



BUSI4370

ANALYTICS SPECIALIZATIONS AND APPLICATIONS

COURSEWORK 2023

Student ID: 20490911

EXECUTIVE SUMMARY

Customer data analysis has been crucial to the success of any organisation since the emergence of digital technologies. The task at hand was to classify customers into distinct categories based on their purchase behaviour and preferences by analysing customer data. This was done in an effort to learn important market information and advice next marketing initiatives. Companies can get deeper insights into the purchasing patterns and preferences of their consumers by utilising machine learning algorithms, which enables them to design more focused campaigns with greater returns on investment.

This research offers a thorough review of customer segmentation, concentrating on how well it can differentiate between various client segments. A sample of transactional data was used for the study, allowing for the discovery of the crucial traits that set the distinct segments apart. Customer segmentation is achieved through a 5-step process: basic statistical analysis, feature selection, clustering and outcome analysis. This culminates in the creation of pen profiles. With the aid of individualised marketing efforts and a better understanding of their audience, firms can target customers more successfully and increase revenue. This report offers a thorough examination of the customers and a finely defined segmentation of the customers. In addition to the summary and results parts, it also provides critical takeaways and marketing advice. The results of this study provide important light on how businesses might successfully reach their target consumers.

FEATURE SELECTION

Data analysis has been used to divide clients into several categories in order to better understand and serve them. Many characteristics from the data have been extracted through this approach and may be used to define clients and guide marketing choices. These attributes include user **average spending, past purchases, categories of products purchased, and more**. Businesses may improve the decisions they make about how to sell their goods and services by utilising these data.

For merchants to improve customer satisfaction and boost revenues, understanding consumer buying behaviour has become more and more crucial. **Average Spend** is a useful statistic that can assist merchants in identifying their most valued consumers. Retailers can better understand their customers' spending habits and identify individuals with the highest spending scores thanks to this feature, which offers information into client purchase behaviour.

Customers' purchasing habits are a key component of any business's success. This feature offers information on the different things that consumers purchase, how frequently they shop, and how much money they spend on each transaction. With the use of this information, organisations may better understand their target market and develop goods and marketing tactics that speak to them.

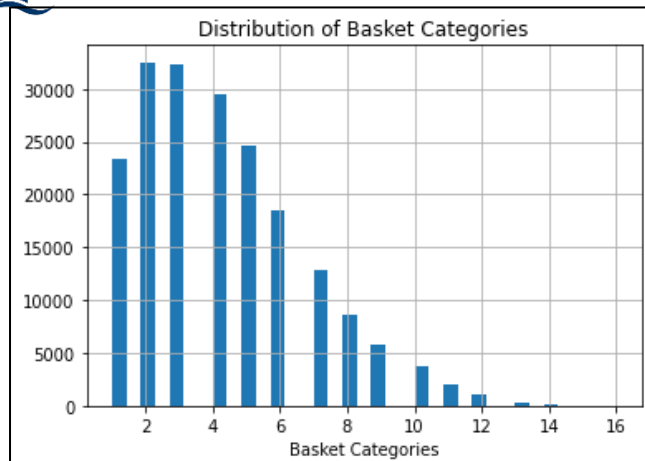


Figure 1: Number of basket categories purchased by consumer

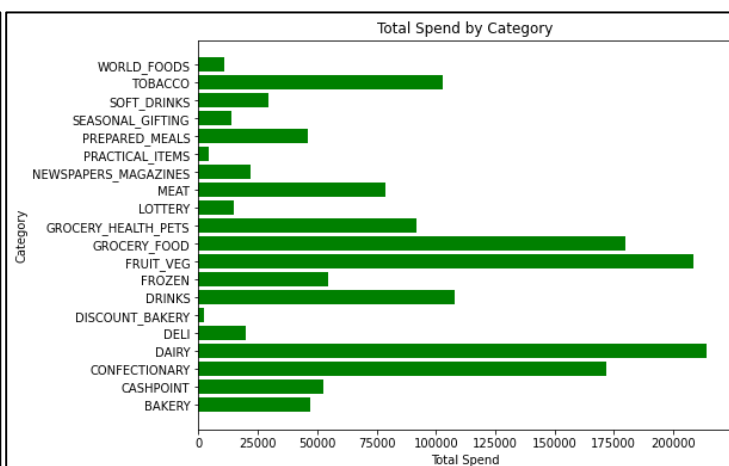


Figure 2: Category of products vs the Total Spend

The graph in Figure 1 shows, the maximum number of consumers purchasing **2 - 4 categories of items**. The graph in Figure 2 shows an analysis of the types of food purchased in a certain area. With careful consideration, it was found that Fruits and Vegetables, Grocery Food, Meat and World Foods can be combined into one category – Grocery Food. This combination allows for more accurate and detailed data collection which can then be used to further understand the purchasing behaviours of consumers in the area. I added four new categories—Alcohol, Bakers, Premade Food and Offer Zone—to the product categories to better organise and manage them. Tobacco, alcohol, and soft drinks were included in the Alcohol category. Dairy and confectionary products were featured in the Bakers category, Prepared meals, Frozen food, and Deli products were included in the Premade Foods category, and lottery, cashpoint and discount bakeries were available in the Offer Zone. This new system enabled me to categorize products more efficiently for customers.

	Premade_Food	Grocery_food	Alcohol	Bakers	Offer_Zone
0	55.98	148.82	50.85	143.69	17.0
1	93.48	80.52	10.76	66.08	0.0
2	75.11	162.52	31.12	132.86	38.0
3	40.78	204.84	124.42	362.06	4.0
4	102.49	100.43	7.23	200.54	0.0
...
2995	156.24	261.97	56.62	163.20	0.0
2996	24.68	39.34	0.67	36.40	30.0
2997	8.77	176.56	29.09	148.39	0.0
2998	17.50	20.09	4.33	35.40	0.0
2999	58.19	209.66	30.56	88.01	235.0

[3000 rows x 5 columns]

Figure 3: Summary of the table

Retailers now have the chance to learn more about consumer behaviour and preferences by using purchase behaviour analysis. Retailers may see trends and develop specialised marketing campaigns and product offers that appeal to their target market by analysing the purchase habits of their customers. They can better serve their consumers' demands and increase ROI as a result.

Successful marketing initiatives depend on an understanding of consumer behaviour and preferences. We have uncovered significant indications that can offer insightful information regarding customer preferences **through analysis of average spending, product categories purchased, and purchase behaviour**. The proper audience may be targeted with the best messages and techniques for maximum impact using the data provided.

Retailers may boost customer loyalty, target, and customise marketing, and better understand client demands by utilising these services.



CUSTOMER BASE SUMMARY

The data's exploratory analysis has yielded some insightful information about the market for the firm. Customers have been shown to be prepared to spend, on average **£ 770** on goods and services, suggesting that there is room for expansion in this market. This study can also assist the business in making defensible choices regarding its marketing plans and product line up.

According to the most recent data, shoppers appear to be concentrating on buying necessities for daily life. As seen above, the most popular **product categories are dairy, grocery foods, confectionery, and fruits and vegetables**. This pattern suggests that consumers are choosing cost- and convenience-effectiveness while making purchases for their essential requirements. Customers' shopping habits are evolving quickly, and many of them are now making purchases from a variety of product categories. Some clients only purchase from one or two categories of items, while others make purchases from **six or more distinct product categories**, showing the broad variation in the number of categories that are purchased.

We can observe that there is a broad variety of client behaviour and preferences when we examine the market considering the features we have designed. Consumers differ in terms of their typical spending, product categories, and purchasing patterns.

The market for the firm appears to be mostly centred on buying daily necessities like groceries and fresh fruit. There is an opportunity to segment the market based on factors including spending propensity and buying behaviour due to the market's diversity and the wide range of customers' purchasing behaviours. The business may maximise its marketing efforts, increase revenue, and foster client loyalty by customising marketing campaigns to certain consumer categories.

Segmentation Methodology

K-means clustering is a potent machine-learning method that effectively divides enormous data sets into more manageable groupings. We can learn a lot about our customers' characteristics—their likes and dislikes—and use that information to create groups for our marketing and sales initiatives. In this research, we'll look at how grouping customers into segments using the K-means clustering technique might provide insightful information about their preferences and activities.

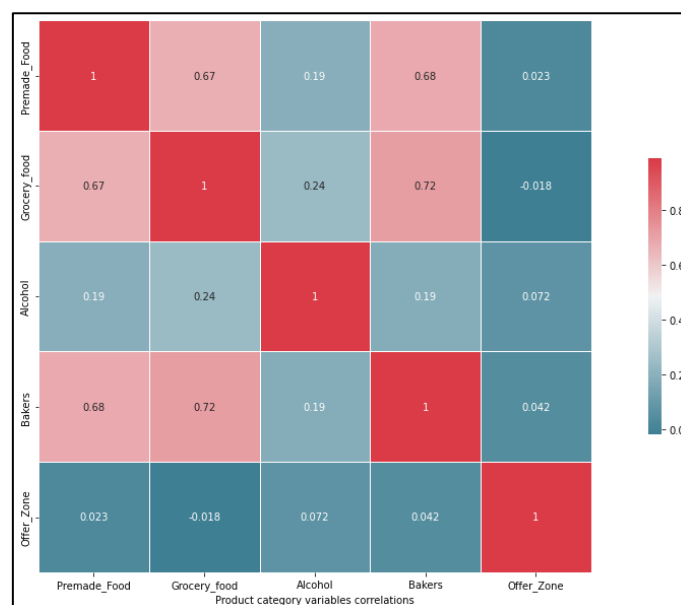


Figure 4: Heatmap of features



This graph in Figure 4 is a heatmap showing the correlations between different variables in the data, which calculates the strength and direction of the relationships between variables. The feature may not be very helpful for summarising clients' spending patterns, according to the association between Bakers, Premade Food, and Grocery Food. Bakers can offer useful insights into client tastes, but it is probably not required to understand purchasing trends. Regrettably, the information utilised to differentiate between various client categories is frequently distorted and contains mistakes. To address this, we have applied a log transform to the data to make it more "normal" and lower the possibility of double counting data that might result in inaccurate segmentation. This not only helps to assure the correctness of our analysis but also helps to obtain accurate findings while using less time and effort.

For feature engineering, we used PCA analysis to draw conclusions. In order to capture the most variety in the data with the fewest number of dimensions feasible, PCA converts the original data into a new collection of variables known as principle components. PCA is an unsupervised machine learning technique. This method allows for the reduction of large, high-dimensional datasets while maintaining the majority of their crucial characteristics. As a result, PCA is a crucial technique for extracting useful insights from complicated datasets.

To create segments, we used the **elbow method**, which is a commonly used technique to infer the number of groups to consider for data clustering. To locate the "elbow" in the graph, we apply K-means clustering to the data for various values of k and visualise the within-cluster sum of squares (WCSS). By locating this location, we can determine the number of clusters that results in the best balance between accuracy and simplicity. By using this technique, companies may properly categorise their customers and adjust their product offers.

The effectiveness of clustering the data points into their designated cluster centres is measured by WCSS. We can observe when the WCSS starts to flatten out and how rapidly it decreases with increasing k by charting the WCSS versus the number of clusters. This aids in figuring out the ideal number of clusters and evaluating the effectiveness of our clustering technique.

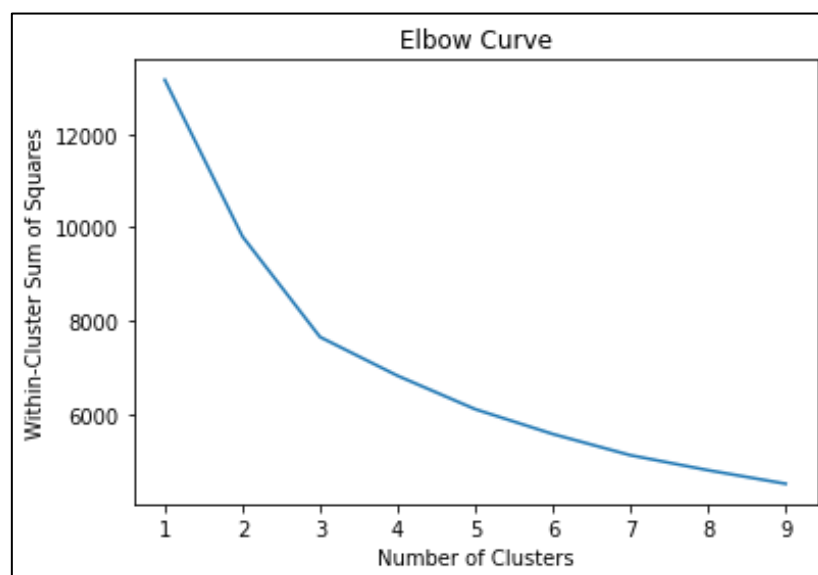


Figure 5: Elbow Curve

According to our research, **the elbow was at k=3**, indicating that three clusters would be the ideal amount to precisely distinguish various groupings. We then **calculated the silhouette score** for the 3-segment solution which resulted in a **value of 0.36**. Nonetheless, **we choose to employ five clusters** to fully comprehend the consumer data and find any unobserved trends. This would help us better represent different client categories and tailor our offerings accordingly.



When working with huge datasets, clustering data points may be a difficult and time-consuming task. Thankfully, the K-means algorithm simplifies this work by combining comparable data points automatically. This technique first initialises clusters, followed by each data point being assigned to the cluster that has the most similar properties. By employing the K-means method, we were able to group together swiftly and reliably the data points based on their average expenditure and the type of things purchased. Gradually, the cluster centres are adjusted based on the average of the data points in each group until they reach a stable state.

The K-means clustering method was effective in classifying customers into distinct groups based on their shopping habits and preferences.

Results

Based on Segmentation Methodology we have created 5 segments of the customers.

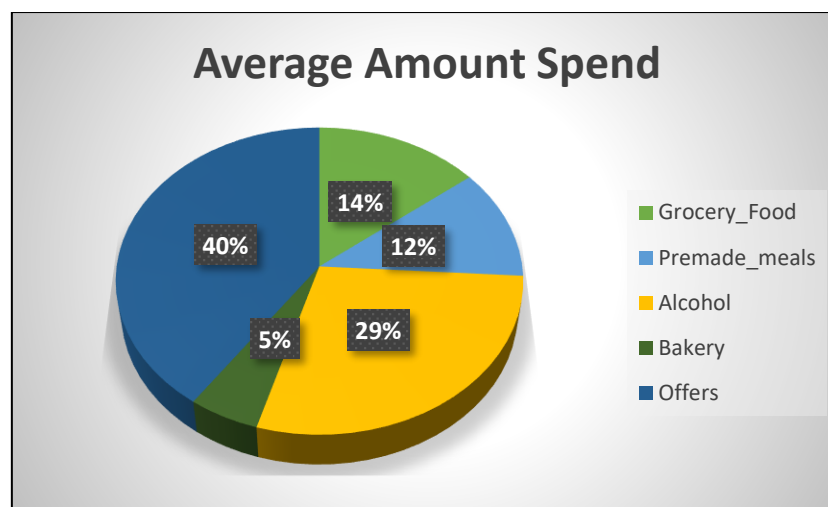


Figure 6: Chart of Average Spend across categories.

Segment 1: Grocery Shoppers

A crucial market for merchants is grocery customers who spend a sizeable amount of their income on fresh produce, pantry essentials, and other food products. These clients typically spend 14% each time they visit and are devoted, regular visitors. Retailers may boost customer happiness and loyalty while also growing their revenues by comprehending the demands of this sector and offering the appropriate goods and services. This portion will be especially intriguing to individuals who are aware of the advantages associated with healthy eating and cooking from scratch. Businesses may begin to develop connections with individuals who respect the quality and freshness of their goods by marketing to consumers who place a high priority on healthy eating and cooking from home.

Segment 2: Offer Zone Shoppers

Customers in this group are strongly driven by deals and discounts and are more likely to buy goods that are on sale or in the offer zone. Businesses may increase their profitability while giving their consumers fantastic prices by focusing on these clients. By offering consumers customised and automatic discounts that can be used right at checkout, these offer zones help customers save money on groceries. The average amount spent by customers on their visit is about 40% with a larger share of that money going towards things in the offer zone. Customers benefit from the time and cost savings while still obtaining the goods they require.



Segment 3: Bakery Lovers

It goes without saying that shoppers love to visit the bakery section. Customers that make a significant amount of their purchases from the bakery, including sweets and dairy goods, make up this market category. These customers frequently seek comfort or pleasure, which they might get in the form of delectable baked products. Businesses can guarantee that they're giving their clients the finest experience by better knowing their requirements and preferences.

They typically spend 5% at the bakery on each trip and spend less in other categories.

Segment 4: Alcohol Enthusiasts

This group of consumers spends a large part of their income on alcoholic beverages and associated products like soft drinks and cigarettes. Businesses may more effectively understand how to appeal to this demographic and give them the items they require by putting the appropriate tactics in place. Due to their significant expenditure on things connected to alcohol, this group of clients distinguishes themselves from the others. With an alcohol feature maximum value of **1125**, this market is likely to be made up of people who like socializing and drinking. Also, they frequently spend less money in other categories, showing that they prioritise buying alcohol over other things.

Segment 5: Pre-made Food Lovers

For buyers who are strapped for time and have little access to cooking supplies, pre-made food items are becoming more and more popular. Customers are offered convenience and variety while also being catered to in terms of their demand for quick and simple meals via ready-to-eat meals, frozen food products, and deli meats. Due to the convenience, they provide as well as the greater availability of such products in supermarkets, this group of customers is turning to pre-made food items more regularly. The fact that this group has the lowest standard deviation of any of the parameters makes it particularly noteworthy. These people are probably busy and value convenience above everything else when it comes to eating. Because they are readily available, pre-made choices are preferred over homemade ones, making them the best option for this group of consumers.

It is known that when it comes to specific product categories, different customer segments have distinctive buying tendencies. We can determine the precise variations in spending across distinct customer categories by employing statistical analysis. Based on their purchasing patterns, consumers have been classified into five main categories, according to the report. While Segment 2 spends more money in the offer zone, Segment 1 is the greatest spender on groceries. Sector 5 spends a lot on pre-processed food. Segment 3 spends the most on bakery items. Finally, Segment 4 spends the most on alcohol items. In today's fiercely competitive industry, clustering analysis has emerged as a crucial tool for retailers. Retailers may better understand their consumers' requirements and preferences by grouping their customers into clusters. As a result, they can customise their marketing efforts to meet the unique requirements and preferences of each consumer cluster, which boosts both customer happiness and sales.

Summary

To understand consumers' wants and habits, it is important to segment them. The K-means clustering technique is a useful tool for categorising consumer groups based on their distinctive traits. Based on our customers' typical spending and buying habits, we were able to establish **five separate groups using clustering analysis**, which gave us insightful information about our customer base. With the help of this data, we can focus our marketing efforts on the most lucrative clients and create very effective campaigns. Marketers may now use data to generate more successful ads by better understanding their target



audiences. Marketers may quickly identify consumer categories and understand each group's distinct buying preferences with the use of data summaries and pen profiles. For instance, "Premade Food Lovers" are more likely than other groups to spend more on pre-packaged meals, allowing marketers to adjust their ads appropriately. Alcohol enthusiasts, on the other hand, are more inclined to buy pricey and speciality alcoholic drinks since their average expenditure is greater; this enables them to partake in a wide range of drinks while still being frugal. As a result, both alcohol manufacturers and merchants find alcohol enthusiasts to be a lucrative market.

It has been determined after a thorough study that **two client groups—alcohol enthusiasts and grocery buyers—are of the highest significance for this business**. These clients provide organisations with a fantastic opportunity to boost sales and foster brand loyalty. Companies may better customise their marketing strategies and appeal to these groups by concentrating on these two markets to realise their full potential. Providing promotional offers to customers is a successful strategy to increase sales and draw in new clients. For a retailer, comprehending the various client groups is essential to developing winning offers. While the latter sector typically tends to be more focused on low prices or discounts, the former section reflects the usual client who searches for value when making purchasing selections. A retailer may design specialised promotional offers that appeal to each group's demands and enhance its sales potential by comprehending these various divisions.

Key take-home points:

- By segmenting a client base using clustering analysis, organisations may create personalised marketing plans that cater to the requirements of each consumer group. This technique proved to be a priceless tool for developing successful advertising.
- In order to create marketing efforts that work, you must thoroughly understand your target audience. Understanding different consumer groups' requirements and preferences can help you develop tactics that will appeal to them and are more likely to result in sales.

Suggestions for further analysis:

Retailers should go a step further and gather extra information on client demographics, lifestyle preferences, and purchase patterns to make sure our marketing efforts are as successful as possible. This will make it possible for them to improve their segmentation and develop specialised marketing plans. Surveys, focus groups, and social media monitoring are ways to get more in-depth information on client requirements and wants. This aids in the consumers' understanding so they can better cater a product or service to suit their requirements. Also, to develop a more comprehensive understanding of market dynamics and consumer behaviour, the organisation may think about combining other data sources, such as market trends.

Big data and advanced analytics have made it possible for organisations to access a range of external data sources to get insight into consumer behaviour, market dynamics, and competitive landscapes. Companies may gain a deeper grasp of their existing markets and future possibilities by adding industry trends and other external contextual elements into their decision-making process. This offers the opportunity for greater understanding that result in wiser strategic choices. Furthermore, it would be beneficial for retailers to regularly gather feedback from their customers to identify and address any shortcomings in their business.