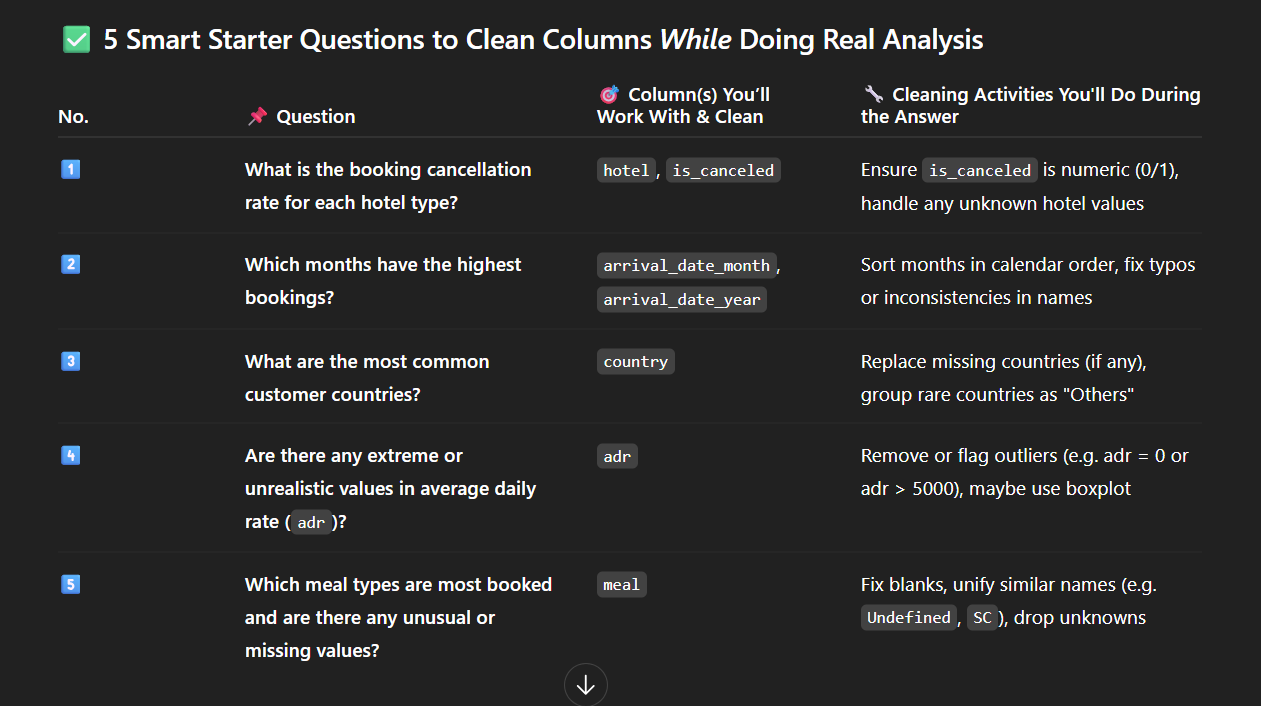
# 📊 Hotel Booking Dataset: Full Column-wise Documentation with Examples

| **No.** | **Column Name** | **Description** | **Example Value** | **Explanation in Simple Terms** |
| --- | --- | --- | --- | --- |
| 1️⃣ | hotel | Type of hotel where the booking was made | City Hotel | City or Resort hotel. Helps compare demand & cancellation between hotel types. |
| 2️⃣ | is\_canceled | Whether the booking was canceled (1) or not (0) | 1 | 1 = booking got canceled, 0 = booking was completed (guest stayed). |
| 3️⃣ | lead\_time | Number of days between booking date and arrival date | 75 | If lead time is 75, guest booked 75 days before their arrival. |
| 4️⃣ | arrival\_date\_year | Year the guest was supposed to arrive | 2016 | Useful for year-wise trend analysis. |
| 5️⃣ | arrival\_date\_month | Month of arrival (in full text) | August | Helpful to check seasonal booking trends. |
| 6️⃣ | arrival\_date\_week\_number | Week number (1–52) of the guest's arrival | 33 | Week-based patterns can be studied using this. |
| 7️⃣ | arrival\_date\_day\_of\_month | Exact day of the month guest arrives | 15 | You can study if mid-month or end-month bookings are high. |
| 8️⃣ | stays\_in\_weekend\_nights | Nights spent during weekends | 2 | If a guest stays Saturday–Sunday, this will be 2. |
| 9️⃣ | stays\_in\_week\_nights | Nights spent during weekdays | 3 | Useful for calculating full trip length. |
| 🔟 | adults | Number of adults included in the booking | 2 | Important for occupancy and revenue metrics. |
| 1️⃣1️⃣ | children | Number of children included (can be missing) | 1 | Missing values may need to be treated (0 if unknown). |
| 1️⃣2️⃣ | babies | Number of babies included in the booking | 0 | Helps hotels provide baby facilities when needed. |
| 1️⃣3️⃣ | meal | Type of meal plan selected | BB | BB = Bed & Breakfast, HB = Half Board, SC = Self Catering, etc. |
| 1️⃣4️⃣ | country | Country code of the guest (ISO 3-letter code) | PRT (Portugal) | Useful for customer demographics and international analysis. |
| 1️⃣5️⃣ | market\_segment | How the booking was made (channel) | Online TA | TA = Travel Agent, Direct = customer booked without third party. |
| 1️⃣6️⃣ | distribution\_channel | Where the booking came from (distribution system) | TA/TO | TA/TO = Travel Agent / Tour Operator, Direct, Corporate, etc. |
| 1️⃣7️⃣ | is\_repeated\_guest | Is this a returning guest? (1 = Yes, 0 = No) | 0 | Used to identify customer loyalty. |
| 1️⃣8️⃣ | previous\_cancellations | Number of previous bookings canceled by this guest | 0 | Helps detect risky or unreliable customers. |
| 1️⃣9️⃣ | previous\_bookings\_not\_canceled | Number of successful past bookings by this guest | 1 | Can help rate the reliability of a customer. |
| 2️⃣0️⃣ | reserved\_room\_type | Room type the guest originally booked | A | Room types are anonymized as letters (A–L). |
| 2️⃣1️⃣ | assigned\_room\_type | Actual room type assigned (might differ from reserved) | B | Useful to check if upgrades/downgrades happened. |
| 2️⃣2️⃣ | booking\_changes | Number of times booking was modified | 1 | If guests frequently change bookings, it may signal indecisiveness. |
| 2️⃣3️⃣ | deposit\_type | Type of deposit paid | No Deposit | Non Refund, Refundable, or No Deposit. |
| 2️⃣4️⃣ | agent | ID of the travel agent who handled the booking (missing = no agent) | 304.0 | Float because of NaNs; useful in agency performance analysis. |
| 2️⃣5️⃣ | company | ID of company making the booking (missing = individual booking) | 223.0 | Very sparse; often NaN unless it's a corporate booking. |
| 2️⃣6️⃣ | days\_in\_waiting\_list | Days the booking waited before being confirmed | 0 | Long waits may affect cancellations. |
| 2️⃣7️⃣ | customer\_type | Type of customer group | Transient | Transient = single stay; Contract = corporate deal; Transient-party = family/friends group. |
| 2️⃣8️⃣ | adr | Average Daily Rate (EUR) paid per night | 100.0 | Price per night per room — helps study hotel pricing trends. |
| 2️⃣9️⃣ | required\_car\_parking\_spaces | Number of car parking spots the guest asked for | 0 | Useful for hotel logistics and parking management. |
| 3️⃣0️⃣ | total\_of\_special\_requests | Number of special requests made (e.g., crib, high floor) | 2 | Higher requests may mean high maintenance or VIP guests. |
| 3️⃣1️⃣ | reservation\_status | Final status: Canceled, No-Show, or Checked-Out | Canceled | Confirms how the booking ended, beyond just is\_canceled. |
| 3️⃣2️⃣ | reservation\_status\_date | Date the final booking status was set | 2016-09-01 | Useful for comparing booking vs. cancellation timing. |



**🔟 Top 10 Intermediate Analysis Questions (With Visual Suggestions & Columns)**

| **🔢** | **🧠 Question** | **🧱 Columns to Use** | **📊 Suggested Visual** |
| --- | --- | --- | --- |
| 1️⃣ | **Which months have the highest booking cancellations?** | arrival\_date\_month, is\_canceled | Bar Plot (canceled count by month) |
| 2️⃣ | **How does the average daily rate (ADR) vary by hotel type?** | hotel, adr | Box Plot or Bar Plot |
| 3️⃣ | **Which market segments cancel the most?** | market\_segment, is\_canceled | Bar Plot (cancellation rate per segment) |
| 4️⃣ | **Is there a relationship between lead time and cancellations?** | lead\_time, is\_canceled | KDE or Box Plot, then use correlation heatmap |
| 5️⃣ | **What is the distribution of stays (weekday vs weekend)?** | stays\_in\_week\_nights, stays\_in\_weekend\_nights | Histogram / KDE plot / Side-by-side bar |
| 6️⃣ | **Which countries have the highest number of bookings?** | country | Bar Plot (Top 10 countries) |
| 7️⃣ | **What is the trend of average ADR by month?** | arrival\_date\_month, adr | Line Plot or Bar Plot (monthly average ADR) |
| 8️⃣ | **Which room types are most often upgraded?** | reserved\_room\_type, assigned\_room\_type | Heatmap or countplot of mismatches |
| 9️⃣ | **Do guests with special requests cancel more or less often?** | total\_of\_special\_requests, is\_canceled | Bar Plot / Box Plot |
| 🔟 | **What is the correlation between numerical columns?** | All numerical columns like adr, lead\_time, stays\_in\_week\_nights, etc. | Heatmap (correlation matrix) |

# NEXT QUESTIONS:

| **No.** | **Question** | **Column(s) Used** | **Purpose / Insight** | **Example / Note** |
| --- | --- | --- | --- | --- |
| 16 | **How many bookings were changed after reservation?** | booking\_changes | Indicates booking reliability and user decision changes. | Count and histogram of booking\_changes. |
| 17 | **How do the number of adults affect the ADR and stay duration?** | adults, adr, stays\_in\_week\_nights, stays\_in\_weekend\_nights | Booking size influence on pricing and duration. | Group by adults, compare average adr, total nights. |
| 18 | **Which customer types are most likely to cancel their bookings?** | customer\_type, is\_canceled | Profiles guest segments by cancellation behavior. | Group by customer\_type and check cancel %. |
| 19 | **What is the cancellation rate among bookings with previous non-canceled records?** | previous\_bookings\_not\_canceled, is\_canceled | Tests customer consistency/loyalty. | Group into bins (e.g., 0, 1–3, >3 bookings). |
| 20 | **Which agents are linked to most cancellations?** | agent, is\_canceled | Evaluate business partners. | Top 10 agents by cancel rate. |
| 21 | **Is there a correlation between number of children and cancellation/ADR?** | children, adr, is\_canceled | Families vs solo travelers insights. | Boxplot of ADR vs children count. |
| 22 | **Does number of babies affect booking length or ADR?** | babies, adr, stays\_in\_week\_nights | For family-specific offers. | Correlation or mean comparison plots. |
| 23 | **Which distribution channel has the most reliable bookings?** | distribution\_channel, is\_canceled | Business channel performance. | Bar chart of cancel % by channel. |
| 24 | **Are companies associated with bulk bookings?** | company, adults + children + babies | B2B customer insight. | Group by company and sum total people. |
| 25 | **What is the ADR distribution for bookings from companies vs individuals?** | company, adr | Check if companies pay more/less. | Compare ADRs for null vs non-null company. |
| 26 | **What percentage of guests are repeated guests and how often do they cancel?** | is\_repeated\_guest, is\_canceled | Loyalty metric. | Pie chart for repeated vs non-repeated. |
| 27 | **How do different reservation statuses affect the ADR?** | reservation\_status, adr | Revenue trends by status types. | Group by reservation\_status, plot average ADR. |
| 28 | **What are the most common reservation statuses?** | reservation\_status | Booking lifecycle analysis. | Countplot of statuses: Check-Out, Canceled, No-Show. |
| 29 | **Does reservation status vary across months?** | reservation\_status, arrival\_date\_month | Cancellation seasonality. | Group by month and plot stacked bar of statuses. |
| 30 | **What is the pattern of reservation confirmation over time?** | reservation\_status\_date, reservation\_status | Booking trends by time. | Time series plot by reservation date. |