TECHNICAL DOCUMENTATION
ON

Hotel Booking Analysis

AI maBetter

SUBMITTED BY: AImabetter Data science Trainee

Thursday, October 6, 2022

Prepared by: Sayesh Ankaram (Cohort Canberra)

Email: sayeshsayesh@gmail.com

ABSTRACT

Various hotel management Businesses often store the data happening inside their systems such as Date, Time, Customer bookings, Origin of Country, Room status, Accommodation status, Most busy months, Least busy months, preferred Meal types etc, and they analyze this data to obtain various patterns by considering past data to predict futuristic data. They aim towards processing & exploring the data in a way that it will be beneficial for their future endeavors and will assist them in understanding market behaviour.

A deep Statistical analysis on these datasets will tend to improve the market value of these organizations.

INTRODUCTION

The Hotel Booking dataset consists of data related to Customer Bookings with respect to their preferences, Country of origin, Booked or canceled, Reservation status, Type of Hotel, Mode of booking, Car Parking, Date-Month-Year of Booking, Type of rooms, Accommodation types and so on.

In short, what this dataset does is, it provides deeper insights towards the hotel data and how things go around in hotel management systems ranging from the preference of the customers to the checked-out status.

This Hotel booking data describes two data sets containing hotel demand information. The first hotel is a Resort hotel and the second hotel is a City hotel. Both data sets are structured similarly with 32 variables describing the 40,060 Resort hotel observations and 79,330 City hotel observations. Each observation represents a hotel reservation. Both data sets include bookings scheduled to arrive between July 1st, 2015 and August 31st, 2017, including bookings that actually arrived and bookings that were canceled.

Moreover This dataset contains booking information for City hotels & Resort hotels. It also includes information such as when the booking was made, length of stay, number of adults, children, babies and the types of rooms, meals and much more.

This hotel dataset can help to answer so many questions like best time of the year to book a hotel room, optimal length of stay in order to get the best daily rates, predicting whether or not a hotel is likely to receive huge amounts of booking requests etc.

This data set contains a single file which compares various booking information between two hotels: City hotel and Resort hotel, it ranges from year 2015 to 2017 with confirmed bookings: 2015 – (21996 bookings), 2016 – (56707 bookings) & 2017 – (40687 bookings).

PROBLEM STATEMENT

The hotel industry is highly volatile, with bookings influenced by a variety of factors such as hotel type, seasonality, day of week, and many others. This emphasizes the importance of analyzing patterns in historical data to assist hotels in better planning. Hotels can use historical data to run various campaigns to boost business. We will use the available data to analyze the factors influencing hotel bookings. These variables can be used to report trends and forecast future bookings.

So, after deep diving into the dataset, we'll get an insight of how the hotel system works, we will be able to gather information and explanations on below crucial questions:

- Which country has the most number of visitors?

- Which country has the least number of visitors?

- What is the percentage of bookings that were both confirmed and canceled?

- Which mode of booking was most frequently used?

- What is the hotel price per stay over a year?

- What are the hotel rates paid by customers per night based on different
- room types?

- How long do people stay in the hotel rooms?

- Which meal type is most preferred and booked by customers?

- Which is the most booked type of Accommodation?

- What is the Reservation Status of hotels?

DATA SYNOPSIS

In order to gain deeper insights about the data, statistical and exploratory data analysis are very important and to do that we need to know the types of variables we're going to be working with, therefore this dataset consists of 2 different types of variables namely:

Categorical Variables:

- agent – ID of the travel agency that made the booking.

- arrival_date_month – Month of arrival date.

- assigned_room_type – Type of room assigned to the booking.
- company – ID of the company/entity that made the booking or responsible for paying the booking.
- country – country of origin.
- customer_type – Customer type based on mode of booking.
- deposit_type – Deposit paid by customer.
- distribution_channel – Booking distribution channel.
- is_canceled – cancellation status: if canceled (1) else (0).
- is_repeated_guest – If guest is repeated.
- market_segment – Market segment distribution.
- meal – Type of meal booked.
- reserved_room_type – Type of room reserved for the booking.
- arrival_date_month – Month of arrival date.
- assigned_room_type – Type of room assigned to the booking.
- reservation_status – status of reservation.

Numerical Variables:

- adr – Average daily rate.

- adults – Number of adults.
- arrival_date_day_of_month – Day of month of arrival date.
- arrival_date_week_number – Week number of the arrival date.
- arrival_date_year – Year of arrival date.
- babies – Number of babies.
- booking_changes – Number of changes made to the booking.
- children – Number of children.
- days_in_waiting_list – Number of days the booking was in waiting list.

- lead_time – Number of days booking got elapsed.
- previous_bookings_not_canceled – Number of previous bookings not canceled.
- previous_cancellations – Number of previous bookings canceled.
- required_car_parking_spaces – Number of car parking spaces required.
- stays_in_weekend_nights – Number of stays in weekend nights.

Scientific Approach

●Instance Creation: Creating an instance of the dataset so that the original dataset will not get disturbed.

●Dataset Description: Obtaining a described info about the dataset we're working with (by using .shape(), .info() & .describe() methods to gain better insights about the data).

●Prime Variable Selection: Selecting a prime variable/feature/column with which large scale insights can be obtained (the prime variable is '*is_canceled*' , for obtaining active data columns with bookings that are not canceled).

●Dataset Processing: Cleaning the data by obtaining Null values, Dropping unnecessary rows and columns with Null values, Altering dtypes of columns.

●Categorical Dataframe Creation: Creating a Categorical dataframe having all the categorical columns in it, so that we can perform various operations to get a clear understanding about each and every aspect related to the hotel data itself.

●EDA & Data Visualizations: Performing exploratory data analysis over the cleaned data in order to get final conclusions and plotting the analyzed data using different graphs.

Conclusion

- There are about 63% of confirmed bookings & 37% of canceled bookings.

- For 1-3 days, majority of people prefer City Hotels and for 4-7 days, majority of people prefer Resort Hotels.

- The Highest Rates for a hotel room in both City & Resort hotels is for Room Type A.

- Hotel rates for Resort hotels are at peak in Summer season (May-August) but in Autumn season (September-November) the rates become cheaper.

- Hotel rates for City hotels have a decent growth in Spring season (March-May) but in Autumn season (September-November) the hotel rates slightly fell down.

- 47.56 % of customers prefer Online mode for hotel booking & 21.17 % of customers prefer Offline mode.
- 76.96 % of customers prefer Breakfast & Bed & 12.6 % of customers prefer Breakfast & Dinner as the Meal Type while Booking hotels.

- 77.85 % of customers prefer Family Type of Accommodation & 65.51 % of customers prefer Couple Type of Accommodation.

- 62.92 % of customers Checked-Out, 36.06 % of customers Canceled the Booking & 1.10 % of customers didn't show up.

- Countries with the most number of Guests are Portugal (Guest Count: 20977) followed by Great Britain (Guest Count: 9668) and France (Guest Count: 8468).

- 87.6 % of customers prefer No – deposit while booking hotel rooms.

- 66.4 % of customers prefer City hotels & 33.5 % of customers prefer Resort hotels.

References

Analytics Vidhya

GeeksforGeeks

Matplotlib.Plotly Documentation