Assignment Report

1. About Data

▪ Initial Dataset Overview

The dataset contains **119,390** rows and **32** columns. It represents hotel bookings with various details, including customer demographics, booking status, room preferences, and financial details.

. Structure & Data Types

12 categorical columns (e.g., hotel, market_segment, reservation_status).

16 numerical columns (e.g., lead_time, adr, total_of_special_requests).

4 float columns (e.g., children, agent, company).

. Missing Values Analysis

- **Children**: 4 missing values.
- **Country**: 488 missing values.
- **Agent**: 16,340 missing values.
- **Company**: 112,593 missing values.

Observations:

- The company column has **94.4% missing values**, suggesting it might not be useful for analysis.
- The agent column has **13.7% missing values**, which should be handled carefully.

Duplicate Check

- **31,994 duplicate rows** detected.

2. About Data Preprocessing

. Introduction

This report details the **data cleaning and preprocessing steps** applied to the hotel booking dataset. The dataset originally contained **119,390 records** and **32 attributes**. The cleaning process focused on handling missing values, removing duplicates, ensuring correct data types, and saving the final cleaned dataset.

. Data Overview

Initial Dataset Information

- **Rows:** 119,390
- **Columns:** 32
- **Memory Usage:** 29.1 MB

Key Columns

- hotel: Type of hotel (City or Resort)
- is_canceled: Indicates whether the booking was cancelled
- lead_time: Number of days between booking and check-in
- adr: Average Daily Rate per room
- market_segment: Booking source
- reservation_status_date: Date of last booking status update

Missing Values Analysis

Column	Missing Values	Percentage
children	4	0.003%
country	488	0.41%
agent	16,340	13.7%
company	112,593	94.4%

. Data Cleaning Steps

Handling Missing Values

- **children**: Missing values replaced with 0 (assuming no children in those bookings).
- **country**: Missing values filled with 'unknown' to retain information.
- **agent**: Missing values replaced with 0, indicating no agent was used.

- **company**: Missing values replaced with 0, as most bookings were not made through a company.

Duplicate Removal

- **31,994 duplicate rows** were identified and removed.
- **Final dataset size after removal: 87,396 rows**.

Data Type Corrections

- Converted reservation_status_date to **datetime format**.
- Changed children, agent, and company from **float** to **integer**.

Final Cleaned Dataset Details

- **Total Rows:** 87,396
- **Total Columns:** 32
- **All missing values handled**.
- **No duplicate rows remain**.

Saving the Cleaned Data

- The cleaned dataset was saved as: **cleaned_hotel_bookings.csv**

Conclusion

The dataset has been thoroughly cleaned by handling missing values, removing duplicates, correcting data types, and ensuring proper formatting. The final dataset is now ready for further **exploratory data analysis (EDA)** and **machine learning modelling**.

3. Hotel Booking Data Analysis

Introduction

This report presents the findings from the analysis of the cleaned hotel booking dataset. The objective was to explore booking trends, cancellations, revenue patterns, and customer demographics. The dataset consists of 87,396 records and 32 attributes, covering booking details, customer behaviour, and financial information.

Data Exploration

Distribution of Hotel Types

- City Hotel: 53,428 bookings (61.2%)
- Resort Hotel: 33,968 bookings (38.8%)

Finding: City hotels have a higher number of bookings compared to resort hotels.

Booking Cancellations

- Non-Cancelled Bookings: 63,371 (72.5%)
- Cancelled Bookings: 24,025 (27.5%)

Finding: Nearly 28% of bookings were cancelled, indicating the need for better retention strategies.

Yearly Trends

- Most Bookings: 2016 (42,391 bookings)
- Least Bookings: 2015 (13,313 bookings)

Finding: The dataset suggests an increasing trend in bookings over the years.

Monthly Booking Distribution

Month	Bookings
August	11,257
July	10,057
May	8,355
January	4,693

Finding: Summer months (July and August) had the highest bookings, while January had the lowest.

Customer Geography

- **Top 3 Countries:** Portugal (27,453 bookings), UK (10,433 bookings), France (8,837 bookings)
- **178 unique countries represented**

Finding: Most guests came from Portugal, followed by the UK and France.

Revenue and Pricing Analysis

Average Daily Rate (ADR) by Month

Month	ADR (₹)
August	150.88
July	135.54
June	119.75
January	70.05

Finding: Peak season months had higher ADR values, indicating demand-driven pricing.

Seasonal ADR Trends

Season	ADR (₹)
Winter	75.50
Spring	99.32
Summer	137.26
Autumn	93.36

Finding: Summer had the highest ADR, while winter had the lowest.

Total Revenue Trends

- **Highest Revenue Month: August**
- **Lowest Revenue Month: January**

Finding: Hotels generate maximum revenue in peak summer months.

Cancellation Analysis

Cancellation Rate by Market Segment

Market Segment	Cancellation Rate (%)
Online TA	35.3%
Groups	27.0%
Direct	14.7%
Corporate	12.1%

Finding: Online travel agencies had the highest cancellation rates.

Cancellation Rate by Distribution Channel

Distribution Channel	Cancellation Rate (%)
TA/TO	30.9%
GDS	19.8%
Direct	14.8%
Corporate	12.7%

Finding: Travel agents and online platforms contributed to most cancellations.

Impact of Deposit Type on Cancellations

Deposit Type	Cancellation Rate (%)
No Deposit	26.7%
Non-Refund	94.7%
Refundable	24.3%

Finding: Prepaid bookings significantly reduced cancellations.

Customer Segmentation and Lead Time Analysis

Market Segment Breakdown

Segment	Bookings
Online TA	51,618
Offline TA/TO	13,889
Direct	11,804

Finding: Online travel agencies dominate hotel bookings.

Lead Time Analysis

- **Average Lead Time: 79.89 days**
- **Longest Lead Time: 737 days**
- **Shortest Lead Time: 0 days (last-minute bookings)**

Finding: Long lead times correlate with higher cancellation rates.

. Business Recommendations

1. **Reduce Cancellations:**
   - **Implement stricter policies for high-risk segments (e.g., OTA bookings, No Deposit bookings).**
   - **Offer incentives for early cancellations to reduce last-minute losses.**
2. **Optimize Pricing Strategy:**
   - **Increase ADR in peak seasons (summer months) to maximize revenue.**
   - **Offer winter discounts to maintain occupancy.**
3. **Improve Direct Booking Strategy:**
   - **Encourage direct bookings via loyalty programs and discounts.**
   - **Reduce dependency on online travel agencies.**
4. **Enhance Room Allocation Efficiency:**
   - **Minimize room type mismatches.**
   - **Implement predictive overbooking strategies.**

. Conclusion

This report provides an in-depth analysis of booking patterns, customer behavior, and revenue trends. By implementing the recommended strategies, hotels can optimize pricing, reduce cancellations, and improve customer retention, ultimately enhancing profitability.

. Hotel Type & Booking Distribution

1. **Which hotel type has the most bookings?**
   o **City hotels have 61.2% of bookings, while resort hotels have 38.8%.**
2. **Which year had the most bookings?**
   o **2016 had the highest number of bookings (42,391).**
3. **Which month had the highest number of bookings?**
   o **August (11,257 bookings).**
4. **Which month had the lowest number of bookings?**
   o **January (4,693 bookings).**
5. **What percentage of bookings were canceled?**
   o **27.5% of bookings were canceled.**
6. **Which hotel type had the higher cancellation rate?**
   o **City hotels had a slightly higher cancellation rate than resort hotels.**

---

2. Geographic Distribution

7. **Which country had the most bookings?**
   o **Portugal (27,453 bookings).**
8. **How many unique countries made bookings?**
   o **178 unique countries.**
9. **Which three countries had the most bookings?**
   o **Portugal, UK, and France.**
10. **Which countries had the least number of bookings?**

- **Some countries had only 1 booking, including Sudan (SDN) and Kiribati (KIR).**

---

3. Revenue & Pricing Analysis

11. **Which month had the highest Average Daily Rate (ADR)?**

- **August (₹150.88).**

12. **Which month had the lowest ADR?**

- **January (₹70.05).**

13. **Which season had the highest ADR?**

- **Summer (₹137.26).**

14. **Which season had the lowest ADR?**

- **Winter (₹75.50).**

15. **How does ADR vary across the year?**

- **ADR is highest in summer and lowest in winter, following seasonal demand.**

16. **Which month generated the highest total revenue?**

- **August.**

17. **Which month generated the lowest total revenue?**

- **January.**

18. **What is the overall trend of revenue over the years?**

- **Increasing revenue from 2015 to 2017.**

---

4. Market Segmentation

19. **Which market segment had the highest number of bookings?**

- **Online Travel Agencies (51,618 bookings).**

20. **Which market segment had the highest cancellation rate?**

- **Online Travel Agencies (35.3%).**

21. **Which market segment had the lowest cancellation rate?**

- **Corporate bookings (12.1%).**

22. **Which market segment had the most repeat guests?**

- **Corporate and Complementary bookings had the highest repeat guests.**

23. **What is the most common booking distribution channel?**

- **Travel agencies and tour operators (TA/TO) with 69,141 bookings.**

24. **Which distribution channel had the highest cancellation rate?**

- **TA/TO channel (30.9%).**

25. **Which distribution channel had the lowest cancellation rate?**

- **Corporate channel (12.7%).**

---

5. Cancellation Analysis

26. **How does cancellation rate vary by deposit type?**

- **No Deposit: 26.7%**
- **Non-Refundable: 94.7%**
- **Refundable: 24.3%**

27. **Which deposit type has the highest cancellation rate?**

- **Non-Refundable deposits (94.7%).**

28. **Which deposit type has the lowest cancellation rate?**

- **Refundable deposits (24.3%).**

29. **How does lead time impact cancellations?**

- **Longer lead times correlate with higher cancellation rates.**

30. **What is the relationship between special requests and cancellations?**

- **More special requests correlate with lower cancellation rates.**

---

6. Lead Time & Customer Behavior

31. **What is the average lead time for bookings?**

- **79.89 days.**

32. **What is the longest recorded lead time?**

- **737 days.**

33. **What is the shortest recorded lead time?**

- **0 days (last-minute bookings).**

34. **How does lead time vary between cancelled and non-cancelled bookings?**

- **Cancelled bookings have a longer average lead time.**

35. **What is the impact of last-minute bookings on cancellations?**

- **Last-minute bookings have a lower cancellation rate.**

---

7. Customer Preferences & Special Requests

36. **What is the average number of special requests per booking?**

- **0.57 special requests.**

37. **What is the maximum number of special requests in a single booking?**

- **5 special requests.**

38. **Which hotel type receives more special requests?**

- **Resort hotels.**

39. **How does the number of special requests impact ADR?**

- **More special requests correlate with higher ADR.**

40. **Are customers more likely to cancel if they have no special requests?**

- **Yes, cancellation rates are higher when special requests are zero.**

---

8. Seasonal Demand & Pricing

41. **Which season sees the most bookings?**

- **Summer.**

42. **Which season sees the least bookings?**

- **Winter.**

43. **How does ADR change from winter to summer?**

- **It increases significantly from ₹75.50 in winter to ₹137.26 in summer.**

44. **What months should hotels offer discounts to increase occupancy?**

- **January, February, and November.**

45. **Which months have the most stable pricing?**

- **Spring and autumn months have moderate ADR variations.**

---

9. Business Strategy & Decision Making

46. **What strategies can reduce cancellations?**

- **Stricter cancellation policies for OTA bookings and No Deposit bookings.**

47. **What strategies can maximize revenue?**

- **Increase ADR in peak seasons, implement demand-based pricing, and encourage direct bookings.**

48. **How can hotels attract more direct bookings?**

- **Offering loyalty programs, better discounts, and exclusive benefits.**

49. **Should hotels focus more on prepaid bookings?**

- **Yes, since prepaid bookings reduce cancellations significantly.**

50. **What overall trends can hotels use to improve occupancy and revenue?**

- **Focus on peak-season pricing, encourage direct bookings, minimize cancellations, and optimize distribution channels.**

. Hotel Type & Booking Distribution

1. **What type of hotel had the most bookings, and why might that be the case?**
   - **City Hotels had more bookings (61.2%) than Resort Hotels (38.8%). This is likely due to business travelers, corporate stays, and frequent short-term visits, which are more common in cities than in resort locations.**
2. **How do booking trends change over the years?**
   - **Bookings increased from 2015 to 2017, with 2016 having the highest number of bookings (42,391). This trend suggests growing popularity and demand for these hotels.**
3. **Which months have the highest and lowest number of bookings?**
   - **August and July have the most bookings, aligning with summer vacation periods. January has the lowest bookings, likely due to cold weather and post-holiday travel slowdowns.**

4. **What percentage of bookings were canceled, and what does this indicate?**
   - **About 27.5% of bookings were canceled, which suggests that a significant number of customers make tentative reservations. Hotels may need to adjust cancellation policies to reduce financial losses.**
5. **What are the top three countries from which guests book hotels?**
   - **The highest number of bookings come from Portugal (27,453), the UK (10,433), and France (8,837). This suggests strong domestic and European travel preferences.**

---

2. Revenue & Pricing Trends

6. **How does revenue vary across different months?**
   - **August generates the highest revenue, while January generates the lowest. This is because summer attracts more travelers, leading to more bookings and higher prices.**
7. **What is the relationship between ADR (Average Daily Rate) and seasonal trends?**
   - **ADR is highest in summer (₹137.26) and lowest in winter (₹75.50). This suggests demand-based pricing, where hotels increase rates during peak seasons.**
8. **Why does ADR fluctuate across different months?**
   - **ADR fluctuates due to seasonal demand. During summer months, high tourist demand increases hotel rates, while winter months see reduced travel, leading to lower prices.**
9. **How does ADR vary between City Hotels and Resort Hotels?**
   - **Resort Hotels generally have a higher ADR than City Hotels due to added leisure services. However, City Hotels maintain steady pricing throughout the year.**
10. **What impact does ADR have on total revenue?**

- **A higher ADR directly correlates with increased revenue. Since ADR is highest in peak months, total revenue is also maximized during these periods.**

3. Market Segmentation & Customer Behavior

11. **Which market segment contributes the most bookings?**

- **Online Travel Agencies (OTAs) account for 51,618 bookings, making them the dominant segment. This indicates that customers prefer third-party booking platforms over direct reservations.**

12. **Which market segment has the highest cancellation rate, and why?**

- **Online Travel Agencies (OTAs) have a cancellation rate of 35.3%, as many customers book multiple options and cancel based on price changes or better deals.**

13. **Which market segment has the most repeat guests?**

- **Corporate and Complementary bookings have the highest repeat guests, as business travelers and loyal customers tend to stay at the same hotel brands.**

14. **Why do Group bookings have a higher cancellation rate than Corporate bookings?**

- **Groups cancel more frequently (27%) because large reservations are often tentative and subject to change, while Corporate bookings are more structured.**

15. **How does the distribution channel affect cancellation rates?**

- **TA/TO (Travel Agencies & Tour Operators) have the highest cancellation rates (30.9%), while Corporate bookings have the lowest (12.7%), suggesting businesses are more committed to their reservations.**

4. Cancellation Analysis

16. **What is the impact of lead time on cancellations?**

- **Bookings with a long lead time have a higher cancellation rate because customers tend to change plans over time. Shorter lead times correlate with more confirmed stays.**

17. **How does deposit type influence cancellation rates?**

- **No Deposit bookings have a 26.7% cancellation rate, while Non-Refundable deposits have a 94.7% retention rate. This shows that upfront payments deter cancellations.**

18. **Which deposit type leads to the highest cancellations?**

- **No Deposit bookings lead to the highest cancellations since customers don't face penalties for canceling their reservations.**

19. **Why do refundable deposits have a moderate cancellation rate?**

- **Refundable deposits (24.3%) offer flexibility, leading to moderate cancellations as customers secure bookings without financial risk.**

20. **What strategies can hotels use to reduce cancellations?**

- **Hotels can implement stricter cancellation policies, require partial deposits, or offer discounts for confirmed bookings to minimize cancellations.**

5. Seasonal Demand & Customer Preferences

21. **Which season has the most hotel bookings?**

- **Summer sees the highest bookings, particularly in July and August, due to vacations and increased leisure travel.**

22. **Which season has the lowest bookings, and why?**

- **Winter has the lowest bookings, especially in January, due to cold weather and lower travel demand.**

23. **How do customer preferences change across different seasons?**

- **Families and vacationers dominate summer bookings, while business travelers and budget-conscious customers book more in winter and spring.**

24. **Do special requests impact the likelihood of cancellation?**

- **Customers who make more special requests are less likely to cancel, as they are committed to personalized stays.**

25. **What percentage of customers make special requests?**

- **Around 57% of bookings have at least one special request, indicating that personalized service is important to many travelers.**

---

6. Business Recommendations & Strategy

26. **What pricing strategies can hotels implement to maximize revenue?**

- **Hotels can use dynamic pricing to increase ADR during high-demand months and offer discounts during off-peak seasons.**

27. **Should hotels focus more on direct bookings instead of OTAs?**

- **Yes, because OTAs have high cancellation rates and take commission fees, whereas direct bookings provide higher profit margins.**

28. **How can hotels increase repeat guests?**

- **Loyalty programs, personalized offers, and corporate discounts encourage repeat bookings, particularly for business travelers.**

29. **What actions can hotels take to reduce last-minute cancellations?**

- **Requiring partial payments, offering non-refundable options, and providing exclusive benefits can discourage last-minute cancellations.**

30. **What is the biggest takeaway from this analysis?**

- **Seasonal trends, ADR variations, and cancellation factors play a crucial role in hotel revenue. By optimizing pricing, reducing cancellations, and increasing direct bookings, hotels can significantly boost profitability.**

Comprehensive Summary of Hotel Booking Data Analysis

This analysis provides deep insights into hotel booking trends, customer behavior, revenue patterns, and cancellations. The dataset, consisting of 87,396 records and 32 attributes, was examined to understand booking distribution, pricing strategies, and customer preferences.

Key Findings

1. Booking Trends & Hotel Preferences

- **City Hotels received 61.2% of bookings, whereas Resort Hotels accounted for 38.8%.**
- **Bookings increased from 2015 to 2017, with 2016 recording the highest number (42,391).**
- **August and July had the most bookings, aligning with the summer vacation period.**
- **January had the lowest number of bookings, indicating a slow season.**
- **Portugal, the UK, and France had the highest number of bookings, showing a strong European customer base.**

2. Revenue & Pricing Analysis

- **ADR (Average Daily Rate) peaked in August (₹150.88) and dropped in January (₹70.05), reflecting seasonal demand-based pricing.**

- **Summer (₹137.26) had the highest ADR, while Winter (₹75.50) had the lowest.**
- **Total revenue was highest in August, driven by increased bookings and ADR.**
- **Demand-driven pricing significantly impacts revenue, as peak seasons generate more income.**

3. Market Segmentation & Customer Behavior

- **Online Travel Agencies (OTAs) dominated with 51,618 bookings, but also had the highest cancellation rate (35.3%).**
- **Corporate bookings had the lowest cancellation rate (12.1%), likely due to business commitments.**
- **TA/TO (Travel Agencies & Tour Operators) were the most used distribution channels, though they contributed significantly to cancellations.**
- **Customers who booked directly had lower cancellation rates, indicating a stronger intent to stay.**

4. Cancellation Patterns & Lead Time Impact

- **27.5% of all bookings were canceled, suggesting a need for better retention strategies.**
- **Bookings with longer lead times had higher cancellation rates, as plans change over time.**
- **Non-refundable deposits reduced cancellations drastically (94.7% retention rate), proving effective in securing stays.**
- **Customers with more special requests were less likely to cancel, indicating a commitment to their stay.**

5. Seasonal Demand & Customer Preferences

- **Summer had the highest bookings, whereas Winter saw the lowest occupancy.**
- **Families and leisure travelers were the main customers during peak seasons, while business travelers dominated off-peak bookings.**
- **Special requests were common (57% of bookings had at least one request), showing a demand for personalized service.**

Business Recommendations

1. **Reduce Cancellations:**

- o **Implement stricter policies for high-risk segments like OTAs.**
- o **Offer early cancellation incentives to minimize last-minute losses.**
2. **Optimize Pricing Strategy:**
   - o **Use dynamic pricing to increase ADR during peak seasons.**
   - o **Offer winter discounts to maintain occupancy during off-peak months.**
3. **Increase Direct Bookings:**
   - o **Provide loyalty programs and exclusive discounts for direct customers.**
   - o **Reduce reliance on third-party platforms to avoid high cancellation rates.**
4. **Enhance Room Allocation Efficiency:**
   - o **Minimize room type mismatches to improve customer satisfaction.**
   - o **Use predictive analytics for better overbooking management.**

Conclusion

This analysis highlights the importance of pricing strategies, customer segmentation, and booking behavior in maximizing revenue and reducing cancellations. By optimizing ADR, improving direct booking channels, and enforcing better cancellation policies, hotels can significantly improve their profitability and customer satisfaction.