**Hotel Booking Analysis**

**Swapnil Bhagawat, Sumit Zoting, Gaurav Dod**

**Bhagyashree Bhoyar, Rahul Katkhede**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

Hotel Booking is a platform for book City hotel and Resort hotel at best time of year or the optimal length of stay in order to get the best daily rate.

Our study used one dataset consisted 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.

Our experiment to help you to discover important factors that govern the bookings.

***Keywords: Hotel booking, City hotel, Resort hotel.***

**1. Problem Statement**

For this project we will be analyzing Hotel Booking data. 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.

The main objective behind this project is to explore and analyze data to discover important factors that govern the bookings and give insights to hotel management, which can perform various campaigns to boost the business and performance.

**1.1 Hotel Booking DF:**

The contents present in Play Store Dataset are:

* **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.

**2. Introductio****n**

### Exploratory Data Analysis (EDA) is for understanding and analyzing the dataset given to get better insights into the data and can be used for taking important business decisions.

### We can summarize the main characteristics of the dataset and can plot the data visually by using Python libraries.

### There are various plots like Bar Chart, Pie Chart, Box Plot, Heatmap and Line Chart etc. by using libraries Matplotlib, Seaborn. We can store our data in data frame by using the Pandas library from Python.

## **3. Steps Involved**

### Data Collection and Understanding

### Data Cleaning and Manipulation

### Exploratory Data Analysis (EDA)

### **Data Collection and Understanding**

There is one dataset which were given hotel booking dataset. It has 119390 rows and 32 columns.

Dataset 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.

### **Data Cleaning and Manipulation**

In hotel booking dataset there is 32 columns in that 4 columns has missing values that columns is company, agent, country and children.

So we replace the missing values of company,agent and children columns with ‘0’ and country column missing value replace wirh ‘other’. After treating missing values next comes to chack for duplicate values.

This dataset also has duplicate values so we drop it, then after that

We created 2 new columns 1)‘Total\_Peoptle’ = It consist Children, adults, babies.

2) ‘Total\_stay’ = It consist weekend nights and weekdays night

### **Exploratory Data Analysis (EDA)**

It is a systematic way of describing and illustrating, condensing, and evaluating data by systematically using statistical or logical methods.

**Data Visualization:**

Python's libraries provide plenty of features with which users can create highly customized, elegant, and interactive plots, making data visualization with Python one of the most widely used features in today's data science environment.

Data Visualization libraries in Python are:

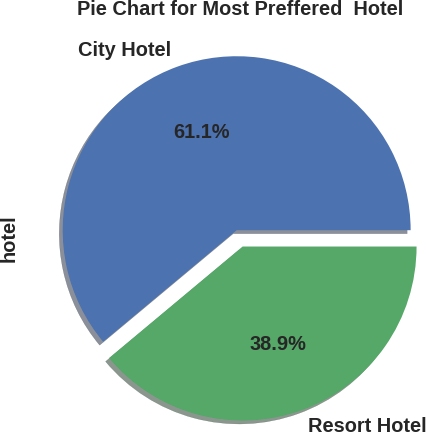
* **Matplotlib:** With a Python library, you can visualize arrays in 2D. NumPy is used to create Matplotlib, which is written in Python. There is a wide variety of plots in Matplotlib, including line, bar, scatter, histogram, etc. that can help us understand trends, patterns, correlations.
* **Seaborn:** A Python library for representing statistics with datasets, Seaborn is a dataset-oriented library. This library is built on matplotlib and creates a variety of visual representations. Pandas data structures are incorporated into it. To create informative visuals, the library performs mapping and aggregation internally

## **4. Visualizations**

**4.1 City Hotel VS Resort Hotel**

**Pie Chart:**

Pie charts can be used to identify proportions of the different components in a given whole



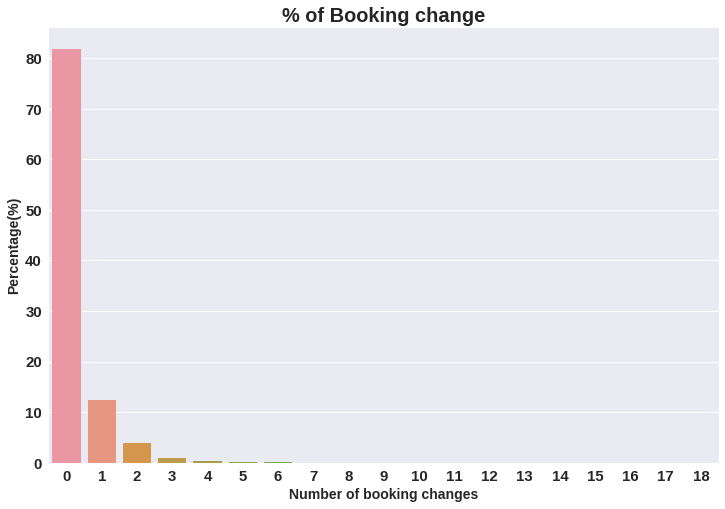
City hotels is the most preferred hotel type

Nearly 61.1% guests preferred City hotel and 38.9 % guests preferred Resort hotel.

**4.2 % Of Booking Change**

**BAR CHART:**

A bar chart is used when we want to compare metric values across different subgroups of the data. If we have a greater number of groups, a bar chart is preferred over a column chart.



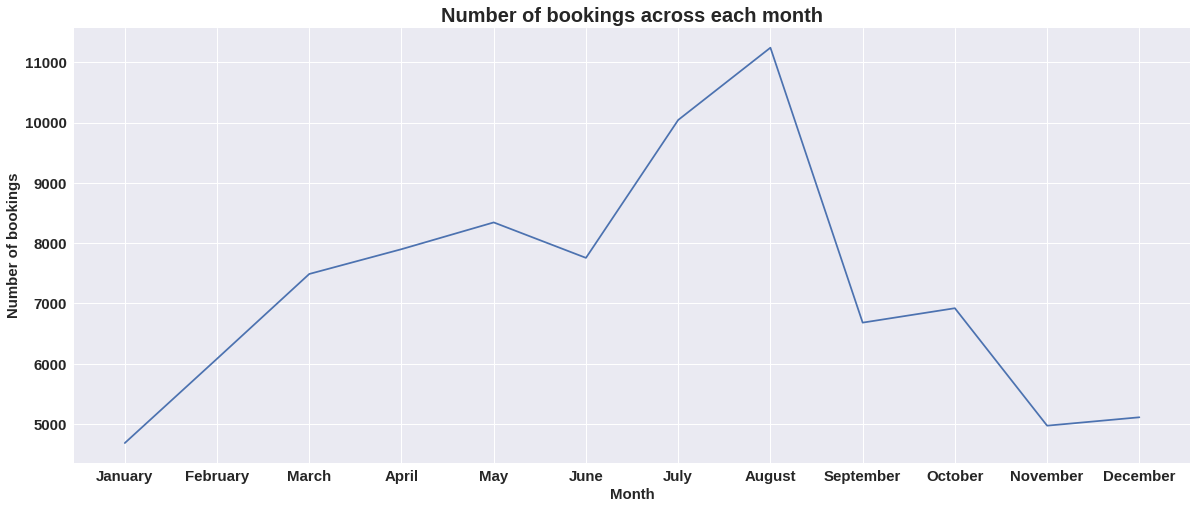
1. The percentage of 0 changes made in the booking was more than 82 %.
2. Percentage of Single changes made was about 10%.

**4.3 Month VS No of booking**

**Line Chart:**

Continuous data points are represented with a line chart.

This visual can be effectively utilized when we want to understand the trend across time.



As we can see in the line chart, from June to September most of the bookings happened, It’s Summer time. After September bookings Starts declining.

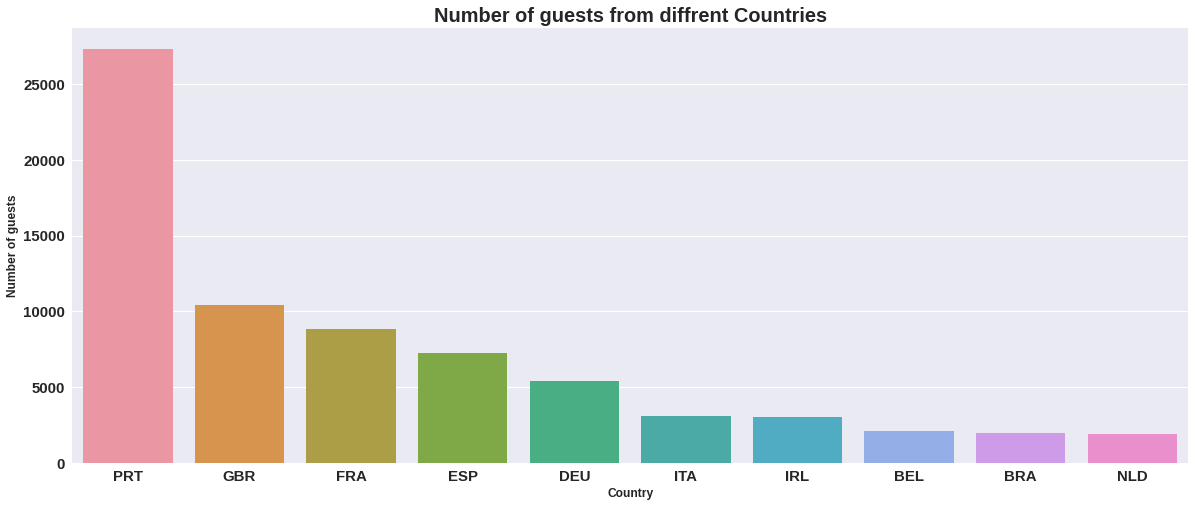
**4.4 ADR Across each Month**

****

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

Thus we can say that, the January, February, March, April ,November and December are the good months for customers to get good adr

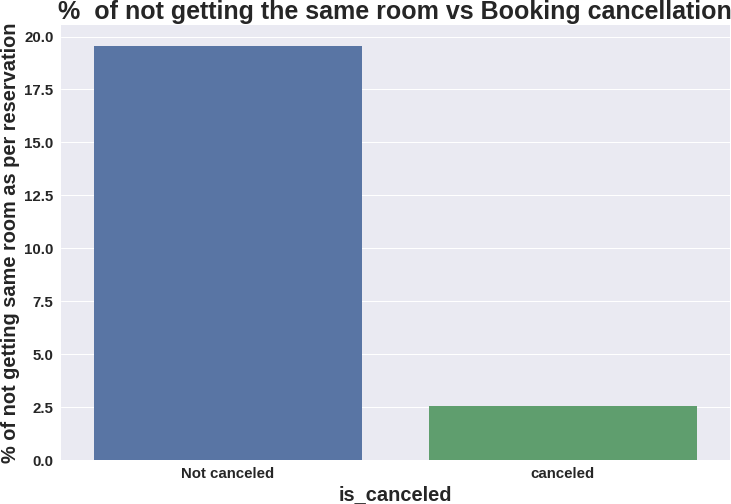
**4.5 Number of guests from different countries**



Maximum number of guests were from Portugal.i.e. more than 25000 guests.

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

**4.6 % of not getting the same room VS Booking cancellation**



Almost 19 % people did not canceled their bookings even after not getting the same room which they reserved while booking hotel. Only 2.5 % people cancelled the booking.

Thus not getting the same room as per reserved room is not the reason for booking cancellations.

**4.7 HEAT MAP**

Heatmap is defined as a graphical representation of data using colors to visualize the value of the matrix. In this, to represent more common values or higher activities brighter colors basically, reddish colors are used, and to represent less common or activity values, darker colors are preferred.

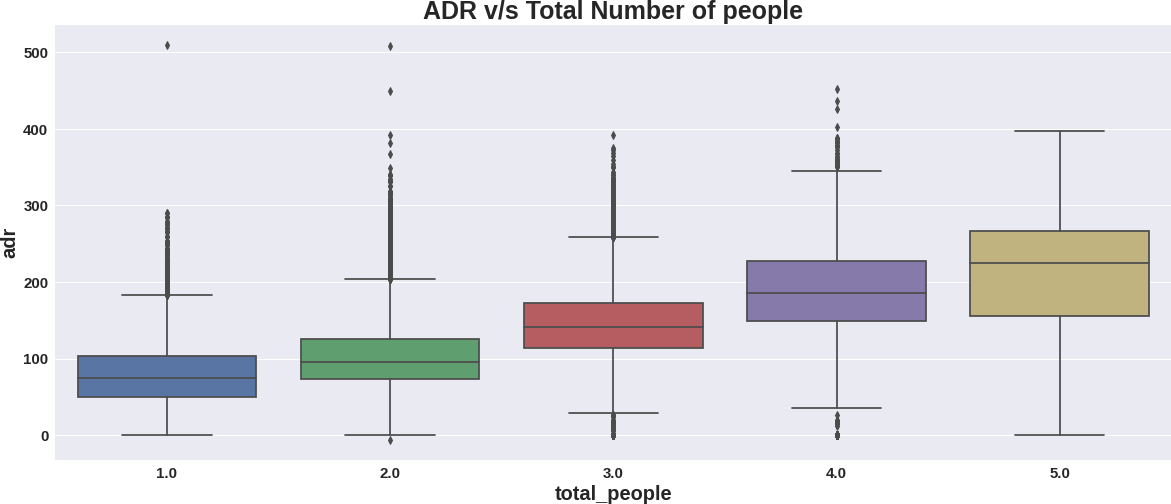
**Corr() function:**

To determine the pairwise correlation of all columns in the data frame, Pandas dataframe.corr() is used.



* 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.
* ADR and total people are highly correlated. That means more the people more will be adr.High adr means high revenue
* is\_repeated\_guest and previous\_bookings Not\_canceled has strong correlation. May be repeated guests are not more likely to cancel their bookings.

**4.8 ADR VS Total Number of people**



As we saw in Correlation heatmap, total people and adr are positively correlated. Thus for 2 people ,adr is almost 100 and for 5 people its more than 200.

Thus more the people more will revenue of the hotels.

**5 CONCLUSION:**

* City hotels are the most preferred hotel type by the guests. We can say City hotel is the busiest hotel.
* 27.5 % bookings were got cancelled out of all the bookings.
* Only 3.9 % people were revisited the hotels. Rest 96.1 % were new guests. Thus retention rate is low.
* The percentage of 0 changes made in the booking was more than 82 %. Percentage of Single changes made was about 10%.
* Most of the customers (91.6%) do not require car parking spaces.
* 79.1 % bookings were made through TA/TO (travel agents/Tour operators).
* BB( Bed & Breakfast) is the most preferred type of meal by the guests.
* 8.Maximum number of guests were from Portugal, i.e. more than 25000 guests.
* Most of the bookings for City hotels and Resort hotel were happened in 2016.
* Average ADR for city hotel is high as compared to resort hotels. These City hotels are generating more revenue than the resort hotels.
* Booking cancellation rate is high for City hotels which almost 30 %.
* Average lead time for resort hotel is high.
* Waiting time period for City hotel is high as compared to resort hotels. That means city hotels are much busier than Resort hotels.
* Resort hotels have the most repeated guests.
* Optimal stay in both the type hotel is less than 7 days. Usually people stay for a week.
* Almost 19 % people did not cancel their bookings even after not getting the same room which they reserved while booking hotel. Only 2.5 % people cancelled the booking.

**References-**

GeeksforGeeks

Kaggle

Analytics Vidhya