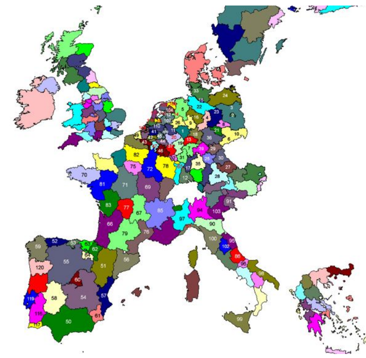


BUMK 742: Advanced Marketing Analytics

Project #4: International Market Segmentation for Global Retailers

Group #1
Shiyun Shen, Yingzheng Xue, Michael Rees, Zihan Cai, Jana Haas, Ziying Yan

November 19th, 2019

Honor Pledge:
We pledge on our honor that we have not given or received any unauthorized assistance on this assignment.

Executive Summary

Whole Foods Market is looking to maximize the customers they are reaching by identifying and positioning for certain segments in Europe. We are helping them achieve that goal by analyzing and interpreting data based on a store image study by the European Commission. We identified the number of segments and analyzed the importance of each attribute to consumers in both segments to draw our conclusions. Between the 2 segments we identified with the help of GLIMMIX, we recommend Whole Foods should enter regions belonging to segment 2, as well as focusing on providing high-quality services and a welcoming atmosphere, for these two attributes are what Whole Foods excels in and are what residents in segment 2 value most.

Introduction and Background

Whole Foods is an American multinational supermarket chain that sells organic, healthy and fresh produce, owns a self-service coffee bar, salad, and snacks and currently is mostly located in relatively middle-to-high-end areas.

Now, Whole Foods is looking to expand into the European market. However, the European market involves many different countries and large areas, which means consumers in this mixed market have different shopping habits and preferences for groceries. Whole Foods must attention to which regions and consumers to target and position towards to be successful in this new market. To stand out among competitors from local and international, Whole Foods asked us to perform market research to identify the segments in these regions and how to optimally target them. Based on all of this information we are striving to identify the market segment based on drivers of store image and determine the best entry strategy for Whole Foods.

Data and Methodology

The data we used was collected by a European Commission sponsored retail store image study. Through mail surveys, households from 107 regions across 7 countries were asked to rate overall store image and several image drivers (Region, Store Image, Service, and Atmosphere) on a 1-7 point Likert scale, with a higher rating being a higher preference. Before that, tests for wording and interpretability of store image were conducted, and researchers used back translation to ensure the content of the questionnaires has the same meaning across languages.

Traditionally, national borders are used to define segments and existing data is mostly based on that. However, because of the increasing globalization trend and the non-geographical influential factors on consumer behaviors, we adopted the new consumer-oriented approach: segmentation based on responses to marketing activities through a Mixture Regression Model using GLIMMIX. Mixture Regression Models are a predictive method to disentangle the dataset and examine the relationship between the variables in the regression model with each latent segment. GLIMMIX is an analytical tool developed by Professor Michel Wedel from the University of Maryland.

Specifically speaking, we ran the Mixture model and chose two segments with the lowest BIC value. Afterward, we ran the model of 2 segments 30 times and determined the parameter estimates with the lowest BIC value. Posterior segment membership probabilities are used as a reference to determine the segment each area belongs in to draw further conclusions.

Key Findings

According to the GLIMMIX graph (Figure 1), when analyzing the results given through GLIMMIX, out of the 5 segments that were tested, the model of two segments showed the lowest BIC value. Based on our results, we found that the sizes of these two segments are different. For all regions, the probability of being in segment 1 is 30.5%. However, the probability of being in segment 2 is 69.5%, which is more than double that of being in segment 1.

The calculated p-values show that all ratings (service, atmosphere, and price) for both segments are statistically significant. In segment 1, if all variables were equal to zero, the overall store image (intercept) is 1.725, which is considerably higher than segment 2. Also in segment 1, if service rating increases by one unit, then the overall store image increases by 0.269. If one unit increases in the atmosphere rating, the overall store image increases by 0.157. The overall store image increases by 0.297 when price increases by one unit. Based on the information above, the most important store image driver for segment 1 is the price as a one-point increase on the scale improves overall image more than a comparable increase in service or atmosphere.

Moreover, in segment 2 the most important store image driver is service, whose estimated value is the largest compared with other store image drivers. If services increase by one unit, the overall store image increases by 0.355. The impact of service in segment 2 is larger than that in segment 1, and it is the largest of any variable for either segment. If the atmosphere increases by one unit, it will result in an overall store image increases by 0.303. The impact of the atmosphere in segment 2 is larger than the impact of the atmosphere in segment 1. If price increases by one unit, the overall store image increases by 0.240, which has less effect on segment 2. Overall, service and atmosphere appear to be more important to customers in segment 2 than price, which matters more for those in segment 1. The overall store image for those in segment 2 is 0.696 when all variables equal zero.

According to the posterior membership segment probability, we divide the regions into two segments by 0.5 to examine their distribution roughly (higher posterior membership probability shows greater confidence that a region is in a certain segment). Among the 107 regions studied, France enjoys the highest percentage of regions in segment 2, which is 87.50%, while Spain has the lowest, only 60%. In Germany, 30 regions belong to segment 2, which is the highest, while there are only 3 in Portugal (See Figure 9). Also, for all the regions, Oberbayem and Schleswig-Holstein in Germany and Liege in Belgium are almost likely to be in segment 1, and Friuli-Venezia Giulia in Italy, Galicia in Spain and Koln in Germany are very likely to be segment 2. Figure 5 shows some of the regions with the most certainty of being in segment 2.

Based on the above, customers living in regions in segment 1 are more likely to focus on price in supermarkets, while customers in segment 2 tend to view supermarkets with better service and atmosphere as better. So, retailers thriving in different areas determine which segment they would do best in, including atmosphere, price, and service. If the retailers do well in atmosphere, segment 2 should be chosen, while segment 1 is better in price. While both segments care about service, segment 2 will react better to increases in service than segment 1.

In order to calculate the optimal x-values for each variable to maximize the effectiveness of our positioning, we used the solver function in excel. The optimal results would obviously be to have a rating of 7 in all categories, but this is not a feasible solution. To account for this, we set an arbitrary maximum that the maximum sum of the 3 x-values can only be 15. We then ran a solver to maximize the predicted value for segment 2, not the overall predicted value because our stores will be in regions that are heavily skewed towards segment 2. The solver solution (see figure 7) yielded an optimal distribution of aiming for a 7 rating on service, a 7 rating on atmosphere, and a 1 rating on price. However, we determined that this sort of extreme positioning may not be reasonable if we ever want to expand to reach segment 1; therefore we modified it to have an ideal rating of 7 still for atmosphere, but lowered the ideal atmosphere rating to 6 and the ideal price rating to 2. This still retains the strong service and atmosphere positioning while not being as extreme.

Conclusion and Recommendations

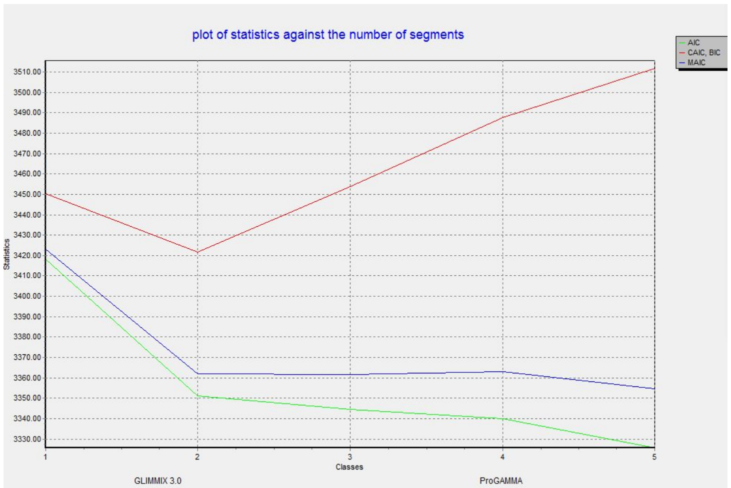
As can be seen from the positioning of Wholefoods, part of its core competence is its welcoming atmosphere and great service, but its price is relatively higher than average, which affects price-sensitive consumers. According to Figure 4, consumers in segment 2 perceive services and atmosphere as more important than price, which to those in segment 1 is the most influential factor in the overall image. As a result, we suggest focusing on segment 2 regions.

We chose segment 2 because Whole Foods is known for its service and atmosphere, not its price, thus Whole Foods is more likely to maintain its high prices by targeting segment 2 and the regions for Whole Foods to enter first should be the ones with the highest posterior membership probability of being in segment 2, which are Friuli-Venezia Giulia in Italy, Koln in Germany, Galicia in Spain, Alentejo in Portugal, and Asturias in Spain. We will position Whole Foods as having the best service in the supermarket category because it is most important to segment 2. If possible, having a good atmosphere is also recommended, because that is the second most important factor for consumers in segment 2. The average rates on service and atmosphere are 5.54 and 5.47 respectively, which WholeFoods should increase or maintain. After establishing stores in these 5 regions, we should evaluate their performance after a year, then possibly expand to the next 5 regions with the highest probability of being in segment 2. Alternatively, we could focus on one country for expansion like France where more regions in France are more likely to be in segment 2 (87.5%) than in segment 1 (12.5%) compared to any other country. We could enter the German market because it has the highest overall number of regions, but the specific population of each region is also needed to know the actual size of the regions in Germany.

One limitation of our memo is that there are no specific criteria of posterior membership segment possibilities, thus we only make estimations. If we know how many new stores Whole Foods plans to build in the area, or its budget and its breakdown, we can come up with more specific conclusions about whether to include a certain region or not. Another limitation is that the only data we have is the store image scale. More factors should be included. Some of the questions we can ask are: how many competitors are there in the chosen region? Are there enough collaborators to provide enough organic food which is the key competence of our brand? What's the attitude of regions to a foreign brand? Our memo is a good reference in terms of consumer behaviors, but not a comprehensive one due to data loss. Besides, we do not know perceived value of Whole Foods, nor overall image or three attributes value. So when making recommendations, we are not able to specify the exact scales Whole Foods should increase or decrease, only making relative comparison to the average. In the future based on the limitations we discovered throughout the process of finalizing some changes there are a couple of recommendations for future research. Having more information on region will allow us to make better conclusions, for example, the population. A highly populated area would be more appealing than a less populated area. Having this information could influence the recommendation greatly.

Appendix

Figure 1



The BIC/CAIC value between different numbers of segments. The lowest number occurs when there are two segments and it shows an upward tendency as the number increases.

Figure 2

```
#classes: 2 current: 0 startnr: 1 iteration: 27
LOG LIKELIHOOD =    -1664.528505316739000
AIC            =     3351.057010633478000
CAIC           =     3421.676789793758000
MAIC           =     3362.057010633478000
BIC            =     3410.676789793758000
Es             =        0.715695748702110
DF             =           11
R-square       =        0.672030
```

Maximum Likelihood Estimation result of the model with 2 segments.

Figure 3

```
#classes:  2 current:  0 startnr:  9 iteration:  28
LOG LIKELIHOOD =    -1664.528504854360000
AIC            =     3351.057009708721000
CAIC           =     3421.676788869002000
MAIC           =     3362.057009708721000
BIC            =     3410.676788869002000
Es             =        0.715697377361644
DF             =           11
R-square       =        0.672030
```

Maximum Likelihood Estimation result of the model with the lowest BIC value among 30 starts.

Figure 4

```
#classes:  2 current:  1 startnr:  9
Independent    Coefficient estimates        STD.ERR            T-value
 Service                   0.268637          0.025494           10.537306
 Atmosphere                0.157097          0.025301            6.209180
 Price                     0.296858          0.020898           14.205074
 intercept                 1.724784          0.127074           13.573031

CLASS SIZE FOR SEGMENT  1 =     0.305204
VARIANCE FOR SEGMENT    1 =     0.612554

#classes:  2 current:  2 startnr:  9
Independent    Coefficient estimates        STD.ERR            T-value
 Service                   0.354843          0.020563           17.256487
 Atmosphere                0.302896          0.020598           14.705301
 Price                     0.239548          0.014525           16.491652
 intercept                 0.695652          0.093259            7.459326
```

Estimation Result for the 2-Segment Model

SEGMENT 1:				
VARIABLE	Estimate	Std. Error	T-Value	P-Value
Service	0.269	0.025	10.538	0.0000
Atmosphere	0.157	0.025	6.209	0.0000
Price	0.297	0.021	14.205	0.0000
intercept	1.725	0.127	13.573	0.0000
Segment Size	0.305			

SEGMENT 2:				
VARIABLE	Estimate	Std. Error	T-value	P-Value
Service	0.355	0.021	17.256	0.0000
Atmosphere	0.303	0.021	14.705	0.0000
Price	0.240	0.015	16.492	0.0000
intercept	0.696	0.093	7.459	0.0000
Segment Size	0.695			

Entropy	0.716
(pseudo) R2	0.672

Parameter estimates of two segments of the model in Figure 3.

Equation for segments:

Segment 1: Store_image = 1.725 + 0.269*service + 0.157*atmosphere +0.297*price

Segment 2: Store_image = 0.696 + 0.355*service + 0.303*atmosphere +0.240*price

Figure 5

countries	regions	posterior in segment 2
Portugal	Alentejo	0.999
Portugal	Algarve	0.951
Portugal	Norte	0.993
Nederland	Drenthe	0.998
Nederland	Gelderland	0.992
Nederland	Noord-Holland	0.967
Italia	Friuli-VeneziaGiulia	1.000
Italia	Lombardia	0.958
Italia	Marche	0.991
France	Alsace	0.999
France	Midi-Pyrenees	0.958
France	Champagne-Ardenne	0.945
Espana	Asturias	0.999
Espana	Galicia	1.000
Espana	Rioja	0.981
Belgique-Belgie	Hainaut	0.955
Belgique-Belgie	LuxembourgB	0.979
Belgique-Belgie	Brabant	0.861
BRDeutschland	Hamburg	0.999
BRDeutschland	Koln	1.000
BRDeutschland	Unterfranken	0.993

In this chart, we pick up the top 3 regions having the highest posterior of segment 2 in their countries.

Figure 6

#classes: 2 current: 0 starter: 9 iteration: 28

CLASS	1	2
1	0.10727	0.89273
2	0.30599	0.69401
3	0.03365	0.96635
4	0.53394	0.46606
5	0.22589	0.77411
6	0.19814	0.80186
7	0.79424	0.20576
8	0.14043	0.85957
9	0.42179	0.57821
10	0.10863	0.89137
11	0.01559	0.98441
12	0.01724	0.98276
13	0.14717	0.85283
14	0.04603	0.95397

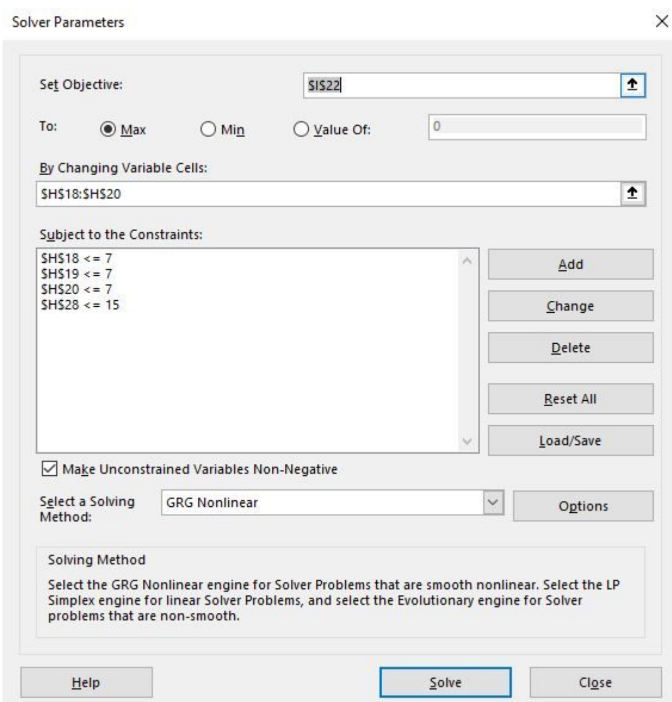
15	0.00094	0.99906
16	0.75260	0.24740
17	0.11415	0.88585
18	0.35006	0.64994
19	0.17741	0.82259
20	0.00050	0.99950
21	0.76727	0.23273
22	0.01073	0.98927
23	0.29028	0.70972
24	0.23228	0.76772
25	0.28809	0.71191
26	0.19377	0.80623
27	0.16539	0.83461
28	1.00000	0.00000
29	0.17061	0.82939
30	0.49395	0.50605
31	0.98021	0.01979
32	0.08504	0.91496
33	1.00000	0.00000
35	0.01810	0.98190
36	0.01021	0.98979
37	0.44147	0.55853
39	0.00676	0.99324
41	0.99739	0.00261
42	0.13868	0.86132
43	0.04530	0.95470
44	1.00000	0.00000
45	0.29710	0.70290
46	0.02131	0.97869
47	0.14652	0.85348
49	1.00000	0.00000
50	0.97520	0.02480
51	0.96036	0.03964
52	0.00090	0.99910
53	0.99989	0.00011
54	0.85601	0.14399
55	0.09888	0.90112
56	0.05481	0.94519
57	0.07573	0.92427

58	0.55154	0.44846
59	0.00050	0.99950
60	0.42784	0.57216
61	0.10418	0.89582
62	0.77408	0.22592
63	0.33992	0.66008
64	0.01920	0.98080
65	0.00141	0.99859
66	0.13308	0.86692
67	0.24752	0.75248
69	0.17032	0.82968
70	0.27912	0.72088
71	0.67368	0.32632
72	0.05465	0.94535
75	0.22641	0.77359
76	0.05663	0.94337
77	0.31858	0.68142
78	0.07149	0.92851
79	0.04171	0.95829
81	0.05623	0.94377
82	0.14124	0.85876
83	0.58547	0.41453
85	0.35303	0.64697
86	0.96128	0.03872
90	0.20234	0.79766
91	0.00049	0.99951
93	0.22514	0.77486
94	0.04198	0.95802
95	0.00895	0.99105
96	0.06671	0.93329
97	0.99628	0.00372
98	0.40897	0.59103
99	0.41179	0.58821
100	0.08045	0.91955
102	0.05844	0.94156
103	0.53503	0.46497
104	0.00191	0.99809
105	0.66218	0.33782
106	0.99892	0.00108

107	0.00845	0.99155
108	0.21513	0.78487
109	0.03552	0.96448
110	0.14782	0.85218
111	0.03302	0.96698
112	0.66094	0.33906
113	0.13909	0.86091
114	0.07181	0.92819
115	0.06063	0.93937
116	0.00078	0.99922
117	0.04920	0.95080
119	0.99928	0.00072
120	0.00722	0.99278

Posterior segment membership probabilities of the two segments of the model in Figure 3

Figure 7



The result of Solvers.

Figure 8

Predictions Based on Model Estimation Result		
(You need to change the X-values to test different positioning strategies)		
SEGMENT 1:		
X-VALUES	b*X	Comments
7	1.880	Scale from 1-7, Average = 5.54
6	0.943	Scale from 1-7, Average =5.47
2	0.594	Scale from 1-7, Average = 5.57
1	1.725	Intercept is always 1!
Predicted Value	5.142	Predicted store image score for segment 1
SEGMENT 2:		
X-VALUES	b*X	Comments
7	2.484	Scale from 1-7, Average = 5.54
6	1.817	Scale from 1-7, Average =5.47
2	0.479	Scale from 1-7, Average = 5.57
1	0.696	Intercept is always 1!
Predicted Value	5.476	Predicted store image score for segment 2
Overall Prediction	5.374	

The values of each attribute when the total value is assumed to be 15.

Figure 9

Country Name	No. of regions studied	No. of regions in segment 2	Percent
Belgique-Belgie	8	5	62.50%
BRDeutschland	39	30	76.92%
Espana	15	9	60.00%
France	16	14	87.50%
Italia	13	10	76.92%
Netherland	12	9	75.00%
Portugal	4	3	75.00%

The percentage of regions belonging to segment 2 in each country