

Capstone Project

EDA- Hotel Booking

Team – Blinders

Team Members

1. UTTAM MISHRA
2. KSHITIJ SINGH

Introduction

- This data contains booking information of various hotels and resorts for a period of three years.
- It also includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies.
- All personal data such as Names , Transaction and Identification information have been removed from Data.

Data summary

- **Data set name-** Hotel Booking Analysis
- **Data type** – Comma Separated Values(CSV) file
- **Data shape**
- **Number of Rows** – 1,19,390 Rows
- **Number of Columns-** 32 Columns

Column description

1. **hotel** : Hotel(Resort Hotel or City Hotel)
2. **is_canceled** : Value indicating if the booking was canceled (1) or not (0)
3. **lead_time** : Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
4. **arrival_date_year** : Year of arrival date
5. **arrival_date_month** : Month of arrival date
6. **arrival_date_week_number** : Week number of year for arrival date
7. **arrival_date_day_of_month** : Day of arrival date
8. **stays_in_weekend_nights** : Number of weekend night
9. **stays_in_week_nights** : Number of week nights (Monday to Friday) the guest stayed at the hotel
10. **adults** : Number of adults
11. **children** : Number of children
12. **babies** : Number of babies

Column description

- 13. **meal** : Type of meal booked. Categories are presented in standard hospitality.
- 14. **country** : Country of traveller.
- 15. **market_segment** : Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”
- 16. **distribution_channel** : Booking distribution channel. The term “TA” means
- 17. **is_repeated_guest** : Value indicating if the booking name was from a repeated guest (1) or not (0)
- 18. **previous_cancellations** : Number of previous bookings that were cancelled by the customer prior to the current booking
- 19. **previous_bookings_not_canceled** : Number of previous bookings not cancelled by the customer prior to the current booking
- 20. **reserved_room_type** : Code of room type reserved. Code is presented instead of designation for anonymity reasons.
- 21. **assigned_room_type** : Code for the type of room assigned to the booking.
- 22. **booking_changes** : Number of changes/amendments made to the booking

Column description

- 23. **deposit_type** : Indication on if the customer made a deposit to guarantee the booking.
- 24. **agent : ID** of the travel agency that made the booking
- 25. **company : ID** of the company/entity that made the booking or responsible for paying the booking.
- 26. **days_in_waiting_list** : Number of days the booking was in the waiting list before it was confirmed
- 27. **customer_type** : Type of booking
- 28. **adr** : Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights
- 29. **required_car_parking_spaces** : Number of car parking spaces required by the customer
- 30. **total_of_special_requests** : Number of special requests made by the customer (e.g. twin bed or high floor)
- 31. **reservation_status** : Reservation last status, assuming one of three categories Canceled – booking was canceled 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

Problem Statement

Have you ever wondered 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 hotel booking dataset can help you explore those questions!

Using this dataset we have to do a Exploratory Data Analysis to find out relevant information to answer above questions.

Data Cleaning

- Initially DataSet had 1,19,390 Rows and 32 Columns.
- There were many null and duplicated values were present in the dataset.
- To get rid of these values we performed required cleaning.
- After cleaning and removing duplicated values the following was the result of clean rows and columns
- Rows = 87,223
- Columns = 31
- After cleaning the data we filled null values.
- Now we will perform EDA on the cleaned data.

1. Finding types of Hotels available

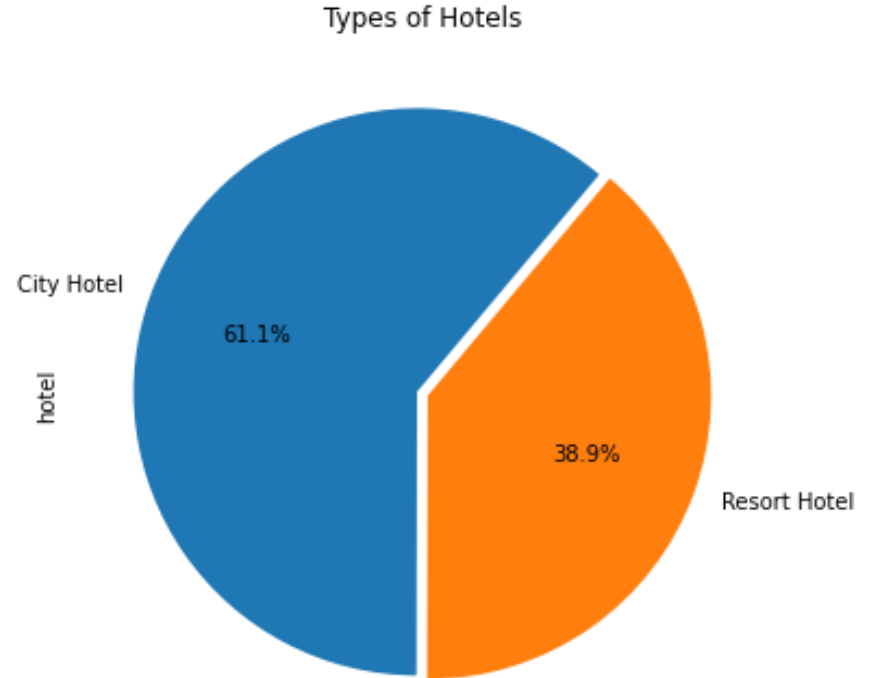
Output-

Number of hotels-

- City Hotel 53272
- Resort Hotel 33951

Conclusion-

As seen in the plot 61% of the hotels are City Hotel



2. Types of Rooms available in all hotels

Output -

A - 56430

D - 17376

E - 6035

F - 2820

G - 2050

B - 996

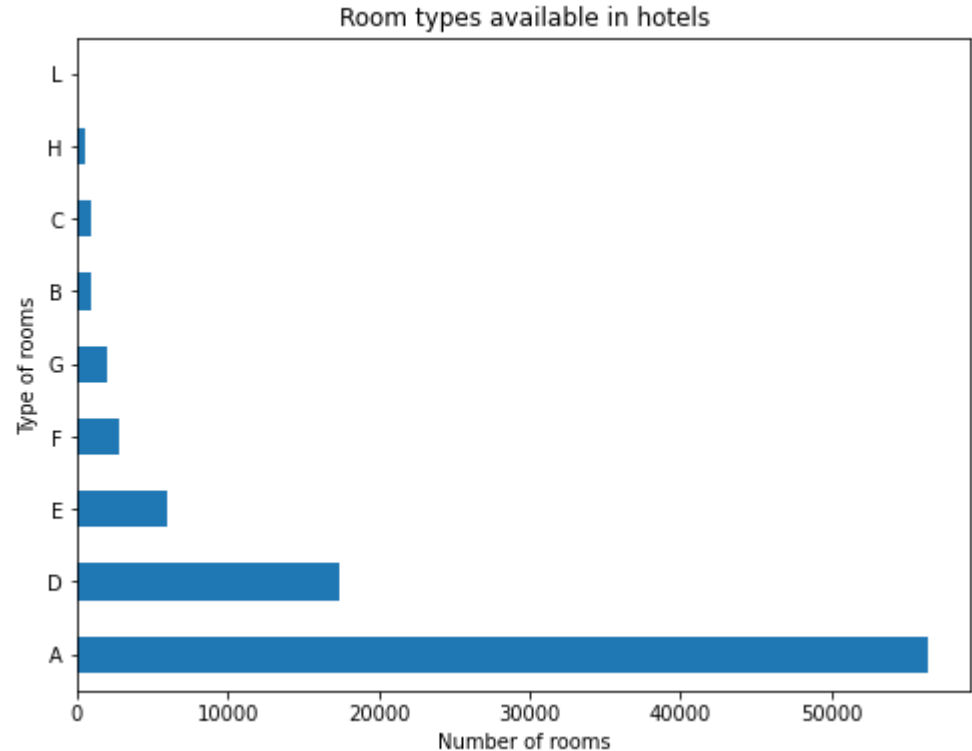
C - 914

H - 596

L - 6

Conclusion-

As seen from the plot Most available type of room is - A



3. Number of People who booked hotel

Number of Adults who booked Hotel

	count	mean	std	min	25%	50%	75%	max
hotel								
City Hotel	53272.0	1.881777	0.535630	0.0	2.0	2.0	2.0	4.0
Resort Hotel	33951.0	1.875703	0.736778	1.0	2.0	2.0	2.0	55.0

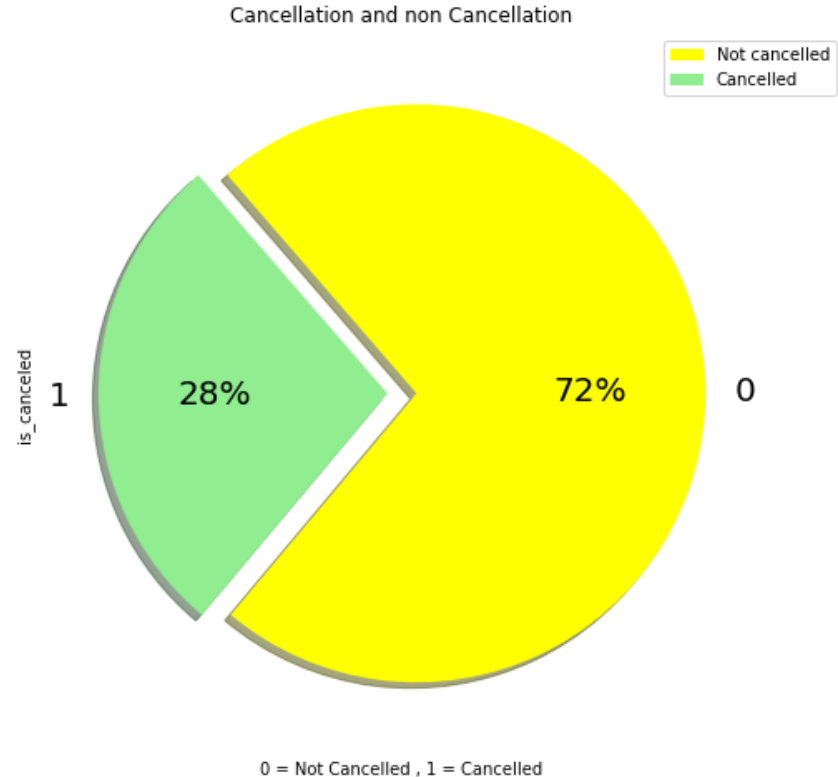
Number of Children who booked Hotel

	count	mean	std	min	25%	50%	75%	max
hotel								
City Hotel	53272.0	0.132227	0.442750	0.0	0.0	0.0	0.0	3.0
Resort Hotel	33951.0	0.149392	0.476558	0.0	0.0	0.0	0.0	10.0

4. Checking the percentage of booking cancellations

Observation-

It is clear from the pie chart that 28% of total bookings were cancelled



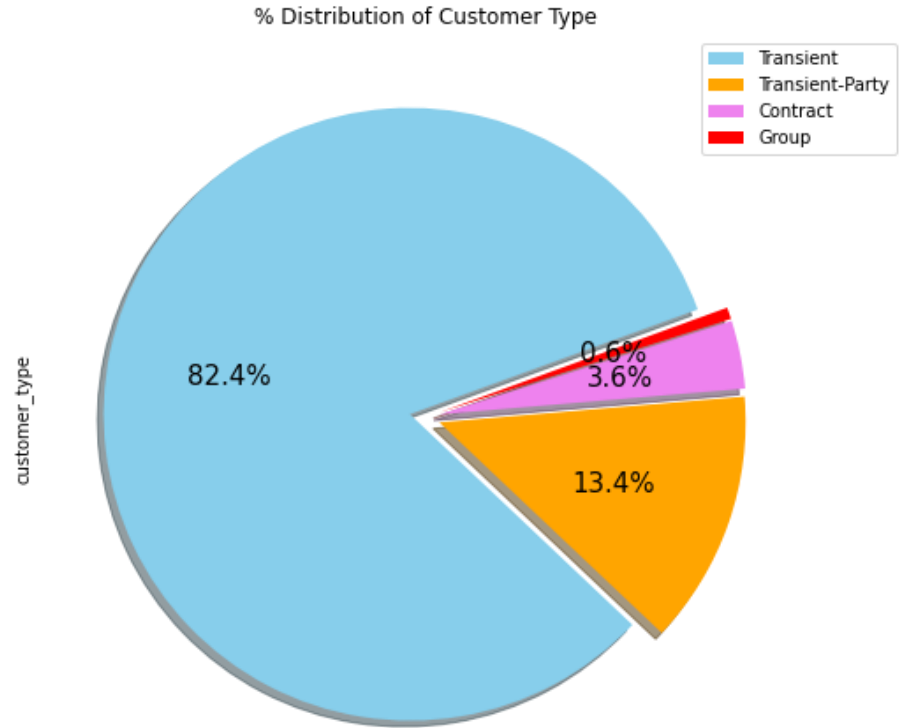
5. Finding the percentage distribution of "Customer Type"

Output -

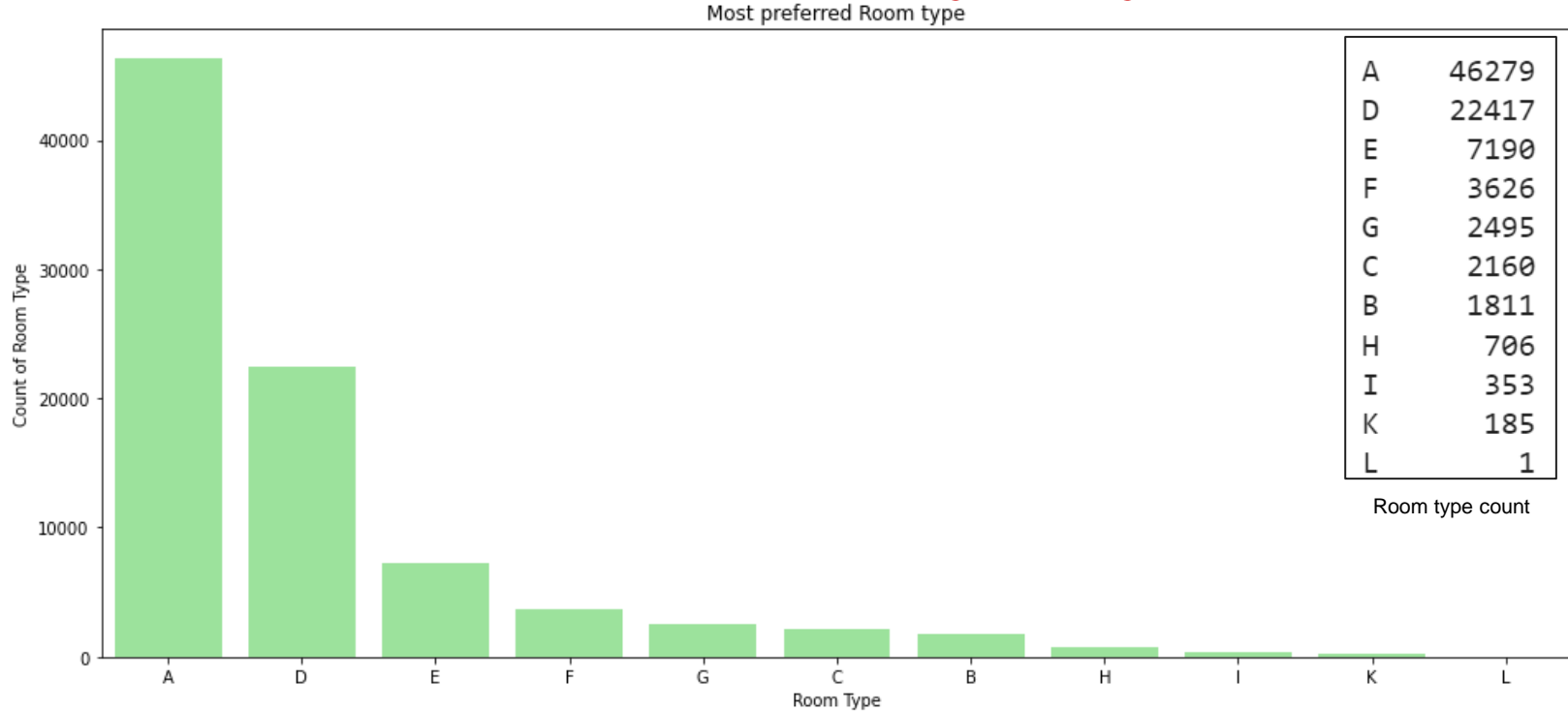
- Transient 71858
- Transient-Party 11689
- Contract 3135
- Group 541

Conclusion –

As seen from pie chart 82.4% of the customers were of Transient category followed by Transient-Party and others.

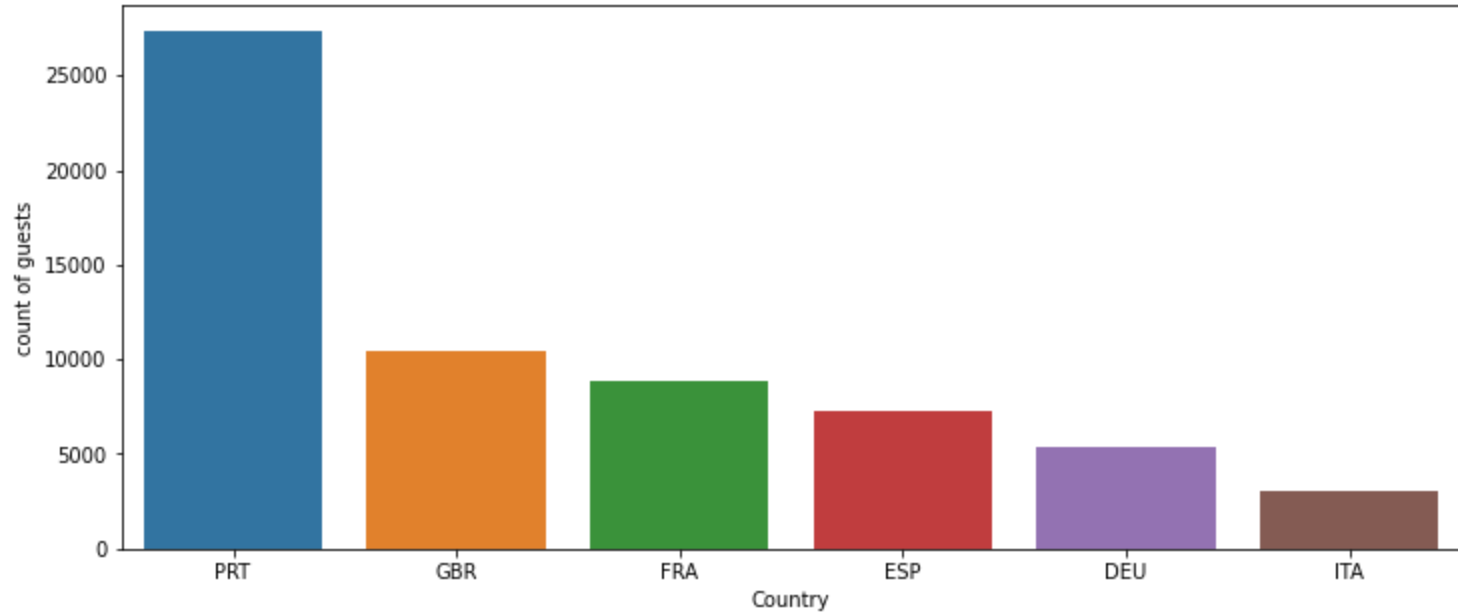


6. The most preferred room type by the customers



Conclusion – Most preferred Room is of 'A'-type

7. From which country the most guests are coming?



Conclusion- Most of the guests are coming from Portugal i.e. more 25000 guests are from Portugal

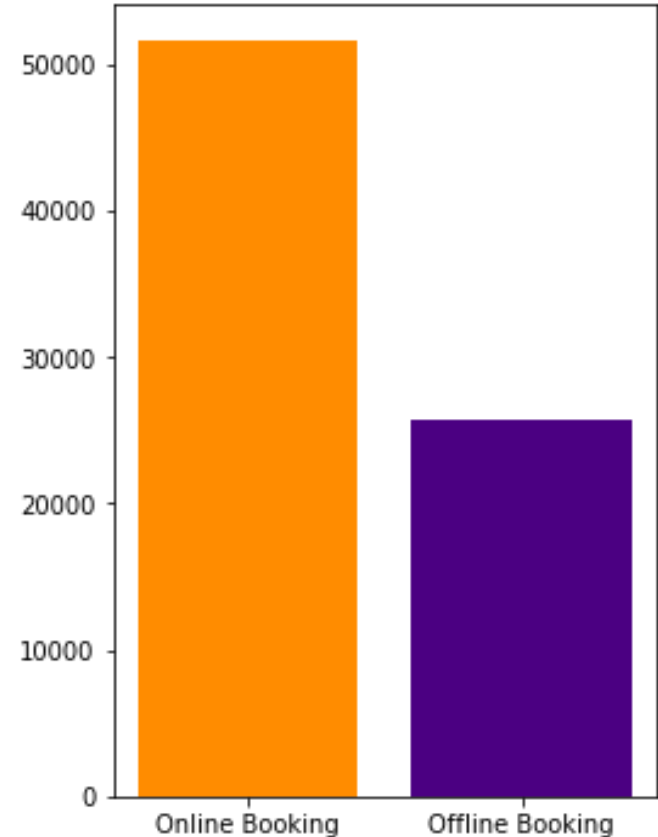
8. Checking booking patterns

Output-

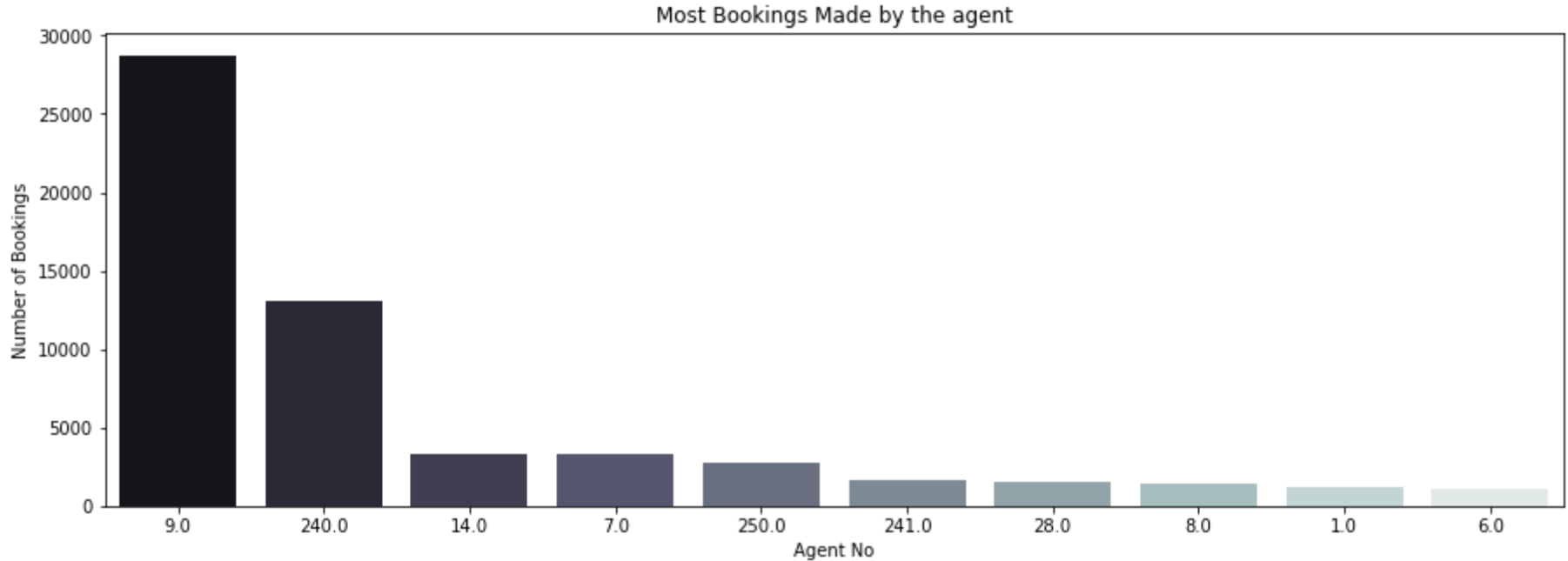
Online TA	51550
Offline TA/TO	13855
Direct	11780
Groups	4922
Corporate	4196
Complementary	692
Aviation	226
Undefined	2

Conclusion-

So, Majority of the customers book through online rather than offline booking



9. Which Agent made the most bookings?



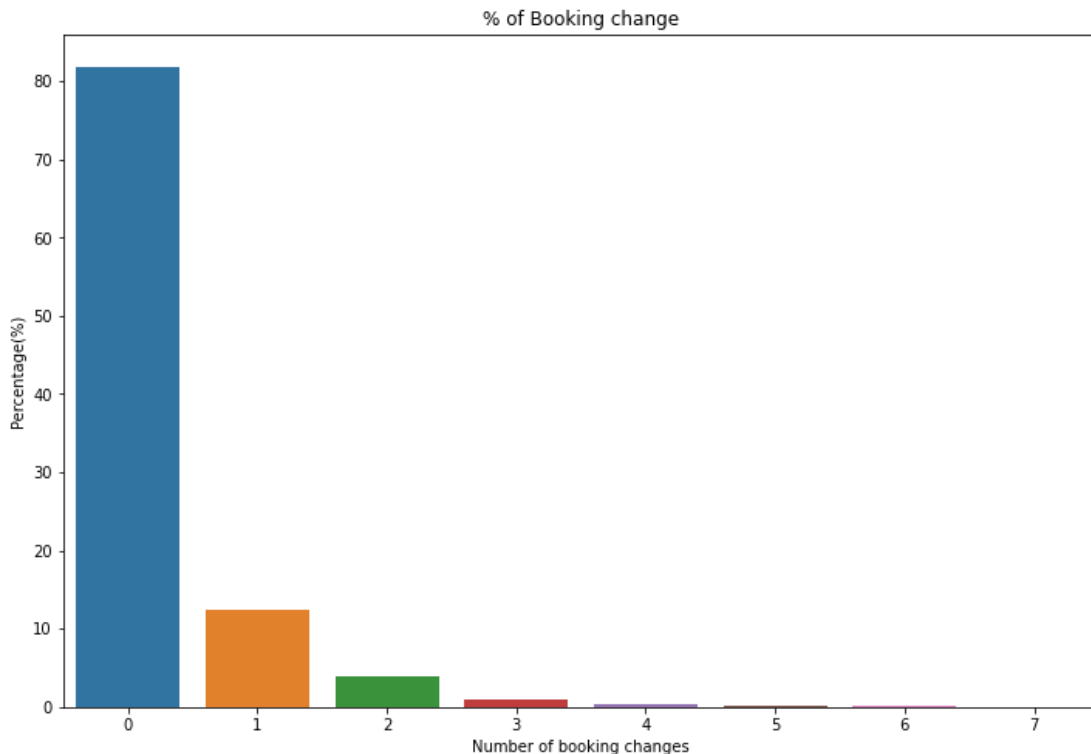
Conclusion - Agent ID no: 9 made most of the bookings

10. What is the percentage of booking changes made by the customer

Output-

	number_booking_changes	Counts
0	0	71418
1	1	10868
2	2	3483
3	3	862
4	4	347
5	5	113
6	6	57
7	7	29

Conclusion- Almost 82% of the bookings were not changed by guests.



11. Which year had the highest bookings

Output -

hotel	arrival_date_year	count
City Hotel	2016	26683
	2017	20063
	2015	6526
Resort Hotel	2016	15627
	2017	11568
	2015	6756

Conclusion- Most bookings were made in 2016 for both types of hotels.

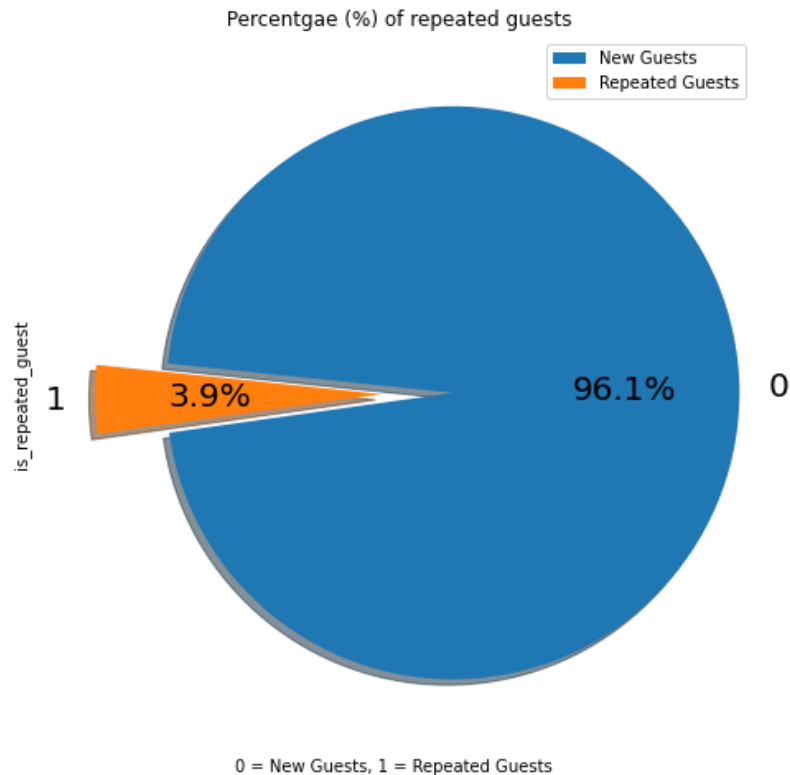


12. Checking the Percentage of repeated guests

Output-

New Guests - 33859
Repeated Guests - 3364

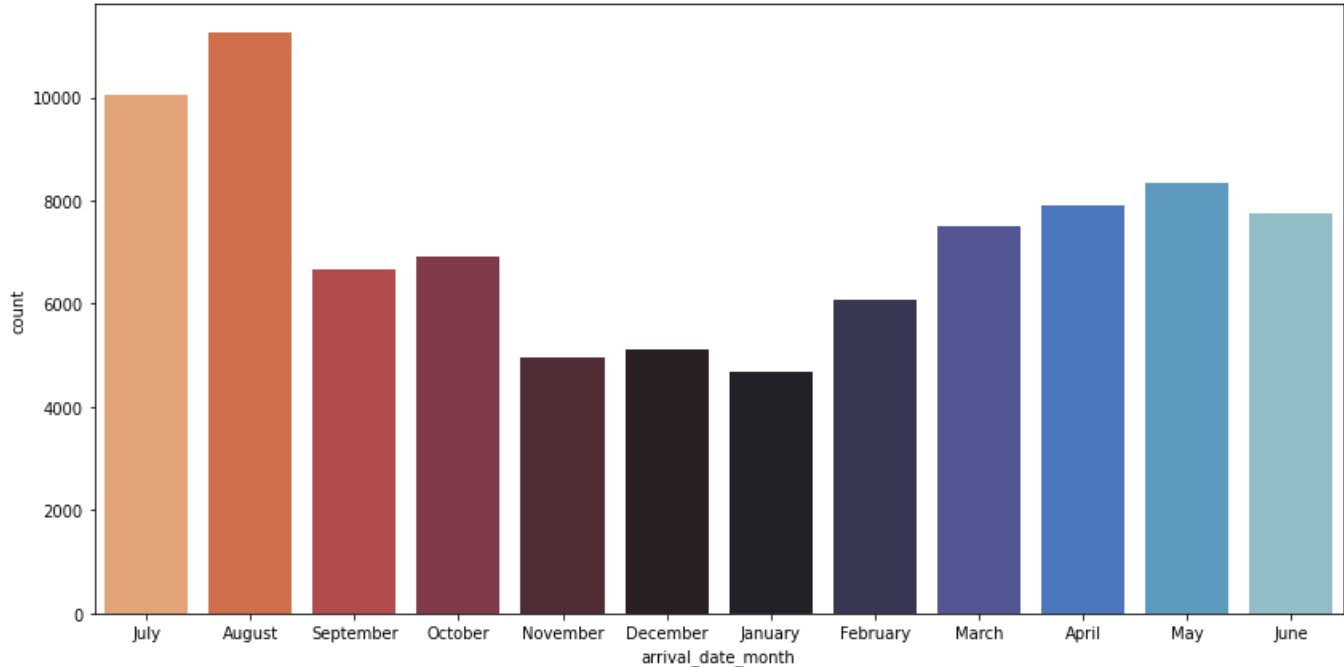
Conclusion – Only 3.9% of the guests are repeated guests.



13. Finding out the busiest month for bookings

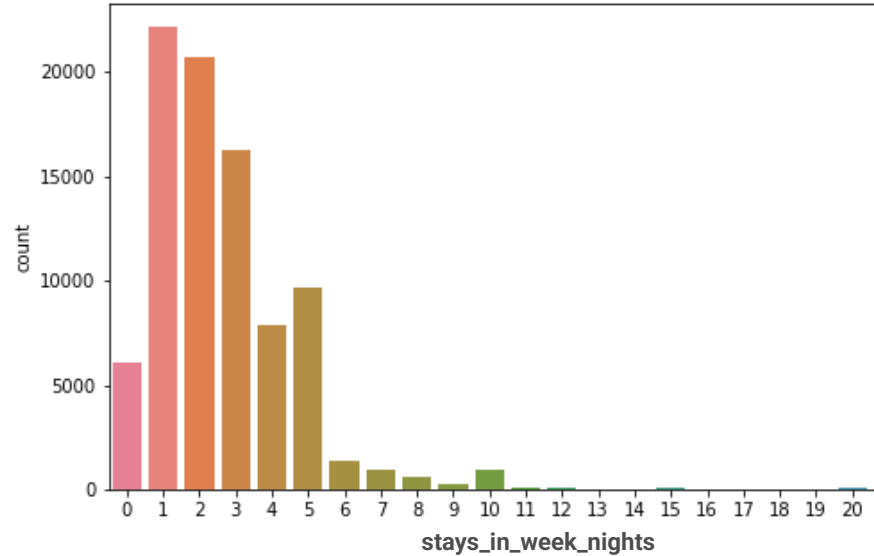
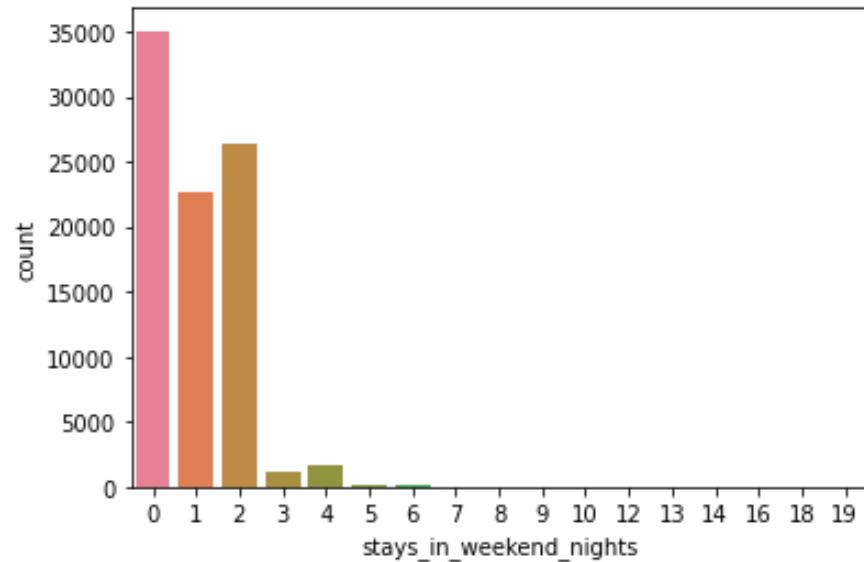
Output- Number of bookings made in each month in descending order

August	11242
July	10042
May	8343
April	7900
June	7756
March	7489
October	6921
September	6681
February	6082
December	5111
November	4971
January	4685



Conclusion- August is busiest month as it has most number of bookings

14. Averages of staying time



Conclusion - People generally prefer long stays on weekdays rather than weekends.

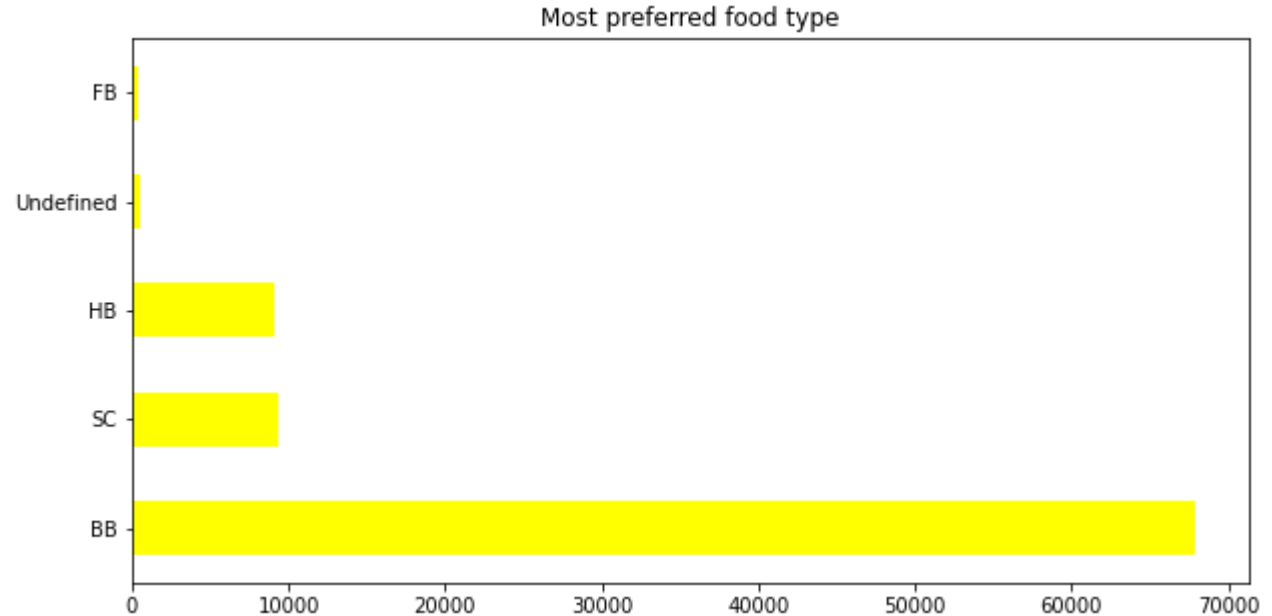
15. Most preferred food types

Output-

BB	67900
SC	9391
HB	9080
Undefined	492
FB	360

Conclusion –

Most people like to eat
BB type of food

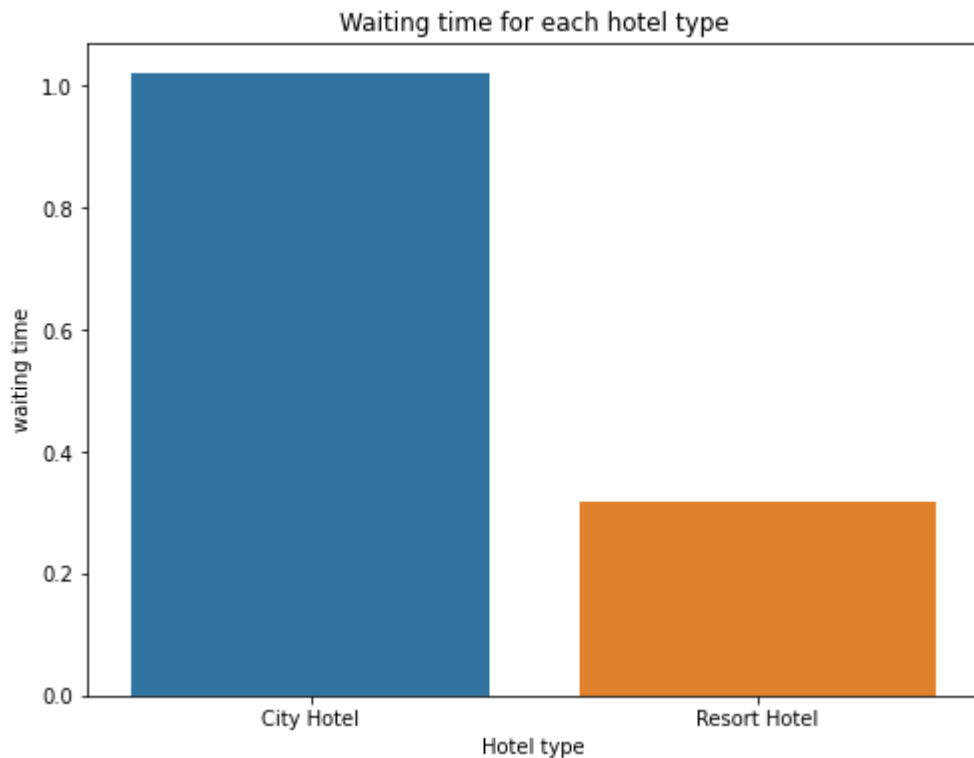


16. Waiting time for each type hotel

Output-

	hotel	days_in_waiting_list
0	City Hotel	1.020104
1	Resort Hotel	0.316810

Conclusion- City hotels have more waiting time than Resort hotels



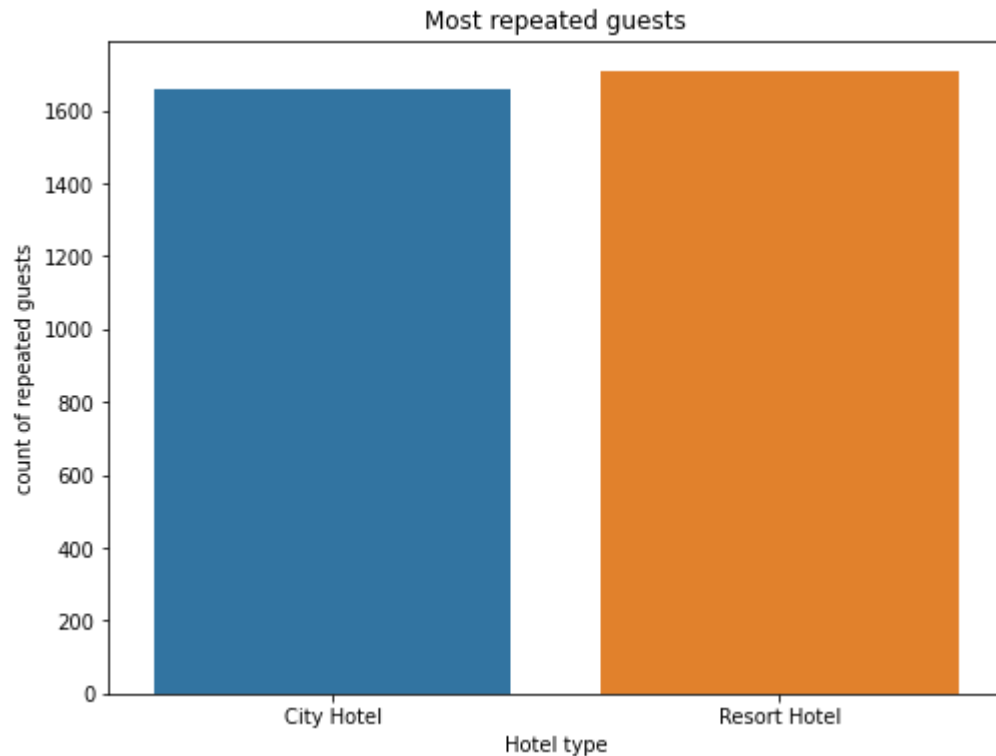
17. Hotels with the most repeated guests

Output –

	hotel	number_of_repated_guests
0	City Hotel	1657
1	Resort Hotel	1707

Conclusion –

It can be seen that there is not much difference in repeated guests in both type of the hotels still Resort hotels have more repeated guests.

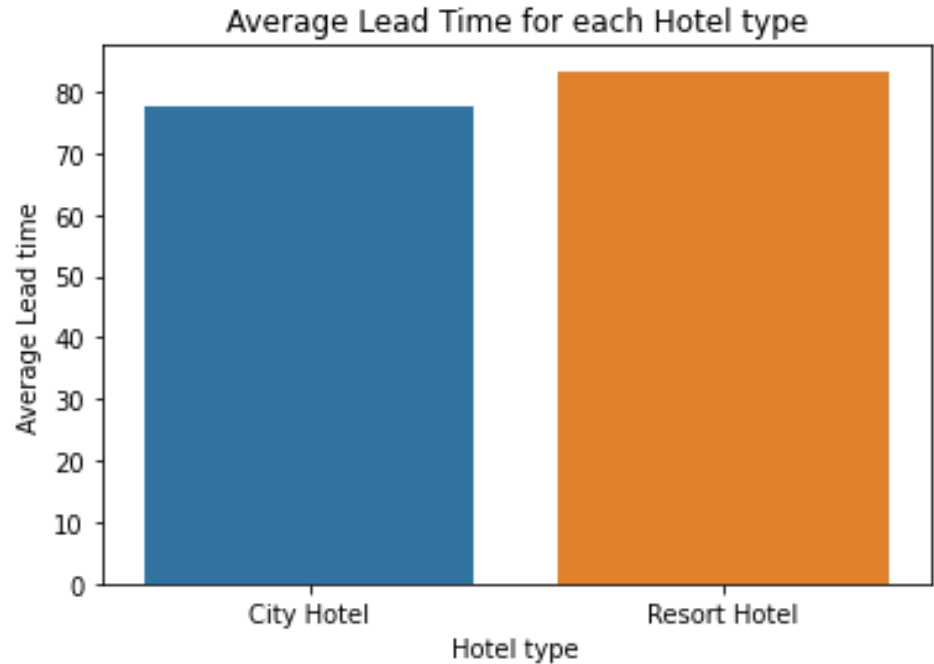


18. Finding hotels with more lead time

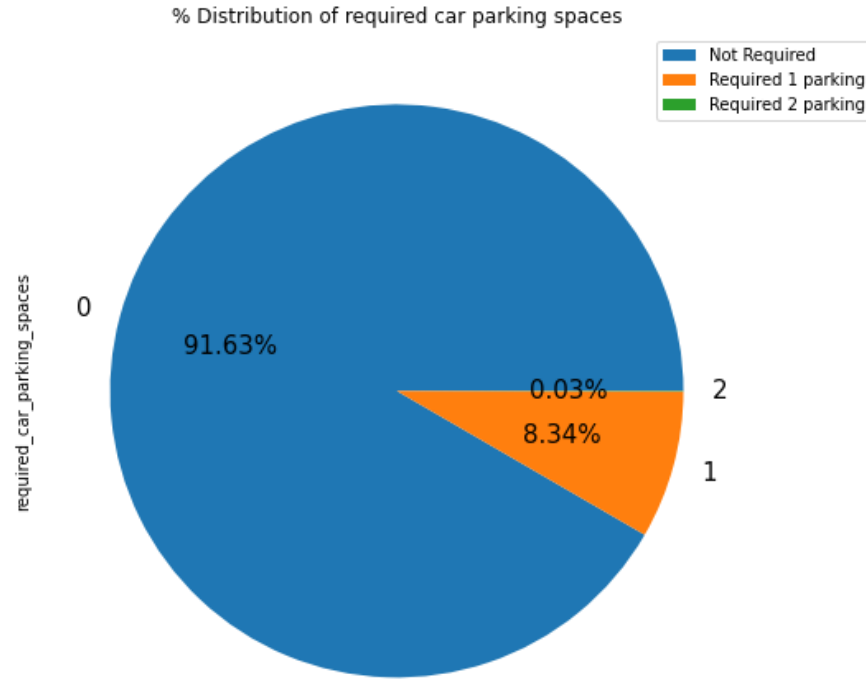
Output –

	hotel	lead_time
0	City Hotel	77.795465
1	Resort Hotel	83.399664

Conclusion – Resort hotels have more lead time than city hotels



19. Checking required car parking spaces



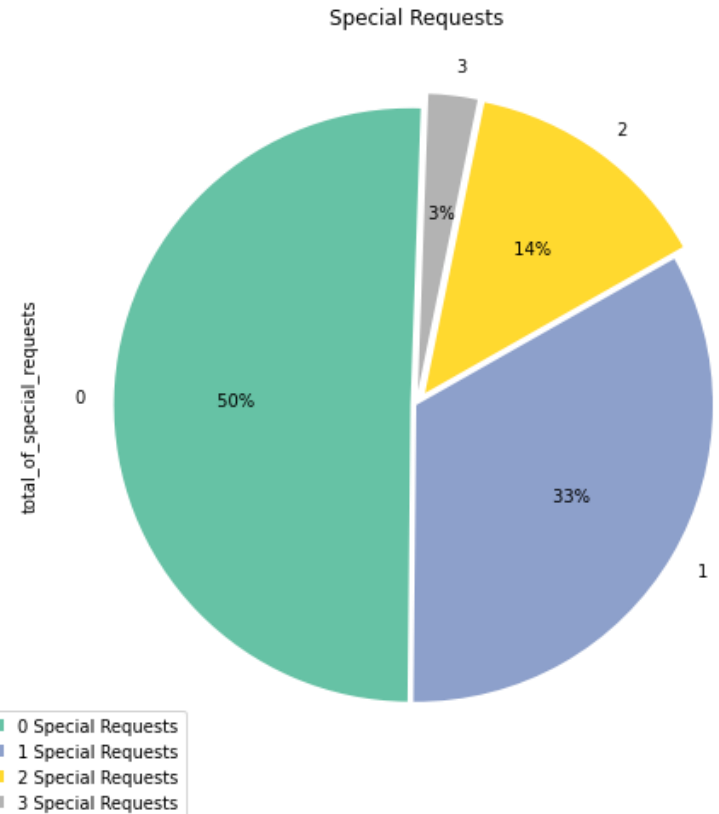
Conclusion- 91.6 % guests did not required the parking space. only 8.3 % guests required only 1 parking space.

20. Finding total number of special requests made by customers

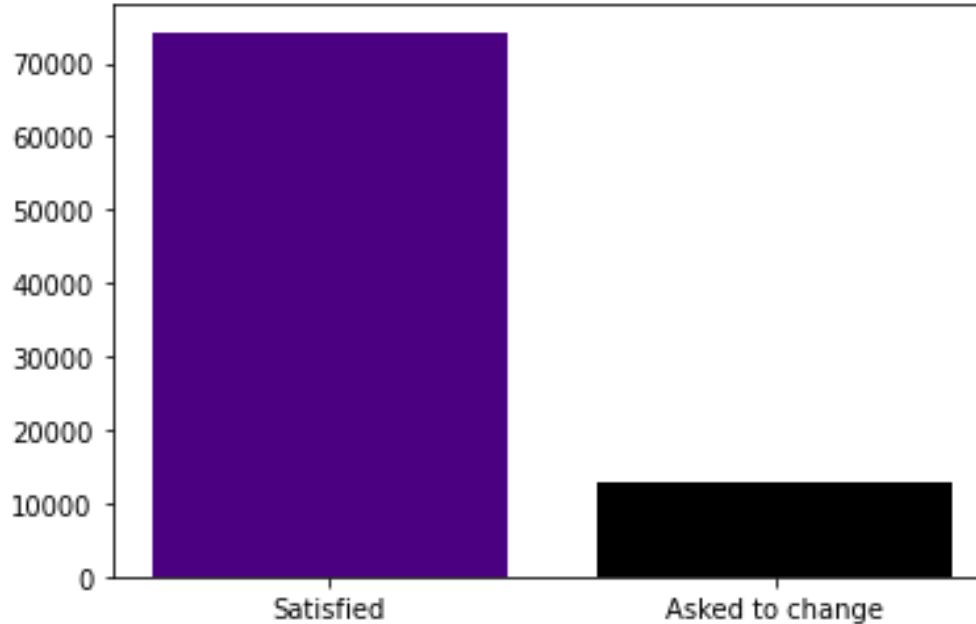
Output — Rounded to integer value.

0	50.0
1	33.0
2	14.0
3	3.0
4	0.0
5	0.0

Conclusion- 50% of the customers did not make any special request.



21. Checking customer satisfaction with rooms



Observation - Most of the guests were satisfied with their assigned rooms.

Conclusion

- 1) People prefer to book City Hotels more as compared to Resort Hotels.
- 2) Almost 28% of pre-booked rooms were cancelled.
- 3) About 82% of guests were Transient i.e. Staying for a short period of time.
- 4) Most travellers were from country – Portugal.
- 5) Bookings done via online channel were more than offline and other channels.
- 6) Busiest year was 2016 across all type of hotels and years.
- 7) August month had most number of bookings.
- 8) Average stay length was one night across all hotels.
- 9) Almost 50% of people did not make any special request.
- 10) Approximately 85% of people were satisfied with their rooms.

Thank you!!