

Hotel EDA

Hotel Exploratory Data Analysis:
Focused on Investigating
Cancelled Reservations



Questioned Asked



Does an increase in Average Daily Cost affect cancellations?

Does prior guest behavior affect cancellations?

Does the type of hotel affect cancellations ?

Does the guest's country affect cancellations?

Does Market Segment effect cancellations?

Recommendations

- To prevent cancellations, hotels should focus on the key segments with the highest rates. For Online Travel Agencies (OTA), implement stricter cancellation policies and offer incentives such as discounts or loyalty points for completed stays. For group bookings, introduce flexible options and enhance communication with group coordinators, along with special deals for non-refundable rates. Strengthen relationships with offline travel agencies by offering exclusive deals and training to encourage commitment. For direct bookings, provide special offers and loyalty programs to attract repeat customers. These strategies can help reduce overall cancellation rates.
- To increase the number of repeated customers, hotels should implement a robust loyalty program that offers significant rewards for repeat stays. These rewards could include discounts on future bookings, free room upgrades, and exclusive amenities. Additionally, personalized communication and offers tailored to previous guests can make them feel valued and encourage them to return.
- For countries with high cancellation rates, such as Portugal (PRT), it is essential to understand the reasons behind the cancellations. Conduct surveys or gather feedback from customers who cancel to identify common issues. Based on the findings, hotels can implement targeted strategies such as flexible booking options, improved booking policies, and localized promotions. Additionally, partnering with local travel agencies to offer attractive packages and ensuring clear communication about cancellation policies can help reduce the high cancellation rate.

Additional Recommendations: Lead Time and booking window

- Restrict very long lead times between booking and the arrival date.
- Consider limiting the arrival date to a maximum of one year, and those who try to book more than a year in advance should receive an error.
- Rather than having customers book a year or more in advance, they could like the hotel and set a reminder for when they want to book the hotel within a year of their expected arrival date.
- This will allow for fewer cancellations due to unknowns that might arise from booking too far in advance. Lead time was one of the features that showed an increase in cancellations as the lead time from the booking date to the arrival date increased.

Additional Recommendations: Promote special request

- According to the correlation matrix special request resulted in a negative correlation.
- Customers who requested more resulted in lower cancelations compared that didn't have special request or only had a few request.

Additional Recommendations: Phase out no deposit

- Non-Deposit resulted the most bookings yet had the most cancelations compared to refundable deposit and non refund.
- Hotels could reduce cancellations by only allowing customers to choose between a refundable deposit and a non-refundable option, while making the refundable deposit a small, affordable amount.

Additional Recommendations: City Hotels

- City hotels resulted in higher bookings compared to resort hotels yet had the highest cancellations.
- Suggest offering more entertainment at or around city hotels to encourage customers not to cancel.
- Some guests might book a city hotel because of a work event or another organized event. If the event gets canceled, they might be prone to cancel their reservations. However, if they know about other entertainment options around, they might be influenced to make it a personal trip to see the museum, zoo, and other attractions.
- Hotels could partner with local organizations to host events like sing nights and other activities to influence those who might cancel because their original event got canceled.

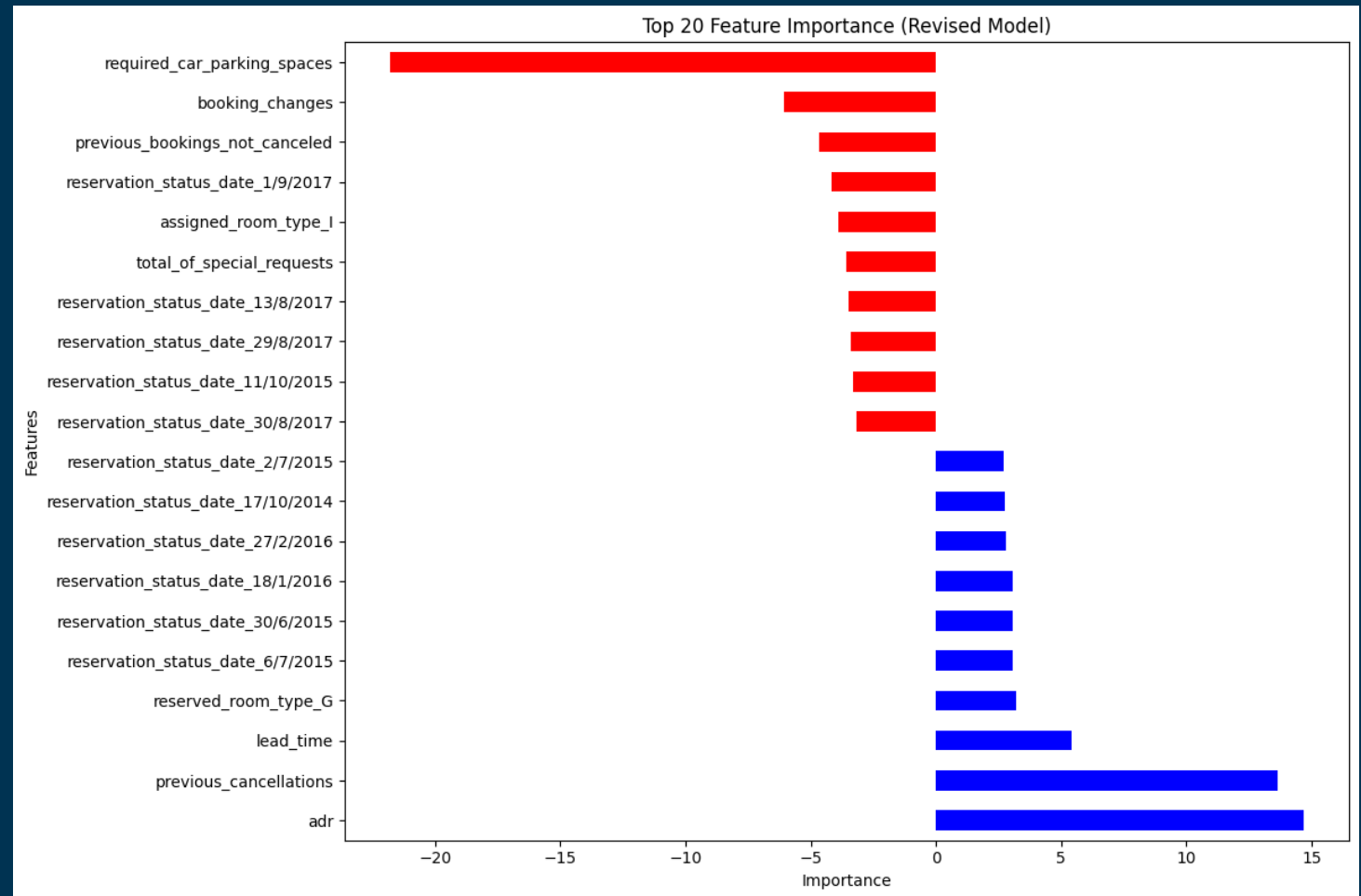
Additional Recommendations: Resorts

- Resorts had lower bookings and lower cancelations
- Around 8 percent of those who booked a reservation, regardless of hotel type, had children. Resorts could leverage this by promoting a family-friendly environment to increase bookings.

Logistic Regress Model: Top Features

Feature Importance:

adr	14.676246
previous_cancellations	13.649661
lead_time	5.441188
reserved_room_type_G	3.185264
reservation_status_date_6/7/2015	3.079710
...	
assigned_room_type_I	-3.900257
reservation_status_date_1/9/2017	-4.157186
previous_bookings_not_canceled	-4.645348
booking_changes	-6.070533
required_car_parking_spaces	-21.785546
Length: 1153, dtype: float64	

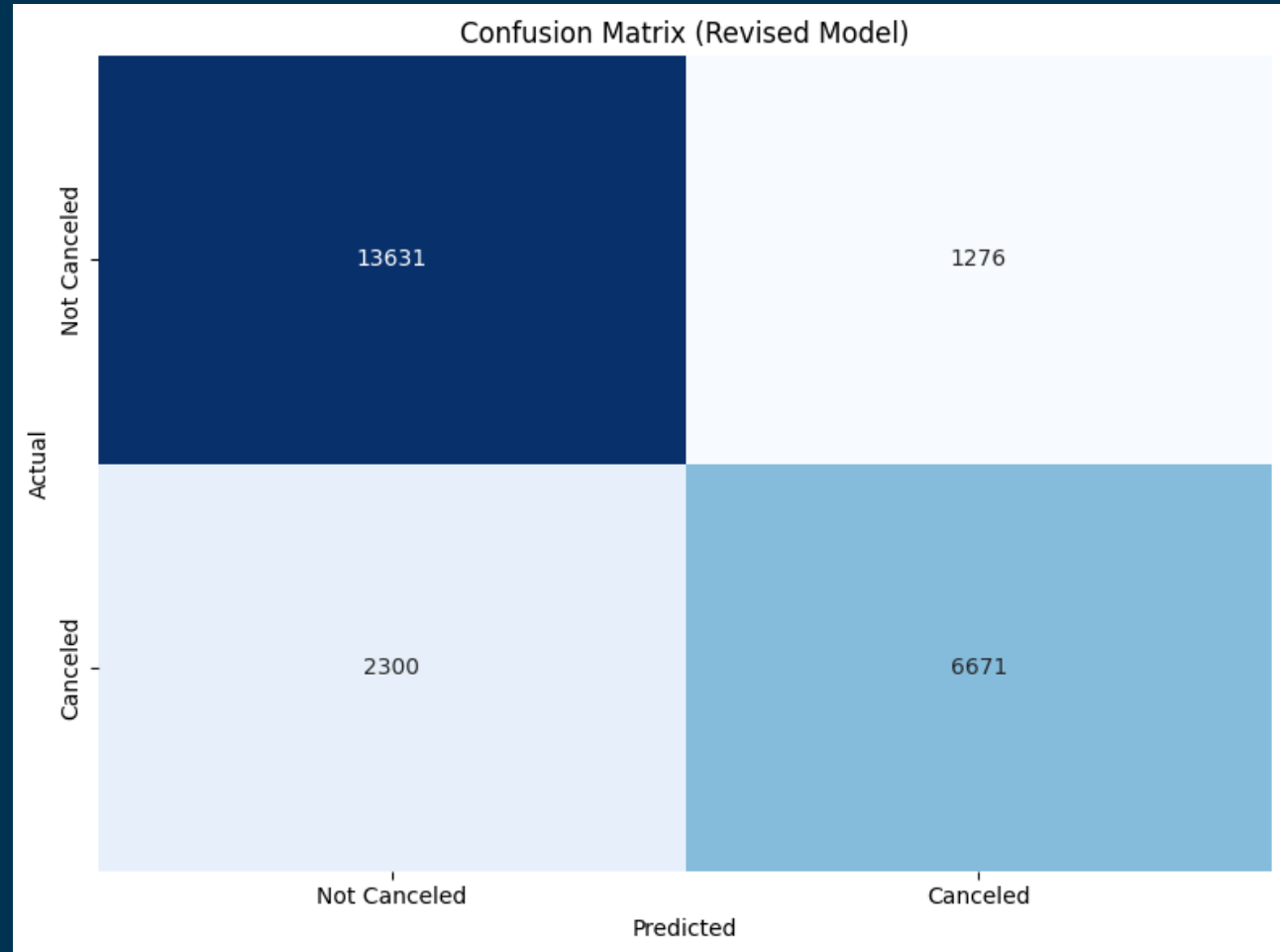


Logistic Regress Revised model:

Accuracy score : 85 %

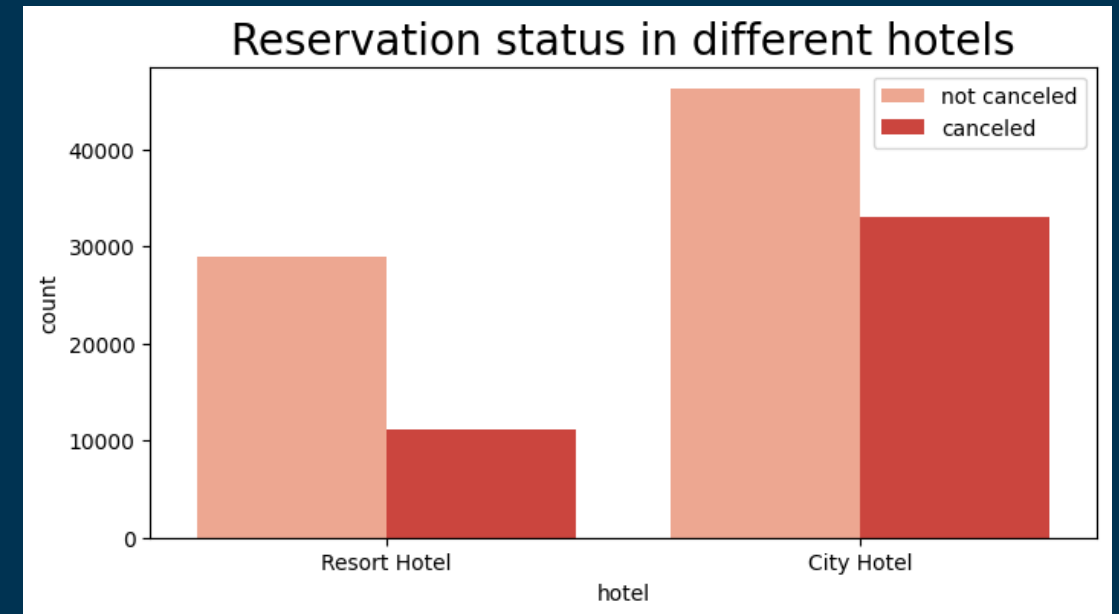
Logistic Regress Confusion Matrix - Revised model:

- True Positives (TP): 6671 bookings were correctly predicted as canceled.
- True Negatives (TN): 13631 bookings were correctly predicted as not canceled.
- False Positives (FP): 1276 bookings were incorrectly predicted as canceled but were not canceled.
- False Negatives (FN): 2300 bookings were incorrectly predicted as not canceled but were canceled.



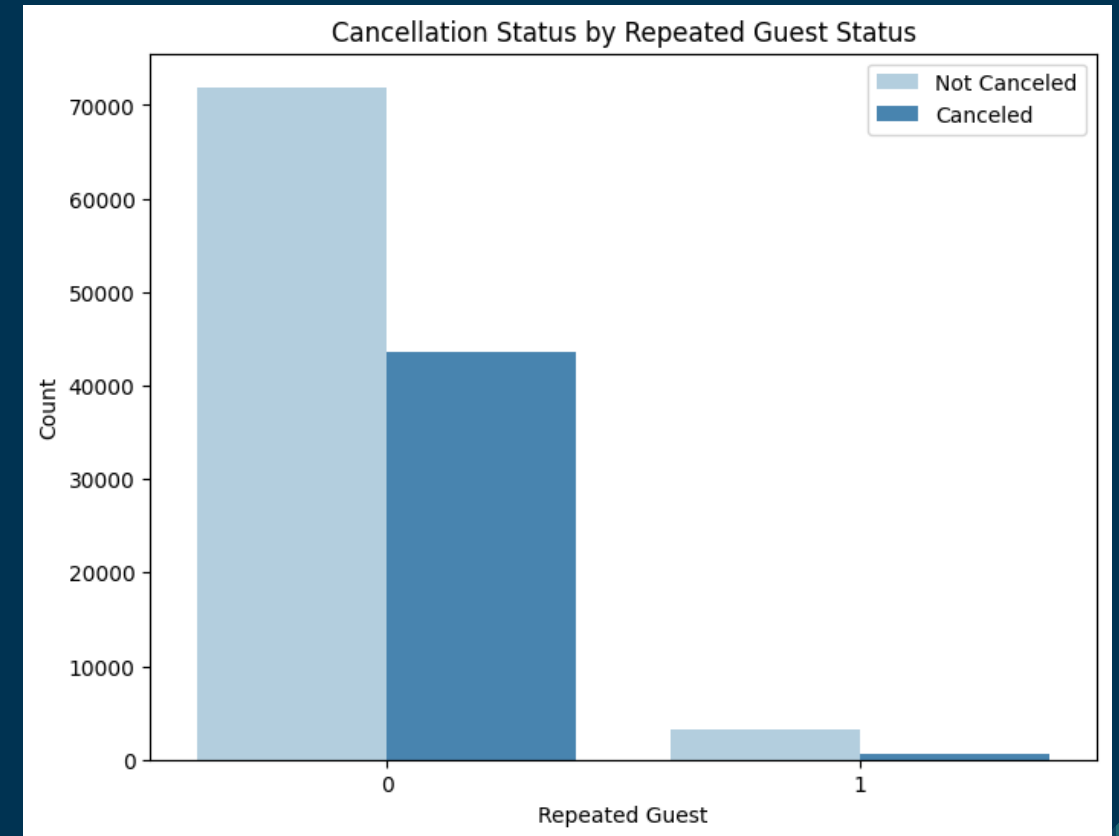
Reservation Status by Hotel

- The following side-by-side bar graph shows the percentage of reservations that are canceled and not canceled.
- The graph shows a significant number of reservations that have been canceled.
- The total cancellation rate is 37 percent, which is a significant loss for the hotel's earnings.
- Resort Hotel had a cancellation rate of 28 percent.
- City Hotel had a significant impact of cancellations with 42 percent.
- It seems the hotel type does play a role in cancellation reservations.



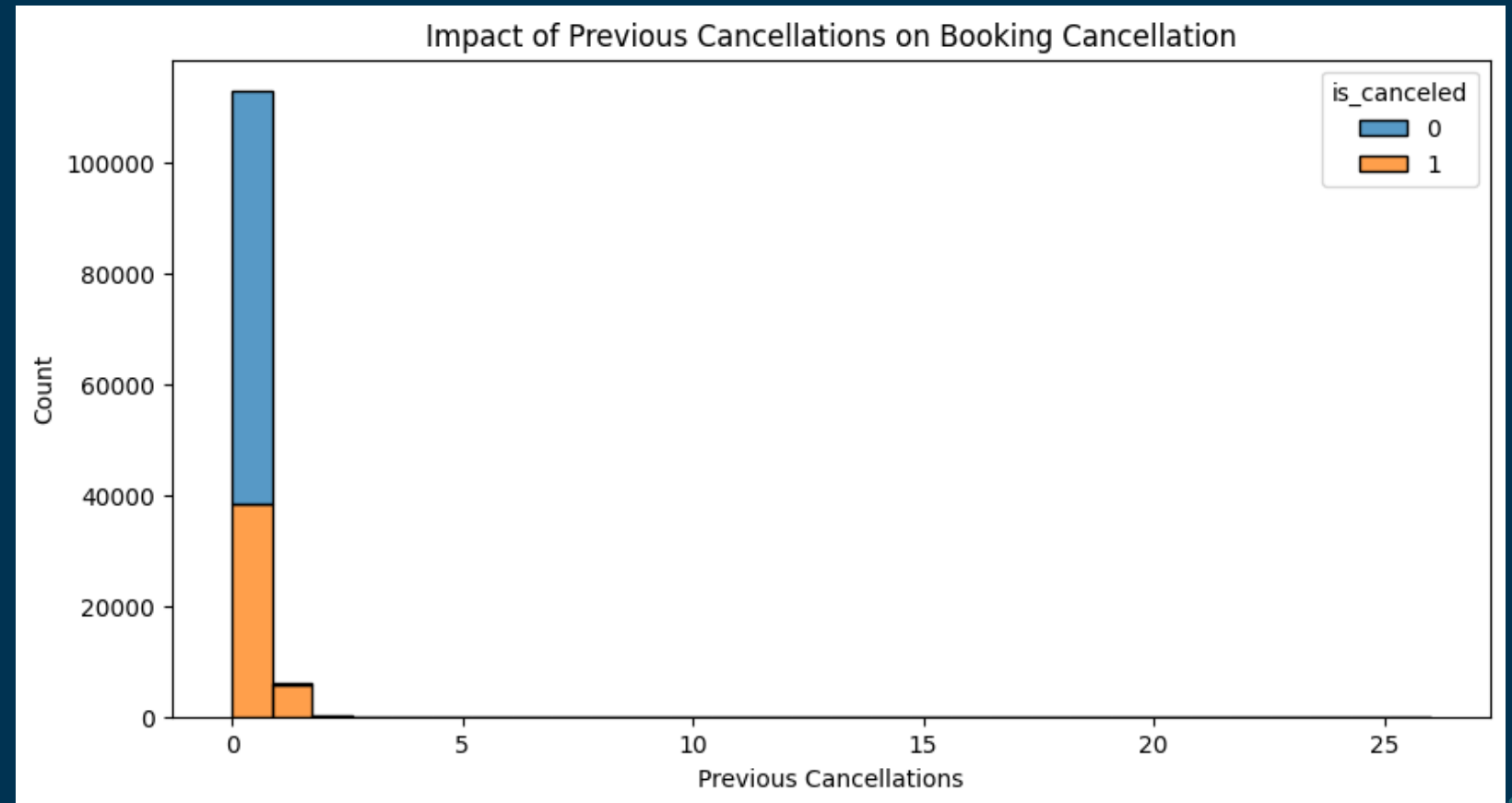
Repeated Guest vs non-Repeated

- There are far fewer repeated guests overall, with a minimal number of cancellations among them.
- The majority of cancellations are among non-repeated guests, with a significant number of both canceled and not canceled reservations
- Repeated guests have a much lower cancellation rate compared to non-repeated guests, indicating that repeat customers are more likely to follow through with their reservations.
- Insights suggest that hotels might benefit from focusing on customer retention and loyalty programs to reduce cancellation rates



No Previous Cancellations: The vast majority of bookings have no previous cancellations, and a substantial portion of these are still canceled.

Higher Previous Cancellations: Customers with a higher number of previous cancellations tend to cancel more frequently, indicating a trend where past behavior predicts future cancellations.



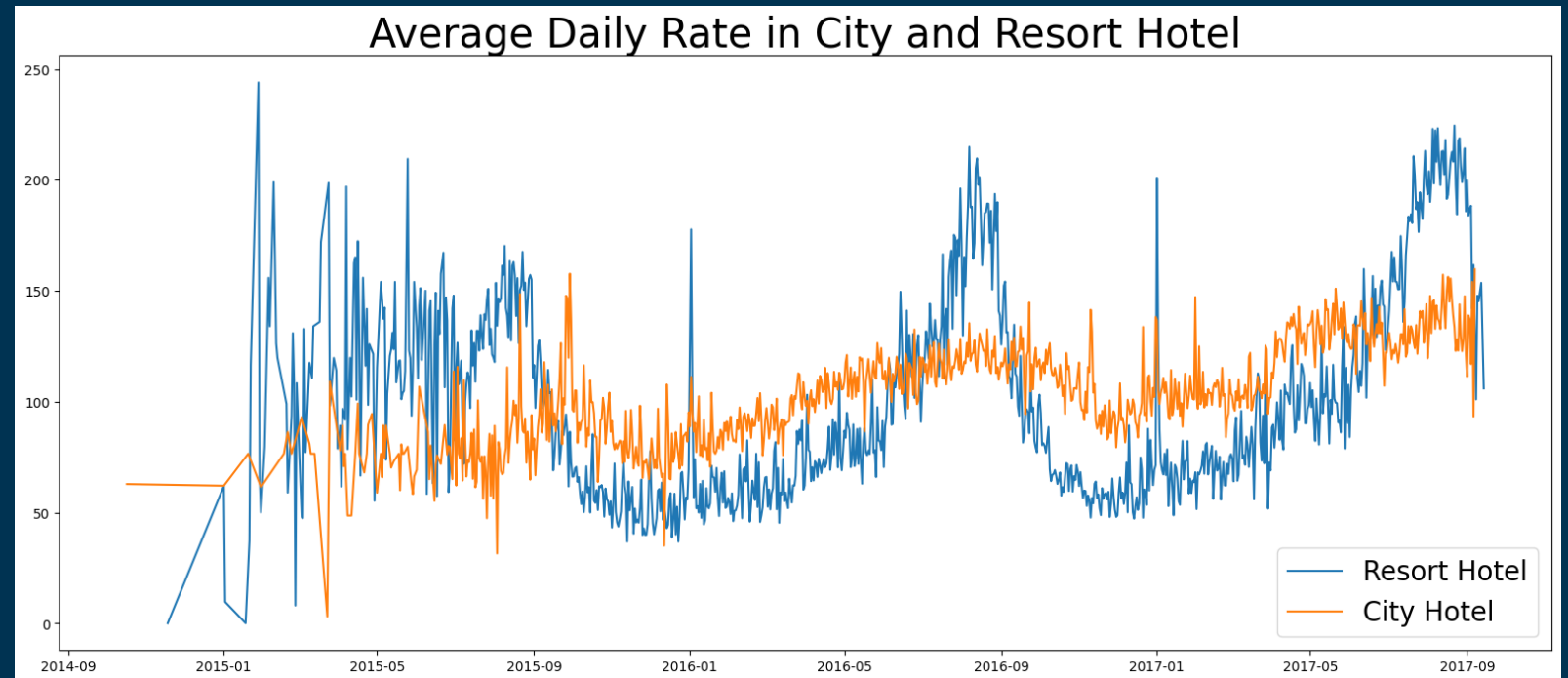
No Previous Bookings Not Canceled: The majority of bookings have no previous bookings that were not canceled, and these have a high number of cancellations.

Higher Previous Bookings Not Canceled: Customers with a history of many successful (not canceled) bookings show a very low likelihood of canceling their current bookings, suggesting that a good track record reduces the likelihood of cancellations.



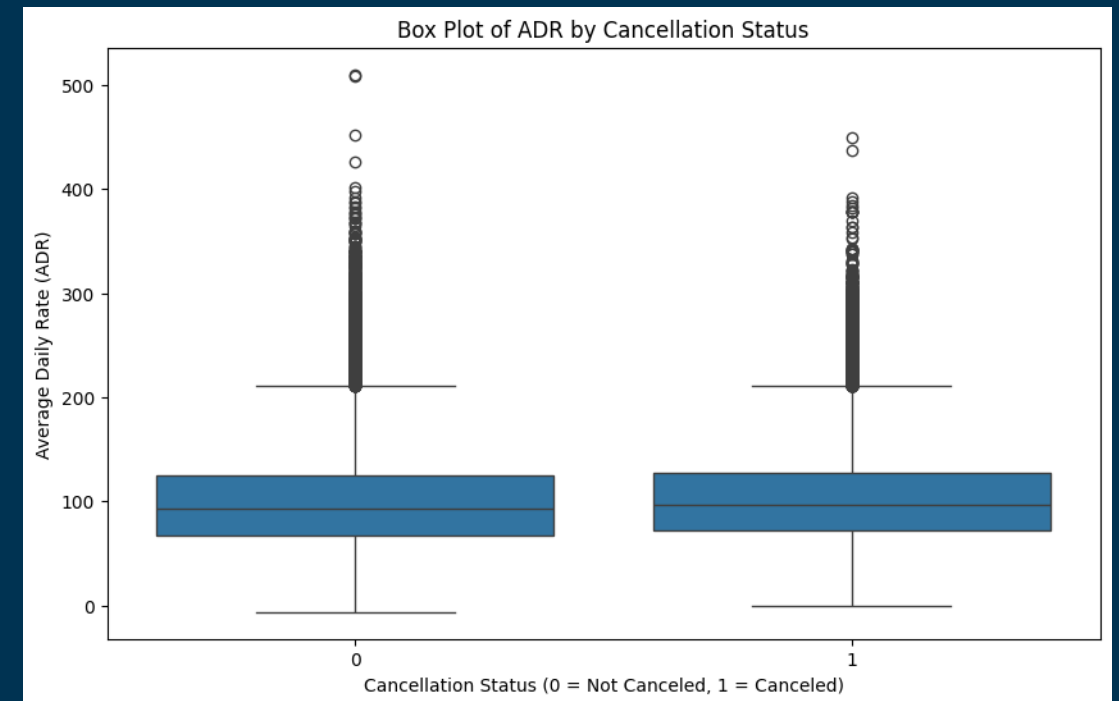
Average Daily Rate comparison of hotel type

- ADR being higher among resorts might indicate that resort hotels are more expensive to operate compared to city hotels.
- City hotels tend to be more stable in ADR and do not fluctuate as much as resort hotels.



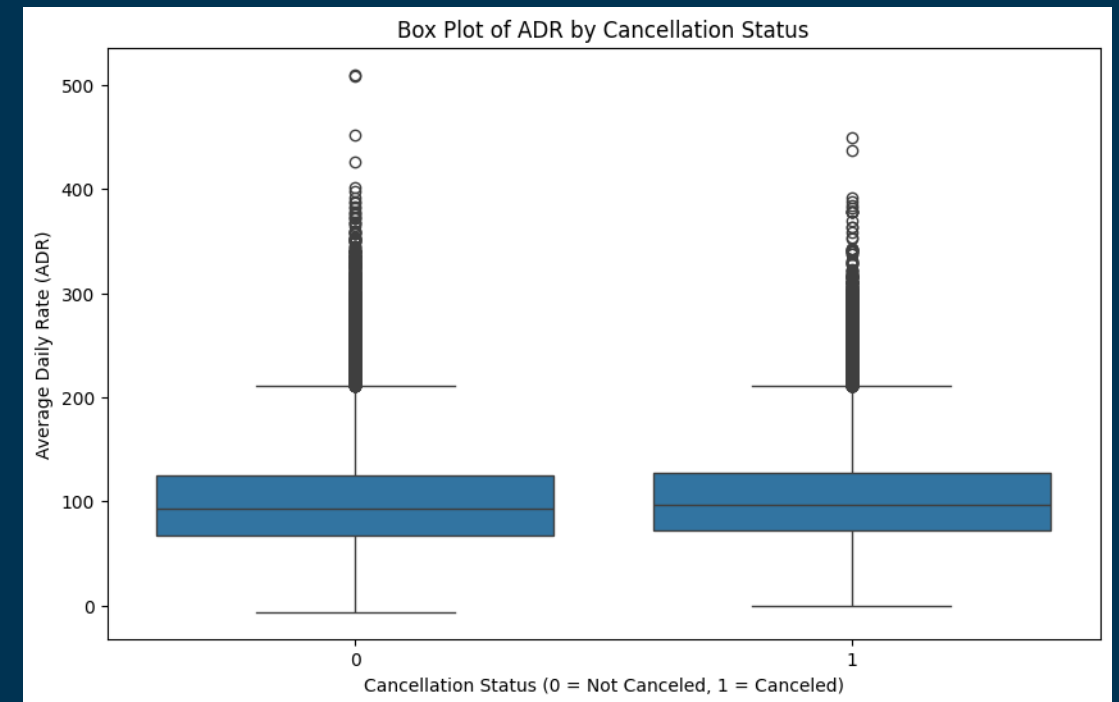
Week Correlation between ADR and Reservation Cancellation

- Median ADR: The median ADR for both canceled (1) and non-canceled (0) reservations is quite similar, suggesting that the typical rate is consistent regardless of cancellation status.
- Interquartile Range (IQR): The IQR, which represents the middle 50% of the data, is also similar for both groups, indicating that the spread of typical ADR values is comparable.
- Outliers: Both canceled and non-canceled reservations have a number of outliers with higher ADR values. There are more outliers on the higher end for canceled reservations, indicating that some of the higher-priced reservations are more likely to be canceled.
- Overall Distribution: The overall distribution of ADR is slightly higher for canceled reservations, but the difference is not substantial.



Week Correlation between ADR and Reservation Cancellation

- A correlation coefficient of 0.0487 suggests a very weak positive correlation between Average Daily Rate (ADR) and cancellations.
- This means that there is a slight tendency for cancellations to increase as ADR increases, but the relationship is not strong.

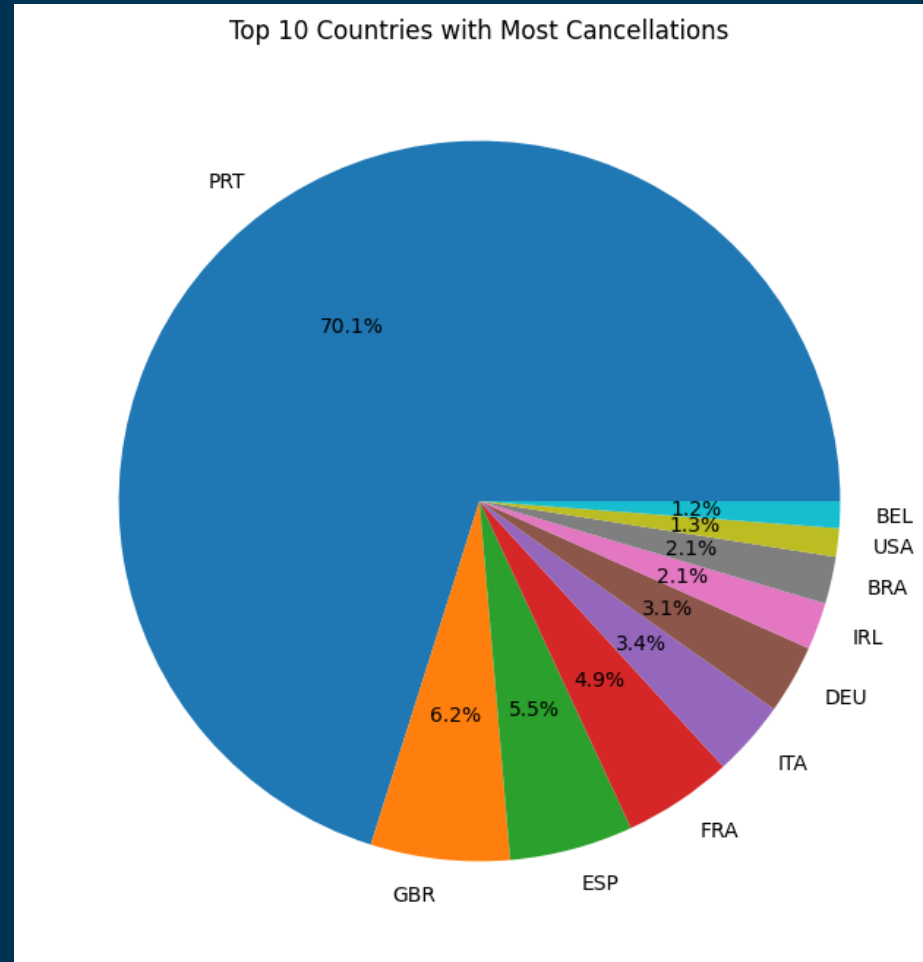


Relationship between guest country and cancelation

Guest come from 177 different countries

PRT has the highest cancelations with 70 percent cancelation rate.

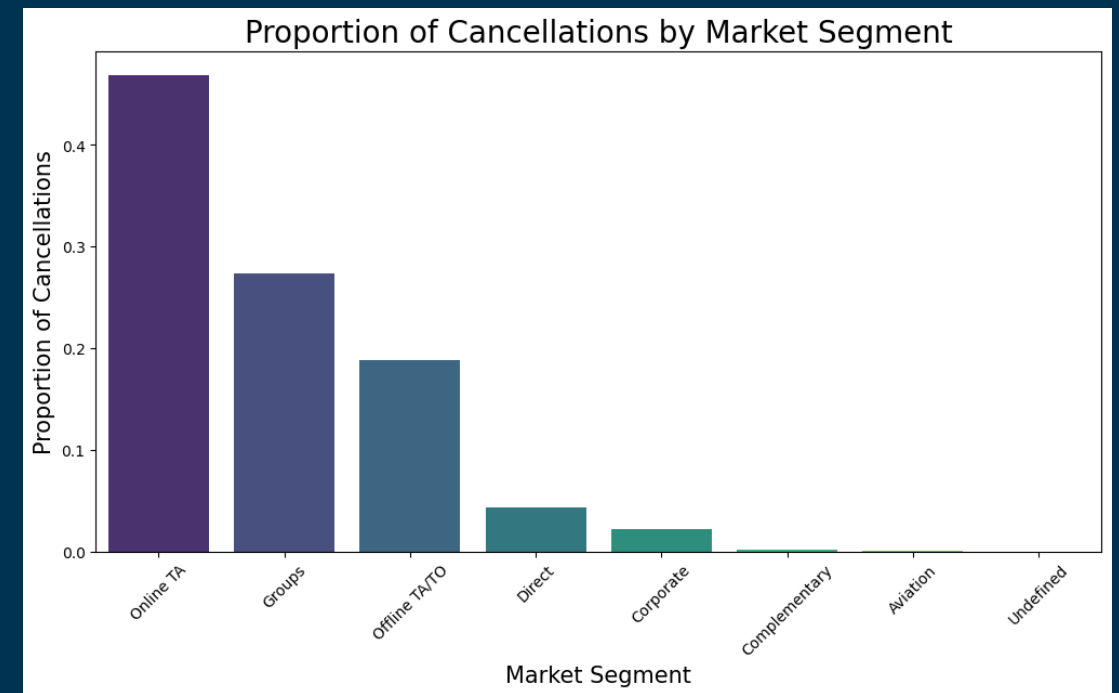
USA and Bell had the least cancelation



Relationship between Market Segment and cancelation

Percentage of cancelation by segment type

- Online TA 0.468964
- Groups 0.273545
- Offline TA/TO 0.187911
- Online Travel Agent make up half of the cancelations.

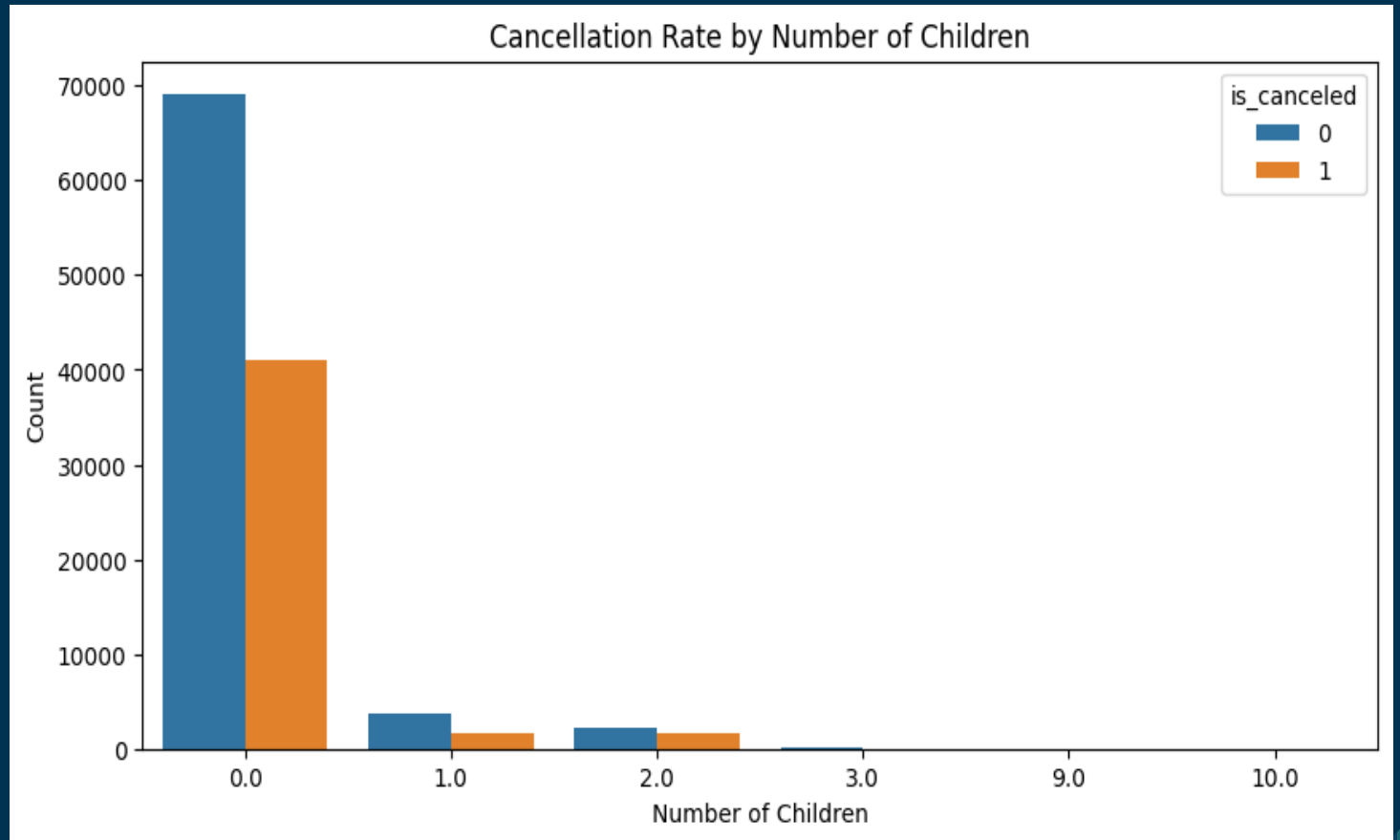


Relationship Between Children and booking

No Children: The majority of bookings have no children, and there is a significant number of cancellations among these bookings.

One and Two Children: Bookings with one or two children have fewer total bookings, with a lower but notable proportion of cancellations.

Three or More Children: There are very few bookings with three or more children, and cancellations in these categories are minimal, indicating that families with more children are less likely to book or cancel.



Relationship Between Children and booking

Percentage of customers booking who have children: 8%

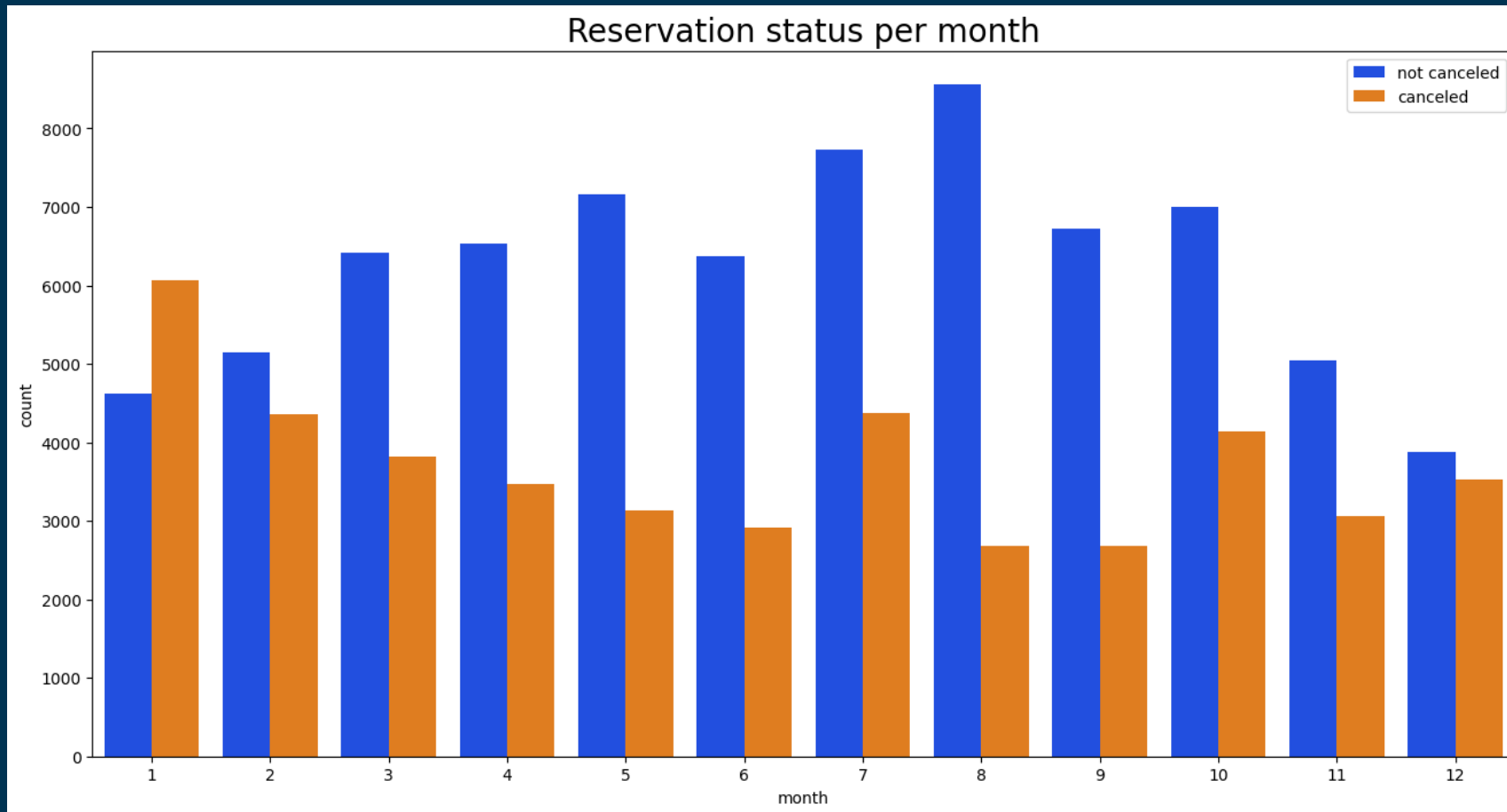
Of the 8% who have children, around 35% cancel.

Whether or not a customer has children doesn't seem to impact cancelations that much sense most customers booking reservations don't have children. Also, children includes babies.

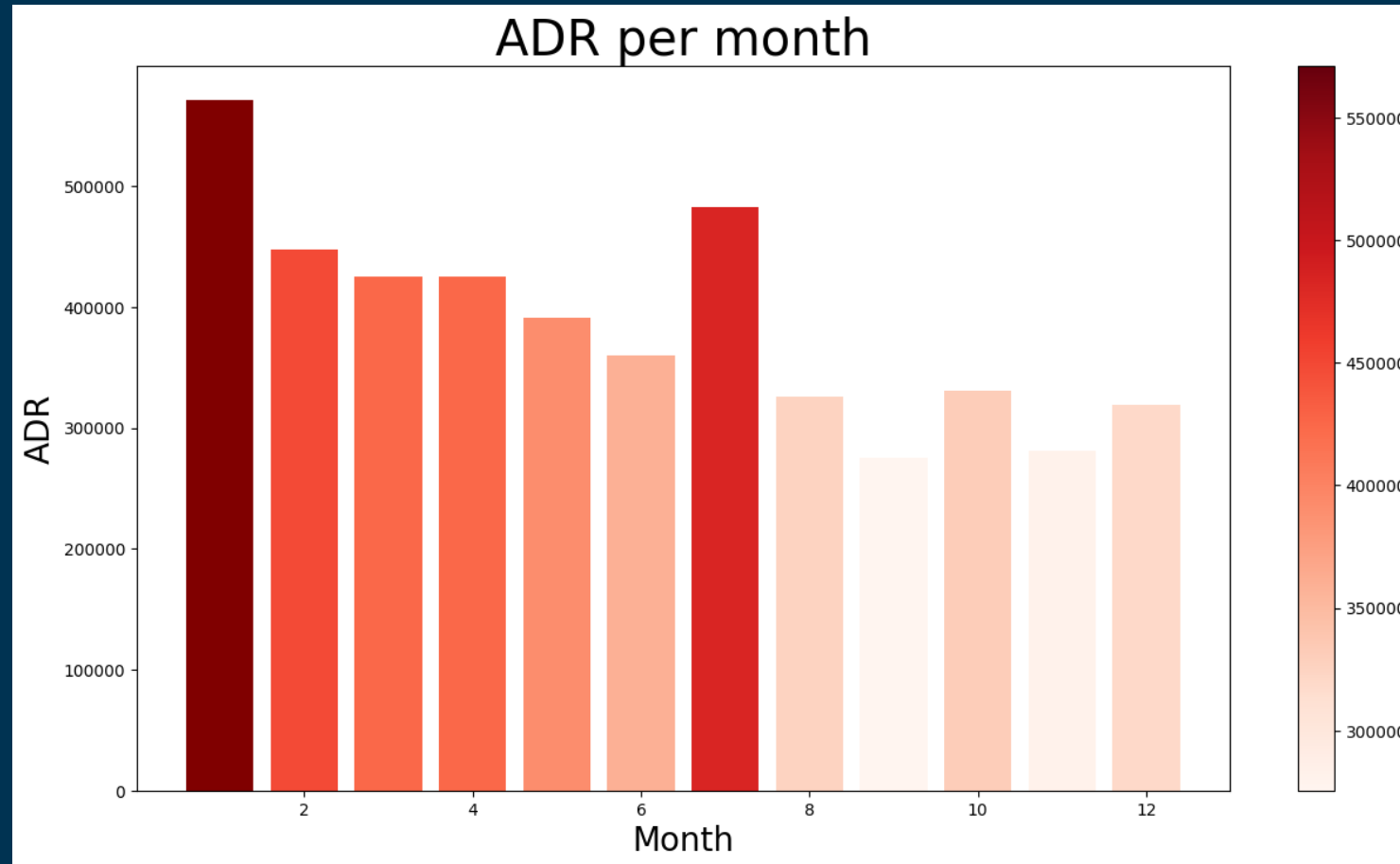
The background is a dark blue gradient. A diagonal line runs from the bottom-left towards the top-right. To the left of this line is a lighter blue area. To the right is the dark blue area. A hatched pattern of thin, parallel lines is visible along the diagonal line.

Thank You

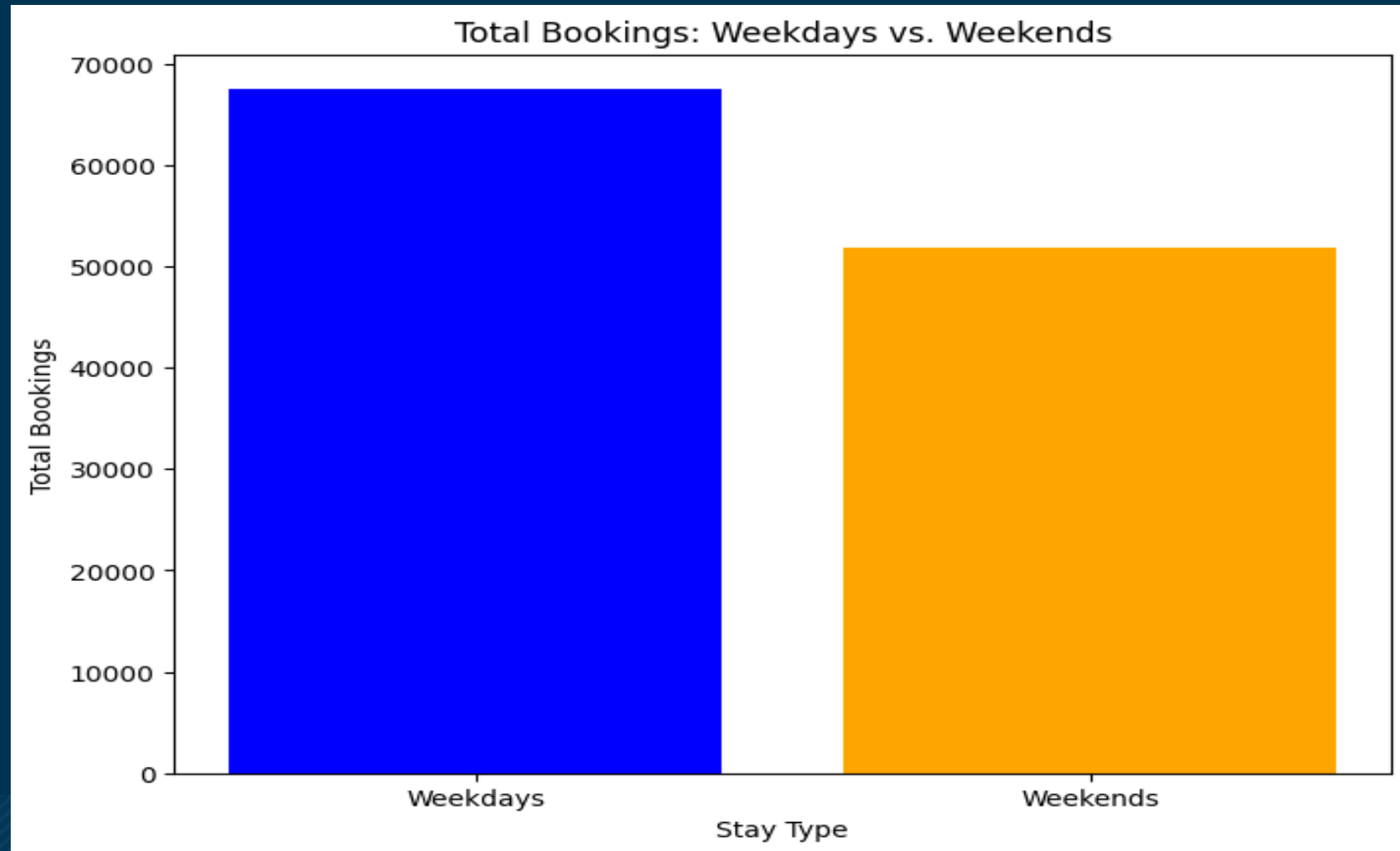
Reservation Status per Moth



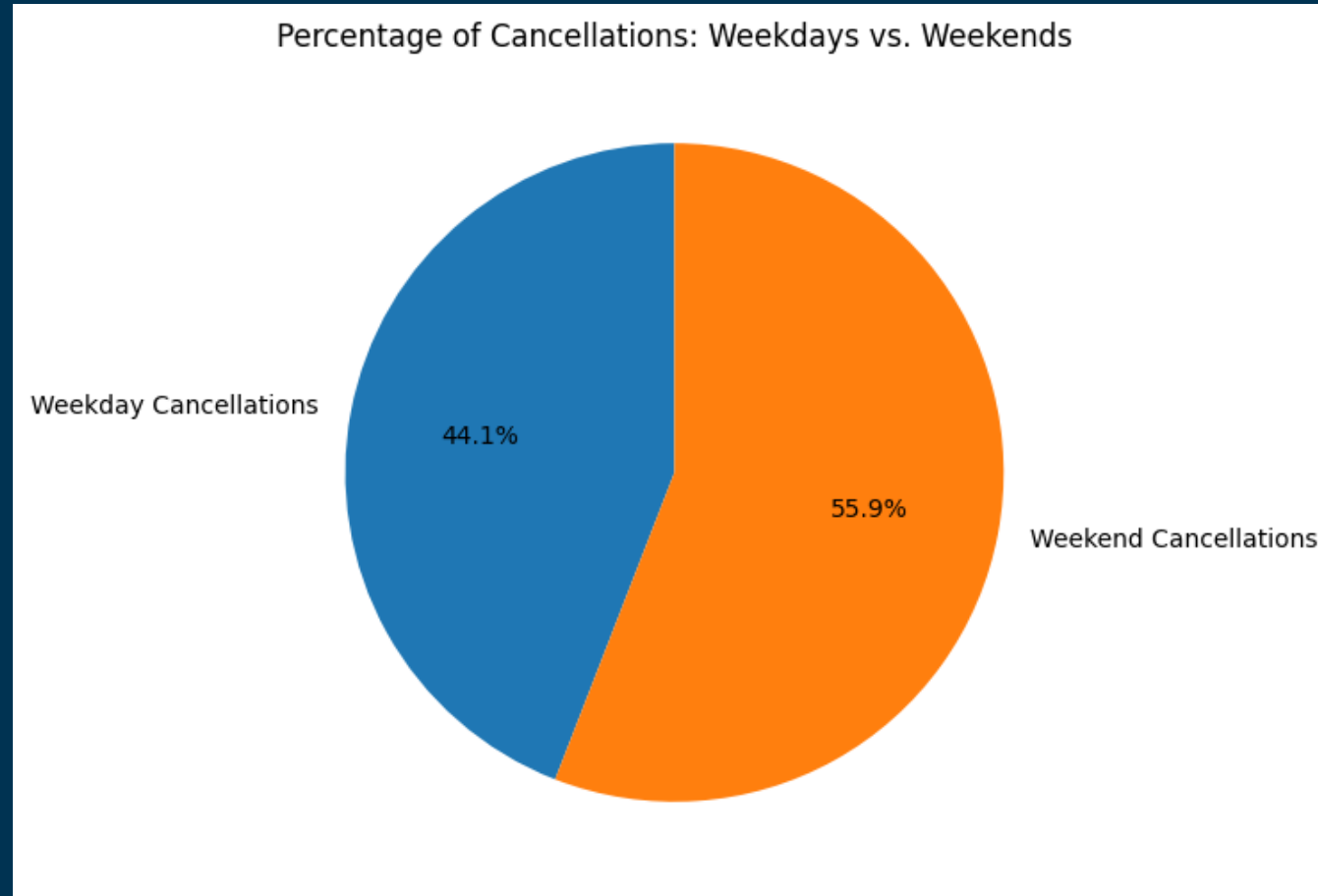
ADR for cancelations by month



Total Reservations by Weekends vs Weekdays

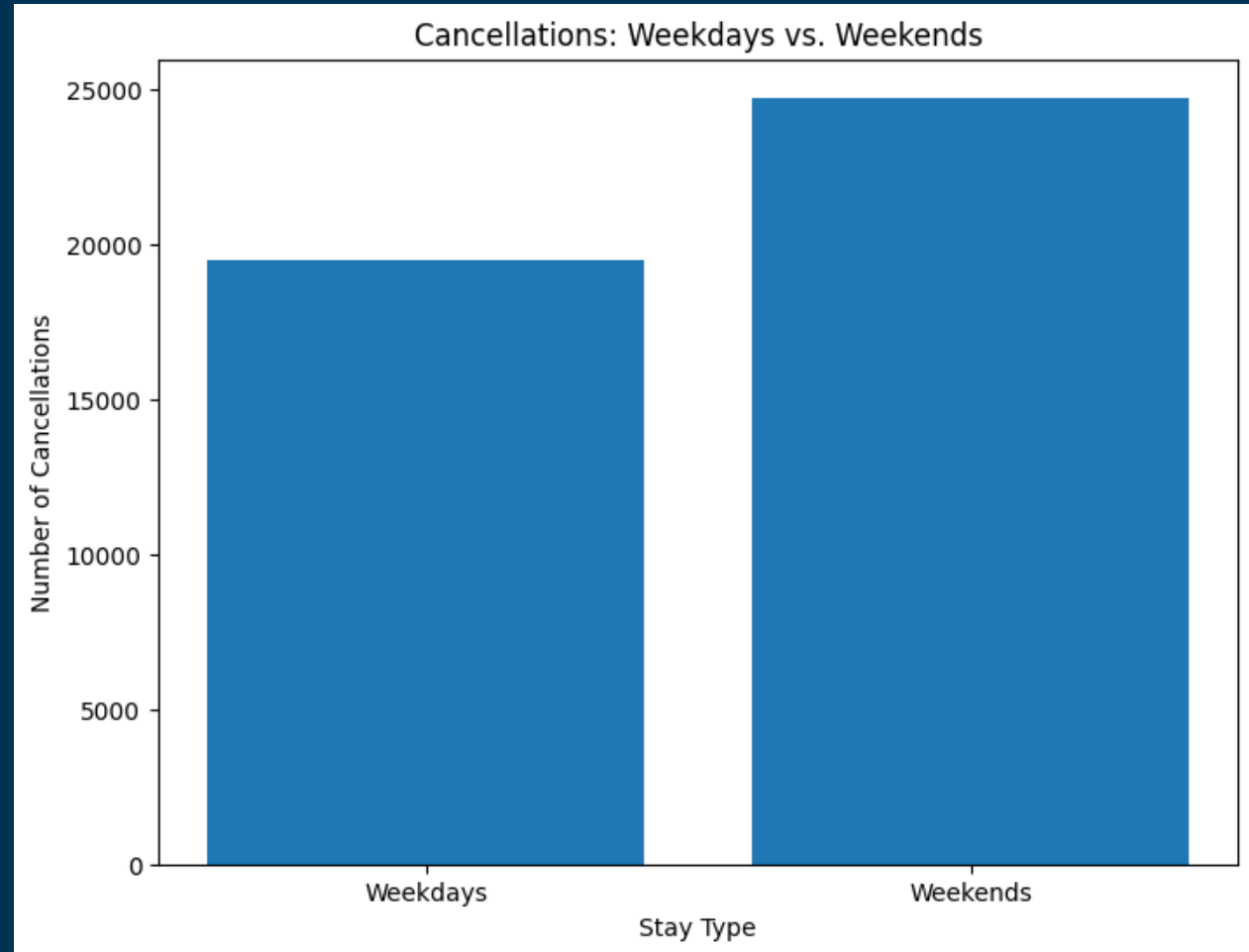


Weekdays vs Weekend by Cancellations Rate

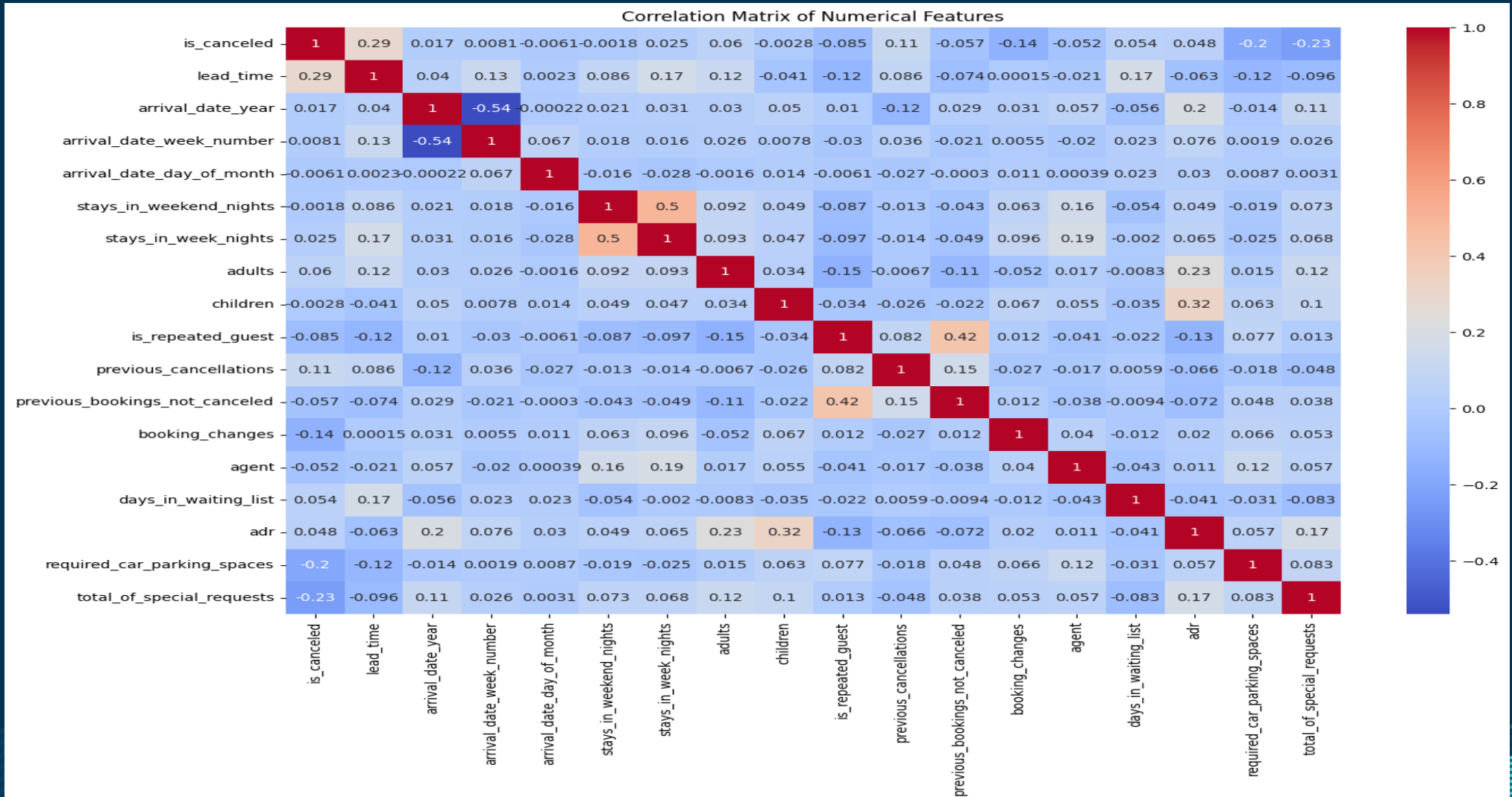


Weekend vs Weekdays by cancellations = true

No real linear relationship seen between ADR and Cancellation Status.



Correlation Matrix for Numerical Features



First logistic regression model including deposit type and reservation status

1. Accuracy score could be due to overfitting and data leaking.
2. Ended up taking deposit type and reservation features
3. The second model scored 85 compared to 100 percent.
4. Recommend to include deposit type back in the model to possibly obtain a 90 something percent.

lassification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	14907
1	1.00	1.00	1.00	8971
accuracy			1.00	23878
macro avg	1.00	1.00	1.00	23878
weighted avg	1.00	1.00	1.00	23878

First logistic regression model including deposit type and reservation status features

- First model had no false positives nor false negatives.

