
BUMK742: ADVANCED MARKETING ANALYTICS

# INTERNATIONAL MARKET SEGMENTATION FOR WHOLE FOODS

GROUP NUMBER: 4

**PREPARED BY**

CHU HSUAN TSAO
I-JU LIN
MENG-LUN WU
NAILA SHARMIN

---

**HONOR PLEDGE**

WE PLEDGE ON OUR HONOR THAT WE HAVE NOT GIVEN OR RECEIVED ANY UNAUTHORIZED ASSISTANCE ON THIS ASSIGNMENT.

## Executive Summary

The objective of this project is to help Whole Foods, a grocery store targeting health-conscious people by selling high-quality organic food, expand into Europe. The data for this project was used to identify market segments based on drivers of store image such as atmosphere, service, and price. We used a mixture regression model to find 2 segments of consumers who have different attitudes toward store image drivers. Segment 1 customers comprise 70% of the European market and perceive service and atmosphere as the most crucial factor in determining store image. On the other hand, Segment 2 customers comprise 30% of the European market and consider price as the most critical factor to store image. Whole Foods should target Segment1 customers because they emphasize more on service and atmosphere. Besides, Whole Foods should enter Germany and France for the market entry strategy because these two countries contain 55% of regions in Segment 1.

## Introduction and Background

Whole Foods is one of the well-known in the US food retail industry. The target customers for Whole Food are primarily people who have higher incomes, are less price-sensitive, are health conscious, and care about how the food is produced. Certificated by verified authoritative organizations, Whole Food always makes sure that its products meet its high standards. It positions itself as a food retail store that provides high-quality organic food to its customers. As the organic food market grows steadily in Europe, Whole Foods found that there might be a chance to start its international expansion in Europe. However, entering the new market is challenging, so Whole Foods asked us to analyze the food retailing data in Europe to gain insight about regions in Europe where customer demand meets Whole Food positioning strategy and identify regions with the best opportunity.

## Data and Methodology

In our study, we collected surveys from households across seven European countries. Each household was asked to rate the overall store image of a frequently visited food retail store and the drivers of the store image by a 1-7 point Likert scale. The surveys were collected through 137 participants across 105 regions. There are a total of 6 variables included in the dataset. (Table 1)

In this project, our primary goal is to identify market segments based on drivers of store image and develop the entry strategy for Whole Foods. To achieve the goal, we want to identify; 1) the drivers that affect food retail store image 2) the number of segments in the European market, 3) overall Whole Foods brand image in each segment 4) the size of segments 5) the distribution of the segments in Europe regions 6) the optimal segment for Whole Foods. We used a mixture regression model generated by GLIMMIX software for this purpose. Few steps we need to do to build the mixture regression model. First, since the store image is scored by the Likert scale, which is a continuous variable, we use linear regression to build the model and assume that it is normally distributed. We take store image as a dependent variable, while three drivers of store image (service, atmosphere, price) as independent variables. Secondly, we assume the number range of the segments in the European market through GLIMMIX and determine the number of segments based on the lowest BIC. After running several models, we chose the best model by the highest $R^2$ value. The higher the $R^2$ is, the better the data fit the model. Thirdly, each region provided in the dataset was assigned to one of the segments based on their posterior membership probability. If one region had a higher posterior membership probability in a specific segment, then it was assigned to that segment. Fourth, we checked the entropy statistic, which falls between 0 and 1, to see whether the segments are well separated. Higher entropy indicates a better separation of segments. Based on the mixture regression model estimation result, we calculate the predicted store image value of Whole Foods in the given segments and determine the best regions to enter by using the Excel Solver function.

## Key Findings

**Selection of Optimal Number of Segments:** We ran several mixture regression models in the software. The numeric values of the models with varied segments can be observed from the statistics tab of the Glimmix software Statistics of 2 to 4 Segments (Pic 1). An optimum number of 2 segments was selected based on the lowest Bayesian Information Criterion (BIC Plot: Pic 2)[1]. After choosing the optimum segment number, several trials with 2 segments set-up were re-run to find the most suitable model with optimum Entropy and Model Fit ($R^2$). The numeric values of the models can be found again under the statistics tab in the software (Pic 3-4). The most favorable 2 segment model primarily has the lowest BIC, highest entropy, and the highest $R^2$ (Pic 3-4).

**Selection of the Optimum Segment:** Based on previous model selection, we figured out the estimates of coefficients for each of the segments of the 2-segments model (Pic 5: Key Finding). All the variable parameters were statistically significant with $p<0.05$. A comparative analysis of both the segments based on descriptive variables reveals that segment 1 has a comparatively lower parameter value for the price, which means people in segment 1 have lower price sensitivity than segment 2 (Table 2). When it comes to retail service and atmosphere, segment 1 has higher parameters respectively (index), indicating that people in segment 1 have a higher preference for store ambiance and service quality. The segment size is also considerably larger (index: ~70%) than segment 2 (index: ~30%). The intercept value of segment 2 is much higher than segment 1 (index), which means people in segment two, in general, are satisfied with the existing retail stores, and the need for a new retail outlet is not that much. The predicted store image on a scale of 1-7 is also higher for segment 1 (Table 2.1: 5.720) than segment 2 (Table 2.1: 5.6150).

[1] BIC known as the most appropriate model-selection criteria because of their high accuracy and precision

The predicted values for each parameter was estimated based on Whole Foods' desired level of store image. E.g. average store image for service and atmosphere is 5.54 and 5.47 respectively on a scale of 1-7. Wholefood as a brand may attempt to achieve a score of 6 which is above average (Table 2.1) to justify its high price image (Table 2.1: 4.5) and to serve segment 1's desire for upscale service and atmosphere. Aiming to achieve a score of 7 will be too expensive for Whole Foods as this would require major investment on store interior and other operational costs which would be risky as they are entering a new market. This will set a standard store image value of ~5.70 (Table 2.1) across both the segments which can be considered pretty good as a new entrant. Also the people under this segment are currently underserved, which is explained by the intercept value (which is lesser than segment 2). On the basis of the analysis, Segment 1 would be the most optimum choice for Wholefood, as this segment fits the box for the brands ideal consumer group (Values store service, atmosphere and comparatively less price sensitive)

**Selection of Optimum Regions:** As the data was collected from 105 regions of Europe across 7 countries, we then assessed how these 2 segments are located across these regions by using Posterior Membership Probability (index). This also reveals the predicted store image in the regions that fall under the same segment. From the result of the Mixture Regression Model (Table 2), the table demonstrates that

**Segment 1**: Occupies 70% of observations, 80 regions, indicating nearly 70% of people in seven European countries belong to segment 1. Value service the most, followed by atmosphere and price. According to the posterior membership probability (Table 2), segment 1 distributes in these seven countries as below (Pic 6). Most of the regions of segment 1 lay in **Germany**, which is across 30 regions. In segment 1, there are 14 regions in France, 10 in Italy, 9 in Netherland and Spain, 5 in Belgium, and 3 in Portugal.

**Segment 2**: Accounts for 30% of the observations, 25 regions, representing 30% of people in seven European countries. Cares about price the most. Next factor is service and then atmosphere. Segment 2 is distributed as the following figure (Pic 6). 7 regions in Germany, 6 in Spain, 3 in Belgium, Italy and Netherland, 2 regions in France and 1 region in Portugal.

Comparison of two segments: We find that segment 1 prefers better service and atmosphere. They are less price sensitive and less favorable to retail stores, compared with segment 2. This finding helped us to understand the variability of store image and set distribution decisions and a roll out plan for Whole Food accordingly.

## Conclusions and Recommendations

We recommend that Whole Foods target Segment 1 region, whose customers consider excellent service quality and a pleasant shopping atmosphere (Whole Foods' strengths). As Whole Foods' products have higher prices than the alternatives, Segment 1 customers emphasize more on service and atmosphere than price, which indicates that Whole Foods will have an advantage in these regions.

To expand and enter the European market successfully, we recommend Whole Foods to enter Europe through Germany and France because they contain 55% of regions in segment 1, which has a higher store image toward Whole Foods. Moreover, the 44 regions (Table 3) in these two countries are in close proximity for efficient distribution. Specifically, we also recommend Whole Foods set up stores in regions with higher store images around the border between Germany and France, such as Alsace, Koln, Unterfranken, Stuttgart, and Freiburg. Taking advantage of the proximity of these regions, Whole Foods should set up a distribution center in these areas to ensure an efficient supply chain, resulting in easier expansion.

If Whole Foods wants to succeed in the European market and realize a roll-out strategy, it should continuously improve and maintain its store image. The result of the mixture regression model indicates that Whole Foods' targeted segment values the service and atmosphere. Therefore, we provide several recommendations to augment services and atmosphere:

1. **Service:** Whole Foods can launch the reward program or the loyalty program to understand customers' needs better and better serve the customers, retaining the targeted segment. Furthermore, Whole Foods should provide transparent information about food sources and the production process, which their targeted segment is concerned about. Plus, improving employee training is also critical for Whole Foods to serve its customers better. Since customers prefer organic food and care about their health, they may ask employees for further information about the products or expect employees to give suggestions.
2. **Atmosphere**: Whole Foods should make stores clean and well-organized to attract customers to shop there. Also, Whole Foods might consider paying attention to the lighting in stores, creating a comfortable and warm atmosphere. Recent EU law has enabled retail chains trademark designs for store layouts ("Designs for retail store layouts may be capable of trademark registration in Europe," 2021). Whole Foods can utilize this golden opportunity to make a grand entry into the European Market with their exceptional store layout.

Still, store image is not only related to price perception, service and atmosphere. Whole Foods should research more on how its targeted segment prefers on some other factors. For example, merchandise assortment, location convenience or store layout.

## Appendices: Tables, Exhibits, Figures

Table 1: Variable Description

| Variable | Details |
|---|---|
| REGION | the ID number for a region<br>(ranging from 1-120, total 105 regions, with a few numbers absent) |
| RESPONDENT | the ID number for a respondent within a region. (=1,2,3...) |
| STORE_IMAGE | Overall store image rating for a respondent's primary food retail outlet, on a 1-7 point scale<br>(missing values are imputed with their corresponding sample averages and thus may contain half points). |
| SERVICE | Store Driver: Service quality perception of a respondent's primary food retail outlet, on a 1-7 point scale |
| ATMOSPHERE | Store Driver: Store atmosphere perception of a respondent's primary food retail outlet, on a 1-7 point scale |
| PRICE | Store Driver: Price perception of a respondent's primary food retail outlet, on a 1-7 point scale |

Table 2: Table of estimation results for two segment model

| Estimation Result for the 2-Segment Model | | | | |
|---|---|---|---|---|
| | | | | |
| SEGMENT 2: | | | | |
| 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 1: | | | | |
| 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 | | | |

Table 2.1: Prediction based on model estimation result

| Predictions Based on Model Estimation Result | | |
|---|---|---|
| SEGMENT1: | | |
| X-VALUES | b*X | Comments |
| 6 | 2.129 | Scale from 1-7, Average = 5.54 |
| 6 | 1.817 | Scale from 1-7, Average =5.47 |
| 4.5 | 1.078 | Scale from 1-7, Average = 5.57 |
| 1 | 0.696 | Intercept is always 1! |
| Predicted Value | 5.720 | Predicted store image score for segment |
| | | |
| SEGMENT 2: | | |
| X-VALUES | b*X | Comments |
| 6 | 1.612 | Scale from 1-7, Average = 5.54 |
| 6 | 0.943 | Scale from 1-7, Average =5.47 |
| 4.5 | 1.336 | Scale from 1-7, Average = 5.57 |
| 1 | 1.725 | Intercept is always 1! |
| Predicted Value | 5.615 | Predicted store image score for segment |
| | | |
| | Overall Pr 5.688 | |

Table 3: Target Regions for Whole Foods

| Country | Region | Region Name |
|---|---|---|
| BRDeutschland | 1 | Arnsberg |
| | 2 | Berlin |
| | 3 | Brandenburg |
| | 5 | Bremen |
| | 6 | Chemnitz |
| | 8 | Dessau |
| | 9 | Detmold |
| | 10 | Dresden |
| | 11 | Dusseldorf |
| | 12 | Freiburg |
| | 13 | Giessen |
| | 14 | Halle |
| | 15 | Hamburg |
| | 17 | Karlsruhe |
| | 18 | Kassel,Landkreis |
| | 19 | Koblenz |
| | 20 | Koln |
| | 22 | Luneburg |
| | 23 | Magdenburg |
| | 24 | Mecklenburg-Vorpommern |
| | 25 | Mittelfranken |
| | 26 | Münster |
| | 27 | Niederbayern |
| | 29 | Oberfranken |
| | 30 | Oberpfalz |
| | 32 | Saarland |
| | 35 | Stuttgart |
| | 36 | Thuringen |
| | 37 | Trier |
| | 39 | Unterfranken |
| France | 65 | Alsace |
| | 66 | Aquitaine |
| | 67 | Auvergne |
| | 69 | Bourgogne |
| | 70 | Bretagne |
| | 72 | Champagne-Ardenne |
| | 75 | IledeFrance |
| | 76 | Languedoc-Roussillon |
| | 77 | Limousin |
| | 78 | Lorraine |
| | 79 | Midi-Pyrenees |
| | 81 | PaysdelaLoire |
| | 82 | Picardie |
| | 85 | Rhone-Alpes |

Pic 1: Statistics of 2 to 4 Segments (Pic 1: Key Finding)

```
#classes: 2 current: 0 startnr: 2 iteration: 23
LOG LIKELIHOOD =    -1665.208552264638000
AIC            =     3352.417104529276000
CAIC           =     3423.036883689557000
MAIC           =     3363.417104529276000
BIC            =     3412.036883689557000
Es             =        0.707273064394265
DF             =           11
R-square       =        0.668853
```

```
#classes: 2 current: 0 startnr: 1 iteration: 26
LOG LIKELIHOOD =    -1665.208552433567000
AIC            =     3352.417104867134000
CAIC           =     3423.036884027414000
MAIC           =     3363.417104867134000
BIC            =     3412.036884027414000
Es             =        0.707272781028258
DF             =           11
R-square       =        0.668853
```

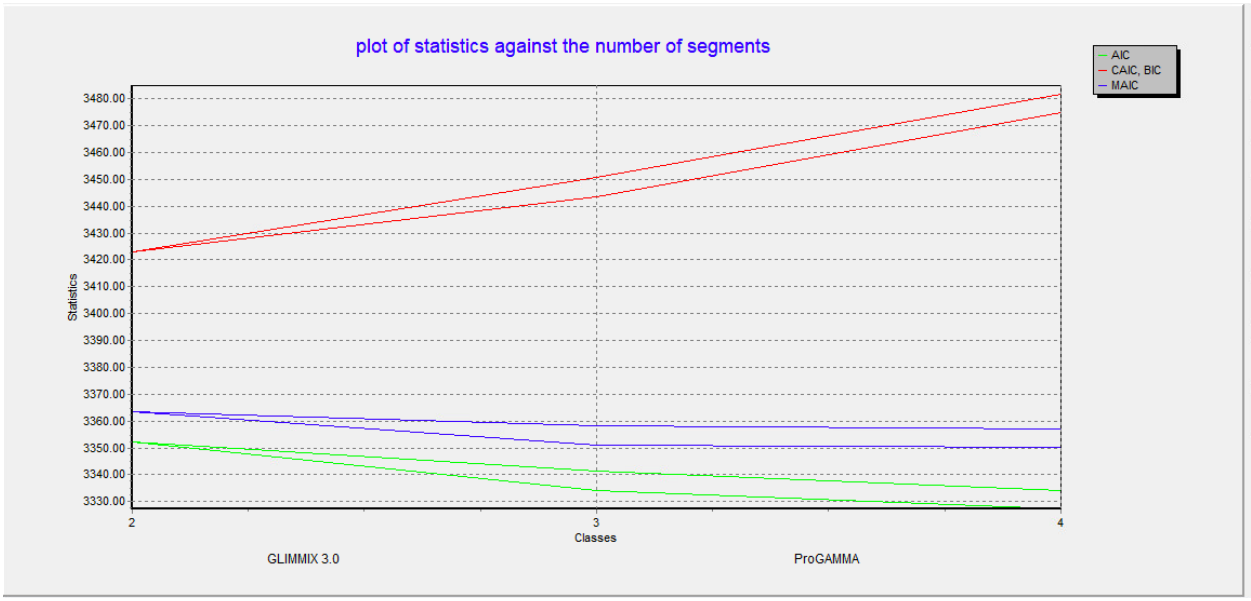
```
#classes: 3 current: 0 startnr: 2 iteration: 42
LOG LIKELIHOOD =    -1654.196857718752000
AIC            =     3342.393715437504000
CAIC           =     3451.533374139756000
MAIC           =     3359.393715437504000
BIC            =     3434.533374139756000
Es             =        0.696654566217745
DF             =           17
R-square       =        0.681344
```

```
#classes: 3 current: 0 startnr: 1 iteration: 32
LOG LIKELIHOOD =    -1655.256026298534000
AIC            =     3344.512052597067000
CAIC           =     3453.651711299318000
MAIC           =     3361.512052597067000
BIC            =     3436.651711299318000
Es             =        0.720205824333602
DF             =           17
R-square       =        0.679929
```

```
#classes: 4 current: 0 startnr: 2 iteration: 82
LOG LIKELIHOOD =    -1645.060360902571000
AIC            =     3336.120721805142000
CAIC           =     3483.780260049364000
MAIC           =     3359.120721805142000
BIC            =     3460.780260049364000
Es             =        0.694870694053094
DF             =           23
R-square       =        0.684662
```

```
#classes: 4 current: 0 startnr: 1 iteration: 61
LOG LIKELIHOOD =    -1645.060359255163000
AIC            =     3336.120718510326000
CAIC           =     3483.780256754548000
MAIC           =     3359.120718510326000
BIC            =     3460.780256754548000
Es             =        0.694723225287251
DF             =           23
R-square       =        0.684649
```

Pic 2: Plot of Statistics Against the #of segments




Pic 3-4: Statistics of 2 segment for 5 starts

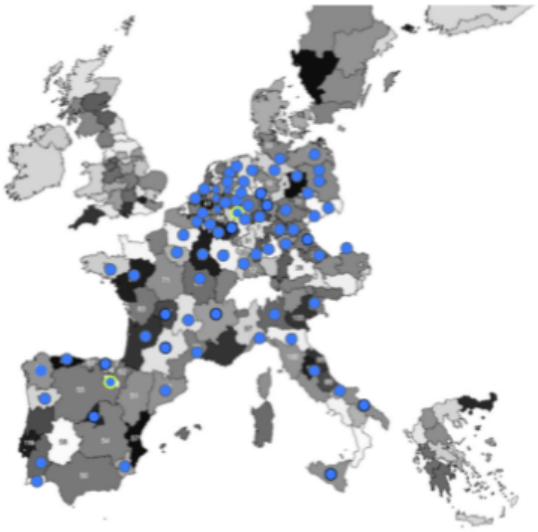
```
#classes: 2 current: 0 startnr: 1 iteration: 30
LOG LIKELIHOOD =    -1664.528503861502000
AIC            =     3351.057007723005000
CAIC           =     3421.676786883286000
MAIC           =     3362.057007723005000
BIC            =     3410.676786883286000
Es             =        0.715700961888594
DF             =           11
R-square       =        0.672030

#classes: 2 current: 0 startnr: 2 iteration: 28
LOG LIKELIHOOD =    -1664.528504239864000
AIC            =     3351.057008479728000
CAIC           =     3421.676787640008000
MAIC           =     3362.057008479728000
BIC            =     3410.676787640008000
Es             =        0.715699581503532
DF             =           11
R-square       =        0.672030

#classes: 2 current: 0 startnr: 3 iteration: 29
LOG LIKELIHOOD =    -1664.528504068160000
AIC            =     3351.057008136320000
CAIC           =     3421.676787296600000
MAIC           =     3362.057008136320000
BIC            =     3410.676787296600000
Es             =        0.715700205698881
DF             =           11
R-square       =        0.672030

#classes: 2 current: 0 startnr: 4 iteration: 16
LOG LIKELIHOOD =    -1665.208554394264000
AIC            =     3352.417108788528000
CAIC           =     3423.036887948808000
MAIC           =     3363.417108788528000
BIC            =     3412.036887948808000
Es             =        0.707269389311568
DF             =           11
R-square       =        0.668853
```

```
#classes: 2 current: 0 startnr: 5 iteration: 31
LOG LIKELIHOOD =    -1664.528503632312000
AIC            =     3351.057007264625000
CAIC           =     3421.676786424906000
MAIC           =     3362.057007264625000
BIC            =     3410.676786424906000
Es             =        0.715701799961586
DF             =           11
R-square       =        0.672029
```

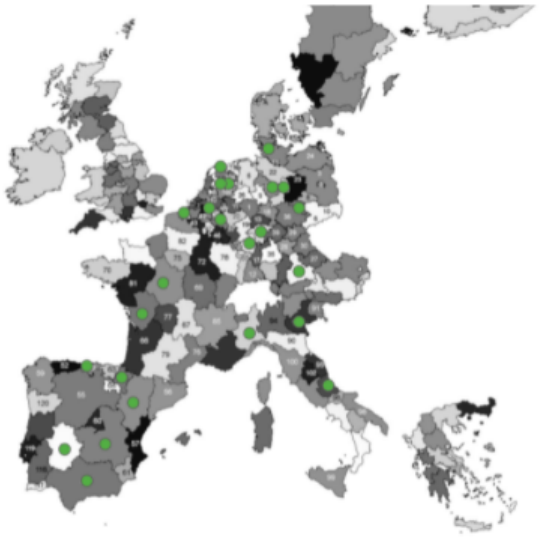
Pic 5: Coefficient and standard error of 2 segments

```
#classes: 2 current: 1 startnr: 1
Independent    Coefficient estimates        STD.ERR              T-value
 Service                    0.354844             0.020211            17.557035
 Atmosphere                 0.302894             0.020462            14.802727
 Price                      0.239548             0.014505            16.514318
 intercept                  0.695662             0.093208             7.463573

CLASS SIZE FOR SEGMENT  1 =    0.694809
VARIANCE FOR SEGMENT    1 =    0.333245

#classes: 2 current: 2 startnr: 1
Independent    Coefficient estimates        STD.ERR              T-value
 Service                    0.268635             0.025493            10.537666
 Atmosphere                 0.157097             0.025300             6.209445
 Price                      0.296859             0.020898            14.205013
 intercept                  1.724787             0.127073            13.573251

CLASS SIZE FOR SEGMENT  2 =    0.305191
VARIANCE FOR SEGMENT    2 =    0.612558
```

Pic 6: Distribution of the 2 segments



Segment 1



Segment 2