# EDA File of Hotel Booking Project


## 1) Finding the dimensions of the dataset

The dataset used consists of 36,285 rows and 17 columns, containing information related to hotel bookings. This data includes both numerical and categorical data types.

```
[9] df.shape
(36285, 17)
```

## 2) Reviewing the first few rows of the dataset

| | Booking_ID | number of adults | number of children | number of weekend nights | number of week nights | type of meal | car parking space | room type | lead time | market segment type | repeated | P-C | P-not-C | average price | special requests | date of reservation | booking status |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | INN00001 | 1 | 1 | 2 | 5 | Meal Plan 1 | 0 | Room_Type 1 | 224 | Offline | 0 | 0 | 0 | 88.00 | 0 | 10/2/2015 | Not_Canceled |
| 1 | INN00002 | 1 | 0 | 1 | 3 | Not Selected | 0 | Room_Type 1 | 5 | Online | 0 | 0 | 0 | 106.68 | 1 | 11/6/2018 | Not_Canceled |
| 2 | INN00003 | 2 | 1 | 1 | 3 | Meal Plan 1 | 0 | Room_Type 1 | 1 | Online | 0 | 0 | 0 | 50.00 | 0 | 2/28/2018 | Canceled |
| 3 | INN00004 | 1 | 0 | 0 | 2 | Meal Plan 1 | 0 | Room_Type 1 | 211 | Online | 0 | 0 | 0 | 100.00 | 1 | 5/20/2017 | Canceled |
| 4 | INN00005 | 1 | 0 | 1 | 2 | Not Selected | 0 | Room_Type 1 | 48 | Online | 0 | 0 | 0 | 77.00 | 0 | 4/11/2018 | Canceled |

To get a better understanding of the dataset structure, we examined the initial rows. This preliminary review helps us gain a general overview of the types and formats of the data in different columns.

### 3) **Arrival Date**

In the dataset, there is a column titled *date of reservation* that indicates the time when the hotel was booked by customers. Using the `datetime` library and the *lead time* (which shows the time gap between booking and the stay date), we calculated the exact start date of the guests' stay. This information is important as it can influence our further analyses, such as investigating seasonal patterns or changes in cancellation rates.

| | Booking_ID | number of adults | number of children | number of weekend nights | number of week nights | type of meal | car parking space | room type | lead time | market segment type | repeated | P-C | P-not-C | average price | special requests | date of reservation | booking status | arrival_date |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | INN00001 | 1 | 1 | 2 | 5 | Meal Plan 1 | 0 | Room_Type 1 | 224 | Offline | 0 | 0 | 0 | 88.00 | 0 | 2015-10-02 | Not_Canceled | 2016-05-13 |
| 1 | INN00002 | 1 | 0 | 1 | 3 | Not Selected | 0 | Room_Type 1 | 5 | Online | 0 | 0 | 0 | 106.68 | 1 | 2018-11-06 | Not_Canceled | 2018-11-11 |
| 2 | INN00003 | 2 | 1 | 1 | 3 | Meal Plan 1 | 0 | Room_Type 1 | 1 | Online | 0 | 0 | 0 | 50.00 | 0 | 2018-02-28 | Canceled | 2018-03-01 |
| 3 | INN00004 | 1 | 0 | 0 | 2 | Meal Plan 1 | 0 | Room_Type 1 | 211 | Online | 0 | 0 | 0 | 100.00 | 1 | 2017-05-20 | Canceled | 2017-12-17 |
| 4 | INN00005 | 1 | 0 | 1 | 2 | Not Selected | 0 | Room_Type 1 | 48 | Online | 0 | 0 | 0 | 77.00 | 0 | 2018-04-11 | Canceled | 2018-05-29 |

### **4) Reviewing feature names**

We reviewed all the columns to ensure that the data is properly defined.

```
df.columns
```

```
Index(['Booking_ID', 'number of adults', 'number of children',
       'number of weekend nights', 'number of week nights', 'type of meal',
       'car parking space', 'room type', 'lead time', 'market segment type',
       'repeated', 'P-C', 'P-not-C', 'average price', 'special requests',
       'date of reservation', 'booking status', 'arrival_date'],
      dtype='object')
```

### **5) Reviewing the data (Features)**

Upon examining the dataset, it was found that most columns contain valid data and appropriate data types. The only exception was the date-related columns, which had 37 null values. At this stage, we decided to remove these rows. These missing data points could negatively affect the results of our analysis, so removing them helped clean the dataset.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36285 entries, 0 to 36284
Data columns (total 18 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Booking_ID                36285 non-null  object
 1   number of adults          36285 non-null  int64
 2   number of children        36285 non-null  int64
 3   number of weekend nights  36285 non-null  int64
 4   number of week nights     36285 non-null  int64
 5   type of meal              36285 non-null  object
 6   car parking space         36285 non-null  int64
 7   room type                 36285 non-null  object
 8   lead time                 36285 non-null  int64
 9   market segment type       36285 non-null  object
 10  repeated                  36285 non-null  int64
 11  P-C                       36285 non-null  int64
 12  P-not-C                   36285 non-null  int64
 13  average price             36285 non-null  float64
 14  special requests          36285 non-null  int64
 15  date of reservation       36248 non-null  datetime64[ns]
 16  booking status            36285 non-null  object
 17  arrival_date              36248 non-null  datetime64[ns]
dtypes: datetime64[ns](2), float64(1), int64(10), object(5)
memory usage: 5.0+ MB
```

```
df.isna().sum()
```

| | 0 |
|---|---|
| Booking_ID | 0 |
| number of adults | 0 |
| number of children | 0 |
| number of weekend nights | 0 |
| number of week nights | 0 |
| type of meal | 0 |
| car parking space | 0 |
| room type | 0 |
| lead time | 0 |
| market segment type | 0 |
| repeated | 0 |
| P-C | 0 |
| P-not-C | 0 |
| average price | 0 |
| special requests | 0 |
| date of reservation | 37 |
| booking status | 0 |
| arrival_date | 37 |

**dtype:** int64

6) **Reviewing the statistical distribution of the data**

For a more detailed analysis, we used the *`describe`* function to obtain a summary of the statistical information of the data. This summary includes parameters such as mean, maximum, minimum, and median values, which provide a better understanding of the data distribution and characteristics.

```
df.describe()
```

| | number of adults | number of children | number of weekend nights | number of week nights | car parking space | lead time | repeated | P-C | P-not-C | average price | special requests |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 36285.000000 | 36285.000000 | 36285.000000 | 36285.000000 | 36285.000000 | 36285.000000 | 36285.000000 | 36285.000000 | 36285.000000 | 36285.000000 | 36285.000000 |
| mean | 1.844839 | 0.105360 | 0.810693 | 2.204602 | 0.030977 | 85.239851 | 0.025630 | 0.023343 | 0.153369 | 103.421636 | 0.619733 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 2.000000 | 0.000000 | 0.000000 | 1.000000 | 0.000000 | 17.000000 | 0.000000 | 0.000000 | 0.000000 | 80.300000 | 0.000000 |
| 50% | 2.000000 | 0.000000 | 1.000000 | 2.000000 | 0.000000 | 57.000000 | 0.000000 | 0.000000 | 0.000000 | 99.450000 | 0.000000 |
| 75% | 2.000000 | 0.000000 | 2.000000 | 3.000000 | 0.000000 | 126.000000 | 0.000000 | 0.000000 | 0.000000 | 120.000000 | 1.000000 |
| max | 4.000000 | 10.000000 | 7.000000 | 17.000000 | 1.000000 | 443.000000 | 1.000000 | 13.000000 | 58.000000 | 540.000000 | 5.000000 |
| std | 0.518813 | 0.402704 | 0.870590 | 1.410946 | 0.173258 | 85.938796 | 0.158032 | 0.368281 | 1.753931 | 35.086469 | 0.786262 |

### 7) Checking for duplicate data

One of the key steps in data preprocessing is identifying duplicate data. Fortunately, there were no duplicate entries in this dataset, ensuring that our analysis will be more accurate and reliable.

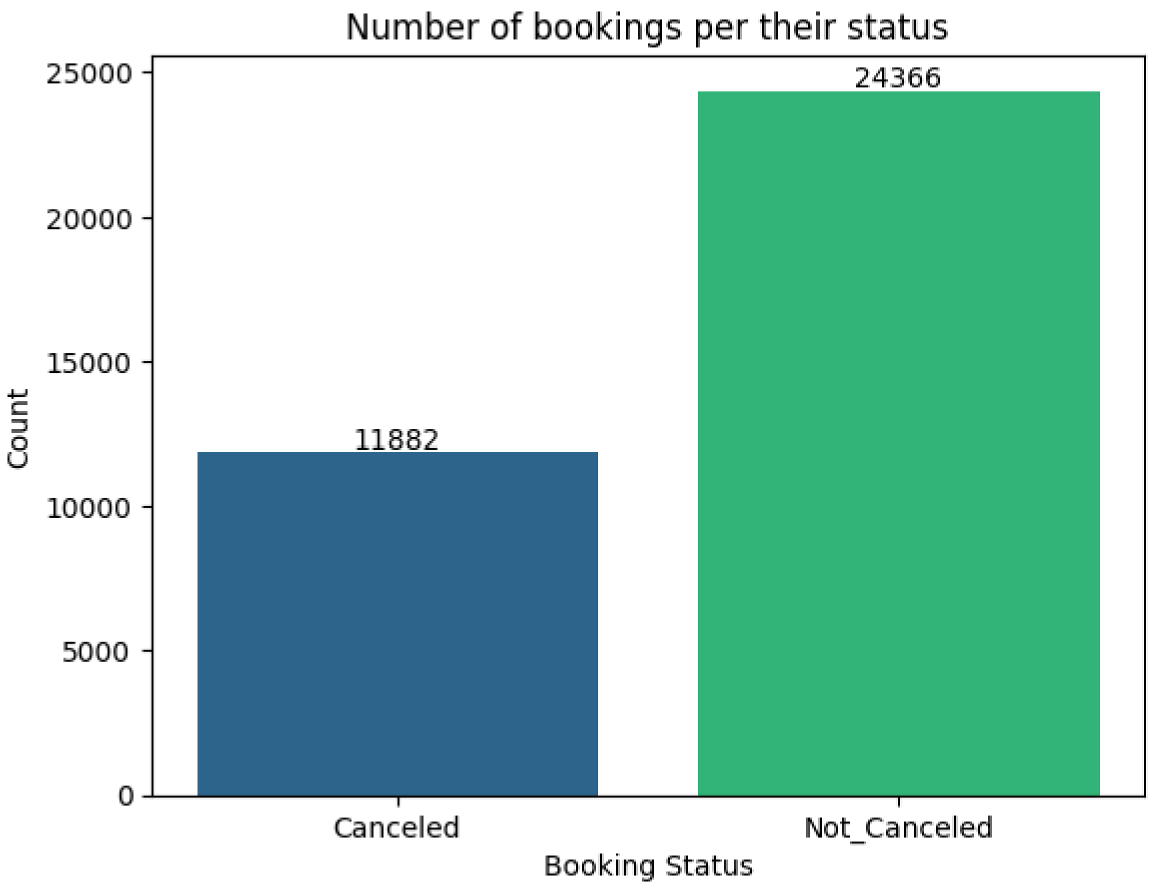
```
[15] df.duplicated().sum()   # Checking for the duplicated rows

0
```

### 8) Main Objective of the Project

The main goal of this project is to thoroughly analyse the different features of the bookings and determine which ones have the greatest impact on whether bookings are cancelled or not. To better understand this, we will first examine the number of *Cancelled* and *Not-Cancelled* bookings.

### 9) Comparison of cancelled and not-cancelled bookings



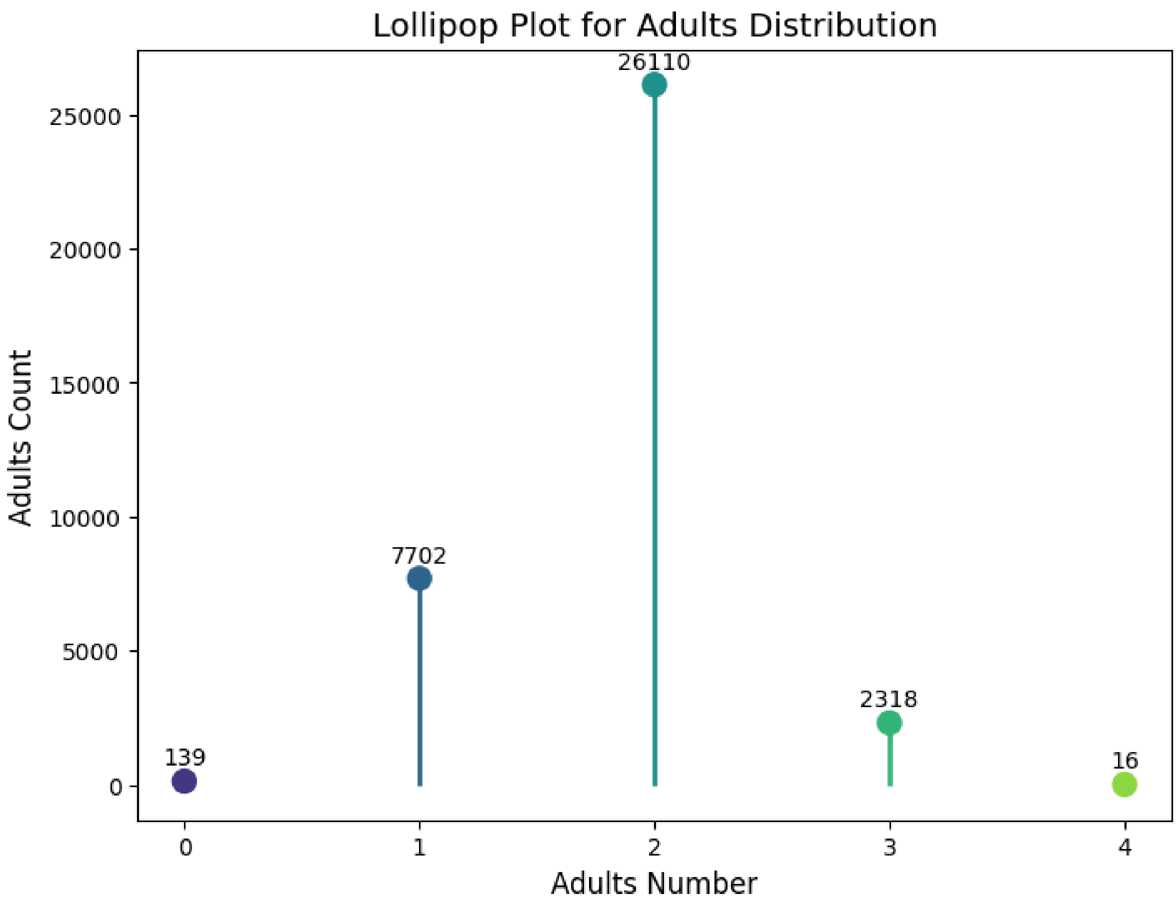

Based on the chart related to booking status, it was found that the number of *Not-Cancelled* bookings is more than double the number of *Cancelled* bookings. This information helps us better understand the existing patterns and see which factors influence customers' decisions to cancel or not cancel their bookings.

**10) Categorization and Analysis of Features**

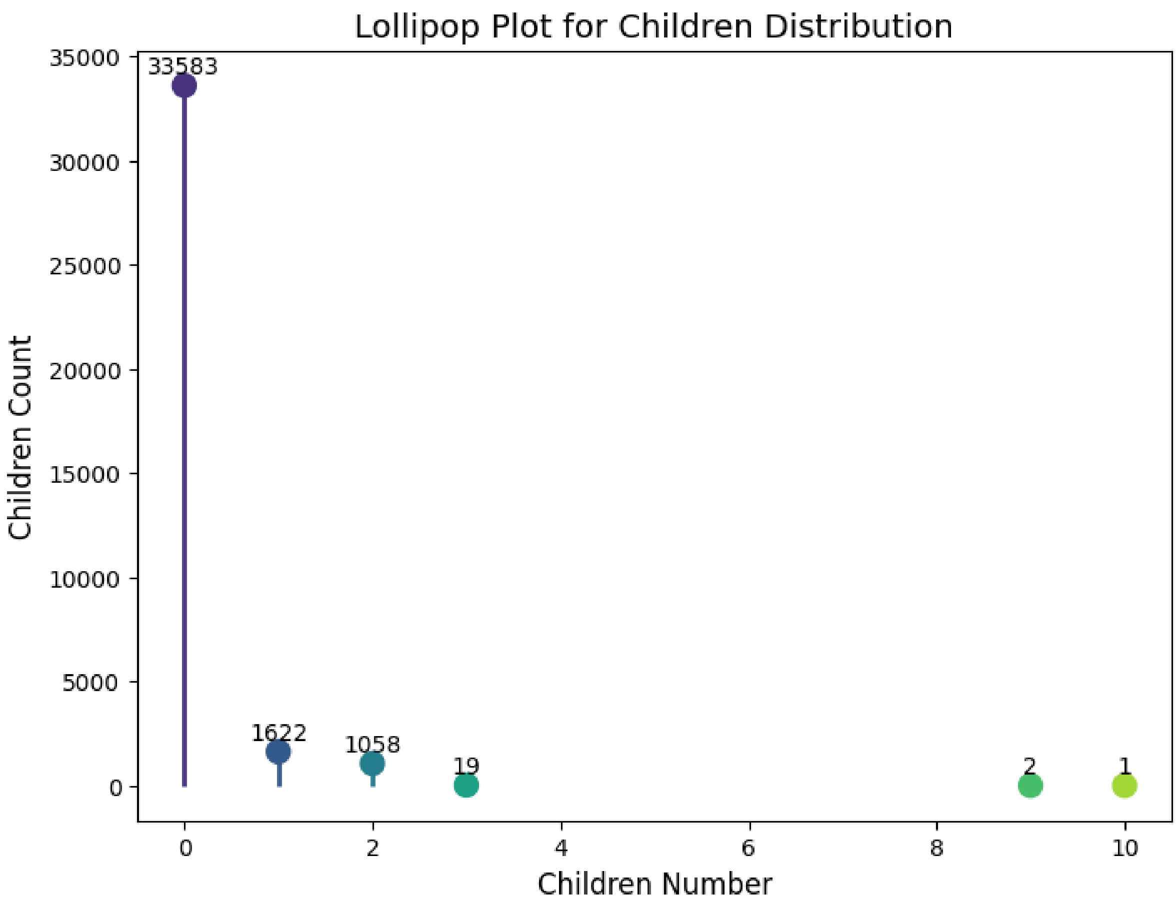
At this stage, we will categorise the various features into different groups and analyse each one in more detail. This categorization allows us to better understand the impact of each feature on whether bookings are cancelled or not.

## 1. Reviewing the number of individuals in bookings:

1) **Number of adults per booking:** The chart related to the number of adults in bookings shows that most bookings include 2 adults. A booking with 0 adults means that all members of the group are children.

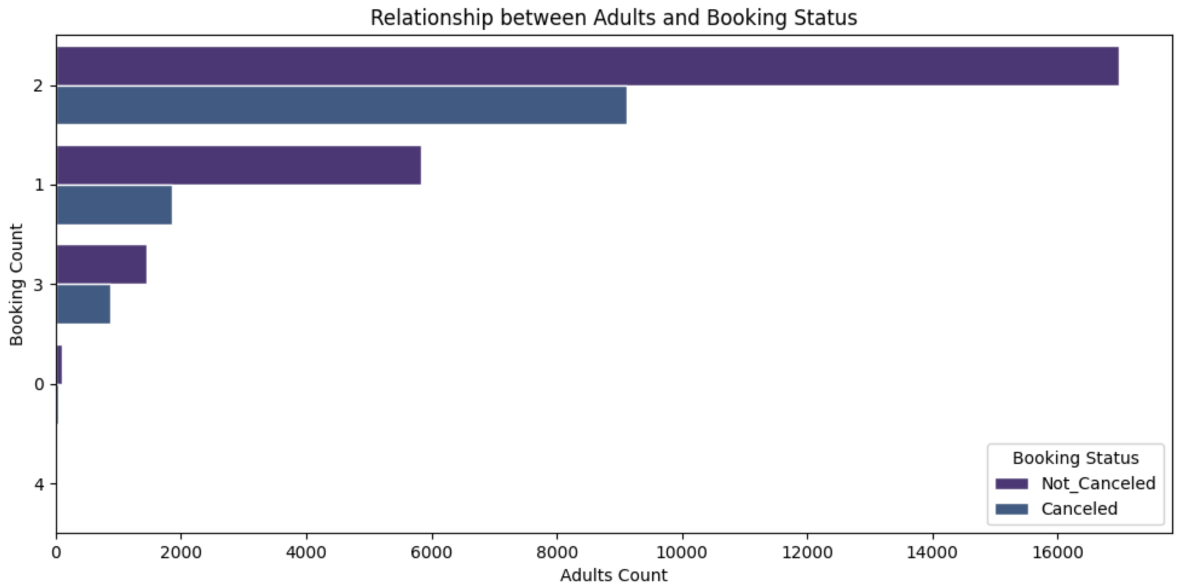


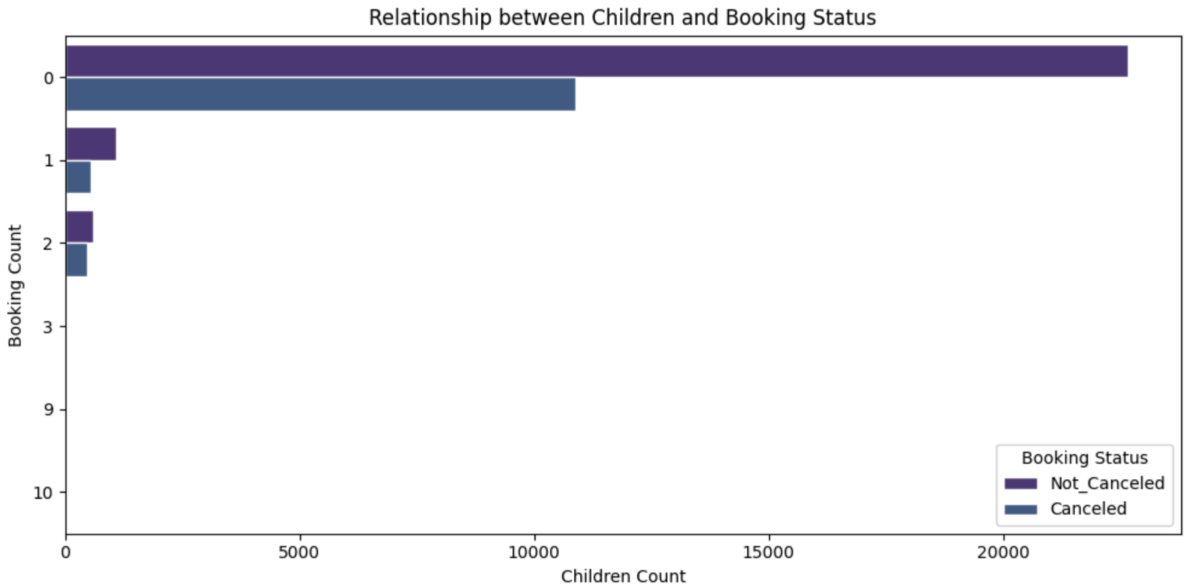

2) **Number of children per booking:** The chart on the number of children shows that most bookings do not include any children. This information is important for analysing family travel patterns and the impact of children on their decision to cancel a booking.




3) **Reviewing the relationship between the number of adults and children with booking cancellations:**

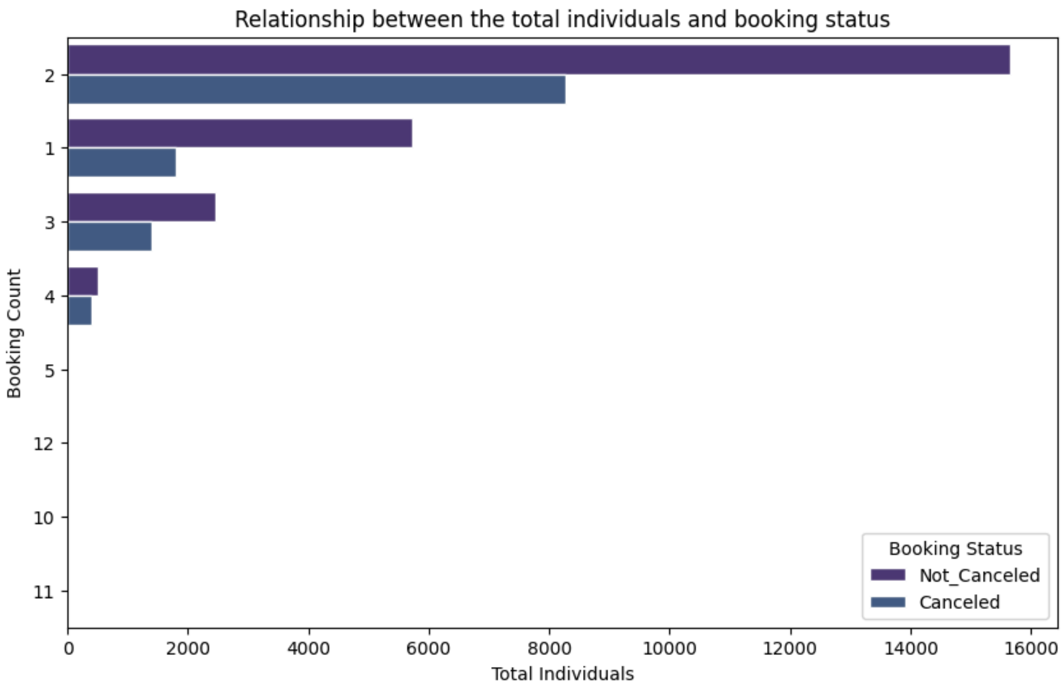
- **Number of adults:** Most bookings are made for 1 or 2 adults, and this group also has the highest number of cancellations.
- **Families without children:** Families without children not only make more bookings, but they also have a higher likelihood of cancelling their bookings.




4) **Reviewing the relationship between the total number of individuals and booking cancellations:**

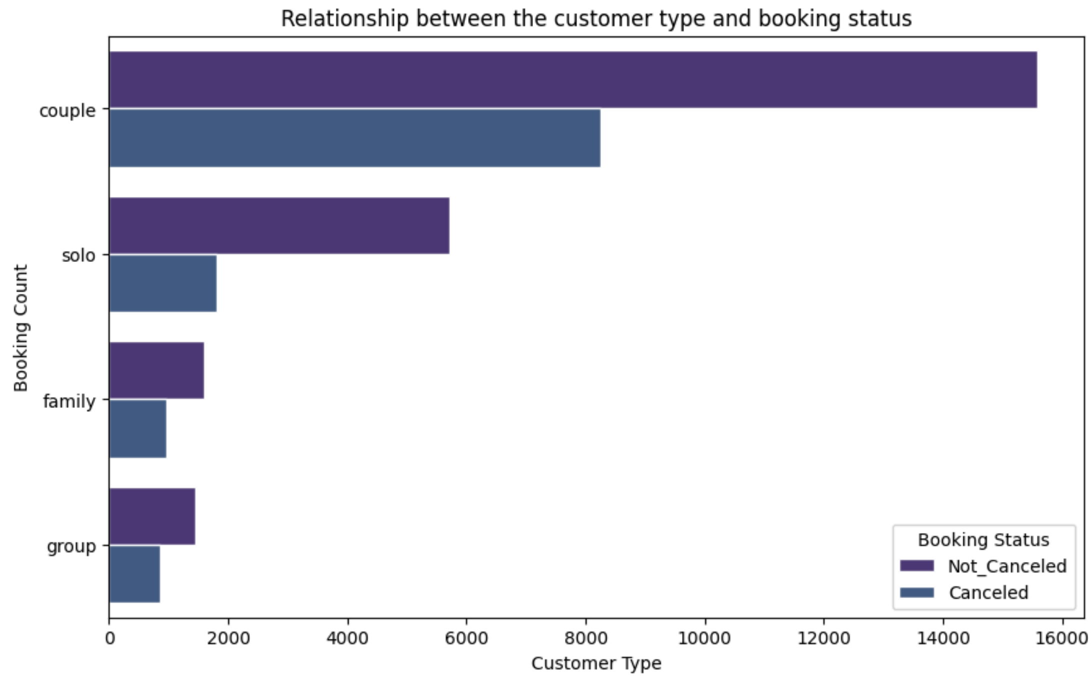
Although two-person bookings are more common, more than half of these bookings are cancelled. In contrast, solo travellers (those travelling alone) are fewer in number, but only one-third of them cancel their bookings. This analysis can help hotels better understand cancellation patterns and create special policies for solo travellers.




5) **Reviewing the customer type and booking cancellations:**

For a better analysis, we categorised the travellers into four groups:

| | |
|---|---|
| Couple | Two adult guests |
| Solo | Solo guests |
| Family | Guests with at least one child |
| Group | Groups with more than two adults |

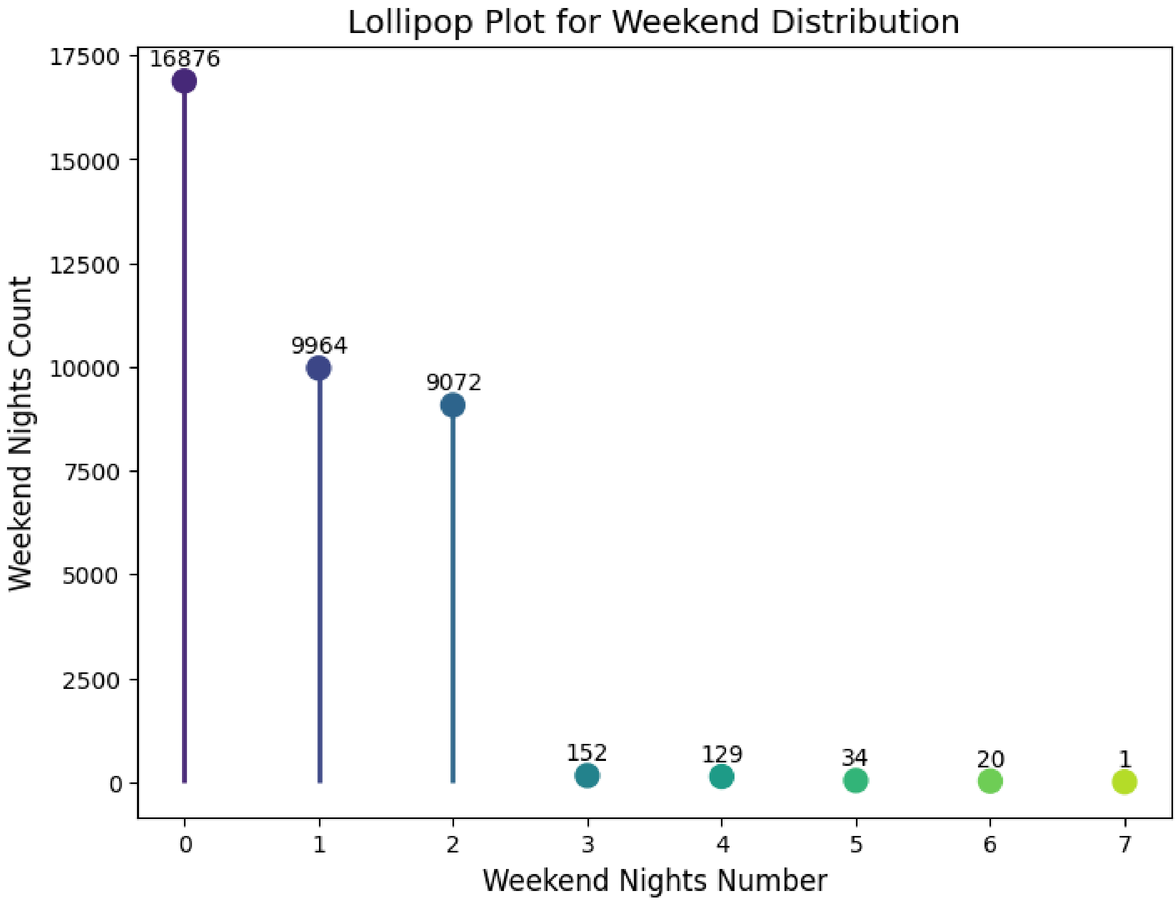



The results show that although couples form the majority of hotel guests, they have a higher likelihood of cancelling their bookings. On the other hand, other groups (especially families and groups) are less likely to cancel.

## 2. Reviewing the number of nights booked:

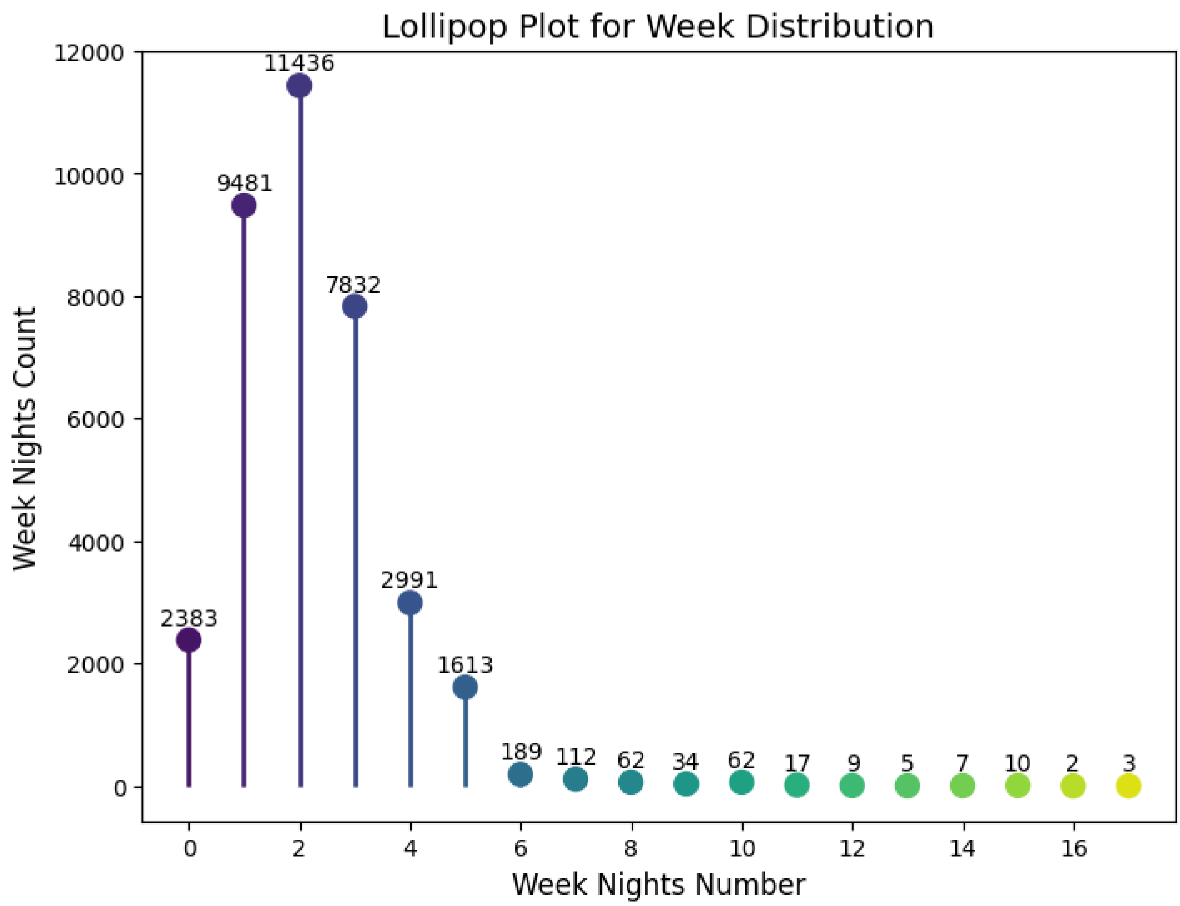
1) **Chart of the number of weekend nights per booking:**

According to the chart of weekend nights per booking, most guests do not include weekend nights in their stay. This indicates that hotel stays mostly happen during weekdays (non-weekends). However, there are longer stays that include more weekend nights. For example, a 7-night weekend stay is associated with guests who have long-term stays, like 24 nights. We will specifically review these guests in more detail in the following charts.




2) **Chart of the number of weekday nights per booking:**

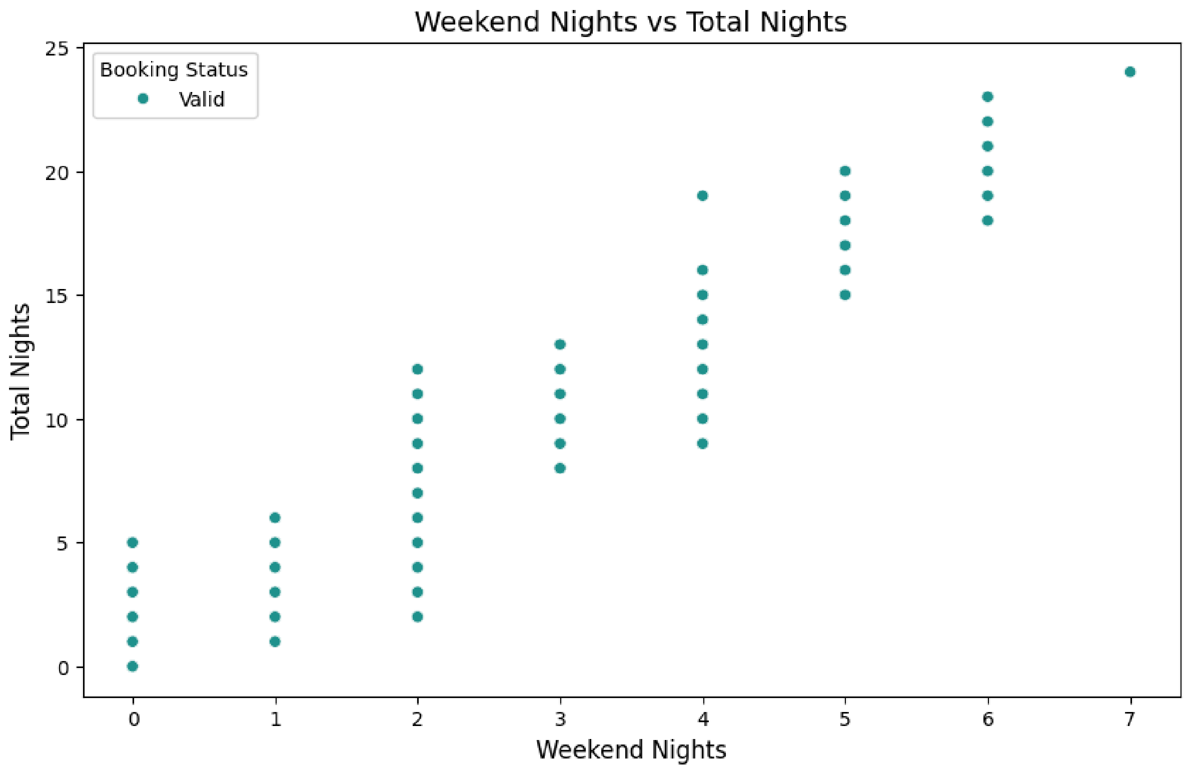
As the chart shows, most weekday nights booked (non-weekend days) range from 1 to 4 nights. Among these, a 2-night stay, with 11,447 groups of guests, is the most popular choice. This pattern suggests that most guests book the hotel for a shorter trip. Additionally, there are longer stays (such as 17 weekdays) associated with guests who have very long stays like 24 nights.



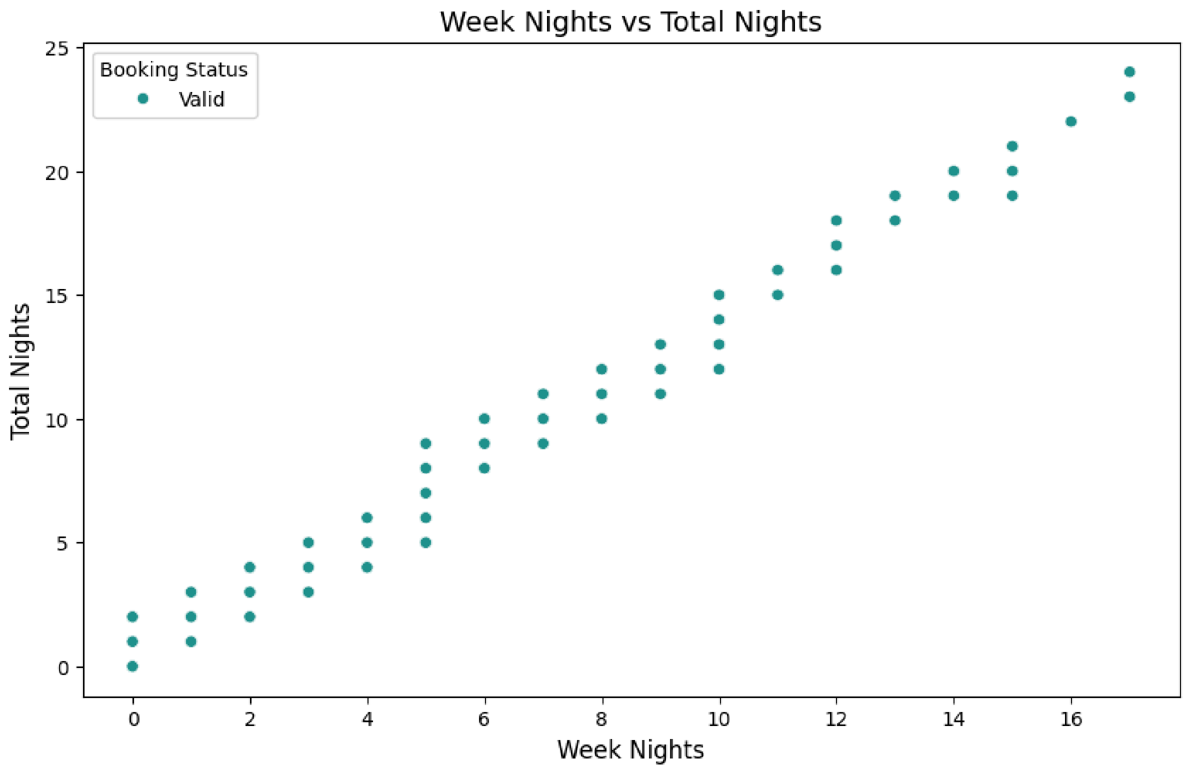

### 3) Chart comparing weekday and weekend stays with the total length of stay:

One standout point in the chart is the 7-night weekend stay. This requires careful review because this number is only logical if the total stay exceeds one week. Therefore, all bookings that include more than 2 weekend nights but have a very short total stay seem invalid and should be removed from the dataset.

According to the chart below, all available data is valid; for instance, the 7-night weekend stays are associated with long-term stays like 24 nights. Therefore, removing invalid data is unnecessary at this stage of preprocessing.

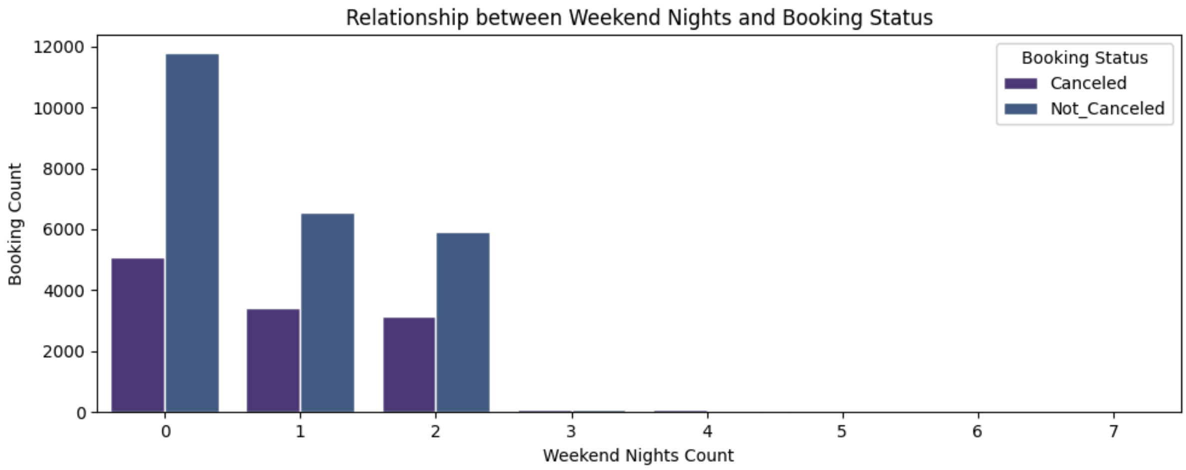


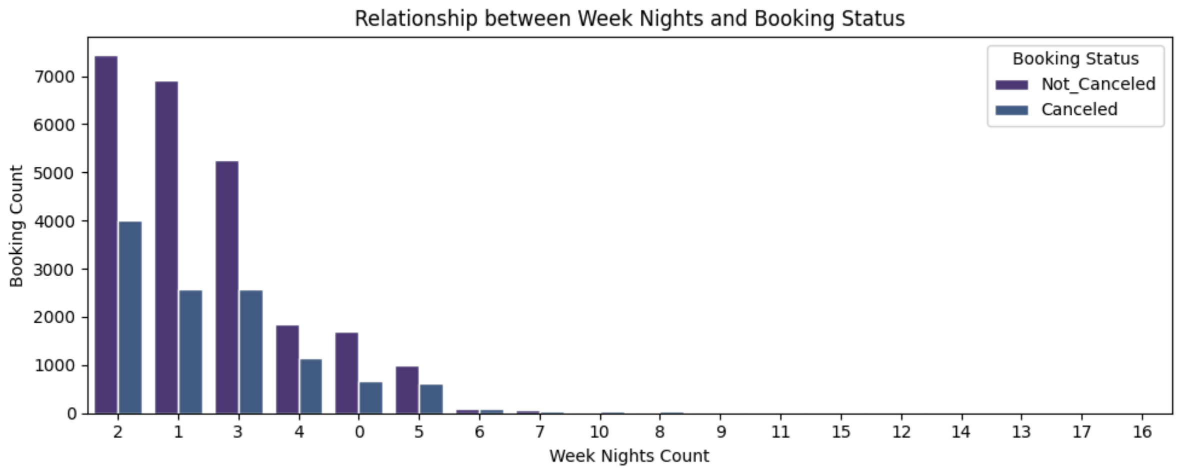

The chart comparing weekday stays with total stays also shows, for example, that guests who booked 5 weekday nights usually have total stays of 5, 6, 7, 8, or 9 nights.




4) **Reviewing the relationship between the number of weekend and weekday nights with booking cancellations:**

- **Number of weekend nights:**
  - ❖ The number of cancelled bookings with 1 or 2 weekend nights is lower compared to bookings without any weekend nights. However, in bookings that include 1 or 2 weekend nights, the number of cancelled bookings is almost half of the non-canceled ones. In contrast, for bookings with no weekend nights, the number of cancellations is less than half of the non-canceled bookings.
  - ❖ On the other hand, bookings without any weekend nights have a higher cancellation rate, meaning that the number of cancelled bookings is almost double that of bookings with 1 or 2 weekend nights.

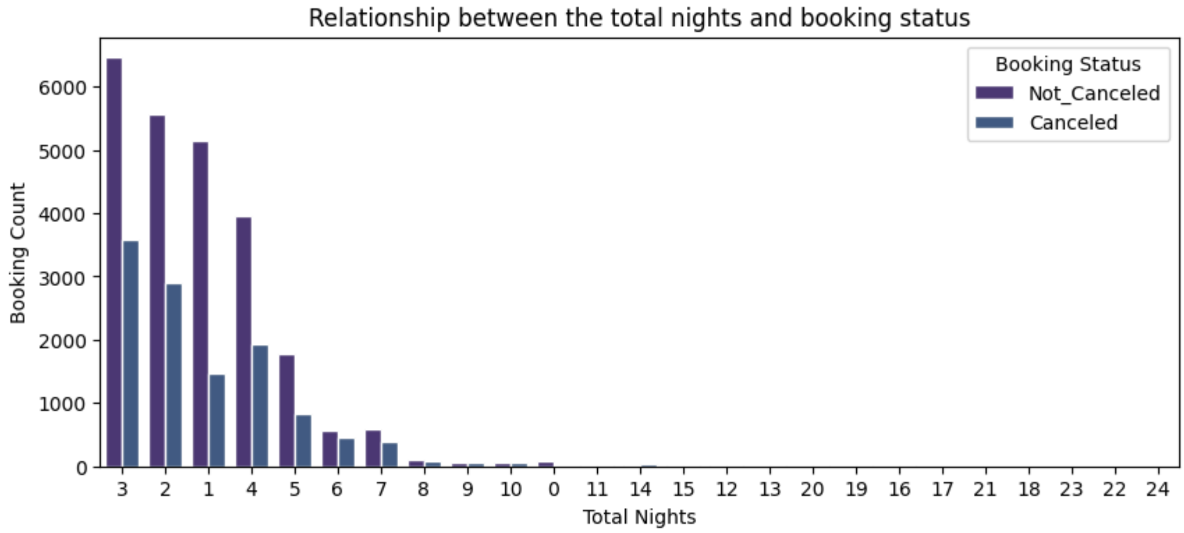



- **Number of weekday nights:**
  - ❖ Most booked stays include at least 2 weekday nights. However, in this group, the number of cancellations is also more than half of the non-canceled bookings. This highlights the importance of weekday nights in customers' decisions to cancel or not.
  - ❖ *Not-cancelled* stays of 1 weekday night are almost equal to those of 2 nights, but the number of cancellations in this group is significantly lower. This could indicate that guests with shorter stays are less likely to cancel due to the limited duration of their stay.

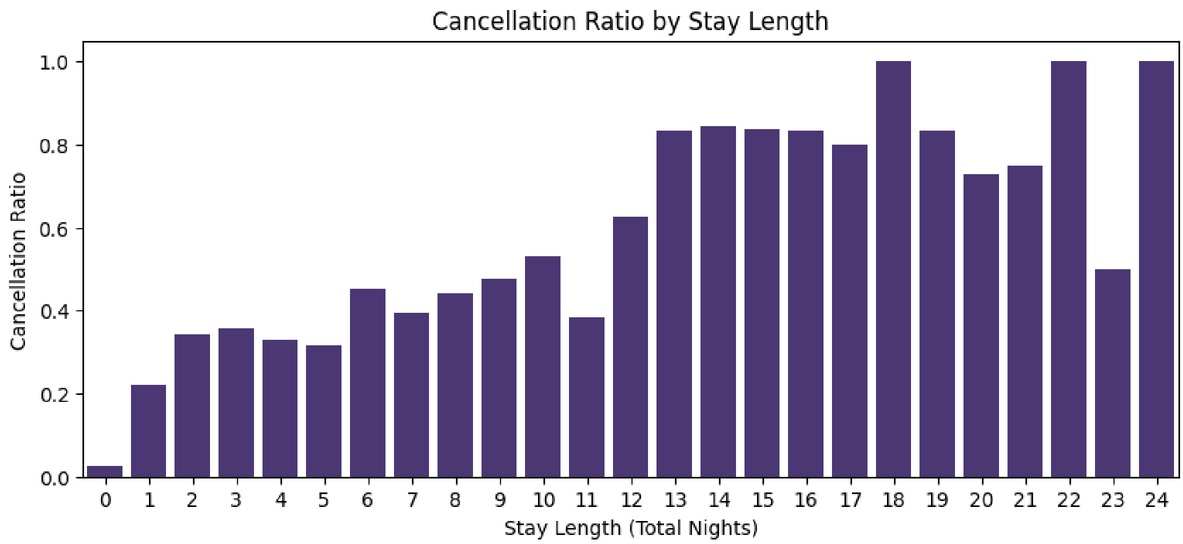
5) **Reviewing the relationship between the total length of stay and hotel booking cancellations:**

According to the chart, the most common stays at the hotel are 3, 2, 1, and 4 nights, respectively. However, for *cancelled* bookings, the order is different. Among them, 1-night bookings have the lowest cancellation rate, while longer stays have a higher cancellation rate. This analysis shows that shorter stays are less prone to cancellation.




6) **Reviewing the ratio of cancellation rate to total length of stay:**

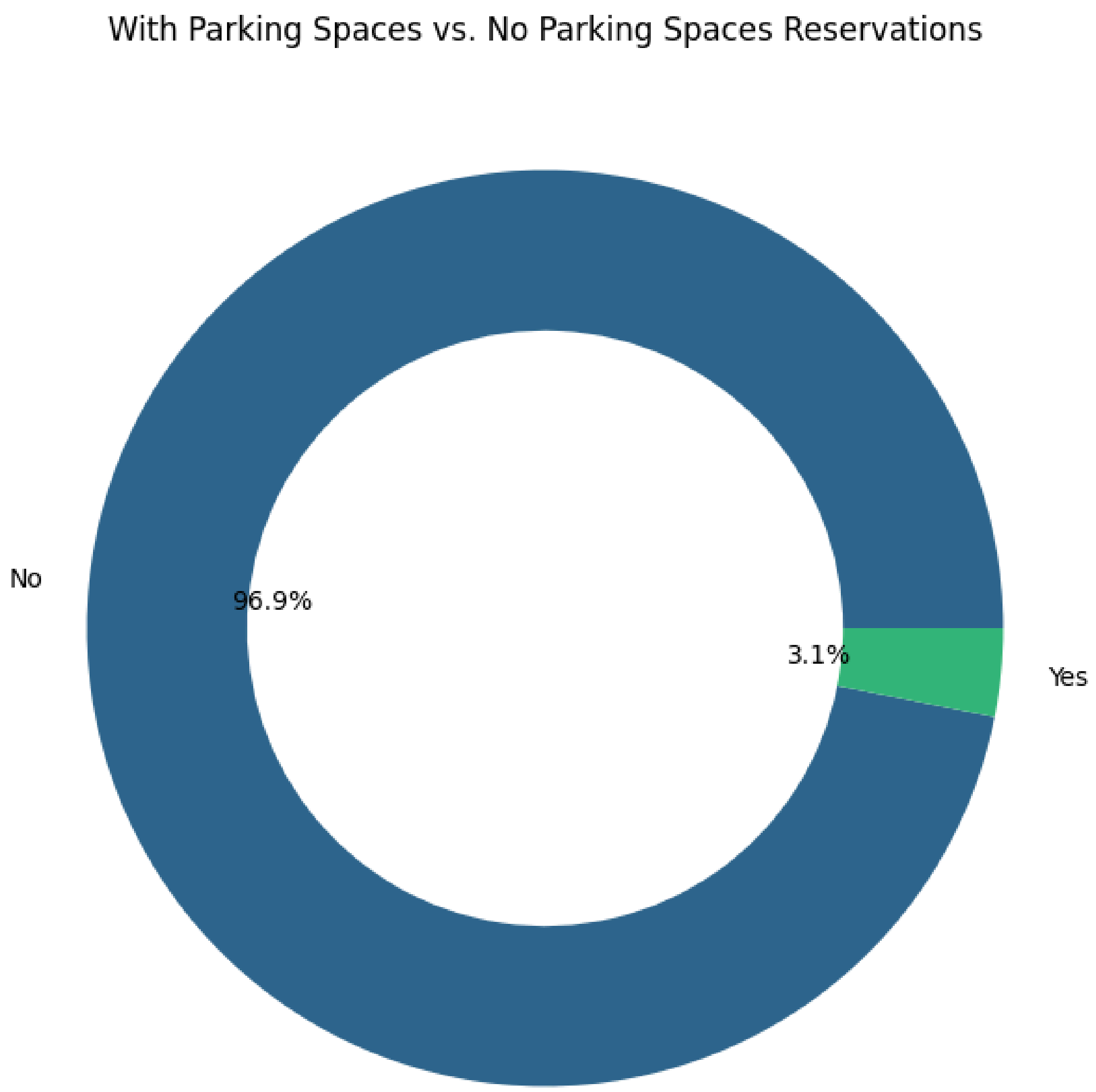
The chart below shows that, for example, only 0.2 of 1-night stays were cancelled. In contrast, almost all stays with a length of 18, 23, and 24 nights were cancelled. This indicates a direct relationship between the length of stay and the likelihood of cancellation. In other words, longer stays are more likely to be cancelled, possibly because customers' travel plans change over time..




## 3. Parking Importance Analysis:

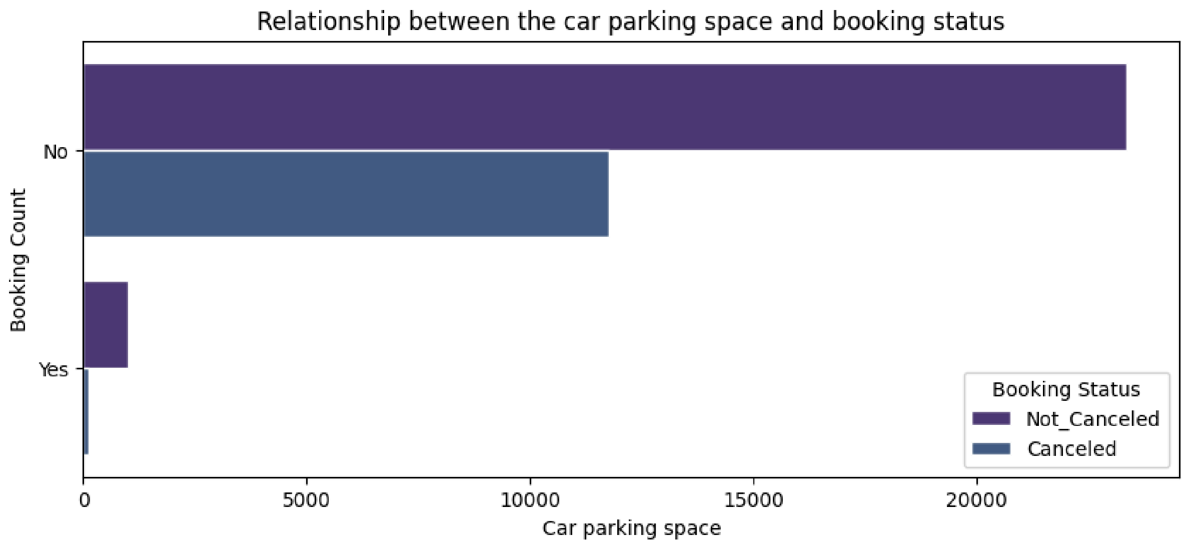
### 1) Chart of the percentage of guests who reserved parking:

Based on the chart, only about 3% of hotel guests have reserved parking. This suggests that most guests are likely travelling via public transportation, such as aeroplanes, trains, or buses. It is also possible that the hotel's location near public transportation stations makes parking less necessary.




2) **Parking Reservation and Booking Cancellation Correlation:**

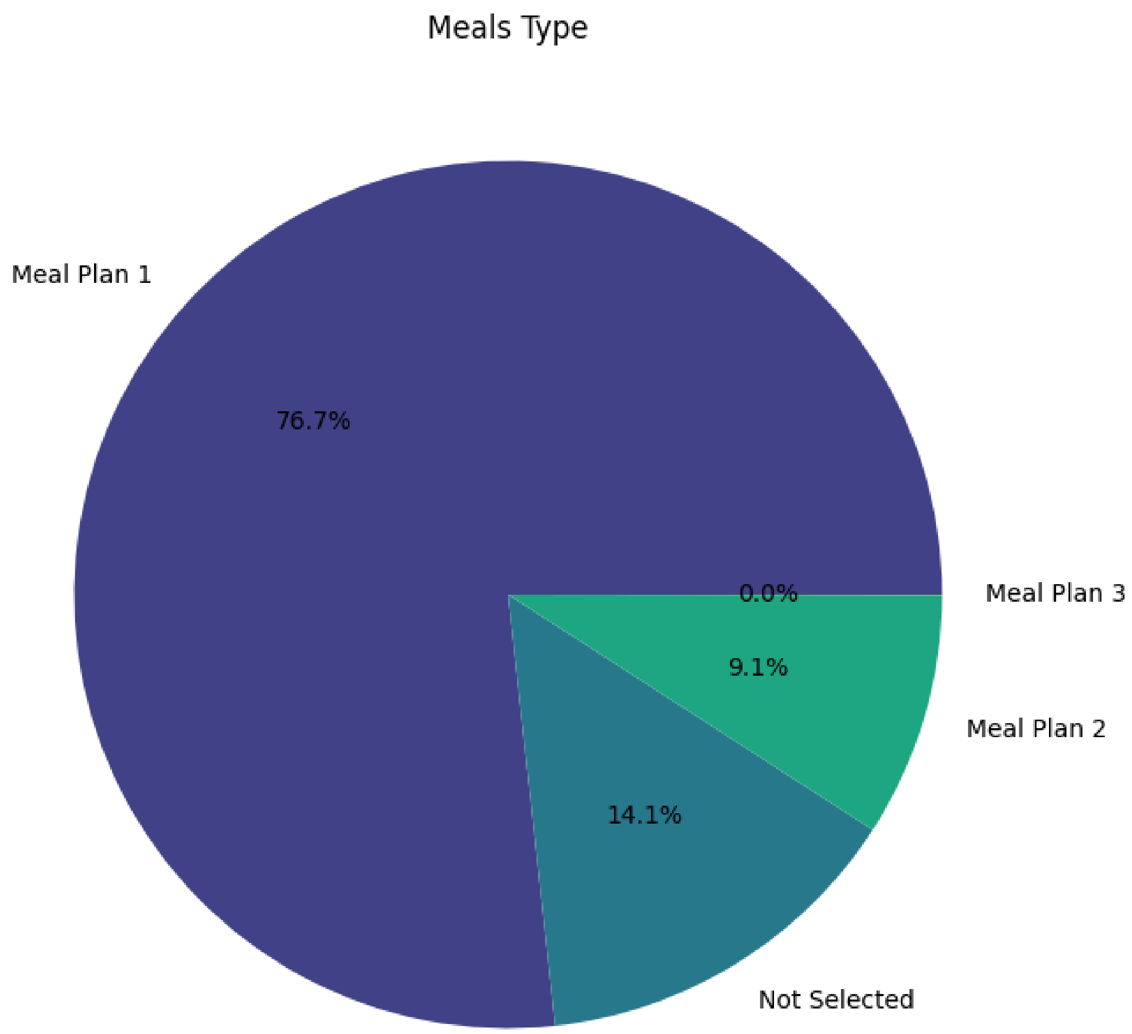
Almost half of the guests who did not reserve parking cancelled their booking. This suggests that reserving parking might be one factor that encourages guests not to cancel their reservation. Interestingly, almost none of the guests who reserved parking cancelled their booking.



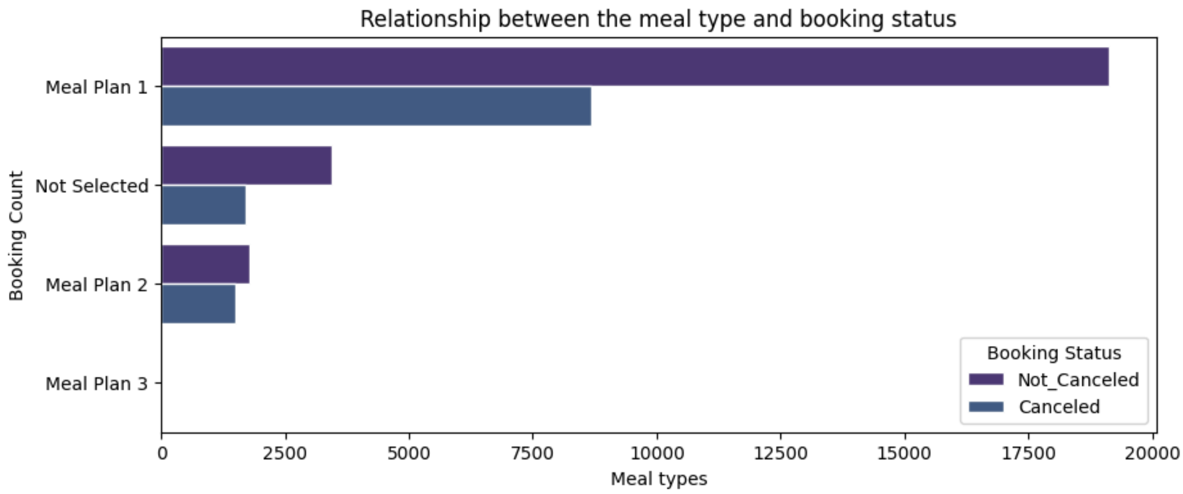

## 4. Importance of Chosen Meal Type:

1) **Chart of the percentage of meals reserved by guests:**

According to the chart, Meal Type 1 is the most popular choice, with about 77% of guests selecting it. On the other hand, about 14% of guests did not reserve any meals. Meal Type 2 is in third place, while Meal Type 3 was not chosen by any guests.

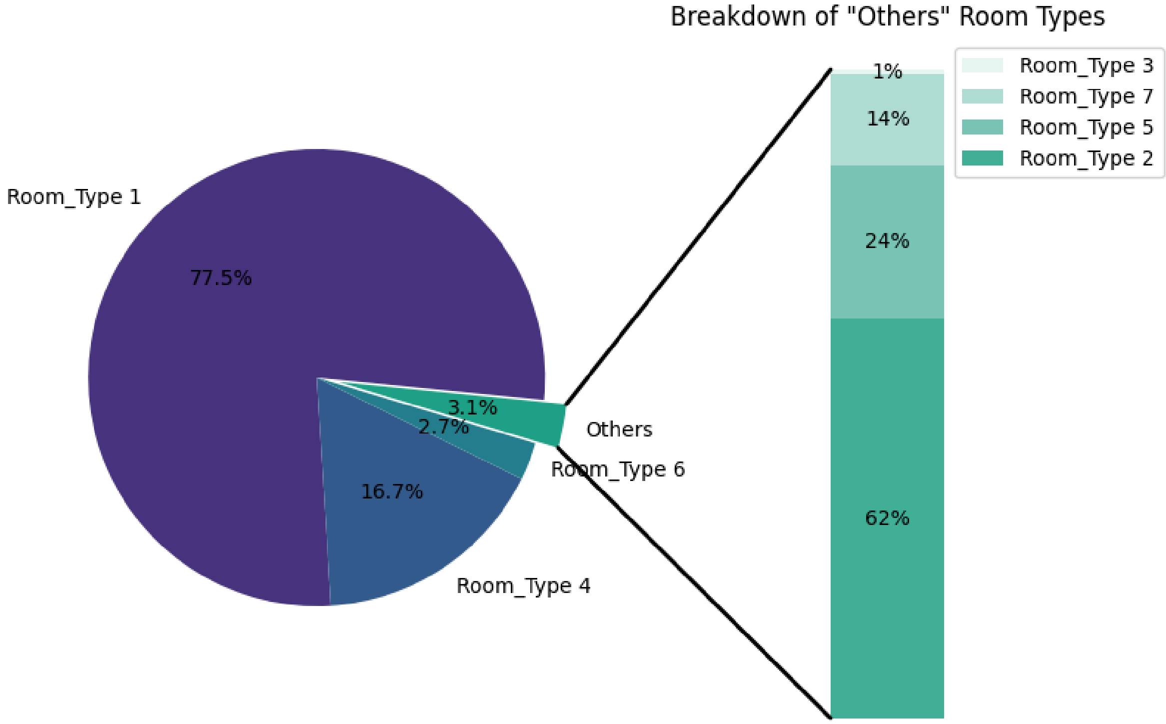



2) **Meal Type and Booking Cancellation Correlation:**
The cancelled bookings of guests who selected Meal Type 1 or did not reserve any meals are slightly less than half of the uncanceled bookings of these groups. Additionally, uncanceled bookings for guests who selected Meal Type 2 slightly exceed the cancelled ones.



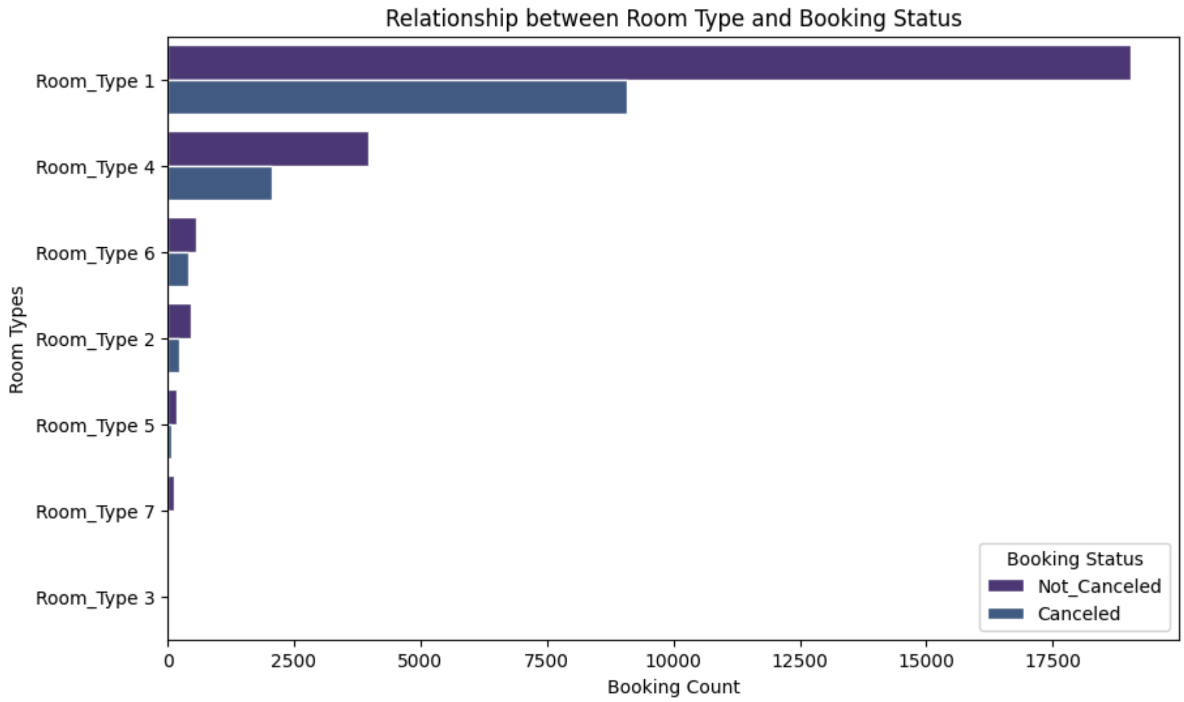

# 5. Importance of Chosen Room Type:

1) **Chart of the percentage of rooms reserved by guests:** About 77.5% of guests reserved Room Type 1, which is likely due to its affordable price. This should be analysed further in subsequent charts. Room Type 4 accounts for about 16.7%, and Room Type 6 for 2.7%, respectively. Room Types 2, 3, 5, and 7 combined make up just over 3% of the total reservations.



2) **Room Type and Booking Cancellation Correlation:**

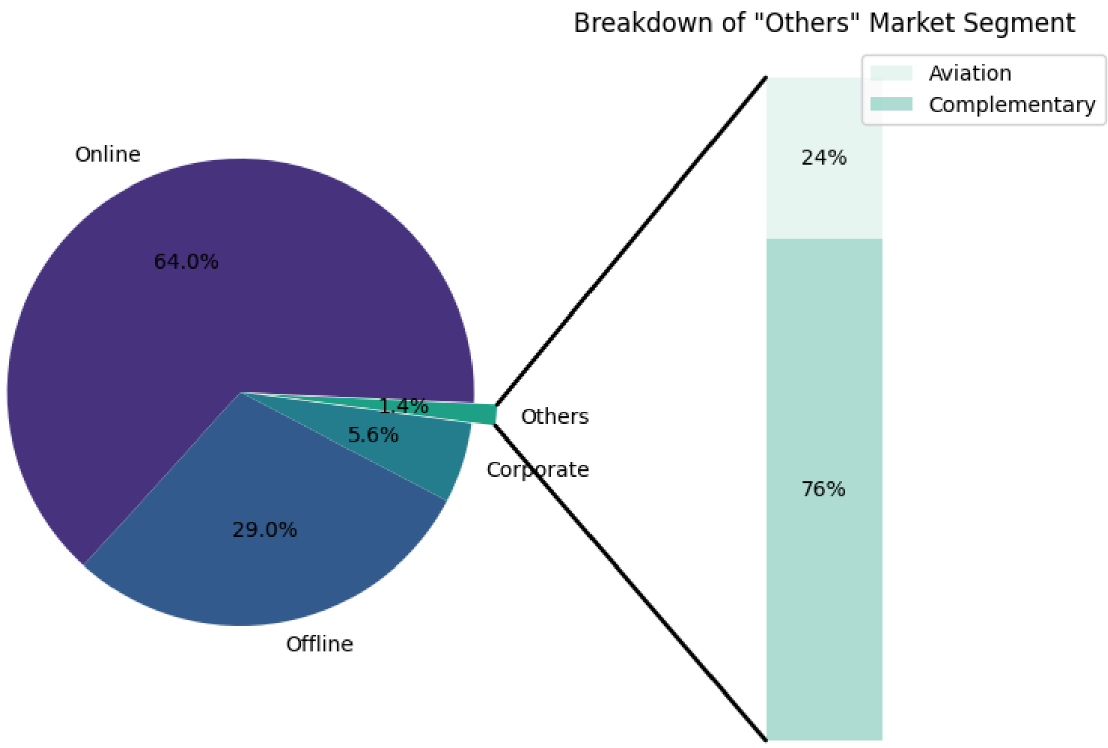
Uncanceled bookings for different room types are almost twice the number of cancelled bookings. This pattern is nearly consistent across all room types..




# 6. Importance of the Booking Method:

1) **Chart of the percentage of booking methods used by guests:**

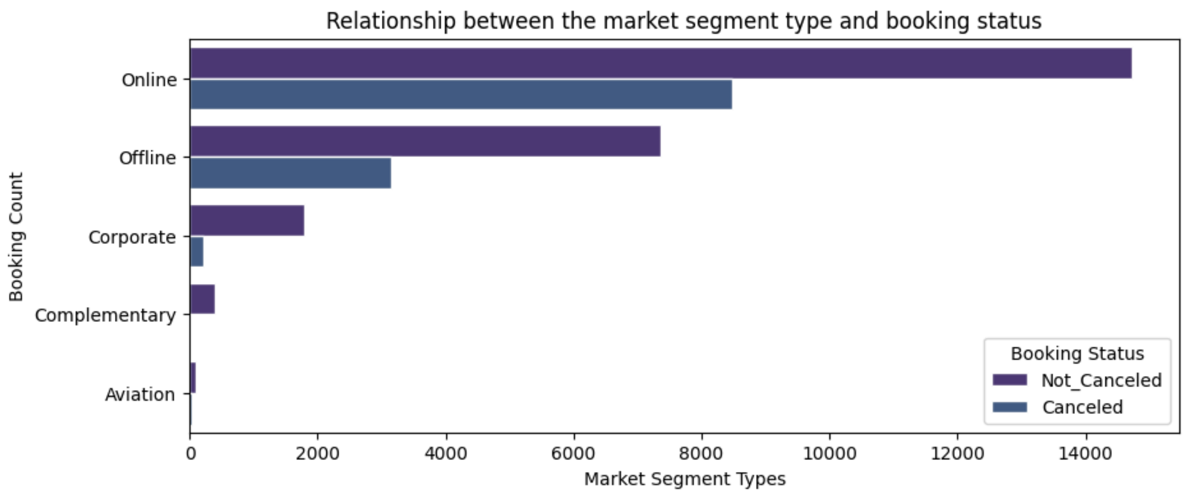
Approximately 64% of guests booked their room online, 29% booked offline, and 5.6% booked through corporate reservations. Only 1.4% of the bookings were made through the Aviation or Complementary methods.




## 2) **Booking Method and Cancellation Correlation:**

| | |
|---|---|
| Online | Bookings made through online travel agencies or hotel websites. |
| Offline | Bookings made through traditional travel agencies or direct phone calls. |
| Aviation | Bookings made through airlines, usually for airline staff or passengers with layovers. |
| Corporate | Bookings made by companies for employees travelling for business purposes. |
| Complimentary | Complimentary stays offered by the hotel, usually for promotional purposes or special guests. |

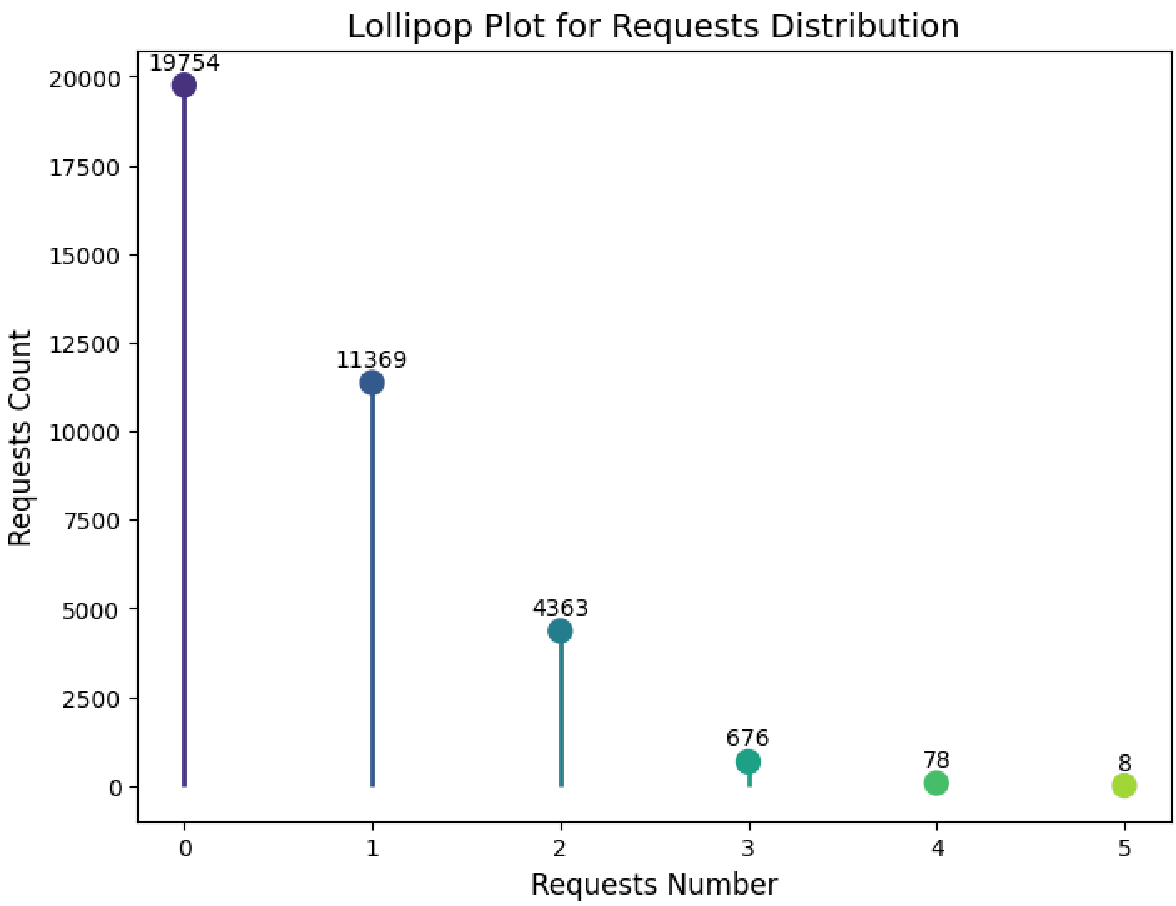
Uncanceled bookings made via online and offline methods are about twice the number of cancelled bookings. For the *Complementary* method, only a very small percentage of cancellations occurred, and for the two other methods, almost none of the bookings were cancelled.




# 7. Importance of Special Requests

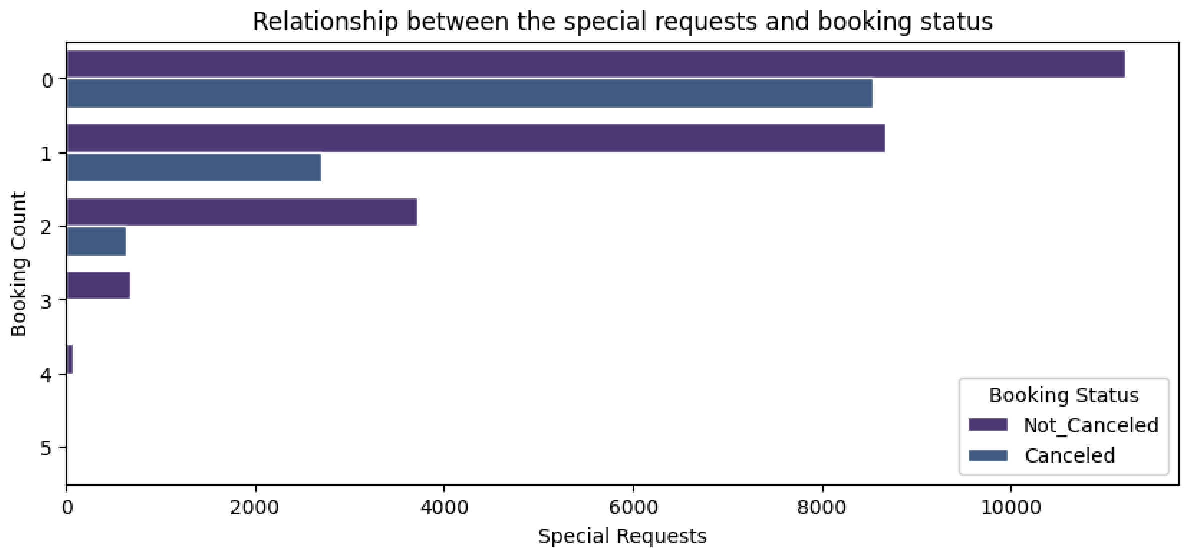
1) **Chart of the number of special requests per booking:**

Most bookings did not have any special requests. This might indicate that most trips are simple and short.




2) **Special Requests and Booking Cancellation Correlation:**

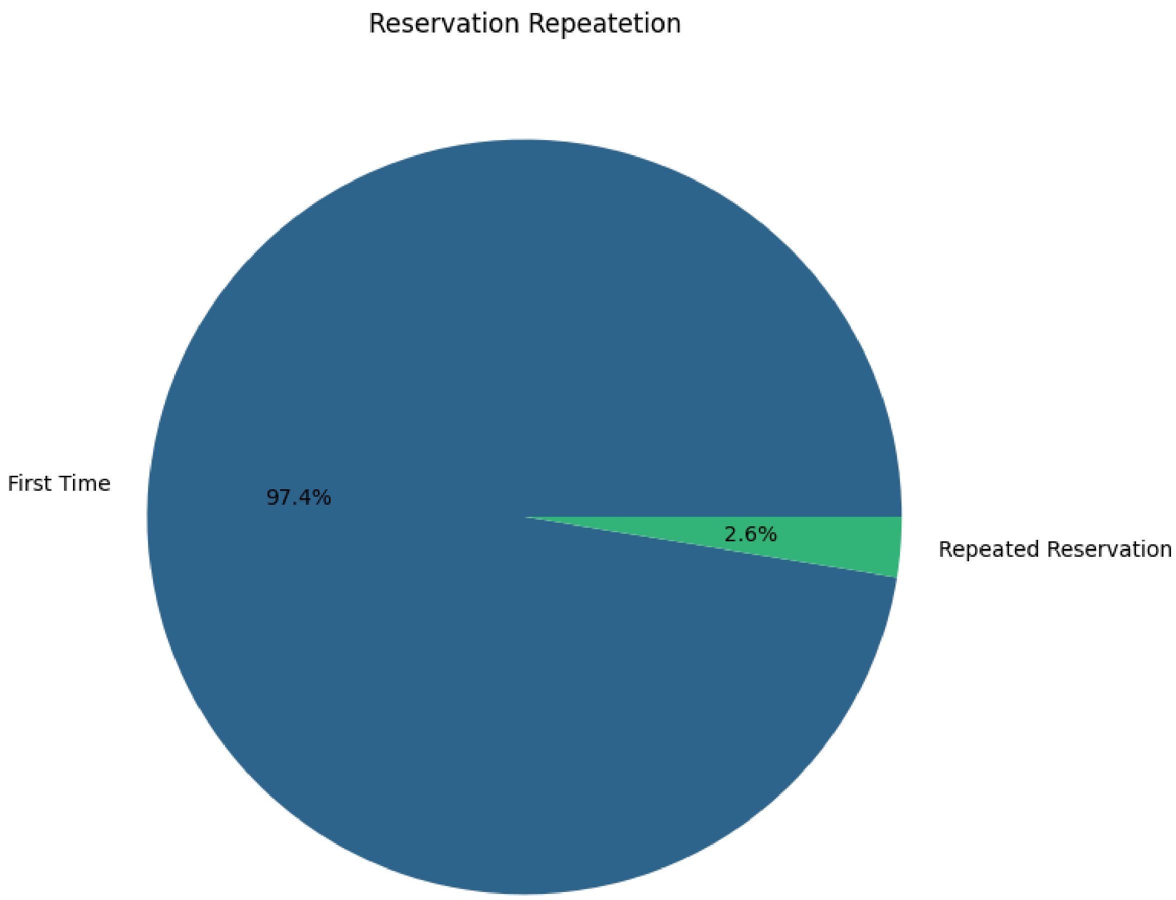
Guests who made no special requests easily cancelled their bookings. However, the more special requests made, the lower the likelihood of cancellation. For example, guests who made more than three special requests did not cancel their bookings at all.




## 8. Importance of Repeat Guests

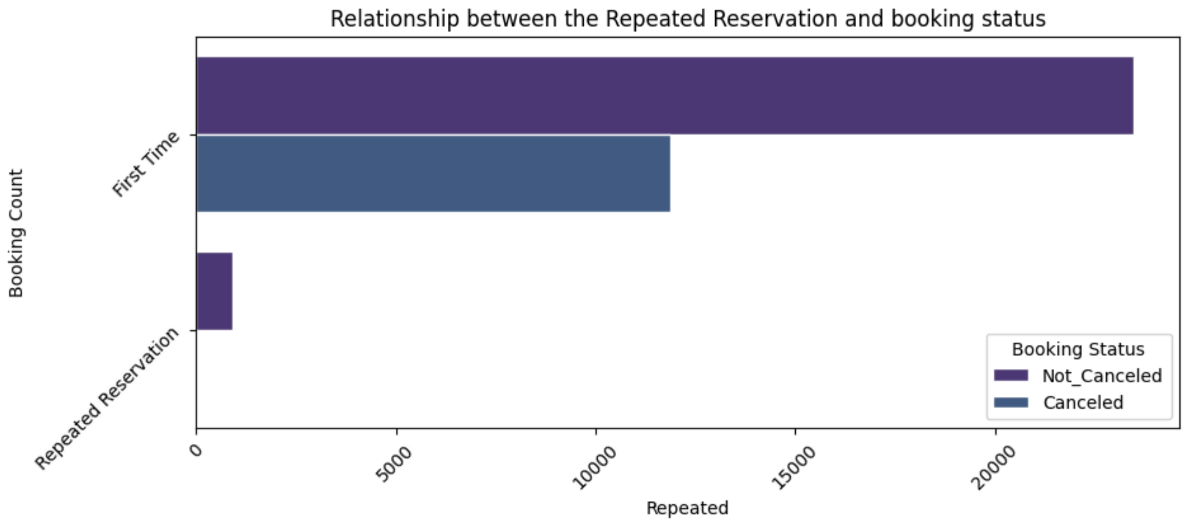
1) **Chart of the percentage of bookings made by repeat guests:**

About 97.4% of guests were staying at the hotel for the first time.




2) **Repeat Guest and Booking Cancellation Correlation:**

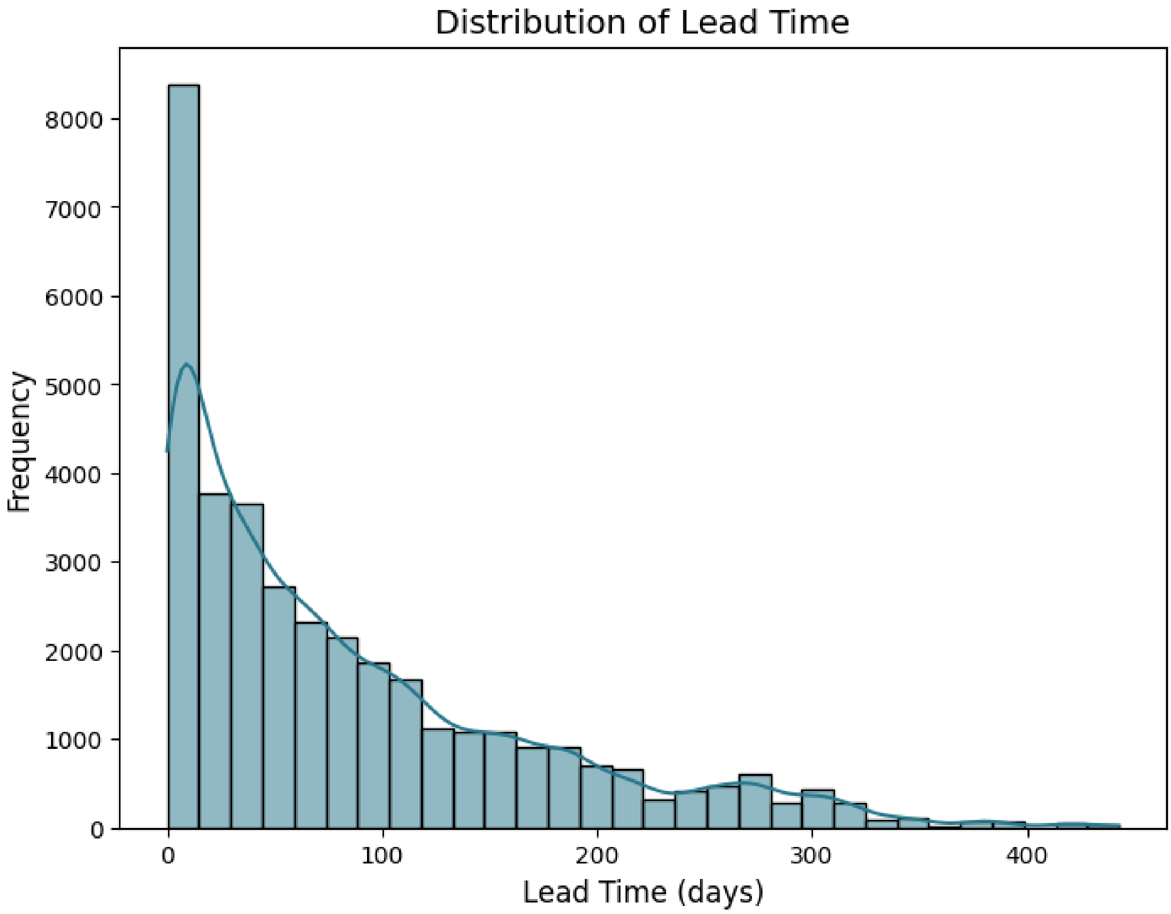
Almost half of the first-time guests' bookings were cancelled, whereas almost none of the repeat guests cancelled their bookings. This suggests that repeat guests have more trust in the hotel, and the likelihood of cancellation among them is significantly lower.




# 9. Importance of the Lead Time:

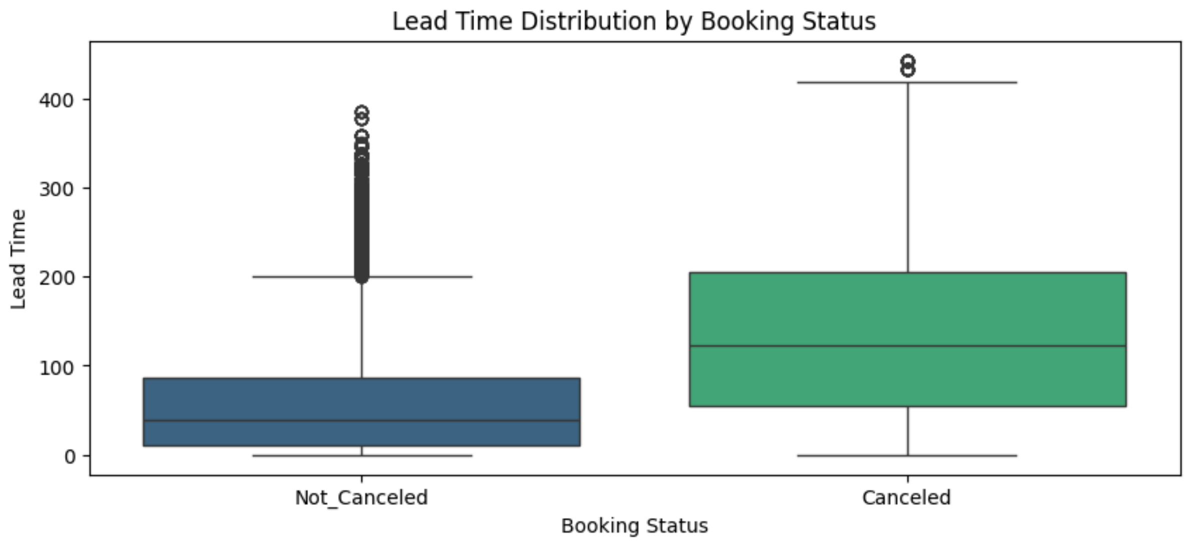
1) **Histogram of the lead time:**

The histogram showing the time gap between booking and the stay date reveals that the majority of guests book their hotel around 20 days before their arrival date. This indicates that most guests prefer to plan their trips close to their travel date and do not need long-term bookings. However, a limited number of guests, particularly those travelling for special reasons or long holidays, may book up to 400 days in advance. This suggests that some guests book early to ensure they reserve a room or special hotel amenities.




2) **Correlation between Lead Time and Booking Cancellation:**

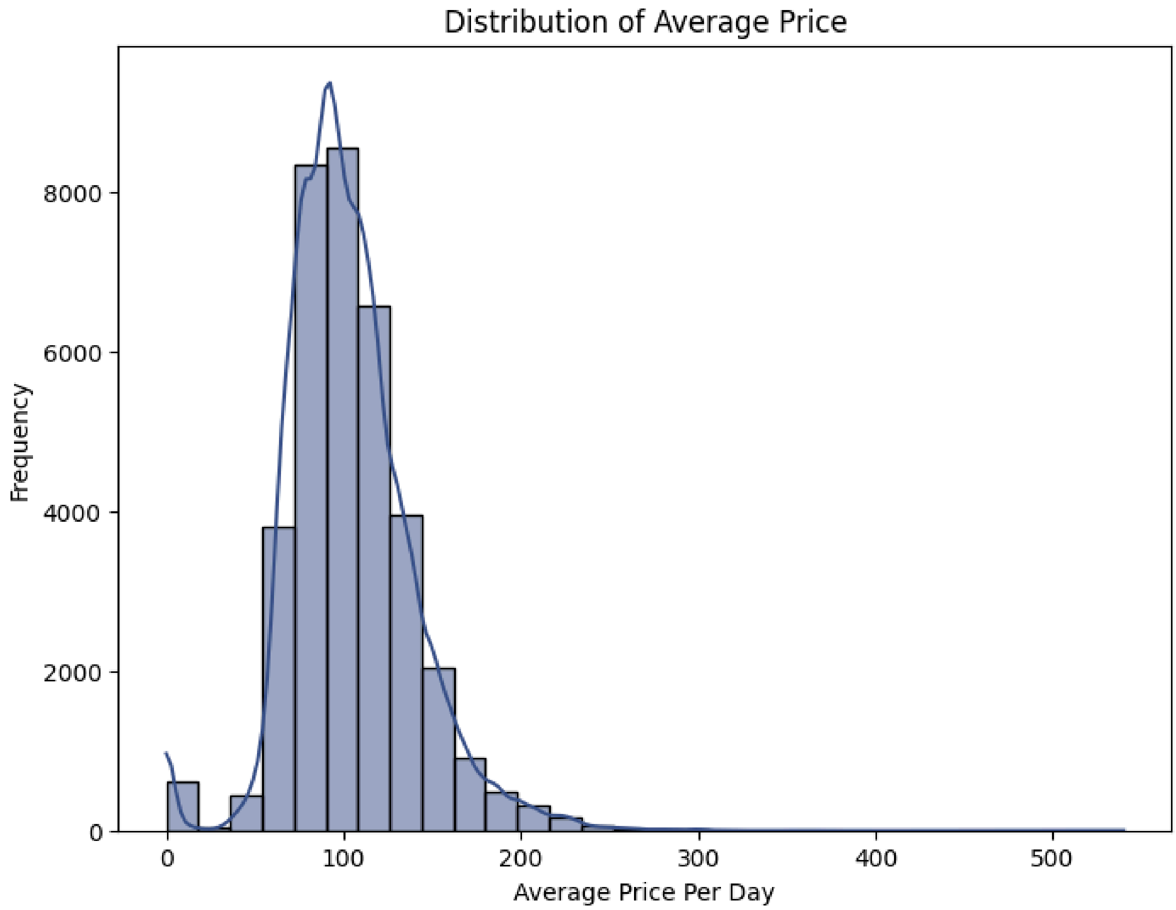
A closer look at the data shows that the longer lead time, the higher the cancellation chance. Bookings made closer to the stay date are less likely to be cancelled. This phenomenon could be due to long-term plans being more susceptible to change, while bookings made closer to the stay are more definitive.




## 10. Importance of Average Price:

1) **Histogram of average price:**

The histogram of payments for bookings shows that most guests paid around $100 for their reservations. A deeper analysis reveals that the highest amount paid by guests for a booking reached $600. This variation in costs may be related to significant differences in room types, hotel amenities, or the length of the stay.



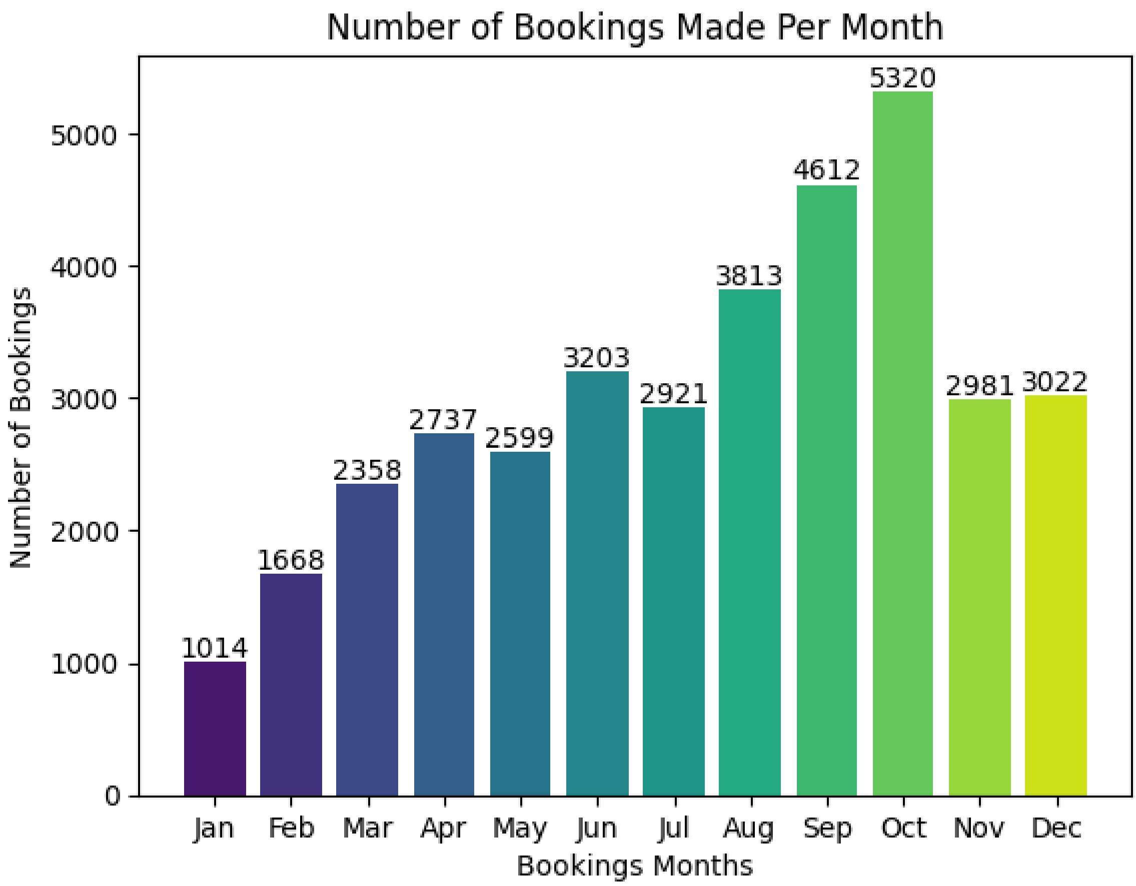

2) **Correlation between average price and booking cancellation:**

Interestingly, the average payment amount for both cancelled and non-canceled stays is almost the same. Therefore, the cost of a booking does not seem to have a direct impact on whether a stay is cancelled or not. This suggests that cancellations are more dependent on other factors such as the guest's planning or the time between booking and stay rather than the amount paid.


# 11. Importance of Time:

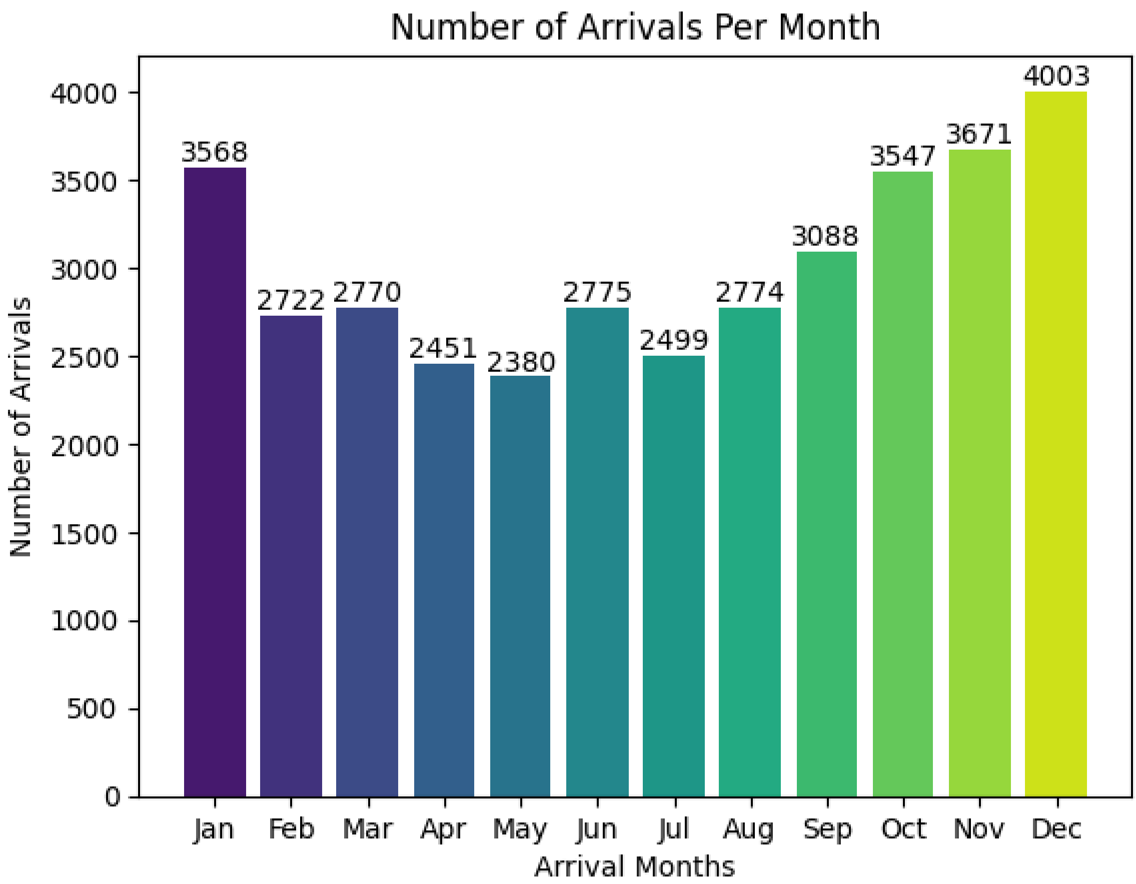
1) **Monthly total bookings chart:**

Most bookings were made in October and September, while the fewest were made in January. This is likely because travellers plan their trips for the New Year several months in advance. Additionally, very few guests book hotels during the New Year's holiday itself.



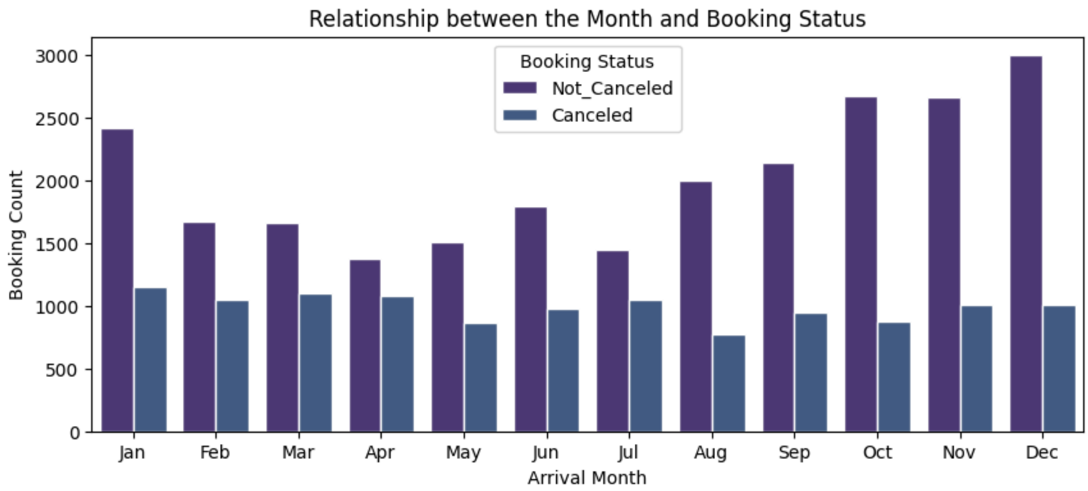

2) **Guest arrivals chart:**

The highest number of guest arrivals occurred in December and January, with the fewest arrivals in May, April, and July. A significant portion of people travel during the New Year holidays, increasing hotel stays during December and January.



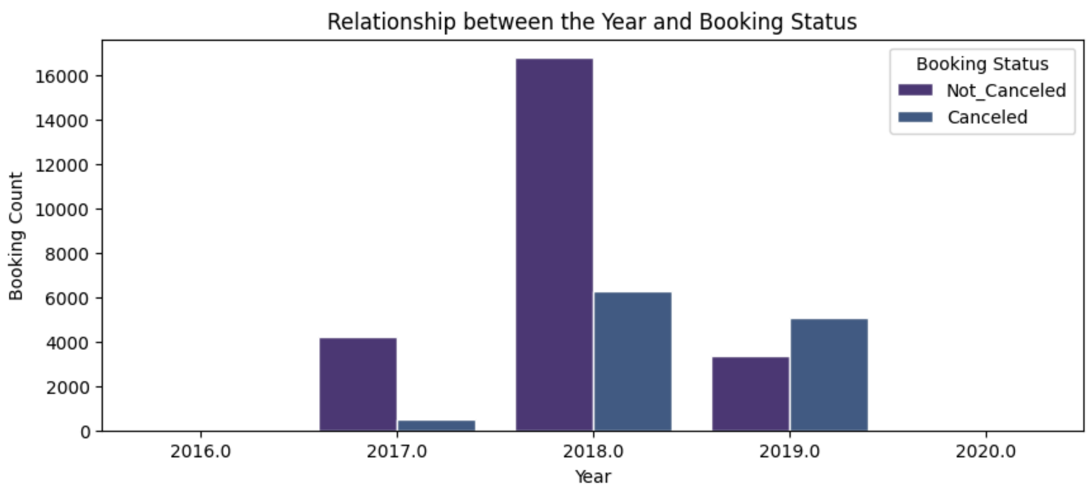

3) **Correlation between Arrival Month and Booking Cancellation:**

Although there are significant differences in the number of *not-cancelled* bookings each month, the number of *cancelled* bookings remains fairly constant across all months, with around 1,000 cancellations each month. This consistency in cancellations might indicate that external factors continuously influence guests' decisions to cancel or keep their reservations, regardless of the season.




4) **Correlation between Year of Arrival and Booking Cancellation:**

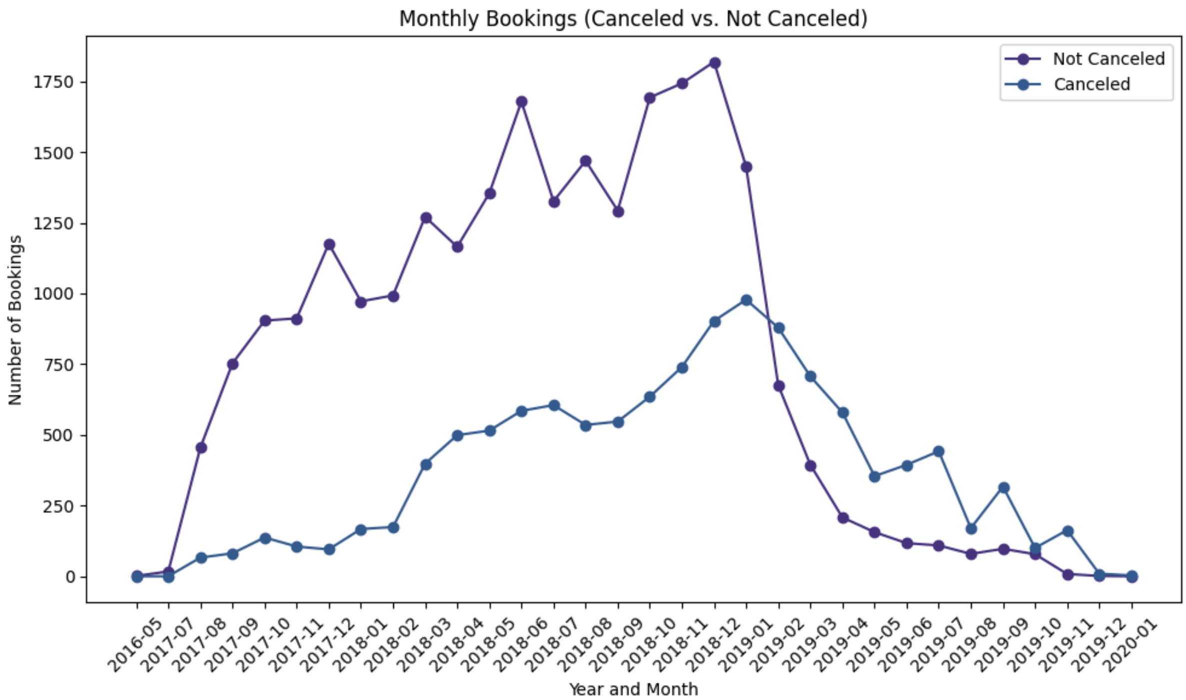
اد The data shows that 2018 had the highest number of confirmed and cancelled bookings. This may be because a large portion of the available data pertains to 2018.




5) **Trend in Booking and Cancellation Over the Entire Period:**

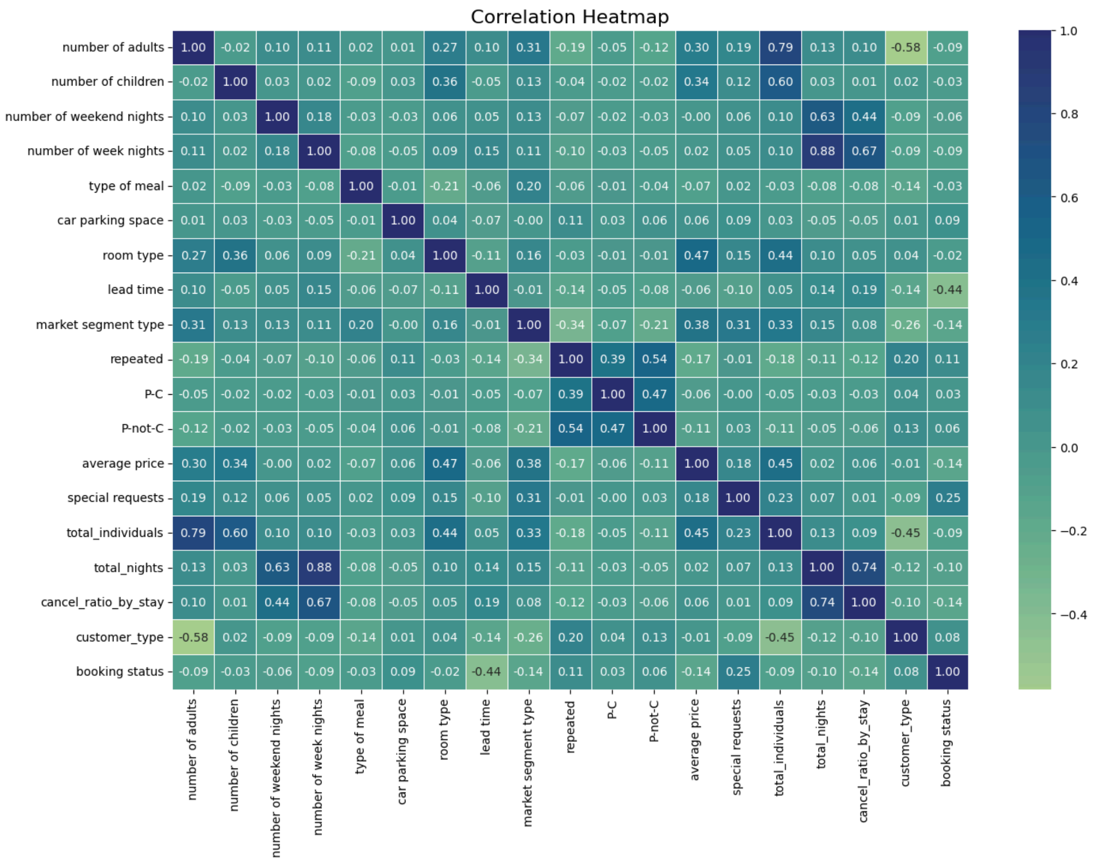
❖ As shown in the chart, until January 2019, both confirmed and cancelled bookings were on the rise, reaching their peak in that month. However, from that point onward, both trends declined.

❖ An interesting observation is that before January 2019, confirmed bookings significantly outnumbered cancellations. After this date, the trend reversed, with cancellations overtaking confirmed bookings. In July and August 2019, the number of both types of bookings became equal and eventually dropped to zero. These changes could be related to factors such as shifts in consumer behaviour, economic fluctuations, or changes in hotel policies.




## 12. Feature Importance and Target Value Chart:

This chart shows the correlation between features and the model target. Some features have a correlation of up to 0.88 with each other. In such cases, if the correlation exceeds 0.7 or 0.8, one of the features should be removed, or dimensionality reduction techniques such as PCA should be used.




# 13. The Relationship Between Customer Type, Meal Type, Room Type, Market Segment Type, Arrival Time, and Average price and Booking Status:

## 1) The Relationship Between Customer Type, Average Price, and Booking Status:

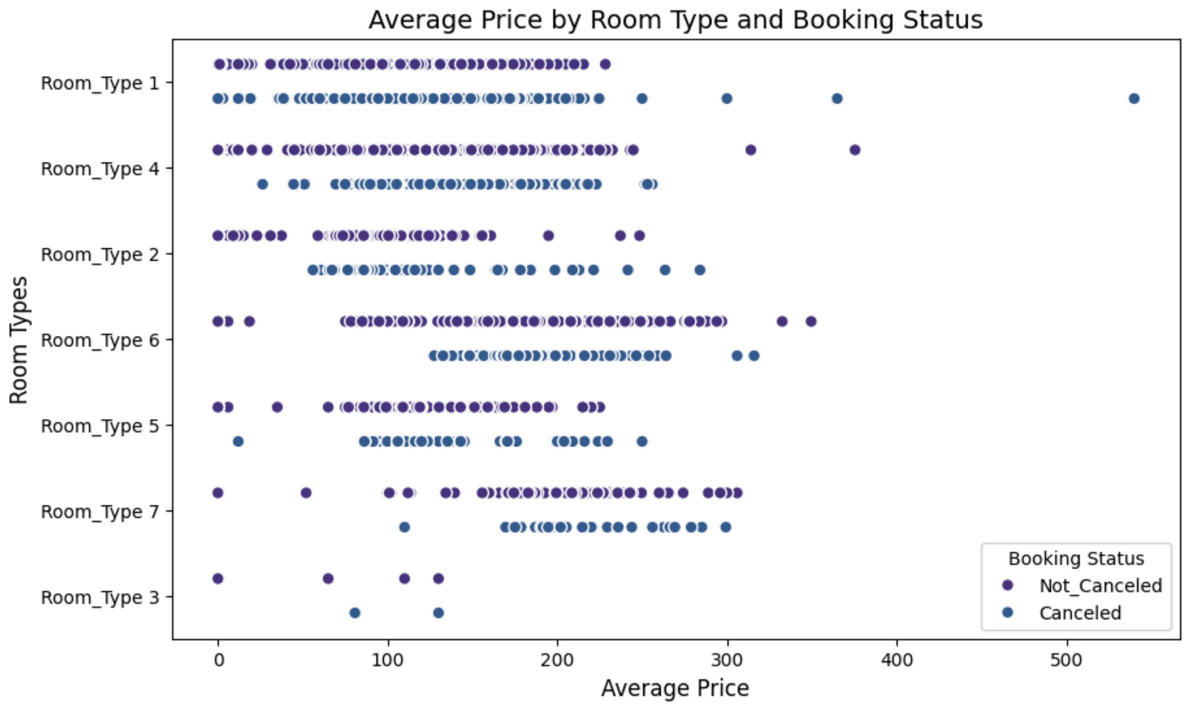
The chart examining the relationship between customer type and average price shows that the four customer types (family, solo, group, and couple) exhibit different behaviours regarding booking cancellations. For the first three groups—families, solo travellers, and couples—the average price of cancelled bookings is higher than that of not-cancelled ones. This may be due to long-term planning changes or personal factors within these groups. Additionally, the highest recorded cost in this dataset ($600) belongs to a couple's booking, which was eventually cancelled. Hence, it seems that customers with higher booking costs are more likely to cancel.


2) **The Relationship Between Meal Type, Average Price, and Booking Status:**

The chart reveals that meal type does not have a significant effect on the cost of cancelled or not-cancelled bookings. The four meal categories examined show limited differences in terms of the costs paid in either cancelled or not-cancelled statuses. This suggests that food costs do not significantly influence the decision to cancel, with passengers paying similar amounts in both cases.

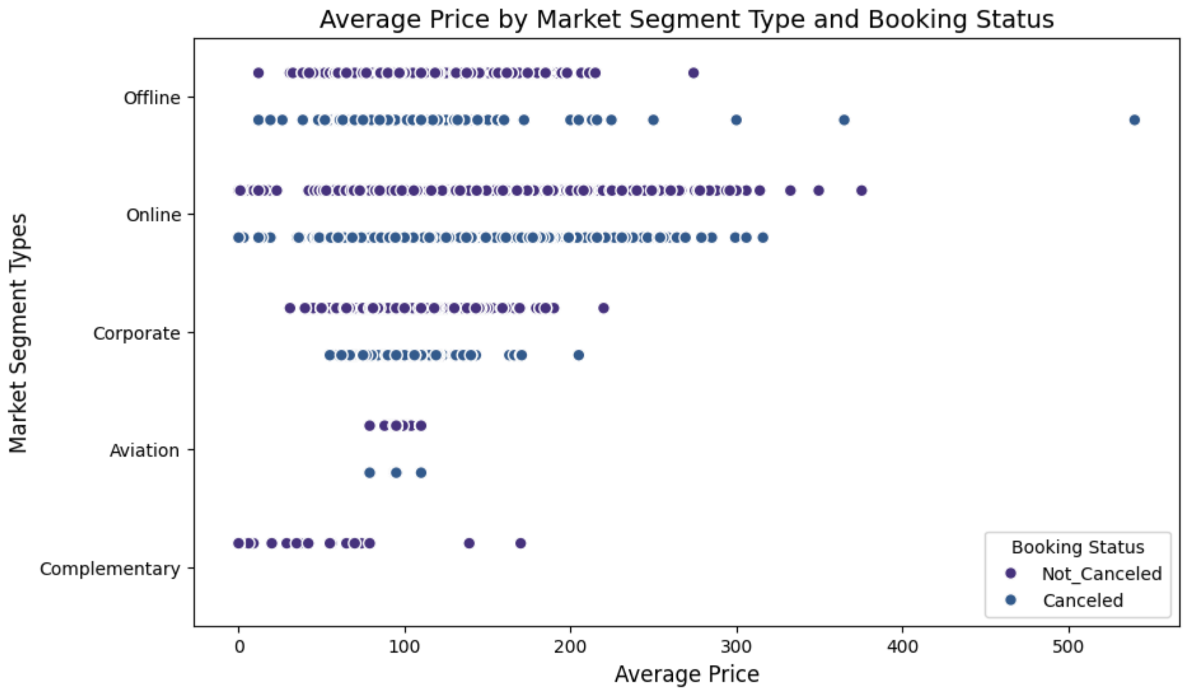

### 3) The Relationship Between Room Type and Average Price, and Booking Status:

The analysis of the relationship between room type and booking status shows that the price range for cancelled bookings in Room Type 1 is broader, indicating that cancellations in this room type are independent of the price. However, in other room types, cancelled bookings tend to be more expensive, suggesting that pricier bookings are more prone to changes or cancellations.



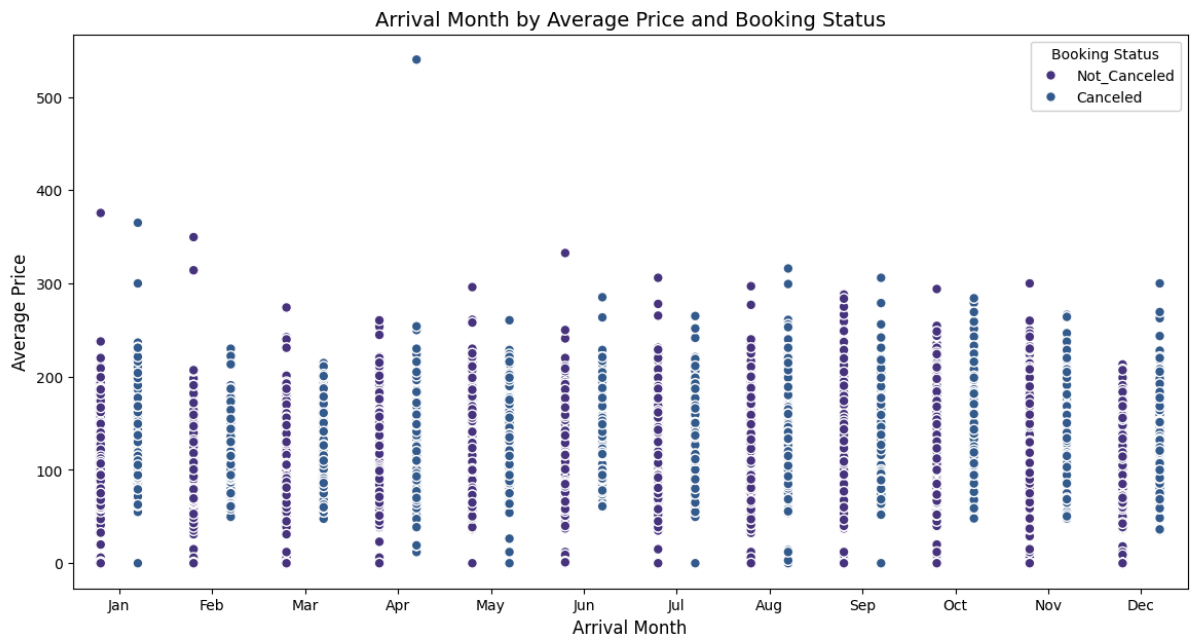

4) **The Relationship Between Market Segment Type, Average Price and Status:**

The booking method (Market Segment Type) can also affect the cost and cancellation likelihood. According to the chart, the bookings cost of cancelled and not-cancelled reservations across different categories is almost the same, except for complementary bookings, none of which were cancelled. In these cases, travellers are less likely to cancel due to receiving free stays or special discounts, highlighting that complementary bookings offer greater financial motivation to stay.



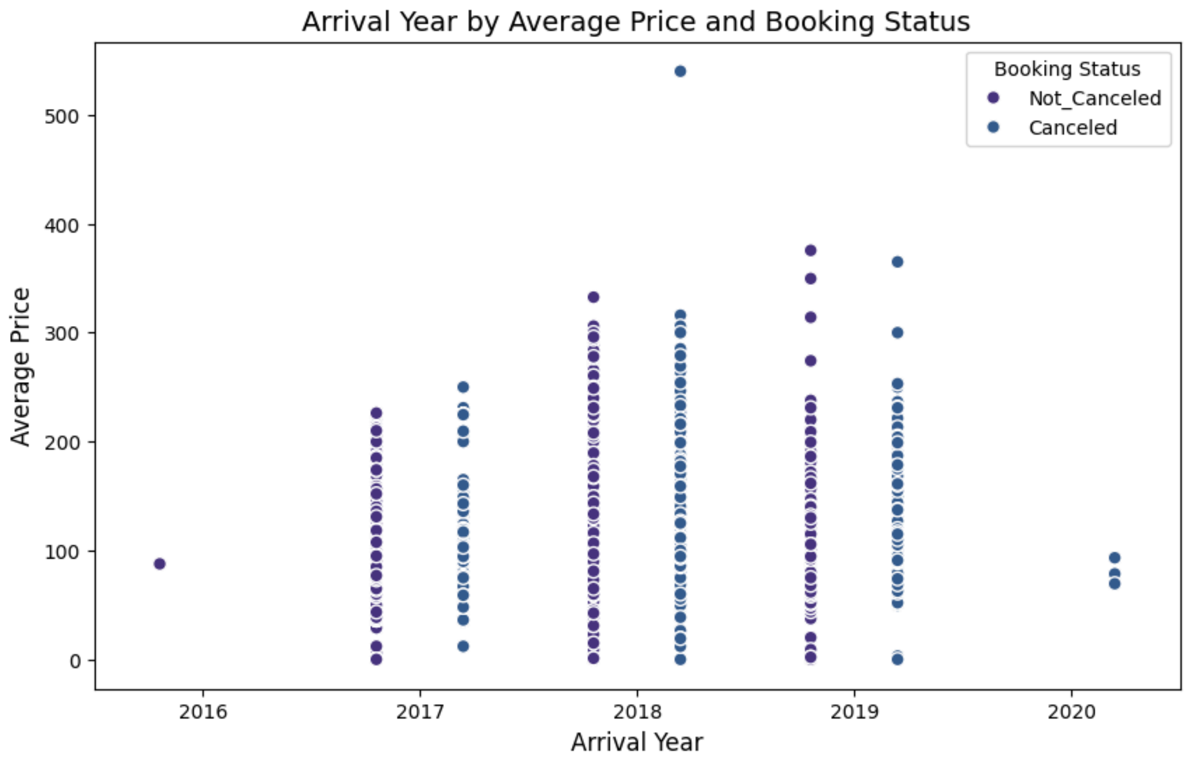

5) **The Relationship Between Entry Month, Average Price, and Booking Status:**

The chart shows that the cost pattern across different months for cancelled and not-cancelled bookings does not significantly differ. However, in half of the months, the initial cost for cancelled bookings is higher. This could be due to travellers being more inclined to cancel expensive bookings in certain months.




6) **The Relationship Between Entry Year, Average Price, Booking Status:**

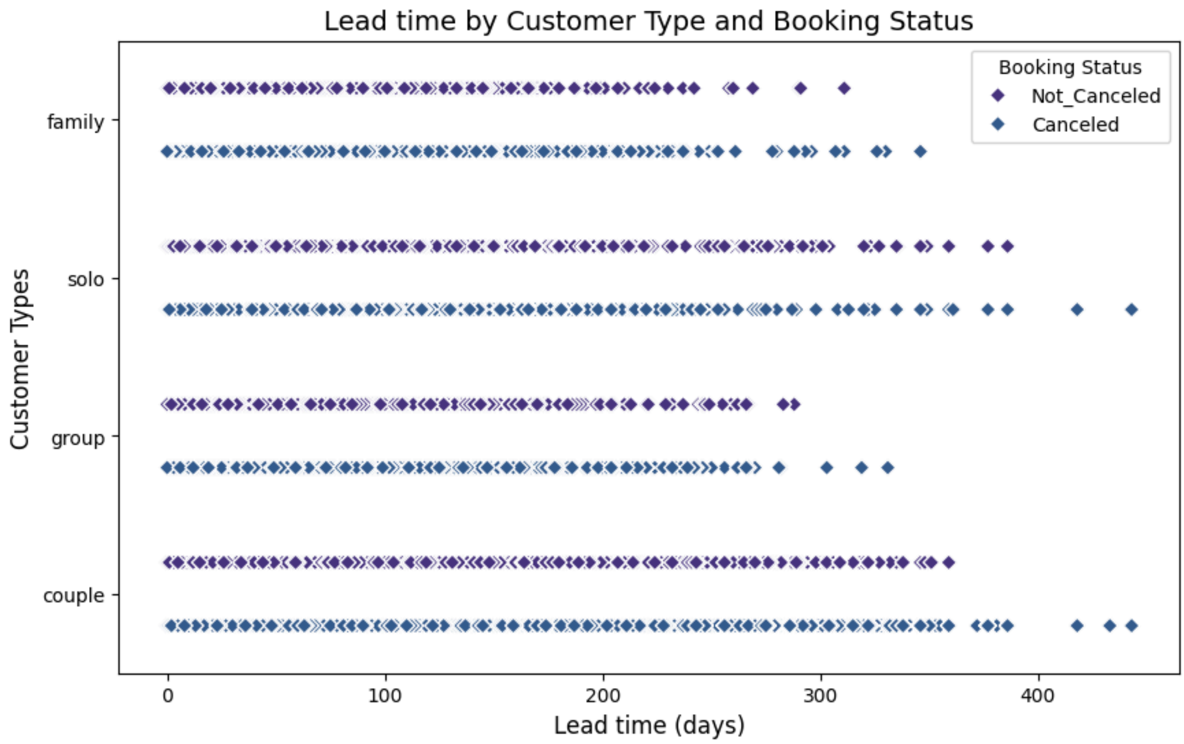
The cost of bookings increased between 2017 and 2019, with a particularly notable rise in 2019. Additionally, the cost pattern for cancelled and non-canceled bookings remains almost identical, indicating that the year alone doesn't greatly influence the decision to cancel. Instead, factors like economic changes and inflation could explain the cost increase.




## 14. The Relationship Between Customer Type, Meal Type, Room Type, Market Segment Type, Arrival Time, and Lead Time and Booking Status:

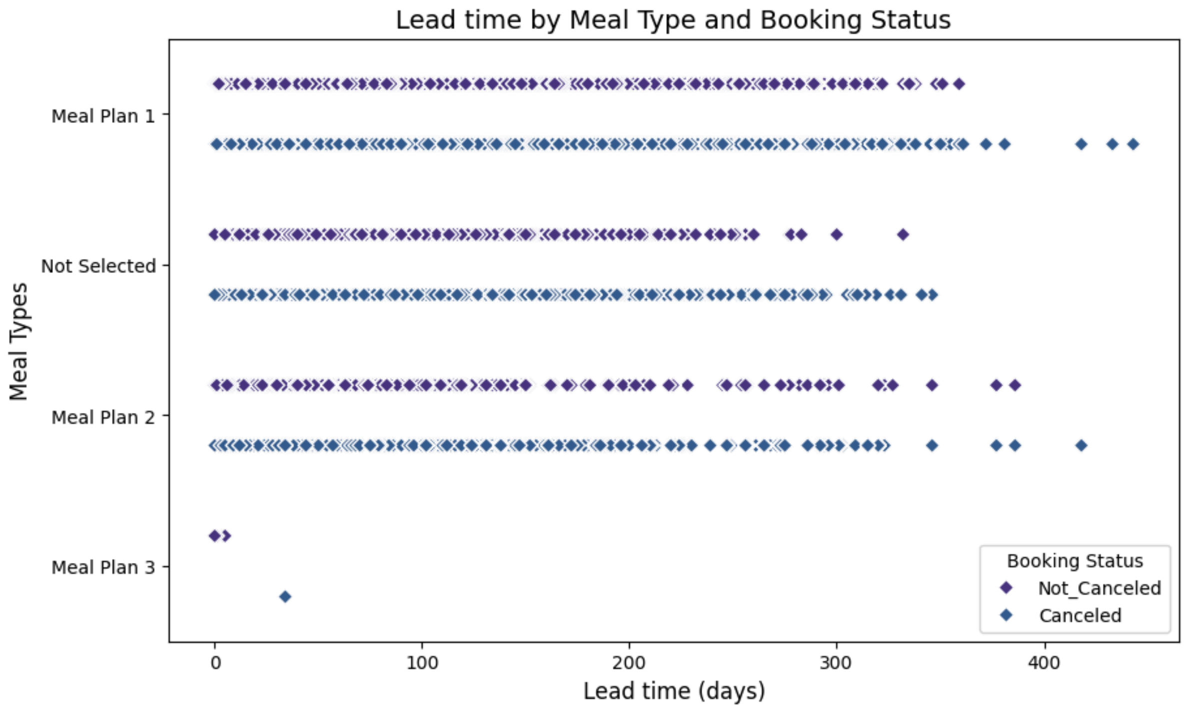
### 1) The Relationship Between Customer Type, Lead Time, and Booking Status:

This chart shows four customer types: family, solo, group, and couple. According to the chart, the lead time is longer for cancelled bookings across all four groups. This difference is especially noticeable in solo and couple groups, indicating a higher likelihood of plans changing if more time remains until the check-in date for these travellers.



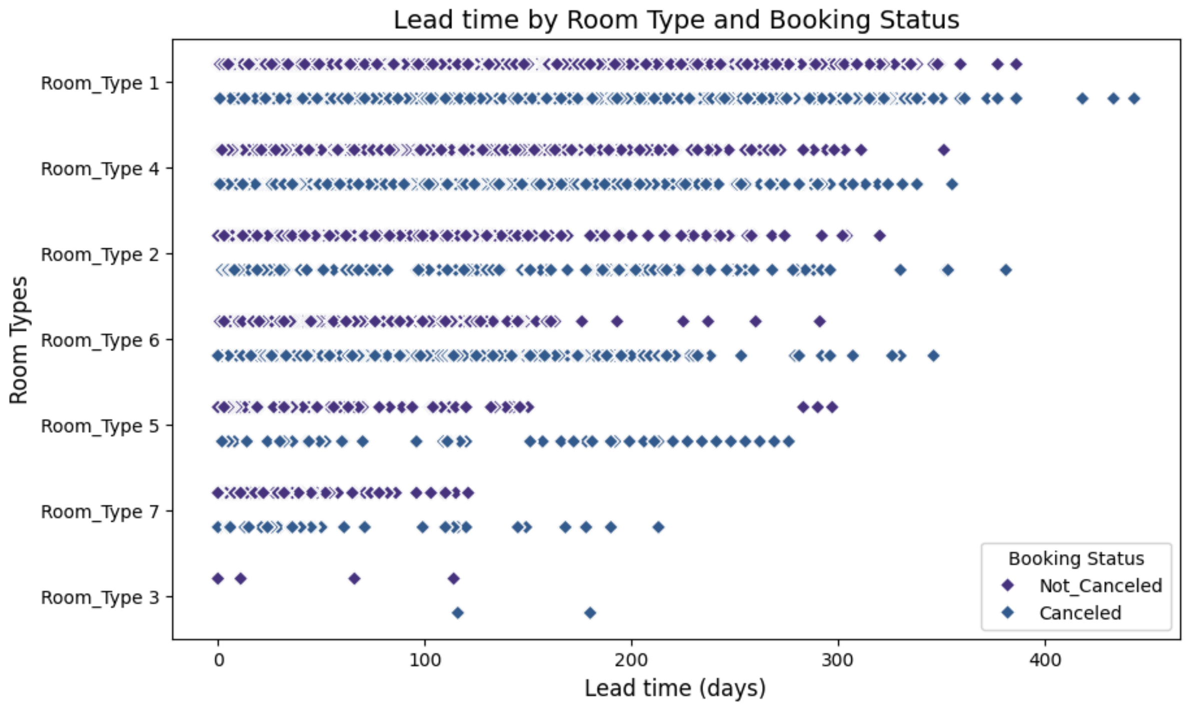

**2) The Relationship Between Meal Type, Lead Time, and Status:**

The chart indicates that lead time is longer for cancelled bookings across all four meal groups. Among them, the lead time for cancelled bookings is greatest in Meal Type 1, suggesting that having more time before the check-in date increases the likelihood of cancellations in this group.



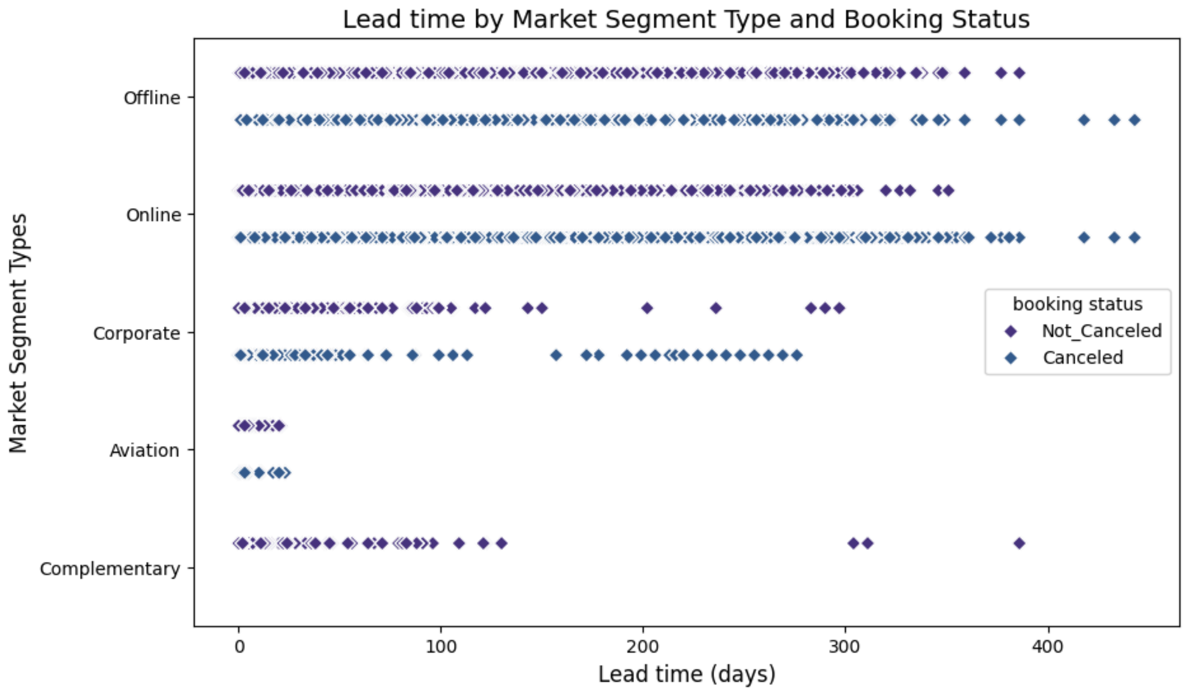

### 3) **The Relationship Between Room Type, Lead Time, and Booking Status:**

According to this chart, the lead time for cancelled bookings is longer than for not-cancelled bookings across all room types. Room Type 1 has the longest waiting period among room types. Additionally, the gap between the lead time for cancelled and not-cancelled bookings is most pronounced in Room Type 7, which could be due to factors like higher cost or the need for more planning. Interestingly, there's a noticeable gap in the booking-to-arrival duration for not-cancelled bookings in Room Type 5. Most travellers book Room Type 5 up to 150 days in advance, but a few make reservations up to 300 days early.



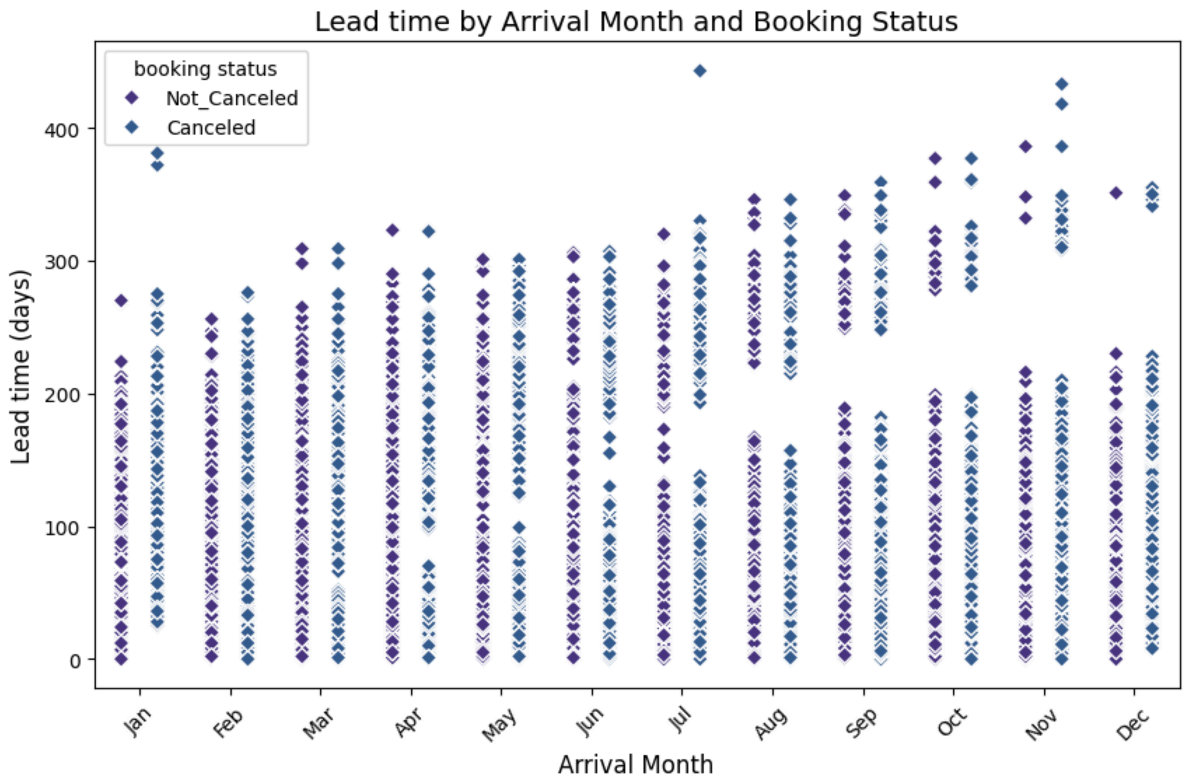

4) **The Relationship Between Market Segment Type, Lead Time, and Booking Status:**

Unlike other features, this section shows that for corporate bookings, the lead time is shorter for cancelled bookings compared to not-cancelled ones. This may be due to quicker decision-making in corporate bookings. Additionally, none of the complementary bookings were cancelled, suggesting that free or discounted stays motivate guests to avoid cancellations.



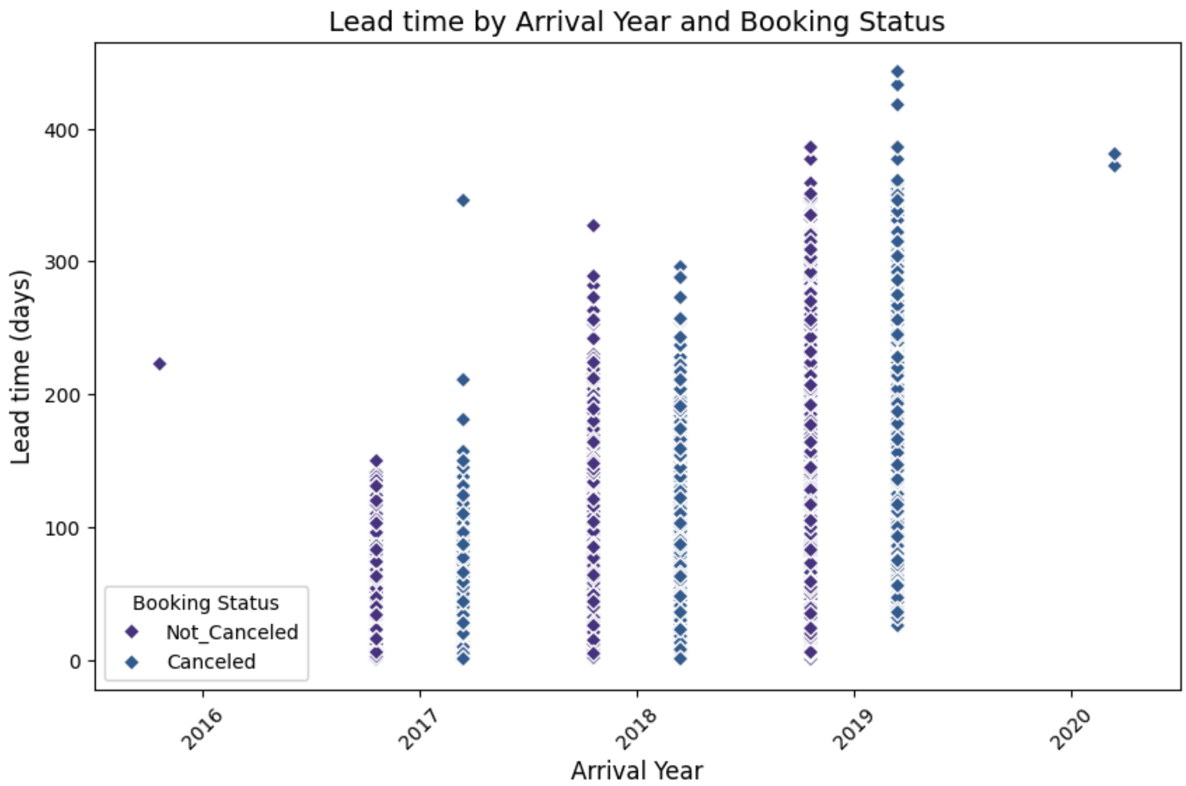

5) **The Relationship Between Entry Month, Lead Time, and Booking Status:**

The chart examining the relationship between entry month and booking status shows that the lead time is roughly the same across months for cancelled and non-canceled bookings, but the duration increases in the final months of the year. From April onward, a time gap is noticeable in the data, which gradually increases and peaks in December. This increase may be due to higher demand and longer-term planning during the holiday season.




6) **The Relationship Between Entry Year, Duration Between Booking and Arrival, and Status:**

The chart shows that the booking-to-arrival duration in 2019 increased significantly compared to the previous two years. This trend may be related to inflation and rising accommodation costs in 2019. In 2018, unlike 2017 and 2019, the booking-to-arrival duration was shorter for cancelled bookings, which may be due to temporary economic or customer behaviour changes during that year.


Lead time by Arrival Year and Booking Status
Lead time (days)
400
300
200
100
0
Booking Status
Not_Canceled
Canceled
2016
2017
2018
2019
2020
Arrival Year