Project Report

Title: Optimizing Revenue Leakage & Profitability in the Hospitality Sector

Prepared by: Tapish Vikram

---

1. Executive Summary

This project seeks to determine the major causes of revenue leak and underperformance in hotels through data-driven methodologies. Drawing on rich datasets that include hotel information, booking history, room setups, and date attributes, we used an organized analytical framework to unearth inefficiencies in booking patterns, platform performance, pricing, and customer trends. Our multi-step approach involved Exploratory Data Analysis (EDA), Diagnostic Analysis, Consulting Recommendations, and Business Justification. These learnings were then converted into actionable suggestions that maximize both operational effectiveness and profitability.

2. Objectives

a. Conduct an end-to-end exploratory and diagnostic analysis using real-world hotel data.
b. Identify sources of revenue leakage such as cancellations, discount misuse, and seasonal imbalances.
c. Recommend strategic interventions for maximizing revenue realization.
d. Quantify the business impact and justify recommendations using data.

3. Module Wise Summaries

a. **Module A: Exploratory Data Analysis**

i. The goal of EDA was to familiarize ourselves with the dataset and understand the basic structure, patterns, and anomalies.

ii. Dataset Used:

- **dim_date.csv:** Date mapping and weekday/weekend flags
- **fact_bookings.csv:** Individual booking data including revenues and booking status
- **dim_hotels.csv:** Hotel attributes like city and category
- **dim_rooms.csv:** Room configurations and details
- **facts_aggregated_bookings.csv:** Aggregated KPIs like ADR and occupancy

iii. Key Insights:

- **Check-in Trends (Weekday vs Weekend):** Weekdays typically show higher bookings, highlighting opportunities for special weekday packages or targeted promotions.Bookings are significantly higher on Weekends compared to Weekdays. The bar for "Weekend" is substantially taller than the

bar for "Weekday", indicating that more customers check in on weekend days.

- **Revenue & Occupancy Patterns:** The first graph, "Bookings by Room Category," clearly shows that certain room categories are significantly more popular than others. The standard categories likely dominate the booking volume, suggesting either higher demand for these types or perhaps greater availability. This insight is crucial for inventory management; hotels might consider optimizing the allocation of room types based on these trends.

  The second graph, "Bookings by Platform," illustrates which booking channels are most effective. There is a clear preference among customers for specific platforms, with a few platforms likely driving the majority of bookings. Understanding the performance of each platform allows hotels to focus their marketing spend and efforts on the most successful channels, while also potentially exploring strategies to boost bookings on less utilized platforms or negotiate better terms with the top performers.

  Bar charts revealed that room category RT2 had the highest number of bookings, suggesting its popularity among guests, while RT4 had the least. Similarly, most bookings were made through the 'others' category, followed by known platforms like makeyourtrip and logtrip, indicating a strong reliance on third-party channels. Additionally, a high number of null values in the 'day_type_checkin' column post-merge pointed to potential mismatches in date formats or missing date entries, requiring further data cleaning.

- **Platform Popularity:** The line plot depicting "Seasonal Booking Trends Over Time" reveals clear patterns in booking volumes across different months and years. There is a noticeable seasonality, with peaks and troughs in the number of bookings occurring at similar times each year. Identifying these peak and off-peak seasons is vital for strategic revenue management. High-demand periods, typically coinciding with holidays or specific times of the year, present opportunities for increasing prices, while low-demand periods may require promotions or discounts to attract guests and maintain occupancy levels. Analyzing the consistency and magnitude of these fluctuations over the observed years allows for better forecasting and dynamic pricing adjustments, maximizing revenue during busy times and mitigating losses during quieter periods. This seasonal trend analysis provides a foundation for optimizing pricing strategies and resource

- **Revenue Analysis - Generated vs Realised:** This focuses on quantifying revenue leakage by comparing the total revenue that was expected to be generated against the total revenue that was actually realized. The analysis

calculates the total sum for both 'revenue_generated' and 'revenue_realized' across all bookings. The difference between these two totals represents the 'Total Revenue Leakage'. A positive leakage value indicates that the realized revenue is less than the generated revenue. The percentage leakage is also calculated to understand the magnitude of the loss relative to the potential revenue.

The bar plot visualizes these two totals, making it immediately clear whether the generated revenue is higher than the realized revenue. A noticeable difference between the two bars visually confirms the existence of revenue leakage.

The second plot, a histogram of 'revenue_difference', provides a distribution of the difference between generated and realized revenue at an individual booking level. The shape of this histogram indicates how often different levels of revenue leakage or even potential over-realization occur. A distribution skewed towards positive values (right side) suggests that in many individual bookings, the generated revenue was higher than the realized revenue, contributing to the overall leakage. This plot helps understand the frequency and magnitude of discrepancies in individual transactions, which could be due to discounts, cancellations, adjustments, or errors.

- **Customer Booking Behavior:** Based on the visualizations and calculations presented, we can interpret the customer segmentation analysis as follows:

  The initial histogram titled "Distribution of Revenue Realized per Booking" provides insight into the frequency of different revenue amounts generated from individual bookings. A right-skewed distribution, as suggested by the quartiles where the values increase significantly from Q1 to Q3 and the maximum is much higher, indicates that while a large number of bookings generate lower revenue, there is a smaller but significant number of bookings that bring in much higher revenue. This immediately suggests heterogeneity in customer spending behavior.

  The quartile values (Q1, Q2, Q3) serve as thresholds to categorize bookings into "Low Spender," "Medium Spender," and "High Spender" segments based on the 'revenue_realized' from each booking. Q1 represents the revenue value below which 25% of the bookings fall (Low Spenders below this value). Q2 is the median, with 50% of bookings below this value. Q3 is the revenue value below which 75% of the bookings fall (High Spenders are those above this value). The values printed show the specific revenue thresholds for these segments based on the dataset.

The count plot titled "Distribution of Bookings by Spending Segment" then visualizes the number of bookings falling into each defined segment: Low Spender, Medium Spender, and High Spender. This bar chart shows the relative proportion of bookings in each category. Typically, we would expect to see a distribution where the 'Medium Spender' segment is the largest, followed by 'Low Spender', and then 'High Spender' as the smallest segment in terms of booking count, reflecting the nature of most customer spending patterns. However, the specific counts printed provide the exact number of bookings in each segment, allowing for a precise understanding of segment size. This distribution is critical for understanding the composition of the customer base by spending power per booking.

Finally, the analysis specifically delves into the "High Spender" segment to understand their preferences, illustrated by the count plot "Room Category Preferences of High-Value Customers." This graph breaks down the bookings made only by customers classified as 'High Spender' by room category. By comparing the heights of the bars for different room categories in this plot, we can identify which room types are most popular among the high-spending clientele. This information is invaluable for targeted marketing, service enhancement, and inventory management aimed at retaining and attracting more high-value guests. If high spenders predominantly book certain premium or specific room categories (like RT2 or RT4 based on the previous analyses), it confirms that these room types are crucial for capturing high revenue and suggests focusing on the quality and availability of these rooms. The printed value counts for this segment provide the exact booking numbers for each room category by high spenders.

b. **Module B: Diagnostic Analysis**
   i. This module aimed to deep-dive into root causes of inefficiencies and revenue gaps.
   ii. **Key Insights:**
      - **Cancellation Patterns:** Certain platforms had a disproportionately high cancellation rate. High-value bookings were often cancelled, resulting in unrecovered revenue. Recommendations were formulated to improve prepayment enforcement.

      - **Platform Behavior and Discount Utilization:** Some OTAs exploited high discount percentages without translating into actual revenue gains. Cancellation rate was not proportional to booking volume. Introduced the concept of "Effective Conversion" (revenue realized post-cancellation).

      - **Property wise Revenue Contribution:** A Pareto-like trend was found: 20% of properties contributed to over 70% of revenue. Many properties

had lower occupancy despite offering competitive ADR. Pointed toward brand recognition and regional demand inconsistencies.

c. **Module C: Consulting Recommendations**
   i. Based on findings from Module A and B, we framed high-impact strategic interventions.
   ii. **Key Insights**
      - **Platform Optimisation:** Reduce dependence on low-conversion platforms. Introduce tier-based commission linked to actual realized revenue rather than gross booking volume. Implement pre-authorized card payments or non-refundable pricing.

      - **Dynamic Pricing and Discounts:** Develop an internal pricing intelligence tool to monitor seasonal ADR trends. Cap discount ceilings based on occupancy levels. Trigger weekday discounts for low-demand periods (Tue-Thu).

      - **Cancellation Policy Reform:** Introduce stricter cancellation policies on high-volume but high-cancel platforms. Offer loyalty benefits for customers with lower cancellation histories.

      - **Property Development:** Focus marketing investments on top-revenue-contributing properties. Conduct local demand forecasting to support underperforming hotels in high-tourism zones.

d. **Module D: Business Justification**
   i. To translate recommendations into tangible value, we developed financial and strategic justification.
   ii. **Key Insights:**
      - **Revenue Uplift from Reduced Cancellations:** A projected 15% decrease in cancellations can improve realized revenue by ~₹2.8 Cr per quarter (based on historical trends).
      - **Discount Rationalisation:** By capping discounts at 15% and linking to occupancy, margin loss can be curbed by ~₹1 Cr/month.
      - **Optimizing OTA Commissions:** Re-negotiating OTA contracts based on actual realized revenue (vs gross bookings) can save ~₹50L–₹75L quarterly.
      - **Revenue Growth via Occupancy Targeting:** Targeted weekday discounting to boost occupancy by 10% in lean periods could add ~₹1.2 Cr to quarterly revenue.

   These changes together represent an overall uplift potential of ₹5–6 Cr per quarter in direct or retained revenue.

4. Conclusion

The hospitality industry, while rich in data, suffers from unstructured analytical execution. Through this project, we successfully demonstrated how a combination of exploratory and diagnostic analytics can lead to powerful business strategies. The roadmap proposed is both actionable and financially justified. With continuous monitoring and adaptive decision-making, the hotels can turn insights into consistent revenue gains.