Capstone Project-1

Hotel Booking Analysis

Sagar Malik

Contents:

Need for Hotels ?

- *HOTELS* AND *HOTEL INDUSTRY*. The primary purpose of hotels is to provide travelers with shelter, food, refreshment, and similar services and goods, offering on a 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, centers 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.

Introduction

- Hotel and lodging businesses are mainstays of their communities, and an important source of quality jobs. Seventy percent of mayors surveyed said hotel jobs provide the most opportunity, good benefits and wages within the tourism industry. They also noted that hotels support their communities through increased tax revenue, capital investment, tourism-related development and promotion, civic leadership, and charitable contributions and sponsorship.

- It can be really appreciated if we have data collected in the booking behavior across many places, because then we will be able to drive great insights which will further help in targeting our audience and many more aspects.

Problem Statement

- Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not 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 analyze the data to discover important factors that govern the bookings.

About the Data

- ***hotel*** : Categorical column with information about the type of hotel (City/Resort).
- ***is_canceled*** : Boolean Type with information about cancellation of 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 guest stayed during the weekend nights.
- ***stays_in_week_nights*** : Number of guest stayed during the week nights.
- ***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, weather the guest has come before or not.
- ***researved_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

Pre-Processing few columns

- Combining reserved_room_type and assigned_room_type to a single column to check weather the room that booked was assigned or not.

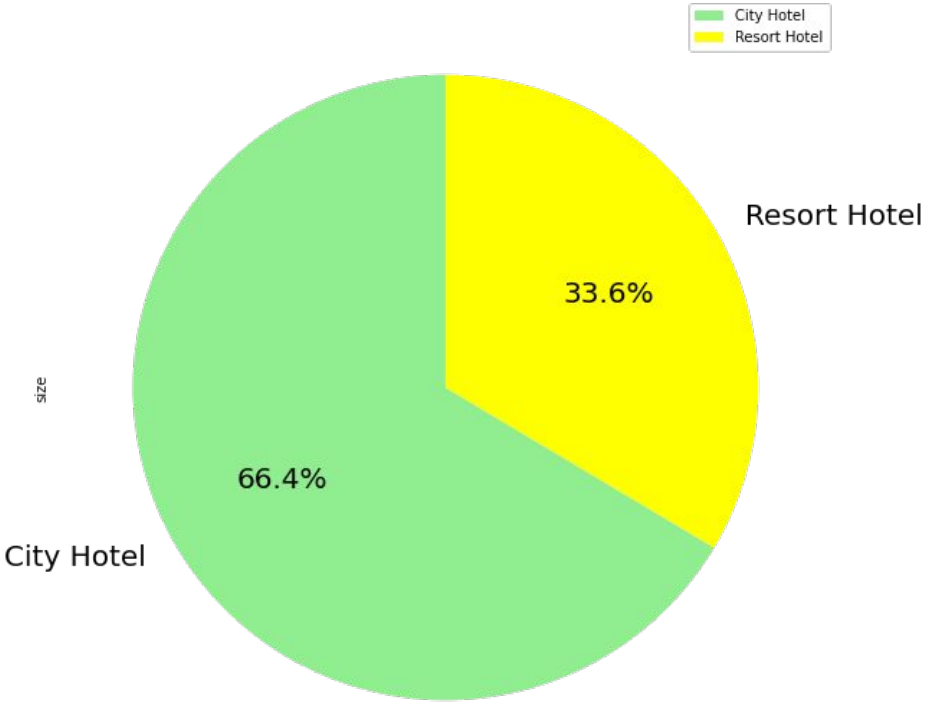
- Reserved room == Assigned room ---> 1
- Reserved room != Assigned room ---> 0

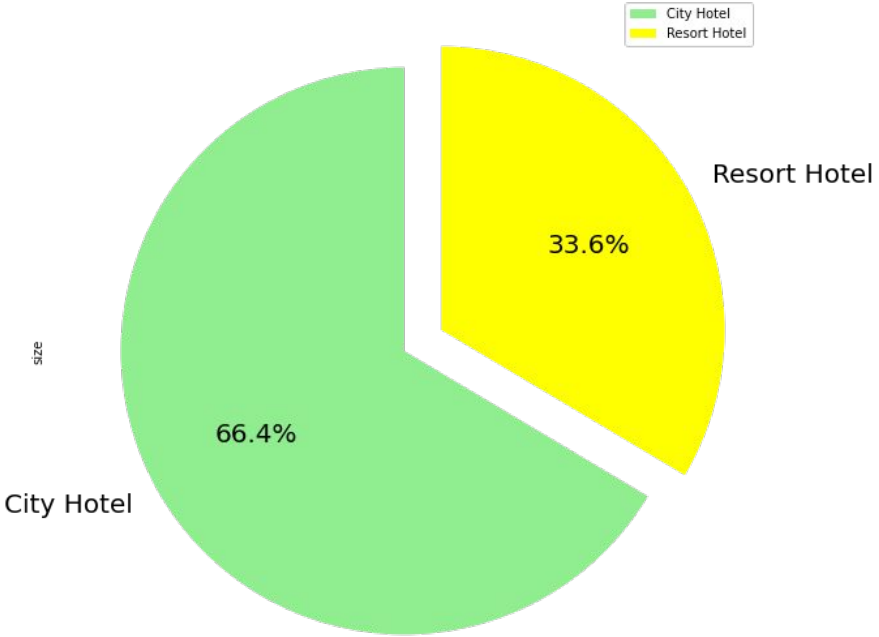
```
df['booking_room'].value_counts()
```

```
1    104473
0     14917
Name: booking_room, dtype: int64
```

EDA

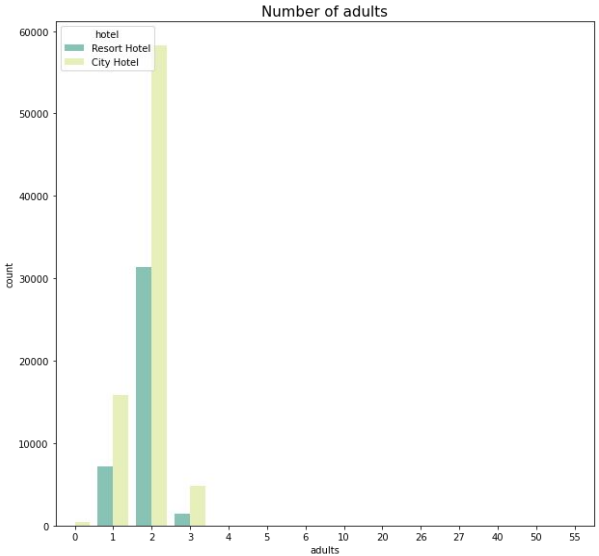
- As clear from the pie-charts 66% booking are from City Hotels and 34% is from Resort Hotel

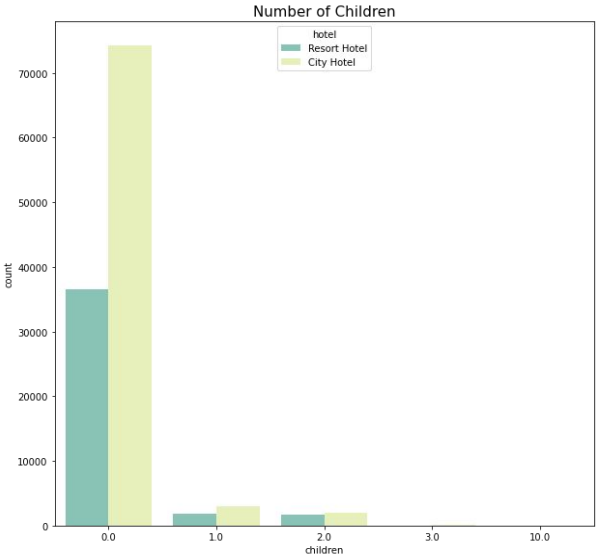


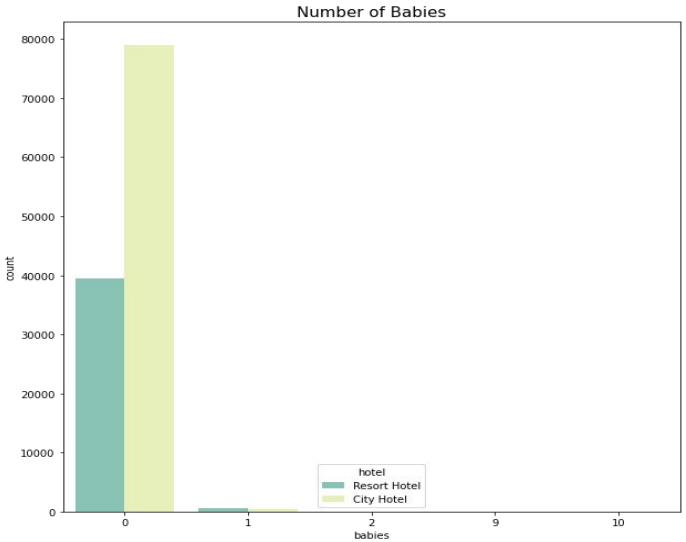


Analysing Different population of visitor in hotels

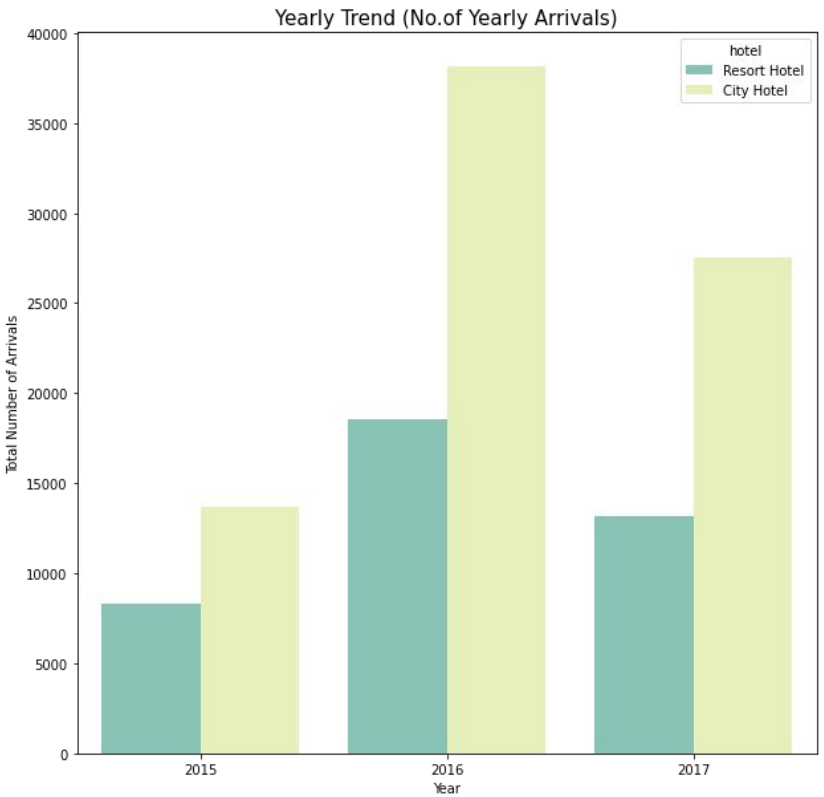
- One thing is clear that our target audience and most majority population are adults.
- And by pair we have give out a hypothesis of couples and youngsters
- people with children and baby have no specific choices of the type of hotels.



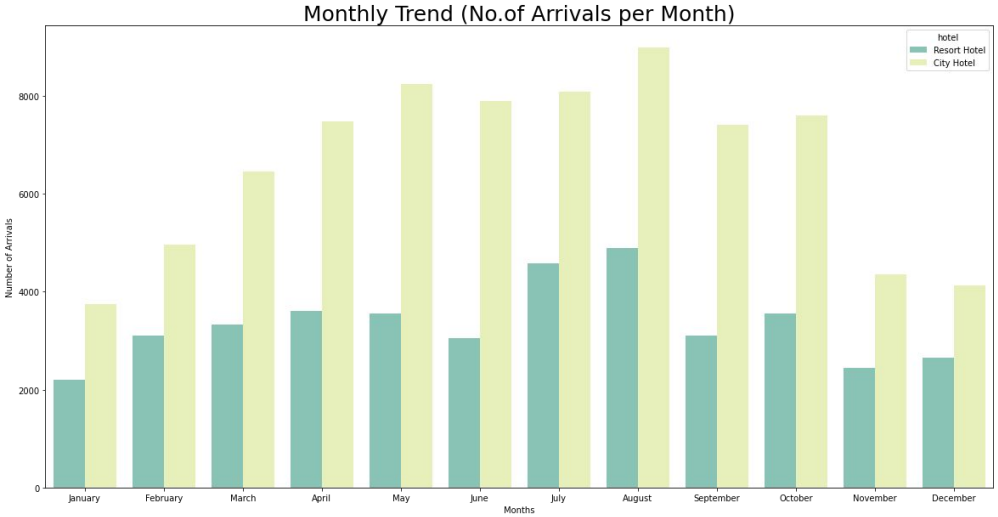




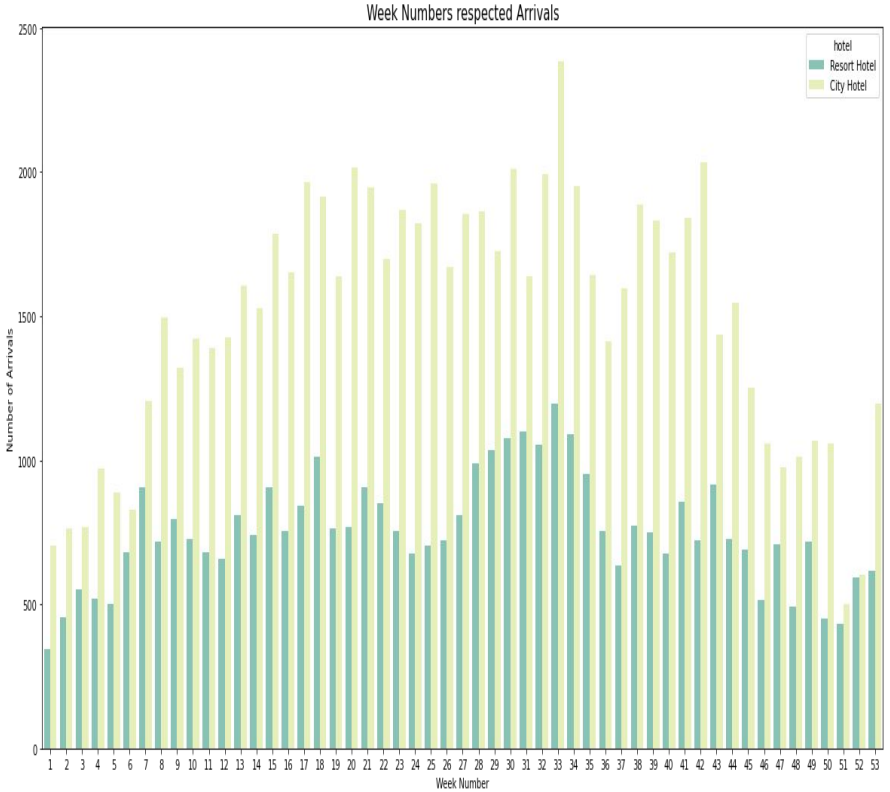
Arrival Trends



- 2016 was the year where the most hotel booking would have take taken place.
- Most booking spikes during the summer season.

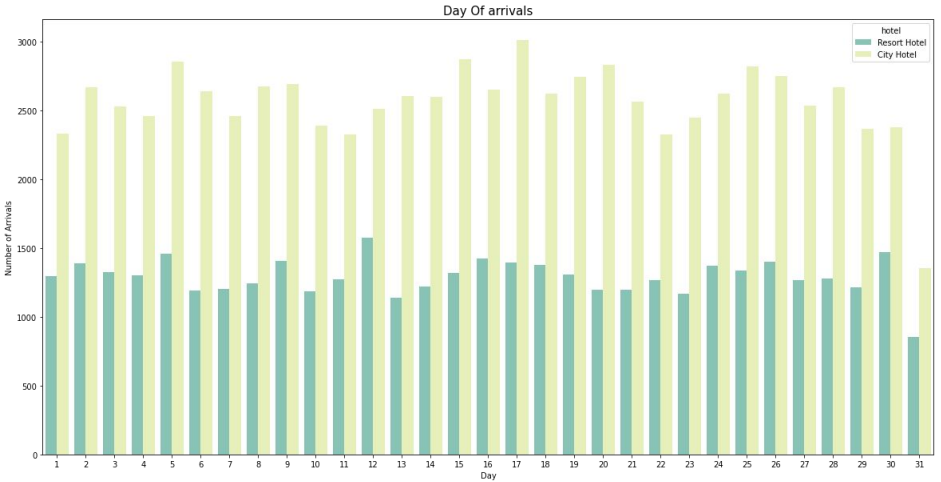


Arrival Trends (Conti.)

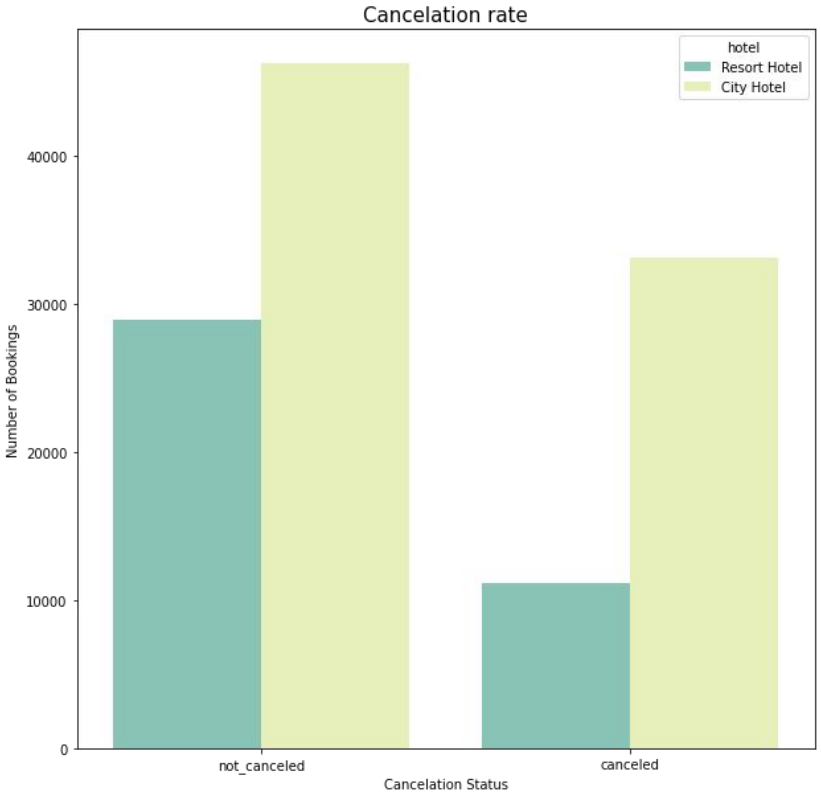


- Arrival_date_day of the month may not be useful as there is no fixed pattern to what date to chosen to travel in a month.

- Talking about the week popularity, it resembles the same insights that the arrivals spiked during the summer times.



Cancellation Rate



- Since 66 percent of the hotels were city hotel thus as expected the the cancelation 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.

EDA (Conti.)

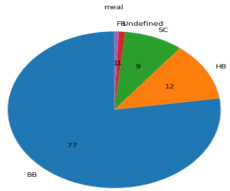
Important Insights from the above visualization

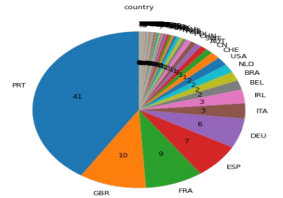
1) Our country pie Chart is little messy but we can easily figure out the top countries are from UK, where 41 percent bookings are from PRT (Portugal), 10 percent from GBR (Great Britain), 9 percent from FRA (France) and 7 percent from ESP (Spain).

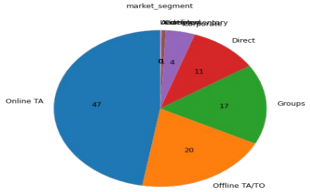
2) Major market_segments are Online TA and Offline TA/TO (Online/Offline Target audience)

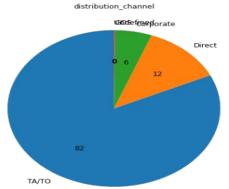
3) Offline TO holds a whooping 82 percent weightage in distribution

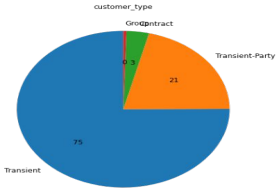
4) 75%, 21% are Transient stays and Transient parties respectively (Transient ~ short duration stays). Thus; marketing special less day offer packages for stay and resort for parties for discounted prices will increase the bookings.

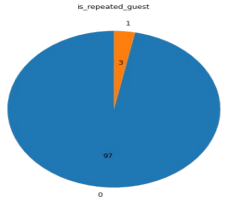


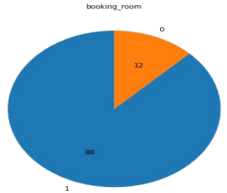


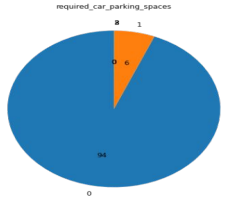


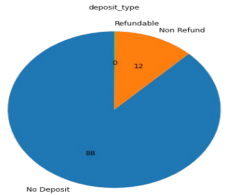


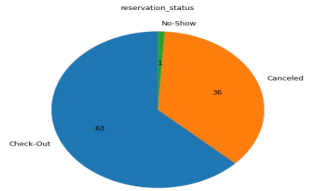








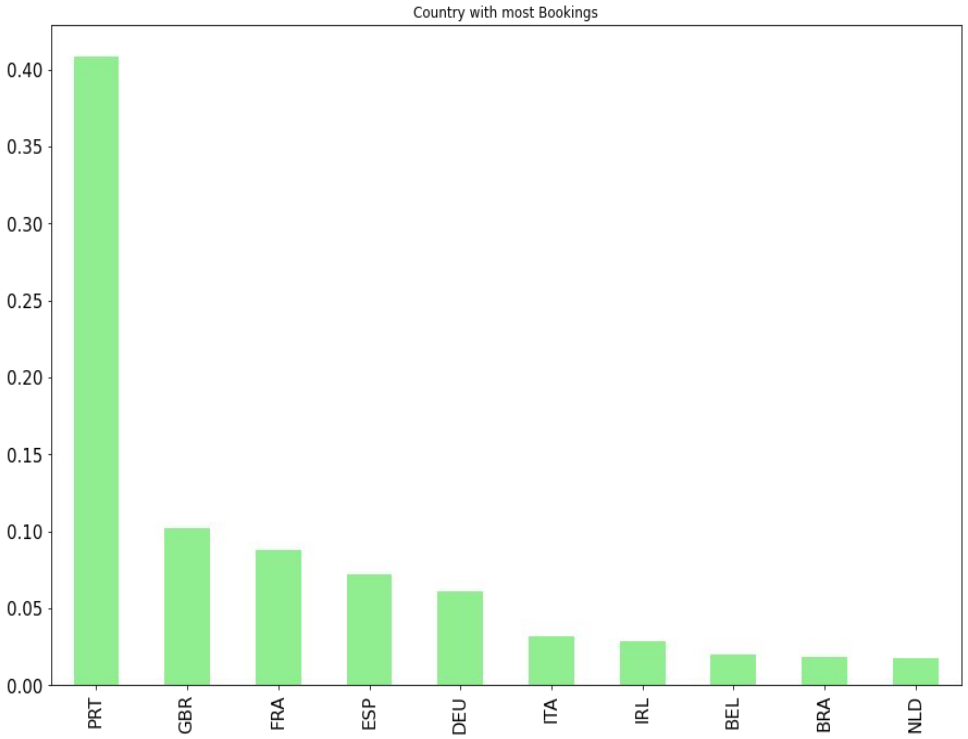




5) The repeated guests are only 3 percent. Hence; this need to be improved - can be through email marketing of discounts or making some kind of membership cards to avail offer if they visit again in certain time period.

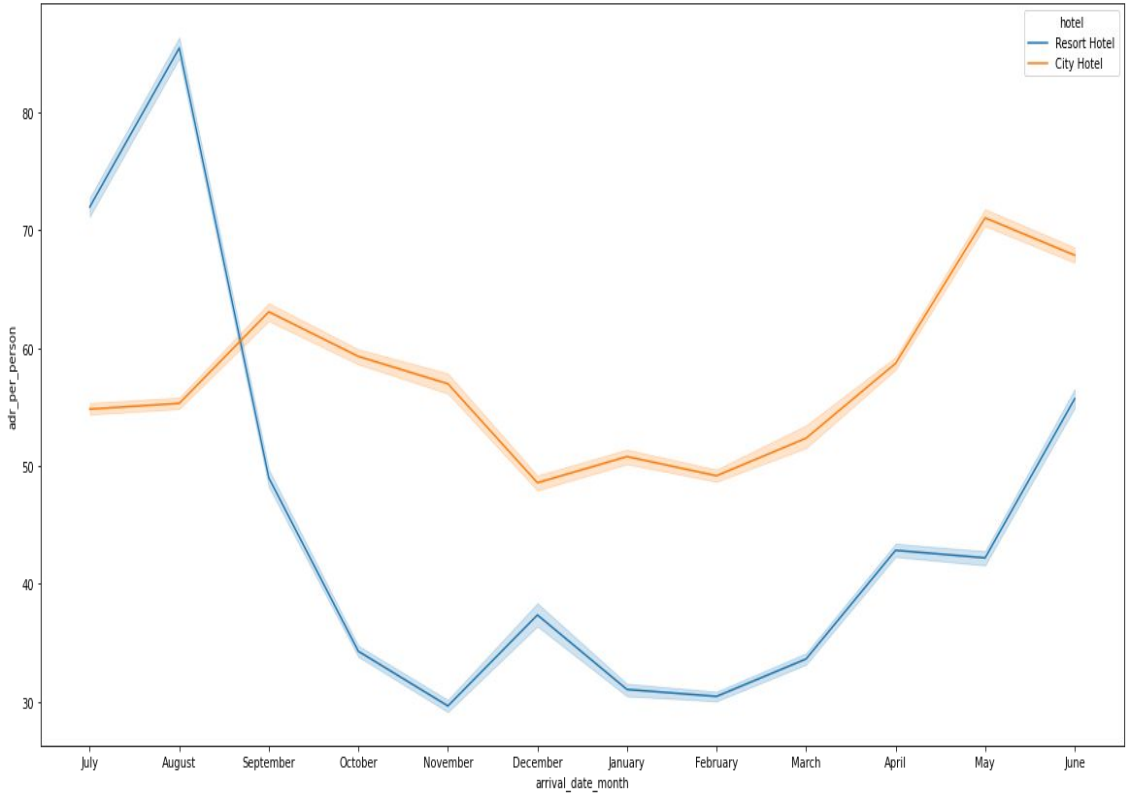
6) 88% of the people were assigned the same room of what they have booked rest were assigned something they didn't ask for, this can have negative impact and hence need to be reduced.

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



(40 percent bookings are from france itself)

Average Daily Rate (adr)



- Average Daily Rate (adr) = Total revenue earned / Total night stays
- Now to find the AVERAGE DAILY RATE PER PERSON, we simply make a little modification in the formula
- ADR (per person) = ADR / [Number of adults] + [Number of children] (Not adding 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 case of city hotels the adr is not much variable

Conclusion

1. City hotels are in the highest demands and are often booked more due to less 'adr' and many other factors
2. The most bookings is from the country "Portugal". Hence we need to focus more on this country.
3. Families with children have no specific choice among the type of hotel
4. People only book for shorter time, hence package modification and discounts for shorter time should be introduced.
5. There is almost no repeated guest, which can be solved by specific membership discounts and other approaches.


Thank You!