

R Notebook

Code ▼

Jaechul Kim

2023-02-17

This is an R Markdown (<http://rmarkdown.rstudio.com>) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*. # Linear models for classification are a type of machine learning algorithm that are used to predict the class of a new observation based on a set of input features.

required library

Hide

```
install.packages("caret")
```

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

<https://cran.rstudio.com/bin/windows/Rtools/>

'C:/Users/leewq/AppData/Local/R/win-library/4.2'의 위치에 패키지(들)을 설치합니다.

(왜냐하면 'lib'가 지정되지 않았기 때문입니다)

'listenv', 'parallelly', 'future', 'globals', 'future.apply', 'numDeriv', 'progressr', 'SQUAREM', 'lava', 'tzdb', 'cpp11', 'proclim', 'timechange', 'iterators', 'Rcpp', 'data.table', 'clock', 'gower', 'hardhat', 'ipred', 'lubridate', 'purrr', 'tidyr', 'timeDate', 'foreach', 'ModelMetrics', 'plyr', 'pROC', 'recipes', 'reshape2'(들)을 또한 설치합니다.

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/listenv_0.9.0.zip'

Content type 'application/zip' length 108518 bytes (105 KB)

downloaded 105 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/parallelly_1.34.0.zip'

Content type 'application/zip' length 328047 bytes (320 KB)

downloaded 320 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/future_1.31.0.zip'

Content type 'application/zip' length 646771 bytes (631 KB)

downloaded 631 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/globals_0.16.2.zip'

Content type 'application/zip' length 107100 bytes (104 KB)

downloaded 104 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/future.apply_1.10.0.zip'

Content type 'application/zip' length 156500 bytes (152 KB)

downloaded 152 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/numDeriv_2016.8-1.1.zip'

Content type 'application/zip' length 116102 bytes (113 KB)

downloaded 113 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/progressr_0.13.0.zip'

Content type 'application/zip' length 385186 bytes (376 KB)

downloaded 376 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/SQUAREM_2021.1.zip'

Content type 'application/zip' length 183063 bytes (178 KB)

downloaded 178 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/lava_1.7.1.zip'

Content type 'application/zip' length 2657415 bytes (2.5 MB)

downloaded 2.5 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/tzdb_0.3.0.zip'

Content type 'application/zip' length 1011488 bytes (987 KB)

downloaded 987 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/cpp11_0.4.3.zip'

Content type 'application/zip' length 321855 bytes (314 KB)

downloaded 314 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/prodlim_2019.11.13.zip'
Content type 'application/zip' length 401937 bytes (392 KB)
downloaded 392 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/timechange_0.2.0.zip'
Content type 'application/zip' length 508615 bytes (496 KB)
downloaded 496 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/iterators_1.0.14.zip'
Content type 'application/zip' length 353335 bytes (345 KB)
downloaded 345 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/Rcpp_1.0.10.zip'
Content type 'application/zip' length 2819232 bytes (2.7 MB)
downloaded 2.7 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/data.table_1.14.6.zip'
Content type 'application/zip' length 2238251 bytes (2.1 MB)
downloaded 2.1 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/clock_0.6.1.zip'
Content type 'application/zip' length 2349325 bytes (2.2 MB)
downloaded 2.2 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/gower_1.0.1.zip'
Content type 'application/zip' length 304819 bytes (297 KB)
downloaded 297 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/hardhat_1.2.0.zip'
Content type 'application/zip' length 805794 bytes (786 KB)
downloaded 786 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/ipred_0.9-13.zip'
Content type 'application/zip' length 389297 bytes (380 KB)
downloaded 380 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/lubridate_1.9.2.zip'
Content type 'application/zip' length 944778 bytes (922 KB)
downloaded 922 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/purrr_1.0.1.zip'
Content type 'application/zip' length 498554 bytes (486 KB)
downloaded 486 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/tidyr_1.3.0.zip'
Content type 'application/zip' length 1292954 bytes (1.2 MB)
downloaded 1.2 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/timeDate_4022.108.zip'
Content type 'application/zip' length 1375172 bytes (1.3 MB)
downloaded 1.3 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/foreach_1.5.2.zip'
Content type 'application/zip' length 148196 bytes (144 KB)
downloaded 144 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/ModelMetrics_1.2.2.2.zip'
Content type 'application/zip' length 476737 bytes (465 KB)
downloaded 465 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/plyr_1.8.8.zip'
Content type 'application/zip' length 1151414 bytes (1.1 MB)
downloaded 1.1 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/pROC_1.18.0.zip'
Content type 'application/zip' length 1157154 bytes (1.1 MB)
downloaded 1.1 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/recipes_1.0.4.zip'
Content type 'application/zip' length 1521424 bytes (1.5 MB)
downloaded 1.5 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/reshape2_1.4.4.zip'
Content type 'application/zip' length 446521 bytes (436 KB)
downloaded 436 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/caret_6.0-93.zip'
Content type 'application/zip' length 3576923 bytes (3.4 MB)
downloaded 3.4 MB

패키지 'listenv'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'parallelly'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'future'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'globals'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'future.apply'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'numDeriv'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'progressr'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'SQUAREM'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'lava'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'tazdb'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'cpp11'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'prodlim'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'timechange'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'iterators'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'Rcpp'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'data.table'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'clock'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'gower'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'hardhat'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'ipred'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'lubridate'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'purrr'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'tidyr'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'timeDate'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'foreach'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'ModelMetrics'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'plyr'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'pROC'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'recipes'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'reshape2'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다
패키지 'caret'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다

다운로드된 바이너리 패키지들은 다음의 위치에 있습니다

C:\Users\leewq\AppData\Local\Temp\RtmpW2J6tF\downloaded_packages

read file and make train and test data

Hide

```
# Set the working directory to the folder where the CSV file is located
setwd("C:\\Users\\leewq\\Downloads\\archive")

# Read the CSV file into a data frame
df <- read.csv("hotel_bookings.csv")
install.packages("e1071")
```

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

<https://cran.rstudio.com/bin/windows/Rtools/>

Warning in install.packages :

패키지 'e1071'가 사용중이므로 설치되지 않을 것입니다

Hide

```
# Remove rows with missing values
install.packages("psych")
```

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

<https://cran.rstudio.com/bin/windows/Rtools/>

Warning in install.packages :
패키지 ‘psych’가 사용중이므로 설치되지 않을 것입니다

Hide

```
install.packages("dplyr")
```

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

<https://cran.rstudio.com/bin/windows/Rtools/>

Warning in install.packages :
패키지 ‘dplyr’가 사용중이므로 설치되지 않을 것입니다

Hide

```
install.packages("ggplot2")
```

Error in install.packages : Updating loaded packages

Hide

```
df <- na.omit(df)

# Set seed for reproducibility
set.seed(123)

# Determine row indices for training and testing sets
train_indices <- sample(1:nrow(df), 0.8*nrow(df), replace = FALSE)
test_indices <- setdiff(1:nrow(df), train_indices)

# Create training and testing sets
train <- df[train_indices, ]
test <- df[test_indices, ]
```

1. Use at least 5 R functions for data exploration, using the training data

Hide

```
summary(train)
```

hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month
Length:95508	Min. :0.0000	Min. : 0.0	Min. :2015	Length:95508
Class :character	1st Qu.:0.0000	1st Qu.: 18.0	1st Qu.:2016	Class :character
Mode :character	Median :0.0000	Median : 69.0	Median :2016	Mode :character
	Mean :0.3712	Mean :104.1	Mean :2016	
	3rd Qu.:1.0000	3rd Qu.:161.0	3rd Qu.:2017	
	Max. :1.0000	Max. :709.0	Max. :2017	
arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights	
Min. : 1.00	Min. : 1.00	Min. : 0.0000	Min. : 0.000	
1st Qu.:16.00	1st Qu.: 8.00	1st Qu.: 0.0000	1st Qu.: 1.000	
Median :28.00	Median :16.00	Median : 1.0000	Median : 2.000	
Mean :27.18	Mean :15.76	Mean : 0.9311	Mean : 2.507	
3rd Qu.:38.00	3rd Qu.:23.00	3rd Qu.: 2.0000	3rd Qu.: 3.000	
Max. :53.00	Max. :31.00	Max. :19.0000	Max. :50.000	
adults	children	babies	meal	country
Min. : 0.000	Min. : 0.0000	Min. : 0.000000	Length:95508	Length:95508
1st Qu.: 2.000	1st Qu.: 0.0000	1st Qu.: 0.000000	Class :character	Class :character
Median : 2.000	Median : 0.0000	Median : 0.000000	Mode :character	Mode :character
Mean : 1.857	Mean : 0.1046	Mean : 0.007947		
3rd Qu.: 2.000	3rd Qu.: 0.0000	3rd Qu.: 0.000000		
Max. :55.000	Max. :10.0000	Max. :10.000000		
market_segment	distribution_channel	is_repeated_guest	previous_cancellations	
Length:95508	Length:95508	Min. :0.00000	Min. : 0.00000	
Class :character	Class :character	1st Qu.:0.00000	1st Qu.: 0.00000	
Mode :character	Mode :character	Median :0.00000	Median : 0.00000	
		Mean :0.03232	Mean : 0.08778	
		3rd Qu.:0.00000	3rd Qu.: 0.00000	
		Max. :1.00000	Max. :26.00000	
previous_bookings_not_canceled	reserved_room_type	assigned_room_type	booking_changes	deposit_type
Min. : 0.0000	Length:95508	Length:95508	Min. : 0.0000	Length:95508
1st Qu.: 0.0000	Class :character	Class :character	1st Qu.: 0.0000	Class :character
Median : 0.0000	Mode :character	Mode :character	Median : 0.0000	Mode :character
Mean : 0.1394			Mean : 0.2204	
3rd Qu.: 0.0000			3rd Qu.: 0.0000	
Max. :72.0000			Max. :21.0000	
agent	company	days_in_waiting_list	customer_type	adr
Length:95508	Length:95508	Min. : 0.000	Length:95508	Min. : -6.38
Class :character	Class :character	1st Qu.: 0.000	Class :character	1st Qu.: 69.50
Mode :character	Mode :character	Median : 0.000	Mode :character	Median : 94.67
		Mean : 2.309		Mean : 101.85
		3rd Qu.: 0.000		3rd Qu.: 126.00
		Max. :391.000		Max. :5400.00
required_car_parking_spaces	total_of_special_requests	reservation_status	reservation_status_date	

Min. :0.00000	Min. :0.0000	Length:95508	Length:95508
1st Qu.:0.00000	1st Qu.:0.0000	Class :character	Class :character
Median :0.00000	Median :0.0000	Mode :character	Mode :character
Mean :0.06269	Mean :0.5718		
3rd Qu.:0.00000	3rd Qu.:1.0000		
Max. :8.00000	Max. :5.0000		

Hide

```
install.packages("ggplot2")
```

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

<https://cran.rstudio.com/bin/windows/Rtools/>

Warning in install.packages :

패키지 ‘ggplot2’가 사용중이므로 설치되지 않을 것입니다

Hide

```
str(train)
```



```
'data.frame': 95508 obs. of 32 variables:
 $ hotel : chr "City Hotel" "City Hotel" "Resort Hotel" "Resort Hotel" ...
 $ is_canceled : int 1 1 0 0 0 1 1 0 1 ...
 $ lead_time : int 158 39 79 49 163 0 164 114 72 160 ...
 $ arrival_date_year : int 2016 2016 2015 2016 2016 2016 2017 2017 2015 2017 ...
 $ arrival_date_month : chr "May" "October" "November" "November" ...
 $ arrival_date_week_number : int 22 41 46 46 34 50 20 2 44 13 ...
 $ arrival_date_day_of_month : int 24 2 9 6 14 10 15 14 28 26 ...
 $ stays_in_weekend_nights : int 0 2 1 2 2 0 1 2 0 2 ...
 $ stays_in_week_nights : int 2 1 5 1 1 1 2 2 3 1 ...
 $ adults : int 1 2 2 2 2 2 1 1 2 2 ...
 $ children : int 0 0 0 0 0 0 0 0 1 ...
 $ babies : int 0 0 0 0 0 0 0 0 0 ...
 $ meal : chr "BB" "BB" "BB" "HB" ...
 $ country : chr "PRT" "ESP" "PRT" "ESP" ...
 $ market_segment : chr "Groups" "Online TA" "Offline TA/TO" "Online TA" ...
 $ distribution_channel : chr "TA/TO" "TA/TO" "TA/TO" "TA/TO" ...
 $ is_repeated_guest : int 0 0 0 0 0 0 0 0 0 ...
 $ previous_cancellations : int 0 0 0 0 0 0 0 0 0 ...
 $ previous_bookings_not_canceled : int 0 0 0 0 0 0 0 0 0 ...
 $ reserved_room_type : chr "A" "A" "E" "A" ...
 $ assigned_room_type : chr "A" "A" "E" "A" ...
 $ booking_changes : int 0 0 0 6 1 0 0 0 0 ...
 $ deposit_type : chr "Non Refund" "No Deposit" "No Deposit" "No Deposit" ...
 ...
 $ agent : chr "37" "9" "314" "240" ...
 $ company : chr "NULL" "NULL" "NULL" "NULL" ...
 $ days_in_waiting_list : int 31 0 0 0 0 0 0 0 58 0 ...
 $ customer_type : chr "Transient" "Transient" "Contract" "Transient" ...
 $ adr : num 130 126 56.2 64.8 114.5 ...
 $ required_car_parking_spaces : int 0 0 0 0 0 0 0 0 0 ...
 $ total_of_special_requests : int 0 1 1 3 1 0 0 0 1 0 ...
 $ reservation_status : chr "Canceled" "Canceled" "Check-Out" "Check-Out" ...
 $ reservation_status_date : chr "2016-01-18" "2016-08-29" "2015-11-15" "2016-11-09" ...
 - attr(*, "na.action")= 'omit' Named int [1:4] 40601 40668 40680 41161
 ..- attr(*, "names")= chr [1:4] "40601" "40668" "40680" "41161"
```

Hide

```
table(train$gender)
```

```
< table of extent 0 >
```

Hide

```
hist(train$lead_time)
```

[Hide](#)

```
numerical_vars <- train %>% select(lead_time, total_of_special_requests, stays_in_week_nights)

cor(numerical_vars)
```

	lead_time	total_of_special_requests	stays_in_week_nights
lead_time	1.00000000	-0.09466659	0.16761168
total_of_special_requests	-0.09466659	1.00000000	0.06847342
stays_in_week_nights	0.16761168	0.06847342	1.00000000

#Create at least 2 informative graphs, using the training data

[Hide](#)

```
ggplot(train, aes(x=hotel, y=adr, fill=hotel)) +
  geom_boxplot() +
  labs(x="Hotel Type", y="Daily Rate", title="Box Plot of Daily Rate by Hotel Type") +
  theme_minimal()
```

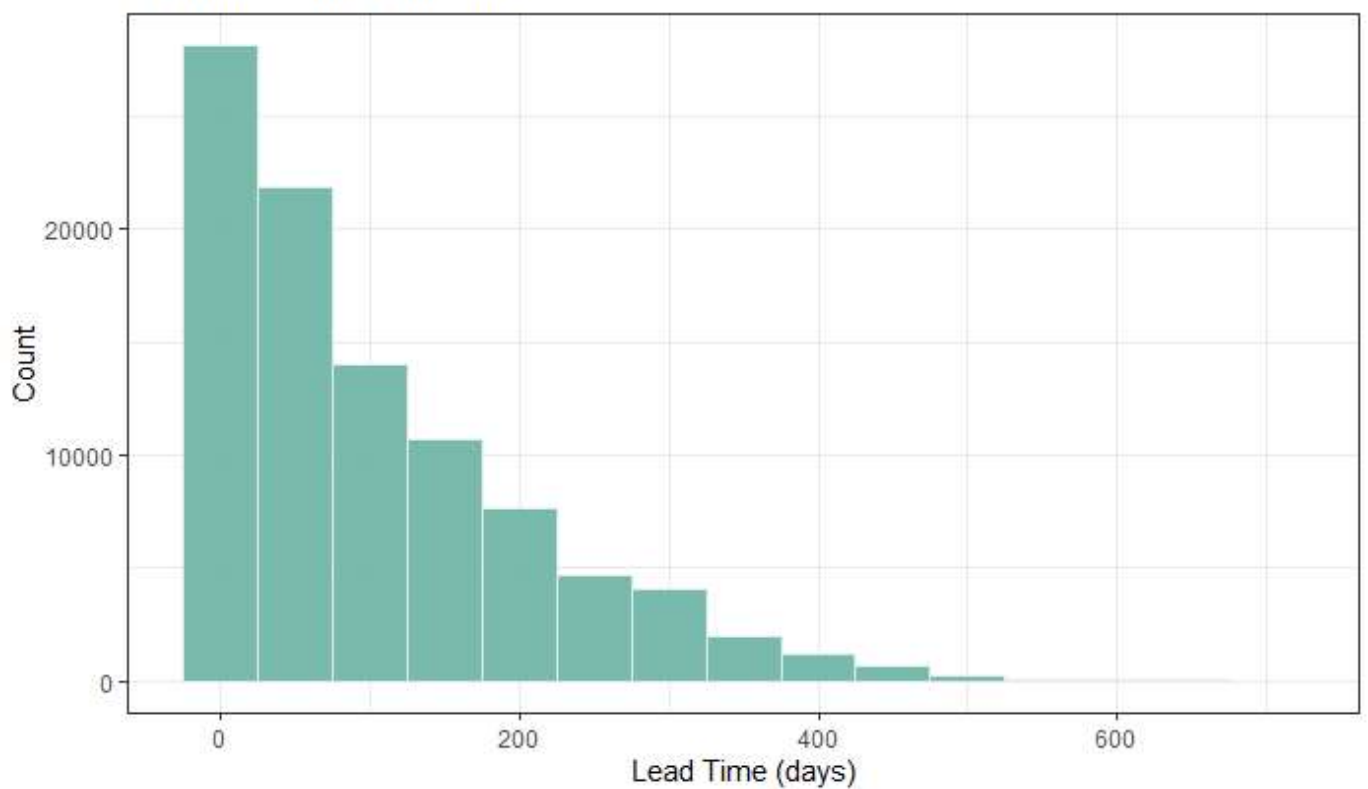
Box Plot of Daily Rate by Hotel Type



Hide

```
ggplot(train, aes(x = lead_time)) +  
  geom_histogram(binwidth = 50, fill = "#69b3a2", color = "#e9ecef", alpha = 0.9) +  
  xlab("Lead Time (days)") +  
  ylab("Count") +  
  ggtitle("Distribution of Lead Time") +  
  theme_bw()
```

Distribution of Lead Time



#Build a logistic regression model and output the summary. Write a thorough explanation of the information in the model summary

Hide

```
# fit logistic regression model
logistic_model <- glm(is_canceled ~ lead_time + stays_in_weekend_nights + stays_in_week_nights, data = train, family = "binomial")

# output model summary
summary(logistic_model)
```

```
Call:
glm(formula = is_canceled ~ lead_time + stays_in_weekend_nights + stays_in_week_nights, family = "binomial", data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2646  -0.8762  -0.7563   1.2288   1.9729

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.095e+00  1.329e-02 -82.377  < 2e-16 ***
lead_time       5.989e-03  7.028e-05  85.217  < 2e-16 ***
stays_in_weekend_nights -4.258e-02  8.083e-03  -5.268  1.38e-07 ***
stays_in_week_nights  -1.632e-02  4.303e-03  -3.793  0.000149 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 125992  on 95507  degrees of freedom
Residual deviance: 117643  on 95504  degrees of freedom
AIC: 117651

Number of Fisher Scoring iterations: 4
```

Using these two classification models, predict and evaluate on the test data using all of the classification metrics discussed in class. Compare the results and indicate why you think these results happened.

Hide

```
# predict using logistic regression model
logistic_pred <- predict(logistic_model, newdata = test, type = "response")
logistic_pred_class <- ifelse(logistic_pred > 0.5, 1, 0)

library(e1071)

# Fit the Naive Bayes model
naivebayes_model <- naiveBayes(is_canceled ~ lead_time + stays_in_weekend_nights + stays_in_week_nights, data = train)

# Print the model summary
summary(naivebayes_model)
```

	Length	Class	Mode
apriori	2	table	numeric
tables	3	-none-	list
levels	2	-none-	character
isnumeric	3	-none-	logical
call	4	-none-	call

Hide

```
# predict using naive Bayes model
naivebayes_pred <- predict(naivebayes_model, newdata = test)
naivebayes_pred_class <- ifelse(naivebayes_pred == "canceled", 1, 0)
# Logistic Regression
logistic_pred <- predict(logistic_model, newdata = test, type = "response")
logistic_pred <- ifelse(logistic_pred > 0.5, 1, 0)
logistic_accuracy <- mean(logistic_pred == test$is_canceled)
logistic_precision <- sum(logistic_pred & test$is_canceled) / sum(logistic_pred)
logistic_recall <- sum(logistic_pred & test$is_canceled) / sum(test$is_canceled)
logistic_f1 <- 2 * (logistic_precision * logistic_recall) / (logistic_precision + logistic_recall)

# Naive Bayes
naivebayes_pred <- predict(naivebayes_model, newdata = test)
naivebayes_accuracy <- mean(naivebayes_pred == test$is_canceled)
naivebayes_table <- table(naivebayes_pred, test$is_canceled)
if ("0" %in% rownames(naivebayes_table)) {
  naivebayes_precision <- naivebayes_table[2,2] / sum(naivebayes_table[2,])
  naivebayes_recall <- naivebayes_table[2,2] / sum(naivebayes_table[,2])
  naivebayes_f1 <- 2 * (naivebayes_precision * naivebayes_recall) / (naivebayes_precision + naivebayes_recall)
} else {
  naivebayes_precision <- 0
  naivebayes_recall <- 0
  naivebayes_f1 <- 0
}

# Output results
cat(sprintf("Logistic Regression:\nAccuracy: %.3f\nPrecision: %.3f\nRecall: %.3f\nF1 Score: %.3f\n\n",
           logistic_accuracy, logistic_precision, logistic_recall, logistic_f1))
```

```
Logistic Regression:
Accuracy: 0.665
Precision: 0.593
Recall: 0.278
F1 Score: 0.379
```

Hide

```
cat(sprintf("Naive Bayes:\nAccuracy: %.3f\nPrecision: %.3f\nRecall: %.3f\nF1 Score: %.3f\n",
           naivebayes_accuracy, naivebayes_precision, naivebayes_recall, naivebayes_f1))
```

```
Naive Bayes:
Accuracy: 0.663
Precision: 0.590
Recall: 0.272
F1 Score: 0.373
```

#Write a paragraph listing the strengths and weaknesses of Naïve Bayes and Logistic Regression.

#Naive Bayes is stronger when there is less data and Logistic Regression is stronger when there is bigger data. Its because Logistic Regression is linear but Naive Bayes is log.

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.