Detailed Report on Market Segmentation For a chain Restaurant using Cluster Analysis

By Kartik Kokane

Introduction

In the last 50 years, businesses have jumped leap and bounds in the marketing sector. Marketing segmentation has been at the forefront for increasing revenue and efficiency of a business (Kotler, 1983). And with the exponential rise in the amount of data generated every day, segmentation methods are getting better by the day. One of the most common ways for segmentation is based on the customer characteristics and behaviours. This data can be used to create homogenous groups of customers that have similar characteristics while maintaining inter-cluster heterogeneity. These segments can be used to analyse specific market areas, which are then targeted with custom marketing strategies tailored to the specific customer group (Aaker, 2007).

Literature Review

Though the descriptive methods yield useful data for segmentation, they are not efficient at revealing patterns between customer groups. Significant research has been conducted on customer attitudes and behaviours but not on the actual reliability of the models. There is little data on the actual potential of these methods to yield consistent results over time (Wedel, 2000). (Crawford-Welch, 1990) suggested that the industry needs to differ from their usual way of descriptive analytics and incorporate more multivariate techniques for segmentation.

Cluster analysis is one of those multivariate techniques that has been found to be valuable in product marketing research and identifying customer characteristics (Arabie, 1994) (Srivastava, 1984). In this method, customers are divided into smaller homogenous groups that show similar characteristics among themselves but have distinct differences from other cluster groups. For a method to be reproducible in different data, the segments should have internal as well as inter-segment stability. The reproducibility of the data depends on whether the data contains true and natural element variables, or the segments created by the clustering method (Dolnicar, 2002). Although cluster analysis is one of the popular segmentation techniques, it requires taking subjective decisions before implementation (Tonks, 2009). One of the most crucial decisions is determining the number of optimal clusters to create.

In this report, we implement and discuss the results of market segmentation in food industry using cluster analysis. A chain restaurant with more than 25000 stores plans to open a store at a new location for which the customer market is to be segmented.

Methodology

The dataset consists of demographic values and survey answers of 1000 customers with their location and consumption habits. To segment the data, initial exploration of the data was undertaken in Tableau. Since the average order size and average order frequency is provided, we create a Customer Value Matrix model which provides us with important

insights on the quadrant a customer belongs to. The Customer Value Matrix denotes four general classification of customer groups based on their consumption habits. Customers in quadrant 1 generally like to consume but not on a frequent basis. Quadrant 2 denotes customers who spend the most on every transaction who also consume more frequently. These customers hold the most value to enterprises since they generate the highest average revenue. Customers in quadrant 3 do not spend a lot on average but do frequent the store more. Customers in quadrant 4 do not really have a fixed pattern and hence are unpredictable to some extent.

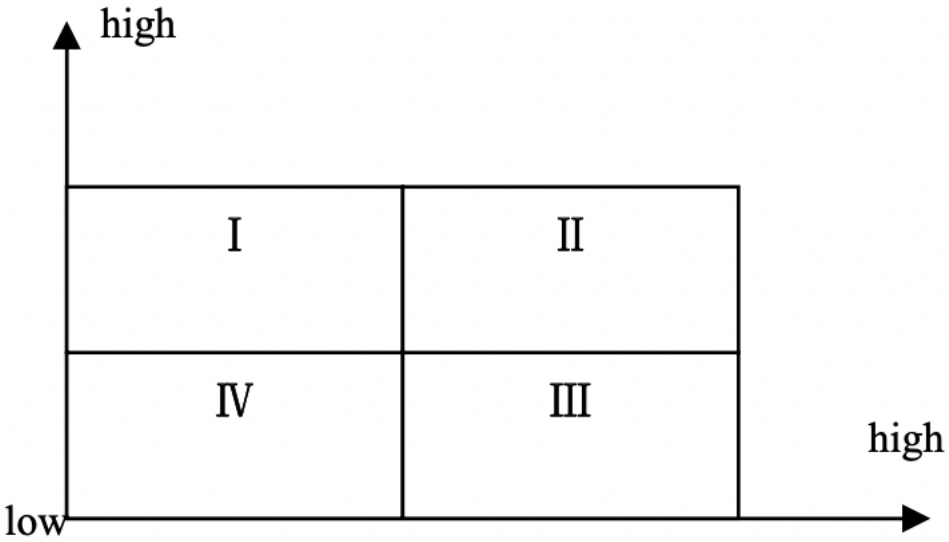


Fig. 1

Before implementing segmentation models on the dataset, we first determine if the data needs to be prepared. The occupation characteristic is divided into 10 different columns with respect to 10 different work occupations, with each field value denoted by either 1 or 0 (binary equivalent of yes or no respectively). These 10 columns are merged into one string column containing the respective occupation value of customers. This reduces the complexity of the dataset and enables us to use these values to categorize our visualisations better. Secondly, since we are using the distance formula to create clusters, all data values are standardized into a range of 0-1. This enables the model to be scaled accurately without any irregularities. A new calculated variable is created by multiplying average order size and avg. order frequency, giving us the average revenue generated by a customer in a year. This variable will be used to plot clusters against revenue enabling us to determine which clusters should be targeted.

Since the optimal number of clusters is unknown, hierarchal clustering is implemented to create a dendrogram which shows us the general hierarchy of clusters. There are different linkage methods for hierarchal clustering namely, complete, average, single and ward, which provide different results depending on the type of data used. To determine the best linkage method, we run a function that calculates the agglomerative coefficient of all linkage methods. Using the best linkage method, the cluster results are analysed by plotting an

elbow-graph. The kink in the graph denotes the optimal cluster number. Once the optimal cluster number is found, we implement K-mean clustering analysis to finally segment the data into the clusters using the calculated value as a measure of K. The segment numbers are stored in a new list and added back to the original dataset. This dataset is then loaded into Tableau again to create the final visualisations.

Results

Segmentation

Initially, we run a function to calculate the agglomerative coefficients of all linkage methods so we could determine the best method. The closer the agglomerative coefficient value is to 1, the better the method fit is for the dataset.

```
  average    single  complete      ward
0.9871487 0.9735004 0.9931839 0.9977127
```

Fig. 2

In fig. 2, we observe that complete and ward methods give the highest value, so we use the ward method to implement hierarchal clustering and plot an elbow graph for the first 10 points. The elbow plot helps determine the optimal cluster number by calculating the Within-Cluster Sum Square (WCSS), which is the squared distance between the data points and cluster centroids.

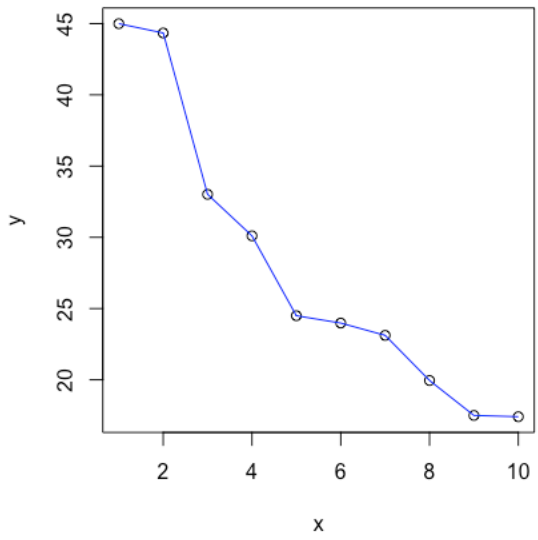


Fig. 3

As seen in fig. 3, the line forms a kink at point 5 which denotes that there will be minimal data loss for cluster numbers more than 5. We take the optimal number as 5 and implement K-means cluster analysis.

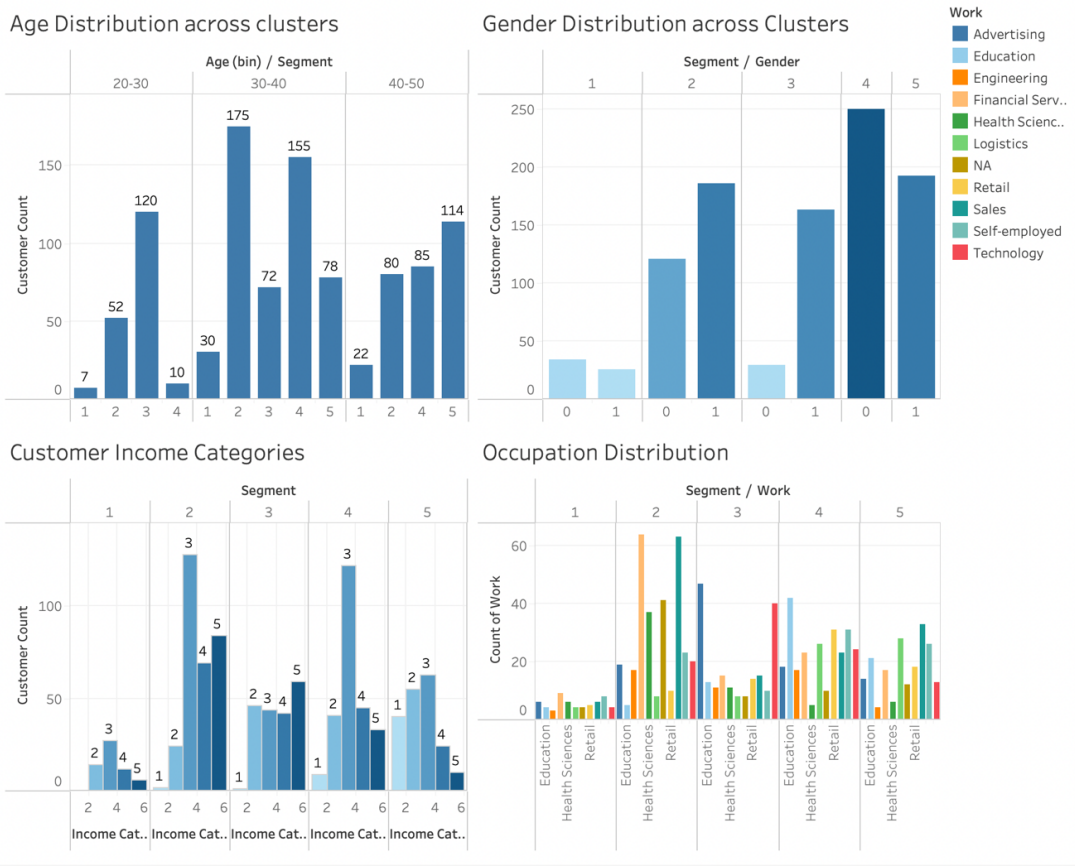


Fig. 4. Demographic Dashboard

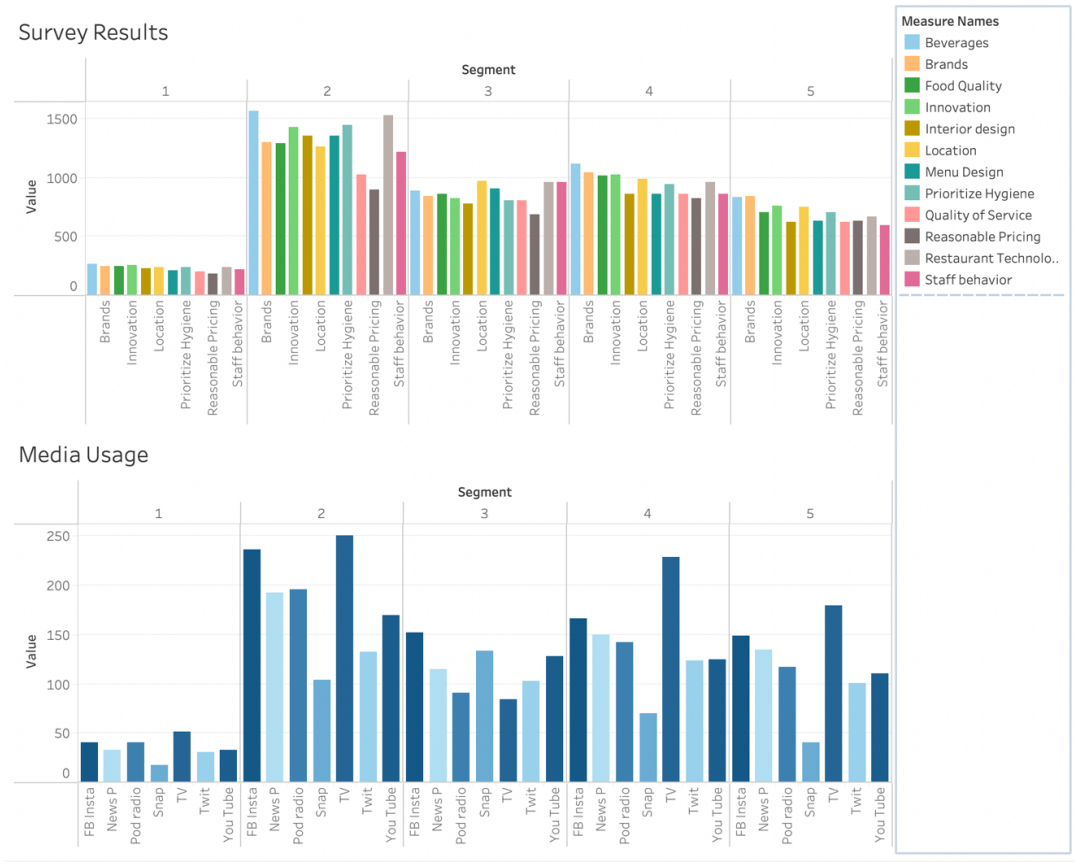


Fig. 5. Psychographic Dashboard

Cluster Profiles

Cluster 1: The customers in cluster 1 lie in an age range of 30-50 with an even distribution in gender. They work in financial services or sales and possibly live in a suburban area. They enjoy watching television but also like to spend time outdoors. They are involved with their local community and like to give back through volunteering. They value convenience when it comes to dining and would prefer to dine at a place that is easily accessible. They are practical and brand loyal, buying products from brands that suit them and with whom they have had positive experience before. They are family oriented and strive for professional success.

Cluster 2: The customers in this group are in their mid 30s. The majority of cluster comprises of females with an undergraduate degree and working in financial services or sales. They like watching television in their free time and are an active user of Instagram/Facebook, often checking in multiple times of the day to stay updated on their friends, brands, or any recent trends they follow.

They prefer quality over quantity, preferring to invest in better products that will last longer. They place a high value on family and try to spend time with their loved ones. They are also passionate about their career and value financial stability.

Cluster 3: This cluster group are again mostly comprised of females in their mid or late 30s. Majority of them work in advertising or technology and have an annual income more than 175,000 USD. This group enjoys dining out and trying new cuisines. They value overall experience when they are dining out with their family or friends, considering staff behaviour, location, and ambience of the restaurant. They are active social media users who like to stay connected to their friends and follow the latest trends and influencers.

They prefer quality over quantity but also like to spend money on trendy items and shop online. They are career oriented and have a strong social responsibility, seeking out brands and companies that align with her values.

Cluster 4: This cluster lies in an age range of 30- 40 with all customers being males. The majority work in the Education Sector and like spending their free time watching television. They like to dine at places that serve different and good quality of beverages. They like quality over quantity and tend to look up online reviews and recommendations.

They work for a good quality of life and seek out experiences that enhance their overall being. They also have a strong social responsibility and seek out brands that align with their values.

Cluster 5: This cluster group only has females who lie in an age range of 40-50. They work in sales or logistics and have a high consumption of television. They are brand loyal and

prefer convenience over anything. They enjoy getting a good deal and like to shop during discount and sale periods. They are career driven but also like to spend quality time with their families.

Targeting & Positioning

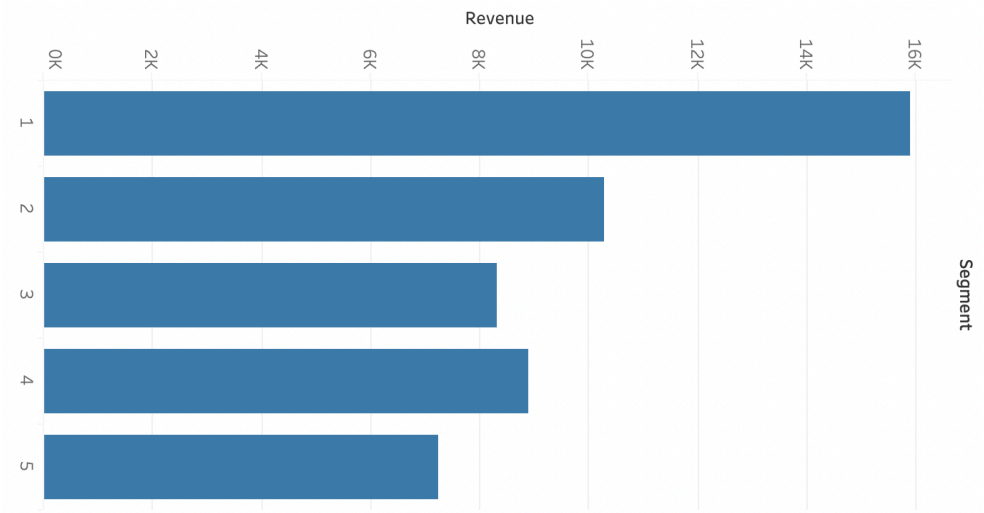
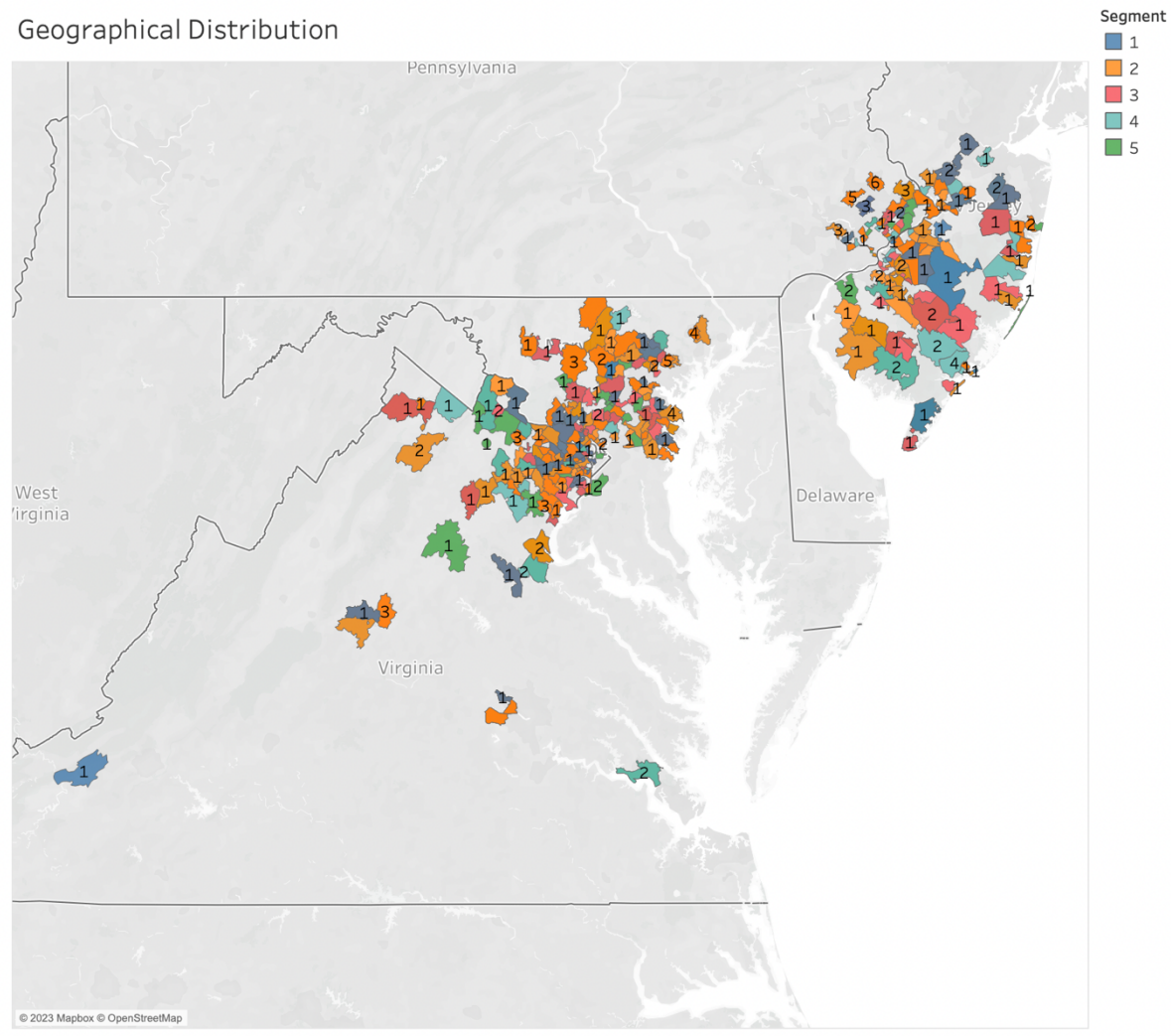
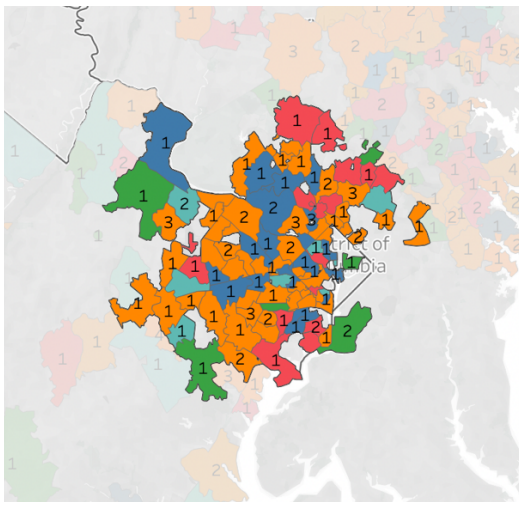


Fig. 6.

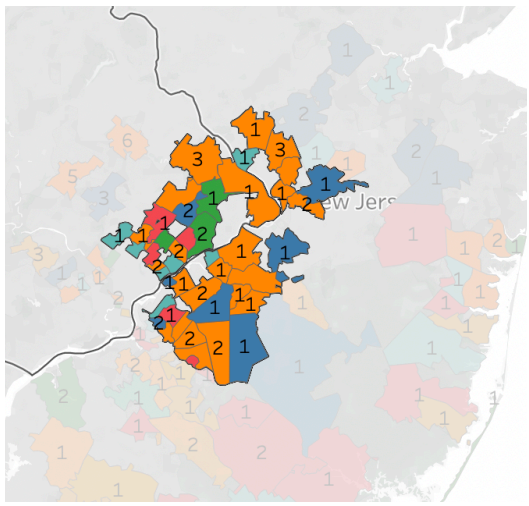
As seen in fig, Segment 1 generates the highest revenue, which suggests the company should prioritize targeting this segment. To appeal to this segment, marketing messages should focus on convenience, practicality, and brand loyalty. Advertisements should emphasize the convenience of shopping at the store and the wide range of practical products available. Television advertising may be particularly effective, as customers are regular viewers of television programs. Messaging should also align with customer's values of family, professional success, and community involvement. Highlighting the company's involvement in local organizations or charitable causes may help to build trust and loyalty with customers. Social media advertising may be less effective for them, as they have light consumption of Facebook and Instagram.

Segment 2 has the second highest average revenue but has the highest density of customers. Hence, successful marketing on this segment could be highly beneficial to the company. To appeal to them, marketing messages should focus on quality, convenience, and staying up-to-date with the latest trends. Advertisements should emphasize the value of investing in high-quality products that will last a long time, while also showcasing the convenience of online shopping and expedited shipping options. Social media advertising should be utilized heavily, particularly on Facebook and Instagram, to reach the customer where they spend much of their time online. Advertisements should be visually appealing and highlight trendy items that are currently popular. Messaging should also emphasize the importance of family and financial stability, aligning with customer's values.





Washington



Philadelphia

Fig. 7.

After plotting customer data-points on a map, it is observed that the densest customer groups are found to be in Washington and Philadelphia. These are densely populated

metropolitan cities in those regions. Hence, the company should consider opening a new store in a metropolitan city, ensuring maximum customer reach.

Conclusion

Although, Cluster analysis is one of the most widely used methods, there were some limitations observed during the process. Some clusters have overlap over other clusters where customers share some common characteristics which makes it harder to develop targeted marketing strategies. Another limitation of this method is that, if the base variables are not selected accurately, the model can have problems differentiating segments adequately. Though customer surveys are a good way to gain insight, they do not reveal all behaviours or attitudes of the customer and need more research to gain a deeper insight.

In conclusion, while cluster analysis can be a useful tool for hotel/restaurant market segmentation, it is important to be aware of its limitations and potential biases when interpreting the results. Careful selection of variables, appropriate clustering algorithms, and additional research to validate the findings can help ensure that the resulting segments are meaningful and useful for marketing strategies. Moreover, being aware of these limitations can help avoid the risk of oversimplifying the market and help develop more effective marketing strategies tailored to specific customer needs and preferences.

Bibliography

Aaker, D. A. K. V. &. D. G. S., 2007. *Marketing research (9th ed),* Hoboken: Wiley: s.n.

Arabie, P. &. H. L., 1994. Cluster analysis in marketing research. *Advanced methods of marketing research (pp. 160-189),* pp. 160-189.

Crawford-Welch, S., 1990. Market segmentation in the hospitality industry.. *Hospitality Research Journal, 14(2), 295-308.,* 14(2), pp. 295-308.

Dolnicar, S., 2002. A review of data-driven market segmentation in tourism. *Journal of Travel and Tourism Marketing,* 12(1), pp. 1-22.

Kotler, P. &. M. G., 1983. *Principles of marketing,* New York: Prentice Hall.

Srivastava, R. K. A. M. I. &. S. A. D., 1984. A customer-oriented approach for determining market structures.. *Journal of Marketing,* Volume 48, pp. 32-45.

Tonks, D. G., 2009. Validity and the design of market segments. *Journal of Marketing Management, 25(3–4), 341–356.,* 25(3-4), pp. 341-356.

Wedel, M. &. K. W., 2000. *Market Segmentation : Conceptual and methodological foundations,* Norwell: Kluwer Academic publishers.

Appendix

R- code:

```
## Install Packages (if needed)
install.packages("cluster")
## Load Packages and Set Seed
library(readxl)
library(psych)
library(dplyr)
library(cluster)

set.seed(1)
setwd("/Users/kartik/Desktop/marketing analytics")
getwd()

# Import Data
seg <- read_excel("Restaurant Data.xlsx") ## Choose retail_segmentation.csv file

#creating new calculated variables for visualization
seg <- seg %>% mutate(revenue = avg_order_freq * avg_order_size)
seg$work = NA

#Combining multiple columns into one to reduce complexity
seg$work = case_when(
  seg$Health  == 1 ~ "Health Sciences",
  seg$Finc == 1 ~ "Financial Services",
  seg$Sales == 1 ~ "Sales",
  seg$Advt  == 1 ~ "Advertising",
  seg$Edu == 1 ~ "Education",
  seg$Cons == 1 ~ "Logistics",
  seg$Eng == 1 ~ "Engineering",
  seg$Tech == 1 ~ "Technology",
  seg$Retail == 1 ~ "Retail",
  seg$SMB == 1 ~ "Self-employed"
)

#initial Visualisation
summary(seg)
describe(seg)
str(seg)

########################################
############ Segmentation ############
########################################
```

```
df1 <- subset(seg, select = c(Age, Income, Gender, Education, avg_order_size,
avg_order_freq, revenue))
#define linkage methods
m <- c( "average", "single", "complete", "ward")
names(m) <- c( "average", "single", "complete", "ward")

#function to compute agglomerative coefficient
ac <- function(x) {
  agnes(df1, method = x)$ac
}

#calculate agglomerative coefficient for each clustering linkage method
sapply(m, ac)

# Run hierarchical clustering with bases variables
seg_hclust <- hclust(dist(scale(df1)), method="ward.D2")
summary(seg_hclust)

# Elbow plot for first 10 segments
x <- c(1:10)
sort_height <- sort(seg_hclust$height,decreasing=TRUE)
y <- sort_height[1:10]
plot(x,y); lines(x,y,col="blue")

# Run k-means with 5 segments
seg_kmeans <- kmeans(x = scale(df1), 5)

# Add segment number back to original data
segment = seg_kmeans$cluster
result <- cbind(seg, segment)

# Export data to a CSV file
write.csv(result, file = file.choose(new=TRUE), row.names = FALSE) ## Name file
segmentation_result.csv
```