3. Cluster of Customers

In this section, respondents are going to be clustered into different groups based on their rating scores on the brands attributes. Different from the previous section, where missing values have been replaced by the brands attribute means, in this part, all the missing values would be replaced with zero, since NA means the respondents have no experience about the rating objective, e.g. they have never consumed Balisto, thus cannot give evaluations to any of the corresponding attributes. By doing so on the one hand, the important information of non-experience on the products has been kept, on the other hand, the distribution of attributes variables with non-zero-ratings are not be influenced.

The cluster results mainly checked **factor representative variables.** From the previous part we got the four factors (Calories, Marketing, Taste and Packaging), from which each factor, one represented variable would be picked out as the benchmark factors to compare among clusters. The variables are selected by going back to the original data set to check the attributes with the highest variance within each group. The factor representative variables tell us from which aspect the respondents’ satisfaction are influenced? E.g. by the taste of the products or the marketing promotion of the companies?

# Calorie,rich,healthful,sweet -> rich (1.563614)

# Commercial, image, chocolaty -> commercial (1.569881)

# Creamy, crunchy, addiction, wrapping -> addiction (1.524589)

# Accessible, handy -> accessible (1.427612)

Afterwards we will go back to the respondents to see the common characteristics of each cluster, e.g. the consumption frequency, age, gender, their favorite brands etc.

**Cluster by Kmeans**

By checking the *NbClust[[1]](#footnote-1)* for determining the best number of clusters, according to the majority rule (proposed by 9 indices), the best number of clusters is 2. By silhouette structure check, 2 cluster solution is 0.61 (medium structure) and 3 cluster solution is 0.49 (weak structure). However, the 2 clusters solution just clustered the respondents into non-experienced and experienced groups, whereas from the interpretation point of view, the 3 clusters offered a better solution. Since we want to know more about e.g. what are the characteristics and consumption behavior within the experienced group? Thus 3 clusters would be chosen, which is the proposed by the second most indices (7 indices).

图片包含 文字, 地图

描述已自动生成

In general, chocolate attracts women more than men. Female have consumed more chocolate bars in different brands than male.

Appendix

**NbClust Package for determining the best number of clusters**

NbClust(attribute.rating[-1], method="kmeans")  
## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*   
## \* Among all indices:   
## \* 9 proposed 2 as the best number of clusters   
## \* 7 proposed 3 as the best number of clusters   
## \* 1 proposed 10 as the best number of clusters   
## \* 1 proposed 11 as the best number of clusters   
## \* 1 proposed 12 as the best number of clusters   
## \* 3 proposed 14 as the best number of clusters   
## \* 2 proposed 15 as the best number of clusters   
##   
## \*\*\*\*\* Conclusion \*\*\*\*\*   
##   
## \* According to the majority rule, the best number of clusters is 2   
##   
##  
## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

## $All.index  
## KL CH Hartigan CCC Scott Marriot TrCovW  
## 2 7.5067 83.8665 15.5234 8.0232 267.6280 9.582151e+18 4619.9354  
## 3 2.6561 61.9385 7.0823 5.0338 357.0975 3.601831e+18 2565.7535  
## 4 2.8523 48.8142 -0.0195 4.5585 432.7966 1.408921e+18 1811.4313  
## 5 0.3905 35.7948 5.7164 3.0340 453.0248 1.468948e+18 2342.6837  
## 6 1.1582 32.6742 5.1216 3.0155 559.5705 2.511452e+17 1829.3550  
## 7 0.5444 30.5412 9.6125 3.1472 641.1974 6.680613e+16 1324.4585  
## 8 1.6617 32.6266 6.5756 4.6697 749.0790 1.008678e+16 923.0276  
## 9 1.8176 33.0341 4.0346 5.6091 847.1322 1.796299e+15 649.2693  
## 10 0.5182 31.9039 8.0826 5.8416 881.2995 1.119746e+15 512.6137  
## 11 1.4125 34.4407 6.4664 7.3592 996.1750 1.361789e+14 395.4782  
## 12 9.4244 36.1379 1.0943 8.4488 1080.4324 3.004941e+13 289.5879  
## 13 0.1095 33.2723 6.9226 7.7966 1117.0910 1.694136e+13 276.4250  
## 14 4.3940 35.9920 1.8811 9.3818 1210.4183 3.038692e+12 203.5956  
## 15 0.1272 34.3212 -8.2907 9.0424 1275.8048 9.433506e+11 197.3841  
## TraceW Friedman Rubin Cindex DB Silhouette Duda Pseudot2  
## 2 701.2186 585.0722 9.5317 0.4110 0.6297 0.6148 0.6921 16.0131  
## 3 529.8598 642.9725 12.6143 0.3316 1.2477 0.4554 0.8335 2.7960  
## 4 460.4722 716.2712 14.5151 0.3909 1.2747 0.4185 4.7663 -21.3352  
## 5 460.6677 708.6422 14.5089 0.3325 1.5443 0.2397 0.8437 2.4081  
## 6 408.7441 908.3555 16.3520 0.3883 1.5427 0.2420 2.4467 -5.9128  
## 7 366.1267 942.2373 18.2554 0.3069 1.4097 0.2647 0.7392 5.2917  
## 8 299.2340 1123.5740 22.3363 0.3217 1.2151 0.2921 0.8535 1.8877  
## 9 258.7270 1377.5219 25.8334 0.3054 1.1564 0.3115 0.5733 14.8836  
## 10 235.5478 1364.7490 28.3755 0.2710 1.1107 0.3224 1.0233 -0.0228  
## 11 195.9528 1553.7318 34.1092 0.3078 0.9853 0.3520 2.0045 -7.0158  
## 12 168.0836 1759.3607 39.7647 0.3025 0.9311 0.3648 12.2626 0.0000  
## 13 163.3786 1984.2909 40.9098 0.3018 0.8620 0.3884 1.9100 0.0000  
## 14 137.6286 2178.4997 48.5640 0.3826 0.7684 0.4250 0.9703 0.3668  
## 15 130.7942 2473.4893 51.1016 0.3528 0.8893 0.4043 1.3645 -1.8699  
## Beale Ratkowsky Ball Ptbiserial Frey McClain Dunn Hubert  
## 2 3.8459 0.5514 350.6093 0.8651 1.7789 0.2305 0.6671 0.0010  
## 3 1.5529 0.4813 176.6199 0.7812 0.5648 0.5025 0.3033 0.0010  
## 4 -5.2665 0.4269 115.1181 0.7844 -7.1921 0.5265 0.3781 0.0010  
## 5 1.3169 0.3830 92.1335 0.5015 0.1110 1.4420 0.1144 0.0010  
## 6 -4.7289 0.3553 68.1240 0.5105 0.0842 1.4694 0.1426 0.0010  
## 7 2.9390 0.3330 52.3038 0.5276 0.1680 1.4749 0.1520 0.0010  
## 8 1.1437 0.3184 37.4042 0.5327 0.0626 1.4969 0.1944 0.0011  
## 9 6.1407 0.3036 28.7474 0.5377 0.6127 1.4733 0.1944 0.0011  
## 10 -0.1011 0.2905 23.5548 0.5026 0.0111 1.7024 0.1859 0.0011  
## 11 -4.0821 0.2806 17.8139 0.5057 -0.9545 1.6719 0.2181 0.0011  
## 12 0.0000 0.2708 14.0070 0.5181 0.9818 1.5801 0.2181 0.0011  
## 13 0.0000 0.2608 12.5676 0.5168 0.0458 1.5881 0.2181 0.0011  
## 14 0.2507 0.2529 9.8306 0.5178 0.7351 1.5771 0.2804 0.0011  
## 15 -2.1580 0.2450 8.7196 0.4876 4.6122 1.7635 0.2804 0.0011  
## SDindex Dindex SDbw  
## 2 0.6262 3.2080 0.4255  
## 3 0.8924 2.8361 0.4331  
## 4 0.9210 2.6789 0.4305  
## 5 1.3230 2.5975 0.3060  
## 6 1.2827 2.4827 0.3431  
## 7 1.2790 2.3101 0.3748  
## 8 1.2576 2.0765 0.2882  
## 9 1.2108 1.9591 0.2708  
## 10 1.2383 1.8455 0.2464  
## 11 1.1215 1.7025 0.1993  
## 12 1.0813 1.6193 0.2022  
## 13 1.0525 1.5579 0.1800  
## 14 0.9402 1.4144 0.1176  
## 15 1.1919 1.3781 0.1179  
##   
## $All.CriticalValues  
## CritValue\_Duda CritValue\_PseudoT2 Fvalue\_Beale  
## 2 0.7539 11.7537 0.0000  
## 3 0.5459 11.6457 0.1140  
## 4 0.4075 39.2588 1.0000  
## 5 0.4549 15.5808 0.2337  
## 6 0.5846 7.1064 1.0000  
## 7 0.6563 7.8570 0.0006  
## 8 0.4075 15.9943 0.3548  
## 9 0.6372 11.3864 0.0000  
## 10 0.2493 3.0108 1.0000  
## 11 0.6139 8.8034 1.0000  
## 12 0.0916 0.0000 NaN  
## 13 0.0916 0.0000 NaN  
## 14 0.6262 7.1631 0.9964  
## 15 0.6002 4.6632 1.0000  
##   
## $Best.nc  
## KL CH Hartigan CCC Scott Marriot  
## Number\_clusters 12.0000 2.0000 15.0000 14.0000 11.0000 3.000000e+00  
## Value\_Index 9.4244 83.8665 10.1718 9.3818 114.8754 3.787409e+18  
## TrCovW TraceW Friedman Rubin Cindex DB  
## Number\_clusters 3.000 3.0000 15.0000 14.0000 10.000 2.0000  
## Value\_Index 2054.182 101.9712 294.9896 -5.1165 0.271 0.6297  
## Silhouette Duda PseudoT2 Beale Ratkowsky Ball  
## Number\_clusters 2.0000 3.0000 3.000 3.0000 2.0000 3.0000  
## Value\_Index 0.6148 0.8335 2.796 1.5529 0.5514 173.9893  
## PtBiserial Frey McClain Dunn Hubert SDindex Dindex  
## Number\_clusters 2.0000 2.0000 2.0000 2.0000 0 2.0000 0  
## Value\_Index 0.8651 1.7789 0.2305 0.6671 0 0.6262 0  
## SDbw  
## Number\_clusters 14.0000  
## Value\_Index 0.1176

1. See Appendix [↑](#footnote-ref-1)