

Tecnura

ISSN: 0123-921X

Universidad Distrital Francisco José de Caldas

Cuadros López, Álvaro Julio; Gonzales Caicedo, Caterine; Jiménez Oviedo, Paola Cristina
Análisis multivariado para segmentación de clientes basada en RFM

Tecnura, vol. 21, núm. 54, 2017, Octubre-Diciembre, pp. 41-51

Universidad Distrital Francisco José de Caldas

DOI: <https://doi.org/10.14483/22487638.12957>

Disponible en: <https://www.redalyc.org/articulo.oa?id=257054861003>



Análisis multivariado para segmentación de clientes basada en RFM

Multivariate analysis for customer segmentation based on RFM

Álvaro Julio Cuadros López¹, Caterine Gonzales Caicedo², Paola Cristina Jiménez Oviedo³

Fecha de recepción: 26 de febrero de 2017

Fecha de aceptación: 28 de agosto de 2017

Cómo citar: Cuadros L., A.J.; Gonzales C., C. y Jiménez O., P.C. (2017). Análisis multivariado para segmentación de clientes basada en RFM. *Tecnura*, 21(54), 41-51. <https://doi.org/10.14483/22487638.12957>

Resumen

Contexto: Para construir una gestión exitosa de relaciones con los clientes, las empresas deben comenzar con la identificación del verdadero valor de los clientes ya que esto proporciona información básica para implementar estrategias de *marketing* más dirigidas y personalizadas. El análisis RFM, un método clásico de análisis, permite mediante tres parámetros de evaluación, analizar el comportamiento de los clientes y establecer segmentos. La adición de un nuevo parámetro en la metodología tradicional es una oportunidad para refinar los posibles resultados de una segmentación de clientes, pues no sólo proporciona un nuevo elemento de evaluación para identificar los clientes más valiosos, sino que permite diferenciar y conocer aún mejor a los clientes.

Método: La propuesta metodológica que se incluye en este artículo permite establecer segmentos de clientes utilizando un método RFM extendido con nuevas variables, seleccionadas mediante análisis multivariados.

Resultados: La propuesta fue aplicada en una empresa en la que se probaron variables, como *ganancia*, *porcentaje de ganancia* y *días de vencimiento de facturación*, y se logró establecer una segmentación de clientes más detallada que el RFM clásico.

Conclusiones: El análisis RFM es un método de amplio uso en la industria por su fácil comprensión y aplicabilidad. Este método puede ser mejorado con el uso de procedimientos estadísticos y nuevas variables, lo que permitirá contar con información más profunda sobre el comportamiento de los clientes y facilitará el diseño de estrategias específicas de marketing.

Palabras clave: Análisis multivariado, CRM, RFM, Segmentación.

Abstract

Context: To build a successful relationship management (CRM), companies must start with the identification of the true value of customers, as this provides basic information to implement more targeted and customized marketing strategies. The RFM methodology, a classic analysis tool that uses three evaluation parameters, allows companies to understand customer behavior, and to establish customer segments. The addition of a new parameter in the traditional technique is an opportunity to refine the possible outcomes of a customer segmentation since it not only provides a new element of evaluation to identify the most valuable customers, but it also makes it possible to differentiate and get to know customers even better.

¹ Ingeniero Industrial, magíster en Administración de Empresas, docente Universidad del Valle. Miembro del grupo de investigación en Logística y Producción. Santiago de Cali, Colombia. Contacto: alvaro.cuadros@correounivalle.edu.co

² Ingeniera industrial. Universidad del Valle. Santiago de Cali, Colombia. Contacto: caterinegcv@yahoo.com

³ Ingeniera industrial. Universidad del Valle. Santiago de Cali, Colombia. Contacto: cristina-jimenovi@hotmail.com

Method: The article presents a methodology that allows to establish customer segments using an extended RFM method with new variables, selected through multivariate analysis..

Results: The proposed implementation was applied in a company in which variables such as profit, profit percentage, and billing due date were tested. Therefore, it was possible to establish a more detailed customer segmentation than with the classic RFM.

Conclusions: the RFM analysis is a method widely used in the industry for its easy understanding and applicability. However, it can be improved with the use of statistical procedures and new variables, which will allow companies to have deeper information about the behavior of the clients, and will facilitate the design of specific marketing strategies.

Keywords: CRM, Multivariate analysis, RFM, Segmentation.

INTRODUCCIÓN

El entorno empresarial es cada vez más competitivo y complejo, y la administración de relaciones con los clientes (*Customer Relationship Management, CRM*) es una estrategia competitiva para comprender a los clientes de una empresa (Kim, Jung, Suh y Hwang, 2006; Chuang y Shen, 2008). Pero la pregunta es: ¿En qué clientes se deben enfocar los esfuerzos para construir relaciones exitosas y competitivas, considerando que no todos ellos tienen la misma importancia para la empresa?

Para determinar este nivel de importancia, la utilización de técnicas de segmentación de clientes resulta útil y decisiva en la identificación de aquellos clientes que son realmente rentables, y permite focalizar más recursos en éstos, maximizando su valor, e igualmente, utilizando óptimamente recursos en función de captación, retención o recuperación de los mismos (Khajvand, Zolafghar, Ashoori y Alizadeh, 2011). Las empresas deben entonces segmentar sus clientes en función de la capacidad de compra y otros condicionamientos, como su solvencia y garantía, debido a que la identificación del valor y la rentabilidad del cliente permiten desplegar estrategias de comercialización más específicas y personalizadas (Santandreu, 2002)

En los últimos años las técnicas de manipulación de las bases de datos han evolucionado a partir de modelos simples RFM (compras recientes (R), frecuencia de las compra (F) y monto de las

compra (M)) (McCarty y Hastak, 2007). Por ejemplo, Shih y Liu (2003) propusieron utilizar ponderados como el ahp para dar pesos a las variables RFM. También, Khajvand *et al.* (2011) propusieron un método en el que asignaron pesos a las variables R, F y M, dependiendo de las características de la industria y adicionalmente agregaron la variable *cantidad de ítems* para segmentar clientes.

Aun cuando el RFM se ha utilizado por más de 50 años, sigue vigente entre los gerentes de mercadeo, ya que es fácil de entender, utilizar e implementar. El análisis RFM se basa en las siguientes observaciones simples que se han hecho una y otra vez a través de múltiples industrias (Kohavi y Parekh, 2004).

Los clientes que han comprado recientemente es probable que respondan mejor a los mensajes. También son más propensos a comprar nuevamente, en comparación con alguien que no ha comprado por un largo tiempo. Para algunas familias de productos donde se genera mucho entusiasmo por la compra, esto en efecto es más cierto.

- Los compradores frecuentes son más propensos a comprar otra vez que los compradores poco frecuentes.
- Los derrochadores suelen responder mejor que los que gastan poco.

Los tres atributos de comportamiento de RFM son extremadamente simples, ya que pueden ser fácilmente calculados para cualquier base de

datos que tenga historial de compras, son fáciles de comprender, y aun así son poderosos en su capacidad predictiva (Kohavi y Parekh, 2004).

El análisis RFM ofrece un punto de partida para identificar el comportamiento de clientes; no obstante, según el tipo de industria o características específicas de una empresa, pueden presentarse otros parámetros de evaluación útiles, que permiten analizar de manera más precisa el comportamiento de éstos. En los últimos años, las técnicas de manipulación de las bases de datos han permitido incluir otras variables al análisis para poder comprender mejor al consumidor, como también crear atributos RFM para categorías específicas de familias de productos e industrias (McCarty y Hastak, 2007).

Según lo anterior, y considerando los estudios que han surgido alrededor de la técnica RFM, este artículo plantea mediante un caso de estudio, un método propuesto de segmentación a partir de la técnica ya existente RFM, que por su simplicidad se ajusta fácilmente a las posibles variables que puede tener un portafolio de clientes. Mediante la inclusión de nuevas variables a la técnica tradicional RFM, la propuesta busca ofrecer los mismos beneficios de la técnica original y adaptarse aún más a las necesidades o características propias de una empresa, como el tipo de productos, servicios, políticas y clientes. La inclusión de las nuevas variables se justifica utilizando técnicas de *análisis multivariante*, con el fin de conocer si las variables están correlacionadas, si son redundantes a los parámetros de la técnica RFM tradicional y si están aportando un valor agregado al método propuesto.

METODOLOGÍA

La metodología seguida para este trabajo consistió en varias etapas partiendo de la preparación de los datos, continuando con la definición, cálculo y selección de las variables a incluir, y finalmente, la agrupación de los clientes con dichas variables. A continuación, se describen los pasos seguidos en la investigación:

Preparación de los datos

Una primera instancia del proceso consiste en la revisión de la información existente de los clientes para identificar posibles errores o información no relevante para el análisis. El caso de estudio se desarrolló en una microempresa de la ciudad de Cali, dedicada a la manufactura y comercialización de productos desechables y plásticos. La empresa tiene un portafolio de 23 líneas de productos que comercializa a mayoristas y minoristas en varias ciudades del suroccidente del país. Tras la depuración, la base de datos de transacciones contó 304 clientes que, durante el periodo de análisis, 8 meses, realizaron 5962 transacciones.

Definición de las variables

Luego de la depuración, se calcularon las variables clásicas del modelo RFM, otras variables para realizar la selección. Estas fueron:

Reciente (R). Compras recientes. Mínimo tiempo en días transcurridos desde la última transacción hasta el final del periodo de análisis.

- Frecuencia (F). Número de veces que cada cliente realizó una compra.
- Monetario (M). Sumatoria de todas las transacciones del periodo.

Adicionalmente, teniendo en cuenta la información disponible y la opinión de los tomadores de decisiones en la empresa, se definieron otras tres variables para probar en la segmentación. Estas fueron:

- Ganancia (G). Representa el valor que, de las ventas, realmente contribuye a la organización. Resulta de multiplicar la razón de margen contributivo que la empresa tiene calculado por familia de productos en el mes, por el valor de la compra realizada por el cliente durante ese mismo mes de esa familia de productos. Despues se suman todas las ganancias por familia

de productos, durante todo el tiempo de análisis para cada cliente como se muestra en la ecuación (1).

$$Ganancia = \sum_{i=1}^n ingreso_{j,i} * \%margen_{j,i} \quad (1)$$

Donde i = mes, j = familia de producto

El margen contributivo está siendo calculado en la empresa por familia de productos como muestra la ecuación (2).

$$\%margen_{j,i} = \frac{ingreso_{j,i} - costo\ de\ variable_{j,i}}{ingreso\ venta_{i,j}} \quad (2)$$

Donde el costo de variable se calcula de acuerdo con la ecuación (3).

$$Costo\ de\ variable_{(i,j)} = compra\ MP_{(i,j)} + inventario_{(i-1,j)} + costo\ MO-inventario_{(i,j)} \quad (3)$$

- % de ganancia (%G). Esta variable representa la ganancia obtenida en términos de porcentajes por un cliente cuando compra. Se calcula como la relación entre la variable *ganancia* y la variable “*monetario*”.
- *Días vencidos* (DV). El número de días vencidos (DV) por cliente hace referencia a la cantidad de días que el cliente tarde en cancelar el valor total de la compra realizada en determinada fecha, el conteo de dichos días comienza al finalizar el plazo que le es otorgado para la cancelación de

la factura, el cual varía dependiendo de si la venta es a crédito o de contado o del sector donde se ubica el cliente.

Cálculo de las variables

A partir de la descripción realizada anteriormente se procedió a calcular los valores de cada variable por cliente. En la tabla 1 se puede ver una fracción de la base de datos.

Tabla 1. Valores iniciales de las variables R, F, M, G, %G, DV

ID CLIENTE	RECENCIA (días)	FRECUENCIA (veces)	MONETARIO (\$)	GANANCIA (\$)	GANANCIA %	D. VENCIDOS (días)
C-0001	22	15	\$3.300.599	\$568.316	17%	0
C-0007	14	15	\$4.535.000	\$1.367.298	30%	4
C-0010	1	9	\$97.800	\$15.887	16%	0
C-0021	2	45	\$3.670.088	\$756.672	21%	96
C-0280	245	1	\$55.000	\$7.150	13%	0
C-0281	97	7	\$1.168.500	\$358.215	31%	29
C-0297	14	2	\$609.600	\$163.831	27%	0
...
...
C-0294	6	6	\$2.087.125	\$610.935	29%	72
C-0282	98	1	\$95.610	\$19.222	20%	0
C-0309	213	1	\$912.000	\$143.035	16%	0
C-0310	18	7	\$3.811.450	\$1.618.339	42%	1
C-0311	84	5	\$487.300	\$151.409	31%	57

Fuente: elaboración propia.

Selección de variables representativas

En primera instancia, teniendo en cuenta que los datos se encontraban expresados en diferentes unidades de medida, se normalizaron para tenerlos en forma homogénea, mediante fórmulas de costo y beneficio.

- Normalización de las variables R y DV: dado que los valores de estas variables influyen negativamente en la calificación del cliente, es decir, a mayores valores de estas variables menor será su calificación, se utiliza la fórmula del costo dada por la ecuación (4).

$$X' = \frac{(X^L - X)}{(X^L - X^S)} \quad (4)$$

Donde:

X' = valor normalizado de la variable

X = valor inicial de la variable

X^L = máximo valor de la variable

X^S = mínimo valor de la variable.

- Normalización de las variables F, M, G y %G: como estas variables influyen positivamente en la calificación del cliente, es decir, a mayores valores de estas variables mayor será su calificación, se utiliza la fórmula del beneficio dada por la ecuación (5).

$$X' = \frac{(X - X^S)}{(X^L - X^S)} \quad (5)$$

Donde:

X' = valor normalizado de la variable

X = valor inicial de la variable

X^L = máximo valor de la variable

X^S = mínimo valor de la variable.

Tras aplicar las ecuaciones (4) y (5) se obtuvo la tabla 2, donde las cifras ya se encuentran normalizadas.

Tabla 2. Valores normalizados de las variables RFM-G-%G-DV

ID CLIENTE	RN	FN	MN	GN	%GN	D.VN
C-0001	0,9194	0,1129	0,1746	0,1074	0,3261	1,0000
C-0002	0,9670	0,0565	0,0104	0,0111	0,4887	1,0000
C-0003	0,8718	0,0565	0,2337	0,2384	0,5101	1,0000
C-0004	0,8388	0,0161	0,0126	0,0090	0,3463	1,0000
C-0005	0,9963	0,3468	0,3617	0,2928	0,4153	1,0000
...
...
C-0308	0,9927	0	0,0438	0,0303	0,3545	0,9969
C-0309	0,2198	0	0,0478	0,0273	0,3013	1,0000
C-0310	0,9341	0,0484	0,2017	0,3051	0,7332	0,9977
C-0311	0,6923	0,0323	0,0253	0,0289	0,5495	0,8655

Fuente: elaboración propia.

Las variables pueden ser dependientes o independientes entre sí, de manera que la selección se hizo mediante técnicas de análisis multivariado para validar que las que se introdujeran en el modelo, realmente ofrecieran información adicional sobre los clientes y no hubiera redundancia.

El análisis multivariado permite el estudio estadístico de datos de una población para resumirlos mediante nuevas variables, encontrar grupos, y clasificar nuevas observaciones en los grupos definidos (Toro, Pérez y Bernal, 2007).

De esta manera, las variables que dependieran de otras resultarían redundantes en el modelo de análisis y se descartarían. Las técnicas utilizadas fueron:

- Matriz de covarianza. Permite identificar si hay relaciones entre variables. En este análisis, si la relación entre las variables es cero, se sabe que no hay relación entre las variables. La ecuación (6) permite calcular la relación entre las variables que es el promedio de los productos de las desviaciones de ambas variables con relación a sus medias respectivas.

$$S_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (6)$$

Como se puede ver en la tabla 3, al calcular las covarianzas, se encontró que en ningún caso la relación es cero, por tanto, aun cuando la relación existe, no se puede discernir si es lo suficientemente fuerte para afectar el análisis a realizar.

Tabla 3. Matriz de covarianzas para medir la relación entre las variables RFM-G-%G-DV

	RN	FN	MN	GN	%GN	D.VN
RN	0,08697984	0,0134849	0,0140787	0,013147	-0,0016044	-0,0052582
FN		0,0119773	0,0079727	0,007211	-0,0007070	-0,0031835
MN			0,0178038	0,015814	0,0002737	-0,0066995
GN				0,015552	0,0022223	-0,0053622
%GN					0,0212727	-0,0004814
D.VN						0,0127100

Fuente: elaboración propia.

- Matriz de correlaciones. Cuando una variable independiente se puede expresar como una combinación lineal de las otras, se considera que la co-linealidad es perfecta y entonces se puede omitir dicha variable. Los valores de la correlación varían entre 1 y -1, y cuando el valor de la correlación entre una variable x y se aproxima

a 0, la relación entre las variables es nula. La matriz de correlaciones se desarrolló utilizando la ecuación (7).

Este análisis permitió identificar una fuerte relación (0,9503638) entre las variables M y G como se ve en la tabla 4.

$$r = \frac{s_{xy}}{s_x s_y}, \quad 0 \quad r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}} \quad (7)$$

Tabla 4. Matriz R, para medir la correlación entre las variables RFM-G-%G-DV

	RN	FN	MN	GN	%GN	D.VN
RN	1	0,4177911	0,3577637	0,3574427	-0,0372997	-0,1581454
FN		1	0,5459694	0,5283655	-0,0442907	-0,2580204
MN			1	0,9503638	0,0140625	-0,4453605
GN				1	0,1221805	-0,3813964
%GN					1	-0,0292789
D.VN						1

Fuente: elaboración propia.

- Co-linealidad. Para comprobar de forma concluyente qué variables en este caso de estudio son independientes y qué variables no, se realizaron dos mediciones más de la co-linealidad:

Tolerancia: se calcula de acuerdo con la ecuación (8).

$$\text{Tolerancia} = (1 - R_j^2) \quad (8)$$

Donde R_j^2 j es el coeficiente de determinación de la variable X_j contra todas las demás X .

El coeficiente de determinación mide la proporción de variabilidad total de la variable dependiente y se obtiene a través de la ecuación (9).

$$R^2 = \frac{S_{xy}^2}{S_x^2 S_y^2} \quad (9)$$

Factor de inflación de la varianza: también conocido como FIV, se calcula de acuerdo con la ecuación (10).

$$FIV = \frac{1}{\text{tolerancia}} \quad (10)$$

Como se puede apreciar en la tabla 5, al calcular la matriz FIV, entre las variables M y G, existe una colinealidad excesiva (10,3296623), es decir, las dos variables están fuertemente relacionadas.

Tabla 5. Factor de inflación de las variables RFM-G-%G-DV

RN	FN	MN	GN	%GN	D.VN
RN	1,2114595	1,1467822	1,1464804	1,0013932	1,0256515
FN		1,4246691	1,3872898	1,0019655	1,0713228
MN			10,3296623	1,0001978	1,2474209
GN				1,0151543	1,1702246
%GN					1,0008580
D.VN					

Fuente: elaboración propia.

Con lo anterior se confirmó que existe una fuerte relación entre las variables M y G, por lo se podía eliminar una del modelo. La decisión fue extraer la variable G, ya que M es parte del modelo tradicional RFM y lo que se buscaba con el proyecto era explorar la adición de nuevas variables y no reemplazarlas. De esta manera, quedaron seleccionadas para realizar el análisis de los clientes, las variables: R, F, M, %G y DV.

Agrupación de los clientes

Con las variables a introducir ya seleccionadas y calculadas, se procedió a realizar la agrupación de clientes a través de *K-means*, con el fin de obtener

los grupos con características de comportamiento similares que permitan detectar aquellos clientes más valiosos. El método *K-means* es una técnica particional no paramétrica que permite hacer conglomerados según un número decidido por el analista y es ampliamente utilizado en minería de datos (Rueda *et al.*, 2016; Durango, Quintero y Ruiz, 2015; Giraldo, León y Gómez, 2013). Este método se desarrolló en diferentes pasos:

Decisión de los k elementos o número de clústers que se desea crear, que representan el centro inicial de cada clúster. Para el proyecto se definió contar con cinco grupos para los cuales se definieron los centroides iniciales, como se ve en la tabla 6.

Tabla 6. Centroides iniciales de cada variable por grupo

CENTROIDES INICIALES	RN	FN	MN	%GN	D.VN
C-0001	0,919414	0,112903	0,174598	0,107400	0,326106
C-0010	0,996337	0,064516	0,004615	0,003377	0,310386
C-0020	0,234432	0,072581	0,026001	0,016219	0,319227
C-0030	0,820513	0,024194	0,009171	0,010996	0,543279
C-0040	0,483516	0,104839	0,052866	0,045711	0,433943

Fuente: elaboración propia.

A continuación, cada una de las instancias es asignada al centro del clúster más cercano de acuerdo con la distancia euclídea que le separa de él. Tras calcular las distancias euclídeas se asignaron los elementos al centro del clúster más cercano, por lo que se contó con una primera aproximación a los grupos. En la tabla 7 se aprecia el número de elementos que inician en cada grupo.

Para cada uno de los clúster así construidos se calcula el centroide de todas sus instancias y estos centroides son tomados como los nuevos centros de sus respectivos clústeres. El proceso se repite

con los nuevos centros de los clúster, hasta que no haya cambios en los grupos con cada iteración, lo que significa que los puntos centrales de los grupos se han establecido y permanecerán inalterables. En la tabla 8 se presentan los resultados de la última iteración en la que no hubo cambios.

RESULTADOS

Finalmente, con los grupos establecidos, se calculó la calificación de cada grupo, como la sumatoria del promedio del valor de las variables

Tabla 7. Número de clientes por cada grupo en la primera iteración.

ITERACIÓN 1	Número de clientes
GRUPO 1	80
GRUPO 2	46
GRUPO 3	41
GRUPO 4	95
GRUPO 5	42

Fuente: elaboración propia.

Tabla 8. Última iteración del *clustering*

ITERACIÓN 12	RN	FN	MN	%GN	D.VN	Número de clientes	CAMBIOS
GRUPO 1	0,964868	0,269795	0,463246	0,429516	0,477570	22	0
GRUPO 2	0,945021	0,107552	0,074290	0,062064	0,406964	107	0
GRUPO 3	0,163041	0,005952	0,009672	0,009891	0,492267	49	0
GRUPO 4	0,891627	0,050000	0,068330	0,079092	0,577217	70	0
GRUPO 5	0,607405	0,018865	0,021403	0,019650	0,427793	56	0

Fuente: elaboración propia.

normalizadas de los clientes que se encuentran en dicho grupo. Los grupos con mejores calificaciones fueron los 1 y 4, mientras que el grupo con la más baja calificación fue el 3, como se puede ver en la tabla 9.

Con el interés de establecer una calificación por intervalos, se estableció una escala con cinco

niveles entre los límites, inferior y superior de las variables normalizadas, como se ve en la tabla 10.

Para calificar las variables de cada grupo, se estableció que calificaciones de 1 y 2 corresponderían a una calificación baja; calificación de 3 corresponde a una media, y calificaciones de 4 y 5 corresponden a una alta. La tabla 11 resume los

Tabla 9. Número de observaciones arrojadas por grupo

ITERACIÓN 12	RN	FN	MN	%GN	D.VN	Número de clientes	ΣTOTAL CALIFICACIONES
GRUPO 1	0,964868	0,269795	0,463246	0,429516	0,477570	22	2,604996
GRUPO 2	0,945021	0,107552	0,074290	0,062064	0,406964	107	1,595890
GRUPO 3	0,163041	0,005952	0,009672	0,009891	0,492267	49	0,680797
GRUPO 4	0,891627	0,050000	0,068330	0,079092	0,577217	70	1,666266
GRUPO 5	0,607405	0,018865	0,021403	0,019650	0,427793	56	1,095115

Fuente: elaboración propia.

Tabla 10. Valor calificación por quintiles

TABLA CALIFICACIÓN		
Límite inferior	Límite superior	Calificación
0	0,2	1
0,2	0,4	2
0,4	0,6	3
0,6	0,8	4
0,8	1	5

Fuente: elaboración propia.

Tabla 11. Resumen resultados de los grupos

GRUPO DE CLIENTES	Número de clientes	R		F		M		%G		DV	
		Mín.	Máx.	Mín.	Máx.	Mín.	Máx.	Mín.	Máx.	Mín.	Máx.
GRUPO 1 Calif	22	- ALTA	42 5	6 BAJA	125 2	\$2.597.355 MEDIA	\$18 852 695 3	14% MEDIA	38% 3	- MEDIA	426 3
GRUPO 4 Calif	70	- ALTA	76 5	1 BAJA	37 1	\$19.050 BAJA	\$4.750.373 1	28% MEDIA	58% 3	1 MEDIA	200 3
GRUPO 2 Calif	107	- ALTA	76 5	1 BAJA	43 1	\$25.600 BAJA	\$4.576.907 1	1% BAJA	28% 1	- MEDIA	207 3
GRUPO 5 Calif	56	63 ALTA	157 4	1 BAJA	21 1	\$12.000 BAJA	\$1.858.600 1	5% BAJA	39% 1	- MEDIA	272 3
GRUPO 3 Calif	49	168 BAJO	273 1	1 BAJA	10 1	\$10.850 BAJA	\$912.000 1	0% BAJA	59% 1	- BAJA	116 2

Fuente: elaboración propia.

resultados, en orden descendente, para cada grupo en función del rango de cada variable de análisis.

En este *ranking*, el grupo 1 concentra los clientes clave para la empresa, con el menor número de clientes, y mejores calificaciones. Tiene un solo atributo con calificación baja, sin embargo es de notar que es el único grupo que no tiene clientes que solo hayan realizado una compra en el periodo de estudio (frecuencia entre 6 y 125). El volumen de las compras de este grupo hace que la ganancia por cliente no sea tan alta (el máximo valor de la %G es de 38 %), pues al ser clientes mayoristas se ofrecen descuentos en el precio por volumen de compra.

El grupo 4, que es el segundo con más clientes asignados, tiene una frecuencia de compra baja y los montos de compra o monetario, también son bastante bajos. Pero tiene una ganancia por cliente bastante alta y es uno de los grupos con la mejor calificación en las variables R y DV. Al analizar con mayor detalle en la base de datos, los clientes de este grupo se encuentran en su mayoría en las afueras de la ciudad, compran periódicamente para abastecer sus negocios y son clientes que suelen pagar de contado. Los productos que suelen comprar son producidos por la empresa, por lo cual representan un alto margen de contribución para la misma, adicionalmente, al estar por fuera de la ciudad, la empresa no tiene tanta competencia con otros proveedores por estos clientes.

El grupo 2 tiene la mayor cantidad de clientes, y las variables muestran similitudes con el grupo 4. Pero los clientes de este grupo no tienen el potencial que se observa en los clientes del grupo 4, ya que la ganancia por cliente no es tan alta. Esto se debe a que compran casi todos los productos ofrecidos por la empresa, y algunos de estos productos tienen un margen por producto bajo, ya que son productos comercializados y no producidos por la empresa, pero son ofrecidos para que el cliente no utilice otro proveedor.

Los dos últimos dos grupos de clientes, 5 y 3, tienen las más bajas calificaciones. Aunque tienen clientes que generaron una ganancia para la

empresa bastante alta, estos clientes no han comprado desde hace más de dos meses.

Aunque análisis adicionales de la información que se tiene de cada grupo pueden aportar más información a tener en cuenta en la toma de decisiones, los anteriores resultados, sirven como punto de partida, para la determinación y diseño de estrategias encaminadas a establecer relaciones más cercanas con los clientes, identificando dentro del conjunto clientes con los que actualmente se está trabajando, a aquellos que pueden ser claves para el crecimiento de la empresa.

CONCLUSIONES

El análisis RFM es ampliamente utilizado por su fácil compresión y aplicabilidad. Permite conocer el comportamiento de compra de los clientes según unos parámetros básicos, en busca de la identificación de aquellos clientes que aportan mayor valor a la empresa. Adicionalmente, este análisis se puede complementar con otras técnicas como técnicas de manipulación de datos, métodos gráficos de visualización, métodos de agrupación, junto con la introducción de nuevas variables, para obtener un análisis adaptado a las necesidades particulares de una empresa o industria.

Las bases de datos de una empresa pueden disponer de gran información sobre sus clientes, lo cual ofrece la posibilidad de nuevos parámetros de evaluación, que pueden integrarse al análisis del comportamiento de los clientes. Pero la integración de nuevos parámetros debe responder a preguntas que sean de interés para empresa y que los parámetros del análisis RFM no responden. En el caso de estudio, se definieron tres parámetros que se ajustaban a unos puntos de interés para de la empresa.

Antes de realizar la adición de nuevos parámetros al análisis RFM, es importante validar si estos realmente aportan un valor agregado al método tradicional, de lo contrario, puede resultar en análisis redundantes y errados de los clientes. En el caso de estudio, una de las variables propuestas, la

G tiene una relación fuerte y positiva (variable dependiente) con la variable tradicional M (variable independiente), es decir que el comportamiento de la variable G se explica mediante el comportamiento de la variable M, razón por lo cual G se excluyó de las variables para realizar la segmentación. Esto no implica que no pueda ser utilizada en análisis posteriores de cada segmento.

Después de obtener la segmentación de los grupos, es importante realizar un análisis de los clientes que los componen, pues permite entender el comportamiento de las variables y así poder diseñar estrategias de *marketing* dirigidas a esos clientes, basadas en las características particulares de cada grupo.

REFERENCIAS

- Chuang, H.M. y Shen, C.C. (2008). A study on the applications of data mining techniques to enhance customer lifetime value — based on the department store industry. *International Conference on Machine Learning and Cybernetics* (pp. 168-173). Kunming: China.
- Durango, C.M.; Quintero, M.E. y Ruiz, C.A. (2015). Metodología para evaluar la madurez de la gestión del conocimiento en algunas grandes empresas colombianas. *Tecnura*, 19(43), 20-36. <https://doi.org/10.14483/udistrital.jour.tecnura.2015.1.a01>
- Giraldo, F.; León, E. y Gómez, J. (2013). Caracterización de flujos de datos usando algoritmos de agrupamiento. *Tecnura*, 17(37), 153-166. <https://doi.org/10.14483/udistrital.jour.tecnura.2013.3.a13>
- Khajvand, M.; Zolfaghari, K.; Ashoori, S. y Alizadeh, S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. *Procedia Computer Science*, 3, 57-63. <https://doi.org/10.1016/j.procs.2010.12.011>
- Kim, S.Y.; Jung, T.S.; Suh, E.H. y Hwang, H.S. (2006). Customer segmentation and strategy development based on customer. *Expert Systems with Applications*, 31(1), 101-107. <https://doi.org/10.1016/j.eswa.2005.09.004>
- Kohavi, R. y Parekh, R. (2004). *Proceedings of the 2004 SIAM International Conference on Data Mining*. Lake Buena Vista.
- McCarty, J.A. y Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of Business Research*, 60(6), 656-662. <https://doi.org/10.1016/j.jbusres.2006.06.015>
- Rueda, J.; Elles, C.; Sánchez, E.; González, A.L. y Rivilas, D.G. (2016). Identificación de patrones de variabilidad climática a partir de análisis de componentes principales, Fourier y clúster k-medias. *Tecnura*, 20(50), 55-68. <https://doi.org/10.14483/udistrital.jour.tecnura.2016.4.a04>
- Santandreu, E. (2002). *Manual del credit manager*. Madrid: Ediciones Gestión 2000.
- Shih, Y.Y. y Liu, C.Y. (2003). A method for customer lifetime value ranking — Combining the analytic hierarchy process and clustering analysis. *Database Marketing & Customer Strategy Management*, 11(2), 159-172. <https://doi.org/10.1057/palgrave.dbm.3240216>
- Toro, E.M.; Pérez, L.P. y Bernal, M.E. (2007). Reducción de la dimensionalidad con componentes principales y técnica de búsqueda de la proyección aplicada a la clasificación de nuevos datos. *Tecnura*, 11(21), 29-40.

