**Hotel Booking Analysis**

**Akash Kumar**

**Abinash Kanungo**

**Kunal Kamble**

**Zueb Ansari**

**Data science trainees,**

**Almabetter, Bangalore**

**Abstract:**

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 from the data.

We will perform exploratory data analysis with python to get insight from the data.

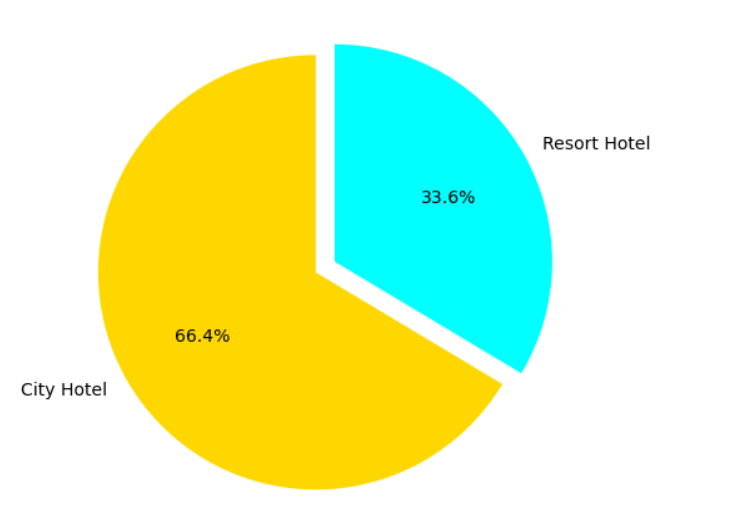
***Keywords: Exploratory data analysis, Hotel booking analysis***

**Problem Statement**

**1. What types of hotels are in the market and how many numbers of hotels in each type?**

There are two types of hotels.

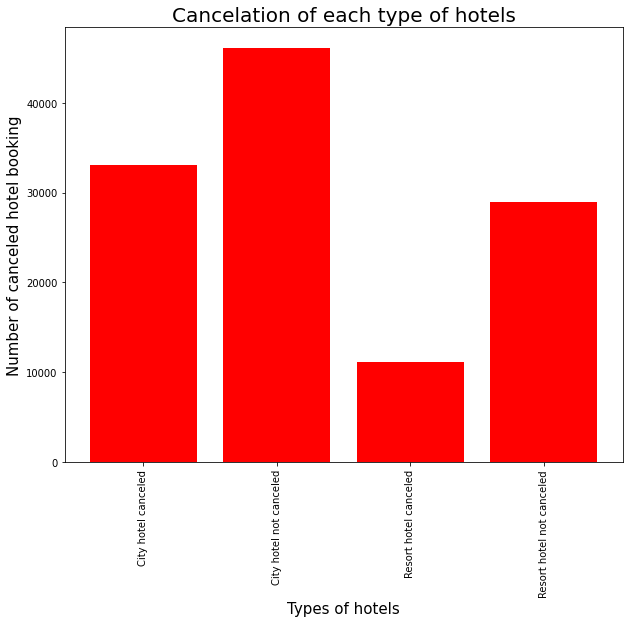
* City Hotels
* Resort Hotels



Out of 119390 Hotel bookings City hotel bookings which is 66.4% whereas Resort bookings makes 33.6%. Here by Count of City Hotel Bookings are morewhen compared to Resort Hotel.

**2. How Many bookings were canceled for different types of hotels?**

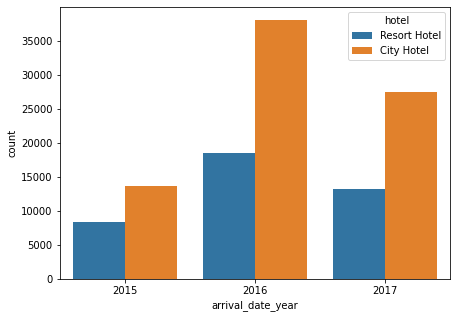
There are four bars in the graph which represent city hotels canceled, city hotels not canceled, resort hotels canceled, resort hotels not canceled.



we can say that city hotels bookings are

More canceled than resort hotels.

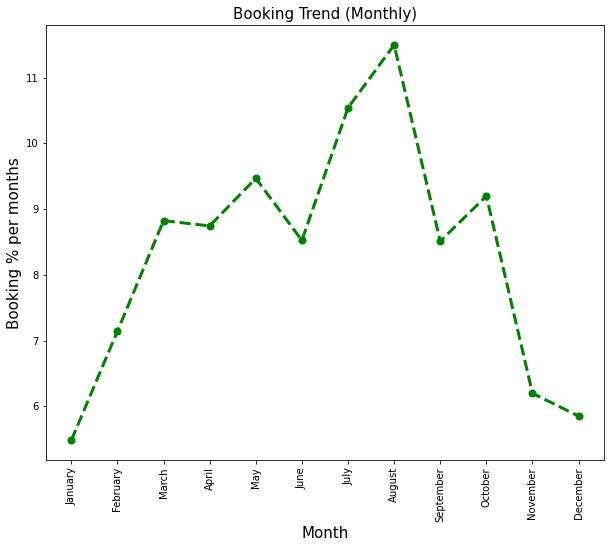
* **Year wise comparison of booking of hotels.**

****

In this we have used a count plot method which is used in seaborn for comparison between years.

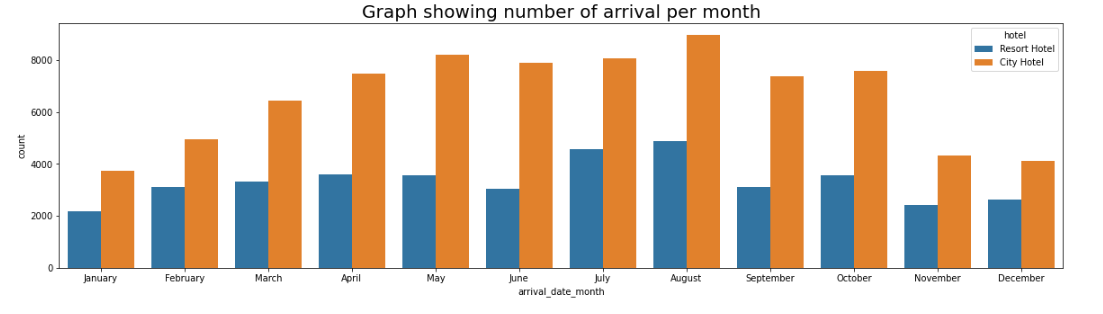
We can say that in 2016 most guests visited both types of hotels.

**3. Which is the busiest month for hotels?**

Busiest month means which month has the most number of bookings.

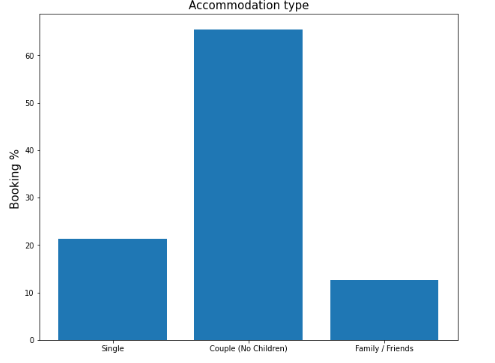
we can say that **August** is the busiest month for hotels, which means august month has the greatest number of bookings**.**

* **Now we will check comparatively for both types of hotels which month is busiest.**

****

For this we have used a count plot imported from seaborn from this comparison. Also August is the busiest month of all.

**4. Which was the most booked accommodation type (Single, Couple, Family)?**

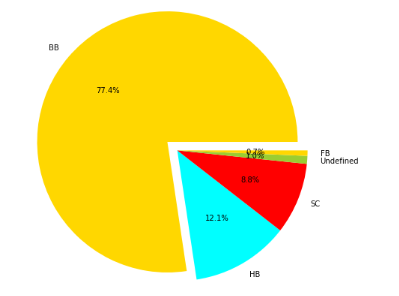
According to analysis there are three types of people (guests) who come to book hotels : singles, couples and families. 

From the analysis we can say thatCouples with no children has greatest number of bookings.

**5. Which type of meal is 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)



In these categories of meals, guests usually like orders BB i.e., Bed & Breakfast.

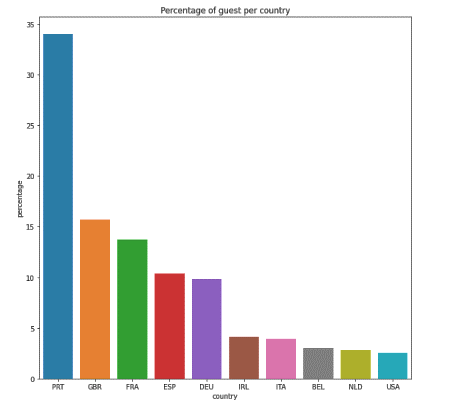
Here we have used pie charts to represent different types of meals ordered by the guests.

BB i.e., Bed & Breakfast**is most preferable meal type of the guests which is 77.4%**

**6. From which country most guests come?**

In this we have firstly fetch the top 10 countries from the given dataset.

Then rename the columns name such as country, number of bookings, percentage.

Then draw graph using bar plot from there we can say that most of the guests come from **PRT i.e., Portugal**.

**7. Which type of market segment is used for booking hotels?**

There are eight type of market segment from guests can book their hotels;

Hotels with respect to number of bookings

Online TA 56408

Offline TA/TO 24182

Groups 19791

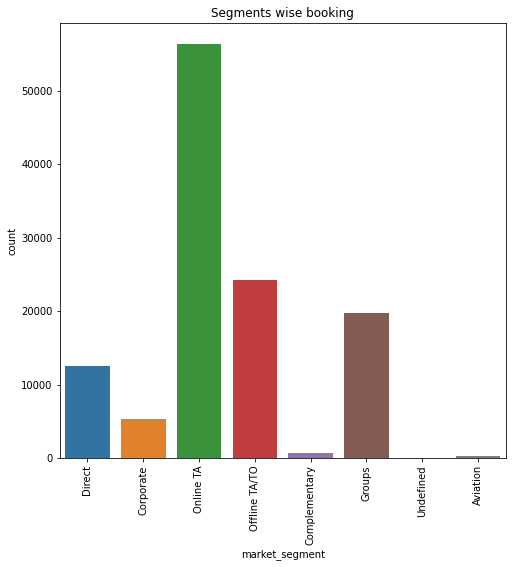
Direct 12582

Corporate 5282

Complementary 728

Aviation 235

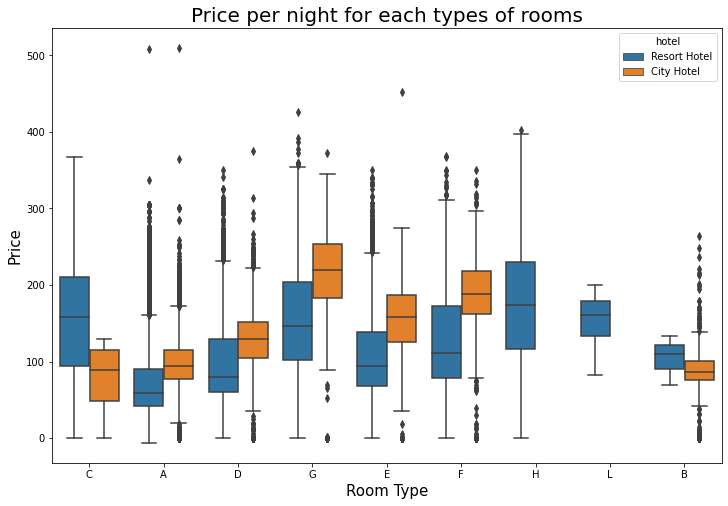
Undefined 2



we can say that **Online TA** market segment is mostly used to book a hotel.

**8. How much do guests pay per night for each type of room?**

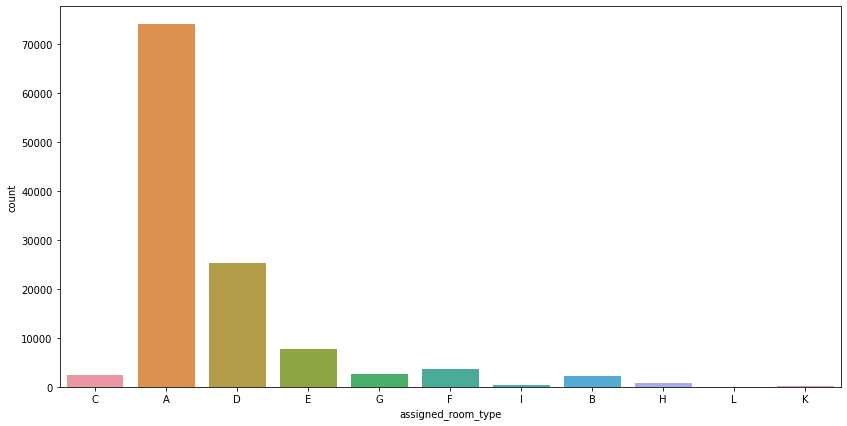
In this we have used box plots to present the types of rooms.



we can say that **'G'** category room of city hotel are much costlier than other.

* **Assigned room type**

In this we have used a bar plot to present the assigned room type from the reserved room type.



we can say that **'A'** type rooms are most assigned room type among all.

**Steps involved:**

* **Exploratory Data Analysis**

After loading the dataset we performed this method by comparing our target variable that is Hotel 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.

* **Null values Treatment**

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

## **Check for the wrong entry**

Our dataset contains 3 categories of guests which are children, adults and babies**.**

If a hotel is booked then there should be any entries in children, adults or babies’ columns, it should not be zero at a time. Hence we drop them at the beginning of our project in order to get a better result.

* **Data analysis and Data visualization**

### Here we do data analysis and data visualization by using matplotlib and seaborn to the graph for better understanding of the insights.

**Conclusion:**

That's it! We reached the end of our exercise.

Starting with loading the data so far we have done EDA, null values cleaning, encoding of categorical columns, data analysis and data visualization.

* According to this analysis, the number of City hotels comparatively more than Resort hotels. Resort Hotels tend to be on the expensive side so most people will just stick with city hotels for affordability reason.
* City hotel bookings are more canceled than resort hotels.
* In 2016 most guests visited both types of hotels compared to other years.
* August is the busiest month for hotels, which means August has the greatest number of bookings.
* August is also the busiest month for both types of hotels.
* Couples (Two Adults) have the greatest number of bookings.
* BB i.e., Bed & Breakfast is most preferable meal type of the guests which is 77.4%
* Most of the guests come from PRT i.e., Portugal.
* Guests use most online TA market segments for hotel booking.
* 'G' category rooms of city hotels are much costlier than others.
* 'A' type rooms are the most assigned room type among all.

**References-**

* Pandas user guide: <https://pandas.pydata.org/docs/user_guide/index.html>
* Matplotlib user guide: <https://matplotlib.org/3.3.1/users/index.html>
* Seaborn user guide & tutorial: <https://seaborn.pydata.org/tutorial.html>
* Tutorials: 1) [https://stackoverflow.com](https://stackoverflow.com/) , 2) [https://www.w3schools.com](https://www.w3schools.com/)