

**UNIVERSIDAD NACIONAL DE INGENIERIA**  
**FACULTAD DE INGENIERÍA ECONÓMICA, ESTADÍSTICA Y CIENCIAS SOCIALES**



**“SEGMENTACIÓN DE CLIENTES EN BASE A SU COMPORTAMIENTO  
DE CONSUMO A TRAVÉS DEL MODELO DE SEGMENTACIÓN  
K-MEANS EN UNA ENTIDAD BANCARIA”**

**INFORME DE SUFICIENCIA PROFESIONAL**

**PARA OBTENER EL TÍTULO PROFESIONAL DE INGENIERO ESTADÍSTICO  
POR LA MODALIDAD DE ACTUALIZACIÓN DE CONOCIMIENTOS**

**Elaborado por:**

**JIMMY STALIN SALAZAR GEBOL**

LIMA – PERÚ  
2015

## **DEDICATORIA**

Porque siempre apoyaron incondicionalmente en mi desarrollo profesional, dedico este trabajo a mis padres:

A Tomasa Gebol (Mi madre)

A Hugo Salazar (Mi padre)

## **AGRADECIMIENTOS**

El desarrollo de este trabajo fue posible gracias al aporte y tiempo invertido de algunas personas.

A la Licenciada Magen Infante Rojas, asesora de tesina, quien con su conocimiento, tiempo y gran experiencia realizó las sugerencias y recomendaciones para el desarrollo exitoso de este trabajo.

A Roberto Samaniego Salcedo, jefe del área de CRM Analítico donde laboro, por haberme orientado en el desarrollo de la metodología de este trabajo y compartir conocimientos estadísticos aplicados al marketing que ayudó al desarrollo de este estudio.

## RESUMEN

La presente investigación plantea como objetivo determinar la segmentación de clientes más adecuada en base al comportamiento de consumo para obtener un mejor direccionamiento en las ofertas comerciales que se hacen en el banco a través de las campañas, así que se vio la necesidad en realizar este estudio debido a que muchas entidades no sabían a ciencia cierta si lo que ofrecían a sus clientes es lo que generalmente les interesaba, se usaron variedades de rubros de consumos que los clientes transaccionan, sin embargo, tener muchas variables pudieron dificultar las agrupaciones, por ello se usó el análisis factorial para la reducción de estas con la técnica de componentes principales, una vez obtenida los factores finales se realizó la primera división que conjuntamente con los experto en el tema de consumo de tarjeta de créditos se decidió que se evaluaría el segmento cibernauta que son aquellos clientes que en su mayoría realizan sus transacciones por el canal internet, ya con el resto de los clientes (tradicional, no usan con frecuencia el canal internet) se empezó a analizar la segmentación K-Means obteniendo el evolutivo de segmentación tomando como punto inicial a la división de 4 segmentos ( $K=4$ ), de esta forma junto a los cálculos de cohesión, distancia externa y silueta (forma) del segmento se determinó la

validación y la segmentación final que consta de 7 segmentos siendo estas las que mejor direccionan a las ofertas. Por otro lado se llevó estos segmentos hacia un análisis de visualización para tener la noción de que segmentos podrían tener características similares en caso se quiera que un cliente pueda trasladarse de un segmento a otro, es así que se usó el Mapa Auto-Organizado de Kohonen.

Palabras claves.- Componentes Principales, análisis factorial, segmentación K-Means, Mapa Auto-Organizado de Kohonen (SOM).

## ÍNDICE

DEDICATORIA .....	ii
AGRADECIMIENTOS .....	iii
RESUMEN .....	iv
CAPITULO I.....	1
1. INTRODUCCION .....	1
CAPITULO II.....	3
1. PLANTEAMIENTO DEL PROBLEMA .....	3
2. PROBLEMA DE INVESTIGACION .....	5
2.1 Problema general .....	5
2.2 Problema específico.....	5
3. OBJETIVO DE INVESTIGACION .....	6
3.1 Objetivo general .....	6
3.2 Objetivo específico .....	6
4. JUSTIFICACION.....	7
4.1 Justificación teórica:.....	7
4.2 Justificación práctica: .....	8
5. DELIMITACIÓN.....	9
6. ANTECEDENTES.....	10
CAPITULO III.....	13
1. MARCO TEORICO .....	13
1.1 Revisión de literatura.....	13
1.2 Bases teóricas .....	16
1.2.1. Reducción de variables .....	16
1.2.2. Componentes principales y análisis factorial.....	18
1.2.3. Análisis Clúster.....	28
1.2.4. Segmentación K-Means.....	31
1.2.5. Análisis de Mapa de visualización de Kohonen.....	41

2. DEFINICION DE TERMINOS.....	43
CAPITULO IV.....	46
1. METODOLOGIA .....	46
CAPITULO V.....	64
1. APLICACIÓN.....	64
1.1 Segmentación de primer público objetivo .....	64
1.2 Segmentación de segundo público objetivo .....	69
1.2.1. Grupo Cibernauta .....	69
1.2.2 Grupo Tradicional (No cibernauta) .....	77
2. RESULTADOS .....	106
2.1 Segmentos Finales .....	106
2.2 Análisis de vecindades entre segmentos.....	107
2.3 Proceso de Réplica o Actualización de segmentos .....	108
CONCLUSIONES .....	110
RECOMENDACIONES .....	111
BIBLIOGRAFIA .....	112
ANEXO I .....	114
ANEXO II .....	118

## ÍNDICE DE GRÁFICOS

GRÁFICO 3.1	Proceso del análisis factorial .....	20
GRÁFICO 3.2	Primer centroide (referencial).....	33
GRÁFICO 3.3	Segundo centroide (K=2).....	34
GRÁFICO 4.1	Prmer público objetivo para el estudio de segmentación .....	48
GRÁFICO 4.2	Segundo público objetivo para el estudio de segmentación .....	49
GRÁFICO 4.3	Dimensión de segmentación .....	51
GRÁFICO 4.4	Proceso de reducción de variables(rubro de consumo) .....	52
GRÁFICO 4.5	Componentes por autovalores I.....	53
GRÁFICO 4.6	Componentes por autovalores II.....	56
GRÁFICO 4.7	Factores finales.....	60
GRÁFICO 4.8	Etapas de segmentación .....	62
GRÁFICO 5.1	K-Means K= 4 segmentos P1.....	64
GRÁFICO 5.2	K-Means K= 5 segmentos P1.....	65
GRÁFICO 5.3	K-Means K= 6 segmentos P1.....	66
GRÁFICO 5.4	K-Means K= 8 segmentos P1.....	66
GRÁFICO 5.5	Segmento para determinar al grupo cibernauta I .....	70
GRÁFICO 5.6	Segmento para determinar al grupo cibernauta II.....	71
GRÁFICO 5.7	Segmento para determinar al grupo cibernauta III.....	72
GRÁFICO 5.8	Segmento para determinar al grupo cibernauta IV .....	73
GRÁFICO 5.9	Segmento para determinar al grupo cibernauta V .....	74
GRÁFICO 5.10	Distribución de clientes cibernautas y tradicionales.....	75
GRÁFICO 5.11	K-Means K= 4 segmentos .....	77
GRÁFICO 5.12	K-Means K= 5 segmentos .....	78
GRÁFICO 5.13	K-Means K= 6 segmentos .....	78
GRÁFICO 5.14	K-Means K= 8 segmentos .....	79
GRÁFICO 5.15	K-Means K= 10 segmentos .....	80
GRÁFICO 5.16	Variables del segmento familia .....	88
GRÁFICO 5.17	Variables del segmento conservador .....	91

<b>GRÁFICO 5.18</b>	<b>Variables del segmento viajero .....</b>	<b>94</b>
<b>GRÁFICO 5.19</b>	<b>Variables del segmento fashion .....</b>	<b>97</b>
<b>GRÁFICO 5.20</b>	<b>Variables del segmento progresista .....</b>	<b>100</b>
<b>GRÁFICO 5.21</b>	<b>Variables del segmento sofisticado.....</b>	<b>103</b>
<b>GRÁFICO 5.22</b>	<b>Segmentación de clientes por transacción de consumo .....</b>	<b>106</b>
<b>GRÁFICO 5.23</b>	<b>Mapa de visualización de segmento de clientes.....</b>	<b>107</b>
<b>GRÁFICO 5.24</b>	<b>Proceso de réplica de segmentación de clientes.....</b>	<b>108</b>

## ÍNDICE DE CUADROS

<b>CUADRO 4.1 Rubros de consumo de tarjeta de crédito .....</b>	<b>52</b>
<b>CUADRO 4.2 Canal de consumo de tarjeta de crédito.....</b>	<b>52</b>
<b>CUADRO 4.3 KMO and Bartlett's Test I.....</b>	<b>54</b>
<b>CUADRO 4.4 Total Variance Explained I.....</b>	<b>55</b>
<b>CUADRO 4.5 KMO and Bartlett's Test II.....</b>	<b>56</b>
<b>CUADRO 4.6 Communalities .....</b>	<b>56</b>
<b>CUADRO 4.7 Total Variance Explained II.....</b>	<b>58</b>
<b>CUADRO 4.8 Rotated Component Matrix.....</b>	<b>59</b>
<b>CUADRO 5.1 Evaluación del segmento – primera propuesta .....</b>	<b>67</b>
<b>CUADRO 5.2 Proporción de transacciones por canal Internet .....</b>	<b>76</b>
<b>CUADRO 5.3 Evaluación del segmento tradicional .....</b>	<b>82</b>
<b>CUADRO 5.4 Comparación del segmento por propuesta.....</b>	<b>83</b>
<b>CUADRO 5.5 Proporción de transacciones de consumo por segmento .....</b>	<b>84</b>
<b>CUADRO 5.6 Distancia de segmentación .....</b>	<b>86</b>
<b>CUADRO 5.7 Proporción de transacciones familia.....</b>	<b>88</b>
<b>CUADRO 5.8 Descripción sociodemográfica familia .....</b>	<b>89</b>
<b>CUADRO 5.9 Descripción de sistema financiero familia .....</b>	<b>90</b>
<b>CUADRO 5.10 Proporción de transacciones conservador .....</b>	<b>91</b>
<b>CUADRO 5.11 Descripción sociodemográfica conservador .....</b>	<b>92</b>
<b>CUADRO 5.12 Descripción de sistema financiero conservador .....</b>	<b>93</b>
<b>CUADRO 5.13 Proporción de transacciones viajero.....</b>	<b>94</b>
<b>CUADRO 5.14 Descripción sociodemográfica viajero .....</b>	<b>95</b>
<b>CUADRO 5.15 Descripción de sistema financiero viajero .....</b>	<b>96</b>
<b>CUADRO 5.16 Proporción de transacciones fashion.....</b>	<b>97</b>
<b>CUADRO 5.17 Descripción sociodemográfica fashion.....</b>	<b>98</b>
<b>CUADRO 5.18 Descripción de sistema financiero fashion.....</b>	<b>99</b>
<b>CUADRO 5.19 Proporción de transacciones progresista.....</b>	<b>100</b>
<b>CUADRO 5.20 Descripción sociodemográfica progresista .....</b>	<b>101</b>

<b>CUADRO 5.21 Descripción de sistema financiero progresista .....</b>	<b>102</b>
<b>CUADRO 5.22 Proporción de transacciones sofisticado .....</b>	<b>103</b>
<b>CUADRO 5.23 Descripción sociodemográfica sofisticado.....</b>	<b>104</b>
<b>CUADRO 5.24 Descripción de sistema financiero sofisticado.....</b>	<b>105</b>

## **CAPITULO I**

### **1. INTRODUCCION**

El presente trabajo tuvo como objetivo presentar el uso de la técnica de segmentación K-Means en clientes de las entidades bancarias, conjuntamente con el análisis de relación de segmentos ayudaron a interpretar con mayor facilidad como dichos segmentos se trasladan entre ellos.

El desarrollo de este trabajo es de suma importancia para las entidades bancarias ya que fue una herramienta para evitar gastos de ofertas hacia clientes que serían muy propensos a no usarlas.

El documento se organiza de cinco capítulos: en el Capítulo I, constituye a la organización del documento. En el Capítulo II, se presenta qué es lo que causó realizar esta investigación, cuál fue el motivo tomando como explicación general y específicas, por otro lado se define el propósito o lo que se quiso lograr, así como que se esperó una vez finalizada la investigación, también se describe la utilidad, las limitaciones al desarrollarlas y los acontecimientos que ocurrieron en relación al tema de segmentación de clientes.

En el Capítulo III, se presenta el Marco Teórico necesaria para el desarrollo de la investigación que consta de la revisión de documentación e información relacionada a la técnica de segmentación, por otro lado se presenta las terminologías que se usa en las entidades bancarias. En el Capítulo IV, se refiere al proceso que seguiría para el desarrollo de la investigación, definiendo la población objetivo, el tratamiento de la información, selección de variables, etc.

En el Capítulo V, constituye a los resultados obtenidos luego del tratamiento de la información y aplicada la técnica de segmentación K-Means.

Por otro lado tenemos a las conclusiones que se obtuvo de la segmentación final conjuntamente con el análisis de relación de segmentos, también se detallaron las recomendaciones para la continuidad de algún trabajo posterior que ayudaría a complementar mas esta investigación.

Finalmente se presentan las correspondientes Referencias Bibliográficas y Anexos.

## **CAPITULO II**

### **1. PLANTEAMIENTO DEL PROBLEMA**

Durante los dos últimos años la importancia de conocer mejor a su cliente en las entidades bancarias era muy esencial ya que eso les permitía saber qué es lo que ellos querían y no que es lo que las entidades querían para el cliente, estudios anteriores realizaban agrupamiento de perfiles mediante técnicas jerárquicas que en teoría eran más sencillas de manejar, sin embargo, se dieron cuenta que cada mes la facturación y la frecuencia de consumo aumentaba en los diferentes rubros y necesitaban alguna herramienta más precisa, por otro lado las áreas de campañas gestionaban por separado a sus clientes el cual no tenían indicadores generales que les permitía saber cuan efectivo fue sus campañas con respecto a los otros, el único sustento de cómo enviaban las ofertas a los clientes era mediante análisis descriptivo y no usaban técnicas adecuadas que les ayude a segmentar en los diferentes rubros que existe en el mercado,

Los datos históricos informan que hace dos años la participación de mercado (**Market Share**) de consumo en el Perú ha aumentado hasta un 20.26% y la facturación de consumo en algunas entidades líderes aumentó hasta un 36.4%, así como el incremento de personas que ingresaron al sistema financiero peruano, por ello, se enfocaron en tener sus herramientas de segmentación para realizar la fidelización de los clientes y no permitir que la competencia pueda adueñarse de ellos.

Por otro lado, el incremento de emisión de tarjetas de crédito llamó la atención de las entidades, ya que una vez que el cliente fuera captado el siguiente paso sería la activación de la tarjeta, que sólo se lograría realizando consumos y para eso se necesitó saber que rubros de consumo le interesa al cliente.

Según la Superintendencia de Banca, Seguros y AFP la participación de consumo (tarjeta de crédito) desde Setiembre del 2013 hasta setiembre del 2014 ha aumentado en un 19.26% lo que genera mayor interés por parte las entidades en invertir investigación de comportamiento de consumo de sus clientes.

**Fuente: SBS (Artículo Sep-2014)**

El número de tarjetas de crédito (crédito de consumo) en Setiembre 2013 hasta Setiembre 2014 tuvo un incremento del 5.3% lo que también causó interés por parte de las entidades.

**Fuente: ASBANC (Publicación Sep-2014)**

## **2. PROBLEMA DE INVESTIGACION**

### **2.1 Problema general**

Ineficiencia en la toma de decisiones comerciales de la empresa debido a la falta de conocimiento del comportamiento de consumo de sus clientes.

### **2.2 Problema específico**

2.2.1 Necesidad en determinar tipos de segmentos de clientes a través de sus rubros de consumo en base a técnicas estadísticas.

2.2.2 Limitación en direccionar ofertas comerciales debido a la carencia de segmentos enfocados con más detalle al comportamiento de consumo.

2.2.3 Carencia de alguna metodología que permita direccionar las ofertas comerciales a través de los segmentos.

### **3. OBJETIVO DE INVESTIGACION**

#### **3.1 Objetivo general**

Contribuir a optimizar las decisiones comerciales de la empresa con los actuales clientes, explorando el comportamiento de consumo.

#### **3.2 Objetivo específico**

3.2.1 Identificar los tipos de segmentos de clientes a través de su comportamiento de consumo aplicando las técnicas estadísticas de agrupamiento multidimensional.

3.2.2 Describir las características de los segmentos hallados en relación al comportamiento de consumo de los clientes.

3.2.3 Proponer una metodología de segmentación de clientes que permita direccionar las actividades comerciales de la empresa hacia dichos segmentos.

## **4. JUSTIFICACION**

### **4.1 Justificación teórica:**

Este trabajo se realizó debido a que muchas técnicas de segmentación K-Means han logrado resultados que les ayuden en su gestión en otras organizaciones distintas a las entidades financieras tales como: en área de salud pública en México, donde el objetivo fue encontrar grupos homogéneos de ciudadanos en distintos municipios que tengan alta tasa de mortalidad por cáncer de pulmón y estómago, obteniendo que para el cáncer de pulmón se manifestaban en grupos de municipios con alta tasa de mortalidad en la región norte y noreste y para el cáncer de estómago se identificó un grupo con alta mortalidad en la región sureste.

Otro caso fue en la cadena de supermercados Tesco en Chile que necesitaban saber que segmentos con respecto al estilo de vida de sus clientes que realizan con frecuencia compras de algunos productos, la información lo obtenían de las boletas de compra y mediante la técnica K-Means identificaron segmentos como: Parrillero, Familiar, Saludables, Gourmet, finalmente observaron que la frecuencia de compra era semanal para la mayoría de clientes.

Así como los casos mencionados, en las entidades financieras generalmente las usaban como segmentación de perfiles, o realizaban

estudios mediante técnicas de segmentación jerárquicas, sin embargo, debido a la importancia y potencia en identificar grupos con las técnicas no jerárquicas el modelo de segmentación multivariado K-Means se usó conjuntamente con información de consumo de clientes el cual nos dará a conocer el comportamiento del cliente más preciso.

#### **4.2 Justificación práctica:**

Esta investigación será útil para las entidades bancarias que estén interesadas en conocer el comportamiento de consumo de sus clientes que usan la tarjeta de crédito, esto con la finalidad de poder fidelizar o aumentar la activación de la tarjeta de crédito de los clientes, por otro lado es importante para direccionar las ofertas que la empresa desea realizar ya sea para las metas en sus campañas de consumo o para aumentar la efectividad de consumo en sus campañas, por ello el modelo de segmentación obtenido ayudaría a saber que clientes son asignados en los diferentes tipos de segmentos.

Por otro lado sería de mucha utilidad para el desarrollo del modelo de perfiles, esto para aquellos clientes que no tuvieran información de consumo y solo se determinarían en base a características generales tales como: demográficas, sociodemográficas, etc.

## 5. DELIMITACIÓN

- Se ha delimitado sólo a personas naturales con tarjeta de crédito vigente.
- El estudio sólo está referido a una entidad bancaria considerada una de las entidades líderes en el mercado peruano.
- El tiempo de registro para la evaluación es de tres meses los cuales los movimientos de consumo son regulares evitando los consumos atípicos para no sesgar la medición y posteriores pruebas estadísticas.
- Aunque no hubo delimitación por nivel socioeconómico, el estudio restringió a los clientes cuya tarjeta de crédito tenga una antigüedad menor de doce meses.
- Solo considera aquellos clientes cuya participación de consumo con su tarjeta de crédito en la entidad bancaria tomada en el estudio sea considerable (mayor al 30%) con respecto al resto de las entidades del sistema financiero.
- El estudio está referido sólo a clientes que hayan realizado por lo menos doce transacciones de consumo en los doce últimos meses.
- El comportamiento de consumo de los clientes esta en base a las transacciones y no por facturación (Monto).

## 6. ANTECEDENTES

En esta sección describiremos algunos trabajos relacionadas con la técnica, variables u objetivos que se desarrolló en esta investigación.

Según **Johana C. y Lina O. (2008)**, presentando su investigación en el XVIII Simposio Colombiano de Estadística el tema de “Segmentación de los solicitantes de crédito a un banco usando análisis factorial múltiple”, donde para diferenciar a los solicitantes de crédito de un banco en la ciudad de Bogotá utilizaron características sociodemográfica y económicas, si bien usaron las técnicas de análisis factorial múltiple para la agrupación de variables, la segmentación K-Means fue la determinante para diferenciar en tres clases de solicitantes dependiendo del porcentaje de clientes y en sus distribuciones de tipos de crédito.

**Muñoz L. (2009)**, en la revista de estudios empresariales de la universidad de Granada publicó la “Caracterización de los clientes de banca electrónica” esto a causa de la inexistencia de trabajos que relacionen los servicios de banca electrónica, aspectos comportamentales y actitudinales, se enfocó en aplicar la técnica de segmentación K-Means para obtener una clasificación de segmentos mas diferenciadas el cual obtuvo cuatro categorías de usuarios de banca electrónica como usuarios: ocasionales, activos, habituales y esporádicos.

Por otro lado **Santiago N. (2010)**, uso la técnica K-Means y ACP usando comportamiento de consumo e información demográfica en una cadena de supermercado el cual pudo diferenciar las clase de clientes como: Tradicional, Familiar, Parrillero, Express, Gourmet y Saludables.

**Ramsler (1982)**, trató de identificar distintos posicionamientos estratégicos de los cien bancos mas grandes del mundo, en términos de diversificación internacional. Aplicando el algoritmo K-Means identifica seis grupos en la banca internacional como gran banca: Global, internacional, internacional media, internacional especifica, domestica mediana y domestica específica.

**Rudolph y Topping (1991)**, en la Industria Financiera Americana se dedicaron a identificar los grupos de entidades financieras, en particular, Cajas de Ahorro y Cooperativas de Crédito, la metodología consistió en tres etapas donde una de ellas es la aplicación del algoritmo K-Means identificaron once grupos estratégicos que comparando con variables estratégicas identificado con conductas anticipada de sus grupos hipotéticos obtuvieron cuatro tipos de segmentos tales como: Tradicional, tradicional modificada, ajuste de vencimiento y no tradicional.

**Gual y Vives (1990)**, identificando los grupos estratégicos en los bancos, describieron sus hipótesis donde mencionan que el proceso de cambio experimentado por el entorno de las entidades, ha llevado a los bancos a seguir cierta especialización distinta en el negocio bancario.

Aplicando el algoritmo K-Means con variables de inversión financiera, cuentas corrientes, créditos, cuentas de ahorros y cartera de valores, obtuvieron tres grupos asimilables como: Banca comercial, banca; por mayor y banca corporativa. Por consiguiente se constata la existencia de grupos dados por la afinidad de especialización.

Según **Kenny M. (2014)**, en su investigación “Sistema para caracterización de perfiles de clientes de la empresa Zona T” destaca la importancia de clasificar a los clientes mediante la metodología de minería de datos usando el algoritmo de segmentación K-Means, donde obtuvo como resultado un nuevo algoritmo de minería de datos y que la empresa tenga un mejor conocimiento de las características y conductas de sus consumidores y mejorar la atención de los mismos.

## **CAPITULO III**

### **1. MARCO TEORICO**

Antes de realizar la técnica estadística de segmentación K-Means, en la preparación de los datos se uso el método de reducción de variables esto para tener facilidad en la interpretación una vez culminada la segmentación, luego se realiza la técnica K-Means y por último la relación entre los segmentos finales el análisis de vecindad de Kohonen.

A continuación revisaremos conceptos relacionado al tema de investigación de algunos autores.

#### **1.1 Revisión de literatura**

Para el uso de reducción de variables se usó el método de componente principal.

Según **Johnson y Wichern (1998)**, el análisis de componente principal suele ser muy útil para identificar variables que estén altamente correlacionadas o que tengan grados de asociación significantes, lo que llamamos multicolinealidad entre variables, en caso se detectaría podría

omitir dichas variables esto para evitar la redundancia de información en el estudio.

Por otro lado, **Flury (1988)**, indica que transformar un grupo de variables correlacionadas en otras variables independientes, encontrar combinaciones lineales a partir de variables con alta o poca variabilidad y considerar una herramienta indicada para reducir las variables originales, son los tres puntos esenciales de las componentes principales (ACP).

Frente a la reducción de dimensiones de un gran número de variables, **Lindsay I. Smith (2002)**, nos indica que las componentes principales también que grupos explican mayor o poca variabilidad del conjunto de variables en estudio, por ello toma interés los indicadores de similitud entre las variables.

Ahora citaremos información con respecto al análisis clúster o conglomerado antes de iniciar la segmentación K-Means.

Según **Johnson S.C. (1967)**, el análisis clúster es una técnica multivariada que se encarga de clasificar a los individuos como grupos homogéneos, por otro lado indica la diferencia entre el análisis discriminante y el análisis clúster donde para el análisis discriminante los grupos son conocidos y lo que se quiere es que las variables que intervienen discriminen los grupos ya existentes, sin embargo, en el

análisis clúster no existe grupos conocidos y lo que se quiere es determinar dichos grupos.

Para el caso de la segmentación K-Means, a veces surge una confusión ya que existen dos métodos distintos con el mismo nombre, por ello:

**Forgy (1965)**, propuso el primer método que consiste básicamente en sucesiones iteraciones hasta obtener las siguientes características de convergencia: representar un grupo por su centro de gravedad, en otras palabras por su vector de promedios, asignar los elementos al grupo cuyo centro de gravedad es la más cercana.

**McQueen (1967)**, propone un método similar a la de Forgy, el cual se representan los grupos por su centro de gravedad, y se examina a cada elemento para asignarlo al grupo más cercano, sin embargo, a diferencia del método Forgy una vez que un elemento fue asignado a un grupo, de inmediato los centros de gravedad se recalculan, mientras que Forgy espera asignar a todos los elementos para luego calcular el nuevo centro, generalmente el método K-Means también es llamado como nube dinámicas, centros móviles, etc. según Mcqueen.

Luego de definir la reducción de variables, los análisis de componentes y la segmentación K-Means, ahora se definirá que relación tendrían segmentos finales.

Según **T. Kohonen (1982)**, nos indica que la técnica de los Mapas Auto-organizados de Kohonen (SOM) consiste en descubrir rasgos comunes, regularidades, correlaciones en la información (similitud) produciendo una fácil visualización en las relaciones de segmentos.

## **1.2 Bases teóricas**

### ***1.2.1. Reducción de variables***

Una de las dificultades que se presentan en una investigación es tener gran cantidad de variables para trabajarlas, es por ello que el investigador puede complicarse al momento de interpretar sus resultados o al momento de realizar el tratamiento de su información con alguna técnica estadística.

Como indica **Stefan Bathe (1994)**, director adjunto de SDR Consulting, el análisis factorial nos permite observar grupos de variables que tienden a correlacionarse entre sí, e identificar las dimensiones que explican dichas correlaciones. Teniendo en cuenta que el agrupamiento implicaría perder cierta información real de las variables originales.

Según **D. Peña (2002)**, la mayor utilidad de los métodos de reducción de variables consiste en que:

- La dimensión de una matriz de información  $X$  sea  $m$ , el método de reducción de variables permite a la información presentar en un espacio adecuado de dimensión más reducida  $k < m$ , donde  $k$  es la dimensión de las nuevas variables.
- Con respecto a la redundancia de variables, el método de reducción de variables permite transformar las variables originales que estuvieron correlacionados en nuevas variables icorrelacionadas.

Por ello frente a estos problema podríamos estar identificando y tratando a las variables fuertemente correlacionadas que estaría redundando información, es por ello que los métodos de reducción de variables ayudan a identificar dichos problemas creando nuevos grupos de variables (factores) como el análisis factorial.

### **1.2.2. Componentes principales y análisis factorial**

Las técnicas de estudio de relaciones entre variables, fueron técnicas desarrolladas inicialmente en el siglo XIX por **Pearson** y ya en los años 30 del siglo XX fueron estudiadas por **Hotelling**.

Durante el proceso de preparación de base, suele suceder un problema al trabajar con muchas variables, esto porque podríamos caer en multicolinealidad el cual nos indica que existen variables que se encuentran correlacionadas y por tanto son informaciones redundantes, es por ello que los componentes principales es un método que ayuda a reducir la dimensionalidad en nuestra información de estudio, a costa que perderíamos una pequeña parte de nuestra información.

Es por ello que el objetivo del análisis de componentes principales (ACP) es resumir un grupo alto de variables en un nuevo conjunto sin perder cierta parte de información original significativa, **(Chuvieco 1996).**

Ahora para determinar las componentes principales, consideremos un espacio de dimensión  $k$  cuya vectores propias se representan como  $\mathbf{a}$  asociados a la matriz de Covarianza  $\mathbf{S}$ . Tener en cuenta que estas direcciones se denominan *Direcciones principales* de la

información en estudio y a las nuevas variables *Componentes principales*.

Por tanto sea la matriz de Información  $\mathbf{X}$  tiene rango  $p$  existiendo tantas variables como componentes principales, obtuvimos los valores propios  $\lambda_1, \dots, \lambda_p$ , de la matriz de varianza y covarianza  $\mathbf{S}$  mediante:

$$|\mathbf{S} - \lambda\mathbf{I}| = 0$$

Luego sus vectores asociados:

$$(\mathbf{S} - \lambda_1\mathbf{I})\mathbf{a}_i = 0$$

Los términos  $\lambda_i$ , son reales cuando  $\mathbf{S}$  es simétrica y positiva, ya que  $\mathbf{S}$  es definida positiva. Entonces por ser  $\mathbf{S}$  simétrica si  $\lambda_i, \lambda_j$ , son distintos entonces sus vectores serán ortogonales.

Sea  $\mathbf{M}$  la matriz donde incluye las  $p$  componentes principales de todos los registros de la base de información en sus columnas, entonces estas nuevas variables se relacionan con las variables iniciales.

$$\mathbf{M} = \mathbf{XA}$$

Donde

$$\mathbf{A}^T\mathbf{A} = \mathbf{I}$$

Luego para obtener las componentes principales realizamos una transformación ortogonal  $\mathbf{A}$  a las variables  $\mathbf{X}$  (ejes iniciales), esto para crear las nuevas variables correlacionadas entre sí, generalmente a ello se llama rotación ortogonal.

Frente a lo definido a nivel general, ¿qué hace el análisis factorial?

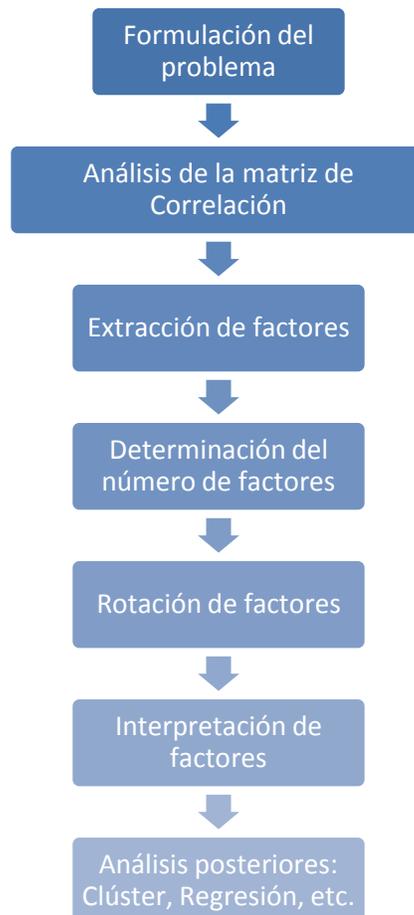
El análisis factorial se encarga de analizar la *varianza común* a todas las variables, partiendo de la matriz de correlaciones.

Los dos enfoques que caben en el análisis factorial son:

1- Analizar “**toda**” la varianza común y no común, el método más usual es el de **Análisis de Componentes Principales**.

2- Analizar solo la varianza común, estimar la varianza que cada ítem tiene en común con los demás (se denominan comunalidades), el esquema de análisis factorial sigue el siguiente proceso.

**Gráfico 3.1: Proceso del análisis factorial**



Fuente: Propio

Como se explicaba los problemas que acontece en un estudio es la cantidad numerosa de variables, que entre ellas podrían estar correlacionadas o no, por ello se sigue un aserie de pruebas para confirmar si es factible o no desarrollar un análisis factorial.

**(a) Análisis de la matriz de correlación:**

La finalidad de analizar la matriz de correlaciones es comprobar si sus características son las adecuadas para realizar un análisis factorial.

Test de esfericidad de Barlett.

Contrasta bajo la hipótesis de normalidad multivariante, si la matriz de correlación de las **p** variables observadas (*p*) es la identidad.

Si una matriz de correlación es la identidad significa que las intercorrelaciones entre las variables son cero.

Siendo la hipótesis nula:

$$H_0: |R_p| = 1$$

$$R_p = I$$

De confirmarse la hipótesis nula entonces afirmaremos que las variables no están intercorrelacionadas..

El test de esfericidad se obtiene mediante la transformación de la determinante de la matriz de correlación.

El estadístico del test esta dado por:

$$d_n = - \left[ n - \frac{(2p + 11)}{6} \right] \sum_{j=1}^p \log(\lambda_j)$$

Donde  $\lambda_j (j = 1, \dots, p)$  son los valores propios de **R** y **n** es el número de individuos de la muestra. Bajo la hipótesis nula, el estadístico se distribuye asintóticamente según una  $X^2_{\left(\frac{p(p-1)}{2}\right)}$ .

#### Medidas de adecuación de la muestra.

El coeficiente de correlaciones parciales es un indicador del grado de relaciones entre dos variables, eliminando la influencia del otro.

Si existe un número elevado de coeficientes de correlaciones parcial distintas de cero, se interpreta que las hipótesis del modelo factorial no son compatibles con los datos.

Con la **Medida de Adecuación de la Muestra KMO** propuesta por Kaiser-Meyer-Olkin se cuantifica este hecho.

$$KMO = \frac{\sum_{j \neq i} \sum_{i \neq j} r^2_{ij}}{\sum_{j \neq i} \sum_{i \neq j} r^2_{ij} + \sum_{j \neq i} \sum_{i \neq j} r^2_{ij(p)}} \quad 0 \leq KMO \leq 1$$

Donde  $r_{ij(p)}$  es el coeficiente de correlación parcial entre (Xi, Xj) eliminando la influencia del resto de las variables.

El índice **KMO** se utiliza para comparar las magnitudes de los coeficientes de correlación parcial, de forma que cuanto más pequeño sea su valor, mayor será el valor de los coeficientes de correlación parciales  $r_{ij(p)}$  y, en consecuencia menos apropiado es realizar un análisis factorial.

Para realizar un análisis factorial, Kaiser-Meyer-Olkin propone:

KMO $\geq$ 0.75	entonces “ <b>Bien</b> ”
KMO $\geq$ 0.50	entonces “ <b>Acceptable</b> ”
KMO $<$ 0.50	entonces “ <b>Inacceptable</b> ”

**(b) Extracción de factores:**

El objetivo del Análisis Factorial (AF) es determinar un número reducido de factores que pueden representar a las variables originales.

El modelo factorial en forma matricial:

$$X = FA' + U$$

Teniendo que cuantificar la matriz **A** de cargas factoriales que explica **X** en función de los factores.

Partiendo del modelo factorial se deduce la llamada *Identidad Fundamental del análisis Factorial*:

$$R_p = AA' + \Psi$$

Donde  $R(p)$  es la matriz de correlación poblacional de las variables

$$(X_1, X_2, \dots, X_p)$$

y

$$\Psi = \text{diag}(\psi_i)$$

Es la matriz diagonal de las especificidades.

El método que se utilizara para la extracción de los factores es la técnica de **Análisis de Componentes Principales**, donde para determinar la proporción de varianza explicada de los factores se basa en los cálculos de sus **eigenvalores** y **comunalidades**.

### **Eigenvalores:**

El cuadrado de una carga factorial indica la proporción de la varianza explicada por un factor en una variable particular.

La suma de los cuadrados de los pesos de cualquier columna de la matriz factorial es lo que denominaremos eigenvalores ( $\lambda$ ), indica la cantidad total de varianza que explica ese factor para las variables consideradas como grupo.

Las cargas factoriales pueden tener como valor máximo 1, por tanto el valor propio como máximo puede alcanzar igual al número de variables.

$$\lambda_1 = P^2_{11} + P^2_{12} + P^2_{13} + \dots + P^2_{1j}$$

$$\lambda_2 = P^2_{21} + P^2_{22} + P^2_{23} + \dots + P^2_{2j}$$

$$\lambda_i = P^2_{i1} + P^2_{i2} + P^2_{i3} + \dots + P^2_{ij}$$

Si dividimos el valor propio entre el número de variables, nos indica la proporción de varianza explicada por el factor.

$$\frac{\lambda_1}{n} \equiv \text{Varianza explicada por primer factor}$$

$$\frac{\lambda_2}{n} \equiv \text{Varianza explicada por segundo factor}$$

### **Comunalidades;**

La comunalidad es representada por la proporción de varianza explicada por los factores comunes en una variable.

La denotaremos como (**h**) (comunalidad) que es la suma de los pesos factoriales al cuadrado en cada una de las filas de la matriz.

La comunalidad no se puede saber hasta que se conocen los factores, este resultado es uno de los problemas del análisis factorial.

Teniendo en cuenta que en el análisis de componentes principales no se supone la existencia de ningún factor común la comunalidad toma como valor inicial **1**.

En otros métodos se utilizan diferentes maneras de estimar la comunalidad inicial.

- Estimando la comunalidad por la mayor correlación en la fila *i*-ésima de la matriz de correlaciones.
- El promedio de los coeficientes de correlación de una variable con todas las demás.
- Calculando a partir de los dos coeficientes mayores de esa variable la siguiente operación:

$$h^2 = \frac{r_{xy}r_{xz}}{r_{yz}}$$

La comunalidad final de cada variable viene dada por:

$$h^2 = P_{1j}^2 + P_{2j}^2 + \dots + P_{kj}^2$$

**(c) Rotación de factores:**

Los ejes se rotan de forma que quede preservada la incorrelación entre los factores. Es decir, los nuevos ejes (ejes rotados) son perpendiculares de igual forma que lo son los factores sin rotar.

La rotación se apoya en el problema de falta de identificabilidad de los factores obtenidos por rotaciones ortogonales, de forma que si  $TT' = T'T = I$ , entonces:

$$X = FA' + U = FTT'A' + U = GB' + U$$

De la ecuación **G** es la rotación de **F**.

Realmente lo que realiza es un giro de ejes, de forma que cambian las cargas factoriales y los factores.

Los métodos empleados en la rotación ortogonal de factores son: *Varimax, Quartimax, Equamax, Oblimin y Promax.*

¿En qué consiste el método Varimax?

El método Varimax es un método de rotación que minimiza el número de variables con cargas altas en un factor, mejorando así la interpretación de los factores.

### **1.2.3. Análisis Clúster**

El análisis clúster o análisis de conglomerados es el análisis de agrupamiento o clasificación que consiste en una técnica exploratoria de información, su agrupación consiste en el grado de asociación o similitud de sus elementos dentro de cada clúster.

Por tanto el análisis clúster descubre las asociaciones que tal vez no sean muy evidentes pero una vez identificados pudieron ser útiles.

Generalmente los pasos para generar los clúster son:

- De la base en estudio para todos los registros se debe tener información en todas sus variables (de los  $n$  casos y  $p$  variables).
- Mediante los indicadores de distancia o similaridad se identifica que observaciones se parecen entre si.
- Mediante los indicadores de distancia o similaridad se crearan los grupos usando los distintos métodos de agrupamiento.
- Luego los grupos creados se comparan y se le asignan una descripción característico del grupo (clúster).

Las medidas de similitud están determinadas por: la distancia euclidiana (**D**).

$$D_{ic} = \sqrt{\sum (X_{ij} - X_{ic})^2}$$

Donde: i: elemento i-ésimo

j: variable j-ésimo

c: centroide del clúster

Otra media es la distancia de Minkowski (**M**).

$$M_{ic} = \left[ \sum (X_{ij} - X_{ic})^2 \right]^{1/n}$$

Donde: i: elemento i-ésimo

j: variable j-ésimo

c: centroide del clúster

En la formación de grupo los clúster pueden someterse a los siguientes análisis:

- **Análisis jerárquico:** Este método inicia siendo cada caso un grupo, luego se van fusionando sucesivamente observando que elemento se encuentra más cerca de un grupo hasta conformar el clúster final.
- **Análisis no jerárquico:** muy por el contrario al método jerárquico, el no jerárquico necesita la asignación de un número de grupo inicial (referencia) de allí cada elemento se irá incorporando en uno de los grupos ya asignados.

#### **1.2.4. Segmentación K-Means**

El clustering de datos es una técnica de exploración de datos que busca objetos de similar características para ser agrupadas en orden para facilitar sus procesos.

La segmentación K-Means es una técnica multivariado que está clasificado como método de agrupamiento no jerárquico, el cual como se definió, consiste en la asignación inicial de un número de grupos (referencia).

Supongamos que  $\Omega$  es el conjunto de  $n$  individuos que queremos agrupar, todos con el mismo peso de ser clasificados y supondremos están compuestas por  $p$  variables cuantitativas  $x_1, x_2, \dots, x_p$ .

Debido a que las variables son cuantitativas, tiene sentido realizar los cálculos de promedios y distancia **Euclidea**.

Por ello tiene sentido que cada grupo este presentado por el centro de gravedad, esto es, para cada individuo ficticio cuyas coordenadas están representadas por los promedios de las variables del individuo asignados a un grupo.

Generalmente se usara la distancia Euclidea clásica.

El proceso del método:

Escogemos  $k$  objetos de  $\Omega$ , que servirán como núcleos iniciales.

Esto es escoger aleatoriamente  $a_1, a_2, \dots, a_k$  en  $\Omega$  sean los centroides:

$$C_1 := \Phi, \dots, C_k := \Phi$$

Se asigna a cada individuo a un grupo con centro de gravedad más cercano, es decir para todo  $i \in \Omega$ , se realiza:

$$si \ d(x_i, g_{k^*}) \leq \{d(x_i, g_k) \text{ para todo } k = 1, \dots, K\}$$

Entonces se asigna  $x_i$  al grupo  $C_{k^*}$ : en el caso que dos grupos alcancen el mínimo, entonces se escoge el de menor índice.

Ahora se calculan los centros de gravedad de la partición.

$$\text{Así, para todo } k \in \{1, \dots, K\} \text{ hacer: } g_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i$$

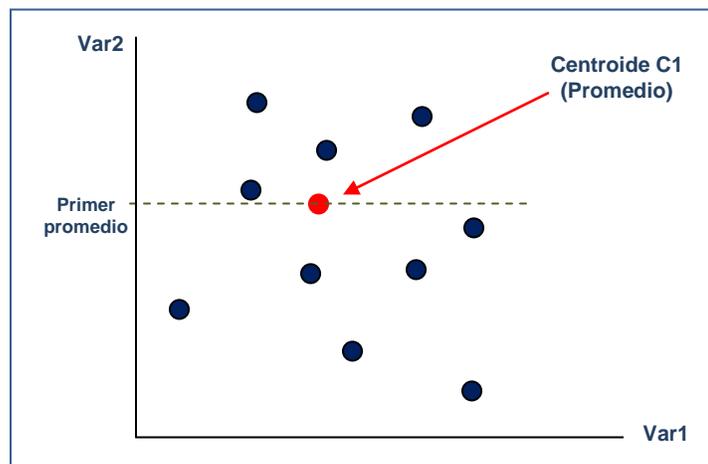
Se calcula el criterio de inercia:

$$W = \sum_{k=1}^K \sum_{i \in C_k} \|x_i - g_k\|^2$$

Ahora en el caso que la variación en el criterio  $W$  entre la iteración anterior y el presente es menor que el referencial dado, o si sobrepasa la cantidad de exploración, entonces se asigna un nuevo número de centro de gravedad.

Frente a todo el método, el método de aplicación de K-Means, dependerá de los centroides asignados al inicio, por ello se recomienda escoger los centros de referencia con criterio.

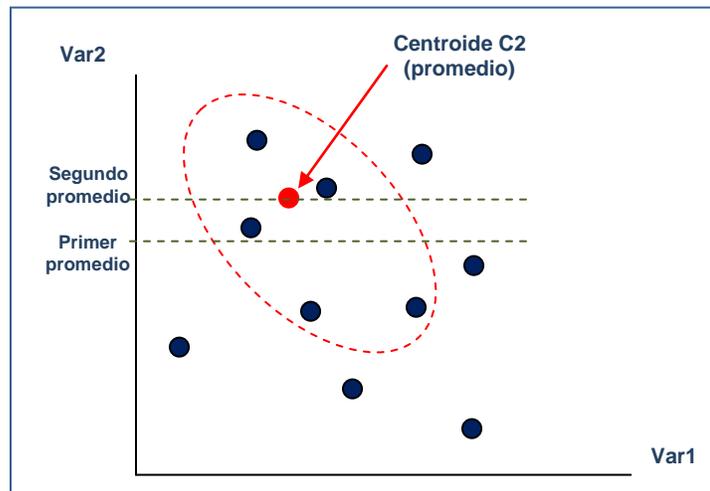
Gráfico 3.2 Primer centroide (referencial)



Fuente: Propia

En el **Gráfico 3.1** nos muestra el centroide inicial que se tomó como referencia, que consiste en el promedio de las variables de toda la población.

**Gráfico 3.3 Segundo centroide (K=2)**



Fuente: Propia

En el **Gráfico 3.2** se observa que una vez que se creó el nuevo grupo con respecto a las distancias más cercanas de los individuos hacia el centroide referencial, el nuevo centro se recalcula para el grupo, y así sucesivamente el método K-Means se recalculará hasta crear tantos grupos sean que cumplan los criterios de similaridad.

Luego de observar la técnica de agrupamiento K-means de manera general, especificaremos los pasos para el desarrollo de cluster K-Means de McQueen.

**(a) Selección de números de clúster:**

Existen varios métodos para selección **K** para el algoritmo **K-Means** y su respectiva técnica de validación de clústering.

Valor de **K** especificado por el usuario.

El algoritmo K-means se implementa en muchos paquetes de software de data mining o análisis de datos, requiere el número de clúster para ser especificado por el usuario para encontrar una agrupación satisfactoria, usualmente el usuario ejecuta el algoritmo con diferentes valores de **K**.

Valor de **K** igualado al número de generadores.

Este método consiste en que el conjunto de datos sintético son usados para probar los algoritmos, estos con frecuencia son creados por una distribución de generadores normal u uniforme. Luego los algoritmos clustering son aplicados a estos conjuntos de datos con el número de clúster igual al número de generadores.

Valor de **K** determinado por medidas estadísticas.

Hay varias medidas estadísticas habilitadas para seleccionar **K**, estas medidas con frecuencia son aplicadas en combinación con clustering probabilístico, estos son calculados con supuestos acerca de distribución lineal de los datos.

### **(a) Factores que afectan la selección de K:**

Una función  $f(k)$  para evaluar el resultado de los clústeres podría ser usado para seleccionar el número de clúster. Los factores de tales funciones deberían ser tomados en cuenta.

#### Distribución interna versus el impacto global.

El Clustering es usado para identificar irregularidades en la distribución de los datos e identificar regiones donde los objetos están concentrados.

En el clúster K-means, la distorsión de un clúster está en función de la población y la distancia entre los objetos y el centro del clúster de acuerdo a:

$$I_j = \sum_{t=1}^{N_j} [d(x_{jt}, w_j)]^2$$

Donde  $I_j$  es la distorsión del clúster  $j$ ,  $w_j$  es el centro del clúster  $j$ ,  $N_j$  es el número de objetos que sigue al clúster  $j$ ,  $x_{jt}$  es el  $t$ -ésimo objeto que sigue al clúster  $j$ , y  $d(x_{jt}, w)$  es la distancia entre el objeto  $x_{jt}$  y el centro  $w_j$  del clúster  $j$ .

Cada clúster tiene su distorsión y su impacto en el conjunto de datos está dado por la suma de todas las distorsiones,

$$S_K = \sum_{j=1}^K I_j$$

Donde  $K$  es el número de clústeres especificado.

**(b) Número de clúster para el clustering K-Means:**

Habíamos mencionado anteriormente que el análisis clúster es usado para identificar irregularidades en la distribución de datos. Cuando la distribución de los datos es uniforme, no hay alguna irregularidad, entonces, el conjunto de datos con distribución uniforme podría ser usado para calibrar y verificar el resultado de los clústeres.

La función de evaluación  $f(k)$  está definida usando la ecuación.

$$f_{(K)} = \begin{cases} 1 & \text{Si } K = 1 \\ \frac{S_K}{S_{K-1}} & \text{Si } S_{K-1} \neq 0, \forall K > 1 \\ \alpha_K & \text{Si } S_{K-1} = 0, \forall K > 1 \\ 1 & \end{cases}$$

$$\alpha_{(K)} = \begin{cases} 1 - \frac{3}{4N_d} & \text{Si } K > 2 \text{ y } N_d > 1 \\ \alpha_{K-1} + \frac{1 - \alpha_{K-1}}{6} & \end{cases}$$

Donde  $S_k$  es la suma de las distorsiones del clúster cuando el número de clúster se  $K$ ,  $N_d$  es el número de atributos del conjunto de datos (número de dimensiones) y  $\alpha_K$  es un peso del factor.

El término  $\alpha_K S_{K-1}$  es un estimador de  $S_K$  basado en  $S_{K-1}$  realizado con los supuestos que los datos tienen una distribución uniforme.

El valor de  $f(k)$  es el ratio de la distorsión real para la distorsión estimada.

Cuando hay grupos de concentración en la distribución de los datos,  $S_K$  será menor que el valor estimado, de modo que  $f(k)$  decrece. Es el más, el más concentrado en la distribución de los datos. Entonces el valor de  $K$  será considerada como un clúster bien definido.

### **(C) El algoritmo continuo K-Means (Clustering McQueen):**

El algoritmo continuo K-means es más rápido que la versión estándar y en el tamaño extenso del conjunto de datos que pueden ser clusterizados.

A diferencia de la versión estándar, el algoritmo estándar escoge el punto inicial de referencia más o menos arbitrariamente. En el algoritmo continuo los puntos de referencia son escogidos mediante una muestra aleatoria de los puntos de la población.

Otra de las diferencias entre los algoritmos K-Means estándar y continuos es cómo son tratados los puntos de los datos. Durante cada iteración completa, el algoritmo estándar examina todos los puntos de datos de manera secuencial, sin embargo, el algoritmo

continuo examina solo una muestra aleatoria de los puntos de datos. Si el conjunto de datos es muy grande y la muestra es representativa del conjunto de datos, el algoritmo debería converger más rápido que un algoritmo que examina cada punto en secuencia.

En efecto, el algoritmo continuo adopta el método MacQueen's de actualizar los centroides durante la partición inicial.

Desde la perspectiva teórica la muestra aleatoria representa el concepto original del algoritmo MacQueen's como un método de clustering de datos en un espacio continuo. En su fórmula, la medida de error  $E_i$  para cada región  $R_i$  está dada por:

$$E_i = \int_{x \in R_i} \rho(x) \|x - z_i\|^2 dx$$

Donde  $\rho(x)$  es la función de densidad de probabilidad, una función definida en el espacio continuo. Y la medida de error total  $E$  esa determinada por la suma de los  $E_i$ .

#### Versión continua del criterio SSQ (Suma de Cuadrados).

Dado: un vector aleatorio  $X$  en  $R$  con distribución conocida  $P$ , densidad  $f(x)$ .

Problema: encontrar una partición óptima  $B = (B_1, \dots, B_k)$  de  $R^p$

Con  $k$  conjunto de clases  $B_i \in R^p$ ,  $i = 1, \dots, k$

Caracterizados por prototipos de clase:  $Z = (z_1, \dots, z_k)$

Luego.

La versión continua con criterio SSQ:

$$G(B) := \sum_{i=1}^k \int_{B_i} \|x - E[X|X \in B_i]\|^2 dP(x) \longrightarrow \min..B$$

Con centroides de clase  $z^*_1 = E[X|X \in B_1], \dots, z^*_k = E[X|X \in B_k]$

Y por tanto de la forma de dos parámetros:

$$G(B, Z) := \sum_{i=1}^k \int_{B_i} \|x - z_i\|^2 dP(x) \longrightarrow \min..B, Z$$

### **1.2.5. Análisis de Mapa de visualización de Kohonen**

El modelo de mapa auto-organizado de Kohonen (SOM) está compuesto por dos capas de neuronas, donde la capa de entrada se encarga de recibir y emitir a la capa de salida quien es la que se encarga de procesar la información y desarrollar los mapas de rasgos semejantes.

El procedimiento para crear las áreas de relación de grupos o clúster, sigue pasos de algoritmo, donde el proceso de aprendizaje es:

Sea un vector  $x$  elegido aleatoriamente de un conjunto de datos, se calcula la distancia (similitud) con respecto al centro global, en este caso podemos usar la distancia Euclidea:

$$\|x - m_c\| = \min \{\|x - m_j\|\}$$

El siguiente paso es cuando se encuentre el vector más próximo entonces el resto de los vectores se actualiza y sus vecinos se encuentran cerca al vector  $x$  en el espacio definido, teniendo en

cuenta que dicho acercamiento es debido a la tasa de aprendizaje del modelo.

Luego, mientras se produce la actualización del proceso, nuevos vectores se asignan al mapa, con eso mientras se realiza la actualización y se van acercando al vector  $x$  entonces la tasa de aprendizaje va decreciendo.

La regla de actualización para el vector de referencia es la siguiente:

$$m_j(t+1) = \begin{cases} m_j(t) + \lambda(t)(x(t) - m_j(t)) & j \in N_c(t) \\ m_j(t) & j \in N_c(t) \end{cases}$$

Este proceso se irá ejecutando hasta que el entrenamiento termine, tener en cuenta que el número de entrenamiento se asignó en un inicio.

Una vez terminada el entrenamiento, el mapa ha de ordenarse en sentido topológico, los  $n$  vectores se ordenan con las  $n$  neuronas adyacentes.

## 2. DEFINICION DE TERMINOS

**2.1 *Market Share (MKS):*** Es un indicador usado en el sistema financiero que representa la participación en el mercado que tuviera una entidad bancaria con respecto al sistema financiero.

**2.2 *Share of Wallet (SOW):*** Es un indicador usado en las entidades bancarias que presentan el uso de su producto con respecto al resto de entidades.

**2.3 *Segmento banco:*** Es el segmento que usa una de las entidades líderes en banca, que constituye grupos en base a la edad e ingreso del cliente.

**2.4 *Cliente:*** Es aquella persona natural que tiene algún producto con una o varias entidades bancarias.

**2.5 *Tarjeta de crédito:*** Es un producto activo cuya función es permitir que se realice compras (consumo) sin la necesidad de tener dinero en efectivo.

- 2.6 Canales de consumo:** Son los diversos establecimientos donde el cliente o no cliente puede realizar consultas u operaciones.
- 2.7 Transacción:** Es la operación que un cliente realiza con su tarjeta de crédito para consumir o retirar dinero de algunos de los canales que dispone la entidad bancaria.
- 2.8 Tarjeta vigente o activa:** Es la tarjeta de crédito que por lo menos ha tenido un consumo o haya realizado alguna transacción.
- 2.9 Rubros de consumo:** Es lo consumido por el cliente una vez realizada su transacción.
- 2.10 Canal ATM:** Es uno de los establecimientos que tienen las entidades bancarias más conocidos como cajeros automáticos para que las personas realicen sus operaciones bancarias sin la necesidad de estar visitando a los bancos.
- 2.11 Cliente cibernauta:** Es aquella persona que dedica mayor parte de sus labores estar conectado en internet, en nuestro caso son personas que realizan mayor parte de sus operaciones bancarias en la web.

**2.12 Cliente tradicional:** es la característica persona común, el cual realiza sus operaciones bancarios en otros establecimientos más personalizado como: ventanilla, ATM, etc.

## **CAPITULO IV**

### **1. METODOLOGIA**

En esta investigación se utilizó información primaria ya que se ha extraído directamente de las fuentes de datos del Banco Internacional del Perú (Data Warehouse) , es una investigación exploratoria debido a que no se tiene una segmentación por comportamiento de consumo de los clientes del banco en estudio.

El público objetivo son aquellos clientes que tienen su tarjeta de crédito activa con una antigüedad mayor o igual de seis meses en el banco en estudio así como otros comportamientos financieros que se detallaran más adelante, las variables y dimensiones consideradas son todos los rubros de consumo (26 variables) y dos dimensiones que tiene categorizada el banco en estudio.

El método de reducción de variables consistió en el análisis factorial multivariado y análisis de componentes principales para la explicación de variabilidad de las variables (ACP), luego se detallarán las dos etapas de segmentación que consiste en diferenciar al grupo cibernauta y grupo tradicional.

Luego para observar el comportamiento entre segmentos se usaron los mapas de visualización de Kohonen que analiza la relación entre segmentos.

Los pasos detallados para la determinación de los segmento de clientes en el banco de estudio son:

**a) Investigación:**

Este trabajo consistió en una investigación exploratoria, ya que el propósito fue determinar que segmentos se obtendrían en base al comportamiento de consumo de los clientes, por otro lado no se tuvo estudios anteriores por comparar, también es transversal y descriptiva.

**b) Público objetivo**

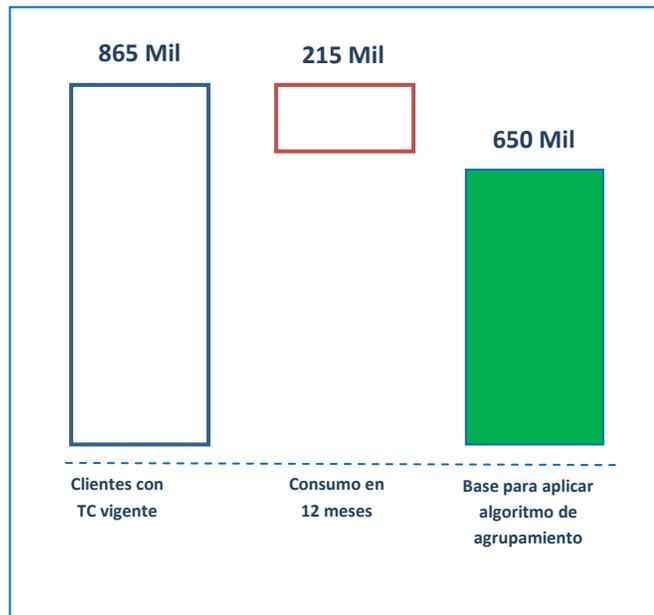
Se desarrollaran dos propuestas el cual elegiremos el que mejor se ajuste a nuestro objetivo.

La primera propuesta consiste en:

- Clientes con tarjeta de crédito activo de una entidad considerada líder en el mercado peruano en Agosto del 2014.
- Que pertenecen a clientes naturales de todos los niveles socioeconómicos.

- El periodo tomado fue de 12 meses, las transacciones de tarjetas de crédito de los clientes fueron evaluados en ese periodo de tiempo.

**Gráfico 4.1 Primer público objetivo en estudio de segmentación**



Fuente: Propia

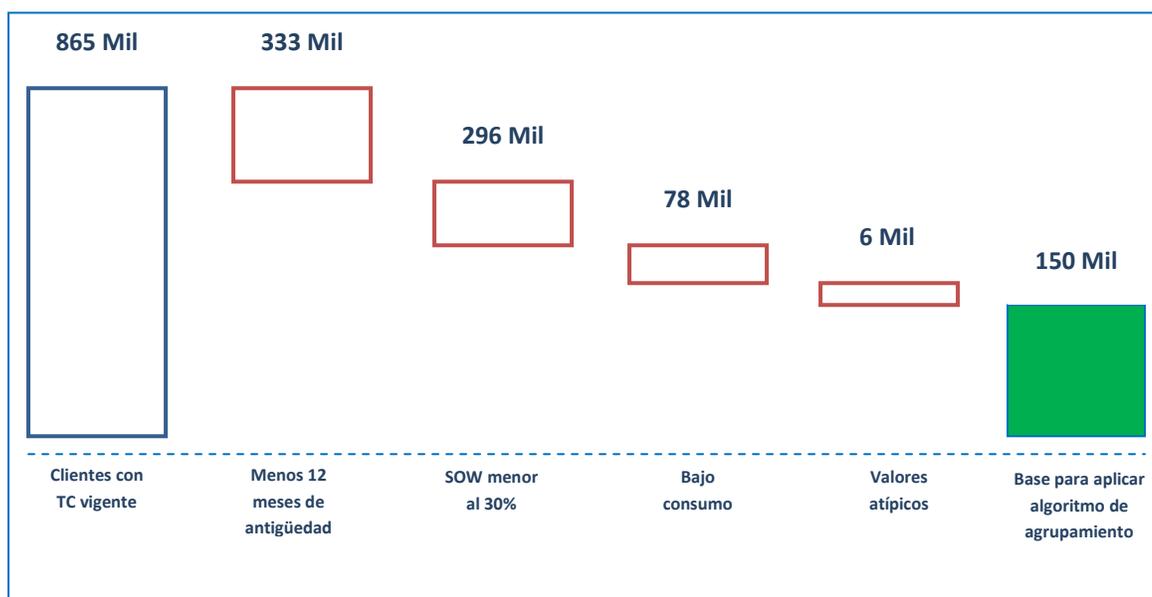
En el **Gráfico 4.1**, nos presenta la información final que se usó para aplicar la técnica de segmentación de la primera propuesta, en donde se partió de la cantidad de clientes total con tarjeta de crédito vigente de una entidad bancaria que es de 865 mil clientes hasta Agosto del 2014, en ello se realizan los siguientes filtros como:

Sólo considerar tarjetas que hayan realizado algún consumo en el lapso de los 12 últimos meses, cuyo titular de la tarjeta de crédito sean personas naturales de cualquier nivel socioeconómico.

Por otro lado la segunda propuesta esta determinada por:

- Clientes con tarjeta de crédito activo de una entidad considerada líder en el mercado peruano en Agosto del 2014.
- Sólo se consideran clientes con tarjeta de crédito mayor o igual a doce meses de antigüedad.
- Que pertenecen a clientes naturales de todos los niveles socioeconómicos.
- El periodo tomado fue de 12 meses, las transacciones de tarjetas de crédito de los clientes fueron evaluados en ese periodo de tiempo.
- Clientes que hayan realizado por lo menos 12 transacciones con su tarjeta de crédito.
- Que hayan tenido una participación significativa con su tarjeta de crédito en la entidad bancaria con un SOW mayor o igual al 30%.
- No se consideraron aquellos clientes con número de transacciones atípicos.

**Gráfico 4.2 Segundo público objetivo para el estudio de segmentación**



Fuente: Propia

En el **Gráfico 4.2**, nos presenta la información final que se usó para aplicar la técnica de segmentación, en donde se partió de la cantidad de clientes total con tarjeta de crédito vigente de una entidad bancaria que es de 865 mil clientes hasta Agosto del 2014, en ello se realizan los siguientes filtros como:

333 mil clientes cuya antigüedad de su tarjeta de crédito es menor a 12 meses, 296 mil clientes cuya participación de consumo en la entidad bancaria con respecto al resto es menor al 30%, 78 mil clientes cuyo consumo no son significativos para ser parte de la segmentación y por ultimo extraemos 6 mil clientes cuyas transacciones se diferencian significativamente del resto de clientes, que son los casos atípicos que podrían alterar el estudio.

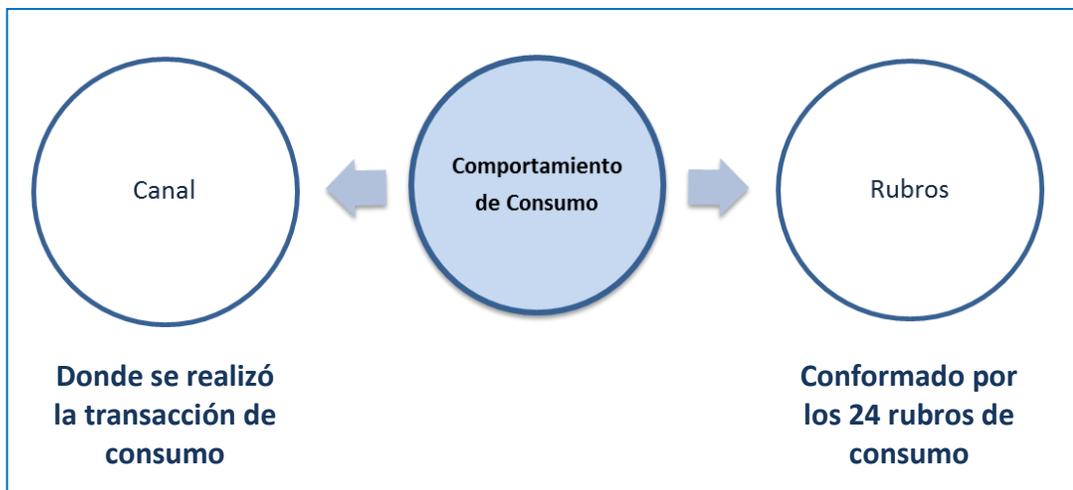
Por tanto la cantidad de clientes que se trabajó para la segmentación fue de 159 mil.

**c) Selección de variables y dimensiones:**

Teniendo en cuenta que se trabajó con los 150 mil clientes, estos están conformados por variables relacionadas a la transacción de consumo y al canal donde lo realizo.

Las dimensiones están conformados por: el canal de consumo y los rubros de consumo.

**Gráfico 4.3 Dimensiones de segmentación**



Fuente: Propia

En el **Gráfico 4.3**, nos muestra las dimensiones que se consideró en el estudio, estas dimensiones se acordaron con los expertos en el producto de tarjetas de crédito, lo que indico que el comportamiento de consumo de los clientes está conformado por el que consume y en donde lo realizó.

A continuación detallaremos que variables intervienen en cada una de las dimensiones definidas.

**Cuadro 4.1 Rubros de consumo de tarjeta de crédito**

Nro	Variable	Descripción
1	Aerol	Aerolíneas
2	Agenc	Agencia de viajes
3	Asoci	Asociaciones
4	Bazar	Tiendas y bazar
5	Casin	Casinos y juegos de azar
6	Contr	Contratistas Generales
7	Dispo	Disposición de efectivo
8	Educa	Educación
9	Empre	Empresas financieras
10	Equip	Equipos y repuestos
11	Estac	Estaciones de Servicio
12	Hotel	Hoteles
13	Inter	Internacional
14	Libre	Librerías
15	Pagod	Pago de servicios
16	Pagos	Pagos Varios
17	Recre	Recreación
18	Resta	Restaurantes
19	Salud	Salud
20	Servi	Servicios Especializados
21	Super	Supermercados
22	Tiend	Tiendas por departamento
23	Trans	Transporte
24	Otras	Otras

Fuente: Banco Internacional del Perú

En el **Cuadro 4.1** describe todas las variables que implica la dimensión rubros de consumo por cliente

**Cuadro 4.2 Canal de consumo de tarjeta de crédito**

Nro	Variable	Descripción
1	Internet	Web
2	Otras	ATM, ventanilla, etc.

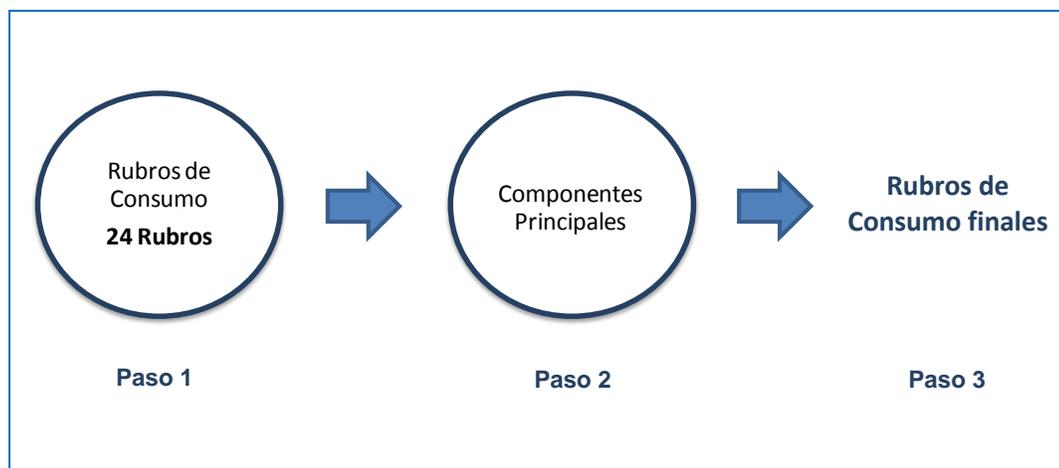
Fuente: Banco Internacional del Perú

En el **Cuadro 4.2**, nos presenta que la dimensión canal de consumo solo presenta 2 variables, donde la variable Otras implica los establecimientos personalizados a diferencia de Internet que es virtual.

Ahora si nos enfocamos en la dimensión rubros de consumo, nos dimos cuenta que existen muchas variables, y esto pudo dificultar la segmentación ya que las transacciones que realizó un clientes pudo estar distribuido en todas.

Por ello usamos la técnica de reducción de variables con el análisis de factores conjuntamente con los componentes principales.

**Gráfico 4.4** Proceso de reducción de variables (rubro de consumo)



Fuente: Propia

En el **Gráfico 4.4**, nos muestra el proceso de cómo reducir las variables en la dimensión rubros de consumo con el objetivo de facilitar la interpretación de todas las variables que explicarían el comportamiento de consumo de los clientes.

En el **Paso 1** nos indica que partimos de las 24 variables de consumo que ya se había definido, en el **Paso 2** iniciamos la técnica de análisis de componentes principales (ACP) para la reducción de variables el cual se desarrollo de la siguiente manera:

Componentes Principales en variables de consumo.

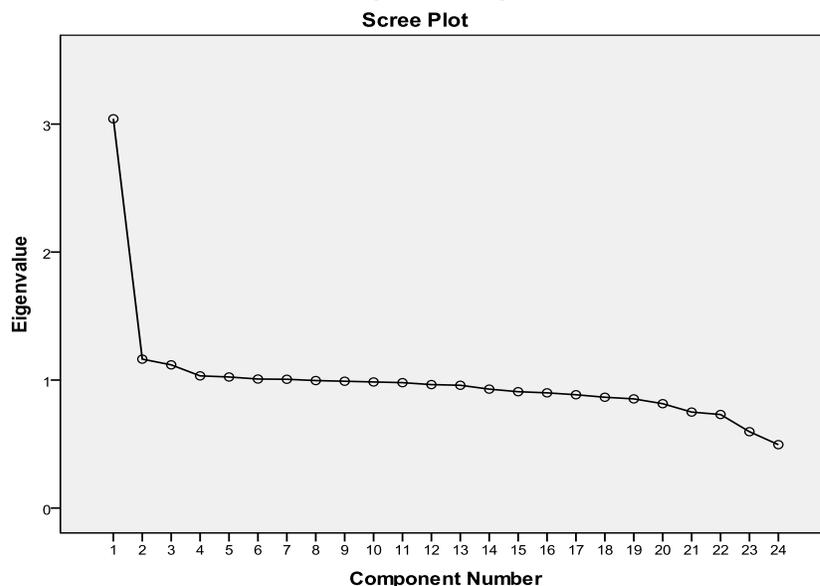
Analizando las variables en el software estadístico SPSS con la cantidad de factores que se obtuvo con autovalores mayores a 1, se obtuvo que la explicación de los factores hacia las variables es muy poco.

**Cuadro 4.3 KMO and Bartlett's Test**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	,839	
Bartlett's Test of Sphericity	Approx. Chi-Square	139618,435
	Df	276
	Sig.	,000

En el **Cuadro 4.3**, nos muestra si fue necesario realizar una reducción de variables agrupando en factores, según el indicador Kaiser-Meyer-Oklin el valor es de 0.839, y la prueba de Barlett rechaza la hipótesis nula ya que  $p\text{-value} < 0.05$ , donde rechaza que la matriz de correlación no se comporta como una matriz identidad lo que implica que si fue necesario realizar el agrupamiento de variables porque existe correlación entre ellas.

**Gráfico 4.5 Componentes por autovalores**



En el **Gráfico 4.5** nos muestra la cantidad de componentes vs el valor de autovalores en este caso >1.

**Cuadro 4.4 Total Variance Explained**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,041	12,672	12,672	3,041	12,672	12,672	2,643	11,011	11,011
2	1,163	4,848	17,519	1,163	4,848	17,519	1,380	5,749	16,760
3	1,119	4,662	22,181	1,119	4,662	22,181	1,175	4,897	21,657
4	1,033	4,304	26,485	1,033	4,304	26,485	1,101	4,588	26,245
5	1,024	4,267	30,752	1,024	4,267	30,752	1,039	4,330	30,575
6	1,008	4,200	34,952	1,008	4,200	34,952	1,030	4,293	34,868
7	1,006	4,192	39,145	1,006	4,192	39,145	1,026	4,276	39,145
8	,996	4,151	43,295						
9	,991	4,129	47,424						
10	,985	4,105	51,528						
11	,980	4,082	55,610						
12	,965	4,019	59,630						
13	,959	3,997	63,627						
14	,929	3,872	67,498						
15	,909	3,788	71,286						
16	,900	3,750	75,036						
17	,885	3,689	78,725						
18	,866	3,607	82,332						
19	,853	3,553	85,886						
20	,815	3,397	89,283						
21	,750	3,124	92,407						
22	,731	3,045	95,451						
23	,596	2,485	97,937						
24	,495	2,063	100,000						

Extraction Method: Principal Component Analysis.

En el **Cuadro 4.4**, si bien de las 24 variables lo reduce en 7 componentes teniendo en cuenta que el autovalor sea mayor que 1, estos no explican lo suficiente a las 24 variables ya que las 7 componentes sólo explican el 39.14% de variabilidad, es aceptable cuando explican a partir del 60%.

**Cuadro 4.5 KMO and Bartlett's Test**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,840
Bartlett's Test of Sphericity	Approx. Chi-Square	139446,920
	Df	276
	Sig.	,000

En el **Cuadro 4.5**, nos muestra otro tratamiento de las variables ya que el primer método no explicaba la variabilidad suficiente de las variables. En este caso el ratio KMO y el test de Barlett indican que es necesario el agrupamiento de variables porque existe correlación entre ellas.

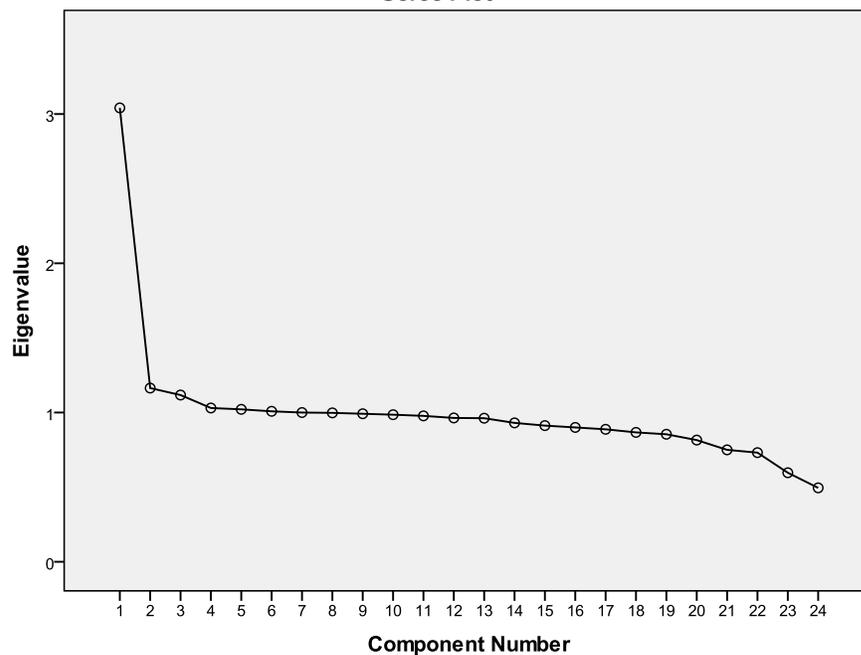
**Cuadro 4.6 Communalities**

	Initial	Extraction
TRX_ASOCI	1,000	,887
TRX_EDUCA	1,000	,936
TRX_CASIN	1,000	,997
TRX_CONTR	1,000	,461
TRX_SALUD	1,000	,383
TRX_SERVI	1,000	,335
TRX_LIBRE	1,000	,273
TRX_RECRE	1,000	,557
TRX_HOTEL	1,000	,452
TRX_PAGOD	1,000	,956
TRX_DISPO	1,000	,966
TRX_BAZAR	1,000	,358
TRX_EMPRE	1,000	,975
TRX_RESTA	1,000	,608
TRX_TIEND	1,000	,434
TRX_TRANS	1,000	,561
TRX_SUPER	1,000	,533
TRX_EQUIP	1,000	,811
TRX_AGENC	1,000	,483
TRX_AEROL	1,000	,523
TRX_INTER	1,000	,826
TRX_ESTAC	1,000	,415
TRX_PAGOS	1,000	,532
TRX_OTRAS	1,000	,998

Extraction Method: Principal Component Analysis.

En el **Cuadro 4.6**, observamos la comunalidad de las variables donde nos indica que proporción de la varianza es explicada por el modelo de factores, por ejemplo en el rubro internacional (TRX\_INTER) la proporción de varianza explicada por el modelo de factores es del 0.82, lo que nos indica que gran parte de la varianza es explicada, sin embargo, el rubro (TRX\_SERVI) es explicado sólo el 0.33 de su varianza por el modelo de factores, de esta manera varios de las rubros de consumo son explicadas gran parte de su varianza como: TRX\_RESTA, TRX\_DISPO, etc.

**Gráfico 4.6 Componentes por autovalores**  
Scree Plot



En el **Gráfico 4.6** nos muestra la cantidad de componentes vs el valor de autovalores, en este caso no solo nos enfocaremos en los autovalores mayor que 1 sino también a los menores, ya que ganaríamos mayor explicación de variabilidad a pesar de incrementar las cantidades de componentes.

**Cuadro 4.7 Total Variance Explained**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,041	12,671	12,671	3,041	12,671	12,671	2,426	10,110	10,110
2	1,164	4,849	17,521	1,164	4,849	17,521	1,395	5,813	15,923
3	1,117	4,655	22,176	1,117	4,655	22,176	1,235	5,145	21,068
4	1,030	4,293	26,469	1,030	4,293	26,469	1,127	4,697	25,765
5	1,021	4,256	30,725	1,021	4,256	30,725	1,029	4,287	30,052
6	1,008	4,201	34,925	1,008	4,201	34,925	1,026	4,273	34,325
7	1,000	4,166	39,092	1,000	4,166	39,092	1,012	4,215	38,541
8	,998	4,157	43,249	,998	4,157	43,249	1,010	4,209	42,750
9	,992	4,132	47,381	,992	4,132	47,381	1,004	4,185	46,935
10	,986	4,106	51,487	,986	4,106	51,487	1,001	4,171	51,105
11	,978	4,073	55,561	,978	4,073	55,561	1,001	4,169	55,275
12	,964	4,015	59,576	,964	4,015	59,576	1,000	4,167	59,442
13	,962	4,008	63,584	,962	4,008	63,584	,994	4,143	63,584
14	,931	3,878	67,462						
15	,912	3,801	71,263						
16	,900	3,751	75,014						
17	,887	3,698	78,712						
18	,867	3,612	82,324						
19	,854	3,560	85,883						
20	,815	3,397	89,280						
21	,750	3,124	92,405						
22	,731	3,047	95,451						
23	,596	2,485	97,937						
24	,495	2,063	100,000						

Extraction Method: Principal Component Analysis.

En el **Cuadro 4.7**, de las 24 variables ha reducido a 13 componentes teniendo en cuenta que el autovalor mayores o menores que 1, estas componentes explican lo suficiente a las 24 variables ya que los 13 componentes explican el 63.58% de variabilidad, es aceptable porque son mayor que el 60%.

**Cuadro 4.8 Rotated Component Matrix<sup>a</sup>**

	Component												
	1	2	3	4	5	6	7	8	9	10	11	12	13
TRX_SUPER	,660	,171	,235	-,056	-,019	,020	,036	,063	,019	,031	-,026	,003	-,046
TRX_TIEND	,611	,098	,064	,033	-,114	-,032	,010	-,046	,052	-,015	-,078	,002	,144
TRX_SALUD	,575	,019	,171	,004	-,014	,107	,029	,057	-,012	-,009	,007	-,009	-,082
TRX_BAZAR	,566	-,002	-,053	,088	,060	-,046	-,047	-,063	-,035	-,002	-,034	,008	,111
TRX_SERVI	,562	,032	-,020	,053	,094	-,012	,010	,034	-,008	-,012	,051	-,013	-,043
TRX_RESTA	,556	,521	,144	-,013	,023	,027	,053	,015	,030	,022	,005	-,001	,021
TRX_LIBRE	,363	-,067	-,029	,017	,165	,229	,227	,007	-,040	,000	,038	,009	,037
TRX_RECRE	,155	,706	-,129	-,029	,010	-,043	,075	,078	,044	,023	,009	,014	-,020
TRX_HOTEL	,015	,644	,130	,044	,043	,051	-,074	-,063	-,050	-,028	,015	-,008	,026
TRX_CONTR	,210	-,160	,597	-,112	,013	-,104	-,014	-,022	,080	-,003	,017	,014	,065
TRX_TRANS	-,104	,188	,579	,366	-,056	-,009	,062	-,036	-,089	-,017	-,076	-,012	,148
TRX_ESTAC	,285	,298	,479	-,045	-,012	,077	,052	,034	,042	,018	-,009	-,007	-,031
TRX_AEROL	,044	,030	-,052	,711	,033	-,003	,033	,010	,032	-,003	-,037	-,013	,086
TRX_AGENC	,089	-,040	,044	,668	,012	,019	-,038	,015	-,007	,012	,063	,024	-,137
TRX_INTER	,039	,083	-,048	,056	,887	-,069	,020	-,043	,023	-,005	-,063	-,007	,114
TRX_PAGOS	,134	-,046	,376	-,046	,412	,220	-,048	,154	-,043	,013	,102	,013	-,333
TRX_ASOCI	,043	,044	-,028	,016	-,042	,936	-,006	-,019	,020	-,002	-,029	-,001	,050
TRX_EDUCA	,062	,022	,048	,000	-,005	-,008	,964	,005	,005	-,005	-,005	,000	-,007
TRX_PAGOD	,023	,015	-,001	,027	-,011	-,015	,006	,976	,002	-,001	-,016	-,001	,046
TRX_EMPRE	-,009	-,001	,039	,021	,011	,013	-,001	,001	,986	-,003	,001	-,001	,001
TRX_CASIN	-,010	,000	,000	,008	-,002	-,002	-,005	-,001	-,003	,998	,009	,005	,005
TRX_DISPO	-,045	,030	-,017	,020	-,038	-,022	,001	-,015	,001	,009	,979	-,004	,036
TRX_OTRAS	-,010	,004	,004	,011	-,003	,001	,001	-,001	-,001	,005	-,004	,999	,002
TRX_EQUIP	,110	,001	,112	-,045	,089	,074	-,008	,063	-,004	,007	,049	,004	,874

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

En el **Cuadro 4.8**, nos muestra las agrupaciones de las variables por componente principal, luego de la rotación con el método Varimax, la interpretación fue más sencilla, las agrupaciones de factores originalmente por la proporción de explicación de la varianza están resaltadas con amarillo, sin embargo, observamos que algunas variables podrían pertenecer a otros componentes ya que la característica de la

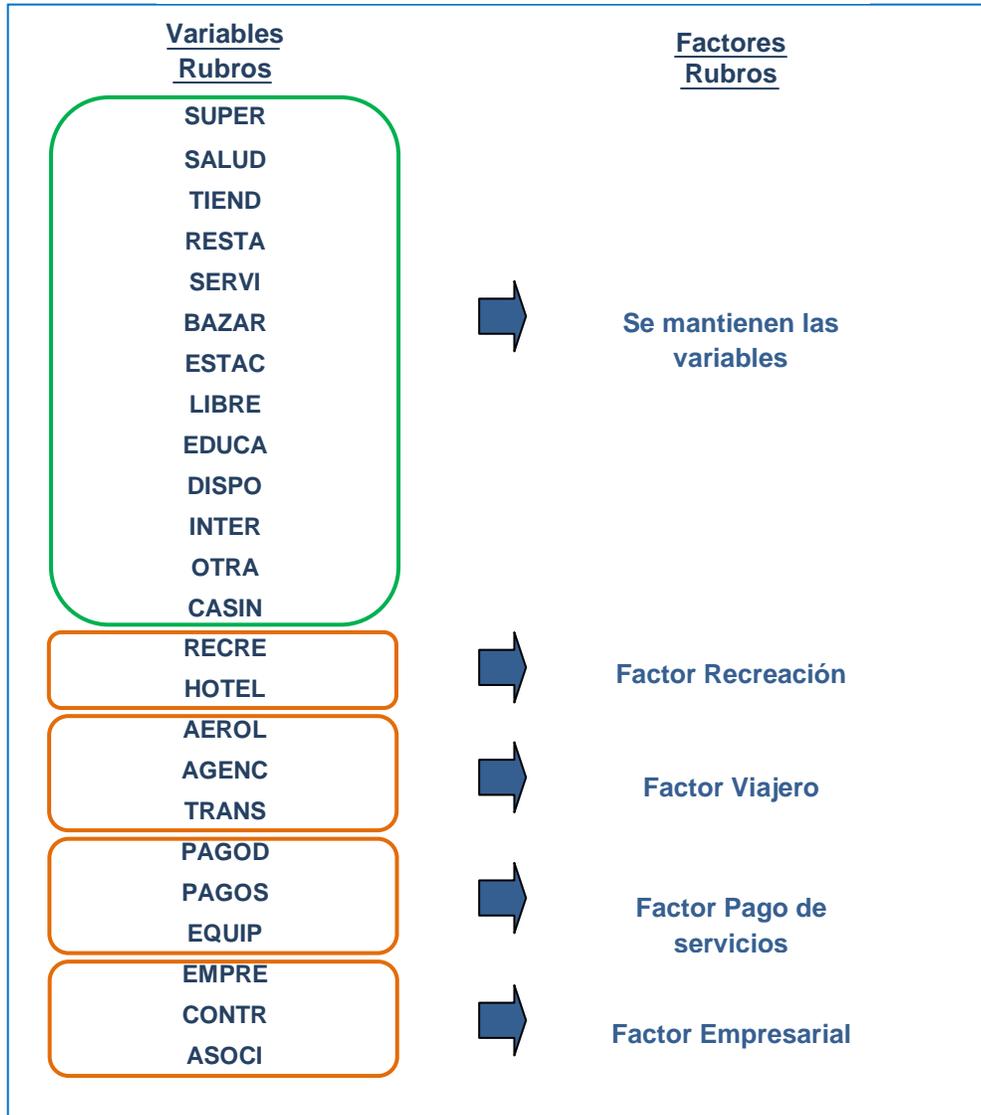
variable se asemejan y la proporción de explicación de su varianza es próxima que la original, también porque expertos en el tema (área de producto de tarjetas de crédito) lo sugirieron.

Es por ello que el rubro de consumo TRX\_ESTAC y TRX\_EDUCA pertenecerían al primer componente, TRX\_TRANS al cuarto componente, TRX\_PAGO y TRX\_EQUIP al octavo componente y TRX\_CONTR, TRX\_ASOCI al noveno componente, el resto se comporta como únicos componentes.

Por otro lado, el primer grupo conformado por los rubros de consumo supermercado (TRX\_SUPER), salud (TRX\_SALUD), tiendas por departamento (TRX\_TIEND), restaurante (TRX\_RESTA), servicio especial (TRX\_SERVI), bazar (TRX\_BAZAR), estación de servicio (TRX\_ESTAC), librería (TRX\_LIBRE) y educación (TRX\_EDUCA). Si bien se agruparon, estos permanecerían como variables ya que por fines comerciales se sugirió mantener estas variables debido a que estos rubros con frecuencia son más dirigidos frente a una gestión de campañas.

En el **Paso 3** ya observamos los grupos finales que fueron:

Gráfico 4.7 Factores finales



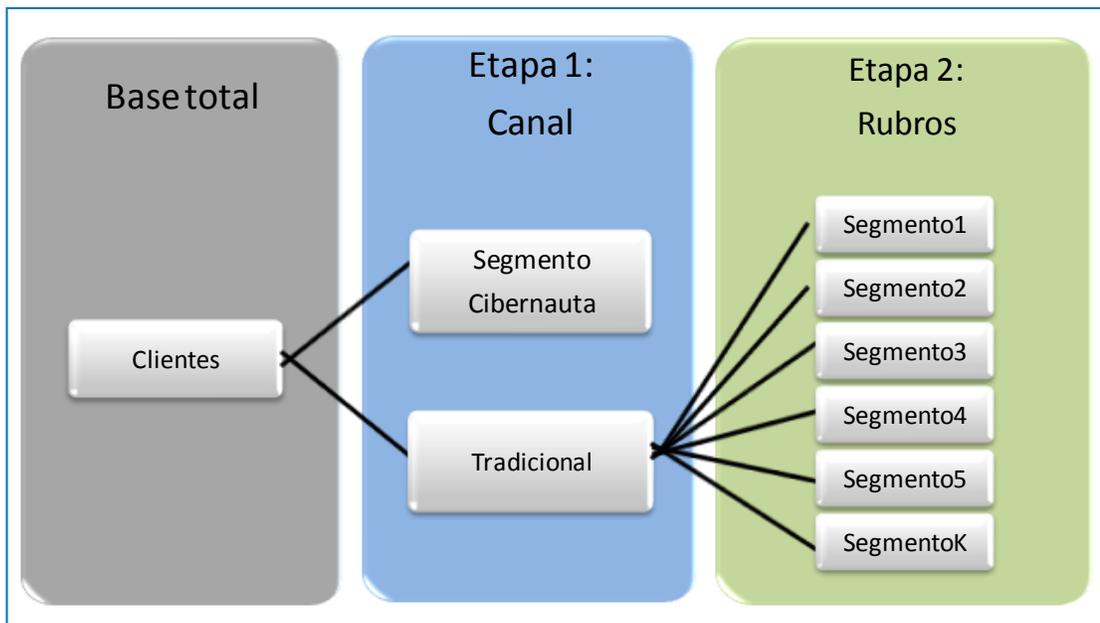
Fuente: Propio

En el **Gráfico 4.7**, se observa que de los 24 rubros se redujeron a **16 rubros** que fueron trabajados para la segmentación, teniendo en cuenta que hubieron algunos factores que mantuvieron sus variables esto por decisiones comerciales.

**d) Etapas de Segmentación:**

Mediante la experiencia de las áreas involucradas en tarjetas de crédito se determinó que existía un grupo cibernauta (personas que realizan sus transacciones en internet) como segmento y el resto se determinaría mediante los rubros de consumo, es por ello que se define dos etapas que consiste en lo siguiente.

**Gráfico 4.8 Etapas de segmentación**



Fuente: Propio

En el **Gráfico 4.8**, nos indica que La primera etapa consiste en determinar qué proporción de los clientes son considerados grupo cibernauta (segmento cibernauta) mediante la variable **Canal** el cual solo esta categorizado como: Internet (usan la web para sus transacciones) y No internet.

La segunda etapa consiste en determinar los segmentos del grupo no cibernauta (Tradicional) el cual se realizara con las transacciones en los rubros de consumo.

***e) Análisis de vecindades (entre segmentos)***

Luego de la determinación del segmento cibernauta y de los otros segmentos obtenidos del grupo tradicional, lo relacionaremos para identificar que segmentos están muy próximos entre ellos en caso se quiera hacer una gestión de que grupos se les podría ofrecer como alternativa otros productos que no sean del segmento que estén definidos.

## CAPITULO V

### 1. APLICACIÓN

#### 1.1 Segmentación de primer público objetivo

En esta primera propuesta de público objetivo, aplicaremos la técnica de agrupamiento para aquellos clientes con tarjeta de crédito que estén activas y que hayan tenido algún consumo.

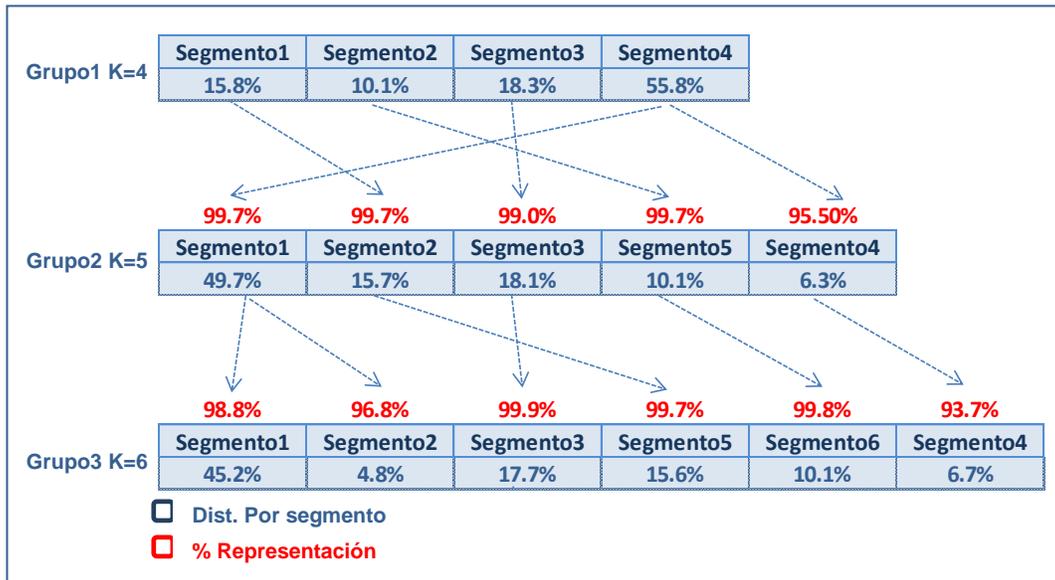
La técnica K-Means realizara una serie de iteraciones para obtener los grupos que mejor se ajusten a nuestro propósito.

En el **Gráfico 5.1**, se observa el inicio del primer agrupamiento, donde iniciamos con cuatro grupos de los 650 Mil clientes.

**Gráfico 5.1 K-Means K=4 segmentos**

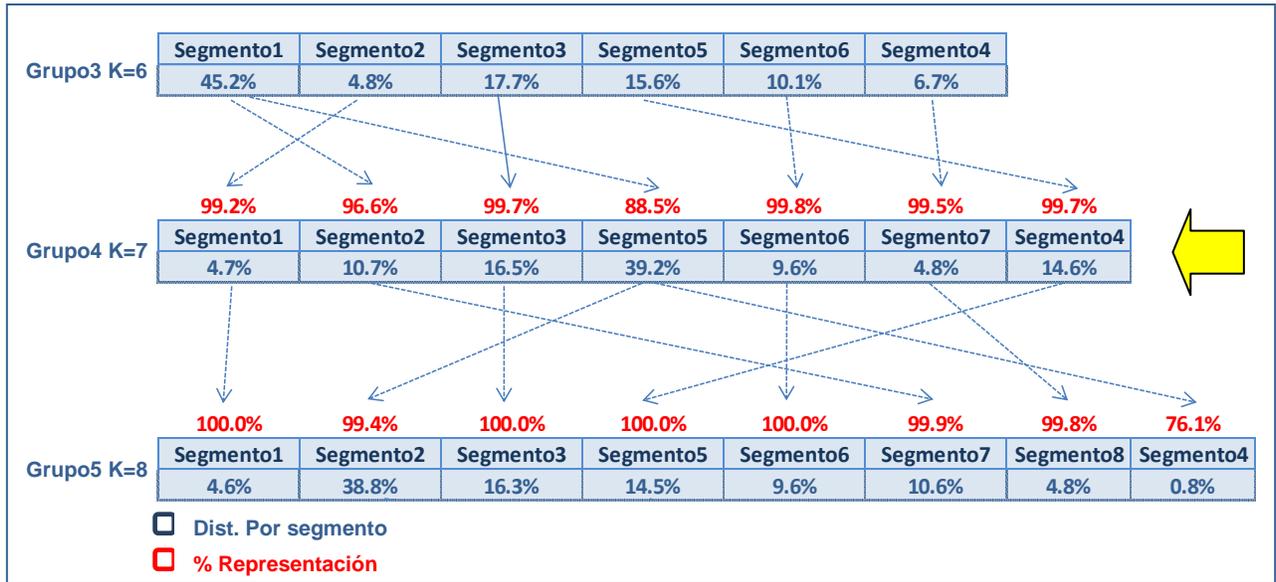
	Segmento1	Segmento2	Segmento3	Segmento4
Grupo1 K=4	15.8%	10.1%	18.3%	55.8%

Gráfico 5.2 K-Means K=5 segmentos



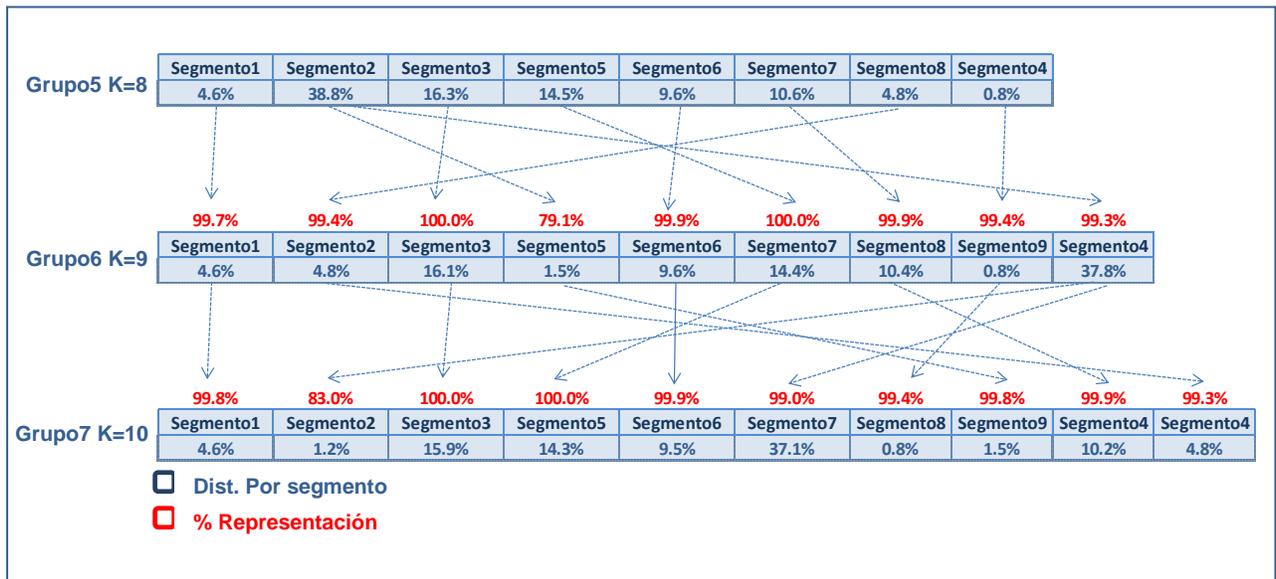
Ahora en el **Gráfico 5.2**, tenemos al Grupo 2 que divide en 5 segmentos, sin embargo, se observa que estos segmentos son incluidos gran parte de su información en los segmentos del Grupo anterior (Grupo 1), por ejemplo, el segmento 1 del Grupo 2 que representa el 49.7% de la población en estudio incluye en un 99.7% el segmento 4 del Grupo 1, para el segmento 1 creado en el Grupo 2 careció de aportación ya que se pudo utilizar los segmentos del Grupo 1.

Gráfico 5.3 K-Means K=6 segmentos



Para el **Gráfico 5.3**, se observa que el Grupo 4 que divide en 7 segmentos a la población en estudio, se crearon segmentos no relevantes que aporten nueva información tal como el segmento 2 donde está incluido el 96,6% de su información en el segmento 1 del Grupo 3.

Gráfico 5.4 K-Means K=8 segmentos



Por último en el **Gráfico 5.4**, se observa que los segmentos del Grupo 6 no aportan información relevante con respecto al Grupo anterior, de la misma forma para el Grupo 7 los segmentos se encuentran incluidos en su mayoría por el Grupo anterior.

A pesar que todos los grupos en esta primera propuesta no se diferencian notablemente, esto tal vez por el público objetivo tomado, se eligió el Grupo 7 como segmentos final ya que a partir de ahí los siguientes grupos no aportaban información que se aprovecharía para el estudio o la distribución de los grupos son muy pequeñas.

**Cuadro 5.1 Evaluación del segmento – Primera propuesta**

<b>GRUPO1 K=4</b>	a) Numero de segmento	<b>4</b>	
	b) Cohesion	0.356	
	c) Exterior	0.004	
	d) Silueta	0.28	
<b>GRUPO2 K=5</b>	a) Numero de segmento	<b>5</b>	
	b) Cohesion	0.35	
	c) Exterior	0.09	
	d) Silueta	0.28	
<b>GRUPO3 K=6</b>	a) Numero de segmento	<b>6</b>	
	b) Cohesion	0.28	
	c) Exterior	0.11	
	d) Silueta	0.29	
<b>GRUPO4 K=7</b>	a) Numero de segmento	<b>7</b>	
	b) Cohesion	0.19	
	c) Exterior	0.125	
	d) Silueta	0.31	
<b>GRUPO5 K=8</b>	a) Numero de segmento	<b>8</b>	
	b) Cohesion	0.21	
	c) Exterior	0.119	
	d) Silueta	0.3	
<b>GRUPO6 K=9</b>	a) Numero de segmento	<b>9</b>	
	b) Cohesion	0.195	
	c) Exterior	0.122	
	d) Silueta	0.298	
<b>GRUPO7 K=10</b>	a) Numero de segmento	<b>10</b>	
	b) Cohesion	0.231	
	c) Exterior	0.128	
	d) Silueta	0.3	

Fuente: Propio

En el **Cuadro 5.1**, observamos con más detalle cómo se comportan los segmentos por cada grupo explorado, nos apoyamos de indicadores que miden la cohesión, distancia exterior y la silueta de cada segmento, esto para determinar la estabilidad y ajuste de los clústeres, en este caso lo que se pudo rescatar fueron los segmentos divididos en 7 por ser diferenciarse de los otros grupos.

## **1.2 Segmentación de segundo público objetivo**

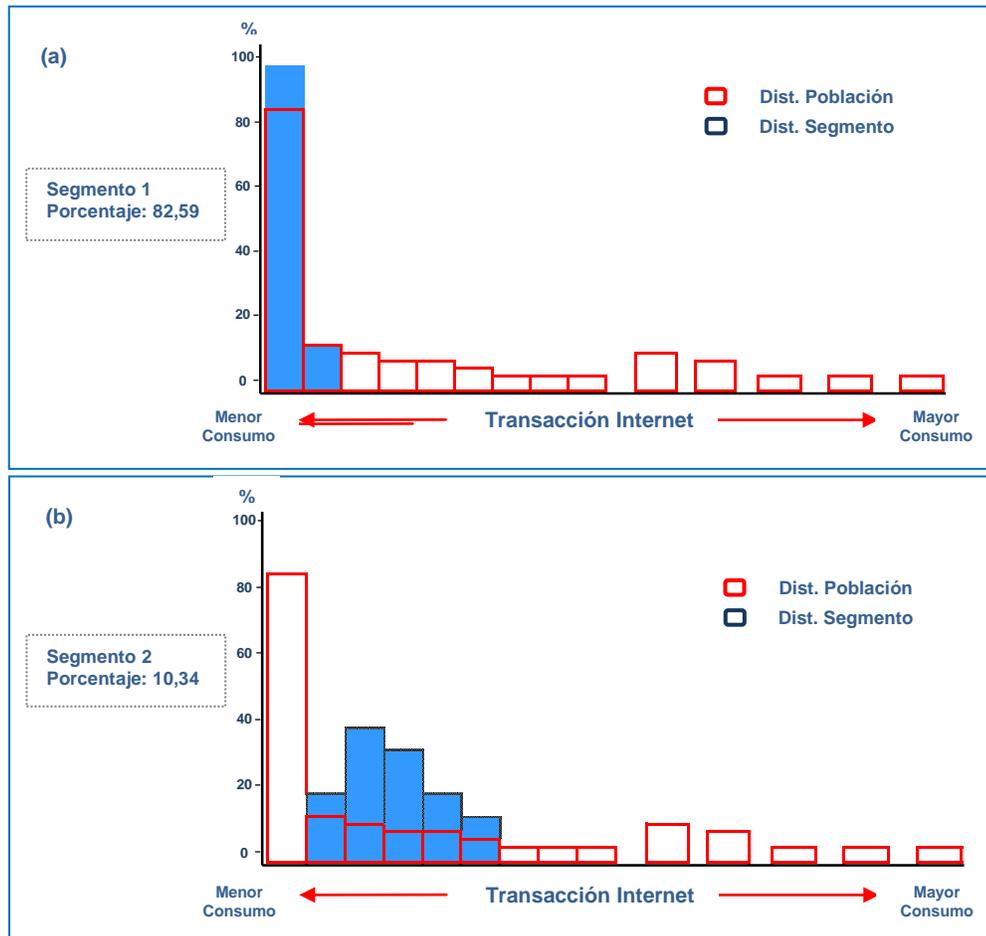
Ahora aplicaremos la segmentación K-Means al público objetivo de la segunda propuesta donde se aplicaron varios criterios para la construcción de la base teniendo en cuenta la experiencia de expertos en el producto de tarjeta de crédito.

### **1.2.1. Grupo Cibernauta**

El propósito para este grupo es saber a partir de cuantas transacciones que realiza un cliente por medio del canal internet se consideraría cibernauta, por ello se usó la técnica de segmentación K-Means para determinar el corte de transacciones más recomendada.

Los resultados de segmentación se analizaron con ayuda del software estadístico SAS Enterprise Miner.

Gráfico 5.5 Segmento para determinar al grupo cibernauta

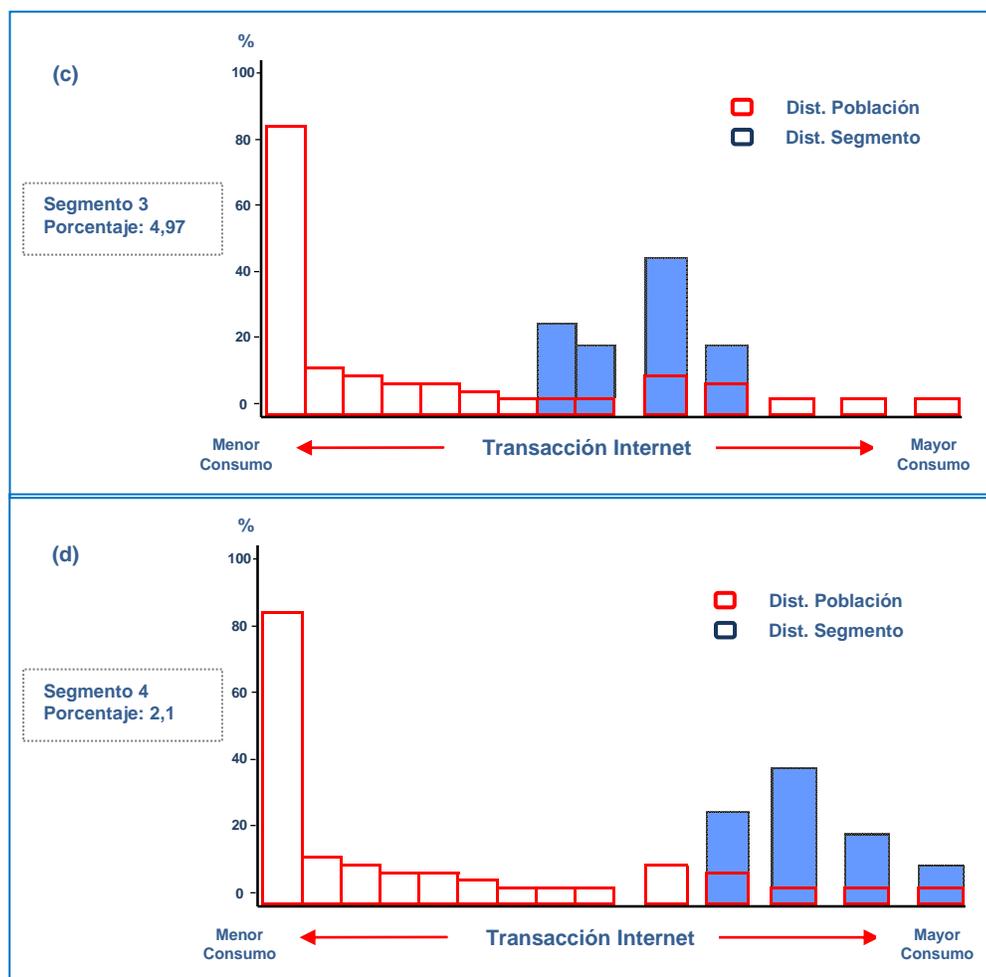


Como primera prueba se usó la variable Internet que nos indica cuantas transacciones el cliente realizó por vía web, la segmentación para determinar al grupo cibernauta se inició tomando 4 segmentos (K=4) esto como referencia.

En el **Gráfico 5.5**, observamos los 2 primeros segmentos que determinó la técnica K-Means, en el **cuadro (a)**, nos indica que los clientes pertenecientes al segmento 1 se encuentran muy por debajo del promedio de la distribución de la población que no realiza sus

transacciones por vía web, en **cuadro (b)**, nos indica que los clientes del segmento 2 realizan pocas transacciones por vía web.

**Gráfico 5.6 Segmento para determinar al grupo cibernauta**

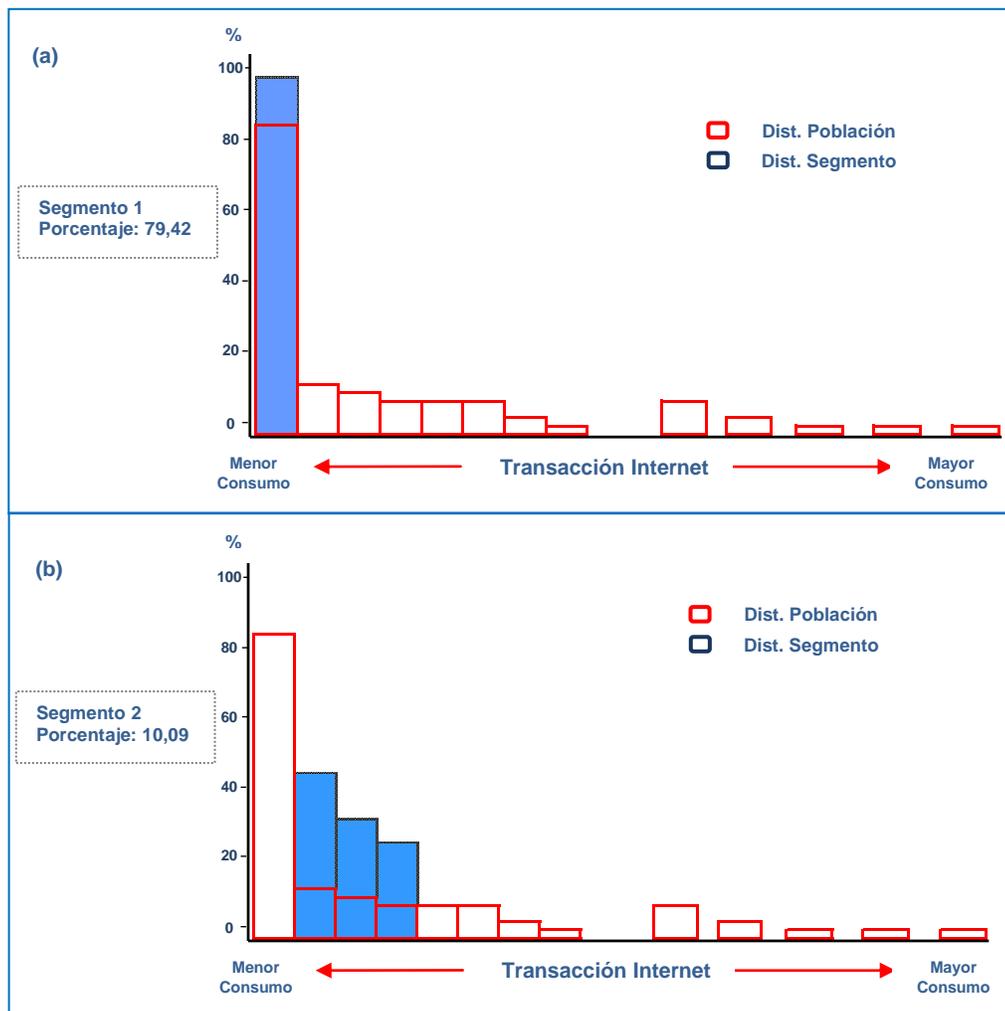


En el **Gráfico 5.6**, observamos los 2 últimos segmentos que determinó la técnica K-Means, en el **cuadro (c)**, nos indica que los clientes pertenecientes al segmento 3 realizan sus transacciones por vía web, en **cuadro (d)**, nos indica que los clientes del segmento 4 realizan las mayores transacciones por vía web, sin embargo, entre el segmento 3 y segmento 4 se observa que existe un contraste mayor por lo que los

segmentos no se comportarían tan independientes eso podría generar confusión.

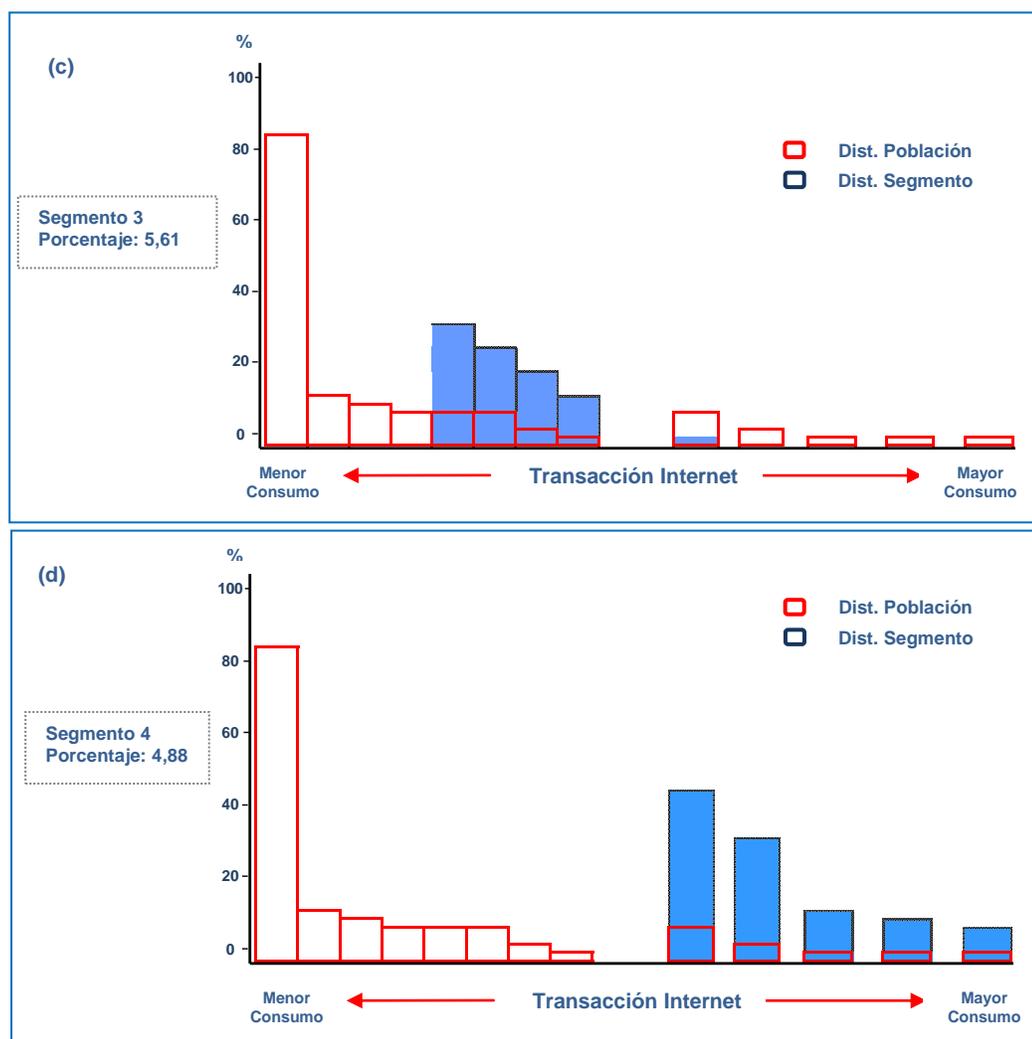
Como segunda prueba usando la variable Internet, la segmentación para determinar al grupo cibernauta se tomó 5 segmentos (K=5).

**Gráfico 5.7 Segmento para determinar al grupo cibernauta**



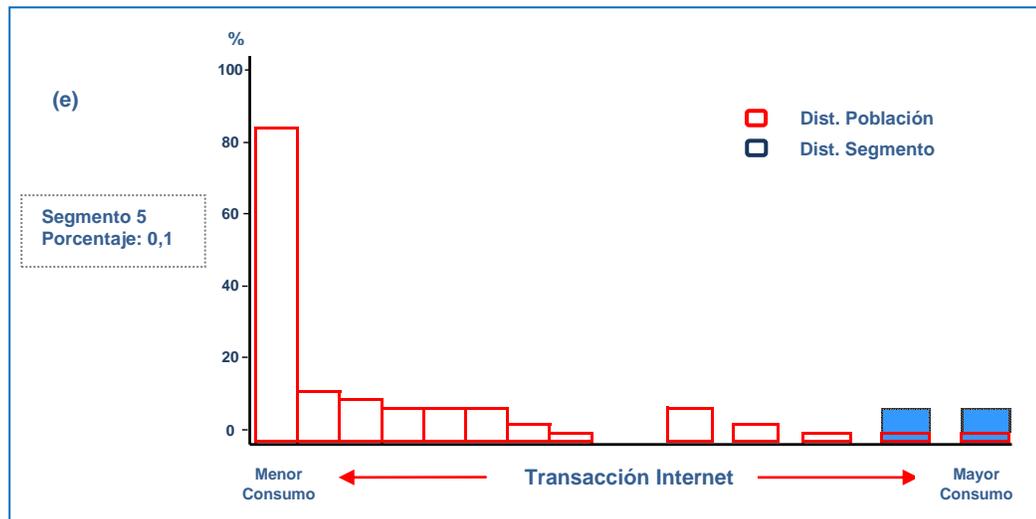
En el **Gráfico 5.7**, observamos los 2 primeros segmentos que determinó la técnica K-Means, en el **cuadro (a)**, nos indica que los clientes pertenecientes al segmento 1 se encuentran muy por debajo del promedio de la distribución de la población que no realiza sus transacciones por vía web y que son la gran parte de la población en estudio (79.4%), en **cuadro (b)**, nos indica que los clientes del segmento 2 realizan pocas transacciones por vía web, sin embargo el traslado hacia ser un grupo cibernauta esta en transición.

**Gráfico 5.8 Segmento para determinar al grupo cibernauta**



En el **Gráfico 5.8** y **Gráfico 5.9**, en el **cuadro (c)** nos indica que los clientes pertenecientes al segmento 3 realizan sus transacciones por vía web, en **cuadro (d)** y el **cuadro (e)** nos indica que los clientes del segmento 4y del segmento 5 realizan las mayores transacciones por vía web, por otro lado, observamos que entre el segmento 3 y el segmento 4 el contraste que existe es mínimo por lo considerar que ambos segmentos se comportaron independientemente.

**Gráfico 5.9 Segmento para determinar al grupo cibernauta**

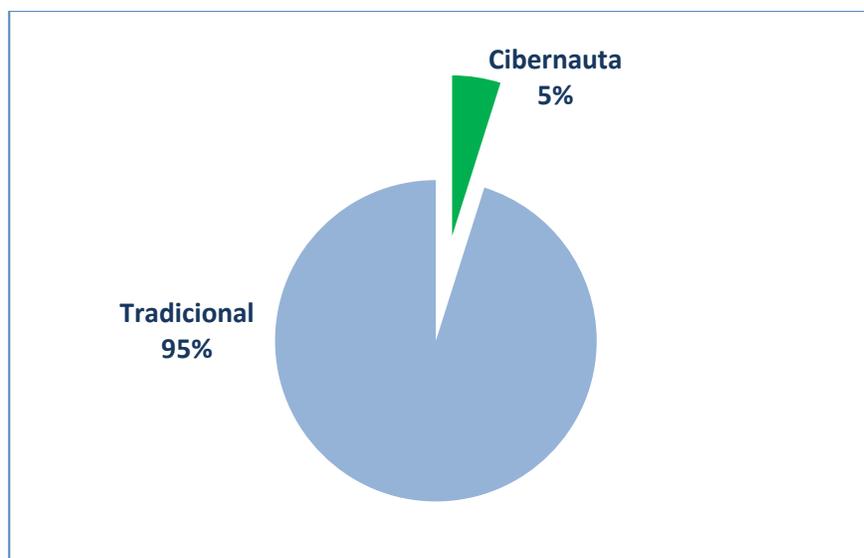


Luego de las 2 pruebas de segmentación con la técnica K-Means hemos observado que la segunda prueba fue la más relevante, ya que el segmento de aquellos clientes que realizan sus transacciones con mayor frecuencia por vía web (**clientes cibernautas**) se distinguen de aquellos que no usan el canal internet o lo realizan con poca frecuencia que suele identificarlos como **clientes tradicionales**.

### Característica del segmento Cibernauta

Luego de seleccionar el segmento adecuado para definir el corte de transacciones para considerar a un cliente cibernauta, en el **Gráfico 5.10**, indica que del total de la población en estudio (150 mil clientes), sólo el 5% de los clientes son considerados cibernauta, el 95% restante pertenecen aquellos clientes que en su mayoría realizan sus transacciones en otros establecimientos como: ventanilla, ATM (cajeros), tiendas, etc.

**Gráfico 5.10 Distribución de clientes cibernautas y tradicionales**



Fuente: Propio

**Cuadro 5.2 Proporción de transacciones por canal Internet**

	Segmento Cibernauta	Población en estudio (Cibernauta + tradicional)
% transacción	42%	4%

Fuente: Propio

En el **Cuadro 5.2**, nos indica como el segmento cibernauta en promedio realizan el 42% de transacciones por vía internet mientras que en la población en estudio solo el **4%** los realiza por internet, también destacar que para ser asignado cliente cibernauta, como mínimo debió realizar el **26%** de sus transacciones totales por el canal internet.

### 1.2.2 Grupo Tradicional (No cibernauta)

Para el caso del grupo tradicional trabajaremos con las variables de rubros de consumo que son 16 rubros luego de la reducción de variables con el análisis factorial, en este caso exploraremos la cantidad de segmentos adecuados para la investigación, por ello analizaremos el evolutivo de segmentos con la técnica de segmentación K-Means asignando como punto inicial K=4 (Referencia).

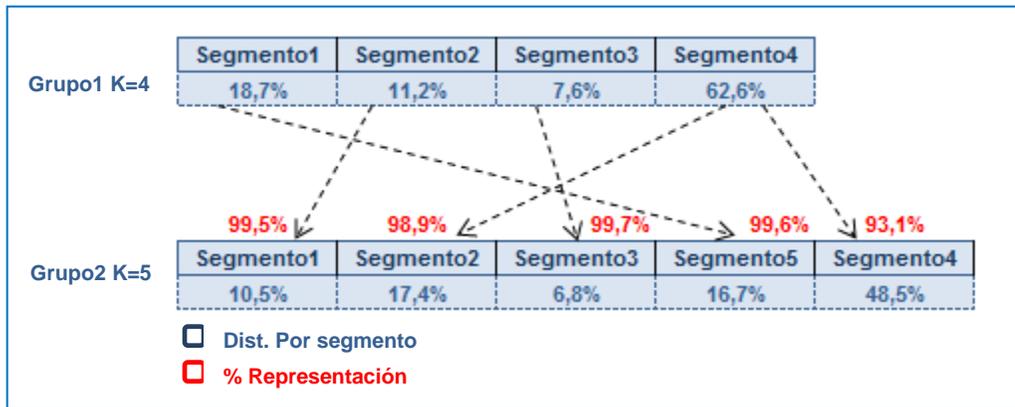
Gráfico 5.11 K-Means K=4 segmentos

	Segmento1	Segmento2	Segmento3	Segmento4
Grupo1 K=4	18.7%	11.2%	7.6%	62.6%

Dist. por segmento

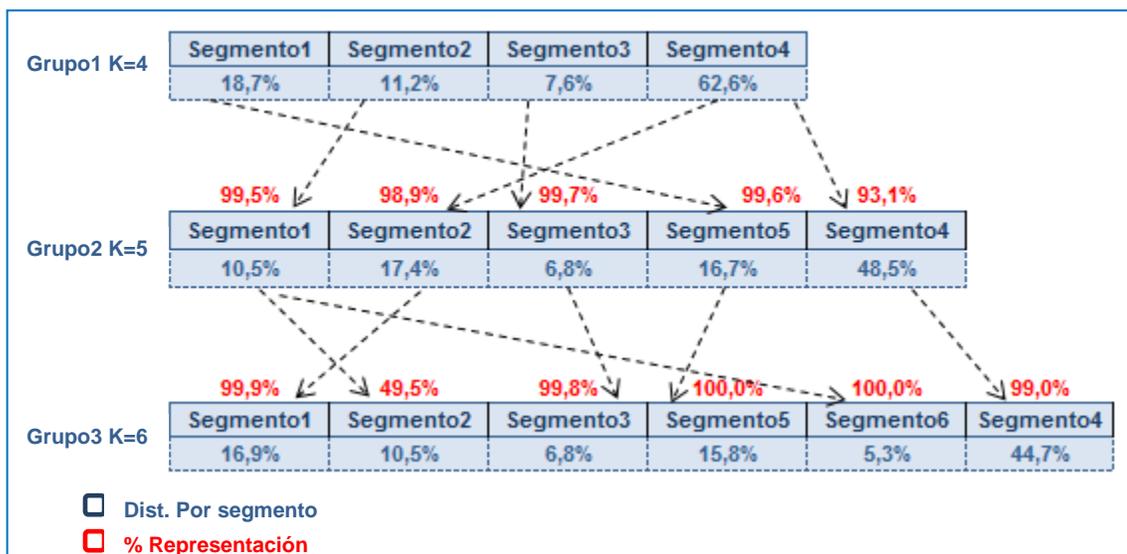
En el **Gráfico 5.11**, nos muestra la primera parte en la exploración de los segmentos donde Grupo1 divide a la población en estudio en 4 segmentos.

Gráfico 5.12 K-Means K=5 segmentos



Ahora en el **Gráfico 5.12**, tenemos al Grupo 2 que divide en 5 segmentos, sin embargo, se observa que estos segmentos son incluidos gran parte de su información en los segmentos del Grupo anterior (Grupo 1), por ejemplo, el segmento 1 del Grupo 2 que representa el 10.5% de la población en estudio se encuentra incluido en un 99.5% en el segmento 2 del Grupo 1, para el segmento 1 creado en el Grupo 2 careció de aportación ya que se pudo utilizar los segmentos del Grupo 1.

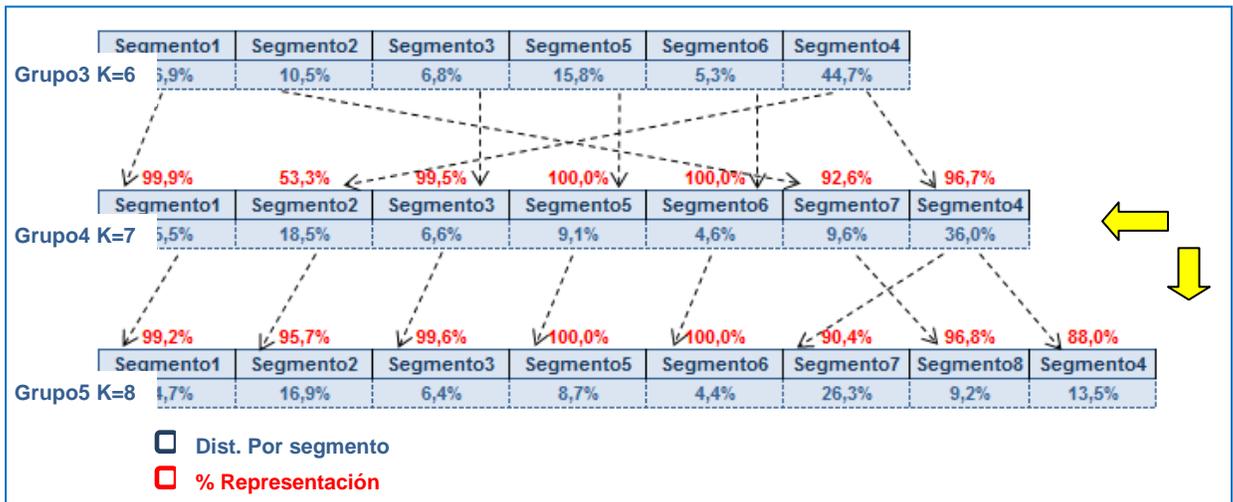
Gráfico 5.13 K-Means K=6 segmentos



Para el **Gráfico 5.13**, se observa que el Grupo 3 que divide en 6 segmentos a la población en estudio, se crearon segmentos relevantes que aportan nueva información tal como el segmento 2 donde solo está incluido el 49,5% de su información en el segmento 1 del Grupo 2.

Esto también nos indica que el segmento 2 estaría compuesto de información de varios segmentos del Grupo 1 a parte que estamos diferenciando aún más nuestra segmentación teniendo más división con información relevante.

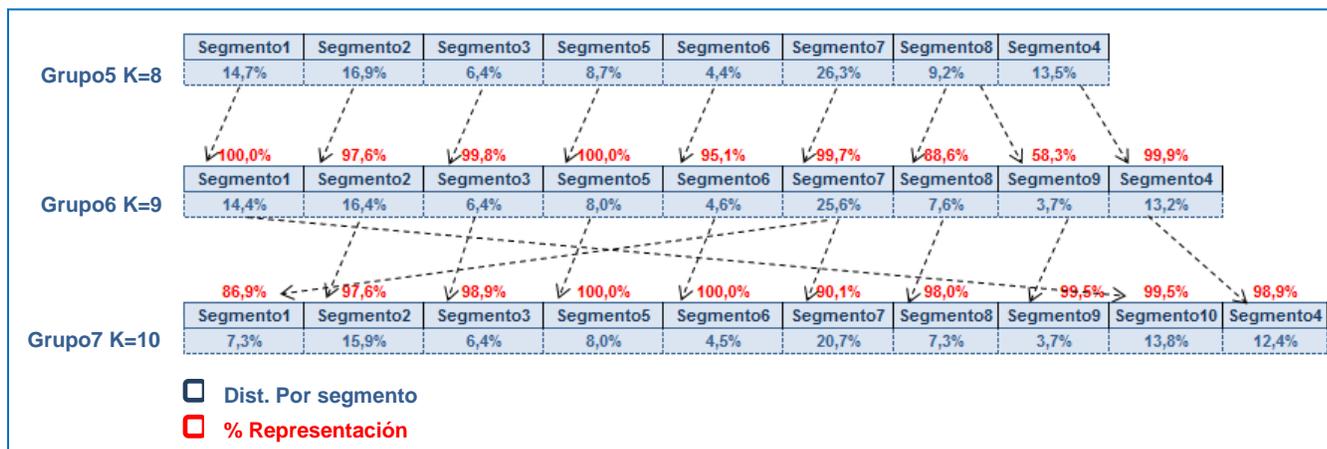
**Gráfico 5.14 K-Means K=8 segmentos**



En el **Gráfico 5.14**, se observa que tanto el Grupo 4 que divide a la población en 7 segmentos y el Grupo 5 que divide en 8 segmentos, ya no estarían ganando información a pesar de tener mayor división, por ejemplo, para el Grupo 5 en el segmento 5 está incluido toda su información en el segmento 5 del Grupo 4, de la misma forma para los

segmentos del Grupo 4 con respecto a los segmentos del Grupo 3 tampoco se ve una aportación de información relevante, como es el caso del segmento 6 del grupo 4 que es el mismo segmento que el grupo 3.

Gráfico 5.15 K-Means K=10 segmentos



Por último en el **Gráfico 5.15**, se observa que los segmentos del Grupo 6 no aportan información relevante con respecto al Grupo anterior, de la misma forma para el Grupo 7 los segmentos se encuentran incluidos en su mayoría por el Grupo anterior.

Por tanto para una segmentación satisfactoria se debe elegir la división de segmentos que aporte información relevante, que se diferencie de los otros segmentos y que cada segmento este incluido de información semejante, por ello se eligió entre el Grupo 4 (7 segmentos) y el Grupo 5 (8 segmentos) como segmentos finales ya que a partir de ahí los siguientes grupos no aportaban información que se aprovecharía para el estudio.

### ***Evaluación de segmentos***

Dado que tenemos que elegir entre 7 segmentos u 8 segmentos, evaluaremos que segmentación es la más adecuada evaluando la estructura que tienen como:

- **Cohesión del segmento (a):** Es la distancia que tiene cada elemento dentro del segmento o grupo con respecto al centroide del segmento, esto indica que tan agrupada se encuentran los elementos dentro del segmento.

$$a_{ic} = \sum_{k=1}^K \frac{|C_k|}{n} \|g_k - g\|^2$$

Donde  $g$ : centro Total

- **Distancia exterior entre segmentos (b):** Es la distancia exterior entre segmentos, esto nos informa que tan diferente pueden ser los segmentos divididos, a mayor distancia mayor independencia entre ellas.

$$b_{ic} = \frac{1}{n} \sum_{k=1}^K \sum_{i \in C_k} \|x_i - g_k\|^2$$

Donde  $g$ : centro Total

- **Silueta o forma del segmento:** es la forma o la pureza que tienen los segmentos entre ellas está compuesta por la distancia interior o inercia inter clase y la inercia intra clase.

$$s_{ic} = (b_{ic} - a_{ic})\max\{a_{ic}, b_{ic}\}$$

**Cuadro 5.3 Evaluación del segmento tradicional**

<b>GRUPO1 K=4</b>	a) Número de segmento	4
	b) Cohesión	0,116
	c) Exterior	0,095
	d) Silueta	0,380
<b>GRUPO2 K=5</b>	a) Número de segmento	5
	b) Cohesión	0,093
	c) Exterior	0,118
	d) Silueta	0,380
<b>GRUPO3 K=6</b>	a) Número de segmento	6
	b) Cohesión	0,088
	c) Exterior	0,123
	d) Silueta	0,365
<b>GRUPO4 K=7</b>	a) Número de segmento	7
	b) Cohesión	0,081
	c) Exterior	0,130
	d) Silueta	0,340
<b>GRUPO5 K=8</b>	a) Número de segmento	8
	b) Cohesión	0,072
	c) Exterior	0,139
	d) Silueta	0,352
<b>GRUPO6 K=9</b>	a) Número de segmento	9
	b) Cohesión	0,086
	c) Exterior	0,146
	d) Silueta	0,355
<b>GRUPO7 K=10</b>	a) Número de segmento	10
	b) Cohesión	0,064
	c) Exterior	0,147
	d) Silueta	0,367



Fuente: Propio

En el **Cuadro 5.3**, nos confirma que la división más apropiada para el grupo tradicional es la división de 8 segmentos (Grupo 5), ya que, la distancia interna en el segmento es menor que la del Grupo 4, así como la distancia externa es mayor y nos indica que los segmentos son más diferenciados entre ellos comparado al grupo 4 (7 segmentos).

**Cuadro 5.4 Comparación de segmentos por propuesta**

Indicador de segmento Primera Propuesta		Indicador de segmento Segunda Propuesta	
<b>GRUPO1 K=4</b>	a) Numero de segmento	<b>4</b>	<b>4</b>
	b) Cohesion	0.356	0.116
	c) Exterior	0.004	0.095
	d) Silueta	0.28	0.38
<b>GRUPO2 K=5</b>	a) Numero de segmento	<b>5</b>	<b>5</b>
	b) Cohesion	0.35	0.093
	c) Exterior	0.09	0.118
	d) Silueta	0.28	0.38
<b>GRUPO3 K=6</b>	a) Numero de segmento	<b>6</b>	<b>6</b>
	b) Cohesion	0.28	0.088
	c) Exterior	0.11	0.123
	d) Silueta	0.29	0.365
<b>GRUPO4 K=7</b>	a) Numero de segmento	<b>7</b>	<b>7</b>
	b) Cohesion	0.19	0.081
	c) Exterior	0.125	0.13
	d) Silueta	0.31	0.34
<b>GRUPO5 K=8</b>	a) Numero de segmento	<b>8</b>	<b>8</b>
	b) Cohesion	0.21	0.072
	c) Exterior	0.119	0.139
	d) Silueta	0.3	0.352
<b>GRUPO6 K=9</b>	a) Numero de segmento	<b>9</b>	<b>9</b>
	b) Cohesion	0.195	0.086
	c) Exterior	0.122	0.146
	d) Silueta	0.298	0.355
<b>GRUPO7 K=10</b>	a) Numero de segmento	<b>10</b>	<b>10</b>
	b) Cohesion	0.231	0.064
	c) Exterior	0.128	0.147
	d) Silueta	0.3	0.367

Fuente: Propio

En el **Cuadro 5.4**, observamos la comparación entre segmentos de las dos propuestas donde la segunda propuesta es más adecuado debido a que tiene menor distancia internan entre segmentos y mayor distancia externa que los diferencia. Por tanto en este estudio continuaremos explorando la segmentación con la segunda propuesta de pública objetivo.

Ahora identificaremos que rubros de consumo representan más en los 8 segmentos seleccionados para el grupo tradicional.

**Cuadro 5.5 Proporción de transacciones de consumo por segmento**

Rubro Consumo	Segmento 1	Segmento 2	Segmento 3	Segmento 4	Segmento 5	Segmento 6	Segmento 7	Segmento 8
INTER	0,8%	2,4%	55,0%	1,9%	1,2%	0,4%	3,9%	1,2%
SUPER	61,1%	17,4%	6,9%	17,9%	4,5%	4,8%	19,5%	12,8%
SALUD	9,5%	37,7%	8,4%	10,3%	77,3%	4,8%	9,5%	14,9%
TIENDA	9,0%	12,1%	6,9%	39,5%	4,1%	3,3%	11,1%	9,1%
RESTA	7,2%	8,6%	6,6%	10,0%	1,8%	2,3%	21,8%	7,7%
SERVICIO	1,4%	3,5%	2,1%	2,8%	1,3%	0,8%	2,8%	2,3%
BAZAR	1,4%	1,9%	1,6%	3,2%	0,6%	0,4%	2,2%	1,3%
LIBRE	0,4%	0,6%	0,6%	0,6%	0,2%	0,1%	0,7%	0,3%
ESTACION	2,0%	3,6%	2,5%	2,2%	0,8%	1,0%	12,3%	3,5%
EDUCA	0,2%	0,4%	0,3%	0,4%	0,1%	0,1%	0,5%	0,3%
RECREACION	0,9%	1,5%	1,9%	1,8%	0,4%	0,6%	3,9%	1,7%
VIAJERO	0,8%	1,7%	2,1%	1,9%	0,8%	0,7%	2,3%	1,6%
EMPRESARIAL	0,5%	0,7%	0,7%	0,9%	0,3%	0,2%	1,0%	0,5%
PAGO_SERV	1,7%	3,2%	2,3%	2,8%	1,1%	1,2%	4,3%	3,0%
DISPO	2,9%	3,9%	1,6%	3,3%	5,4%	78,9%	2,8%	38,6%
CASIN	0,2%	0,8%	0,3%	0,3%	0,1%	0,5%	1,2%	1,2%
OTROS	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%

Fuente: Propio

En el **Cuadro 5.5**, nos muestra que rubros de consumo es representativa para cada segmento donde:

- *Segmento1* está representado por el rubro **Supermercado**.
- *Segmento2* está representado por el rubro **Salud**.
- *Segmento3* está representado por el rubro **Internacional**.
- *Segmento4* está representado por el rubro **Tienda por departamento**.
- *Segmento5* está representado por el rubro **Salud**
- *Segmento6* está representado por el rubro **Disposición de efectivo**.
- *Segmento7* está representado por varios rubros.
- *Segmento8* está representado por el rubro **Disposición de efectivo**.

Sin embargo, tanto el Segmento 2 como el Segmento 5 los representa el rubro Salud, lo que podrían tener características similares, al igual que los segmentos 6 con el segmento 8.

Para ambos casos decidimos agruparlos en un solo grupo, teniendo las distancias de los segmentos observaremos que segmentos se encuentran tan cerca como para agruparlos.

Para calcular la distancia entre segmentos, usamos la distancia euclidiana:

$$D_{ic} = \sqrt{\sum (X_{ij} - X_{ic})^2}$$

**Cuadro 5.6 Distancia de segmentación**

	Segmento1	Segmento2	Segmento3	Segmento4	Segmento5	Segmento6	Segmento7	Segmento8
Segmento1	0							
Segmento2	0,52235	0						
Segmento3	0,76694	0,61513	0					
Segmento4	0,52968	0,38842	0,63464	0				
Segmento5	0,88561	<b>0,43168</b>	0,87808	0,77485	0			
Segmento6	0,94952	0,83737	0,94956	0,85537	103,283	0		
Segmento7	0,45601	0,32645	0,55885	0,32658	0,73798	0,8149	0	
Segmento8	0,60279	0,41929	0,66	0,47289	0,71659	<b>0,40229</b>	0,41674	0

Fuente: Propio

Efectivamente en el **Cuadro 5.6**, comprobamos que el segmento 5 y el segmento 2 con una distancia de 0.431 se encuentran más cerca comparado con el resto de segmentos, por lo que consideramos que sus características son similares y por tanto conformarían un solo segmento, de la misma forma el segmento 6 y el segmento 8 con una distancia externa de 0.402 están más cerca comparado con el resto de segmentos, por ello solo conformarían un solo segmento.

Ahora de los 8 segmentos que representaba el grupo tradicional, ahora estará conformada por 6 segmentos que son:

- *Segmento1* está representado por el rubro **Supermercado**.
- *Segmento2* está representado por el rubro **Salud**.
- *Segmento3* está representado por el rubro **Internacional**.
- *Segmento4* está representado por el rubro **Tienda por departamento**.
- *Segmento5* está representado por varios rubros
- *Segmento6* está representado por el rubro **Disposición de efectivo**.

Considerando variables sociodemográficas y las características que tienen los rubros de consumo comparando con la información sociodemográfica de la población, se definieron los segmentos de clientes.

## 1.2.1. Característica del Grupo Tradicional

### a) Segmento Familia

Gráfico 5.16 Variables del segmento Familia



Fuente: Propio

El segmento Familia se caracteriza por aquellos clientes que velan por el bienestar de las personas que les rodean, es hogareña, suelen ser personas maduras con hijos, etc. Tal como se observara en la caracterización del segmento. En el **Gráfico 5.16**, nos indica como el rubro Supermercado está por encima del promedio de los que consumen en toda la población de estudio.

Cuadro 5.7 Proporción de transacciones familia

	Segmento Familia	Población en estudio (Tradicional)
% transacción	61%	21%

Fuente: Propio

En el **Cuadro 5.7** nos muestra que los clientes del segmento **Familia** realizan el 61% de todas sus transacciones en consumo por supermercados en comparación con el total de clientes que consumen solo el 21% de todas sus transacciones.

*Caracterización del **segmento Familia**:*

**Cuadro 5.8 Descripción sociodemográfica familia**

	Segmento (promedio)	Población (promedio)
<b>Transacción</b>	2,8	3,3
<b>Consumo</b>	<b>S/. 335</b>	S/. 509
<b>Sueldo</b>	<b>S/. 1,729</b>	S/. 2,239
<b>Edad</b>	45	41
<b>Lima</b>	72,0%	69,4%
<b>Maculino</b>	<b>45,0%</b>	53,3%
<b>Segmento</b>		
<b>Banco:</b>		
<b>S1A</b>	3%	8%
<b>S1BC</b>	6%	12%
<b>S2</b>	22%	21%
<b>S3</b>	27%	32%
<b>S4</b>	3%	5%
<b>S5</b>	<b>29%</b>	18%
<b>S6</b>	<b>10%</b>	4%

Fuente: Propio

En el **Cuadro 5.8**, observamos la información descriptiva sociodemográfica del segmento comparado con la información de la población donde en este grupo se encuentran los clientes con un sueldo promedio bajo, generalmente son femeninas (55%), se trata de personas adultas y el monto consumido suele ser menor que el consumo promedio de la población (S/. 509).

Por otro lado, en este grupo se encuentran los clientes que pertenecen al segmento banco S5 y S6 que son los grupos bajos económicamente y de edad mayor, este grupo se asemeja a las personas mayores que podría tratarse de madres de bajos recursos que sólo consumen lo necesario para su hogar.

**Cuadro 5.9 Descripción de sistema financiero familia**

	Segmento (promedio)	Población (promedio)
<b>Antigüedad</b>	<b>54</b>	49
<b>Num_emp_TC</b>	3	3
<b>Saldo_TC</b>	S/. 3,246	S/. 5,402
<b>Saldo_VEH</b>	S/. 23,410	S/. 24,824
<b>Saldo_PP</b>	S/. 3,041	S/. 5,136
<b>Saldo_HIP</b>	S/. 83,269	S/. 118,896
<b>Tenencia_TC</b>	65,0%	71,0%
<b>Tenencia_VEH</b>	1,0%	3,0%
<b>Tenencia_PP</b>	<b>22,0%</b>	28,0%
<b>Tenencia_HIP</b>	<b>3,0%</b>	6,0%

Fuente: Propio

También observamos la información que tiene este segmento en el sistema financiero, donde en el **Cuadro 5.9** observamos que son clientes que tienen experiencia por tener en promedio 54 meses en el sistema, sin embargo, la proporción de clientes que tienen préstamos personales y préstamos hipotecarios están por debajo que la población, lo que nos indica que en este grupo en su mayoría no aprueban para algún préstamo o no están en las condiciones de adquirirlas ya sea por bajos recursos, bajo sueldo, etc.

## b) Segmento Conservador

Gráfico 5.17 Variables del segmento Conservador



Fuente: Propio

El segmento Conservador se caracteriza por aquellos clientes que solo realizan consumos en momentos de emergencias, se preocupa por si mismo, suelen ser jóvenes con ingresos regulares o bajos. En el **Gráfico 5.17**, nos indica como el rubro Salud está por encima del promedio de los que consumen en toda la población de estudio.

Cuadro 5.10 Proporción de transacciones

	Segmento Conservador	Población en estudio (Tradicional)
% transacción	51%	20%

Fuente: Propio

En el **Cuadro 5.10** nos muestra que los clientes del segmento **Conservador** realizan el 51% de todas sus transacciones en consumo por pago de seguros, farmacia, consulta médica en comparación con el total de clientes que consumen sólo el 20% de todas sus transacciones.

Caracterización del **segmento Conservador**:

**Cuadro 5.11 Descripción sociodemográfica conservador**

	Segmento (promedio)	Población (promedio)
Transacción	2,6	3,3
Consumo	<b>S/. 341</b>	S/. 509
Sueldo	2,098	2,239
Edad	44	41
Lima	66,0%	69,4%
Maculino	<b>50,6%</b>	53,3%
<b>Segmento Banco:</b>		
S1A	5%	8%
S1BC	11%	12%
S2	22%	21%
S3	33%	32%
S4	3%	5%
S5	<b>20%</b>	18%
S6	<b>6%</b>	4%

Fuente: Propio

En el **Cuadro 5.11**, observamos la información descriptiva sociodemográfica del segmento comparado con la información de la población donde en este grupo se encuentran los clientes con un sueldo promedio medio, generalmente es indistinto si es femenino o masculino (50%), se trata de personas adultas y el monto consumido (S/.341) suele ser menor que el consumo promedio de la población.

Por otro lado, en este grupo se encuentran los clientes que pertenecen al segmento banco S5 y S6 que son los grupos bajos económicamente y de edad mayor, este grupo se asemeja a las personas mayores que podría tratarse de padres con recursos medio que sólo consumen en casos de emergencia.

**Cuadro 5.12 Descripción de sistema financiero conservador**

	Segmento (promedio)	Población (promedio)
<b>Antigüedad</b>	49	49
<b>Num_emp_TC</b>	3	3
<b>Saldo_TC</b>	5,011	5,402
<b>Saldo_VEH</b>	23,707	24,824
<b>Saldo_PP</b>	4,468	5,136
<b>Saldo_HIP</b>	<b>102,593</b>	118,896
<b>Tenencia_TC</b>	70,0%	71,0%
<b>Tenencia_VEH</b>	2,0%	3,0%
<b>Tenencia_PP</b>	28,0%	28,0%
<b>Tenencia_HIP</b>	<b>5,0%</b>	6,0%

Fuente: Propio

También observamos la información que tiene este segmento en el sistema financiero, donde en el **Cuadro 5.12** se observa que son clientes que tienen igual experiencia que la población por tener en promedio 49 meses en el sistema, sin embargo, la proporción de clientes que tienen préstamos hipotecarios (5%) están por debajo que la población (6%), lo que nos indica que en este grupo se encuentran clientes que en su mayoría no califican para créditos de alta capacidad financiera (crédito hipotecario y vehicular).

### c) Segmento Viajero

Gráfico 5.18 Variables del segmento Viajero



Fuente: Propio

El segmento Viajero se caracteriza por aquellos clientes con altos ingresos que suelen realizar turismo, se encuentran en el rango de edad mayor. En el **Gráfico 5.18**, nos indica como el rubro Internacional está por encima del promedio de los que consumen en toda la población de estudio.

Cuadro 5.13 Proporción de transacciones viajero

	Segmento Viajero	Población en estudio (Tradicional)
% transacción	55%	5%

Fuente: Propio

En el **Cuadro 5.13**, nos muestra que los clientes del segmento **Viajero** realizan el 55% de todas sus transacciones en consumo de viajes, compras en el exterior, suelen ser los clientes con mayores ingresos en comparación con el total de clientes que consumen sólo el 5% de todas sus transacciones.

Caracterización del **segmento Viajero**:

**Cuadro 5.14 Descripción sociodemográfica viajero**

	Segmento (promedio)	Población (promedio)
Transacción	4,8	3,3
Consumo	S/. 1.064	S/. 509
Sueldo	5,664	2,239
Edad	39	41
Lima	84,0%	69,4%
Maculino	55,5%	53,3%
<b>Segmento</b>		
<b>Banco:</b>		
S1A	8%	8%
S1BC	20%	12%
S2	10%	21%
S3	37%	32%
S4	16%	5%
S5	6%	18%
S6	3%	4%

Fuente: Propio

En el **Cuadro 5.14**, observamos la información descriptiva sociodemográfica del segmento comparado con la información de la población donde en este grupo se encuentran los clientes con alto sueldo (S/. 5.664), generalmente son de Lima (84%), se trata de personas no muy adultas y el monto mensual consumido (S/.1.064) suele ser mayor que el consumo promedio de la población.

Por otro lado, en este grupo se encuentran los clientes que pertenecen al segmento banco S4 y S1BC que son los grupos top por ser económicamente alto, este grupo se asemeja a las personas jóvenes y adultas con recursos altos que consumen por satisfacción como viajes, vacaciones, etc.

**Cuadro 5.15 Descripción de sistema financiero viajero**

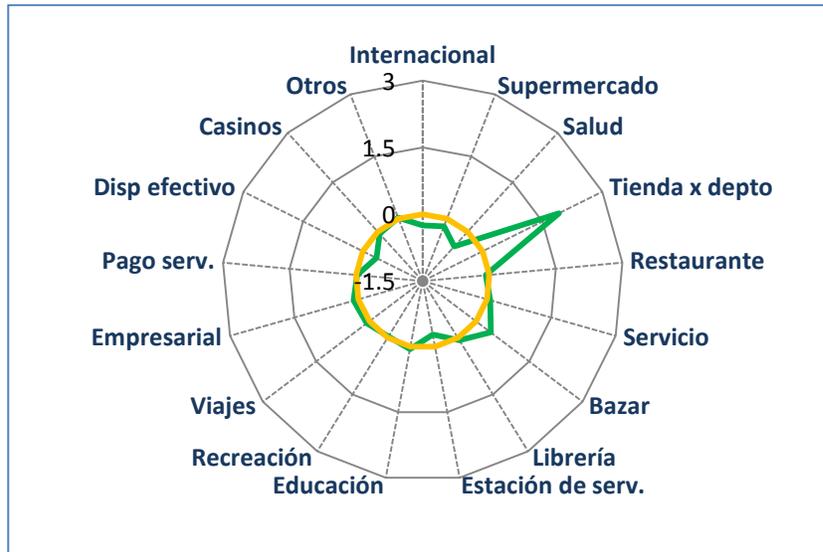
	Segmento (promedio)	Población (promedio)
Antigüedad	65	49
Num_emp_TC	3	3
Saldo_TC	7,551	5,402
Saldo_VEH	28,914	24,824
Saldo_PP	15,682	5,136
Saldo_HIP	176,703	118,896
Tenencia_TC	68,0%	71,0%
Tenencia_VEH	7,0%	3,0%
Tenencia_PP	22,0%	28,0%
Tenencia_HIP	13,0%	6,0%

Fuente: Propio

Con respecto a la información que tiene este segmento en el sistema financiero, en el **Cuadro 5.15** se observa que son clientes que tienen alta experiencia que la población por tener en promedio 65 meses en el sistema, sin embargo, la proporción de clientes que tienen préstamos personales (22%) están por debajo que la población (28%), pero la proporción de clientes con crédito hipotecario y vehicular están por encima que la población, lo que nos indica que en este grupo se encuentran clientes que califican para créditos de alta capacidad financiera (crédito hipotecario y vehicular) y no necesitan de créditos personales.

#### d) Segmento Fashion

**Gráfico 5.19 Variables del segmento Fashion**



Fuente: Propio

El segmento Fashion se caracteriza por aquellos clientes jóvenes con ingresos regulares o bajos y generalmente son. En el **Gráfico 5.19**, nos indica como el rubro Tienda por departamento está por encima del promedio de los que consumen en toda la población de estudio.

**Cuadro 5.16 Proporción de transacciones fashion**

	Segmento Fashion	Población en estudio (Tradicional)
% transacción	39%	13%

Fuente: Propio

En el **Cuadro 5.16** nos muestra que los clientes del segmento **Fashion** realizan el 39% de todas sus transacciones en consumo en tiendas por departamento, suelen ser clientes que están pendientes de la moda y por lo general destacan las mujeres en comparación con el total de clientes que consumen sólo el 13% de todas sus transacciones.

Caracterización del **segmento Fashion:**

**Cuadro 5.17 Descripción sociodemográfica fashion**

	Segmento (promedio)	Población (promedio)
Transacción	2,8	3,3
Consumo	<b>S/. 490</b>	S/. 509
Sueldo	<b>1,988</b>	2,239
Edad	<b>38</b>	41
Lima	63,0%	69,4%
Maculino	<b>32,0%</b>	53,3%
<b>Segmento Banco:</b>		
S1A	<b>9%</b>	8%
S1BC	9%	12%
S2	20%	21%
S3	32%	32%
S4	2%	5%
S5	<b>25%</b>	18%
S6	<b>3%</b>	4%

Fuente: Propio

En el **Cuadro 5.17**, observamos la información descriptiva sociodemográfica del segmento comparado con la información de la población donde en este grupo se encuentran los clientes con sueldo medio (S/. 1.988), generalmente son de Lima (63%), en su mayoría se trata de mujeres jóvenes y el monto mensual consumido (S/.490) suele ser casi igual que el consumo promedio de la población.

Por otro lado, en este grupo se encuentran los clientes que pertenecen al segmento banco S5 y S1A que son los grupos consumidores por ser económicamente medio y jóvenes, este grupo se asemeja a las personas jóvenes y adultas con recursos medios que consumen por satisfacción como tiendas por departamento.

**Cuadro 5.18 Descripción de sistema financiero fashion**

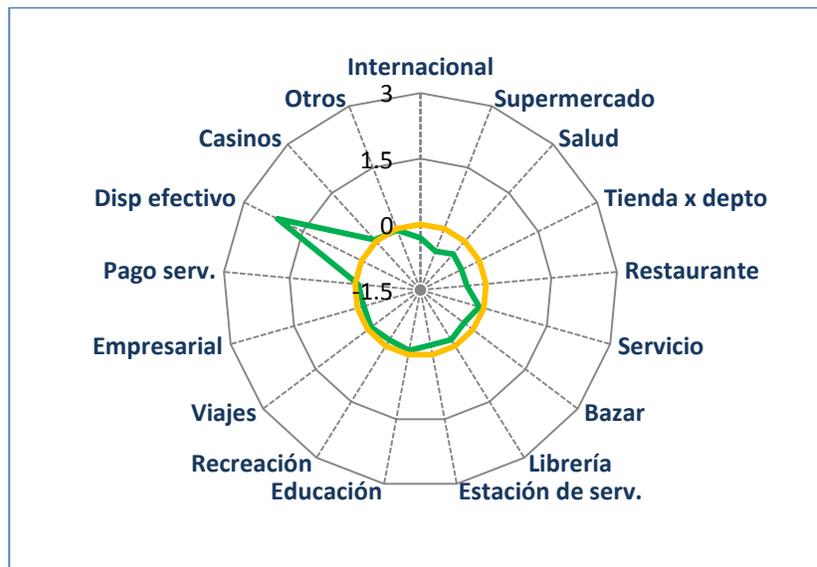
	Segmento (promedio)	Población (promedio)
Antigüedad	<b>45</b>	49
Num_emp_TC	3	3
Saldo_TC	4,501	5,402
Saldo_VEH	23,971	24,824
Saldo_PP	<b>4,210</b>	5,136
Saldo_HIP	102,861	118,896
Tenencia_TC	71,0%	71,0%
Tenencia_VEH	2,0%	3,0%
Tenencia_PP	<b>26,0%</b>	28,0%
Tenencia_HIP	5,0%	6,0%

Fuente: Propio

En el sistema financiero, en el **Cuadro 5.18** se observa que son clientes que tienen poca experiencia comparado con la población por tener en promedio 45 meses en el sistema, sin embargo, la proporción de clientes que tienen tarjetas de crédito (71%) es igual que la población, pero la proporción de clientes con crédito personal (26%) están por debajo que la población, lo que nos indica que en este grupo se encuentran clientes que no califican para créditos de alta capacidad financiera, es por ello que se sostienen de las tarjetas de crédito.

## e) Segmento Progresista

Gráfico 5.20 Variables del segmento Progresista



Fuente: Propio

El segmento Progresista se caracteriza por aquellos clientes con ansias de superación, suelen ser los pequeños negociantes que buscan disposición de efectivo para su crecimiento. En el **Gráfico 5.20**, nos indica como el rubro Disposición de efectivo por departamento está por encima del promedio de los que consumen en toda la población de estudio.

Cuadro 5.19 Proporción de transacciones progresista

	Segmento Progresista	Población en estudio (Tradicional)
% transacción	51%	10%

Fuente: Propio

En el **Cuadro 5.19**, nos muestra que los clientes del segmento **Progresista** realizan el 51% de todas sus transacciones en adquirir dinero ya sea por negocio o no controlarse en sus consumos y/o gastos

en comparación con el total de clientes que consumen sólo el 10% de todas sus transacciones.

*Caracterización del **segmento Progresista**:*

**Cuadro 5.20 Descripción sociodemográfica progresista**

	Segmento (promedio)	Población (promedio)
Transacción	2,5	3,3
Consumo	<b>S/. 648</b>	S/. 509
Sueldo	<b>1,908</b>	2,239
Edad	<b>40</b>	41
Lima	54,0%	69,4%
Maculino	<b>64,0%</b>	53,3%
<b>Segmento Banco:</b>		
S1A	<b>9%</b>	8%
S1BC	9%	12%
S2	<b>28%</b>	21%
S3	31%	32%
S4	<b>2%</b>	5%
S5	<b>16%</b>	18%
S6	4%	4%

Fuente: Propio

En el **Cuadro 5.20**, observamos la información descriptiva sociodemográfica del segmento comparado con la información de la población donde en este grupo se encuentran los clientes con sueldo medio (S/ 1.908), generalmente son de Lima (54%), en su mayoría se trata de hombres adultos y el monto mensual consumido (S/.648) suele ser mayor que el consumo promedio de la población.

Por otro lado, en este grupo se encuentran los clientes que pertenecen al segmento banco S2 y S1A que son los grupos consumidores por ser

económicamente medio, jóvenes y adultos, este grupo se asemeja a las personas jóvenes y adultas con recursos medios que consumen por necesidad de préstamo como adquirir disposición de efectivo.

**Cuadro 5.21 Descripción de sistema financiero progresista**

	Segmento (promedio)	Población (promedio)
<b>Antigüedad</b>	<b>41</b>	49
<b>Num_emp_TC</b>	3	3
<b>Saldo_TC</b>	7,587	5,402
<b>Saldo_VEH</b>	24,231	24,824
<b>Saldo_PP</b>	5,044	5,136
<b>Saldo_HIP</b>	<b>86,205</b>	118,896
<b>Tenencia_TC</b>	<b>76,0%</b>	71,0%
<b>Tenencia_VEH</b>	1,0%	3,0%
<b>Tenencia_PP</b>	<b>40,0%</b>	28,0%
<b>Tenencia_HIP</b>	<b>4,0%</b>	6,0%

Fuente: Propio

En el sistema financiero, en el **Cuadro 5.21**, se observa que son clientes que tienen poca experiencia comparado con la población por tener en promedio 41 meses en el sistema, sin embargo, la proporción de clientes que tienen tarjetas de crédito (76%) es mayor que la población al igual que la proporción de préstamos personales (40%) es mayor que la población (28%), lo que nos indica que en este grupo se encuentran clientes que sienten la necesidad de adquirir disposición de efectivo o algún préstamo personal y que son clientes que califican para estos créditos.

## f) Segmento Sofisticado

**Gráfico 5.21 Variables del segmento Sofisticado**



Fuente: Propio

El segmento Sofisticado se caracteriza por aquellos clientes entre ingresos regulares y altos, suelen ser personas más cordiales, conocedoras del sistema financiero, etc. En el **Gráfico 5.21**, nos indica como varios rubros están por encima del promedio de los que consumen en toda la población de estudio.

**Cuadro 5.22 Proporción de transacciones sofisticado**

	Segmento Sofisticado	Población en estudio (Tradicional)
% transacción	64%	51%

Fuente: Propio

En el **Cuadro 5.22**, nos muestra que los clientes del segmento **Sofisticados** realizan el 64% de todas sus transacciones, transaccionan

en diferentes rubros en proporciones similares en comparación con el total de clientes que consumen sólo el 51% de todas sus transacciones.

*Caracterización del **segmento Sofisticado**:*

**Cuadro 5.23 Descripción sociodemográfica sofisticado**

	Segmento (promedio)	Población (promedio)
Transacción	5,4	3,3
Consumo	<b>S/. 694</b>	S/. 509
Sueldo	<b>3,315</b>	2,239
Edad	<b>39</b>	41
Lima	76,8%	69,4%
Maculino	<b>63,2%</b>	53,3%
<b>Segmento</b>		
<b>Banco:</b>		
S1A	<b>11%</b>	8%
S1BC	<b>16%</b>	12%
S2	18%	21%
S3	32%	32%
S4	<b>8%</b>	5%
S5	<b>13%</b>	18%
S6	<b>2%</b>	4%

Fuente: Propio

En el **Cuadro 5.23**, observamos la información descriptiva sociodemográfica del segmento comparado con la información de la población donde en este grupo se encuentran los clientes con sueldo mayor (S/ 3.315), generalmente son de Lima (76%), en su mayoría se trata de hombres adultos y el monto mensual consumido (S/.694) suele ser mayor que el consumo promedio de la población.

Por otro lado, en este grupo se encuentran los clientes que pertenecen al segmento banco S4, S1BC y S1A que son los grupos top y consumidores

por ser económicamente altos, jóvenes y adultos, este grupo se asemeja a las personas jóvenes y adultas con recursos altos que consumen en diversos rubros.

**Cuadro 5.24 Descripción de sistema financiero sofisticado**

	Segmento (promedio)	Población (promedio)
<b>Antigüedad</b>	49	49
<b>Num_emp_TC</b>	3	3
<b>Saldo_TC</b>	6,285	5,402
<b>Saldo_VEH</b>	<b>25,127</b>	24,824
<b>Saldo_PP</b>	7,302	5,136
<b>Saldo_HIP</b>	<b>138,478</b>	118,896
<b>Tenencia_TC</b>	71,0%	71,0%
<b>Tenencia_VEH</b>	<b>5,0%</b>	3,0%
<b>Tenencia_PP</b>	27,0%	28,0%
<b>Tenencia_HIP</b>	<b>8,0%</b>	6,0%

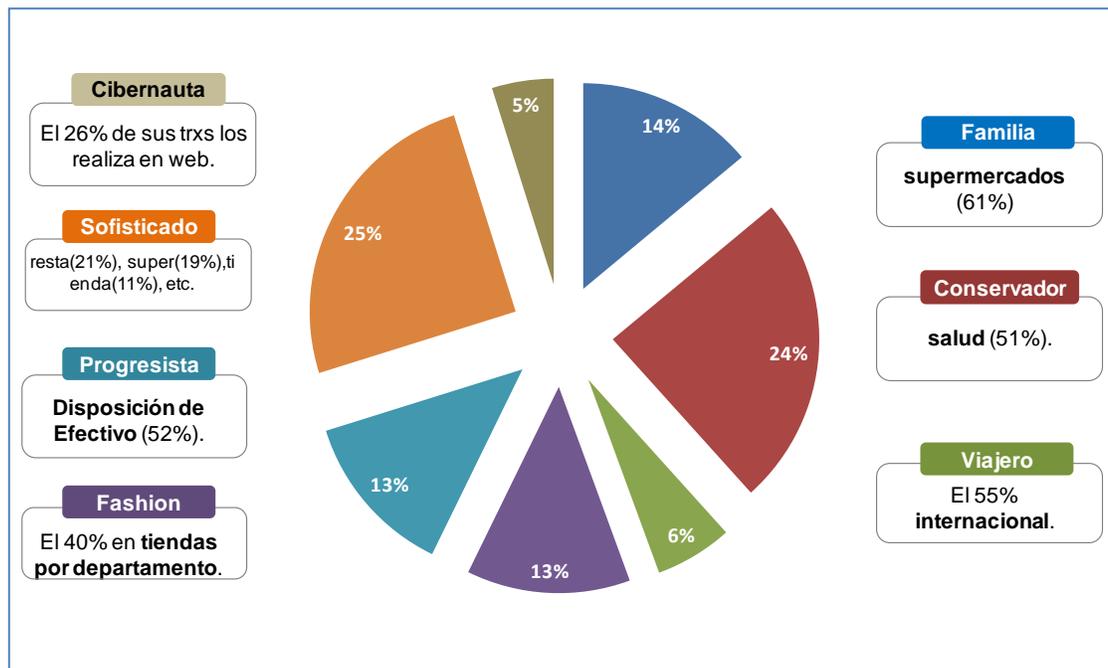
Fuente: Propio

La descripción en el sistema financiero de este segmento se observa en el **Cuadro 5.24** que son clientes que tienen igual experiencia comparado con la población por tener en promedio 49 meses en el sistema, sin embargo, la proporción de clientes que tienen créditos vehiculares (5%) y créditos hipotecarios (8%) están por encima que la población, lo que nos indica que en este grupo se encuentran clientes que no necesitan de algún crédito personal o disposición de efectivo sino que invierten en créditos hipotecarios o vehicular que están calificados para adquirir créditos de alta capacidad financiera.

## 2. RESULTADOS

### 2.1 Segmentos Finales

**Gráfico 5.22 Segmentación de clientes por transacción de consumo (Tradicional – Cibernauta)**

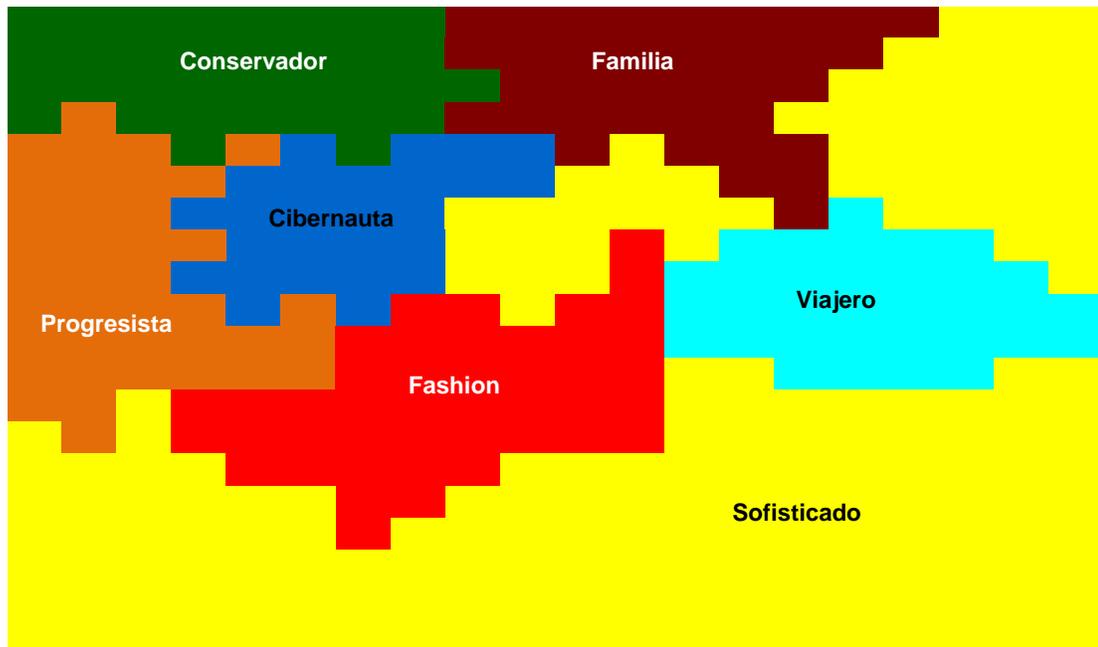


Fuente: Propio

En el **Gráfico 5.22**, se observa los 7 segmentos finales que determinan la segmentación de clientes por comportamiento de consumo, donde está involucrado el segmento cibernauta y los 6 segmentos del grupo tradicional (no usan con frecuencia el internet para realizar sus transacciones).

## 2.2 Análisis de vecindades entre segmentos

Gráfico 5.23 Mapa de visualización de segmentos de clientes



Fuente: Propio

Ya determinado los 7 segmentos finales, ahora en el **Gráfico 5.23**, muestra el **mapa de auto-organizado de Kohonen** el cual nos ayuda a visualizar como se relacionan los segmentos, por ejemplo, se comprueba como el segmento Viajero que es un perfil de alto ingreso no requiere de disposición de efectivo (Progresista) que es un segmento donde destacan aquellos clientes con necesidad de adquirir dinero, tal como se observa en el mapa ambos segmentos se encuentran lejos.

Por otro lado el segmento Conservador que se caracteriza por tener ingresos regulares se ve que en momentos puede tener la necesidad de adquirir dinero, tal como se observa en el mapa ambos segmentos se encuentran juntos.

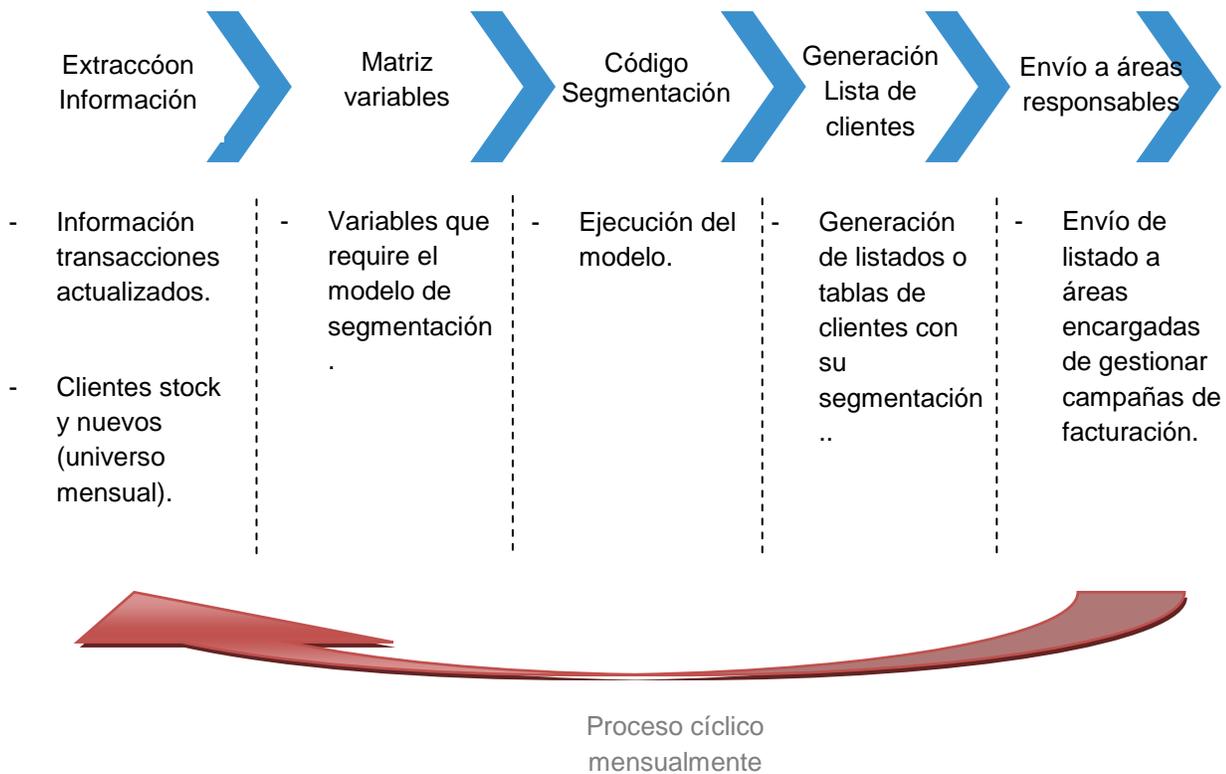
### **2.3 Proceso de Réplica o Actualización de segmentos**

Luego de haber identificado los segmentos, perfil de segmentos y la relación que existe entre ellas, surge la necesidad de implementar este modelo con miras a la gestión comercial. Por ello se desarrollo un proceso el cual ordena toda la implementación de la réplica (actualización) del modelo de segmentación.

El proceso empezará con la confirmación que la información de transacciones de tarjeta de crédito se encuentre actualizada, esto proviene de la base origen de cada entidad bancaria, luego se procederá a construir la matriz de variables según lo que el modelo necesitaría, luego nos aseguramos que la información este muy bien construida mediante la validación de la matriz y por ultimo ejecutamos el modelo que segmentará a los clientes actuales y a los nuevos que se incorporan a la entidad bancaria que cuenten por lo menos con una tarjeta de crédito.

Con ello se generará el listado de clientes con su segmento identificado a través del consumo que realiza, y se enviaran a las aéreas responsables de monitorear los indicadores de consumo o facturación en tarjeta de crédito.

**Gráfico 5.24 Proceso de réplica de segmentación de clientes**



Fuente: Propio

Tener en cuenta que el proceso de réplica se actualiza mensualmente debido a que las campañas se realizan en esos periodos y por tanto necesitan de información más actualizada posible.

## CONCLUSIONES

- a) Para la optimización y mejores decisiones comerciales se contribuyó en el desarrollo de la segmentación bajo el comportamiento de consumo de los clientes de una entidad bancaria.
- b) Los segmentos de clientes a través del comportamiento de consumo están determinados por 7 segmentos, 1 segmento cibernauta y 6 segmentos del grupo tradicional.
- c) Se destaca que para un mejor direccionamiento de las ofertas comerciales se tiene que enfocar a los segmentos caracterizados como: Familia, Conservador, Viajero, Progresista, Sofisticado y el Cibernauta.
- d) Se desarrolló la metodología de segmentación de consumo que realizan los clientes basada en la segmentación K-means conjuntamente con el análisis de vecindad entre segmentos (mapa auto-organizado de Kohonen) que se actualizará mensualmente para fines comerciales.

## RECOMENDACIONES

- a) Para medir la performance del modelo de segmentación de clientes en las campañas, se recomienda realizar el seguimiento durante los tres meses posteriores.
- b) Si bien cierto que las transacciones y el monto consumido están correlacionados, se recomienda realizar un modelo usando monto de rubros de consumo y no transacciones como el presentado.
- c) El modelo de segmentación está dirigido a clientes con suficiente información de consumo (transacción), para aquellos que tienen poca o ningún consumo podrían usar los segmentos como target para luego desarrollar un modelo de perfil con variables globales.
- d) Para próximos estudios de comportamiento de consumo podrían incluir más dimensiones como el canal de consumo más detallado (POS, ATM, TITENDAS, etc.). para potenciar la segmentación de comportamiento.

## BIBLIOGRAFIA

- GORSUCH, R. (1983). Factor Analysis. Second Edition. LEAAFIFI, A.A. and CLARK, V. (1996) Computer-Aided Multivariate Analysis. Third Edition. Texts in Statistical Science. Chapmanand Hall.
- EVERITT, B. And GRAHAM, D. (1991). Applied Multivariate Data Analysis.
- MARDIA, K.V., KENT, J.T. y BIBBY, J.M. (1994).Multivariate Analysis. Academic Press.
- KLINE, R.B. (1998). Principles and Practice of Structural Equation Modeling. The Guilford Press.
- Pollard, D. (1981), “Strong Consistency of K-Means Clustering ”.The Annals of Statistics, Vol 9 No 1, 135–140.
- Data Mining: Practical Machine Learning Tools and Techniques, I. H. Witten, E. Frank & M. A. Hall, 3ª. Edición, Morgan Kaufmann Publishers, Elsevier, USA, 2011.Págs 285 –288.

- Bottou, L. & Bengio, Y. (1995) Convergence properties of the K-means algorithms, in *Advances in Neural Information Processing Systems*, G. Tesauro, D. Touretzky & T. Leen, eds., 7, The MIT Press, 585–592.
- Ultsch, Alfred; Siemon, H. Peter (1990), *Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis*.
- Lindsay I Smith, “A tutorial on Principal Components Analysis”, February 26, 2002.

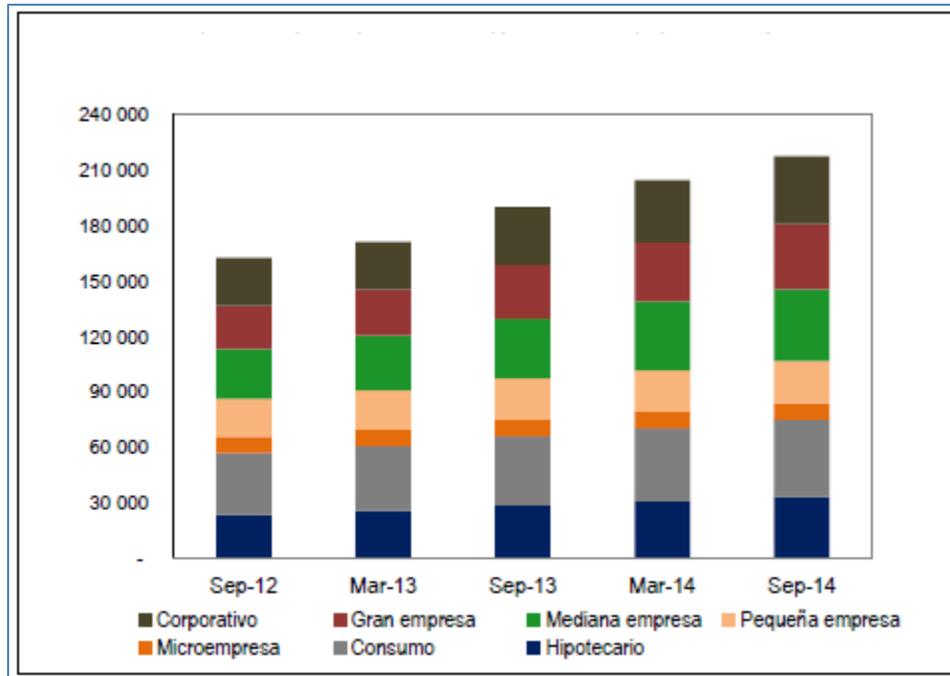
## ANEXO I

### Número de Créditos de Consumo

Empresas	2012	2013	2014	Crecimiento 2014-2013
B. Continental	505.019	508.214	488.600	-3,9%
B. de Comercio	1.824	1.471	1.229	-16,5%
B. de Crédito del Perú (con sucursales en el exterior)	979.268	1.127.228	1.174.754	4,2%
B. Financiero	144.610	174.470	179.571	2,9%
B. Interamericano de Finanzas	23.945	38.532	52.647	36,6%
Scotiabank Perú (con sucursales en el exterior)	278.248	250.778	261.156	4,1%
Citibank	140.048	131.844	123.646	-6,2%
Interbank (con sucursales en el exterior)	1.058.062	973.290	997.375	2,5%
Mibanco	7.568	5.672	50	-99,1%
B. GNB	39.007	25.940	19.909	-23,2%
B. Falabella Perú	1.302.935	1.342.302	1.416.056	5,5%
B. Santander Perú	-	-	-	-
B. Ripley	968.239	1.054.449	1.209.168	14,7%
B. Azteca Perú	44.417	96.933	127.127	31,1%
Deutsche Bank Perú	-	-	-	-
B. Cencosud	8.301	471.295	477.549	1,3%
B. ICBC	-	-	-	-
<b>TOTAL BANCA MÚLTIPLE</b>	5.501.491	6.202.418	6.528.837	5,3%

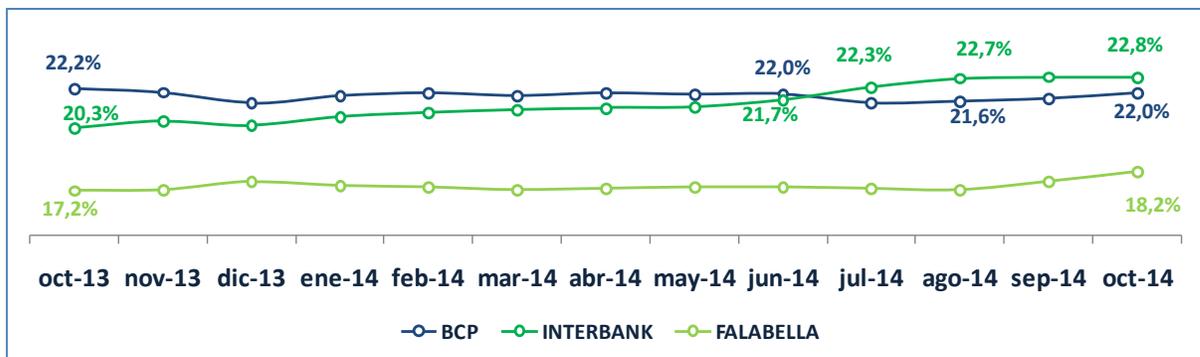
*Fuente: Reporte 7, número de Tarjetas de Crédito, Banco Internacional del Perú*

**Sistema Financiero: Evolución de los créditos por tipo  
(En millones de S/.)**



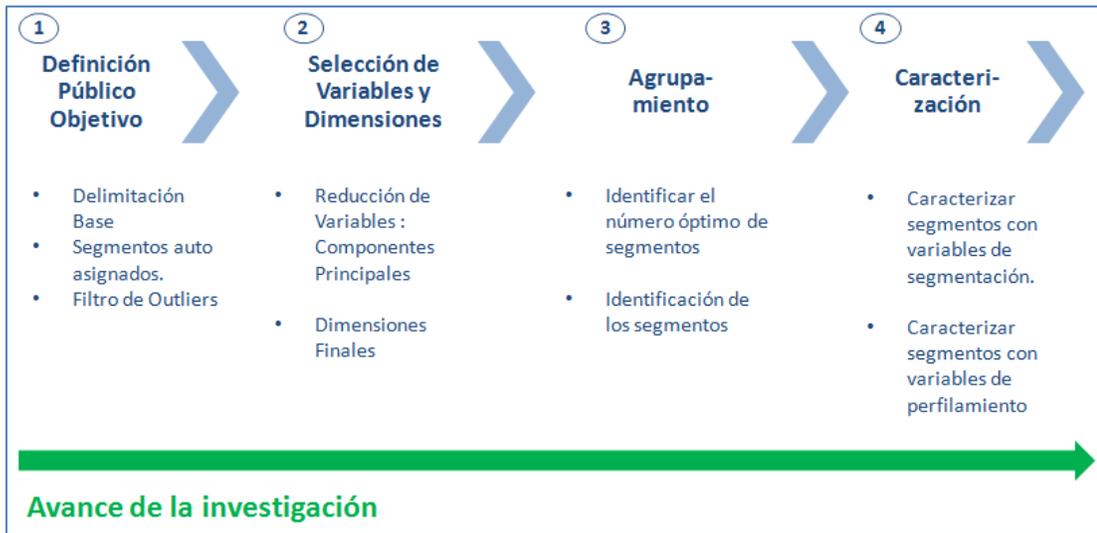
Fuente: SBS (Artículo Sep-2014)

**Participación de mercado de tarjetas de crédito en el Perú**



Fuente: Banco Internacional del Perú

## Metodología de investigación



Fuente: propio

## Mapa conceptual de Segmentos Banco

Grupos de clientes que tienen necesidades y comportamientos similares y está basado en las variables de ingresos, edad y situación laboral (Dependientes e Independientes).

	Edad < 21	Dependientes			Independientes	Retirado > 65
		Edad >= 21 y < 31	Edad >= 31 y < 46	Edad >= 46 y <= 65		
Ingresos >= S/.10K		S4 - Afluente Masivo			S4 **	S4 *
Ingresos >= S/.4K y < S/.10K		S1A - Joven	1BC - Potencial		S3 - Independiente	
Ingresos >= S/.1.2K y < S/.4K			S2 - Consumo			
Ingresos >= S/.750 y < S/.1.2K		S5 - Consumo Inicial				
Ingresos < S/.750		S6 - Aspirante				

(\*) Corte saldos pasivos de S/. 20 mil en los últimos 6 meses

(\*\*) Corte en S/.20 mil de ingresos

Fuente: Banco Internacional del Perú

**Diccionario de variables de rubros de consumo  
y demográfico**

Variable	Descripción	Fuente
Codunicocli	Identidad del cliente en la entidad financiera	DW
Codsbs	Identidad del cliente en el sistema financiero	DW
Edad	Edad del cliente	DW
Sexo	Sexo del cliente	DW
Departamento	Procedencia del cliente	DW
Sueldo	Ingreso bruto del cliente	DW
Segmento_Banco	Segmento banco (Edad - Ingreso)	DW
Saldo_SSFF_TC	Saldo de tarjeta de crédito en el sistema financiero	DW
Saldo_SSFF_PP	Saldo de préstamos personales en el sistema financiero	DW
Ant	Aerolíneas Lan Chile, Taca Internacional, Aeroméxico	DW
Trx_Internet	Aerolíneas Lan Chile, Taca Internacional, Aeroméxico	DW
Trx_No_Internet	Aerolíneas Lan Chile, Taca Internacional, Aeroméxico	DW
Trx_Aerol	Aerolíneas Lan Chile, Taca Internacional, Aeroméxico	DW
Trx_Agenc	Agencias de viaje	DW
Trx_Asoci	Organizaciones sin fines de lucro, colegios profesionales	DW
Trx_Bazar	Tiendas mayoristas,tiendas de regalo,florerias, ferreterias,cosméticos	DW
Trx_Casin	Casinos y juegos de azar	DW
Trx_Contr	Material y acabado de construccion, lavanderia, talleres de reparación de veh	DW
Trx_Dispo	Disp efectivo atm, ventanilla.	DW
Trx_Educa	Academias, universidades, estudios superiores, colegios y nidos.	DW
Trx_Empre	Empresas financieras, transferencias de fondos	DW
Trx_Equip	Equipos de telecomunicaciones, computadoras, periféricos y software	DW
Trx_Estac	Grifos y servicios, grifos gasol.	DW
Trx_Hotel	Alojamiento, hoteles, moteles y resorts, marriot.	DW
Trx_Inter	Todo rubro	DW
Trx_Libre	Librerías, imprentas y editoriales, útiles de escritorio.	DW
Trx_Pagod	Pago de servicios	DW
Trx_Pagos	Pago de impuestos, pagos de cuotas a clubes, suscripciones.	DW
Trx_Recre	Teatros, museos, espectáculos de temporada, discotecas, peñas, bares	DW
Trx_Resta	Restaurantes, cafeterias, snacks, servicio express.	DW
Trx_Salud	Venta seguros y reaseguros, farmacias, clinicas y centros médicos	DW
Trx_Servi	Servicio de consultoria, peluqueria,servicio público, servicio veterinario	DW
Trx_Super	Supermercados, minimercado, confiteria y dulceria, panaderia y pasteleria	DW
Trx_Tiend	Tienda por departamento, zapateria, tienda de ropa, boutique,	DW
Trx_Trans	Transporte terrestre, venta repuesto para vehiculos, transporte aéreo	DW
Trx_Otras	(Vacío), Centro de servicios de metalurgia.	DW

Fuente: propio

## ANEXO II

### Base de datos para la segmentación K-Means

PERIODO	CODUNICOCLI	CODSBS	CODDOC	EDAD	SUELDO	SEGMENTO_BANCO	INTER	SUPER
201408	CU927899	CO251743	COD577857	45	S/. 791	5	0,0000	0,0000
201408	CU471715	CO612885	COD401841	26	S/. 2025	1A	0,0000	0,9796
201408	CU255599	CO792883	COD579156	39	S/. 791	5	0,0000	0,4286
201408	CU815449	CO653067	COD645281	32	S/. 2637	2	0,0000	0,1707
201408	CU909676	CO795039	COD627973	30	S/. 1873	3	0,0000	0,5000
201408	CU939969	CO866649	COD479463	54	S/. 3684	3	0,0167	0,0000
201408	CU771467	CO827947	COD283717	63	S/. 2160	3	0,0000	0,0000
201408	CU839289	CO678358	COD863732	45	S/. 1096	3	0,0000	0,1935
201408	CU602589	CO464718	COD725225	49	S/. 13620	3	0,0000	0,1111
201408	CU101122	CO985227	COD224560	43	S/. 1808	3	0,0000	1,0000
201408	CU192859	CO991527	COD710649	57	S/. 2125	3	0,0000	0,1200
201408	CU721267	CO206878	COD453109	32	S/. 2536	2	0,0000	0,1148
201408	CU693946	CO120625	COD223950	50	S/. 7400	3	0,0000	0,4615
201408	CU164860	CO466570	COD439732	60	S/. 7972	3	0,0000	0,0000
201408	CU866465	CO490470	COD590695	25	S/. 1765	3	0,0000	0,3962
201408	CU725306	CO448797	COD163197	34	S/. 1965	3	0,0000	0,0765
201408	CU611494	CO290958	COD319803	38	S/. 4500	1BC	0,0000	0,0000
201408	CU893337	CO802724	COD680237	74	S/. 1221	6	0,0370	0,0000
201408	CU545137	CO799517	COD389913	72	S/. 1068	6	0,5870	0,0870
201408	CU319706	CO259240	COD226543	38	S/. 2639	2	0,0000	0,0000
201408	CU921293	CO751232	COD450710	32	S/. 2214	2	0,0000	0,0000
201408	CU805616	CO332518	COD400561	25	S/. 806	5	0,0000	0,3333
201408	CU499966	CO428148	COD336258	52	S/. 4093	3	0,3810	0,2381
201408	CU361947	CO746825	COD600365	30	S/. 962	5	0,0000	0,0526
201408	CU295781	CO323425	COD161036	28	S/. 1909	3	0,0464	0,2405
201408	CU144890	CO568059	COD248519	33	S/. 2964	3	0,0000	0,2800
201408	CU397383	CO354498	COD381928	59	S/. 1150	2	0,5714	0,0000
201408	CU986685	CO387416	COD217264	31	S/. 2164	2	0,0000	1,0000
201408	CU401114	CO627829	COD965371	60	S/. 2913	2	0,0000	0,3056
201408	CU337582	CO692319	COD239961	27	S/. 3237	1A	0,0000	0,0000
201408	CU122009	CO221916	COD397011	33	S/. 9643	1BC	0,0000	0,1250
201408	CU303136	CO420859	COD801588	29	S/. 806	5	0,0000	0,1667
201408	CU565694	CO981195	COD500144	25	S/. 1265	1A	0,0000	0,1000
201408	CU816299	CO881167	COD969109	61	S/. 2085	2	0,0000	0,1250
201408	CU386203	CO631730	COD218190	28	S/. 1997	1A	0,0000	0,0000
201408	CU807291	CO394802	COD545894	58	S/. 11392	4	0,0000	0,0000
201408	CU858272	CO506456	COD898864	28	S/. 4946	1A	0,8571	0,0357
201408	CU498356	CO995034	COD560853	59	S/. 1690	3	0,0000	0,5000

SALUD	TIENDA	RESTA	SERVICIO	BAZAR	ESTACION	LIBRE	EDUCA	RECREACION	VIAJERO
0,0000	0,1429	0,5714	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,1429
0,0000	0,0204	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,0000	0,2857	0,0714	0,0000	0,0714	0,0000	0,0714	0,0000	0,0000	0,0000
0,3171	0,1463	0,1951	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,2000	0,0000	0,2167	0,0333	0,0000	0,0000	0,0167	0,0167	0,0667	0,0000
1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,0968	0,2903	0,0000	0,0000	0,0323	0,2258	0,0000	0,0000	0,0000	0,0000
0,2963	0,2222	0,3704	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,0400	0,4000	0,0800	0,0000	0,0400	0,2800	0,0400	0,0000	0,0000	0,0000
0,0656	0,0492	0,2295	0,0000	0,0164	0,1803	0,0000	0,0164	0,0164	0,0000
0,3077	0,0769	0,0000	0,0000	0,1538	0,0000	0,0000	0,0000	0,0000	0,0000
0,0000	0,2000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,1132	0,3208	0,0189	0,0189	0,0189	0,0566	0,0000	0,0000	0,0000	0,0000
0,1093	0,0601	0,3607	0,0000	0,0164	0,2077	0,0000	0,0000	0,1148	0,0437
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,1481	0,2593	0,1852	0,1111	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,0217	0,0326	0,1087	0,0000	0,0109	0,0000	0,0000	0,0109	0,0543	0,0000
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,0000	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,3333	0,0000	0,3333	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,0000	0,0952	0,0000	0,1429	0,0476	0,0000	0,0000	0,0000	0,0952	0,0000
0,7368	0,0000	0,0526	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,1579
0,1814	0,2489	0,0253	0,0084	0,0422	0,0000	0,0000	0,0084	0,0000	0,0042
0,3600	0,2400	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,0000	0,0000	0,1429	0,0000	0,0000	0,1429	0,0000	0,0000	0,1429	0,0000
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,0694	0,5417	0,0139	0,0000	0,0417	0,0000	0,0000	0,0000	0,0139	0,0000
0,4000	0,6000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,7500	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,4375	0,1250	0,0417	0,0208	0,0208	0,0000	0,0000	0,0000	0,0000	0,0000
0,0000	0,4000	0,0000	0,3000	0,0000	0,0000	0,1000	0,1000	0,0000	0,0000
0,7500	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
0,0909	0,4545	0,0000	0,1818	0,0000	0,2727	0,0000	0,0000	0,0000	0,0000
0,0000	0,0000	0,0357	0,0000	0,0000	0,0000	0,0000	0,0000	0,0357	0,0000
0,0000	0,5000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

PAGO_SERV	EMPRESARIAL	DISPO	CASIN	OTROS	INTERNET	NOINTERNET	TRX_TOTAL	MTO_TOTAL
0,0000	0,1429	0,0000	0,0000	0,0000	0,0000	1,0000	7	S/. 3842,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	98	S/. 5104,0
0,0000	0,0000	0,0714	0,0000	0,0000	0,0000	1,0000	14	S/. 1734,0
0,0976	0,0000	0,0732	0,0000	0,0000	0,0732	0,9268	41	S/. 2683,0
0,0000	0,0000	0,5000	0,0000	0,0000	0,0000	1,0000	2	S/. 2322,0
0,1167	0,0000	0,3167	0,0000	0,0000	0,1167	0,8833	60	S/. 9710,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	3	S/. 88,0
0,1613	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	31	S/. 4933,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	27	S/. 3038,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	4	S/. 1154,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	25	S/. 3220,0
0,2459	0,0492	0,0164	0,0000	0,0000	0,1967	0,8033	61	S/. 22909,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	13	S/. 1914,0
0,4000	0,0000	0,4000	0,0000	0,0000	0,4000	0,6000	5	S/. 5755,0
0,0000	0,0000	0,0000	0,0566	0,0000	0,0000	1,0000	53	S/. 4862,0
0,0000	0,0109	0,0000	0,0000	0,0000	0,0000	1,0000	183	S/. 11189,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0	S/. ,0
0,2593	0,0000	0,0000	0,0000	0,0000	0,2593	0,7407	27	S/. 4664,0
0,0326	0,0000	0,0543	0,0000	0,0000	0,0109	0,9891	92	S/. 18480,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0	S/. ,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	1	S/. 1079,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	3	S/. 76,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,3810	0,6190	21	S/. 4970,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	19	S/. 1479,0
0,1941	0,0000	0,0000	0,0000	0,0000	0,1941	0,8059	237	S/. 26071,0
0,0800	0,0400	0,0000	0,0000	0,0000	0,0000	1,0000	25	S/. 5331,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	7	S/. 1524,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	11	S/. 4106,0
0,0139	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	72	S/. 5605,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	5	S/. 1314,0
0,0000	0,0000	0,1250	0,0000	0,0000	0,0000	1,0000	16	S/. 1427,0
0,1667	0,0000	0,0208	0,0000	0,0000	0,0000	1,0000	48	S/. 9886,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	10	S/. 1470,0
0,0000	0,0000	0,0000	0,0000	0,1250	0,0000	1,0000	8	S/. 50,0
0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	1,0000	2	S/. 1051,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	11	S/. 2567,0
0,0357	0,0000	0,0000	0,0000	0,0000	0,0714	0,9286	28	S/. 2827,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	2	S/. 2211,0

## Código de segmentación K-Means en SAS

```
*****
***          GRUPO CIBERNAUTA (SEGMENTACION)
*****;

DATA PERFIL.JSG_BASE_UNIV_SEGMENTO_FIN_01;
SET PERFIL.JSG_BASE_UNIV_SEGMENTO_FIN_01;

*** Begin Class Look-up, Standardization, Replacement ;
drop _dm_bad; _dm_bad = 0;

*** Omitted Cases;
if _dm_bad then do;
    SEGMENTO1_CIBER = .; Distance = .;
    goto CLUS10vlex ;
end; *** omitted;

*** Compute Distances and Cluster Membership;
label SEGMENTO1_CIBER = 'Id de segmento' ;
label Distance = 'Distancia' ;
array CLUS10vads [5] _temporary_;
drop _vqclus _vqmvar _vqnvar;
_vqmvar = 0;
do _vqclus = 1 to 5; CLUS10vads [_vqclus] = 0; end;
if not missing( TRX1_INTERNET ) then do;
    CLUS10vads [1] + ( TRX1_INTERNET - 0.57932489294171 )**2;
    CLUS10vads [2] + ( TRX1_INTERNET - 0.00225345544256 )**2;
    CLUS10vads [3] + ( TRX1_INTERNET - 0.0750366727331 )**2;
    CLUS10vads [4] + ( TRX1_INTERNET - 0.18646181462454 )**2;
    CLUS10vads [5] + ( TRX1_INTERNET - 0.34580436796685 )**2;
end;
else _vqmvar + 0.01066937907316;
_vqnvar = 0.01066937907316 - _vqmvar;
if _vqnvar <= 1.2129679672812E-15 then do;
    SEGMENTO1_CIBER = .; Distance = .;
end;
else do;
    SEGMENTO1_CIBER = 1; Distance = CLUS10vads [1];
    _vqfzdst = Distance * 0.99999999999988; drop _vqfzdst;
    do _vqclus = 2 to 5;
        if CLUS10vads [_vqclus] < _vqfzdst then do;
            SEGMENTO1_CIBER = _vqclus; Distance = CLUS10vads [_vqclus];
            _vqfzdst = Distance * 0.99999999999988;
        end;
    end;
end;
end;
CLUS10vlex ;;
RUN;
```

```

*****
***          GRUPO TRADICIONAL (SEGMENTACION)
*****;

DATA PERFIL.JSG_BASE_UNIV_SEGMENTO_FIN_01;
SET PERFIL.JSG_BASE_UNIV_SEGMENTO_FIN_01;

*** Begin Class Look-up, Standardization, Replacement ;
drop _dm_bad; _dm_bad = 0;

*** Omitted Cases;
if _dm_bad then do;
    SEGMENTO2_TRAD = .; Distance = .;
    goto CLUS15vlex ;
end; *** omitted;

*** Compute Distances and Cluster Membership;
label SEGMENTO2_TRAD = 'Id de segmento' ;
label Distance = 'Distancia' ;
array CLUS15vads [8] _temporary_;
drop _vqclus _vqmvar _vqnvar;
_vqmvar = 0;
do _vqclus = 1 to 8; CLUS15vads [_vqclus] = 0; end;
if not missing( BAZAR ) then do;
    CLUS15vads [1] + ( BAZAR - 0.01373298958412 )**2;
    CLUS15vads [2] + ( BAZAR - 0.01921582474762 )**2;
    CLUS15vads [3] + ( BAZAR - 0.01598567554832 )**2;
    CLUS15vads [4] + ( BAZAR - 0.0318474193778 )**2;
    CLUS15vads [5] + ( BAZAR - 0.00574194288672 )**2;
    CLUS15vads [6] + ( BAZAR - 0.00429425016661 )**2;
    CLUS15vads [7] + ( BAZAR - 0.02239343102174 )**2;
    CLUS15vads [8] + ( BAZAR - 0.01279876977284 )**2;
end;
else _vqmvar + 0.0010449545576;
if not missing( CASIN ) then do;
    CLUS15vads [1] + ( CASIN - 0.00240380302989 )**2;
    CLUS15vads [2] + ( CASIN - 0.00809461186263 )**2;
    CLUS15vads [3] + ( CASIN - 0.00309332254171 )**2;
    CLUS15vads [4] + ( CASIN - 0.00340855447307 )**2;
    CLUS15vads [5] + ( CASIN - 0.00138609189974 )**2;
    CLUS15vads [6] + ( CASIN - 0.00458496266741 )**2;
    CLUS15vads [7] + ( CASIN - 0.01159523907566 )**2;
    CLUS15vads [8] + ( CASIN - 0.01158577528241 )**2;
end;
else _vqmvar + 0.00256942108261;
if not missing( DISPO ) then do;
    CLUS15vads [1] + ( DISPO - 0.0294429815691 )**2;
    CLUS15vads [2] + ( DISPO - 0.03892741341676 )**2;
    CLUS15vads [3] + ( DISPO - 0.01591564051893 )**2;
    CLUS15vads [4] + ( DISPO - 0.03264048063659 )**2;
    CLUS15vads [5] + ( DISPO - 0.05370880993972 )**2;
    CLUS15vads [6] + ( DISPO - 0.78968927513084 )**2;
    CLUS15vads [7] + ( DISPO - 0.02844962385353 )**2;
    CLUS15vads [8] + ( DISPO - 0.38711921754629 )**2;
end;
else _vqmvar + 0.03796936771276;
if not missing( EDUCA ) then do;

```

```

    CLUS15vads [1] + ( EDUCA - 0.00206684909323 )**2;
    CLUS15vads [2] + ( EDUCA - 0.0038529533704 )**2;
    CLUS15vads [3] + ( EDUCA - 0.00315492674571 )**2;
    CLUS15vads [4] + ( EDUCA - 0.0044824927687 )**2;
    CLUS15vads [5] + ( EDUCA - 0.00120213391015 )**2;
    CLUS15vads [6] + ( EDUCA - 0.00108041615494 )**2;
    CLUS15vads [7] + ( EDUCA - 0.00545912628527 )**2;
    CLUS15vads [8] + ( EDUCA - 0.00263226650329 )**2;
end;
else _vqmvar + 0.00021505486933;
if not missing( EMPRESARIAL ) then do;
    CLUS15vads [1] + ( EMPRESARIAL - 0.00454404457061 )**2;
    CLUS15vads [2] + ( EMPRESARIAL - 0.00712998537695 )**2;
    CLUS15vads [3] + ( EMPRESARIAL - 0.0071677581464 )**2;
    CLUS15vads [4] + ( EMPRESARIAL - 0.00927619682692 )**2;
    CLUS15vads [5] + ( EMPRESARIAL - 0.00276174356009 )**2;
    CLUS15vads [6] + ( EMPRESARIAL - 0.00177628218251 )**2;
    CLUS15vads [7] + ( EMPRESARIAL - 0.01010164614223 )**2;
    CLUS15vads [8] + ( EMPRESARIAL - 0.00488741308538 )**2;
end;
else _vqmvar + 0.00041492122455;
if not missing( ESTACION ) then do;
    CLUS15vads [1] + ( ESTACION - 0.02027292426929 )**2;
    CLUS15vads [2] + ( ESTACION - 0.03603361680332 )**2;
    CLUS15vads [3] + ( ESTACION - 0.02543833201576 )**2;
    CLUS15vads [4] + ( ESTACION - 0.02168392373586 )**2;
    CLUS15vads [5] + ( ESTACION - 0.0084958105558 )**2;
    CLUS15vads [6] + ( ESTACION - 0.01024260875093 )**2;
    CLUS15vads [7] + ( ESTACION - 0.12374065410223 )**2;
    CLUS15vads [8] + ( ESTACION - 0.03476220667544 )**2;
end;
else _vqmvar + 0.01085791467982;
if not missing( INTER ) then do;
    CLUS15vads [1] + ( INTER - 0.00802317834569 )**2;
    CLUS15vads [2] + ( INTER - 0.02379338780344 )**2;
    CLUS15vads [3] + ( INTER - 0.55008917929184 )**2;
    CLUS15vads [4] + ( INTER - 0.01879860174735 )**2;
    CLUS15vads [5] + ( INTER - 0.01184384624289 )**2;
    CLUS15vads [6] + ( INTER - 0.00438233689422 )**2;
    CLUS15vads [7] + ( INTER - 0.03948549712232 )**2;
    CLUS15vads [8] + ( INTER - 0.01187775962196 )**2;
end;
else _vqmvar + 0.02182539718558;
if not missing( LIBRE ) then do;
    CLUS15vads [1] + ( LIBRE - 0.00358672601032 )**2;
    CLUS15vads [2] + ( LIBRE - 0.00598702050737 )**2;
    CLUS15vads [3] + ( LIBRE - 0.0057644424932 )**2;
    CLUS15vads [4] + ( LIBRE - 0.00609553332632 )**2;
    CLUS15vads [5] + ( LIBRE - 0.0015508963196 )**2;
    CLUS15vads [6] + ( LIBRE - 0.00083247155465 )**2;
    CLUS15vads [7] + ( LIBRE - 0.00690711934502 )**2;
    CLUS15vads [8] + ( LIBRE - 0.00272437106815 )**2;
end;
else _vqmvar + 0.00030432601317;
if not missing( OTROS ) then do;
    CLUS15vads [1] + ( OTROS - 0.00022367235524 )**2;
    CLUS15vads [2] + ( OTROS - 0.00013259700468 )**2;
    CLUS15vads [3] + ( OTROS - 0.00026083238737 )**2;
    CLUS15vads [4] + ( OTROS - 0.00019011282794 )**2;
    CLUS15vads [5] + ( OTROS - 0.00004875049082 )**2;

```

```

        CLUS15vads [6] + ( OTROS - 0.00003422679944 )**2;
        CLUS15vads [7] + ( OTROS - 0.0001436643256 )**2;
        CLUS15vads [8] + ( OTROS - 0.00009603274599 )**2;
end;
else _vqmvar + 3.6982005002428E-6;
if not missing( PAGO_SERV ) then do;
    CLUS15vads [1] + ( PAGO_SERV - 0.01743973182505 )**2;
    CLUS15vads [2] + ( PAGO_SERV - 0.03221122429743 )**2;
    CLUS15vads [3] + ( PAGO_SERV - 0.02351467494671 )**2;
    CLUS15vads [4] + ( PAGO_SERV - 0.0276268412362 )**2;
    CLUS15vads [5] + ( PAGO_SERV - 0.01142776881068 )**2;
    CLUS15vads [6] + ( PAGO_SERV - 0.01162481667872 )**2;
    CLUS15vads [7] + ( PAGO_SERV - 0.04284248784168 )**2;
    CLUS15vads [8] + ( PAGO_SERV - 0.03043357252404 )**2;
end;
else _vqmvar + 0.00313673564154;
if not missing( RECREACION ) then do;
    CLUS15vads [1] + ( RECREACION - 0.00863771146559 )**2;
    CLUS15vads [2] + ( RECREACION - 0.01543256771663 )**2;
    CLUS15vads [3] + ( RECREACION - 0.01922796617502 )**2;
    CLUS15vads [4] + ( RECREACION - 0.01853832725178 )**2;
    CLUS15vads [5] + ( RECREACION - 0.00439017874632 )**2;
    CLUS15vads [6] + ( RECREACION - 0.00606075642031 )**2;
    CLUS15vads [7] + ( RECREACION - 0.03905084070263 )**2;
    CLUS15vads [8] + ( RECREACION - 0.01665110133601 )**2;
end;
else _vqmvar + 0.00178960703892;
if not missing( RESTA ) then do;
    CLUS15vads [1] + ( RESTA - 0.07184336981615 )**2;
    CLUS15vads [2] + ( RESTA - 0.08654622806163 )**2;
    CLUS15vads [3] + ( RESTA - 0.06646598859302 )**2;
    CLUS15vads [4] + ( RESTA - 0.10055919486534 )**2;
    CLUS15vads [5] + ( RESTA - 0.01811443289354 )**2;
    CLUS15vads [6] + ( RESTA - 0.02255224248253 )**2;
    CLUS15vads [7] + ( RESTA - 0.21858380549305 )**2;
    CLUS15vads [8] + ( RESTA - 0.07636751408834 )**2;
end;
else _vqmvar + 0.01327274106595;
if not missing( SALUD ) then do;
    CLUS15vads [1] + ( SALUD - 0.09503025274253 )**2;
    CLUS15vads [2] + ( SALUD - 0.37497137248114 )**2;
    CLUS15vads [3] + ( SALUD - 0.0834984787997 )**2;
    CLUS15vads [4] + ( SALUD - 0.10305809020451 )**2;
    CLUS15vads [5] + ( SALUD - 0.77074283315835 )**2;
    CLUS15vads [6] + ( SALUD - 0.0474793912078 )**2;
    CLUS15vads [7] + ( SALUD - 0.09427784652295 )**2;
    CLUS15vads [8] + ( SALUD - 0.14946250023177 )**2;
end;
else _vqmvar + 0.0517781202396;
if not missing( SERVICIO ) then do;
    CLUS15vads [1] + ( SERVICIO - 0.01423551762249 )**2;
    CLUS15vads [2] + ( SERVICIO - 0.03460674232251 )**2;
    CLUS15vads [3] + ( SERVICIO - 0.02121493446012 )**2;
    CLUS15vads [4] + ( SERVICIO - 0.0283243662462 )**2;
    CLUS15vads [5] + ( SERVICIO - 0.01293162131791 )**2;
    CLUS15vads [6] + ( SERVICIO - 0.00772319392799 )**2;
    CLUS15vads [7] + ( SERVICIO - 0.02831666292649 )**2;
    CLUS15vads [8] + ( SERVICIO - 0.02349900703657 )**2;
end;
else _vqmvar + 0.0030487312553;

```

```

if not missing( SUPER ) then do;
  CLUS15vads [1] + ( SUPER - 0.61045372690133 )**2;
  CLUS15vads [2] + ( SUPER - 0.17480482392609 )**2;
  CLUS15vads [3] + ( SUPER - 0.06931632538934 )**2;
  CLUS15vads [4] + ( SUPER - 0.17944763928892 )**2;
  CLUS15vads [5] + ( SUPER - 0.04581864094783 )**2;
  CLUS15vads [6] + ( SUPER - 0.04809659061665 )**2;
  CLUS15vads [7] + ( SUPER - 0.19515044646702 )**2;
  CLUS15vads [8] + ( SUPER - 0.12826461112424 )**2;
end;
else _vqmvar + 0.04275590058255;
if not missing( TIENDA ) then do;
  CLUS15vads [1] + ( TIENDA - 0.09048473527532 )**2;
  CLUS15vads [2] + ( TIENDA - 0.12126326806606 )**2;
  CLUS15vads [3] + ( TIENDA - 0.06920885705289 )**2;
  CLUS15vads [4] + ( TIENDA - 0.39477852967171 )**2;
  CLUS15vads [5] + ( TIENDA - 0.04183600023129 )**2;
  CLUS15vads [6] + ( TIENDA - 0.03305665464618 )**2;
  CLUS15vads [7] + ( TIENDA - 0.11044708453538 )**2;
  CLUS15vads [8] + ( TIENDA - 0.09071517426118 )**2;
end;
else _vqmvar + 0.01859792117289;
if not missing( VIAJERO ) then do;
  CLUS15vads [1] + ( VIAJERO - 0.00758088873093 )**2;
  CLUS15vads [2] + ( VIAJERO - 0.01700931523376 )**2;
  CLUS15vads [3] + ( VIAJERO - 0.02091195183672 )**2;
  CLUS15vads [4] + ( VIAJERO - 0.01925498502293 )**2;
  CLUS15vads [5] + ( VIAJERO - 0.00800523559235 )**2;
  CLUS15vads [6] + ( VIAJERO - 0.00649095287762 )**2;
  CLUS15vads [7] + ( VIAJERO - 0.02308005882876 )**2;
  CLUS15vads [8] + ( VIAJERO - 0.01612514985695 )**2;
end;
else _vqmvar + 0.00148084471616;
_vqnvar = 0.21106565723891 - _vqmvar;
if _vqnvar <= 4.0792158109315E-13 then do;
  SEGMENTO2_TRAD = .; Distance = .;
end;
else do;
  SEGMENTO2_TRAD = 1; Distance = CLUS15vads [1];
  _vqfzdst = Distance * 0.99999999999988; drop _vqfzdst;
  do _vqclus = 2 to 8;
    if CLUS15vads [_vqclus] < _vqfzdst then do;
      SEGMENTO2_TRAD= _vqclus; Distance = CLUS15vads [_vqclus];
      _vqfzdst = Distance * 0.99999999999988;
    end;
  end;
  Distance = sqrt(Distance * (0.21106565723891 / _vqnvar));
end;
CLUS15vlex ;;
RUN;

*****
***          BASE FINAL SEGMENTACION
*****

DATA PERFIL.JSG_BASE_UNIV_SEGMENTO_FIN;
SET PERFIL.JSG_BASE_UNIV_SEGMENTO_FIN_01;
DROP Distance;
RUN;

```