UNIVERSIDAD NACIONAL DE INGENIERÍA

FACULTAD DE INGENIERÍA ECONÓMICA Y CIENCIAS SOCIALES

ESCUELA PROFESIONAL DE ESTADÍSTICA

"ANÁLISIS ESTADÍSTICO MULTIVARIANTE PARA LA SEGMENTACIÓN DE LOS CLIENTES DE TELEFÓNICA"

TESIS

PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO ESTADÍSTICO:

Presentado por: **GARGATE OBREGÓN, SAMUEL GUSTAVO y LINDO HUERTAS, ELSA MARÍA.**

LIMA - PERÚ

2006

A mis padres Samuel y Luisa por todo el apoyo que me dieron en cada etapa de mi vida, de niño por la educación impartida, de adolescente por sus consejos y comprensión, de adulto por siempre tener una frase de aliento y de amor.

A mis hermanas Zoila y Marita por confiar siempre en mi y por haber compartido grandes experiencias.

A mi jefe Cesar por sus enseñanzas, su paciencia y su apoyo incondicional.

Samuel

A mi esposo, Daniel, por todo su apoyo incondicional, por su aliento constante y su amor.

A mis hijitas: Andrea, Daniela y Alejandra, por ser motivo constante de mi esfuerzo y a Dios por haberme permitido tenerlas.

A mis padres, que me inculcaron los principios fundamentales para mi vida.

Elsa

AGRADECIMIENTOS

A nuestro asesor Ms. EM Víctor Valdivieso Benavides y al Prof.: Víctor Sánchez, por todos sus consejos y paciencia para la realización del presente trabajo.

A nuestro amigo Hernán Garrafa Aragón, por su apoyo moral y sus consejos siempre oportunos.

A todos nuestros profesores que nos transmitieron sus enseñanzas y conocimientos para alcanzar nuestros objetivos profesionales.

ÍNDICE

INTRODUCCIÓN

En las últimas décadas se ha producido un gran crecimiento del uso de las técnicas estadísticas multivariantes en todos los campos de la investigación científica. Podrían darse muchas razones para este uso creciente, pero quizás las más importantes sean: que en la mayoría de investigaciones científicas, es necesario analizar relaciones simultáneas entre tres ó más variables, para lo cual muchas veces se siguen procesos iterativos, durante los cuales se añaden y eliminan continuamente variables, además la complejidad de los fenómenos analizados hacen que sean muchas las variables implicadas y por ello, las investigaciones se vuelven necesariamente multivariantes; la otra razón fundamental es el desarrollo de ordenadores con capacidad de almacenamiento y potencia de procesamiento suficiente, acompañados de programas cada vez más fáciles de usar.

Supongamos ahora que el responsable de marketing de una empresa tiene una base de datos con las características sociodemográficas de sus clientes: edad, nivel de instrucción, nivel de ingresos, tipo de ocupación, estado civil, miembros por familia, etc. Este directivo se plantea si podría dividir a sus clientes en subgrupos que tuvieran características sociodemográficas similares entre sí, pero que fueran lo más diferentes posible unos subgrupos de otros. Si esto fuera así, el directivo de marketing podría, por ejemplo, diseñar campañas de publicidad distintas para cada grupo, con creatividades diferentes o empleando diarios, revistas o canales de televisión distintas de acuerdo al grupo al que se dirige la campaña.

El análisis de conglomerados, al que también se denomina comúnmente análisis cluster, es una técnica diseñada para clasificar distintas observaciones en grupos de modo que: cada grupo (conglomerado o cluster) sea homogéneo respecto a las variables utilizadas para caracterizarlos, es decir, que cada observación contenida en él sea parecida a todas las que estén incluidas en ese grupo y que además todos los grupos formados sean lo más distintos posible unos de otros respecto a las variables consideradas.

Acorde con los tiempos actuales de constante cambio, niveles de competencia, la empresa de telecomunicaciones tiene la necesidad de segmentar a sus clientes, para lo cual en el presente estudio se hizo uso de la técnica correspondiente del análisis multivariado, obteniendo finalmente el presente informe.

RESUMEN EJECUTIVO

El presente siglo XXI en el que vivimos: el ritmo al que se logran avances tecnológicos, el desarrollo económico, la competitividad que actualmente existe a todo nivel, el cambio constante en todas las áreas en este mundo globalizado obliga a todas las empresas a tener el mayor conocimiento posible de sus clientes y el mercado al cual se dirigen. Este conocimiento del cliente requiere conocer sus necesidades, costumbres, patrones de consumo, hábitos, etc. hasta su nivel de satisfacción. Todo ello con el objetivo de conocer a los mejores clientes, manteniéndolos en la empresa, lograr que nuevos clientes, mantenerlos en la empresa, lograr que nuevos clientes se incorporen, evitar la migración de clientes a otras empresas de la competencia. Para ello deberán implementarse las campañas de marketing dirigidos a grupos de clientes que tengan características similares. Surge por ello la necesidad de identificar previamente estos grupos de clientes: es decir llevar a cabo un proceso de segmentación de los clientes, para conocer estos grupos con las mismas preferencias y así direcciones. Las campañas de marketing; lo cual permitirá ahorrar otros y un trato especial (diferenciado) a sus clientes.

En el presente trabajo se vieron las técnicas del análisis multivariado, ya que se cuenta con una gran cantidad de variables y nuestro interés fue analizar las relaciones de interdependencia entre estas variables para obtener una metodología que nos permite segmentar a todos los clientes de Telefónica. Teniendo especial interés en aquellos que no indican la rentabilidad para la empresa.

Este informe consta de 5 capítulos:

- Capítulo I: en el que se indica la población objetivo, la definición del problema, los objetivos del presente trabajo, así como las hipótesis, metodología a crear.
- Capítulo II: es el Marco Teórico que respalda nuestro Trabajo de Investigación.
- Capítulo III: es el resultado del análisis factorial que nos permite reducir variables a través de 5 factores que resultaron.
- Capítulo IV: es el proceso y resultado de la segmentación, que se hizo de dos formas, usando dos criterios: el de patrón de consumo y el de valor.

  * En la segmentación por patrón de consumo se excluyeron a los de bajo consumo (apáticos) y quedaron los que tienen consumo de Voz a Internet: resultado finalmente 5 grupos (5 clusters), de los cuales los grupos 2 y 3 corresponden a los clientes internautas con un 20% y un 52% respectivamente, finalmente se hace una microsegmentación.

  * * En la segmentación por valor, se usa como indicador referencial el margen comercial, y sobre la base de nuestro conocimiento y experiencia en el área de Telefonía, decidimos usar 5 segmentos: Oro, Plata, Bronce, Plomo y Destructores, resultarón:

    - Oro : con más de 150 soles (9% de clientes de la muestra)
    - Plata : entre 101 y 150 soles (11% de clientes de la muestra)
    - Bronce: entre 61 y 100 soles (31% de clientes de la muestra)

  - Plomo: mayor a O y menor a 60 nuevos soles (48% de clientes de la muestra)
  - Destructores: menor a 0 nuevos soles (1% de clientes de la muestra)

- Capítulo V las 7 conclusiones y recomendaciones finalmente se logró segmentar a todos los clientes de Telefónica del Perú bajo dos criterios: el patrón de consumo y el de valor, cuando los técnicos del análisis multivariado siendo esta metodología aplicada a situaciones futuras y empresas similares que deseen direccional sus promociones a través de campañas de marketing.

CAPITULO I

1.1 JUSTIFICACIÓN

El estudio del Análisis de Segmentación permitirá mostrar, con suficiente detalle, los aspectos conceptuales y matemáticos del Análisis Factorial, Análisis Cluster y Matriz de Kohonen.

Si los datos son cuantitativos, podemos estudiar también los denominados **métodos de reducción de dimensiones, análisis de componentes principales y análisis factorial.** Los cuales se utilizan para analizar interrelaciones entre un nº elevado de variables cuantitativas, explicando dichas interrelaciones en términos de un nº menor de variables denominadas factores (si no son observables) o componentes principales (si son observables).

Si los datos son cualitativos, estudiaremos el **Análisis cluster** o también denominado **método de agrupación**, cuyo objetivo es **agrupar** una

muestra de individuos o variables en un nº pequeño de grupos de forma que las observaciones de un mismo grupo sean muy similares entre sí y muy diferentes del resto. A diferencia del análisis discriminante se desconoce el nº y la composición de dichos grupos.

Análisis discriminante el cual proporciona reglas de clasificación en los grupos establecidos para las nuevas observaciones.

Con el presente trabajo se iniciará el desarrollo de investigaciones para identificar análisis estadísticos y metodologías apropiadas, aplicables al estudio y conocimiento de factores asociados a la clasificación de nuevos clientes en las diferentes empresas privadas en la cual su objetivo sean los consumidores (clientes).

1.2 OBJETIVOS

Objetivos Principales

- Segmentar a los Clientes de Telefónica, para así ofrecerles un servicio diferenciado, promociones especiales (Análisis Cluster).
- Identificar a los clientes candidatos a ser dados de baja, de cada categoría, luego de hacer previamente la clasificación correspondiente.

Objetivos Secundarios

- Clasificar a los clientes de acuerdo a sus patrones de comportamientos.
- Tener un control adecuado de nuestros clientes.
- Generar campañas diferenciadas por segmentos.

1.3 TIPO DE ESTUDIO

Es un estudio de investigación aplicada que tiene como propósito mejorar la clasificación de los nuevos clientes de Telefónica del Perú. Además permitirá estructurar conocimientos publicados sobre Análisis de Segmentación y validar su uso para la clasificación de nuevos clientes.

La aplicación de la técnica, al caso indicado, constituye un estudio explicativo puesto que está dirigido a explicar un conjunto de variables y las relaciones entre éstas en los Clientes de Telefónica del Perú, ejerciendo un control directo sobre los clientes sin modificar el comportamiento de las variables.

CAPITULO II

2.1. MARCO CONCEPTUAL DE LA INVESTIGACIÓN

Para desarrollar satisfactoriamente el tema elegido, se presentarán progresivamente los contenidos de tópicos básicos previos en la siguiente secuencia:

2.1.1. Tipos y Métodos de Análisis Multivariante

2.1.2. Análisis Factorial

2.1.3. Análisis Cluster

2.1.4. Análisis Discriminante

2.1.5. Matriz de Kohonen

A fin de inducir en el desarrollo del Análisis de Segmentación, a continuación se incluye el avance de los aspectos citados.

2.1.1. TIPOS Y MÉTODOS DE ANÁLISIS MULTIVARIANTE

Los que vamos a tratar, se pueden clasificar en 2 grandes grupos:

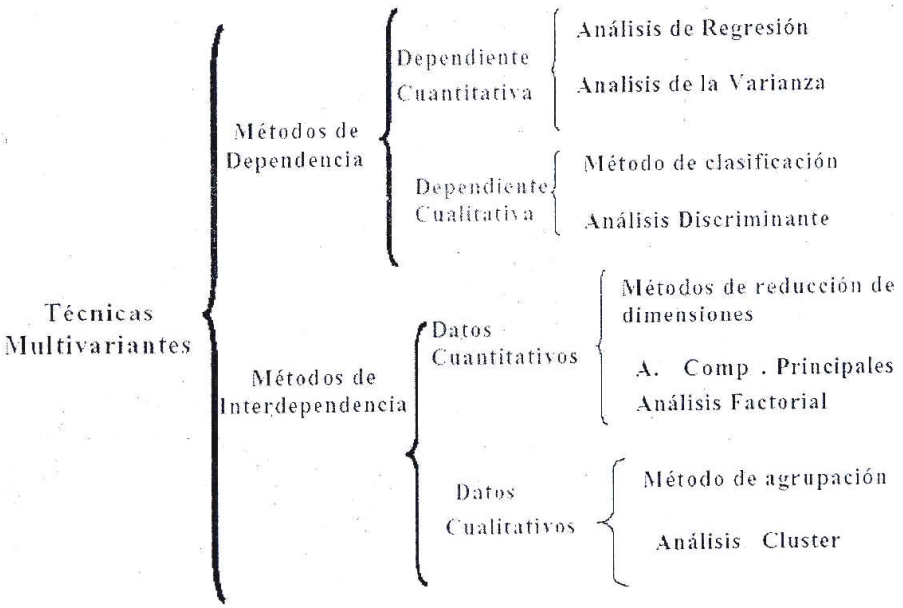
* **Métodos de dependencia**

Suponen que las variables analizadas están divididas en dos grupos: las **variables dependientes** o previctorias, y las **independientes** o explicativas.

El objeto de estos métodos consiste en determinar si el conjunto de variables independientes afecta al de dependientes y de qué forma.

*** Métodos de interdependencia**

No distinguen entre variables dependientes e independientes y su objetivo consiste en identificar que variables están relacionadas, cómo lo están y por qué.



- El **análisis multivariante** es el conjunto de técnicas estadísticas cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes, en el sentido de que hay variables medidas para cada individuo u objeto estudiado.

- Su razón de ser radica en un **mejor entendimiento del fenómeno objeto de estudio**, obteniendo información que los métodos estadísticos univariantes y bivariantes son incapaces de conseguir.

2.1.2 ANÁLISIS FACTORIAL

El Análisis Factorial ocupa un lugar primordial entre los métodos de análisis de datos, debido a las representaciones geométricas de los datos que transforman en distancias euclidianas las proximidades estadísticas entre los elementos. El objetivo del Análisis Factorial consiste en estudiar la estructura de una nube de puntos, resultado de las observaciones de las variables para cada objeto; permite encontrar una manera de condensar la información contenida en un conjunto de variables originales en un conjunto más pequeño de dimensiones o factores con una mínima pérdida de información. La obtención de estos factores están íntimamente ligada a la matriz de varianzas y covarianzas o matriz de correlación.

El Análisis Factorial permite extraer dimensiones latentes de las preferencias de ciertos productos y marcas, a nivel de consumidor, determinar imágenes de marca en función de atributos y analizar la estructura interna de una series de atributos entre si, aplicados o no a una marca concreta.

K. Pearson (1901) y C. Sperman (1904); Hotelling, Thurstone e Guttman. Algunas aplicaciones.

(i) Inteligencia en Psicometría (Spearman Inteligencia General)

(ii) Desempeño en Deportes

(iii) Productividad en Economía

Considere el siguiente modelo:

$$(X-\mu)_{px1} = L_{pxm} F_{mx1} + \varepsilon_{px1}$$

Donde:

X_{px1} : es el vector de variables aleatorias;

μ_{px1} : es el vector de medias X_{px1}

L_{pxm} : es el matriz de pesos o cargas factoriales;

F_{mx1} : es el vector de factores comunes;

ε_{px1} : es el vector de factores específicos.

Similar con el modelo de regresión múltiple, mas F es no observable.

(i) E (F_{mx1}) = O_{mx1};

(ii) E (ε_{px1}) = O_{px1};

(iii) Cov (F_{mx1}) = E (F_{mx1} F'_{mx1}) = I_{mxm}

(iv) Cov (ε_{px1}) = E (ε_{px1} ε'_{px1}) = $p\Psi p$ =

$$\begin{pmatrix} \Psi_1 & 0 & \ldots & 0 \\ 0 & \Psi_2 & \ldots & 0 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 0 & 0 & \ldots & \Psi_p \end{pmatrix}$$

(v) Cov (ε_{px1} , F'_{mx1}) = E (ε_{px1} F'_{mx1}) = 0_{pxm}

De las suposiciones se obtiene

$$\sum = LL' + \Psi$$

$$Cov(X, F) = L \quad ó \quad Cov\left(X_j, F_{k_j}\right) = l\,jk,\ (j \neq k = 1,2,..., p)$$

$$\sigma_{jj} = Var(X_j) = l^2{}_{j1} + l^2{}_{j2} + ... + l^2{}_{jm} + \Psi_j = h_j^2 + \Psi j,\ \left(j = 1,2,....., p\right)$$

Donde $h_j^2 = l_{j1}^2 + l_{j2}^2 + ... + l_{jm}^2$ es la comunidad y Ψj es la especificidad.

En la práctica, las variables originales generalmente son estandarizadas. Así, trabajaremos con la matriz de correlaciones R, en lugar de la matriz de varianzas y covarianzas $\sum$. Luego R = LL' + Ψ

Corr (X, F) = L o Corr (X_j, Fk_j) = ljk, (j $\neq$ k = 1,2,……,p)

$$1 = p_{jj} = l_{j1}^2 + l_{j2}^2 + ... + l_{jm}^2 + \Psi_j = h_j^2 + \Psi_j, \left(j = 1,2,....., p\right)$$

Donde $h_j^2 = l_{j1}^2 + l_{j2}^2 + ... + l_{jm}^2$ es la comunidad y Ψj es la especificidad.

Métodos de Estimación

El propósito del análisis factorial consiste en encontrar los mejores valores l_{ij}, los cuales representan las correlaciones entre los factores y los pesos y permitan reproducir los valores de X_{ik} más próximos a los observados y que indiquen claramente que variables pertenecen a los factores identificados. Existen diferentes métodos de estimación de estos pesos, entre ellos podemos citar a: método de componentes principales, método de ejes principales, método alfa, método de máxima verosimilitud, etc. el método más común es el de componentes principales, en el cual los pesos de los factores se obtienen de la descomposición espectral de la matriz de correlación. Los criterios para

seleccionar el número de factores, tenemos: porcentaje de variación total explicada, criterio de Catell, criterio de Kaiser. Como la meta original es encontrar la matriz de los pesos que permita indicar que variables están relacionadas a cada uno de los factores, es posible mejorar la solución inicial con una rotación de factores de tal manera que los pesos se aproximen a uno o a cero. La rotación mantiene la información total pero la reasigna a través de los factores y facilita la interpretación de estos.

Componentes Principales Por descomposición espectral

$$R = Q \wedge Q' = Q \wedge^{1/2} \wedge^{1/2} Q' = L_1 L_1^1$$

Donde $L_1 = \left(\sqrt{\lambda_1}\, u_1 \quad \sqrt{\lambda_2}\, u_2 \quad ..... \quad \sqrt{\lambda_P}\, u_P\right)$

Eliminando las columnas correspondientes a los autovalores más pequeños, o sea, aquellas que contribuyen poco a R, tenemos:

$$R \cong LL'$$

donde $L = \left(\sqrt{\lambda_1}\, u_1 \quad \sqrt{\lambda_2}\, u_2 \quad .... \quad \sqrt{\lambda_m}\, u_m\right)$ con m < p.

Si incluimos la variación del factor específico (matriz de especificidad) tenemos

$$R \cong LL' + \Psi$$

Las estimaciones de L, que son las cargas factoriales, está dada por

$$\hat{L} = \left(\sqrt{\hat{\lambda}_1}\,\hat{u}_1 \quad \sqrt{\hat{\lambda}_2}\,\hat{u}_2 \quad ..... \quad \sqrt{\hat{\lambda}_m}\,\hat{u}_m\right)$$

$y\ \hat{R} = \hat{L}\hat{L}' + \hat{\psi}.\ En\ consecuencia: \hat{\psi}_i = 1 - \hat{h}_i^2,\ (i = 1, 2, ..., p)$

que es la estimación de los elementos de la matriz de especificidad.

Para verificar la proporción de la variabilidad acumulada por los m primeros factores utilizamos el siguiente criterio.

$$\begin{cases} \dfrac{\sum_{i=1}^{m} \hat{\lambda}_j}{s_{11} + s_{22} + \ldots\ldots + s_{pp}} & \text{para análisis factorial de S} \\ \dfrac{\sum_{i=1}^{m} \hat{\lambda}_j}{p} & \text{para análisis factorial de R} \end{cases}$$

En términos prácticos, se puede retener aquellos factores asociados a los autovalores mayores que 1. Si la proporción de la variabilidad explicada por los autovalores mayores que 1 fuera baja, se toma los demás factores hasta que se consiga una proporción adecuada de la variabilidad, generalmente mayor a 60%.

Rotación de los Factores: Cuando m > 1, podemos obtener diferentes valores para las cargas factoriales, sin afectar las propiedades estadísticas. Para eso, sea T una matriz ortogonal mxm, o sea T'T = T'T = I.

El modelo factorial puede ser re-escrito como:

$$X - \mu = LF + \varepsilon = LTT'F + \varepsilon = L^*F^* + \varepsilon$$

$$con\ L^* = LT\ \ e\ \ F^* = T'F.$$

Las suposiciones del modelo factorial son satisfechas, y la descomposición de la matriz de correlaciones R puede expresarse también en términos de L* y F*:

$$R = LL' + \psi = LTT'L' + \psi = L^*L^{*'} + \psi.$$

2.1.3 ANÁLISIS CLUSTER

Objetivos

- Formación de grupos, a partir de las observaciones.
- Dentro de los grupos se reúnan las observaciones más homogéneas.
- Los grupos obtenidos sean lo más heterogéneos posibles entre si.
- Agrupar variables.
- Origen **en la Biología, Botánica. Necesidad de agrupar las distintas especies de animales y vegetales conocidas en familias homogéneas.**

Aplicaciones

- Segmentación de mercados: Identificación de grupos de consumidores con comportamientos semejantes.
- Identificar hábitos de compra, grupos de productos competitivos, mediante la valoración de los productos competidores y la posterior agrupación de los mismos.
- Identificar localidades para ser usadas como mercados de prueba. Identificar localidades que reúnen las características adecuadas.

Grupos mutuamente excluyentes basándose en la similaridad de las variables usadas. Calculando medidas de semejanzas o de diferencias.

Nota: No existe ningún tipo de restricción o condición que deben cumplir los datos. Es conveniente algunas veces, realizar algún tipo de estandarización de las variables antes de ser incluidas en el análisis.

Pasos Básicos

¿Qué variables sirven como base para la formación de los grupos?

¿De qué modo las distancias entre las observaciones son medidas?

¿Qué criterios serían utilizados para unir las observaciones dentro de los grupos?

Educación $\leftarrow$ años de experiencia a ingresos

estudiantes., # profesores $\leftarrow$ agrupar colegios

Criterios de Agrupación.- Distintas formas que se puede medir la similaridad o diferencia entre las observaciones, a partir de las cuales establecer un método de agregación de grupos.

Variables continuas: Utilizan medidas de disimilaridad, llamadas distancias, que cumplen los siguientes axiomas.

$d_{aa} > = 0$

$d_{ab} = d_{ba}$

$d_{ab} < = d_{ac} + d_{cb}$

Distancia euclideana al cuadrado, City Block, Chebyshev, Mahalanobis, Distancia generalizada de Minkowsky.

Variables continuas: Medidas de similaridad, basadas en el coeficiente de correlación: Pearson, Spearman, Kendall.

*** Frecuencias**: Chi-cuadrado, Phi-cuadrado

* **Dictómicas**: Roger y Tanimoto, Sokal y Snealth, etc.

Métodos de Agregación

Permite agrupar a los individuos más parecidos, llegando a grupos de individuos homogéneos entre si y diferentes grupo a grupo. Son Normas o criterios que deben cumplir las observaciones iniciales. Existen dos clases: Jerárquicos y Aglomerativos.

Métodos Jerárquicos

- Resulta en un conjunto de particiones, que van desde un grupo por observación hasta la inclusión de todas las observaciones en un solo grupo.
- Permiten una clasificación tanto de los individuos, como de las variables, y los resultados obtenidos se pueden representar gráficamente en dendogramas.
- No garantiza una solución óptima para un número determinado de grupos.
- Los métodos aglomerativos parten de tantos grupos como observaciones para llegar finalmente a la obtención de un único grupo.

- Los métodos disociativos parten de un único grupo para después proceder a la división del mismo en subgrupos.

Clasificación de los criterios utilizados en los métodos jerárquicos

- Distancia mínima: Eslabón simple
- Distancia máxima: Eslabón completo
- Promedio entre grupos (Eslabón promedio entre grupos)
- Media ponderada (Eslabón promedio dentro de grupos)
- Centroide
- Mediana
- Ward

Métodos No jerárquicos

- Llamados de Partición u óptimos
- Número de grupos fijados de antemano.
- Clasificación de las observaciones dentro de algunos de los grupos, mediante un proceso de optimización.
- Proceso de asignación es iterativo y no permanente entre las distintas iteraciones.

Método de k medias

- Permite la reasignación de un individuo a distintos grupos en distintos pasos del proceso, hasta llegar a la solución óptima.

- Permiten la introducción de información sobre los posibles centroides iniciales de cada uno de los grupos, datos que pueden ser extraídos de los resultados de un cluster de tipo jerárquico.

Diferencias entre los métodos jerárquicos y los de optimización.

Jerárquicos

- No existen la definición del número de grupos.
- Proceso iterativo.
- Provee de distintos tipos de resultados.
- Requiere de cantidad de cálculos, limitando su utilización para muestras grandes.
- Aplicable sobre las observaciones y sobre las variables.

De optimización

- Exige el número de grupos.
- Proporcionan índices que indican el número óptimo de grupos.
- Proporcionan los centroides de grupos.
- Permiten seleccionar variables para la interpretación de los grupos.
- Soluciones de tipo óptimo.
- Sólo puede aplicarse sobre las observaciones.

2.1.4 ANÁLISIS DISCRIMINANTE

Uno de los tópicos importantes del análisis multivariado es la técnica de "**ANALISIS DISCRIMINANTE**", cuyo interés principal

es clasificar a cada individuo (u objeto de la muestra) o un grupo de ellos en una de las categorías, grupos o poblaciones de referencia.

Una característica típica del análisis discrimante es que casi siempre trabaja con dos conjuntos de observaciones conocidas, como:

- Muestra de trabajo.
- Muestra de prueba.

La diferencia entre estas dos conjuntos de observaciones son las siguientes: La muestra de trabajo está formada por aquellas observaciones cuyo origen es conocido (es decir se conoce de que población provienen). Por otro lado la muestra de prueba está formado por observaciones para las cuales no se conocen su origen y que deben ser clasificadas en algunas de las poblaciones.

Los criterios de discriminación adoptados frecuentemente, son empleados para reducir la dimensión. La clasificación consiste en la identificación de la categoría o grupo al cual pertenece el nuevo individuo, teniendo en consideración sus características observadas; es por ello que a veces se le conoce como análisis de agrupamiento.

Cuando esas características son mediciones numéricas la designación de los grupos se llama discriminación y la combinación de mediciones recibe el nombre de función discriminante.

Específicamente, en la discriminación se intenta describir de manera gráfica (en 2 o 3 dimensiones) o algebraicamente mediante funciones llamadas discriminantes, los aspectos que permiten diferenciar a los individuos u objetos de varias poblaciones.

Cuando las funciones de discriminación son lineales o combinaciones lineales de las variables originales escogidas convenientemente, podemos proporcionar información importante, tales combinaciones simplifican la estructura de la matriz de covarianza, facilitando la interpretación de los datos.

Los métodos mas utilizados en problemas prácticos de discriminación y clasificación son el de FISHER (1936) el de razón de verosimilitud y el de BAYES, los cuales permiten obtener funciones de discriminación lineales (caso homocedástico) o cuadráticas (caso heterocedástico) en las mediciones del individuo a ser clasificado.

2.1.5 MAPAS AUTO-ORGANIZADOS DE KOHONEN: MATRIZ DE KOHONEN

Introducción

En los últimos quince años, las redes neuronales artificiales (RNA) han emergido como una potente herramienta para el modelado estadístico orientada principalmente al reconocimiento de patrones –tanto en la vertiente de clasificación como de predicción. Las RNA poseen una serie de

características admirables, tales como la habilidad para procesar datos con ruido o incompletos, la alta tolerancia a fallos que permite a la red operar satisfactoriamente con neuronas o conexiones dañadas y la capacidad de responder en tiempo real debido a su paralelismo inherente.

Actualmente, existen unos 40 paradigmas de RNA que son usados en diversos campos de aplicación (Taylor, 1996; Arbib, Erdi y Szentagothai, 1997; Sarle, 1998). Entre estos paradigmas, podemos destacar la red *backpropagation* (Rumelhart, Hinton y Williams, 1986) y los mapas autoorganizados de Kohonen (Kohonen, 1982a, 1982b).

La red backpropagation, mediante un esquema de aprendizaje supervisado, ha sido utilizada satisfactoriamente en la clasificación de patrones y la estimación de funciones.

En el presente capítulo nos proponemos describir otro de los sistemas neuronales más conocidos y empleados, los mapas autoorganizados de Kohonen. Este tipo de red neuronal, mediante un aprendizaje no supervisado, puede ser de gran utilidad en el campo del análisis exploratorio de datos, debido a que son sistemas capaces de realizar análisis de clusters, representar densidades de probabilidad y proyectar un espacio de alta dimensión mucho menor.

Los mapas autoorganizados de Kohonen

En 1982 Teuvo Kohonen presentó un modelo de red denominado mapas autoroganizados o SOM = Mapas autoorganizados, basados en ciertas evidencias descubiertas a nivel cerebral y con un gran potencial de

aplicabilidad práctica. Este tipo de red se caracteriza por poseer un aprendizaje no supervisado competitivo. Vamos a ver en qué consiste este tipo de aprendizaje.

A diferencia de lo que sucede en el aprendizaje supervisado, en el no supervisado (o autoorganizado) no existe ningún maestro externo que indique si la red neuronal está operando correcta o incorrectamente, pues no se dispone de ninguna salida objetivo hacia la cual la red neuronal deba tender. Así, durante el proceso de aprendizaje la red autoorganizada debe descubrir por sí misma rasgos comunes, regularidades, correlaciones o categorías en los datos de entrada, e incorporarlos a su estructura interna de conexiones. Se dice, por tanto, que las neuronas deben autoorganizarse en función de los estímulos (datos) procedentes de exterior.

Dentro del aprendizaje no supervisado existe un grupo de modelos de red caracterizados por poseer un aprendizaje competitivo. En el aprendizaje competitivo las neuronas compiten unas con otras con el fin de llevar a cabo una tarea dada. Con este tipo de aprendizaje, se pretende que cuando se presente a la red un patrón de entrada, sólo una de las neuronas de salida (o un grupo de vecinas) se active. Por tanto, las neuronas compiten por activarse, quedando finalmente una como neurona vencedora y anuladas el resto, que son forzadas a sus valores de respuesta mínimos.

El objetivo de este aprendizaje es caracterizar (clusterizar) los datos que se introducen en la red. De esta forma, las informaciones similares son clasificadas formando parte de la misma categoría y, por tanto, deben activar la misma neurona de salida. Las clases o categorías deben ser creadas por la

propia red, puesto que se trata de un aprendizaje no supervisado, a través de las correlaciones entre los datos de entrada.

Fundamentos biológicos

Se ha observado que en el córtex de los animales superiores aparecen zonas donde las neuronas detectoras de rasgos se encuentran topológicamente ordenadas (Kohonen, 1989, 1990); de forma que las informaciones captadas del entorno a través de los órganos sensoriales, se representan internamente en forma de mapas bidimensionales. Por ejemplo, en el área somatosensorial, las neuronas que reciben señales de sensores que se encuentran próximos en la piel se sitúan también próximas en el córtex, de manera que reproducen –de forma aproximada–, el mapa de la superficie de la piel en una zona de la corteza cerebral. En el sistema visual se han detectado mapas del espacio visual en zonas del cerebro. Por lo que respecta al sentido del oído, existen en el cerebro áreas que representan mapas tonotópicos, donde los detectores de determinados rasgos relacionados con el tono de un sonido se encuentran ordenados en dos dimensiones (Martín del Brío y Sanz, 1997).

Aunque en gran medida esta organización neuronal está predeterminada genéticamente, es probable que parte de ella se origine mediante el aprendizaje. Esto sugiere, por tanto, que el cerebro podría poseer la capacidad inherente de formar mapas topológicos de las informaciones recibidas del exterior (Kohonen, 1982a)

Por otra parte, también se ha observado que la influencia que una neurona ejerce sobre las demás es función de la distancia entre ellas, siendo muy

pequeña cuando están muy alejadas. Así, se ha comprobado que en determinados primates se producen interacciones laterales de tipo excitatorio entre neuronas próximas en un radio de 50 a 100 micras, de tipo inhibitorio en una corona circular de 150 a 400 micras de anchura alrededor del circulo anterior, y de tipo excitatorio muy débil, prácticamente nulo, desde ese punto hasta una distancia de varios centímetros. Este tipo de interacción tiene la forma típica de un sombrero mejicano como veremos más adelante.

En base a este conjunto de evidencias, el modelo de red autoorganizado presentado por Kohonen pretende mimetizar de forma simplificada la capacidad del cerebro de formar mapas topológicos a partir de las señales recibidas del exterior.

Arquitectura

Un modelo SOM está compuesto por dos capas de neuronas. La capa de entrada (formada por N neuronas, una parte por cada variable de entrada) se encarga de recibir y transmitir a la capa de salida la información procedente del exterior. La capa de salida (formada por M neuronas) es la encargada de procesar la información y formar el mapa de rasgos. Normalmente, las neuronas de la capa de salida se organizan en forma de mapa bidimensional aunque a veces también se utilizan capas de una sola dimensión (cadena lineal de neuronas) o de tres dimensiones (paralepípedo)

Las conexiones entre las dos capas que forman la red son siempre hacia delante, es decir, la información se propaga desde la capa de entrada hacia la capa de salida. Cada neurona de entrada y está conectada con cada una de

las neuronas de salida y mediante un peso W_{ji}. De esta forma, las neuronas de salida tienen asociado un vector de pesos W_j llamado vector de referencia (o codebook), debido a que constituye el vector prototipo (o promedio) de la categoría representada por la neurona de salida j.

Entre las neuronas de la capa de salida, puede decirse que existen conexiones laterales de excitación e inhibición implícitas, pues aunque no estén conectadas, cada una de estas neuronas va as tener cierta influencia sobre sus vecinas. Esto se consigue a través de un proceso de competición entre las neuronas y de la aplicación de una función denominada de vecindad como veremos más adelante.

Algoritmo

En el algoritmo asociado al modelo SOM podemos considerar, por un lado, una etapa de funcionamiento donde se presenta, ante la red entrenada, un patrón de entrada y éste se asocia a la neurona o categoría cuyo vector de referencia es el más parecido y, por otro lado, una etapa de entrenamiento o aprendizaje donde se organizan las categorías que forman el mapa mediante un proceso no supervisado a partir de las relaciones descubiertas en el conjunto de los datos de entrenamiento.

Etapa de funcionamiento

Cuando se presenta un patrón p de entrada X_p: X_{p1},…., X_{pi}, …..X_{pN}, éste se transmite directamente desde la capa de entrada hacia la capa de salida. En esta capa, cada neurona calcula la similitud entre el vector de entrada X_p y su

propio vector de pesos W_j o vector de referencia según una cierta medida de distancia o criterio de similitud establecido. A continuación, simulando un proceso competitivo, se declara vencedora la neurona cuyo vector de pesos es el más similar al de entrada.

La siguiente expresión matemática representa cuál de las M neuronas se activará al presentar el patrón de entrada X_p:

$$y_{pj} = \begin{cases} 1 \min \left\| X_p - W_j \right\| \\ 0 \ resto \end{cases}$$

Donde y_{pj} representa la salida o el grado de activación de las neuronas de salida en función del resultado de la competición (1 = neurona vencedora, 0 = neurona no vencedora), $||X_p – W_j||$ representa una medida de similitud entre el vector o patrón de entrada X_p: X_{p1}, ……..X_{pN} y el vector de pesos $_{Wj}$: W_{j1}, …W_{ji}, …W_{jN}, de las conexiones entre cada una de las neuronas de entrada y la neurona de salida j. En el siguiente apartado veremos las medidas de similitud más comúnmente utilizadas. En cualquier caso, la neurona vencedora es la que presenta la diferencia mínima.

En esta etapa de funcionamiento, lo que pretende es encontrar el vector de referencia más parecido al vector de entrada para averiguar qué neurona es la vencedora y, sobre todo, en virtud de las interacciones excitatorias e inhibitorias que existen entre las neuronas, para averiguar en que zona del espacio bidimensional de salida se encuentra tal neurona. Por tanto, lo que hace la red SOM es realizar una tarea de clasificación, ya que la neurona de salida activada ante una entrada representa la clase a la que pertenece dicha información de entrada. Además, como ante otra entrada parecida se activa la

misma neurona de salida, u otra cercana a la anterior, debido a la semejanza entre las clases, se garantiza que las neuronas topológicamente próximas sean sensibles a entradas físicamente similares. Por ese motivo, la red es espacialmente útil para establecer relaciones, desconocidos previamente, entre conjuntos de datos (Hilera y Martínez, 1995).

Etapa de aprendizaje

Se debe advertir, en primer lugar, que no existe un algoritmo de aprendizaje totalmente estándar para la red SOM. Sin embargo, se trata de un procedimiento bastante robusto ya que el resultado final es en gran medida independiente de los detalles de su realización concreta. En consecuencia, trataremos de exponer el algoritmo más habitual asociado a este modelo (Kohonen, 1982a, 1982b, 1989, 1995).

El algoritmo de aprendizaje trata de establecer, mediante la presentación de un conjunto de patrones de entrenamiento, las diferentes categorías (una por neurona de salida) que servirán durante la etapa de funcionamiento para realizar clasificaciones de nuevos patrones de entrada.

De forma simplificada, el proceso de aprendizaje se desarrolla de la siguiente manera. Una vez presentado y procesado un vector de entrada, se establece a partir de una medida de similitud, la neurona vencedora, esto es, la neurona de salida cuyo vector de pesos es el más parecido respecto al vector de entrada. A continuación, el vector de pesos asociados a la neurona vencedora se modifica de manera que se parezca un poco más al vector de entrada. De este modo, ante el mismo patrón de entrada, dicha neurona responderá en el futuro

todavía con más intensidad. El proceso se repite para un conjunto de patrones de entrada los cuales son presentados repetidamente a la red, de forma que al final los diferentes vectores de pesos sintonizan con uno o varios patrones de entrada y, por tanto, con dominios específicos del espacio de entrada. Si dicho espacio está dividido en grupos, cada neurona se especializará en uno de ellos, y la operación esencial de la red se podrá interpretar como un análisis de clusters.

La siguiente interpretación geométrica (Masters, 1993) del proceso de aprendizaje puede resultar interesante para comprender la operación de la red SOM. El efecto de la regla de aprendizaje no es otro que acercar de forma iterativa el vector de pesos de la neurona de mayor actividad (ganadora) al vector de entrada. Así, en cada iteración el vector de pesos de la neurona vencedora rota hacia el de entrada, y se aproxima a él en una cantidad que depende del tamaño de una tasa de aprendizaje.

Al finalizar el aprendizaje, el vector de referencia de cada neurona de salida se corresponderá con el vector de entrada que consigue activar la neurona correspondiente. En el caso de existir más patrones de entrenamiento que neuronas de salida, como en el ejemplo expuesto, más de un patrón deberá asociarse con la misma neurona, es decir, pertenecerán a la misma clase. En tal caso, los pesos que componen el vector de referencia se obtienen como un promedio (centroide) de dichos patrones.

Además de este esquema de aprendizaje competitivo, el modelo SOM aporta una importante novedad, pues incorpora relaciones entre las neuronas próximas en el mapa. Para ello, introduce una función denominada zona de

vecindad que define un entorno alrededor de la neurona ganadora actual (vecindad); su efecto es que durante el aprendizaje se actualizan tanto los pesos de la vencedora como los de las neuronas pertenecientes a su vecindad. De esta manera, en el modelo SOM se logra que neuronas próximas sintonicen con patrones similares, quedando de esta manera reflejada sobre el mapa una cierta imagen del orden topológico presente en el espacio de entrada.

Una vez entendida la forma general de aprendizaje del modelo SOM, vamos a expresar este proceso de forma matemática. Recordemos que cuando se presenta un patrón de entrenamiento, se debe identificar la neurona de salida vencedora, esto es, la neurona cuyo vector de pesos sea el más parecido al patrón presentado. Un criterio de similitud muy utilizado es la distancia euclídea que viene dado por la siguiente expresión:

$$\min \left\| X_p - W_j \right\| = \min \sum_{i=1}^{N} \left(x_{pi} - W_{ji} \right)^2$$

De acuerdo con este criterio, dos vectores serán más similares cuanto menor sea su distancia.

Una medida de similitud alternativa más simple que la euclídea, es la correlación o producto escalar:

$$\min \left\| X_p - W_j \right\| = i = 1 \sum^{N} X_{pi} \; W_{ji}$$

Según la cual, dos vectores serán más similares cuanto mayor sea su correlación.

Identificada la neurona vencedora mediante el criterio de similitud, podemos pasar a modificar su vector de pesos asociado y el de sus neuronas vecinas, según la regla de aprendizaje:

$$\Delta W_{ji}(n+1) = a(n)\left(x_{pi} - w_{ji}(n)\right) \; para \; j \in Zona_{j^*}(n)$$

Donde n hace referencia al número de ciclos o iteraciones, esto es, el número de veces que ha sido presentado y procesado todo el juego de patrones de entrenamiento $\propto$ (n) es la tasa de aprendizaje que, con un valor inicial entre 0 y 1, decrece con el número de iteraciones (n) del proceso de aprendizaje. $Zona_{j^*}(n)$ es la zona de vecindad alrededor de la neurona vencedora j* en la que se encuentran las neuronas cuyos pesos son actualizados. Al igual que la tasa de aprendizaje, el tamaño de esta zona normalmente se va reduciendo paulatinamente en cada iteración, con lo que el conjunto de neuronas que pueden considerarse vecinas cada vez es menor.

Tradicionalmente el ajuste de los pesos se realiza después de presentar cada vez un patrón de entrenamiento, como se muestra en la regla de aprendizaje expuesta. Sin embargo, hay autores (Masters, 1993) que recomiendan acumular los incrementos calculados para cada patrón de entrenamiento y, una vez presentados todos los patrones, actualizar los pesos a partir del promedio de incrementos acumulados. Mediante este procedimiento se evita que la dirección del vector de pesos vaya oscilando de un patrón a otro y acelera la convergencia de los pesos de la red.

En el proceso general de aprendizaje suelen considerarse dos fases. En la primera fase, se pretende organizar los vectores de pesos en el mapa. Para ello, se comienza con una tasa de aprendizaje y un tamaño de vecindad

grandes, para luego ir reduciendo su valor a medida que avanza el aprendizaje. En la segunda fase, se persigue el ajuste fino del mapa, de modo que los vectores de pesos se ajusten más a los vectores de entrenamiento. El proceso es similar al anterior aunque suele ser más largo, tomando la tasa de aprendizaje constante e igual a un pequeño valor (por ejemplo, 0.01) y un radio de vecindad constante e igual a 1.

No existe un criterio objetivo acerca del número total de iteraciones necesarias para realizar un buen entrenamiento del modelo. Sin embargo, el número de iteraciones debería ser proporcional al número de neuronas del mapa (a más neuronas, son necesarias más iteraciones) e independientemente del número de variables de entrada. Aunque 500 iteraciones por neurona es una cifra adecuada, de 50 a 100 suelen ser suficientes para la mayor parte de los problemas (Kohone, 1990).

Fases en la aplicación de los mapas autoorganización

En el presente apartado, pasamos a describir las diferentes fases necesarias para la aplicación de los mapas autoorganizados a un problema tipo de agrupamiento de patrones.

Iniciación de los pesos

Cuando un mapa autoorganizado es diseñado por primera vez, se deben asignar valores a los pesos a partir de los cuales comenzar la etapa de entrenamiento. En general, no existe discusión en este punto y los pesos se inicializan con pequeños valores aleatorios, por ejemplo, entre -1 y 1 ó entre 0 y

1 (Kohonen, 1990), aunque también se pueden inicializar con valores nulos (Martín del Brío y Serrano, 1993) o a partir de una selección aleatoria de patrones de entrenamiento (SPSS Inc., 1997).

Entrenamiento de la red

Vista la manera de modificar los vectores de pesos de las neuronas a partir del conjunto de entrenamiento, se van a proporcionar una serie de consejos prácticos acerca de tres parámetros relacionados con el aprendizaje cuyos valores óptimos no pueden conocerse *a priori* dado un problema.

Medida de similitud

Hemos visto las dos medidas de similitud más ampliamente utilizadas a la hora de establecer la neurona vencedora ante la presentación de un patrón de entrada, tanto en la etapa de funcionamiento como en la etapa de aprendizaje de la red. Sin embargo, se debe advertir que el criterio de similitud y la regla de aprendizaje que se utilicen en el algoritmo deben ser métricamente compatibles. Si esto no es así, estaríamos utilizando diferentes métricas para la identificación de la neurona vencedora y para la modificación del vector de pesos asociados, lo que podría causar problemas en el desarrollo del mapa (Demartines y Blayo, 1992).

La distancia euclídea y la regla de aprendizaje presentada son métricamente compatibles y, por tanto, no hay problema. Sin embargo, la correlación o producto escalar y la regla de aprendizaje presentada no son compatibles, ya que dicha regla procede de la métrica euclídea y la correlación solamente es

compatible con esta métrica si se utilizan vectores normalizados (en cuyo caso distancia euclídea y correlación coinciden). Por tanto, si utilizamos la correlación como criterio de similitud, deberíamos utilizar vectores normalizados; mientras que si utilizamos la distancia euclídea, esto no será necesario (Martín del Brío y Sanz, 1997). Finalmente, independientemente del criterio de similitud utilizado, se recomienda que el rango de posibles valores de las variables de entrada sea el mismo, por ejemplo, entre -1 y 1 ó entre 0 y 1 (Masters, 1993).

Tasas de aprendizaje

Como ya se ha comentado, $\propto$(n) es la tasa de aprendizaje que determina la magnitud del cambio en los pesos ante la presentación de un patrón de entrada. La tasa de aprendizaje, con un valor inicial entre 0 y 1, por ejemplo, 0.6, decrece con el número de iteraciones (n), de forma que cuando se ha presentado un gran número de veces todo el juego de patrones de aprendizaje, su valor es prácticamente nulo, con lo que la modificación de los pesos es insignificante. Normalmente, la actualización de este parámetro se realiza mediante una de las siguientes funciones (Hilera y Martínez, 1995):

$$\alpha(n) = \frac{1}{n} \qquad \alpha(n) = \alpha_j \left(1 - \frac{n}{\alpha_2}\right)$$

Siendo $\propto_1$ un valor de 0.1 ó 0.2 y $\propto_2$ un valor próximo al número total de la iteraciones del aprendizaje. Suele tomarse un valor $\propto_2$ = 10000.

El empleo de una u otra función no influye en exceso en el resultado final.

Zona de vecindad

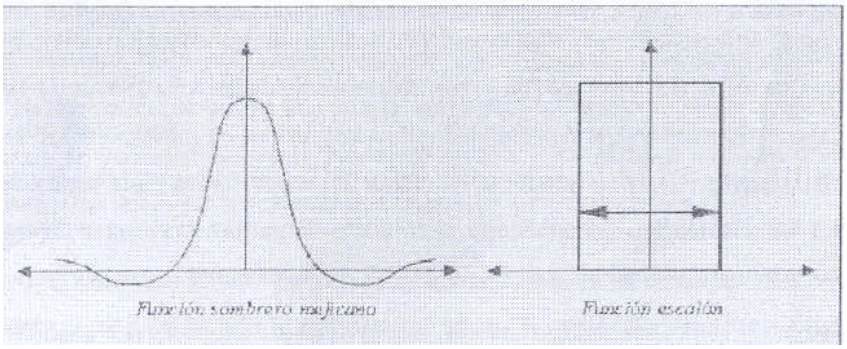
La zona de vecindad ($Zona_{j^*}(n)$) es una función que define en cada iteración n si una neurona de salida pertenece o no a la vecindad de la vencedora j*. La vecindad es simétrica y centrada en j*, pudiendo adoptar una forma circular, cuadrada, hexagonal o cualquier otro polígono regular.

En general, $Zona_{j^*}(n)$ decrece a medida que avanza el aprendizaje y depende de un parámetro denominado radio de vecindad R(n), que representa el tamaño de la vecindad actual.

La función de vecindad más simple y utilizada es la de tipo escalón. En este caso, una neurona j pertenece a la vecindad de la ganadora j* solamente si su distancia es inferior o igual a R(n). Con este tipo de función, las vecindades adquieren una forma (cuadrada, circular, hexagonal, etc.) de bordes nítidos, en torno a la vencedora; por lo que en cada iteración únicamente se actualizan las neuronas que distan de la vencedora menos o igual a R(n).

También se utilizan a veces funciones gaussianas o en forma de sombrero mejicano (figura), continuas y derivables en todos sus puntos, que al delimitar vecindades decrecientes en el dominio espacial establecen niveles de pertenencia en lugar de fronteras nítidas.

Figura de la Campana Gaussiana



Formas de la función de vecindad

La función en forma de sombrero mejicano se basa en el tipo de interacción que se produce entre ciertas neuronas del córtex comentado al inicio del documento. Con esta función, una neurona central emite señales excitatorias a una pequeña vecindad situada a su alrededor. A medida que aumenta la distancia lateral desde la neurona central, el grado de excitación disminuye hasta convertirse en una señal inhibitoria. Finalmente, cuando la distancia es considerablemente grande la neurona central emite una débil señal excitatoria. Por su parte, la función escalón supone una simplificación de la función en forma de sombrero mejicano y, como hemos visto, define de forma discreta la vecindad de neuronas que participan en el aprendizaje.

La zona de vecindad posee una forma definida, pero como hemos visto, su radio varía con el tiempo. Se parte de un valor inicial R_0 grande, por ejemplo, igual al diámetro total del mapa (SOM_PAK, 1996, Koski, Alanen, Komu et al., 1996), que determina vecindades amplias, con el fin de lograr la ordenación global del mapa. R(n) disminuye monótonamente con el tiempo, hasta alcanzar un valor final de $R_f = 1$, por el que solamente se actualizan los pesos de la neurona vencedora y las adyacentes. Una posible función de actualización de R(n) es la siguiente (Martín del Brío y Sanz, 1997):

$$R_{(n)} = R_0 + \left(R_f - R_0\right)\frac{n}{n_R}$$

Donde n es la iteración y n_R el número de iteraciones para alcanzar R_f.

Evaluación del ajuste del mapa

En los mapas autoorganizados, el conjunto de vectores de pesos finales va a depender entre otros factores, del valor de los pesos aleatorios iniciales, el valor de la tasa de aprendizaje, el tipo de función de vecindad utilizando y la tasa de reducción de estos dos últimos parámetros. Como es obvio, debe existir un mapa óptimo que represente de forma fiel las relaciones existentes entre el conjunto de patrones de entrenamiento. El mapa más adecuado será aquel cuyos vectores de pesos se ajusten más al conjunto del error cuantificador promedio a partir de la media de $|| X_p – W_{j*}||$, esto es, la media de la diferencia (por ejemplo, la distancia euclídea) entre cada vector de entrenamiento y el vector de pesos asociado a su neurona vencedora (SOM_PAK, 1996). La expresión del error cuantificador promedio utilizada en nuestras simulaciones es la siguiente:

$$\mathrm{E}rror_{medio} = \frac{\sum_{p=1}^{P}\sum_{i=1}^{N}\left(x_{pi} - w_{j*i}\right)^2}{P}$$

Por tanto, con el objeto de obtener un mapa lo más adecuado posible, deberíamos comenzar el entrenamiento en múltiples ocasiones, cada vez utilizando una configuración de parámetros de aprendizaje diferentes. Así, el mapa que obtenga el error cuantificador promedio más bajo será el seleccionado para pasar a la fase de funcionamiento normal de la red.

Visualización y funcionamiento del mapa

Una vez seleccionado el mapa óptimo, podemos pasar a la fase de visualización observando en qué coordenadas del mapa se encuentra la

neurona asociada a cada patrón de entrenamiento. Esto nos permite proyectar el espacio multidimensional de entrada en un mapa bidimensional y, en virtud de la similitud entre las neuronas vecinas, observar los clusters o agrupaciones de datos organizados por la propia red. Por este motivo, el modelo de mapa autoorganizado es especialmente útil para establecer relaciones, desconocidas previamente, entre conjuntos de datos.

En la fase de funcionamiento, la red puede actuar como un clasificador de patrones ya que la neurona de salida activada ante una entrada nueva representa la clase a la pertenece dicha información de entrada. Además, como ante otra entrada parecida se activa la misma activa de salida, u otra cercana a la anterior, debido a la semejanza entre clases, se garantiza que las neuronas topológicamente próximas sean sensibles a entradas físicamente similares.

CAPITULO III

3.1. Segmentación por patrón de Consumo:

3.1.1 Construcción de la matriz de datos

Composición de la matriz de datos:

La información considerada se ordena en una matriz de datos compuesta por cinco grandes grupos de variables, siendo las siguientes:

- Variables de identificación: Es una columna con la variable Código Único del Cliente (CUC). La variable de agrupación es el CUC, no la línea. Un CUC puede agrupar varias líneas propiedad de un mismo titular.
- Variables de segmentación: Incluye todas las variables disponibles referentes a minutos y número de llamadas (tipo de llamada, horario, día de la semana, etc.).
- Variables de productos: Incluye las variables de tenencia de productos y servicios y las variables de productos y servicios dados de baja.

- Variables de caracterización: Incluye todas las variables disponibles (número de líneas, antigüedad de la línea,...).
- Variables de valor: Incluye las variables de facturación y otras que dan origen a las variables de la Función Valor (Margen Bruto y Margen Comercial).

En esta matriz, para cada cliente, se reflejan los valores totales de todas las líneas que el cliente posee en las 532 variables origen. Ver detalle de composición de la matriz en Anexo 1 *Matriz de variables originales.*

Los valores de las variables origen consideran periodos diferentes debido a limitaciones de disponibilidad de la información en las bases de datos de La Empresa de Telecomunicaciones.

Por lo tanto tenemos:

- Variables de identificación, antigüedad, ubicación geográfica, nivel socioeconómico y otros datos generales del cliente: noviembre 2005
- Variables de valores medios de número de llamadas y tráfico expresado en minutos reales (saliente, prepago, entrante y datos) referidos al histórico de 3 meses: septiembre – noviembre 2005
- Variables de tenencia de productos referidos al histórico de un año: La tenencia actual de productos y/o servicios hace referencia al mes de diciembre del 2005. El detalle histórico de altas y bajas toma como referencia el año 2004, de enero a diciembre
- Variables de facturación por tráfico medido de 1 año: enero a diciembre 2005.
- Variables de abono de servicios (cargo fijo por SVA): enero 2005.

- Variables de costo para construir la cuenta de resultados del cliente que provienen de las distintas unidades involucradas en los respectivos procesos:
  - Costos de interconexión: noviembre 2005
  - Costos de información y atención a clientes: noviembre 2005.
  - Costos de asistencia técnica y facturación: noviembre 2005.

3.1.2 Sumatoria de las variables de segmentación, productos, caracterización y valor

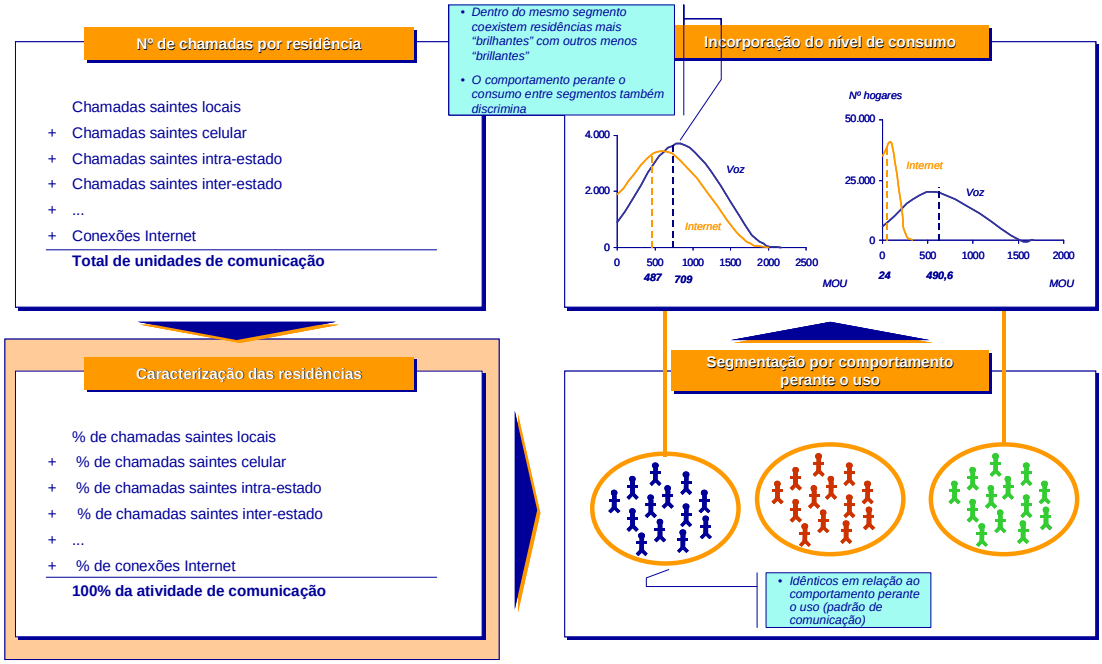
Es importante señalar que toda la información obtenida del Datawarehouse es a nivel de línea telefónica. Mientras no se disponga de la información estandarizada a nivel de hogar se deberá agrupar toda la información a nivel de cliente de la siguiente manera:

a) Variables de segmentación: se suma la información de tráfico y cantidad de llamadas.
b) Variables de productos: se suma el total de líneas telefónicas y de servicios. En el caso de los packs de productos, se consideran de manera total y desagregada, es decir, para cada línea, se indica que posee tanto un pack como cada uno de los productos que conforman dicho pack.
c) Variables de caracterización: las variables de nivel socioeconómico, departamento, distrito, ciudad y URA poseen criterios distintos para su agrupación. Estos se detallan en el Anexo 3 *Agrupación de variables internas a nivel de cliente*
d) Variables de valor: se suman los valores de cada línea, obteniendo el valor total por cliente.

3.1.3 Proceso de construcción de las variables de segmentación

Una vez construida la matriz de datos, se puede proceder a la construcción de las variables de segmentación.

Gráfico 1.3-1 Fases de la segmentación: construcción de las variables de segmentación



Las variables proporcionadas por la matriz, juntamente con aquellas variables calculadas a partir de las variables originales, son medidas en términos absolutos (número de minutos, número de llamadas, etc.). Sin embargo, el objetivo de la segmentación es la identificación de pautas de consumo, por lo que las diferencias absolutas entre clientes que siguen una pauta común de consumo determinada no interesan inicialmente. Es decir, dos clientes que tienen una proporción de destino de llamadas idéntico, pero tienen diferente número de llamadas, a efectos de la segmentación, son clientes idénticos en pauta de comportamiento. Para ello, se ha aislando el factor de cantidad de número de

llamadas de la segmentación –cuánto consumen- centrándonos en el cómo el cliente consume.

Por esta razón, los valores de las variables deben ser relativizados, expresando las variables en porcentajes sobre el total del consumo del cliente. De esta manera, se eliminan las diferencias de cantidad entre los clientes centrándose en el comportamiento de uso, el cómo esos clientes consumen, no el cuánto consumen.

Gráfico 1.3-2 Caracterización de los clientes en función de su actividad de comunicación

Siguiendo el ejemplo anterior, es posible formar 5 bloques de variables que suman 100% cada uno:

- Llamadas salientes, entrantes y conexiones a Internet (bloque original)
- Llamadas salientes por destino y entrantes por origen
- Llamadas en horarios normal y reducido

- Llamadas en día laboral, feriado, sábado y domingo
- Llamadas durante la mañana, la tarde, la noche y la madrugada

A continuación se detalla el cálculo de las variables de segmentación (ver el Anexo 1 *Matriz de variables original*):

a) Composición de llamadas según bloque original:

Tabla 1.3-1 Cálculo de las variables de número total de llamadas y porcentaje– bloque original

Descripción	Nombre de la variable	Composición
Total llamadas	tllamada	v39 + v40 + v87 + v88 + v143 + v144 + v165
Total llamadas salientes	llamsal	v39 + v40 + v87 + v88
Total llamadas entrantes	llament	v143 + v144
Total conexiones Internet	llamint	v165
Porcentaje llamadas salientes	pllasal	(llamsal / tllamada) *100
Porcentaje llamadas entrantes	pllaent	(llament / tllamada) *100
Porcentaje conexiones Internet	pllaint	(llamint / tllamada) *100

b) Composición de llamadas según destino y origen

Tabla 1.3-2 Cálculo de las variables porcentaje de llamadas en función del origen o destino

Descripción	Nombre de la variable	Composición
Porcentaje conexiones Internet	pllaint	(llamint / tllamada) *100
Porcentaje de llamadas salientes –incluidas llamadas prepago- a números fijos locales (incluye números intra y extrared)	pslocfij	(v19 + v20 + v71 + v72) / tllamada)*100
Porcentaje de llamadas salientes –incluidas llamadas prepago- a números móviles locales	psmovloc	((v21 + v73) / tllamada)*100
Porcentaje de llamadas salientes –incluidas llamadas prepago- a números móviles larga distancia nacional	psmovldn	((v22 + v74) / tllamada)*100
Porcentaje de llamadas salientes –incluidas llamadas prepago- a números fijos larga distancia nacional (incluye números intra y extrared)	psfijldn	((v23 + v24 + v25 + v75 + v76) / tllamada)*100
Porcentaje de llamadas salientes –incluidas llamadas prepago- a números de larga distancia internacional	psfijldi	((v26 + v28 + v77) / tllamada)*100
Porcentaje de llamadas salientes –incluidas llamadas prepago- a la casilla de buzón de voz (para rescatar mensajes de su casilla)	psbuz159	((v29 + v79) / tllamada)*100
Porcentaje de llamadas salientes –incluidas llamadas prepago- a la casilla de buzón de voz (para dejar mensajes de la casilla)	psbuz158	((v29a + v79a) / tllamada)*100
Porcentaje de llamadas salientes –incluidas llamadas prepago- a números de atención a clientes (102, 103, 104)	pscliente	((v30 + v80) / tllamada)*100
Porcentaje de llamadas salientes –incluidas llamadas prepago- a números gratuitos	psgratui	(v31 / tllamada)*100
Porcentaje de llamadas salientes –incluidas llamadas prepago- a números de tarifa diferenciada	psdifere	((v32 + v81) / tllamada)*100
Porcentaje de llamadas entrantes desde números fijos locales (incluye números intra y extrared)	pefijloc	((v123 +v124) / tllamada)*100
Porcentaje de llamadas entrantes desde números móviles locales	pemovloc	(v125 / tllamada)*100
Porcentaje de llamadas entrantes desde números móviles larga distancia nacional	pemovldn	(v126 / tllamada)*100
Porcentaje de llamadas entrantes desde números fijos larga distancia nacional (incluye números intra y extrared)	pefijldn	((v127 + v128) / tllamada)*100
Porcentaje de llamadas entrantes desde números larga distancia internacional	peldi	(v129 / tllamada)*100
Porcentaje de llamadas entrantes larga distancia nacional con cobro revertido	pecobrev	(v132 / tllamada)*100
Porcentaje de llamadas entrantes desde teléfonos públicos	petups	(v137 / tllamada)*100

c) Composición de llamadas según bloque horario

Tabla 1.3-3 Cálculo de las variables porcentaje de llamadas en función del bloque horario

Descripción	Nombre de la variable	Composición
Porcentaje de llamadas salientes en horario normal	psnormal	((v39+ v87) / tllamada)*100
Porcentaje de llamadas salientes en horario reducido	psreduci	((v40 + v88) / tllamada)*100
Porcentaje de llamadas entrantes en horario normal	penormal	v143*100 / tllamada
Porcentaje de llamadas entrantes en horario reducido	pereduci	v144*100 / tllamada
Porcentaje de conexiones a Internet en horario normal	pinormal	v171*100 / tllamada
Porcentaje de conexiones a Internet en horario reducido	pireduci	v172*100 / tllamada

Dado que las operadoras móviles y de larga distancia poseen distintas franjas horarias para la tarifa reducida, se ha determinado que la franja horaria de la Empresa de Telecomunicaciones sea la que rija a las demás operadoras.

d) Composición de llamadas según día de la semana

Tabla 1.3-4 Cálculo de las variables porcentaje de llamadas en función del día

Descripción	Nombre de la variable	Composición
Porcentaje de llamadas salientes realizadas en día laboral	pslabora	((v48 + v96) / tllamada)*100
Porcentaje de llamadas salientes realizadas en día feriado	psferiad	((v49 + v97)/ tllamada)*100
Porcentaje de llamadas salientes realizadas en día domingo	psdoming	((v50 + v98) / tllamada)*100
Porcentaje de llamadas salientes realizadas en día sábado	pssabado	((v51 + v99) / tllamada)*100
Porcentaje de llamadas entrantes recibidas en día laboral	pelabora	v152*100 / tllamada
Porcentaje de llamadas entrantes recibidas en día feriado	peferiad	v153*100 / tllamada
Porcentaje de llamadas entrantes recibidas en día domingo	pedoming	v154*100 / tllamada
Porcentaje de llamadas entrantes recibidas en día sábado	pesabado	v155*100 / tllamada
Porcentaje de conexiones a Internet realizadas en día laboral	pilabora	v180*100 / tllamada
Porcentaje de conexiones a Internet realizadas en día feriado	piferiad	v181*100 / tllamada
Porcentaje de conexiones a Internet realizadas en día domingo	pidoming	v182*100 / tllamada
Porcentaje de conexiones a Internet realizadas en día sábado	pisabado	v183*100 / tllamada

e) Composición de llamadas según momento del día

Tabla 1.3-5 Cálculo de las variables porcentaje de llamadas en función del momento del día

Descripción	Nombre de la variable	Composición
Porcentaje de llamadas salientes realizadas durante la mañana	psmañana	((v56 + v104) / tllamada)*100
Porcentaje de llamadas salientes realizadas durante la tarde	pstarde	((v57 + v105) / tllamada)*100
Porcentaje de llamadas salientes realizadas durante la noche	psnoche	((v58 + v106) / tllamada)*100
Porcentaje de llamadas salientes realizadas durante la madrugada	psmadrug	((v59 + v107) / tllamada)*100
Porcentaje de llamadas entrantes recibidas durante la mañana	pemañana	v160*100 / tllamada
Porcentaje de llamadas entrantes recibidas durante la tarde	petarde	v161*100 / tllamada
Porcentaje de llamadas entrantes recibidas durante la noche	penoche	v162*100 / tllamada
Porcentaje de llamadas entrantes recibidas durante la madrugada	pemadrug	v163*100 / tllamada

Los momentos del día en la Empresa de Comunicaciones se dividen de la siguiente manera:

- Mañana: 08:00 a 11:59 horas
- Tarde: 12:00 a 17:59 horas
- Noche: 18:00 a 23:59 horas
- Madrugada: 00:00 a 07:59 horas

f) Composición de llamadas sumarizadas:

Tabla 1.3-6 Cálculo de las variables porcentaje de llamadas en función del día y en función del momento del día sumarizadas

Descripción	Nombre de la variable	Composición
Porcentaje de llamadas en día laboral	pllalabo	((v48 + v96 + v152 + v180) / tllamada)*100
Porcentaje de llamadas en día feriado	pllaferi	((v49 + v97 + v153 + v181) / tllamada)*100
Porcentaje de llamadas en día domingo	plladomi	((v50 + v98 + v154 + v182) / tllamada)*100
Porcentaje de llamadas en día sábado	pllasaba	((v51 + v99 + v155 + v183)/ tllamada)*100
Porcentaje de llamadas durante la mañana	pllamaña	(V56 + v104 + v160) / tllamada)*100
Porcentaje de llamadas durante la tarde	pllatard	(v57 + v105 + v161) / tllamada)*100
Porcentaje de llamadas durante la noche	pllanoch	(v58 + v106 + v162) / tllamada)*100
Porcentaje de llamadas durante la madrugada	pllamadr	(v59 + v107 + v163) / tllamada)*100

Es importante señalar que este último bloque de variables, junto al bloque de variables *porcentaje de llamadas en función del origen o destino,* fueron los que finalmente se emplearon para el proceso de segmentación por patrón de consumo. El resto de bloques de variables también se empleó en diversos análisis y para completar la información disponible en la toma de decisiones.

3.1.4 Proceso de limpieza de la matriz

Una vez generada la información que constituye la matriz y la construcción de las variables calculadas que resumen variables originales, es necesario llevar a cabo un proceso de limpieza de la matriz para garantizar la calidad y fiabilidad de los resultados de la segmentación.

El proceso de limpieza de la matriz se estructura en cuatro etapas fundamentales realizadas sobre la matriz original de 57.299 clientes:

a) Identificación de variables con valores missing, es decir, donde se carece de datos en las variables de segmentación. Es necesario la identificación de clientes con valores missing para no considerarlos en la segmentación y evitar afectar a los resultados.

    a. En la matriz entregada no se encuentra ningún cliente con valores missing por lo que no fue necesario la extracción de la matriz de ningún cliente.

b) Eliminación de clientes con consumo de cero llamadas en las variables de segmentación:

    Estos clientes son denominados “muertos” y se extraen de la matriz de datos porque su inclusión causa distorsiones al no presentar ningun patrón de consumo. Se identifican 367 clientes con consumo cero (0,64% de la muestra) que no se considerarán en la segmentación. Para seleccionar los clientes con consumo cero, se toma la variable *total de llamadas* (tllamada) como criterio de selección. SPSS identifica estos clientes y crea una variable adicional (muertos) donde se indica si el cliente tiene consumo cero (if tllamada = 0 muertos = 0) o no (if v198 > 0 muertos = 1).

c) Eliminación de los datos extremos (outliers):

    datos extremos (outliers) es un individuo que presenta en sus variables algún error de medición. Luego de realizar un análisis descriptivo (media, mediana, desviación estándar, máximos y mínimos, y percentile), se detecta que la variable *número de llamadas a buzón de voz* (v29) presenta valores excesivamente altos.

Para determinar un umbral de corte, se realiza un análisis de frecuencias y se identifica que el punto de inflexión ocurre en las 226 llamadas (a partir de esta cantidad las llamadas al buzón de voz se elevan aun ritmo mucho mayor). De esta manera se identifican 59 clientes que superan esta marca y se eliminan del análisis (0,10% de la muestra). SPSS identifica estos clientes y crea una variable adicional (buz159) donde se indica si el cliente es outlier (if v29 >= 226 buz159 = 0) o no (if v29 < 226 buz159 = 1).

d) Eliminación de clientes con antigüedad menor a 6 meses:

La petición original de variables considera a clientes con una antigüedad mayor a 6 meses. Algunos clientes no cumplen con esta condición inicial, por lo que se procede a su eliminación.

Se detecta mediante la variable *antigüedad del cliente* (v199) que 37 clientes incumplen con la condición (0,7% de la muestra).

SPSS identifica estos clientes y crea una variable adicional (antigued) donde se indica si el cliente cumple la condicion de antigüedad (if v199 >= 6 antigued = 1) o no (if v199 < 6 antigued = 0).

e) Eliminación de clientes con posible comportamiento similar al de una empresa:

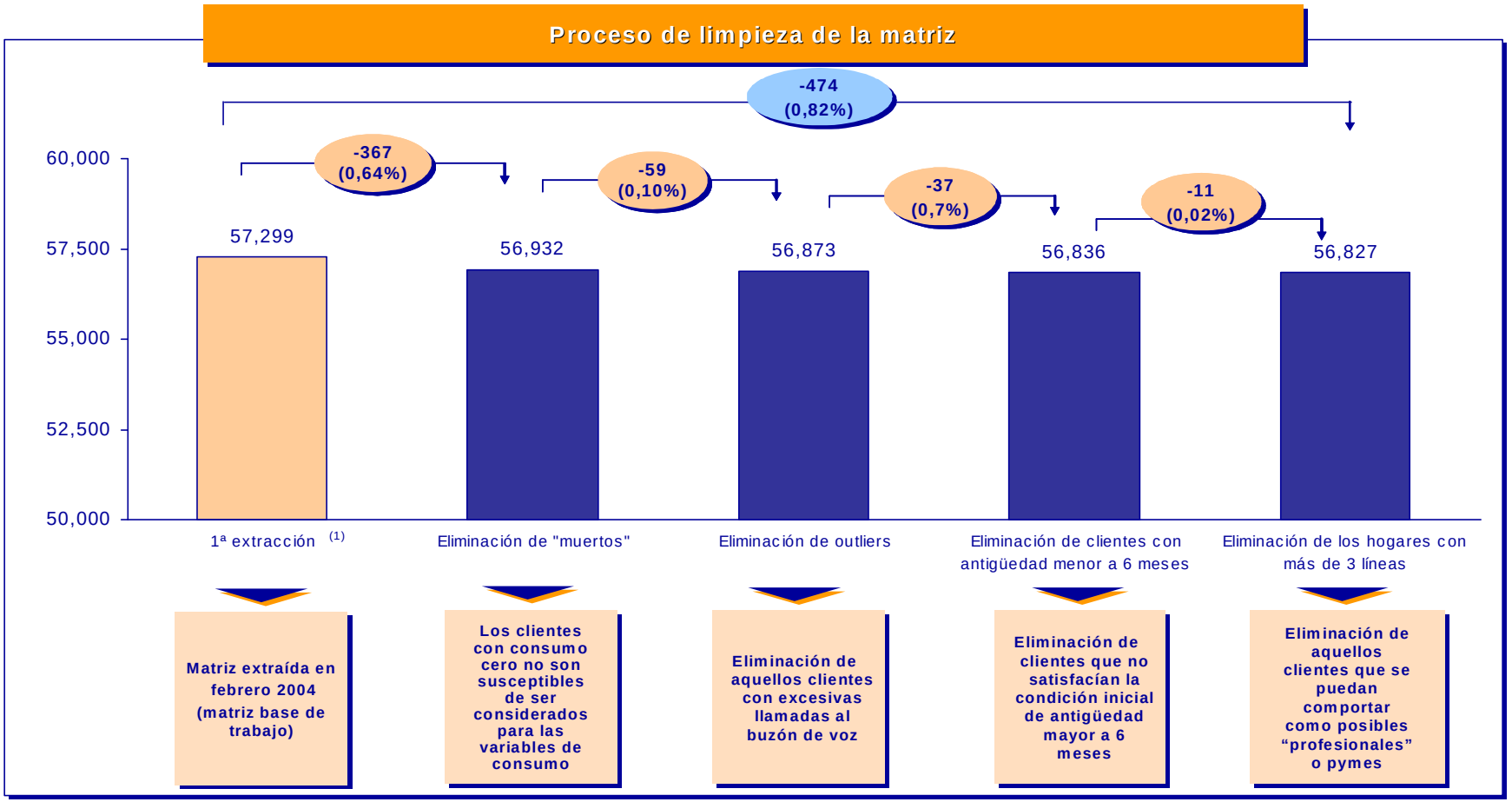
No considerar aquellos clientes con más de 3 líneas. En este proceso de identificación de potenciales empresas se identifican un total de 11 clientes con más de 3 líneas (0,02% de la muestra) que no se considerarán en la segmentación residencial. Para seleccionar los clientes con más de 3 líneas, se toma la variable *número de líneas de cada hogar* (v198) como criterio de selección. SPSS identifica estos

clientes y crea una variable adicional (residenc) donde se indica si el cliente tiene más de 3 líneas (if v198 > 3 residenc = 0) o no (if v198 < 3 residenc = 1).

Tras el proceso de limpieza, se consigue una matriz de trabajo con un total de 56.827 clientes donde no existen variables con missing en las variables de segmentación, no existen clientes con consumo cero en las variables de segmentación, no existen outliers, no existen clientes con menos de 6 meses de antigüedad y no existen clientes que podrían ser empresas.

A continuación se muestra el esquema de la fase de limpieza de la matriz.

Gráfico 1.4-1 Proceso de limpieza de la matriz de datos



3.1.5 Proceso de segmentación por patrón de consumo

El objetivo de la segmentación por patrón de consumo es identificar pautas diferenciales de consumo entre los clientes, por lo que como primer paso se identifican los clientes de bajo consumo (apáticos) y en una segunda fase aquellos clientes que consumen voz e Internet.

a) Identificación de clientes con consumo de bajo umbral

Una vez construidas las variables de segmentación y la matriz de datos limpia, es necesario separar los clientes con un bajo consumo (apáticos) para evitar que su consideración en la segmentación afecte a los segmentos de los clientes que efectivamente consumen.

Existen dos motivos para la no consideración de los clientes de bajo consumo en la segmentación por patrón de consumo:

i) Estadísticamente, para la identificación de patrones de comportamiento se necesita que los vectores que forman las variables de cada cliente sean sólidos y no contengan ruido. Esto se consigue sólo si el nivel de consumo que presenta cada cliente es suficientemente alto para que permita identificar una determinada pauta de consumo. Por ejemplo, un cliente que realiza 100 llamadas y estas son internacionales, es aceptable, con un nivel de confianza alto, pensar que la llamada 101 será también internacional. En cambio, un cliente que realiza 2 llamadas, una local y otra internacional, es complejo suponer cuál será el destino de la tercera llamada.

ii) Desde la perspectiva de negocio, se debe tener en cuenta que el objetivo de la segmentación es identificar diferentes pautas de consumo para el diseño de acciones comerciales. Estas acciones

comerciales pretenden fomentar uno o varios aspectos de ese consumo, por ejemplo incentivar las llamadas a larga distancia, las conexiones a Internet, etc. En el caso de los clientes de bajo consumo, el objetivo de negocio para ellos no es inducir un determinado aspecto de ese consumo, sino incentivar el consumo, cualquiera sea su destino.

Para ello, con el objetivo de identificar el umbral mínimo que determine un segmento de clientes apáticos, se desarrolla un análisis de sensibilidad con la variable *total llamadas* (tllamada).

a) Construcción de diferentes escenarios para la variable total de llamadas, calculando una curva de sensibilidad de 30, 45 y 60 llamadas sobre la variable total llamadas (tllamada).

b) De la tabla Tabla 3.1.5-1, se ha identificando como umbral mínimo óptimo aquellos clientes con 45 llamadas o menos (3.057 clientes o 5,38% de la muestra limpia). El umbral de 45 llamadas es una cantidad fácilmente manejable y explicable a la organización al representar una llamada y media por día (tanto saliente, entrante e Internet). A continuación se muestran los resultados obtenidos en SPSS:

Tabla 1.5-2 Escenarios de identificación del umbral de apáticos

Umbral de llamadas mínimas	Número de hogares	% sobre la muestra
Total llamadas <= 30	**1.761**	**3,10%**
Total llamadas <= 45	**3.057**	**5,38%**
Total llamadas <= 60	**4.711**	**8,29%**

c) De ahora en adelante, para la segmentación, los clientes con menor consumo a las 45 llamadas no son considerados y, para tal efecto, se

seleccionan a través del SPSS y su función Select Cases. De esa manera, se crea una nueva variable filter_$ que otorga el valor 1 a los clientes cuyo valor de total llamadas (tllamada) es mayor a 45 llamadas y 0 si el valor de tllamada es menor o igual a 45 llamadas.

b) Identificación de clientes internautas

Una vez identificados los clientes de bajo umbral, el siguiente ejercicio es identificar aquellos clientes con consumo de voz e Internet debido a su importancia específica para La Empresa de Telecomunicaciones. Este es un paso previo a la segmentación de patrón de consumo debido a que si no se hiciera previamente, estos clientes quedarían diluidos en segmentos de voz. Esto es así debido a que las variables que explican las llamadas de voz (17 variables) son mucho más numerosas que las variables de Internet (1 variable), por lo que las características de Internet del cliente quedarían diluidas en su patrón de consumo de voz.

Para la identificación de los clientes con consumo de voz e Internet, el consumo de Internet se considera respecto al total de las comunicaciones que realiza el cliente, es decir, siendo el 100% la suma de llamadas entrantes, salientes y conexiones a Internet a través del Dial Up.

Para ello se utiliza el algoritmo K-means no teniendo ya en cuenta aquellos clientes considerados como apáticos.

Existen múltiples métodos de análisis de grupos que se pueden agrupar en dos grandes familias: los Métodos no jerárquicos o partitivos y los métodos jerárquicos.

- Métodos no jerárquicos: estos algoritmos acostumbran a utilizarse en la clasificación de individuos, consumidores... en función de sus

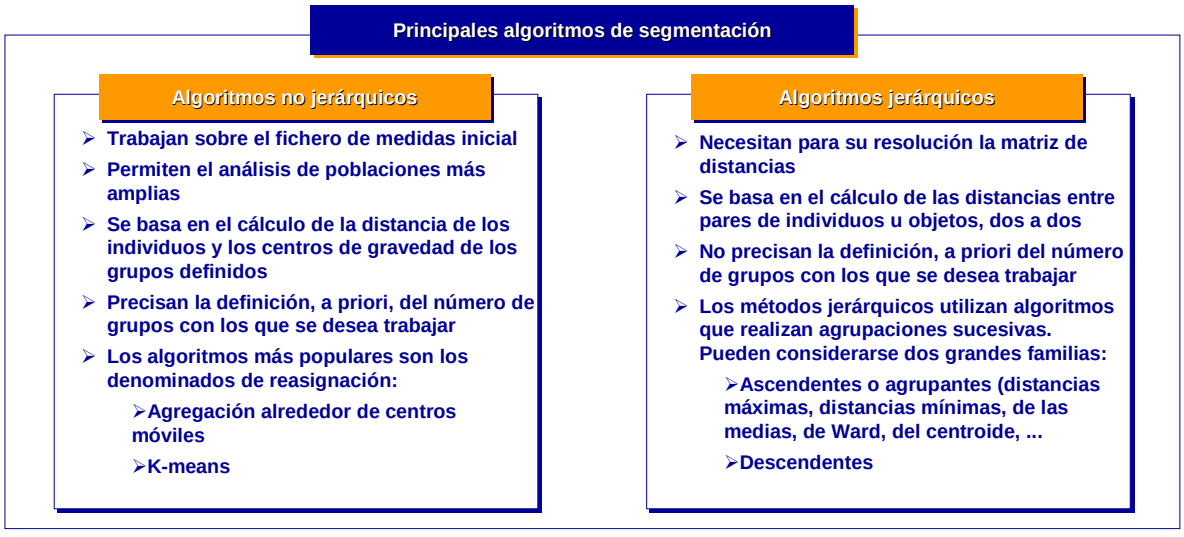
particularidades de comportamientos confesado o real, estilo de vida, características socio demográficas

- Métodos jerárquicos: estos son muy utilizados en la clasificación de marcas, productos,...en función de sus similitudes respecto a un conjunto de atributos o características

El criterio del investigador y su conocimiento del campo de estudio son fundamentales para decidir qué algoritmos de clasificación utilizar y con qué número final de segmentos o grupos quedarse. Una clasificación muy fina distinguirá mucho, pero proporcionará, generalmente, poca explicación. Contrariamente, pocos grupos, quizás no distinguirán lo suficiente.

En la segmentación de La Empresa de Telecomunicaciones se utilizo el algoritmo no jerárquico K-means, que construye clusters a partir de las distancias euclídeas entre los clientes. Las ventajas de utilizar este algoritmo no jerárquico las podemos ver el gráfico siguiente.

Gráfico 1.5-1 Comparativo de las ventajas de los algoritmos no jerárquicos y los jerárquicos.



a) Con el objetivo de identificar los potenciales segmentos, se desarrolla un análisis de clusters de 3, 4, 5, 6 y 7 grupos utilizando las variables % de llamadas salientes (pllasal), % de llamadas entrantes (pllaent) y % de conexiones a Internet (pllaint). Como metodología estándar de trabajo se realizan de 3 a 7 segmentos, aumentando el número en función de que a mayor número de segmentos se añade riqueza explicativa. Como mínimo siempre se realiza una segmentación más del número de segmentos que se intuyan.

Para cada uno de los grupos, se analiza el peso de cada una de las variables del centro de masas de cada segmento considerando las llamadas salientes, llamadas entrantes y conexiones a Internet. Esto se realiza haciendo foco en el comportamiento de cada segmento respecto al consumo de Internet para conocer aquellos individuos con mayor peso de la variable *Porcentaje de conexiones Internet* (pllaint).

A continuación se muestran en tablas los resultados de los análisis de clusters donde se puede observar la distribución de Total llamadas entre los destinos Internet (pllaint), voz saliente (pllasal) y voz entrante (pllaent) para cada uno de los grupos formados y el número de individuos que forman cada cluster.

Tabla 1.5-3 Outputs SPSS: 3 Clusters para la identificación de Internautas

Final Cluster Centers

	Cluster		
	1	2	3
PLLASAL	20.35	42.05	65.56
PLLAENT	79.56	56.69	32.24
PLLAINT	.09	1.26	2.20

Number of Cases in each Cluster

Cluster	1	19318.000
	2	24753.000
	3	9699.000
Valid		53770.000
Missing		.000

Tabla 1.5-4 Outputs SPSS: 4 Clusters para la identificación de Internautas

Final Cluster Centers

	Cluster			
	1	2	3	4
PLLASAL	42.11	35.80	66.47	20.12
PLLAENT	57.49	31.63	32.83	79.80
PLLAINT	.40	32.57	.71	.08

Number of Cases in each Cluster

Cluster	1	24337.000
	2	1117.000
	3	9395.000
	4	18921.000
Valid		53770.000
Missing		.000

Tabla 1.5-5 Outputs SPSS: 5 Clusters para la identificación de Internautas

Final Cluster Centers

	Cluster				
	1	2	3	4	5
PLLASAL	42.12	41.82	26.84	20.10	66.82
PLLAENT	57.64	38.31	20.59	79.82	32.70
PLLAINT	.24	19.87	52.57	.08	.48

Number of Cases in each Cluster

Cluster	1	24110.000
	2	1298.000
	3	319.000
	4	18883.000
	5	9160.000
Valid		53770.000
Missing		.000

Tabla 1.5-6 Outputs SPSS: 6 Clusters para la identificación de Internautas

Final Cluster Centers

	Cluster					
	1	2	3	4	5	6
PLLASAL	40.55	54.85	37.02	25.03	17.60	80.93
PLLAENT	37.30	44.72	62.76	19.73	82.33	18.47
PLLAINT	22.14	.43	.22	55.24	.06	.60

Number of Cases in each Cluster

Cluster	1	1180.000
	2	13444.000
	3	20998.000
	4	271.000
	5	14942.000
	6	2935.000
Valid		53770.000
Missing		.000

Tabla 1.5-7 Outputs SPSS: 7 Clusters para la identificación de Internautas

Final Cluster Centers

	Cluster						
	1	2	3	4	5	6	7
PLLASAL	17.53	25.07	35.98	53.02	81.21	36.91	54.70
PLLAENT	82.41	19.97	42.23	27.98	18.51	62.90	44.95
PLLAINT	.06	54.96	21.79	19.00	.28	.19	.35

Number of Cases in each Cluster

Cluster	1	14838.000
	2	275.000
	3	847.000
	4	538.000
	5	2886.000
	6	20869.000
	7	13517.000
Valid		53770.000
Missing		.000

b) Una vez identificados los clusters se analiza cómo se rompen los grupos al añadir un nivel más. De este modo se puede identificar aquellos grupos consistentes y los que no son permitiendo entender

cómo se van configurando los segmentos. En SPSS se utiliza la función de Crosstab. Se desarrolla Crosstabs entre los 3, 4, 5, 6 y 7 clusters anteriores para analizar cómo se "rompen" cada uno de los grupos identificados, con especial énfasis en aquellos grupos con mayor peso de la variable *Porcentaje de conexiones a Internet* (pllaint).

En las tablas se puede observar la cantidad de clientes que cambian de un segmento a otro así como su peso relativo dentro del segmento. Por ejemplo en la Tabla 1.5-7 podemos observar que 23.947 clientes del segmento 2 pasan el segmento 1 y que los 18.921 clientes que forman el segmento 4 (4 clusters) provienen del segmento 1.

Tabla 1.5-8 Outputs SPSS: Crosstab de 3 a 4 Clusters para identificación de Internautas

Primera corrida 3 * Primera corrida 4 Crosstabulation

			Primera corrida 4				
			1	2	3	4	Total
Primera corrida 3	1	Count	390	7		18921	19318
		% within Primera corrida 4	1.6%	.6%		100.0%	35.9%
		% of Total	.7%	.0%		35.2%	35.9%
	2	Count	23947	662	144		24753
		% within Primera corrida 4	98.4%	59.3%	1.5%		46.0%
		% of Total	44.5%	1.2%	.3%		46.0%
	3	Count		448	9251		9699
		% within Primera corrida 4		40.1%	98.5%		18.0%
		% of Total		.8%	17.2%		18.0%
Total		Count	24337	1117	9395	18921	53770
		% within Primera corrida 4	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	45.3%	2.1%	17.5%	35.2%	100.0%

Tabla 1.5-9 Outputs SPSS: Crosstab de 4 a 5 Clusters para identificación de Internautas

Primera corrida 4 * Primera corrida 5 Crosstabulation

			Primera corrida 5					
			1	2	3	4	5	Total
Primera corrida 4	1	Count	24013	324				24337
		% within Primera corrida 5	99.6%	25.0%				45.3%
		% of Total	44.7%	.6%				45.3%
	2	Count		796	319	1	1	1117
		% within Primera corrida 5		61.3%	100.0%	.0%	.0%	2.1%
		% of Total		1.5%	.6%	.0%	.0%	2.1%
	3	Count	58	178			9159	9395
		% within Primera corrida 5	.2%	13.7%			100.0%	17.5%
		% of Total	.1%	.3%			17.0%	17.5%
	4	Count	39			18882		18921
		% within Primera corrida 5	.2%			100.0%		35.2%
		% of Total	.1%			35.1%		35.2%
Total		Count	24110	1298	319	18883	9160	53770
		% within Primera corrida 5	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	44.8%	2.4%	.6%	35.1%	17.0%	100.0%

Tabla 1.5-10 Outputs SPSS: Crosstab de 5 a 6 Clusters para identificación de Internautas

Primera corrida 5 * Primera corrida 6 Crosstabulation

			Primera corrida 6						
			1	2	3	4	5	6	Total
Primera corrida 5	1	Count	13	7043	17054				24110
		% within Primera corrida 6	1.1%	52.4%	81.2%				44.8%
		% of Total	.0%	13.1%	31.7%				44.8%
	2	Count	1103	189	4			2	1298
		% within Primera corrida 6	93.5%	1.4%	.0%			.1%	2.4%
		% of Total	2.1%	.4%	.0%			.0%	2.4%
	3	Count	47			271		1	319
		% within Primera corrida 6	4.0%			100.0%		.0%	.6%
		% of Total	.1%			.5%		.0%	.6%
	4	Count	1		3940		14942		18883
		% within Primera corrida 6	.1%		18.8%		100.0%		35.1%
		% of Total	.0%		7.3%		27.8%		35.1%
	5	Count	16	6212				2932	9160
		% within Primera corrida 6	1.4%	46.2%				99.9%	17.0%
		% of Total	.0%	11.6%				5.5%	17.0%
Total		Count	1180	13444	20998	271	14942	2935	53770
		% within Primera corrida 6	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	2.2%	25.0%	39.1%	.5%	27.8%	5.5%	100.0%

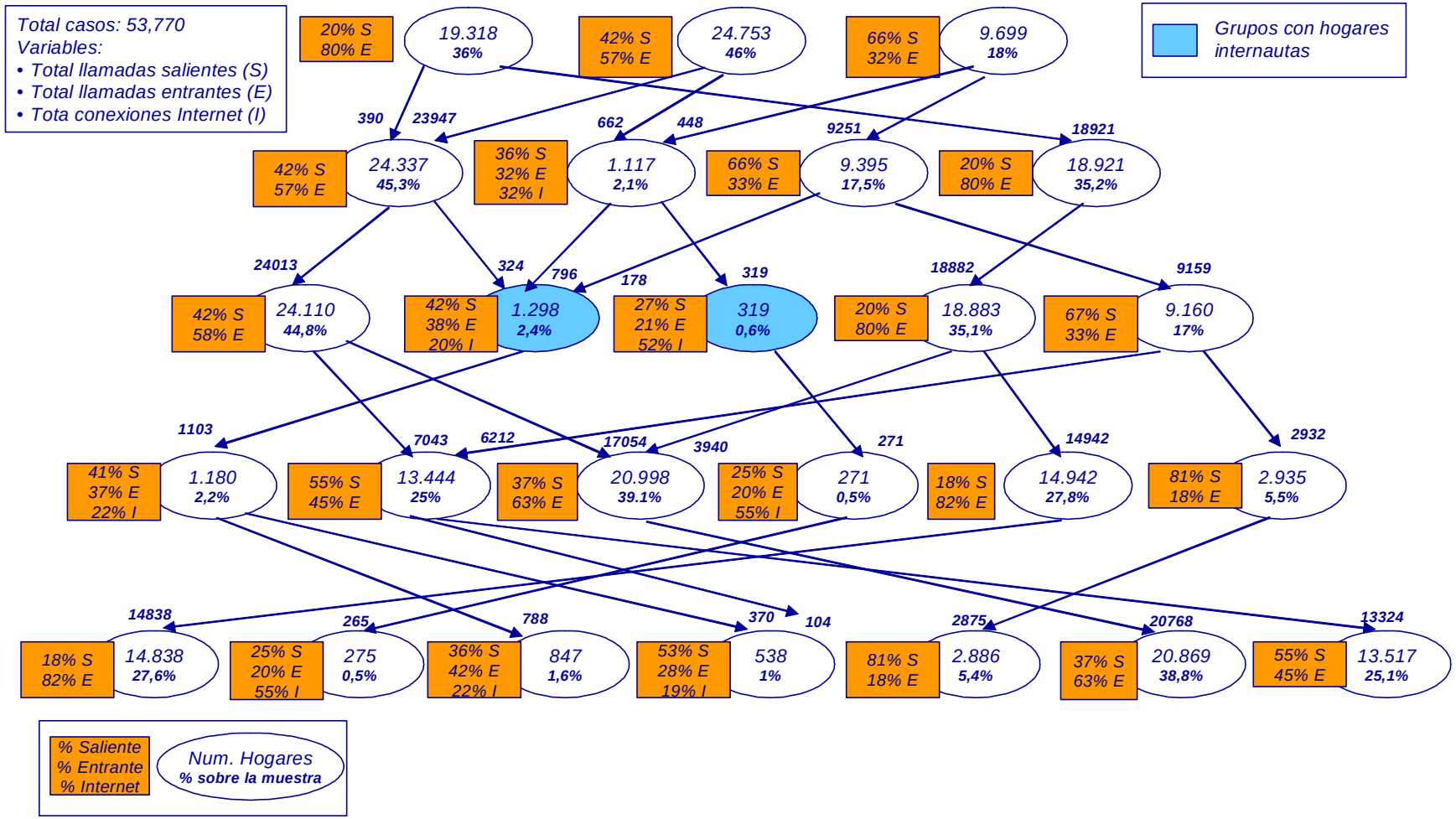
Tabla 1.5-11 Outputs SPSS: Crosstab de 6 a 7 Clusters para identificación de Internautas

Primera corrida 6 * Primera corrida 7 Crosstabulation

			Primera corrida 7							
			1	2	3	4	5	6	7	Total
Primera corrida 6	1	Count		10	788	370			12	1180
		% within Primera corrida 7		3.6%	93.0%	68.8%			.1%	2.2%
		% of Total		.0%	1.5%	.7%			.0%	2.2%
	2	Count			5	104	11		13324	13444
		% within Primera corrida 7			.6%	19.3%	.4%		98.6%	25.0%
		% of Total			.0%	.2%	.0%		24.8%	25.0%
	3	Count			49			20768	181	20998
		% within Primera corrida 7			5.8%			99.5%	1.3%	39.1%
		% of Total			.1%			38.6%	.3%	39.1%
	4	Count		265	2	4				271
		% within Primera corrida 7		96.4%	.2%	.7%				.5%
		% of Total		.5%	.0%	.0%				.5%
	5	Count	14838		3			101		14942
		% within Primera corrida 7	100.0%		.4%			.5%		27.8%
		% of Total	27.6%		.0%			.2%		27.8%
	6	Count				60	2875			2935
		% within Primera corrida 7				11.2%	99.6%			5.5%
		% of Total				.1%	5.3%			5.5%
Total		Count	14838	275	847	538	2886	20869	13517	53770
		% within Primera corrida 7	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	27.6%	.5%	1.6%	1.0%	5.4%	38.8%	25.1%	100.0%

A continuación se detalla el resumen de los resultados de los análisis realizados en SPSS.

Gráfico 1.5-2 Formación de los grupos para la identificación de los internautas



De las construcciones, finalmente, se decide seleccionar la segmentación que identifica 5 grupos, ya que con los cuatro segmentos se pierde algunos internautas que se diluyen en los grupos de voz y con seis segmentos se introducen clientes con bajo nivel de Internet.

Fueron seleccionados como clientes internautas los grupos 2 y 3 del análisis de 5 clusters debido a elevado % de la variable *Porcentaje de conexiones a internet* (pllaint), como se puede observar en el gráfico anterior.

a. Los grupos 2 (1.298 clientes) y 3 (319 clientes) son grupos relativamente sólidos en referencia a Internet, con un 20% y un 52%, respectivamente, sobre la variable *Porcentaje conexiones Internet* (pllaint). Para el caso concreto de los clientes del grupo 2, observamos que provienen de los grupos 1, 2 y 3 del análisis de 4 clusters.

   Se realiza un descriptivo que arroja la media de las variables *Porcentaje de conexiones Internet* (pllaint), *Número de conexiones a Internet* (v165) y *Tráfico de Internet* (v164), sus máximos y sus mínimos, para aquellos clientes que alimentan al grupo 2, con la finalidad de asegurar que estos clientes son efectivamente internautas. A continuación se muestran los resultados obtenidos en SPSS para decidir qué clientes serán finalmente considerados como internautas.

Tabla 1.5-12 Output SPSS: Clientes que pasan del grupo 1 (4 clusters) al grupo 2 (5 clusters)

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Porcentaje conexiones Internet	324	5.74	21.83	12.3496	3.32120
Total conexiones Internet	324	10.33	220.67	57.3539	33.96236
Tráfico internet (dial-up)	324	38.46	13410.08	1389.1984	1650.04419
Valid N (listwise)	324				

Tabla 1.5-13 Output SPSS: Clientes que pasan del grupo 3 (4 clusters) al grupo 2(5 clusters)

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Porcentaje conexiones Internet	178	6.01	24.09	12.5853	3.39147
Total conexiones Internet	178	7.67	240.67	58.3783	36.06569
Tráfico internet (dial-up)	178	22.89	12557.23	1388.4231	1650.87629
Valid N (listwise)	178				

Finalmente, y ya que la media de la variable *Porcentaje de conexiones a Internet* (pllaint) es relativamente alta, se consideran como internautas a todos los clientes del grupo 2 de 5 clusters. En total existen 1.617 clientes (2,84% de la muestra limpia considerando también los apáticos) considerados finalmente como internautas.

Una vez identificados los internautas se analiza si existe un patrón de consumo distinto dentro de este grupo. La segmentación se produce con las variables Número de llamadas cerradas (*voz saliente, voz entrante y conexiones a Internet*) y con las variables de momento del día y día de la semana (estos dos últimos bloques agrupados, es decir, sumarizadas las variables de voz entrante, saliente e Internet para cada caso), analizando si hay diferentes segmentos influenciados por el consumo de voz.

Para ello, se realiza el mismo ejercicio de creación de clusters y crosstab ya desarrollado anteriormente sólo para los 1.617 clientes. Se selecciona en SPSS estos clientes mediante la función Select Cases y se desarrolla un ejercicio de 2 y 3 clusters.

A continuación se muestran los resultados:

Tabla 1.5-14 Output SPSS: 2 clusters para los clientes internautas

Final Cluster Centers

	Cluster	
	1	2
Porcentaje llamadas salientes	26.75	41.77
Porcentaje llamadas entrantes	20.43	38.27
Porcentaje conexiones Internet	52.83	19.96
Porcentaje llamadas mañana	36.34	37.32
Porcentaje llamadas tarde	22.64	23.33
Porcentaje llamadas noche	34.47	33.21
Porcentaje llamadas madrugada	6.54	6.14
Porcentaje llamadas día laboral	65.62	70.00
Porcentaje llamadas feriado	1.94	1.77
Porcentaje llamadas domingo	16.70	13.31
Porcentaje llamadas sábado	15.74	14.92

Number of Cases in each Cluster

Cluster	1	313.000
	2	1304.000
Valid		1617.000
Missing		.000

Tabla 1.5-155 Output SPSS: 3 clusters para los clientes internautas

Final Cluster Centers

	Cluster		
	1	2	3
Porcentaje llamadas salientes	40.30	23.27	42.06
Porcentaje llamadas entrantes	32.37	18.58	40.66
Porcentaje conexiones Internet	27.34	58.16	17.28
Porcentaje llamadas mañana	31.48	37.79	40.56
Porcentaje llamadas tarde	22.01	22.87	24.04
Porcentaje llamadas noche	39.49	32.92	29.75
Porcentaje llamadas madrugada	7.02	6.42	5.65
Porcentaje llamadas día laboral	64.54	66.26	72.87
Porcentaje llamadas feriado	2.10	1.91	1.57
Porcentaje llamadas domingo	16.13	16.68	11.87
Porcentaje llamadas sábado	17.23	15.15	13.68

Number of Cases in each Cluster

Cluster	1	543.000
	2	224.000
	3	850.000
Valid		1617.000
Missing		.000

Tabla 1.5-16 Output SPSS crosstab entre 2 clusters y 3 clusters internautas

Cuarta corrida version 2 clus 2 * Cuarta corrida version 2 clus 3 Crosstabulation

			Cuarta corrida version 2 clus 3			
			1	2	3	Total
Cuarta corrida version 2 clus 2	1	Count	86	224	3	313
		% within Cuarta corrida version 2 clus 2	27.5%	71.6%	1.0%	100.0%
		% within Cuarta corrida version 2 clus 3	15.8%	100.0%	.4%	19.4%
		% of Total	5.3%	13.9%	.2%	19.4%
	2	Count	457		847	1304
		% within Cuarta corrida version 2 clus 2	35.0%		65.0%	100.0%
		% within Cuarta corrida version 2 clus 3	84.2%		99.6%	80.6%
		% of Total	28.3%		52.4%	80.6%
Total		Count	543	224	850	1617
		% within Cuarta corrida version 2 clus 2	33.6%	13.9%	52.6%	100.0%
		% within Cuarta corrida version 2 clus 3	100.0%	100.0%	100.0%	100.0%
		% of Total	33.6%	13.9%	52.6%	100.0%

Como se puede observar en las tablas anteriores, si se desarrolla un clúster de 2 grupos se aprecian diferentes comportamientos de uso, ya que el segmento 1 tiene un mayor peso de Internet que el segmento 2 (ver *Porcentaje conexiones Internet*). En cambio, al desarrollar un clúster de 3 grupos esta diferencia de comportamientos no es tan evidente como en el caso anterior (los segmentos 1 y 3 de 3 clusters presentan una menor diferencia).

Por lo tanto, se decide diferenciar al grupo de internautas en dos segmentos:

1.- Segmento 1 (313 clientes): al que se va a denominar "internautas" debido al alto peso de su comportamiento relativo a tráfico de Internet (*Porcentaje llamadas Internet*: 53%).

2.- Segmento 2 (1.304 clientes): al que se va a denominar "conectados" debido a su comportamiento más homogéneo que el segmento 1 entre el peso relativo de tráfico Internet (20%) y de voz (80%).

c) <u>Segmentación de los clientes sin considerar bajo consumo ni internautas</u>

   a. Una vez separados los clientes Internautas y los de bajo consumo a través de la función de SPSS "Select cases", se desarrolla un ejercicio de clusters para las variables principales de la segmentación: tipo de llamada, día y momento del día para los clientes no internautas. Se segmenta el resto de clientes definidos por patrones esencialmente de voz, utilizando la apertura de las variables por destino/origen, día de la semana y momento del día.

La apertura por franja horaria no es utilizada ya que está altamente correlacionada con el momento del día.

Las variables consideras son:

- % conexiones Internet
- % llamadas saliente a fijo local
- % llamadas salientes a móviles
- % llamadas salientes a movil LDN
- % llamadas salientes a fijo LDN
- % llamadas salientes a LDI
- % llamadas salientes a buzon de voz 159
- % llamadas salientes a buzon de voz 158
- % llamadas salientes a plataforma de atencion a clientes
- % llamadas salientes a # gratuitos
- % llamadas salientes a # de tarifa diferenciada
- % llamadas entrantes fijo local
- % llamadas entrantes movil fijo local
- % llamadas entrantes movil fijo LDN
- % llamadas entrantes fijo LDN
- % llamadas entrantes LDI
- % llamadas cobro revertido
- % llamadas entrantes teléfono público

- % llamadas en día laboral
- % llamadas día feriado
- % llamadas en sábado
- % llamadas en domingo
- % llamadas en mañana
- % llamadas en tarde
- % llamadas en noche
- % llamadas en madrugada

A continuación se muestran los resultados obtenidos. En cada uno de los clúster se observa para cada grupo la distribución de llamadas totales por tipo de llamada, por día de la semana y por momento del día.

Tabla 1.5-17 Output SPSS: 3 Clusters para la identificación de segmentos

Final Cluster Centers

	Cluster		
	1	2	3
Porcentaje conexiones Internet	.09	.41	.18
Porcentaje llamadas salientes a fijo local	14.03	41.68	21.97
Porcentaje llamadas salientes a movil local	1.32	4.49	2.35
Porcentaje llamadas salientes a movil LDN	.06	.18	.22
Porcentaje llamadas salientes a fijo LDN	.47	1.44	1.54
Porcentaje llamadas salientes LDI	.06	.37	.16
Porcentaje llamadas salientes a buzon de voz 159	2.98	4.08	8.40
Porcentaje llamadas salientes a buzon de voz 158	1.06	1.34	1.95
Porcentaje llamadas salientes a plat. atencion a clientes (102 103 104)	.66	1.29	1.48
Porcentaje llamadas salientes a # gratuitos	.38	.57	.90
Porcentaje llamadas salientes a # de tarifa diferenciada	.00	.00	.00
Porcentaje llamadas entrantes fijo local	50.69	30.67	32.17
Porcentaje llamadas entrantes movil fijo local	4.13	3.94	3.83
Porcentaje llamadas entrantes movil fijo LDN	.04	.05	.08
Porcentaje llamadas entrantes fijo LDN	3.98	2.74	7.25
Porcentaje llamadas entrantes LDI	.16	.19	.29
Porcentaje llamadas cobro revertido	.01	.01	.01
Porcentaje llamadas entrantes telf. público	19.89	6.54	17.21
Porcentaje llamadas mañana	36.50	39.16	36.05
Porcentaje llamadas tarde	23.22	23.80	22.49
Porcentaje llamadas noche	33.32	30.66	33.46
Porcentaje llamadas madrugada	6.96	6.38	8.00
Porcentaje llamadas día laboral	69.74	72.15	68.83
Porcentaje llamadas feriado	1.64	1.80	1.90
Porcentaje llamadas domingo	13.35	11.85	14.12
Porcentaje llamadas sábado	15.26	14.19	15.15

Number of Cases in each Cluster

Cluster	1	16490.000
	2	16410.000
	3	19253.000
Valid		52153.000
Missing		.000

Tabla 1.5-18 Output SPSS: 4 Clusters para la identificación de segmentos

Final Cluster Centers

	Cluster			
	1	2	3	4
Porcentaje conexiones Internet	.11	.48	.08	.27
Porcentaje llamadas salientes a fijo local	14.00	46.93	14.23	29.92
Porcentaje llamadas salientes a movil local	1.38	4.79	1.36	3.33
Porcentaje llamadas salientes a movil LDN	.05	.24	.18	.16
Porcentaje llamadas salientes a fijo LDN	.44	1.80	1.15	1.28
Porcentaje llamadas salientes LDI	.07	.44	.08	.23
Porcentaje llamadas salientes a buzon de voz 159	2.85	6.10	6.62	5.37
Porcentaje llamadas salientes a buzon de voz 158	1.11	1.62	1.49	1.58
Porcentaje llamadas salientes a plat. atencion a clientes (102 103 104)	.65	1.78	1.17	1.15
Porcentaje llamadas salientes a # gratuitos	.36	.82	.79	.58
Porcentaje llamadas salientes a # de tarifa diferenciada	.00	.01	.00	.00
Porcentaje llamadas entrantes fijo local	54.35	23.03	35.88	36.57
Porcentaje llamadas entrantes movil fijo local	4.44	3.46	3.13	4.49
Porcentaje llamadas entrantes movil fijo LDN	.04	.05	.06	.06
Porcentaje llamadas entrantes fijo LDN	3.54	2.94	8.02	4.07
Porcentaje llamadas entrantes LDI	.15	.19	.23	.25
Porcentaje llamadas cobro revertido	.01	.01	.01	.01
Porcentaje llamadas entrantes telf. público	16.46	5.30	25.51	10.66
Porcentaje llamadas mañana	39.63	39.90	33.93	36.93
Porcentaje llamadas tarde	24.18	23.53	21.79	23.34
Porcentaje llamadas noche	29.93	29.92	35.82	32.77
Porcentaje llamadas madrugada	6.27	6.65	8.47	6.96
Porcentaje llamadas día laboral	71.75	72.12	67.49	70.31
Porcentaje llamadas feriado	1.45	1.86	1.95	1.82
Porcentaje llamadas domingo	12.05	11.94	15.03	13.00
Porcentaje llamadas sábado	14.75	14.08	15.53	14.87

Number of Cases in each Cluster

Cluster	1	9984.000
	2	8642.000
	3	13391.000
	4	20136.000
Valid		52153.000
Missing		.000

Tabla 1.5-19 Output SPSS: 5 Clusters para la identificación de segmentos

Final Cluster Centers

	Cluster				
	1	2	3	4	5
Porcentaje conexiones Internet	.48	.29	.14	.07	.21
Porcentaje llamadas salientes a fijo local	47.33	26.40	20.75	10.31	28.57
Porcentaje llamadas salientes a movil local	4.98	3.36	1.96	1.05	2.74
Porcentaje llamadas salientes a movil LDN	.22	.18	.29	.08	.10
Porcentaje llamadas salientes a fijo LDN	1.69	1.40	1.84	.55	.86
Porcentaje llamadas salientes LDI	.45	.23	.19	.04	.16
Porcentaje llamadas salientes a buzon de voz 159	4.29	3.36	15.32	3.46	3.21
Porcentaje llamadas salientes a buzon de voz 158	1.36	1.29	2.90	1.06	1.22
Porcentaje llamadas salientes a plat. atencion a clientes (102 103 104)	1.51	1.00	2.26	.70	.84
Porcentaje llamadas salientes a # gratuitos	.65	.49	1.46	.45	.41
Porcentaje llamadas salientes a # de tarifa diferenciada	.00	.00	.00	.00	.00
Porcentaje llamadas entrantes fijo local	24.89	40.32	24.58	47.72	41.92
Porcentaje llamadas entrantes movil fijo local	3.59	4.22	3.56	3.63	4.58
Porcentaje llamadas entrantes movil fijo LDN	.05	.05	.11	.04	.05
Porcentaje llamadas entrantes fijo LDN	2.73	4.28	9.06	5.51	3.29
Porcentaje llamadas entrantes LDI	.19	.18	.37	.16	.22
Porcentaje llamadas cobro revertido	.01	.01	.02	.01	.01
Porcentaje llamadas entrantes telf. público	5.59	12.94	15.18	25.17	11.63
Porcentaje llamadas mañana	39.12	46.10	34.67	35.91	32.39
Porcentaje llamadas tarde	23.46	24.70	22.38	22.65	22.74
Porcentaje llamadas noche	30.70	23.41	34.54	33.77	37.69
Porcentaje llamadas madrugada	6.72	5.79	8.41	7.67	7.19
Porcentaje llamadas día laboral	71.69	75.60	68.05	68.80	67.95
Porcentaje llamadas feriado	1.91	1.50	1.91	1.72	1.91
Porcentaje llamadas domingo	12.16	9.67	14.85	14.01	14.43
Porcentaje llamadas sábado	14.24	13.23	15.19	15.46	15.71

Number of Cases in each Cluster

Cluster	1	8688.000
	2	9326.000
	3	7978.000
	4	12701.000
	5	13460.000
Valid		52153.000
Missing		.000

Tabla 1.5-20 Output SPSS: 6 Clusters para la identificación de segmentos

Final Cluster Centers

	Cluster					
	1	2	3	4	5	6
Porcentaje conexiones Internet	.15	.07	.51	.31	.24	.08
Porcentaje llamadas salientes a fijo local	22.23	13.07	49.41	28.20	31.04	12.33
Porcentaje llamadas salientes a movil local	1.76	1.32	5.09	3.68	3.14	1.13
Porcentaje llamadas salientes a movil LDN	.18	.17	.23	.21	.12	.04
Porcentaje llamadas salientes a fijo LDN	1.29	1.05	1.75	1.68	1.04	.36
Porcentaje llamadas salientes LDI	.19	.06	.46	.25	.20	.05
Porcentaje llamadas salientes a buzon de voz 159	22.37	4.40	3.95	3.45	3.34	2.97
Porcentaje llamadas salientes a buzon de voz 158	3.80	1.19	1.33	1.31	1.28	1.10
Porcentaje llamadas salientes a plat. atencion a clientes (102 103 104)	2.86	.95	1.55	1.08	.90	.61
Porcentaje llamadas salientes a # gratuitos	1.87	.62	.66	.50	.44	.35
Porcentaje llamadas salientes a # de tarifa diferenciada	.00	.00	.00	.00	.00	.00
Porcentaje llamadas entrantes fijo local	22.66	37.73	23.49	37.67	38.85	57.06
Porcentaje llamadas entrantes movil fijo local	4.02	2.94	3.47	4.16	4.48	4.73
Porcentaje llamadas entrantes movil fijo LDN	.11	.05	.05	.06	.05	.04
Porcentaje llamadas entrantes fijo LDN	6.31	8.03	2.71	4.60	3.52	3.44
Porcentaje llamadas entrantes LDI	.36	.19	.18	.21	.24	.16
Porcentaje llamadas cobro revertido	.02	.01	.01	.01	.01	.01
Porcentaje llamadas entrantes telf. público	9.81	28.13	5.14	12.61	11.11	15.53
Porcentaje llamadas mañana	34.61	35.29	39.59	45.86	32.32	37.02
Porcentaje llamadas tarde	22.89	22.02	23.51	24.52	22.56	23.84
Porcentaje llamadas noche	34.51	34.19	30.20	23.70	37.80	32.77
Porcentaje llamadas madrugada	8.00	8.49	6.69	5.92	7.32	6.37
Porcentaje llamadas día laboral	68.02	68.21	72.03	75.37	67.78	70.39
Porcentaje llamadas feriado	1.82	1.95	1.90	1.54	1.97	1.43
Porcentaje llamadas domingo	14.87	14.49	11.95	9.82	14.55	12.99
Porcentaje llamadas sábado	15.30	15.35	14.12	13.26	15.71	15.19

Number of Cases in each Cluster

Cluster	1	4710.000
	2	10991.000
	3	6808.000
	4	9213.000
	5	12938.000
	6	7493.000
Valid		52153.000
Missing		.000

Tabla 1.5-21 Output SPSS: 7 Clusters para la identificación de segmentos

Final Cluster Centers

	Cluster						
	1	2	3	4	5	6	7
Porcentaje conexiones Internet	.07	.15	.07	.29	.39	.59	.20
Porcentaje llamadas salientes a fijo local	11.03	22.13	12.23	26.00	40.60	57.48	26.75
Porcentaje llamadas salientes a movil local	1.05	1.81	1.25	3.33	4.60	5.43	2.55
Porcentaje llamadas salientes a movil LDN	.04	.19	.19	.20	.19	.25	.11
Porcentaje llamadas salientes a fijo LDN	.33	1.31	1.10	1.54	1.58	1.89	.90
Porcentaje llamadas salientes LDI	.05	.19	.06	.22	.40	.44	.14
Porcentaje llamadas salientes a buzon de voz 159	2.85	22.08	4.42	3.60	2.75	6.57	3.71
Porcentaje llamadas salientes a buzon de voz 158	1.05	3.73	1.17	1.37	1.05	1.95	1.36
Porcentaje llamadas salientes a plat. atencion a clientes (102 103 104)	.59	2.85	.95	1.07	1.05	2.35	.88
Porcentaje llamadas salientes a # gratuitos	.35	1.85	.63	.52	.46	1.03	.45
Porcentaje llamadas salientes a # de tarifa diferenciada	.00	.00	.00	.00	.00	.00	.00
Porcentaje llamadas entrantes fijo local	58.09	22.93	37.14	39.07	32.13	13.56	40.93
Porcentaje llamadas entrantes movil fijo local	4.66	3.99	2.56	4.19	3.85	3.24	4.79
Porcentaje llamadas entrantes movil fijo LDN	.04	.11	.04	.06	.05	.04	.06
Porcentaje llamadas entrantes fijo LDN	3.50	6.36	8.60	4.68	3.19	2.17	3.77
Porcentaje llamadas entrantes LDI	.16	.36	.17	.20	.21	.17	.25
Porcentaje llamadas cobro revertido	.01	.02	.01	.01	.01	.01	.01
Porcentaje llamadas entrantes telf. público	16.14	9.93	29.39	13.65	7.48	2.83	13.16
Porcentaje llamadas mañana	37.12	34.61	36.08	46.31	38.76	38.59	31.34
Porcentaje llamadas tarde	23.86	22.89	22.13	24.63	23.69	23.18	22.14
Porcentaje llamadas noche	32.64	34.47	33.17	23.13	30.99	30.97	39.14
Porcentaje llamadas madrugada	6.37	8.03	8.63	5.94	6.56	7.25	7.38
Porcentaje llamadas día laboral	70.43	68.02	68.59	75.49	71.84	71.16	66.96
Porcentaje llamadas feriado	1.41	1.82	1.95	1.52	1.84	2.02	1.95
Porcentaje llamadas domingo	12.96	14.87	14.26	9.71	12.00	12.58	15.13
Porcentaje llamadas sábado	15.20	15.29	15.20	13.28	14.33	14.25	15.96

Number of Cases in each Cluster

Cluster	1	6512.000
	2	4708.000
	3	9402.000
	4	7806.000
	5	9669.000
	6	2471.000
	7	11585.000
Valid		52153.000
Missing		.000

b. Posteriormente se desarrolla un ejercicio de crosstab entre los diferentes grupos obtenidos, para observar cómo se rompen los grupos. De este modo, es posible entender cómo se van configurando los segmentos. El resultado del ejercicio se muestra en el siguiente resultado de Crosstabs donde se puede observar cómo cada grupo se "parte" en la composición de un nivel más de cluster. Por ejemplo, en la *Tabla 1.5-21 Crosstabs de 3 a 4 clusters* podemos ver cómo los individuos del primer grupo de 3 clusters se reparten en 3 de los 4 grupos de 4 clusters, distribuyéndose un 57,6% en el uno, un 0% en el 2, un 33,3% en el 3 y un 9% en el 4. Adicionalmente, en la tabla se puede observar de qué grupos del nivel 3 cluster se conforman los grupos del nivel 4 clusters. Así, el grupo 2 de 4 Clusters es formado por el 96,5% proveniente del grupo 2, y 3,5% de clientes provienen del grupo 3.

Tabla 1.5-22 Output SPSS: Crosstabs de 3 a 4 clusters para la identificación de segmentos

Quinta corrida 2 ver 2 clus 3 * Quinta corrida 2 ver 2 clus 4 Crosstabulation

			Quinta corrida 2 ver 2 clus 4				
			1	2	3	4	Total
Quinta corrida 2 ver 2 clus 3	1	Count	9505		5493	1492	16490
		% within Quinta corrida 2 ver 2 clus 3	57.6%		33.3%	9.0%	100.0%
		% within Quinta corrida 2 ver 2 clus 4	95.2%		41.0%	7.4%	31.6%
		% of Total	18.2%		10.5%	2.9%	31.6%
	2	Count	116	8336		7958	16410
		% within Quinta corrida 2 ver 2 clus 3	.7%	50.8%		48.5%	100.0%
		% within Quinta corrida 2 ver 2 clus 4	1.2%	96.5%		39.5%	31.5%
		% of Total	.2%	16.0%		15.3%	31.5%
	3	Count	363	306	7898	10686	19253
		% within Quinta corrida 2 ver 2 clus 3	1.9%	1.6%	41.0%	55.5%	100.0%
		% within Quinta corrida 2 ver 2 clus 4	3.6%	3.5%	59.0%	53.1%	36.9%
		% of Total	.7%	.6%	15.1%	20.5%	36.9%
Total		Count	9984	8642	13391	20136	52153
		% within Quinta corrida 2 ver 2 clus 3	19.1%	16.6%	25.7%	38.6%	100.0%
		% within Quinta corrida 2 ver 2 clus 4	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	19.1%	16.6%	25.7%	38.6%	100.0%

Tabla 1.5-23 Output SPSS: Crosstabs de 4 a 5 clusters para la identificación de segmentos

Quinta corrida 2 ver 2 clus 4 * Quinta corrida 2 ver 2 clus 5 Crosstabulation

			Quinta corrida 2 ver 2 clus 5					
			1	2	3	4	5	Total
Quinta corrida 2 ver 2 clus 4	1	Count		2589	11	5630	1754	9984
		% within Quinta corrida 2 ver 2 clus 4		25.9%	.1%	56.4%	17.6%	100.0%
		% within Quinta corrida 2 ver 2 clus 5		27.8%	.1%	44.3%	13.0%	19.1%
		% of Total		5.0%	.0%	10.8%	3.4%	19.1%
	2	Count	7400	545	664		33	8642
		% within Quinta corrida 2 ver 2 clus 4	85.6%	6.3%	7.7%		.4%	100.0%
		% within Quinta corrida 2 ver 2 clus 5	85.2%	5.8%	8.3%		.2%	16.6%
		% of Total	14.2%	1.0%	1.3%		.1%	16.6%
	3	Count		567	3943	7070	1811	13391
		% within Quinta corrida 2 ver 2 clus 4		4.2%	29.4%	52.8%	13.5%	100.0%
		% within Quinta corrida 2 ver 2 clus 5		6.1%	49.4%	55.7%	13.5%	25.7%
		% of Total		1.1%	7.6%	13.6%	3.5%	25.7%
	4	Count	1288	5625	3360	1	9862	20136
		% within Quinta corrida 2 ver 2 clus 4	6.4%	27.9%	16.7%	.0%	49.0%	100.0%
		% within Quinta corrida 2 ver 2 clus 5	14.8%	60.3%	42.1%	.0%	73.3%	38.6%
		% of Total	2.5%	10.8%	6.4%	.0%	18.9%	38.6%
Total		Count	8688	9326	7978	12701	13460	52153
		% within Quinta corrida 2 ver 2 clus 4	16.7%	17.9%	15.3%	24.4%	25.8%	100.0%
		% within Quinta corrida 2 ver 2 clus 5	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	16.7%	17.9%	15.3%	24.4%	25.8%	100.0%

Tabla 1.5-24 Output SPSS: Crosstabs de 5 a 6 clusters para la identificación de segmentos

Quinta corrida 2 ver 2 clus 5 * Quinta corrida 2 ver 2 clus 6 Crosstabulation

			Quinta corrida 2 ver 2 clus 6						
			1	2	3	4	5	6	Total
Quinta corrida 2 ver 2 clus 5	1	Count	218		6781	558	1131		8688
		% within Quinta corrida 2 ver 2 clus 5	2.5%		78.1%	6.4%	13.0%		100.0%
		% within Quinta corrida 2 ver 2 clus 6	4.6%		99.6%	6.1%	8.7%		16.7%
		% of Total	.4%		13.0%	1.1%	2.2%		16.7%
	2	Count	33	339		8025	8	921	9326
		% within Quinta corrida 2 ver 2 clus 5	.4%	3.6%		86.0%	.1%	9.9%	100.0%
		% within Quinta corrida 2 ver 2 clus 6	.7%	3.1%		87.1%	.1%	12.3%	17.9%
		% of Total	.1%	.7%		15.4%	.0%	1.8%	17.9%
	3	Count	4416	2217	27	436	880	2	7978
		% within Quinta corrida 2 ver 2 clus 5	55.4%	27.8%	.3%	5.5%	11.0%	.0%	100.0%
		% within Quinta corrida 2 ver 2 clus 6	93.8%	20.2%	.4%	4.7%	6.8%	.0%	15.3%
		% of Total	8.5%	4.3%	.1%	.8%	1.7%	.0%	15.3%
	4	Count	9	7631		10		5051	12701
		% within Quinta corrida 2 ver 2 clus 5	.1%	60.1%		.1%		39.8%	100.0%
		% within Quinta corrida 2 ver 2 clus 6	.2%	69.4%		.1%		67.4%	24.4%
		% of Total	.0%	14.6%		.0%		9.7%	24.4%
	5	Count	34	804		184	10919	1519	13460
		% within Quinta corrida 2 ver 2 clus 5	.3%	6.0%		1.4%	81.1%	11.3%	100.0%
		% within Quinta corrida 2 ver 2 clus 6	.7%	7.3%		2.0%	84.4%	20.3%	25.8%
		% of Total	.1%	1.5%		.4%	20.9%	2.9%	25.8%
Total		Count	4710	10991	6808	9213	12938	7493	52153
		% within Quinta corrida 2 ver 2 clus 5	9.0%	21.1%	13.1%	17.7%	24.8%	14.4%	100.0%
		% within Quinta corrida 2 ver 2 clus 6	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	9.0%	21.1%	13.1%	17.7%	24.8%	14.4%	100.0%

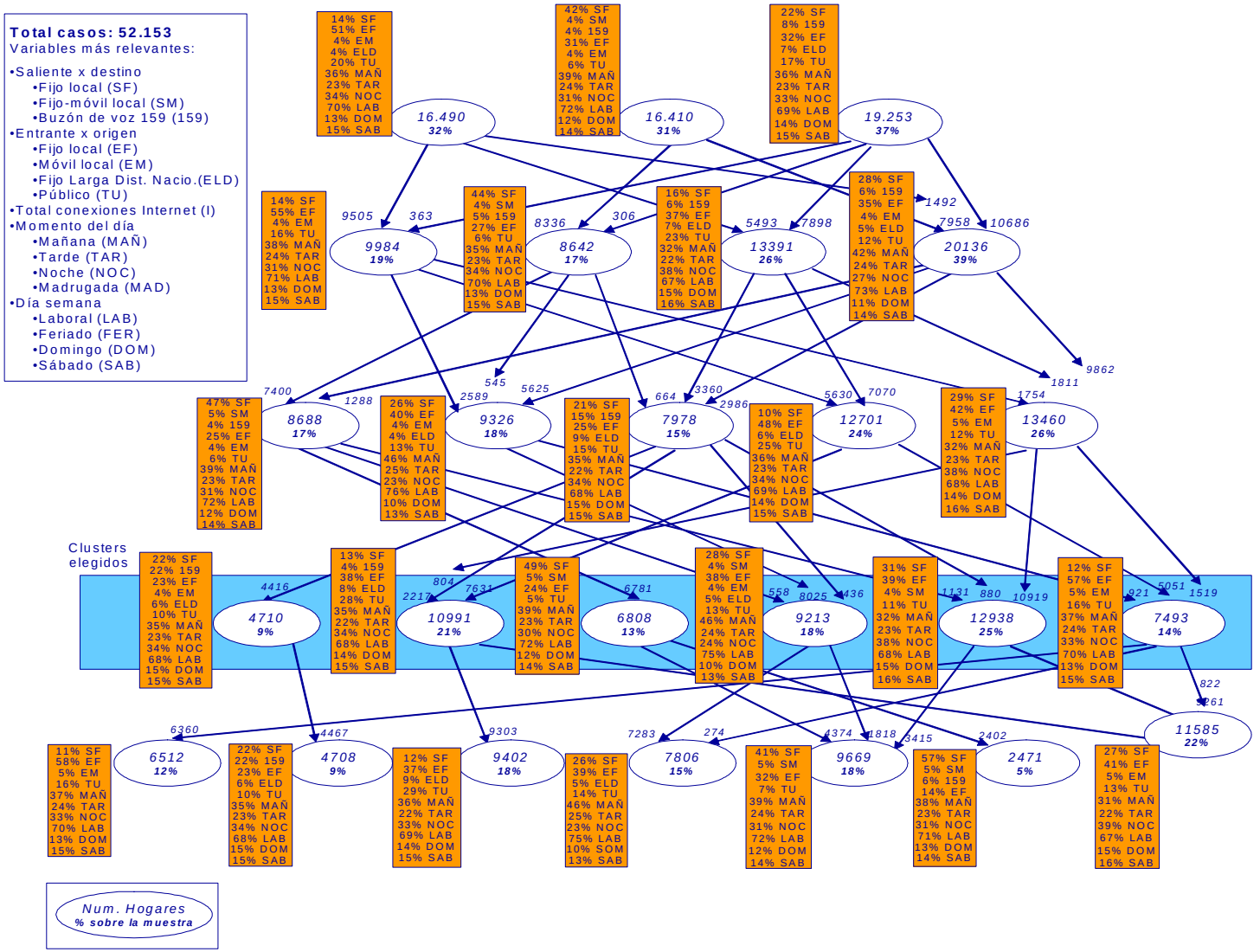
Tabla 1.5-25 Output SPSS: Crosstabs de 6 a 7 clusters para la identificación de segmentos

Quinta corrida 2 ver 2 clus 6 * Quinta corrida 2 ver 2 clus 7 Crosstabulation

			Quinta corrida 2 ver 2 clus 7							
			1	2	3	4	5	6	7	Total
Quinta corrida 2 ver 2 clus 6	1	Count		4467	22	8	61	67	85	4710
		% within Quinta corrida 2 ver 2 clus 6		94.8%	.5%	.2%	1.3%	1.4%	1.8%	100.0%
		% within Quinta corrida 2 ver 2 clus 7		94.9%	.2%	.1%	.6%	2.7%	.7%	9.0%
		% of Total		8.6%	.0%	.0%	.1%	.1%	.2%	9.0%
	2	Count	152	54	9303	101	1		1380	10991
		% within Quinta corrida 2 ver 2 clus 6	1.4%	.5%	84.6%	.9%	.0%		12.6%	100.0%
		% within Quinta corrida 2 ver 2 clus 7	2.3%	1.1%	98.9%	1.3%	.0%		11.9%	21.1%
		% of Total	.3%	.1%	17.8%	.2%	.0%		2.6%	21.1%
	3	Count		25		7	4374	2402		6808
		% within Quinta corrida 2 ver 2 clus 6		.4%		.1%	64.2%	35.3%		100.0%
		% within Quinta corrida 2 ver 2 clus 7		.5%		.1%	45.2%	97.2%		13.1%
		% of Total		.0%		.0%	8.4%	4.6%		13.1%
	4	Count		37	38	7283	1818		37	9213
		% within Quinta corrida 2 ver 2 clus 6		.4%	.4%	79.1%	19.7%		.4%	100.0%
		% within Quinta corrida 2 ver 2 clus 7		.8%	.4%	93.3%	18.8%		.3%	17.7%
		% of Total		.1%	.1%	14.0%	3.5%		.1%	17.7%
	5	Count		124	3	133	3415	2	9261	12938
		% within Quinta corrida 2 ver 2 clus 6		1.0%	.0%	1.0%	26.4%	.0%	71.6%	100.0%
		% within Quinta corrida 2 ver 2 clus 7		2.6%	.0%	1.7%	35.3%	.1%	79.9%	24.8%
		% of Total		.2%	.0%	.3%	6.5%	.0%	17.8%	24.8%
	6	Count	6360	1	36	274			822	7493
		% within Quinta corrida 2 ver 2 clus 6	84.9%	.0%	.5%	3.7%			11.0%	100.0%
		% within Quinta corrida 2 ver 2 clus 7	97.7%	.0%	.4%	3.5%			7.1%	14.4%
		% of Total	12.2%	.0%	.1%	.5%			1.6%	14.4%
Total		Count	6512	4708	9402	7806	9669	2471	11585	52153
		% within Quinta corrida 2 ver 2 clus 6	12.5%	9.0%	18.0%	15.0%	18.5%	4.7%	22.2%	100.0%
		% within Quinta corrida 2 ver 2 clus 7	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	12.5%	9.0%	18.0%	15.0%	18.5%	4.7%	22.2%	100.0%

Como resumen de los análisis de Clusters y Crosstabs se muestra el siguiente gráfico. En él se reflejan cada uno de los niveles de clusters con su descripción de las variables de segmentación y los movimientos de los clientes.

Gráfico 1.5-3 Composición de los clusters para la identificación de segmentos de voz



Una vez analizado en detalle cada uno de los diferentes clusters desde una perspectiva de negocio, se decide tomar la decisión de seleccionar a los 6 grupos del sexto nivel. La decisión de cuántos clúster resultan adecuados para la segmentación considera el criterio y la experiencia del investigador, pero sobre todo la facilidad con que cada uno de los clusters define un comportamiento diferencial y intuitivamente identificable.

Esto, en el caso de la segmentación de la Empresa de Telecomunicaciones, se operativiza de la siguiente manera:

- En el clúster de 6 grupos aparece un nuevo grupo con las siguientes características:
  - 12% llamadas a saliente fijo
  - 57% llamadas entrantes de fijo
  - 5% llamadas entrantes de móvil
  - 16% llamadas de teléfonos públicos
  - 37% llamadas durante la mañana
  - 24% llamadas durantes la tarde
  - 33% llamadas durante la noche
  - 70% llamadas durante días laborales
  - 13% llamadas durante días domingo
  - 15% llamadas durante días sábados

  Este segmento es fácilmente clasificable como el segmento “Racionales” o “Entrantes” debido a su alto peso de llamadas entrantes de fijo en día y horario laborable. Este segmento que en el clúster de 5 grupos no aparece

(o puede ser confundido con el segmento de “Públicos), es interesante comercialmente, puesto que es fácil diseñar acciones sobre él.

- En el clúster de 7 grupos, aparece un grupo nuevo con las siguientes características:
  - 57% llamadas a saliente fijo
  - 5% llamadas saliente a móvil
  - 6% llamadas saliente a buzón de voz 159
  - 14% llamadas entrantes de fijo
  - 38% llamadas durante la mañana
  - 23% llamadas durantes la tarde
  - 31% llamadas durante la noche
  - 71% llamadas durante días laborales
  - 13% llamadas durante días domingo
  - 14% llamadas durante días sábados

  Este segmento, que se alimenta casi enteramente del segmento “Próximos”, intensifica las características de ese segmento, pero contiene sólo a un 5% de la muestra. Sus características y la cantidad de clientes que contiene no lo hacen destacar sobre los otros segmentos para consideralo como un segmento sustancialmente diferente a los existentes. La utilización del clúster de 7 grupos no nos aporta valor comercial adicional respecto al clúster de 6 grupos, puesto que el nuevo grupo que aparece es una escisión de un grupo sólido de los de 6 clústers.

En cualquier caso, el hecho de seleccionar una segmentación de 6 grupos de voz no condiciona a que en un futuro no se pueda a elegir otro número de

segmentos. Si los valores de las variables de la segmentación varían o se le encuentra sentido comercial a uno o varios niveles más de segmentación, la segmentación debe ser modificada. En cualquier caso, la decisión deberá estar siempre guiada por el valor comercial que el número de segmentos identificados aporte, es decir, que todos los segmentos identificados sean comprensibles y con valor comercial para diseñar acciones específicas.

Por lo tanto, luego de seleccionar 6 clusters de voz, 2 segmentos de orientación hacia el Internet y 1 segmento de clientes apáticos o de bajo consumo, finalmente se configura una segmentación de 9 segmentos en total.

A continuación se ofrece el detalle de cada uno de los segmentos finales y la media de la muestra para cada una de las variables de segmentación.

Tabla 1.5-26Caracterización de los segmentos identificados por las variables de segmentación

		Clusters									
	Variables de segmentación	1	2	3	4	5	6	7	8	9	Muestra
Totales	% conexiones Internet	0.15%	0.07%	0.51%	0.31%	0.24%	0.08%	52.83%	19.96%	0.32%	0.97%
	% llamadas salientes	56.56%	22.84%	64.44%	40.38%	41.50%	18.95%	26.75%	41.77%	53.59%	39.31%
	% llamadas entrantes	43.30%	77.08%	35.04%	59.31%	58.26%	80.97%	20.43%	38.27%	46.09%	59.72%
Destino y origen de llamada	% conexiones Internet	0.15%	0.07%	0.51%	0.31%	0.24%	0.08%	52.83%	19.96%	0.32%	0.97%
	% llam sal fijo local	22.23%	13.07%	49.41%	28.20%	31.04%	12.33%	18.94%	30.45%	30.65%	26.01%
	% llam sal móvil local	1.76%	1.32%	5.09%	3.68%	3.14%	1.13%	2.63%	4.44%	2.07%	2.70%
	% llam sal móvil LDN	0.18%	0.17%	0.23%	0.21%	0.12%	0.04%	0.18%	0.22%	0.25%	0.16%
	% llam. fijo fijo LDN	1.29%	1.05%	1.75%	1.68%	1.04%	0.36%	1.31%	1.50%	2.37%	1.24%
	% llam. fijo fijo LDI	0.19%	0.06%	0.46%	0.25%	0.20%	0.05%	0.48%	0.42%	0.32%	0.21%
	% llam. sal. buzón voz 159	22.37%	4.40%	3.95%	3.45%	3.34%	2.97%	1.57%	2.44%	9.05%	5.44%
	% llam. sal. buzón voz 158	3.80%	1.19%	1.33%	1.31%	1.28%	1.10%	0.80%	1.08%	5.82%	1.70%
	% llam. sal. a Plat. Client.	2.86%	0.95%	1.55%	1.08%	0.90%	0.61%	0.59%	0.85%	1.74%	1.18%
	% llam. sal. a # gratuitos	1.87%	0.62%	0.66%	0.50%	0.44%	0.35%	0.25%	0.35%	1.31%	0.66%
	% llam. sal. a # tarifa difer.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%
	% llam. ent. fijo local	22.66%	37.73%	23.49%	37.67%	38.85%	57.06%	14.05%	26.80%	28.86%	36.71%
	% llam. ent. móvil fijo local	4.02%	2.94%	3.47%	4.16%	4.48%	4.73%	1.61%	3.55%	4.30%	3.96%
	% llam. ent. móvil fijo LDN	0.11%	0.05%	0.05%	0.06%	0.05%	0.04%	0.03%	0.03%	0.05%	0.05%
	% llam. ent. fijo LDN	6.31%	8.03%	2.71%	4.60%	3.52%	3.44%	1.54%	2.45%	6.24%	4.80%
	% llam. ent. fijo LDI	0.36%	0.19%	0.18%	0.21%	0.24%	0.16%	0.12%	0.15%	0.25%	0.22%
	% llam. cobro revertido	0.02%	0.01%	0.01%	0.01%	0.01%	0.01%	0.00%	0.00%	0.01%	0.01%
	% llam. ent. telf. público	9.81%	28.13%	5.14%	12.61%	11.11%	15.53%	3.07%	5.29%	6.39%	13.97%
		100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Día de la semana	% llamadas dia laboral	68.02%	68.21%	72.03%	75.37%	67.78%	70.39%	65.62%	70.00%	68.35%	70.04%
	% llamadas dia feriado	1.82%	1.95%	1.90%	1.54%	1.97%	1.43%	1.94%	1.77%	1.79%	1.79%
	% llamadas dia domingo	14.87%	14.49%	11.95%	9.82%	14.55%	12.99%	16.70%	13.31%	14.50%	13.26%
	% llamadas dia sábado	15.30%	15.35%	14.12%	13.26%	15.71%	15.19%	15.74%	14.92%	15.36%	14.91%
		100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Momento del día	% llamadas mañana	34.61%	35.29%	39.59%	45.86%	32.32%	37.02%	36.34%	37.32%	37.58%	37.19%
	% llamada tarde	22.89%	22.02%	23.51%	24.52%	22.56%	23.84%	22.64%	23.33%	23.71%	23.16%
	% llamadas noche	34.51%	34.19%	30.20%	23.70%	37.80%	32.77%	34.47%	33.21%	30.65%	32.46%
	% llamadas madrugada	8.00%	8.49%	6.69%	5.92%	7.32%	6.37%	6.54%	6.14%	8.06%	7.18%
		100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

d) Reasignación de segmentos considerando la posesión Speedy

Debido a la imposibilidad de la operadora, hoy por hoy, de medir el tráfico de datos, se incluyen dentro de los segmentos de voz, clientes que tienen contratada banda ancha (Speedy). En consecuencia, es necesario reasignar los clientes que tienen Speedy debido a su clara orientación al uso de Internet. El hecho de tener Speedy implica consumo de Internet, y en consecuencia, este cliente debería estar asignado a alguno de los segmentos que consumen Internet.

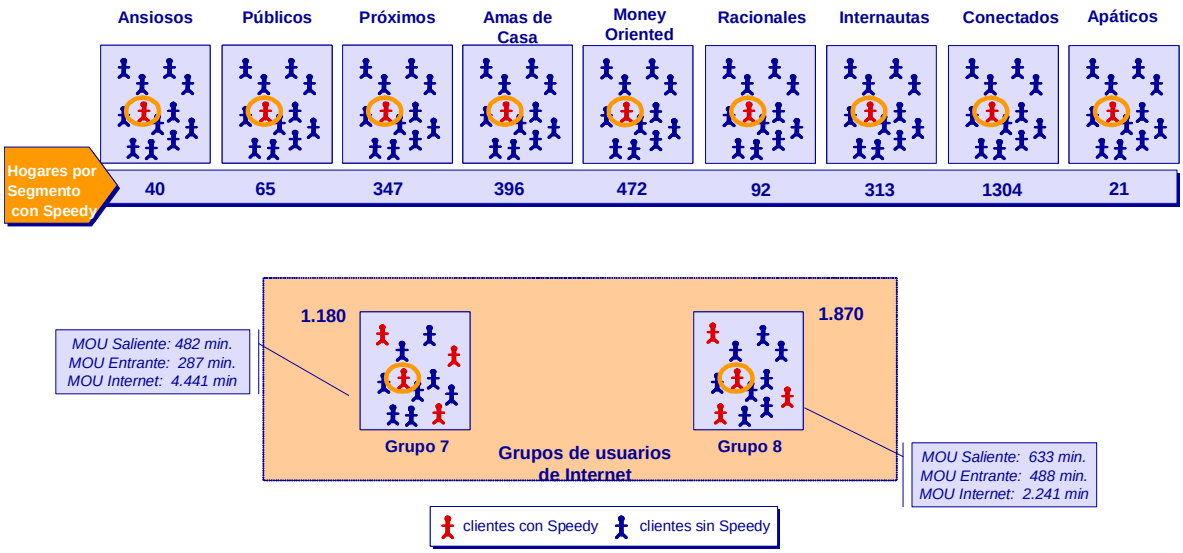
La identificación de esos clientes con tenencia de Speedy se realiza a posteriori de la segmentación por patrón de consumo de los grupos de voz para conocer la distribución por segmento de los clientes con Speedy y así, poder identificar la estrategia más conveniente.

En el caso que los clientes con Speedy se encontraran localizados todos en un mismo segmento, se habría identificado un nuevo segmento internauta. Si los clientes con Speedy se encontrasen repartidos por todos los segmentos, reflejando una alta penetración de este producto en cada uno de los segmentos, deberían ser clasificados en tantos segmentos de Internet como segmentos de voz de los que provienen. En el caso concreto de La Empresa de Telecomunicaciones, la penetración general es baja, por lo que es suficiente la reclasificación de estos individuos desde los segmentos originales de voz a los segmentos internautas hallados. Esto es, identificar a todos los clientes con Speedy y asignarlos en uno de los dos segmentos internautas: Internautas (segmento 7) o Conectados (segmento 8).

a) Para identificar los clientes con tenencia de Speedy se analiza la variable *Tenencia de Speedy* (v184s7), y se identifican aquellos

clientes donde esta variable es mayor que 0 (lo que indica la tenencia de este producto). Se crea una variable filtro (bananc) donde a los clientes con Speedy se les asigna el valor 1 y a los que no 0, identificándose los clientes que poseen Speedy.

Gráfico 1.5-4 Selección de clientes con tenencia de Speedy en los segmentos de voz



b) Una vez identificados los clientes con tenencia de Speedy, se asignan a los grupos 7 y 8 en función del consumo de voz de cada uno de ellos. El criterio de reasignación se basa en la comparación entre el consumo de voz de cada uno de los clientes con Speedy y la media de consumo de voz de los nueve grupos (1.120 minutos). La imposibilidad de obtener el tráfico de datos obliga a utilizar el tráfico de voz como variable para la reasignación:

   - Aquellos clientes con Speedy con consumo de voz menor a 1.120 minutos son asignados al grupo 7, “Internautas”, al suponer que el posible tráfico de Internet a través de Speedy representa un porcentaje alto sobre un MOU de voz bajo.

- Aquellos clientes con Speedy con consumo de voz mayor a 1.120 minutos son asignados al grupo 8, “Conectados”, por considerar que el tráfico de datos en porcentaje es menor al de voz debido al elevado MOU que los clientes representan.

Este proceso de identificación y reclasificación de los clientes entre segmentos se lleva a cabo para todos los grupos identificados (seis grupos de voz, un grupo de clientes internautas, los conectados, y el grupo de clientes apáticos), excepto para los clientes del grupo 7 (internautas), que se mantienen en el mismo grupo (para realizar esta operación en SPSS ver Anexo *Resumen Sintaxis SPSS*).

El proceso de asignación de clientes con Speedy de los segmentos de voz a los segmentos de internautas descrito puede ilustrarse de la siguiente manera.

Gráfico 1.5-5Reclasificación de los clientes con Speedy en los segmentos internautas

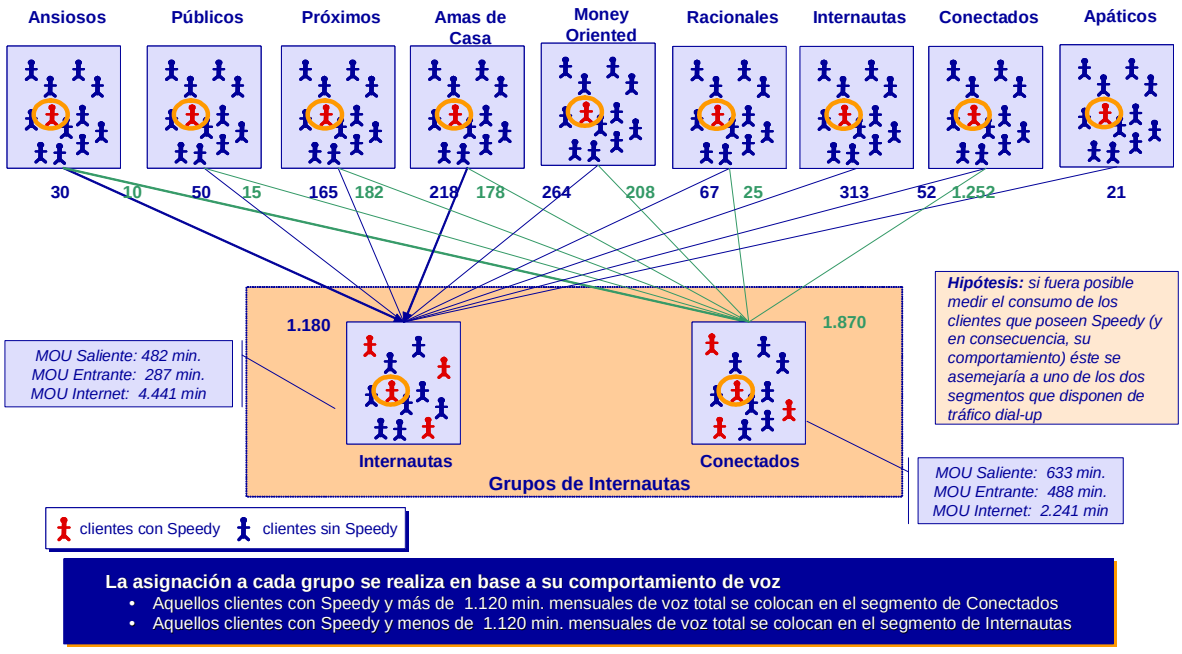


Tabla 1.5-27 Output SPSS: Crosstab entre segmentos antes de la reclasificación de clientes de Speedy y los segmentos después de la reclasificación de clientes con Speedy.

GRPSNEW * Grupos finales de uso Crosstabulation

			Grupos finales de uso									
			1.00	2.00	3.00	4.00	5.00	6.00	10.00	20.00	99.00	Total
Grupos antes de la reclasificación de Speedy	1.00	Count	4670						30	10		4710
		% within GRPSNEW	99.2%						.6%	.2%		100.0%
		% within Grupos finales de uso	100.0%						2.5%	.5%		8.3%
		% of Total	8.2%						.1%	.0%		8.3%
	2.00	Count		10926					50	15		10991
		% within GRPSNEW		99.4%					.5%	.1%		100.0%
		% within Grupos finales de uso		100.0%					4.2%	.8%		19.3%
		% of Total		19.2%					.1%	.0%		19.3%
	3.00	Count			6461				165	182		6808
		% within GRPSNEW			94.9%				2.4%	2.7%		100.0%
		% within Grupos finales de uso			100.0%				14.0%	9.7%		12.0%
		% of Total			11.4%				.3%	.3%		12.0%
	4.00	Count				8817			218	178		9213
		% within GRPSNEW				95.7%			2.4%	1.9%		100.0%
		% within Grupos finales de uso				100.0%			18.5%	9.5%		16.2%
		% of Total				15.5%			.4%	.3%		16.2%
	5.00	Count					12466		264	208		12938
		% within GRPSNEW					96.4%		2.0%	1.6%		100.0%
		% within Grupos finales de uso					100.0%		22.4%	11.1%		22.8%
		% of Total					21.9%		.5%	.4%		22.8%
	6.00	Count						7401	67	25		7493
		% within GRPSNEW						98.8%	.9%	.3%		100.0%
		% within Grupos finales de uso						100.0%	5.7%	1.3%		13.2%
		% of Total						13.0%	.1%	.0%		13.2%
	10.00	Count							313			313
		% within GRPSNEW							100.0%			100.0%
		% within Grupos finales de uso							26.5%			.6%
		% of Total							.6%			.6%
	20.00	Count							52	1252		1304
		% within GRPSNEW							4.0%	96.0%		100.0%
		% within Grupos finales de uso							4.4%	67.0%		2.3%
		% of Total							.1%	2.2%		2.3%
	99.00	Count							21		3036	3057
		% within GRPSNEW							.7%		99.3%	100.0%
		% within Grupos finales de uso							1.8%		100.0%	5.4%
		% of Total							.0%		5.3%	5.4%
Total		Count	4670	10926	6461	8817	12466	7401	1180	1870	3036	56827
		% within GRPSNEW	8.2%	19.2%	11.4%	15.5%	21.9%	13.0%	2.1%	3.3%	5.3%	100.0%
		% within Grupos finales de uso	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	8.2%	19.2%	11.4%	15.5%	21.9%	13.0%	2.1%	3.3%	5.3%	100.0%

c) Tras el proceso de asignación de clientes de Speedy, se vuelve a calcular el peso de las variables de segmentación en los segmentos de voz, con un descriptivo que es mostrado a continuación. Cabe señalar

que las características de los grupos internautas (conectados e internautas) no varían ya que no es posible medir el tráfico de datos y asumimos que se asemejará al de sus colegas dial up del segmento.

Tabla 1.5-28 Caracterización de los segmentos finales de las variables de segmentación

		Clusters									
	Variables de segmentación	1	2	3	4	5	6	7	8	9	Muestra
Totales	% conexiones Internet	0.14%	0.07%	0.52%	0.30%	0.24%	0.08%	52.83%	19.96%	0.30%	0.97%
	% llamadas salientes	56.54%	22.82%	64.47%	40.20%	41.38%	18.94%	26.75%	41.77%	53.56%	39.31%
	% llamadas entrantes	43.32%	77.11%	35.02%	59.50%	58.38%	80.98%	20.43%	38.27%	46.14%	59.72%
Destino y origen de llamada	% conexiones Internet	0.14%	0.07%	0.52%	0.30%	0.24%	0.08%	52.83%	19.96%	0.30%	0.97%
	% llam sal fijo local	22.23%	13.05%	49.56%	28.12%	31.03%	12.31%	18.94%	30.45%	30.63%	26.01%
	% llam sal móvil local	1.69%	1.31%	4.87%	3.52%	3.01%	1.12%	2.63%	4.44%	2.05%	2.70%
	% llam sal móvil LDN	0.17%	0.17%	0.23%	0.21%	0.12%	0.04%	0.18%	0.22%	0.25%	0.16%
	% llam. fijo fijo LDN	1.27%	1.04%	1.77%	1.68%	1.04%	0.35%	1.31%	1.50%	2.33%	1.24%
	% llam. fijo fijo LDI	0.19%	0.06%	0.43%	0.24%	0.19%	0.05%	0.48%	0.42%	0.31%	0.21%
	% llam. sal. buzón voz 159	22.45%	4.41%	4.02%	3.50%	3.37%	2.99%	1.57%	2.44%	9.10%	5.44%
	% llam. sal. buzón voz 158	3.78%	1.20%	1.34%	1.33%	1.28%	1.11%	0.80%	1.08%	5.86%	1.70%
	% llam. sal. a Plat. Client.	2.87%	0.95%	1.57%	1.08%	0.90%	0.61%	0.59%	0.85%	1.73%	1.18%
	% llam. sal. a # gratuitos	1.88%	0.62%	0.67%	0.51%	0.44%	0.35%	0.25%	0.35%	1.30%	0.66%
	% llam. sal. a # tarifa difer.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%
	% llam. ent. fijo local	22.67%	37.74%	23.41%	37.70%	38.91%	57.09%	14.05%	26.80%	28.82%	36.71%
	% llam. ent. móvil fijo local	4.00%	2.94%	3.38%	4.06%	4.42%	4.69%	1.61%	3.55%	4.33%	3.96%
	% llam. ent. móvil fijo LDN	0.11%	0.05%	0.05%	0.06%	0.05%	0.04%	0.03%	0.03%	0.05%	0.05%
	% llam. ent. fijo LDN	6.32%	8.03%	2.77%	4.67%	3.55%	3.45%	1.54%	2.45%	6.27%	4.80%
	% llam. ent. fijo LDI	0.36%	0.19%	0.18%	0.21%	0.24%	0.16%	0.12%	0.15%	0.24%	0.22%
	% llam. cobro revertido	0.02%	0.01%	0.01%	0.01%	0.01%	0.01%	0.00%	0.00%	0.01%	0.01%
	% llam. ent. telf. público	9.83%	28.15%	5.22%	12.80%	11.21%	15.55%	3.07%	5.29%	6.42%	13.97%
		100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	100%	100%
Día de la semana	% llamadas dia laboral	68.00%	68.20%	71.88%	75.24%	67.70%	70.37%	65.62%	70.00%	68.31%	70.04%
	% llamadas dia feriado	1.82%	1.95%	1.91%	1.55%	1.98%	1.43%	1.94%	1.77%	1.79%	1.79%
	% llamadas dia domingo	14.88%	14.49%	12.05%	9.90%	14.61%	13.00%	16.70%	13.31%	14.52%	13.26%
	% llamadas dia sábado	15.30%	15.36%	14.15%	13.31%	15.71%	15.20%	15.74%	14.92%	15.38%	14.91%
		100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	100%	100%
Momento del día	% llamadas mañana	34.61%	35.30%	39.61%	45.85%	32.29%	37.01%	36.34%	37.32%	37.53%	37.19%
	% llamada tarde	22.88%	22.01%	23.44%	24.47%	22.51%	23.83%	22.64%	23.33%	23.69%	23.16%
	% llamadas noche	34.51%	34.19%	30.17%	23.67%	37.82%	32.77%	34.47%	33.21%	30.69%	32.46%
	% llamadas madrugada	8.00%	8.50%	6.78%	6.01%	7.37%	6.39%	6.54%	6.14%	8.10%	7.18%
		100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Sombreadas se destacan, en la tabla anterior, las variables que mejor caracterizan cada segmento.

3.1.6 Discriminante, calidad de la segmentación

Una vez realizada la segmentación por patrón de comportamiento se comprueba la bondad del proceso de asignación de clientes a los diferentes segmentos. Para ello se utiliza la función del discriminante que:

- permite constatar si los resultados obtenidos mediante técnicas de análisis tipológico se reproducen en función de las variables de formación más discriminantes, con un análisis discriminante, y
- determina las variables que tienen mayor influencia en la formación de los grupos.

Una de las funciones del Análisis Discriminante es que pretende *explicar* la pertenencia de un cliente a un segmento dado en función de las Variables Independientes (de segmentación disponibles). Esto es, crea un modelo explicativo (expresado por la ecuación de "función discriminante") compuesto por las Variables Independientes (variables de segmentación) capaces de discriminar de forma significativa entre los casos pertenecientes a las categorías de la Variables Dependientes (Segmento al que corresponden).

El Análisis Discriminante se realiza antes de la reasignación de los clientes con tenencia de Speedy ya que lo que se está midiendo es la bondad de la segmentación. El segmento de “Apáticos” no se analiza debido a que este

segmento se crea ad hoc no siguiendo un proceso de segmentación no jerárquico.

Del análisis discriminante se obtienen los siguientes outputs:

a) Funciones discriminantes: Se obtienen (n-1) funciones lineales (8 segmentos – 1 = 7) dado que la octava función discriminante es linealmente dependiente.

Tabla 1.6-1Matriz de coeficientes de las Funciones discriminantes

Standardized Canonical Discriminant Function Coefficients

	Function						
	1	2	3	4	5	6	7
Porcentaje conexiones Internet	1.000	-.098	.066	.229	.011	.037	.024
Porcentaje llamadas salientes fijo fijo local	.216	1.017	.391	.623	-.187	.003	.159
Porcentaje llamadas salientes fijo movil local	.019	.253	.114	.401	.025	.494	-.489
Porcentaje llamadas salientes fijo movil LDN	.003	.017	.028	.097	.005	-.070	-.210
Porcentaje llamadas salientes fijo fijo LDN	.020	.146	.079	.379	.083	.420	.215
Porcentaje llamadas salientes fijo fijo LDI	.006	.067	.005	.080	-.008	-.217	-.231
Porcentaje llamadas salientes a buzon de voz 159	.038	.295	-.525	1.070	.138	.059	-.040
Porcentaje llamadas salientes a buzon de voz 158	.020	.103	.072	.233	.042	.238	.124
Porcentaje llamadas salientes a plat. atencion a clientes (102 103 104)	.015	.087	.035	.285	.024	-.188	-.041
Porcentaje llamadas salientes a # gratuitos	.016	.086	.009	.221	-.002	-.038	-.003
Porcentaje llamadas salientes a # de tarifa diferenciada	-.004	.001	-.001	.010	-.004	.066	-.147
Porcentaje llamadas entrantes fijo local	-.003	.012	.721	1.350	-.063	.143	.090
Porcentaje llamadas entrantes movil fijo local	.034	.247	.116	.612	.006	.478	.185
Porcentaje llamadas entrantes movil fijo LDN	.007	.035	-.020	.064	.019	.027	.151
Porcentaje llamadas entrantes fijo LDN	.033	.159	.021	.539	.008	.406	.003
Porcentaje llamadas entrantes fijo LDI	.009	.056	-.011	.102	.025	.192	.302
Porcentaje llamadas cobro revertido	.003	.009	.000	.028	.005	.028	.054
Porcentaje llamadas mañana	.015	.087	.167	-.088	.501	.024	-.121
Porcentaje llamadas tarde	-.002	.016	.080	.007	.052	-.013	-.169
Porcentaje llamadas noche	-.011	-.035	-.009	-.019	-.393	.155	-.504
Porcentaje llamadas día laboral	-.004	.034	.046	-.032	.301	-.008	-.018
Porcentaje llamadas feriado	.003	.001	-.013	.024	-.003	.362	.386
Porcentaje llamadas domingo	.000	-.031	-.040	-.019	-.060	-.206	.341

Con esta matriz de coeficientes se pueden construir cada una de las 7 funciones discriminantes que se generan.

Como la fórmula de las funciones es F1= $a_{11}Var_1 + a_{21} Var_2 + a_{31} Var_3 +.....+ a_{i1} V_i$ se puede construir la primera función discriminante, siendo esta:

F1= *pllaint* + 0,216*psfijloc* + 0,019*psmovloc* + 0,003*psmovldn* + 0,020*psfijldn* + 0,006*psfijldi* + 0,038*psbuz159* + 0,020*psbuz158* + 0,015*psclient* + 0,016*psgratui* – 0,004*psdifere* – 0,003*pefijloc* + 0,034*pemovloc* + 0,007*pemovldn* + 0,033*pefijldn* + 0,009*peldi* + 0,003*pecobrev* + 0,015*pllamaña* – 0,002*pllatard* – 0,011*pllanoch* – 0,004*pllalabo* + 0,003*pllaferi*

Las tres primeras funciones discriminantes explican el 89,5% de la varianza total como se muestra en la tabla Enginevalues

Tabla 1.6-2Matriz de Valores Propios

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	5.653[a]	52.9	52.9	.922
2	2.719[a]	25.5	78.4	.855
3	1.193[a]	11.2	89.5	.738
4	.613[a]	5.7	95.3	.617
5	.496[a]	4.6	99.9	.576
6	.007[a]	.1	100.0	.085
7	.001[a]	.0	100.0	.033

a. First 7 canonical discriminant functions were used in the analysis.

Las principales funciones discriminantes vienen explicadas principalmente por las variables marcadas con asterisco en la matriz de estructura.

Tabla 1.6-3 Output SPSS Matriz de estructura

Structure Matrix

	Function						
	1	2	3	4	5	6	7
Porcentaje conexiones Internet	.979*	-.199	-.009	.015	.016	.020	-.029
Porcentaje llamadas salientes fijo fijo local	.106	.874*	.239	-.138	-.224	-.220	.105
Porcentaje llamadas salientes a buzon de voz 159	-.034	.067	-.754*	.490	.151	-.247	-.012
Porcentaje llamadas entrantes fijo local	-.176	-.486	.660*	.425	-.074	-.223	.048
Porcentaje llamadas salientes a buzon de voz 158	-.011	.038	-.235*	.179	.053	.126	.117
Porcentaje llamadas entrantes telf. público [a]	-.158	-.514	-.172	-.756*	.038	-.197	.020
Porcentaje llamadas noche	.000	-.085	-.141	.078	-.818*	.165	-.225
Porcentaje llamadas mañana	.009	.090	.171	-.077	.806*	-.069	.152
Porcentaje llamadas día laboral	-.008	.092	.172	-.008	.521*	.059	-.385
Porcentaje llamadas domingo	.011	-.087	-.180	.010	-.459*	-.146	.413
Porcentaje llamadas sábado [a]	-.001	-.061	-.058	.039	-.289*	-.021	.028
Porcentaje llamadas tarde	.002	.027	.066	.079	.157*	-.077	-.093
Porcentaje llamadas entrantes movil fijo local	-.014	.006	.048	.125	-.018	.400*	.066
Porcentaje llamadas salientes fijo fijo LDN	.015	.074	-.024	-.079	.100	.384*	.032
Porcentaje llamadas entrantes fijo LDN	-.047	-.112	-.202	-.224	.089	.326*	.099
Porcentaje llamadas salientes a plat. atencion a clientes (102 103 104)	-.005	.078	-.204	.088	.071	-.316*	-.036
Porcentaje llamadas salientes a # gratuitos	-.009	.028	-.184	.076	.045	-.192*	-.020
Porcentaje llamadas salientes fijo movil local	.039	.183	.092	-.062	.039	.469	-.576*
Porcentaje llamadas feriado	.005	.024	-.061	-.105	-.157	.273	.391*
Porcentaje llamadas entrantes fijo LDI	-.005	.010	-.040	.031	-.005	.276	.299*
Porcentaje llamadas salientes fijo fijo LDI	.027	.078	.012	-.012	.021	-.120	-.276*
Porcentaje llamadas madrugada [a]	-.016	-.037	-.118	-.100	-.130	-.063	.234*
Porcentaje llamadas salientes fijo movil LDN	.008	.031	-.026	-.061	.055	.106	-.223*
Porcentaje llamadas entrantes movil fijo LDN	-.003	.009	-.035	.024	.012	.099	.180*
Porcentaje llamadas salientes a # de tarifa diferenciada	.002	.005	-.003	-.002	.006	.068	-.151*
Porcentaje llamadas cobro revertido	-.003	.000	-.029	.005	.010	.019	.050*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

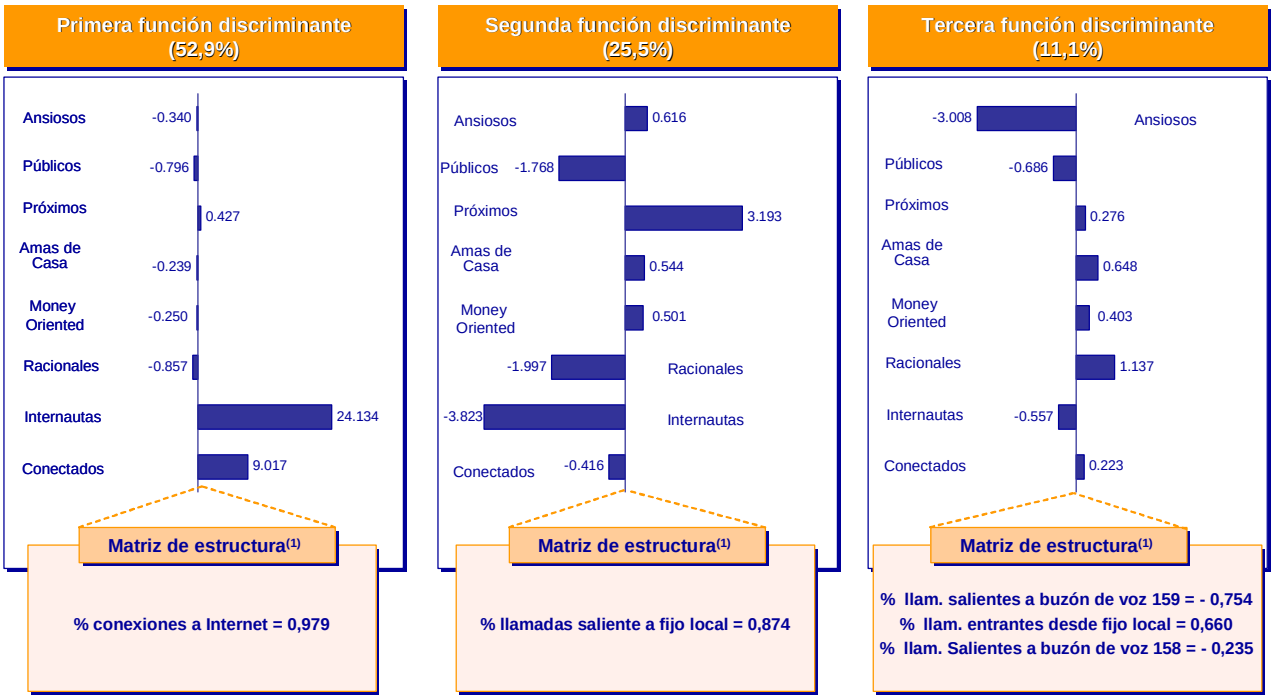
*. Largest absolute correlation between each variable and any discriminant function

a. This variable not used in the analysis.

Tal y como se observa en la tabla anterior, la Función discriminante 1 es explicada principalmente por la variable *Porcentaje de conexiones a Internet* (pllaint) y la función discriminante 2 está explicada principalmente por la variable *Porcentaje de llamadas a fijo local* (psfijloc).

En resumen, el discriminante explica un 92,1 % de los casos, mostrándose las tres primeras funciones discriminantes.

Gráfico 1.6-1 Bondad de la segmentación de la Empresa de Telecomunicaciones mostrada por el discriminante



(1) Correlaciones intra-grupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas

Estas funciones discriminantes permiten reclasificar los individuos asignados a los segmentos proporcionando una medición de la bondad de la segmentación.

A continuación se muestra la matriz de confusión donde se comparan los segmentos otorgados por la segmentación y los segmentos otorgados por el discriminante para poder comparar la calidad de la segmentación.

Tabla 1.6-4 Matriz de confusión

Classification Results[b,c]

			Predicted Group Membership								
		DICRIMIN	1	2	3	4	5	6	7	8	Total
Original	Count	1	4151	150	30	136	226	1	0	16	4710
		2	200	10233	0	69	215	250	0	24	10991
		3	135	1	6366	133	90	0	0	83	6808
		4	177	346	104	8362	110	68	0	46	9213
		5	123	131	185	292	11988	165	0	54	12938
		6	42	369	0	52	63	6951	1	15	7493
		10	0	0	0	0	0	0	306	7	313
		20	2	0	58	39	44	0	16	1145	1304
	%	1	88.1	3.2	.6	2.9	4.8	.0	.0	.3	100.0
		2	1.8	93.1	.0	.6	2.0	2.3	.0	.2	100.0
		3	2.0	.0	93.5	2.0	1.3	.0	.0	1.2	100.0
		4	1.9	3.8	1.1	90.8	1.2	.7	.0	.5	100.0
		5	1.0	1.0	1.4	2.3	92.7	1.3	.0	.4	100.0
		6	.6	4.9	.0	.7	.8	92.8	.0	.2	100.0
		10	.0	.0	.0	.0	.0	.0	97.8	2.2	100.0
		20	.2	.0	4.4	3.0	3.4	.0	1.2	87.8	100.0
Cross-validated[a]	Count	1	4145	151	30	138	229	1	0	16	4710
		2	202	10229	0	69	217	250	0	24	10991
		3	135	1	6364	133	92	0	0	83	6808
		4	177	346	104	8359	111	70	0	46	9213
		5	123	133	186	296	11981	165	0	54	12938
		6	42	370	0	52	63	6950	1	15	7493
		10	0	0	0	0	0	0	306	7	313
		20	2	0	58	39	44	0	16	1145	1304
	%	1	88.0	3.2	.6	2.9	4.9	.0	.0	.3	100.0
		2	1.8	93.1	.0	.6	2.0	2.3	.0	.2	100.0
		3	2.0	.0	93.5	2.0	1.4	.0	.0	1.2	100.0
		4	1.9	3.8	1.1	90.7	1.2	.8	.0	.5	100.0
		5	1.0	1.0	1.4	2.3	92.6	1.3	.0	.4	100.0
		6	.6	4.9	.0	.7	.8	92.8	.0	.2	100.0
		10	.0	.0	.0	.0	.0	.0	97.8	2.2	100.0
		20	.2	.0	4.4	3.0	3.4	.0	1.2	87.8	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 92.1% of original grouped cases correctly classified.

c. 92.0% of cross-validated grouped cases correctly classified.

El resultado del discriminante de la segmentación realizada en La Empresa de Telecomunicaciones es de 92,1%, considerándose como una segmentación excelente.

3.1.7 Proceso de caracterización de los segmentos

Los segmentos se pasan a enriquecer con la información disponible de cada segmento, permitiendo un conocimiento más amplio, imprescindible en la futura definición de acciones y propuestas de productos para cada uno de ellos.

a) Caracterización interna: variables disponibles al ser información propia de La Empresa de Telecomunicaciones. De todas las variables disponibles se seleccionaron como más relevantes las siguientes 18 variables:

- Ingresos medios (nuevos soles)
- Margen Comercial medio (nuevos soles)
- Total tráfico voz saliente (minutos)
- Total tráfico voz entrante (minutos)
- Total tráfico datos (minutos)
- Número total llamadas salientes
- Número total llamadas entrantes
- Número total conexiones Internet
- Duración llamadas salientes (minutos)
- Duración llamadas entrantes (minutos)
- Duración conexiones Internet(minutos)
- Segmento actual (%)
    - VIP
    - Medio
    - Masivo
- # medio de líneas por cliente

- Antigüedad del cliente (meses)
- % clientes con líneas prepago
- Nivel socioeconómico
- Ubicación geográfica (%)
    - Lima
    - Provincias
- Servicios activos por cliente

Se ha realizado un análisis descriptivo de cada una de las variables para cada uno de los segmentos. Los resultados de la caracterización interna se muestran a continuación:

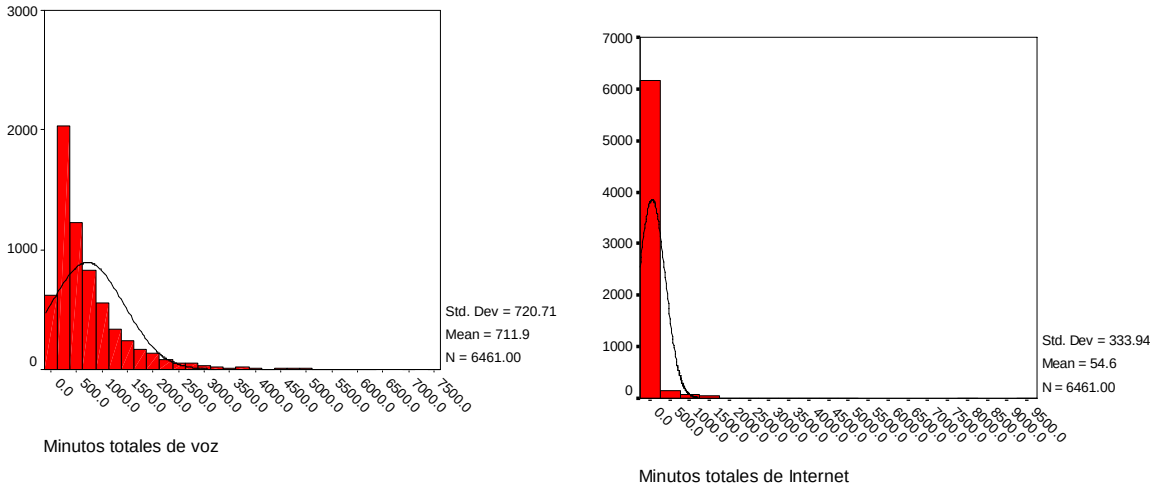
Tabla 1.7-1 Caracterización interna de los segmentos finales

	Clusters									
Variables de caracterización interna	1	2	3	4	5	6	7	8	9	Muestra
Ingresos medios (n. soles)	73.9	69.2	139.4	127.4	103.2	75.4	246.8	280.8	48.4	104.4
Margen comercial medio (n. soles)	55.4	55.3	96.7	91.8	78.0	60.6	198.6	200.9	41.3	78.3
Minutos totales salientes	144.5	94.4	436.7	257.8	267.4	95.7	481.2	632.6	23.6	223.8
Minutos totales entrantes	260.3	418.0	276.3	422.3	422.1	491.4	287.3	488.3	28.5	385.3
Minutos totales Internet	11.6	4.4	54.5	25.0	21.6	5.2	4440.6	2241.3	1.5	94.8
Llamadas totales salientes	101.8	54.0	175.3	129.0	109.6	49.9	120.4	196.0	13.0	100.9
Llamadas totales entrantes	84.2	184.1	101.9	187.2	154.4	210.5	95.7	176.8	12.1	155.8
Conexiones totales Internet	0.4	0.3	2.6	1.5	1.0	0.3	234.5	92.4	0.1	4.3
Duración llamadas salientes	1.4	1.8	2.5	2.0	2.4	1.9	4.0	3.2	1.8	2.1
Duración llamadas entrantes	2.8	2.2	2.8	2.3	2.7	2.4	3.0	2.8	2.3	2.5
Duración conexiones Internet	14.3	12.6	16.4	13.7	16.3	15.7	24.8	24.5	17.2	18.7
Segmento actual (%)										
VIP	3%	2%	14%	10%	5%	2%	43%	42%	2%	7%
Medio	20%	24%	49%	46%	43%	24%	39%	49%	12%	35%
Masivo	77%	74%	37%	44%	52%	74%	18%	10%	85%	58%
# medio líneas/cliente	1.03	1.03	1.08	1.12	1.06	1.04	1.13	1.24	1.01	1.07
Antigüedad del cliente (meses)	61	73	146	127	116	99	129	160	79	106
% clientes c/líneas prepago	78%	63%	37%	45%	51%	62%	12%	14%	67%	53%
Nivel socioeconómico (%)										
A	2%	0%	11%	8%	5%	3%	13%	20%	3%	5%
B	16%	10%	31%	25%	26%	23%	35%	39%	15%	22%
C	9%	11%	10%	10%	9%	8%	7%	7%	11%	10%
D	42%	48%	24%	29%	28%	29%	20%	14%	42%	33%
E	6%	7%	3%	4%	3%	3%	2%	2%	7%	4%
K	1%	1%	1%	2%	1%	1%	2%	1%	1%	1%
N	23%	21%	19%	22%	27%	31%	21%	16%	20%	24%
O	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%
Ubicación geográfica (%)										
Lima	58%	49%	65%	63%	71%	73%	76%	79%	50%	63%
Provincias	42%	51%	35%	37%	29%	27%	24%	21%	50%	37%
Servicios activos/cliente (media)	1.3	0.9	1.3	1.2	1.1	0.9	2.6	2.3	0.8	1.2

b) Se analizó, también para cada segmento, el nivel de consumo tanto de voz como de Internet. Para ello, se ha obtenido un histograma para cada uno de los segmentos.

A continuación ejemplo de curvas de consumo para voz e Internet del segmento 3 “Próximos”

Gráfico 1.7-1 Histogramas de consumo Internet y de voz



c) Se identificó la cartera de productos por segmento, identificando el nivel de penetración para cada uno de los productos. Para ello, se ha calculado un descriptivo para cada uno de las variables de productos, donde se tomaba como ratio de penetración, la suma total de productos del segmento para un producto determinado / total de clientes del segmento.

Tabla 1.7-2 Penetración por segmento de los productos

	Clusters									
Penetración de productos	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Muestra**
Bloqueos	2.0%	2.4%	5.3%	4.1%	4.1%	2.5%	12.0%	12.7%	1.2%	3.8%
Cambios	2.5%	1.8%	4.5%	3.4%	2.5%	2.2%	9.2%	6.0%	2.1%	2.9%
Duplicado de Recibo Telefónico	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%
No figurar en Guía Telefónica	1.7%	0.8%	5.6%	3.0%	2.6%	2.2%	6.4%	8.0%	1.8%	2.7%
Reconexión del Servicio	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Traslado de la línea principal	2.3%	0.9%	3.7%	2.5%	1.8%	1.2%	5.0%	4.1%	1.4%	2.1%
Speedy	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	75.9%	35.2%	0.0%	2.7%
Tarifa Plana	0.0%	0.0%	0.1%	0.1%	0.1%	0.0%	2.2%	2.7%	0.1%	0.2%
Cyberbonos	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.7%	0.7%	0.1%	0.1%
Llamada en Espera	6.8%	3.3%	14.6%	11.4%	8.6%	4.9%	26.4%	28.0%	2.1%	8.7%
Conferencia Tripartita	3.8%	2.2%	7.9%	6.1%	4.6%	2.4%	14.2%	12.8%	0.7%	4.6%
Identificación de llamadas	6.6%	2.5%	11.8%	9.2%	7.3%	3.6%	21.9%	23.0%	1.5%	7.2%
Facturación detallada	2.3%	1.9%	6.4%	5.9%	4.4%	2.4%	12.8%	16.5%	0.8%	4.3%
Facturación detallada B	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Desvío por ocupado	0.1%	0.2%	0.2%	0.4%	0.3%	0.1%	0.9%	1.4%	0.0%	0.3%
Desvío por ausencia	0.1%	0.6%	0.5%	0.7%	0.5%	0.3%	1.1%	1.8%	0.2%	0.5%
Transferencia de Llamadas	2.6%	1.1%	6.4%	4.9%	3.2%	1.7%	11.3%	10.5%	0.8%	3.5%
Línea Directa	0.1%	0.2%	0.2%	0.2%	0.1%	0.1%	0.3%	0.5%	0.1%	0.2%
Servipack1	1.5%	0.8%	2.2%	1.8%	1.5%	0.7%	3.9%	4.2%	0.1%	1.5%
Servipack2	0.8%	0.3%	1.4%	1.4%	1.1%	0.4%	3.5%	3.8%	0.1%	1.0%
Servipack3	2.3%	0.5%	3.2%	2.1%	1.6%	0.7%	6.7%	7.3%	0.3%	1.8%
Seguripack	0.0%	0.0%	0.4%	0.3%	0.2%	0.1%	1.0%	1.9%	0.1%	0.3%
Memovox	93.0%	70.8%	56.1%	64.5%	65.1%	68.4%	50.3%	57.2%	68.2%	67.4%

3.2. Segmentación por valor

3.2.1. Cálculo de los segmentos de valor

a) Construcción de cuenta de resultados por cliente

El valor del cliente vendrá determinado por el margen comercial que el cliente aporte a La Empresa de Telecomunicaciones. Ver detalle de variables en el Anexo 1 *Matriz de variables original.*

Gráfico 0-1 Cuenta de resultados de un cliente residencial de La Empresa de Telecomunicaciones

En el caso de La Empresa de Telecomunicaciones, las variables para la elaboración de la Función Valor provinieron de distintas Gerencias. CUNE no dispone en este momento de la información de costos al nivel de detalle que la Función Valor requiere.

Se usaron::

a) Variables de ingresos por tráfico tarifado

b) Variables de ingresos por tráfico prepago

c) Variables de por abonos fijos mensuales

Los Costos de Interconexión fueron provistos por la Gerencia Correspondiente. Los Costos Variables Directos fueron provistos de la siguiente manera:

a) Costos de Información y Atención a Clientes: Gerencia de Atención a Clientes
b) Costos de Reparaciones: Gerencia de Operaciones y Redes de Atención al Cliente
c) Costos de Facturación: Gerencia de Facturación

Para el proceso de segmentación de valor, se decide utilizar como variable base el Margen comercial, el cuál se calcula mediante la siguiente fórmula:

Ingresos por tráfico tarifado = *Facturación por tráfico local saliente a fijo* (vi1) + *Facturación por tráfico de datos* (vi2) + *Facturación por tráfico larga distancia nacional saliente a fijo* (vi3) + *Facturación por tráfico larga distancia nacional saliente a fijo -otras operadoras-* (vi4) + *Facturación por tráfico larga distancia nacional saliente a movil -Movistar-* (vi5) + *Facturación por tráfico larga distancia nacional saliente a móviles -otras operadoras móviles-* (vi6) + *Facturación por tráfico larga distancia internacional* (vi7) + *Facturación por tráfico larga distancia internacional -otras operadoras-* (vi8) + *Facturación por tráfico a cobrar entrante* (vi9) + *Facturación por tráfico local saliente fijo a móvil –Movistar-* (vi10) + *Facturación por tráfico local saliente fijo a móvil -otras operadoras móviles-* (vi11) + *Facturación por EVAS* (vi28)

Ingresos por tráfico prepago = *Facturación por tráfico local saliente a fijo c/ tarjeta prepago* (vi20) + *Facturación por tráfico local saliente a movil c/tarjeta prepago* (vi21) + *Facturación por tráfico larga distancia nacional saliente a fijo c/tarjeta prepago* (vi22) + *Facturación por tráfico larga distancia nacional saliente a movil c/tarjeta prepago* (vi23) + *Facturación por tráfico larga distancia nacional internacional c/tarjeta prepago* (vi24)

Ingresos por cuotas fijas = *Facturación por cuota mensual de la línea* (vi12) + *Facturación por cuota mensual de servicios de valor agregado* (vi13) + *Facturación por cuota mensual de Speedy* (vi14) + *Facturación de otros conceptos* (vi16) + *Facturación de Pago Adelantado* (vi27)

Ingresos de Interconexión = 0,0593x *Tráfico saliente de fijo a fijo larga distancia nacional realizado a través de otras operadoras* (v10) + 0,0416x *Tráfico saliente de fijo a larga distancia internacional realizado a través de otras operadoras* (v13) + 0,0416x *Tráfico entrante desde fijo local extra-red* (v109) + 0,041676x *Tráfico entrante desde móvil local* (v110) + 0,1321x *Tráfico entrante desde móvil larga distancia nacional* (v111) + 0,0422x *Tráfico entrante desde fijo larga distancia nacional extra-red* (v113) + 0,1932x *Tráfico entrante desde larga distancia internacional* (v114)

Costos de Interconexión = 0,0416x *Tráfico saliente de fijo a fijo local extra-red* (v5) + 0,91x *Facturación por tráfico local saliente fijo a móvil –*

Movistar- (vi10) + *Facturación por tráfico local saliente fijo a móvil -otras operadoras móviles-* (vi11)) + 0,7227x *Tráfico saliente de fijo a móvil larga distancia nacional* (v7) + 0,0416x *Tráfico saliente de fijo a fijo larga distancia nacional extra-red* (v9) + 0,2753x *Tráfico saliente de fijo a larga distancia internacional* (v11) + 0,0416x *Tráfico saliente de fijo a fijo local extra-red c/tarjeta 147* (v61) + 0,8625x *Tráfico saliente de fijo a móvil local c/tarjeta 147* (v62) + 0,7082x *Tráfico saliente de fijo a móvil larga distancia nacional c/tarjeta 147* (v63) + 0,0416x *Tráfico saliente de fijo a fijo larga distancia nacional extra-red c/tarjeta 147* (v65) + 0,2753x *Tráfico saliente de fijo a larga distancia internacional c/tarjeta 147* (v66) + 0,0416x *Tráfico entrante desde cobro revertido* (v117)

Transferencias de dinero = 0,093 x *Facturación por EVAS* (vi28) x *Nºllamadas saliente de fijo a números de tarifa diferenciada* (v32) + *Facturación por tráfico larga distancia nacional saliente a fijo -otras operadoras-* (vi4) + *Facturación por tráfico larga distancia internacional -otras operadoras-* (vi8)

Costos variables directos = 0,4801x *Nº de Facturas* (vi26) + 0.275x *Nº de llamadas que ha realizado al 103* (v209a) +0,9039x *Nº de llamadas que ha realizado 104* (v209b) +7,7460x *Nº de llamadas que ha realizado al 102* (v209)

Margen comercial = *Ingresos por tráfico tarifado + Ingresos por tráfico prepago + Ingresos por cuotas fijas + Ingresos de Interconexión – Costes de Interconexión – Transferencias de dinero – Costos variables directos*

Esta nueva variable se calcula para toda la muestra y se adiciona a la matriz creando una nueva columna (mar_come).

Los pasos para la segmentación por valor fueron los siguientes:

Los valores de corte de los segmentos son números redondos fácilmente comunicables y entendibles en toda la organización por lo que no deben ser complejos. No se usa 87,34; sí se puede usar 85 o 90, por ejemplo.

a) Como resultado, se hallan los cortes de la función valor en La Empresa de Telecomunicaciones que definen los segmentos de valor:
    - Oro: más de 150 nuevos soles (9% de los clientes de la muestra)
    - Plata: entre 101 y 150 nuevos soles (11,3% de los clientes de la muestra)
    - Bronce: entre 61 y 100 nuevos soles (31,4% de los clientes de la muestra)
    - Plomo: mayor que 0 y menor que 60 nuevos soles (47,6% de los clientes de la muestra)
    - Destructores: menor a 0 soles (1% de los clientes de la muestra)

A continuación se observa esa distribución en el Output de Frecuencial de SPSS

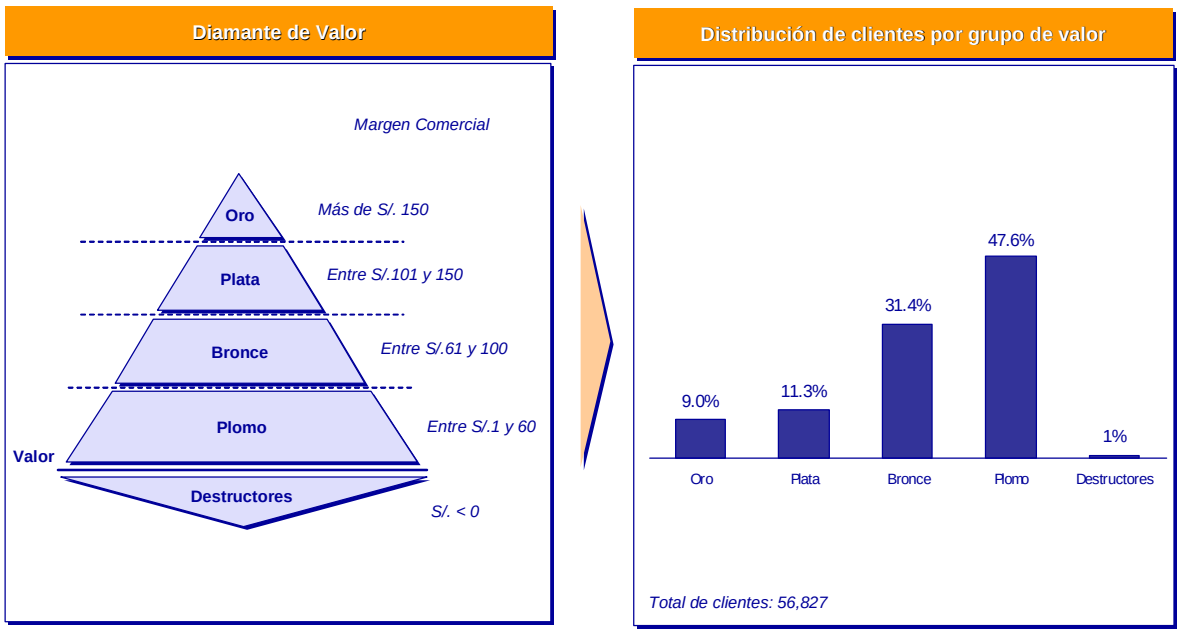
Tabla 2.1-1 Output SPSS: Frecuencial grupos de Valor

Grupos de valor

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Oro	5095	9.0	9.0	9.0
	Plata	6410	11.3	11.3	20.2
	Bronce	17829	31.4	31.4	51.6
	Plomo	27077	47.6	47.6	99.3
	Destructores	416	.7	.7	100.0
	Total	56827	100.0	100.0	

A continuación se muestra un gráfico resumen del resultado de la segmentación de Valor

Gráfico 0-2 Distribución de la muestra de Telesp por Segmento de Valor



3.2.2. Cruces de Segmentos por patrón de consumo con Segmentos de valor

Los grupos por valor identificados se cruzaron con los segmentos por patrón de consumo para ver su composición.

a) Se aplicó un crosstab entre las variables que definen los grupos de valor y de uso, lo que permite analizar la composición de los segmentos por

patrón de consumo respecto al valor y los segmentos por Valor respecto al patrón de consumo.

Tabla 2.2.1 Output Crosstab Segmentos por patrón de consumo con segmento de valor

Grupos de valor * Grupos finales de uso Crosstabulation

			Grupos de valor					
			Oro	Plata	Bronce	Plomo	Destructores	Total
Grupos finales de uso	1.00	Count	170	324	1089	2920	167	4670
		% within Grupos de valor	3.3%	5.1%	6.1%	10.8%	40.1%	8.2%
		% within Grupos finales de uso	3.6%	6.9%	23.3%	62.5%	3.6%	100.0%
		% of Total	.3%	.6%	1.9%	5.1%	.3%	8.2%
	2.00	Count	197	573	2765	7318	73	10926
		% within Grupos de valor	3.9%	8.9%	15.5%	27.0%	17.5%	19.2%
		% within Grupos finales de uso	1.8%	5.2%	25.3%	67.0%	.7%	100.0%
		% of Total	.3%	1.0%	4.9%	12.9%	.1%	19.2%
	3.00	Count	922	1168	2428	1897	46	6461
		% within Grupos de valor	18.1%	18.2%	13.6%	7.0%	11.1%	11.4%
		% within Grupos finales de uso	14.3%	18.1%	37.6%	29.4%	.7%	100.0%
		% of Total	1.6%	2.1%	4.3%	3.3%	.1%	11.4%
	4.00	Count	1061	1339	3422	2953	42	8817
		% within Grupos de valor	20.8%	20.9%	19.2%	10.9%	10.1%	15.5%
		% within Grupos finales de uso	12.0%	15.2%	38.8%	33.5%	.5%	100.0%
		% of Total	1.9%	2.4%	6.0%	5.2%	.1%	15.5%
	5.00	Count	756	1670	5235	4772	33	12466
		% within Grupos de valor	14.8%	26.1%	29.4%	17.6%	7.9%	21.9%
		% within Grupos finales de uso	6.1%	13.4%	42.0%	38.3%	.3%	100.0%
		% of Total	1.3%	2.9%	9.2%	8.4%	.1%	21.9%
	6.00	Count	207	502	2194	4458	40	7401
		% within Grupos de valor	4.1%	7.8%	12.3%	16.5%	9.6%	13.0%
		% within Grupos finales de uso	2.8%	6.8%	29.6%	60.2%	.5%	100.0%
		% of Total	.4%	.9%	3.9%	7.8%	.1%	13.0%
	10.00	Count	714	297	152	16	1	1180
		% within Grupos de valor	14.0%	4.6%	.9%	.1%	.2%	2.1%
		% within Grupos finales de uso	60.5%	25.2%	12.9%	1.4%	.1%	100.0%
		% of Total	1.3%	.5%	.3%	.0%	.0%	2.1%
	20.00	Count	1046	504	284	33	3	1870
		% within Grupos de valor	20.5%	7.9%	1.6%	.1%	.7%	3.3%
		% within Grupos finales de uso	55.9%	27.0%	15.2%	1.8%	.2%	100.0%
		% of Total	1.8%	.9%	.5%	.1%	.0%	3.3%
	99.00	Count	22	33	260	2710	11	3036
		% within Grupos de valor	.4%	.5%	1.5%	10.0%	2.6%	5.3%
		% within Grupos finales de uso	.7%	1.1%	8.6%	89.3%	.4%	100.0%
		% of Total	.0%	.1%	.5%	4.8%	.0%	5.3%
Total		Count	5095	6410	17829	27077	416	56827
		% within Grupos de valor	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% within Grupos finales de uso	9.0%	11.3%	31.4%	47.6%	.7%	100.0%
		% of Total	9.0%	11.3%	31.4%	47.6%	.7%	100.0%

A continuación dos gráficos muestran los cruces entre los segmentos de valor y patrón de consumo.

Gráfico 2.2-1 Descomposición de los segmentos de patrón de comportamiento entre los segmentos de valor

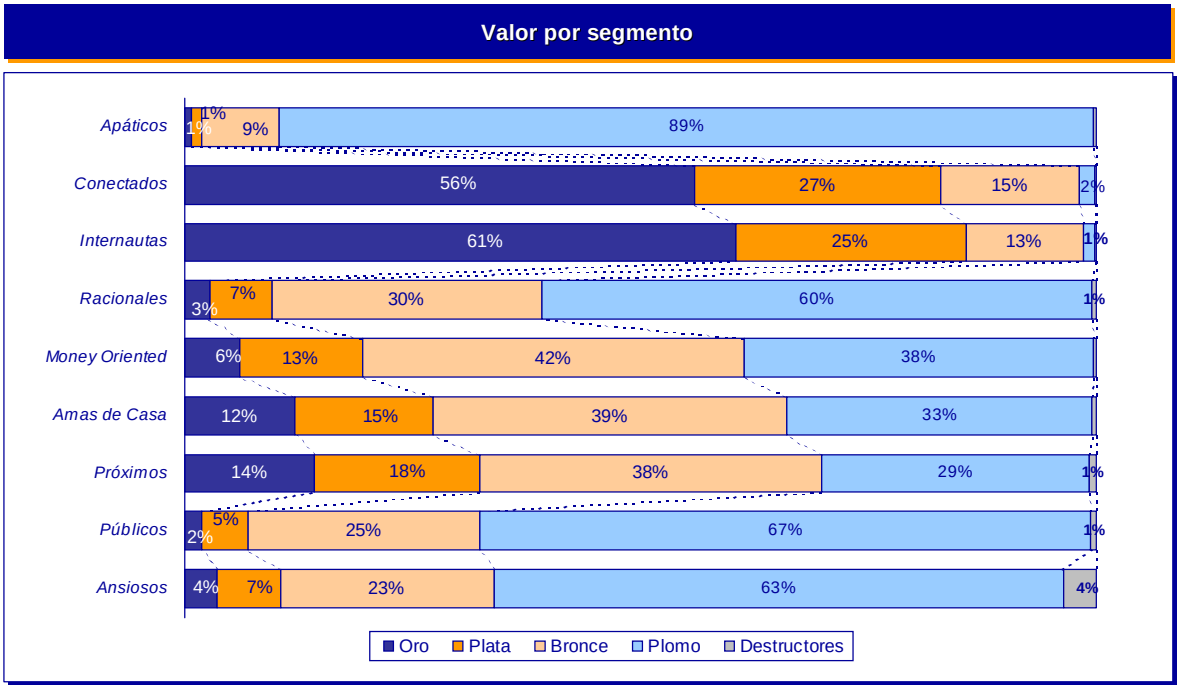
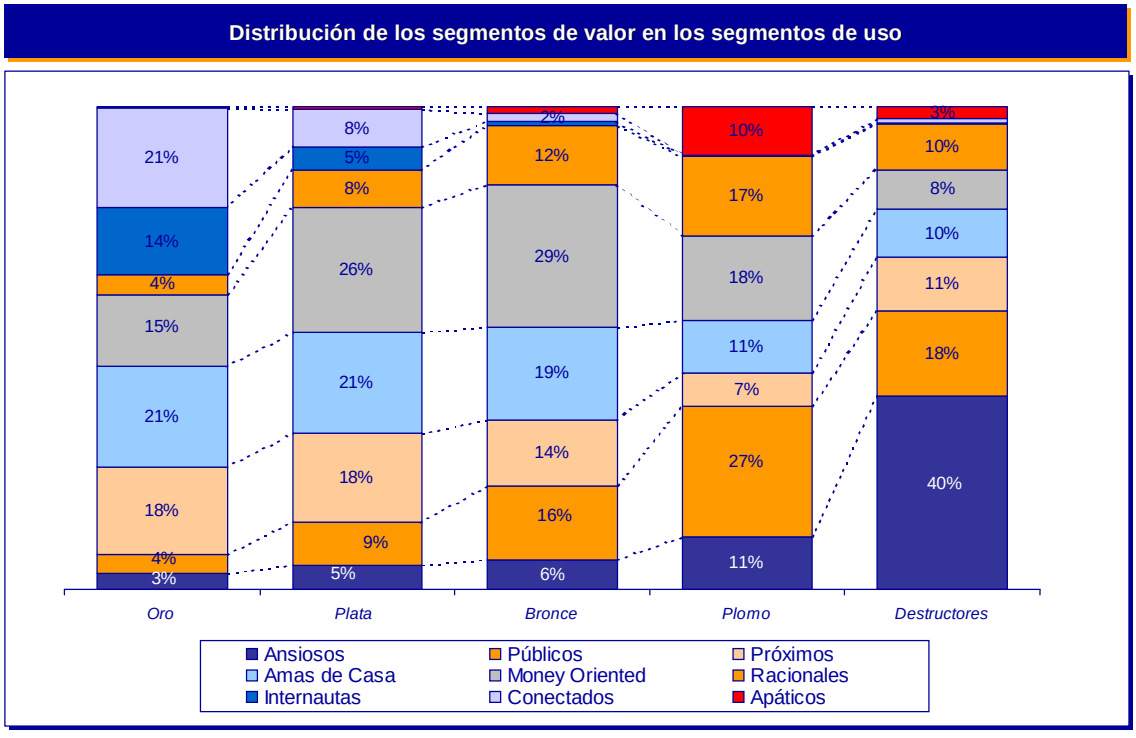


Gráfico 2.2-2 Descomposición de los segmentos de valor entre los segmentos de patrón de comportamiento



b) Se analizó la media de los ingresos totales y del margen comercial para cada uno de los grupos de uso. Mediante un análisis descriptivo se calcula la media de la variable *Ingresos Totales* (ing_to) y de la variables *Margen Comercial* (mar_come) para cada uno de los grupos.

A continuación puede observarse un ejemplo del output del análisis descriptivo, en este caso el del segmento 1, mediante la función Olap Cubes

Tabla 2.2-2 Output Olab Cubes de la variable Ingresos Totales y Margen Comercial

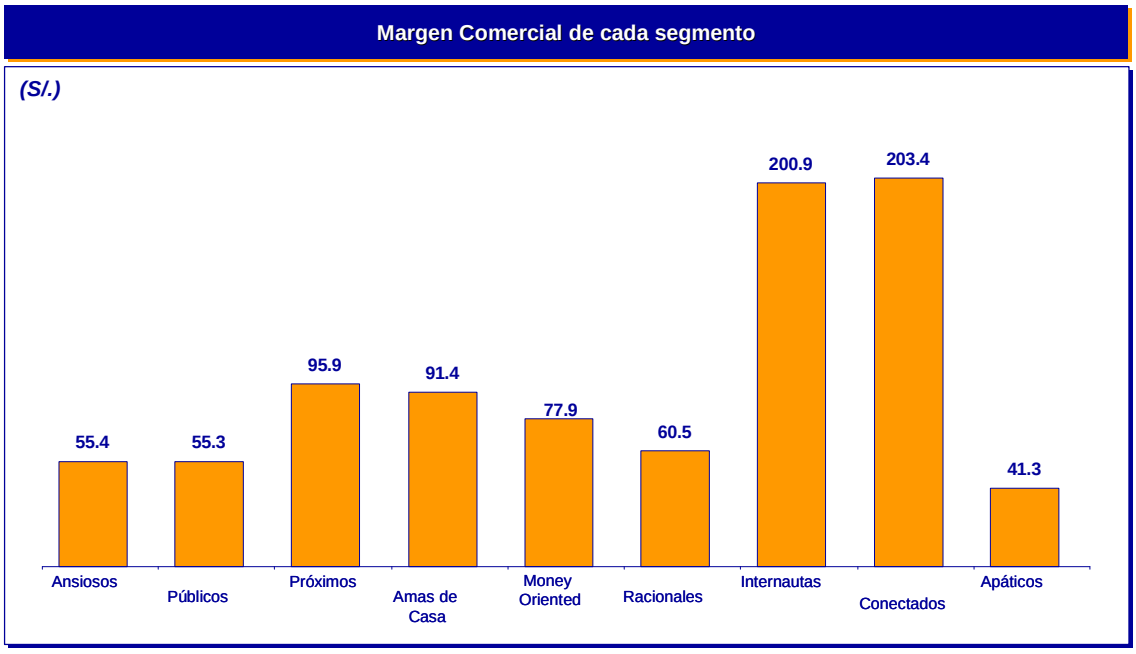
OLAP Cubes

Grupos finales segmentacion: 1.00

	Sum	N	Mean	Std. Deviation	% of Total Sum	% of Total N
Ingreso total	345268	4670	73.93	63.85	5.8%	8.2%
Margen comercial	258744	4670	55.41	47.32	5.8%	8.2%

En el siguiente gráfico se muestra la media del margen comercial para cada uno de los segmentos de patrón de consumo

Gráfico 2.2-3 Media de margen comercial por segmento



c) Composición del valor para cada segmento por patrón de consumo.

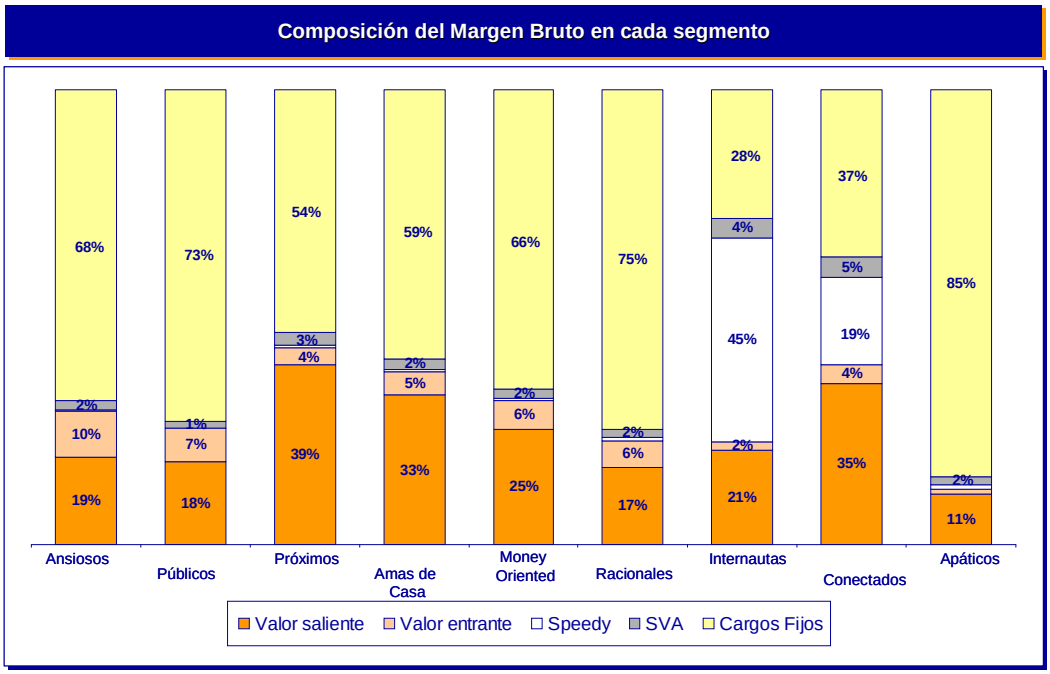
Para cada segmento se ha calculado el desglose del margen bruto. Para ello, en primer lugar, se ha calculado los siguientes conceptos:

- Valor saliente: ingresos por tráfico tarifado + ingresos por tráfico prepago – costes de interconexión – transferencia de dinero
- Valor entrante: ingresos de interconexión
- Speedy : abono mensual por Speedy
- SVA: cargos fijos por servicios de valor agregado
- Cargos fijos: renta mensual + otros conceptos + cuotas de pago adelantado

Una vez obtenido el margen bruto de cada concepto se calcula el margen bruto total como la suma de todos estos conceptos. Obviamente, el margen bruto ha de ser el mismo que el calculado de forma conjunta y recogido en la variable *Margen bruto* (mar_brut).

De este modo es posible analizar para cada segmento, cuánto aporta cada uno de los conceptos al margen comercial total.

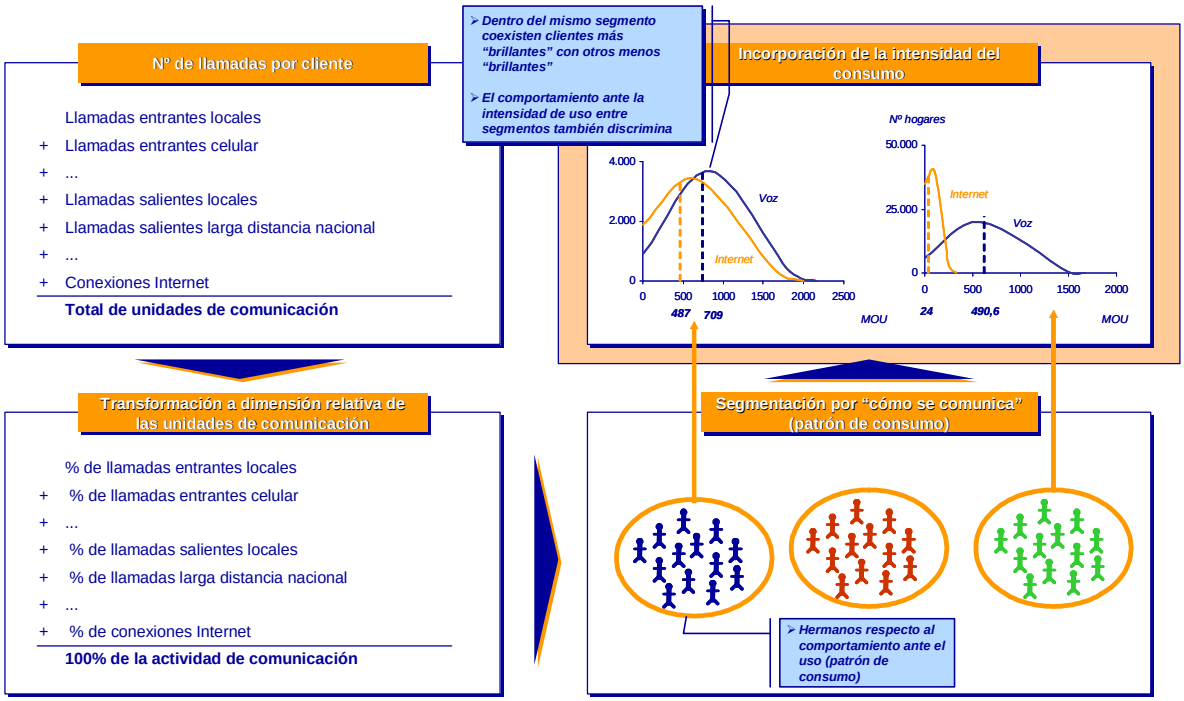
Gráfico 2.2-4 Descomposición del margen bruto de cada segmento por concepto



3.2.3 Microsegmentación

Una vez definidos los segmentos por uso dando respuesta a la pregunta ¿cómo consume?, se lleva a cabo un proceso de microsegmentación de cada uno de los segmentos para de este modo dar respuesta al ¿cuánto consumen?

Gráfico 2.3-1 Metodología de la segmentación, entender primero como los clientes consumen para después incorporar cuánto consumen



La dimensión "cuánto" se ha trabajado con tres niveles de margen comercial dado que esta variable se encuentra altamente correlacionada con el nivel de consumo. La ventaja de utilizar la variable margen radica en que es habitualmente utilizada en la definición de estrategias comerciales y en la definición de las políticas de relación con el cliente.

Para ello, cada segmento se ha dividido en tres microsegmentos:

a) Alto consumo: (incluye el segmento Oro) – margen comercial mayor a 150 nuevos soles

b) Consumo medio: (incluye los segmentos plata y bronce) – margen comercial entre 61 y 150 nuevos soles

c) Bajo consumo: (incluye el segmento plomo) – margen comercial entre 0 y 60 soles

De este modo, existen un total de 27 microsegmentos (9x3), con el siguiente peso sobre el total de la muestra, que se obtiene de realizar la función frecuencial sobre la variable microsegmentos:

Gráfico 2.3-2 Cuadro resumen del peso relativo de los microsegmentos

Distribución de los 27 microsegmentos

Segmentos de valor

Segmentos por patrón de consumo	Oro	Plata / Bronce	Plomo	TOTAL
Ansiosos	0,3%	2,5%	5,2%	**8,0%**
Públicos	0,3%	5,9%	13,0%	**19,2%**
Próximos	1,6%	6,4%	3,4%	**11,4%**
Amas de casa	1,9%	8,4%	5,2%	**15,6%**
Money Oriented	1,3%	12,2%	8,5%	**22,0%**
Racionales	0,4%	4,8%	7,9%	**13,0%**
Internautas	1,3%	0,8%	0,0%	**2,1%**
Conectados	1,9%	1,4%	0,1%	**3,3%**
Apáticos	0,0%	0,5%	4,8%	**5,4%**
TOTAL	**9,0%**	**42,7,0%**	**47,6%**	**100,0%**

- ***Acciones de fidelización***
- ***Modelos de atención personalizados***

- ***Acciones de cross selling***
- ***Acciones de hostigamiento promocional***

- ***Acciones de activación***
- ***Acciones de hostigamiento promocional***

En la matriz de trabajo SPSS, se crea una columna adicional, denominada microsegmento, con el valor del microsegmento al que pertenece cada uno de los clientes. El nuevo valor del microsegmento se obtiene de realizar la función if en SPSS, cruzando los valores de las columnas del segmento por patrón de consumo y al segmento de valor al que pertenece el cliente. Esto es, si el cliente pertenece al grupo por patrón de consumo 1 y de valor alto consumo, entonces en la nueva columna, el cliente toma el valor 11, si pertenece al segmento por patrón de consumo 1 y a medio consumo, entonces toma el valor 12, si pertenece al segmento por patrón de consumo 1 y al de bajo valor entonces toma el valor 13 y así sucesivamente para los 27 microsegmentos.

Para cada uno de los microsegmentos, se lleva a cabo un proceso de caracterización similar al realizado para la segmentación por patrón de consumo y comentado en el apartado anterior. Para obtener los resultados, se calcula la media para cada microsegmento de las variables de caracterización a través de un análisis descriptivo.

El resultado de la caracterización de las variables internas se puede observar en el siguiente cuadro resumen.

Gráfico 2.3-3 Variables de caracterización interna por microsegmento

Variables de caracterización interrnas	Clusters												Muestra
	1			2			3			4			
	Bajo valor	Valor medio	Alto valor	Bajo valor	Valor medio	Alto valor	Bajo valor	Valor medio	Alto valor	Bajo valor	Valor medio	Alto valor	
Ingresos medios (nuevos soles)	47.1	109.0	288.4	50.0	101.7	246.5	58.8	127.9	348.2	60.9	120.9	340.1	104.4
Margen comercial medio (nuevos soles)	36.2	85.6	212.8	40.3	81.0	202.5	42.5	91.5	228.7	45.4	89.0	234.7	78.3
Minutos totales salientes	93.9	208.4	493.8	65.9	149.2	227.4	159.3	414.5	1096.0	103.5	259.2	674.7	223.8
Minutos totales entrantes	150.1	397.6	1143.7	321.0	570.0	1474.2	93.5	276.7	655.8	199.6	446.8	931.2	385.3
Minutos totales Internet	0.3	19.9	147.6	0.4	10.8	47.7	1.8	40.7	219.5	0.5	15.4	137.9	94.8
Llamadas totales salientes	82.1	125.7	226.6	43.7	73.4	104.1	79.5	164.5	413.1	62.4	129.2	311.9	100.9
Llamadas totales entrantes	66.0	111.9	192.2	158.0	233.9	315.3	41.3	100.5	232.3	102.5	192.7	398.6	155.8
Conexiones totales Internet	0.0	0.8	4.8	0.0	0.6	2.8	0.1	1.8	11.4	0.0	1.0	7.4	4.3
Duración llamadas salientes (min.)	1.1	1.7	2.4	1.5	2.2	2.5	2.0	2.6	2.9	1.7	2.1	2.3	2.1
Duración llamadas entrantes (min.)	2.3	3.7	6.2	2.0	2.5	5.1	2.3	2.9	3.0	2.0	2.5	2.6	2.5
Duración conexiones Internet (min.)	5.4	15.5	19.8	5.4	15.3	18.9	13.8	16.4	16.7	7.8	12.8	15.8	18.7
Segmento actual (%)													
VIP	0%	5%	38%	0%	4%	26%	0%	8%	66%	0%	6%	55%	7%
Medio	9%	41%	38%	13%	48%	44%	23%	68%	30%	23%	62%	40%	35%
Masivo	90%	54%	24%	87%	47%	29%	77%	24%	4%	77%	32%	5%	58%
# medio líneas/hogar	1.01	1.06	1.24	1.00	1.07	1.30	1.00	1.04	1.40	1.01	1.09	1.59	1.07
Antigüedad del hogar (meses)	47	87	131	64	91	97	99	160	189	96	137	170	106
% clientes c/líneas prepago	87%	60%	46%	70%	48%	52%	59%	29%	20%	55%	41%	39%	53%
Nivel socioeconómico (%)													
A	1%	4%	9%	0%	1%	0%	3%	11%	31%	2%	8%	23%	5%
B	12%	23%	36%	8%	13%	25%	20%	34%	38%	18%	27%	30%	22%
C	11%	7%	5%	12%	9%	7%	12%	9%	7%	13%	9%	9%	10%
D	49%	30%	21%	51%	42%	30%	37%	21%	10%	37%	26%	19%	33%
E	6%	5%	2%	8%	6%	6%	4%	3%	1%	5%	3%	3%	4%
K	1%	1%	3%	1%	1%	2%	1%	1%	1%	1%	2%	2%	1%
N	20%	30%	24%	19%	27%	28%	23%	19%	11%	23%	24%	13%	24%
O	1%	1%	1%	1%	1%	2%	1%	1%	1%	1%	1%	1%	1%
Ubicación geográfica (%)													1.16
Lima	49%	74%	86%	43%	61%	72%	52%	68%	80%	51%	68%	74%	63%
Provincias	51%	26%	14%	57%	39%	28%	48%	32%	20%	49%	32%	26%	37%
Servicios activos/hogar (media)	1.05	1.64	2.59	0.82	1.05	1.56	0.85	1.31	2.23	0.80	1.25	2.26	1.16

Gráfico 2.3-4 Variables de caracterización interna por microsegmento (cont´)

Variables de caracterización interrnas	Clusters												Muestra
	5			6			7			8			
	Bajo valor	Valor medio	Alto valor	Bajo valor	Valor medio	Alto valor	Bajo valor	Valor medio	Alto valor	Bajo valor	Valor medio	Alto valor	
Ingresos medios (nuevos soles)	60.3	112.9	284.3	50.5	101.3	278.1	79.1	143.9	319.6	85.1	149.0	392.7	104.4
Margen comercial medio (nuevos soles)	46.9	85.9	205.8	41.3	82.2	212.1	42.5	112.2	260.6	43.8	109.6	279.9	78.3
Minutos totales salientes	143.0	302.1	721.3	59.0	145.0	237.9	806.0	473.2	464.5	518.2	533.7	781.4	223.8
Minutos totales entrantes	240.1	481.6	1030.6	353.0	669.7	1148.5	275.3	258.9	352.1	261.1	397.7	636.8	385.3
Minutos totales Internet	1.0	22.5	144.6	0.8	7.5	73.0	3779.1	4473.3	4362.3	1219.2	2058.2	2584.9	94.8
Llamadas totales salientes	69.8	121.6	250.5	35.7	68.2	111.0	170.9	110.6	135.7	121.3	164.5	248.2	100.9
Llamadas totales entrantes	103.4	172.1	315.1	159.0	272.3	508.2	118.0	93.4	98.8	102.1	150.4	221.7	155.8
Conexiones totales Internet	0.1	1.1	6.0	0.0	0.4	3.2	289.4	239.2	216.2	61.5	81.3	111.1	4.3
Duración llamadas salientes	2.1	2.6	3.0	1.7	2.3	2.5	4.6	4.2	3.5	3.3	3.1	3.3	2.1
Duración llamadas entrantes	2.3	2.9	3.5	2.3	2.6	2.9	2.7	2.7	3.5	2.3	2.7	3.0	2.5
Duración conexiones Internet	7.6	16.4	19.7	13.5	16.1	16.7	14.5	24.1	27.3	15.5	24.0	25.7	18.7
Segmento actual													
VIP	0%	4%	43%	0%	3%	40%	0%	7%	67%	0%	7%	69%	7%
Medio	18%	60%	48%	12%	44%	45%	38%	61%	26%	39%	77%	27%	35%
Masivo	82%	37%	9%	88%	54%	15%	63%	33%	7%	61%	16%	3%	58%
# medio líneas/hogar	1.00	1.05	1.50	1.00	1.08	1.43	1.00	1.05	1.19	1.00	1.05	1.42	1.07
Antigüedad del hogar (meses)	85	131	176	81	125	146	94	127	131	99	146	172	106
% clientes c/líneas prepago	62%	44%	41%	69%	52%	46%	25%	12%	12%	27%	12%	15%	53%
Nivel socioeconómico (%)													
A	1%	5%	19%	1%	5%	15%	0%	8%	17%	3%	13%	27%	5%
B	19%	29%	39%	19%	30%	37%	19%	32%	37%	39%	37%	41%	22%
C	10%	8%	6%	9%	7%	4%	13%	10%	5%	6%	8%	5%	10%
D	37%	24%	14%	35%	20%	14%	50%	23%	17%	24%	18%	10%	33%
E	4%	3%	1%	4%	2%	1%	13%	2%	2%	0%	2%	2%	4%
K	1%	1%	1%	1%	2%	1%	6%	1%	2%	3%	1%	2%	1%
N	27%	29%	19%	30%	32%	27%	0%	22%	20%	24%	21%	13%	24%
O	1%	1%	1%	1%	1%	1%	0%	1%	1%		1%	1%	1%
Ubicación geográfica (%)													1.16
Lima	63%	75%	85%	67%	81%	85%	44%	69%	81%	70%	73%	84%	63%
Provincias	37%	25%	15%	33%	19%	15%	56%	31%	19%	30%	27%	16%	37%
Servicios activos/hogar (media)	0.81	1.18	2.15	0.81	1.09	1.72	1.63	1.91	3.15	0.85	1.52	3.08	1.16

CAPITULO IV

4.1. Matriz de Kohonen

Si estamos atentos, la vida cotidiana nos proporciona abundantes ejemplos de mapas autoorganizados. Por ejemplo, el primer día de curso los alumnos de una determinada asignatura se sientan de forma aleatoria en las sillas del aula. Conforme va avanzando el curso, podemos observar como se van recolocando, sentándose en función de sus afinidades: grupos de chicos, grupos de chicas, los empollones, los de las últimas filas, las “parejas”...

Las matrices o mapas de Kohonen pertenecen al grupo de mapas auto-asociativos (SOM: Self Organisising Maps). Desarrollados por Teuvo Kohonen desde 1989, se basan en redes neuronales no supervisadas -el alumno aprende por sí mismo, sin profesor que le corrige-. No se dispone de patrones de

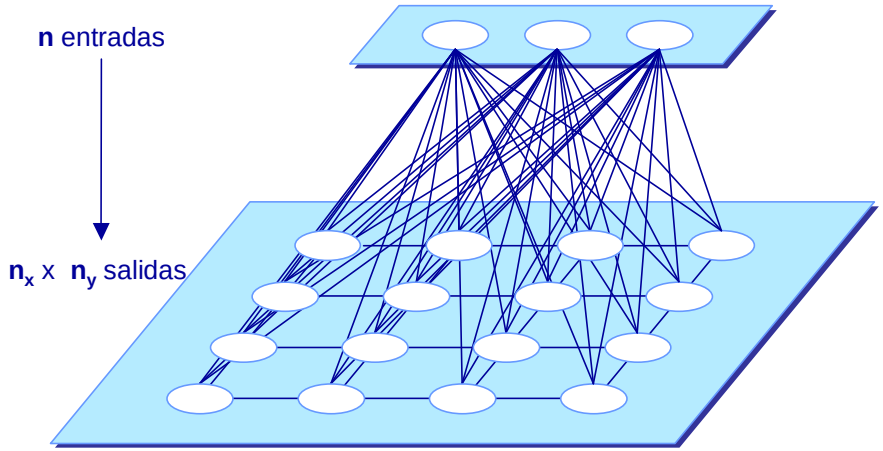
entrenamiento, la información relevante hay que buscarla dentro de los patrones de entrada-. Los algoritmos no supervisados están normalmente basado en algún método de competición entre las neuronas.

Si bien este tipo de redes neuronales ha demostrado su eficacia en tareas de clasificación, reducción de dimensiones y extracción de rasgos, su utilidad más importante se relaciona con la clasificación de información o el agrupamiento de patrones por tipos o clases.

La idea básica del modelo es crear una imagen de un espacio multidimensional de entrada en un espacio de salida de menor dimensionalidad.

Se trata de un modelo con dos capas de neuronas, una de entrada y otra de procesamiento. Las neuronas de la primera capa se limitan a recoger y canalizar la información. La segunda capa está conectada a la primera a través de los pesos sinápticos y realiza una proyección no lineal del espacio multidimensional de entrada, preservando las características esenciales de estos datos en forma de relaciones de vecindad. El resultado final es la creación del llamado mapa auto-organizado donde se representan los rasgos más sobresalientes del espacio de entrada *(ver Gráfico 4.1- 1)*.

Gráfico 4.1-1 Metodología de formación de un mapa auto-organizado



Y / X	0	1	2	3	4	5	6	7	8	9	10	11
0	N: 766 M_COM_DN: 14.723 INTERNET: 47% FF S: 28% FF E: 7% FM E: 5%	N: 509 M_COM_DN: 20.923 FF E: 61% INTERNET: 31% FF E: 6% FM S: 5%	N: 736 M_COM_DN: 13.871 FF S: 55% INTERNET: 13% FF E: 8%	N: 347 M_COM_DN: 15.957 FF E: 60% FM S: 15% INTERNET: 13% FF E: 11%	N: 538 M_COM_DN: 12.506 FF E: 61% FM S: 24% FF S: 6% FN S: 6%	N: 527 M_COM_DN: 12.930 FM S: 23% FF S: 22% FF E: 15% FF C: 7% FM E: 7%	N: 505 M_COM_DN: 11.865 FM S: 42% FF S: 22% FN S: 6%	N: 519 M_COM_DN: 15.148 FF S: 55% FM S: 23%	N: 531 M_COM_DN: 12.892 FF S: 84% FM S: 24% H.Laboral: 69% H.Tarde: 36%	N: 494 M_COM_DN: 11.303 FF S: 69% FM S: 17% FF E: 7% H.Laboral: 61% H.Noche: 37%	N: 392 M_COM_DN: 10.439 FF S: 67% FM S: 12% FF E: 7% H.Laboral: 54% H.Noche: 36%	N: 631 M_COM_DN: 9.936 FF S: 74% FF E: 8% FM S: 6% H.Laboral: 69% H.Tarde: 33%
1	N: 223 M_COM_DN: 13.381 FF S: 39% INTERNET: 25% FN S: 12% FN E: 11% FM S: 6%	N: 200 M_COM_DN: 20.714 FF E: 61% INTERNET: 20% FF S: 6% FN S: 3% FN E: 7%	N: 289 M_COM_DN: 14.898 FF S: 54% INTERNET: 14% FF E: 9% FN E: 7%	N: 198 M_COM_DN: 16.263 FF S: 55% FM S: 15% FF E: 9% FN E: 8% INTERNET: 6%	N: 432 M_COM_DN: 13.841 FF S: 59% FF E: 10% FN S: 6% FF C: 9%	N: 284 M_COM_DN: 14.462 FF S: 49% FM S: 23% FF E: 9% FF C: 9% FN E: 6%	N: 300 M_COM_DN: 13.592 FF S: 51% FM S: 22% FF C: 7% FN E: 66%	N: 325 M_COM_DN: 12.952 FF S: 23%	N: 402 M_COM_DN: 12.563 FF S: 68% FM S: 24% H.Laboral: 71% H.Tarde: 34%	N: 307 M_COM_DN: 13.438 FF S: 73% FM S: 11% FF E: 9% H.Laboral: 61% H.Mañana: 33%	N: 242 M_COM_DN: 12.774 FF S: 77% FF E: 7% FM S: 5% H.Laboral: 60% H.Tarde: 33%	N: 507 M_COM_DN: 9.819 FF S: 85% FF E: 7% H.Laboral: 67% H.Noche: 33%
2	N: 486 M_COM_DN: 17.833 FN S: 33% FN E: 14% FM E: 14% FM S: 8% FF S: 5%	N: 293 M_COM_DN: 15.777 FF S: 42% FF E: 15% FN S: 11% FF C: 6% FM S: 6%	N: 323 M_COM_DN: 13.478 FF S: 53% FF E: 13% FF C: 6% FN S: 7% FM E: 6%	N: 445 M_COM_DN: 13.821 FF S: 61% FN S: 15% FM E: 7% FF S: 5%	N: 380 M_COM_DN: 15.513 FF S: 57% FN S: 12% FN E: 9% FM S: 7%	N: 413 M_COM_DN: 18.736 FF S: 52% FN S: 11% FN E: 10% FM S: 10% H.Laboral: 79%	N: 386 M_COM_DN: 15.612 FF S: 58% FM S: 16% FF E: 8% FN S: 8% H.Laboral: 53% H.Mañana: 41%	N: 413 M_COM_DN: 12.741 FF S: 66% FM S: 17% FF E: 7% H.Laboral: 79% H.Mañana: 44%	N: 409 M_COM_DN: 12.303 FF S: 73% FM S: 14% H.Laboral: 73% H.Mañana: 42%	N: 574 M_COM_DN: 12.096 FF S: 78% FM S: 11% FF E: 4% H.Laboral: 71% H.Mañana: 41%	N: 431 M_COM_DN: 11.776 FF S: 84% FF E: 7% H.Laboral: 66% H.Mañana: 36%	N: 732 M_COM_DN: 10.394 FF S: 88% FF E: 6% H.Laboral: 67% H.Noche: 37%
3	N: 371 M_COM_DN: 14.019 FN S: 26% FN E: 24% FF S: 22% FM S: 9% FM E: 7%	N: 245 M_COM_DN: 13.622 FF S: 36% FF E: 26% FN S: 12% FF C: 7% FM S: 7%	N: 365 M_COM_DN: 12.834 FF S: 55% FN S: 22% FM S: 10% FF E: 7%	N: 390 M_COM_DN: 13.074 FF S: 58% FN S: 15% FM E: 13%	N: 396 M_COM_DN: 12.130 FF S: 62% FN S: 9% H.Laboral: 74%	N: 448 M_COM_DN: 17.535 FF S: 57% FN S: 12% FN E: 12% FM E: 6% H.Laboral: 79%	N: 339 M_COM_DN: 15.750 FF S: 59% FM S: 19% FN S: 9% FM E: 7% H.Laboral: 50%	N: 262 M_COM_DN: 13.260 FF S: 69% FM S: 18% FF E: 6% H.Laboral: 95% H.Mañana: 52%	N: 393 M_COM_DN: 12.639 FF S: 79% FF E: 7% H.Laboral: 74% H.Mañana: 54%	N: 567 M_COM_DN: 12.313 FF S: 85% FF E: 6% H.Laboral: 73% H.Mañana: 54%	N: 431 M_COM_DN: 11.374 FF S: 87% FF E: 5%	N: 540 M_COM_DN: 10.915 FF S: 88% FF E: 4%
4	N: 474 M_COM_DN: 13.612 FN S: 40% FN E: 25% FM S: 13% FF S: 7% FM E: 7	N: 323 M_COM_DN: 17.564 FF S: 35% FF E: 26% FN S: 11% FN E: 7% FM E: 7	N: 329 M_COM_DN: 13.545 FF S: 63% FM S: 13% FN S: 10% FF E: 9%	N: 403 M_COM_DN: 14.348 FF S: 56% FF E: 18% FN S: 13% FM E: 5%	N: 391 M_COM_DN: 14.944 FF S: 53% FN S: 11% FM S: 9% H.Laboral: 72%	N: 351 M_COM_DN: 14.774 FF S: 56% FN S: 14% FM S: 8% H.Laboral: 85% H.Mañana: 50%	N: 357 M_COM_DN: 12.924 FF S: 63% FM S: 8% H.Laboral: 55% H.Mañana: 55%	N: 396 M_COM_DN: 12.331 FF S: 73% FM S: 7% H.Laboral: 85% H.Mañana: 55%	N: 469 M_COM_DN: 12.279 FF S: 82% FF E: 7% H.Laboral: 86% H.Mañana: 56%	N: 455 M_COM_DN: 11.897 FF S: 9% H.Laboral: 81% H.Mañana: 44%	N: 418 M_COM_DN: 11.458 FF S: 87% FF E: 5% H.Laboral: 78% H.Tarde: 40%	N: 691 M_COM_DN: 11.561 FF S: 87% FF E: 4%
5	N: 368 M_COM_DN: 7.523 FN S: 53% FN E: 11% FM S: 11% FN S: 5%	N: 217 M_COM_DN: 9.671 FM E: 26% FF S: 24% FN E: 22% FN S: 9%	N: 477 M_COM_DN: 13.087 FF S: 34% FM S: 26% FN E: 22% FM S: 9%	N: 377 M_COM_DN: 16.322 FM E: 28% FN E: 14% FM S: 9% FF S: 7%	N: 387 M_COM_DN: 16.478 FN S: 57% FF E: 8% FN E: 8% FM S: 7%	N: 311 M_COM_DN: 14.762 FF S: 54% FN S: 14% FM S: 9% H.Laboral: 81%	N: 436 M_COM_DN: 12.577 FF S: 61% FM S: 8% H.Laboral: 61% H.Mañana: 53%	N: 367 M_COM_DN: 14.468 FF S: 70% FM S: 8% H.Laboral: 81% H.Mañana: 50%	N: 437 M_COM_DN: 15.667 FF S: 76% FM S: 8% H.Laboral: 86% H.Mañana: 47%	N: 456 M_COM_DN: 15.315 FF S: 80% FM S: 6% H.Laboral: 87% H.Tarde: 46%	N: 338 M_COM_DN: 12.294 FF S: 81% FM S: 5% H.Laboral: 81% H.Tarde: 50%	N: 422 M_COM_DN: 10.774 FF S: 84% FM S: 4%
6	N: 517 M_COM_DN: 3.844 FN S: 28% FM E: 26%	N: 256 M_COM_DN: 7.960 FM E: 45% FF S: 19% FN S: 13% FN S: 6%	N: 324 M_COM_DN: 11.602 FM E: 28% FF S: 26% FN S: 11% FF E: 11% FM S: 7%	N: 266 M_COM_DN: 15.869 FF S: 37% FM E: 27% FN E: 14% FM S: 11% FN S: 7%	N: 326 M_COM_DN: 13.837 FF S: 49% FF E: 22% FN S: 9% FN E: 6%	N: 450 M_COM_DN: 13.117 FF S: 54% FN S: 24% FM S: 5% H.Laboral: 84% H.Mañana: 49%	N: 462 M_COM_DN: 12.458 FF S: 61% FM S: 8% H.Laboral: 83% H.Mañana: 44%	N: 372 M_COM_DN: 14.390 FF S: 71% FM S: 9% H.Laboral: 84% H.Tarde: 45%	N: 419 M_COM_DN: 14.612 FF S: 74% FM S: 4% H.Laboral: 86% H.Tarde: 47%	N: 435 M_COM_DN: 13.541 FF S: 81% FF E: 9% FM S: 5% H.Laboral: 83% H.Tarde: 47%	N: 320 M_COM_DN: 12.336 FF S: 83% FF E: 7% FM S: 5%	N: 570 M_COM_DN: 11.134 FF S: 87% FF E: 7% FM S: 5%
7	N: 451 M_COM_DN: 3.151 FN S: 28% FF S: 24%	N: 191 M_COM_DN: 5.687 FF E: 29% FM E: 29% FF S: 22% FF C: 10% FN S: 9%	N: 298 M_COM_DN: 9.512 FF E: 27% FM E: 26% FF S: 25% FM E: 10% FN S: 7%	N: 307 M_COM_DN: 12.149 FF S: 37% FM E: 27% FN E: 13% FM S: 9% FF E: 5%	N: 496 M_COM_DN: 12.329 FF S: 55% FF C: 10% FN S: 5%	N: 466 M_COM_DN: 11.983 FF S: 57% FF E: 22% H.Laboral: 90% H.Tarde: 41%	N: 502 M_COM_DN: 12.509 FF S: 67% FF E: 27% FM S: 5%	N: 522 M_COM_DN: 12.887 FF S: 82% FM S: 8% FF E: 3% H.Laboral: 80% H.Tarde: 47%	N: 491 M_COM_DN: 13.702 FF S: 88% FF E: 18%	N: 497 M_COM_DN: 13.758 FF S: 92% FF E: 12% FM S: 5%	N: 355 M_COM_DN: 12.889 FF S: 93% FN S: 6%	N: 648 M_COM_DN: 12.992 FF S: 94% FN S: 5%
8	N: 361 M_COM_DN: 1.967 FF S: 55% FN E: 23% FM E: 14%	N: 204 M_COM_DN: 3.009 FF S: 37% FM E: 17% FN E: 10% FN S: 7%	N: 237 M_COM_DN: 8.273 FF S: 23% FM E: 23% FN E: 14% FF E: 9% FN S: 6%	N: 354 M_COM_DN: 9.963 FF S: 38% FF E: 28% FM S: 5%	N: 393 M_COM_DN: 10.311 FF S: 47% FF E: 39% FN S: 5%	N: 445 M_COM_DN: 12.071 FF S: 51% FF E: 31% FM S: 5%	N: 422 M_COM_DN: 12.521 FF S: 56% FF E: 27% FN S: 4%	N: 467 M_COM_DN: 12.691 FF S: 62% FF E: 23%	N: 441 M_COM_DN: 13.366 FF S: 68% FF E: 19%	N: 495 M_COM_DN: 13.179 FF S: 72% FF E: 14% H.Laboral: 76% H.Tarde: 36%	N: 404 M_COM_DN: 13.306 FF S: 77% FF E: 12% FM S: 5% H.Laboral: 76% H.Noche: 36%	N: 489 M_COM_DN: 12.715 FF S: 81% FF E: 8% FM S: 4% H.Laboral: 79% H.Noche: 41%
9	N: 859 M_COM_DN: 1.317 FN E: 7%	N: 271 M_COM_DN: 4.662 FF E: 71% FF S: 12% FN S: 5%	N: 458 M_COM_DN: 6.273 FF E: 59% FF S: 27% FN S: 5%	N: 402 M_COM_DN: 8.144 FF E: 48% FF S: 38% FM S: 5%	N: 507 M_COM_DN: 10.838 FF E: 63% FF S: 29% FN S: 5%	N: 563 M_COM_DN: 10.597 FF S: 69% FF E: 22% FM S: 6%	N: 550 M_COM_DN: 11.351 FF S: 76% FF E: 20% FM S: 5% H.Laboral: 47% H.Tarde: 36%	N: 621 M_COM_DN: 10.749 FF S: 82% FF E: 14% H.Laboral: 84% H.Tarde: 37%	N: 630 M_COM_DN: 10.669 FF S: 87% FF E: 10% H.Laboral: 86% H.Mañana: 36%	N: 580 M_COM_DN: 11.202 FF S: 70% FF E: 7% H.Laboral: 71% H.Mañana: 36%	N: 423 M_COM_DN: 11.477 FF S: 70% FF E: 14% FN E: 4% H.Laboral: 74% H.Noche: 42%	N: 512 M_COM_DN: 10.769 FF S: 73% FF E: 9% FM S: 4% H.Laboral: 73% H.Noche: 54%

Nota: el margen comercial directo neto está expresado en pesos chilenos

De igual manera que una cámara es capaz de representar en dos dimensiones un espacio de tres dimensiones. Gracias a ello, es posible, al contemplar una fotografía, hacerse idea de qué hay en una habitación, en un paisaje...

La red neuronal de Kohonen realiza una fotografía de un espacio *n* dimensional, de tal forma que se conserva la topología: los clientes/hogares/empresas que están cercanos en el espacio de *n* dimensiones aparecen próximos en el mapa autoorganizado. De este modo, al contemplar el mapa, tenemos una idea de cómo están situados en el espacio original de *n* dimensiones.

La matriz de Kohonen está formada por una matriz rectangular de neuronas *(ver Gráfico 3.1-1)*, de modo que las relaciones entre los patrones de entrada son visibles, en forma de relaciones de vecindad, de manera más simple. Cada neurona aprende por sí misma a reconocer un determinado tipo de patrón de entrada. En el espacio de salida la topología esencial de entrada queda preservada, de manera que neuronas próximas en el mapa aprenden a reconocer patrones de entrada similares, cuyas imágenes, por lo tanto, aparecerán cercanas en el mapa creado. Este espacio de salida se representa por una capa discreta de neuronas generalmente ordenadas formando una matriz rectangular. Si volvemos al ejemplo con el que iniciábamos el capítulo, los alumnos sentados en las sillas son como neuronas alojadas en la estructura reticular.

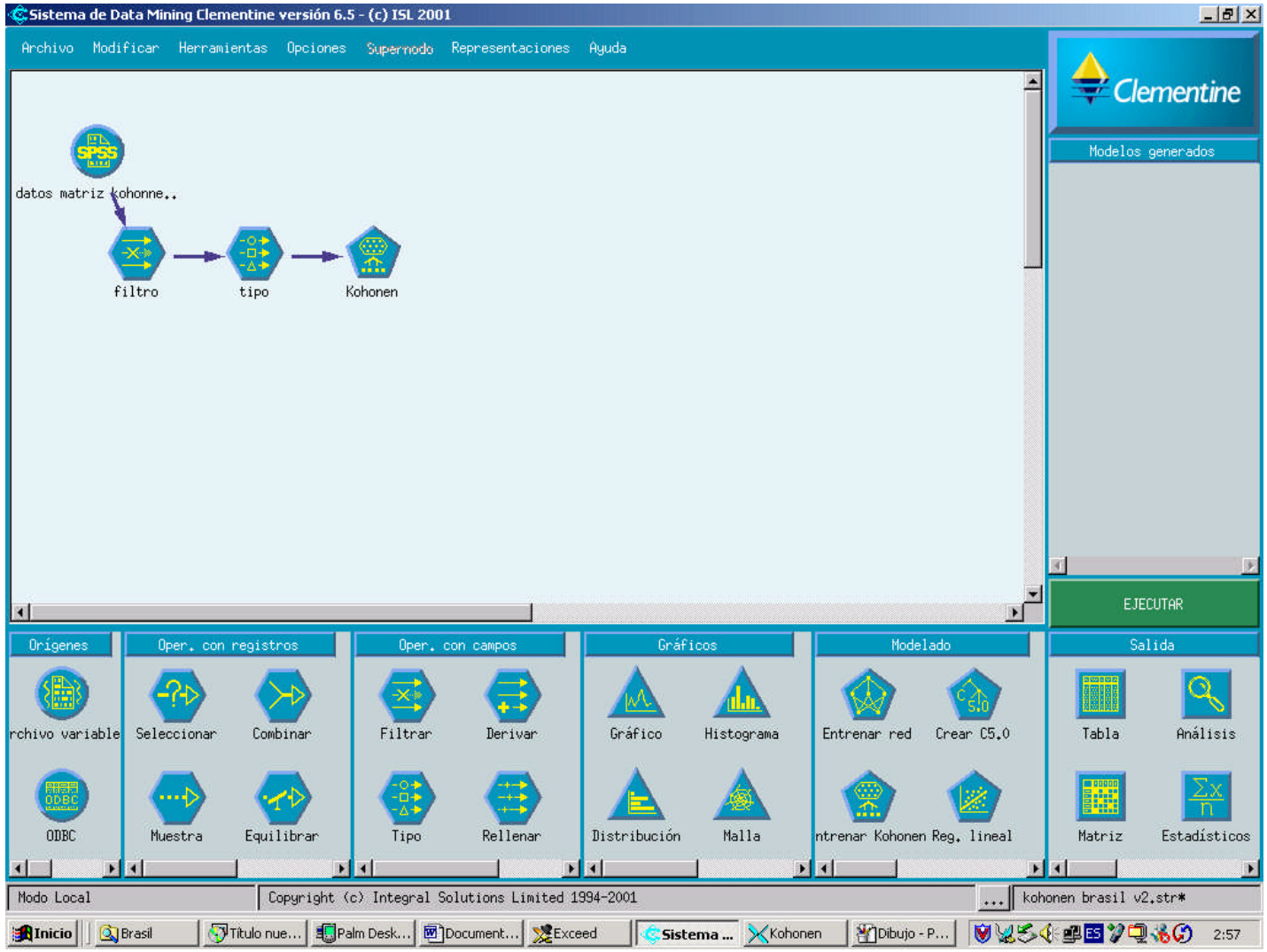
El sistema neuronal descrito puede ser de gran utilidad otras variables de caracterización en el mapa auto-organizado, pudiéndose estudiar caminos de migración y/o evoluciones temporales de los diferenetes clientes/hogares... A priori, son muchas las aplicaciones que puede tener, como análisis exploratorio de datos o mediante su integración en un sistema de ayuda a la toma de decisiones.

4.2. Resultados de la aplicación

Para realizar la segmentación de Kohonen, se ha utilizado el programa Clementine SPSS. Este software es muy visual, lo que nos permite ver con claridad que la preparación para la segmentación de Kohonen es relativamente sencilla.

En el gráfico siguiente podemos ver, cómo se presenta en el Clementine la información para poder ser ejecutado.

Gráfico 4.2-1 Diagrama de análisis



Como puede observarse, y empezando por la izquierda arriba, se realiza un link a la matriz original, donde están incluidas las variables de segmentación (las mismas que se han utilizado para realizar el k-means), el código de cliente, el margen comercial, y el clúster obtenido en la segmentación k-means. Debe observarse que los segmentos del análisis k-means corresponden a los grupos obtenidos antes de la reasignación de los clientes que disponen de Speedy. Se recuerda que El Análisis Discriminante se realiza antes de la reasignación de los clientes con tenencia de Speedy ya que lo que se está midiendo es la bondad de la segmentación. El segmento de “Apáticos” no se analiza debido a que este segmento se crea ad hoc no siguiendo un proceso de segmentación no jerárquico.

Después de hacer el link con la base original, se filtran. Es decir, se seleccionan solamente aquellas variables que van a participar de la segmentación de Kohonen que son exactamente las mismas que las utilizadas en la segmentación k-means.

Estas variables son:

- % conexiones Internet
- % llamadas saliente a fijo local
- % llamadas salientes a móviles
- % llamadas salientes a movil LDN
- % llamadas salientes a fijo LDN
- % llamadas salientes a LDI
- % llamadas salientes a buzon de voz 159
- % llamadas salientes a buzon de voz 158

- % llamadas salientes a plataforma de atencion a clientes
- % llamadas salientes a # gratuitos
- % llamadas salientes a # de tarifa diferenciada
- % llamadas entrantes fijo local
- % llamadas entrantes movil fijo local
- % llamadas entrantes movil fijo LDN
- % llamadas entrantes fijo LDN
- % llamadas entrantes LDI
- % llamadas cobro revertido
- % llamadas entrantes teléfono público
- % llamadas en día laboral
- % llamadas día feriado
- % llamadas en sábado
- % llamadas en domingo
- % llamadas en mañana
- % llamadas en tarde
- % llamadas en noche
- % llamadas en madrugada

Posteriormente se determina el tipo de cada variable (real, entera, marca, ...) y determinar su dirección (entrada, salida, ambos,...). Para el caso de La Empresa de Telecomunicaciones las variables utilizadas -de segmentación- son rangos reales de entrada entre 0 y 1.

Ahora la matriz ya está casi preparada para ser segmentada mediante una red neuronal de Kohonen.

Antes de ejecutar la segmentación, debe indicarse el número de celdas en que se quiere realizar la segmentación. La decisión de utilizar una determinada dimensión de la matriz es similar a la de seleccionar un determinado número de segmentos utilizando un algoritmo k-means. Una vez más, el criterio del investigador y su conocimiento del campo de estudio son fundamentales.

Es conveniente establecer inicialmente una semilla aleatoria, para asegurar que siempre que ejecutemos la segmentación obtengamos los mismos resultados.

Una vez ejecutado el proceso el resultado obtenido es una marca de X y una marca de Y para cada cliente considerado. Esta X e Y, son las coordenadas que nos definen cada cuadrado del plano al que pertenece cada cliente analizado.

Gráfico 4.2-2Tabla de resultados

PLS_E_81	PLS_E_12	PLS_E_24	PLS_E_LA	PLS_E_DO	PLS_E_SA	$KX-Kohonen	$KY-Koho
0.385	0.374	0.228	0.698	0.17	0.132	1	8
0.273	0.345	0.338	0.734	0.144	0.122	3	9
0.305	0.433	0.244	0.744	0.113	0.144	2	1
0.089	0.422	0.456	0.672	0.106	0.222	3	9
0.458	0.284	0.239	0.841	0.085	0.075	4	6
0.358	0.37	0.233	0.798	0.07	0.132	3	5
0.446	0.327	0.217	0.727	0.138	0.136	5	3
0.435	0.426	0.139	0.833	0.12	0.046	3	6
0.073	0.122	0.78	0.829	0.146	0.024	11	9
0.282	0.321	0.391	0.622	0.167	0.212	10	9
0.211	0.546	0.222	0.735	0.189	0.076	1	0
0.335	0.266	0.335	0.814	0.091	0.095	2	7
0.156	0.443	0.348	0.795	0.119	0.086	4	0
0.451	0.41	0.131	0.811	0.09	0.098	4	7
0.256	0.462	0.282	0.615	0.077	0.308	9	1

Con estas coordenadas es posible representar en la matriz 8x8 todos los clientes de la muestra. Observación: para mejorar la rapidez y efectividad de realización del mapa de Kohonen, es aconsejable traspasarse los resultados de la matriz a un software más manejable -como el SPSS- para este tipo de operaciones.

Gráfico 4.2-3 Matriz de Kohonen

Y/X	0	1	2	3	4	5	6	7
0	N: 1.177 FF local E: 35% FF local S: 33% F TELF. PUB. E: 9% FM local S: 5% FM local E: 5% H.Laboral: 82% H.Mañana: 54%	N: 817 FF local S: 43% FF local E: 31% FM local S: 5% F TELF. PUB. E: 5% H.Laboral: 78% H.Mañana: 50%	N: 831 FF local S: 50% FF local E: 27% FM local S: 5% F TELF. PUB. E: 5% H.Laboral: 75% H.Mañana: 46%	N: 911 FF local S: 56% FF local E: 20% FM local S: 5% F TELF. PUB. E: 4% FM local E: 3%	N: 512 FF local S: 63% FB.Voz S 159: 8% FF local E: 7% FM local S: 5% FM local E: 4%	N: 392 FF local S: 55% FB.Voz S 159: 25% F AC S: 5% FB.Voz S 158: 5%	N: 102 FF local: 43% FF local E: 21% INTERNET: 11% FM local S: 7% F TELF. PUB. E: 4%	N: 3.622 FF local S: 30% FF local E: 29% INTERNET: 15% F TELF. PUB. E: 7% FM local S: 5%
1	N:771 FF local S: 30% FF local E: 30% F TELF. PUB. E: 12% FF LDN E: 7% FM local S: 5% FB.Voz S 159: 4% H.Laboral: 75%	N: 616 FF local S: 37% FF local E: 32% F TELF. PUB. E: 8% FM local S: 5% FM local E: 4% H.Laboral: 74%	N: 810 FF local S:43% FF local E: 32% F TELF. PUB. E: 7% FM local S: 5% FM local E: 4% H.Laboral: 73% H.Mañana: 40%	N: 727 FF local S: 48% FF local E: 29% F TELF. PUB. E: 7% FM local S: 4% H.Laboral: 70% H.Mañana: 36%	N: 407 FF local S: 51% FF local E: 20% F TELF. PUB. E: 6% FM local S: 5% FB.Voz S 159: 4% H.Noche: 39%	N: 307 FF local S: 43% FB.Voz S 159: 19% FF local E: 11% FM local E: 7% FB.Voz S 158: 4% F AC S: 4%	N: 157 FF local S: 33% FF local E: 32% F TELF. PUB. E: 8% FM local E: 6% FM local S: 6% FB.Voz S 159: 5%	N: 109 FF local E: 39% FF local S: 27% F TELF. PUB. E: 10% FB.Voz S 159: 5% FM local E: 5%
2	N:1.187 FF local E: 30% FF local S: 29% F TELF. PUB. E: 17% FF LDN E: 7% FM local S: 4% H.Laboral: 70% H.Mañana: 42%	N: 780 FF local E: 34% FF local S: 33% F TELF. PUB. E: 11% FM local E: 4% FF LDN E: 4% H.Laboral: 71% H.Mañana: 40%	N: 1.072 FF local S: 39% FF local E: 37% F TELF. PUB. E: 7% FM local S: 4% H.Mañana: 38%	N: 893 FF local S: 43% FF local E: 36% F TELF. PUB. E: 8% FM local S: 4% H.Noche: 36%	N: 906 FF local S: 43% FF local E: 29% F TELF. PUB. E: 9% FM local S: 5% FF LDN E: 3% H.Noche: 42%	N: 719 FF local S: 35% FF local E: 25% FB.Voz S 159: 10% F TELF. PUB. E: 8% FM local E: 5% FF LDN E: 4% H.Noche: 42%	N: 798 FF local S: 29% FF local E: 23% FB.Voz S 159: 21% F TELF. PUB. E: 6% FM local E: 5% FB.Voz S 158: 5%	N: 1.356 FB.Voz S 159: 32% FF local E: 25% FF local S: 18% F TELF. PUB. E: 8% FB.Voz S 158: 3%
3	N: 1.375 FF local E: 38% FF local S: 24% F TELF. PUB. E: 19% FF LDN E: 5% H.Laboral: 70% H.Mañana: 41%	N: 766 FF local E: 40% FF local S: 29% F TELF. PUB. E: 13% FM local E: 4% H.Laboral: 71% H.Mañana: 39%	N: 1.103 FF local E: 43% FF local S: 34% F TELF. PUB. E: 7% FM local E: 4% H.Laboral: 72% H.Mañana: 37%	N: 842 FF local E: 41% FF local S: 37% F TELF. PUB. E: 8% H.Laboral: 69% H.Noche: 38%	N: 867 FF local E: 36% FF local S: 35% F TELF. PUB. E:12% FM local E: 4% H.Laboral: 66% H.Noche: 44%	N: 746 FF local E: 30% FF local S: 28% F TELF. PUB. E: 12% FB.Voz S 159: 8% FM local E: 6% FF LDN E: 5% H.Laboral: 65%	N: 586 FF local E: 27% FF local S: 22% FB.Voz S 159: 14% F TELF. PUB. E: 12% FF LDN E: 7% FM local E: 6%	N: 705 FF local E: 24% FB.Voz S 159: 21% FF local S: 15% F TELF. PUB. E: 15% FF LDN E: 9% FM local E: 4%
4	N: 1.051 FF local E: 43% FF local S: 21% F TELF. PUB. E: 18% FM local E: 4% FF LDN E: 4% H.Laboral: 73% H.Mañana: 44%	N: 852 FF local E: 45% FF local S: 26% F TELF. PUB. E: 12% FM local E: 5% H.Laboral: 73% H.Mañana: 40%	N: 934 FF local E: 47% FF local S: 29% F TELF. PUB. E: 8% FM local E: 5% H.Laboral: 71% H.Mañana: 36%	N: 801 FF local E: 46% FF local S: 30% F TELF. PUB. E: 9% FM local E: 4% H.Laboral: 68% H.Noche: 40%	N: 793 FF local E: 41% FF local S: 28% F TELF. PUB. E: 14% FM local E: 5% H.Laboral: 64% H.Mañana: 44%	N: 940 FF local E: 34% FF local S: 24% F TELF. PUB. E: 18% FF LDN E: 6% FB.Voz S 159: 6% H.Mañana: 42%	N: 699 FF local E: 29% FF local S: 20% F TELF. PUB. E: 18% FF LDN E: 11% FB.Voz S 159: 6% FM local E: 4%	N: 905 FF local E: 22% FF LDN E: 20% F TELF. PUB. E: 17% FF local S: 17% FB.Voz S 159: 8% FF LDN S: 3%
5	N: 1.413 FF local E: 52% FF local S: 18% F TELF. PUB. E: 14% FM local E: 5% H.Laboral: 78% H.Mañana: 49%	N: 875 FF local E: 53% FF local S: 21% F TELF. PUB. E: 10% FM local E: 5% H.Laboral: 73% H.Mañana: 41%	N: 884 FF local E: 52% FF local S: 23% F TELF. PUB. E: 9% FM local E: 5% H.Laboral: 70% H.Mañana: 36%	N: 899 FF local E: 49% FF local S: 22% F TELF. PUB. E: 12% FM local E: 5% H.Laboral: 66% H.Noche: 40%	N: 885 FF local E: 44% FF local S: 21% F TELF. PUB. E: 18% FM local E: 5% H.Laboral: 65% H.Noche: 42%	N: 1.103 FF local E: 37% F TELF. PUB. E: 23% FF local S: 20% FF LDN E: 5% FB.Voz S 159: 5% H.Noche: 39%	N: 786 FF local E: 33% F TELF. PUB. E: 25% FF local S: 19% FF LDN E: 10%	N: 1.245 F TELF. PUB. E: 29% FF local E: 25% FF LDN E: 19% FF local S: 15%
6	N: 6 FF local E: 58% F TELF. PUB. E: 29% H.Laboral: 78% H.Mañana: 56%	N: 164 FF local E: 68% F TELF. PUB. E: 11% FF local S: 9% H.Laboral: 78% H.Mañana: 47%	N: 794 FF local E: 57% FF local S: 16% F TELF. PUB. E: 10% FM local E: 5% H.Laboral: 70%	N: 622 FF local E: 52% F TELF. PUB. E: 15% FF local S: 15% FM local E: 5% H.Laboral: 67%	N: 792 FF local E: 46% F TELF. PUB. E: 22% FF local S: 14% FM local E: 6%	N: 830 FF local E: 41% F TELF. PUB. E: 27% FF local S: 15% FF LDN E: 5%	N: 697 FF local E: 37% F TELF. PUB. E: 29% FF local S: 14% FF LDN E: 7%	N: 895 F TELF. PUB. E: 33% FF local E: 32% FF LDN E: 13% FF local S:12%
7	N: 677 FF local E: 63% F TELF. PUB. E: 24% H.Laboral: 71% H.Mañana: 37%	N: 31 FF local E: 63% F TELF. PUB. E: 26% H.Laboral: 69% H.Mañana: 35%	N: 1.177 FF local E: 66% F TELF. PUB. E: 14% FF local S: 7%	N: 782 FF local E: 56% F TELF. PUB. E: 21% FF local S: 7% FM local E: 5%	N: 939 FF local E: 50% F TELF. PUB. E: 28% FF local S: 7% FM local E: 5%	N: 1.037 FF local E: 44% F TELF. PUB. E: 33% FF local S: 7% FF LDN E: 5%	N: 931 FF local E: 44% F TELF. PUB. E: 32% FF local S: 9%	N: 1.362 FF local E: 42% F TELF. PUB. E: 33% FF local S: 10% FF LDN E: 5%

Se ha añadido para cada celda las principales variables de segmentación (la explicación del mismo se discute más adelante). Cabe resaltar que la matriz representa la proximidad de los segmentos con la proximidad física de éstos en la matriz.

Al objeto de combinar las bondades de ambos tipos de segmentaciones realizadas hasta el momento:

- Análisis tipológico o de grupos que define segmentos homogéneos por patrón de consumo utilizando un algoritmo k-means
- Red neuronal de Kohonen que define relaciones de proximidad entre los individuos que forman parte de cada una de las celdas

Se ha tratado de proyectar los segmentos obtenidos en el análisis tipológico sobre la matriz 8x8 de Kohonen. El objetivo es representar en el plano, con coordenadas X e Y, la ubicación de cada uno de los segmentos identificados en la segmentación k-means.

En esta línea, buscando la máxima eficiencia se ha desarrollado una sintaxis en SPSS que nos permite identificar cada cuadrante resultante (con coordenadas X e Y, recordemos), en un solo número. Esta sintaxis se describe a continuación:

```
IF x=0 & Y =0  KOH = 01.
IF x=1 & Y =0  KOH = 02.
IF x=2 & Y =0  KOH = 03.
IF x=3 & Y =0  KOH = 04.
IF x=4 & Y =0  KOH = 05.
IF x=5 & Y =0  KOH = 06.
IF x=6 & Y =0  KOH = 07.
IF x=7 & Y =0  KOH = 08.
IF x=0 & Y =1 KOH = 11.
IF x=1 & Y =1  KOH = 12.
IF x=2 & Y =1  KOH = 13.
```

IF x=3 & Y =1 KOH = 14.
IF x=4 & Y =1 KOH = 15.
IF x=5 & Y =1 KOH = 16.
IF x=6 & Y =1 KOH = 17.
IF x=7 & Y =1 KOH = 18.
IF x=0 & Y =2 KOH = 21.
IF x=1 & Y =2 KOH = 22.
IF x=2 & Y =2 KOH = 23.
IF x=3 & Y =2 KOH = 24.
IF x=4 & Y =2 KOH = 25.
IF x=5 & Y =2 KOH = 26.
IF x=6 & Y =2 KOH = 27.
IF x=7 & Y =2 KOH = 28.
IF x=0 & Y =3 KOH = 31.
IF x=1 & Y =3 KOH = 32.
IF x=2 & Y =3 KOH = 33.
IF x=3 & Y =3 KOH = 34.
IF x=4 & Y =3 KOH = 35.
IF x=5 & Y =3 KOH = 36.
IF x=6 & Y =3 KOH = 37.
IF x=7 & Y =3 KOH = 38.
IF x=0 & Y =4 KOH = 41.
IF x=1 & Y =4 KOH = 42.
IF x=2 & Y =4 KOH = 43.
IF x=3 & Y =4 KOH = 44.
IF x=4 & Y =4 KOH = 45.
IF x=5 & Y =4 KOH = 46.
IF x=6 & Y =4 KOH = 47.
IF x=7 & Y =4 KOH = 48.
IF x=0 & Y =5 KOH = 51.
IF x=1 & Y =5 KOH = 52.
IF x=2 & Y =5 KOH = 53.
IF x=3 & Y =5 KOH = 54.
IF x=4 & Y =5 KOH = 55.
IF x=5 & Y =5 KOH = 56.
IF x=6 & Y =5 KOH = 57.
IF x=7 & Y =5 KOH = 58.
IF x=0 & Y =6 KOH = 61.
IF x=1 & Y =6 KOH = 62.
IF x=2 & Y =6 KOH = 63.
IF x=3 & Y =6 KOH = 64.
IF x=4 & Y =6 KOH = 65.
IF x=5 & Y =6 KOH = 66.
IF x=6 & Y =6 KOH = 67.

IF x=7 & Y =6 KOH = 68.
IF x=0 & Y =7 KOH = 71.
IF x=1 & Y =7 KOH = 72.
IF x=2 & Y =7 KOH = 73.
IF x=3 & Y =7 KOH = 74.
IF x=4 & Y =7 KOH = 75.
IF x=5 & Y =7 KOH = 76.
IF x=6 & Y =7 KOH = 77.
IF x=7 & Y =7 KOH = 78.
EXECUTE.

Una vez ejecutada la sintaxis, tenemos una nueva variable llamada KOH, que nos identifica su situación en la matriz, según la nomenclatura expresada en la sintaxis.

Tabla 4.2-1Tabla de nomenclatura

Y\X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	0	1	2	3	4	5	6	7
1	10	11	12	13	14	15	16	17
2	20	21	22	23	24	25	26	27
3	30	31	32	33	34	35	36	37
4	40	41	42	43	44	45	46	47
5	50	51	52	53	54	55	56	57
6	60	61	62	63	64	65	66	67

Posteriormente, realizamos un crosstabs de cada uno de los segmentos obtenidos por k-means -recordemos son los segmentos obtenidos en la segmentación, es decir antes de realizar el proceso de reasignación de los clientes que poseen Speedy- y la variable nueva variable KOH, indicando que nos muestre los % de las filas.

De este modo, observamos el porcentaje de los individuos pertenecientes a una determinada celda corresponde a cada uno de los segmentos obtenidos en el análisis de grupos. En función de esto, es posible proyectar cada uno de los segmentos originales en la matriz 8x7 de Kohonen.

El resultado de ese crosstabs es el siguiente:

Tabla 4.2-2 Crosstab

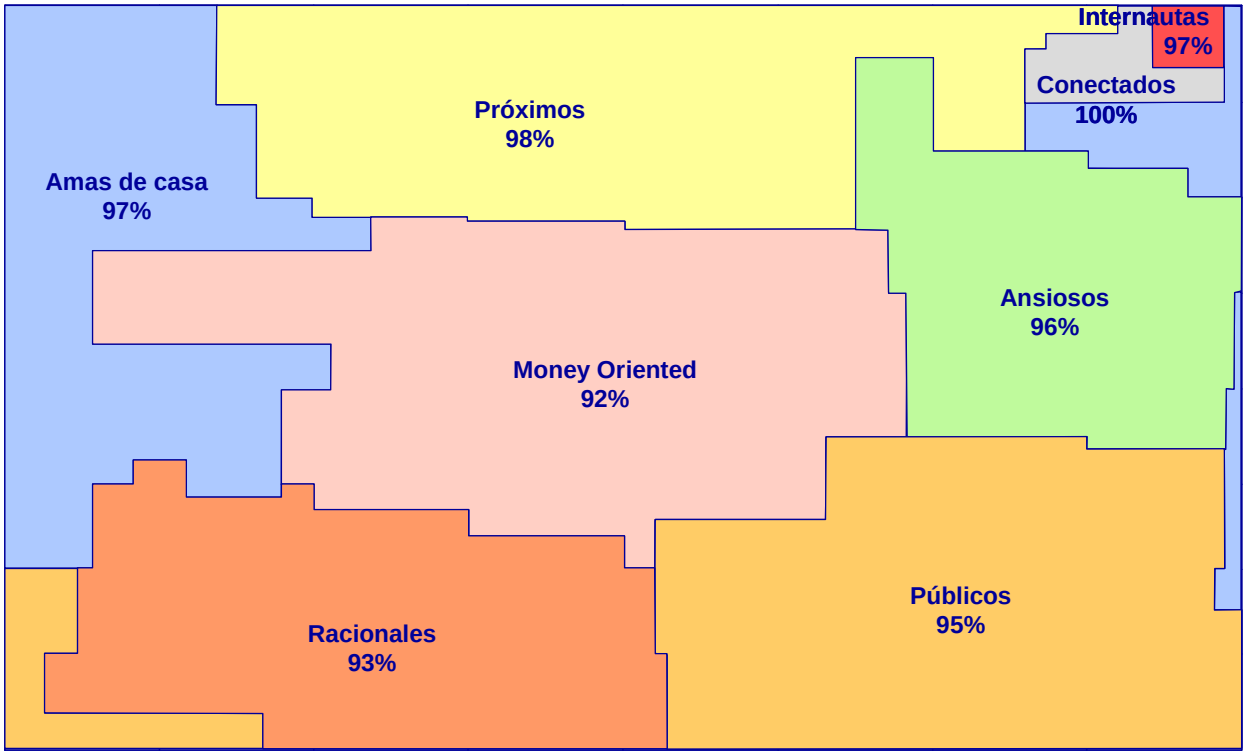
KOH * GRPSNEW Crosstabulation

% within KOH

		GRPSNEW								
		1	2	3	4	5	6	10	20	Total
KOH	,00		26,3%		,3%		73,4%			100,0%
	1,00		55,6%				44,4%			100,0%
	2,00		87,5%			,3%	12,2%			100,0%
	3,00		9,3%			6,5%	84,2%			100,0%
	4,00					,1%	99,9%			100,0%
	5,00				,4%		99,6%			100,0%
	6,00	5,3%	6,6%		18,4%	48,7%	21,1%			100,0%
	7,00	3,0%	3,1%	15,4%	15,1%	16,5%	3,0%	8,5%	35,4%	100,0%
	10,00		55,0%				45,0%			100,0%
	11,00		65,9%				34,1%			100,0%
	12,00		76,4%				23,6%			100,0%
	13,00		22,0%			18,6%	59,4%			100,0%
	14,00					15,3%	84,7%			100,0%
	15,00	,3%			16,3%	1,8%	81,6%			100,0%
	16,00	4,6%	4,1%	13,8%	34,4%	34,9%	8,2%			100,0%
	17,00	8,6%	3,2%	37,6%	34,4%	16,1%				100,0%
	20,00		90,9%		5,5%		3,7%			100,0%
	21,00		76,1%		9,6%		14,2%			100,0%
	22,00		56,0%		7,6%		36,4%			100,0%
	23,00		22,5%		6,1%	37,7%	33,7%			100,0%
	24,00	,1%			6,0%	84,7%	9,2%			100,0%
	25,00				52,9%	45,9%	1,2%			100,0%
	26,00				100,0%					100,0%
	27,00				100,0%					100,0%
	30,00		94,6%		5,4%					100,0%
	31,00		72,7%		27,2%		,1%			100,0%
	32,00		34,8%		64,9%	,4%				100,0%
	33,00		2,2%		52,2%	45,3%	,3%			100,0%
	34,00				8,4%	91,6%				100,0%
	35,00	,2%			41,8%	58,0%				100,0%
	36,00			2,6%	97,4%					100,0%
	37,00			2,7%	97,3%					100,0%
	40,00	16,7%	66,6%	,4%	15,1%	1,1%	,1%			100,0%
	41,00	3,2%	53,7%	1,1%	35,0%	7,0%				100,0%
	42,00	,9%	30,5%	,8%	38,0%	29,8%				100,0%
	43,00		,9%	,7%	14,9%	83,5%				100,0%
	44,00	,1%			4,0%	95,8%				100,0%
	45,00	,9%		22,8%	10,6%	65,7%				100,0%
	46,00	,7%		81,9%	17,1%	,3%				100,0%
	47,00	,5%		89,6%	9,8%					100,0%
	50,00	89,3%	2,6%		8,1%					100,0%
	51,00	66,4%	19,5%	,4%	7,7%	5,9%				100,0%
	52,00	3,0%	35,0%		,1%	61,9%				100,0%
	53,00		1,6%			98,4%				100,0%
	54,00	,1%				99,9%				100,0%
	55,00	2,1%		57,4%		40,5%				100,0%
	56,00	2,2%		97,8%						100,0%
	57,00	2,3%		97,7%						100,0%
	60,00	99,1%	,2%		,7%		,1%			100,0%
	61,00	89,9%	5,0%		,7%	4,1%	,3%			100,0%
	62,00	19,1%	27,8%			52,8%	,3%			100,0%
	63,00	,5%	3,2%			96,3%				100,0%
	64,00	1,9%				98,1%				100,0%
	65,00	6,0%		63,2%		30,8%				100,0%
	66,00	6,5%		93,5%						100,0%
	67,00	19,9%		80,1%						100,0%
Total		8,8%	20,4%	12,7%	17,1%	24,1%	13,9%	,6%	2,4%	100,0%

Estimamos la bondad de la proyección calculando el porcentaje de clientes de cada segmento original que se encuentran dentro del área de proyección.

Gráfico 4.2-4 Proyección de los segmentos en un plano



Finalmente, se procede a caracterizar cada uno de los cuadrantes de la matriz. Para eso, realizaremos un análisis descriptivo de las variables de segmentación y el margen comercial, con la variable KOH, mediante un OLAP Cubes en SPSS, que no añadimos aquí dado que no cabe en una página.

El resultado es el siguiente:

Gráfico 4.2-5 Mapa topográfico de valor

N: 1.177 MAR_COME: 129,19 FF local E: 35% FF local S: 33% F TELF. PUB. E: 9% FM local S: 5% FM local E: 5% H.Laboral: 82% H.Mañana: 54%	N: 817 MAR_COME: 127,55 FF local S: 43% FF local E: 31% FM local S: 5% F TELF. PUB. E: 5% H.Laboral: 78% H.Mañana: 50%	N: 831 MAR_COME: 114,34 FF local S: 50% FF local E: 27% FM local S: 5% F TELF. PUB. E: 5% H.Laboral: 75% H.Mañana: 46%	N: 911 MAR_COME: 101,77 FF local S: 56% FF local E: 20% FM local S: 5% F TELF. PUB. E: 4% FM local E: 3%	N: 512 MAR_COME: 78,74 FF local S: 63% FB.Voz S 159: 8% FF local E: 7% FM local S: 5% FM local E: 4%	N: 392 MAR_COME: 46,15 FF local S: 55% FB.Voz S 159: 25% F AC S: 5% FB.Voz S 158: 5%	N: 102 MAR_COME: 142,66 FF local: 43% FF local E: 21% INTERNET: 11% FM local S: 7% F TELF. PUB. E: 4%	N: 3.622 MAR_COME: 148,31 FF local S: 30% FF local E: 29% INTERNET: 15% F TELF. PUB. E: 7% FM local S: 5%
N:771 MAR_COME: 107,05 FF local S: 30% FF local E: 30% F TELF. PUB. E: 12% FF LDN E: 7% FM local S: 5% FB.Voz S 159: 4% H.Laboral: 75%	N: 616 MAR_COME: 106,41 FF local S: 37% FF local E: 32% F TELF. PUB. E: 8% FM local S: 5% FM local E: 4% H.Laboral: 74%	N: 810 MAR_COME: 110,18 FF local S:43% FF local E: 32% F TELF. PUB. E: 7% FM local S: 5% FM local E: 4% H.Laboral: 73% H.Mañana: 40%	N: 727 MAR_COME: 97,83 FF local S: 48% FF local E: 29% F TELF. PUB. E: 7% FM local S: 4% H.Laboral: 70% H.Mañana: 36%	N: 407 MAR_COME: 91,00 FF local S: 51% FF local E: 20% F TELF. PUB. E: 6% FM local S: 5% FB.Voz S 159: 4% H.Noche: 39%	N: 307 MAR_COME: 51,43 FF local S: 43% FB.Voz S 159: 19% FF local E: 11% FM local E: 7% FB.Voz S 158: 4% F AC S: 4%	N: 157 MAR_COME: 133,42 FF local S: 33% FF local E: 32% F TELF. PUB. E: 8% FM local E: 6% FM local S: 6% FB.Voz S 159: 5%	N: 109 MAR_COME: 118,86 FF local E: 39% FF local S: 27% F TELF. PUB. E: 10% FB.Voz S 159: 5% FM local E: 5%
N:1.187 MAR_COME: 81,19 FF local E: 30% FF local S: 29% F TELF. PUB. E: 17% FF LDN E: 7% FM local S: 4% H.Laboral: 70% H.Mañana: 42%	N: 780 MAR_COME: 90,11 FF local E: 34% FF local S: 33% F TELF. PUB. E: 11% FM local E: 4% FF LDN E: 4% H.Laboral: 71% H.Mañana: 40%	N: 1.072 MAR_COME: 109,50 FF local S: 39% FF local E: 37% F TELF. PUB. E: 7% FM local S: 4% H.Mañana: 38%	N: 893 MAR_COME: 94,35 FF local S: 43% FF local E: 36% F TELF. PUB. E: 8% FM local S: 4% H.Noche: 36%	N: 906 MAR_COME: 85,16 FF local S: 43% FF local E: 29% F TELF. PUB. E: 9% FM local S: 5% FF LDN E: 3% H.Noche: 42%	N: 719 MAR_COME: 82,55 FF local S: 35% FF local E: 25% FB.Voz S 159: 10% F TELF. PUB. E: 8% FM local E: 5% FF LDN E: 4% H.Noche: 42%	N: 798 MAR_COME: 58,13 FF local S: 29% FF local E: 23% FB.Voz S 159: 21% F TELF. PUB. E: 6% FM local E: 5% FB.Voz S 158: 5%	N: 1.356 MAR_COME: 42,37 FB.Voz S 159: 32% FF local E: 25% FF local S: 18% F TELF. PUB. E: 8% FB.Voz S 158: 3%
N: 1.375 MAR_COME: 74,10 FF local E: 38% FF local S: 24% F TELF. PUB. E: 19% FF LDN E: 5% H.Laboral: 70% H.Mañana: 41%	N: 766 MAR_COME: 83,28 FF local E: 40% FF local S: 29% F TELF. PUB. E: 13% FM local E: 4% H.Laboral: 71% H.Mañana: 39%	N: 1.103 MAR_COME: 96,53 FF local E: 43% FF local S: 34% F TELF. PUB. E: 7% FM local E: 4% H.Laboral: 72% H.Mañana: 37%	N: 842 MAR_COME: 82,39 FF local E: 41% FF local S: 37% F TELF. PUB. E: 8% H.Laboral: 69% H.Noche: 38%	N: 867 MAR_COME: 76,40 FF local E: 36% FF local S: 35% F TELF. PUB. E:12% FM local E: 4% H.Laboral: 66% H.Noche: 44%	N: 746 MAR_COME: 77,00 FF local E: 30% FF local S: 28% F TELF. PUB. E: 12% FB.Voz S 159: 8% FM local E: 6% FF LDN E: 5% H.Laboral: 65%	N: 586 MAR_COME: 72,09 FF local E: 27% FF local S: 22% FB.Voz S 159: 14% F TELF. PUB. E: 12% FF LDN E: 7% FM local E: 6%	N: 705 MAR_COME: 48,96 FF local E: 24% FB.Voz S 159: 21% FF local S: 15% F TELF. PUB. E: 15% FF LDN E: 9% FM local E: 4%
N: 1.051 MAR_COME: 75,93 FF local E: 43% FF local S: 21% F TELF. PUB. E: 18% FM local E: 4% FF LDN E: 4% H.Laboral: 73% H.Mañana: 44%	N: 852 MAR_COME: 85,14 FF local E: 45% FF local S: 26% F TELF. PUB. E: 12% FM local E: 5% H.Laboral: 73% H.Mañana: 40%	N: 934 MAR_COME: 82,86 FF local E: 47% FF local S: 29% F TELF. PUB. E: 8% FM local E: 5% H.Laboral: 71% H.Mañana: 36%	N: 801 MAR_COME: 75,46 FF local E: 46% FF local S: 30% F TELF. PUB. E: 9% FM local E: 4% H.Laboral: 68% H.Noche: 40%	N: 793 MAR_COME: 70,06 FF local E: 41% FF local S: 28% F TELF. PUB. E: 14% FM local E: 5% H.Laboral: 64% H.Mañana: 44%	N: 940 MAR_COME: 70,55 FF local E: 34% FF local S: 24% F TELF. PUB. E: 18% FF LDN E: 6% FB.Voz S 159: 6% H.Mañana: 42%	N: 699 MAR_COME: 75,02 FF local E: 29% FF local S: 20% F TELF. PUB. E: 18% FF LDN E: 11% FB.Voz S 159: 6% FM local E: 4%	N: 905 MAR_COME: 74,69 FF local E: 22% FF LDN E: 20% F TELF. PUB. E: 17% FF local S: 17% FB.Voz S 159: 8% FF LDN S: 3%
N: 1.413 MAR_COME: 78,18 FF local E: 52% FF local S: 18% F TELF. PUB. E: 14% FM local E: 5% H.Laboral: 78% H.Mañana: 49%	N: 875 MAR_COME: 74,57 FF local E: 53% FF local S: 21% F TELF. PUB. E: 10% FM local E: 5% H.Laboral: 73% H.Mañana: 41%	N: 884 MAR_COME: 72,17 FF local E: 52% FF local S: 23% F TELF. PUB. E: 9% FM local E: 5% H.Laboral: 70% H.Mañana: 36%	N: 899 MAR_COME: 69,96 FF local E: 49% FF local S: 22% F TELF. PUB. E: 12% FM local E: 5% H.Laboral: 66% H.Noche: 40%	N: 885 MAR_COME: 63,37 FF local E: 44% FF local S: 21% F TELF. PUB. E: 18% FM local E: 5% H.Laboral: 65% H.Noche: 42%	N: 1.103 MAR_COME: 62,42 FF local E: 37% F TELF. PUB. E: 23% FF local S: 20% FF LDN E: 5% FB.Voz S 159: 5% H.Noche: 39%	N: 786 MAR_COME: 61,85 FF local E: 33% F TELF. PUB. E: 25% FF local S: 19% FF LDN E: 10%	N: 1.245 MAR_COME: 63,44 F TELF. PUB. E: 29% FF local E: 25% FF LDN E: 19% FF local S: 15%
N: 6 MAR_COME: 62,58 FF local E: 58% F TELF. PUB. E: 29% H.Laboral: 78% H.Mañana: 56%	N: 164 MAR_COME: 66,57 FF local E: 68% F TELF. PUB. E: 11% FF local S: 9% H.Laboral: 78% H.Mañana: 47%	N: 794 MAR_COME: 61,78 FF local E: 57% FF local S: 16% F TELF. PUB. E: 10% FM local E: 5% H.Laboral: 70%	N: 622 MAR_COME: 61,34 FF local E: 52% F TELF. PUB. E: 15% FF local S: 15% FM local E: 5% H.Laboral: 67%	N: 792 MAR_COME: 59,36 FF local E: 46% F TELF. PUB. E: 22% FF local S: 14% FM local E: 6%	N: 830 MAR_COME: 56,22 FF local E: 41% F TELF. PUB. E: 27% FF local S: 15% FF LDN E: 5%	N: 697 MAR_COME: 53,76 FF local E: 37% F TELF. PUB. E: 29% FF local S: 14% FF LDN E: 7%	N: 895 MAR_COME: 52,42 F TELF. PUB. E: 33% FF local E: 32% FF LDN E: 13% FF local S:12%
N: 677 MAR_COME: 69,62 FF local E: 63% F TELF. PUB. E: 24% H.Laboral: 71% H.Mañana: 37%	N: 31 MAR_COME: 73,24 FF local E: 63% F TELF. PUB. E: 26% H.Laboral: 69% H.Mañana: 35%	N: 1.177 MAR_COME: 52,58 FF local E: 66% F TELF. PUB. E: 14% FF local S: 7%	N: 782 MAR_COME: 52,18 FF local E: 56% F TELF. PUB. E: 21% FF local S: 7% FM local E: 5%	N: 939 MAR_COME: 48,53 FF local E: 50% F TELF. PUB. E: 28% FF local S: 7% FM local E: 5%	N: 1.037 MAR_COME: 46,40 FF local E: 44% F TELF. PUB. E: 33% FF local S: 7% FF LDN E: 5%	N: 931 MAR_COME: 47,47 FF local E: 44% F TELF. PUB. E: 32% FF local S: 9%	N: 1.362 MAR_COME: 49,25 FF local E: 42% F TELF. PUB. E: 33% FF local S: 10% FF LDN E: 5%

Más de 140 *De 120 a 130* *De 100 a 110* *De 80 a 90* *De 60 a70* *De 40 a 50*

De 130 a 140 *De 110 a 120* *De 90 a 100* *De 70 a 80* *De 50 a 60*

CAPITULO V

5.1. CONCLUSIONES

Segmentos Comportamentales

Ansiosos

- ✓ Realizan la mayor cantidad de llamadas al buzón de voz
- ✓ Reciben llamadas desde fijos LDN
- ✓ Poseen la mayor penetración de líneas prepago

Públicos

- ✓ Reciben la mayor proporción de llamadas originadas en teléfonos públicos
- ✓ Reciben la mayor proporción de llamadas LDN

Próximos

- ✓ Realizan la mayor proporción de llamadas salientes a teléfonos fijos

Amas de casa

- ✓ Realizan la mayor proporción de llamadas salientes durante la

mañana y durante los días laborales

Money Oriented

- ✓ Realizan sus llamadas sobre todo durante el horario nocturno y durante los días feriado y domingo

Racionales

- ✓ Tienen la mayor proporción de llamadas entrantes

Internautas

- ✓ Presentan la mayor proporción de conexiones a Internet

Conectado

- ✓ Hogares con una alta proporción de conexiones a Internet

Apático

- ✓ Son aquellos hogares que en total. contabilizan hasta 45 llamadas mensuales (saliente. prepago, entrante y conexiones a Internet)

Segmentos de valor

- ✓ El segmento de valor Oro contiene los clientes más valiosos *para* nuestra empresa. los cuales deben tener una atención diferenciada. Aplicar CRM con ellos.
- ✓ El segmento de valor Plata contiene clientes valiosos a los cuales debemos de realizar campañas de fidelización para así intentar que en un periodo no muy lejano puedan migrar a un segmento Oro.
- ✓ Los segmento de valor Bronce y Plomo agrupan los clientes que generan bajo valor para la empresa pero en grupo representan ingresos significativos, por lo cual debemos de cuidarlos

incentivando el uso de nuestros servicios para que mejoren su rentabilidad.

- ✓ El segmento de valor Destructores agrupa todos aquellos clientes que representan costos para la empresa, a estos clientes solo debemos mantenerlos y no gastar en marketing ni en campañas.

5.2. RECOMENDACIONES

En la segmentación por valor, el margen comercial de cada hogar debe ser calculado mensualmente con la media móvil mensual de las variables que intervienen en el cálculo del margen comercial y del valor presente de cliente.

En la segmentación por patrones de consumo y dado que los segmentos tienen la propiedad de robustez, no es necesario volver a hacer una nueva segmentación antes de transcurridos de ocho a doce meses.

Independientemente de esto, sí puede ser interesante ir reasignando los individuos a los segmentos cada cierto periodo de tiempo para ver si se han producido migraciones utilizando las Funciones Lineales de Fisher.

En la primera fase del proyecto, sin embargo, se recomienda no abordar esta metodología de análisis de migraciones y concentrarse en la consolidación dentro de la organización de los segmentos y el lanzamiento con éxito de las primeras acciones identificadas.

Más adelante, una vez alcanzado este objetivo, adquiere mucho interés el análisis de las migraciones y puede ser realizado cada tres meses.

BIBLIOGRAFÍA

[1] Calvo Gomez, Félix 1993, Técnicas Estadísticas Multivariadas. Universidad de Deusro. Bilbao.

[2] Ezequiel Uriel Jiménez y Joaquín Aldás Manzano: 2005, Análisis Multivariante Aplicado. Thomson Editores Spain. Madrid – España

[3] García Santesmares, José M. 1994, Introducción a las técnicas del Análisis Multivariable Aplicadas a Ciencias Sociales. Centro de Investigaciones Sociológicas. Madrid.

[4] Hair, Anderson: Tathan, Black 1999, Análisis Multivariante. Prentice Hall Iberia SRL. Madrid.

[5] Jhonson, Richard A. ; Wichern, Dean W. 1992. Tercera Edición, Applied Multivariate Statistical Analysis. Prentice Hall, Englewood Cliffs. New Jersey.

[6] Pardo Merino , Antonio; Ruiz Días, Miguel Ángel 2002. Primera edición en español, SPSS 11. Guía para el análisis de datos.

[7] Tafur Portilla, Raul 1995. La tesis Universitaria.

Páginas Web

[8] www.estadístico.com/técnicasdeanálisismultivariante/

[9] www.campus5.com/análisiscluster/

[10] www.psicologia.com/tutorialsobreredesneuronalesartificiales/