

Análisis RFM en Cadena de Farmacias

Una cadena de farmacias quiere segmentar a su base de clientes en base al nivel de gasto. La idea de los ejecutivos es poder establecer acciones diferenciadas a los clientes según la calidad de estos en términos del valor que significa para la compañía. Por ejemplo, algunos clientes pueden tener una actividad muy baja o casi nula porque operan con la competencia y una campaña de reactivación puede ser muy significativa para ellos; otros clientes pueden mantener altos niveles de gasto en los locales de la farmacia y una campaña de fidelización puede ser lo relevante para estos clientes de manera de mantener en el tiempo los niveles de gasto.

Se requiere determinar o predecir el comportamiento de compra de los clientes en el futuro. Por ejemplo, un cliente que ha realizado muchas compras en los últimos meses pero que no presenta actividad en las últimas semanas puede ser debido a que el cliente se ha vuelto inactivo y podría ser que en el futuro no haga transacciones con la compañía. Una mirada retrospectiva lo muestra como un buen cliente, pero una prospectiva muestra que se ha vuelto inactivo.

Las empresas de retail tienen, usualmente, bases de datos de las transacciones de sus clientes que miden la calidad prospectiva de los clientes mediante tres indicadores denominados RFM. Estos indicadores se definen de la siguiente manera

- R: Recencia. Tiempo desde la última transacción
- F: Frecuencia o número de transacciones en un período de tiempo
- M: Monto promedio por transacción gastado en un período de tiempo

El período de tiempo utilizado en la definición de F y M depende del ciclo de vida de los productos. Mientras menor el ciclo de vida de los productos, mayor es la frecuencia de transacciones en un período de tiempo y usualmente es menor el período definido en el cálculo de F y M. Estos indicadores, RFM se pueden calcular para cada cliente y son la base para segmentación. Por ejemplo, suponga que un cliente tiene la siguiente actividad durante los últimos 12 meses,

Mes	1	2	3	4	5	6	7	8	9	10	11	12
Compras (M\$)				25				25;70				

Entonces, considerando un período de 12 meses desde el mes 13, R = 5 (la última transacción se hizo hace 5 meses), F = 3 (se han realizado 3 transacciones) y M = 40 (el monto promedio de una transacción es M\$ 40). El archivo "Datos Farmacias SO.xlsx" tiene una base de datos de 10.578 clientes para los cuales se quiere hacer una segmentación según los valores de RFM. Use el programa R y considere las diferencias en escala de las variables para hacer una segmentación estandarizando las variables.

Las librerías que debiera usar en este ejercicio son las siguientes:

```
library(ggplot2)  
library(cluster)
```

1. (1 pts.) Lea los datos con alguno de los siguientes comandos y luego visualice los datos con los comandos **head** y **summary**

Pontificia Universidad Católica de Chile
Escuela de Administración

```
datos=read.table("C:/...../Datos Farmacia SO.csv",header=TRUE,dec=".",sep=",")
datos=read.table("clipboard",header=TRUE,dec=".",sep="\t")
head(datos)
summary(datos)
```

Grafique las tres variables de segmentación RFM mediante un **ggplot y geom_histogram**. Genere visualizaciones atractivas.

2. (2 pts.) Seleccionar las variables de segmentación RFM (**subset**) y, como las variables RFM tienen unidades de medida muy distintas, estandarizar las variables (**scale**) antes de utilizar el comando **kmeans** que genera los segmentos.

```
datos1= datos %>% select("Recencia","Frecuencia","Monto")
datos2=as.data.frame(scale(datos1))
fit=kmeans(datos2,k,nstart=10)
```

Desafío: Grafique las distancias para los distintos valores del número de segmentos mediante un **ggplot**.

¿Cuál es el número de segmentos? Explique.

3. (2 pts.) Calcule la segmentación final, con el valor de k óptimo encontrado. Describa los segmentos.

```
fit=kmeans(datos2,k,nstart=10)
fit$centers
fit$size
```

Incorpore la variable **fit\$cluster** al **data.frame** inicial y describa los segmentos en base a las variables demográficas. Describa los segmentos en términos de las variables demográficas usando **table**.

4. (1 pts.) Proponga acciones de marketing para cada segmento.