Causal Inference: Hotel Booking Cancellations in Portugal

Ming Zhao

Abstract

This project investigated the relationship between assigning a different room type from the one reserved and booking cancellations in the hotel industry. Contrary to the hypothesis, the study revealed that discrepancies in room types did not cause cancellations; instead, it decreased the likelihood of cancellations. The results were obtained through rigorous matching techniques. Additionally, sensitivity analysis further confirmed the robustness of the results by addressing potential hidden biases. These findings contribute to the understanding of factors influencing booking cancellations and have implications for hotel management.

1. Introduction

Booking cancellations can be a significant issue for hotels because cancellations directly impact hotels' revenue. What drives customers to cancel their reservations? From the customer's perspective, changed plans, unforeseen circumstances, better deals, and unmet expectations are the primary factors. The former two factors are beyond hotels' control, and they apply equally to every hotel. However, for the latter two factors, hotels are likely to find ways to address them. For instance, they can provide competitive discounts frequently and avoid unnecessary discrepancies between requests and assignments. Nevertheless, what happens when the discrepancies are inevitable due to restricted resources? Does these have an impact on booking cancellations, and if so, to what extent? This project focuses on the discrepancies in room types and aims to determine whether assigning a different room from what was reserved leads to a booking cancellation.

2. Dataset Description

To investigate the impact of assigning a different room on booking cancellations, the Hotel Booking Demand Dataset was utilized. The dataset consists of 32 variables and 119,390 observations, providing booking information for two real hotels located in Portugal from October 2014 to September 2017. The booking information includes customers' profiles, requests, reservations, and arrivals, as well as hotel features and management. All the data are shared by the hotels and obtained from their Property Manage System databases. (For an overview of the variables in the dataset, see Table A in the Appendix, which contains a detailed description of each variable.)

3. Proposed Methods

To determine whether assigning a different room type leads to booking cancellations, an ideal study would be a randomized experiment. In such an experiment, customers would be randomly assigned to either the treatment group, which receives the room they reserved, or the control group, which receives a different room type. By comparing the cancellation rates between the two groups, we could obtain a clear understanding of the impact of room discrepancies on booking cancellations. However, conducting a randomized experiment in this context is not feasible due to potential high costs and risks for hotels, such as revenue loss and damage to reputation. Give these circumstances, we must rely on observational data, which introduce confounders and biases that can impact the relationship between the treatment and the outcome. To establish a causal relationship, quasi-experimental methods are necessary. These methods involve carefully selecting control subjects to control for the covariates between the treatment and control groups, aiming to minimize hidden biases.

The proposed method for this project is multivariate matching. Firstly, a logistic regression model is used to estimate the propensity score, which represents the probability of being assigned a different room type. The observed covariates are included as predictors in the regression model. In multivariate matching, the goal is to find matched sets of subjects who are close in terms of their propensity score and covariates. To achieve this, the robust Mahalanobis distance is employed, which adjusts the covariance matrix using ranks. This approach helps to reduce the influence of outliers and prevents heavily correlated covariates from having excessive influence. The robust Mahalanobis distance, to some extent, assigns equal weight to all covariates, regardless of whether they are difficult to balance. However, the propensity score is primarily affected by covariates that differ the most between the treated and control groups or are difficult to balance. To emphasize these covariates, a propensity score caliper is added to the distance. If the difference in propensity scores between two individuals exceeds a certain threshold, a penalty is given to the distance between them; otherwise, the distance uses the robust Mahalanobis distance.

To pair each treated subject with one or more control subjects, optimal matching is applied. Optimal matching minimizes the total distance with matched sets, ensuring the best possible balance between the treated and control groups. These matching techniques address unobserved confounders, enhancing the robustness of the treatment effect. Additionally, sensitivity analysis is conducted to test the assumption of being free of biases and further validate the robustness of the treatment effect estimation.

4. Data Analysis

4.1 Analytical Sample

This project focuses on analyzing the bookings of customers from Portugal (n = 48,590) and bookings in the year 2016 (n = 56,707), resulting in a final sample size of 22,321. Seven variables are excluded from the analysis: variables related to week and date (`reservation_status_date`, `arrival_date_day_of_month`, `arrival_date_week_number`), variables overlapping with other variables (`reservation_status`, `distribution_channel`), and variables containing ID information (`agent`, `company`). As a result, there are no missing data in the dataset.

Two new variables are generated by summarizing the existing variables. The first variable represents the total number of stayed nights (`total_stays` = `stays_in_week_nights` + `stays_in_weekend_nights`). The second variable represents the total number of guests (`guests` = `adults` + `children` + `babies`). Furthermore, a new variable is created to indicate whether the hotel assigned a different room type compared to the one reserved. It takes “yes” if the reserved room type (`reserved_room_type`) differs from the assigned room type (`assigned_room_type`), and "no" otherwise.

Overall, this analysis contains 19 variables, including four binary variables, five categorical variables, and ten numeric variables. Of these variables, 17 are covariates. To incorporate categorical covariates into this analysis, they are converted into dummy variables, resulting in a total of 38 covariates.

4.2 Descriptive Statistics

The variables of interest in this analysis are the treatment and the outcome. The treatment refers to if a different type of room was assigned as compared to what the customer had reserved, while the outcome pertains to if the booking was cancelled. Looking at the contingency table for the two variables, it is found that more than half of bookings were cancelled, yet only 2.4% of the cancelled bookings had a different room type assigned. Additionally, 13.2% of the bookings were assigned a different room, and of these bookings, 10.0% were cancelled.

In summary, first, the overall cancellation rate was relatively high, indicating that there are likely other primary factors contributing to cancellations aside from the assignment of a different room. Second, the rate of assigning a different room was not very high and the cancellation rate following the assignment of a different room was also not as high as initially expected. Furthermore, there is no evidence to suggest that the cancellations after a different room was assigned were due to the room discrepancies.

Based on these observations, it is necessary to conduct a more rigorous examination of the treatment effect. To achieve this, a technique related to experimental design is required, which can provide a better understanding of the causal relationship between assigning a different room and booking cancellations. By utilizing an appropriate experimental design, more conclusive insights on the impact of the room discrepancies can be obtained.

Table 1. Contingency Table for Outcome and Treatment

		Cancelled			
		Yes	No	**Total**	Prob-Vector
Different Room	Yes	295	2651	2946	(0.100, 0.900)
	No	12110	7265	19375	(0.625, 0.375)
	Total	12405	9916	22321	(0.556, 0.444)
	Prob-Vector	(0.024, 0.976)	(0.733, 0.267)	(0.132, 0.868)	

4.3 Preliminary Inference about Treatment Effect

Before involving any experimental design, or by simply assuming that it was a randomized experiment, I conducted Fisher's exact test. My hypothesis was that the relationship between the treatment and the outcome is positive, meaning that assigning a different room by the hotel would increase booking cancellations. However, the test showed that the effect was not positive (p-value = 1). Subsequently, I used logistic regression to examine the treatment effect and discovered that the effect was negative (coefficient = -2.707). Since this is not a randomized experiment but rather an observational study, the exploration on the relationship requires the use of a quasi-experimental method.

4.4 Raw Comparison of Treated and Control Groups

In an observational study, it is expected that there will be imbalances in the characteristics between the treatment group and the control group. Before further investigation of the treatment effect, I assessed the comparability of the treated and control groups using t-tests. Table 2 presents the results of the comparison for each covariate. As expected, the treated and control groups significantly differ for most of the covariates.

Table 2. Baseline Comparison between Treated and Control Groups

	Covariate Mean		
Covariate	Treatment Group	Control Group	p-value
resort_hotel	0.613	0.335	< 0.01
is_repeated_guest	0.175	0.053	< 0.01
previous_cancellations	0.046	0.114	< 0.01
previous_bookings_not_canceled	0.696	0.297	< 0.01
lead_time	37.415	132.292	< 0.01
total_stays	1.926	2.949	< 0.01
guests	1.694	1.862	< 0.01
booking_changes	0.345	0.141	< 0.01
days_in_waiting_list	1.526	7.785	< 0.01
adr	68.940	92.376	< 0.01
required_car_parking_spaces	0.145	0.058	< 0.01
total_of_special_requests	0.512	0.326	< 0.01
arrival_date_month_February	0.140	0.087	< 0.01
arrival_date_month_March	0.125	0.088	< 0.01
arrival_date_month_April	0.096	0.106	0.087
arrival_date_month_May	0.082	0.105	< 0.01
arrival_date_month_June	0.071	0.115	< 0.01
arrival_date_month_July	0.052	0.073	< 0.01
arrival_date_month_August	0.033	0.068	< 0.01
arrival_date_month_September	0.054	0.089	< 0.01
arrival_date_month_October	0.068	0.095	< 0.01
arrival_date_month_November	0.096	0.074	< 0.01
arrival_date_month_December	0.079	0.056	< 0.01
meal_Undefined	0.004	0.016	< 0.01
meal_SC	0.037	0.035	0.714

meal_HB	0.077	0.120	< 0.01
meal_FB	0.007	0.012	< 0.01
market_segment_Corporate	0.200	0.074	< 0.01
market_segment_Direct	0.210	0.105	< 0.01
market_segment_Groups	0.058	0.270	< 0.01
market_segment_Offline.TA.TO	0.186	0.288	< 0.01
market_segment_Online.TA	0.307	0.250	< 0.01
market_segment_Others	0.002	0.003	0.243
deposit_type_Non.Refund	0.009	0.343	< 0.01
deposit_type_Refundable	0.004	0.001	< 0.01
customer_type_Group	0.008	0.002	< 0.01
customer_type_Transient	0.776	0.785	0.295
customer_type_Transient.Party	0.211	0.205	0.401

4.5 Multivariate Matching and Examining Balance

Under the assumption that no unobserved confounders exist, I produced matched pairs to balance the observed covariates between the treatment and control groups. The matching method employed is the robust Mahalanobis distance with a propensity score caliper. The treatment and control groups consist of 2,946 and 19,375 observations, respectively. Therefore, first, I implemented one treated subject being matched with one control subject (1:1), then one treated subject being matched with two control subjects(1:2). Finally, one treated subject being matched with three control subjects (1:3) was attempted. I stopped further matching at this point because the 1:3 matching fails to adequately balance the covariates between the treatment and control groups.

Specifically, to assess the balance, I used standardized mean differences (SMD), which measure the mean difference between treatment and control groups in units of standard deviation. A balance is considered acceptable if the absolute SMD is less than 0.2. After the initial matching for the 1:1 pairs, the highest SMD obtained was 0.131 for `booking_changes`. To improve the matching, I implemented an asymmetric penalty for this variable, following the approach suggested by Yu and Rosenbaum (2019). As a result, the SMD for `booking_changes` decreased to 0.011, while the balance on other variables remained acceptable. The Figure 1 visualizes the changes in SMDs from pre-matching to 1:1 matching for each covariate.

After the initial matching for the 1:2 pairs, the largest SMDs are 0.244, -0.175, and 0.122 for the variables, `resort_hotel`, `adr`, and `is_repeated_guest`, respectively. To address this imbalance, asymmetric calipers were added for these three variables. Consequently, although the balance on these variables improved, some other variables become imbalanced. To adjust the newly imbalanced variables, asymmetric calipers were further added for ` total_stays` and ` market_segment_Corporate`. With these adjustments, the overall balance across covariates reached an acceptable level. Figure 2 provides a visual representation of the changes in SMDs from pre-matching to 1:2 matching.

After the initial matching for the 1:3 pairs, the same three variables, as mentioned above in the 1:2 matching, had absolute SMDs greater than 0.2. To address this imbalance, asymmetric calipers were once again applied for these variables. However, their balance remained unacceptable and the balance on other variables fluctuated. I thus attempted to indtroduce more asymmetric calipers for additional variables, but it turned out achieving balance across all the covariate simultaneously was not possible. One potential reason for this is the insufficient number of observations in certain categories. As a consequence, the 1:3 matching was unsuccessful and considering the circumstances, further matching seemed unreasonable.

Figure 1. Love Plot for 1:1 Matching

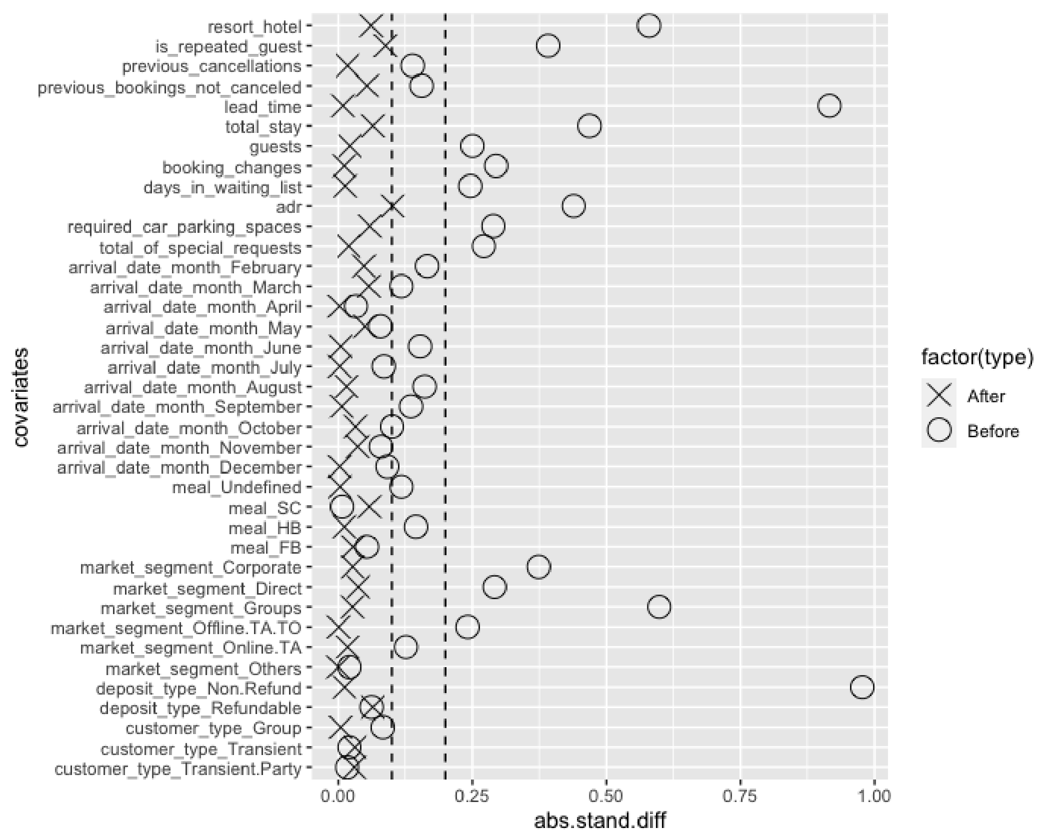
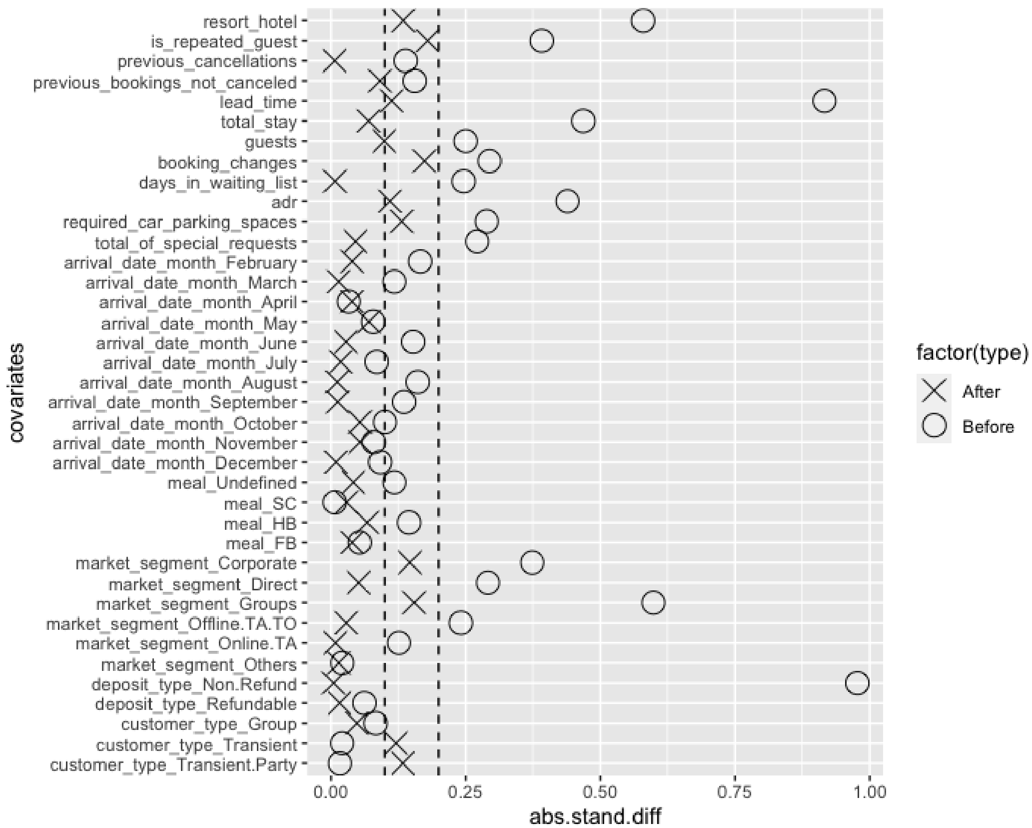


Figure 2. Love Plot for 1:2 Matching



4.6 Inference about Treatment after Matching

Since both 1:1 and 1:2 matching results achieved an acceptable level, I then employed Fisher's exact test and logistic regression to re-examine the treatment effect. Based on the 1:1 matched pairs, the test showed that the relationship was not positive (p-value = 1) and the regression revealed that the effect was negative (coefficient = -1.028). Similarly, based on the 1:2 matched pairs, the results are consistent, but the effect size obtained from logistic regression increased (coefficient = -1.455). Although the results agreed on the direction of the effect, the effect size for the 1:1 matching data appeared more convincing because of its smaller overall standardized differences. As a result, the matching process concluded with the 1:1 matching, which achieved a better balance in the observed covariates.

4.7 Sensitivity Analysis for Matched Pairs

To evaluate the assumption of no unmeasured confounding or the robustness of the inferenced treatment effect, I performed a sensitivity analysis on the 1:1 matched pairs. (The contingency table for the data can be found in Table B in the Appendix). Specifically, McNemar's test was performed, where the test statistic is the number of discordant pairs in which the treated unit has the outcome. Upon examining the matched data, out of the 2944 pairs (two subjects were removed due to the lack of overlap), 738 pairs had only one subject that cancelled its booking. Among these discordant pairs, 167 pairs in which the customer had been assigned a different room subsequently cancelled their booking, while 571 pairs in which the customer had been assigned the same room but still cancelled their booking.

The sensitivity analysis using McNemar’s test statistics yielded a sensitivity parameter of one. It suggests that the study is free of hidden bias, indicating that within each pair, there is an equal likelihood for each subject to be treated. This result supports the assumption of no unmeasured confounding and enhance the robustness of the inferred treatment effect.

5. Conclusion and Discussion

This project examined the impact of assigning a different room type from the one reserved on booking cancellations in the hotel industry. Through rigorous matching techniques and data analysis, the study found that discrepancies in room types led to a decrease in cancellations rather than an increase. Specifically, I employed the robust Mahalanobis distance with a propensity score caliper to balance observed covariates between the treatment and control groups. The 1:1 and 1:2 matching results achieved an acceptable balance, with the overall standardized mean differences favoring the 1:1 matching. Contrary to the initial hypothesis, both Fisher’s exact test and logistic regression indicated that the treatment effect was not positive. Furthermore, the results demonstrated robustness when a sensitivity analysis for McNemar's test statistics was conducted to account for hidden biases.

The possible reasons for the observed negative treatment effect include, first, the customers being assigned a better room. This improved room type may have increased customer satisfaction and reduced the desire to cancel the booking. Second, the assignment of a different room may also indicate effective communication and personalized service, creating a positive perception of the hotel and decreasing the likelihood of cancellations. Third, it is possible that customers themselves requested a different room type and had their needs met could contribute to the decreased cancellations. However, it is important to note that these explanations are speculative. Additional information and further research will be necessary for a more comprehensive understanding of this observed negative treatment effect.

6. References

R. Yu and P. R. Rosenbaum. Directional penalties for optimal matching in observational studies. Biometrics, 75(4):1380-1390, 2019.

N. Antonio, A. d. Almeida, and L. Nunes. Hotel booking demand datasets. Data in Brief, 22:41-49, 2019.

7. Appendix

Table A. Variables Description

Variable	Type	Description
hotel	binary	Hotel type: resort hotel or city hotel
is_canceled	binary	If the booking was canceled
is_repeated_guest	binary	If the booking name was from a repeated guest
previous_cancellations	binary	Number of previous bookings that were cancelled by the customer prior to the current booking
previous_bookings_not_canceled	binary	Number of previous bookings not cancelled by the customer prior to the current booking
lead_time	numeric	Number of days between booking date and arrival date
stays_in_weekend_nights	numeric	Number of weekend nights the guest stayed or booked
stays_in_week_nights	numeric	Number of week nights the guest stayed or booked
adults	numeric	Number of adults
children	numeric	Number of children
babies	numeric	Number of babies
booking_changes	numeric	Number of changes made to the booking from the moment of booking to the moment of check-in or cancellation
days_in_waiting_list	numeric	Number of days the booking was in the waiting list before it was confirmed
adr	numeric	Average daily rate
required_car_parking_spaces	numeric	Number of car parking spaces required
total_of_special_requests	numeric	Number of special requests
arrival_date_year	categorical	Year of arrival date
arrival_date_month	categorical	Month of arrival date
meal	categorical	Type of meal booked: Undefined, SC, BB, HB, or FB
country	categorical	Country of origin
market_segment	categorical	Market segment designation: complementary, corporate, direct, groups, offline TA/TO, online TA, or others
reserved_room_type	categorical	Code for the room type reserved: A, B, C, D, E, F, G, H, L, or P
assigned_room_type	categorical	Code for the room type assigned to the booking: A, B, C, D, E, F, G, H, L, or P
deposit_type	categorical	Deposit type: no deposit, non refund, or refundable

customer_type	categorical	Type of booking: contract, group, transient, or transient-party
agent	categorical	ID of the travel agency that made the booking
company	categorical	ID of the company that made the booking
reservation_status	categorical	Last status of reservation: canceled, check-out, no-show
reservation_status_date	categorical	Date at which the last status of reservation was set
arrival_date_week_number	categorical	Week number of arrival date
arrival_date_day_of_month	categorical	Day of arrival date
distribution_channel	categorical	Booking distribution channel: corporate, direct, GDS, TA/TO, undefined

Table B. Contingency Table for Outcome and Treatment after 1:1 Matching

		Cancelled			
		Yes	No	**Total**	Prob-Vector
Different Room	Yes	295	2649	2944	(0.100, 0.900)
	No	699	2245	2944	(0.237, 0.763)
	Total	994	4894	5888	(0.169, 0.831)
	Prob-Vector	(0.297, 0.703)	(0.541, 0.459)	(0.5, 0.5)	