K-means Clustering

```
library(class)
library(MASS)
library(dplyr)

library(readxl)
```

Loading Data & Data Pre-processing:

```
prospects <- read_xls("prospect.xls")

prospects <- prospects %>%
  select(-ID,-LOC)

str(prospects)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    4701 obs. of  7 variables:
##  $ AGE      : num  37 46 45 38 34 69 46 28 37 46 ...
##  $ INCOME   : num  57 71 65 50 44 60 42 63 59 57 ...
##  $ SEX      : chr  "F" "M" "M" "F" ...
##  $ MARRIED  : num  0 1 1 0 0 0 1 0 1 1 ...
##  $ OWNHOME  : num  0 0 1 0 0 0 0 1 0 1 ...
##  $ CLIMATE  : chr  "20" "20" "20" "10" ...
##  $ FICO>=700: num  0 0 1 0 0 0 1 1 1 1 ...
```

```
#Converting into factor type
prospects$SEX <- as.factor(prospects$SEX)
prospects$CLIMATE <- as.factor(prospects$CLIMATE)

summary(prospects)
```

```
##       AGE            INCOME         SEX          MARRIED
##  Min.   :18.00   Min.   : 15.00   F   :2161   Min.   :0.0000
##  1st Qu.:38.00   1st Qu.: 35.00   M   :2434   1st Qu.:0.0000
##  Median :44.00   Median : 50.00   NA's: 106   Median :1.0000
##  Mean   :44.23   Mean   : 47.69               Mean   :0.5785
##  3rd Qu.:50.00   3rd Qu.: 61.00               3rd Qu.:1.0000
##  Max.   :75.00   Max.   :116.00               Max.   :1.0000
##  NA's   :106     NA's   :106                  NA's   :106
##     OWNHOME       CLIMATE     FICO>=700
##  Min.   :0.0000   10: 871   Min.   :0.0000
##  1st Qu.:0.0000   20:2932   1st Qu.:0.0000
##  Median :0.0000   30: 898   Median :0.0000
##  Mean   :0.3277             Mean   :0.4135
##  3rd Qu.:1.0000             3rd Qu.:1.0000
##  Max.   :1.0000             Max.   :1.0000
##  NA's   :106                NA's   :106
```

Creating Dummy variables for Factor variables

```r
# Creating dummy variables for SEX and CLIMATE

library(caret)
```

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

```
## Loading required package: ggplot2
```

```r
var_onehot <- c('SEX', "CLIMATE")

# One Hot Encoding

dummys <- dummyVars(" ~ .", data = prospects[,var_onehot])

dummy_cats <- data.frame(predict(dummys, newdata = prospects[,var_onehot]))

prospects <- prospects %>%
  select(-SEX, -CLIMATE)

  prospects <- cbind(prospects, dummy_cats)

  summary(prospects)
```

```
##       AGE             INCOME          MARRIED          OWNHOME
##  Min.   :18.00   Min.   : 15.00   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:38.00   1st Qu.: 35.00   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :44.00   Median : 50.00   Median :1.0000   Median :0.0000
##  Mean   :44.23   Mean   : 47.69   Mean   :0.5785   Mean   :0.3277
##  3rd Qu.:50.00   3rd Qu.: 61.00   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :75.00   Max.   :116.00   Max.   :1.0000   Max.   :1.0000
##  NA's   :106     NA's   :106      NA's   :106      NA's   :106
##    FICO>=700          SEX.F            SEX.M          CLIMATE.10
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :1.0000   Median :0.0000
##  Mean   :0.4135   Mean   :0.4703   Mean   :0.5297   Mean   :0.1853
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  NA's   :106      NA's   :106      NA's   :106
##    CLIMATE.20       CLIMATE.30
##  Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.000
##  Median :1.0000   Median :0.000
##  Mean   :0.6237   Mean   :0.191
##  3rd Qu.:1.0000   3rd Qu.:0.000
##  Max.   :1.0000   Max.   :1.000
##
```

Missing values imputation using KNN:

```
pp <- prospects

#imputing missing values using KNN

library(DMwR)

## Loading required package: grid

## Registered S3 method overwritten by 'xts':
##   method     from
##   as.zoo.xts zoo

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

knn_prospects <- knnImputation(pp, k=10, scale = T, meth = "weighAvg", distData = NULL)
anyNA(knn_prospects)

## [1] FALSE
```

Scaling or Normalizing Data

```
prospects <- knn_prospects

# Scaling data

options(digits=2)
normalize <- function(x) {
return ((x - min(x)) / (max(x) - min(x)))
}

scaled_prospects <- as.data.frame(lapply(prospects[1:10], normalize))

summary(scaled_prospects)

##       AGE           INCOME        MARRIED        OWNHOME
##  Min.   :0.00   Min.   :0.00   Min.   :0.00   Min.   :0.00
##  1st Qu.:0.35   1st Qu.:0.20   1st Qu.:0.00   1st Qu.:0.00
##  Median :0.46   Median :0.35   Median :1.00   Median :0.00
##  Mean   :0.46   Mean   :0.32   Mean   :0.58   Mean   :0.33
##  3rd Qu.:0.56   3rd Qu.:0.45   3rd Qu.:1.00   3rd Qu.:1.00
##  Max.   :1.00   Max.   :1.00   Max.   :1.00   Max.   :1.00
##    FICO..700        SEX.F          SEX.M       CLIMATE.10
##  Min.   :0.00   Min.   :0.00   Min.   :0.00   Min.   :0.00
##  1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00
##  Median :0.00   Median :0.00   Median :1.00   Median :0.00
##  Mean   :0.42   Mean   :0.47   Mean   :0.53   Mean   :0.19
##  3rd Qu.:1.00   3rd Qu.:1.00   3rd Qu.:1.00   3rd Qu.:0.00
##  Max.   :1.00   Max.   :1.00   Max.   :1.00   Max.   :1.00
##    CLIMATE.20     CLIMATE.30
##  Min.   :0.00   Min.   :0.00
```

```
##  1st Qu.:0.00   1st Qu.:0.00
##  Median :1.00   Median :0.00
##  Mean   :0.62   Mean   :0.19
##  3rd Qu.:1.00   3rd Qu.:0.00
##  Max.   :1.00   Max.   :1.00
```

Part a.)

```
set.seed(7)
kmeans4 <- kmeans(scaled_prospects, centers=4, nstart =100)
kmeans4
```

```
## K-means clustering with 4 clusters of sizes 898, 1361, 871, 1571
##
## Cluster means:
##    AGE INCOME MARRIED OWNHOME FICO..700 SEX.F SEX.M CLIMATE.10 CLIMATE.20
## 1 0.46   0.30    0.58    0.23      0.42 0.485  0.52          0          0
## 2 0.47   0.30    0.60    0.34      0.39 1.000  0.00          0          1
## 3 0.46   0.30    0.57    0.31      0.38 0.443  0.56          1          0
## 4 0.46   0.37    0.57    0.39      0.46 0.018  0.98          0          1
##   CLIMATE.30
## 1          1
## 2          0
## 3          0
## 4          0
##
## Clustering vector:
##    [1] 2 4 4 3 4 1 2 2 4 4 4 2 1 1 4 3 3 3 4 4 2 4 2 4 3 4 1 4 4 4 2 4 4 1
##   [35] 1 2 4 3 1 2 4 2 1 4 4 4 2 3 3 3 1 4 4 4 1 2 1 2 4 2 1 2 3 4 1 4 2 3
##   [69] 3 3 3 4 2 2 2 4 3 4 2 1 2 2 3 3 2 2 2 2 3 1 2 3 1 1 1 2 4 4 3 4 2 4
##  [103] 2 4 3 4 1 4 2 4 4 3 3 4 3 3 4 3 3 4 3 3 4 1 4 4 4 1 3 4 4 4 2 4 2 3
##  [137] 3 2 2 2 2 1 4 1 4 4 4 2 4 3 2 4 4 3 4 2 1 4 1 2 4 1 4 2 1 3 2 3 2 1
##  [171] 3 4 4 3 2 2 3 4 2 4 1 3 4 4 1 1 2 2 3 3 4 3 4 3 2 2 2 3 1 4 2 4 1 3
##  [205] 4 3 2 4 2 1 3 2 2 2 4 2 3 3 4 3 2 4 3 1 4 3 1 3 2 1 4 2 3 2 4 4 3 3
##  [239] 3 2 2 4 4 4 4 4 4 4 4 2 4 1 2 4 2 3 3 2 3 2 4 4 1 4 4 1 2 3 2 4 4 2
##  [273] 2 1 3 4 4 3 4 4 2 4 1 3 2 1 4 2 3 4 4 1 4 2 4 2 2 3 1 2 1 1 4 2 3 2
##  [307] 4 4 1 1 2 2 3 1 2 4 1 3 1 2 2 2 1 4 4 2 4 3 4 3 4 4 1 4 3 4 1 2 4 2
##  [341] 3 4 4 4 2 2 3 1 2 4 2 2 4 4 3 1 4 2 4 4 2 4 3 4 4 2 1 1 3 4 1 2 4 3
##  [375] 3 3 2 2 4 4 4 4 4 3 2 2 4 3 2 3 3 4 3 2 3 4 2 2 2 2 4 1 2 1 2 2 1 4
##  [409] 1 1 1 3 2 1 2 2 2 1 2 1 2 3 2 1 4 4 3 2 2 4 4 1 2 2 3 4 4 4 3 2 3 1
##  [443] 4 2 2 3 4 3 4 2 3 3 1 3 4 4 1 1 2 4 1 1 2 3 2 2 4 3 4 4 4 3 1 4 4 3
##  [477] 1 2 4 2 4 1 4 3 2 1 2 1 2 4 4 3 4 3 1 1 3 4 1 2 2 3 1 1 4 4 1 3 4 4
##  [511] 1 2 3 1 4 2 3 1 2 4 4 4 3 2 4 4 3 3 4 2 1 2 1 4 3 3 1 1 3 2 4 2 1 3
##  [545] 2 1 3 4 1 4 3 4 2 2 4 4 3 1 4 2 4 4 2 4 4 3 2 4 2 2 2 2 2 4 4 2 3 2
##  [579] 3 4 2 3 2 1 4 4 1 4 4 2 2 4 4 2 3 2 4 2 4 2 3 2 2 4 4 4 4 2 3 2 3 1
##  [613] 1 3 4 2 1 2 4 4 2 1 4 2 2 1 4 3 4 4 4 4 4 1 1 4 4 2 4 4 1 4 4 1 1 3
##  [647] 3 4 3 3 4 2 2 3 4 2 4 4 2 2 2 4 4 2 1 1 2 3 2 1 4 3 2 3 4 4 1 1 4 3
##  [681] 4 4 2 1 2 4 2 4 3 2 3 4 3 4 2 2 4 1 1 4 3 4 2 2 3 1 3 2 4 4 4 2 1 3
##  [715] 4 1 2 2 4 2 4 1 1 2 1 1 4 2 4 2 4 4 3 4 3 3 1 1 2 4 4 1 4 3 2 4 3 4
##  [749] 2 3 3 3 2 2 4 4 3 2 2 4 4 4 4 2 2 4 4 1 4 2 1 2 2 2 2 1 4 2 4 4 2 1
##  [783] 2 3 1 2 3 2 1 1 2 4 4 3 2 2 4 4 2 4 2 2 2 3 3 2 3 1 3 3 1 1 3 4 2 1
##  [817] 3 4 4 3 2 1 4 2 2 2 3 4 3 2 4 4 2 4 2 1 4 3 2 4 1 2 1 4 2 4 2 3 2 2
##  [851] 2 1 4 1 2 4 2 4 1 4 2 3 1 2 4 2 3 3 4 4 3 2 2 3 3 3 3 4 1 4 2 3 3 3
##  [885] 4 1 4 2 2 2 2 1 3 3 1 4 2 4 1 4 4 4 1 1 4 3 2 4 4 4 3 3 4 2 4 3 4 4
##  [919] 4 4 4 1 4 2 2 1 1 4 4 1 1 4 4 2 1 2 1 4 2 1 3 4 3 3 2 4 4 4 4 1 4 2
##  [953] 2 4 2 1 3 1 4 2 2 2 2 3 1 3 3 4 2 2 3 2 4 1 4 1 4 3 1 1 1 4 2 2 3 4
```

```
##  [987] 3 4 1 1 2 4 4 3 2 2 4 4 1 1 2 2 4 3 3 2 4 3 4 4 4 4 1 3 1 1 1 1 2 1
## [1021] 4 1 4 4 4 1 3 1 2 4 2 1 4 4 1 1 1 2 4 3 2 3 2 2 4 4 1 2 4 4 3 1 2 3
## [1055] 3 3 2 1 1 1 3 2 4 2 2 1 4 1 1 3 3 2 1 1 3 4 3 4 1 4 1 4 1 1 1 3 4 2
## [1089] 4 2 1 3 1 1 4 4 1 4 2 4 4 2 1 2 1 4 4 4 3 1 3 4 3 4 4 2 3 1 4 1 2 4
## [1123] 4 1 3 3 4 1 1 2 4 1 4 2 1 1 4 1 4 4 2 1 2 2 2 3 3 4 1 1 4 1 2 1 2 4
## [1157] 4 4 3 4 4 3 4 1 3 4 2 2 3 1 3 3 1 3 4 2 1 1 4 1 1 1 2 2 1 4 1 3 3 2
## [1191] 4 4 3 2 4 4 4 2 2 4 1 3 4 1 3 4 3 4 1 4 3 3 2 2 3 2 4 2 4 2 3 4 4 3
## [1225] 4 4 2 1 1 2 3 3 3 1 4 2 2 1 4 3 2 3 2 1 2 4 4 4 2 2 2 4 4 3 4 3 4 3
## [1259] 1 3 2 2 2 2 1 4 1 4 2 1 2 2 2 3 4 4 4 4 1 4 2 2 3 2 3 2 1 2 1 4 4 1
## [1293] 4 2 1 4 1 4 2 2 3 4 2 1 3 4 3 4 3 3 1 4 2 1 3 2 2 4 4 4 4 4 1 4 2 4
## [1327] 3 4 4 1 3 1 2 2 2 1 1 2 2 1 3 4 3 4 4 2 3 3 3 2 2 4 2 4 2 1 2 1 2 2
## [1361] 3 3 2 3 4 4 4 1 1 3 4 3 2 4 4 4 4 1 4 3 1 4 4 4 1 4 1 2 2 2 2 2 4 4
## [1395] 3 4 4 4 4 3 1 1 2 1 4 3 1 2 3 3 3 1 4 3 4 3 4 2 2 2 2 2 2 3 1 4 3 2
## [1429] 2 4 4 3 1 2 1 2 4 3 1 3 2 2 4 4 2 4 2 1 2 4 3 1 1 2 2 2 4 2 3 2 2 3
## [1463] 2 4 2 2 3 2 4 4 2 2 1 1 4 4 3 1 2 3 4 2 4 4 2 2 4 4 3 4 1 1 2 1 3 1
## [1497] 4 1 3 3 4 3 3 4 4 2 2 1 1 4 3 1 1 4 4 2 4 4 3 4 3 2 4 3 1 4 4 2 2 2
## [1531] 3 4 3 4 4 4 2 2 1 3 4 3 1 2 2 1 2 4 4 2 4 1 4 1 3 1 3 2 2 2 3 3 2 1
## [1565] 1 4 3 4 1 3 4 1 2 1 2 1 4 3 2 3 1 1 3 2 2 1 1 1 1 4 4 4 1 1 2 3 2 2
## [1599] 3 4 4 2 4 2 4 2 1 2 3 4 4 2 4 4 4 4 1 2 3 1 2 1 2 2 2 2 1 4 4 4 2 3
## [1633] 4 4 4 1 1 4 4 2 2 1 3 4 2 4 2 4 2 2 3 4 2 3 1 2 1 4 1 3 2 4 1 3 2 3
## [1667] 1 1 4 4 4 3 2 4 4 2 4 2 1 2 4 2 3 2 4 4 2 1 4 3 1 1 2 4 2 1 4 1 2 2
## [1701] 2 4 2 4 3 2 2 3 3 4 1 4 2 4 4 1 1 2 4 4 4 3 4 1 1 4 1 3 3 2 2 2 4 2
## [1735] 4 3 3 1 2 4 4 4 4 1 3 1 1 4 3 1 1 4 4 4 2 1 3 1 2 2 4 4 3 1 2 4 1 2
## [1769] 2 1 2 2 4 4 2 2 3 4 4 1 1 4 2 4 1 4 4 4 4 3 1 4 3 4 2 4 2 1 4 2 1 4
## [1803] 4 3 3 2 1 4 2 1 4 1 1 4 3 2 2 3 4 3 4 3 4 2 3 2 1 2 4 4 2 4 2 4 2 4
## [1837] 4 2 2 4 4 3 4 2 4 1 2 2 1 2 4 1 4 2 1 2 4 2 2 4 4 1 4 4 1 4 3 4 3 4
## [1871] 1 4 3 4 4 3 4 3 3 4 4 2 2 2 3 1 4 4 3 2 3 2 3 4 2 4 1 4 1 2 3 2 1 2
## [1905] 4 4 3 4 1 2 4 2 2 1 2 1 1 3 1 3 4 4 2 1 4 2 4 4 1 2 3 2 3 3 4 2 2 2
## [1939] 3 4 1 4 3 2 2 2 4 2 1 4 4 3 4 3 3 2 2 2 4 2 2 2 4 4 3 3 3 2 2 4 4 4
## [1973] 2 3 4 1 4 2 4 4 4 1 1 2 2 2 1 4 2 2 1 4 1 1 2 3 1 4 4 4 2 1 1 1 4 4
## [2007] 1 2 4 2 2 1 3 3 2 4 2 1 2 4 2 2 4 2 4 4 1 1 2 3 4 3 2 4 2 3 3 4 4 2
## [2041] 1 1 2 3 4 2 1 1 2 4 1 1 3 4 3 2 2 2 4 4 4 4 4 1 3 4 3 4 3 1 2 2 2 4
## [2075] 4 1 1 4 3 3 2 4 1 1 2 4 4 4 4 1 4 4 1 2 2 4 2 2 4 2 4 2 4 4 3 3 3 3
## [2109] 2 2 3 2 4 3 3 3 2 2 2 3 1 4 3 2 3 2 4 1 3 2 4 2 1 1 3 4 2 2 2 3 3 4
## [2143] 1 2 1 1 1 4 4 2 1 1 1 2 4 4 1 1 4 1 4 2 2 4 2 4 2 2 4 3 2 2 2 2 4 4
## [2177] 1 2 3 4 3 4 4 4 3 2 2 3 3 3 3 3 3 2 1 4 2 4 1 4 4 4 1 2 1 4 1 4 4 1
## [2211] 4 2 3 2 4 3 2 4 2 2 4 1 4 4 4 3 2 3 2 1 4 4 4 4 3 4 2 2 4 2 3 2 2 1
## [2245] 4 3 3 1 2 3 1 3 2 2 2 2 2 1 3 1 4 2 1 3 4 3 2 2 3 1 2 3 4 4 1 2 2 4
## [2279] 4 4 2 1 2 2 2 2 1 2 1 3 4 4 3 2 4 3 4 2 2 3 3 1 1 4 3 3 1 1 3 2 4 4
## [2313] 4 3 3 1 1 1 1 4 3 4 3 4 4 2 4 1 2 2 4 4 4 3 3 3 4 3 3 2 1 2 2 3 4 3
## [2347] 3 1 4 2 3 2 2 4 2 2 2 4 2 2 4 1 2 1 4 3 4 1 1 2 2 2 1 4 1 1 1 3 2 1
## [2381] 1 3 4 3 3 3 4 4 1 1 2 3 4 3 1 2 2 2 4 3 2 4 4 3 4 4 4 4 3 2 4 4 3 3
## [2415] 1 2 4 3 4 3 3 4 4 4 4 3 2 4 3 2 4 4 4 4 4 1 2 2 4 1 2 1 4 4 3 3 2 2
## [2449] 4 4 4 4 2 3 4 4 4 4 3 1 4 3 4 2 4 4 3 4 2 1 4 1 3 4 2 3 1 1 4 3 4 1
## [2483] 1 1 4 4 3 2 1 2 1 4 4 4 4 2 4 3 1 3 4 1 4 2 3 2 4 2 2 2 1 2 4 2 2 1
## [2517] 4 2 2 2 4 4 4 4 4 3 4 4 4 2 1 2 4 2 2 4 2 4 3 2 2 4 1 3 4 1 1 1 4 1
## [2551] 2 1 2 4 2 2 3 3 2 2 1 2 1 4 3 3 4 4 4 4 4 3 4 1 2 2 4 1 2 2 4 4 2 3
## [2585] 1 1 2 2 4 4 2 2 4 4 2 4 2 1 4 4 2 4 4 1 4 3 3 2 3 2 4 2 2 3 3 4 1 1
## [2619] 2 4 4 3 3 4 4 3 3 1 2 1 4 3 1 2 3 2 3 1 4 1 3 2 2 2 2 4 3 4 2 1 1 2
## [2653] 4 3 4 4 1 2 3 4 4 2 4 3 4 3 3 1 3 2 3 3 1 4 2 3 4 2 4 3 2 2 4 2 2 3
## [2687] 1 4 2 2 4 3 1 4 1 1 3 3 2 4 4 1 2 4 3 3 4 1 4 4 2 1 1 4 3 2 2 1 4 2
## [2721] 2 3 3 2 1 4 3 3 3 2 3 4 4 1 4 2 4 4 4 2 1 2 2 4 2 1 1 4 4 3 4 4 3 3
## [2755] 4 1 4 3 3 2 4 2 4 4 2 4 2 4 3 2 1 2 2 2 4 2 4 3 4 1 4 4 1 1 4 1 2 3
## [2789] 4 4 2 4 1 3 4 1 1 2 4 3 3 4 4 1 4 2 2 4 1 3 2 4 4 4 4 2 2 4 3 4 4 4
## [2823] 4 4 2 2 2 2 4 4 4 2 2 4 3 3 2 1 1 3 4 1 4 2 3 3 3 2 3 3 2 2 2 4 4 2
```

```
## [2857] 4 2 3 2 2 1 3 4 1 3 2 2 4 4 3 1 4 1 2 2 4 1 3 2 1 4 2 2 2 3 4 1 2 4
## [2891] 4 2 3 4 4 2 4 2 2 4 1 4 1 4 2 1 1 4 2 2 2 4 2 4 2 2 4 4 1 1 3 2 2 3
## [2925] 4 2 4 4 1 4 2 4 1 2 3 2 1 2 2 1 4 4 3 3 2 1 4 4 4 4 1 2 3 4 2 2 2 4
## [2959] 2 2 2 3 3 4 4 4 2 2 2 4 3 4 1 3 2 1 4 4 1 1 1 4 3 4 4 2 3 1 3 2 1 4
## [2993] 2 3 4 4 2 4 3 2 1 4 3 4 4 1 2 2 4 2 2 2 4 2 3 2 4 4 4 3 4 4 2 1 1 1
## [3027] 1 3 1 2 1 3 2 4 3 4 1 4 2 4 4 4 3 4 4 4 3 4 4 2 1 4 2 4 2 4 1 4 4 2
## [3061] 4 2 3 1 4 1 2 2 2 3 4 2 4 3 4 4 4 2 1 3 4 2 3 2 4 1 4 1 1 4 2 1 4 2
## [3095] 4 2 2 4 4 4 4 2 2 3 2 4 3 2 4 2 3 3 2 4 1 2 1 1 2 4 2 1 1 1 1 3 4 4
## [3129] 4 4 1 1 4 4 4 3 2 4 3 4 4 4 1 2 2 2 4 4 2 4 4 1 2 3 4 3 1 1 3 4 4 3
## [3163] 2 4 4 3 1 4 4 3 4 4 4 4 4 1 1 1 4 2 2 4 2 2 3 2 4 4 4 3 2 2 1 4 2 1 1
## [3197] 1 4 2 2 3 2 3 3 2 2 4 1 2 4 4 2 4 2 4 2 4 4 2 2 4 1 3 3 4 2 3 4 3 1
## [3231] 1 4 3 2 3 2 2 4 4 2 4 4 1 3 2 3 3 2 1 4 2 3 3 2 1 3 4 4 3 2 2 4 4 4
## [3265] 4 1 4 2 1 1 4 2 2 1 1 1 3 1 3 4 4 4 4 2 4 4 4 4 4 2 2 3 3 2 1 2 2 2
## [3299] 4 4 4 4 4 4 1 4 4 3 3 2 1 2 3 2 1 1 4 4 3 1 3 1 2 4 2 2 4 1 4 2 4 2 4
## [3333] 1 4 1 4 3 4 3 4 4 1 3 3 2 1 4 3 2 3 2 2 3 1 4 3 4 2 4 2 2 2 4 3 2 2
## [3367] 3 2 2 4 2 4 1 2 1 2 2 4 4 2 1 3 4 2 1 4 4 1 2 2 1 2 2 4 2 4 2 4 2 2
## [3401] 4 3 2 2 4 4 2 2 1 1 4 2 3 2 4 3 2 2 1 2 3 2 2 2 3 4 4 3 2 1 1 2 3 2
## [3435] 3 3 1 3 3 2 4 4 4 3 4 1 4 4 2 1 2 3 3 4 3 1 2 2 2 1 4 2 1 4 4 4 1 4
## [3469] 2 1 4 2 3 1 3 4 4 1 2 4 1 2 3 2 2 4 4 2 3 2 2 2 1 1 3 3 4 4 4 4 2 4
## [3503] 4 4 2 2 4 4 1 2 4 2 2 4 4 1 4 3 3 3 1 1 3 1 3 4 3 4 4 2 3 1 3 4 4 4
## [3537] 2 3 1 2 2 2 2 2 3 2 3 1 3 4 2 1 2 1 4 2 4 1 2 2 3 4 3 2 4 2 4 4 4 4
## [3571] 4 2 4 2 4 4 2 2 4 4 2 4 3 1 2 2 4 4 1 4 4 4 4 1 1 4 1 1 4 1 2 1 1 3
## [3605] 4 2 1 3 4 3 2 3 4 3 2 1 2 2 4 4 1 2 2 2 4 4 4 1 4 1 4 4 4 2 2 1 2 1
## [3639] 1 3 4 1 1 2 2 4 2 1 2 4 1 1 3 1 4 1 2 2 1 4 2 2 4 2 1 3 4 1 3 1 4 3
## [3673] 2 2 4 4 1 1 1 1 4 1 3 2 4 1 4 4 2 4 2 2 4 4 3 3 3 1 2 2 3 2 2 2 2 4
## [3707] 2 2 1 3 3 3 4 2 4 2 4 4 2 1 3 3 3 2 1 4 2 2 1 2 4 3 3 4 2 2 2 2 4 2
## [3741] 4 2 1 2 4 1 4 2 2 1 1 3 2 2 1 2 2 1 4 4 2 3 1 1 1 2 4 4 2 2 2 2 4 3
## [3775] 4 2 4 2 4 2 1 2 4 1 3 1 4 2 4 2 2 1 4 2 2 2 3 4 3 1 3 3 2 4 1 3 2 2
## [3809] 2 3 3 3 4 4 1 2 4 4 3 1 4 4 2 4 4 2 4 2 4 1 1 2 1 3 4 2 1 2 4 2 3 4
## [3843] 2 2 4 4 2 1 1 4 4 4 1 4 2 2 3 4 4 3 3 1 4 4 3 4 2 4 2 1 1 2 3 3 1 4
## [3877] 3 4 2 4 4 2 2 3 2 2 2 1 1 4 3 2 4 1 4 4 3 1 3 4 2 2 2 1 4 2 4 4 3 4
## [3911] 4 2 4 1 4 4 2 3 2 3 2 3 4 2 2 4 3 4 1 2 1 4 2 4 1 3 4 2 3 2 4 3 1 4
## [3945] 4 4 3 1 4 3 2 3 3 4 3 3 2 4 2 3 4 2 1 2 4 3 4 2 4 4 2 3 4 1 4 4 4 3
## [3979] 3 1 1 3 4 2 2 2 1 4 4 3 4 3 4 2 1 4 1 2 2 2 1 4 4 3 1 3 1 4 4 3 3 4
## [4013] 1 4 2 2 4 2 4 2 4 3 3 1 2 3 2 4 3 3 3 4 2 2 3 2 1 3 4 4 2 4 4 2 2 2
## [4047] 2 2 3 2 4 2 1 2 4 2 2 2 1 1 1 4 4 2 1 2 4 2 2 2 2 1 4 2 4 2 1 3 1 2
## [4081] 4 3 2 3 4 1 1 2 2 4 3 4 2 2 4 4 4 1 4 1 2 4 3 4 1 3 2 3 1 1 1 1 4 1
## [4115] 4 3 2 4 4 2 3 2 2 3 1 4 1 2 2 3 4 2 1 2 1 4 2 4 1 3 3 2 2 3 1 4 4 4
## [4149] 1 4 2 4 2 4 2 3 1 1 3 2 3 4 2 1 2 4 4 1 4 3 1 3 3 2 3 3 4 4 4 1 3 2
## [4183] 2 1 2 2 4 2 4 1 1 4 3 4 4 4 4 4 3 4 2 2 4 2 2 2 1 2 3 4 2 2 1 2 4 4
## [4217] 4 2 1 3 2 1 4 1 3 2 4 2 2 4 4 4 2 3 2 1 4 2 3 2 4 3 1 2 3 3 2 4 2 4
## [4251] 4 3 4 4 4 4 4 1 4 4 4 2 3 2 1 4 2 2 4 4 4 2 1 4 4 1 4 2 1 1 4 2 2 1 3
## [4285] 2 2 4 3 3 4 4 4 1 1 2 1 4 2 4 4 2 2 4 3 4 2 2 3 1 3 1 2 4 2 2 4 2 4
## [4319] 2 2 3 3 1 2 4 2 1 4 4 3 2 3 2 2 4 1 3 4 2 2 4 3 4 2 4 2 4 4 2 4 2 1
## [4353] 4 2 4 4 4 4 4 3 3 2 4 4 4 1 1 4 4 1 1 2 3 4 3 1 1 1 2 1 2 2 2 4 2 4 2
## [4387] 3 4 4 1 4 1 2 4 4 2 3 2 3 1 2 1 4 3 3 4 2 3 4 2 4 4 4 4 4 2 1 2 4 3
## [4421] 2 3 2 2 3 4 3 2 4 2 4 1 1 4 1 1 1 2 1 2 4 1 1 2 2 1 1 2 3 1 1 2 2 1
## [4455] 2 2 2 3 1 3 4 3 1 4 4 2 4 2 4 3 4 1 1 4 3 4 1 4 4 1 1 4 3 1 4 2 1 4
## [4489] 1 3 4 3 3 1 1 2 3 3 2 1 2 3 3 4 4 2 3 1 4 2 3 4 4 1 4 1 1 3 2 3 1 3
## [4523] 3 4 4 1 4 2 3 2 4 4 3 2 4 4 2 2 1 4 4 4 1 4 1 3 1 3 1 4 1 1 4 3 2 2
## [4557] 2 1 3 2 2 2 4 3 4 2 2 1 4 3 2 2 1 1 4 4 2 1 3 4 1 1 4 1 1 3 3 4 2 3
## [4591] 4 1 2 4 3 4 4 1 1 2 4 4 3 2 3 4 4 4 2 4 2 4 1 1 4 4 1 4 3 4 2 2 2 3
## [4625] 3 3 2 2 4 3 4 4 2 4 2 4 3 4 2 2 4 3 2 4 3 2 2 4 2 3 2 1 2 3 2 2 2 1
## [4659] 1 3 3 2 4 3 4 2 4 1 2 4 2 1 1 2 4 4 2 1 2 4 3 4 2 1 3 1 4 1 2 2 4 2
## [4693] 2 1 4 2 3 2 2 1 2
```

```
##
## Within cluster sum of squares by cluster:
## [1] 1073 1025 1073 1205
##  (between_SS / total_SS =  47.6 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

Part b.)

```
kmeans4$centers
```

```
##    AGE INCOME MARRIED OWNHOME FICO..700 SEX.F SEX.M CLIMATE.10 CLIMATE.20
## 1 0.46   0.30    0.58    0.23      0.42 0.485  0.52          0          0
## 2 0.47   0.30    0.60    0.34      0.39 1.000  0.00          0          1
## 3 0.46   0.30    0.57    0.31      0.38 0.443  0.56          1          0
## 4 0.46   0.37    0.57    0.39      0.46 0.018  0.98          0          1
##   CLIMATE.30
## 1          1
## 2          0
## 3          0
## 4          0
```

```
kmeans4$withinss
```

```
## [1] 1073 1025 1073 1205
```

```
kmeans4$betweenss
```

```
## [1] 3975
```

```
kmeans4$size
```

```
## [1]  898 1361  871 1571
```

```
cluster1 <- prospects[kmeans4$cluster == 1,]
summary(cluster1)
```

```
##       AGE          INCOME      MARRIED         OWNHOME       FICO>=700
##  Min.   :18   Min.   :15   Min.   :0.00   Min.   :0.00   Min.   :0.00
##  1st Qu.:39   1st Qu.:28   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00
##  Median :44   Median :51   Median :1.00   Median :0.00   Median :0.00
##  Mean   :44   Mean   :45   Mean   :0.58   Mean   :0.23   Mean   :0.42
##  3rd Qu.:50   3rd Qu.:56   3rd Qu.:1.00   3rd Qu.:0.00   3rd Qu.:1.00
##  Max.   :75   Max.   :87   Max.   :1.00   Max.   :1.00   Max.   :1.00
##      SEX.F          SEX.M        CLIMATE.10   CLIMATE.20   CLIMATE.30
##  Min.   :0.00   Min.   :0.00   Min.   :0    Min.   :0    Min.   :1
##  1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0    1st Qu.:0    1st Qu.:1
##  Median :0.00   Median :1.00   Median :0    Median :0    Median :1
##  Mean   :0.48   Mean   :0.52   Mean   :0    Mean   :0    Mean   :1
##  3rd Qu.:1.00   3rd Qu.:1.00   3rd Qu.:0    3rd Qu.:0    3rd Qu.:1
##  Max.   :1.00   Max.   :1.00   Max.   :0    Max.   :0    Max.   :1
```

```
cluster2 <- prospects[kmeans4$cluster == 2,]
summary(cluster2)
```

```
##       AGE          INCOME      MARRIED       OWNHOME       FICO>=700
##  Min.   :18   Min.   :15   Min.   :0.0   Min.   :0.00   Min.   :0.00
##  1st Qu.:38   1st Qu.:33   1st Qu.:0.0   1st Qu.:0.00   1st Qu.:0.00
##  Median :44   Median :43   Median :1.0   Median :0.00   Median :0.00
##  Mean   :45   Mean   :45   Mean   :0.6   Mean   :0.34   Mean   :0.39
##  3rd Qu.:51   3rd Qu.:58   3rd Qu.:1.0   3rd Qu.:1.00   3rd Qu.:1.00
##  Max.   :75   Max.   :89   Max.   :1.0   Max.   :1.00   Max.   :1.00
##      SEX.F        SEX.M     CLIMATE.10   CLIMATE.20   CLIMATE.30
##  Min.   :1   Min.   :0   Min.   :0    Min.   :1    Min.   :0
##  1st Qu.:1   1st Qu.:0   1st Qu.:0    1st Qu.:1    1st Qu.:0
##  Median :1   Median :0   Median :0    Median :1    Median :0
##  Mean   :1   Mean   :0   Mean   :0    Mean   :1    Mean   :0
##  3rd Qu.:1   3rd Qu.:0   3rd Qu.:0    3rd Qu.:1    3rd Qu.:0
##  Max.   :1   Max.   :0   Max.   :0    Max.   :1    Max.   :0
```

```
cluster3 <- prospects[kmeans4$cluster == 3,]
summary(cluster3)
```

```
##       AGE          INCOME       MARRIED        OWNHOME       FICO>=700
##  Min.   :18   Min.   : 15   Min.   :0.00   Min.   :0.00   Min.   :0.00
##  1st Qu.:37   1st Qu.: 34   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00
##  Median :44   Median : 44   Median :1.00   Median :0.00   Median :0.00
##  Mean   :44   Mean   : 46   Mean   :0.57   Mean   :0.31   Mean   :0.38
##  3rd Qu.:51   3rd Qu.: 58   3rd Qu.:1.00   3rd Qu.:1.00   3rd Qu.:1.00
##  Max.   :75   Max.   :116   Max.   :1.00   Max.   :1.00   Max.   :1.00
##      SEX.F          SEX.M        CLIMATE.10   CLIMATE.20   CLIMATE.30
##  Min.   :0.00   Min.   :0.00   Min.   :1    Min.   :0    Min.   :0
##  1st Qu.:0.00   1st Qu.:0.00   1st Qu.:1    1st Qu.:0    1st Qu.:0
##  Median :0.00   Median :1.00   Median :1    Median :0    Median :0
##  Mean   :0.44   Mean   :0.56   Mean   :1    Mean   :0    Mean   :0
##  3rd Qu.:1.00   3rd Qu.:1.00   3rd Qu.:1    3rd Qu.:0    3rd Qu.:0
##  Max.   :1.00   Max.   :1.00   Max.   :1    Max.   :0    Max.   :0
```

```
cluster4 <- prospects[kmeans4$cluster == 4,]
summary(cluster4)
```

```
##       AGE          INCOME       MARRIED        OWNHOME       FICO>=700
##  Min.   :18   Min.   : 16   Min.   :0.00   Min.   :0.00   Min.   :0.00
##  1st Qu.:38   1st Qu.: 43   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00
##  Median :43   Median : 55   Median :1.00   Median :0.00   Median :0.00
##  Mean   :44   Mean   : 53   Mean   :0.57   Mean   :0.39   Mean   :0.46
##  3rd Qu.:50   3rd Qu.: 64   3rd Qu.:1.00   3rd Qu.:1.00   3rd Qu.:1.00
##  Max.   :75   Max.   :115   Max.   :1.00   Max.   :1.00   Max.   :1.00
##      SEX.F          SEX.M        CLIMATE.10   CLIMATE.20   CLIMATE.30
##  Min.   :0.00   Min.   :0.60   Min.   :0    Min.   :1    Min.   :0
##  1st Qu.:0.00   1st Qu.:1.00   1st Qu.:0    1st Qu.:1    1st Qu.:0
##  Median :0.00   Median :1.00   Median :0    Median :1    Median :0
##  Mean   :0.02   Mean   :0.98   Mean   :0    Mean   :1    Mean   :0
##  3rd Qu.:0.00   3rd Qu.:1.00   3rd Qu.:0    3rd Qu.:1    3rd Qu.:0
##  Max.   :0.40   Max.   :1.00   Max.   :0    Max.   :1    Max.   :0
```

Characteristics of Cluster members:

Cluster 1: Average age = 44, Average Income = 45K, lowest number of homeowners in this cluster, FICO (Credit Score): Second Highest among 4 clusters, CLIMATE Code for all samples is 30.

The size of the cluster 1 is: 898, The variance within cluster is: 1073

Cluster 2: Average age = 45, Average Income = 45K, more number of homeowners in this cluster (2nd Highest among 4 clusters), FICO (Credit Score): Second Lowest among 4 clusters, CLIMATE Code for all samples is 20.

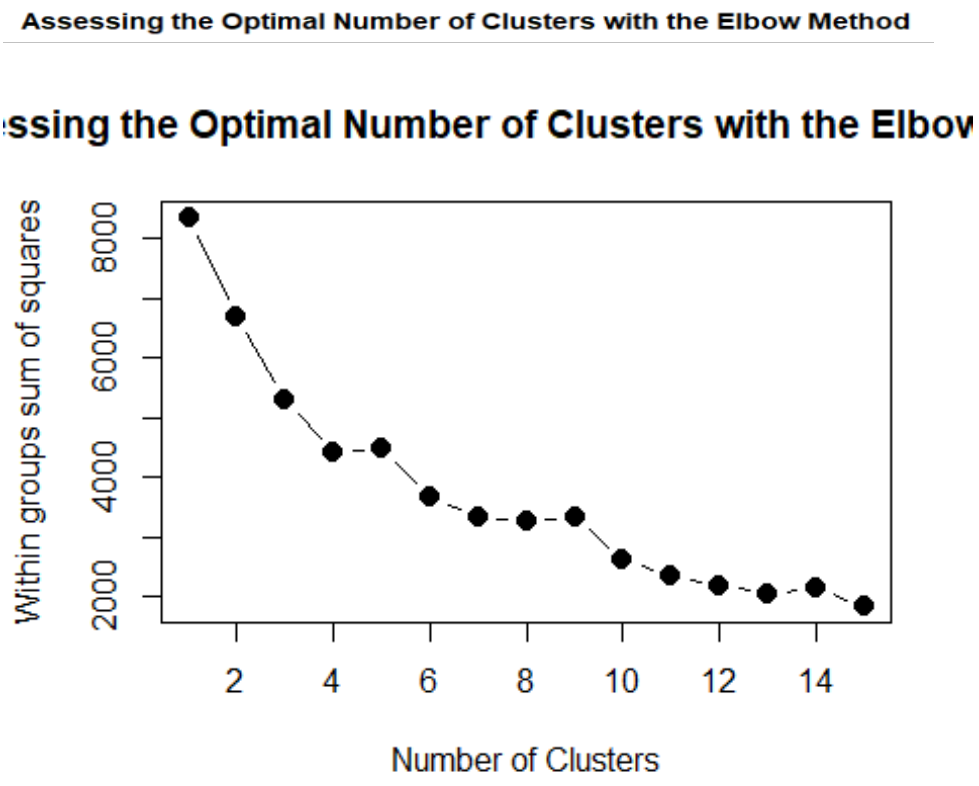
The size of the cluster 2 is: 1361, The variance within cluster is: 1025

Cluster 3: Average age = 44, Average Income = 46K, less number of homeowners in this cluster (3rd Highest among 4 clusters), FICO (Credit Score): Lowest among 4 clusters, CLIMATE Code for all samples is 10.
The size of the cluster 3 is: 871, The variance within cluster is: 1073

Cluster 4: Average age = 44, Average Income = 53K (Highest income samples), Highest number of homeowners in this cluster, FICO (Credit Score): Highest among 4 clusters, CLIMATE Code for all samples is 20.
The size of the cluster 4 is: 1571, The variance within cluster is: 1205

Part c. We use scree plot and elbow-point to find the best number of clusters.

```
pp <- scaled_prospects
wss <- (nrow(pp)-1)*sum(sapply(pp,var))
for (i in 1:15)
  wss[i] <- sum(kmeans(pp, centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of Clusters", ylab="Within groups sum of squares",
     main="Assessing the Optimal Number of Clusters with the Elbow Method", pch=20, cex=2
)
```



From the Scree plot, I think six clusters (k=6) is the best option.

Part d.)

```
set.seed(7)
kmeans6 <- kmeans(scaled_prospects, centers=6, nstart =100)

library(cluster)
ss <- silhouette(kmeans6$cluster, dist(scaled_prospects))
mean(ss[ ,3])

## [1] 0.38
```

The Silhoutte measure for 6 clusters, is 0.38. Higher the Silhoutte coefficient, the better the clusters.

```
set.seed(7)
kmeans4 <- kmeans(scaled_prospects, centers=4, nstart =100)

library(cluster)
ss4 <- silhouette(kmeans4$cluster, dist(scaled_prospects))
mean(ss4[ ,3])

## [1] 0.31
```

For 4 clusters Silhouette coefficient comes to be 0.31. So solution with six clusters is better.