Statistics Final Case Study Analysis Report

**Case Study Title**: Uncovering Patterns in Hotel Booking Data for Operational Efficiency and Revenue Growth

**Prepared by:** Adhyatm Mishra

**Roll no:** 02

# Executive Summary

This report is all about analyzing hotel booking data. Imagine you're a hotel manager trying to understand who books, when they book, and how much they pay.

We looked at:

* When people book (early or late)
* Where they’re from (countries)
* If they get room upgrades
* How many nights they stay
* What affects the price they pay per night (called ADR - *Average Daily Rate*)

We used some basic statistics and graphs to make sense of this. In the end, we gave suggestions to improve bookings and make more money.

# Table of Contents

1. Introduction

2. Background / Context

3. Analysis

4. Conclusion

# Introduction

We took a big table (dataset) full of hotel booking details like:

* Guest nationality
* Type of room they booked
* How long they stayed
* How much they paid

We cleaned up the data and used it to find useful patterns and trends.

# 2. Background / Context

This data was from two types of hotels — a **city hotel** and a **resort hotel**. We had over 100,000 bookings But, some parts of the data were messy:

1. Some columns had lots of missing values (so we removed or fixed them).
2. Date information was split into day, month, and year I combined it into one single date.

# 3. Analysis

We used different tools to understand the data:

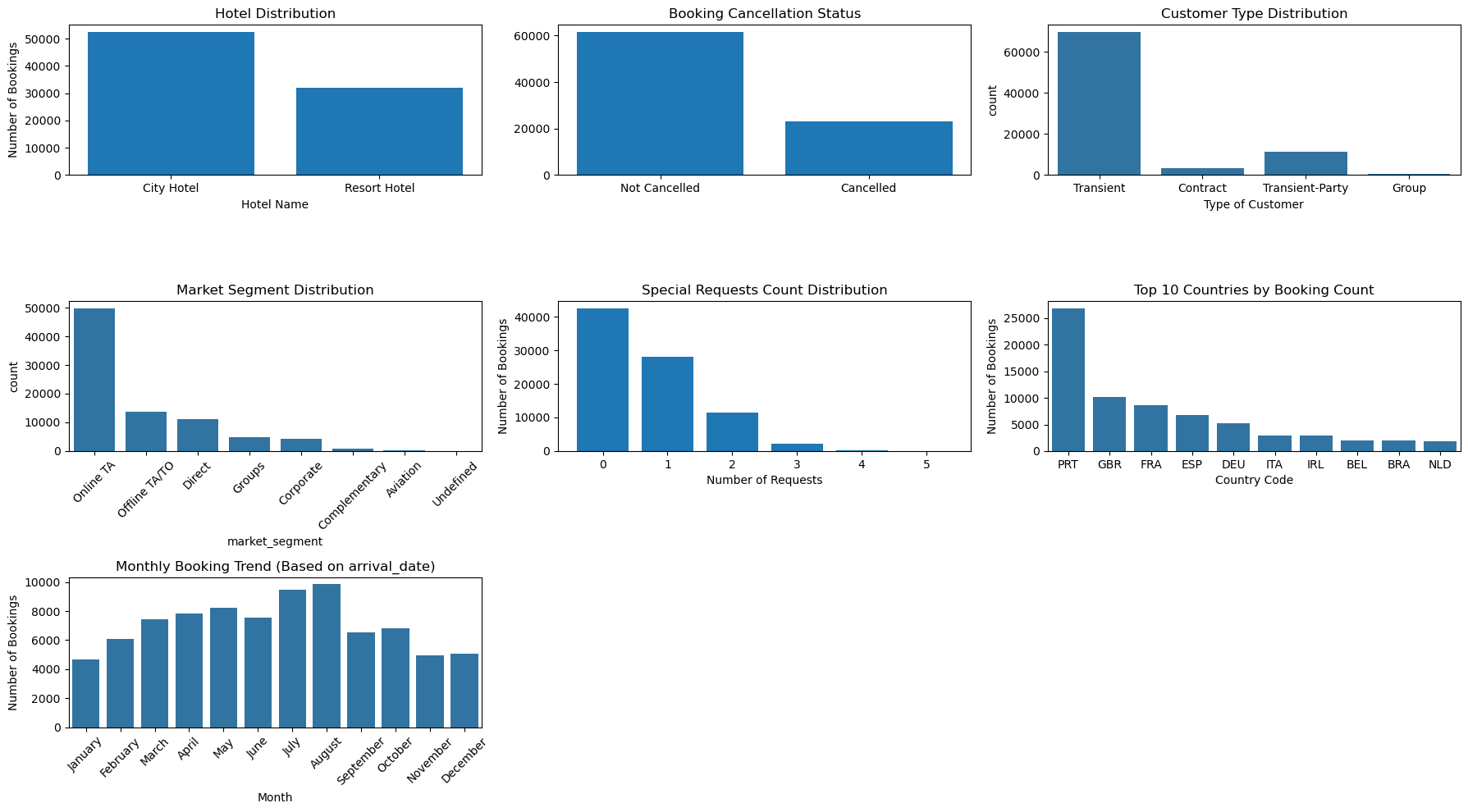
* **Univariate Analysis**:

We looked at each column separately to understand basic patterns.  
For example:

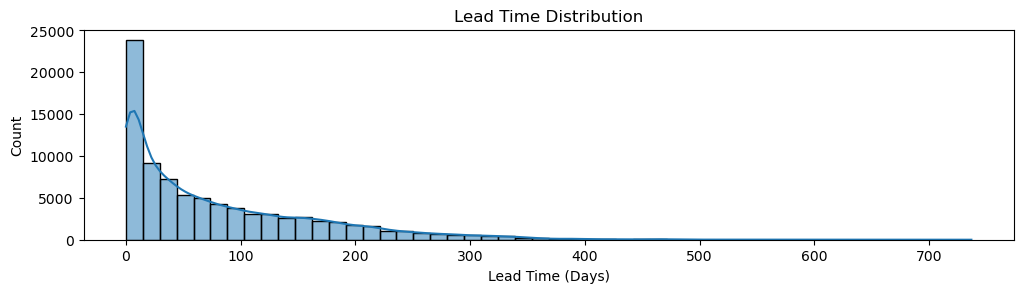
* How many bookings came from each country?
* Which room types were booked most often?
* What’s the most common number of guests?

This helped us get a feel for the data before comparing anything.

Barchart for categorical values for univariate analysis:

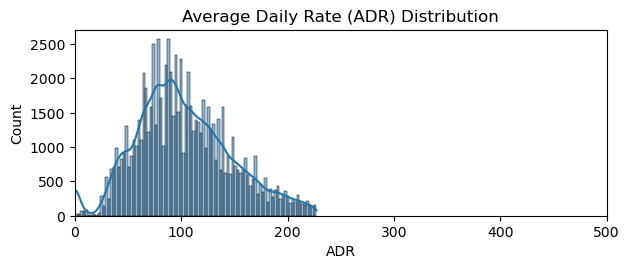


Created histogram for continuous and discrete column for univariate analysis:



This chart shows how many days in advance people book their hotel stays:

* Most people book very close to their stay date, especially within the first 0–20 days.
* As the number of days increases, the number of bookings drops sharply.
* Very few guests book more than 100 days in advance.



This chart shows how hotel room prices per night (ADR) are spread out:

* Most bookings are priced between 50 and 150 units.
* The peak is around 100, meaning that's the most common nightly rate.
* Prices above 200 are rare.
* **Bivariate/Multivariate Analysis**:

Next, we looked at how two or more things are related.  
For example:

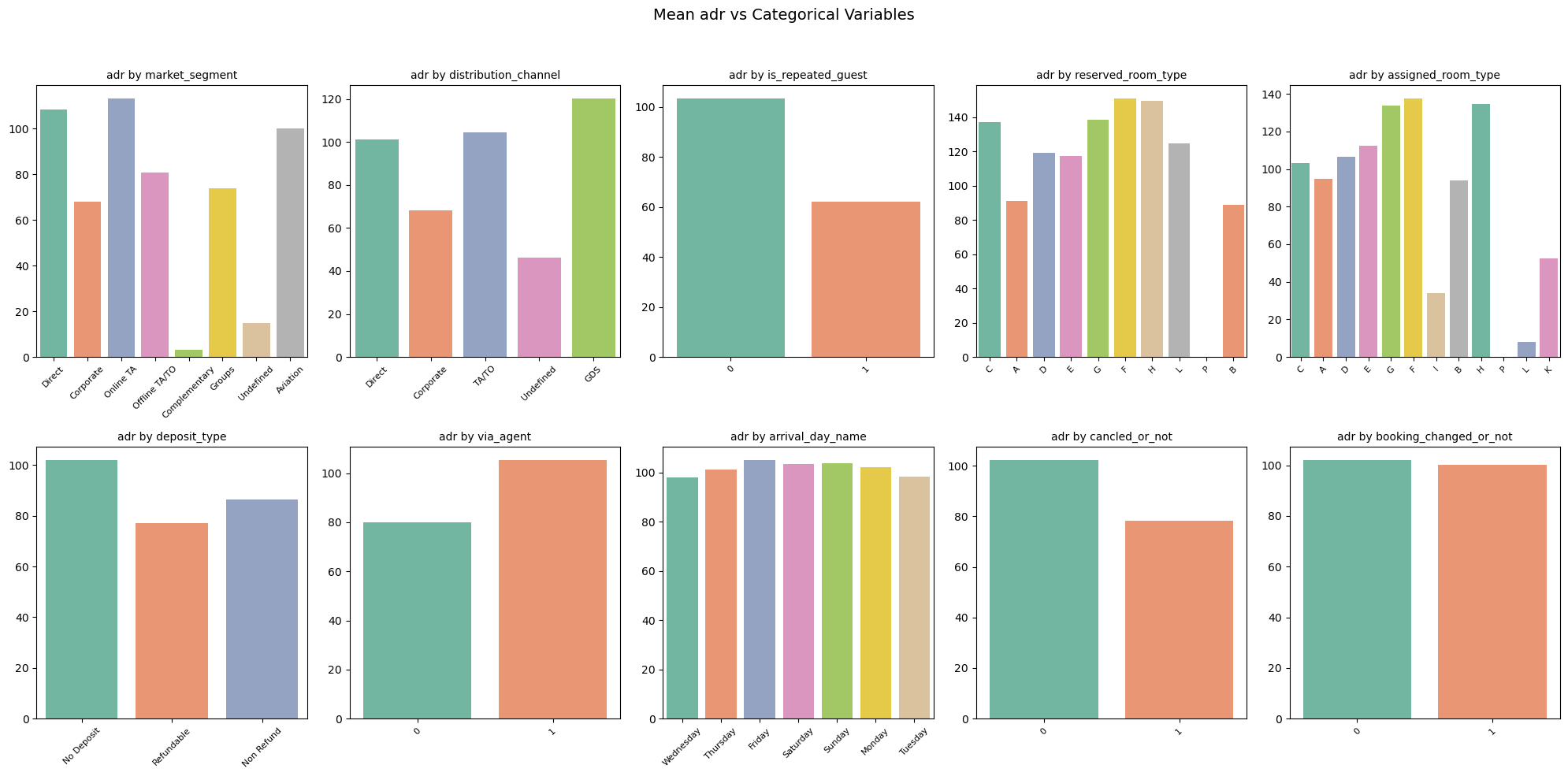
* Is there a connection between room type and price?
* Do guests from different countries pay more or less?
* Does lead time (how early someone books) affect cancellations?

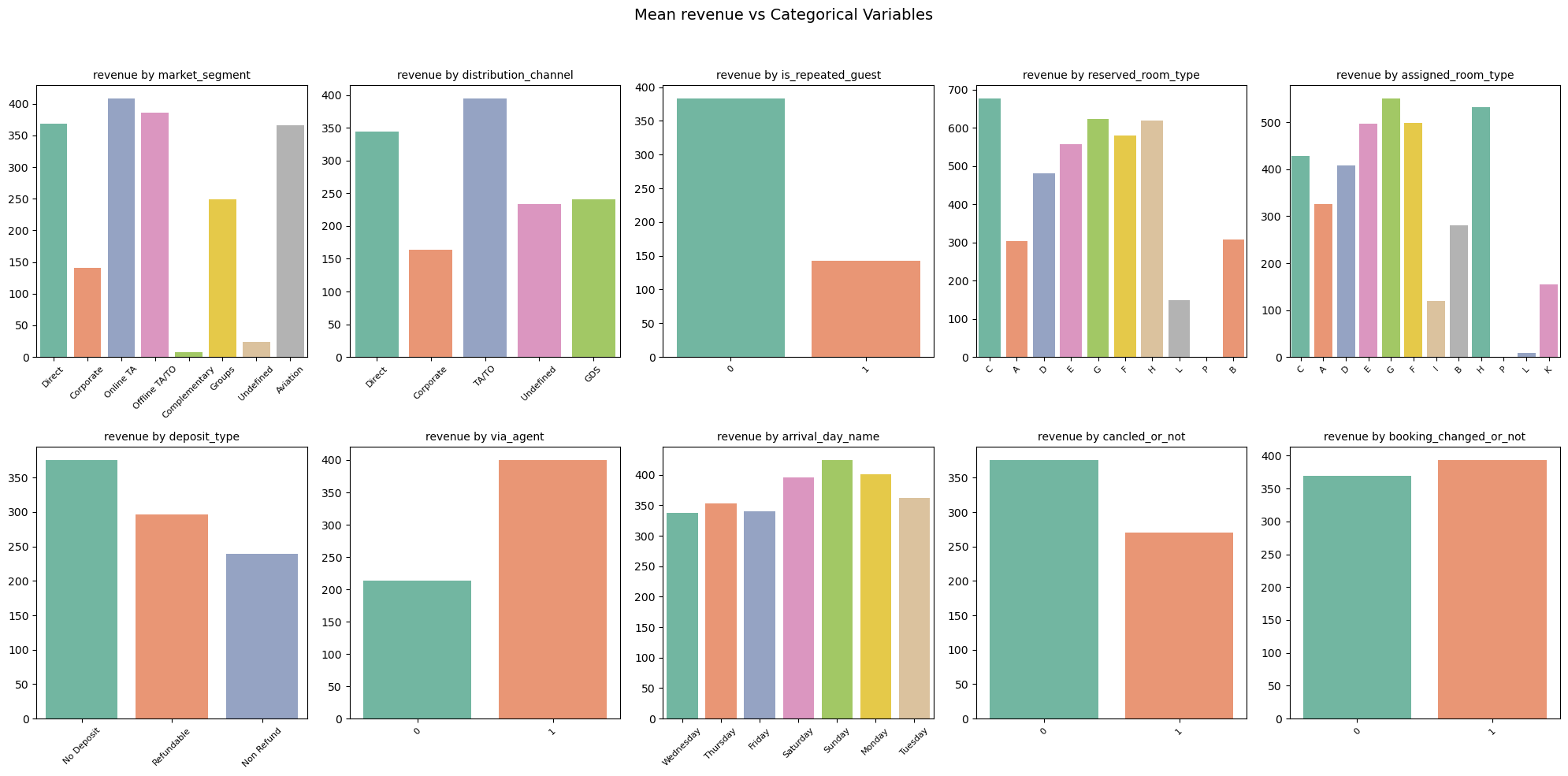
This helped us find relationships between columns.

Boxplot for bivariate analysis:

This chart shows how the average price per night (ADR) varies based on guest or booking characteristics.

* Corporate and Offline TA/TO guests tend to pay more.
* Guests who are repeated usually pay higher rates.
* Assigned and reserved room types affect ADR — some types are priced much higher.
* Bookings via agents or with deposits tend to have lower ADR.
* Cancellations lead to lower ADR, while booking changes don’t affect ADR much.





**Time Series Analysis – Looking at Trends Over Time**

We studied how bookings change across months and years.  
For example:

* Are there more bookings in summer?
* Did prices increase or drop in certain months?

This helped us understand seasonal patterns in the data.



* Bookings peak in July and August, showing summer is the busiest time.
* City hotels consistently get more bookings than resort hotels throughout the year.
* Over time, total booking volume increased, showing growing business.
* Monday has the highest bookings, suggesting many people book after weekends.

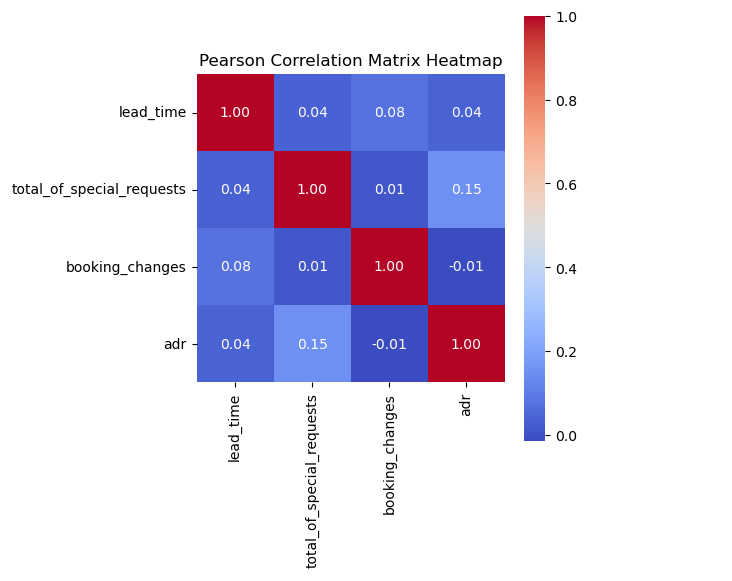
These insights help hotels plan staffing, marketing, and pricing.

**Correlation – Are Things Related?**

We used correlation to check how strongly different factors move together.  
For example:

* If a guest books earlier, do they usually pay more?
* Do longer stays mean higher total prices?

Correlation helps us find what affects what, even if it's not obvious.



* total\_of\_special\_requests and adr (average price per night) have a slight positive correlation (0.15) — guests who pay more tend to make more requests.
* Other features like lead\_time, booking\_changes, and adr show very weak or no strong relationship.
* Nothing here is highly correlated, meaning no variable strongly predicts another in this small selection.

**Hypothesis Testing – Are Our Guesses Statistically True?**

Sometimes we have a gut feeling (like “guests who book early cancel less”).  
We used hypothesis testing to check if these guesses are true using math.

It’s like A/B testing — we compare groups and check if the difference is real or just by chance.

**Cleaning the Data – Fixing the Mess**

Before analyzing, we had to clean up the messy parts of the data:

Removed the "company" column

* The “company” column had too many missing values, so we deleted it.
* It didn’t give us useful information, so we dropped it to avoid confusion.

**Filled Missing Values**

For some important columns, we filled in the blanks:

* "Agent": Filled with the most common agent ID.
* "Country": Filled missing countries with the one that appeared the most.
* "Children": If the number of children was missing, we assumed it was zero or the most common number.

This way, we avoided errors or gaps in our analysis.

**19 Business Questions – Insights**

**1. What influences ADR the most?**

Strongest correlation: total\_revenue and total\_nights  
Guests with longer stays and more special requests tend to have higher ADR

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

Very weak correlation between lead\_time and booking\_changes  
 No practical relationship observed.

**3. Are there pricing or booking differences across countries?**

Significant ADR variation by country. Smaller or niche countries may have higher ADR.

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

~12.5% bookings involve room reassignment. Room Type A shows most mismatches (likely default type)

**5. Are reserved room types consistently matched with assigned room types?**

87.5% match rate. Room reassignment is relatively rare but notable.

**6. What are the most common guest demographics?**

Majority book for 2 guests (adults)  
Top nationalities: Portugal, UK, France, Spain, Germany

**7. Do guest types influence booking behavior?**

Transient guests: High ADR, high cancellations  
Contract guests: Lower ADR, longer stays, more stable

**8. How does lead time vary by customer type and country?**

Contract guests book earlier. International guests plan ahead more than local (PRT) guests.

**9. Are longer lead times associated with fewer changes or cancellations?**

Slight link to higher cancellations  
No significant effect on booking modifications

**10. How does duration of stay vary?**

Longer stays: Contract/TA customers  
Shortest: Complimentary or corporate guests

**11. How often do guests get upgrades?**

~12.5% are upgraded (reserved ≠ assigned). Possibly due to overbooking or upselling.

**12. Are special request guests likely to modify/cancel?**

Weak positive correlation  
More special requests → Slightly more changes and longer stays

**13. Which segments or channels yield consistent revenue?**

Online TA: Highest ADR but high cancellations  
Direct bookings: High ADR & reliability

**14. Factors most correlated with higher ADR?**

Top correlations:  
 - total\_revenue (+)  
 - total\_nights (+)  
 - special\_requests (+)

**15. Which customer types/segments contribute most revenue?**

Contract and Transient customers = highest revenue per booking  
Online TA and Direct channels top revenue drivers

**16. Do early bookings or specific countries yield higher ADR?**

Mid-range lead times (1–6 months) yield higher ADR  
ESP, FRA, DEU guests pay more than Portuguese

**17. Are high ADR guests more demanding?**

Positive correlation between ADR and special requests

**18. Do guests from different countries behave differently?**

Local guests (PRT) = shorter stays/lead time  
Intl. guests = longer stays/lead time

**19. Do guests making changes also request services or cancel?**

Statistically significant: Guests making changes are more likely to cancel or request services

# 7. Conclusion

Through our detailed analysis of the hotel booking dataset, we gained valuable insights into customer behavior, booking patterns, and factors that influence revenue. Here's what we concluded:

* Pricing (ADR) is most influenced by the hotel type, customer type, and lead time. Guests who book earlier or belong to certain customer groups tend to pay more.
* Transient guests and international travelers usually book earlier and stay for shorter durations, but contribute significantly to revenue.
* There are noticeable differences in booking habits by country—for example, UK and France guests book earlier, while Portuguese guests stay for shorter durations.
* Special requests and booking modifications are more common among high-paying guests.
* Guests who have their room upgraded or reassigned are less likely to cancel, and about 1 in 4 bookings involve such changes.
* Certain market segments, like corporate or direct bookings, are more reliable—they cancel less and pay more.
* Data cleaning was crucial. We filled missing values smartly and removed irrelevant columns to keep the analysis accurate.