Report

Fahim Ahmed

06-01-2026

Title: Hotel Booking Insights for Revenue Growth and Cancellation Reduction

Business Task: Management wants to increase revenue and reduce cancellations for the upcoming year

Introduction

The dataset used in this project is from Kaggle, https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand.

The business task is about a hotel business that wants to increase revenue and reduce cancellations using data to make a data-driven decision.

The tool used is SQL Server because the dataset is rather large for a spreadsheet project.

The strategy is to use SQL queries to analyse booking trends, and revenue drivers by hotel type, market segment, and lead time.

The dataset is composed of 32 columns and 119,391 columns.

The dataset does not contain any customer identifiable data, and ADR is used as a proxy for booking value.

Data preparation and cleaning

Some early tables alterations are necessary. There are null values that are not possible to update because NULL is recorded as a string instead of a null value. So, the tables ‘agents’ and ‘children’ was set to NULL from the string null. There were cases where the string was ‘NA’, which has also been handled the same way.

There are no duplicated values in the dataset, and the final number of rows is 119,390.

Analysis

Task 1: Cancellation rates by customer type, market segment and lead time

For this task, the question is what characteristics are associated with higher cancellation rates.

The metrics to consider here include the customer type, market segment and the lead time buckets.

In this dataset, there are 4 types of customers, which include transient (meaning individual travelers), transient-party (meaning a small group), contract (meaning a booking via long-term contracts, usually from companies and agencies), and group (meaning a large group booking). This matters because contract customer often have penalties from agreements, group bookings are rather volatile when it comes to confirmation, and transient bookings are price-sensitive and flexible.

When it comes to lead times, shorter lead times indicate an urgent travel, and longer lead times indicate speculative planning. So, dividing the lead times in specific patterns helps understand how customers would behave. Longer lead times tend to be unpredictable because likely customers are going to cancel.

There are 8 market segments, which are Online TA (travel agency), Direct, Corporate, Groups, Complementary, Offline TA/TO, Aviation, and Undefined. The difference between these are that Online TAs customers shop aggressively due to its ease of booking process, Direct customers have brand loyalty, Corporate bookings have fixed schedules, Offline TA/TO are just traditional travel agencies, and Aviation refers to airline crew stays.

In this task, segments with the highest cancellation rates were considered. From the results, Transient and Transient-party customers bookings exhibit higher cancellation rates compared to other customer types. Contract bookings also exhibit elevated cancellation rates, likely because their longer booking lead times. The results also show that longer lead times may be a factor for higher cancellation rates, especially when the booking was made at least 30 days earlier. The reason why transient and transient-parties have higher cancellation rates is because these customers are booking for leisure and are more price-sensitive. Contract booking appearing in the results is interesting because these bookings are made very far in advance and cancellations would be made because of operations, not related to behaviour. This does not mean that contract bookings are bad, and it's not relevant to the task at hand.

	customer_type	market_segment	lead_time_bucket	total_bookings	canceled_bookings	cancellation_rate
1	Contract	Groups	91+ days	727	705	0.97
2	Transient	Groups	91+ days	6583	6384	0.97
3	Transient	Groups	31-90 days	1489	1445	0.97
4	Transient	Groups	8-30 days	223	188	0.84
5	Transient	Corporate	91+ days	105	73	0.70
6	Transient	Offline TA/TO	91+ days	7420	4407	0.59
7	Transient	Corporate	31-90 days	387	218	0.56
8	Contract	Online TA	91+ days	193	100	0.52
9	Transient	Online TA	91+ days	18759	9505	0.51
10	Transient-Party	Corporate	91+ days	204	100	0.49
11	Transient	Online TA	31-90 days	15259	6333	0.42
12	Transient-Party	Groups	91+ days	6311	2291	0.36
13	Transient	Groups	0-7 days	132	47	0.36
14	Transient	Online TA	8-30 days	9885	3325	0.34
15	Contract	Online TA	31-90 days	379	121	0.32
16	Transient	Offline TA/TO	31-90 days	3510	1104	0.31
17	Transient-Party	Offline TA/TO	31-90 days	1941	562	0.29
18	Contract	Online TA	8-30 days	504	132	0.26
19	Transient-Party	Offline TA/TO	91+ days	4960	1308	0.26
20	Transient-Party	Groups	8-30 days	1327	347	0.26
21	Transient-Party	Groups	31-90 days	2449	621	0.25
22	Transient-Party	Corporate	31-90 days	479	122	0.25
23	Transient	Direct	91+ days	2290	521	0.23
24	Transient-Party	Direct	31-90 days	238	51	0.21

Figure 1, output query for the first task

The strategy to reduce cancellations would be to apply stricter cancellation policies for long lead time bookings, introduce partial prepayment for transient bookings, and encourage shorter lead time bookings via pricing incentives.

Task 2: Revenue maximization

The objective of this task is to identify hotel types, market segments, and months that underperform in booking value in order to highlight opportunities for pricing or promotional optimization. The dataset does not contain the actual revenue, however, 'adr' (Average Daily Rate) was used as a proxy to estimate booking value.

The estimated booking value was calculated as ADR multiplied by the total length of stay, both weekdays and weekend nights, using non-canceled bookings only.

The steps used include:

- Calculating the average ADR and average estimated booking value for each hotel type to establish a performance baseline
- Analye booking value across market segments within each hotel type to identify channels that consistently underperform
- Evaluated monthly booking value patterns to identify underperforming months where pricing could be improved
- Segment-month performance was compared against hotel-wide average

	hotel	market_segment	arrival_date_month	total_bookings	avg_estimated_booking_value_in_euros	hotel_avg_booking_value_in_euros	relative_performance_vs_hotel_avg
1	Resort Hotel	Complementary	August	1	4.00	410.40	-0.99
2	Resort Hotel	Complementary	June	4	8.00	410.40	-0.98
3	City Hotel	Complementary	February	3	7.00	318.82	-0.98
4	City Hotel	Complementary	October	1	6.00	318.82	-0.98
5	Resort Hotel	Complementary	October	3	18.00	410.40	-0.96
6	City Hotel	Complementary	January	3	16.00	318.82	-0.95
7	City Hotel	Complementary	April	1	18.00	318.82	-0.94
8	Resort Hotel	Complementary	May	1	26.00	410.40	-0.94
9	Resort Hotel	Complementary	December	6	41.25	410.40	-0.90
10	Resort Hotel	Corporate	December	129	48.16	410.40	-0.88
11	Resort Hotel	Corporate	January	175	54.09	410.40	-0.87
12	Resort Hotel	Complementary	April	1	70.50	410.40	-0.83
13	City Hotel	Complementary	December	2	54.00	318.82	-0.83
14	Resort Hotel	Corporate	February	253	84.34	410.40	-0.79
15	Resort Hotel	Corporate	April	188	91.24	410.40	-0.78
16	Resort Hotel	Complementary	November	4	89.50	410.40	-0.78
17	City Hotel	Complementary	July	1	70.00	318.82	-0.78
18	Resort Hotel	Corporate	May	129	96.73	410.40	-0.76
19	Resort Hotel	Complementary	February	4	104.50	410.40	-0.75
20	City Hotel	Complementary	November	6	78.67	318.82	-0.75

Figure 2, output query for the second task

In this dataset, the hotel type and the specific months do not show significant patterns to give an insight. This means that city and resort hotels maintain relatively stable pricing structures, and booking value differences are not strongly driven by hotel type alone. Seasonal effects exist, but they are already embedded in ADR, and so not extreme enough to dominate booking value. This suggests that the hotels have an effective seasonal pricing strategies in place.

Underperformance is primarily driven by market segments based on this dataset. Complementary bookings generate low estimated booking value due to their non-revenue nature.

The strategy should be to focus revenue optimization on channel strategy and not on seasonal pricing. This means not over-adjusting prices by month or force discounts when they are not needed. So, the focus is on reducing reliance on low-value segments (such as Complementary) and encouraging higher value channels such as Corporate and Direct. The strategy can also be tied with the first task by introducing minimum stay rules for low-value segments or a higher minimum price for low-value segments.

Task 3: Booking trends

The objective of this task is identify booking trends by hotel type and month. Monthly booking demand was ranked by hotel type to identify peak and off-peak periods. The pattern here is that the summer months are the busiest.

	hotel	arrival_date_month	total_bookings	demand_rank
1	City Hotel	August	8983	1
2	City Hotel	May	8232	2
3	City Hotel	July	8088	3
4	City Hotel	June	7894	4
5	City Hotel	October	7605	5
6	City Hotel	April	7480	6
7	City Hotel	September	7400	7
8	City Hotel	March	6458	8
9	City Hotel	February	4965	9
10	City Hotel	November	4357	10
11	City Hotel	December	4132	11
12	City Hotel	January	3736	12

Figure 3, output for the third task, specific to City Hotel type

13	Resort Hotel	August	4894	1
14	Resort Hotel	July	4573	2
15	Resort Hotel	April	3609	3
16	Resort Hotel	May	3559	4
17	Resort Hotel	October	3555	5
18	Resort Hotel	March	3336	6
19	Resort Hotel	September	3108	7
20	Resort Hotel	February	3103	8
21	Resort Hotel	June	3045	9
22	Resort Hotel	December	2648	10
23	Resort Hotel	November	2437	11
24	Resort Hotel	January	2193	12

Figure 4, output for the third task, specific to Resort Hotel type

The strategy will be focused more on operations where during peak periods more staff should be hired for the summer season.

Conclusion

This analysis used SQL to examine hotel booking data with the goal of supporting cancellation reduction, booking value optimization, and operational planning. The results show that booking behaviour is strongly influenced by lead time and customer type, with longer lead times and transient customer segments exhibiting higher cancellation risk.

Estimated booking value analysis, using ADR as proxy, indicated underperformance is primarily driven by market segment rather than hotel type or seasonality. Complementary and certain intermediary channels consistently generate lower booking value, while pricing across months appears relatively stable, which suggests that seasonal pricing strategies are already aligned with demand.

Clear seasonal patterns were observed in booking volume, which enabled to identify peak and off-peak periods for each hotel type. These demand trends support operational strategies focused on staffing, capacity planning, and targeted marketing rather than aggressive pricing adjustments.

Overall, the findings suggest that the greatest opportunities for improvement lie in managing cancellation risk through lead time based policies, optimizing channel mix to improve booking quality, and aligning operational resources with predictable demand cycles.