**Hotel Booking analysis**

**Sundar K**

**Data science trainee**

**AlmaBetter**

|  |
| --- |
| **Contents** |

|  |  |
| --- | --- |
| **1.** | **Introduction** |
| **2.** | **Problem Statement** |
| **3.** | **Data Description** |
| **4.** | **Data Features** |
| **5.** | **Steps involved** |
| **6 .** | **Exploratory Data Analysis** |
| **7.** | **Result** |
| **8.** | **Conclusion** |

1. **Introduction**

When we hear about hotel booking we used many application for hotel booking. 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, cleaniness of hotel room and of course, availability of free wifi. 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).

1. **Problem statement**

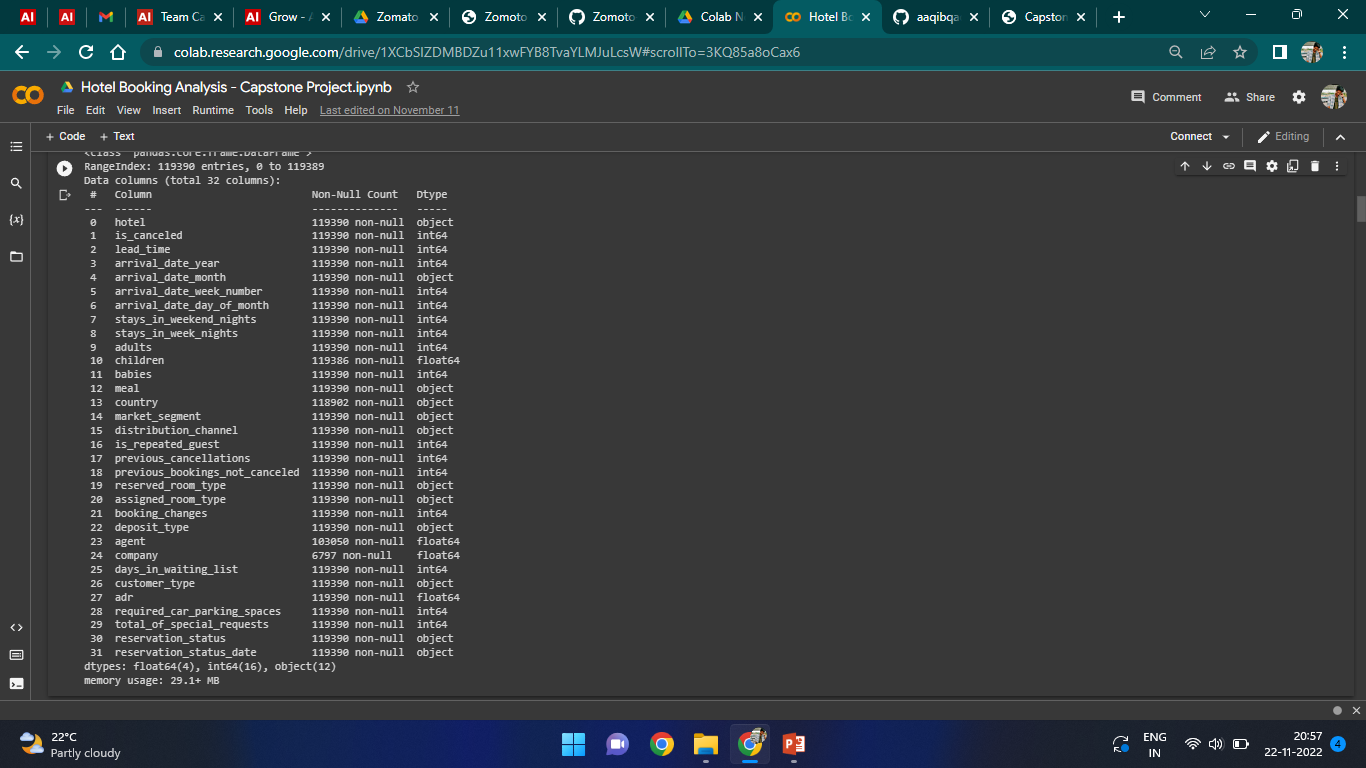
This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. All personally identifying information has been removed from the data.

Explore and analyze the data to discover important factors that govern the bookings.

1. **Data Description**

* **Hotel**
* H1: Resort hotel
* H2: City hotel
* **is\_canceled**
* 1: Canceled
* 0: Not canceled
* **lead\_time**
* No of days thaat elapsed between entering date of booking into property management system and arrival date
* **arrival\_date\_year**
* Year of arrival date (2015-2017)
* **arrival\_date\_month**
* Month of arrival date (Jan - Dec)
* **arrival\_date\_week\_numberr**
* Week number of year for arrival date (1-53)
* **arrival\_date\_day\_of\_month**
* Day of arrival date
* **stays\_in\_weekend\_nights**
* No of weekend nights (Sat/Sun) the guest stayed or booked to stay at the hotel
* **stays\_in\_week\_nights**
* No of week nights (Mon - Fri) the guest stayed or booked to stay at the hotel
* **Adults**
* **Children**
* **Babies**
* **meal**
* Type of meal booked. 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)
* **country**
* **market\_segment** (a group of people who share one or more common characteristics, lumped together for marketing purposes)
* TA: Travel agents
* TO: Tour operators
* **distribution\_channel** (A distribution channel is a chain of businesses or intermediaries through which a good or service passes until it reaches the final buyer or the end consumer)
* TA: Travel agents
* TO: Tour operators
* **is\_repeated\_guest** (value indicating if the booking name was from repeated guest)
* 1: Yes
* 0: No
* **previous\_cancellations**
* Number of previous bookings that were cancelled by the customer prior to the current booking
* **previous\_bookings\_not\_canceled**
* Number of previous bookings not 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 to hotel operation reasons (e.g. overbooking) or by customer request. Code is presented instead of designation for anonymity reasons.
* **booking\_changes**
* Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation
* **deposit\_type**
* Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: No Deposit – no deposit was made; Non Refund – a deposit was made in the value of the total stay cost; Refundable – a deposit was made with a value under the total cost of stay.
* **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. ID is presented instead of designation for anonymity reasons
* **day\_in\_waiting\_list**
* Number of days the booking was in the waiting list before it was confirmed to the customer
* **customer\_type**
* Contract - when the booking has an allotment or other type of contract associated to it;
* Group – when the booking is associated to a group;
* Transient – when the booking is not part of a group or contract, and is not associated to other transient booking;
* Transient-party – when the booking is transient, but is associated to at least other transient booking
* **adr (average daily rate)**
* average daily rate = SumOfAllLodgingTransactionTotalNumberOfStayingNight
* **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**
* Canceled – booking was canceled by the customer;
* Check-Out – customer has checked in but already departed;
* No-Show – customer did not check-in and did inform the hotel of the reason why
* **reservation\_status\_date**
* Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to understand when was the booking canceled or when did the customer checked-out of the hotel

1. **Data Features**

****

1. **Steps Involved**

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.

* **Null values Treatment**

Our dataset contains a large number of null values which might tend to disturb our accuracy hence we dropped them at the beginning of our project inorder to get a better result.

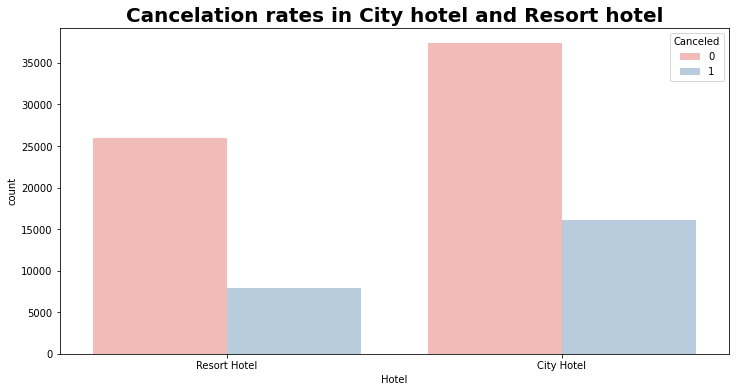
* **Encoding of categorical columns**

We used One Hot Encoding to produce binary integers of 0 and 1 to encode our categorical features because categorical features that are in string format cannot be understood by the machine and needs to be converted to numerical format.

1. **Exploratory Data Analysis**

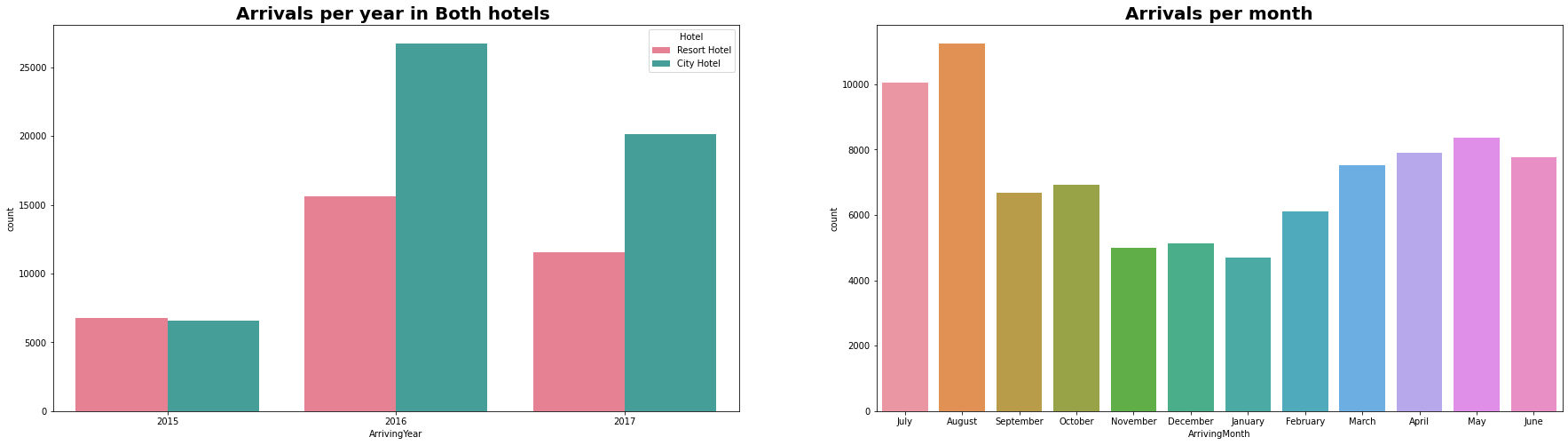
* **Cancelation rates**

Comparisons of total bookings and cancellations are conducted in this section.

****

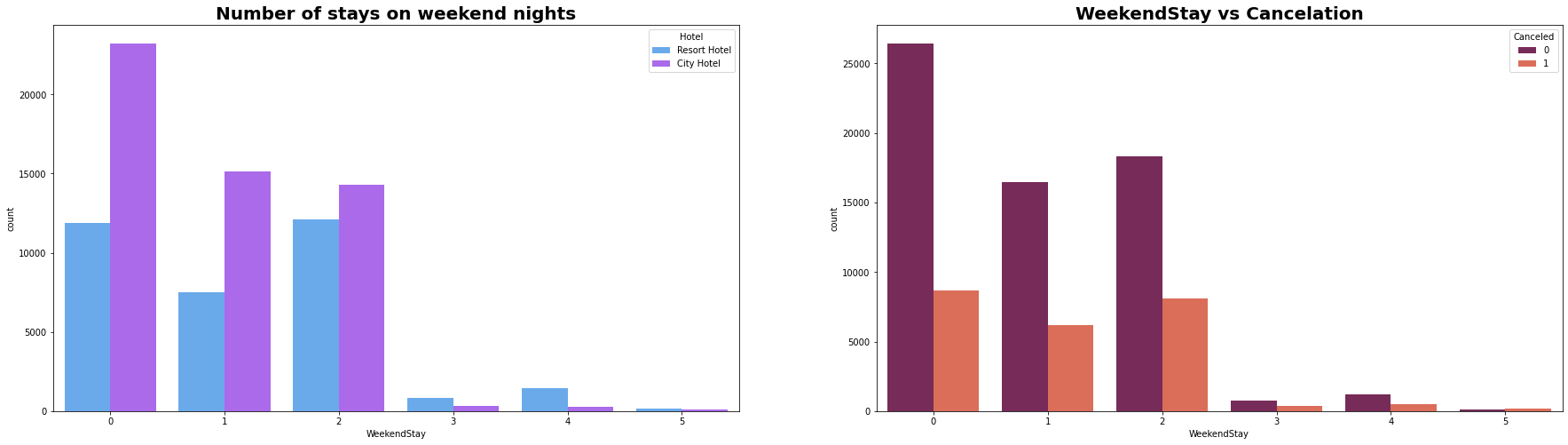
* **Arrivals per year and in month**

Comparisons of arrivals per year in Both hotels and arrivals per months



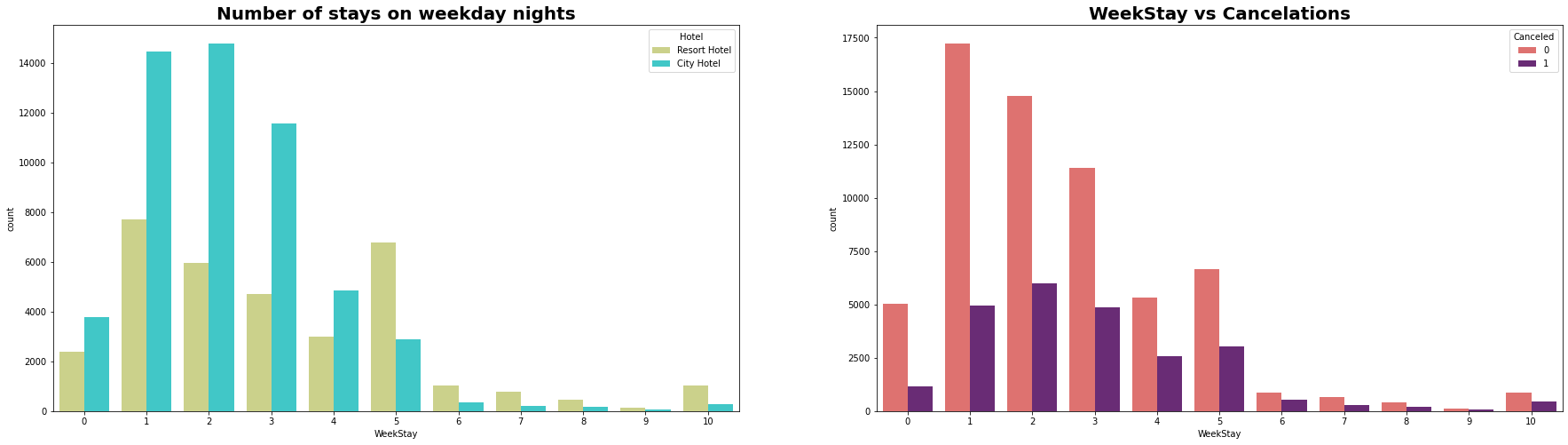
* **No of weekend stays**

Comparisons of weekend day and night stays in both the hotels



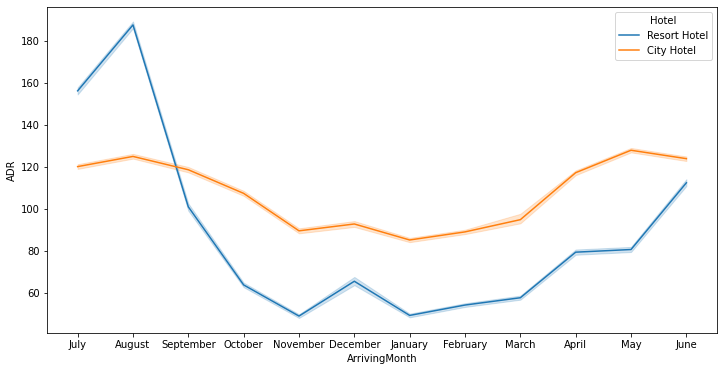
* **No of weekday stays**

Comparisons of weekday day and night in both the hotels



1. **Result**

This is the result after comparing both the hotels weekend stays , weekday stays , cancellations , yearly customer visit etc…



1. **Conclusion**

● Around 60% bookings are for City hotel and 40% bookings are for Resort hotel, therefore City Hotel is busier than Resort hotel. Also the overall adr of City hotel is slightly higher than Resort hotel.

● Mostly guests stay for less than 5 days in hotel and for longer stays Resort hotel is preferred.

● Both hotels have significantly higher booking cancellation rates and very few guests less than 3 % return for another booking in City hotel. 5% guests return for stay in Resort hotel.

● Most of the guests came from european countries, with most no. of guest coming from Portugal.

● Guests use different channels for making bookings out of which most preferred way is TA/TO.

● For hotels higher adr deals come via GDS channel, so hotels should increase their popularity on this channel.

● Almost 30% of bookings via TA/TO are cancelled.

● Not getting same room as reserved, longer lead time and waiting time do not affect cancellation of bookings. Although different room allotment do lowers the adr.

● July- August are the most busier and profitable months for both of hotels.

● Within a month, adr gradually increases as month ends, with small sudden rise on weekends.

● Couples are the most common guests for hotels, hence hotels can plan services according to couples needs to increase revenue.

● More number of people in guests results in more number of special requests.

● Bookings made via complementary market segment and adults have on average high no. of special request.

● For customers, generally the longer stays (more than 15 days) can result in better deals in terms of low adr