# :: Capstone Project Hotel Booking Analysis - EDA







#### **Table of contents**

Agenda of Data
O1 Analysis

Benefits for doing this **Hotel Booking Analysis** Project

Data Description & Summary

Dataset Summary & Description of all the data

Hotel Booking

Data-Introduction

Short introduction of the **Dataset** 

**EDA (Exploratory Data Analysis)** 

04

**Data Visualization**, Overall Stats and **Conclusion** 

#### **Agenda of Data Analysis**

- Have you ever wondered when the best time 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 which hotel likely to receive a disproportionately high number of special requests?

• • • • • •

So, this hotel booking dataset can help us to explore all of this questions!



This Dataset contains data that compares various booking information between two hotels, City Hotel and Resort Hotel. So, here i will be using the data to analyze the factors affecting the hotel bookings. These factors can be used for reporting trends and predict the future bookings.

This Dataset contains the booking data of hotels from year 2015-2017.

hotel	There are two types of hotels, one is City Hotel and another is Resort Hotel
is_canceled	Here 0 and 1 value indicates booking was cancelled (1) or not (0)
lead_time	Time-lapse between reservation and actual arrival date
arrival_date_year	Year of arrival date
arrival_date_month	Month of arrival date
arrival_date_week_number	Week number of arrival date
arrival_date_day_of_month	Day of arrival date
stays_in_weekend_nights	Number of weekend nights (Saturday or Sunday) spent at the hotel by the guests

stays_in_week_nights	Number of weeknights (Monday to Friday) spent at the hotel by the guests
adults	Number of adults among guests
children	Number of children among guests
babies	Number of babies among guests
meal	Type of meal booked
country	Country of guests
market_segment	Designation of market segment
distribution_channel	Name of booking distribution channel

is_repeated_guest	If the booking was from 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 of room type assigned
booking_changes	Number of changes/ amendments made to the booking
deposit_type	Type of the deposit made by the guest
agent	ID of travel agent who made the booking

company	ID of the company that made the booking
days_in_waiting_list	Number of days the booking was in the waiting list
customer_type	Type of customer, assuming one of four categories
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
reservation_status	Reservation status (Canceled, Check-Out or No-Show)
reservation_status_date	Date at which the last reservation status was updated

#### **Data Summary**

- This data set contains a single file which compares various booking information between two hotels: City Hotel and Resort Hotel. 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.
- The dataset contains a total of 119390 rows and 32 columns.
- All the columns are divided into three dtypes: Object, float64 and int64.
- This dataset does have duplicated values as well as null values. There are total of
   31994 duplicate values and four columns have missing values/ null values.
- The maximum number of missing values are from 'Company' column then followed by 'Agent', 'Country' and 'Children' columns. The 'Children' column consists of only 4 null values, while 'Company' column consists of 112593 null values.

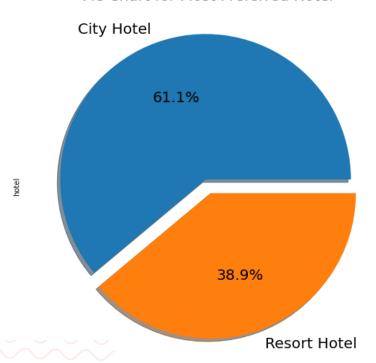
#### **Points for Discussion**

- Top Most Preferred Hotel & Hotel Rooms
- ADR (Average Daily Rate)
- Repeated Vs Non-Repeated Guests
- Requirement of Car Parking Space
- Most Preferred Meal
- Distribution Channel Vs ADR
- Top Booking Months & Year
- Optimal Stay Length
- Confirmation Vs Cancellation
- Mostly Arrived Customers/ Visitors
- Overall Stats
- Conclusion



#### **Top Most Preferred Hotel**

Pie Chart for Most Preferred Hotel



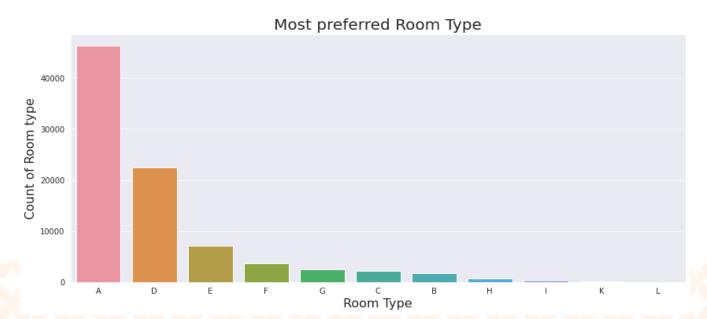
#### **Insights found**

From the chart, we got to know that **City Hotel** is **most preferred hotel** by the guests. Thus **City Hotel** has **maximum bookings**. **61.1% guests** are preferred **City Hotel**, while only **38.9% guests** have shown interest in **Resort Hotel**.



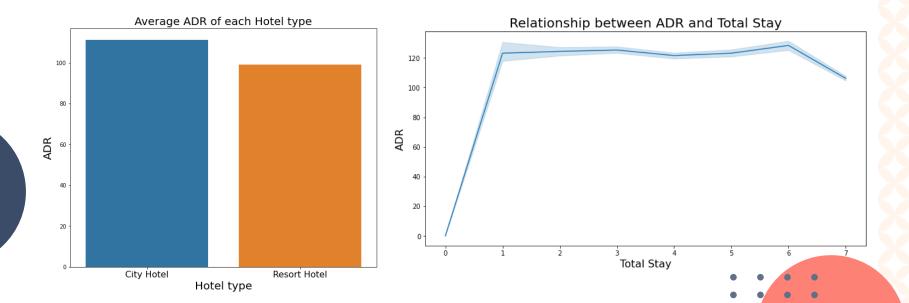


- It is found that the **most preferred Room** type is 'A'. So, majority of the guests have **shown interest** in this **room type**.
- There are **positive impacts** because 'A', 'D', 'E' is **more preferred** by **guest** due to **better services** offered in room type.



Average Daily Rate
City Hotels are generating more revenues than the Resort Hotels, because City

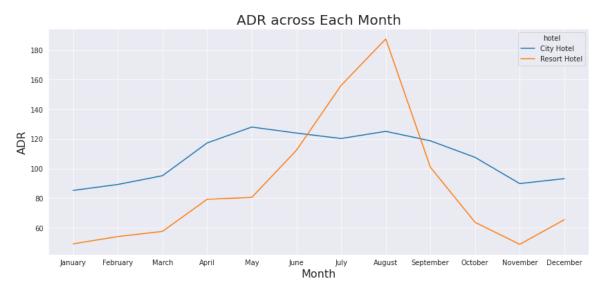
- City Hotels are generating more revenues than the Resort Hotels, because City hotel has the highest ADR. More the ADR, more will be the revenue.
- From the **line chart**, we have found that as the **total stay** increases the **ADR** is also getting **high**. So, **ADR** is directly **proportional** to **total stay**.



# **Average Daily Rate**

#### > ADR across months

• For Resort Hotel, ADR is high in the months of June, July, August as compared to City Hotels. The reason may be that customers/people want to spent their summer vacation in Resort Hotels.

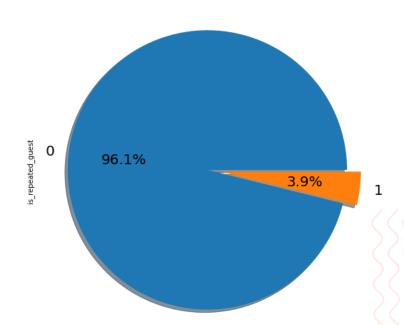


### **Repeated Vs Non-Repeated Guests**

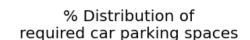
#### **Insights found**

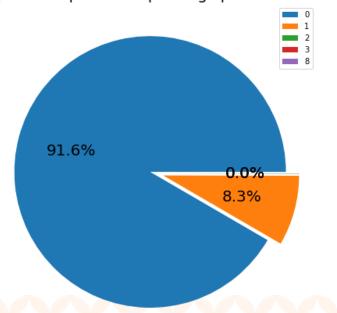
- The pie chart show the percentage of repeated guests or not (where 0 is not repeated guest and 1 is repeated guest)
- Repeated guests are very few which is only 3.9% while 96.1% guests are not returning to the same hotel.
- The guests management should take feedbacks from guests and try to improve the services.

Percentage (%) of repeated guests



## **Requirement of Car Parking Space**





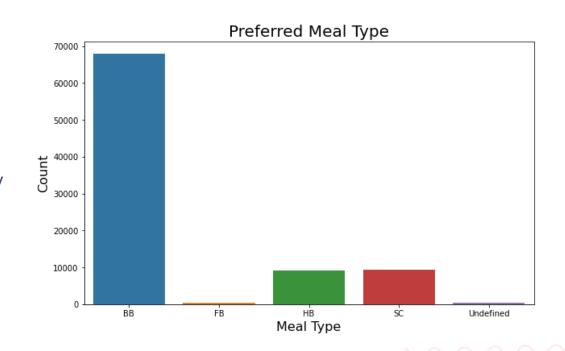
#### **Insights found**

- This chart shows that 91.6% guests did not required the parking space.
   Only 8.3% guests required the parking space.
- The demand for car parking area is less. It can be said that hotels need to work less on car parking spaces as only 1 car parking space was required by 8.3% of guests.

#### **Most Preferred Meal**

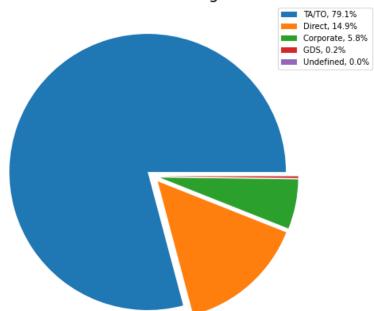
#### **Insights found**

The most preferred meal type by the guests is BB (Bed and Breakfast) while HB (Half Board) and SC (Self Catering) are equally preferred.



#### **Maximum used Distribution Channel**

Mostly used Distribution Channel for Hotel Bookings



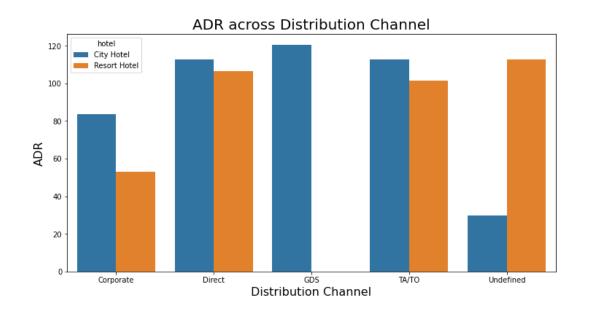
#### **Insights found**

'TA/TO' has been mostly (79.1%) used for booking hotels. Direct market segment of 14.9%, Corporates market segment of 5.8%, GDS market segment of only 0.2% and rest unidentified are 0%.



#### **Distribution Channel Vs ADR**

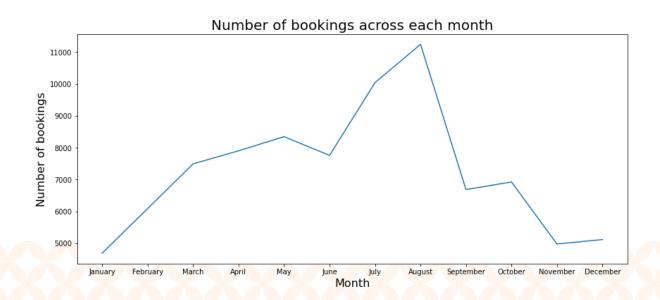
 'Direct' and 'TA/TO' have almost equally contribution in ADR in both type of hotels. While, GDS has highly contributed in ADR in 'City Hotel' type. GDS needs to increase Resort Hotel bookings.





### **Top Booking Months**

- From this graph, we can say that July and August months had the most Bookings.
   As, July and August generally surrounds in and near the summer vacation.
- Hotels should be well prepared for the month of July and August as maximum bookings takes place for this month.



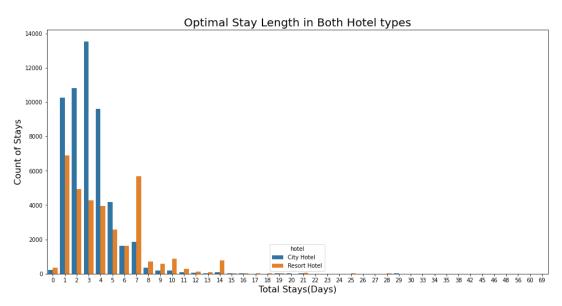
## **Top Booking Year**

• It can be summarised that in the year 2016 both the hotel saw a massive increase in their bookings and by far the year 2016 is the year of the highest bookings of both hotel. In 2016 and 2017 the City hotel is having the highest number of bookings but in 2015 the Resort hotel is having the highest number of bookings.



# **Optimal Stay Length**

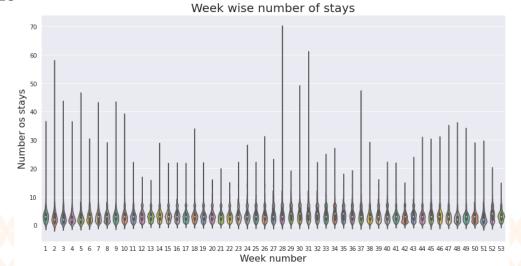
- We have found that the Optimal stay in both the type hotel is less than 7 days. So, after that staying numbers have declined drastically.
- Customers usually prefers a one week stay in a hotel. So, hotels need to work efficiently
  in these seven days so that customers would return to the same hotel again.





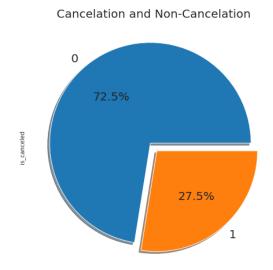
# Optimal Stay Length > Week Wise Number of Stay

- From the week **28 to 31**, it has shown the **highest days** of **stay** whereas from the week 1 to 11 has shown a very steady trend in the number of stays and also the week 18 to 22 has shown the least number of stays by the visitors in aggregate of all 3 years 2015, 2016 and 2017.



#### **Confirmation Vs Cancellation**

- More than 1/4th of the overall bookings i.e. approx 27.5% of the tickets was got canceled.
- We can clearly deduce from the 2nd graph that the City hotel is having greater number of bookings as compared to Resort hotel. But, the cancelation percentage is also high of the City Hotel.





## **Mostly Arrived Customers/ Visitors**

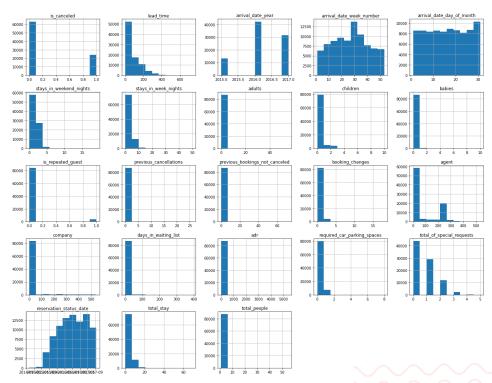
• It can be summarised that the **Transient** type of **customers** visit the **most** whereas the visitors who are in group comes in the category of **least visitors**.



#### **Overall Stats**

#### **Insights found**

- Maximum guest came in the year 2016.
- Maximum arrival week number is 30.
- Maximum arrival happens in the last of the month.
- Maximum guests comes with no children.
- There is very less requirement of Car parking spaces.



#### **Overall Stats**

								Co	-rel	latio	n o	f the	e co	lum	ns								
Hoolmon	is_canceled	1.00	0.18	0.09	0.00	0.01	0.06	0.08	0.08	0.07	-0.02	-0.09	0.05	-0.05	-0.09	-0.00	-0.08	0.00	0.13	-0.18	-0.12	0.09	0.10
Heatmap	lead_time	0.18	1.00	0.14	0.10	0.01	0.24	0.31	0.14	0.03	-0.00	-0.15	0.01	-0.08	0.08	0.08	-0.08	0.13	0.02	-0.09	0.03	0.32	0.13
	arrival_date_year	0.09	0.14	1.00	-0.51	-0.01	0.01	0.00	0.04	0.04	-0.02	0.02	-0.05	0.03	0.01	-0.00	0.02	-0.03	0.18	-0.04	0.06	0.00	0.05
	arrival_date_week_number	0.00	0.10	-0.51	1.00	0.09	0.03	0.03			0.01								0.10	0.01	0.05	0.03	0.03
	arrival_date_day_of_month	0.01	0.01		0.09				-0.00		-0.00								0.02	0.01	-0.00	-0.03	0.01
stays_in_weekend_nights		0.06	0.24			-0.02		0.55	0.09		0.01									-0.04		0.78	0.09
	stays_in_week_nights	0.08	0.31	0.00		-0.03			0.10		0.02									-0.04	0.04	0.95	0.10
adults		0.08	0.14	0.04							0.02									0.01			0.80
	children	0.07	0.03	0.04		0.02	0.03				0.02									0.04		0.03	0.60
	babies										1.00			_									0.17
	is_repeated_guest										-0.01			0.44			0.20				-0.00		-0.16
	previous_cancellations										-0.01						0.03					-0.02	
previo	ous_bookings_not_canceled		-0.08								-0.01						0.13						-0.11
	booking_changes	-0.09		0.01			0.03				0.08						0.09		0.01		0.02		0.00
	agent		-0.08								-0.01											-0.08	
	days in waiting list	0.00	0.13	-0.03	0.02						-0.01											-0.01	
	adr	0.13	0.13	0.18	0.10	0.01	0.04				0.02												0.38
n	equired car parking spaces			-0.04	0.10	0.01	-0.04	-0.04				0.07					0.04			1.00	0.05		0.03
	total of special requests	-0.12		0.06			0.03				0.09						-0.11						0.13
	total stay	0.09	0.32	0.00		-0.03	_		0.11					-0.06			-0.08			-0.05	0.04		0.11
	total people	0.10	0.13	0.05	0.03	0.01	0.09	0.10	0.80	0.60				-0.11						0.03	0.13		1.00
		is_canceled	lead_time	arrival_date_year	ate_week_number	y_of_month	in_weekend_nights	ys_in_week_nights	adults	children	babies	eated_guest	cancellations	ings_not_canceled	booking_changes	agent	company	waiting_list	adr	ar_parking_spaces	special requests	total_stay	total_people
				arriva	ate_we	late_day	in wee	ys_in_v				is rep	rious c	ings_n	booki			lays_in		ar_pari	of spec		

#### **Overall Stats**

#### **Insights found**

- > From Heatmap
  - is\_canceled and total\_stay are negatively correlated. This means
     customers are unlikely to cancel their bookings if they don't get the same
     room as per reserved room.
  - **lead\_time** and **total\_stay** is **positively correlated**. This means more **the stay** of customer is, more will be the **lead time**.
  - adults, childrens and babies are correlated to each other. This indicates more the people, more will be ADR.
  - is\_repeated guest and previous bookings not canceled have a strong correlation. This may be due to the reason that repeated guests are not more interested to cancel their bookings.

#### **Conclusion**

- **City hotels** are the **most preferred hotel** type by the guests. So, we can say that **City hotels** are the **busiest hotel** in **comparison** to the **resort hotel**.
- The average **ADR** of **city hotels** is **higher** as compared to the resort hotels. So, it can be said that these **City hotels** are generating **more revenue** than the **resort hotels**.
- The total stay of guests is directly proportional to the adr. So, higher the days of stay, the higher will be ADR and revenue as well.
- The percentage of repeated guests is very low. Only **3.9**% people had revisited the hotels. Rest **96.1**% were **new guests**. So, **retention rate** is much **low**.
- The **percentage** of required **car parking spaces** is very **low**. This means less car parking spaces **don't affect** the business much. Most of the customers **(91.6%)** do **not require** car parking spaces.
- Among different types of meals, BB (Bed & Breakfast) is the most preferred type
  of meal by the guests. So, guests love to opt for this meal type.

#### **Conclusion**

- 'Direct' and 'TA/TO' have almost equally contribution in ADR in both type of hotels. While, GDS has highly contributed in ADR in 'City Hotel' type.
- Optimal stay length in both the hotel types (City and Resort Hotel) is less than 7 days. Usually people stay for a week. So, after 1 week, the optimal stay length declined drastically.
- Most number of bookings have taken place in the month of July and August.
   July and August are the favourite months of guests to visit different places.
- The mostly used distribution channel for booking is 'TA/TO'. 79.1% bookings
  were made through TA/TO (travel agents/ tour operators).
- While calculating ADR across different month, it is found that for Resort hotel,
   ADR is high in the months of June, July, August as compared to City Hotels.
- Almost **1/4th** of the total bookings is **canceled**. Approx, **27.5% bookings** have got **canceled** out of all the **bookings**.
- Majority of the guests have shown interest in the room type 'A'. Room type 'A' is the most preferred room type.

