Classification and Clustering of Clients of a Company Dedicated to the Distribution of Auto Parts in the Metropolitan Area of Monterrey

Leonardo G. Hernández Landa(✉), Rosa E. Mata Martinez, Argelia Vargas Moreno, and Arlethe Y. García Aguilar

Facultad de Ciencias Químicas, Universidad Autónoma de Nuevo León, 66647 Monterrey, NL, Mexico
leonardo.hernandezln@uanl.edu.mx

Abstract. In this research work, we have collaborated with a company with a local presence dedicated to the sale and distribution of auto parts. The main problem is the lack of an optimized system of distribution to customers, which is capable of accelerating the delivery process. Particularly in this work, we have developed a method to calculate the correct clustering of customers by areas considering the density and volume of order. The result is a classification and assignment of a zone to each customer by the method of k-medoid considering two measures of performance of the literature in order to take the best.

Keywords: Clustering · Level of service · Logistics

1 Introduction

The study was conducted in a company dedicated to the sale and distribution of automotive parts, this company has positioned itself in the domestic market with 11 branches strategically located in Culiacan, Guadalajara, Mexico City, Merida, Monterrey, Puebla, Villahermosa, Tampico, Toluca, and Leon Guanajuato. This company is one of the most important auto parts wholesalers in Mexico, serving more than 7,000 automotive parts stores throughout the country.

In the city of Monterrey N.L. this company is one of the main wholesalers and they are looking to consolidate as the biggest supplier of the city based on their growth strategy.

Excellence in the Supply Chain is pursued, a fundamental part of customer service is efficiency in the delivery of purchase orders to customers. In the Monterrey branch, delivery is made employing different types of fleet. It has pick-up trucks, motorcycles, and cars that distribute throughout the Monterrey metropolitan area and in cities within its coverage area.

J. A. Marmolejo-Saucedo et al. (Eds.): COMPSE 2020, LNICST 359, pp. 3–11, 2021.
https://doi.org/10.1007/978-3-030-69839-3_1

The company seeks to define the best distribution alternative that will improve delivery times and take advantage of the resources it currently has adequately and sustainably.

The project has three main stages, which are listed below, and the process is detailed.

- First stage. Analysis of the current state of the distribution system and classification of customers.
- Second stage. Analysis of distribution alternatives and development of distribution algorithms.
- Third stage. Control and improvement.

1.1 Literature Review

The central feature of the old economy was the mass production and consumption of commodities. The modern economy is based on the production and consumption of increasingly differentiated goods and services [1]. [2] point out in their findings that the value of a service product is largely defined by perceptions of quality. Therefore, service consumers seem to place more importance on the quality of a service than on the costs associated with its acquisition. [3] in her research determines that service quality acts on service loyalty through customer satisfaction.

On the subject of distribution [4] suggest that clients evaluate the process of placing an order by considering the design, information accuracy, privacy, functionality and ease of use of a system. The quality of this process, in turn, positively affects their perceptions of the quality of the transaction outcome. The quality of the transaction outcome subsequently affects satisfaction evaluations. In the case of a problem, the way the retailer handles service recovery has a positive impact on satisfaction, and satisfaction measured the relationship from recovery and quality of results to purchase intentions. [5] in his exploration finds that service quality improvement initiatives must begin with defining customer needs and preferences, and their related quality dimensions.

That is why the first stage begins with the analysis of customer behavior, the value chain and thus finding the most important parameters and variables to establish the appropriate model. The classification of clients by location is one of the techniques that are used to make a correct routing of vehicles for the distribution of products.

[6] developed a cost estimation based on clustering to facility location and demonstrate the first step is to have an adequate technique for classifying clients through their geographic location, access, concentration and importance.

[7] performs a classification using data mining such that it identifies the high-profit, high-value and low-risk customers. [8] optimize the distribution of products with drones, the first step is to classify clients partitioning of delivery locations into small clusters, identifying a focal point per cluster, and routing through all focal points.

2 Methodology

2.1 Analysis of the Current State of the Service

The Monterrey branch currently has a portfolio of approximately 400 clients of which more than 250 have regular activity. Most of the clients' activities are not constant, however the recurrent clients are frequent. Figure 1 shows that the 10 clients with the highest participation represent 32.4% of the total activity, the 20 most frequent clients represent 45% of the activity and the 50 most frequent clients represent 63% of the total activity.

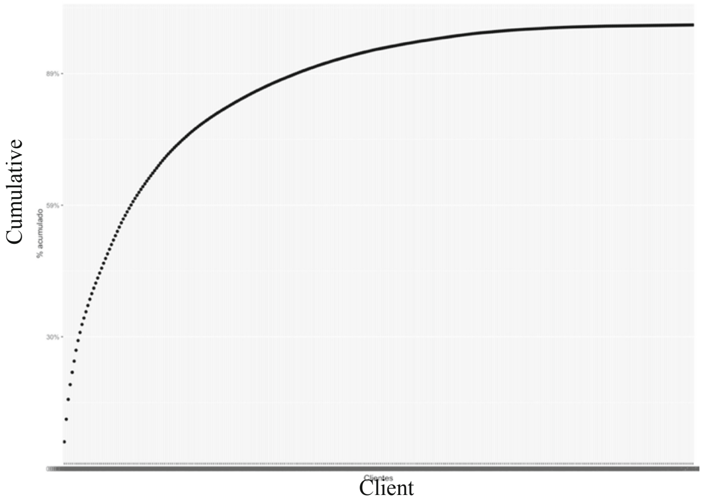

Fig. 1. Accumulated activity by client

The average daily activity per customer is 5 lines, you can see in Fig. 2. The average number of orders processed by the branch is 150 per working day, and in Fig. 3 it can be seen that demand is maintained on working days at high peaks. This results in an average of 700 to 800 products being shipped per day in the week.

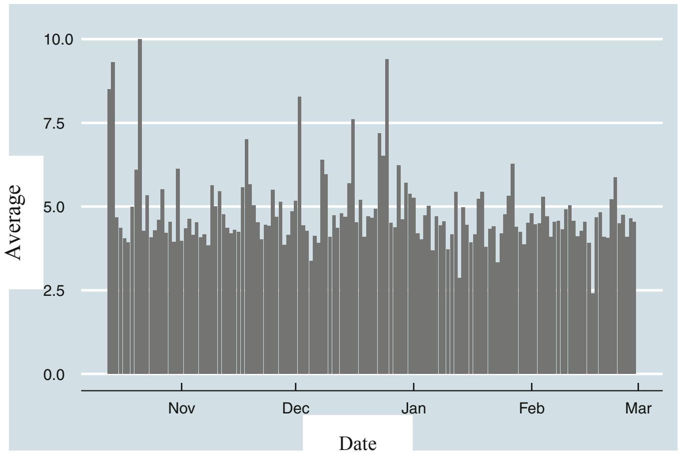

Fig. 2. Average number of products ordered per day per customer

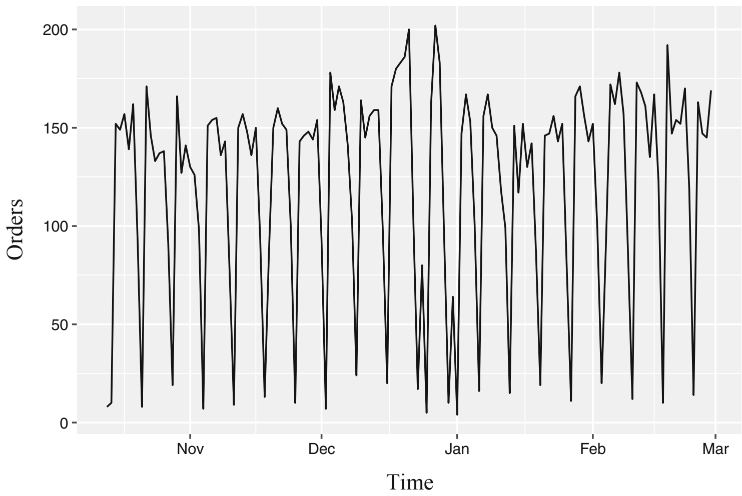

Fig. 3. Average order trend per day

This trend represents a challenge for customer service both in the internal warehouse process and in the distribution of products to customers.

2.2 Classification and Clustering

The actual process starts with the generation of a purchase order, the product stock is checked in the system, consequently the purchase order is passed to the warehouse department to collect the product, the order invoice is generated, it is assigned to a driver for distribution together with other purchase orders and finally it is delivered to the customer in route.

One of the needs in the distribution process is to speed up the process of assigning routes to customer orders and reduce delivery times. There are currently 11 identified customer zones in the metropolitan area. However, the current distribution of customers is not adequate, since there are areas with a high density of customers and others with a low density of customers. The current process only recognizes a geographic distribution per municipality in the area but does not consider density by zones. Having a large number of customers makes it necessary to redistribute customer areas and create better routes in order to improve service times. Within the desirable characteristics is that it is dynamic, flexible and precise.

As a background there are many works in the scientific literature that have managed to classify clients through the use of data analysis and mathematical techniques. [9] make a compilation until 2009 of the methods of classification and grouping of clients, highlighting the most used techniques among which are neural networks, decision trees, association rules, Markov chains, etc. Each method used is dependent on the structure of the customers and the objective of the classification. For our case we want to group the customers based on the following aspects: Geographic location, sales volume, geographic dispersion and workload balancing for the fleet. In [10] a complete guide to geospatial information management in points of interest is presented.

A collection of the precise geospatial information of the customers is made in order to analyze their location and visualize the distribution challenge. Figure 4 shows only the main customers distributed throughout the city of Monterrey and its metropolitan area.

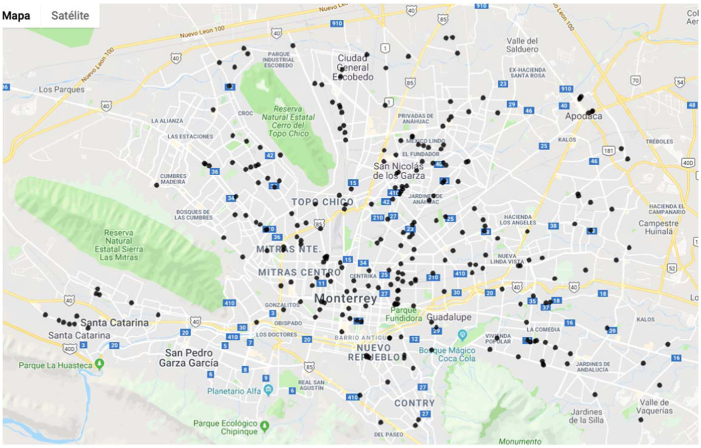

Fig. 4. Physical location of the clients of the Monterrey branch

At first glance, the density of the client concentration is not noticeable, since the concentration of the dots is lost in the dotted display, so Fig. 5 shows the heat map where the green color represents a higher degree of client concentration and is degraded to purple. Greater concentration is observed in the central zone of the municipality of Monterrey and in the central zone of the municipality of San Nicolas de los Garza. The center of Santa Catarina, Apodaca and Guadalupe also show a higher intensity with respect to the rest of the metropolitan area. It can be seen how the municipality of Santa Catarina is separated from the density map and it is important to take this into account when making the classification.

For the clustering of clients, it is decided to use the partition around the medoids with estimation of the number of groups. The method is described in [11] where k representative objects, called medoids (selected from the data set) are searched for that can serve as *prototypes.* for the group instead of the means to allow the use of other arbitrary dissimilarities and arbitrary input domains, using the absolute error criterion (Total deviation) as a target.

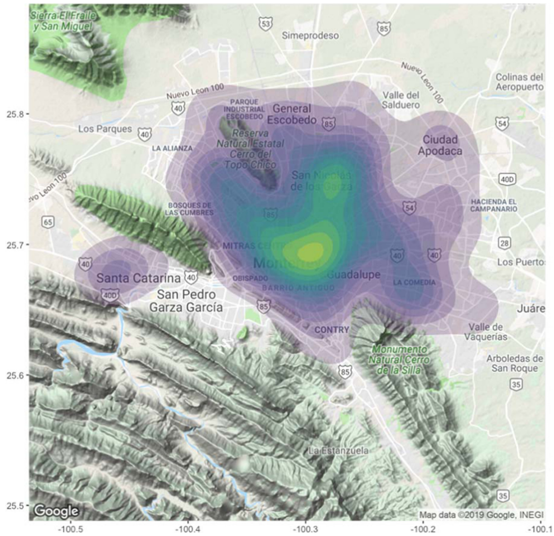

Fig. 5. Heat map of customer concentration density in the city of Monterrey

$$TD := \sum_{i=1} \sum_{x_{j \in C_i}} d\left(x_j, m_i\right)$$

which is the sum of the dissimilarities of each point $x_j \in C_i$ to the medoid of each cluster. If we use the Euclidean squared distance function (i.e., $d(x, m) = |x - m|_2^2$),), we almost get the usual SSQ target used by k-means, except that k-medoids is free to choose any $\mu_i \in \mathbb{R}^d$, while in k-median $m_i \in C_i$ must be one of the original data points.

For Euclidean squared distances and Bregman divergences, the arithmetic mean is the optimal choice for μ and a fixed group assignment. For distance L_1 (that is, $\sum |x_i - y_i|$), also called Manhattan distance, the median of the components is a better choice. The medoid of a C set is defined as the object with the smallest sum of differences (or, equivalently, the smallest average) from all other objects in the set

$$\text{medoid}(C) := \arg\min \sum_{x_i \in C} d\left(x_i, x_j\right)$$

the "Partition Around Medoids" (PAM) algorithm consists of two phases, one of construction and another of improvement by means of exchange that optimizes the grouping. It should be noted that this algorithm works on $O(k)$.

For this article we consider the performance measures: Average silhouette width ASW [12] and [13] index CH.

ASW and CH attempt to balance a small within-cluster heterogeneity (which increases) and a large between-cluster heterogeneity (which increases with k as well) in a supposedly optimal way [14].

3 Results

The clustering algorithm is executed on a computer with an Intel core i5 processor with 6 GB of RAM. The development of the cleaning, selection and classification algorithm is executed in the R language, creating a variation of clustering scenarios of the customers to verify the sensitivity of the algorithm to reclassify the groups in cases of fleet size change or the generation of a new clustering strategy. Instances from 4 clusters up to 20 clusters were executed in order to analyze the different alternatives and reach a conclusion on how many clusters for the client classifications are necessary.

As you can see in Table 1 the best metrics are between 9 and 12 clusters according to the AWS and CH metrics. That directly adjusts to the size of the fleet of vehicles that the company has to carry out the delivery of the products.

Table 1. Results by number of cluster VS performance measures

Number of clusters	ASW	CH
4	0.3923259	190.9851
5	0.3856825	198.5562
6	0.3662694	186.4601
7	0.3674657	191.7803
8	0.3857323	192.7776
9	0.3582335	183.0448
10	0.3995877	184.6448
11	0.3812556	176.7437
12	0.3740707	174.2456
13	0.3810907	172.2084
14	0.3942427	184.6819
15	0.4011587	181.5634
16	0.4085602	184.3999
17	0.4145238	255.751
18	0.4040372	254.3008
19	0.414239	261.3025
20	0.4102043	260.5144

Figure 6 is the final result of grouping customers by the available fleet as can be seen in the colors so that the supplier can find a better way to order and make the process much faster.

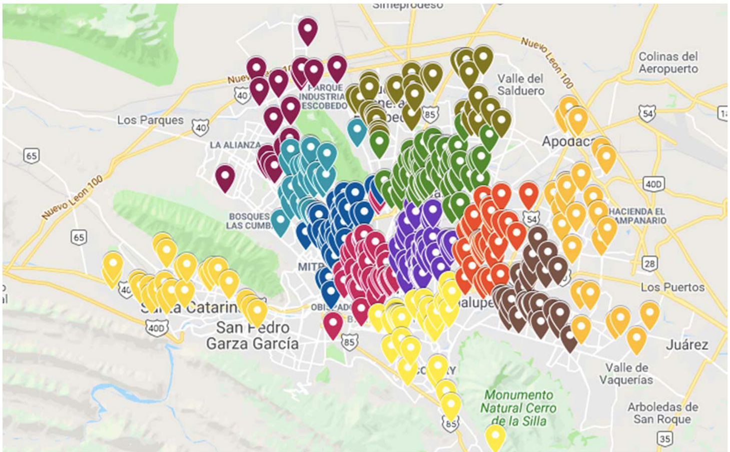

Fig. 6. Final grouping with 11 client clusters

4 Discussion and Further Work

After the analysis, a methodology was established for the classification and grouping of clients for stage one of the projects. The final document provided to the company is a database supported with google maps so they can manipulate the details of each client and create a better routing. The result will speed up the process of route generation and distribution of products in order to reduce loading time. In the same way, this classification will allow the generation of optimal routes for the distribution units, since a lower sectorization will allow to speed up the route search algorithms when considering smaller groups of customers. The efficiency of the algorithm was demonstrated by maintaining the level of homogeneity in the variation of the number of clusters and it is expected that in the future the level of activity of the customer will be introduced to improve the algorithm.

References

1. Fornell, C., Johnson, M.D., Anderson, E.W., et al.: The American customer satisfaction index: nature, purpose, and findings. J. market. **60**(4), 7–18 (1996)
2. Cronin, J.J., Brady, M.K., Tomas, G., Hult, M.: Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. J. Retail. **76**, 193–218 (2000)
3. Caruana, A.: Service loyalty Service loyalty the effects of service quality and the mediating role of customer satisfaction. Eur. J. Mark. **36**, 309–0566 (2002). https://doi.org/10.1108/03090560210430818
4. Collier, J., Bienstock, C.: Measuring service quality in e-retailing. Artic J. Serv. Res. (2006). https://doi.org/10.1177/1094670505278867

5. Yang, Z., Kim, D., Jun, M., Kong, H.: Customers' perceptions of online retailing service quality and their satisfaction Implementation of a cross-border charity project: institutional distance partnership arrangement and project performance view project customers' perceptions of online retailing service quality and their satisfaction. Artic Int. J. Qual. Reliab. Manag. (2004). https://doi.org/10.1108/02656710410551728
6. Klose, A., Klose, A.: Using Clustering Methods in Problems of Combined Location and Routing. https://doi.org/10.1007/978-3-642-80117-4_71
7. Rajagopal, S.: Customer data clustering using data mining technique. Int. J. Database Manag. Syst. (IJDMS 3) (2011). https://doi.org/10.5121/ijdms.2011.3401
8. Salama, M., Srinivas, S.: Joint optimization of customer location clustering and drone-based routing for last-mile deliveries. Transp. Res. Part C Emerg. Technol. **114**, 620–642 (2020). https://doi.org/10.1016/j.trc.2020.01.019
9. Ngai, E.W.T., Xiu, L., Chau, D.C.K.: Application of data mining techniques in customer relationship management: a literature review and classification. Expert Syst. Appl. **36**, 2592–2602 (2009)
10. Baddeley, A., Rubak, E., Turner, R., et al.: Spatial Point Patterns. Chapman and Hall/CRC (2015)
11. Schubert, E., Rousseeuw, P.J.: Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms (2018)
12. Van der Laan, M.J., Pollard, K.S., Bryan, J.: A new partitioning around medoids algorithm. J. Stat. Comput. Simul. **73**, 575–584 (2003). https://doi.org/10.1080/0094965031000136012
13. Caliñski, T., Harabasz, J.: A dendrite method foe cluster analysis. Commun. Stat. **3**, 1–27 (1974). https://doi.org/10.1080/03610927408827101
14. Hennig, C., Liao, T.F.: How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. J. R. Stat. Soc. Ser. C (Appl. Stat.) **62**, 309–369 (2013). https://doi.org/10.1111/j.1467-9876.2012.01066.x