

TRABAJO FIN DE MASTER



ANÁLISIS DEL MODELO RFM SEGÚN EL MÉTODO CONVENCIONAL Y EL MÉTODO DE LAS 2-TUPLAS

MASTER EN MINERÍA DE DATOS E INTELIGENCIA DE NEGOCIOS
UNIVERSIDAD COMPLUTENSE DE MADRID

TUTOR: RAMÓN ALBERTO CARRASCO
ALUMNO: ROSA MARÍA MAS DÍAZ

FECHA: 17/06/2016



ÍNDICE

1. INTRODUCCIÓN	4
2. OBJETIVOS.....	5
3. PRELIMINARES. EL MODELO RFM	7
4. APLICACIÓN PRÁCTICA A E-COMMERCE DE RFM CONVENCIONAL VS MODELO 2-TUPLAS.....	10
4.1. MÉTODO CONVENCIONAL.....	14
4.2. MÉTODO DE LAS 2-TUPLAS	21
5. ANÁLISIS DE CLUSTERS SOBRE EL <i>RFM</i>	26
5.1. ANÁLISIS CLUSTERS PARA EL MÉTODO CONVENCIONAL.....	26
5.2. ANÁLISIS CLUSTERS PARA EL MÉTODO 2-TUPLAS	29
6. ACCIONES FUTURAS.....	31
7. CONCLUSIONES	32
8. BIBLIOGRAFIA.....	37
ANEXOS.....	39



RESUMEN

En el siguiente Trabajo de Fin de Máster se pone en práctica la Minería de Datos (*Data Mining*), llevando a cabo una investigación de *CRM (Customer Relationship Management)* en la cual se analizan los comportamientos de compra de los clientes de una empresa que comercializa solo por internet (*online*). Este negocio es de origen español y mediante estos análisis podremos saber principalmente cuántos tipos de clientes posee y cómo son sus hábitos de compra para poder clasificarlos.

Para ello, utilizaremos la segmentación *RFM (Recency, Frequency, Monetary)* que la calcularemos mediante dos metodologías muy importantes que son el *Método Convencional* y el *Método de las 2-Tuplas*.

En el primer método realizaremos una clasificación numérica mediante quintiles que se numerarán de 1 a 5 tanto para la *Recencia*, la *Frecuencia* y el *Valor Monetario*, con los que podremos determinar el comportamiento de compra de cada cliente.

En el segundo método veremos otra clasificación de los clientes más precisa, más detallada y con la ventaja que ofrece un valor lingüístico para poder entender mejor a que cluster pertenece cada cliente.

Finalmente, realizaremos unos análisis de clusters con el método de “*K-medias*” con diferentes segmentos (entre 5 y 7 segmentos) que nos permitirán distinguir cuántos tipos de clientes tiene este negocio y cómo son con respecto a su hábito de compra. Todo esto con el fin de dar respuesta a este negocio sobre cómo es el comportamiento de compra de cada cliente, cuáles son los más importantes, cuáles son los menos importantes, cuántos han dejado de comprar, etc.



1. INTRODUCCIÓN

El comercio electrónico ha evolucionado durante las últimas décadas, en el momento actual hay millones de transacciones electrónicas, es decir, compras que se realizan por internet (*online*). Los negocios que venden sus productos por internet se denominan *e-commerce*, y conllevan a millones de compras *online*. La gran ventaja que tienen estos negocios es que pueden comprar sus productos cualquier persona del mundo que tenga algún dispositivo que se pueda conectar a internet. De hecho, los consumidores pueden realizar las compras en cualquier lugar, ya sea con sus ordenadores en sus hogares, o bien con sus *Smartphones* en la calle, en una cafetería, en la playa, etc.

Según un estudio realizado por la “Asociación Danesa de Comercio Electrónico en 1999” (Jakob Nielsen, 1999), los principales motivos por los que la gente compra *online* son los siguientes:

- La comodidad y facilidad de uso
- Rapidez
- Eficiencia
- Placer de la compra en Internet
- Información detallada sobre los productos
- Ausencia de presión del vendedor

Estos negocios *online* luchan cada día por conseguir más clientes y fidelizarlos, y llegan a conseguir miles de clientes que visitan su web y acaban realizando una o varias compras.

Es por ello, que para este proyecto se pretende analizar el comportamiento de compra de los clientes de un negocio *online* real, en concreto, creado en España desde hace aproximadamente 3 años, el cual pone a disposición del consumidor productos y servicios locales con diferentes rangos de descuento, aumentando la capacidad de compra de los usuarios.

Además, elegimos estudiar a este comercio electrónico porque hasta el momento ellos no conocen cuál es el comportamiento de compra de sus clientes, es decir, vamos a analizar cada cuánto compran sus clientes, qué compran, cuánto compran, si repiten o



no compra año tras año, cuántos clientes tienen fieles, cuántos han dejado de comprar, etc.

Por tanto, vamos a utilizar la minería de datos para extraer toda la información posible y así poder darles respuesta a todas estas preguntas con el fin de que puedan mejorar su negocio en términos monetarios y en términos de satisfacción de sus clientes.

Para poder dar respuesta a estas preguntas vamos a utilizar un modelo llamado **RFM** que consiste en medir tres variables clave muy importantes:

- cuándo fue la última vez que compró cada cliente (*Recency*)
- cuántas veces ha comprado (*Frequency*)
- cuánto dinero se ha gastado en total (*Monetary*).

Además, para poder llevarlo a cabo, utilizaremos el software “R” que es una herramienta estadística muy útil para la minería de datos; aunque el modelo *RFM* también se podría calcular con más herramientas estadísticas como son: *SQL*, *ORACLE*, *SPSS*, *SAS*, entre otros.

2. OBJETIVOS

En cuanto a los objetivos de este estudio, planteamos lo que pretendemos conseguir con esta investigación aplicando el *RFM* para poder ayudar a este negocio.

Como hemos comentado, en esta investigación queremos dar a conocer cómo se comportan los clientes de esta empresa, y además, cuáles son los más y los menos fieles.

Por lo tanto, nuestro **objetivo principal** es: *conocer cuáles son los clientes más y menos rentables de este e-commerce segmentándolos por características similares de hábitos de compra.*



Ya que para este negocio lo más importante es conseguir fidelizar al menos el 70% de sus clientes dentro del sector en el que compran frecuentemente, y el 30% restante que sean nuevos.

También pretendemos conseguir algunos **objetivos secundarios** como los siguientes:

- *Conocer qué tipos de clientes tiene esta empresa y cuántos.*
- *Saber cuántos han dejado de comprar frecuentemente.*
- *Que la empresa pueda conocer aquellos segmentos de clientes que se gastan más y menos dinero.*
- *Poder aumentar el engagement en aquellos segmentos que son más esporádicos.*

Para este negocio es también importante incentivar al mayor número de compradores posible a que diversifiquen sus sectores de compra.

Entendemos como “sector de compra” la clasificación de sus productos por tipologías diferentes como son por ejemplo:

- Gastronomía
- Salud y Belleza
- Viajes y eventos
- Formación
- Servicios profesionales
- Shopping

Por lo tanto, aplicando esta técnica de segmentación de clientes, el *RFM*, podremos responder a estos objetivos analizando las siguientes preguntas:

- Conocer cuántos clientes tiene la empresa hasta el momento.
- Saber con qué frecuencia realizan una o varias compras.
- Saber si esos clientes que compran volverán a comprar posteriormente.
- Saber a qué clientes vale la pena retener.
- Saber qué clientes pueden responder mejor a una determinada promoción.



De esta forma se pueden establecer estrategias por cliente y/o segmento que:

- Hagan al cliente volver a comprar cuanto antes.
- Que consigan incrementar su ticket medio, es decir, que compren más productos.
- Que aumenten la frecuencia de visita y compra en la web.

Además, también podremos saber en qué clientes deberían invertir más, incentivando su compra, y en cuales no merece la pena.

3. PRELIMINARES. EL MODELO RFM

El modelo *RFM* es una metodología con la que podemos segmentar clientes según características similares. Se creó hace más de 75 años, principalmente para los vendedores directos, para poder satisfacer a las finanzas mejorando los beneficios. Fue muy popular para los pioneros del marketing de base de datos (Stan Rapp, Tom Collins, David Pastor, Arthur Hughes, etc).

Este modelo contempla la *Recencia*, *Frecuencia* y *Valor Monetario* para cada cliente a partir del cual determinamos el comportamiento o evolución de compra de éstos. Lo gran ventaja es que es un modelo fácil de entender, de explicar e implementar.

El *RFM* sigue la premisa de que *“los más propensos a comprar son aquellos que han comprado más recientemente, con más frecuencia y gastan más dinero”*.

Se basa en la conocida “Ley de Pareto” o del “80/20” enunciada por el economista italiano Vilfredo Pareto, en el siglo XIX.

El *RFM* se aplica sobre esta “Ley de Pareto” y se refiere a que *“el 80% de las compras las realizan el 20% de los clientes”*.

Podemos ver en la Figura 1 cómo se dividen a los clientes en quintiles (20%), siendo el último 20% el que más rentabilidad genera:

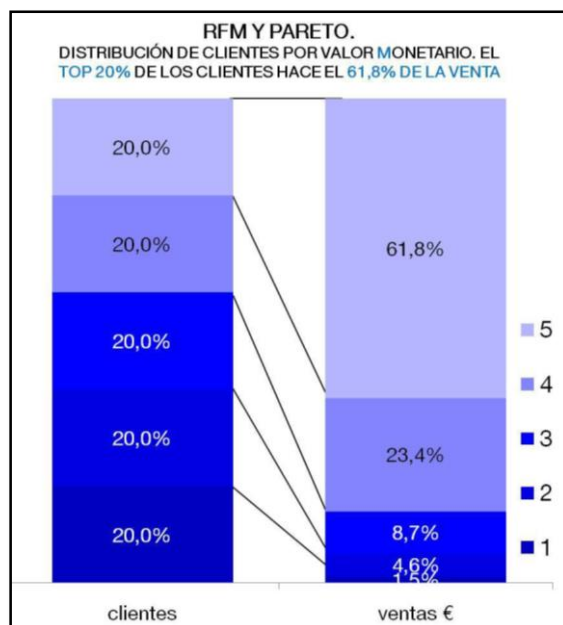


Figura 1: RFM y PARETO, Distribución clientes

Como hemos comentado anteriormente:

- Con la *Recencia (Recency)* medimos los **días** que han pasado desde hoy (o cualquier fecha a futuro) hasta la fecha en que realizó su última compra.
- Con la *Frecuencia (Frequency)* medimos el **número de compras** que ha hecho cada cliente en total.
- Y el *Valor Monetario (Monetary)* es la suma total de **cantidad de dinero** que el cliente lleva gastado en sus compras.

Con el *RFM* podemos construir escalas, basadas en estas tres variables, dando a cada cliente un valor según el percentil en el que se encuentre. Lo más común es escalar por quintiles, es decir, tanto a la *Recencia* como a la *Frecuencia* como al *Valor Monetario* les asignamos un valor del 1 al 5, siendo 1 la peor puntuación y 5 la mejor, como una escala de “Likert”.

Por tanto, para cada una de las tres variables los clientes cuya última compra sea más reciente, más frecuente y cuya suma de gastos sea la más elevada tendrán una puntuación de **(R=5) (F=5) (M=5)**. Con estas puntuaciones se dividen a los clientes en 20%, tratándose de quintiles.



Un ejemplo para entenderlo mejor es el siguiente: *Un cliente que estuviera en el último 20% de los que más recientemente han comprado y en el segundo 20% por frecuencia de compra y en el primer 20% de los que menos se han gastado, se le asignaría el segmento 521, es decir, (R=5) (F=2) (M=1).*

Lo podemos visualizar en la Figura 2:

SEGMENTOS RFM (hasta 5*5*5 =125 segmentos diferentes)					
5	Más Reciente	5	Más Frecuente	5	Mayor gasto
4		4		4	
3		3		3	
2		2		2	
1	Más Antiguo	1	Menos Frecuente	1	Menos gasto

Ejemplo: "5 2 1"

Figura 2: Segmentos RFM

De esta manera podemos afirmar que:

- Los clientes que compran de forma reciente son más favorables a comprar que aquellos que no lo han hecho últimamente.
- Los clientes que compran más frecuentemente están más dispuestos a comprar nuevamente que aquellos que han hecho una o dos compras.
- Los clientes que gastan más, están más dispuestos a comprar nuevamente.
- Los clientes más valiosos son aquellos que pueden llegar hacerlo aún más.

Pero el *RFM*, como todo, no es 100% perfecto, porque no sirve para detectar lo siguiente:

- El cambio migratorio de cada cliente de un valor a otro (de un quintil a otro quintil) cada vez que se actualice el análisis. Sólo nos dice el número de clientes en un segmento "x", pero no quiénes. Es difícil comprender por qué hay clientes



que de un mes a otro o de un año a otro, cambian su frecuencia de compra, su recencia y su gasto.

- Además, tampoco detalla cuándo un cliente es nuevo (de hace una semana, o de hace un mes), porque se puede concluir que si la *Frecuencia* es baja y la *Recencia* es alta podría ser nuevo, pero hemos comprobado en este estudio que se mezclan con clientes con fecha de registro muy antigua que han empezado a comprar este año por primera vez. Realmente no es nuevo en la base de datos pero sí en comprar.

Entonces, la manera de ver evoluciones o cambios migratorios, o de saber de qué tipo de cliente se trata en cada caso, sería cruzar la información del *RFM* con la base de datos original y ver en detalle todo su historial.

Para este análisis, vamos a calcular el *RFM* con dos métodos diferentes que son: el “*Método Convencional*” y el “*Método de las 2-Tuplas*” que definiremos a continuación.

4. APLICACIÓN PRÁCTICA A E-COMMERCE DE RFM CONVENCIONAL VS MODELO 2-TUPLAS

De este negocio *online* que vamos a analizar, hemos extraído la base de datos de compras de su *Back-Office*, que es una herramienta donde tienen toda la información principal de su negocio, con un período de tiempo desde el 1 de enero de 2013 hasta el 1 de junio de 2016.

Hasta la fecha existen 15.497 usuarios únicos que han realizado alguna compra. Las variables necesarias para el cálculo del *RFM* son las siguientes:

- Identificador de cliente: es un código numérico para poder identificar a los usuarios.



- Código de la transacción: es un código alfanumérico único por cada compra. Si en una compra hay varios productos iguales o diferentes, a todos estos se les asigna el mismo código de compra.
- Fechas de compras: son todas las fechas de compras por usuario y transacción.
- Precio o gasto: es el valor monetario de cada compra por fecha y usuario.

De tal manera que un cliente está almacenado en la base de datos más de una vez siempre que haya comprado varias veces aunque sea en fechas distintas y con códigos de compras diferentes según la fecha.

Antes de empezar con el análisis del *RFM*, realizaremos un pequeño análisis descriptivo previo para cada variable con el fin de conocer cómo se distribuyen estos clientes.

- **Para la variable “Id Cliente”:**

Obtenemos el siguiente histograma, en la Figura 3, donde podemos ver que la mayoría de los clientes tienen el “id” entre “30000” y “35000”, y poco más de 400 clientes tienen un “id” hasta “15000”. El “id” se asigna de menor a mayor según se van registrando clientes nuevos, por lo que hay más clientes compradores que se registraron sobre el 2015 aproximadamente que mas antiguos.

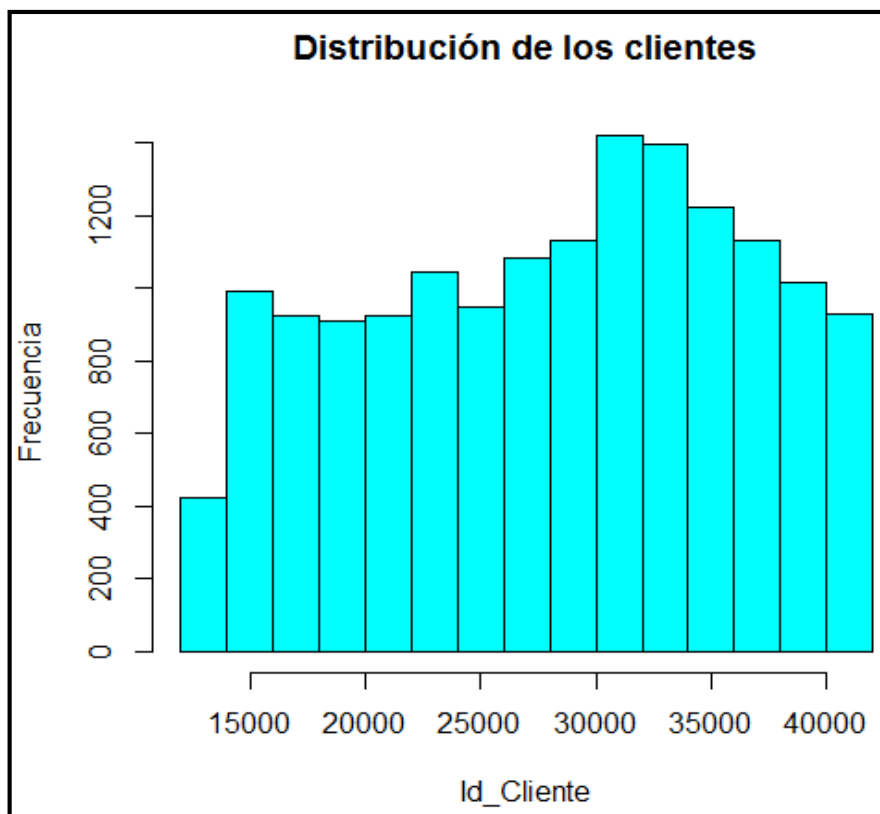


Figura 3: Distribución de la frecuencia de los clientes

- Para la variable “Código de Transacción”:

Con “R” podemos obtener la siguiente tabla de estadísticos descriptivos, en la Figura 4, para identificar principalmente cuáles son los valores mínimo y máximo de compras que se han hecho, y cuál es la media de compra:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.0	1.0	2.0	3.04	3.0	125.0

Figura 4: Tabla de estadísticos para la variable “Código de Transacción”

Como vemos, el mínimo de compras es 1 y el máximo 125, y la media es de 2 compras.

Además, obtenemos el siguiente histograma que nos muestra la frecuencia de número de compras que se han realizado en total. Podemos ver en la Figura 5 que la mayoría de los clientes han realizado entre 0 y 10 compras, y una gran minoría han hecho compras superiores hasta 125 que es el máximo.

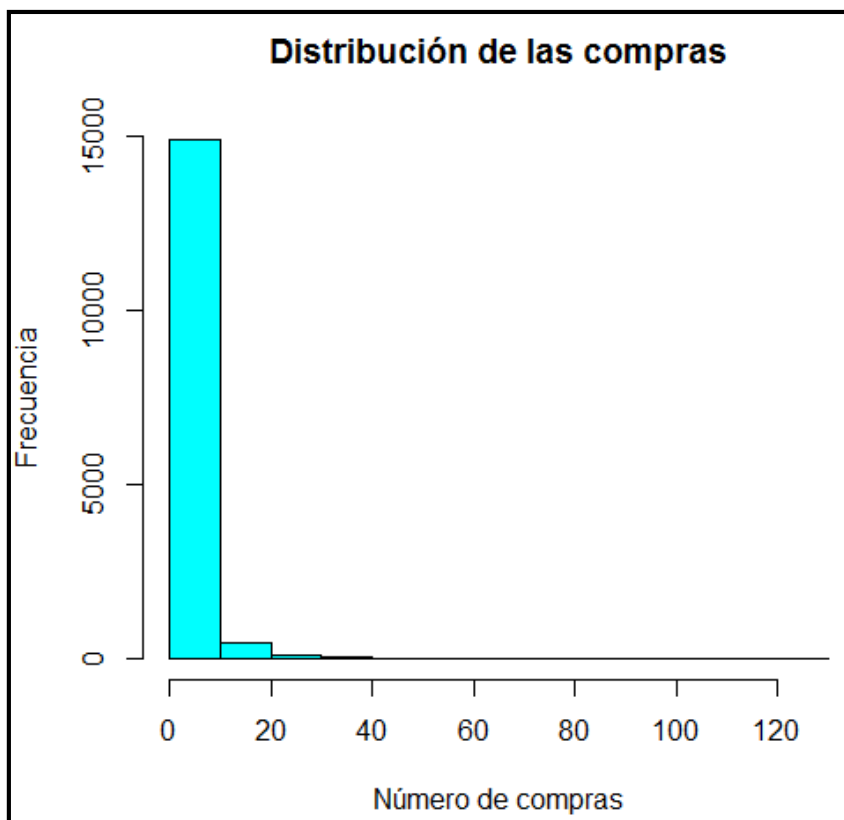


Figura 5: Gráfico de distribución de la variable “Código de Transacción”

- **Para la variable “Fecha”:**

Como hemos comentado antes, la fecha mínima es el 1 de enero de 2013 y la fecha máxima es el 1 de junio de 2016. Además, la fecha en la que más compras se han realizado es el 4 de marzo de 2016 donde hubo 468 compras.

- **Para la variable “Precio o gasto”:**

Para la variable “Precio” obtenemos también las medidas estadísticas más comunes para saber cuál es el precio mínimo que se han gastado estos clientes y el precio máximo:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2	5	12	17,9	20	279

Figura 6: Tabla de estadísticos de la variable “Precio”

Podemos ver en la Figura 6, que el precio mínimo son 2€ y el máximo que ha pagado algún cliente son 279€. Además, el precio medio ronda los 17,9€ a lo largo de todo este período.



Ahora que ya conocemos mejor cómo es esta base de datos, procedemos a realizar el análisis *RFM*.

4.1. MÉTODO CONVENCIONAL

El *Método Convencional* es una metodología para el calcular el *RFM* y para obtener la segmentación de los clientes. Como hemos comentado anteriormente, la manera más común de calcularlo es haciendo particiones en quintiles (20%), pero también se puede calcular haciendo particiones en deciles (10%), etc., según el mejor ajuste para cada base de datos.

En este estudio, vamos a calcularlo con quintiles porque es como mejor se ajusta a nuestros clientes, y los clasificaremos con un valor del 1 al 5.

Esta base de datos se extrae por defecto en un fichero “Excel” el cual tenemos que importarlo en el programa “R” para poder realizar los análisis. Una vez importado, podemos ver las 6 primeras filas en la Figura 7:

Fecha	Compras	Cliente	Precio
15/01/2013	af5ad4e8eea5af89544b0d42ae83c0	13153	32
15/01/2013	b582012d9768fadca92f67ad5d7133	13167	32
15/01/2013	b582012d9768fadca92f67ad5d7133	13167	32
15/01/2013	b582012d9768fadca92f67ad5d7133	13167	32
15/01/2013	d660a568405dce2c989e9c5df5bf94	13174	32
15/01/2013	d660a568405dce2c989e9c5df5bf94	13174	32

Figura 7: Muestra de la base de datos original

Solo mostramos las 6 primeras filas como muestra, y como podemos ver:

- El cliente “13153” ha realizado una compra el día 15 de enero de 2013 y se ha gastado 32 euros.
- El cliente “13176” ha hecho tres compras el mismo día 15 de enero de 2013 y se ha gastado 32 euros en cada compra ese día.
- Y el cliente “13174” ha realizado 2 compras del mismo importe y el mismo día.



El problema que se presenta ahora es hacer reconocer en el programa “R” que la variable “Fecha” sea reconocida como formato de fecha, ya que lo reconoce a priori como carácter o texto, para poder calcular la *Recencia*. Para ello, con la función “as.Date” lo definimos.

Una vez definida, procedemos a calcular las tres variables, y para ello, utilizaremos el paquete “sqlite” que nos permitirá hacer una consulta como si fuera “SQL” para que sea más práctico y sencillo.

- Para la *Recencia*, escribimos “min(días)” para que nos devuelva el mínimo de días que han pasado entre la fecha actual, que es el 02/06/2016, y la del último día que ha comprado cada cliente por última vez.
- Para la *Frecuencia*, escribimos “count(Transacciones)”, es decir, sólo necesitamos saber cuántas transacciones (compras) ha hecho cada cliente.
- Por último, para el *Valor Monetario*, calcularemos “sum(Precio)”, es decir, la suma total de gastos de cada cliente.

Una vez calculadas estas tres variables, lo que obtenemos es una tabla, que vemos en la Figura 8:

Cliente	Recencia	Frecuencia	Monetary
1	210	15	158
2	401	2	98
3	6	125	2682
4	50	58	939
5	1234	1	32
6	359	19	213

Figura 8: RFM valores agrupados

En la Figura 8 se muestran sólo los valores del *RFM* para los mismos 6 clientes de la Figura 7.

A priori, podemos ver que por ejemplo, el Cliente 3 tiene una *Recencia* muy baja de 6 días desde que realizó su última compra; su *Frecuencia* es de 125 veces que ha comprado y su *Monetary* son 2.682€. Como conclusión, podríamos deducir que es



un cliente “Muy bueno” ya que hace muy poco que compró por última vez, ha comprado muchas veces y se ha gastado mucho dinero.

Por otro lado, el Cliente 5 tiene una *Recencia* muy alta de 1.234 días desde que compró por última vez, solamente ha hecho una compra y se ha gastado 32€. Por tanto, intuimos que este cliente compró una vez hace más de 3 años y no ha vuelto a repetir más, por lo que es un cliente “perdido”.

Pero esta visión es muy genérica y no vemos realmente una clasificación de clientes hasta ahora. Para ello, utilizaremos el *Método Convencional* con el fin de poder clasificarlos.

Vamos a empezar calculando quintiles sobre estos clientes tanto para la *Recencia*, *Frecuencia* y *Valor Monetario*.

Una vez calculados obtenemos una tabla como la que vemos en la Figura 9:

Cliente	Recencia	Frecuencia	Monetary
1	4	5	5
2	4	2	5
3	5	5	5
4	5	5	5
5	1	1	3
6	4	5	5

Figura 9: Quintiles aplicados por el Método Convencional

La Figura 9 muestra los quintiles asociados a los mismos clientes que vimos en las Figuras 7 y 8.

Podemos ver que los clientes 1, 3, 4 y 6 son muy buenos porque tienen todos los valores muy altos, ya que del 1 al 5 los valores cercanos a 1 son peores y los más cercanos a 5 son mejores.

El Cliente 5 al tener la *Recencia* muy alta y *Frecuencia* muy baja en la Figura 8 ahora está clasificado con un “1”.

Con este método ya hemos agrupado a los clientes en quintiles y ahora es más intuitivo entender su comportamiento de compra en los últimos años.



Además, podemos hacer un gráfico de dispersión para ver cómo se distribuyen los quintiles sobre todos los clientes. Con “R” realizamos los gráficos y los vemos en las Figuras 10, 11 y 12:

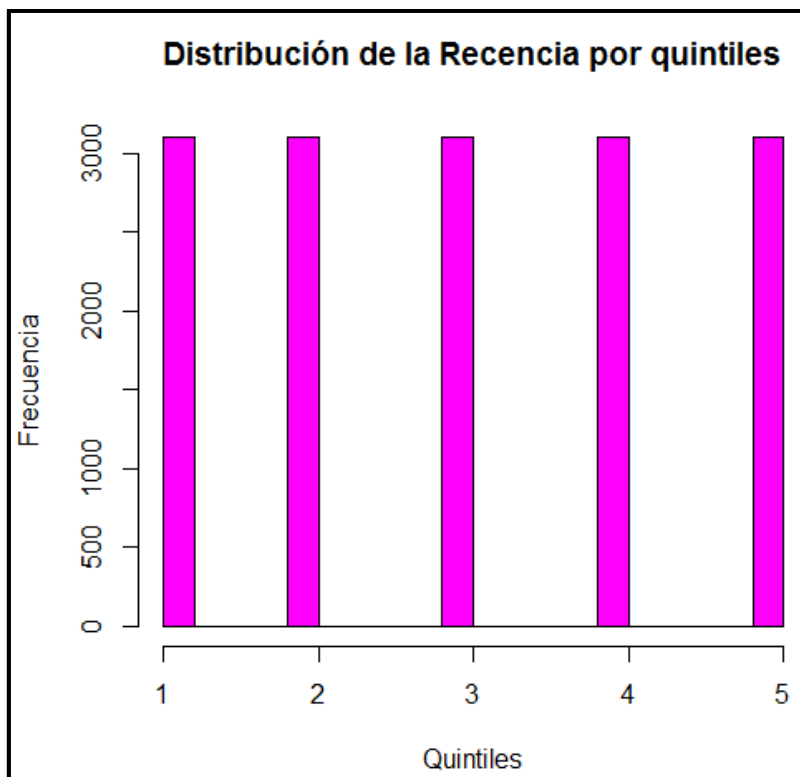


Figura 10: Distribución de los quintiles según la Recencia

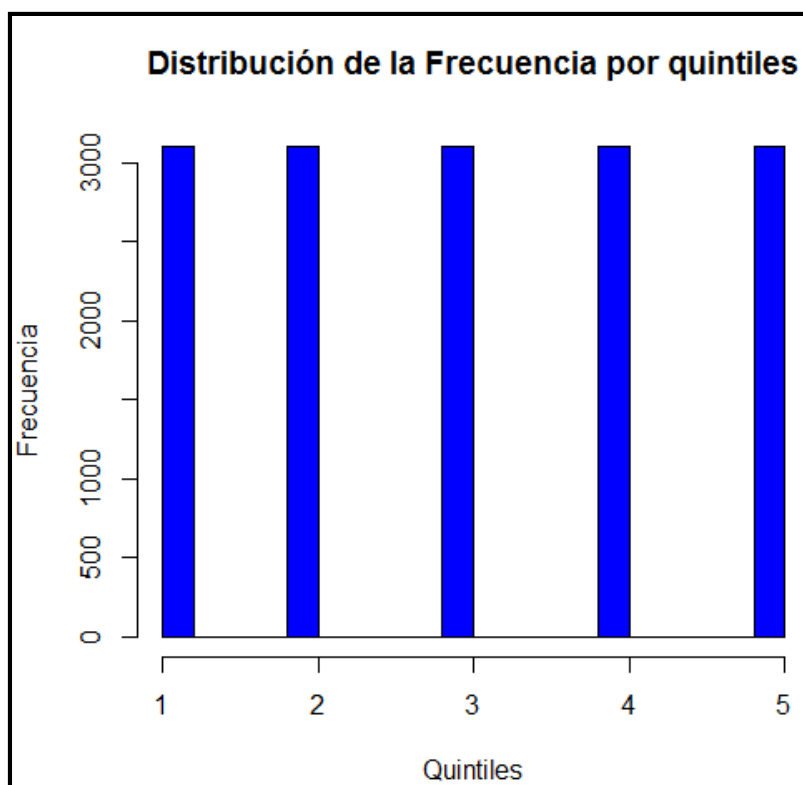


Figura 11: Distribución de los quintiles según la Frecuencia

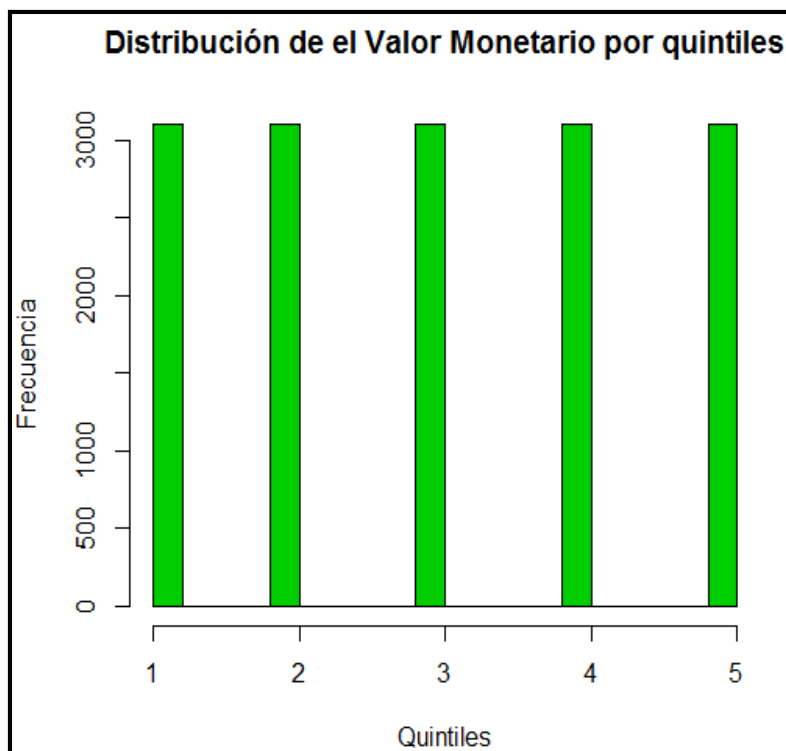


Figura 12: Distribución de los quintiles según el Valor Monetario

En los tres gráficos anteriores podemos observar que los quintiles están bien calculados puesto que las tres variables se distribuyen uniformemente. Las particiones por el 20% recogen como máximo a 3.100 clientes.

Pero también nos planteamos la siguiente pregunta: *¿qué significa realmente una Recencia igual a 5, una Frecuencia igual a 4 y un Valor Monetario igual a 3?*

Es importante conocer cómo se han establecido los cortes de los quintiles en la base de datos para saber entre qué números se asigna cada quintil. Por ello, vamos a ver los valores mínimos y máximos para cada una de las variables y así sabremos en torno a qué números significa una *Recencia* igual a 5 ó a 1; al igual que para la *Frecuencia* y para el *Valor Monetario*:

RECENCY									
1		2		3		4		5	
Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
1002	1234	719	1002	461	718	202	461	1	202

Figura 13: Valores mínimos y máximos para cada quintil de la Recencia



FREQUENCY									
1		2		3		4		5	
Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
1	1	1	2	2	2	2	4	4	125

Figura 14: Valores mínimos y máximos para cada quintil de la Frecuencia

MONETARY									
1		2		3		4		5	
Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
2	16	16	25	25	39	39	69	69	2682

Figura 15: Valores mínimos y máximos para cada quintil del Valor Monetario

Por ejemplo, podemos ver que para la *Recencia* el quintil 5 oscila entre 1 día y 202 días que compraron por última vez. En la *Frecuencia*, el quintil 5 contempla el mayor número de compras entre 4 y 125. Existe mucha variabilidad en este quintil pero es porque muy pocas personas hicieron tantas compras, lo vimos en la Figura 5. Y en el *Valor Monetario* el quintil 5 oscila entre 69€ y 2.682€, por el mismo motivo muy pocos clientes se han gastado tanto dinero.

Además, en la Figura 16 podemos visualizar el conteo de todos los clientes por *Recencia*, *Frecuencia* y *Valor Monetario* y por quintiles para ver cómo se distribuyen:



REGENCY	FRECUENCY	MONETARY				
		1	2	3	4	5
1	1	949	337	408	208	86
	2	16	55	77	46	18
	3	137	116	98	51	84
	4	30	50	78	93	50
	5		5	9	41	57
2	1	205	296	265	152	125
	2	153	94	112	76	36
	3	123	117	187	227	117
	4	86	72	62	122	164
	5	11	28	34	57	178
3	1	12	15	11	6	5
	2	314	314	193	179	126
	3	121	289	233	171	71
	4	99	71	78	149	118
	5		35	137	126	227
4	1		4	3	1	2
	2	222	209	110	156	57
	3	111	292	148	177	169
	4	105	94	162	130	139
	5		101	120	243	344
5	1	3	3	3	1	
	2	144	77	89	117	109
	3	3	10	16	19	13
	4	256	262	265	216	148
	5		153	202	335	656

Figura 16: Visualización Conjunta de los clientes por quintiles y por el Método Convencional

La Figura 16 indica el recuento de clientes por cada quintil asignado a la *Recencia*, a la *Frecuencia* y al *Valor Monetario*.

En esta tabla podemos ver que este *e-commerce* tiene 949 clientes con (R=1) (F=1) (M=1), y 656 clientes con (R=5) (F=5) (M=5), por lo que concluimos que este negocio tiene más clientes “malos” que “buenos” desgraciadamente.

A su vez, podemos concluir que la mayoría de los clientes se distribuyen entre “muy fieles” o “nada fieles”.

Además, podemos concluir más situaciones como:

- No existen clientes que tengan una *Recencia* muy baja (R=1), es decir, que hayan comprado hace mucho tiempo, con una *Frecuencia* muy alta (F=5), es decir, que hayan comprado muchas veces.
- Tampoco existen clientes con (R=5) y (F=5) y que se gasten poco dinero (M=1), ya que aunque no se gasten mucho dinero, lo bueno sería que repitieran compra.



Como vemos, podemos obtener hasta 125 segmentos ($5*5*5$) de clientes que nos permiten saber cuántos existen de cada tipo para después accionar sobre ellos.

Ya hemos visto cómo es la metodología del *Método Convencional* y lo que conseguimos con esta mecánica. Ahora vamos a realizar otro análisis *RFM* de segmentación de clientes con el *Método de las 2-Tuplas* que es mucho más preciso y eficaz y lo detallamos a continuación.

4.2. MÉTODO DE LAS 2-TUPLAS

Otra manera de calcular el *RFM* es mediante el *Método de las 2-Tuplas*, el cual introduce mejoras con respecto al *Método Convencional* ya que aporta una parte lingüística que con el otro método no obtenemos, y además, es mucho más preciso ya que en vez de solo cuantificar a los clientes con valores de 1 a 5, nos aporta el valor exacto con decimales. Por lo tanto, sabemos mediante la parte lingüística a qué cluster pertenece y con la parte decimal su valor exacto.

Para realizar el análisis del *RFM* por el *Método de las 2-Tuplas* hemos tenido como base, y seguido paso a paso, las aportaciones del artículo *A 2-tuple Fuzzy Linguistic Approach Applied to RFM Models. An Application to Customer Segmentation* escrito por Ramón Alberto Carrasco, María Francisca Blasco, Jesús García-Madariaga, Enrique Herrera-Viedma.

Este modelo sirve también para detectar segmentos de clientes al igual que el método anterior pero siguiendo otra mecánica diferente. Se ha utilizado en muchas aplicaciones comerciales y de gestión y lleva a cabo procesos de "computación con palabras" sin la pérdida de información que son típicos de otros enfoques lingüísticos difusos.

Consta principalmente de dos partes: una parte lingüística y una parte numérica. Estas dos partes las definimos a continuación. La metodología a seguir es la siguiente:



Primero, dividimos en cien partes iguales a los clientes con una previa normalización de los datos a los que después los clasificaremos de 1 a 5, para que sea comparable con el método anterior cuando aplicábamos los quintiles.

Una vez aplicada la normalización y la clasificación, detallamos todos los parámetros y funciones que determinan este método:

1. Consideramos un conjunto $S = \{s_1, \dots, s_T\}$ como un término lingüístico y “ b ” $\in [0, T]$ que es un valor que representa el resultado de una operación de agregación simbólica y es obtenida con la siguiente función:

$$\Delta: [0, T] \rightarrow S \times [-0.5, 0.5)$$

$$\Delta(b) = (s_i, \alpha), \text{ with } \begin{cases} s_i, & i = \text{round}(b) \\ \alpha = b - i, & \alpha \in [-0.5, 0.5). \end{cases}$$

Donde la función $\Delta(b)$ engloba el conjunto “ S ” + la parte “ α ” que explicamos a continuación:

- “ b ” es el valor real que obtenemos con la clasificación de la normalización, es un número con decimales.
- “ i ” es el redondeo al entero más cercano del valor de “ b ”.
- “ α ” es la parte decimal que está entre (-0,5 y 0,5).

Podemos entenderlo mejor con un ejemplo: *Si un cliente tiene un valor “ b ” de 3,7; “ i ” será el valor redondeado, es decir, será 4, y “ α ” es $4 - 3,7 = 0,3$ que es la desviación de ese cliente para llegar a tener un valor de 4.*

Además, para todo Δ , también existe la función inversa $\Delta^{-1}(b) = i + \alpha$, que devuelve el valor inicial, es decir, devuelve el valor de “ b ” que en el ejemplo anterior sería igual a 3,7.

2. Añadido al punto 1, daremos un significado semántico, que será una manera de clasificarlos, que definimos dentro del conjunto $S = \{s_1, \dots, s_T\}$, donde $T=5$: $s_1 =$ “VeryLow (VL)”, $s_2 =$ “Low (L)”, $s_3 =$ “Medium (M)”, $s_4 =$ “High (H)” y $s_5 =$ “Very High (VH)”; lo vemos en la Figura 17:

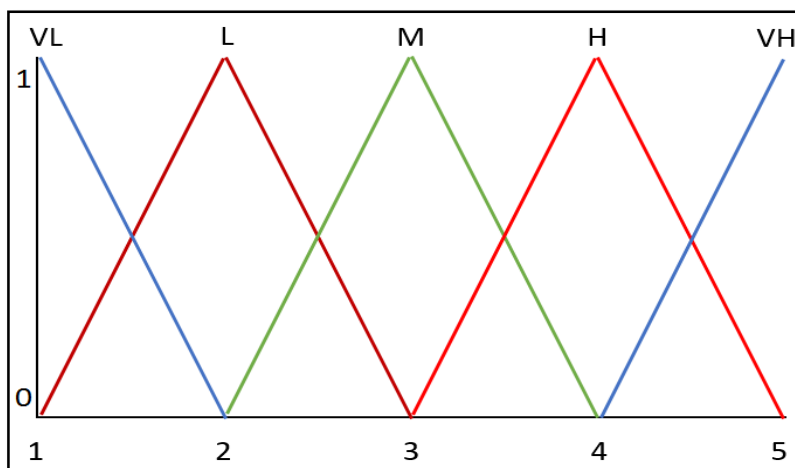


Figura 17: Definición del conjunto S

La Figura 17 indica cómo se distribuyen los segmentos del conjunto “ S ”, es decir, si los cuantificamos el subconjunto s_1 será “1”, s_2 será “2”, s_3 será “3”, s_4 será “4” y s_5 será “5”.

Por lo tanto, como decíamos al principio, la primera parte corresponde a los valores lingüísticos que son de “VL” a “VH”, y la parte numérica será el valor de “ α ” que nos indica la desviación del cluster al que pertenecería el cliente con respecto a su cluster inicial.

Con “ R ” podemos obtener cada uno de estos valores para cada cliente:

- Los valores de “ b ”:

Cliente	Recencia	Frecuencia	Monetary
1	4,15	4,96	4,80
2	3,63	2,54	4,52
3	5,00	5,00	5,00
4	4,80	5,00	5,00
5	1,00	1,00	3,02
6	3,75	5,00	4,92

Figura 18: Datos obtenidos de “ b ”

Los valores de “ b ” son los resultados para cada cliente una vez normalizados y clasificados de 1 a 5, donde 1 es muy malo, y 5 muy bueno.



- Los valores de “i”:

Cliente	Recencia	Frecuencia	Monetary
1	4	5	5
2	4	3	5
3	5	5	5
4	5	5	5
5	1	1	3
6	4	5	5

Figura 19: Datos obtenidos de “i”

Estos son los valores de “b” redondeados al entero más cercano.

- Los valores de “α”:

Cliente	Recencia	Frecuencia	Monetary
1	0,15	-0,04	-0,20
2	-0,37	-0,46	-0,48
3	0,00	0,00	0,00
4	-0,20	0,00	0,00
5	0,00	0,00	0,02
6	-0,25	0,00	-0,08

Figura 20: Datos obtenidos de “α”

Estas son las desviaciones, es decir, la diferencia entre su valor real y el redondeado para saber cuánto dista del cluster que se le asigna al cluster que debería pertenecer realmente.

Por ejemplo, el Cliente 2 tiene en *Recencia* una desviación de -0,37 lo que quiere decir que es una desviación alta ya que las desviaciones se mueven entre -0,5 y 0,5.

Pero aparte de la desviación necesitamos obtener el **valor lingüístico** para identificar el tipo de cluster al que se refiere.

Cliente	Recencia	Frecuencia	Monetary
1	(H,0.15)	(VH,-0.04)	(VH,-0.2)
2	(H,-0.37)	(M,-0.46)	(VH,-0.48)
3	(VH,0)	(VH,0)	(VH,0)
4	(VH,-0.2)	(VH,0)	(VH,0)
5	(VL,0)	(VL,0)	(M,0.02)
6	(H,-0.25)	(VH,0)	(VH,-0.08)

Figura 21: Resultado RFM mediante el Método 2-Tuplas



En la Figura 21 vemos el resultado final del *Método de 2-Tuplas* que nos indica lo siguiente:

El Cliente 3 es “Very High” en las tres variables con una desviación de 0 lo que quiere decir que claramente que es “Very High” 100% sin ninguna desviación. Si recordamos en el método anterior el mismo cliente tenía las tres variables clasificadas con un 5 que es la mayor puntuación y en este método también vemos que el resultado es el mismo.

El Cliente 4, se diferencia del 3 en la *Recencia*, donde existe una pequeña desviación de -0,2 lo que indica que no es “Very High” 100% y lo podríamos clasificar como “High” que es una categoría menos puesto que es negativo. Si lo comparamos con el *Método Convencional* este cliente también tenía la *Recencia* clasificada con un 5 y ahora vemos la precisión que tiene este método puesto que no se clasifica al 100% con un 5 en *Recencia*.

El Cliente 2 sin embargo, tiene las desviaciones muy altas en las tres variables, por lo que en un principio se clasifica como “High” en *Recencia*, “Medium” en *Frecuencia* y “Very High” en *Monetary*, debería ser clasificado como “Medium” en *Recencia*, “Low” en *Frecuencia* y “High” en *Monetary*, puesto que todas las desviaciones son negativas.

Los Clientes 1, 5 y 6 tienen las desviaciones muy pequeñas por lo que pertenecen bien a los valores lingüísticos que se indican.

Ya hemos visto cómo funcionan estos dos métodos, cuáles son sus procedimientos, cómo acaban clasificando y cómo se interpretan.

A continuación, vamos a hacer un análisis de clusters para agrupar e identificar comportamientos de compra similares como comentábamos al principio.



5. ANÁLISIS DE CLUSTERS SOBRE EL RFM

Vamos a proceder a realizar el análisis de cluster para cada método estudiado. En este análisis veremos los promedios o centroides obtenidos para cada variable (*Recencia, Frecuencia y Valor Monetario*) con los que definiremos las características de cada cluster.

Además, obtenemos las siguientes columnas:

- La columna C_s corresponde al número de clientes que pertenecen a ese cluster. La suma de los clientes en cada cluster debe coincidir con el total de clientes.
- Los valores de *Recency Score*, *Frequency Score* y *Monetary Score* son los centroides de cada cluster, es decir, los valores promedios.
- En la columna Customer Type definiremos qué tipo de clientes son en cada cluster.

5.1. ANÁLISIS CLUSTERS PARA EL MÉTODO CONVENCIONAL

A continuación, vamos a ver las diferencias entre los diferentes clusters de clientes variando el número de clusters de 5 a 7, por lo que al aumentar el número de clusters se detallará más y veremos cuáles son las características que mejor los definen.

Los valores de los centroides oscilan de 1 a 5 ya que es la clasificación de quintiles hecha previamente. Entonces, para poder comentar los resultados cuanto más cerca esté el promedio de 1 peor será ese cluster y cuanto más cerca esté de 5 mejor será. Al igual si ronda el valor 2,5 significará un hábito de compra normal o medio.

Procedemos a analizar los siguientes clusters:

- **5 clusters:**

S	C_s	Recency Score	Frequency Score	Monetary Score	CustomerType
Cluster 1	3850	1.39	1.41	1.99	WORST
Cluster 2	2337	4.14	4.34	2.17	GOOD
Cluster 3	3230	3.68	2.36	2.14	MEDIUM
Cluster 4	2854	1.98	3.21	4.19	LOST
Cluster 5	3226	4.32	4.38	4.61	VIP

Figura 22: Análisis 5 clusters. Método Convencional



En este caso:

- El **Cluster 1** recoge a 3.850 clientes donde tanto la *Recencia* como la *Frecuencia* como el *Valor Monetario* son los más bajos por lo que es el grupo de los peores clientes.
- El **Cluster 2** recoge a 2.337 clientes donde la *Recencia* y la *Frecuencia* son muy altas pero el *Valor Monetario* es medio, por lo que aunque no se gasten mucho dinero son clientes que repiten compra, así que podemos clasificarlos como “Buenos”.
- El **Cluster 3** recoge a 3.230 clientes donde la *Recencia* es medio buena y la *Frecuencia* y el *Valor monetario* son medios, por lo que los podemos clasificar como “medios o normales o de uso de compra normal”.
- El **Cluster 4** recoge a 2.854 clientes donde la *Recencia* es muy baja pero su *Frecuencia* y *Valor Monetario* es bueno, por lo que intuimos que se refiere a clientes que hemos perdido hace tiempo y eran buenos.
- El **Cluster 5** recoge a 3.226 clientes donde todas las variables tienen mayores promedios que el resto por lo que los clasificaremos como “VIP o mejores”.

- **6 clusters:**

S	C _s	Recency Score	Frequency Score	Monetary Score	CustomerType
Cluster 1	2350	3.73	2.41	3.71	MEDIUM
Cluster 2	2451	4.47	4.73	4.62	VIP
Cluster 3	2472	2.00	3.96	3.91	LOST
Cluster 4	4001	1.97	1.83	1.41	WORST
Cluster 5	1631	1.50	1.22	3.63	SLEPT
Cluster 6	2592	4.45	3.90	2.01	LOYAL

Figura 23: Análisis 6 clusters. Método Convencional

Con 6 clusters hemos detectado dos posibles clusters nuevos que son:

- el **Cluster 5**, con 1.695 clientes (el grupo más pequeño) que hemos denominado “SLEPT” que hace referencia a los que están dormidos, es decir, compraban al



principio y se gastaban una cantidad razonable de dinero pero de repente dejaron de tener frecuencia en la compra.

- el **Cluster 6**, con 2.592 clientes que denominamos “LOYAL”, es decir, son los fieles que compran recientemente y con buena frecuencia, pero se gastan poco dinero. Lo importante es que repitan compra aunque sea con poco gasto.

- **7 clusters:**

S	C _s	Recency Score	Frequency Score	Monetary Score	CustomerType
Cluster 1	3175	1.35	1.16	1.99	WORST
Cluster 2	1896	2.07	3.60	2.23	MEDIUM
Cluster 3	2483	3.70	2.37	1.77	BAD
Cluster 4	2992	4.62	4.62	4.18	VIP
Cluster 5	1000	4.70	4.22	1.65	LOYAL
Cluster 6	1634	2.15	4.27	4.61	LOST
Cluster 7	2317	3.04	2.18	4.26	SPORADIC

Figura 24: Análisis 7 clusters. Método Convencional

Con 7 clusters obtenemos dos grupos diferentes que son:

- El **Cluster3**, con 2.483 clientes, lo denominamos “BAD”, es decir, son aquellos que son más o menos recientes pero tienen muy baja la frecuencia y el gasto.
- El **Cluster7**, con 2.317 clientes, lo denominamos “SPORADIC” que son los que compran esporádicamente, es decir, solo es fiel a un tipo de promoción o producto.

Con **7 segmentos** sería suficiente para tener un conocimiento más claro sobre qué tipos de clientes tiene este *e-commerce* y cuáles son las características más importantes que los definen.

Ahora procedemos a realizar el análisis de clusters para el *Método de las 2-Tuplas*.



5.2. ANÁLISIS CLUSTERS PARA EL MÉTODO 2-TUPLAS

Para el *Método de las 2-Tuplas* clasificaremos también a los diferentes clientes variando el número de clusters de 5 a 7 para ajustarnos mejor a los comportamientos de compra.

La interpretación de los clusters es igual que para los clusters obtenidos con el *Método Convencional*, es decir, nos muestra por cada *R*, *F* y *M* el valor promedio o centroide con el valor lingüístico y la desviación media. La diferencia es que gracias a la parte lingüística que aporta este método podremos determinar más claramente el tipo de cluster al que pertenecen.

- **5 clusters:**

S	C _s	Recency Score	Frequency Score	Monetary Score	Customer Type
Cluster 1	3541	(H,0.36)	(M,0.28)	(VL,0.23)	GOOD
Cluster 2	3142	(VH,-0.48)	(VH,0)	(VH,0)	VIP
Cluster 3	1966	(VH,0)	(M,-0.1)	(H,0.17)	MEDIUM
Cluster 4	3239	(L,-0.49)	(L,0.3)	(H,0.18)	SLEPT
Cluster 5	3609	(VL,0)	(VL,0)	(VL,0)	WORST

Figura 25: Análisis 5 clusters. Método 2-Tuplas

Con 5 clusters la interpretación es la siguiente:

- El **Cluster 1** recoge a 3.541 clientes, y lo clasificamos como “GOOD” porque su *Recencia* es “High”, su *Frecuencia* es “Medium” y su *Monetary* es “Very Low”. Estas tres variables tienen una desviación positiva media medio-alta que nos indica que están más cerca del cluster siguiente al que de verdad pertenecen.
- El **Cluster 2** recoge a 3.142 clientes y son los “VIP”, es decir, los mejores, puesto que son “Very High” en todas las variables, y además el promedio de las desviaciones es muy pequeña.

El **Cluster 3** recoge a 1.966 clientes y lo clasificamos como “MEDIUM” ya que presentan un hábito de compra normal. Además, su promedio de las desviaciones es muy pequeño también.



- El **Cluster 4** recoge a 3.239 clientes de los cuales parece que su comportamiento se ha quedado “dormido” porque han dejado de comprar hace bastante tiempo. Los clasificamos como “SLEPT”. En *Recencia* la desviación roza la máxima desviación en negativo lo que indica que debería clasificarse como “Very Low” en vez de como “Low”.
- Por último, el **Cluster 5** recoge a 3.609 clientes que son todos “Very Low” sin promedio de desviaciones por lo que por ello lo clasificamos como “WORST”.

- **6 clusters:**

S	C _s	Recency Score	Frequency Score	Monetary Score	Customer Type
Cluster 1	2702	(VH,0)	(VH,0)	(VH,0)	VIP
Cluster 2	2771	(VH,-0.17)	(H,0.1)	(L,0.31)	GOOD
Cluster 3	1710	(L,0.16)	(VH,0)	(H,0.4)	LOST
Cluster 4	2647	(M,0.36)	(M,-0.41)	(VL,0)	MEDIUM
Cluster 5	2714	(VL,0)	(VL,0)	(VL,0.24)	WORST
Cluster 6	2953	(L,0.18)	(L,0.01)	(H,0.1)	SLEPT

Figura 26: Análisis 6 clusters. Método 2-Tuplas

En cuanto a 6 clusters, hemos obtenido un cluster nuevo que es:

- El **Cluster 3** clasificado como “LOST” puesto que tenían buena *Frecuencia* y buen *Valor Monetario* y dejaron de comprar hace tiempo. Por lo tanto, a este cluster, al igual que comentamos para el *Método Convencional* sería conveniente reactivarles para volver a frecuentarles su hábito de compra.

- **7 clusters:**

S	C _s	Recency Score	Frequency Score	Monetary Score	CustomerType
Cluster 1	2494	(M,0.2)	(L,0.33)	(VL,0)	BAD
Cluster 2	1738	(L,0.17)	(VH,-0.33)	(H,0.4)	LOST
Cluster 3	2497	(VL,0)	(VL,0)	(VL,0.08)	WORST
Cluster 4	2299	(VH,0)	(VH,0)	(VH,0)	VIP
Cluster 5	2409	(VH,-0.27)	(H,-0.02)	(L,-0.23)	LOYAL
Cluster 6	1643	(H,0.21)	(M,-0.21)	(H,0.32)	MEDIUM
Cluster 7	2417	(L,-0.25)	(L,-0.21)	(H,0.04)	SLEPT

Figura 27: Análisis 7 clusters. Método 2-Tuplas



Por último, con 7 clusters, en comparación con el *Método Convencional*, se cambia el cluster “SPORADIC” por el de “SLEPT” por el comportamiento de compra que conllevan.

Además, se añade el cluster nuevo, el *Cluster 5* con 2.409 clientes, que clasificamos como “LOYAL” o más fieles, y son los que compran mucho aun gastándose poco dinero, pero su *Frecuencia* es muy buena.

Como vemos, estos resultados son muy parecidos a los del *Método Convencional*, ya que siempre existen los “muy buenos”, los “muy malos” y los de comportamientos irregulares.

Para este método también elegiremos quedarnos con **7 clusters** que detallan más información de cada cliente.

6. ACCIONES FUTURAS

Una vez clasificados los clientes en segmentos por comportamientos de compra parecidos, proponemos a este *e-commerce* hacer algunas acciones futuras para algunos tipos de segmentos más destacables como por ejemplo:

- Para los más importantes que son los que se encuentran en un nivel de compromiso con la compañía (R=5) (F=5) (M=5), sería conveniente enviarles:
 - Campañas exclusivas.
 - Descuentos especiales.
 - Muestras de productos.
 - Encuestas de satisfacción.
 - Obtener testimoniales.
 - Acciones que recompensen su lealtad.

Con el fin de hacerles saber a estos clientes que la empresa les valora por ser tan fieles con la marca. Así se siempre se sentirán muy satisfechos con la marca.



- A aquellos que se encuentran en la mitad (R=3) (F=3) (M=3) se les puede enviar:
 - Acciones que les hagan subir al segmento (R=5) (F=5) (M=5).
 - Descuentos atractivos.
 - Sorteos.

Con el fin de incentivar el hábito de compra y mejorar el engagement con la marca, y también crear mayor posicionamiento en estos clientes.

- A aquellos que se encuentran en la parte inferior (R=1) (F<=3) (M<=3)
 - Reactivación de valor.
 - Descuentos atractivos pero que amarren en el tiempo.
 - Encuestas de Satisfacción para identificar problemas.

Con el fin de que vuelvan a hacer alguna compra y como mínimo subirlos al segmento “MEDIUM” ó (R=3), (F=3) y (M=3).

7. CONCLUSIONES

A continuación, vamos a detallar las conclusiones más relevantes que hemos obtenido con esta investigación dando respuesta a las preguntas que planteábamos al comienzo con el fin de dar a conocer a este negocio *online* el comportamiento de compra de sus clientes.

1.- ¿Cuántos clientes son los más fieles?

Si los miramos en el *Método Convencional*, en la Figura 16, vimos que los más fieles, son aquellos que están clasificados como (R=5), (F=5) y (M=5), que es el segmento más estricto, son 656 clientes (un 4,2%). Además, haciendo referencia a la “Ley de Pareto” este 20% de clientes supone el 80% de los beneficios. En la Figura 28 los vemos:

		MONETARY
RECENCY	FREQUENCY	5
5	5	656

Figura 28: Zoom Figura 16



Además, en los análisis de 7 clusters, en el *Método Convencional* el grupo “VIP” engloba a más clientes con hábitos de compra muy buenos y son 2.992 clientes (un 19,3%). Sin embargo en el *Método de las 2-Tuplas* el segmento “VIP” lo forman 2.299 clientes (14,8%) y vemos que son menos consumidores ya que este método más preciso.

2.- ¿Cuántos clientes son los menos fieles?

Si lo miramos en el *Método Convencional*, en la Figura 16, también vimos que los menos fieles son los que están clasificados como (R=1), (F=1) y (M=1), que es el segmento más estricto, y son 949 clientes (un 6,1%), que son más que los menos fieles. En la Figura 29 los vemos:

		MONETARY
RECENCY	FRECUENCY	1
1	1	949

Figura 29: Zoom Figura 16

Además, en los análisis de 7 clusters, en el *Método Convencional* el segmento “WORST” contiene a 3.175 clientes (un 20,5%), mientras que en el *Método de las 2-Tuplas* el mismo segmento contiene a 2.497 clientes (16,1%), y vemos que también son menos que en el método anterior por su precisión.

3.- ¿Cuántos clientes han dejado de comprar recientemente y tienen una Frecuencia y un Valor Monetario bajos?

En la Figura 16 del *Método Convencional* serán aquellos clientes que tengan (R=4), (F<=3) y (M<=3) y son 1.099 clientes (un 7,09%). En la Figura 30 los vemos:

		MONETARY		
RECENCY	FRECUENCY	1	2	3
4	1		4	3
	2	222	209	110
	3	111	292	148

Figura 30: Zoom Figura 16



En los análisis de 7 clusters, según el *Método Convencional* serán los que pertenecen al segmento “BAD” y son 2.483 clientes (un 16%), mientras que el mismo segmento en el *Método de las 2-Tupas* son 2.494 (un 16,1%). En este caso, este último método contiene a 11 clientes más que el anterior, pero en porcentaje es igual.

4.- ¿Cuántos clientes que dejaron de comprar hace más tiempo tenían buena frecuencia de compra con un gasto bueno?

Son aquellos clientes que cumplen ($R \leq 3$), ($F \geq 3$) y ($M \geq 3$). En la Figura 16 del *Método Convencional* vemos que en total son 3.019 clientes (un 19,4%). En la Figura 31 los vemos:

RECENCY	FRECUENCY	MONETARY		
		3	4	5
1	3	98	51	84
	4	78	93	50
	5	9	41	57
2	3	187	227	117
	4	62	122	164
	5	34	57	178
3	3	233	171	71
	4	78	149	118
	5	137	126	227

Figura 31: Zoom Figura 16

También, en los análisis de 7 clusters este segmento hace referencia al cluster “LOST” donde según el *Método Convencional* son 1.634 clientes (un 10,5%), mientras que en el *Método de las 2-Tupas* son 1.738 clientes (un 11,2%).

5.- ¿Cuántos clientes tienen un hábito de compra normal o medio?

Son todos aquellos clientes que tienen según el *Método Convencional* ($R=3$), ($F=3$) y ($M=3$), que es el segmento más estricto para indicar un hábito de compra medio, y son 233 (un 1,5%). En la Figura 32 los vemos:



		MONETARY
RECENCY	FRECUENCY	3
3	3	233

Figura 32: Zoom Figura 16

También este segmento corresponde, en los análisis de 7 clusters, a los clusters “MEDIUM”. En el *Método Convencional* son 1.896 clientes (un 12,2%), y en el *Método de las 2-Tuplas* son 1.643 clientes (un 10,6%).

6.- ¿Cuántos clientes son los que más dinero se han gastado en los dos últimos años?

Corresponden aquellos clientes que tienen ($R \geq 3$) y ($M=5$) porque ya vimos que cuando la *Recencia* es igual a 3 contempla hasta los 2 años. Son en total 2.184 clientes (un 6%). En la Figura 33 los vemos:

		MONETARY
RECENCY	FRECUENCY	5
3	1	5
	2	126
	3	71
	4	118
	5	227
4	1	2
	2	57
	3	169
	4	139
	5	344
5	1	
	2	109
	3	13
	4	148
	5	656

Figura 33: Zoom Figura 16

7.- ¿Cuántos clientes son los que más compras han realizado en los últimos 6 meses?

Los últimos 6 meses corresponden según la Figura 16 del *Método Convencional* a ($R=5$) y ($F=5$), y son 1.346 clientes (un 8,6%). En la Figura 34 los vemos:



		MONETARY				
RECENCY	FRECUENCY	1	2	3	4	5
5	5		153	202	335	656

Figura 34: Zoom Figura 16

8.- ¿Cuántos clientes que se gastaban bastante dinero están ahora “dormidos”?

Este es un cluster importante para poder reactivar ya que generaban bastante beneficio, pero por alguna razón dejaron de comprar. En los análisis de 7 clusters para el *Método de las 2-Tuplas* este segmento lo formaban 2.417 clientes (un 15,6%).

9.- ¿Cuántos clientes tienen un hábito de compra más esporádico?

En el análisis de 7 clusters según el *Método Convencional*, hemos clasificado un cluster como “SPORADIC” refiriéndonos a una frecuencia de compra más esporádica. Esto quiere decir que estos clientes compran según les motive más un tipo de producto o de oferta. En total son 2.317 clientes que suponen un 15%.

Con el *Método de las 2-Tuplas* no apreciamos bien este cluster, el más parecido podría ser el cluster “MEDIUM”.

10.- ¿Cuántos clientes son hasta ahora muy fieles aunque su Valor Monetario no sea muy alto?

Estos clientes son muy continuos en sus compras, es decir, tienen una *Frecuencia y Recencia* altas, aunque no se gasten mucho dinero.

Corresponden al cluster “LOYAL” en el análisis de 7 cluster que los obtuvimos tanto para el *Método Convencional* como para el *Método de las 2-Tuplas*.

En el primer método son 1.000 clientes que suponen un 6,5%, y en el segundo método son 2.409 clientes (un 15,5%). En este caso, hay un 2,4% más de clientes que cumplen estas características en el *Método de las 2-Tuplas*.



Como conclusión general, tanto el *Método Convencional* como el *Método de las 2-Tuplas* son dos buenos métodos para calcular un modelo *RFM*. La distinción es que el de las *2-Tuplas* es mucho más preciso, claro y detallado porque devuelve el valor del quintil más exacto con los decimales, que nos hacen ver realmente a qué cluster debe pertenecer cada cliente y además tiene la ventaja de que cuenta con una interpretación semántica, esto es, especificamos para cada usuario etiquetas lingüísticas y no escalas numéricas, que nos ayudan a entender mejor el tipo de cluster al que pertenecen.

Concluimos que todo lo realizado son primeros pasos para tener una visión general de cómo está funcionando este negocio *online*, conociendo principalmente cómo es el comportamiento de sus clientes, cuántos son más rentables y menos rentables, etc., y así puedan personalizar determinadas acciones para los diferentes segmentos. Más adelante, sería conveniente volver a repetir todos estos análisis para comparar resultados.

8. BIBLIOGRAFIA

1. Carrasco, R. A., Blasco, M. F., & García-Madariaga, J. (s. f.). *A 2-tuple Fuzzy Linguistic Approach Applied to RFM Models. An Application to Customer Segmentation*.
2. Carrasco, R. A., Blasco, M. F., & Herrera-Viedma, E. (2015). *A 2-tuple Fuzzy Linguistic RFM Model and Its Implementation*. *Procedia Computer Science*, 55, 1340-1347. <http://doi.org/10.1016/j.procs.2015.07.118>
3. Herrera, F., & Martínez, L. (2000). *A 2-tuple fuzzy linguistic representation model for computing with words*. *Fuzzy Systems, IEEE Transactions on*, 8(6), 746–752.
4. Andalucía, E. (s. f.). *Fideliza a tus clientes con la Estrategia RFM | El Blog de ESIC Andalucía. Recuperado a partir de <http://esic.edu/andaluciablog/estrategia-rfm/>*



5. Córdoba, G. (2011, febrero 10). *Análisis RFM en retail. Empezando a segmentar clientes (I)*. Recuperado a partir de <http://www.unica360.com/analisis-rfm-en-retail-empezando-a-segmentar-clientes-i>
6. IBM Knowledge Center - *Análisis de RFM*. (s. f.). Recuperado 10 de junio de 2016, a partir de http://www.ibm.com/support/knowledgecenter/es/SSLVMB_21.0.0/com.ibm.spss.statistics.help/rfm_intro.xml.htm
7. *Rethinking RFM*. (s. f.). Recuperado 10 de junio de 2016, a partir de <https://www.ama.org/publications/MarketingInsights/Pages/rethinking-rfm.aspx>
8. *RFM: herramienta poderosa | mercadeo.com*. (s. f.). Recuperado a partir de <http://www.mercadeo.com/blog/2010/01/rfm-herramienta-poderosa/>
9. *Segmentación Efectiva V1.0*. (18:39:29 UTC). Recuperado a partir de <http://es.slideshare.net/solvisconsulting/segmentacin-efectiva-v10-presentation>
10. *What is RFM analysis (recency, frequency, monetary)? - Definition from WhatIs.com*. (s. f.). Recuperado 10 de junio de 2016, a partir de <http://searchdatamanagement.techtarget.com/definition/RFM-analysis>
11. Jakob Nielsen. (1999, febrero 7). *Why People Shop on the Web*. Recuperado 10 de junio de 2016, a partir de <https://www.nngroup.com/articles/why-people-shop-on-the-web/>
12. Di Michele, Pepe (2007). *La metodologia di segmentazione RFM*. Recuperado 10 de junio de 2016, a partir de http://www.webdieci.com/docs/innovazione/85_PMI_trevigiane/ac_metodologia-rfm_1235.asp



ANEXOS

#Importar fichero:#

```
rm(list=ls())
```

```
df <- read.csv("C:/Users/rmmas/Desktop/RFM/k means/oferplan.csv", header = T, sep=";")
```

```
head(df)
```

#Dar formato de fecha:#

```
df$Fecha <- as.Date(df$Fecha, "%d/%m/%Y")
```

```
head(df)
```

#Calcular los días:#

```
df$hoy <- as.Date(c('2016-06-02'))
```

```
head(df)
```

```
df$dias <- c(df$hoy - df$Fecha)
```

```
head(df)
```

#SQL:#

```
library(sqldf)
```

```
library(proto)
```

```
library(gsubfn)
```

```
library(RSQLite)
```

```
library(DBI)
```

```
library(tcltk)
```

```
df1 <- sqldf ('select Cliente,min(dias) as Recency,count(*) as Frequency,
```

```
sum(Precio) as Monetary
```

```
from df group by cliente')
```

```
head(df1)
```

#Método Convencional: cálculo de quintiles:#



```
library(dplyr)

Rscore <- 6-ntile(df1$Recency, 5)

Fscore <- ntile(df1$Frequency, 5)

Mscore <- ntile(df1$Monetary, 5)

df2<-data.frame(Rscore,Fscore,Mscore)

head(df2)

hist(df2$Rscore)

hist(df2$Fscore)

hist(df2$Mscore)

h<-data.frame(df1$Recency,df1$Frequency,df1$Monetary,df2$Rscore,df2$Fscore,
df2$Mscore)
```

#Método 2-Tuplas: Hacemos 100 tiles:#

```
Rscore2 <- 101-ntile(df1$Recency, 100)

Fscore2 <- ntile(df1$Frequency, 100)

Mscore2 <- ntile(df1$Monetary, 100)

df3<-data.frame(Rscore2,Fscore2,Mscore2)

head(df3)
```

```
hist(df3$Rscore2)

hist(df3$Fscore2)

hist(df3$Mscore2)
```

#Normalizamos:#



```
library(dprep)

df3$norm=znorm(df3)

Normaliza<-mmnorm(df3,minval=0,maxval=4)

head(Normaliza)

b<-data.frame(Normaliza$Rscore2, Normaliza$Fscore2, Normaliza$Mscore2) #el valor real

head(b)

i<-data.frame(round(b$Normaliza.Rscore2),round(b$Normaliza.Fscore2),
round(b$Normaliza.Mscore2)) #el valor redondeado

head(i)

alfa<-b-i #la parte decimal

head(alfa)

S1<-i+1 #para que sea de 1 a 5, en vez de 0 a 4

head(S1)

R2tuplas <- factor(S1$round.b.Normaliza.Rscore2,

levels = c("1", "2", "3", "4", "5" ),

labels = c("VL", "L", "M", "H", "VH"))

F2tuplas <- factor(S1$round.b.Normaliza.Fscore2,

levels = c("1", "2", "3", "4", "5" ),

labels = c("VL", "L", "M", "H", "VH"))

M2tuplas <- factor(S1$round.b.Normaliza.Mscore2,

levels = c("1", "2", "3", "4", "5" ),
```



```
labels = c("VL", "L", "M", "H", "VH"))
```

#Concatenamos:#

```
R<-paste("R2tuplas,",round(alfa$Normaliza.Rscore2,2),",", sep="")
```

```
F<-paste("F2tuplas,",round(alfa$Normaliza.Fscore2,2),",", sep="")
```

```
M<-paste("M2tuplas,",round(alfa$Normaliza.Mscore2,2),",", sep="")
```

#Unimos:#

```
final<- data.frame(df1$Cliente,R,F,M)
```

```
head(final)
```

#Método de k means:#

```
#####
```

#Metodo convencional:#

```
kmeansdf2 <- kmeans(df2,7) #Aquí especificamos 5, 6 y 7 clusters
```

```
str(kmeansdf2)
```

```
RFMclusterconv<-data.frame(kmeansdf2$centers)
```

```
Recency_Score_conv<-RFMclusterconv$Rscore
```

```
Frequency_Score_conv<-RFMclusterconv$Fscore
```

```
Monetary_Score_conv<-RFMclusterconv$Mscore
```

```
S<- c("Cluster-1","Cluster-2", "Cluster-3", "Cluster-4", "Cluster-5", "Cluster-6", "Cluster-7")
```



```

C<-kmeansdf2$size

finalclusterconv<- data.frame(S,C,round(Recency_Score_conv,2),
round(Frequency_Score_conv,2),round(Monetary_Score_conv,2))

finalclusterconv

#Metodo 2 tuplas:#

kmeansb <- kmeans(b, 5) #Aquí especificamos 5, 6 y 7 clusters

str(kmeansb)

cluster2tuplas<-data.frame(kmeansb$size,kmeansb$centers)

RFMcluster<-data.frame(kmeansb$centers)

RFMcluster$norm=znorm(RFMcluster)

NormalizaMeans<-mmnorm(RFMcluster,minval=1,maxval=5)

head(NormalizaMeans)

bcluster<-data.frame(NormalizaMeans$Normaliza.Rscore2,
NormalizaMeans$Normaliza.Fscore2, NormalizaMeans$Normaliza.Mscore2)

head(bcluster)

icluster<-data.frame(round(bcluster$NormalizaMeans.Normaliza.Rscore2),
round(bcluster$NormalizaMeans.Normaliza.Fscore2),
round(bcluster$NormalizaMeans.Normaliza.Mscore2)) #el valor redondeado

alfacluster<-bcluster-icluster #la parte decimal

```



```
Rscorecluster <- factor(icluster$round.bcluster.NormalizaMeans.Normaliza.Rscore2,
levels = c("1", "2", "3", "4", "5" ),
labels = c("VL", "L", "M", "H", "VH"))
```

```
Fscorecluster<- factor(icluster$round.bcluster.NormalizaMeans.Normaliza.Fscore2,
levels = c("1", "2", "3", "4", "5" ),
labels = c("VL", "L", "M", "H", "VH"))
```

```
Mscorecluster<- factor(icluster$round.bcluster.NormalizaMeans.Normaliza.Mscore2,
levels = c("1", "2", "3", "4", "5" ),
labels = c("VL", "L", "M", "H", "VH"))
```

#Concatenamos:#

```
Recency_Score <-paste(" ",Rscorecluster,"",
round(alfacluster$NormalizaMeans.Normaliza.Rscore2,2)," ", sep="")
Frequency_Score<-paste(" ",Fscorecluster,"",
round(alfacluster$NormalizaMeans.Normaliza.Fscore2,2),"", sep="")
Monetary_Score<-paste(" ",Mscorecluster,"",
round(alfacluster$NormalizaMeans.Normaliza.Mscore2,2),"",sep="")
```

#Unimos:#

```
S<- c("Cluster-1", "Cluster-2", "Cluster-3", "Cluster-4", "Cluster-5")
C<-kmeansb$size
finalcluster<- data.frame(S,C,Recency_Score,Frequency_Score,Monetary_Score)
finalcluster
```