# Table of Contents



# Table of Figures

# 1. Introduction:

Supermarket chains generate tons of data every day that needs to be transformed into useful interpretable information to help in the decision-making process and to devise strategies that ensure their success and survival in the market. Numerous techniques are hence used to assist in the process of the generation of valuable knowledge and trends. In this report, the data set of a supermarket chain has been analysed to recommend future strategies for their existing products, customers, and potential customers. The report will present the following sections: knowledge acquisition was first conducted to extract relevant information from existing literature to help in the analysis of the supermarket chain's data. This was followed by the development of models using RapidMiner. Two techniques were used to create the models, the results of which are demonstrated in this report along with explanations and justifications for each model developed. Future strategies were then recommended to help the supermarket chain improve its performance. Finally, the report is concluded with a discussion of the results obtained and a comparison of those results with the literature gathered in the knowledge acquisition process.

# 2. Knowledge Acquisition:

The success and survival of any business in the market depends on its ability to make business decisions that meet and satisfy its customers' demands (Kauffmann et al. 2019, p. 1). According to Mitchell and Pavur (2002, p. 58), interpreting and modifying data collected into knowledge is of great importance in order to apply it to business operations for its decision-making processes. This is further supported by Baierle et al. (2019, pp. 17-18) in their research, where it is stated that reliably and efficiently sorting through data, plays a key role in aiding businesses to understand where changes need to be implemented in their current processes.

## 2.1. Importance of Customer Satisfaction in Relation to Customer Loyalty:

Businesses make huge investments in research about the markets that they operate in and perform analytical customer relationship management (aCRM) in order to have a better understanding of the customers' demands and preferences (Kauffmann et al. 2019, p. 1; Ranjan & Bhatnagar 2011, pp. 132-136). Deshwal (2016, p. 944) claims that there have been many models constructed for customer segmentation with emphasis on gender from which it has been established that males and females have different experiences and satisfaction levels when it comes to shopping. Understanding those experiences helps in interpreting the customers' demands and this tends to improve business management and decision-making, and thus, ensures customer loyalty (Kauffmann et al. 2019, p. 1; Ranjan & Bhatnagar 2011, pp. 132-136). Research conducted by Wicker (2016, p. 2) also highlights the importance of customer loyalty and retention to increase the success of a supermarket, stating that a customer who is loyal to a supermarket is more likely to make multiple trips per week and spend more on average per trip.

## 2.2. Strategies for Stock Management:

The results generated from the market research conducted help retailers in the process of decision-making regarding the products that should be shelved and the strategies they should implement that would be well-received by customers and generate the profit targeted by the business (Kauffmann et al. 2019, pp. 1-2; Lejeune 2001, p. 375). This is further supported by Ghochani et al. (2013, pp. 59-60) who stated that market-based research enables businesses to make predictions regarding various elements that aid decision-making. An example of this would be predicting consumer sales to help in determining the quantity of stock needed for each product category, which in turn allows for more accurate planning of the investments required for stock management (Ghochani et al. 2013, p. 60).

## 2.3. Product Placements in Supermarkets:

Product Placement strategies are another important element that can be successfully determined through market research, they can hugely influence the customers' purchasing decisions and help maximize sales (Kauffmann et al 2019, pp. 1-2). Dhar, Hoch, and Kumar (2001, p. 179), as well as Neff (2008) emphasized the importance of instore display in drawing the attention of customers to certain products which can be more effective than common promotional schemes, such as price cuts and special offers. According to Han (2018, p. 1), each display has its importance in gaining the customers' attention and each display's impact only lasts as long as the customer is within effective range of the display. However, on the other hand several surveys conducted by Inman, Winer, and Ferraro (2009, p. 21) have shown that dramatically increasing the shelf space for any product without changing its price or location can comparatively impact product sales to a greater extent where it was observed that the sales of the product experienced an increase of 19% to 39% with more shelf space.

### 2.4. Customer Response to Price Fluctuations:

Finally, another important strategy that businesses implement is the effective application of promotional strategies such as discounts and offers. The decisions regarding such strategies can only be successfully made via prior customer and market research. According to Stead et al. (2017, p. 525), businesses that analyse the customers' purchasing history are able to devise effective promotion strategies to merchandize specific products to increase their sales and maintain the customers' loyalty rates (Stead et al. 2017, p. 525). Chen, Choubey, and Singh (2021), support the effectiveness of this strategy and suggest that customers do tend to buy discounted products even if they have high price sensitivity and low predilections.

## 3. Data Analysis Techniques:

### 3.1. Technique 1 – Classification Using Decision Trees:

#### *3.1.1. The Models:*

##### *3.1.1.1 Model 1 – Loyalty and Churn Based on Lifestyle:*

For the models presented below, a customer has been defined as either *loyal* or *churn*. *Loyal* has been denoted as a customer who has shopped more than once at a particular supermarket, and *churn* is described as a customer who shopped once and did not return. To aid in the business decision-making, there will be a focus on highlighting the different aspects that are more likely to lead to customer loyalty or churn and then solutions can be made based on the data presented.

Firstly, consider the city with the highest loyalty rate. In Figure 1 the red bar represents loyalty, and the blue bar represents churn; the thicker the bar indicates there is a larger number of customers that fall into the category. The decision tree made predictions dictating whether a customer who shopped in a particular city would be more likely to stay loyal or churn. In this tree, the churn predictions are 77% accurate and the loyalty predictions are 62% accurate.

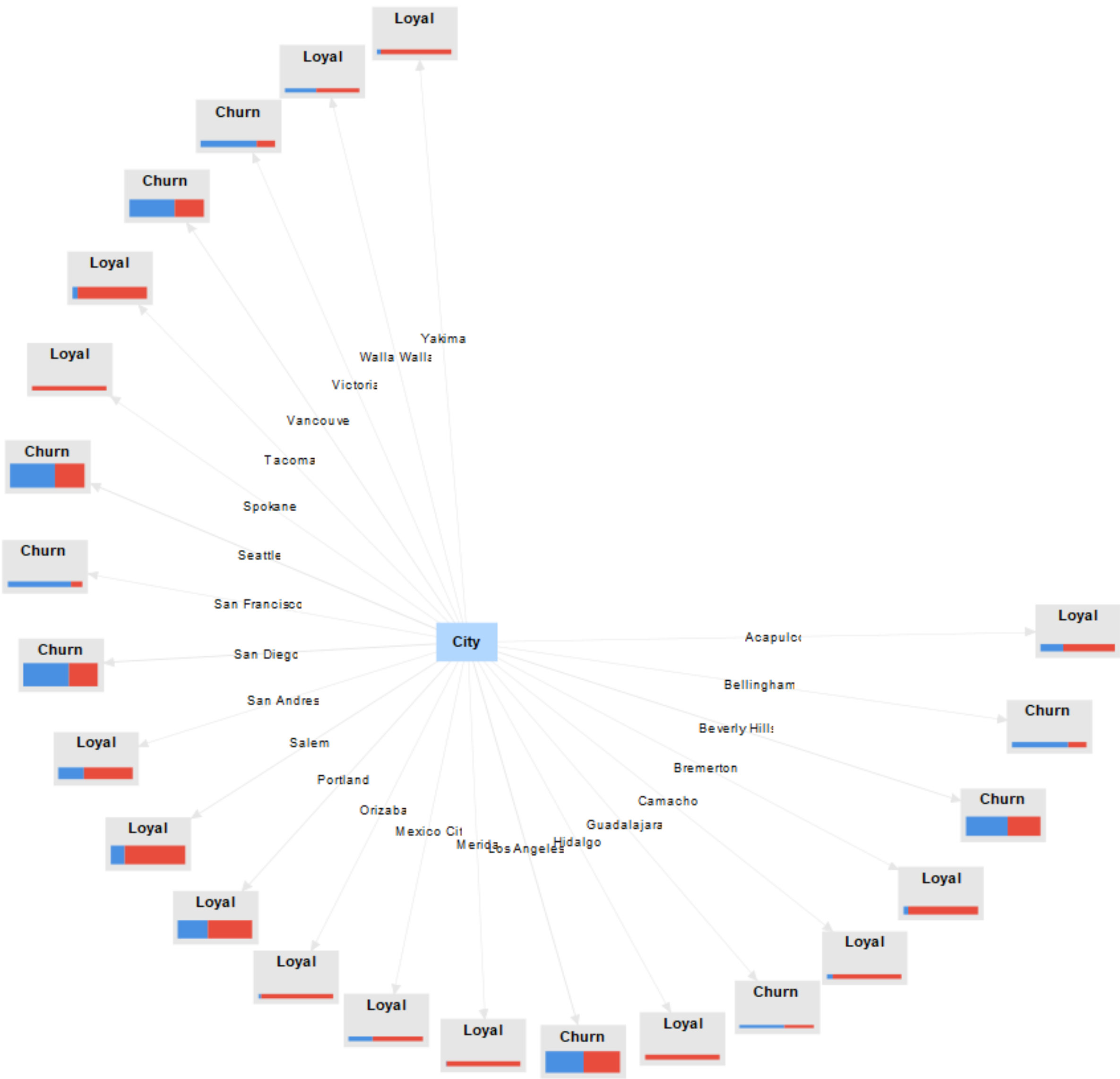


*Figure 1: Predicted loyalty and churn from each city*

The key information from the decision tree is to obtain which city has a higher number of loyal customers, then a further in-depth analysis can be conducted on those cities to discover the trends among customers that lead to higher loyalty. From Figure 1, it can be seen that Hidalgo, Merida, Yakima, Camacho, San Andres, Tacoma, Orizaba, Salem, Bremerton, and Acapulco have the highest and most overwhelming customer loyalty. The cities with the highest churn are Vancouver, Victoria, San Francisco, Bellingham, Seattle, and San Diego. By comparing the differences between a city with a high churn rate against a city with a high loyalty, then it is possible to identify the key difference. This helps in manipulating the information and turning into customer retaining strategies.

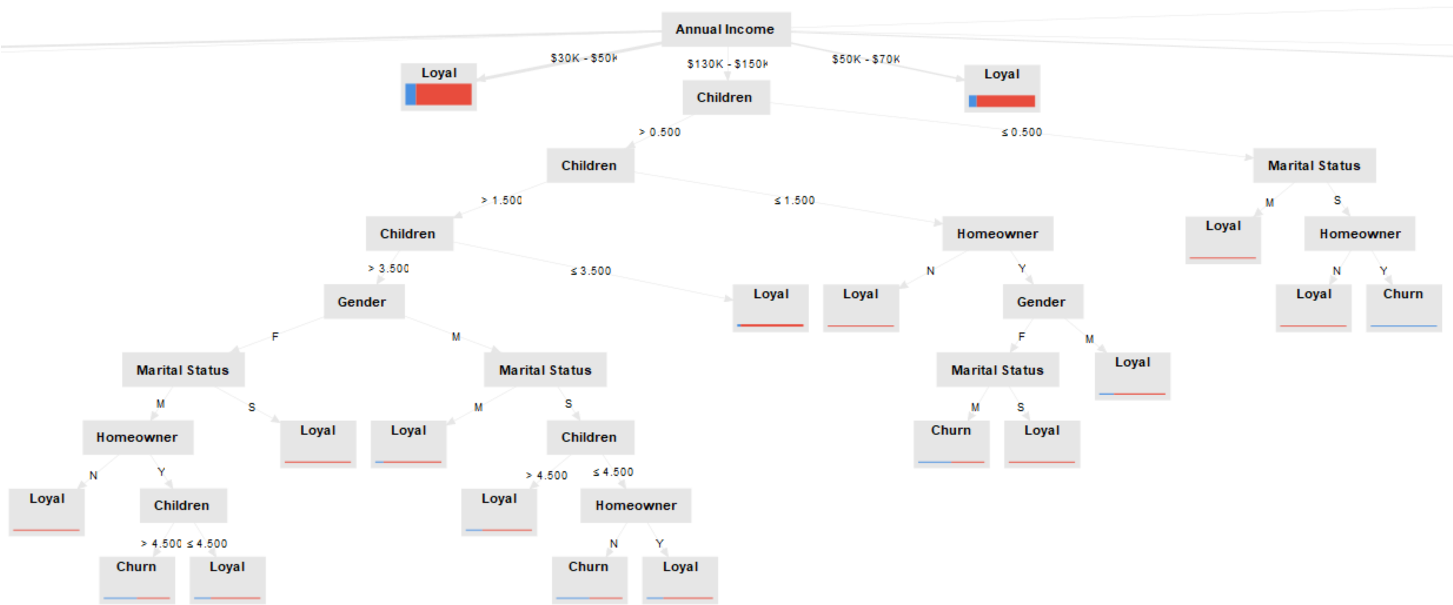


*Figure 2: Lifestyle factors from the highest loyalty cities that may lead to churn*

A model, Figure 2, was generated with 82% accuracy that showed different lifestyle factors that may lead to churn or loyalty. The model compares annual income, children, homeownership, marital status, and gender to consider what may lead a customer to be more likely to stay loyal. The model considered 1785 customers from the cities that were identified to have the highest loyalty and within Figure 1, it is duly noted that those who earn between $30 - $70K a year are significantly more likely to stay loyal. Customers within that annual income bracket compose 48.3% of all customers within the example set. Comparatively, using the same model and exchanging highly loyal cities for highly churn cities from Figure 1, within Figure 3, there was an abundance of different aspects of a customer's lifestyle that might lead them to churn. The sample set is composed of 1910 customers and has notable flags for churn. For example, a woman who earns $30K-$50K and is a homeowner in a high churn city is significantly more likely to churn, this leaf contains 10% of all customers within the example set. Similarly, a person who earns $50K-$70K, is not a homeowner, and has 3 or fewer children is also most likely to churn. Although Figure 3 has an accuracy of 60% (6 in 10 customers), it can be used as an indicator for audiences that may need targeted advertising.

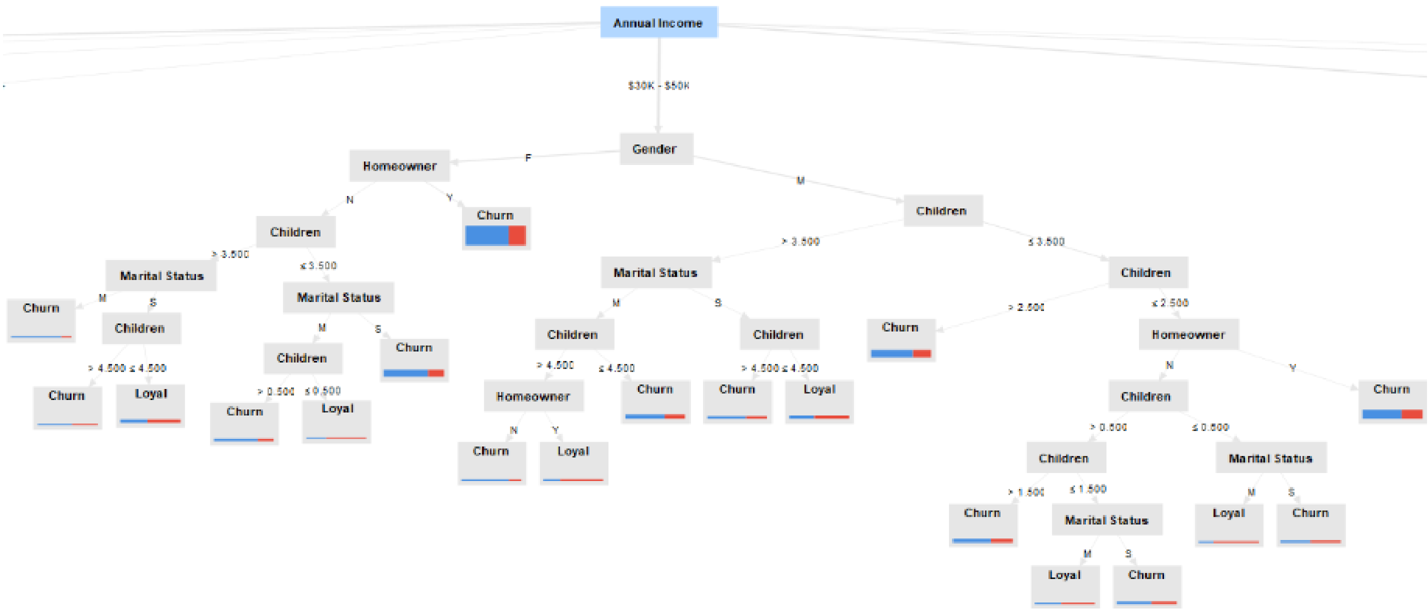


***Figure 3:*** *Lifestyle factors from the highest churn cities that may lead to churn*

When considering churn and loyalty more accurate results are found when comparing similar groups. For example, when considering all cities that are 'in the middle' (there is not a significant churn or loyalty rate) the results tended to be less accurate as the trees became overfitted with too much data; meaning that there were no distinct trends to predict. Therefore, for more accurate results, it is better to consider the two trees and their respective differences.

### *3.1.1.2 Model 2 – Loyalty and Churn Based on Sales Data:*

When considering a customer's loyalty or churn, there is also the impact a shopping experience can have that can influence a customer's desire to return to a supermarket or to churn. In the model below, Figure 4, the model showcases how the amount spent, units sold, and product family can sway a customer's loyalty. The model has an 83% accuracy, dictating that 8 out of every 10 people will be more likely to stay loyal if they purchase 2 or more products and spend less than $47.26 in any product family.

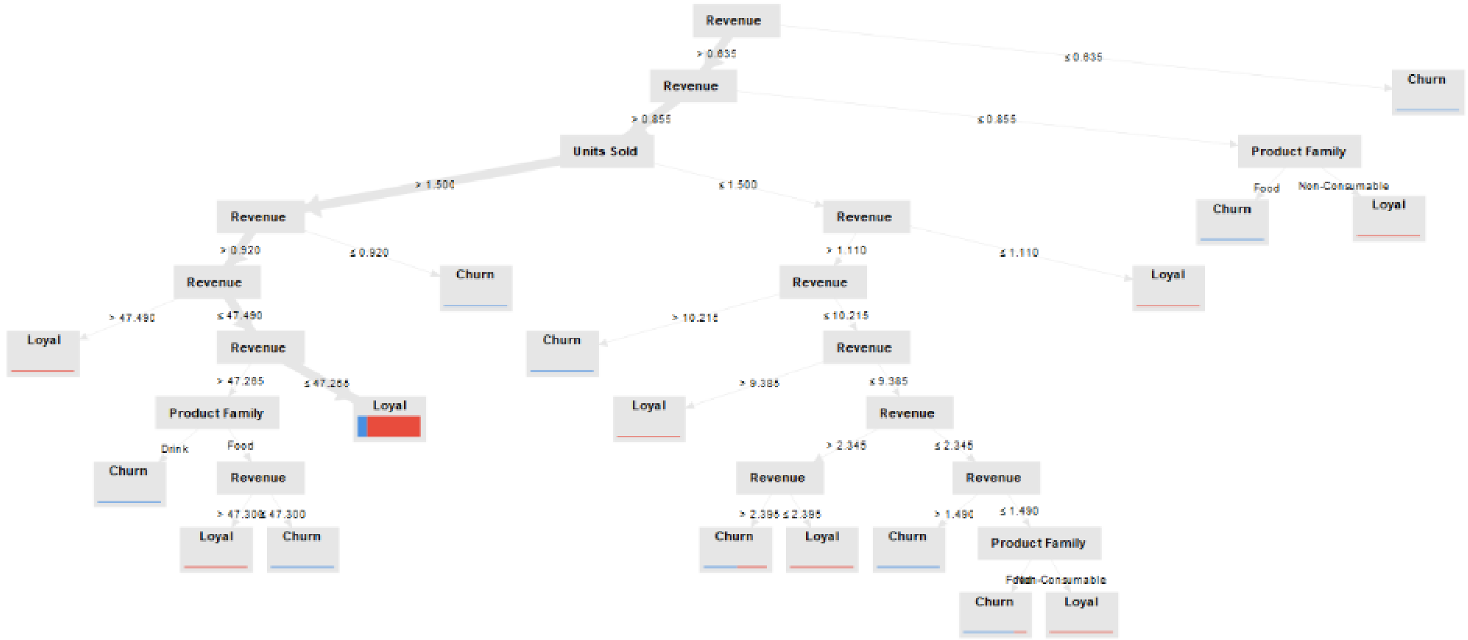


***Figure 4:*** *How revenue, units sold, and product family during a transaction may affect churn and loyalty*

Another model, Figure 5, considers the units sold in a particular product department and indicates, similar to Figure 4, that the more items a customer purchases, the more likely they are to remain loyal. The model has an accuracy of 83%, meaning that it would be true for 8 out of every 10 customers. When a customer purchases only one item they are significantly more likely to churn, if they purchase between two and seven products then they are highly likely to remain loyal. This reinforces the fact that a customer who is loyal is more likely to purchase more units during a transaction.

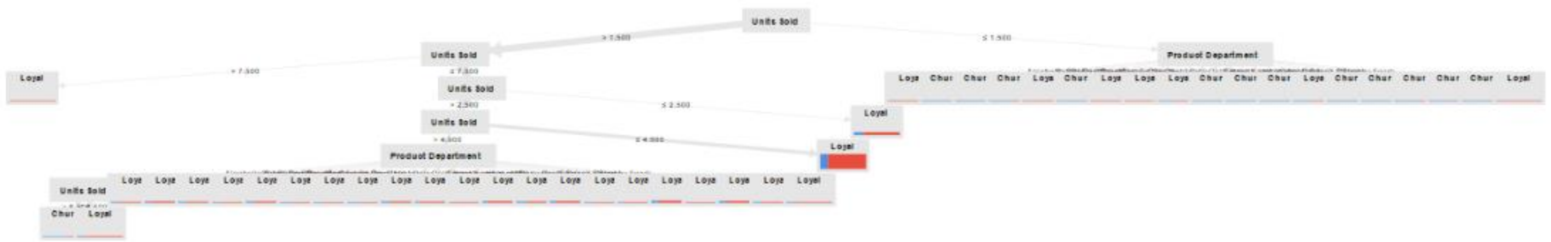


***Figure 5:*** *Likelihood of churn predicted by units sold and product department*

### 3.1.2. Justification:

Decision trees are widely used for an array of different purposes; however, they have a significant appeal when predicting what may affect trends or what particular changes may impact a business. Decision trees, as the name suggests, are a 'tree-like' system that branches into different options, much like saying "if this, then that" multiple times until an outcome has been determined for a particular set of data.

Decision trees are simplistic in nature, often with minimal guidance needed they can accurately predict outcomes depending on a particular branch or leaf. When considering what may retain a customer to stay loyal to a business there are many factors involved and decision trees allow an interpreter to visualise and consider the different aspects that may lead a potential customer to end up at a certain branch. Decision trees are effective at analysing discrete or continuous data which aids in the ability to understand the dataset given. As the variables that will be used within this report are of discrete nature, they will work sufficiently. Due to their ease of use, there is no 'clean up' required to interpret the data gained from a decision tree; they can be read as soon as they are produced.

Decision trees and other data mining methods have been used frequently when considering data handling for businesses such as supermarkets. By providing multiple visual options in a decision tree, the data is user-friendly and accessible.

## 3.2. Technique 2 – Clustering Using K-means:

### *3.2.1. The Models:*

#### ***3.2.1.1 Model 1 – Demographical Customer Segmentation:***

Identifying and having awareness of the target customers' demographics is crucial information for any business as it helps in understanding the clients' preferences. In this case study, a model representing the demographical customer segmentation of the supermarket chain's 5404 customers was produced to analyse their demographics. The clients were distributed among five clusters according to their annual income, gender, marital status, and homeownership. Figure 6 demonstrates the results of this model from which it can be deduced that most of the supermarket chain's customers are single homeowners, regardless of gender, with an annual income of $30K - $50K. This signifies the suitability of the products that are sold by the supermarket and their prices for both males and females in the middle class as they satisfy their needs and demands. On the other hand, those with an annual income of $90K+ cannot be considered part of the supermarket chain's target customers as they represent a small portion of the business's total number of customers, reaching no more than 0.5% of them.


***Figure 6:*** *Demographical Customer Segmentation*

### *3.2.1.2 Model 2 – Geographic Sales Categorization:*

Using the same settings for the k-means clustering operator in Model 1 (Figure 6), another model, for which the results were divided into several figures as shown below, was produced to determine the geographical locations and the number of purchases made by each cluster in Figure 6.

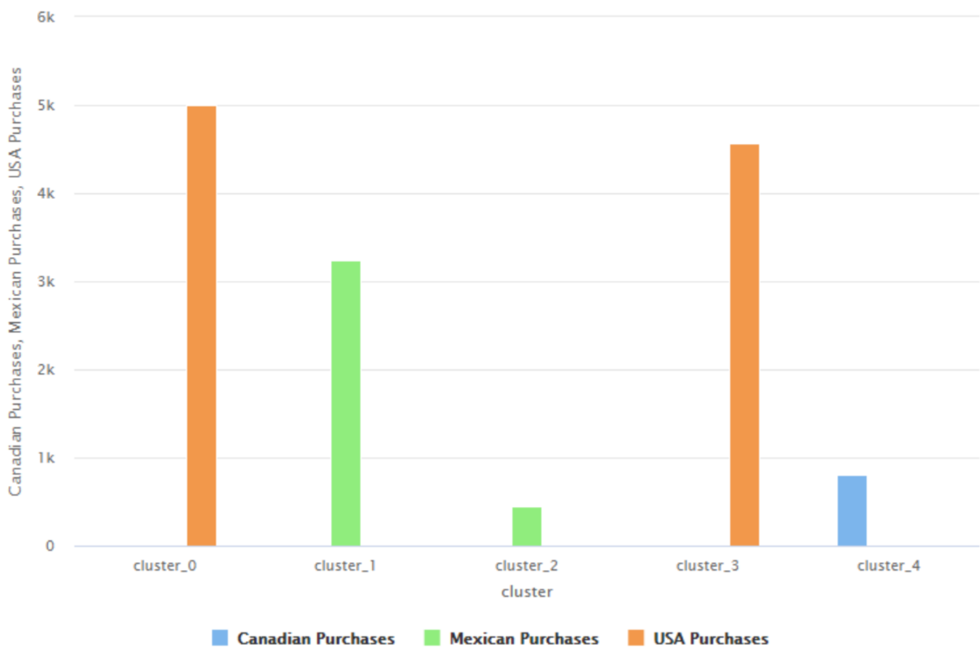


***Figure 7:*** *The Number of Purchases of Each Country*

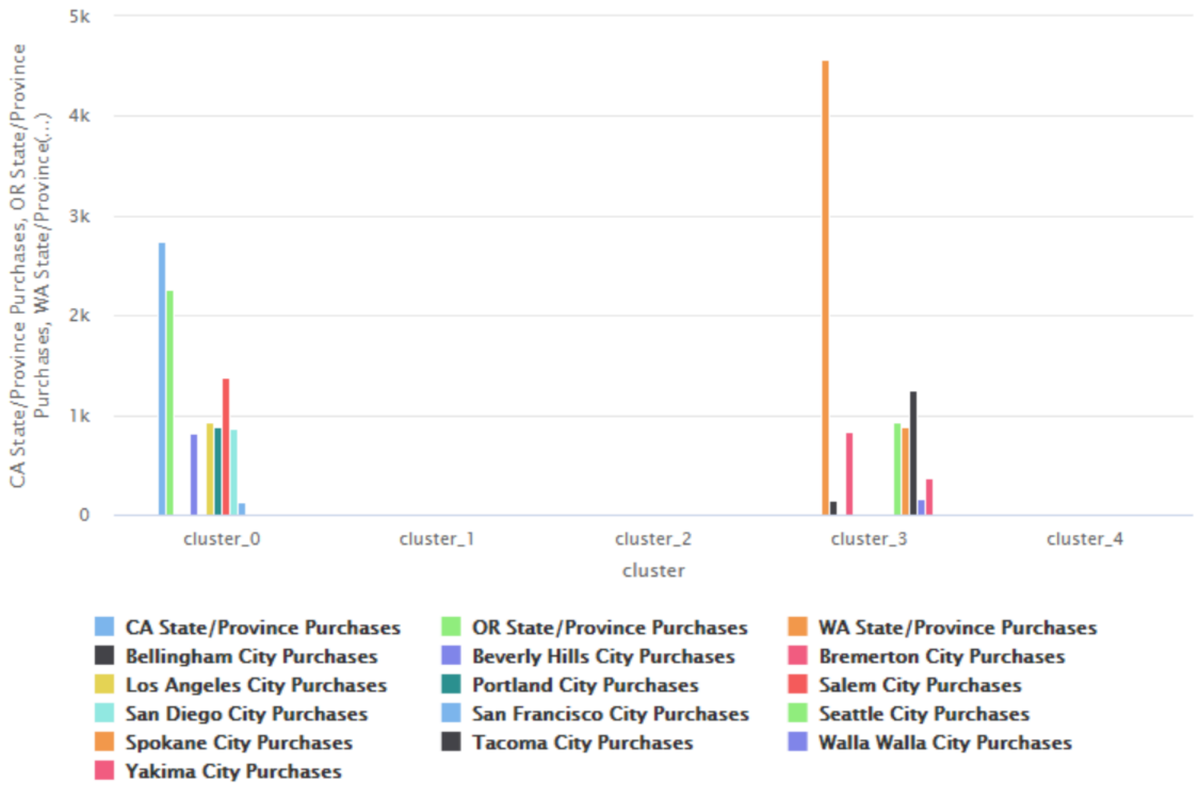


***Figure 8:*** *The Number of Purchases of the American States and Cities*

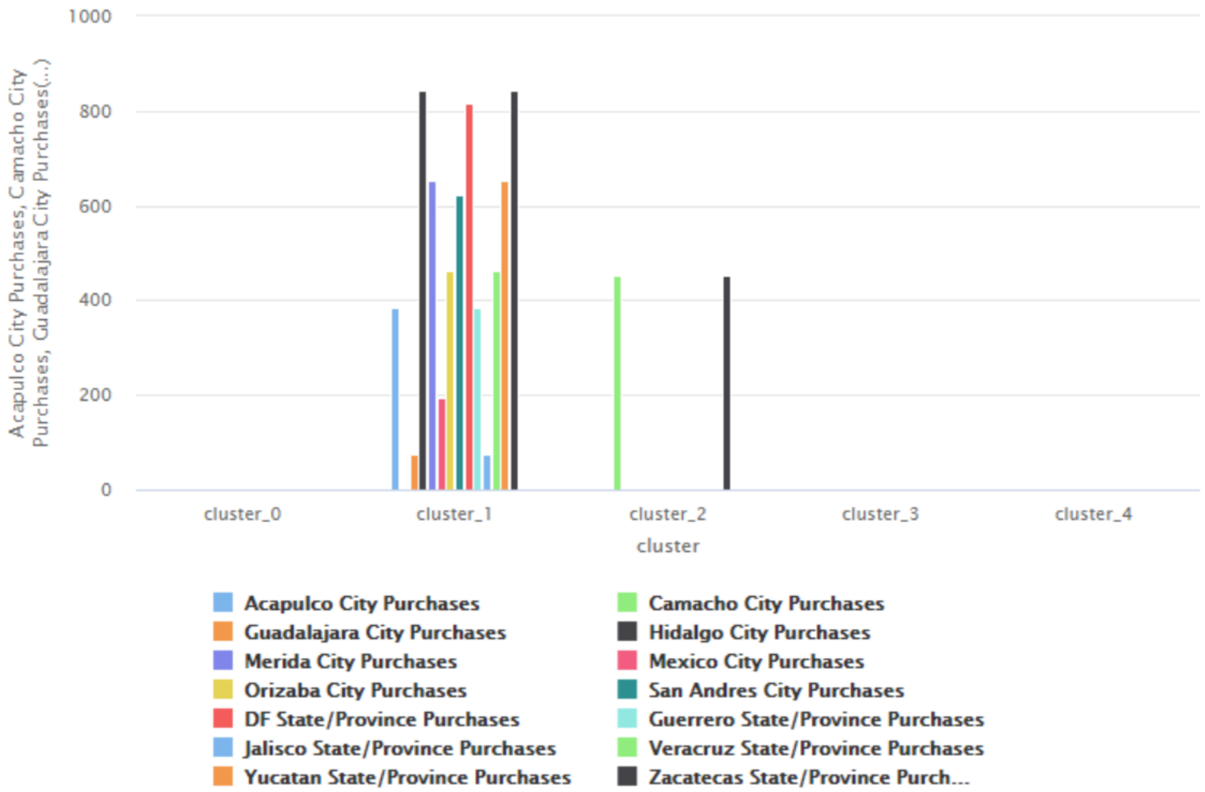


***Figure 9:*** *The Number of Purchases of the Mexican States and Cities*

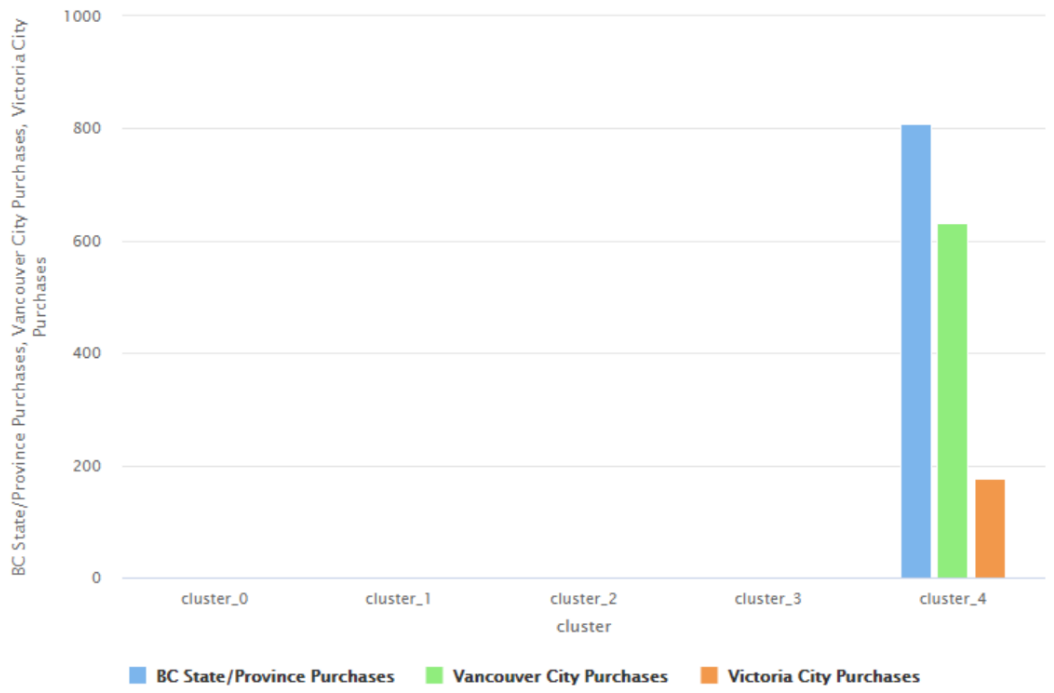


***Figure 10:*** *The Number of Purchases of the Canadian State and Cities*

According to Figure 7, cluster_0 and cluster_3 consist of only American purchases, which are the highest compared to the Canadian and the Mexican ones, reaching about 10K purchases combined. However, Figure 7 suggests that cluster_3 does not have the highest number of purchases indicating that it might have a higher customer churn rate compared to cluster_0.

Figure 7 also illustrates cluster_1 and cluster_2 representing the Mexican purchases with cluster_1 having the third-highest purchases and cluster_2 having the lowest purchases among all five clusters. This indicates that there are many churn customers in that cluster and the reason for this might be because of its expensive prices. However, it can be observed that all the Mexican purchases combined are still higher than the Canadian ones which are represented by cluster_4. This is because the supermarket chain has 13 American branches across three states (refer to Figures 8), eight Mexican stores in six states (refer to Figures 9), and only two Canadian shops in one state (refer to Figure 10). This implies that the American branches generate the highest revenue followed by the Mexican ones and the Canadian stores come in last.

#### *3.2.1.3 Model 3 – Product Popularity Categorization:*

The last model created for this data set using k-means is for determining and comparing the popularity of product categories for each cluster by summing up their number of purchases. The figure below shows this in the form of a chart:

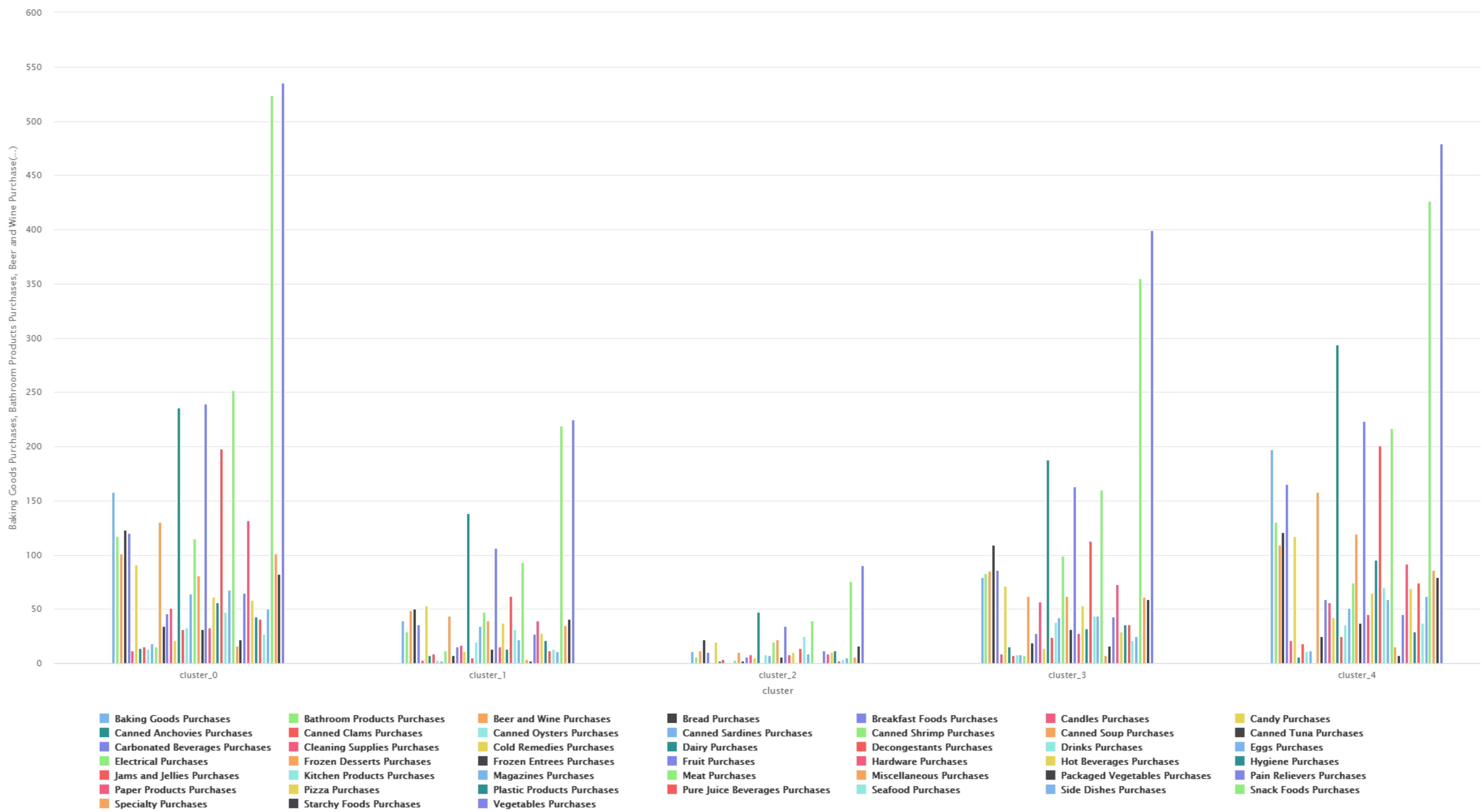


***Figure 11:*** *Product Categories Comparison*

As shown in Figure 11, the most popular product category for all clusters is Vegetables, as they are used in everyday cooking regardless of the customers' dietary requirements. On the other hand, each cluster has its own least popular product category. Cluster_0's is Candles. For cluster_1, it is Packaged Vegetables and Canned Sardines. For cluser_2, Candles, Canned Oysters, Canned Sardines, Decongestants, and Miscellaneous had zero purchases. The least number of purchases in cluster_3 belong to the Canned Clams and Canned Shrimp, with Canned Shrimp being the least popular product category for customers segmented into cluster_4. However, it can also be observed that all these products are in the 5% least popular product categories in all clusters, which indicates that the supermarket chain's management needs to assess the reasons behind this percentage. The causes could be a result of the high prices and absence of sales, poor packaging and product size, inaccurate product placements, and off-target advertisements.

### *3.2.1. Justification:*

According to Cammack et al. (2008), K-means clustering is an algorithm designed for grouping similar data by constructing user-defined clusters. It is an unsupervised clustering technique used by many businesses in different industries due to its numerous advantages making it an ideal segmentation technique for many businesses (HolyPython 2021). It is known to be a fast, easy-to-use and interpret technique with high performance that does not need much guidance or labelled data (Google Developers 2021; HolyPython 2021). It can also easily handle large data sets and is very flexible to the addition of new information (Google Developers 2021; HolyPython 2021). However, there is no specific method to determine the user-specified number of clusters, the K value, hence determining the accuracy is very difficult (Trevino 2016).

K-means segments the businesses' data and produces new valuable information about their customers' demographics and purchase rates. This helps in understanding the clients' demands and in the categorization of the products (Model 1 and Model 3). As a result, it is a cost-effective method to make decisions about the products, their prices, and advertisements, which can increase the success rate of products, decrease the customer churn rate, and assist in building robust relationships between the businesses and their clients (Hubbard 2019; RingCentral 2022; Sagar 2019). It also helps in the identification of the differences between the stores' rates of purchases; thus, details about the success rates of the different stores can be perceived and compared (Model 2).

## 4. Results and Discussion:

The two main techniques implemented to analyse the supermarket chain's data in this report were classification using decision trees and k-means clustering. RapidMiner and its operators were used to retrieve, filter, and change the forms of the data to produce decision trees and clusters of data to predict the rate of churn and loyalty of customers, segment them demographically, categorize sales geographically, and categorize products according to their popularity.

Cities that were shown to have customers with the highest loyalty were Hidalgo, Merida, Yakima, Camacho, San Andres, Tacoma, Orizaba, Salem, Bremerton, and Acapulco. Whereas the cities with the highest churn are Vancouver, Victoria, San Francisco, Bellingham, Seattle, and San Diego. It can be seen that although America has the greatest number of stores, only the cities Yakima, Tacoma, Salem and Bremerton have a high number of loyal customers. In comparison Mexico has a higher number of cities with loyal customers namely Hidalgo, Merida, San Andres, Orizaba. The difference in store loyalty could be store specific which has been supported by multiple studies that have suggested that store attributes significantly affect store choice and loyalty (Solgaard & Hansen, 2003; Chang & Tu, 2005; Sinha & Banerjee 2004) which in turn influence the sales rate. Although the American stores have the highest number of purchases and subsequently, the highest revenue generated, some of them have lower purchase numbers than the others, indicating that they might have a higher number of churn customers. This might be because products are being sold at higher prices in those branches. This has been supported by previous studies which prove that price is an important stimulus of customer satisfaction (Hunneman et al. 2015) and change in satisfaction level has a critical relationship to store loyalty (Bhat & Singh, 2017). In terms of generated income, Mexico has the second highest while Canada has the least. In a similar fashion to America, Mexico has two times less units sold than Canada, although the generated income is almost the same.

The majority of the supermarket's consumers include those who have an annual income between $30K and $70K, i.e., middle class and are single homeowners irrespective of their gender. Study conducted by Singh & Sao (2015, pp 45-53) supports that gender has no impact on purchasing for customers of unorganized retail. Only a handful of customers fall into the category of having remuneration greater than 90K. It has also been noticed from the decision tree classification that lower income customers are most likely to churn. Furthermore, another set of results obtained based on customer purchases depict that higher the units per transaction of consumers imply they are more likely to be loyal. The above result is also endorsed by previous research that implies customers with more income tend to purchase more (Anic & Radas, 2006). Correspondingly, it was also noticed that if the number of purchased goods is between two to seven and less than $47.26 is spent in any category, they are more likely to be loyal whereas if only one product is purchased, they are more likely to churn. Lifestyle factors such as marital status, number of children were also found to be reasons that contribute to customer churn. However, there was no distinct trends predicted for cities that could not be classified as significantly loyal or churn. Contrastingly, a study conducted by Rahman, Md Chand, Gupta, and Kumar (2021, pp. 1-18) suggests that store loyalty has no significant difference among single and married people. However, the results, i.e., lifestyle factors like marital status, number of children, and relationship with store loyalty, are consistent with studies conducted by Sinha et al., as well as Vasudeva and Chawla (2002; 2019) which further support the finding that store loyalty is subject to both store-related as well as shopper-related variables.

Candles, canned oysters, canned clams, canned shrimp, canned Sardines, decongestants, and packaged vegetables were observed to belong to the 5% least popular product categories in all clusters modelled. On the other hand, vegetables were found to be the most popular product category. Results obtained through market research regarding the popularity of products enable supermarkets to find the reasons for lower sales and to determine the required stocks necessary for each category of products as noted previously (Kauffmann et al. 2019, pp. 1-2; Lejeune 2001, p. 375).

## 5. Future Strategies Recommended:

After the analysis and creation of models using the data set provided by the supermarket chain, a list of recommended future strategies has been assembled to aid the business in strategizing their product and customer plans in order to attract new clients and maintain their customer base:

- ***Advertisement Campaigns:***

K-means can be used to perform various types of customer segmentation that can help in understanding their preferences and purchase behaviour. The evaluation of this data results in the assembly of advertisement campaigns that are customized to attract the target customers.

- ***Discounts and Offers:***

The assessment of the popularity of products helps in determining their sales. The least popular products that have little to no purchases can have frequent discounts or offers, such as buy one get one free and buy two for a lower price, to increase their popularity and reduce their stocks. Popular products can have seasonal or monthly sales in order to keep their high rate of purchases. Discounts can also be created according to the customers' ethnicities and religions.

- ***Distribution Strategies:***

With the help of k-means, the supermarket chain can produce distribution strategies that depend on understanding the customers' demands, the purchase rates, and the popularity of the products. This increases the target clients' accessibility to the products, and hence increases the revenue and the success rates of the stores with low performance and purchases.

- ***Feedback and Product Reviews:***

Collecting data from customers about their opinions helps in generating data that can be used to construct marketing strategies, distribution strategies, and product placement strategies.

- ***Increasing Units Per Transaction:***

This technique highlighted that a customer is more likely to be loyal if they have a higher UPT (units per transaction). A strategy recommendation is to increase the customers basket size, which can be done in a multitude of ways. Targeted product placement near point-of-sale systems, deals that include buying more than one item (such as "buy one get one free") and encouraging salesclerks to ask prompting questions (such as "We currently have a deal of dishwashing liquid, if you buy two you get one half price").

- *Product Pricing:*

The models comparing customer lifestyle can give a good indication of what customers may lead to churn, and by having an understanding of their needs, decisions can be made regarding advertising or producing sales and deals that may apply to specific customer bases. For example, if it is noted that a customer is more likely to churn if they are within a lower income bracket, it may be a worthwhile opportunity to supply lower-cost items.

- *Sales Prediction:*

Having knowledge of the customers' behavioural purchases and the number of units sold of each product category can help in predicting the number of products that need to be ordered from the suppliers. This is particularly helpful with perishable food that has short shelf life.

- *Successful Products and Marketing:*

By assessing the sales trends of customers there is an indication of what leads to a positive customer experience and what may lead to a negative experience. If a product is leaving a lasting negative impression that leads a customer to churn, then it would be a worthwhile measure to opt to sell a different product. However, if there is a product that often sells to loyal customers, a marketing campaign can be employed to promote that particular product as it may be more likely to lead to a positive experience. For example, if customers are purchasing a dishwashing liquid and having a negative experience with the product they may not want to return, however, if they purchase a different brand and have a positive experience, they may be more likely to remain loyal. These distinctions can be made and decisions regarding stock and marketing can be applied.

- *Using Team Mentoring:*

Comparing the loyalty and churn of various cities allows business managers to gain insight into what stores are more successful than others. Decisions can be made regarding these specific stores to increase their success, such as moving managers from successful stores and having them mentor teams from stores with high churn rates.

## 6. Conclusion:

This report demonstrated the importance of analysing data to develop business strategies for a supermarket chain. Models have been designed using the provided data set to obtain useful pieces of information. We first acquired key information from decision trees to determine which cities have a high number of loyal customers. This enabled an in-depth analysis for discovery of trends among customers that lead to high loyalty. Then, another model which presented the demographical segmentation of the supermarket chain's customers was produced to analyse their demographics. Finally, two other models segmenting customers according to their geographical locations and number of purchases, followed by popularity of product categories were also created.

However, this study has few limitations due to the limited availability of data. For instance, no trends were identified for cities that were neither loyal nor churn. Data for an in-depth study would require information directly from the customers by using surveys and questionnaires which can be used for a comprehensive analysis to examine customer loyalty to find store variables such as layout, accessibility, services along with a wider range of consumer related variables and demographics including age that affect loyalty rate of supermarkets. The provided data set is also outdated and dates back to 2012 with the newest of it being from 2014. This decreases the level of usefulness of the generated results to the supermarket. Furthermore, subsequent studies on factors such as pricing, marketing strategy, and store branding with respect to customer loyalty can also be examined. This would further help in identification of trends for the supermarket market chain's data that could be used for implementing new strategies to improve the productivity of business processes.

## 7. References:


Anic, ID, & Radas, S 2006, 'The Role of Satisfaction and Demographic Factors in Building Store Loyalty', *Privredna kretanja i ekonomska politika*, vol.108, 67.

Baierle, IC, Sellitto, MA, Frozza, R, Schaefer, JL & Habekost, AF 2019, 'An artificial intelligence and knowledge-based system to support the decision-making process in sales', *South African journal of industrial engineering*, vol. 30, no. 2, pp. 17–25.

Bhat, I.H, & Singh, S 2017, 'Exploring the relationship between store image and store loyalty of an organized grocery retail', *International Journal of Research in Commerce & Management*, vol. 8, no.10.

Cammack, R, Atwood, T, Campbell, P, Parish, H, Smith, A, Vella, F & Stirling, J (eds) 2008, '*k*-means clustering', in *Oxford Dictionary of Biochemistry and Molecular Biology*, 2nd edn, Oxford University Press.

Chang, CH, & Tu, CY 2005, 'Exploring store image, customer satisfaction and customer loyalty relationship: evidence from Taiwanese hypermarket industry', *Journal of American Academy of Business*, vol.7. no.2, pp.197-202.

Dhar, SK, Hoch, SJ & Kumar, N 2001, 'Effective category management depends on the role of the category', *Journal of retailing*, vol. 77, no. 2, pp. 165–184.

Deshwal, P 2016, 'Customer experience quality and demographic variables (age, gender, education level, and family income) in retail stores', *International journal of retail & distribution management*, vol. 44, no. 9, pp. 940–955.

Ghochani, M, Afzalian, M, Siamak, G & Siavash, G 2013, 'Simulation of customer behavior using artificial neural network techniques', *International journal of information, business and management*, vol. 5, no. 2, pp. 59–59.

Google Developers 2021, *k-Means Advantages and Disadvantages*, Google Developers, viewed 13 May 2022, <https://developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages>.

Han, Y 2018, 'Essays on Dynamics of Consumers' In-store Shopping Behavior', PhD thesis, Indiana University.

HolyPython 2021, *K-Means Pros & Cons*, HolyPython, viewed 13 May 2022, <https://holypython.com/k-means/k-means-pros-cons/>.

Hubbard, L 2019, *Why Is Identifying the Target Market so Important to a Company?*, CHRON, viewed 13 May 2022, <https://smallbusiness.chron.com/identifying-target-market-important-company-76792.html>.

Hunneman, A, Verhoef, PC, & Sloot, LM 2017, 'The moderating role of shopping trip type in store satisfaction formation', *Journal of Business Research*, vol.78, pp.133-142.

Inman, JJ, Winer, RS & Ferraro, R 2009, ‘The Interplay among Category Characteristics, Customer Characteristics, and Customer Activities on In-Store Decision Making’, *Journal of marketing*, vol. 73, no. 5, pp. 19–29.

Kauffmann, E, Peral, J, Gil, D, Ferrández, A, Sellers, R & Mora, H 2019, ‘Managing marketing decision-making with sentiment analysis: an evaluation of the main product features using text data mining’, *Sustainability (Basel, Switzerland)*, vol. 11, no. 15, pp. 1-19.

Lejeune, MAP 2001, ‘Measuring the impact of data mining on churn management’, *Internet research*, vol. 11, no. 5, pp. 375–387.

Mitchell, D & Pavur, R 2002, ‘Using modular neural networks for business decisions’, *Management decision*, vol. 40, no. 1, pp. 58–63.

Neff, J 2008, *In-Store Displays Are More Effective Than Price Cuts*, Advertising Age, viewed 01 May 2022, <http://adage.com/article/news/store-displays-effective-price-cuts/132767/>.

Rahman, M, Md Chand, R, Kumar, J & Gupta, A 2020, ‘Does Consumer Demographics Effects Store Loyalty In Departmental Stores?’, *Academy of Marketing Studies Journal*, vol. 24, no.4, pp.1-18.

Ranjan, J & Bhatnagar, V 2011, ‘Role of knowledge management and analytical CRM in business: data mining based framework’, *The learning organization*, vol. 18, no. 2, pp. 131–148.

RingCentral 2022, *Customer Segmentation*, RignCentral, viewed 13 May 2022, <https://www.ringcentral.co.uk/gb/en/blog/definitions/customer-segmentation/>.

Sagar, A 2019, *Customer Segmentation Using K Means Clustering*, Toward Data Science, viewed 05 May 2022, <https://towardsdatascience.com/customer-segmentation-using-k-means-clustering-d33964f238c3>.

Singh, S & Sao, A 2015, ‘The Impact of Demographic Factor on Consumer Buying Behaviour: A Study of Retail Sector in Delhi NCR’, *International Journal on Customer Relations*, vol. 3, no. 1, pp. 45-53.

Sinha, PK, Banerjee, A, & Uniyal, D.P. 2002,’ Deciding where to buy: Store choice behaviour of Indian shoppers’, *Vikalpa*, vol. 27, no.2, pp.13-28.

Solgaard, HS & Hansen, T 2003,’ A hierarchical Bayes model of choice between supermarket formats’, *Journal of Retailing and Consumer Services*, vol.10, pp. 169-180.

Stead, M, MacKintosh, AM, Findlay, A, Sparks, L, Anderson, AS, Barton, K & Eadie, D 2017, ‘Impact of a targeted direct marketing price promotion intervention (Buywell) on food-purchasing behaviour by low income consumers: a randomised controlled trial’, *Journal of human nutrition and dietetics*, vol. 30, no. 4, pp. 524–533.

Trevino, A 2016, ‘Introduction to K-means Clustering’, *Oracle AI & Data Science Blog*, weblog post, 6 December, viewed 13 May 2022, <https://blogs.oracle.com/ai-and-datascience/post/introduction-to-k-means-clustering#:~:text=Business%20Uses,groups%20in%20complex%20data%20sets.>.

Vasudeva, S, & Chawla, S 2019, ‘Does Gender, Age and Income Moderate the Relationship Between Mobile Banking Usage and Loyalty?’, *International Journal of Online Marketing*, vol.9, no.4, pp. 1-18.

Wicker, KD 2016, ‘A study of customer value and loyalty in the supermarket industry’, PhD thesis, Capella University.