

UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERIA INDUSTRIAL

METODOLOGÍA PARA REALIZAR PREDICCIÓN DE FUGA DE CLIENTES EN UNA
EMPRESA DE RETAIL

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

PRISCILA DENISSE BASTÍAS LARRAÍN

**PROFESOR GUÍA:
RICARDO MONTOYA MOREIRA**

**MIEMBROS DE LA COMISIÓN:
ALEJANDRA PUENTE CHANDÍA
HA NUI BAEK**

**SANTIAGO DE CHILE
OCTUBRE 2009**

METODOLOGÍA PARA REALIZAR PREDICCIÓN DE FUGA DE CLIENTES EN UNA EMPRESA DE RETAIL

La industria del retail está compuesta por distintos tipos de locales, tales como las tiendas de especialidad, los supermercados y las tiendas por departamento. Esta memoria se desarrollará en un holding que posee locales de diferentes tipos, entre los cuales están los dos últimos mencionados. Por a la alta competitividad que existe en esta industria, es importante conocer el comportamiento de los clientes y realizar un programa de fidelización, con el objetivo de motivarlos a establecer o mantener una relación con la empresa, la cual se da de manera no contractual. Debido a esto último es que no se cuenta con herramientas que indiquen cuando los clientes se fugan, por lo que el objetivo principal de este trabajo será predecir cuando esto ocurre.

Lo primero será definir un criterio mediante el cual se considere a un cliente como fugado, como una forma de trabajar con la relación no contractual entre los clientes y la empresa. Esto se realizará con datos transaccionales, basándose en las variables RFM. Luego se aplicarán una modificación del modelo Pareto/NBD y las Cadenas de Markov Ocultas para definir un conjunto de clientes sobre los cuales es conveniente aplicar acciones de marketing porque se encuentran en vías de fugarse. El modelo Pareto/NBD estimará la probabilidad que éstos poseen de permanecer activos, teniéndose la ventaja de que el modelo fue adaptado según el criterio de fuga definido y que se ha usado una variación de él, de modo que se simplifiquen los cálculos computacionales. Sin embargo, se tiene la desventaja de tener que definir de manera manual los estados con los cuales se identifica a los clientes que requieren acciones de marketing, en base a la probabilidad calculada, lo cual puede resultar poco exacto. Por su lado, las Cadenas de Markov Ocultas estimarán las probabilidades de que los clientes se encuentren en los diferentes estados, lo cual es una ventaja, ya que éstos quedan definidos por el modelo y mediante ellos se identifican los clientes a los que hay que aplicar acciones de marketing. Sin embargo, se tiene la desventaja de que, como estos estados son ocultos, sólo se cuenta con una probabilidad de que los clientes pertenezcan a ellos, lo cual podría hacer que se les clasifique erróneamente.

Al comparar los errores absolutos, se tiene que éstos son cercanos al 30% en promedio en las Cadenas de Markov Ocultas mientras que en Pareto/NBD se acercan al 10% en promedio. Además, las Cadenas de Markov clasifican a la mayoría de los clientes en un sólo estado, lo cual es poco discriminador para identificar a aquellos próximos a fugarse. Por esto, la caracterización de los clientes se realizó en base a Pareto/NBD, obteniendo que cerca del 15% de ellos debieran ser abordados para evitar su fuga.

Por otro lado, se tiene que una de las principales razones de la fuga de los clientes es que sus necesidades no son atendidas, por lo tanto, la recomendación es preocuparse de esto, ya que cuando sienten que los están considerando, el porcentaje de clientes que se fuga disminuye. Luego, las acciones de marketing para aplicar al 15% de los clientes seleccionados deberían ir enfocadas a mejorar la satisfacción de ellos.

Finalmente, un trabajo futuro interesante sería calcular el valor de los clientes y compararlo con la probabilidad que poseen de fugarse, para medir que tan conveniente es retenerlos, en cuanto al valor que le entrega a la empresa.

Agradecimientos

Al finalizar esta etapa, no puedo dejar de agradecer a las personas que, ya sea directa o indirectamente, me han apoyado y acompañado durante el transcurso de ésta.

A mi familia, por toda la preocupación que me han demostrado durante este proceso, por estar conmigo en todo momento y sobre todo por entregarme todo el amor que tienen en sus corazones. Especialmente a mis padres, por todo lo que han hecho por mí y mis hermanos, por el esfuerzo que han puesto en entregarnos la mejor educación y los valores para poder ser mejores personas. Sin ustedes nada de esto sería posible.

A Alexis, por acompañarme en este proceso con toda la paciencia del mundo, tanto en mis buenos como en mis malos momentos y por entregarme todo el amor y el apoyo que alguien podría necesitar.

A todos mis amigos y compañeros que de una u otra forma influyeron en mí durante este proceso. Especialmente a Paula, Javiera y Alejandra, que han sido un pilar fundamental para mí en esta etapa de 6 años y algo más e incluso desde antes, gracias por ser mis amigas incondicionales, por la paciencia que me han tenido y por todo el cariño que me han entregado.

A mis profesores, que me guiaron durante este proceso y que siempre estuvieron dispuestos a entregarme toda la ayuda y el apoyo necesario para salir adelante durante el transcurso de éste.

A la gente de la empresa, por haberme dado la oportunidad de realizar este trabajo en ella y por contar siempre con la mejor disposición tanto conmigo como con mis compañeros.

Y finalmente, no puedo dejar de agradecerle a Dios por permitirme llegar a esta etapa de mi vida, por la fortaleza y la sabiduría que me ha entregado durante todo este proceso y por las hermosas personas que ha puesto en mi camino y que han aportado a que sea lo que hoy soy.

INDICE DE CONTENIDOS

| | |
|---|-----------|
| 1. ANTECEDENTES GENERALES | 1 |
| 1.1. Acerca de la Empresa | 2 |
| 2. DESCRIPCIÓN DEL PROYECTO | 3 |
| 3. OBJETIVOS | 4 |
| 3.1. Objetivo General | 4 |
| 3.2. Objetivos Específicos | 4 |
| 4. MARCO CONCEPTUAL..... | 4 |
| 4.1. Definición de Cliente Fugado | 4 |
| 4.2. Modelos de Predicción de Fuga de Clientes..... | 5 |
| 4.2.1. Modelo Pareto/NBD | 6 |
| 4.2.2. Cadenas de Markov | 8 |
| 4.2.3. Cadenas de Markov Ocultas | 9 |
| 4.2.4. Eficiencia de los modelos..... | 12 |
| 5. METODOLOGÍA..... | 13 |
| 6. ALCANCES | 14 |
| 7. RESULTADOS ESPERADOS..... | 14 |
| 8. SELECCIÓN Y PREPROCESAMIENTO DE DATOS | 15 |
| 8.1. Descripción de la Base de Datos en General..... | 15 |
| 8.2. Preprocesamiento de Datos | 17 |
| 8.3. División en Base de Entrenamiento y Base de Testeo..... | 19 |
| 9. DESCRIPCIÓN DE VARIABLES | 19 |
| 9.1. Descripción Base de Datos Limpia..... | 19 |
| 9.2. Descripción Base de Datos Entrenamiento | 20 |
| 9.3. Variables Calculadas..... | 20 |
| 10.DEFINICIÓN CRITERIO DE FUGA | 26 |
| 11.MODELO PARETO/NBD..... | 29 |
| 11.1. Error Modelo Pareto/NBD | 35 |
| 12.CADENAS DE MARKOV OCULTAS | 36 |

| | |
|--|----|
| 12.1. Error Cadenas de Markov Ocultas | 44 |
| 12.2. Comparación Cadenas de Markov Ocultas con Pareto/NBD | 44 |
| 13. ANÁLISIS DE RESULTADOS..... | 46 |
| 14. CARACTERIZACIÓN CLIENTES | 47 |
| 15. POSIBLES ACCIONES DE MARKETING | 48 |
| 16. CONCLUSIONES Y RECOMENDACIONES..... | 50 |
| 16.1. Sobre los datos..... | 50 |
| 16.2. Sobre los modelos..... | 51 |
| 16.3. Sobre los errores | 52 |
| 16.4. Sobre los resultados obtenidos | 52 |
| 16.5. Trabajos Futuros | 53 |
| 17. BIBLIOGRAFÍA | 55 |
| 18. ANEXOS..... | 57 |

1. ANTECEDENTES GENERALES

En la actualidad existe una alta competitividad en los mercados, por lo que se hace necesario contar con herramientas que puedan marcar una diferencia con respecto a los competidores. En ese sentido, entender el comportamiento de los clientes es fundamental para poder realizar acciones que sean atractivas para ellos y los hagan preferir a la empresa frente al resto. Estas acciones pueden estar dirigidas a diferentes estrategias, ya sea para captar nuevos clientes, generar fidelidad en los actuales o, así mismo, evitar que algunos dejen de serlo. Esto último es lo que se denomina migración o fuga de clientes y la predicción de ellos es el tema central que se abordará en esta memoria.

La importancia de realizar un modelo que pueda predecir cuando un cliente se encuentra más propenso a terminar la relación con la empresa, radica tanto en un asunto de costos como también en una forma de comprender el comportamiento de los clientes y generar estrategias de marketing que se encuentren acorde con éste. Debido a la gran competencia existente en el mercado, conseguir nuevos clientes es una tarea cada vez más difícil. Los costos de realizar programas para captarlos pueden ser muy elevados, lo cual justifica cuidar a los clientes actuales. Sin embargo, no todos los clientes poseen el mismo comportamiento de compra, algunos son más leales a la empresa, mientras que otros tienen una alta probabilidad de fugarse. Lo anterior deja en evidencia el hecho de que las acciones de marketing deben ser diferentes dependiendo del tipo de cliente al cual vayan dirigidas.

En términos generales, la fuga de clientes puede ser de dos tipos: contractual y no contractual. La primera es cuando se establece un contrato cliente-empresa, en el cual ambas partes se comprometen a cumplir ciertos compromisos, y cuando el cliente no quiere seguir consumiendo los productos o servicios ofrecidos por la empresa da término al contrato existente. Por lo tanto, en este caso, la fuga del cliente es conocida por parte de la empresa, ya que existe un contrato de por medio que indica cuando la relación finaliza. En cambio, en el segundo caso, no existe un contrato de por medio que finalice una vez que el cliente no quiere seguir consumiendo los productos de la empresa. En este caso, el cliente simplemente deja de relacionarse con la empresa, sin tener la necesidad de dar previo aviso a ésta, por lo tanto, ella no tiene la certeza de que el cliente se encuentra fugado.

Este trabajo se centrará en la industria del retail, en la cual se da una relación no contractual. Por esta razón, se hace necesario primero definir qué criterio se usará para considerar un cliente fugado, para luego establecer un modelo que indique el momento en el cual un cliente debe ser abordado con el objetivo de evitar su fuga a futuro, cuando esto sea conveniente para la empresa.

Para comprender mejor el contexto en el cual se desarrollará este trabajo, se presenta a continuación una descripción de la empresa.

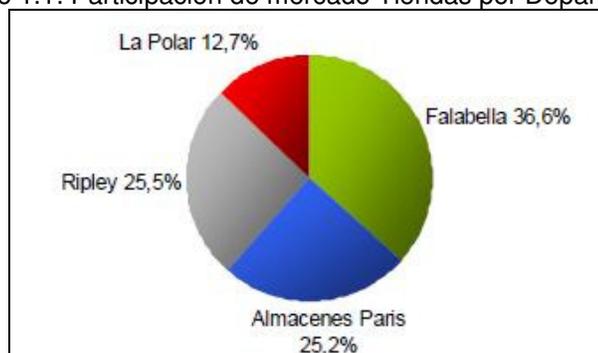
1.1. Acerca de la Empresa

La industria del retail en Chile ha adquirido cada vez más importancia para el progreso del país. Esto se puede afirmar, ya que durante el año 2006 representó el 21,2% del PIB, con un monto total de US\$28.260 MM¹.

Esta industria puede dividirse en tres grandes grupos. En el primer grupo se tienen las tiendas especializadas, tales como las farmacias, las ferreterías, tiendas de calzado, etc. El segundo grupo lo conforman los supermercados y las tiendas de mejoramiento del hogar, mientras que el tercer grupo está compuesto por las tiendas por departamento o multitiendas. Estas últimas son aquellas en las cuales se vende una variedad de productos, los cuales son separados por sectores dentro de la tienda, según sus categorías.

Dentro de este último sector de la industria se pueden identificar cuatro grandes operadores: Falabella, Paris, Ripley y La Polar. La participación de mercado de cada uno de ellos en 2007 se puede observar en el Gráfico 1.1.

Gráfico 1.1: Participación de mercado Tiendas por Departamento



Fuente: www.falabella.com

La empresa con la que se trabajará cuenta con tiendas en varios grupos de la industria. En términos generales, se puede mencionar que cuenta con 37 locales de las Tiendas por Departamento, 62 locales de Mejoramiento del Hogar y 21 locales de los supermercados en la industria nacional. Además de encontrarse en Chile, posee una sólida presencia en Perú, Argentina y Colombia, teniendo locales de los tres tipos mencionados anteriormente en estos países. En total, la empresa posee 175 locales a nivel latinoamericano, contando con una utilidad del ejercicio de un valor de aproximadamente \$217 MM².

Cabe mencionar el hecho de que la empresa ofrece servicios también fuera de la industria del retail, sin embargo, estos no son su principal negocio, por lo que serán mencionados a futuro como: "No-Retail".

¹ Fuente: INE.

² Fuente: Memoria Anual 2007.

Últimamente, una de las principales preocupaciones dentro de la empresa es entender el comportamiento de los clientes, de manera de poder enfocar sus estrategias en ellos, para que se sientan atraídos a establecer o mantener una relación con la empresa.

Sin embargo, hoy en día la empresa no posee la capacidad de identificar a sus clientes fugados, ya que la finalización de la relación cliente-empresa se produce de manera no contractual.

Debido a lo anterior, se hace necesario generar una metodología mediante la cual se pueda identificar a los clientes que se encuentran propensos a fugarse, con el objetivo de evitar que esto ocurra.

Durante el transcurso de este trabajo se desarrollará una metodología mediante la cual ellos podrán ser identificados, consiguiendo así poder aplicar ciertas estrategias sobre éstos de manera de poder retenerlos.

Los diferentes modelos a usar serán detallados en la sección Marco Conceptual.

2. DESCRIPCIÓN DEL PROYECTO

Durante el último tiempo la empresa se ha preocupado de desarrollar herramientas mediante las cuales pueda entender el comportamiento de los clientes, de modo de seguir acciones que vayan acorde con éste. En este ámbito, es que existe una preocupación de su parte por saber cuáles de sus clientes están más cercanos a fugarse y poder aplicar acciones para retenerlos.

En ese sentido, la empresa podría aplicar sobre todos sus clientes acciones que busquen retenerlos, sin embargo, cuando la probabilidad de que éstos se fuguen es pequeña, aplicarlas sobre ellos no agregará un gran valor a la firma y la hará incurrir en costos innecesarios. Luego, se hace importante identificar a los clientes con probabilidad más alta de fugarse, para aplicar solamente sobre ellos las acciones de retención.

Como en la empresa la fuga de clientes se da de manera no contractual, es muy difícil identificar a los clientes fugados sin tener una herramienta que apoye esta labor. Por esta razón, se hace necesario desarrollarla, así como definir los criterios bajo los cuales un cliente puede ser considerado en estado de fuga.

Tomando en cuenta los argumentos mencionados, esta memoria busca desarrollar una metodología que permita predecir la posible fuga de los clientes, de modo de identificar a aquellos sobre los cuales es conveniente para la empresa realizar acciones de retención.

3. OBJETIVOS

3.1. Objetivo General

Diseñar una metodología mediante la cual se pueda detectar cuales son los clientes que poseen una mayor probabilidad de fuga.

3.2. Objetivos Específicos

- Definir un criterio mediante el cual un cliente pueda ser considerado como fugado.
- Realizar una caracterización de los clientes, a modo de ver si se encuentran fugados o no, y saber con qué probabilidad podrían fugarse. Aplicar diferentes metodologías para llevar a cabo este punto y analizar cual se adapta mejor a las condiciones de la empresa.
- Caracterizar de manera más detallada a aquellos segmentos de clientes que tengan mayor probabilidad de fugarse.

4. MARCO CONCEPTUAL

Frente a la temática de la predicción de fuga de clientes, existen diversas formas de abordar este tema.

En el momento de decidir de qué manera abordarlo, es necesario fijarse si se trata de un caso de fuga contractual o uno no contractual, ya que de esto depende en gran manera del modelo a usar para realizar la predicción de fuga de los clientes.

A continuación se definen algunas maneras de abordar este tema, tanto por la definición de un cliente fugado como los posibles modelos mediante los cuales se puede predecir la fuga de un cliente.

4.1. Definición de Cliente Fugado

En este caso, se tiene una fuga del tipo no contractual. Lo primero es definir un criterio mediante el cual un cliente pueda ser considerado como fugado por parte de la empresa.

Para esto se puede usar el método RFM, el cual se define a continuación.

- Método RFM

Este método caracteriza a los clientes en base a tres variables:

1. Recency (R): es el tiempo transcurrido desde la última compra.
2. Frequency (F): esta variable mide la velocidad con la que compra el cliente, por lo tanto puede ser estimada de dos formas: el tiempo transcurrido entre transacciones o el número de transacciones realizadas en un período de tiempo. Para el caso de esta memoria, se usó la primera definición.
3. Monetary Value (M): es el monto promedio de las transacciones dentro de un período de tiempo.

Basándose en estas variables, se plantean dos posibles métodos para definir a un cliente como fugado:

1. Método R:
Observando la variable R, se puede analizar cuál es la probabilidad de que un cliente vuelva a realizar una transacción dado el valor que tenga esta variable, basándose en el comportamiento histórico de los clientes. En base a esto se podría definir un valor máximo para esta variable, de modo que si un cliente sobrepasa este valor, entonces será considerado como fugado para efectos de la empresa.
2. Método R/F:
Aquí lo que se busca es hacer una comparación entre las variables R y F. Si la razón R/F tiene un valor menor que 1, significa que el cliente ha realizado su última compra en un tiempo menor que su frecuencia habitual, por lo tanto es poco probable que se haya fugado. En cambio, si ese valor es mucho mayor que 1, quiere decir que la probabilidad de fuga de ese cliente es alta, ya que ha pasado un tiempo mucho mayor a su frecuencia habitual desde su última compra. Por lo tanto, al igual que en el caso anterior, se podría definir un valor máximo para la razón R/F, el cual indicará que los clientes se encuentran fugados cuando lo sobrepasen.

La variable M entrega un valor promedio de las compras del cliente, por lo tanto, observándola en conjunto con las otras dos variables, se puede estimar si el cliente es valioso para la empresa y así saber si vale la pena invertir en su retención.

4.2. Modelos de Predicción de Fuga de Clientes

Debido a la condición de fuga no contractual, la metodología a usar para la predicción de fuga de clientes una adaptación de la Pareto/NBD. Los resultados que se obtengan serán comparados con los obtenidos mediante la metodología de Cadenas de Markov Ocultas, para saber cuál de los dos es el más adecuado para este caso. Ambas metodologías serán detalladas a continuación.

4.2.1. Modelo Pareto/NBD³

Es un modelo que busca pronosticar el comportamiento futuro de los clientes basándose en métodos probabilísticos. Con él se pueden estimar tanto la probabilidad que tiene un cliente de permanecer activo, así como el número de transacciones futuras o su Lifetime Value. Para efectos de esta memoria, sólo se usará para calcular la probabilidad de permanecer activo.

Para realizar estas estimaciones, el modelo utiliza la cantidad de transacciones realizadas en un período de tiempo, el largo de este período y el instante en el cual se realiza la última compra dentro de ese período. Toda esta información es requerida para cada cliente.

Este modelo considera los siguientes supuestos:

- a) El número de transacciones realizadas por los clientes activos siguen un proceso de Poisson de tasa λ , por lo tanto, el tiempo entre transacciones se distribuye de manera exponencial de media $1/\lambda$.
- b) Una vez que un cliente es considerado inactivo o fugado no puede volver a ser considerado como activo.
- c) Existen independencia de la variable M tanto con la variable R como con la variable F.
- d) Un cliente se encuentra activo por un período no observable τ , el cual se distribuye según una exponencial de tasa μ . Una vez que transcurre este período el cliente pasa al estado de inactivo.
- e) La tasa de compra de los clientes (λ) sigue una distribución Gamma de parámetros r (forma) y α (escala).
- f) La tasa de fuga de los clientes (μ) sigue una distribución Gamma de parámetros s (forma) y β (escala).
- g) Las tasas λ y μ varían de manera independiente entre los clientes.

Luego, para estimar el comportamiento futuro de los clientes se quiere encontrar los parámetros r, α , s y β , en vez de λ y μ . Estos valores se pueden obtener aplicando máxima verosimilitud al logaritmo de la función de probabilidad, entregándole la información requerida, detallada anteriormente.

³ Fuente: Fader, Peter; Hardie, Bruce; Lee, K. (2005), "Counting Your Customers, the Easy Way: An Alternative to Pareto/NBD Model", *Marketing Science* v24 issue 2, p275-284.

Algunas de las limitaciones de este modelo son:

- a) Las variables consideradas en este caso (RFM) no toman en cuenta otros aspectos de los clientes, tales como la preferencia que éstos tienen sobre ciertos productos.
- b) El comportamiento futuro sólo será pronosticado de manera similar al observado.

Sin embargo, para las estimaciones de la probabilidad de permanecer activo este modelo utiliza una función hipergeométrica gaussiana, la cual produce cierta complejidad en los cálculos. Por esta razón se utilizará una variación del modelo Pareto/NBD, el Beta Geométrico/NBD (BG/NBD), el cual, además de tener los supuestos del modelo Pareto/NBD, supone que la probabilidad de que un cliente deje de estar activo justo después de haber realizado una transacción sigue una distribución Beta(a,b). Luego, tomando este supuesto en conjunto con el que dice que las transacciones realizadas siguen un proceso de Poisson, se calcula la probabilidad de que un individuo específico deje de estar activo, lo cual se generaliza tomando la esperanza de éste. Luego, para calcular los parámetros necesarios, se toma el logaritmo de esta esperanza y se le calcula la máxima verosimilitud. En este caso, los parámetros estimados son r , α , a y b , y en base a ellos se puede estimar la probabilidad de que un cliente cualquiera permanezca activo.

La función con la cual se calcula esta probabilidad es la siguiente:

$$P(\text{alive}) = \frac{1}{\left\{ 1 + \delta_{x>0} \frac{a}{b+x-1} \left(\frac{\alpha+T}{\alpha+t_x} \right)^{r+x} \right\}}$$

donde

- x =número de transacciones.
- t_x =fecha última transacción.
- T =largo del período.
- a y b = parámetros de la función Beta.
- α =parámetro de escala de la tasa de transacciones de los clientes.

Los valores de x , t_x , y T se pueden obtener de los datos entregados por la empresa, sin embargo, los valores de a , b y α deben ser estimados.

La estimación de estos parámetros, como se mencionó anteriormente, se realiza mediante el cálculo de la máxima verosimilitud del logaritmo de la siguiente ecuación:

$$L(r, \alpha, a, b / X = x, t_x, T) = \frac{B(a, b+x)}{B(a, b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)\Gamma(\alpha+T)^{r+x}} + \delta_{x>0} \frac{B(a+1, b+x-1)}{B(a, b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)\Gamma(\alpha+t_x)^{r+x}}$$

donde:

- $B(p_1, p_2)$ = Función Beta de parámetros p_1 y p_2 en el caso generalizado.
- $\Gamma(p_1, p_2)$ = Función Gamma de parámetros p_1 y p_2 en el caso generalizado.
- r, α = Parámetros de forma y escala de la tasa de transacciones de los clientes.
- x = número de transacciones.
- t_x = última transacción realizada en el período de estudio.
- T = largo del período de estudio.

Esta ecuación es la función de probabilidad de un cliente escogido aleatoriamente con historia de compra $X=x, t_x, T$.

Además, este modelo supone que los clientes se encuentran activos al principio del período de estudio. Para que este supuesto se dé, se toma el inicio del período en el punto donde el cliente realiza su primera transacción, por lo tanto el largo de los períodos de estudio va a depender de cada cliente.

4.2.2. Cadenas de Markov⁴

Es un proceso estocástico cuyas probabilidades describen la manera en que evolucionará un sistema. Tiene las siguientes características:

- a) Posee la propiedad de considerar sólo el estado actual en que se encuentra el sistema para realizar la estimación, por lo tanto, es independiente de todos los estados pasados.
- b) Son flexibles y son capaces de modelar el ciclo de vida de los clientes, usando las variables RFM.
- c) Quedan definidas por sus estados y sus probabilidades de transición. Estos conceptos son definidos a continuación:
 - Estados: define los niveles entre los cuales un cliente puede transitar de un período a otro. En este caso, posibles niveles de la cadena de Markov serían “cliente leal” o “cliente fugado”.
 - Probabilidad de transición: es la probabilidad de que un cliente pase de un estado a otro en un período de tiempo.
- d) Las probabilidades salientes de cada estado deben sumar 1.

⁴ Fuente: Pfeifer, P.; Carraway, R. (2000), “Modeling Customer Relationships as Markov Chains”, *Journal of Interactive Marketing* v14, n°2: 43-55.

- e) Si la probabilidad de que un cliente se mantenga en el mismo estado durante el próximo período es p , la probabilidad de que se mantenga en el mismo estado después de N períodos es p^N .
- f) Cuando un cliente cae en el estado de “fugado”, no puede salir de ahí.

Luego, con este método se pueden identificar cuáles son los clientes que se encuentran más propensos a encontrarse en el estado de “fugados” y por lo tanto, se pueden aplicar sobre ellos acciones de marketing diferentes a las cuales se deben aplicar a aquellos clientes que encuentran en el estado “leales”.

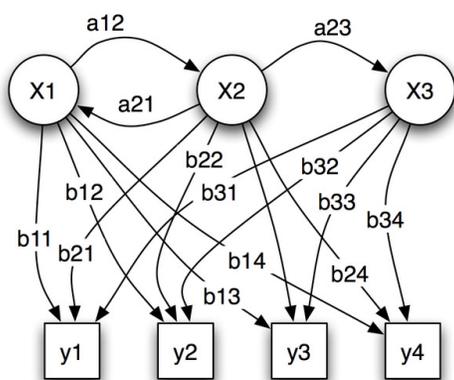
4.2.3. Cadenas de Markov Ocultas⁵

Es un modelo estocástico en el cual se asume que el sistema que está siendo modelado es un proceso de Markov que posee estados no observados.

En un modelo de Markov regular los estados son directamente visibles al observador, por lo que también lo son las probabilidades de transición entre ellos. En una Cadena de Markov Oculta, los estados no son visibles, sin embargo, los resultados visibles dependen directamente de los estados ocultos.

Al igual que las Cadenas de Markov regulares, posee la propiedad markoviana, que indica que los resultados obtenidos para un período dependen sólo de lo ocurrido durante el período anterior, no considerando lo ocurrido en períodos pasados.

Un ejemplo de Cadena de Markov Oculta es el siguiente:



donde:

- X = Estados Ocultos
- Y = Estados Visibles
- a_{ij} = probabilidad de pasar del estado oculto i al estado oculto j
- b_{ij} = probabilidad de llegar al estado visible j desde el estado oculto i

⁵ Fuente 1: Rabiner, Lawrence (1989), "A Tutorial on Hidden Markov Model and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, 77(2), 321-340.

Fuente 2: Netzer, O.; Lattin, J.; Srinivasan, V. (2007), "A Hidden Markov Model of Customer Relationship Dynamics", *Marketing Science*, 27(2), 185-204.

Para que una Cadena de Markov Oculta quede definida se necesita conocer lo siguiente:

1. Número de Estados Ocultos
2. Número de Estados Visibles
3. Matriz de transición de los Estados Ocultos. Las probabilidades salientes deben sumar 1. Cuando un cliente cae en el estado “fugado” no puede salir de ahí. En este caso, se define el estado 3 como “fugados”.
4. Distribución de Probabilidad de los Estados Observados
5. Distribución de probabilidad del Estado Inicial

Una de las principales diferencias que posee con las Cadenas de Markov regulares es el hecho de que no se conocen los estados a utilizar, pero se puede saber cuál es el número de estados indicado a usar gracias a los criterios BIC y/o AIC, los cuales se definen a continuación.

- Bayesian Information Criterion (BIC)

Es un criterio de selección de modelos, el cual considera el ajuste que poseen los modelos a la realidad y el número de parámetros que tiene cada modelo. De esta manera, su valor disminuye a medida que el ajuste mejora, pero aumenta cuando el número de parámetros lo hace, es decir, favorece el ajuste del modelo, pero penaliza su sobreajuste.

El cálculo de este indicador se realiza de la siguiente manera:

$$BIC = -2 * \ln(L) + k * \ln(n)$$

donde

- n= tamaño de la muestra
- k= numero de parámetros
- L= máximo valor de la función de probabilidad del modelo

Luego, la cantidad de estados ocultos a usar será aquella cuyo modelo posea el menor valor del BIC.

- Akaike Information Criterion

Similar al BIC, es un criterio de selección de modelos creado por Akaike. Al igual que el BIC, considera el ajuste del modelo a la realidad, penalizando el sobreajuste debido al aumento del número de parámetros.

Su principal diferencia con el BIC, es que el anterior es un poco más estricto que éste en la penalización.

Su cálculo se realiza como sigue:

$$AIC = 2 * k - 2 * \ln(L)$$

donde

- k = número de parámetros
- L = máximo valor de la función de probabilidad del modelo.

Por lo tanto, el modelo que posea el mínimo valor del AIC indica el número de parámetros que sería adecuado usar. El número de parámetros debería coincidir con lo indicado por el BIC.

- Regresión Logit

La regresión logística se usa para modelar la probabilidad de que un evento ocurra en la presencia de diversos factores. Su uso se da generalmente en ciencias médicas y sociales. En este caso se usará para las probabilidades relacionadas con las Cadenas de Markov Ocultas.

Esta regresión analiza los datos distribuidos binomialmente de la forma:

$$Y_i \sim B(p_i, n_i) \text{ para } i=1, \dots, m \text{ períodos}$$

donde

- n_i =son los números de ensayo de Bernoulli. Estos números son conocidos.
- p_i =son las probabilidades de que ocurra un evento. Estas probabilidades son desconocidas.

Un ejemplo para n_i podría ser la cantidad de clientes que tiene una empresa en el período i , mientras que p_i sería la probabilidad de que los clientes realicen compras.

Entonces, el modelo es obtenido en base a los valores de n_i y de las variables independientes que pudieran contener información acerca de la probabilidad final. Considerando estas variables como un vector x_i , el modelo queda de la siguiente forma:

$$p_i = E\left(\frac{Y_i}{n_i} \mid X_i\right)$$

Luego, los logits de las probabilidades de ocurrencia de un evento son modelados como una función lineal de X, de la forma:

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

siendo los parámetros β_j estimados mediante máxima verosimilitud.

Finalmente, las probabilidades de que un evento ocurra son estimadas de la siguiente forma:

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}}$$

Por lo tanto, está será la manera que se usará para el cálculo de las probabilidades en el modelo de Cadenas de Markov Ocultas.

4.2.4. Eficiencia de los modelos

Para medir la eficiencia de los modelos a usar, se puede utilizar las siguientes medidas de error.

- **MAPE**: Error porcentual absoluto medio.

Mide el error lineal producido por las predicciones de los modelos de manera porcentual.

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|\text{Real} - \text{Predicción}|}{|\text{Real}|}$$

- **MAE**: Error absoluto medio.

Mide el error lineal de las predicciones hechas por los modelos.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\text{Real} - \text{Predicción}|$$

- **MSE**: Error cuadrático medio.

Mide el error del pronóstico de los modelos de manera cuadrática, por lo tanto, es más severo cuando el error de pronóstico es muy grande.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\text{Real} - \text{Predicción})^2$$

Otra forma de medir la eficiencia de los modelos es mediante el uso de indicadores tales como las curvas ROC (Receiver Operating Characteristic) o las Lift Curves. En las primeras se mide la cantidad de errores que tiene el pronóstico hecho, ya que grafica la cantidad de clientes que fueron catalogados como fugados del total de los que realmente se fugan versus los que fueron catalogados como no fugados del total de los que realmente no se fugaron.

Por otro lado, las curvas Lift miden el desempeño del modelo, comparando la cantidad de clientes fugados encontrados con y sin el uso del modelo.

En base a los modelos usados se analizará cual es la mejor opción para medir la eficiencia de los modelos, de modo de poder escoger el que mejor predice la fuga de clientes para la empresa.

5. METODOLOGÍA

La metodología a seguir para desarrollar este estudio será la siguiente:

- **Revisión Bibliográfica:** se realizará una revisión de artículos relacionados con el tema de predicción de fuga de clientes, así como diversas tesis que hablen sobre el tema.
- **Aprendizaje y descripción de los diferentes modelos a usar:** junto con el estudio bibliográfico realizado, se escogerán los modelos a usar durante el desarrollo de este trabajo, tanto para la identificación de futuros clientes fugados, así como para los posibles criterios de definición de fuga a usar. Los modelos serán descritos en la sección Marco Conceptual. El objetivo es tener claridad acerca de los supuestos y los requerimientos que tiene cada uno de ellos.
- **Pre-procesamiento de Datos y Selección de las variables a usar:** de los datos disponibles, es probable que existan algunos que no agreguen valor a la predicción de fuga de clientes, por lo tanto se hará una selección de aquellos que sean más relevantes para los modelos a usar. Una vez hecho esto, se analizará la calidad de los datos y se realizará la limpieza que sea necesaria de manera que queden aptos para aplicar sobre ellos los diferentes modelos asociados a este estudio. Se les denominará datos procesados.
- **Definición de cliente fugado:** en base a lo anterior se analizarán los criterios mediante los cuales un cliente puede ser considerado en estado de fuga. Luego, en conjunto con la empresa se definirá cual es el más adecuado a este caso.
- **Aplicación de modelos:** A los datos procesados se les aplicará el modelo de Pareto/NBD y el modelo de Cadenas de Markov Ocultas. Luego, se analizarán los resultados obtenidos en cada uno de ellos.
- **Comparación de modelos y elección del más adecuado:** para cada modelo se realizará una estimación de la eficiencia asociada a ellos, de manera de poder compararlos. En base a esto se escogerá el más adecuado para el uso en la empresa.
- **Caracterización de los clientes según los resultados obtenidos:** usando el modelo escogido se hará una caracterización del comportamiento de los clientes.

6. ALCANCES

- La definición de un cliente fugado se hará en conjunto con la empresa, de manera que sea acorde a su realidad.
- Tanto los criterios de fuga como los modelos a usar se desarrollarán según el tipo de uso de la tarjeta. En base a esto se tendrán cuatro grupos: Tienda por Departamento, Mejoramiento del Hogar, Supermercados y No-Retail.
- El grupo No-Retail está conformado por: Alianzas, Avances y Viajes. Dentro de las Alianzas se considerarán sólo aquellas que no son descontadas automáticamente de la cuenta de los clientes. Esto significa que se excluyen los Pagos Automáticos de Cuentas (PAC) y los Seguros.
- La información a usar serán los datos transaccionales proporcionados por la empresa. No se utilizará ningún otro mecanismo de recopilación de información como focus groups o encuestas.
- El modelo entregará información acerca de las probabilidades de fuga de los diferentes grupos de clientes que posean comportamientos similares. Las probabilidades de fuga no se entregarán desagregadamente a nivel de cada cliente.
- Debido a la cantidad de clientes que posee la empresa, sólo se utilizará una muestra de ellos para realizar este trabajo.

7. RESULTADOS ESPERADOS

Se pretende contar con los siguientes entregables:

- Definición de un criterio mediante el cual un cliente pueda ser considerado en estado de fuga para la empresa.
- Una metodología mediante la cual se pueda caracterizar el comportamiento de los clientes con respecto a su fuga.
- Una caracterización de los diferentes grupos de clientes que poseen similar comportamiento al respecto.
- La probabilidad de fuga de cada uno de los segmentos caracterizados anteriormente.

- La documentación de los procedimientos utilizados para obtener los resultados, con el fin de que se puedan implementar en la empresa.
- Recomendaciones de acciones que busquen evitar la fuga de los clientes en base a los resultados obtenidos.

8. SELECCIÓN Y PREPROCESAMIENTO DE DATOS

A continuación se presenta una breve descripción de los datos seleccionados, los cuales se utilizarán durante el transcurso de este trabajo. Estos datos fueron proporcionados por la empresa.

8.1. Descripción de la Base de Datos en General

Se cuenta con información transaccional desde Enero de 2004 hasta Marzo de 2009 de un holding de retail. Las variables que se encuentran disponibles por cliente, ya que fueron entregadas por la empresa, se pueden observar en la Tabla 8.1.

Tabla 8.1: Variables Disponibles

| VARIABLES | DESCRIPCIÓN | TIPO DE VARIABLE |
|---------------|--|------------------|
| ID_CLIENTE | Es la identificación de cada cliente. Es utilizada sólo para este caso y no corresponde a la identificación interna de los clientes en la empresa. | Numérica |
| ID_AGNO_MES | Indica de manera numérica el mes y el año en los cuales los clientes poseen transacciones. | Numérica |
| DESC_AGNO_MES | Indica con palabras el mes y el año en el cual los clientes poseen transacciones. | Cadena |
| MONTO | Suma de los montos de las transacciones realizadas cada mes. | Numérica |
| N_BOLETAS | Número de boletas registradas por cada cliente durante el mes correspondiente. | Numérica |

Fuente: Elaboración Propia

A partir de esta información se calcularán las variables que sean requeridas por cada uno de los modelos a aplicar.

En la Tabla 8.2 se observa un resumen de los datos disponibles, entregados por las variables descritas en la Tabla 8.1.

Tabla 8.2: Resumen de Datos

| | |
|-------------------------|--------------|
| Número de clientes | 207.112 |
| Número de transacciones | 17.722.582 |
| Monto Total | MM\$ 510.419 |

Fuente: Elaboración Propia

Las transacciones fueron realizadas en todas las empresas pertenecientes a este holding, debido a lo cual existen 7 tipos de usos para la tarjeta. Sin embargo, para fines de facilitar el análisis, estos usos fueron divididos en 4 grupos, los cuales aparecen en la Tabla 8.3.

Tabla 8.3: Tipos de Uso de la Tarjeta del Holding

| | Tipo de Uso |
|--------|-------------------------|
| Tipo 1 | Tienda por Departamento |
| Tipo 2 | Mejoramiento del Hogar |
| Tipo 3 | Supermercado |
| Tipo 4 | No-Retail |

Fuente: Elaboración Propia

La distribución de los valores de la Tabla 8.2 a través de los diferentes tipos de uso de la tarjeta se encuentra en la Tabla 8.4.

Tabla 8.4: Resumen de Datos según Tipo de Uso

| TIPO DE USO | NÚMERO DE CLIENTES | NÚMERO DE TRANSACCIONES | MONTO TOTAL |
|-------------|--------------------|-------------------------|--------------|
| Tipo 1 | 207.112 | 6.553.053 | MM\$ 200.000 |
| Tipo 2 | 207.112 | 2.836.819 | MM\$ 100.000 |
| Tipo 3 | 207.112 | 477.331 | MM\$10.419 |
| Tipo 4 | 207.112 | 7.855.379 | MM\$ 200.000 |

Fuente: Elaboración Propia

Al observar el número de transacciones, se puede inferir que el Tipo 1 y el Tipo 4 son los usos que más frecuentes de la tarjeta, ya que poseen los más elevados números de transacciones. Por el contrario, el Tipo 3 es el uso que posee menos transacciones, lo que hace suponer que la tarjeta es menos usada en este tipo de tienda.

Cabe mencionar que el Tipo 3 es “Supermercados”, por lo tanto, el hecho de que el número de transacciones en este tipo de tiendas sea menor no necesariamente indica que ellos realicen menos compras aquí, sino mas bien que sus transacciones muchas veces tienen un monto bajo, el cual puede ser cancelado en efectivo y, como este estudio considera sólo las compras realizadas con la tarjeta de la empresa, las que son en efectivo no se encuentran dentro de los datos disponibles. De hecho, si bien el monto en este tipo de uso es el menor, no es un valor despreciable, lo que da para pensar que en general las pocas transacciones que se realizan en este tipo de tienda con la tarjeta poseen un monto no tan bajo.

Por otro lado, el principal negocio de la empresa es el de Tipo 1, por lo que llama un poco la atención que no lidere el número de transacciones.

Lo anterior podría explicarse con el hecho que el Tipo 4 está compuesto por 3 diferentes usos de la tarjeta: Alianzas, Avances y Viajes. Estos tres tipos se agruparon por conveniencia para la empresa, ya que son usos fuera del ámbito del retail, que es el principal rubro de ésta.

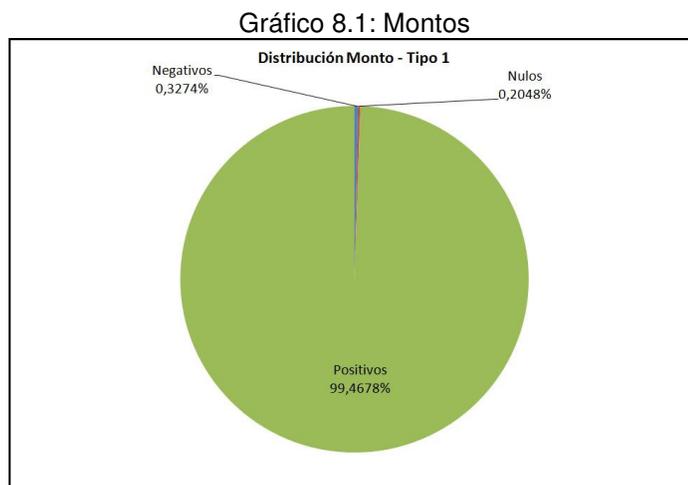
Además, se tiene que, entre los usos de la tarjeta en las tiendas del área del retail, las que pertenecen al Tipo 1 tienen un monto mayor que las otras, por lo que se considera el uso más importante. Esto era de esperarse, debido al funcionamiento de la compañía. Cabe mencionar también que el monto cuando la tarjeta se usa en las tiendas que no son del área de retail, es decir, de Tipo 4, es tan alto como cuando su uso es en las tiendas de Tipo 1. Esto puede deberse al hecho de que dentro de esta clasificación entran varios usos, tales como las alianzas o los viajes, por lo que al agregar todos los valores resulta una cifra importante. Sin embargo, como su aporte es similar al uso de la tarjeta en las tiendas de Tipo 1, resulta interesante también estudiar el comportamiento de los clientes en este caso.

8.2. Preprocesamiento de Datos

Para que los resultados que entregarán los modelos sean confiables es necesario que los datos que se les entreguen sean los correctos.

En este caso, se hace necesario observar las variables “Monto” y “N_Boletas”, ya que son las que podrían presentar errores. En particular, hay que comprobar que no existan valores nulos ni negativos en estas variables.

En el gráfico 8.1 se observa la distribución de los montos en cuanto a si son positivos, negativos o nulos, para el uso de la tarjeta en las tiendas de Tipo 1.



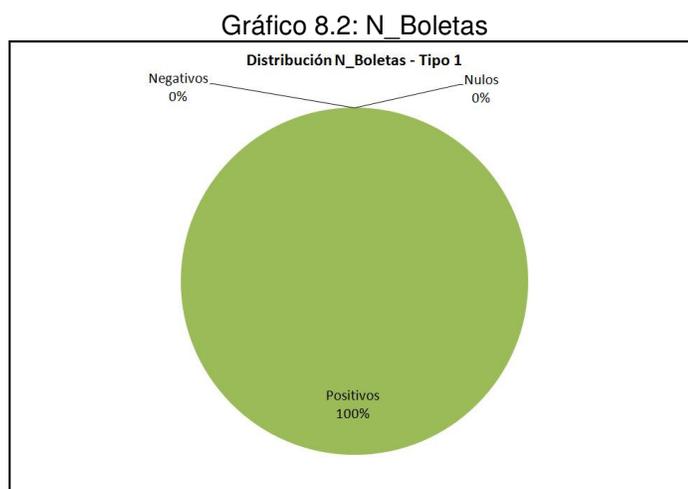
Fuente: Elaboración Propia

Como se puede observar, existe un 0,33% de transacciones que poseen montos negativos, las cuales corresponden a devoluciones hechas por los clientes. Además, hay 0,21% de las cuales son valores nulos, las que corresponden a cambios en la tienda o a errores del sistema. En total ambos valores conforman el 0,54% de las transacciones realizadas. Para realizar un mejor estudio, se ha decidido eliminar estos datos, debido al bajo porcentaje que significan.

Para la mayoría de los usos de la tarjeta se tiene una situación similar, por lo tanto se ha decidido eliminar los montos negativos y nulos para cualquier uso de la tarjeta.

Cabe mencionar el hecho de que la empresa informó que aquellas transacciones con montos nulos generalmente se deben a movimientos internos de la empresa, por lo que es positivo eliminar estos datos del estudio para enfocarse a lo que en realidad ocurre con los clientes.

Con respecto a la variable “N_Boletas”, en el gráfico 8.2 se puede observar la distribución de sus valores en cuanto a si son positivos, negativos o nulos, para el uso de la tarjeta en tiendas del Tipo 1. En él se puede observar que no hay valores negativos ni nulos para esta variable.



Fuente: Elaboración Propia

Para los otros tipos de uso de la tarjeta, existe un porcentaje mínimo de valores negativos y/o nulos, similar a lo ocurrido con los montos.

Por otro lado, cuando esta variable posee un valor negativo o nulo, es porque claramente se ha encontrado un error, lo más probable que de tipeo, ya que no se pueden realizar transacciones negativas y cuando no se realiza ninguna transacción la variable no figura con valor cero, sino que se encuentra en blanco.

Los gráficos de distribución de los montos y las boletas para el uso de la tarjeta en tiendas del tipo 2, 3 y 4 se encuentran en el Anexo 1.

Como se pudo observar anteriormente, los valores erróneos dentro de los datos, ya sean nulos o negativos, son un porcentaje muy pequeño, por lo cual serán eliminados del estudio.

8.3. División en Base de Entrenamiento y Base de Testeo

Para el desarrollo de este estudio se usará una base de datos para entrenar los modelos, es decir, para que éstos calculen los parámetros que necesitan, así como la matriz de probabilidades de transición, en el caso que corresponda. Esta será la base “Entrenamiento”.

Por otro lado, se tiene la base “Testeo”, en la cual se prueba el modelo ocupando los parámetros calculados con la base “Entrenamiento” y se observan los resultados obtenidos.

En Data Mining se suele usar una partición de 80%-20%, esto es, el 80% de la base original se define como base “Entrenamiento” y el 20% restante se ocupa como base “Testeo”.

En este caso, la partición se realizará con respecto a los períodos con los que se cuenta. En total, se tienen datos de 63 meses, lo que se traduce en aproximadamente 5 años, por lo tanto, se tomará el último año de datos, es decir, los últimos 12 meses disponibles como base “Testeo” y todos los meses anteriores serán la base “Entrenamiento”. Cada mes será considerado un período.

9. DESCRIPCIÓN DE VARIABLES

En esta sección se realizará una descripción de las variables a usar, esto es, después de haber realizado el preprocesamiento sobre los datos disponibles.

9.1. Descripción Base de Datos Limpia

Al eliminar los valores nulos y negativos de las variables “Monto” y “N_Boletas” disminuye el número de clientes que compra en cada tipo de uso de la tarjeta, ya que algunos de ellos figuraban sólo con estos valores.

En la tabla 9.1 se tiene un resumen de la información disponible en la Base de Datos preprocesada. En ella se puede observar claramente que las tiendas del Tipo 3 son el menor uso que los clientes le dan a la tarjeta, debido a la explicación que fue dada anteriormente. En este punto también influye el hecho de que los datos para las tiendas del Tipo 3 comienzan en Septiembre 2004, ya que antes de eso la empresa no se movía en el rubro de los supermercados.

Tabla 9.1: Resumen Base de Datos Preprocesada

| TIPO DE USO | NÚMERO DE CLIENTES | NÚMERO DE TRANSACCIONES | MONTO TOTAL |
|-------------|--------------------|-------------------------|--------------|
| Tipo 1 | 207.112 | 6.515.057 | MM\$ 200.000 |
| Tipo 2 | 165.181 | 2.836.758 | MM\$ 100.000 |
| Tipo 3 | 37.578 | 477.331 | MM\$10.419 |
| Tipo 4 | 178.869 | 7.855.014 | MM\$ 200.000 |

Fuente: Elaboración Propia

En resumen, se puede ver un comportamiento similar al encontrado antes de limpiar la base en cuanto al número de transacciones y los montos de éstas, variando principalmente la cantidad de clientes que las realizan.

9.2. Descripción Base de Datos Entrenamiento

Los datos disponibles para la Base de Datos “Entrenamiento” son los mismos de la Base de Datos anterior, sólo que esta vez se consideran sólo las transacciones hechas entre Enero 2004 y Marzo 2008.

En la Tabla 9.2 se presenta un resumen de los datos disponibles para la base “Entrenamiento”, separada por tipo de uso de la tarjeta. Nuevamente se observa una disminución en el número de clientes con respecto a la Base de Datos considerada hasta Marzo 2009. Esto se debe a que existen clientes que poseen su primera compra después de Marzo 2008, por lo tanto no están siendo considerados en esta Base de Datos.

Tabla 9.2: Resumen de datos – Base de Datos “Entrenamiento”

| TIPO DE USO | NÚMERO DE CLIENTES | NÚMERO DE TRANSACCIONES | MONTO TOTAL |
|-------------|--------------------|-------------------------|--------------|
| Tipo 1 | 189.634 | 5.452.884 | MM\$ 200.000 |
| Tipo 2 | 161.733 | 2.324.383 | MM\$86.347 |
| Tipo 3 | 29.748 | 304.219 | MM\$6.469 |
| Tipo 4 | 176.858 | 6.637.881 | MM\$ 200.000 |

Fuente: Elaboración Propia

Teniendo en cuenta lo anterior, es normal que tanto los valores de las transacciones como de los montos sean menores que en el caso anterior.

9.3. Variables Calculadas

Además de las variables que fueron entregadas por la empresa, se requieren nuevas variables para el desarrollo de los modelos. Estas variables son: Recency, Frecuency, Monetary Value, Primera Compra, Compra, Número de Transacciones.

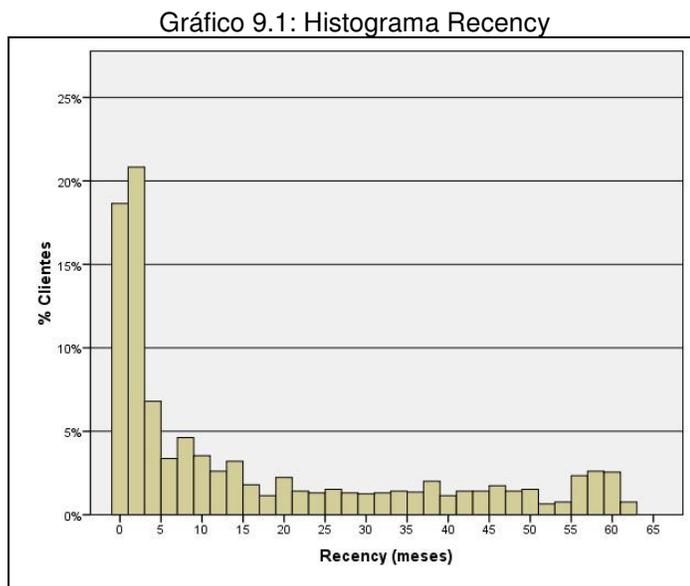
A continuación se realizará una descripción de cada una de las variables calculadas a partir de las variables iniciales.

- Recency:

El recency se define como el tiempo transcurrido desde la última compra. Para su cálculo se fue viendo mes a mes si los clientes poseían o no compras el mes anterior, por lo que, cuando esta variable tiene valor cero indica que el cliente realizó compra durante el mes inmediatamente anterior, cuando el valor es 1 la última compra fue 2 meses atrás y así sucesivamente. Por ejemplo, si se considera el recency del mes de Noviembre 2008, cuando toma valor cero significa que el cliente realizó compras en el mes de Octubre 2008, cuando su valor es 1 la última compra del cliente fue realizada el mes de Septiembre 2008 y si el recency es 3 la última compra fue hecha en Agosto 2008.

Para efectos de este estudio, esta variable fue nombrada como “R”.

En el Gráfico 9.1 se presenta la distribución de R a través de los clientes para el mes de Marzo 2009 cuando el uso de la tarjeta es en las tiendas del Tipo 1. En él se puede observar que cerca del 20% de los clientes posee recency 0 ó 1, lo cual indica que realizaron compras hace poco, ya sea el último mes o el anterior.



Fuente: Elaboración Propia

Se observa además que, a partir del recency 5, menos del 5% de los clientes posee cada uno de los recencys, lo cual indica que son más propensos a tener recencys más pequeños.

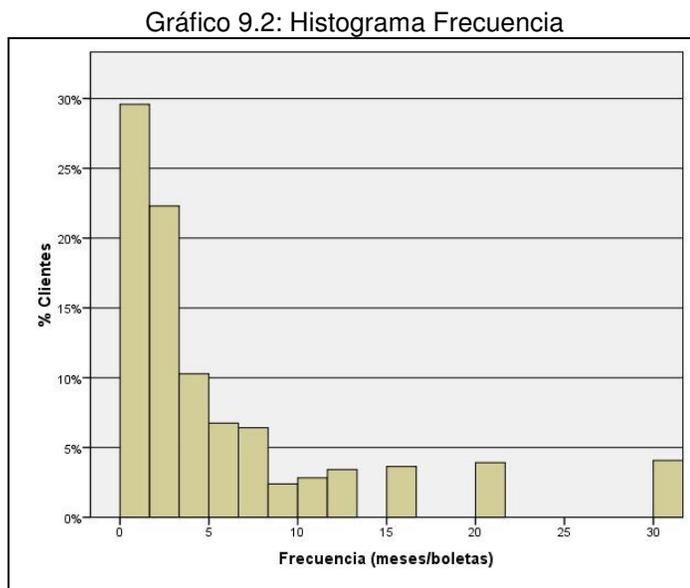
Este patrón de comportamiento se presenta a través de todos los meses cuando el uso de la tarjeta es en las tiendas de Tipo 1.

El comportamiento para los otros tipos de uso de la tarjeta sigue un patrón similar al observado en el gráfico 9.1, disminuyendo a distintas velocidades el porcentaje de clientes a medida que va aumentando el recency. Los gráficos de la distribución del recency a través de los clientes para el mes de Marzo de 2009 para los otros tipos de uso de la tarjeta se pueden observar en el Anexo 2.

- Frecuency:

Esta variable se define como el tiempo promedio transcurrido entre transacciones. La principal razón para tomar esta definición de frecuency es para que pueda ser comparada con el recency, como una opción para definir un criterio de fuga más adelante. Similar al caso del recency, esta variable fue nombrada como “F” para efectos de este estudio.

En el Gráfico 9.2 se presenta la distribución de F a través de los clientes en el mes de Marzo 2009 para el uso de la tarjeta en las tiendas del Tipo 1.



Fuente: Elaboración Propia

Similar al caso del Recency, se tiene que la mayoría de los clientes realiza transacciones con una frecuencia menor 5, lo cual indica que entre una compra y otra no suelen pasar más de 5 meses.

Sin embargo, a diferencia de lo que ocurre con el recency, una vez que el valor de F va aumentando, existen puntos donde no se encuentra ningún cliente.

Cabe mencionar que lo observado en el gráfico 9.2 se da de manera similar para el resto de los períodos en estudio y para los diferentes tipos de uso de la tarjeta.

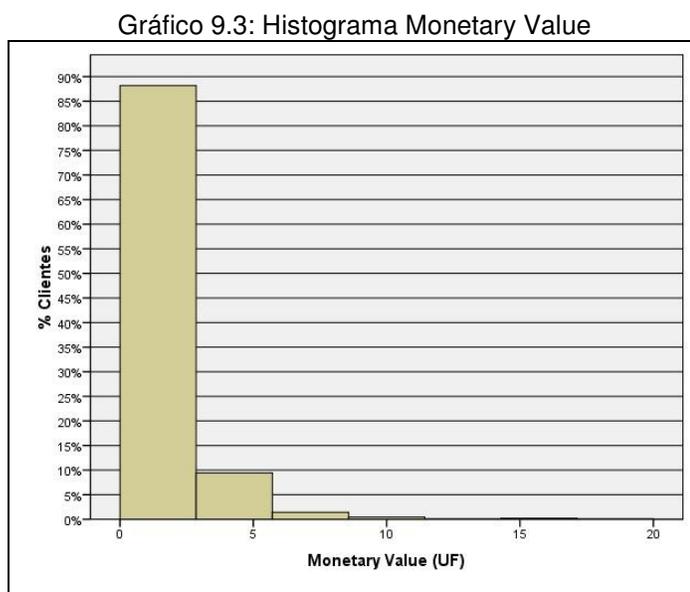
En el Anexo 3 se encuentran las distribuciones de las frecuencias para los usos de la tarjeta en las tiendas de los tipos 2, 3 y 4.

- Monetary Value

Este valor se define como el monto promedio de las transacciones de los clientes, por lo que mes a mes se fue calculando este promedio con todos los meses anteriores disponibles. Para este estudio, esta variable fue nombrada como “M”.

Además, debido al orden de los valores de esta variable, es que éstos se transformaron de pesos a UF⁶, de modo que pudiera ser comparable con el resto de las variables.

En el Gráfico 9.3 se presenta la distribución de la variable M a través de los clientes para el mes de Marzo 2009, para el uso de la tarjeta en las tiendas de Tipo 1.



Fuente: Elaboración Propia

Del gráfico 9.3 se desprende que los montos promedios de las transacciones de más del 80% de los clientes son menores a 5 UF.

Además, se tiene que el porcentaje de clientes va disminuyendo a medida que el monto promedio va aumentando, por lo que se tiende a pensar que los clientes son más propensos a realizar compras de montos no muy elevados.

Al igual que en los casos anteriores, este patrón es seguido por los clientes para en el resto de los meses del período en estudio, así como para los otros tipos de uso de la tarjeta.

Gráficos de las mismas características para los otros tipos de uso de la tarjeta se encuentran en el Anexo 4.

⁶ Se consideró la UF=\$21.000 como un valor aproximado.

- Primera Compra

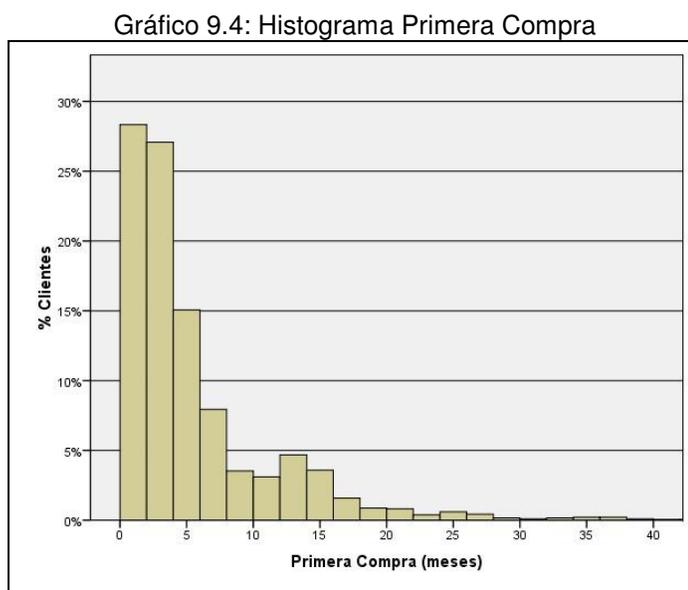
El primer mes en el cual se cuenta con los datos es Enero 2004. Sin embargo, no todos los clientes realizan compras ese mes en los diferentes tipos de uso de la tarjeta. Por lo tanto, la variable “Primera Compra” indica en qué mes, dentro de los que se encuentran disponibles, cada cliente efectúa su primera compra.

Cabe destacar el hecho de que esta variable no indica que el cliente no haya realizado compras antes de Enero 2004, ya que no se cuenta con información de los períodos anteriores. Por lo tanto, cuando dos clientes tienen como valor de esta variable, por ejemplo, “Junio 2004” se podría tener que uno de ellos haya realizado compras antes del período de estudio, mientras que el otro sea un cliente nuevo, como podrían ser 2 clientes nuevos o 2 antiguos.

En el Gráfico 9.4 se muestra la distribución de esta variable a través de los clientes para el uso de la tarjeta en las tiendas del Tipo 1. En él se puede observar el hecho de que las primeras compras de los clientes suelen producirse en los primeros períodos del estudio, lo cual hace pensar que los clientes no son nuevos, si no que vienen comprando desde antes. Sin embargo, esto no puede asegurarse, debido a que no se cuenta con información transaccional de antes de Enero 2004.

Además, se tiene que para los otros tipos de uso de la tarjeta, esta variable sigue el mismo patrón de lo observado en el gráfico 9.4.

En el Anexo 5 se encuentra la distribución de esta variable a través de los clientes para los otros tipos de uso de la tarjeta.



Fuente: Elaboración Propia

- Compra

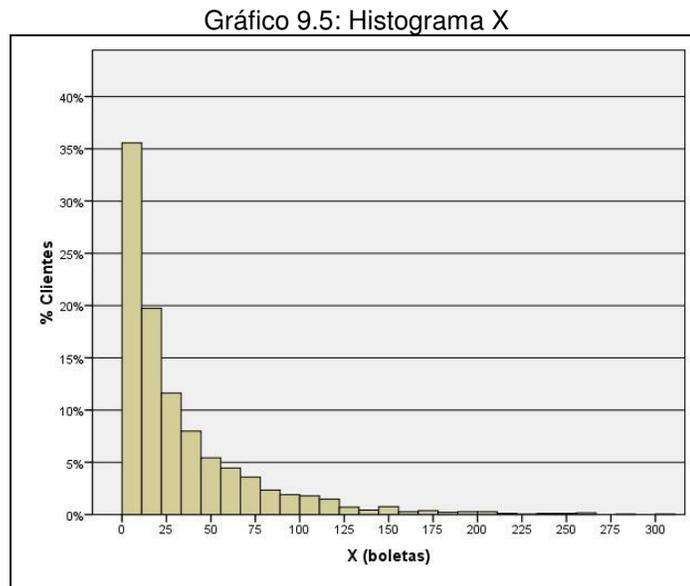
Esta variable es calculada mes a mes e indica si el cliente realiza o no transacciones durante el mes correspondiente. Es una variable binaria. Toma valor 1 cuando se observa compra y cero cuando no.

Esta variable será usada por un lado, para estimar el error de los modelos, ya que indica quienes en realidad compraron y quiénes no. Y por otro lado, será usada para las Cadenas de Markov Ocultas, ya que la compra es la acción que se puede observar de los clientes, por lo que el hecho de si compra o no, serán los estados observados de ese modelo.

- Número de transacciones

Esta variable indica la suma de las transacciones realizadas por los clientes desde Enero 2004 hasta el mes en el cual se está calculando su valor. Para efectos de este estudio, esta variable será llamada "X".

En el Gráfico 9.5 se muestra la distribución de esta variable a través de los clientes durante el mes de Marzo de 2009 para el uso de la tarjeta en las tiendas de Tipo 1.



Fuente: Elaboración Propia

En el gráfico 9.5 se muestra que el número de boletas suele ser bajo para la mayoría de los clientes. Esto podría explicarse por el hecho de que los clientes tienen la posibilidad de comprar con otros medios de pago, tales como efectivo o cheque, por lo que no todas sus boletas quedan registradas.

En el Anexo 6 se encuentran las distribuciones de esta variable para el mes de Marzo 2009 para los otros tipos de uso de la tarjeta.

10. DEFINICIÓN CRITERIO DE FUGA

Como en este caso se tiene un tipo de fuga no contractual, se hace necesario definir un criterio bajo el cual un cliente pueda ser considerado como fugado. Como se mencionó en la sección Marco Conceptual, se utilizarán 2 posibles criterios de fuga, basados en las variables RFM, los cuales son:

- Caso 1: Recency

En este caso se toma un valor máximo del Recency de los clientes, el cual, al ser sobrepasado, el cliente se considera como fugado.

- Caso 2: R/F

En este caso se hace una comparación entre el Recency y el Frecuency de los clientes, de manera de ver si la velocidad con la cual se encuentran realizando transacciones los clientes es la que habitualmente tienen. Al igual que en el caso anterior, aquí se define un valor para R/F, el cual, si es sobrepasado, el cliente se considera fugado.

Tipo de Uso: Tipo 1⁷:

Para el caso del uso de la tarjeta en las tiendas por departamento, se tiene el siguiente análisis acerca de los criterios de fuga mencionados anteriormente:

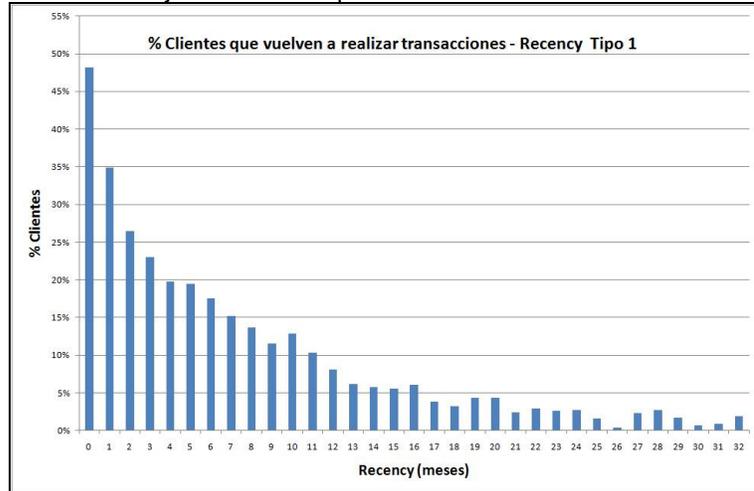
- Caso 1:

En el gráfico 10.1 se ilustra el porcentaje de clientes que realiza transacciones en el período siguiente de cuando tuvieron un cierto valor para su Recency. Este cálculo fue realizado utilizando la información histórica entre Enero 2004 y Marzo 2009.

En él se observa, por ejemplo, que cuando los clientes poseen Recency=0 mes, sobre el 45% de ellos realiza transacciones en el período siguiente, por lo tanto, se puede afirmar que ese no sería un corte adecuado para el criterio de fuga, debido al alto porcentaje de clientes que realiza transacciones después de haber tenido ese recency. Por otro lado, cuando el recency toma un valor de 17 o más, el porcentaje de clientes que vuelve a comprar es menos a un 5%, lo cual pareciera ser un valor más razonable para que el cliente sea considerado como fugado.

⁷ El análisis de ambos criterios de fuga para los otros tipos de uso se encuentra en los Anexo 7.

Gráfico 10.1: Porcentaje de clientes que vuelven a realizar transacciones-Recency



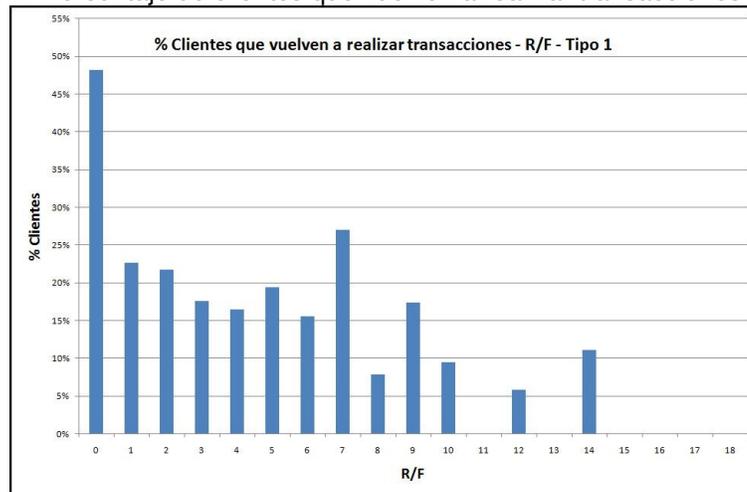
Fuente: Elaboración Propia

- Caso 2:

A continuación se presenta el Gráfico 10.2, el cual posee las mismas características que el anterior, pero ahora se usa para visualizar lo que ocurre tomando como criterio de fuga el caso 2 (R/F).

En él se observa claramente que sobre el 45% de los clientes vuelve a realizar transacciones cuando la razón R/F es cero. Esto era de esperarse, ya que cuando la variable R es cero se tiene el mismo porcentaje de clientes que vuelve a comprar, como se puede observar en el gráfico 10.1. Cuando R/F es 1 o más el porcentaje de clientes baja a cerca de 20%. Luego de eso se mantiene entre 20% y 15%, hasta que toma el valor 7, después del cual no sigue ningún patrón aparente.

Gráfico 10.2: Porcentaje de clientes que vuelven a realizar transacciones-R/F



Fuente: Elaboración Propia

- Criterio de Fuga Definido:

Finalmente, en conjunto con la empresa, se decidió que el criterio de fuga a usar sería el Caso 1, debido a que satisface las necesidades de la empresa con respecto a la identificación de los clientes fugados.

Se determinó para cada tipo de uso de la tarjeta un corte de Recency, de manera que si el cliente lo alcanza o lo sobrepasa, será considerado como fugado. Se realizó esto en lugar de poner un Recency límite igual para todos los tipos de uso debido a que éstos tienen diferentes características, por lo que no hubiera sido óptimo realizarlo en conjunto para todos.

Los Recency cortes, así como el porcentaje de clientes fugados actualmente por tipo de uso de la tarjeta se encuentra en la Tabla 10.1. El porcentaje de clientes fugados que en ella se encuentra es para el mes de Marzo de 2009. Además, el porcentaje de clientes que vuelve a comprar que aparece en ella, es un valor calculado con los datos históricos.

Cabe mencionar el hecho de que, como los criterios están definidos según el tipo de uso de la tarjeta, es posible que un cliente pueda clasificarse como fugado en un tipo de uso puede no estarlo en otro. El objetivo principal de esto es saber hacia que área enfocar las acciones de marketing que serán aplicadas sobre cada cliente para evitar que se fugue, las cuales pueden ser muy diferentes cuando la fuga se produce, por ejemplo, en las Tiendas por Departamento versus las fugas de las tiendas de Mejoramiento del Hogar.

Tabla 10.1: Criterio de fuga y Porcentaje de clientes fugados

| TIPO DE USO | REGENCY CORTE | PORCENTAJE FUGADOS (Marzo 2009) | % FUGADOS QUE VUELVEN A COMPRAR (Histórico) |
|-------------|---------------|---------------------------------|---|
| Tipo 1 | 12 | 41% | 8,1% |
| Tipo 2 | 12 | 49,9% | 7,2% |
| Tipo 3 | 3 | 67% | 13% |
| Tipo 4 | 7 | 56% | 9,8% |

Fuente: Elaboración Propia

Como se puede observar, para el caso de las tiendas del Tipo 1, considerando el mes de Marzo de 2009, se tiene que el 41% de los clientes aproximadamente poseen Recency 12 o más, por lo tanto ese porcentaje de clientes es el que se considera como fugado en la situación actual.

Además, se tiene que el porcentaje de clientes que se considera fugado y no se encuentra en ese estado históricamente es del orden del 8% al 13%. Esto es considerado un porcentaje pequeño en la empresa, debido a la gran cantidad de clientes que poseen.

Para el caso de uso de la tarjeta en las tiendas del Tipo 3 se tiene un porcentaje mucho mayor de clientes que vuelven a comprar una vez que fueron considerados como fugados. A su vez, el porcentaje de clientes que puede ser considerado como fugado en la actualidad para este tipo de uso de la tarjeta es muy alto, lo cual podría explicarse debido al hecho de que los clientes no siempre realizan sus transacciones con la tarjeta, debido a que los montos de compra son más pequeños, por lo cual pueden realizarlas en efectivo, datos de los cuales no se dispone. Además, como se observó en la descripción de las variables, este tipo de uso es el que cuenta con menos número de transacciones, menores montos y menos usuarios, lo cual se traduce en menos información para este tipo de uso.

Sin embargo, el porcentaje de clientes que vuelve a realizar compras que se observa en la Tabla 10.1, es un valor histórico, por lo que sería interesante saber cual es el porcentaje real de los clientes que vuelven a comprar una vez que se les consideró como fugados.

Como ejemplo, se verá que ocurre cuando se toman los clientes fugados en Noviembre de 2008 y se mide que porcentaje realiza compras entre Diciembre 2008 y Marzo 2009. Estos valores se encuentran en la Tabla 10.2.

Tabla 10.2: Porcentajes de clientes que vuelven a realizar compras

| | % Clientes Fugados Noviembre 2008 | % Fugados que compran entre Dic-08 y Mar-09 |
|--------|--------------------------------------|--|
| Tipo 1 | 40% | 5% |
| Tipo 2 | 48% | 2% |
| Tipo 3 | 65% | 3% |
| Tipo 4 | 54% | 4% |

Fuente: Elaboración Propia

Como se puede observar, los porcentajes de clientes que vuelven a comprar después de ser considerados como fugados es muy bajo, por lo tanto se puede decir que el criterio de fuga definido es bueno, ya que cumple con tener un error menor al 10%.

11. MODELO PARETO/NBD

Este modelo calcula la probabilidad de que un cliente permanezca activo después que realizó una compra.

Para realizar este cálculo se utiliza el número de transacciones realizadas por los clientes, la fecha de la última transacción del cliente dentro del período en estudio y el largo de este período.

El número de transacciones es una variable que ya fue calculada. Para el cálculo de la fecha de la última compra, se usa la variable “compra” que toma el valor 1 el mes en que se observa compra por parte del cliente y el valor 0, por lo tanto, la fecha de la última compra será indicada por el último mes donde esta variable tenga el valor 1.

Con respecto al largo del período, para realizar su cálculo se tomará el largo inicial del período en estudio y se le restarán la cantidad de meses que transcurren entre Enero 2004, que es donde comienzan siempre los periodos, hasta la fecha de la primera compra.

Para este modelo se confeccionó un programa en Matlab, el cual recibe las variables mencionadas en los párrafos anteriores y calcula los parámetros necesarios para el cálculo de la probabilidad de que un cliente se encuentre activo, así como esta probabilidad.

Inicialmente se le entregaron al modelo los datos disponibles de todos los clientes que poseían compras dentro del período de estudio. Sin embargo, al momento de realizar los cálculos, el programa tenía problemas para calcular la probabilidad de permanecer activo cuando el número de transacciones de los clientes era muy grande, debido a que la función Gamma tiende a infinito cuando esto ocurre.

Por esta razón, se decidió no incorporar en el modelo a los clientes que tuvieran un número de transacciones superior a 100, para así evitar los problemas anteriores. Estos clientes se considerarán como clientes leales, debido a que sus transacciones son muy superiores al promedio, el cual es 32.

En la Tabla 11.1 se muestra el porcentaje de clientes que es considerado como leal y que se fuga. En ella se puede observar el bajo porcentaje de clientes que son considerados como leales y que luego se fugan, por lo tanto se puede decir que no hay problema con considerar a aquellos clientes como leales.

Tabla 11.1: Porcentaje de Leales que se fugan

| TIPO DE USO | % LEALES QUE SE FUGAN |
|-------------|-----------------------|
| Tipo 1 | 2,75% |
| Tipo 2 | 0% |
| Tipo 3 | 0% |
| Tipo 4 | 0% |

Fuente: Elaboración Propia

Para realizar los cálculos, el modelo Pareto/NBD estima 4 parámetros (r , α , a , b), de los cuales dos se encuentran asociados a la distribución de la tasa de compra de los clientes y los otros dos a la distribución de su tasa de fuga.

Sin embargo, en este caso se realizó una modificación al modelo, de tal manera de poder incluir el criterio de fuga en él. Por esta razón, se agregaron 2 nuevos parámetros, los cuales se usan de la siguiente manera:

$$r = r_1 + r_2 * t_x + r_3 * (t_x - CF)^2$$

donde CF es el criterio de fuga definido para cada tipo de uso de la tarjeta de la empresa.

De esta manera, se estima el valor de r_1 , r_2 , r_3 , además de α , a y b y se calcula el valor de r con estos parámetros para poder estimar la probabilidad de permanecer activo.

Como se mencionó en la sección Marco Conceptual, la estimación de los parámetros se realiza mediante estimación de máxima verosimilitud de

$$L(r, \alpha, a, b / X = x, t_x, T) = \frac{B(a, b+x)}{B(a, b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)\Gamma(\alpha+T)^{r+x}} + \delta_{x>0} \frac{B(a+1, b+x-1)}{B(a, b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)\Gamma(\alpha+t_x)^{r+x}}$$

donde:

- $B(p_1, p_2)$ = Función Beta de parámetros p_1 y p_2 en el caso generalizado.
- $\Gamma(p_1, p_2)$ = Función Gamma de parámetros p_1 y p_2 en el caso generalizado.
- r, α = Parámetros de forma y escala de la tasa de transacciones de los clientes.
- a y b = parámetros de la función Beta.
- x = número de transacciones.
- t_x = última transacción realizada en el período de estudio.
- T = largo del período de estudio.

Para realizar esta estimación se utilizó en Matlab la función *fminsearch*, a la cual se le entregan los valores iniciales de los parámetros y la función que se quiere maximizar. Para este caso, los valores iniciales de los parámetros se consideraron como 0,001 para todos ellos. Uno de los beneficios de ocupar esta función en vez de otras de Matlab, es el hecho de que no necesita que se definan cotas superiores ni inferiores para los parámetros.

Finalmente, los parámetros estimados por el modelo Pareto/NBD para el uso de la tarjeta en las tiendas del Tipo 1 se pueden observar en la Tabla 11.2

Tabla 11.2: Parámetros a usar por el modelo Pareto/NBD

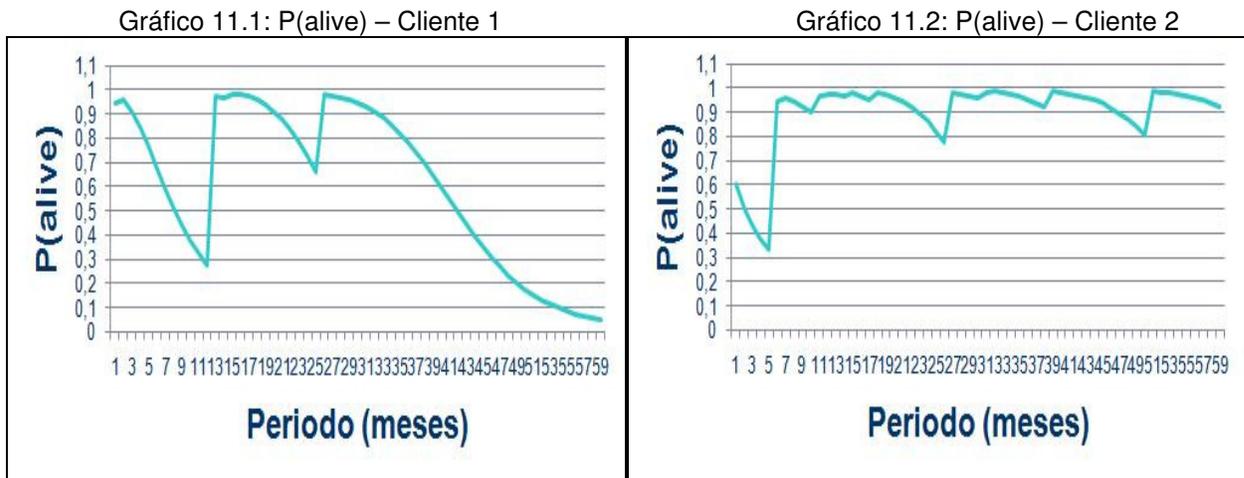
| | |
|----------|----------------|
| r_1 | -0,69931 |
| r_2 | 0,029162 |
| r_3 | 0,000142 |
| α | exp(0,915733) |
| a | exp(-1,682186) |
| b | exp(1,663019) |

Fuente: Elaboración Propia

Los parámetros que estima el modelo para los otros tipos de uso de la tarjeta se encuentran en el Anexo 8.

Luego, con los valores de r , α , a y b se calculó la probabilidad de que los clientes permanezcan activos para cada tipo de uso de la tarjeta de la empresa.

En los gráficos 11.1 y 11.2 se muestra la probabilidad de permanecer activo para dos clientes diferentes escogidos de manera aleatoria, como una manera de ejemplificar el comportamiento de esta probabilidad para cada cliente. El cliente 1, del gráfico 11.1, se encuentra fugado en el último período graficado, el que corresponde a Noviembre 2008, mientras que el cliente 2, del gráfico 11.2, no cae dentro del criterio de fuga definido.



Fuente: Elaboración Propia

Fuente: Elaboración Propia

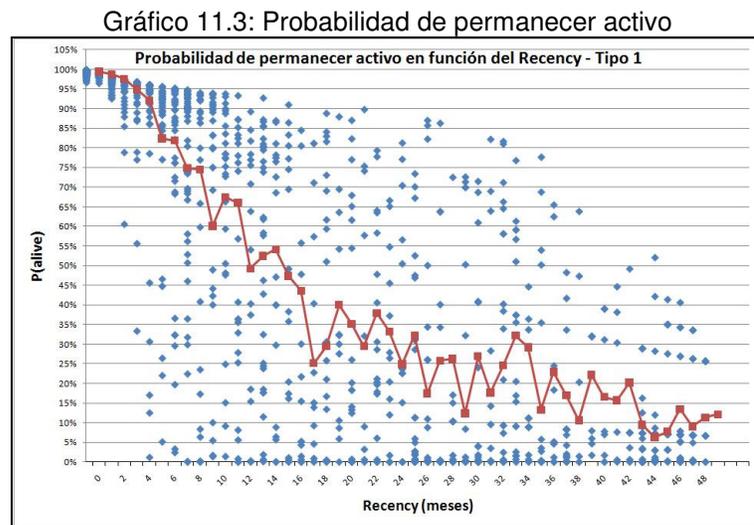
Como se puede observar en el Gráfico 11.1, la probabilidad de permanecer activo del cliente 1 al final del período va decreciendo hasta un valor muy bajo. Por otro lado, en el Gráfico 11.2 se puede observar que la probabilidad de permanecer activo del cliente 2 sigue manteniéndose alta al final del período en estudio, lo cual concuerda con el hecho de que no se encuentra fugado en ese instante.

Además, cabe destacar de los gráficos 11.1 y 11.2 que los instantes en que la curva sube, llegando la probabilidad de permanecer activo muy cercana a 1, son los períodos en los cuales los clientes realizan compras, por lo que además se puede concluir que el cliente 2 es un cliente mucho más frecuente que el 1, por lo que no es de extrañar que se encuentre activo al final del período mientras el otro cae dentro del criterio para ser considerado como fugado.

Para el caso general de todos los clientes, se tiene que en el gráfico 11.3 se muestra la probabilidad de permanecer activo en función del Recency observado por cada uno de ellos, cuando el uso de la tarjeta es en las tiendas del Tipo 1, para el mes de Marzo 2008. En él se puede observar que para cada uno de los recencys se tienen diferentes probabilidades de permanecer activo. Esto se debe principalmente al hecho de que

esta probabilidad depende del largo del período en estudio, el cual variará según el momento en el cual el cliente haya realizado su primera compra, por lo tanto el largo del período en estudio será diferente para cada cliente, a pesar de que puedan tener igual recency o igual número de transacciones.

Además, en el gráfico 11.3 se observa el promedio de las probabilidades de permanecer activo por recency a través de una línea roja. Si bien esta línea no es recta, se puede afirmar que a medida que el recency va aumentando, la probabilidad promedio de que el cliente permanezca activo va disminuyendo, lo cual era de esperarse.



Por otro lado, se tiene una concentración de puntos antes del recency indicado como punto de fuga, el cual es 12 para este tipo de uso, sobre una probabilidad aproximada de un 80%, dispersándose un poco más los puntos después de eso.

Es importante señalar que las probabilidades de permanecer activo con respecto al Recency para los otros meses pertenecientes al estudio poseen características muy similares a lo observado en el Gráfico 11.3, por lo que ocupar este mes como muestra es una buena representación de los resultados.

Con respecto a los otros tipos de uso de la tarjeta, se puede mencionar el hecho de que siguen el mismo patrón observado en el Gráfico 11.3, sólo que varía la velocidad en la cual va disminuyendo la probabilidad de permanecer activo a medida que el recency crece.

Gráficos de similares características para los otros tipos de uso de la tarjeta se pueden encontrar en el Anexo 9.

Luego, lo que se busca es encontrar un punto, basándose en la probabilidad recién calculada, que indique en que instante sería conveniente aplicar sobre los clientes

acciones de marketing, con el objetivo de evitar que se fuguen a futuro. Para esto, se comparará la probabilidad de permanecer activo calculada con el modelo de Pareto/NBD, con las compras que tienen los clientes después que se les calculó la probabilidad de modo de saber que porcentaje de ellos realizan compras según el rango de probabilidad al cual pertenezcan.

Esta comparación se realizará para los últimos 12 meses, de modo de tener una visión general con respecto al comportamiento del cliente con plazo un año a partir de la fecha donde se calculó su probabilidad de permanecer activo.

Debido a que se cuentan con datos hasta Marzo 2009, se realizó el cálculo de la probabilidad de permanecer activo en Marzo 2008 y se comparó con el porcentaje de compras desde Abril 2008 hasta Marzo 2009. Los resultados obtenidos en esta comparación se observan en la Tabla 11.3.

Tabla 11.3: Porcentaje de clientes que realiza compra el año siguiente en base a su Probabilidad de permanecer activo

| P(alive) | % Clientes que realiza compra los siguientes 12 meses |
|----------|---|
| 0-0,1 | 9,714% |
| 0-0,2 | 11,085% |
| 0-0,3 | 12,449% |
| 0-0,4 | 14,313% |
| 0-0,5 | 16,216% |
| 0-0,6 | 16,978% |
| 0-0,7 | 17,516% |
| 0-0,8 | 19,742% |
| 0-0,9 | 23,797% |

Fuente: Elaboración Propia

Lo que se busca es tener cerca de un 15% de error, ya que así se estaría considerando al 85% de los clientes, lo cual es un valor aceptable para la empresa. Debido a esto, es que se consideró un punto de corte cuando la $P(\text{alive}) \leq 0,5$, ya que se tiene que de los clientes que están dentro de este rango, alrededor de 16% posee compra los próximos 12 meses, lo cual indica que el resto se fuga en ese período, razón por la cual se hace interesante aplicar acciones de marketing sobre ellos para evitar que esto pase.

Se realizó el mismo análisis para los otros tipos de uso de la tarjeta, el cual se puede observar en el Anexo 10.

En la Tabla 11.4 se puede observar el resumen de los puntos de corte para los distintos tipos de uso de la tarjeta, según el análisis realizado anteriormente.

Tabla 11.4: Puntos de Corte – Modelo Pareto/NBD

| Tipo de Uso | Punto de Corte | % Clientes que realizan compras los siguientes 12 meses |
|-------------|----------------------------|---|
| Tipo 1 | $P(\text{alive}) \leq 0,5$ | 16% |
| Tipo 2 | $P(\text{alive}) \leq 0,2$ | 15,2% |
| Tipo 3 | $P(\text{alive}) \leq 0,1$ | 22,4% |
| Tipo 4 | $P(\text{alive}) \leq 0,6$ | 14,9 |

Fuente: Elaboración Propia

Se puede observar en la Tabla 11.4 que el Tipo 3 de uso de la tarjeta tiene un punto de corte cercano al 20%. Esto se definió así por el hecho de que para considerar el 15% se hubiera tenido que tomar una probabilidad demasiado pequeña. Lo anterior puede deberse al hecho de que el recency con el cual los clientes son etiquetados como fugados es considerablemente menor que en el resto de los casos, y al hecho de que es el tipo de uso que posee menos clientes, por lo tanto, la información con que se cuenta es menor que en los otros tipos de uso de la tarjeta.

Luego, en base a lo obtenido anteriormente, se definen 4 grupos, con los cuales se caracterizará a los clientes. En la Tabla 11.5 se observan los grupos y los criterios que se usaron para definirlos.

Tabla 11.5: Grupos Modelo Pareto/NBD

| GRUPOS | CARACTERISTICAS |
|-----------|-----------------------------|
| Leales | $X > 100$ |
| Normales | $P(\text{alive}) > P.C.$ |
| Propensos | $P(\text{alive}) \leq P.C.$ |
| Fugados | $R \geq C.F.$ |

Fuente: Elaboración Propia

Donde se tiene que X es la variable que indica el número de transacciones; $P.C.$ es el punto de corte definido en la Tabla 11.4; $C.F.$ es el criterio de fuga definido anteriormente para cada tipo de uso de la tarjeta y R es el recency del cliente.

Por lo tanto, se tiene que a los clientes que hay que atacar para evitar su posible fuga, son aquellos que se encuentran en el grupo de los propensos.

11.1. Error Modelo Pareto/NBD

Las posibles medidas de error descritas en la sección Marco Conceptual consideran el hecho de que se conoce el valor actual de lo que se está estimando, que en este caso es la probabilidad de que un cliente permanezca activo.

Sin embargo, como se tiene una fuga no contractual, no se cuenta con la información real sobre la probabilidad que tienen los clientes de fugarse, por lo que se comparará el promedio de las probabilidades calculadas por periodo (lo que será considerado como la estimación), con el porcentaje de clientes que efectivamente se fugan ese periodo (lo que será considerado como lo que ocurre en realidad).

Luego, en la Tabla 11.6 se pueden observar los valores del Error Absoluto (MAE) y del Error Cuadrático (MSE) para el modelo de Pareto/NBD, para los diferentes tipos de uso de la tarjeta.

Tabla 11.6: Errores asociados al Modelo Pareto/NBD

| TIPO DE USO | MAE | MSE |
|-------------|------|------|
| Tipo 1 | 7,3% | 1,2% |
| Tipo 2 | 16% | 4,5% |
| Tipo 3 | 15% | 3,7% |
| Tipo 4 | 6,7% | 0,8% |

Fuente: Elaboración Propia

Al observar la Tabla 11.6, se puede decir que el modelo Pareto/NBD se ajusta bastante bien a la realidad, y que los valores de sus errores son muy bajos.

Cabe mencionar el hecho que en los 2 tipos de uso que posee más transacciones la empresa (Tipo1 y Tipo 4) es donde mejor se ajusta el modelo.

Por otro lado, si se considera el análisis realizado anteriormente, gracias al cual se determinó el punto en donde se tiene que empezar a aplicar acciones sobre los clientes, se tiene que los porcentajes de error en ese caso son de alrededor de 15%, lo cual, a pesar de ser mayor a lo obtenido en la Tabla 11.6, no es un resultado desalentador.

12. CADENAS DE MARKOV OCULTAS

Este es el segundo enfoque propuesto para caracterizar la fuga de clientes. Con este modelo se tienen tanto estados ocultos como estados observados donde se pueden encontrar los clientes. Sin embargo, sólo se considerarán los estados ocultos para realizar el análisis acerca de en qué instante sería conveniente empezar a aplicar acciones de marketing sobre los clientes para evitar su posible fuga. Esto se hará a través de la probabilidad que poseen los clientes de estar en los estados ocultos diferentes al “fugados”.

Como se menciona en la sección Marco Conceptual, se necesita lo siguiente para que una Cadena de Markov Oculta quede definida:

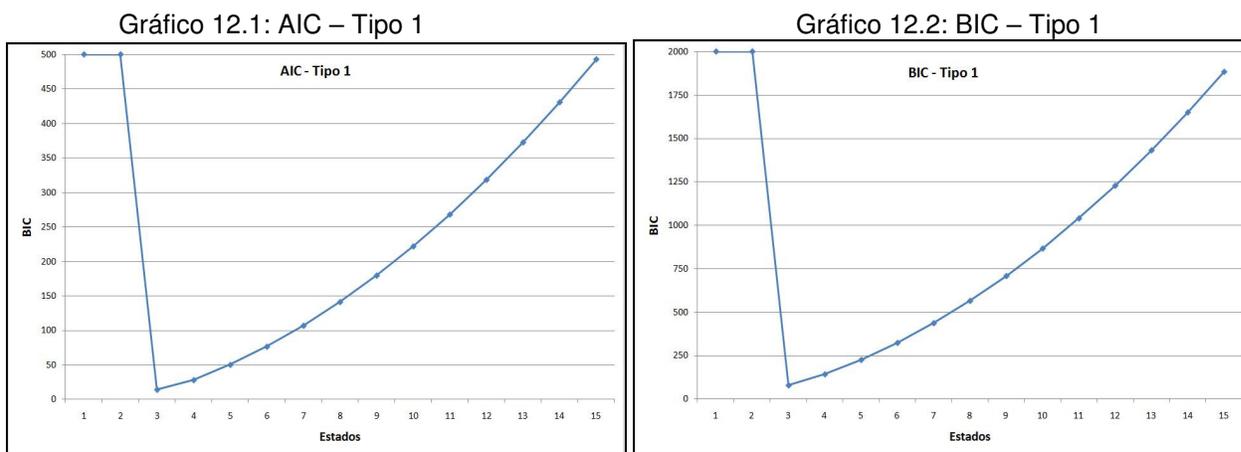
- Número de Estados Ocultos del modelo.
- Número de Estados Observados del modelo.
- Matriz de Transición de los Estados Ocultos.
- Distribución de Probabilidad de los Estados Observados.
- Distribución del Estado Inicial.

A continuación se detallará la manera en que se calcularon cada uno de los pasos anteriores y se definirá la Cadena de Markov Oculta en base a los resultados obtenidos en cada paso.

- Número de Estados Ocultos

Para calcular el número de Estados Ocultos se utilizó tanto el criterio AIC como el BIC, ya que arrojaron el mismo número de estados a usar para cada uno de los tipos de uso de la tarjeta.

En los Gráficos 12.1 y 12.2, así como en la Tabla 12.1, se observan los valores obtenidos para los criterios AIC y BIC respectivamente, para el uso de la tarjeta en las tiendas de Tipo 1.



Fuente: Elaboración Propia

Fuente: Elaboración Propia

Tabla 12.1: AIC y BIC - Tipo 1

| ESTADOS | AIC | BIC |
|---------|-----------|-----------|
| 1 | inf | inf |
| 2 | inf | inf |
| 3 | 14,543384 | 80,74711 |
| 4 | 28,394815 | 144,25134 |
| 5 | 50,530098 | 227,07337 |
| 6 | 76,832279 | 325,09625 |
| 7 | 107,19244 | 438,21107 |
| 8 | 141,50094 | 566,30819 |

Fuente: Elaboración Propia

Cuando se tiene 1 y 2 estados, tanto el AIC como el BIC tienden a infinito. Esto puede deberse al hecho de que con tan pocos estados el modelo se ajusta muy poco a la realidad. Además, una vez que se tiene un valor mínimo para ambos criterios, sus valores van aumentando paulatinamente a medida que se agregan más estados, lo cual era de esperarse.

Para el caso del uso de la tarjeta del Tipo 1, se tiene que el menor valor, tanto para el AIC como para el BIC se da cuando se tiene el modelo con 3 estados, por lo tanto, este será el número de estados que se usará para este tipo de uso de la tarjeta.

Para los otros tipos de uso de la tarjeta se tiene un comportamiento similar del AIC y del BIC. En todos los casos estos valores concuerdan entre sí al designar el número de estados óptimos a usar, que, coincidentemente, para todos los tipos de uso de la tarjeta son 3 estados.

Los gráficos del AIC y BIC para los otros tipos de uso de la tarjeta se encuentran en el Anexo 11.

- Número de Estados Observados

En este caso, se puede observar mensualmente si los clientes realizan o no compras, el monto de éstas y la cantidad de transacciones efectuadas por mes. Basándose en esto, se tendrá que los estados observados de los clientes serán si realiza o no compra.

- Matriz de Transición de los Estados Ocultos

Las probabilidades de transición de los estados ocultos fueron calculadas con regresión logit. Además se definió que el último estado es el de fuga, por lo tanto es un estado recurrente.

En la Tabla 12.2 se presenta la matriz de transición para el uso de la tarjeta en las tiendas del tipo 1, mientras que en la Tabla 12.3 se tienen los nombres asignados a cada uno de los estados, basándose en lo observado en la Tabla 12.2.

Se tiene que el estado 3 fue definido como “fugados”, el cual es considerado un estado absorbente, debido al supuesto de que una vez que el cliente cae en este estado no puede salir de él. Por esta razón es que la probabilidad de que un cliente se encuentre en el estado 3 una vez que se encuentra en él es siempre 1.

Además, debido a que uno de los estados ocultos fue definido como “fugados”, es interesante definir los otros en base a alguna característica que pueda hacerlos llegar a ese estado.

Se tiene que una de las principales razones por las cuales los clientes terminan su relación con la empresa es porque no se sienten completamente satisfechos con lo que reciben de ella⁸.

Tabla 12.2: Matriz de Transición Estados Ocultos

| | Estado 1 | Estado 2 | Estado 3 |
|----------|----------|----------|----------|
| Estado 1 | 0,0042 | 0,993 | 0,0028 |
| Estado 2 | 0,0011 | 0,989 | 0,0099 |
| Estado 3 | 0 | 0 | 1 |

Fuente: Elaboración Propia

Tabla 12.3: Nombres de los Estados Ocultos.

| ESTADO | NOMBRE ASIGNADO |
|----------|-----------------|
| Estado 1 | Satisfecho |
| Estado 2 | Semi Satisfecho |
| Estado 3 | Fugado |

Fuente: Elaboración Propia

Por lo tanto, los estados ocultos fueron definidos basándose en el nivel de satisfacción de los clientes, asumiendo que mientras más alta fuera la probabilidad de caer en el estado de fuga, menos satisfechos se sentían.

Al observar la Tabla 12.2, se puede afirmar que tanto el estado 1 como el 2 poseen una probabilidad similar de pasar al estado de fuga, siendo el segundo estado un poco más probable que el primero.

Además cabe mencionar el hecho de que los clientes tienen una alta probabilidad de encontrarse en el estado 2, ya que cuando se encuentran en el primero tienden a irse hacia el segundo y al encontrarse en éste la tendencia es a quedarse en él.

Si se observan las tablas 12.2 y 12.3 en conjunto, se puede ver que los clientes no suelen estar “satisfechos”. Esto podría significar que los clientes son muy sensibles a cualquier condición adversa que se presente y, por otro lado, que se tiene bastante competencia en el mercado, por lo que si el cliente se fuga de la empresa puede seguir consumiendo productos similares obtenidos en la competencia. Esto último también podría explicar la importancia que le dan los clientes a la atención de sus necesidades, ya que tienen muy disponible la opción de irse a la competencia.

Lo anterior hace pensar que los clientes podrían tener una buena disposición frente a las acciones de marketing que se puedan realizar sobre ellos, siempre y cuando éstas les sean útiles para sentirse más satisfechos, ya que son muy susceptibles a las condiciones que se les presentan.

⁸ Fuente: Manual de Marketing Directo
www.marketinet.com

Para los otros tipos de uso de la tarjeta se tienen comportamientos similares al observado en el Tipo 1. Las matrices de transición de los estados ocultos para estos tipos de uso pueden observarse en el Anexo 12.

- Distribución de Probabilidad de los Estados Observados

La probabilidad de observar compra para un cliente queda definida por:

$$L = \pi * m * q * m * q * m * q * m \dots$$

donde:

- π = probabilidad inicial.
- q = matriz de transición de los estados ocultos.
- m = matriz diagonal que indica la distribución de probabilidad condicional.

La probabilidad inicial se encuentra definida en el siguiente punto, mientras que la matriz de transición fue definida en el punto anterior.

Para el caso de “m” se tiene que es una matriz diagonal que indica la probabilidad que tiene cada cliente que no se encuentra en el estado de fuga de permanecer en el estado en que se encuentra actualmente. Esta matriz se tiene para cada cliente por cada período, por lo que no se mostrará aquí.

- Distribución del Estado Inicial

La probabilidad inicial tomada para todos los tipos de uso de la tarjeta es la siguiente:

$$\pi = [1 \ 0 \ 0]$$

Al tomar esta probabilidad inicial se asume que todos los clientes parten en el Estado 1.

Este supuesto puede realizarse ya que no se tiene información sobre el estado en el que parten los clientes y se considera que inicialmente todos los clientes son iguales.

- Caracterización de la Cadena de Markov Oculta y Probabilidad de Encontrarse en los Estados Ocultos

Basándose en los resultados obtenidos en los pasos anteriores, las Cadenas de Markov Ocultas a usar para los diferentes tipos de uso de la tarjeta quedan caracterizadas por los datos que se encuentran en la Tabla 12.4.

Tabla 12.4: Resumen Cadenas de Markov Ocultas

| | |
|------------------------------|---|
| Estados Ocultos | Satisfecho Semi-Satisfecho Fugado |
| Estados Observados | Compra No-Compra |
| Matriz de Transición | q |
| Distribución de probabilidad | L |
| Probabilidad Inicial | π |

Fuente: Elaboración Propia

Para cada tipo de uso de la tarjeta se genera una Cadena de Markov Oculta. En este caso coincidió que las 4 tuvieran la misma cantidad de estados ocultos, pero eso no es un requisito. Por lo tanto, con los datos de cada uno de los usos de la tarjeta, se estimó para cada uno de los clientes la respectiva probabilidad de encontrarse en los estados ocultos.

El cálculo de esta probabilidad se realiza de la siguiente manera:

$$P(s) = \pi * m_1 * q_2 * m_2 * q_3 * m_3 \dots * q_{t-1} * m_{t-1} * q_{ts} * m_s / L$$

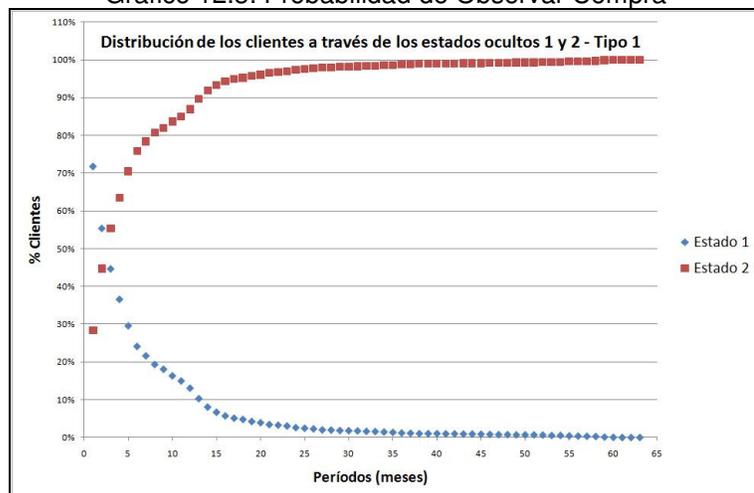
donde:

- $P(s)$ = probabilidad de encontrarse en el estado oculto s .
- π = probabilidad inicial.
- q_t = matriz de transición de los estados ocultos del período t para todos los estados ocultos.
- m_t = matriz diagonal que indica la distribución de probabilidad condicional del período t .
- q_{ts} = columna de la matriz de transición de los estados ocultos del período t para el estado oculto s .
- m_{ts} = distribución de probabilidad condicional del período t para el estado oculto s .
- L = distribución de probabilidad de los estados observados.

Luego, los clientes serán clasificados en estado 1 ó 2, dependiendo de en cuál de ellos es más probable que se encuentren.

En el Gráfico 12.3 se tiene esta clasificación a través de los períodos comprendidos en este estudio para el uso de la tarjeta en las tiendas de Tipo 1. En él se puede observar que la tendencia de los clientes es a encontrarse en el estado 2, el cual fue denominado como “Semi Satisfechos”.

Gráfico 12.3: Probabilidad de Observar Compra



Fuente: Elaboración Propia

Lo anterior hace sentido debido a que, como se observa en la matriz de probabilidades de transición en la Tabla 12.2, cuando los clientes se encuentran en el estado 1 tienden a cambiarse al 2 y los que ya se encuentran en él poseen la tendencia de permanecer ahí, por lo que es de esperar que un gran porcentaje de clientes se encuentre en ese estado a través del tiempo.

En este modelo se tienen comportamientos similares a través de los diferentes tipos de uso que se le dé a la tarjeta, en el sentido de que se observa una tendencia a que la mayoría de los clientes se encuentre en un sólo estado. Solamente para el caso del uso de la tarjeta en las tiendas del tipo 3, se tiene que los clientes no se mantienen en el mismo grupo, cambiando de un estado a otro cuando pasan de un período a otro.

Para el uso de la tarjeta del tipo 2, se tiene que la mayoría de los clientes tiende a estar en el estado 1. Al igual que en el caso del tipo 1, al mirar la matriz de transición se observa que los clientes que se encuentran en este estado tienen una alta probabilidad de permanecer en él, por lo que no es extraño que la mayoría de los clientes se encuentre en él. Sin embargo, llama la atención el hecho de que, en este caso, el estado 1 es denominado como “Satisfechos”, teniendo el estado 2 una probabilidad cercana al 80% de que los clientes se fuguen, lo cual muestra que, si bien los clientes suelen estar satisfechos, cuando no lo están por completo lo más probable es que terminen por fugarse si es que no se realizan acciones con el objetivo de que vuelvan a encontrarse en el estado 1.

Para el uso de de la tarjeta en las tiendas de tipo 3, como se mencionó anteriormente, se observa que, para ambos estados, cuando el porcentaje de clientes es muy alto en un período, será bajo al siguiente y si es bajo, será alto al siguiente. Este fenómeno se explica muy bien con la matriz de transición de los estados ocultos, ya que en ella se puede observar la tendencia de los clientes de pasar siempre de un estado a otro, teniendo una probabilidad muy baja de permanecer en el estado en que se encuentran. Esto significa que los clientes son muy propensos a pasar de “satisfechos” a “semi satisfechos” y en la inversa, por lo que son muy sensibles a las condiciones que encuentren cuando efectúen sus transacciones.

Finalmente, para el uso de la tarjeta en las tiendas de tipo 4, se tiene que el mayor porcentaje de clientes se encuentra en el estado 2, el cual fue denominado como “satisfechos”. Sin embargo, también se observa que al pasar el tiempo los porcentajes de clientes que se clasifican en el estado 1 y en el estado 2 se van acercando, por lo que se podría esperar que en un futuro sean muy cercanos. Esto se debe a que no existe una tendencia clara de los clientes a permanecer en un estado o a pasar al otro, ya que tienen probabilidades similares para ambos casos, lo cual se puede observar en la matriz de transición respectiva⁹. Además, se observa una probabilidad similar de pasar al estado de los fugados, siendo la del estado 1 un poco más alta que la del estado 2.

En el Anexo 13 se pueden observar los gráficos de las probabilidades de encontrarse en los estados ocultos para los tipos de uso de la tarjeta 2, 3 y 4.

Luego, en base a lo anterior, se definió que el conjunto de clientes al cual se le deberían aplicar acciones de marketing para evitar su fuga es el que pertenece al estado denominado “Semi Satisfecho”, ya que en ellos se observa la mayor probabilidad de pasar al estado de fugados, según lo visto en las matrices de transiciones respectivas.

En la Tabla 12.5 se encuentra un resumen del porcentaje de clientes que fue clasificado en el estado 1 y 2 para los diferentes tipos de uso de la tarjeta, para el mes de Marzo de 2008.

Tabla 12.5: % Clientes en cada estado

| TIPO DE USO | % Clientes en estado 1 | % Clientes en estado 2 |
|-------------|------------------------|------------------------|
| Tipo 1 | 0% | 100% |
| Tipo 2 | 99,5% | 0,5% |
| Tipo 3 | 93,5% | 6,5% |
| Tipo 4 | 21,3% | 78,7% |

Fuente: Elaboración Propia

Se desprende de la Tabla 12.5 que para los tipos de uso 1 y 4, la mayoría de los clientes quedan clasificados en el estado 1, mientras que en los tipos 2 y 3, la mayoría se clasifica en el estado 2.

Sin embargo, los estados 1 y 2 no son nombrados de igual forma para los diferentes tipos de uso de la tarjeta, siendo el estado 1 “Satisfechos” para los tipos 1 y 2 y “Semi Satisfechos” para los tipos 3 y 4. Luego, en base a esto, se puede decir que para el uso de la tarjeta en las tiendas de tipo 2 y 4, la mayoría de los clientes se encuentra en el estado “Satisfecho”, mientras que para los tipos 1 y 3 la mayoría se encuentra en “Semi Satisfechos”.

⁹ Las matrices de transición para los tipos de uso de la tarjeta en las tiendas del tipo 2, 3 y 4 se encuentran en el Anexo 12.

12.1. Error Cadenas de Markov Ocultas

En este caso, se tiene la probabilidad de pasar al estado “Fugados” para los estados “Satisfechos” y “Semi Satisfechos” para cada tipo de uso de la tarjeta, lo cual se comparará con el porcentaje real de fugados por período.

Luego, en la Tabla 12.7 se pueden observar los valores del Error Absoluto (MAE) y del Error Cuadrático (MSE) para el modelo de Cadenas de Markov Ocultas, para los diferentes tipos de uso de la tarjeta, considerando la probabilidad de pasar al estado de “Fugados” desde el estado “Satisfechos” y desde el “Semi Satisfechos”.

Tabla 12.7: Errores asociados a las Cadenas de Markov Ocultas

| TIPO DE USO | MAE | | MSE | |
|-------------|-------------|------------------|-------------|------------------|
| | Satisfechos | Semi Satisfechos | Satisfechos | Semi Satisfechos |
| Tipo 1 | 22,25% | 20,8% | 6,56% | 6,26% |
| Tipo 2 | 58,2% | 23,9% | 37,3% | 8,68% |
| Tipo 3 | 40,5% | 46,34% | 18,8% | 24,6% |
| Tipo 4 | 15,16% | 15,38% | 3,66% | 3,35% |

Fuente: Elaboración Propia

Al observar la Tabla 12.7, se tiene, para el caso de los tipos de uso 1, 3 y 4, que los errores encontrados para cada tipo de uso son similares cuando se comparan entre los estados “Satisfechos” y “Semi Satisfechos”, lo cual no ocurre en el uso de tipo 2, donde el estado “Semi Satisfechos” posee notoriamente menor error al estado “Satisfechos”.

Además cabe mencionar el hecho de que el uso de tipo 4 de la tarjeta es el que mejor se ajusta a las Cadenas de Markov Ocultas, ya que es el que posee el menor error asociado.

12.2. Comparación Cadenas de Markov Ocultas con Pareto/NBD

Tomando como base los resultados obtenidos con las Cadenas de Markov Ocultas, se realizó una comparación de éstos con los arrojados por el modelo Pareto/NBD, como una forma de poder visualizar con cuál de ellos se obtienen resultados más cercanos a lo esperado y por lo tanto, facilitar la elección del mejor modelo para este problema.

Se observó el comportamiento para las Cadenas de Markov Ocultas con los mismos 2 clientes tomados para el modelo Pareto/NBD, de los gráficos 11.1 y 11.2. Para el cliente 1, el cual se encontraba fugado en Noviembre 2008, se tiene que los primeros períodos se encontraba en el estado 1, para luego de 6 períodos pasar al estado 2 y permanecer ahí el resto de los períodos. Por otro lado, el cliente 2, que no se encontraba fugado en Noviembre 2008, se encuentra siempre clasificado en el estado 2. No es extraño que ambos clientes se encuentren la mayor parte del tiempo en el segundo estado, ya que, como se observó en el gráfico 12.3, la mayoría de los clientes se encuentran en ese estado a través del tiempo.

Además, similar a lo realizado en el caso de Pareto/NBD, se midió que porcentaje de los clientes que pertenecen al grupo al cual hay que aplicarle acciones de marketing, esto es, a los que se encuentran en el estado “Semi Satisfechos”, y que vuelven a realizar transacciones dentro de los siguientes 12 meses.

En la Tabla 12.6 se observa el resumen de los porcentajes de clientes que pertenecen a cada uno de los estados en Marzo 2008 y que vuelven a realizar compras los siguientes 12 meses.

Tabla 12.6: Porcentaje de clientes que vuelve a realizar compras.

| TIPO DE USO | % QUE VUELVE ESTANDO EN EL ESTADO “Satisfechos” | % QUE VUELVE ESTANDO EN EL ESTADO “Semi Satisfechos” |
|-------------|---|--|
| Tipo 1 | ¹⁰ | 58,3% |
| Tipo 2 | 47,9% | 100% |
| Tipo 3 | 77,8% | 51,2% |
| Tipo 4 | 53,1% | 45,5% |

Fuente: Elaboración Propia

Como se puede observar en la Tabla 12.6, los porcentajes de clientes que realizan compras los siguientes 12 meses para el estado “Semi Satisfechos” son bastante altos, considerando que este valor puede ser usado como una medida de error. Sin embargo, lo que más llama la atención es el caso del uso de la tarjeta en las tiendas de tipo 2, donde el porcentaje de clientes que vuelve a realizar compras en los siguientes 12 meses cuando se encuentra en el estado “Semi Satisfechos” es mayor que este porcentaje para los clientes que se encuentran en el estado “Satisfechos”. Esto puede deberse al pequeño porcentaje que pertenece al estado “Semi Satisfechos” para este tipo de uso, el cual comprende el 0,5% de los clientes.

Finalmente, se puede decir que los porcentajes de clientes que, perteneciendo al estado de los “Semi Satisfechos”, realizan compras los siguientes 12 meses para los usos de la tarjeta en las tiendas del tipo 1, 3 y 4 son similares, por lo que no se puede asegurar a priori que alguno de ellos se ajuste mejor a las Cadenas de Markov Ocultas que el otro.

Por otro lado, al comparar estos errores con los encontrados para el caso de Pareto/NBD (Tabla 11.7), se tiene que los de las Cadenas de Markov Ocultas son mayores que los obtenidos en ese caso.

¹⁰ No existen clientes en Marzo de 2008 que sean clasificados en el estado “Satisfechos” para el uso de la tarjeta de Tipo 1.

13. ANÁLISIS DE RESULTADOS

Al desarrollar ambos modelos se observaron diferentes cosas en cada uno de ellos.

Para el caso del modelo Pareto/NBD, se observa un comportamiento relativamente similar de los clientes a través de los diferentes usos de la tarjeta. Además, se tiene el hecho de que el comportamiento estimado tiene relación directa con lo que se esperaba encontrar, esto es, que a medida que el recency creciera, la probabilidad de que un cliente se encuentre activo vaya disminuyendo.

Por otro lado, para el caso de las Cadenas de Markov Ocultas, se tienen comportamientos específicos para cada uno de los tipos de uso de la tarjeta, pero con algunas características en común, como el hecho de que la mayoría de los clientes se encuentre en un solo estado. Como en este caso la probabilidad de pasar al estado de fuga era igual para todos los clientes dependiendo del tipo de uso de la tarjeta y de si se encontraban en el estado "Satisfechos" o "Semi Satisfechos", no se pudo comparar con el recency de cada cliente como se hizo con el modelo Pareto/NBD.

Se pudo observar también que los errores encontrados fueron más pequeños para el caso del modelo de Pareto/NBD que para las Cadenas de Markov Ocultas, tanto para los valores MAE y MSE, como para los criterios que se definieron de manera de comenzar a aplicar acciones de marketing sobre los clientes que lo cumplieran con el fin de evitar posibles fugas.

El hecho de que los errores de las Cadenas de Markov sean en general mayores que para el modelo Pareto/NBD, se podría explicar en parte por el hecho de que éste último se adaptó para cada criterio de fuga definido según el tipo de uso de la tarjeta, cosa que no se realizó con las Cadenas de Markov Ocultas debido a su complejidad.

Cabe mencionar además que el modelo de Pareto/NBD fue creado con el objetivo principal de estimar el Lifetime Value de los clientes en una empresa de retail, mientras que en general, las Cadenas de Markov Ocultas se usan para otras áreas, como el reconocimiento de palabras o de cadenas genéticas, por lo que ha sido poco probado en el ámbito del retail. Esto podría explicar por otro lado que los resultados obtenidos con este último no fueran mejores que los de Pareto/NBD.

En definitiva, por todo lo expuesto anteriormente, el modelo elegido para realizar la predicción de fuga de clientes y definir un punto donde empezar a ejercer acciones de marketing sobre ellos es el Pareto/NBD.

14. CARACTERIZACIÓN CLIENTES

Finalmente, como el modelo escogido fue el Pareto/NBD, se realizará la caracterización de los clientes para el mes de Marzo 2009 con este modelo. Esto puede ser considerado como la situación actual, ya que este es el último mes donde se poseen datos.

Lo primero es tener claro cuántos son los clientes que se encuentran fugados, ya que el modelo considera un grupo para este tipo de clientes.

En la Tabla 14.1 se muestran los porcentajes de fugados según el tipo de uso de la tarjeta, para el mes de Marzo 2009, que es el último mes del que se tienen datos.

Por otro lado, se tiene el grupo de los leales, que corresponde a los clientes que poseen más de 100 transacciones durante el período de estudio.

Tabla 14.1: Porcentaje de clientes fugados – Situación Actual

| TIPO DE USO | % FUGADOS |
|-------------|-----------|
| Tipo 1 | 41,05% |
| Tipo 2 | 49,9% |
| Tipo 3 | 67% |
| Tipo 4 | 56,1% |

Fuente: Elaboración Propia

En la Tabla 14.2 se presenta el porcentaje de clientes que se encuentra en el grupo de los leales para cada tipo de uso de la tarjeta.

Tabla 14.2: Porcentaje clientes leales – Situación Actual

| TIPO DE USO | % LEALES |
|-------------|----------|
| Tipo 1 | 7,3% |
| Tipo 2 | 1,8% |
| Tipo 3 | 2,9% |
| Tipo 4 | 10,2% |

Fuente: Elaboración Propia

Luego, se tienen el grupo de los normales y de los propensos, en donde se indica el porcentaje de clientes sobre los que habrá que aplicar acciones de retención para evitar que se fuguen.

En la Tabla 14.3 se observan los porcentajes de clientes que pertenecen a cada uno de los grupos mencionados en el párrafo anterior, para cada tipo de uso de la tarjeta.

Tabla 14.3: Porcentajes de clientes en el grupo “Normales” y “Propensos”

| TIPO DE USO | % NORMALES | % PROPENSOS |
|-------------|------------|-------------|
| Tipo 1 | 31,9% | 19,7% |
| Tipo 2 | 35,5% | 12,8% |
| Tipo 3 | 25,5% | 4,5% |
| Tipo 4 | 15% | 18,6% |

Fuente: Elaboración Propia

Luego, al observar la Tabla 14.3, se tiene que los clientes que pertenecen al grupo de “Propensos” son aquellos que tienen una mayor probabilidad de fugarse y que, por lo tanto, se hace necesario aplicar sobre ellos acciones de marketing con el objetivo de retenerlos. Se puede mencionar además, el hecho de que el mayor porcentaje de clientes que se encuentran propensos a fugarse es cuando se usa la tarjeta en las tiendas de Tipo 1, por lo tanto sería bueno poner especial atención en este uso de la tarjeta.

15. POSIBLES ACCIONES DE MARKETING

Para saber acerca de la clase de acciones que deben realizarse sobre los clientes para evitar su posible fuga, se hace necesario entender cuáles son los motivos por los que un cliente termina su relación con la empresa.

Esta labor puede resultar compleja, ya que, por lo general, los clientes toman una decisión de ese tipo basándose en impresiones, sentimientos o sensaciones que les son producidas durante el transcurso de su relación con la empresa.

Sin embargo, existen algunos estudios que tratan de explicar estos motivos que suelen ser intangibles en el momento de la compra.

En la Tabla 15.1 se presentan las razones por las cuales un cliente puede llegar a fugarse. En ella se puede observar que el principal motivo por el cual los clientes se fugan es porque sus necesidades no son atendidas, en otras palabras, porque no se sienten completamente satisfechos.

En realidad, el problema central que se tiene es el hecho de que los clientes que se encuentran insatisfechos no siempre hacen el respectivo reclamo, por lo que la empresa no se entera del motivo de la insatisfacción del cliente y, por lo tanto, no hace nada para remediarla.

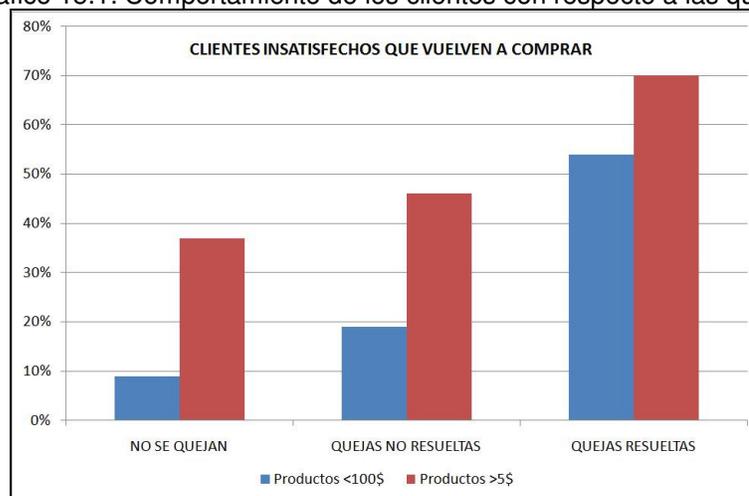
Tabla 15.1: Razones por las cuales los clientes se fugan

| RAZONES DE FUGA | PORCENTAJE DE CLIENTES |
|--------------------------------|------------------------|
| Indiferencia a sus necesidades | 68% |
| Descontentos crónicos | 10% |
| Precio | 9% |
| Recomendaciones de amigo | 5% |
| Cambio natural | 4% |
| Cambio de ubicación | 3% |
| Fallecimiento | 1% |

Fuente: Manual de Marketing Directo¹¹

En el Gráfico 15.1 se puede observar al comportamiento de los clientes son respecto a si realizan o no reclamos cuando se sienten insatisfechos y lo que ocurre cuando sus necesidades son atendidas.

Gráfico 15.1: Comportamiento de los clientes con respecto a las quejas



Fuente: Manual de Marketing Directo¹²

Como se puede observar en el Gráfico 15.1, cuando un cliente se encuentra insatisfecho, la probabilidad de que vuelva a comprar es mucho menor cuando no se queja que cuando lo hace. Aún más, cuando sus quejas son atendidas, su probabilidad de volver a comprar es más alta que cuando no los son.

Con todo lo anterior, queda clara la importancia de que exista una instancia en la cual el cliente puede generar una queja o reclamo, de manera que la empresa tenga conocimiento acerca de las necesidades insatisfechas de sus clientes y pueda realizar acciones para satisfacerlas.

¹¹ Fuente:

http://www.marketinet.com/ebooks/manual_de_marketing_directo_interactivo/manual_de_marketing_directo_interactivo.php?pg=22

¹² Fuente:

http://www.marketinet.com/ebooks/manual_de_marketing_directo_interactivo/manual_de_marketing_directo_interactivo.php?pg=22

Como una medida para mejorar el sistema en el cual los clientes realizaban sus quejas, una empresa multinacional instaló una línea 800 en unos de los países en donde se encuentra ubicada, de manera que los clientes pudieran realizar sus quejas a ese número y fuera de manera gratuita.

Como recomendación, se podría realizar una encuesta, la cual tendría 2 objetivos principales. Por un lado se quiere saber en qué cosas los clientes no se encuentran 100% satisfechos, de modo de poder mejorar esos puntos, y por otro lado, sería interesante conocer acerca del comportamiento de los clientes con respecto a las quejas, por ejemplo, saber qué tan frecuente es que se encuentren insatisfechos, cuantas de esas veces realizan un reclamo y los motivos por los cuales no dejan una queja cada vez que se encuentran insatisfechos. También se podría obtener una información acerca de cuál es el medio mediante el cual los clientes prefieren realizar sus quejas, de modo de poder motivar al cliente a indicar su necesidad no atendida, ya que, como se mencionó anteriormente, el porcentaje de clientes que se fugan disminuye cuando éstos expresan sus necesidades y aún más cuando sienten que esas necesidades fueron atendidas.

16. CONCLUSIONES Y RECOMENDACIONES

Una vez finalizado este trabajo se pueden obtener varias conclusiones respecto a él, las cuales serán detalladas a continuación.

16.1. Sobre los datos

Una vez que se realizó la limpieza de la Base de Datos, se tiene que ésta varió principalmente el número de clientes que usaba la tarjeta para cada uno de sus distintos usos, por lo que la existencia de clientes que no realizaban compras en algún tipo de uso de la tarjeta era la principal razón de que la Base de Datos original no pudiera ser ocupada directamente en los modelos.

Debido a lo anterior, se demuestra la importancia de realizar una buena limpieza en la Base de Datos, ya que, de los clientes que aparecían en un comienzo con valores perdidos, varios no realizaban compras en alguno de los tipos de uso de la tarjeta, por lo que se le hubiera considerado como fugado, cuando en realidad nunca ha utilizado la tarjeta en tiendas de ese tipo. Esto sólo hubiera producido ruido en los resultados.

Por lo tanto, el preprocesamiento de los datos evitó un mayor error en la estimación de los modelos, por lo que sus resultados son más representativos de la realidad.

Cabe mencionar el hecho de que para el uso de la tarjeta en tiendas del Tipo 3, se tiene una cantidad de clientes mucho menor que para los otros tipos de uso. Por esta razón,

si se pudiera incentivar a los clientes a usar la tarjeta en este tipo de tiendas, se podría contar con más información sobre su comportamiento, lo cual haría que los modelos se aproximaran más a la realidad.

16.2. Sobre los modelos

Si bien los modelos utilizados en este trabajo estimaban cosas diferentes, ambos se enfocaban en analizar el mismo punto: la fuga de los clientes.

El modelo Pareto/NBD estima la probabilidad de que un cliente permanezca activo, mientras que con las Cadenas de Markov Ocultas se estimaron las probabilidades de que los clientes se encuentren en los diferentes estados ocultos. Estas estimaciones miden cosas distintas, sin embargo, ambas se utilizaron para determinar un conjunto de clientes sobre los cuales es conveniente realizar acciones de marketing, ya que se encuentran en vías de fugarse. Por esta razón, es que los modelos pueden ser comparados según la clasificación dada a los clientes más que por las probabilidades estimadas.

Si bien en las matrices de transición de las Cadenas de Markov Ocultas se obtienen las probabilidades de pasar al estado “Fugados” desde los otros dos estados, esto no es una medida muy comparable con la probabilidad de permanecer activo estimada con el modelo Pareto/NBD, ya que para este último, las probabilidades se estiman de manera individual para cada cliente, mientras que en las Cadenas de Markov se da una probabilidad grupal. Además, en éstas últimas se clasifica a los clientes en el estado en que es más probable que se encuentren, pero no se tiene una completa certeza de que sea así, lo cual puede producir algunos errores. Esto no pasa con el modelo Pareto/NBD, ya que en éste no se tienen varios grupos donde clasificar a los clientes, ellos simplemente se fugan o no lo hacen, lo cual simplifica en algún grado los resultados y su interpretación.

Cabe mencionar que con el modelo Pareto/NBD desde un comienzo se observó un comportamiento que seguía los patrones de lo esperado, lo cual facilitó la definición del grupo de clientes sobre los cuales era conveniente aplicar acciones de marketing. Esto podría justificarse por el hecho de que este modelo se utiliza generalmente para problemas en el área del retail, por lo que, con este trabajo, se comprueba su efectividad en ese rubro. Debido a esto, se decidió incluir el criterio de fuga en este modelo, lo cual hizo que los resultados obtenidos fueran aún más cercanos a lo esperado.

Por su parte, en las Cadenas de Markov Ocultas, debido a la estructura que poseen, se definió que los clientes sobre los cuales se debían realizar acciones de marketing eran aquellos que pertenecían en estado “Semi Satisfechos”, ya que éste posee una mayor probabilidad de que sus integrantes pasen al estado “Fugados”. Sin embargo, esto puede conllevar algunos errores, ya que los clientes fueron clasificados según la probabilidad que poseen de encontrarse en ese estado, lo cual no es posible corroborar debido a que los estados son ocultos. Cabe mencionar además el hecho de que al

realizar esta clasificación se tiene que la mayoría de los clientes se encuentra agrupado en un solo estado, por lo que este modelo se considera poco discriminador.

Similar a lo realizado en el modelo Pareto/NBD, se hizo la prueba de modificar las Cadenas de Markov Ocultas incluyendo el criterio de fuga en ellas para cada tipo de uso de la tarjeta, sin obtener mejores resultados que con el modelo original. Por esta razón se decidió no modificarlo, ya que no agregaba valor al estudio y sólo hubiera complejizado más un modelo que por naturaleza ya lo era.

En definitiva, quedó demostrado que abordar un problema con un modelo más complejo no siempre va a entregar una mejor solución. Lo anterior se observó cuando se intentó incluir el criterio de fuga en las Cadenas de Markov Ocultas, así como en el hecho de que el Pareto/NBD genera mejores resultados en este caso, siendo un modelo más sencillo que las Cadenas de Markov Ocultas.

16.3. Sobre los errores

Al tener una relación no contractual entre el cliente y la empresa, no se conocen las probabilidades de fuga reales de los clientes. Por esta razón, es que se decidió tomar como referencia los porcentajes de clientes que se encontraban fugados basándose en el criterio de fuga definido anteriormente. Luego, este valor fue comparado con las probabilidades promedio estimadas con los modelos, para poder cuantificar los errores de cada uno de ellos.

De esta forma, al calcular el MAE y el MSE se encontró que el modelo Pareto/NBD obtenía errores más pequeños que las Cadenas de Markov Ocultas. Esto era esperable, ya que en el instante en que se definieron los grupos de clientes sobre los cuales era conveniente realizar acciones de marketing, se vio que el porcentaje de ellos que volvía a realizar compras con el modelo Pareto/NBD era menor que en el caso de las Cadenas de Markov Ocultas.

Luego, tomar como medida de error el porcentaje de clientes que vuelve a comprar de aquellos a los que se les debe realizar acciones de marketing, es tan válido en este caso como tomar los valores del MAE y el MSE, ya que con ambos criterios se obtienen mejores resultados con el modelo Pareto/NBD.

16.4. Sobre los resultados obtenidos

Como se ha mencionado anteriormente, los resultados en este caso se aproximaron más a la realidad con el modelo Pareto/NBD que con las Cadenas de Markov Ocultas. Sin embargo, esto no quiere decir que siempre ocurra de esta forma, de hecho, es de esperar que, por ejemplo, para la identificación de cadenas genéticas las Cadenas de Markov Ocultas arrojen menos errores que el modelo Pareto/NBD, ya que ésta es una de las áreas donde se tiene más experiencia en su uso.

Además, cabe mencionar que con ambos modelos se obtuvieron resultados que iban acorde con lo que se esperaba, por lo que, a pesar de tener un modelo que tiene errores mayores que el otro, con ambos se podría realizar una caracterización de los clientes, asumiendo que una se aproximará más a la realidad que la otra.

En ese sentido, cabe destacar el hecho de que el modelo Pareto/NBD discrimina mejor a los clientes que las Cadenas de Markov Ocultas para el caso de la predicción de fuga de clientes, lo que hace suponer que de utilizarse este modelo para estimar el número de compras a futuro o el Lifetime Value se deberían obtener también buenos resultados.

Es interesante destacar que, debido a las características del modelo Pareto/NBD, incluir el criterio de fuga en él resultó beneficioso, ya que hizo que los resultados mejoraran. Por lo tanto, de utilizarlo para realizar las estimaciones mencionadas en el párrafo anterior, se podría considerar la opción de incluir también en ellas el criterio de fuga.

Finalmente, se puede mencionar el hecho de que el modelo Pareto/NBD se usa generalmente para la industria del retail, por lo que lo obtenido en este estudio corrobora los resultados históricos del modelo. Esta puede ser una de las razones que explique por qué Pareto/NBD entrega mejores resultados que las Cadenas de Markov Ocultas, las cuales están recientemente incorporándose al trabajo en el área del retail.

16.5. Trabajos Futuros

Los trabajos a futuro que se desprenden de esta memoria se encuentran relacionados con el estudio del comportamiento de fuga del cliente y sus formas de evitarla.

Por un lado, se tiene el conocimiento de que la compañía cuenta con un estudio acerca del valor de los clientes, por lo tanto, al cruzar esa información con los resultados entregados en esta tesis, se podrían identificar a aquellos clientes que son más valiosos para la empresa y que se encuentran en vías de fugarse, por lo que es de gran importancia aplicar sobre ellos acciones de marketing con el objetivo de retenerlos. En base a esto, se deberían designar más recursos para evitar la fuga de estos clientes que la de aquellos que, a pesar de encontrarse prontos a fugarse, no le agregan un mayor valor a la empresa.

Por otro lado, como se sabe que una de las principales razones de la fuga de los clientes es que sienten que sus necesidades no son atendidas, es decir, se encuentran insatisfechos, un posible trabajo futuro podría ir enfocado a descubrir cuáles son las necesidades menos abordadas de los clientes y qué es lo que les falta para llegar a sentirse satisfechos. Lograr un consenso en ese sentido es una ardua tarea, debido a que la diversidad de los clientes indica que es muy posible que cada uno de ellos tenga una necesidad diferente. Por lo tanto, al realizar esta labor se podría analizar cuáles son las necesidades más recurrentes de los clientes o se podrían priorizar las necesidades de los clientes que son más valiosos para la empresa, dependiendo de los recursos que se encuentren disponibles para esto. El objetivo principal de conocer los motivos por los

cuales los clientes no se están sintiendo satisfechos es para poder generar acciones de marketing que vayan enfocadas a solucionar estos puntos.

Finalmente, y otorgándole más importancia a las variables utilizadas en este estudio, otra posible trabajo futuro sería incluir en el modelo de Pareto/NBD variables demográficas, ya que en el desarrollo de esta memoria sólo se utilizaron variables transaccionales. De esta manera se podría analizar si los resultados mejoran al incluir otro tipo de información en el modelo.

17. BIBLIOGRAFÍA

- Baek, Ha Nui (2006), "Estimación de Customer Lifetime Value a Nivel de Clientes Usando Variables Socio-Demográficas y Transaccionales", Memoria Ingeniería Civil Industrial, Universidad de Chile.
- Fader, Peter; Hardie, Bruce; Lee, K. (2005), "Counting Your Customers, the Easy Way: An Alternative to Pareto/NBD Model", *Marketing Science* v24 issue 2, p275-284.
- Fader, Peter; Hardie, Bruce; Lee, K. (2005), "Implementing the BG/NBD Model for Customer Base Analysis en Excel", [en línea]
<http://www.brucehardie.com/notes/004/bgnbd_spreadsheet_note.pdf>
- Fader, Peter; Hardie, Bruce; Lee, K. (2005), "Computing P(alive) Using the BG/NBD Model", [en línea]
<http://www.brucehardie.com/notes/021/palive_for_BGNBD.pdf>
- Falabella, "Presentación Encuentro Inversionistas Capital – Santander", [en línea]
<http://www.falabella.com/pdf/Release/2006/Presentacion_Santander_Capital_v3_20070430.pdf>
- Instituto Nacional de Estadísticas, [en línea]
<www.ine.cl>
- Marín, Pablo (2005), "Estimación de Lifetime Value Basada en Comportamiento Transaccional", Memoria Ingeniería Industrial; Universidad de Chile.
- Marketinet, "Manual de Marketing Directo", [en línea]
<http://www.marketinet.com/ebooks/manual_de_marketing_directo_interactivo/manual_de_marketing_directo_interactivo.php?pg=22>
- Neslin, Scott A.; Gupta, Sunil; Kamakura, Wagner; Lu, Junxiang; Mason, Charlotte H. (2006), "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churns Models", *Journal of Marketing Research*, v18, 204-211.
- Netzer, O.; Lattin, J.; Srinivasan, V. (2007), "A Hidden Markov Model of Customer Relationship Dynamics", *Marketing Science*, 27(2), 185-204.
- Pfeifer, P.; Carraway, R. (2000), "Modeling Customer Relationships as Markov Chains", *Journal of Interactive Marketing* v14, n°2: 43-55.

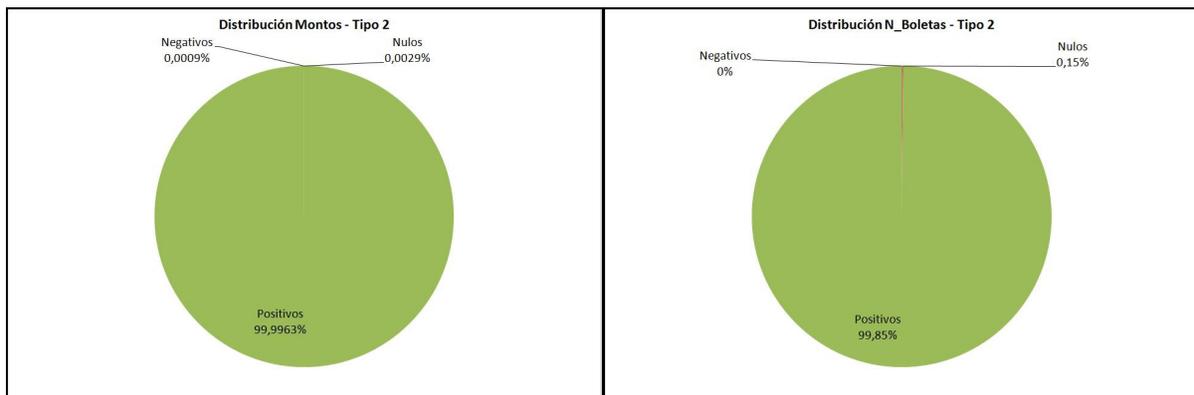
- *Rabiner, Lawrence (1989), "A Tutorial on Hidden Markov Model and Selected Applications in Speech Recognition", Proceedings of the IEEE, 77(2), 321-340.*
- *Segovia, Carolina (2005), "Caracterización del Proceso de Fuga de Clientes de un Retail Banking utilizando Información Transaccional", Memoria Ingeniería Civil Industrial, Universidad de Chile.*
- *Urzúa, Pedro (2007), "Estimación de Customer Lifetime Value Mediante Técnicas Supervisadas de Data Mining en una Empresa de Retail", Memoria Ingeniería Civil Industrial, Universidad*

18. ANEXOS

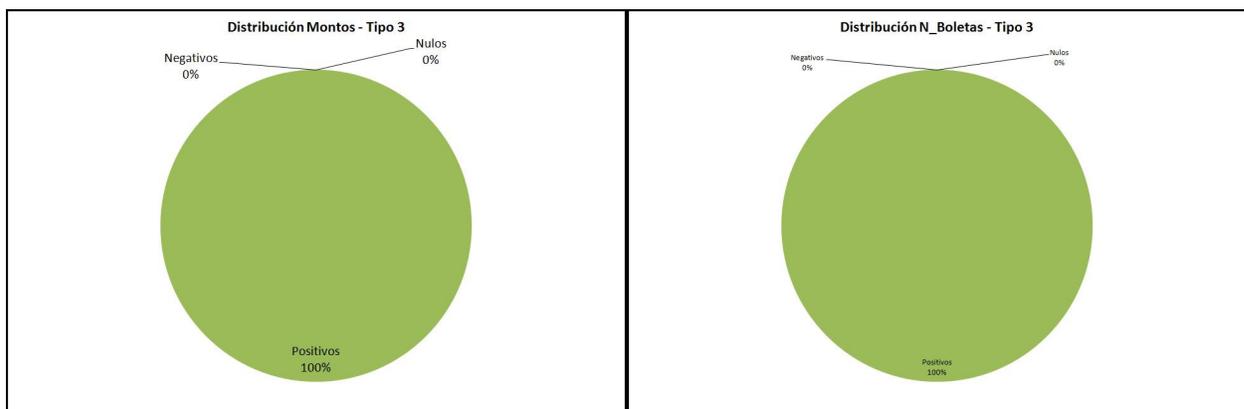
- **ANEXO 1: Distribución de las variables “Monto” y “N Boletas” en cuanto a si son positivos, negativos o nulos, para el uso de la tarjeta en tiendas del tipo 2, 3 y 4.**

A continuación se presenta la distribución de las variables Monto y N_Boletas con respecto a si son negativos, nulos o positivos. Esto se muestra para el Tipo 2, 3 y 4 de uso de la tarjeta,

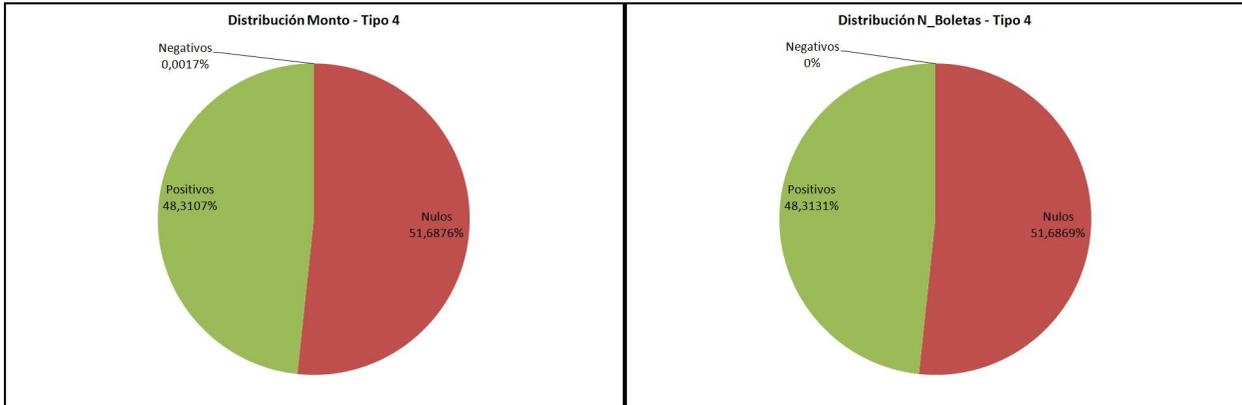
Gráficos Montos y N Boletas - Tipo 2



Gráficos Montos y N Boletas - Tipo 3



Gráficos Montos y N Boletas - Tipo 4



Como se puede observar en los gráficos anteriores, sólo para el caso en el que la tarjeta se usa para “No-Retail” se tiene un número significativo de valores nulos tanto en boletas como en monto. Sin embargo, se tiene que para las boletas, todos los valores nulos o negativos son erróneos, ya que no existe un significado claro para ellos.

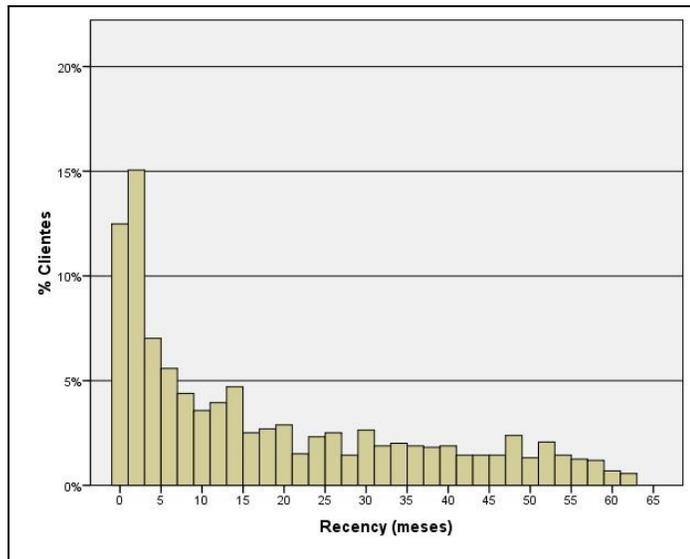
Viéndolo desde ese punto de vista y considerando la similitud de los porcentajes, existe una alta probabilidad de que los montos nulos correspondan a las boletas erróneas, por lo tanto podrían ser eliminadas sin perder información importante.

Por otro lado, se tiene que el uso de la tarjeta donde ocurre esta anomalía no es de los negocios principales de la empresa, por lo que no es muy preocupante esta situación.

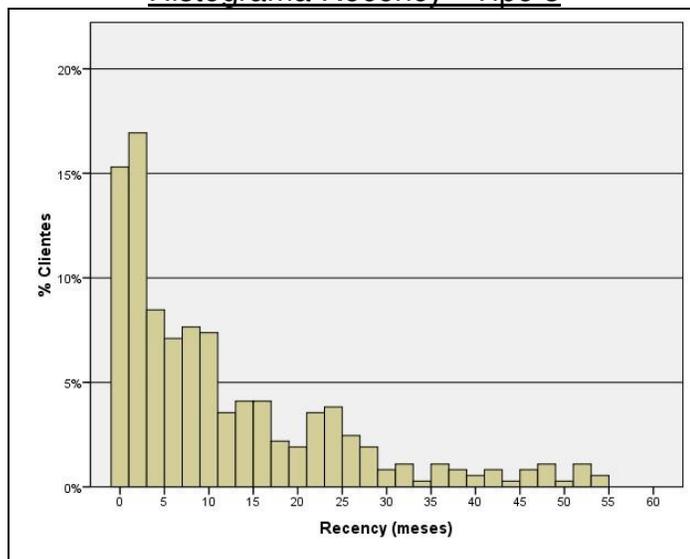
- **ANEXO 2: Distribución del Recency a través de los clientes para los usos de la tarjeta en las tiendas de tipo 2, 3 y 4.**

A continuación se presentan las distribuciones del recency para los usos de la tarjeta en las tiendas de los tipos 2, 3 y 4.

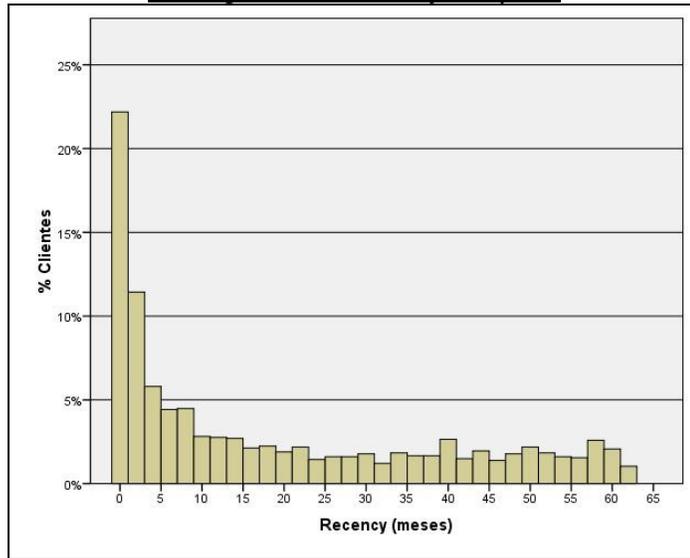
Histograma Recency - Tipo 2



Histograma Recency - Tipo 3



Histograma Recency - Tipo 4



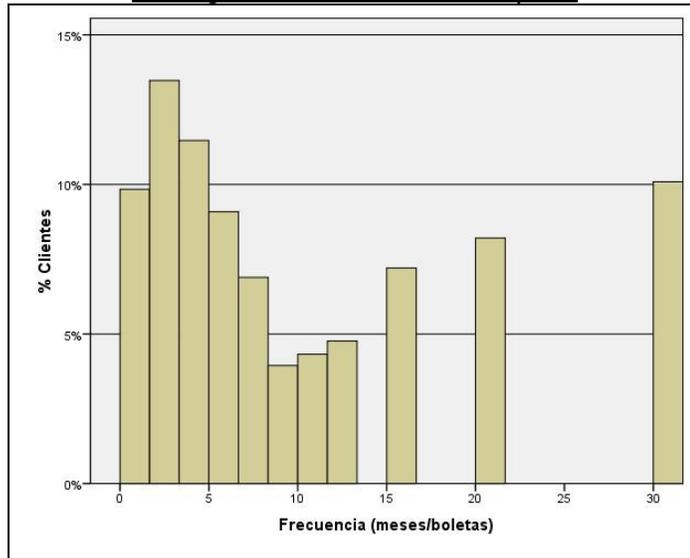
En los gráficos anteriores se ve claramente que la mayoría de los clientes posee un recency pequeño para todos los tipos de uso que posea la tarjeta.

De igual manera, se tiene que menos del 5% de los clientes por tipo de uso de la tarjeta tienen recencys muy grandes.

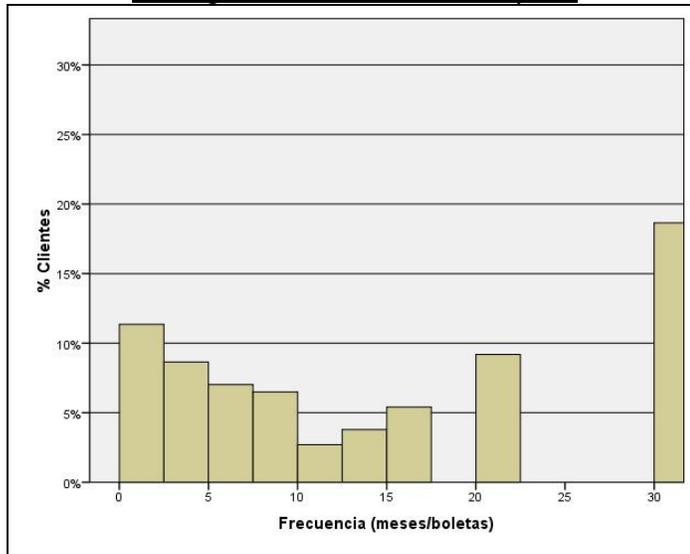
- **ANEXO 3: Distribución de la Frecuencia a través de los clientes para los usos de la tarjeta en las tiendas de los tipos 2, 3 y 4.**

A continuación se presentan las distribuciones de las frecuencias para los usos de las tarjetas en las tiendas de tipo 2, 3 y 4.

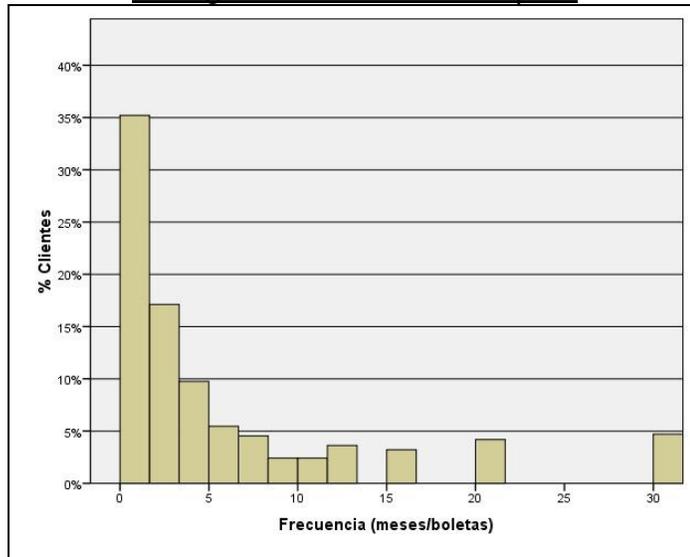
Histograma Frecuencia - Tipo 2



Histograma Frecuencia - Tipo 3



Histograma Frecuencia - Tipo 4



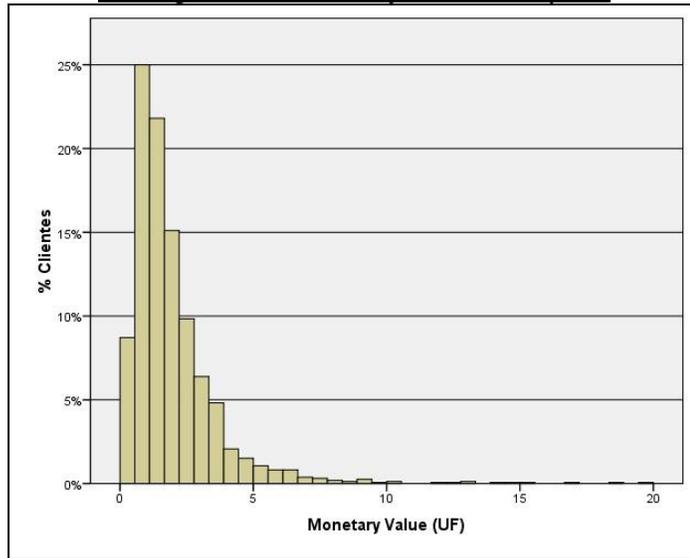
Al observar los histogramas anteriores, se tiene que la mayoría sigue la tendencia de tener un mayor porcentaje de clientes cuando la frecuencia es baja y contar con frecuencias más altas donde no se encuentran clientes.

Sin embargo, se tiene que para el Tipo 3, el porcentaje más alto de clientes se da cuando la frecuencia es muy alta.

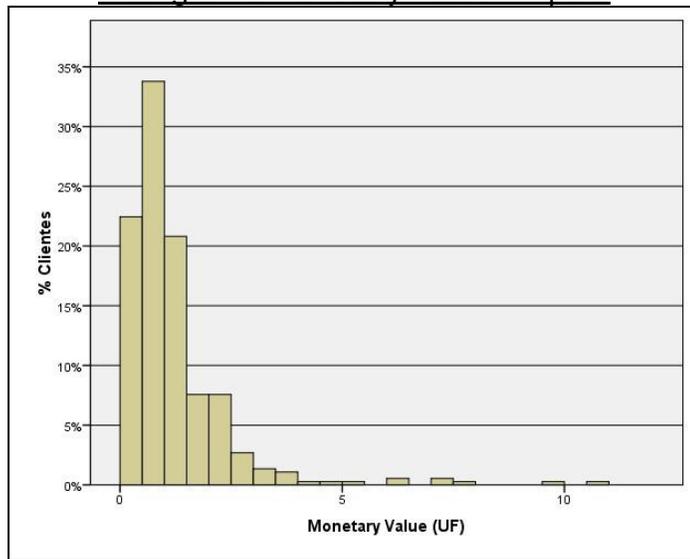
- **ANEXO 4: Distribución del monto promedio a través de los clientes para los usos de la tarjeta en las tiendas de tipo 2, 3 y 4.**

A continuación se presenta la distribución de los montos promedios para los usos de la tarjeta en las tiendas del tipo 2, 3 y 4.

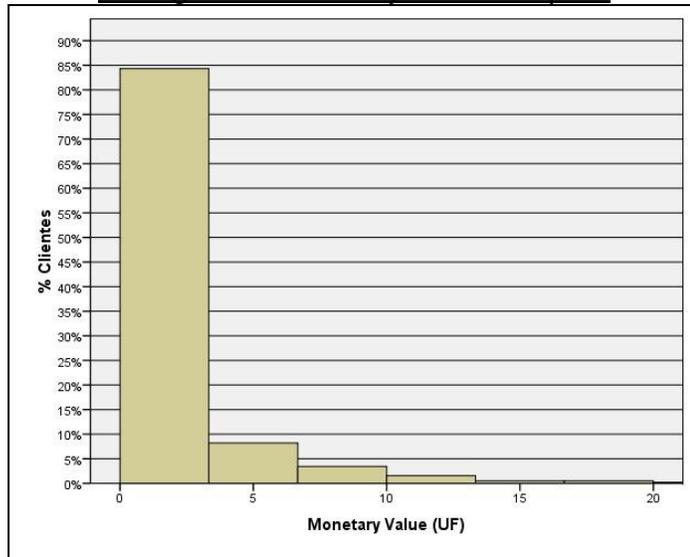
Histograma Monetary Value - Tipo 2



Histograma Monetary Value - Tipo 3



Histograma Monetary Value - Tipo 4



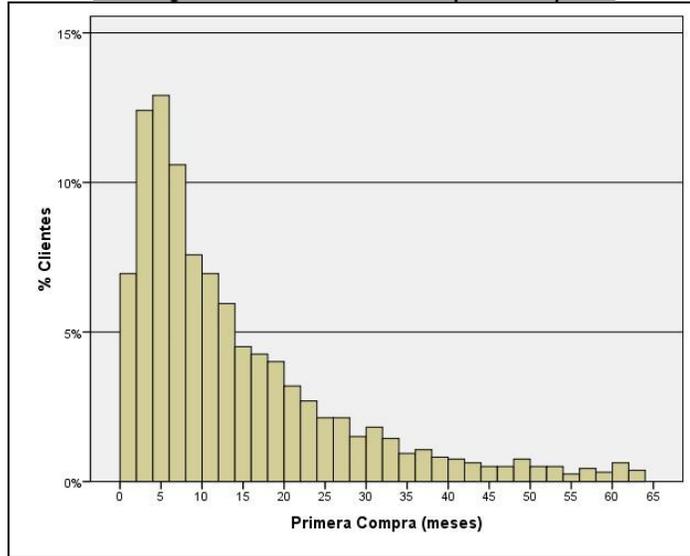
En esta caso, al observar los histogramas, se tiene el hecho de que la mayoría de los clientes poseen un monto promedio de compras relativamente pequeño, siendo muy pocos los clientes que poseen un monto promedio de compras mayor a 5 UF.

Lo anterior ocurre para todos los tipos de usos que se le puede dar a la tarjeta.

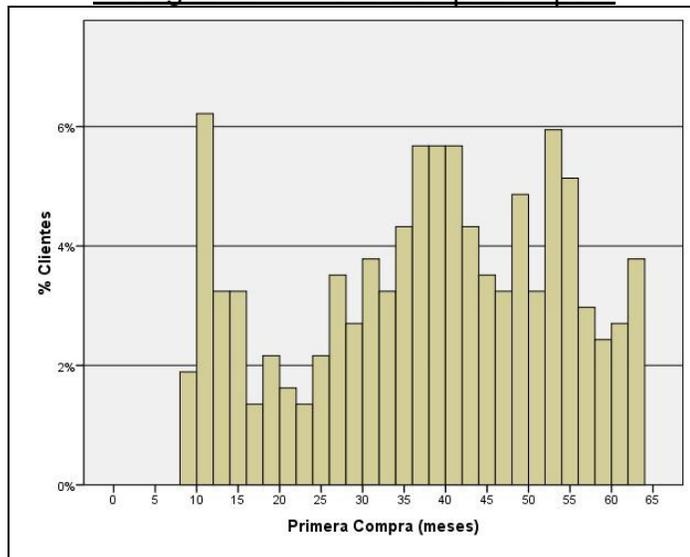
- **ANEXO 5: Distribución de la variable “Primera Compra” para el uso de la tarjeta en las tiendas de tipo 2, 3 y 4.**

A continuación se presentan los histogramas para la variable “Primera Compra” para el uso de la tarjeta en las tiendas de los tipos 2, 3 y 4.

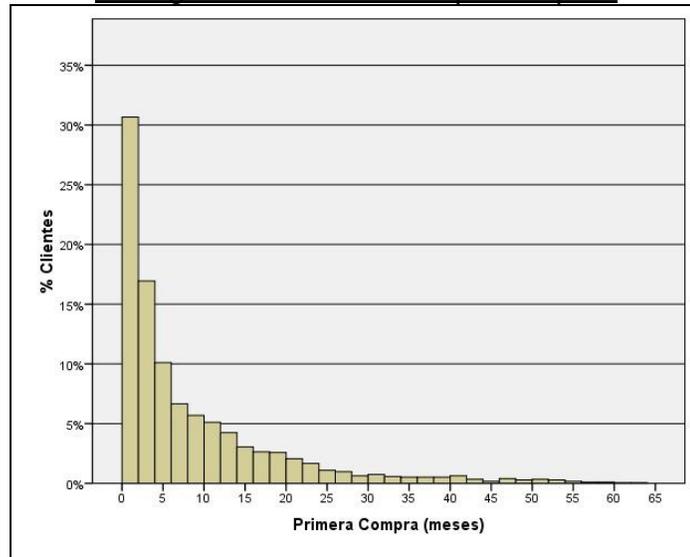
Histograma Primera Compra - Tipo 2



Histograma Primera Compra - Tipo 3



Histograma Primera Compra - Tipo 4



En este gráfico del Tipo 3 no aparece nadie con primera compra antes del período 9, ya que los datos para este tipo de uso de la tarjeta comienzan en Septiembre 2004.

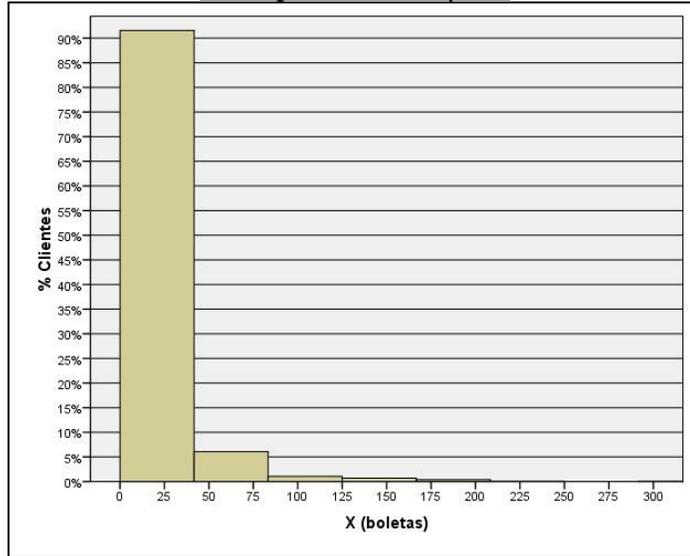
Para los otros dos tipos, el 2 y el 4, se tiene que el mayor porcentaje de clientes tiene primera compra al inicio del período de estudio, lo que hace suponer que son clientes que vienen comprando desde antes.

Por otro lado, para el Tipo 3 no se da esta tendencia, probablemente porque, como antes de Septiembre 2004 no se podían hacer compras con la tarjeta en este tipo de tienda, por lo que todos los clientes que poseen compras pueden ser considerados como nuevos para el uso de la tarjeta en supermercados.

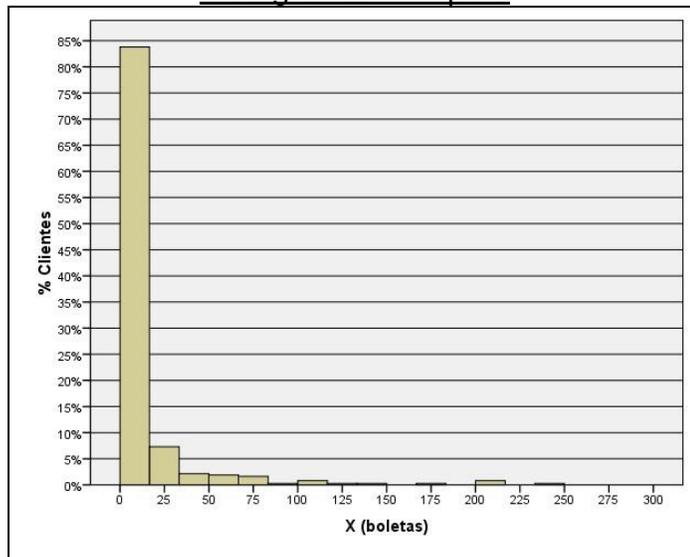
- **ANEXO 6: Distribución de la variable “X” para los usos de la tarjeta en las tiendas de los tipos 2, 3 y 4.**

Se presenta a continuación la distribución del número de boletas para los diferentes tipos de uso de la tarjeta.

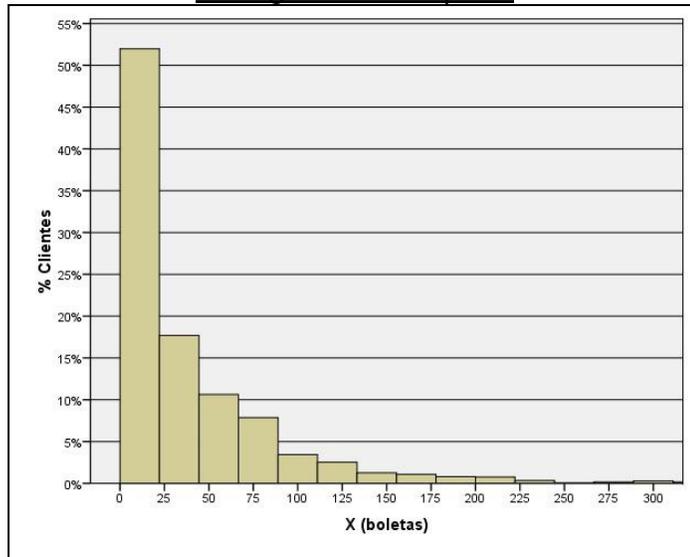
Histograma X - Tipo 2



Histograma X - Tipo 3



Histograma X - Tipo 4:



Al igual que en el caso del uso de la tarjeta en las tiendas del Tipo 1, el número de boletas para los otros tipos de uso es bajo para la mayoría de los clientes.

- **ANEXO 7: Criterios de Fuga para diferentes tipos de uso de la tarjeta.**

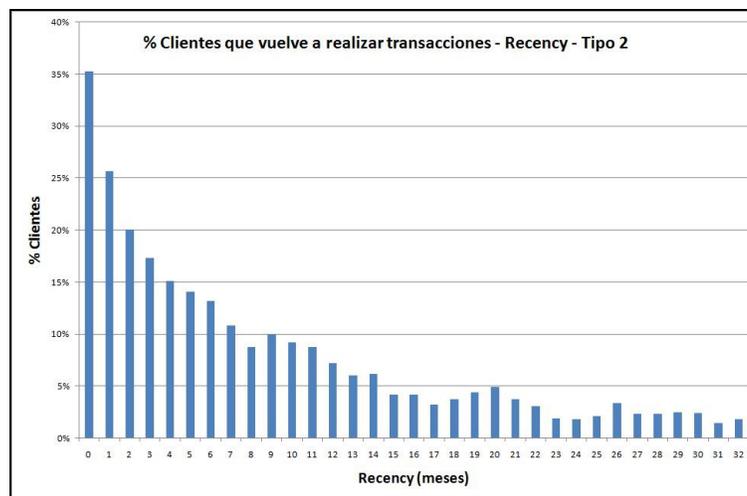
A continuación se presentan los criterios de fuga, separados por tipo de uso de la tarjeta.

Tipo 2

Para el caso del uso de la tarjeta de Tipo 2, se tiene el siguiente análisis acerca de los criterios de fuga definidos en la sección “Marco Teórico”.

- Caso 1:

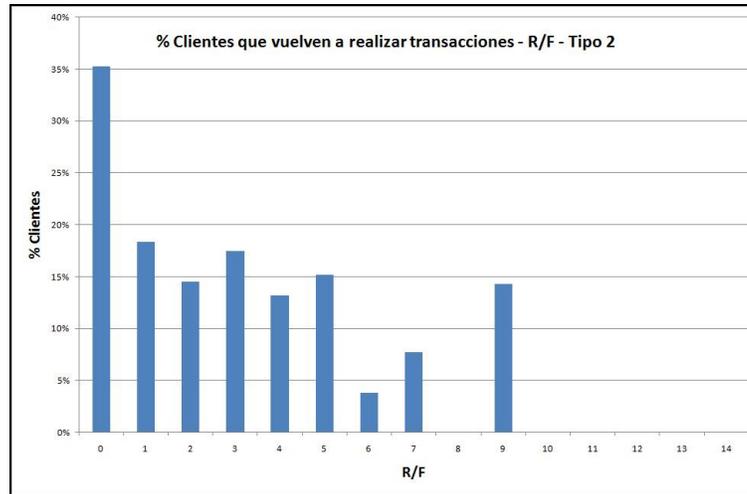
En el siguiente gráfico se ilustra el porcentaje de clientes que realiza transacciones en el período siguiente de cuando tuvieron un cierto valor para su Recency.



En él se observa, por ejemplo, que cuando los clientes poseen Recency=0, cerca del 35% realiza transacciones en el período siguiente. Por otro lado, después del Recency 15 sólo el 5% de los clientes vuelven a realizar transacciones. Sin embargo, al tomar estos últimos valores como punto de corte se estaría esperando más de un año para decidir que el cliente se encuentra fugado, lo cual no pareció satisfacer a la empresa, razón por la cual el punto de corte asignado es 12 meses.

- Caso 2:

A continuación se presenta un gráfico de las mismas características que el anterior, pero ahora para visualizar lo que ocurre tomando como criterio de fuga el caso 2 (R/F).



Aquí se observa claramente que cerca del 35% de clientes vuelve a realizar transacciones cuando la razón R/F es menor cero. Esto era de esperarse, ya que esto ocurre cuando el Recency es cero, por lo que se da la misma situación que con el otro criterio de fuga.

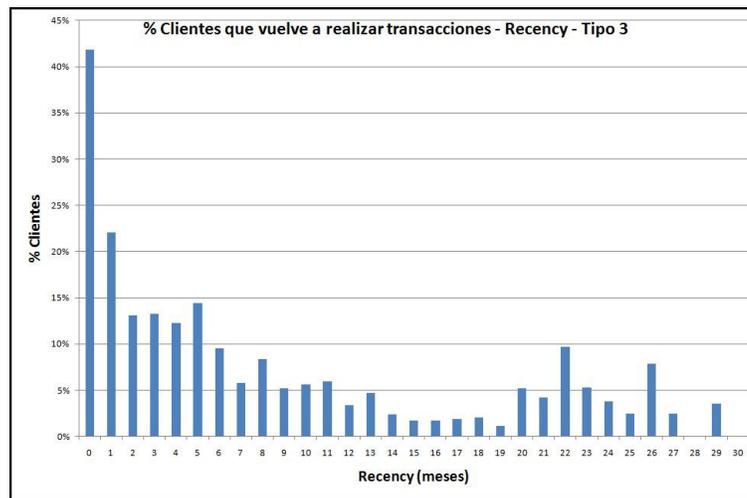
Cuando el Recency es 1 o más, entre un 18% a un 13% de los clientes vuelve a realizar compras, por lo menos hasta que esta razón toma el valor 5. Después de eso, los valores de R/F no parecen seguir un patrón claro.

Tipo 3:

Para el caso del uso del Tipo 3, se tiene el siguiente análisis acerca de los criterios de fuga definidos en la sección “Marco Teórico”.

- Caso 1:

En el siguiente gráfico se ilustra el porcentaje de clientes que realiza transacciones en el período siguiente de cuando tuvieron un cierto valor para su Recency.

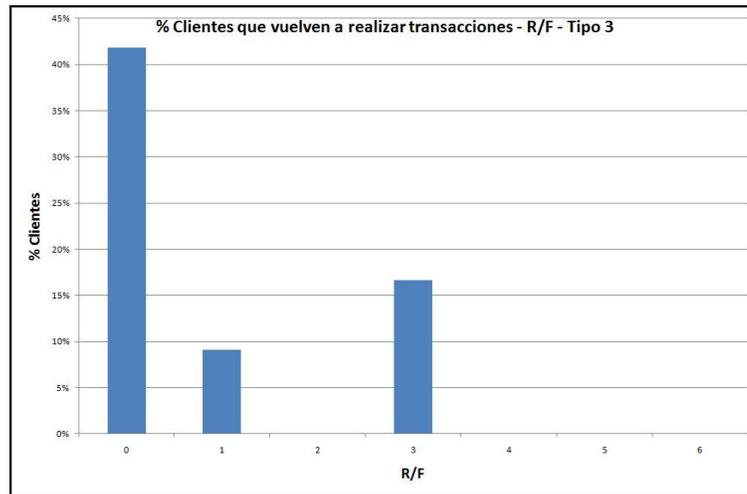


En este caso se observa algo un poco diferente al resto de los casos, ya que aun con un recency alto se sigue teniendo que cerca del 5% de los clientes vuelve a comprar. Además se tiene que cuando el recency es cero más del 40% de los clientes vuelve a comprar, mientras que cuando el recency es 1 se cerca del 22%, por lo que tiene una baja brusca.

Finalmente se optó por poner el criterio de fuga en recency=3, donde se tiene que cerca de un 13% de los clientes vuelve a comprar.

- Caso 2:

A continuación se presenta un gráfico de las mismas características que el anterior, pero ahora para visualizar lo que ocurre tomando como criterio de fuga el caso 2 (R/F).



Como se puede observar, se tiene que cerca del 40 % de las personas que tienen R/F cero vuelven a comprar, al igual que con el criterio de fuga anterior. Luego se tiene que sólo cuando $R/F=1$ y $R/F=3$ existe un porcentaje de clientes que vuelve.

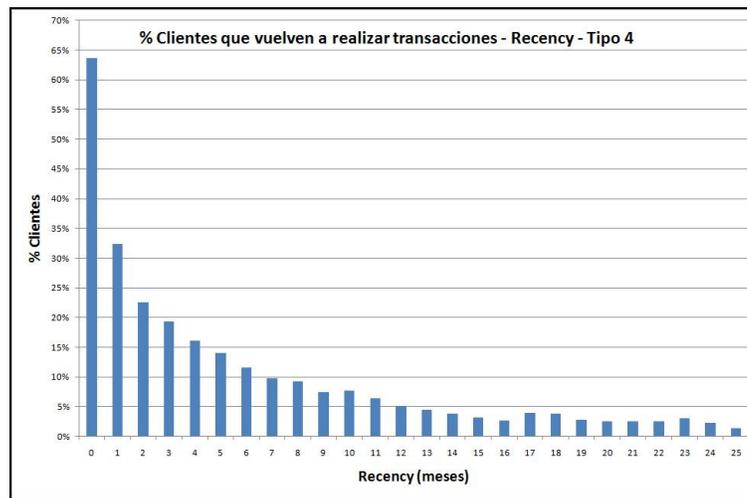
De todas maneras, no es mucho lo que se puede desprender de este gráfico, debido a que cuenta con muy poca información. Esto viene principalmente del hecho de que este tipo de uso era el menos preferido por los clientes para la tarjeta, por lo que la cantidad de clientes que en realidad tenía transacciones en tiendas de este tipo era muy pequeña en comparación con los otros tipos de uso que tiene la tarjeta.

Tipo 4:

Para el uso de la tarjeta del Tipo 4, se tiene el siguiente análisis acerca de los criterios de fuga mencionados anteriormente:

- Caso 1:

En el siguiente gráfico se ilustra el porcentaje de clientes que realiza transacciones en el período siguiente de cuando tuvieron un cierto valor para su Recency.

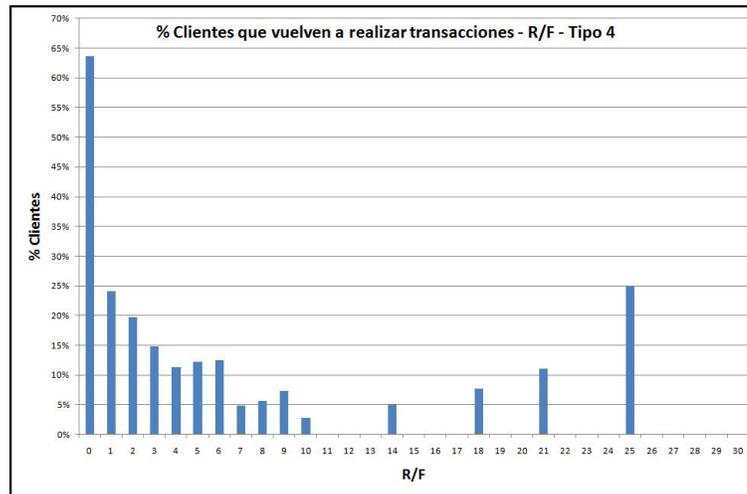


En él se observa, por ejemplo, que cuando los clientes poseen Recency=0, casi el 65% de los clientes realiza transacciones en el período siguiente, por lo tanto, se puede afirmar que ese no sería un corte adecuado para el criterio de fuga, debido a que este porcentaje es muy alto. Por otro lado, cuando el recency es 12 o más, el porcentaje de clientes que vuelven a realizar transacciones es cerca del 5%, lo cual pareciera ser un valor más razonable para que el cliente sea considerado como fugado.

En este caso, el criterio de fuga seleccionado es R=7, ya que se buscaba que el porcentaje de personas que volviera a realizar transacciones fuera cerca del 10%.

- Caso 2:

A continuación se presenta un gráfico de las mismas características que el anterior, pero ahora para visualizar lo que ocurre tomando como criterio de fuga el caso 2 (R/F).



Al igual que en los casos anteriores, se tiene que cuando R/F=0, casi el 65% de los clientes vuelve a realizar compras, lo cual también ocurre con el otro criterio de fuga, ya que esto pasa cuando el Recency es cero.

Después de eso, el porcentaje cae bruscamente a 25% y sigue disminuyendo paulatinamente. Pasado R/F=10, ya no se observa un patrón claro.

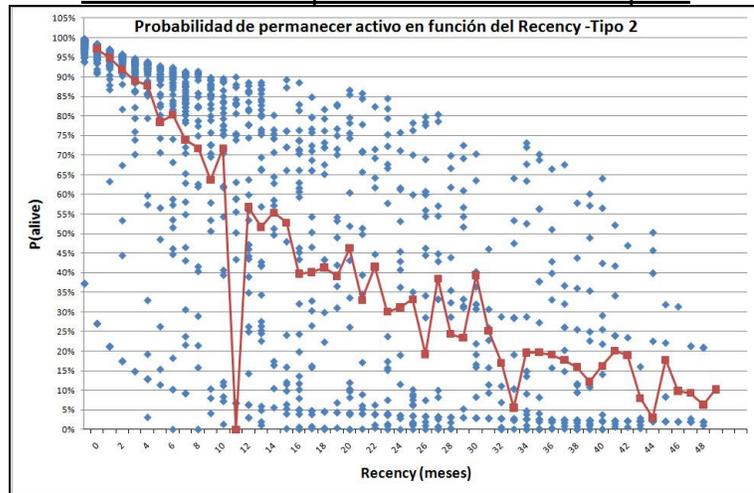
- **ANEXO 8: Parámetros a usar por el modelo Pareto/NBD**

| | r1 | r2 | r3 | α | a | b |
|--------|-----------|-----------|-----------|----------------|----------------|----------------|
| Tipo 1 | -0,69931 | 0,029162 | 0,000142 | exp(0,915733) | exp(-1,682186) | exp(1,663019) |
| Tipo 2 | -0,400608 | -0,00263 | -0,000887 | exp(-0,423561) | exp(-1,144674) | exp(1,578454) |
| Tipo 3 | 3,379176 | -0,085332 | -0,001982 | exp(0,393881) | exp(-2,434129) | exp(-0,952494) |
| Tipo 4 | -0,73486 | 0,037352 | 0,000316 | exp(1,0780612) | exp(-1,392173) | exp(0,093825) |

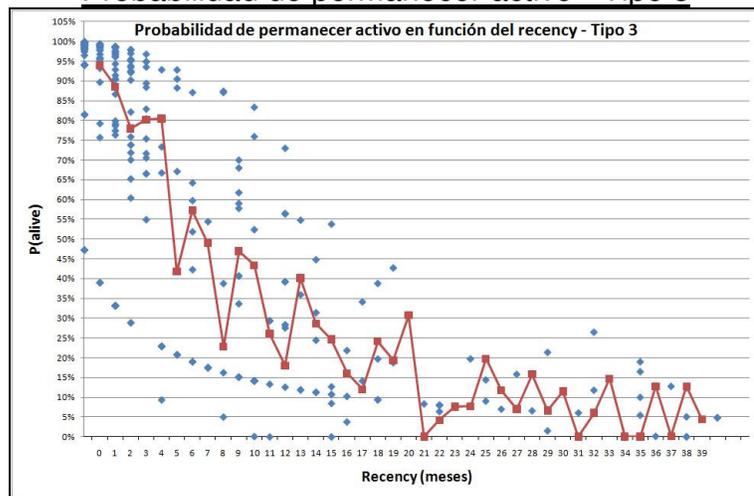
- **ANEXO 9: Probabilidad de permanecer activo para el uso de la tarjeta en las tiendas de los tipos 2, 3 y 4.**

A continuación se presentan los gráficos de las probabilidades de permanecer activos estimadas con el modelo Pareto/NBD. Se muestran los gráficos para los usos en las tiendas de tipo 2, 3 y 4.

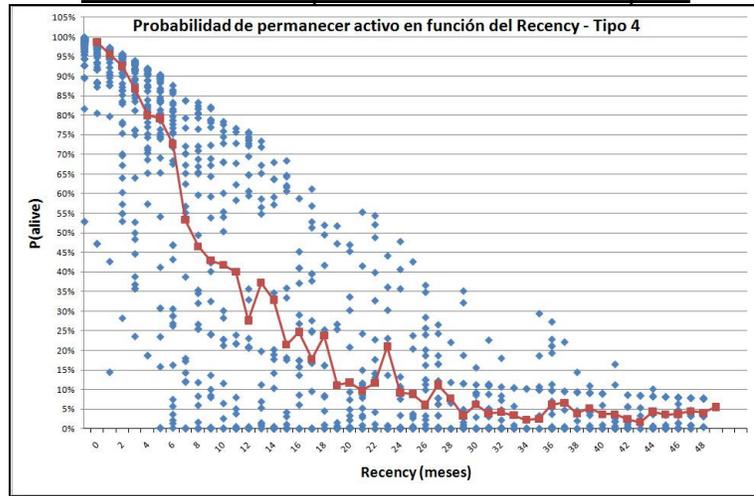
Probabilidad de permanecer activo - Tipo 2



Probabilidad de permanecer activo - Tipo 3



Probabilidad de permanecer activo - Tipo 4



Al observar las probabilidades de permanecer activo estimadas por el modelo Pareto/NBD, se tiene una gran dispersión de éstos, pero tendiendo a disminuir a medida que el recency crece. Esto queda mucho más claro al observar la línea que grafica la probabilidad promedio por recency, la cual, a pesar de no ser lineal, tiene una clara tendencia a la baja.

- **ANEXO 10: Comparación porcentaje de clientes que compran dentro de los próximos 12 meses en base a la probabilidad de permanecer activos, calculada con el modelo Pareto/NBD.**

Se presentan a continuación las probabilidades de que los clientes permanezcan activos comparadas con el porcentaje de clientes que posee compra los siguientes 12 meses.

P(alive) vs % Clientes - Tipo 2

| P(alive) | % Clientes que realiza compra los siguientes 12 meses |
|----------|---|
| 0-0,1 | 13,02% |
| 0-0,2 | 15,22% |
| 0-0,3 | 16,84% |
| 0-0,4 | 17,98% |
| 0-0,5 | 18,62% |
| 0-0,6 | 20,04% |
| 0-0,7 | 20,44% |
| 0-0,8 | 23,02% |
| 0-0,9 | 26,22% |

P(alive) vs % Clientes - Tipo 3

| P(alive) | % Clientes que realiza compra los siguientes 12 meses |
|----------|---|
| 0-0,1 | 22,45% |
| 0-0,2 | 27,78% |
| 0-0,3 | 28,41% |
| 0-0,4 | 30,93% |
| 0-0,5 | 31,48% |
| 0-0,6 | 31,90% |
| 0-0,7 | 31,40% |
| 0-0,8 | 32,14% |
| 0-0,9 | 34,71% |

P(alive) vs % Clientes - Tipo 4

| P(alive) | % Clientes que realiza compra los siguientes 12 meses |
|----------|---|
| 0-0,1 | 9,77% |
| 0-0,2 | 12,35% |
| 0-0,3 | 12,65% |
| 0-0,4 | 13,99% |
| 0-0,5 | 14,41% |
| 0-0,6 | 14,96% |
| 0-0,7 | 16,37% |
| 0-0,8 | 17,97% |
| 0-0,9 | 20,78% |

La idea de este análisis es definir un punto en el cual se debiera comenzar a atacar a los clientes porque es muy probable que se fuguen.

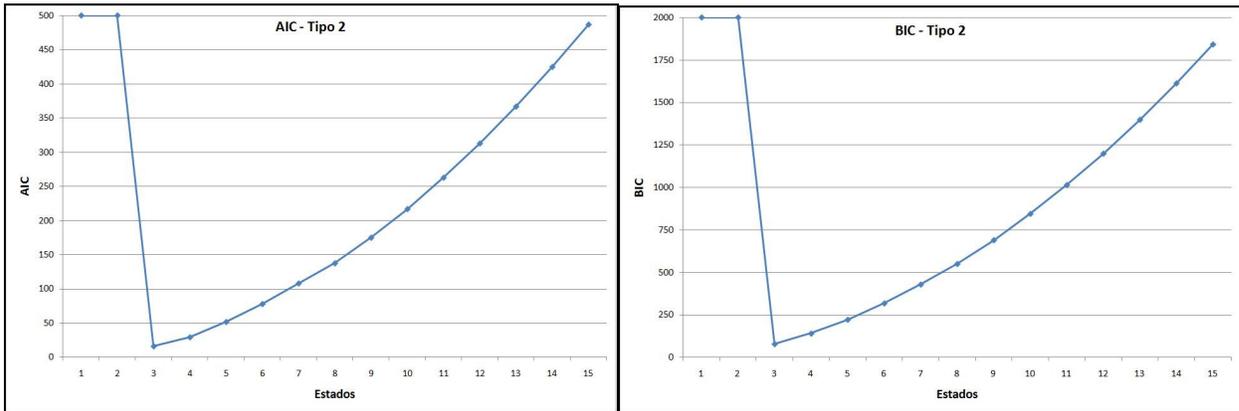
Se estimó que cuando el porcentaje de clientes que vuelven a comprar los siguientes 12 meses era cercano al 15% era un buen punto para comenzar a aplicar acciones de marketing sobre los clientes, ya que el 85% de ellos se fuga.

Por esta razón se tiene que los puntos definidos para los usos de la tarjeta en las tiendas de tipo 2, 3 y 4 son 0,2 – 0,1 – 0,6 respectivamente.

- **ANEXO 11: AIC y BIC para el uso de la tarjeta en las tiendas de tipo 2, 3 y 4.**

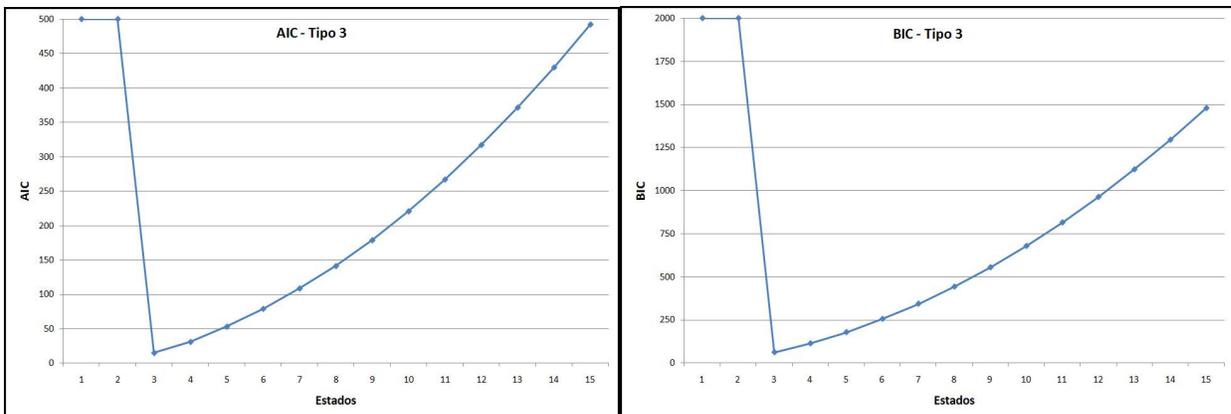
A continuación se presentan los gráficos de los valores del AIC y BIC para los usos de la tarjeta en las tiendas de tipo 2, 3 y 4.

AIC y BIC - Tipo 2:



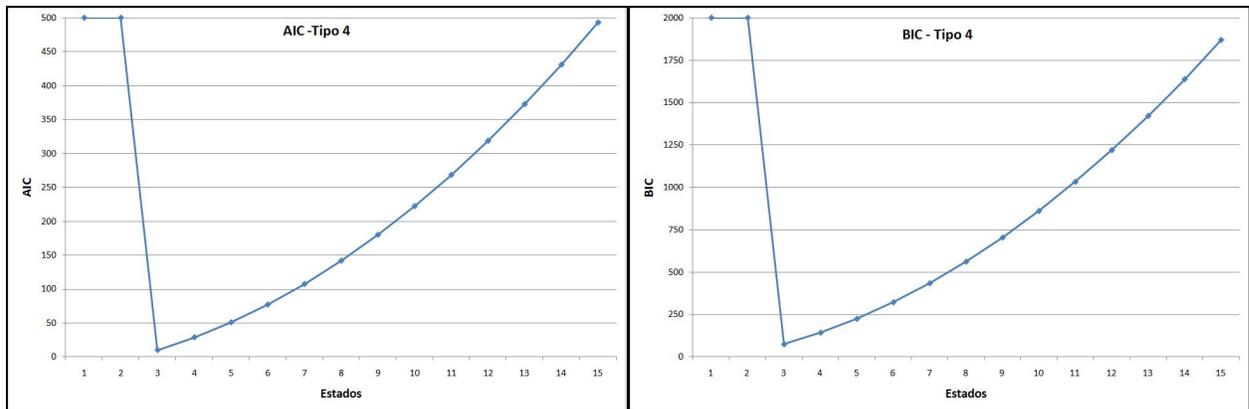
| ESTADOS | AIC | BIC |
|---------|-----------|-----------|
| 1 | inf | inf |
| 2 | inf | inf |
| 3 | 16,137937 | 80,641006 |
| 4 | 29,321806 | 142,20218 |
| 5 | 51,729248 | 223,73743 |
| 6 | 77,967207 | 319,85372 |
| 7 | 108,06004 | 430,57539 |
| 8 | 137,50128 | 551,39597 |

AIC y BIC - Tipo 3:



| ESTADOS | AIC | BIC |
|---------|-----------|-----------|
| 1 | inf | inf |
| 2 | inf | inf |
| 3 | 15,433834 | 62,395871 |
| 4 | 31,385351 | 113,56891 |
| 5 | 53,617711 | 178,84981 |
| 6 | 79,392497 | 255,50013 |
| 7 | 108,90767 | 343,71785 |
| 8 | 141,46387 | 442,8036 |

AIC y BIC - Tipo 4:



| ESTADOS | AIC | BIC |
|---------|-----------|-----------|
| 1 | inf | inf |
| 2 | inf | inf |
| 3 | 9,9274649 | 75,474044 |
| 4 | 28,648302 | 143,35482 |
| 5 | 50,589041 | 225,37992 |
| 6 | 76,94186 | 322,74153 |
| 7 | 107,30197 | 435,03487 |
| 8 | 141,61045 | 562,201 |

Se puede observar en los gráficos que tanto los valores del AIC como del BIC es mínimo cuando se tiene un modelo con 3 estados.

Además, cabe mencionar el hecho de que en todos los tipos de uso de la tarjeta, cuando se tienen 1 y 2 estados, tanto el AIC como el BIC tiende a infinito.

- **ANEXO 12: Matrices de Transición de los Estados Ocultos para el uso de la tarjeta en las tiendas de tipo 2, 3 y 4.**

Se presentan a continuación las matrices de transición de los estados ocultos para los usos de la tarjeta en las tiendas del tipo 2, 3 y 4.

Matriz de Transición - Tipo 2

| | Estado 1 | Estado 2 | Estado 3 |
|----------|-----------|-----------|-----------|
| Estado 1 | 0,9952 | 0,0002 | 0,0046 |
| Estado 2 | 0,1623657 | 0,0000004 | 0,8376339 |
| Estado 3 | 0 | 0 | 1 |

Para este caso se tiene que el estado 2 es el que tiene más probabilidades de pasar al estado de fuga, por lo tanto los nombres de los estados serán:

Estado 1: Satisfechos

Estado 2: Semi Satisfechos

Matriz de Transición - Tipo 3

| | Estado 1 | Estado 2 | Estado 3 |
|----------|----------|----------|----------|
| Estado 1 | 0,0127 | 0,8904 | 0,0969 |
| Estado 2 | 0,970801 | 0,000006 | 0,029193 |
| Estado 3 | 0 | 0 | 1 |

En este caso se tiene que la mayor probabilidad de pasar al estado de fuga se da en el estado 1, por lo que los nombres de los estados serán:

Estado 1: Semi Satisfecho

Estado 2: Satisfecho

Matriz de Transición - Tipo 4

| | Estado 1 | Estado 2 | Estado 3 |
|----------|----------|----------|----------|
| Estado 1 | 0,3338 | 0,2695 | 0,3967 |
| Estado 2 | 0,2357 | 0,4205 | 0,3438 |
| Estado 3 | 0 | 0 | 1 |

Para este último caso se tiene que el estado 1 tiene la mayor probabilidad de pasar al estado de fuga, por lo que los nombres de los estados serán:

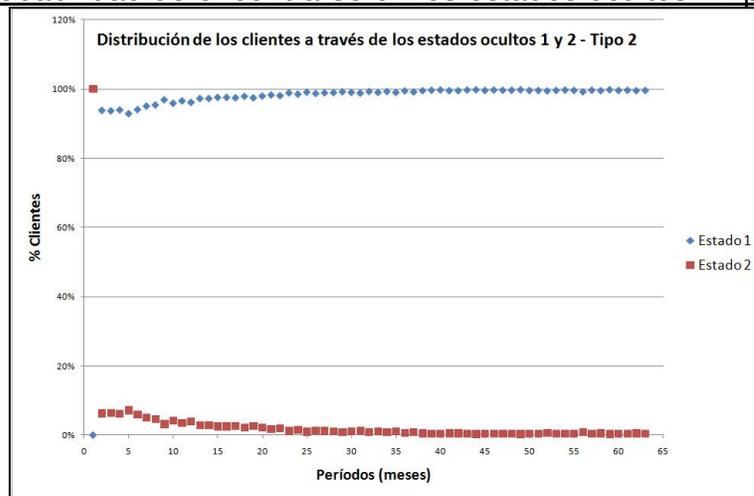
Estado 1: Semi Satisfecho

Estado 2: Satisfecho

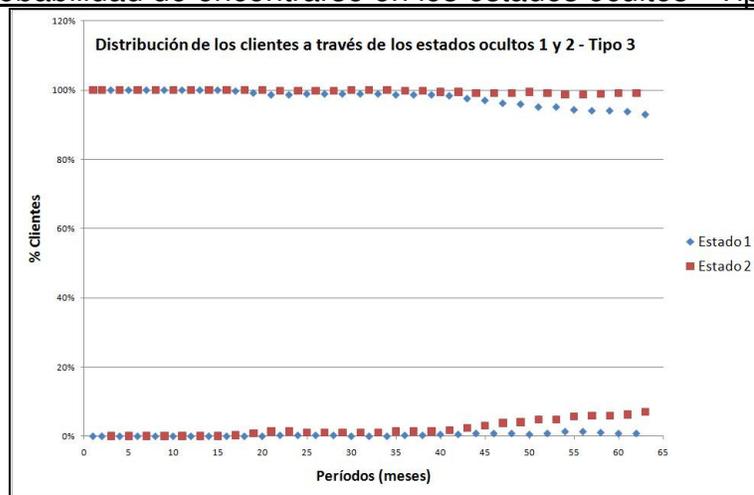
- **ANEXO 13: Gráficos de las Probabilidades de encontrarse en los Estados Ocultos para el uso de la tarjeta en las tiendas de tipo 2, 3 y 4. Cálculos realizados con Cadenas de Markov Ocultas.**

A continuación se presentan la distribución de los clientes a través de las probabilidades de encontrarse en los estados ocultos en los distintos períodos comprendidos dentro de este estudio, para los usos en las tiendas del tipo 2, 3 y 4.

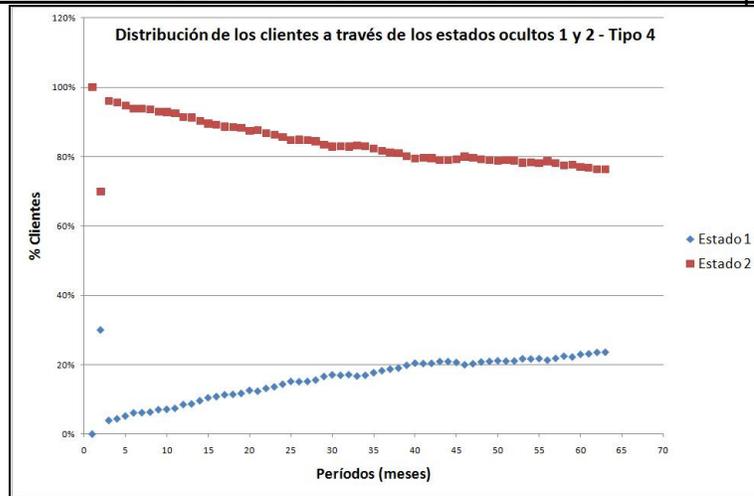
Probabilidad de encontrarse en los estados ocultos - Tipo 2



Probabilidad de encontrarse en los estados ocultos - Tipo 3



Probabilidad de encontrarse en los estados ocultos - Tipo 4



Para el Tipo 2, se tiene que se mantienen parejos a través de los períodos el porcentaje de clientes que fue clasificado en cada uno de los estados ocultos diferentes al “fugados”.

Para el Tipo 3, se tiene que los porcentajes de clientes que pertenecen a un estado o a otro se mantiene en a través de los períodos, pero se va intercambiando los estados, esto es, si en el período “t” el estado 1 tenía cerca del 100% y el estado 2 casi 0%, en el período “t+1” se tiene que el estado 2 tiene cerca del 100% de los clientes y el estado 1 un porcentaje cercano a cero.

Para el Tipo 4 se tiene que la probabilidad encontrarse en el estado 2 posee un porcentaje de clientes mayor que el estado 1. Sin embargo, al ir aumentado los períodos, estos porcentajes se van acercando, por lo que no sería extraño que más adelante ambos estados tengan porcentajes de clientes similares.