Report

🏨 Project Name: Exploratory Data Analysis on Hotel Data

📁 Project Summary

This project analyzes a hotel booking dataset containing records from two hotel types — **City Hotel** and **Resort Hotel** — spanning three years (2015–2017). The dataset includes detailed information on booking dates, length of stay, cancellations, guest demographics (including children and babies), revenue (adr), special requests, food preferences, and booking channels (agents and distribution platforms). It also provides country-wise booking data, enabling geographic analysis.

The data is segmented by hotel type, allowing for comparative analysis between city and resort hotels. Before visualization, the dataset undergoes thorough **data cleaning** (handling nulls and duplicates) and **data wrangling** to prepare it for meaningful insights.

The analysis is conducted both **individually for each hotel** and **comparatively**, using charts and dashboards to highlight trends and performance metrics.

🎯 Problem Statement

Hotels face ongoing challenges in maintaining profitability and operational efficiency due to the complex dynamics of:

- Revenue management
- Occupancy optimization
- Booking cancellations
- Guest preference analysis
- Dynamic pricing decisions

High cancellation rates — especially in city hotels — disrupt forecasting and resource planning. Limited understanding of guest behavior restricts upselling and loyalty-building opportunities. Pricing decisions are further complicated by fluctuating demand, reliance on third-party booking platforms, and competitive pressures.

Without a **data-driven strategy** that integrates booking trends, customer segmentation, and operational metrics, hotels risk:

- Revenue leakage
- Inefficient resource allocation
- Missed opportunities to improve guest satisfaction and brand loyalty

---

About dataset

1. In total we got to know that we have 119390 columns and 32 rows in our data set.
2. Total duplicates values count is of 31994
3. After removing duplicates we got 87396 cleaned records
4. There are four columns in dataset that has the null values they are company, agent, country, children
5. Identified that there are two columns stay_in_week_nights and stay_in_weekend_nights so we need to add both the column to get the Total_number_of_days
6. Identified that there are there column as Adults, children, babies so we need to add the values to get the total_guests

---

Variable description

1. Booking ID = PK it will identify every record uniquely
2. Hotel = which type of hotel is this (City, Resort)
3. is_cancled = cancellation status (0 = not cancled, 1= cancled after booking)
4. lead_time = Number of days before arrival of guests
5. Arrival_date_year= it is the arrival year
6. Arrival_date_month= it is the arrival month
7. Arrival_date_week_number = it is the arrival week
8. Arrival_date_day_of_month = it is the arrival day/date
9. stay_in_weekend_nights = It tell us how many people stayed on weekend.
10. stay_in_week_nights = It tell us how many people stayed during weekend.
11. Adults = This tell how many adults stayed in hotel
12. Children = How many children among the guests

13. Babies = Total babies amont guests
14. Meal = Type of meal choosen by guests
15. Country = From which country guest belong
16. Market_segment = Guest belong from which market segment (Online TA, Offline TA/TO, Groups, Direct, Corporate, Compliementary, Aviation)
17. Distribution_channel = Name of booking distribution channel (TA/TO, Direct, corporate, GDS, Undefined)
18. is_repeated_guests = It show the status that if the guest revisited the hotel (0 = not repeated, 1= Repeated)
19. previous_cancellation = If the guest canceled the booking previously (0= not canceled, 1= Canceled)
20. previous_booking_not_canceled = if guest not canceled the previous booking
21. Reserved_room_type = the category of room that guest has booked for
22. assigned_room_type = the category of room the guest got after check in.
23. booking_changes = The changes that were made to the booking
24. deposit_type = in which form the guest has given the deposite (Refundable, non-refunddable, no deposite)
25. Agent = which agent has made the booking
26. company = from which company the booking was made
27. days_in_waitinng_list = Number of day the booking was in the waiting list.
28. customer_type = Type of customer (Transient, Transient-party, contract, group)
29. ADR = Average daily rate
30. required_car_parking_spaces = Number of car parking spaces required by the customer
31. total_of_special_request = number of special request made by the customer
32. Reservation_status = Reservation status(checkout, canceled, no-show)
33. reservation_status_date = date at which the last reservation status was updated

---

Manipulations made in data

1. Added column named total_stays
2. Added column named total_num_people
3. total_num_people were having some records which were zero so we have removed those records as it signifies that no bookings were made.

📊 Data Visualization and Story Telling

Question 1. Which hotel have a greater number of booking (City or Resort Hotel)

Question 2. Which year had highest bookings (City hotels or Resort Hotels)

Question 3. Which Type Of hotel have Highest average revenue or ADR (City hotels or Resort Hotels)

Question 4. What is the cancellation rate by hotel type.

Question 5. Total Guest by hotel.

Question 6. Meals Preferred by guests.

Question 7. Bookings By Market segment.

Question 8. Which month has a greater number of bookings.

Question 9. Counts of bookings by country.

Question 10. Which is the most preferred room type.

Question 11. Repeated guest by hotels

Question 12. Count of guest by customer type

Question 13. Count of booking by reservation status

Question 14. Count of repeated per hotel

Question 15. What is the parking distribution of required car parking space

Question 16. Relationship between ADR and Total stay.

Question 17. Plot correlation Heatmap

🏨 Recommendations / Solutions

1. Optimize Pricing and Promotions

- **Data Used:** adr, lead_time, arrival_date_month, market_segment
- **Solution:**
    - Implement dynamic pricing models based on seasonality and booking lead times.
    - Offer targeted promotions to segments with lower occupancy (e.g., corporate vs. leisure).
    - Use dashboards to monitor ADR trends and adjust promotions in real time.
    - Impact: Maximizes revenue by aligning room rates with demand patterns.

2. Reduce Cancellations and Pending Bookings

- **Data Used:** is_canceled, deposit_type, lead_time, previous_cancellations
- **Solution:**
    - Build predictive models to identify high-risk cancellations.
    - Introduce flexible deposit policies or incentives for guests with long lead times.
    - Track cancellation rates by country and distribution channel to mitigate risks.
    - Impact: Improves booking reliability and reduces revenue leakage.

3. Boost Inventory and Room Mix

- **Data Used:** reserved_room_type, assigned_room_type, stays_in_week_nights, stays_in_weekend_nights
- **Solution:**
    - Analyze mismatches between reserved and assigned room types to optimize allocation.
    - Forecast demand for specific room categories and adjust inventory accordingly.
    - Use dashboards to visualize occupancy trends by room type and season.
    - Impact: Ensures optimal room utilization and reduces overbooking issues.

4. Focus on Guest Satisfaction

- **Data Used:** total_of_special_requests, customer_type, reservation_status
- **Solution:**

- Track how special requests correlate with repeat bookings.
- Segment guests by type (transient, group, contract) and tailor services.
- Integrate survey/review data with booking dataset for deeper insights.
- Impact: Enhances guest experience, leading to better reviews and loyalty.

5. Retention & Loyalty

- **Data Used:** is_repeated_guest, customer_type, market_segment
- **Solution:**
  - Identify repeat guests and analyze their booking behavior.
  - Offer loyalty perks (discounts, upgrades, exclusive deals).
  - Track retention rates across distribution channels to refine loyalty programs.
  - Impact: Builds long-term customer relationships and stabilizes revenue streams.

6. Upsell Opportunities

- **Data Used:** meal, required_car_parking_spaces, total_of_special_requests
- **Solution:**
  - Design upsell packages (meal plans, parking, spa services) based on guest demand.
  - Use clustering to identify guest profiles likely to purchase add-ons.
  - Highlight upsell opportunities in dashboards (e.g., % of guests requesting extras).
  - Impact: Increases ancillary revenue beyond room bookings.

7. Operational Planning

- **Data Used:** arrival_date_week_number, arrival_date_day_of_month, stays_in_week_nights
- **Solution:**
  - Forecast occupancy by week/month to plan staffing and inventory.
  - Align housekeeping schedules with peak check-in/check-out days.
  - Use time-series analysis to anticipate demand spikes (holidays, events).
  - Impact: Improves efficiency, reduces costs, and ensures smooth operations.

🏁 Conclusion

1. Data-driven decisions for sales, operations, and marketing

- By analyzing columns like adr (pricing), lead_time (booking behavior), and market_segment (customer source), managers can make informed choices instead of relying on intuition.
- Example: Adjusting room rates during peak months or targeting promotions to segments with lower occupancy.
- Impact: Increases profitability and ensures resources are allocated efficiently.

2. Actionable insights that boost business performance

- Correlation analysis and dashboards reveal patterns such as cancellation risks (is_canceled vs. lead_time) or upsell opportunities (meal, special_requests).
- These insights can be turned into concrete actions like flexible deposit policies or bundled packages.
- Impact: Transforms raw data into strategies that directly improve revenue and guest satisfaction.

3. Dashboarding helps hotel managers gain valuable visibility into daily operations

- Visual tools (heatmaps, KPIs, Power BI dashboards) allow managers to monitor occupancy, cancellations, and revenue in real time.
- Example: A dashboard showing daily arrivals vs. cancellations helps managers adjust staffing and inventory quickly.
- Impact: Enhances decision-making speed and operational control.

4. The results support smarter pricing

- Analysis of adr across seasons and segments highlights when to increase or decrease rates.
- Predictive models can recommend optimal pricing strategies based on demand forecasts.
- Impact: Maximizes revenue while maintaining competitiveness in the market.

5. Improved guest experience

- Tracking special_requests, customer_type, and repeat bookings (is_repeated_guest) helps tailor services to guest needs.
- Example: Offering loyalty perks or personalized packages based on guest history.
- Impact: Higher satisfaction leads to better reviews, repeat visits, and stronger brand reputation.

6. Proper marketing and inventory management

- Data on market_segment, distribution_channel, and reserved_room_type helps identify which channels and room types perform best.
- Marketing campaigns can be aligned with demand trends, while inventory can be adjusted to avoid mismatches.
- Impact: Reduces wasted resources and ensures maximum occupancy.

7. Long-term strategic growth within the competitive hospitality market

- Combining all insights — pricing, cancellations, loyalty, upselling, and operations — creates a holistic strategy.
- Continuous monitoring ensures the hotel adapts to changing market conditions.
- Impact: Builds resilience, sustainable growth, and a competitive edge in the hospitality industry.