**Hotel Booking Analysis**

**Priya Bodke, Awaze Mohammad**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

This data article describes two datasets with hotel demand data. One of the hotels is a Resort hotel and other is City hotel. Both datasets share the same structure,with 32 variables(columns) describing the observations.

This project contains the real world data record of hotel bookings of a hotel containing details like bookings, cancellation’s, market segments, guest details and their requirement’s etc. from 2015 to 2017.

Main aim of the project is to understand and visualize dataset from hotel and customer point of view i.e.

* Best time to book hotel
* Peak season
* Most preferred hotel type
* Cancellation rate
* Percentage of repeated guests
* Market segments
* Distribution channels etc.

Our team has worked on the hotel industry dataset to get a comprehensive image of the hospitality industry. From basic data visualizations to complex multivariate analysis, we have performed data analysis at every level of the dataset to draw out key insights about the industry.

**Problem Statement:**

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions!

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 analyse the data to discover important factors that govern the bookings.

**Data Summary:**

We have analyse and understand the dataset. we have found that dataset contains 32 columns and 119390 rows.

Some duplicate rows are present so we will drop those duplicate rows, Also we have seen some Nan values, Which will require cleaning and handling of those Nan values.

The dataset spans over three years - 2015, 2016 and 2017.

1.hotel: Denotes the type of hotel - Resort hotel or city hotel

2.is\_canceled: Indicates if the booking was canceled (1) or otherwise (0)

3.lead\_time: Number of days between the entering date of the booking and the arrival date

4.arrival\_date\_year: Year of arrival date

5.arrival\_date\_month: Month of arrival date

6.arrival\_date\_month: The month the customer arrived at the hotel

7.arrival\_date\_week\_number: Week number of the year for arrival

Date

8.arrival\_date\_day\_of\_month: Day of arrival date

9.stays\_in\_weekend\_nights: Number of weekend nights

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

11.adults: Number of adults

12.children: Number of children

13.babies: Number of babies

14.meal: Type of meal booked

15.country: Country of origin

16.market\_segment: Different market segments

17.distribution\_channel: Booking distribution channels

18.is\_repeated\_guest: Is repeated guest (1) or not (0)

19.previous\_cancellations: Number of previous bookings cancelled by the

customer before current booking

20.previous\_booking\_not\_canceled: Number of previous bookings not

cancelled by the customer before current booking

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

22.assigned\_room\_type: Code of the assigned room type reserved

23.deposite\_type: No Deposit, Non Refund, Refundable

24.agent: Id of travel agency that made booking

26.company: Id of the company that made booking

27.days\_in\_waiting\_list: The number of days the booking was in waiting

28.customer\_type: Types of customer

29.adr: Average Daily Rate

30.required\_car\_parking\_spaces: No of car parking spaces required by the customer

31.total\_of\_special\_requests: Number of special request made by customer

32.reservation\_status: Reservation of last status

**Work Flow:**

**1.Understand Problem Statement:**

Our first step is to understand the problem statement. It gives us an explanation that describes the issue that we need to solve. So good understanding of problem helps us to outline the scope of a dataset.

This data set contains booking information for a city hotel and a resort hotel and we have to explore and analyse the data to discover important factors that govern the bookings.

**2.Data Collection and Understanding:**

Here we have to understand the dataset on which we have to perform an EDA.

1.First we will check shape of the dataset which will give us total number of rows and columns present in dataset. This dataset contains 32 columns and 119390 rows.

2.We will find descriptive statistics summary of a given data frame.

3.Checking data type of every column in the dataset.

4.We have seen some Nan values in the dataset while data understanding.

**3.Data Cleaning and Manipulation:**

1.First we will check if duplicate rows are present in the dataset or not using a Duplicated() function. Where True means Duplicate rows are present.

2.So total 31994 duplicate rows are present in the data set which we need to drop using Drop\_Duplicates() function.

3.We have already seen some Nan values, Which will require cleaning and handling of those Nan values.

4.Using isnull() function will check total null values present in the dataset

5.we found Nan values in 'childrean','country','agent','company' columns.

6.Looking into the nature of our dataset column 'company' is insignificant

to our data analysis. So Dropping column that are not significant.

7.Filling the null values of column ‘children’ & ‘agent’ by ‘0’ and column ‘country’ by ‘other’

**4.EDA(Exploratory Data Analysis):**

EDA is used to take insights from data. It refers to the critical process of performing initial investigations on data as to discover patterns, to check hypothesis and to check assumptions with the help of summary statistics and graphical representations.

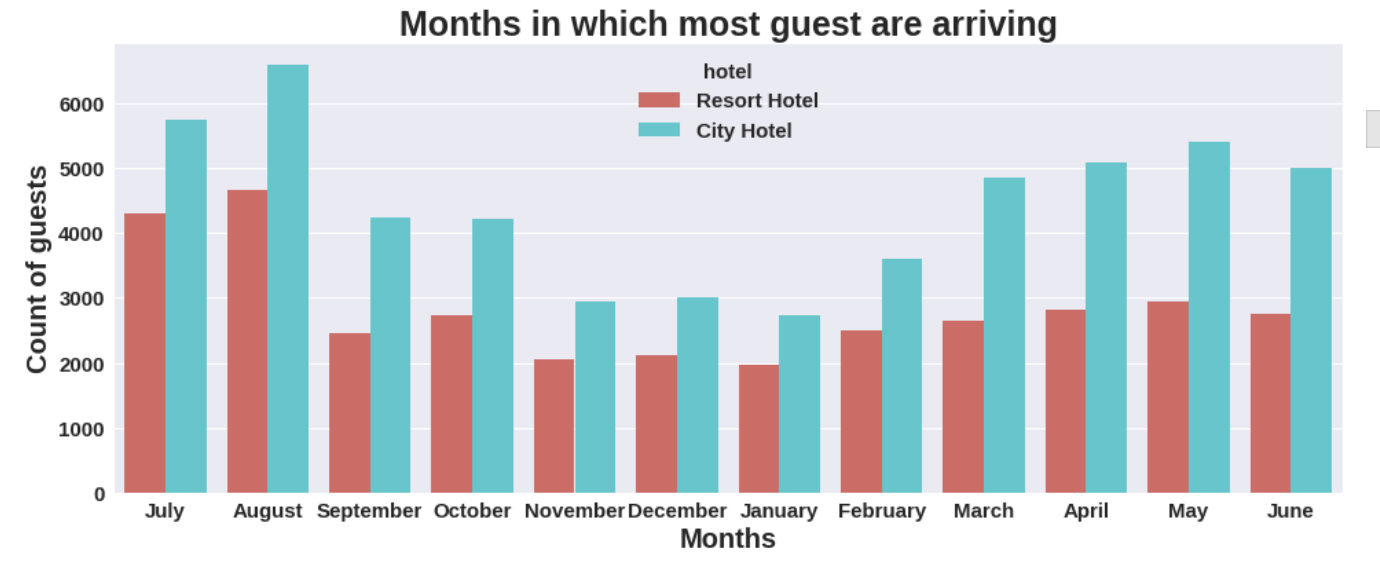
In this, we will understand the data frame and will decide our target variables (Important Columns) based upon which we were going to conduct further analysis.

Here we use seaborn and matplotlib libraries to visualize the summary.

We have perform EDA on dataset as shown below



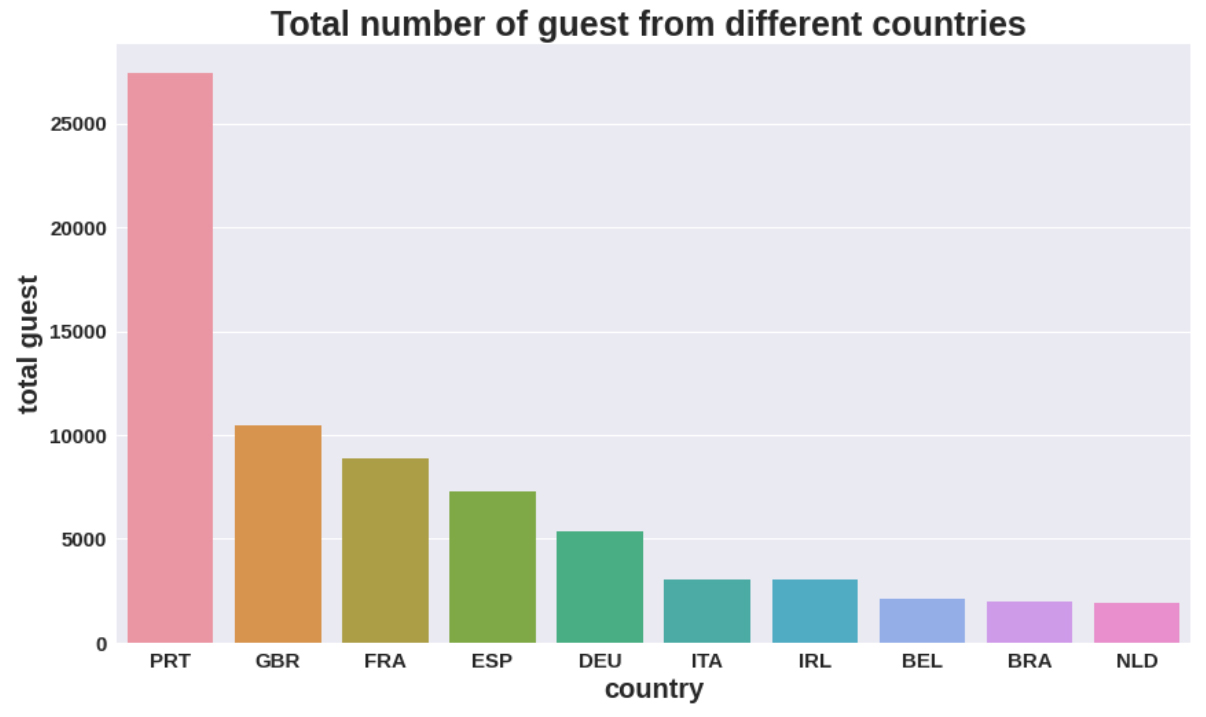
City hotel is the most preferred hotel type by the guests.

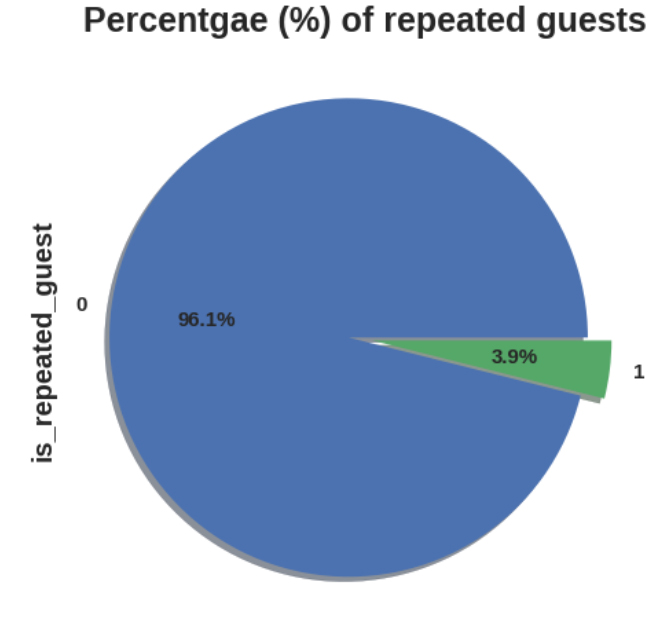


August month has the most guests bookings in both type of hotel Bookings are also high in May and July (ie.more than 8000)

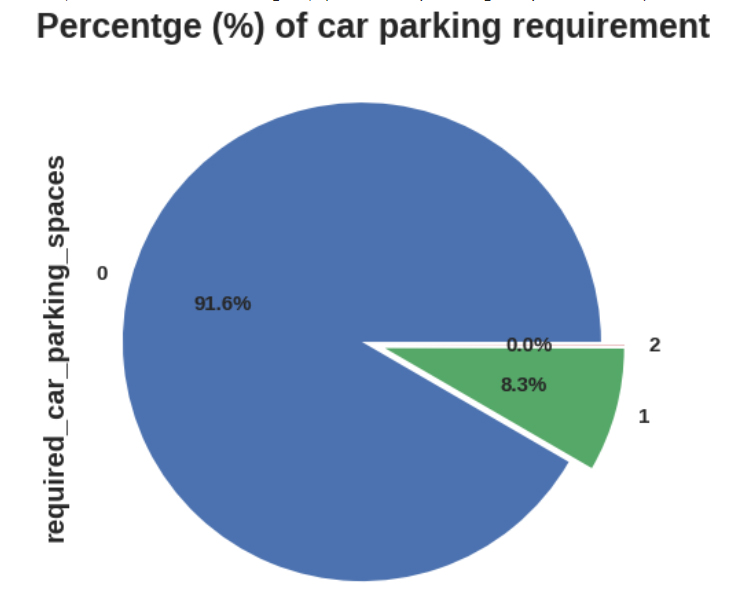


Graph shows the bookings for both city hotel and resort hotel is high in the year 2016and very less in 2015

It shows top 10 countries from where most of the guest are arriving as per the graph most number of guests coming from PRT(Portugal) followed by GBR(Great Britain) and FRA(France)



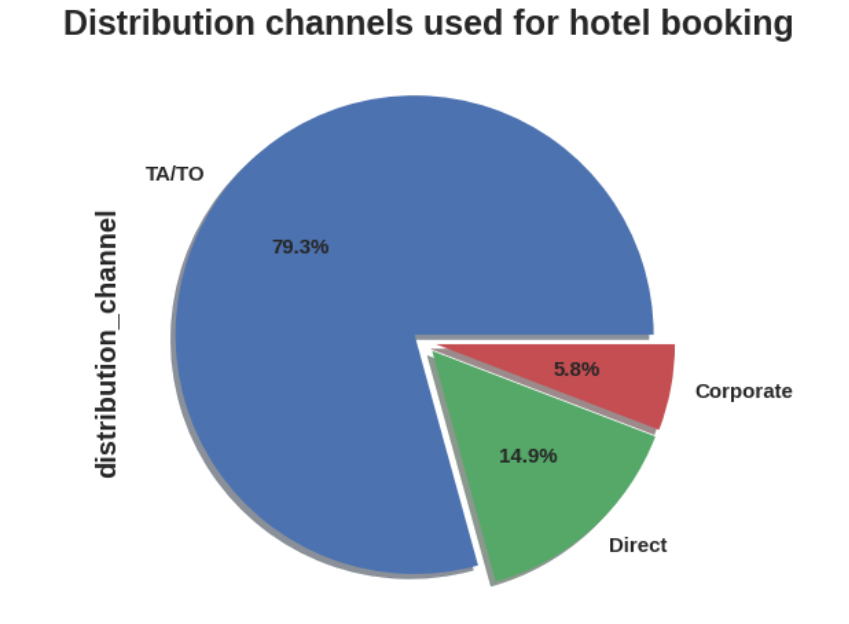
Very few guests are repeated i.e. 3.9% are repeated, Remaining 96.1% were new guests.



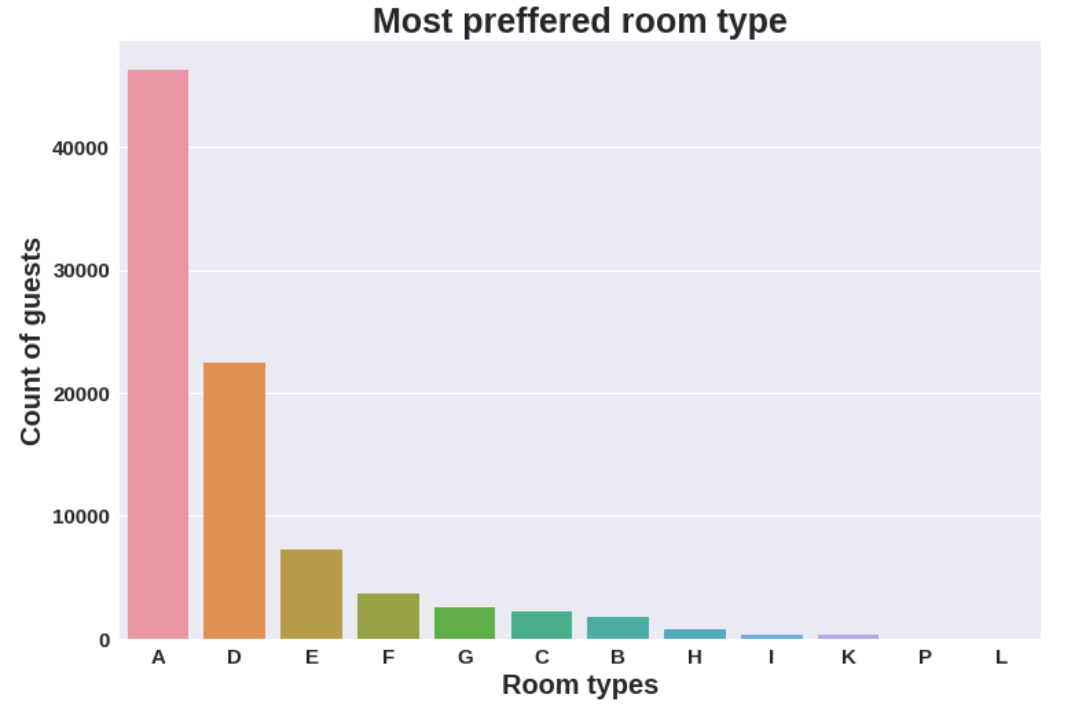
Most of the customers do not required parking spaces only 8.3% requires single parking space.



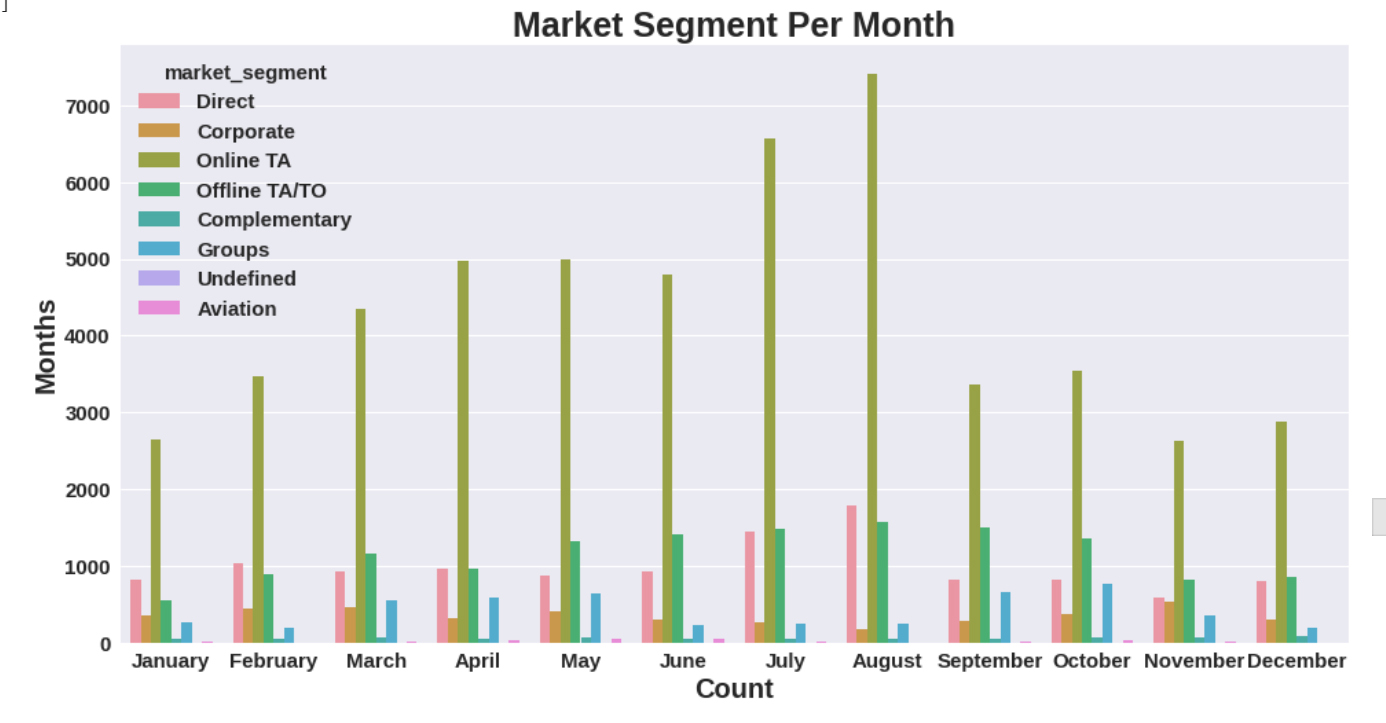
City hotel has the high booking cancellation rate than the Resort hotel.



79.3% Bookings were made through TA/TO, Second most channel is Direct.

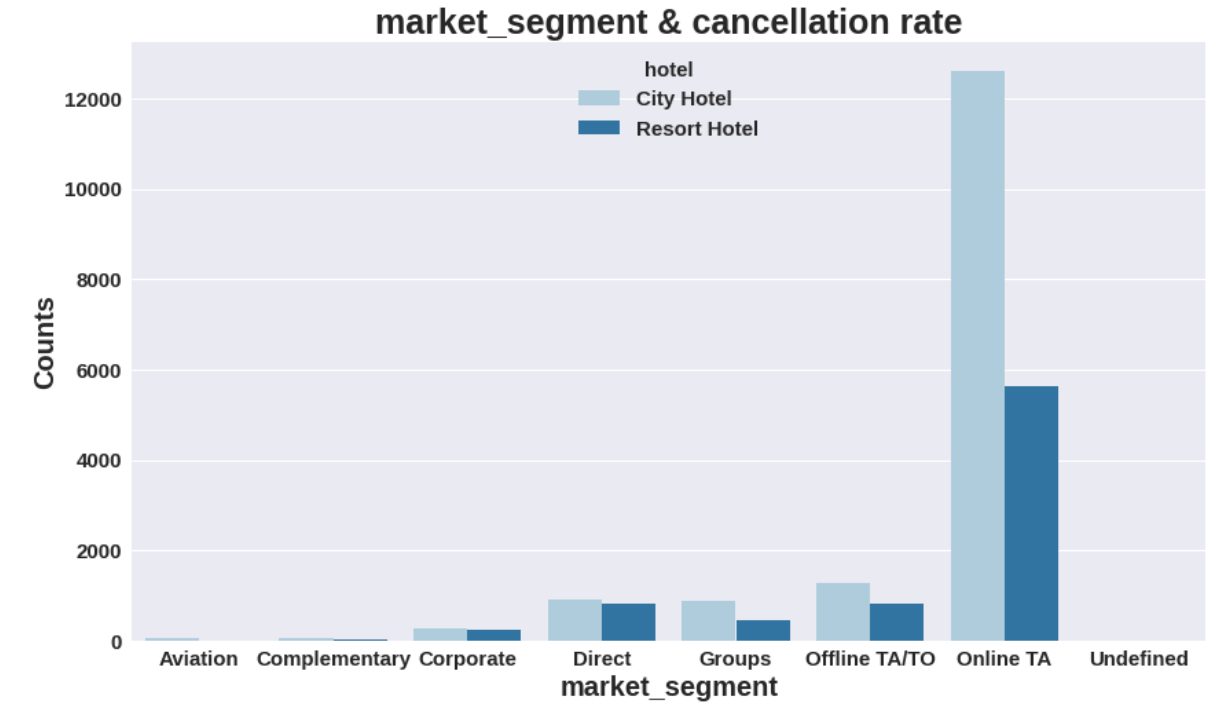


Most guests preferred room type is A, Second most Preferred is D

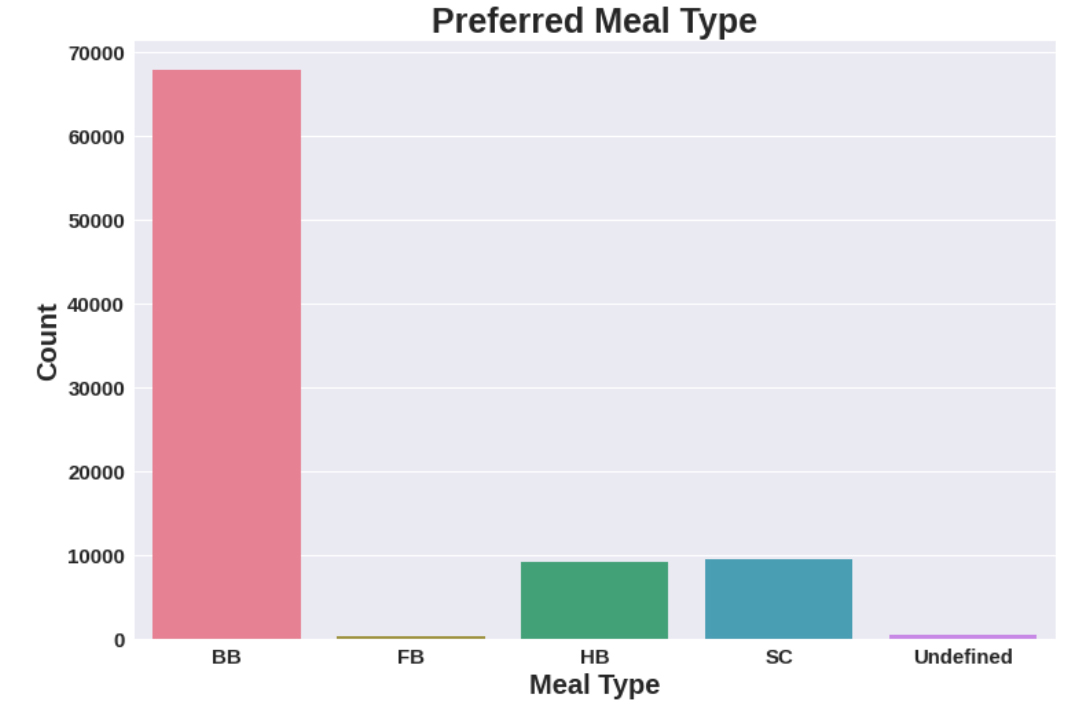
Most of the hotels are booked by online TA(travel agent) in the August month



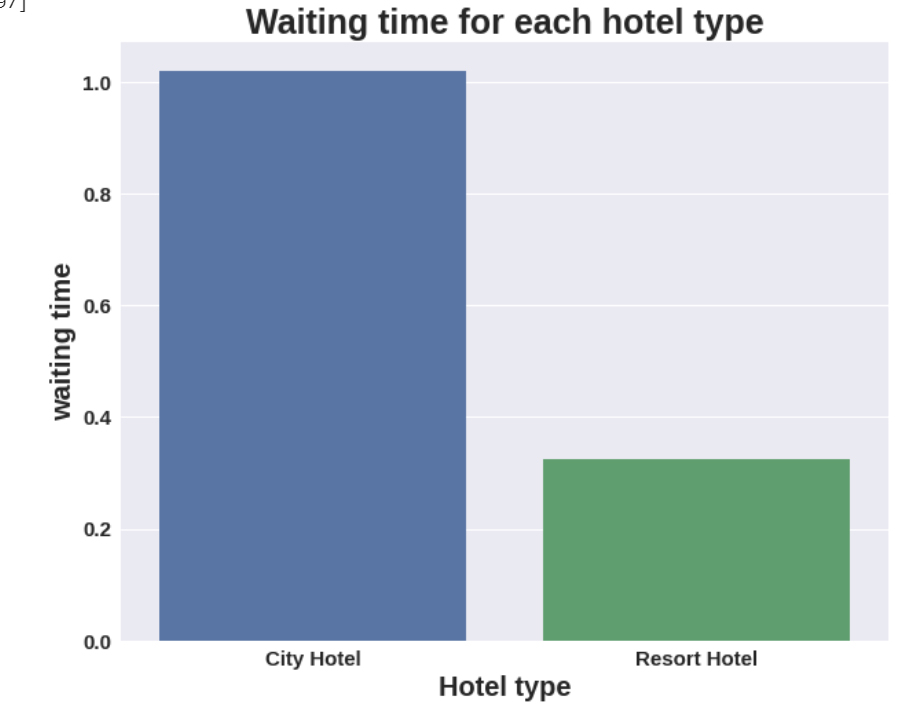
Avg ADR for city hotel is high as compared to resort hotel, These city hotels are generating more revenue than Resort hotel



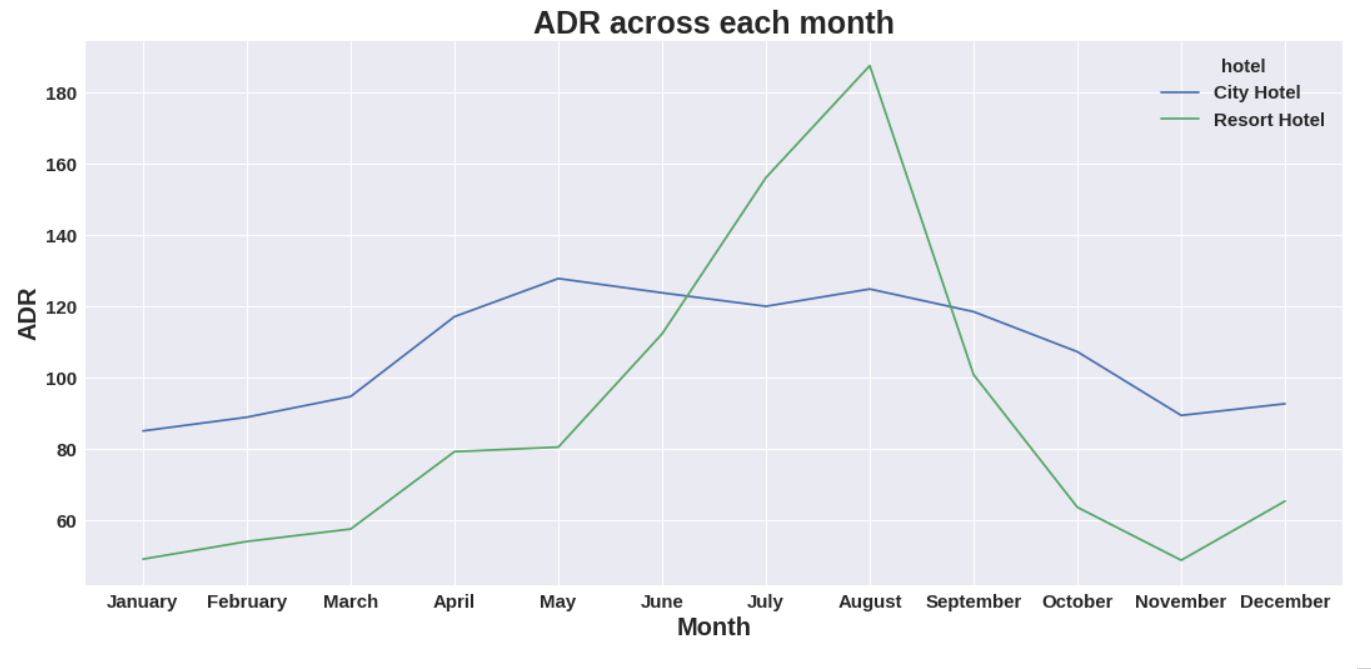
Online TA(travel agent) has the highest cancellation rate in both type of hotels.



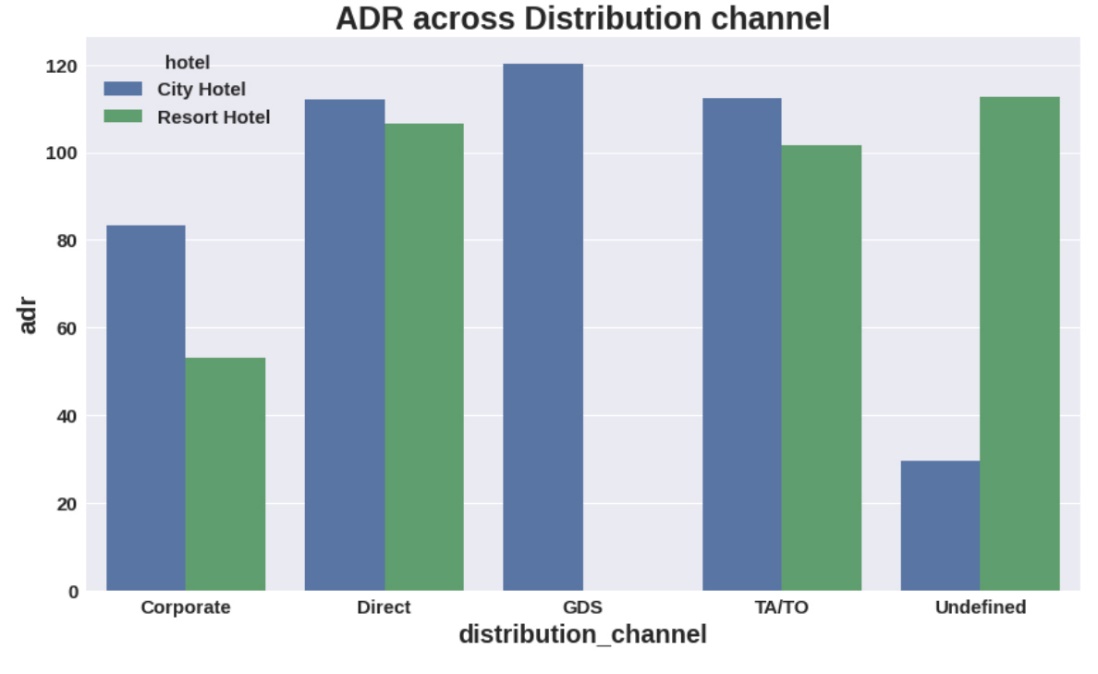
BB(Bread and breakfast) is most preferred meal type by the guest.FB(full bed) is least preferred



City hotel has longer waiting time as compared to Resort hotel,So city hotels are much busier than the resort hotel



Resort hotels had the highest ADR in July and August than the City hotels. But in other months ADR of Resort hotel was less than the City hotels.



Graph shows GDS is more effective in city hotels but it needs to improve in Resort hotel as well. TA/TO and Direct is almost equally contributed in both type of hotel for increasing their income.

**Conclusion:**

1. Resort hotel needs to work more on their marketing strategy and they should promote their hotel more on social media platforms in order to increase the number of guests.

2. Resort hotels could also reduce prices to increases booking percentages.

3. Avg ADR for city hotel is high as compared to resort hotel, These city hotels are generating more revenue than Resort hotel

4.Given that hotels do not have more repeated guests, hotels may implement extra benefits offer for repeated guest, to attract them.

5. August month has the most guests bookings in both type of hotel Bookings are also high in May and July (ie.more than 8000)

6. Although city hotels have more bookings, they also tend to have more cancellations so to prevent this they could take advance money during vacation. They could also apply minimum charge policies This would ensure most bookings to not being cancelled.

7.We also realize that the high rate of cancellations can be due high no deposit policies.

8. City hotel has longer waiting time as compared to Resort hotel,

So city hotels are much busier than the resort hotel.

9. BB(Bread and breakfast) is most preferred meal type by the guest, FB(full bed) is least preferred.

10. 82.1% Bookings were made through TA/TO,Second most channel is Direct.

11. It shows top 10 countries from where most of the guest are arriving as per the graph most number of guests coming from PRT(Portugal) followed by GBR(Great Britain) and FRA(France)

12.Year 2016 is the peak year for hotel booking and 2015 has very less bookings.

13. The hotel reservation tends to peak during July and August.

14. Graph shows different market segments across months, According to this graph in every month online TA segment effects the most.

15.City Hotels has longer waiting period than the Resort Hotel.

16. Online T/A(Travel agent) has the highest cancellation in both type of hotel.

17. The most preferred Room type is "A".

18.As the graph shows very few guests required car parking.

19. Graph shows GDS is more effective in city hotels but it needs to improve in Resort hotel as well.TA/TO and Direct is almost equally contributed in both type of hotel for increasing their income.