

# CAPSTONE PROJECT

## Hotel Bookings Analysis

Rajat Sharma

# Problem Statements

For this project we will analyzing hotel booking data.

- This data contains booking information for a city hotel and resort hotel and includes information such as when the booking was made length of the stay the number of adults children's and or babies and the number of available parking spaces.
- Hotel industry is a very volatile industry and the bookings depend on our factors and many more the main objective behind this project the bookings and give insights to hotel management on which they can perform various campaigns to boost their performance of business.



# Workflows

We have divided the data into three parts

## ❖ Data collection :-

Dataset has been provided by the Alma Better <https://grow.almabetter.com/data-science/projects/Hotel-Booking-Analysis>

## ❖ Data loading:-

Dataset was loaded on google drive was loaded on goggle colab ('content/Hotel Bookings raw.csv')

## ❖ Data Cleaning and manipulation:- For data cleaning numpy, pandas, and matplotlib is used

❖ Removing the duplicates:- Analysing the data we found out of 119390 rows 31994 rows are duplicates so we removed them.

❖ Removing the company column with most nan values:- Analysing the company column we found it has only 5259 values are not nan remaining does not provide any significant data . So we deleted the column.

## ❖ Beginning with EDA :-



# Data Description:



**hotel** : Hotel(Resort Hotel or City Hotel)

**is\_canceled** : Value indicating if the booking was canceled (1) or not (0)

**lead\_time**: Number of days that elapsed between the entering date of the booking into the PMS and the arrival

**datearrival\_date\_year** : Year of arrival

**datearrival\_date\_month** : Month of arrival

**datearrival\_date\_week\_number** : Week number of year for arrival

**datearrival\_date\_day\_of\_month** : Day of arrival

**datestays\_in\_weekend\_nights** : Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the

**hotelstays\_in\_week\_nights** : Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel

**adults**: Number of adults

**children**: Number of children

**babies**: Number of babies

**meal**: Type of meal booked. Categories are presented in standard hospitality meal packages:

**country**: Country of origin.

**market\_segment** : Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”

**distribution\_channel** : Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”

**is\_repeated\_guest** : Value indicating if the booking name 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

**bookingreserved\_room\_type** : Code of room type reserved. Code is presented instead of designation for anonymity reasons.

**Assigned\_room\_type** : Code for the type of room assigned to the booking.

**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** : Indication on if the customer made a deposit to guarantee the booking.

**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.

**days\_in\_waiting\_list** : Number of days the booking was in the waiting list before it was confirmed to the

**customercustomer\_type** : Type of booking, 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 (e.g. twin bed or high floor)

**reservation\_status** : Reservation last status, assuming one of three categories

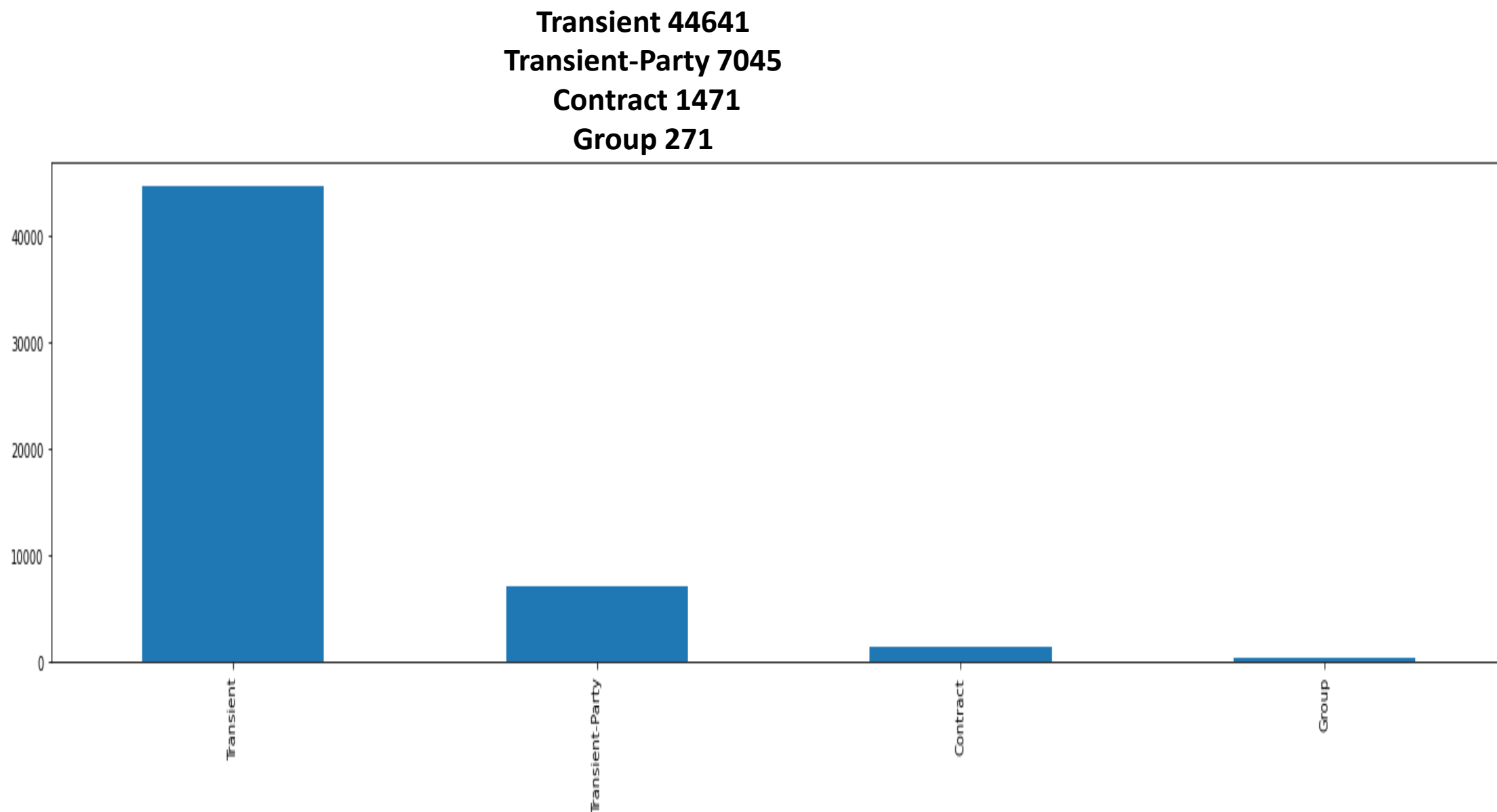
**Canceled** – booking was cancelled by the customer  
**Check-Out** – customer has checked in but already departed  
**No-Show** – customer did not check-in and did inform the hotel of the reason why

**reservation\_status\_date** : Date at which the last status was set. This variable can be used in conjunction with the Reservation Status to understand when was the booking cancelled or when did the customer checked-out of the hotel

## **For EDA analysis we were solving these problem statements**

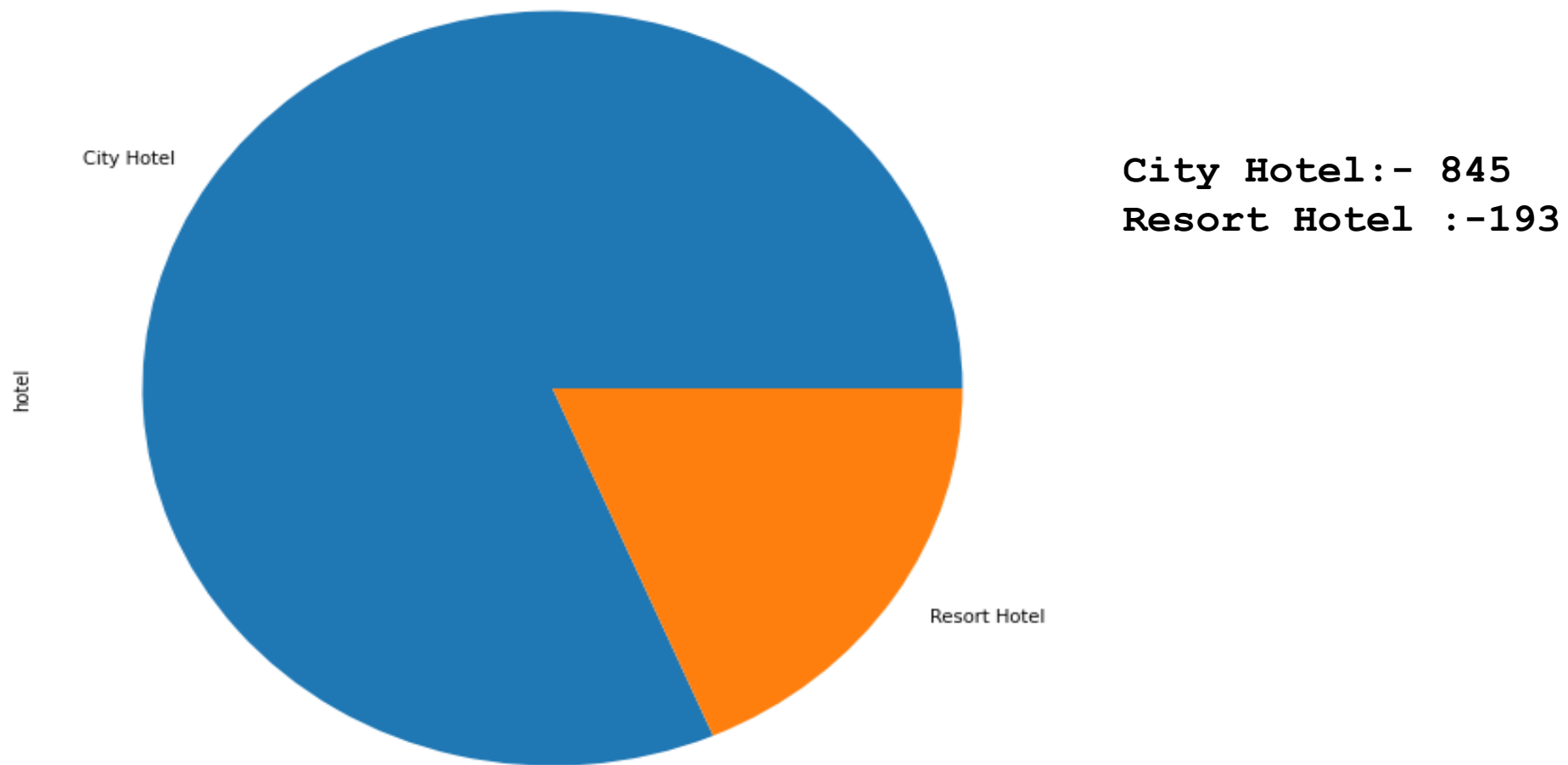


- Question 1. Which customer type preferred the most City hotel?**
- Question 2. Which type of hotel non refunded the most?**
- Question 3. Which distribution channel is mostly used for booking Resort hotel?**
- Question 4. Which type of hotel is preferred by customer having babies?**
- Question 5. Which type of hotel is most preferred for week night?**
- Question 6. Which year has highest booking and cancellation based on market segment (Direct)?**
- Question 7. Which food is most preferred by customer based on top 5 country?**
- Question 8. Which Month is favorable for booking ?**
- Question 9. Which are the top five agents for Resort hotel?**
- Question 10. Hotel data based on month where repeated guest is 0:**



From the above graph shows that transient customer most preferred the city hotel.

## Question 2. Which type of hotel non refunded the most?



**This graph shows that city hotel has non refunded the most 845 compared to Resort Hotel 193**



### Question 3. Which distribution channel is mostly used for booking Resort hotel?

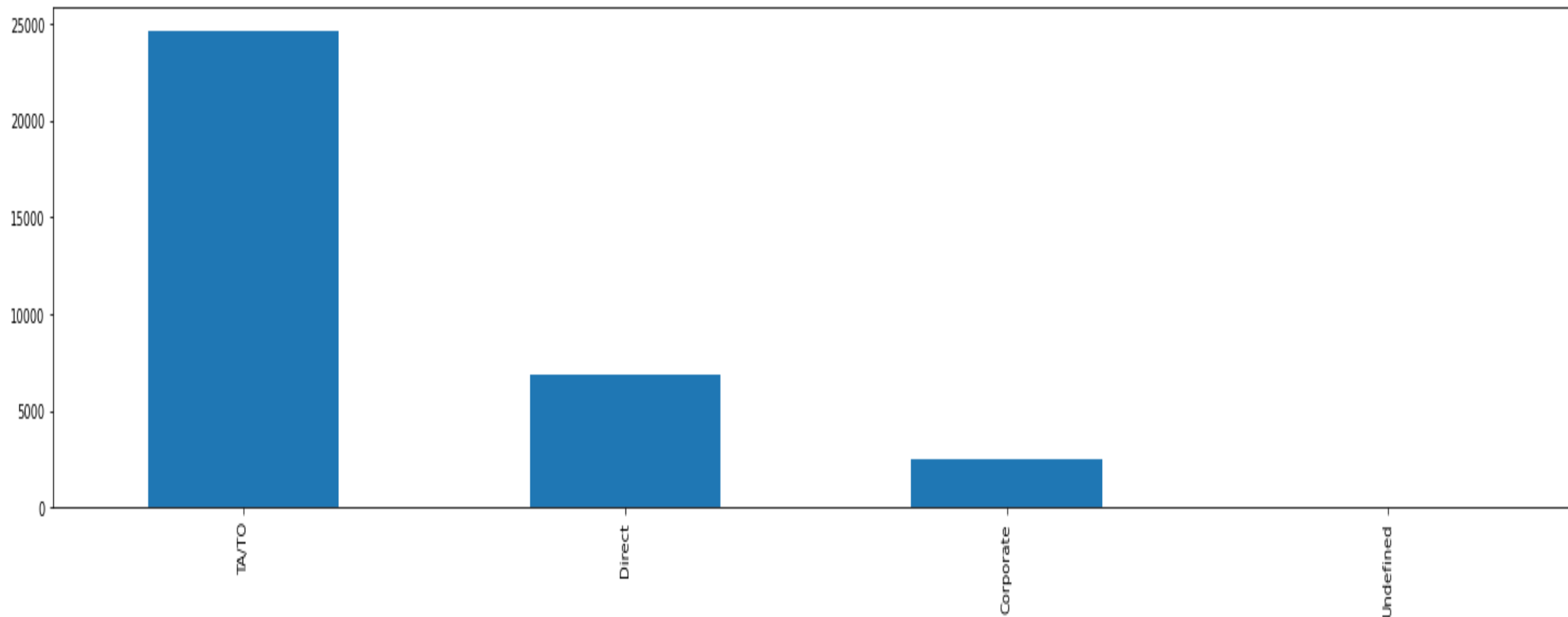


TA/TO 24592

Direct 6901

Corporate 2474

Undefined 1

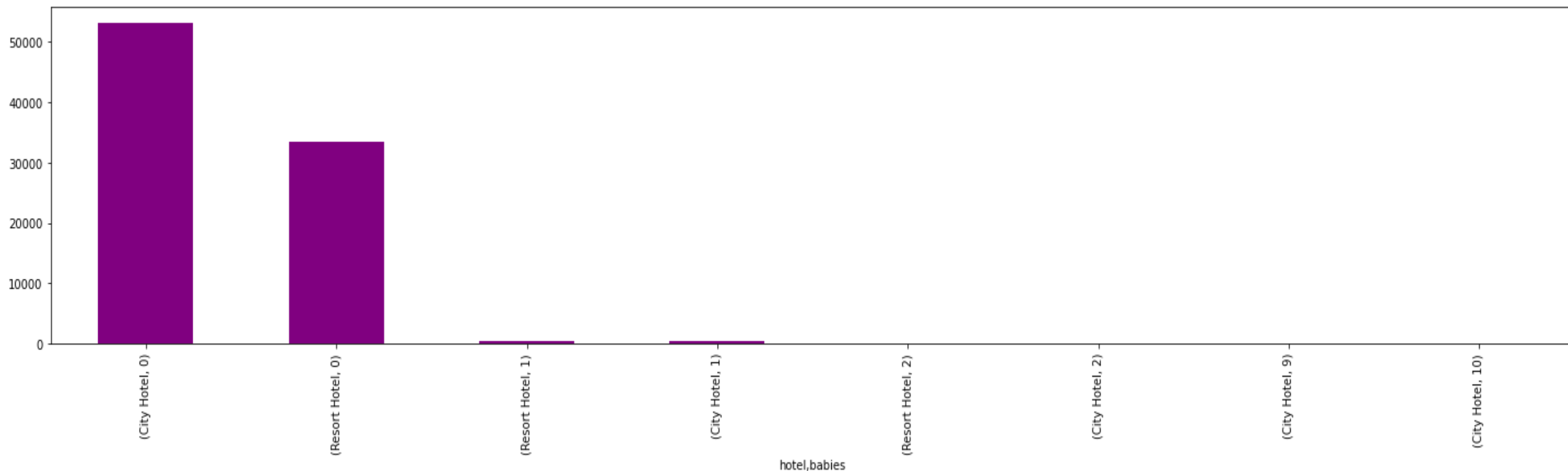


**This bar graph shows that TA/TO is most preferred for City Hotel 24592**

Question 4. Which type of hotel is preferred by customer having babies?

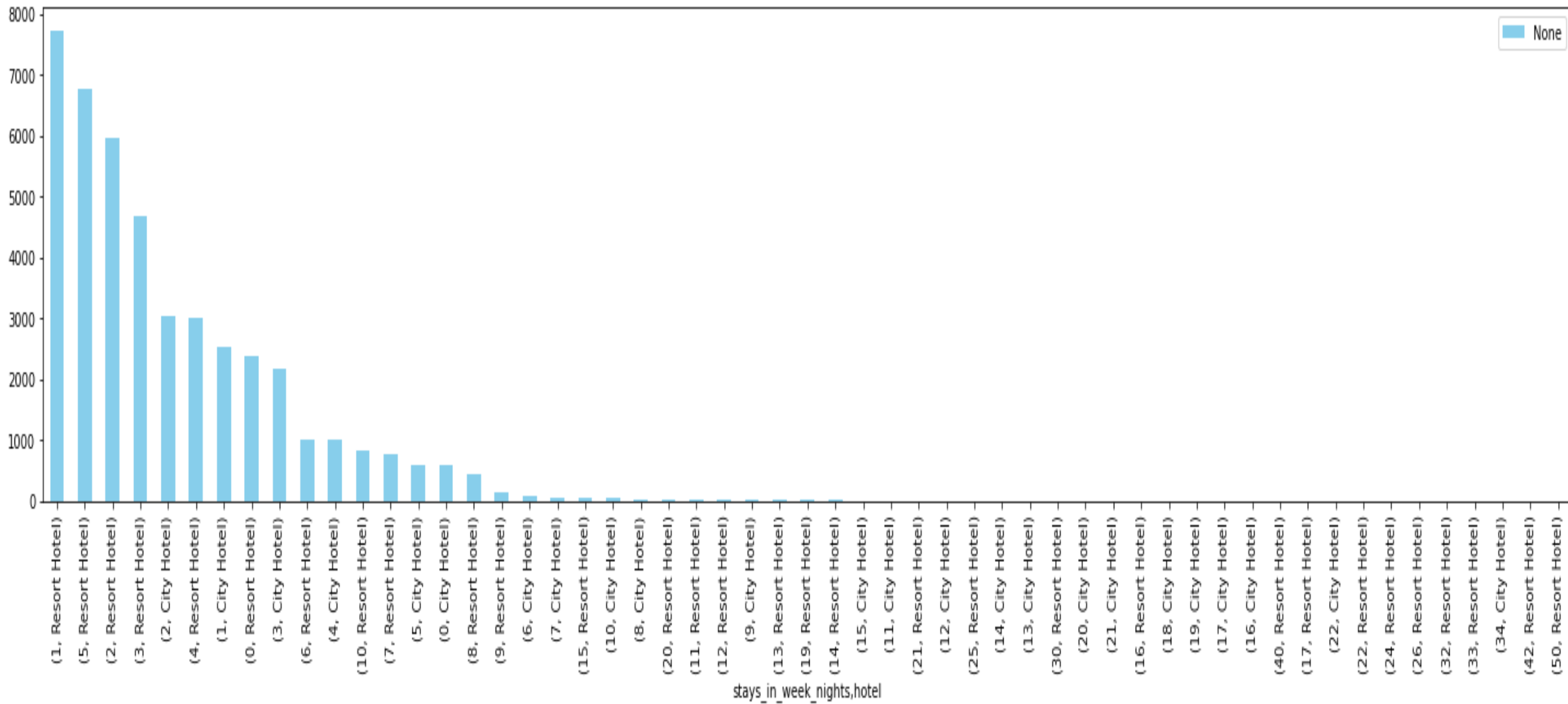


hotel	babies	
City Hotel	0	53059
Resort Hotel	0	33423
	1	536
City Hotel	1	361
Resort Hotel	2	9
City Hotel	2	6



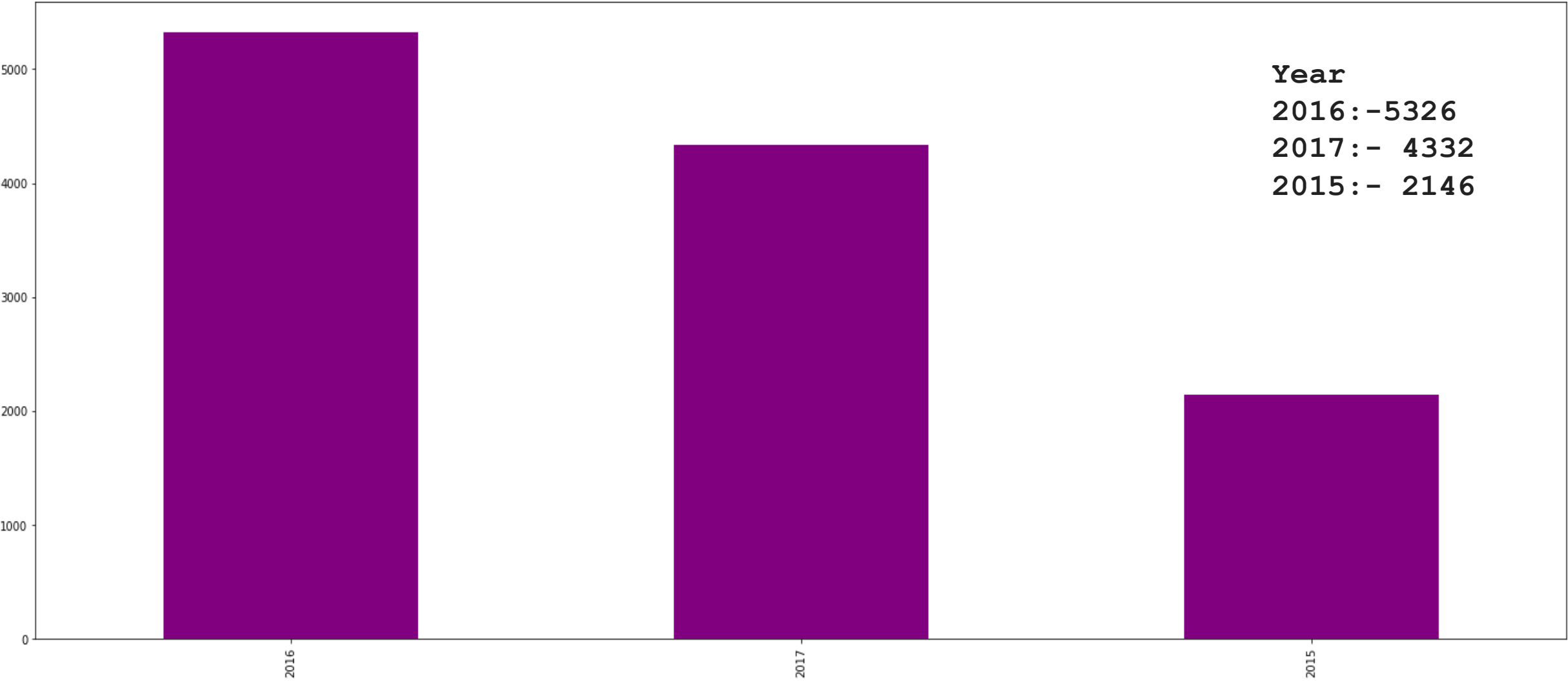
From the above graph shows that City hotel is most preferred by the customer having babies.

stays_in_week_nights	hotel	
1	Resort Hotel	7720
5	Resort Hotel	6773
2	Resort Hotel	5953
3	Resort Hotel	4690
2	City Hotel	3028
4	Resort Hotel	3010
1	City Hotel	2541
0	Resort Hotel	2378
3	City Hotel	2174
6	Resort Hotel	1020
4	City Hotel	1007
10	Resort Hotel	839
7	Resort Hotel	774
5	City Hotel	600
0	City Hotel	585
8	Resort Hotel	458
9	Resort Hotel	150
6	City Hotel	94
7	City Hotel	55
15	Resort Hotel	50
10	City Hotel	42
8	City Hotel	35
20	Resort Hotel	31
11	Resort Hotel	24
12	Resort Hotel	22
9	City Hotel	21



**This Graph shows that Resort hotel is most preferred for week nights.**

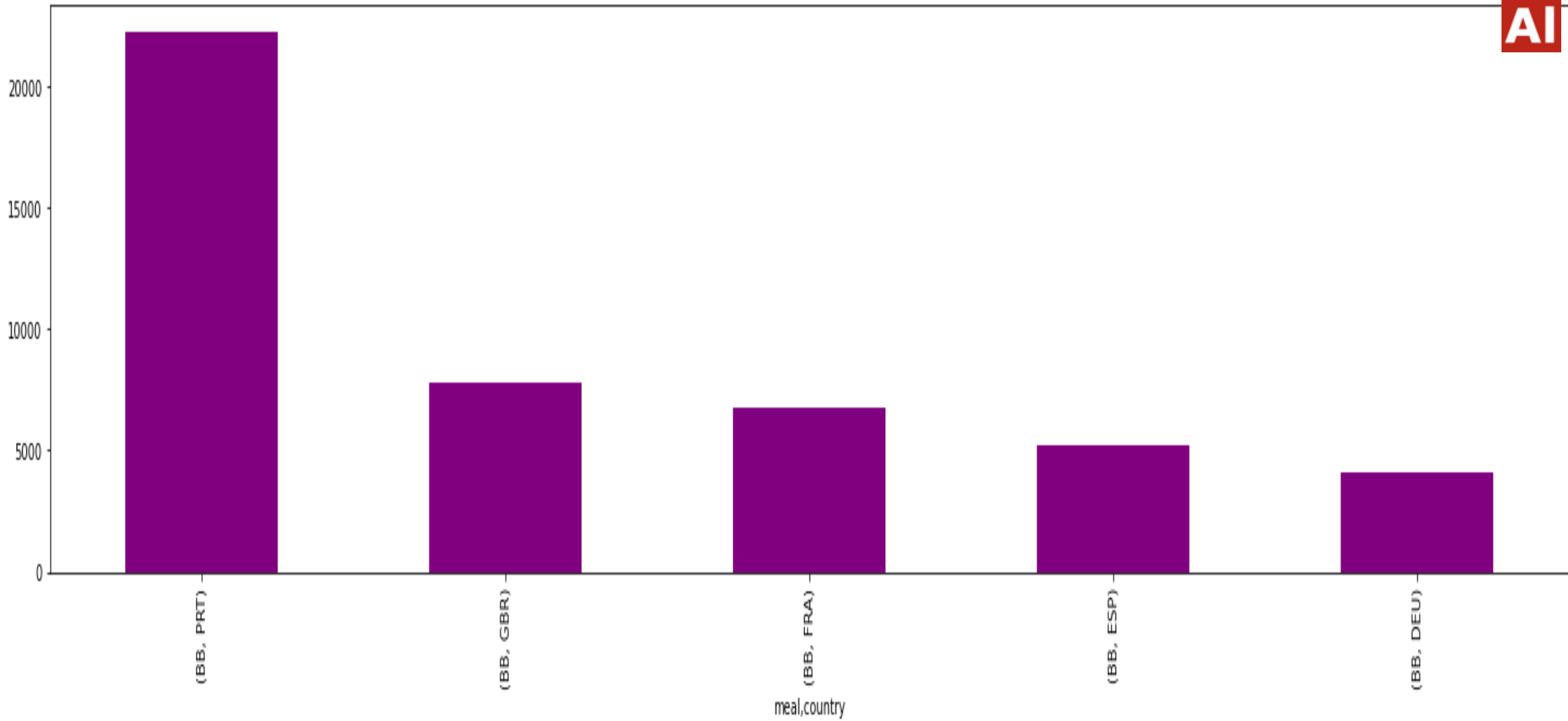
Question 6. Which year has highest booking and cancellation based on market segment (Direct)?



This Graph shows that 2016 has highest booking and cancellation based on market segment (Direct).

Question 7. preferred food by customer based on top 5 country?

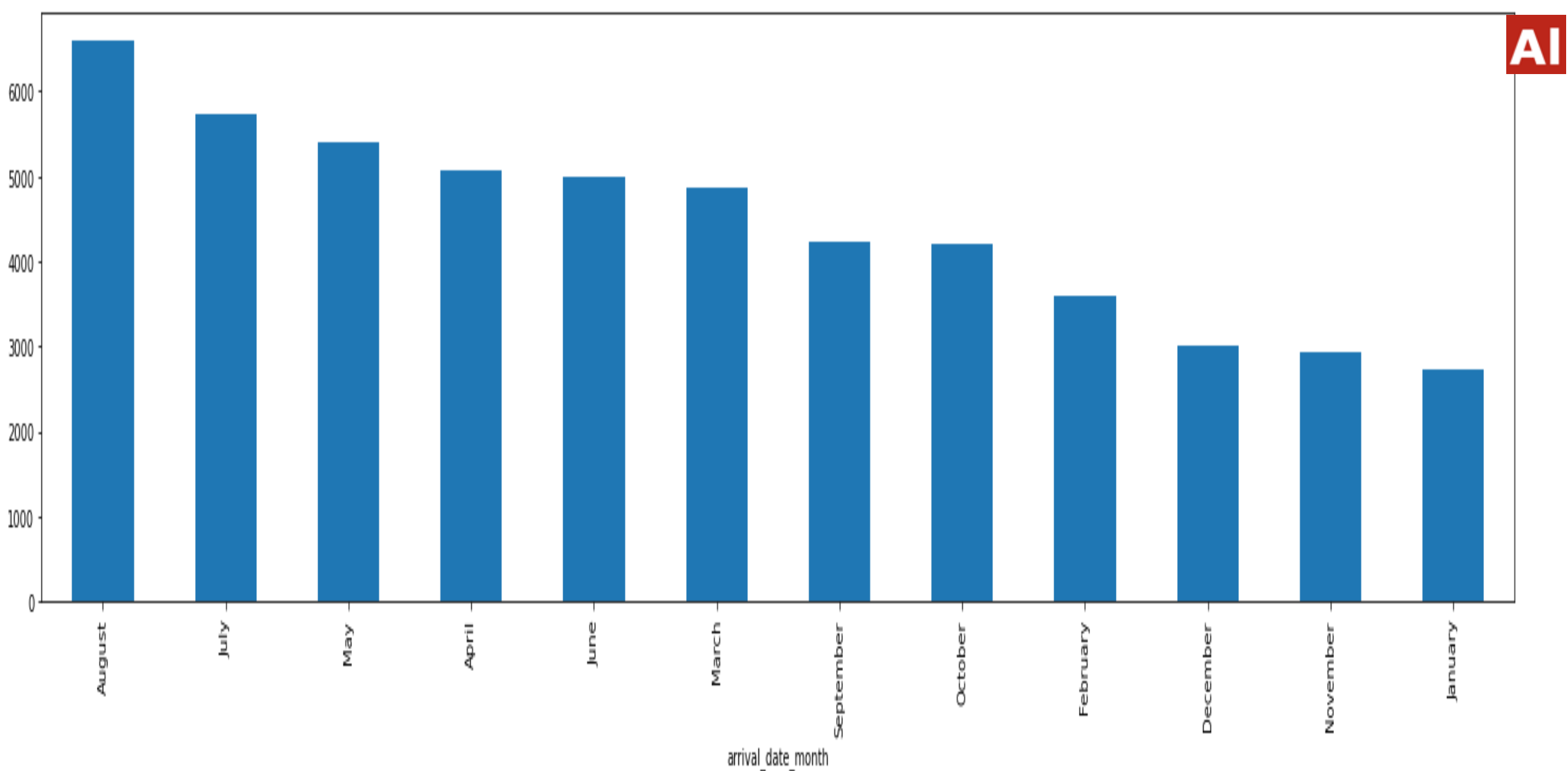
meal	country	
BB	PRT	22283
	GBR	7796
	FRA	6769
	ESP	5178
	DEU	4059



The graph shows that Portugal preferred the BB food in top 5 countries.

arrival_date_month	
August	6591
July	5744
May	5413
April	5080
June	5005
March	4856
September	4240
October	4208
February	3605
December	3014
November	2942
January	2730



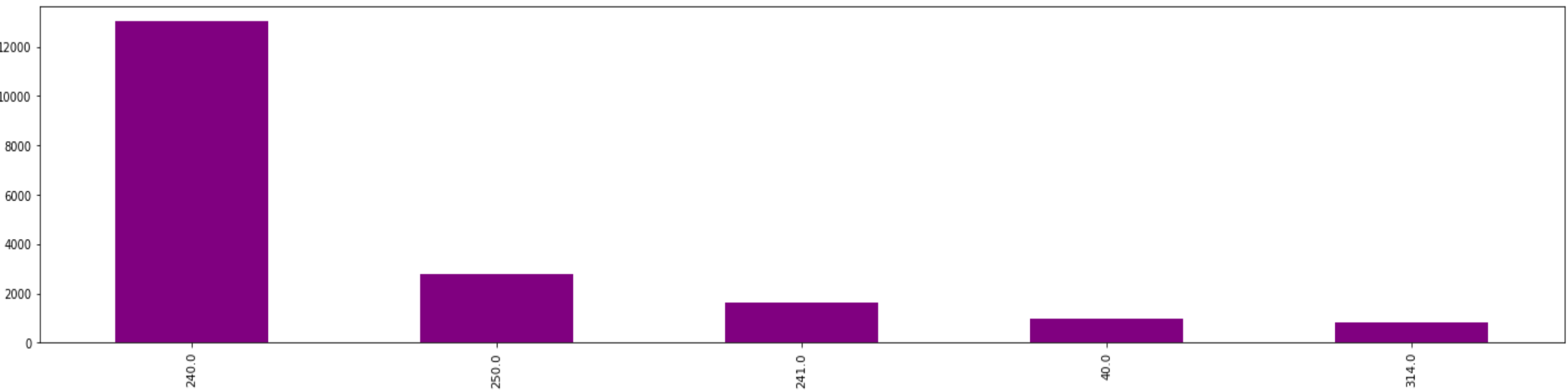


**This Graph shows that August month is most favorable month for booking.**

Question 9. Which are the top five agents for Resort hotel?



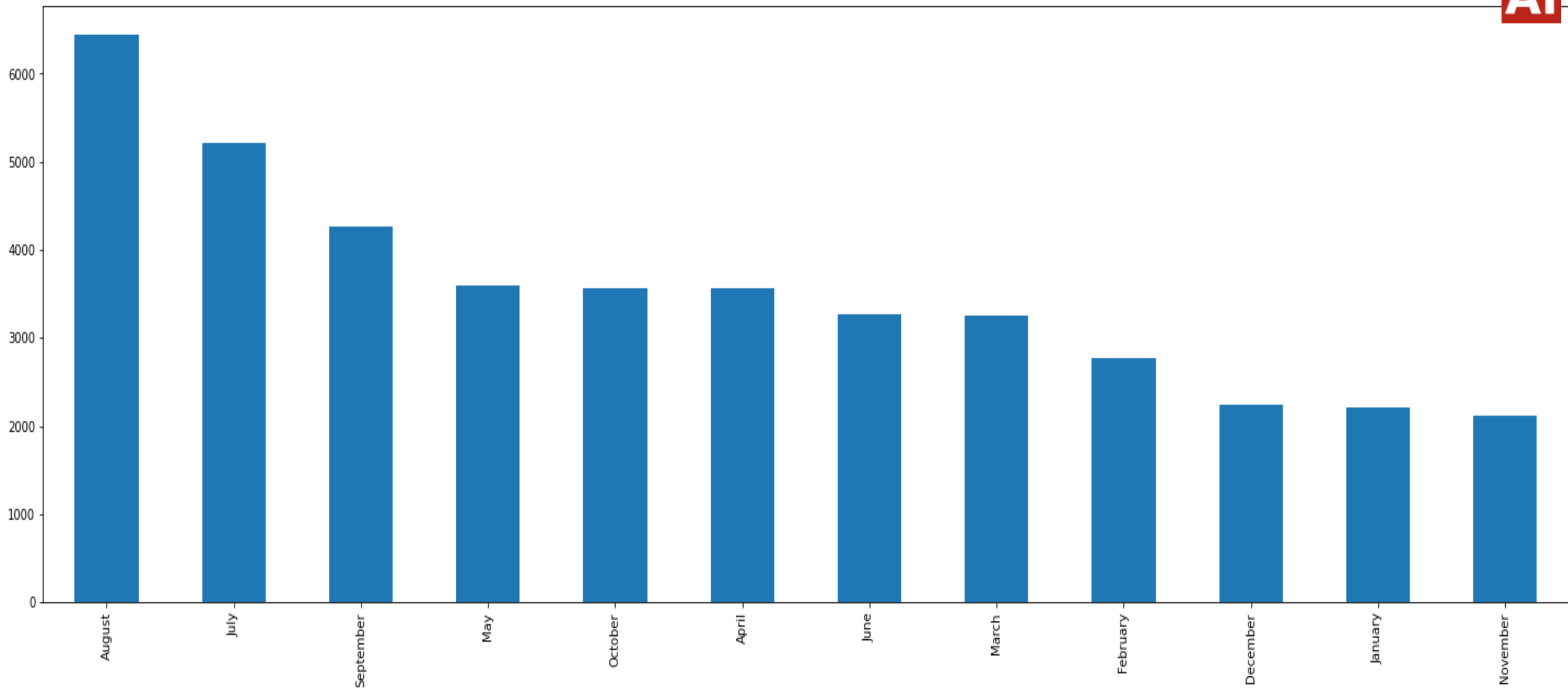
Agent	
240.0	13020
250.0	2778
241.0	1644
40.0	954
314.0	844



This Graph shows that 240.0 is the top agent for booking resort hotel.

Question 10. Hotel data based on month where repeated guest is 0:

August	6445
July	5219
September	4259
May	3595
October	3559
April	3559
June	3273
March	3259
February	2764
December	2237
January	2213
November	2114



**This graph shows that in the month of August no repeated guest.**

# Conclusion:-



1. Transient customer most preferred the city hotel.
2. City Hotel not refunded the most 845 in comparison of resort hotel 193.
3. TA/TO is most preferred distribution channel for booking for resort hotels.
4. City hotel is most preferred by customers having babies as compared to resort hotel.
5. City Hotel is most preferred for week night as compared to resort hotel.
6. in the year 2016 has highest booking and cancellation based on market segment (Direct).
7. Portugal preferred the BB food in top 5 countries.
8. August month is most favorable month for booking.
- 9- 240.0 is the top agent for booking resort hotel.
- 10- August month has no repeated guest with 0.

From the above conclusions we can say that City hotels are the most preferred hotel type by the guests. We can say City hotel is the busiest hotel.

Thank you