**Hotel Management Project (R.I.S.E)**  
**Revenue, Insights, Strategy & Evaluation**

An Integrated Analytical Report on Financial Trends, Guest Demographics, and Marketing Efficiency in Hotel Management

Report prepared by:

Shahad Abdulrahman Saleh Alshumrakh

Date:

29/04/2025

**Abstract**

The study presents a comprehensive analysis of hotel booking data, focusing on revenue trends, customer behavior, and market segmentation. Data was sourced from Kaggle and refined using Excel to align with the project’s objectives, with further methodological details outlined in the Data Collection section. Key variables include booking lead time, cancellation status, customer type, deposit type, and total revenue paid. Critical insights were uncovered regarding how lead time and booking behavior influence cancellations and revenue, alongside patterns in repeat guests, market segments, and room preferences, supporting strategic marketing efforts and revenue optimization. To enhance the accessibility of these insights, a dynamic and user-friendly dashboard was developed, designed with soft, eye-friendly colors and an intuitive layout. Users can easily view key metrics through KPI Cards displaying Total Revenue, Total Cancellations, Revenue Loss due to cancellations, the Most Frequent Customer Type, and Top Countries contributing to bookings. The dashboard offers visualizations that track the hotel’s historical and current performance across revenue, customer behavior, booking behavior, and cancellation insights, including cancellation rates analyzed by day and month. Guest profiles are clearly highlighted, enabling better targeting and marketing strategies. The dashboard is structured to ensure every visual directly serves the management’s decision-making needs, providing an intuitive experience that transforms complex data into clear, actionable insights. Ultimately, this project bridges in-depth analysis with practical visualization, offering hotel management a powerful tool to drive operational and financial success.

**1 Introduction**

In recent years, data analysis has become an essential tool in the hospitality sector, enabling hotels to make more informed decisions, improve operational efficiency, and enhance customer satisfaction. While many contemporary studies focus on big data and predictive analytics, small-scale data analysis remains highly valuable—particularly for initial investigations and decision support.  
This paper presents an analytical examination of hotel booking data, conducted using tools such as Microsoft Excel for data cleaning, transformation, and visualization. The goal is to identify meaningful patterns related to financial performance, guest demographics, and marketing efficiency. By uncovering significant trends within the data, the research aims to offer practical insights that can assist hotel management in enhancing service quality, optimizing revenue generation, and supporting long-term profitability.

**1.1 Purpose of the Project**

The aim of this project is to conduct a structured and extensive study of hotel booking data to identify patterns and trends that influence operational outcomes. The objective of this research is to clarify how booking patterns, customer types, and market segments affect hotel income and customer retention. The findings are meant to assist with strategic decision-making in areas like as pricing strategies, marketing initiatives, and customer relationship management.

**1.2 Purpose of this Document**

The purpose of this paper is to present the whole analytical process, including data gathering methods, methodologies, and significant discoveries. It provides a detailed record of the methodology used, the insights discovered, and the conclusions derived from the data analysis. This document also explains the manner in which the study could impart hotel management techniques to improve operational and financial performance.

**1.3 Overview of this Document**

This document is divided into sections to maintain clarity and logical flow. It begins with an overview of the study and its aims, followed by a literature evaluation that situates the research within current knowledge. The methodology section goes over the tools and methodologies utilized, including Agile methodology principles. The data collecting and preparation techniques are discussed next, followed by the main analysis and discussion of findings. Finally, the report provides a summary of significant findings and recommendations for hotel management, as well as references.

**1.4 Literature Review**

Numerous research studies in the fields of hospitality, tourism, and hotel management have recently emphasized the relevance of data analysis in these sectors. Although much of the research has focused on big data—currently the leading strategy for generating meaningful insights and solutions—working with smaller datasets also holds significant value. Analyzing small-scale data is particularly useful for conducting preliminary tests and evaluations that help in developing strategies later scalable to larger datasets. The following studies demonstrate the significant impact of data analysis approaches in the hospitality sector, leading to numerous valuable insights.

A study conducted by **M. P. Legg and S. R. Cho** [1] aims to examine the integration of **Customer Lifetime Value (CLV) models** to enhance performance in the hospitality sector. The researchers employed **cluster analysis** and **Markov chains** to propose strategies and techniques for understanding customer behavior in a dynamic, non-contractual setting. By using loyalty data and statistical methods, they successfully categorized customers into distinct segments and predicted their future behavior. The results demonstrated that CLV offers more accurate insights into customer profitability and supports the development of long-term engagement strategies. This report aligns with the previous study in emphasizing the strategic importance of customer data analysis for enhancing decision-making in hotel management. It employs practical tools to explore financial trends, guest demographics, and marketing efficiency.

This document by **R. Anubala [2]** studies the role of **predictive analytics** in transforming hotel operations and enhancing guest experiences. The research analyzed customer data using tools such as **machine learning** and **Internet of Things (IoT)** technologies, leading to improved decision-making, more efficient operations, and dynamic pricing strategies. The study also addresses challenges in the data analysis industry, including concerns about **data privacy**, **ethical issues**, and the need for adequate training. The findings emphasize the advantages of incorporating predictive analytics in the hospitality sector, showing that it can significantly improve customer satisfaction. This study supports the conducted report by emphasizing the importance of data analysis for enhancing operational efficiency and customer satisfaction in hotels. While Anubala studies advanced techniques such as machine learning and IoT, this paper focuses on fundamental data methodologies for identifying visitor behavior patterns and financial trends, setting the foundations for the potential use of predictive analytics in hotel operations.

**2 Methodology**

Due to the changing nature of project requirements during the execution phase, this project used an Agile methodology. Changes in data preparation requirements and improvements made during dashboard testing required adaptability in both planning and implementation.   
Agile methodology is an adaptive and iterative approach to project management and product development that emphasizes continuous improvement, collaboration, and flexibility to change. Instead of following a strict linear approach, Agile divides the project into smaller, more manageable chunks known as iterations or sprints, which generally run one to four weeks and focus on generating a workable version of a part of the final product.

**2.1 Key characteristics of the Agile methodology:**

1. **Iterative Development**: Work is separated into cycles, with feedback gathered at the conclusion of each iteration to guide the next phases.
2. **Flexibility**: Collaboration allows for the evolution of requirements and solutions, which are readily altered when new requirements or obstacles arise.
3. **Continuous Feedback**: Stakeholders are regularly consulted, generating feedback that informs continuing progress.
4. **Collaboration and Communication**: Agile facilitates open communication among team members and stakeholders to ensure that goals are corresponded.
5. **Focus on Deliverables**: Each iteration produces a useful or tested version of the product, allowing for early detection of issues and progressive refinement.

The Agile methodology enabled the project team to react fast to changes, respond effectively to stakeholder demands, and guarantee that the final analytical outputs remained relevant and useful. This technique allowed for flexibility at crucial phases such as data cleaning, metric calculation, and dashboard construction, especially when new patterns or improvement opportunities emerged.

**2.2 Analytical Approach:**

1. **Descriptive Analysis** The dataset's key components have been summarized to offer an overall picture of hotel booking behaviors and revenue trends.
2. **Exploratory Analysis** The objective was to find deeper ties and patterns that might inform strategic recommendations for hotel management.

**Microsoft Excel** was the major tool used throughout the study's execution, from data cleaning and restructuring to custom measure calculation and insight visualization. Excel's adaptability and wide range of capabilities, including Pivot Tables, IF statements, and custom formulae, enabled the project to handle the information in an agile and iterative manner.



Figure 2.1: Agile Methodology obtained from an article at nexapp website [3].

The figure [2.1] illustrates the Agile process as a continuous cycle consisting of Plan, Design, Develop, Test, Deploy, and Review stages. After each review, feedback is incorporated and the cycle continues, allowing for continual modification and adaptability. This iterative approach ensures responsiveness to change, faster feedback integration, and continuous improvement until the final launch.

## **3 Data Collection and Preparation**

### 3.1 Data Source

The study is based on secondary data collected from the Kaggle [4] platform, which includes a publicly available hotel booking dataset. N. Antonio, A. Almeida [5] published the Hotel Booking Demand Datasets (Data in Brief, 2019), from which the dataset derives. It includes booking data from two hotels in Portugal, a resort hotel (H1) and a city hotel (H2), for bookings made between July 1, 2015 and August 31, 2017.

### 3.2 Data Gathering and Processing

The original data was retrieved from the hotels' Property Management Systems (PMS) using T-SQL queries run in SQL Server Management Studio. Data collection according to strict guidelines to protect integrity and prevent data leakage:

1. Booking, customer, transaction, and distribution channel tables were used.
2. Timestamps were adjusted to reflect information available only up to the day before the guest's arrival.
3. Variables such as **ADR**, **deposit type**, and **booking changes** were engineered based on transactional and modification records.

The dataset underwent initial cleaning and anonymization using R, ensuring missing values were treated appropriately and personal information was removed to protect guest privacy.

### 3.3 Data Preparation in Excel

After initial processing, the dataset was imported into Microsoft Excel for further preparation to meet the specific analytical objectives of this project. This included:

* Deleting irrelevant columns to reduce noise.
* Creating new calculated fields, such as **total revenue** and **refunded amount**.
* Applying key Excel functions and tools:
  1. **Remove Duplicates** to eliminate redundant records.
  2. **Sort & Filter** to isolate relevant data subsets.
  3. **Conditional Formatting** to highlight anomalies and missing data.
  4. **IF Statements and Logical Functions** for value categorization.
  5. **Custom Formulas** for advanced metric calculation.
  6. **Pivot Tables and Slicers** for interactive data summarization and dynamic filtering.
  7. **Text-to-Columns** to split combined fields for cleaner analysis.

### 3.4 Time Scope

The dataset's time frame allows for examination of trends throughout numerous seasons and two complete years, giving a solid foundation for discovering repeating patterns and strategic insights.

**4 Data Analysis & Findings**

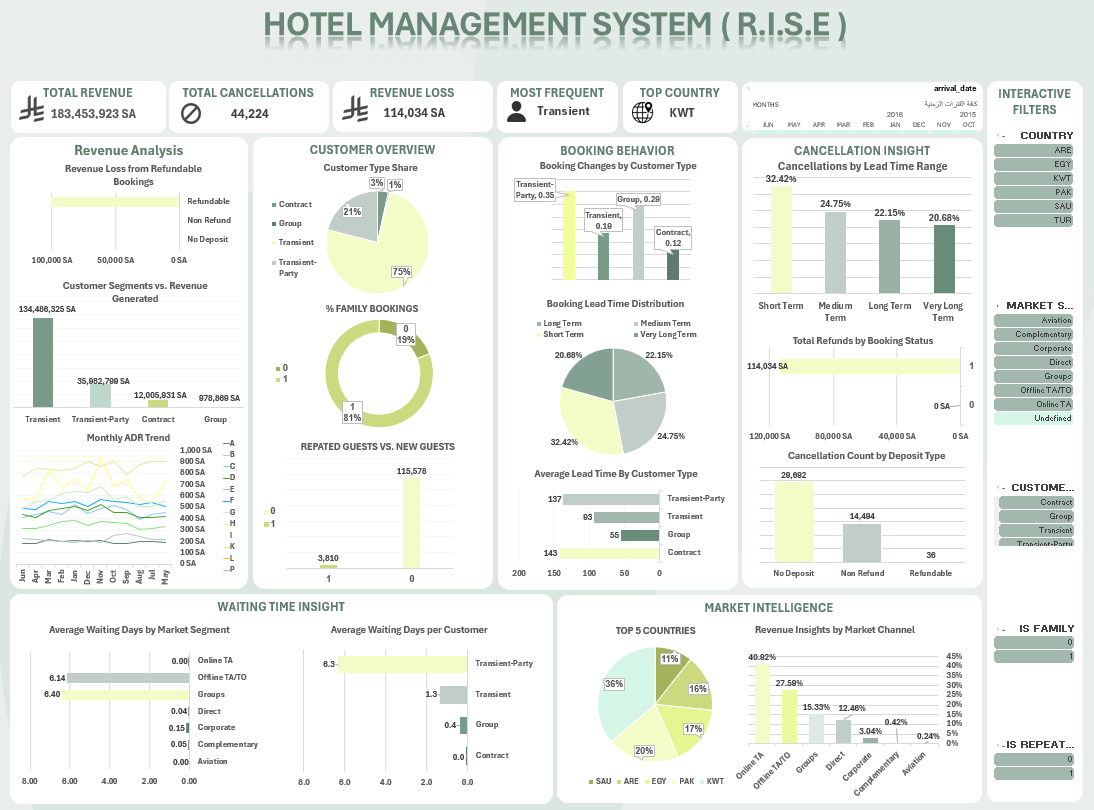


Figure 4.1 Hotel Management System (R.I.S.E) dashboard.

**4.1 Revenue Analysis**

After conducting an in-depth analysis of the revenue section, several key findings emerged:

**4.1.1 Revenue Loss from Refundable Bookings**  
The first chart demonstrates that revenue loss is mostly due to refundable reservations. In contrast, non-refundable and no-deposit bookings result in virtually no loss. This shows that customers who choose refundable bookings are more likely to cancel, potentially resulting in a considerable loss of revenue. To address this, the hotel may think about providing benefits for non-refundable bookings or changing the refundable policy to promote more commitment from guests.

**4.1.2 Customer Segments vs. Revenue Generated**  
The second chart shows that transient customers contribute the most funds, representing almost 115 million SA. This is followed by the Transient-Party and Contract segments, while Group Bookings contributing the least.   
This insight emphasizes the necessity of targeting Transient and Transient-Party customers in marketing and retention efforts, since they are the most significant customer base for revenue growth.

**4.1.3 Monthly ADR Trend**  
The last chart shows seasonal variations in the Average Daily Rate (ADR), with notable peaks in February and May. The ADRs for Deluxe and Premium rooms are consistently higher, while Accessible and Standard rooms are on the lower end.   
This study can help influence pricing strategies, such as advertising high-ADR room types during peak seasons and providing targeted discounts or packages for lower-ADR rooms during off-peak periods.

**4.2 Customer Overview**

This section presents an overview of the hotel's customer base, emphasizing on the distribution of customer types, the presence of family bookings, and the interaction between new and returning customers. These data assist identify crucial target segments and possibilities for enhancing visitor retention strategies.

**4.2.1 Customer Type Share**  
Statistics demonstrate that transitory customers dominating the booking environment, accounting for nearly 75% of all reservations. This demonstrates a high reliance on individual visitors rather than visitors from formal contracts or groups. Group reservations account for approximately 21% of overall revenue, while contract and transient-party clients account for only 3% and 1%, respectively. These proportions indicate that the hotel is mostly adapting to impulsive, short-stay guests rather than corporate or group-based guests.

**4.2.2 Family Bookings**  
An examination of family participation in bookings finds that families made most reservations—around 81%. Only 19% of bookings were for alone visitors, which may indicate that the hotel's services and facilities are more suited to families or couples than alone travelers. This trend suggests a possible area for enhancement if the hotel wants to attract more lone travelers. Individual travelers must have access to suitable services and advantages at the hotel. It is also essential to develop key standards that demand families to maintain a quiet environment with assigned floors exclusively for their rentals.

**4.2.3 Repeated vs. New Guests**  
The analysis of new and returning guests revealed a significant variation. New customers accounted for 115,580 bookings, while returning visitors contributed merely 3,810 bookings. This big variation indicates a large number of first-time guests, but it also indicates a lack of loyalty from customers. Improving customer service and establishing customized efforts to retain visitors may assist enhance the rate of repeat stays.

**4.3 Booking Behavior**

This section studies how visitors act during the booking process, with an emphasis on the frequency of booking changes, planning horizons, and the variation in lead time based on customer type. These trends can guide operational planning and revenue estimations.

**4.3.1 Booking Changes by Customer Type**  
Transient-party guests were found to make the most modifications to their bookings, with an average of 0.35 per booking. Group clients trailed with an average of 0.29, while transitory and contract guests had lower change rates of 0.19 and 0.12, respectively. This shows that certain guest segments, notably those traveling in informal groups, are more adaptable and may require more dynamic reservation management.

**4.3.2 Booking Lead Time Distribution**  
The lead time distribution shows a generally balanced trend, with short-notice bookings remaining the most common. Short-term bookings accounted for 32.42% of the total, followed by medium-term (24.75%), long-term (22.15%), and very long-term (20.68%). This spread emphasizes the importance of both short-term and long-term booking techniques to suit various guest planning patterns.

**4.3.3 Average Lead Time by Customer Type**  
When comparing how long in advance guests book based on their customer type, noticeable variations appear. Contract customers make the earliest reservations, with an average lead time of 143 days. Transient-party customers, on average, plan 137 days in advance. In contrast, group guests book 93 days in advance, whereas transitory consumers book closer to their arrival date, with an average lead time of only 55 days. These insights are important for targeting marketing efforts and predicting demand by segment.

**4.4 Cancellation Insight**   
Identifying the patterns and behaviors that cause booking cancellations is critical for reducing revenue loss and increasing operational efficiency. This section investigates the effects of lead time, deposit policy, and booking status on cancellation rates. By studying these factors, we acquire actionable insights that can assist with construct more effective cancellation rules and customer efforts for reducing financial risk and improving booking stability.

**4.4.1 Cancellations by Lead Time Range**  
The chart shows how the time of a reservation impacts its risk of cancelation. According to the study, short-term bookings are the most frequently cancelled, and cancellations tend to decrease as lead time grows. This pattern shows that customers who book closer to their stay date may be less committed or more willing to modify their arrangements. Understanding this pattern may assist in the development of short-term booking cancelation avoidance strategies, such as stricter rules or earlier booking benefits.

**4.4.2 Total Refunds by Booking Status**  
This section shows the total value of refunds processed, segmented by booking status. Most of the refund value is due to refundable bookings, demonstrating a significant possibility of revenue loss. Monitoring refund amounts assists in determining the financial effect of cancellation policies and encourages decision-making on how to enhance refund rules to reduce loss of revenue.

**4.4.3 Cancellation Count by Deposit Type**  
This chart analyzes the number of cancellations for various deposit types. Bookings with no deposit have the greatest cancellation rate, followed by non-refundable bookings, but refundable bookings have a surprisingly low cancellation rate. These findings highlight the importance of deposit policies in guest commitment and suggest that enforcing deposit requirements might significantly minimize cancellations.

**4.5 Waiting Time Insight**

This section focuses on the waiting periods that guests experience before obtaining a room, either by market sector or customer type. Understanding wait times is essential for improving customer satisfaction and operational efficiency.

**4.5.1 Average Waiting Days by Market Segment**  
The chart shows that guests that book using Offline TA/TO and Online TA have the longest average waiting durations, around 6.4 and 6.1 days, respectively. This means that third-party booking channels may create delays, potentially due to processing or coordination issues. Meanwhile, categories such as Corporate, Complementary, and direct bookings have short to no wait times, indicating streamlined and prioritized processes.

**4.5.2 Average Waiting Days per Customer**

Transient-Party customers had the longest average wait time of any customer class, at nearly 6.3 days, while Transient and Group customers have substantially shorter wait times. Contract customers exhibit almost little wait time, most likely due to pre-arranged plans. This suggests that Transient-Party customers may require additional attention to decrease delays and improve their booking experience.

**4.6 Market Intelligence**

This section gives insight into the geographical distribution of visitors as well as the market channels that contribute to revenue, assisting in the development of marketing and outreach strategies.

**4.6.1 Top 5 Countries**  
According to the pie chart, the bulk of hotel visitors (38%) come from Kuwait, followed by Saudi Arabia, the UAE, Egypt, and Pakistan. This data is critical in determining the hotel's primary markets, allowing it to adjust services, advertising materials, and even language options according to customer preferences. Furthermore, targeting marketing and advertising efforts in these high-performing nations, particularly Kuwait, may greatly increase outreach, visitor engagement, and income through customized campaigns.

**4.6.2 Revenue Insights by Market Channel**  
According to the bar chart, Online Travel Agencies (OTAs) account for more than 40% of income, followed by Offline TA/TO (28%). Direct bookings and groups contribute significantly less, whilst Corporate, Complementary, and Aviation channels have little influence on revenue. The dependence on OTAs and agents indicates a potential to enhance direct booking techniques, such as offering discounts or loyalty benefits.

**5 Discussion**

The study of hotel booking data provided crucial insights into customer behavior, revenue trends, and market performance that may be used to make strategic decisions. One of the main findings was that transient customers represent the majority of bookings and heavily contribute to the average daily rate (ADR), particularly during peak seasons. This indicates that they are a crucial target for advertising campaigns.

When booking patterns were studied, it became clear that bookings with long lead times were more likely to be cancelled, especially among customers of Online Travel Agents (OTA). This trend emphasizes the need to examine cancellation rules as well as incorporate more flexible or incentive-based policies for long-term bookings.   
The findings also showed that non-refundable, no deposit bookings have a lower cancellation rate, emphasizing the need of motivating customers to consider these alternatives through strategic pricing or package deals.

The analysis of market channels demonstrates that Online Travel Agencies (Online TA) are the primary source of bookings, accounting for 47.30% of total market share and 40.19% of total revenue. This suggests that the property is significantly relying on third-party web channels for customer acquisition and visibility.   
Offline Travel Agencies/Tour Operators (Offline TA/TO) also play a significant part, accounting for 20.29% of bookings and 28.33% of income, highlighting the significance of maintaining good relationships with conventional travel agencies and tour operators.

Groups and direct bookings proceed, with 16.59% and 10.56% of the market share, respectively. Although direct bookings now account for just 12% of total revenue, enhancing direct booking channels (e.g., official website, contact center, loyalty programs) might lead to enhanced profitability, as direct bookings often have less commission costs than OTAs.   
It is crucial to highlight that channels such as Corporate and Complementary bookings have minimal effect, accounting for a very small amount of both booking volume and revenue.

The significant dependence on Online TA has both advantages and disadvantages. While it provides enormous visibility and accessibility to a wide audience, it also experiences significant commission costs, which may have an influence on total profit margins. As a result, a balanced strategy that continues to employ Online TA platforms while aggressively seeking to increase direct bookings has the potential to improve long-term financial success and customer loyalty.   
Furthermore, the research reveals that investing in corporate partnerships, enhancing group sales efforts, and targeting specific offline and online segments with tailored marketing initiatives.

An interesting finding from the study is the low percentage of repeat guests, indicating that the hotel is now depending mainly on first-time stays.   
This is an enormous opportunity for growth: by focusing on increasing guest satisfaction, introducing loyalty programs, and providing personalized benefits for future stays, the hotel can develop a wider base of repeat customers.  
Increasing the number of returning clients not only improves revenue, but also eliminates the marketing costs engaged with acquiring new customers, leading in more sustainable, long-term growth.

The geographical distribution of bookings shows that Kuwait has the most guests (38%), followed by Egypt (20%), Pakistan (17%), the United Arab Emirates (14%), and Saudi Arabia (11%).   
The concentration of bookings from certain countries represents a significant opportunity for focused advertising campaigns. Focused efforts in Kuwait, Egypt, and the UAE, where guest interaction is already high, could assist to build market presence, enhance loyalty programs, and enhance sales.  
Overall, the findings of the study present an outstanding basis for:   
  
• Optimize pricing strategies for certain customer types and market segments.   
• Boost loyalty with customized benefits.   
• Improve marketing strategies based on segment performance and demographic patterns.   
• Reduce cancellations through strengthened booking conditions and deposit benefits.

**6 Recommendations**

**6.1 Enhancing Booking Channels and Maximizing Revenue**

**6.1.1 Strengthen Direct Booking Channels**  
To promote direct bookings, the hotel should advertise its website, app, and contact center by offering additional benefits such as discounts, loyalty awards, and flexible cancellations for direct customers. This will strengthen guest loyalty, reduce dependency on online travel agencies, and increase profitability.

**6.1.2 Optimize Online TA Partnerships**  
Establish strong connections with major online travel agencies (OTAs) since they constitute most bookings and revenue. Negotiate better commission rates where possible and maintain competitive visibility through dynamic pricing and targeted promotions.

**6.1.3 Expand Corporate and Group Sales Efforts**  
Corporate and group segments are now underperforming. To increase volume in these higher-yield channels, launch ads aimed at corporate clients and event organizers offering customized packages, meeting facilities, and group discounts.

**6.1.4 Diversify the Channel Mix**  
Expanding booking sources reduces the risks of being overly reliant on one channel. Explore opportunities in emerging markets and enhance partnerships with offline travel agencies, especially in places where offline bookings remain dominant.

**6.1.5 Monitor and optimize** **Channel Performance Regularly**  
Implement a system to track booking and revenue trends by channel monthly. This will enable the hotel to respond immediately to changes in customer behavior and optimize the marketing strategy effectively.

**6.2 Marketing Strategies Recommendations**

Based on the comprehensive analysis of the hotel’s booking data, customer preferences, and revenue patterns, the following marketing strategies are recommended to optimize performance, enhance customer experience, and increase overall revenue:

**6.2.1** **Targeted Promotions for Key Customer Segments**  
The data reveals distinct booking behaviors among various customer types. It is recommended to tailor promotional offers specifically for frequent customers, loyalty members, and high-spending segments. Implementing targeted discounts or value-added packages could encourage repeat bookings and attract high-value guests.

**6.2.2 Dynamic Pricing Strategy**  
By utilizing data-driven pricing, the hotel can optimize room rates according to demand fluctuations, seasonality, and competitor pricing. This can involve offering time-sensitive discounts during off-peak periods, as well as premium pricing during high-demand seasons. The goal is to maximize revenue per available room (RevPAR) by aligning prices with customer willingness to pay.

**6.2.3 Enhanced Guest Experience and Upselling Opportunities**  
Insights into customer preferences suggest that offering personalized upselling opportunities such as room upgrades, meal packages, and premium services (e.g., spa, exclusive tours) can significantly increase revenue. It is recommended to implement a targeted upselling strategy at various touchpoints of the guest journey, from booking to check-out.

**6.2.4 Improved Marketing for Family and Accessible Rooms**  
There is a growing demand for family-friendly and accessible room options. To capture this segment, we recommend increasing visibility of these room types through specific marketing channels, such as family-oriented social media campaigns or partnerships with accessibility-focused travel agencies.

**6.2.5 Leveraging Data for Effective Advertising**  
The analysis of booking channels and customer demographics reveals the need to focus advertising efforts on the most profitable segments and channels. Paid advertisements, email marketing, and social media campaigns should be designed to target key demographics with tailored messaging that speaks to their preferences and needs.

**6.2.6 Retention Strategies for Cancellations and No-Show Reservations**  
Data also highlights a significant number of cancellations, especially for refundable reservations. To mitigate this, we recommend introducing incentives for customers to retain their bookings or choose non-refundable options at a discounted rate. Additionally, implementing stricter policies for high-risk periods can reduce the impact of no-shows on overall revenue.

**6.2.7 Seasonal and Event-based Campaigns**  
Seasonal trends and special events, both within the hotel and in the surrounding area, can be leveraged to create limited-time offers. These campaigns could include discounted rates for early bookings, special packages during local festivals, or promotional deals for off-season travel.

These strategies aim to align marketing efforts with the underlying data insights, thereby fostering increased guest satisfaction, enhanced brand loyalty, and optimized revenue generation. By integrating these approaches, the hotel can not only stay competitive but also lead in providing exceptional customer experiences and profitable growth.

**7 Conclusions**

The hotel booking studies revealed significant insights on customer behavior, revenue trends, and market dynamics. While online travel agencies presently dominate booking channels, there is significant opportunity to increase direct bookings and client loyalty. Furthermore, the lack of return guests emphasizes the need for more effective retention strategies. The hotel may promote long-term development, boost profitability, and reinforce its market position by fine-tuning its advertising strategies, optimizing relationships, and focusing on the guest experience.

**8 References**

|  |  |
| --- | --- |
| [1] | R. Anubala, “The Future of Hospitality: Predictive Analytics in Hotel Management.” |
| [2] | T. Webb, S. R. Cho, P. D. Mba, and M. P. Msc, *Customer Lifetime Value: A Data Science Approach for Hospitality Applications*. |
| [3] | P.-J. L. Michaud, “Agile software development: everything you need to know,” *Nexapp.ca*, 25-Mar-2024. |
| [4] | *Kaggle.com*. [Online]. Available: https://www.kaggle.com/code/jillanisofttech/data-anaylsis-hotel-booking-project/input. [Accessed: 29-Apr-2025]. |
| [5] | N. Antonio, A. de Almeida, and L. Nunes, “Hotel booking demand datasets,” *Data Brief*, vol. 22, pp. 41–49, 2019. |

**9 Appendix**

**9.1. Data Fields Used**

The following variables were extracted from the hotel booking dataset to support the analysis:

|  |  |  |  |
| --- | --- | --- | --- |
| Field Name | Description | Values / Examples | Data Type |
| is\_canceled | Indicates whether the booking is cancelled or not | 1 = Cancelled, 0 = Not Cancelled | Integer (0/1) |
| lead\_time | Number of days between booking and arrival | 0 to 737 days | Integer |
| arrival\_date\_year | Calendar year of arrival | 2015, 2016, 2017 | Integer |
| arrival\_date\_month | Month of arrival | Jan, Feb, ..., Dec | Text |
| arrival\_date\_week\_number | Week number of arrival | 1 to 53 | Integer |
| arrival\_date\_day\_of\_month | Day of arrival | 1 to 31 | Integer |
| total\_stay | Total days booked to stay | 1 to 69 | Integer |
| stays\_in\_weekend\_nights | Weekend nights booked | 0 to 19 | Integer |
| stays\_in\_week\_nights | Week nights booked | 0 to 50 | Integer |
| is\_family | Whether the booking is by a family | 0 = Not Family, 1 = Family | Integer (0/1) |
| total\_guests | Total number of guests in booking | 0 to 55 | Integer |
| adults | Number of adults in booking | 0 to 55 | Integer |
| children | Number of children in booking | 0 to 10 | Integer |
| babies | Number of babies in booking | 0 to 10 | Integer |
| meal | Type of meal plan booked | BB, HB, SC, FB, Undefined | Text |
| country | Country of origin (ISO code) | PRT, GBR, USA, etc. | Text |
| market\_segment | Source of booking | Corporate, Direct, Online TA, Offline TA/TO, Groups | Text |
| distribution\_channel | Channel of distribution | TA/TO, Direct, GDS, Corporate, Undefined | Text |
| is\_repeated\_guest | Whether guest is repeated | 0 = Not Repeated, 1 = Repeated | Integer (0/1) |
| previous\_cancellations | Number of previous cancellations | 0 to 26 | Integer |
| previous\_bookings\_not\_canceled | Number of previous non-canceled bookings | 0 to 72 | Integer |
| reserved\_room\_type | Type of room reserved (code) | A, D, E, F, G | Text |
| assigned\_room\_type | Type of room assigned (code) | A, D, E, F, G | Text |
| booking\_changes | Number of booking changes | 0 to 21 | Integer |
| refunded\_amount | Amount refunded (SAR) | 0, 600, 14,000 | Float |
| deposit\_type | Type of deposit | No Deposit, Non-Refundable, Refundable | Text |
| agent | Travel agency ID | 0 = No agency, else positive integer | Integer |
| company | Company ID for booking | 0 = No company, else positive integer | Integer |
| days\_in\_waiting\_list | Days on waiting list | 0 to 391 | Integer |
| review | Customer rating of stay | 0 to 10 | Integer |
| amount\_paid | Total amount paid (SAR) | Non-negative values | Float |
| assigned\_room\_full\_name | Full name of assigned room | "Superior Room", "Standard Room", etc. | Text |
| customer\_type | Classification of customer | Contract, Transient, Transient-party, Group | Text |
| adr | Average Daily Rate (revenue per room) | Non-negative values (SAR) | Float |
| required\_car\_parking\_spaces | Number of car parking spaces requested | 0, 1, 2, etc. | Integer |
| total\_of\_special\_requests | Total special requests made | 0, 1, 2, etc. | Integer |
| reservation\_status | Current status of reservation | Cancelled, Check-out, No-show | Text |
| reservation\_status\_date | Date when reservation status last updated | Date format (YYYY-MM-DD) | Date |
| arrival\_date | Guest arrival date | Date format (YYYY-MM-DD) | Date |
| lead\_time\_group | Grouped lead time | Short, Medium term, Long term, Very long term | Text |

**9.2 Formulas and Calculations Used**

To calculate some of the columns in this dataset, a set of formulas and logical operations were used. Each formula is explained below along with a brief description of its purpose.

**9.2.1 Total Stay**

Formula: =IF(I + H = 0, 1, I + H)

1. **I** = Stays in week nights
2. **H** = Stays in weekend nights

**Description:** Calculates the total number of nights a guest stays by adding both weekday and weekend stays. If both are zero, we consider a minimum stay of 1 night.

**9.2.2 Is Family**

Formula: =IF(K >= 2, 1, 0)

1. **K** = Total guests

**Description:** Classifies bookings as "Family" if the total number of guests is 2 or more, otherwise it is not considered a family booking.

**9.2.3 Total Guests**

Formula: =N + M + L

1. **N** = Babies
2. **M** = Children
3. **L** = Adults

**Description:** Calculates the total number of guests per booking by summing up adults, children, and babies.

**9.2.4 Reviews**

Formula: =MIN(10, MAX(1, IF(A=1, 2, 6) + IF(W=V, 1, 0) + IF(AJ>0, 1, 0) - IF(X>2, 1, 0) - IF(Y>0, 2, 0)))

1. **A** = Is cancelled
2. **W** = Assigned room
3. **V** = Reserved room
4. **AJ** = Total special requests
5. **X** = Booking changes
6. **Y** = Refunded amount

**Description:** Simulates a review score between 1 and 10 based on several booking factors like cancellation status, room match, number of changes, special requests, and refunds.

**9.2.5 Amount Paid**

Formula: = =IF(AND(OR(Z="No Deposit",Z="Refundable"),A=1),0,(VLOOKUP(AF,roompricing,2,FALSE)+IF(O="BB",80,IF(O="HB",180,IF(O="FB",280,IF(O="AI",400,0)))))\*G\*1.15)

1. **A** = Is cancelled
2. **Z** = Deposit type
3. **AF** = Reserved room full name
4. **O** = Meal type
5. **G** = Total stay
6. **roompricing** = Room prices and full name table

**Description:** This formula calculates the total amount paid by a guest by first retrieving the base price of the reserved room using a VLOOKUP from the 'roompricing' table. It then adds the price of the selected meal plan based on the meal type. After summing the room price and meal cost, the total is multiplied by the number of stay nights (G) and adjusted by a 15% tax. If the reservation was canceled and the deposit type was either "No Deposit" or "Refundable," the total amount is set to zero.

**9.2.6 ADR (Average Daily Rate)**

Formula: =IF(A = 1, 0, VLOOKUP(TRIM(AF), roompricing, 2, FALSE))

1. **A** = Is cancelled
2. **AF** = Reserved room full name
3. **roompricing** = Table with room prices and full names

**Description:** Calculates the average daily rate (ADR) based on the reserved room type by looking up its price from the 'roompricing' table.  
If the booking is canceled (A = 1), the ADR is automatically set to **0**.  
The room name in AF is trimmed to remove any extra spaces before performing the lookup to ensure accurate matching.

**9.2.7 Lead Time Group**

Formula: =IF(B <=30,"Short Term", IF(B <= 90,"Medium Term", IF(B <= 180,"Long Term","Very Long Term")))

1. **B** = Lead time (number of days between booking date and arrival date)

**Description:** Categorizes bookings based on the lead time into Short Term, Medium Term, Long Term, and Very Long Term groups to help understand booking patterns.

**9.2.7 Refunded Amount**

Formula: =IF(AND(OR(Z="Refundable", Z="No Deposit"), A=1), AF, 0)

1. **Z** = Deposit type
2. **A** = Is cancelled
3. **AF** = Amount Paid

**Description:** Calculates the amount refunded to the customer. If the deposit type is "Refundable" or "No Deposit" and the booking was canceled, the refunded amount equals the amount paid. Otherwise, it is set to zero.

**9.3 Data Cleaning Steps**

1. Removed records with missing or invalid values in key fields like Lead Time, Amount Paid, and ADR to ensure data reliability.
2. Replaced "#VALUE!" and other formula errors by cross-checking calculations and ensuring correct data types (e.g., dates, numbers).
3. Deleted unnecessary columns, such as the "Hotel" field which contained repetitive information, to streamline the dataset and focus on relevant variables.
4. Added calculated columns, including "Amount Paid" and "Refunded Amount," to align the dataset with the analysis objectives.
5. Standardized naming conventions for categorical fields such as Customer Type and Market Segment to maintain consistency.
6. Checked for duplicate entries and removed redundant bookings to prevent data skew.
7. Reviewed numerical fields to treat or remove significant outliers that could distort results.
8. Used the **TRIM** function to remove extra spaces from textual fields, ensuring clean and consistent values for accurate lookups and comparisons.

**9.4 Tools and Methods**

1. Excel Pivot Tables and Pivot Charts were used extensively for aggregating data and creating visual insights.
2. Applied formulas such as VLOOKUP, and logical IF conditions to calculate key performance indicators.
3. Conditional Formatting was implemented to highlight anomalies, trends, and important metrics during analysis.

**9.5 Supporting Tables and Charts:**

The charts and tables in this section were utilized to support the study by aiding in both the planning and analysis of the overall project.

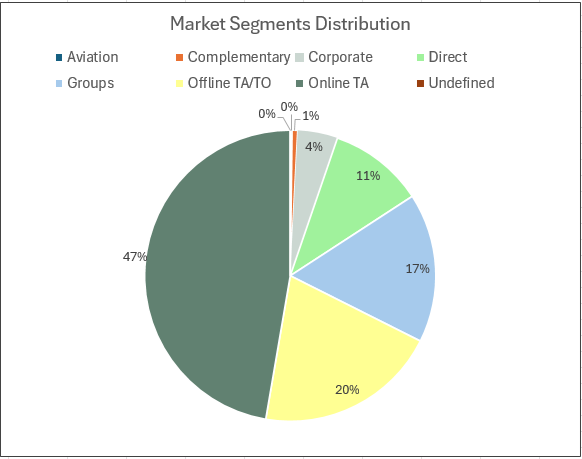


Figure 9.1: Market Segments Distribution

The pie chart in Fig 9.1 illustrates the distribution of different market segments in hotel bookings. Most bookings (47%) are made through Online Travel Agencies (Online TA), followed by Offline TA/TO (20%) and Groups (17%). Direct bookings account for 11%, while Corporate, Complementary, Aviation, and Undefined segments represent smaller proportions, each less than 5%.

~~صورة تحتوي على نص, لقطة شاشة, الخط, رقم

تم إنشاء الوصف تلقائياً~~

Figure 9.2: Room Pricing Table

Figure 9.2 shows a table listing the full room names according to the hotel classifications, along with their corresponding pricing. Each price was determined based on the standard hotel rates in Saudi Arabia.

صورة تحتوي على نص, لقطة شاشة, التصميم, مستطيل

تم إنشاء الوصف تلقائياً

Figure 9.3: Guests Category Analysis

The Fig 9.3 presents a 3D column chart which shows the total number of guests by category. Adults represent the vast majority with a total of 221,636, followed by children with 12,403, and babies with only 949. This highlights that hotel bookings are predominantly made by or for adults.