**EDA on Hotel Bookings**

***Aravind Dasarapu***

***Pedagadi Prasanthi***

***Data Science Trainees***

***Alma Better***

# **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 etc,. All personally identifying information is from the data.**

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

# 

# **Problem Statement:**

**We have been provided a dataset of the hotel industry. Our primary goal is to understand the data and use it for our analysis and build techniques to draw out key features. Our secondary goal is to draw out actionable insights from our analysis and give conclusions about key aspects of the dataset such as cancellation rate, distribution channels, country of origin etc,.**

# **Data Summary :**

**Based upon the initial assessment we found that the data was pretty much clean except for some missing values in a few columns. Upon 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**

## **Provided Columns :**

● 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

# **Data Cleaning :**

**We dropped rows in which columns having less null values and whole columns in which more null values are present.**

#### Rows dropped with respect to columns having null values :

**‘children’**

**‘country’**

#### Columns dropped with respect to percentage of null values:

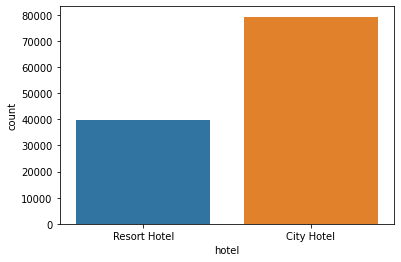
**‘agent’**

**‘company’**

**We have handled outliers using boxplots to restrict the data domain to remove the skewness from the dataset, which can be seen in the notebook file. We added new columns which can be important with respect to the given dataset.**

### **Insights :**

**i) Which type of hotel has more bookings ?**



**Out of total 119390 bookings 79302 are of city hotels and 40088 are of resort hotels**

**ii) How many bookings are canceled ?**

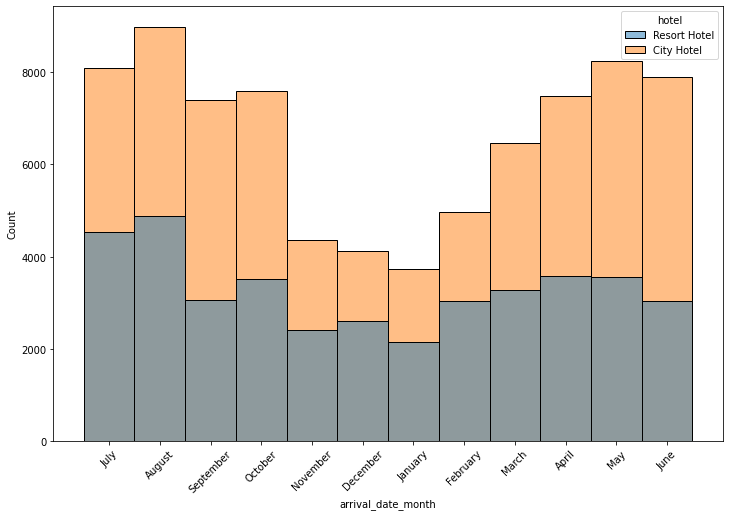
****

**Out of all bookings nearly 37 percent of bookings are being canceled.**

**iii)Which type of customers are booking more and Which type of hotel has more cancellations ?**

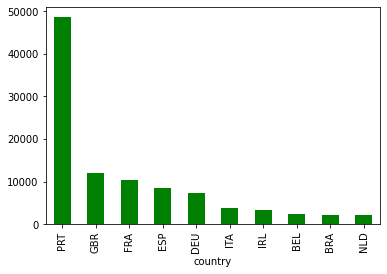
**From the plots in the notebook file we concluded that Transient bookings are more in number compared to other types of bookings, they represent 75% of the total visitors. City hotels are more prone to cancellations. Hence hotels should provide some offers focusing on transient customers to decrease cancellations**.

**iv) Which month has the most number of bookings and least number of bookings ?**

****

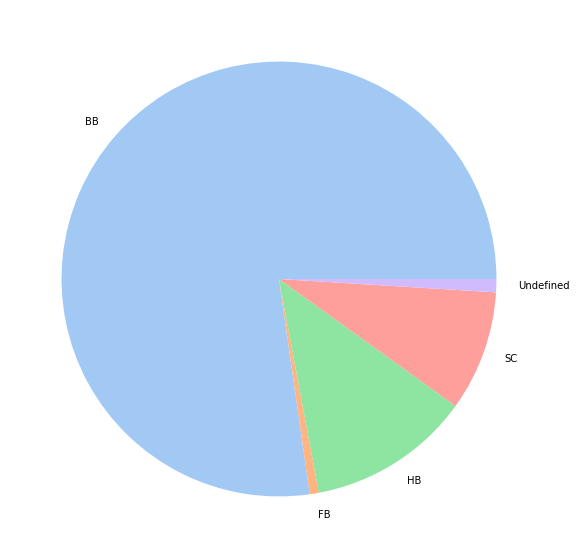
**Out of all months we see august having highest number of bookings and january having least number of bookings total hotels combined**

**v) Which countries cover the majority of customers ?**

****

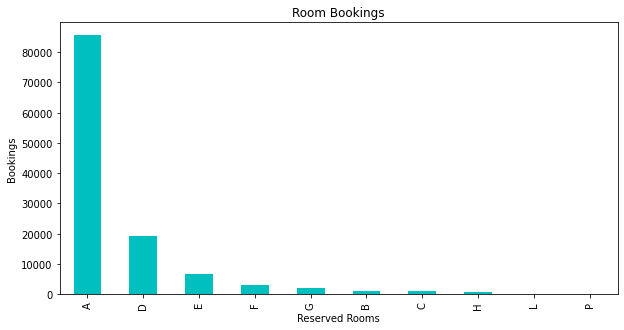
**We found that a huge number of visitors are from Western Europe namely Portugal, France, Great Britain. And city hotels have maximum bookings. And it seems 2016 to be the year where the hotel bookings are highest.**

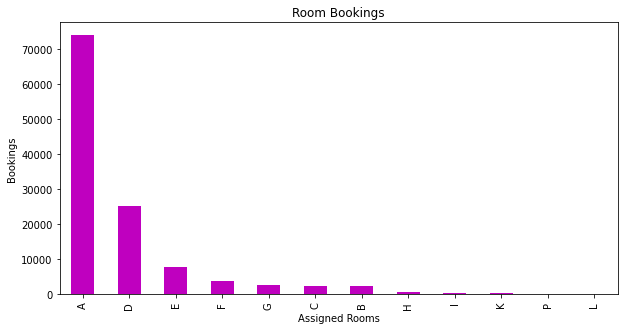
**vi) Which type of meals are preferred by customers ?**

****

**As we can see the majority of customers prefer Bed and Breakfast(BB) meal type**

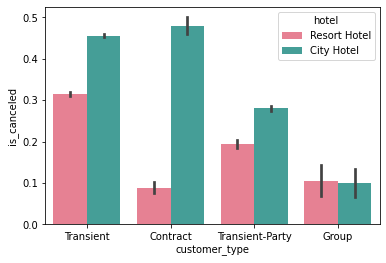
**vii) Which major type of rooms are reserved and assigned to customers ?**

****

****

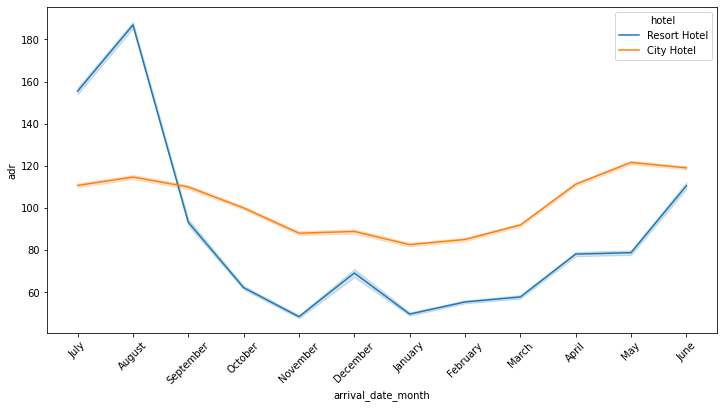
**We can see that majority of customers prefer ‘A’ room type**

**viii) Which type of customers prefer which type of hotel ?**

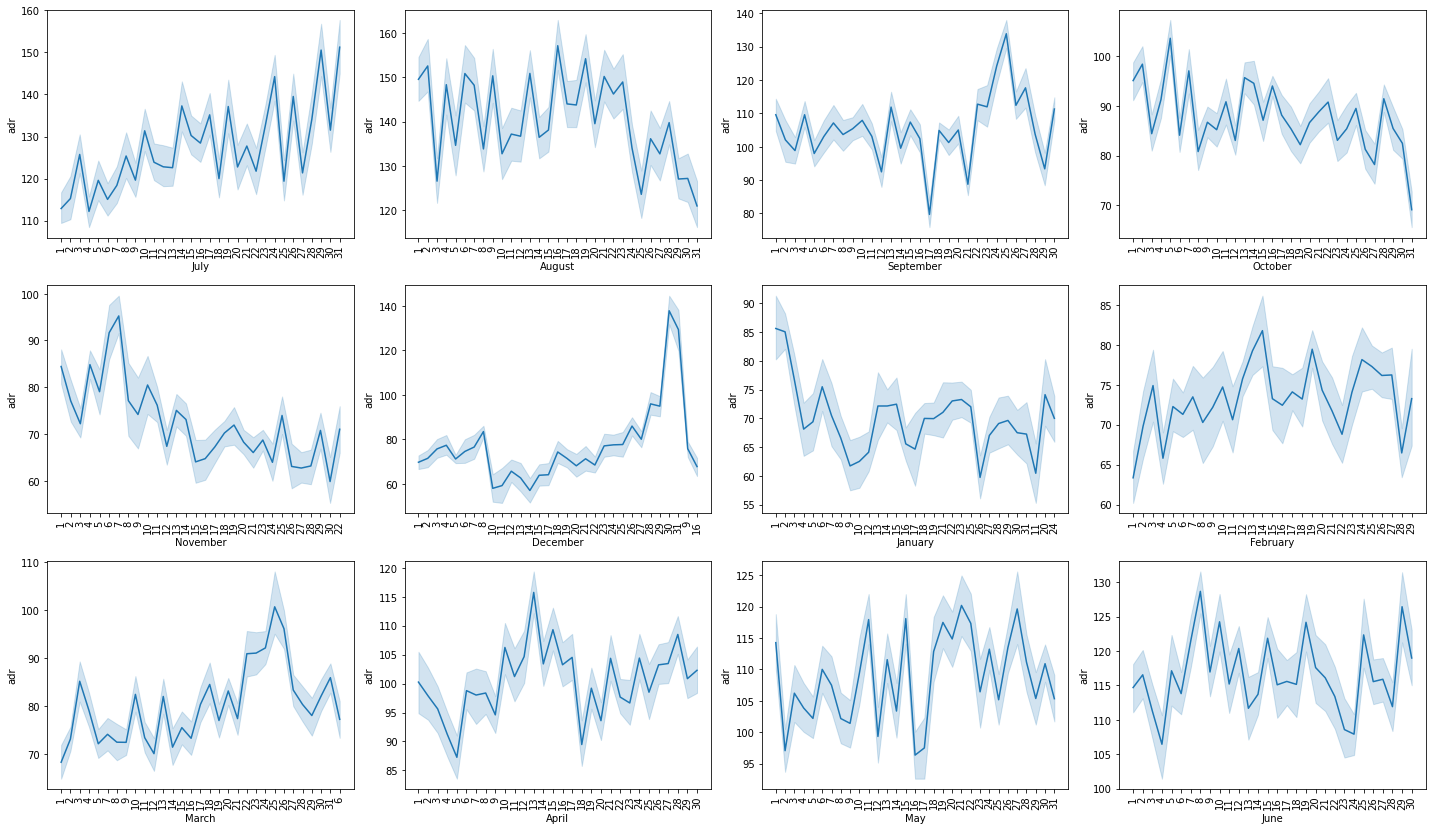
****

**We can say the most guests are staying in city hotels compared to resort hotels, but guests of group type is a little more in resort hotel than city hotels.**

**ix) How does the ADR(average daily rate) fluctuate throughout the year with respect to hotels ?**

****

**For resort hotels, the average daily rate is more expensive during August, July. For city hotels, the average daily rate is more expensive from September to June compared to resort hotels.**

****

**ADR(Average Daily Rates) fluctuations throughout the days in the months**

# **Conclusions :**

* **The majority of guests come from western europe countries.**
* **The majority of reservations are for city hotels.**
* **The number of repeated guests is too low.**
* **The majority of reservations convert into successful transactions.**
* **The many bookings are made in 2016 compared to 2015 and 2017**
* **The majority of Bookings are made through Online Travel Agents.**
* **The month of highest occupation is august with 11.65% of the reservations. The month of least occupation is January with 4.94% of the reservations.**
* **The Bed & Breakfast option is the most popular, with a frequency of 77.26%.**
* **Transients are the most common customer type, they represent 75% of the total customers.**
* **More guests are staying at city hotels on weekdays.**

# **References :**

* Stack Exchange : [Data Science Stack Exchange](https://datascience.stackexchange.com/)
* Stack Overflow : <https://stackoverflow.com/>
* Pandas : [User Guide — pandas 1.5.2 documentation (pydata.org)](https://pandas.pydata.org/docs/user_guide/index.html#user-guide)
* Seaborn : [User guide and tutorial — seaborn 0.12.1 documentation (pydata.org)](http://seaborn.pydata.org/tutorial.html)