

Market segmentation

Kristen Lowe and Siboney Cardoso

2024-08-13

```
# import R libraries
library(tidyverse)
library(cluster)
library(foreach)
library(ClusterR)
```

```
## Warning: package 'ClusterR' was built under R version 4.3.3
```

```
library(mosaic)
library(reshape2)
```

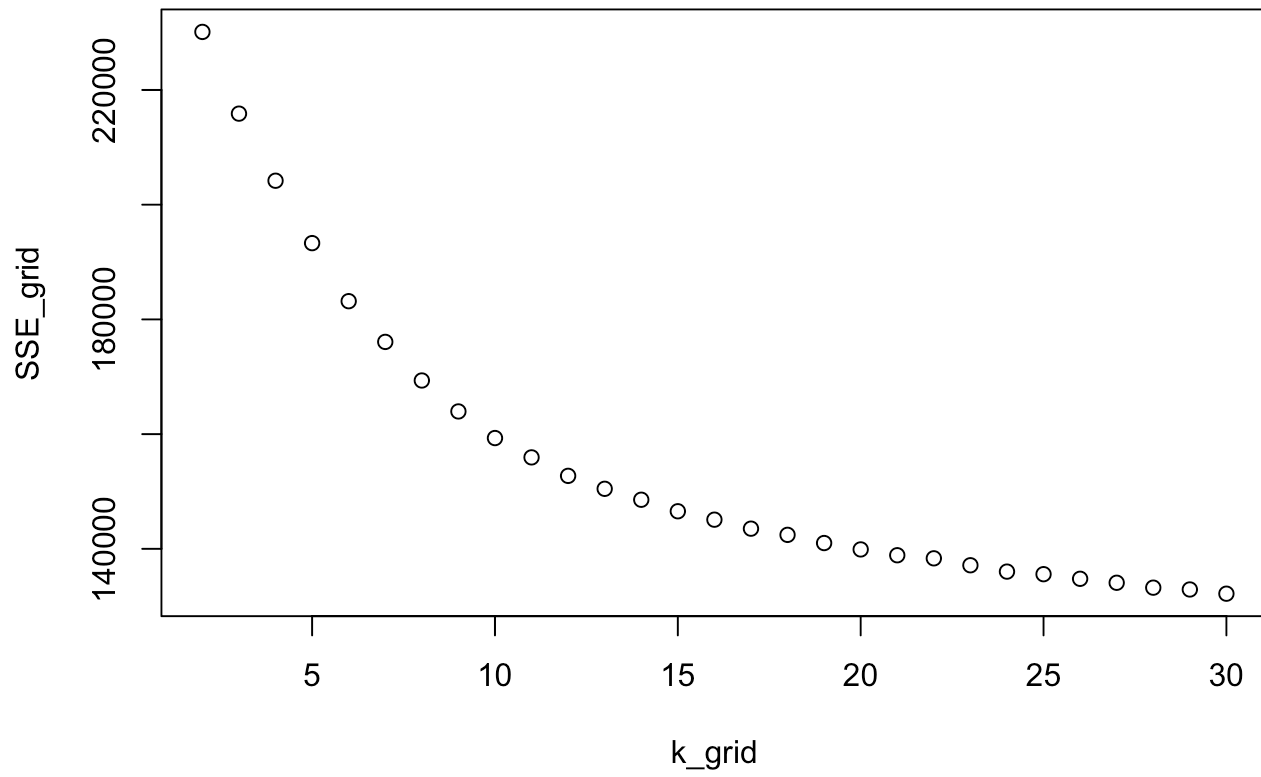
```
social_marketing <- read_csv('data/social_marketing.csv', show_col_types = FALSE) %>%
  dplyr::select(-chatter, -uncategorized, -adult, -spam)
```

```
## New names:
## • `` -> `...1`
```

```
# find optimal number of clusters
X = scale(social_marketing[,2:33]) # cluster on measurables
k_grid = seq(2, 30, by=1)
SSE_grid = foreach(k = k_grid, .combine='c') %do% {
  cluster_k = kmeans(X, k, nstart=50)
  cluster_k$tot.withinss
}
```

```
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
## Warning: did not converge in 10 iterations
```

```
plot(k_grid, SSE_grid)
```



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

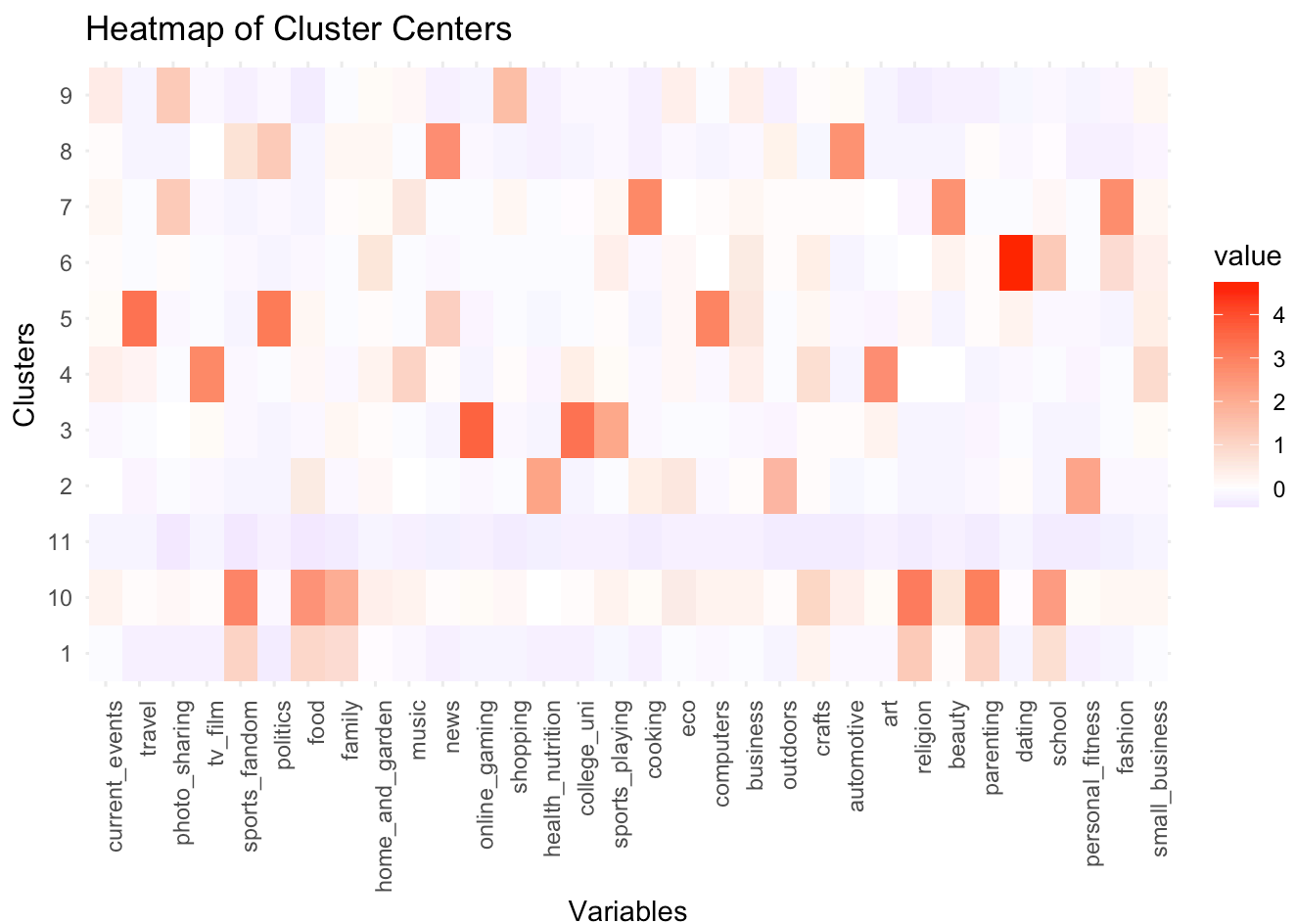
# Run k-means with 11 clusters and 25 starts
clust1 = kmeans(X, 11, nstart=25)

# Convert the cluster centers into a data frame
centers_df <- as.data.frame(clust1$centers)

# Add a column for cluster numbers
centers_df$cluster <- rownames(centers_df)

# Melt the data frame, keeping the Cluster column as an id variable
centers_melted <- melt(centers_df, id.vars = "cluster", variable.name = "variable", value.name = "value")

# Plot the heatmap
ggplot(centers_melted, aes(x = variable, y = cluster, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint = 0) +
  labs(title = "Heatmap of Cluster Centers", x = "Variables", y = "Clusters") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



```

column_sums <- colSums(social_marketing[,2:33], na.rm = TRUE)

# Convert the sums into a data frame for easier merging
sums_df <- data.frame(
  variable = names(column_sums),
  prevalence = column_sums
)

centers_melted %>%
  group_by(variable) %>%
  slice_max(order_by = value, n = 1) %>%
  select(variable, cluster, max_value = value) %>%
  mutate(cluster = as.numeric(cluster)) %>%
  arrange(cluster) %>%
  left_join(sums_df, by = "variable")

```

```

## # A tibble: 32 × 4
## # Groups:   variable [32]
##   variable      cluster max_value prevalence
##   <chr>         <dbl>     <dbl>     <dbl>
## 1 health_nutrition      2      2.23      20235
## 2 eco                  2      0.574      4038
## 3 outdoors             2      1.73      6169
## 4 personal_fitness     2      2.17     11524
## 5 online_gaming        3      3.58      9528
## 6 college_uni          3      3.29     12213
## 7 sports_playing       3      2.13      5038
## 8 tv_film              4      2.79      8436
## 9 music                4      1.04      5354
## 10 art                 4      2.67      5713
## # i 22 more rows

```

For the purpose of this analysis, I will define “market segment” as a cluster of interests. Based on my analysis, there are 10 possible clusters that NutrientH2O could target:

Cluster 1: Health & Wellness Enthusiasts

- Variables in Cluster: Health nutrition, eco-friendly lifestyles, outdoor activities, personal fitness.
- Description of Cluster: This cluster focuses on maintaining a healthy lifestyle, emphasizing nutrition, fitness, and outdoor activities. They are also environmentally conscious, showing interest in eco-friendly practices.

Cluster 2: News & Automotive Followers

- Variables in Cluster: News, automotive topics.
- Description of Cluster: This cluster is focused on staying informed on current events and strongly interested in automotive trends and technologies.

Cluster 3: Creative & Artistic Pursuits

- Variables in Cluster: TV and film, crafts, art.
- Description of Cluster: This cluster is made up of individuals who are deeply engaged in creative and artistic activities. They enjoy crafts, appreciate various art forms, and have a keen interest in TV and film.

Cluster 4: Lifestyle & Aesthetic Enthusiasts

- Variables in Cluster: Photo sharing, cooking, beauty, fashion.
- Description of Cluster: This group is interested in lifestyle and aesthetics, focusing on sharing moments visually, exploring culinary arts, and following beauty and fashion trends.

Cluster 5: Entrepreneurial & Music Lovers

- Variables in Cluster: Music, small business.
- Description of Cluster: This cluster combines entrepreneurial individuals and music enthusiasts. They will likely be involved in small business ventures and have a strong passion for music.

Cluster 6: Young & Active

- Variables in Cluster: Online gaming, college/university life, sports playing.
- Description of Cluster: This cluster represents a younger, active demographic that is engaged in online gaming, involved in college or university life, and enjoys playing sports.

Cluster 7: Informed & Tech-Savvy Travelers

- Variables in Cluster: Travel, politics, computers, business.
- Description of Cluster: This group is well-informed, tech-savvy, and enjoys traveling. They are interested in politics, technology, and business-related topics.

Cluster 8: Family & Faith-Oriented

- Variables in Cluster: Sports fandom, food, family, religion, parenting, school.
- Description of Cluster: This cluster is oriented towards family life and faith, showing interest in sports, food, parenting, and religious activities. They also engage in school-related activities, likely reflecting a connection to children's education.

Cluster 9: Shoppers & Current Events Followers

- Variables in Cluster: Current events, shopping.
- Description of Cluster: This group keeps up with current events and enjoys shopping. They are likely to stay informed while being interested in consumer trends and products.

Cluster 10: Home & Relationship Focused

- Variables in Cluster: Home and garden, dating.
- Description of Cluster: This cluster focuses on personal life, particularly in home improvement and gardening. They also have a strong interest in dating and relationship-building activities.

If NutrientH2O wants to hone its messages more to its current social-media audience, they should create online-advertising campaigns centered around:

- Health & Wellness Enthusiasts
- Lifestyle & Aesthetic Enthusiasts
- Informed & Tech-Savvy Travelers
- Family & Faith-Oriented
- Shoppers & Current Events Followers.