HW2

Lauren Stover

3/7/2021

#Question 1, Plot 1

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

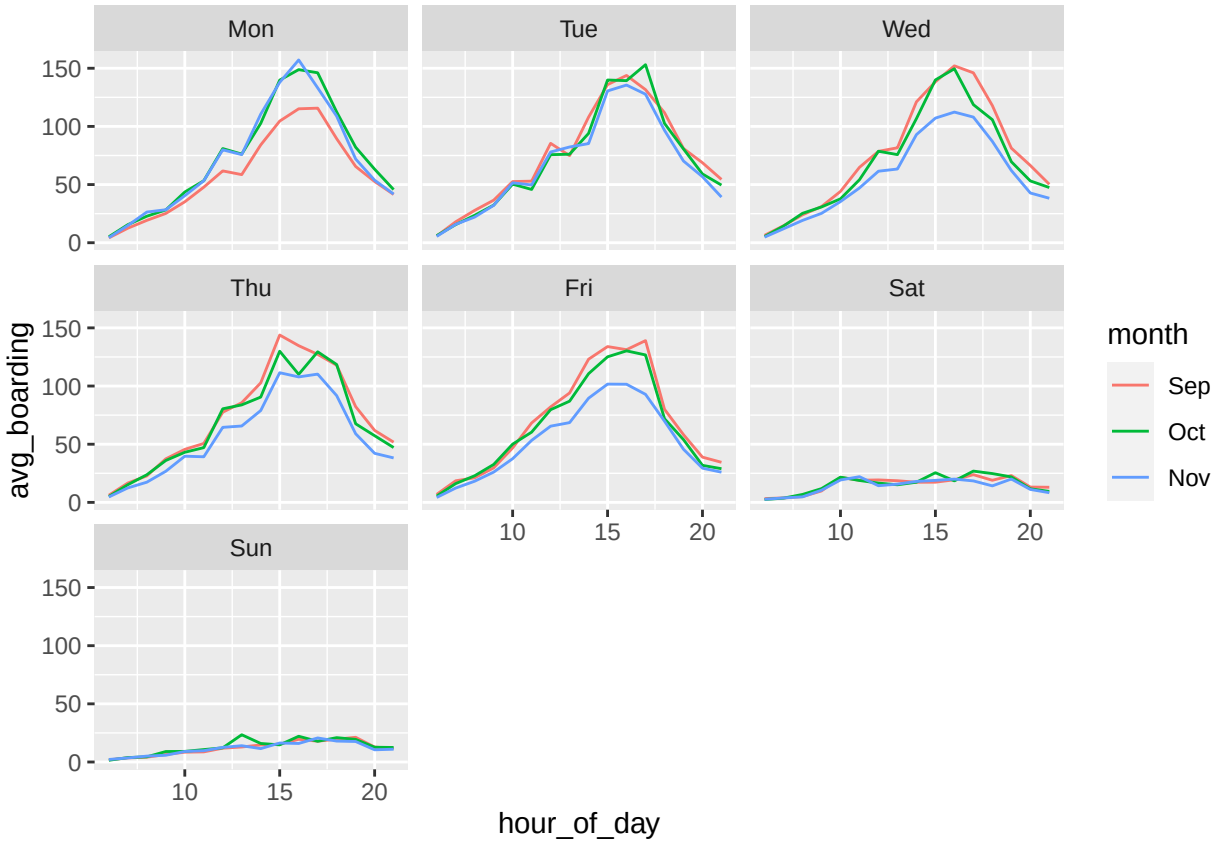
```
library(dbplyr)
```

```
##
## Attaching package: 'dbplyr'

## The following objects are masked from 'package:dplyr':
##
##     ident, sql
```

```
library(readxl)
library(ggplot2)
capmetro_UT <- read_excel("~/Documents/UTX/DataMining/ECO395M-master/data/capmetro_UT.xls")
# Recode the categorical variables in sensible, rather than alphabetical, order
capmetro_UT = mutate(capmetro_UT,
               day_of_week = factor(day_of_week,
                 levels=c("Mon", "Tue", "Wed","Thu", "Fri", "Sat", "Sun")),
               month = factor(month,
                 levels=c("Sep", "Oct","Nov")))
metroavg=capmetro_UT %>%
  group_by(day_of_week,hour_of_day,month) %>%
  summarise(avg_boarding=mean(boarding))
```

```
## `summarise()` has grouped output by 'day_of_week', 'hour_of_day'. You can override using the `.groups
```

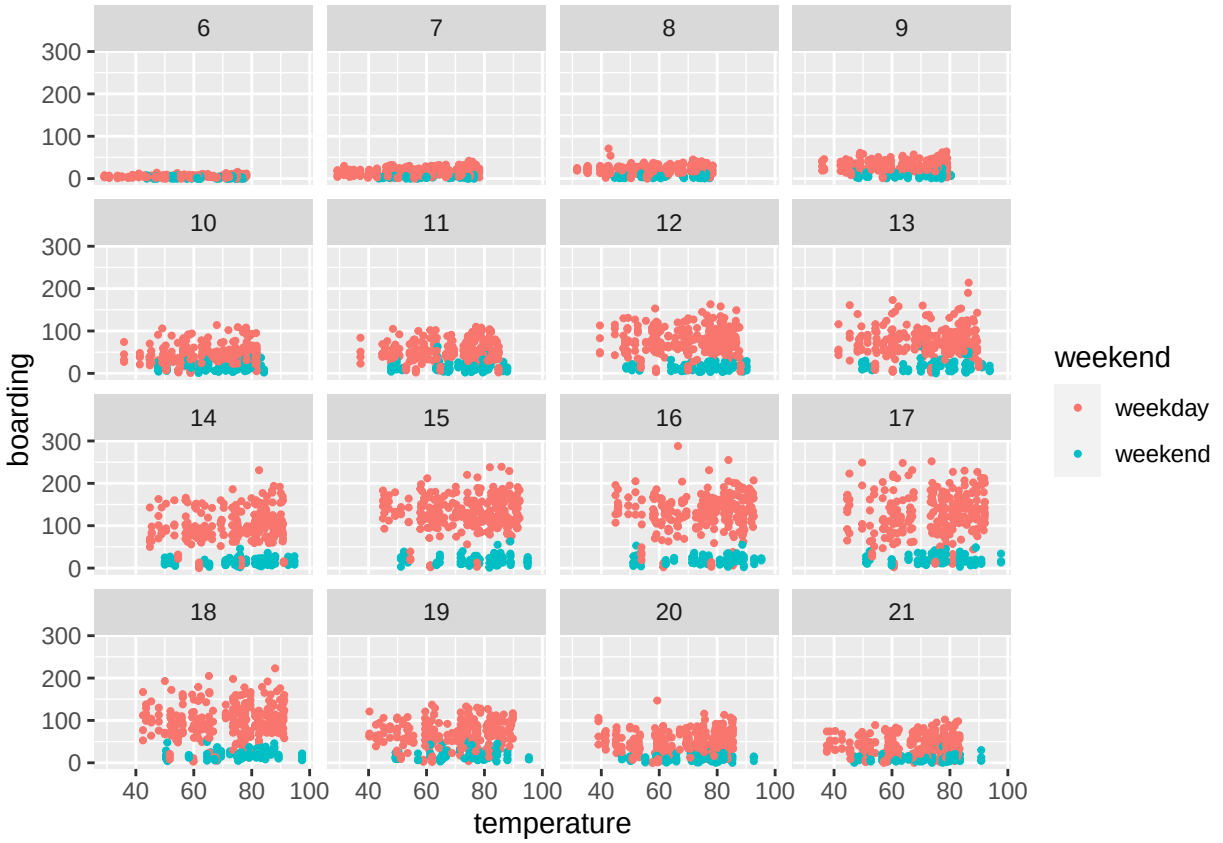
```
plot1 <- ggplot(metroavg) +
  geom_line(aes(x=hour_of_day, y=avg_boarding, color=month))

plot1+facet_wrap(. ~ day_of_week)
```



Here we can see that the weekday average boarding for each month has a similar behavior, but weekends are much lower. This is likely due to the fact that students are not needing rides to and from classes as much on the weekend. We can also see that the months seem to follow a similar pattern, no matter what day it is. This is interesting because in November it seems like there are less average boardings than September and October, but I suppose in Texas the weather is probably quite hot in September and October and a little cooler in November. Either that or people skip class more, which I hope is not the case.

#Question 1, Plot 2

```
plot2 <-ggplot(capmetro_UT, aes(x=temperature, y=boarding, color=weekend)) +
    geom_point(size=0.75)

plot2+facet_wrap(. ~ hour_of_day)
```



Each point represents a 15 minute window. We see that temperature does not affect boarding, they seem to mostly be clustered between 40 and 80 daily evenly. The hour of the day has more of an effect than temperature.

#Question 2

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v tibble  3.1.0     v purrr   0.3.4
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dbplyr::ident() masks dplyr::ident()
## x dplyr::lag()    masks stats::lag()
## x dbplyr::sql()   masks dplyr::sql()
```

```
library(ggplot2)
library(rsample)  # for creating train/test splits
library(caret)
```

```
## Loading required package: lattice
```

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

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

```
library(parallel)
library(foreach)
```

```
##
## Attaching package: 'foreach'

## The following objects are masked from 'package:purrr':
##
##     accumulate, when
```

```
library(modelr)
library(mosaic)
```

```
## Registered S3 method overwritten by 'mosaic':
##   method                           from
##   fortify.SpatialPolygonsDataFrame ggplot2

##
## The 'mosaic' package masks several functions from core packages in order to add
## additional features.  The original behavior of these functions should not be affected by this.

##
## Attaching package: 'mosaic'

## The following object is masked from 'package:Matrix':
##
##     mean

## The following object is masked from 'package:modelr':
##
##     resample

## The following object is masked from 'package:caret':
##
##     dotPlot

## The following object is masked from 'package:purrr':
##
##     cross

## The following object is masked from 'package:ggplot2':
##
##     stat
```

```
## The following objects are masked from 'package:dplyr':
##
##     count, do, tally

## The following objects are masked from 'package:stats':
##
##     binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##     quantile, sd, t.test, var

## The following objects are masked from 'package:base':
##
##     max, mean, min, prod, range, sample, sum
```

```
data(SaratogaHouses)

####
# Compare out-of-sample predictive performance
####

# Split into training and testing sets
saratoga_split = initial_split(SaratogaHouses, prop = 0.8)
saratoga_train = training(saratoga_split)
saratoga_test = testing(saratoga_split)

# Fit to the training data
# Sometimes it's easier to name the variables we want to leave out
# The command below yields exactly the same model.
# the dot (.) means "all variables not named"
# the minus (-) means "exclude this variable"
lm1 = lm(price ~ lotSize + bedrooms + bathrooms, data=saratoga_train)
lm2 = lm(price ~ . - pctCollege - sewer - waterfront - landValue - newConstruction, data=saratoga_train)
lm3 = lm(price ~ (. - pctCollege - sewer - waterfront - landValue - newConstruction)^2, data=saratoga_tr

coef(lm1) %>% round(0)
```

```
## (Intercept)     lotSize    bedrooms   bathrooms
##        2883       13138       18754       75368
```

```
coef(lm2) %>% round(0)
```

```
##            (Intercept)                lotSize                    age
##                  49703                   7379                     89
##             livingArea               bedrooms             fireplaces
##                     96                 -16612                   1398
##              bathrooms                  rooms heatinghot water/steam
##                  18621                   3010                 -12402
##        heatingelectric           fuelelectric                fueloil
##                  -6122                 -11903                  -6202
##           centralAirNo
##                 -18016
```

```
coef(lm3) %>% round(0)
```

```
##                           (Intercept)                               lotSize
##                                  4265                                 17025
##                                   age                             livingArea
##                                  -270                                    84
##                              bedrooms                             fireplaces
##                                 -5118                                -14447
##                             bathrooms                                 rooms
##                                 47546                                   902
##               heatinghot water/steam                       heatingelectric
##                                 46255                                  8703
##                          fuelelectric                              fueloil
##                                -18931                                 63110
##                          centralAirNo                          lotSize:age
##                                 18860                                  -222
##                    lotSize:livingArea                     lotSize:bedrooms
##                                    -9                                  6820
##                    lotSize:fireplaces                    lotSize:bathrooms
##                                 -8429                                  -941
##                         lotSize:rooms      lotSize:heatinghot water/steam
##                                   229                                20861
##               lotSize:heatingelectric                lotSize:fuelelectric
##                                  7742                                -1824
##                       lotSize:fueloil                lotSize:centralAirNo
##                                  5319                               -17295
##                        age:livingArea                         age:bedrooms
##                                     0                                   74
##                        age:fireplaces                        age:bathrooms
##                                    17                                  150
##                             age:rooms          age:heatinghot water/steam
##                                     1                                  228
##                   age:heatingelectric                   age:fuelelectric
##                                   934                                -709
##                          age:fueloil                   age:centralAirNo
##                                  -122                                  239
##                   livingArea:bedrooms                livingArea:fireplaces
##                                     2                                   18
##                  livingArea:bathrooms                     livingArea:rooms
##                                   -10                                    4
##     livingArea:heatinghot water/steam          livingArea:heatingelectric
##                                    -2                                  -60
##                livingArea:fuelelectric                   livingArea:fueloil
##                                    65                                  -40
##                livingArea:centralAirNo                 bedrooms:fireplaces
##                                   -22                                -7638
##                    bedrooms:bathrooms                       bedrooms:rooms
##                                 -1509                                -2035
##       bedrooms:heatinghot water/steam            bedrooms:heatingelectric
##                                  1764                                58637
##                  bedrooms:fuelelectric                     bedrooms:fueloil
##                                -49873                                -10824
##                  bedrooms:centralAirNo                  fireplaces:bathrooms
```

```
##                                7823                                4314
##                   fireplaces:rooms   fireplaces:heatinghot water/steam
##                               -1422                              -13156
##          fireplaces:heatingelectric             fireplaces:fuelelectric
##                               77969                              -61918
##                 fireplaces:fueloil            fireplaces:centralAirNo
##                               15004                               18607
##                    bathrooms:rooms    bathrooms:heatinghot water/steam
##                                1168                               -9851
##           bathrooms:heatingelectric              bathrooms:fuelelectric
##                              -11515                              -14033
##                  bathrooms:fueloil             bathrooms:centralAirNo
##                                4211                              -19498
##        rooms:heatinghot water/steam               rooms:heatingelectric
##                               -5140                              -24890
##                 rooms:fuelelectric                        rooms:fueloil
##                               20731                                4244
##                rooms:centralAirNo heatinghot water/steam:fuelelectric
##                                1053                              108430
##        heatingelectric:fuelelectric      heatinghot water/steam:fueloil
##                               11290                              -31253
##             heatingelectric:fueloil heatinghot water/steam:centralAirNo
##                                -223                               -8010
##        heatingelectric:centralAirNo           fuelelectric:centralAirNo
##                               58979                              -51246
##               fueloil:centralAirNo
##                               -1841
```

```
# Predictions out of sample
# Root mean squared error
rmse(lm1, saratoga_test)
```

```
## [1] 77878.29
```

```
#86024.66
rmse(lm2, saratoga_test)
```

```
## [1] 70360.61
```

```
#73803.68
rmse(lm3, saratoga_test)
```

```
## [1] 67292.41
```

```
#76782.72

#add different variables, interactions, for linear model another linear model and transfromations to sh
lm4 = lm(price ~ lotSize + bedrooms + bathrooms+ newConstruction + rooms + waterfront+ age + livingArea
lm5 = lm(price ~ . - pctCollege - sewer - centralAir - landValue - fireplaces - fuel - heating, data=sa
lm6 = lm(price ~ (. - pctCollege - sewer - centralAir - landValue - fireplaces - fuel - heating)^2, dat

coef(lm4) %>% round(0)
```

```
##       (Intercept)           lotSize          bedrooms         bathrooms
##            143937              5528            -17146             22309
## newConstructionNo             rooms      waterfrontNo               age
##             19125              3511            -143261               -37
##        livingArea
##               101
```

```
coef(lm5) %>% round(0)
```

```
##       (Intercept)           lotSize               age        livingArea
##            143937              5528               -37               101
##          bedrooms         bathrooms             rooms      waterfrontNo
##            -17146             22309              3511            -143261
## newConstructionNo
##             19125
```

```
coef(lm6) %>% round(0)
```

```
##                     (Intercept)                         lotSize
##                           22597                         -258800
##                             age                      livingArea
##                           26448                              -9
##                        bedrooms                       bathrooms
##                           99209                          136130
##                           rooms                     waterfrontNo
##                          -15952                         -61280
##               newConstructionNo                     lotSize:age
##                           28233                            -224
##              lotSize:livingArea                lotSize:bedrooms
##                             -10                            2401
##              lotSize:bathrooms                   lotSize:rooms
##                           -2828                            1948
##            lotSize:waterfrontNo       lotSize:newConstructionNo
##                          274688                           -2587
##                  age:livingArea                   age:bedrooms
##                               0                              69
##                   age:bathrooms                       age:rooms
##                            -110                              10
##                age:waterfrontNo          age:newConstructionNo
##                             147                          -26151
##             livingArea:bedrooms           livingArea:bathrooms
##                              -2                              17
##                livingArea:rooms         livingArea:waterfrontNo
##                               5                              45
##    livingArea:newConstructionNo             bedrooms:bathrooms
##                               7                           -8231
##                  bedrooms:rooms           bedrooms:waterfrontNo
##                           -3457                          -32600
##      bedrooms:newConstructionNo                 bathrooms:rooms
##                          -39604                             653
##           bathrooms:waterfrontNo     bathrooms:newConstructionNo
##                         -173405                           60355
##               rooms:waterfrontNo         rooms:newConstructionNo
```

```
##                           20853                           -1256
## waterfrontNo:newConstructionNo
##                             NA
```

```
# Predictions out of sample
# Root mean squared error
rmse(lm4, saratoga_test)
```

```
## [1] 66615.73
```

```
#67916.38
rmse(lm5, saratoga_test)
```

```
## [1] 66615.73
```

```
#67916.38
rmse(lm6, saratoga_test)
```

```
## Warning in predict.lm(model, data): prediction from a rank-deficient fit may be
## misleading
```

```
## [1] 69853.15
```

```
#67395.49

# KNN with K = 40
knn40 = knnreg(price ~ lotSize + bedrooms + bathrooms+ newConstruction + rooms + waterfront+ age + livi
rmse(knn40, saratoga_test)
```

```
## [1] 73966.14
```

```
# 80591.01

#KNN with K = 50
knn50 = knnreg(price ~ lotSize + bedrooms + bathrooms+ newConstruction + rooms + waterfront+ age + livi
rmse(knn50, saratoga_test)
```

```
## [1] 74192.85
```

```
#81222.5

#KNN with K = 20
knn20 = knnreg(price ~ lotSize + bedrooms + bathrooms+ newConstruction + rooms + waterfront+ age + livi
rmse(knn20, saratoga_test)
```

```
## [1] 74981.56
```

```
#80835.27

#This would indicate the best K is between K=20 and K=40, let's perhaps try K=30
# KNN with K = 30
knn30 = knnreg(price ~ lotSize + bedrooms + bathrooms+ newConstruction + rooms + waterfront+ age + livi
rmse(knn30, saratoga_test)
```

```
## [1] 74244.93
```

```
#80380.6
#K=30 did in fact produce the lowest RMSE so far
```

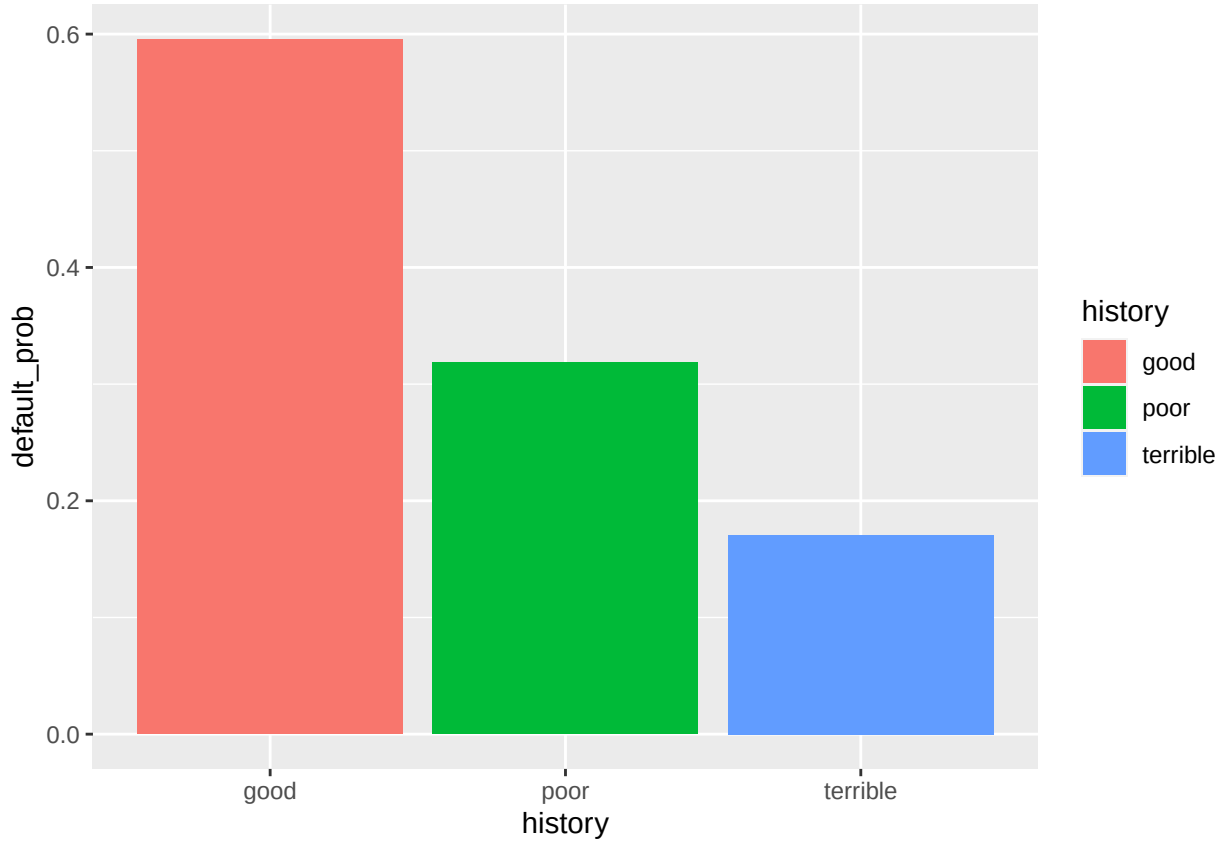
To create an effective price-modeling strategy, one should consider two various kinds of regression models: K-nearest neighbor design and linear regression models. Above, one can see that the RMSE for various linear regression models is lower than the training data set when tested. This demonstrates that the linear model performs well. One can also observe that the K-nearest neighbors RMSE's are above the linear model RMSE's, indicating that the linear model is the best prediction of pricing.

#Question 3

```
library(tidyverse)
library(readxl)
german_credit <- read_excel("~/Documents/UTX/DataMining/ECO395M-master/data/german_credit.xls")
```

```
## New names:
## * `` -> ...1
```

```
loans1 <-  german_credit %>%
  group_by(history) %>%
  summarize(default_prob = sum(Default == 1)/n())
ggplot(data=loans1) +
  geom_col(mapping = aes(x = history, y = default_prob, fill = history))
```




```
# a simple linear probability model
lm1 = glm(Default ~ history+duration+amount+installment+age+purpose+foreign, data=german_credit)
german_credit$lm1_pred = predict(lm1)


# in-sample accuracy?
yhat_train = ifelse(predict(lm1) >= 0.5, 1, 0)
table(y=german_credit$Default, yhat=yhat_train)
```

```
##    yhat
## y     0   1
##   0 651  49
##   1 215  85
```

```
# yhat
#y   0   1
# 0 651  49
# 1 215  85
```

```
fpr2=49/(49+651)
tpr2=85/(85+215)

fpr2
```

```
## [1] 0.07
```

```
#0.07
tpr2
```

```
## [1] 0.2833333
```

```
#0.2833333
```

The history variable seems to be a large component of predicting defaults. I think this is because of the way the bank used the sampling method. The false positive rate for the model is extremely low at 7% but the true positive rate is extremely low at only 28.3%. This means that the model does not predict defaults very well. The bank should perhaps reconsider their sampling scheme by taking a random sample and not adding in potential new data. The amoutn of defaults should be lower by nature, so there is no need to add.

#Question 4

```
library(tidyverse)
library(ggplot2)
library(naivebayes)
```

```
## naivebayes 0.9.7 loaded
```

```
library(modelr)
library(rsample)
library(foreach)
library(caret)
library(parallel)
library(readxl)
hotels_dev <- read_csv("~/Documents/UTX/DataMining/ECO395M-master/data/hotels_dev.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double(),
##   hotel = col_character(),
##   meal = col_character(),
##   market_segment = col_character(),
##   distribution_channel = col_character(),
##   reserved_room_type = col_character(),
##   assigned_room_type = col_character(),
##   deposit_type = col_character(),
##   customer_type = col_character(),
##   required_car_parking_spaces = col_character(),
##   arrival_date = col_date(format = "")
## )
## i Use `spec()` for the full column specifications.
```

```
hotels_val <- read_csv("~/Documents/UTX/DataMining/ECO395M-master/data/hotels_val.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
```

```
##   .default = col_double(),
##   hotel = col_character(),
##   meal = col_character(),
##   market_segment = col_character(),
##   distribution_channel = col_character(),
##   reserved_room_type = col_character(),
##   assigned_room_type = col_character(),
##   deposit_type = col_character(),
##   customer_type = col_character(),
##   required_car_parking_spaces = col_character(),
##   arrival_date = col_character()
## )
## i Use `spec()` for the full column specifications.
```

```
#Model Building#
####Compare out-of-sample predictive performance

# Split into training and testing sets
hotel1_split = initial_split(hotels_dev, prop = 0.8)
hotel1_train = training(hotel1_split)
hotel1_test = testing(hotel1_split)

# Fit to the training data
# Sometimes it's easier to name the variables we want to leave out
# The command below yields exactly the same model.
# the dot (.) means "all variables not named"
# the minus (-) means "exclude this variable"
lm7 = lm(children ~ market_segment + adults + customer_type + is_repeated_guest, data=hotel1_train)
lm8 = lm(children ~ . - arrival_date, data=hotel1_train)
lm9 = lm(children ~ (market_segment + adults + customer_type + is_repeated_guest + meal + reserved_room_

coef(lm7) %>% round(0)
```

```
##                  (Intercept)  market_segmentComplementary
##                            0                            0
##      market_segmentCorporate         market_segmentDirect
##                            0                            0
##         market_segmentGroups  market_segmentOffline_TA/TO
##                            0                            0
##      market_segmentOnline_TA                       adults
##                            0                            0
##           customer_typeGroup       customer_typeTransient
##                            0                            0
## customer_typeTransient-Party            is_repeated_guest
##                            0                            0
```

```
coef(lm8) %>% round(0)
```

```
##                         (Intercept)                  hotelResort_Hotel
##                                   0                                  0
##                           lead_time            stays_in_weekend_nights
##                                   0                                  0
##               stays_in_week_nights                              adults
```

```
##                                0                                0
##                           mealFB                           mealHB
##                                0                                0
##                           mealSC                    mealUndefined
##                                0                                0
##       market_segmentComplementary           market_segmentCorporate
##                                0                                0
##              market_segmentDirect              market_segmentGroups
##                                0                                0
##       market_segmentOffline_TA/TO           market_segmentOnline_TA
##                                0                                0
##        distribution_channelDirect           distribution_channelGDS
##                                0                                0
##         distribution_channelTA/TO                 is_repeated_guest
##                                0                                0
##            previous_cancellations     previous_bookings_not_canceled
##                                0                                0
##               reserved_room_typeB              reserved_room_typeC
##                                0                                1
##               reserved_room_typeD              reserved_room_typeE
##                                0                                0
##               reserved_room_typeF              reserved_room_typeG
##                                0                                0
##               reserved_room_typeH              reserved_room_typeL
##                                1                                0
##               assigned_room_typeB              assigned_room_typeC
##                                0                                0
##               assigned_room_typeD              assigned_room_typeE
##                                0                                0
##               assigned_room_typeF              assigned_room_typeG
##                                0                                0
##               assigned_room_typeH              assigned_room_typeI
##                                0                                0
##               assigned_room_typeK                   booking_changes
##                                0                                0
##            deposit_typeNon_Refund             deposit_typeRefundable
##                                0                                0
##              days_in_waiting_list                customer_typeGroup
##                                0                                0
##            customer_typeTransient       customer_typeTransient-Party
##                                0                                0
##                average_daily_rate required_car_parking_spacesparking
##                                0                                0
##         total_of_special_requests
##                                0
```

```
coef(lm9) %>% round(0)
```

```
##                                          (Intercept)
##                                                    0
##                                market_segmentComplementary
##                                                    0
##                                   market_segmentCorporate
##                                                    0
```

```
##                                 market_segmentDirect
##                                                    0
##                                market_segmentGroups
##                                                    0
##                           market_segmentOffline_TA/TO
##                                                    0
##                               market_segmentOnline_TA
##                                                    0
##                                               adults
##                                                    0
##                                   customer_typeGroup
##                                                    0
##                                customer_typeTransient
##                                                    0
##                           customer_typeTransient-Party
##                                                    0
##                                    is_repeated_guest
##                                                    0
##                                               mealFB
##                                                    1
##                                               mealHB
##                                                    0
##                                               mealSC
##                                                    0
##                                        mealUndefined
##                                                    0
##                                   reserved_room_typeB
##                                                    1
##                                   reserved_room_typeC
##                                                    1
##                                   reserved_room_typeD
##                                                    0
##                                   reserved_room_typeE
##                                                    0
##                                   reserved_room_typeF
##                                                    1
##                                   reserved_room_typeG
##                                                    1
##                                   reserved_room_typeH
##                                                    1
##                                   reserved_room_typeL
##                                                    0
##                                       booking_changes
##                                                    0
##                        required_car_parking_spacesparking
##                                                    0
##                                 total_of_special_requests
##                                                    0
##                        market_segmentComplementary:adults
##                                                    0
##                            market_segmentCorporate:adults
##                                                    0
##                               market_segmentDirect:adults
##                                                    0
```

```
##                                 market_segmentGroups:adults
##                                                           0
##                           market_segmentOffline_TA/TO:adults
##                                                           0
##                              market_segmentOnline_TA:adults
##                                                           0
##                  market_segmentComplementary:customer_typeGroup
##                                                           0
##                      market_segmentCorporate:customer_typeGroup
##                                                           0
##                         market_segmentDirect:customer_typeGroup
##                                                           0
##                         market_segmentGroups:customer_typeGroup
##                                                           0
##                  market_segmentOffline_TA/TO:customer_typeGroup
##                                                           0
##                      market_segmentOnline_TA:customer_typeGroup
##                                                          NA
##              market_segmentComplementary:customer_typeTransient
##                                                          -1
##                  market_segmentCorporate:customer_typeTransient
##                                                           0
##                     market_segmentDirect:customer_typeTransient
##                                                           0
##                     market_segmentGroups:customer_typeTransient
##                                                           0
##              market_segmentOffline_TA/TO:customer_typeTransient
##                                                           0
##                  market_segmentOnline_TA:customer_typeTransient
##                                                           0
##        market_segmentComplementary:customer_typeTransient-Party
##                                                           0
##            market_segmentCorporate:customer_typeTransient-Party
##                                                           0
##               market_segmentDirect:customer_typeTransient-Party
##                                                           0
##               market_segmentGroups:customer_typeTransient-Party
##                                                           0
##        market_segmentOffline_TA/TO:customer_typeTransient-Party
##                                                           0
##            market_segmentOnline_TA:customer_typeTransient-Party
##                                                          NA
##                   market_segmentComplementary:is_repeated_guest
##                                                           0
##                       market_segmentCorporate:is_repeated_guest
##                                                           0
##                          market_segmentDirect:is_repeated_guest
##                                                           0
##                          market_segmentGroups:is_repeated_guest
##                                                           0
##                   market_segmentOffline_TA/TO:is_repeated_guest
##                                                           0
##                       market_segmentOnline_TA:is_repeated_guest
##                                                           0
```

```
##                           market_segmentComplementary:mealFB
##                                                            0
##                               market_segmentCorporate:mealFB
##                                                            0
##                                  market_segmentDirect:mealFB
##                                                            0
##                                  market_segmentGroups:mealFB
##                                                            0
##                           market_segmentOffline_TA/TO:mealFB
##                                                            0
##                               market_segmentOnline_TA:mealFB
##                                                           NA
##                           market_segmentComplementary:mealHB
##                                                            0
##                               market_segmentCorporate:mealHB
##                                                            0
##                                  market_segmentDirect:mealHB
##                                                            0
##                                  market_segmentGroups:mealHB
##                                                            0
##                           market_segmentOffline_TA/TO:mealHB
##                                                            0
##                               market_segmentOnline_TA:mealHB
##                                                           NA
##                           market_segmentComplementary:mealSC
##                                                            0
##                               market_segmentCorporate:mealSC
##                                                            0
##                                  market_segmentDirect:mealSC
##                                                            0
##                                  market_segmentGroups:mealSC
##                                                            0
##                           market_segmentOffline_TA/TO:mealSC
##                                                            0
##                               market_segmentOnline_TA:mealSC
##                                                           NA
##                     market_segmentComplementary:mealUndefined
##                                                           -1
##                         market_segmentCorporate:mealUndefined
##                                                            0
##                            market_segmentDirect:mealUndefined
##                                                            0
##                            market_segmentGroups:mealUndefined
##                                                            0
##                     market_segmentOffline_TA/TO:mealUndefined
##                                                            0
##                         market_segmentOnline_TA:mealUndefined
##                                                           NA
##                  market_segmentComplementary:reserved_room_typeB
##                                                            0
##                      market_segmentCorporate:reserved_room_typeB
##                                                            0
##                        market_segmentDirect:reserved_room_typeB
##                                                            0
```

```
##                        market_segmentGroups:reserved_room_typeB
##                                                               0
##                  market_segmentOffline_TA/TO:reserved_room_typeB
##                                                               0
##                      market_segmentOnline_TA:reserved_room_typeB
##                                                              NA
##                  market_segmentComplementary:reserved_room_typeC
##                                                              -1
##                      market_segmentCorporate:reserved_room_typeC
##                                                               0
##                         market_segmentDirect:reserved_room_typeC
##                                                               0
##                         market_segmentGroups:reserved_room_typeC
##                                                               0
##                  market_segmentOffline_TA/TO:reserved_room_typeC
##                                                               0
##                      market_segmentOnline_TA:reserved_room_typeC
##                                                              NA
##                  market_segmentComplementary:reserved_room_typeD
##                                                               0
##                      market_segmentCorporate:reserved_room_typeD
##                                                               0
##                         market_segmentDirect:reserved_room_typeD
##                                                               0
##                         market_segmentGroups:reserved_room_typeD
##                                                               0
##                  market_segmentOffline_TA/TO:reserved_room_typeD
##                                                               0
##                      market_segmentOnline_TA:reserved_room_typeD
##                                                               0
##                  market_segmentComplementary:reserved_room_typeE
##                                                               0
##                      market_segmentCorporate:reserved_room_typeE
##                                                               0
##                         market_segmentDirect:reserved_room_typeE
##                                                               0
##                         market_segmentGroups:reserved_room_typeE
##                                                               0
##                  market_segmentOffline_TA/TO:reserved_room_typeE
##                                                               0
##                      market_segmentOnline_TA:reserved_room_typeE
##                                                               0
##                  market_segmentComplementary:reserved_room_typeF
##                                                              -1
##                      market_segmentCorporate:reserved_room_typeF
##                                                               0
##                         market_segmentDirect:reserved_room_typeF
##                                                               0
##                         market_segmentGroups:reserved_room_typeF
##                                                               0
##                  market_segmentOffline_TA/TO:reserved_room_typeF
##                                                               0
##                      market_segmentOnline_TA:reserved_room_typeF
##                                                              NA
```

```
##                 market_segmentComplementary:reserved_room_typeG
##                                                               0
##                     market_segmentCorporate:reserved_room_typeG
##                                                              -1
##                        market_segmentDirect:reserved_room_typeG
##                                                               0
##                        market_segmentGroups:reserved_room_typeG
##                                                              -1
##                 market_segmentOffline_TA/TO:reserved_room_typeG
##                                                               0
##                     market_segmentOnline_TA:reserved_room_typeG
##                                                              NA
##                 market_segmentComplementary:reserved_room_typeH
##                                                              -1
##                     market_segmentCorporate:reserved_room_typeH
##                                                              NA
##                        market_segmentDirect:reserved_room_typeH
##                                                               0
##                        market_segmentGroups:reserved_room_typeH
##                                                              NA
##                 market_segmentOffline_TA/TO:reserved_room_typeH
##                                                              NA
##                     market_segmentOnline_TA:reserved_room_typeH
##                                                              NA
##                 market_segmentComplementary:reserved_room_typeL
##                                                              NA
##                     market_segmentCorporate:reserved_room_typeL
##                                                              NA
##                        market_segmentDirect:reserved_room_typeL
##                                                              NA
##                        market_segmentGroups:reserved_room_typeL
##                                                              NA
##                 market_segmentOffline_TA/TO:reserved_room_typeL
##                                                              NA
##                     market_segmentOnline_TA:reserved_room_typeL
##                                                              NA
##                     market_segmentComplementary:booking_changes
##                                                               0
##                         market_segmentCorporate:booking_changes
##                                                               0
##                            market_segmentDirect:booking_changes
##                                                               0
##                            market_segmentGroups:booking_changes
##                                                               0
##                     market_segmentOffline_TA/TO:booking_changes
##                                                               0
##                         market_segmentOnline_TA:booking_changes
##                                                               0
##  market_segmentComplementary:required_car_parking_spacesparking
##                                                               0
##      market_segmentCorporate:required_car_parking_spacesparking
##                                                               0
##         market_segmentDirect:required_car_parking_spacesparking
##                                                               0
```

```
##         market_segmentGroups:required_car_parking_spacesparking
##                                                               0
##  market_segmentOffline_TA/TO:required_car_parking_spacesparking
##                                                               0
##      market_segmentOnline_TA:required_car_parking_spacesparking
##                                                               0
##           market_segmentComplementary:total_of_special_requests
##                                                               0
##               market_segmentCorporate:total_of_special_requests
##                                                               0
##                  market_segmentDirect:total_of_special_requests
##                                                               0
##                  market_segmentGroups:total_of_special_requests
##                                                               0
##           market_segmentOffline_TA/TO:total_of_special_requests
##                                                               0
##               market_segmentOnline_TA:total_of_special_requests
##                                                               0
##                                        adults:customer_typeGroup
##                                                               0
##                                   adults:customer_typeTransient
##                                                               0
##                               adults:customer_typeTransient-Party
##                                                               0
##                                      adults:is_repeated_guest
##                                                               0
##                                                  adults:mealFB
##                                                               0
##                                                  adults:mealHB
##                                                               0
##                                                  adults:mealSC
##                                                               0
##                                             adults:mealUndefined
##                                                               0
##                                        adults:reserved_room_typeB
##                                                               0
##                                        adults:reserved_room_typeC
##                                                               0
##                                        adults:reserved_room_typeD
##                                                               0
##                                        adults:reserved_room_typeE
##                                                               0
##                                        adults:reserved_room_typeF
##                                                               0
##                                        adults:reserved_room_typeG
##                                                               0
##                                        adults:reserved_room_typeH
##                                                               0
##                                        adults:reserved_room_typeL
##                                                               0
##                                           adults:booking_changes
##                                                               0
##                         adults:required_car_parking_spacesparking
##                                                               0
```

```
##                              adults:total_of_special_requests
##                                                             0
##                          customer_typeGroup:is_repeated_guest
##                                                             0
##                       customer_typeTransient:is_repeated_guest
##                                                             0
##                    customer_typeTransient-Party:is_repeated_guest
##                                                             0
##                                  customer_typeGroup:mealFB
##                                                             0
##                               customer_typeTransient:mealFB
##                                                             0
##                           customer_typeTransient-Party:mealFB
##                                                             0
##                                  customer_typeGroup:mealHB
##                                                             0
##                               customer_typeTransient:mealHB
##                                                             0
##                           customer_typeTransient-Party:mealHB
##                                                             0
##                                  customer_typeGroup:mealSC
##                                                             0
##                               customer_typeTransient:mealSC
##                                                             0
##                           customer_typeTransient-Party:mealSC
##                                                             0
##                             customer_typeGroup:mealUndefined
##                                                             0
##                          customer_typeTransient:mealUndefined
##                                                             0
##                       customer_typeTransient-Party:mealUndefined
##                                                             0
##                          customer_typeGroup:reserved_room_typeB
##                                                             0
##                       customer_typeTransient:reserved_room_typeB
##                                                             0
##                    customer_typeTransient-Party:reserved_room_typeB
##                                                             0
##                          customer_typeGroup:reserved_room_typeC
##                                                             0
##                       customer_typeTransient:reserved_room_typeC
##                                                             0
##                    customer_typeTransient-Party:reserved_room_typeC
##                                                             0
##                          customer_typeGroup:reserved_room_typeD
##                                                             0
##                       customer_typeTransient:reserved_room_typeD
##                                                             0
##                    customer_typeTransient-Party:reserved_room_typeD
##                                                             0
##                          customer_typeGroup:reserved_room_typeE
##                                                             0
##                       customer_typeTransient:reserved_room_typeE
##                                                             0
```

```
##                customer_typeTransient-Party:reserved_room_typeE
##                                                                0
##                          customer_typeGroup:reserved_room_typeF
##                                                                0
##                       customer_typeTransient:reserved_room_typeF
##                                                                0
##                customer_typeTransient-Party:reserved_room_typeF
##                                                                0
##                          customer_typeGroup:reserved_room_typeG
##                                                                0
##                       customer_typeTransient:reserved_room_typeG
##                                                                0
##                customer_typeTransient-Party:reserved_room_typeG
##                                                                0
##                          customer_typeGroup:reserved_room_typeH
##                                                                0
##                       customer_typeTransient:reserved_room_typeH
##                                                                0
##                customer_typeTransient-Party:reserved_room_typeH
##                                                               NA
##                          customer_typeGroup:reserved_room_typeL
##                                                               NA
##                       customer_typeTransient:reserved_room_typeL
##                                                               NA
##                customer_typeTransient-Party:reserved_room_typeL
##                                                               NA
##                                customer_typeGroup:booking_changes
##                                                                0
##                           customer_typeTransient:booking_changes
##                                                                0
##                      customer_typeTransient-Party:booking_changes
##                                                                0
##            customer_typeGroup:required_car_parking_spacesparking
##                                                                0
##       customer_typeTransient:required_car_parking_spacesparking
##                                                                0
## customer_typeTransient-Party:required_car_parking_spacesparking
##                                                                0
##                      customer_typeGroup:total_of_special_requests
##                                                                0
##                customer_typeTransient:total_of_special_requests
##                                                                0
##          customer_typeTransient-Party:total_of_special_requests
##                                                                0
##                                           is_repeated_guest:mealFB
##                                                               -1
##                                           is_repeated_guest:mealHB
##                                                                0
##                                           is_repeated_guest:mealSC
##                                                                0
##                                  is_repeated_guest:mealUndefined
##                                                                0
##                             is_repeated_guest:reserved_room_typeB
##                                                                0
```

```
##                           is_repeated_guest:reserved_room_typeC
##                                                               0
##                           is_repeated_guest:reserved_room_typeD
##                                                               0
##                           is_repeated_guest:reserved_room_typeE
##                                                               0
##                           is_repeated_guest:reserved_room_typeF
##                                                               0
##                           is_repeated_guest:reserved_room_typeG
##                                                               0
##                           is_repeated_guest:reserved_room_typeH
##                                                               0
##                           is_repeated_guest:reserved_room_typeL
##                                                              NA
##                               is_repeated_guest:booking_changes
##                                                               0
##            is_repeated_guest:required_car_parking_spacesparking
##                                                               0
##                     is_repeated_guest:total_of_special_requests
##                                                               0
##                                    mealFB:reserved_room_typeB
##                                                              NA
##                                    mealHB:reserved_room_typeB
##                                                               0
##                                    mealSC:reserved_room_typeB
##                                                               0
##                               mealUndefined:reserved_room_typeB
##                                                              NA
##                                    mealFB:reserved_room_typeC
##                                                               0
##                                    mealHB:reserved_room_typeC
##                                                               0
##                                    mealSC:reserved_room_typeC
##                                                               0
##                               mealUndefined:reserved_room_typeC
##                                                               0
##                                    mealFB:reserved_room_typeD
##                                                               0
##                                    mealHB:reserved_room_typeD
##                                                               0
##                                    mealSC:reserved_room_typeD
##                                                               0
##                               mealUndefined:reserved_room_typeD
##                                                               0
##                                    mealFB:reserved_room_typeE
##                                                               0
##                                    mealHB:reserved_room_typeE
##                                                               0
##                                    mealSC:reserved_room_typeE
##                                                               0
##                               mealUndefined:reserved_room_typeE
##                                                               0
##                                    mealFB:reserved_room_typeF
##                                                              -1
```

```
##                                 mealHB:reserved_room_typeF
##                                                          0
##                                 mealSC:reserved_room_typeF
##                                                          0
##                            mealUndefined:reserved_room_typeF
##                                                         -1
##                                 mealFB:reserved_room_typeG
##                                                          0
##                                 mealHB:reserved_room_typeG
##                                                          0
##                                 mealSC:reserved_room_typeG
##                                                         -1
##                            mealUndefined:reserved_room_typeG
##                                                          0
##                                 mealFB:reserved_room_typeH
##                                                         NA
##                                 mealHB:reserved_room_typeH
##                                                          0
##                                 mealSC:reserved_room_typeH
##                                                         NA
##                            mealUndefined:reserved_room_typeH
##                                                         NA
##                                 mealFB:reserved_room_typeL
##                                                         NA
##                                 mealHB:reserved_room_typeL
##                                                         NA
##                                 mealSC:reserved_room_typeL
##                                                         NA
##                            mealUndefined:reserved_room_typeL
##                                                         NA
##                                      mealFB:booking_changes
##                                                          0
##                                      mealHB:booking_changes
##                                                          0
##                                      mealSC:booking_changes
##                                                          0
##                                 mealUndefined:booking_changes
##                                                          0
##                       mealFB:required_car_parking_spacesparking
##                                                          0
##                       mealHB:required_car_parking_spacesparking
##                                                          0
##                       mealSC:required_car_parking_spacesparking
##                                                          0
##                  mealUndefined:required_car_parking_spacesparking
##                                                          0
##                                  mealFB:total_of_special_requests
##                                                          0
##                                  mealHB:total_of_special_requests
##                                                          0
##                                  mealSC:total_of_special_requests
##                                                          0
##                             mealUndefined:total_of_special_requests
##                                                          0
```

```
##                             reserved_room_typeB:booking_changes
##                                                               0
##                             reserved_room_typeC:booking_changes
##                                                               0
##                             reserved_room_typeD:booking_changes
##                                                               0
##                             reserved_room_typeE:booking_changes
##                                                               0
##                             reserved_room_typeF:booking_changes
##                                                               0
##                             reserved_room_typeG:booking_changes
##                                                               0
##                             reserved_room_typeH:booking_changes
##                                                               0
##                             reserved_room_typeL:booking_changes
##                                                              NA
##          reserved_room_typeB:required_car_parking_spacesparking
##                                                               0
##          reserved_room_typeC:required_car_parking_spacesparking
##                                                               0
##          reserved_room_typeD:required_car_parking_spacesparking
##                                                               0
##          reserved_room_typeE:required_car_parking_spacesparking
##                                                               0
##          reserved_room_typeF:required_car_parking_spacesparking
##                                                               0
##          reserved_room_typeG:required_car_parking_spacesparking
##                                                               0
##          reserved_room_typeH:required_car_parking_spacesparking
##                                                               0
##          reserved_room_typeL:required_car_parking_spacesparking
##                                                              NA
##                    reserved_room_typeB:total_of_special_requests
##                                                               0
##                    reserved_room_typeC:total_of_special_requests
##                                                               0
##                    reserved_room_typeD:total_of_special_requests
##                                                               0
##                    reserved_room_typeE:total_of_special_requests
##                                                               0
##                    reserved_room_typeF:total_of_special_requests
##                                                               0
##                    reserved_room_typeG:total_of_special_requests
##                                                               0
##                    reserved_room_typeH:total_of_special_requests
##                                                               0
##                    reserved_room_typeL:total_of_special_requests
##                                                              NA
##                 booking_changes:required_car_parking_spacesparking
##                                                               0
##                         booking_changes:total_of_special_requests
##                                                               0
##    required_car_parking_spacesparking:total_of_special_requests
##                                                               0
```

```
# Predictions out of sample
# Root mean squared error
rmse(lm7, hotel1_test)
```

```
## [1] 0.2653464
```

```
#0.2684471
rmse(lm8, hotel1_test)
```

```
## [1] 0.2330057
```

```
#0.232954
rmse(lm9, hotel1_test)
```

```
## Warning in predict.lm(model, data): prediction from a rank-deficient fit may be
## misleading
```

```
## [1] 0.2288163
```

```
#0.2279379
validation2 = predict(lm9, newdata=hotels_dev)
```

```
## Warning in predict.lm(lm9, newdata = hotels_dev): prediction from a rank-
## deficient fit may be misleading
```

```
yhat_test2 = ifelse(validation2 >= 0.5, 1, 0)
confusion_in=table(y=hotels_dev$children, yhat1=yhat_test2)
confusion_in
```

```
##    yhat1
## y       0     1
##   0 40798   567
##   1  2189  1446
```

```
#yhat1
#y       0     1
#0 40810   555
#1  2208  1427

fpr1=0.013417
tpr1=0.393572
```

```
#model validation: Step 1 - validate
lm10 = lm(children ~ (market_segment + adults + customer_type + is_repeated_guest + meal + reserved_roo
validation1 = predict(lm10, newdata=hotels_val)
```

```
## Warning in predict.lm(lm10, newdata = hotels_val): prediction from a rank-
## deficient fit may be misleading
```

```
yhat_test = ifelse(validation1 >= 0.5, 1, 0)
confusion_out=table(y=hotels_val$children, yhat=yhat_test)
confusion_out
```

```
##    yhat
## y      0    1
##   0 4543   54
##   1  231  171
```

```
#yhat
#y      0    1
#0 4543   54
#1  231  171
#error rate = (250+83)/ 4999 or about 5.7% error, accuracy of 94.3% accuracy
fpr=0.011747
tpr=0.425373
```

```
library(PRROC)
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:mosaic':
##
##     cov, var
```

```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
library(ROCit)
```

```
##
## Attaching package: 'ROCit'
```

```
## The following object is masked from 'package:mosaic':
##
##     logit
```

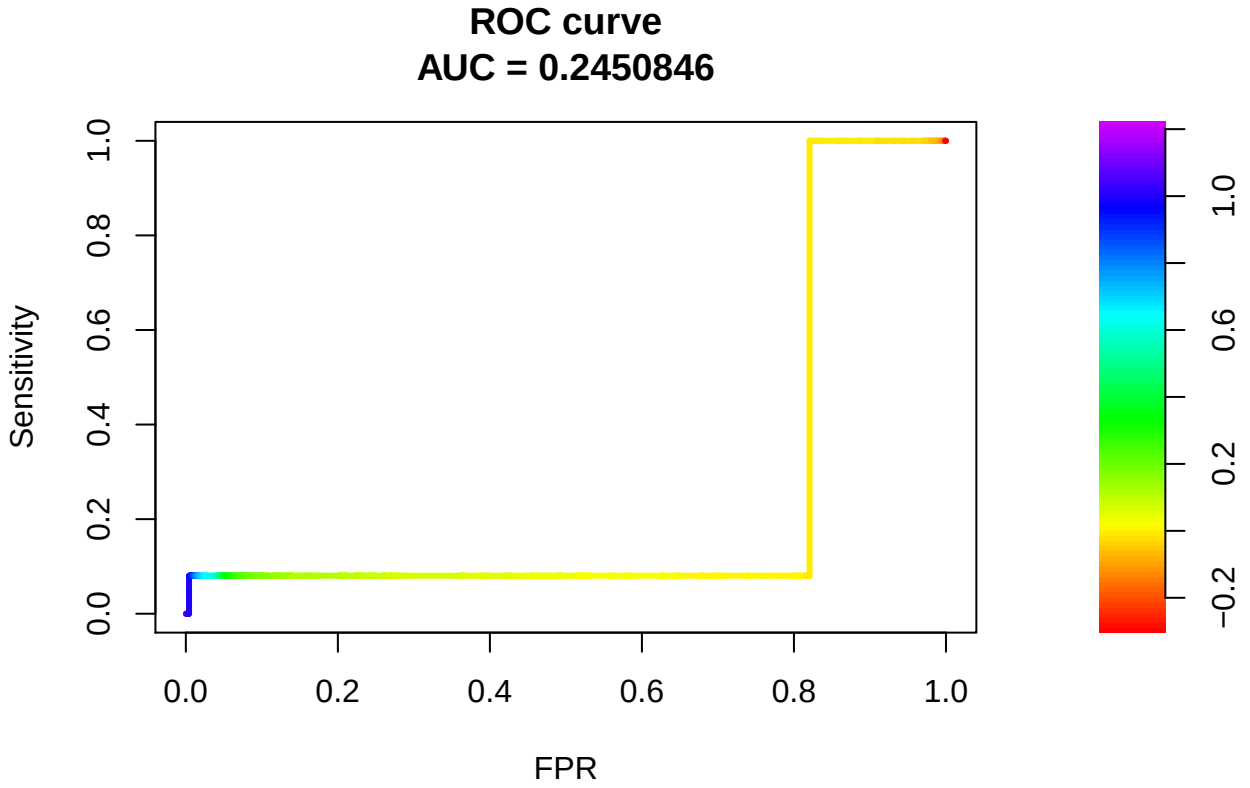
```
library(ROCR)
# check imbalance on training set
table(hotels_val$children)
```

```
##
##    0    1
## 4597  402
```

```
# model estimation using logistic regression
lm10 = lm(children ~ (market_segment + adults + customer_type + is_repeated_guest + meal + reserved_roo

# prediction on training set
pred.lm10 <- predict(lm10, newdata=hotels_val)
```

```
## Warning in predict.lm(lm10, newdata = hotels_val): prediction from a rank-
## deficient fit may be misleading
```

```
# add the ROC curve (test set)
roc1=roc.curve(hotels_val$children, pred.lm10, curve=TRUE)

plot(roc1)
```




```
#model validation: Step 2 - folds
K_folds = 20

hotels_val = hotels_val %>%
  mutate(fold_id = rep(1:K_folds, length=nrow(hotels_val)) %>% sample)


# now loop over folds
rmse_cv = foreach(fold = 1:K_folds, .combine='c') %do% {
  lm10 = knnreg(children ~ (market_segment + adults + customer_type + is_repeated_guest + meal + reserve
```

```
                  data=filter(hotels_val), k=20)
  modelr::rmse(lm10, data=filter(hotels_val))
}

rmse_cv
```

```
##  [1] 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592
## [10] 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592
## [19] 0.24592 0.24592
```

```
#[1] 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592
#[8] 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592
#[15] 0.24592 0.24592 0.24592 0.24592 0.24592 0.24592
#mean(rmse_cv)  # mean CV error
#0.24592
sd(rmse_cv)/sqrt(K_folds)
```

```
## [1] 0
```

```
#0
```