**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**

Web application for hotel reservation. We will take the selection criteria from user and display the hotels list for user basing on the criteria. User can book the room if there is availability of the rooms in that particular hotel. There are three different types of user roles for the application they are administrator hotel agent and normal user. Following are the actions provided for each user.

**Normal user:**

 Can register for the site

 Search the hotel details basing on the criteria.

 Book the hotel room

 Can modify the self-details.

**Hotel Agent:**

 Can register for the site

 Can add/update the details of the hotel.

**Administrator:**

 Will approve the new hotel details added to the application

 Can delete the user/hotel details.

**Minimal Features**:

 Adding Hotel information such as hotel name, location, number of rooms , facilities etc. to the database

 Listing the hotels based on different criteria selected by the user.

 User able to select a hotel and book a room.

 Booking permitted only if there are rooms available

 Displaying the reservation status.

 Registration of users.

 Update user details.

 Modify hotel details.

 Approval of the details entered by the hotel agent.

 Delete user/hotel details by admin

**Additional Features**:

 Taking Feedback from user.

 Rating the hotels based on the feedback.

**2.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

**3. Steps involved:**

**3.1**. **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.

**3.2. 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.

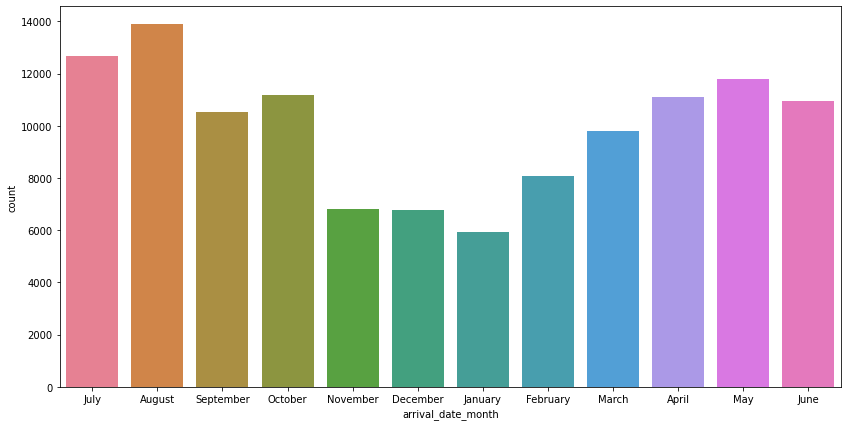
**3.3. EDA:**

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:

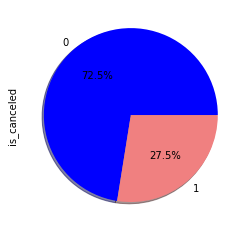
* Bar Plot.
* Histogram.
* Scatter Plot.
* Pie Chart.
* Line Plot.
* Heatmap.
* Box Plot

**Confirmed bookings in a month:-**

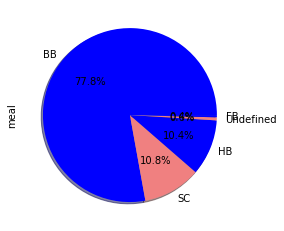
**August Month 13877 is high confirmed booking hotel. July is 2nd hight 12661 booking confirmed.**

**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.**



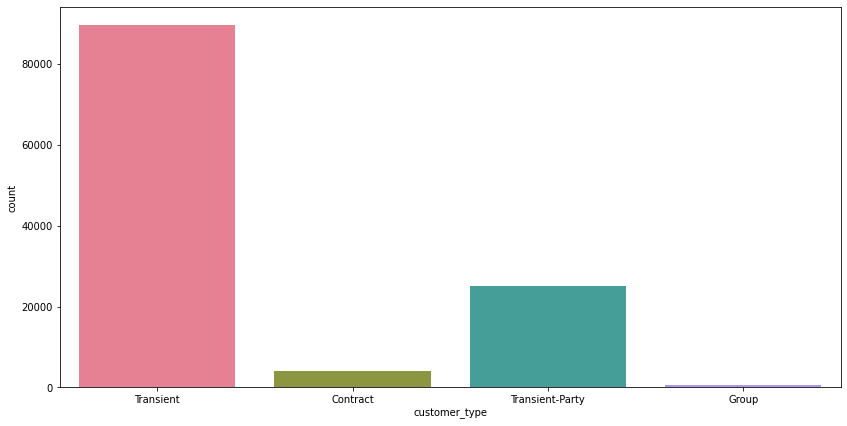
**Type of meals booked:-**

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

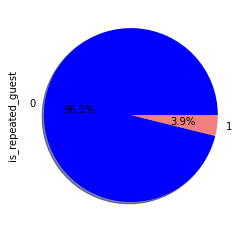
**Undefined/SC — no meal package; BB — Bed & Breakfast; HB — Half board (breakfast and one other meal — usually dinner); FB — Full board (breakfast, lunch and dinner)**

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

**Customer type are having most booking**

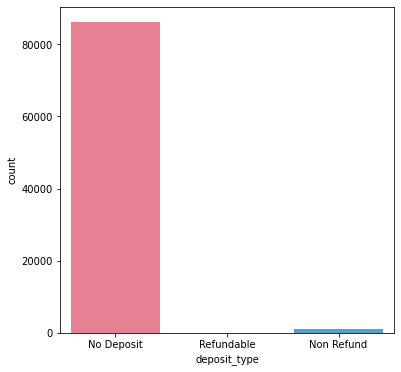
****

**Repeated guest:-**

****

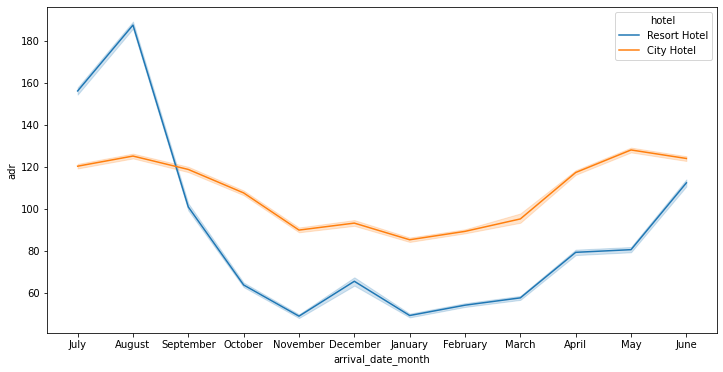
**Almost 4% of guest are coming repeatedly.**

**Deposit Type Indication:-**

****

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

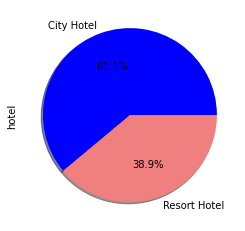
**Months having most expensive hotels:-**

****

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

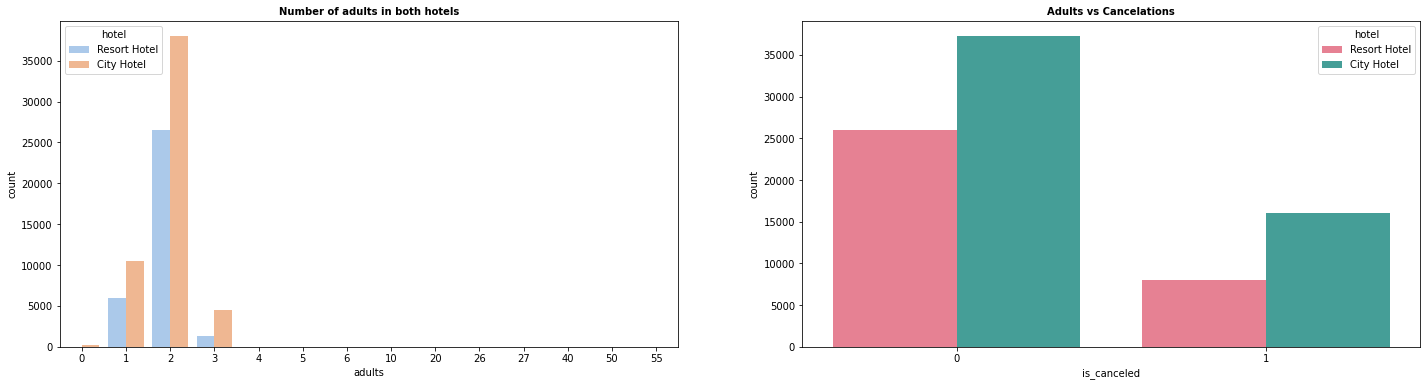
**For city hotels, the average daily rate is more expensive during august, july, june and may.**

**Hotel type are having most reservation:-**

****

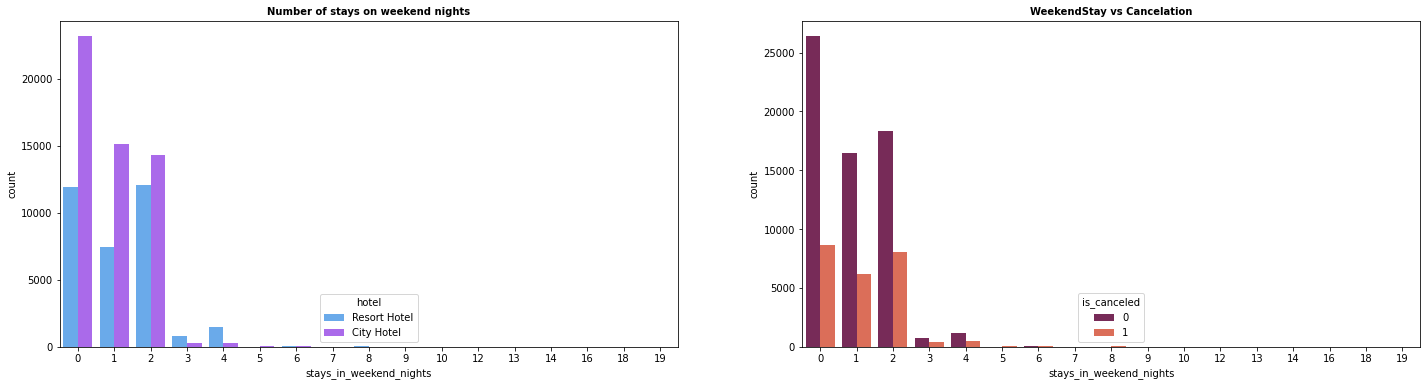
**City hotel is having most prefer customers.**

**Hotels are preferred by adults:-**



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

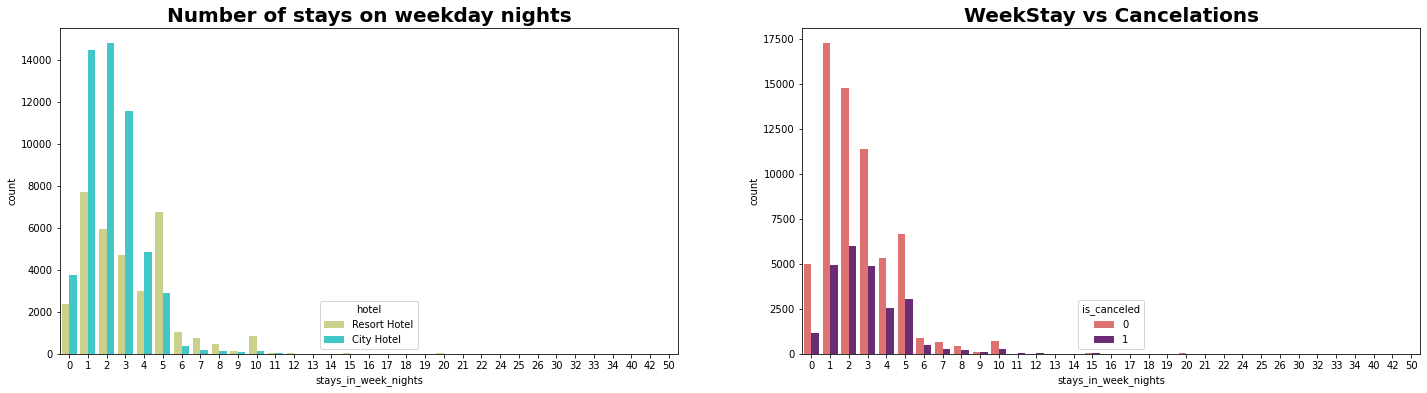
**Hotels are having most booking in a weekend nights and then cancelled:-**

****

**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.**

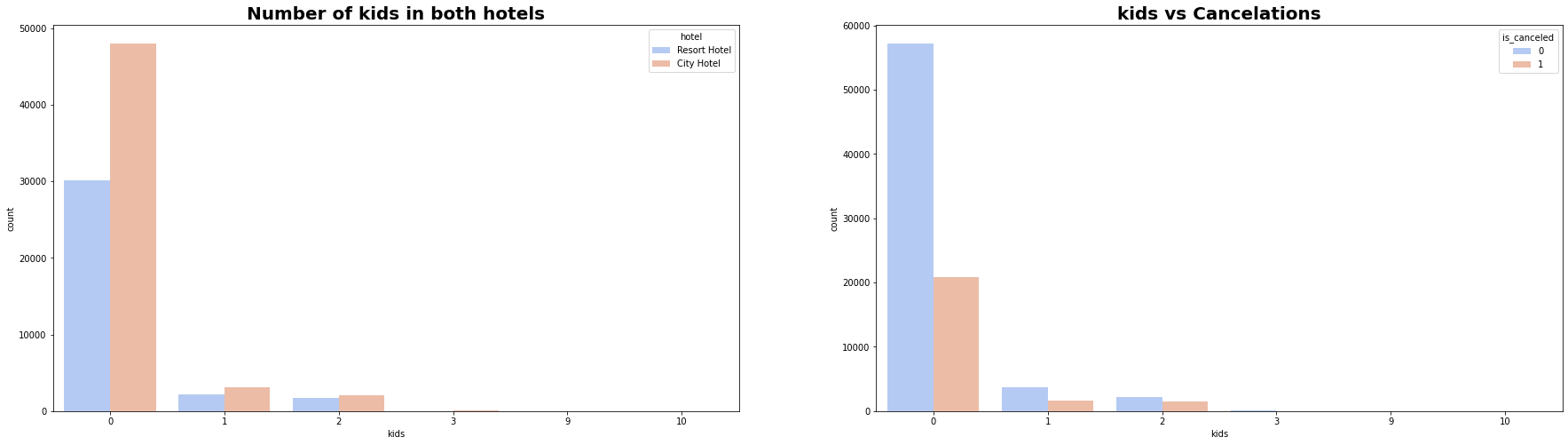
**Hotels are having most booking in a weekdays night and then cancelled:-**

****

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/>