**Hotel Booking Analysis**

**Samara Reddy**

**Data science trainee,**

**Alma Better, Bangalore.**

**Abstract:**

The hotel industry is a is a very volatile industry and the bookings depends on so many factors and also it becoming very big industry everywhere in world, which has a very peculiar set of booking arrangements with different distribution. In this project I had worked on the hotel industry dataset contains a single file which compares various booking information between two hotels: a city hotel and a resort hotel to get a comprehensive image of the hospitality industry. From basic data visualizations to complex multivariate analysis, I had performed data analysis at every level of the dataset to draw out key insights about the industry.

First, we did inspection on our dataset, and cleaning of dataset like finding duplicate rows, null values and did exploratory data analysis on the data set to get more information from the dataset visually from the large dataset and drawn some insights from data visualization.

***Keywords: Hotel, data cleaning, data inspection, data visualization and exploratory data analysis.***

**1.Problem Statement:**

Give 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 required parking spaces, among other things.

To Explore and analyze the data to discover important factors that govern the bookings and and give insights to hotel management, which can perform various campaigns to boost the business and performance.

**2.Dataset info:**

In order to understand our data, we can look at each variable and try to understand their meaning and relevance to this problem. The hotel booking dataset have 119390 rows and 32 columns having all the columns with data type of object, int, float, or bool.

**3.Dataset Summary:**

The main objective of Exploratory data analysis is to understand trend and behaviour of guest in hotel bookings. For staring analysing the data it’s very important to understanding our data. So we had hotel Booking analysis data. Which had 119390 rows and 32 columns. So, let’s understand this 32columns.

**Hotel: Resort Hotel or City Hotel**

**is\_canceled : Value indicating if the booking was canceled (1) or not (0)**

**lead\_time : Number of days that elapsed between the entering date of the booking and the arrival date**

**arrival\_date\_year : Year of arrival date**

**arrival\_date\_month : Month of arrival date**

**arrival\_date\_week\_number : Week number of year for arrival date arrival\_date\_day\_of\_month : Day of arrival date**

**stays\_in\_weekend\_nights : Number of weekend nights**

**stays\_in\_week\_nights : Number of week nights.**

**adults : Number of adults**

**children : Number of children**

**babies : Number of babies**

**meal : Type of meal booked.**

**country : Country of origin.**

**market\_segment : Market segment designation. (TA/TO)**

**distribution\_channel : Booking distribution channel.(T/A/TO)**

**is\_repeated\_guest : is 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 : Number of previous bookings not cancelled by the customer prior to the current booking**

**reserved\_room\_type : Code of room type reserved.**

**assigned\_room\_type : Code for the type of room assigned to the booking.**

**booking\_changes : Number of changes made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation**

**deposit\_type : No Deposit, Non Refund , Refundable.**

**agent : ID of the travel agency that made the booking**

**company : ID of the company/entity that made 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 customer. Contract, Group, transient, Transient party.**

**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**

**4. Introduction:**

The **hotel Booking** is must have for hotels these days. Today on the internet so many **hotel booking apps** are available like Oyo, trivago, goibibo many more. **Hotel booking process** must be flawless and user-friendly to attract more customers and increase sales. Every hotel’s goal is to increase their business and to earn the profit.

Hotel booking Analysis discovers the more information about a hotel which is situated in a particular area and you can also select a hotel according to your demands and choice.

**5. Steps involved:**

**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.

**Data Cleaning**: 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:

Remove duplicate rows

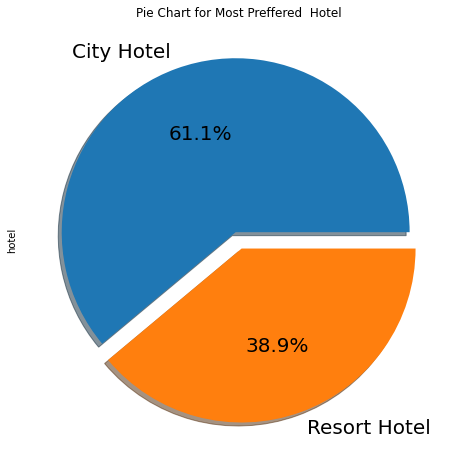
Handling Null values –

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.

**Exploratory Data Analysis (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:

**Which hotel type is the most preferred by the guests?**

****

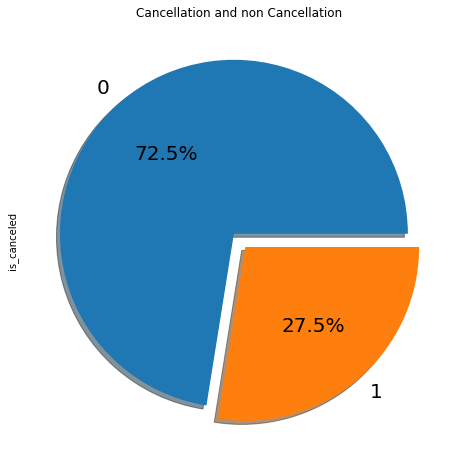
From above Pie Chart we can say City hotel is the busiest hotel

**What is the Percentage of repeated guests**

## 

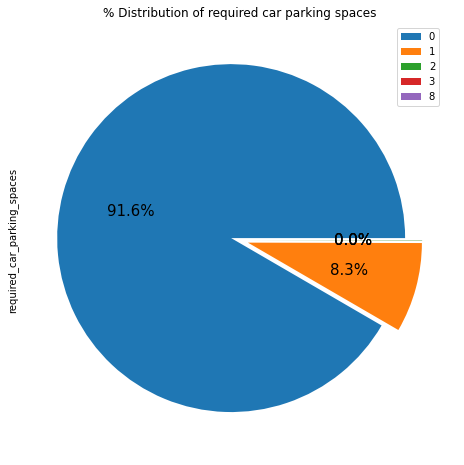
## From above Pie Chart we can say Repeated guests are very few which only 3.9 %.

**What is the percentage of cancellation**



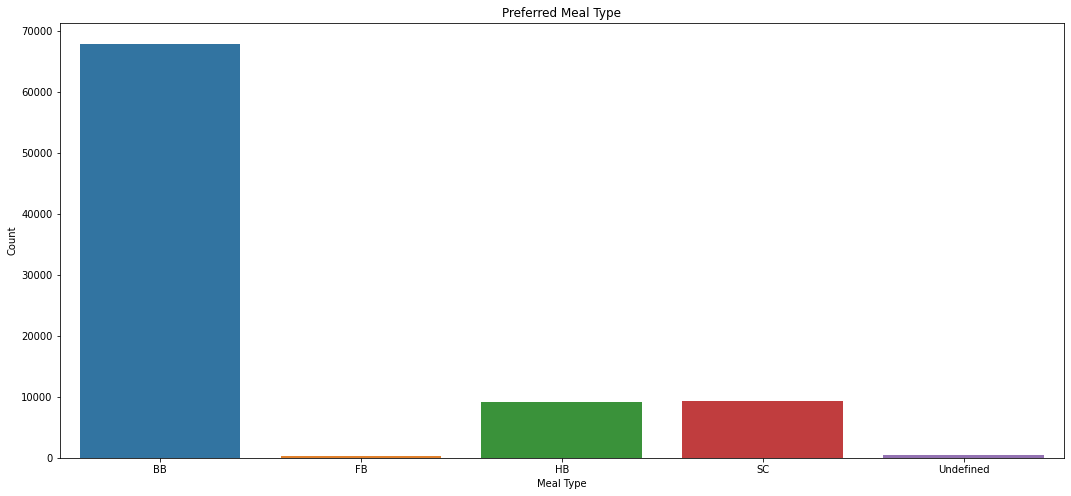
From above Pie Chart we can say 27.5 % of the bookings were cancelled.

**What is the percentage distribution of required\_car\_parking\_spaces**



From above Pie Chart we can say 91.6 % guests did not required the parking space. only 8.3 % guests required only 1 parking space**.**

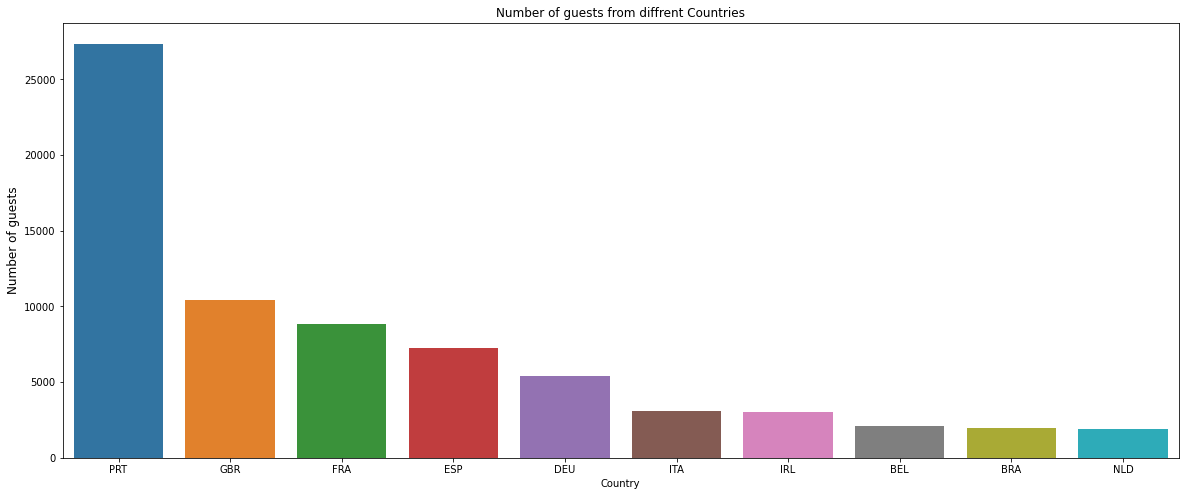
**Which type of food is mostly preferred by the guests**



From above Bar Chart we can say So the most preferred meal type by the guests is BB( Bed and Breakfast)

HB- (Half Board) and SC- (Self Catering) are equally preferred.

**From which country the most guests are coming?**

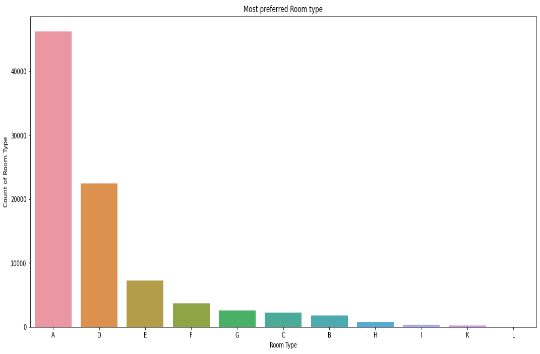
****

From above Pie Chart we can say

Most of the guests are coming from portugal i.e more 25000 guests are from Portugal

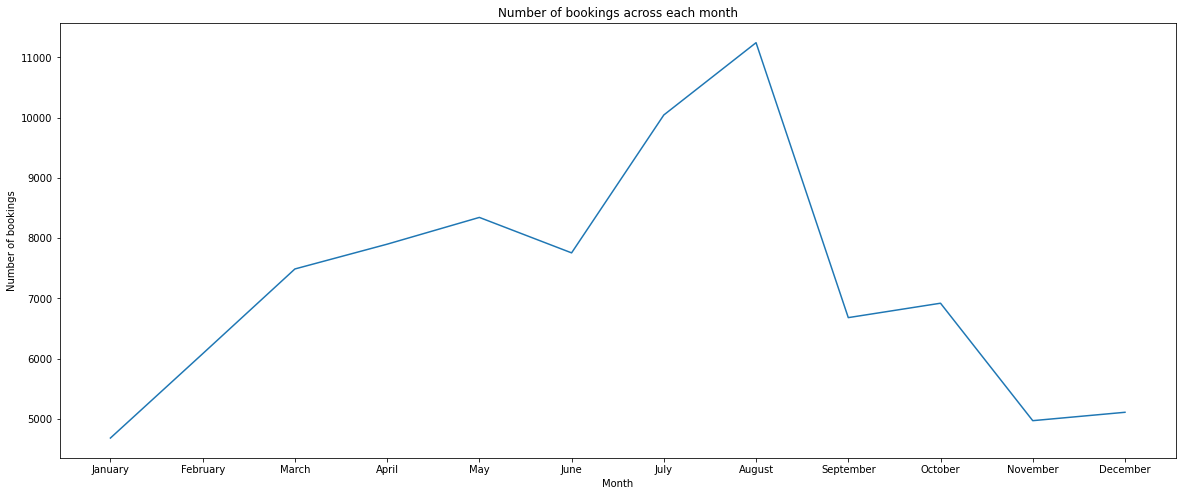
After Portugal, GBR(Great Brittan),France and Spain are the countries from where most of the guests came

**Which is the most preferred room type by the customers**



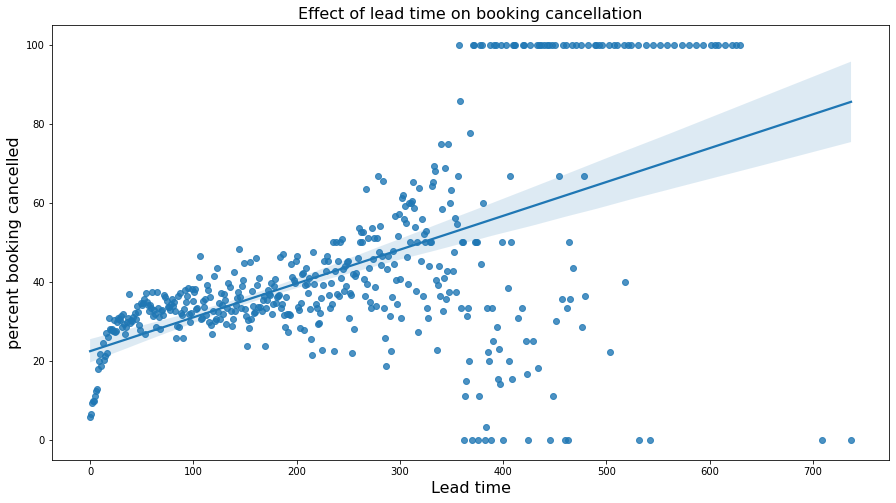
From above Pie Chart we can say the most preferred Room type is "A".

**In which month most of the bookings happened?**

****

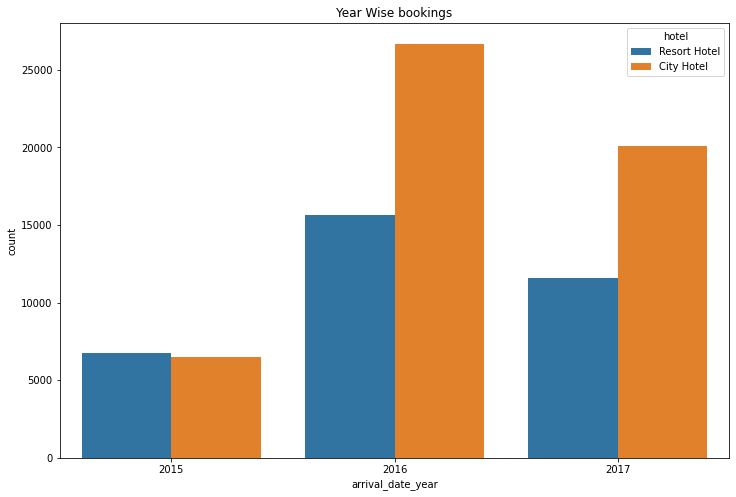
From above Line Chart we can say both August and July months had the most Bookings. Summer vacation can be the reason for the bookings.

**Let’s check Effect of lead time on booking cancellation**

****

From above Scatter plot we can say There is likelihood of lesser correlation between lead time and booking cancellation. So, more lead time is not correlated to cancellation of booking

**Which year had the highest bookings?**

****

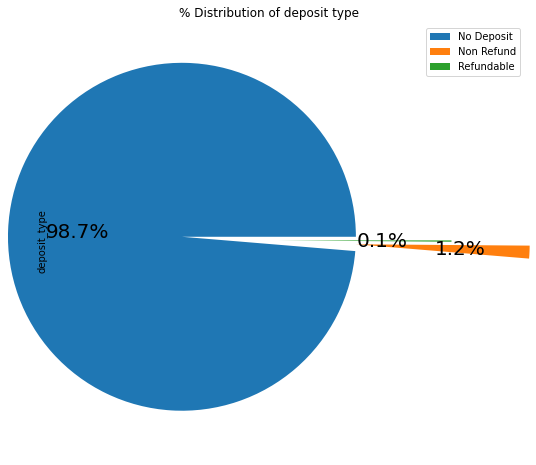
From above Chart we can say 2016 had the highest bookings. 2015 had less 7000 bookings. overall City hotels had the most of the bookings**.**

**Checking Which hotel type has the more lead time?**

****

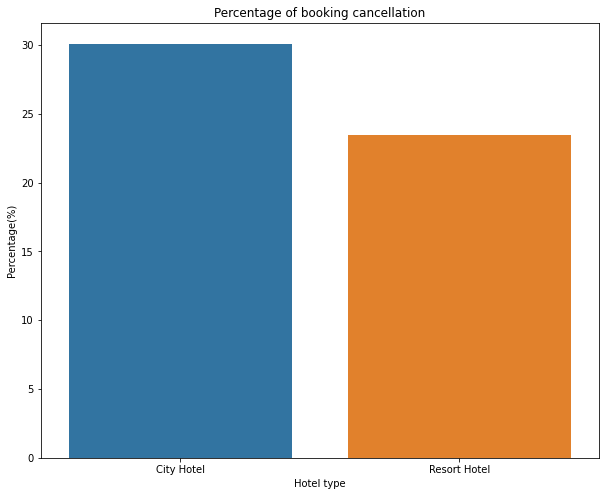
From above Bar Chart we can say Resort hotels has slightly high avg lead time. That means customers plan their trips very early.

**What is Percentage distribution of Deposit type**



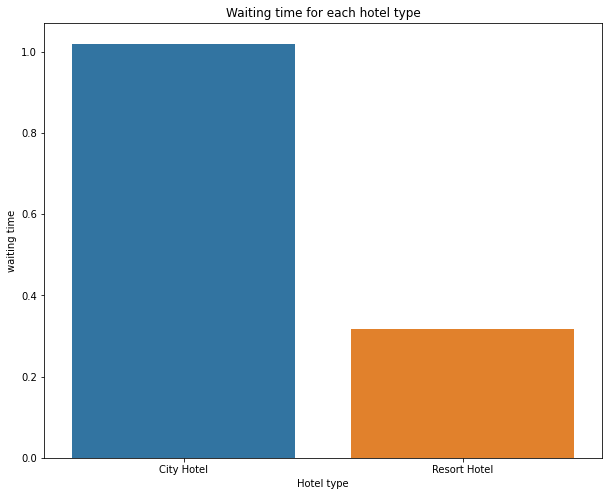
98.7 % of the guests prefer "No deposit" type of deposit.

**Let’s see Which hotel has highest percentage of booking cancellation?**

****

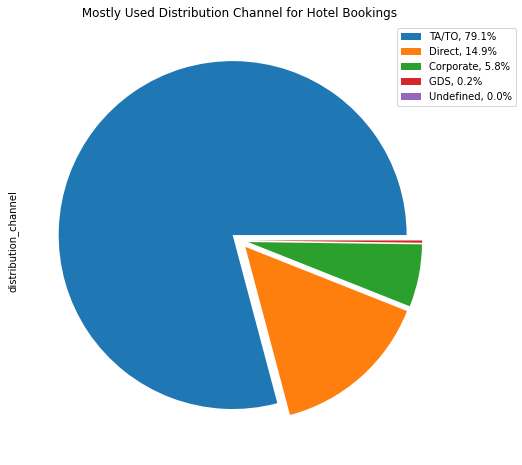
From above Bar Chart we can say Percentage of Booking cancellation rate is high for City hotels which almost 30 %.

**Which hotel has longer waiting time?**

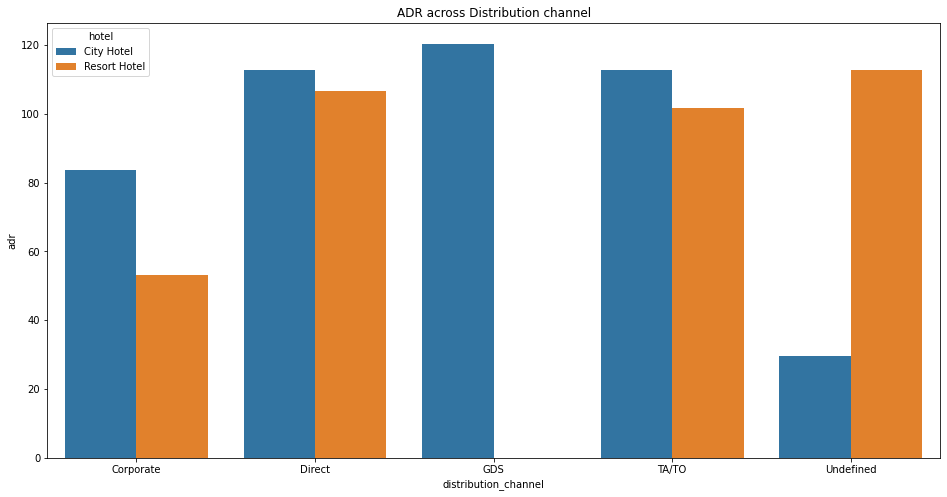
****

From above Bar Chart we can say City Hotels has longer waiting period than the Resort Hotels. Thus we can say that City Hotels are much busier than the Resort Hotels

**Which Distribution channel is mostly used for hotel bookings**

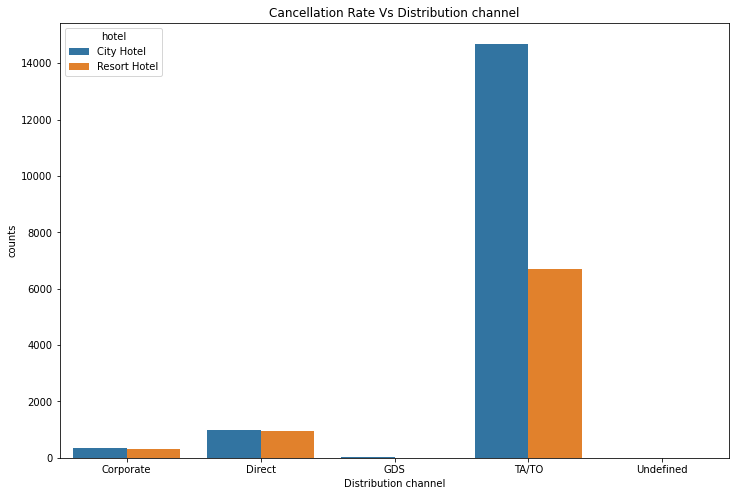
'TA/TO' is mostly (79.1%) used for booking hotels**.**

**Which distribution channel contributed more to adr in order to increase the income?**

****

* Corporate- These are corporate hotel booing companies which makes bookings possible.
* GDS-A GDS is a worldwide conduit between travel bookers and suppliers, such as hotels and other accommodation providers. It communicates live product, price and availability data to travel agents and online booking engines, and allows for automated transactions.
* Direct- means that bookings are directly made with the respective hotels
* TA/TO- means that bookings are made through travel agents or travel operators.
* Undefined- Bookings are undefined. may be customers made their bookings on arrival.

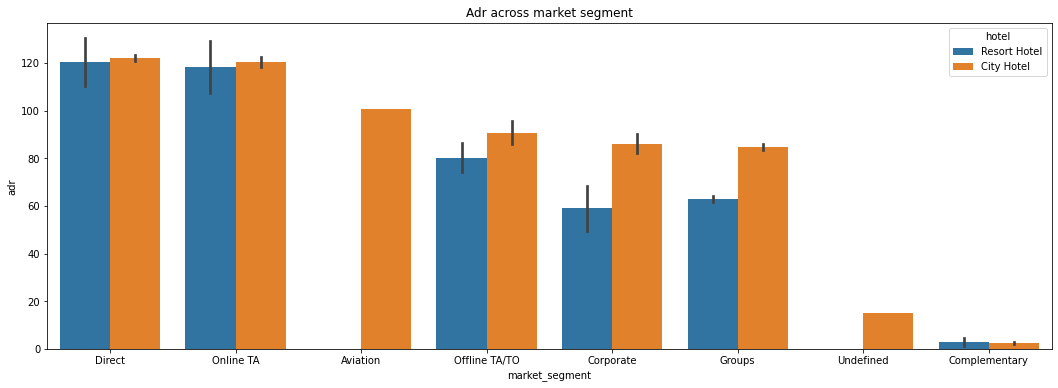
**Which distribution channel has the highest cancellation rate?**



In "TA/TO", City hotels has the high cancellation rate compared to resort hotels

In "direct" both the hotels has almost same cancellation rate**.**

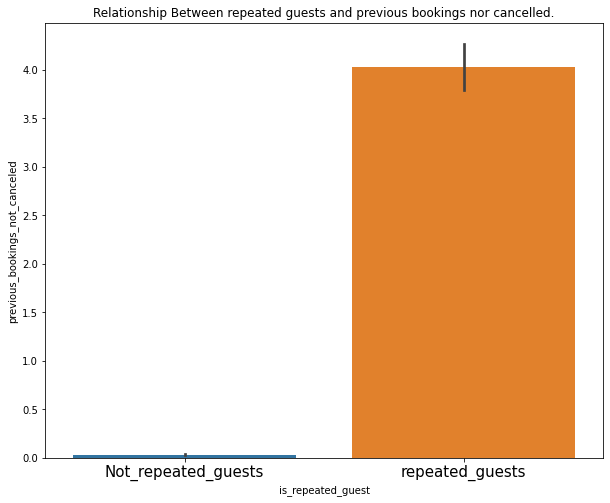
**Let’s see ADR across different market segment**



'Direct' and 'Online TA' are contributing the most in both types of hotels.

Aviation segment should focus on increasing the bookings of 'Resort Hotel'

**Let’s check Relationship between the repeated guests and previous bookings not cancel-ed**

Not Repeated guests are more likely to cancel their bookings.

**Correlation of the columns**

****From correlation figure we can observe

is\_canceled and same\_room\_alloted\_or\_not are negatively corelated. That means customer is unlikely to cancel his bookings if he don't get the same room as per reserved room. We have visualized it above.

## lead\_time and total\_stay is positively corelated.That means more is the stay of cutsomer more will be the lead time.

## adults,childrens and babies are corelated to each other. That means more the people more will be adr.

## is\_repeated guest and previous bookings not canceled has strong corelation. may be repeated guests are not more likely to cancel their bookings**.**

**Conclusion:**

In this project I explored the choice-set model as a theoretical lens to explore online hotel booking behaviours by using multiple methods (i.e., observation and survey methods). The findings of this research shed light on the dynamic patterns of online hotel decision-making process and identify important factors (i.e., internal and external information sources) across sequential stages of the choice-set model.

* City hotel is the busiest hotel.
* Repeated guests are very few which only 3.9 %.
* 27.5 % of the bookings were cancelled.
* 91.6 % guests did not required the parking space. only 8.3 % guests required only 1 parking space.
* most preferred meal type by the guests is BB( Bed and Breakfast).
* HB- (Half Board) and SC- (Self Catering) are equally preferred
* Most of the guests are coming from portugal i.e more 25000 guests are from Portugal.
* After Portugal, GBR(Great Brittan),France and Spain are the countries from where most of the guests came.
* both August and July months had the most Bookings. Summer vacation can be the reason for bookings.
* 2016 had the highest bookings. 2015 had less 7000 bookings. overall City hotels had the most of the bookings.
* Resort hotels has slightly high avg lead time. That means customer plan their trips very early.
* In "TA/TO", City hotels has the high cancellation rate compared to resort hotels.
* In "direct" both the hotels has almost same cancellation rate.
* Not Repeated guests are more likely to cancel their bookings.