

Hotel Booking Cancellations Analysis

Sacha Cohen & Loup Theuret

December 6, 2020

Wording A hotel chain wants to identify signs that can help foretell if guests are going to cancel their room bookings. By doing so, they can better anticipate the occurrence of such events and adjust their room prices and cancellation policies accordingly.

In a hotel chain, 32.8% of the clients cancel their reservation. Our purpose is to understand what is behind this number and advise some policies the hotel could implement to lower this cancellation rate.

1 Data set presentation

The data set is composed of features that can help us to identify the determinants of booking cancellation. After cleaning the data set, we started by conducting a univariate analysis. We tried to answer to the following questions :

- Is the booking made at a resort or city hotel ?
- When do people arrive (months in particular) ?
- How long are they staying ?
- What type of meal do they order ?
- From what country they come ?

In other words, we wanted first to learn more about the clients characteristics and the hotel features. Our 4576 bookings were made in two categories of hotels : city hotels (63%) and resort hotels (37%). The hotel welcomes customers from 90 countries, but the most frequent visitors come from Portugal.

Customers can have various booking types : associated to a contract (Contract), or a group (Group), not associated to anything (Transient) or transient and associated to at least another transient booking (Transient Party). There is a clear majority of Transient customers in our data set (almost 76%). To guarantee the booking, guests can : make no deposit, make a deposit in the value of the total stay cost or make a deposit with a value under the total cost of stay. The no-deposit option is chosen 92% of the time.

Over the year, the hotel chain has always more than 200 reservations per month but there is a peak above 500 in August. On average, over the whole week, clients stay 3 nights.

We can also wonder what are the specificities of the reservation.

In general, clients tend to book rooms A and D, which correspond to mid-range rooms (room types range from A to K). They can choose among four room services : Bed & Breakfast, Half Board (breakfast + one other meal), Full Board (breakfast + lunch + dinner) or nothing. The first option is the most selected one (76%). On the whole, guests are not very demanding : most of them do not ask for special requests and for the most part, they do not require any parking space (93%). The cost of the stay is between 0 and 382 euros. The price follows a normal distribution of mean 103 and standard deviation of 50.

Finally, the market segments are composed of 7 categories : online TA (51%), offline TA/TO (18%), Direct (11%), Group (13%), Complementary (4%), Corporate (0.6%) and Aviation (0.2%). The distribution channels are composed of : TA/TO (80%), Direct (15%), Corporate (5.4%) and GDS (0.17%).

2 What determines cancellation ?

To have a broad picture of the key determinants of cancellation, we studied the interaction of the data set variables with cancellation. We systematically compared the distribution of cancellation for each variable to the distribution of cancellation in the whole data set (32.8%). However, combining these features with the ones of a third category might sometimes be useful to understand their behaviour.

It was the first step of our analysis in order to divide our variables in three categories :

1. The ones that, alone, have no impact : the distribution of cancellation and non cancellation is the same as for the data set.
2. The ones that have a clear cut impact : we can directly draw conclusions in terms of policies ; no need for further analysis.
3. The ones for which we cannot manage to draw immediate conclusions in terms of policy for the hotel ; hence, we need further analysis.

As aforementioned, not all features have an explicit impact. For instance, booking at a City or at a Resort hotel does not change much the probability of cancelling. Another example is the meal variable. For each type of room service, we have around 33% of cancellation. Again, it is close to the distribution of the data set.

In contrast, some variables are obvious predictors of cancellation. It is the case for parking space and deposit type. Those who required one parking space never cancel and guests that chose the non-refund option cancel 99.4% of the time.

Special requests and booking changes are good indicators too. On one hand, whereas clients with no special request tend to cancel more, as the number of request increases, the probability to cancel decreases. On the other hand, the higher the number of booking changes, the lower the cancellation rate. Hence, for no special request the probability of cancellation is 40% (which is greater than 32.8%). But for one the probability is 23,75%, for two it is 23,19% ...

Last, if we pay attention to the arrival date month, we notice that the month with the highest cancellation rate is July (42%) and the lowest is November (23%).

Following this bi-variate study, we could decide which variables needed further explanations. The results for the number of nights are not straightforward. Indeed, the original distribution

of cancellation is respected for 3,4,5 and 7 nights (meaning around 32% of cancellation). Nonetheless, it is not the case for 1 (low rate) and 2 (high rate) nights for example. We also decided to deepen the analysis for customer types. Transient tend to cancel relatively more (36%) whereas Group, Contract and Transient-Party relatively less (9%, 29% and 19% respectively).

As we mentioned in the previous part, a significant number of customers comes from Portugal and we observe that it is the country with the highest cancellation rate. More precisely, Portuguese guests cancel as much as they do not. Except for Italy, the cancellation rate is below 30% for every country. Since the clientele is mainly Portuguese, we thought it would be enlightening to have a deeper understanding of what could explain this high rate.

With respect to assigned room type, a majority of guest books rooms A and D. Cancellations are numerous for the former and few for the latter compared to the whole data-set. We chose to give further explanations especially to compare it with reserved room type and with the average daily rate in what follows. The last two features are market segments and distribution channels. Here are the results for each market segment :

name	not cancel	cancel
Aviation	0.90	0.10
Complementary	0.83	0.17
Corporate	0.86	0.14
Direct	0.87	0.13
Groups	0.51	0.49
Offline TA/TO	0.73	0.27
Online TA	0.63	0.37

We see that the cancellation rate is relatively higher for Groups and Online TA and lower for the other market segments. For distribution channels, only TA/TO cancel relatively often.

In a nutshell, the further analysis required can be summarized by the following questions :

1. What is the impact of market segment, distribution channel, customer type.... on the number of nights ?
2. How are customer types, number of nights, market segment... distributed across the variables of the data set?
3. How is country related to the other variables ?
4. What is the impact of price on the variables yet to analyse ?

3 Multivariate Analysis

3.1 Average Daily Rate

We wanted first to learn more about the price. Especially its variation across time and across types of rooms. We interacted it with the assigned room types, which distribution is almost the same as reserved ones. As we can see on the graph (Figure 1), the most expensive rooms are G and H. Room A and D are mid-range types of rooms.

Then, we examined the evolution with respect to the month of arrival. We see that the price increases from January to August and then plummets. This trend is driven by the Resort Hotel. In contrast, City Hotel's prices are quite stable over time.

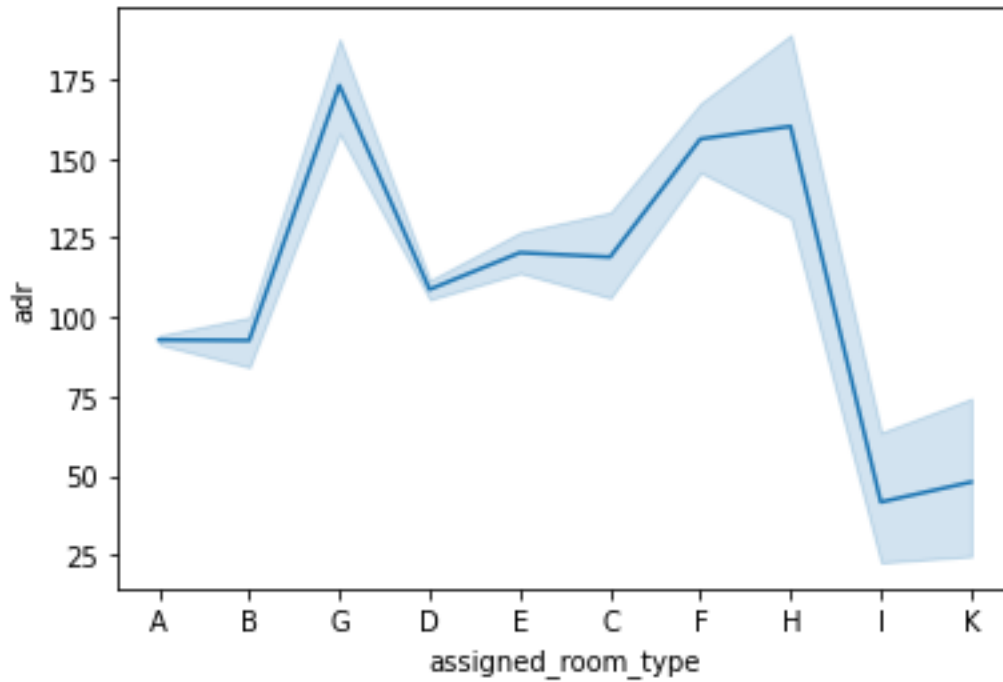


Figure 1: Average daily rate and Types of room

3.2 The number of nights

In City hotels, the average number of nights is 3, and more than 4 in Resort Hotels. The number of nights remains constant for the City hotels, whereas the total number of nights in Resort Hotels goes up from January to July and then tends to decrease afterwards. Furthermore, in these hotels, customers who spent the most nights are those who ordered Half Board meals (more than 5 nights on average).

Whereas Portugal is the country for which we have the highest number of customers, it is actually the country with the lowest number of nights spent on average. Ireland and Great Britain guests are the only ones that stay more than 4 nights on average. Considering both types of hotels, we observe the same pattern across countries. In addition, the analysis of the types of customers allows us to conclude that the customers of type Contract stay on average longer than the other (5 nights against 3).

When it comes to market segment, we find that Corporate, Complementary and Direct segments spend on average less nights than other segments, who spend 3 nights on average. Concerning distribution channel, those who reserve through the Corporate channel (and GDS) have a shorter stay than those who book through the other two channels (who spend a little more than 3 nights on average).

3.3 Customer type

The next variable of interest is the type of customer. All types have an increasing booking rate roughly from November to August. But whereas the number of Transient customers plummets in October and November, the number of Transient-Party and Contract ones increases slightly during these months. Contrary to Transient and Transient-Party customers, Contract customers (as well as Groups) reserve a lot more in Resort Hotels.

37% of Half Board meals requests come from Transient-Party customers, while they do not represent more than 20% of any other type of meal, and the proportion of them who reserve a A-type room is extremely high, compared to other types of customers. Finally, the percentage of them using Non-Refund schemes is low (2,4%) against other types (around 10%).

3.4 Country

If we look now closely to the origin country (with significant number of reservations), we can observe that, in the general increasing-price trend from November to August. Spain is the country which pays by far the most in summer and Portugal is the one which pays the least overall. It is also the country of origin for which the customers pay the lowest price in City Hotels (90 against 100 on average). The price is lower in all countries for Resort Hotels.

Portugal is also the country where the market segments Offline TA/TO, Direct, Groups and Corporate proportions are the highest. Corporate is solely represented by this country. If we focus on the type of customer, the proportion of Transient-Party is higher in Portugal and in Germany. Another noteworthy feature is that it is almost the only country of origin for which Non-Refund is used.

Finally, the vast majority of A-type rooms reservations comes from Portugal.

3.5 Market Segment

When we analyse the market segment, a first question we may want to answer to is : Are all the market segments paying on average the same price ? We recall that we have a large majority of online TA in the data set. These are the ones that pay the highest price on average (118) followed by Direct and Aviation (117 and 104 respectively). In addition, the lowest average price is paid by Complementary (5). It is a puzzling result because the price is extremely low. After verifying in the data set we found that many prices are set to 0. We assume that this category of market segment has some discounts.

Another question could be : In which type of hotel are they going to ? All market segments, except Groups, book in majority the City hotel. But we can note that for each hotel type, the main clientele comes from the Online TA market segment. If we now spot the rooms reserved, we remark that Groups and Offline TA/TO tend to book more room A than room D.

Only Online TA/TO and Groups choose the Non-refund option.

3.6 Distribution Channel

In our data set, not all distribution channels pay the same price. Actually, Corporate pay less during the year. They are also the ones, with Direct, that book more in the Resort Hotel (more than 50%). In contrast, GDS and TA/TO reserve in the City Hotel (more than 60%). In terms of rooms, A and D are still highly demanded for all the distribution channels.

Then, interacting the distribution channels with the market segments, we discover that Online TA accounts for a great part of Corporate, TA/TO, and Direct distribution channels.

cancel their reservations a lot (especially compared to British ones).

In addition, the increase of cancellations during summer as well as the high cancellation rate for A-type rooms are driven by this country.

For all other countries, it is noticeable that Transient-Party customers are almost not cancelling at all. These Transient-party customers prefer Half-Board meals, being the ones that order them the most. They also come more often in October and November.

We also managed to see that Resort Hotels were less expensive and that the customers who choose these hotels stay longer. But, no conclusion can be drawn from this interaction on the cancellation rate. Plus, we have no inference on cancellation considering the price customers pay.

One last remark to make is that customers whose assigned room type is different from their reserved have an extremely low rate of cancellation (5%). As this phenomenon concerns mainly changes from A to D, and given that D-type is more expensive than A-type, people who are "upgraded" cancel a lot less.

5.2 Policy recommendations

As a consequence of these conclusions, we can recommend the following policies :

1. Advertise the availability of car-parking spaces and the possibility of granting special requests.
2. If possible, remove the non-refund deposit type.
3. Insist on the possibility of Half Board meals to attract the Transient-Party customers (in autumn especially).
4. If not too costly, propose "upgrading" room types.
5. Promote the hotel in other countries than Portugal.
6. For Portuguese customers, favour distribution channels other than TA/TO and target Direct or Corporate market segments.