**Hotel Booking Analysis**

**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

**1.Problem Statement**

For this project we are doing analysis on Hotel bookings data from 2015 to 2017.

It 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.

Our main aim is to find out which important factors influences Citi and Resort hotel for bookings, cancellation and other parameters. For this we will explore and visualise the data to discover important factors that the bookings using exploratory data analysis.

Here We will find out the percentage of booking done in different hotels, total bookings done in different Years, Total Number of Booking Cancelled in different months, room types and market segment .

**Given data set has different columns of variables crucial for hotel bookings.**

**Some of them are:**

**hotel:** Type of Hotels

**is\_cancelled :** cancelled or not | Values[0,1], where 0 indicates not cancelled.

**lead\_time :** The time between reservation and actual arrival.

**stays\_in\_weekend\_nights:** The number of weekend nights stay per reservation

**stays\_in\_weekday\_nights:** The number of weekday nights stay per reservation.

**meal:** Meal preferences per reservation. (BB,FB,HB,SC,Undefined)

**Country:** The origin country of guest.

**market\_segment:** A group of people who share one or more similar characteristics

- TA: Travel agents

- TO: Tour operators

**distribution\_channel:** The medium through booking was made.[Direct,Corporate,TA/TO,undefined,GDS.]

**Is\_repeated\_guest:** Shows if the guest is who has arrived earlier or not.Values[0,1]-->0 indicates no and 1 indicated yes person is repeated guest.

**2. Introduction**

While dealing with hotel bookings in different popular market players some crucial factors contribute like type of hotel, room type, price, when the booking was made, length of stay, among other things.

Our aim here is to  explore and analyze the data to discover important factors that govern the bookings using Exploratory Data analysis technique to get useful insights.

## 3. Types of Hotels

## 1) Citi Hotel 2) Resort Hotel

## This shows how location matter for booking . Citi hotel has more bookings compare to resort hotel.

## Time of Booking Hotels

## The City hotel has more guests in July, August, September and In December and January there are less visitors. Thus, customers can get good deal on bookings in July and August in city hotel.

## Guest numbers for the Resort hotel go down slighty in November , December, January. More guest are in August, September.

## As we can see most bookings were made from July to August. And the least bookings were

## made at the start and end of the year.

## Stay for Number of days

## Most people stay for one, two, or three. More than 60% of guests come under these three options.

## Most of the customers preferred to stay for One, two, three days in City Hotel and One and 5 days in Resort Hotel.

## longer the stay length, the best price customer will get.

## Revenue per month per hotel

## Revenue of Resort Hotel are maximum in the month of August and least in the month of January and there is no high fluctuation in the price of city Hotels throughout the Year.

## Country and Guests

## Portugal, UK and France, Spain and Germany are the top countries from most guests come , more than 80% come from these 5 countries.

**. Steps involved:**

## Data Preprocessing:

## Data cleaning : To start with the very basic of data cleaning, let's find out if any of the columns have any Null or missing values and replace those with integers. As we can see from above column company,agent,country and children contains missing values. We can drop 'company' column as it contains too many missing values. In Children column we have 4 missing values, so we are filling those missing values with average value of children column and we are type-casting float values to integer value. Removing outlier data from dataset.

## Converting Datatype: Let’s check the datatype of each column in our dataset. There are some columns like children, company, and agent, that are float type but their values are only in integers.

## Exploratory Data Analysis and Visualisation : Now let’s do the fun part, extract the information from our data, plot the different graphs and try to answer our questions and get some useful insights.

**Conclusion:**

1. Almost 35% of bookings were canceled.
2. More than 60% of the population booked the City hotel.
3. More than double bookings were made in 2016, compared to the previous year. But the bookings decreased by almost 15% next year.
4. Most bookings were made from July to August. And the least bookings were made at the start and end of the year.
5. Portugal, the UK, and France, Spain and Germany are the top countries from most guests come, more than 80% come from these 5 countries.
6. Most people stay for one, two, or three.  
   -> For Resort hotel, the most popular stay duration is three, two, one, and four days respectively.  
   -> For City hotel, most popular stay duration is one, two, seven(week), and three respectively
7. Couple (or 2 adults) is the most popular accommodation type. So hotels can make arrangement plans accordingly