# Hotel Booking Analysis by EDA

**Debashish Das, Lucky Jain, Vivek Katolkar**
**Data science trainees,**
**AlmaBetter, Bangalore**

## Abstract:

To understand the Hotel Booking firstly, we have to know some parameters like the main few things I will usually consider include prices per night, distance of hotel from attractions and restaurants, availability of free breakfasts, scenery in hotel room, cleanliness of hotel room and of course, availability of free Wi-Fi. In this dataset, we ae able to know different types of bookings (i.e. type of hotel, duration of stay, types of visitors, types of booking, etc).

**Keywords: EDA Hotel booking Analysis**

## 1.Problem Statement

We are provided with data of hotel bookings
we had analysis data on following questions

- How many confirmed bookings are there in a month?
- Which is the month getting most visitors?
- Which Distribution Channel is mostly preferred?
- Which Market segment designation most Booking
- What is the maximum lead time in each hotel?
- Which Numbers of guests who did not booked hotels in advance?
- Which is the busiest year?
- How many cancelled are there after booking?
- Which countries are having most visitors?
- Which is the Meal Type of meal most booked
- Which customer type are having most booking?
- How may repeated guest are coming?
- How many guests arrived year-wise?
- How many numbers of car parking spaces required by the customer?
- How many families' member per reservation?
- How many customers have booked and then cancelled?
- Which months are having most expensive hotels?
- Which type of hotels are preferred by adults.
- Which type of hotels are having most booking in a weekend night and then cancelled?
- Which type of hotels are having most booking in a weekdays night and then cancelled?
- How many kids are preferred in hotel?
- How many people have been registered in the hotel?
- How many car parking spaces have been used?
- How many numbers of car parking spaces? required by the customer?
- What is the Babies favourite and least favourite meals?
- How many guests arrived date-wise?
- Guest arrival trend week wise is as follows

## 2.Data description:

The main objective of Exploratory data analysis is to understand trend and behaviour of guest in hotel bookings. For that first we will need to understand what every feature in data means. The data table consists of 119,390 rows and 32 columns. Our analysis starts with defining each column and our understanding for each column mentioned below:

- hotel: Hotel type (City hotels, Resort hotels)
- is_canceled: value indicates if the booking is cancelled or not.
- lead_time: How long in advance the booking was made.
- arrival_date_year: Customer arrival year.
- arrival_date_month: In which month of the year customer visited hotel.
- arrival_date_week_number: In which week of the year customer arrived.

- arrival_date_day_of_month: Date of the month customer visited hotel.
- stays_in_weekend_nights: Customer stayed or booked to stay in hotel during weekend nights.
- stays_in_week_nights: Customer stayed in hotel during week nights.
- adults: Number of adults.
- children: number of children.
- babies: Number of babies.
- meal: Type of meal booked.:
- country: Country of origin of customer.
- market_segment:
- where the bookings came from
- distribution channel:
- Booking distribution channel. The term "TA" means "Travel Agents" and "TO" means "Tour Operators"
- is_repeated_guest: Value indicating if the booking name was from a repeated guest (1) or not (0).
- previous cancellations: Number of previous bookings that were cancelled by the customer prior to the current booking.
- previous_bookings_not_canceled: umber of previous bookings that were cancelled by the customer prior to the current booking.
- reserved_room_type:
- Code of room type reserved. Code is presented instead of designation for anonymity reasons
- . • assigned_room_type:
- Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due
- booking_changes:
- Number of changes/amendments made to the booking from the moment the booking was entered on the PMS.
- deposit_type: Indication on if the customer made a deposit to guarantee the booking.
- agent:
- ID of the travel agency that made the booking.
- company:
- ID of the company/entity that made the booking or responsible for paying the booking.
- days_in_waiting_list: Number of days the booking was in the waiting list before it was confirmed to the customer.
- customer_type: Type of booking, assuming one of four categories.
- adr: Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights.
- required_car_parking_spaces: Number of car parking spaces required by the customer.
- total_of_special_requests: Number of special requests made by the customer (e.g. twin bed or high floor).
- reservation_status: Reservation last status, assuming one of three categories: Canceled –booking was cancelled by the customer;
- CheckOut:
- customer check out from hotel,
- No show:
- Customer did not check-in hotel and informed hotel with reason.
- reservation_status_date:
- Date at which the last status was set. This variable can be used in conjunction with the Reservation Status to understand when was the booking cancelled or when did the customer checked out of the hotel.

## 3.Cleaning Data

Cleaning data cleaning data is crucial step before EDA as it will remove the ambiguous data that can affect the outcome of EDA.

While cleaning data we will perform following steps:

1) Remove duplicate rows

2) Handling missing values.

3) Convert columns to appropriate datatypes.

4) Adding important columns

# 4. Steps involved:

## 4.1. Data Wrangling:

After loading the dataset, we performed this method by cleaning, organizing, and transforming raw data into the desired format which makes us to understand the data clearly. This process helped us to tackle the unwanted data, to produce accurate results, to make better decision.

## 4.2. Null Value Treatment:

Our data set contains a small number of null values; still we have treated the null values by filling with zeros in order to produce more accurate results.

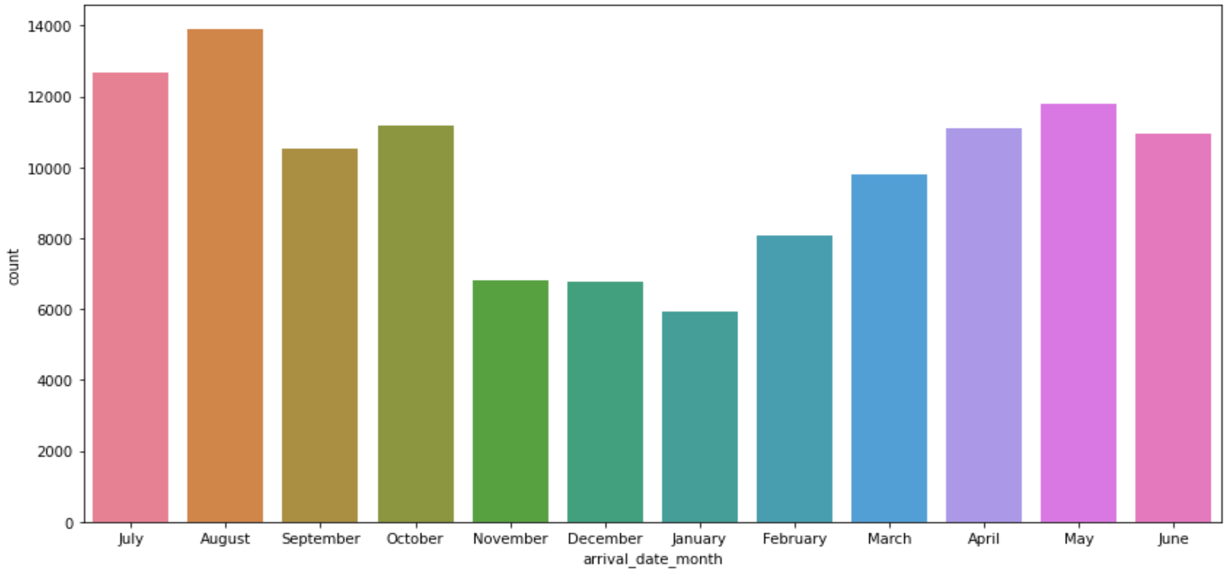
## 4.3. EDA:

After loading the dataset, we performed this method by comparing our target variable that is booking analysis with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

Mainly performed using Matplotlib and Seaborn library and the following graph and plots had been used:

- Bar Plot.
- Histogram.
- Scatter Plot.
- Pie Chart.
- Line Plot.
- Heatmap.
- Box Plot

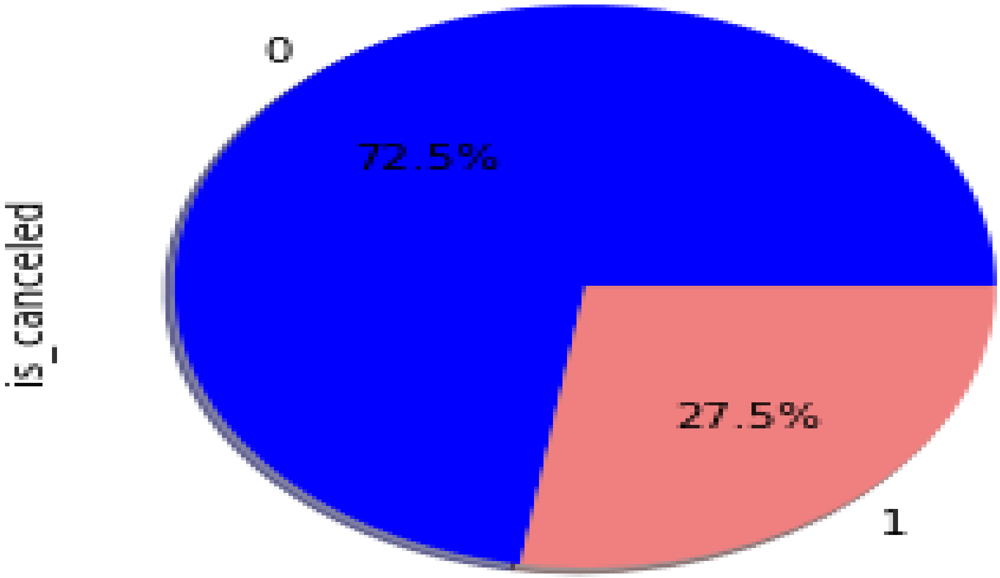
**Confirmed bookings in a month:-**



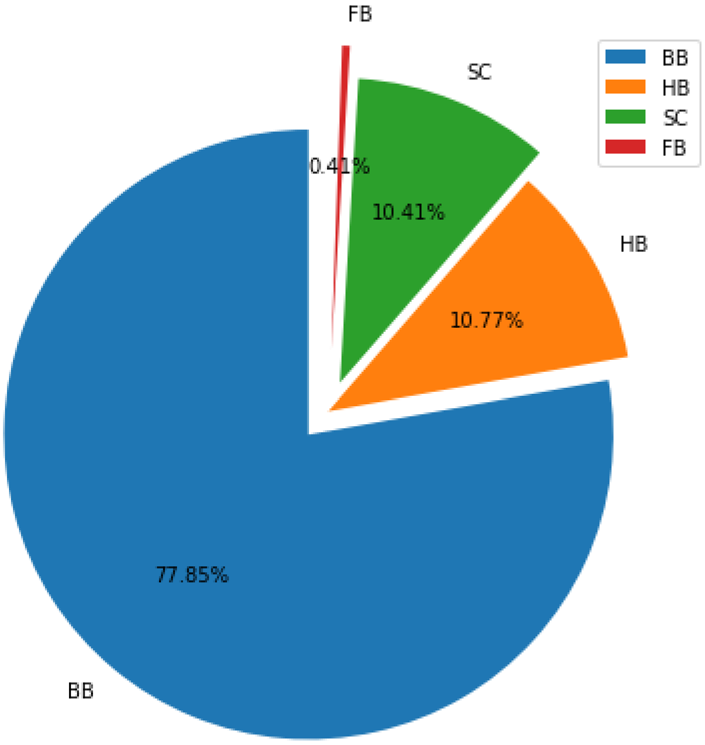
**August Month 13877 is high confirmed booking hotel. July is 2nd hight 12661 booking confirmed.**

## Cancelled bookings:-

**According to the pie chart, 72.5% of bookings were not cancelled and 27.5 % of the bookings were cancelled at the Hotel.**



## Type of meals booked:-

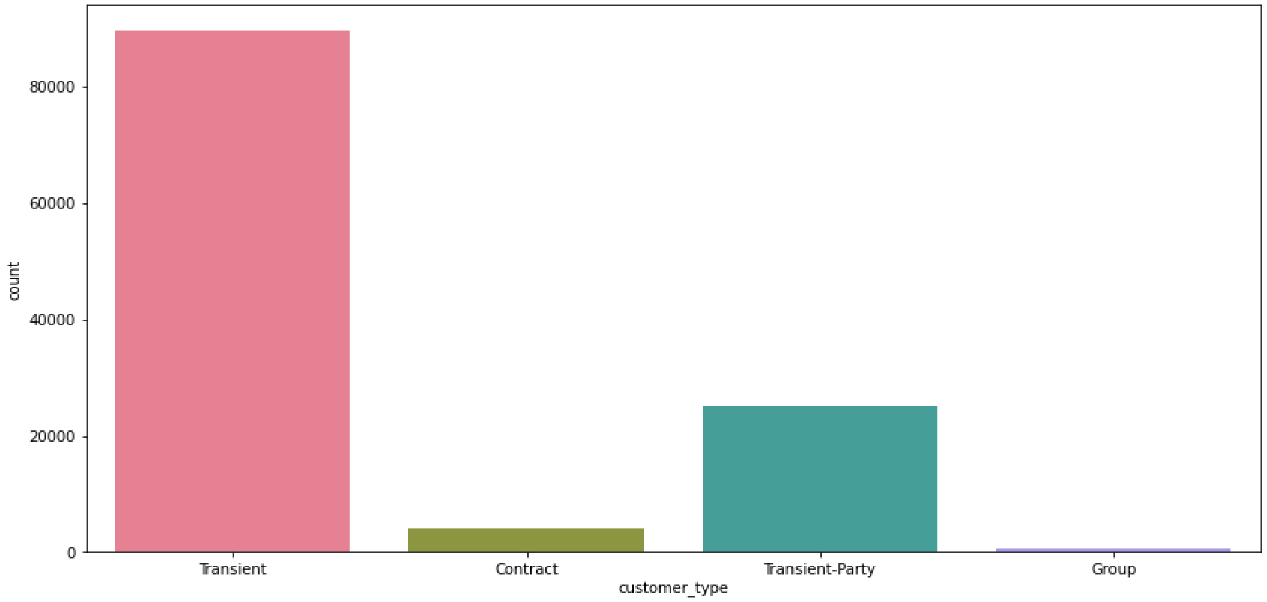


**Categories are presented in standard hospitality meal packages:**

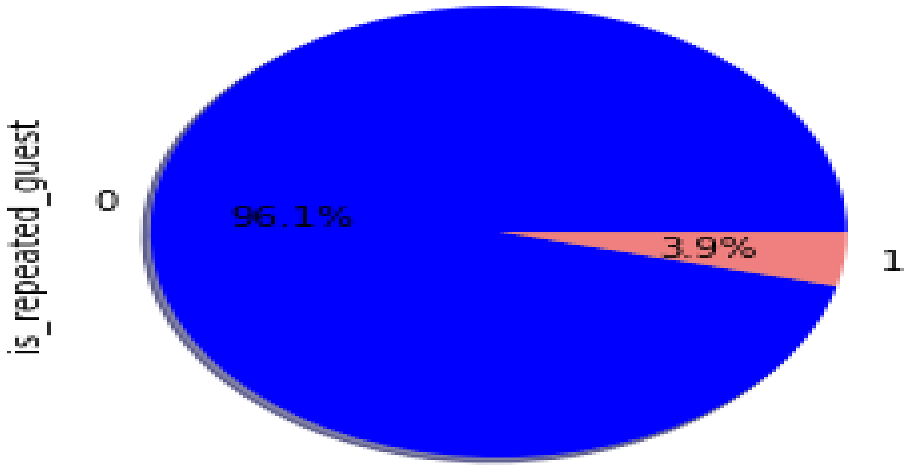
**Undefined/SC — no meal package; BB — Bed & Breakfast; HB — Half board (breakfast and one other meal — usually dinner); FB — Full board (breakfast, lunch and dinner)**

**-77.8% people prefer (BB — Bed & Breakfast) meal**

## Customer type are having most booking

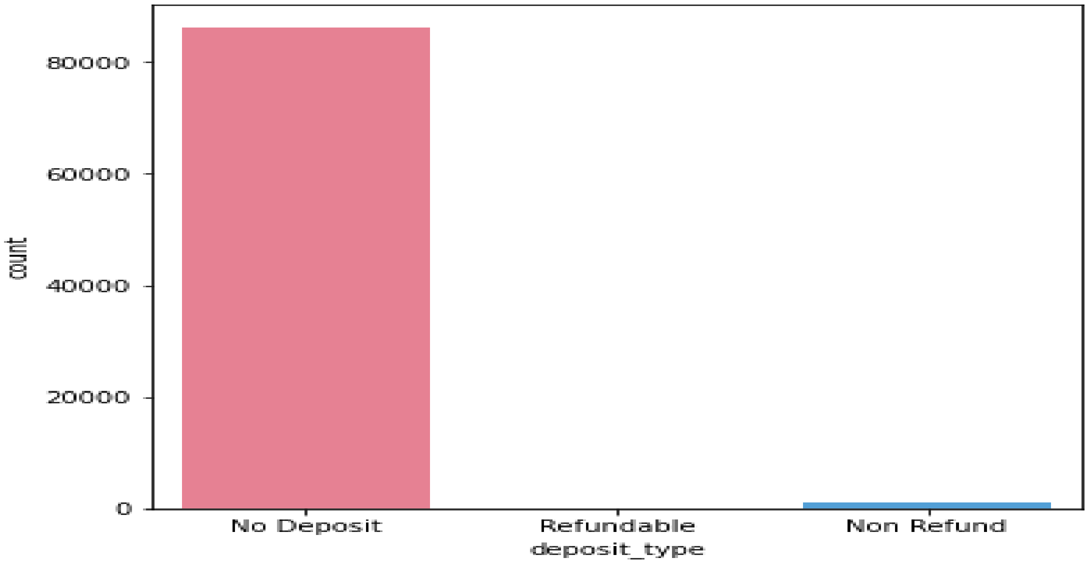


## Repeated guest:-



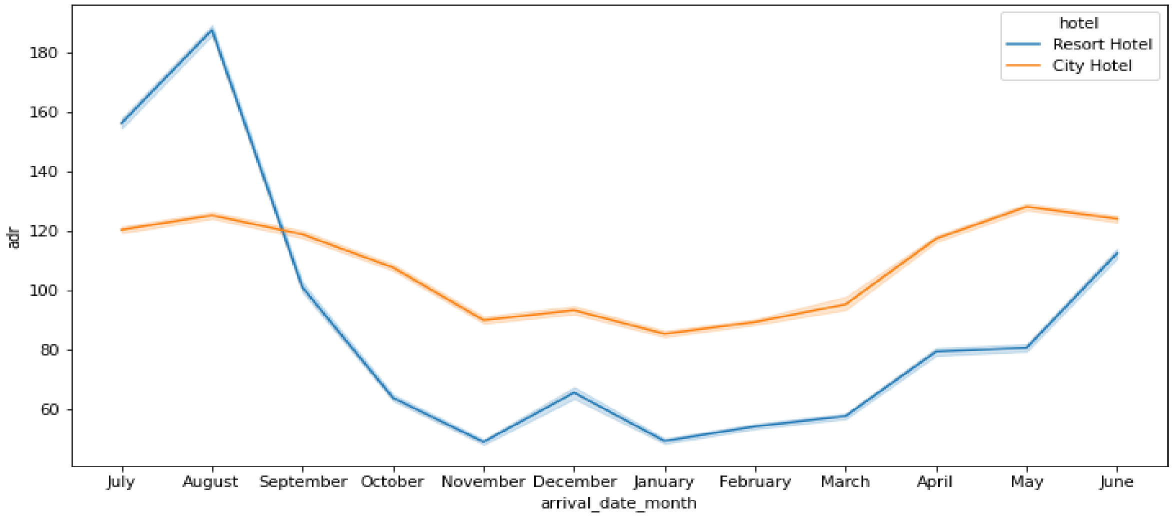
**Almost 4% of guest are coming repeatedly.**

## Deposit Type Indication:-



**Almost 99% of Customer are having major booking without deposit**.

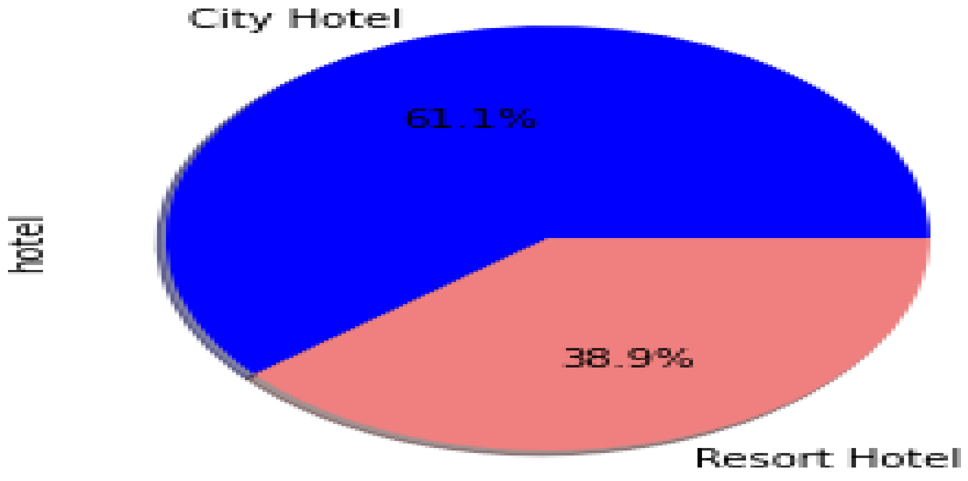
## Months having most expensive hotels:-



**For resort hotels, the average daily rate is more expensive during august, July and September.**

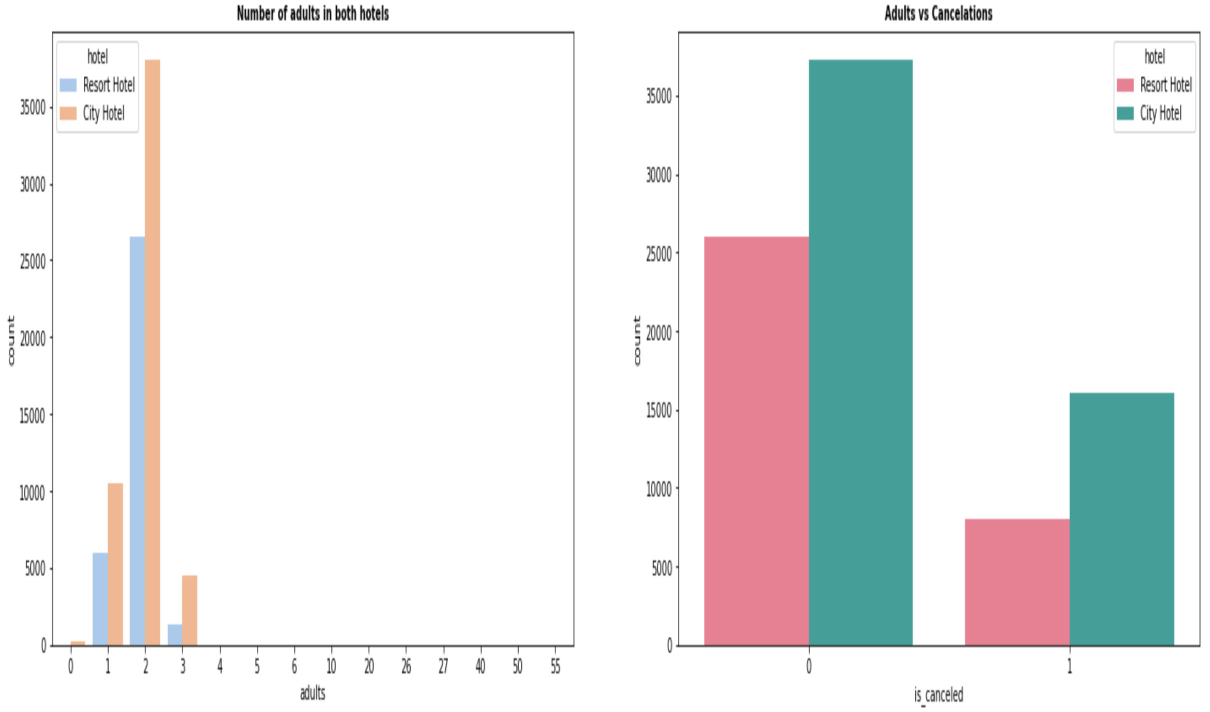
**For city hotels, the average daily rate is more expensive during august , July ,June and may.**

## Hotel type are having most reservation:-



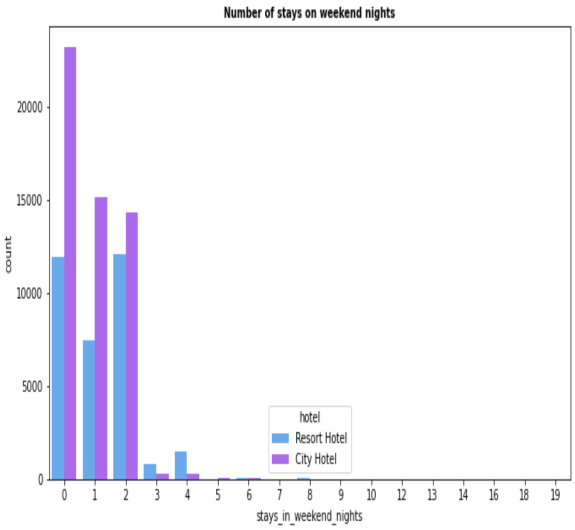
**City hotel is having most prefer customers.**

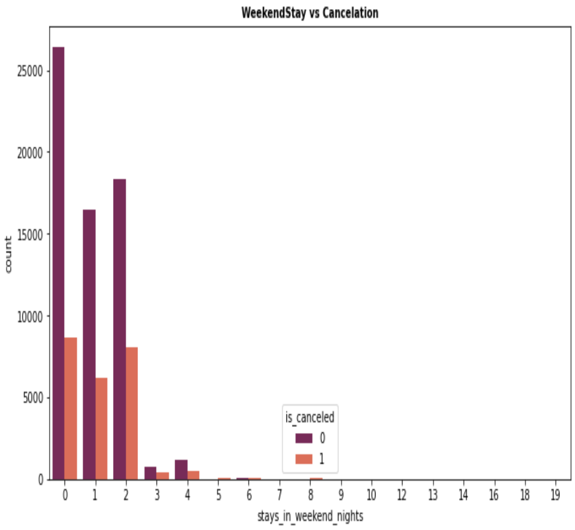
## Hotels are preferred by adults:-



**Adult prefer more city hotel and adult cancelled are less.**

## Hotels are having most booking in a weekend nights and then cancelled:-

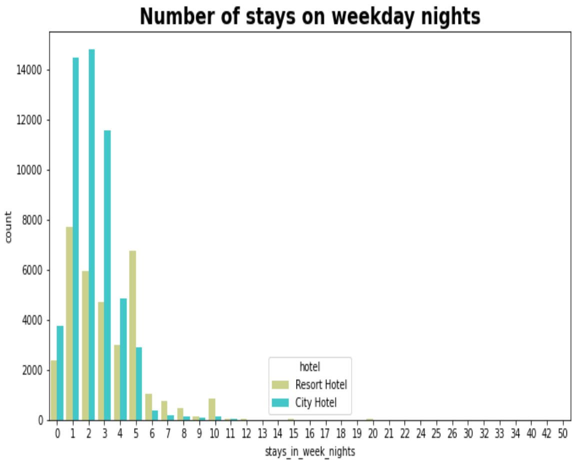


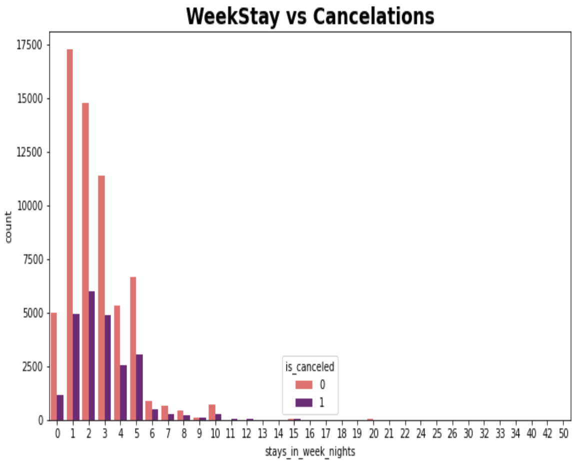


**In the first graph we can see that most of the weekend nights were booked in City Hotel**

**Second plot shows most of weekend nights which were booked were not cancelled.**

## Hotels are having most booking in a weekdays night and then cancelled:-

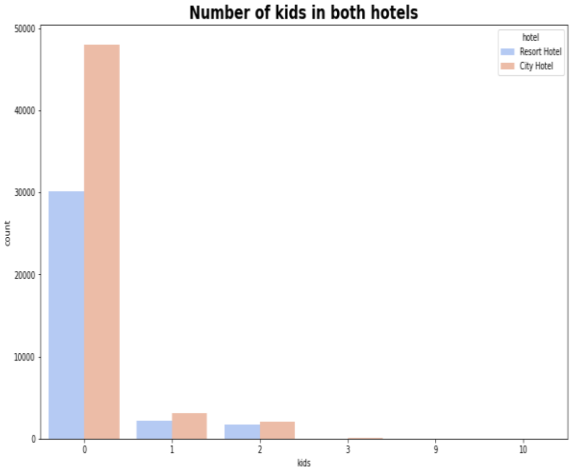


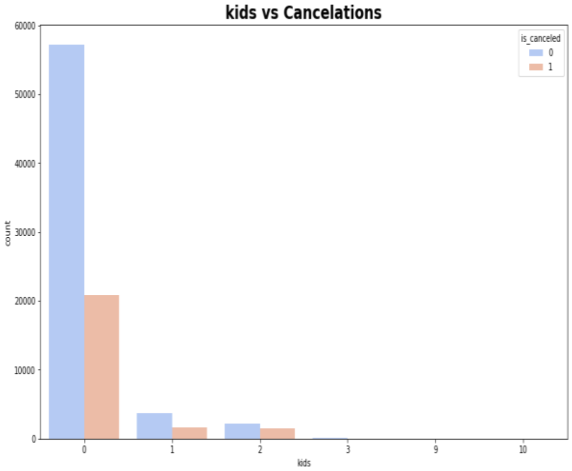


1) Weekday night stays were more in City Hotel.

2) Less cancelations were observed.

# Kids are preferred in a Hotel :-





**1) Most visitors were arrived in pair with no kids and preferred**

**2) City hotel over resort hotel**

**3) visitors who had 1 or 2 children also preferred city hotel**

# Conclusion:-

1) Majority of the guests are from Western Europe. We should spend a significant amount of our budget on those areas. Encourage Direct bookings by offering special discounts

2) Majority of the hotels are booked by city hotels. Definitely need to spend the most targeting fund on those hotels.

3) The number of repeated guests are too low. we should target our advertisement on guests to increase returning guests.

4) The majority of reservations converts into successful transactions.

5) We have also realise that the high rate of cancellations can be due to high no deposit policies.

6) We should also target months between May to Aug. Those are peak months due to the summer period.

**References-**

1. Pandas user guide: https://pandas.pydata.org/docs/user_guide/index.html

2. Matplotlib user guide: https://matplotlib.org/3.3.1/users/index.html

3. Seaborn user guide & tutorial: https://seaborn.pydata.org/tutorial.html

4. Analytics Vidya :- https://www.analyticsvidhya.com/blog/2022/04/exploratory-data-analysis-eda-in-python/