## ❖ INTRODUCTION

The hospitality industry is a broad group of businesses that provide services to customers. It's focused on the satisfaction of customers and providing specific experiences for them.

It includes hotels, tourism agencies, restaurants and bars. The hospitality industry is unique because it relies so heavily on discretionary income and free time. The goal of the hospitality industry is to provide customers with an enjoyable experience. Whether that enjoyment comes from eating a good meal, relaxing in a luxurious spa, or getting a good night's rest away from home

Hotel Booking process starts with the booking inquiry, checking the room availability, price per night, cancellations, lead time, meal preference, and among others. Hotel can be booked via online or through trave agent depending upon the customers preference. However online booking of the hotel is most famous type of booking method.

## 1. Problem statement:

The main objective of this project is to understand the features that plays an important role in deciding the booking factor of different hotel types. We formulated important questions like what is the average daily rate, what is the cancellation rate, what is the number of special requests, which hotel type is most booked, which months are busiest etc. which in the end will lead to meaningful insights.

## 2. Methodology

- **Exploratory Data Analysis (EDA)**

EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better.

In statistics, exploratory data analysis is an approach of analysing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.

We performed exploratory data analysis to decide the factors that impacts the hotel bookings.

## 3. Data understanding

The dataset contains the booking information of city hotel and resort hotel. There are 119390 entries and 32 columns present in the dataset. The hotel bookings are seen from 1st July of 2015 till 31st August 2017 with the customers effectively arriving and cancellation of bookings.

- CITY HOTEL

It provides accommodation and meals to travellers. city Hotels are often located near major transportation corridors like freeways, airports, or main streets.

- RESORT HOTEL

The resort hotel is a luxury facility that is intended primarily for vacationers and is usually located near special attractions, such as beaches and seashores, scenic or historic areas, ski parks, or spas

Columns present in the dataset

- **Hotel**
  - Resort hotel
  - City hotel

- **is_canceled**
  - 1: Cancelled
  - 0: Not cancelled
- **lead_time**
  - No of days that elapsed between entering date of booking into property management system and arrival date
- **arrival_date_year**
  - Year of arrival date (2015-2017)
- **arrival_date_month**
  - Month of arrival date

- **arrival_date_week_numberr**
  - Week number of year for arrival date (1-53)
- **arrival_date_day_of_month**
  - Day of the arrival date
- **stays_in_weekend_nights**
  - Number of weekend nights (Sat/Sun)

- **stays_in_week_nights**
  - Number of week nights (Mon - Fri)
- **Adults**
- **Children**
- **Babies**
- **meal**
  - 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)
- **country**
- **market_segment**
  - TA: Travel agents (online or offline)
  - TO: Tour operators
  - Direct
  - Corporate

    - Groups
    - Complementary
    - Aviation
- **distribution_channel**
    - TA: Travel agents
    - TO: Tour operators
    - Direct
    - Corporate
    - GDS
    - Undefined
- **is_repeated_guest** (value indicating if the booking name was from repeated guest)
    - 1: Yes
    - 0: No

- **previous_cancellations**
    - Number of previous bookings that were cancelled by the customer

- **previous_bookings_not_canceled**
    - Number of previous bookings not cancelled by the customer prior to the current booking
- **reserved_room_type**
    - The room type reserved for the customers.

- **assigned_room_type**
    - The room assigned to the customers.
- **booking_changes**
    - Number of changes/amendments 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 – no deposit was made;

    - Non-Refund – a deposit was made in the value of the total stay cost;
    - Refundable – a deposit was made with a value under the total cost of stay.
- **agent -**ID of the travel agency that made the booking
- **company**
    - ID of the company/entity that made the booking or responsible for paying the booking.
- **day_in_waiting_list**
    - Number of days the booking was in the waiting list before it was confirmed to the customer.

- **customer_type:**
    - Contract - when the booking has an allotment or other type of contract associated to it;
    - Group – when the booking is associated to a group;
    - Transient – when the booking is not part of a group or contract, and is not associated to other transient booking;
    - Transient-party – when the booking is transient, but is associated to at least other transient booking
- **adr (average daily rate)**
- **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**
    - Cancelled – booking was cancelled by the customer;
    - Check-Out – customer has checked in but already departed;
    - No-Show – customer did not check-in and did inform the hotel of the reason why
- **reservation_status_date**
    - Date at which the last status was set.

## 4. Data Wrangling

Data wrangling is the process of cleaning the dataset from null and duplicate values. EDA performed on the clean dataset results in a better visualisation of different features and the data interpretation is more accurate.

- **Loading the dataset:**

We are using Google Collab which allows users to write and execute arbitrary python code through the browser and its well suited for data analysis and machine learning

The Hotel booking dataset is provided by Almabetter.

- **Libraries used:**

For loading and visualising the dataset we used following libraries:

**Numpy:** NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, Fourier transform, and matrices

**Pandas:** Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

**Matplotlib:** Matplotlib is a cross-platform, data visualization and graphical plotting library for Python

**Seaborn:** It is used for data visualization and exploratory data analysis. Seaborn works easily with data frames and the Pandas library. The graphs created can also be customized easily.

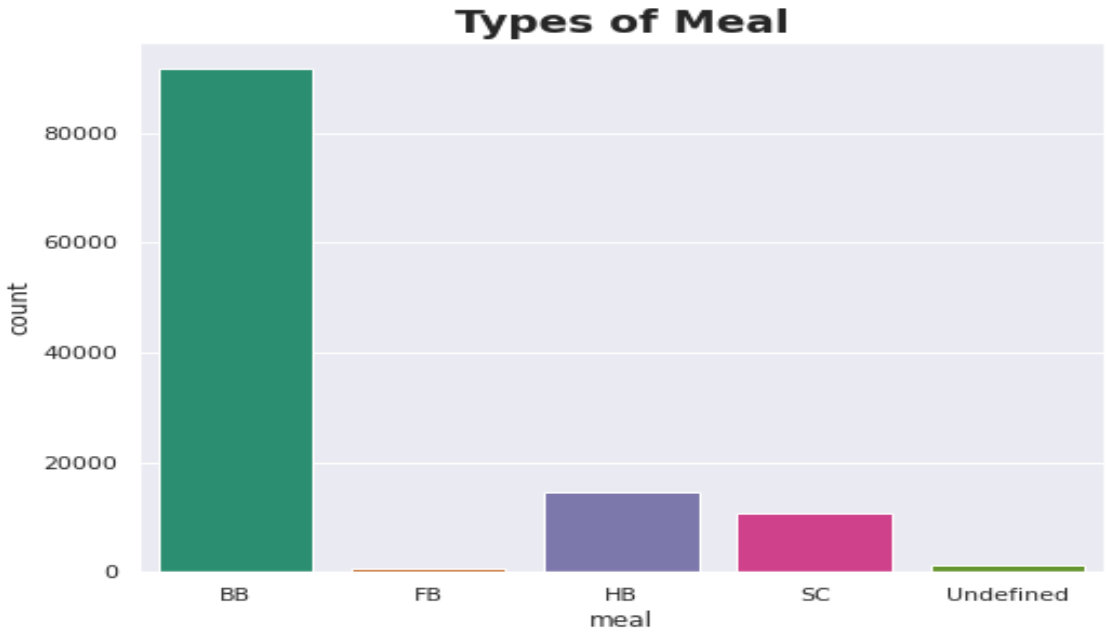
**Missingno:** It provides a series of visualisations to understand the presence and distribution of missing data within a pandas data frame.

## 5.Data visualisation

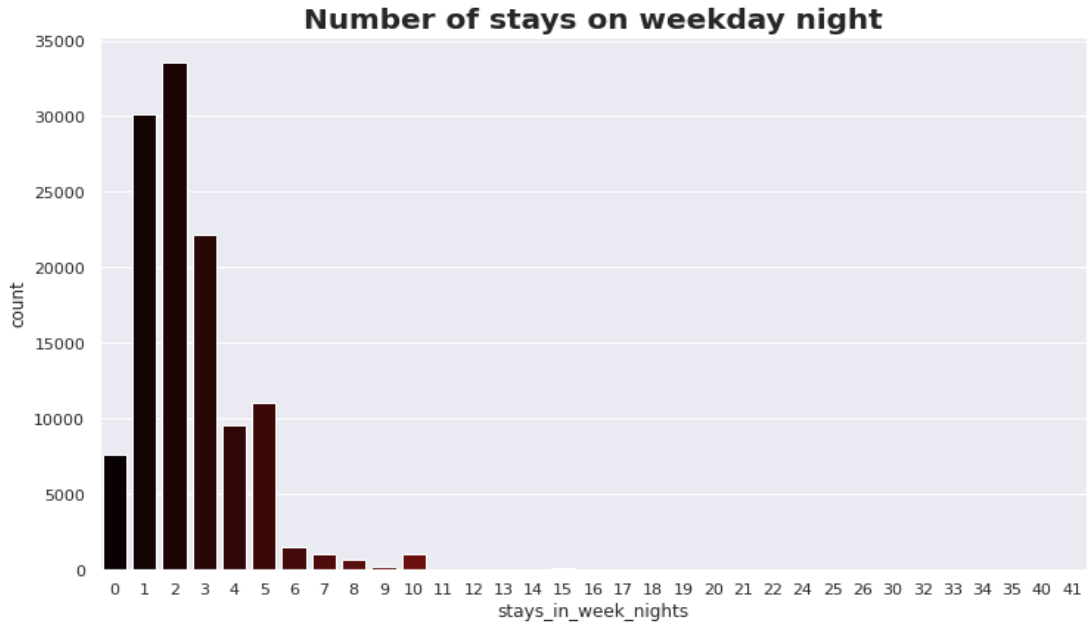
Once we cleared the data our next step is to visualise the data for a clear understanding of different features.

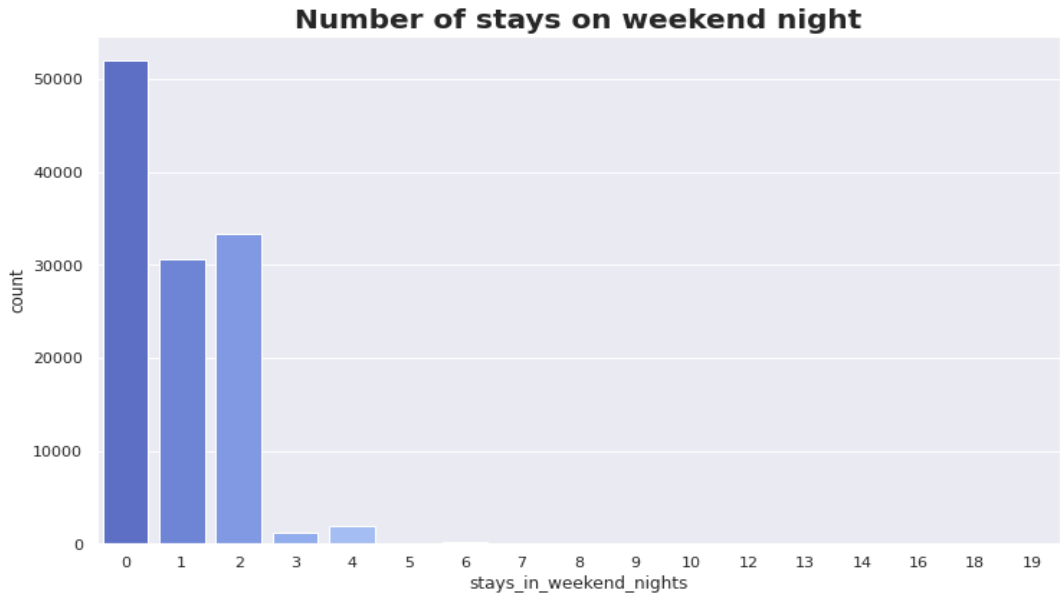
Below are few highlights from the analysis:

**A) Most preferred meal type:**

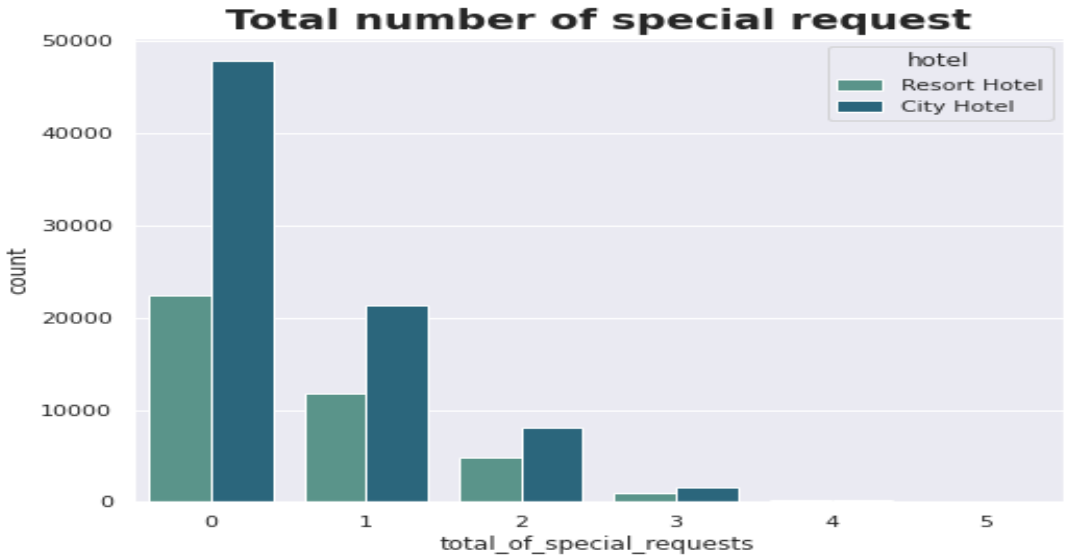



**B) Number of stay on weekday and weekend night:**






## C) Special request



# 6. CHALLENGES FACED

- **Size:** Big dataset often leads to problem like poor data quality, solving the wrong problem, skills shortage, dated data and inability to operationalize insights etc. there are 32 variables present in the given dataset and finding out the most important factors and their relevance was difficult.

- **Null (NaN) values:** There are four columns in our dataset that contains null values. They are agent, company, country and children which were filled with 0. NaN values in textual fields like company for instance were filled with ‘unknown’.

## 7. Conclusion:

After performing the EDA on the given dataset of hotel booking, we understood the important factors that govern the hotel booking. This will help the decision makers to plan accordingly for improving the business. The insights that we drew are as follows:

- Majority of the hotels booked are city hotel. City hotels are cheaper than resort hotel, thus explaining such high bookings than resort hotel.
- Number of special requests made in City hotel are higher than resort hotel. If the special request is met than people will eventually book the same type of hotel for the next time.
- The stay on weekday night is higher than weekend night. The average stay on weekend night is 1 whereas the average stay on weekday night is 3. More focus should be done on improving the quality for bookings made during weekdays.
- Bed and Breakfast (BB) is the most preferred meal type of the customers.
- The busiest months for hotel bookings are July and August and least bookings are made during winter time. Summer time is busiest for hotel bookings and slowest during summer.

- Number of non-repeated guest are very less. Companies should focus on having the customer book for a second time.
- Cancellation rate is very high. No pre-requisite of deposition could be explained as one of the main reasons that leads to high cancellation rate.
- Maximum customers come from western European countries.