MATH 324 Final Project

Amy Cao, Amy Folkestad, Anker Hojgaard and Jaxson Stathis

1. Project Description

Gold Mine Resorts is looking to understand the predictors for guests who cancel their reservations. The industry average for cancellation rates is twenty percent; however Gold Mine Resorts experienced a 32.8% cancellation rate for reservations between 2015 and 2018. Cancelled reservations represent a total of $4.3 million dollars in lost revenue for this time frame.

Description of the dataset

The dataset includes 36,285 rows of booking data, spanning the course of four years (2015-2018). Data collected includes the group size (adult and children), the stay length (count of weekend and weeknight stays), guest upgrades (parking, meal plans, room types and count of special requests), booking details (price and reservation method), the date of stay, year, and status (whether or not the guest cancelled the reservation).

Our team observed there was only one row of data for year 2015 and 2016. We additionally noticed there only 5 data points for meal plan 3. The variables: GroupSize, StayLength, HasChildren, HasRequests and Online were created from the existing dataset. GroupSize was an aggregation of count of adults and children, StayLength was an aggregation of Weekend and Weeknights. HasChildren, HasRequests and Online created binary, true/false (0 and 1) variables from the count of children, requests and instances where the market was online.

The status column (cancelled versus not cancelled), is the response variable of interest in this analysis.

Goal of study

The goal of this analysis is to understand the best predictors for guest cancellations in order to develop company policy in response to predicted cancellations.

1.1 Research Objectives

Objective 1:

The first objective will explore and define the most important numeric variable to predict the percentage that a group will cancel their reservation.

Objective 2:

The second objective will explore and define the most important category type variable for predicting the percentage chance a group will cancel their reservation.

Objective 3:

The final goal of our statistical analysis will combine our understanding of numeric and category type variables in order to create a statistical model to predict the percent chance that a customer will cancel.

1.2 Variables

The response variable of interest is status, whether or not a client cancelled their booking. The possible explanatory variables include both categorical and quantitative options. The number of requests from a client was treated as a categorical variable. Additionaly a new variable called "HasRequests" was created. It is a binary variable with "1" indicating the customer had a special request or "0" otherwise. We made a similar to transformation to the children column, creating a variable "HasChildren" indicating whether or not there were children in the group. We also made the variables "StayLength" which was the length of the stay (Weekends and weeknights), and "GroupSize" which was the size of the group (Adults plus Children).

Table of variable names and types

Variable	Description of Variable	Variable Type
Status	Cancellation status of booking	C
Meal	Type of meal plan selected	C
Parking	Parking option selection	C
RoomType	Type of room booked	C
Market	Market segment of booking	C
Month	Month of booking date	C
HasChildren	Indicator for children as guests	C
HasRequests	Indicator for special requests	C
Online	Indicator for an online booking	C
Requests	Number of special requests made	Q
Adults	Number of adults in the booking	Q
Children	Number of children in the booking	Q
GroupSize	Number of people in the group	Q
StayLength	Total length of stay (nights)	Q
Weekends	Number of weekend nights stayed	Q
Weeknights	Number of weeknights stayed	Q
LeadTime	Num. of days bw booking & arrival	Q
AvgPrice	Avg. room price (week of booking)	Q
Year	Year of the reservation	Q

2. Detailed Exploratory Data Analysis (EDA)

Quantitative EDA

For our quantitative exploration we used side by side box plots for every variable to status. Looking at the side by side box plot, the lead time seems to have the largest mean variation when looking at a difference in status. This indicates it might be the best predictor for predicting cancellation rates. Conversely the variables: GroupSize, StayLength, AvgPrice, Adults, and Children do not seem to be strong individual predictors for cancellation rates.

Table 2: Quantitative Explanatory Variables

Status	Mean_LeadTime	Mean_AvgPrice	Mean_GroupSize	Mean_StayLength
Canceled	139.220	110.580	2.034	3.280
Not_Canceled	58.934	99.933	1.909	2.886

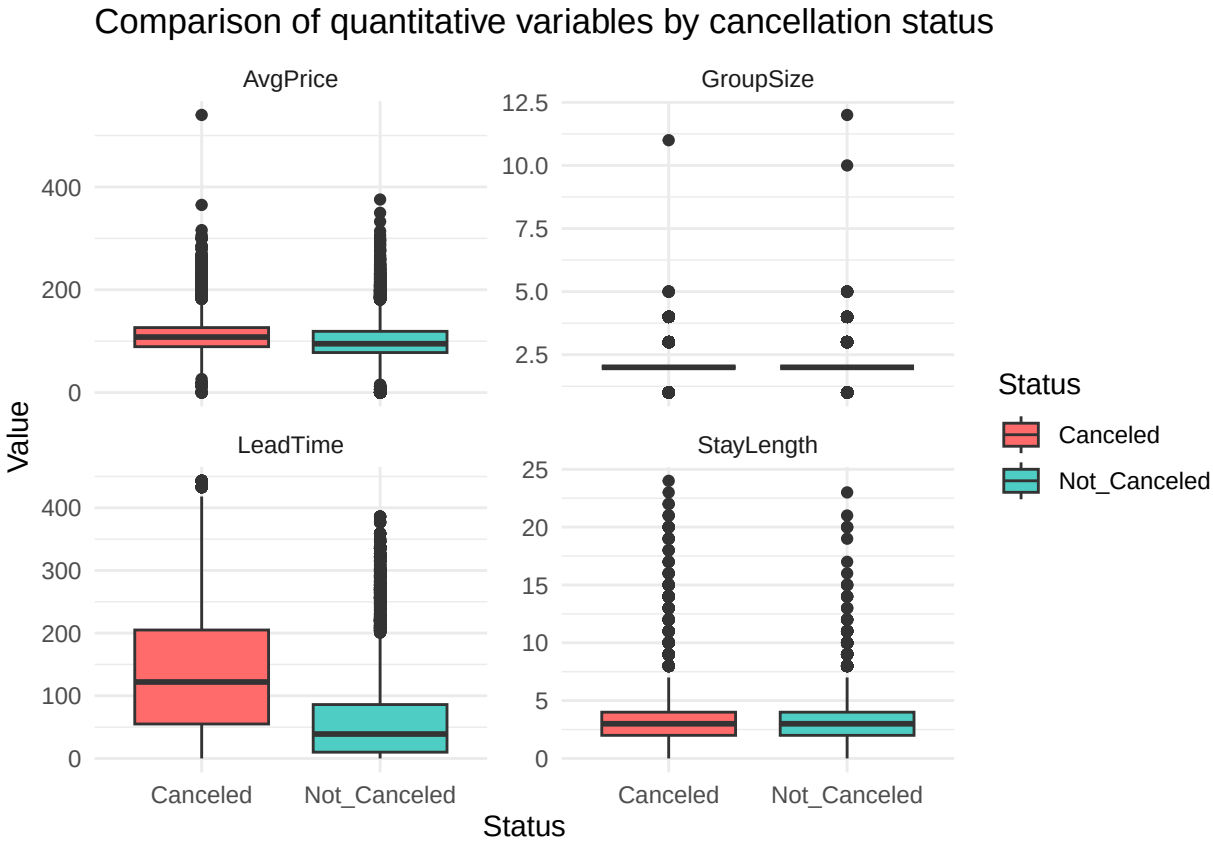
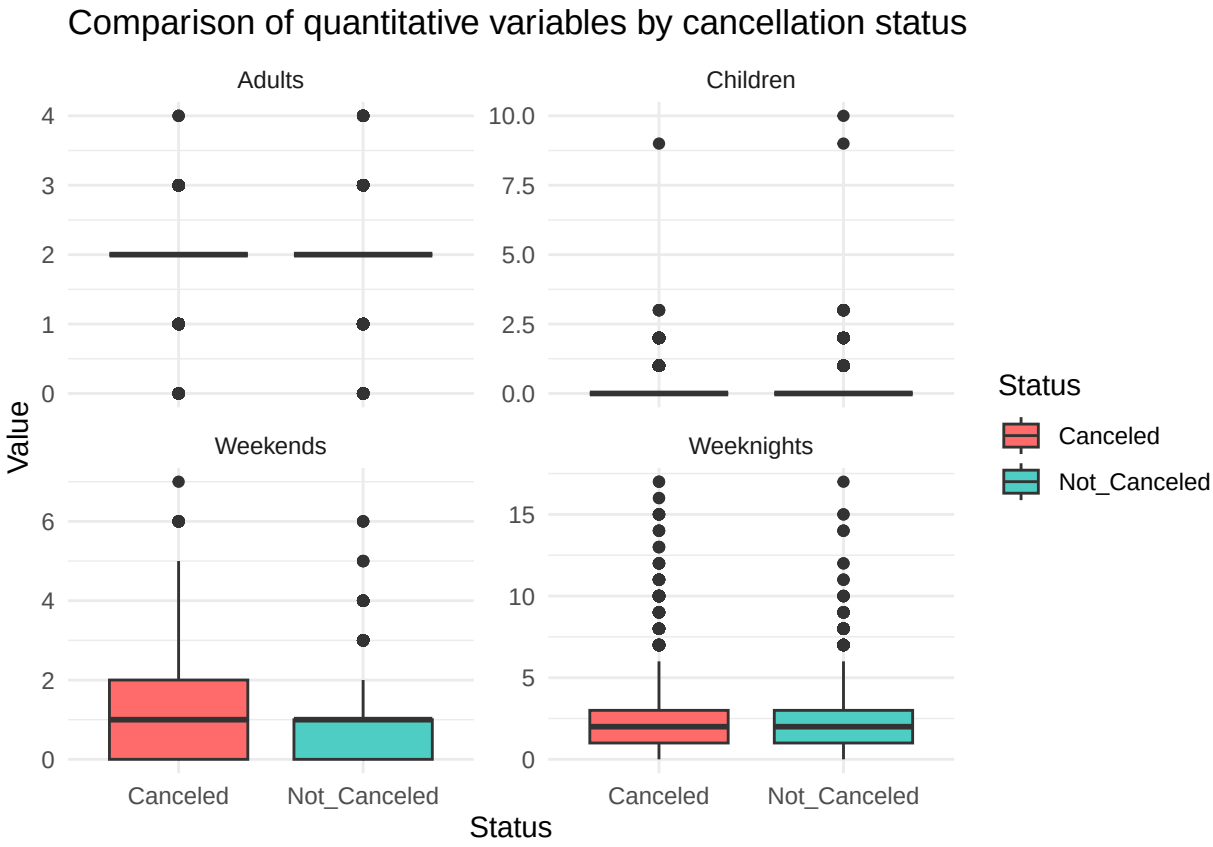
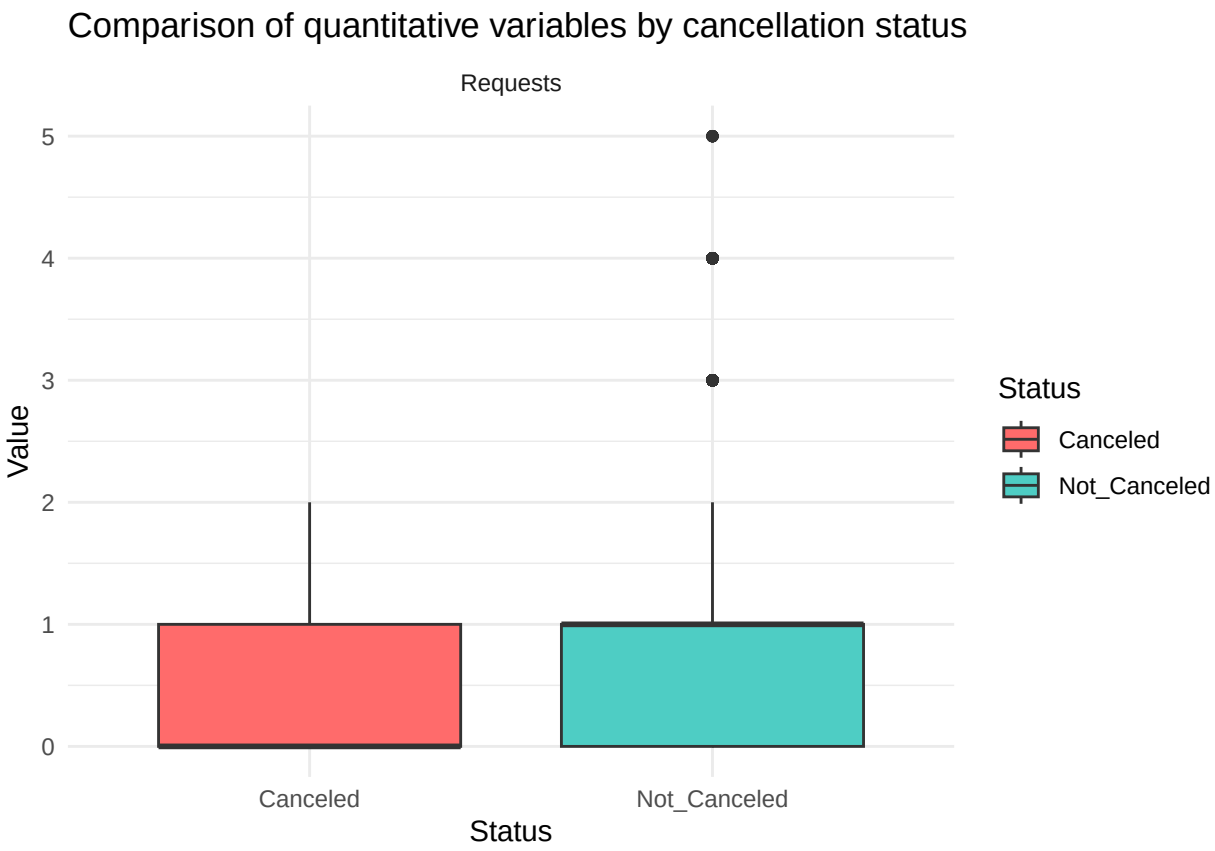


Table 3: Quantitative Explanatory Variables Table 2

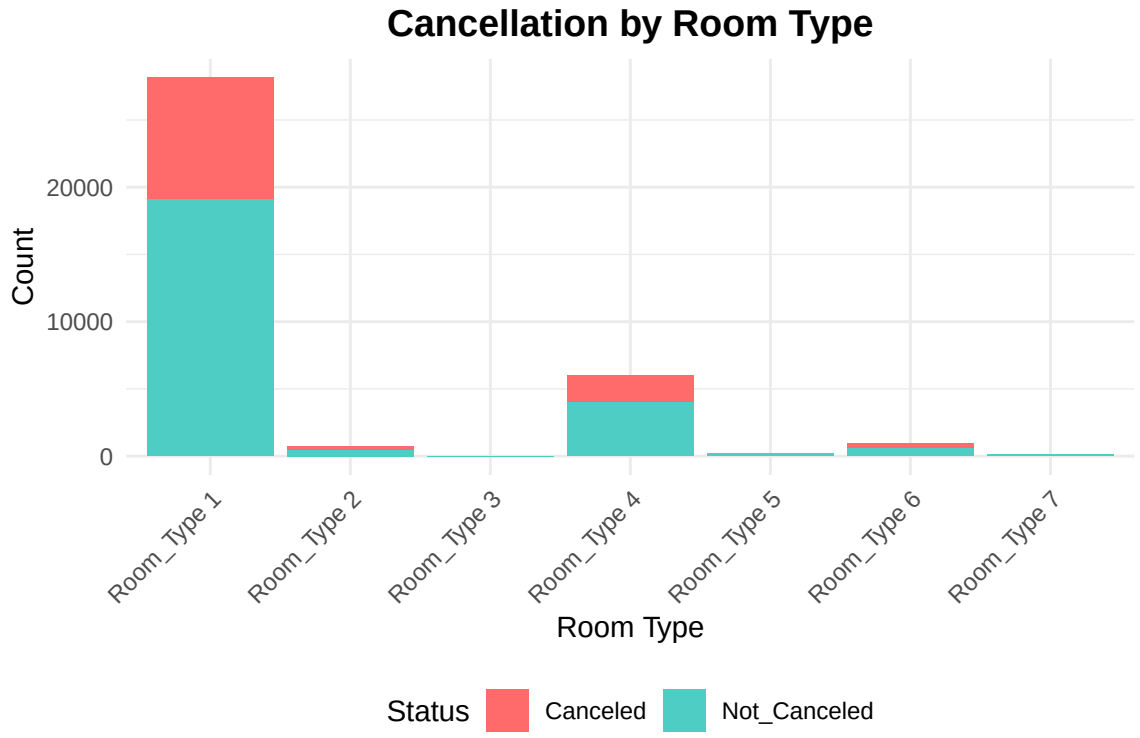
Status	Mean_LeadTime	Mean_AvgPrice	Mean_GroupSize	Mean_StayLength
Canceled	1.909	0.124	0.887	2.392
Not_Canceled	1.813	0.096	0.773	2.113

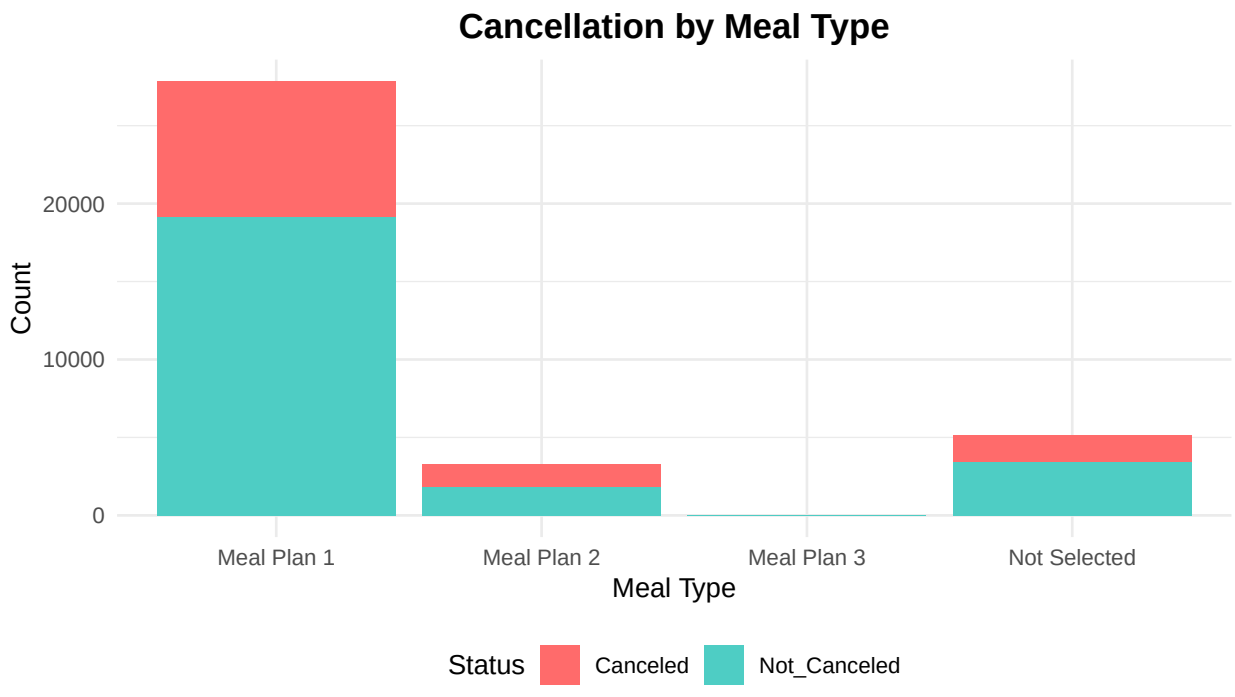


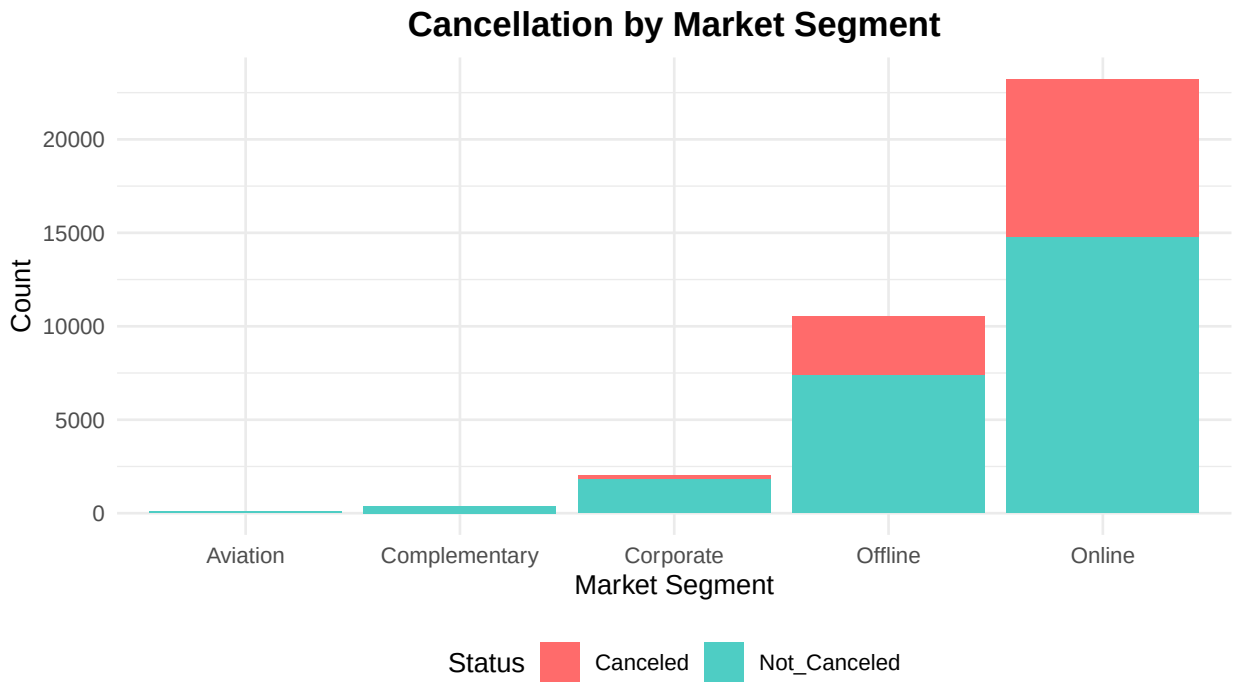

Comparison of quantitative variables by cancellation status
Requests
Value
5
4
3
2
1
0
Canceled
Not_Canceled
Status
Status
Canceled
Not_Canceled

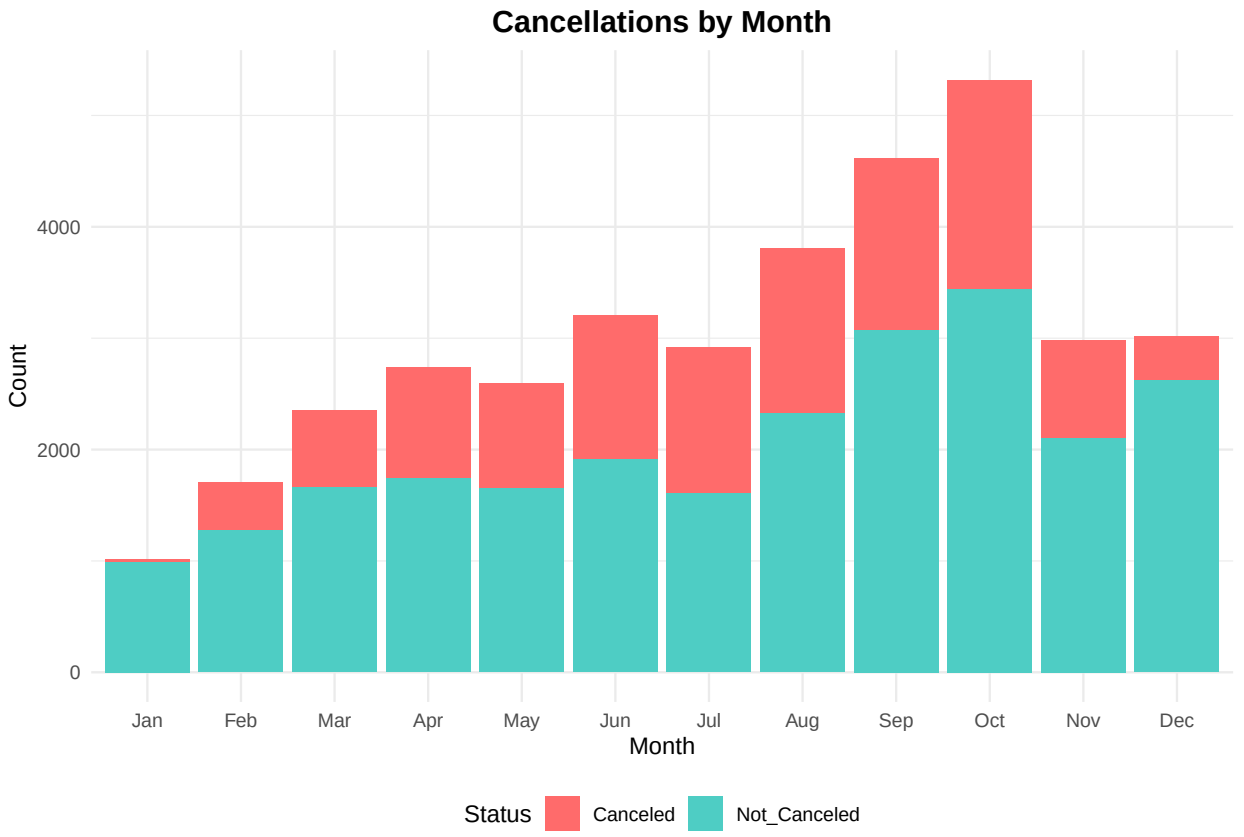
Categorical EDA

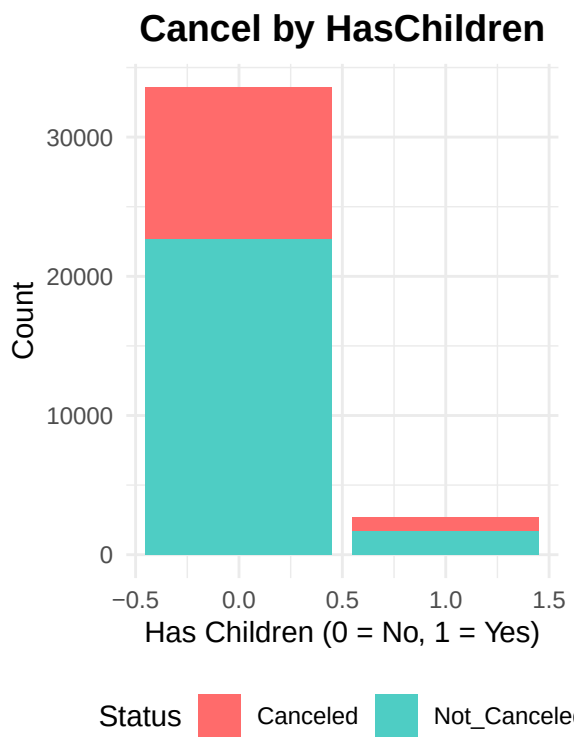
For our categorical exploration, we investigated: RoomType, Meal, Market, Month, HasRequest, Online (new variable where Market = Online), or if the group had children. Looking at the side by side bar plots we have found the variables: HasRequests, Month, Market, and Online to all be individual indicators of cancellation rates. Looking at the visualizations, the best indicators appear to be: HasRequests, Month, and Market. Other important observations include the very low count for many of the room types, low counts of meal plan three, and low counts for the Aviation market type. These limited data points might be an issue if using these categories to create a model, as they might not be sufficient to cover the conditions.

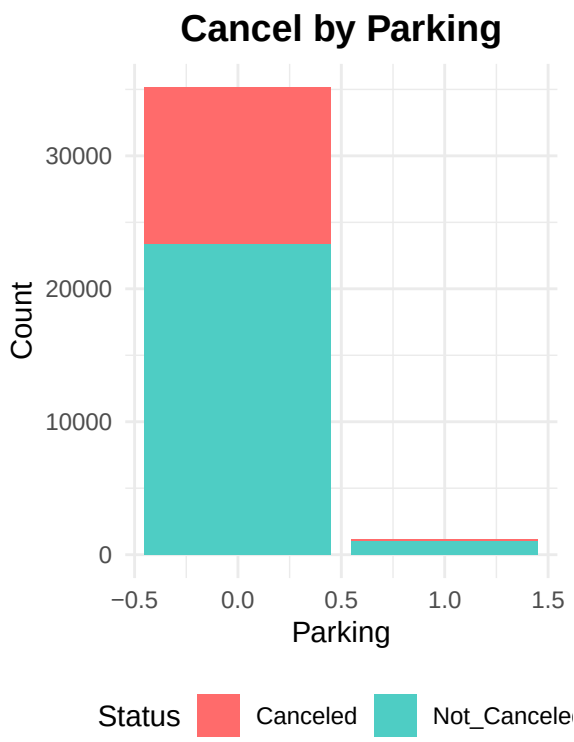


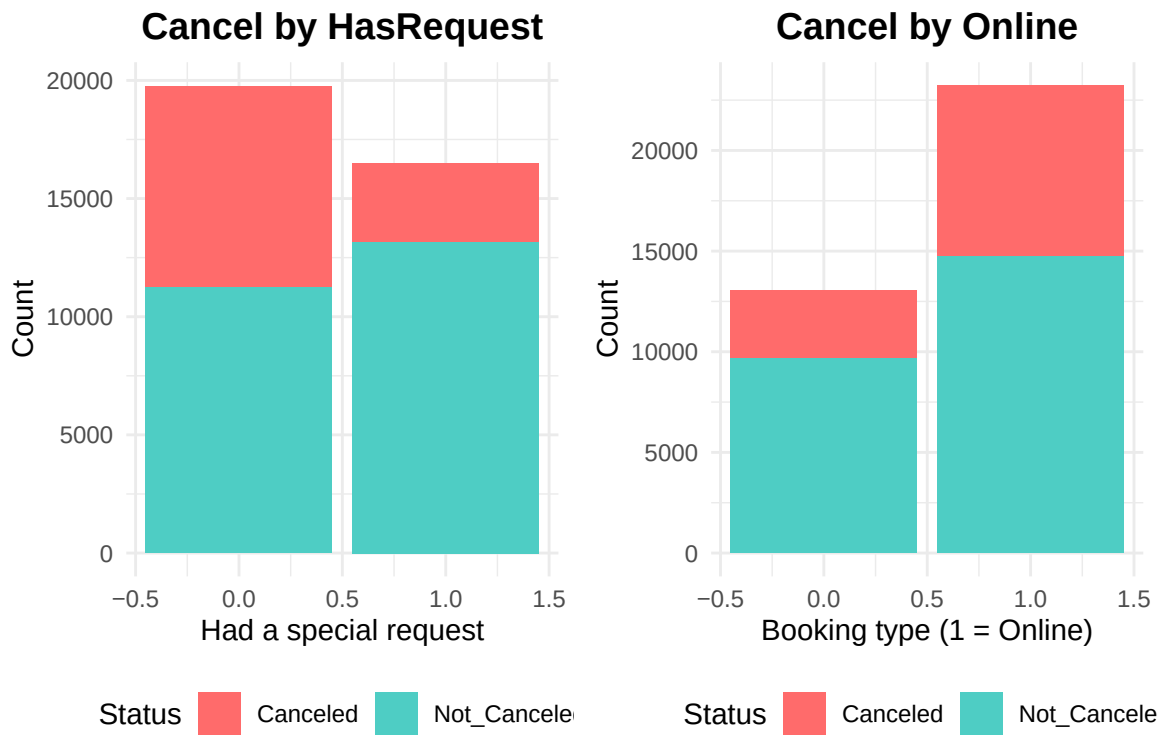

Cancellation by Meal Type
Count
20000
10000
0
Meal Plan 1
Meal Plan 2
Meal Plan 3
Not Selected
Meal Type
Status
Canceled
Not_Canceled


Cancellation by Market Segment
Count
20000
15000
10000
5000
0
Aviation
Complementary
Corporate
Offline
Online
Market Segment
Status
Canceled
Not_Canceled


Cancellations by Month
Count
4000
2000
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
Status
Canceled
Not_Canceled


Cancel by HasChildren
Count
30000
20000
10000
0
−0.5
0.0
0.5
1.0
1.5
Has Children (0 = No, 1 = Yes)
Status
Canceled
Not_Cancele


Cancel by Parking
Count
30000
20000
10000
0
−0.5
0.0
0.5
1.0
1.5
Parking
Status
Canceled
Not_Cancele

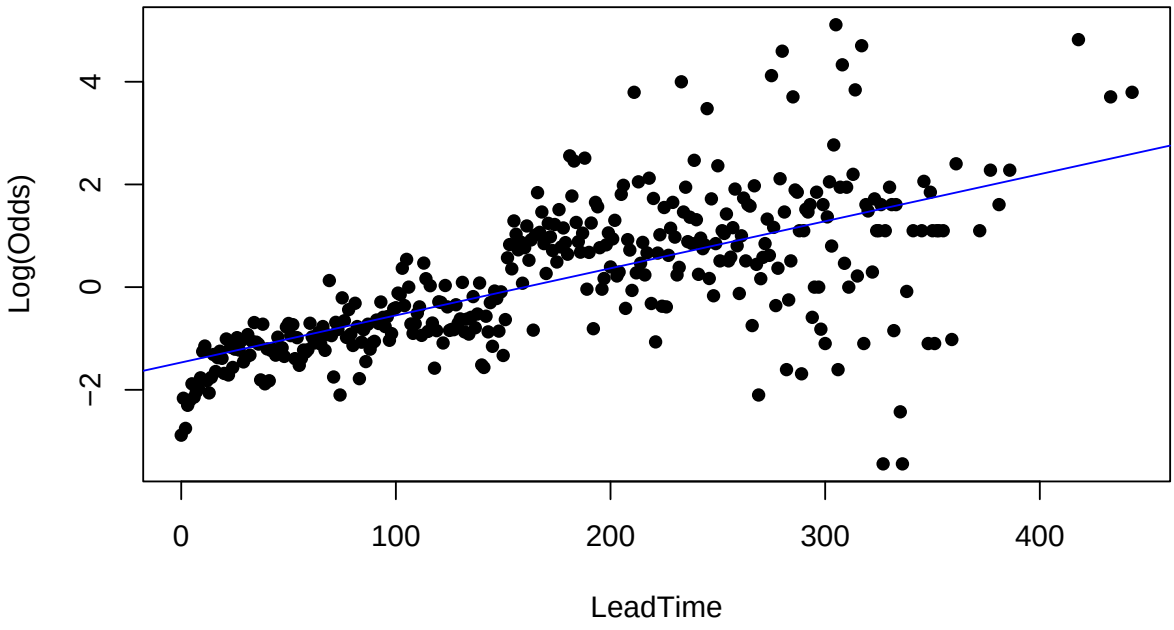


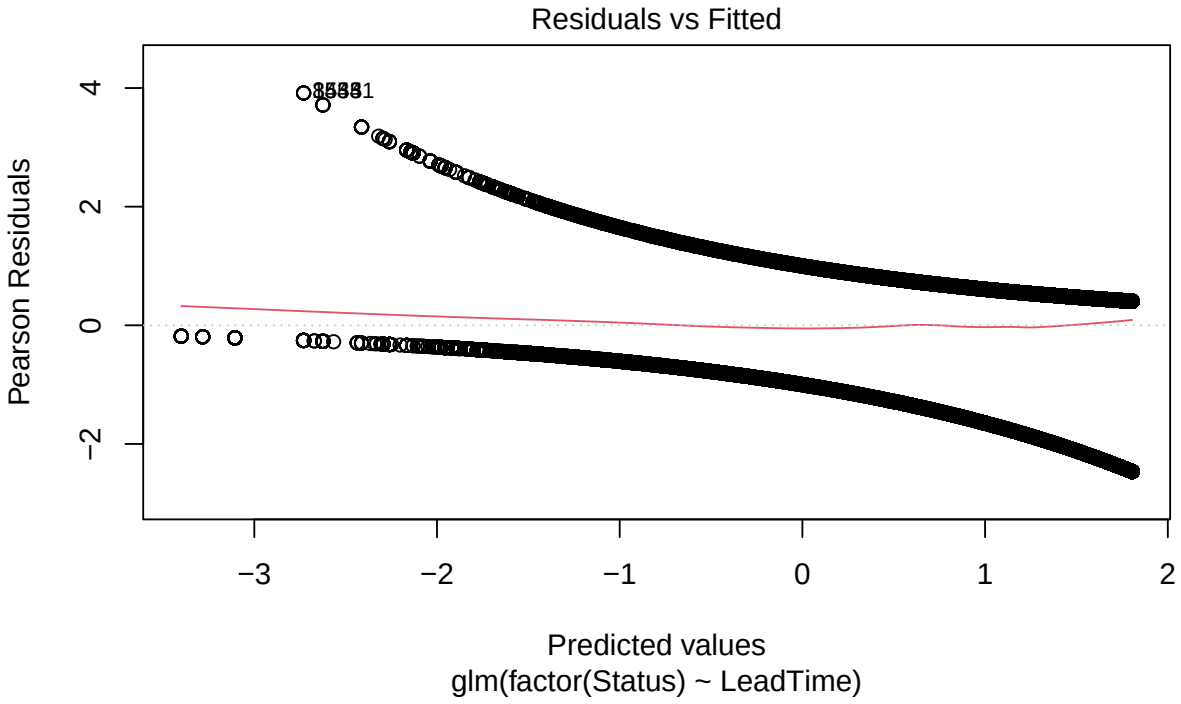
3. Statistical Analysis

3.1 Objective 1

The metrics used to determine the best quantitative predictor were AIC and residual deviance. All of the individual models created from each variable were statistically significant, with most having a p-value less than $2 * 10^{-16}$. However, the AIC and residual deviance for the generalized linear model predicting Status with LeadTime was significantly lower than the models for AvgPrice, StayLength, GroupSize, Children, Adults, Weeknights, Weekdays, and Requests, showing that LeadTime is the greatest individual quantitative predictor for Status. Additional the AIC for Groupsize was lower than both individual models for Children and Adults. In the same way, StayLength was a better indicator than Weekends or Weekdays.

For the conditions of the model using LeadTime, there is slight concern about the equal variance condition. Due to slight flaring of the logit plot, we attempted a variety of transformations (square root, logarithmic, inverse), but this had almost no impact on the conditions and made the model unnecessarily complicated. Additionally there were issues of equal variance for all the quantitative variables. After this analysis, we decided that LeadTime is the best individual quantitative predictor for status.


Log(Odds)
4
2
0
−2
0
100
200
300
400
LeadTime


Residuals vs Fitted
Pearson Residuals
4
2
0
−2
−3
−2
−1
0
1
2
Predicted values
glm(factor(Status) ~ LeadTime)

Quantitative Model

LeadTime has a a p-value of less than $2 * 10^{-16}$, we have strong evidence that LeadTime is a statistically significant predictor for cancellation rates.

Logit function for Status using LeadTime:

$$\widehat{log(\frac{\pi}{1-\pi})} = 1.8046260 - 0.0117484(LeadTime)$$

Using the logit function above, for every 1 day increase in LeadTime, the predicted logit of cancellation rate decreases by 0.011748.

Odds for Status using LeadTime:

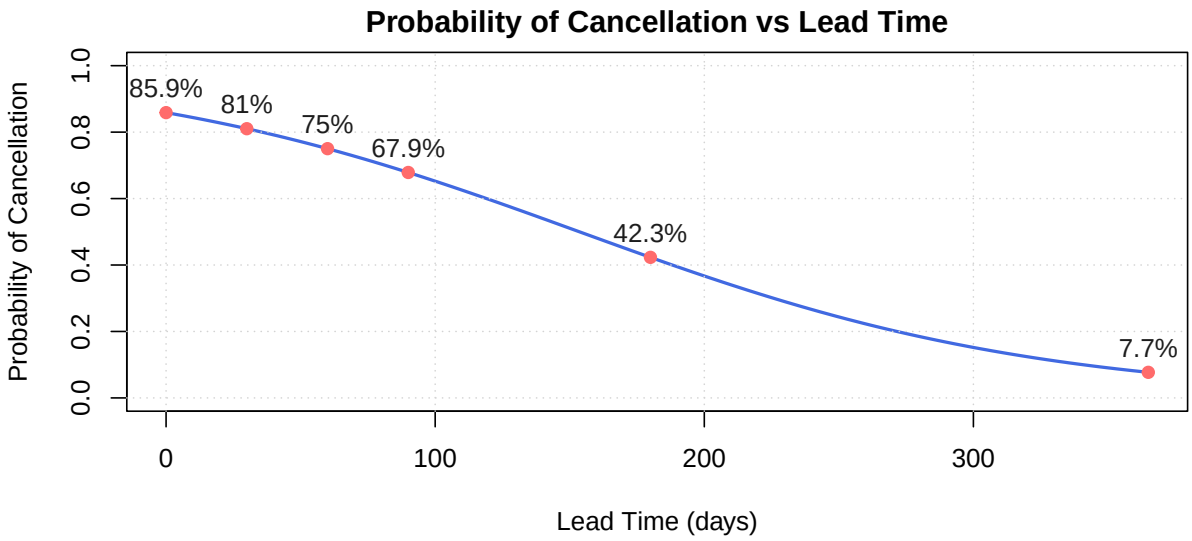
$$odds(Status) = e^{1.8046260 - 0.0117484(LeadTime)}$$

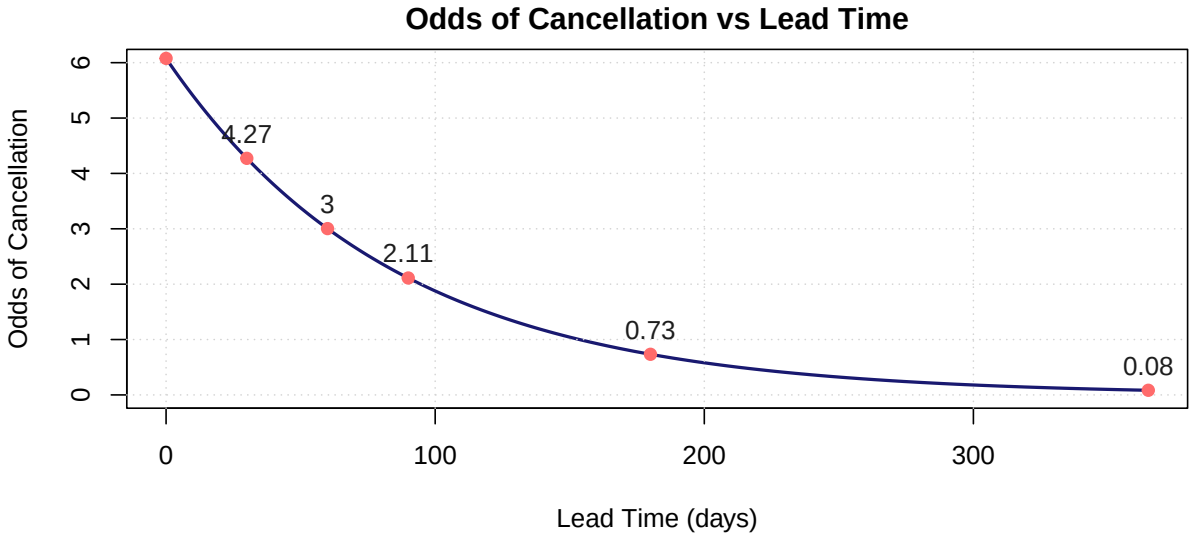
$$oddsratio = e^{\widehat{\beta_1}} = e^{-0.0117484} = 0.988324$$

Therefore, for every 1 day increase in LeadTime, the predicted odds of cancelling the hotel booking decreases by a factor of 0.988. To further Gold Mine Resorts's understanding of these statistics, Visuals of the probabilities and odds of cancellation are included below.

Probability function for Status using LeadTime:

$$p(Status) = \widehat{\pi} = \frac{e^{1.8046260 - 0.0117484(LeadTime)}}{1 - e^{1.8046260 - 0.0117484(LeadTime)}}$$


Probability of Cancellation vs Lead Time
Probability of Cancellation
0.0
0.2
0.4
0.6
0.8
1.0
85.9%
81%
75%
67.9%
42.3%
7.7%
0
100
200
300
Lead Time (days)


Odds of Cancellation vs Lead Time
Odds of Cancellation
0
1
2
3
4
5
6
4.27
3
2.11
0.73
0.08
0
100
200
300
Lead Time (days)

3.2 Objective 2

Looking at the AIC for the categorical models, HasRequests was shown to be the best single categorical explanatory variable for predicting cancellations. This is within the expectations of our exploratory data analysis, as it was also followed by Month and Market.

For our categorical analysis, we transformed Requests (a quantitative variable) into a categorical variable, HasRequests, that indicates whether or not a given booking made a nonzero number of special requests. As evidenced by the model output for HasRequests, we note that the p-value for the variable's coefficient is less than $2 * 10^{-16}$, meaning that HasRequests is statistically significant for predicting cancellation rate. Using a confidence interval, we are 95% certain that for bookings with special requests, the logarithmic likelihood of a cancellation will increase by a factor of between 1.05 and 1.14.

The conditions for our model are met because we have many instances for guests making requests and guests not making requests, as indicated by table 9 (Has Requests) in the appendix.

Logit function for Status using HasRequests:

$$\widehat{log(\frac{\pi}{1-\pi})} = 0.27328 + 1.09756(HasRequests)$$

Probability function for Status using HasRequests:

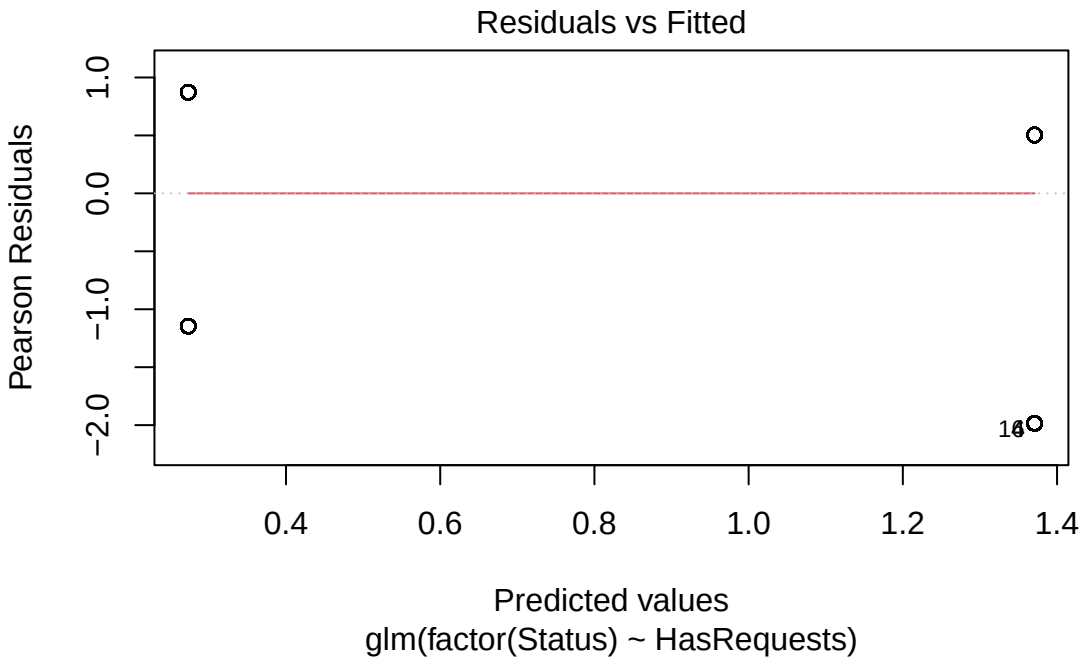
$$p(Status) = \widehat{\pi} = \frac{e^{0.27328+1.09756(HasRequests)}}{1 - e^{0.27328+1.09756(HasRequests)}}$$

Odds for Status using HasRequests:

$$odds(Status) = e^{0.27328+1.09756(HasRequests)}$$

$$oddsratio = e^{\widehat{\beta_1}} = e^{1.09756} = 2.996845$$

HasRequests is a binary categorical variable. When HasRequests = 1 (at least special request was made), the predicted odds of not cancelling are approximately 3 times higher than when HasRequests = 0 (no special requests were made).

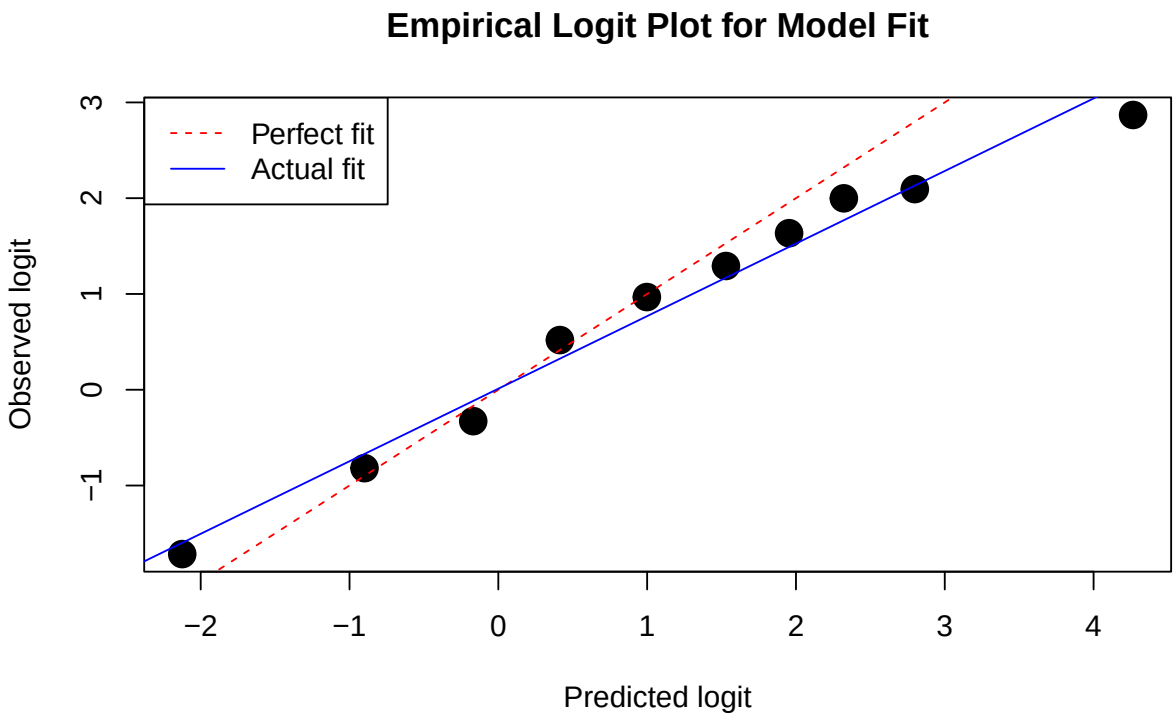

Residuals vs Fitted
Pearson Residuals
1.0
0.0
−1.0
−2.0
0.4
0.6
0.8
1.0
1.2
1.4
Predicted values
glm(factor(Status) ~ HasRequests)

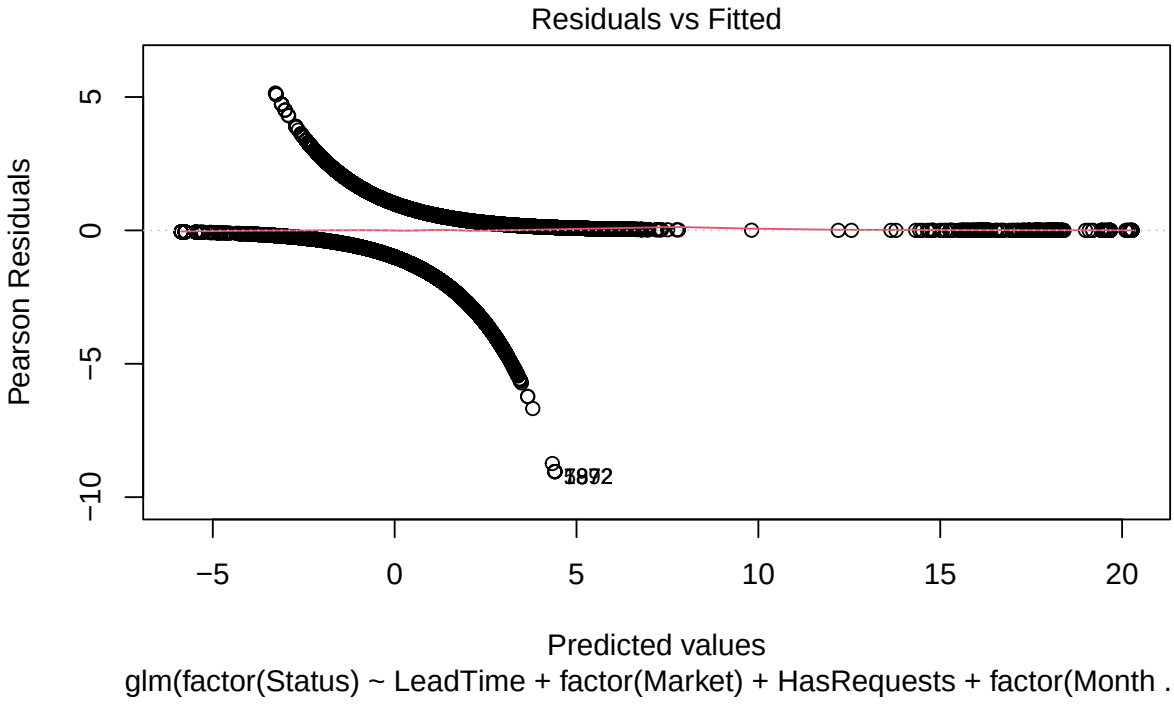
3.3 Objective 3

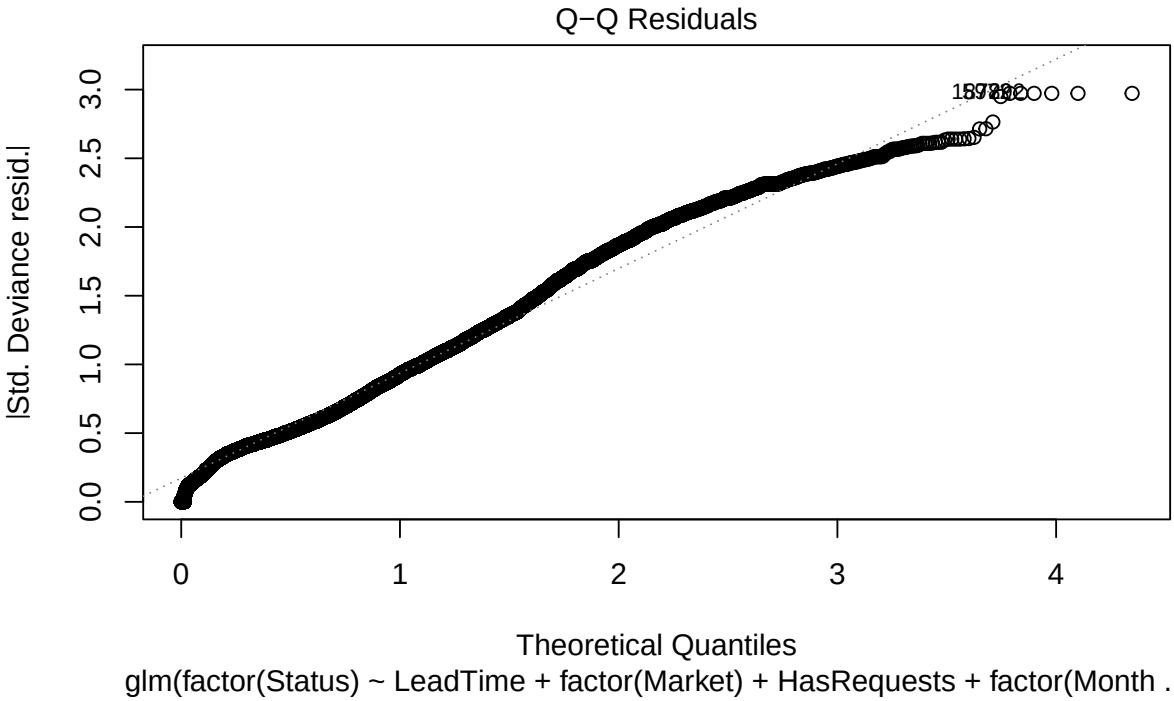
When observing the ANOVA table, the HasChildren variable was not statistically significant. We replaced the HasChildren variable with the Children variable and observed a large p value for this created indicator column as well (greater than .2 in both instances). We additionally removed Year from the statistical model, since this historical data point is not useful for future predictions.

Using forward step analysis, the AIC values recommended using all the remaining variables, including LeadTime, HasRequests, GroupSize, Market, Month, AvgPrice, Parking, RoomType, StayLength, and Meal. The AIC of this model was 30,674.

We used the output from an ANOVA table to eliminate variables one at a time, starting with GroupSize (the highest p value). The final model we selected used LeadTime, Market, HasRequests, Month, AvgPrice, and StayLength. While our AIC increased slightly to 31,109, we reduced the variable count from 10 to 6. Simplifying the model to six variables is far easier to utilize for improvements to Gold Mine Resorts marketing procedures, and only changed the AIC value by 1.4%. Additionally, our model had good variance and normality, in addition to fitting closely to the expected results (as indicated by the empirical logit plot).




Residuals vs Fitted
Pearson Residuals
5
0
−5
−10
−5
0
5
10
15
20
Predicted values
glm(factor(Status) ~ LeadTime + factor(Market) + HasRequests + factor(Month ...


Q−Q Residuals
|Std. Deviance resid.|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
Theoretical Quantiles
glm(factor(Status) ~ LeadTime + factor(Market) + HasRequests + factor(Month ...

Cross Validation

After running two cross validation tests, we observed the percent difference in AIC to be -.326% and .753%. This metric shows our model does not have any indicators for overfitting.

Key Quantitative Predictors:*

To summarize, the business implication is that longer lead times, higher prices, and longer stays correlate with higher cancellation risk. In addition, LeadTime, AvgPrice, and StayLength have respective p-values of <2e-16, <2e-16, and 7.49e-16, indicating that these variables are useful for predicting cancellation rates.

$$oddsratio_{LeadTime} = e^{-0.0161356} = 0.984$$

For each additional day in lead time, the predicted odds of not cancelling decrease by a factor of 0.984, holding all other variables constant.

$$oddsratio_{AvgPrice} = e^{-0.0145549} = 0.985$$

For each $1 increase in average price, the predicted odds of not cancelling decrease by a factor of 0.985, holding all other variables constant.

$$oddsratio_{StayTime} = e^{-0.0643062} = 0.938$$

For each additional night in the hotel, the predicted odds of not cancelling decrease by a factor of 0.938, holding all other variables constant.

Key Categorical Predictors:

To summarize, the business implication is that corporate/offline bookings are more reliable than aviation bookings. In addition, the months of February to November have a higher cancellation risk than January. Specifically, the peak cancellation risk occurs in the month of February where the predicted odds of not cancelling decrease by a factor of 0.059 compared to January (holding all other variables constant).

$$oddsratio_{HasRequests} = e^{2.0247462} = 7.57$$

Holding all other variables constant, when a guest makes at least one special request the predicted odds of not cancelling are approximately 8 times higher than those without special requests. With a p-value <2e-16, this variable is useful to predict Status.

$$oddsratio_{Market(Corporate)} = e^{1.0866028} = 2.96$$

Holding all other variables constant, when Gold Mine Resorts has a corporate booking the predicted odds of not cancelling are approximately 3 times higher than an aviation booking (baseline). With a p-value of 3.95e-07, this segment is useful in predicting Status.

$$oddsratio_{Market(Offline)} = e^{1.6324226} = 5.12$$

Holding all other variables constant, when Gold Mine Resorts has an offline booking the predicted odds of not cancelling are approximately 5 times higher than an aviation booking (baseline). With a p-value of 8.80e-16, this segment is useful in predicting Status.

$$oddsratio_{Market(Online)} = e^{-0.2199951} = 0.8$$

Holding all other variables constant, when Gold Mine Resorts has an online booking the predicted odds of not cancelling decrease by a factor of 0.8 compared to an aviation booking (baseline). However, with a p-value of 0.27308, this segment is not statistically significant in predicting Status.

$$oddsratio_{Month(Feb-Nov)} \approx e^{-2.28} = 0.102$$

Holding all other variables constant, when there is a booking made between February to November (all β's are relatively similar in range) the predicted odds of not cancelling decrease by a factor of 0.102 when compared to bookings made in January (baseline). Each individual month is statistically significant in predicting Status with p-values <2e-16.

$$oddsratio_{Month(Dec)} \approx e^{-0.0643} = 0.55$$

Holding all other variables constant, when there is a booking made in December the predicted odds of not cancelling decrease by a factor of 0.55 when compared to bookings made in January (baseline). December is a statistically significant segment of Month in predicting Status with a p-value of 0.00674.

4. Conclusions

Objective 1:

Our final conclusion is that LeadTime is the best single numeric predictor for cancellation rates, even when considering new variables synthesized from the dataset. For every one day increase in lead time, the predicted odds of cancelling the hotel booking decreases slightly. Guests with very large lead times are the most likely to keep their reservation.

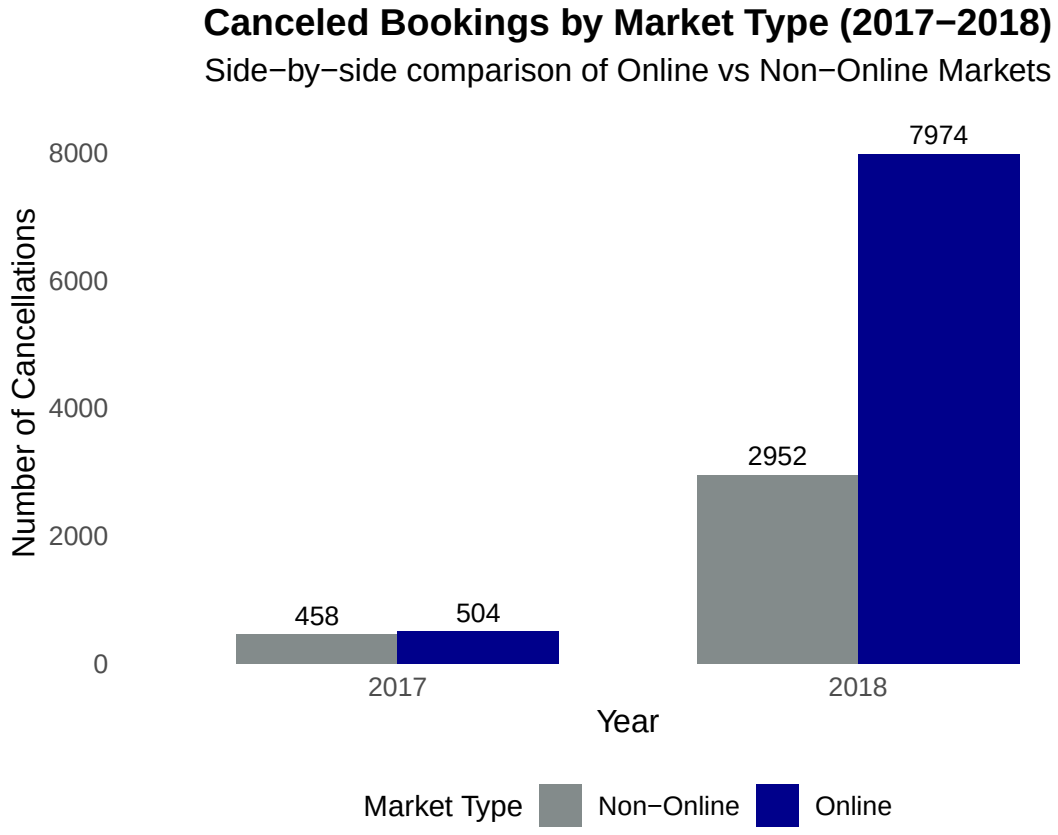
Objective 2:

When considering only the categories of factors available from the dataset or synthesized, we found that HasRequests is the most useful variable for predicting cancellation rates. If a guest has a reservation request, the predicted of odds of a cancellation decreases by a factor of three.

Objective 3:

With all of the original and synthesized variables available for our final model, we determined the most useful model for predicting cancellation rates included the predictors of LeadTime, Market, Month, AvgPrice, and StayLength. Gold Mine Resorts should focus on these factors to minimize cancellations in the future.

While this model is a good starting point for predicting hotel cancellations, we noticed that there was a significant increase in total reservations between 2017 and 2018. Further, 2018 had a significant spike in online reservations that led to cancellations. If Gold Mine Resorts has access to further details about the sources for online bookings, this may prove to be a beneficial research study in the future. Other factors impacting the economy between 2017 and 2018 may also be useful for predicting cancellations.



5. Appendix

Below are tables used in our exploratory data analysis, followed by the statistical output of our models and cross validation.

Data Tables

Table 4: Room Type

RoomType	Status	count	proportion
Room_Type 1	Canceled	9076	0.323
Room_Type 1	Not_Canceled	19062	0.677
Room_Type 2	Canceled	228	0.329
Room_Type 2	Not_Canceled	464	0.671
Room_Type 3	Canceled	2	0.286
Room_Type 3	Not_Canceled	5	0.714
Room_Type 4	Canceled	2069	0.341
Room_Type 4	Not_Canceled	3990	0.659
Room_Type 5	Canceled	72	0.272
Room_Type 5	Not_Canceled	193	0.728
Room_Type 6	Canceled	406	0.420
Room_Type 6	Not_Canceled	560	0.580
Room_Type 7	Canceled	36	0.228
Room_Type 7	Not_Canceled	122	0.772

Table 5: Meal Option

Meal	Status	count	proportion
Meal Plan 1	Canceled	8681	0.312
Meal Plan 1	Not_Canceled	19161	0.688
Meal Plan 2	Canceled	1507	0.456
Meal Plan 2	Not_Canceled	1799	0.544
Meal Plan 3	Canceled	1	0.200
Meal Plan 3	Not_Canceled	4	0.800
Not Selected	Canceled	1700	0.331
Not Selected	Not_Canceled	3432	0.669

Table 6: Market

Market	Status	count	proportion
Aviation	Canceled	37	0.296
Aviation	Not_Canceled	88	0.704
Complementary	Not_Canceled	391	1.000
Corporate	Canceled	220	0.109
Corporate	Not_Canceled	1797	0.891
Offline	Canceled	3154	0.299
Offline	Not_Canceled	7377	0.701
Online	Canceled	8478	0.365
Online	Not_Canceled	14743	0.635

Table 7: Parking

Parking	Status	count	proportion
0	Canceled	11775	0.335
0	Not_Canceled	23386	0.665
1	Canceled	114	0.101
1	Not_Canceled	1010	0.899

Table 8: Month

Month	Status	count	proportion
Jan	Canceled	24	0.024
Jan	Not_Canceled	990	0.976
Feb	Canceled	431	0.253
Feb	Not_Canceled	1274	0.747
Mar	Canceled	700	0.297
Mar	Not_Canceled	1658	0.703
Apr	Canceled	996	0.364
Apr	Not_Canceled	1741	0.636
May	Canceled	949	0.365
May	Not_Canceled	1650	0.635
Jun	Canceled	1291	0.403
Jun	Not_Canceled	1912	0.597
Jul	Canceled	1314	0.450
Jul	Not_Canceled	1607	0.550
Aug	Canceled	1488	0.390
Aug	Not_Canceled	2325	0.610
Sep	Canceled	1539	0.334
Sep	Not_Canceled	3073	0.666
Oct	Canceled	1880	0.353
Oct	Not_Canceled	3440	0.647
Nov	Canceled	875	0.294
Nov	Not_Canceled	2106	0.706
Dec	Canceled	402	0.133
Dec	Not_Canceled	2620	0.867

Table 9: Has Requests

HasRequests	Status	count	proportion
0	Canceled	8547	0.432
0	Not_Canceled	11233	0.568
1	Canceled	3342	0.202
1	Not_Canceled	13163	0.798

Table 10: Has Children

HasChildren	Status	count	proportion
0	Canceled	10885	0.324
0	Not_Canceled	22698	0.676
1	Canceled	1004	0.372
1	Not_Canceled	1698	0.628

Table 11: Online Booking

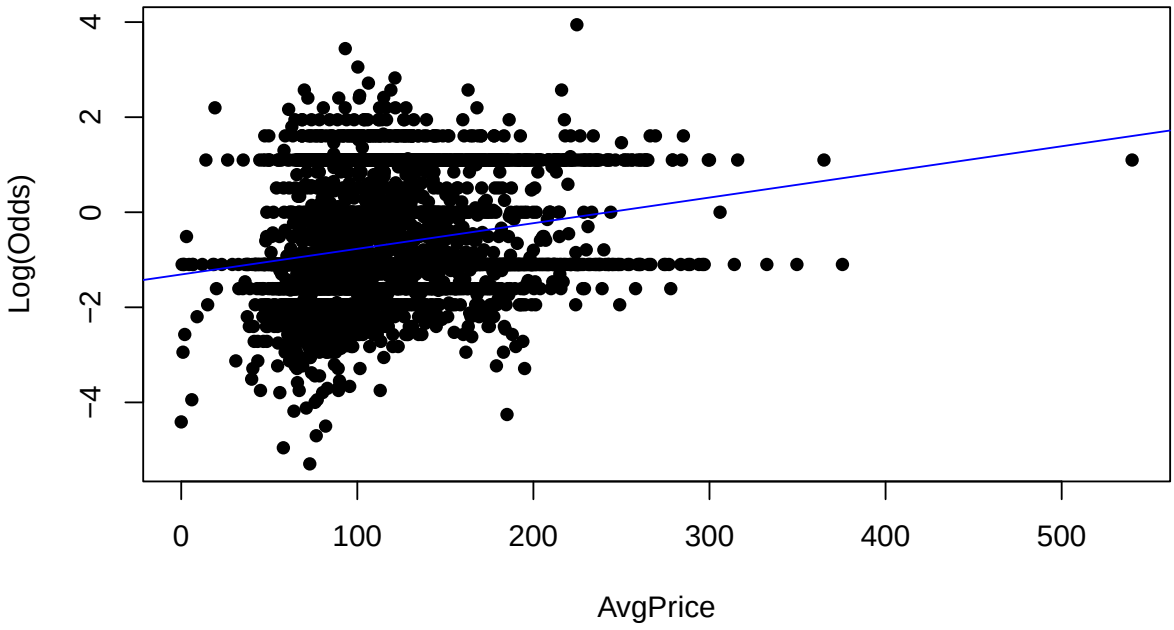
Online	Status	count	proportion
0	Canceled	3411	0.261
0	Not_Canceled	9653	0.739
1	Canceled	8478	0.365
1	Not_Canceled	14743	0.635

Table 12: Requests

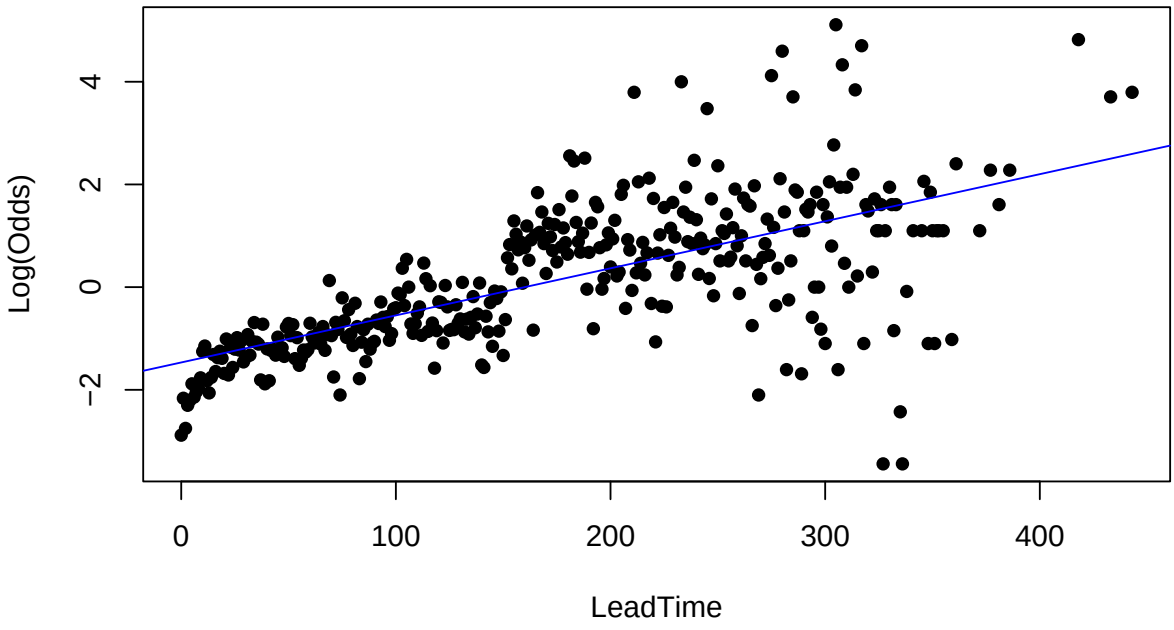
Requests	Status	count	proportion
0	Canceled	8547	0.432
0	Not_Canceled	11233	0.568
1	Canceled	2705	0.238
1	Not_Canceled	8674	0.762
2	Canceled	637	0.146
2	Not_Canceled	3727	0.854
3	Not_Canceled	676	1.000
4	Not_Canceled	78	1.000
5	Not_Canceled	8	1.000

Quantitative Variable Models

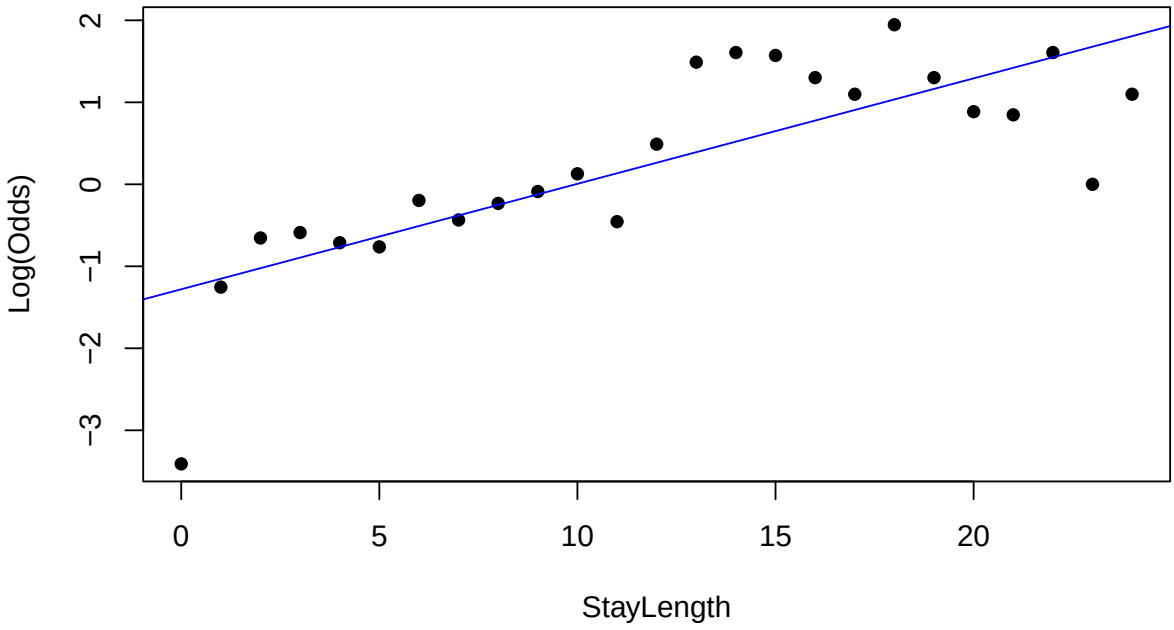
```
##
## Call:
## glm(formula = factor(Status) ~ AvgPrice, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.6315518  0.0365165   44.68   <2e-16 ***
## AvgPrice    -0.0086872  0.0003261  -26.64   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 45167  on 36283  degrees of freedom
## AIC: 45171
##
## Number of Fisher Scoring iterations: 4
```



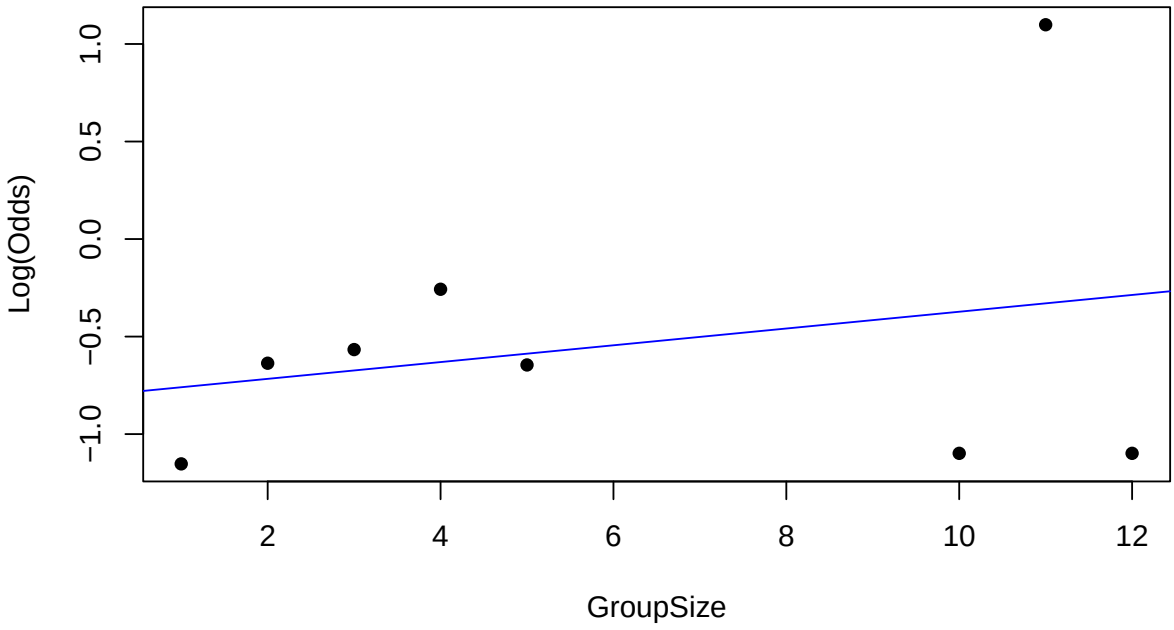

```
##
## Call:
## glm(formula = factor(Status) ~ LeadTime, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.8046260  0.0194124   92.96   <2e-16 ***
## LeadTime    -0.0117484  0.0001593  -73.73   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 38894  on 36283  degrees of freedom
## AIC: 38898
##
## Number of Fisher Scoring iterations: 4
```



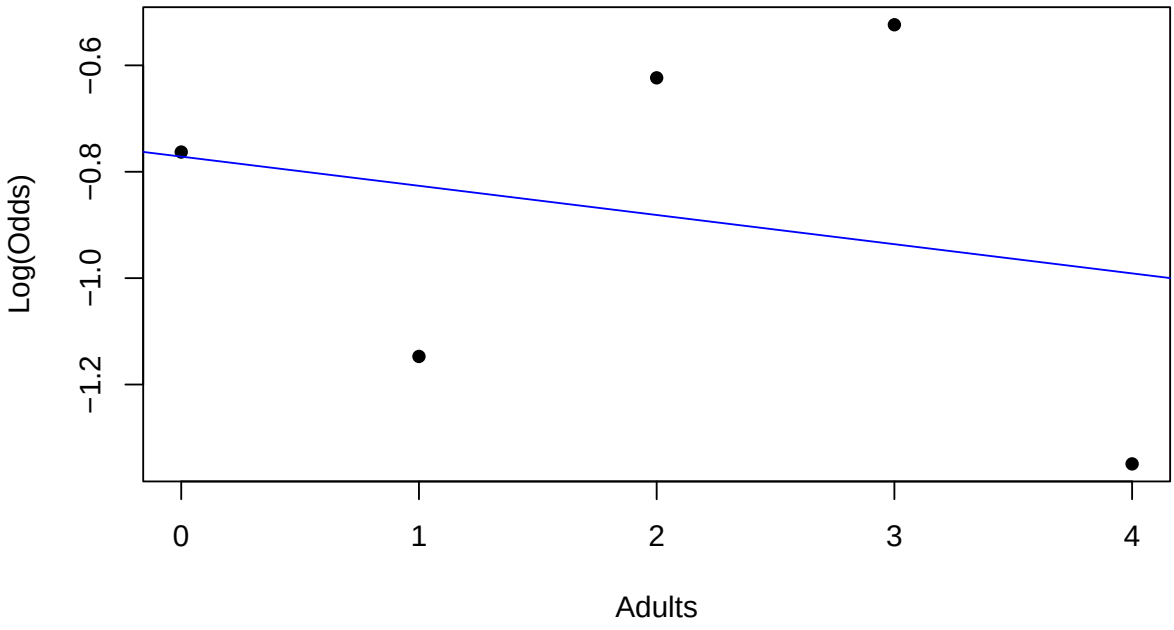

```
##
## Call:
## glm(formula = factor(Status) ~ StayLength, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.086290   0.022432   48.43   <2e-16 ***
## StayLength  -0.119786   0.006249  -19.17   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 45524  on 36283  degrees of freedom
## AIC: 45528
##
## Number of Fisher Scoring iterations: 4
```



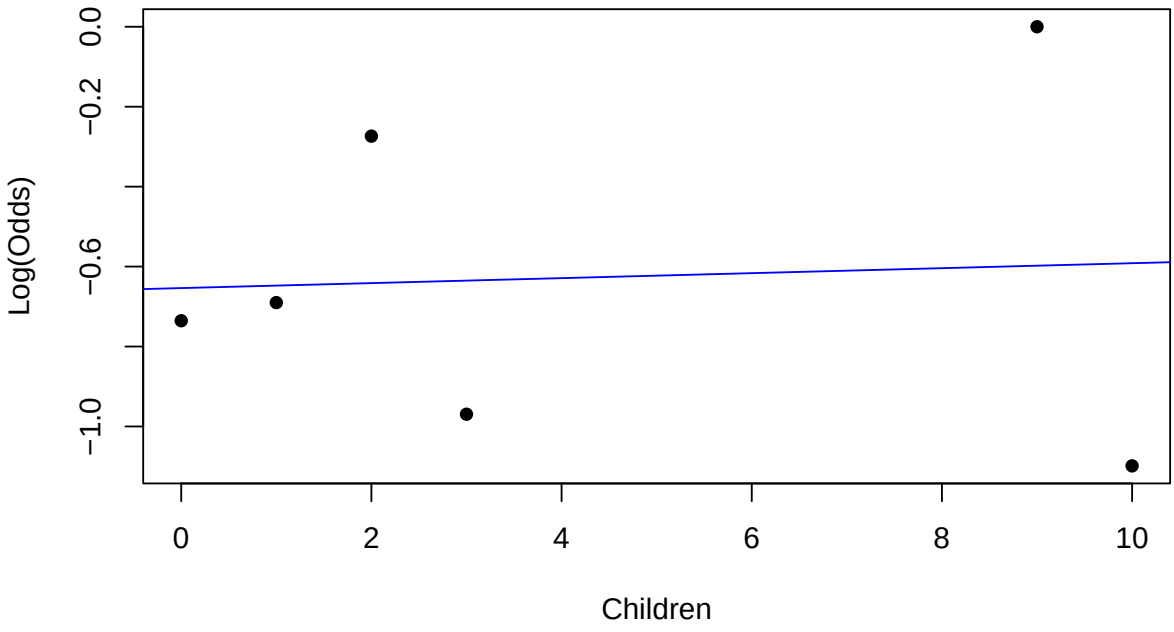

```
##
## Call:
## glm(formula = factor(Status) ~ GroupSize, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.29287    0.03602   35.89   <2e-16 ***
## GroupSize   -0.29141    0.01721  -16.93   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 45612  on 36283  degrees of freedom
## AIC: 45616
##
## Number of Fisher Scoring iterations: 4
```



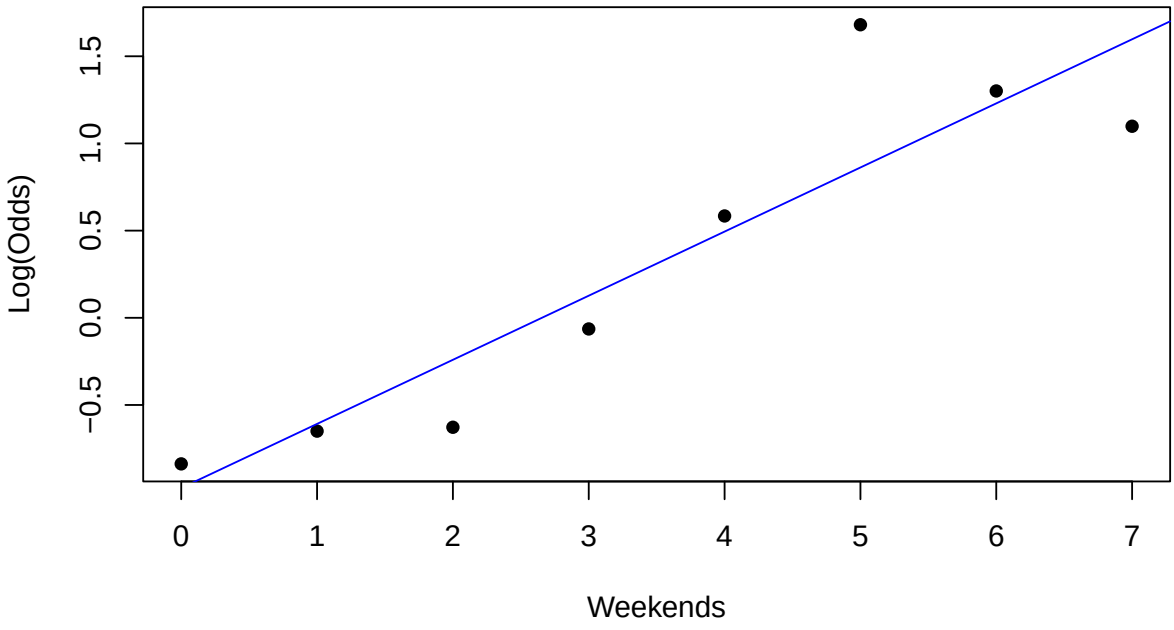

```
##
## Call:
## glm(formula = factor(Status) ~ Adults, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.39588    0.04303   32.44   <2e-16 ***
## Adults      -0.36366    0.02211  -16.45   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 45624  on 36283  degrees of freedom
## AIC: 45628
##
## Number of Fisher Scoring iterations: 4
```



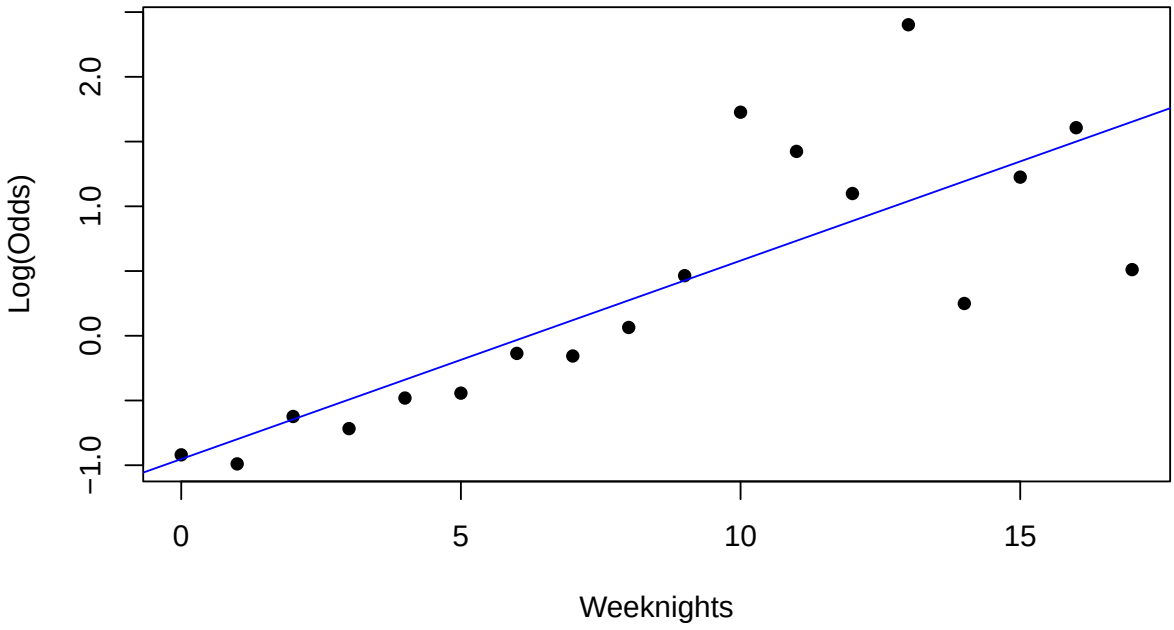

```
##
## Call:
## glm(formula = factor(Status) ~ Children, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.73716    0.01160  63.573  < 2e-16 ***
## Children    -0.16764    0.02686  -6.241 4.34e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 45863  on 36283  degrees of freedom
## AIC: 45867
##
## Number of Fisher Scoring iterations: 4
```



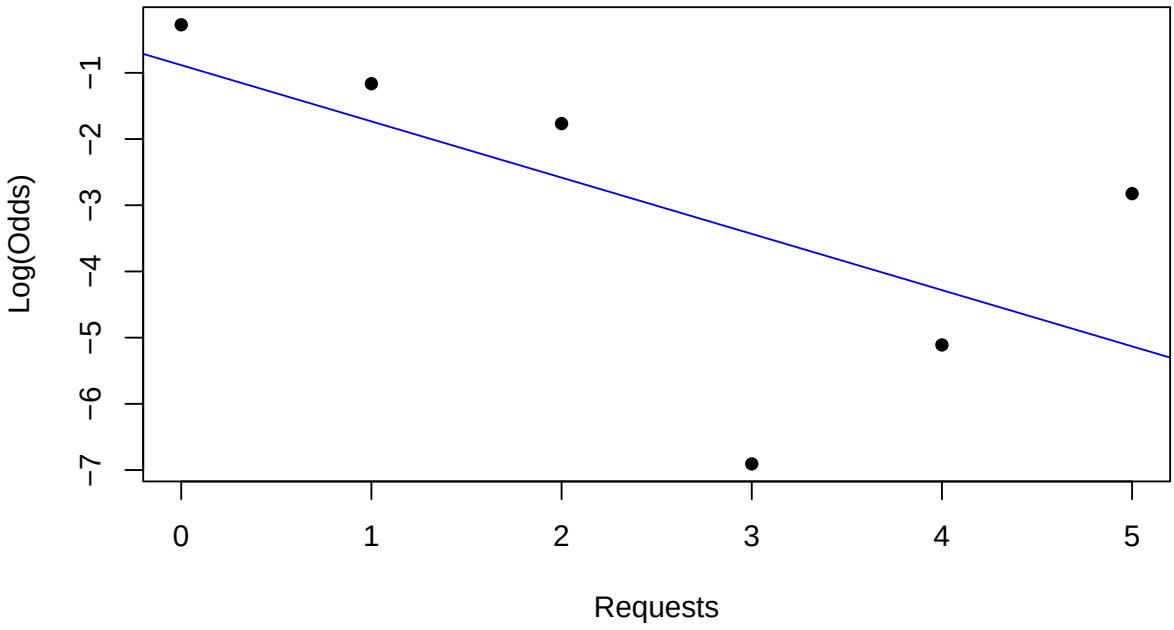

```
##
## Call:
## glm(formula = factor(Status) ~ Weekends, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.84213    0.01555   54.15   <2e-16 ***
## Weekends    -0.14869    0.01272  -11.69   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 45765  on 36283  degrees of freedom
## AIC: 45769
##
## Number of Fisher Scoring iterations: 4
```




```
##
## Call:
## glm(formula = factor(Status) ~ Weeknights, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.02668    0.02119   48.46   <2e-16 ***
## Weeknights  -0.13710    0.00788  -17.40   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 45595  on 36283  degrees of freedom
## AIC: 45599
##
## Number of Fisher Scoring iterations: 4
```




```
##
## Call:
## glm(formula = factor(Status) ~ Requests, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.27765    0.01392   19.95   <2e-16 ***
## Requests     0.84692    0.01826   46.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 43272  on 36283  degrees of freedom
## AIC: 43276
##
## Number of Fisher Scoring iterations: 4
```


Log(Odds)
−1
−2
−3
−4
−5
−6
−7
0
1
2
3
4
5
Requests

Categorical Variable Models

```
##
## Call:
## glm(formula = factor(Status) ~ RoomType, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)           0.74206    0.01275  58.187  < 2e-16 ***
## RoomTypeRoom_Type 2 -0.03152    0.08188  -0.385  0.70022
## RoomTypeRoom_Type 3  0.17423    0.83676   0.208  0.83506
## RoomTypeRoom_Type 4 -0.08534    0.02994  -2.850  0.00437 **
## RoomTypeRoom_Type 5  0.24396    0.13868   1.759  0.07856 .
## RoomTypeRoom_Type 6 -0.42048    0.06642  -6.331 2.44e-10 ***
## RoomTypeRoom_Type 7  0.47844    0.19010   2.517  0.01184 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 45845  on 36278  degrees of freedom
## AIC: 45859
##
## Number of Fisher Scoring iterations: 4
```

```
##
## Call:
## glm(formula = factor(Status) ~ Parking, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.68616    0.01130   60.72   <2e-16 ***
## Parking      1.49535    0.09944   15.04   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 45574  on 36283  degrees of freedom
## AIC: 45578
##
## Number of Fisher Scoring iterations: 4
```

```
##
## Call:
## glm(formula = factor(Status) ~ Market, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
```

```
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.86642    0.19594   4.422 9.78e-06 ***
## MarketComplementary 14.69965   73.60288   0.200    0.842
## MarketCorporate      1.23383    0.20855   5.916 3.29e-09 ***
## MarketOffline       -0.01672    0.19709  -0.085    0.932
## MarketOnline        -0.31313    0.19641  -1.594    0.111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 44879  on 36280  degrees of freedom
## AIC: 44889
##
## Number of Fisher Scoring iterations: 14

##
## Call:
## glm(formula = factor(Status) ~ Meal, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        0.79174    0.01294  61.197  < 2e-16 ***
## MealMeal Plan 2   -0.61463    0.03724 -16.505  < 2e-16 ***
## MealMeal Plan 3    0.59455    1.11811   0.532  0.59490
## MealNot Selected -0.08923    0.03236  -2.758  0.00582 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 45634  on 36281  degrees of freedom
## AIC: 45642
##
## Number of Fisher Scoring iterations: 4

##
## Call:
## glm(formula = factor(Status) ~ Month, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.7197     0.2066  18.006   <2e-16 ***
## MonthFeb     -2.6358     0.2140 -12.319   <2e-16 ***
## MonthMar     -2.8574     0.2114 -13.514   <2e-16 ***
## MonthApr     -3.1612     0.2104 -15.027   <2e-16 ***
## MonthMay     -3.1665     0.2106 -15.039   <2e-16 ***
## MonthJun     -3.3269     0.2097 -15.865   <2e-16 ***
## MonthJul     -3.5184     0.2099 -16.762   <2e-16 ***
## MonthAug     -3.2734     0.2092 -15.645   <2e-16 ***
```

```
## MonthSep     -3.0281     0.2089 -14.494   <2e-16 ***
## MonthOct     -3.1155     0.2086 -14.938   <2e-16 ***
## MonthNov     -2.8413     0.2105 -13.501   <2e-16 ***
## MonthDec     -1.8452     0.2134  -8.646   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 44226  on 36273  degrees of freedom
## AIC: 44250
##
## Number of Fisher Scoring iterations: 6
```

```
##
## Call:
## glm(formula = factor(Status) ~ HasRequests, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.27328    0.01435   19.04   <2e-16 ***
## HasRequests  1.09756    0.02411   45.53   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 43686  on 36283  degrees of freedom
## AIC: 43690
##
## Number of Fisher Scoring iterations: 4
```

```
##
## Call:
## glm(formula = factor(Status) ~ HasChildren, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.73489    0.01166  63.033  < 2e-16 ***
## HasChildren -0.20943    0.04148  -5.049 4.45e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 45901  on 36284  degrees of freedom
## Residual deviance: 45876  on 36283  degrees of freedom
## AIC: 45880
##
## Number of Fisher Scoring iterations: 4
```

Combined Variable and Final Models

```
## Start:  AIC=45902.98
## factor(Status) ~ 1
##
##                     Df Deviance   AIC
## + LeadTime           1    38894 38898
## + HasRequests        1    43686 43690
## + factor(Month)     11    44226 44250
## + factor(Market)     4    44879 44889
## + AvgPrice           1    45167 45171
## + StayLength         1    45524 45528
## + factor(Parking)    1    45574 45578
## + GroupSize          1    45612 45616
## + Adults             1    45624 45628
## + factor(Meal)       3    45634 45642
## + factor(RoomType)   6    45845 45859
## + Children           1    45863 45867
## <none>                    45901 45903
##
## Step:  AIC=38898.17
## factor(Status) ~ LeadTime
##
##                     Df Deviance   AIC
## + HasRequests        1    37078 37084
## + factor(Market)     4    37267 37279
## + AvgPrice           1    37491 37497
## + factor(Month)     11    37671 37697
## + factor(RoomType)   6    38630 38646
## + GroupSize          1    38661 38667
## + factor(Parking)    1    38697 38703
## + factor(Meal)       3    38750 38760
## + Children           1    38773 38779
## + Adults             1    38787 38793
## + StayLength         1    38812 38818
## <none>                    38894 38898
##
## Step:  AIC=37084.44
## factor(Status) ~ LeadTime + HasRequests
##
##                     Df Deviance   AIC
## + factor(Market)     4    33257 33271
## + AvgPrice           1    35004 35012
## + factor(Month)     11    35794 35822
## + GroupSize          1    36436 36444
## + factor(RoomType)   6    36591 36609
## + Adults             1    36701 36709
## + factor(Meal)       3    36829 36841
## + Children           1    36840 36848
## + StayLength         1    36909 36917
## + factor(Parking)    1    36970 36978
## <none>                    37078 37084
##
## Step:  AIC=33271.15
```

```
## factor(Status) ~ LeadTime + HasRequests + factor(Market)
##
##                     Df Deviance   AIC
## + factor(Month)    11    32063 32099
## + AvgPrice          1    32215 32231
## + factor(Parking)   1    33083 33099
## + GroupSize         1    33119 33135
## + factor(RoomType)  6    33133 33159
## + factor(Meal)      3    33162 33182
## + Adults            1    33189 33205
## + Children          1    33198 33214
## + StayLength        1    33230 33246
## <none>                   33257 33271
##
## Step:  AIC=32098.81
## factor(Status) ~ LeadTime + HasRequests + factor(Market) + factor(Month)
##
##                     Df Deviance   AIC
## + AvgPrice          1    31134 31172
## + factor(Parking)   1    31885 31923
## + GroupSize         1    31899 31937
## + factor(RoomType)  6    31949 31997
## + factor(Meal)      3    31967 32009
## + Adults            1    31988 32026
## + Children          1    31989 32027
## + StayLength        1    32014 32052
## <none>                   32063 32099
##
## Step:  AIC=31172.45
## factor(Status) ~ LeadTime + HasRequests + factor(Market) + factor(Month) +
##     AvgPrice
##
##                     Df Deviance   AIC
## + factor(Parking)   1    30890 30930
## + factor(RoomType)  6    31024 31074
## + factor(Meal)      3    31043 31087
## + StayLength        1    31069 31109
## + Children          1    31124 31164
## + Adults            1    31132 31172
## <none>                   31135 31173
## + GroupSize         1    31134 31174
##
## Step:  AIC=30929.5
## factor(Status) ~ LeadTime + HasRequests + factor(Market) + factor(Month) +
##     AvgPrice + factor(Parking)
##
##                     Df Deviance   AIC
## + factor(RoomType)  6    30773 30825
## + factor(Meal)      3    30800 30846
## + StayLength        1    30834 30876
## + Children          1    30879 30921
## + Adults            1    30886 30928
## <none>                   30890 30930
## + GroupSize         1    30889 30931
```

```
##
## Step:  AIC=30824.65
## factor(Status) ~ LeadTime + HasRequests + factor(Market) + factor(Month) +
##     AvgPrice + factor(Parking) + factor(RoomType)
##
##                Df Deviance   AIC
## + StayLength    1    30696 30750
## + factor(Meal)  3    30712 30770
## + GroupSize     1    30764 30818
## + Adults        1    30766 30820
## <none>               30773 30825
## + Children      1    30772 30826
##
## Step:  AIC=30749.68
## factor(Status) ~ LeadTime + HasRequests + factor(Market) + factor(Month) +
##     AvgPrice + factor(Parking) + factor(RoomType) + StayLength
##
##                Df Deviance   AIC
## + factor(Meal)  3    30619 30679
## + GroupSize     1    30689 30745
## + Adults        1    30690 30746
## <none>               30696 30750
## + Children      1    30695 30751
##
## Step:  AIC=30679.42
## factor(Status) ~ LeadTime + HasRequests + factor(Market) + factor(Month) +
##     AvgPrice + factor(Parking) + factor(RoomType) + StayLength +
##     factor(Meal)
##
##             Df Deviance   AIC
## + GroupSize  1    30613 30675
## + Adults     1    30616 30678
## + Children   1    30617 30679
## <none>            30619 30679
##
## Step:  AIC=30674.51
## factor(Status) ~ LeadTime + HasRequests + factor(Market) + factor(Month) +
##     AvgPrice + factor(Parking) + factor(RoomType) + StayLength +
##     factor(Meal) + GroupSize
##
##            Df Deviance   AIC
## <none>           30613 30675
## + Adults    1    30612 30676
## + Children  1    30612 30676


##
## Call:  glm(formula = factor(Status) ~ LeadTime + HasRequests + factor(Market) +
##     factor(Month) + AvgPrice + factor(Parking) + factor(RoomType) +
##     StayLength + factor(Meal) + GroupSize, family = binomial(link = "logit"),
##     data = hotel_bookings)
##
## Coefficients:
##                 (Intercept)                     LeadTime
##                     5.15790                     -0.01651
```

```
##                 HasRequests  factor(Market)Complementary
##                     2.02060                     17.07502
##     factor(Market)Corporate        factor(Market)Offline
##                     1.14786                      1.91428
##        factor(Market)Online             factor(Month)Feb
##                     0.06250                     -2.83161
##            factor(Month)Mar             factor(Month)Apr
##                    -2.52033                     -2.28402
##            factor(Month)May             factor(Month)Jun
##                    -1.90397                     -2.13202
##            factor(Month)Jul             factor(Month)Aug
##                    -1.98030                     -1.92694
##            factor(Month)Sep             factor(Month)Oct
##                    -1.73622                     -1.97145
##            factor(Month)Nov             factor(Month)Dec
##                    -2.39601                     -0.50040
##                   AvgPrice             factor(Parking)1
##                    -0.01911                      1.58065
## factor(RoomType)Room_Type 2  factor(RoomType)Room_Type 3
##                     0.28519                      0.17572
## factor(RoomType)Room_Type 4  factor(RoomType)Room_Type 5
##                     0.25314                      0.75040
## factor(RoomType)Room_Type 6  factor(RoomType)Room_Type 7
##                     0.85804                      1.45018
##                  StayLength      factor(Meal)Meal Plan 2
##                    -0.07773                     -0.15018
##     factor(Meal)Meal Plan 3     factor(Meal)Not Selected
##                   -10.60678                     -0.36584
##                   GroupSize
##                    -0.07558
##
## Degrees of Freedom: 36284 Total (i.e. Null);  36254 Residual
## Null Deviance:       45900
## Residual Deviance: 30610     AIC: 30670

## Analysis of Deviance Table (Type II tests)
##
## Response: factor(Status)
##                   LR Chisq Df Pr(>Chisq)
## factor(Meal)         76.2  3  < 2.2e-16 ***
## factor(Parking)     238.1  1  < 2.2e-16 ***
## factor(RoomType)    102.7  6  < 2.2e-16 ***
## LeadTime           7710.7  1  < 2.2e-16 ***
## factor(Market)     1954.3  4  < 2.2e-16 ***
## AvgPrice            916.4  1  < 2.2e-16 ***
## HasRequests        4125.0  1  < 2.2e-16 ***
## factor(Month)      1109.8 11  < 2.2e-16 ***
## StayLength           90.7  1  < 2.2e-16 ***
## GroupSize                0
## Adults                   0
## Children                 0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: factor(Status)
##                LR Chisq Df Pr(>Chisq)
## LeadTime         7937.8  1  < 2.2e-16 ***
## factor(Market)   2553.0  4  < 2.2e-16 ***
## HasRequests      4287.6  1  < 2.2e-16 ***
## factor(Month)    1097.4 11  < 2.2e-16 ***
## AvgPrice          945.0  1  < 2.2e-16 ***
## StayLength         65.1  1  7.202e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] 31109.38
```

Cross Validation

```
##
## Call:
## glm(formula = factor(Status) ~ LeadTime + factor(Market) + HasRequests +
##     factor(Month) + AvgPrice + StayLength, family = binomial(link = "logit"),
##     data = train_data)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  4.6354612  0.4029889  11.503  < 2e-16 ***
## LeadTime                    -0.0164280  0.0003068 -53.551  < 2e-16 ***
## factor(Market)Complementary 12.4838067 91.6140098   0.136 0.891612
## factor(Market)Corporate      1.0526272  0.3117396   3.377 0.000734 ***
## factor(Market)Offline        1.7030397  0.2984467   5.706 1.15e-08 ***
## factor(Market)Online        -0.1691951  0.2952272  -0.573 0.566576
## HasRequests                  1.9873485  0.0472468  42.063  < 2e-16 ***
## factor(Month)Feb            -2.6980334  0.2803294  -9.625  < 2e-16 ***
## factor(Month)Mar            -2.4009449  0.2766204  -8.680  < 2e-16 ***
## factor(Month)Apr            -2.2037771  0.2747055  -8.022 1.04e-15 ***
## factor(Month)May            -1.8574544  0.2766521  -6.714 1.89e-11 ***
## factor(Month)Jun            -2.1050696  0.2750478  -7.653 1.96e-14 ***
## factor(Month)Jul            -1.9278522  0.2760302  -6.984 2.86e-12 ***
## factor(Month)Aug            -1.8656157  0.2746965  -6.792 1.11e-11 ***
## factor(Month)Sep            -1.6685899  0.2743934  -6.081 1.19e-09 ***
## factor(Month)Oct            -1.9260882  0.2727406  -7.062 1.64e-12 ***
## factor(Month)Nov            -2.3392606  0.2763377  -8.465  < 2e-16 ***
## factor(Month)Dec            -0.6113595  0.2835020  -2.156 0.031048 *
## AvgPrice                    -0.0141549  0.0006904 -20.504  < 2e-16 ***
## StayLength                  -0.0602561  0.0113038  -5.331 9.79e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 22912  on 18141  degrees of freedom
## Residual deviance: 15570  on 18122  degrees of freedom
## AIC: 15610
```

```
##
## Number of Fisher Scoring iterations: 14
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: factor(Status)
##
## Terms added sequentially (first to last)
##
##
##                Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                           18141      22912
## LeadTime        1   3628.0     18140      19284 < 2.2e-16 ***
## factor(Market)  4    802.2     18136      18482 < 2.2e-16 ***
## HasRequests     1   1899.8     18135      16582 < 2.2e-16 ***
## factor(Month)  11    549.5     18124      16032 < 2.2e-16 ***
## AvgPrice        1    433.6     18123      15599 < 2.2e-16 ***
## StayLength      1     28.4     18122      15570 9.691e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] -0.003261449
```

```
##
## Call:
## glm(formula = factor(Status) ~ LeadTime + factor(Market) + HasRequests +
##     factor(Month) + AvgPrice + StayLength, family = binomial(link = "logit"),
##     data = train_data)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  4.705e+00  4.013e-01  11.725  < 2e-16 ***
## LeadTime                    -1.611e-02  3.047e-04 -52.872  < 2e-16 ***
## factor(Market)Complementary  1.477e+01  1.476e+02   0.100   0.9203
## factor(Market)Corporate      8.707e-01  3.095e-01   2.813   0.0049 **
## factor(Market)Offline        1.537e+00  2.960e-01   5.192 2.09e-07 ***
## factor(Market)Online        -3.017e-01  2.928e-01  -1.030   0.3028
## HasRequests                  2.047e+00  4.762e-02  42.996  < 2e-16 ***
## factor(Month)Feb            -2.518e+00  2.812e-01  -8.955  < 2e-16 ***
## factor(Month)Mar            -2.218e+00  2.771e-01  -8.007 1.18e-15 ***
## factor(Month)Apr            -2.088e+00  2.755e-01  -7.580 3.45e-14 ***
## factor(Month)May            -1.802e+00  2.773e-01  -6.497 8.19e-11 ***
## factor(Month)Jun            -2.076e+00  2.756e-01  -7.535 4.89e-14 ***
## factor(Month)Jul            -1.914e+00  2.761e-01  -6.930 4.20e-12 ***
## factor(Month)Aug            -1.794e+00  2.752e-01  -6.518 7.11e-11 ***
## factor(Month)Sep            -1.693e+00  2.750e-01  -6.156 7.47e-10 ***
## factor(Month)Oct            -1.833e+00  2.734e-01  -6.704 2.03e-11 ***
## factor(Month)Nov            -2.248e+00  2.769e-01  -8.118 4.75e-16 ***
## factor(Month)Dec            -3.543e-01  2.852e-01  -1.242   0.2142
## AvgPrice                    -1.430e-02  6.786e-04 -21.073  < 2e-16 ***
## StayLength                  -7.167e-02  1.126e-02  -6.365 1.96e-10 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 22912  on 18141  degrees of freedom
## Residual deviance: 15483  on 18122  degrees of freedom
## AIC: 15523
##
## Number of Fisher Scoring iterations: 15
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: factor(Status)
##
## Terms added sequentially (first to last)
##
##
##                Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                           18141      22912
## LeadTime        1   3565.1     18140      19347 < 2.2e-16 ***
## factor(Market)  4    768.8     18136      18578 < 2.2e-16 ***
## HasRequests     1   2023.4     18135      16555 < 2.2e-16 ***
## factor(Month)  11    577.6     18124      15977 < 2.2e-16 ***
## AvgPrice        1    453.0     18123      15524 < 2.2e-16 ***
## StayLength      1     40.7     18122      15483  1.82e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] 0.007529658
```