**Hotel Booking Analysis**

**Priyanka K,**

**Data Science Trainee,**

**AlmaBetter**

**Abstract**

This study explores the factors that govern the bookings of hotels. The dataset contains various information such as how long the guest stayed, deposit type used during the booking, market segment of the guest, etc., The dataset has 119390 observations and 32 features describing each booking. The dataset has booking data for City hotel and Resort Hotel over three years, from 2015 through 2017.

Exploratory data analysis helps in finding whether different factors have any influence over hotel bookings.

***Keywords: EDA, Data analysis, hotel booking, revenue generation, ADR***

**1. Problem Statement**

Exploratory Data Analysis is a practicable tool for a better understanding of data and its effects on other aspects of the business. With higher competition in the hotel industry, Hotel Booking Analysis would be helpful to get insights into the factors that govern bookings.

To do an Exploratory Data Analysis on this hotel data, we use python to get insights that would increase the booking, ways to reduce cancellations, and areas to focus on more.

In this project, we do analyse and visualize different attributes that contribute to booking and cancelation, areas to concentrate more, which segment brings in more business, etc.,

**2.Business Objective**

The main objective is to generate more revenue for the hotel and increase bookings.

**3.Data Summary**

There are 32 different features for a single observation in the dataset. The different features are

* **Hotel:** Whether the booking is for City hotel or Resort hotel
* **Is\_canceled: Is** the booking canceled or not
* **Lead\_time:**The number of days elapsed between the booking and the arrival date
* **Arrival\_date\_year:** Year of arrival date
* **Arrival\_date\_month:** Month of the arrival date
* **Arrival\_date\_week\_number:** The week number for which the guest is going to visit.
* **Arrival\_date\_day\_of\_month:** Day of the arrival date
* **Stays\_in\_weekend\_nights:** Number of weekend night stay
* **Stays\_in\_week\_nights:** Number of weekday night stay
* **Adults:** Number of adults
* **Children:** Number of children
* **Babies:** Number of babies
* **Meal:** Type of meal preferred
* Country: Country code of the guest
* **Market\_segment:** The market segment of the booking
* **Distribution\_channel:** By which market segment customer access the stay
* **Is\_repeated\_guest:** Whether the guest stays for the first time or not
* **Previous\_cancellations:** Are there any previous cancellations
* **Previous\_bookings\_not\_canceled:** Count of the prior bookings canceled
* **Reserved\_room\_type:** Room type preferred by the guest
* **Assigned\_room\_type:** Assigned room for the guest
* **Booking\_changes:** Count of changes made to the booking
* **Deposit\_type:** Deposit type opted for the booking
* Agent: Agent data for the booking
* **Company:** Company to which the guest belongs
* **Days\_in\_waiting\_list:** Number of days on the waiting list
* **Customer\_type:** Customer type to which the booking belongs
* **ADR:** Revenue generated by the hotel through this booking
* **Required\_car\_parking\_spaces:** Is car parking is required
* **Total\_of\_special\_requests:** Number of special requests by the guest.
* **Reservation\_status**: Reservation status of the booking
* **Reservation\_status\_date:** Date of the reservation status for the booking

**4.Steps Involved**

* **Framing the right questions**

Framing the right questions is one of the most crucial parts of any analysis. The right question drives us to obtain a solution to the problem statement, which in turn helps in attaining the business objective.

* **Data Cleaning**

Cleaning of data involves dealing with missing values and duplicates in the dataset. Removing duplicates and missing value treatment provides more accurate results for the problem. Our Hotel Dataset has 31994 duplicates and missing values in three columns. The duplicates are removed and missing values are replaced with more appropriate values.

* **Exploratory Data Analysis**

Explanatory Data Analysis helps in a better understanding of the dataset, such as the relationship between variables, identifying the variable important for the problem statement, a feature that influences another variable in the dataset, etc.,

The different approaches adopted for EDA are univariate analysis, bivariate analysis, and multivariate analysis.

* **Univariate analysis** - Analyzing a single variable in a dataset is Univariate analysis.
* **Bivariate analysis** - Bivariate analysis is slightly more analytical than Univariate analysis. When the data set contains two variables, to know the relationship between the two variables, bivariate analysis guides us in exploring the relationship.
* **Multivariate analysis** - When there is a relationship between more than two variables, multivariate analysis comes into play.

Exploratory Data Analysis gives us a better picture of the dataset and helps in finding the solution or corrective measures for the problem.

* **Visualizing the insights**

Visualization brings down even complex analysis data into an easily understandable visual representation. Visuals highlight the trends and outliers in the analysis output. A good visualization tells a story, removing the noise from data and highlighting information. Python has libraries such as matplotlib, seaborn, plotly, etc., for visualizing data.

* **Driving conclusions**

The final step of any analysis is to find an actionable insight from the given data to attain the business objective. The conclusion should answer all the questions framed during the start of the Exploratory Data Analysis.

**5.Findings from Analysis**

* The number of bookings is high for city hotels compared to Resort hotels.
* The percentage of bookings canceled is 27.5%, and 72.5% remains unchanged.
* Transient type of customers makes up the majority of the bookings.
* TA/TO, Direct, and Corporate is the distribution channel for the hotel.
* Maximum bookings are with no deposit type, which is one of the factors for cancellation.
* The most preferred room in both the hotel is a type A room.
* The most preferred meal is BB meal.
* The maximum number of special requests is 5. The majority of the bookings don't have any special requests.
* European countries like Portugal, the United Kingdom, France, and Spain tend to make more bookings.
* The cancellations show consistent growth year on year alongside the increase in bookings for the hotel.
* The highest revenue generated is in August compared to other months.
* The market segment with a higher booking is Online TA, but also with a higher number of cancellations compared to other market segments.
* Staying less than ten days has more bookings and generates higher revenue.
* Days on the waiting list is not having a relationship with cancellation.

**6.Solutions to Business Objective**

* The number of bookings and adr is high in August. We can promote bookings in other months with exciting offers while the bookings are low.
* The number of booking cancellations is due to the no-deposit option in the deposit type. Implement deposit rate for bookings will reduce cancellations.
* The contribution of marketing channels is low other than the online channel. Develop a marketing strategy to promote bookings.
* Most of the bookings are through online channels, so effective advertisement and online presence drive more bookings.

**7.Conclusion**

That is a wrap. We reached the end of the project. In this project, we loaded the data into python and cleaned the dataset by removing duplicates and dealing with missing values. Then we did an exploratory analysis of the data to find some key factors that govern hotel booking. With the help of visualization, we unraveled the pattern and trends in the dataset. Finally, we have drawn some actionable insights to improve hotel bookings and generate more revenue.

**References**

* + Towards Data Science
  + Geeks For Geeks
  + Analytics Vidhya