STA 380 Part 2 Q4 Market Segmentation

Vivek Mehendiratta

8/9/2021

Market Segmentation

Required Libraries

```
# required libraries
library(tidyverse)
```

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

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.0     v forcats 0.5.1
```

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

```
library(ggplot2)
library(corrplot)
```

```
## corrplot 0.90 loaded
```

Data import and Exploration

```
mkt = read.csv('https://raw.githubusercontent.com/jgscott/STA380/master/data/social_marketing.csv')
```

```
head(mkt)
```

```
##           X chatter current_events travel photo_sharing uncategorized tv_film
## 1 hmjoe4g3k       2              0      2             2             2       1
## 2 clk1m5w8s       3              3      2             1             1       1
## 3 jcsovtak3       6              3      4             3             1       5
## 4 3oeb4hiln       1              5      2             2             0       1
## 5 fd75x1vgk       5              2      0             6             1       0
```

```
## 6 h6nvj91yp       6             4      2            7           0       1
##   sports_fandom politics food family home_and_garden music news online_gaming
## 1             1        0    4      1               2     0    0             0
## 2             4        1    2      2               1     0    0             0
## 3             0        2    1      1               1     1    1             0
## 4             0        1    0      1               0     0    0             0
## 5             0        2    0      1               0     0    0             3
## 6             1        0    2      1               1     1    0             0
##   shopping health_nutrition college_uni sports_playing cooking eco computers
## 1        1               17           0              2       5   1         1
## 2        0                0           0              1       0   0         0
## 3        2                0           0              0       2   1         0
## 4        0                0           1              0       0   0         0
## 5        2                0           4              0       1   0         1
## 6        5                0           0              0       0   0         1
##   business outdoors crafts automotive art religion beauty parenting dating
## 1        0        2      1          0   0        1      0         1      1
## 2        1        0      2          0   0        0      0         0      1
## 3        0        0      2          0   8        0      1         0      1
## 4        1        0      3          0   2        0      1         0      0
## 5        0        1      0          0   0        0      0         0      0
## 6        1        0      0          1   0        0      0         0      0
##   school personal_fitness fashion small_business spam adult
## 1      0               11       0              0    0     0
## 2      4                0       0              0    0     0
## 3      0                0       1              0    0     0
## 4      0                0       0              0    0     0
## 5      0                0       0              1    0     0
## 6      0                0       0              0    0     0
```

```
summary(mkt)
```

```
##       X                chatter       current_events      travel
##  Length:7882        Min.   : 0.000   Min.   :0.000   Min.   : 0.000
##  Class :character   1st Qu.: 2.000   1st Qu.:1.000   1st Qu.: 0.000
##  Mode  :character   Median : 3.000   Median :1.000   Median : 1.000
##                     Mean   : 4.399   Mean   :1.526   Mean   : 1.585
##                     3rd Qu.: 6.000   3rd Qu.:2.000   3rd Qu.: 2.000
##                     Max.   :26.000   Max.   :8.000   Max.   :26.000
##  photo_sharing    uncategorized      tv_film      sports_fandom
##  Min.   : 0.000   Min.   :0.000   Min.   : 0.00   Min.   : 0.000
##  1st Qu.: 1.000   1st Qu.:0.000   1st Qu.: 0.00   1st Qu.: 0.000
##  Median : 2.000   Median :1.000   Median : 1.00   Median : 1.000
##  Mean   : 2.697   Mean   :0.813   Mean   : 1.07   Mean   : 1.594
##  3rd Qu.: 4.000   3rd Qu.:1.000   3rd Qu.: 1.00   3rd Qu.: 2.000
##  Max.   :21.000   Max.   :9.000   Max.   :17.00   Max.   :20.000
##     politics           food            family        home_and_garden
##  Min.   : 0.000   Min.   : 0.000   Min.   : 0.0000   Min.   :0.0000
##  1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.:0.0000
##  Median : 1.000   Median : 1.000   Median : 1.0000   Median :0.0000
##  Mean   : 1.789   Mean   : 1.397   Mean   : 0.8639   Mean   :0.5207
##  3rd Qu.: 2.000   3rd Qu.: 2.000   3rd Qu.: 1.0000   3rd Qu.:1.0000
##  Max.   :37.000   Max.   :16.000   Max.   :10.0000   Max.   :5.0000
##      music              news        online_gaming       shopping
```

```
##  Min.   : 0.0000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
##  1st Qu.: 0.0000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000
##  Median : 0.0000   Median : 0.000   Median : 0.000   Median : 1.000
##  Mean   : 0.6793   Mean   : 1.206   Mean   : 1.209   Mean   : 1.389
##  3rd Qu.: 1.0000   3rd Qu.: 1.000   3rd Qu.: 1.000   3rd Qu.: 2.000
##  Max.   :13.0000   Max.   :20.000   Max.   :27.000   Max.   :12.000
##  health_nutrition  college_uni     sports_playing      cooking
##  Min.   : 0.000   Min.   : 0.000   Min.   :0.0000   Min.   : 0.000
##  1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.: 0.000
##  Median : 1.000   Median : 1.000   Median :0.0000   Median : 1.000
##  Mean   : 2.567   Mean   : 1.549   Mean   :0.6392   Mean   : 1.998
##  3rd Qu.: 3.000   3rd Qu.: 2.000   3rd Qu.:1.0000   3rd Qu.: 2.000
##  Max.   :41.000   Max.   :30.000   Max.   :8.0000   Max.   :33.000
##       eco           computers          business         outdoors
##  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 :0.0000   Median : 0.0000
##  Mean   :0.5123   Mean   : 0.6491   Mean   :0.4232   Mean   : 0.7827
##  3rd Qu.:1.0000   3rd Qu.: 1.0000   3rd Qu.:1.0000   3rd Qu.: 1.0000
##  Max.   :6.0000   Max.   :16.0000   Max.   :6.0000   Max.   :12.0000
##      crafts         automotive           art             religion
##  Min.   :0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.000
##  1st Qu.:0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.000
##  Median :0.0000   Median : 0.0000   Median : 0.0000   Median : 0.000
##  Mean   :0.5159   Mean   : 0.8299   Mean   : 0.7248   Mean   : 1.095
##  3rd Qu.:1.0000   3rd Qu.: 1.0000   3rd Qu.: 1.0000   3rd Qu.: 1.000
##  Max.   :7.0000   Max.   :13.0000   Max.   :18.0000   Max.   :20.000
##      beauty          parenting           dating            school
##  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 : 0.0000   Median : 0.0000
##  Mean   : 0.7052   Mean   : 0.9213   Mean   : 0.7109   Mean   : 0.7677
##  3rd Qu.: 1.0000   3rd Qu.: 1.0000   3rd Qu.: 1.0000   3rd Qu.: 1.0000
##  Max.   :14.0000   Max.   :14.0000   Max.   :24.0000   Max.   :11.0000
##  personal_fitness    fashion        small_business        spam
##  Min.   : 0.000   Min.   : 0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median : 0.000   Median : 0.0000   Median :0.0000   Median :0.00000
##  Mean   : 1.462   Mean   : 0.9966   Mean   :0.3363   Mean   :0.00647
##  3rd Qu.: 2.000   3rd Qu.: 1.0000   3rd Qu.:1.0000   3rd Qu.:0.00000
##  Max.   :19.000   Max.   :18.0000   Max.   :6.0000   Max.   :2.00000
##      adult
##  Min.   : 0.0000
##  1st Qu.: 0.0000
##  Median : 0.0000
##  Mean   : 0.4033
##  3rd Qu.: 0.0000
##  Max.   :26.0000
```

```
colnames(mkt)
```

```
##  [1] "X"                "chatter"          "current_events"   "travel"
##  [5] "photo_sharing"    "uncategorized"    "tv_film"          "sports_fandom"
##  [9] "politics"         "food"             "family"           "home_and_garden"
```

```
## [13] "music"            "news"             "online_gaming"    "shopping"
## [17] "health_nutrition" "college_uni"      "sports_playing"   "cooking"
## [21] "eco"              "computers"        "business"         "outdoors"
## [25] "crafts"           "automotive"       "art"              "religion"
## [29] "beauty"           "parenting"        "dating"           "school"
## [33] "personal_fitness" "fashion"          "small_business"   "spam"
## [37] "adult"
```

To perform cluster analysis on the data, I will be using K-Means Clustering approach. Following columns shall be removed to perform cluster analysis : * *X* : Since it is unique value for each user, doesn't make much sense to keep it * *chatter* : Random values which doesn't fit in any column. It won't lie in any segment. * *adult* : Target variable, no need in cluster analysis * *spam* : Target variable

```
drop_columns = c("X", "chatter", "adult", "spam", "uncategorized")
mkt_km = mkt[, !(names(mkt) %in% drop_columns)]
```

Data Standardization

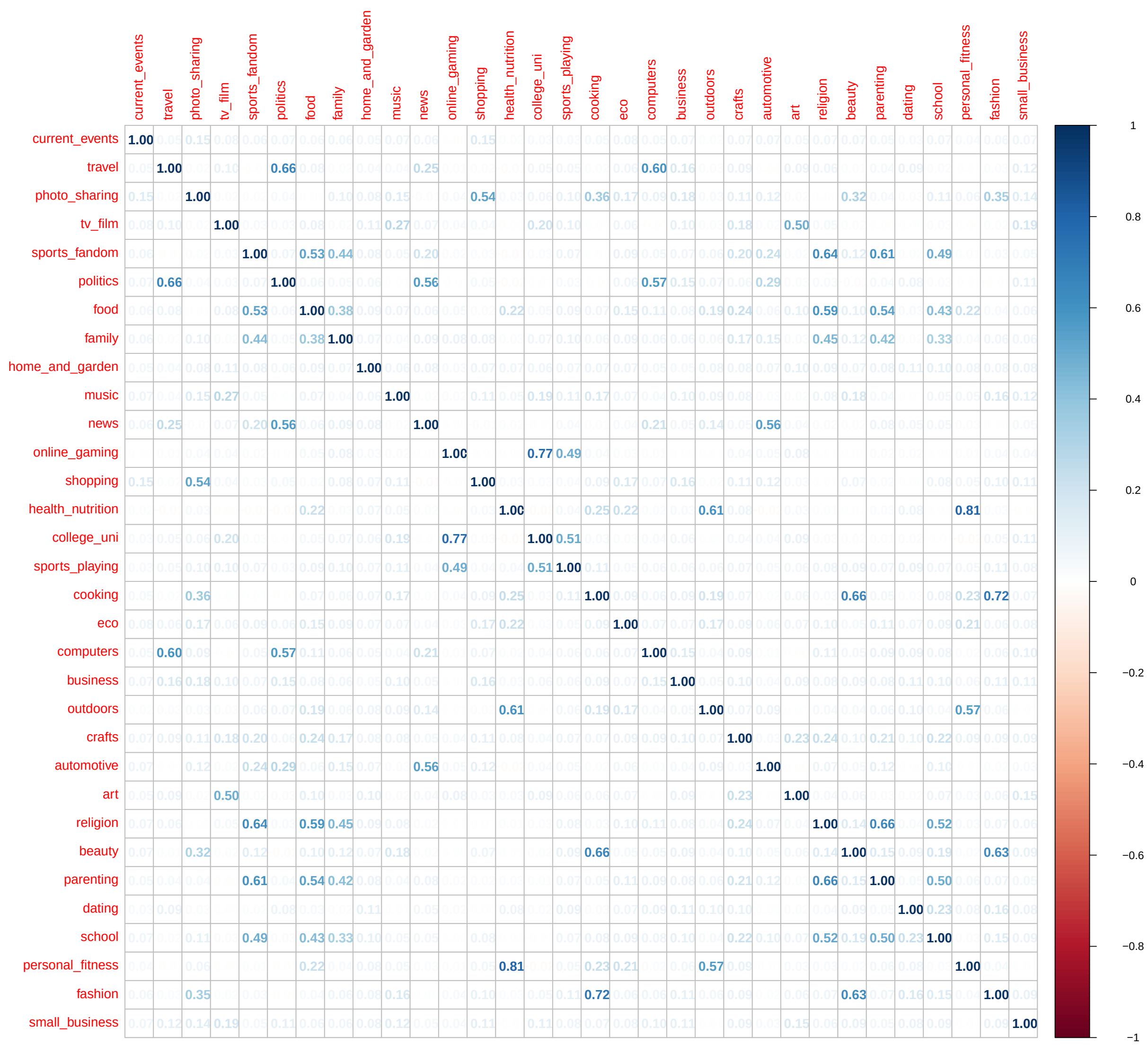
```
set.seed(1)

mkt_scaled = scale(mkt_km)
scaled_mean = attr(mkt_scaled, "scaled:center")
scaled_sd = attr(mkt_scaled, "scaled:scale")

cat("Scaled Mean :\n", scaled_mean, "\n\n")
```

```
## Scaled Mean :
##  1.526262 1.585004 2.696777 1.070287 1.594012 1.788632 1.397488 0.863867 0.52068 0.6792692 1.205532
```

```
cat("Scaled SD :\n", scaled_sd)
```

```
## Scaled SD :
##  1.26889 2.28553 2.73151 1.658783 2.160917 3.031113 1.775557 1.132562 0.7366913 1.030015 2.10078 2.6
```

```
corrplot(cor(mkt_km), method = "number")
```




```r
cr = cor(mkt_km)
cr[upper.tri(cr, diag=TRUE)] <- NA
cr = reshape2::melt(cr, na.rm=TRUE, value.name="corr")
```

```r
cr = cr %>% arrange(desc(corr))
head(cr, 10)
```

```
##                Var1             Var2      corr
## 1  personal_fitness health_nutrition 0.8099024
## 2       college_uni     online_gaming 0.7728393
## 3           fashion          cooking 0.7214027
## 4            beauty          cooking 0.6642389
## 5          politics           travel 0.6602100
```

```
## 6         parenting         religion 0.6555973
## 7          religion    sports_fandom 0.6379748
## 8           fashion           beauty 0.6349739
## 9          outdoors health_nutrition 0.6082254
## 10        parenting    sports_fandom 0.6077181
```

Let's see if there is any highly negatively correlated variables ?

```
tail(cr, 10)
```

```
##                 Var1             Var2         corr
## 487 personal_fitness       automotive -0.009861229
## 488 health_nutrition    sports_fandom -0.011229255
## 489           beauty         politics -0.011292710
## 490         shopping             news -0.011813142
## 491 health_nutrition           travel -0.011922499
## 492             news    photo_sharing -0.011980028
## 493 health_nutrition         politics -0.016851900
## 494 personal_fitness      college_uni -0.021526868
## 495       automotive health_nutrition -0.023824999
## 496      college_uni health_nutrition -0.027778856
```

The data has almost positive correlation with personal_fitness and health_nutrition being top correlated variables.

Clustering

```
#k-means clustering with 10 clusters
cl = kmeans(mkt_scaled, 10, nstart=25)
```

Distribution of columns in each cluster :

```
for(i in c(1:10)){
  a = mkt_scaled[which(cl$cluster == i),]
  cat("cluster No :", i, "\n")
  print(sort(colSums(a[, 2:ncol(a)]), decreasing = T)[0:5])
  cat("\n")
}
```

```
## cluster No : 1
## health_nutrition personal_fitness         outdoors              eco
##        1723.5995        1673.6716        1337.3631         444.8355
##             food
##         358.1745
##
## cluster No : 2
##      religion     parenting sports_fandom          food        school
##      1537.241      1453.985      1402.637      1241.678      1116.322
##
```

```
## cluster No : 3
##    travel  politics computers      news  business
## 1150.7524 1094.7797 1026.3450  403.0183  193.5624
##
## cluster No : 4
##          news    automotive      politics sports_fandom      outdoors
##     1129.1330     1109.6628      522.3851      285.8661      130.5772
##
## cluster No : 5
##          dating          school         fashion home_and_garden        business
##       961.48250       257.65186       165.00401       123.93517        93.99723
##
## cluster No : 6
## home_and_garden          dating          travel  small_business         tv_film
##       -697.6195       -717.1647       -717.4607       -729.4961       -732.8135
##
## cluster No : 7
##       shopping  photo_sharing            eco       business small_business
##      1383.2380      1083.5140       313.5060       308.0282       180.0068
##
## cluster No : 8
##   online_gaming    college_uni sports_playing            art         family
##     1280.84978     1176.42546      761.82617      100.34261       74.81764
##
## cluster No : 9
##       cooking       fashion        beauty photo_sharing         music
##     1351.8754     1291.6365     1262.5378      606.4583      261.1135
##
## cluster No : 10
##        tv_film            art          music small_business         crafts
##      1119.9018      1087.3378       408.0082       340.4638       312.3090
```

FInding the optimal value of `k` for K-Means Clustering

```r
kmean_withinss = function(k) {
    cl = kmeans(mkt_scaled, k)
    return (cl$tot.withinss)
}
```

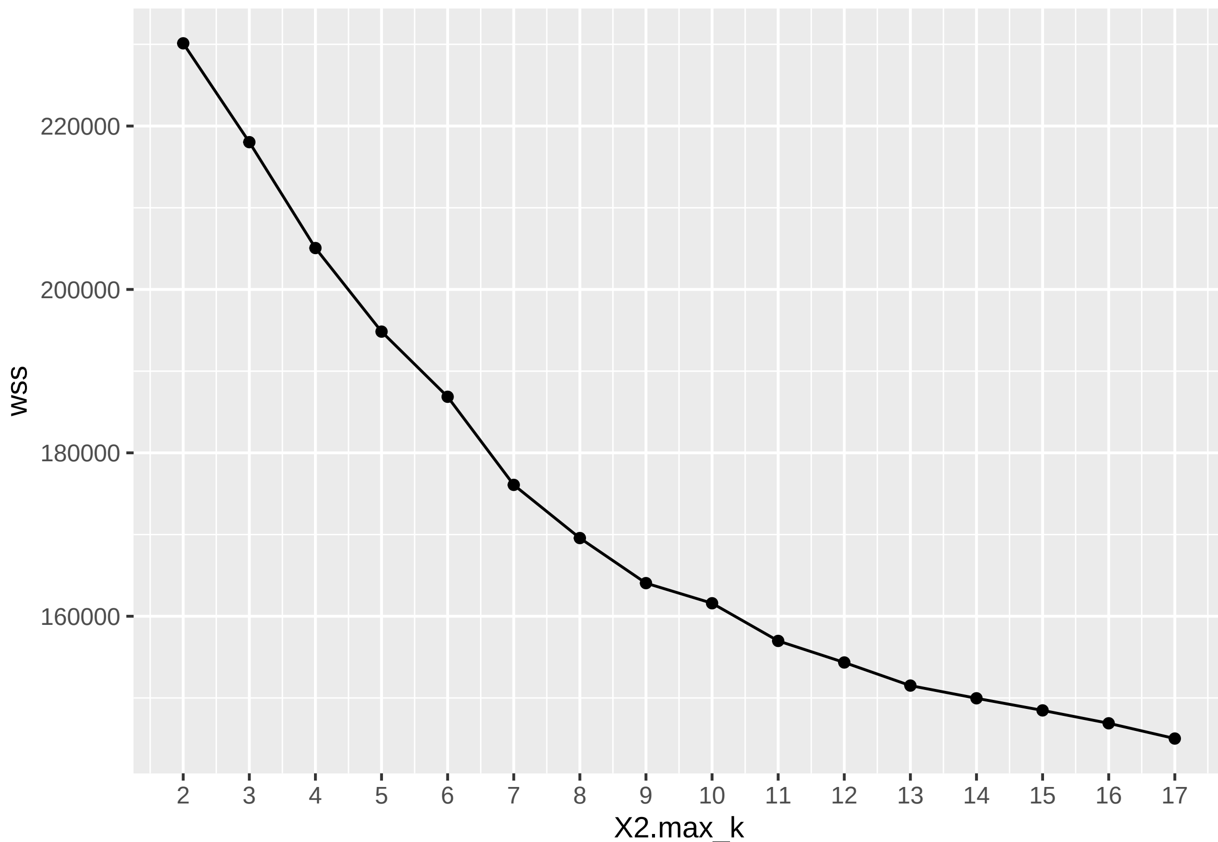
```r
kmean_withinss(2)
```

```
## [1] 230126.3
```

```r
# Setting maximum cluster
max_k = 17

# Run algorithm over a range of k
wss = sapply(2:max_k, kmean_withinss)
```

```r
# Create a data frame to plot the graph
elbow = data.frame(2:max_k, wss)
```

```r
# Plot the graph with gglop
ggplot(elbow, aes(x = X2.max_k, y = wss)) +
    geom_point() +
    geom_line() +
    scale_x_continuous(breaks = seq(1, 20, by = 1))
```




```
Optimal K : 11
```

```r
cl2 = kmeans(mkt_scaled, 11)
```

```r
cl2$size
```

```
##  [1]  406 2958  333  475  773  332  207  742  493  357  806
```

```r
cl2$centers
```

```
##    current_events      travel photo_sharing     tv_film sports_fandom
## 1     0.320938006  0.22360477  -0.064826061  2.78182343    -0.1232928
## 2    -0.213590027 -0.23628704  -0.391849450 -0.21784749    -0.4196009
## 3     0.150884065  3.35469238  -0.087981011 -0.05323465    -0.1928970
## 4     0.185865451 -0.05423302   1.274044265 -0.14390534    -0.2310476
```

```
## 5    -0.008972617 -0.14954763  -0.080327927 -0.14141792    -0.2126277
## 6     0.257033435  0.02738383   0.099982080  0.01573371     2.8181193
## 7     0.228674393 -0.10588734  -0.003949145  0.14692873     0.9591551
## 8    -0.032380390 -0.24003876  -0.241767129 -0.22274043     0.8969955
## 9    -0.080643116  0.01117634  -0.408062709 -0.18299705     0.1813070
## 10   -0.099064617 -0.03903056  -0.007947047  0.11129557    -0.1400771
## 11    0.417348007 -0.20873209   1.268805006 -0.13586680    -0.2599607
##       politics        food      family home_and_garden       music
## 1  -0.08953507  0.12570797 -0.12554914      0.30292518  1.05268159
## 2  -0.35275067 -0.43863788 -0.36694790     -0.21438415 -0.23312399
## 3   3.15683496  0.18881102 -0.06540680      0.06364612 -0.04430576
## 4  -0.13168713 -0.20963828  0.03469202      0.13053240  0.54847884
## 5  -0.19658678  0.45372618 -0.08654688      0.16421514  0.01497723
## 6  -0.13596532  2.55143716  2.04567717      0.30310645  0.20318640
## 7   1.58700805  0.01012132  0.40172071      0.30964429 -0.06851705
## 8  -0.32242659  0.77957676  0.64021384      0.05608127 -0.12432471
## 9   0.76836936 -0.32211303 -0.08576376     -0.03495497 -0.11398195
## 10 -0.16776682 -0.11027933  0.19439696      0.06128252 -0.04758875
## 11 -0.14884410 -0.36851064 -0.06931828      0.11171252  0.14515801
##             news online_gaming    shopping health_nutrition   college_uni
## 1   0.0311333715   -0.16568903  0.06524657     -0.155743844  0.3884490977
## 2  -0.3937753478   -0.23079628 -0.38122154     -0.301204785 -0.2550077859
## 3   1.1372265549   -0.15815889 -0.03429858     -0.160225068 -0.0227797122
## 4  -0.0878144993   -0.02208645  0.22235767     -0.046093234 -0.0130821056
## 5  -0.0762828284   -0.11284520 -0.02001205      2.233773274 -0.2021646942
## 6  -0.0003390711    0.04782170  0.10612132     -0.006246506 -0.0004442485
## 7   3.6113911500    0.01756615  0.03578973     -0.204597279 -0.0812763600
## 8  -0.3012001529   -0.18502194 -0.18471010     -0.290723343 -0.2287373329
## 9   1.3253773295   -0.22941255 -0.35879401     -0.315641316 -0.2736780374
## 10 -0.2018388171    3.56301718 -0.12389230     -0.180986291  3.2968075318
## 11 -0.2868238809   -0.17465387  1.66000397     -0.269102086 -0.1100081496
##    sports_playing     cooking          eco   computers     business    outdoors
## 1     0.102241570 -0.14166674  0.112008394 -0.14727534  0.384778207 -0.08301346
## 2    -0.266043012 -0.31922147 -0.279949828 -0.26512106 -0.253875890 -0.33126098
## 3     0.031257065 -0.18071607  0.188835184  3.08550962  0.603135023 -0.03880471
## 4     0.199387967  2.81474474  0.001789846  0.08155416  0.221842775  0.03173526
## 5    -0.001438182  0.41577891  0.571409394 -0.08526116  0.070688824  1.73916341
## 6     0.252547432  0.09799398  0.426197644  0.24390381  0.237018413  0.02777690
## 7     0.003417884 -0.13891921  0.125225435 -0.07109773 -0.046128616  0.39936993
## 8    -0.076357308 -0.28199510 -0.007233723 -0.11382311  0.005749579 -0.19805380
## 9    -0.176979433 -0.32474102 -0.312450926 -0.17196643 -0.207002195  0.04216845
## 10    2.124316512 -0.12769891 -0.072394554 -0.08008116 -0.117718966 -0.14917819
## 11   -0.071448238 -0.22990087  0.366024480 -0.03066727  0.377857649 -0.30449153
##         crafts  automotive         art     religion      beauty   parenting
## 1   0.75250634 -0.20359491  2.65523623 -0.004804794  0.00316938 -0.19840196
## 2  -0.31946838 -0.37681616 -0.23836888 -0.386332859 -0.28204035 -0.39360856
## 3   0.21403232 -0.13924221 -0.14993816  0.146210469 -0.18276075  0.04596353
## 4   0.08496596  0.02128744  0.01125560 -0.160870113  2.64136849 -0.07868402
## 5   0.06056328 -0.17659083 -0.06293974 -0.172784873 -0.20563868 -0.09492768
## 6   0.98355954  0.28548383  0.09123722  3.037995485  0.57816023  2.98373701
## 7   0.03676916  3.62179028 -0.06532900 -0.097760504 -0.04717130  0.20493124
## 8   0.33036198 -0.11025153 -0.11976520  1.072778773  0.06878891  0.90658974
## 9  -0.30622463  1.03617436 -0.26056674 -0.293466703 -0.29578803 -0.18100069
## 10  0.01659165  0.05482292  0.27028514 -0.210739655 -0.22726786 -0.14403095
```

```
## 11  0.05804818  0.09547426 -0.21790274 -0.320664415 -0.25820271 -0.27886500
##          dating      school personal_fitness     fashion small_business
## 1  -0.06579198 -0.04080484      -0.14807400 -0.02102720     0.82659114
## 2  -0.18032845 -0.38773896      -0.32183869 -0.28989249    -0.22307170
## 3   0.34755868 -0.07238898      -0.14341511 -0.17057956     0.39836269
## 4   0.02284559  0.15829115      -0.01880198  2.73188065     0.18473016
## 5   0.18690205 -0.16595203       2.16959272 -0.10567143    -0.12554257
## 6  -0.02704837  2.32222893       0.08840405  0.18143542     0.19654776
## 7   0.10530528  0.26054596      -0.17804805 -0.13023340    -0.05956360
## 8   0.60759234  0.93385344      -0.28960337 -0.03276985     0.01839844
## 9  -0.12001187 -0.26710864      -0.29667955 -0.31651753    -0.18971062
## 10 -0.01693717 -0.22175033      -0.18628456 -0.07013031     0.10842956
## 11 -0.13570734 -0.08224078      -0.21480399 -0.15012483     0.23465885
```

```
for(i in c(1:10)){
  a = mkt_scaled[which(cl$cluster == i),]
  cat("cluster No :", i, "\n")
  print(names(sort(colSums(a[, 2:ncol(a)]), decreasing = T))[1:10])
  cat("\n")
}
```

```
## cluster No : 1
##  [1] "health_nutrition" "personal_fitness" "outdoors"         "eco"
##  [5] "food"             "cooking"          "home_and_garden"  "crafts"
##  [9] "business"         "dating"
##
## cluster No : 2
##  [1] "religion"        "parenting"       "sports_fandom"   "food"
##  [5] "school"          "family"          "crafts"          "beauty"
##  [9] "eco"             "home_and_garden"
##
## cluster No : 3
##  [1] "travel"         "politics"       "computers"      "news"
##  [5] "business"       "small_business" "dating"         "crafts"
##  [9] "eco"            "food"
##
## cluster No : 4
##  [1] "news"            "automotive"      "politics"        "sports_fandom"
##  [5] "outdoors"        "family"          "home_and_garden" "parenting"
##  [9] "school"          "tv_film"
##
## cluster No : 5
##  [1] "dating"          "school"          "fashion"         "home_and_garden"
##  [5] "business"        "crafts"          "sports_playing"  "small_business"
##  [9] "beauty"          "eco"
##
## cluster No : 6
##  [1] "home_and_garden" "dating"          "travel"          "small_business"
##  [5] "tv_film"         "music"           "online_gaming"   "art"
##  [9] "computers"       "business"
##
## cluster No : 7
##  [1] "shopping"        "photo_sharing"   "eco"             "business"
##  [5] "small_business"  "music"           "home_and_garden" "automotive"
```

```
##  [9] "crafts"          "computers"
##
## cluster No : 8
##  [1] "online_gaming"   "college_uni"     "sports_playing"  "art"
##  [5] "family"          "tv_film"         "small_business"  "home_and_garden"
##  [9] "automotive"      "crafts"
##
## cluster No : 9
##  [1] "cooking"        "fashion"        "beauty"         "photo_sharing"
##  [5] "music"          "business"       "shopping"       "sports_playing"
##  [9] "small_business" "school"
##
## cluster No : 10
##  [1] "tv_film"         "art"             "music"           "small_business"
##  [5] "crafts"          "college_uni"     "business"        "home_and_garden"
##  [9] "travel"          "food"
```

Insights

Some clusters turned out to be meaningful and informative, below are the some categories which can be clubbed together :

- **Cluster 1** contains categories like `personal_fitness`, `nutrition_health` and `outdoors` can be taken posts related to fitness.
- **Cluster 2** contains categories like `parenting`, `food`, `school`, `sports`, `religion`, `crafts`. These can be considered as posts from educational institutions.

- **Cluster 3** contains categories like `travel`, `politics`, `computers`, `news` which clearly show that the posts belong to news.
- **Cluster 6** contains categories like `home_and_garden`, `dating`, `travel`, `small_business`, `tv_film`,`music` can be related lifestyle posts.
- **Cluster 8** contains categories like `online_gaming`, `college_uni`, `sports_playing` can be related college online ganing event.

Association rule mining