
Status Report: IST 687

Sinem Aytac • Emmanuel Victor Kamya • Abhijith Vamadev

Overview

- Data Structure
 - Data Preparatory Work
 - EDA
 - Maps
 - Machine Learning - Logistic Regression
 - Final Recommendation /Conclusion
-

Data (40,600,20)

- **IsCanceled:** Categorical Value indicating if the booking was canceled (1) or not (0)
 - **LeadTime:** Integer, Number of days that elapsed between the entering date of the booking into and the arrival date
 - **StaysInWeekendNights:** Integer, Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
 - **StaysInWeekNights:** Integer, Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
 - **Adults:** Integer, Number of adults
 - **Children:** Integer, Number of children
 - **Babies:** Integer, Number of babies
 - **Meal:** Categorical, Type of meal booked. Categories are presented in standard hospitality meal packages: Undefined/SC – no meal package; BB – Bed & Breakfast; HB – Half board (breakfast and one other meal – usually dinner); FB – Full board (breakfast, lunch and dinner)
 - **Country:** Categorical, Country of origin. Categories are represented in the ISO 3155–3:2013 format
 - **MarketSegment:** Categorical, Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”
 - **IsRepeatedGuest:** Categorical, Value indicating if the booking name was from a repeated guest (1) or not (0)
 - **PreviousCancellations:** Integer, Number of previous bookings that were cancelled by the customer prior to the current booking
 - **PreviousBookingsNotCanceled:** Integer, Number of previous bookings not cancelled by the customer prior to the current booking
 - **ReservedRoomType:** Categorical, Code of room type reserved. Code is presented instead of designation for anonymity reasons
 - **AssignedRoomType:** Categorical, Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request. Code is presented instead of designation for anonymity reasons
 - **BookingChanges:** Integer, Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation
 - **DepositType:** Categorical, Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: No Deposit – no deposit was made. Non Refund – a deposit was made in the value of the total stay cost. Refundable – a deposit was made with a value under the total cost of stay.
 - **CustomerType:** Categorical, Type of booking, assuming one of four categories: Contract - when the booking has an allotment or other type of contract associated to it; Group – when the booking is associated to a group; Transient – when the booking is not part of a group or contract, and is not associated to other transient booking; Transient-party – when the booking is transient, but is associated to at least other transient booking
 - **RequiredCardParkingSpaces:** Integer, Number of car parking spaces required by the customer
 - **TotalOfSpecialRequests:** Integer, Number of special requests made by the customer (e.g. twin bed or high floor)
-

Data Preparatory Work

- Checked for null values - No null values present
 - Checked for negative values - No negative values present
-

Data Preparatory Work

- Checked for null values - No null values present
 - Checked for negative values - No negative values present
 - Created a new variable - $\text{StayInNights} = \text{StayInWeekendNights} + \text{StayInWeekNights}$
-

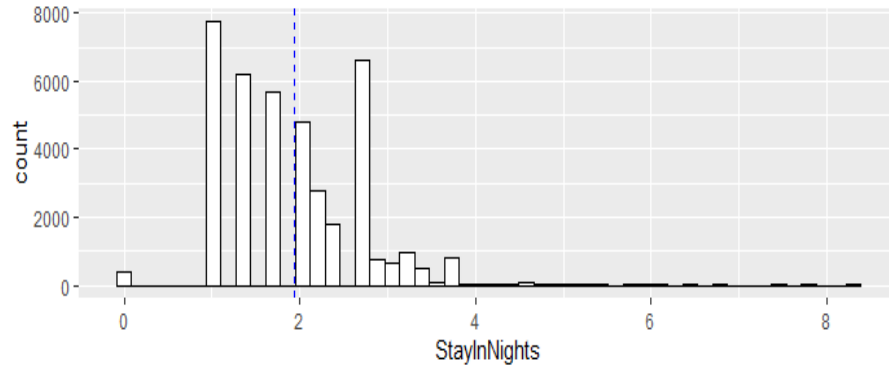
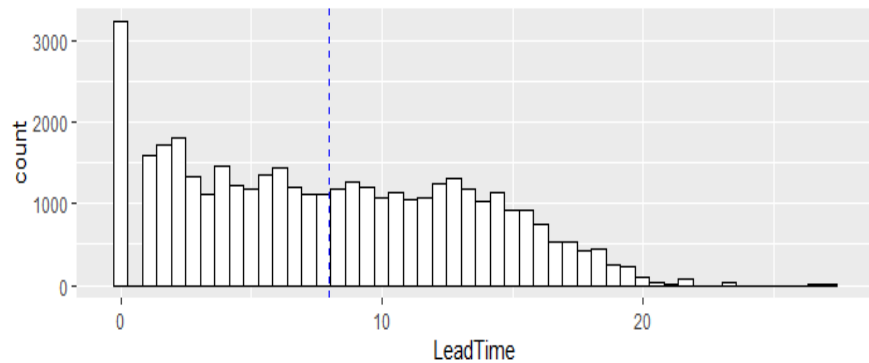
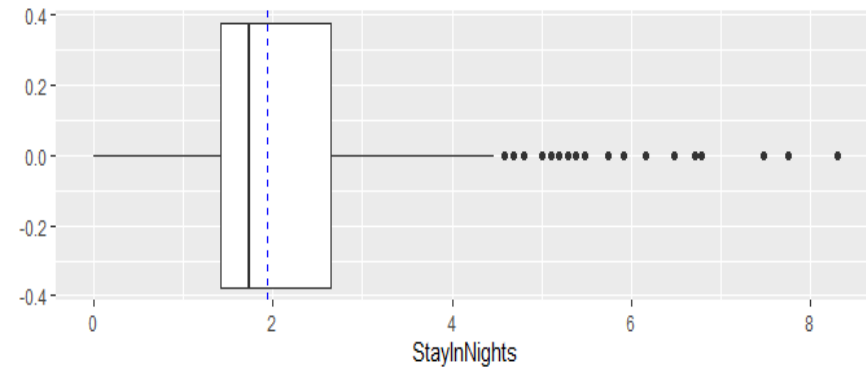
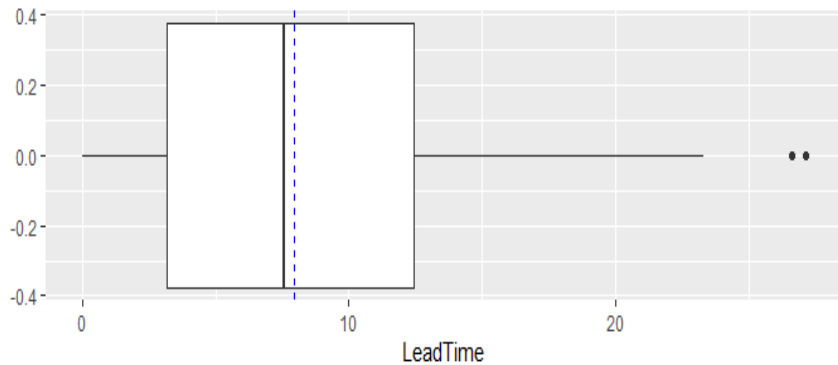
Data Preparatory Work

- Checked for null values - No null values present
 - Checked for negative values - No negative values present
 - Created a new variable - $\text{StayInNights} = \text{StayInWeekendNights} + \text{StayInWeekNights}$
 - Separated variables into numeric and categorical:
 - Numeric: LeadTime, StaysInWeekendNights, StaysInWeekNights, StayInNights, Adults, Children, Babies, PreviousCancellations, PreviousBookingsNotCanceled, BookingChanges, RequiredCarParkingSpaces, TotalOfSpecialRequests
 - Categorical: IsCanceled, Meal, Country, MarketSegment, IsRepeatedGuest, ReservedRoomType, AssignedRoomType, DepositType, CustomerType
-

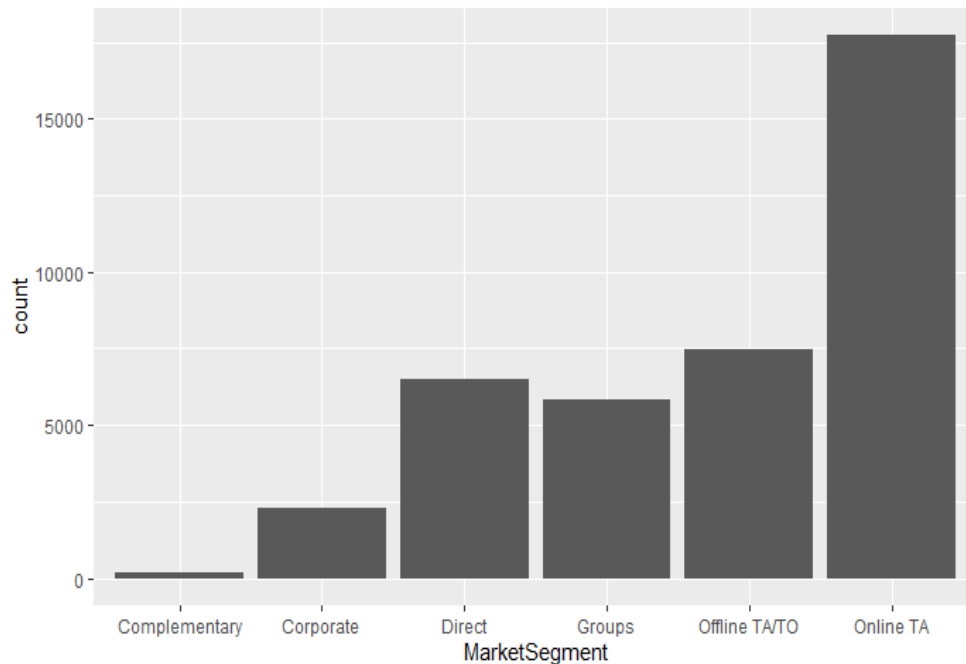
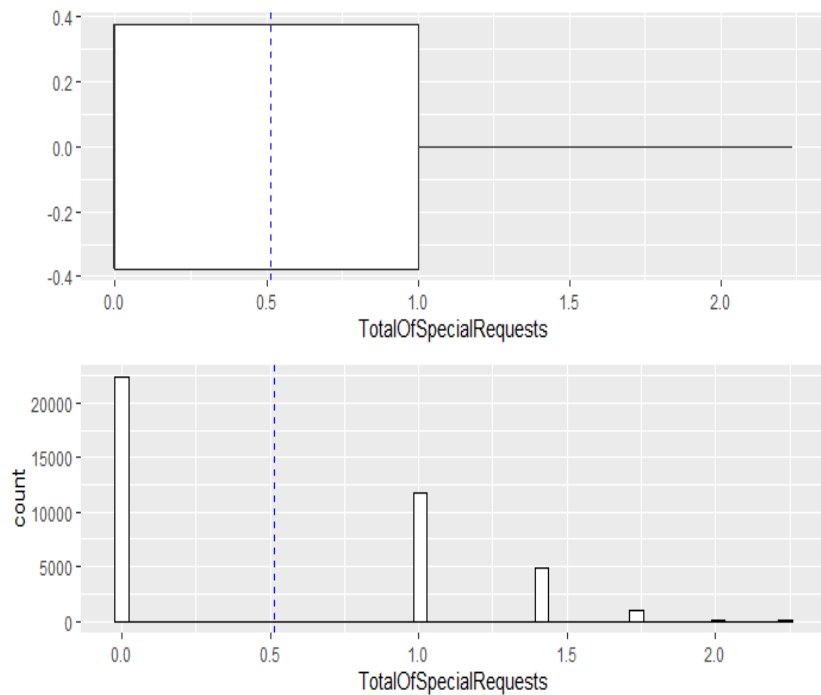
Data Preparatory Work

- Checked for null values - No null values present
 - Checked for negative values - No negative values present
 - Created a new variable - $\text{StayInNights} = \text{StayInWeekendNights} + \text{StayInWeekNights}$
 - Separated variables into numeric and categorical:
 - Numeric: LeadTime, StaysInWeekendNights, StaysInWeekNights, StayInNights, Adults, Children, Babies, PreviousCancellations, PreviousBookingsNotCanceled, BookingChanges, RequiredCarParkingSpaces, TotalOfSpecialRequests
 - Categorical: IsCanceled, Meal, Country, MarketSegment, IsRepeatedGuest, ReservedRoomType, AssignedRoomType, DepositType, CustomerType
 - Used Square Root transformation for all the numerical variables.
-

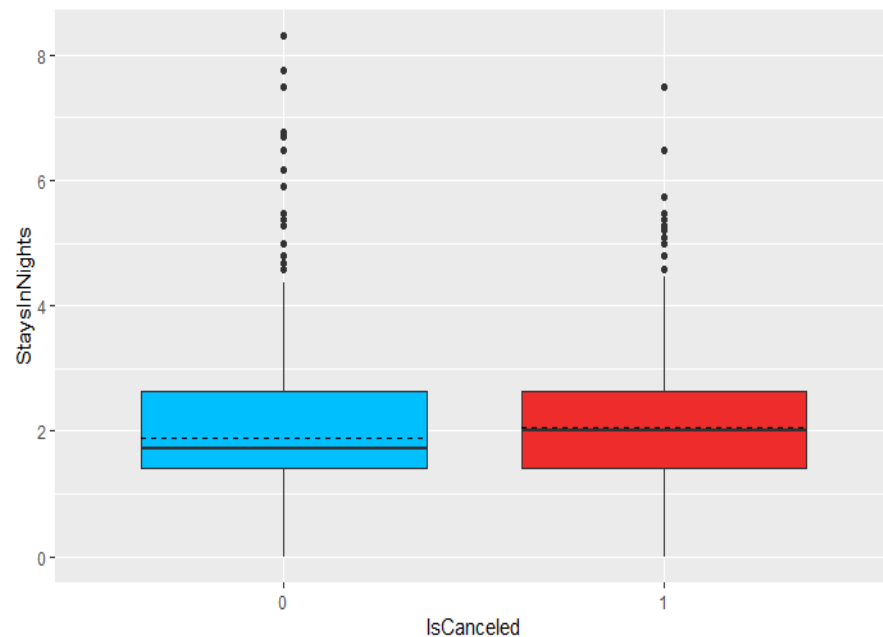
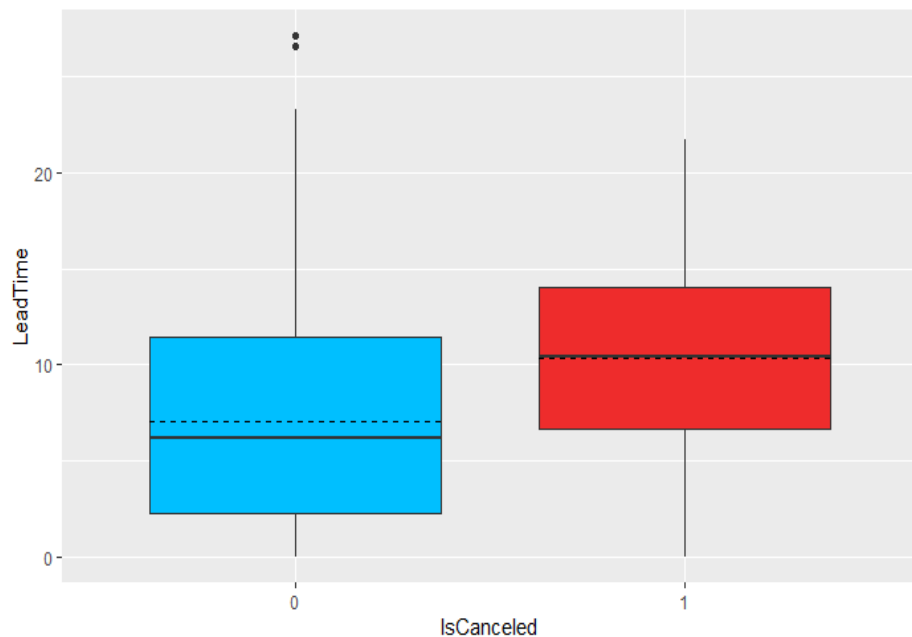
Univariate Analysis



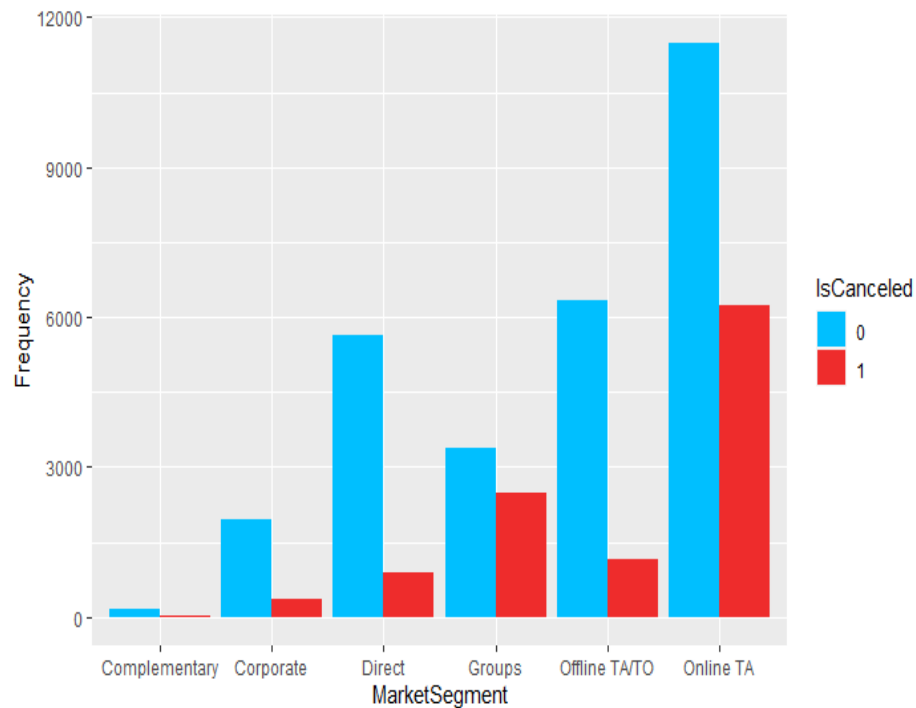
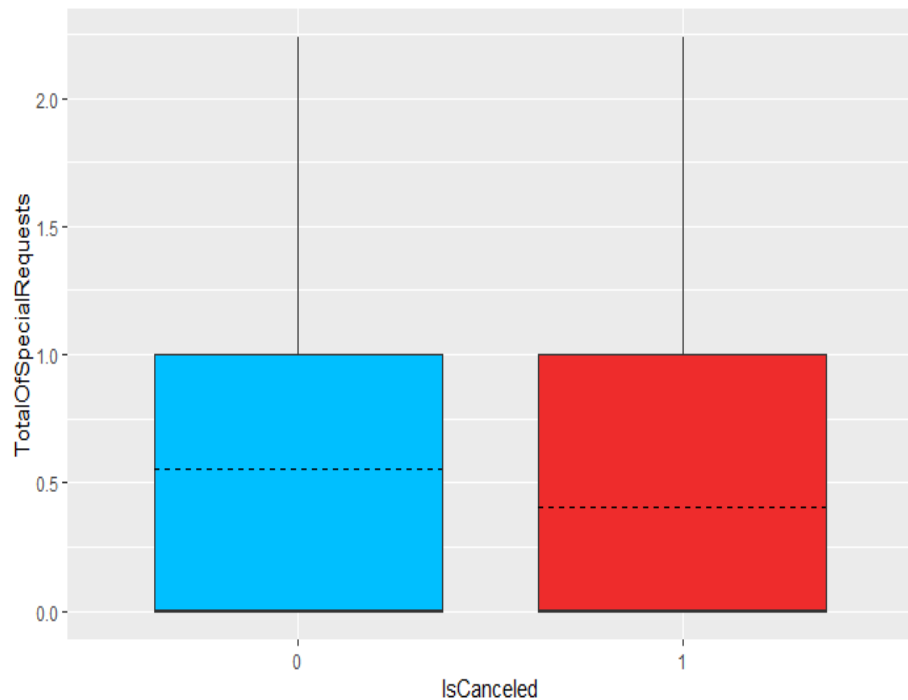
Univariate



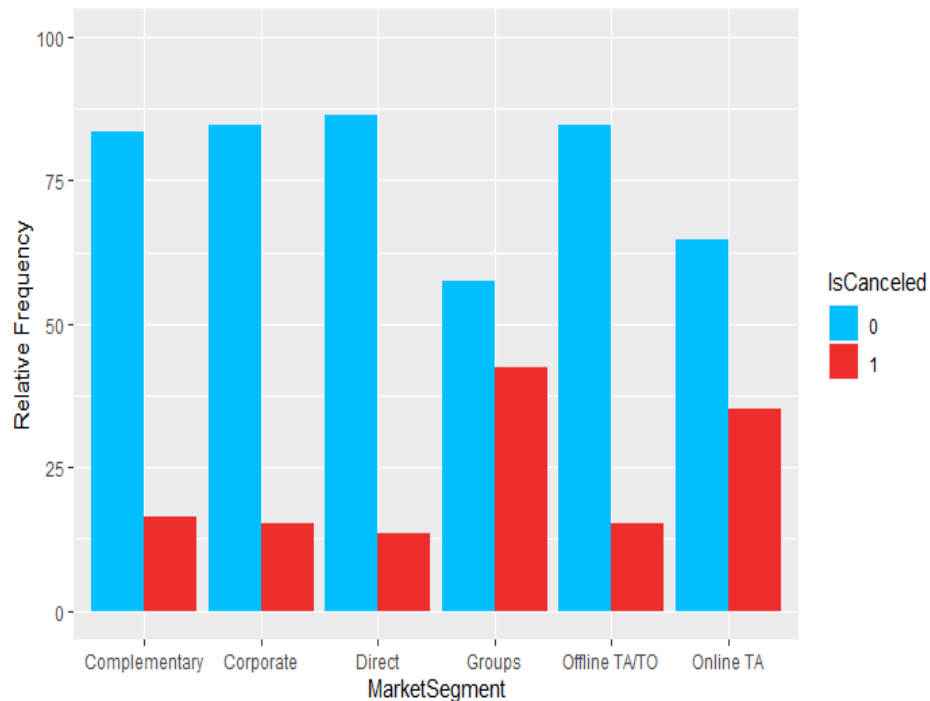
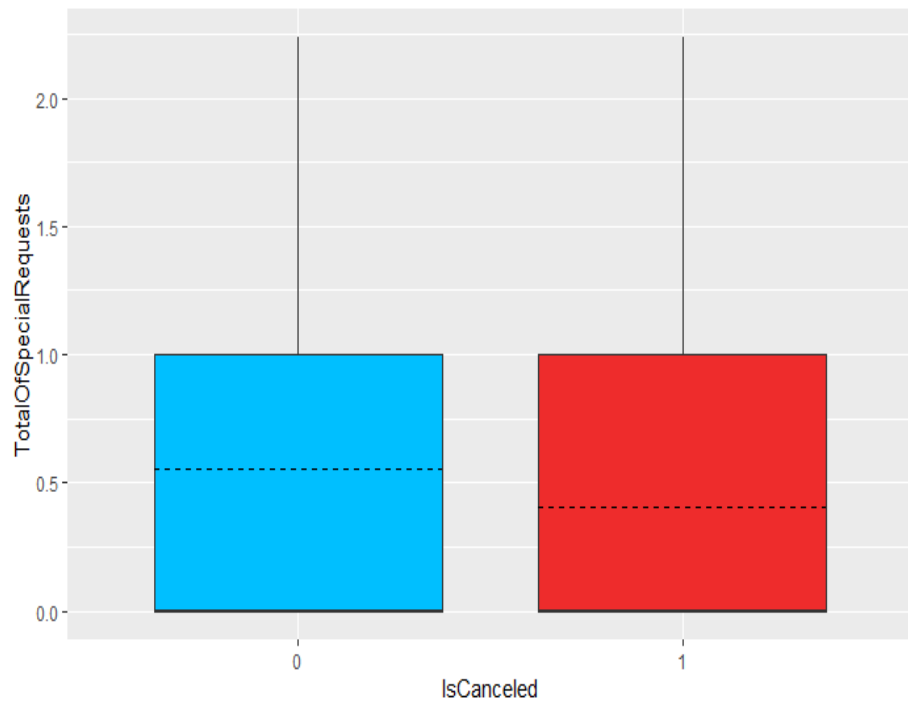
Bivariate Analysis



Bivariate Analysis



Bivariate Analysis

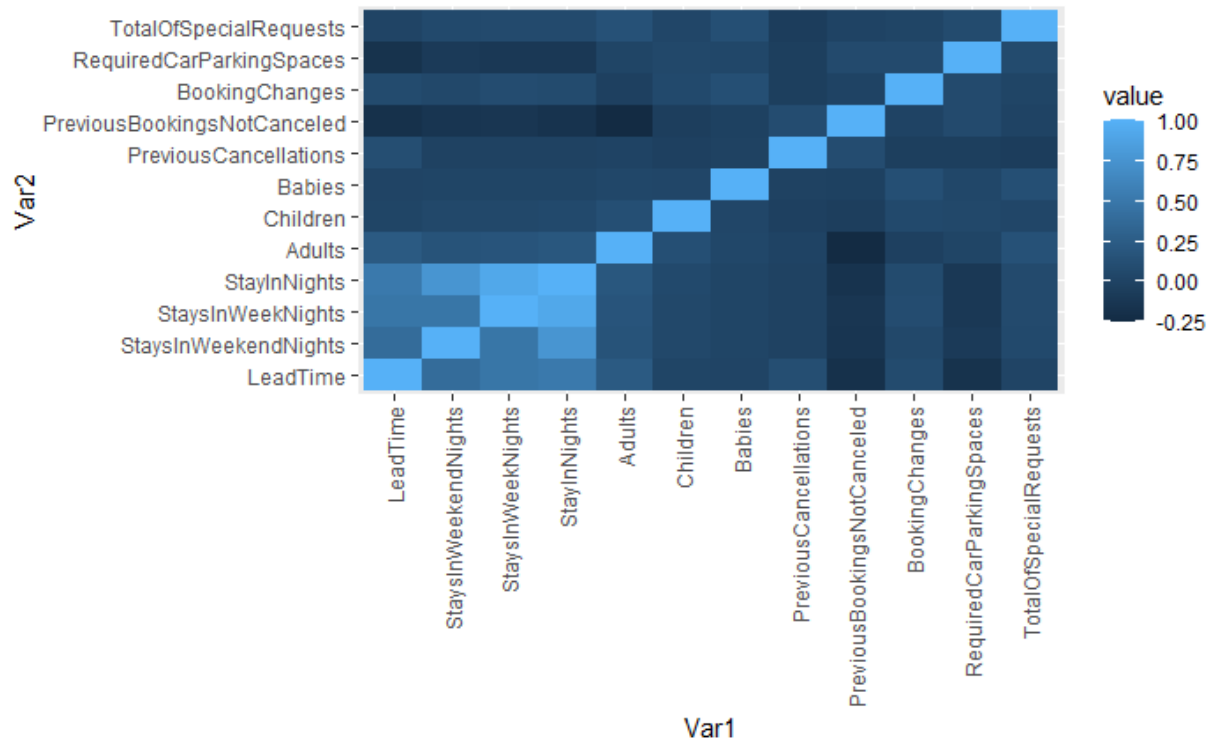


Bivariate Analysis

Heatmap visualization showing the bivariate analysis of 12 variables. The color scale represents the value, ranging from -0.25 (dark blue) to 1.00 (light blue). The diagonal elements are 1.00, indicating perfect self-correlation. The strongest positive correlations (light blue) are between 'LeadTime' and 'StayInWeekendNights', 'StayInWeekNights', and 'StayInWeekendNights'. 'LeadTime' also shows a moderate positive correlation with 'TotalOfSpecialRequests'. 'StayInWeekendNights' and 'StayInWeekNights' show moderate positive correlations with 'TotalOfSpecialRequests'. 'PreviousCancellations' and 'PreviousBookingsNotCanceled' show moderate negative correlations (dark blue) with 'TotalOfSpecialRequests'. 'RequiredCarParkingSpaces' shows a moderate negative correlation with 'TotalOfSpecialRequests'. 'Babies' and 'Children' show moderate positive correlations with 'TotalOfSpecialRequests'. 'Adults' shows a moderate negative correlation with 'TotalOfSpecialRequests'. 'BookingChanges' shows a moderate positive correlation with 'TotalOfSpecialRequests'. 'TotalOfSpecialRequests' shows a moderate positive correlation with 'LeadTime' and 'StayInWeekendNights'. 'LeadTime' shows a moderate negative correlation with 'PreviousCancellations' and 'PreviousBookingsNotCanceled'. 'StayInWeekendNights' shows a moderate negative correlation with 'PreviousCancellations' and 'PreviousBookingsNotCanceled'. 'StayInWeekNights' shows a moderate negative correlation with 'PreviousCancellations' and 'PreviousBookingsNotCanceled'. 'Adults' shows a moderate negative correlation with 'PreviousCancellations' and 'PreviousBookingsNotCanceled'. 'Children' shows a moderate negative correlation with 'PreviousCancellations' and 'PreviousBookingsNotCanceled'. 'Babies' shows a moderate negative correlation with 'PreviousCancellations' and 'PreviousBookingsNotCanceled'. 'PreviousCancellations' and 'PreviousBookingsNotCanceled' show a moderate positive correlation with each other. 'BookingChanges' shows a moderate positive correlation with 'PreviousCancellations' and 'PreviousBookingsNotCanceled'. 'RequiredCarParkingSpaces' shows a moderate positive correlation with 'PreviousCancellations' and 'PreviousBookingsNotCanceled'. 'TotalOfSpecialRequests' shows a moderate positive correlation with 'LeadTime' and 'StayInWeekendNights'.

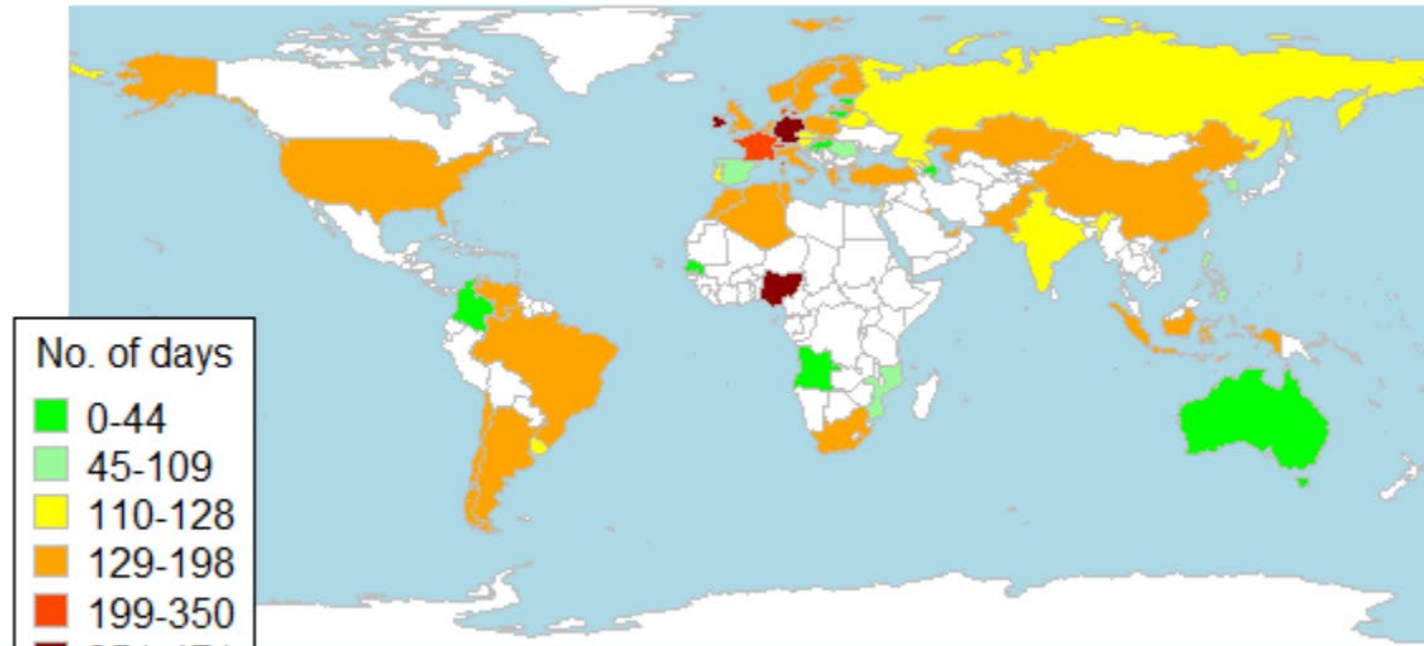
Variables (Var1):

- LeadTime
- StayInWeekendNights
- StayInWeekNights
- StayInWeekendNights
- Adults
- Children
- Babies
- PreviousCancellations
- PreviousBookingsNotCanceled
- BookingChanges
- RequiredCarParkingSpaces
- TotalOfSpecialRequests

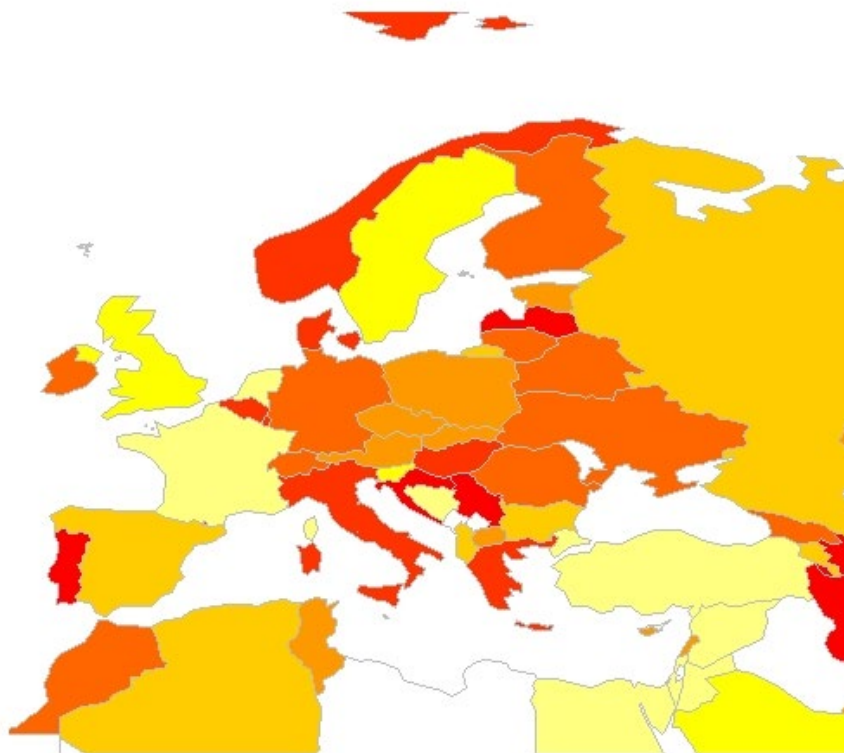


Maps

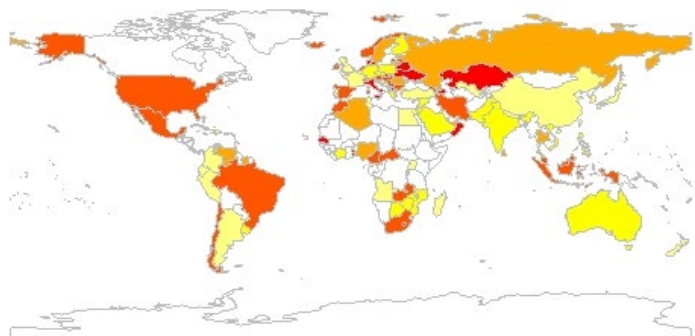
Map showing lead time for cancellations



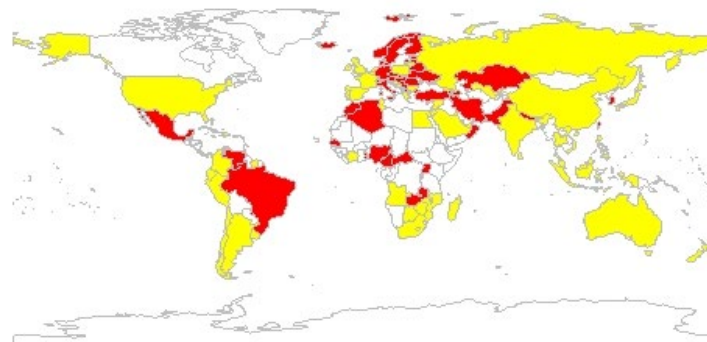
High LeadTime cancellations in Europe



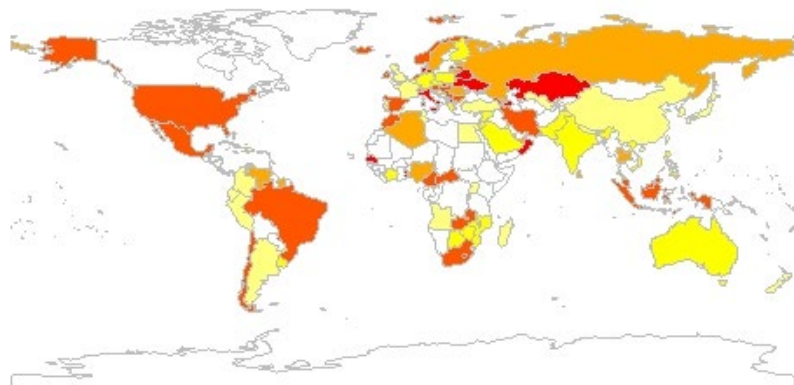
Stay In Week Nights



Stay In Weekend Nights



Stay In Week Nights

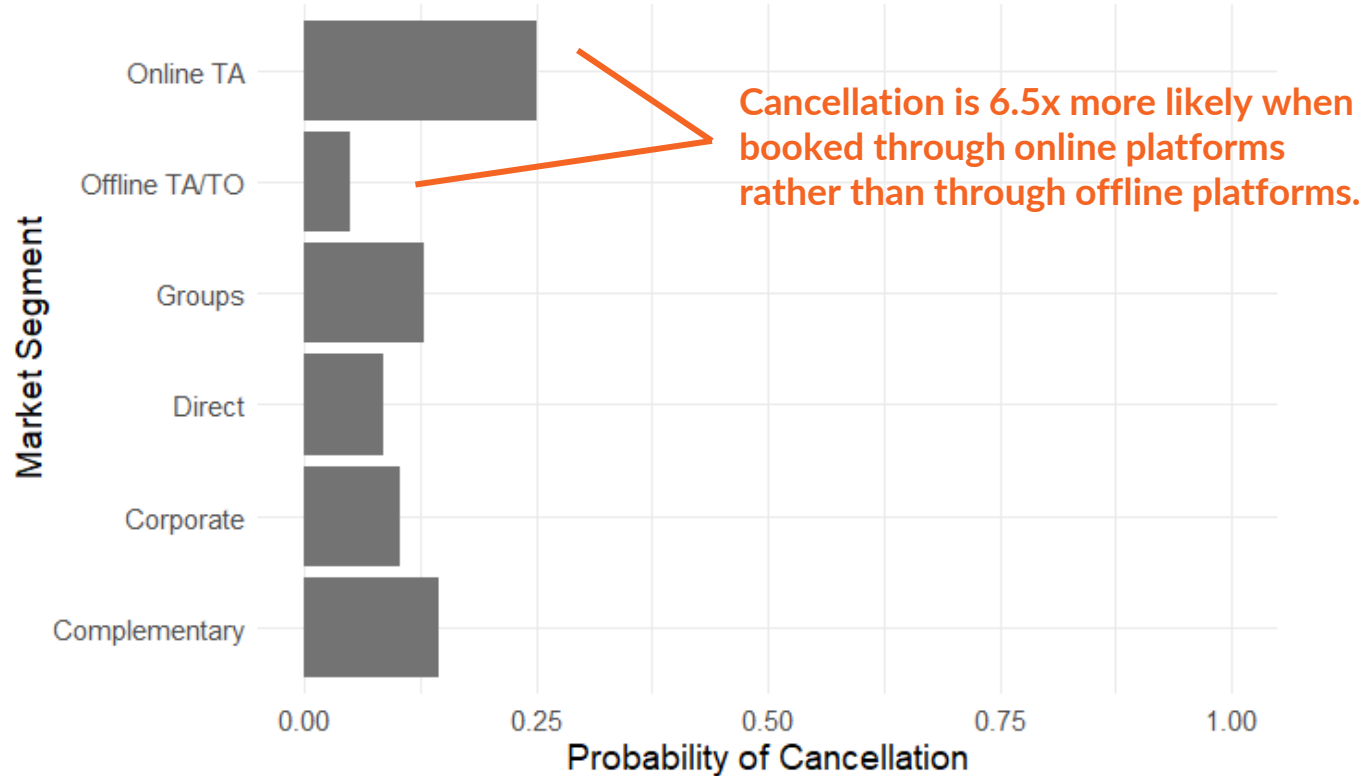


Machine Learning: Logistic Regression

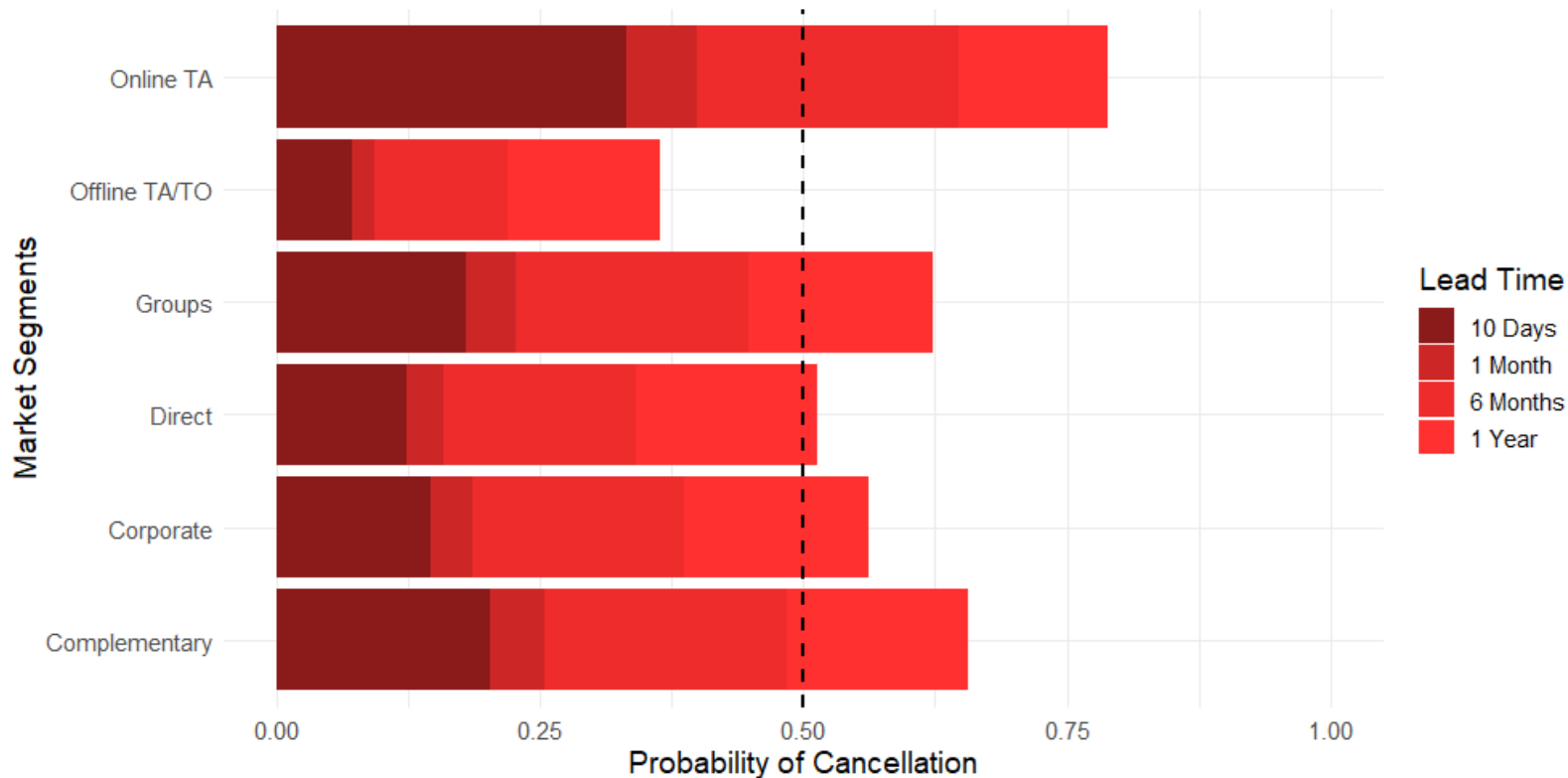
Logistic Regression

- Lead Time
 - Overall Stays in Nights
 - Total # of Special Requests
 - Market Segment
-

Market Segment



Lead Time



Number of Nights Stayed

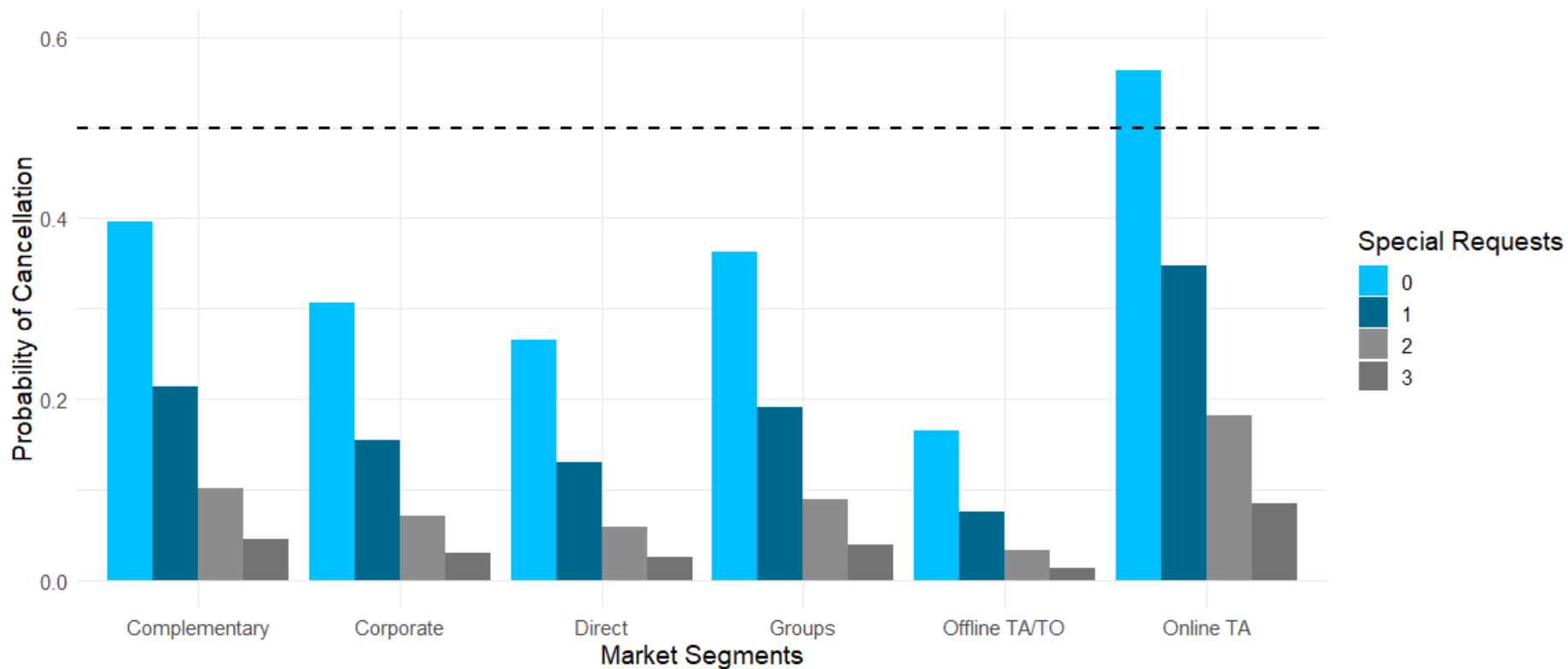
- ★ Every one more night booked to stay in a hotel (regardless of week day or weekend) increases the odds for cancellation by around 7%.
-

Total Number of Special Requests

- ★ Every additional special request by customers decreases the odds for cancellation by around 58%.
-

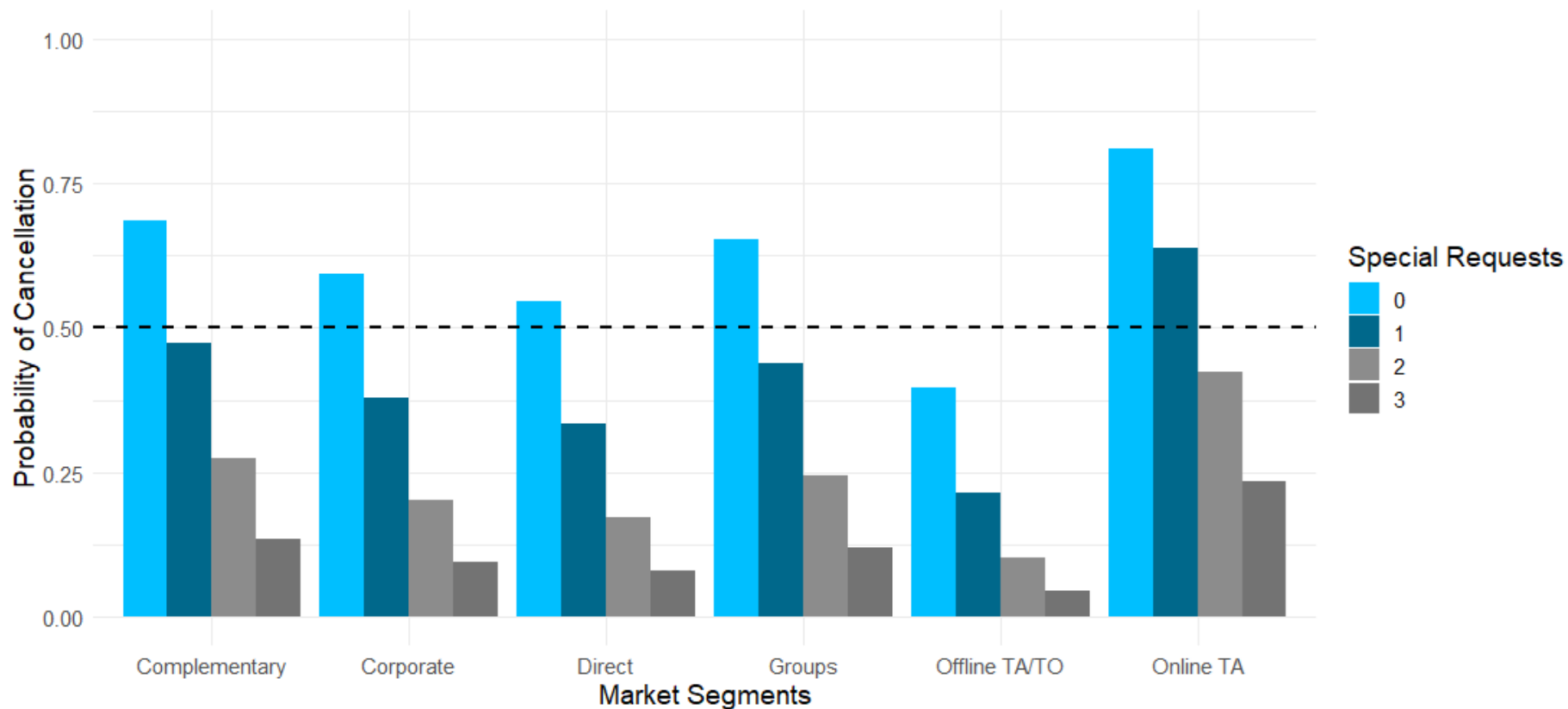
When Lead Time is on average;
Stays in Night is on average.

Total Number of Special Requests



**When Lead Time is a year;
Stays in Night is on average.**

Total Number of Special Requests



Conclusions

Cancellations are expected MORE:

- When a hotel reservation is made through *online platforms*.
- When *a year* has passed since the booking.

Cancellations are expected **LESS:**

- With each **special request** made by customer.

Thank you!
