

# Market Segmentation

## Market Segmentation

The goal of this report is to provide NutrientH2O interesting market segments identified from an analysis of the tweets by Twitter users who follow the NutrientH2O account. Our team decided to conduct a market segmentation using clustering. NutrientH2O's twitter followers were clustered together into 7 groups based on the words they used in their tweets. Our team provides additional information on how NutrientH2O could categorize these clusters to better understand their social-media audience. The report is broken down into an outline of steps taken and the conclusions derived from the analysis.

### Step 1: Cleaning the data

In this step, we read the file information into a dataframe and removed the spam and adult columns as these are likely from bots and do not represent actual twitter users. We ran the cluster analysis once including the uncategorized variable and then ran it again without the variable. We found that the clustering was more effective without the variable so we left it out of our final analysis.

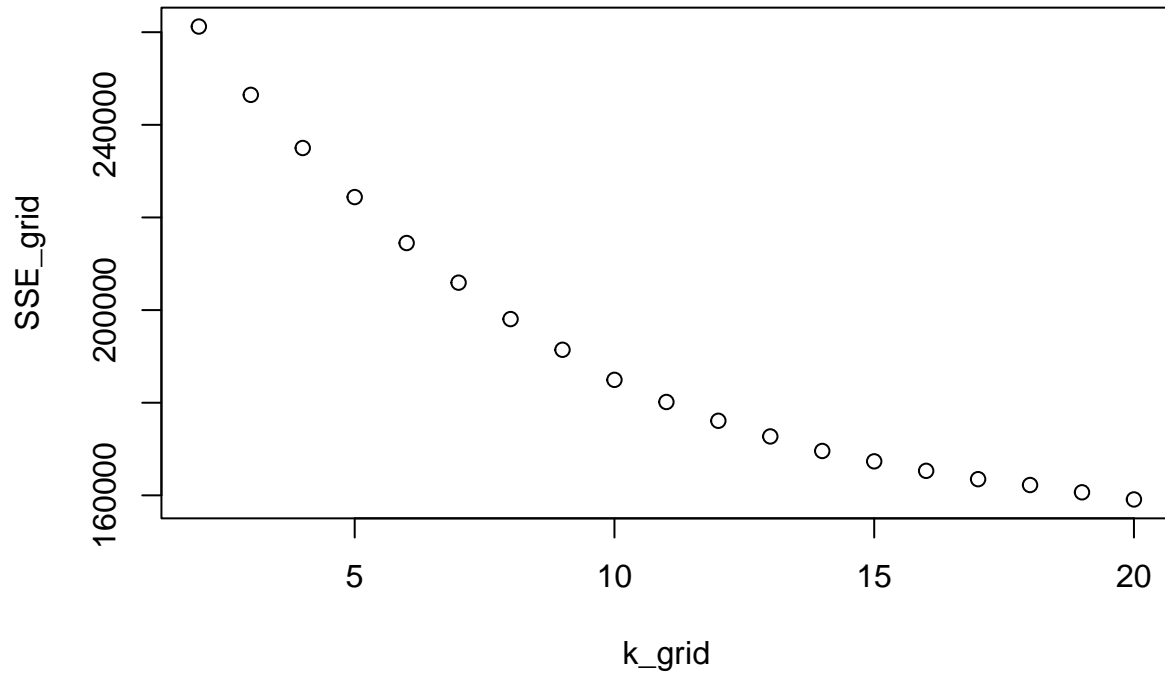
```
##      X          chatter      current_events      travel
## Length:7882      Min.   : 0.000      Min.   :0.000      Min.   : 0.000
## Class :character  1st Qu.: 2.000      1st Qu.:1.000      1st Qu.: 0.000
## Mode  :character  Median : 3.000      Median :1.000      Median : 1.000
##                               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			
##	current_events	travel	photo_sharing	tv_film
##	Min. :0.000	Min. : 0.000	Min. : 0.000	Min. : 0.00
##	1st Qu.:1.000	1st Qu.: 0.000	1st Qu.: 1.000	1st Qu.: 0.00
##	Median :1.000	Median : 1.000	Median : 2.000	Median : 1.00
##	Mean :1.526	Mean : 1.585	Mean : 2.697	Mean : 1.07
##	3rd Qu.:2.000	3rd Qu.: 2.000	3rd Qu.: 4.000	3rd Qu.: 1.00
##	Max. :8.000	Max. :26.000	Max. :21.000	Max. :17.00

## sports_fandom	politics	food	family
## Min. : 0.000	Min. : 0.000	Min. : 0.000	Min. : 0.0000
## 1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.0000
## Median : 1.000	Median : 1.000	Median : 1.000	Median : 1.0000
## Mean : 1.594	Mean : 1.789	Mean : 1.397	Mean : 0.8639
## 3rd Qu.: 2.000	3rd Qu.: 2.000	3rd Qu.: 2.000	3rd Qu.: 1.0000
## Max. :20.000	Max. :37.000	Max. :16.000	Max. :10.0000
## home_and_garden	music	news	online_gaming
## Min. :0.0000	Min. : 0.0000	Min. : 0.000	Min. : 0.000
## 1st Qu.:0.0000	1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 0.000
## Median :0.0000	Median : 0.0000	Median : 0.000	Median : 0.000
## Mean :0.5207	Mean : 0.6793	Mean : 1.206	Mean : 1.209
## 3rd Qu.:1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.000	3rd Qu.: 1.000
## Max. :5.0000	Max. :13.0000	Max. :20.000	Max. :27.000
## shopping	health_nutrition	college_uni	sports_playing
## Min. : 0.000	Min. : 0.000	Min. : 0.000	Min. :0.0000
## 1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.:0.0000
## Median : 1.000	Median : 1.000	Median : 1.000	Median :0.0000
## Mean : 1.389	Mean : 2.567	Mean : 1.549	Mean :0.6392
## 3rd Qu.: 2.000	3rd Qu.: 3.000	3rd Qu.: 2.000	3rd Qu.:1.0000
## Max. :12.000	Max. :41.000	Max. :30.000	Max. :8.0000
## cooking	eco	computers	business
## Min. : 0.000	Min. :0.0000	Min. : 0.0000	Min. :0.0000
## 1st Qu.: 0.000	1st Qu.:0.0000	1st Qu.: 0.0000	1st Qu.:0.0000
## Median : 1.000	Median :0.0000	Median : 0.0000	Median :0.0000
## Mean : 1.998	Mean :0.5123	Mean : 0.6491	Mean :0.4232
## 3rd Qu.: 2.000	3rd Qu.:1.0000	3rd Qu.: 1.0000	3rd Qu.:1.0000
## Max. :33.000	Max. :6.0000	Max. :16.0000	Max. :6.0000
## outdoors	crafts	automotive	art
## 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 : 0.7827	Mean :0.5159	Mean : 0.8299	Mean : 0.7248
## 3rd Qu.: 1.0000	3rd Qu.:1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.0000
## Max. :12.0000	Max. :7.0000	Max. :13.0000	Max. :18.0000
## religion	beauty	parenting	dating
## Min. : 0.000	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000
## 1st Qu.: 0.000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000
## Median : 0.000	Median : 0.0000	Median : 0.0000	Median : 0.0000
## Mean : 1.095	Mean : 0.7052	Mean : 0.9213	Mean : 0.7109
## 3rd Qu.: 1.000	3rd Qu.: 1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.0000
## Max. :20.000	Max. :14.0000	Max. :14.0000	Max. :24.0000
## school	personal_fitness	fashion	small_business
## Min. : 0.0000	Min. : 0.000	Min. : 0.0000	Min. :0.0000
## 1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 0.0000	1st Qu.:0.0000
## Median : 0.0000	Median : 0.000	Median : 0.0000	Median :0.0000
## Mean : 0.7677	Mean : 1.462	Mean : 0.9966	Mean :0.3363
## 3rd Qu.: 1.0000	3rd Qu.: 2.000	3rd Qu.: 1.0000	3rd Qu.:1.0000
## Max. :11.0000	Max. :19.000	Max. :18.0000	Max. :6.0000

**Step 2:** We centered and scaled the data and then extracted the centers and scales from the rescaled data.

**Step 3:** Created an Elbow Plot to identify the best number of clusters. The elbow plot showed that  $k=7$  is the best number of clusters to use for the dataset.



**## Step 4:** We decided to graph the top 10 factors in each cluster based on the size of the centroid computed for each factor in that specific cluster. This created 7 graphs that show the most important factors used in the cluster

Cluster 1 “Outdoors Health Enthusiasts”: This cluster showed the highest centroid values for health\_nutrition, personal\_fitness,cooking, and outdoors. All of these features group together to paint the picture of this segment consisting of followers that prioritize healthy eating and living and also enjoy spending time outdoors. These followers can be targeted by NutrientH20 with messaging focused on healthy living for outdoor enthusiasts.

Cluster 2 “Millennial Women”: This cluster had high feature values for cooking, photo\_sharing, fashion, and beauty which leads us to believe this cluster is likely made up of women. We can segment this group further by looking at the next highest value which is photo-sharing. This narrows the age range of the women in this segment by focusing on users that tweet about photo-sharing on social media. These followers are likely to be younger millennials.

Cluster 3 “Politics Enthusiasts”: This cluster showed the highest centroid values for politics, news, travel, and photo\_sharing. These followers enjoy staying up to date on all things political and likely intake a large amount of daily news. These followers also have an interest in travel and technology. With this information, NutrientH20 can target these followers by marketing their message on news and travel related apps and websites.

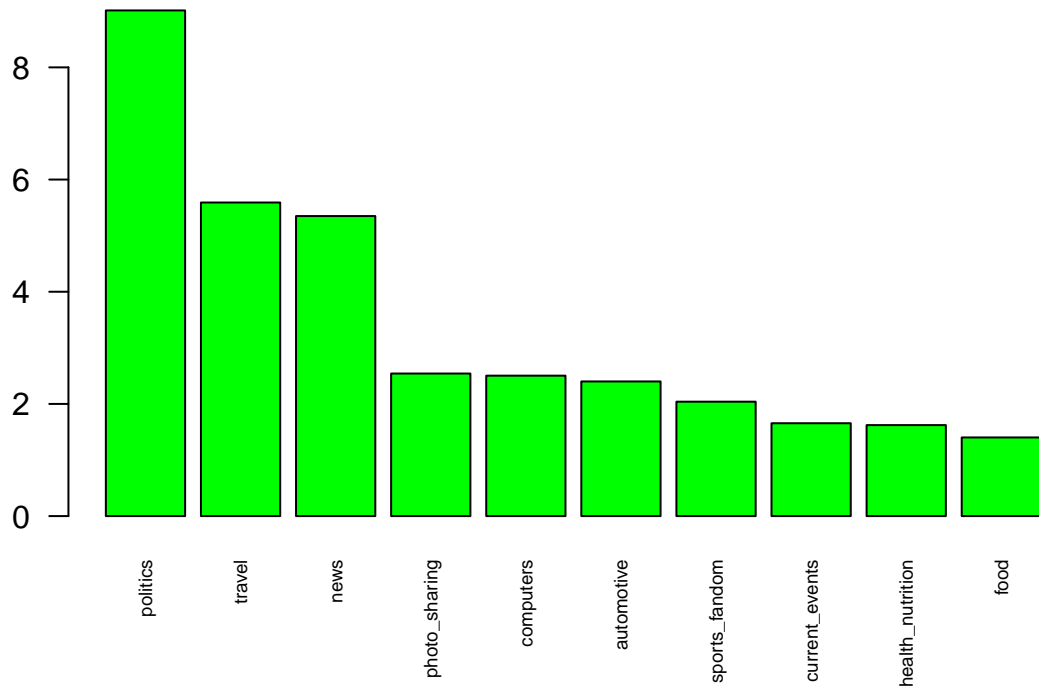
Cluster 4 “Families”: This cluster consists of users that tweeted frequently about sports\_fandom,religion,food, parenting, school,photo\_sharing,and family. All of these features can be grouped together to segment these followers into a group that is likely parents of school-aged children.

Cluster 5 “Fine Arts Enthusiasts”: This cluster showed the highest centroid values for tweets associated with tv\_film and art. These followers appreciate the fine arts and NutrientH20 can target these followers with fine arts focused messaging.

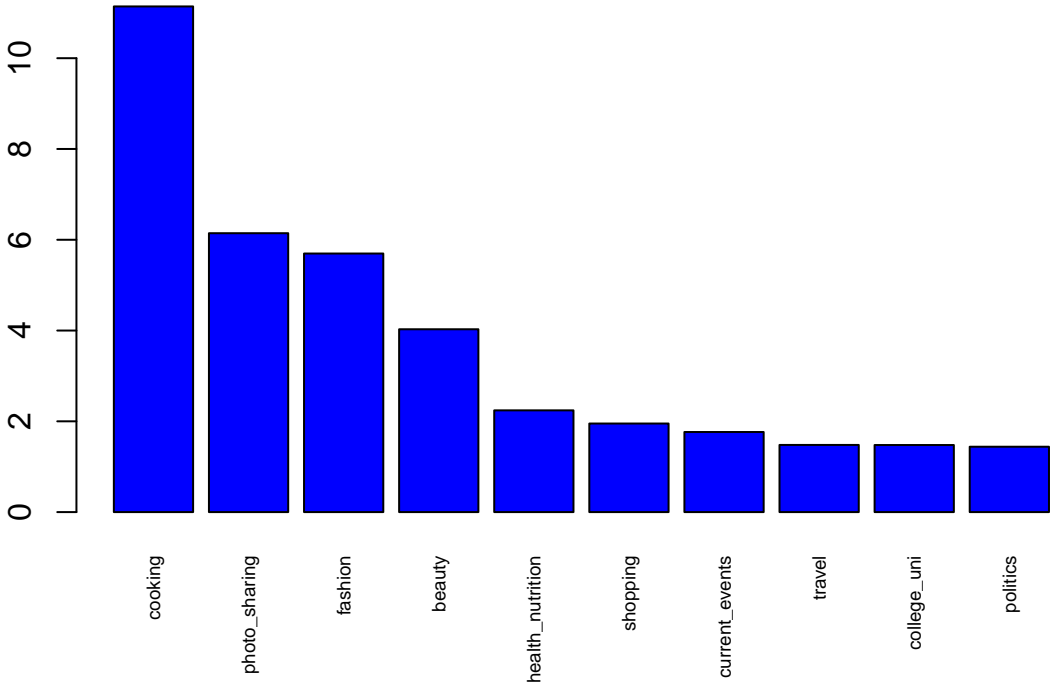
Cluster 6 “Women that Enjoy Shopping and the News”: This cluster showed the highest centroid values for photo\_sharing, shopping, and current\_events. The followers in this cluster are similar to Cluster 2 but they have a stronger affinity for current events and that makes them easier to message to, through outlets where they consume news and potentially shop. An example would be on major social media platforms through sponsored advertisements.

Cluster 7 “College Students that enjoy sports video games”: This cluster had high centroid values for online-gaming,college-uni, photo\_sharing, and sports\_playing indicating that these factors were strong predictors for a cluster. The plot below shows that these users can be clustered into a market segment representing college students that enjoy online-gaming entertainment. NutrientH20 clearly has a segment of Twitter followers that can be clustered based on these attributes found in their tweets.

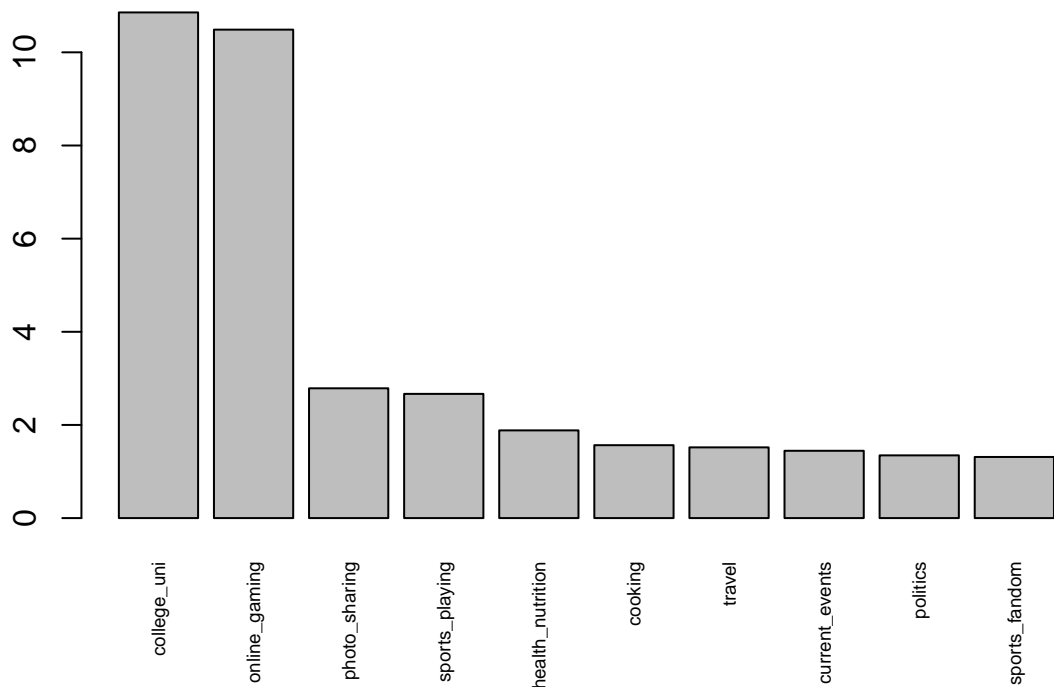
## Cluster 1



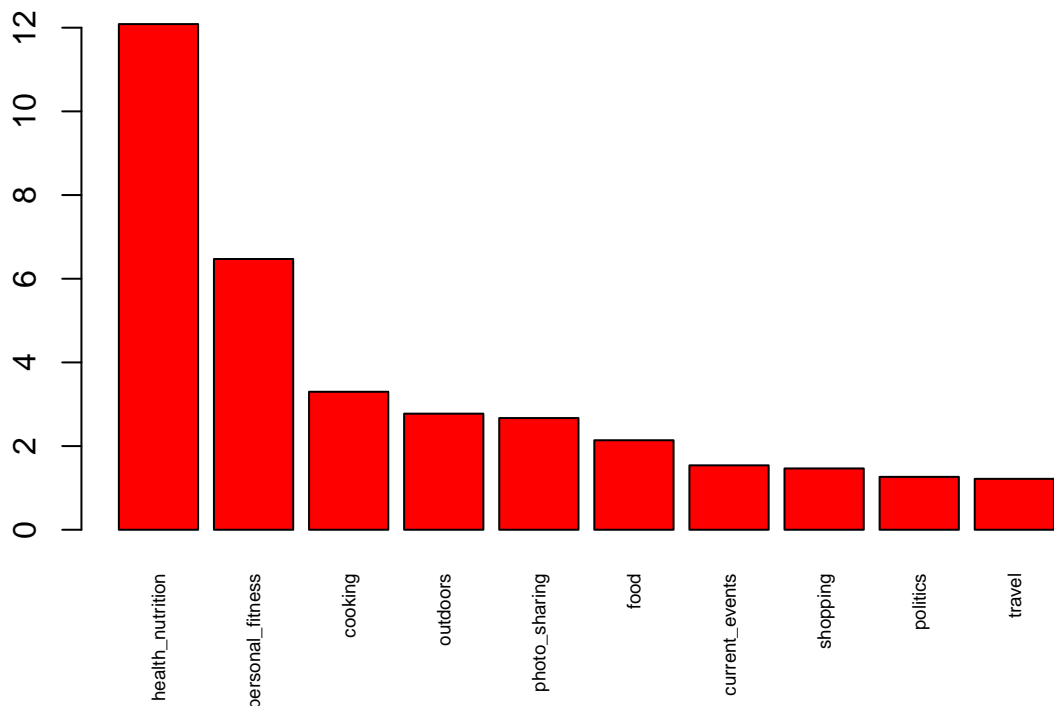
Cluster 2



Cluster 3

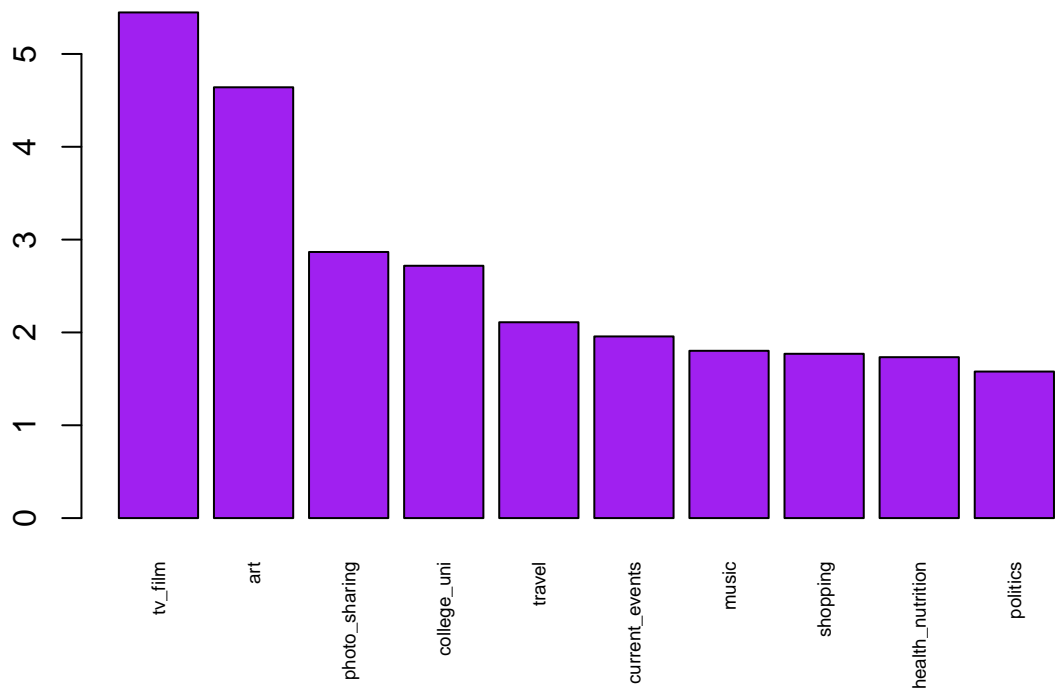


Cluster 4

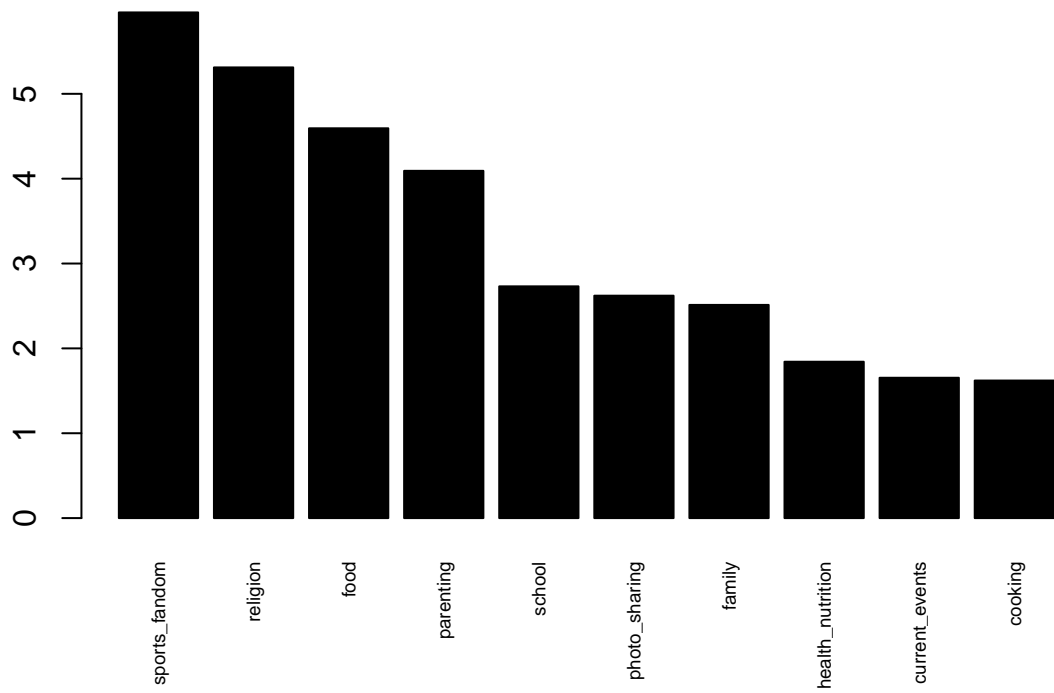




Cluster 5



Cluster 6



## Cluster 7

