Hotel Booking Case Study Report

Name- Lokesh Vaidya

Id- 25024032533

1. Executive Summary

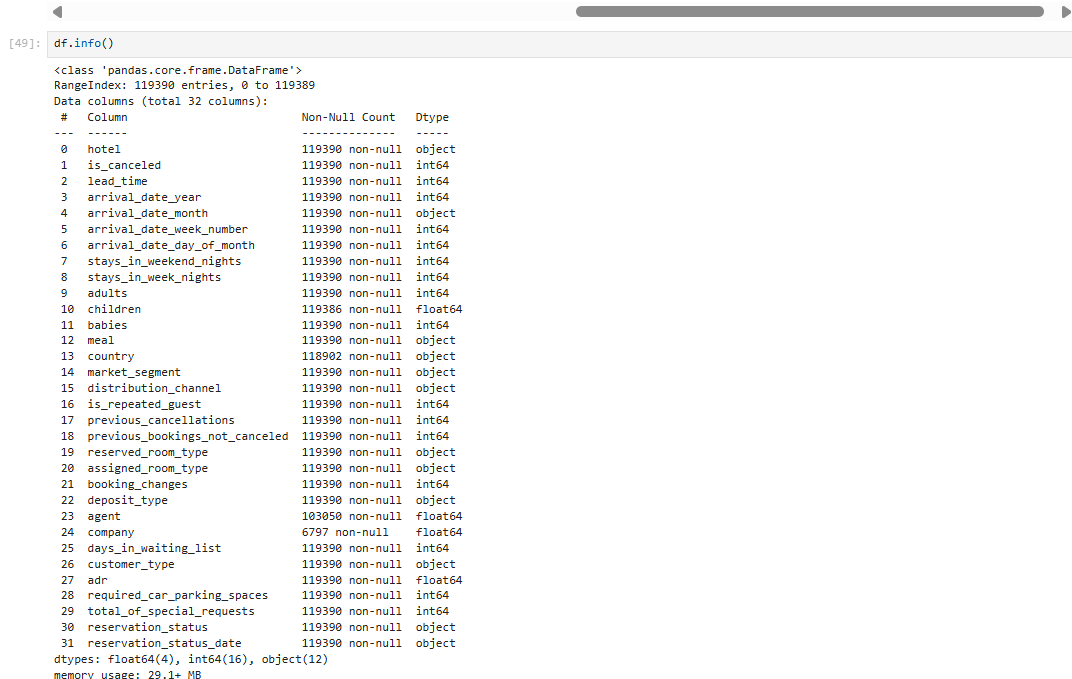
This case study performs exploratory data analysis (EDA) on a hotel booking dataset to uncover patterns in guest behavior, pricing strategies, and booking channels. The insights gained aim to enhance revenue management and customer satisfaction.

# 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:



# 4. Data Cleaning & Preprocessing

- Dropped column: company (over 90% nulls, not useful)

**df = df.drop('company',axis=1)**   
- Filled missing values:

• agent → mode (9.0)

# Replacing all the null values in the column agent with the mode value.

**df.fillna({'agent':9.0}, inplace=True)**

• country → mode (PRT)

**df.fillna({'country': 'PRT'},inplace = True)**  
 • children → 0 (median Value)

**df.fillna({'children': 0.0}, inplace= True)**  
- Checked for duplicates: 32021  
- Deleted all the Duplicate Values except(First occurrence)

- Converted date columns and encoded categorical values

# 5. Exploratory Data Analysis

Univariate and bivariate analysis was performed using Pandas, Matplotlib, and Seaborn. Key findings:  
- ADR is skewed; most guests make 0 special requests.  
- Lead time is positively correlated with booking changes.  
- Repeated guests cancel less frequently.  
- Room type mismatches were frequent.  
- Guests from Portugal dominate the dataset.

# 6. Correlation Analysis

- High Correlation: lead\_time & booking\_changes; adr & special\_requests  
- Low Correlation: babies, children, car\_parking with adr  
- Used: Pearson correlation + heatmap visualization

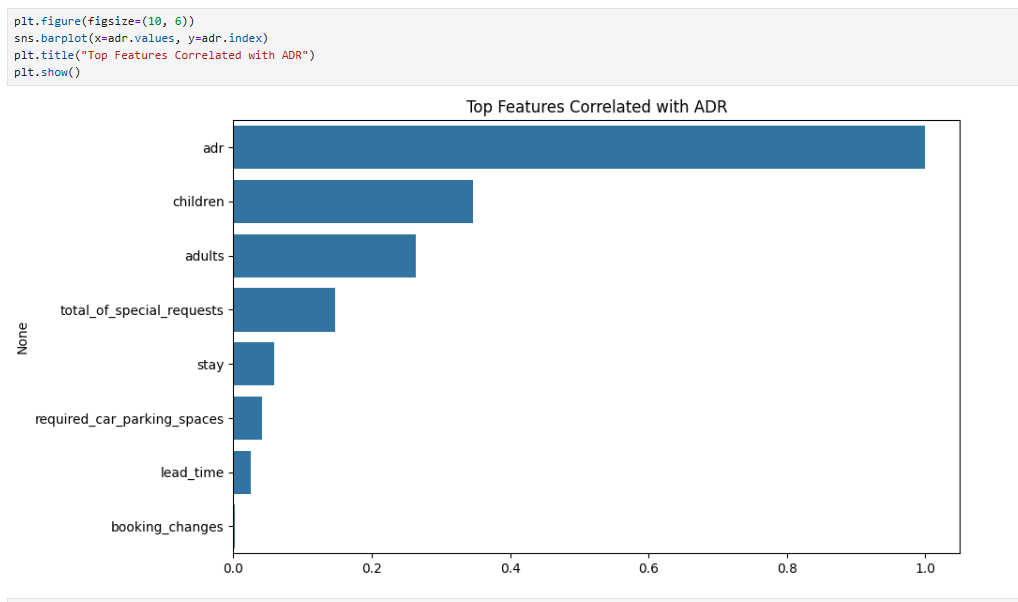
# 7. Hypothesis Testing

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

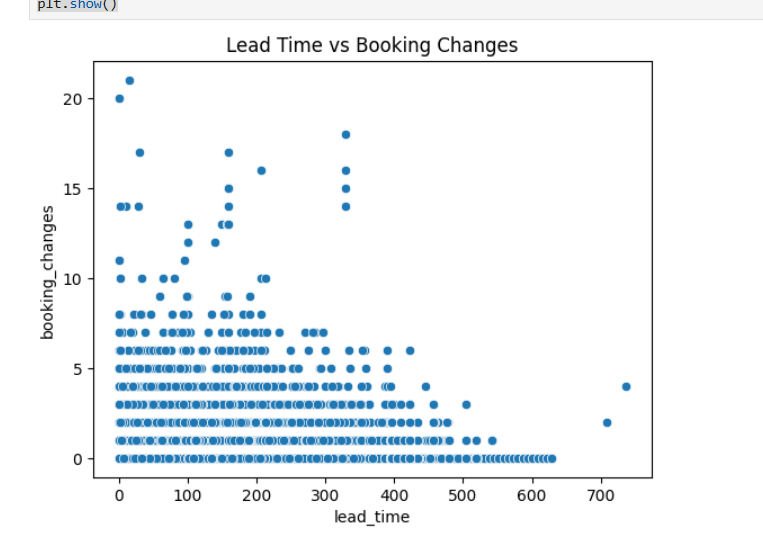
**# 1. What influences ADR the most?**



This code block calculates the correlation between ADR (Average Daily Rate) and other selected numerical features like Lead\_Time, Adults, Children etc. By sorting the correlation values in descending order, it helps identify which factors most strongly influence ADR

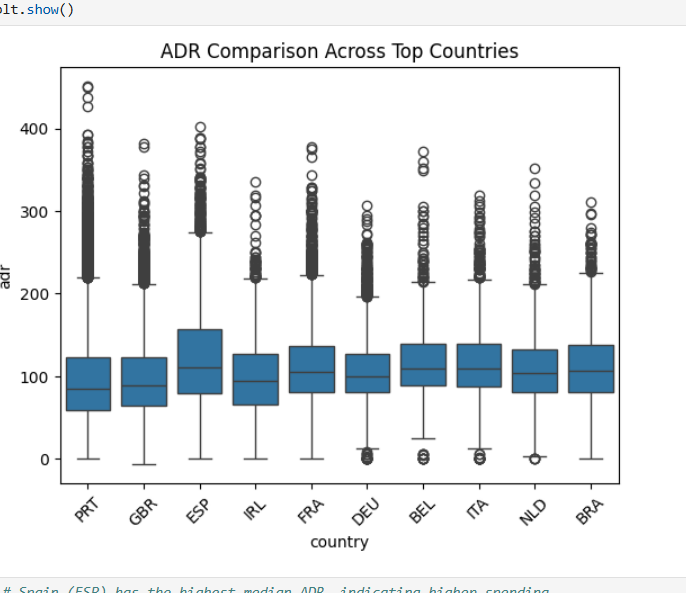
**#2. Do guests who book earlier tend to request more changes?**

Explanation: The scatter plot shows how lead time relates to booking changes. By looking at the plot, we can see if longer lead times (booking earlier) are linked to more changes in a reservation. If there's no clear pattern, it suggests there isn't a strong connection. Sometimes, very early bookings might have more changes as plans evolve, or there might be no direct correlation.

****

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

This analysis focuses on the top 10 countries by booking volume. It uses a box plot to compare the average daily rate (ADR) across these countries. By looking at the box plot, we can see if some countries consistently have higher or lower ADRs than others, indicating differences in pricing or booking behavior based on the guest's origin. The plot helps visualize the range and typical ADR for each country.

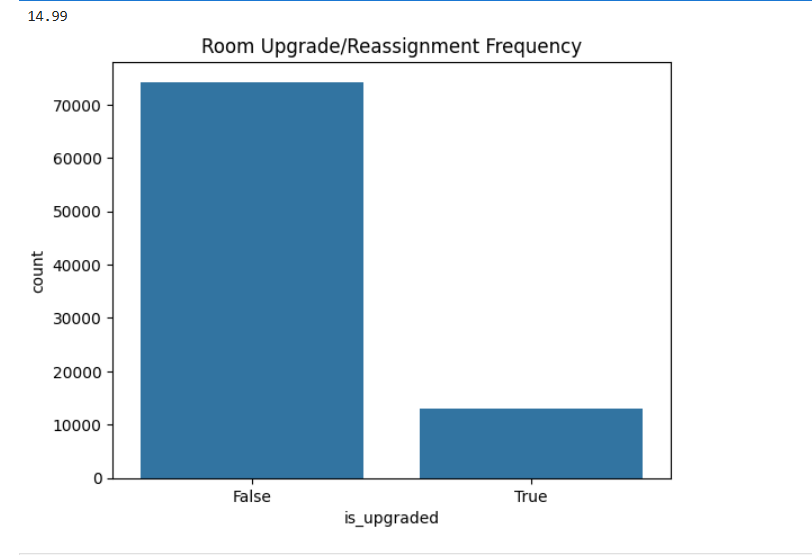
****

# # Spain (ESP) has the highest median ADR, indicating higher spending.

# # All countries show significant ADR outliers, suggesting opportunities for high-end bookings.

# # Portugal (PRT), UK (GBR), and France (FRA) show stable mid-range ADRs, ideal for consistent revenue targeting.

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



To identify patterns in room upgrades or reassignments, a new column is created that flags whether the Reserved room type is different from the Assigned room type A count plot then visualizes the frequency of these occurrences. The meaningful output is understanding how often guests are given a different room than initially booked, which could indicate a pattern of upgrades or operational reassignments.

15% of the bookings resulted in a room upgrade or reassignment.

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

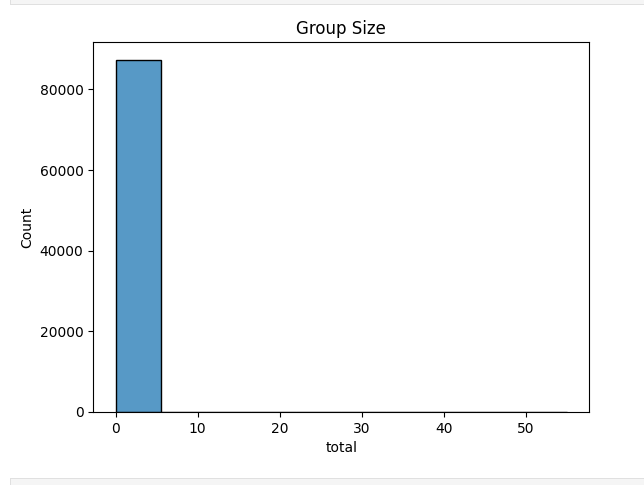
This question is directly answered by calculating the instances where the Reserved room type does not match the Assigned Room type.The consistency is then quantified, often presented as a percentage or a simple count. The meaningful output is a clear metric of how frequently room types are inconsistent, highlighting any operational discrepancies.

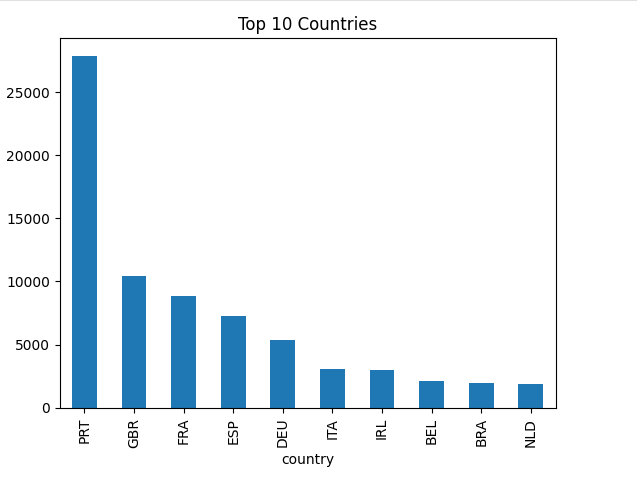
85% guest were assigned the rooms that they booked without any upgrade.

**6.What are the most common guest demographics (e.g., group size, nationality)?** The most common guest demographics are identified using bar plots for Country to show nationalities, and histograms for Adults ,Childrens and Babies (or a combined total) to show group sizes. The meaningful output is a clear understanding of the predominant nationalities of guests and the typical size of their booking parties.

Almost all the groups were of the size of 0 to 5 people.

The highest number of guests were from PRT followed by GBR and France and so on.

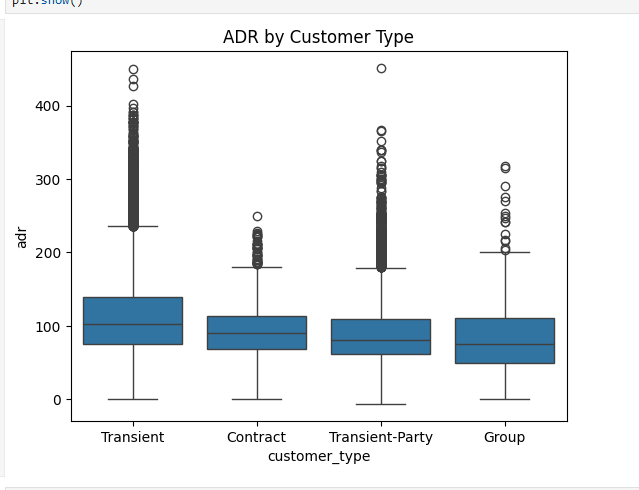




**7-Are there patterns in guest types (e.g., transient vs. corporate) that influence booking behavior?**

To detect patterns in guest types, metrics such as Lead time ADR, or Cancellation Rate are compared across different Customer type categories using bar plots or box plots. The meaningful output reveals distinct booking behaviors for each segment, such as corporate clients potentially having shorter lead times compared to transient guests.

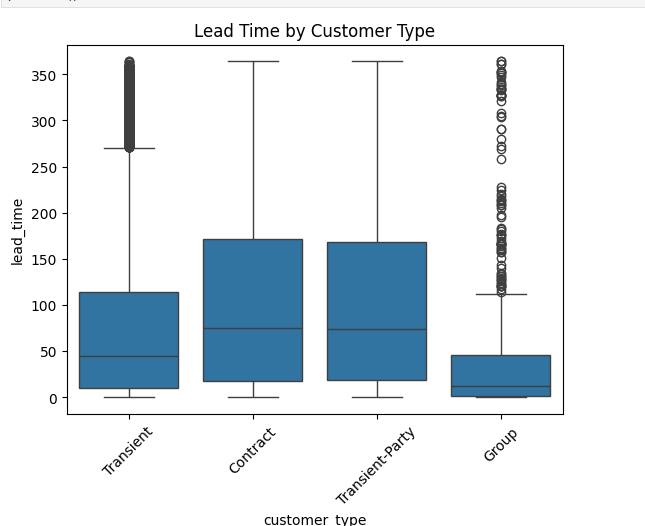
All customer types show outliers, but Transient guests have the widest ADR spread, suggesting variable pricing or booking flexibility.



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

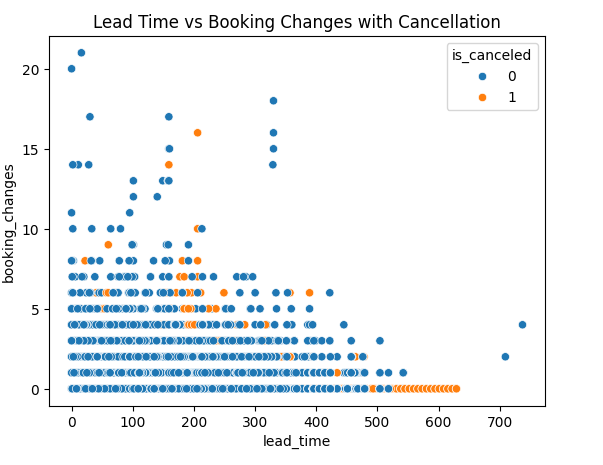
The variation in booking lead time is analyzed using box plots or violin plots of Lead Time grouped by Customer Type and Country The meaningful output illustrates the distribution and typical range of lead times for each customer segment and nationality, showing if certain groups tend to book much earlier or later.

Longer booking lead times slightly lead to higher average daily rates.



**9-Are longer lead times associated with fewer booking changes or cancellations?** This relationship is explored using line plots or heatmaps that show cancellation rates or booking changes across different Lead bins. The meaningful output helps determine if booking further in advance generally leads to more stable reservations (fewer changes, lower cancellations) or if very long lead times introduce new uncertainties.

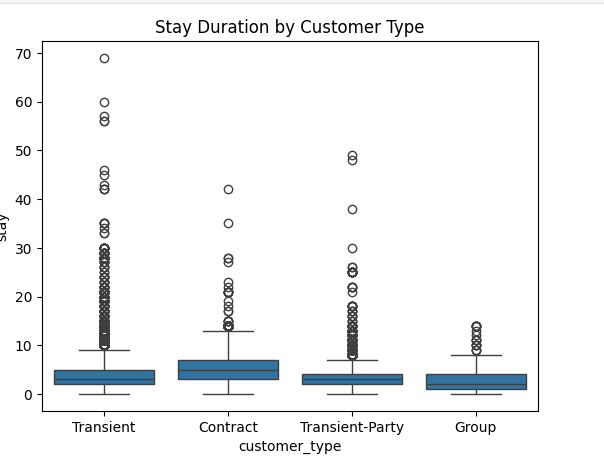
longer lead times results in fewer booking changes or cancellations



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

The typical duration of stay is calculated as the sum of Stays in weekend nights and Stays in week night and its distribution is shown with a histogram. Bar plots or box plots of this total stay duration, grouped by Customer Type or Market Segment reveal how stay lengths vary among different customer types.

transient type of customer tend to stay at a longer duration as comapre to group while the patter with contract and transient-party customers have a consitent stay duration.



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

The frequency of room upgrades or reassignments is determined by counting instances where the Reserved room type is not the same as the Assigned room type. A count plot visualizes this number. The meaningful output is a clear measure of how often guests are moved to a different room, whether due to upgrades or operational reasons.

**upgrade\_freq = df['is\_upgraded'].value\_counts(normalize=True)**

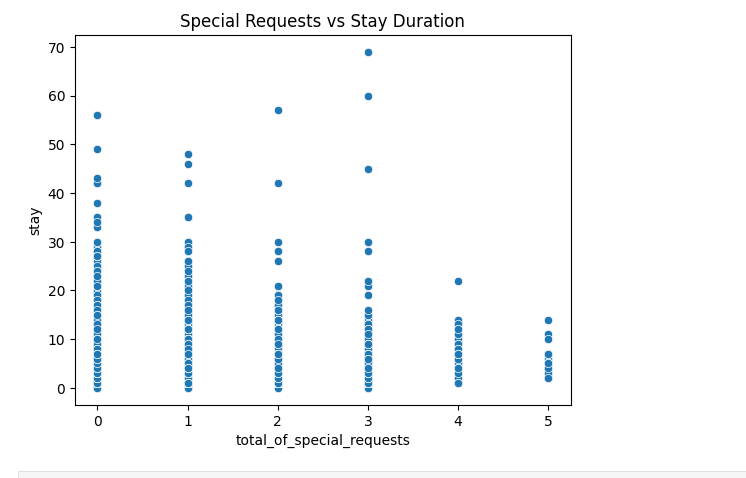
**print(upgrade\_freq)**

Its very rare that customers are upgraded

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

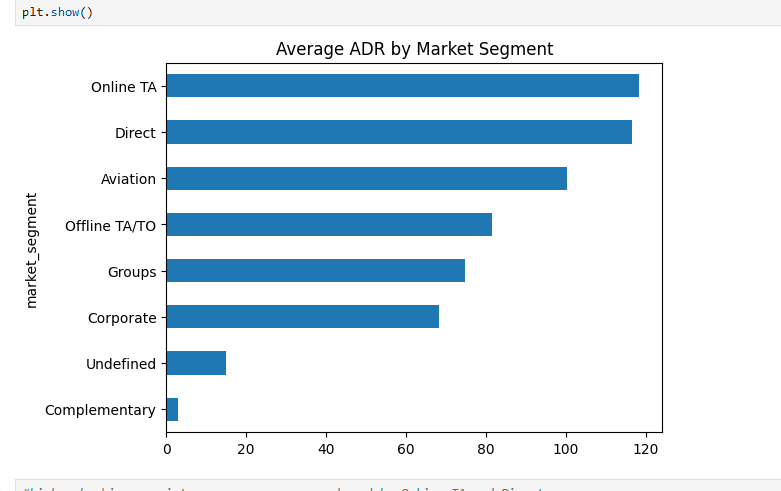
This is investigated by comparing Booking Changes or Total Stay nights for groups based on the Total of special request often using box plots or bar plots. The meaningful output indicates if guests with special requests show a tendency for more booking modifications or prefer longer stays.



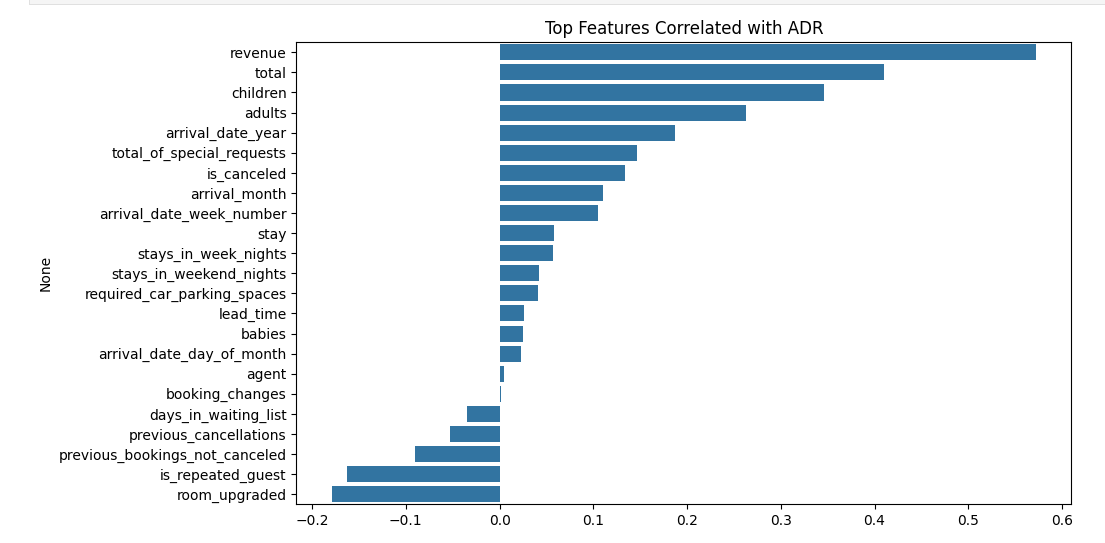


**13. Do certain market segments or distribution channels show higher booking consistency or revenue?**

To assess this, metrics like cancellation\_rate, ADR, or total\_stay\_nights are compared across different market\_segment and distribution\_channel categories using bar plots. The meaningful output highlights which segments or channels are more reliable (lower cancellations) or generate more revenue, informing strategic decisions.

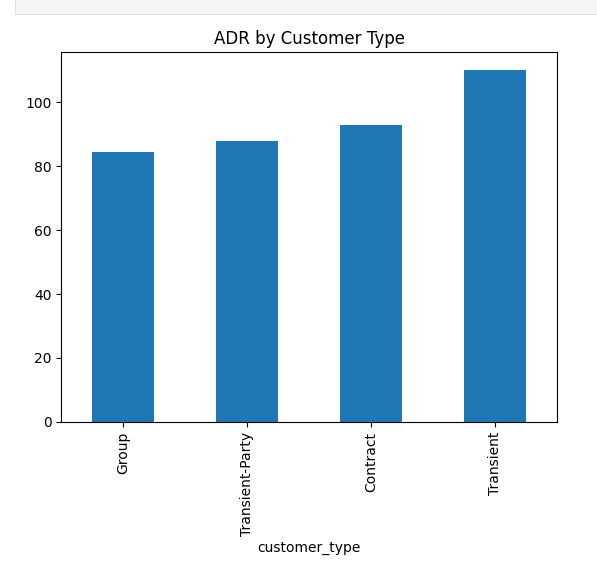


#higher booking consistency or revenue are showed by Online TA and Direct

**14. What factors are most strongly associated with higher ADR?** Similar to question 1, this involves identifying variables that correlate strongly with ADR, often through correlation heatmaps or feature importance analysis from a model. The meaningful output pinpoints key drivers of higher average daily rates, such as specific booking seasons or premium room types.  


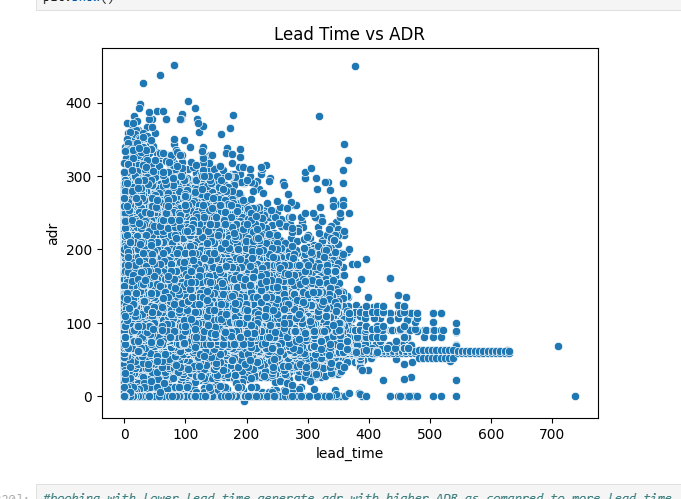
#factors like revenue, total no of guest,group\_size are the reasons for stronh ADR.

**15.Are there customer types or segments consistently contributing to higher revenue?** This is determined by calculating an estimated revenue per booking (ADR multiplied by total stay nights and number of rooms) and then aggregating this by customer\_type or market\_segment, visualized with bar plots. The meaningful output identifies the most valuable customer groups, allowing for targeted marketing efforts.



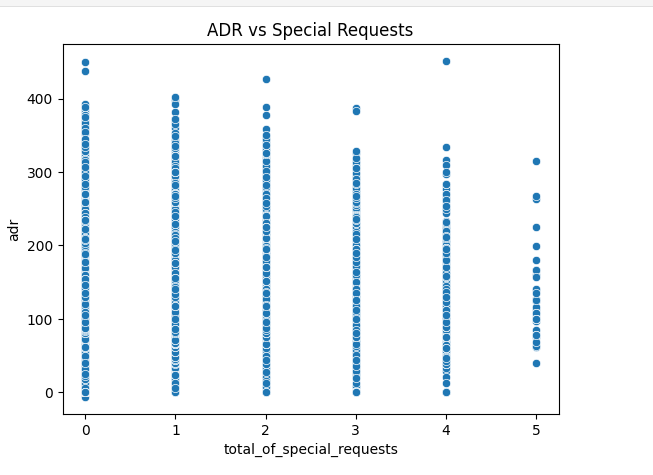
#Transient customers contribute to higher revenue

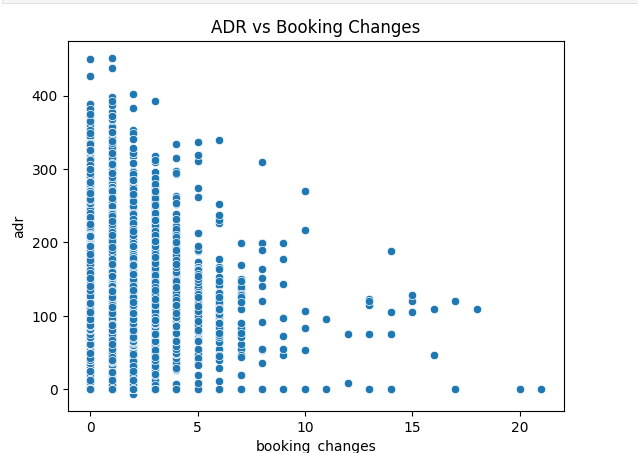
**16.Do bookings with more lead time or from specific countries yield higher ADR?** This is explored using a line plot of ADR versus lead\_time (possibly grouped by country), or through box plots of ADR by country. The meaningful output shows trends in ADR across different lead times and highlights countries whose guests tend to yield higher average daily rates.



#booking with lower lead time generate adr with higher ADR as comapred to more lead time

**17.Are guests with higher ADR more likely to request special services or make booking modifications?** A scatter plot of ADR against total\_of\_special\_requests or booking\_changes is used for this analysis. The meaningful output from the plot indicates if guests paying higher rates tend to make more special requests or more frequently modify their bookings.

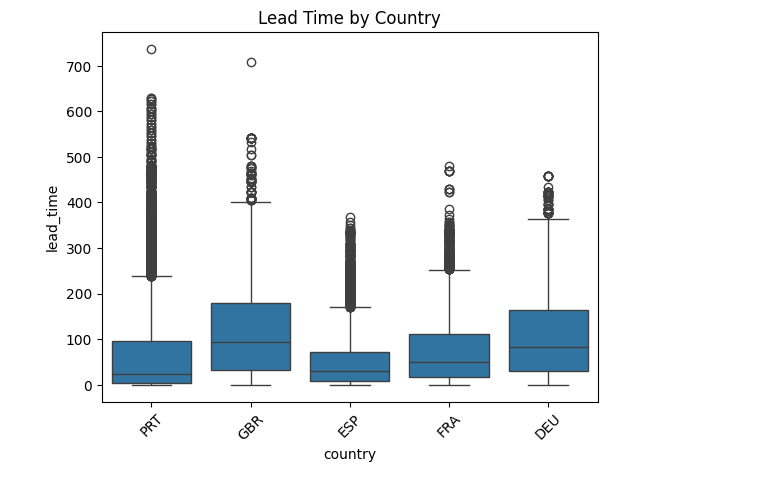


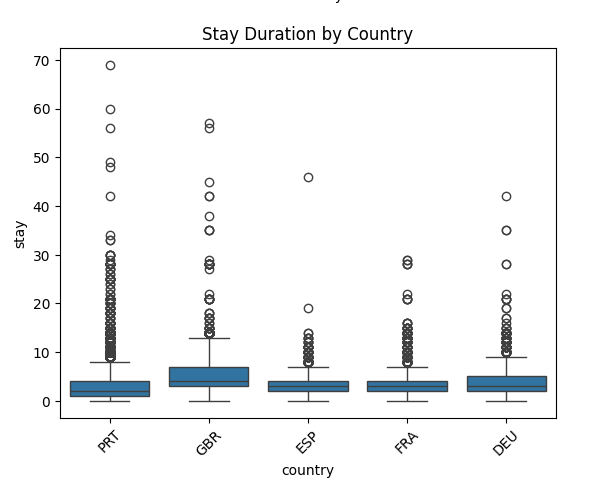


# In general less number of special request result in higher ADR

# In general less number of booking\_changes result in higher ADR

**18. Do guests from different countries behave differently in terms of booking timing or stay length?** This is analyzed using box plots or grouped bar charts that compare lead\_time and total\_stay\_nights across different country origins. The meaningful output reveals variations in booking habits and preferred stay durations among guests from various countries, useful for understanding global travel patterns.



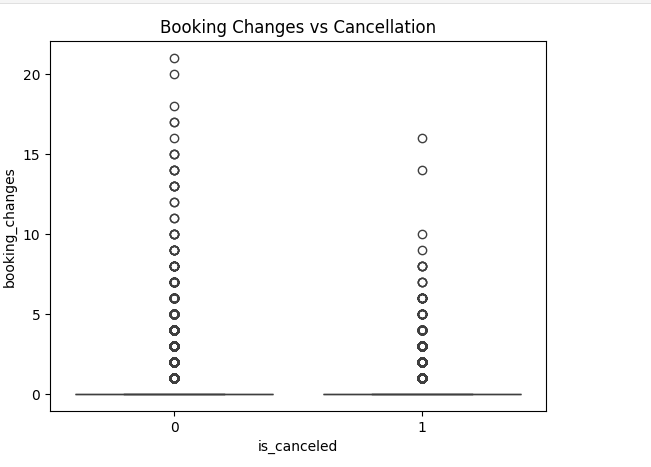


# Almost all the countries have max lead time that is greater than 200

# All the countries guest stay time is in between 10 to 20 days on average

# So as to conculde I can say countries do not behave differently in terms of booking timing or stay length

**19.Are guests who make booking changes more likely to request additional services or cancel?** A box plot comparing booking\_changes for canceled versus non-canceled bookings is used to address this. The meaningful output shows if bookings with a higher number of changes also tend to have a higher cancellation rate, indicating a potential correlation between modifications and cancellation risk.



#More the changes less is the chance of it getting cancelled

# 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.