HOTEL BOOKING ANALYSIS

Team Mates- Rutvina Papadkar, Sagar Khairkar, Nayan Rewatkar

**AlmaBetter Capstone Project**

# Abstract:

The hotel industry has a very convenient ways of booking arrangements with different distribution channels. We worked on the hotel industry dataset to get a comprehensive image of the hospitality industry. From basic data visualizations to complex analysis, we have performed data analysis at every level of the dataset to draw out key insights about the industry.

We are grateful to Alma Better for giving us the opportunity to work on such an exciting project.

**PROBLEM STATEMENT:**

Main motive in hotel industry is hospitality because you have to be able to identify the problems guests are having, create a plan of action to resolve them and follow through with guests afterward to ensure their satisfaction. We have been provided a dataset of the hotel industry of the region in which our hotels aim to begin its initial operations. Our primary goal is to understand the market of this particular region and try to use our data analysis techniques to draw out key features of the market. Our secondary goal is to draw out actionable insights from our analysis and give conclusions about key aspects of the market such as cancellation rate, distribution channels, and alike.

# Data Summary:

On performing primary assessment, we found that the data was much clean except for some missing values in some columns. On using the info() method, we draw out the following key insights about the data:-

1. The dataset has a shape of (119390, 32) which means that it contains approximately 1.2 lakh rows and 32 columns.
2. Our Dataset has 4 columns with float64 dtype, 16 columns with int64 dtype, and 12 columns with object dtype.
3. In our Dataset, we observed null values in the following columns:
   * 4 null values in the children column
   * 488 null values in the country column
   * 16,340 null values in the agent column
   * 112,593 null values in the company column

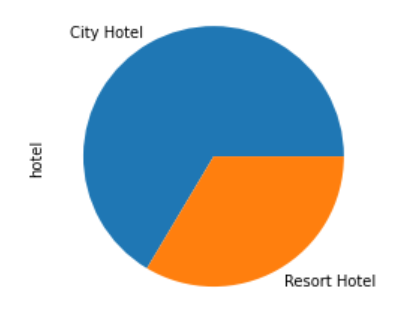
Following are the column names present in Dataset:

* hotel
* is\_canceled
* lead\_time
* arrival\_date\_year
* arrival\_date\_month
* arrival\_date\_week\_number
* arrival\_date\_day\_of\_month
* stays\_in\_weekend\_nights
* stays\_in\_week\_nights
* adults
* children
* babies
* meal
* country
* market\_segment
* distribution\_channel
* is\_repeated\_guest
* previous\_cancellations
* previous\_bookings\_not\_canceled
* reserved\_room\_type
* assigned\_room\_type
* booking\_changes
* deposit\_type
* agent
* company
* days\_in\_waiting\_list
* Customer\_type
* adr
* required\_car\_parking\_spaces
* total\_of\_special\_requests
* reservation\_status
* reservation\_status\_date

# Steps involved in the Data Analysis: -

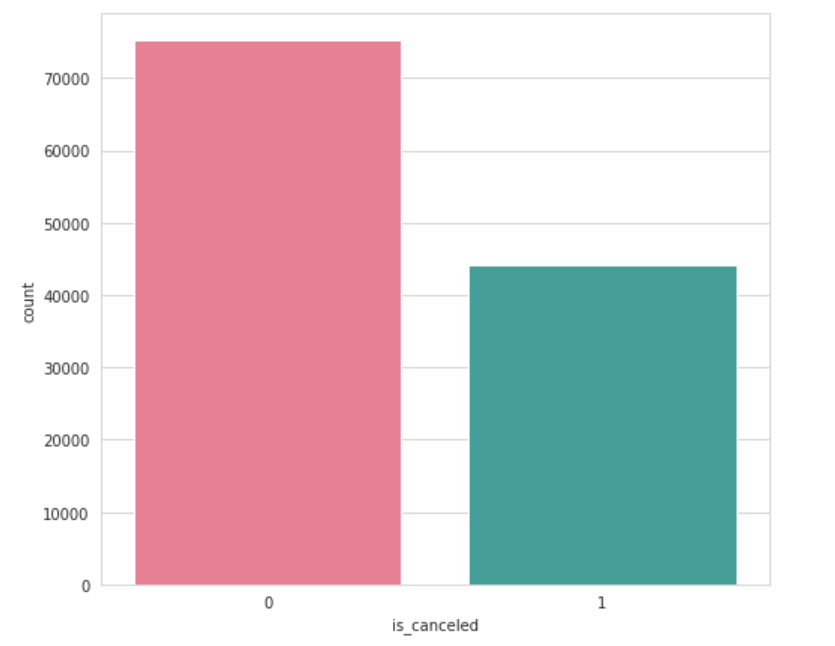
1. Data Collection
2. Data filtering
3. Data cleaning
4. Exploratory data analysis
5. Data visualization

**Type of Hotels**

****

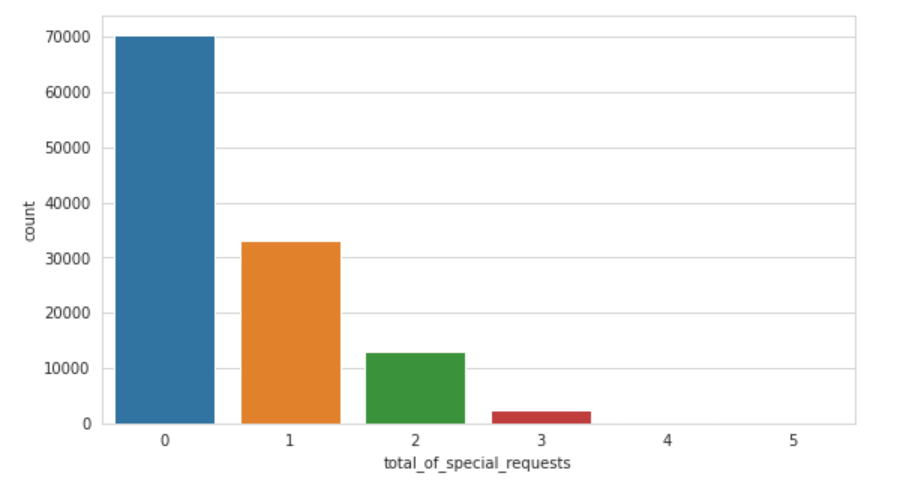
City hotel has booked in majority than Resort hotel

**Calculate the booking cancellation percentages**

****

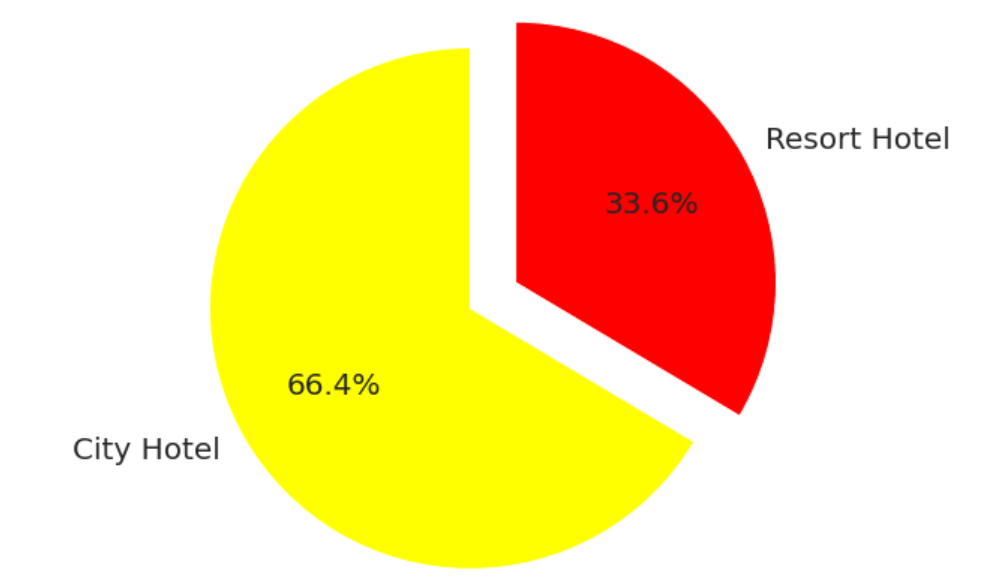
***We can conclude that the 47% of cancellation were done in a year***

**Special Request**

****

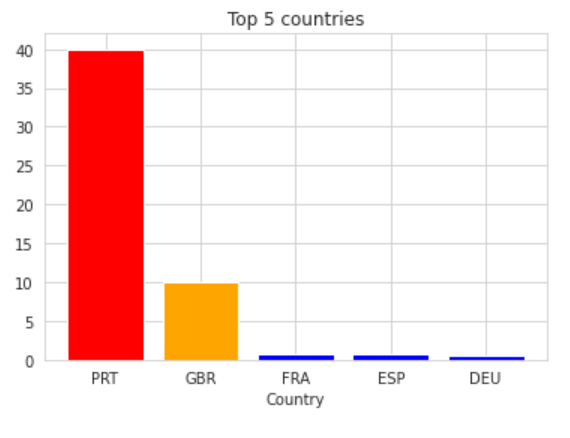
we can see 1st special request has made 26% of total bookings. 2nd special request has made 10% among all and 1% for 3rd special request.

**Booking ratio between City hotel and Resort hotel**

****

City Hotel has highest ratio of booking compare to resort hotel

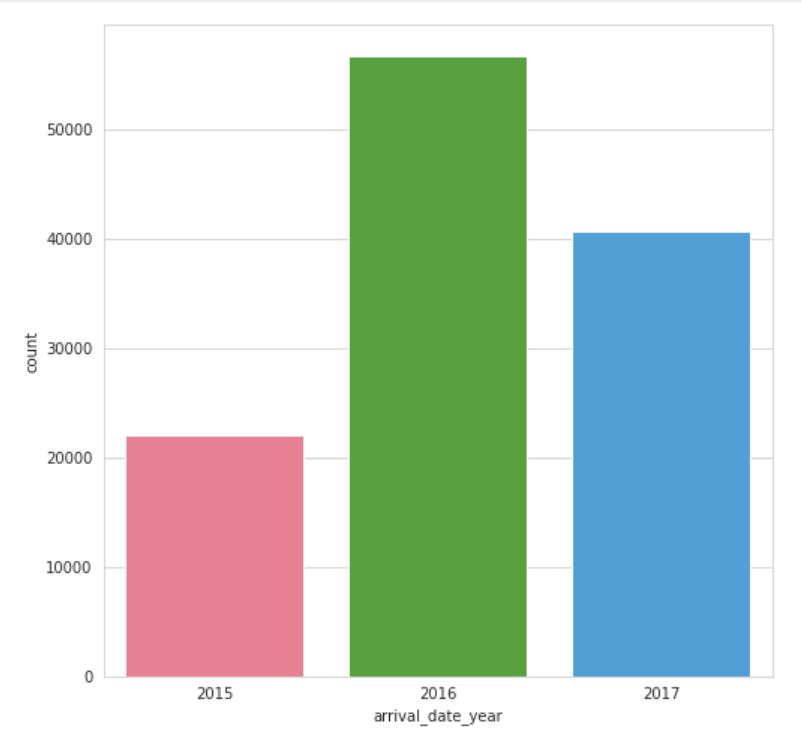
**Country origin of most guest**

****

Portugal has highest 40.69% of the guest, followed by Great Britain with 10.15%, France, Spain and Germany.

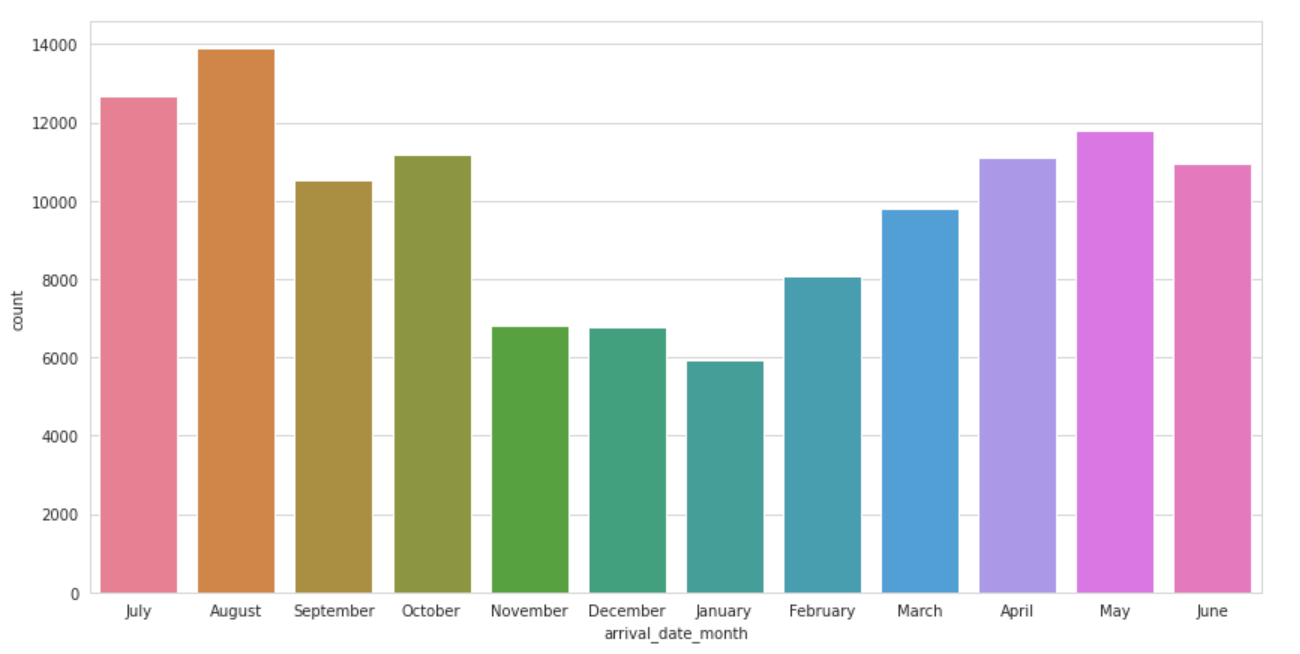
Majority of the guests are from European Countries

**Booking Per Year**

****

* There has many arrivals in the year 2016 than any other year.

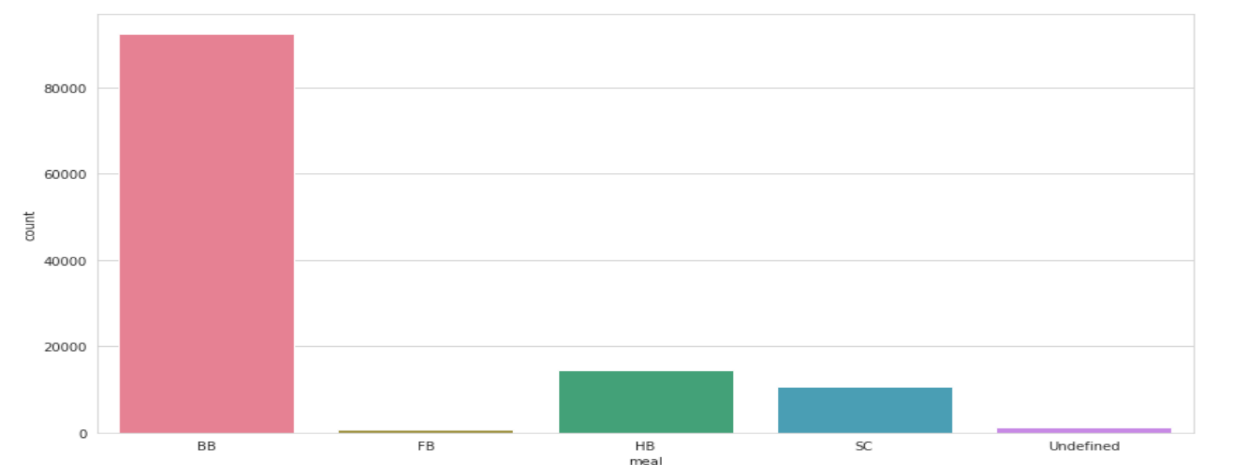
**Busiest Month for Hotels**

****

The Month with highest booking is August which has 11.62% reservation. The month with lowest booking is January which has 4.96% of reservation.

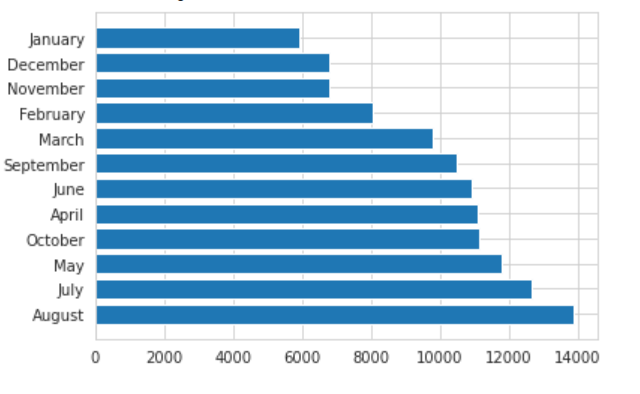
We can say that the busiest month for hotels is August which has 11.62% of reservation

**Meal Type**

****

* BB( Bed & Breakfast) is the most preferred type of meal by the guests.
* Full Board i.e. FB is least preferred.
* HB (Half Board) and SC(Self Catering) are equally preferred.

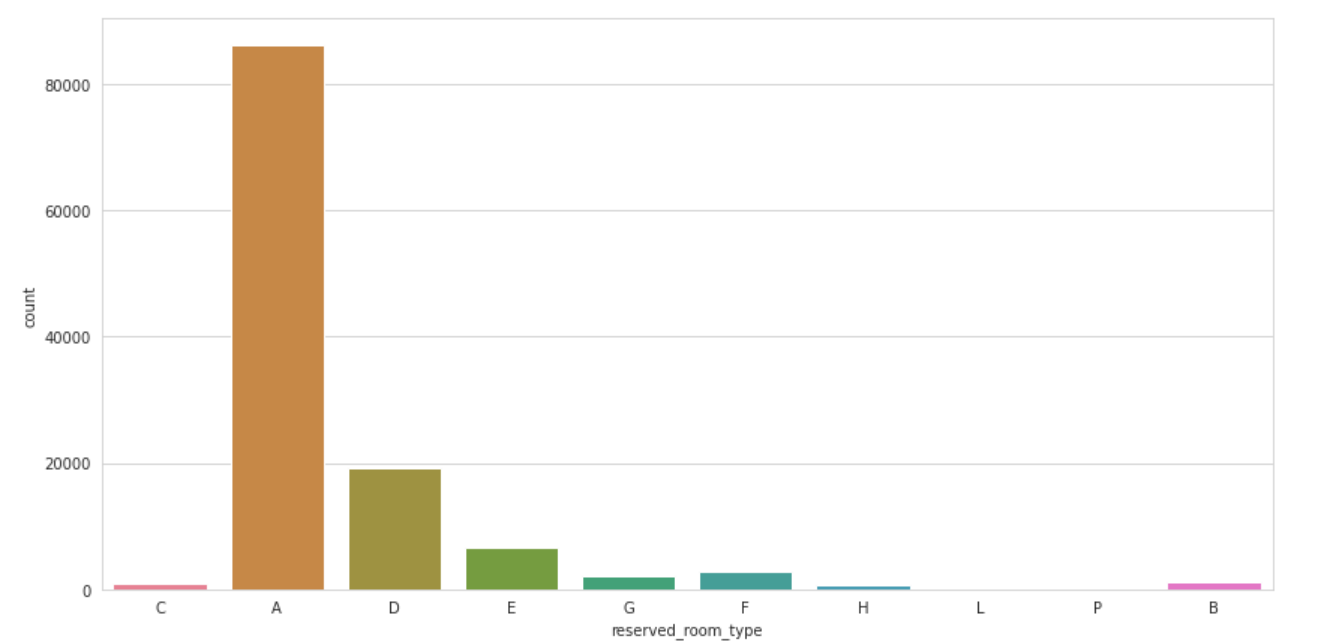
**Number of travelers in various month**

****

**Peak Months are August, July and May.**

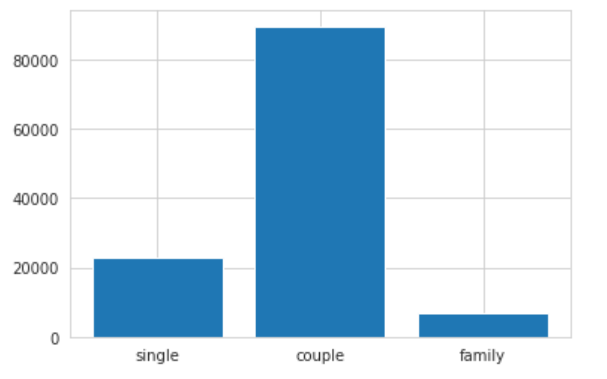
* Maximum number of travellers travels in August month.
* Minimum number of travellers travels in January month.

**Room Types**

****

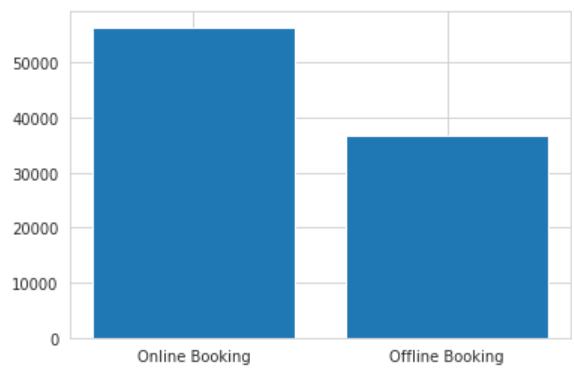
The room type “A” is selected by most of the guest

**Room booked for family, couple and single**

****

we can say that most of rooms are booked for couples

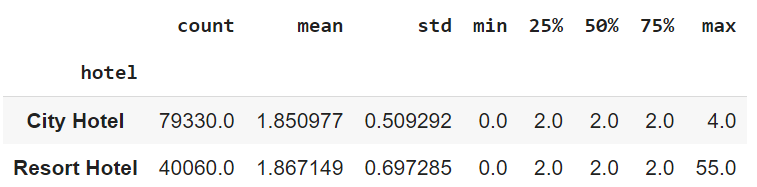
**Online Vs Offline Booking**

****

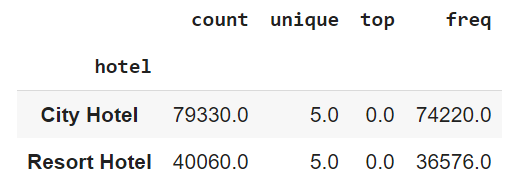
customers prefer to book online hotel rather than offline one.

**People who booked the hotel**

**Bookings Done by Adults**

****

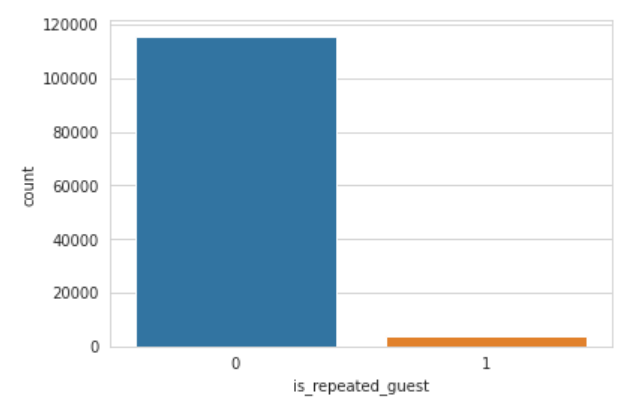
**Bookings done by Children’s**

****

● It seems that mean values for adults and children are higher.

● This means that resort hotels are better choice for large families.

**Repeated Guests**

****

The number of guests rarely visited the same hotel.

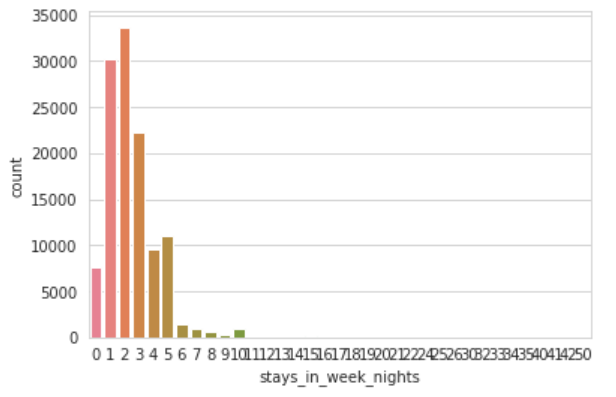
**Reservation Status**

****

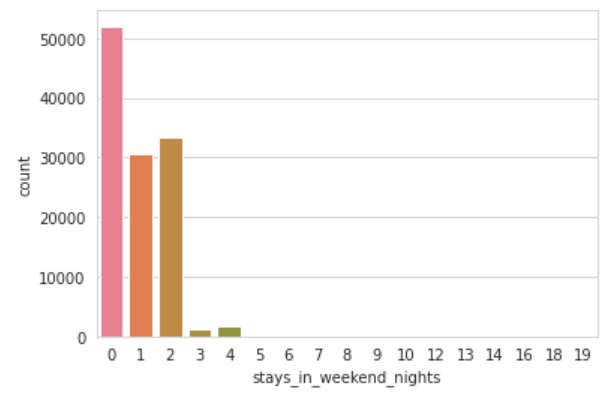
50% of guest has canceled their booking which is very high.

**Weekend Vs Weekdays**

**Average stays on weekdays**

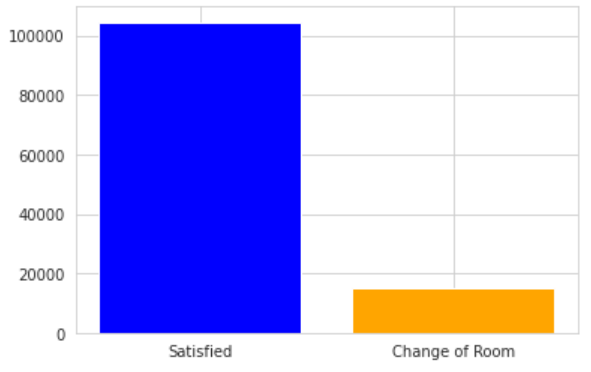
****

**Average stays on weekend**

****

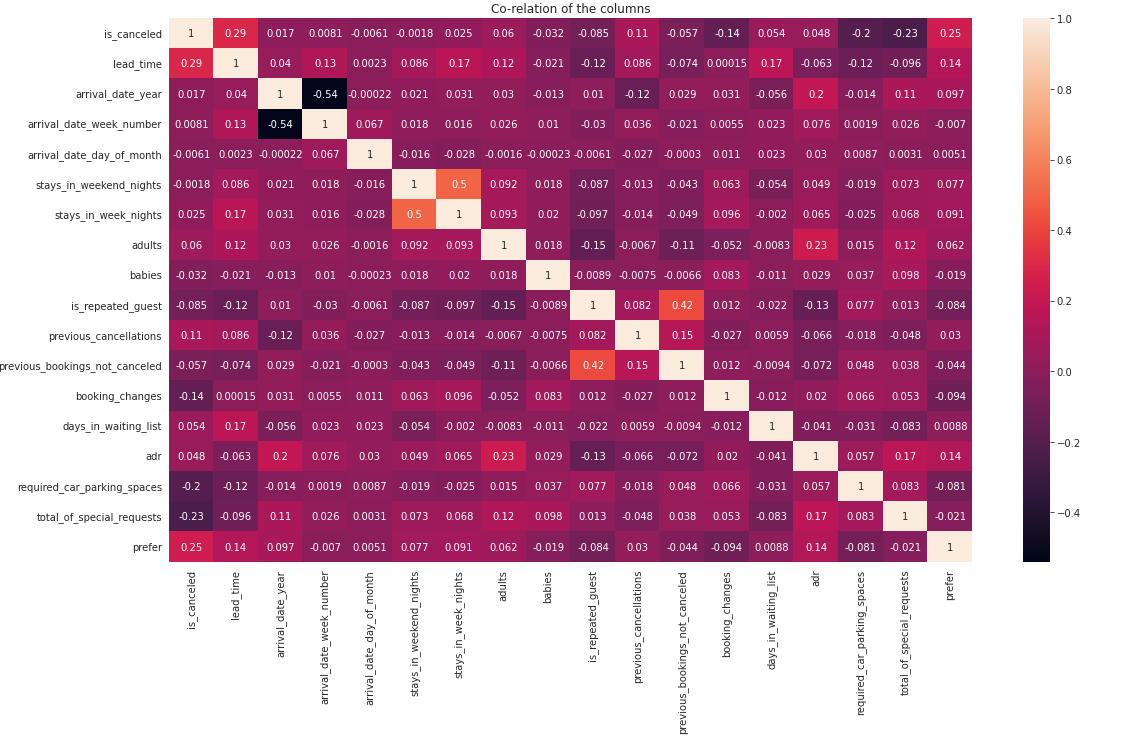
The count for booking on weekdays is higher than that of weekends with more than 50000.

**Customers satisfactory**



Majority of the customers has got the room type they have booked

**Co-relation of Columns**

****

* Is\_canceled and same\_room\_alloted\_or\_not are negatively correlated. Not getting the same room as per reserved room is not the reason for booking cancellations.
* lead-time and total stay is positively correlated means more is the stay of customer more will be the lead time.

**Conclusion**

* Hence we conclude that the hotel bookings are expected to get increase in summer season than any other season. So, if you want to enjoy the family time then you can book in any other seasons except summer.
* It seems like people are staying for mostly 1 to 3 days, so for increasing the business you can offer the discount on long stays.
* Many bookings were done in online mode because there are many hotel booking applications are offering good discount and it is flexible to book remotely.