**MARKETING ANALYSIS**

**GROUP PROJECT FINAL REPORT**

**TOPIC: MARKETING RESEARCH FOR THE LAUNCH OF A NEW FAST-FOOD RESTAURANT**

**By: Chan Jun Yi**

**CONTENTS**

**Page**

**1.0 Introduction**

**1.1Background 3**

**1.2 Objectives 4**

**2.0 Methods**

**2.1 Target audience 5**

**2.2 Data Collection Procedures 6**

**2.3 Sampling Design 6**

**2.4 Statistical Methods 7**

**3.0 Results**

**3**

**.1 Customer Analysis 8 - 15**

**3.2 Food Preferences 16 - 30**

**3.3 Service Preferences 30 - 36**

**3.4 Competitor Analysis 37 - 46**

**4.0 Limitations 47**

**5.0 Conclusion 48 - 50**

**1.0 Introduction**

Launching a new business in today’s competitive generation can be extremely daunting, especially one in the food and beverage industry which already has large giants and established companies dominating it. Knowing the market and analysing where your product nee7ds to be placed, hence, becomes integral to the success of your new enterprise. Our final project focuses on delivering key metrics of product placement after thorough research and analysis of the market and competition.

**1.1 Background**

In this project, we have decided to launch a restaurant in Hong Kong, specialising in the sale of fast-food products like burgers. For this purpose, we have conducted market survey and research to determine and finalise on the various aspects of the restaurant along with the food and services we could potentially offer. We initiated our project by collecting data through online questionnaires (Google Forms) and tabulated the results to create our data set. After processing and cleaning the data, we implemented our statistical knowledge acquired throughout this course to perform customer and competitor analysis, as well as analysis on food and service preference. We used the R program to run various techniques like factor and cluster analysis, mds and correspondence analysis. After studying and interpreting the results generated, we will propose suitable implementation strategies that will not only allow us to make better decisions but also help us achieve our objectives in the most efficient manner.

**1.2 Objectives**

The objective of this marketing research is to investigate consumer behaviour with regards to their preferences in a restaurant in order to launch a fast-food chain through collecting data and interpreting their results. We will analyse the data to answer questions such as:

* What are the most popular food items to determine the different options of burger patty, sauce and side dish we will serve in the restaurant.
* Who will our target audience be and what services suit them best. Do they prefer to dine in or get food delivered? Do they prefer to be serviced by staff or self-service booths where they can order and collect food themselves.
* What is the best geographical location for the restaurant and the most favourable opening hours depending on where and when the highest footfall is, while maintaining a central location for ease of delivery if we choose to provide that service.
* What is the most optimal price ie. a competitive price to attract customers while also making a profit. We wanted to study what is the average monthly expense of customers, along with the amount they spend in fast food restaurants and the frequency of visit to establish the ideal price range of our dishes.
* Who are our competitors that are currently operating in the market? We wanted to identify the gaps and opportunities available in the market and position our business to be more distinctive in the market.

**2.0 Methods**

In this section, we will discuss our target audience, data collection procedures, sampling design and sampling methods used.

**2.1 Target audience**

Age structure of the population and geographic shifts in the population affect the growth of fast food companies. Millennials and Generations Z’s tend to be more interested in food that can be prepared and served fast. When this audience is aware of the serving time, they will be more likely to spend money on fast-food. Hence, these individuals serve as high potential customers. Apart from age structure, people in more developed cities (MDCs) are busier and often skip meals. In addition, since they generally have higher living standards, they tend to spend more. If they are aiming for a better quality of life, they are willing to spend on delivery service which will save them time as compared to picking up or taking away the food. Since the fast-food industry currently seems to be lacking behind on target marketing, our company will emphasize more on the younger generation and those who work from home or office. Thus, our new fast-food restaurant will primarily focus on attracting university students.

**2.2 Data Collection Procedures**

For data collection, we will be using the survey method or the questionnaire method to gather information. This is because of the unique characteristic of the survey method in which data is collected by asking relevant questions to the respondents. Ideally, we would have liked to conduct a few interviews with people, but as the COVID-19 pandemic still persists, conducting face-to-face surveys was not very feasible. Our only choice was to collect data by using questionnaires in google forms through different platforms such as whatsapp, instagram, facebook and emails.

**2.3 Sampling Design**

We generated a google form with a set of questions and sent it through social media asking people to answer the questions. By doing this, we did not require a random sample since the responses depended on the participants agreeing to answer the questions - they could choose whether to participate in the survey or not. Our survey collected different variables such as:

* Demographic identifiers like age, gender, monthly expense and year of study.
* Important attributes in selecting a restaurant like price, taste, nutritional value.
* Food preferences including choice of burger patty, sauce and side dishes.
* Service preferences including the favoured districts for fast-food restaurants.
* Spending habits along with frequency and time of visit (breakfast, lunch, dinner etc).
* Lastly, it asked customers’ opinions on potential competitors like McDonald’s, KFC, Burger King, Marrybrown, Shake Shack, MOS Burger and Five Guys.

The Google Form can be accessed via this link <https://forms.gle/p79tvgG3Hdx5Tkrv5>

The survey, on a point scale, measured the importance of various aspects relating to food and service quality to better understand customer needs and preferences. The reason for using a scale measure was for easy data analysis in R.

We fixed our sample size to be around 100 and we did attain this number of responses. Since our sample size was small, it was not advisable to use the probability sampling method such as a simple random sampling method or the stratified sampling methods for our survey. This is because the probability of each member’s inclusion in the sample was not fixed and would have created misleading data. We therefore used the non-probability sampling method which is the convenient sampling method. This method was conducted by taking samples from a group of people who were easy to contact or reach out to. Since our target customers were university students, this study was not impacted by the poor representation of this sampling method.

**2.4 Statistical Methods**

We will be using the statistical knowledge that we have learned in the STAT3613 course so far to analyse all the data collected from our questionnaire. This includes statistical methods mentioned in the proposal like cluster analysis, factor analysis, and multidimensional scaling. For customer segmentation, we will first implement factor analysis to identify the underlying dimensions by the relationships of variables. Using the result from factor analysis, we will use cluster analysis to classify similar respondents into a cluster to target customers according to their needs. We will also be using correspondence analysis in the competitor analysis to better visualise other restaurants and their various variables on the same perceptual map. We will explain the details in the next section.

**3.0 Results**

Analysis was conducted in four major sections namely customer analysis, food preferences, service preferences and competitor analysis. The results and interpretations are provided under each section for that specific aspect of the restaurant.

**3.1 Customer Analysis**

**3.1.1 Attributes of Fast-Food Restaurants**

In the first section of our questionnaire, we ask the respondent to rank their importance on the various attributes of a fast-food restaurant, which are price, service, taste, nutritional / health value, delivery services and meal options. By implementing factor and cluster analysis, we can carry out customer segmentation based on the attributes. The result will be able to give us a clear direction on how we should position our restaurant.

We begin our factor analysis by examining the correlation among the 6 attributes, plotting a scree plot and observing the cumulative variance explained by each number of factors.

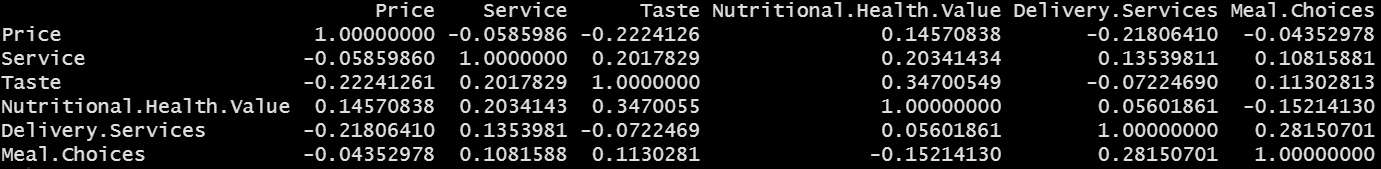


Image 3.1.1.1 Correlation Matrix of the 6 Attributes

|  |  |
| --- | --- |
| Image 3.1.1.2 Scree Plot | Image 3.1.1.3 Cumulative Variance |

We apply various criteria to determine a number of factors. First, according to the latent root criterion, we choose the number of factors with eigenvalues greater than 1 as they are considered significant. In image 3.1.1.2, it is clear that the eigenvalues for the 3rd factor is greater than 1 while that of the 4th factor is lower than 1. Therefore, 3 factors are suggested based on the latent root criteria.

Then, we use the percentage of variance criterion, which is based on the cumulative percentages of the variance explained by successive factors. By referring to Image 3.1.1.3, we can see that with 3 factors, the model can explain approximately 67% of the total variance, which is sufficient. In conclusion, we will use the 3-factor model for our factor analysis.

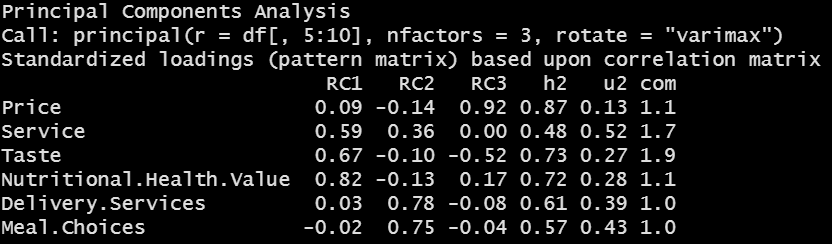


Image 3.1.1.4 Initial Factor Loadings for 3-factor Solution

We obtain the initial factor loadings for 3-factor solution, which is shown in Image 3.1.1.4 above after carrying out the principal component analysis with varimax rotation employed. We can then deduce the variables explained by each factor. In the first factor, we can see that nutritional / health value, taste and service variables have significantly higher factor loadings compared to other variables. Therefore, factor 1 can be the factor representing these 3 variables, which is the factor for the overall quality of a fast-food restaurant . Then, in the second factor, we can see that both delivery services and meal choices variables have factor loadings of over 0.7, indicating that factor 2 represents these 2 variables. Last but not least, the price variable has the highest factor loadings in the 3rd factor, thus we can conclude that factor 3 is the price factor.

After we have obtained the result of the factor analysis, we proceed our customer segmentation process with cluster analysis. We start off the analysis by measuring the squared Euclidean distance for each factor as a measure of (dis)similarity. Then, we implemented Ward's method since it is one of the hierarchical build-up methods.

|  |  |
| --- | --- |
| Image 3.1.1.5 Dendrogram | Image 3.1.1.6 Plot of Distances of Concern |

To determine the number of clusters, we refer to the dendrogram in Image 3.1.1.5 and the plot of distances in Image 3.1.1.6. Since there is a jump of distance from the 4-cluster solution to the 3-cluster solution, which indicates the 4-cluster solution is merging clusters with large differences, we thus decide to apply the 4-cluster solution for our cluster analysis. We then proceed our cluster analysis by implementing K-means clustering.

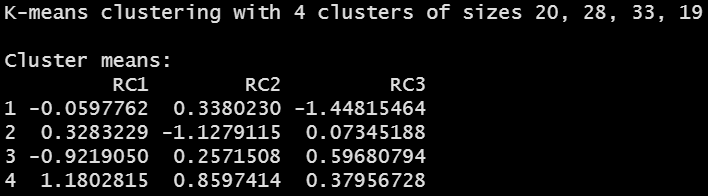


Image 3.1.1.7 Cluster Means for Each Factor

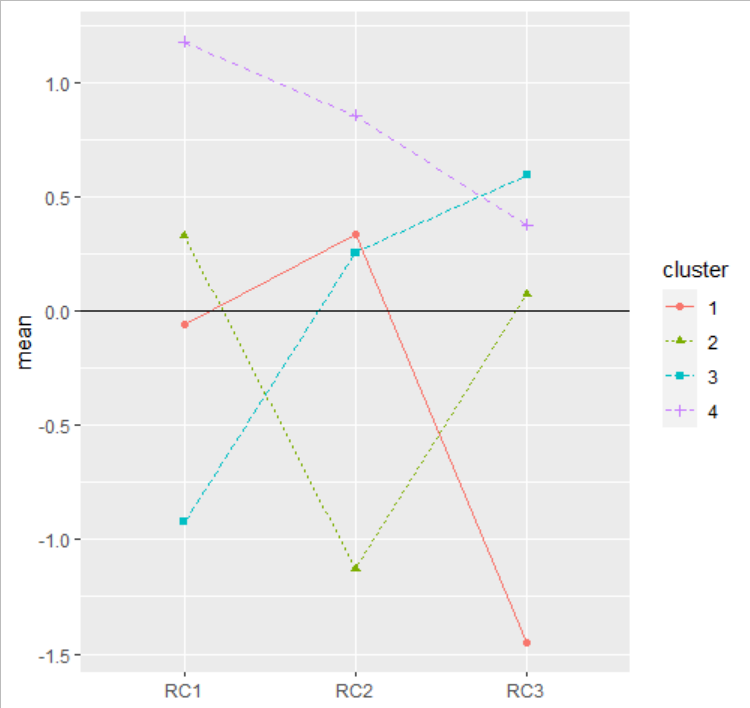


Image 3.1.1.8 Profiles Plot for Mean Values

On the next step, we try to interpret each cluster by their means and the profiles plot. For Cluster 1, we can see that it has the lowest mean in Factor 3, indicating it puts the least importance on the price factor. Cluster 2 has the lowest mean in Factor 2, so delivery and meal options do not matter much for this cluster. Then, Cluster 3 has lowest mean in Factor 1 but highest mean in Factor 2, which shows that this cluster does not really emphasize on the overall quality of a fast-food restaurant except the price. Last but not least, Cluster 4 has positive means on all factors, hence this cluster is picky on their fast-food restaurant choice and puts attention on all variables.

Lastly, we try to relate the cluster with the demographics of the respondents.

|  |  |
| --- | --- |
| Image 3.1.1.9 Mean of Age and Monthly Expense for Each Cluster | Image 3.1.1.10 Barplot on Proportion of Cluster for Each Gender |

Since Cluster 1’s monthly expense is the highest, respondents in this cluster do not put emphasis on the price factor. As for Cluster 2, the number of respondents is quite balanced in terms of gender. We can then see that Cluster 3 has the lowest monthly expense among all clusters, hence the respondents in this cluster only look into price when deciding fast-food restaurants. We can also see that the proportion of male in Cluster 3 is higher, indicating male respondents can accept fast-food of low or mid-tier in quality. Since Cluster 4 has the second highest monthly expense, respondents in this cluster will prefer fast-food restaurants with higher quality overall.

**3.1.2 Spending Habits**

In this section, we will examine the spending habits of the respondents so we can decide on the price of our menu.

|  |  |
| --- | --- |
| Image 3.1.2.1 Boxplot for Spending Amount for a Single Portion in a Fast-Food Restaurant | Image 3.1.2.2 Boxplot for Spending Amount for Delivery Charges |

|  |  |  |
| --- | --- | --- |
| Statistics | Spending Amount for a Single Portion | Spending Amount for Delivery Charges |
| Minimum | 15.00 | 0.00 |
| 1st Quarter | 40.00 | 5.00 |
| Median | 50.00 | 10.00 |
| 3rd Quarter | 55.00 | 15.00 |
| Maximum | 100.00 | 70.00 |
| Mean | 48.30 | 12.06 |
| Standard Deviation | 13.637 | 11.485 |

Table 3.1.2.3 Summary Statistics for Spending Amount for Single Portion and Delivery Charges

First, we analyze the spending amount of respondents for a single portion and delivery charges based on the summary statistics and plot out the boxplots for clear illustration. We can see that on average, respondents usually spend around HKD48.30 for a single portion and around HKD12.06 for delivery charges. On the other hand, referring to Image 3.1.2.1 and 3.1.2.2, we can see that respondents typically are only willing to pay less for delivery charges, while there is greater variability for the spending amount for a single portion. It is because respondents are willing to pay for food with higher quality, while the quality of delivery charges is standard.

**3.1.3 Other Habits**

|  |  |
| --- | --- |
| Image 3.1.3.1 Barplot for Companion When Eating in Fast-Food Restaurant | Image 3.1.3.2 Barplot for Meal Type for Fast Food |

Next, we try to examine other habits of respondents when spending in a fast-food restaurant. We plot a barplot to have a better illustration of the frequency for each variable. In Image 3.1.3.1, most of the respondents spend their time with their friends when eating in a fast-food restaurant, which is usual since all respondents are university students. In Image 3.1.3.2, most of the respondents have fast-food as their lunch. This is because since the lunch break for university students is only 1 hour, fast-food is the optimal choice for students who need to rush for classes.

**3.2 Food Preference**

**3.2.1 Patty**

We asked the respondents to rank their preference of different flavours of a burger patty, which are beef, chicken, pork, vegan, fish and other seafood. By implementing factor and cluster analysis, we can carry out customer segmentation based on the preferences.

We begin with the correlation among the 6 patty options, plotting a scree plot and observing the cumulative variance explained by each number of factors.

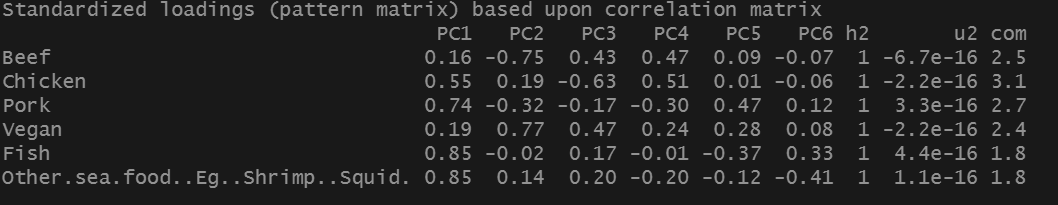
****

Image 3.2.1.1 Correlation Matrix of the 6 Patty options

|  |  |
| --- | --- |
| Image 3.2.1.2 Scree Plot | Image 3.2.1.3 Cumulative Variance |

According to the latent root criterion, we choose the number of factors with eigenvalues greater than 1 as they are considered significant. In image 3.2.1.2, we see that the eigenvalues for the 2nd factor is greater than 1 while that of the 3rd factor is lower than 1. However, the 3rd factor is closer to 1 compared to 2nd factor. Therefore, 3 factors are suggested based on the latent root criteria.

Then, we use the percentage of variance criterion, which is based on the cumulative percentages of the variance explained by successive factors. By referring to Image 3.2.1.3, we can see that with 3 factors, the model can explain approximately 76% of the total variance. Thus, we will use the 3-factor model for our factor analysis.

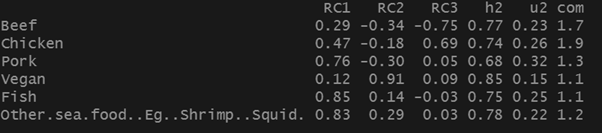


Image 3.2.1.4 Initial Factor Loadings for 3-factor Solution

We obtain the initial factor loadings for a 3-factor solution, and we can then deduce the variables explained by each factor. In the first factor, we can see that fish and other seafood have significantly higher factor loadings compared to other variables. Therefore, factor 1 can be the factor for seafood patty. Then, in the second factor, we can see that vegan have factor loadings of over 0.9, indicating that factor 2 is the factor for vegan patty. Last but not least, both beef and chicken variables have the highest factor loadings in different directions in the 3rd factor, thus we can conclude that factor 3 is the meat factor.

After we have obtained the result of the factor analysis, we proceed our customer segmentation process with cluster analysis. We start off the analysis by measuring the squared Euclidean distance for each factor as a measure of (dis)similarity. Then, we implemented Ward's method since it is one of the hierarchical build-up methods.

|  |  |
| --- | --- |
| Image 3.2.1.5 Dendrogram | Image 3.2.1.6 Plot of Distances of Concern |

To determine the number of clusters, we refer to the dendrogram in Image 3.2.1.5 and the plot of distances in Image 3.2.1.6. Since there is a jump of distance from the 5-cluster solution to the 4-cluster solution, it indicates the 5-cluster solution is merging clusters with large differences. We thus decide to apply the 5-cluster solution for our cluster analysis. We then proceed our cluster analysis by implementing K-means clustering.

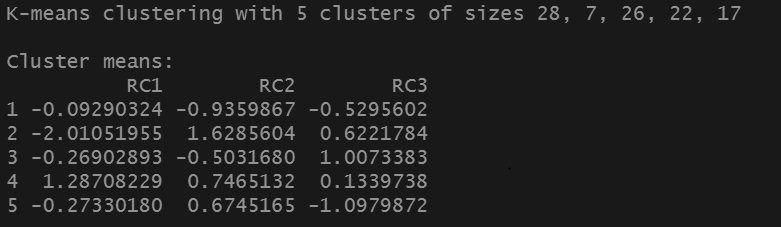


Image 3.2.1.7 Cluster Means for Each Factor

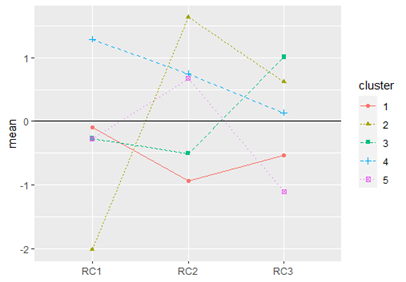


Image 3.2.1.8 Profiles Plot for Mean Values

On the next step, we try to interpret each cluster by their means and the profiles plot. For Cluster 1, we can see that it has the lowest mean in Factor 2, indicating it puts the least importance on the vegan factor. Cluster 2 has the highest mean in Factor 2, so we can say that this cluster is vegetarian. Then, Cluster 3 has the highest mean in Factor 3, which shows that this cluster has to choose between chicken or beef. This cluster can be said to be religious people as some religions do not allow eating beef. Next, Cluster 4 has positive means on all factors and highest in Factor 1, hence this cluster is seafood lovers and puts attention on all variables. Last but not least, Cluster 5 has average mean among 3 factors with high Factor 2. This cluster can be classified as a healthy balanced cluster.

Lastly, we try to relate the cluster with the demographics of the respondents.

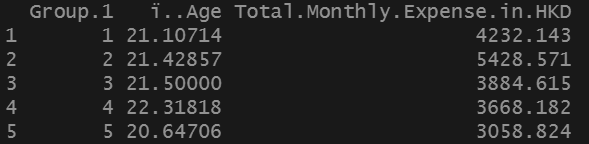


Image 3.2.1.9 Mean of Age and Monthly Expense for Each Cluster

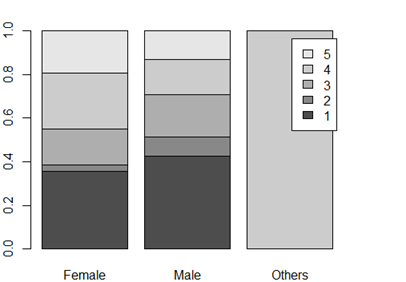


Image 3.2.1.10 Barplot on Proportion of Cluster for Each Gender

For Cluster 1, the number of respondents are quite balanced in terms of gender. Since Cluster 2’s monthly expense is the highest, respondents in this cluster do not put emphasis on the price factor. This is because vegetarians tend to have only vegan food options as their choice so they are less concerned about the price. We can then see that Cluster 3 has the third highest monthly expense among all clusters, hence the respondents in this cluster does not look into price when deciding the type of patty as they cannot eat beef and mostly go for the chicken option. Since Cluster 4 and Cluster 5 have the second lowest and lowest monthly expense, respondents in these clusters will prefer fast-food restaurants with any price range as they are not picky and they are good with different types of patties.

**3.2.2 Sauces**

Alongside patty preference, we also asked the respondents to rank their preference of different sauces they like on a burger such as BBQ sauce, Thousand Island, Tomato sauce, Mustard, Chilli sauce and sweet onion. By implementing factor and cluster analysis, we can carry out customer segmentation based on the preferences.

We begin with the correlation among the 6 sauces, plotting a scree plot and observing the cumulative variance explained by each number of factors.

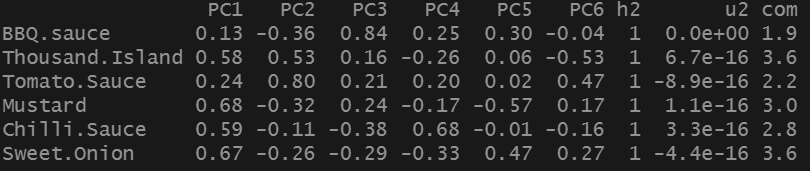
****

Image 3.2.2.1 Correlation Matrix of the 6 sauces

|  |  |
| --- | --- |
| Image 3.2.2.2 Scree Plot | Image 3.2.2.2 Cumulative Variance |

According to the latent root criterion, we choose the number of factors with eigenvalues greater than 1 as they are considered significant. In image 3.2.2.2, we see that the eigenvalue for the 3rd factor is greater than and nearest to 1 while that of the 4th factor is lower than 1. Therefore, 3 factors are suggested based on the latent root criteria.

Then, we use the percentage of variance criterion, which is based on the cumulative percentages of the variance explained by successive factors. By referring to Image 3.2.2.3, we observe that with 3 factors, the model can explain approximately 66% of the total variance. Thus, we will use the 3-factor model for our factor analysis.

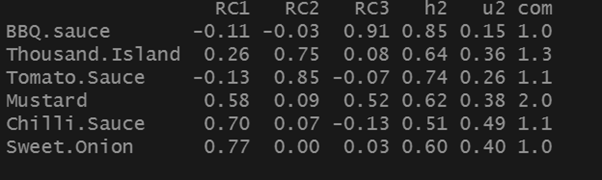


Image 3.2.2.4 Initial Factor Loadings for 3-factor Solution

We obtain the initial factor loadings for a 3-factor solution, and we can then deduce the variables explained by each factor. In the first factor, we can see that chilli sauce and sweet onion have significantly higher factor loadings compared to other variables. Therefore, factor 1 can be the factor for these 2 variables with strong spices. Then, in the second factor, we can see that tomato sauce and thousand islands have a high factor loading, indicating that factor 2 represents the sour sauce variables. Lastly, BBQ sauce and mustard have the highest factor loadings in the 3rd factor, thus we can conclude that factor 3 likes smoked BBQ.

After we have obtained the result of the factor analysis, we proceed our customer segmentation process with cluster analysis. We start off the analysis by measuring the squared Euclidean distance for each factor as a measure of (dis)similarity. Then, we implemented Ward's method since it is one of the hierarchical build-up methods.

|  |  |
| --- | --- |
| Image 3.2.2.5 Dendrogram | Image 3.2.2.6 Plot of Distances of Concern |

To determine the number of clusters, we refer to the dendrogram in Image 3.2.2.5 and the plot of distances in Image 3.2.2.6. Since there is a jump of distance from the 5-cluster solution to the 4-cluster solution, it indicates the 5-cluster solution is merging clusters with large differences. We thus decide to apply the 5-cluster solution for our cluster analysis. We then proceed our cluster analysis by implementing K-means clustering.

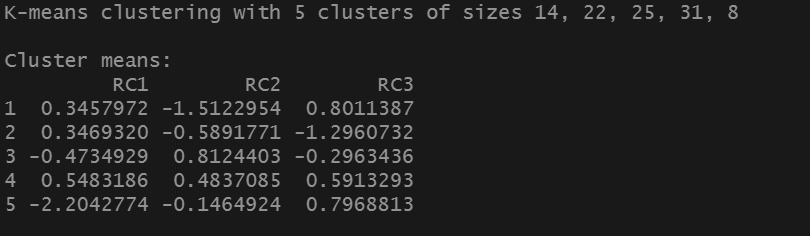


Image 3.2.2.7 Cluster Means for Each Factor

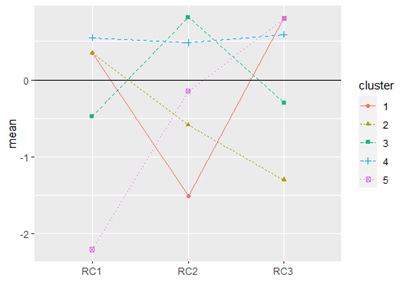


Image 3.2.2.8 Profiles Plot for Mean Values

Next, we try to interpret each cluster by their means and the profiles plot. For Cluster 1, we can see that it has the lowest mean in Factor 2, indicating that they dislike the sour sauces. Cluster 2 has a higher mean in Factor 1 and lowest mean in Factor 3, so we can say that this cluster has a liking towards chilli sauce and dislikes BBQ. Then, Cluster 3 has the highest mean in Factor 2, which shows that this cluster likes the sour sauces. Next, Cluster 4 has positive means on all factors and highest in Factor 1, hence this cluster has a liking for all sauces with a strong preference for chilli sauce. Lastly, Cluster 5 has the lowest average mean for Factor 1 and highest mean for Factor 3, indicating that they hate chilli sauce and love smoked BBQ.

Lastly, we try to relate the cluster with the demographics of the respondents.

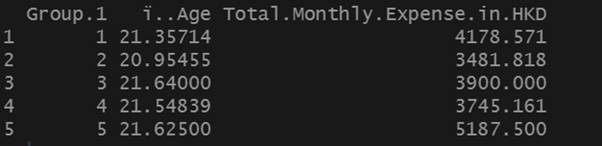


Image 3.2.2.9 Mean Age and Monthly Expense for Each Cluster

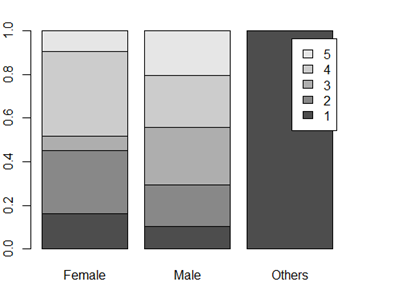


Image 3.2.2.10 Barplot on Proportion of Cluster for Each Gender

We observed that Cluster 5 has the highest monthly expense of HKD5187.5 per month. This is because this cluster only likes BBQ sauce which is expensive in the market. The other cluster with the highest mean for Factor 3 (BBQ sauce) was cluster 1, has the second highest monthly expense as it also enjoys chilli sauce in addition to BBQ. Cluster 2 which has the least preference for BBQ sauce, also has the least monthly expenditure. Cluster 4 is the next lowest as it likes all sauces equally so the expenditure balances out across different sauces.

**3.2.3 Side dishes**

The third and last preference we asked the respondents to rank was the choice of side dish including fries, potato wedges, soft drink, salad and corn. By implementing factor and cluster analysis, we can carry out customer segmentation based on the preferences.

We begin with the correlation among the 6 side dishes, plotting a scree plot and observing the cumulative variance explained by each number of factors.

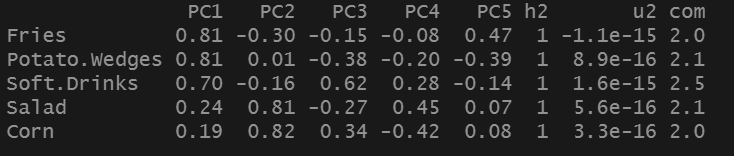
****

Image 3.2.3.1 Correlation Matrix of the 5 side dishes

|  |  |
| --- | --- |
| Image 3.2.3.2 Scree Plot | Image 3.2.3.3 Cumulative Variance |

According to the latent root criterion, we choose the number of factors with eigenvalues greater than 1 as they are considered significant. In image 3.2.3.2, we see that the eigenvalues for the 2nd factor is greater than 1 while that of the 3rd factor is lower than 1. However, the 3rd factor is closer to 1 compared to 2nd factor. Therefore, 3 factors are suggested based on the latent root criteria.

Then, we use the percentage of variance criterion, which is based on the cumulative percentages of the variance explained by successive factors. By referring to Image 3.2.3.3, we can see that with 3 factors, the model can explain approximately 82% of the total variance. Thus, we will use the 3-factor model for our factor analysis.

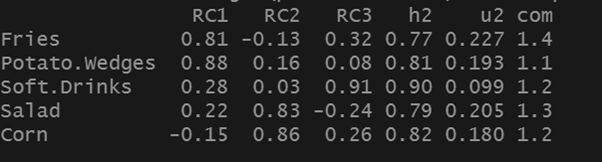
****

Image 3.2.3.4 Initial Factor Loadings for 3-factor Solution

We obtain the initial factor loadings for a 3-factor solution, and we can then deduce the variables explained by each factor. In the first factor, we can see that potato wedges and fries have significantly higher factor loadings compared to other variables. Therefore, factor 1 can be the factor for these 2 variables of fried dishes. Then, in the second factor, we can see that corn and salad have factor loadings of around 0.85, indicating that factor 2 represents the healthy option variable. Lastly, the third factor is soft drinks with a factor loadings of 0.91.

After we have obtained the result of the factor analysis, we proceed our customer segmentation process with cluster analysis. We start off the analysis by measuring the squared Euclidean distance for each factor as a measure of (dis)similarity. Then, we implemented Ward's method since it is one of the hierarchical build-up methods.

|  |  |
| --- | --- |
| Image 3.2.3.5 Dendrogram | Image 3.2.3.6 Plot of Distances of Concern |

To determine the number of clusters, we refer to the dendrogram in Image 3.2.3.5 and the plot of distances in Image 3.2.3.6. Since there is a jump of distance from the 5-cluster solution to the 4-cluster solution, it indicates the 5-cluster solution is merging clusters with large differences. We thus decide to apply the 5-cluster solution for our cluster analysis. We then proceed our cluster analysis by implementing K-means clustering.

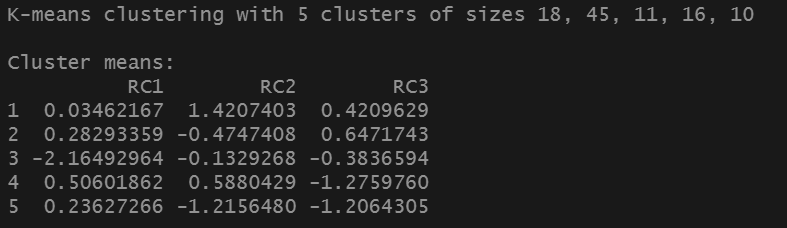


Image 3.2.3.7 Cluster Means for Each Factor

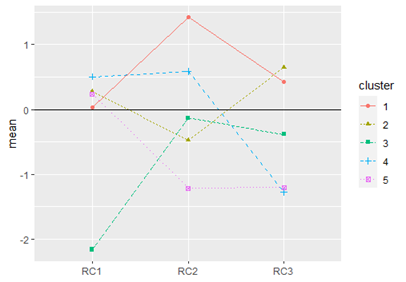


Image 3.2.3.8 Profiles Plot for Mean Values

On the next step, we try to interpret each cluster by their means and the profiles plot. For Cluster 1, we can see that it has the highest mean in Factor 2, indicating that this cluster places emphasis on healthy eating. Cluster 2 has the highest mean in Factor 3, so we can say that this cluster likes soft drinks, although they prefer all kinds of side sets. Then, Cluster 3 has the lowest mean in Factor 1 and below 0 mean in general, which shows that this cluster doesn’t particularly enjoy a side dish especially those which are fried. Next, Cluster 4 has similar means in Factor 1 and Factor 2, hence this cluster prefers a balanced diet. Last but not least, Cluster 5 has the lowest average mean in Factor 2. This cluster can be classified as an unhealthy cluster.

Lastly, we try to relate the cluster with the demographics of the respondents.

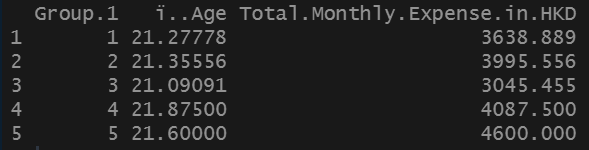


Image 3.2.3.9 Mean of Age and Monthly Expense for Each Cluster

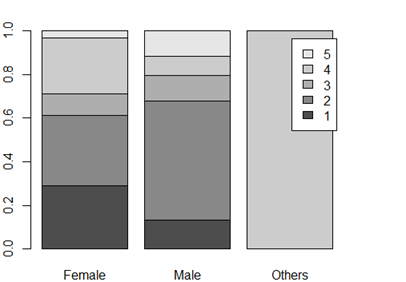


Image 3.2.3.10 Barplot on Proportion of Cluster for Each Gender

We notice that clusters with unhealthy eating habits such as cluster 5 spend the most on fast-food with side dishes. Cluster 4 which enjoys a more balanced eating habit but prefers side nonetheless (be it fried or healthy) has the next highest monthly expense. Similarly, Cluster 2 prefers soft drinks and thus has the third highest expenditure. On the contrary, clusters which don’t like side dishes in general like Cluster 3 have the lowest expenditure. Cluster 1, which has healthy eating and prefers a healthy side, has the second lowest monthly expenditure after the cluster that doesn’t like side dishes at all.

**3.3 Service Analysis**

**3.3.1 Service Preference**

Along with food preferences, we also asked the respondents to rank their service preferences which included wauter service, self-service, takeaway and delivery. By implementing factor and cluster analysis, we can carry out customer segmentation based on the preferences.

We begin with the correlation among the 4 service options, plotting a scree plot and observing the cumulative variance explained by each number of factors.

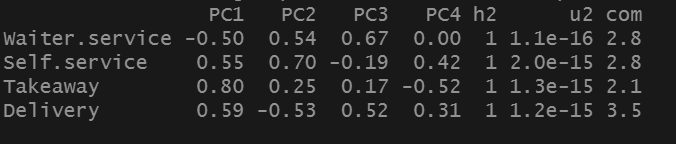
****

Image 3.3.1.1 Correlation Matrix of the 4 Service options

|  |  |
| --- | --- |
| Image 3.3.1.2 Scree Plot | Image 3.3.1.3 Cumulative Variance |

According to the latent root criterion, we choose the number of factors with eigenvalues greater than 1 as they are considered significant. In image 3.3.1.2, we see that the eigenvalues for the 2nd factor is greater than 1 while that of the 3rd factor is lower than 1. Therefore, 2 factors are suggested based on the latent root criteria.

Then, we use the percentage of variance criterion, which is based on the cumulative percentages of the variance explained by successive factors. By referring to Image 3.3.1.3, we can see that with 2 factors, the model can explain approximately 67% of the total variance. Thus, we will use the 2-factor model for our factor analysis.

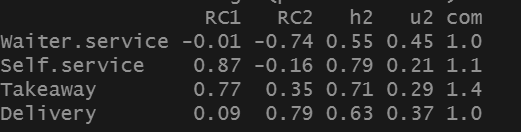


Image 3.3.1.4 Initial Factor Loadings for 2-factor Solution

We obtain the initial factor loadings for a 2-factor solution, and we can then deduce the variables explained by each factor. In the first factor, we can see that self-service and takeaway have significantly higher factor loadings compared to other variables. Therefore, factor 1 can be the factor for these 2 variables indicating independence. In the second factor, both delivery and waiter-service variables have the highest factor loadings in different directions. Thus we can conclude that factor 2 is the delivery factor.

After we have obtained the result of the factor analysis, we proceed our customer segmentation process with cluster analysis. We start off the analysis by measuring the squared Euclidean distance for each factor as a measure of (dis)similarity. Then, we implemented Ward's method since it is one of the hierarchical build-up methods.

|  |  |
| --- | --- |
| Image 3.3.1.5 Dendrogram | Image 3.3.1.6 Plot of Distances of Concern |

To determine the number of clusters, we refer to the dendrogram in Image 3.3.1.5 and the plot of distances in Image 3.3.1.6. We decided to apply the 2-cluster solution for our cluster analysis. We then proceed our cluster analysis by implementing K-means clustering.

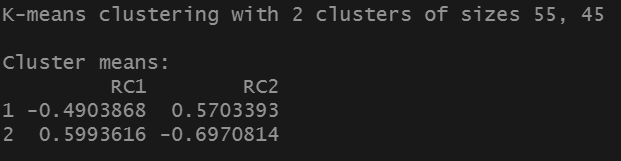


Image 3.3.1.7 Cluster Means for Each Factor

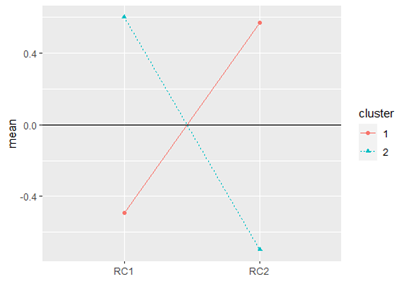


Image 3.3.1.8 Profiles Plot for Mean Values

On the next step, we try to interpret each cluster by their means and the profiles plot. For Cluster 1, we can see that it has the low mean in Factor 1 and high mean in Factor 2, indicating that this cluster is lazier compared to others and prefers delivery service or waiter service. Cluster 2 has the high mean in Factor 1 and low mean in Factor 2, which shows that this cluster consists of independent people who like self-service or come to pick up their own food.

Lastly, we try to relate the cluster with the demographics of the respondents.

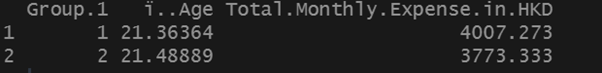


Image 3.3.1.9 Mean of Age and Monthly Expense for Each Cluster

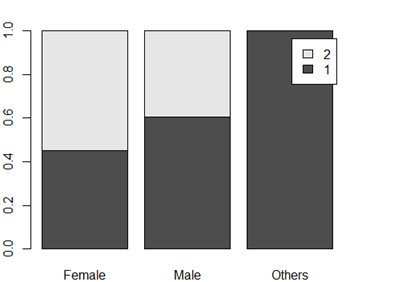


Image 3.3.1.10 Barplot on Proportion of Cluster for Each Gender

We can clearly see that Cluster 1 has higher monthly expenses due to the fact that they need to pay for the services they want i.e delivery charge or service fee. Cluster 1, on the other hand, has less monthly expenditure as they don’t pay for services. In terms of gender, there doesn’t seem to be much of a disparity in terms of independant and service preference. Having said that, females are still higher in cluster 2 indicating that they prefer independence in comparison to males who prefer to avail the services offered by the restaurants.

**3.3.2 Location Preference**

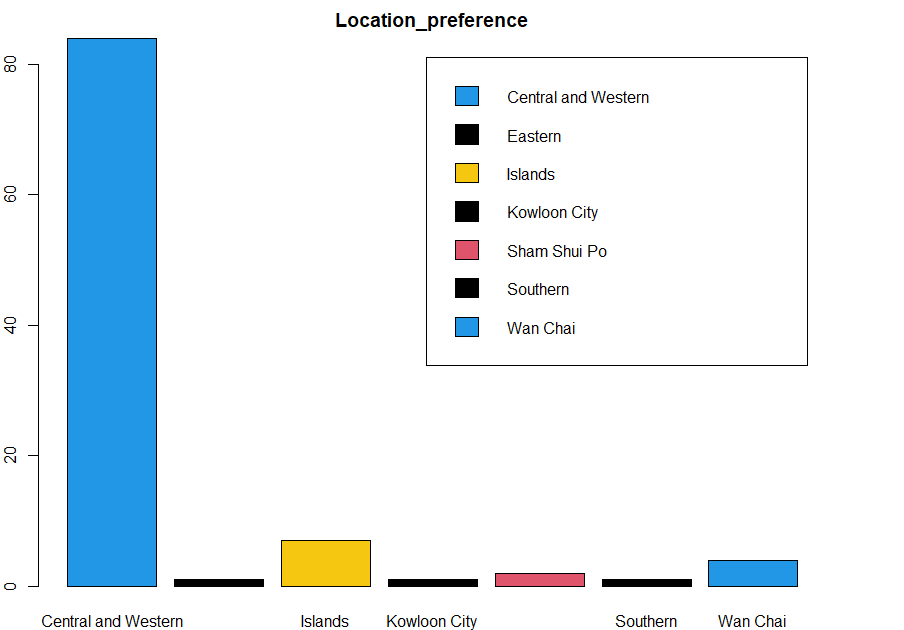
****

Image3.3.2.1 Barplot on location preference

We can clearly see that almost all the respondents prefer to have meals in Central and Western, since all of our respondents study in HKU. Therefore, we should set the location of our restaurant to be in the area around Central and Western. If we are planning to have more shops, the shops should be in areas around Hong Kong island as the Islands and Wan Chai are the second and third most preferable meal places.

**3.4 Competitor Analysis**

**3.4.1 Preference Scores**

In the final section of the survey, we asked our respondents about their views on some potential competitors of the new restaurant: McDonald’s, KFC, Burger King, Marrybrown, Shake Shack, MOS Burger and Five Guys.

In the survey, the respondents rated each of these restaurants on how much they would recommend them on a 7-point scale, with 7 being the most recommended and 1 being the least.

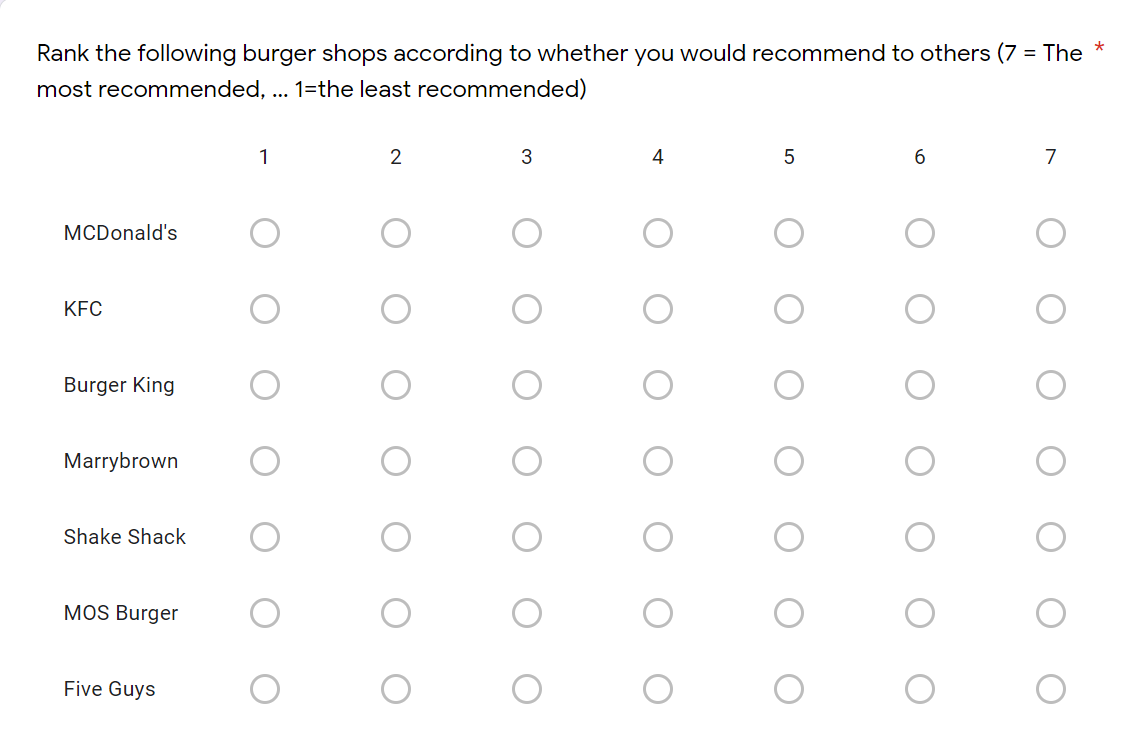


Image 3.4.1.1 Survey Question

With these scores, we conducted multidimensional unfolding in R using the cmdpref function in the cmdpref library. For the purpose of this analysis, four responses which had the same rating for all restaurants had to be excluded as the function returned an error with them present. The statistical analysis was applied on the remaining 96 responses.

With the function, applied we get two-dimensional coordinates for each of the seven restaurants and ideal vectors for the 96 respondents.

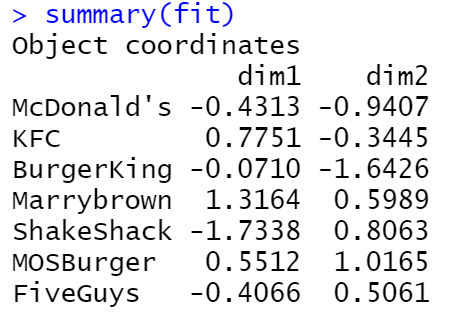


Image 3.4.1.2 Coordinates for the restaurants

With this data we create a perceptual map.

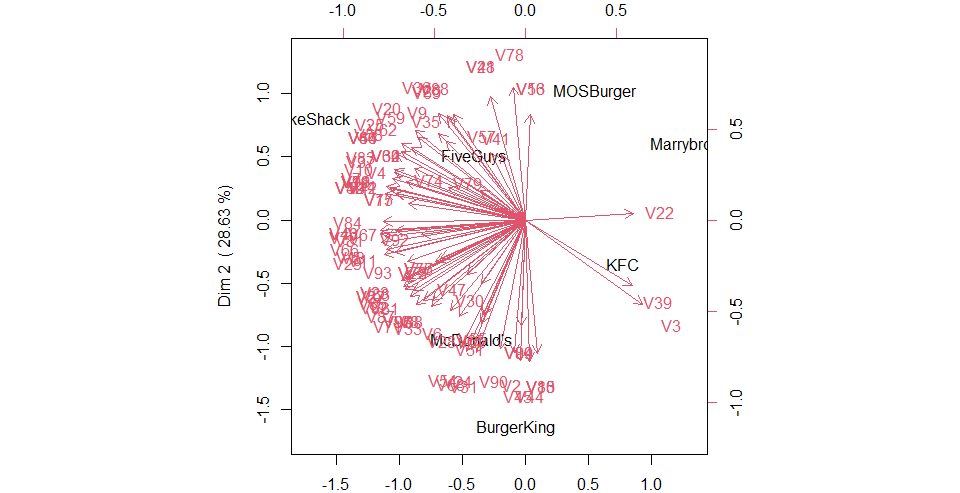


Image 3.4.1.3 Perceptual Map for Respondents and Restaurants

From Image 3.4.1.3, it is clear that the vast majority of respondents show a preference towards Shake Shack, McDonald’s, Five Guys and Burger King. Very few show preference towards KFC, Marrybrown and MOS Burger.

It is also possible to identify direct competitors to the restaurants: Shake Shack and Five Guys; McDonald’s and Burger King; MOS Burger and Marrybrown. KFC does not have a clear competitor, perhaps, because it is a fried chicken focused restaurant unlike the others.

**3.4.2 Attributes**

In this section we will explore the different qualities or features that the respondents attributed to each of the restaurants. In the survey, we let them choose as many attributes they felt the restaurants had from the options: acceptable price, good service, good taste, high nutritional value, good delivery service, and large number of meal choices.

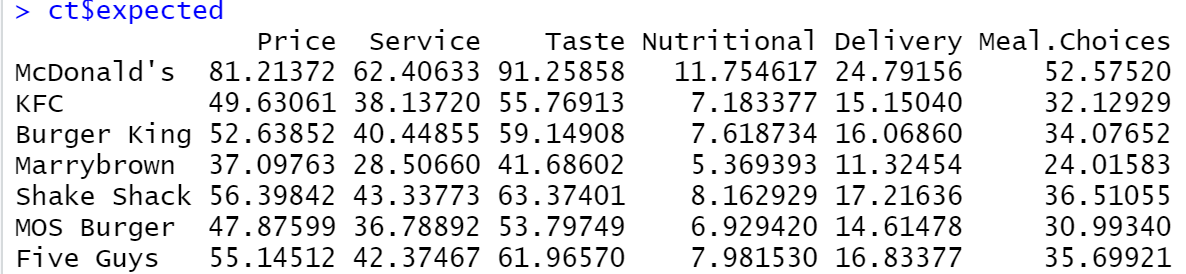
### 

Image 3.4.2.1 Survey question asking respondents for the attributes of each restaurant

The table below shows the results of this question, with each field representing the number of respondents who felt the restaurant had the respective attribute. These results tell us that, for example, that 98% of the respondents felt that McDonald’s had an acceptable price while only 21% felt that way about Shake Shack.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Acceptable Price | Good Service | Good Taste | High Nutritional Value | Good Delivery Service | Large number of meal choices |
| McDonald's | 98 | 38 | 72 | 0 | 48 | 68 |
| KFC | 75 | 26 | 44 | 3 | 21 | 29 |
| Burger King | 63 | 29 | 65 | 8 | 11 | 34 |
| Marrybrown | 58 | 23 | 25 | 4 | 10 | 28 |
| Shake Shack | 21 | 75 | 86 | 7 | 7 | 29 |
| MOS Burger | 35 | 39 | 57 | 23 | 8 | 29 |
| Five Guys | 30 | 62 | 78 | 10 | 11 | 29 |

By applying a chi-squared test on this data we found that the p-value is <2.2e-16. This suggests that there is a significant relationship between the rows and columns.



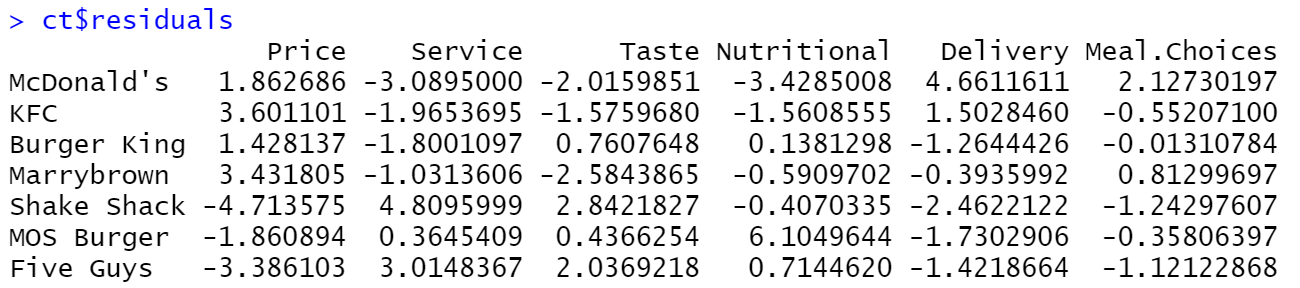
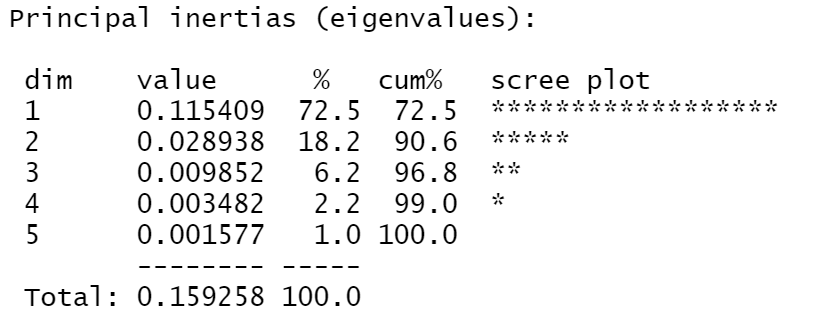


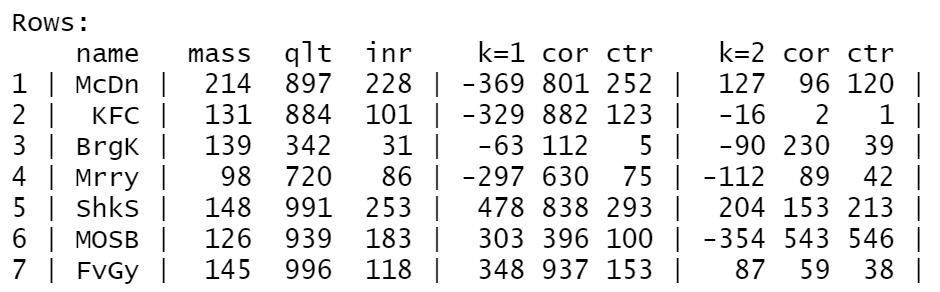
Image 3.4.2.2 Expected and residual values from the chi-squared test

The residual values from the chi-squared test also give insight into each restaurant’s relative strengths and weaknesses. For instance, McDonald’s has a very good perception among respondents on its delivery service while its food has a reputation of having a poor nutritional value.

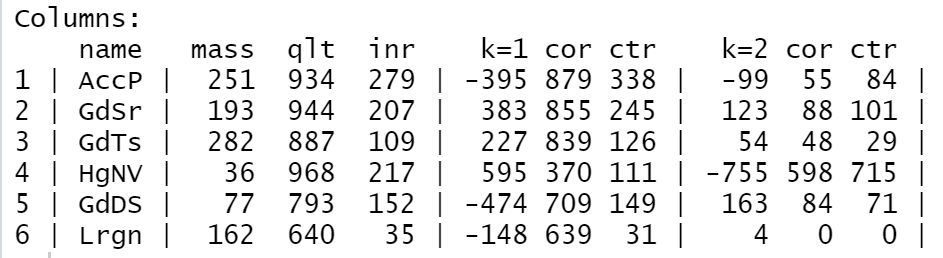
On this data, correspondence analysis was applied using the ca function in R.



Two dimensions account for 90.6% of the variance.



The largest contributors for the first dimension are Shake Shack and McDonald’s which are also on the opposite ends. Likewise, for the second dimension, MOS Burger and Shake Shake are the biggest contributors.



For the columns, the first dimension is being mostly influenced by price, while the second is being influenced by nutritional value.

By plotting the coordinates of the rows and columns, we get a perceptual map.

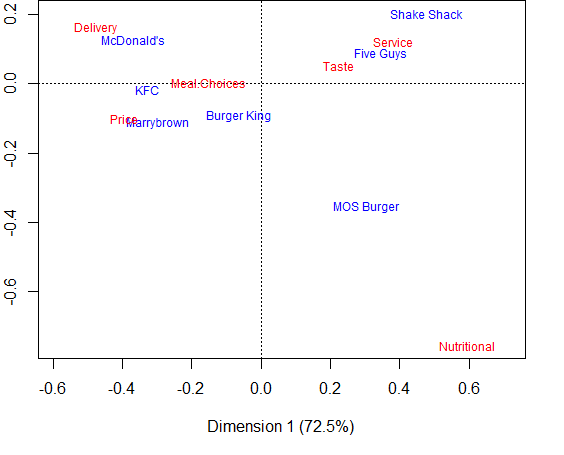


Image 3.4.2.3 Perceptual Map

From the map, we can see each restaurant’s strengths and competitors based on their location. There seem to be three distinct groups with different attributes in it. In the top-left section, McDonald’s, KFC, Marrybrown and Burger King are closely placed along with the price, delivery and meal choices attributes. This suggests that these restaurants follow the traditional fast food principles of affordability, efficiency, and appealing to a wide range of people.

On the other hand, in the top-right section of the map, Shake Shack and Five Guys are positioned along with service and taste attributes, which suggests that these are the more premium fast food restaurants among this group. These two restaurants had the lowest chi-squared residual scores but had the highest in terms of taste and service. Finally, in the bottom-left section, MOS Burger is the only restaurant which does not seem to be facing any close competition from the other restaurants. It is also not closely positioned with any of the attributes. However, it is the closest restaurant to the nutritional value attribute. Note that in the dataset used for this analysis, MOS Burger was the only restaurant to receive over 20 responses associating it with this attribute. Its position may be explained by the fact that it is a non-American brand which serves burgers inspired by Japanese cuisine.

**3.4.3 Grouping**

Finally, we asked our respondents to place the competitors into groups based on their similarities. They assigned a number to each restaurant to represent their group as shown in Image 3.4.3.1.

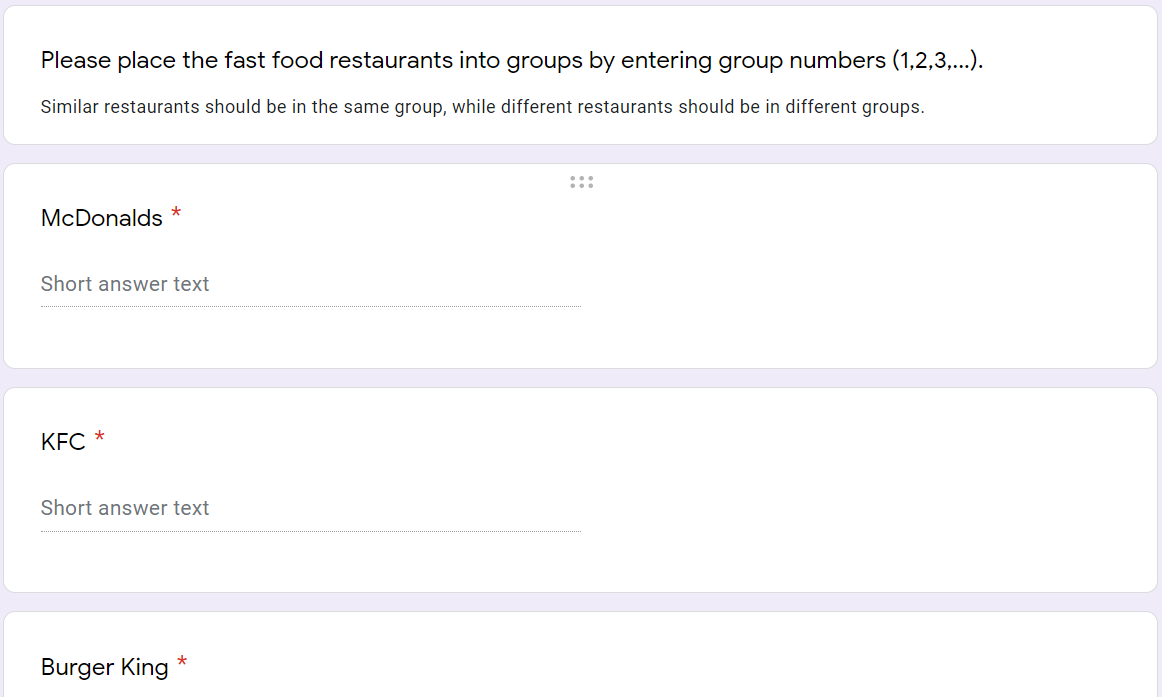


Image 3.4.3.1 The respondents were asked to group the restaurants

This data was standardized and a distance matrix was constructed using squared Euclidean distance as shown in Image 3.4.3.2..

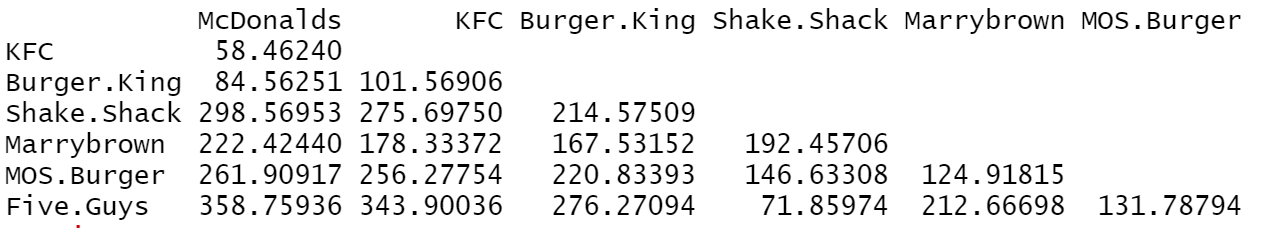


Image 3.4.3.2 Distance matrix

The *mds* function of the *smacof* library was applied to the matrix to conduct multidimensional scaling. Two dimensions were chosen along with ordinal transformation. This was because ordinal transformation was a better fit with a stress value of 0.002 compared to interval transformation which had a stress value of 0.04.

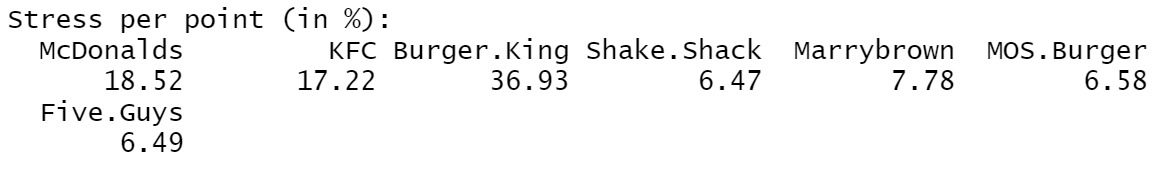
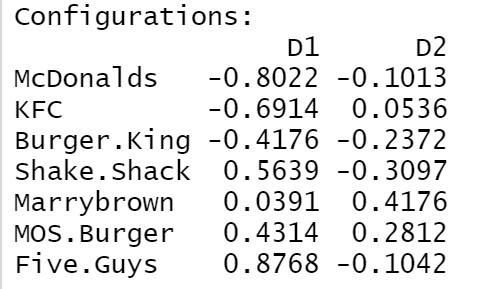
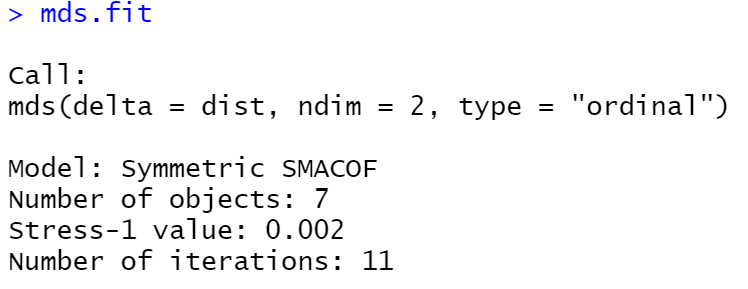


Image 3.4.3.3 Results of MDS

Using the coordinates from the above results, a perceptual map was plotted.

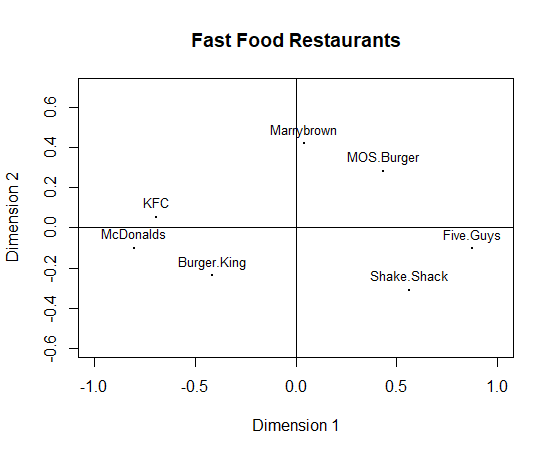


Image 3.4.3.4 Perceptual Map

In this map, there seem to be three groups of restaurants suggesting a perceived similarity between them. KFC, McDonald’s and Burger King are positioned close to each other in the left, MOS Burger and Marrybrown in the top, and Five Guys and Shake Shack located in the right.

By referring to the attribute scores in the previous section 3.4.2, a reasonable guess can be made as to what the dimensions of this map represent. Dimension 1 appears to be positioning the brands based on their perceived affordability, with cheaper restaurants such as McDonald’s and KFC on the left, and premium restaurants such as Shake Shack and Five Guys on the right.

On the other hand, the second dimension seems to be classifying these brands on their perceived “Western-ness” or “Americanness”. With the exception of MOS Burger (Japanese) and Marrybrown (Malaysian), all other restaurants are American brands. The non-American restaurants have menus with elements of their country’s cuisine, differentiating them from the others.

With this information, McDonald’s, KFC and Burger King can be positioned as affordable American restaurants, MOS Burger and MarryBrown as non-American restaurants, and Shake Shack and Five Guys as premium American restaurants.

**4.0 Limitations**

There are several limitations that are worth to be taken into consideration. First, in our project, due to time and manpower constraints, we have only collected 100 responses as the representative for our target audience. Since there are in total 30,629 of current students in The University of Hong Kong, the margin of error will be 8.21% given the confidence interval is 90%. Therefore, we have to tolerate a high amount of error, which may affect our result and final recommendation. In order to achieve a margin of error of 5%, we have to collect at least 270 responses.

Second, the responses collected may not be a good representation of the whole population. Since all of our group members are international students, we only manage to spread our questionnaire in our circle of friends, which are also all international students. Therefore, we are not able to collect enough data from local and Mainland students, so the responses collected may not represent our target population. If there is more time available and physical class is allowed, we can conduct a random face-to-face survey in the university compound to get a better representation of the population.

Third, among all of our respondents, 68 of them are males, which are significantly higher compared to females. This imbalance in gender may lead to the overall result to be more skewed towards the preference of males instead of females. Last but not least, we have received feedback from some of the respondents saying that they do not recognise some of the competitors listed in our questionnaire. This problem may affect our competitor analysis as the respondents may rank the competitors in random.

**5.0 Conclusion**

In the final section, we will make decisions on our business based on the analysis we implemented above. The position of our fast-food restaurant will focus on offering fast-foods that are the best value of money for university students. It is because in the customer analysis, we can see that 33% of the respondents put high importance on the price factor, which makes sense as students have limited budgets and prefer food with a cheap price tag without compromising much on the quality. According to the analysis of spending habits, the average price of the meals of our fast-food restaurant should be similar to the mean of spending amount on a single portion by the respondents, which is HKD48.30.

Besides, we can offer relevant and effective promotion by using the results of our analysis. Since there are 56% of respondents who dine in fast food restaurants together with their friends, we can offer promotions that encourage friends to purchase their meal together. For example, we can offer a 15% off for the 2nd set meal and 30% off for the 3rd set meal for our opening promotion. Moreover, more promotion should be held during the lunch period, since 48% of the respondents will visit fast-food restaurants for lunch. Examples of promotions include offering lunch value sets during 12pm - 2pm to customers with valid student ID cards. We believe that these promotions can persuade students with limited budgets to give a first try on our restaurant together with their friends and further stimulate growth in our sales.

With regards to what patty options to serve in our restaurant, we have decided to have higher fish and seafood options available with a lower focus on vegan food since there are 28% of respondents in cluster 1 which is the highest and only 7% of respondents in cluster 2.

Preference for sauces was quite diversified so we plan to have all the different sauces in our meal set as 31% of the respondents love all the sauces. However, since respondents preferred sour and chilli sauces more (50%), we will invest in developing some unique flavours relating to these two sauces while the BBQ sauce can be the regular one as only 8% of the respondents prefer that.

Lastly, for side dishes, the cluster that preferred soft drinks has the highest proportion of respondents so we will definitely include a variety of drink options with our meals. They also displayed a preference for other side dishes so we will include a few options for the healthy and fried dishes as well.

While analysing the service preference, we discovered that both clusters have almost an equal percentage of respondents (55% in cluster 1 and 45% in cluster 2). So we have decided to provide self-service options for people who like independence and delivery for people who prefer availing the lazier option. To determine the charges of the delivery service, we can refer to the boxplot of Spending Amount for Delivery Charges in Image 3.1.2.2. We should set the delivery charges similar to the mean of the boxplot, which is HKD12.06. With both self-service and delivery options, we will attract all kinds of customers including busy individuals who might need quick processing time or those only who can only rely on delivery for meals.

In terms of location of the restaurant, it will be around the area of Central and Western since almost most of the respondents will like to have meals at these areas according to the results of the survey.

Competition in the fast food industry is fierce with many multinational brands offering a wide range of options for customers. In order to succeed, it is important to have a competitive advantage and a differentiating factor for the new restaurant. Thus we recommend that the new restaurant be positioned as an affordable restaurant with innovative cuisine. Most competitors for the restaurant in the perceptual map (Image 3.4.3.4) are perceived to have a more American cuisine. The few restaurants which have non-American cuisines are not in the affordable segment of the map. This presents an opportunity to capture a less explored market segment, while also scoring high on price and taste factors. By serving seafood patties inspired by the local cuisine, along with unique sauces, the restaurant has an opportunity to compete with established fast food brands.