

▼ Final Project

Rules of Engagement:

This is an **honor system assignment**: You may consult your professor, your lab instructor, the textbook, and material on the Internet at any time. You may not consult, collaborate, or seek assistance from any other human besides your professor and lab instructor. Your attribution statement, at the top of your R-code file, should reflect these constraints.

```
dataFile <- "https://intro-datascience.s3.us-east-2.amazonaws.com/Resort01.csv"

#Rebecca Candee
#1

library(tidyverse)
bookings <- read_csv("https://intro-datascience.s3.us-east-2.amazonaws.com/Resort01.csv")

Warning message in system("timedatectl", intern = TRUE):
"running command 'timedatectl' had status 1"
— Attaching packages — tidyverse 1.3.1 —

✔ ggplot2 3.3.5      ✔ purrr 0.3.4
✔ tibble 3.1.6       ✔ dplyr 1.0.9
✔ tidyr 1.2.0        ✔ stringr 1.4.0
✔ readr 2.1.2        ✔ forcats 0.5.1

— Conflicts — tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()     masks stats::lag()

Rows: 40060 Columns: 20
— Column specification —
Delimiter: ","
chr (7): Meal, Country, MarketSegment, ReservedRoomType, AssignedRoomType, ...
dbl (13): IsCanceled, LeadTime, StaysInWeekendNights, StaysInWeekNights, Adu...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message

#The data frame has data that displays information about hotel bookings and cancelatio
#The data frame has 40060 observations
#The data frame has 20 variables.
#Adults, cancellations, and repeat stays are all variables.

#2
```

```
sum(is.na(bookings))
#There are no missing values
```

```
0
```

```
#3
```

```
table(bookings$ReservedRoomType)
#The table shows the room type reserved.
#The most popular room is room type A which has 23399.
```

```
table(bookings$IsRepeatedGuest)
#38383 reservations are not repeating guests.
#1778 reservations are repeating guests.
```

```
table(bookings$IsCanceled)
#28938 bookings were not canceled.
#11122 bookings were canceled.
```

A	B	C	D	E	F	G	H	L	P
23399	3	918	7433	4982	1106	1610	601	6	2

0	1
38282	1778

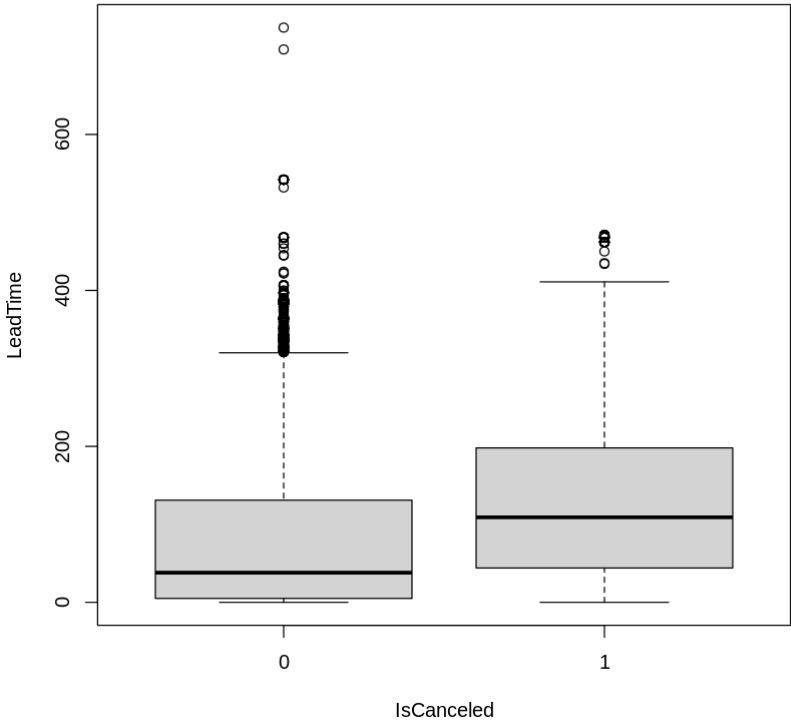
0	1
28938	11122

```
#4
```

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

```
boxplot(LeadTime ~ IsCanceled, data=bookings)
#The higher number of days in LeadTime results in less cancellations.
boxplot(StaysInWeekNights ~ IsCanceled, data=bookings)
#The more times you stay monday to friday results in more cancellations.
boxplot(PreviousCancellations ~ IsCanceled, data=bookings)
#This shows that someone who has previously canceled is more likely to cancel in the 1
```

Installing package into `'/usr/local/lib/R/site-library'`
(as `'lib'` is unspecified)





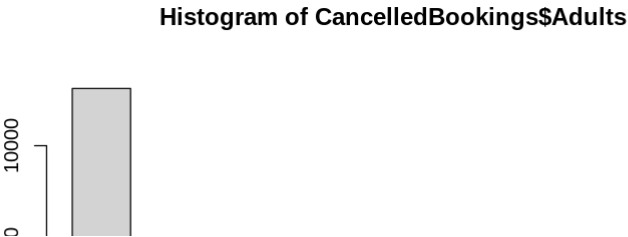
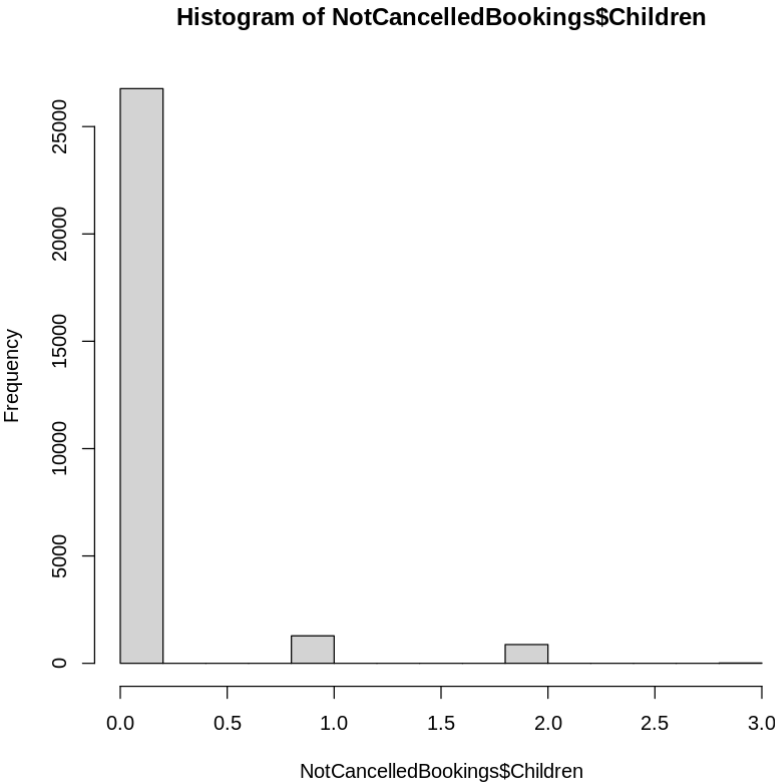
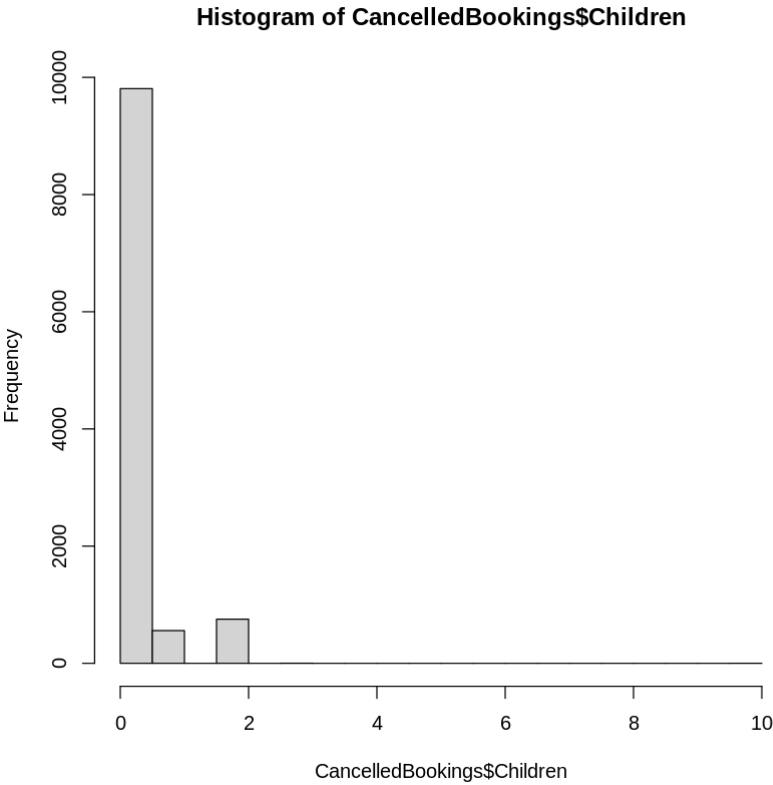
#5

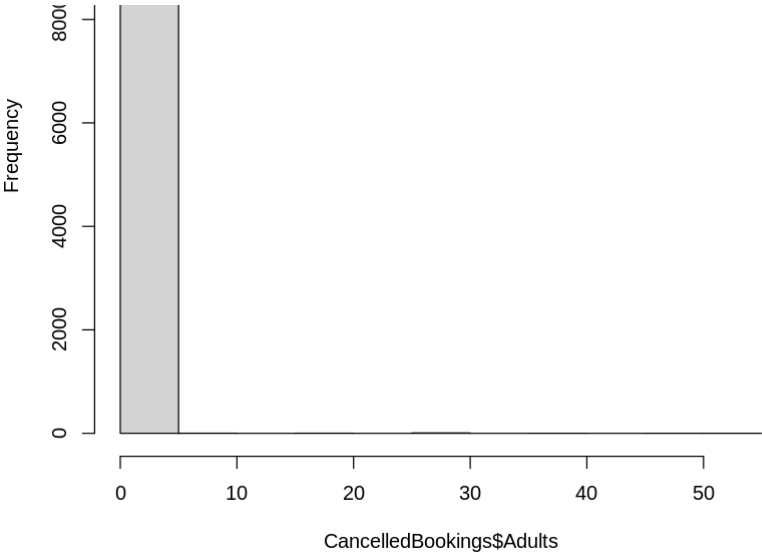
```
bookings %>% filter(bookings$IsCanceled==1) ->CancelledBookings
bookings %>% filter(bookings$IsCanceled==0) ->NotCancelledBookings
```

```
hist(CancelledBookings$Children)
hist(NotCancelledBookings$Children)
#The bookings with children are cancelled more often.
```

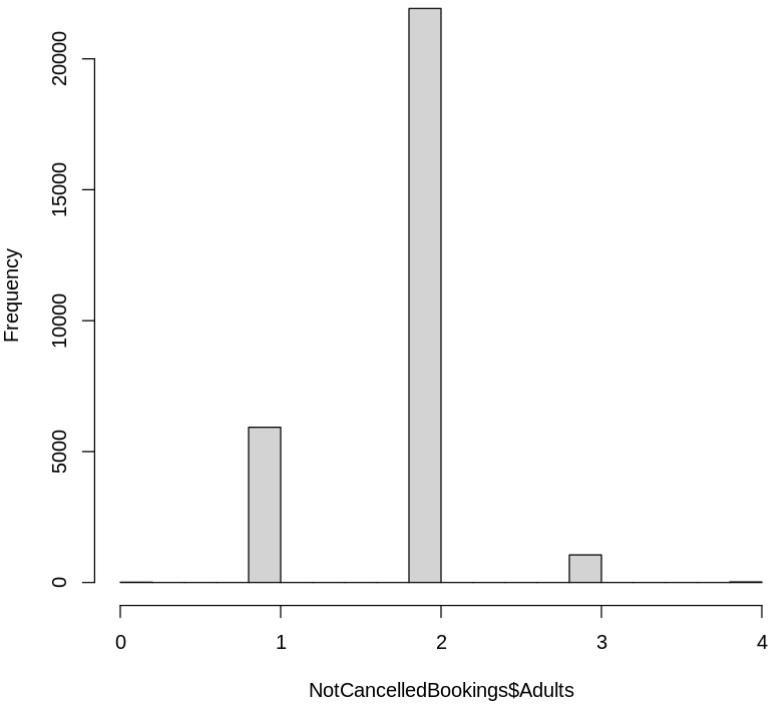
```
hist(CancelledBookings$Adults)
hist(NotCancelledBookings$Adults)
#In the cancelled histogram adults in large parties cancel more often.
#In the other histogram it showed parties of two are least likely to cancel.
```

```
hist(CancelledBookings$Babies)
hist(NotCancelledBookings$Babies)
#There is not a big difference between cancelled and not cancelled bookings for babies
```

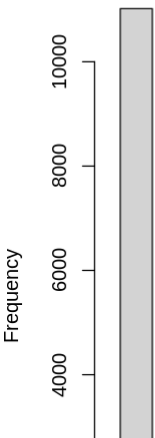


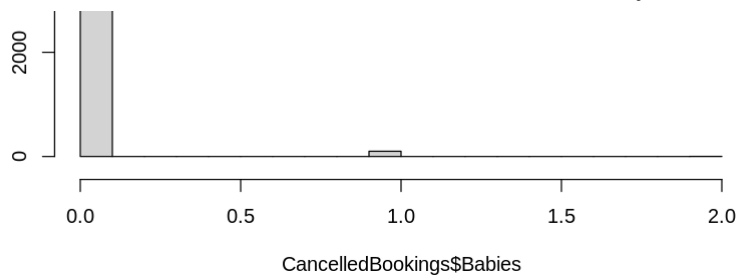


Histogram of NotCancelledBookings\$Adults



Histogram of CancelledBookings\$Babies





#6



```
which.max(table(bookings$Country))
```

```
#Portugal is the country with the largest number number of bookings.
```

PRT: 97

#7



```
country <- aggregate(IsCanceled ~ Country, data = bookings, sum)
names(country)[names(country) == "Country"] <- 'name'
names(country)[names(country) == "IsCanceled"] <- 'NotCancelledBookings'
which.max(country$numCanceled)
country
#Portugal has the highest number of cancellations.
```

A data.frame: 126 x 2

name	NotCancelledBookings
<chr>	<dbl>
AGO	7
ALB	1
AND	3
ARE	8
ARG	9
ARM	0
AUS	15
AUT	34
AZE	1
BDI	0
BEL	59
BGR	0
BHR	1
BHS	0
BIH	0
BLR	3
BRA	101
BWA	0
CAF	0
CHE	112
CHL	3
CHN	9
CIV	0
CMR	0
CN	96
COL	4
COM	0
CPV	0
CRI	0

CUB	0
:	:
PRT	7438
QAT	0
ROU	32
RUS	61
SAU	0
SEN	1
SGP	0
SMR	0
SRB	0
SUR	0
SVK	3
SVN	1
SWE	73
SYC	1
SYR	0
TGO	0

#8

TUN	1
-----	---

```
install.packages("rworldmap")
library(rworldmap)
sPDF <-
  joinCountryData2Map(country, joinCode="ISO3", nameJoinColumn="name")
map<-mapCountryData(sPDF, nameColumnToPlot='NotCancelledBookings',
  catMethod="logFixedWidth")
```

```
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
```

```
also installing the dependencies 'dotCall64', 'gridExtra', 'spam', 'viridis', 'm
```

```
Loading required package: sp
```

```
### Welcome to rworldmap ###
```

```
For a short introduction type :          vignette('rworldmap')
```

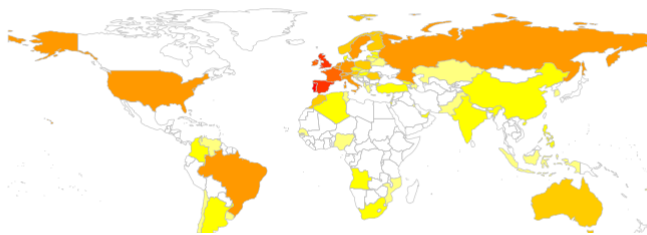
```
123 codes from your data successfully matched countries in the map
```

```
3 codes from your data failed to match with a country code in the map
```

```
120 codes from the map weren't represented in your data
```

```
zero values are replaced with NA as they can't be logged in catMethod=logFixedWid
```

NotCancelledBookings



```
#The Portugal, United Kingdom and Spain locations have the least cancellations in rela
```

```
#9
```

```
bookCat <- data.frame(meal=as.factor(bookings$Meal),
                      marketSegment=as.factor(bookings$MarketSegment),
                      isRepeatedGuest=as.factor(bookings$IsRepeatedGuest),
                      assignedRoom=as.factor(bookings$AssignedRoomType),
                      customerType=as.factor(bookings$CustomerType),
                      bookingChanges=as.factor(bookings$BookingChanges>0),
                      canceled=as.factor(bookings$IsCanceled))
```

```
install.packages(c("arules", "arulesViz"))
library("arules", "arulesViz")
bookCat1 <- as(bookCat, "transactions")
itemFrequencyPlot(bookCat1)
```



```
rules <- apriori(bookCat,parameter=list(supp=0.01, conf=0.5), control = list(verbose=F
```

```
#StaysInWeekNights is a rule that supports the fact that people are more
#likely people to cancel when staying from monday to friday.
```

```
#PreviousCancelations is a rule that states if the customer cancelled in the past
#that they will most likely cancel in the future as well.
```

```
#11
```

```
book <- data.frame(leadTime=bookings$LeadTime,
                   staysWeekend=bookings$StaysInWeekendNights,
                   staysWeek=bookings$StaysInWeekNights,
                   adults=bookings$Adults,
                   children=bookings$Children,
                   babies=bookings$Babies,
                   prevCancellations=bookings$PreviousCancellations,
                   specialRequests=bookings$TotalOfSpecialRequests,
                   canceled=as.factor(bookings$IsCanceled))
```

```
install.packages("caret")
library(caret)
trainList <-
  createDataPartition(y=book$canceled, p=.70,list = FALSE)
```

```
trainSet <- book[trainList,]
testSet <- book[-trainList,]
```

```
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
```

```
also installing the dependencies 'listenv', 'parallelly', 'future', 'globals', 'r'
```

```
Loading required package: lattice
```

```
Attaching package: 'caret'
```

```
The following object is masked from 'package:purrr':
```

```
lift
```

#12

```
install.packages("e1071")
install.packages("rpart")
install.packages("rpart.plot")
install.packages("kernlab")
library(e1071)
library(rpart)
library(rpart.plot)
library(arules)
library(kernlab)
library(caret)

svm <- ksvm(canceled ~ ., data=trainSet, method="svmRadial", prepProc=c("center","scale"))
svm
predict(svm)
svmPred <- predict(svm, testSet, type="response")
confusionMatrix(testSet$canceled,svmPred)
table(testSet$canceled,svmPred)

library(rpart)
model <- rpart(canceled ~., data = testSet, method = 'class', parms=list(split = 'info'))

model
install.packages('rpart.plot')
library(rpart.plot)

rpart.plot(model)
```



Confusion Matrix and Statistics

	Reference	
Prediction	0	1
0	19872	385
1	6341	1445

Accuracy : 0.7602
 95% CI : (0.7551, 0.7651)
 No Information Rate : 0.9347
 P-Value [Acc > NIR] : 1

Kappa : 0.2179

Mcnemar's Test P-Value : <2e-16

Sensitivity : 0.7581
 Specificity : 0.7896
 Pos Pred Value : 0.9810
 Neg Pred Value : 0.1856
 Prevalence : 0.9347
 Detection Rate : 0.7086
 Detection Prevalence : 0.7224
 Balanced Accuracy : 0.7739

'Positive' Class : 0

	svmPred	
	0	1
0	19872	385
1	6341	1445

n= 28043

node), split, n, loss, yval, (yprob)
 * denotes terminal node

```

1) root 28043 7786 0 (0.72235495 0.27764505)
  2) leadTime< 14.5 7943 682 0 (0.91413823 0.08586177) *
  3) leadTime>=14.5 20100 7104 0 (0.64656716 0.35343284)
    6) prevCancellations< 0.5 19454 6493 0 (0.66623831 0.33376169)
      12) specialRequests>=0.5 9060 2414 0 (0.73355408 0.26644592) *
      13) specialRequests< 0.5 10394 4079 0 (0.60756206 0.39243794)
        26) children< 0.5 9454 3511 0 (0.62862281 0.37137719) *
        27) children>=0.5 940 372 1 (0.39574468 0.60425532) *
        7) prevCancellations>=0.5 646 35 1 (0.05417957 0.94582043) *
Installing package into '/usr/local/lib/R/site-library'
```

#13

#Dear CEO, I reviewed the data regarding your hotel, and after analyzing it I have some
 #One of the first things I noticed was that the reservations with people of 5 or more
 #than groups of less than 5. After analyzing the data I have a suggestion for this. I
 #implemented a security deposit for reservations of large parties. This will ensure the
 #higher regard because more money is on the line. This will make sure people do not cost
 #the company money as well as time and effort. Something else that I noticed from you

#bookings in the country of Portugal. To improve cancellation rates and bookings, I su
#those in the Portugal location. It is clear that this is the most sought after and de
#to strive to make all locations resemble the one in Portugal. I believe that this wil
#an international level. The Portugal location will act as a great model because it al
#hotel.

#Thank you for taking the time to read my message and I hope that my suggestions can b
#I thoroughly enjoyed interpreting the data I had available and I hope you have a grea

#Best,

#Rebecca

✓ 0s completed at 3:34 PM

