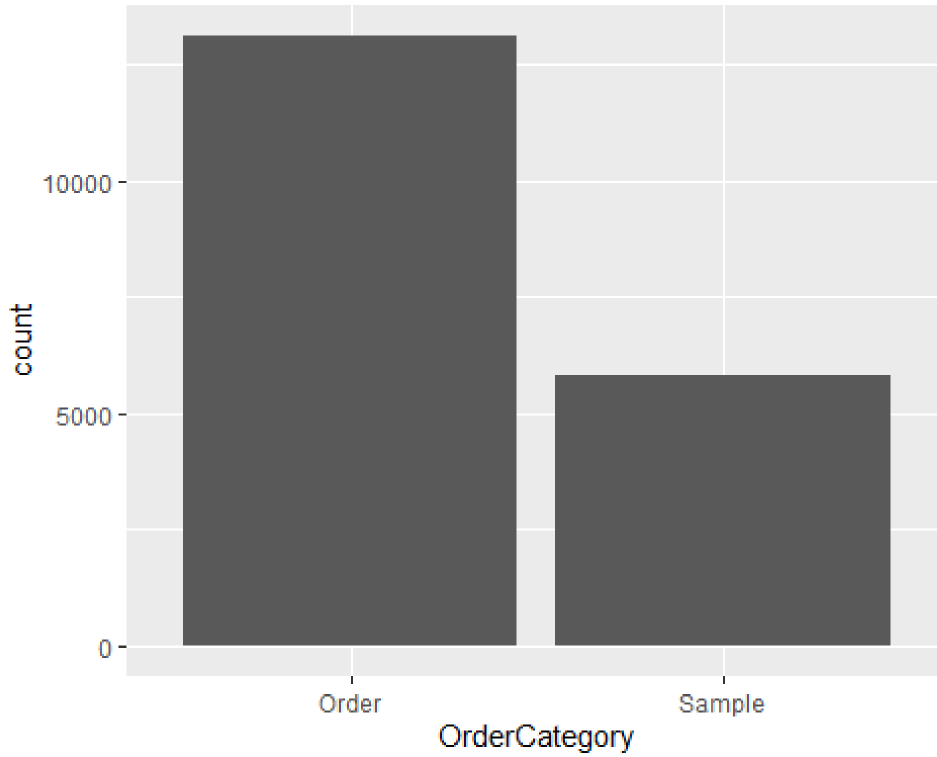
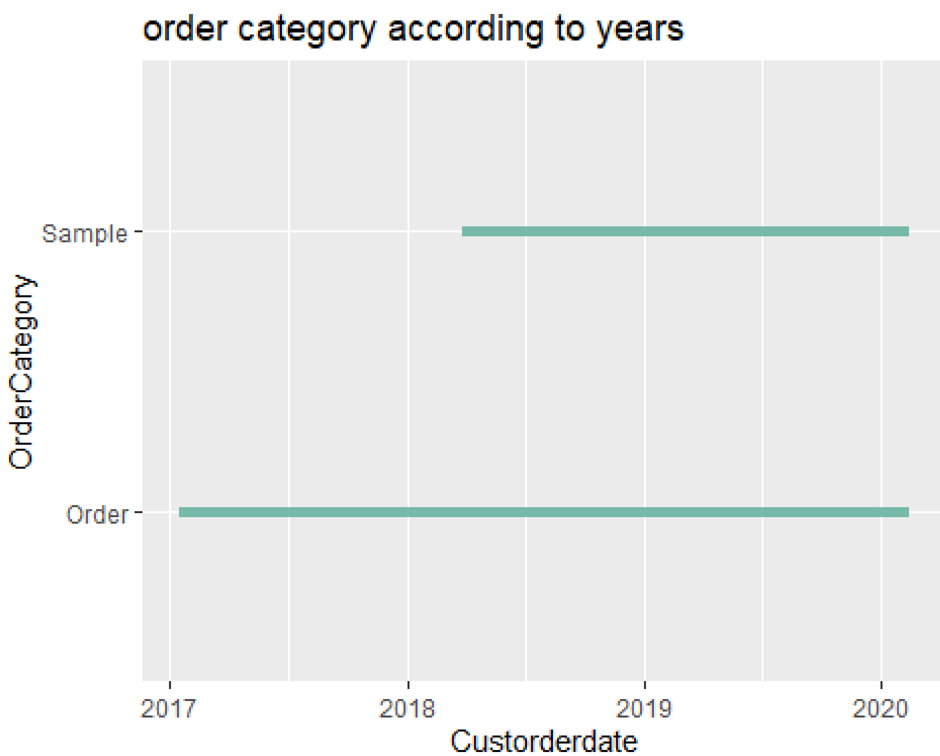
CHAMPO CARPETS: PERFORMING EDA ON DATASET (DERIVING INSIGHTS AND CLEANING)

```
CC <- read_excel("C:/Users/Bodake/Desktop/UIC Sem 2/Data Mining IDS-
572/R/champo.xlsx")
View(CC)

mydata <- CC
mydata <- drop_na(CC)
str(mydata)

## tibble [18,946 x 16] (S3: tbl_df/tbl/data.frame)
##  $ OrderType      : chr [1:18946] "Area Wise" "Area Wise" "Area Wise"
"Area Wise" ...
##  $ OrderCategory  : chr [1:18946] "Order" "Order" "Order" "Order" ...
##  $ CustomerCode   : chr [1:18946] "H-1" "H-1" "H-1" "H-1" ...
##  $ CountryName    : chr [1:18946] "USA" "USA" "USA" "USA" ...
##  $ CustomerOrderNo: chr [1:18946] "1873354" "1873354" "1873354" "1918436"
...
##  $ Custorderdate  : POSIXct[1:18946], format: "2017-01-16" "2017-01-16"
...
##  $ UnitName       : chr [1:18946] "Ft" "Ft" "Ft" "Ft" ...
##  $ QtyRequired    : num [1:18946] 2 2 2 5 5 4 6 16 2 4 ...
##  $ TotalArea      : num [1:18946] 6 9 54 54 71.2 ...
##  $ Amount         : num [1:18946] 12 18 108 270 356 ...
##  $ ITEM_NAME      : chr [1:18946] "HAND TUFTED" "HAND TUFTED" "HAND
TUFTED" "HAND TUFTED" ...
##  $ QualityName    : chr [1:18946] "TUFTED 30C HARD TWIST" "TUFTED 30C HARD
TWIST" "TUFTED 30C HARD TWIST" "TUFTED 30C HARD TWIST" ...
##  $ DesignName     : chr [1:18946] "OLD LONDON [3715]" "OLD LONDON [3715]"
"OLD LONDON [3715]" "OLD LONDON [3715]" ...
##  $ ColorName      : chr [1:18946] "BEIGE" "BEIGE" "BEIGE" "BEIGE" ...
##  $ ShapeName      : chr [1:18946] "REC" "REC" "REC" "REC" ...
##  $ AreaFt         : num [1:18946] 6 9 54 54 71.2 ...

count(mydata)

## # A tibble: 1 x 1
##       n
##   <int>
## 1 18946

ggplot(data=mydata) + geom_bar(mapping=aes(x=OrderCategory))
```



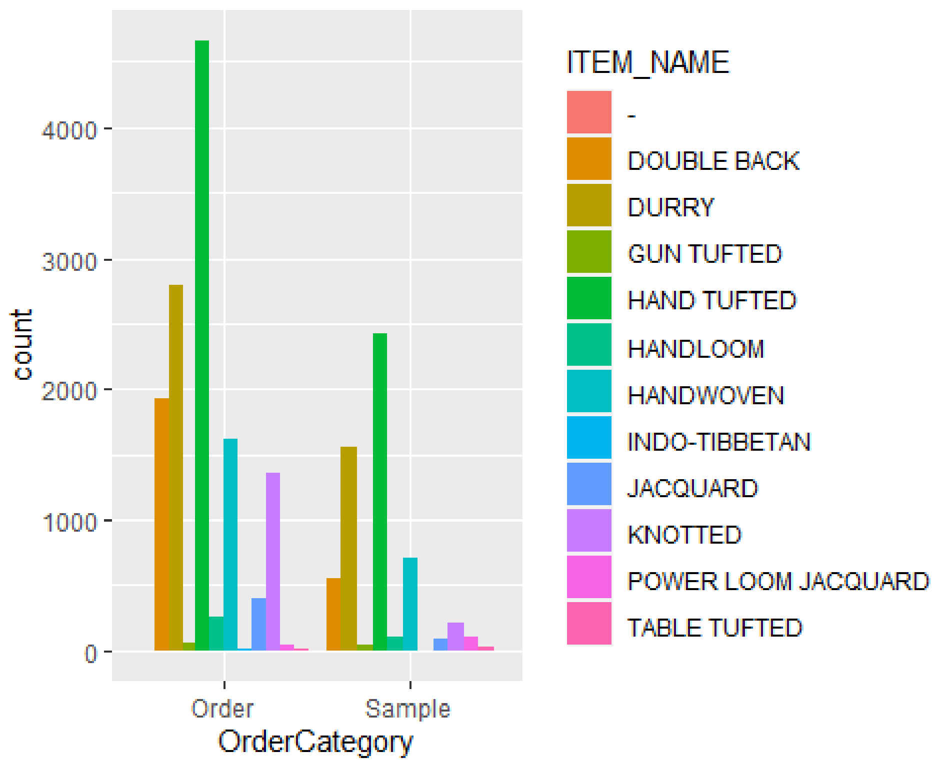
Here, number of 'Orders' in `Order Category` is higher than 'Sample' category.

```
ggplot(mydata, aes(x=Custorderdate, y=OrderCategory)) +
  geom_line( color="#69b3a2", size=2, alpha=0.9, linetype=1.5) +
  ggtitle("order category according to years")
```



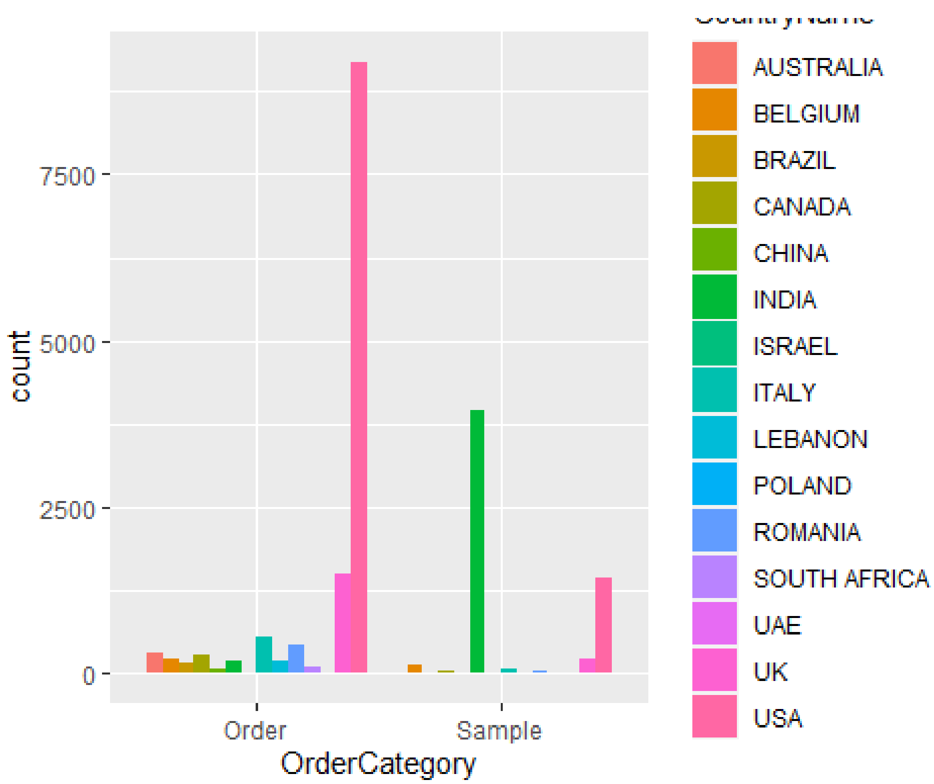
Here, there were no samples sent out in year 2017

```
ggplot(mydata, aes(x = OrderCategory, fill =
                     ITEM_NAME)) +
  geom_bar(position = "dodge")
```



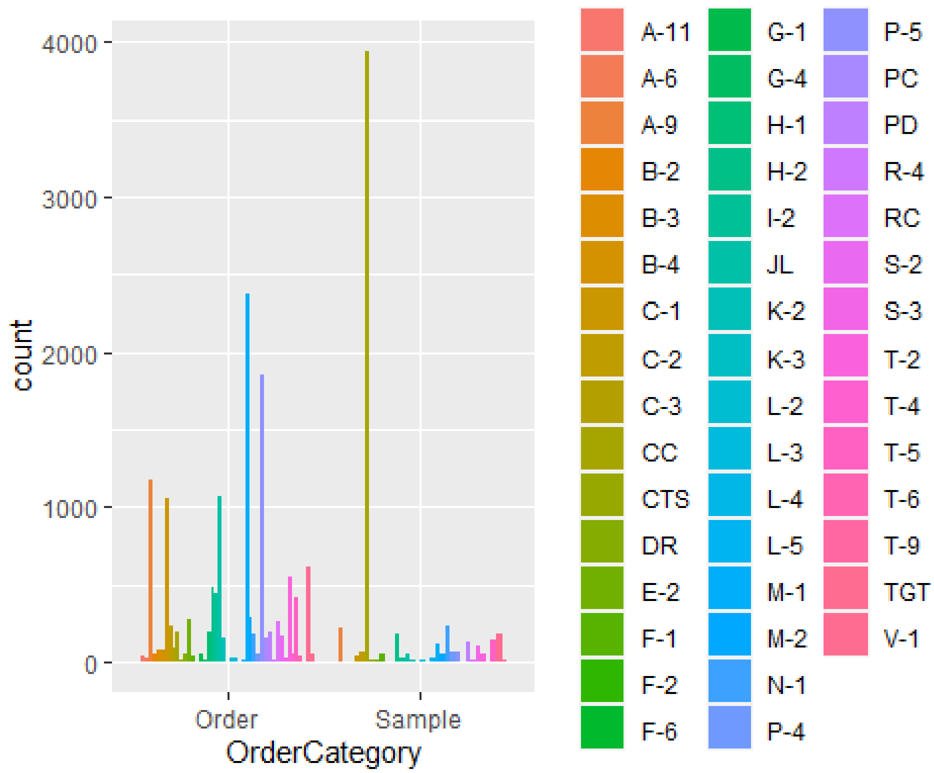
Here, as the "Hand-Tufted" carpets use base-material and entitled the least production effort, hence, generates maximum revenue for the organization in both order and sample category

```
ggplot(mydata, aes(x = OrderCategory, fill =
                     CountryName)) +
  geom_bar(position = "dodge")
```



The maximum consumption of product was in USA followed by India and UK. As, many Indians use Durries variants of kilim.

```
ggplot(mydata, aes(x = OrderCategory, fill =
                     CustomerCode)) +
  geom_bar(position = "dodge")
```



From the graph we can infer that more samples were sent out to the customers with customer code 'CC' and a greater number of orders we placed by customers with customer code between 'L-5' and 'N-1'.

These are some of the insights that could be derived from the Dataset.

--END OF INSIGHTS SECTION--

2) What kind of analytics and machine learning algorithms (e.g., classification, regression, clustering, recommender systems and etc) can be used by Champo Carpets to solve their problems, and in general for value creation? Justify your choices.

Answer:

- Champo carpets can use various analytics and machine learning algorithms to solve their problems such as **Decision Tree, Random Forest, Logistic Regression, K-means Clustering, Neural Networks etc.**
- All these algorithms/methods can be implemented on the Champo Dataset in this assignment.
- Then by calculating the accuracy of the different models, the model with the highest accuracy would be the best for Champo Carpets business.

CHAMPO CARPETS: DESCISION TREE MODEL

```
library(stats)
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.1.2

##
## 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(readxl)

## Warning: package 'readxl' was built under R version 4.1.2

library(tidyr)
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.1.2

library(ISLR)

## Warning: package 'ISLR' was built under R version 4.1.2

library(rpart)
library(rpart.plot)

## Warning: package 'rpart.plot' was built under R version 4.1.2

library(psych)

## Warning: package 'psych' was built under R version 4.1.2

##
## Attaching package: 'psych'

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

Sampledata <- read_excel("C:/Users/Bodake/Desktop/UIC Sem 2/Data Mining IDS-5
72/R/sample.xlsx")
View(Sampledata)
```

```
mydata <- Sampledata
mydata

## # A tibble: 5,820 x 25
##    CustomerCode CountryName   USA    UK Italy Belgium Romania Australia In
dia
##    <chr>        <chr>       <dbl> <dbl> <dbl>   <dbl>   <dbl>     <dbl> <d
bl>
##  1 A-11         USA             1     0     0       0       0         0
0
##  2 A-11         USA             1     0     0       0       0         0
0
##  3 A-11         USA             1     0     0       0       0         0
0
##  4 A-11         USA             1     0     0       0       0         0
0
##  5 A-9          USA             1     0     0       0       0         0
0
##  6 A-9          USA             1     0     0       0       0         0
0
##  7 A-9          USA             1     0     0       0       0         0
0
##  8 A-9          USA             1     0     0       0       0         0
0
##  9 A-9          USA             1     0     0       0       0         0
0
## 10 A-9          USA             1     0     0       0       0         0
0
## # ... with 5,810 more rows, and 16 more variables: QtyRequired <dbl>,
## #   ITEMNAME <chr>, HandTufted <dbl>, Durry <dbl>, DoubleBack <dbl>,
## #   HandWoven <dbl>, Knotted <dbl>, Jacquard <dbl>, Handloom <dbl>,
## #   Other <dbl>, ShapeName <chr>, REC <dbl>, Round <dbl>, Square <dbl>,
## #   AreaFt <dbl>, OrderConversion <dbl>

summary(mydata)

##  CustomerCode       CountryName             USA               UK
##  Length:5820        Length:5820        Min.   :0.0000   Min.   :0.00000
##  Class :character   Class :character   1st Qu.:0.0000   1st Qu.:0.00000
##  Mode  :character   Mode  :character   Median :0.0000   Median :0.00000
##                                        Mean   :0.2491   Mean   :0.03512
##                                        3rd Qu.:0.0000   3rd Qu.:0.00000
##                                        Max.   :1.0000   Max.   :1.00000
##                                        NA's   :39       NA's   :39
##      Italy             Belgium            Romania            Australia
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.00778   Mean   :0.02283   Mean   :0.00346   Mean   :0.00173
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
```

```
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##  NA's   :39        NA's   :39        NA's   :39        NA's   :39
##      India         QtyRequired        ITEMNAME           HandTufted
##  Min.   :0.0000   Min.   :  1.000   Length:5820        Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:  1.000   Class :character   1st Qu.:0.0000
##  Median :1.0000   Median :  1.000   Mode  :character   Median :0.0000
##  Mean   :0.6817   Mean   :  1.975                      Mean   :0.4167
##  3rd Qu.:1.0000   3rd Qu.:  1.000                      3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :200.000                      Max.   :1.0000
##  NA's   :39
##      Durry          DoubleBack        HandWoven         Knotted
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.2686   Mean   :0.09519   Mean   :0.1211   Mean   :0.03729
##  3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.0000   Max.   :1.00000
##
##     Jacquard          Handloom           Other            ShapeName
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Length:5820
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   Class :character
##  Median :0.00000   Median :0.0000   Median :0.00000   Mode  :character
##  Mean   :0.01443   Mean   :0.0177   Mean   :0.02904
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.00000
##
##       REC             Round               Square             AreaFt
##  Min.   :0.0000   Min.   :0.000000   Min.   :0.00000   Min.   :  0.6667
##  1st Qu.:1.0000   1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:  6.0000
##  Median :1.0000   Median :0.000000   Median :0.00000   Median : 11.0000
##  Mean   :0.9864   Mean   :0.009794   Mean   :0.00378   Mean   : 21.5558
##  3rd Qu.:1.0000   3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.: 39.8125
##  Max.   :1.0000   Max.   :1.000000   Max.   :1.00000   Max.   :480.0000
##
##  OrderConversion
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.2009
##  3rd Qu.:0.0000
##  Max.   :1.0000
##
```

```
str(mydata)
```

```
## tibble [5,820 x 25] (S3: tbl_df/tbl/data.frame)
##  $ CustomerCode   : chr [1:5820] "A-11" "A-11" "A-11" "A-11" ...
##  $ CountryName    : chr [1:5820] "USA" "USA" "USA" "USA" ...
##  $ USA            : num [1:5820] 1 1 1 1 1 1 1 1 1 1 ...
##  $ UK             : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
```

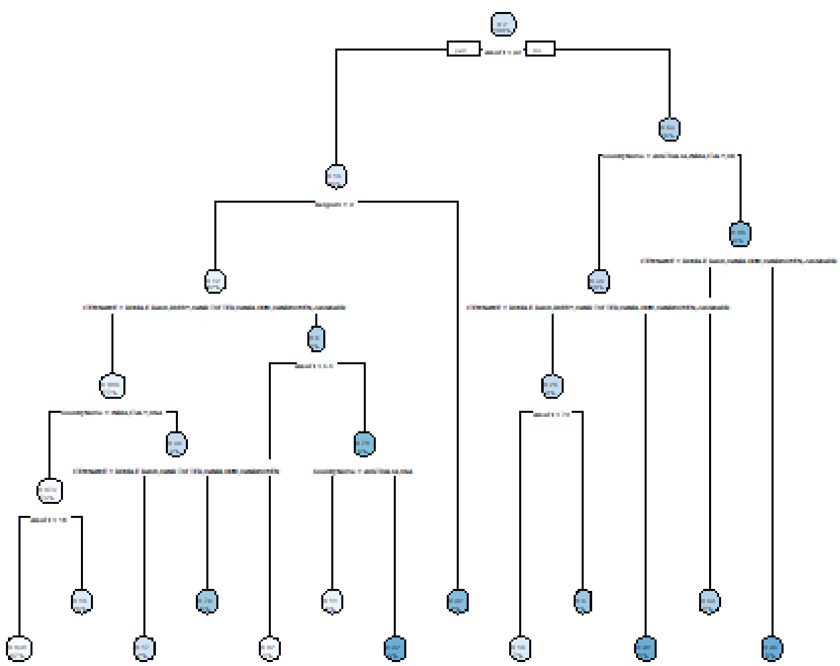
```
##  $ Italy          : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Belgium        : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Romania        : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Australia      : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ India          : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ QtyRequired    : num [1:5820] 2 2 2 1 5 5 5 5 11 2 ...
##  $ ITEMNAME       : chr [1:5820] "DURRY" "DOUBLE BACK" "JACQUARD" "JACQUAR
D" ...
##  $ HandTufted     : num [1:5820] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Durry          : num [1:5820] 1 0 0 0 0 0 0 0 0 0 ...
##  $ DoubleBack     : num [1:5820] 0 1 0 0 1 1 1 1 0 0 ...
##  $ HandWoven      : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Knotted        : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Jacquard       : num [1:5820] 0 0 1 1 0 0 0 0 1 0 ...
##  $ Handloom       : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Other          : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ ShapeName      : chr [1:5820] "REC" "REC" "REC" "REC" ...
##  $ REC            : num [1:5820] 1 1 1 1 1 1 1 1 1 1 ...
##  $ Round          : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Square         : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ AreaFt         : num [1:5820] 4 4 4 42.3 1 ...
##  $ OrderConversion: num [1:5820] 0 0 0 1 0 0 0 0 0 0 ...

mydata <- na.omit(mydata)
mydata <- drop_na(Sampledata)
mydata

## # A tibble: 5,781 x 25
##    CustomerCode CountryName   USA    UK Italy Belgium Romania Australia In
dia
##    <chr>        <chr>       <dbl> <dbl> <dbl>   <dbl>   <dbl>     <dbl> <d
bl>
##  1 A-11         USA             1     0     0       0       0         0
0
##  2 A-11         USA             1     0     0       0       0         0
0
##  3 A-11         USA             1     0     0       0       0         0
0
##  4 A-11         USA             1     0     0       0       0         0
0
##  5 A-9          USA             1     0     0       0       0         0
0
##  6 A-9          USA             1     0     0       0       0         0
0
##  7 A-9          USA             1     0     0       0       0         0
0
##  8 A-9          USA             1     0     0       0       0         0
0
##  9 A-9          USA             1     0     0       0       0         0
0
```

```
## 10 A-9          USA             1     0     0       0       0         0
0
## # ... with 5,771 more rows, and 16 more variables: QtyRequired <dbl>,
## #   ITEMNAME <chr>, HandTufted <dbl>, Durry <dbl>, DoubleBack <dbl>,
## #   HandWoven <dbl>, Knotted <dbl>, Jacquard <dbl>, Handloom <dbl>,
## #   Other <dbl>, ShapeName <chr>, REC <dbl>, Round <dbl>, Square <dbl>,
## #   AreaFt <dbl>, OrderConversion <dbl>

data <- mydata[-1]
```

Training and testing data in 85% and 15% using gini index and info gain

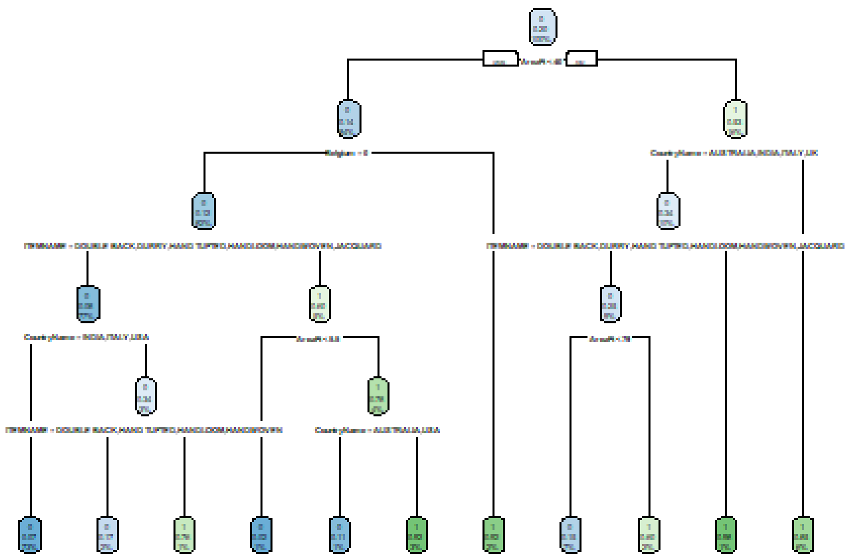
```
set.seed(134)
indx <- sample(2, nrow(data), replace = TRUE, prob = c(0.7, 0.3))
train <- data[indx == 1, ]
nrow(train)

## [1] 4075

test <- data[indx == 2, ]
nrow(test)

## [1] 1706

nrow(train)/nrow(data)

## [1] 0.7048953

myFormula = OrderConversion ~ .
myTree <- rpart(myFormula , data = train)
print(myTree)

## n= 4075
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 4075 655.5912000 0.20147240
##    2) AreaFt< 39.90625 3423 408.3628000 0.13847500
##      4) Belgium< 0.5 3333 345.1311000 0.11731170
##        8) ITEMNAME=DOUBLE BACK,DURRY,HAND TUFTED,HANDLOOM,HANDWOVEN,JACQUA
RD 3121 241.6687000 0.08458827
##         16) CountryName=INDIA,ITALY,USA 2990 202.9595000 0.07324415
##           32) AreaFt< 15.875 1929  71.1612200 0.03836185 *
##           33) AreaFt>=15.875 1061 125.1838000 0.13666350 *
##         17) CountryName=AUSTRALIA,ROMANIA,UK 131  29.5419800 0.34351150
##           34) ITEMNAME=DOUBLE BACK,HAND TUFTED,HANDLOOM,HANDWOVEN 93  13.2
473100 0.17204300 *
##           35) ITEMNAME=DURRY,JACQUARD 38   6.8684210 0.76315790 *
##        9) ITEMNAME=GUN TUFTED,KNOTTED,POWER LOOM JACQUARD,TABLE TUFTED 212
50.9198100 0.59905660
##         18) AreaFt< 5.5 51   0.9803922 0.01960784 *
```

```
##          19) AreaFt>=5.5 161  27.3913000 0.78260870
##            38) CountryName=AUSTRALIA,USA 28   2.6785710 0.10714290 *
##            39) CountryName=INDIA,UK 133   9.2481200 0.92481200 *
##       5) Belgium>=0.5 90   6.4555560 0.92222220 *
##     3) AreaFt>=39.90625 652 162.3236000 0.53220860
##       6) CountryName=AUSTRALIA,INDIA,ITALY,UK 406  91.4113300 0.34236450
##        12) ITEMNAME=DOUBLE BACK,DURRY,HAND TUFTED,HANDLOOM,HANDWOVEN,JACQUA
RD 355  66.6873200 0.25070420
##          24) AreaFt< 79 275  34.8872700 0.14909090 *
##          25) AreaFt>=79 80  19.2000000 0.60000000 *
##        13) ITEMNAME=GUN TUFTED,KNOTTED,POWER LOOM JACQUARD 51   0.9803922 0
.98039220 *
##       7) CountryName=BELGIUM,ROMANIA,USA 246  32.1300800 0.84552850
##        14) ITEMNAME=DOUBLE BACK,HANDLOOM,HANDWOVEN,JACQUARD 51  12.7451000
0.49019610 *
##        15) ITEMNAME=DURRY,HAND TUFTED,KNOTTED 195  11.2615400 0.93846150 *

rpart.plot(myTree)
```



USING GINI INDEX

```
mytree <- rpart(myFormula, data = train,method="class",, parms=list(split="g
ini"), control = rpart.control(minsplit = 10, minbucket = 10), cp=0.03)
rpart.plot(mytree)
```



```
pred_train <- predict(mytree, data=train, type = "class", parms=list(split="
gini"), control = rpart.control(minsplit = 10, minbucket = 10), cp=0.03)
mean(train$OrderConversion != pred_train)

## [1] 0.09251534

pred_test <- predict(mytree, newdata1=test, type = "class", parms=list(split=
                                                                         "gin
i"), control = rpart.control(minsplit = 10, minbucket = 10), cp=0.03)
mean(test$OrderConversion != pred_test)

## Warning in `!=.default`(test$OrderConversion, pred_test): longer object le
ngth
## is not a multiple of shorter object length

## Warning in is.na(e1) | is.na(e2): longer object length is not a multiple o
f
## shorter object length

## [1] 0.2807362
```

```
table_mat <- table(train$OrderConversion, pred_train)
table_mat
```

```
##    pred_train
##         0    1
##   0 3157   97
##   1  280  541
```

```
accuracy_Test <- sum(diag(table_mat)) / sum(table_mat)
print(paste('Accuracy for test', accuracy_Test))
```

```
## [1] "Accuracy for test 0.907484662576687"
```

```
printcp(mytree)
```

```
##
## Classification tree:
## rpart(formula = myFormula, data = train, method = "class", parms = list(sp
lit = "g\nini"),
##     control = rpart.control(minsplit = 10, minbucket = 10), cp = 0.03)
##
## Variables actually used in tree construction:
## [1] AreaFt      Belgium     CountryName ITEMNAME
##
## Root node error: 821/4075 = 0.20147
##
## n= 4075
##
##         CP nsplit rel error  xerror     xstd
## 1 0.103532      0   1.00000 1.00000 0.031187
## 2 0.092570      2   0.79294 0.79172 0.028470
## 3 0.059683      3   0.70037 0.70767 0.027186
## 4 0.055420      4   0.64068 0.65895 0.026383
## 5 0.026797      6   0.52984 0.53471 0.024107
## 6 0.019488      7   0.50305 0.50914 0.023591
## 7 0.012180      8   0.48356 0.50426 0.023491
## 8 0.010000     10   0.45920 0.50305 0.023465
```

```
printcp(mytree)
```

```
##
## Classification tree:
## rpart(formula = myFormula, data = train, method = "class", parms = list(sp
lit = "g\nini"),
##     control = rpart.control(minsplit = 10, minbucket = 10), cp = 0.03)
##
## Variables actually used in tree construction:
## [1] AreaFt      Belgium     CountryName ITEMNAME
##
## Root node error: 821/4075 = 0.20147
##
```

```
## n= 4075
##
##          CP nsplit rel error  xerror     xstd
## 1 0.103532      0   1.00000 1.00000 0.031187
## 2 0.092570      2   0.79294 0.79172 0.028470
## 3 0.059683      3   0.70037 0.70767 0.027186
## 4 0.055420      4   0.64068 0.65895 0.026383
## 5 0.026797      6   0.52984 0.53471 0.024107
## 6 0.019488      7   0.50305 0.50914 0.023591
## 7 0.012180      8   0.48356 0.50426 0.023491
## 8 0.010000     10   0.45920 0.50305 0.023465
```

LOGISTIC REGRESSION

```
mydata = subset(mydata, select = -c(2))
head(mydata)
```

```
## # A tibble: 6 x 24
##   CustomerCode   USA    UK Italy Belgium Romania Australia India QtyRequir
ed
##   <chr>        <dbl> <dbl> <dbl>   <dbl>   <dbl>     <dbl> <dbl>       <db
l>
## 1 A-11             1     0     0       0       0         0     0
2
## 2 A-11             1     0     0       0       0         0     0
2
## 3 A-11             1     0     0       0       0         0     0
2
## 4 A-11             1     0     0       0       0         0     0
1
## 5 A-9              1     0     0       0       0         0     0
5
## 6 A-9              1     0     0       0       0         0     0
5
## # ... with 15 more variables: ITEMNAME <chr>, HandTufted <dbl>, Durry <dbl
>,
## #   DoubleBack <dbl>, HandWoven <dbl>, Knotted <dbl>, Jacquard <dbl>,
## #   Handloom <dbl>, Other <dbl>, ShapeName <chr>, REC <dbl>, Round <dbl>,
## #   Square <dbl>, AreaFt <dbl>, OrderConversion <dbl>
```

```
mydata$ITEMNAME <- as.numeric(factor(as.matrix(mydata$ITEMNAME)))
mydata$ShapeName <- as.numeric(factor(as.matrix(mydata$ShapeName)))
mydata$CustomerCode <- as.numeric(factor(as.matrix(mydata$CustomerCode)))


sapply(mydata, sd)
```

```
##     CustomerCode             USA              UK           Italy          Be
lgium
##       6.06258072      0.43252452      0.18408647      0.08789121       0.149
38512
```

```
##          Romania        Australia              India      QtyRequired          ITE
MNAME
##       0.05872169       0.04155853       0.46585067       5.70186587       2.102
92013
##       HandTufted            Durry       DoubleBack        HandWoven           Kn
otted
##       0.49320195       0.44310911       0.29343256       0.32725777       0.190
08945
##         Jacquard         Handloom            Other        ShapeName
REC
##       0.11391193       0.13039081       0.16799129       0.15742051       0.116
10789
##            Round           Square           AreaFt OrderConversion
##       0.09881476       0.06157713      21.58072448       0.39935834
```

Converting the target variable into categorical variable

```
mydata$OrderConversion  = factor(mydata$OrderConversion)
mylogit = glm(OrderConversion ~ ., data = mydata, family = "binomial")
summary(mylogit)

##
## Call:
## glm(formula = OrderConversion ~ ., family = "binomial", data = mydata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.0003  -0.5838  -0.2655  -0.1937   2.9829
##
## Coefficients: (4 not defined because of singularities)
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.771625   1.504858  -1.842   0.0655 .
## CustomerCode -0.084170   0.009133  -9.216  < 2e-16 ***
## USA           1.950556   0.110686  17.622  < 2e-16 ***
## UK            3.114674   0.229502  13.571  < 2e-16 ***
## Italy         1.205744   0.528207   2.283   0.0224 *
## Belgium       7.172141   0.393417  18.230  < 2e-16 ***
## Romania       3.526021   0.507606   6.946 3.75e-12 ***
## Australia    -0.398053   0.834774  -0.477   0.6335
## India               NA         NA      NA       NA
## QtyRequired   0.012970   0.006547   1.981   0.0476 *
## ITEMNAME      0.245822   0.059973   4.099 4.15e-05 ***
## HandTufted   -3.137228   0.323028  -9.712  < 2e-16 ***
## Durry        -2.466272   0.421541  -5.851 4.90e-09 ***
## DoubleBack   -2.481866   0.496060  -5.003 5.64e-07 ***
## HandWoven    -4.341886   0.294290 -14.754  < 2e-16 ***
## Knotted      -1.440230   0.266355  -5.407 6.40e-08 ***
## Jacquard     -4.266561   0.404220 -10.555  < 2e-16 ***
## Handloom     -3.470214   0.401869  -8.635  < 2e-16 ***
## Other               NA         NA      NA       NA
```

```
## ShapeName     1.116868   0.652405   1.712   0.0869 .
## REC           0.303105   0.890790   0.340   0.7337
## Round               NA         NA      NA       NA
## Square              NA         NA      NA       NA
## AreaFt        0.058131   0.002399  24.229  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5771.2  on 5780  degrees of freedom
## Residual deviance: 3637.7  on 5761  degrees of freedom
## AIC: 3677.7
##
## Number of Fisher Scoring iterations: 6
```

Confidence using standard errors

```
confint(mylogit)

## Waiting for profiling to be done...

##                     2.5 %      97.5 %
## (Intercept)  -5.686117076  0.25024041
## CustomerCode -0.102237222 -0.06642042
## USA           1.734398031  2.16843937
## UK            2.664437197  3.56480443
## Italy         0.122120250  2.21010343
## Belgium       6.440979573  7.99411018
## Romania       2.501229177  4.51413502
## Australia    -2.254510950  1.13814790
## India                  NA          NA
## QtyRequired  -0.000407948  0.02552246
## ITEMNAME      0.129111972  0.36510007
## HandTufted   -3.772908583 -2.49999503
## Durry        -3.295205445 -1.63350703
## DoubleBack   -3.457581395 -1.50382220
## HandWoven    -4.925722829 -3.76963709
## Knotted      -1.973329270 -0.92676053
## Jacquard     -5.095712313 -3.50430456
## Handloom     -4.272190943 -2.69308427
## Other                  NA          NA
## ShapeName    -0.207331667  2.36942268
## REC          -1.447360027  2.06103259
## Round                  NA          NA
## Square                 NA          NA
## AreaFt        0.053499641  0.06290785

confint.default(mylogit)
```

```
##                      2.5 %      97.5 %
## (Intercept)  -5.7210927484  0.17784195
## CustomerCode -0.1020704785 -0.06626957
## USA           1.7336156454  2.16749643
## UK            2.6648588133  3.56448967
## Italy         0.1704777114  2.24101016
## Belgium       6.4010575211  7.94322513
## Romania       2.5311320999  4.52091068
## Australia    -2.0341787328  1.23807358
## India                   NA          NA
## QtyRequired   0.0001392121  0.02580140
## ITEMNAME      0.1282774504  0.36336659
## HandTufted   -3.7703506333 -2.50410506
## Durry        -3.2924779752 -1.64006656
## DoubleBack   -3.4541252522 -1.50960773
## HandWoven    -4.9186835743 -3.76508842
## Knotted      -1.9622750283 -0.91818438
## Jacquard     -5.0588162594 -3.47430484
## Handloom     -4.2578622515 -2.68256608
## Other                   NA          NA
## ShapeName    -0.1618212298  2.39555746
## REC          -1.4428119266  2.04902117
## Round                   NA          NA
## Square                  NA          NA
## AreaFt        0.0534288937  0.06283371
```

Odd Ratio

```
exp(coef(mylogit))

##  (Intercept) CustomerCode          USA           UK        Italy       Belg
ium
## 6.256024e-02 9.192749e-01 7.032597e+00 2.252609e+01 3.339242e+00 1.302631e
+03
##      Romania    Australia        India  QtyRequired     ITEMNAME    HandTuf
ted
## 3.398847e+01 6.716267e-01           NA 1.013055e+00 1.278672e+00 4.340295e
-02
##        Durry   DoubleBack    HandWoven      Knotted     Jacquard      Handl
oom
## 8.490076e-02 8.358707e-02 1.301196e-02 2.368733e-01 1.402996e-02 3.111037e
-02
##        Other    ShapeName          REC        Round       Square        Are
aFt
##           NA 3.055270e+00 1.354056e+00           NA           NA 1.059854e
+00

exp(cbind(OR = coef(mylogit), confint(mylogit)))

## Waiting for profiling to be done...
```

```
##                         OR        2.5 %       97.5 %
## (Intercept)  6.256024e-02 3.392741e-03 1.284334e+00
## CustomerCode 9.192749e-01 9.028154e-01 9.357374e-01
## USA          7.032597e+00 5.665516e+00 8.744626e+00
## UK           2.252609e+01 1.435987e+01 3.533254e+01
## Italy        3.339242e+00 1.129890e+00 9.116659e+00
## Belgium      1.302631e+03 6.270207e+02 2.963452e+03
## Romania      3.398847e+01 1.219748e+01 9.129856e+01
## Australia    6.716267e-01 1.049248e-01 3.120983e+00
## India                  NA           NA           NA
## QtyRequired  1.013055e+00 9.995921e-01 1.025851e+00
## ITEMNAME     1.278672e+00 1.137818e+00 1.440658e+00
## HandTufted   4.340295e-02 2.298511e-02 8.208541e-02
## Durry        8.490076e-02 3.706043e-02 1.952436e-01
## DoubleBack   8.358707e-02 3.150587e-02 2.222789e-01
## HandWoven    1.301196e-02 7.257478e-03 2.306043e-02
## Knotted      2.368733e-01 1.389933e-01 3.958339e-01
## Jacquard     1.402996e-02 6.122944e-03 3.006768e-02
## Handloom     3.111037e-02 1.395118e-02 6.767190e-02
## Other                  NA           NA           NA
## ShapeName    3.055270e+00 8.127500e-01 1.069122e+01
## REC          1.354056e+00 2.351904e-01 7.854076e+00
## Round                  NA           NA           NA
## Square                 NA           NA           NA
## AreaFt       1.059854e+00 1.054957e+00 1.064929e+00

newdata1 = with(mydata, data.frame(CustomerCode = mean(CustomerCode),
                                   QtyRequired  = mean(QtyRequired),
                                   ITEM_NAME    = factor(1:12),
                                   ShapeName    = mean(ShapeName),
                                   AreaFt       = mean(AreaFt)))

newdata1

##    CustomerCode QtyRequired ITEM_NAME ShapeName   AreaFt
## 1      8.854869    1.977859         1  1.017471 21.63013
## 2      8.854869    1.977859         2  1.017471 21.63013
## 3      8.854869    1.977859         3  1.017471 21.63013
## 4      8.854869    1.977859         4  1.017471 21.63013
## 5      8.854869    1.977859         5  1.017471 21.63013
## 6      8.854869    1.977859         6  1.017471 21.63013
## 7      8.854869    1.977859         7  1.017471 21.63013
## 8      8.854869    1.977859         8  1.017471 21.63013
## 9      8.854869    1.977859         9  1.017471 21.63013
## 10     8.854869    1.977859        10  1.017471 21.63013
## 11     8.854869    1.977859        11  1.017471 21.63013
## 12     8.854869    1.977859        12  1.017471 21.63013

Pred <- predict(mylogit, newdata2 = newdata1, type = "response")
```

Difference between deviance for the two models = test statistic

```
with(mylogit, null.deviance - deviance)

## [1] 2133.524
```

The degrees of freedom for the difference between the two models is equal to the number of predictor variables in the mode, and can be obtained using:

```
with(mylogit, df.null, df.residual)

## [1] 5780
```

Finally, the p-value can be obtained using

```
with(mylogit, pchisq(null.deviance - deviance, df.null - df.residual, lower.t
ail = FALSE))

## [1] 0

class <- ifelse(Pred >= 0.5, "YES", "NO")
```

Accuracy using the Discission Tree Model is around 90.74%

--END OF DESCICION TREE MODEL--

CHAMPO CARPETS – RANDOM FOREST MODEL

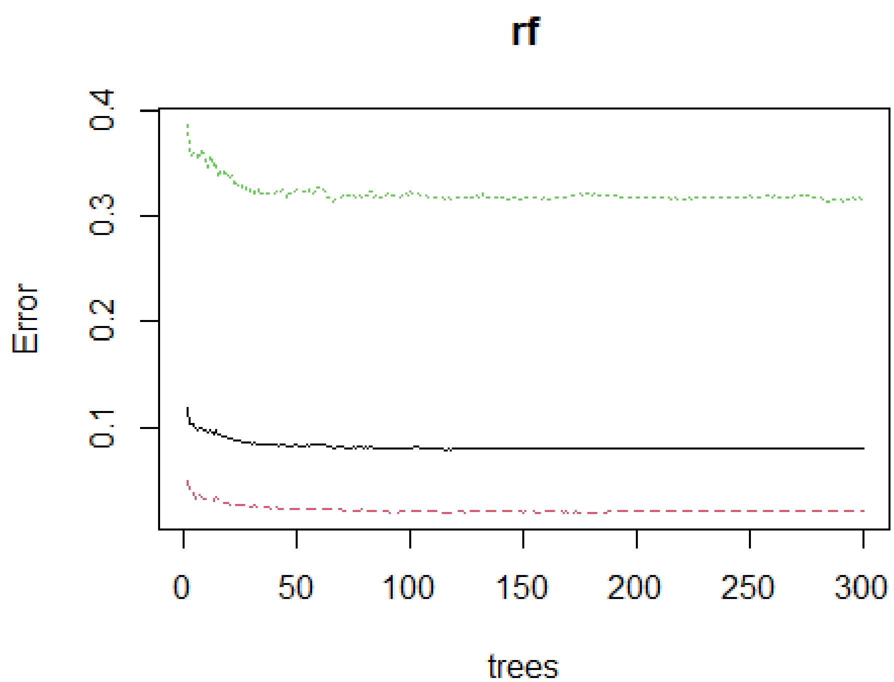
```
Sampledata <- read_excel("C:/Users/Bodake/Desktop/UIC Sem 2/Data Mining IDS-5
72/R/sample.xlsx")
View(Sampledata)
mydata <- Sampledata

str(mydata)

## tibble [5,820 x 25] (S3: tbl_df/tbl/data.frame)
##  $ CustomerCode   : chr [1:5820] "A-11" "A-11" "A-11" "A-11" ...
##  $ CountryName    : chr [1:5820] "USA" "USA" "USA" "USA" ...
##  $ USA            : num [1:5820] 1 1 1 1 1 1 1 1 1 1 ...
##  $ UK             : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Italy          : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Belgium        : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Romania        : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Australia      : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ India          : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ QtyRequired    : num [1:5820] 2 2 2 1 5 5 5 5 11 2 ...
##  $ ITEMNAME       : chr [1:5820] "DURRY" "DOUBLE BACK" "JACQUARD" "JACQUAR
D" ...
##  $ HandTufted     : num [1:5820] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Durry          : num [1:5820] 1 0 0 0 0 0 0 0 0 0 ...
##  $ DoubleBack     : num [1:5820] 0 1 0 0 1 1 1 1 0 0 ...
##  $ HandWoven      : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Knotted        : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Jacquard       : num [1:5820] 0 0 1 1 0 0 0 0 1 0 ...
##  $ Handloom       : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Other          : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ ShapeName      : chr [1:5820] "REC" "REC" "REC" "REC" ...
##  $ REC            : num [1:5820] 1 1 1 1 1 1 1 1 1 1 ...
##  $ Round          : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Square         : num [1:5820] 0 0 0 0 0 0 0 0 0 0 ...
##  $ AreaFt         : num [1:5820] 4 4 4 42.3 1 ...
##  $ OrderConversion: num [1:5820] 0 0 0 1 0 0 0 0 0 0 ...

sum(is.na(mydata))

## [1] 273

mydata[is.na(mydata)] = 0
sum(is.na(mydata))

## [1] 0

sapply(mydata,function(x)sum(is.na(x)))
```

```
##    CustomerCode     CountryName             USA              UK           
Italy 
##               0               0               0               0               
0 
##         Belgium         Romania       Australia           India     QtyReq
uired 
##               0               0               0               0               
0 
##        ITEMNAME      HandTufted           Durry      DoubleBack        Hand
Woven 
##               0               0               0               0               
0 
##         Knotted        Jacquard        Handloom           Other         Shap
eName 
##               0               0               0               0               
0 
##             REC           Round          Square          AreaFt OrderConve
rsion 
##               0               0               0               0               
0

mydata <- as.data.frame(mydata)
mydata <- subset (mydata, select = -c(USA, UK, Italy, Belgium, Romania, Austr
alia, India))
str(mydata)

mydata$OrderConversion <-as.factor(factor(as.matrix(mydata$OrderConversion)))
```

#Random Forest Model

```
set.seed(100)
rf <- randomForest(mydata$OrderConversion ~ ., data = mydata, ntree = 300, mt
ry 
                   = sqrt(ncol(mydata)-1), proximity = T, importance = T )
print(rf)

## 
## Call:
##  randomForest(formula = mydata$OrderConversion ~ ., data = mydata,      nt
ree = 300, mtry = sqrt(ncol(mydata) - 1), proximity = T,      importance = T) 
##                Type of random forest: classification
##                      Number of trees: 300
## No. of variables tried at each split: 4
## 
##         OOB estimate of  error rate: 8.09%
## Confusion matrix:
##      0   1 class.error
## 0 4550 101  0.02171576
## 1  370 799  0.31650984
```

```
attributes(rf)

## $names
##  [1] "call"            "type"            "predicted"       "err.rate"
##  [5] "confusion"       "votes"           "oob.times"       "classes"
##  [9] "importance"      "importanceSD"    "localImportance" "proximity"
## [13] "ntree"           "mtry"            "forest"          "y"
## [17] "test"            "inbag"           "terms"
##
## $class
## [1] "randomForest.formula" "randomForest"

plot(rf)
```




```
p1 <- predict(rf, mydata)
confusionMatrix(p1, mydata$OrderConversion)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 4585  325
##          1   66  844
##
##                Accuracy : 0.9328
##                  95% CI : (0.9261, 0.9391)
##     No Information Rate : 0.7991
```

```
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.7718
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9858
##             Specificity : 0.7220
##          Pos Pred Value : 0.9338
##          Neg Pred Value : 0.9275
##              Prevalence : 0.7991
##          Detection Rate : 0.7878
##    Detection Prevalence : 0.8436
##       Balanced Accuracy : 0.8539
##
##        'Positive' Class : 0
##
```

Hence, we can say that the ACCURACY OF RANDOM FOREST MODEL ON OUR DATA SET IS Around 93.3%

```
rf$importance

##                         0            1 MeanDecreaseAccuracy MeanDecreaseGi
ni
## CustomerCode 0.0646396504 0.0954731843         0.0708230649       137.6328
50
## CountryName  0.0604618670 0.1396365557         0.0764096629       227.2380
58
## QtyRequired  0.0126190009 0.0187168546         0.0138396952        59.2091
02
## ITEMNAME     0.0317481579 0.1123979410         0.0480179342       158.8418
82
## HandTufted   0.0099520814 0.0392259754         0.0158561771        20.8640
11
## Durry        0.0073486691 0.0163196586         0.0091820266        12.9136
27
## DoubleBack   0.0028084925 0.0031394463         0.0028756082         6.7083
94
## HandWoven    0.0036159754 0.0130353377         0.0055149885        10.8905
96
## Knotted      0.0063744450 0.0297972179         0.0110864751        36.4349
19
## Jacquard     0.0007098712 0.0001888175         0.0006036450         3.0705
36
## Handloom     0.0011495385 0.0016574721         0.0012533988         8.3225
12
## Other        0.0060933853 0.0284706373         0.0105960845        53.9485
06
## ShapeName    0.0006676833 0.0016274978         0.0008615113         4.3851
```

```
76
## REC         0.0008059448 0.0017874371         0.0010039013          3.7482
23
## Round       0.0002526424 0.0010040740         0.0004028903          2.2161
22
## Square      0.0002057407 0.0006069963         0.0002872994          2.2244
54
## AreaFt      0.0537991030 0.2040063657         0.0841027039        350.2552
06
```

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

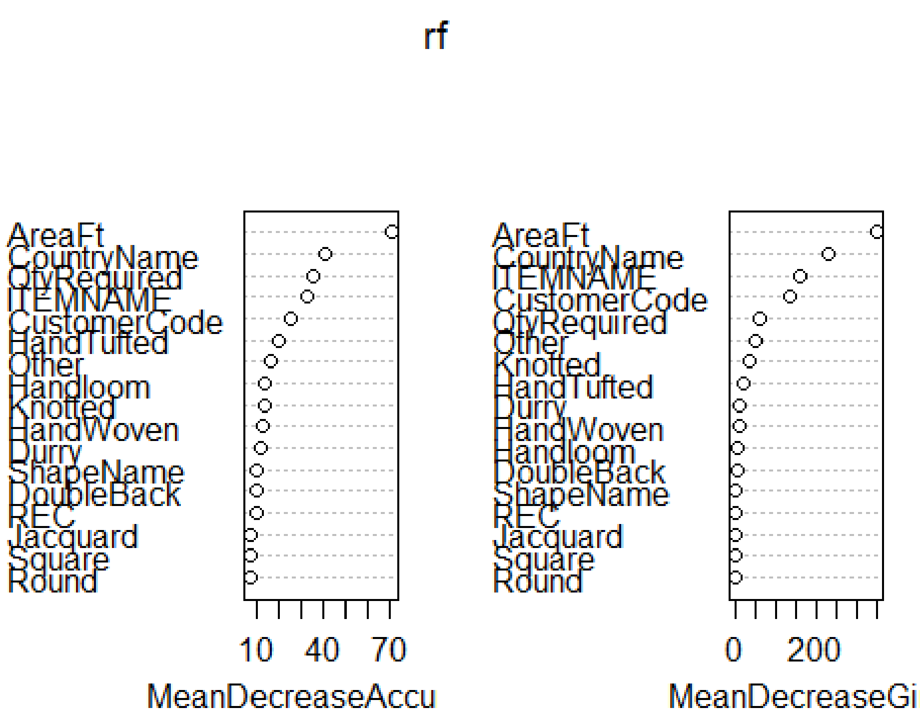
```
##                MeanDecreaseAccuracy
## CustomerCode              25.388151
## CountryName               41.369526
## QtyRequired               35.330187
## ITEMNAME                  32.495191
## HandTufted                19.806019
## Durry                     12.171852
## DoubleBack                10.019951
## HandWoven                 13.239589
## Knotted                   13.745132
## Jacquard                   7.663893
## Handloom                  14.064707
## Other                     16.796032
## ShapeName                 10.236109
## REC                        9.808721
## Round                      6.955230
## Square                     7.489018
## AreaFt                    71.507265
```

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

```
##                MeanDecreaseGini
## CustomerCode         137.632850
## CountryName          227.238058
## QtyRequired           59.209102
## ITEMNAME             158.841882
## HandTufted            20.864011
## Durry                 12.913627
## DoubleBack             6.708394
## HandWoven             10.890596
## Knotted               36.434919
## Jacquard               3.070536
## Handloom               8.322512
## Other                 53.948506
## ShapeName              4.385176
## REC                    3.748223
## Round                  2.216122
## Square                 2.224454
## AreaFt               350.255206
```

```r
varImpPlot(rf, scale=FALSE)
```


```r
varImpPlot(rf)
```



```
## [299,] 0.08041237 0.02128575 0.3156544
## [300,] 0.08092784 0.02171576 0.3165098

ntree=100
rf$err.rate[ntree,1]

##        OOB
## 0.08127148

head(rf$predicted)

## 1 2 3 4 5 6
## 0 0 0 0 0 0
## Levels: 0 1

ind <- sample(2, nrow(mydata), replace = T, prob = c(0.7, 0.3))
Train <- mydata[ind == 1, ]
Validation <- mydata[ind == 2, ]
pr.err <- c()

rf$predicted


## Levels: 0 1

table(rf$predicted, mydata$OrderConversion, dnn = c("Predicted", "Actual"))

##          Actual
## Predicted    0    1
##         0 4550  370
##         1  101  799

confusionMatrix(rf$predicted, mydata$OrderConversion)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 4550  370
##          1  101  799
##
##                Accuracy : 0.9191
##                  95% CI : (0.9118, 0.926)
##     No Information Rate : 0.7991
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.7242
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9783
##             Specificity : 0.6835
```

```
##          Pos Pred Value : 0.9248
##          Neg Pred Value : 0.8878
##              Prevalence : 0.7991
##          Detection Rate : 0.7818
##    Detection Prevalence : 0.8454
##       Balanced Accuracy : 0.8309
##
##        'Positive' Class : 0
##
```

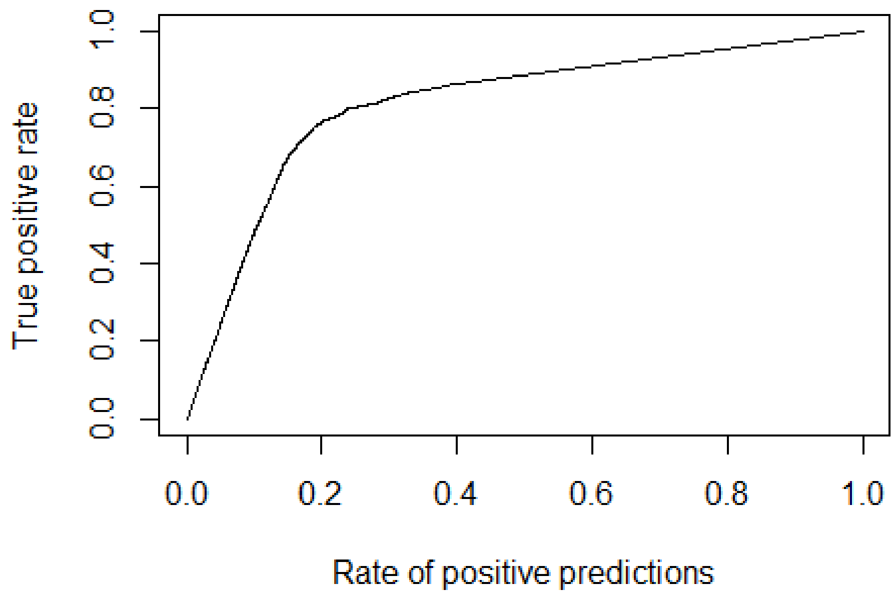
```
library(ROCR)


pred <- prediction(rf$votes[,2], mydata$OrderConversion)

pred
```

```
## A prediction instance
##   with 5820 data points
```

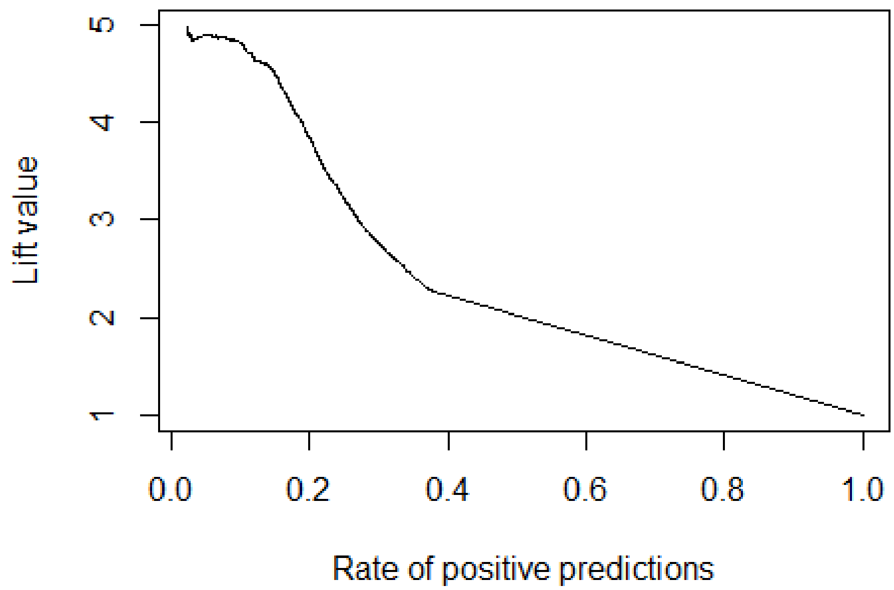
```
#Gain Chart
per <- performance(pred, "tpr","rpp")
per
```

```
## A performance instance
##   'Rate of positive predictions' vs. 'True positive rate' (alpha: 'Cutoff'
)
##   with 1191 data points
```

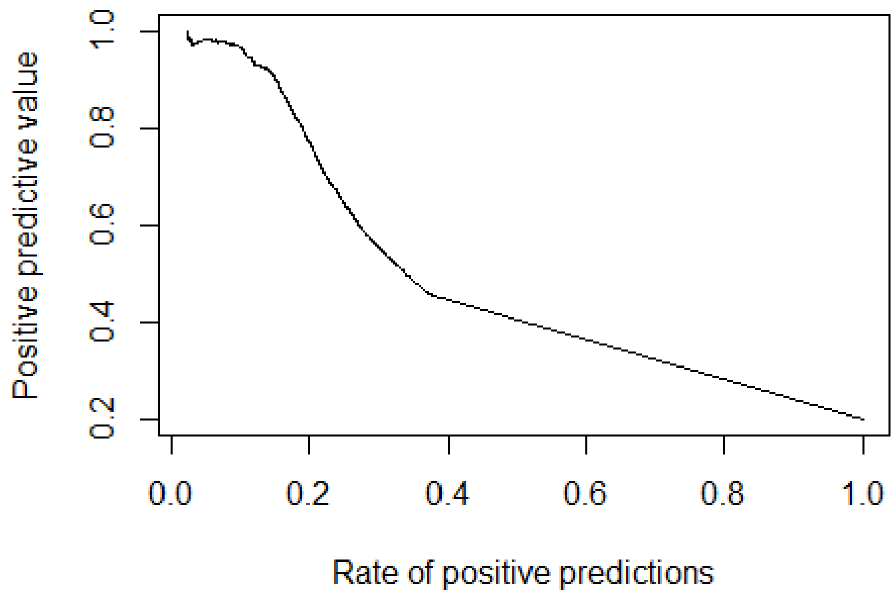
```
plot(per)
```



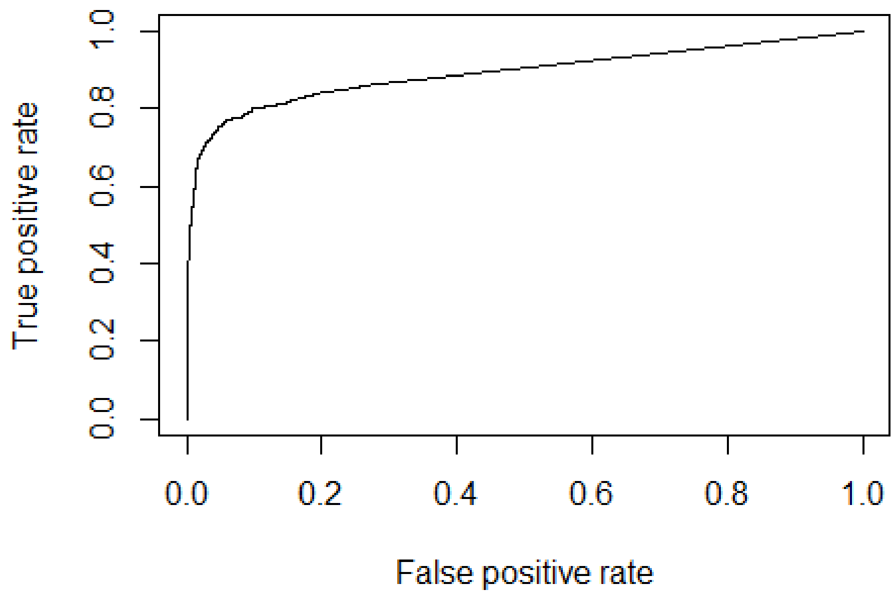

```
#Lift Chart
perf <- performance(pred, "lift","rpp")
perf

## A performance instance
##   'Rate of positive predictions' vs. 'Lift value' (alpha: 'Cutoff')
##   with 1191 data points

plot(perf)
```




```
#Response Chart
perfo <- performance(pred, "ppv","rpp")
perfo

## A performance instance
##   'Rate of positive predictions' vs. 'Positive predictive value' (alpha: '
Cutoff')
##   with 1191 data points

plot(perfo)
```




```r
#ROC Curve
perfom <- performance(pred, "tpr","fpr")
perfom

## A performance instance
##   'False positive rate' vs. 'True positive rate' (alpha: 'Cutoff')
##   with 1191 data points

plot(perfom)
```




```r
#AUC
auc <- performance(pred, "auc")
auc

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

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

## [1] 0.8915729
```

Hence, we can say that the ACCURACY OF RANDOM FOREST MODEL ON OUR DATA SET IS Around 93.3%

---END OF RANDOM FOREST MODEL ANALYSIS---

CHAMPO CARPETS: LOGISTIC RERESSION MODEL ON SAMPLE DATA

```
mydata <- read_excel("C:/Users/Bodake/Desktop/UIC Sem 2/Data Mining IDS-
572/R/sample.xlsx")
mydata <- na.omit(mydata)
mydata <- drop_na(mydata)

mydata = subset(mydata, select = -c(2))
head(mydata)

## # A tibble: 6 x 24
##   CustomerCode   USA    UK Italy Belgium Romania Australia India
QtyRequired
##   <chr>        <dbl> <dbl> <dbl>   <dbl>   <dbl>     <dbl> <dbl>
<dbl>
## 1 A-11             1     0     0       0       0         0     0
2
## 2 A-11             1     0     0       0       0         0     0
2
## 3 A-11             1     0     0       0       0         0     0
2
## 4 A-11             1     0     0       0       0         0     0
1
## 5 A-9              1     0     0       0       0         0     0
5
## 6 A-9              1     0     0       0       0         0     0
5
## # ... with 15 more variables: ITEMNAME <chr>, HandTufted <dbl>, Durry
<dbl>,
## #   DoubleBack <dbl>, HandWoven <dbl>, Knotted <dbl>, Jacquard <dbl>,
## #   Handloom <dbl>, Other <dbl>, ShapeName <chr>, REC <dbl>, Round <dbl>,
## #   Square <dbl>, AreaFt <dbl>, OrderConversion <dbl>

mydata$ITEMNAME <- as.numeric(factor(as.matrix(mydata$ITEMNAME)))
mydata$ShapeName <- as.numeric(factor(as.matrix(mydata$ShapeName)))
mydata$CustomerCode <- as.numeric(factor(as.matrix(mydata$CustomerCode)))
str(mydata)

## tibble [5,781 x 24] (S3: tbl_df/tbl/data.frame)
##  $ CustomerCode   : num [1:5781] 1 1 1 1 2 2 2 2 2 2 ...
##  $ USA            : num [1:5781] 1 1 1 1 1 1 1 1 1 1 ...
##  $ UK             : num [1:5781] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Italy          : num [1:5781] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Belgium        : num [1:5781] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Romania        : num [1:5781] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Australia      : num [1:5781] 0 0 0 0 0 0 0 0 0 0 ...
##  $ India          : num [1:5781] 0 0 0 0 0 0 0 0 0 0 ...
##  $ QtyRequired    : num [1:5781] 2 2 2 1 5 5 5 5 11 2 ...
```

```
##  $ ITEMNAME       : num [1:5781] 2 1 8 8 1 1 1 1 8 4 ...
##  $ HandTufted     : num [1:5781] 0 0 0 0 0 0 0 0 0 1 ...
##  $ Durry          : num [1:5781] 1 0 0 0 0 0 0 0 0 0 ...
##  $ DoubleBack     : num [1:5781] 0 1 0 0 1 1 1 1 0 0 ...
##  $ HandWoven      : num [1:5781] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Knotted        : num [1:5781] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Jacquard       : num [1:5781] 0 0 1 1 0 0 0 0 1 0 ...
##  $ Handloom       : num [1:5781] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Other          : num [1:5781] 0 0 0 0 0 0 0 0 0 0 ...
##  $ ShapeName      : num [1:5781] 1 1 1 1 1 1 1 1 1 1 ...
##  $ REC            : num [1:5781] 1 1 1 1 1 1 1 1 1 1 ...
##  $ Round          : num [1:5781] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Square         : num [1:5781] 0 0 0 0 0 0 0 0 0 0 ...
##  $ AreaFt         : num [1:5781] 4 4 4 42.3 1 ...
##  $ OrderConversion: num [1:5781] 0 0 0 1 0 0 0 0 0 0 ...
##  - attr(*, "na.action")= 'omit' Named int [1:39] 227 4284 4285 4286 4287
4288 4289 4290 4291 4292 ...
##   ..- attr(*, "names")= chr [1:39] "227" "4284" "4285" "4286" ...

sapply(mydata, sd)

##    CustomerCode             USA              UK           Italy
Belgium
##      6.06258072      0.43252452      0.18408647      0.08789121
0.14938512
##         Romania       Australia           India     QtyRequired
ITEMNAME
##      0.05872169      0.04155853      0.46585067      5.70186587
2.10292013
##      HandTufted           Durry      DoubleBack       HandWoven
Knotted
##      0.49320195      0.44310911      0.29343256      0.32725777
0.19008945
##        Jacquard        Handloom           Other       ShapeName
REC
##      0.11391193      0.13039081      0.16799129      0.15742051
0.11610789
##           Round          Square          AreaFt OrderConversion
##      0.09881476      0.06157713     21.58072448      0.39935834
```

Converting the target variable into categorical variable

```
mydata$OrderConversion = factor(mydata$OrderConversion)
mylogit = glm(OrderConversion ~ ., data = mydata, family = "binomial")
summary(mylogit)

##
## Call:
## glm(formula = OrderConversion ~ ., family = "binomial", data = mydata)
##
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -3.0003  -0.5838  -0.2655  -0.1937   2.9829
##
## Coefficients: (4 not defined because of singularities)
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.771625   1.504858  -1.842   0.0655 .
## CustomerCode -0.084170   0.009133  -9.216  < 2e-16 ***
## USA           1.950556   0.110686  17.622  < 2e-16 ***
## UK            3.114674   0.229502  13.571  < 2e-16 ***
## Italy         1.205744   0.528207   2.283   0.0224 *
## Belgium       7.172141   0.393417  18.230  < 2e-16 ***
## Romania       3.526021   0.507606   6.946 3.75e-12 ***
## Australia    -0.398053   0.834774  -0.477   0.6335
## India               NA         NA      NA       NA
## QtyRequired   0.012970   0.006547   1.981   0.0476 *
## ITEMNAME      0.245822   0.059973   4.099 4.15e-05 ***
## HandTufted   -3.137228   0.323028  -9.712  < 2e-16 ***
## Durry        -2.466272   0.421541  -5.851 4.90e-09 ***
## DoubleBack   -2.481866   0.496060  -5.003 5.64e-07 ***
## HandWoven    -4.341886   0.294290 -14.754  < 2e-16 ***
## Knotted      -1.440230   0.266355  -5.407 6.40e-08 ***
## Jacquard     -4.266561   0.404220 -10.555  < 2e-16 ***
## Handloom     -3.470214   0.401869  -8.635  < 2e-16 ***
## Other               NA         NA      NA       NA
## ShapeName     1.116868   0.652405   1.712   0.0869 .
## REC           0.303105   0.890790   0.340   0.7337
## Round               NA         NA      NA       NA
## Square              NA         NA      NA       NA
## AreaFt        0.058131   0.002399  24.229  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5771.2  on 5780  degrees of freedom
## Residual deviance: 3637.7  on 5761  degrees of freedom
## AIC: 3677.7
##
## Number of Fisher Scoring iterations: 6
```

Performing ANOVA to analyze table of deviance

```
anova(mylogit, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: OrderConversion
##
```

```
## Terms added sequentially (first to last)
##
##
##              Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                         5780     5771.2
## CustomerCode  1    54.52      5779     5716.7 1.537e-13 ***
## USA           1    32.82      5778     5683.9 1.010e-08 ***
## UK            1    19.97      5777     5663.9 7.859e-06 ***
## Italy         1     0.40      5776     5663.5 0.5269861
## Belgium       1   471.78      5775     5191.7 < 2.2e-16 ***
## Romania       1    14.92      5774     5176.8 0.0001124 ***
## Australia     1     0.91      5773     5175.9 0.3413791
## India         0     0.00      5773     5175.9
## QtyRequired   1     4.98      5772     5170.9 0.0256835 *
## ITEMNAME      1   327.88      5771     4843.0 < 2.2e-16 ***
## HandTufted    1     0.11      5770     4842.9 0.7374292
## Durry         1     6.91      5769     4836.0 0.0085843 **
## DoubleBack    1   159.91      5768     4676.1 < 2.2e-16 ***
## HandWoven     1   101.80      5767     4574.3 < 2.2e-16 ***
## Knotted       1     0.49      5766     4573.8 0.4824902
## Jacquard      1    55.01      5765     4518.8 1.199e-13 ***
## Handloom      1    45.19      5764     4473.6 1.792e-11 ***
## Other         0     0.00      5764     4473.6
## ShapeName     1     5.10      5763     4468.5 0.0239327 *
## REC           1     1.05      5762     4467.5 0.3054274
## Round         0     0.00      5762     4467.5
## Square        0     0.00      5762     4467.5
## AreaFt        1   829.77      5761     3637.7 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

fitted.results <- predict(mylogit,mydata)

fitted.results <- ifelse(fitted.results > 0.5,1,0)

misClasificError <- mean(fitted.results != mydata$OrderConversion)
print(paste('Accuracy',1-misClasificError))

## [1] "Accuracy 0.877183878221761"
```

Hence, we can conclude that the accuracy of Logistic Regression Model on CHAMPO Sample dataset is around 88%

--END OF LOGISTIC REGRESSION MODEL--

CHAMPO CARPETS: NEURAL NETWORK MODEL ON SAMPLE DATA

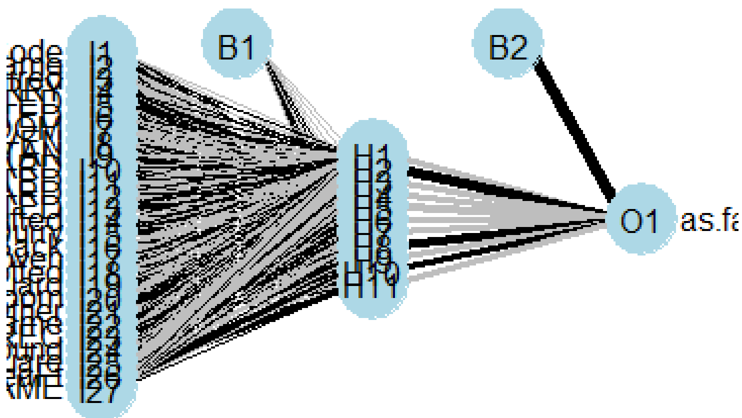
```
Data <- read_excel("C:/Users/Bodake/Desktop/UIC Sem 2/Data Mining IDS-572/R/s
ample.xlsx")
sum(is.na(Data)) #checking for NA's

## [1] 273

Data <- na_mean(Data)#Replacing the NA's with mean
sum(is.na(Data)) #checking if NA's were substitued correctly

## [1] 0

Data$Target <- as.factor(ifelse(Data$OrderConversion>0, "Conversion", "No_Con
version"))
df <- as.data.frame(Data)
df <- subset (df, select = -c(USA, UK, Italy, Belgium, Romania, Australia, In
dia,OrderConversion)) #Removing countries as they are already included in `Co
untryName` coloumn.
str(df)

## 'data.frame':    5820 obs. of  18 variables:
##  $ CustomerCode: chr  "A-11" "A-11" "A-11" "A-11" ...
##  $ CountryName : chr  "USA" "USA" "USA" "USA" ...
##  $ QtyRequired : num  2 2 2 1 5 5 5 5 11 2 ...
##  $ ITEMNAME    : chr  "DURRY" "DOUBLE BACK" "JACQUARD" "JACQUARD" ...
##  $ HandTufted  : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ Durry       : num  1 0 0 0 0 0 0 0 0 0 ...
##  $ DoubleBack  : num  0 1 0 0 1 1 1 1 0 0 ...
##  $ HandWoven   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Knotted     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Jacquard    : num  0 0 1 1 0 0 0 0 1 0 ...
##  $ Handloom    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Other       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ ShapeName   : chr  "REC" "REC" "REC" "REC" ...
##  $ REC         : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ Round       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Square      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ AreaFt      : num  4 4 4 42.3 1 ...
##  $ Target      : Factor w/ 2 levels "Conversion","No_Conversion": 2 2 2 1
2 2 2 2 2 2 ...
```

Now that we have cleaned our sample data by replacing the null values and removing redundant data, we can get started with implementing Neural Network model on our dataset

```
myscale <- function(x) {
  (x - min(x)) / (max(x) - min(x))
```

```
}

df <- df %>% mutate_if(is.numeric, myscale)
#categorical variables remain untouched

set.seed(1234)
df$CustomerCode<- as.numeric(factor(as.matrix(df$CustomerCode)))
df$CountryName<- as.numeric(factor(as.matrix(df$CountryName)))
df$ITEM_NAME<- as.numeric(factor(as.matrix(df$ITEMNAME)))
df$ShapeName<- as.numeric(factor(as.matrix(df$ShapeName)))

ind <- sample(2, nrow(df), replace = T, prob = c(0.7, 0.3))
train <- df[ind == 1, ]
test <- df[ind == 2, ]


#Neural Net Model


#str(df)
#summary(df)

sapply(lapply(df, unique), length)

## CustomerCode  CountryName  QtyRequired     ITEMNAME   HandTufted         Du
rry
##           34           14           41           11            2
2
##   DoubleBack    HandWoven      Knotted     Jacquard     Handloom         Ot
her
##            2            2            2            2            2
2
##    ShapeName          REC        Round       Square       AreaFt        Tar
get
##            3            2            2            2          139
2
##    ITEM_NAME
##           11

neuralnModel <- nnet(as.factor(Target) ~ ., data = train,hidden=3 ,linout = F
,size = 11, decay = 0.01,
                     maxit = 1000)

## # weights:  320
## initial  value 2187.883744
## iter  10 value 1780.107594
## iter  20 value 1672.149351
## iter  30 value 1578.914927
## iter  40 value 1474.603456
## iter  50 value 1302.278617
```

```
## iter  60 value 1262.190067
## iter  70 value 1235.031910
## iter  80 value 1206.816505……
…………
## stopped after 1000 iterations

plotnet(neuralnModel)
```




```
neuraln.preds = predict(neuralnModel, test)
neuraln.preds.class = as.factor(predict(neuralnModel, test, type = "class"))
#neuraln.preds.class
CM <- table(neuraln.preds.class, test$Target, dnn = c("predicted","actual"))
print(CM)

##                  actual
## predicted         Conversion No_Conversion
##   Conversion             239            35
##   No_Conversion          106          1334
```

We print our Confusion matrix and observe the values

```
error_metric = function(CM)
{
  TN = CM[1,1]
  TP = CM[2,2]
  FN = CM[1,2]
  FP = CM[2,1]
  recall = (TP)/(TP+FN)
```

```
  precision =(TP)/(TP+FP)
  falsePositiveRate = (FP)/(FP+TN)
  falseNegativeRate = (FN)/(FN+TP)
  error =(FP+FN)/(TP+TN+FP+FN)
  modelPerf <- list("precision" = precision,
                    "recall" = recall,
                    "falsepositiverate" = falsePositiveRate,
                    "falsenegativerate" = falseNegativeRate,
                    "error" = error)
  return(modelPerf)
}
outPutlist <- error_metric(CM)
library(plyr)

df <- ldply(outPutlist, data.frame)
setNames(df, c("", "Values"))

##                         Values
## 1         precision 0.92638889
## 2            recall 0.97443389
## 3 falsepositiverate 0.30724638
## 4 falsenegativerate 0.02556611
## 5             error 0.08226371
```

Now that we have successfully executed the Neural Network model on our dataset we get the following information

Precision is about 92.31% Recall is around 97.37% FPR is around 32.1% FNR is around 2.62% and Error rate is around 8.57%

We can calculate Accuracy as 1-error i.e., 1-0.082 = 0.918 = 91.8%

Hence, we can conclude that the Accuracy of Neural Network Model on our Dataset is around 91.8%

--END OF NEURAL NETWORK MODEL--

4. Discuss clustering algorithms that can be used for segmenting Champo Carpets's customers. Please justify your choices. Discuss what distance and similarity measures is suitable in this case.

Answer:

Clustering is a type of Unsupervised learning. Clustering algorithms are used to make small clusters for large dataset. When these clusters are made the items within clusters are like each other and close to each other in distance. The similarity should be high within each cluster and there should be low similarity between different clusters.

The different types of clustering methods are:

- Partition-based Algorithms (K-means algorithm)
- Hierarchy-based Algorithms

This distance can be calculated using various methods such as:

- Euclidian method
- Distance method

For our assignment we have used the Partition-based (K -means clustering) algorithm to segment the customers into various groups. It is the best choice when our dataset contains categorical variables. It is also less complex, fast, and highly scalable. We haven't used Hierarchical clustering as the as our data set does not contain variables that are hierarchical.

To calculate the distance, we have used Euclidian method over Manhattan as in Euclidian method the distance between any two objects is not dependent or affected by addition of any other new objects. When utilizing the Manhattan distance metric, the distortion is greater than when using the Euclidean distance metric. And it's well known that the optimal grouping of data-items is achieved with the least amount of distortion.

5. Discuss the data strategy for building customer segmentation using clustering. What are the benefits Champo Carpets can expect from clustering? Hint: Data strategy should clearly identify the data that should be used and how it should be used, including any feature engineering that may be performed before the model building.

Answer:

For customer segmentation we have implemented the K means clustering algorithm on the dataset. We used the sheet named "Data for clustering" under the Champo Carpets dataset as it consists of only unique records. Hence, we don't need to perform any data cleaning on it before processing it. To begin with we will first need to omit null values if any to begin with clustering. The next step would be to normalise the data. We can use the min-max function to do so. The next step would be to convert all the variables to numeric type which is a requirement of performing clustering. After this, we need to create clusters by using the 'kmean' function by specifying the dataset we are using, number of clusters we want to create(using ‘centres’ parameter) and number of centroids we want to start with(using the ‘nstart’ parameter). The desired clusters can be now visualised using the ‘fviz_cluster’ function under the ‘factoextra’ library. The next step would be to make a scree plot for various values of k and use both ‘WSS’ and ‘Silhouette’ methods to evaluate the model. Once the scree plot is created for both evaluation measures, we can select the best value for k using the elbow method. Once we have decided the best value of k, we can create a model using that value for k to get the best possible results.

CHAMPO CARPETS: K-MEANS CLUSTERING

```
library(dplyr)

library(readxl)
library(tidyr)
library(ggplot2)

library(ISLR)
library(rpart)
library(rpart.plot)
library(psych)

library(ggplot2)
library(Rcpp)
library(fastDummies)

library(factoextra)

library(klaR)

library(purrr)
library(ggplot2)
library(gridExtra)

Clustering_data <- read_excel("Clustering_data.xlsx")
View(Clustering_data)
mydata <- Clustering_data
mydata

mydata$Row_Labels  <- as.numeric(factor(as.matrix(mydata$Row_Labels)))
summary(mydata)

##    Row_Labels Sum of QtyRequired Sum of TotalArea    Sum of Amount
##  Min.   : 1   Min.   :     2     Min.   :     1.35   Min.   :      329
##  1st Qu.:12   1st Qu.:   565     1st Qu.:   376.77   1st Qu.:    39701
##  Median :23   Median :  1566     Median :  2120.00   Median :   116778
##  Mean   :23   Mean   : 12978     Mean   : 13056.59   Mean   :   698210
##  3rd Qu.:34   3rd Qu.: 11146     3rd Qu.:  8451.56   3rd Qu.:   426626
##  Max.   :45   Max.   :183206     Max.   :209725.22   Max.   :11341053
##      DURRY            HANDLOOM       DOUBLE BACK        JACQUARD
##  Min.   :     0   Min.   :   0.0   Min.   :   0.0   Min.   :  0.00
##  1st Qu.:     0   1st Qu.:   0.0   1st Qu.:   0.0   1st Qu.:  0.00
##  Median :   289   Median :   0.0   Median :   0.0   Median :  0.00
##  Mean   :  7103   Mean   : 185.5   Mean   : 407.9   Mean   : 89.42
##  3rd Qu.:  1560   3rd Qu.:   0.0   3rd Qu.: 175.0   3rd Qu.: 72.00
##  Max.   :139618   Max.   :3673.0   Max.   :5439.0   Max.   :714.00
##   HAND TUFTED      HAND WOVEN         KNOTTED          GUN TUFTED
##  Min.   :    0   Min.   :    0.0   Min.   :   0.0   Min.   :  0.000
##  1st Qu.:    0   1st Qu.:    0.0   1st Qu.:   0.0   1st Qu.:  0.000
```

```
##  Median :  510   Median :    0.0   Median :   0.0   Median :  0.000
##  Mean   : 3651   Mean   :  867.7   Mean   : 365.8   Mean   :  8.133
##  3rd Qu.: 3544   3rd Qu.:  269.0   3rd Qu.:  18.0   3rd Qu.:  0.000
##  Max.   :60685   Max.   :14314.0   Max.   :9502.0   Max.   :195.000
##  Powerloom Jacquard  INDO TEBETAN
##  Min.   :   0.0     Min.   : 0.0000
##  1st Qu.:   0.0     1st Qu.: 0.0000
##  Median :   0.0     Median : 0.0000
##  Mean   : 216.7     Mean   : 0.7111
##  3rd Qu.:   0.0     3rd Qu.: 0.0000
##  Max.   :9753.0     Max.   :20.0000
```

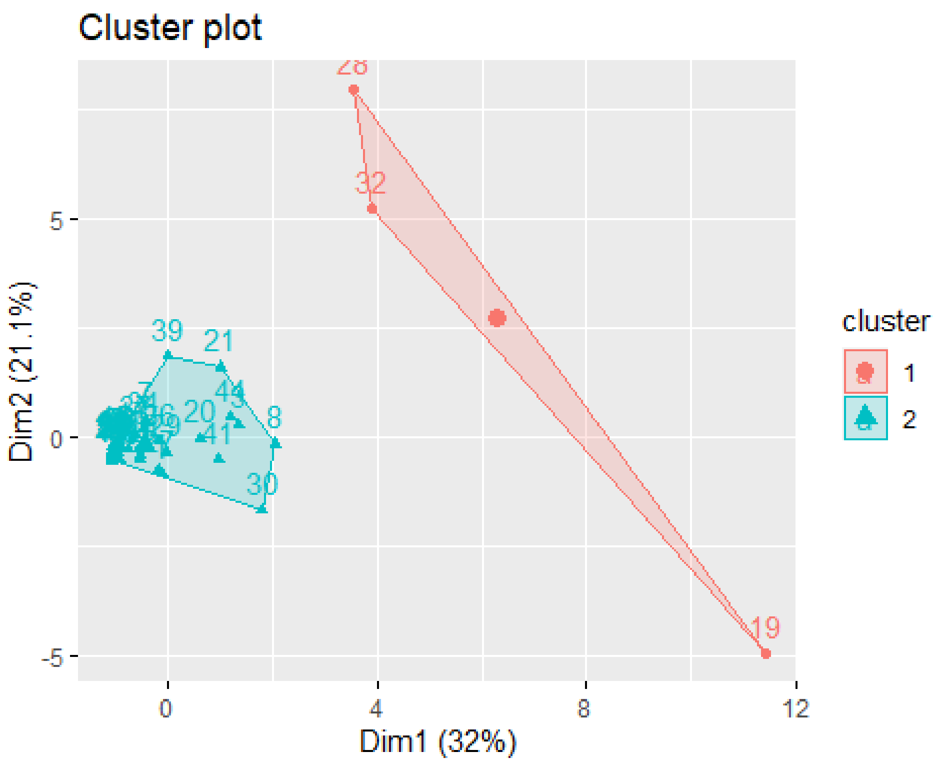
```
sum(is.na(mydata))
```

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

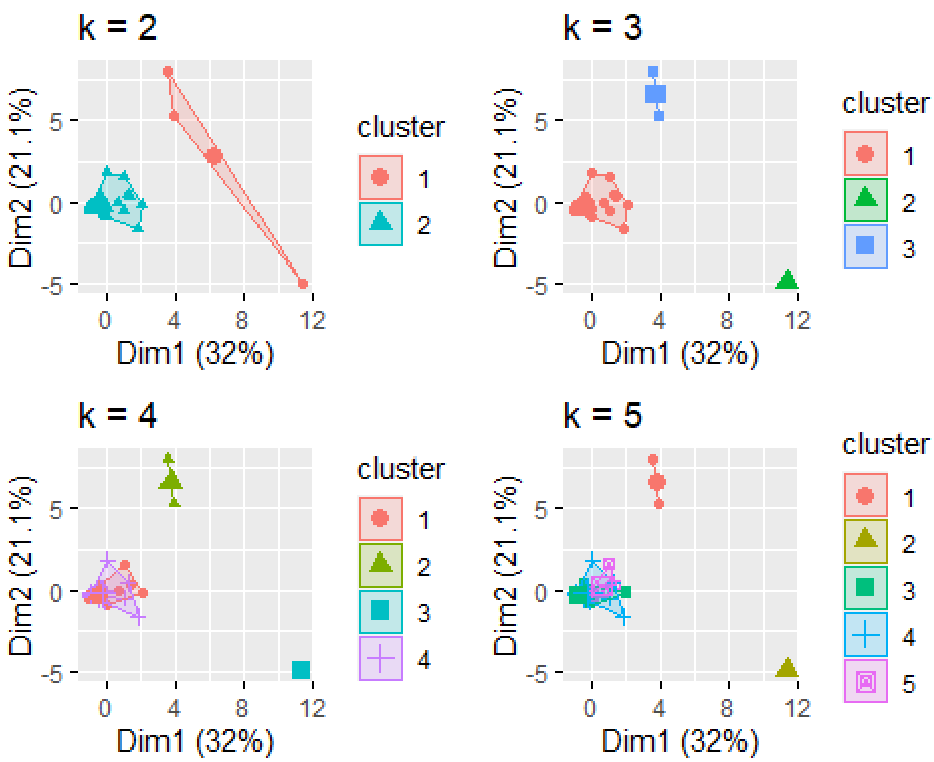
```
mydata <- na.omit(mydata)

myscale <- function(x) {
  (x - min(x)) / (max(x) - min(x))
}
df <- mydata %>% mutate_if(is.numeric, myscale)

kmModel <- kmeans(df, centers = 2, nstart = 100)

kmModel
```

```
## K-means clustering with 2 clusters of sizes 3, 42
## 
## Cluster means:
##   Row_Labels Sum of QtyRequired Sum of TotalArea Sum of Amount      DURRY
## 1  0.5757576          0.45163133       0.49095247    0.25954037 0.39638394
## 2  0.4945887          0.04362532       0.03162803    0.04739456 0.02619529
##     HANDLOOM DOUBLE BACK  JACQUARD HAND TUFTED HAND WOVEN     KNOTTED
## 1 0.44432344  0.62082491 0.4495798  0.17390898 0.26605654 0.460534624
## 2 0.02238342  0.03600977 0.1020742  0.05203569 0.04594403 0.008346614
##   GUN TUFTED Powerloom Jacquard INDO TEBETAN
## 1 0.33333333          0.3333333   0.00000000
## 2 0.02087912          0.0000000   0.03809524
## 
## Clustering vector:
##  [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2
2 2 2
## [39] 2 2 2 2 2 2 2
## 
## Within cluster sum of squares by cluster:
## [1]  4.805835 12.093430
##  (between_SS / total_SS =  21.8 %)
## 
## Available components:
```

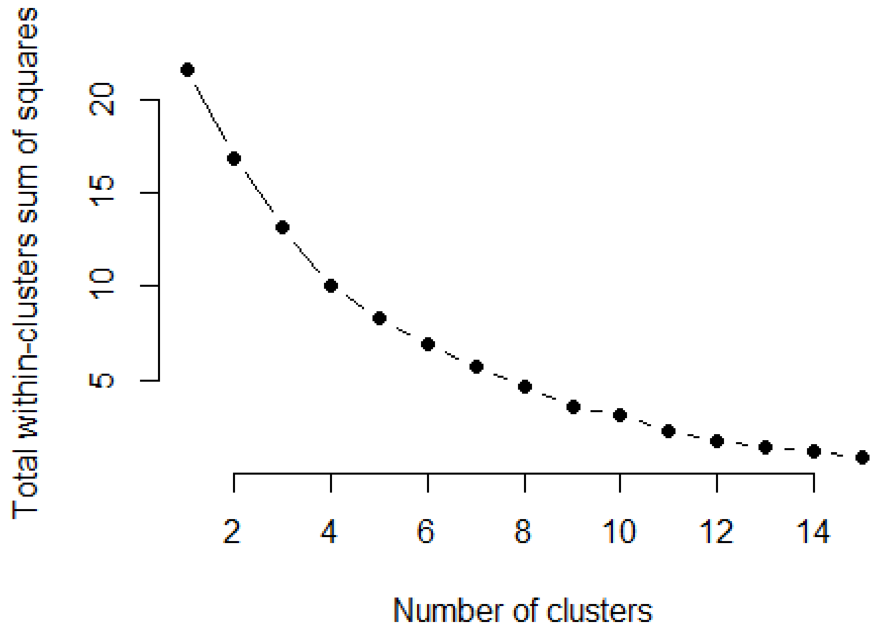
```
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withi
nss"
## [6] "betweenss"    "size"         "iter"         "ifault"

model <- lm(df$Row_Labels ~ .,  data = df)
cooksD <- cooks.distance(model)
influential <- cooksD[(cooksD > (3 * mean(cooksD, na.rm = TRUE)))]
influential

##     <NA>       31       32       44
##       NA 67.03154 14.44764 14.54626

kmModel$cluster

##  [1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2
2 2 2
## [39] 2 2 2 2 2 2 2

kmModel$centers

##   Row_Labels Sum of QtyRequired Sum of TotalArea Sum of Amount      DURRY
## 1  0.5757576         0.45163133       0.49095247    0.25954037 0.39638394
## 2  0.4945887         0.04362532       0.03162803    0.04739456 0.02619529
##     HANDLOOM DOUBLE BACK  JACQUARD HAND TUFTED HAND WOVEN     KNOTTED
## 1 0.44432344  0.62082491 0.4495798  0.17390898 0.26605654 0.460534624
## 2 0.02238342  0.03600977 0.1020742  0.05203569 0.04594403 0.008346614
##   GUN TUFTED Powerloom Jacquard INDO TEBETAN
## 1 0.33333333          0.3333333   0.00000000
## 2 0.02087912          0.0000000   0.03809524

kmModel$withinss

## [1]  4.805835 12.093430

kmModel$tot.withinss

## [1] 16.89927

kmModel$betweenss

## [1] 4.717582

kmModel$size

## [1]  3 42

kmModel$iter

## [1] 1

fviz_cluster(kmModel, data = df)
```



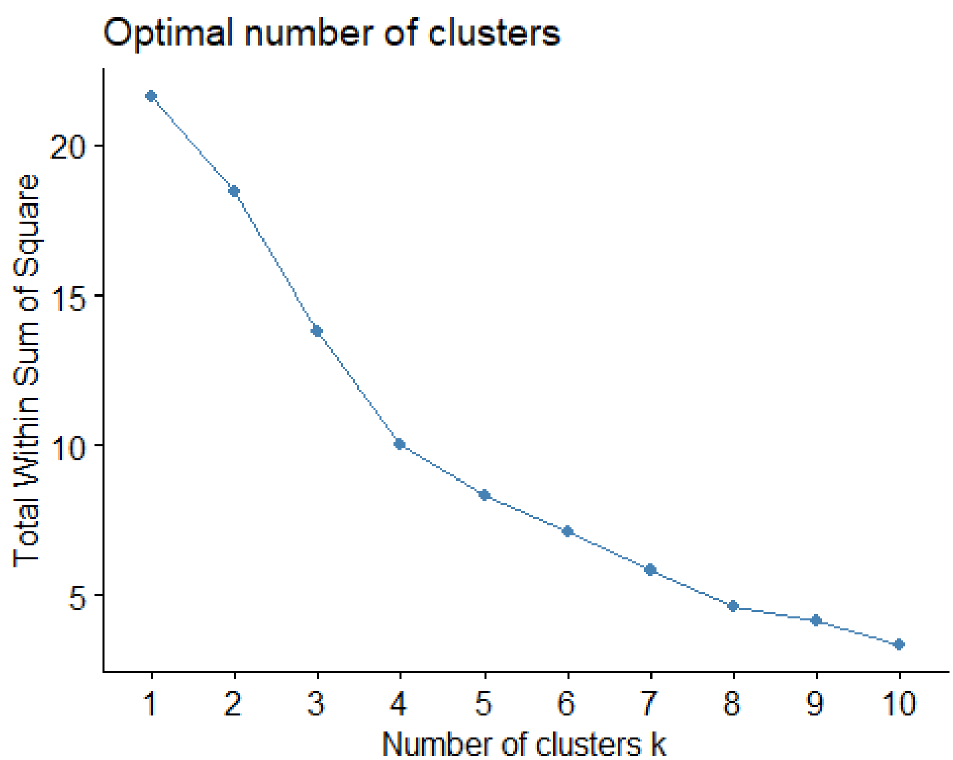

```r
kmMode2 <- kmeans(df, centers = 3, nstart = 100)
kmMode3 <- kmeans(df, centers = 4, nstart = 100)
kmMode4 <- kmeans(df, centers = 5, nstart = 100)

fp1 <- fviz_cluster(kmModel, geom = "point", data = df) + ggtitle("k = 2")
fp2 <- fviz_cluster(kmMode2, geom = "point", data = df) + ggtitle("k = 3")
fp3 <- fviz_cluster(kmMode3, geom = "point", data = df) + ggtitle("k = 4")
fp4 <- fviz_cluster(kmMode4, geom = "point", data = df) + ggtitle("k = 5")

library(gridExtra)
grid.arrange(fp1, fp2, fp3, fp4, nrow = 2)
```



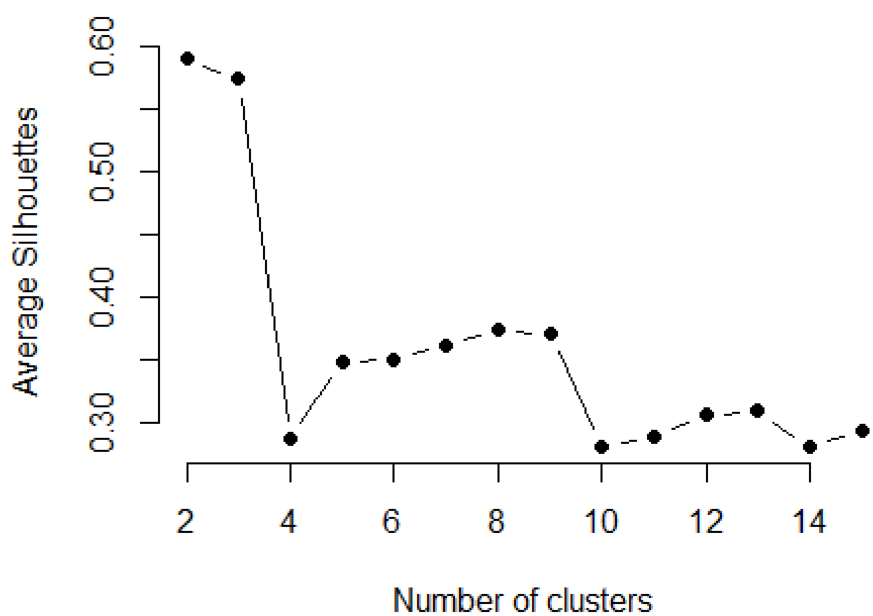

```
set.seed(123)
wss <- function(k) {
  kmeans(df, centers = k, nstart = 100)$tot.withinss
}

k.values <- 1:15
wss_values <- map_dbl(k.values, wss)

plot(k.values, wss_values,
     type="b", pch = 19, frame = FALSE,
     xlab="Number of clusters",
     ylab="Total within-clusters sum of squares")
```



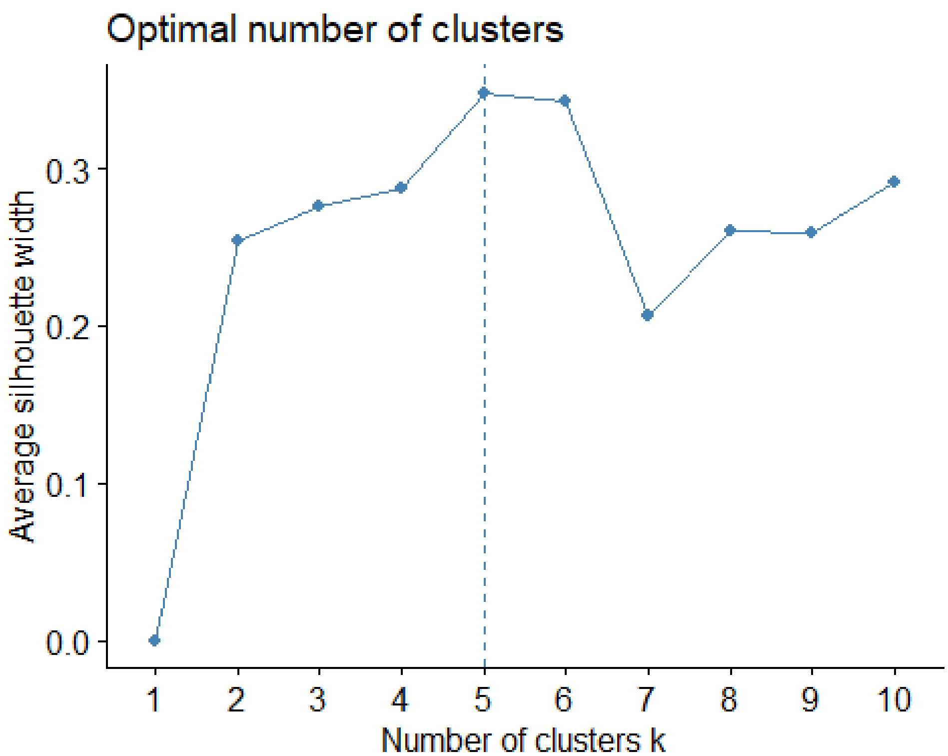

```
set.seed(123)
fviz_nbclust(df, kmeans, method = "wss")
```



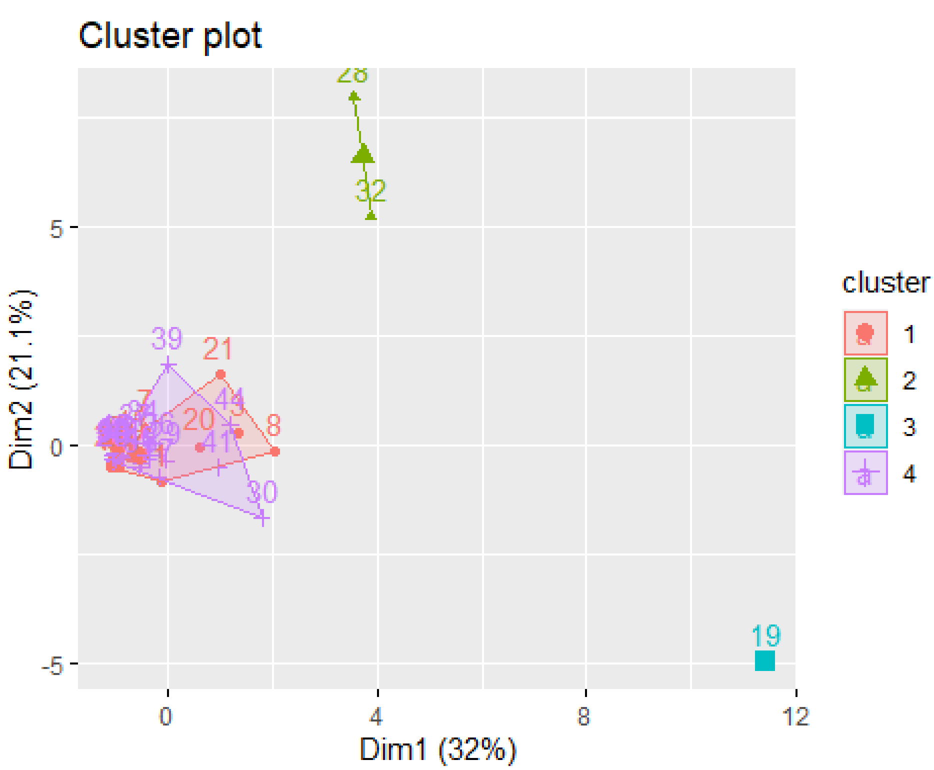
```
library(cluster)
avgsil <- function(k) {
  kmModel <- kmeans(df, centers = k, nstart = 100)
  ss <- silhouette(kmModel$cluster, dist(df))
  mean(ss[, 3])
}

k.values <- 2:15
avgsil_values <- map_dbl(k.values, avgsil)
map_dbl

## function (.x, .f, ...)
## {
##     .f <- as_mapper(.f, ...)
##     .Call(map_impl, environment(), ".x", ".f", "double")
## }
## <bytecode: 0x000000002790ad90>
## <environment: namespace:purrr>

plot(k.values, avgsil_values,
     type = "b", pch = 19, frame = FALSE,
     xlab = "Number of clusters",
     ylab = "Average Silhouettes")
```




```
fviz_nbclust(df, kmeans, method = "silhouette")
```




```
fviz_cluster(kmMode3, data = df)
```



```
df %>%
  mutate(Cluster = kmMode3$cluster) %>%
  group_by(Cluster) %>%
  summarise_all("mean")
```

```
## # A tibble: 4 x 15
##   Cluster Row_Labels `Sum of QtyRequir~ `Sum of TotalAre~ `Sum of Amount`
DURRY
##     <int>      <dbl>              <dbl>             <dbl>           <dbl>
<dbl>
## 1       1      0.255             0.0291            0.0409          0.0245
0.0163
## 2       2      0.659             0.177             0.690           0.222
0.0946
## 3       3      0.409             1                 0.0930          0.335
1
## 4       4      0.785             0.0612            0.0204          0.0751
0.0382
## # ... with 9 more variables: HANDLOOM <dbl>, DOUBLE BACK <dbl>, JACQUARD <
dbl>,
## #   HAND TUFTED <dbl>, HAND WOVEN <dbl>, KNOTTED <dbl>, GUN TUFTED <dbl>,
## #   Powerloom Jacquard <dbl>, INDO TEBETAN <dbl>
```

```
kmodes(df, modes = 4, iter.max = 10, weighted = FALSE)
```

```
## Warning in kmodes(df, modes = 4, iter.max = 10, weighted = FALSE): data ha
s
## numeric coloumns with more than 30 different levels!
```

```
## K-modes clustering with 4 clusters of sizes 15, 28, 1, 1
##
## Cluster modes:
##   Row_Labels Sum of QtyRequired Sum of TotalArea Sum of Amount       DURRY
## 1 0.13636364        0.003285954      0.002515319   0.004974449 0.000000000
## 2 0.00000000        0.000000000      0.000000000   0.000000000 0.000000000
## 3 0.29545455        0.006309906      0.013449351   0.010273535 0.002062771
## 4 0.04545455        0.103278313      0.255690036   0.140357085 0.025677205
##   HANDLOOM DOUBLE BACK JACQUARD HAND TUFTED HAND WOVEN   KNOTTED GUN TUFTE
D
## 1        0  0.00000000        0 0.002191645 0.00125751 0.0000000
0
## 2        0  0.00000000        0 0.000000000 0.00000000 0.0000000
0
## 3        0  0.01176687        0 0.013281701 0.00000000 0.0000000
0
## 4        0  0.03217503        1 0.193062536 0.14782730 0.0649337
0
##   Powerloom Jacquard INDO TEBETAN
## 1                  0            0
## 2                  0            0
## 3                  0            0
```

```
## 4                  0            0
##
## Clustering vector:
##  [1] 2 2 4 2 2 2 1 1 2 2 2 2 2 3 2 2 2 1 1 1 1 2 2 2 2 2 2 1 1 2 1 1 2 2 2
1 1 1
## [39] 1 2 1 2 2 2 2
##
## Within cluster simple-matching distance by cluster:
## [1] 132 172   0   0
##
## Available components:
## [1] "cluster"    "size"       "modes"      "withindiff" "iterations"
## [6] "weighted"
```

CHAMPO CARPETS: RECOMMENDATION SYSTEM FOR CHAMPO CARPETS

```
RSD <- read_excel("C:/Users/Bodake/Desktop/UIC Sem 2/Data Mining IDS-
572/R/RSD.xlsx")
mydata <- RSD
mydata<- as.data.frame(mydata)
sum(is.na(mydata))

## [1] 0

#str(mydata)
mydata[sapply(mydata, is.character)] <- lapply(mydata[sapply(mydata,
is.character)],as.factor)

df <-mydata[2:21]
str(mydata)

## 'data.frame':    20 obs. of  21 variables:
##  $ Customer   : Factor w/ 20 levels "A-11","A-9","C-1",..: 6 14 10 2 4 8
12 20 3 18 ...
##  $ HandTufted : num  26612 2352 2697 11716 3816 ...
##  $ DoubleWowen: num  3000 5340 3085 2116 14314 ...
##  $ Durry      : num  139618 25997 412 3585 37042 ...
##  $ DoubleBack : num  0 4691 5439 175 0 ...
##  $ Knotted    : num  0 9502 3626 617 0 ...
##  $ Jacquared  : num  550 353 60 714 0 231 0 100 0 0 ...
##  $ Handloom   : num  3673 138 1085 0 0 ...
##  $ Other      : num  9753 0 245 0 0 ...
##  $ Rectangle  : num  132756 47951 16619 18923 55172 ...
##  $ Square     : num  0 315 30 0 0 0 0 0 2 13 ...
##  $ Round      : num  50540 107 0 0 0 ...
##  $ Purple     : num  0 0 0 0 0 55 0 0 0 0 ...
##  $ Gray       : num  0 1863 1216 273 2007 ...
##  $ Navy       : num  16235 0 53 1843 1601 ...
##  $ PINK       : num  0 0 138 552 0 316 0 875 0 0 ...
##  $ BLUE       : num  15537 6218 69 705 0 ...
##  $ BLUSHPINK  : num  0 0 0 0 0 0 0 250 0 0 ...
##  $ NEUTRAL    : num  0 0 0 1832 0 ...
##  $ TAN        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ NAVY       : num  16235 0 53 1843 1601 ...

getCosine <- function(x,y)
{
  this.cosine <- sum(x*y) / (sqrt(sum(x*x)) * sqrt(sum(y*y)))
  return(this.cosine)
}
df.similarity  <- matrix(NA,
nrow=ncol(df),ncol=ncol(df),dimnames=list(colnames(df),colnames(df)))
#df.similarity
```

```
for(i in 1:ncol(df)) {

  for(j in 1:ncol(df)) {

    df.similarity[i,j] <- getCosine(as.matrix(df[i]),as.matrix(df[j]))
  }
}


df.similarity <- as.data.frame(df.similarity)

df.neighbours <- matrix(NA,
nrow=ncol(df.similarity),ncol=11,dimnames=list(colnames(df.similarity)))
for(i in 1:ncol(df))
{
  df.neighbours[i,] <-
(t(head(n=11,rownames(df.similarity[order(df.similarity[,i],decreasing=TRUE),
][i]))))
}
df.neighbours[1:20]

##  [1] "HandTufted"  "DoubleWowen" "Durry"       "DoubleBack"  "Knotted"
##  [6] "Jacquared"   "Handloom"    "Other"       "Rectangle"   "Square"
## [11] "Round"       "Purple"      "Gray"        "Navy"        "PINK"
## [16] "BLUE"        "BLUSHPINK"   "NEUTRAL"     "TAN"         "Navy"
```

Creating a Helper Function

```
getScore <- function(history, similarities)
{
  x <- sum(history*similarities)/sum(similarities)

}
```

Creating a Matrix

```
holder <- matrix(NA,nrow=nrow(mydata),ncol=ncol(mydata)-
1,dimnames=list((mydata$Customer),colnames(mydata[-1])))
for(i in 1:nrow(holder))
{

  for(j in 1:ncol(holder))
  {

    Customer <- rownames(holder)[i]
    product <- colnames(holder)[j]

    if(as.integer(mydata[mydata$Customer==Customer,product]) == 1)
    {
      holder[i,j]<-""
```

```
    } else {

      topN<-
((head(n=11,(df.similarity[order(df.similarity[,product],decreasing=TRUE),][p
roduct]))))
      topN.names <- as.character(rownames(topN))
      topN.similarities <- as.numeric(topN[,1])

      topN.similarities<-topN.similarities[-1]
      topN.names<-topN.names[-1]

      topN.purchases<- mydata[,c("Customer",topN.names)]
      topN.CustomerPurchases<-
topN.purchases[topN.purchases$Customer==Customer,]
      topN.CustomerPurchases <-
as.numeric(topN.CustomerPurchases[!(names(topN.CustomerPurchases) %in%
c("Customer"))])


      holder[i,j]<-
getScore(similarities=topN.similarities,history=topN.CustomerPurchases)

    }
  }
}

mydata.Customer.scores <- holder


topN<-
((head(n=11,(df.similarity[order(df.similarity[,product],decreasing=TRUE),][p
roduct]))))
topN.names <- as.character(rownames(topN))
topN.similarities <- as.numeric(topN[,1])

topN.similarities<-topN.similarities[-1]
topN.names<-topN.names[-1]

topN.purchases<- mydata[,c("Customer",topN.names)]
topN.CustomerPurchases<-topN.purchases[topN.purchases$Customer==Customer,]
topN.CustomerPurchases <-
as.numeric(topN.CustomerPurchases[!(names(topN.CustomerPurchases) %in%
c("Customer"))])
holder[i,j]<-
getScore(similarities=topN.similarities,history=topN.CustomerPurchases)

mydata.Customer.scores.holder <- matrix(NA,
nrow=nrow(mydata.Customer.scores),ncol=100,dimnames=list(rownames(mydata.Cust
omer.scores)))
```

```
for(i in 1:nrow(mydata.Customer.scores))
{
  mydata.Customer.scores.holder[i,] <-
names(head(n=100,(mydata.Customer.scores[,order(mydata.Customer.scores[i,],decreasing=TRUE)])[i,]))
}

mydata.Customer.scores.holder
```

Following are the recommendations for the Consumers:

```
##       [,1]          [,2]          [,3]          [,4]          [,5]
## H-2   "Other"       "Handloom"    "BLUE"        "Navy"        "NAVY"
## P-5   "Navy"        "NAVY"        "Other"       "TAN"         "Round"
## M-1   "Rectangle"   "TAN"         "Square"      "DoubleWowen" "Knotted"
## A-9   "TAN"         "BLUE"        "Navy"        "NAVY"        "Round"
## C-2   "Square"      "Durry"       "PINK"        "Knotted"     "Rectangle"
## JL    "TAN"         "DoubleWowen" "Gray"        "Square"      "Navy"
## N-1   "PINK"        "Square"      "Knotted"     "Purple"      "DoubleBack"
## T-5   "BLUSHPINK"   "Jacquared"   "Gray"        "Durry"       "PINK"
## C-1   "Durry"       "Handloom"    "Round"       "DoubleWowen" "Other"
## T-2   "DoubleWowen" "Durry"       "Other"       "Handloom"    "Round"
## I-2   "Knotted"     "DoubleBack"  "Rectangle"   "TAN"         "BLUE"
## PD    "Purple"      "NEUTRAL"     "DoubleWowen" "BLUE"        "Other"
## L-5   "DoubleWowen" "Other"       "HandTufted"  "Round"       "Navy"
## M-2   "PINK"        "Square"      "Rectangle"   "Purple"      "Knotted"
## RC    "DoubleWowen" "BLUE"        "Navy"        "NAVY"        "Square"
## P-4   "DoubleWowen" "HandTufted"  "Round"       "Navy"        "NAVY"
## T-4   "Durry"       "HandTufted"  "BLUSHPINK"   "Jacquared"   "Gray"
## PC    "TAN"         "BLUE"        "Navy"        "NAVY"        "Durry"
## A-11  "Other"       "Round"       "BLUE"        "Navy"        "NAVY"
## CC    "Round"       "Handloom"    "Other"       "DoubleWowen" "HandTufted"
##       [,6]         [,7]         [,8]          [,9]          [,10]
## H-2   "HandTufted" "Round"      "DoubleWowen" "Jacquared"   "TAN"
## P-5   "Handloom"   "BLUE"       "Jacquared"   "BLUSHPINK"   "Knotted"
## M-1   "Gray"       "Purple"     "BLUE"        "DoubleBack"  "Durry"
## A-9   "Handloom"   "Other"      "DoubleWowen" "Durry"       "HandTufted"
## C-2   "DoubleBack" "Purple"     "NEUTRAL"     "DoubleWowen" "BLUE"
## JL    "NAVY"       "HandTufted" "BLUE"        "Handloom"    "Knotted"
## N-1   "TAN"        "BLUE"       "Navy"        "NAVY"        "Other"
## T-5   "Square"     "Knotted"    "Rectangle"   "DoubleBack"  "NEUTRAL"
## C-1   "HandTufted" "NEUTRAL"    "Gray"        "Square"      "BLUSHPINK"
## T-2   "HandTufted" "Square"     "Knotted"     "Gray"        "Purple"
## I-2   "Navy"       "NAVY"       "Durry"       "DoubleWowen" "Other"
## PD    "Round"      "Navy"       "NAVY"        "Handloom"    "HandTufted"
## L-5   "NAVY"       "Handloom"   "BLUE"        "BLUSHPINK"   "Gray"
## M-2   "DoubleBack" "TAN"        "BLUE"        "Navy"        "NAVY"
## RC    "Handloom"   "Other"      "HandTufted"  "Round"       "Durry"
```

```
## P-4  "Handloom"   "BLUE"        "Other"       "Jacquared"   "TAN"
## T-4  "PINK"       "NEUTRAL"     "Square"      "Rectangle"   "Knotted"
## PC   "Round"      "Handloom"    "Other"       "DoubleWowen" "NEUTRAL"
## A-11 "HandTufted" "Jacquared"   "DoubleWowen" "Durry"       "TAN"
## CC   "NEUTRAL"    "BLUSHPINK"   "Jacquared"   "Gray"        "Rectangle"
##      [,11]        [,12]          [,13]         [,14]         [,15]
## H-2  "Rectangle"  "Durry"        "Gray"        "BLUSHPINK"   "Square"
## P-5  "Durry"      "DoubleBack"   "PINK"        "Purple"      "Rectangle"
## M-1  "HandTufted" "Navy"         "NAVY"        "Round"       "Other"
## A-9  "Jacquared"  "BLUSHPINK"    "NEUTRAL"     "Gray"        "PINK"
## C-2  "Other"      "HandTufted"   "Handloom"    "Navy"        "NAVY"
## JL   "Other"      "Round"        "Durry"       "BLUSHPINK"   "Jacquared"
## N-1  "Round"      "Handloom"     "Durry"       "HandTufted"  "DoubleWowen"
## T-5  "Purple"     "DoubleWowen"  "TAN"         "BLUE"        "Other"
## C-1  "PINK"       "Jacquared"    "Purple"      "DoubleBack"  "Rectangle"
## T-2  "Jacquared"  "PINK"         "BLUSHPINK"   "NEUTRAL"     "DoubleBack"
## I-2  "Round"      "HandTufted"   "Handloom"    "Gray"        "Jacquared"
## PD   "Gray"       "Jacquared"    "BLUSHPINK"   "TAN"         "Square"
## L-5  "Jacquared"  "TAN"          "Square"      "PINK"        "Knotted"
## M-2  "Handloom"   "Other"        "DoubleWowen" "Round"       "Durry"
## RC   "Gray"       "NEUTRAL"      "BLUSHPINK"   "PINK"        "Purple"
## P-4  "BLUSHPINK"  "Gray"         "Durry"       "Square"      "Rectangle"
## T-4  "Purple"     "DoubleBack"   "TAN"         "BLUE"        "Navy"
## PC   "BLUSHPINK"  "PINK"         "HandTufted"  "Gray"        "Jacquared"
## A-11 "Handloom"   "Gray"         "BLUSHPINK"   "Knotted"     "Square"
## CC   "Purple"     "Knotted"      "DoubleBack"  "TAN"         "BLUE"
##      [,16]        [,17]          [,18]        [,19]          [,20]
## H-2  "PINK"       "Knotted"      "NEUTRAL"    "DoubleBack"   "Purple"
## P-5  "NEUTRAL"    "DoubleWowen"  "Square"     "Gray"         "HandTufted"
## M-1  "PINK"       "BLUSHPINK"    "Handloom"   "Jacquared"    "NEUTRAL"
## A-9  "Square"     "Rectangle"    "Purple"     "Knotted"      "DoubleBack"
## C-2  "Round"      "Gray"         "TAN"        "Jacquared"    "BLUSHPINK"
## JL   "Purple"     "PINK"         "DoubleBack" "NEUTRAL"      "Rectangle"
## N-1  "Jacquared"  "Rectangle"    "NEUTRAL"    "BLUSHPINK"    "Gray"
## T-5  "Round"      "Handloom"     "Navy"       "NAVY"         "HandTufted"
## C-1  "Knotted"    "TAN"          "BLUE"       "Navy"         "NAVY"
## T-2  "Rectangle"  "TAN"          "BLUE"       "Navy"         "NAVY"
## I-2  "BLUSHPINK"  "NEUTRAL"      "Square"     "PINK"         "Purple"
## PD   "PINK"       "Knotted"      "Durry"      "Rectangle"    "DoubleBack"
## L-5  "Durry"      "Rectangle"    "DoubleBack" "Purple"       "NEUTRAL"
## M-2  "HandTufted" "BLUSHPINK"    "Gray"       "NEUTRAL"      "Jacquared"
## RC   "Jacquared"  "Knotted"      "DoubleBack" "Rectangle"    "TAN"
## P-4  "Knotted"    "PINK"         "DoubleBack" "Purple"       "NEUTRAL"
## T-4  "NAVY"       "Round"        "Other"      "DoubleWowen"  "Handloom"
## PC   "Purple"     "Rectangle"    "Knotted"    "DoubleBack"   "Square"
## A-11 "Purple"     "DoubleBack"   "Rectangle"  "PINK"         "NEUTRAL"
## CC   "Navy"       "NAVY"         "Durry"      "Square"       "PINK"
##      [,21]         [,22]          [,23]          [,24]          [,25]
## H-2  "Other"       "Handloom"     "BLUE"         "Navy"         "NAVY"
## P-5  "Navy"        "NAVY"         "Other"        "TAN"          "Round"
```

```
## M-1  "Rectangle"   "TAN"          "Square"       "DoubleWowen" "Knotted"
## A-9  "TAN"         "BLUE"         "Navy"         "NAVY"        "Round"
## C-2  "Square"      "Durry"        "PINK"         "Knotted"     "Rectangle"
## JL   "TAN"         "DoubleWowen" "Gray"         "Square"      "Navy"
## N-1  "PINK"        "Square"       "Knotted"      "Purple"      "DoubleBack"
## T-5  "BLUSHPINK"   "Jacquared"    "Gray"         "Durry"       "PINK"
## C-1  "Durry"       "Handloom"     "Round"        "DoubleWowen" "Other"
## T-2  "DoubleWowen" "Durry"        "Other"        "Handloom"    "Round"
## I-2  "Knotted"     "DoubleBack"  "Rectangle"    "TAN"         "BLUE"
## PD   "Purple"      "NEUTRAL"      "DoubleWowen" "BLUE"        "Other"
## L-5  "DoubleWowen" "Other"        "HandTufted"  "Round"       "Navy"
## M-2  "PINK"        "Square"       "Rectangle"    "Purple"      "Knotted"
## RC   "DoubleWowen" "BLUE"         "Navy"         "NAVY"        "Square"
## P-4  "DoubleWowen" "HandTufted"  "Round"        "Navy"        "NAVY"
## T-4  "Durry"       "HandTufted"  "BLUSHPINK"    "Jacquared"   "Gray"
## PC   "TAN"         "BLUE"         "Navy"         "NAVY"        "Durry"
## A-11 "Other"       "Round"        "BLUE"         "Navy"        "NAVY"
## CC   "Round"       "Handloom"     "Other"        "DoubleWowen" "HandTufted"
##      [,26]        [,27]         [,28]         [,29]         [,30]
## H-2  "HandTufted" "Round"       "DoubleWowen" "Jacquared"   "TAN"
## P-5  "Handloom"   "BLUE"        "Jacquared"   "BLUSHPINK"   "Knotted"
## M-1  "Gray"       "Purple"      "BLUE"        "DoubleBack"  "Durry"
## A-9  "Handloom"   "Other"       "DoubleWowen" "Durry"       "HandTufted"
## C-2  "DoubleBack" "Purple"      "NEUTRAL"     "DoubleWowen" "BLUE"
## JL   "NAVY"       "HandTufted" "BLUE"        "Handloom"    "Knotted"
## N-1  "TAN"        "BLUE"        "Navy"        "NAVY"        "Other"
## T-5  "Square"     "Knotted"     "Rectangle"   "DoubleBack"  "NEUTRAL"
## C-1  "HandTufted" "NEUTRAL"     "Gray"        "Square"      "BLUSHPINK"
## T-2  "HandTufted" "Square"      "Knotted"     "Gray"        "Purple"
## I-2  "Navy"       "NAVY"        "Durry"       "DoubleWowen" "Other"
## PD   "Round"      "Navy"        "NAVY"        "Handloom"    "HandTufted"
## L-5  "NAVY"       "Handloom"    "BLUE"        "BLUSHPINK"   "Gray"
## M-2  "DoubleBack" "TAN"         "BLUE"        "Navy"        "NAVY"
## RC   "Handloom"   "Other"       "HandTufted"  "Round"       "Durry"
## P-4  "Handloom"   "BLUE"        "Other"       "Jacquared"   "TAN"
## T-4  "PINK"       "NEUTRAL"     "Square"      "Rectangle"   "Knotted"
## PC   "Round"      "Handloom"    "Other"       "DoubleWowen" "NEUTRAL"
## A-11 "HandTufted" "Jacquared"   "DoubleWowen" "Durry"       "TAN"
## CC   "NEUTRAL"    "BLUSHPINK"   "Jacquared"   "Gray"        "Rectangle"
##      [,31]        [,32]          [,33]         [,34]         [,35]
## H-2  "Rectangle"  "Durry"        "Gray"        "BLUSHPINK"   "Square"
## P-5  "Durry"      "DoubleBack"   "PINK"        "Purple"      "Rectangle"
## M-1  "HandTufted" "Navy"         "NAVY"        "Round"       "Other"
## A-9  "Jacquared"  "BLUSHPINK"    "NEUTRAL"     "Gray"        "PINK"
## C-2  "Other"      "HandTufted"   "Handloom"    "Navy"        "NAVY"
## JL   "Other"      "Round"        "Durry"       "BLUSHPINK"   "Jacquared"
## N-1  "Round"      "Handloom"     "Durry"       "HandTufted" "DoubleWowen"
## T-5  "Purple"     "DoubleWowen" "TAN"         "BLUE"        "Other"
## C-1  "PINK"       "Jacquared"    "Purple"      "DoubleBack" "Rectangle"
## T-2  "Jacquared"  "PINK"         "BLUSHPINK"   "NEUTRAL"     "DoubleBack"
```

```
## I-2  "Round"      "HandTufted"  "Handloom"    "Gray"       "Jacquared"
## PD   "Gray"       "Jacquared"   "BLUSHPINK"   "TAN"        "Square"
## L-5  "Jacquared"  "TAN"         "Square"      "PINK"       "Knotted"
## M-2  "Handloom"   "Other"       "DoubleWowen" "Round"      "Durry"
## RC   "Gray"       "NEUTRAL"     "BLUSHPINK"   "PINK"       "Purple"
## P-4  "BLUSHPINK"  "Gray"        "Durry"       "Square"     "Rectangle"
## T-4  "Purple"     "DoubleBack"  "TAN"         "BLUE"       "Navy"
## PC   "BLUSHPINK"  "PINK"        "HandTufted"  "Gray"       "Jacquared"
## A-11 "Handloom"   "Gray"        "BLUSHPINK"   "Knotted"    "Square"
## CC   "Purple"     "Knotted"     "DoubleBack"  "TAN"        "BLUE"
##      [,36]        [,37]         [,38]        [,39]         [,40]
## H-2  "PINK"       "Knotted"     "NEUTRAL"    "DoubleBack"  "Purple"
## P-5  "NEUTRAL"    "DoubleWowen" "Square"     "Gray"        "HandTufted"
## M-1  "PINK"       "BLUSHPINK"   "Handloom"   "Jacquared"   "NEUTRAL"
## A-9  "Square"     "Rectangle"   "Purple"     "Knotted"     "DoubleBack"
## C-2  "Round"      "Gray"        "TAN"        "Jacquared"   "BLUSHPINK"
## JL   "Purple"     "PINK"        "DoubleBack" "NEUTRAL"     "Rectangle"
## N-1  "Jacquared"  "Rectangle"   "NEUTRAL"    "BLUSHPINK"   "Gray"
## T-5  "Round"      "Handloom"    "Navy"       "NAVY"        "HandTufted"
## C-1  "Knotted"    "TAN"         "BLUE"       "Navy"        "NAVY"
## T-2  "Rectangle"  "TAN"         "BLUE"       "Navy"        "NAVY"
## I-2  "BLUSHPINK"  "NEUTRAL"     "Square"     "PINK"        "Purple"
## PD   "PINK"       "Knotted"     "Durry"      "Rectangle"   "DoubleBack"
## L-5  "Durry"      "Rectangle"   "DoubleBack" "Purple"      "NEUTRAL"
## M-2  "HandTufted" "BLUSHPINK"   "Gray"       "NEUTRAL"     "Jacquared"
## RC   "Jacquared"  "Knotted"     "DoubleBack" "Rectangle"   "TAN"
## P-4  "Knotted"    "PINK"        "DoubleBack" "Purple"      "NEUTRAL"
## T-4  "NAVY"       "Round"       "Other"      "DoubleWowen" "Handloom"
## PC   "Purple"     "Rectangle"   "Knotted"    "DoubleBack"  "Square"
## A-11 "Purple"     "DoubleBack"  "Rectangle"  "PINK"        "NEUTRAL"
## CC   "Navy"       "NAVY"        "Durry"      "Square"      "PINK"
##      [,41]         [,42]         [,43]         [,44]         [,45]
## H-2  "Other"       "Handloom"    "BLUE"        "Navy"        "NAVY"
## P-5  "Navy"        "NAVY"        "Other"       "TAN"         "Round"
## M-1  "Rectangle"   "TAN"         "Square"      "DoubleWowen" "Knotted"
## A-9  "TAN"         "BLUE"        "Navy"        "NAVY"        "Round"
## C-2  "Square"      "Durry"       "PINK"        "Knotted"     "Rectangle"
## JL   "TAN"         "DoubleWowen" "Gray"        "Square"      "Navy"
## N-1  "PINK"        "Square"      "Knotted"     "Purple"      "DoubleBack"
## T-5  "BLUSHPINK"   "Jacquared"   "Gray"        "Durry"       "PINK"
## C-1  "Durry"       "Handloom"    "Round"       "DoubleWowen" "Other"
## T-2  "DoubleWowen" "Durry"       "Other"       "Handloom"    "Round"
## I-2  "Knotted"     "DoubleBack"  "Rectangle"   "TAN"         "BLUE"
## PD   "Purple"      "NEUTRAL"     "DoubleWowen" "BLUE"        "Other"
## L-5  "DoubleWowen" "Other"       "HandTufted"  "Round"       "Navy"
## M-2  "PINK"        "Square"      "Rectangle"   "Purple"      "Knotted"
## RC   "DoubleWowen" "BLUE"        "Navy"        "NAVY"        "Square"
## P-4  "DoubleWowen" "HandTufted"  "Round"       "Navy"        "NAVY"
## T-4  "Durry"       "HandTufted"  "BLUSHPINK"   "Jacquared"   "Gray"
## PC   "TAN"         "BLUE"        "Navy"        "NAVY"        "Durry"
```

```
## A-11 "Other"       "Round"       "BLUE"        "Navy"        "NAVY"
## CC   "Round"       "Handloom"    "Other"       "DoubleWowen" "HandTufted"
##      [,46]        [,47]        [,48]         [,49]         [,50]
## H-2  "HandTufted" "Round"      "DoubleWowen" "Jacquared"   "TAN"
## P-5  "Handloom"   "BLUE"       "Jacquared"   "BLUSHPINK"   "Knotted"
## M-1  "Gray"       "Purple"     "BLUE"        "DoubleBack"  "Durry"
## A-9  "Handloom"   "Other"      "DoubleWowen" "Durry"       "HandTufted"
## C-2  "DoubleBack" "Purple"     "NEUTRAL"     "DoubleWowen" "BLUE"
## JL   "NAVY"       "HandTufted" "BLUE"        "Handloom"    "Knotted"
## N-1  "TAN"        "BLUE"       "Navy"        "NAVY"        "Other"
## T-5  "Square"     "Knotted"    "Rectangle"   "DoubleBack"  "NEUTRAL"
## C-1  "HandTufted" "NEUTRAL"    "Gray"        "Square"      "BLUSHPINK"
## T-2  "HandTufted" "Square"     "Knotted"     "Gray"        "Purple"
## I-2  "Navy"       "NAVY"       "Durry"       "DoubleWowen" "Other"
## PD   "Round"      "Navy"       "NAVY"        "Handloom"    "HandTufted"
## L-5  "NAVY"       "Handloom"   "BLUE"        "BLUSHPINK"   "Gray"
## M-2  "DoubleBack" "TAN"        "BLUE"        "Navy"        "NAVY"
## RC   "Handloom"   "Other"      "HandTufted"  "Round"       "Durry"
## P-4  "Handloom"   "BLUE"       "Other"       "Jacquared"   "TAN"
## T-4  "PINK"       "NEUTRAL"    "Square"      "Rectangle"   "Knotted"
## PC   "Round"      "Handloom"   "Other"       "DoubleWowen" "NEUTRAL"
## A-11 "HandTufted" "Jacquared"  "DoubleWowen" "Durry"       "TAN"
## CC   "NEUTRAL"    "BLUSHPINK"  "Jacquared"   "Gray"        "Rectangle"
##      [,51]        [,52]         [,53]         [,54]        [,55]
## H-2  "Rectangle"  "Durry"       "Gray"        "BLUSHPINK"  "Square"
## P-5  "Durry"      "DoubleBack"  "PINK"        "Purple"     "Rectangle"
## M-1  "HandTufted" "Navy"        "NAVY"        "Round"      "Other"
## A-9  "Jacquared"  "BLUSHPINK"   "NEUTRAL"     "Gray"       "PINK"
## C-2  "Other"      "HandTufted"  "Handloom"    "Navy"       "NAVY"
## JL   "Other"      "Round"       "Durry"       "BLUSHPINK"  "Jacquared"
## N-1  "Round"      "Handloom"    "Durry"       "HandTufted" "DoubleWowen"
## T-5  "Purple"     "DoubleWowen" "TAN"         "BLUE"       "Other"
## C-1  "PINK"       "Jacquared"   "Purple"      "DoubleBack" "Rectangle"
## T-2  "Jacquared"  "PINK"        "BLUSHPINK"   "NEUTRAL"    "DoubleBack"
## I-2  "Round"      "HandTufted"  "Handloom"    "Gray"       "Jacquared"
## PD   "Gray"       "Jacquared"   "BLUSHPINK"   "TAN"        "Square"
## L-5  "Jacquared"  "TAN"         "Square"      "PINK"       "Knotted"
## M-2  "Handloom"   "Other"       "DoubleWowen" "Round"      "Durry"
## RC   "Gray"       "NEUTRAL"     "BLUSHPINK"   "PINK"       "Purple"
## P-4  "BLUSHPINK"  "Gray"        "Durry"       "Square"     "Rectangle"
## T-4  "Purple"     "DoubleBack"  "TAN"         "BLUE"       "Navy"
## PC   "BLUSHPINK"  "PINK"        "HandTufted"  "Gray"       "Jacquared"
## A-11 "Handloom"   "Gray"        "BLUSHPINK"   "Knotted"    "Square"
## CC   "Purple"     "Knotted"     "DoubleBack"  "TAN"        "BLUE"
##      [,56]        [,57]         [,58]        [,59]         [,60]
## H-2  "PINK"       "Knotted"     "NEUTRAL"    "DoubleBack"  "Purple"
## P-5  "NEUTRAL"    "DoubleWowen" "Square"     "Gray"        "HandTufted"
## M-1  "PINK"       "BLUSHPINK"   "Handloom"   "Jacquared"   "NEUTRAL"
## A-9  "Square"     "Rectangle"   "Purple"     "Knotted"     "DoubleBack"
## C-2  "Round"      "Gray"        "TAN"        "Jacquared"   "BLUSHPINK"
```

```
## JL   "Purple"     "PINK"        "DoubleBack" "NEUTRAL"     "Rectangle"
## N-1  "Jacquared"  "Rectangle"   "NEUTRAL"    "BLUSHPINK"   "Gray"
## T-5  "Round"      "Handloom"    "Navy"       "NAVY"        "HandTufted"
## C-1  "Knotted"    "TAN"         "BLUE"       "Navy"        "NAVY"
## T-2  "Rectangle"  "TAN"         "BLUE"       "Navy"        "NAVY"
## I-2  "BLUSHPINK"  "NEUTRAL"     "Square"     "PINK"        "Purple"
## PD   "PINK"       "Knotted"     "Durry"      "Rectangle"   "DoubleBack"
## L-5  "Durry"      "Rectangle"   "DoubleBack" "Purple"      "NEUTRAL"
## M-2  "HandTufted" "BLUSHPINK"   "Gray"       "NEUTRAL"     "Jacquared"
## RC   "Jacquared"  "Knotted"     "DoubleBack" "Rectangle"   "TAN"
## P-4  "Knotted"    "PINK"        "DoubleBack" "Purple"      "NEUTRAL"
## T-4  "NAVY"       "Round"       "Other"      "DoubleWowen" "Handloom"
## PC   "Purple"     "Rectangle"   "Knotted"    "DoubleBack"  "Square"
## A-11 "Purple"     "DoubleBack"  "Rectangle"  "PINK"        "NEUTRAL"
## CC   "Navy"       "NAVY"        "Durry"      "Square"      "PINK"
##      [,61]         [,62]         [,63]         [,64]         [,65]
## H-2  "Other"       "Handloom"    "BLUE"        "Navy"        "NAVY"
## P-5  "Navy"        "NAVY"        "Other"       "TAN"         "Round"
## M-1  "Rectangle"   "TAN"         "Square"      "DoubleWowen" "Knotted"
## A-9  "TAN"         "BLUE"        "Navy"        "NAVY"        "Round"
## C-2  "Square"      "Durry"       "PINK"        "Knotted"     "Rectangle"
## JL   "TAN"         "DoubleWowen" "Gray"        "Square"      "Navy"
## N-1  "PINK"        "Square"      "Knotted"     "Purple"      "DoubleBack"
## T-5  "BLUSHPINK"   "Jacquared"   "Gray"        "Durry"       "PINK"
## C-1  "Durry"       "Handloom"    "Round"       "DoubleWowen" "Other"
## T-2  "DoubleWowen" "Durry"       "Other"       "Handloom"    "Round"
## I-2  "Knotted"     "DoubleBack"  "Rectangle"   "TAN"         "BLUE"
## PD   "Purple"      "NEUTRAL"     "DoubleWowen" "BLUE"        "Other"
## L-5  "DoubleWowen" "Other"       "HandTufted"  "Round"       "Navy"
## M-2  "PINK"        "Square"      "Rectangle"   "Purple"      "Knotted"
## RC   "DoubleWowen" "BLUE"        "Navy"        "NAVY"        "Square"
## P-4  "DoubleWowen" "HandTufted"  "Round"       "Navy"        "NAVY"
## T-4  "Durry"       "HandTufted"  "BLUSHPINK"   "Jacquared"   "Gray"
## PC   "TAN"         "BLUE"        "Navy"        "NAVY"        "Durry"
## A-11 "Other"       "Round"       "BLUE"        "Navy"        "NAVY"
## CC   "Round"       "Handloom"    "Other"       "DoubleWowen" "HandTufted"
##      [,66]        [,67]        [,68]         [,69]         [,70]
## H-2  "HandTufted" "Round"      "DoubleWowen" "Jacquared"   "TAN"
## P-5  "Handloom"   "BLUE"       "Jacquared"   "BLUSHPINK"   "Knotted"
## M-1  "Gray"       "Purple"     "BLUE"        "DoubleBack"  "Durry"
## A-9  "Handloom"   "Other"      "DoubleWowen" "Durry"       "HandTufted"
## C-2  "DoubleBack" "Purple"     "NEUTRAL"     "DoubleWowen" "BLUE"
## JL   "NAVY"       "HandTufted" "BLUE"        "Handloom"    "Knotted"
## N-1  "TAN"        "BLUE"       "Navy"        "NAVY"        "Other"
## T-5  "Square"     "Knotted"    "Rectangle"   "DoubleBack"  "NEUTRAL"
## C-1  "HandTufted" "NEUTRAL"    "Gray"        "Square"      "BLUSHPINK"
## T-2  "HandTufted" "Square"     "Knotted"     "Gray"        "Purple"
## I-2  "Navy"       "NAVY"       "Durry"       "DoubleWowen" "Other"
## PD   "Round"      "Navy"       "NAVY"        "Handloom"    "HandTufted"
## L-5  "NAVY"       "Handloom"   "BLUE"        "BLUSHPINK"   "Gray"
```

```
## M-2   "DoubleBack" "TAN"        "BLUE"        "Navy"        "NAVY"
## RC    "Handloom"   "Other"      "HandTufted"  "Round"       "Durry"
## P-4   "Handloom"   "BLUE"       "Other"       "Jacquared"   "TAN"
## T-4   "PINK"       "NEUTRAL"    "Square"      "Rectangle"   "Knotted"
## PC    "Round"      "Handloom"   "Other"       "DoubleWowen" "NEUTRAL"
## A-11  "HandTufted" "Jacquared"  "DoubleWowen" "Durry"       "TAN"
## CC    "NEUTRAL"    "BLUSHPINK"  "Jacquared"   "Gray"        "Rectangle"
##       [,71]        [,72]         [,73]         [,74]        [,75]
## H-2   "Rectangle"  "Durry"       "Gray"        "BLUSHPINK"  "Square"
## P-5   "Durry"      "DoubleBack"  "PINK"        "Purple"     "Rectangle"
## M-1   "HandTufted" "Navy"        "NAVY"        "Round"      "Other"
## A-9   "Jacquared"  "BLUSHPINK"   "NEUTRAL"     "Gray"       "PINK"
## C-2   "Other"      "HandTufted"  "Handloom"    "Navy"       "NAVY"
## JL    "Other"      "Round"       "Durry"       "BLUSHPINK"  "Jacquared"
## N-1   "Round"      "Handloom"    "Durry"       "HandTufted" "DoubleWowen"
## T-5   "Purple"     "DoubleWowen" "TAN"         "BLUE"       "Other"
## C-1   "PINK"       "Jacquared"   "Purple"      "DoubleBack" "Rectangle"
## T-2   "Jacquared"  "PINK"        "BLUSHPINK"   "NEUTRAL"    "DoubleBack"
## I-2   "Round"      "HandTufted"  "Handloom"    "Gray"       "Jacquared"
## PD    "Gray"       "Jacquared"   "BLUSHPINK"   "TAN"        "Square"
## L-5   "Jacquared"  "TAN"         "Square"      "PINK"       "Knotted"
## M-2   "Handloom"   "Other"       "DoubleWowen" "Round"      "Durry"
## RC    "Gray"       "NEUTRAL"     "BLUSHPINK"   "PINK"       "Purple"
## P-4   "BLUSHPINK"  "Gray"        "Durry"       "Square"     "Rectangle"
## T-4   "Purple"     "DoubleBack"  "TAN"         "BLUE"       "Navy"
## PC    "BLUSHPINK"  "PINK"        "HandTufted"  "Gray"       "Jacquared"
## A-11  "Handloom"   "Gray"        "BLUSHPINK"   "Knotted"    "Square"
## CC    "Purple"     "Knotted"     "DoubleBack"  "TAN"        "BLUE"
##       [,76]        [,77]         [,78]        [,79]         [,80]
## H-2   "PINK"       "Knotted"     "NEUTRAL"    "DoubleBack"  "Purple"
## P-5   "NEUTRAL"    "DoubleWowen" "Square"     "Gray"        "HandTufted"
## M-1   "PINK"       "BLUSHPINK"   "Handloom"   "Jacquared"   "NEUTRAL"
## A-9   "Square"     "Rectangle"   "Purple"     "Knotted"     "DoubleBack"
## C-2   "Round"      "Gray"        "TAN"        "Jacquared"   "BLUSHPINK"
## JL    "Purple"     "PINK"        "DoubleBack" "NEUTRAL"     "Rectangle"
## N-1   "Jacquared"  "Rectangle"   "NEUTRAL"    "BLUSHPINK"   "Gray"
## T-5   "Round"      "Handloom"    "Navy"       "NAVY"        "HandTufted"
## C-1   "Knotted"    "TAN"         "BLUE"       "Navy"        "NAVY"
## T-2   "Rectangle"  "TAN"         "BLUE"       "Navy"        "NAVY"
## I-2   "BLUSHPINK"  "NEUTRAL"     "Square"     "PINK"        "Purple"
## PD    "PINK"       "Knotted"     "Durry"      "Rectangle"   "DoubleBack"
## L-5   "Durry"      "Rectangle"   "DoubleBack" "Purple"      "NEUTRAL"
## M-2   "HandTufted" "BLUSHPINK"   "Gray"       "NEUTRAL"     "Jacquared"
## RC    "Jacquared"  "Knotted"     "DoubleBack" "Rectangle"   "TAN"
## P-4   "Knotted"    "PINK"        "DoubleBack" "Purple"      "NEUTRAL"
## T-4   "NAVY"       "Round"       "Other"      "DoubleWowen" "Handloom"
## PC    "Purple"     "Rectangle"   "Knotted"    "DoubleBack"  "Square"
## A-11  "Purple"     "DoubleBack"  "Rectangle"  "PINK"        "NEUTRAL"
## CC    "Navy"       "NAVY"        "Durry"      "Square"      "PINK"
##       [,81]         [,82]         [,83]         [,84]         [,85]
```

```
## H-2  "Other"       "Handloom"    "BLUE"        "Navy"        "NAVY"
## P-5  "Navy"        "NAVY"        "Other"       "TAN"         "Round"
## M-1  "Rectangle"   "TAN"         "Square"      "DoubleWowen" "Knotted"
## A-9  "TAN"         "BLUE"        "Navy"        "NAVY"        "Round"
## C-2  "Square"      "Durry"       "PINK"        "Knotted"     "Rectangle"
## JL   "TAN"         "DoubleWowen" "Gray"        "Square"      "Navy"
## N-1  "PINK"        "Square"      "Knotted"     "Purple"      "DoubleBack"
## T-5  "BLUSHPINK"   "Jacquared"   "Gray"        "Durry"       "PINK"
## C-1  "Durry"       "Handloom"    "Round"       "DoubleWowen" "Other"
## T-2  "DoubleWowen" "Durry"       "Other"       "Handloom"    "Round"
## I-2  "Knotted"     "DoubleBack"  "Rectangle"   "TAN"         "BLUE"
## PD   "Purple"      "NEUTRAL"     "DoubleWowen" "BLUE"        "Other"
## L-5  "DoubleWowen" "Other"       "HandTufted"  "Round"       "Navy"
## M-2  "PINK"        "Square"      "Rectangle"   "Purple"      "Knotted"
## RC   "DoubleWowen" "BLUE"        "Navy"        "NAVY"        "Square"
## P-4  "DoubleWowen" "HandTufted"  "Round"       "Navy"        "NAVY"
## T-4  "Durry"       "HandTufted"  "BLUSHPINK"   "Jacquared"   "Gray"
## PC   "TAN"         "BLUE"        "Navy"        "NAVY"        "Durry"
## A-11 "Other"       "Round"       "BLUE"        "Navy"        "NAVY"
## CC   "Round"       "Handloom"    "Other"       "DoubleWowen" "HandTufted"
##      [,86]        [,87]        [,88]         [,89]         [,90]
## H-2  "HandTufted" "Round"      "DoubleWowen" "Jacquared"   "TAN"
## P-5  "Handloom"   "BLUE"       "Jacquared"   "BLUSHPINK"   "Knotted"
## M-1  "Gray"       "Purple"     "BLUE"        "DoubleBack"  "Durry"
## A-9  "Handloom"   "Other"      "DoubleWowen" "Durry"       "HandTufted"
## C-2  "DoubleBack" "Purple"     "NEUTRAL"     "DoubleWowen" "BLUE"
## JL   "NAVY"       "HandTufted" "BLUE"        "Handloom"    "Knotted"
## N-1  "TAN"        "BLUE"       "Navy"        "NAVY"        "Other"
## T-5  "Square"     "Knotted"    "Rectangle"   "DoubleBack"  "NEUTRAL"
## C-1  "HandTufted" "NEUTRAL"    "Gray"        "Square"      "BLUSHPINK"
## T-2  "HandTufted" "Square"     "Knotted"     "Gray"        "Purple"
## I-2  "Navy"       "NAVY"       "Durry"       "DoubleWowen" "Other"
## PD   "Round"      "Navy"       "NAVY"        "Handloom"    "HandTufted"
## L-5  "NAVY"       "Handloom"   "BLUE"        "BLUSHPINK"   "Gray"
## M-2  "DoubleBack" "TAN"        "BLUE"        "Navy"        "NAVY"
## RC   "Handloom"   "Other"      "HandTufted"  "Round"       "Durry"
## P-4  "Handloom"   "BLUE"       "Other"       "Jacquared"   "TAN"
## T-4  "PINK"       "NEUTRAL"    "Square"      "Rectangle"   "Knotted"
## PC   "Round"      "Handloom"   "Other"       "DoubleWowen" "NEUTRAL"
## A-11 "HandTufted" "Jacquared"  "DoubleWowen" "Durry"       "TAN"
## CC   "NEUTRAL"    "BLUSHPINK"  "Jacquared"   "Gray"        "Rectangle"
##      [,91]        [,92]         [,93]         [,94]        [,95]
## H-2  "Rectangle"  "Durry"       "Gray"        "BLUSHPINK"  "Square"
## P-5  "Durry"      "DoubleBack"  "PINK"        "Purple"     "Rectangle"
## M-1  "HandTufted" "Navy"        "NAVY"        "Round"      "Other"
## A-9  "Jacquared"  "BLUSHPINK"   "NEUTRAL"     "Gray"       "PINK"
## C-2  "Other"      "HandTufted"  "Handloom"    "Navy"       "NAVY"
## JL   "Other"      "Round"       "Durry"       "BLUSHPINK"  "Jacquared"
## N-1  "Round"      "Handloom"    "Durry"       "HandTufted" "DoubleWowen"
## T-5  "Purple"     "DoubleWowen" "TAN"         "BLUE"       "Other"
```

```
## C-1  "PINK"        "Jacquared"   "Purple"      "DoubleBack" "Rectangle"
## T-2  "Jacquared"   "PINK"        "BLUSHPINK"   "NEUTRAL"    "DoubleBack"
## I-2  "Round"       "HandTufted"  "Handloom"    "Gray"       "Jacquared"
## PD   "Gray"        "Jacquared"   "BLUSHPINK"   "TAN"        "Square"
## L-5  "Jacquared"   "TAN"         "Square"      "PINK"       "Knotted"
## M-2  "Handloom"    "Other"       "DoubleWowen" "Round"      "Durry"
## RC   "Gray"        "NEUTRAL"     "BLUSHPINK"   "PINK"       "Purple"
## P-4  "BLUSHPINK"   "Gray"        "Durry"       "Square"     "Rectangle"
## T-4  "Purple"      "DoubleBack"  "TAN"         "BLUE"       "Navy"
## PC   "BLUSHPINK"   "PINK"        "HandTufted"  "Gray"       "Jacquared"
## A-11 "Handloom"    "Gray"        "BLUSHPINK"   "Knotted"    "Square"
## CC   "Purple"      "Knotted"     "DoubleBack"  "TAN"        "BLUE"
##      [,96]         [,97]         [,98]        [,99]         [,100]
## H-2  "PINK"        "Knotted"     "NEUTRAL"    "DoubleBack"  "Purple"
## P-5  "NEUTRAL"     "DoubleWowen" "Square"     "Gray"        "HandTufted"
## M-1  "PINK"        "BLUSHPINK"   "Handloom"   "Jacquared"   "NEUTRAL"
## A-9  "Square"      "Rectangle"   "Purple"     "Knotted"     "DoubleBack"
## C-2  "Round"       "Gray"        "TAN"        "Jacquared"   "BLUSHPINK"
## JL   "Purple"      "PINK"        "DoubleBack" "NEUTRAL"     "Rectangle"
## N-1  "Jacquared"   "Rectangle"   "NEUTRAL"    "BLUSHPINK"   "Gray"
## T-5  "Round"       "Handloom"    "Navy"       "NAVY"        "HandTufted"
## C-1  "Knotted"     "TAN"         "BLUE"       "Navy"        "NAVY"
## T-2  "Rectangle"   "TAN"         "BLUE"       "Navy"        "NAVY"
## I-2  "BLUSHPINK"   "NEUTRAL"     "Square"     "PINK"        "Purple"
## PD   "PINK"        "Knotted"     "Durry"      "Rectangle"   "DoubleBack"
## L-5  "Durry"       "Rectangle"   "DoubleBack" "Purple"      "NEUTRAL"
## M-2  "HandTufted"  "BLUSHPINK"   "Gray"       "NEUTRAL"     "Jacquared"
## RC   "Jacquared"   "Knotted"     "DoubleBack" "Rectangle"   "TAN"
## P-4  "Knotted"     "PINK"        "DoubleBack" "Purple"      "NEUTRAL"
## T-4  "NAVY"        "Round"       "Other"      "DoubleWowen" "Handloom"
## PC   "Purple"      "Rectangle"   "Knotted"    "DoubleBack"  "Square"
## A-11 "Purple"      "DoubleBack"  "Rectangle"  "PINK"        "NEUTRAL"
## CC   "Navy"        "NAVY"        "Durry"      "Square"      "PINK"
```

8. What will be your final recommendation to Champo Carpets?

Answer:

- Some recommendations to Champo carpets would be:
    [1] Send out more samples and focus on **A-9 and C-1 Customer code** as they are more likely to place order amongst all other customers.
    [2] Send out more samples to customers from **USA and UK.**
    [3] Producing and sending out samples of **'Hand Tufted'** and **'Durry'** as these are the highly in demand and have more chances of conversion.
    [4] Sending out samples with Quality **'TUFTED 60C+VISC 2/16 5PLY'** and Design Name **'ZILLAH'** can be highly effective as it has the highest number of orders.
    [5] Manufacturing more products in the color **'100 Multi'.**
    [6] Sending out and manufacturing carpets in **'Rectangular'** shape.