# IDS_HW3_NY

Navya Yadagiri 674788385,Sayali Bonawale 656488690 ,Jona 651224838

3/17/2022

## Retention Modeling at Scholastic Travel Company (A)

####Installing the necessary packages

```
#install.packages("lubridate")
library(lubridate)

##
## Attaching package: 'lubridate'

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

library(skimr)

## Warning: package 'skimr' was built under R version 4.1.3

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(devtools)
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse
1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v stringr 1.4.0
## v tidyr   1.1.4     v forcats 0.5.1
## v readr   2.0.2

## -- Conflicts ------------------------------------------
tidyverse_conflicts() --
## x lubridate::as.difftime() masks base::as.difftime()
## x lubridate::date()        masks base::date()
## x dplyr::filter()          masks stats::filter()
## x lubridate::intersect()   masks base::intersect()
## x dplyr::lag()             masks stats::lag()
## x lubridate::setdiff()     masks base::setdiff()
## x lubridate::union()       masks base::union()
```

```
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 4.1.3
```

```
## randomForest 4.7-1
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:psych':
##
##     outlier
```

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

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
#ools::install_github("ropensci/visdat")
library(visdat)
```

```
## Warning: package 'visdat' was built under R version 4.1.3
```

```
library("funModeling")
```

```
## Warning: package 'funModeling' was built under R version 4.1.3
```

```
## Loading required package: Hmisc
```

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

```
## Loading required package: survival
## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following object is masked from 'package:psych':
##
##     describe

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units

## funModeling v.1.9.4 :)
## Examples and tutorials at livebook.datascienceheroes.com
##  / Now in Spanish: librovivodecienciadedatos.ai

library("Hmisc")
library("rpart")
library("caret")

##
## Attaching package: 'caret'

## The following object is masked from 'package:survival':
##
##     cluster

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

library("rpart.plot")
```

####Changing the datatype. We converted numerical variable to numerical values and categorical to categorical values

```
original_dataset <- readxl::read_excel("Scholastic Travel.xlsx", sheet =
"Exhibit 1 -- Data");

head(original_dataset)

## # A tibble: 6 x 56
##      ID Program.Code From.Grade To.Grade Group.State Is.Non.Annual.  Days
##   <dbl> <chr>        <chr>      <chr>    <chr>                <dbl> <dbl>
## 1     1 HS           4          4        CA                       0     1
## 2     2 HC           8          8        AZ                       0     7
## 3     3 HD           8          8        FL                       0     3
## 4     4 HN           9          12       VA                       1     3
## 5     5 HD           6          8        FL                       0     6
## 6     6 HC          10         12        LA                       0     4
## # ... with 49 more variables: Travel.Type <chr>, Departure.Date <dttm>,
## #   Return.Date <dttm>, Deposit.Date <dttm>, Special.Pay <chr>, Tuition
<dbl>,
## #   FRP.Active <dbl>, FRP.Cancelled <dbl>, FRP.Take.up.percent. <dbl>,
## #   Early.RPL <chr>, Latest.RPL <chr>, Cancelled.Pax <dbl>,
## #   Total.Discount.Pax <dbl>, Initial.System.Date <chr>, Poverty.Code
<chr>,
## #   Region <chr>, CRM.Segment <chr>, School.Type <chr>,
## #   Parent.Meeting.Flag <dbl>, MDR.Low.Grade <chr>, MDR.High.Grade <chr>,
...
```

```
#There are many variables whose data types needs to be changed, and also
contains NA Values
summary(original_dataset)
```

```
##        ID       Program.Code        From.Grade          To.Grade
##  Min.   :   1   Length:2389        Length:2389        Length:2389
##  1st Qu.: 598   Class :character   Class :character   Class :character
##  Median :1195   Mode  :character   Mode  :character   Mode  :character
##  Mean   :1195
##  3rd Qu.:1792
##  Max.   :2389
##
##  Group.State        Is.Non.Annual.       Days        Travel.Type
##  Length:2389        Min.   :0.000   Min.   : 1.000   Length:2389
##  Class :character   1st Qu.:0.000   1st Qu.: 4.000   Class :character
##  Mode  :character   Median :0.000   Median : 5.000   Mode  :character
##                     Mean   :0.154   Mean   : 4.575
##                     3rd Qu.:0.000   3rd Qu.: 5.000
##                     Max.   :1.000   Max.   :12.000
##
##  Departure.Date                  Return.Date
##  Min.   :2011-01-14 00:00:00   Min.   :2011-01-14 00:00:00
##  1st Qu.:2011-04-09 00:00:00   1st Qu.:2011-04-12 00:00:00
##  Median :2011-05-17 00:00:00   Median :2011-05-20 00:00:00
##  Mean   :2011-05-07 18:20:38   Mean   :2011-05-11 11:57:53
##  3rd Qu.:2011-06-07 00:00:00   3rd Qu.:2011-06-10 00:00:00
##  Max.   :2011-06-30 00:00:00   Max.   :2011-07-05 00:00:00
##
##   Deposit.Date                 Special.Pay           Tuition
##  Min.   :2009-09-25 00:00:00   Length:2389        Min.   :  79
##  1st Qu.:2010-10-15 00:00:00   Class :character   1st Qu.:1174
##  Median :2010-10-28 00:00:00   Mode  :character   Median :1700
##  Mean   :2010-10-24 19:42:37                      Mean   :1615
##  3rd Qu.:2010-11-05 00:00:00                      3rd Qu.:2048
##  Max.   :2011-10-30 00:00:00                      Max.   :4200
##
##    FRP.Active     FRP.Cancelled    FRP.Take.up.percent.  Early.RPL
##  Min.   :  0.00   Min.   : 0.000   Min.   :0.0000       Length:2389
##  1st Qu.:  6.00   1st Qu.: 1.000   1st Qu.:0.4550       Class :character
##  Median : 12.00   Median : 2.000   Median :0.6000       Mode  :character
##  Mean   : 16.87   Mean   : 3.306   Mean   :0.5707
##  3rd Qu.: 23.00   3rd Qu.: 4.000   3rd Qu.:0.7270
##  Max.   :257.00   Max.   :45.000   Max.   :1.0000
##
##   Latest.RPL        Cancelled.Pax    Total.Discount.Pax
Initial.System.Date
##  Length:2389        Min.   : 0.000   Min.   : 0.000     Length:2389
##  Class :character   1st Qu.: 2.000   1st Qu.: 1.000     Class :character
##  Mode  :character   Median : 4.000   Median : 2.000     Mode  :character
##                     Mean   : 4.807   Mean   : 2.954
##                     3rd Qu.: 6.000   3rd Qu.: 4.000
##                     Max.   :39.000   Max.   :47.000
##
##  Poverty.Code          Region          CRM.Segment        School.Type
##  Length:2389        Length:2389        Length:2389        Length:2389
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Parent.Meeting.Flag MDR.Low.Grade      MDR.High.Grade
##  Min.   :0.0000      Length:2389        Length:2389
##  1st Qu.:1.0000      Class :character   Class :character
##  Median :1.0000      Mode  :character   Mode  :character
##  Mean   :0.8589
##  3rd Qu.:1.0000
##  Max.   :1.0000
##
##  Total.School.Enrollment Income.Level       EZ.Pay.Take.Up.Rate
##  Min.   :  19.0          Length:2389        Min.   :0.0000
##  1st Qu.: 360.0          Class :character   1st Qu.:0.1000
##  Median : 597.0          Mode  :character   Median :0.2000
##  Mean   : 648.4                             Mean   :0.2079
##  3rd Qu.: 825.8                             3rd Qu.:0.2920
##  Max.   :3990.0                             Max.   :1.7500
##  NA's   :91
##  School.Sponsor   SPR.Product.Type   SPR.New.Existing        FPP
##  Min.   :0.0000   Length:2389        Length:2389        Min.   :  2.0
##  1st Qu.:0.0000   Class :character   Class :character   1st Qu.: 12.0
##  Median :0.0000   Mode  :character   Mode  :character   Median : 23.0
##  Mean   :0.1059                                         Mean   : 31.3
##  3rd Qu.:0.0000                                         3rd Qu.: 41.0
##  Max.   :1.0000                                         Max.   :286.0
##
##    Total.Pax      SPR.Group.Revenue NumberOfMeetingswithParents
##  Min.   :  2.00   Min.   :  79      Min.   :0.000
##  1st Qu.: 14.00   1st Qu.:1174      1st Qu.:1.000
##  Median : 26.00   Median :1700      Median :1.000
##  Mean   : 34.25   Mean   :1615      Mean   :1.102
##  3rd Qu.: 44.00   3rd Qu.:2048      3rd Qu.:1.000
##  Max.   :313.00   Max.   :4200      Max.   :2.000
##
##  FirstMeeting       LastMeeting        DifferenceTraveltoFirstMeeting
##  Length:2389        Length:2389        Length:2389
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  DifferenceTraveltoLastMeeting SchoolGradeTypeLow SchoolGradeTypeHigh
##  Length:2389                   Length:2389        Length:2389
##  Class :character              Class :character   Class :character
##  Mode  :character              Mode  :character   Mode  :character
##
##
##
##
##  SchoolGradeType    DepartureMonth     GroupGradeTypeLow
GroupGradeTypeHigh
##  Length:2389        Length:2389        Length:2389        Length:2389
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  GroupGradeType     MajorProgramCode   SingleGradeTripFlag
##  Length:2389        Length:2389        Min.   :0.0000
##  Class :character   Class :character   1st Qu.:0.0000
##  Mode  :character   Mode  :character   Median :1.0000
##                                        Mean   :0.5567
##                                        3rd Qu.:1.0000
##                                        Max.   :1.0000
##
##  FPP.to.School.enrollment   FPP.to.PAX     Num.of.Non_FPP.PAX
##  Length:2389              Min.   :0.6000   Min.   : 0.000
##  Class :character         1st Qu.:0.8824   1st Qu.: 1.000
##  Mode  :character         Median :0.9091   Median : 2.000
##                           Mean   :0.9007   Mean   : 2.954
##                           3rd Qu.:0.9333   3rd Qu.: 4.000
##                           Max.   :1.0000   Max.   :47.000
##
##  SchoolSizeIndicator Retained.in.2012.
##  Length:2389         Min.   :0.0000
##  Class :character    1st Qu.:0.0000
##  Mode  :character    Median :1.0000
##                      Mean   :0.6074
##                      3rd Qu.:1.0000
##                      Max.   :1.0000
##
```

```
colsNumerical <-
c("Days","Tuition","FRP.Active","FRP.Cancelled","FRP.Take.up.percent.","Cance
lled.Pax","Total.Discount.Pax","Total.School.Enrollment","EZ.Pay.Take.Up.Rate
","FPP","Total.Pax","SPR.Group.Revenue","NumberOfMeetingswithParents","Differ
enceTraveltoFirstMeeting","DifferenceTraveltoLastMeeting","FPP.to.School.enro
llment","FPP.to.PAX","Num.of.Non_FPP.PAX")

#There are 18 numerical
length(colsNumerical)
```

```
## [1] 18
```

```
colsCategorical <-
c("Program.Code","From.Grade","To.Grade","Group.State","Is.Non.Annual.","Trav
el.Type","Special.Pay","Poverty.Code","Region","CRM.Segment","School.Type","P
arent.Meeting.Flag","MDR.Low.Grade","MDR.High.Grade","Income.Level","School.S
ponsor","SPR.Product.Type","SPR.New.Existing","SchoolGradeTypeLow","SchoolGra
deTypeHigh","SchoolGradeType","DepartureMonth","GroupGradeTypeLow","GroupGrad
eTypeHigh","GroupGradeType","MajorProgramCode","SingleGradeTripFlag","SchoolS
izeIndicator","Retained.in.2012.")


length(colsCategorical)
```

```
## [1] 29
```

```
#Changing the dataset columns to the actual dataypes
dataset <- data.frame(original_dataset)
dataset[colsNumerical] <- lapply(original_dataset[colsNumerical], as.numeric)
```

```
## Warning in lapply(original_dataset[colsNumerical], as.numeric): NAs
introduced
## by coercion

## Warning in lapply(original_dataset[colsNumerical], as.numeric): NAs
introduced
## by coercion

## Warning in lapply(original_dataset[colsNumerical], as.numeric): NAs
introduced
## by coercion
```

```
dataset[colsCategorical] <- lapply(original_dataset[colsCategorical],
as.factor)
```

```
glimpse(dataset)

## Rows: 2,389
## Columns: 56
## $ ID                             <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,
12, ~
## $ Program.Code                   <fct> HS, HC, HD, HN, HD, HC, SG, FN, CC,
HD,~
## $ From.Grade                     <fct> 4, 8, 8, 9, 6, 10, 11, 9, 8, 8, 8,
8, 8~
## $ To.Grade                       <fct> 4, 8, 8, 12, 8, 12, 12, 9, 8, 8, 8,
8, ~
## $ Group.State                    <fct> CA, AZ, FL, VA, FL, LA, MA, MX, AZ,
TX,~
## $ Is.Non.Annual.                 <fct> 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0,
0, ~
## $ Days                           <dbl> 1, 7, 3, 3, 6, 4, 6, 8, 8, 4, 4, 4,
6, ~
## $ Travel.Type                    <fct> A, A, A, B, T, A, A, A, A, A, A, A,
A, ~
## $ Departure.Date                 <dttm> 2011-01-14, 2011-01-14, 2011-01-
15, 20~
## $ Return.Date                    <dttm> 2011-01-14, 2011-01-21, 2011-01-
17, 20~
## $ Deposit.Date                   <dttm> 2010-08-30, 2009-11-15, 2010-10-
15, 20~
## $ Special.Pay                    <fct> NA, CP, NA, NA, NA, NA, NA, NA, CP,
NA,~
## $ Tuition                        <dbl> 424, 2350, 1181, 376, 865, 2025,
1977, ~
## $ FRP.Active                     <dbl> 25, 9, 17, 0, 40, 9, 16, 10, 30,
51, 47~
## $ FRP.Cancelled                  <dbl> 3, 9, 6, 0, 8, 4, 4, 0, 0, 1, 1, 0,
6, ~
## $ FRP.Take.up.percent.           <dbl> 0.424, 0.409, 0.708, 0.000, 0.494,
0.90~
## $ Early.RPL                      <chr> "40266", "40106", "40297", "NA",
"40266~
## $ Latest.RPL                     <chr> "40402", "40400", "40406", "NA",
"40402~
## $ Cancelled.Pax                  <dbl> 3, 11, 6, 1, 9, 3, 5, 1, 0, 1, 1,
0, 6,~
## $ Total.Discount.Pax             <dbl> 4, 3, 3, 0, 8, 1, 2, 1, 4, 6, 4, 5,
1, ~
## $ Initial.System.Date            <chr> "40263", "40088", "40206", "40470",
"40~
## $ Poverty.Code                   <fct> B, C, C, NA, D, C, NA, NA, NA, NA,
NA, ~
## $ Region                         <fct> Southern California, Other, Other,
Othe~
## $ CRM.Segment                    <fct> 4, 10, 10, 7, 10, 8, 8, 7, 5, 5,
10, 10~
## $ School.Type                    <fct> PUBLIC, PUBLIC, PUBLIC, CHD,
PUBLIC, PU~
## $ Parent.Meeting.Flag            <fct> 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1,
1, ~
## $ MDR.Low.Grade                  <fct> K, 7, 6, NA, 6, 10, 9, NA, 6, PK,
K, PK~
## $ MDR.High.Grade                 <fct> 5, 8, 8, NA, 8, 12, 12, NA, 12, 8,
12, ~
## $ Total.School.Enrollment        <dbl> 927, 850, 955, NA, 720, 939, 225,
NA, 5~
## $ Income.Level                   <fct> Q, A, O, NA, C, I, G, NA, K, K, O,
L, Q~
## $ EZ.Pay.Take.Up.Rate            <dbl> 0.170, 0.091, 0.042, 0.000, 0.383,
0.10~
## $ School.Sponsor                 <fct> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
0, ~
## $ SPR.Product.Type               <fct> CA History, East Coast, East Coast,
Eas~
## $ SPR.New.Existing               <fct> EXISTING, EXISTING, EXISTING,
EXISTING,~
## $ FPP                            <dbl> 59, 22, 24, 18, 81, 10, 25, 13, 52,
66,~
## $ Total.Pax                      <dbl> 63, 25, 27, 18, 89, 11, 27, 14, 56,
72,~
## $ SPR.Group.Revenue              <dbl> 424, 2350, 1181, 376, 865, 2025,
1977, ~
## $ NumberOfMeetingswithParents    <dbl> 1, 2, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1,
1, ~
## $ FirstMeeting                   <chr> "40402", "40134", "40434", "NA",
"40414~
## $ LastMeeting                    <chr> "40402", "40417", "40434", "NA",
"40414~
## $ DifferenceTraveltoFirstMeeting <dbl> 155, 423, 124, NA, 145, 91, 63,
138, 14~
## $ DifferenceTraveltoLastMeeting  <dbl> 155, 140, 124, NA, 145, 91, 63,
138, 14~
## $ SchoolGradeTypeLow             <fct> Elementary, Middle, Middle, High,
Middl~
## $ SchoolGradeTypeHigh            <fct> Elementary, Middle, Middle, High,
Middl~
## $ SchoolGradeType                <fct> Elementary->Elementary, Middle-
>Middle,~
## $ DepartureMonth                 <fct> January, January, January, January,
Jan~
## $ GroupGradeTypeLow              <fct> K, Middle, Middle, Undefined,
Middle, H~
## $ GroupGradeTypeHigh             <fct> Elementary, Middle, Middle,
Undefined, ~
## $ GroupGradeType                 <fct> K->Elementary, Middle->Middle,
Middle->~
## $ MajorProgramCode               <fct> H, H, H, H, H, H, S, I, C, H, C, H,
H, ~
## $ SingleGradeTripFlag            <fct> 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1,
1, ~
## $ FPP.to.School.enrollment       <dbl> 0.06364617, 0.02588235, 0.02513089,
NA,~
## $ FPP.to.PAX                     <dbl> 0.9365079, 0.8800000, 0.8888889,
1.0000~
## $ Num.of.Non_FPP.PAX             <dbl> 4, 3, 3, 0, 8, 1, 2, 1, 4, 6, 4, 5,
1, ~
## $ SchoolSizeIndicator            <fct> L, L, L, NA, M-L, L, S, NA, S-M, M-
L, M~
## $ Retained.in.2012.              <fct> 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1,
1, ~

#Relabeling
levels(dataset$Retained.in.2012.) <- c("Not Retained","Retained")

colnames(dataset)

##  [1] "ID"                             "Program.Code"
##  [3] "From.Grade"                     "To.Grade"
##  [5] "Group.State"                    "Is.Non.Annual."
##  [7] "Days"                           "Travel.Type"
##  [9] "Departure.Date"                 "Return.Date"
## [11] "Deposit.Date"                   "Special.Pay"
## [13] "Tuition"                        "FRP.Active"
## [15] "FRP.Cancelled"                  "FRP.Take.up.percent."
## [17] "Early.RPL"                      "Latest.RPL"
## [19] "Cancelled.Pax"                  "Total.Discount.Pax"
## [21] "Initial.System.Date"            "Poverty.Code"
## [23] "Region"                         "CRM.Segment"
## [25] "School.Type"                    "Parent.Meeting.Flag"
## [27] "MDR.Low.Grade"                  "MDR.High.Grade"
## [29] "Total.School.Enrollment"        "Income.Level"
## [31] "EZ.Pay.Take.Up.Rate"            "School.Sponsor"
## [33] "SPR.Product.Type"               "SPR.New.Existing"
## [35] "FPP"                            "Total.Pax"
## [37] "SPR.Group.Revenue"              "NumberOfMeetingswithParents"
## [39] "FirstMeeting"                   "LastMeeting"
## [41] "DifferenceTraveltoFirstMeeting" "DifferenceTraveltoLastMeeting"
## [43] "SchoolGradeTypeLow"             "SchoolGradeTypeHigh"
## [45] "SchoolGradeType"                "DepartureMonth"
## [47] "GroupGradeTypeLow"              "GroupGradeTypeHigh"
## [49] "GroupGradeType"                 "MajorProgramCode"
## [51] "SingleGradeTripFlag"            "FPP.to.School.enrollment"
## [53] "FPP.to.PAX"                     "Num.of.Non_FPP.PAX"
## [55] "SchoolSizeIndicator"            "Retained.in.2012."
```

```
levels(dataset$Retained.in.2012.)
```

```
## [1] "Not Retained" "Retained"
```

```
#Performing the Basic NA
#This indicates that there are 1081 NA values in the dataset

describe(dataset)
```

```
## Warning in all.is.numeric(names(weights), "vector"): NAs introduced by
coercion

## Warning in all.is.numeric(names(weights), "vector"): NAs introduced by
coercion

## Warning in all.is.numeric(names(weights), "vector"): NAs introduced by
coercion

## Warning in all.is.numeric(names(weights), "vector"): NAs introduced by
coercion
```

```
## dataset
##
##  56  Variables      2389  Observations
## --------------------------------------------------------------------------
------
## ID
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2389        0     2389        1     1195    796.7    120.4    239.8
##      .25      .50      .75      .90      .95
##    598.0   1195.0   1792.0   2150.2   2269.6
##
## lowest :    1    2    3    4    5, highest: 2385 2386 2387 2388 2389
## --------------------------------------------------------------------------
------
## Program.Code
##        n  missing distinct
##     2389        0       28
##
## lowest : CC  CD  CN  CVP FN , highest: SD  SG  SK  SM  ST
## --------------------------------------------------------------------------
------
## From.Grade
##        n  missing distinct
##     2389        0       11
##
## lowest : 10 11 12 3  4 , highest: 6  7  8  9  NA
##
```

```
## Value         10    11    12     3     4     5     6     7     8     9
NA
## Frequency     24    32    10     5   160    94   226   515  1121    75
127
## Proportion 0.010 0.013 0.004 0.002 0.067 0.039 0.095 0.216 0.469 0.031
0.053
## --------------------------------------------------------------------------
------
## To.Grade
##        n  missing distinct
##     2389        0       11
##
## lowest : 10 11 12 3  4 , highest: 6  7  8  9  NA
##
## Value         10    11    12     3     4     5     6     7     8     9
NA
## Frequency     15    23   130     1   132    63    57    75  1646    97
150
## Proportion 0.006 0.010 0.054 0.000 0.055 0.026 0.024 0.031 0.689 0.041
0.063
## --------------------------------------------------------------------------
------
## Group.State
##        n  missing distinct
##     2389        0       54
##
## lowest : AB AK AL AR AZ, highest: VT WA WI WV WY
## --------------------------------------------------------------------------
------
## Is.Non.Annual.
##        n  missing distinct
##     2389        0        2
##
## Value          0     1
## Frequency   2021   368
## Proportion 0.846 0.154
## --------------------------------------------------------------------------
------
## Days
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2389        0       12    0.925    4.575     1.49        2        3
##      .25      .50      .75      .90      .95
##        4        5        5        6        7
##
## lowest :  1  2  3  4  5, highest:  8  9 10 11 12
##
## Value          1     2     3     4     5     6     7     8     9    10
11
## Frequency     77    84   269   621   907   264   111    34     6     5
10
```

```
## Proportion 0.032 0.035 0.113 0.260 0.380 0.111 0.046 0.014 0.003 0.002
0.004
##
## Value         12
## Frequency      1
## Proportion 0.000
## ---------------------------------------------------------------------------
------
## Travel.Type
##        n  missing distinct
##     2389        0        4
##
## Value           A     B     N     T
## Frequency   2014   367     2     6
## Proportion 0.843 0.154 0.001 0.003
## ---------------------------------------------------------------------------
------
## Departure.Date
##          n    missing   distinct       Info       Mean        Gmd
.05
##       2389          0        144          1 2011-05-08    3431599 2011-03-
12
##        .10        .25        .50        .75        .90        .95
## 2011-03-19 2011-04-09 2011-05-17 2011-06-07 2011-06-15 2011-06-20
##
## lowest : 2011-01-14 2011-01-15 2011-01-16 2011-01-17 2011-01-18
## highest: 2011-06-26 2011-06-27 2011-06-28 2011-06-29 2011-06-30
## ---------------------------------------------------------------------------
------
## Return.Date
##          n    missing   distinct       Info       Mean        Gmd
.05
##       2389          0        143          1 2011-05-11    3439247 2011-03-
15
##        .10        .25        .50        .75        .90        .95
## 2011-03-22 2011-04-12 2011-05-20 2011-06-10 2011-06-19 2011-06-25
##
## lowest : 2011-01-14 2011-01-17 2011-01-20 2011-01-21 2011-01-23
## highest: 2011-06-30 2011-07-01 2011-07-02 2011-07-03 2011-07-05
## ---------------------------------------------------------------------------
------
## Deposit.Date
##          n    missing   distinct       Info       Mean        Gmd
.05
##       2389          0        135      0.993 2010-10-25    2777421 2010-09-
30
##        .10        .25        .50        .75        .90        .95
## 2010-10-01 2010-10-15 2010-10-28 2010-11-05 2010-11-19 2010-12-10
##
## lowest : 2009-09-25 2009-11-15 2009-11-17 2010-01-07 2010-01-10
## highest: 2011-04-08 2011-04-15 2011-04-20 2011-06-01 2011-10-30
## ---------------------------------------------------------------------------
------
## Special.Pay
##        n  missing distinct
##     2387        2        4
##
## Value          CP    FR    NA    SA
## Frequency      70   293  1917   107
## Proportion 0.029 0.123 0.803 0.045
## ---------------------------------------------------------------------------
------
## Tuition
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2389        0     1230        1     1615    722.7    449.0    629.8
##      .25      .50      .75      .90      .95
##   1174.0   1700.0   2048.0   2329.0   2522.6
##
## lowest :   79  100  119  143  149, highest: 3628 3799 3884 4199 4200
## ---------------------------------------------------------------------------
------
## FRP.Active
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2389        0       93    0.999    16.87    16.05        0        2
##      .25      .50      .75      .90      .95
##        6       12       23       36       47
##
## lowest :   0   1   2   3   4, highest: 124 139 149 160 257
## ---------------------------------------------------------------------------
------
## FRP.Cancelled
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2389        0       29     0.98    3.306    3.543        0        0
##      .25      .50      .75      .90      .95
##        1        2        4        8       10
##
## lowest :  0  1  2  3  4, highest: 27 28 30 32 45
## ---------------------------------------------------------------------------
------
## FRP.Take.up.percent.
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2389        0      476        1   0.5707   0.2543    0.000    0.250
##      .25      .50      .75      .90      .95
##    0.455    0.600    0.727    0.833    0.898
##
## lowest : 0.000 0.013 0.020 0.028 0.029, highest: 0.947 0.952 0.960 0.966
1.000
## ---------------------------------------------------------------------------
------
## Early.RPL
##        n  missing distinct 
##     2389        0      142 
## 
## lowest : 39920 39923 39934 39939 39961, highest: 40452 40459 40465 40490
NA
## --------------------------------------------------------------------------
------
## Latest.RPL 
##        n  missing distinct 
##     2389        0      216 
## 
## lowest : 39979 40045 40050 40065 40066, highest: 40595 40599 40606 40609
NA
## --------------------------------------------------------------------------
------
## Cancelled.Pax 
##        n  missing distinct     Info     Mean      Gmd      .05      .10 
##     2389        0       34     0.99    4.807    4.593        0        0 
##      .25      .50      .75      .90      .95 
##        2        4        6       10       14 
## 
## lowest :  0  1  2  3  4, highest: 33 34 37 38 39
## --------------------------------------------------------------------------
------
## Total.Discount.Pax 
##        n  missing distinct     Info     Mean      Gmd      .05      .10 
##     2389        0       26    0.943    2.954    2.467        1        1 
##      .25      .50      .75      .90      .95 
##        1        2        4        6        8 
## 
## lowest :  0  1  2  3  4, highest: 22 26 27 29 47
## --------------------------------------------------------------------------
------
## Initial.System.Date 
##        n  missing distinct 
##     2389        0      297 
## 
## lowest : 39905 39920 39933 39939 39961, highest: 40599 40600 40606 40607
NA
## --------------------------------------------------------------------------
------
## Poverty.Code 
##        n  missing distinct 
##     1790      599        6 
## 
## lowest : 0 A B C D, highest: A B C D E
## 
## Value          0     A     B     C     D     E
## Frequency      4   265   961   507    36    17
## Proportion 0.002 0.148 0.537 0.283 0.020 0.009
## --------------------------------------------------------------------------
------
## Region
##        n  missing distinct
##     2389        0        6
##
## lowest : Dallas              Houston             Northern California Other
Pacific Northwest
## highest: Houston             Northern California Other
Pacific Northwest   Southern California
##
## Value                      Dallas             Houston Northern California
## Frequency                     163                 145                 275
## Proportion                  0.068               0.061               0.115
##
## Value                       Other   Pacific Northwest Southern California
## Frequency                    1165                 198                 443
## Proportion                  0.488               0.083               0.185
## --------------------------------------------------------------------------
------
## CRM.Segment
##        n  missing distinct
##     2389        0       12
##
## lowest : 1  10 11 2  3 , highest: 6  7  8  9  NA
##
## Value          1    10    11     2     3     4     5     6     7     8
9
## Frequency     77   914    13    47    11   228   788    94   111    93
9
## Proportion 0.032 0.383 0.005 0.020 0.005 0.095 0.330 0.039 0.046 0.039
0.004
##
## Value         NA
## Frequency      4
## Proportion 0.002
## --------------------------------------------------------------------------
------
## School.Type
##        n  missing distinct
##     2389        0        4
##
## Value                      Catholic                   CHD Private non-
Christian
## Frequency                       163                   257
151
## Proportion                    0.068                 0.108
0.063
##
## Value                        PUBLIC
## Frequency                   1818
## Proportion                 0.761
## --------------------------------------------------------------------------
------
## Parent.Meeting.Flag
##        n  missing distinct
##     2389        0        2
##
## Value          0     1
## Frequency    337  2052
## Proportion 0.141 0.859
## --------------------------------------------------------------------------
------
## MDR.Low.Grade
##        n  missing distinct
##     2321       68       12
##
## lowest : 1  10 2  3  4 , highest: 7  8  9  K  PK
##
## Value          1    10     2     3     4     5     6     7     8     9
K
## Frequency      8     3     2    12    17    96   888   348    14   104
428
## Proportion 0.003 0.001 0.001 0.005 0.007 0.041 0.383 0.150 0.006 0.045
0.184
##
## Value         PK
## Frequency    401
## Proportion 0.173
## --------------------------------------------------------------------------
------
## MDR.High.Grade
##        n  missing distinct
##     2389        0       13
##
## lowest : 1  10 11 12 2 , highest: 6  7  8  9  NA
##
## Value          1    10    11    12     2     3     4     5     6     7
8
## Frequency      2     3     3   358     1     1     6    99   110    25
1659
## Proportion 0.001 0.001 0.001 0.150 0.000 0.000 0.003 0.041 0.046 0.010
0.694
##
## Value          9    NA
## Frequency     54    68
## Proportion 0.023 0.028
## --------------------------------------------------------------------------
------
## Total.School.Enrollment
##        n  missing distinct     Info     Mean      Gmd      .05      .10 
##     2298       91      893        1    648.4    415.8    165.0    220.0 
##      .25      .50      .75      .90      .95 
##    360.0    597.0    825.8   1082.3   1300.0 
## 
## lowest :   19   36   50   52   56, highest: 3100 3200 3600 3700 3990
## ---------------------------------------------------------------------------
------
## Income.Level 
##        n  missing distinct 
##     2327       62       22 
## 
## lowest : A  B  C  D  E , highest: P3 P4 P5 Q  Z 
## ---------------------------------------------------------------------------
------
## EZ.Pay.Take.Up.Rate 
##        n  missing distinct     Info     Mean      Gmd      .05      .10 
##     2389        0      371    0.997   0.2079   0.1682   0.0000   0.0000 
##      .25      .50      .75      .90      .95 
##   0.1000   0.2000   0.2920   0.4000   0.4826 
## 
## lowest : 0.000 0.008 0.011 0.012 0.016, highest: 0.786 0.800 1.000 1.205
1.750
## ---------------------------------------------------------------------------
------
## School.Sponsor 
##        n  missing distinct 
##     2389        0        2 
## 
## Value          0     1
## Frequency   2136   253
## Proportion 0.894 0.106
## ---------------------------------------------------------------------------
------
## SPR.Product.Type 
##        n  missing distinct 
##     2389        0        6 
## 
## lowest : CA History    Costa Rica    East Coast    IL History   
International
## highest: Costa Rica    East Coast    IL History    International Science      
## 
## Value         CA History    Costa Rica    East Coast    IL History
## Frequency            175            46          2005             5
## Proportion         0.073         0.019         0.839         0.002
## 
## Value      International       Science
## Frequency             15           143
## Proportion         0.006         0.060
## ---------------------------------------------------------------------------
```

```
------
## SPR.New.Existing 
##        n  missing distinct 
##     2389        0        2 
##                             
## Value      EXISTING      NEW
## Frequency      1607      782
## Proportion    0.673    0.327
## --------------------------------------------------------------------------
------
## FPP 
##        n  missing distinct     Info     Mean      Gmd      .05      .10 
##     2389        0      146        1     31.3    27.35      5.0      7.0 
##      .25      .50      .75      .90      .95 
##     12.0     23.0     41.0     65.0     82.6 
## 
## lowest :   2   3   4   5   6, highest: 222 230 243 257 286
## --------------------------------------------------------------------------
------
## Total.Pax 
##        n  missing distinct     Info     Mean      Gmd      .05      .10 
##     2389        0      159        1    34.25    29.53        6        8 
##      .25      .50      .75      .90      .95 
##       14       26       44       70       89 
## 
## lowest :   2   3   4   5   6, highest: 250 251 262 276 313
## --------------------------------------------------------------------------
------
## SPR.Group.Revenue 
##        n  missing distinct     Info     Mean      Gmd      .05      .10 
##     2389        0     1230        1     1615    722.7    449.0    629.8 
##      .25      .50      .75      .90      .95 
##   1174.0   1700.0   2048.0   2329.0   2522.6 
## 
## lowest :   79  100  119  143  149, highest: 3628 3799 3884 4199 4200
## --------------------------------------------------------------------------
------
## NumberOfMeetingswithParents 
##        n  missing distinct     Info     Mean      Gmd 
##     2389        0        3    0.749    1.102   0.6107 
##                             
## Value          0     1     2
## Frequency    337  1471   581
## Proportion 0.141 0.616 0.243
## --------------------------------------------------------------------------
------
## FirstMeeting 
##        n  missing distinct 
##     2389        0      208 
## 
```

```
## lowest : 39945 40057 40084 40085 40108, highest: 40682 40683 40800 40821
NA
## --------------------------------------------------------------------------
------
## LastMeeting
##        n  missing distinct
##     2389        0      173
##
## lowest : 39945 40057 40184 40213 40233, highest: 40689 40708 40800 40821
NA
## --------------------------------------------------------------------------
------
## DifferenceTraveltoFirstMeeting
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2052      337      342        1    262.1    85.21    165.0    182.0
##      .25      .50      .75      .90      .95
##    208.0    250.0    287.0    386.0    411.4
##
## lowest : -204 -188   14   22   25, highest:  598  604  623  651  749
## --------------------------------------------------------------------------
------
## DifferenceTraveltoLastMeeting
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2052      337      251        1      229    54.06    154.6    173.0
##      .25      .50      .75      .90      .95
##    196.8    233.0    261.0    275.0    285.0
##
## lowest : -204 -188  -17   -4    9, highest:  455  456  530  651  749
## --------------------------------------------------------------------------
------
## SchoolGradeTypeLow
##        n  missing distinct
##     2389        0        4
##
## Value      Elementary       High     Middle  Undefined
## Frequency         259        141       1862        127
## Proportion      0.108      0.059      0.779      0.053
## --------------------------------------------------------------------------
------
## SchoolGradeTypeHigh
##        n  missing distinct
##     2389        0        4
##
## Value      Elementary       High     Middle  Undefined
## Frequency         196        265       1778        150
## Proportion      0.082      0.111      0.744      0.063
## --------------------------------------------------------------------------
------
## SchoolGradeType
##        n  missing distinct
##     2389        0        9
##
## lowest : Elementary->Elementary Elementary->High       Elementary->Middle
Elementary->Undefined  High->High
## highest: High->High             Middle->High           Middle->Middle
Middle->Undefined      Undefined->Undefined
##
## Elementary->Elementary (196, 0.082), Elementary->High (4, 0.002),
## Elementary->Middle (57, 0.024), Elementary->Undefined (2, 0.001), High-
>High
## (141, 0.059), Middle->High (120, 0.050), Middle->Middle (1721, 0.720),
## Middle->Undefined (21, 0.009), Undefined->Undefined (127, 0.053)
## --------------------------------------------------------------------------
------
## DepartureMonth
##        n  missing distinct
##     2389        0        6
##
## lowest : April    February January  June     March
## highest: February January  June     March    May
##
## Value          April February  January     June    March      May
## Frequency        534       49        9      903      387      507
## Proportion     0.224    0.021    0.004    0.378    0.162    0.212
## --------------------------------------------------------------------------
------
## GroupGradeTypeLow
##        n  missing distinct
##     2389        0        6
##
## lowest : Elementary High       K          Middle     PK
## highest: High       K          Middle     PK         Undefined
##
## Value      Elementary       High          K     Middle         PK
Undefined
## Frequency         135        107        428       1250        401
68
## Proportion      0.057      0.045      0.179      0.523      0.168
0.028
## --------------------------------------------------------------------------
------
## GroupGradeTypeHigh
##        n  missing distinct
##     2389        0        4
##
## Value      Elementary       High     Middle  Undefined
## Frequency         109        418       1794         68
## Proportion      0.046      0.175      0.751      0.028
## --------------------------------------------------------------------------
------
```

```
## GroupGradeType
##        n  missing distinct
##     2389        0       13
##
## lowest : Elementary->Elementary Elementary->High       Elementary->Middle
High->High             K->Elementary
## highest: Middle->Middle         PK->Elementary         PK->High
PK->Middle             Undefined->Undefined
## ---------------------------------------------------------------------------
------
## MajorProgramCode
##        n  missing distinct
##     2389        0        4
##
## Value           C     H     I     S
## Frequency    135  2049    16   189
## Proportion 0.057 0.858 0.007 0.079
## ---------------------------------------------------------------------------
------
## SingleGradeTripFlag
##        n  missing distinct
##     2389        0        2
##
## Value           0     1
## Frequency   1059  1330
## Proportion 0.443 0.557
## ---------------------------------------------------------------------------
------
## FPP.to.School.enrollment
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2298       91     1909        1  0.06618  0.06494 0.006667 0.009982
##      .25      .50      .75      .90      .95
## 0.020787 0.045256 0.087517 0.142105 0.192817
##
## lowest : 0.000922084 0.001196172 0.001383126 0.001385681 0.001437470
## highest: 0.470588235 0.480769231 0.500000000 1.440000000 2.052631579
## ---------------------------------------------------------------------------
------
## FPP.to.PAX
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2389        0      306    0.999   0.9007  0.05018   0.8000   0.8333
##      .25      .50      .75      .90      .95
##   0.8824   0.9091   0.9333   0.9474   0.9592
##
## lowest : 0.6000000 0.6250000 0.6666667 0.6923077 0.7058824
## highest: 0.9777778 0.9782609 0.9807692 0.9848485 1.0000000
## ---------------------------------------------------------------------------
------
## Num.of.Non_FPP.PAX
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2389        0       26    0.943    2.954    2.467        1        1
##      .25      .50      .75      .90      .95
##        1        2        4        6        8
##
## lowest :  0  1  2  3  4, highest: 22 26 27 29 47
## --------------------------------------------------------------------------
------
## SchoolSizeIndicator
##        n  missing distinct
##     2298       91        4
##
## Value          L   M-L     S   S-M
## Frequency    597   594   507   600
## Proportion 0.260 0.258 0.221 0.261
## --------------------------------------------------------------------------
------
## Retained.in.2012.
##        n  missing distinct
##     2389        0        2
##
## Value      Not Retained     Retained
## Frequency           938         1451
## Proportion        0.393        0.607
## --------------------------------------------------------------------------
------
```

```
sum(is.na(dataset[,]))
```

```
## [1] 1678
```

```
skim(dataset)
```

*Data summary*

| | |
|---|---|
| Name | dataset |
| Number of rows | 2389 |
| Number of columns | 56 |
| Column type frequency: | |
| character | 5 |
| factor | 29 |
| numeric | 19 |
| POSIXct | 3 |
| Group variables | None |

**Variable type: character**

| skim_variable | n_missing | complete_rate | min | max | empty | n_unique | whitespace |
|---|---:|---:|---:|---:|---:|---:|---:|
| Early.RPL | 0 | 1 | 2 | 5 | 0 | 142 | 0 |
| Latest.RPL | 0 | 1 | 2 | 5 | 0 | 216 | 0 |
| Initial.System.Date | 0 | 1 | 2 | 5 | 0 | 297 | 0 |
| FirstMeeting | 0 | 1 | 2 | 5 | 0 | 208 | 0 |
| LastMeeting | 0 | 1 | 2 | 5 | 0 | 173 | 0 |

**Variable type: factor**

| skim_variable | n_missing | complete_rate | ordered | n_unique | top_counts |
|---|---:|---:|---|---:|---|
| Program.Code | 0 | 1.00 | FALSE | 28 | HD: 1430, HC: 274, HS: 131, CD: 114 |
| From.Grade | 0 | 1.00 | FALSE | 11 | 8: 1121, 7: 515, 6: 226, 4: 160 |
| To.Grade | 0 | 1.00 | FALSE | 11 | 8: 1646, NA: 150, 4: 132, 12: 130 |
| Group.State | 0 | 1.00 | FALSE | 54 | CA: 718, TX: 308, WA: 147, IL: 104 |
| Is.Non.Annual. | 0 | 1.00 | FALSE | 2 | 0: 2021, 1: 368 |
| Travel.Type | 0 | 1.00 | FALSE | 4 | A: 2014, B: 367, T: 6, N: 2 |
| Special.Pay | 2 | 1.00 | FALSE | 4 | NA: 1917, FR: 293, SA: 107, CP: 70 |
| Poverty.Code | 599 | 0.75 | FALSE | 6 | B: 961, C: 507, A: 265, D: 36 |
| Region | 0 | 1.00 | FALSE | 6 | Oth: 1165, Sou: 443, Nor: 275, Pac: 198 |
| CRM.Segment | 0 | 1.00 | FALSE | 12 | 10: 914, 5: 788, 4: 228, 7: 111 |
| School.Type | 0 | 1.00 | FALSE | 4 | PUB: 1818, CHD: 257, Cat: 163, Pri: 151 |
| Parent.Meeting.Flag | 0 | 1.00 | FALSE | 2 | 1: 2052, 0: 337 |
| MDR.Low.Grade | 68 | 0.97 | FALSE | 12 | 6: 888, K: 428, PK: 401, 7: 348 |
| MDR.High.Grade | 0 | 1.00 | FALSE | 13 | 8: 1659, 12: 358, |

| skim_variable | n_missing | complete_rate | ordered | n_unique | top_counts |
|---|---|---|---|---|---|
| | | | | | 6: 110, 5: 99 |
| Income.Level | 62 | 0.97 | FALSE | 22 | Q: 283, O: 266, L: 214, P: 212 |
| School.Sponsor | 0 | 1.00 | FALSE | 2 | 0: 2136, 1: 253 |
| SPR.Product.Type | 0 | 1.00 | FALSE | 6 | Eas: 2005, CA : 175, Sci: 143, Cos: 46 |
| SPR.New.Existing | 0 | 1.00 | FALSE | 2 | EXI: 1607, NEW: 782 |
| SchoolGradeTypeLow | 0 | 1.00 | FALSE | 4 | Mid: 1862, Ele: 259, Hig: 141, Und: 127 |
| SchoolGradeTypeHigh | 0 | 1.00 | FALSE | 4 | Mid: 1778, Hig: 265, Ele: 196, Und: 150 |
| SchoolGradeType | 0 | 1.00 | FALSE | 9 | Mid: 1721, Ele: 196, Hig: 141, Und: 127 |
| DepartureMonth | 0 | 1.00 | FALSE | 6 | Jun: 903, Apr: 534, May: 507, Mar: 387 |
| GroupGradeTypeLow | 0 | 1.00 | FALSE | 6 | Mid: 1250, K: 428, PK: 401, Ele: 135 |
| GroupGradeTypeHigh | 0 | 1.00 | FALSE | 4 | Mid: 1794, Hig: 418, Ele: 109, Und: 68 |
| GroupGradeType | 0 | 1.00 | FALSE | 13 | Mid: 1103, K->: 305, PK-: 266, Mid: 147 |
| MajorProgramCode | 0 | 1.00 | FALSE | 4 | H: 2049, S: 189, C: 135, I: 16 |
| SingleGradeTripFlag | 0 | 1.00 | FALSE | 2 | 1: 1330, 0: 1059 |
| SchoolSizeIndicator | 91 | 0.96 | FALSE | 4 | S-M: 600, L: 597, M-L: 594, S: 507 |
| Retained.in.2012. | 0 | 1.00 | FALSE | 2 | Ret: 1451, Not: 938 |

**Variable type: numeric**

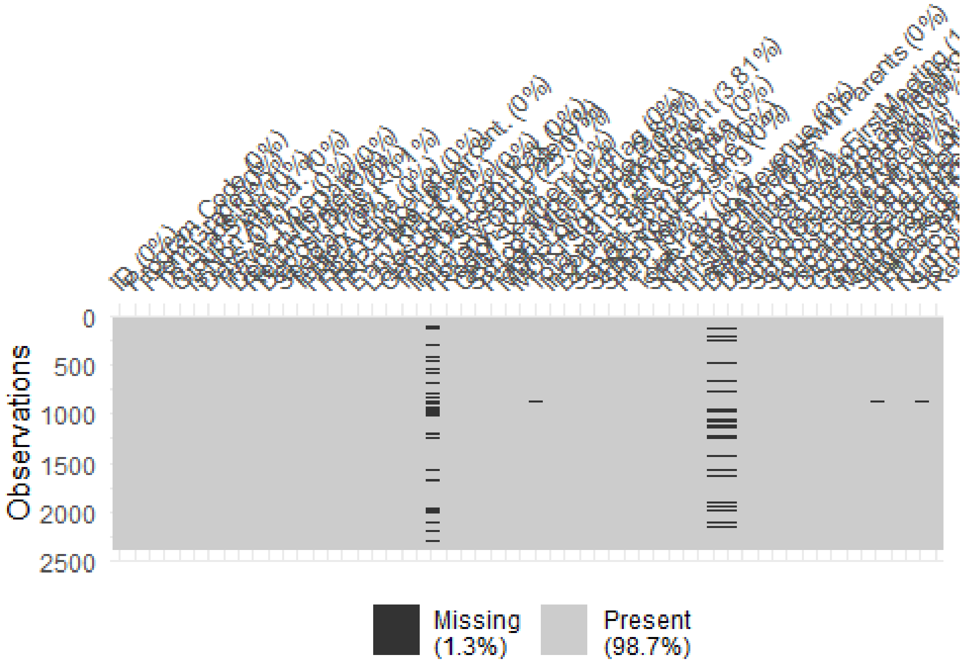
| skim_variable | n_missing | complete_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
|---|---|---|---|---|---|---|---|---|---|---|
| ID | 0 | 1.00 | 1195.00 | 689.79 | 1.0 | 598.00 | 1195.00 | 1792.00 | 2389.00 | ▇▇▇▇▇ |
| Days | 0 | 1.00 | 4.58 | 1.43 | 1.0 | 4.00 | 5.00 | 5.00 | 12.00 | ▂▇▂▁▁ |
| Tuition | 0 | 1.00 | 1615.22 | 645.10 | 79.0 | 1174.00 | 1700.00 | 2048.00 | 4200.00 | ▂▅▇▁▁ |
| FRP.Active | 0 | 1.00 | 16.87 | 16.94 | 0.0 | 6.00 | 12.00 | 23.00 | 257.00 | ▇▁▁▁▁ |
| FRP.Cancelled | 0 | 1.00 | 3.31 | 3.68 | 0.0 | 1.00 | 2.00 | 4.00 | 45.00 | ▇▁▁▁▁ |
| FRP.Take.up.percent. | 0 | 1.00 | 0.57 | 0.23 | 0.0 | 0.46 | 0.60 | 0.73 | 1.00 | ▂▂▇▇▂ |
| Cancelled.Pax | 0 | 1.00 | 4.81 | 4.66 | 0.0 | 2.00 | 4.00 | 6.00 | 39.00 | ▇▂▁▁▁ |
| Total.Discount.Pax | 0 | 1.00 | 2.95 | 2.88 | 0.0 | 1.00 | 2.00 | 4.00 | 47.00 | ▇▁▁▁▁ |
| Total.School.Enrollment | 91 | 0.96 | 648.36 | 411.73 | 19.0 | 360.00 | 597.00 | 825.75 | 3990.00 | ▇▂▁▁▁ |
| EZ.Pay.Take.Up.Rate | 0 | 1.00 | 0.21 | 0.16 | 0.0 | 0.10 | 0.20 | 0.29 | 1.75 | ▇▂▁▁▁ |
| FPP | 0 | 1.00 | 31.30 | 29.13 | 2.0 | 12.00 | 23.00 | 41.00 | 286.00 | ▇▁▁▁▁ |
| Total.Pax | 0 | 1.00 | 34.25 | 31.59 | 2.0 | 14.00 | 26.00 | 44.00 | 313.00 | ▇▁▁▁▁ |
| SPR.Group.Revenue | 0 | 1.00 | 1615.22 | 645.10 | 79.0 | 1174.00 | 1700.00 | 2048.00 | 4200.00 | ▂▅▇▁▁ |
| NumberOfMeetingswithParents | 0 | 1.00 | 1.10 | 0.61 | 0.0 | 1.00 | 1.00 | 1.00 | 2.00 | ▂▁▇▁▃ |
| DifferenceTraveltoFirstMeeting | 337 | 0.86 | 262.08 | 79.52 | -204.0 | 208.00 | 250.00 | 287.00 | 749.00 | ▁▁▇▂▁ |
| DifferenceTraveltoLastMeeting | 337 | 0.86 | 228.98 | 53.64 | -204.0 | 196.75 | 233.00 | 261.00 | 749.00 | ▁▁▇▁▁ |
| FPP.to.School.enrollment | 91 | 0.96 | 0.07 | 0.08 | 0.0 | 0.02 | 0.05 | 0.09 | 2.05 | ▇▁▁▁▁ |
| FPP.to.PAX | 0 | 1.00 | 0.90 | 0.0 | 0.6 | 0.88 | 0.91 | 0.93 | 1.00 | ▁▁ |

| skim_variable | n_missing | complete_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
|---|---|---|---|---|---|---|---|---|---|---|
| Num.of.Non_FPP.PAX | 0 | 1.00 | 2.95 | 5<br>2.8<br>8 | 0.0 | 1.00 | 2.00 | 4.00 | 47.00 | ▁▇<br>▇▁▁<br>▁▁ |

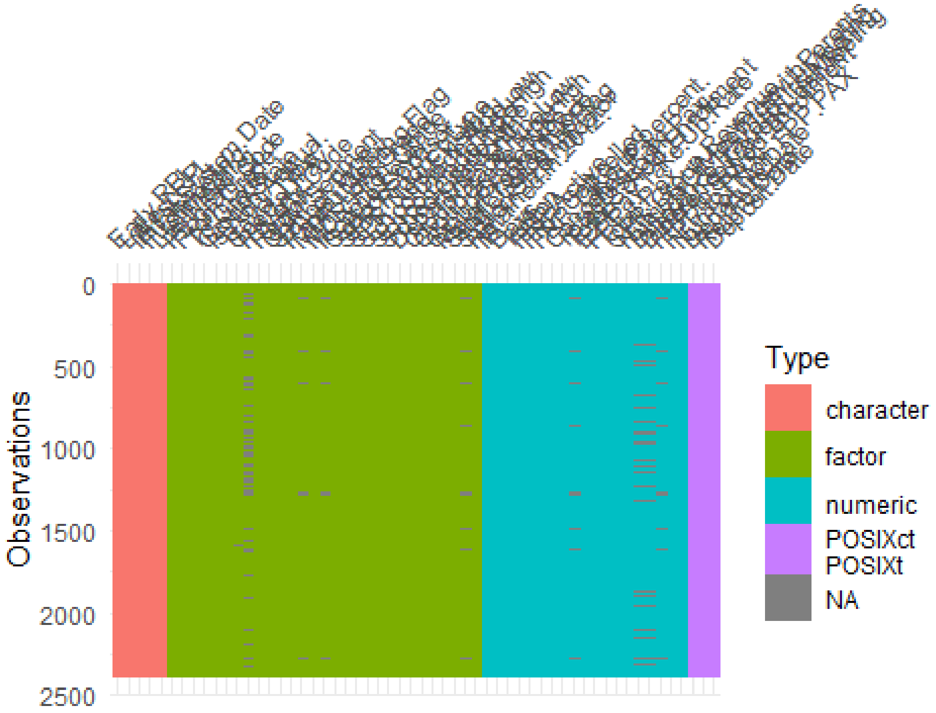
**Variable type: POSIXct**

| skim_variable | n_missing | complete_rate | min | max | median | n_unique |
|---|---|---|---|---|---|---|
| Departure.Date | 0 | 1 | 2011-01-14 | 2011-06-30 | 2011-05-17 | 144 |
| Return.Date | 0 | 1 | 2011-01-14 | 2011-07-05 | 2011-05-20 | 143 |
| Deposit.Date | 0 | 1 | 2009-09-25 | 2011-10-30 | 2010-10-28 | 135 |

```
vis_miss(dataset)
```



```
vis_dat(dataset)
```



```
#As we can see there are many NA values present in:
#1.MDR.LowGrade
#2.Poverty_Code
#3.Income.Level
#4.SchoolSizeIndicator
#5.TotalSchoolEnrollment




#Selecting all the numerical variables and the categorical variables
#Replacing the NA with the mean in numerical variables
numerical_columns <- c(7,13,14,15,16,19,20,29,31,35,36,37,38,41,42,52,53,54)
length(numerical_columns)

## [1] 18

for(i in numerical_columns)
  dataset[,i][is.na(dataset[,i])] <- mean(dataset[,i],na.rm = T)

sum(is.na(dataset[numerical_columns]))

## [1] 0

#Replacing the NA Values for factor columns

#Finding the mode of the factor columns using a function
getmode <- function(v){
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v,uniqv)))]
}

glimpse(dataset)
```

```
## Rows: 2,389
## Columns: 56
## $ ID                             <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 
12, ~
## $ Program.Code                   <fct> HS, HC, HD, HN, HD, HC, SG, FN, CC, 
HD,~
## $ From.Grade                     <fct> 4, 8, 8, 9, 6, 10, 11, 9, 8, 8, 8, 
8, 8~
## $ To.Grade                       <fct> 4, 8, 8, 12, 8, 12, 12, 9, 8, 8, 8, 
8, ~
## $ Group.State                    <fct> CA, AZ, FL, VA, FL, LA, MA, MX, AZ, 
TX,~
## $ Is.Non.Annual.                 <fct> 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 
0, ~
## $ Days                           <dbl> 1, 7, 3, 3, 6, 4, 6, 8, 8, 4, 4, 4, 
6, ~
## $ Travel.Type                    <fct> A, A, A, B, T, A, A, A, A, A, A, A, 
A, ~
## $ Departure.Date                 <dttm> 2011-01-14, 2011-01-14, 2011-01-
15, 20~
## $ Return.Date                    <dttm> 2011-01-14, 2011-01-21, 2011-01-
17, 20~
## $ Deposit.Date                   <dttm> 2010-08-30, 2009-11-15, 2010-10-
15, 20~
## $ Special.Pay                    <fct> NA, CP, NA, NA, NA, NA, NA, NA, CP, 
NA,~
## $ Tuition                        <dbl> 424, 2350, 1181, 376, 865, 2025, 
1977, ~
## $ FRP.Active                     <dbl> 25, 9, 17, 0, 40, 9, 16, 10, 30, 
51, 47~
## $ FRP.Cancelled                  <dbl> 3, 9, 6, 0, 8, 4, 4, 0, 0, 1, 1, 0, 
6, ~
## $ FRP.Take.up.percent.           <dbl> 0.424, 0.409, 0.708, 0.000, 0.494, 
0.90~
## $ Early.RPL                      <chr> "40266", "40106", "40297", "NA", 
"40266~
## $ Latest.RPL                     <chr> "40402", "40400", "40406", "NA", 
"40402~
## $ Cancelled.Pax                  <dbl> 3, 11, 6, 1, 9, 3, 5, 1, 0, 1, 1, 
0, 6,~
## $ Total.Discount.Pax             <dbl> 4, 3, 3, 0, 8, 1, 2, 1, 4, 6, 4, 5, 
1, ~
## $ Initial.System.Date            <chr> "40263", "40088", "40206", "40470", 
"40~
```

```
## $ Poverty.Code                   <fct> B, C, C, NA, D, C, NA, NA, NA, NA,
NA, ~
## $ Region                         <fct> Southern California, Other, Other,
Othe~
## $ CRM.Segment                    <fct> 4, 10, 10, 7, 10, 8, 8, 7, 5, 5,
10, 10~
## $ School.Type                    <fct> PUBLIC, PUBLIC, PUBLIC, CHD,
PUBLIC, PU~
## $ Parent.Meeting.Flag            <fct> 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1,
1, ~
## $ MDR.Low.Grade                  <fct> K, 7, 6, NA, 6, 10, 9, NA, 6, PK,
K, PK~
## $ MDR.High.Grade                 <fct> 5, 8, 8, NA, 8, 12, 12, NA, 12, 8,
12, ~
## $ Total.School.Enrollment        <dbl> 927.0000, 850.0000, 955.0000,
648.3586,~
## $ Income.Level                   <fct> Q, A, O, NA, C, I, G, NA, K, K, O,
L, Q~
## $ EZ.Pay.Take.Up.Rate            <dbl> 0.170, 0.091, 0.042, 0.000, 0.383,
0.10~
## $ School.Sponsor                 <fct> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
0, ~
## $ SPR.Product.Type               <fct> CA History, East Coast, East Coast,
Eas~
## $ SPR.New.Existing               <fct> EXISTING, EXISTING, EXISTING,
EXISTING,~
## $ FPP                            <dbl> 59, 22, 24, 18, 81, 10, 25, 13, 52,
66,~
## $ Total.Pax                      <dbl> 63, 25, 27, 18, 89, 11, 27, 14, 56,
72,~
## $ SPR.Group.Revenue              <dbl> 424, 2350, 1181, 376, 865, 2025,
1977, ~
## $ NumberOfMeetingswithParents    <dbl> 1, 2, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1,
1, ~
## $ FirstMeeting                   <chr> "40402", "40134", "40434", "NA",
"40414~
## $ LastMeeting                    <chr> "40402", "40417", "40434", "NA",
"40414~
## $ DifferenceTraveltoFirstMeeting <dbl> 155.0000, 423.0000, 124.0000,
262.0838,~
## $ DifferenceTraveltoLastMeeting  <dbl> 155.0000, 140.0000, 124.0000,
228.9781,~
## $ SchoolGradeTypeLow             <fct> Elementary, Middle, Middle, High,
Middl~
## $ SchoolGradeTypeHigh            <fct> Elementary, Middle, Middle, High,
Middl~
## $ SchoolGradeType                <fct> Elementary->Elementary, Middle-
>Middle,~
## $ DepartureMonth                 <fct> January, January, January, January,
Jan~
```

```
## $ GroupGradeTypeLow              <fct> K, Middle, Middle, Undefined,
Middle, H~
## $ GroupGradeTypeHigh             <fct> Elementary, Middle, Middle,
Undefined, ~
## $ GroupGradeType                 <fct> K->Elementary, Middle->Middle,
Middle->~
## $ MajorProgramCode               <fct> H, H, H, H, H, H, S, I, C, H, C, H,
H, ~
## $ SingleGradeTripFlag            <fct> 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1,
1, ~
## $ FPP.to.School.enrollment       <dbl> 0.06364617, 0.02588235, 0.02513089,
0.0~
## $ FPP.to.PAX                     <dbl> 0.9365079, 0.8800000, 0.8888889,
1.0000~
## $ Num.of.Non_FPP.PAX             <dbl> 4, 3, 3, 0, 8, 1, 2, 1, 4, 6, 4, 5,
1, ~
## $ SchoolSizeIndicator            <fct> L, L, L, NA, M-L, L, S, NA, S-M, M-
L, M~
## $ Retained.in.2012.              <fct> Retained, Retained, Retained, Not
Retai~

#So we are using the mode of the categorical variables for replacing the NA
Values in categorical values
#Replacing the NA in From.Grade

#From.Grade
dataset$From.Grade <- as.numeric(dataset$From.Grade)

dataset$From.Grade[is.na(dataset$From.Grade)] <- getmode(dataset$From.Grade)

dataset$From.Grade <- as.factor(dataset$From.Grade)

#To.Grade
dataset$To.Grade <- as.numeric(dataset$To.Grade)

dataset$To.Grade[is.na(dataset$To.Grade)] <- getmode(dataset$To.Grade)

dataset$To.Grade <- as.factor(dataset$To.Grade)



#Income level
dataset$Income.Level <- as.numeric(dataset$Income.Level)

dataset$Income.Level[is.na(dataset$Income.Level)] <-
getmode(dataset$Income.Level)

dataset$Income.Level <- as.factor(dataset$Income.Level)
```

```
#MDR.High.Grade
dataset$MDR.High.Grade <- as.numeric(dataset$MDR.High.Grade)

dataset$MDR.High.Grade[is.na(dataset$MDR.High.Grade)] <-
getmode(dataset$MDR.High.Grade)

dataset$MDR.High.Grade <- as.factor(dataset$MDR.High.Grade)

#MDR.Low.Grade
dataset$MDR.Low.Grade <- as.numeric(dataset$MDR.Low.Grade)

dataset$MDR.Low.Grade[is.na(dataset$MDR.Low.Grade)] <-
getmode(dataset$MDR.Low.Grade)

dataset$MDR.Low.Grade <- as.factor(dataset$MDR.Low.Grade)

#sum(is.na(dataset$MDR.Low.Grade))

#Poverty.Code
dataset$Poverty.Code <- as.numeric(dataset$Poverty.Code)

dataset$Poverty.Code[is.na(dataset$Poverty.Code)] <-
getmode(dataset$Poverty.Code)

dataset$Poverty.Code <- as.factor(dataset$Poverty.Code)

sum(is.na(dataset$Poverty.Code))
```

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

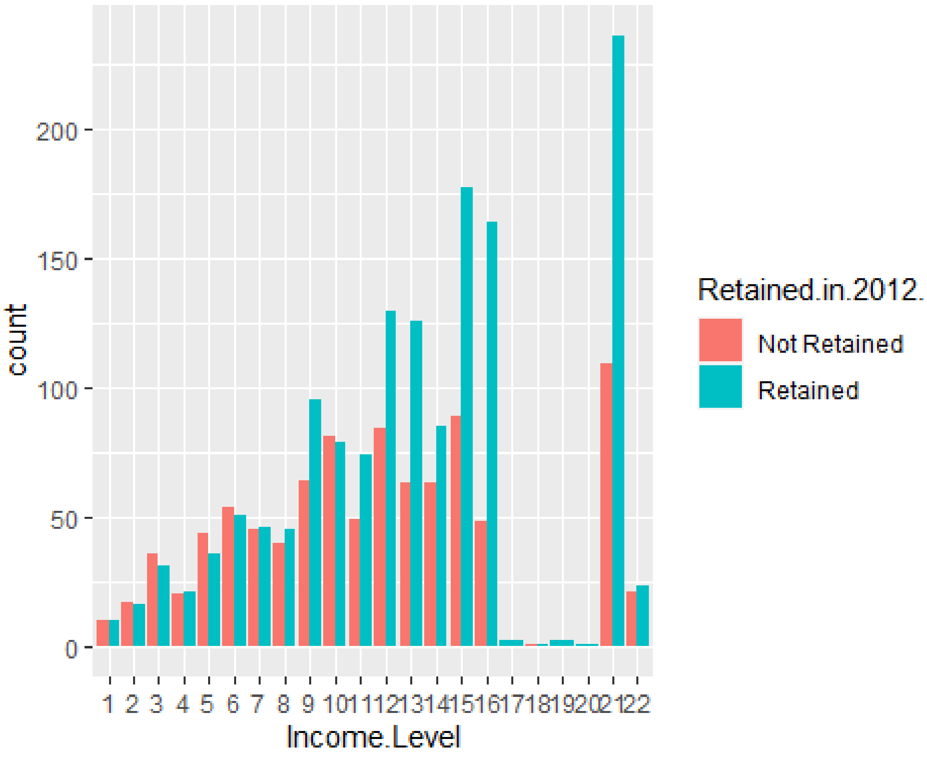
```
##SchoolGradeTypeLow
dataset$SchoolGradeTypeLow <- as.numeric(dataset$SchoolGradeTypeLow)
dataset$SchoolGradeTypeLow[is.na(dataset$SchoolGradeTypeLow)] <-
getmode(dataset$SchoolGradeTypeLow)
dataset$SchoolGradeTypeLow <- as.factor(dataset$SchoolGradeTypeLow)

#SchoolGradeTypeHigh
dataset$SchoolGradeTypeHigh <- as.numeric(dataset$SchoolGradeTypeHigh)
dataset$SchoolGradeTypeHigh[is.na(dataset$SchoolGradeTypeHigh)] <-
getmode(dataset$SchoolGradeTypeHigh)
dataset$SchoolGradeTypeHigh <- as.factor(dataset$SchoolGradeTypeHigh)

#SchoolGradeType
dataset$SchoolGradeType <- as.numeric(dataset$SchoolGradeType)
dataset$SchoolGradeType[is.na(dataset$SchoolGradeType)] <-
getmode(dataset$SchoolGradeType)
dataset$SchoolGradeType <- as.factor(dataset$SchoolGradeType)
```

```r
#DepartureMonth
dataset$DepartureMonth <- as.numeric(dataset$DepartureMonth)
dataset$DepartureMonth[is.na(dataset$DepartureMonth)] <-
getmode(dataset$DepartureMonth)
dataset$DepartureMonth <- as.factor(dataset$DepartureMonth)



#replace_labels(dataset$Special.Pay,labels=c("Not Applicable" =
tagged_na("NA")))

#GroupGradeTypeLow
dataset$GroupGradeTypeLow <- as.numeric(dataset$GroupGradeTypeLow)
dataset$GroupGradeTypeLow[is.na(dataset$GroupGradeTypeLow)] <-
getmode(dataset$GroupGradeTypeLow)
dataset$GroupGradeTypeLow <- as.factor(dataset$GroupGradeTypeLow)


dataset$GroupGradeTypeHigh <- as.numeric(dataset$GroupGradeTypeHigh)
dataset$GroupGradeTypeHigh[is.na(dataset$GroupGradeTypeHigh)] <-
getmode(dataset$GroupGradeTypeHigh)
dataset$GroupGradeTypeHigh <- as.factor(dataset$GroupGradeTypeHigh)

dataset$GroupGradeType <- as.numeric(dataset$GroupGradeType)
dataset$GroupGradeType[is.na(dataset$GroupGradeType)] <-
getmode(dataset$GroupGradeType)
dataset$GroupGradeType <- as.factor(dataset$GroupGradeType)

dataset$MajorProgramCode <- as.numeric(dataset$MajorProgramCode)
dataset$MajorProgramCode[is.na(dataset$MajorProgramCode)] <-
getmode(dataset$MajorProgramCode)
dataset$MajorProgramCode <- as.factor(dataset$MajorProgramCode)

dataset$SingleGradeTripFlag <- as.numeric(dataset$SingleGradeTripFlag)
dataset$SingleGradeTripFlag[is.na(dataset$SingleGradeTripFlag)] <-
getmode(dataset$SingleGradeTripFlag)
dataset$SingleGradeTripFlag <- as.factor(dataset$SingleGradeTripFlag)

dataset$SchoolSizeIndicator <- as.numeric(dataset$SchoolSizeIndicator)
dataset$SchoolSizeIndicator[is.na(dataset$SchoolSizeIndicator)] <-
getmode(dataset$SchoolSizeIndicator)
dataset$SchoolSizeIndicator <- as.factor(dataset$SchoolSizeIndicator)

#replace_labels(dataset,)

dataset$Departure.Date <- as.Date(dataset$Departure.Date)

dataset$Return.Date <- as.Date(dataset$Return.Date)
```

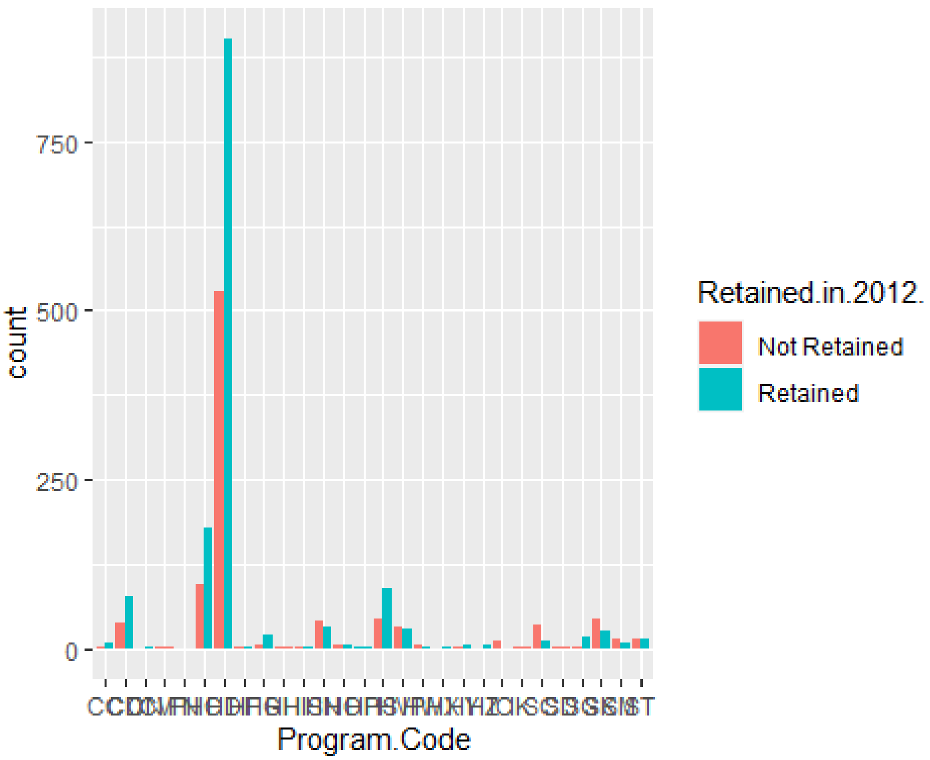
```
dataset$Deposit.Date <- as.Date(dataset$Deposit.Date)

#Changing the date columns to date
dataset$Initial.System.Date <- as.numeric(dataset$Initial.System.Date )

## Warning: NAs introduced by coercion

dataset$Initial.System.Date <- as.Date(dataset$Initial.System.Date, origin =
"1899-12-30")

dataset$Latest.RPL <- as.numeric(dataset$Latest.RPL)

## Warning: NAs introduced by coercion

dataset$Latest.RPL <- as.Date(dataset$Latest.RPL, origin = "1899-12-30")

dataset$FirstMeeting <- as.numeric(dataset$FirstMeeting)

## Warning: NAs introduced by coercion

dataset$FirstMeeting <- as.Date(dataset$FirstMeeting, origin = "1899-12-30")

dataset$LastMeeting <- as.numeric(dataset$LastMeeting)

## Warning: NAs introduced by coercion

dataset$LastMeeting <- as.Date(dataset$LastMeeting, origin = "1899-12-30")

#ncol(dataset_WithNONA)

ncol(dataset)

## [1] 56

#actual_data <- dataset_WithNONA

#We will not consider the date columns for any models construction, since we
have another Departure Month column that
#gives the information regarding that.
```

###Graphical Representation

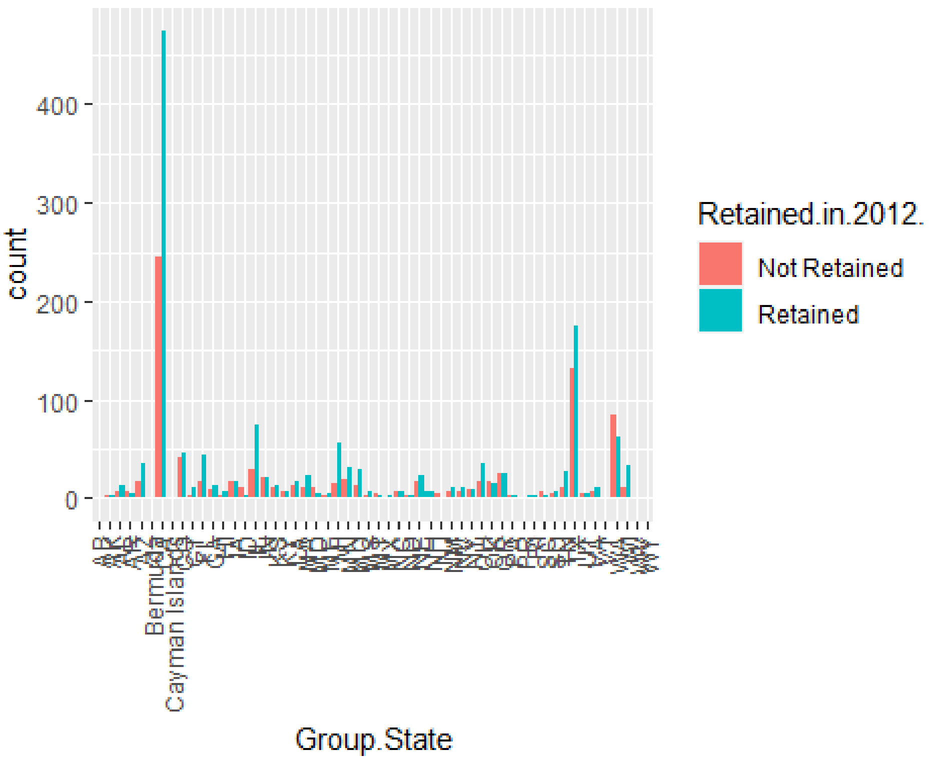
```
#Income Level
ggplot(data=dataset) +
geom_bar(mapping=aes(fill=Retained.in.2012.,x=Income.Level),position="dodge")
```



```
#As we can see the higher income levels have retained rate more than the
lower income levels, so we can say thay the income level does have an impact
on the target variable



#As we can see in the below graph, the Program code which have the highest
retained rate is HC, HT,HS and CD
ggplot(data=dataset) +
geom_bar(mapping=aes(fill=Retained.in.2012.,x=Program.Code),position="dodge")
```



```
#Our High target countries are
#1.California
#2.Texas
#3.Washington
#4.Illinois
ggplot(data=dataset) +
geom_bar(mapping=aes(fill=Retained.in.2012.,x=Group.State),position="dodge")
+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



##Chi-Square Test - for all variables ## If if the p-value is below your threshold of significance (typically p < 0.05), you can reject the null hypothesis

##A pvalue higher than 0.05 (greater than 0.05) is not statistically significant and indicates strong evidence for the null hypothesis. This means we retain the null hypothesis and reject the alternative hypothesis.

```
### Chi-square for all variables
chisq.test(dataset$Program.Code, dataset$Retained.in.2012., correct = FALSE)
#p-value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$Program.Code, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$Program.Code and dataset$Retained.in.2012.
## X-squared = 116.78, df = 27, p-value = 3.89e-13

chisq.test(dataset$From.Grade, dataset$Retained.in.2012., correct = FALSE)#p-
value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$From.Grade, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$From.Grade and dataset$Retained.in.2012.
## X-squared = 421.53, df = 10, p-value < 2.2e-16
```

```
chisq.test(dataset$To.Grade, dataset$Retained.in.2012., correct = FALSE) #p-
value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$To.Grade, dataset$Retained.in.2012., correct
=
## FALSE): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$To.Grade and dataset$Retained.in.2012.
## X-squared = 164.22, df = 10, p-value < 2.2e-16
```

```
chisq.test(dataset$Group.State, dataset$Retained.in.2012., correct = FALSE)
#p-value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$Group.State, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$Group.State and dataset$Retained.in.2012.
## X-squared = 124.47, df = 53, p-value = 1.11e-07
```

```
chisq.test(dataset$Is.Non.Annual., dataset$Retained.in.2012., correct =
FALSE) #p-value is less than 0.05 reject null hypothesis
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$Is.Non.Annual. and dataset$Retained.in.2012.
## X-squared = 364.55, df = 1, p-value < 2.2e-16
```

```
chisq.test(dataset$Days, dataset$Retained.in.2012., correct = FALSE) #p-value
is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$Days, dataset$Retained.in.2012., correct =
FALSE):
## Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
```

```
## data:  dataset$Days and dataset$Retained.in.2012.
## X-squared = 53.769, df = 11, p-value = 1.301e-07
```

```
chisq.test(dataset$Travel.Type, dataset$Retained.in.2012., correct = FALSE)
#p-value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$Travel.Type, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$Travel.Type and dataset$Retained.in.2012.
## X-squared = 16.135, df = 3, p-value = 0.001064
```

```
chisq.test(dataset$Departure.Date, dataset$Retained.in.2012., correct =
FALSE) #p-value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$Departure.Date, dataset$Retained.in.2012., :
Chi-
## squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$Departure.Date and dataset$Retained.in.2012.
## X-squared = 250.81, df = 143, p-value = 6.527e-08
```

```
chisq.test(dataset$Return.Date, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 so significant accept Null hypothesis
```

```
## Warning in chisq.test(dataset$Return.Date, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$Return.Date and dataset$Retained.in.2012.
## X-squared = 231.67, df = 142, p-value = 2.975e-06
```

```
chisq.test(dataset$Deposit.Date, dataset$Retained.in.2012., correct = FALSE)
#p-value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$Deposit.Date, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
```

```
## data:  dataset$Deposit.Date and dataset$Retained.in.2012.
## X-squared = 185.23, df = 134, p-value = 0.002251

chisq.test(dataset$Tuition, dataset$Retained.in.2012., correct = FALSE) #p-
value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$Tuition, dataset$Retained.in.2012., correct
=
## FALSE): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$Tuition and dataset$Retained.in.2012.
## X-squared = 1271.5, df = 1229, p-value = 0.1944

chisq.test(dataset$FRP.Active, dataset$Retained.in.2012., correct = FALSE)
#p-value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$FRP.Active, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$FRP.Active and dataset$Retained.in.2012.
## X-squared = 276.23, df = 92, p-value < 2.2e-16

chisq.test(dataset$FRP.Cancelled, dataset$Retained.in.2012., correct =
FALSE)#Not Significant -  strong evidence for null hypothesis no association

## Warning in chisq.test(dataset$FRP.Cancelled, dataset$Retained.in.2012., :
Chi-
## squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$FRP.Cancelled and dataset$Retained.in.2012.
## X-squared = 33.176, df = 28, p-value = 0.2293

chisq.test(dataset$FRP.Take.up.percent., dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$FRP.Take.up.percent.,
dataset$Retained.in.2012., :
## Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
```

```
## data:  dataset$FRP.Take.up.percent. and dataset$Retained.in.2012.
## X-squared = 550.2, df = 475, p-value = 0.009562

chisq.test(dataset$Early.RPL, dataset$Retained.in.2012., correct = FALSE)#p-
value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$Early.RPL, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$Early.RPL and dataset$Retained.in.2012.
## X-squared = 190.16, df = 141, p-value = 0.003664

chisq.test(dataset$Latest.RPL, dataset$Retained.in.2012., correct = FALSE)#p-
value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$Latest.RPL, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$Latest.RPL and dataset$Retained.in.2012.
## X-squared = 278.77, df = 214, p-value = 0.001911

chisq.test(dataset$Cancelled.Pax, dataset$Retained.in.2012., correct =
FALSE)#Not Significant -  strong evidence for null hypothesis no association

## Warning in chisq.test(dataset$Cancelled.Pax, dataset$Retained.in.2012., :
Chi-
## squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$Cancelled.Pax and dataset$Retained.in.2012.
## X-squared = 33.719, df = 33, p-value = 0.4325

chisq.test(dataset$Total.Discount.Pax, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$Total.Discount.Pax,
dataset$Retained.in.2012., :
## Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
```

```
## data:  dataset$Total.Discount.Pax and dataset$Retained.in.2012.
## X-squared = 187.83, df = 25, p-value < 2.2e-16

chisq.test(dataset$Initial.System.Date, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$Initial.System.Date,
dataset$Retained.in.2012., :
## Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$Initial.System.Date and dataset$Retained.in.2012.
## X-squared = 435.93, df = 295, p-value = 1.708e-07

chisq.test(dataset$Poverty.Code, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$Poverty.Code, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$Poverty.Code and dataset$Retained.in.2012.
## X-squared = 38.421, df = 5, p-value = 3.106e-07

chisq.test(dataset$Region, dataset$Retained.in.2012., correct = FALSE)#p-
value is less than 0.05 reject null hypothesis

##
##  Pearson's Chi-squared test
##
## data:  dataset$Region and dataset$Retained.in.2012.
## X-squared = 36.425, df = 5, p-value = 7.807e-07

chisq.test(dataset$CRM.Segment, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$CRM.Segment, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$CRM.Segment and dataset$Retained.in.2012.
## X-squared = 154.13, df = 11, p-value < 2.2e-16
```

```
chisq.test(dataset$School.Type, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis

##
##  Pearson's Chi-squared test
##
## data:  dataset$School.Type and dataset$Retained.in.2012.
## X-squared = 14.228, df = 3, p-value = 0.00261

chisq.test(dataset$Parent.Meeting.Flag, dataset$Retained.in.2012., correct =
FALSE)#Not Significant - strong evidence for null hypothesis no association

##
##  Pearson's Chi-squared test
##
## data:  dataset$Parent.Meeting.Flag and dataset$Retained.in.2012.
## X-squared = 1.0022, df = 1, p-value = 0.3168

chisq.test(dataset$MDR.Low.Grade, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$MDR.Low.Grade, dataset$Retained.in.2012., :
Chi-
## squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$MDR.Low.Grade and dataset$Retained.in.2012.
## X-squared = 92.837, df = 11, p-value = 4.625e-15

chisq.test(dataset$MDR.High.Grade, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$MDR.High.Grade, dataset$Retained.in.2012., :
Chi-
## squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$MDR.High.Grade and dataset$Retained.in.2012.
## X-squared = 84.994, df = 12, p-value = 4.563e-13

chisq.test(dataset$Total.School.Enrollment, dataset$Retained.in.2012.,
correct = FALSE)#p-value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$Total.School.Enrollment,
## dataset$Retained.in.2012., : Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
```

```
##
## data:  dataset$Total.School.Enrollment and dataset$Retained.in.2012.
## X-squared = 968.43, df = 893, p-value = 0.03984
```

```
chisq.test(dataset$Income.Level, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$Income.Level, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$Income.Level and dataset$Retained.in.2012.
## X-squared = 85.119, df = 21, p-value = 1.106e-09
```

```
chisq.test(dataset$EZ.Pay.Take.Up.Rate, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$EZ.Pay.Take.Up.Rate,
dataset$Retained.in.2012., :
## Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$EZ.Pay.Take.Up.Rate and dataset$Retained.in.2012.
## X-squared = 450.32, df = 370, p-value = 0.002662
```

```
chisq.test(dataset$School.Sponsor, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$School.Sponsor and dataset$Retained.in.2012.
## X-squared = 34.814, df = 1, p-value = 3.627e-09
```

```
chisq.test(dataset$SPR.Product.Type, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$SPR.Product.Type, dataset$Retained.in.2012.,
:
## Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$SPR.Product.Type and dataset$Retained.in.2012.
## X-squared = 64.032, df = 5, p-value = 1.779e-12
```

```
chisq.test(dataset$SPR.New.Existing, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 so significant accept Null hypothesis
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$SPR.New.Existing and dataset$Retained.in.2012.
## X-squared = 325.16, df = 1, p-value < 2.2e-16
```

```
chisq.test(dataset$FPP, dataset$Retained.in.2012., correct = FALSE)#p-value
is less than 0.05 so significant accept Null hypothesis
```

```
## Warning in chisq.test(dataset$FPP, dataset$Retained.in.2012., correct =
FALSE):
## Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$FPP and dataset$Retained.in.2012.
## X-squared = 327.24, df = 145, p-value = 4.159e-16
```

```
chisq.test(dataset$Total.Pax, dataset$Retained.in.2012., correct = FALSE)#p-
value is less than 0.05 so significant accept Null hypothesis
```

```
## Warning in chisq.test(dataset$Total.Pax, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$Total.Pax and dataset$Retained.in.2012.
## X-squared = 343.4, df = 158, p-value = 8.902e-16
```

```
chisq.test(dataset$SPR.Group.Revenue, dataset$Retained.in.2012., correct =
FALSE)#Not Significant Accept Null Hypothesis
```

```
## Warning in chisq.test(dataset$SPR.Group.Revenue,
dataset$Retained.in.2012., :
## Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$SPR.Group.Revenue and dataset$Retained.in.2012.
## X-squared = 1271.5, df = 1229, p-value = 0.1944
```

```
chisq.test(dataset$NumberOfMeetingswithParents, dataset$Retained.in.2012.,
correct = FALSE)#Not Significant, Accept Null Hypothesis
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$NumberOfMeetingswithParents and dataset$Retained.in.2012.
## X-squared = 8.0861, df = 2, p-value = 0.01754

chisq.test(dataset$FirstMeeting, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$FirstMeeting, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$FirstMeeting and dataset$Retained.in.2012.
## X-squared = 287.53, df = 206, p-value = 0.00015

chisq.test(dataset$LastMeeting, dataset$Retained.in.2012., correct =
FALSE)##p-value is less than 0.05 reject null hypothesis

## Warning in chisq.test(dataset$LastMeeting, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$LastMeeting and dataset$Retained.in.2012.
## X-squared = 255.54, df = 171, p-value = 2.964e-05

chisq.test(dataset$DifferenceTraveltoFirstMeeting, dataset$Retained.in.2012.,
correct = FALSE)#Not Significant

## Warning in chisq.test(dataset$DifferenceTraveltoFirstMeeting,
## dataset$Retained.in.2012., : Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dataset$DifferenceTraveltoFirstMeeting and
dataset$Retained.in.2012.
## X-squared = 366.38, df = 342, p-value = 0.1746

chisq.test(dataset$DifferenceTraveltoLastMeeting, dataset$Retained.in.2012.,
correct = FALSE) #Not Significant

## Warning in chisq.test(dataset$DifferenceTraveltoLastMeeting,
## dataset$Retained.in.2012., : Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$DifferenceTraveltoLastMeeting and dataset$Retained.in.2012.
## X-squared = 281.24, df = 251, p-value = 0.09202
```

```
chisq.test(dataset$SchoolGradeTypeLow, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$SchoolGradeTypeLow and dataset$Retained.in.2012.
## X-squared = 78.368, df = 3, p-value < 2.2e-16
```

```
chisq.test(dataset$SchoolGradeTypeHigh, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$SchoolGradeTypeHigh and dataset$Retained.in.2012.
## X-squared = 144.26, df = 3, p-value < 2.2e-16
```

```
chisq.test(dataset$SchoolGradeType, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$SchoolGradeType, dataset$Retained.in.2012.,
: Chi-
## squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$SchoolGradeType and dataset$Retained.in.2012.
## X-squared = 168.43, df = 8, p-value < 2.2e-16
```

```
chisq.test(dataset$DepartureMonth, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$DepartureMonth, dataset$Retained.in.2012., :
Chi-
## squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$DepartureMonth and dataset$Retained.in.2012.
## X-squared = 85.954, df = 5, p-value < 2.2e-16
```

```
chisq.test(dataset$GroupGradeTypeLow, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$GroupGradeTypeLow and dataset$Retained.in.2012.
## X-squared = 87.771, df = 5, p-value < 2.2e-16
```

```
chisq.test(dataset$GroupGradeTypeHigh, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$GroupGradeTypeHigh and dataset$Retained.in.2012.
## X-squared = 63.205, df = 3, p-value = 1.214e-13
```

```
chisq.test(dataset$GroupGradeType, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$GroupGradeType, dataset$Retained.in.2012., :
Chi-
## squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$GroupGradeType and dataset$Retained.in.2012.
## X-squared = 122.05, df = 12, p-value < 2.2e-16
```

```
chisq.test(dataset$MajorProgramCode, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$MajorProgramCode and dataset$Retained.in.2012.
## X-squared = 56.326, df = 3, p-value = 3.579e-12
```

```
chisq.test(dataset$SingleGradeTripFlag, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
##
##  Pearson's Chi-squared test
##
## data:  dataset$SingleGradeTripFlag and dataset$Retained.in.2012.
## X-squared = 496.48, df = 1, p-value < 2.2e-16
```

```
chisq.test(dataset$FPP.to.School.enrollment, dataset$Retained.in.2012.,
correct = FALSE)# Accept Null hypothesis
```

```
## Warning in chisq.test(dataset$FPP.to.School.enrollment,
## dataset$Retained.in.2012., : Chi-squared approximation may be incorrect
```

```
## 
##  Pearson's Chi-squared test
## 
## data:  dataset$FPP.to.School.enrollment and dataset$Retained.in.2012.
## X-squared = 1863.3, df = 1909, p-value = 0.7687
```

```
chisq.test(dataset$FPP.to.PAX, dataset$Retained.in.2012., correct = FALSE)#p-
value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$FPP.to.PAX, dataset$Retained.in.2012.,
correct =
## FALSE): Chi-squared approximation may be incorrect
```

```
## 
##  Pearson's Chi-squared test
## 
## data:  dataset$FPP.to.PAX and dataset$Retained.in.2012.
## X-squared = 392.18, df = 305, p-value = 0.0005437
```

```
chisq.test(dataset$Num.of.Non_FPP.PAX, dataset$Retained.in.2012., correct =
FALSE)#p-value is less than 0.05 reject null hypothesis
```

```
## Warning in chisq.test(dataset$Num.of.Non_FPP.PAX,
dataset$Retained.in.2012., :
## Chi-squared approximation may be incorrect
```

```
## 
##  Pearson's Chi-squared test
## 
## data:  dataset$Num.of.Non_FPP.PAX and dataset$Retained.in.2012.
## X-squared = 187.83, df = 25, p-value < 2.2e-16
```

###Random Forest

```
ntree <- 100
set.seed(123)

attach(dataset)
colnames(dataset)
```

```
##  [1] "ID"                                 "Program.Code"
##  [3] "From.Grade"                         "To.Grade"
##  [5] "Group.State"                        "Is.Non.Annual."
##  [7] "Days"                               "Travel.Type"
##  [9] "Departure.Date"                     "Return.Date"
## [11] "Deposit.Date"                       "Special.Pay"
## [13] "Tuition"                            "FRP.Active"
## [15] "FRP.Cancelled"                      "FRP.Take.up.percent."
## [17] "Early.RPL"                          "Latest.RPL"
## [19] "Cancelled.Pax"                      "Total.Discount.Pax"
## [21] "Initial.System.Date"                "Poverty.Code"
## [23] "Region"                             "CRM.Segment"
## [25] "School.Type"                    "Parent.Meeting.Flag"
## [27] "MDR.Low.Grade"                  "MDR.High.Grade"
## [29] "Total.School.Enrollment"        "Income.Level"
## [31] "EZ.Pay.Take.Up.Rate"            "School.Sponsor"
## [33] "SPR.Product.Type"               "SPR.New.Existing"
## [35] "FPP"                            "Total.Pax"
## [37] "SPR.Group.Revenue"              "NumberOfMeetingswithParents"
## [39] "FirstMeeting"                   "LastMeeting"
## [41] "DifferenceTraveltoFirstMeeting" "DifferenceTraveltoLastMeeting"
## [43] "SchoolGradeTypeLow"             "SchoolGradeTypeHigh"
## [45] "SchoolGradeType"                "DepartureMonth"
## [47] "GroupGradeTypeLow"              "GroupGradeTypeHigh"
## [49] "GroupGradeType"                 "MajorProgramCode"
## [51] "SingleGradeTripFlag"            "FPP.to.School.enrollment"
## [53] "FPP.to.PAX"                     "Num.of.Non_FPP.PAX"
## [55] "SchoolSizeIndicator"            "Retained.in.2012."
```

```
#random_forest_data <- subset(dataset,select=-c(School.Type,FPP.to.School.enrollment,DifferenceTraveltoFirstMeeting,DifferenceTraveltoLastMeeting,Parent.Meeting.Flag,NumberOfMeetingswithParents,SchoolGradeType,Days,GroupGradeType,Group.State,SchoolSizeIndicator))
random_forest_data <- subset(dataset, select = -
c(1,9,10,11,12,17,18,21,39,40))

dataset <- random_forest_data

colnames(dataset)
```

```
##  [1] "Program.Code"                   "From.Grade"
##  [3] "To.Grade"                       "Group.State"
##  [5] "Is.Non.Annual."                 "Days"
##  [7] "Travel.Type"                    "Tuition"
##  [9] "FRP.Active"                     "FRP.Cancelled"
## [11] "FRP.Take.up.percent."           "Cancelled.Pax"
## [13] "Total.Discount.Pax"             "Poverty.Code"
## [15] "Region"                         "CRM.Segment"
## [17] "School.Type"                    "Parent.Meeting.Flag"
## [19] "MDR.Low.Grade"                  "MDR.High.Grade"
## [21] "Total.School.Enrollment"        "Income.Level"
## [23] "EZ.Pay.Take.Up.Rate"            "School.Sponsor"
## [25] "SPR.Product.Type"               "SPR.New.Existing"
## [27] "FPP"                            "Total.Pax"
## [29] "SPR.Group.Revenue"              "NumberOfMeetingswithParents"
## [31] "DifferenceTraveltoFirstMeeting" "DifferenceTraveltoLastMeeting"
## [33] "SchoolGradeTypeLow"             "SchoolGradeTypeHigh"
## [35] "SchoolGradeType"                "DepartureMonth"
## [37] "GroupGradeTypeLow"              "GroupGradeTypeHigh"
## [39] "GroupGradeType"                 "MajorProgramCode"
## [41] "SingleGradeTripFlag"            "FPP.to.School.enrollment"
## [43] "FPP.to.PAX"                     "Num.of.Non_FPP.PAX"
## [45] "SchoolSizeIndicator"            "Retained.in.2012."
```

```
random_forest_data <- subset(random_forest_data,select=-c(4))
#random_forest_data <- dataset
sum(is.na(dataset))
```

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

```
myFormula = Retained.in.2012.~ .

rf <- randomForest(myFormula, data = random_forest_data, mtry =
sqrt(ncol(random_forest_data)-1), ntree = 300,
proximity = T, importance = T)

print(rf)
```

```
##
## Call:
##  randomForest(formula = myFormula, data = random_forest_data,      mtry =
sqrt(ncol(random_forest_data) - 1), ntree = 300, proximity = T,
importance = T)
##                Type of random forest: classification
##                      Number of trees: 300
## No. of variables tried at each split: 7
##
##         OOB estimate of  error rate: 20.47%
## Confusion matrix:
##              Not Retained Retained class.error
## Not Retained          627      311   0.3315565
## Retained              178     1273   0.1226740
```

```
#rf$proximity, 10

#Assigning the importance for each variable
rf$importance
```

```
##                                 Not Retained      Retained
MeanDecreaseAccuracy
## Program.Code                    1.220190e-02  4.300084e-04
5.047320e-03
## From.Grade                      3.263587e-02  1.013142e-02
1.893242e-02
## To.Grade                        6.658087e-03  3.357332e-03
4.626117e-03
## Is.Non.Annual.                  5.583996e-02  3.770173e-02
4.485063e-02
## Days                            8.698600e-04  7.790761e-04
8.231958e-04
## Travel.Type                     5.036824e-04  2.325070e-04
3.357493e-04
```

```
## Tuition                          4.408248e-03  5.210003e-03
4.908133e-03
## FRP.Active                       8.798991e-03  1.143639e-02
1.044779e-02
## FRP.Cancelled                   -9.235923e-04  2.819168e-03
1.343264e-03
## FRP.Take.up.percent.            -7.676785e-04  6.330729e-03
3.553310e-03
## Cancelled.Pax                    3.354124e-04  2.131821e-03
1.423160e-03
## Total.Discount.Pax               4.329104e-03  6.038958e-03
5.364577e-03
## Poverty.Code                     1.603008e-03  5.870625e-04
9.779989e-04
## Region                           3.684297e-03  6.408279e-04
1.856043e-03
## CRM.Segment                      8.736190e-03  9.105081e-04
3.995742e-03
## School.Type                      1.572909e-05  7.672982e-04
4.760891e-04
## Parent.Meeting.Flag              4.061180e-04 -2.056672e-04
3.180288e-05
## MDR.Low.Grade                    4.938093e-03 -3.787405e-04
1.727630e-03
## MDR.High.Grade                   8.179645e-03 -1.194183e-04
3.094759e-03
## Total.School.Enrollment          8.541012e-03  6.654695e-03
7.370701e-03
## Income.Level                     6.038436e-03  4.827293e-04
2.669529e-03
## EZ.Pay.Take.Up.Rate              3.475157e-04  3.265628e-03
2.125007e-03
## School.Sponsor                   2.084921e-04  8.338308e-04
5.873610e-04
## SPR.Product.Type                 6.773083e-04  1.334851e-04
3.422242e-04
## SPR.New.Existing                 3.871758e-02  1.622576e-02
2.505550e-02
## FPP                              1.106853e-02  2.524507e-02
1.969092e-02
## Total.Pax                        1.009093e-02  2.279866e-02
1.780919e-02
## SPR.Group.Revenue                2.739252e-03  5.478128e-03
4.396920e-03
## NumberOfMeetingswithParents      6.988646e-04 -6.729757e-05
2.207130e-04
## DifferenceTraveltoFirstMeeting  1.486359e-03  2.021776e-03
1.801702e-03
## DifferenceTraveltoLastMeeting   1.442132e-03  1.904877e-03
1.707757e-03
```

```
## SchoolGradeTypeLow              5.976916e-04  4.110387e-04
4.791291e-04
## SchoolGradeTypeHigh             3.434559e-03  2.254719e-03
2.701198e-03
## SchoolGradeType                 5.170628e-03  2.662106e-03
3.633769e-03
## DepartureMonth                  2.259155e-03  7.101453e-04
1.327865e-03
## GroupGradeTypeLow               3.792819e-03  8.850008e-04
2.017611e-03
## GroupGradeTypeHigh              2.247692e-03  1.085481e-04
9.330802e-04
## GroupGradeType                  1.010773e-02  4.008388e-04
4.222387e-03
## MajorProgramCode                3.320017e-04 -1.592303e-04
3.305450e-05
## SingleGradeTripFlag             4.426686e-02  1.905422e-02
2.889830e-02
## FPP.to.School.enrollment        4.114832e-03  1.092227e-02
8.250622e-03
## FPP.to.PAX                      3.988260e-03  3.674691e-03
3.794787e-03
## Num.of.Non_FPP.PAX              3.662207e-03  6.301064e-03
5.277022e-03
## SchoolSizeIndicator             2.689817e-03  2.049014e-03
2.294723e-03
##                                MeanDecreaseGini
## Program.Code                          33.628268
## From.Grade                            67.268281
## To.Grade                              19.427584
## Is.Non.Annual.                        84.059914
## Days                                  10.883878
## Travel.Type                            2.463068
## Tuition                               30.651917
## FRP.Active                            33.626157
## FRP.Cancelled                         18.680562
## FRP.Take.up.percent.                  30.208816
## Cancelled.Pax                         20.222592
## Total.Discount.Pax                    16.481201
## Poverty.Code                          11.601516
## Region                                21.982345
## CRM.Segment                           25.133001
## School.Type                            6.971828
## Parent.Meeting.Flag                    2.534605
## MDR.Low.Grade                         18.682364
## MDR.High.Grade                        14.233641
## Total.School.Enrollment               39.929538
## Income.Level                          87.363529
## EZ.Pay.Take.Up.Rate                   26.444318
## School.Sponsor                         1.689846
## SPR.Product.Type                       2.705500
## SPR.New.Existing                      59.594879
## FPP                                   44.565041
## Total.Pax                             41.782578
## SPR.Group.Revenue                     30.090820
## NumberOfMeetingswithParents            7.337064
## DifferenceTraveltoFirstMeeting        30.496488
## DifferenceTraveltoLastMeeting         28.906695
## SchoolGradeTypeLow                     2.960485
## SchoolGradeTypeHigh                    8.847103
## SchoolGradeType                       16.644411
## DepartureMonth                        17.161511
## GroupGradeTypeLow                      9.987827
## GroupGradeTypeHigh                     4.143843
## GroupGradeType                        29.463884
## MajorProgramCode                       2.050545
## SingleGradeTripFlag                   81.567513
## FPP.to.School.enrollment              35.369481
## FPP.to.PAX                            28.761127
## Num.of.Non_FPP.PAX                    17.008883
## SchoolSizeIndicator                   15.129393
```

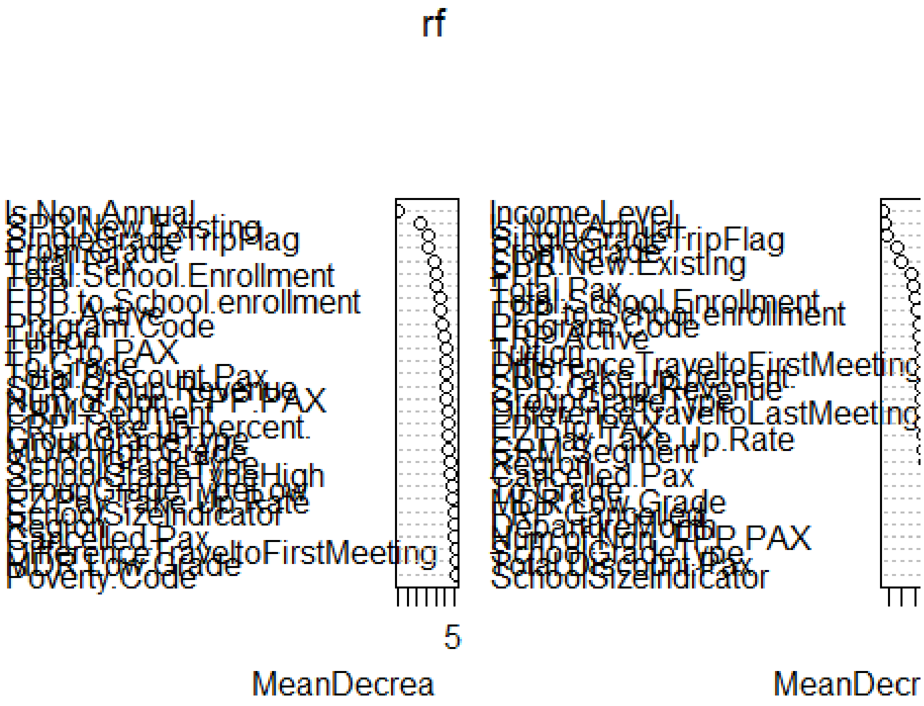
```
importance(rf, type = 1)
```

```
##                                MeanDecreaseAccuracy
## Program.Code                             11.0851356
## From.Grade                               18.6621369
## To.Grade                                  9.7211390
## Is.Non.Annual.                           34.5214555
## Days                                      4.2540955
## Travel.Type                               2.7777099
## Tuition                                  10.3977381
## FRP.Active                               12.4548924
## FRP.Cancelled                             4.4946422
## FRP.Take.up.percent.                      8.2880686
## Cancelled.Pax                             4.6836457
## Total.Discount.Pax                        9.6265104
## Poverty.Code                              4.5639953
## Region                                    5.4739246
## CRM.Segment                               8.7005997
## School.Type                               2.5074311
## Parent.Meeting.Flag                       0.2701960
## MDR.Low.Grade                             4.5873801
## MDR.High.Grade                            7.9747881
## Total.School.Enrollment                  15.0617824
## Income.Level                              4.4980561
## EZ.Pay.Take.Up.Rate                       5.8724183
## School.Sponsor                            4.5333766
## SPR.Product.Type                          2.6532781
## SPR.New.Existing                         22.7368412
## FPP                                      14.6903238
## Total.Pax                                15.5305461
## SPR.Group.Revenue                         9.5103847
## NumberOfMeetingswithParents               1.1829277
## DifferenceTraveltoFirstMeeting            4.6322275
## DifferenceTraveltoLastMeeting             4.3541164
## SchoolGradeTypeLow                        3.1565617
## SchoolGradeTypeHigh                       6.8278294
## SchoolGradeType                           7.5747245
## DepartureMonth                            4.3830641
## GroupGradeTypeLow                         6.6182605
## GroupGradeTypeHigh                        4.1682443
## GroupGradeType                            8.1491358
## MajorProgramCode                          0.3915151
## SingleGradeTripFlag                      19.2883887
## FPP.to.School.enrollment                 12.8722654
## FPP.to.PAX                                9.8741586
## Num.of.Non_FPP.PAX                        9.4487831
## SchoolSizeIndicator                       5.6557600
```

```
importance(rf, type = 2)
```

```
##                                 MeanDecreaseGini
## Program.Code                           33.628268
## From.Grade                             67.268281
## To.Grade                               19.427584
## Is.Non.Annual.                         84.059914
## Days                                   10.883878
## Travel.Type                             2.463068
## Tuition                                30.651917
## FRP.Active                             33.626157
## FRP.Cancelled                          18.680562
## FRP.Take.up.percent.                   30.208816
## Cancelled.Pax                          20.222592
## Total.Discount.Pax                     16.481201
## Poverty.Code                           11.601516
## Region                                 21.982345
## CRM.Segment                            25.133001
## School.Type                             6.971828
## Parent.Meeting.Flag                     2.534605
## MDR.Low.Grade                          18.682364
## MDR.High.Grade                         14.233641
## Total.School.Enrollment                39.929538
## Income.Level                           87.363529
## EZ.Pay.Take.Up.Rate                    26.444318
## School.Sponsor                          1.689846
## SPR.Product.Type                        2.705500
## SPR.New.Existing                       59.594879
## FPP                                    44.565041
## Total.Pax                              41.782578
## SPR.Group.Revenue                         30.090820
## NumberOfMeetingswithParents                7.337064
## DifferenceTraveltoFirstMeeting            30.496488
## DifferenceTraveltoLastMeeting             28.906695
## SchoolGradeTypeLow                         2.960485
## SchoolGradeTypeHigh                        8.847103
## SchoolGradeType                           16.644411
## DepartureMonth                            17.161511
## GroupGradeTypeLow                          9.987827
## GroupGradeTypeHigh                         4.143843
## GroupGradeType                            29.463884
## MajorProgramCode                           2.050545
## SingleGradeTripFlag                       81.567513
## FPP.to.School.enrollment                  35.369481
## FPP.to.PAX                                28.761127
## Num.of.Non_FPP.PAX                        17.008883
## SchoolSizeIndicator                       15.129393
```

```
#From this we can conclude the Income.Level, Is.Non.Annual and
SPR.NewExisting and Total.PAX are of higher importance, so we should focus to
get more retained rate
varImpPlot(rf)
```



```
rf$err.rate[ntree,1]
```

```
##       OOB
## 0.2046882
```

```
rf$predicted
```

```
##            1            2            3            4            5
6
##     Retained     Retained     Retained     Retained     Retained Not
Retained
##            7            8            9           10           11
12
## Not Retained     Retained     Retained     Retained     Retained
Retained
##           13           14           15           16           17
18
##     Retained     Retained     Retained     Retained     Retained
Retained
##           19           20           21           22           23
24
##     Retained     Retained     Retained     Retained     Retained
Retained
##           25           26           27           28           29
30
##     Retained     Retained     Retained     Retained     Retained
Retained
##           31           32           33           34           35
36
##     Retained     Retained Not Retained     Retained     Retained
Retained
##           37           38           39           40           41
42
##     Retained Not Retained     Retained     Retained     Retained
Retained
##           43           44           45           46           47
48
##     Retained     Retained     Retained     Retained     Retained
Retained
##           49           50           51           52           53
54
## Not Retained     Retained     Retained     Retained     Retained
Retained
##           55           56           57           58           59
60
##     Retained Not Retained Not Retained     Retained     Retained
Retained
##           61           62           63           64           65
66
##     Retained     Retained Not Retained Not Retained     Retained Not
Retained
##           67           68           69           70           71
72
##     Retained     Retained     Retained     Retained     Retained
Retained
```

```
##           73           74           75           76           77 
78 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##           79           80           81           82           83 
84 
##     Retained     Retained     Retained Not Retained     Retained 
Retained 
##           85           86           87           88           89 
90 
##     Retained Not Retained     Retained     Retained     Retained Not 
Retained 
##           91           92           93           94           95 
96 
##     Retained     Retained     Retained Not Retained     Retained 
Retained 
##           97           98           99          100          101 
102 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          103          104          105          106          107 
108 
##     Retained     Retained     Retained     Retained Not Retained Not 
Retained 
##          109          110          111          112          113 
114 
##     Retained     Retained Not Retained     Retained     Retained 
Retained 
##          115          116          117          118          119 
120 
##     Retained     Retained Not Retained     Retained     Retained 
Retained 
##          121          122          123          124          125 
126 
##     Retained     Retained Not Retained     Retained     Retained Not 
Retained 
##          127          128          129          130          131 
132 
##     Retained     Retained Not Retained     Retained     Retained 
Retained 
##          133          134          135          136          137 
138 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          139          140          141          142          143 
144 
##     Retained     Retained Not Retained     Retained     Retained Not 
Retained 
##          145          146          147          148          149 
150 
```

```
##     Retained Not Retained     Retained     Retained     Retained 
Retained 
##          151          152          153          154          155 
156 
##     Retained Not Retained Not Retained     Retained     Retained 
Retained 
##          157          158          159          160          161 
162 
##     Retained     Retained     Retained     Retained Not Retained Not 
Retained 
##          163          164          165          166          167 
168 
##     Retained Not Retained     Retained Not Retained     Retained Not 
Retained 
##          169          170          171          172          173 
174 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          175          176          177          178          179 
180 
##     Retained     Retained     Retained     Retained     Retained Not 
Retained 
##          181          182          183          184          185 
186 
## Not Retained     Retained     Retained Not Retained Not Retained 
Retained 
##          187          188          189          190          191 
192 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          193          194          195          196          197 
198 
## Not Retained Not Retained     Retained Not Retained Not Retained Not 
Retained 
##          199          200          201          202          203 
204 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          205          206          207          208          209 
210 
## Not Retained     Retained     Retained Not Retained     Retained Not 
Retained 
##          211          212          213          214          215 
216 
## Not Retained     Retained     Retained Not Retained     Retained Not 
Retained 
##          217          218          219          220          221 
222 
##     Retained Not Retained Not Retained     Retained     Retained Not 
Retained
```

```
##          223          224          225          226          227
228
##     Retained Not Retained     Retained     Retained     Retained
Retained
##          229          230          231          232          233
234
##     Retained     Retained     Retained Not Retained Not Retained
Retained
##          235          236          237          238          239
240
##     Retained     Retained     Retained     Retained     Retained
Retained
##          241          242          243          244          245
246
##     Retained Not Retained     Retained     Retained     Retained
Retained
##          247          248          249          250          251
252
## Not Retained     Retained     Retained     Retained     Retained
Retained
##          253          254          255          256          257
258
##     Retained     Retained Not Retained     Retained     Retained
Retained
##          259          260          261          262          263
264
##     Retained     Retained     Retained     Retained     Retained Not
Retained
##          265          266          267          268          269
270
## Not Retained Not Retained     Retained Not Retained     Retained Not
Retained
##          271          272          273          274          275
276
##     Retained     Retained Not Retained Not Retained     Retained Not
Retained
##          277          278          279          280          281
282
##     Retained Not Retained Not Retained     Retained Not Retained Not
Retained
##          283          284          285          286          287
288
##     Retained     Retained Not Retained Not Retained     Retained
Retained
##          289          290          291          292          293
294
##     Retained     Retained     Retained     Retained     Retained Not
Retained
##          295          296          297          298          299
300
```

```
## Not Retained     Retained     Retained     Retained     Retained Not 
Retained 
##          301          302          303          304          305 
306 
## Not Retained Not Retained Not Retained     Retained     Retained Not 
Retained 
##          307          308          309          310          311 
312 
##     Retained Not Retained Not Retained     Retained     Retained Not 
Retained 
##          313          314          315          316          317 
318 
##     Retained     Retained     Retained Not Retained     Retained 
Retained 
##          319          320          321          322          323 
324 
## Not Retained Not Retained     Retained     Retained Not Retained 
Retained 
##          325          326          327          328          329 
330 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          331          332          333          334          335 
336 
##     Retained     Retained     Retained     Retained     Retained Not 
Retained 
##          337          338          339          340          341 
342 
##     Retained Not Retained Not Retained     Retained     Retained 
Retained 
##          343          344          345          346          347 
348 
##     Retained Not Retained     Retained     Retained     Retained 
Retained 
##          349          350          351          352          353 
354 
##     Retained     Retained     Retained Not Retained     Retained 
Retained 
##          355          356          357          358          359 
360 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          361          362          363          364          365 
366 
##     Retained Not Retained Not Retained     Retained     Retained Not 
Retained 
##          367          368          369          370          371 
372 
##     Retained Not Retained     Retained Not Retained Not Retained 
Retained
```

```
##          373          374          375          376          377 
378 
## Not Retained     Retained     Retained Not Retained Not Retained 
Retained 
##          379          380          381          382          383 
384 
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##          385          386          387          388          389 
390 
## Not Retained Not Retained     Retained     Retained     Retained 
Retained 
##          391          392          393          394          395 
396 
##     Retained     Retained Not Retained     Retained     Retained 
Retained 
##          397          398          399          400          401 
402 
## Not Retained     Retained     Retained Not Retained Not Retained 
Retained 
##          403          404          405          406          407 
408 
##     Retained     Retained     Retained     Retained Not Retained Not 
Retained 
##          409          410          411          412          413 
414 
## Not Retained Not Retained     Retained Not Retained     Retained 
Retained 
##          415          416          417          418          419 
420 
## Not Retained     Retained Not Retained     Retained Not Retained 
Retained 
##          421          422          423          424          425 
426 
##     Retained     Retained Not Retained     Retained     Retained Not 
Retained 
##          427          428          429          430          431 
432 
## Not Retained     Retained Not Retained Not Retained Not Retained 
Retained 
##          433          434          435          436          437 
438 
## Not Retained     Retained Not Retained     Retained Not Retained 
Retained 
##          439          440          441          442          443 
444 
## Not Retained     Retained Not Retained Not Retained     Retained 
Retained 
##          445          446          447          448          449 
450
```

```
##     Retained     Retained     Retained     Retained Not Retained Not
Retained
##          451          452          453          454          455
456
##     Retained     Retained     Retained     Retained     Retained
Retained
##          457          458          459          460          461
462
##     Retained     Retained     Retained     Retained     Retained
Retained
##          463          464          465          466          467
468
##     Retained     Retained     Retained     Retained     Retained
Retained
##          469          470          471          472          473
474
## Not Retained     Retained Not Retained Not Retained     Retained
Retained
##          475          476          477          478          479
480
##     Retained Not Retained Not Retained     Retained Not Retained
Retained
##          481          482          483          484          485
486
##     Retained     Retained     Retained     Retained     Retained
Retained
##          487          488          489          490          491
492
## Not Retained     Retained     Retained     Retained     Retained Not
Retained
##          493          494          495          496          497
498
##     Retained Not Retained     Retained Not Retained     Retained
Retained
##          499          500          501          502          503
504
##     Retained     Retained     Retained Not Retained     Retained Not
Retained
##          505          506          507          508          509
510
##     Retained     Retained     Retained Not Retained     Retained
Retained
##          511          512          513          514          515
516
##     Retained     Retained     Retained     Retained     Retained
Retained
##          517          518          519          520          521
522
##     Retained     Retained     Retained     Retained Not Retained
Retained
```

```
##          523          524          525          526          527 
528 
##     Retained Not Retained Not Retained     Retained     Retained Not 
Retained 
##          529          530          531          532          533 
534 
## Not Retained Not Retained     Retained     Retained     Retained Not 
Retained 
##          535          536          537          538          539 
540 
##     Retained     Retained     Retained Not Retained Not Retained Not 
Retained 
##          541          542          543          544          545 
546 
##     Retained     Retained Not Retained Not Retained     Retained Not 
Retained 
##          547          548          549          550          551 
552 
##     Retained     Retained     Retained     Retained Not Retained 
Retained 
##          553          554          555          556          557 
558 
##     Retained     Retained     Retained     Retained Not Retained Not 
Retained 
##          559          560          561          562          563 
564 
##     Retained     Retained     Retained Not Retained Not Retained 
Retained 
##          565          566          567          568          569 
570 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          571          572          573          574          575 
576 
##     Retained     Retained     Retained     Retained     Retained Not 
Retained 
##          577          578          579          580          581 
582 
##     Retained Not Retained     Retained     Retained     Retained Not 
Retained 
##          583          584          585          586          587 
588 
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##          589          590          591          592          593 
594 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          595          596          597          598          599 
600 
```

```
##     Retained Not Retained     Retained     Retained     Retained 
Retained 
##          601          602          603          604          605 
606 
##     Retained     Retained     Retained     Retained     Retained Not 
Retained 
##          607          608          609          610          611 
612 
## Not Retained     Retained     Retained     Retained     Retained Not 
Retained 
##          613          614          615          616          617 
618 
##     Retained     Retained     Retained Not Retained     Retained 
Retained 
##          619          620          621          622          623 
624 
##     Retained     Retained     Retained     Retained     Retained Not 
Retained 
##          625          626          627          628          629 
630 
##     Retained     Retained     Retained     Retained Not Retained Not 
Retained 
##          631          632          633          634          635 
636 
##     Retained     Retained     Retained     Retained Not Retained Not 
Retained 
##          637          638          639          640          641 
642 
##     Retained     Retained     Retained Not Retained Not Retained Not 
Retained 
##          643          644          645          646          647 
648 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          649          650          651          652          653 
654 
## Not Retained     Retained Not Retained     Retained     Retained 
Retained 
##          655          656          657          658          659 
660 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          661          662          663          664          665 
666 
##     Retained Not Retained Not Retained     Retained     Retained 
Retained 
##          667          668          669          670          671 
672 
## Not Retained     Retained Not Retained     Retained Not Retained 
Retained
```

```
##          673          674          675          676          677
678
##     Retained     Retained     Retained Not Retained     Retained
Retained
##          679          680          681          682          683
684
## Not Retained     Retained     Retained     Retained     Retained Not
Retained
##          685          686          687          688          689
690
## Not Retained Not Retained     Retained     Retained     Retained
Retained
##          691          692          693          694          695
696
##     Retained     Retained     Retained     Retained Not Retained
Retained
##          697          698          699          700          701
702
##     Retained     Retained     Retained     Retained     Retained
Retained
##          703          704          705          706          707
708
##     Retained     Retained Not Retained Not Retained Not Retained
Retained
##          709          710          711          712          713
714
##     Retained     Retained     Retained     Retained     Retained
Retained
##          715          716          717          718          719
720
## Not Retained     Retained     Retained     Retained     Retained Not
Retained
##          721          722          723          724          725
726
##     Retained     Retained     Retained Not Retained Not Retained
Retained
##          727          728          729          730          731
732
## Not Retained     Retained     Retained Not Retained     Retained Not
Retained
##          733          734          735          736          737
738
##     Retained Not Retained     Retained     Retained     Retained
Retained
##          739          740          741          742          743
744
##     Retained Not Retained Not Retained Not Retained     Retained
Retained
##          745          746          747          748          749
750
```

```
##     Retained Not Retained     Retained     Retained     Retained Not
Retained
##          751          752          753          754          755
756
##     Retained Not Retained Not Retained     Retained Not Retained Not
Retained
##          757          758          759          760          761
762
##     Retained     Retained     Retained Not Retained Not Retained
Retained
##          763          764          765          766          767
768
##     Retained     Retained Not Retained     Retained     Retained
Retained
##          769          770          771          772          773
774
##     Retained Not Retained Not Retained Not Retained     Retained Not
Retained
##          775          776          777          778          779
780
## Not Retained Not Retained Not Retained     Retained     Retained
Retained
##          781          782          783          784          785
786
##     Retained Not Retained Not Retained     Retained     Retained Not
Retained
##          787          788          789          790          791
792
##     Retained     Retained Not Retained     Retained     Retained Not
Retained
##          793          794          795          796          797
798
##     Retained Not Retained     Retained     Retained Not Retained
Retained
##          799          800          801          802          803
804
##     Retained     Retained     Retained     Retained Not Retained
Retained
##          805          806          807          808          809
810
##     Retained     Retained Not Retained Not Retained Not Retained
Retained
##          811          812          813          814          815
816
##     Retained     Retained     Retained     Retained Not Retained
Retained
##          817          818          819          820          821
822
##     Retained Not Retained     Retained     Retained     Retained Not
Retained
```

```
##          823          824          825          826          827
828
##     Retained     Retained Not Retained     Retained Not Retained
Retained
##          829          830          831          832          833
834
##     Retained Not Retained Not Retained Not Retained Not Retained
Retained
##          835          836          837          838          839
840
##     Retained     Retained     Retained Not Retained Not Retained
Retained
##          841          842          843          844          845
846
##     Retained Not Retained Not Retained     Retained Not Retained
Retained
##          847          848          849          850          851
852
## Not Retained Not Retained Not Retained     Retained Not Retained Not
Retained
##          853          854          855          856          857
858
##     Retained Not Retained     Retained     Retained     Retained
Retained
##          859          860          861          862          863
864
##     Retained Not Retained     Retained Not Retained     Retained Not
Retained
##          865          866          867          868          869
870
##     Retained     Retained     Retained     Retained     Retained
Retained
##          871          872          873          874          875
876
##     Retained Not Retained     Retained     Retained     Retained
Retained
##          877          878          879          880          881
882
##     Retained     Retained Not Retained Not Retained     Retained Not
Retained
##          883          884          885          886          887
888
##     Retained     Retained     Retained     Retained     Retained
Retained
##          889          890          891          892          893
894
## Not Retained Not Retained     Retained     Retained     Retained
Retained
##          895          896          897          898          899
900
```

```
##     Retained     Retained     Retained Not Retained     Retained 
Retained 
##          901          902          903          904          905 
906 
##     Retained     Retained Not Retained     Retained     Retained Not 
Retained 
##          907          908          909          910          911 
912 
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##          913          914          915          916          917 
918 
##     Retained     Retained     Retained Not Retained     Retained 
Retained 
##          919          920          921          922          923 
924 
##     Retained     Retained     Retained     Retained Not Retained 
Retained 
##          925          926          927          928          929 
930 
## Not Retained Not Retained Not Retained Not Retained     Retained Not 
Retained 
##          931          932          933          934          935 
936 
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##          937          938          939          940          941 
942 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          943          944          945          946          947 
948 
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##          949          950          951          952          953 
954 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##          955          956          957          958          959 
960 
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##          961          962          963          964          965 
966 
##     Retained     Retained Not Retained Not Retained Not Retained 
Retained 
##          967          968          969          970          971 
972 
##     Retained     Retained     Retained     Retained     Retained 
Retained
```

```
##          973          974          975          976          977
978
##     Retained     Retained     Retained     Retained Not Retained
Retained
##          979          980          981          982          983
984
##     Retained     Retained     Retained Not Retained     Retained
Retained
##          985          986          987          988          989
990
## Not Retained     Retained     Retained     Retained     Retained
Retained
##          991          992          993          994          995
996
##     Retained     Retained     Retained     Retained Not Retained
Retained
##          997          998          999         1000         1001
1002
##     Retained     Retained     Retained     Retained     Retained
Retained
##         1003         1004         1005         1006         1007
1008
##     Retained     Retained Not Retained     Retained Not Retained Not
Retained
##         1009         1010         1011         1012         1013
1014
##     Retained Not Retained     Retained     Retained Not Retained
Retained
##         1015         1016         1017         1018         1019
1020
##     Retained     Retained     Retained Not Retained     Retained
Retained
##         1021         1022         1023         1024         1025
1026
##     Retained     Retained     Retained     Retained Not Retained
Retained
##         1027         1028         1029         1030         1031
1032
##     Retained Not Retained     Retained     Retained Not Retained
Retained
##         1033         1034         1035         1036         1037
1038
##     Retained     Retained     Retained     Retained Not Retained
Retained
##         1039         1040         1041         1042         1043
1044
##     Retained     Retained     Retained Not Retained     Retained
Retained
##         1045         1046         1047         1048         1049
1050
```

```
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1051         1052         1053         1054         1055 
1056 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1057         1058         1059         1060         1061 
1062 
##     Retained     Retained     Retained     Retained Not Retained 
Retained 
##         1063         1064         1065         1066         1067 
1068 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1069         1070         1071         1072         1073 
1074 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1075         1076         1077         1078         1079 
1080 
## Not Retained Not Retained     Retained     Retained     Retained 
Retained 
##         1081         1082         1083         1084         1085 
1086 
## Not Retained     Retained     Retained     Retained Not Retained 
Retained 
##         1087         1088         1089         1090         1091 
1092 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1093         1094         1095         1096         1097 
1098 
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##         1099         1100         1101         1102         1103 
1104 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1105         1106         1107         1108         1109 
1110 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1111         1112         1113         1114         1115 
1116 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1117         1118         1119         1120         1121 
1122 
##     Retained     Retained     Retained     Retained     Retained Not 
Retained
```

```
##         1123         1124         1125         1126         1127
1128
##     Retained     Retained     Retained     Retained     Retained Not
Retained
##         1129         1130         1131         1132         1133
1134
##     Retained     Retained     Retained     Retained     Retained Not
Retained
##         1135         1136         1137         1138         1139
1140
##     Retained Not Retained     Retained     Retained     Retained
Retained
##         1141         1142         1143         1144         1145
1146
## Not Retained Not Retained     Retained     Retained     Retained
Retained
##         1147         1148         1149         1150         1151
1152
##     Retained     Retained     Retained     Retained     Retained
Retained
##         1153         1154         1155         1156         1157
1158
## Not Retained     Retained     Retained     Retained     Retained Not
Retained
##         1159         1160         1161         1162         1163
1164
##     Retained     Retained     Retained Not Retained     Retained
Retained
##         1165         1166         1167         1168         1169
1170
##     Retained     Retained     Retained     Retained     Retained
Retained
##         1171         1172         1173         1174         1175
1176
##     Retained     Retained     Retained Not Retained Not Retained
Retained
##         1177         1178         1179         1180         1181
1182
## Not Retained     Retained     Retained     Retained     Retained
Retained
##         1183         1184         1185         1186         1187
1188
##     Retained     Retained Not Retained Not Retained     Retained
Retained
##         1189         1190         1191         1192         1193
1194
##     Retained     Retained     Retained     Retained     Retained
Retained
##         1195         1196         1197         1198         1199
1200
```

```
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##         1201         1202         1203         1204         1205 
1206 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1207         1208         1209         1210         1211 
1212 
##     Retained Not Retained     Retained     Retained     Retained 
Retained 
##         1213         1214         1215         1216         1217 
1218 
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##         1219         1220         1221         1222         1223 
1224 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1225         1226         1227         1228         1229 
1230 
## Not Retained     Retained     Retained Not Retained Not Retained 
Retained 
##         1231         1232         1233         1234         1235 
1236 
##     Retained     Retained     Retained     Retained     Retained Not 
Retained 
##         1237         1238         1239         1240         1241 
1242 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1243         1244         1245         1246         1247 
1248 
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##         1249         1250         1251         1252         1253 
1254 
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##         1255         1256         1257         1258         1259 
1260 
##     Retained     Retained Not Retained     Retained     Retained 
Retained 
##         1261         1262         1263         1264         1265 
1266 
##     Retained     Retained     Retained Not Retained     Retained Not 
Retained 
##         1267         1268         1269         1270         1271 
1272 
##     Retained     Retained     Retained     Retained     Retained 
Retained
```

```
##         1273         1274         1275         1276         1277 
1278 
##     Retained Not Retained     Retained Not Retained     Retained 
Retained 
##         1279         1280         1281         1282         1283 
1284 
##     Retained     Retained     Retained Not Retained     Retained 
Retained 
##         1285         1286         1287         1288         1289 
1290 
##     Retained     Retained     Retained     Retained     Retained Not 
Retained 
##         1291         1292         1293         1294         1295 
1296 
## Not Retained     Retained     Retained Not Retained     Retained 
Retained 
##         1297         1298         1299         1300         1301 
1302 
## Not Retained Not Retained     Retained     Retained     Retained 
Retained 
##         1303         1304         1305         1306         1307 
1308 
##     Retained Not Retained     Retained     Retained Not Retained 
Retained 
##         1309         1310         1311         1312         1313 
1314 
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##         1315         1316         1317         1318         1319 
1320 
## Not Retained Not Retained     Retained Not Retained     Retained 
Retained 
##         1321         1322         1323         1324         1325 
1326 
##     Retained     Retained     Retained Not Retained Not Retained 
Retained 
##         1327         1328         1329         1330         1331 
1332 
##     Retained     Retained     Retained     Retained Not Retained 
Retained 
##         1333         1334         1335         1336         1337 
1338 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1339         1340         1341         1342         1343 
1344 
##     Retained Not Retained     Retained Not Retained     Retained 
Retained 
##         1345         1346         1347         1348         1349 
1350 
```

```
## Not Retained Not Retained     Retained     Retained     Retained 
Retained 
##         1351         1352         1353         1354         1355 
1356 
## Not Retained Not Retained     Retained Not Retained     Retained 
Retained 
##         1357         1358         1359         1360         1361 
1362 
##     Retained     Retained     Retained Not Retained     Retained Not 
Retained 
##         1363         1364         1365         1366         1367 
1368 
## Not Retained Not Retained     Retained Not Retained Not Retained 
Retained 
##         1369         1370         1371         1372         1373 
1374 
##     Retained     Retained     Retained Not Retained     Retained Not 
Retained 
##         1375         1376         1377         1378         1379 
1380 
##     Retained     Retained Not Retained Not Retained     Retained Not 
Retained 
##         1381         1382         1383         1384         1385 
1386 
##     Retained     Retained Not Retained     Retained     Retained Not 
Retained 
##         1387         1388         1389         1390         1391 
1392 
## Not Retained Not Retained Not Retained Not Retained Not Retained 
Retained 
##         1393         1394         1395         1396         1397 
1398 
##     Retained Not Retained     Retained     Retained     Retained 
Retained 
##         1399         1400         1401         1402         1403 
1404 
##     Retained     Retained     Retained     Retained Not Retained 
Retained 
##         1405         1406         1407         1408         1409 
1410 
##     Retained     Retained Not Retained     Retained Not Retained 
Retained 
##         1411         1412         1413         1414         1415 
1416 
##     Retained Not Retained Not Retained     Retained Not Retained Not 
Retained 
##         1417         1418         1419         1420         1421 
1422 
##     Retained Not Retained Not Retained     Retained Not Retained 
Retained
```

```
##         1423         1424         1425         1426         1427 
1428 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1429         1430         1431         1432         1433 
1434 
## Not Retained Not Retained     Retained     Retained     Retained 
Retained 
##         1435         1436         1437         1438         1439 
1440 
##     Retained     Retained     Retained     Retained Not Retained 
Retained 
##         1441         1442         1443         1444         1445 
1446 
##     Retained Not Retained Not Retained Not Retained Not Retained 
Retained 
##         1447         1448         1449         1450         1451 
1452 
##     Retained Not Retained Not Retained     Retained     Retained 
Retained 
##         1453         1454         1455         1456         1457 
1458 
##     Retained     Retained Not Retained     Retained Not Retained Not 
Retained 
##         1459         1460         1461         1462         1463 
1464 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1465         1466         1467         1468         1469 
1470 
## Not Retained Not Retained     Retained     Retained Not Retained 
Retained 
##         1471         1472         1473         1474         1475 
1476 
##     Retained Not Retained Not Retained Not Retained Not Retained Not 
Retained 
##         1477         1478         1479         1480         1481 
1482 
##     Retained Not Retained Not Retained     Retained Not Retained 
Retained 
##         1483         1484         1485         1486         1487 
1488 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1489         1490         1491         1492         1493 
1494 
##     Retained     Retained     Retained     Retained     Retained Not 
Retained 
##         1495         1496         1497         1498         1499 
1500 
```

```
##     Retained Not Retained     Retained Not Retained     Retained 
Retained 
##         1501         1502         1503         1504         1505 
1506 
##     Retained     Retained Not Retained Not Retained     Retained Not 
Retained 
##         1507         1508         1509         1510         1511 
1512 
##     Retained Not Retained Not Retained Not Retained Not Retained 
Retained 
##         1513         1514         1515         1516         1517 
1518 
## Not Retained Not Retained Not Retained Not Retained Not Retained Not 
Retained 
##         1519         1520         1521         1522         1523 
1524 
## Not Retained     Retained Not Retained     Retained     Retained 
Retained 
##         1525         1526         1527         1528         1529 
1530 
## Not Retained     Retained     Retained Not Retained     Retained Not 
Retained 
##         1531         1532         1533         1534         1535 
1536 
##     Retained Not Retained Not Retained Not Retained     Retained 
Retained 
##         1537         1538         1539         1540         1541 
1542 
## Not Retained Not Retained     Retained Not Retained     Retained Not 
Retained 
##         1543         1544         1545         1546         1547 
1548 
##     Retained     Retained     Retained Not Retained Not Retained Not 
Retained 
##         1549         1550         1551         1552         1553 
1554 
## Not Retained     Retained     Retained     Retained Not Retained 
Retained 
##         1555         1556         1557         1558         1559 
1560 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1561         1562         1563         1564         1565 
1566 
##     Retained     Retained     Retained     Retained Not Retained 
Retained 
##         1567         1568         1569         1570         1571 
1572 
##     Retained     Retained Not Retained     Retained     Retained Not 
Retained
```

```
##         1573         1574         1575         1576         1577
1578
## Not Retained     Retained     Retained Not Retained Not Retained
Retained
##         1579         1580         1581         1582         1583
1584
## Not Retained Not Retained     Retained     Retained     Retained
Retained
##         1585         1586         1587         1588         1589
1590
##     Retained     Retained     Retained     Retained     Retained
Retained
##         1591         1592         1593         1594         1595
1596
##     Retained Not Retained     Retained     Retained     Retained
Retained
##         1597         1598         1599         1600         1601
1602
##     Retained     Retained     Retained Not Retained Not Retained Not
Retained
##         1603         1604         1605         1606         1607
1608
##     Retained     Retained     Retained     Retained     Retained
Retained
##         1609         1610         1611         1612         1613
1614
## Not Retained Not Retained     Retained Not Retained Not Retained
Retained
##         1615         1616         1617         1618         1619
1620
## Not Retained Not Retained     Retained Not Retained Not Retained Not
Retained
##         1621         1622         1623         1624         1625
1626
##     Retained     Retained     Retained Not Retained Not Retained
Retained
##         1627         1628         1629         1630         1631
1632
##     Retained     Retained     Retained     Retained Not Retained
Retained
##         1633         1634         1635         1636         1637
1638
##     Retained Not Retained     Retained     Retained Not Retained Not
Retained
##         1639         1640         1641         1642         1643
1644
##     Retained     Retained     Retained     Retained Not Retained
Retained
##         1645         1646         1647         1648         1649
1650
```

```
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1651         1652         1653         1654         1655 
1656 
##     Retained     Retained Not Retained Not Retained Not Retained 
Retained 
##         1657         1658         1659         1660         1661 
1662 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1663         1664         1665         1666         1667 
1668 
##     Retained     Retained     Retained     Retained     Retained Not 
Retained 
##         1669         1670         1671         1672         1673 
1674 
##     Retained     Retained Not Retained Not Retained     Retained 
Retained 
##         1675         1676         1677         1678         1679 
1680 
## Not Retained Not Retained     Retained     Retained Not Retained 
Retained 
##         1681         1682         1683         1684         1685 
1686 
##     Retained Not Retained     Retained Not Retained     Retained 
Retained 
##         1687         1688         1689         1690         1691 
1692 
##     Retained Not Retained Not Retained     Retained     Retained 
Retained 
##         1693         1694         1695         1696         1697 
1698 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         1699         1700         1701         1702         1703 
1704 
## Not Retained Not Retained     Retained Not Retained     Retained Not 
Retained 
##         1705         1706         1707         1708         1709 
1710 
##     Retained Not Retained     Retained Not Retained Not Retained Not 
Retained 
##         1711         1712         1713         1714         1715 
1716 
## Not Retained     Retained Not Retained Not Retained Not Retained Not 
Retained 
##         1717         1718         1719         1720         1721 
1722 
##     Retained Not Retained Not Retained Not Retained     Retained Not 
Retained
```

```
##          1723          1724          1725          1726          1727 
1728 
##     Retained Not Retained Not Retained      Retained Not Retained 
Retained 
##          1729          1730          1731          1732          1733 
1734 
##     Retained Not Retained Not Retained Not Retained Not Retained Not 
Retained 
##          1735          1736          1737          1738          1739 
1740 
## Not Retained Not Retained      Retained      Retained Not Retained Not 
Retained 
##          1741          1742          1743          1744          1745 
1746 
## Not Retained      Retained      Retained      Retained Not Retained Not 
Retained 
##          1747          1748          1749          1750          1751 
1752 
##     Retained Not Retained      Retained      Retained      Retained 
Retained 
##          1753          1754          1755          1756          1757 
1758 
##     Retained      Retained      Retained      Retained Not Retained 
Retained 
##          1759          1760          1761          1762          1763 
1764 
##     Retained Not Retained Not Retained Not Retained      Retained Not 
Retained 
##          1765          1766          1767          1768          1769 
1770 
##     Retained Not Retained Not Retained Not Retained      Retained 
Retained 
##          1771          1772          1773          1774          1775 
1776 
## Not Retained      Retained      Retained      Retained      Retained Not 
Retained 
##          1777          1778          1779          1780          1781 
1782 
## Not Retained      Retained Not Retained Not Retained Not Retained 
Retained 
##          1783          1784          1785          1786          1787 
1788 
## Not Retained      Retained      Retained Not Retained      Retained 
Retained 
##          1789          1790          1791          1792          1793 
1794 
## Not Retained      Retained      Retained Not Retained      Retained 
Retained 
##          1795          1796          1797          1798          1799 
1800 
```

```
##     Retained     Retained     Retained Not Retained Not Retained 
Retained 
##         1801         1802         1803         1804         1805 
1806 
##     Retained     Retained Not Retained     Retained Not Retained 
Retained 
##         1807         1808         1809         1810         1811 
1812 
##     Retained     Retained     Retained Not Retained Not Retained Not 
Retained 
##         1813         1814         1815         1816         1817 
1818 
## Not Retained     Retained Not Retained Not Retained     Retained 
Retained 
##         1819         1820         1821         1822         1823 
1824 
## Not Retained     Retained     Retained Not Retained     Retained 
Retained 
##         1825         1826         1827         1828         1829 
1830 
##     Retained Not Retained     Retained Not Retained Not Retained 
Retained 
##         1831         1832         1833         1834         1835 
1836 
##     Retained     Retained Not Retained     Retained Not Retained 
Retained 
##         1837         1838         1839         1840         1841 
1842 
##     Retained Not Retained     Retained     Retained     Retained Not 
Retained 
##         1843         1844         1845         1846         1847 
1848 
##     Retained Not Retained Not Retained Not Retained     Retained Not 
Retained 
##         1849         1850         1851         1852         1853 
1854 
##     Retained Not Retained Not Retained Not Retained Not Retained 
Retained 
##         1855         1856         1857         1858         1859 
1860 
## Not Retained     Retained Not Retained Not Retained Not Retained Not 
Retained 
##         1861         1862         1863         1864         1865 
1866 
## Not Retained     Retained     Retained     Retained     Retained Not 
Retained 
##         1867         1868         1869         1870         1871 
1872 
## Not Retained Not Retained     Retained     Retained Not Retained 
Retained
```

```
##         1873         1874         1875         1876         1877 
1878 
##     Retained     Retained     Retained Not Retained Not Retained Not 
Retained 
##         1879         1880         1881         1882         1883 
1884 
##     Retained     Retained Not Retained     Retained     Retained 
Retained 
##         1885         1886         1887         1888         1889 
1890 
##     Retained     Retained     Retained     Retained Not Retained 
Retained 
##         1891         1892         1893         1894         1895 
1896 
##     Retained Not Retained     Retained     Retained Not Retained Not 
Retained 
##         1897         1898         1899         1900         1901 
1902 
## Not Retained     Retained     Retained     Retained     Retained 
Retained 
##         1903         1904         1905         1906         1907 
1908 
##     Retained     Retained     Retained     Retained Not Retained Not 
Retained 
##         1909         1910         1911         1912         1913 
1914 
##     Retained Not Retained     Retained Not Retained Not Retained 
Retained 
##         1915         1916         1917         1918         1919 
1920 
## Not Retained     Retained Not Retained Not Retained     Retained Not 
Retained 
##         1921         1922         1923         1924         1925 
1926 
##     Retained Not Retained     Retained Not Retained Not Retained Not 
Retained 
##         1927         1928         1929         1930         1931 
1932 
##     Retained     Retained     Retained Not Retained     Retained Not 
Retained 
##         1933         1934         1935         1936         1937 
1938 
## Not Retained Not Retained Not Retained Not Retained Not Retained 
Retained 
##         1939         1940         1941         1942         1943 
1944 
## Not Retained     Retained Not Retained     Retained     Retained 
Retained 
##         1945         1946         1947         1948         1949 
1950
```

```
## Not Retained Not Retained     Retained     Retained     Retained
Retained
##         1951         1952         1953         1954         1955
1956
## Not Retained     Retained     Retained     Retained Not Retained Not
Retained
##         1957         1958         1959         1960         1961
1962
##     Retained Not Retained     Retained Not Retained     Retained Not
Retained
##         1963         1964         1965         1966         1967
1968
## Not Retained     Retained     Retained     Retained     Retained
Retained
##         1969         1970         1971         1972         1973
1974
##     Retained     Retained     Retained Not Retained Not Retained Not
Retained
##         1975         1976         1977         1978         1979
1980
##     Retained Not Retained Not Retained Not Retained Not Retained
Retained
##         1981         1982         1983         1984         1985
1986
##     Retained Not Retained Not Retained     Retained     Retained Not
Retained
##         1987         1988         1989         1990         1991
1992
## Not Retained     Retained Not Retained     Retained Not Retained
Retained
##         1993         1994         1995         1996         1997
1998
##     Retained Not Retained Not Retained Not Retained Not Retained
Retained
##         1999         2000         2001         2002         2003
2004
##     Retained Not Retained     Retained     Retained     Retained
Retained
##         2005         2006         2007         2008         2009
2010
##     Retained     Retained Not Retained     Retained Not Retained
Retained
##         2011         2012         2013         2014         2015
2016
##     Retained Not Retained Not Retained     Retained     Retained Not
Retained
##         2017         2018         2019         2020         2021
2022
##     Retained     Retained     Retained     Retained     Retained
Retained
```

```
##         2023         2024         2025         2026         2027 
2028 
##     Retained     Retained     Retained     Retained     Retained Not 
Retained 
##         2029         2030         2031         2032         2033 
2034 
## Not Retained Not Retained     Retained     Retained Not Retained 
Retained 
##         2035         2036         2037         2038         2039 
2040 
## Not Retained     Retained Not Retained     Retained     Retained 
Retained 
##         2041         2042         2043         2044         2045 
2046 
##     Retained Not Retained Not Retained     Retained     Retained Not 
Retained 
##         2047         2048         2049         2050         2051 
2052 
##     Retained Not Retained Not Retained     Retained Not Retained Not 
Retained 
##         2053         2054         2055         2056         2057 
2058 
## Not Retained     Retained Not Retained Not Retained Not Retained 
Retained 
##         2059         2060         2061         2062         2063 
2064 
## Not Retained Not Retained Not Retained     Retained     Retained 
Retained 
##         2065         2066         2067         2068         2069 
2070 
## Not Retained     Retained Not Retained     Retained     Retained 
Retained 
##         2071         2072         2073         2074         2075 
2076 
## Not Retained     Retained     Retained     Retained     Retained Not 
Retained 
##         2077         2078         2079         2080         2081 
2082 
##     Retained     Retained Not Retained     Retained Not Retained Not 
Retained 
##         2083         2084         2085         2086         2087 
2088 
##     Retained     Retained Not Retained Not Retained Not Retained 
Retained 
##         2089         2090         2091         2092         2093 
2094 
## Not Retained     Retained     Retained     Retained Not Retained Not 
Retained 
##         2095         2096         2097         2098         2099 
2100 
```

```
##     Retained     Retained     Retained Not Retained     Retained 
Retained 
##         2101         2102         2103         2104         2105 
2106 
##     Retained     Retained Not Retained     Retained Not Retained Not 
Retained 
##         2107         2108         2109         2110         2111 
2112 
##     Retained Not Retained Not Retained Not Retained Not Retained Not 
Retained 
##         2113         2114         2115         2116         2117 
2118 
## Not Retained     Retained     Retained Not Retained Not Retained 
Retained 
##         2119         2120         2121         2122         2123 
2124 
## Not Retained Not Retained Not Retained Not Retained Not Retained Not 
Retained 
##         2125         2126         2127         2128         2129 
2130 
## Not Retained Not Retained     Retained Not Retained     Retained 
Retained 
##         2131         2132         2133         2134         2135 
2136 
##     Retained     Retained Not Retained     Retained     Retained Not 
Retained 
##         2137         2138         2139         2140         2141 
2142 
##     Retained Not Retained     Retained     Retained     Retained 
Retained 
##         2143         2144         2145         2146         2147 
2148 
##     Retained     Retained Not Retained     Retained Not Retained Not 
Retained 
##         2149         2150         2151         2152         2153 
2154 
##     Retained Not Retained Not Retained Not Retained     Retained 
Retained 
##         2155         2156         2157         2158         2159 
2160 
## Not Retained     Retained     Retained Not Retained Not Retained 
Retained 
##         2161         2162         2163         2164         2165 
2166 
##     Retained Not Retained Not Retained     Retained     Retained 
Retained 
##         2167         2168         2169         2170         2171 
2172 
## Not Retained     Retained     Retained Not Retained Not Retained Not 
Retained
```

```
##         2173         2174         2175         2176         2177 
2178 
##     Retained     Retained     Retained     Retained     Retained Not 
Retained 
##         2179         2180         2181         2182         2183 
2184 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         2185         2186         2187         2188         2189 
2190 
## Not Retained     Retained Not Retained     Retained     Retained 
Retained 
##         2191         2192         2193         2194         2195 
2196 
##     Retained Not Retained Not Retained Not Retained Not Retained Not 
Retained 
##         2197         2198         2199         2200         2201 
2202 
## Not Retained     Retained Not Retained     Retained     Retained 
Retained 
##         2203         2204         2205         2206         2207 
2208 
##     Retained     Retained     Retained     Retained     Retained 
Retained 
##         2209         2210         2211         2212         2213 
2214 
##     Retained     Retained Not Retained     Retained     Retained 
Retained 
##         2215         2216         2217         2218         2219 
2220 
## Not Retained Not Retained     Retained     Retained     Retained 
Retained 
##         2221         2222         2223         2224         2225 
2226 
## Not Retained Not Retained Not Retained     Retained Not Retained 
Retained 
##         2227         2228         2229         2230         2231 
2232 
## Not Retained     Retained Not Retained     Retained     Retained Not 
Retained 
##         2233         2234         2235         2236         2237 
2238 
##     Retained     Retained     Retained     Retained Not Retained 
Retained 
##         2239         2240         2241         2242         2243 
2244 
## Not Retained Not Retained Not Retained     Retained     Retained 
Retained 
##         2245         2246         2247         2248         2249 
2250 
```

```
## Not Retained Not Retained Not Retained     Retained     Retained Not
Retained
##         2251         2252         2253         2254         2255
2256
##     Retained Not Retained Not Retained Not Retained Not Retained
Retained
##         2257         2258         2259         2260         2261
2262
## Not Retained     Retained Not Retained     Retained     Retained Not
Retained
##         2263         2264         2265         2266         2267
2268
## Not Retained     Retained     Retained     Retained Not Retained
Retained
##         2269         2270         2271         2272         2273
2274
##     Retained Not Retained Not Retained Not Retained     Retained
Retained
##         2275         2276         2277         2278         2279
2280
## Not Retained Not Retained     Retained     Retained     Retained Not
Retained
##         2281         2282         2283         2284         2285
2286
## Not Retained Not Retained Not Retained Not Retained     Retained
Retained
##         2287         2288         2289         2290         2291
2292
##     Retained Not Retained Not Retained Not Retained Not Retained
Retained
##         2293         2294         2295         2296         2297
2298
## Not Retained Not Retained     Retained Not Retained Not Retained Not
Retained
##         2299         2300         2301         2302         2303
2304
## Not Retained Not Retained Not Retained Not Retained Not Retained
Retained
##         2305         2306         2307         2308         2309
2310
## Not Retained     Retained Not Retained Not Retained Not Retained Not
Retained
##         2311         2312         2313         2314         2315
2316
##     Retained Not Retained Not Retained Not Retained     Retained Not
Retained
##         2317         2318         2319         2320         2321
2322
##     Retained Not Retained     Retained Not Retained Not Retained
Retained
```

```
##         2323         2324         2325         2326         2327
2328
##     Retained Not Retained Not Retained Not Retained Not Retained
Retained
##         2329         2330         2331         2332         2333
2334
## Not Retained Not Retained     Retained Not Retained     Retained Not
Retained
##         2335         2336         2337         2338         2339
2340
##     Retained Not Retained Not Retained     Retained     Retained Not
Retained
##         2341         2342         2343         2344         2345
2346
##     Retained     Retained Not Retained Not Retained Not Retained Not
Retained
##         2347         2348         2349         2350         2351
2352
##     Retained     Retained     Retained     Retained Not Retained
Retained
##         2353         2354         2355         2356         2357
2358
##     Retained Not Retained     Retained     Retained Not Retained Not
Retained
##         2359         2360         2361         2362         2363
2364
## Not Retained Not Retained Not Retained     Retained     Retained Not
Retained
##         2365         2366         2367         2368         2369
2370
## Not Retained     Retained Not Retained     Retained Not Retained
Retained
##         2371         2372         2373         2374         2375
2376
## Not Retained Not Retained     Retained Not Retained Not Retained Not
Retained
##         2377         2378         2379         2380         2381
2382
## Not Retained     Retained     Retained Not Retained Not Retained
Retained
##         2383         2384         2385         2386         2387
2388
## Not Retained Not Retained     Retained     Retained Not Retained Not
Retained
##         2389
##     Retained
## Levels: Not Retained Retained

# Confusion matrix
Confusion_Matrix_Random <- table(rf$predicted,
random_forest_data$Retained.in.2012., dnn = c("Predicted", "Actual"))
Confusion_Matrix_Random
```

```
##                Actual
## Predicted      Not Retained Retained
##   Not Retained          627      178
##   Retained              311     1273
```

```
library(caret)
confusionMatrix(rf$predicted, random_forest_data$Retained.in.2012., positive
= "Retained")
```

```
## Confusion Matrix and Statistics
## 
##                Reference
## Prediction      Not Retained Retained
##   Not Retained          627      178
##   Retained              311     1273
##                                           
##                Accuracy : 0.7953          
##                  95% CI : (0.7786, 0.8113)
##     No Information Rate : 0.6074          
##     P-Value [Acc > NIR] : < 2.2e-16       
##                                           
##                   Kappa : 0.5598          
##                                           
##  Mcnemar's Test P-Value : 2.384e-09       
##                                           
##             Sensitivity : 0.8773          
##             Specificity : 0.6684          
##          Pos Pred Value : 0.8037          
##          Neg Pred Value : 0.7789          
##              Prevalence : 0.6074          
##          Detection Rate : 0.5329          
##    Detection Prevalence : 0.6630          
##       Balanced Accuracy : 0.7729          
##                                           
##        'Positive' Class : Retained        
## 
```

```
## The Sensitivity(postive results out of which are actually postive) is
0.8773 ~ 87% (which is a good percentage) and Specificity and 0.6684 ~ 67%
and the accuracy is 79% with the postive class as "Retained"
```

### We can say that using the random forest model, we are getting an OOB error rate of 0.2046882 ~ 20% and accuracy of 79%

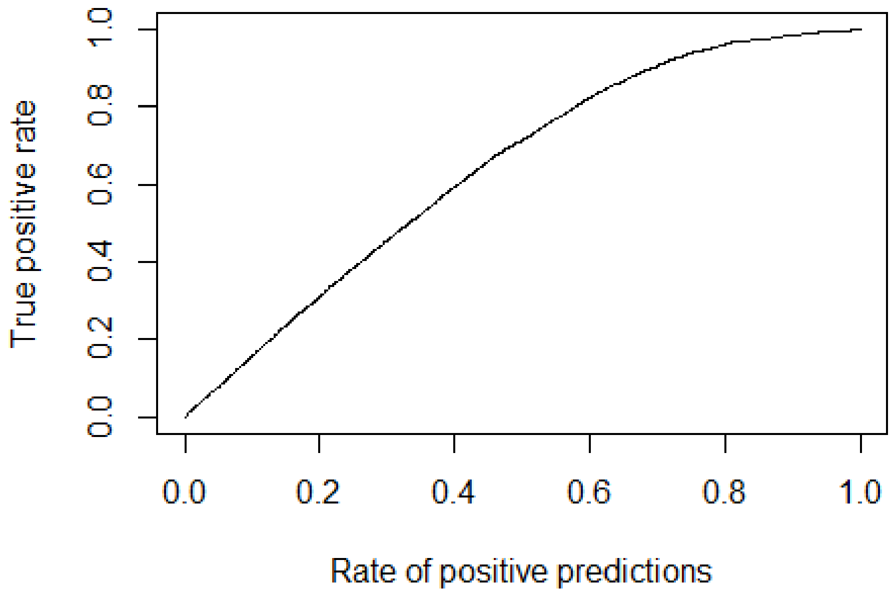
## Drawing evaluation charts

```
library(ROCR)
pred <- prediction(rf$votes[, 2],random_forest_data$Retained.in.2012.)
```

## Gain Chart

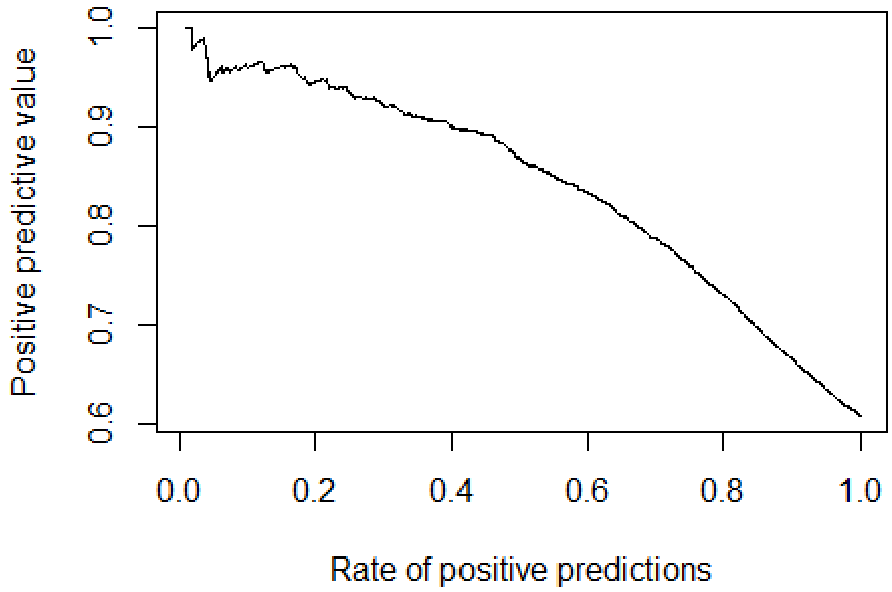
###Gain chart presents the percentage of captured positive responses as a function of selected percentage of a sample. ####Which is actually in our case

```
perf <- performance(pred, "tpr", "rpp")
plot(perf)
```



## Response Chart

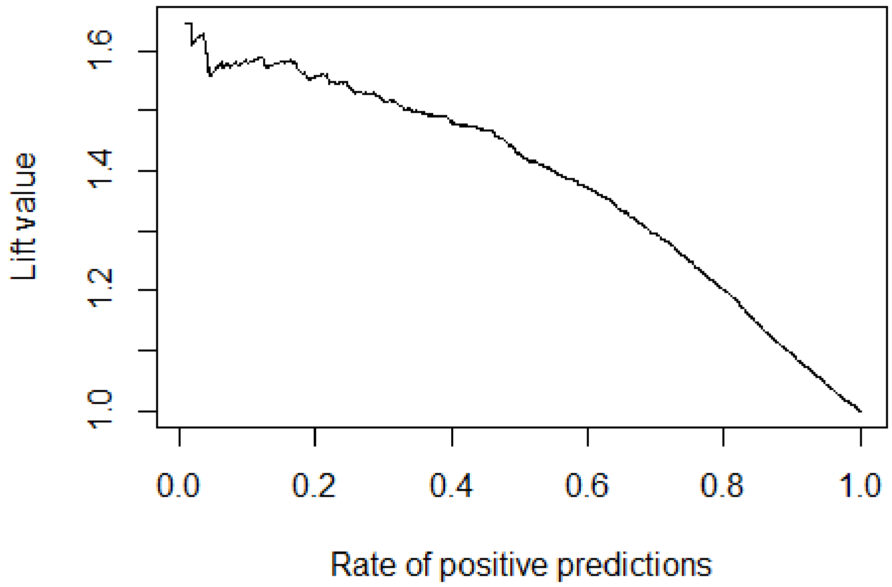
```
perf <- performance(pred, "ppv", "rpp")
plot(perf)
```



## Lift Chart

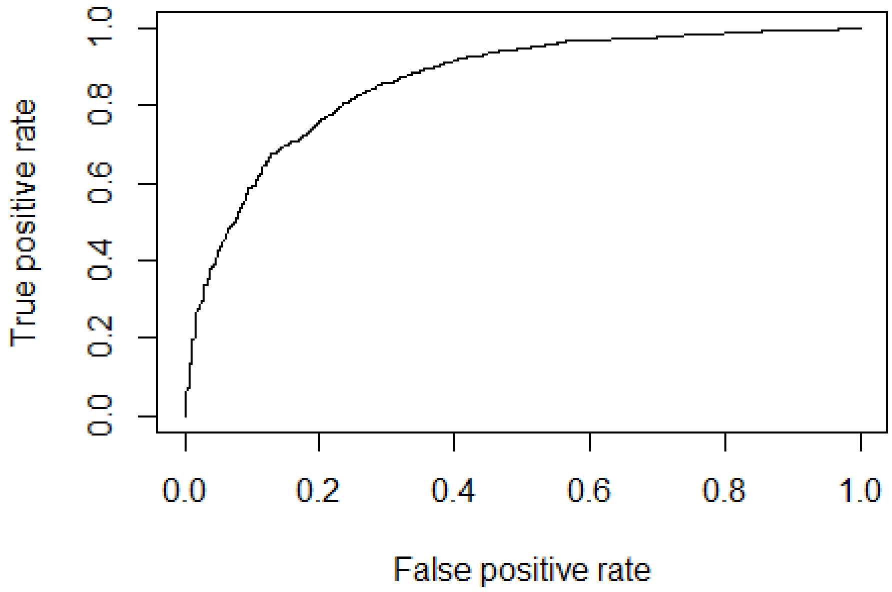
###The lift chart measures effectiveness of our predictive classification model comparing it with the baseline model.

```
perf <- performance(pred, "lift", "rpp")
plot(perf)
```



### ROC Curve - We can conclude that we have a smaller false alarm and also has higehr recall,captures more retained(positve)

```
perf <- performance(pred, "tpr", "fpr")
plot(perf)
```



## auc

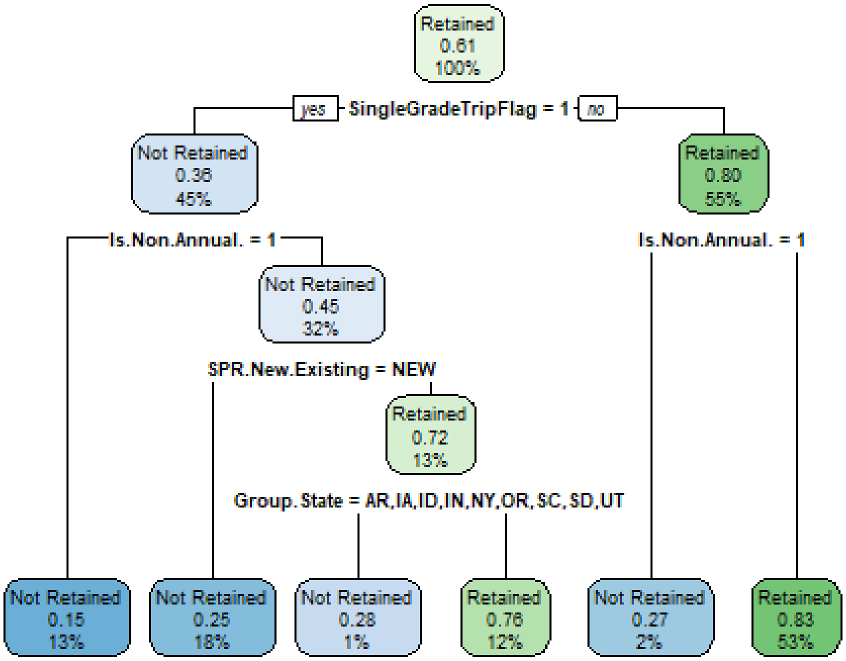
##Since the AUC is 0.86 and the graph clearly shows the the model is accurate and a good model

```
auc <- performance(pred, "auc")
auc

## A performance instance
##   'Area under the ROC curve'

auc <- unlist(slot(auc, "y.values"))
auc

## [1] 0.8616949
```

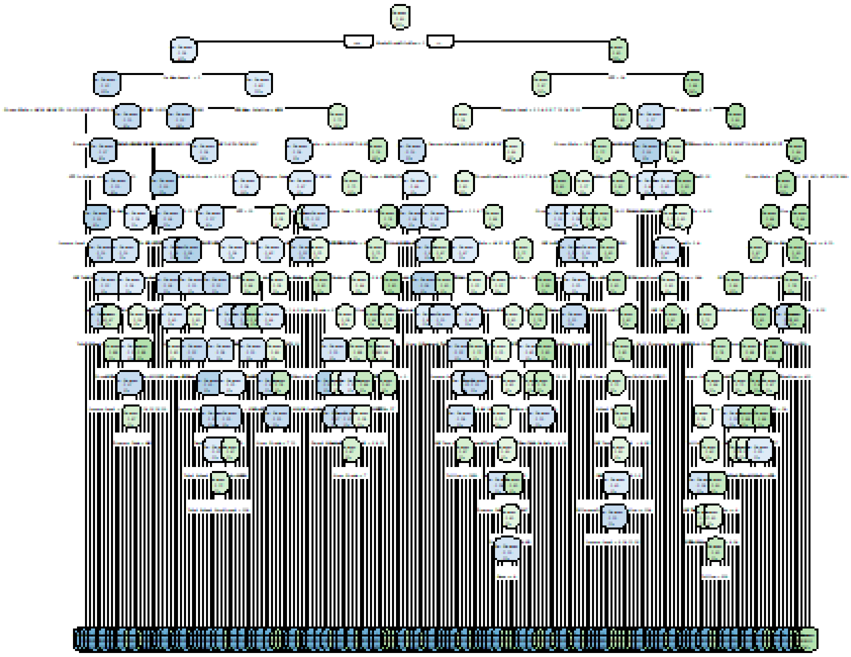
##Constructing the decision Tree

```
sum(is.na(dataset))

## [1] 0

# decision_tree_data <- subset(dataset,select=-
c(School.Type,FPP.to.School.enrollment,DifferenceTraveltoFirstMeeting,Differe
nceTraveltoLastMeeting,Parent.Meeting.Flag,NumberOfMeetingswithParents,School
GradeType,Days,GroupGradeType,Group.State,SchoolSizeIndicator))
decision_tree_data <- dataset
```

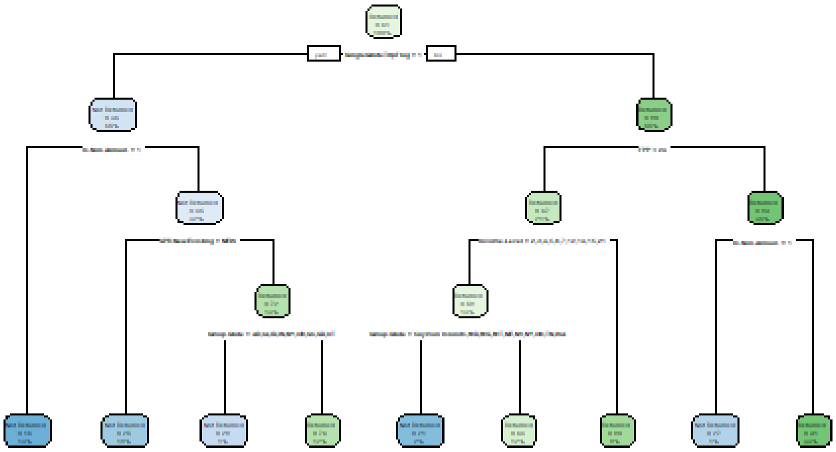
```
set.seed(134)
indx <- sample(2, nrow(decision_tree_data), replace = TRUE, prob =
c(0.8,0.2))
train <- decision_tree_data[indx == 1, ]
test <- decision_tree_data[indx == 2, ]

#Ratio of the train and test data size
nrow(train)/nrow(test) #-> 1925/464

## [1] 4.104701

#Constructing the tree:
mytree <-rpart(Retained.in.2012. ~ ., data = train, method = 'class')

print(mytree)

## n= 1921
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
##  1) root 1921 754 Retained (0.3925039 0.6074961)
##    2) SingleGradeTripFlag=1 855 310 Not Retained (0.6374269 0.3625731)
##      4) Is.Non.Annual.=1 244  36 Not Retained (0.8524590 0.1475410) *
##      5) Is.Non.Annual.=0 611 274 Not Retained (0.5515548 0.4484452)
##       10) SPR.New.Existing=NEW 352  88 Not Retained (0.7500000 0.2500000)
*
##       11) SPR.New.Existing=EXISTING 259  73 Retained (0.2818533 0.7181467)
##         22) Group.State=AR,IA,ID,IN,NY,OR,SC,SD,UT 25   7 Not Retained
(0.7200000 0.2800000) *
##         23)
Group.State=AL,AZ,CA,CO,CT,FL,IL,KS,LA,MA,MD,MI,MN,MO,MS,MT,NC,ND,NE,NH,NM,NV
,OH,OK,TN,TX,VA,WA,WI 234  55 Retained (0.2350427 0.7649573) *
##    3) SingleGradeTripFlag=2 1066 209 Retained (0.1960600 0.8039400)
##      6) Is.Non.Annual.=1 48  13 Not Retained (0.7291667 0.2708333) *
##      7) Is.Non.Annual.=0 1018 174 Retained (0.1709234 0.8290766) *

#The percentage of number of Retained is greater than the number of people
not retained
prop.table(table(decision_tree_data$Retained.in.2012.))

##
## Not Retained     Retained
##    0.3926329    0.6073671

rpart.plot(mytree)
```



```
#Constructing the Full decision tree
tree_model2 <- rpart(Retained.in.2012. ~ ., train, parms = list(split =
"information"), control = rpart.control(minbucket = 0,minsplit = 0,cp = -1))

#As we can see the entire tree does not give us a full information, lets tune
the hyperparameters in the rpart.control
rpart.plot(tree_model2)

## Warning: labs do not fit even at cex 0.15, there may be some overplotting
```



```
#Examining the complexity of the plot

printcp(tree_model2)

##
## Classification tree:
## rpart(formula = Retained.in.2012. ~ ., data = train, parms = list(split =
"information"),
##     control = rpart.control(minbucket = 0, minsplit = 0, cp = -1))
##
## Variables actually used in tree construction:
##  [1] Cancelled.Pax                CRM.Segment
##  [3] Days                         DepartureMonth
##  [5] DifferenceTraveltoFirstMeeting DifferenceTraveltoLastMeeting
##  [7] EZ.Pay.Take.Up.Rate          FPP
##  [9] FPP.to.PAX                   FPP.to.School.enrollment
## [11] From.Grade                   FRP.Active
## [13] FRP.Cancelled                FRP.Take.up.percent.
## [15] Group.State                  GroupGradeType
## [17] Income.Level                 Is.Non.Annual.
## [19] MDR.High.Grade               MDR.Low.Grade
## [21] Parent.Meeting.Flag          Poverty.Code
## [23] Program.Code                 Region
## [25] School.Sponsor               School.Type
## [27] SchoolSizeIndicator          SingleGradeTripFlag
## [29] SPR.New.Existing             To.Grade
## [31] Total.Discount.Pax           Total.Pax
## [33] Total.School.Enrollment        Tuition
##
## Root node error: 754/1921 = 0.3925
##
## n= 1921
##
##               CP nsplit rel error  xerror     xstd
## 1   0.31167109      0 1.0000000 1.00000 0.028385
## 2   0.07493369      1 0.6883289 0.68833 0.025812
## 3   0.01458886      3 0.5384615 0.53846 0.023732
## 4   0.00961538      4 0.5238727 0.54907 0.023901
## 5   0.00795756      8 0.4854111 0.55703 0.024026
## 6   0.00663130     10 0.4694960 0.57692 0.024329
## 7   0.00563660     13 0.4496021 0.57162 0.024249
## 8   0.00530504     17 0.4270557 0.57427 0.024289
## 9   0.00464191     18 0.4217507 0.57692 0.024329
## 10  0.00397878     20 0.4124668 0.59019 0.024524
## 11  0.00353669     24 0.3965517 0.59682 0.024619
## 12  0.00331565     36 0.3262599 0.61671 0.024898
## 13  0.00298408     46 0.2904509 0.62069 0.024953
## 14  0.00265252     55 0.2559682 0.63793 0.025184
## 15  0.00221043     78 0.1896552 0.65119 0.025356
## 16  0.00198939     85 0.1697613 0.64456 0.025270
## 17  0.00176835     99 0.1419098 0.65385 0.025389
## 18  0.00132626    108 0.1233422 0.66180 0.025490
## 19  0.00088417    149 0.0649867 0.70955 0.026057
## 20  0.00066313    194 0.0026525 0.72546 0.026233
## 21 -1.00000000    198 0.0000000 0.72546 0.026233

#As we can see the root node train error is 0.39 ~ 40%
#As we can see the if the cp value is 0.00928382, we are getting an xerror of
0.53846

tree_model3 <- rpart(Retained.in.2012. ~ ., train, method = "class", parms =
list(split = "information"), control = rpart.control(minsplit = 3,cp =
0.00928382))

 #Tuning the Hyperparamters
tree_model4 <- rpart(myFormula, train, parms = list(split = "information"),
control = rpart.control(minbucket = 5,minsplit = 3,cp = 0.00928382))
rpart.plot(tree_model4)
```



```
printcp(tree_model4)

##
## Classification tree:
## rpart(formula = myFormula, data = train, parms = list(split =
"information"),
##     control = rpart.control(minbucket = 5, minsplit = 3, cp = 0.00928382))
##
## Variables actually used in tree construction:
## [1] FPP                  Group.State          Income.Level
## [4] Is.Non.Annual.       SingleGradeTripFlag SPR.New.Existing
##
## Root node error: 754/1921 = 0.3925
##
## n= 1921
##
##          CP nsplit rel error  xerror     xstd
## 1 0.3116711      0   1.00000 1.00000 0.028385
## 2 0.0749337      1   0.68833 0.68833 0.025812
## 3 0.0145889      3   0.53846 0.53846 0.023732
## 4 0.0096154      4   0.52387 0.55438 0.023985
## 5 0.0092838      8   0.48541 0.55040 0.023922

#Train Error:
predTrain1 <- predict(tree_model4, data = train, type = 'class')
trainError <- mean(predTrain1 != train$Retained.in.2012.)
trainError <- mean(train$Retained.in.2012.!=predTrain1)
```

```
#The train data error is estimated to be 46%
trainError
```

```
## [1] 0.1905258
```

```
# #Confusion Matrix:
# # Building the confusion matrix
#
# confu_matr <- table(train$Retained.in.2012., predTrain1)
# confu_matr
#
# #Accuracy of the Model Train data
# #Accuracy of the Model Train data
# #For the Accuracy, the success rate or the accuracy of the model can be easily calculated:
#
# acc_Test <- sum(diag(confu_matr)) / sum(confu_matr)
# acc_Test
#
# #Recall of the model
# rec_matr<-confu_matr[2,2]/(confu_matr[2,1]+confu_matr[2,2])
# rec_matr
#
# #Precision of the model
# prec_matr <- confu_matr[2,2]/(confu_matr[1,2]+confu_matr[2,2])
# prec_matr
#
# printcp(mytree)

#TestError:

predTest <- predict(tree_model4, newdata = test, type='class')
testError <- mean(test$Retained.in.2012. == predTest)

#Test Error is 56.7%
testError
```

```
## [1] 0.8034188
```

```
#
#
# Confusion_Matrix_Function <- function(actualValues, predictedValues)
# {
#   funcMatrix <- table(actual = actualValues, pred = predictedValues)
#   print(funcMatrix)
#   TN <- funcMatrix[1,1]
#   FP <- funcMatrix[1,2]
#   FN <- funcMatrix[2,1]
#   TP <- funcMatrix[2,2]
#   Sensitivity <- TP/(TP + FN)
#   Specificity <- TN/(TN + FP)
#   Precision <- TP/(TP + FP)
#   print(paste("Sensitivity = ", round(Sensitivity, 4)))
#   print(paste("Specificity = ", round(Specificity, 4)))
#   print(paste("Precision = ", round(Precision, 4)))
# }
#
# length(test$Retained.in.2012.)
# length(predTest)
#
# Confusion_Matrix_Function(test$Retained.in.2012.,predTest)
```

###Recommendations:

###1. Focusing on Prpgram.Code like HT, HS, HD all are which History programs that runs in states like Texas ###2.Focusing on target areas whose parent have higher income levels ###3. Focusing on metroplitan areas like California, Texas, Washington and Illinois ###4. Our machine learning model (random forest) with an accuracy of around 79% conclude that having to continue the programs specified above might increase the retained rate in the year 2013 ###5.From this we can conclude the Income.Level, Is.Non.Annual and SPR.NewExisting and Total.PAX are of higher importance, so we should focus to get more retained rate