

**Identificación de límites para la
categorización de clientes según su nivel de
compra en el Puesto de Ventas de Zamorano**

Federico Jaller Jaramillo

**Escuela Agrícola Panamericana, Zamorano
Honduras**

Noviembre, 2013

ZAMORANO
CARRERA DE ADMINISTRACIÓN DE AGRONEGOCIOS

Identificación de límites para la categorización de clientes según su nivel de compra en el Puesto de Ventas de Zamorano

Proyecto especial de graduación presentado como requisito parcial para optar
al título de Ingeniero en Administración de Agronegocios en el
Grado Académico de Licenciatura

Presentado por

Federico Jaller Jaramillo

Zamorano, Honduras
Noviembre, 2013

Identificación de límites para la categorización de clientes según su nivel de compra en el Puesto de Ventas de Zamorano

Presentado por:

Federico Jaller Jaramillo

Aprobado:

Wolfgang Pejuán, M.Sc.
Asesor principal

Ernesto Gallo, M.B.A., M.Sc.
Director
Departamento de Administración de Agronegocios.

Raúl Zelaya, Ph.D.
Decano Académico

Identificación de límites para la categorización de clientes según su nivel de compra en el Puesto de Ventas de Zamorano.

Federico Jaller Jaramillo

Resumen: El Puesto de Ventas de Zamorano cuenta con una base de datos de clientes que compran con su código, ya sean estudiantes o empleados, la cual no había sido analizada para la segmentación de los clientes. Se utilizó por primera vez la base de datos con tal fin. El presente estudio separó a los clientes en tres categorías utilizando el análisis de conglomerados de K-medias, y se realizó de dos maneras:

1) separadamente empleados y estudiantes y 2) empleados y estudiantes como un solo grupo. Para crear las categorías de clientes se utilizaron tres variables: promedio de ventas trimestrales, ganancias generadas trimestralmente por cliente, y actitud hacia los descuentos. Con el análisis de K-medias realizado se crearon tres conglomerados de clientes. A cada uno de los conglomerados se le calculó la media de compras trimestral, y posteriormente se calculó el punto medio entre cada media de cada uno de los conglomerados de clientes para obtener los límites. Los dos límites encontrados para las categorías bajo, medio y alto para los empleados fueron: L. 705 y L. 2,619; Los dos límites encontrados para las categorías bajo, medio y alto para los estudiantes fueron: L. 1,229 y L. 3,375; Los dos límites encontrados para las categorías bajo, medio y alto para los empleados y estudiantes juntos fueron: L. 441 y L. 2,046.

Palabras clave: Conglomerados, K-medias, segmentación.

Summary: Zamorano's grocery store has a database that registers its sales to customers who pay with their Student or employee ID number, which had not been analyzed for client segmentation. The present study used the *K-means Method for Analyzing Clusters* to divide clients into three different categories, based on this information. The clients were separated into categories in two ways:

1) students and employees separately, and 2) students and employees as a group. In order to categorize the clients, three variables were taken into account: quarterly average sales, quarterly revenue generated per customer, and reaction to discounts. Based on the K-means analysis, three clusters of customers were created. The quarterly mean purchases were calculated for each of the cluster. Then, the midpoints between the means of each of them were calculated to establish the limits of the categories. The two limits that separate the categories low, medium and high for employees were: L. 705 and L. 2,619. The two limits that separate the categories low, medium and high for students were: L. 1,229 and L. 3,375. The two limits that separate the categories low, medium and high for students and employees as a group were: L. 441 and L. 2,046.

Key words: Cluster, K-means, segmentation.

CONTENIDO

Portadilla.....	i
Página de firmas	ii
Resumen	iii
Contenido	iv
Índice de cuadros, figuras y anexos.....	v
1 INTRODUCCIÓN.....	1
2 MATERIALES Y MÉTODOS.....	4
3 RESULTADOS Y DISCUSIÓN.....	11
4 CONCLUSIONES.....	17
5 RECOMENDACIONES.....	18
6 LITERATURA CITADA.....	19
7 ANEXOS.....	20

ÍNDICE DE CUADROS, FIGURAS Y ANEXOS

Cuadros	Página
1. Seis combinaciones en la elaboración de conglomerados según el tipo de agrupación de cliente y el método de categorización de las variables continuas (ventas y utilidad trimestral).....	9
2. Estadísticos de todos los clientes (empleados y estudiantes) del Puesto de Ventas de Zamorano en cada categoría, utilizando el método de categorización en base a percentiles.....	12
3. Estadísticos de clientes (empleados) del Puesto de Ventas de Zamorano en cada categoría, utilizando el método de categorización en base a percentiles.....	13
4. Estadísticos de clientes (estudiantes) del Puesto de Ventas de Zamorano en cada categoría, utilizando el método de categorización en base a percentiles.....	14
5. Estadísticos de todos los clientes (empleados y estudiantes) del Puesto de Ventas de Zamorano en cada categoría, utilizando el método de categorización en base al análisis de conglomerados K-medias.	15
6. Límites propuestos para la categorización de los clientes del Puesto de Ventas de Zamorano, según la clasificación del cliente.....	16
Figuras	Página
1. Representación gráfica de la primera iteración en el método K-medias.	5
2. Representación gráfica de la segunda iteración en el método K-medias.	5
Anexos	Página
1. Número de clientes en cada categoría según el tipo de agrupación y método empleado de categorización de la variable continua promedio de compras trimestral.....	20
2. Número de clientes en cada categoría según el tipo de agrupación y método empleado de categorización de la variable continua promedio de utilidad trimestral.....	20

1. INTRODUCCIÓN

La segmentación es la manera como las empresas crean grupos de clientes con características o comportamientos similares entre sí, para proveerle un trato especial a cada grupo, generando valor, tanto para el cliente como para la empresa.

Al definir las características o comportamientos de los clientes, logramos identificar y posteriormente establecer los diferentes segmentos de clientes. Con estos segmentos se busca establecer relaciones duraderas con los mismos, una relación que busca ser redituable; es decir una relación ganar, ganar.

En la presente investigación se utilizó el método de segmentación conductual; en específico, el método por utilización o por las compras del cliente, dando uso a la bases de datos. No obstante, las bases de datos no sólo son útiles para la segmentación por utilización, hoy en día son sumamente importantes en las campañas de marketing. Según Kotler podemos realizar una campaña de marketing utilizando las bases de datos como apoyo o como base de la campaña. Este es una forma de marketing que él determina como “marketing de base de datos.” A continuación se presenta la definición y los cinco tipos de usos que le podemos dar a las bases de datos.

“El marketing de base de datos es el proceso de confección, mantenimiento y utilización de las bases de datos de clientes, así como de otras bases de datos (productos, proveedores, vendedores, etc.) con el fin de ponerse en contacto y operar con los clientes, y de establecer relaciones sólidas con ellos” (Kotler 2006).

Siguiendo la línea de Kotler en cuanto a las bases de datos, él propone que estas pueden ser utilizadas de cinco diferentes maneras:

- Para identificar clientes potenciales.
- Para decidir qué clientes deberían recibir una oferta especial.
- Para aumentar la lealtad de los clientes
- Para reactivar las compras de los clientes
- Para evitar graves errores.

Con la segmentación por utilización se espera alcanzar dos metas: la primera es proporcionar el más alto nivel de servicio a los clientes más rentables de la empresa y promoverla entre los clientes ocasionales o moderados; y la segunda meta es la maximización de las ventas en todos los grupos o segmentos de clientes con los que cuenta la empresa.

“Los clientes se colocan en grupos basados en actitudes comunes, estilos de vida y comportamiento pasado de compra. Esta técnica ofrece a la empresa las siguientes ventajas:

1. Un esquema de clasificación significativa para agrupar a los clientes, basado en los clientes reales de la empresa.
2. La capacidad de reducir grandes volúmenes de datos de los clientes a unos cuantos grupos concisos y utilizables.
3. La capacidad de asignar un número de código de grupo a cada cliente en la base de datos. Cada número se basa en las compras reales del cliente así como en otras características (e.g., dirección, cantidad gastada, pago a crédito o al contado).
4. La capacidad de medir el crecimiento y la migración de los clientes a través del tiempo y de un grupo a otro, lo que permite la evaluación de los programas de marketing.
5. La capacidad de usar la base de datos para desarrollar varios grupos con base en los diferentes beneficios de uso” (Kenneth E. Clow 2010).

El Puesto de Ventas de Zamorano cuenta con una amplia base de datos sobre los clientes que compran y son identificados con su código los estudiantes o empleados, sin embargo, no son identificados los demás clientes con un código. Dicha base de datos está conformada por las compras que fueron realizadas por estas personas en los años 2010, 2011 y 2012.

La identificación de límites para la categorización de clientes según su nivel de compra en el Puesto de Ventas Zamorano conlleva los siguientes objetivos:

- Identificar factores o atributos asociados a la rentabilidad del Puesto de Ventas necesarios para la categorización de los clientes disponibles en la base de datos.
- Organizar, desarrollar y estructurar la base de datos según los factores y atributos asociados.
- Crear categorías de clientes según los factores y atributos asociados.
- Categorizar los clientes del Puesto de Ventas de Zamorano según su comportamiento de compra.

La principal limitante del estudio es el lugar donde se realizó, ya que al trabajar con una base de datos de clientes, los resultados que obtenidos sólo serán válidos para el lugar donde compran dichos clientes. Específicamente los resultados sólo serán válidos para el Puesto de Ventas de Zamorano. Se estudiaron los clientes que compran con su código de identificación ya sean estudiantes o empleados. Adicionalmente, la data manejada es delimitada, tres años de recopilación de información.

Hasta el momento dicha base de datos no ha sido analizada, ni utilizada, para crear segmentos, grupos o conjuntos de clientes del Puesto de Ventas. Tampoco se ha utilizado para crear una campaña de marketing que tenga como base o fundamento, la utilización de la dicha base de datos.

En el estudio que se presenta a continuación, se utilizó la base de datos del Puesto de Ventas para clasificar los clientes en categorías dependiendo de su nivel de compra. Con

lo anterior, en el futuro se podrá realizar una campaña de marketing en la que logren hacer una mejor administración a la relación con los clientes, logrando mantener los clientes más rentables (estrella) y cultivar los menos rentables.

Con el estudio se buscó crear los límites de las categorías en las cuales se agruparon a los clientes, puesto que los clientes fueron categorizados según su nivel de compra. Para lograr dicha categorización de los clientes, se utilizó un método de análisis de conglomerados mutuamente excluyentes, con lo cual se puede asegurar que un cliente sólo podrá estar en una y solo una categoría. El análisis a utilizar fue el análisis de conglomerados de K-medias.

2. MATERIALES Y MÉTODOS

En el presente estudio se trabajó con la base de datos del Puesto de Ventas de Zamorano, con el fin de segmentar los clientes de acuerdo a su promedio habitual de consumo. Usando el análisis de conglomerados K-medias, los clientes fueron ubicados en tres categorías. Para lograr establecer las categorías de los clientes fue necesario calcular variables y adicionalmente fue necesario realizar transformación de algunas de ellas, debido a que la forma original como se encontraba la información en la base de datos, no permitiría hacer los conglomerados.

Análisis de conglomerados K-medias. El Análisis de conglomerados K-medias es un método de conglomerados mutuamente excluyentes, es decir, cada observación sólo puede pertenecer a una y sólo una categoría o conglomerado. Para cada uno de los conglomerados o categorías se encuentra un centroide, es decir el punto medio del conglomerado. El método K-medias encuentra tantos centroides como conglomerados se especifiquen. Luego se estiman las distancias entre cada una de las observaciones y dichos centroides. Seguidamente, ubica cada observación en el conglomerado con el centroide más cercano. Finalmente, se repite el procedimiento iterativamente hasta que el resultado no cambia, es decir, hasta que se encuentra el punto medio del conglomerado.

Para la elaboración de los conglomerados usando el análisis de conglomerados K-medias, Peter Tryfos (2001), recomienda seguir los siguientes pasos:

1. Especificar el número de conglomerados arbitraria o deliberadamente.
2. Calcular el centroide de cada conglomerado y la distancia entre cada observación y el centroide. Este cálculo del centroide se hace por medio de iteraciones, es decir, se repite varias veces el procedimiento de cálculo hasta encontrar el punto más céntrico entre las observaciones de dicho conglomerado. Si la observación está más cerca del centroide de un conglomerado al que no pertenece, debe ser reasignado al conglomerado más cercano.
3. Repetir el paso 2 hasta que todas las observaciones estén más cerca del centroide del conglomerado que pertenecen. Este proceso se debe repetir hasta que los centroides no cambien (iteraciones), ya que estos, cambian de acuerdo con las observaciones que los rodean (observaciones del conglomerado). Recordemos que el centroide puede ser visto también como el punto de gravedad del conglomerado. Por ello, éste varía de acuerdo con las observaciones que lo rodean.

4. Si el número de conglomerados no puede ser especificado con confianza desde el principio, se repiten los pasos 1-3 con diferentes números de conglomerados y se evalúan los resultados.

El proceso de ubicación de las observaciones en el conglomerado es repetitivo, ya que los centroides pueden ir variando su ubicación a medida que se dan las iteraciones. Por ello encontramos que las observaciones pueden ir cambiando de conglomerado a medida que cambian los centroides. Como se mencionó anteriormente, se debe ubicar cada observación en el conglomerado más cercano (Figura 1 y 2).

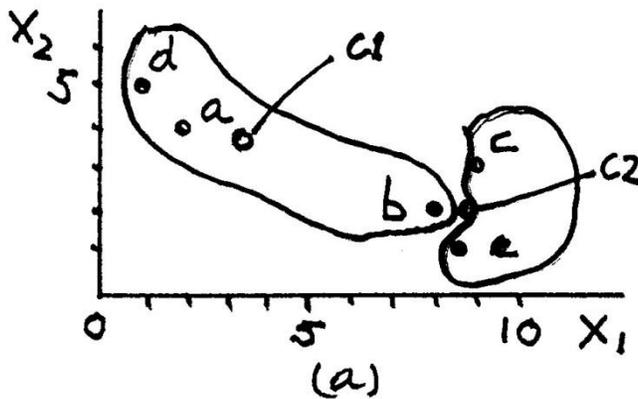


Figura 1. Representación gráfica de la primera iteración en el método K-medias.
Fuente: Tryfos, 2001.

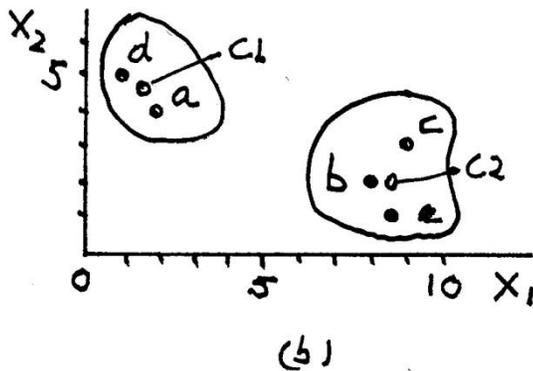


Figura 2. Representación gráfica de la segunda iteración en el método K-medias.
Fuente: Tryfos, 2001.

En las figuras anteriores las letras X_1 y X_2 representan las variables; a , b , c , d , e , representan las observaciones y C_1 y C_2 representan los centroides de los conglomerados. Adicionalmente se puede ver gráficamente como varían los centroides de la Figura 1 (primera iteración) a la Figura 2 (segunda iteración) de acuerdo a las iteraciones y la

reacomodación de las observaciones. Se observa claramente cómo al transferir la observación b del conglomerado uno al conglomerado dos, varían los dos centroides, y dicha variación de los centroides se presenta por las iteraciones.

Organización de datos. En la presente investigación se comenzó a trabajar con la base de datos que comprende las ventas realizadas entre el 2010 y el 2012 en el Puesto de Ventas de Zamorano. Esta es una base de datos de 2,315,286 transacciones (ventas), la cual incluye variables como: cantidad vendida, valor de la venta, margen de ganancia, precio de venta, fecha de venta y cliente, entre otras. Vale la pena resaltar que la base de datos contabiliza no la venta total realizada al cliente en cada visita, sino, cada producto vendido, es decir, si el cliente adquirió tres artículos en una sola visita, en la base de datos aparecen tres artículos comprados por el cliente.

Lo primero que se realizó fue un filtro que permitió separar los clientes que se podían identificar, es decir que las compras hubiesen sido realizadas con el código, ya fuese de un empleado o de un estudiante. Este procedimiento se realizó para poder contar con clientes identificables y poder observar su comportamiento. Con este ejercicio se redujo la base de datos de 2,315,286 a 810,844 ventas realizadas a clientes que pagaron con su código de identificación.

Al tener la base de datos con los clientes que se podían identificar, se procedió a dividir los clientes en dos grupos, los empleados y los estudiantes, con el fin de estudiar el comportamiento de cada tipo de cliente. Sin embargo, el estudio se realizó de tres formas según la conformación de grupos a analizar: el primer grupo estaba conformado por empleados y alumnos como un solo conjunto, el segundo grupo lo conformaban solamente los empleados y el tercer grupo se conformaba por los estudiantes solamente.

El siguiente paso fue la identificación de las compras realizadas con descuento. Para este procedimiento se utilizó el programa Excel de la siguiente manera: Se identificaron las fechas de todos los descuentos realizados del 2010 al 2012, y luego se relacionó producto y fecha de venta, con el producto y fecha de descuentos. Es decir que identificara todas las ventas con descuento. Se concluyó que 6,195 ventas (1%) fueron realizadas con descuento y 804,649 (99%) fueron ventas sin ningún descuento. El siguiente paso que se realizó fue migrar los datos al programa Statistical Package for the Social Sciences (SPSS) en cual se realizó todo el estudio estadístico.

Creación de variables. Para realizar el estudio fue necesario calcular variables, ya que en la base de datos existía la información pero no de la manera que se requería para realizar el estudio. En total se calcularon cuatro variables para poder llegar a las tres variables utilizadas en el estudio para poder generar los conglomerados.

La primera variable que se obtuvo fue la variable en la cual se segmentaron las ventas del año de acuerdo con los periodos académicos de Zamorano, para así, separar las ventas por periodos. Cada uno de estos segmentos de tiempo consta de 17 semanas calendario. Con ello logramos segmentar los tres años en nueve intervalos de tiempo diferentes.

La siguiente variable que se calculó fue la aceptación (importancia) de los descuentos para cada cliente. Esta variable fue calculada tomando el promedio de las cantidades compradas de un mismo producto por cada cliente con descuento, menos el promedio de las cantidades compradas del mismo producto pero sin el descuento. Se hizo el análisis en unidades físicas de producto y no en unidades monetarias, ya que la cantidad de dinero iba a ser afectada por el descuento y por ende se presentaría un error. Se calculó la diferencia de producto comprado con descuento y producto comprado a precio estándar por cada cliente. Si dicha diferencia es mayor a cero nos indica que para el cliente es importante el descuento y modifica su comportamiento de compra, ya que las compras fueron mayores los días que existió el descuento. Así creamos una variable dicotómica la cual nos muestra la aceptación (importancia) de los descuentos para los clientes.

Las otras dos variables, ventas y utilidad promedio por trimestre, fueron calculadas de la siguiente forma: lo que se buscó hacer fue crear una sumatoria de las ventas y la utilidad generada por cada cliente, en cada uno de los nueve intervalos de tiempo anteriormente descritos. De esta forma se puede conocer cuánto fue el total de ingreso por cada cliente en cada uno de los trimestres y, cuanto de este dinero fue la utilidad generada. Así podemos saber cuáles son los clientes que más ingresos generan y cuáles de éstos son los más rentables. Recordemos que los que más gastan no siempre son los más rentables, debido a que el margen de ganancia de los artículos varía.

Al finalizar el cálculo de las variables de aceptación (importancia) de los descuentos, ventas totales por trimestre y utilidad generada por trimestre, se deben transformar estas para realizar el análisis de conglomerados y obtener las tres categorías de clientes.

Trasformación de variables. Las variables de ventas y utilidad por trimestre deben ser modificadas, ya que estas son variables continuas y deben ser transformar a variables categóricas. Esta transformación es necesaria ya que una de las tres variables es categórica. Para poder correr el conglomerado es necesario que las variables contengan la misma composición numérica, es decir, sean del mismo tipo (en este caso variables categóricas). La transformación de variables se realizó por dos métodos diferentes. El primer método de categorización utilizado fue por el método de percentiles. El segundo método que se utilizó fue el método de categorización por medio de la utilización del análisis de conglomerados K-medias.

Para realizar la transformación de variable continua a variable categórica, es necesario crear categorías dentro de las variable continuas (ventas y utilidad por trimestre), para poder agrupar a los clientes que se asemejan entre sí; con ello podemos generar distancias establecidas en cada uno de los grupos para poder generar el “conglomerado”. La forma como generamos la categorización de estas dos variables es la siguiente:

- 1) Se debe establecer el criterio bajo el cual se estableció el número de categorías para las variables. Se estableció que la categorización de las variables fuese en tres grupos o categorías para que fuese congruente con el sistema de cobro del Puesto de Ventas de Zamorano. Esto es debido a que el sistema actual que maneja el

Puesto de Ventas de Zamorano permite clasificar a los clientes hasta en tres grupos o categorías. Al establecer dicho número de categorías, estamos siguiendo la misma línea de segmentación que se podría manejar actualmente en el Puesto de Ventas.

- 2) Se debe calcular la variable promedio de compra por periodo académico para cada cliente. Esto es con el fin de estandarizar las compras por periodo académico ya que los clientes pueden variar su promedio de consumo de un periodo académico a otro y así cambiar su categoría. Por ello, es esencial hacer el estudio en promedios académicos y no en compras totales. Este cálculo se realizó tomando la sumatoria de sus compras y dividiendo sobre el número de periodo académico del estudio, que en este caso son nueve.
- 3) Utilizar uno de los dos métodos de categorización: categorización por percentiles y por K-medias.

Categorización en base a percentiles. El procedimiento para crear los tres grupos en las variables ventas y utilidad por trimestre en base a percentiles fue el siguiente: se tomaron como el primer grupo clientes que estuviesen por bajo del percentil 25, el segundo grupo clientes que estuviesen entre los percentiles 25 y 75, y el tercer grupo de clientes que estuviesen arriba del percentil 75.

Seguidamente se identificaron las tres categorías según los niveles de compra como: categoría B, los clientes con promedio de compras trimestrales bajo; categoría M, clientes con promedios de compras medio; y categoría A, clientes con promedios de compras altos. Así mismo, de igual manera se realizó la categorización para la variable de utilidad generada por clientes, la categoría B, es para clientes que generan una utilidad baja, la categoría M, es para los clientes que generan una utilidad media y la categoría A, para los clientes que genera una alta utilidad.

Con ello se puede establecer que las dos variables continuas con las que se obtuvieron, se convirtieron en variables categóricas. Cada una de las variables se dividieron en tres categorías distintas; B “bajo”; M “medio”; A “alto”.

Categorización usando análisis de conglomerados K-medias¹. En el segundo método para la categorización de las variables continuas, se utilizó el análisis de conglomerados de K-medias. El procedimiento se realizó tomando las variables compra y utilidad promedio trimestral y se corrió un análisis de conglomerados de K-medias de tres conglomerados, con ello las variables se categorizaron en tres categorías para obtener de los clientes tres conglomerados.

¹ Se utiliza un K-medias dentro del método general de K-medias. Es decir, se utiliza dos veces el método de K-medias.

Los procedimientos anteriormente descritos fueron los métodos por los cuales se transformaron las variables continuas (promedio de ventas y utilidad trimestral) a variables categóricas. De esta manera las tres variables del estudio: aceptación a los descuentos, promedio de ventas y utilidad trimestral, cuentan con la misma composición numérica, es decir, las tres son variables categóricas.

Elaboración de los conglomerados (clúster). Los conglomerados fueron elaborados en el programa SPSS utilizando el método K-medias. Se eligió este tipo de análisis de conglomerados ya que es un análisis mutuamente excluyente para cada uno de los elementos, es decir, un elemento sólo puede estar en una y sólo una categoría.

Se realizó el conglomerado para tres categorías, ya que el sistema de cobro del Puesto de Ventas de Zamorano permite categorizar a los clientes en tres diferentes categorías. Este procedimiento fue realizado para cada uno de los grupos de clientes definidos anteriormente, es decir, empleados y estudiantes juntos, estudiantes solamente y empleados solamente. Dichos conglomerados se realizaron haciendo uso de las dos formas de categorizar las variables continuas, por el método de percentiles y por el método de K-medias. Las diferentes formas de realizar los conglomerados se pueden observar a continuación (Cuadro 1).

Cuadro 1. Seis combinaciones en la elaboración de conglomerados según el tipo de agrupación de cliente y el método de categorización de las variables continuas (ventas y utilidad trimestral).

Agrupación de cliente	Método de categorización de las variables continuas a categóricas	
	Percentiles	K-medias
Juntos (empleados y estudiantes)	1	4
Solamente empleados	2	5
Solamente estudiantes	3	6

Los resultados del primer y cuarto conglomerado (estudiantes y empleados juntos) fueron evaluados para definir cuál de los dos métodos de categorización de las variables es el más apropiado. Luego se elaboraron los siguientes dos conglomerados (estudiantes y empleados por separado) utilizando el mejor método de categorización de las variables. Los parámetros de evaluación de los conglomerados son distribución de los datos y coeficiente de variación, ya que estos son indicadores de la homogeneidad de los datos dentro de los conglomerados.

Creación de límites entre categorías. Teniendo identificados a todos los clientes en cada una de tres categorías, se procedió a calcular la media de los promedios de compra trimestrales de los clientes por cada una de las categorías. Al tener las medias de las compras de cada una de las categorías, se procedió a encontrar los puntos medios entre

dichas medias de las categorías. La forma como se hallaron los puntos medios entre las medias fue tomar la suma de las medias de cada una de las categorías colindantes y dividirla entre dos. Estos valores se convirtieron en los nuevos límites para categorizar los clientes según su nivel de compra. Este proceso fue realizado para cada uno de los grupos de clientes generados, es decir, empleados y estudiantes juntos, estudiantes solamente y empleados solamente por los dos tipos de categorización de las variables continuas.

3. RESULTADOS Y DISCUSIÓN

Se realizaron 6 conglomerados diferentes, de los cuales se presentan a continuación 4 resultados, los cuales se dividen en dos grupos de resultados debido a que la segmentación de las variables continuas para convertirlas en variables categóricas, se realizó por dos métodos diferentes. Tanto el método de percentiles como el método de K-medias, se evaluaron para identificar cuál de los dos es el más apto para la categorización de los estudiantes y los empleados por separado. Se identificó que el método más apropiado es el método de percentiles, por ello los conglomerados para estudiantes y empleados por separado, se realizó con la variable categorizada por medio de percentiles.

En total se registraron 2420 clientes, que compran con su cuenta personal de Zamorano, dentro de estos clientes hay estudiantes y empleados. Estos son los clientes que compraron entre el año 2010 y 2012. Los clientes que no compran con su cuenta personal o son externos no pueden ser identificados.

Resultados obtenidos usando percentiles para la categorización de las variables compras y utilidad trimestral. En el primer método se usó una categorización arbitraria de los datos, en donde se tomaron las variables y se segmentaron en tres grupos: del percentil 0 hasta el 25, del percentil 25 al 75, y mayores al 75. Con este método los resultados son los siguientes: El primer grupo de clientes conformado por personas que su promedio de consumo trimestral es menor que L. 412 (categoría bajo). El segundo grupo con las personas que su promedio de consumo trimestral es mayor que L. 441 y menor que L. 2,046 (categoría medio). El tercer y último grupo es el conformado por los clientes que su promedio de consumo trimestral es mayor que L. 2,046 (categoría alto). Con ello, se tienen definidos los límites de cada una de las categorías de los clientes. A continuación se presentan los estadísticos de cada uno de los grupos anteriormente mencionados (Cuadro 2).

² Tasa de cambio oficial lempiras por dólar septiembre 2013 fue de L. 20.5/US\$

Cuadro 2. Estadísticos de todos los clientes (empleados y estudiantes) del Puesto de Ventas de Zamorano en cada categoría, utilizando el método de categorización en base a percentiles.

Estadístico	Categoría		
	Bajo	Medio	Alto
N	940	1096	384
Media	192.82	1029.04	4135.05
Mediana	184.19	944.06	3095.06
Desviación	125.9	439.37	2916.95
Mínimo	2.34	441.35	2047.7
Máximo	439.65	2043.13	24531.36
Coefficiente de variación	65%	42%	70%

Todos los clientes están clasificados en tres categorías. No se identificaron categorías intermedias o combinadas, es decir, no se obtuvieron categorías con compras altas y utilidad baja, o compras bajas con utilidad alta. Las categorías identificadas para cada uno de los conglomerados corridos son:

- 1) Categoría baja: son clientes que su promedio de compras trimestral es bajo, igualmente la utilidad que genera el cliente es baja. No obstante son clientes que modifican su nivel de compra de acuerdo con los descuentos, algo que nos hace pensar que son clientes que solo compran artículos en descuento.
- 2) Categoría media: son clientes que su promedio de compra es medio, la utilidad que este cliente genera es también una utilidad media. Es un cliente que no modifica su consumo de acuerdo con los descuentos, es un cliente que consume generalmente lo mismo sin importar si el producto está o no en descuento.
- 3) Categoría alta: estos son los clientes estrellas; son los clientes que su promedio de compra es alto igualmente el promedio de utilidad generado es alto. No obstante son clientes que modifican su nivel de compra de acuerdo con los descuentos, es decir, incrementan sus compras cuando existe el descuento

La primera categoría comprende el 39% de los clientes, con una media de consumo de L. 192.82 y un coeficiente de variación de 65%. Se puede observar adicionalmente que el rango de esta categoría es de L. 437.31. La segunda categoría comprende el 45% de los clientes con una media de consumo de L. 1029.04 y un coeficiente de variación 42% con un rango de esta categoría es de L. 1601.78, su coeficiente variación es el menor de las tres categorías. La tercera y última categoría comprende el 16% de los clientes, con una media de L. 4135,05 y un coeficiente de variación de 70% con un rango de L. 22,483.66. Se puede observar que los coeficientes de variación de las categorías alto y bajo son coeficientes más altos que el coeficiente de la categoría medio, al ser este coeficiente relativamente más bajo nos indica que esta categoría es la más homogénea de las tres. Adicionalmente se puede observar que el coeficiente de variación más elevado lo tiene la categoría alto, lo cual nos indica que esta categoría es la menos homogénea.

Clasificación de los clientes en empleados solamente. Al categorizar clientes en empleados se pudo identificar que el número de empleados que compran con su cuenta en el Puesto de Ventas de Zamorano son el 25% de los clientes (623 clientes). Adicionalmente al categorizar dicho grupo de empleados se identificó que el primer grupo de clientes está conformado por personas que su promedio trimestral de consumo es menor que L. 705 (categoría bajo). El segundo grupo lo conforman los clientes que su promedio trimestral de consumo es mayor que L. 705 y menor que L. 2,619 (categoría medio). El tercer y último grupo son los clientes que su promedio de compra trimestral es mayor que L. 2,619 (categoría alto). A continuación se presenta un cuadro con la información estadística de cada una de las categorías creadas (Cuadro 3).

Cuadro 3. Estadísticos de clientes (empleados) del Puesto de Ventas de Zamorano en cada categoría, utilizando el método de categorización en base a percentiles.

Estadístico	Categorías		
	Bajo	Medio	Alto
N	248	229	146
Media	266.37	1,487.39	5,967.75
Mediana	218.13	1,380.99	4,695.76
Desviación	204.91	545.79	3,910.19
Mínimo	2.34	708.59	2,623.98
Máximo	700.19	2,617.96	24,531.36
Coefficiente de variación	76%	36%	65%

La primera categoría (bajo) contiene el 40% de los clientes, con una media de consumo trimestral de L. 266.37 y un coeficiente de variación de 76% con un rango de L. 697.85. La segunda categoría contiene el 36% de los clientes, con una media de consumo de L. 1,487.39 y un coeficiente de variación de 36% con un rango de L. 1,909.37. La tercera y última categoría está compuesta por 24% de los clientes con una media de L. 5,967.75 Lempiras y un coeficiente de variación de 65% con un rango de L. 21,907.38. Se puede observar que los coeficientes de las categorías medio y alto son coeficientes bajos, lo cual nos indica que estas categorías son bien homogéneas. Adicionalmente se puede observar que el coeficiente de variación más elevado lo tiene la categoría bajo, lo cual nos indica que esta categoría no es homogénea.

Clasificación de los clientes en estudiantes solamente. Al categorizar los clientes en estudiantes, se pudo identificar que el porcentaje de estudiantes que compran con su cuenta en el Puesto de Ventas de Zamorano son el 75% de los clientes (1,797 clientes). Adicionalmente al categorizar dicho grupo de estudiantes se identificó que el primer grupo de clientes está conformado por clientes que su consumo trimestral es menor que L. 1,228.53 (categoría bajo). El segundo grupo lo conforman los clientes que su promedio trimestral de consumo es mayor que L. 1,228.53 y menor que L. 3,374.83 (categoría

medio). El tercer y último grupo son los clientes que su promedio de compra trimestral es mayor a L. 3,374.83 (categoría alto). Con ello se tienen definido los límites de las categorías en la cuales se agruparan los clientes. A continuación se presenta un cuadro con la información estadística de cada una de las categorías (Cuadro 4).

Cuadro 4. Estadísticos de clientes (estudiantes) del Puesto de Ventas de Zamorano en cada categoría, utilizando el método de categorización en base a percentiles.

Estadístico	Categorías		
	Bajo	Medio	Alto
N	1,356	392	49
Media	452.63	1,980.82	4,667.78
Mediana	383.88	1,813.47	4,592.33
Desviación	334.52	585.78	1,099.55
Mínimo	2.4	1,235.8	3,378.47
Máximo	1,228.15	3,363.13	7,787.56
Coefficiente de variación	73%	29%	23%

La primera categoría está conformada por el 75% de los clientes, con una media de consumo de L. 452.63 y un coeficiente de variación de 73% con un rango de L. 1,225.75. La segunda categorías está conformada por 22% de los clientes, con una media de consumo trimestral de L. 1,980.82 y un coeficiente de variación de 29% con un rango de L. 2,127.33 la tercera y última categoría está conformada por el 3% de los clientes, con una media de consumo trimestral de L. 4,667.78 un coeficiente de variación de 23%, con un rango de L. 4,409.09. Se puede observar que los coeficientes de las categorías medio y alto son coeficientes bajos, lo cual nos indica que estas categorías son bien homogéneos. Adicionalmente se puede observar que el coeficiente de variación más elevado lo tiene la categoría bajo, lo cual nos indica que esta categoría no es homogénea.

Resultados obtenidos usando la categorización de las variables compras y utilidad trimestral en base a análisis K-medias. En el segundo método utilizado para la categorización de los clientes del Puesto de Ventas de Zamorano se procedió a que los datos se expresaran por sí solos. La transformación de las variables continuas a variables categóricas se realizó por medio análisis de conglomerados de K-medias. Con ello se categorizaron las variables compras y utilidad promedio trimestral en tres grupos y se realizó el conglomerado para finalmente definir los tres grupos de clientes.

El primer grupo de clientes resultó conformado por personas que su promedio de consumo trimestral es menor que L. 1,765.9 (categoría bajo). El segundo grupo lo conformarán las personas que su promedio de consumo trimestral es mayor que L. 1,765.9 y menor que L. 7,870.54 (categoría medio). El tercer y último grupo es el conformado por los clientes que su promedio de consumo trimestral es mayor que 7,870.54 L. (categoría alto). A continuación se presentan los estadísticos de cada uno de los grupos anteriormente mencionados (Cuadro 5).

Cuadro 5. Estadísticos de todos los clientes (empleados y estudiantes) del Puesto de Ventas de Zamorano en cada categoría, utilizando el método de categorización en base al análisis de conglomerados K-medias.

Estadísticos	Categorías		
	Bajo	Medio	Alto
N	1,955	437	28
Media	590.67	3,194.61	1,2362.03
Mediana	465.16	2,692.64	1,0668.51
Desviación	4,77.36	1,408.47	4,369.13
Mínimo	2.34	1,768.01	7,920.55
Máximo	1,764.75	7,858.66	2,4531.36
Coefficiente de variación	80%	44%	35%

La primera categoría abarca el 81% de los clientes, con una media de consumo de L. 590.67, un coeficiente de variación de 80% y con un rango de L. 1,1762.41. La segunda categoría contiene el 18% de los clientes, con una media de consumo de L. 3194.61, con un coeficiente de variación de 44%, y con un rango de L. 6090. La tercera y última categoría comprende el 1% de los clientes con una media de L. 12,362.03, un coeficiente de variación de 35% y con un rango de L. 16610.81 a pesar que el rango es el más amplio de los tres es el que menor coeficiente de variación tiene. Se puede observar que los coeficientes de las categorías medio y alto son coeficientes bajos, lo cual nos indica que estas categorías son homogéneas. Adicionalmente se puede observar que el coeficiente de variación más elevado lo tiene la categoría bajo, lo cual nos indica que esta categoría no es homogénea.

Es de notar que los datos que fueron categorizados por medio del método de percentiles se distribuyeron de una manera normal. Se generaron así unos coeficientes de variación más bajos que los presentados por los datos categorizados por medio del método de K-medias. Lo cual nos indica que los conglomerados son más homogéneos entre sí con el método de percentiles.

Basados en la información anterior, y al decidir cuál de los dos métodos de categorización de las variables continuas es el más conveniente, se concluye que en este caso es la categorización por percentiles. Debido a ello para la elaboración de los conglomerados de estudiantes y empleados por se parado, el método seleccionado fue la categorización por percentiles.

Los límites de las categorías de clientes se encontraron para los siguientes grupos: empleados solamente, estudiantes solamente y todos juntos. Estos límites se calcularon utilizando el método de categorización de las variables continuas a categóricas de percentiles. A continuación se presentan los límites de cada uno de los grupos anteriormente mencionados (Cuadro 6).

Cuadro 6. Límites propuestos para la categorización de los clientes del Puesto de Ventas de Zamorano, según la clasificación del cliente.

Grupo de clientes	Límites		
	Categoría baja	categoría media	categoría alta
Empleados	0 - 705	706 - 2,619	Mayor a 2,620
Estudiantes	0 - 1,229	1,229 - 3,375	Mayor a 3,375
Todos juntos	0 - 441	441 - 2,046	Mayor a 2,046

Es de notar que los límites encontrados para el grupo de solamente empleados y solamente estudiante, son límites distintos entre sí. Esto nos sugiere que contamos con dos grupos de clientes diferentes. Siendo lo anteriormente dicho respaldado por la diferencia que existe entre los límites de los empleados y estudiantes solamente y los límites encontrados para todos juntos. No obstante los coeficientes de variación son menores para los grupos de estudiantes y profesores por separado, lo cual es un indicativo de mayor uniformidad en los grupos de clientes.

Al definir los límites de las categorías de los clientes, se podrá establecer un programa de fidelización, que permita identificar a cada uno de los clientes del Puesto de Ventas de Zamorano. Al contar con dichos límites se podrá incluir a cada uno de los clientes es una categoría y por ende se prestará un mejor servicio a cada uno de ellos, logrando así mantener los clientes más rentables y promocionar a los menos rentables.

4. CONCLUSIONES

- Los factores y atributos asociados a la rentabilidad del Puesto de Ventas identificados fueron: utilidad trimestral por cliente, ventas trimestrales por cliente y aceptación (importancia) de los descuentos.
- El método más apropiado para la categorización de las variables continuas, compras y utilidad promedio trimestral generadas por los clientes del Puesto de Ventas de Zamorano, es el método de percentiles debido a que este genera una mejor distribución de los datos.
- El método más apropiado para la categorización de los clientes del Puesto de Ventas de Zamorano, es generando la clasificación entre estudiantes y empleados, ya que los datos se distribuyen de una mejor manera y se generan grupos más homogéneos entre sí.

5. RECOMENDACIONES

- Hacer uso de los límites establecidos en el proyecto con el método de percentiles y con ello categorizar los clientes del Puesto de Ventas de Zamorano.
- Establecer un sistema de identificación de clientes, con el cual se pueda conocer el comportamiento de compra de cada individuo y no solamente de estudiantes y empleados para poder así categorizar a cada uno de los clientes en una de las categorías encontradas.
- Hacer la categorización de los clientes por separado, es decir, estudiantes y empleados por separado, ya que de esta manera los grupos de clientes se comportan de una manera más similar al ser más homogéneos.
- Crear un sistema de fidelización de clientes con el cual se puedan mantener los clientes más rentables (estrellas) y estimular a los clientes menos rentables.

6. LITERATURA CITADA

Clow, K. E., Baack, D. 2010. Publicidad, promoción y comunicación integral en marketing. Trad. Maria del Pilar Carril Villarreal. 4ta edición. México. Prentice Hall. 482 p.

Kotler, P., Lane Keller, K. 2006. Dirección de Marketing. Duodécima edición. México. Editorial Pearson. 816 p.

Tryfos, P. 2001. Methods for Business Analysis and Forecasting: Text & Cases. illustrated edition. USA. Wiley. 592 p.

Tan, P-N., Steinbach, M., Kumar, V. 2006. Introduction to Data Mining, Cluster Analysis: Basic Concepts and Algorithms (en línea) consultado 4 de Julio 2013. Disponible en <http://www-users.cs.umn.edu/~kumar/dmbook/ch8.pdf>

7. ANEXOS

Anexo 1. Número de clientes en cada categoría según el tipo de agrupación y método empleado de categorización de la variable continua promedio de compras trimestral.

Cliente y tipo de categorización	Número clientes Categoría baja	Número clientes Categoría media	Número clientes Categoría alta	Asimetría
Empleados K-medias	19	109	495	-1.9
Empleados Percentiles	134	240	249	-0.3
Estudiantes K-medias	0	102	1,695	-3.8
Estudiantes Percentiles	471	970	356	0.7
Todos K-medias	19	211	2,190	-3.2
Todos Percentiles	605	1,210	605	0

Anexo 2. Número de clientes en cada categoría según el tipo de agrupación y método empleado de categorización de la variable continua promedio de utilidad trimestral.

Cliente y tipo de categorización	Número clientes Categoría baja	Número clientes Categoría media	Número clientes Categoría alta	Asimetría
Empleados K-medias	21	472	130	0.6
Empleados Percentiles	157	255	211	-0.1
Estudiantes K-medias	1	1,559	237	2.1
Estudiantes Percentiles	448	955	394	0.3
Todos K-medias	22	2,032	367	1.5
Todos Percentiles	605	1,210	605	0