

Hotel Booking Analysis

By: Prerna Kashyap

Other members: Chandan Prasad, Meet Dave, Prerna Kashyap, Lalith Kumar PS

Team: Data Driven

Cohort Enlighten

AlmaBetter Capstone Project

Abstract:

Everyone wants to own a successful business and in hospitality business every owner thinks to own a hotel which is running with full capacity and generating high revenue.

Here in this project we have analyzed data related to Hotel Booking and pointed out some factors which may help the owners to increase the booking. The 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. Also there are other important information that we have figured out by analyzing this data.

Through our analysis we have tried our best to provide all possible conclusions which may help the hotelier to increase their revenue. There are various possible reasons of getting less revenue for a particular hotel viz. Hotel facilities, Hotel location, Service, Rate and others.

1. Problem Statement

The dataset contains booking information for two different hotels. One being City Hotel and another Resort Hotel along with information on various booking criteria such as booking season, time of booking, length of stay, number of adults, children and babies, parking spaces, pricing data, market segment and many more.

Primary objective is to explore and inspect the dataset; and discover important features using Exploratory Data Analysis that can govern bookings and help hotels penetrate deep into market, thereby attracting more customers. Analyze booking and pricing trends to draw out insights to make intelligent business decision.

Secondary objective is help the customers in deciding best period to visit places while availing low accommodation cost benefits.

The dataset contains following features:

- hotel
- is_canceled
- lead_time
- arrival_date_year
- arrival_date_month
- arrival_date_week_number
- arrival_date_day_of_month
- stays_in_weekend_nights
- stays_in_week_nights
- adults
- children
- babies
- meal
- country
- market_segment
- distribution_channel
- is_repeated_guest
- previous_cancellations
- previous_bookings_not_canceled
- reserved_room_type
- assigned_room_type
- booking_changes
- deposit_type
- agent
- company
- days_in_waiting_list
- customer_type
- adr
- required_car_parking_spaces
- total_of_special_requests
- reservation_status
- reservation_status_date

2. Introduction

Over past couple of decades, travel and tourism industry has seen massive surge

in term of profit. With more people having better lifestyle and resources to travel, Hotel businesses have been growing very rapidly and there is scope for more. Hotel business is lucrative yet very volatile business segment. It's dependent on multiple criteria like type of hotel, business seasons, customer segments, multiple modes of booking, amenities, services and many more. Starting a new Hotel business requires thorough data analysis on Industry, Customers, Competitors and many more. Which is what we are going to do in this project. Thorough Exploratory Data Analysis on guests, hotel type, bookings trend, pricing trend, distribution channels, people stay, etc. to draw out insights and make intelligent business decision.

3. Motivation

To try and answer following questions from data:

- What are the hotel preferences of guests?
- From which country do the majority of guests visit the hotel?
- What is the booking trend of hotels round the year?
- What is the best season to leverage benefits of low cost accommodation?
- Which market segment prospers more compared to others?
- What is the effect of different booking channels on reservation status?

- Is there any correlation between Booking, Pricing, Stay length, Cancellation, Parking and Guest revisiting with each other?
- What is the most preferred stay length?

4. Steps involved:

I. Data exploration and inspection

After importing essential libraries and loading dataset, we explored and inspected the initial raw data. Gained information on size of dataset, each column's Dtype and non-null count. Go through columns data description and unique values in categorical columns.

II. Data Cleaning

Data cleaning involves

- **Handling Null values:** Handled null values in column 'agent' by replacing those with new agent ID 0 assuming those bookings were reserved without any agent; replacing nulls in column 'children' with 0 assuming guests had no children in their group and finally replacing null in column 'country' with 'other' assuming guests belonged to country other than available list.
- **Dropping irrelevant columns and rows:** Dropped irrelevant column 'company' which had 94% null data and rows with 0 guest total.

- **Parsing date:** Parsed required date column in string format to datetime format
- **Adding new feature columns:** 3 new columns were added for new feature analysis namely 'total_stay' and 'total_people'.

Finally, we have cleaned and processed data to work on to gain insights for better business decision.

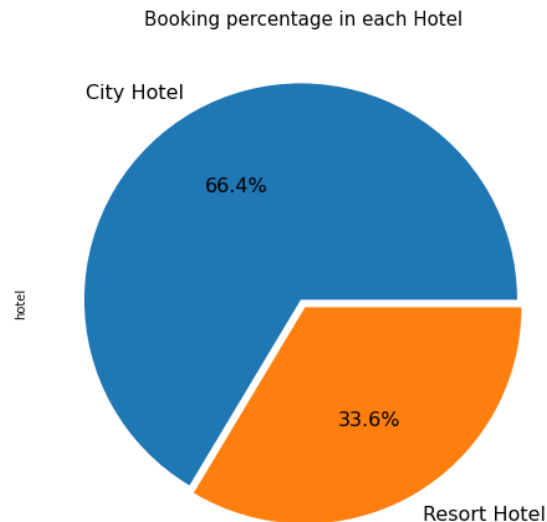
III. Exploratory Data Analysis and Data Visualization

To work on our problem statement, we start our initial analysis on one variable, then move onto relational analysis between two variables and gradually increase our analysis and include multiple variables and establish relation between them.

We will try to address each questions we framed earlier one by one.

a. Booking percentage of different type of Hotels

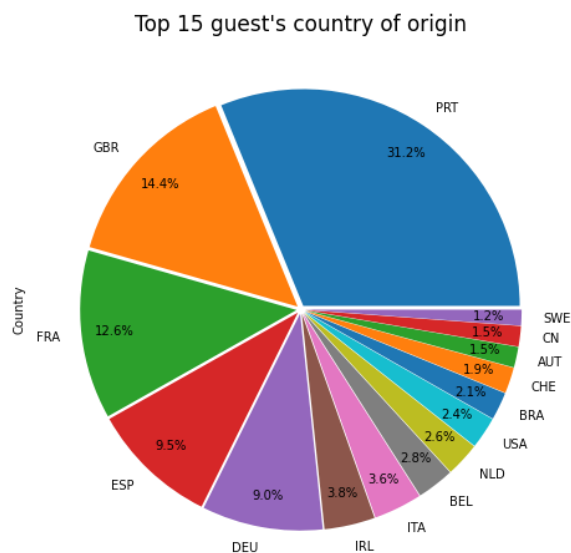
Approach: We fetched value count for types of resort



Inference: 2/3rd of total guest prefer City Hotel

b. Home country of majority of guests

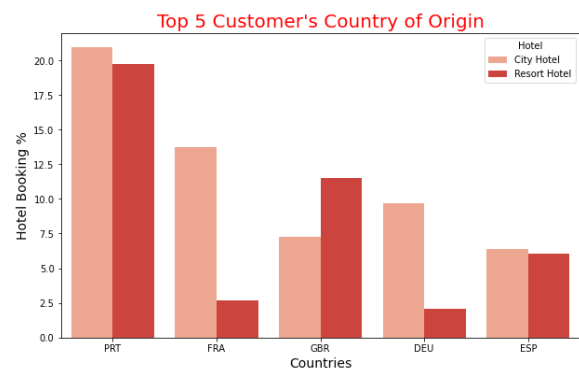
Approach: We featured a variable that groups 'country' by counts, sorts by values and fetches data for top 15 countries.



Inference: Most of guest visiting these hotels are from Portugal and other European countries namely Britain, France, Spain and Germany.

c. Hotel preference of guest from Top 5 Countries

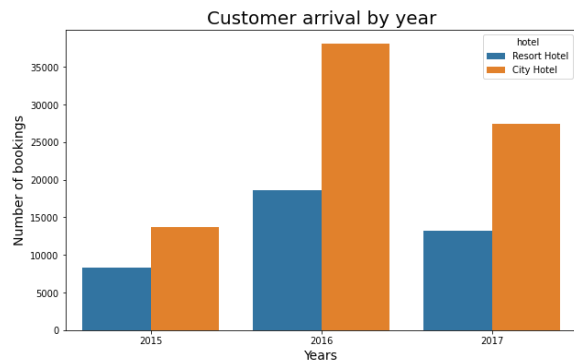
Approach: Sliced top 15 countries to just 5 countries and fetch hotel preferences of guest from these countries



Inference: Guest from southern European countries like Portugal and Spain prefer City Hotel and Resort Hotel equally. Guest from northern European countries like France and Germany prefer City Hotel a lot more than Resort Hotel. Guest from Britain prefers lavish Resort hotels

d. Overview of guest's visit over different years

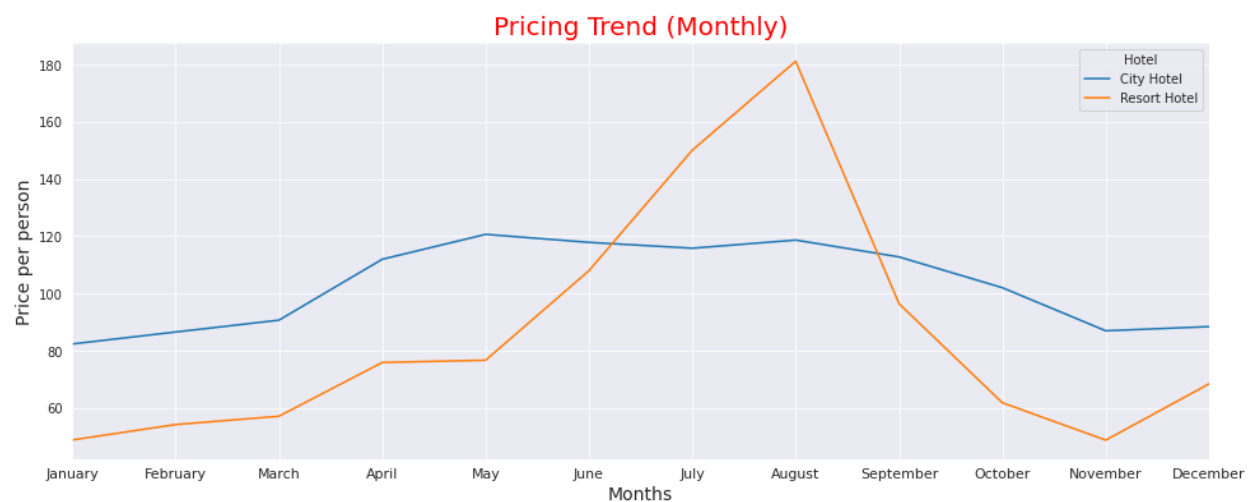
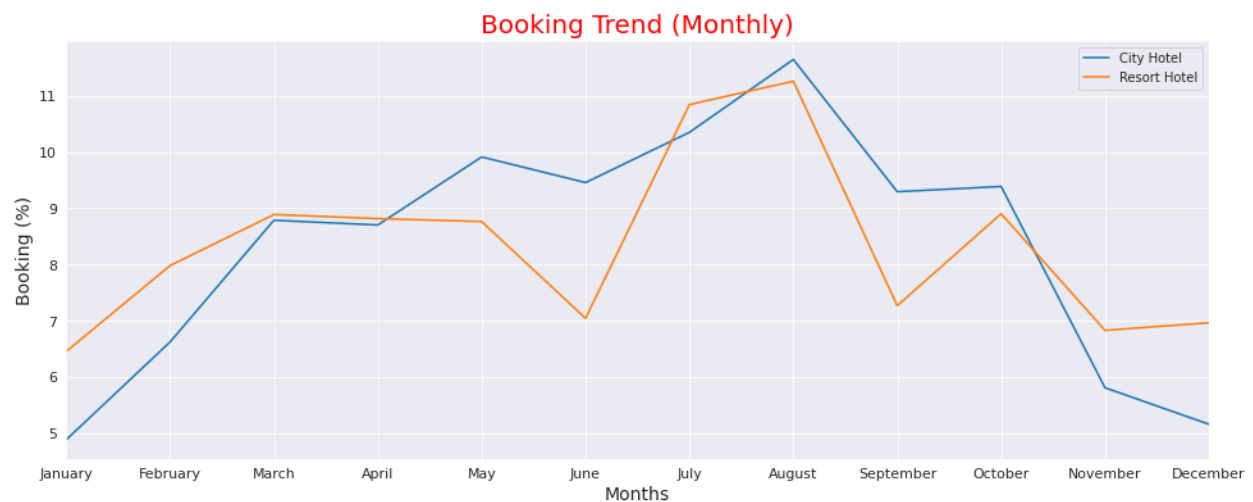
Inference: We fetched value count of booking reservation each year for different hotels



Inference: 2016 was the year where number of hotel booking was highest followed by total booking in 2017 and 2015.

e. Booking and Price trend round the year

Approach: We fetched booking count data and pricing per person data, grouped it by month with hue on city hotel and resort hotel and plotted a line graph to analyze booking trend and pricing trend respectively.

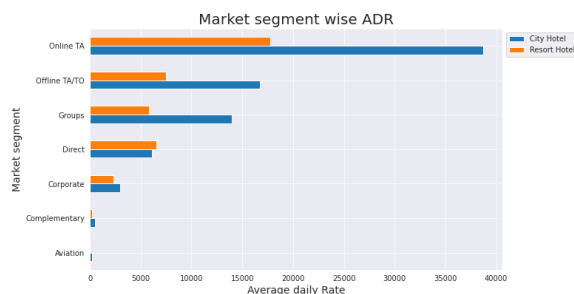


Inference: From Booking trend it can be inferred that around 11.5% of total reservations throughout year are coming from August whereas January has the least reservation of mere 5%.

Pricing trend is highly correlated with booking trend indicating that price for Resort hotels during peak season hikes to nearly 300% compared to off-season. Meanwhile, pricing trend for City hotels suggests almost same pricing throughout the year with low fluctuation during busy period from May to August

f. Average booking rate of different Market Segments

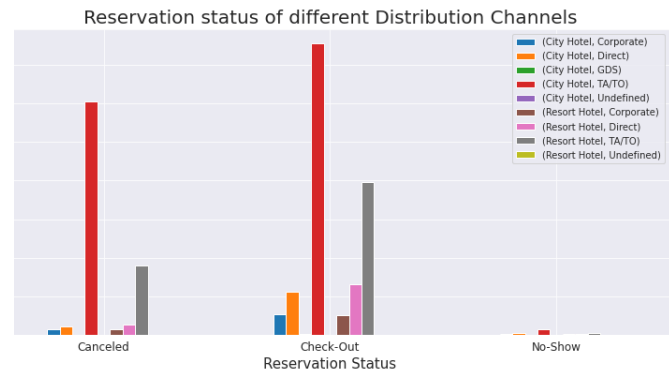
Approach: Grouping by different market segment and hotel; and plot average daily rate



Inference: Inspecting different market segments, it was concluded that online travel agency holds monopoly as both hotels are getting the most of booking from online travel agency (around 79%)

g. Reservation status from different Distribution Channels

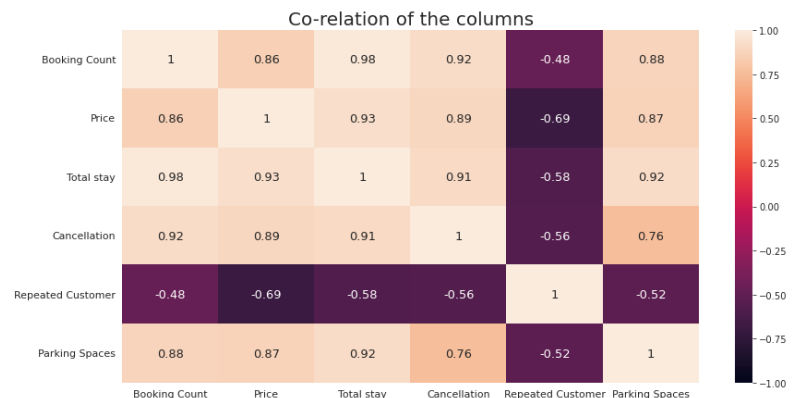
Approach: Grouping by different market segment and hotel; and analyzing different reservation status plot.



Inference: Bookings and Cancellations from both Hotels are more from Travel agency (TA/TO). Guest visiting both Hotels directly and via Corporate are less likely to cancel their booking

h. Correlation between different booking criteria

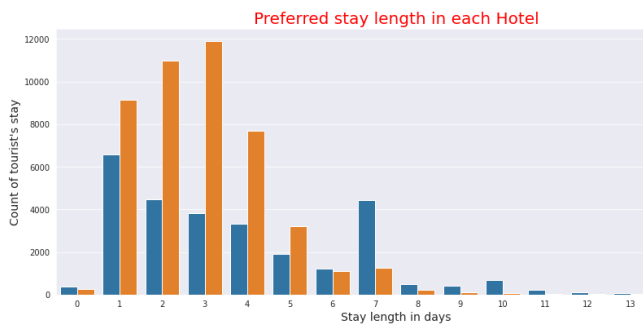
Approach: We carried out multiple aggregation on multiple booking relevant criteria to analyze correlation between them via Heatmap.



Inference: There is high positive correlation between Booking, Pricing, Total Stay, Cancellations and Parking spaces whereas negative correlation with Repeated guests. There is firm correlation between Parking space and Cancellation inferring that people are more likely to cancel their booking if Parking space is not available.

i. Guest's stay length

Approach: Values count on our new feature column 'total_stay' to get insight on guest's preferred stay length.



Inference: Ideally guest prefer to stay 1-4 days in both hotels but 7 days stay at Resort hotel is also a popular choice among guests.

5. Conclusion

- Majority of the guest prefer City Hotel over Resort Hotel
City Hotel : 66.4%
Resort Hotel : 33.4%
- 2/3rd of total guest prefer City Hotel.
- Most of guest visiting these City hotels and Resort hotels are from Portugal and other European

countries namely Britain, France, Spain and Germany.

- Considering top 5 countries, it is evident that:
 - Guest from southern European countries like Portugal and Spain prefer City Hotel and Resort Hotel equally
 - Guest from northern European countries like France and Germany prefer City Hotel a lot more than Resort Hotel
 - Guest from Britain prefers lavish Resort hotels
 - This makes us believe that people from different region of Europe prefer different type of accommodations and comforts.
- From hotel's statistical data, we can notice that larger families which includes children prefers Resort Hotel more than City Hotel.
- Peak visiting season is from mid June to August because of summer breaks in Europe
- Off season is from November to February because of cold weather throughout Europe
- Guests can consider visiting these hotels during month of June and September to enjoy decent weather with almost full availability of hotels accommodation.
- Pricing trend is highly correlated with booking trend when it comes to Resort hotel
- Pricing trend indicates that during Peak season, price for Resort

hotels is triple compared to off-season

- k. Pricing trend City hotels suggests almost same pricing throughout the year with bit of fluctuation during May to August
- l. We can see that Resort hotel and City hotel are getting the most of booking from Online travel agency and may be in future it will be monopoly by them. Hence hotel owners should promote more in different market segments.
- m. We can see that Bookings and Cancellations from both Hotels are more from Travel agency (TA/TO)
- n. And cancellation from people travelling via corporate distribution channel is less.
- o. For short stays guests prefer city hotels. Guests do prefer Resort hotel for short stays but it is most preferred for longer stays.
- p. There is high positive correlation between Booking, Pricing, Total Stay, Cancellations and Parking spaces whereas negative correlation with Repeated guests. With increase in Booking --> Pricing, Total stay and Parking spaces occupation increases but increase in Pricing leads to repeated Customers not visiting again.
- q. There is firm correlation between Parking space and Cancellation inferring that people are more likely to cancel their booking if Parking space is not available.

r.