BUSINESS REPORT

PREPARED FOR :

Global Hotels and Resorts(GHR)

TEAM BUSINESS AS USUAL
Rochester Institute of Technology

Harshitha B Nagaraj (Team Leader)
Master’s in Data Science | hb2548@rit.edu

Rahul Kengeri
Master’s in Data Science | rk1087@rit.edu

RIT Business Analytics Competition 2024

EXECUTIVE OVERVIEW

1

Background

Global Hotels and Resorts (GHR) is a top European hotel brand . With a recent acquisition of a prominent hotel in Lisbon, Portugal, GHR aims to harness the full revenue potential of this new property. Recognizing the importance of strategic optimization, GHR has enlisted our team's expertise to devise tailored solutions to this endeavor.

Overview

After conducting an extensive analysis of booking data from **2015 to 2017**, we have compiled actionable recommendations for Global Hotels and Resorts (GHR) to enhance its profitability. This report delves into GHR's **revenue potential** and highlights areas where improvements can be made, specifically focusing on its cancellation policy, upgrade policy, and resource planning. The main concern for GHR is to **increase its revenue** and it also wishes to have an in-depth understanding of various factors affecting the revenue. They also want to identify the **exact revenue loss** happening from free upgrades and cancellations.

Through our analysis, we estimate that out of all upgrades, GHR loses money on **62.5%** of upgrades [Fig. A]. We talk about upgrade loss in more detail in [Page 4]. To mitigate this loss, we propose leveraging our Upgrade Prediction Model, which can help forecast future upgrades and identify high-upgrade probability customers. By utilizing this model, GHR can strategically allocate resources and minimize revenue loss associated with upgrades. We developed our **Upgrade Prediction Model** using **CatBoost**, a boosting algorithm known for its effectiveness in handling imbalanced data by iteratively training weak learners to correct misclassifications. After rigorous optimization and hyperparameter tuning, our model achieved an impressive F1 score of 0.40. For Global Hotels and Resorts (GHR), we recommend implementing two variations of the same model, each tailored to serve a distinct purpose and usage within GHR's operations. Further details on these model variations can be found on [Page 5].

Our analysis of **canceled bookings** at GHR reveals some key findings such as cancellation rates are higher for specific booking channels (OTAs), specific market segments, and specific months. Additionally, customer history and pre-payment behavior also influence cancellation likelihood. We have highlighted all factors affecting cancellations in [Page 6] and also provided our recommendations to mitigate losses due to such cancellations.

DATA PREPARATION

2

Data Cleaning

The dataset consists of 94,364 rows documenting booking reservations at GHR from 2015 to 2017, with each row containing 28 columns of relevant information. Before analysis, thorough data cleaning and preprocessing were conducted, encompassing checks for data types, missing values, null entries, and duplicates. Null values were addressed by either imputation with suitable values or mode substitution. Additionally, we identified five instances where the sum of 'Number of Children', 'Number of Babies', and 'Number of Adults' equaled zero, and these rows were subsequently removed from the dataset.

Feature Engineering

We introduced several new columns to enrich the dataset with valuable insights. One of these additions was '**LengthOfStay**,' calculated as the sum of 'StaysInWeekendNights' and 'StaysInWeekNights'. We also introduced '**is_upgraded,**' assigning values of 0 for not upgraded and 1 for upgraded bookings (if alphabetically higher). Another column, '**TotalGuests**' represented the combined count of children, babies, and adults. '**is_family**' identified bookings including children or babies. We segmented **'ADR_category'** into four quartile-based categories for effective rate analysis. '**DepartureDate**' was derived by adding the length of stay to the arrival date. '**CancellationTiming**' measured days between 'ReservationStatusDate' and 'ArrivalDate' for cancellations. **'FinalReservationStatus' c**ategorized canceled bookings into last-minute (<= 3 days) or advanced cancellations. Also, changed the status of no-show bookings into three types (advance, last-minute cancellation, late check-in). We define status as **late check in** if the ReservationStatusDate is between ArrivalDate and DepartureDate.Further, '**Length_of_stay_category**' and '**Lead_time_category**' segmented lengths of stay and lead times into useful categories based on intuition. Additionally, **'BookingDate'** was derived by subtracting the lead time from the arrival date. Finally, 'MarketCategory' strategically categorized 'Market Segment' into 'Business' (e.g., 'Aviation,' 'Corporate,' 'Complementary') and 'Leisure' (e.g., 'Direct,' 'Groups,' 'Offline TA/TO,' 'Online TA').

Our target variable is '**is_upgraded**'. Prior to model training, we applied **One Hot Encoding** and **Label Encoding** to all categorical columns and also dropped some columns to **avoid data leakage.**

For all our data cleaning, processing and to create charts and visualizations we use **Python** notebooks along with libraries like Pandas, numpy and matplotlib.

LET'S TALK MONEY

3

Method

The calculation of Potential Revenue for GHR is based on a formula recommended by the other consulting team, with our own tweaks :

Potential Revenue = Revenue Gain (from hotel stays + from non-refundable cancellations) - Revenue Loss from free upgrades - Revenue Loss from last-minute cancellations.

We define '**upgrade loss**' as the loss incurred when a customer is upgraded from a lower room type to a higher one (including within the same tier, such as from A to B). Free upgrades may occur when the hotel has overbooked lower-tier rooms and has unsold inventory of higher-tier rooms, or to celebrate special occasions for loyal customers. Our analysis also has revealed that some **upgrades have been profitable** for GHR. This can happen when the average ADR for the higher-tier room is lower than that for the lower-tier room, due to a lack of demand for higher-tier rooms when the booking was made by a customer.

Under GHR's current cancellation policy, **the non-refundable deposit,** which equals the total cost of the stay for the customer, is retained regardless of whether the cancellation is made in advance or at the last minute. Therefore, this non-refundable deposit contributes to revenue.

Revenue Gain

Year	Revenue From Hotel Stays	Revenue From Non-refundable cancellations	Total Revenue
2015	€ 2,021,325.35	€ 596,973.94	€ 2,618,299.29
2016	€ 6,782,078.58	€ 1,146,835.34	€ 7,928,913.92
2017	€ 8,322,469.19	€ 1,554,516.51	€ 9,876,985.70

Total Revenue (2015 - 2017)

€ 20.4 M

Upgrade Loss

To determine whether an upgrade given to a customer at GHR resulted in a **loss or profit,** we consider several important factors. Imagine a scenario where a customer was upgraded from a lower room type **"A,"** to a higher room type, **"G."** This upgrade might happen because room A was either fully booked or the customer received a complimentary upgrade. Because of the **free upgrade**, the customer pays a lower ADR for room G than what room G typically sells for. This situation is what we refer to as an **"upgrade loss".** The greater the upgrade, the higher the revenue loss. To assess whether this upgrade was a loss or a profit, we need to establish the typical selling price of room G during the same time period. We do this by analyzing other bookings for room type G that weren't upgraded but share similar characteristics with this upgraded booking.

To identify which characteristics have the most significant influence on the ADR, we use a statistical method called **Spearman's Rank Test**. This test helps us determine correlations between different factors like Booking Date, Total Guests, Deposit Type, etc. Once we identify the most correlated characteristics, we use them to find similar bookings to the customer who received the upgrade. Next, we calculate the average ADR for these similar bookings. If the average ADR for these bookings turns out to be higher than what the upgraded customer is paying, it indicates a **loss for GHR.** If the average ADR is lower or similar to what the upgraded customer is paying, it signifies a **profit**. The gains have been **subtracted from the losses** for this calculation. For yearly calculation see [Fig. B].

Upgrade Loss (2015 - 2017)

€ 272,202.05

Cancellation Loss

Table 1 illustrates the revenue gain or loss resulting from booking cancellations. In cases of Advance Cancellation, we are assuming that the room will be sold again and hence the loss is 0. For Last Minute Cancellation, the loss GHR incurs is the cost of stay. However the customer pays a one-night fee per GHR's current cancellation policy. For yearly breakup, see [Fig. B].

	Advance Cancel	Last Minute Cancel
No Deposit	0	(ADR * LengthOfStay) - ADR
Refundable	0	(ADR * LengthOfStay) - ADR
Non-Refund	-(ADR * LengthOfStay)	-(ADR * LengthOfStay)

Cancellation Loss (2015 - 2017)

€ 1.2 M

Potential Revenue = €20,424,198.90 − € 272,202.05 − € 1,228,895.04

€18,923,101.80 is the Potential Revenue (2015-2017) of GHR.

PREDICTIVE MODEL FOR UPGRADES

4

Initially exploring various machine learning models, we settled on **boosting algorithms** due to their ability to handle our **imbalanced dataset** containing 85,731 instances of "Not Upgraded" and 8,475 instances of "Upgrades". Experimenting with AdaBoost, GradientBoost, XGBoost, and CatBoost, we found XGBoost and CatBoost to be the most promising, initially yielding F1 scores between 0.05 to 0.23. To address the data imbalance, we applied a combination of **oversampling and undersampling techniques**, slightly improving the ratio. Subsequent model runs with the resampled data showed CatBoost to be slightly superior. After fine-tuning, we achieved a **final F1 score of 0.40 [Fig. R] with the CatBoost model.**

Use Cases

Through our analysis, we've pinpointed **two practical applications** for leveraging the predictive model within GHR. Each application has a slight variation of the same model:

1. Estimating the Number of Future Upgrades

This application is all about predicting how many upgrades GHR will have in the coming month. By passing in the next month's data to the predictive model, GHR can figure out the total number of upgrades expected. This helps GHR estimate how much **revenue** they might lose. To check if they're overbooking rooms more than before, they can compare this predicted loss with data from previous months. If the predicted loss is higher, it means GHR might be **overbooking** more or might be offering more complementary upgrades [Page 4]. When setting up this model, we prioritize parameters that optimize **recall**. Why? Because high recall means we catch most upgrades, even if it means sometimes we mistakenly think an upgrade is needed when it isn't. But this also means that the calculated revenue loss will also be higher, showing us the maximum potential loss GHR could face from upgrades.

2. Predicting the Probability of Customer Upgrade

In this scenario, we focus on evaluating the likelihood of individual customers upgrading. The model not only provides predictions but also assigns probabilities to each class (Not Upgraded, Upgraded) [Fig. Q]. If the probability of an upgrade is notably high (additional analysis needed to set a threshold), GHR can anticipate the likelihood of that specific customer upgrading. When configuring this model, our focus was on optimizing **precision** [Fig. S]. Why? Because precision ensures that when the model predicts an upgrade, it's highly likely to be correct. This means fewer false positives, resulting in more **accurate resource allocation and strategic planning** for the customer's arrival.

OPTIMIZING CANCELLATION POLICY

5

Factors Influencing Cancellation

1. Timing

In the **Stacked Bar Chart (Figure 1)**, we've analyzed the Arrival Month and Reservation Final Status to show the proportion of canceled bookings. This includes both Last-minute and Advance cancellations, while the Check-Out category encompasses both regular Check-outs and Late Check-ins.

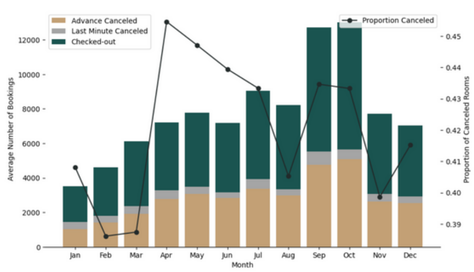


Figure 1 : Canceled proportion

April-June

has the highest proportion of cancellations.

Lead Time

Longer Lead times leads to higher advance cancellations.

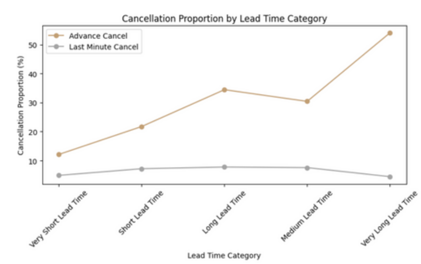


Figure 2 : Lead Time proportion

Longer lead times lead to higher advance cancellations, however, lead time has no effect on last-minute cancellers. From [Fig. C], reservations with longer length of stay have a higher cancellation rate. This can lead to a higher room unoccupancy and also a significant loss of revenue. Therefore, GHR must consider revisiting its cancellation policy for bookings with longer length of stays.

2. Market Segment

By understanding the different types of customers and markets, GHR can improve its cancellation policy. This means tailoring cancellation rules and incentives to suit the needs and behaviors of each customer segment, ultimately maximizing retention and satisfaction.

Although TA/TO has the lowest proportion of last minute cancellers (Figure 3), they have the highest cancellation rate overall [Fig. D], possibly due to the convenience of online booking and rate comparison. These platforms often offer flexible cancellation policies. Meanwhile, repeated guests exhibit lower cancellation rates, suggesting a higher level of commitment to their reservations.[Fig. E]

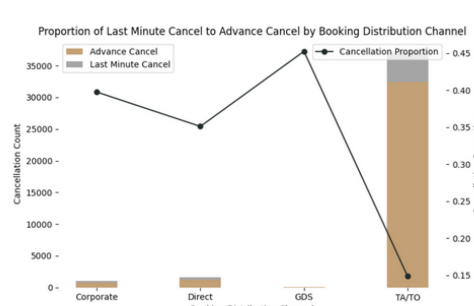


Figure 3 : BDC proportion

When examining GHR's customer market segment [Fig. F], it's evident that the cancellation rate is notably high for group bookings. This poses a risk of unoccupied rooms and revenue loss. Therefore, GHR needs to implement a distinct policy for group bookings to mitigate this issue.

Based on customer histories, we estimate Reservation Cancellation probability using logistic regression [Formula]. **Figure 4** indicates that as past cancellations increase, so does the probability of future cancellations, aligning with intuition. Moreover, as Booking Changes rise, cancellation probability drops [Fig. G]. This suggests that individuals who make changes to their stay, to suit their needs tend not to cancel as GHR has satisfied their requirements.

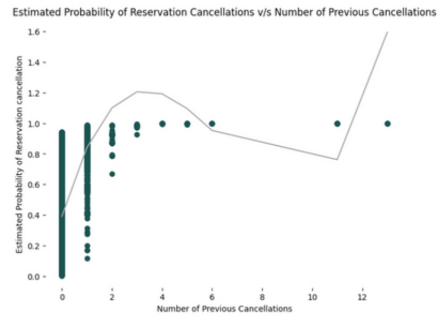


Figure 4 : Previous cancel Analysis

3. Money

Analysis of GHR Customers' spending on room reservations reveals that cheaper rooms are canceled more frequently than expensive ones [Fig. H]. However, there is a higher proportion of last-minute cancellations for Executive Room types[Fig. I]. Additionally, customers who don't pay a deposit before arrival tend to cancel their reservations more often. Thus, GHR should review its deposit policy to mitigate cancellations. [Fig. P]

Our Recommendations

These are our recommended cancellation policy amendments that we recommend to GHR based on the above analyses.

- Introduce more **Non-refundable rates** from April to June (Figure 1) as these months have a higher cancellation rate, offering a **10-20% discount.** This strategy helps guarantee revenue for GHR.
- For customers with refundable deposits, who cancel last minute, GHR can return that deposit as a **voucher** which can be used for a future stay at GHR.
- For bookings with **longer length of stays**, and customers seeking a fully refundable option for advance cancellations, GHR must implement a policy requiring an **additional fixed fee** for flexible cancellation privileges.
- To address the surge in OTA cancellations, GHR should **re-negotiate contracts** with high-cancellation **booking companies listed** in the [Fig. J], prone to **last-minute cancellations**.
- Establish **clear communication** with the group leader for group bookings to prevent cancellations. Implement a **phased deposit system**, starting with a small deposit and followed by incremental payments closer to the arrival date, to cover against cancellations and ease the initial financial burden on the group.
- Implementing a **loyalty point system** where guests earn points with each successful booking, redeemable for future stays. This encourages repeat guests, reduces cancellations, and incentivizes customers to maintain their bookings.

MAXIMIZING POTENTIAL REVENUE

6

Refining ADR

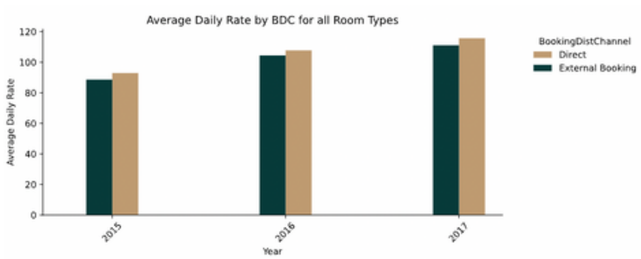


Figure 5 : Rate Parity Analysis

The **Rate Parity Analysis** (Figure 5) shows that room rates are similar for Direct and External Bookings (GDS, Corporate, and TA/TO). However, with fewer Direct Bookings [Fig. K], revenue loss occurs due to OTA commission fees, which can reach 15-30% of the room rate.

Strategy

Rate parity is a contractual agreement between the hotel and its Distribution agents that its room rate will remain the same across all booking channels including the properties' own website. After the recent amendment to *Portuguese competition law* [Link] which prevents OTAs from banning hotels from offering lower prices. Therefore we must take advantage of this new law and price direct bookings strategically lower to attract more direct bookings which will increase revenue.

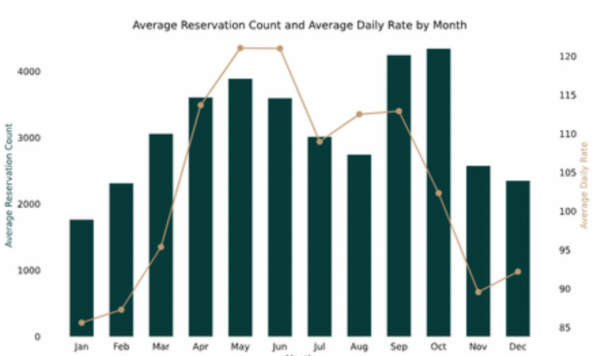


Figure 6 : Demand Analysis

Figure 6 highlights a notable surge in hotel demand during September to October, despite it being outside Portugal's peak tourism season (May-July). This trend likely stems from tourist guides advocating off-season travel, benefiting from favorable weather and reduced hotel prices.

Strategy

Considering the increased demand seen during off-peak months like September to October, it's essential for GHR to adapt its pricing strategy promptly. By raising prices strategically during this period, GHR can make the most of the surge in demand. GHR's maintenance of high ADR during peak seasons is commendable. Additionally, exploring bundled offers might further stimulate bookings during peak periods.

Improving Room Occupancy

Observing **Figure 7**, we notice that despite an extended duration of stay, the average daily rate (ADR) remains relatively steady. It's also evident that both Business and Leisure Customers tend to favor brief stays (1-3 days) [Fig. L]. Additionally, a slight majority of Leisure customers also opt for stays lasting 4-7 days, the goal of GHR should be to improve the percentage of customers in this category.

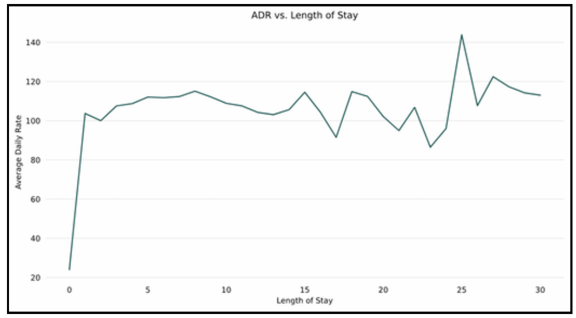


Figure 7 : Length of Stay Analysis

Strategy
One simple approach could be to provide reduced rates for extended stays. However, GHR could enhance customer experience by partnering with and advertising local events, offering discounted passes along with the room rates. This approach not only promotes nearby events but also encourages longer stays. Given GHR's focus on leisure travelers, potential partnerships could include Festas de Lisboa and Rock in Rio Lisboa. For business travelers, GHR might consider collaborating with Web Summit.[Link]

When it comes to maximizing occupancy rate, attracting group bookings can be a game changer. Group Bookings not only ensure a higher occupancy rate but also offer an opportunity to generate additional revenue through various services and amenities. As we can see from **Figure 8**, Group Travelers (Group and Contract) on average make more special requests than individual travelers (Transient and Transient-Party).

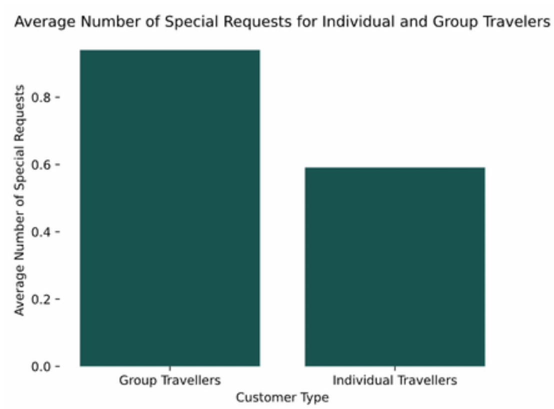
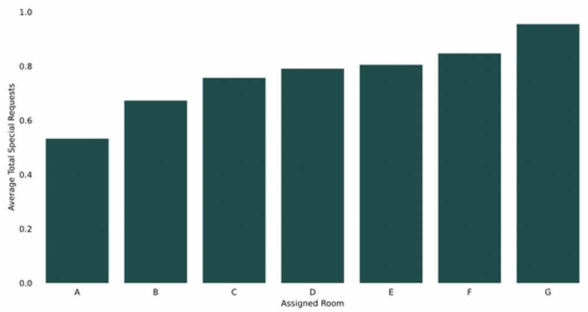


Figure 8 : Average Number of Special Requests

Strategy
Based on [Fig. M], it's evident that the majority of group bookings occur between August and November. Therefore, GHR needs to ensure all amenities are serviced and well-kept before this period to maximize group travelers' utilization and boost revenue. Lisbon is also a very popular wedding destination, especially during June and September [Link]. However, GHR has low group bookings during this time. To improve this, GHR can offer wedding packages featuring discounted room rates, complimentary upgrades for the bride and groom, and a dedicated event coordinator for a seamless experience. Highlighting the value of these packages can effectively attract group bookings and improve occupancy rates.

Improving Resource Planning



From **Figure 9**, it's apparent that customers booking Luxury Rooms (E, F, G) tend to make more special requests on average compared to customers booking Deluxe rooms (A, B, C, D). Additionally, customers who book Room Types B and C typically make numerous booking changes [Fig. N].

Figure 9 : Special Requests for Room Types

Strategy
For luxury room bookings, proactively communicate with customers before their arrival via email or chat to address any special requests and ensure that their needs are met. This proactive approach helps in better preparation and resource planning for GHR. Additionally, for room types B and C which experience frequent booking changes, gather guest feedback and note down any changes made to bookings. Incorporate these insights into room allocation processes to minimize future booking changes for these room types.

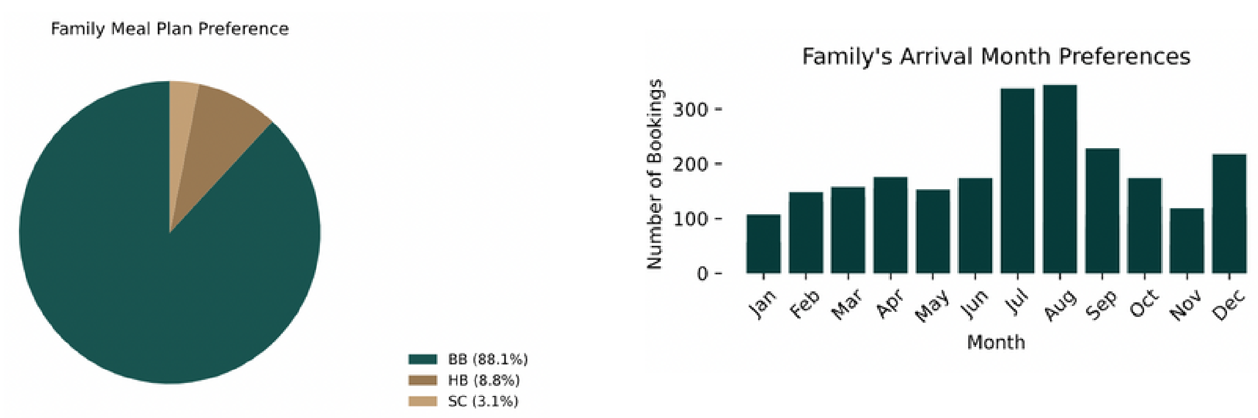


Figure 10 : Family Meal Plan Analysis

From **Figure 10**, Guests with children or babies prefer to book a meal plan more than not having a meal plan at all when compared to Adults-Only.[Fig. O]

Strategy
Discounted meal plans are a compelling option for families, offering the convenience of pre-booked meals and alleviating the stress often associated with dining arrangements during vacations. Notably, **Figure 10** highlights a surge in family reservations during July and August, coinciding with Portugal's school summer holidays [Link]. Recognizing this trend, GHR can enhance its operational efficiency by strategically contracting additional kitchen staff and ensuring ample grocery supplies during these peak months.

APPENDIX-1

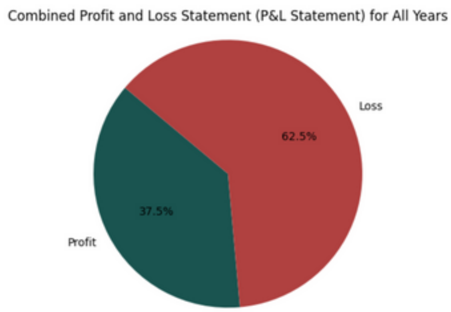


Fig. A

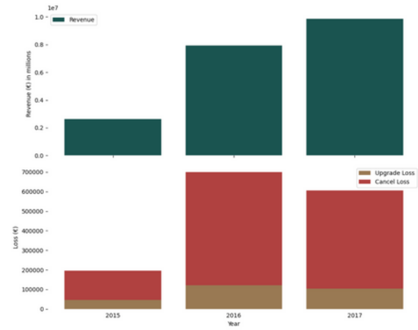


Fig. B

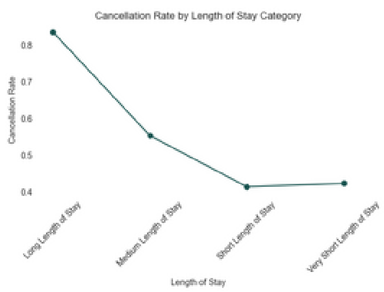


Fig. C

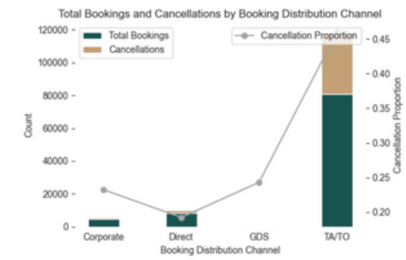


Fig. D

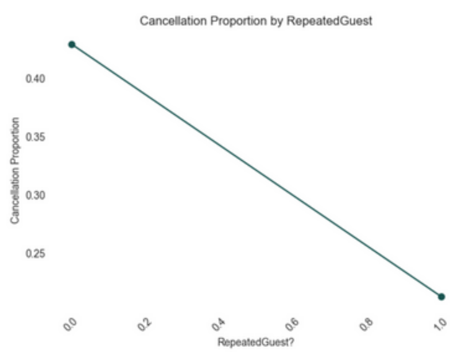


Fig. E

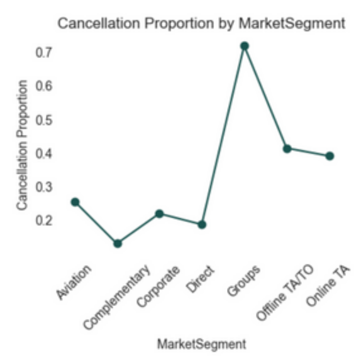


Fig. F

Count of booking changes:

$$\text{booking\_changes\_counts} = df['BookingChanges'].value\_counts()$$

Count of cancellations for each booking change value:

$$\text{cancelled\_counts}(x) = \text{Number of entries in } cancelled\_df \text{ where 'BookingChanges'} = x$$

PMF Calculation:

$$pmf(x) = \frac{\text{cancelled\_counts}(x)}{N}$$

Where:

- N is the total number of entries in the dataset ($N = len(df)$).

Formula

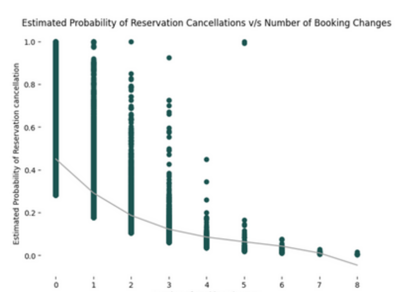


Fig. G

APPENDIX-2

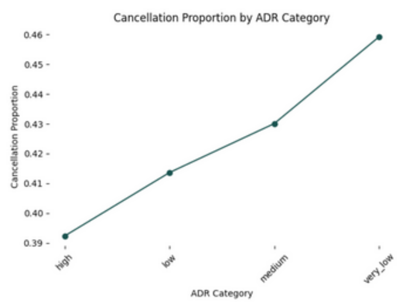


Fig. H

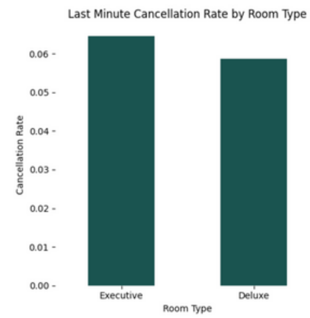


Fig. I

Fig. J

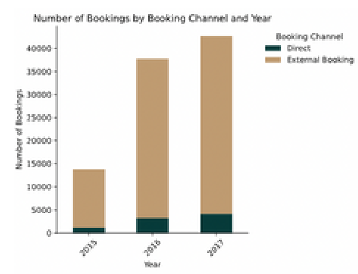


Fig. K

```
Business Market - Length of Stay Category:
length_of_stay_category
Long Length of Stay          0.000000
Medium Length of Stay        1.189609
Short Length of Stay         6.870600
Very Short Length of Stay   91.939791
Name: Business, dtype: float64

Leisure Market - Length of Stay Category:
length_of_stay_category
Long Length of Stay          0.013989
Medium Length of Stay        1.368905
Short Length of Stay        29.724221
Very Short Length of Stay   68.892886
Name: Leisure, dtype: float64
```

Fig. L

Fig. M

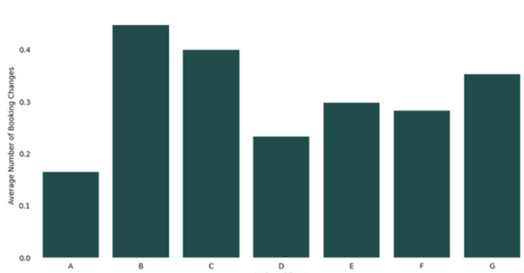


Fig. N

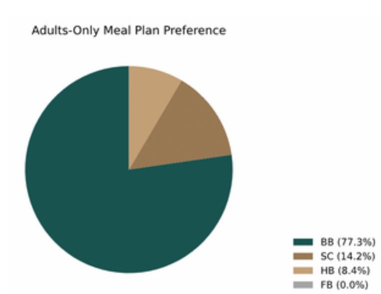


Fig. O

APPENDIX-3

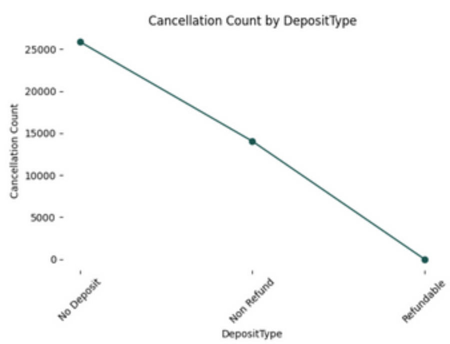


Fig. P

DataFrame with Predictions and Probabilities:

	Prediction	Probability of 0	Probability of 1
0	0	0.785989	0.214011
1	0	0.917348	0.082652
2	0	0.874519	0.125481
3	0	0.528159	0.471841
4	0	0.866958	0.133042
...	...	...	...
28257	0	0.522644	0.477356
28258	0	0.966720	0.033280
28259	0	0.956736	0.043264
28260	0	0.936967	0.063033
28261	0	0.987049	0.012951

Fig. Q

Classification Report:

	precision	recall	f1-score	support
0	0.95	0.88	0.91	25719
1	0.31	0.54	0.40	2543
accuracy			0.85	28262
macro avg	0.63	0.71	0.65	28262
weighted avg	0.89	0.85	0.87	28262

Fig. R

Classification Report:

	precision	recall	f1-score	support
0	0.94	0.94	0.94	25719
1	0.40	0.37	0.39	2543
accuracy			0.89	28262
macro avg	0.67	0.66	0.66	28262
weighted avg	0.89	0.89	0.89	28262

Fig. S