

Group21_BA_Capstone_Module1

Leila Allahham, Seok Lee, Andrew See, Amani Akhtar, Maitri Patel

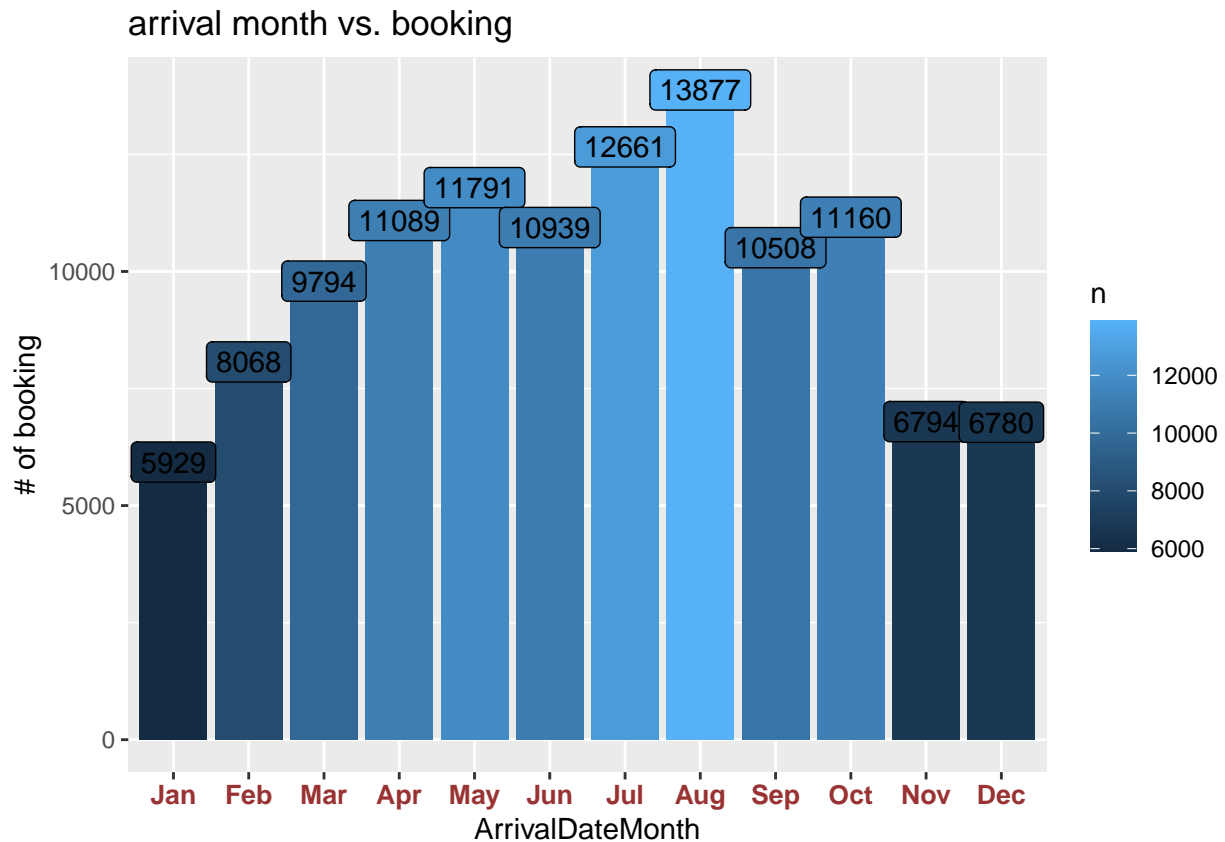
10/10/2020

Contents

Question 1 - Distribution of number of bookings by arrival month	2
Question 2 - Distribution of number of bookings by customer segment	3
Question 3 - Distribution of cancellation by hotel and month	4
Question 4 - Relationship of lead time and cancellation	5
Question 5 & 6 - Effects of deposit on cancellations/ by hotel	6
Question 7 - Effects of deposit on cancellations by customer segments	7
Question 8 - Distribution of the number of nights spent at the two hotels for each month	9
Question 9 - Distribution of ADR by hotel, month, and distribution channel	10
Question 10 - Relationship between repeated guest and cancellations	12

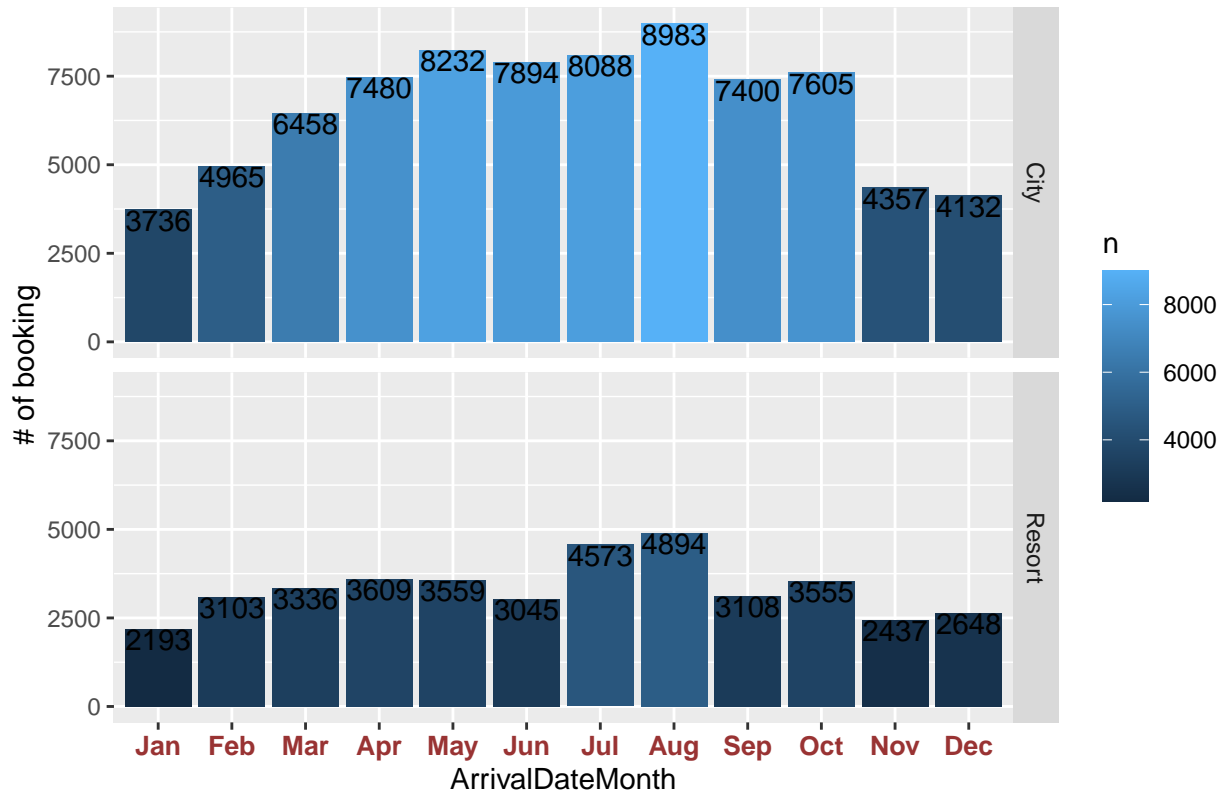
Question 1 - Distribution of number of bookings by arrival month

The team noticed from the graph that July and August have the highest booking, but we have to keep in mind that here we're not differentiating the hotel "City, Resort" which may not reflect the true peak month for each hotel type. but overall and based on the aggregated booking for city and resort hotels we can say that Jan, Nov, and Dec are the off-peak season and Jul & Aug are the peak season which makes sense since Portugal is located in the northern hemisphere and summer starts at June 21th, winter starts at December 21th . One interesting aspect for further analysis is the drop in booking in June as seen in the graph despite the upward trend in the previous months. The distribution is almost normal with left skewness.



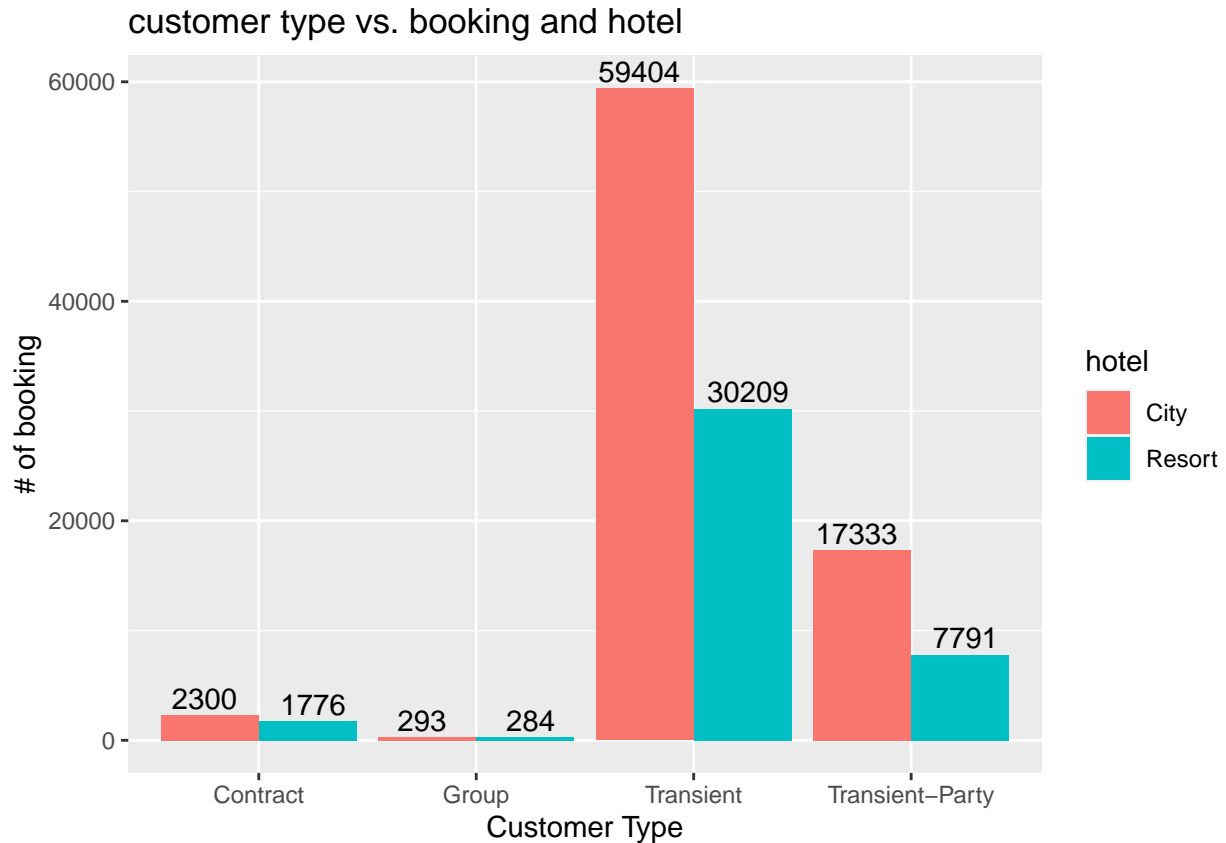
To get a better picture, let's differentiate between the two hotel types: we notice that overall "City" hotel has higher reservations than "Resort" hotel, Aug is still the highest booked month for both. Dec, Nov, and Jan are still the off-peak season for both types as well. for the "Resort" the booking is almost equal all months of the year except for Jul, and Aug when there's an increase, while in the "City" the booking varies between the different months.

arrival month vs. booking and hotel



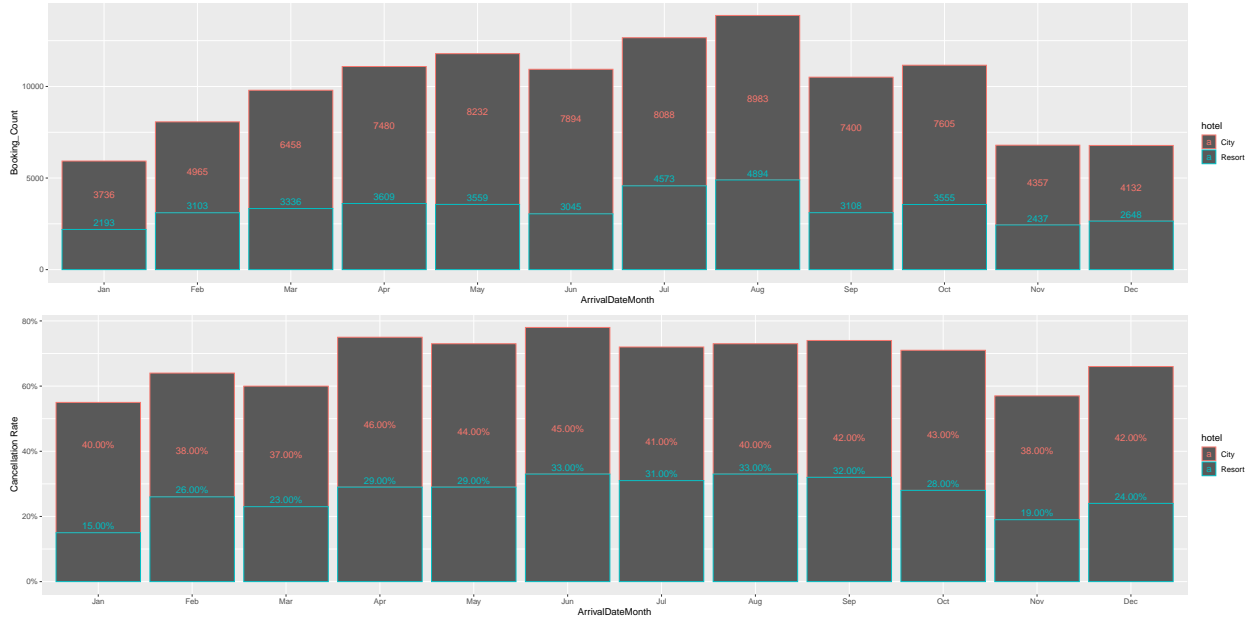
Question 2 - Distribution of number of bookings by customer segment

From the graph, the team notices that “Transient” booking has the largest share of number of bookings in both “City” and “Resort” hotels and as we saw earlier the booking in “City” hotel is higher than it is in “Resort” hotel which indicate that most of the revenues comes from transient customer segment since transient customers are those who book at the last minute and walk-in guest with short stays and it’s known that the earlier you book the lower the price you get. Transient-part has the second highest booking and the “City” hotel still have a higher booking number than the “Resort” hotel. the second important segment in hotel customers is the “contract” customers, those who are corporate customers with business agreements with the hotel and are considered a secured income to hotels because if the hotel has a certain company agreement it knows how much it will yield from it. “Group” customer segment constitute the lowest aggregate booking in both “City” and “Resort” hotels which means they contribute the least to hotel revenue.



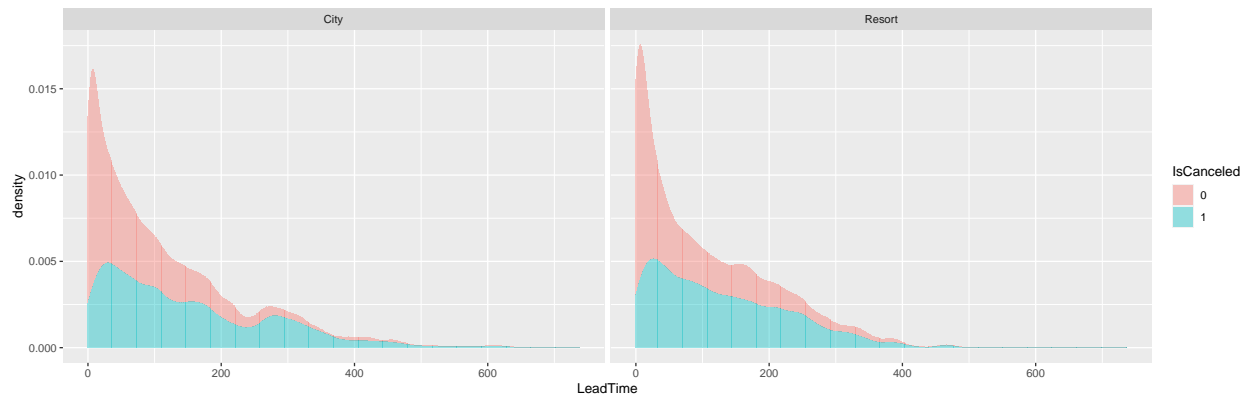
Question 3 - Distribution of cancellation by hotel and month

The team sees that the cancellation rate in “City” hotel ranges from 37% to 45% and it is higher than the cancellation rate in “Resort” hotel that ranges from 15% to 33%. June in both hotels has a high a cancellation rate 45% and 33% for “City” and “Resort” respectively and at the same time has a lower booking compared to the month before “May” and the month after “July” in both hotels this means that we should focus on June in our promotions to drive higher revenues since the occupancy is really low. we don’t have the overall capacity of the hotel but if we multiply the booking number by the cancellation rate we’ll get the lost occupancy for each month, this number can be used to for exploit overbooking.



Question 4 - Relationship of lead time and cancellation

Team notices that most bookings that are made before 400 days are canceled which makes sense because those bookings are made before more than a year. the behavior in the “Resort” and “City” hotels are consistent the higher the lead time the higher the probability of cancellation. All bookings with more than 450 days lead time in “Resort” hotel are canceled and all bookings with more than 500 days lead time in “City” hotel are canceled.



Question 5 & 6 - Effects of deposit on cancellations/ by hotel

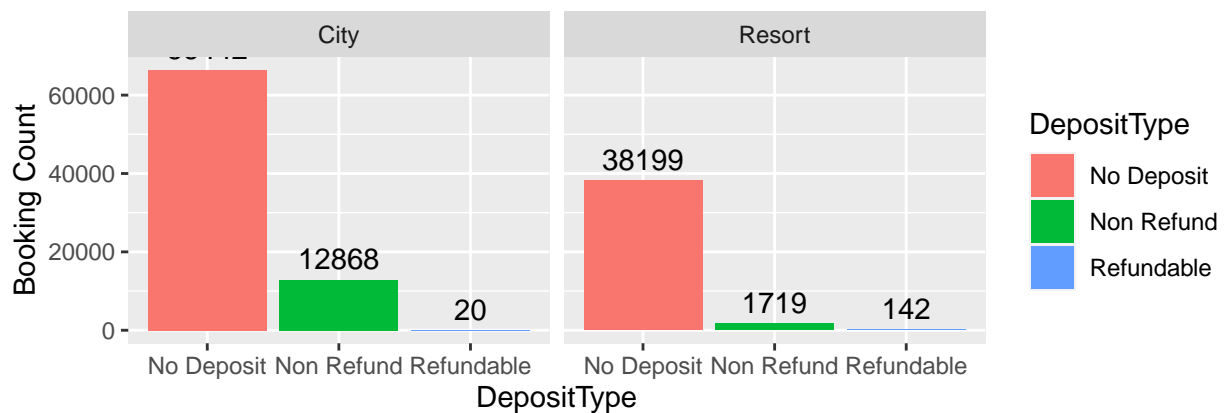
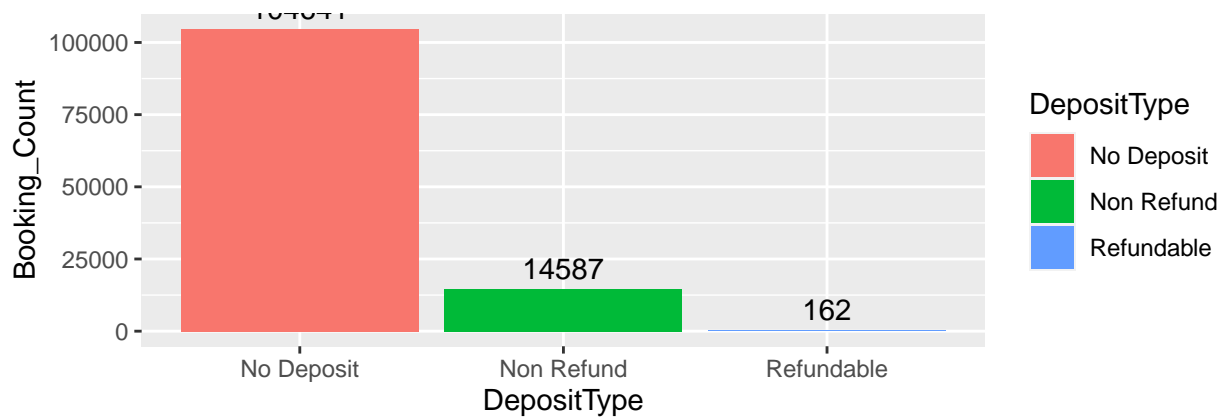
The prevailing deposit option to reserve hotels is no deposit on both city and resort hotels. City has higher booking counts. City hotels exhibit a higher cancellation rate than resorts. Non-refundable deposits show the highest cancellation rate overall. There's less hotels that accommodate refundable options in the city than resorts, but it exhibits the second highest cancellation. Deposit is a significant explanatory variable to explain 22.94 % of variation in dependent variable. cancellation ($p < 2e-16$, R-square= 0.2294)

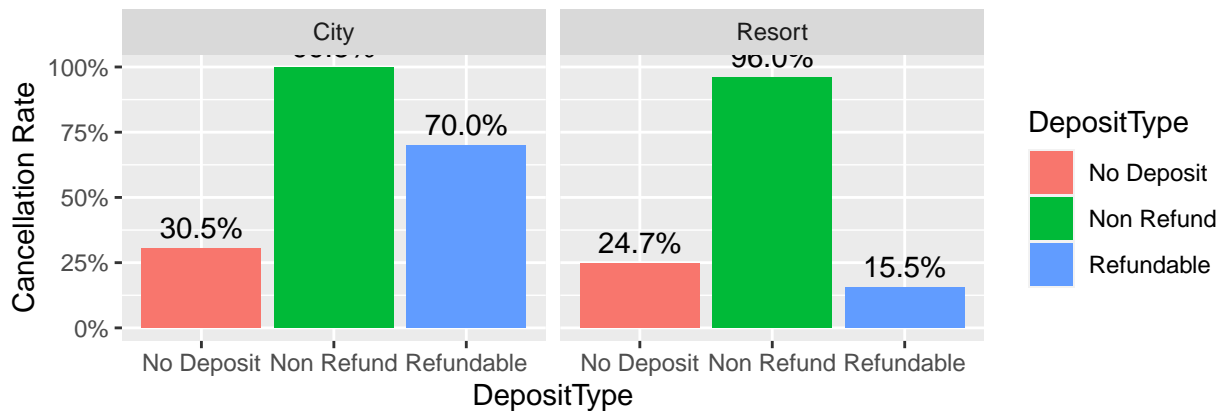
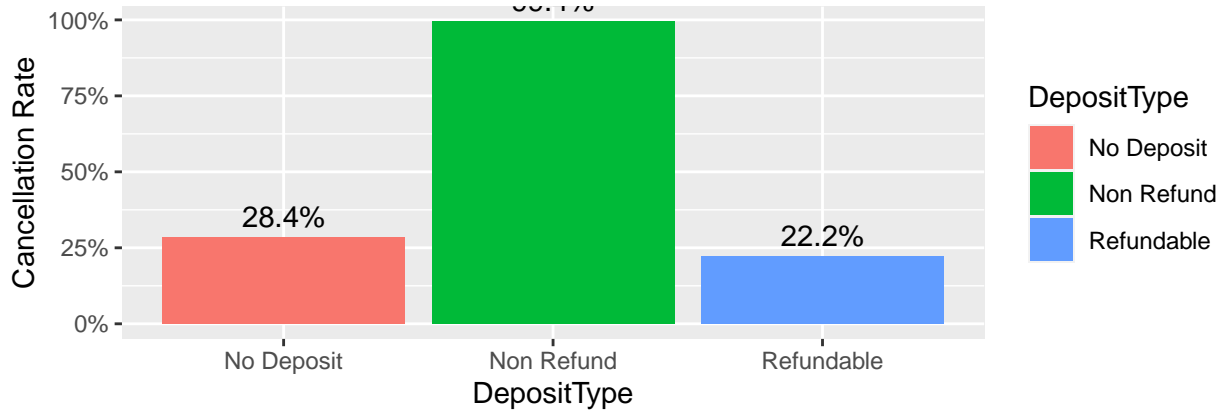
Business insights:

- The stricter refund policy, the more cancellation.
- The more choice of hotel in the city, the more cancellation.
- City hotels exhibit higher cancellation for refundable than no deposit.
- Deposit is a significant to explain the cancellation.

```
$y
[1] "Booking Count"
```

```
attr("class")
[1] "labels"
```

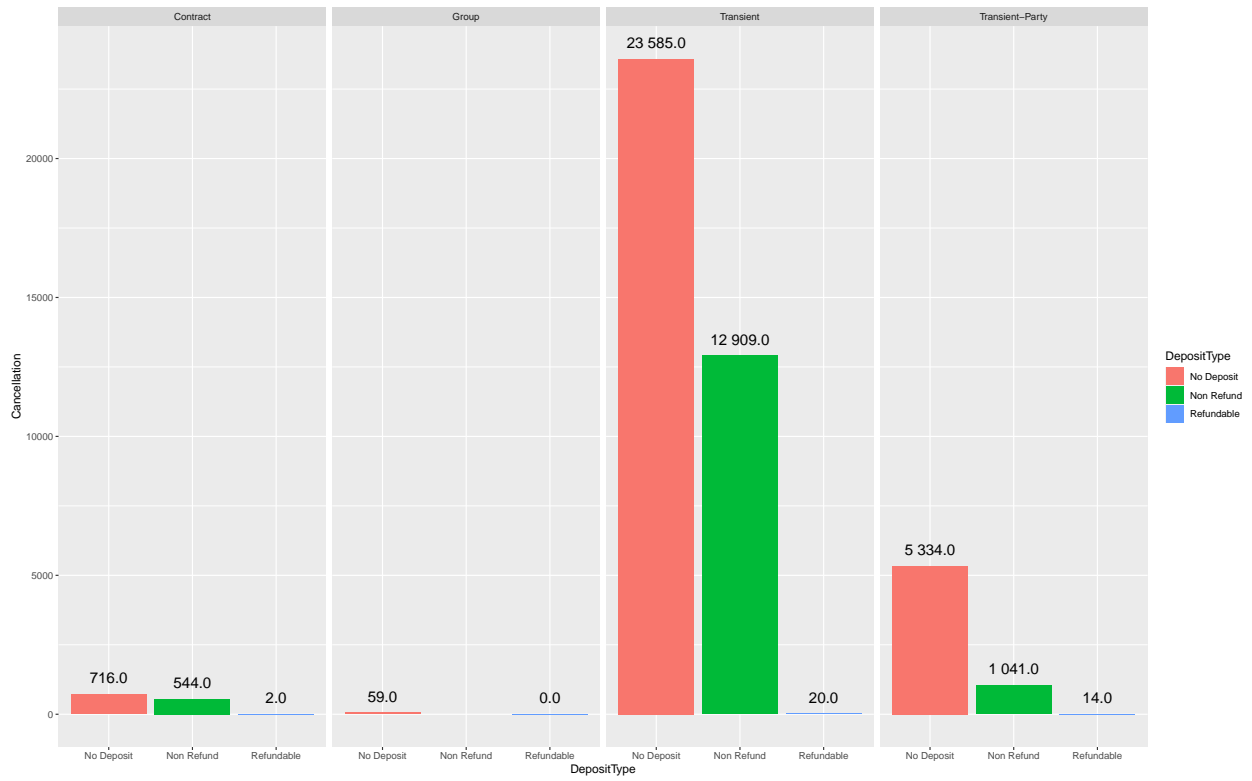
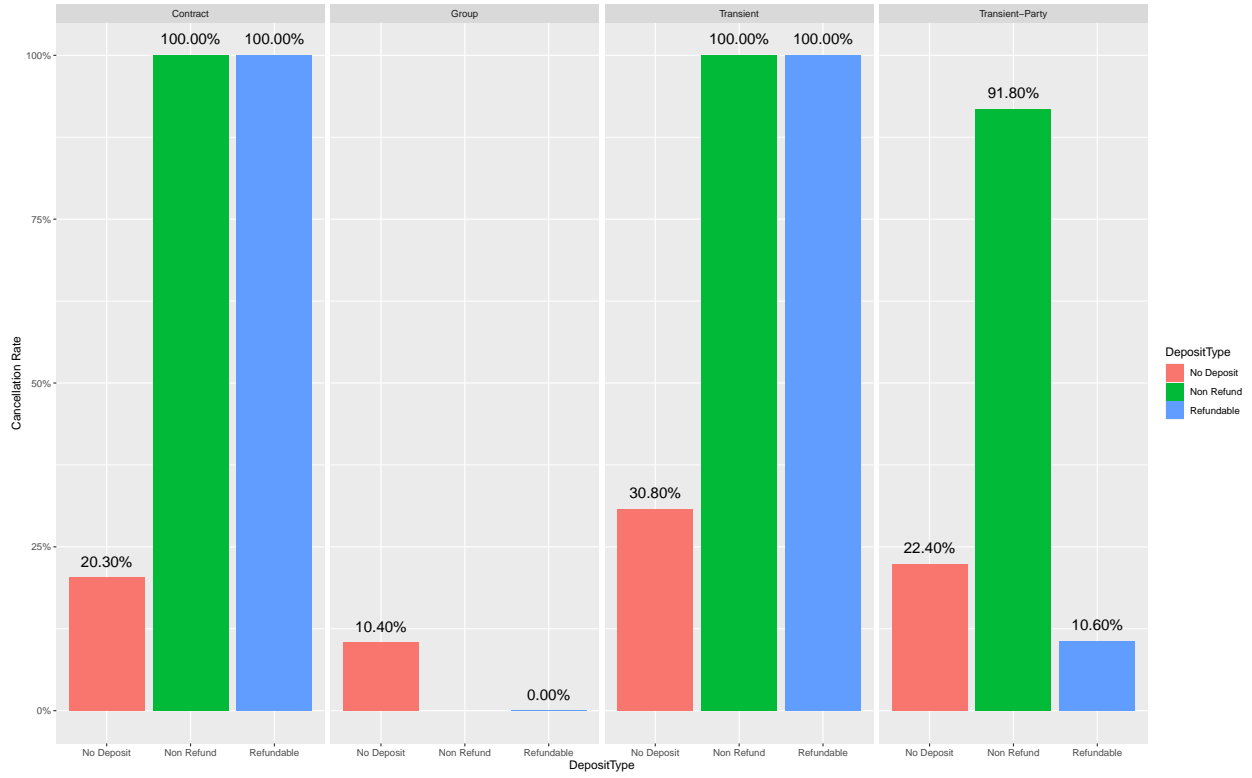




Question 7 - Effects of deposit on cancellations by customer segments

we see that 100% of contract customer cancel their bookings when the deposit is non-refundable or refundable which means that most contract customers book under no-deposit option, to look deeper into the data we plotted the number of cancellations per customer segment and for contract customers with non-refundable deposit booking the number was 544 (we suspected that the number is really small like “1” that is why the percentage is 100%) and based on the data from Question 2 we know that the total number of contract customers in both of the hotels is 4,076 customer, so 544 contribute to almost the eights of 4,076 but it’s a big enough number while if we look at refundable deposit for contract customers we see that there’s only 2 and the cancellation percentage is 100% but this is a really small number that can’t be used to infer or generalize that 100% of refundable deposit contract customers will cancel their bookings. for transient customers with a refundable deposit the percentage of cancellation is 100% but if we look again at the numbers of cancellation we find a small number “20” while in Question 2 we saw that transient customers in both hotels are 89,613 so we can’t rely on this percentage to say for sure 100% of transient customers with refundable deposit will cancel. the case is not the same for transient customers with non refundable deposit since the percentage 100% reflects the actions of 12,909 customers out of 89,613 so we would consider this percentage in our interpretation.

the transient-party with refundable deposit cancellation percentage of 10.6% is also not representative since it only reflects the action of 14 customers out of 25,124 (the number is taken from Q2) and the same can be said for transient-party with non-refundable that have a cancellation percentage of 91.8% that represent only 1,041 customers out of 25,124.

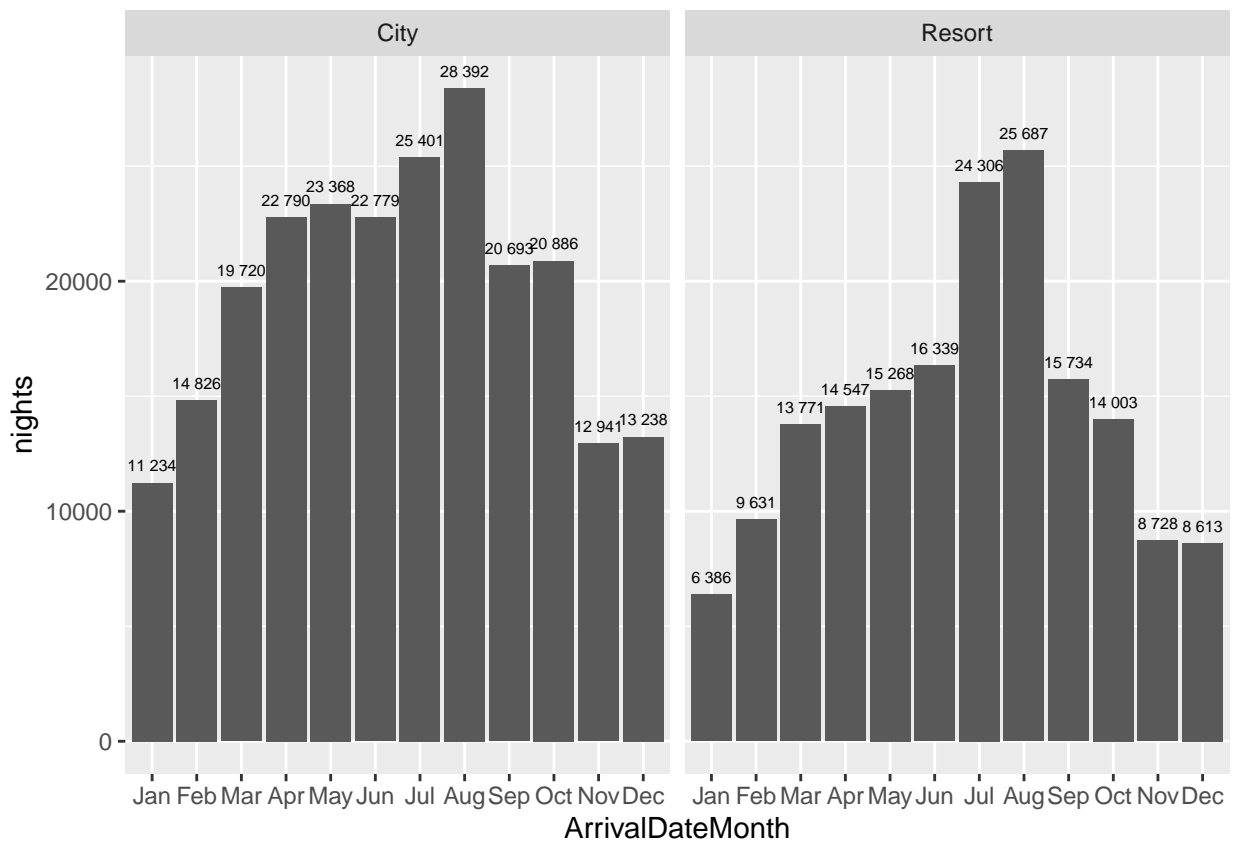


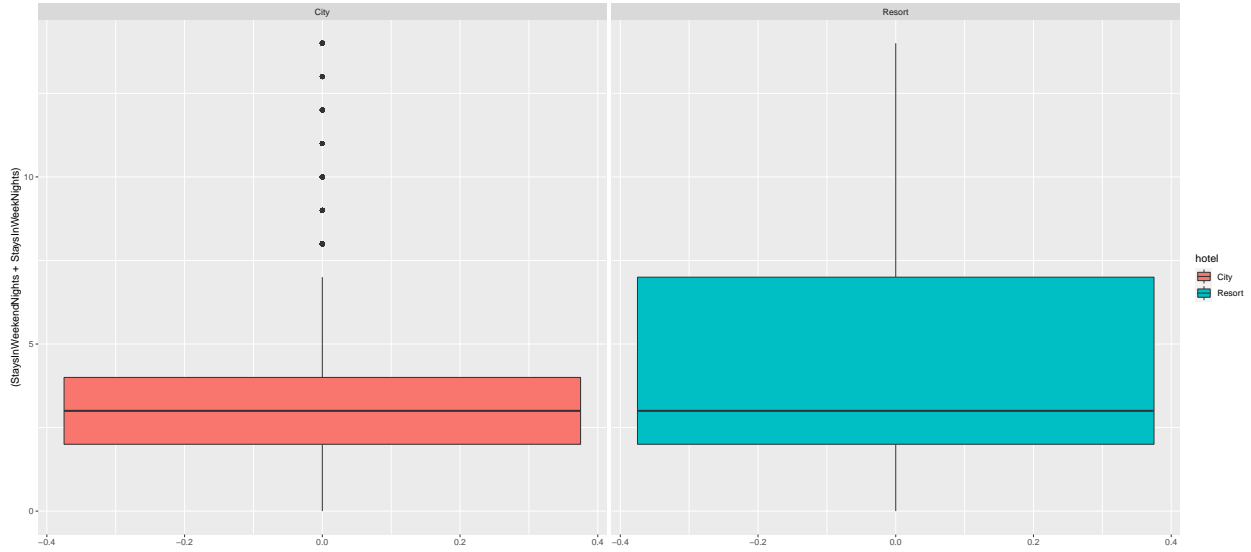
Question 8 - Distribution of the number of nights spent at the two hotels for each month

We start by finding the maximum number of days a customer spent at the hotel, since the number is 69 that is longer than a month and we want to visualize the data for each month

max stay	City	Resort
xxx	yyy	

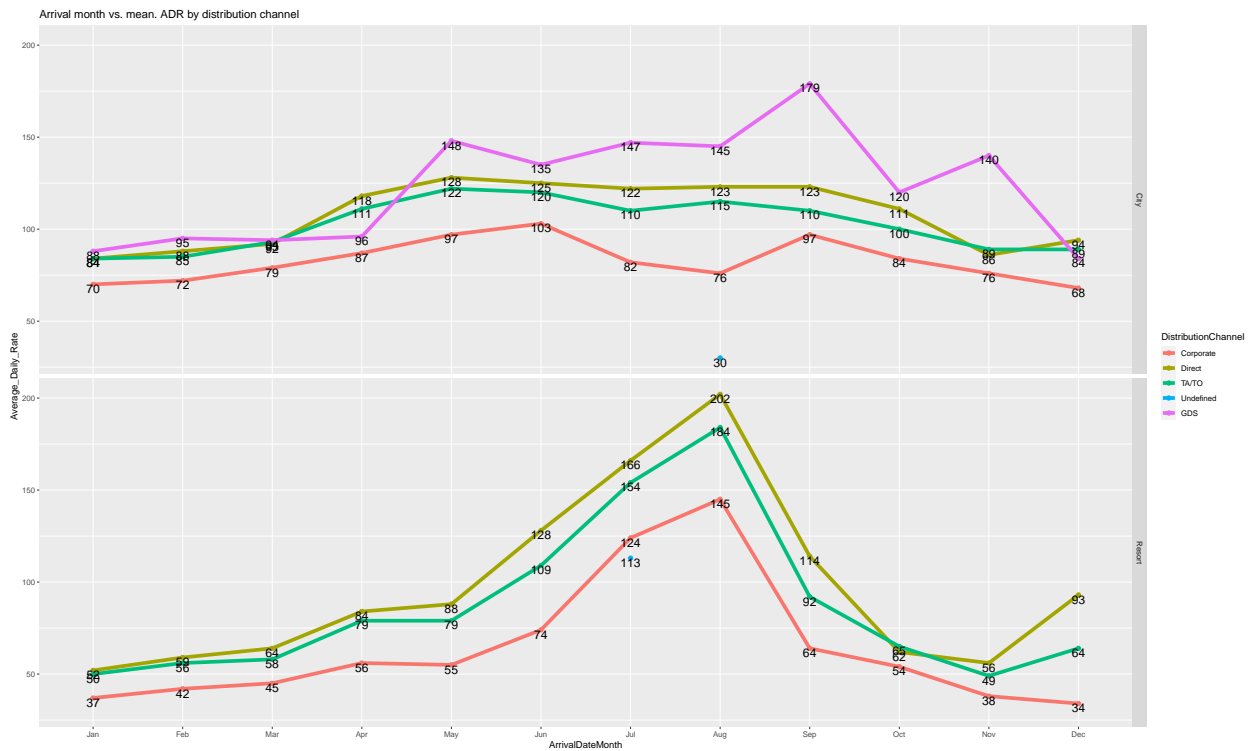
we draw a bar chart for each month on x-axis and the number of nights spent in that month on the y-axis. the findings are similar to Question 1 where we found that the maximum booking happens in August, here we find that the maximum stayed nights are also in August



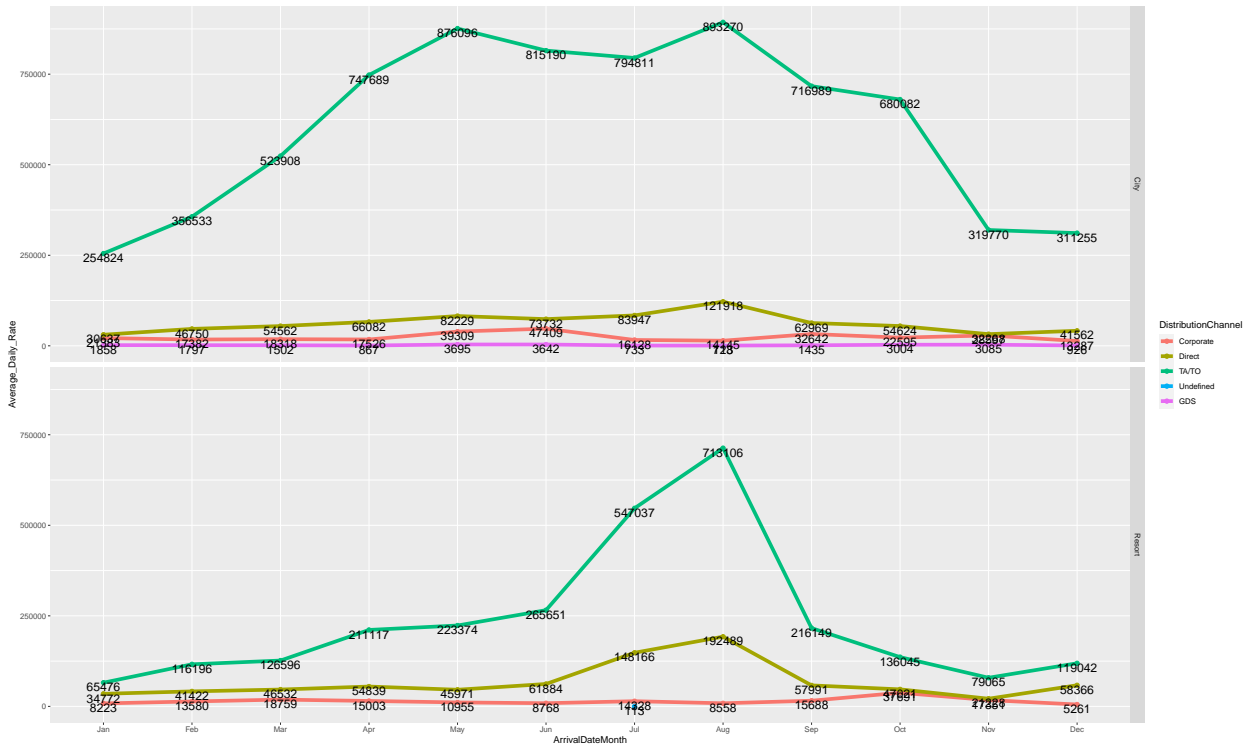
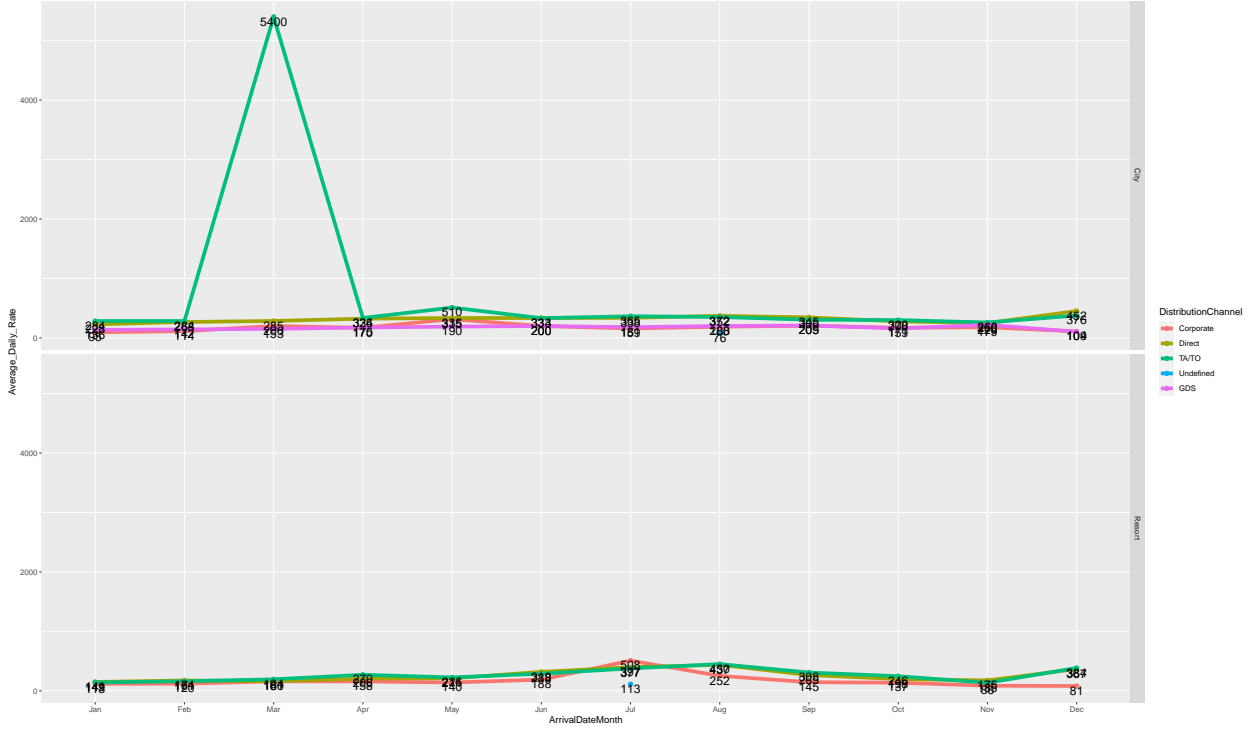


Question 9 - Distribution of ADR by hotel, month, and distribution channel

If we take the average value of ADR per hotel and month and compare it between the distribution channels (namely: Direct, Corporate, TA/TO, Undefined, GDS) we find that “corporate” channel charges the minimum average ADR fares over the whole year in both “City” and “Resort” hotels, while the rates for “Direct” and “undefined” are similar with a slightly higher rate for “Direct” in both hotels “City” and “Resort”. “GDS” channel only exists in “City” hotel and charges a higher rate compared to all other channels over the year except in April.



Arrival month vs. max. ADR by distribution channel



Question 10 - Relationship between repeated guest and cancellations

The output is intuitive, we see that if the customer is a repeated customer its cancellation rate is lower than a new customer and this is can be attributed to loyalty among other reasons.

