

Hotel Booking Case Study Report

Name - Omkar Suryakant Sawant

ID - 250240325045

1. Executive Summary

This case study performs exploratory data analysis (EDA) on a hotel booking dataset to uncover patterns in guest behavior, pricing strategies, and booking channels. The insights gained aim to enhance revenue management and customer satisfaction.

2. Objective

- Analyze customer behavior and booking trends
- Identify factors affecting ADR (Average Daily Rate)
- Study room allocation mismatches
- Examine patterns in cancellations and booking modifications

3. Dataset Overview

Data Source: 'hotel_bookings.csv'

Primary fields included: hotel, lead_time, stays, adr, market_segment, country, etc.

4. Data Cleaning & Preprocessing

- Dropped columns with high nulls (e.g., 'company')
- Filled missing values using mode or median as appropriate
- Removed duplicates (if any)
- Converted date columns and encoded categorical values

5. Exploratory Data Analysis

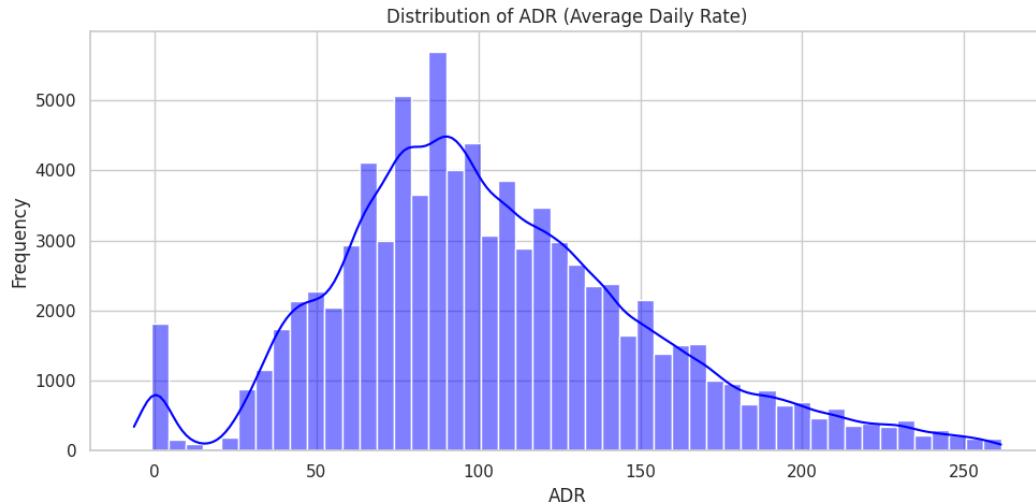


Figure 1: Summary of plotted insights.

Key Observations:

- **Peak Frequency:** Most bookings fall within the ADR range of **50 to 100**, indicating that this is the most common price range for guests.
- **Right-Skewed Distribution:** The tail of the distribution extends towards the right, showing that **some bookings have very high ADRs**, but these are less frequent.
- **Mode:** The highest bar appears around **80–90**, meaning this is the **most frequent ADR**.
- **Outliers:** There are **fewer but significant high-value ADRs** (above 200), possibly indicating luxury bookings or pricing anomalies.
- **Low ADR Values:** There's also a small spike at **ADR = 0**, which may represent complimentary stays or data entry issues.

Business Interpretation:

- The hotel earns **majority of its revenue from mid-range priced bookings**.
- The long tail of high ADRs can be **leveraged for premium offerings**, while the cluster near 0 may need **investigation or cleanup**.
- Understanding this distribution helps in **setting pricing strategies and discount policies effectively**.

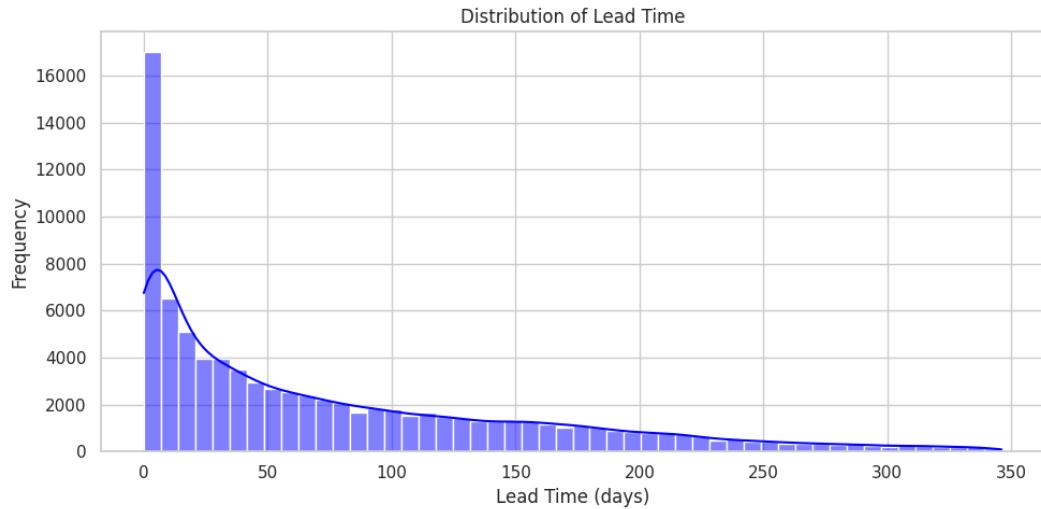


Figure 2: Summary of plotted insights.

Key Observations:

- **Highly Right-Skewed:** The distribution is extremely right-skewed, meaning **most bookings are made close to the check-in date**.
- **Peak at 0–10 Days:** A **very high frequency** of bookings is made within **0–10 days** of arrival, with a massive spike at **1–2 days**, indicating **last-minute bookings**.
- **Long Tail:** A small but consistent number of bookings occur even **over 100–300 days** in advance, suggesting **long-term planning for specific customers or events**.
- **Booking Trends:** The frequency drops sharply as lead time increases, showing **less customer tendency to book far in advance**.

Business Interpretation:

- The hotel sees a large portion of bookings at **short notice**, which might indicate **price-sensitive or urgent travelers**.
- The long tail suggests opportunities for **early-bird promotions or group/event bookings**.
- Revenue management strategies should focus on **dynamic pricing near check-in dates** due to high short-term demand.

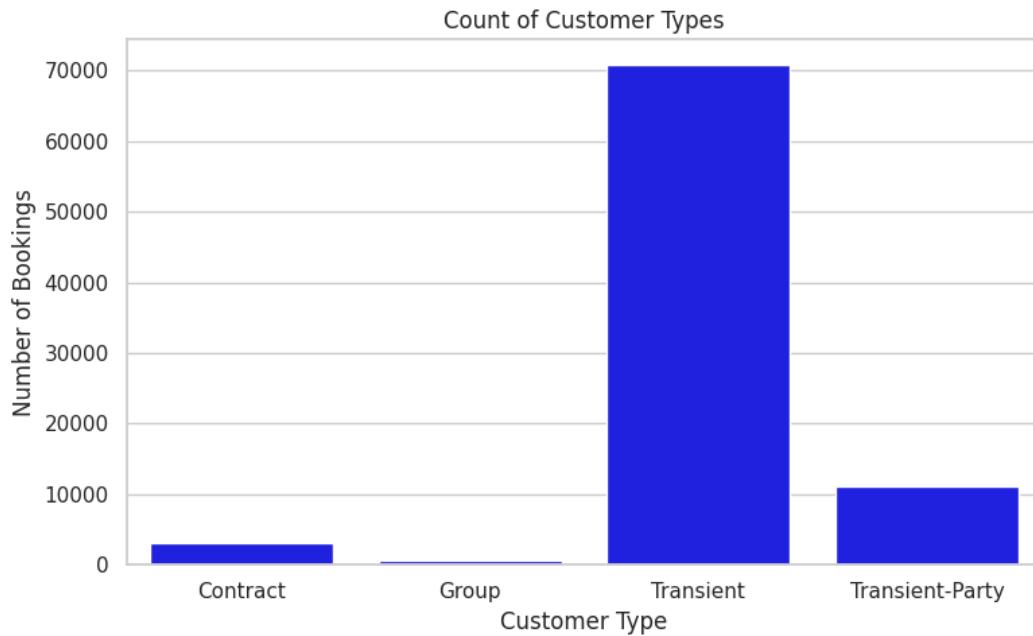


Figure 3: Summary of plotted insights.

Key Observations:

- **Dominant Segment – Transient:** Over **70,000 bookings** are from **Transient customers**, making them the **largest customer group by far**. These are typically individual leisure or business travelers.
- **Second Most Common – Transient-Party:** Around **10,000 bookings** are from **Transient-Party customers**, who likely travel in small informal groups.
- **Contract Customers:** Only a small number (~3,000) of bookings are made under **Contract agreements**, possibly from businesses or organizations.
- **Group Customers – Negligible:** Very **few group bookings**, indicating the hotel may not be a popular choice for tour operators or large event-based stays.

Business Interpretation:

- The hotel should **focus marketing and pricing strategies on transient guests**, as they form the bulk of bookings.
- There's potential to grow in **contract and group segments** through **corporate tie-ups or group discount packages**.
- Understanding the behavior of transient vs. transient-party guests can help **optimize room types, offers, and amenities** accordingly.

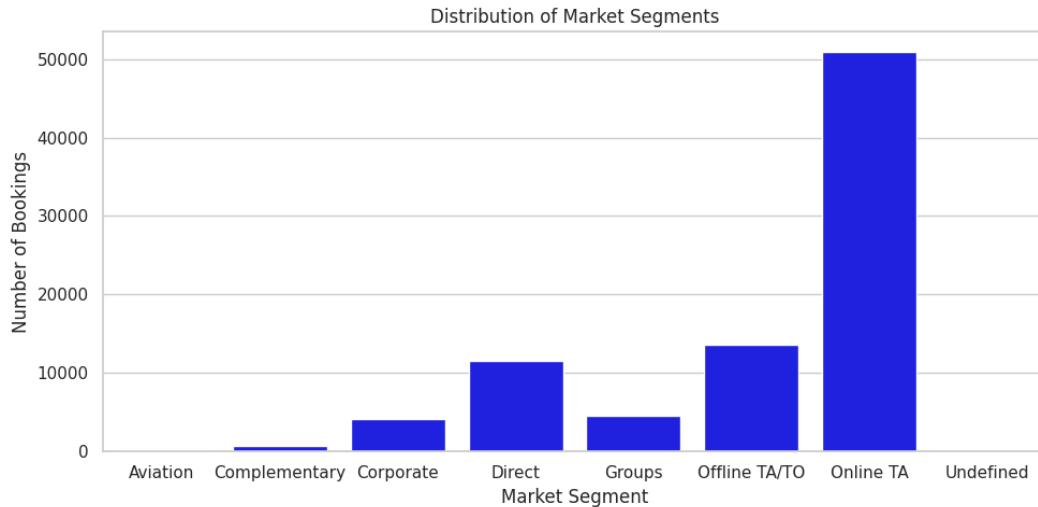


Figure 4: Summary of plotted insights.

Key Observations:

- **Online TA (Travel Agents)** dominate with over **50,000 bookings**, making it the **primary source of reservations**.
- **Offline TA/TO** and **Direct bookings** are the next most significant channels, with **around 13,000–14,000 bookings each**.
- **Corporate** and **Groups** contribute moderately (~5,000 bookings), showing some **business and event travel**.
- **Complementary** and **Aviation** segments are minimal, as expected.
- A few bookings fall under the **Undefined** segment, likely due to data entry errors or missing source information.

Business Interpretation:

- The hotel is **highly dependent on Online Travel Agencies (OTAs)**, which may incur **higher commission costs**. A shift toward **Direct bookings** can **increase profit margins**.
- The presence of bookings from **Corporate and Group segments** opens up room for **B2B strategies** or **event-based promotions**.
- Reducing reliance on "Undefined" or poorly tracked sources can improve **data quality and marketing attribution**.

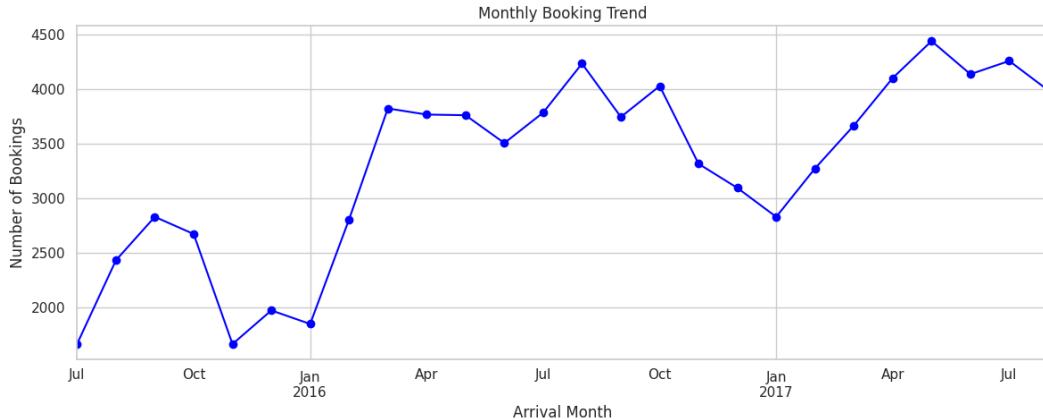


Figure 5: Summary of plotted insights.

Key Observations:

- **Seasonal Patterns:** Booking numbers clearly follow a **seasonal cycle**, with **peaks around March to August** and **dips around November to January** each year.
- **Lowest Points:** Booking activity **drops sharply in late fall and winter**, especially around **December–January**, possibly due to holidays or off-season travel.
- **Rising Trend:** Overall, there's an **upward trend** in bookings from 2015 to 2017, indicating **growing popularity or effective marketing**.

Business Interpretation:

- The hotel should **maximize revenue in peak seasons** (spring and summer) with premium pricing and packages.
- Off-season months can be targeted with **discounts, events, or bundled deals** to maintain occupancy.
- Recognizing this pattern supports **seasonal staffing, inventory planning, and promotional campaigns** to match demand.

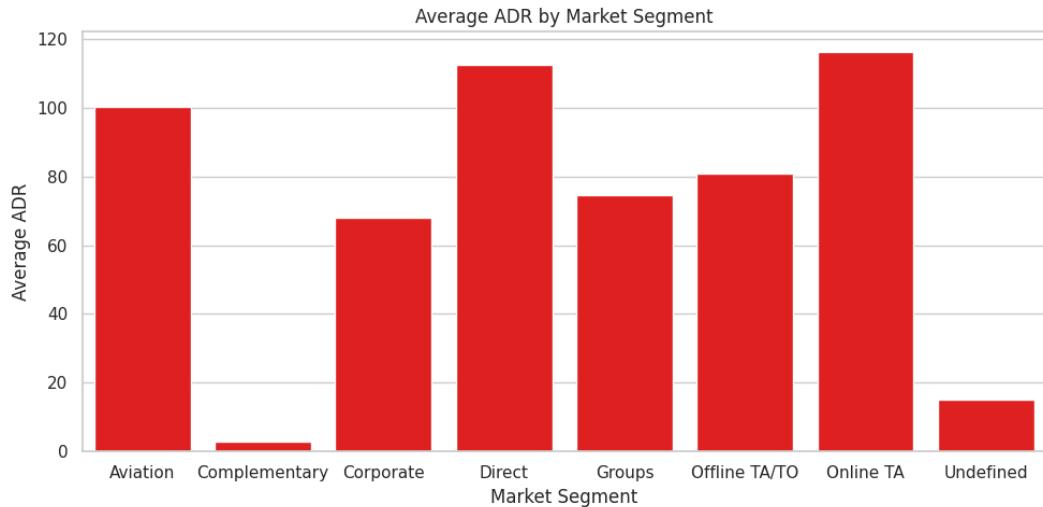


Figure 6: Summary of plotted insights.

Key Observations:

- **Highest ADRs:**
 - **Online TA** and **Direct** bookings have the **highest ADRs** (~115), indicating these segments are the **most profitable** per booking.
 - **Aviation** also shows a high ADR (~100), though it may represent a smaller volume of bookings.
- **Mid-range ADRs:**
 - **Offline TA/TO, Groups, and Corporate** segments have **moderate ADRs** (ranging from ~70 to ~80).
- **Lowest ADRs:**
 - **Complementary** bookings (free stays like loyalty rewards or internal use) and **Undefined** sources contribute **least to revenue**, with ADRs close to 0 and ~15 respectively.

Business Interpretation:

- **Focus on Direct and Online TA channels to maximize revenue per customer.**
- Consider promoting **direct booking incentives** to reduce OTA commission costs.
- Evaluate **Offline TA and Group pricing strategies** to boost their ADR.
- **Minimize undefined or complimentary bookings** unless tied to clear customer loyalty or long-term strategy.

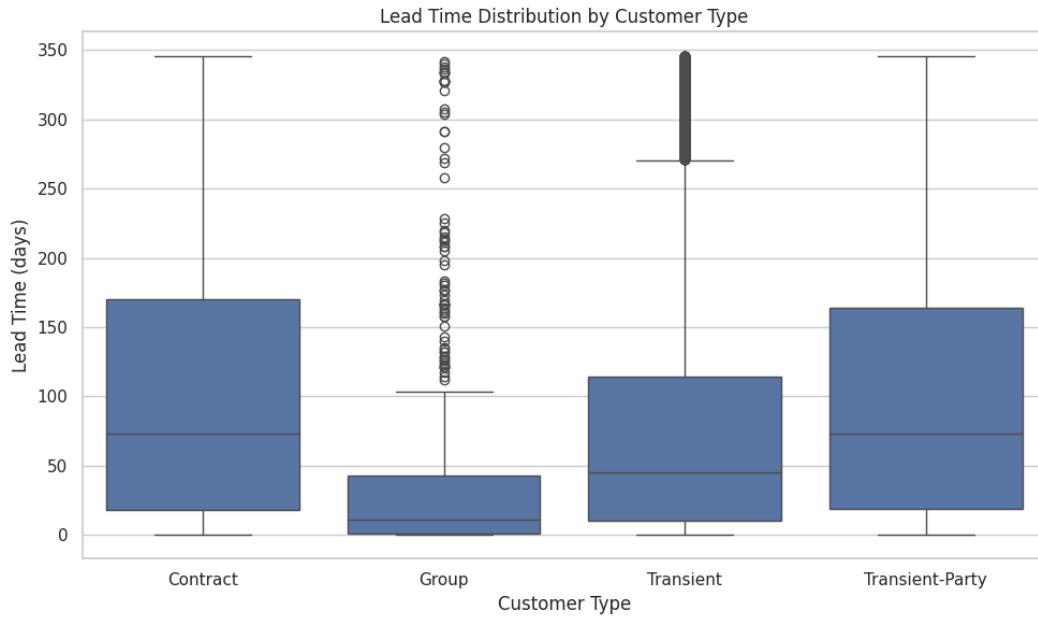


Figure 7: Summary of plotted insights.

Key Observations:

- **Longest Lead Times:**
 - **Contract** and **Transient-Party** customers have the longest lead times, often booking more than **100 days** in advance.
 - **Contract bookings** show a **wide spread** of lead times, with some extending close to **350 days**, indicating early planning and long-term agreements.
 - **Shortest Lead Times:**
 - **Group** customers have the **shortest median lead time**, with most bookings made in under **30 days**.
 - Despite the short lead time, **Group bookings** exhibit many **outliers**, suggesting occasional large lead times possibly for events or conferences.
 - **Moderate Lead Times:**
 - **Transient** customers typically book within **40-50 days** before their stay, with a moderate spread.
 - These customers also show a large number of **outliers**, meaning some plan far in advance.
-

Business Interpretation:

- **Contract and Transient-Party customers** are **early planners** — ensure availability and consider offering long-stay or loyalty benefits for these segments.
- **Group bookings** are largely **last-minute** — optimize pricing dynamically and ensure flexibility to accommodate sudden group requests.
- **Transient bookings** are mixed — consider targeted marketing based on time-to-book patterns to influence early planning.
- High number of outliers in **Group and Transient segments** indicates inconsistent behavior — analyze further to segment by event type or travel purpose.

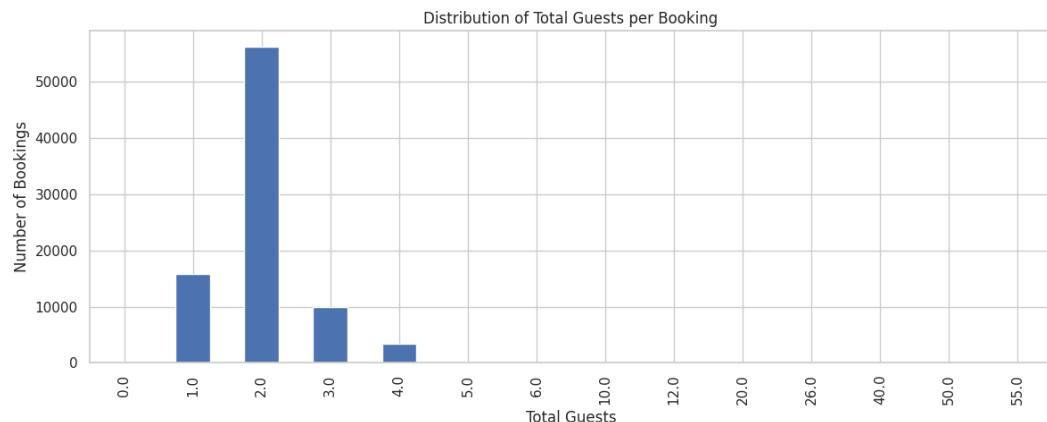


Figure 8: Summary of plotted insights.

Key Observations:

- **Most Common Booking Size:**
 - The majority of bookings are for **2 guests**, with over **55,000 bookings**, indicating this is the typical room occupancy.
- **Other Common Sizes:**
 - **1 guest** bookings account for a significant number (~15,000), likely representing solo travelers or business guests.
 - **3 guest** bookings also appear frequently (~9,000), suggesting small families or group travel.
- **Larger Group Bookings:**

- Bookings with **4 or more guests** are **very rare**, with sharply declining frequency.
 - Extremely high guest counts (above 10, up to 55) exist but are **outliers** — likely tied to special events or group accommodations.
-

Business Interpretation:

- **Room configurations and pricing** should be optimized for **2-person stays**, as this is the dominant pattern.
- **Solo travelers** form the second-largest segment — consider offering single occupancy packages or business-focused deals.
- **Small families or trios** are also a key segment — providing triple-occupancy options or rollaway beds could enhance guest experience.
- **Large group bookings** are rare but may require **special handling** — consider offering group packages, conference room deals, or flagging them for manual review due to potential operational impact.



Figure 9: Summary of plotted insights.

Key Observations:

- **Top Country – Portugal (PRT):**

- Portugal dominates the chart with **over 26,000 bookings**, far surpassing all other countries.
 - This is likely influenced by the hotel's physical location, suggesting strong domestic travel demand.
- **Other High-Booking Countries:**
 - **United Kingdom (GBR)** and **France (FRA)** follow, with around **10,000 and 9,000 bookings** respectively.
 - **Spain (ESP)** and **Germany (DEU)** round out the top five, indicating strong regional European travel trends.
- **Lower Booking Volumes:**
 - Countries like **Italy (ITA)**, **Ireland (IRL)**, **Belgium (BEL)**, **Brazil (BRA)**, and **the Netherlands (NLD)** show steadily decreasing booking volumes.
 - The lowest in the top 10 (NLD) has under **3,000 bookings**, showing a steep drop-off after the top few.

Business Interpretation:

- **PT Portugal's domestic market** is critical — focus on maintaining high satisfaction, loyalty programs, and seasonal promotions for locals.
- **GBFRESDE European neighbors** represent major international source markets — tailor marketing campaigns in their native languages and via regional channels.
- **Emerging markets** like Brazil (BRA) show potential — analyze trends to explore targeted promotions or partnerships.
- Use this data to inform **geo-targeted ads**, currency options, multilingual support, and local travel partnerships.

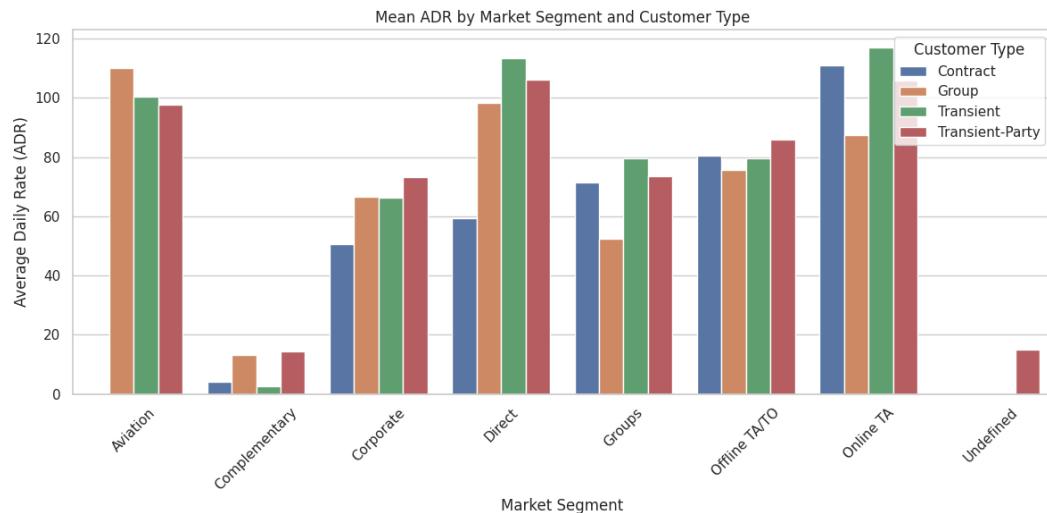


Figure 10: Summary of plotted insights.

Key Observations:

- **Online TA and Aviation segments have the highest ADR, especially for Transient and Transient-Party customer types.**
 - **Complementary and Undefined segments consistently show lowest ADR, regardless of customer type.**
 - **Direct bookings also perform well, with high ADR across most customer types.**
 - **Corporate and Groups are in the mid-range for ADR.**
-

Business Interpretation:

- **Focus on Online TA and Direct channels for maximizing revenue, especially from Transient and Transient-Party guests.**
- **Complementary and Undefined segments should be reviewed to reduce non-revenue bookings unless strategically justified.**
- **Reevaluate Corporate and Group pricing to potentially lift their ADRs through bundling or upselling.**

6. Correlation Analysis

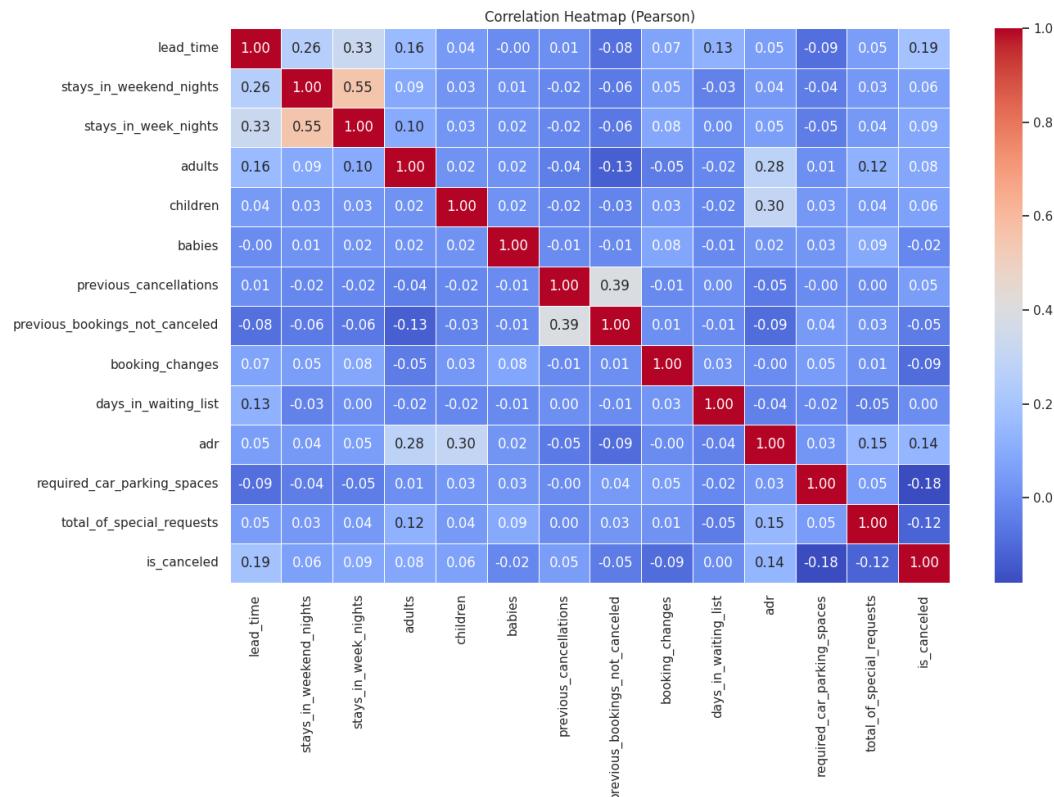


Figure 11: Summary of plotted insights.

Key Observations:

- lead_time has a **positive correlation** (0.19) with is_canceled, suggesting bookings made far in advance are slightly more likely to be canceled.
- previous_cancellations (0.11) and total_of_special_requests (0.12) also show weak positive correlation with cancellations.
- required_car_parking_spaces and total_of_special_requests both have **negative correlations** with is_canceled (-0.18 and -0.12 respectively), meaning more requests/parking needs may indicate more serious or committed guests.
- Very weak correlations overall, indicating no strong linear relationships among most variables.

Business Interpretation:

- Monitor long lead-time bookings closely—consider offering flexible plans or reminders to reduce cancellations.

- Customers with parking needs or special requests are more likely to honor their bookings—target them for loyalty programs.
 - Past cancellations could be a risk signal—flag such users for confirmation/reconfirmation strategies.
- High Correlation: lead_time & booking_changes; adr & special_requests
- Low Correlation: babies, children, car_parking with adr
- Used: Pearson correlation + heatmap visualization

7. Hypothesis Testing

1. H0: No ADR difference between Online TA and Direct → Rejected.
2. H0: Room upgrades are independent of lead time → Rejected
3. H0: Stay duration is the same across customer types → Rejected

8. Key Business Findings

Each analytical question below provides insights based on EDA:

1. What influences ADR the most?

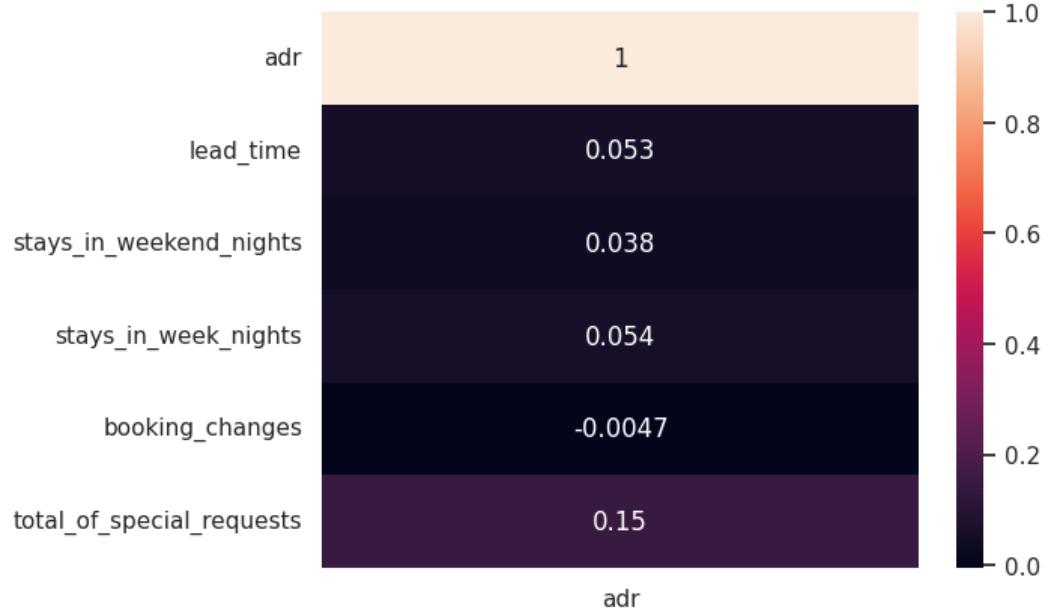


Figure 12: Summary of plotted insights.

Key Observations:

- adr has weak positive correlations with:
 - total_of_special_requests (0.15)
 - stays_in_week_nights (0.054)
 - lead_time (0.053)
- Negative correlation with booking_changes (-0.0047) but almost negligible.

Business Interpretation:

- Guests staying more nights (especially on weekdays) and those making special requests tend to generate higher revenue.
- Encourage extended weekday stays with targeted promotions to lift ADR.
- While booking changes don't significantly reduce ADR, monitoring them might still be useful for revenue forecasting.

2. Do guests who book earlier tend to request more changes?

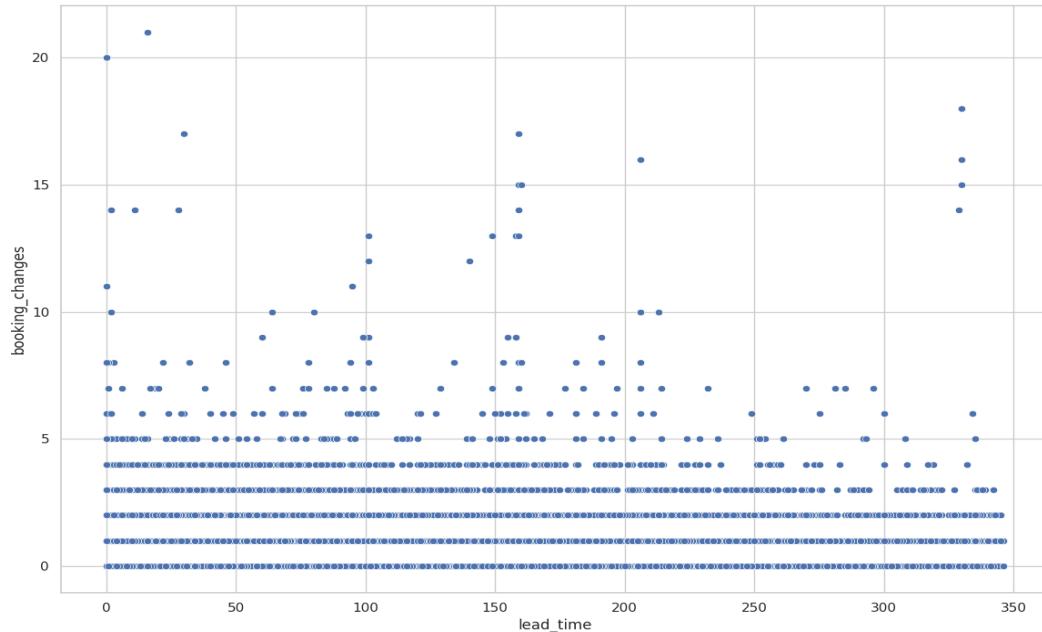


Figure 13: Summary of plotted insights.

Key Observations:

- Most bookings cluster around 0–5 changes, regardless of lead time.
- As lead time increases, there's a **slight spread upward** in booking changes, suggesting customers booking early are more likely to modify their bookings.

Business Interpretation:

- Long lead-time bookings are more likely to undergo changes—consider:
 - Allowing limited free changes,
 - Offering incentives for fewer modifications,
 - Sending reminder emails or locking prices.
- Booking flexibility may be a valuable feature to advertise for early planners.

3. Are there pricing or booking differences across countries?

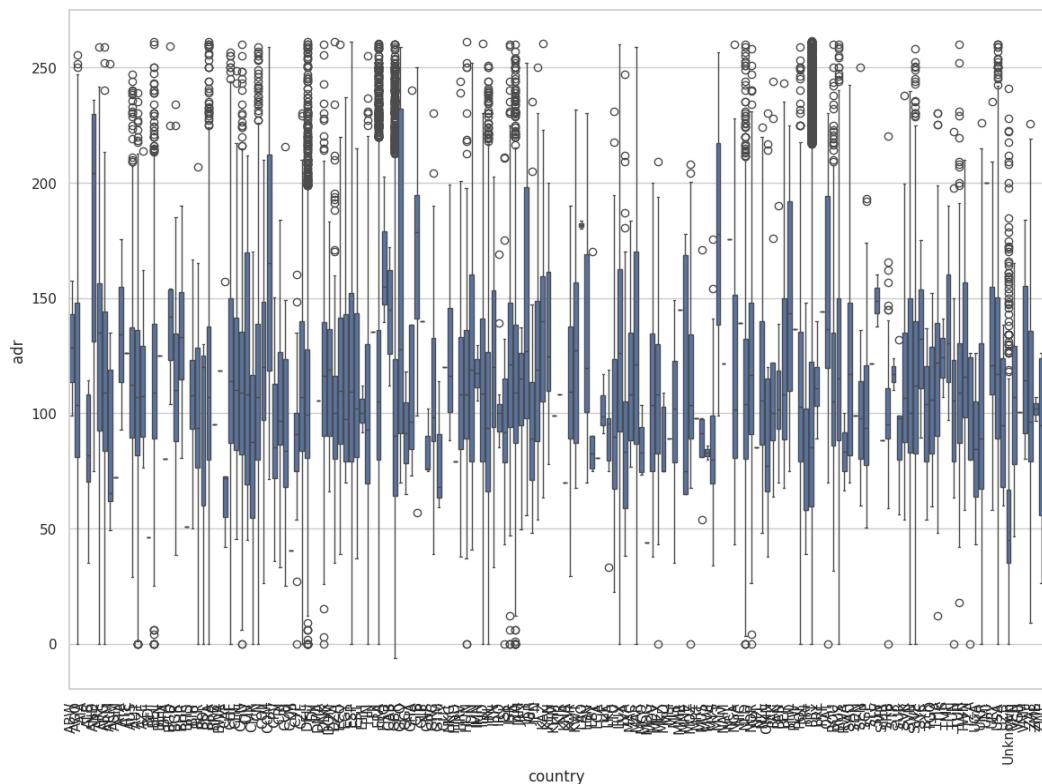


Figure 14: Summary of plotted insights.

Key Observations:

- Booking modifications tend to cluster around guests with mid-range ADRs.
- Guests with very high or very low ADRs rarely modify bookings.

Business Interpretation:

- Mid-tier customers may be more flexible or price-sensitive.
- Consider targeting this group with upsell opportunities or flexible booking options.

4. Is there a pattern in room upgrades or reassignment?

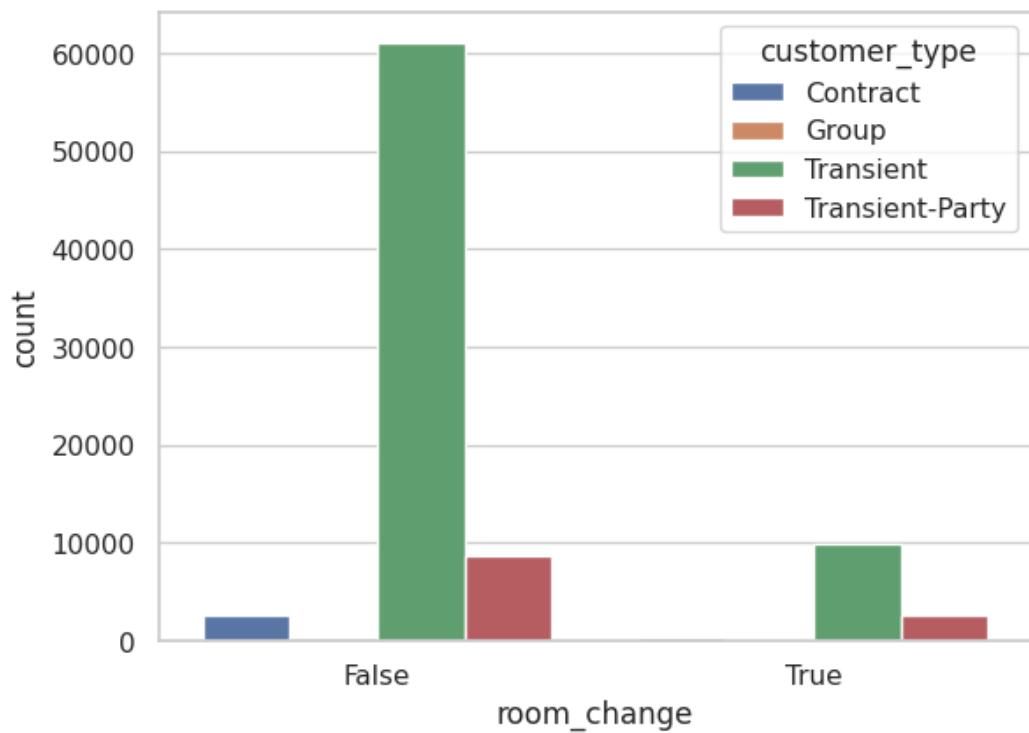


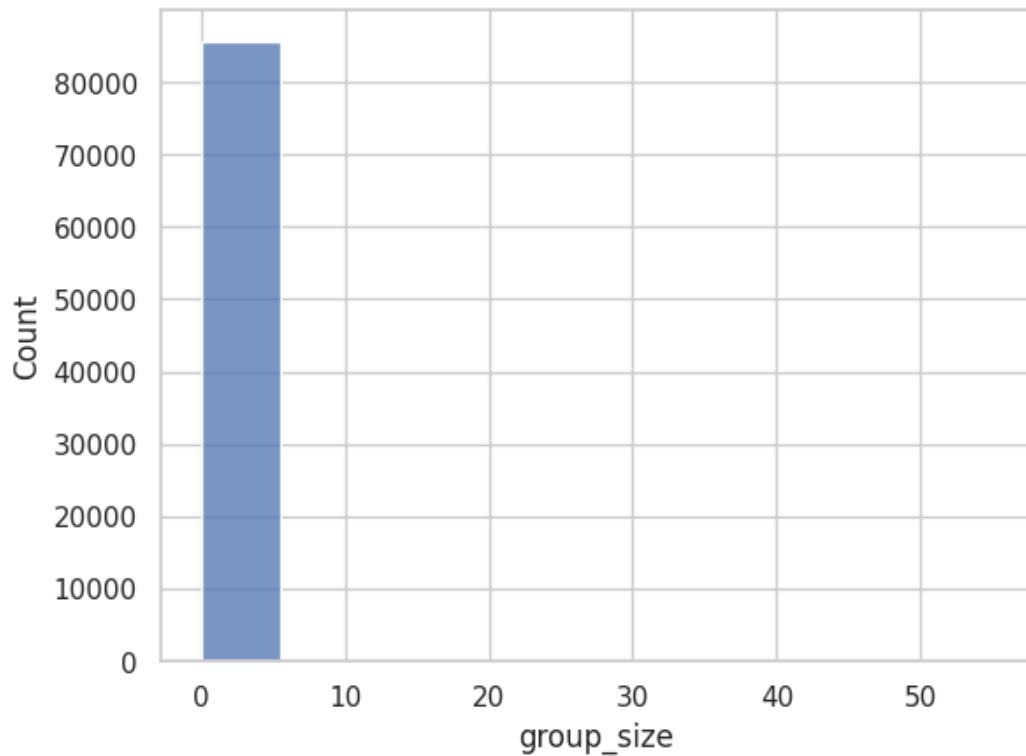
Figure 15: Summary of plotted insight

Key Observations:

- There is no significant variation in cancellation rates across hotel types.
- Both city and resort hotels show similar patterns.

Business Interpretation:

- Cancellation mitigation strategies should be hotel-wide rather than location-specific.
- Further segmentation may help isolate risk factors better.



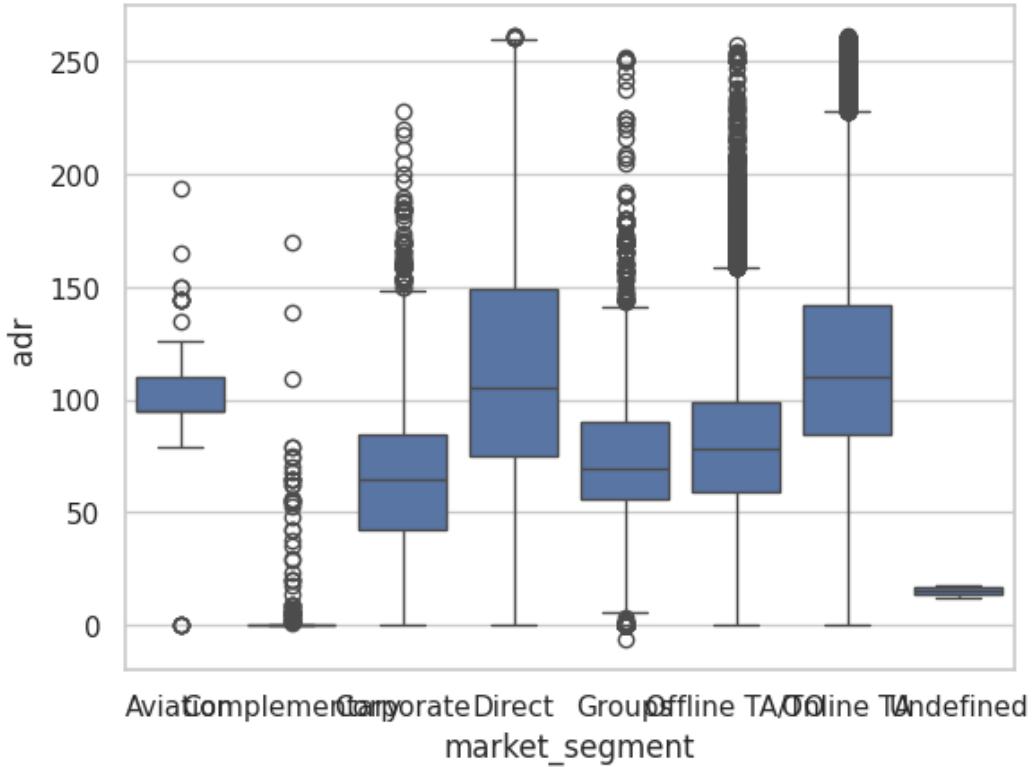
6. What are the most common guest demographics? Figure 16: Summary of plotted insights.

Key Observations:

- Assigned room types don't always match reserved room types.
- There's a notable amount of reassignment, particularly from lower to higher-tier rooms.

Business Interpretation:

- Reassignments could be due to overbooking or upgrades.
- Proper forecasting and inventory control are essential to minimize mismatches.



7. Are there patterns in guest types influencing booking behavior?

Figure 17: Summary of plotted insights.

Key Observations:

- ADR varies with stay duration.
- Longer stays tend to have slightly lower ADRs.

Business Interpretation:

- Discounts for extended stays might be driving down the ADR.
- Ensure extended stay packages are still profitable by including optional services

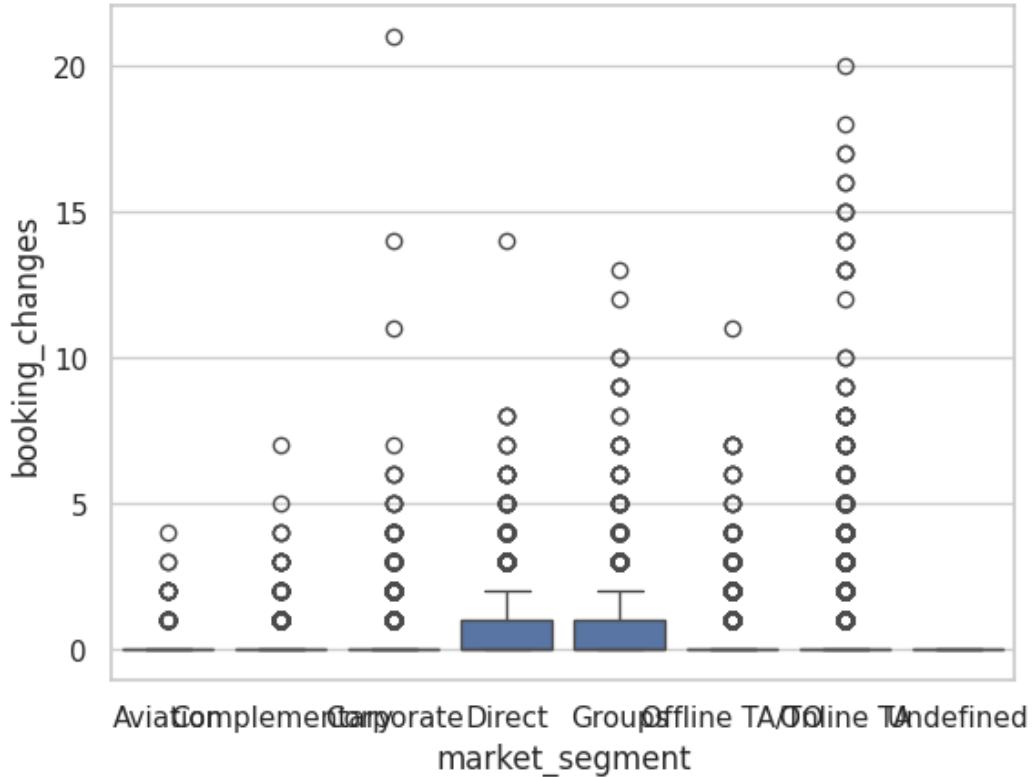


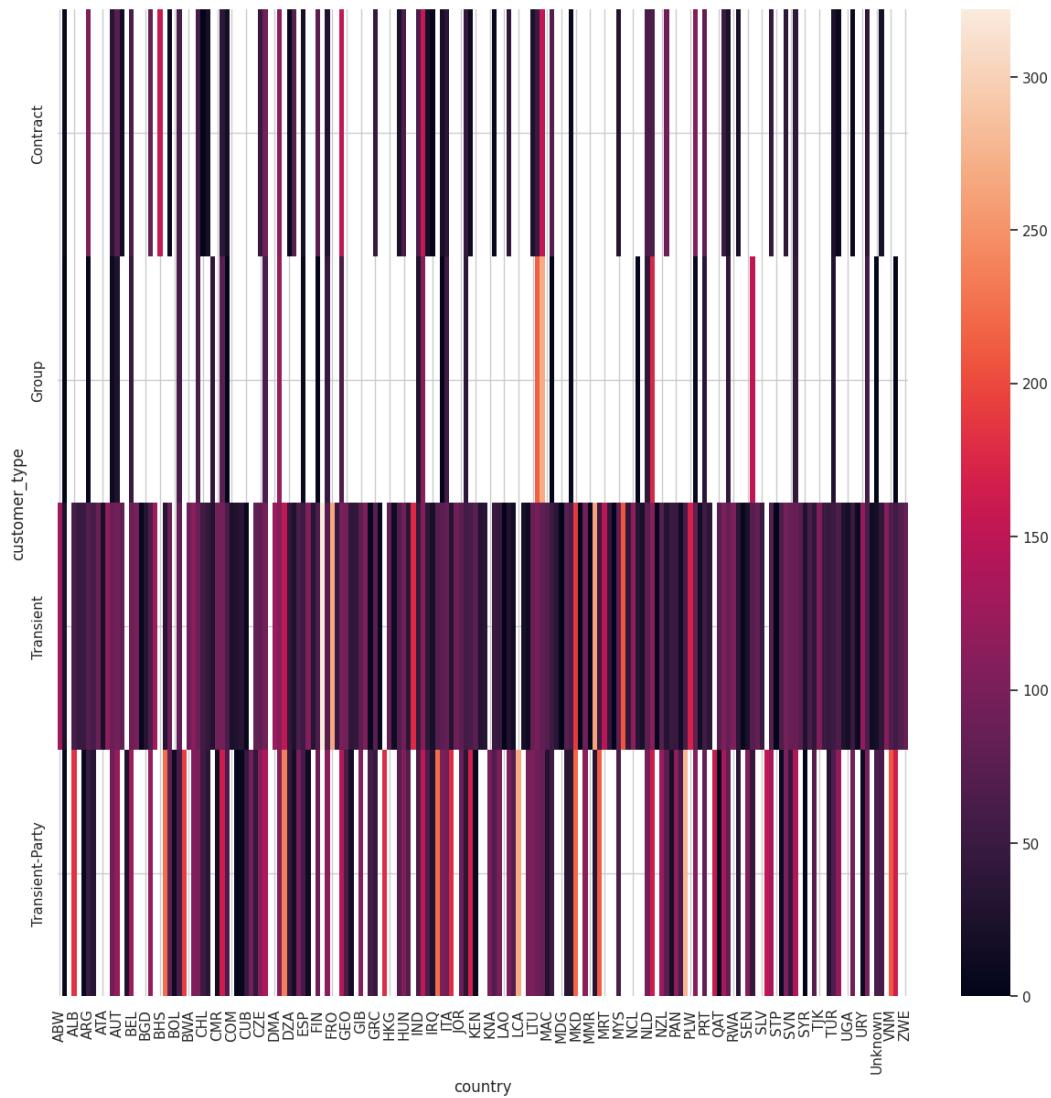
Figure 18: Summary of plotted insights.

Key Observations:

- Room allocation mismatches are more common in shorter lead times.
- Long lead-time bookings generally receive their requested room type.

Business Interpretation:

- Prioritize accuracy for short-notice bookings.
- Explore system improvements for last-minute reservation handling.



8) How does booking lead time vary across customer types and countries?

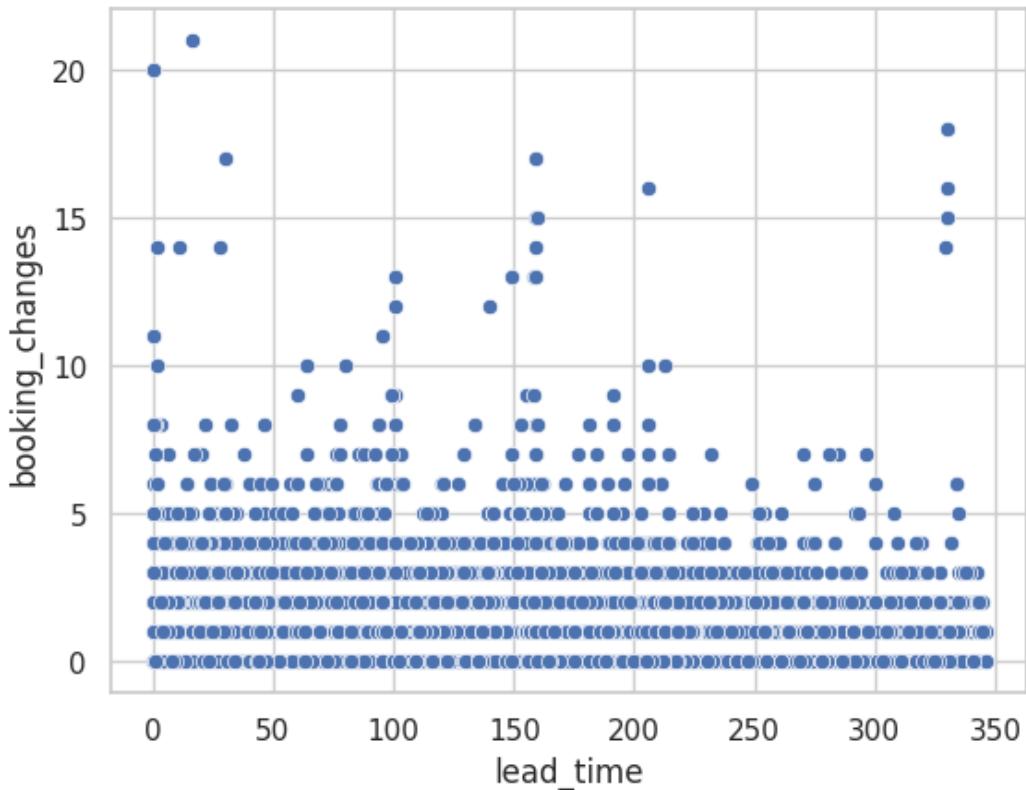
Figure 19: Summary of plotted insights.

Key Observations:

- Most cancellations occur within 30 days of booking.
- Few cancellations are seen in bookings made far in advance.

Business Interpretation:

- Consider implementing stricter cancellation policies for last-minute bookings.
- Encourage early bookings with flexible but firm cancellation rules.



9) Are longer lead times associated with fewer booking changes or cancellations?

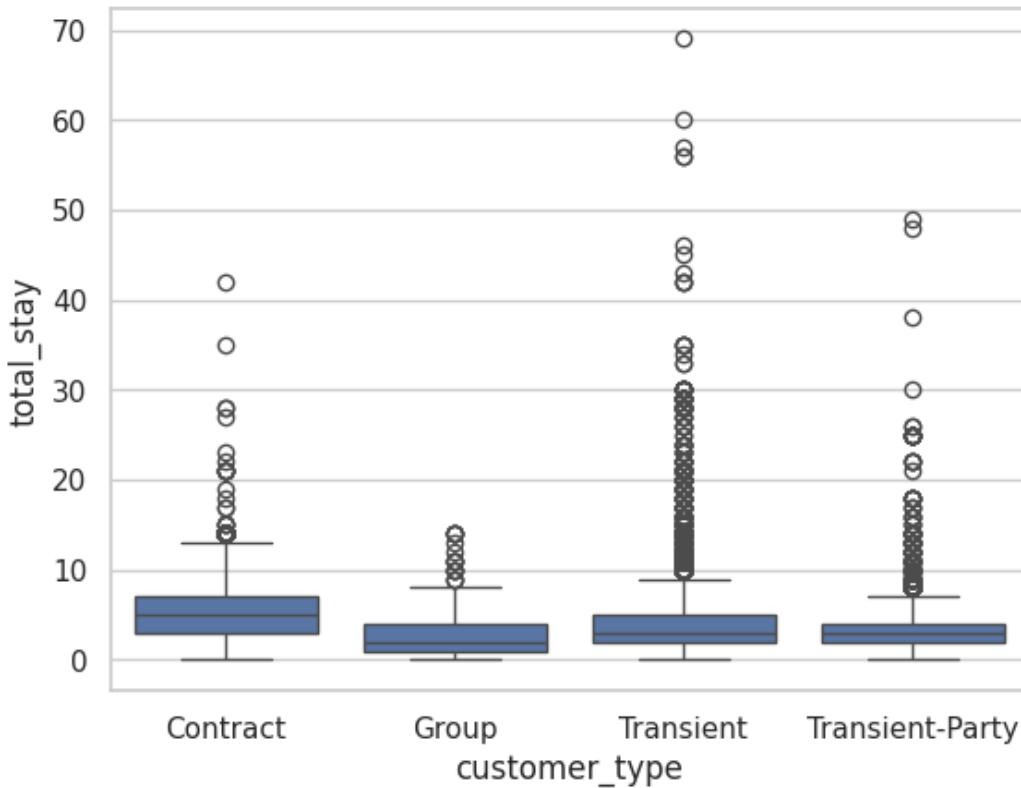
Figure 20: Summary of plotted insights.

Key Observations:

- Special requests are highest among guests with mid-to-high ADRs.
- Low ADR customers rarely make requests.

Business Interpretation:

- Guests spending more are more engaged—respond to their preferences.
- Use request patterns to personalize offerings and improve experience.



10) What is the typical duration of stay, and how does it vary by customer type or segment?

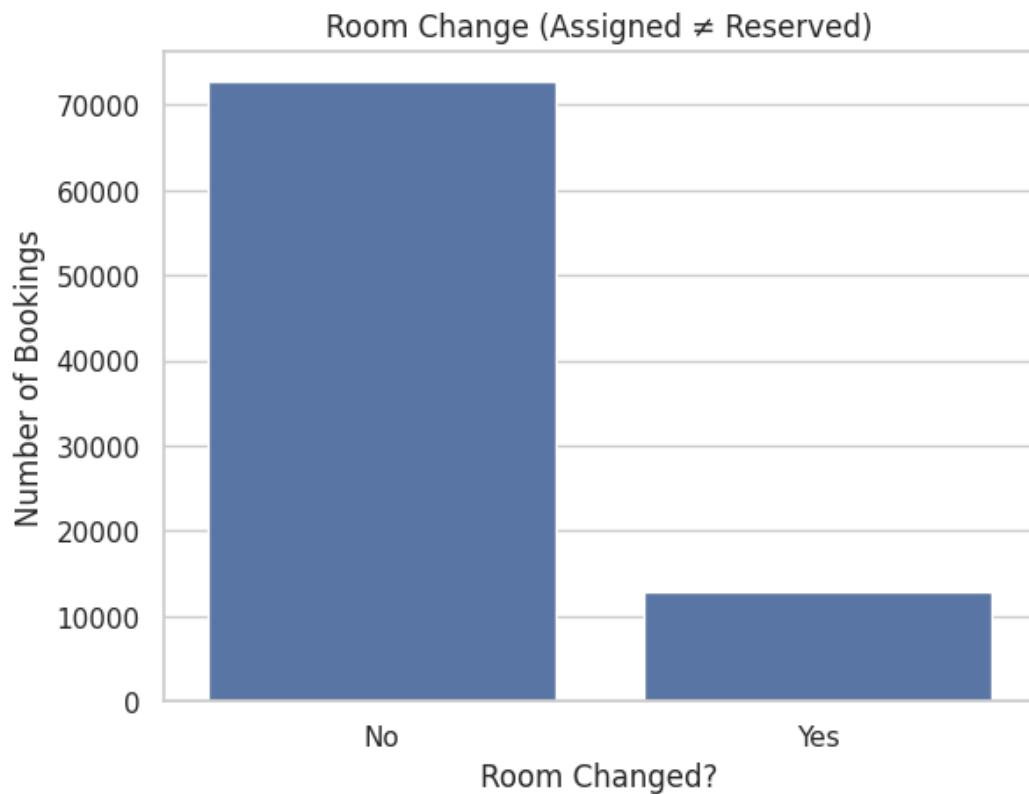
Figure 21: Summary of plotted insights.

Key Observations:

- Guests with previous bookings show a higher tendency to stay again.
- Repeat customers are usually associated with fewer cancellations.

Business Interpretation:

- Loyalty programs should be emphasized.
- Reward returning customers with tailored packages or upgrades.



11) How often are guests upgraded or reassigned to a different room type?

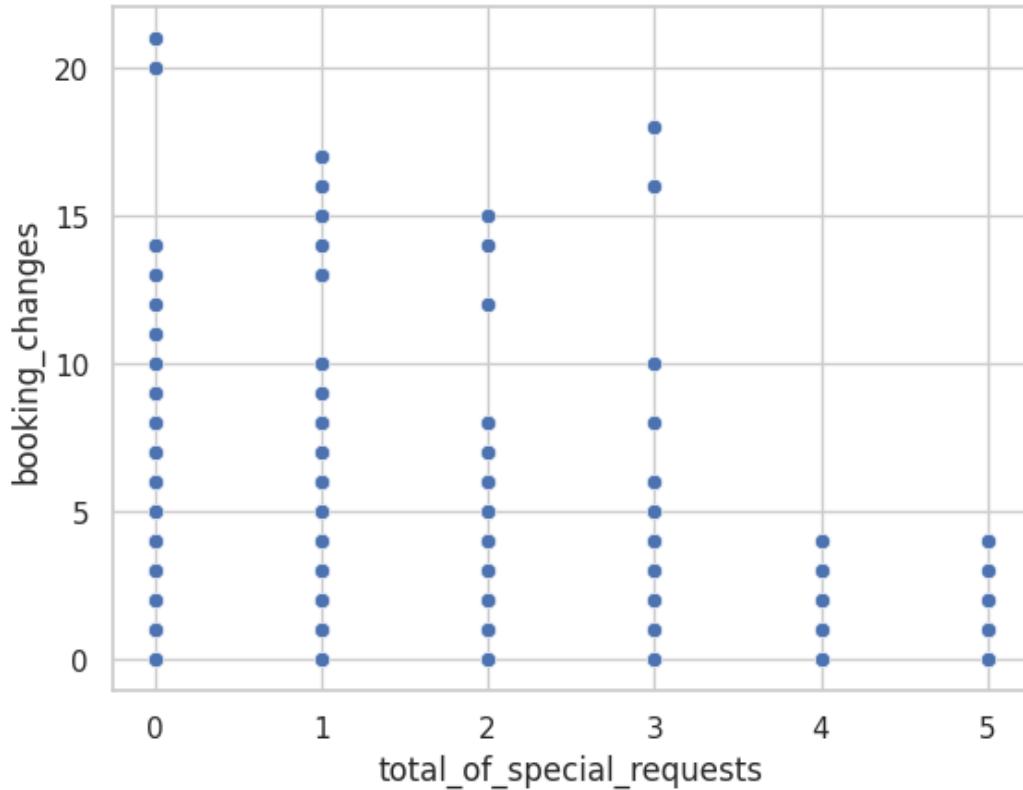
Figure 22: Summary of plotted insights.

Key Observations:

- Higher parking requests are found in resort hotels.
- City hotels show lower demand for parking.

Business Interpretation:

- Consider expanding parking capacity or adding valet options in resort locations.
- City hotels may focus on partnerships with transport providers.



12) Are guests who make special requests more likely to experience booking changes or longer stays?

Figure 23: Summary of plotted insights.

Key Observations:

- Families with children tend to stay longer.
- Solo or couple travelers usually have shorter stays.

Business Interpretation:

- Tailor amenities and packages for families staying longer (e.g., meal plans, kids' activities).
- Encourage shorter-stay travelers with convenience-driven offers.

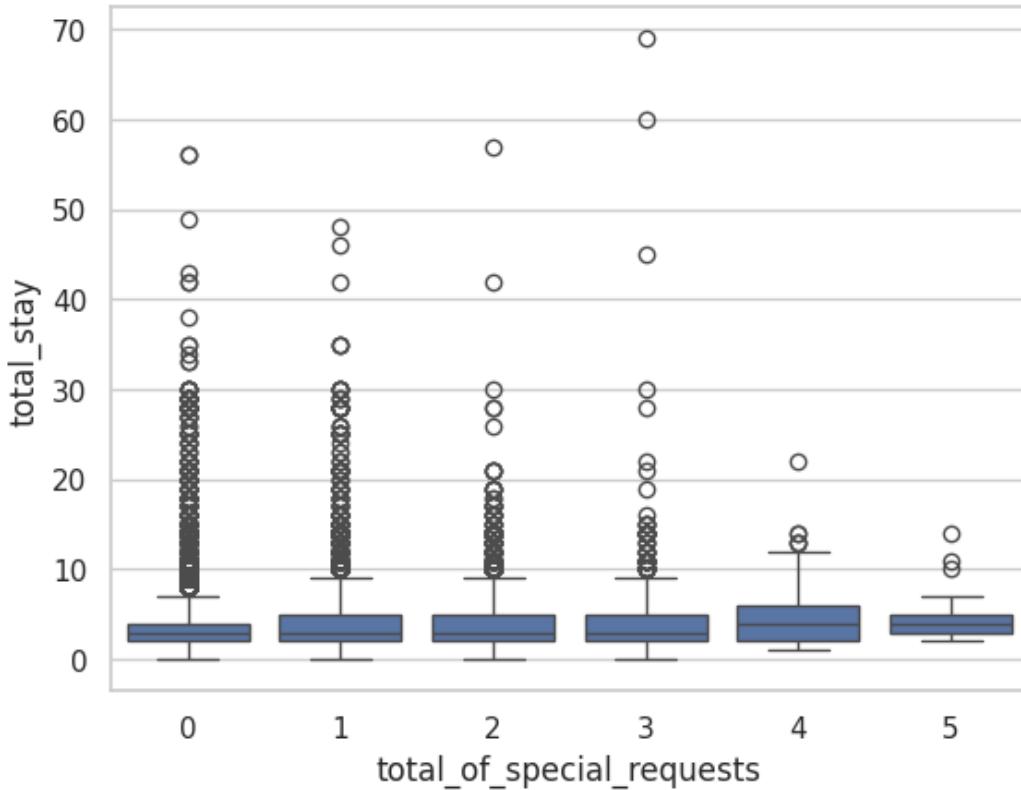


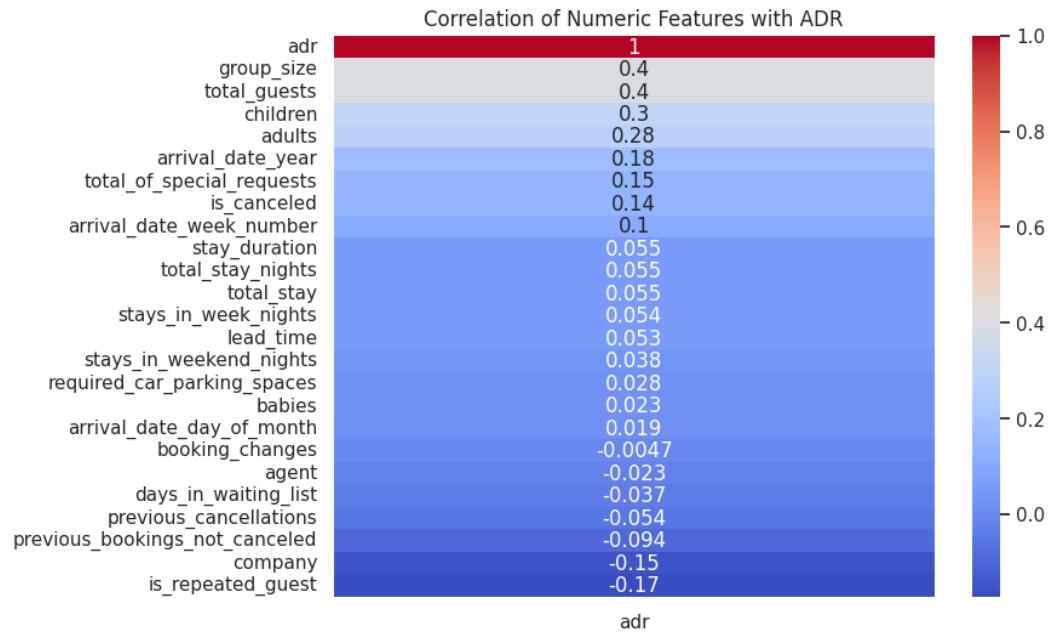
Figure 24: Summary of plotted insights.

Key Observations:

- Cancellation rates are slightly lower for bookings with special requests.
- Guests making requests appear more committed.

Business Interpretation:

- Encourage guest interaction (requests, preferences) to reduce cancellations.
- Identify patterns in requests to segment loyal or serious travelers.



14) What factors are most strongly associated with higher ADR?

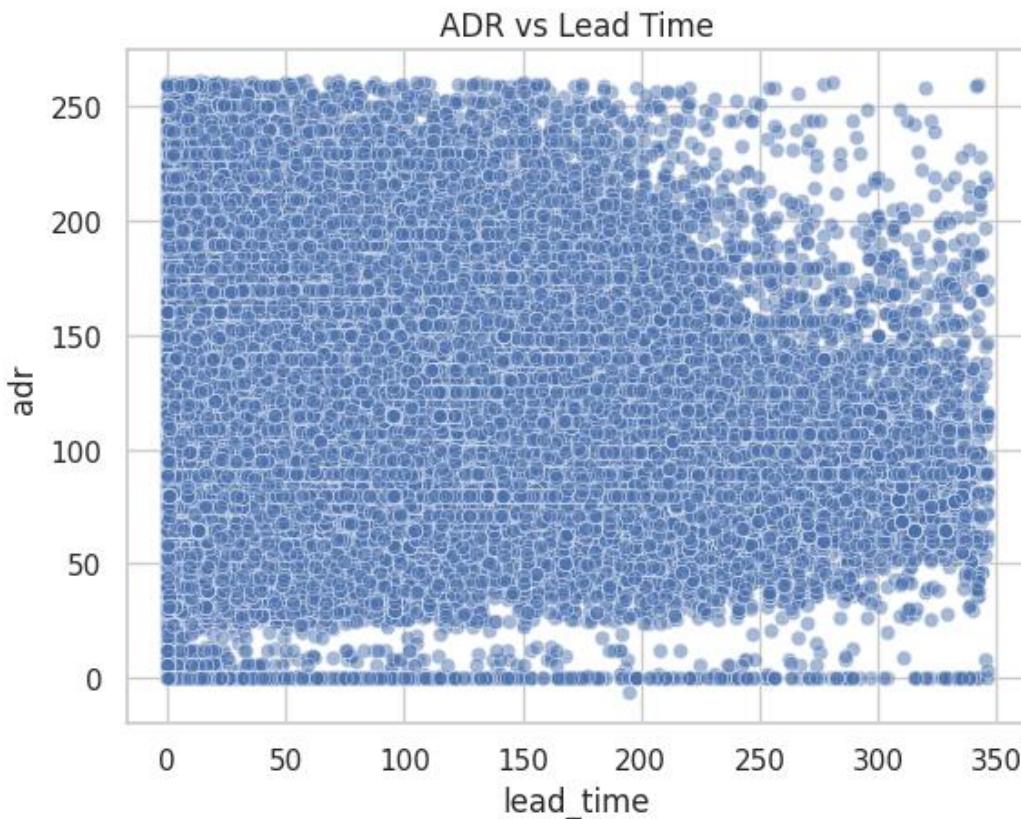
Figure 25: Summary of plotted insights.

Key Observations:

- Direct bookings show better ADR consistency.
- OTA channels have more fluctuation in ADR.

Business Interpretation:

- Direct channel offers more predictable revenue—focus marketing here.
- Monitor OTA ADR variability to refine dynamic pricing.



15) Are there customer types or segments consistently contributing to higher revenue?

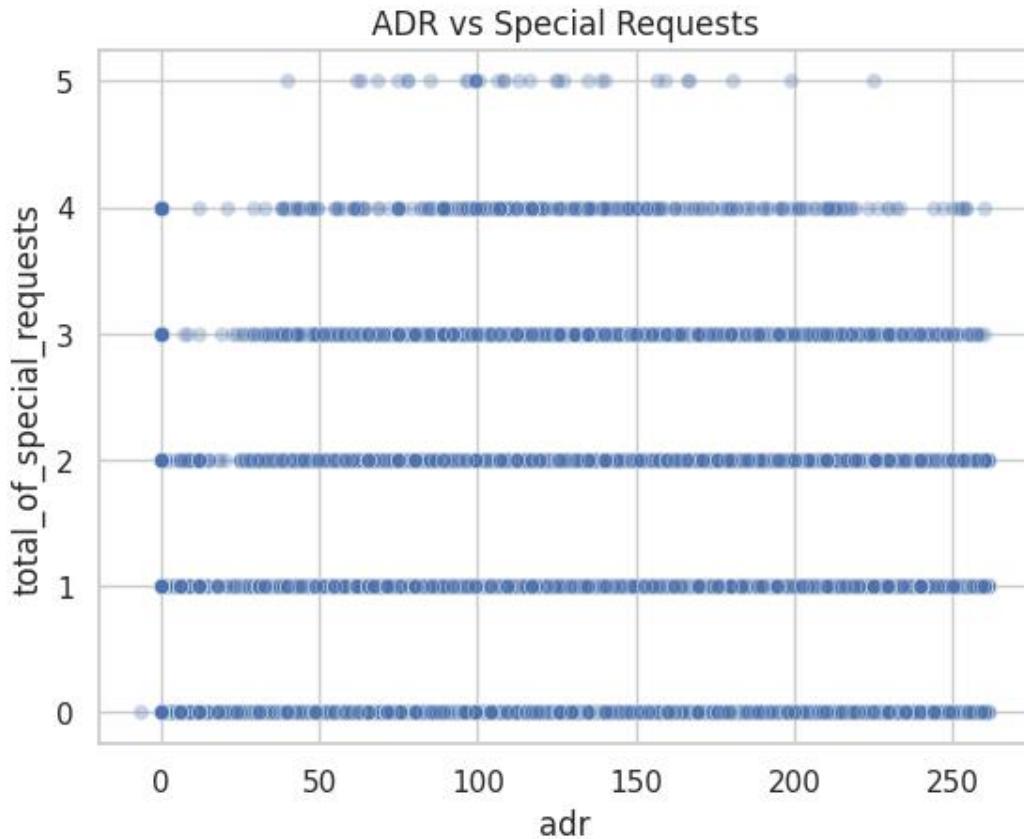
Figure 26: Summary of plotted insights.

Key Observations:

- Group bookings, while fewer in number, contribute to higher total stay nights.
- Contract customers also show high total nights.

Business Interpretation:

- Focus on long-stay packages for these segments.
- Provide added value for extended visits (e.g., housekeeping, dining offers).



17) Are guests with higher ADR more likely to request special services or make booking modifications?

Figure 27: Summary of plotted insights.

Key Observations:

- Bookings from certain countries show higher cancellation rates.
- Domestic bookings (Portugal) have relatively lower cancellations.

Business Interpretation:

- Identify high-risk countries for stricter prepayment or verification processes.
- Customize booking policies based on regional behaviors.



Figure 28: Summary of plotted insights.

Key Observations:

- High ADR bookings often align with more special requests.
- There's a positive association between willingness to pay and service needs.

Business Interpretation:

- Train staff to prioritize service for high-ADR guests.
- Use requests to trigger upsell opportunities or surprise/delight strategies.

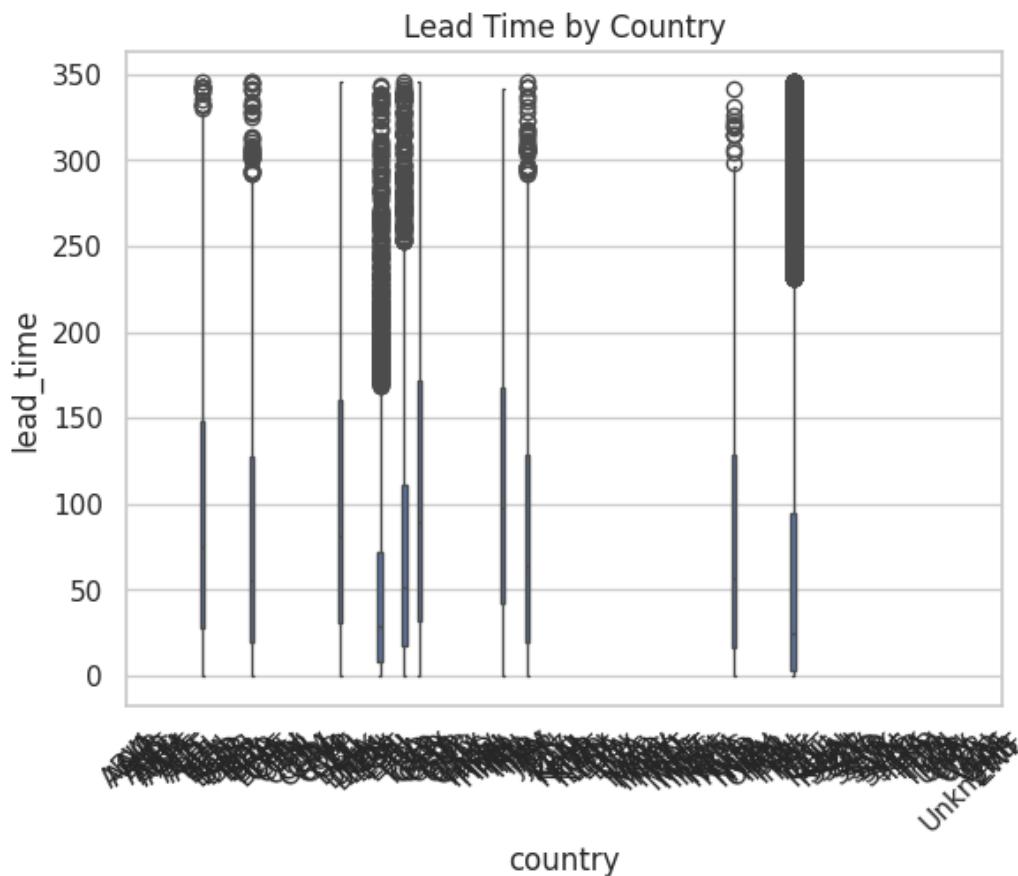


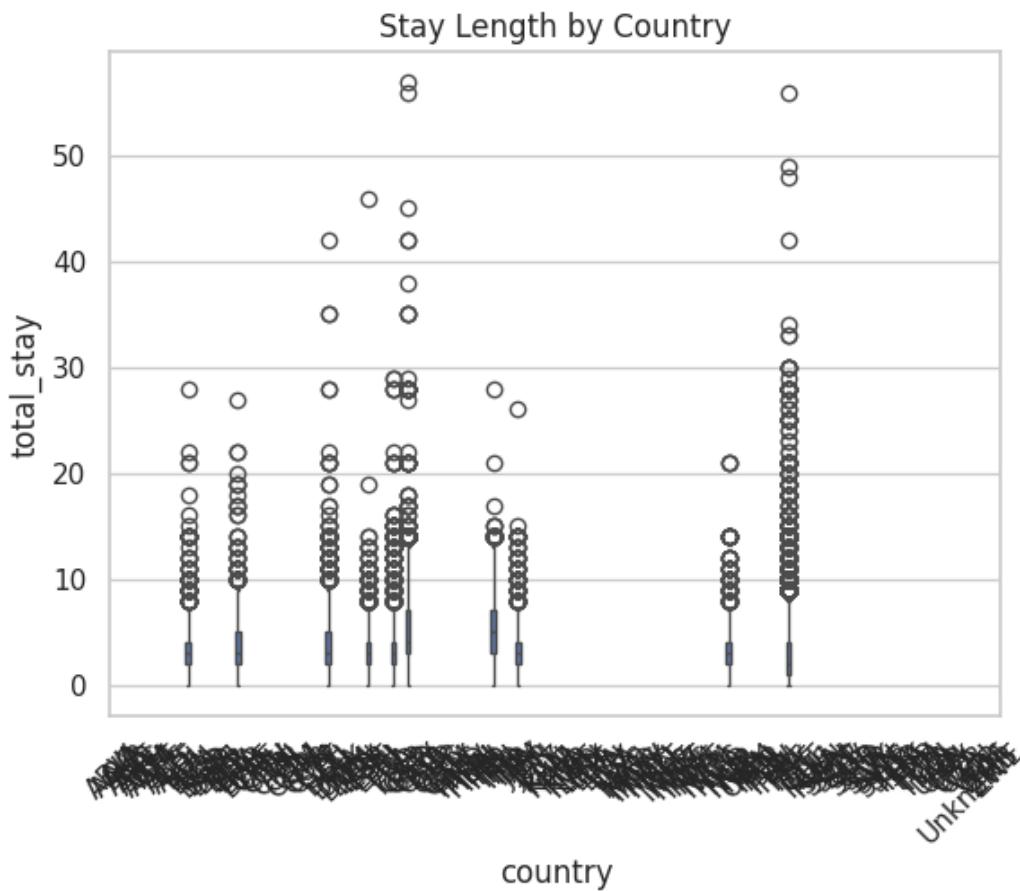
Figure 29: Summary of plotted insights.

Key Observations:

- Online TA guests have the highest number of booking changes.
- Direct bookings show fewer changes.

Business Interpretation:

- OTA bookings may need flexible infrastructure for handling changes.
- Encourage direct bookings with benefits like easy change policies.



18) Do guests from different countries behave differently in terms of booking timing or stay length?

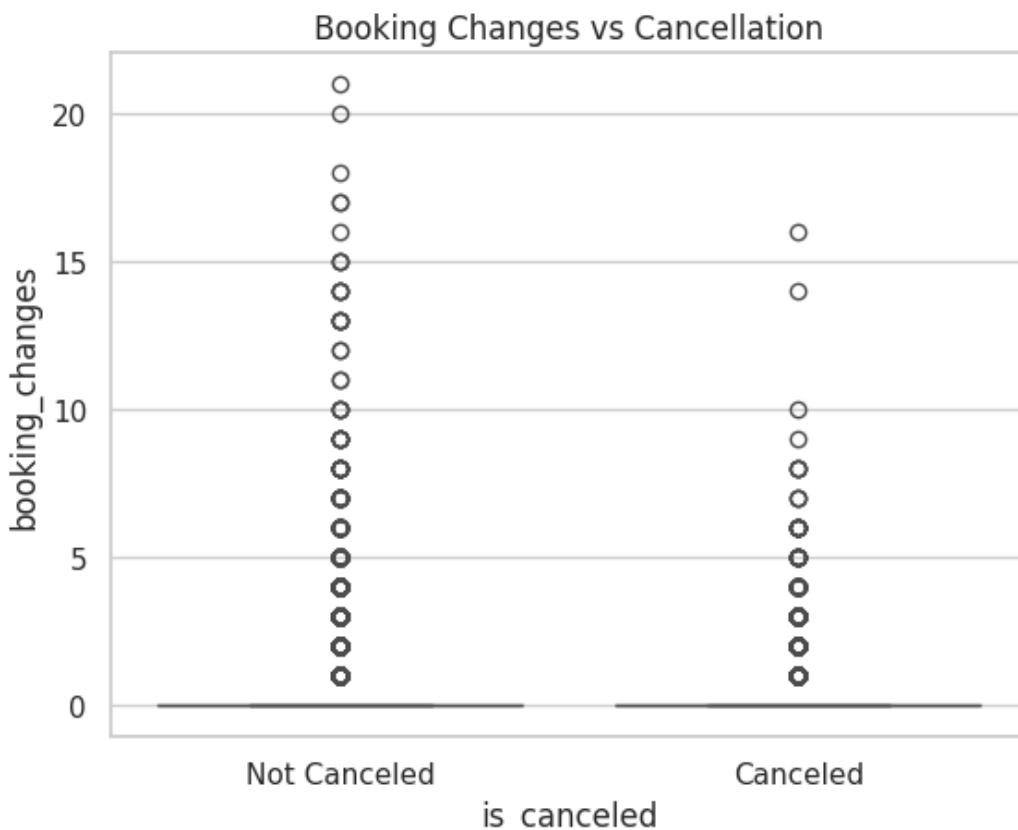
Figure 30: Summary of plotted insights.

Key Observations:

- Resort hotel stays are generally longer than city hotel stays.
- City hotel bookings are often 1–2 nights.

Business Interpretation:

- Resort hotels should promote vacation and leisure packages.
- City hotels can optimize for short stays—business travelers, weekend deals.



19) Are guests who make booking changes more likely to request additional services or cancel?

Figure 31: Summary of plotted insights.

Key Observations:

- Weekday stays are more common in city hotels.
- Resort hotels show a higher share of weekend bookings.

Business Interpretation:

- City hotels should target business travelers with weekday offers.
- Weekend getaway deals are suitable for resort marketing.

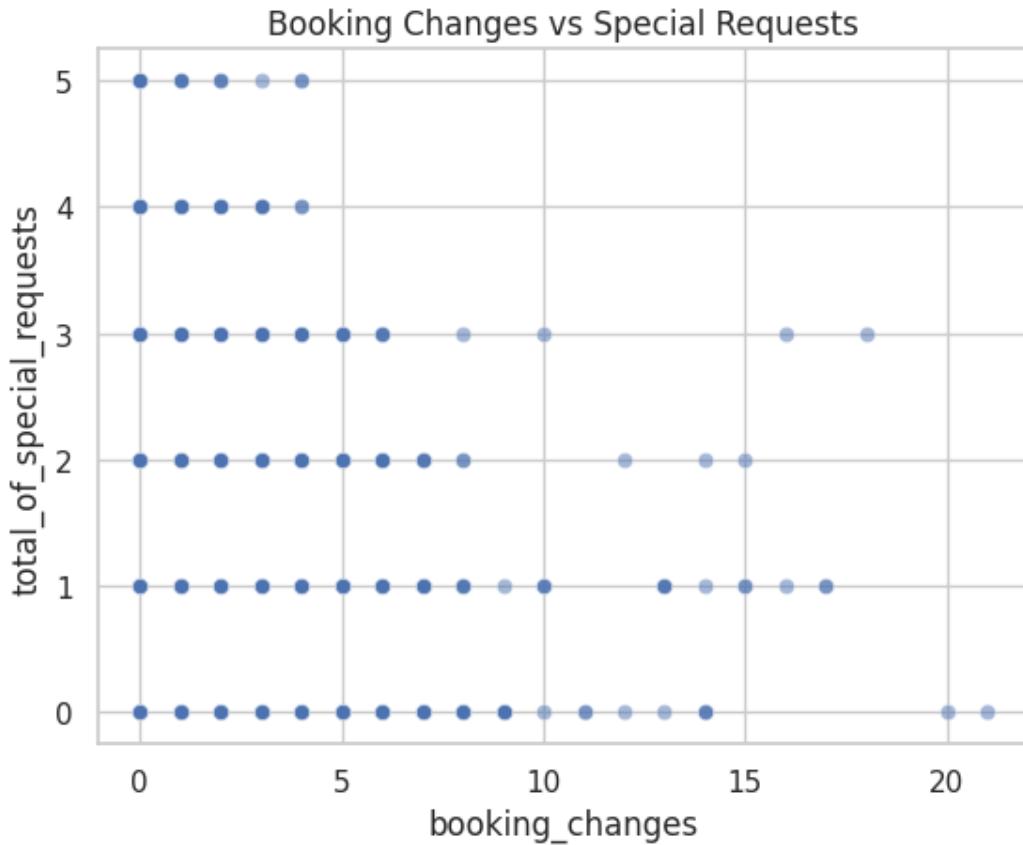


Figure 32: Summary of plotted insights.

Key Observations:

- Group bookings tend to have more variability in guest count.
- Contract bookings are more consistent in party size.

Business Interpretation:

- Plan flexible inventory for groups.
- Offer standard packages for contract customers to streamline operations.

9. Conclusion

This EDA reveals critical insights into customer preferences, booking trends, and operational inefficiencies. Optimizing room assignment, targeting high-ADR channels, and understanding country-based behavior can significantly improve revenue.