

# **CAPSTONE PROJECT**

## **HOTEL BOOKING ANALYSIS**



### **Team Members**

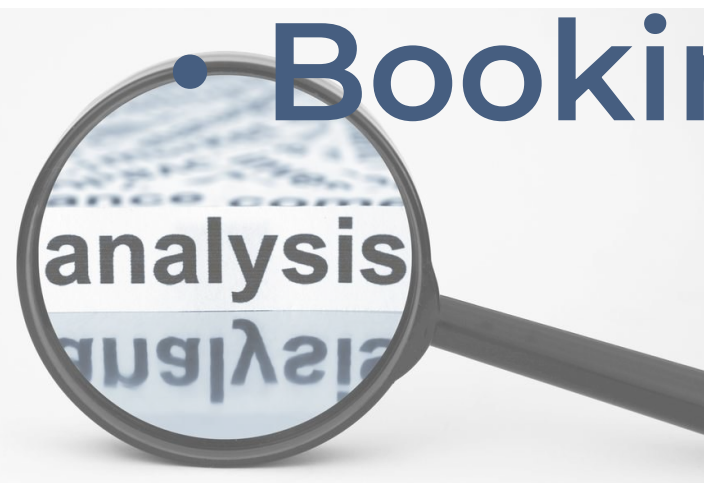
**Palash Pathak**

**Yogesh Dubey**

# **CONTENT**

## **Analysis Based On :**

- Time
- Hotels
- Visitors
- Country
- Bookings



- 1.Data Summary
- 2.Data Cleaning
- 3.Analysis Of Data
- 4.Challenges
- 5.Conclusion

# **DATASUMMARY**

**Data Set Name** : Hotel Booking Database including information of visitors bookings & cancellations around the world from 2015 to 2017

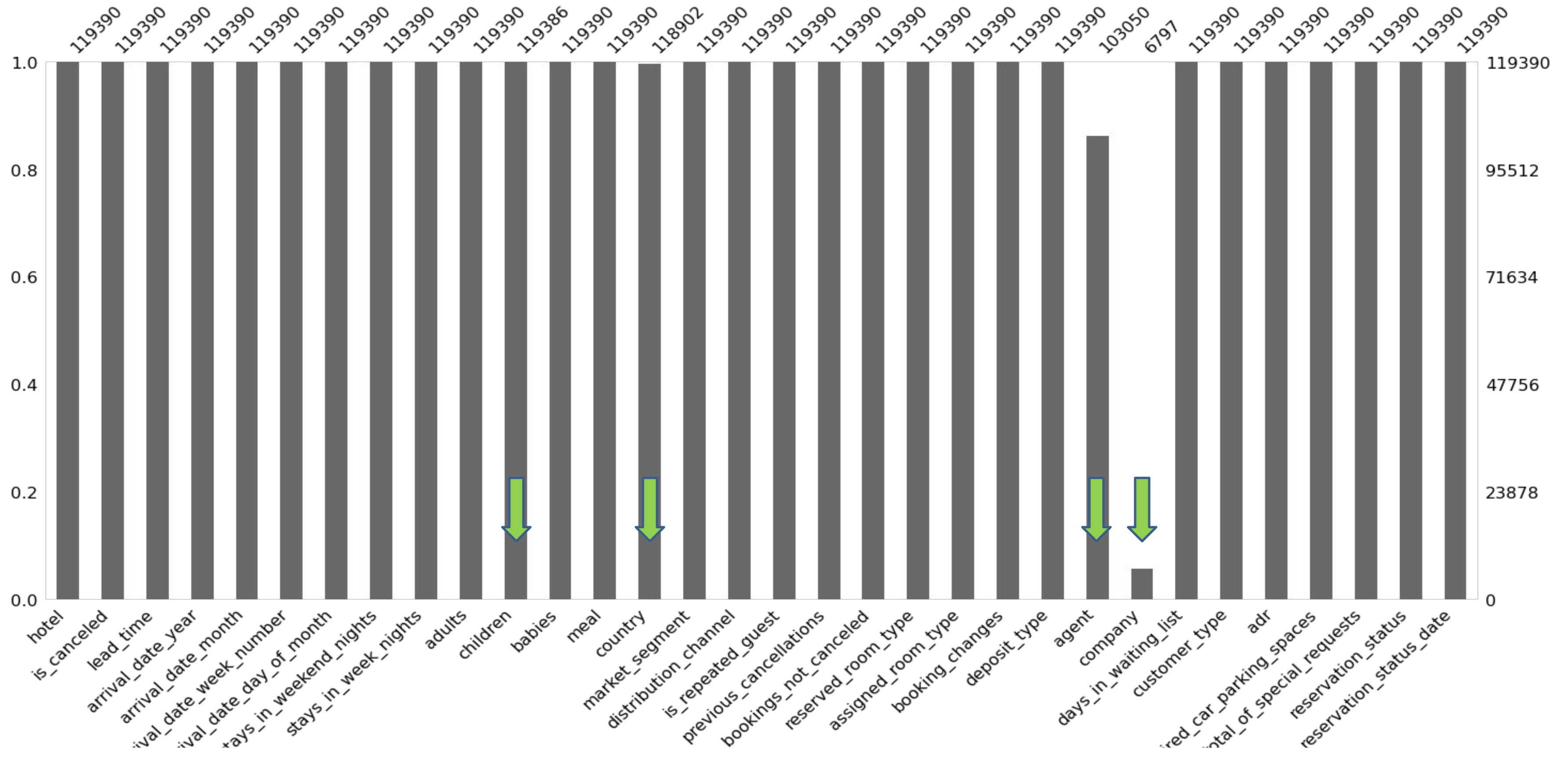
**Shape :**

- Rows - 119390
- Column - 32



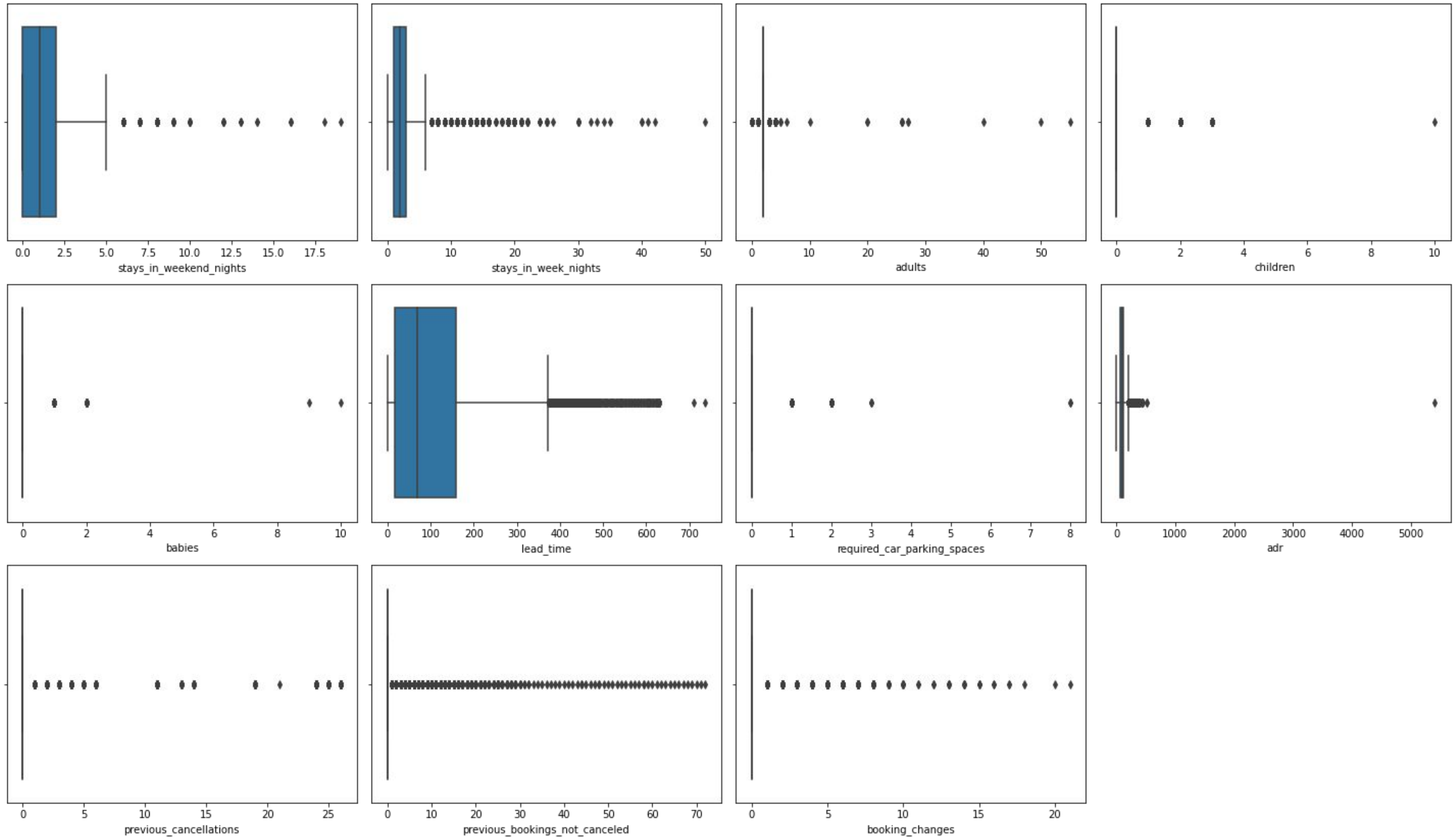
Important Columns : hotel, is\_canceled ,arrival\_date\_year, arrival\_date\_month,stays\_in\_weekend\_nights,stays\_in\_week\_nights,adults, children,babies,meal,country,market\_segment,distribution\_channel, customer\_type,adr,required\_car\_parking\_spaces,total\_of\_special\_requests

# WHERE ARE MISSING VALUES

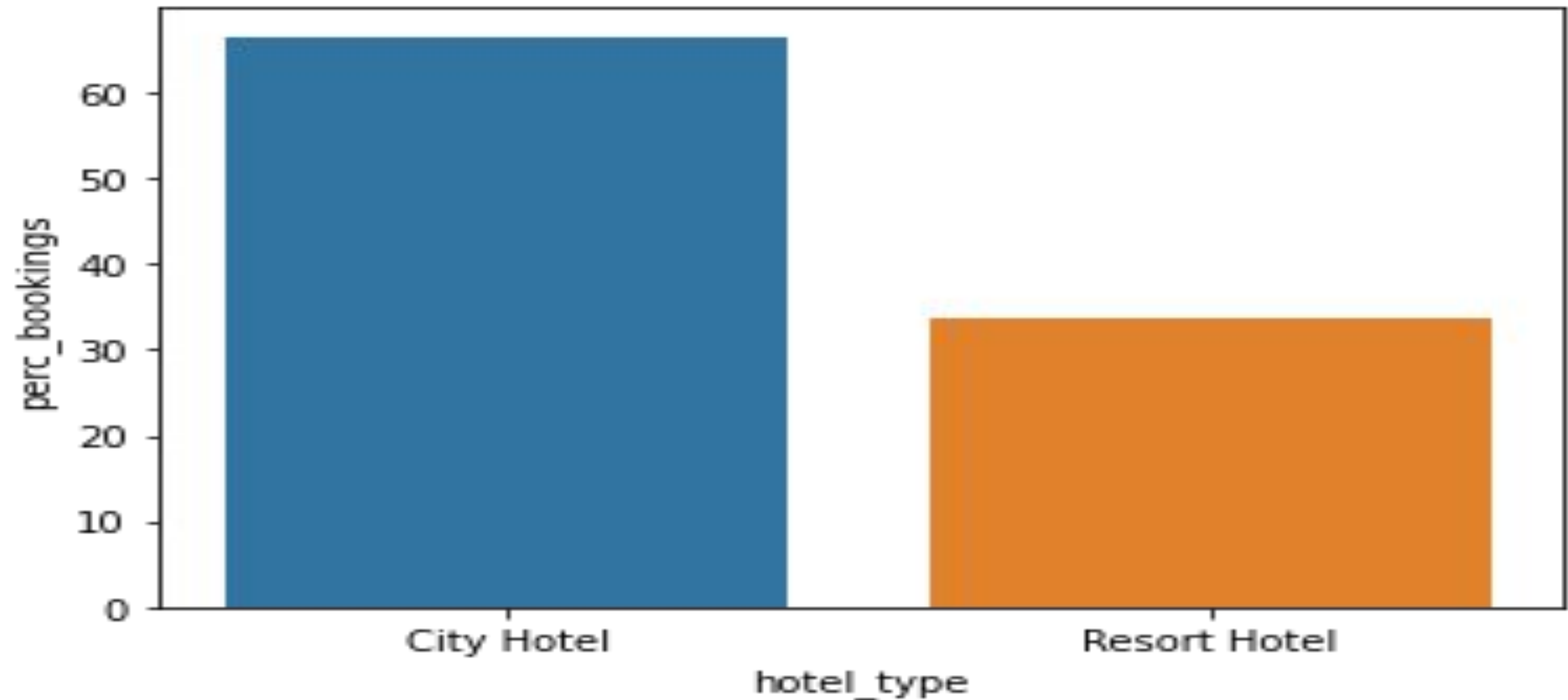




# Outliers

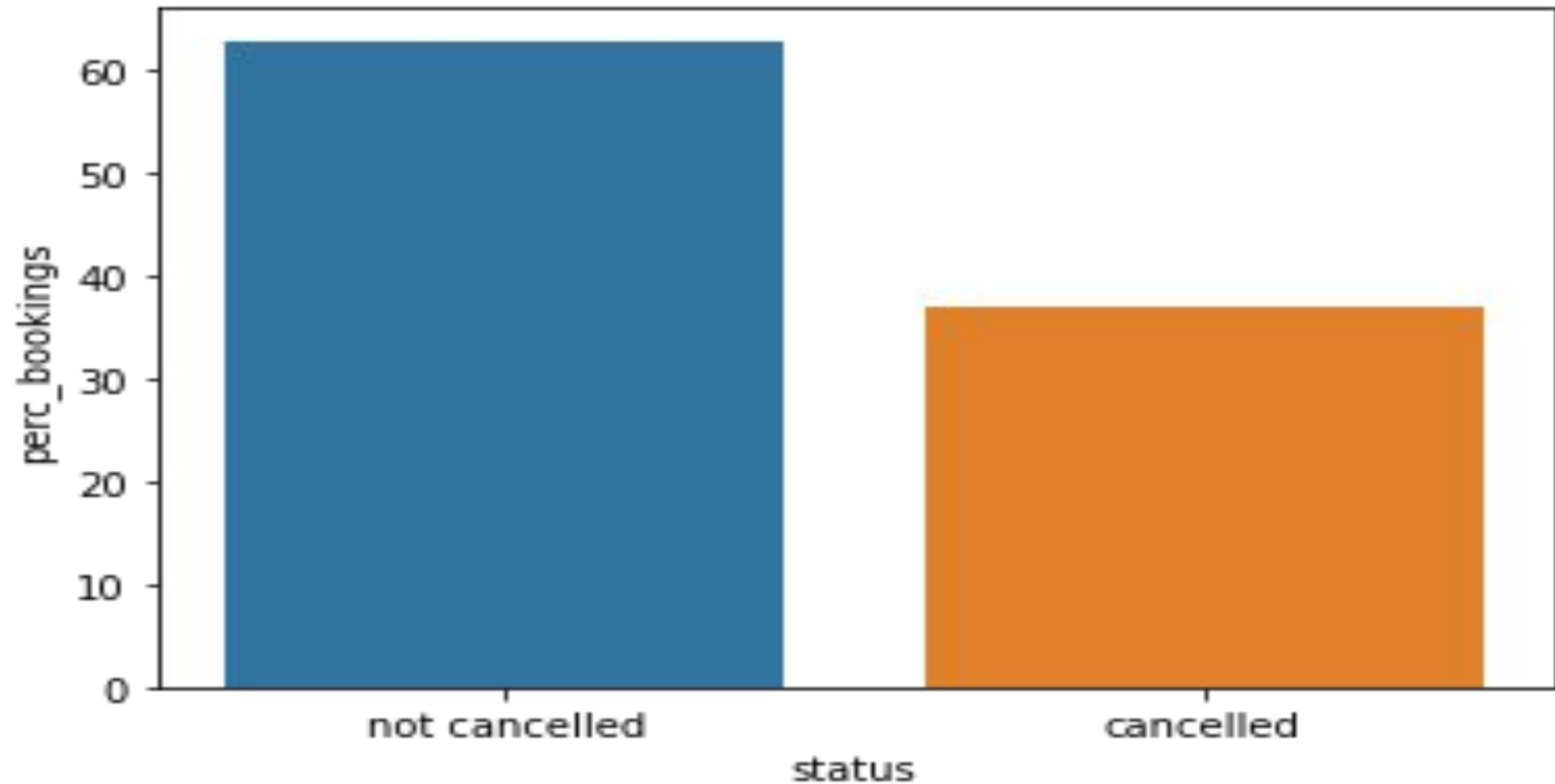


**As we have two categories of hotel, lets compare them.**



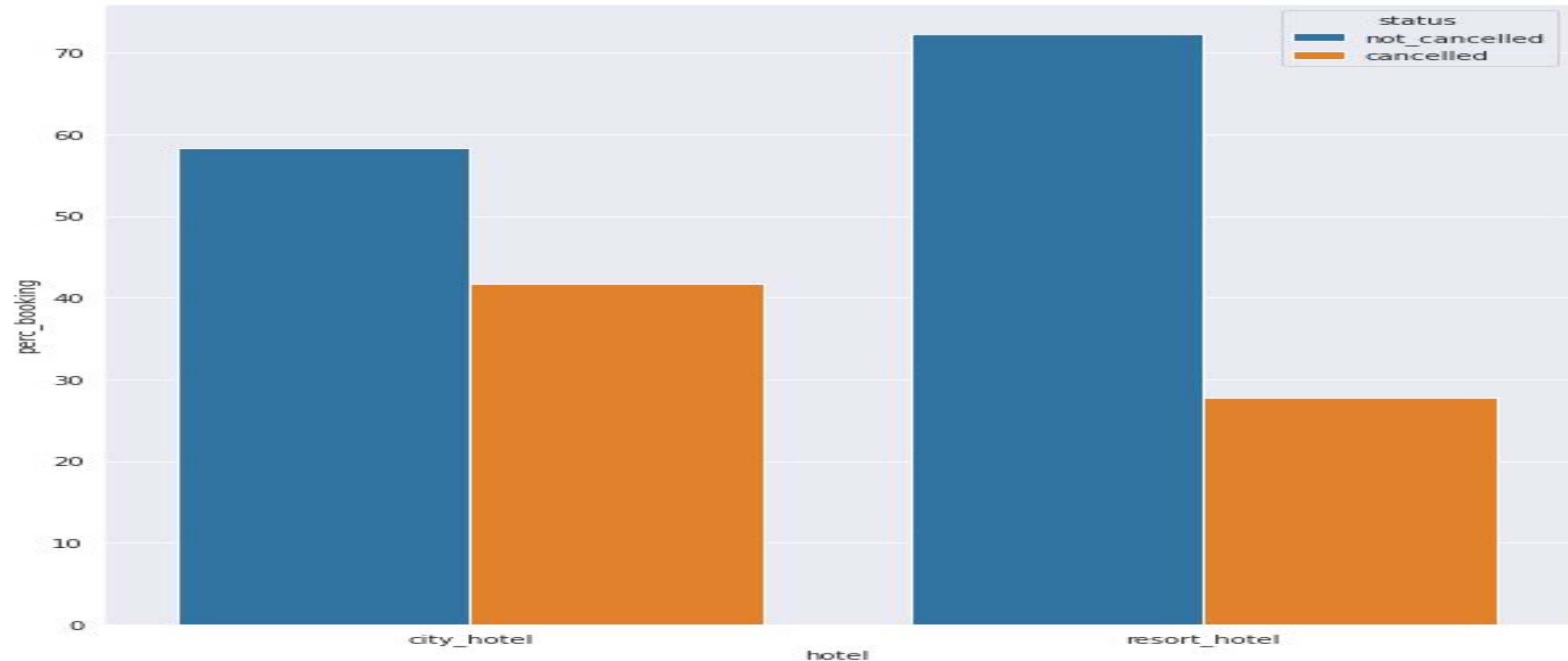
**SO, CITY HOTELS ARE BOOKED ALMOST TWICE THE NO OF TIMES  
THAT RESORT HOTELS ARE BOOKED**

# Total cancelled vs non-cancelled booking



We can comment that almost 37% bookings are cancelled by visitors.

# Cancelled vs non-cancelled booking for hotel and resort

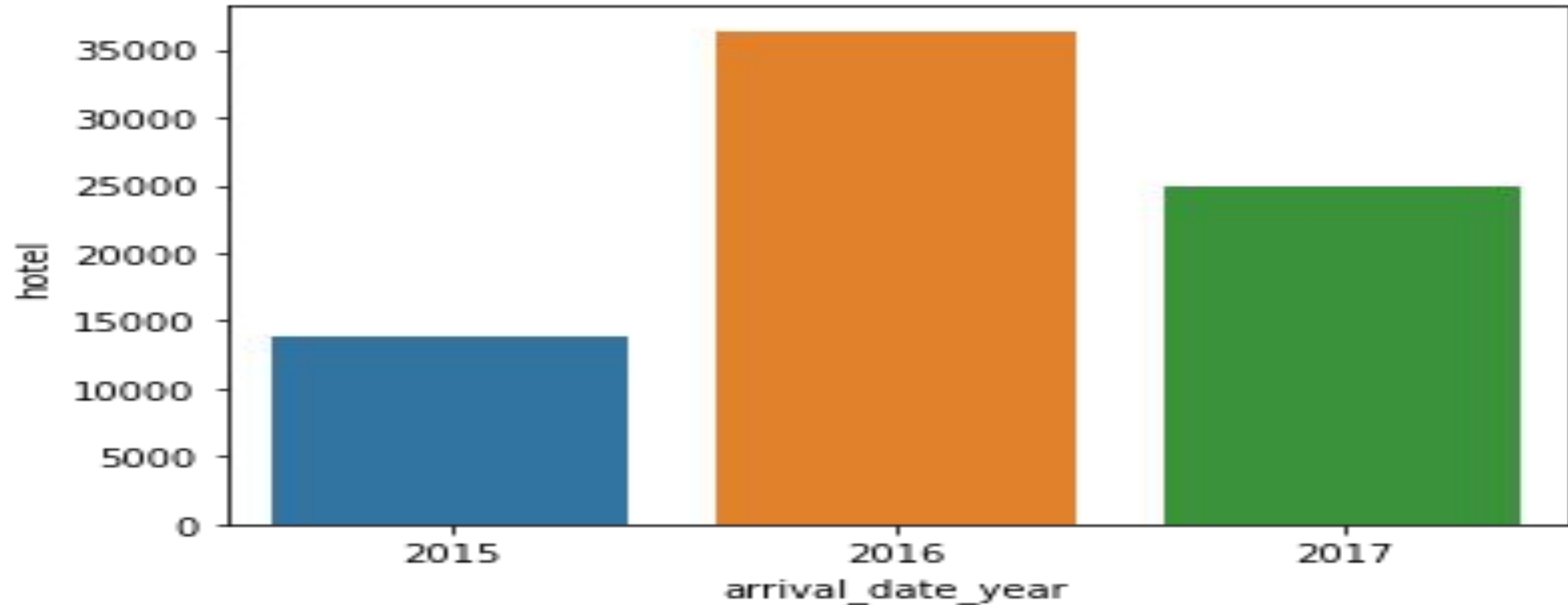


We can comment from above results that almost 40% of city hotel are cancelled while for resort hotels , less than 30% are cancelled.. So, city hotels are cancelled more often than resort hotels



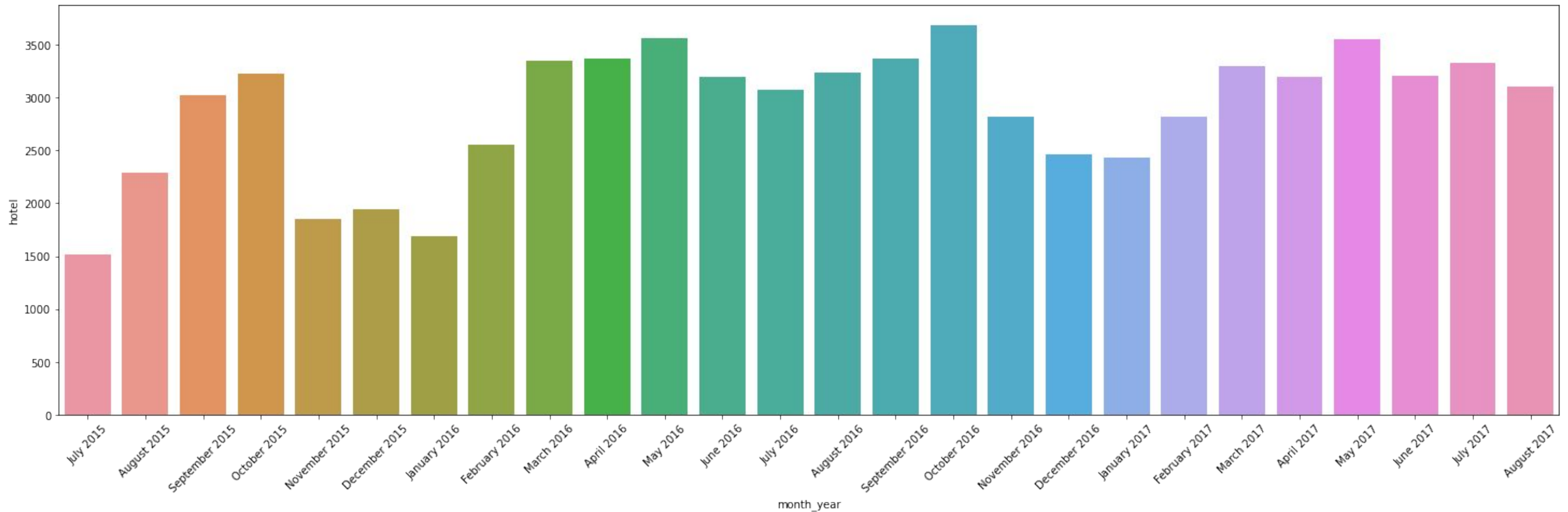
**Using only the non-cancelled bookings for further study.**

**Comparing bookings made year-wise.**



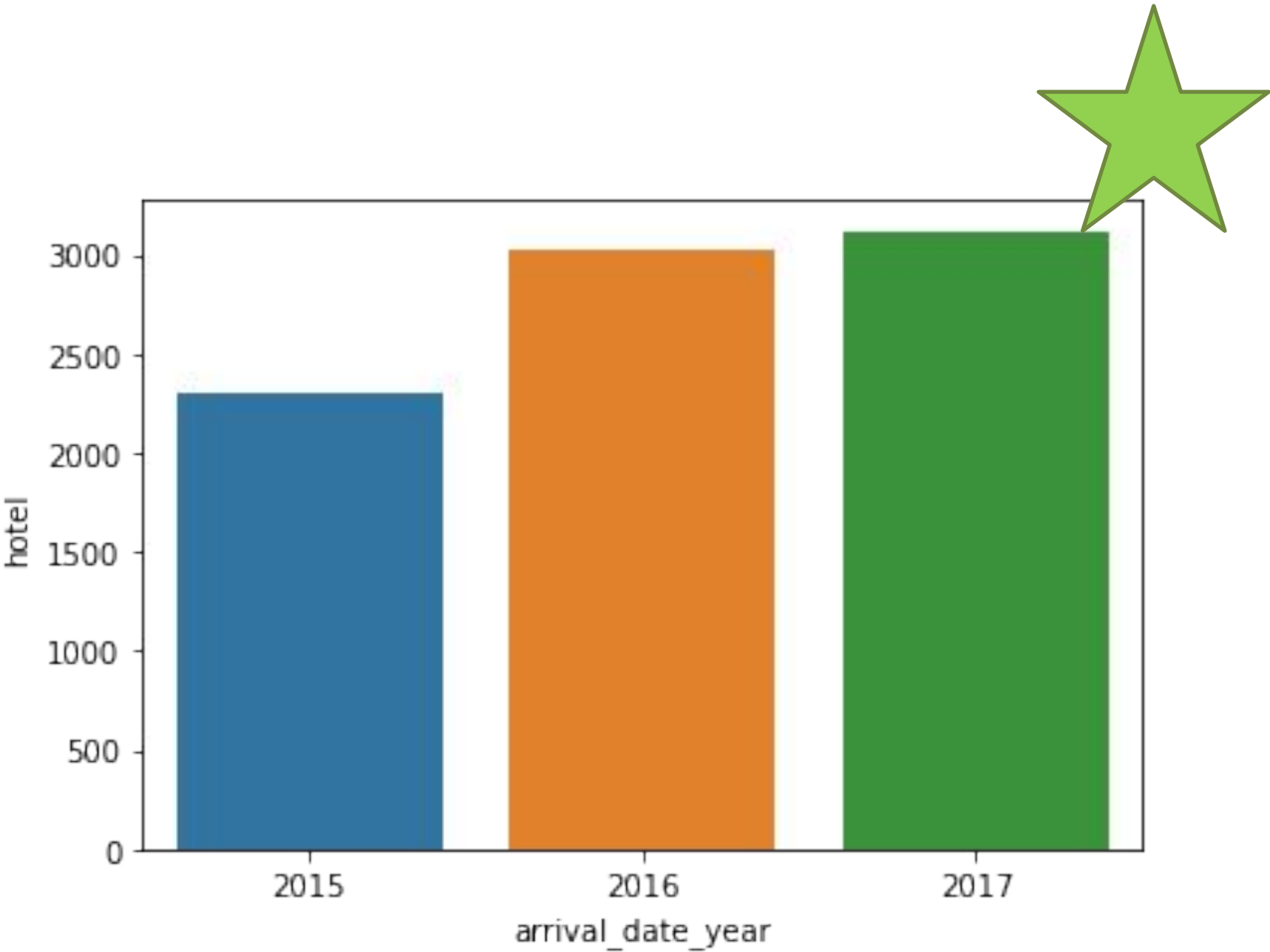
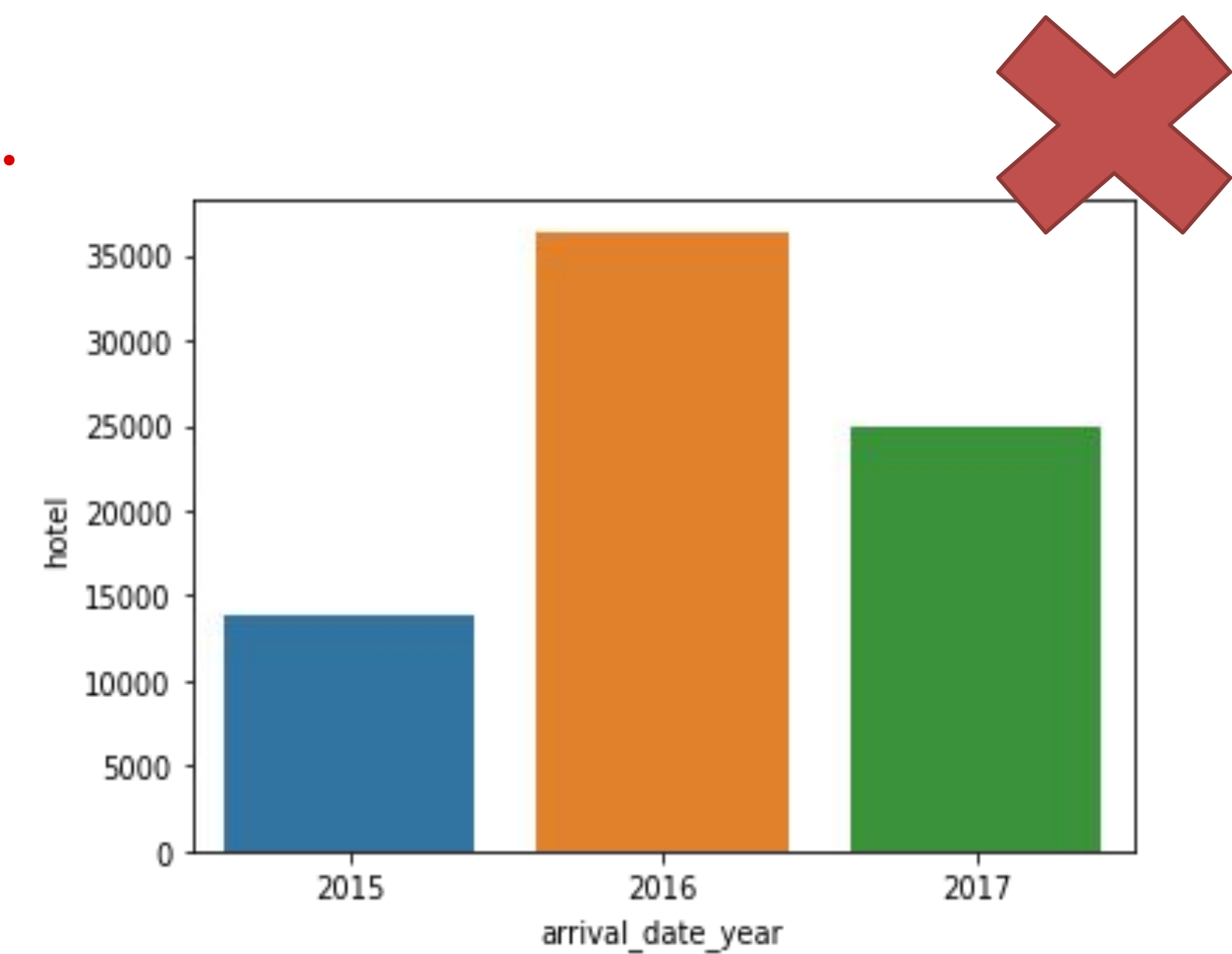
So the bookings almost doubled in the year 2016 as compared to previous year but in 2017 it decreased considerably, Lets check further why this happened...

# Checking the bookings monthwise.

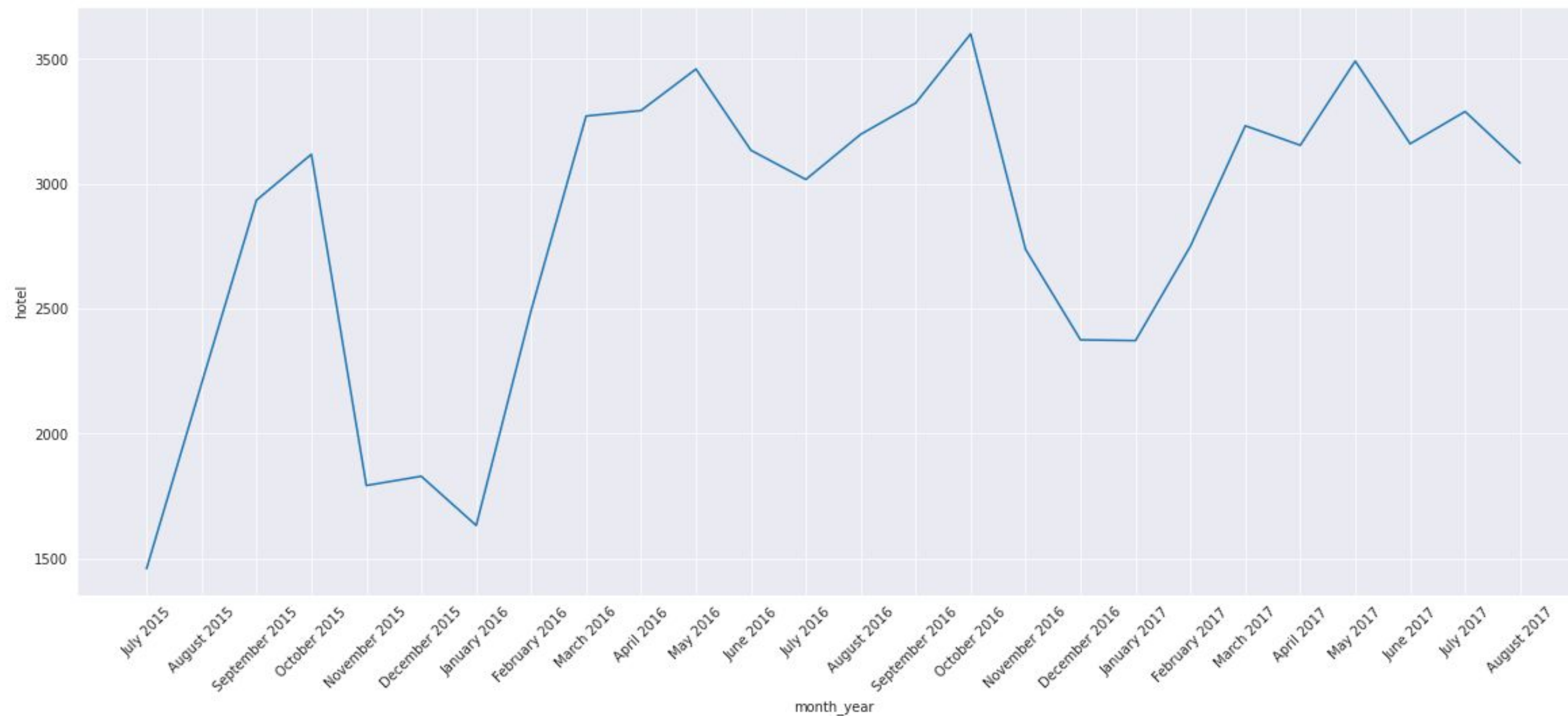


So we can see that the successful bookings have increased year to year but the growth from 2015 to 2016 is relatively large as compared to that from 2016 to 2017.

**Lets normalize the data and check the yearly trend.**  
**So, after checking the results, we can comment that the price has overall increasing trend from 2015 to 2017**



# Lets find the busiest months



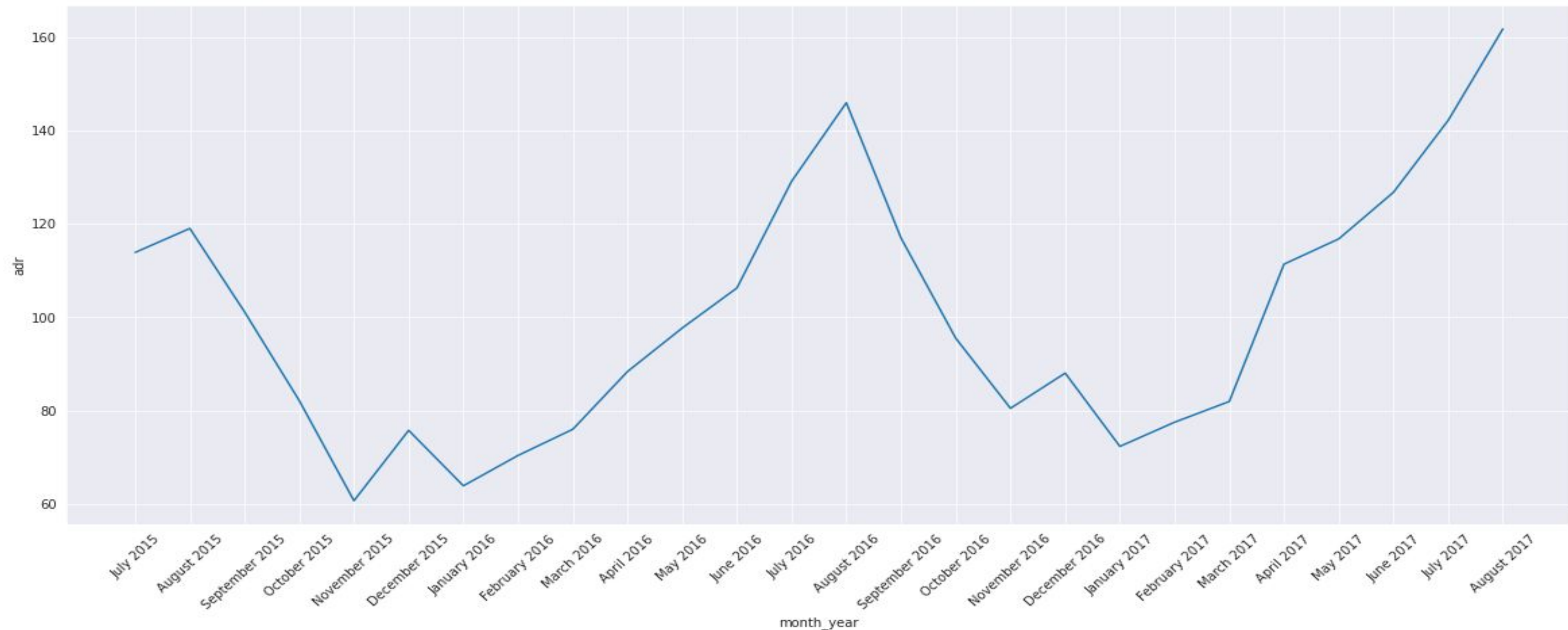
From March to June are the busiest months!! -- Peak Season

Also , it is evident that the number of visitors are lowest at year ending and starting time.. --- Off-season

Also, there is spike in number of visitors in the month of October when the weather is quite pleasant.

While, the number of visitors decreases at **year ending** i.e. in the winter season and also monsoon season experiences less visitors.

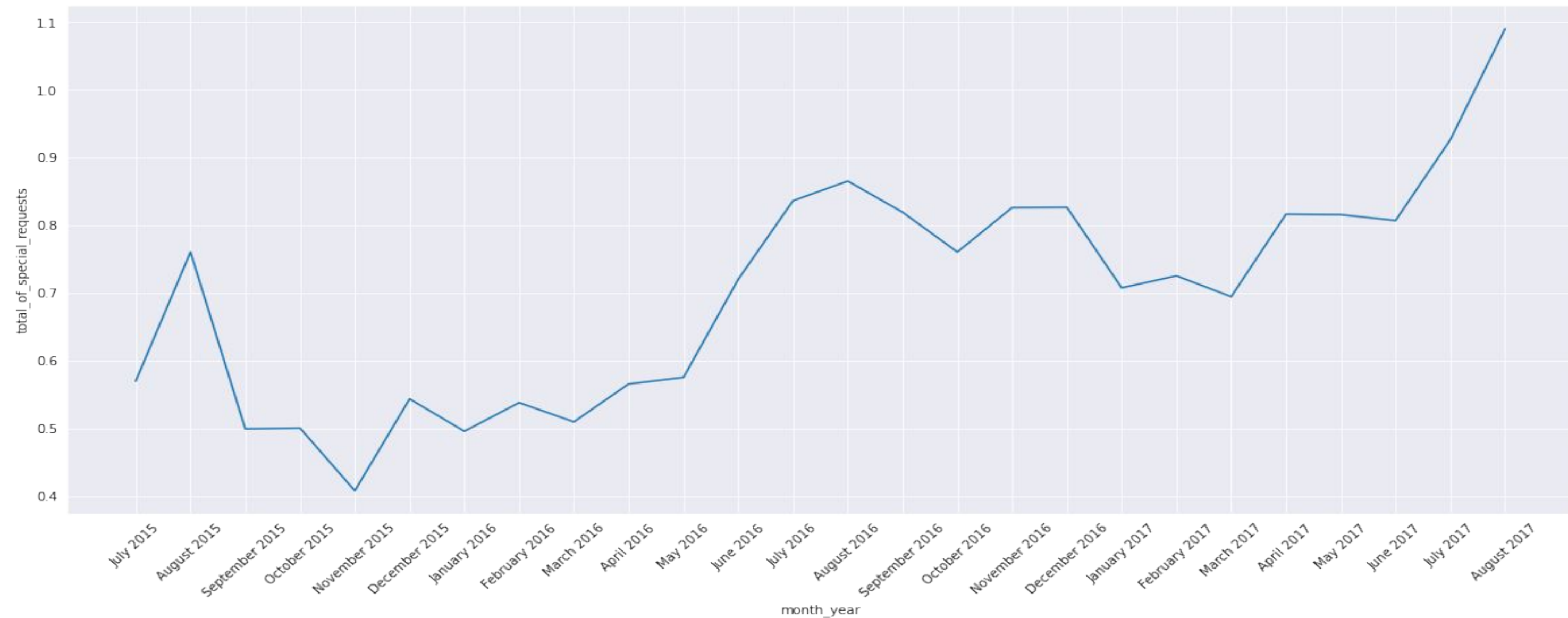
# Lets check average ADR for each month



As we can see from the trend that the prices are increasing with time as expected

Average rates are high during July-August months while they are lowest in Dec-jan which is the off season

# Lets check when hotels are receiving lots of Special requests



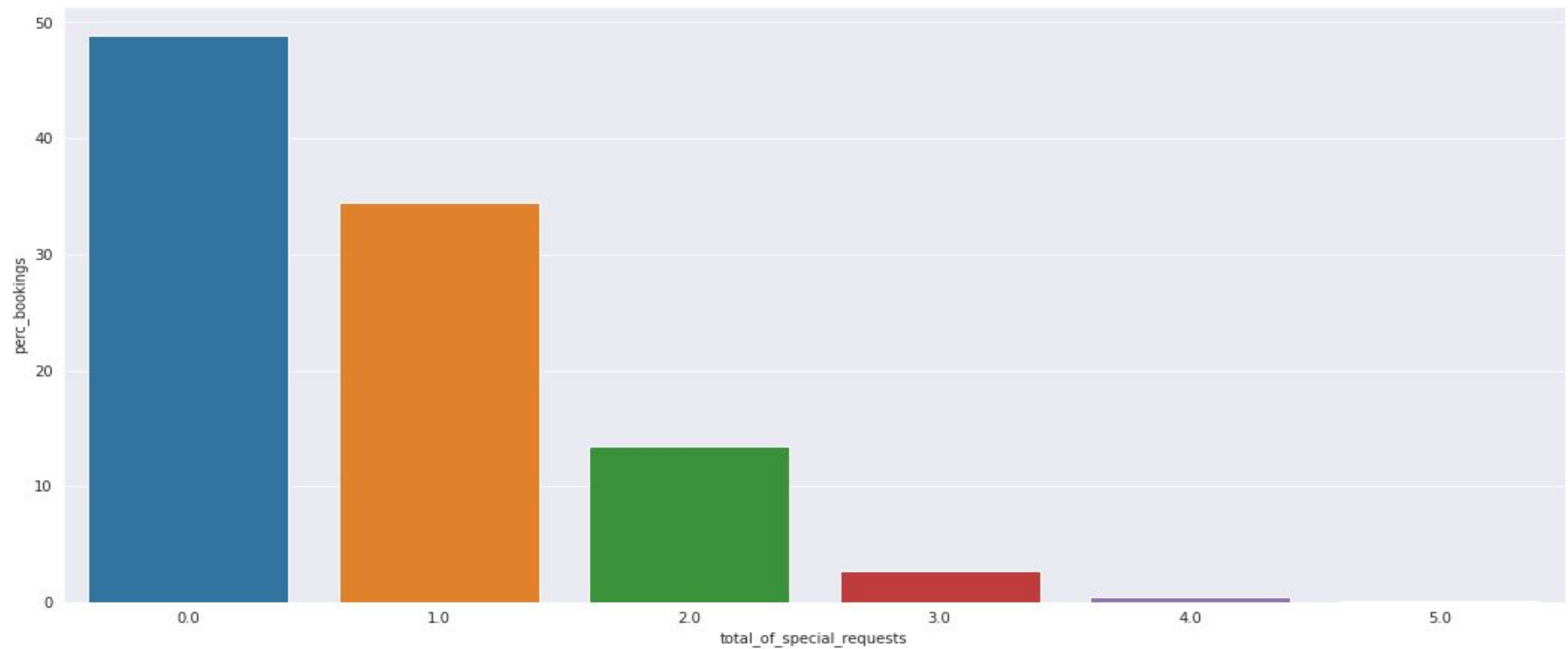
The average monthly special requests are increasing with time and are highest in **July-August**

Also as expected requests are lowest in off season i.e Dec-Jan.

So, the special requests and rates are both showing similar seasons.



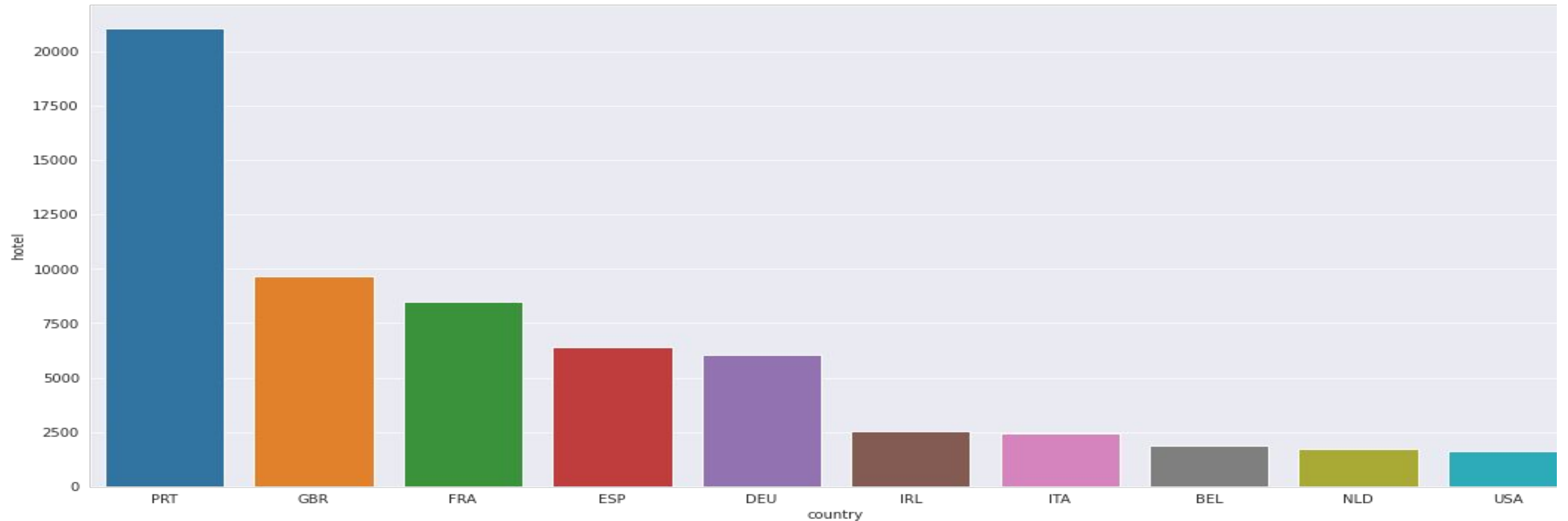
Lets also check how many requests are received from each visitors



So, more than half of visitors have at least 1 special request

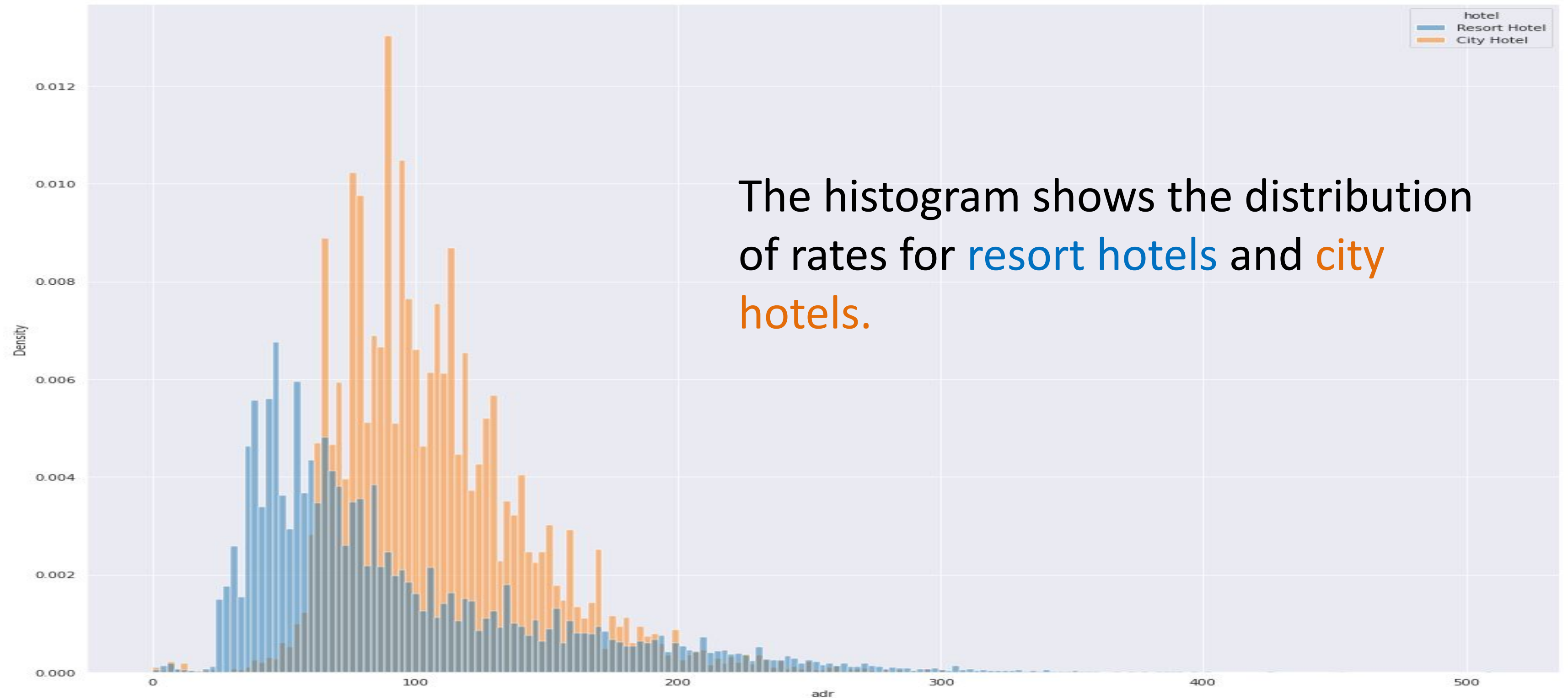
# Lets find maximum visitors are from which countries.

As the no. of countries are 177, which is quite large to plot in graph , we will check the top 10 countries



The top three countries in order are Portugal(PRT) ,United Kingdom(GRT) and France(FRA)

# Lets check the range of rates of hotels



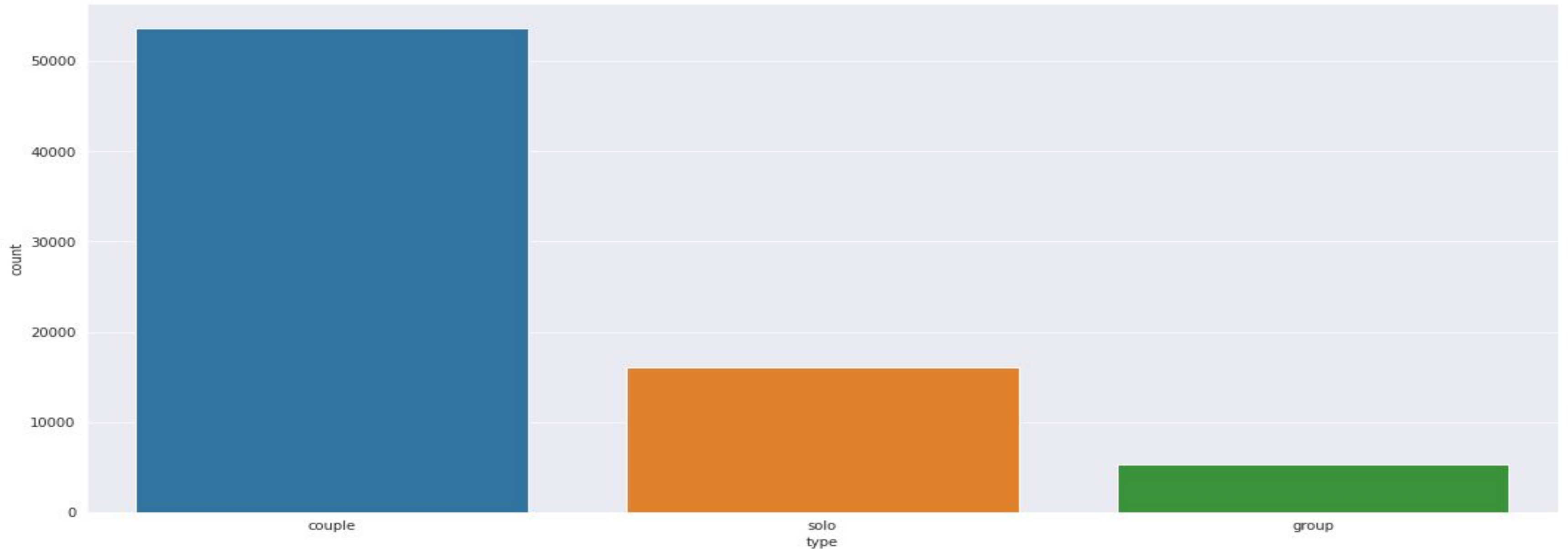
It is evident that there are more bookings for resort in lower price range as compared to city hotels. In medium to high price range, city hotels have more bookings. For higher prices, again resort hotels are booked more than city hotels

# Lets compare the type of visitors

Dividing into three categories: **Solo, Couple, Family/Friends**

For two adult visitors without children/babies , we will treat them as a couple

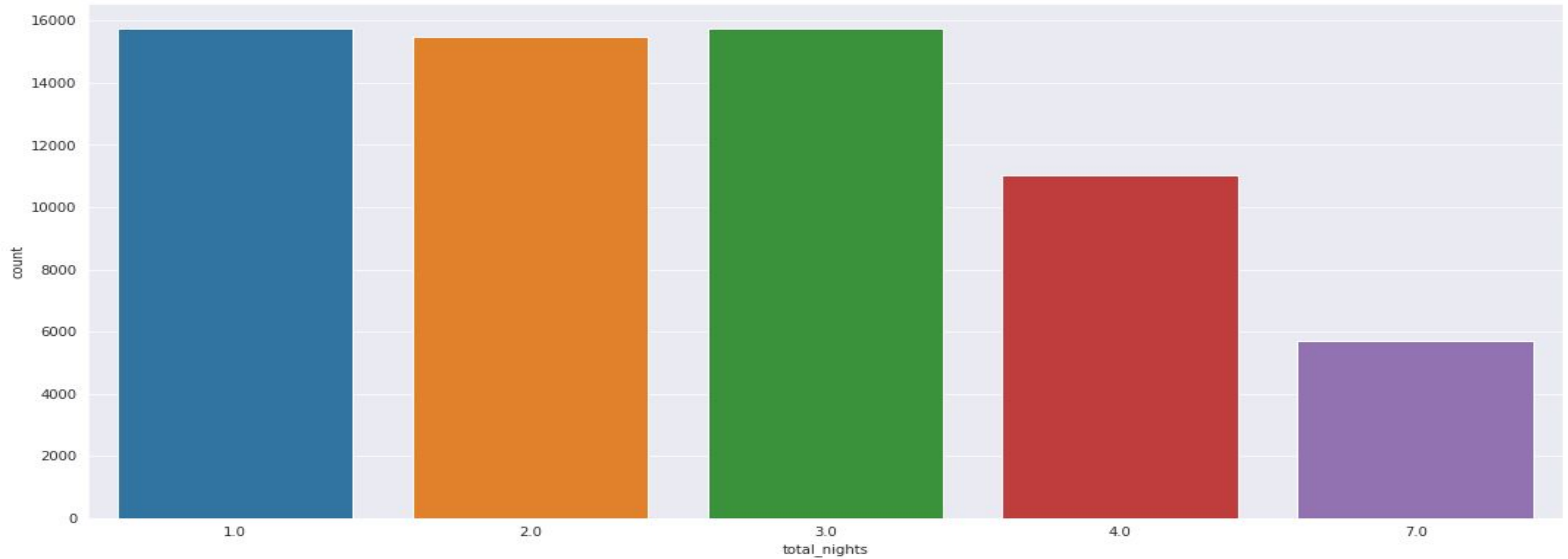
For more than 2 visitors it can be family or group of friends or relatives



We can comment that couples are the most frequent visitors, in fact the no of couple are almost three times solo travellers .. followed by group which are the least frequent ones.

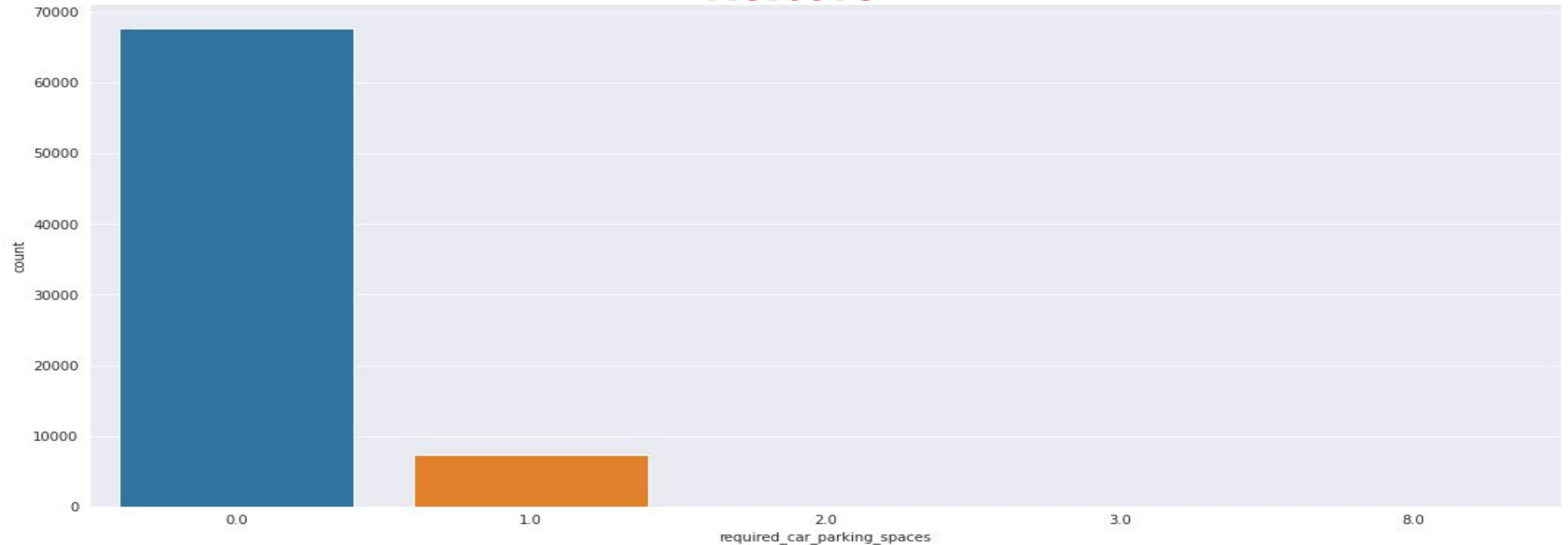
we have found one interesting co-incidence that all bookings for visitors more than 4 are cancelled.

# How long do people stay ?



So, majority of visitors stay for upto 3 days

# Lets check how many cars parking spaces are required by visitors

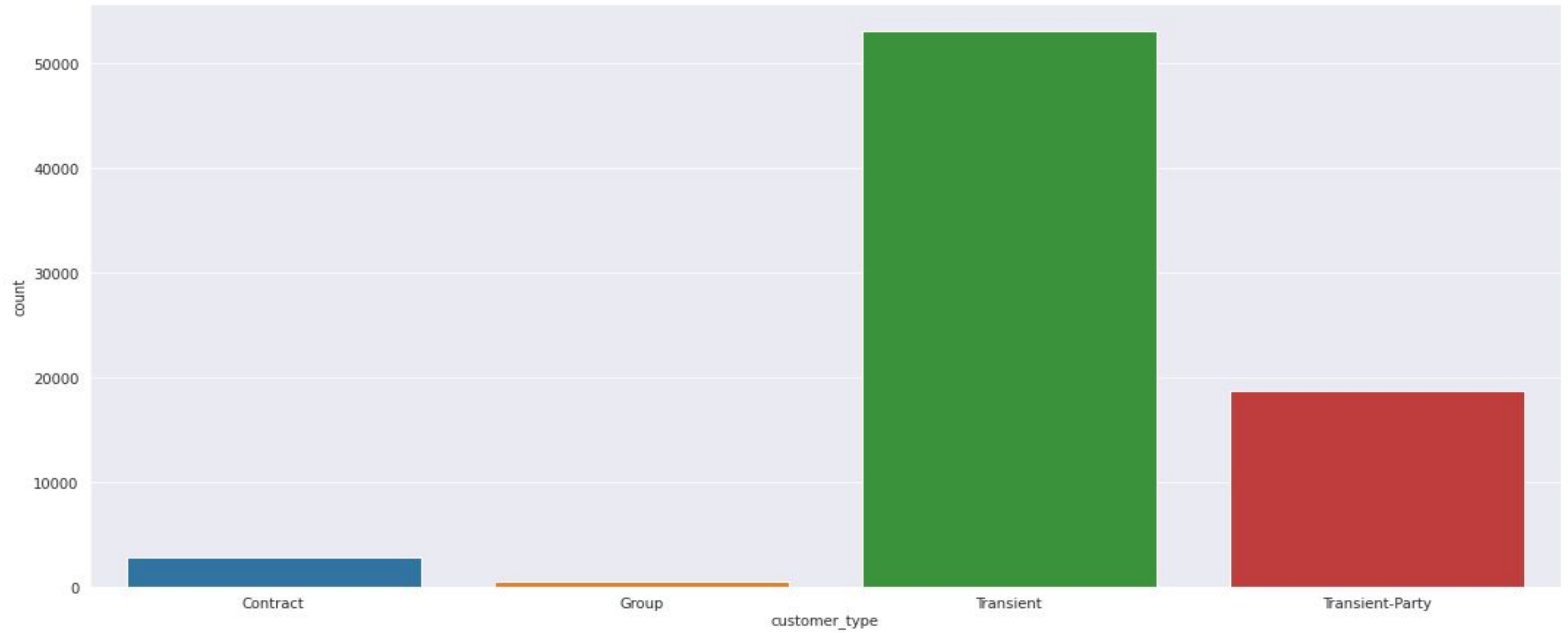


So there are very rare visitors who require more than 1 car parking spaces.



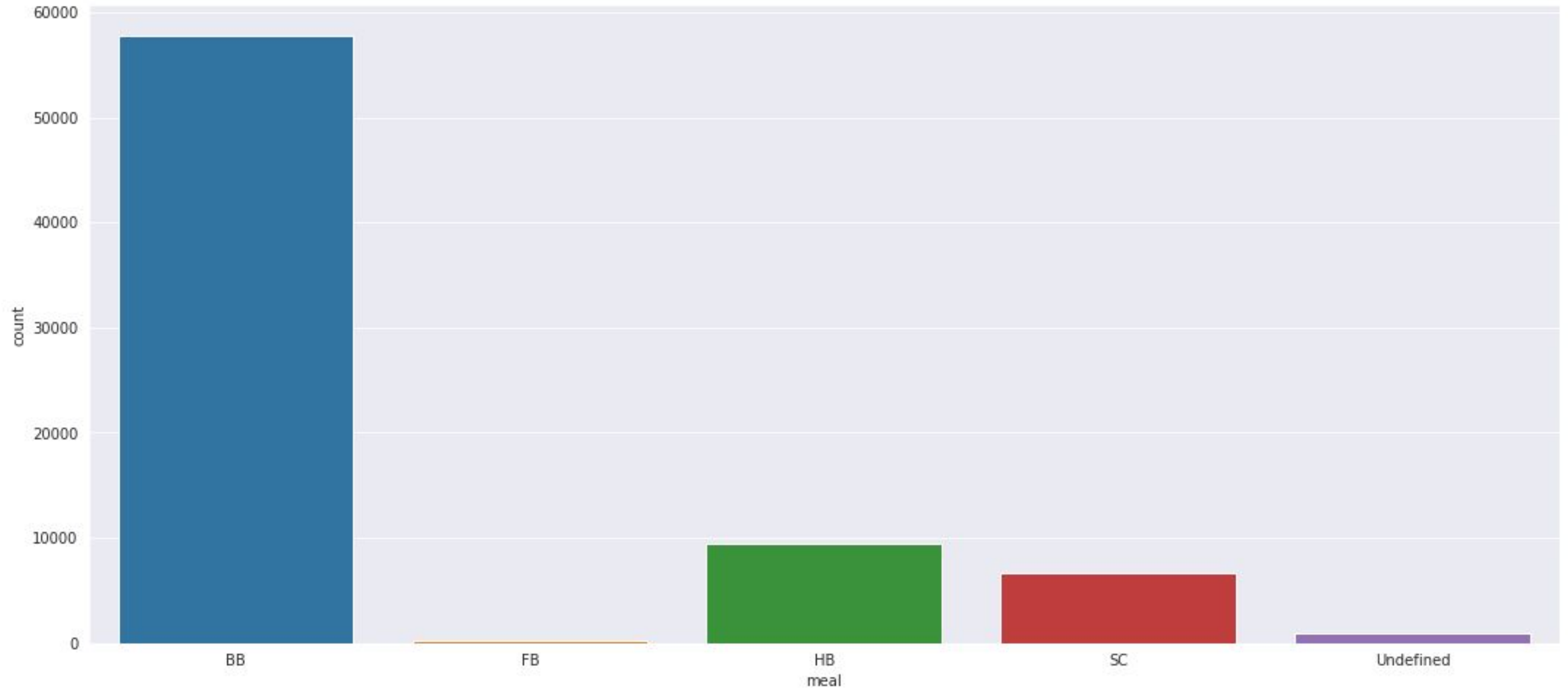
# Lets check the majority type of visitors

We have four types here : **Transient**, **Contract**, **Transient-Party**, **Group**



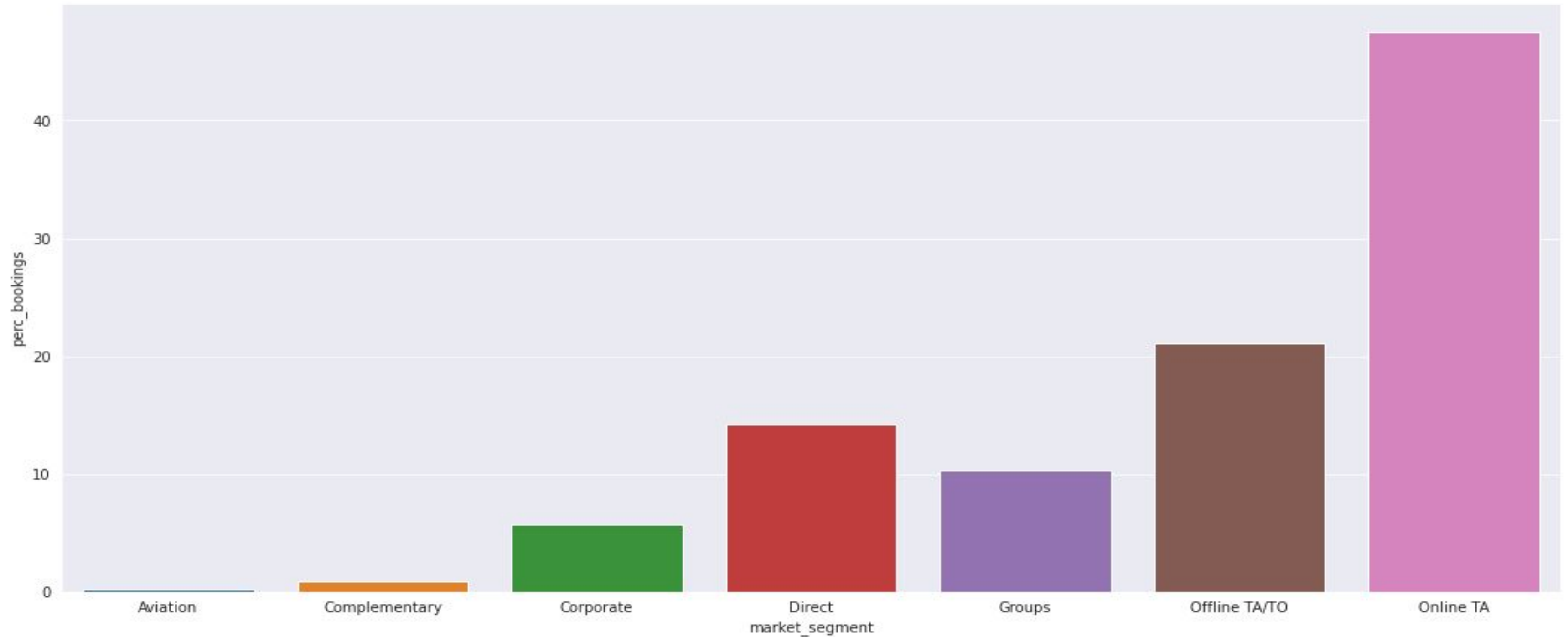
So, majority of visitors are Transient Type.

# Lets check what type of meal do visitors prefer



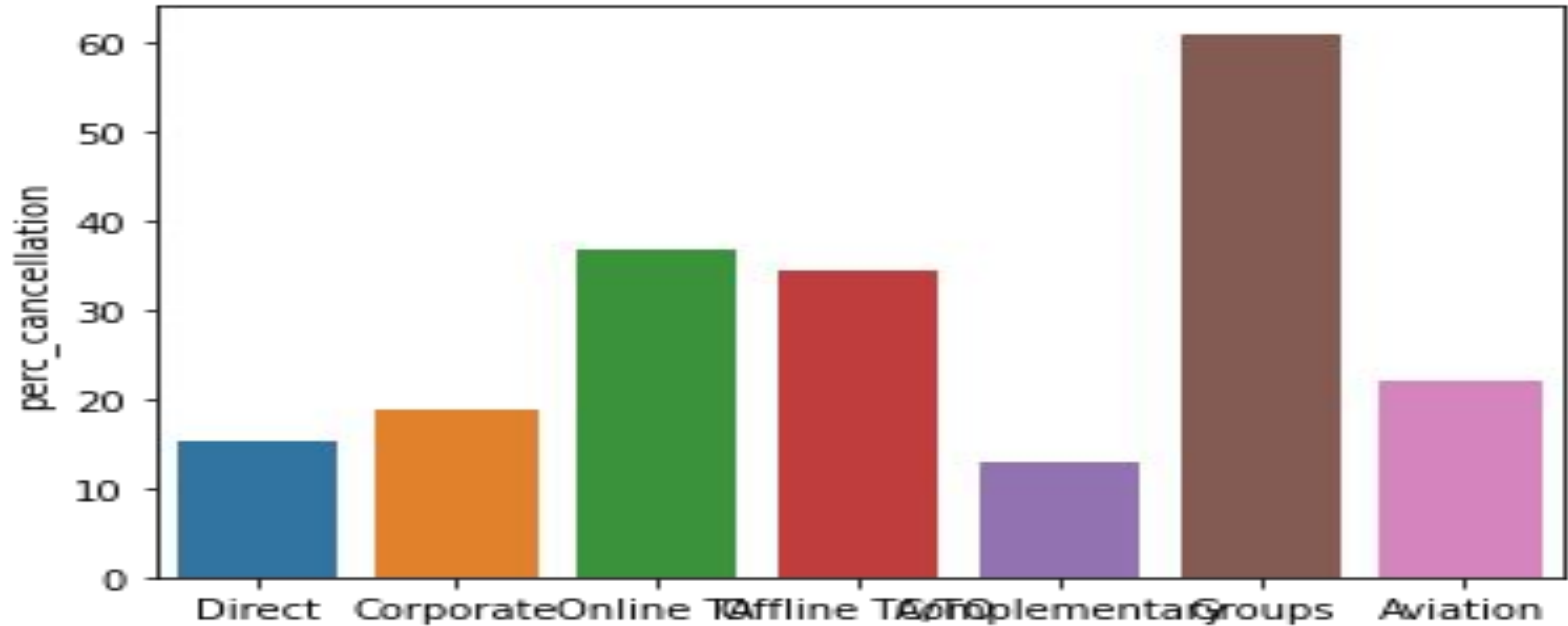
So, most of the visitors prefer breakfast only...

# Lets find which market segment brings in more visitors



So, we can conclude that almost half of the visitors come from bookings by online travel agencies..

# Lets check relation between cancellation and market segment



So here we can say that Groups bookings are most-frequently cancelled more than 60% of times , while the complementary and direct bookings are rarely cancelled..

only 13% and 15% respectively

# Challenges

- To understand many of the terms in the data sets like lead-time, types of market-segments, distribution channels, country, etc as most of them were abbreviated.
- As the entire dataset has two types of hotels : City and Resort, we had to do analysis for each type simultaneously.
- Dealing with the null values and erroneous entries such that negative ADR , etc.
- Lots of outliers in the data.

CHALLENGE



# Conclusion

- We had data of more than 1 lakh hotel bookings by people from various regions, having two types of hotel.

Following are the **takeaway** points that would be handy to ameliorate hospitality services as well as beneficial to a user beforehand booking.

- The peak season is Mar-June (summer) and the off-season is winter (Dec-Jan).
- Apart from this there is spike in visitors in Oct month.
- The hotel rates are highest around August month and lowest in winters.
- Bookings are fewer at year-ending & start of year, and the rates are also the lowest in the period.
- Hotel received high no of special requests from visitors in August month.
- Groups bookings are most-frequently cancelled while direct and complimentary bookings are rarely cancelled.
- Online travel agencies bring in most bookings.
- Most of the visitors prefer breakfast only as their meal-inclusive option
- Majority of visitors stay for up to 3 days
- There are very rarely any visitors who require more than 1 car parking spaces.
- Most frequent bookings are done by couples.
- Maximum bookings are from Portugal , United Kingdom and France
- Bookings are increasing with time and 2016 has seen noticeable increase in bookings which is relatively less in 2017 .
- City hotels are booked almost twice the no of times that Resort hotels are booked
- City hotels are cancelled more often than Resort hotels



**Thank You**