**Analysis of Sales Report of a Clothes Manufacturing Outlet**

This is an R HTML document. When you click the **Knit HTML** button a web page will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

**install.packages**("readxl")

*## Installing package into '/usr/local/lib/R/site-library'*

*## (as 'lib' is unspecified)*

**## Warning in install.packages("readxl"): 'lib = "/usr/local/lib/R/site-library"'**

**## is not writable**

**## Error in install.packages("readxl"): unable to install packages**

**install.packages**("plyr")

*## Installing package into '/usr/local/lib/R/site-library'*

*## (as 'lib' is unspecified)*

**## Warning in install.packages("plyr"): 'lib = "/usr/local/lib/R/site-library"' is**

**## not writable**

**## Error in install.packages("plyr"): unable to install packages**

**install.packages**("dplyr")

*## Installing package into '/usr/local/lib/R/site-library'*

*## (as 'lib' is unspecified)*

**## Warning in install.packages("dplyr"): 'lib = "/usr/local/lib/R/site-library"' is**

**## not writable**

**## Error in install.packages("dplyr"): unable to install packages**

**install.packages**("caTools")

*## Installing package into '/usr/local/lib/R/site-library'*

*## (as 'lib' is unspecified)*

**## Warning in install.packages("caTools"): 'lib = "/usr/local/lib/R/site-library"'**

**## is not writable**

**## Error in install.packages("caTools"): unable to install packages**

**install.packages**("e1071")

*## Installing package into '/usr/local/lib/R/site-library'*

*## (as 'lib' is unspecified)*

**## Warning in install.packages("e1071"): 'lib = "/usr/local/lib/R/site-library"' is**

**## not writable**

**## Error in install.packages("e1071"): unable to install packages**

**install.packages**("caret")

*## Installing package into '/usr/local/lib/R/site-library'*

*## (as 'lib' is unspecified)*

**## Warning in install.packages("caret"): 'lib = "/usr/local/lib/R/site-library"' is**

**## not writable**

**## Error in install.packages("caret"): unable to install packages**

**install.packages**("randomForest")

*## Installing package into '/usr/local/lib/R/site-library'*

*## (as 'lib' is unspecified)*

**## Warning in install.packages("randomForest"): 'lib = "/usr/local/lib/R/site-**

**## library"' is not writable**

**## Error in install.packages("randomForest"): unable to install packages**

**install.packages**("pscl")

*## Installing package into '/usr/local/lib/R/site-library'*

*## (as 'lib' is unspecified)*

**## Warning in install.packages("pscl"): 'lib = "/usr/local/lib/R/site-library"' is**

**## not writable**

**## Error in install.packages("pscl"): unable to install packages**

*# import necessary libraries*

**library**(readxl)

*# to read excel*

**library**(plyr)

**library**(dplyr)

*##*

*## Attaching package: 'dplyr'*

*## The following objects are masked from 'package:plyr':*

*##*

*## arrange, count, desc, failwith, id, mutate, rename, summarise,*

*## summarize*

*## The following objects are masked from 'package:stats':*

*##*

*## filter, lag*

*## The following objects are masked from 'package:base':*

*##*

*## intersect, setdiff, setequal, union*

**library**(caTools)

**library**(e1071)

**library**(caret)

*## Loading required package: lattice*

*## Loading required package: ggplot2*

**library**(randomForest)

*## randomForest 4.6-14*

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

*##*

*## Attaching package: 'randomForest'*

*## The following object is masked from 'package:ggplot2':*

*##*

*## margin*

*## The following object is masked from 'package:dplyr':*

*##*

*## combine*

**library**(ggplot2)

**library**(lattice)

**library**(readxl)

attrib1 <- **read\_excel**("Attribute DataSet.xlsx")

**View**(attrib1)

**## Warning in View(attrib1): unable to open display**

**## Error in .External2(C\_dataviewer, x, title): unable to start data viewer**

**library**(readxl)

dresssale1 <- **read\_excel**("Dress Sales.xlsx")

**View**(dresssale1)

**## Warning in View(dresssale1): unable to open display**

**## Error in .External2(C\_dataviewer, x, title): unable to start data viewer**

attrib1 = **read\_excel**('Attribute DataSet.xlsx')

dresssale1 = **read\_excel**('Dress Sales.xlsx')

*#remove Dress\_ID column*

attrib2 = attrib1[2:14]

dresssale2 = dresssale1[2:24]

*# check the unique values for each columns*

**lapply**(attrib2, unique)

## $Style

## [1] "Sexy" "Casual" "vintage" "Brief" "cute" "bohemian"

## [7] "Novelty" "Flare" "party" "sexy" "work" "OL"

## [13] "fashion"

##

## $Price

## [1] "Low" "High" "Average" "Medium" "very-high" "low"

## [7] "high" NA

##

## $Rating

## [1] 4.6 0.0 4.5 5.0 4.7 4.8 4.3 4.0 4.4 4.9 4.2 3.6 3.7 4.1 3.5 1.0 3.0

##

## $Size

## [1] "M" "L" "XL" "free" "S" "small" "s"

##

## $Season

## [1] "Summer" "Automn" "Spring" "Winter" "spring" "winter" NA "summer"

## [9] "Autumn"

##

## $NeckLine

## [1] "o-neck" "v-neck" "boat-neck" "peterpan-collor"

## [5] "ruffled" "turndowncollor" "slash-neck" "mandarin-collor"

## [9] "open" "sqare-collor" "Sweetheart" "sweetheart"

## [13] NA "Scoop" "halter" "backless"

## [17] "bowneck" "NULL"

##

## $SleeveLength

## [1] "sleevless" "Petal" "full" "butterfly"

## [5] "short" "threequarter" "halfsleeve" "cap-sleeves"

## [9] "turndowncollor" "threequater" "capsleeves" "sleeveless"

## [13] "sleeevless" "half" "urndowncollor" "thressqatar"

## [17] "NULL" "sleveless"

##

## $waiseline

## [1] "empire" "natural" "null" NA "princess" "dropped"

##

## $Material

## [1] "null" "microfiber" "polyster" "silk"

## [5] "chiffonfabric" "cotton" "nylon" "other"

## [9] "milksilk" "linen" "rayon" "lycra"

## [13] "mix" "acrylic" "spandex" "lace"

## [17] "modal" "cashmere" NA "viscos"

## [21] "knitting" "sill" "wool" "model"

## [25] "shiffon"

##

## $FabricType

## [1] "chiffon" "null" "broadcloth" "jersey" "other"

## [6] "batik" "satin" "flannael" "worsted" "woolen"

## [11] "poplin" "dobby" "knitting" "flannel" "tulle"

## [16] "sattin" NA "organza" "lace" "Corduroy"

## [21] "wollen" "knitted" "shiffon" "terry"

##

## $Decoration

## [1] "ruffles" "null" "embroidary" "bow" "lace"

## [6] "beading" "sashes" "hollowout" "pockets" "sequined"

## [11] "applique" "button" "Tiered" "rivet" "feathers"

## [16] "flowers" "pearls" NA "pleat" "crystal"

## [21] "ruched" "draped" "tassel" "plain" "none"

## [26] "cascading"

##

## $`Pattern Type`

## [1] "animal" "print" "dot" "solid" "null" "patchwork"

## [7] "striped" "geometric" "plaid" "leopard" "floral" "character"

## [13] NA "splice" "leapord" "none"

##

## $Recommendation

## [1] 1 0

*# values checking*

attrib2$Style<-**mapvalues**(attrib2$Style, **c**('sexy'), **c**('Sexy'))

attrib2$Price<-**mapvalues**(attrib2$Price, **c**('low','high'), **c**('Low','High'))

attrib2$Size<-**mapvalues**(attrib2$Size, **c**('s','small'), **c**('S','S'))

attrib2$Season<-**mapvalues**(attrib2$Season, **c**('spring','summer','Automn','winter'), **c**('Spring','Summer','Autumn','Winter'))

attrib2$NeckLine<-**mapvalues**(attrib2$NeckLine, **c**('sweetheart'), **c**('Sweetheart'))

attrib2$SleeveLength<-**mapvalues**(attrib2$SleeveLength, **c**('sleevless','sleeevless','sleveless','threequater','thressqatar','urndowncollor'), **c**('sleeveless','sleeveless','sleeveless','threequarter','threequarter','turndowncollar'))

attrib2$FabricType<-**mapvalues**(attrib2$FabricType, **c**('shiffon','sattin','wollen','flannael','knitting'), **c**('chiffon','satin','woolen','flannel','knitted'))

attrib2$Decoration<-**mapvalues**(attrib2$Decoration, **c**('embroidary','sequined','ruched','none'), **c**('embroidery','sequins','ruche','null'))

attrib2$`Pattern Type`<-**mapvalues**(attrib2$`Pattern Type`, **c**('none','leapord'), **c**('null','leopard'))

*# factoring*

**sapply**(attrib2,class)

## Style Price Rating Size Season

## "character" "character" "numeric" "character" "character"

## NeckLine SleeveLength waiseline Material FabricType

## "character" "character" "character" "character" "character"

## Decoration Pattern Type Recommendation

## "character" "character" "numeric"

cols<-**c**("Style","Price","Size","Season","NeckLine","SleeveLength","waiseline","Material",

"FabricType","Decoration","Pattern Type","Recommendation")

attrib2[cols]<-**lapply**(attrib2[cols],factor)

**sapply**(attrib2,class)

## Style Price Rating Size Season

## "factor" "factor" "numeric" "factor" "factor"

## NeckLine SleeveLength waiseline Material FabricType

## "factor" "factor" "factor" "factor" "factor"

## Decoration Pattern Type Recommendation

## "factor" "factor" "factor"

*# OR ###*

*#attrib2$Style <- factor(attrib2$Style,levels=unique(attrib2$Style))*

*#table(attrib2$Style)*

*#attrib2$Price <- factor(attrib2$Price,levels=unique(attrib2$Price))*

*#summary(attrib2$Price)*

*#attrib2$Recommendation <- sapply(attrib2$Recommendation, factor)*

*# count of missing values in attrib2 dataset*

**colSums**(**is.na**(attrib2))

## Style Price Rating Size Season

## 0 2 0 0 2

## NeckLine SleeveLength waiseline Material FabricType

## 1 0 1 1 1

## Decoration Pattern Type Recommendation

## 1 1 0

*# fill missing Value with mode # Categorical data. So, no use of mean & median. So, use Mode. # The mode is the value that has highest number of occurrences in a set of data. # R does not have a standard in-built function to calculate mode. So we create a user function to calculate mode of a data set in R.*

getmode <- **function**(v) {

uniqv <- **unique**(v)

uniqv[**which.max**(**tabulate**(**match**(v, uniqv)))]

}

*# fill missing Value with mode*

attrib2$Price[**is.na**(attrib2$Price) ==TRUE] <- **getmode**(attrib2$Price)

attrib2$Season[**is.na**(attrib2$Season) ==TRUE] <- **getmode**(attrib2$Season)

attrib2$NeckLine[**is.na**(attrib2$NeckLine) ==TRUE] <- **getmode**(attrib2$NeckLine)

attrib2$waiseline[**is.na**(attrib2$waiseline) ==TRUE] <- **getmode**(attrib2$waiseline)

attrib2$Material[**is.na**(attrib2$Material) ==TRUE] <- **getmode**(attrib2$Material)

attrib2$FabricType[**is.na**(attrib2$FabricType) ==TRUE] <- **getmode**(attrib2$FabricType)

attrib2$Decoration[**is.na**(attrib2$Decoration) ==TRUE] <- **getmode**(attrib2$Decoration)

attrib2$`Pattern Type`[**is.na**(attrib2$`Pattern Type`) ==TRUE] <- **getmode**(attrib2$`Pattern Type`)

attrib2data <- **data.frame**(attrib2)

**str**(attrib2data)

## 'data.frame': 500 obs. of 13 variables:

## $ Style : Factor w/ 12 levels "bohemian","Brief",..: 10 3 11 2 4 1 3 7 6 1 ...

## $ Price : Factor w/ 5 levels "Average","High",..: 3 3 2 1 3 3 1 1 1 3 ...

## $ Rating : num 4.6 0 0 4.6 4.5 0 0 0 0 0 ...

## $ Size : Factor w/ 5 levels "free","L","M",..: 3 2 2 2 3 3 5 1 1 1 ...

## $ Season : Factor w/ 4 levels "Autumn","Spring",..: 3 3 1 2 3 3 3 1 2 3 ...

## $ NeckLine : Factor w/ 16 levels "backless","boat-neck",..: 7 7 7 7 7 16 7 7 16 16 ...

## $ SleeveLength : Factor w/ 13 levels "butterfly","cap-sleeves",..: 10 8 4 4 1 10 4 9 9 10 ...

## $ waiseline : Factor w/ 5 levels "dropped","empire",..: 2 3 3 3 3 2 4 3 2 3 ...

## $ Material : Factor w/ 24 levels "acrylic","cashmere",..: 14 9 17 20 3 14 4 17 4 15 ...

## $ FabricType : Factor w/ 18 levels "batik","broadcloth",..: 3 10 10 3 3 10 10 2 2 3 ...

## $ Decoration : Factor w/ 24 levels "applique","beading",..: 20 20 13 8 3 13 13 12 2 13 ...

## $ Pattern.Type : Factor w/ 13 levels "animal","character",..: 1 1 10 10 3 10 11 7 11 7 ...

## $ Recommendation: Factor w/ 2 levels "0","1": 2 1 1 2 1 1 1 1 2 2 ...

*#Dresses dataset*

**head**(dresssale2)

## # A tibble: 6 x 23

## `29/8/2013` `31/8/2013` `41314` `41373` `41434` `41495` `41556` `41617`

## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>

## 1 2114 2274 2491 2660 2727 2887 2930 3119

## 2 151 275 570 750 813 1066 1164 1558

## 3 6 7 7 7 8 8 9 10

## 4 1005 1128 1326 1455 1507 1621 1637 1723

## 5 996 1175 1304 1396 1432 1559 1570 1638

## 6 4 5 11 13 13 13 16 18

## # … with 15 more variables: `14/9/2013` <chr>, `16/9/2013` <chr>,

## # `18/9/2013` <chr>, `20/9/2013` <chr>, `22/9/2013` <chr>, `24/9/2013` <dbl>,

## # `26/9/2013` <dbl>, `28/9/2013` <dbl>, `30/9/2013` <dbl>, `41315` <dbl>,

## # `41374` <dbl>, `41435` <dbl>, `40400` <dbl>, `41557` <dbl>, `41618` <dbl>

**names**(dresssale2)

## [1] "29/8/2013" "31/8/2013" "41314" "41373" "41434" "41495"

## [7] "41556" "41617" "14/9/2013" "16/9/2013" "18/9/2013" "20/9/2013"

## [13] "22/9/2013" "24/9/2013" "26/9/2013" "28/9/2013" "30/9/2013" "41315"

## [19] "41374" "41435" "40400" "41557" "41618"

*# Update columns name in dresssale2 dataset*

*# Change any column names you want to, all at once*

**colnames**(dresssale2)[**colnames**(dresssale2) %in% **c**("41314","41373","41434","41495","41556","41617","41315","41374","41435","40400","41557","41618")] <- **c**("2/9/2013","4/9/2013", "6/9/2013","8/9/2013","10/9/2013","12/9/2013","2/10/2013","4/10/2013","6/10/2013","8/10/2013","10/10/2013","12/10/2013")

*# names(dresssale2) <- c("29/8/2013", "31/8/2013", "2/9/2013", "41373" , "41434" , "41495" , "41556" , "41617",*

*# "14/9/2013", "16/9/2013" ,"18/9/2013", "20/9/2013", "22/9/2013", "24/9/2013", "26/9/2013", "28/9/2013",*

*# "30/9/2013", "41315", "41374", "41435", "40400" , "41557" , "41618")*

**head**(dresssale2)

## # A tibble: 6 x 23

## `29/8/2013` `31/8/2013` `2/9/2013` `4/9/2013` `6/9/2013` `8/9/2013`

## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>

## 1 2114 2274 2491 2660 2727 2887

## 2 151 275 570 750 813 1066

## 3 6 7 7 7 8 8

## 4 1005 1128 1326 1455 1507 1621

## 5 996 1175 1304 1396 1432 1559

## 6 4 5 11 13 13 13

## # … with 17 more variables: `10/9/2013` <dbl>, `12/9/2013` <chr>,

## # `14/9/2013` <chr>, `16/9/2013` <chr>, `18/9/2013` <chr>, `20/9/2013` <chr>,

## # `22/9/2013` <chr>, `24/9/2013` <dbl>, `26/9/2013` <dbl>, `28/9/2013` <dbl>,

## # `30/9/2013` <dbl>, `2/10/2013` <dbl>, `4/10/2013` <dbl>, `6/10/2013` <dbl>,

## # `8/10/2013` <dbl>, `10/10/2013` <dbl>, `12/10/2013` <dbl>

*# Convert all variable types to numeric*

dresssale2 <- **as.data.frame**(**apply**(dresssale2, 2, as.numeric))

**## Warning in apply(dresssale2, 2, as.numeric): NAs introduced by coercion**

**## Warning in apply(dresssale2, 2, as.numeric): NAs introduced by coercion**

**## Warning in apply(dresssale2, 2, as.numeric): NAs introduced by coercion**

**## Warning in apply(dresssale2, 2, as.numeric): NAs introduced by coercion**

**## Warning in apply(dresssale2, 2, as.numeric): NAs introduced by coercion**

**## Warning in apply(dresssale2, 2, as.numeric): NAs introduced by coercion**

*# mean row*

dresssale2 = **as.matrix**(dresssale2)

k <- **which**(**is.na**(dresssale2), arr.ind=TRUE)

dresssale2[k] <- **rowMeans**(dresssale2, na.rm=TRUE)[k[,1]]

dresssale2 = **as.data.frame**(dresssale2)

*# sum all values on row on (total sales column)-New column created*

dresssale2$total\_sales = **rowSums**(dresssale2)

**head**(dresssale2)

## 29/8/2013 31/8/2013 2/9/2013 4/9/2013 6/9/2013 8/9/2013 10/9/2013 12/9/2013

## 1 2114 2274 2491 2660 2727 2887 2930 3119

## 2 151 275 570 750 813 1066 1164 1558

## 3 6 7 7 7 8 8 9 10

## 4 1005 1128 1326 1455 1507 1621 1637 1723

## 5 996 1175 1304 1396 1432 1559 1570 1638

## 6 4 5 11 13 13 13 16 18

## 14/9/2013 16/9/2013 18/9/2013 20/9/2013 22/9/2013 24/9/2013 26/9/2013

## 1 3204 3277 3321 3386 3479 3554 3624

## 2 1756 1878 1985 2106 2454 2710 2942

## 3 10 10 10 10 11 11 11

## 4 1746 1783 1796 1812 1845 1878 1892

## 5 1655 1681 1743 1824 1919 2032 2156

## 6 19 20 20 21 22 25 25

## 28/9/2013 30/9/2013 2/10/2013 4/10/2013 6/10/2013 8/10/2013 10/10/2013

## 1 3706 3746 3795 3832 3897 3923 3985

## 2 3258 3354 3475 3654 3911 4024 4125

## 3 11 11 11 11 11 11 11

## 4 1914 1924 1929 1941 1952 1955 1959

## 5 2252 2312 2387 2459 2544 2614 2693

## 6 26 26 26 26 27 27 27

## 12/10/2013 total\_sales

## 1 4048 75979

## 2 4277 52256

## 3 11 223

## 4 1963 39691

## 5 2736 44077

## 6 27 457

*#Merged data*

merged\_data <- **data.frame**(attrib2 ,dresssale2)

**head**(merged\_data)

## Style Price Rating Size Season NeckLine SleeveLength waiseline

## 1 Sexy Low 4.6 M Summer o-neck sleeveless empire

## 2 Casual Low 0.0 L Summer o-neck Petal natural

## 3 vintage High 0.0 L Autumn o-neck full natural

## 4 Brief Average 4.6 L Spring o-neck full natural

## 5 cute Low 4.5 M Summer o-neck butterfly natural

## 6 bohemian Low 0.0 M Summer v-neck sleeveless empire

## Material FabricType Decoration Pattern.Type Recommendation X29.8.2013

## 1 null chiffon ruffles animal 1 2114

## 2 microfiber null ruffles animal 0 151

## 3 polyster null null print 0 6

## 4 silk chiffon embroidery print 1 1005

## 5 chiffonfabric chiffon bow dot 0 996

## 6 null null null print 0 4

## X31.8.2013 X2.9.2013 X4.9.2013 X6.9.2013 X8.9.2013 X10.9.2013 X12.9.2013

## 1 2274 2491 2660 2727 2887 2930 3119

## 2 275 570 750 813 1066 1164 1558

## 3 7 7 7 8 8 9 10

## 4 1128 1326 1455 1507 1621 1637 1723

## 5 1175 1304 1396 1432 1559 1570 1638

## 6 5 11 13 13 13 16 18

## X14.9.2013 X16.9.2013 X18.9.2013 X20.9.2013 X22.9.2013 X24.9.2013 X26.9.2013

## 1 3204 3277 3321 3386 3479 3554 3624

## 2 1756 1878 1985 2106 2454 2710 2942

## 3 10 10 10 10 11 11 11

## 4 1746 1783 1796 1812 1845 1878 1892

## 5 1655 1681 1743 1824 1919 2032 2156

## 6 19 20 20 21 22 25 25

## X28.9.2013 X30.9.2013 X2.10.2013 X4.10.2013 X6.10.2013 X8.10.2013 X10.10.2013

## 1 3706 3746 3795 3832 3897 3923 3985

## 2 3258 3354 3475 3654 3911 4024 4125

## 3 11 11 11 11 11 11 11

## 4 1914 1924 1929 1941 1952 1955 1959

## 5 2252 2312 2387 2459 2544 2614 2693

## 6 26 26 26 26 27 27 27

## X12.10.2013 total\_sales

## 1 4048 75979

## 2 4277 52256

## 3 11 223

## 4 1963 39691

## 5 2736 44077

## 6 27 457

**str**(merged\_data)

## 'data.frame': 500 obs. of 37 variables:

## $ Style : Factor w/ 12 levels "bohemian","Brief",..: 10 3 11 2 4 1 3 7 6 1 ...

## $ Price : Factor w/ 5 levels "Average","High",..: 3 3 2 1 3 3 1 1 1 3 ...

## $ Rating : num 4.6 0 0 4.6 4.5 0 0 0 0 0 ...

## $ Size : Factor w/ 5 levels "free","L","M",..: 3 2 2 2 3 3 5 1 1 1 ...

## $ Season : Factor w/ 4 levels "Autumn","Spring",..: 3 3 1 2 3 3 3 1 2 3 ...

## $ NeckLine : Factor w/ 16 levels "backless","boat-neck",..: 7 7 7 7 7 16 7 7 16 16 ...

## $ SleeveLength : Factor w/ 13 levels "butterfly","cap-sleeves",..: 10 8 4 4 1 10 4 9 9 10 ...

## $ waiseline : Factor w/ 5 levels "dropped","empire",..: 2 3 3 3 3 2 4 3 2 3 ...

## $ Material : Factor w/ 24 levels "acrylic","cashmere",..: 14 9 17 20 3 14 4 17 4 15 ...

## $ FabricType : Factor w/ 18 levels "batik","broadcloth",..: 3 10 10 3 3 10 10 2 2 3 ...

## $ Decoration : Factor w/ 24 levels "applique","beading",..: 20 20 13 8 3 13 13 12 2 13 ...

## $ Pattern.Type : Factor w/ 13 levels "animal","character",..: 1 1 10 10 3 10 11 7 11 7 ...

## $ Recommendation: Factor w/ 2 levels "0","1": 2 1 1 2 1 1 1 1 2 2 ...

## $ X29.8.2013 : num 2114 151 6 1005 996 ...

## $ X31.8.2013 : num 2274 275 7 1128 1175 ...

## $ X2.9.2013 : num 2491 570 7 1326 1304 ...

## $ X4.9.2013 : num 2660 750 7 1455 1396 ...

## $ X6.9.2013 : num 2727 813 8 1507 1432 ...

## $ X8.9.2013 : num 2887 1066 8 1621 1559 ...

## $ X10.9.2013 : num 2930 1164 9 1637 1570 ...

## $ X12.9.2013 : num 3119 1558 10 1723 1638 ...

## $ X14.9.2013 : num 3204 1756 10 1746 1655 ...

## $ X16.9.2013 : num 3277 1878 10 1783 1681 ...

## $ X18.9.2013 : num 3321 1985 10 1796 1743 ...

## $ X20.9.2013 : num 3386 2106 10 1812 1824 ...

## $ X22.9.2013 : num 3479 2454 11 1845 1919 ...

## $ X24.9.2013 : num 3554 2710 11 1878 2032 ...

## $ X26.9.2013 : num 3624 2942 11 1892 2156 ...

## $ X28.9.2013 : num 3706 3258 11 1914 2252 ...

## $ X30.9.2013 : num 3746 3354 11 1924 2312 ...

## $ X2.10.2013 : num 3795 3475 11 1929 2387 ...

## $ X4.10.2013 : num 3832 3654 11 1941 2459 ...

## $ X6.10.2013 : num 3897 3911 11 1952 2544 ...

## $ X8.10.2013 : num 3923 4024 11 1955 2614 ...

## $ X10.10.2013 : num 3985 4125 11 1959 2693 ...

## $ X12.10.2013 : num 4048 4277 11 1963 2736 ...

## $ total\_sales : num 75979 52256 223 39691 44077 ...

*# spliting dataset*

**set.seed**(100)

spl = **sample.split**(merged\_data$Recommendation, SplitRatio = 0.7)

train = **subset**(merged\_data, spl==TRUE)

test = **subset**(merged\_data, spl==FALSE)

**print**(**dim**(train));

## [1] 350 37

**print**(**dim**(test))

## [1] 150 37

*#Classification - Predict recommendation*

*#First model (Naive Bayes): # non-linear model # simple & fast*

**options**(scipen = 999)

naive\_model = **naiveBayes**(Recommendation ~.,data = train) *# build model # . means all column*

**confusionMatrix**(train$Recommendation, **predict**(naive\_model,train), positive = '1') *# create confusion Matrix*

## Confusion Matrix and Statistics

##

## Reference

## Prediction 0 1

## 0 136 67

## 1 41 106

##

## Accuracy : 0.6914

## 95% CI : (0.6401, 0.7394)

## No Information Rate : 0.5057

## P-Value [Acc > NIR] : 0.000000000001409

##

## Kappa : 0.3817

##

## Mcnemar's Test P-Value : 0.01614

##

## Sensitivity : 0.6127

## Specificity : 0.7684

## Pos Pred Value : 0.7211

## Neg Pred Value : 0.6700

## Prevalence : 0.4943

## Detection Rate : 0.3029

## Detection Prevalence : 0.4200

## Balanced Accuracy : 0.6905

##

## 'Positive' Class : 1

##

naive\_predict = **predict**(naive\_model,test) *# predict test set*

**table**(naive\_predict,test$Recommendation) *# create table*

##

## naive\_predict 0 1

## 0 50 30

## 1 37 33

*# Support vector machine (SVM): # Linear model # complex*

svm\_model = **svm**(Recommendation ~.,train) *# build model*

**confusionMatrix**(train$Recommendation,**predict**(svm\_model),positive = '1') *# create confusion Matrix*

## Confusion Matrix and Statistics

##

## Reference

## Prediction 0 1

## 0 203 0

## 1 141 6

##

## Accuracy : 0.5971

## 95% CI : (0.5437, 0.6489)

## No Information Rate : 0.9829

## P-Value [Acc > NIR] : 1

##

## Kappa : 0.047

##

## Mcnemar's Test P-Value : <0.0000000000000002

##

## Sensitivity : 1.00000

## Specificity : 0.59012

## Pos Pred Value : 0.04082

## Neg Pred Value : 1.00000

## Prevalence : 0.01714

## Detection Rate : 0.01714

## Detection Prevalence : 0.42000

## Balanced Accuracy : 0.79506

##

## 'Positive' Class : 1

##

svm\_predict = **predict**(svm\_model,test) *# predict test set*

**table**(svm\_predict,test$Recommendation) *# create table*

##

## svm\_predict 0 1

## 0 85 61

## 1 2 2

*# Third model (Random Forest)*

randomForest\_model = **randomForest**(x = train, y = train$Recommendation,ntree =800) *# build model*

**confusionMatrix**(train$Recommendation,**predict**(randomForest\_model),positive = '1') *# create confusion Matrix*

## Confusion Matrix and Statistics

##

## Reference

## Prediction 0 1

## 0 203 0

## 1 0 147

##

## Accuracy : 1

## 95% CI : (0.9895, 1)

## No Information Rate : 0.58

## P-Value [Acc > NIR] : < 0.00000000000000022

##

## Kappa : 1

##

## Mcnemar's Test P-Value : NA

##

## Sensitivity : 1.00

## Specificity : 1.00

## Pos Pred Value : 1.00

## Neg Pred Value : 1.00

## Prevalence : 0.42

## Detection Rate : 0.42

## Detection Prevalence : 0.42

## Balanced Accuracy : 1.00

##

## 'Positive' Class : 1

##

randomForest\_predict = **predict**(randomForest\_model,test) *# predict test set*

**table**(randomForest\_predict,test$Recommendation ) *# create table*

##

## randomForest\_predict 0 1

## 0 87 0

## 1 0 63

You can also embed plots, for example: