**Hotel Booking Analysis**

**Akash raj**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

This data article describes two datasets with hotel demand data.

One of the hotels (H1) is a resort hotel and the other is a city hotel (H2).

Both datasets share the same structure, with 31 variables describing the 40060 observations

of H1 and 79330 observations of H2.

Each observation represents a hotel booking.

Both datasets comprehend bookings due to arrive between the 1st of July of 2015 and the 31st of August 2017,

Including bookings that effectively arrived and bookings that were canceled. Since this is hotel real data,

all data elements pertaining hotel or costumer

identification were deleted. Due to the scarcity of real business data for scientific and educational purposes,

these datasets can have an important role

for research and education in revenue management, machine learning, or data mining, as well as in other fields

**1.Problem Statement**

**Through this analysis, we were able to answer key business questions which are stated below:**

* 1) Which type of hotel is mostly prefered by the guests?
* 2) What is the percentage of cancellation?
* 3) What is the Percentage of repeated guests?
* 4) What is the percentage distribution of "Customer Type"?
* 5)what is the percentage distribution of required\_car\_parking\_spaces?
* 6)What is Percentage distribution of Deposit type?
* 7) Which type of food is mostly preferred by the guests?
* 8) From which country the most guests are coming?
* 9) Which is the most preferred room type by the customers?
* 10)In which month most of the bookings happened?
* 11) Which Distribution channel is mostly used for hotel bookings?
* 12) Which year had the highest bookings?

**2. Introduction**

This data set contains 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.

**Goal:** The goal of this notebook is to practice EDA and figure out the standard patterns of booking.

**3. Steps involved:**

* **Loading the data sets : - Hotel booking**
* Importing Essential Libraries : - NumPy, Pandas, Seaborn and Matplotlib
* Data Cleaning: - Null values, Finding and removing Outliers, Removing duplicate data.
* EDA (Exploratory Data Analysis) : - Analyzing the data sets to  summarize their main characteristics using statistical graphics  and data visualizations method.

**Data Collection and Understanding .**

1. hotel :- Hotel(Resort Hotel or City Hotel)

2. is\_canceled :-Value indicating if the booking was canceled (1) or not (0)

3. lead\_time :- Number of days that elapsed between the entering date of the booking into the PMS and the arrival date

4. arrival\_date\_year :- Year of arrival date

5. arrival\_date\_month :- Month of 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) the guest stayed or booked to stay at the hotel

9. stays\_in\_week\_nights :- Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel

10. adults :- Number of adults

11. children :- Number of children

12. babies :- Number of babies

13. meal :- Type of meal booked. Categories are presented in standard hospitality meal packages:\*

14. country :- Country of origin.`

15. market\_segment : -Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”

16. distribution\_channel :- Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”

17. is\_repeated\_guest : -Value indicating if the booking name was from a repeated guest (1) or not (0)\*

18. previous\_cancellations :- Number of previous bookings that were cancelled by the customer prior to the current booking\*

19. previous\_bookings\_not\_canceled :-Number of previous bookings not cancelled by the customer prior to the current booking\*

20. reserved\_room\_type :- Code of room type reserved. Code is presented instead of designation for anonymity reasons.\*

21. \*\*assigned\_room\_type\*\* : \*Code for the type of room assigned to the booking.\*

22. \*\*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\*

23. \*\*deposit\_type\*\* : \*Indication on if the customer made a deposit to guarantee the booking.\*

24. \*\*agent\*\* : \*ID of the travel agency that made the booking\*

25. \*\*company\*\* : \*ID of the company/entity that made the booking or responsible for paying the booking.\*

26. \*\*days\_in\_waiting\_list\*\* : \*Number of days the booking was in the waiting list before it was confirmed to the customer\*

27. \*\*customer\_type\*\* : \*Type of booking, 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 (e.g. twin bed or high floor)\*

31. \*\*reservation\_status\*\* : \*Reservation 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 – customer did not check-in and did inform the hotel of the reason why

32. \*\*reservation\_status\_date\*\* : \*Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus

to understand when was the booking canceled or when did the customer checked-out of the hotel\*

**References-**

1. GeeksforGeeks
2. stack overflow