Group 2

Cellula Internship

Group 2

Hotel Exploratory Data Analysis

Supervised by:
Eng. Mohamed Alghani

Team Names:
Eng. Waleed Shaban (Team Leader)
Eng. Omar Nezar
Eng. Mustafa Kamel
Eng. Mazen Ahmed
Eng. Ahmed Abdelaziz
Eng. Abdelrahman Farraag
Eng. Malak Amgad

OUR AGENDA

01

INTRODUCTION

02

Data Processing and Visualization

03

Conclusion


01
INTRODUCTION

Grand Hotel


02
Data Processing
and Visualization

Data Overview

Data loading and reading

Basic data exploration

	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
5	INN00006	1	0	0	2	Meal Plan 2	0	Room_Type 1	346	Offline	0	0	0	100.00	1	9/13/2016	Canceled

Data Overview

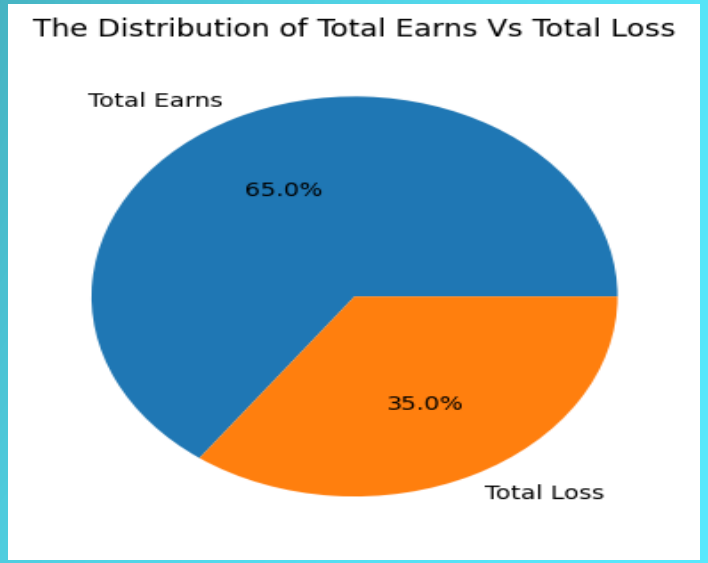
-The data structure showed 17 features and 36285 entries between 2017-07 and 2018-12
-10 int , 1 float , 6 strings

Features: 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 ,repeated , P-C ,P-not-C,average price ,special requests , date of reservation, booking status

The First Thing I Want to Share with You
We have both good and bad news for you.

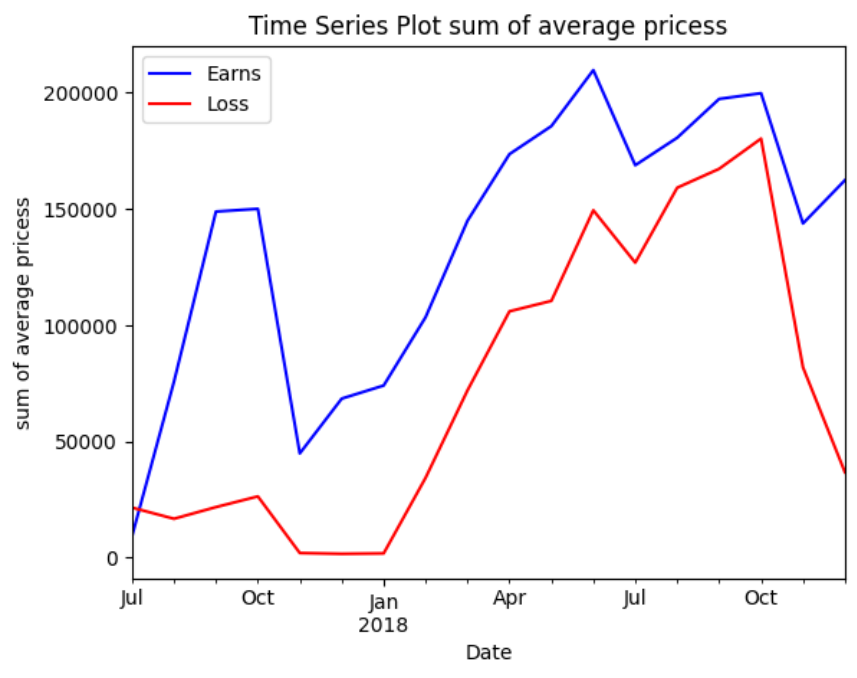
You have achieved

- **Total Earns: $2.4 M**
- **Total Loss: $1.3 M**

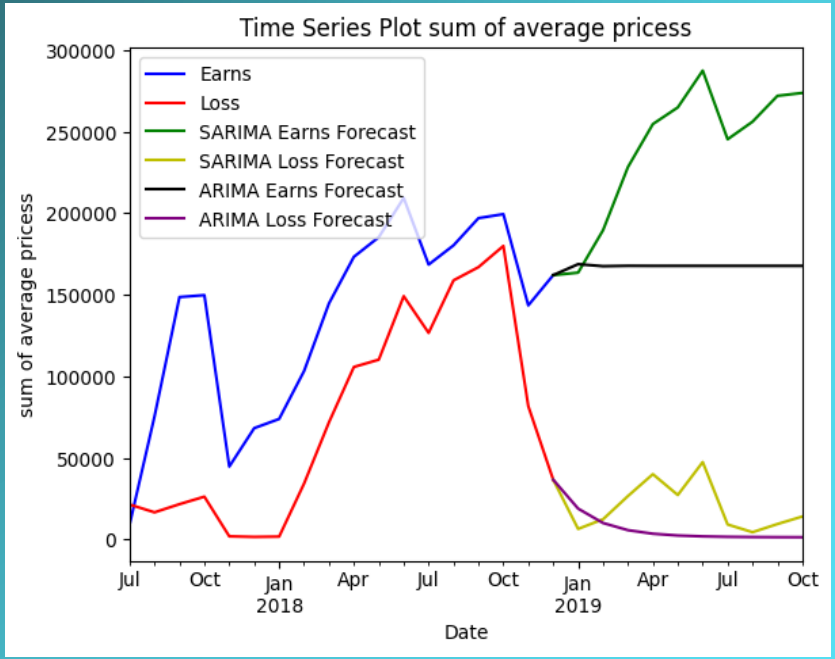


As you’ve seen before, you earned $2.4M,
but also lost $1.3M due to canceled reservations.
This could have increased your earnings by approximately 53%.
It is crucial to investigate the reasons behind these cancellations
to improve future performance.

Also, I want to highlight:
You have achieved high earnings over the past two months and significantly reduced losses from canceled reservations, as shown below.

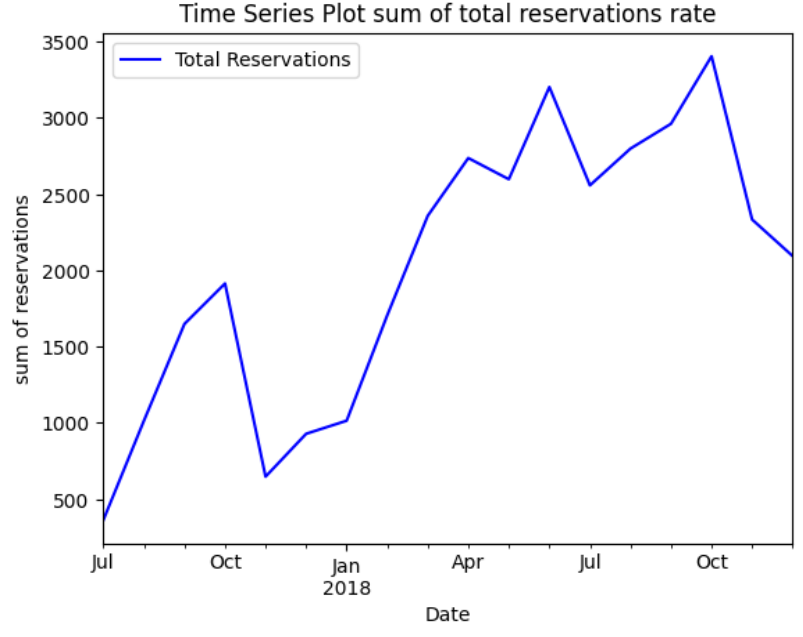


Additionally, using time series forecasting, the models predict that you will achieve higher earnings in the upcoming 10 months, as shown below. However, the amount of training data is limited, so this prediction comes with some uncertainty.



You might think this is good news because the cancellation rate is decreasing, but that’s not the case. In reality, you’re becoming less popular and losing guests.

This is because the canceled reservations are not converting into completed bookings, and guests are losing interest in the hotel.
As shown below, the total number of hotel reservations has decreased over the past two months, from 3.5k reservations to 2k.
This indicates that the hotel is becoming less popular and losing guests.



Therefore, it is crucial to understand why the hotel is becoming less popular and losing guests. Identifying these reasons will help you improve popularity and increase earnings.

The good news is that we've identified a problem, and we can work on solving it. The first step in problem-solving is recognizing that there is an issue and understanding it.

So, let's dive into the data and discover the underlying problems.

Descriptive Statistics

Average Booking Profile:

- **Adults per Booking:** 1.84
- **Children per Booking:** 0.11
- **Weekend Nights:** 0.81
- **Week Nights:** 2.20
- **Total Guests:** 1.95
- **Special Requests:** 0.62

Booking Price Insights:

- **Average Price per Booking:** $103.42
- **Price per Night:** $47.15
- **Price per Night per Guest:** $27.30

Descriptive Statistics

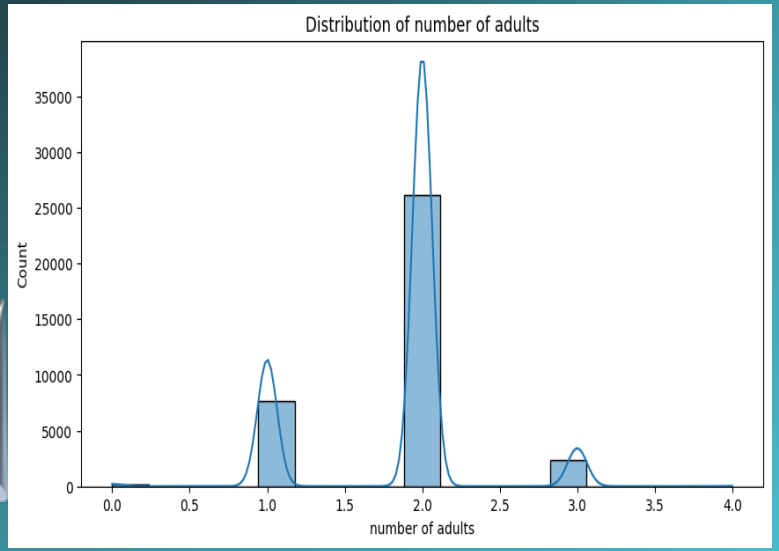
Lead Time:

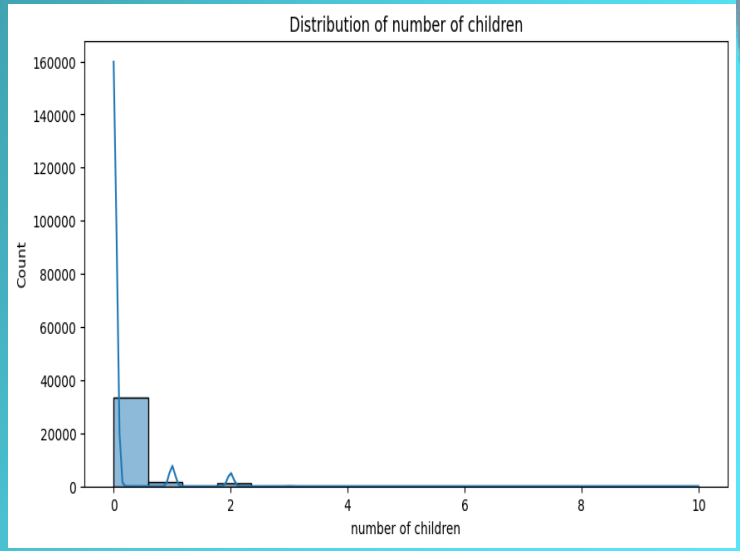
- Average lead time is **85 days**, with bookings made as early as 443 days in advance.

Cancellation Rates:

- **15%** of bookings were not canceled, while **2%** were canceled.
- **Car Parking:**
- Only **3%** of bookings included a parking space.
- **Booking Duration:**
- Most bookings are between **2-4 nights**.

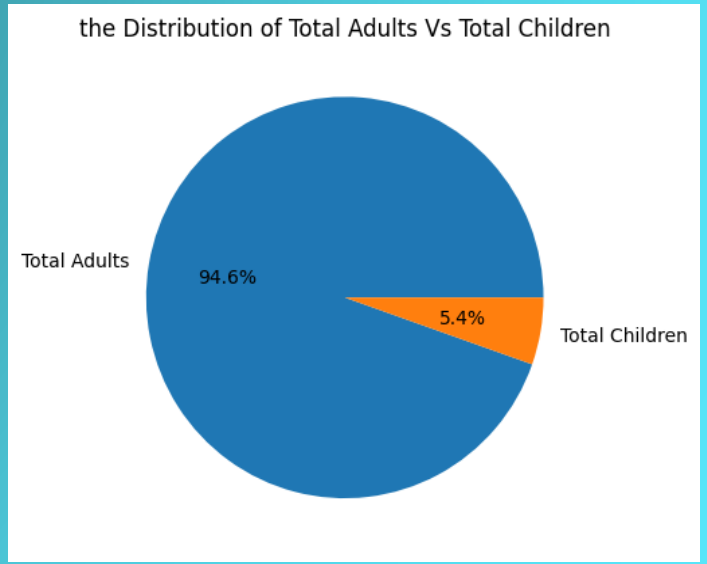
Distribution of Adults Vs Children





Distribution of Adults Vs Children

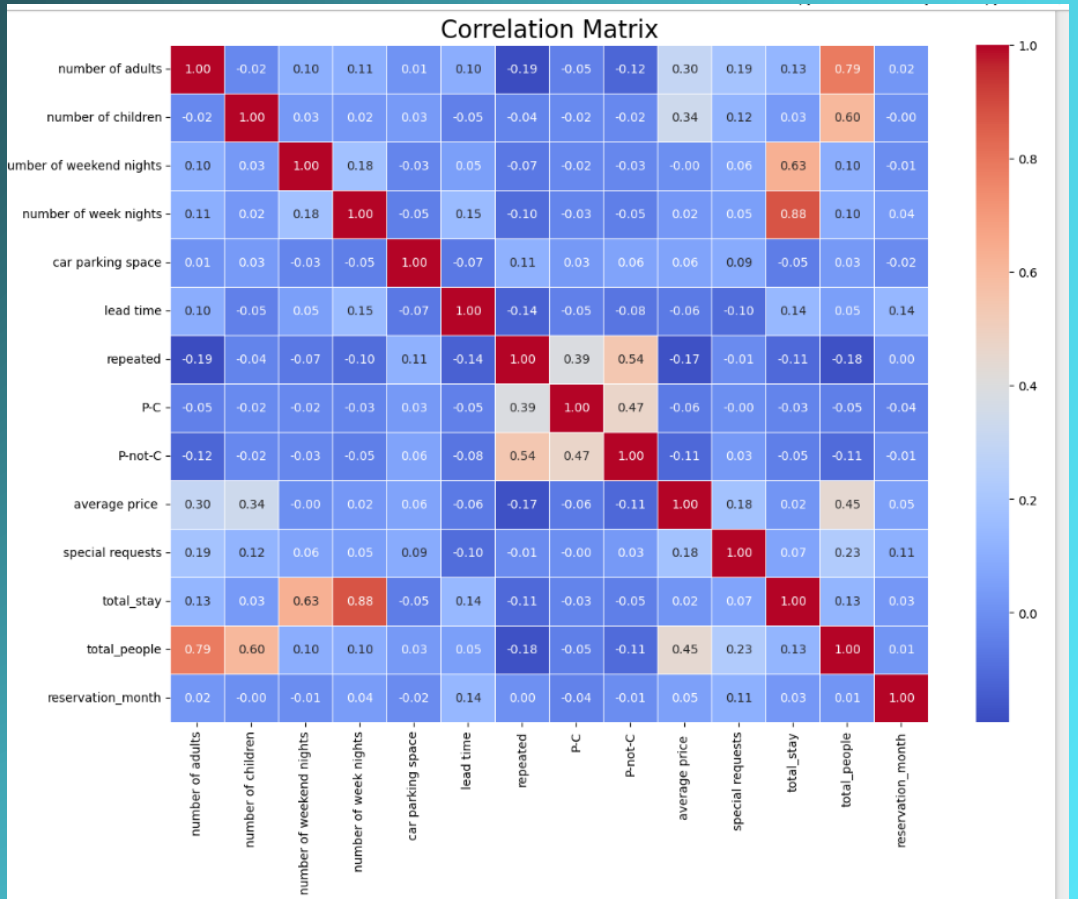
- **The total guests of the hotel is:** 70763
- **The total children of the hotel is:** 3823
- **The total adults of the hotel is:** 66940
- **The total children percent of the hotel is:** 5.4%
- **The total adults percent of the hotel is:** 9



As observed from the descriptive statistics and the plots, the average number of adults per booking is about 2, indicating that most of your guests are couples. You can leverage this insight to create customized special offers for couples, potentially increasing your popularity.

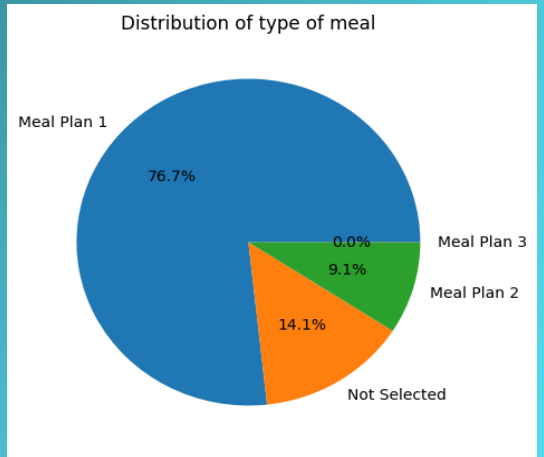
However, the data also shows that the number of children per booking is almost zero. This suggests that guests with children are not interested in your hotel. To address this, consider adding more family-friendly services and amenities to attract these guests.

Correlation Matrix



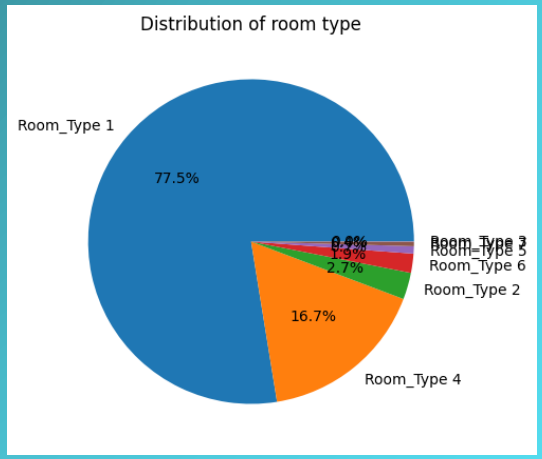
Distribution of meal plans

From the distribution of meal plans, we can see that most guests prefer Meal Plan 1 over the others, while Meal Plan 3 is almost unused. This suggests there might be an issue with Meal Plan 3. We should consider redesigning it into a more appealing option that offers guests greater variety, which could enhance our popularity.

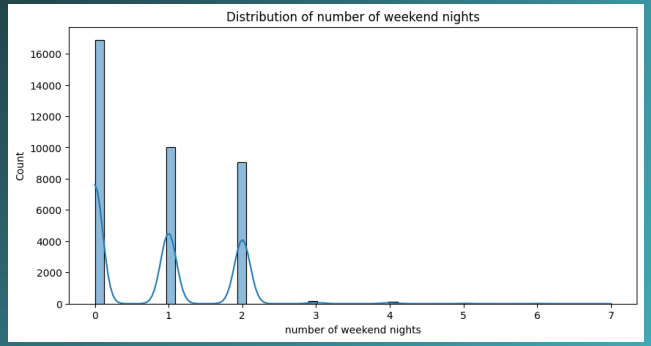


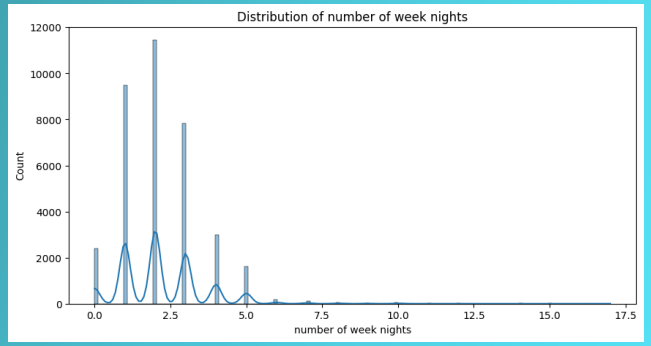
Distribution of room type

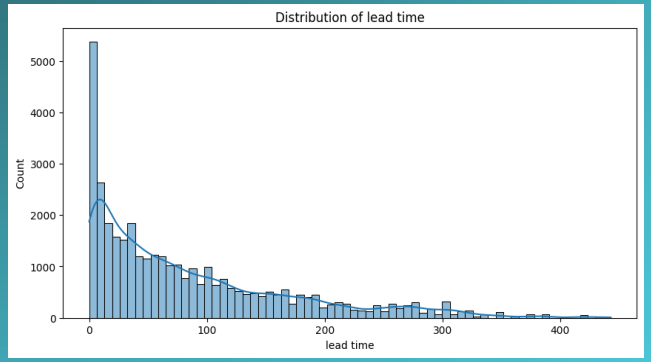
From the distribution of room types, we observe that most guests prefer Room Type 1, while Rooms Type 2, 3, 4, and 5 account for less than 5% and are almost unused. This suggests there might be an issue with these less popular room types. However, it's challenging to definitively identify the problem due to insufficient data. The imbalance could be due to a limited number of available rooms for these types. Therefore, we need additional data on the count of each room type to make a clearer judgment.

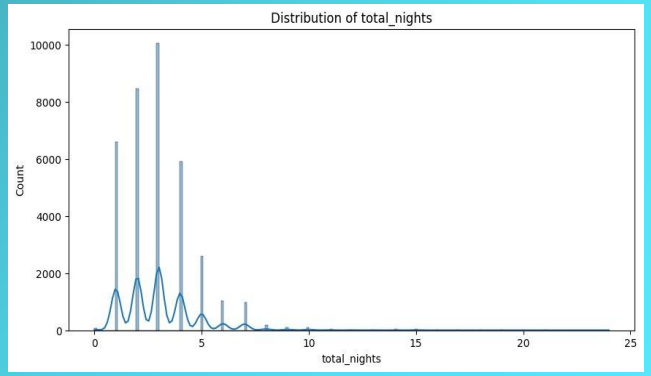


Data Distributions

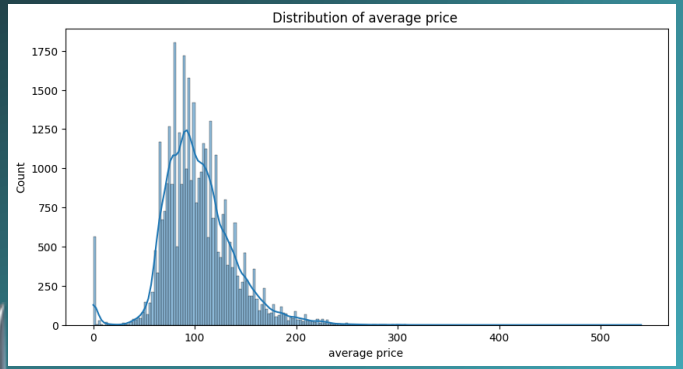


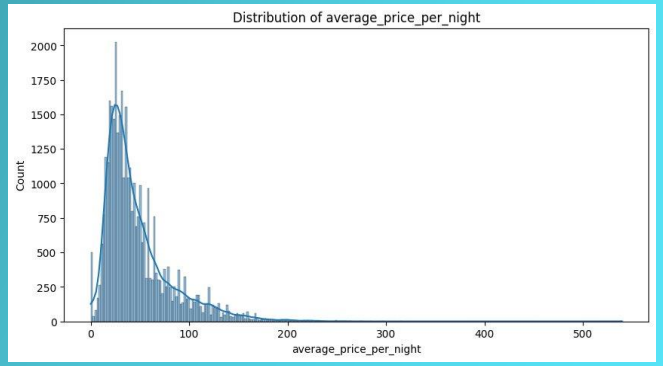


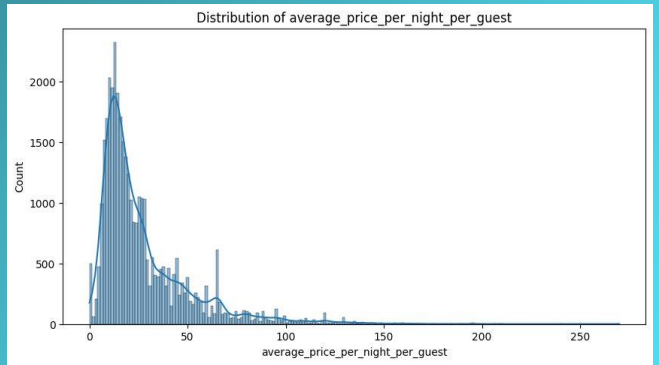




Prices Distribution

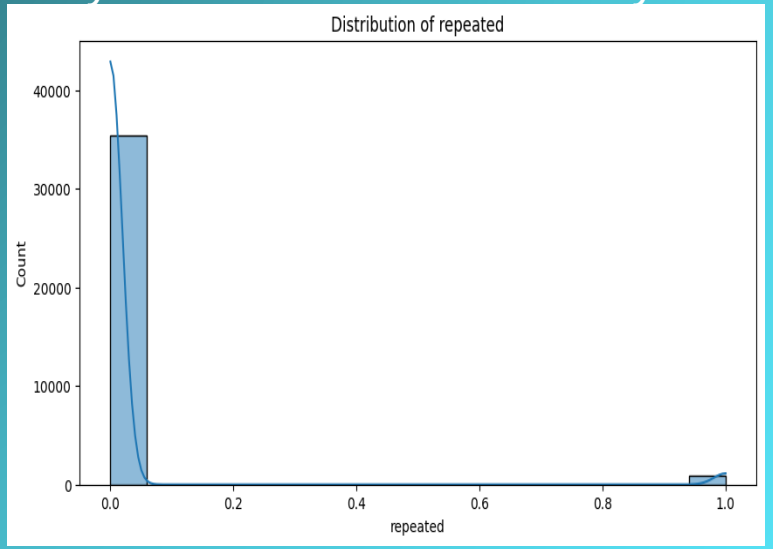




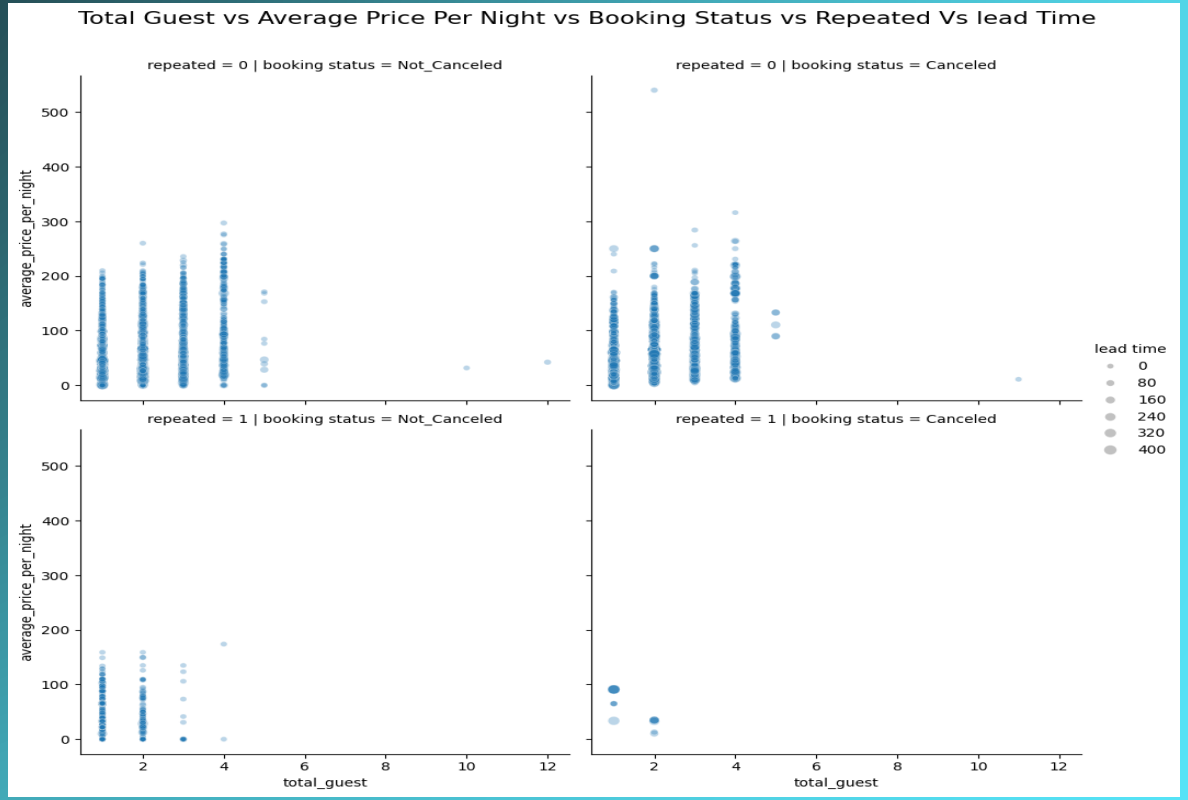


Distribution of repeated

From the plot below, we can see that the number of repeated reservations is relatively low compared to the total reservations. However, we cannot make a definitive judgment based solely on this data. It is important to consider when the hotel started serving guests, as the low number of repeat reservations might be due to the hotel being recently established and currently attracting many new guests.

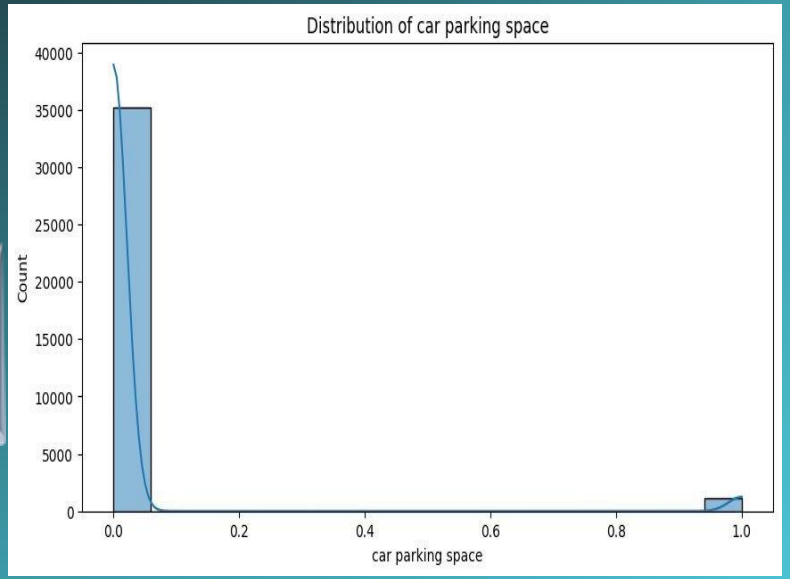


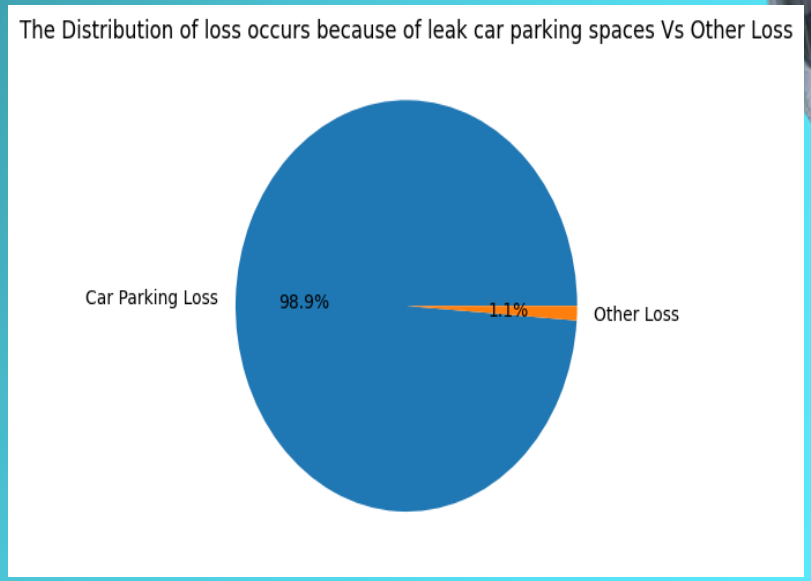
Distribution of Repeated vs Booking Status



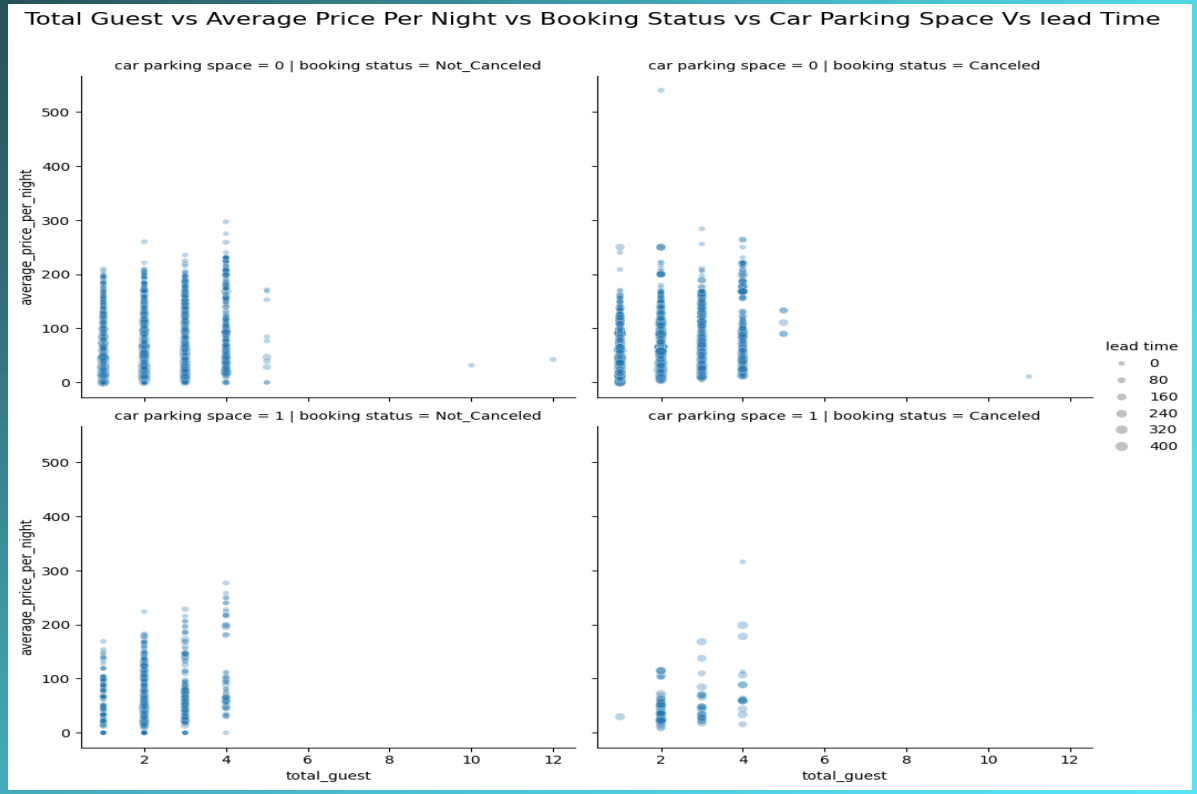
From the previous plots, we can see that repeated reservations are almost complete, but there are still some cancellations. We need to investigate the reasons for these cancellations, as they are causing us to lose loyal guests.

Distribution of Car Parking Space Vs Car Parking Space Loss





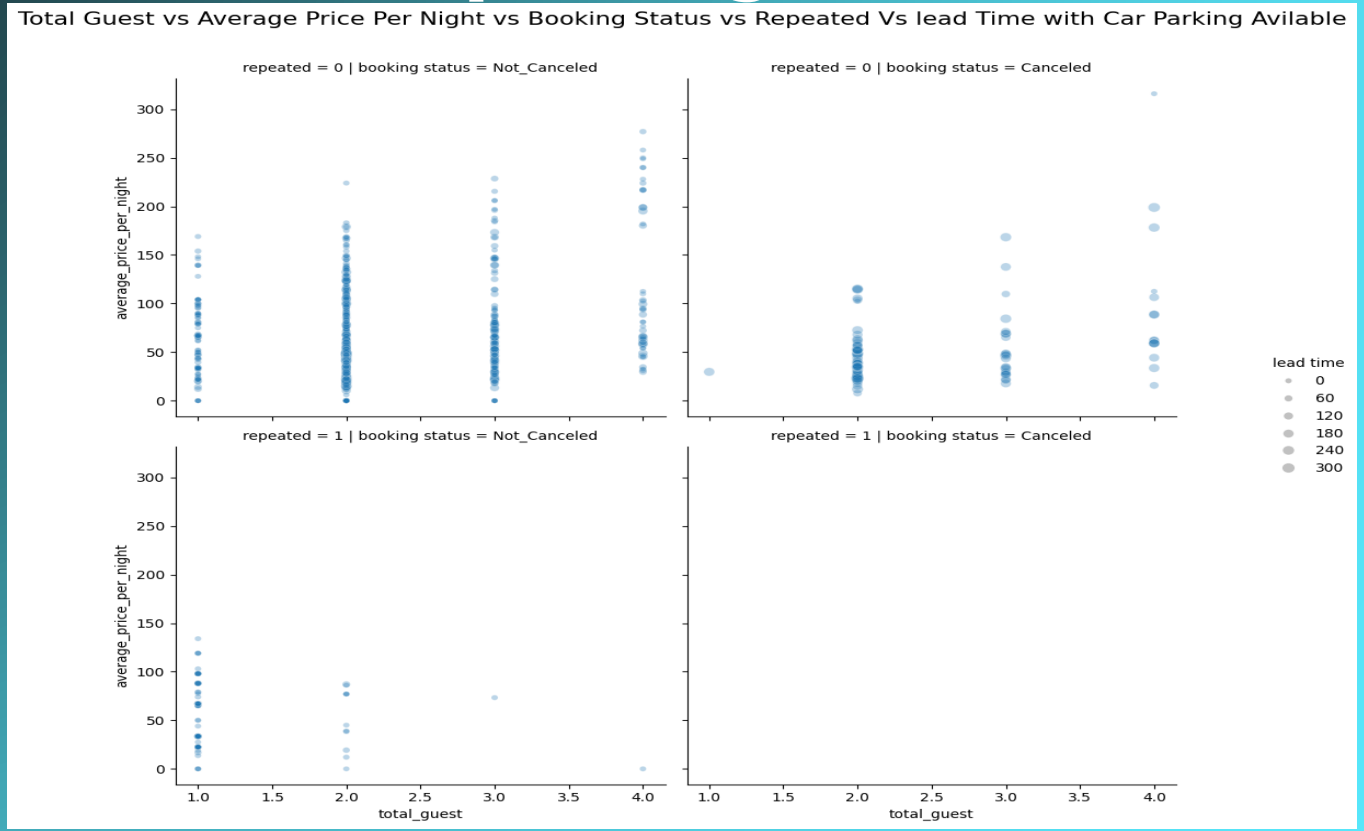
Distribution of Car parking Space Vs Booking Status



From the previous plots, we can see that the main cause of most cancellations is the lack of parking space, with 98.9% of canceled reservations attributed to this issue. In the next slide, we will observe that this is also the cause of all cancellations for repeated reservations, making it a significant factor in guest loss.

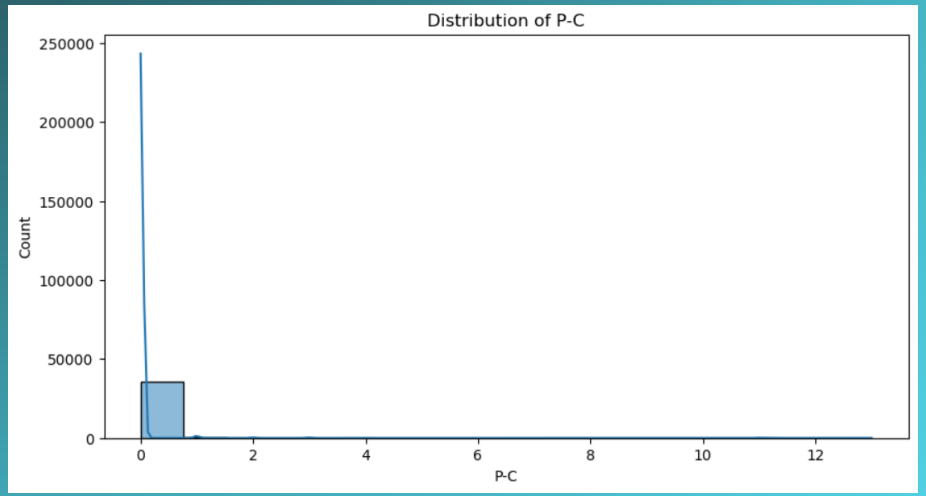
It is crucial to address this problem by providing more parking spaces to prevent further cancellations and retain guests.

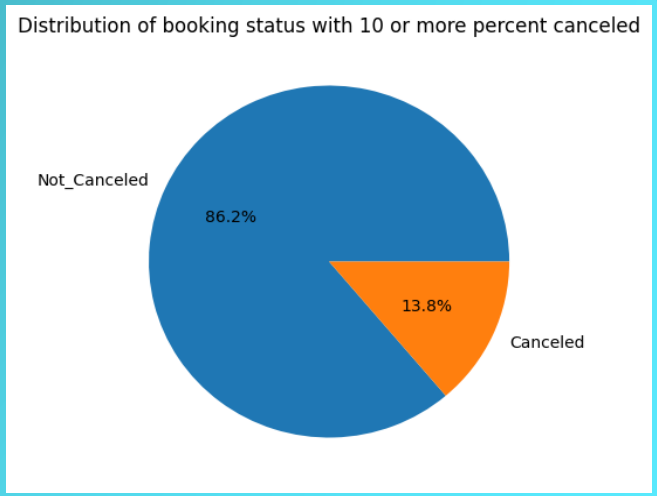
Distribution of Repeated Vs Booking Status when carparking is Available



Probability of cancelation

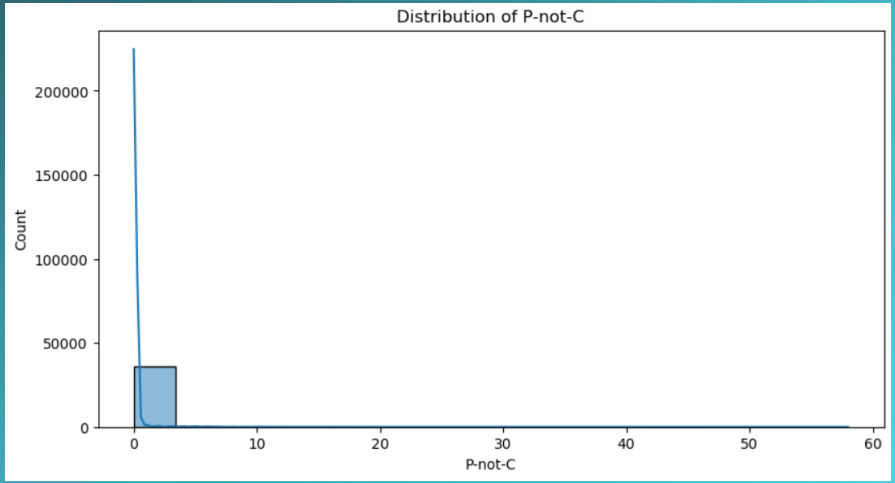
As we can see, the probability of cancellation shows a little significant relationship with the booking status. However, the data is highly skewed, with most values being zero. The plot on the right is based on 29 data points, representing all cases where the cancellation probability is 10% or higher.

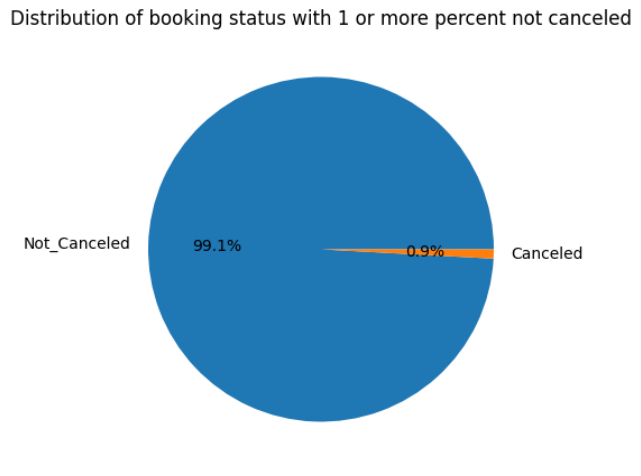




Probability of not cancelation

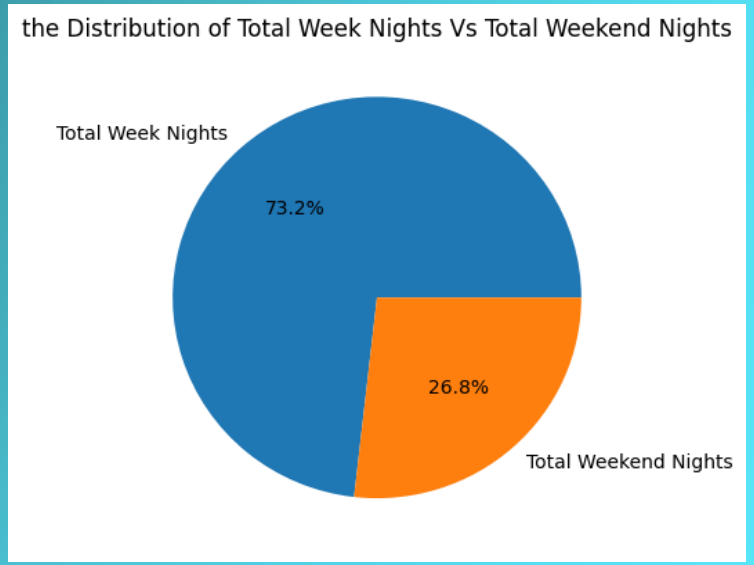
As we can see, the probability of non-cancellation shows a strong significant relationship with the booking status. However, the data is also highly skewed, with most values being zero. The plot on the right is based on 812 data points, representing all cases where the non-cancellation probability is 1% or higher.





Distribution of Week Nights Vs Weekend Nights

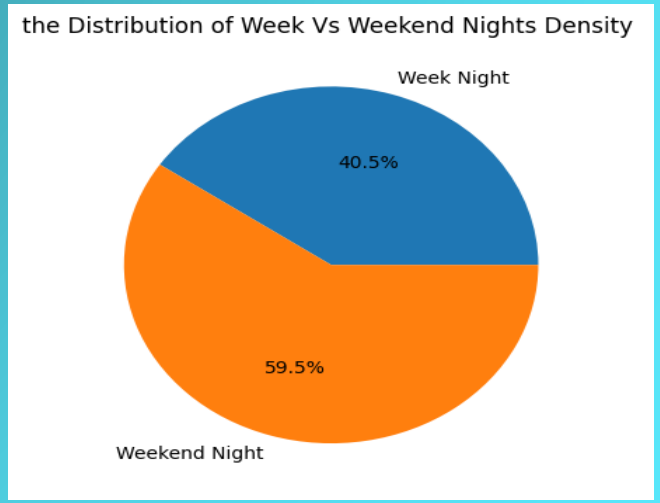
- **The total nights of the hotel is:** 70419
- **The total weekend nights of the hotel is:** 18866
- **The total week nights of the hotel is:** 51553
- **The weekend nights percent of the hotel is:** 26.79%
- **The week nights percent of the hotel is:** 73.21



Distribution of Week Nights Vs Weekend Nights Density

As we can see in the plots below, the guest density is higher on weekend nights compared to weeknights. It is important to effectively manage this increased density to ensure smooth operations and guest satisfaction.

- **The total weekend nights guests of the hotel is:** 39044
- **The total week nights guests of the hotel is:** 66419
- **The weekend nights guests density of the hotel is:** 19522.0
- **The week nights guests density the hotel is:** 13283



Conclusion

From the analysis of the data, several key insights have emerged that can guide the hotel in improving its operations and guest satisfaction.

Firstly, the majority of guests are couples, and tailoring special offers for them could boost popularity. However, the lack of family-friendly services seems to be a barrier for guests with children, and addressing this could attract more family-oriented visitors.

Further, the distribution of meal plans reveals that Meal Plan 1 is favored by most guests, while Meal Plan 3 is largely unused. Redesigning Meal Plan 3 to offer more appealing options may increase its usage.

Similarly, the underutilization of certain room types suggests potential issues that should be investigated, though more data on room availability is needed to make a definitive judgment.

One of the most critical findings is the high cancellation rate due to a lack of parking space. With **98.9%** of cancellations linked to this issue, it is vital to increase parking capacity to retain guests and prevent further losses. Additionally, while the number of repeated reservations is low, this could be attributed to the hotel's recent establishment, and improving services to loyal guests should remain a priority.

Lastly, the data reveals that guest density is significantly higher on weekend nights, suggesting a need to manage these peak periods efficiently to maintain service quality. By addressing these identified issues, the hotel can enhance its reputation, retain guests, and increase earnings.

TO DO List

Must to solve : leak of car parking spaces.

Should to solve: more improvement for family-friendly services, meal plan redesign.

Nice to solve: tailoring special offers for couples, provide detailed data on the counts per room type, weekend nights guest density.


Thank You!