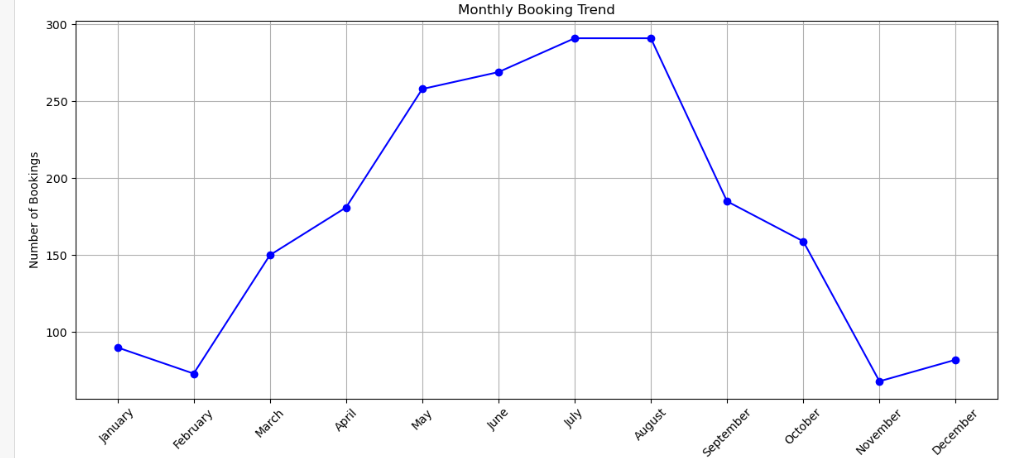
**HOTEL-PROJECT**

**PART – 1**

* **Task 1 – Data Preprocessing & Cleaning**

# ****Task 2 - Booking Pattern Analysis****

## **2.1 - What are the peak booking periods for the hotel?**

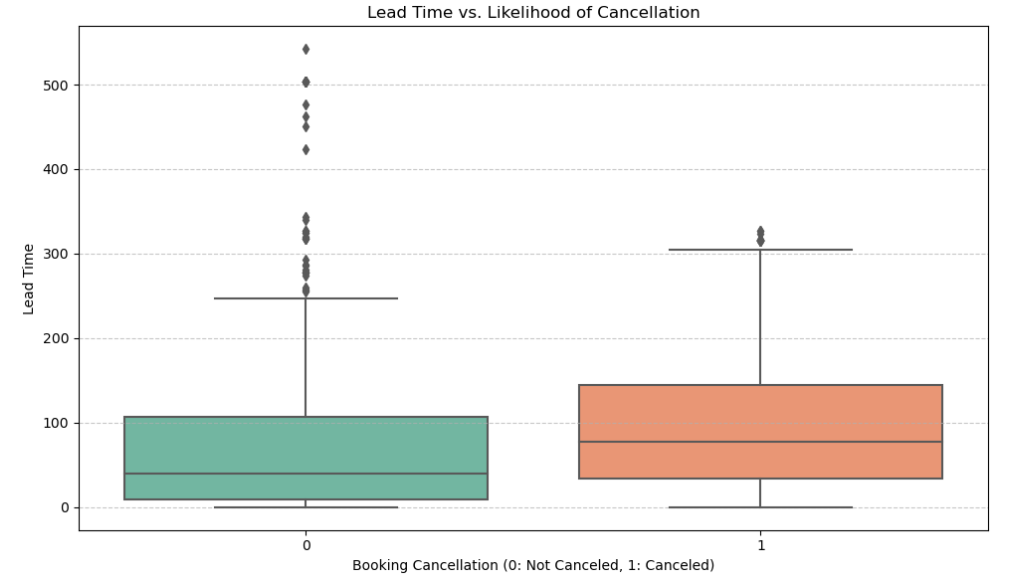


## **Interpretation:**

This analysis represents the monthly booking trends.

* The month of **July** has the highest booking trend followed by **June & August**.
* The month of **November** has the lowest booking trend followed by **February & December**.

## **2.2 - Is there a relationship between lead time and the likelihood of a booking being cancelled?**

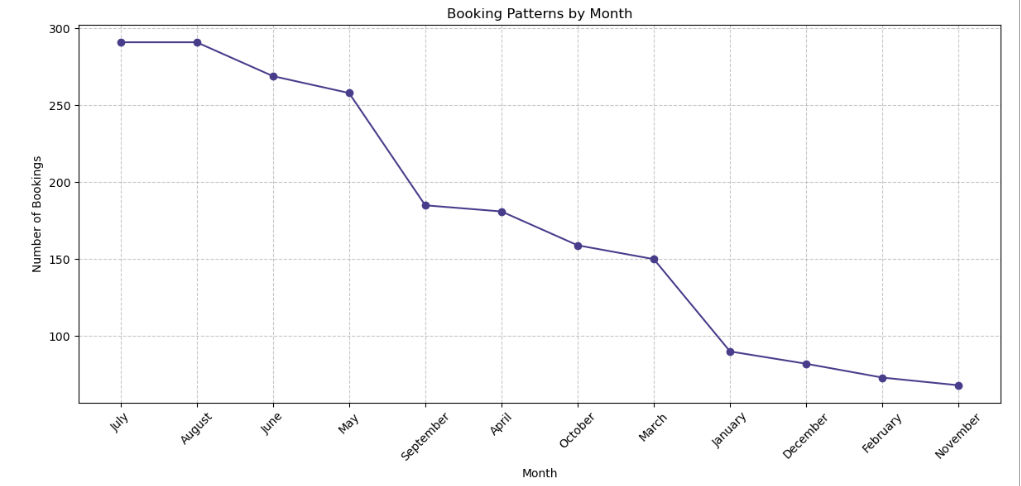
****

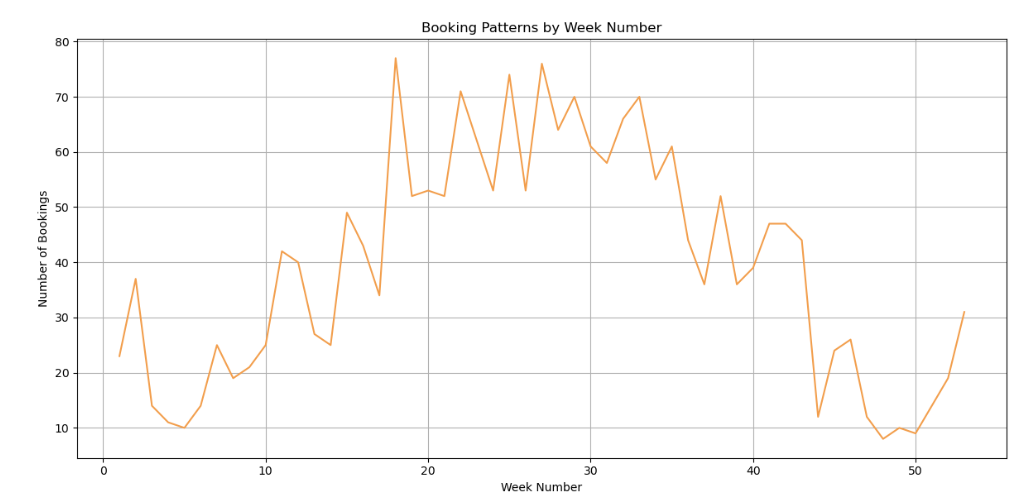
## **Interpretation:**

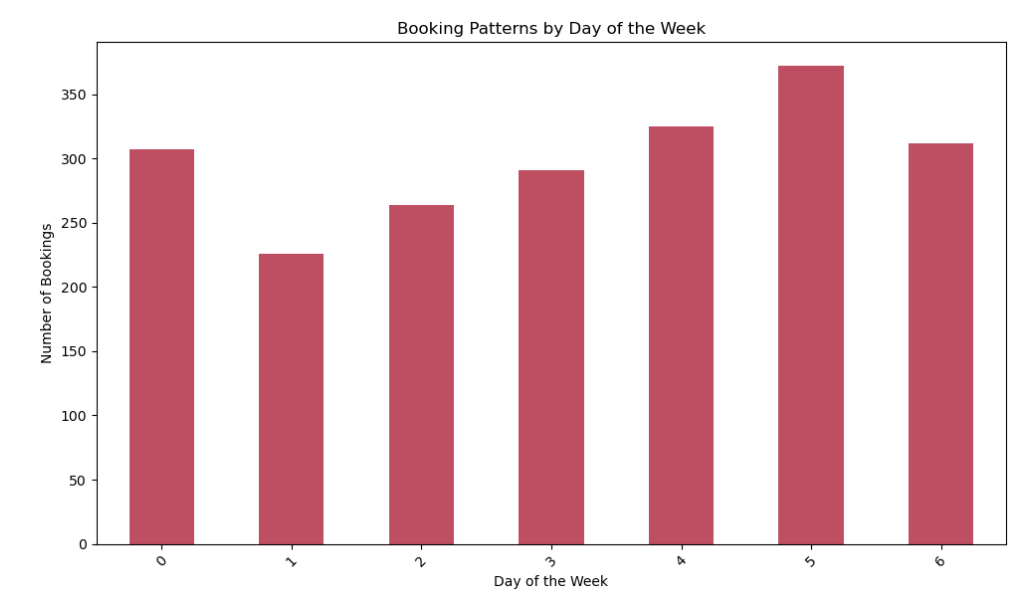
This analysis helps in understanding the relationship between lead time and booking cancellations, which can be valuable for managing reservations and predicting cancellation rates 0 indicates that the booking was not canceled. 1 indicates that the booking was canceled.

* The middle line in the **booking cancellation - 0 Not Cancelled** indicates that the mean is greater than median and the data values are exceptionally high values towards top.
* The middle line in the **booking cancellation - 1 Cancelled** indicates that the median is greater than mean and the outliers data values are exceptionally high values towards top and these outliers are low compared to **Not cancelled**.

## **2.3 - How do booking patterns vary by month, week, and day of the week?**



****

****

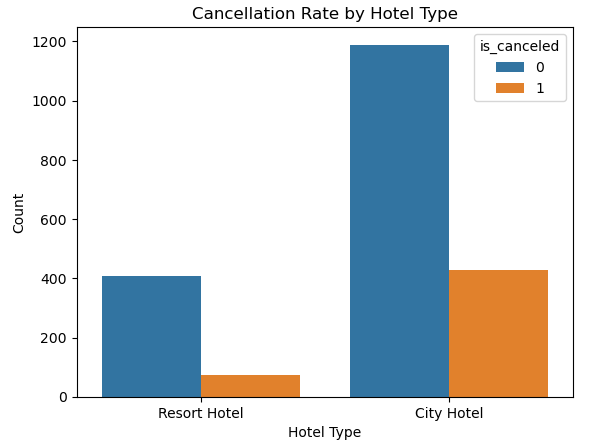
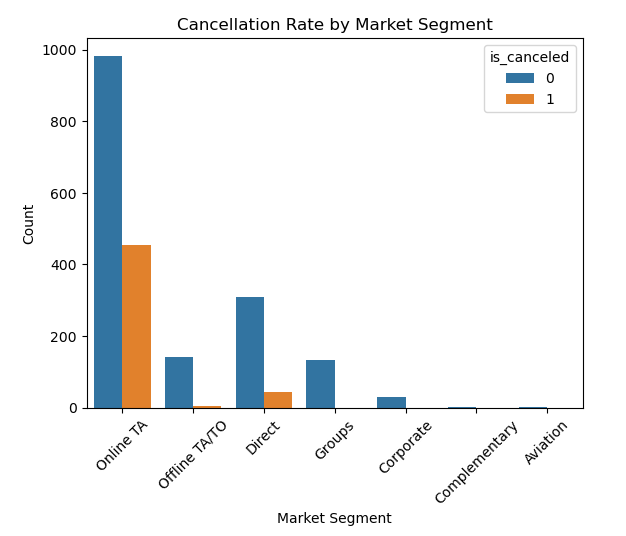
## **Interpretation:**

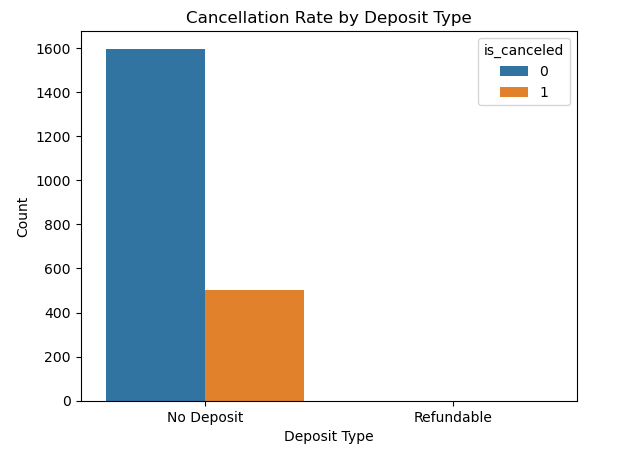
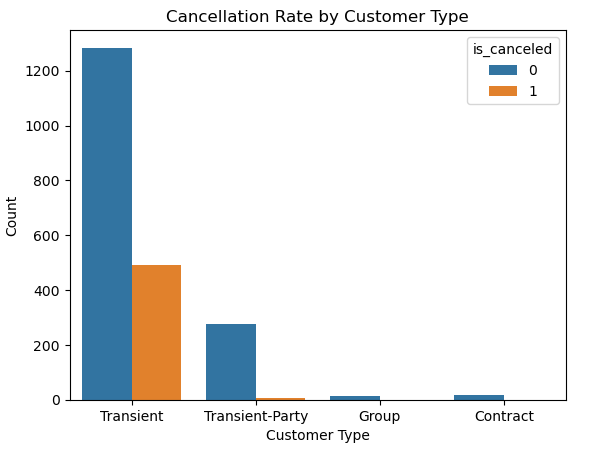
This analysis is to visualize booking patterns by different time intervals: month, week number, and day of the week.

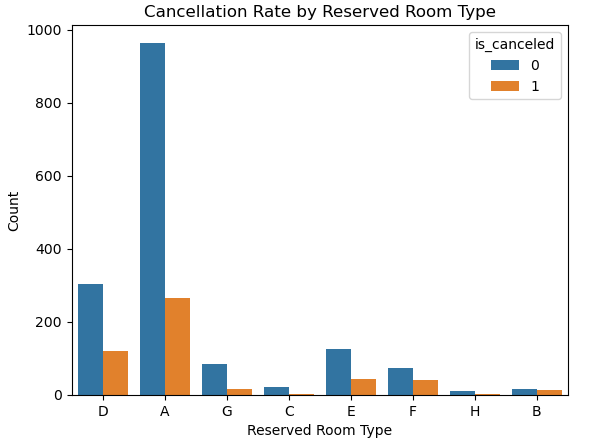
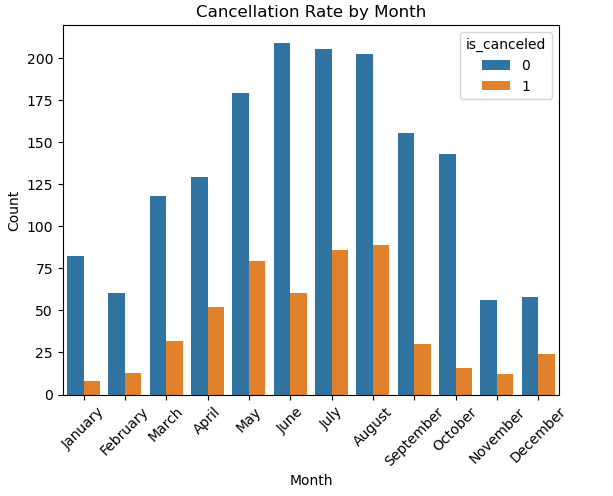
* In the first plot we can observe the seasonal variation in booking activity, with certain months - **July, August & June** experiencing higher booking volumes compared to others.
* The second plot is to identify any weekly patterns in booking behavior, where number - **17** has the higher trend compared to others.
* The third plot is to identify the distribution of bookings across different days of the week, where the day - **5** is tend to be higher compared to others.

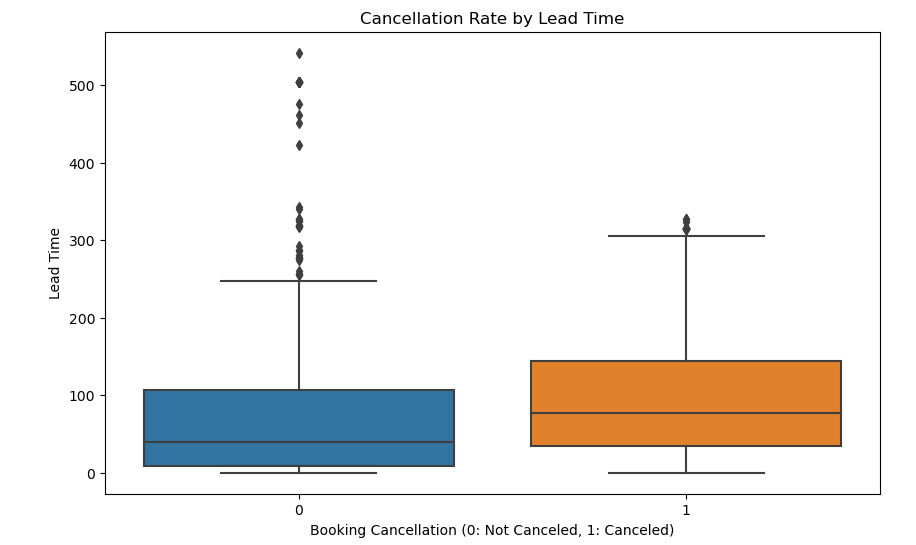
# ****Task 3 - Booking Cancellation Analysis****

## **3.1 - What factors influence booking cancellations?**

** **

** **

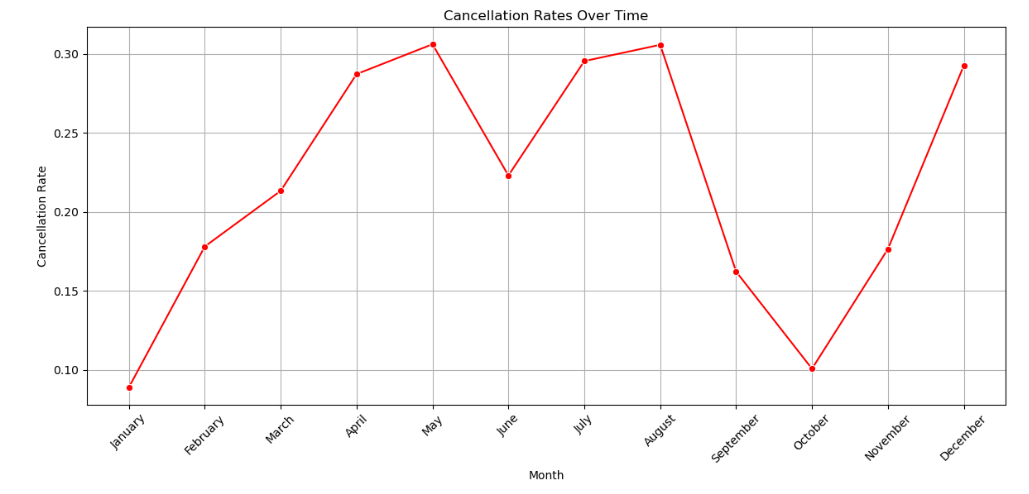
****

## **Interpretation:**

This analysis is to visualize the cancellation patterns to improve overall booking performance.

* The first plot displays the cancellation rate for different types of hotels where **City Hotel** has the highest Non Cancellation & Cancellation rate compared to **Resort Hotel**.
* The second plot displays the cancellation rate for various market segments, where **Online TA** has the highest Non Cancellation & Cancellation rate compared to other market segments.
* The third plot displays the cancellation rate based on the type of deposit made, where **No-Deposit** has the highest Non Cancellation & Cancellation rate compared to other deposit made.
* The fourth plot displays the cancellation rate by different customer types, where **Transient** has the highest Non Cancellation & Cancellation rate compared to other customer type.
* The fifth plot displays the cancellation rate for different types of reserved rooms, where **A** has the highest Non Cancellation & Cancellation rate compared to other reserved rooms.
* The sixth plot displays the cancellation rate across different months of the month, where **June** has the highest Non Cancellation & Cancellation rate compared to other months.
* The seventh plot displays the cancellation rate based on the lead time between booking and arrival, where **Non Cancelled** marked lower values compared to **Cancelled** which has the higher values.

## **3.2 - Are there specific trends or patterns in cancellations over time?**

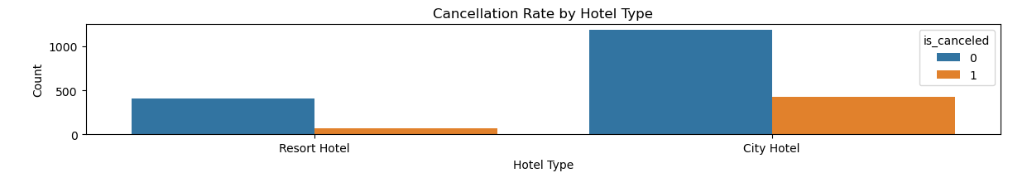
****

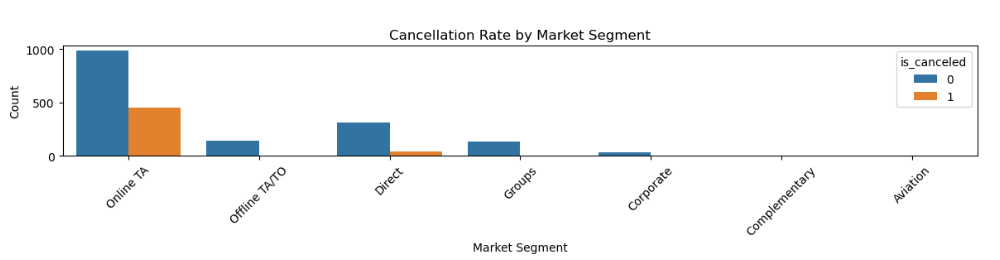
## **Interpretation:**

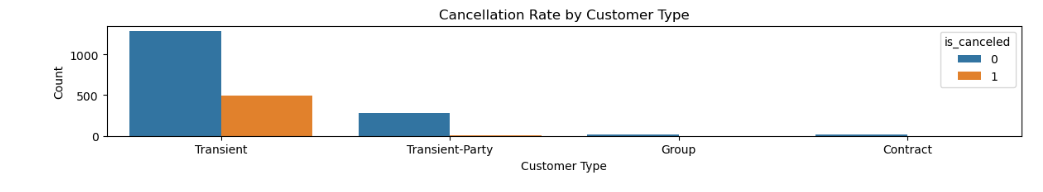
This analysis is to display the the cancellation rates over time, specifically across different months of the year.

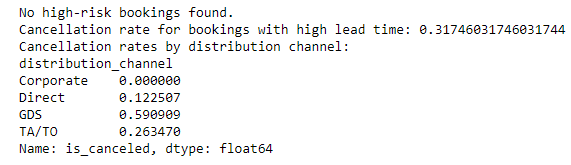
* The month - **May** has the highest trend of cancellation of booking followed by **August & July**.
* The month - **January** has the lowest trend of cancellation of booking followed by **October & September**.

## **3.3 - Can we identify high-risk bookings that are more likely to be cancelled?**









## **Interpretation:**

This analysis is to display the cancellation rates across different features in the dataset and perform additional analyses on high-risk bookings, cancellation rates based on lead time, and cancellation rates by distribution channel.

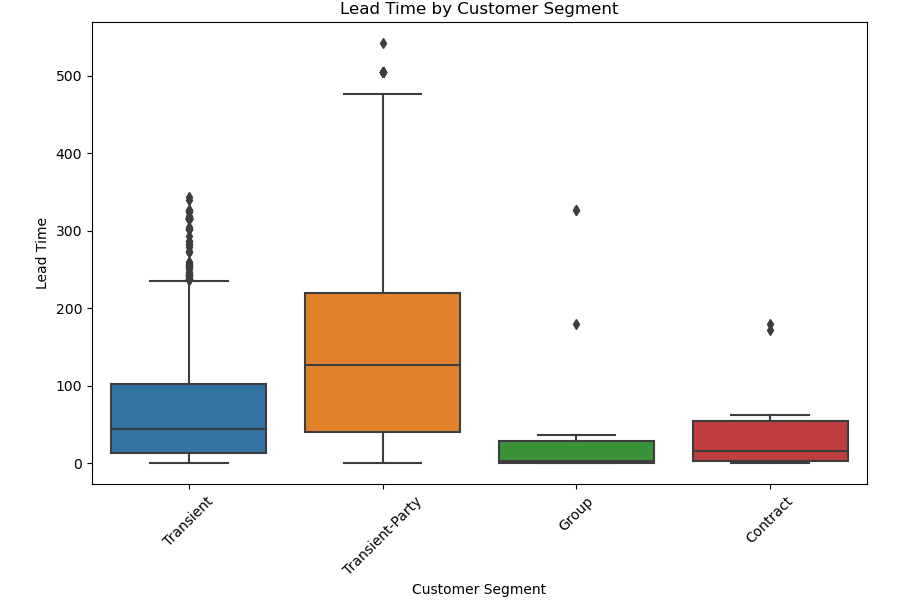
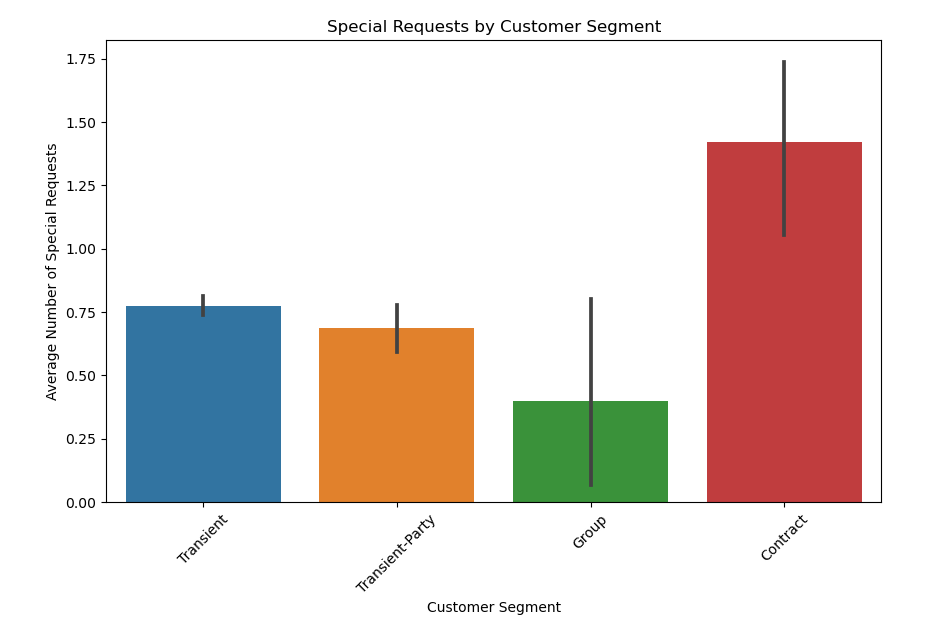
Bookings with a cancellation rate above 50% are identified as high-risk bookings.

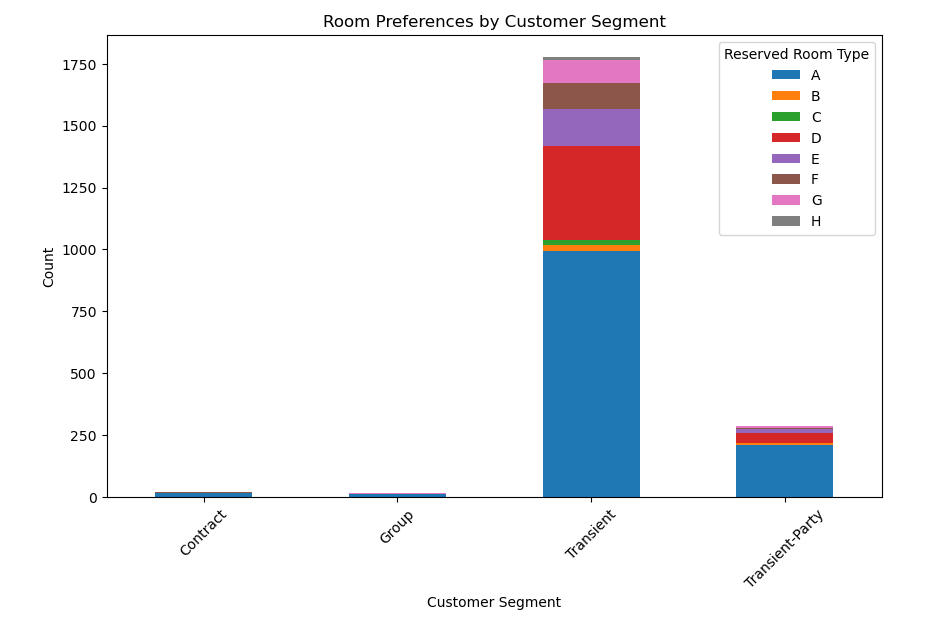
If high-risk bookings are found, they are printed and visualized using a bar plot to show the cancellation rates for each combination of hotel type and market segment.

* The first plot is to display the cancellation rate by hotel type, where **City Hotel** has the highest Non Cancellation and Cancellation rate compared to **Resort Hotel**.
* The second plot is to display the cancellation rate by Market Segment, where **Online TA** has the highest Non Cancellation and Cancellation rate compared to **Resort Hotel**.
* The third plot is to display the cancellation rate by customer type, where **Transient** has the highest Non Cancellation and Cancellation rate compared to **Resort Hotel**.
* The fourth analysis is cancellation rate by distribution channel, where **GDS** has the highest mean value compared to other distribution channel.

# ****Task 4 - Customer Behavioral Segmentation****

## **4.1 - Are there distinct patterns in the lead time, special requests, or room preferences for different customer segments?**

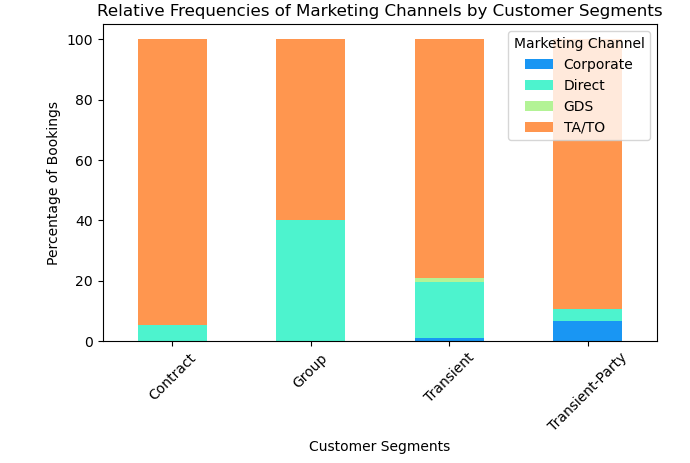


## **Interpretation:**

This analysis helps to visualize the lead time, special requests, and room preferences across different customer segments.

* The first plot displays the lead time by customer segment, where **Transient-Part** has no skewness but has the outliers at top but are not exceptionally high compared to others.
* The second plot displays the special requests by customer segment, where **Contract** highest average number of special requests.
* The third plot displays the room preference by customer segment, where **Transient** has the highest count of room preference.

## **4.2 - Which marketing channels are most effective for reaching specific customer segments?**

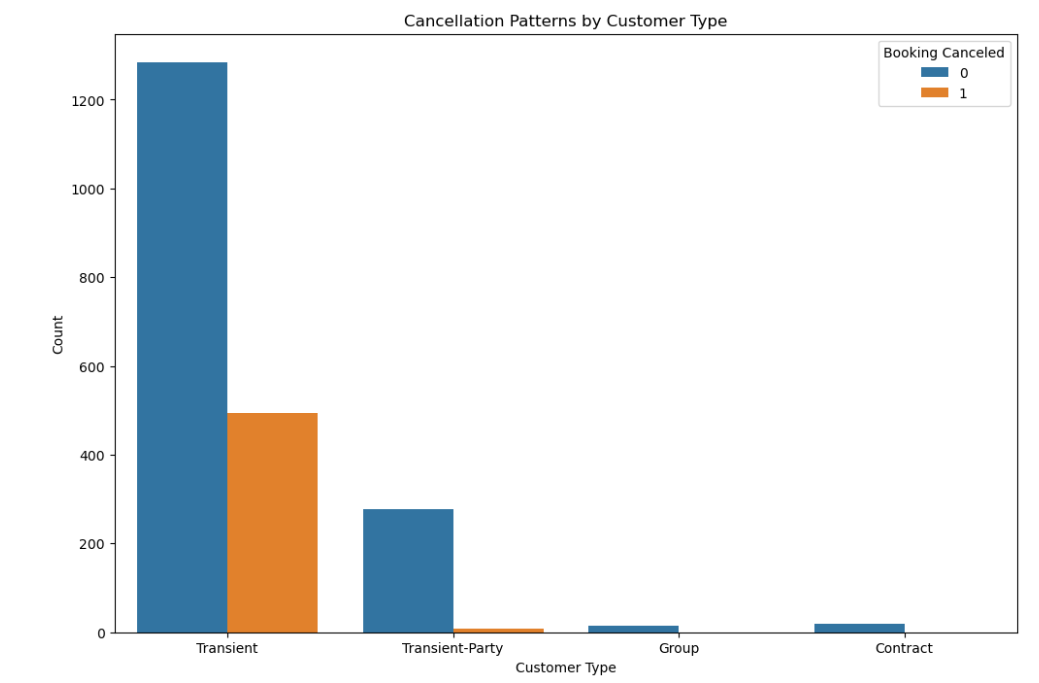
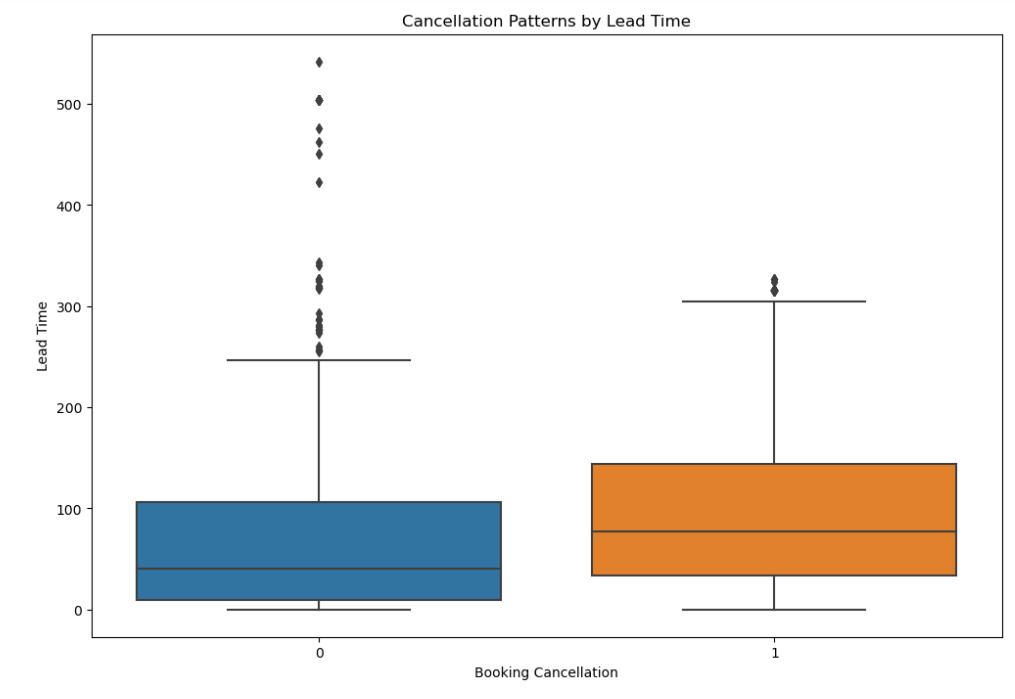
****

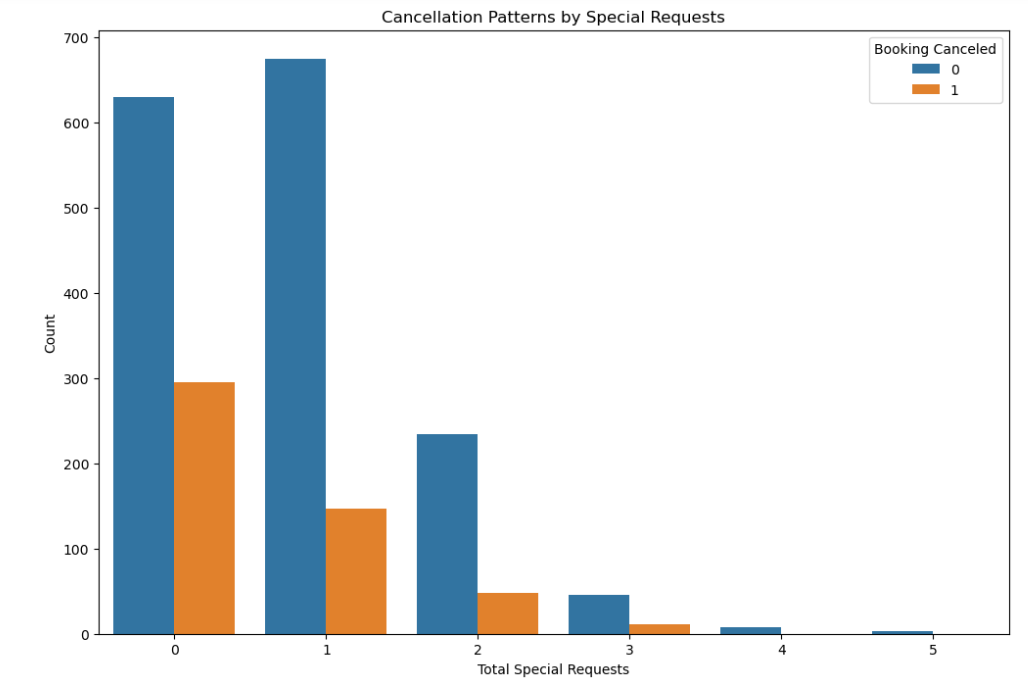
## **Interpretation:**

This analysis is to display the relative frequency of marketing channels by customer segments.

* The customer segment - **Group** has the highest percentage of marketing channel - **TA/TO & Direct**.
* The customer segment - **Contract** has the highest percentage of marketing channel - **TA/TO** and lower percentage of **Direct**.
* The customer segment - **Transient** has the highest percentage of marketing channel - **TA/TO** and lower percentage of **Direct, GDS & Corporate**.
* The customer segment - **Transient - Party** has the highest percentage of marketing channel - **TA/TO** and lower percentage of **Direct & Corporate**.

## **4.3 - How can marketing strategies be customized to resonate with specific customer segments, considering factors such as previous cancellations, booking lead time, and special requests?**

** **

****

## **Interpretation:**

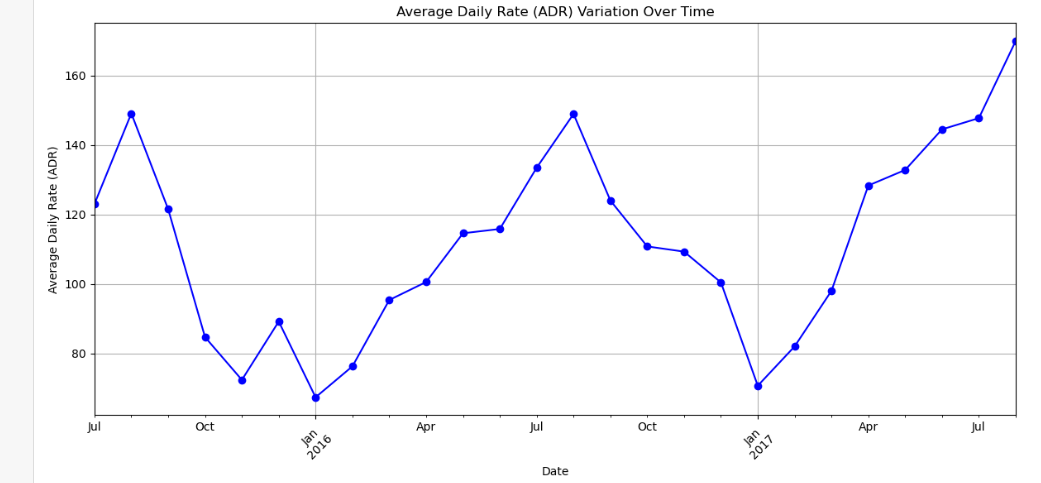
This analysis is to display the cancellation patterns across different features.

* The first plot is to display the cancellation patterns by customer type, where **Transient** has the highest Non Cancellation and Cancellation trend compared to other customer types.
* The second plot is to display the cancellation patterns by lead time, where **Non Cancellation** has the lower values compared to **Cancellation**, but the outliers varies between them, where Non cancellation has the higher values at the top and the Cancellation has the lower values at the top.
* The third plot is to display the cancellation patterns by special requests, where **Total Request - 0** has the highest Non Cancellation & Cancellation trend compared other special requests.

**PART – 2**

# ****Task 1 - Revenue Management****

## **1.1 - How does the Average Daily Rate (ADR) vary over time?**

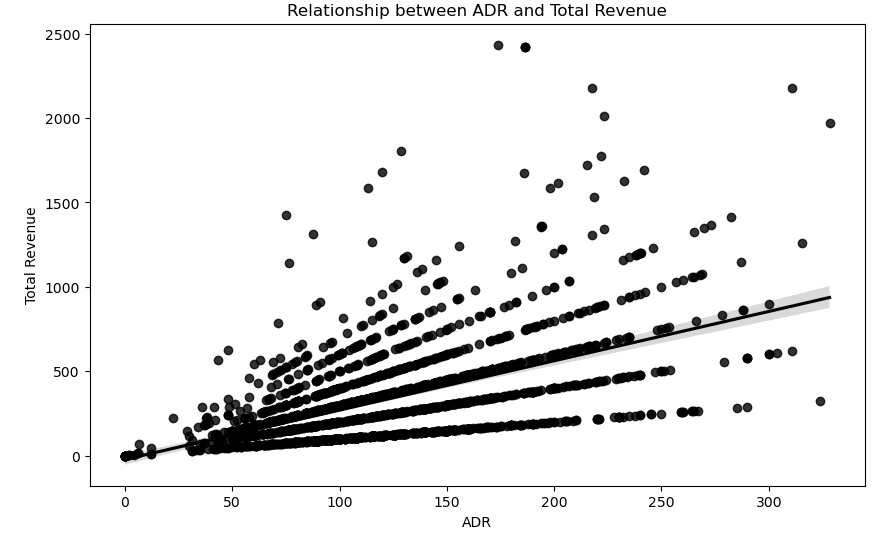


## **Interpretation:**

This analysis is to display the average daily rate (ADR) variation over time.

* The month - **July 2017** marked the highest trend/variation over time.
* The month - **Jan 2016 & 2017** marked the lowest trend/variation over time.

## **1.2 - Can we identify pricing strategies that maximize revenue?**

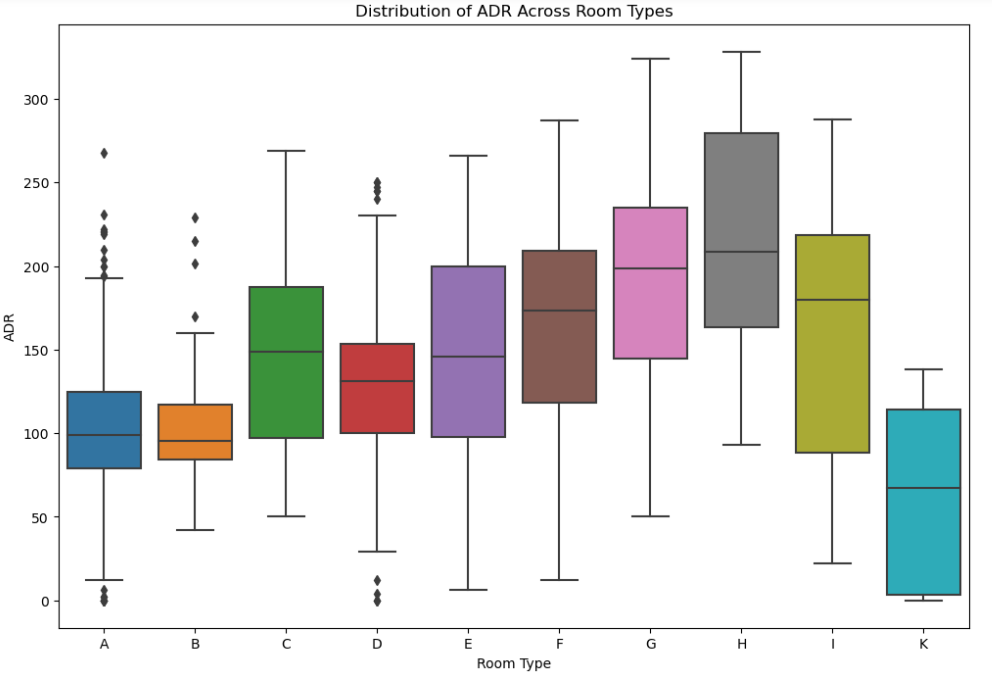
****

## **Interpretation:**

This analysis is to display the relationship between ADR and Total Revenue.

* The trend line displays the **Weak Positive Correlation** between ADR and Total Revenue generated.

## **1.3 - Are there specific room types associated with higher ADR?**

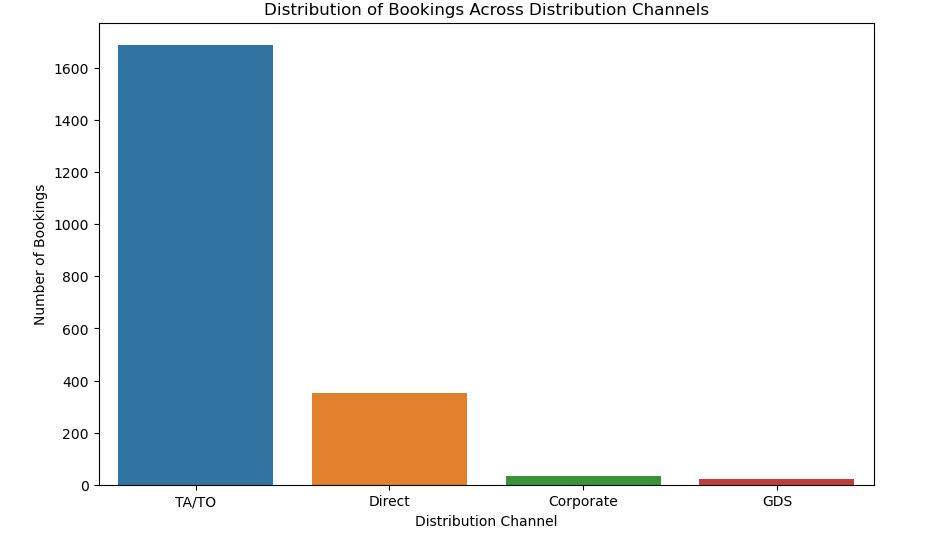
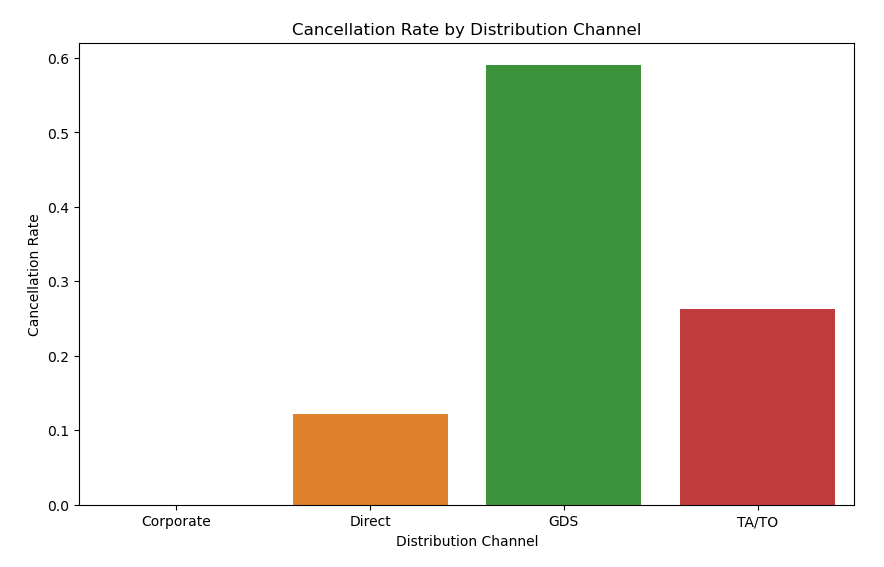
****

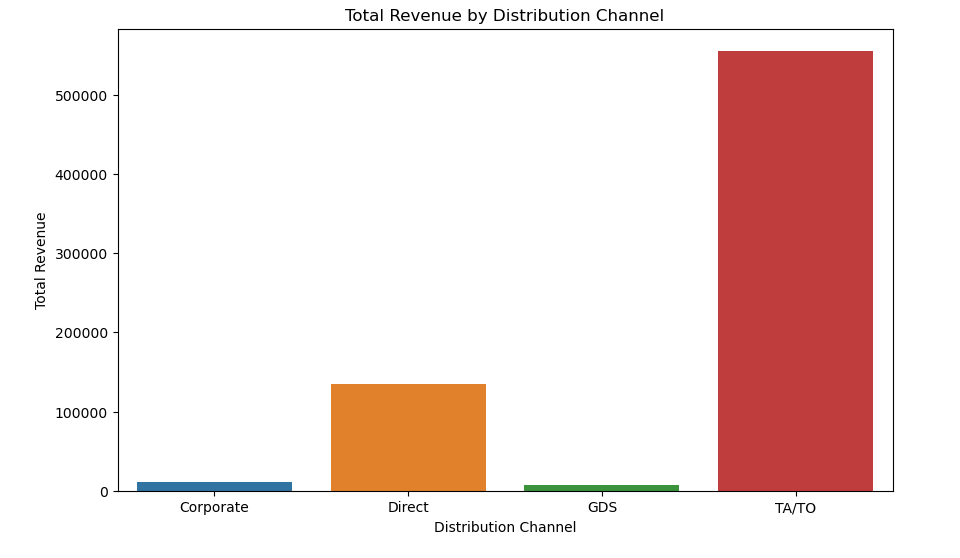
## **Interpretation:**

This analysis helps to display the distribution of ADR across Room Type, where rooms - **A & D** has the higher and lower median values and outliers which lied high and low values at top and bottom of the box plot.

# ****Task 2 - Operational Efficiency****

## **2.1 - What is the effectiveness of different booking distribution channels?**

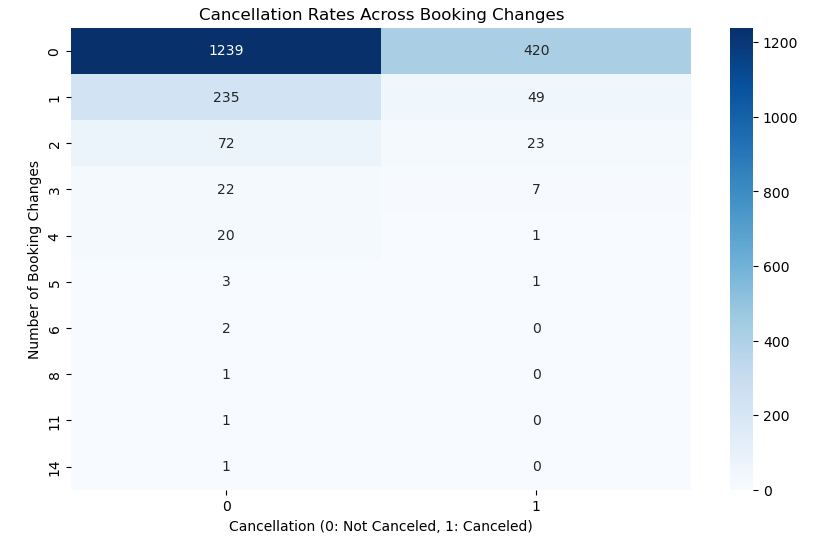
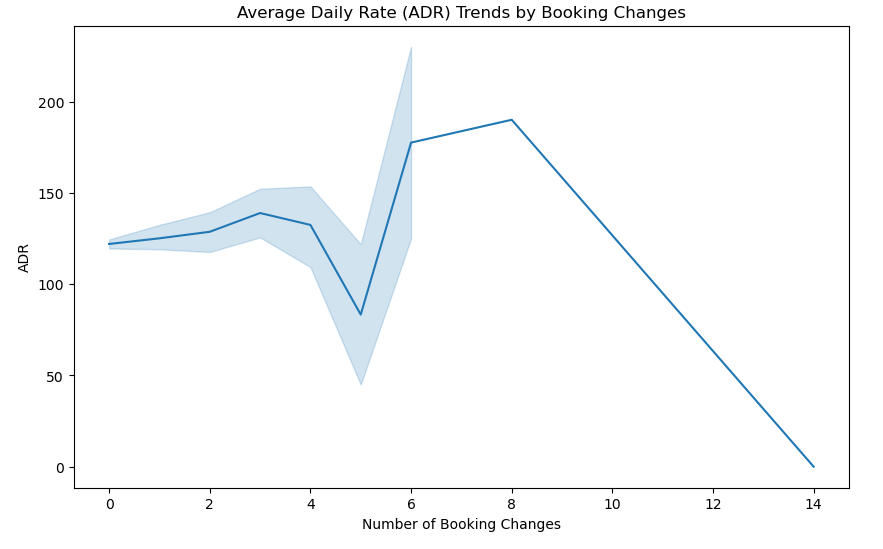


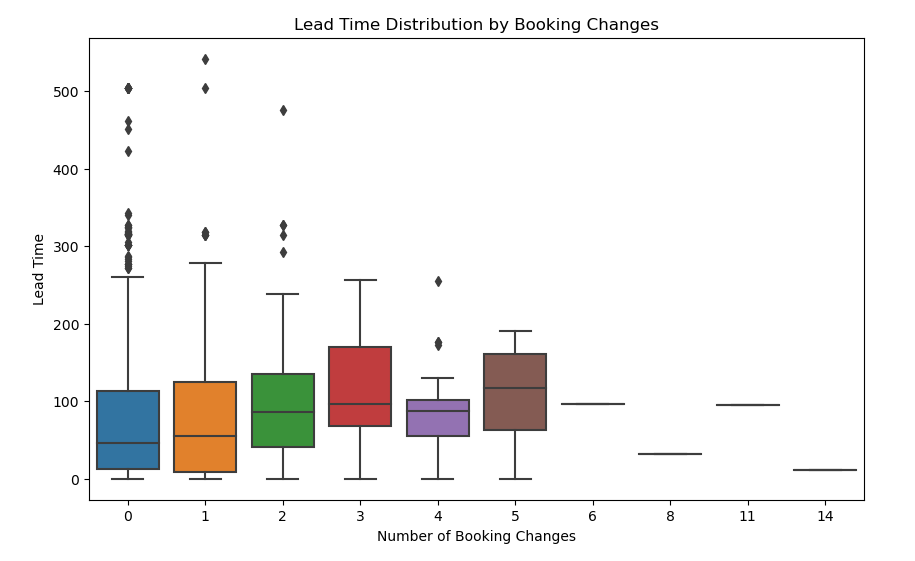
## **Interpretation:**

This analysis helps to display the effectiveness of different booking distribution channels.

* The first plot displays the distribution of booking across distribution channels, where **TA/TO** has the highest number of bookings.
* The second plot displays the cancellation rate by different distribution channels, where **GDS** has the highest cancellation rate compared to other distribution channels.
* The third plot displays the Total revenue by Distribution channels, where **TA/TO** had generated the highest revenue compared to other distribution channels.

## **2.2 - How do booking changes impact hotel operations?**

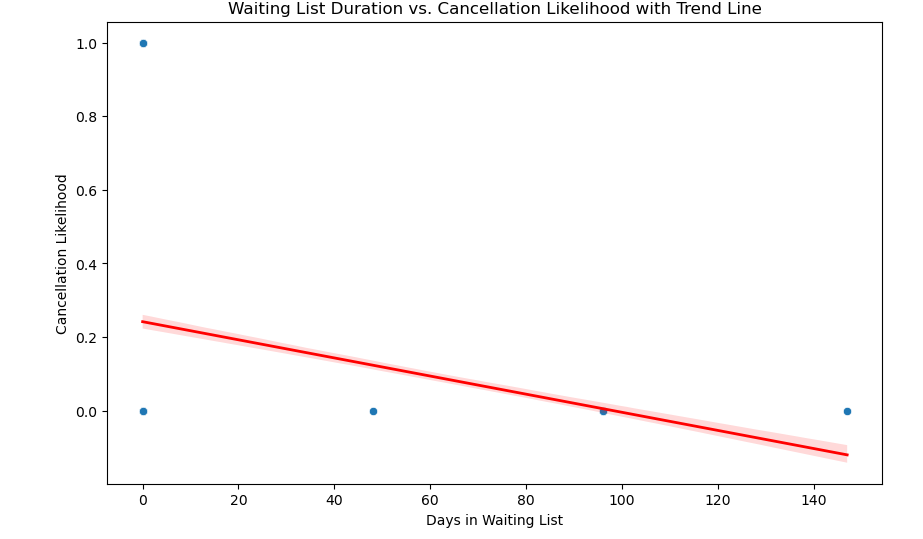


## **Interpretation:**

This analysis helps to display the booking changes impact on hotel operations.

* The first plot displays the cancellation rate across booking changes, where **Non Cancelled** rate is higher i.e., **1239** compared to **Cancelled** rate i.e., **420**.
* The second plot displays the ADR trend by booking changes, where it is higher in **8** number of booking and lower at **14** number of booking.
* The third plot displays the lead time distribution by booking changes, where **2** have no skewness and has the lower values of outliers compared to other distribution channels booking changes.

## **2.3 - Is there a correlation between the duration a booking remains on the waiting list and the likelihood of cancellation?**



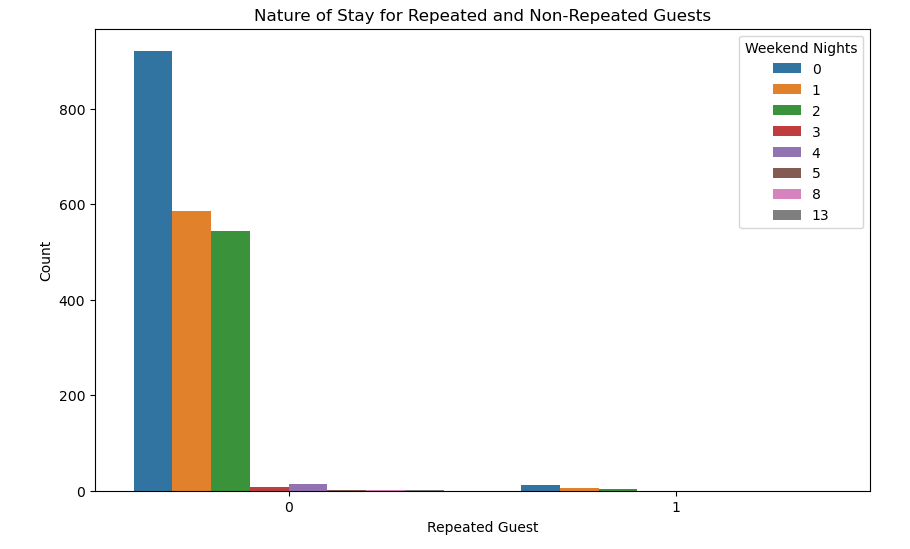
## **Interpretation:**

This analysis helps to display the correlation between the duration a booking remains on the waiting list and the likelihood of cancellation

* The trend line shows that the relation is **Weak Negative Correlation**.

# ****Task 3 - Loyalty Programs and Repeat Business****

## **3.1 - What is the behavior of repeated guests with respect to their nature of stay?**

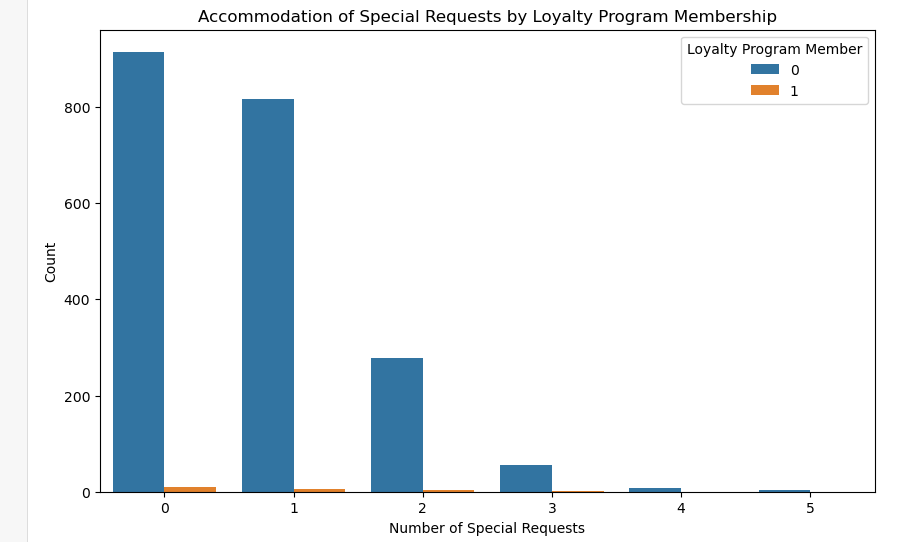


## **Interpretation:**

This analysis helps to display the behavior of repeated guests with respect to their nature of stay.

* The nature of stay for repeated and non-repeated guests is displayed by week nights, where repeated guests are higher than compared to non repeated.

## **3.2 - How does the success of loyalty programs relate to the accommodation of special requests from customers?**

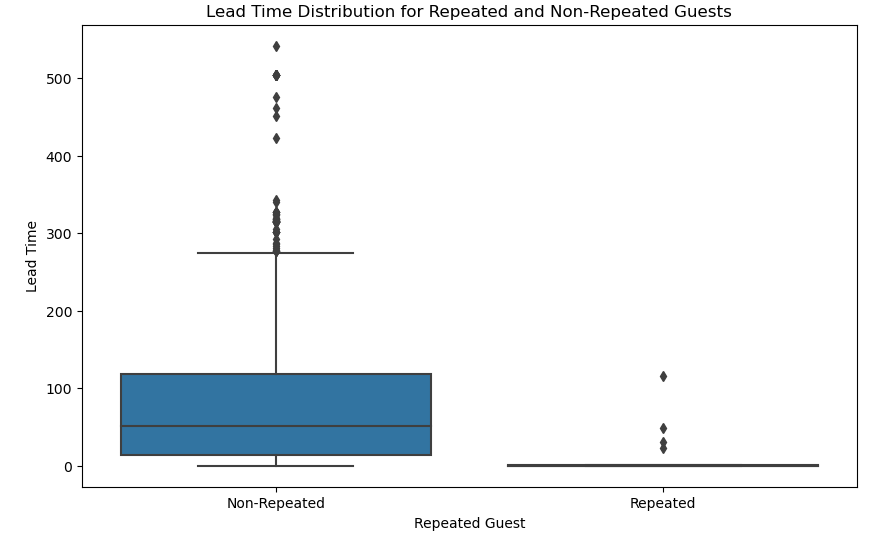


## **Interpretation:**

This analysis is to display the loyalty programs relate to the accommodation of special requests from customers.

The loyalty program member **0** has the highest number of special requests compared to loyalty program member **1**.

## **3.3 - What factors contribute to repeat business? Also, what is the distribution of lead time for repeated and non-repeated guests?**



## **Interpretation:**

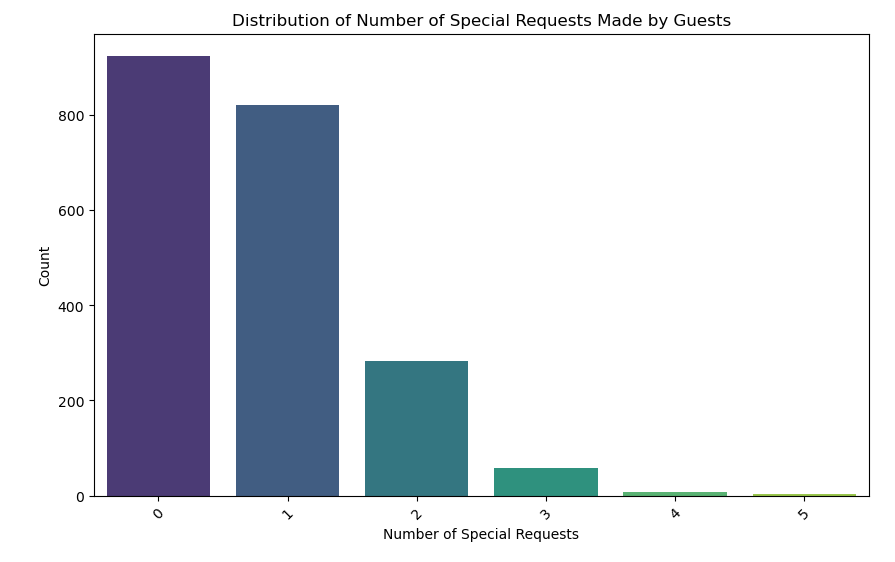
This analysis helps to understand the factors contribute to repeat business and the distribution of lead time for repeated and non-repeated guests.

* The middle line displays the median is lower than mean value in **Non Repeated** guests and have the exceptionally high values of outliers compared to **Repeated** guests.

**PART – 3**

# ****Task 1 - Customer Satisfaction****

## **1.1 - What is the distribution of the number and types of special requests made by guests?**

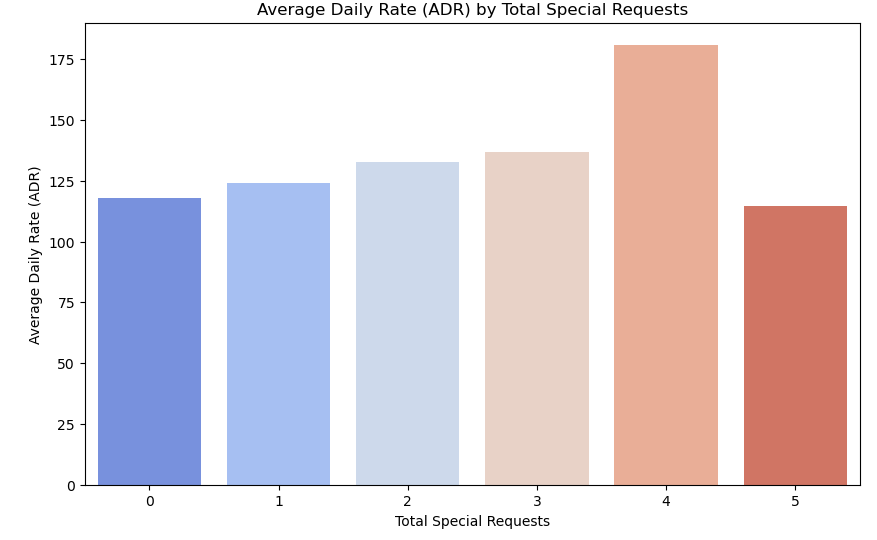
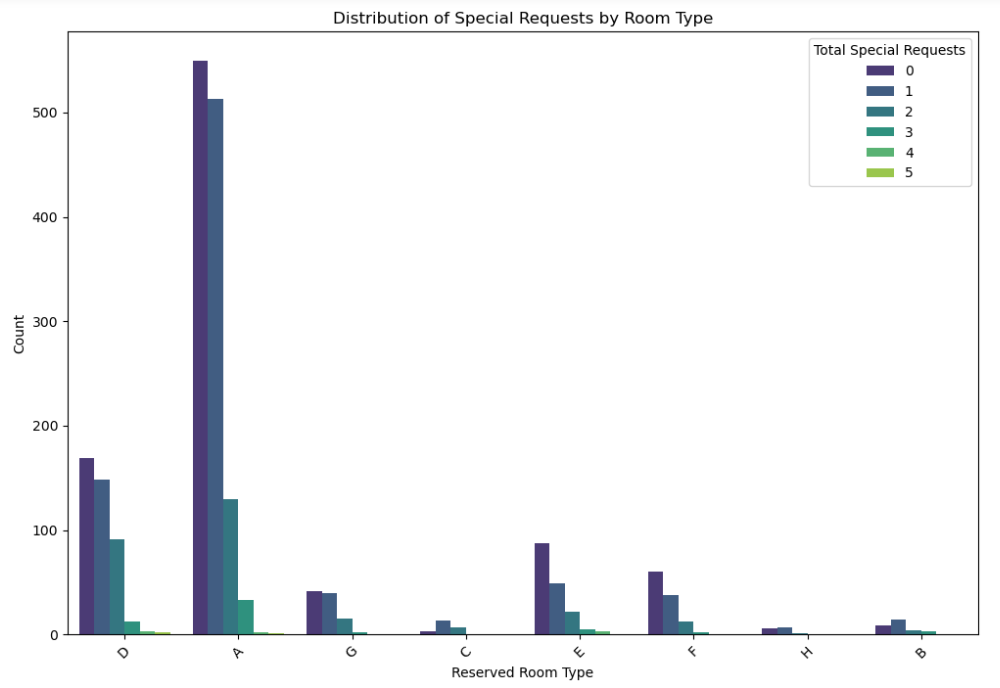
****

## **Interpretation:**

This analysis displays the distribution of the number and types of special requests made by guests.

* The distribution of number of special requests made by gusts is higher in count **0** i.e., number of special requests.
* The distribution of number of special requests made by gusts is lower in count **5** i.e., number of special requests.

## **1.2 - Is there any relation between special requests made by customers and the average daily rate? Additionally, explore customer preferences and expectations for different room types.**

** **

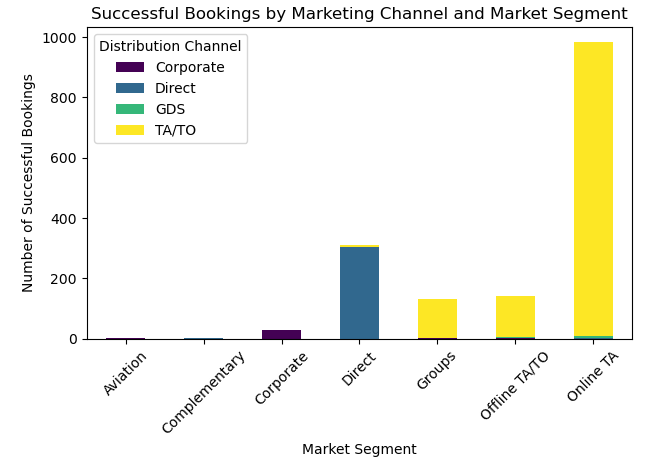
## **Interpretation:**

This analysis helps in understanding the relation between special requests made by customers and the average daily rate and exploring the customer preferences and expectations for different room types.

* The first plot displays the ADR by total special requests, where total special request **4** has the highest average daily rate (ADR).
* The second plot displays the distribution of special requests by room type, where room - **A** has the highest total special requests.

# ****Task 2 - Marketing and Sales Optimization****

## **2.1 - Which marketing channels and market segments contribute the most to successful bookings?**



## **Interpretation:**

This analysis helps to understand the marketing channels and market segments contribute the most to successful bookings.

* The market segment - **Online TA** has the highest distribution channel - **TA/TO** and lowest **Direct & GDS**.
* The market segment - **Direct** has the highest distribution channel - **Direct** and lowest **TA/TO.**

## **2.2 - Which amenities or services have the highest impact on the average daily rate (ADR)?**

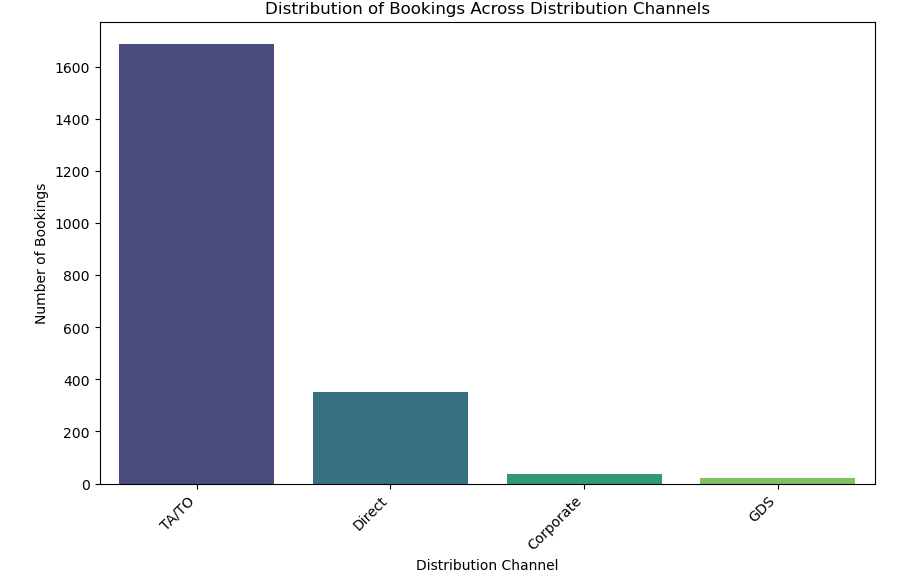
## 

## **Interpretation:**

This analysis helps in understanding the amenities or services which have the highest impact on the average daily rate (ADR).

* The room - **H** has the highest non parking space.
* The room - **B** has the highest parking space and lowest non parking space.

## **2.3 - What is the distribution of bookings across various distribution channels?**

****

## **Interpretation:**

This analysis helps in understanding the distribution of bookings across various distribution channels

* The number of bookings is high in **TA/TO** distribution channel compared to other distribution channels.

# ****Summary****

## **Part 1 -**

## **Task 1 - Data Cleaning**

* Data Preprocessing and Cleaning

## **Task 2 - Marketing and Sales Optimization**

### **2.1**

* The month of **July** has the highest booking trend followed by **June & August**.
* The month of **November** has the lowest booking trend followed by **February & December**.

### **2.2**

* The middle line in the **booking cancellation - 0 Not Cancelled** indicates that the mean is greater than median and the data values are exceptionally high values towards top.
* The middle line in the **booking cancellation - 1 Cancelled** indicates that the median is greater than mean and the outliers data values are exceptionally high values towards top and these outliers are low compared to **Not cancelled**.

### **2.3**

* In the first plot we can observe the seasonal variation in booking activity, with certain months - **July, August & June** experiencing higher booking volumes compared to others.
* The second plot is to identify any weekly patterns in booking behavior, where number - **17** has the higher trend compared to others.
* The third plot is to identify the distribution of bookings across different days of the week, where the day - **5** is tend to be higher compared to others.

## **Task 3 - Booking Cancellation Analysis**

### **3.1**

* The first plot displays the cancellation rate for different types of hotels where **City Hotel** has the highest Non Cancellation & Cancellation rate compared to **Resort Hotel**.
* The second plot displays the cancellation rate for various market segments, where **Online TA** has the highest Non Cancellation & Cancellation rate compared to other market segments.
* The third plot displays the cancellation rate based on the type of deposit made, where **No-Deposit** has the highest Non Cancellation & Cancellation rate compared to other deposit made.
* The fourth plot displays the cancellation rate by different customer types, where **Transient** has the highest Non Cancellation & Cancellation rate compared to other customer type.
* The fifth plot displays the cancellation rate for different types of reserved rooms, where **A** has the highest Non Cancellation & Cancellation rate compared to other reserved rooms.
* The sixth plot displays the cancellation rate across different months of the month, where **June** has the highest Non Cancellation & Cancellation rate compared to other months.
* The seventh plot displays the cancellation rate based on the lead time between booking and arrival, where **Non Cancelled** marked lower values compared to **Cancelled** which has the higher values.

### **3.2**

* The month - **May** has the highest trend of cancellation of booking followed by **August & July**.
* The month - **January** has the lowest trend of cancellation of booking followed by **October & September**.

### **3.3**

* The first plot is to display the cancellation rate by hotel type, where **City Hotel** has the highest Non Cancellation and Cancellation rate compared to **Resort Hotel**.
* The second plot is to display the cancellation rate by Market Segment, where **Online TA** has the highest Non Cancellation and Cancellation rate compared to **Resort Hotel**.
* The third plot is to display the cancellation rate by customer type, where **Transient** has the highest Non Cancellation and Cancellation rate compared to **Resort Hotel**.
* The fourth analysis is cancellation rate by distribution channel, where **GDS** has the highest mean value compared to other distribution channel.

## **Task 4 - Customer Behavioral Segmentation**

### **4.1**

* The first plot displays the lead time by customer segment, where **Transient-Part** has no skewness but has the outliers at top but are not exceptionally high compared to others.
* The second plot displays the special requests by customer segment, where **Contract** highest average number of special requests.
* The third plot displays the room preference by customer segment, where **Transient** has the highest count of room preference.

### **4.2**

* The customer segment - **Group** has the highest percentage of marketing channel - **TA/TO & Direct**.
* The customer segment - **Contract** has the highest percentage of marketing channel - **TA/TO** and lower percentage of **Direct**.
* The customer segment - **Transient** has the highest percentage of marketing channel - **TA/TO** and lower percentage of **Direct, GDS & Corporate**.
* The customer segment - **Transient - Party** has the highest percentage of marketing channel - **TA/TO** and lower percentage of **Direct & Corporate**.

### **4.3**

* The first plot is to display the cancellation patterns by customer type, where **Transient** has the highest Non Cancellation and Cancellation trend compared to other customer types.
* The second plot is to display the cancellation patterns by lead time, where **Non Cancellation** has the lower values compared to **Cancellation**, but the outliers varies between them, where Non cancellation has the higher values at the top and the Cancellation has the lower values at the top.
* The third plot is to display the cancellation patterns by special requests, where **Total Request - 0** has the highest Non Cancellation & Cancellation trend compared other special requests.

## **Part 2 -**

## **Task 1 - Revenue Management**

### **1.1**

* The month - **July 2017** marked the highest trend/variation over time.
* The month - **Jan 2016 & 2017** marked the lowest trend/variation over time.

### **1.2**

* The trend line displays the **Weak Positive Correlation** between ADR and Total Revenue generated.

### **1.3**

This analysis helps to display the distribution of ADR across Room Type, where rooms - **A & D** has the higher and lower median values and outliers which lied high and low values at top and bottom of the box plot.

## **Task 2 - Operational Efficiency**

### **2.1**

* The first plot displays the distribution of booking across distribution channels, where **TA/TO** has the highest number of bookings.
* The second plot displays the cancellation rate by different distribution channels, where **GDS** has the highest cancellation rate compared to other distribution channels.
* The third plot displays the Total revenue by Distribution channels, where **TA/TO** had generated the highest revenue compared to other distribution channels.

### **2.2**

* The first plot displays the cancellation rate across booking changes, where **Non Cancelled** rate is higher i.e., **1239** compared to **Cancelled** rate i.e., **420**.
* The second plot displays the ADR trend by booking changes, where it is higher in **8** number of booking and lower at **14** number of booking.
* The third plot displays the lead time distribution by booking changes, where **2** have no skewness and has the lower values of outliers compared to other distribution channels booking changes.

### **2.3**

* The trend line shows that the relation is **Weak Negative Correlation**.

## **Task 3 - Loyalty Programs and Repeat Business**

### **3.1**

* The nature of stay for repeated and non-repeated guests is displayed by week nights, where repeated guests are higher than compared to non repeated.

### **3.2**

The loyalty program member **0** has the highest number of special requests compared to loyalty program member **1**.

### **3.3**

* The middle line displays the median is lower than mean value in **Non Repeated** guests and have the exceptionally high values of outliers compared to **Repeated** guests.

## **Part 3 -**

## **Task 1 - Customer Satisfaction**

### **1.1**

* The distribution of number of special requests made by gusts is higher in count **0** i.e., number of special requests.
* The distribution of number of special requests made by gusts is lower in count **5** i.e., number of special requests.

### **1.2**

* The first plot displays the ADR by total special requests, where total special request **4** has the highest average daily rate (ADR).
* The second plot displays the distribution of special requests by room type, where room - **A** has the highest total special requests.

## **Task 2 - Marketing and Sales Optimization**

### **2.1**

* The market segment - **Online TA** has the highest distribution channel - **TA/TO** and lowest **Direct & GDS**.
* The market segment - **Direct** has the highest distribution channel - **Direct** and lowest **TA/TO**

### **2.2**

* The room - **H** has the highest non parking space.
* The room - **B** has the highest parking space and lowest non parking space.

### **2.3**

* The number of bookings is high in **TA/TO** distribution channel compared to other distribution channels.

**Conclusion**

* July consistently has the highest booking trend, while November experiences the lowest.
* Cancellation rates vary across different factors, with transient customer segments showing the highest cancellation rates.
* May has the highest trend of booking cancellations, while January has the lowest.
* Transient customer segments tend to cancel bookings more frequently.
* Contract segments request the most special services, while transient segments prefer specific room types.
* TA/TO is the primary distribution channel across different customer segments.
* July 2017 marks the highest variation in booking trends, while January 2016 and 2017 have the lowest.
* There's a weak positive correlation between average daily rate (ADR) and total revenue.
* GDS distribution channel has high cancellation rates but contributes less to total revenue.
* Different booking changes lead to fluctuations in ADR.
* Non-repeated guests tend to have higher variability in special requests compared to repeated guests.
* Room type 'A' receives the highest total special requests.
* Online TA is the primary distribution channel across different market segments.
* Room 'H' uses non-parking space most, while room 'B' uses parking space most.

**Thank you**