of the company that made the booking
26. Days\_in\_waiting\_list : Number of days the booking was in the waiting list
27. Customer\_type: Type of customer assuming one of four categories
28. Adr: Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights
29. Required\_car\_parking\_spaces: Number of car parking spaces required by the customer
30. Total\_of\_special\_requests: Number of special requests made by the customer
31. Reservation\_status: Reservation status (Cancelled, Check out or No Show)
32. Reservation\_status\_date: Date at which the last reservation status was updated

Since downloaded data is cleaned one, we tried to introduce few incorrect records to the existing Kaggle data for perform cleanup.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Re$0rt Hotel | 0 | 12 | 2015 | July | 27 | 1 | 0 | 1 | 2 | 0 | 0 | BB | IRL | Offline TA/TO |  | TA/TO |
| C! TY Hotel | 0 | 37 | 2015 | July | 27 | 1 | 0 | 4 | 2 | 1 | 0 | HB | PRT | Offline TA/TO |  | TA/TO |
| Resort Hotel | 0 | 85 | 2015 | NULL | 25 | NULL | NULL | 1 | 2 | 1 | 0 | HB | USA | Online TA |  | TA/TO |
| City Hotel | 0 | 12 | 2015 | July | 20 | 1 | 0 | 1 | 2 | 1 | 0 | HB | NULL | Offline TA/TO |  | TA/TO |
| Resort Hotel | 0 | 85 | 2015 | July | 27 | 1 | 0 | 4 | 2 | 1 | 0 | NULL | IRL | Online TA |  | TA/TO |
| Resort Hotel | 0 | 75 | NULL | July | 20 | 1 | 0 | 4 | 2 | 1 | 0 | HB | PRT | Online TA |  | TA/TO |

Data Cleaning:

* Removed filed “company” as it had 93% NULL values.
* 13% Agent filed value had NULL values, replaced it with mean value.
* Removed two records where we had special characters (hotel – Re$0rt hotel and C!TY hotel)
* We have cleaned rows that has NULL values for below fields:

arrival\_date\_month

country

meal

arrival\_date\_year

* We have created two more extra fields.

1. Total\_duration = stays\_in\_weekend\_nights+ stays\_in\_week\_nights
2. Revenue = adr\* Total\_duration (Revenue indicates total revenue as per reserved date)

|  |  |
| --- | --- |
|  |  |

Graphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generated

Questionaries:

1 Which hotel produces the most revenue and the elements that influence it. **SYEDA**

Differentiate between the hotel types for cities and resorts.

Parking space and distribution method are all factors.

2 which is most popular? Booking distribution channel (compare the concept of market segment) MARIYA

3 Which market sector experiences the highest rate of cancellations of reservation? what are the causes of the increased number of cancellations of reservations? **AMRIT**

4 What is the ideal length of stay for guests at each type of hotel? comparisons between the weekends and weekdays **DEVI**

5 Which of the top five Countries had the most frequent visitors to both hotels? **MARIYA**

6 In Which month most of the hotel booking done? **DEVI**

7. Which hotel type has more lead time? **SYEDA**

Compare the ratio of reservation status and waiting list.