**Hotel Booking Analysis Summery**

|  |
| --- |
| Name: Bhupathi Sampath,  Email: samastha849@gmail.com |
| Github Link:- https://github.com/BhupathiSampath/Hotel-booking-analysis |
| **Problem Statement:**   * Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay 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!​ * 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. All personally identifying information has been removed from the data.​ * Explore and analyse the data to discover important factors that govern the bookings.   **Approaches:**   * Cleaning the dataset by removing duplicated data and filling the missing and null values.​ * Adding the new columns by performing feature engineering.​ * Data sub setting using python pandas library.​ * Visualizing the subsets of data from the sub setting step by using matplotlib and seaborn libraries.   **Conclusions:**   * City hotels are most preferred hotels. Thus, City hotels are busiest than Resorts.​ * Transient customer type is more percentage of booking which is **82.4%** and Group type is low which is **0.6%.**​ * Most of the guests are preferring the rooms "A"(Code of room type). So, Code "A" type rooms can be increased to increase the bookings.​ * 27.5% bookings has been cancelled.​ * Distribution channel "TA/TO" has more cancellations with 89.13%.​ * Less than 150 days waiting list has cancelled and there is not cancelled bookings also more in less than 150 days waiting list. Hence, we can say days in waiting list is not much affecting the cancelation.​ * Lead time also not much affecting the cancelation of bookings.​ * Same room allotted and not allotted are almost similar. Hence, not getting same room is affecting the daily "adr". Guests who are not getting the same room are paying the less adr compared to same room allotted.​ * Agent with ID Number: 9 has done more bookings.​ * There is only few guest are repeated which is 3.9%.​ * 91.6% guests did not require the parking space where only 8.3% of guests required only 1 parking space.​ * Distribution channel "TA/TO" has done more bookings and used for early bookings.​ * Distribution channels 'Direct' and 'TA/TO' are contributing the most in both types of hotels. GDS distribution channel should focus on increasing the bookings of 'City Hotel'.​ * Market segment "Direct" has the high adr in both Resort and City hotels where "Complementary" has less. Aviation segment can focus on City hotel.​ * Bookings has been increasing till the mid of year and we can see in August bookings went to highest. Hence, Most of the people are planning trips in august month.​ * Average adr rises from beginning of year up to middle of year and reaches peak in August and then lowers to the end of year. But hotels do make some good deals with high adr at end of year also.​ * Mostly bookings are done by couples (although we are not sure that they are couple as data doesn't talk about that).​ * It is clear from graph that there is a sudden surge in arrival number 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. |