

Data Analysis Report on Hotel Bookings

Data on a Big Picture

The file of hotel_bookings.csv contain the data of two different hotels in Portugal, one in the city and other in the surroundings.

There were 119,390 rows of data. Each row represents a single booking process with many different attributes. Some of them explain what kind of room they are booking, what sort of meal options they want, by when they are arriving, does the guest needs a parking lot, and many other. The entire table structure is fully explained in detail at the end of the report, as an appendix. Data collection dates from July 2015 up to August 2017.

Motive of the Data Analysis

The focus of this analysis is to discover what are the main drivers that increase ADR (Room revenue / # of Rooms Sold) on the hotel.

ADR is a very common ratio to study profit performance by room. ADR does not tell what is the net margin, neither it tell what is the breakeven point. ADR informs about how much money is a single booking process earning for the hotel. The higher it is, the more profitable it is.

Process on the General Analysis

First, I perform a general analysis on each attribute on the data. For that, I used MYSQL. The code along with the results are in a separate report called General_View

Second, I performed a more in-depth analysis on the factors that most contribute or affects to ADR. I've done this analysis also in MYSQL and the results are in a separate report called "Ways to increase the ratio ADR". This report is complementary to this one.

Last, I performed a graphical analysis on the main factors that has more important impact on the ADR ratio. Graphics can easily give a more accurate idea rather than numbers in a table.

This report shows the results in graphics, detail and factual data, and general conclusions on the subject.

Results

Through investigating all attributes. I've performed a more in-depth analysis over the most important factor that impact "ADR". These are:

- Room Type.
- Arrival Month of the Guest.
- Guest with Children and Babies or not.
- Special Requests
- Total Staying Nights
- Type of Meal

ADR Distribution in relation to different months of the year:

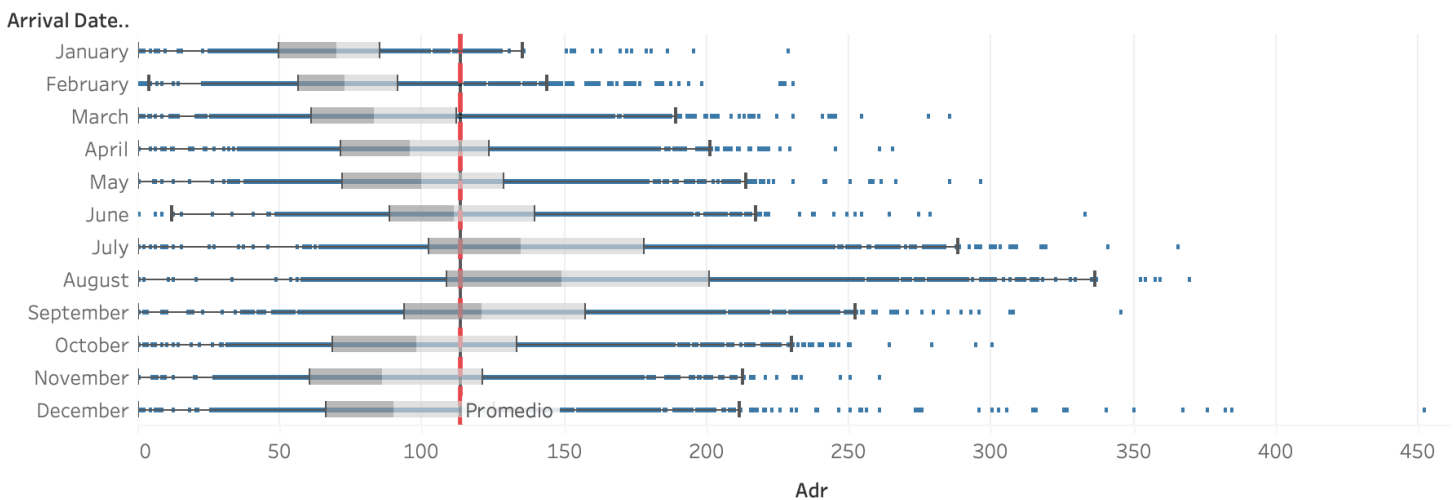
" ADR is way more profit favorable on the summer season, than in the rest of the year."

- Is in the months of June, July, August and September, when ADR is stronger. The median of ADR distribution in these months are stronger or at least equal to the total yearly average ADR.

These months represents the summer season. The strongest month is August with a median on ADR distribution of 148.8 euros against the yearly average of 113 euros.

- From October till May, the ADR decreases. The median on these months are considerably lower than the yearly average of 113 euros. In January, the weakest month, average ADR is 69.3 euros.
- (*Graphics remarks*). Below are the horizontal boxplots where it is showed in detail how average ADR is distributed in relation to months. The red line in the center represents the total yearly average ADR. I've only used data from the year 2016, because is the only complete time-frame dataset with all the months of the year. Some months in the year 2015 and 2017 are missing, and considering them would be leading to misrepresentation of the real behavior of ADR between seasons, especially because the performance on booking have been increasing year by year.

Average ADR & Arrival Month in the year 2016



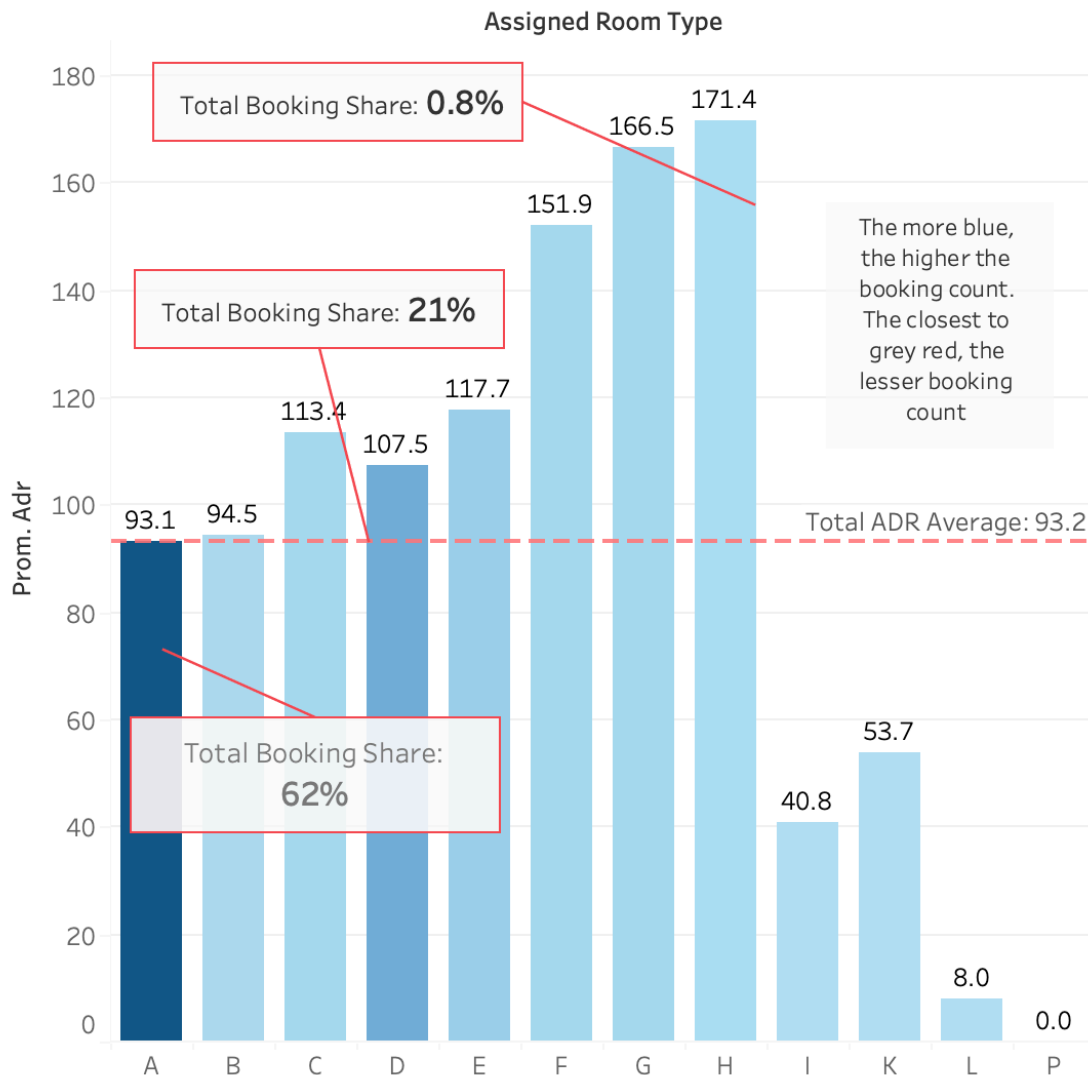
Impact of "Room Type" on ADR

"Room Type F, G and H are the ones that delivers a higher ADR. Nevertheless, they are also the less used by the guests. Instead, the room type A and D are the most booked, but they only offer an average ADR or relatively close to it."

"Total Average ADR across all type of rooms is the same as the average ADR of room type A, which is the most booked type of room."

- The total average of ADR of all room types is 93 euros.
- The most occupied Room Type is "A" with 62% of all bookings. But this type of room only gives an average ADR of 93 euros.
- The second most occupied type of room is "D" with 21% of all bookings. It has an average ADR of 107 euros.
- The room type that has given the highest ADR is room type "H", with an average of 171 euros. That is almost twice what the room type "A" gives. Nevertheless, only 712 bookings have asked for this type of room, that is only 0.8% of all bookings.
- Other type of rooms that gives top ADR are room type "F" and room type "G". Both types of room, together, just represents 5% of all bookings.
- (*Graphics remarks*). The quantity of bookings is in relation to the color. The most blue color is the type of room most booked, and the grey blue is the least. The thin dotted red line is the average ADR across all room types.

Average ADR (y-axis) & Assigned Room Type (x-axis)



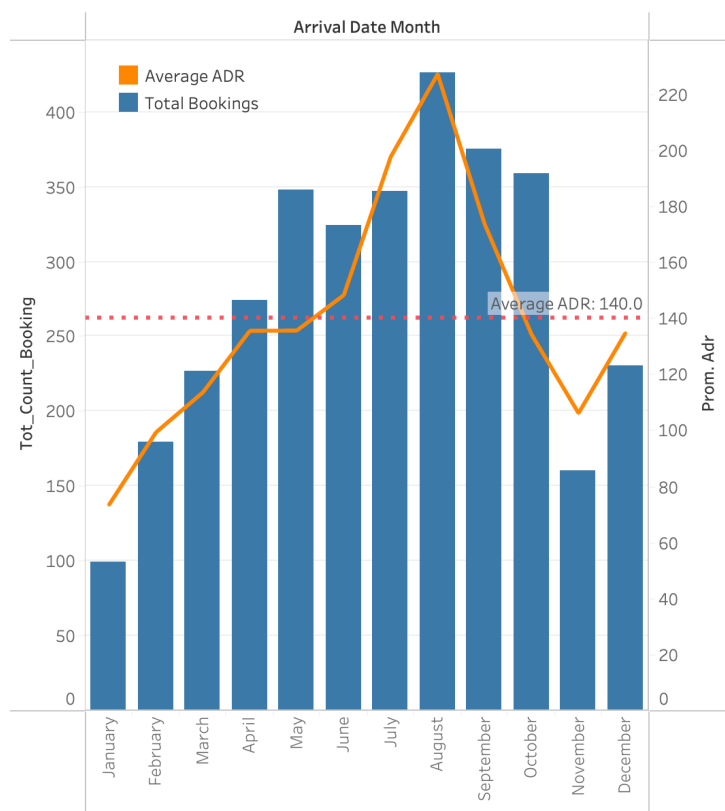
Impact of "Room Type" across the Months of the year on average ADR:

"The behavior on booking of certain type of rooms across the months of a year is very similar to the behavior of total booking in all rooms across the same months. Summer Season is when the peak of booking on any type of room is reached. But in slow months, like January or November or March, the bookings on more expensive room like type H, decreases more than any other regular type of room."

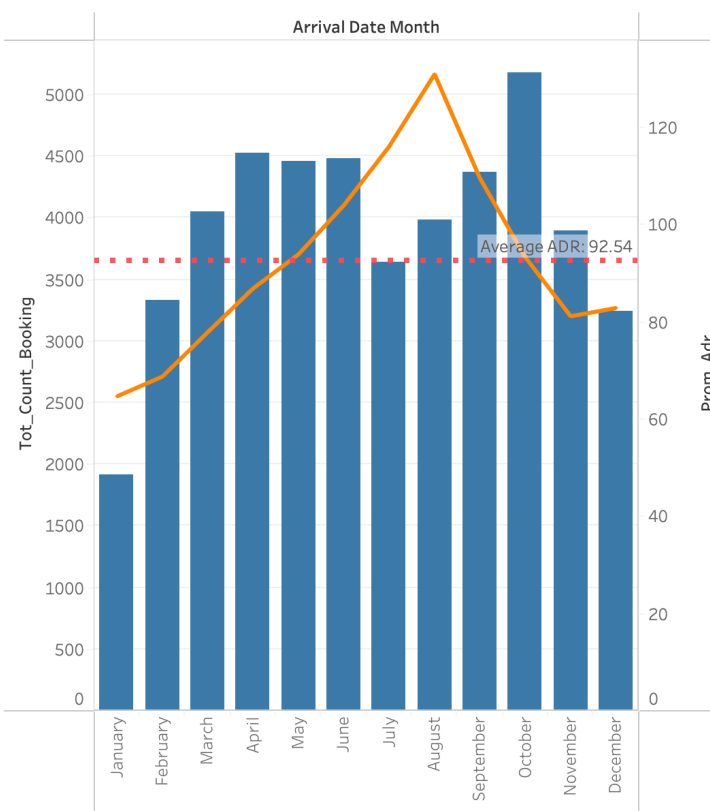
- The rooms type "F", "G" and "H" are more occupied through hot seasons. In the coldest and slowest month of January, the occupancy could decrease up to 4 times from the occupancy peak in August.
- The average ADR also drops to 73 euros from a peak of 227 euros in the summer season. That is almost 3 times the drop. Also, for this expensive type of room, the ADR fell below year average from October to May. This behavior on ADR would could signal that Hotels gives a much bigger discount on slow seasons rather than in any other types of room.

- The room type “A” and “D” are also more occupied through the summer season. The average ADR drops to 64 euros in slow seasons from a peak of 131 euros in the summer, that’s twice the drop. This could signal that discount on slow seasons are not as significant rather than room types “F”, “G” and “H” which average ADR fell 3 times.
- Average ADR on room type “A” and “D” remains above average from May to October.
- Yearly Average ADR on more expensive room like “F”, “G” and “H” remains at 140 euros. And for room type “A” and “D” remains at 92.5 euros.
- *(Graphics remarks).* The histogram on the left represents the booking (blue bars) of summed up rooms type F, G and H and the histogram on the right represents the booking of summed up room type A and D. The orange line is the average ADR of each month and the dotted red line is the yearly average ADR according to the selected room type. I have only taken in consideration the months in 2016 since the periods of 2015 and 2017 have missing months.

How Room Type F, G & H is occupied through months.



How Room Type A & D is occupied through months.



Impact of “Room Type” and Adult Guests with Children & Babies or without them:

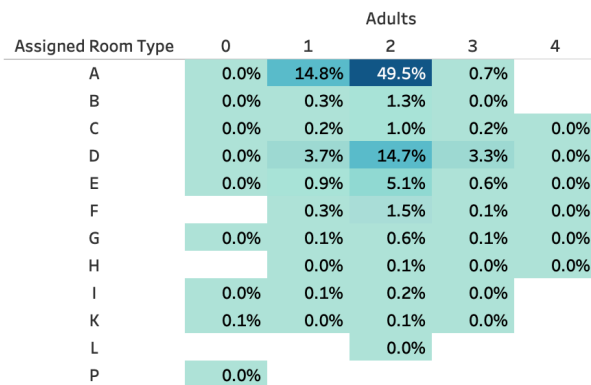
“Room Type F, G and H are commonly occupied by Guests with Children or Babies, rather than Room Type A and D which are 99% occupied by Guests with no Children or Babies at all”.

“Independently of the type of room booked. Guests with Children or Babies increases the average ADR more than Guests without them”

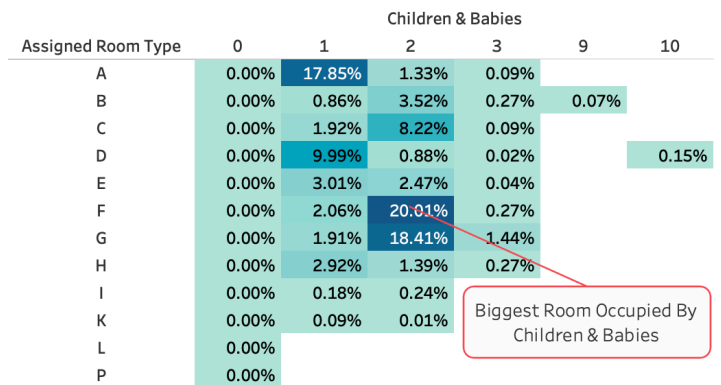
- 48% of all Booking with children or babies occupies room type “F”, “G” and “H”. In contrast, only 2.8% of all bookings for Adults with no children or babies chooses room type “F”, “G” and “H”.
- Room Types “A” and “D” are 99% occupied by adults with no children or babies. This percentage decreases to 56% when room type is “F”, 39% when room type is “G” and to 30% when room type is “H”.

- Guests with no Children or Babies that booked room type “F”, “G” and “H” aren’t booking those rooms because there’s a lack of rooms type “A” or “D”, especially in summer seasons where demand is high. The fifth graphic tells that Guests without children booked more times room type “F”, “G” and “H” in slow seasons rather than in summer season.
- Average ADR for Guests with Children or Babies (line orange) is always higher than the average ADR of Guests without Children or Babies (line Blue) choosing the same type of room.
- (*Graphics Remarks*). The last 4 graphics are split for Guests with Children or Babies (orange) and Guests without Children or Babies (blue). The percentages on the first two graphics shows representation on total booking for Guests without Children and in the next chart for Guests with Children or Babies.

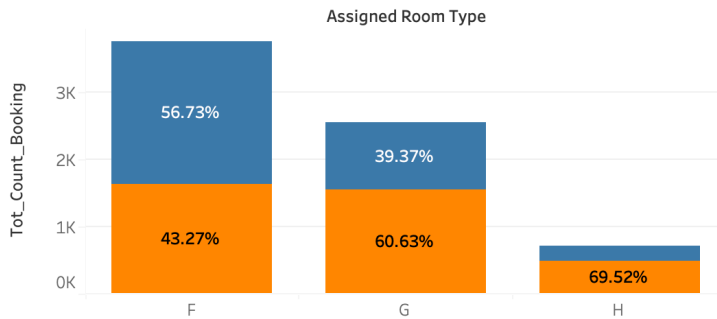
% of Adults in Room Type (No Children)



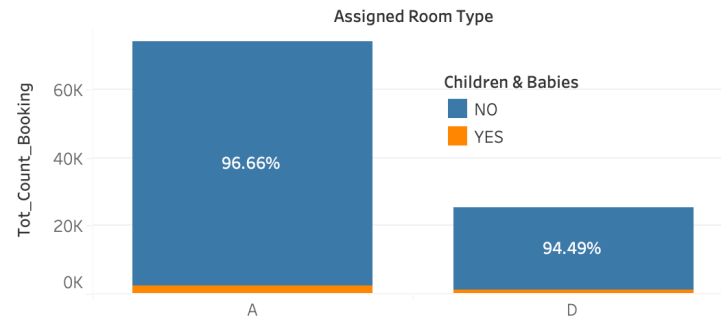
% of Children - Babies in Room Type



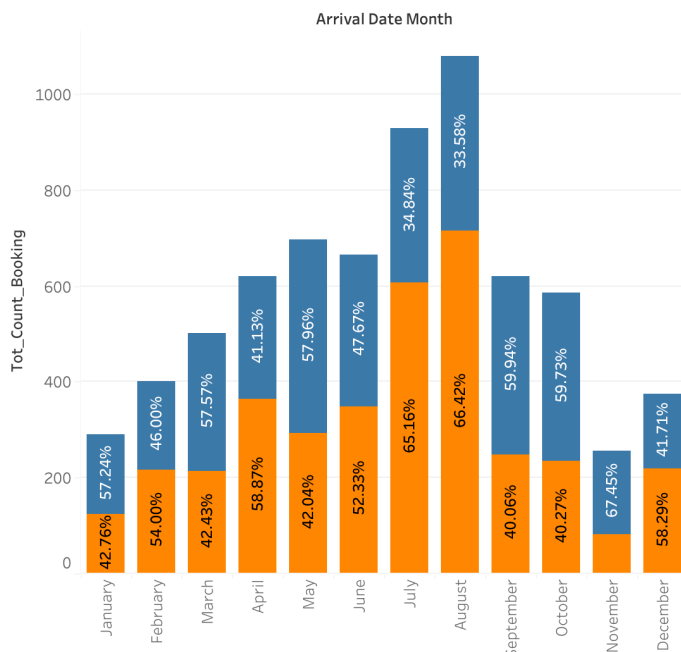
Room Type F, G and H & Children or Not



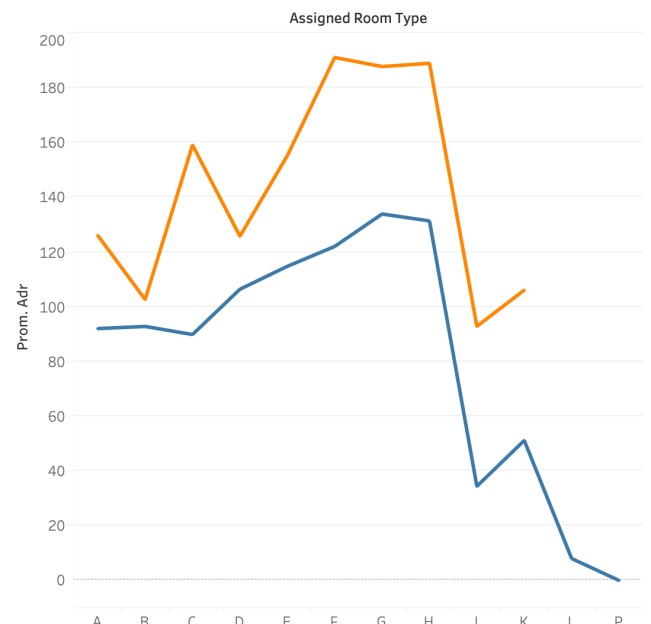
Room Type A and D & Children or Not



Room Type F, G, H & Guests across Months



Room Type VS Average ADR & Children or Not

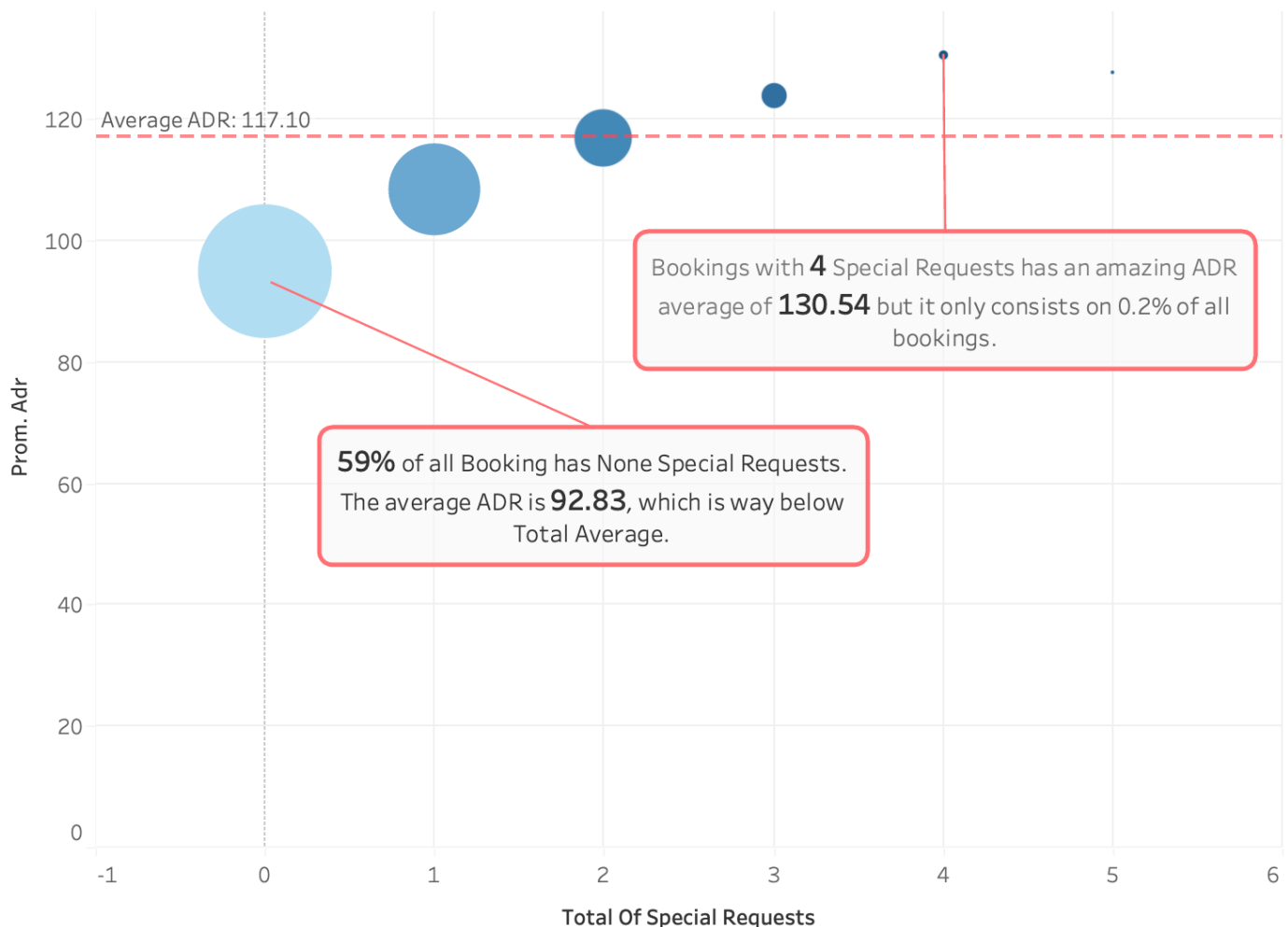


Impact of Special Requests on Average ADR:

“ The more Special Requests there are, the more increase has the average ADR. Nevertheless, Special Requests are not common.”

- In regards to Special Requests. 97% of all Bookings has a maximum of 2 Special Requests and these bookings remain at or below ADR Average of 117.10 euros.
- 59% of all bookings has no Special Requests at all and their average ADR is set at 92.83 euros.
- Only 2% of all bookings have 3 Special Requests with an ADR average of 123.85 euros. The remaining 1% of all bookings has 4 or more Special Requests with an ADR average around 130 euros.
- There's no relationship over the quantity of Special Requests and the number of Guests with babies or children. In fact, guests with no Children or Babies makes 89% of the total Special Requests.
- *(Graphics remarks.)* The size on the first graph is related to the sum of booking according to how many Special Requests have been made. The color light blue has the least average ADR and the darkest blue has the most average ADR.

Special Requests vs Average ADR and Booking Count



Children/Babies & Special Req

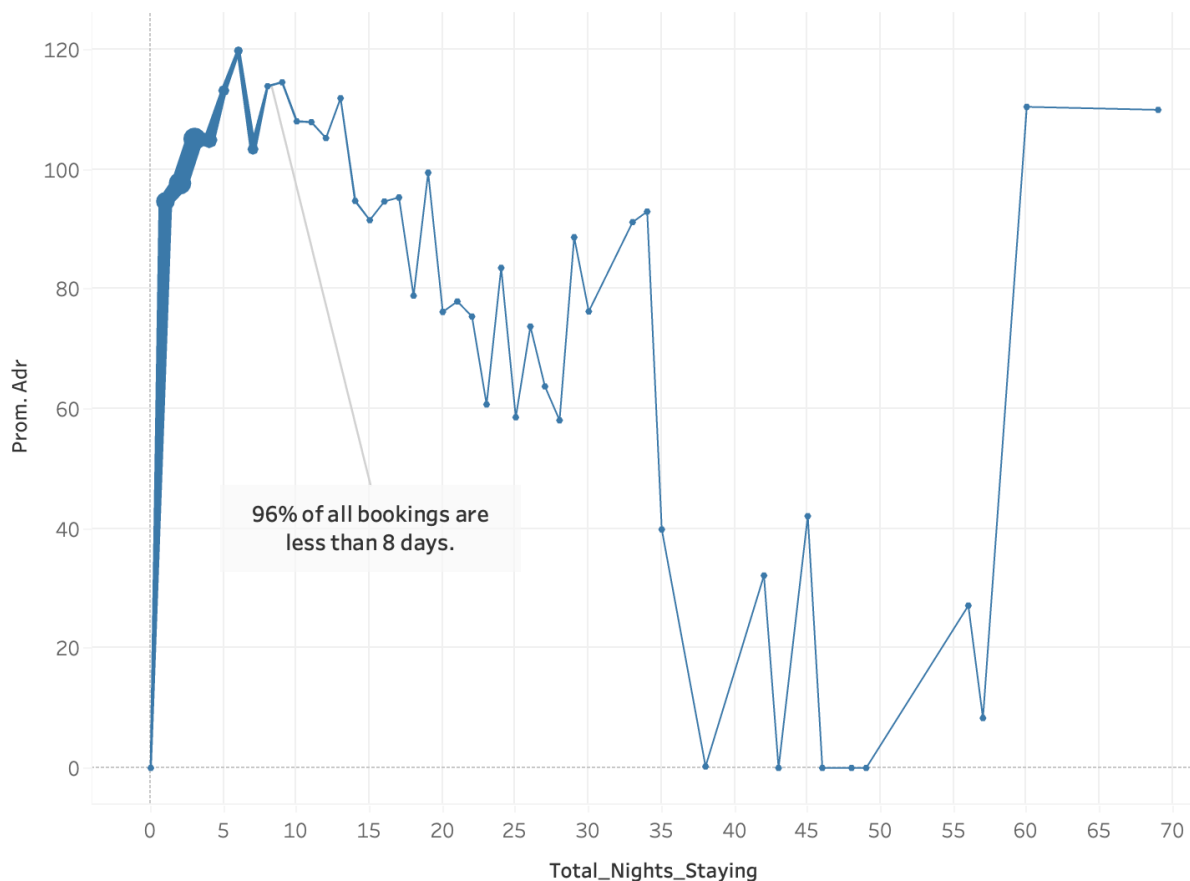
| Total Of Special Requests | Children & Babies | | | | |
|---------------------------|-------------------|-------|-------|-------|-------|
| | 0 | 1 | 2 | 3 | 10 |
| 1 | 61.71% | 3.65% | 2.27% | 0.07% | 0.00% |
| 2 | 22.59% | 2.70% | 1.09% | 0.04% | |
| 3 | 3.97% | 0.79% | 0.30% | 0.02% | |
| 4 | 0.56% | 0.10% | 0.04% | | |
| 5 | 0.07% | 0.01% | 0.01% | | |
| Total general | 88.89% | 7.25% | 3.71% | 0.13% | 0.00% |

Impact of Total of Nights Staying on Average ADR:

“ Total Number of Staying Nights has a slightly negative correlation with average ADR. This can be pushed by lowering tariffs for Guests who are willing to stay more than 8 nights in order to avoid idle capacity in rooms.”

- 96% of all bookings made has a maximum Night Staying of 8 nights.
- There's a downtrend on Average ADR when the total “Nights Staying” starts exceeding the 10 nights.
- Average ADR remains high above 100 euros when the total of “Nights” oscillates from 3 up to 13. For total Nights Staying above 2 weeks, the average ADR starts decreasing.
- (Graphics Remarks.) The line is thicker where there are more bookings made and is thinner where there are only a couple of hundreds bookings or dozens or fewer than that. The x-axis relates to Continuous Nights Staying and y-axis represents average ADR.

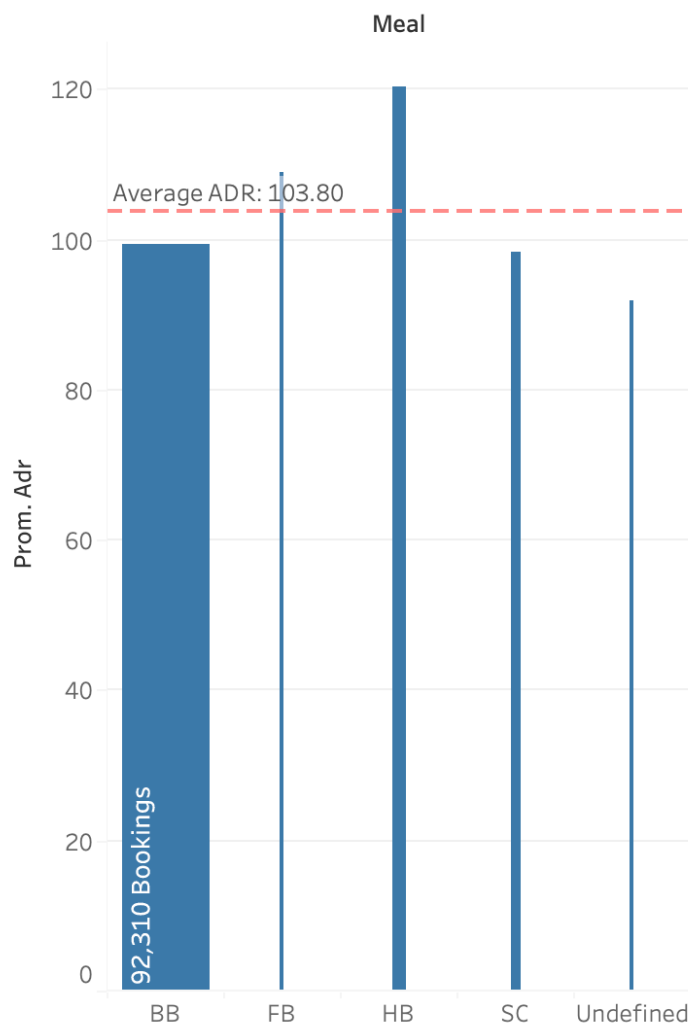
Staying Nights vs Average ADR & Booking Count



Impact of Meal Types on Average ADR:

- There isn't much oscillation on the average ADR in relation to different kind of meals. Obviously, there are differences, but not as notorious as other attributes.
- The meal "HB" which correspond to Half-Meal is what contributes the most with a higher ADR. It represents 12% of all bookings and has an average ADR of 120 euros.
- The meal "BB" which corresponds to Bed & Breakfast Meal is the most type of meal booked with a representation of 77% of all bookings. It delivers an average ADR of 99 euros which is slightly below total ADR average of 103 euros.
- (Graphics Remarks.) The graphic x-axis has the different type of Meals and the y-axis has a continuous Average ADR. The size of each bar is in direct relation to the total Booking according to each type of Meal.

Type of Meal VS Average ADR & Booking Count



Final Conclusions

- The attributes that affects the more and rises the average ADR, are booking rooms on the **summer season**, the increasing amount of **“Special Requests”**, the **type of the room** booked when is “F”, “G” or “H” and if welcomed Guests are bringing **children or babies**.
- **Meal** options don't have significant impact on ADR and the total number of **“Staying Nights”** seem to be positively impact the average ADR if a Guest is staying from 3 up to 13 nights. Staying for a period longer than 2 weeks has a negative impact on average ADR.

APPENDIX

Table Structure

| Column | Description |
|---------------------------|---|
| hotel | Hotel (H1 = Resort Hotel or H2 = 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 date |
| arrival_date_year | Year of arrival date |
| arrival_date_month | Month of arrival date |
| arrival_date_week_number | Week number of year for arrival date |
| arrival_date_day_of_month | Day of arrival date |
| stays_in_weekend_nights | Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel |
| stays_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: Undefined/SC – no meal package; BB – Bed & Breakfast; HB – Half board (breakfast and one other meal – usually dinner); FB – Full board (breakfast, lunch and dinner) |
| country | Country of origin. Categories are represented in the ISO 3155–3:2013 format |
| 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_cancelled | Number of previous bookings not cancelled by the customer prior to the current booking |
| reserved_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. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request. Code is presented instead of designation for anonymity reasons |
| 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. This variable can assume three categories: No Deposit – no deposit was made; Non Refund – a deposit was made in the value of the total stay cost; Refundable – a deposit was made with a value under the total cost of stay. |
| 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. ID is presented instead of designation for anonymity reasons |
| days_in_waiting_list | Number of days the booking was in the waiting list before it was confirmed to the customer |
| customer_type | Type of booking, assuming one of four categories: Contract - when the booking has an allotment or other type of contract associated to it; Group – when the booking is associated to a group; Transient – when the booking is not part of a group or contract, and is not associated to other transient booking; Transient-party – when the booking is transient, but is associated to at least other transient booking |
| 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 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 |
| reservation_status_date | Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to understand when was the booking canceled or when did the customer checked-out of the hotel |