Hotel Booking Analysis

**Aarti Gade**

**Kunal Badgujar**

**Vivek Tanagawade**

**Data science trainees, Alma Better, Bangalore**

# Abstract:

Hotels plays a major role in tourism industry. People travel around the world and to relax they definitely need a hotel to stay and have food of different cultures. We had concern of many things related to hotels during our journey.

They are, when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests?

In order to achieve this, we can use data visualization method with several datasets and predict the possibility of the best outcome for the customer to be get satisfied.

# Problem Statement

Data of different hotels (i.e., excluding the personal information) like 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 are gathered. Using this information obtain the results which helps the customer to solve different problems like when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate?

What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests.

This approach builds a customer friendly platform in order to find the best results and solve the multiple problems, The data includes.

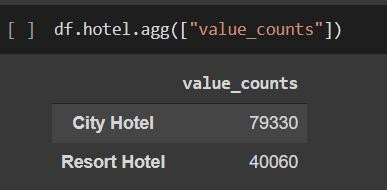
* + Number of hotels available.
  + How many people are cancelled the booking?
  + Lead Time
  + Arrival date year
  + Arrival date month
  + Arrival date week number
  + Arrival date day of month
  + Stays in weekend nights
  + Stays in week nights
  + Adults, children, babies,
  + Meal, country, market 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

# Introduction

Hotel industry is a very volatile industry and the bookings depend on variety of factors such as type of hotels, seasonality, days of week and many more. This makes analyzing the patterns available in the past data more important to help the hotels plan better. Using the historical data, hotels can perform various campaigns to boost the business. We can do EDA to predict the future bookings, most engaged months of coming year, additional facilities which can attract more customers and based upon the data

# 3. List of observations

1. In the given dataset, to find which hotel contain maximum bookings. This can be done by plotting the graphs.



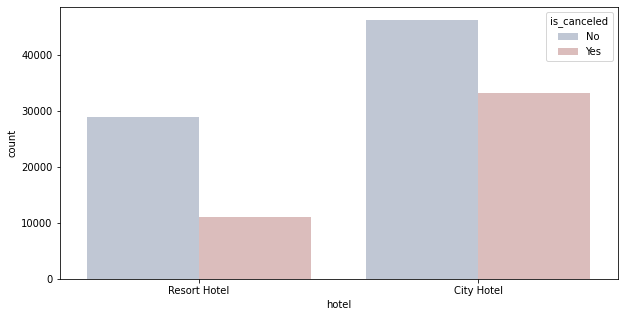
In the above figure we get the value counts for each hotel bookings.



After plotting the graph, we get this figure as an output, this is a bar graph which are labelled with City hotel and Resort Hotel.

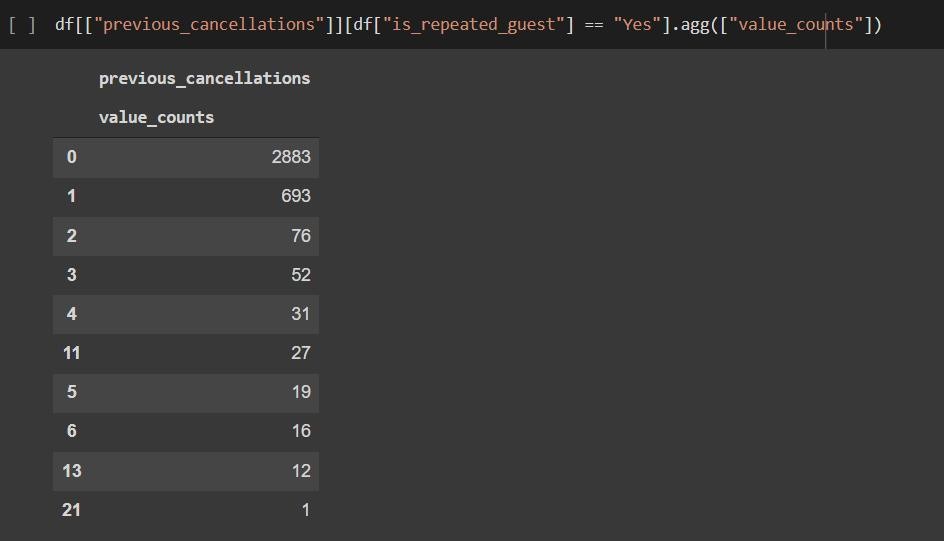
1. In the same way we can plot the graph for bar graph, i.e. for the hotel have the maximum number of cancellation of bookings

In the above figure we can observe that we group each hotel with is canceled either yes or no and get the total count respectively.

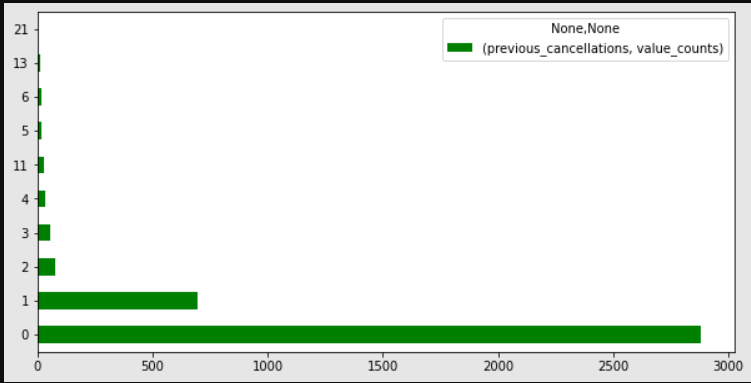


This is the graph which indicates the person is cancelled the booking for Resort or City hotel. We have the count of cancelled and not cancelled list of each hotel from previous figure.

1. To count the previous cancellation by repeated guests. In order to get this, we need to check is\_repeated\_guest is Yes or No.

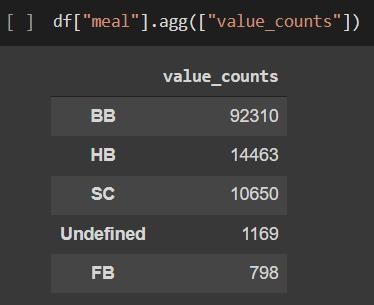


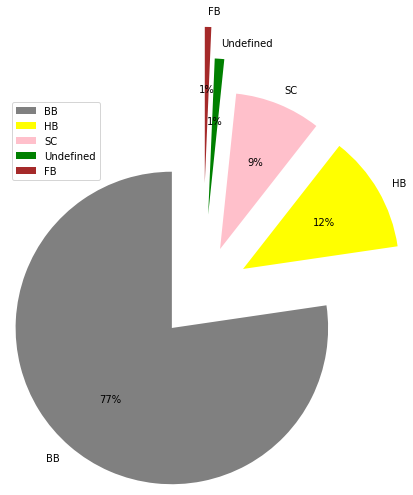
The above figure indicates previous cancellation and their total value counts respectively.



Plotting the graph for the above table which provides total value counts for previous cancellation.

1. Now we need to analyze the preference of the guest like what they basically prefer.

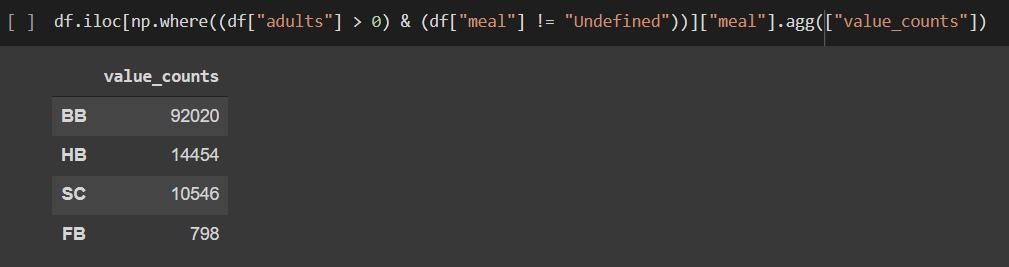


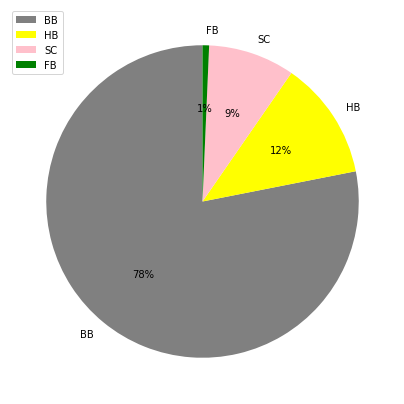
In the above figure we can see that total meal counts are displayed for BB, HB, SC FB and Undefined respectively.

In this figure a pie graph is plotted for the previous data table which provides total meal counts for respective data frames.

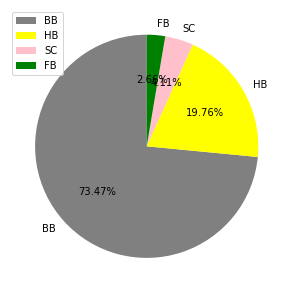
Like the same way,

* Adults favourite and least favourite meal is.

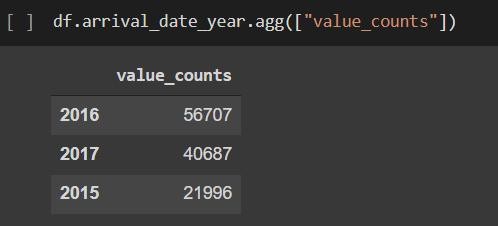




* The babies favorite and least favorite meals



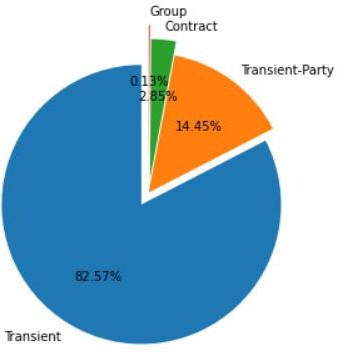
1. The most busy year could be found by using arrival date year field. Using agg function we get the total value count for arrival date year.



There are three data fields obtained after running the above code. 2016,2017 and 2016 along with their value count.



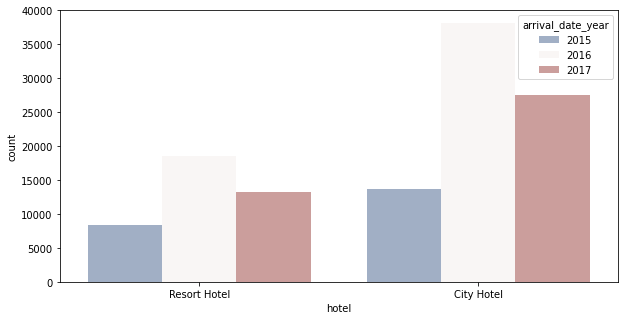
Upon plotting a graph for the arrival date year total value count we get this as a output.

1. How many guests arrived year-wise could be found by grouping hotel into arrival date year and fetching total view count by using agg function.

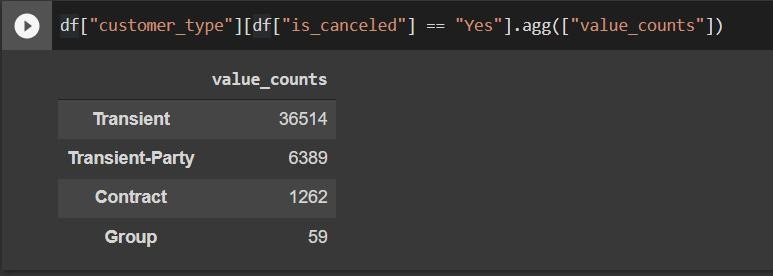
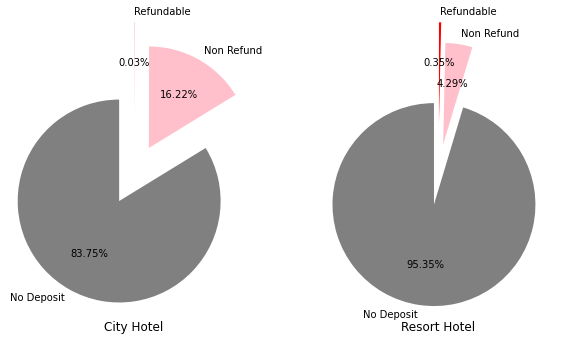


The above figure shows the total view count for arrival date year for respective hotels.

Upon plotting the graph we get,



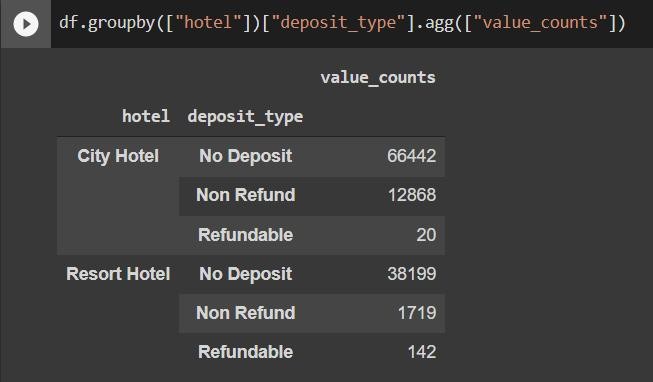
1. Which type of customers are more likely to cancel the booking, this can be get by mapping customer type with is cancelled field is yes.



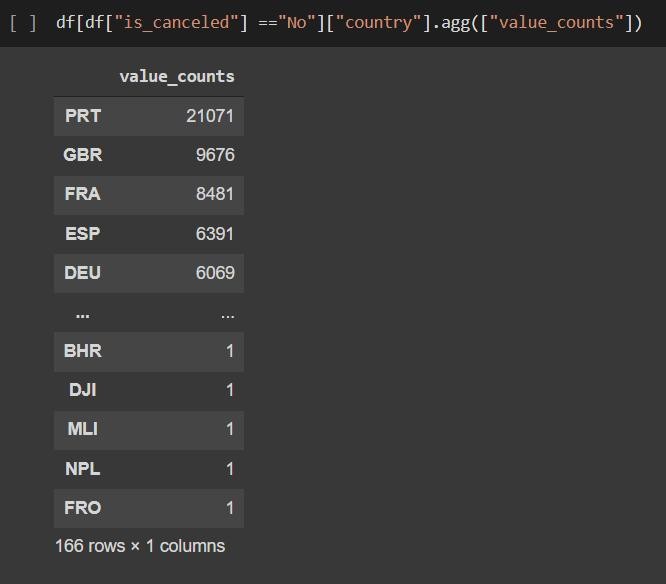
The above figure gives the total count for the customers are more likely to cancel the booking.

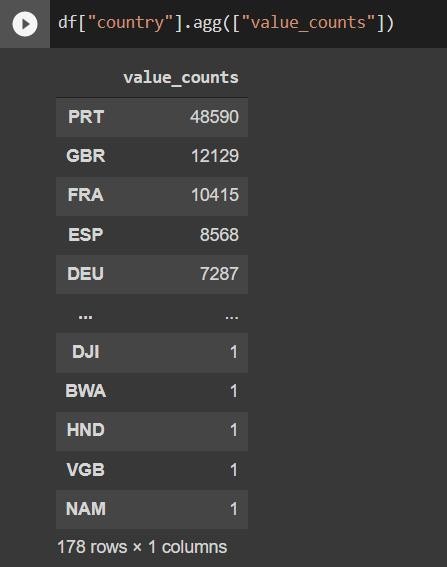
Above graphs are obtained from the table results.

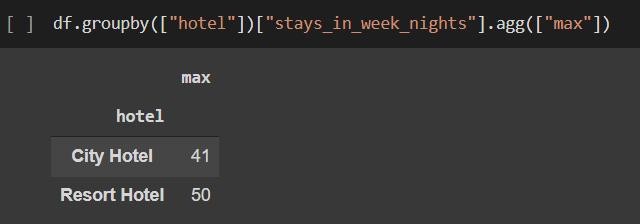
1. Deposit Type hotel-wise, to get this group by is to be done for the hotel and deposit type and finding the total value count.



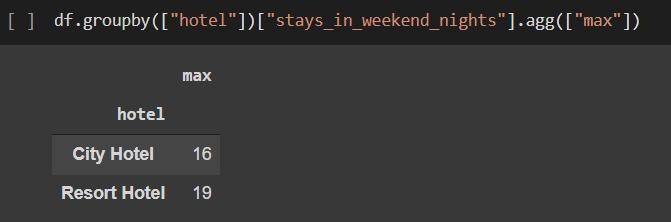
1. Number of guests who had not cancelled their booking, this can be achieved by using Is Cancelled is No and mapping with country. Using agg function total value count is obtained.



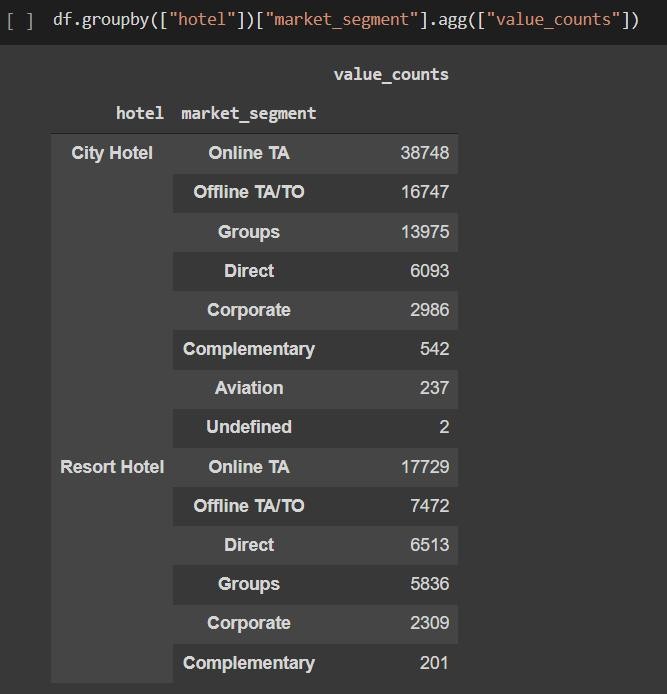
1. From where the most guests coming, this can be easily achieved by total value count for each country.
2. Maximum number of stays in week nights in each hotel, this can be achieved by grouping by hotel with stays\_in\_week\_night field and finding the max value by agg function.

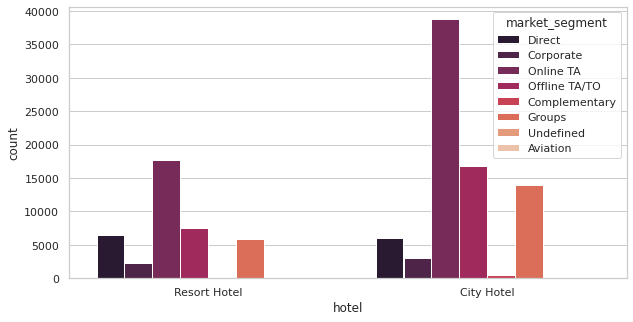


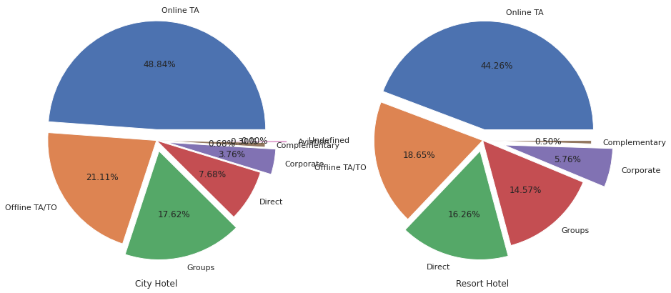
1. Maximum number of stays in weekend nights in each hotel, this can be achieved by grouping by hotel with stays\_in\_week end night field and finding the max value by agg function.



1. From where most of the bookings were made, this can be achieved by grouping by hotel with market segment field and finding the value count by agg function.







The above graphs are plotted by the results obtained from the table in previous figure.

# Conclusion:

That's all there is to it! We have reached the end of our exercise. The data has been loaded, null values have been treated, categorical columns encoded, and major reasons that govern hotel bookings have been identified, along with steps to increase them.

1. Around 60% bookings are for city hotels and around 40% bookings are for resort hotels therefore city hotel is busier than resort hotel ,also overall adr of city hotel is slightly higher than resort hotel.

2. The majority of reservations are for city hotels.

3. The number of repeated guests is too low.

4. Most of the bookings either in the canceled or checkout done by online TA.

5. City hotels and resort hotels maximum number of bookings by online TA.

6. More visitors are from western europe, namely Portugal,France, Great Britain, and Spain being the highest.

7. Families with children have no particular preference for the hotel type.

8 disproportionately high number of special requests? etc. Are easily identified and a proper way of analyzing the huge dataset and converting into easily understandable method will be a better choice.

## **Challenges:**

(1) There was a lot of duplicate data.

(2) Data was present in wrong datatype format.

(3) Choosing appropriate visualization techniques to use was difficult.

(4) A lot of null values were there in the dataset.

**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>