

Hotel Booking Analysis

Sagar Malik

Data science trainee,

AlmaBetter

Abstract:

The objective of this work is to analyze the Hotel booking patterns for two types of Hotels i.e., city and resort using exploratory data analysis, in order to acquire valuable information about booking patterns and choices of people.

The work has been carried out upon the Hotel Booking Data (HBD), which is an open database containing a list of information about various hotel bookings From 2015 - 2017.

Keywords: *exploratory data analysis, Hotel Booking, City Hotel, Resort Hotel*

1. Problem Statement

The HBD includes systematic data on Hotel Booking records from 2015 - 2017 of city hotels as well as resort hotels from various parts of the world and mostly the countries from the united kingdom. The data is maintained and built over a 3-year span to help tourism business and hotel booking take advantage of it, using its insights accordingly.

When we talk about travelling the first thing that comes to mind is where will you stay? In this project, I focus on Hotel bookings in different types of hotels and at different

times of the year. Also deciding the best time to travel by analyzing the dataset provided by START (Hotel Booking Analysis) to explore meaningful patterns and statistics.

- hotel: Categorical column with information about the type of hotel (City/Resort).
- is_canceled: Boolean Type with information about cancellation of the booking.
- lead_time: The time they are scheduled to arrive at the hotel.
- arrival_date_year: Which has the guest arrived.
- arrival_date_month: Which month of the year the guest checked in.
- arrival_date_week_number: Which week of the month the guest checked in.
- arrival_date_day_of_month: which date of the month the guest checked in.
- stays_in_weekend_nights: Number of guests stayed during the weekend nights.
- stays_in_week_nights: Number of guests stayed during the weeknights.
- adults: counts of adults.
- children: count of children.
- babies: count of babies.
- country: Codes names of the countries where the hotels are.
- market_segment: dividing potential guests into groups based on a set of shared characteristics.

- `distribution_channel`: Different channels of a hotel to receive bookings.
- `is_repeated_guest`: Boolean type, whether the guest has come before or not.
- `reserved_room_type`: room types reserved during booking.
- `assigned_room_type`: room assigned to the guest.
- `customer_type`: categorical type column holding information about the type of group.
- `ADR`: average daily rate.
- `reservation_status_date`: information regarding date of check-in/check out.

2. Introduction

The primary purpose of hotels is to provide travellers with shelter, food, refreshment, and similar services and goods, offering on commercial basis things that are customarily furnished within households but unavailable to people on a journey away from home. Historically hotels have also taken on many other functions, serving as business exchanges, centres of sociability, places of public assembly and deliberation, decorative showcases, political headquarters, vacation spots, and permanent residences. The hotel as an institution, and hotels as an industry, transformed travel all over the world, hastened the settlement of the continent and extended the influence of urban culture.

3. Factors Affecting Hotels

Many popular hotels are built next to major attractions. Visitors to theme parks, big cities and vacation spots need a place to

stay, so they'll pay more if they get to book a room close to what they want to do. Most other hotels are in smaller cities or beside remote stretches of highways. If that's the case for you, keep an eye out for developments in your area.

New attractions like a restaurant or mall could be a reason to raise the price of your room rates. If yours is the only hotel in the area next to a popular restaurant chain, people will book with you because they know they can get their favourite food right next door.

Over the past decade, the economy has grown at a steady rate as it recovered from the 2008 recession. When it flourishes, people have more money to spend, so it's reasonable to increase room rates. Monitor the local, national and global markets to understand if your guests are in a place to spend more money than in previous months.

Hotel tech goes beyond what kind of TV guests have access to and how fast the Wi-Fi is. Software programs can increase room rates and pull in more guests at the same time. In the modern age of technology, people will pay more for guaranteed cybersecurity, especially when they trust a hotel with their payment information. Clarifying how you protect personal data will build trust with visitors and give them more reason to pay a bit higher for a room.

Cleanliness makes or breaks a hotel's reputation. Poor hygiene results in terrible online reviews, which turn more potential guests away. Prevent this from happening by upgrading your hotel's standard of

cleanliness and teaching staff how to inspect a room for pests.

If a resulting problem occurs — which is sometimes beyond anyone's control — rid your property infestations by relying on a professional exterminator. The faster hotel operators stop pest problems, the sooner you can get back to increased room rates that match the higher standards of your property.

4. HBD Dataset

The dataset used in this project is called the HB (Hotel Booking) dataset, which is being recorded for over 3-years.

5. Steps involved in Exploratory Data Analysis:

5.1 Project Structure

Implementation of this project involves system design, backend design, visual design, and user interface. It includes the overall design plan of analysis which explains how each parameter is correlated with others.

5.2 Data Cleaning

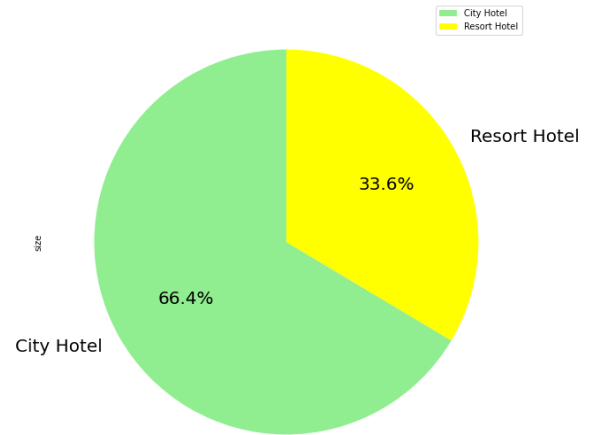
Our dataset contains 32 columns in which there are a lot of numbers columns that are irrelevant. I choose only those columns which are useful.

Analysis

This section consists of details regarding the visual results.

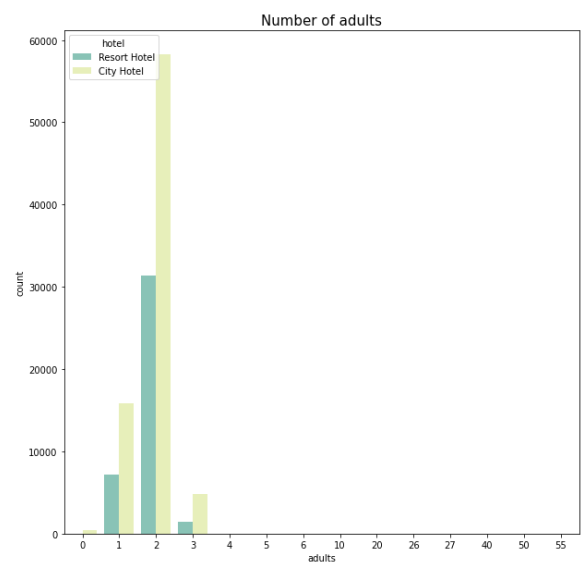
6. Analysis

• Distribution Type



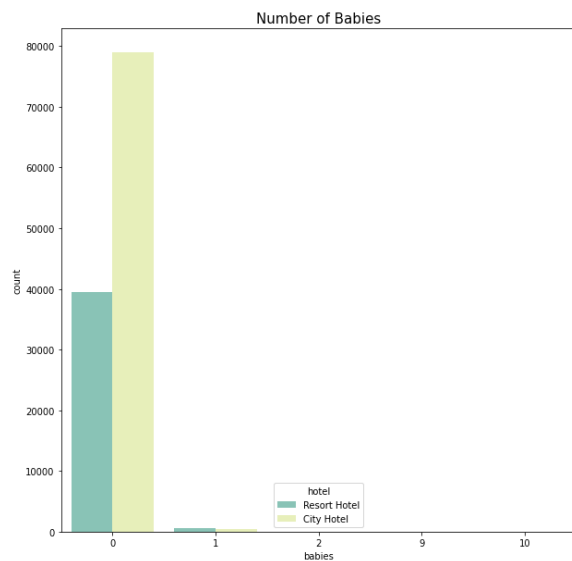
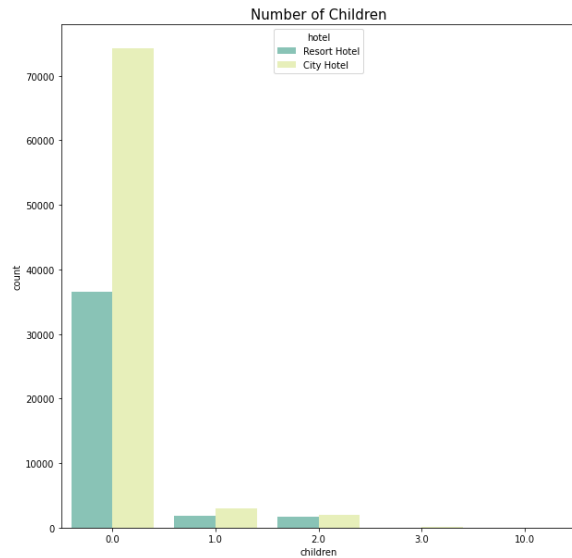
In our dataset, there are more numbers of City Hotels than that resort hotels, maybe because resorts are more expensive and lie on the outskirts.

• Total number of adults, children and babies count



Summarizing from the above chart we notice that the most number of

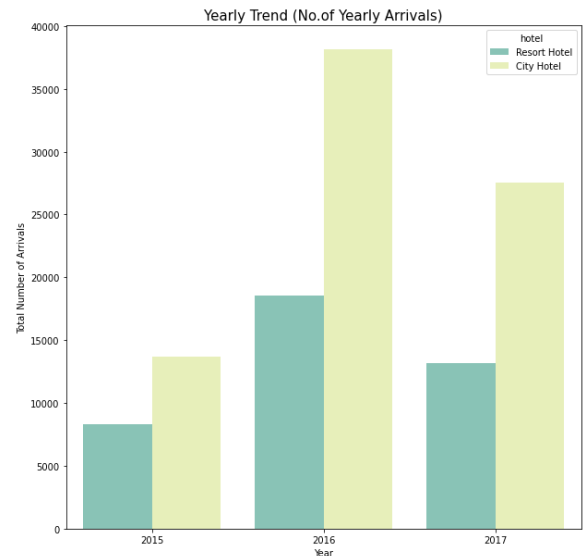
bookings/check-in in either of the hotel type is done by a pair of adult, maybe newlyweds couples can be the most attraction.



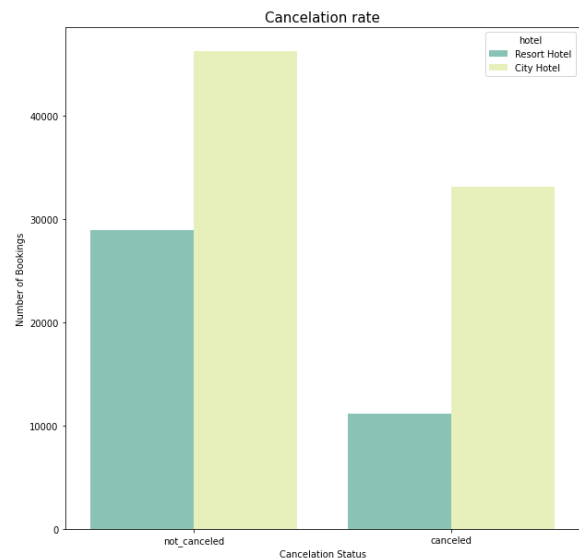
Summarising the above two charts gives multiple insights about the family with children and babies with them during the time of check-in.

It is very clear from the chart itself that families with children and babies have very fewer bookings and they don't have a specific choice between the type of hotel.

● Number of yearly arrivals



2016 was the year when the most hotel booking would have to take place, and with hotel booking comes the rate of cancellation of the booking among the type of the hotel, which is an important factor to visualize.

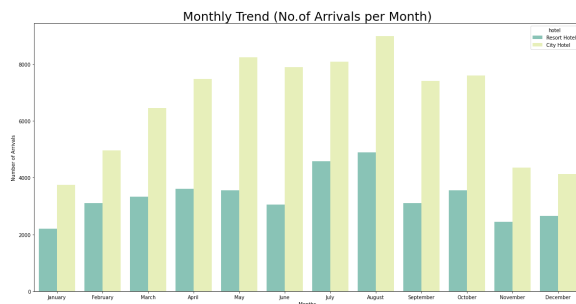


With the number of bookings, we also have a huge cancellation count.

Since 66 per cent of the hotels were city hotels thus as expected the cancellation in the city hotels were higher but if we take the

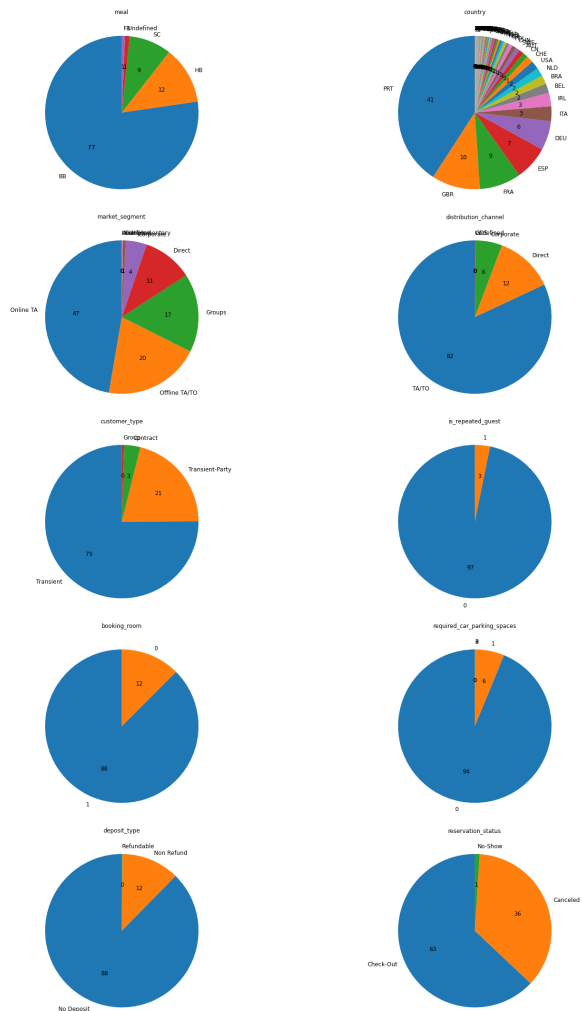
rate of cancellation into consideration we see that there is not much noticeable difference between the both of them compared to the numbers and weightage

- **Monthly Trend (Number of arrivals)**



From the above chart, we observe that most arrivals are taking place from April to August as it explains "Summer Season" is the most preferred season for a vacation and for families with children find time during the Summer breaks of their kids. Thus; we have a more specific duration to target the audience.

- **Distribution of the leftover important columns**



Important Insights from the above visualization

Our country pie Chart is a little messy but we can easily figure out the top countries are from the UK, where 41 per cent bookings are from PRT (Portugal), 10 per cent from GBR (Great Britain), 9 per cent from FRA (France) and 7 per cent from ESP (Spain).

Major market_segments are Online TA and Offline TA/TO (Online/Offline Target audience)

Offline TO holds a whopping 82 per cent weightage in distribution

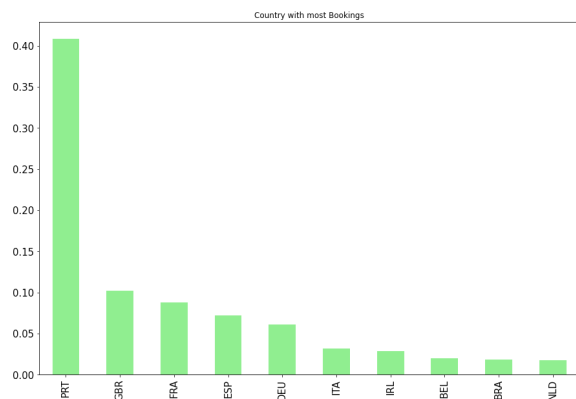
75%, 21% are Transient stays and Transient parties respectively (Transient ~ short duration stays). Thus; marketing special less day offer packages for stay and resort for parties for discounted prices will increase the bookings.

The repeated guests are only 3 per cent. Hence; this need to be improved - can be through email marketing of discounts or making some kind of membership cards to avail offer if they visit again in a certain time period.

88% of the people were assigned the same room as what they have booked rest were assigned something they didn't ask for, this can have a negative impact and hence need to be reduced.

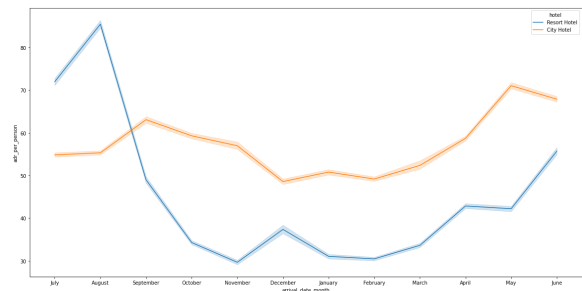
7) We see that there is 88 per cent of hotels take no deposit fees and thus this may be the reason why there are so many cancellations.

● Top Countries with most bookings



40 per cent are from France itself.

● Average daily rate per person



Average Daily Rate (adr) = Total revenue earned / Total night stays

ADR (per person) = ADR / [Number of adults] + [Number of children] (Not adding a number of babies since the count is very low compared to the dataset)

The adr_per_person has huge spikes at some points because that is what resorts are for (extra luxury)

In the case of city hotels, the ADR is not much variable.

7. Conclusion:

City hotels are in the highest demand and are often booked more due to less 'ADR' and many other factors.

Most bookings are from the country "Portugal". Hence we need to focus more on this country.

Families with children have no specific choice among the type of hotel

People only book for a shorter time, hence package modification and discounts for a shorter time should be introduced.

There is almost no repeated guest, which can be solved by specific membership discounts and other approaches.

References-

1. Towardsdatascience
2. HBD codebook
3. Google