

Market Segmentation

STA380 Exam 2

2023-08-14

Market Segmentation

```
# importing dataset
social = read.csv('C:/Users/ianwm/Downloads/social_marketing.csv', header=TRUE)
set.seed(1)
```

```
# centering and scaling data
X = social[,-(1)]
X = scale(X, center=TRUE, scale=TRUE)
```

```
# Extract the centers and scales from the rescaled data (which are named attributes)
mu = attr(X,"scaled:center")
sigma = attr(X,"scaled:scale")
```

```
# Run k-means with 6 clusters and 25 starts
clust1 = kmeans(X, 6, nstart=25)
```

```
# What are the clusters?
round(clust1$center, 3)
```

```
##   chatter current_events travel photo_sharing uncategorized tv_film
## 1  -0.020      -0.065 -0.213      -0.147      -0.090  -0.044
## 2  -0.013       0.026 -0.150      -0.015       0.164  -0.052
## 3  -0.044       0.118 -0.105      -0.029      -0.072  -0.008
## 4   0.020      -0.014  0.001       0.048       0.112   0.442
## 5   0.042       0.111  1.762      -0.057      -0.040   0.078
## 6   0.169       0.199 -0.039       1.253       0.516   0.009
##  sports_fandom politics   food family home_and_garden music  news
## 1    -0.288    -0.257 -0.354 -0.256    -0.110 -0.118 -0.244
## 2    -0.199    -0.176  0.412 -0.079     0.156  0.058 -0.048
## 3     1.996    -0.204  1.789  1.442     0.168  0.054 -0.080
## 4    -0.116    -0.165 -0.075  0.187     0.128  0.331 -0.192
## 5     0.195     2.366  0.025  0.044     0.123 -0.037  1.958
## 6    -0.194    -0.114 -0.178  0.048     0.161  0.565 -0.070
##  online_gaming shopping health_nutrition college_uni sports_playing cooking
## 1    -0.231    -0.061      -0.328    -0.224     -0.226  -0.331
## 2    -0.135     0.037       2.101    -0.208     -0.031   0.374
## 3    -0.075     0.046      -0.160    -0.123      0.101  -0.118
```

```
## 4      3.086  -0.016      -0.182      3.080      2.000  -0.154
## 5      -0.142  -0.005      -0.206     -0.080      -0.010  -0.215
## 6      -0.053   0.381      -0.064     -0.004      0.182   2.570
##      eco computers business outdoors crafts automotive art religion beauty
## 1 -0.160   -0.233   -0.125   -0.315 -0.187   -0.182 -0.063   -0.298 -0.265
## 2  0.527   -0.073   0.068   1.620  0.090   -0.121  0.009   -0.174 -0.211
## 3  0.195    0.069   0.116   -0.081  0.701    0.165  0.085    2.179  0.294
## 4 -0.032   -0.056   0.030   -0.101  0.105    0.046  0.309   -0.130 -0.198
## 5  0.106    1.547   0.357   0.111  0.153    1.111 -0.004   -0.034 -0.174
## 6  0.086    0.072   0.287   0.034  0.151    0.054  0.137   -0.118  2.389
##      parenting dating school personal_fitness fashion small_business spam adult
## 1   -0.305 -0.094 -0.244      -0.333 -0.264      -0.096  0.004  0.008
## 2   -0.106  0.183 -0.144      2.070 -0.107      -0.069  0.003  0.007
## 3    2.069  0.039  1.631      -0.114  0.007      0.104 -0.014 -0.004
## 4   -0.148  0.017 -0.203      -0.189 -0.054      0.222  0.032  0.033
## 5    0.017  0.201 -0.035      -0.192 -0.179      0.239 -0.007 -0.092
## 6   -0.065  0.157  0.197      -0.046  2.498      0.276 -0.036  0.019
```

We can see some clusters starting to form here. While cluster 1 does not have any features that stand out, cluster 2 appears to be very focused on health and nutrition, the outdoors and personal fitness. Cluster 3 focuses on sports fandom, food, family, religion, parenting and school, while cluster 4 is more focused on online gaming, college, and playing sports. Cluster 5 likes travel, politics, news, computers and cars, and finally cluster 6 likes photo sharing, cooking and beauty. While these are some good clusters, the one cluster without any big interests tells us to try a model with fewer clusters.

```
# Run k-means with 5 clusters and 25 starts
clust2 = kmeans(X, 5, nstart=25)
```

```
# What are the clusters?
round(clust2$center, 3)
```

```
##      chatter current_events travel photo_sharing uncategorized tv_film
## 1  -0.212      -0.129 -0.234      -0.314      -0.166  -0.149
## 2  -0.006      0.103  1.848      -0.109      -0.046   0.031
## 3  -0.087      -0.011 -0.155      -0.052      0.148  -0.104
## 4   0.719      0.275 -0.037      1.031      0.457   0.487
## 5  -0.086      0.117 -0.099      -0.065      -0.077  -0.005
##      sports_fandom politics food family home_and_garden music news
## 1   -0.298   -0.273 -0.366 -0.271      -0.171 -0.202 -0.248
## 2    0.191    2.433  0.035  0.049      0.124 -0.049  1.932
## 3   -0.192   -0.181  0.435 -0.076      0.140  0.021 -0.031
## 4   -0.165   -0.132 -0.155  0.041      0.267  0.573 -0.128
## 5    2.050   -0.199  1.820  1.495      0.170  0.049 -0.067
##      online_gaming shopping health_nutrition college_uni sports_playing cooking
## 1   -0.127   -0.241      -0.329   -0.157      -0.207  -0.340
## 2   -0.072   -0.056      -0.203   -0.027      0.029  -0.193
## 3   -0.063   -0.006      2.157   -0.159      0.019   0.433
## 4    0.428    0.734      -0.176    0.585      0.503   0.865
## 5    0.025   -0.001      -0.155   -0.032      0.145  -0.087
##      eco computers business outdoors crafts automotive art religion beauty
## 1 -0.236   -0.253   -0.210   -0.316 -0.258   -0.207 -0.158   -0.304 -0.291
## 2  0.096    1.640    0.322    0.106  0.101    1.067 -0.057   -0.030 -0.153
## 3  0.531   -0.082    0.028    1.660  0.083   -0.129 -0.017   -0.167 -0.169
```

```
## 4  0.233    -0.021    0.383   -0.103  0.289      0.092  0.442   -0.164  0.861
## 5  0.172     0.076    0.121   -0.069  0.705      0.172  0.107    2.256  0.313
##   parenting dating school personal_fitness fashion small_business spam adult
## 1   -0.312 -0.170 -0.306      -0.344 -0.295      -0.181 -0.006 -0.006
## 2    0.021  0.205 -0.046      -0.196 -0.159      0.226  0.012 -0.092
## 3   -0.105  0.162 -0.175      2.119 -0.077      -0.131  0.005 -0.014
## 4   -0.135  0.308  0.162      -0.143  0.966      0.453  0.015  0.058
## 5    2.123 -0.011  1.658      -0.098  0.028      0.094 -0.013  0.018
```

Unfortunately, cluster 1 still does not have any major interests. Cluster 2 appears to be similar to the cluster 5 of the last model. Cluster 3 appears similar to cluster 2 of the last model, while cluster 4 appears somewhat similar to cluster 6. Finally, cluster 5 appears to be similar to cluster 3 of the last model. We will try with 4 clusters to see if we get any definite segments.

```
# Run k-means with 4 clusters and 25 starts
clust3 = kmeans(X, 4, nstart=25)
```

```
# What are the clusters?
round(clust3$center, 3)
```

```
##   chatter current_events travel photo_sharing uncategorized tv_film
## 1   0.551      0.212 -0.074      0.802      0.418  0.362
## 2  -0.082      0.121 -0.106     -0.054     -0.071 -0.011
## 3  -0.207     -0.121 -0.229     -0.298     -0.150 -0.150
## 4   0.002      0.103  1.769     -0.092     -0.032  0.044
##   sports_fandom politics   food family home_and_garden music  news
## 1   -0.177   -0.156  0.069  0.041      0.268  0.485 -0.113
## 2    2.021   -0.199  1.809  1.462      0.173  0.046 -0.079
## 3   -0.301   -0.275 -0.337 -0.271     -0.155 -0.195 -0.245
## 4    0.207    2.376  0.036  0.058      0.122 -0.045  1.942
##   online_gaming shopping health_nutrition college_uni sports_playing cooking
## 1    0.253    0.580      0.661    0.356      0.397  0.878
## 2    0.024    0.009     -0.086   -0.033      0.131 -0.077
## 3   -0.101   -0.226     -0.232   -0.136     -0.192 -0.312
## 4   -0.026   -0.049     -0.120   -0.006      0.070 -0.173
##   eco computers business outdoors crafts automotive  art religion beauty
## 1  0.418   -0.016   0.308   0.544  0.257    0.019  0.331  -0.165  0.655
## 2  0.182    0.080   0.117  -0.028  0.692    0.162  0.098   2.230  0.302
## 3 -0.214   -0.249  -0.193  -0.241 -0.237   -0.210 -0.145  -0.303 -0.290
## 4  0.102    1.549   0.319   0.181  0.113    1.122 -0.028  -0.037 -0.145
##   parenting dating school personal_fitness fashion small_business spam adult
## 1   -0.114  0.331  0.101      0.675  0.770      0.298  0.021  0.062
## 2    2.100 -0.026  1.630      -0.028  0.024      0.086  0.000  0.012
## 3   -0.312 -0.158 -0.308      -0.248 -0.290     -0.169 -0.012 -0.013
## 4    0.026  0.189 -0.038      -0.113 -0.145      0.222  0.023 -0.091
```

While we see some of the same patterns, as the previous models, the clusters appear to be getting mixed together, as the differences between them are decreasing. We will try again with 3 and 7 clusters before making any final decisions.

```
# Run k-means with 7 clusters and 25 starts
clust4 = kmeans(X, 7, nstart=25)
```

```
# What are the clusters?
round(clust4$center, 7)
```

```
##      chatter current_events      travel photo_sharing uncategorized      tv_film
## 1 -0.1334059    -0.0133547 -0.1407332    -0.1013731     0.1613325 -0.0962921
## 2  1.2071272     0.3450481 -0.1138719     0.8646392     0.2740657  0.5025940
## 3 -0.3516643    -0.1813725 -0.2333687    -0.4128490    -0.1798150 -0.1763021
## 4 -0.0602621     0.0978990  1.9083535    -0.1567518    -0.0629867  0.0220097
## 5 -0.0346151     0.1934744 -0.0350764     1.2107190     0.4788466 -0.0517009
## 6 -0.0778571    -0.0791135  0.0023157    -0.0271410     0.0241361  0.2306169
## 7 -0.1342319     0.1236560 -0.0944369    -0.0930115    -0.0767554 -0.0099703
## sports_fandom  politics      food      family home_and_garden      music
## 1   -0.1846348 -0.1708365  0.4477798 -0.0695537     0.1565714  0.0255538
## 2   -0.1762663 -0.1249693 -0.2102371 -0.0638358     0.2278357  0.3818686
## 3   -0.2998158 -0.2735302 -0.3612486 -0.2888725    -0.1972282 -0.2188896
## 4    0.1983611  2.4890856  0.0395085  0.0244744     0.0949310 -0.0515096
## 5   -0.1880640 -0.1102192 -0.1813812  0.0382228     0.1228989  0.5254621
## 6   -0.1177227 -0.1534691 -0.0766139  0.2106365     0.1164517  0.0156780
## 7    2.0749435 -0.2074492  1.8401506  1.4989562     0.1744856  0.0408622
##      news online_gaming      shopping health_nutrition college_uni
## 1 -0.0180075    -0.1169951 -0.0476893     2.1935623    -0.2080994
## 2 -0.1717833    -0.1665905  1.0992985     -0.2379759     0.0320960
## 3 -0.2406532    -0.2338971 -0.3744902     -0.3186548    -0.2600824
## 4  1.9972403    -0.1343027 -0.1159615     -0.2100165    -0.0924792
## 5 -0.0536411    -0.0338021  0.2167477     -0.0461449    -0.0287623
## 6 -0.1658378     3.4968415 -0.1213795     -0.1735220     3.2787475
## 7 -0.0682112    -0.0724422 -0.0330634     -0.1431495    -0.1218004
## sports_playing      cooking      eco      computers      business      outdoors
## 1   -0.0223334  0.4043421  0.5383762 -0.0868962  0.0417850  1.7172280
## 2    0.0041717 -0.2364094  0.2968865 -0.0625222  0.4226069 -0.1998257
## 3   -0.2658371 -0.3352165 -0.2754225 -0.2670697 -0.2601905 -0.3115912
## 4   -0.0149484 -0.2188525  0.0975311  1.6859223  0.3141832  0.1020928
## 5    0.1755654  2.7795498  0.0154230  0.0564613  0.2295884  0.0469532
## 6    2.1741601 -0.1275756 -0.0457535 -0.0521197 -0.0584846 -0.1256253
## 7    0.1088912 -0.1075248  0.1723632  0.0780751  0.0934958 -0.0717424
##      crafts automotive      art      religion      beauty      parenting      dating
## 1  0.0786748 -0.1173372 -0.0110358 -0.1588146 -0.2082830 -0.0899105  0.1615272
## 2  0.2381394  0.0953934  0.3997682 -0.2137133 -0.1430289 -0.1876371  0.3326001
## 3 -0.2891666 -0.2295050 -0.1919952 -0.2995036 -0.2794432 -0.3193942 -0.1937023
## 4  0.1034033  1.0436606 -0.0754026 -0.0101082 -0.1739894  0.0412305  0.1903679
## 5  0.1167411  0.0254276  0.1035859 -0.1164944  2.5467901 -0.0757335  0.0211458
## 6  0.0911279  0.0692916  0.3127718 -0.1610301 -0.2184321 -0.1348583  0.0079711
## 7  0.7208306  0.1566311  0.0872791  2.2644277  0.3168402  2.1511127  0.0090641
##      school personal_fitness      fashion small_business      spam      adult
## 1 -0.1641290     2.1640278 -0.1183815    -0.1361522  0.0118021  0.0214577
## 2  0.0943253    -0.2076863 -0.0591803     0.4005244  0.0254130  0.0380292
## 3 -0.3205293    -0.3370369 -0.2988347    -0.2108938 -0.0073171 -0.0070966
## 4 -0.0441723    -0.1947985 -0.1857440     0.2202647 -0.0194033 -0.0939184
## 5  0.1515470    -0.0256434  2.6497404     0.1974903 -0.0312405  0.0250458
## 6 -0.2128103    -0.1898662 -0.0585681     0.1361314  0.0517628 -0.0054003
## 7  1.6798457    -0.0909612  0.0119301     0.0897204 -0.0097580  0.0092426
```

```
# Run k-means with 3 clusters and 25 starts
clust5 = kmeans(X, 3, nstart=25)
```

```
# What are the clusters?
round(clust5$center, 3)
```

```
## chatter current_events travel photo_sharing uncategorized tv_film
## 1 -0.228 -0.133 -0.150 -0.310 -0.156 -0.155
## 2 -0.092 0.138 -0.021 -0.076 -0.083 -0.017
## 3 0.557 0.253 0.351 0.738 0.388 0.362
## sports_fandom politics food family home_and_garden music news
## 1 -0.280 -0.150 -0.325 -0.261 -0.156 -0.192 -0.129
## 2 1.983 -0.096 1.791 1.421 0.197 0.041 -0.002
## 3 -0.105 0.379 0.070 0.065 0.284 0.425 0.297
## online_gaming shopping health_nutrition college_uni sports_playing cooking
## 1 -0.110 -0.248 -0.206 -0.148 -0.187 -0.293
## 2 0.025 -0.011 -0.087 -0.033 0.121 -0.082
## 3 0.243 0.572 0.505 0.350 0.384 0.702
## eco computers business outdoors crafts automotive art religion beauty
## 1 -0.208 -0.183 -0.187 -0.209 -0.240 -0.146 -0.155 -0.303 -0.278
## 2 0.183 0.151 0.129 -0.022 0.673 0.177 0.091 2.200 0.294
## 3 0.408 0.363 0.379 0.486 0.297 0.267 0.321 -0.134 0.525
## parenting dating school personal_fitness fashion small_business spam adult
## 1 -0.302 -0.152 -0.305 -0.224 -0.280 -0.179 -0.024 -0.028
## 2 2.054 -0.023 1.608 -0.030 0.019 0.087 -0.003 0.001
## 3 -0.081 0.358 0.095 0.525 0.634 0.378 0.056 0.064
```

7 clusters yields similar clusters to what we have already seen, while 3 clusters seems to blur the clusters even further. So far 5, 6 and 7 clusters seems to reveal the best models. We see the same 5 clusters in the 6 and 7 cluster models. However, there is an additional cluster added with the 7 cluster model. We will run the model with 8 clusters to verify if this cluster is legitimate.

```
# Run k-means with 8 clusters and 25 starts
clust6 = kmeans(X, 8, nstart=25)
```

```
# What are the clusters?
round(clust6$center, 8)
```

```
## chatter current_events travel photo_sharing uncategorized
## 1 -0.13209289 -0.01725352 -0.14015774 -0.09786320 0.15847221
## 2 -0.08989052 -0.08280204 -0.01460203 -0.03125173 0.02851118
## 3 0.07205880 0.27684711 0.28872703 -0.09071828 0.11259854
## 4 -0.06017664 0.08978878 1.89626032 -0.15659240 -0.05994898
## 5 -0.35122913 -0.18081595 -0.23335202 -0.41392650 -0.18036005
## 6 -0.13515892 0.12335634 -0.09600738 -0.09104774 -0.07832912
## 7 -0.03455631 0.18899938 -0.03847151 1.21287095 0.47893929
## 8 1.20941359 0.34294131 -0.11677538 0.87145494 0.27218463
## tv_film sports_fandom politics food family home_and_garden
## 1 -0.09731398 -0.1905392 -0.1740972 0.44480568 -0.06829666 0.15410419
## 2 0.23120417 -0.1176990 -0.1687362 -0.08153528 0.20657508 0.11578559
## 3 -0.11619101 0.1406567 0.1505274 0.04049422 -0.05999555 0.23510191
## 4 0.02406822 0.2058784 2.4817780 0.03993536 0.02718130 0.09843043
```

## 5	-0.17594444	-0.3010094	-0.2746318	-0.36178105	-0.28842434	-0.19941681
## 6	-0.00807081	2.0679074	-0.2080877	1.84228092	1.50303323	0.17563815
## 7	-0.04824820	-0.1873870	-0.1096930	-0.18544140	0.03930472	0.12672209
## 8	0.50446606	-0.1761408	-0.1232787	-0.20834523	-0.06682025	0.22310478
##	music	news	online_gaming	shopping	health_nutrition	
## 1	0.01685412	-0.01820310	-0.11672680	-0.04340157	2.19376201	
## 2	0.01062914	-0.17027270	3.49868120	-0.13413473	-0.18176505	
## 3	0.01418264	-0.00069020	0.08935906	-0.23782264	0.05086059	
## 4	-0.05209774	2.00449341	-0.13560059	-0.11580059	-0.20618720	
## 5	-0.21834878	-0.24313963	-0.23496655	-0.37376616	-0.31853237	
## 6	0.04208496	-0.06668843	-0.07135369	-0.03045397	-0.14514445	
## 7	0.52902379	-0.05051285	-0.03128372	0.22011614	-0.05419201	
## 8	0.38630139	-0.17392306	-0.16677439	1.10763731	-0.23736825	
##	college_uni	sports_playing	cooking	eco	computers	business
## 1	-0.21034603	-0.02173242	0.40999850	0.52654720	-0.08458197	0.04504632
## 2	3.27231832	2.17772652	-0.12782763	-0.04776128	-0.06879379	-0.05397805
## 3	0.12732753	-0.11129036	-0.05898219	0.44799948	0.29753290	-0.34600901
## 4	-0.09455937	-0.01224925	-0.21731789	0.08823982	1.66893445	0.31568319
## 5	-0.26042051	-0.26458294	-0.33565323	-0.27848458	-0.26793882	-0.25923820
## 6	-0.11994126	0.10886442	-0.10648067	0.17713099	0.08044460	0.09339702
## 7	-0.02683493	0.17005379	2.77791735	0.01313539	0.05458161	0.23614496
## 8	0.03292957	0.00584717	-0.23570670	0.29908500	-0.06430564	0.42701298
##	outdoors	crafts	automotive	art	religion	beauty
## 1	1.71974505	0.07763626	-0.11854503	-0.01040201	-0.16196682	-0.2047910
## 2	-0.12811424	0.08703901	0.06884124	0.29893703	-0.16619409	-0.2220229
## 3	0.29780310	0.21793373	0.12453564	0.33167537	0.12070179	-0.1007019
## 4	0.10330706	0.09864196	1.05345477	-0.07579853	-0.01089024	-0.1734107
## 5	-0.31369757	-0.28945796	-0.23169223	-0.19370768	-0.29913956	-0.2785233
## 6	-0.07199692	0.71633541	0.15473126	0.08681482	2.25902379	0.3173796
## 7	0.02818953	0.11780271	0.02750858	0.10307683	-0.11599618	2.5561090
## 8	-0.20115976	0.23699637	0.09110647	0.39741938	-0.21389919	-0.1455980
##	parenting	dating	school	personal_fitness	fashion	
## 1	-0.09309025	0.16152376	-0.16907678	2.15970326	-0.11078845	
## 2	-0.14176053	0.00976279	-0.22299946	-0.19097773	-0.06054701	
## 3	0.18658414	-0.00952824	0.09244824	0.12183236	-0.02044987	
## 4	0.04228275	0.18950417	-0.04319682	-0.19615054	-0.18693445	
## 5	-0.32039093	-0.19322282	-0.31946340	-0.33641448	-0.29900303	
## 6	2.14991505	0.01058691	1.67272910	-0.09156854	0.01276527	
## 7	-0.07544236	0.02332680	0.15448524	-0.03569165	2.64693174	
## 8	-0.18919225	0.33009982	0.09609958	-0.20782120	-0.05884783	
##	small_business	spam	adult			
## 1	-0.13007568	-0.07768727	0.01303733			
## 2	0.14607287	-0.07768727	-0.01711754			
## 3	0.31428826	12.41886450	3.75022215			
## 4	0.21625971	-0.07768727	-0.10711639			
## 5	-0.21246952	-0.07768727	-0.02636218			
## 6	0.08872439	-0.07768727	-0.01375608			
## 7	0.18750523	-0.07768727	-0.00312293			
## 8	0.39342046	-0.07768727	-0.01312237			

The 8 cluster model confirms the validity of this cluster, and also finds a new cluster, however this cluster consists of spam and adult content, and therefore it is not an audience that Nutrient H2O needs to be concerned with. Before, we confirm these clusters as the final segments of the market, we will finally run the model 10 and 50 starts, along with the 8 clusters.

```
# Run k-means with 8 clusters and 10 starts
clust7 = kmeans(X, 8, nstart=10)
```

```
# What are the clusters?
round(clust7$center, 8)
```

```
##          chatter current_events      travel photo_sharing uncategorized
## 1 -0.03871235    0.18971529 -0.03931613    1.21072449    0.47785535
## 2 -0.06017664    0.08978878  1.89626032   -0.15659240   -0.05994898
## 3 -0.13343134   -0.02069711 -0.14167823   -0.10482470    0.15862651
## 4 -0.08989052   -0.08280204 -0.01460203   -0.03125173    0.02851118
## 5  1.20941359    0.34294131 -0.11677538    0.87145494    0.27218463
## 6  0.07205880    0.27684711  0.28872703   -0.09071828    0.11259854
## 7 -0.13515892    0.12335634 -0.09600738   -0.09104774   -0.07832912
## 8 -0.35068759   -0.18048704 -0.23307415   -0.41322326   -0.18089462
##          tv_film sports_fandom    politics      food      family home_and_garden
## 1 -0.05056656   -0.1886254 -0.1109180 -0.18012263  0.03961888    0.12612095
## 2  0.02406822    0.2058784  2.4817780  0.03993536  0.02718130    0.09843043
## 3 -0.09935842   -0.1902784 -0.1735978  0.44301113 -0.06978972    0.14920359
## 4  0.23120417   -0.1176990 -0.1687362 -0.08153528  0.20657508    0.11578559
## 5  0.50446606   -0.1761408 -0.1232787 -0.20834523 -0.06682025    0.22310478
## 6 -0.11619101    0.1406567  0.1505274  0.04049422 -0.05999555    0.23510191
## 7 -0.00807081    2.0679074 -0.2080877  1.84228092  1.50303323    0.17563815
## 8 -0.17528460   -0.3010461 -0.2747449 -0.36292525 -0.28850229   -0.19870341
##          music      news online_gaming    shopping health_nutrition
## 1  0.52629343 -0.05162092  -0.03218652  0.21842537   -0.04928907
## 2 -0.05209774  2.00449341  -0.13560059 -0.11580059   -0.20618720
## 3  0.01553804 -0.01790818  -0.11843243 -0.04541805    2.19285130
## 4  0.01062914 -0.17027270   3.49868120 -0.13413473   -0.18176505
## 5  0.38630139 -0.17392306  -0.16677439  1.10763731   -0.23736825
## 6  0.01418264 -0.00069020   0.08935906 -0.23782264    0.05086059
## 7  0.04208496 -0.06668843  -0.07135369 -0.03045397   -0.14514445
## 8 -0.21827457 -0.24334394  -0.23466447 -0.37367811   -0.32130469
## college_uni sports_playing    cooking      eco    computers    business
## 1 -0.02746727    0.17082981  2.77449085  0.01049963  0.05552511  0.23285396
## 2 -0.09455937   -0.01224925 -0.21731789  0.08823982  1.66893445  0.31568319
## 3 -0.21241540   -0.02664617  0.40303122  0.52371488 -0.08737433  0.04259732
## 4  3.27231832    2.17772652 -0.12782763 -0.04776128 -0.06879379 -0.05397805
## 5  0.03292957    0.00584717 -0.23570670  0.29908500 -0.06430564  0.42701298
## 6  0.12732753   -0.11129036 -0.05898219  0.44799948  0.29753290 -0.34600901
## 7 -0.11994126    0.10886442 -0.10648067  0.17713099  0.08044460  0.09339702
## 8 -0.26003467   -0.26403368 -0.33596193 -0.27830565 -0.26778021 -0.25874319
##      outdoors    crafts    automotive      art    religion    beauty
## 1  0.04483224  0.11726982  0.02504273  0.10333233 -0.11573921  2.5528937
## 2  0.10330706  0.09864196  1.05345477 -0.07579853 -0.01089024 -0.1734107
## 3  1.70474336  0.07346748 -0.12036930 -0.01125960 -0.16479614 -0.2106915
## 4 -0.12811424  0.08703901  0.06884124  0.29893703 -0.16619409 -0.2220229
## 5 -0.20115976  0.23699637  0.09110647  0.39741938 -0.21389919 -0.1455980
## 6  0.29780310  0.21793373  0.12453564  0.33167537  0.12070179 -0.1007019
## 7 -0.07199692  0.71633541  0.15473126  0.08681482  2.25902379  0.3173796
## 8 -0.31462382 -0.28897699 -0.23116389 -0.19387235 -0.29875580 -0.2783801
##      parenting    dating    school personal_fitness    fashion
## 1 -0.07494779  0.02168735  0.15301040   -0.03145531  2.64727953
```

```
## 2 0.04228275 0.18950417 -0.04319682 -0.19615054 -0.18693445
## 3 -0.09583234 0.16013070 -0.17190333 2.15661608 -0.11989160
## 4 -0.14176053 0.00976279 -0.22299946 -0.19097773 -0.06054701
## 5 -0.18919225 0.33009982 0.09609958 -0.20782120 -0.05884783
## 6 0.18658414 -0.00952824 0.09244824 0.12183236 -0.02044987
## 7 2.14991505 0.01058691 1.67272910 -0.09156854 0.01276527
## 8 -0.32017211 -0.19309149 -0.31900417 -0.33860499 -0.29881087
## small_business spam adult
## 1 0.19094651 -0.07768727 -0.00397453
## 2 0.21625971 -0.07768727 -0.10711639
## 3 -0.13564479 -0.07768727 0.01215880
## 4 0.14607287 -0.07768727 -0.01711754
## 5 0.39342046 -0.07768727 -0.01312237
## 6 0.31428826 12.41886450 3.75022215
## 7 0.08872439 -0.07768727 -0.01375608
## 8 -0.21200322 -0.07768727 -0.02608652
```

```
# Run k-means with 8 clusters and 50 starts
clust8 = kmeans(X, 8, nstart=50)
```

```
# What are the clusters?
round(clust8$center, 8)
```

```
## chatter current_events travel photo_sharing uncategorized
## 1 -0.08989052 -0.08280204 -0.01460203 -0.03125173 0.02851118
## 2 -0.13515892 0.12335634 -0.09600738 -0.09104774 -0.07832912
## 3 1.20941359 0.34294131 -0.11677538 0.87145494 0.27218463
## 4 -0.13209289 -0.01725352 -0.14015774 -0.09786320 0.15847221
## 5 -0.35122913 -0.18081595 -0.23335202 -0.41392650 -0.18036005
## 6 0.07205880 0.27684711 0.28872703 -0.09071828 0.11259854
## 7 -0.06017664 0.08978878 1.89626032 -0.15659240 -0.05994898
## 8 -0.03455631 0.18899938 -0.03847151 1.21287095 0.47893929
## tv_film sports_fandom politics food family home_and_garden
## 1 0.23120417 -0.1176990 -0.1687362 -0.08153528 0.20657508 0.11578559
## 2 -0.00807081 2.0679074 -0.2080877 1.84228092 1.50303323 0.17563815
## 3 0.50446606 -0.1761408 -0.1232787 -0.20834523 -0.06682025 0.22310478
## 4 -0.09731398 -0.1905392 -0.1740972 0.44480568 -0.06829666 0.15410419
## 5 -0.17594444 -0.3010094 -0.2746318 -0.36178105 -0.28842434 -0.19941681
## 6 -0.11619101 0.1406567 0.1505274 0.04049422 -0.05999555 0.23510191
## 7 0.02406822 0.2058784 2.4817780 0.03993536 0.02718130 0.09843043
## 8 -0.04824820 -0.1873870 -0.1096930 -0.18544140 0.03930472 0.12672209
## music news online_gaming shopping health_nutrition
## 1 0.01062914 -0.17027270 3.49868120 -0.13413473 -0.18176505
## 2 0.04208496 -0.06668843 -0.07135369 -0.03045397 -0.14514445
## 3 0.38630139 -0.17392306 -0.16677439 1.10763731 -0.23736825
## 4 0.01685412 -0.01820310 -0.11672680 -0.04340157 2.19376201
## 5 -0.21834878 -0.24313963 -0.23496655 -0.37376616 -0.31853237
## 6 0.01418264 -0.00069020 0.08935906 -0.23782264 0.05086059
## 7 -0.05209774 2.00449341 -0.13560059 -0.11580059 -0.20618720
## 8 0.52902379 -0.05051285 -0.03128372 0.22011614 -0.05419201
## college_uni sports_playing cooking eco computers business
## 1 3.27231832 2.17772652 -0.12782763 -0.04776128 -0.06879379 -0.05397805
## 2 -0.11994126 0.10886442 -0.10648067 0.17713099 0.08044460 0.09339702
```



```

## 3  0.03292957    0.00584717 -0.23570670  0.29908500 -0.06430564  0.42701298
## 4 -0.21034603   -0.02173242  0.40999850  0.52654720 -0.08458197  0.04504632
## 5 -0.26042051   -0.26458294 -0.33565323 -0.27848458 -0.26793882 -0.25923820
## 6  0.12732753   -0.11129036 -0.05898219  0.44799948  0.29753290 -0.34600901
## 7 -0.09455937   -0.01224925 -0.21731789  0.08823982  1.66893445  0.31568319
## 8 -0.02683493    0.17005379  2.77791735  0.01313539  0.05458161  0.23614496
##      outdoors      crafts  automotive      art      religion      beauty
## 1 -0.12811424  0.08703901  0.06884124  0.29893703 -0.16619409 -0.2220229
## 2 -0.07199692  0.71633541  0.15473126  0.08681482  2.25902379  0.3173796
## 3 -0.20115976  0.23699637  0.09110647  0.39741938 -0.21389919 -0.1455980
## 4  1.71974505  0.07763626 -0.11854503 -0.01040201 -0.16196682 -0.2047910
## 5 -0.31369757 -0.28945796 -0.23169223 -0.19370768 -0.29913956 -0.2785233
## 6  0.29780310  0.21793373  0.12453564  0.33167537  0.12070179 -0.1007019
## 7  0.10330706  0.09864196  1.05345477 -0.07579853 -0.01089024 -0.1734107
## 8  0.02818953  0.11780271  0.02750858  0.10307683 -0.11599618  2.5561090
##      parenting      dating      school personal_fitness      fashion
## 1 -0.14176053  0.00976279 -0.22299946      -0.19097773 -0.06054701
## 2  2.14991505  0.01058691  1.67272910      -0.09156854  0.01276527
## 3 -0.18919225  0.33009982  0.09609958      -0.20782120 -0.05884783
## 4 -0.09309025  0.16152376 -0.16907678      2.15970326 -0.11078845
## 5 -0.32039093 -0.19322282 -0.31946340      -0.33641448 -0.29900303
## 6  0.18658414 -0.00952824  0.09244824      0.12183236 -0.02044987
## 7  0.04228275  0.18950417 -0.04319682      -0.19615054 -0.18693445
## 8 -0.07544236  0.02332680  0.15448524      -0.03569165  2.64693174
##      small_business      spam      adult
## 1  0.14607287 -0.07768727 -0.01711754
## 2  0.08872439 -0.07768727 -0.01375608
## 3  0.39342046 -0.07768727 -0.01312237
## 4 -0.13007568 -0.07768727  0.01303733
## 5 -0.21246952 -0.07768727 -0.02636218
## 6  0.31428826 12.41886450  3.75022215
## 7  0.21625971 -0.07768727 -0.10711639
## 8  0.18750523 -0.07768727 -0.00312293

```

Both the model with 10 starts and the one with 50 starts support the existence of the clusters we saw earlier. These clusters seem to represent segments of the market, with each segment having a group of correlated interests. One segment is interested in health, the outdoors and personal fitness. Another seems to like chatter and shopping. There is one segment that enjoys travel, politics, news, computers and cars. Another artistic segment of Twitter focuses on photo sharing, cooking, beauty and fashion. A fifth segment likes online gaming, sports playing and college. The final cluster is interested in sports fans, food, family, religion, parenting and school. Many of these clusters make logical sense, as many of the interests in each segment are somewhat linked. Nutrient H2O can use the information to make advertisements towards people who share an interest that would be correlated to their product. For example, if they have a product that is healthy and high in nutrition value, they could market ads for that product towards people who post about content related to the outdoors or their personal fitness.