

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS

DEPARTAMENTO DE MATEMÁTICAS

PROYECTO DE GRADUACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

“MAGÍSTER EN SEGUROS Y RIESGOS FINANCIEROS”

TEMA

“DETERMINACIÓN DEL PERFIL DE COMPORTAMIENTO DE CLIENTES DE
UNA INSTITUCIÓN FINANCIERA MEDIANTE TÉCNICAS ESTADÍSTICAS
MULTIVARIANTES”

AUTOR:

JESSENIA ELEANA HUACHISACA VERA

Guayaquil – Ecuador

2016

DEDICATORIA

A mis padres, y hermanos de quienes siempre he tenido, tengo y tendré apoyo incondicional y motivación para el cumplimiento de mis objetivos personales y profesionales.

A mis amigos, quienes desinteresadamente aportaron con su granito de arena, sea con conocimiento, momentos o palabras de aliento para la culminación de esta nueva meta.

AGRADECIMIENTO

A Dios, por guiarme en este proceso, por iluminar mi camino y por darme las fuerzas para cumplir mis sueños enseñándome que sus tiempos son perfectos.

A mis padres y hermanos por ser mi apoyo, y mi inspiración a siempre ser mejor, por confiar, creer y tener fe en mí.

A mi directora de tesis que con gran paciencia siempre me guio de la mejor manera.

A mis amigos, familiares y compañeros que estuvieron a mi lado en el trayecto para el cumplimiento de ésta meta.

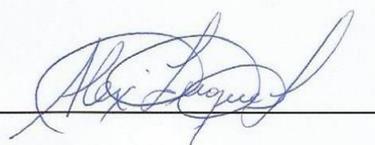
TRIBUNAL DE GRADUACIÓN



M.Sc. Marlon Manyá Orellana
PRESIDENTE DEL TRIBUNAL



Sandra García Bustos, P.hD.
DIRECTOR DE PROYECTO



MPC. Alex Luque Letechi
VOCAL DEL TRIBUNAL

FIRMA DE AUTOR



Ing. Jessenia Huachisaca Vera

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Graduación, me (nos) corresponde(n) exclusivamente; el patrimonio intelectual del mismo, corresponde exclusivamente a la **Facultad de Ciencias Naturales y Matemáticas, Departamento de Matemáticas** de la Escuela Superior Politécnica del Litoral.



Ing. Jessenia Huachisaca Vera

RESUMEN EJECUTIVO

El presente proyecto pretende otorgarle a la institución financiera un perfil de comportamiento de clientes guía, con el cual les permita conocer las condiciones de riesgo que representa cada cliente de acuerdo a sus características propias y habituales.

Los factores de riesgos identificados serán los parámetros que permitirán evaluar las características de cada cliente de acuerdo a los productos o servicios ofrecidos por la institución, la ubicación geográfica desde el punto de vista de la nacionalidad como del lugar donde se encuentre generando transacciones.

Dicho procedimiento está enfocado en el inicio de las relaciones contractuales del cliente con la institución financiera, es decir la aceptación de los mismos por parte de la institución en concordancia con las actividades que realizan, el origen de los recursos financieros y el patrimonio que administran.

Posteriormente, se analizarán las variables demográficas y socioeconómicas mediante técnicas multivariadas para conocer la incidencia que dichas variables tendrán en la construcción del perfil.

Finalmente, se seleccionará el perfil que mejor segmentación nos dé con el fin de prevenir y mitigar los riesgos que pueden representar los clientes o posibles clientes para la institución financiera.

CONTENIDO

DEDICATORIA	<i>i</i>
AGRADECIMIENTO	<i>ii</i>
TRIBUNAL DE GRADUACIÓN	<i>iii</i>
FIRMA DE AUTOR	<i>iv</i>
DECLARACIÓN EXPRESA	<i>v</i>
RESUMEN EJECUTIVO	<i>vi</i>
ÍNDICE DE TABLAS	<i>ix</i>
ÍNDICE DE ILUSTRACIONES	<i>x</i>
CAPÍTULO I	<i>1</i>
INTRODUCCIÓN	<i>1</i>
1.1 Delimitación del problema	<i>1</i>
1.2 Formulación del problema	<i>2</i>
1.3 Justificación	<i>2</i>
1.4 Objetivo General	<i>3</i>
1.5 Objetivos Específicos	<i>3</i>
1.6 Alcance	<i>3</i>
CAPÍTULO II	<i>4</i>
MARCO TEÓRICO	<i>4</i>
2.1 Segmentación del Mercado Financiero	<i>4</i>
2.1.1 Definiciones de segmentación de mercado	<i>5</i>
2.2 Criterios o factores de Segmentación	<i>5</i>
2.2.1 Criterios o factores demográficos y socioeconómicos	<i>7</i>
2.2.2 Criterios o factores geográficos	<i>7</i>
2.2.3 Criterios o factores mercadológicos y conductuales	<i>7</i>
2.3 Técnicas estadísticas para segmentar	<i>8</i>
2.3.1 Análisis de contingencia	<i>9</i>

2.3.2	Regresión Logística.....	10
2.3.3	Árbol de Decisión	12
<i>CAPÍTULO III.....</i>		13
<i>ASPECTOS METODOLÓGICOS DE LOS DATOS.....</i>		13
2.1	Variables que se usarán en el análisis.....	13
<i>CAPÍTULO IV.....</i>		33
<i>RESULTADOS.....</i>		33
4.1	Análisis de Contingencia	33
4.2	Regresión Logística	41
4.3	Árboles de Decisión	47
<i>CONCLUSIONES.....</i>		55
<i>RECOMENDACIONES.....</i>		57
<i>BIBLIOGRAFÍA</i>		58

ÍNDICE DE TABLAS

TABLA 1 FRECUENCIA PRODUCTO SERVICIO	14
TABLA 2 ESTADÍSTICO DE VARIABLE INGRESOS	15
TABLA 3 FRECUENCIA INGRESOS (AGRUPADOS).....	16
TABLA 4 ACTIVIDADES ECONÓMICAS POR RIESGO	25
TABLA 5 FRECUENCIA DE ACTIVIDAD ECONÓMICA.....	25
TABLA 6 FRECUENCIA DE NACIONALIDADES	27
TABLA 7 FRECUENCIA DE SEXO	29
TABLA 8 ESTADÍSTICO DE LA VARIABLE PATRIMONIO.....	31
TABLA 9 FRECUENCIA DE PATRIMONIO (AGRUPADO).....	32
TABLA 10 TABLA DE CONTINGENCIA INGRESOS - PRODUCTOS/SERVICIOS.....	34
TABLA 11 PRUEBA CHI-CUADRADO INGRESO - PRODUCTO/SERVICIO	34
TABLA 12 TABLA DE CONTINGENCIA INGRESO - ACTIVIDAD ECONÓMICA.....	35
TABLA 13 PRUEBA CHI-CUADRADO INGRESO - ACTIVIDAD ECONÓMICA	36
TABLA 14 TABLA DE CONTINGENCIA INGRESOS - NACIONALIDAD	37
TABLA 15 PRUEBA CHI-CUADRADO INGRESO - NACIONALIDAD	37
TABLA 16 TABLA DE CONTINGENCIA INGRESO - PATRIMONIO.....	38
TABLA 17 PRUEBA CHI-CUADRADO INGRESO - PATRIMONIO	39
TABLA 18 TABLA DE CONTINGENCIA INGRESO - SEXO	39
TABLA 19 PRUEBA CHI-CUADRADO INGRESO - SEXO	40
TABLA 20 CODIFICACIÓN DE VARIABLE PRODUCTO/SERVICIO.....	42
TABLA 21 CODIFICACIÓN VARIABLES INDEPENDIENTES	42
TABLA 22 TABLA DE CLASIFICACIÓN	43
TABLA 23 VARIABLES EN LA ECUACIÓN - MODELO 1	43
TABLA 24 VARIABLES INCLUIDAS EN EL MODELO	44
TABLA 25 PRUEBA DE HOSMER Y LEMESHOW	45
TABLA 26 TABLA DE OBSERVADOS VS ESPERADOS	46
TABLA 27 TABLA DE CLASIFICACIÓN	46
TABLA 28 RESUMEN DEL MODELO - ÁRBOL DE DECISIÓN	47
TABLA 29 TABLA DE CLASIFICACIÓN CHAID – MODELO 4.....	48
TABLA 30 TABLA DE RIESGO - MODELO 4.....	48

ÍNDICE DE ILUSTRACIONES

ILUSTRACIÓN 1 CRITERIOS SEGMENTACIÓN	6
ILUSTRACIÓN 2 HISTOGRAMA DE PRODUCTOS SERVICIOS	14
ILUSTRACIÓN 3 HISTOGRAMA INGRESOS (AGRUPADOS)	16
ILUSTRACIÓN 4 HISTOGRAMA DE ACTIVIDAD ECONÓMICA	26
ILUSTRACIÓN 6 HISTOGRAMA DE NACIONALIDADES	28
ILUSTRACIÓN 7 HISTOGRAMA DE SEXO	30
ILUSTRACIÓN 8 HISTOGRAMA DE PATRIMONIO (AGRUPADO)	32
ILUSTRACIÓN 9 ÁRBOL DE DECISIÓN - NODO 0	49
ILUSTRACIÓN 10 ÁRBOL DE DECISIÓN – NODO 1	50
ILUSTRACIÓN 11 ÁRBOL DE DECISIÓN - NODO 4	50
ILUSTRACIÓN 12 ÁRBOL DE DECISIÓN - NODO 5	51
ILUSTRACIÓN 13 ÁRBOL DE DECISIÓN - NODO 2	52
ILUSTRACIÓN 14 ÁRBOL DE DECISIÓN - NODO 6	52
ILUSTRACIÓN 15 ÁRBOL DE DECISIÓN - NODO 7	53
ILUSTRACIÓN 16 ÁRBOL DE DECISIÓN - NODO 8	53
ILUSTRACIÓN 17 ÁRBOL DE DECISIÓN - NODO 9	54

CAPÍTULO I

INTRODUCCIÓN

Las instituciones financieras por años han gestionado el riesgo de los clientes mediante el conocimiento que tengan sobre ellos antes de su aceptación y durante su transaccionalidad en la institución.

Si bien los perfiles de clientes son parte de los controles que permiten determinar operaciones sospechosas mediante las transacciones que realizan, establecer un perfil de comportamiento corresponde a determinar criterios de riesgos mediante el análisis de variables demográficas y socioeconómicas y que puedan resultar sospechosas o inusuales.

Basados en la intervención del ente regulador en el sistema financiero con la finalidad de proteger y salvaguardar los intereses del público en general, ha establecido la normativa de obligatorio cumplimiento con los parámetros mínimos que se debe considerar para establecer el perfil de comportamiento.

En este sentido, la definición de un perfil de comportamiento mediante el uso de técnicas multivariadas permitirá reducir la subjetividad en los parámetros a considerar para perfilar los clientes mediante la construcción de una herramienta útil.

1.1 DELIMITACIÓN DEL PROBLEMA

En la actualidad, las instituciones financieras están llamadas a participar activamente en la aplicación de procesos preventivos mediante la recopilación, confirmación y actualización de información del cliente para conocer su perfil de riesgo previo a establecer una relación comercial con la institución.

Por ésta razón se levantará el perfil de comportamiento de los clientes ya aceptados por la institución financiera con la finalidad de categorizarlos y establecer los parámetros que servirán para gestionar los riesgos de los futuros clientes.

1.2 FORMULACIÓN DEL PROBLEMA

- ¿Tienen las variables demográficas y socioeconómicas de una persona influencia en el perfil de comportamiento para aceptar o no su ingreso en la institución financiera?
- ¿Cuáles son las condiciones de riesgo que representa cada cliente de acuerdo a sus características propias y habituales frente a los productos y servicios que ofrece la institución financiera?

1.3 JUSTIFICACIÓN

El giro de cada institución financiera se basa en los productos y servicios que ofrece y el perfil de cada posible cliente, esto hace necesaria la determinación de los parámetros propios de la institución que permita categorizar los factores de riesgo y el perfil de comportamiento de sus clientes.

En este proyecto se usarán técnicas estadísticas multivariadas para diseñar un perfil que la institución financiera podrá usar para detectar clientes inusuales.

Los datos para el análisis de la información se procesarán mediante el software SPSS.

1.4 OBJETIVO GENERAL

Determinar el perfil de comportamiento de los clientes de la institución financiera de acuerdo a los productos y servicios, actividad económica y ubicación geográfica, mediante técnicas multivariadas que responden a su histórico de clientes.

1.5 OBJETIVOS ESPECÍFICOS

1. Analizar la influencia de los factores socioeconómicos y demográficos en el perfil de comportamiento de clientes.
2. Determinar el modelo que mejor ajuste y que describa la relación entre las variables de respuesta y las variables explicativas.

1.6 ALCANCE

Para la determinación del perfil de comportamiento utilizaremos la información de la base de datos de clientes de la Institución Financiera, siendo esto los datos declarados por el cliente al inicio de la relación comercial.

Dentro de las variables a considerar tenemos:

- ✓ Ubicación geográfica
- ✓ Nacionalidad
- ✓ Actividad que desempeña
- ✓ Patrimonio
- ✓ Sueldo / Ingreso
- ✓ Producto / Servicio

CAPÍTULO II

MARCO TEÓRICO

2.1 SEGMENTACIÓN DEL MERCADO FINANCIERO

Para la adecuada aplicación de los requerimientos en el ámbito de gestión de riesgos en lo que el documento se refiere es importante conocer cómo se ha manejado y se maneja el sector financiero en el ofrecimiento de sus productos y servicios.

El entorno financiero ha sufrido grandes transformaciones a través del tiempo, el acelerado cambio tecnológico que ha ido propiciando la automatización de información, el alto grado de exigencia rentable, confiable y de mejores servicios por parte de los clientes, la aparición de productos sustitutos, competidores no financieros nacionales e internacionales, las restricciones legales y normativas de control existentes, constituyen los factores determinantes en los cambios que se han dado en la competitividad del sector.

Actualmente, este sector exige que sean instituciones integradas, con procesos que les permitan diversificar sus unidades de negocio y sus carteras, con tecnología adecuada para las innovaciones que vayan teniendo, y creativas en sus estrategias de marketing de acuerdo a los nuevos sistemas de distribución de sus productos y servicios.

Ante ésta realidad, las instituciones financieras deben considerar la adaptación de sus estructuras organizativas a una cultura corporativa acorde a la competitividad y regulación del sector, como segmentar sus clientes en operaciones de economía domésticas o particulares y las operaciones de las empresas.

2.1.1 Definiciones de segmentación de mercado

No existe una única forma de segmentar mercados, por eso las ventajas y limitaciones en función del producto – mercado que se considere dependerá de los objetivos que dicha segmentación persiga (Grover & Srinivasan, 1987).

Desde la perspectiva del marketing, el concepto de segmentación responde a "la subdivisión en grupos o segmentos de clientes de características homogéneas que hacen que los esfuerzos de marketing sean similares durante un determinado periodo de tiempo, para que puedan considerarse como mercado objetivo de una combinación distinta de marketing mix"

Para determinar los grupos de clientes homogéneos es necesaria la utilización de técnicas estadísticas, por lo que varios autores definen la segmentación de mercado como "la división del mercado a través de diferentes procedimientos estadísticos, en una serie de subconjuntos o segmentos homogéneos, en función de diferentes variables o características que se toman como referencia para explicar un determinado fenómeno" (Salgado Carrión, 1981)

Varios autores consideran la segmentación como estrategia de marketing (Frank, Massy, & Wind, 1972), otros que lo definen como estrategia comercial lo consideran como una técnica de investigación dirigida a identificar segmentos de clientes con objeto de priorizar las asignaciones de los recursos (Mahajan & Jain, 1978).

En base a la normativa ecuatoriana y el objetivo que persigue el presente trabajo se conceptualizará la segmentación como "el resultado para definir, identificar, clasificar y analizar adecuadamente los grupos de los clientes, en función de sus características y criterios de riesgos adoptados" (Resolución JB-2013-2535, 2013 (Superintendencia de Bancos y Seguros, 2013)

2.2 CRITERIOS O FACTORES DE SEGMENTACIÓN

En la segmentación de mercado se debe considerar las variables que se utilizarán ya que el objetivo es lograr dividir el mercado en grupos homogéneos

o heterogéneos entre sí, logrando que los clientes se parezcan más entre ellos conforme a sus características.

Como nuestro objetivo es obtener un perfil de comportamiento de clientes, y existe una relación directa entre los criterios y características que definen la segmentación de mercado de consumidores, clasificaremos las principales variables de acuerdo al diseño metodológico sugerido por (Kotler, 1988), y complementarlo con los factores y subfactores que constan en los procedimientos emitidos por la Superintendencia de Bancos:

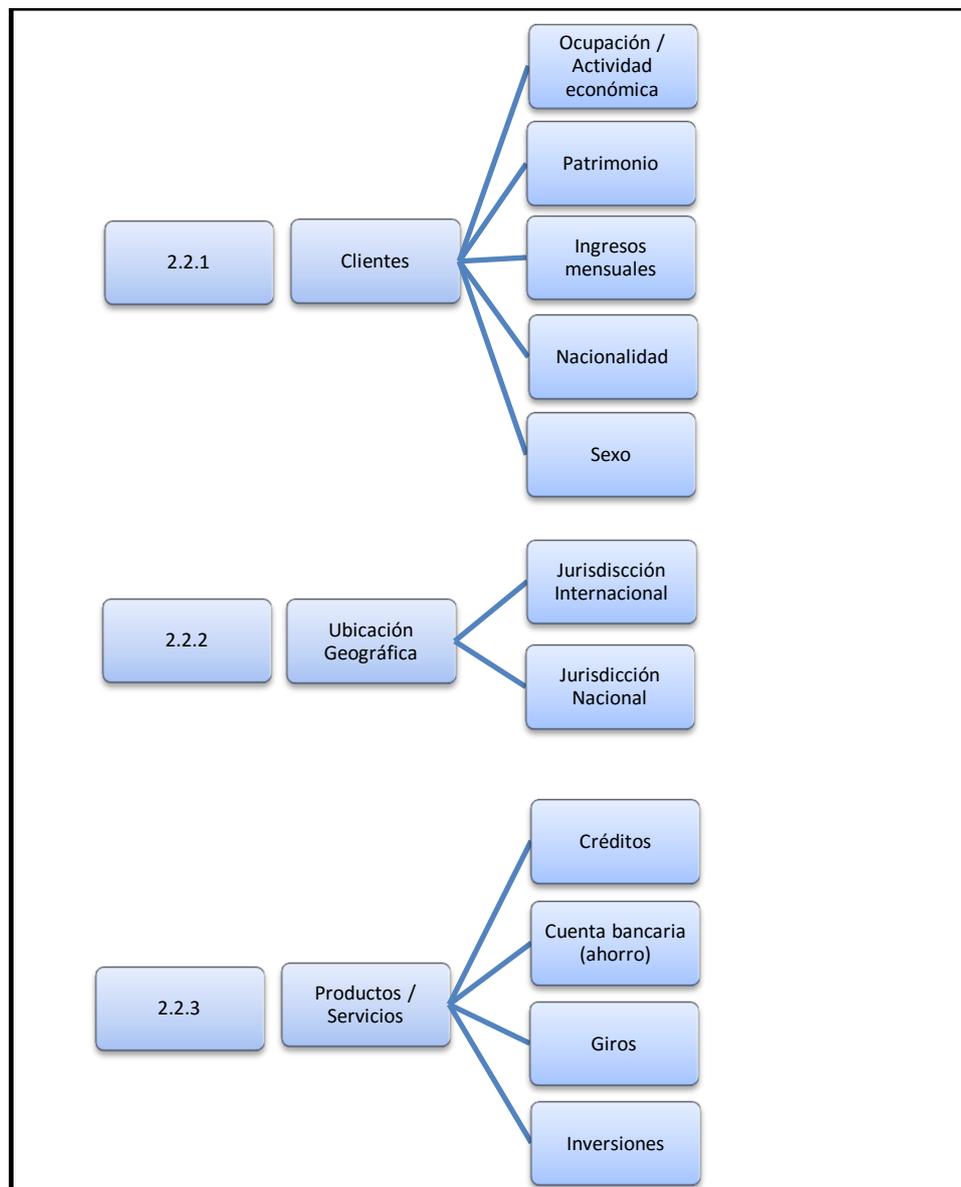


Ilustración 1 Criterios Segmentación

Fuente: Elaboración propia.

2.2.1 Criterios o factores demográficos y socioeconómicos

Los subfactores que contiene esta categoría representan las variables más fáciles de examinar y medir en un grupo de clientes y las que permitirán la diferenciación entre los segmentos que se establezcan.

Basados en la hipótesis que las características socioeconómicas de una persona generan diferencia entre los clientes, para nuestro análisis definiremos y consideraremos las siguientes variables:

- Actividad económica
- Nacionalidad
- Ubicación geográfica
- Productos o servicios
- Sexo

2.2.2 Criterios o factores geográficos

La ubicación geográfica de los clientes o futuros clientes responde a otro criterio de segmentación ya que estos pueden influir por razones económicas, culturales, entre otras.

Al tratarse de una institución financiera las operaciones se desarrollan, a nivel nacional e internacional porque la institución en análisis posee varias agencias a nivel nacional y corresponsales en diferentes países, ésta variable contendrá información de la jurisdicción del cliente.

2.2.3 Criterios o factores mercadológicos y conductuales

Estos criterios hacen referencia a diversos factores relacionados a la compra o consumo de los productos o servicios, como por ejemplo: tipo de usuario, volumen de consumo, canales de acceso, etc.

La variable que se utilizará en éste segmento corresponden a los productos o servicios que mantienen en su portafolio, y se los establecerá como:

- Giros o remesas
- Cuenta bancaria (ahorro por depósito de ingresos)
- Créditos
- Inversiones

La segmentación del producto se fundamenta en las diferencias que los individuos perciben entre los distintos productos que están a su disposición (Barnett, 1969).

La independencia de los criterios a los que se recurra en la segmentación, la evolución de las necesidades y las condiciones de los individuos a lo largo de los años establece cambios en cortos períodos de tiempo, por lo tanto se debe actualizar el perfil de comportamiento constantemente.

En un estudio de segmentación realizado en el mercado minorista bancario por (Calantone & Sawyer, 1978) se observó que sólo el 28,8% de los individuos de la muestra permanecían en el mismo segmento de mercado durante dos años consecutivos.

(Pride & Ferrell, 1982) Consideran que la selección de las variables de segmentación representa una fase crítica de la división de cualquier mercado. La elección de una variable inadecuada limita las oportunidades de elaborar una estrategia exitosa. Por ello, dichos autores establecen la necesidad de que los criterios utilizados se relacionen con los usos y necesidades del consumidor o con su comportamiento hacia el producto.

2.3 TÉCNICAS ESTADÍSTICAS PARA SEGMENTAR

Con éste antecedente, empezaremos a definir la parte estadística y el análisis de datos.

El análisis multivariante estudia, analiza, representa e interpreta los datos que resultan de observar más de una variable estadística sobre una muestra de individuos. Las variables observables son homogéneas y correlacionadas, sin que alguna predomine sobre las demás. (Cuadras, 2014).

Sólo a través de éstas técnicas se logra examinar adecuadamente las relaciones múltiples de las variables para llegar a una comprensión de la toma de decisiones más completa y realista (Hair, Anderson, Tatham, & Black, 1999).

Considerar la naturaleza de las variables dependientes e independientes y el objetivo que persiguen, es una forma de clasificarlas. Las variables dependientes constituyen el núcleo central de la investigación, y las variables independientes son las que se utilizan para explicar el comportamiento de las variables dependientes (Uriel, 1995).

Entre las técnicas más conocidas y cuya aplicación se considerará en el presente trabajo, describimos las siguientes (Uriel, 1995).

2.3.1 Análisis de contingencia

El análisis de contingencia nos permite conocer la relación de dos variables entre sí mediante la elaboración de tablas de contingencia, en donde una ocupa las filas y otra las columnas.

La intersección entre una fila y una columna da lugar a una celda o casilla, cuya frecuencia observada es n_{ij} .

Se contrasta la hipótesis nula que presupone la independencia entre ambas variables, mediante el estadístico χ^2 de Pearson.

H0: Las variables son independientes

H1: Existe relación de dependencia entre las variables

El estadístico observado se define como:

$$\sum_{i=1}^k \sum_{j=1}^m \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \chi_{(k-1)(m-1)}^2$$

Si la hipótesis nula H_0 es verdadera, seguirá una distribución χ^2 con $(k - 1)(m - 1)$ grados de libertad y $e_{ij} > 5 / 1 \leq i \leq k; 1 \leq j \leq m$.

2.3.2 Regresión Logística

El modelo de regresión logística es la estimación de la probabilidad de un suceso o categoría que dependerán de otras covariables, es decir es un modelo de predicción de probabilidad de ocurrencia de una variable categórica y dicotómica.

Es una técnica estadística que se usa para modelar resultados binarios y que toma los valores de 0 y 1, Si – No, Verdadero - Falso.

Partiendo del modelo de regresión clásico

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e$$

Donde, β_0 es el término constante, $\beta_i; i = 1, 2, \dots, n$ son los parámetros asociados al modelo y e es el error.

Para la aplicación de este modelo debe cumplir el supuesto de linealidad, por lo tanto es necesario realizar una transformación logarítmica (Logit), ya que al ser la variable dependiente binaria se podrían presentar ciertos inconvenientes:

- Obtener predicciones fuera de intervalo (0,1)
- e no cumpla hipótesis de homocedasticidad
- No se puede esperar un buen ajuste cuando la variable de respuesta es binaria.

Para lo cual se utilizará el modelo logit transformando la variable respuesta a partir de una función f estrictamente creciente:

$$Y = f(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n) + e$$

Donde f es la función logística, i.e.:

$$f(z) = \exp(z) / (1 + \exp(z))$$

Por tanto, tendremos que:

$$E[Y] = P(Y = 1) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}$$

El método de máxima verosimilitud es usado para la estimación de los parámetros del modelo *logit*, y para interpretar los coeficientes beta, es necesario fijarse en los signos de los estimadores:

- Si el signo del estimador es positivo, entonces la variable explicativa asociada a este coeficiente crece, por lo tanto la variable respuesta $P(Y = 1)$ también crece.
- Si el signo del estimador es negativo, entonces cuando la variable explicativa asociada a este coeficiente crece, la variable respuesta disminuye.

Para la interpretación de los estimadores, también se utilizará el cociente de probabilidades odds, definido de la siguiente manera:

$$Odds = \frac{P(Y = 1)}{1 - P(Y = 1)}$$

$$Odds = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)$$

Tomando logaritmos en la expresión anterior, obtenemos una expresión lineal para el modelo:

$$Logit[P(Y = 1)] = \ln\left(\frac{P(Y = 1)}{1 - P(Y = 1)}\right)$$

$$Logit[P(Y = 1)] = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

La siguiente relación representa el cambio en odds cuando hay un incremento unitario en una de las variables explicativas, manteniendo las demás constantes, (el obtenido tras realizar el incremento y el anterior al mismo); dicha variación está dada por esta expresión:

$$Odds\ ratio = \frac{odds2}{odds1} = \exp(\beta_i)$$

2.3.3 Árbol de Decisión

Los árboles de decisión, conocidos como algoritmos de segmentación CHAID, consisten en un algoritmo de árbol estadístico y multidireccional que estudia los datos y genera subconjuntos precisos y homogéneos con respecto a la variable dependiente.

Se elige la variable independiente (predictora o endógena) que presenta la relación más fuerte con las variables dependientes.

Los árboles de decisión son técnicas utilizadas para la predicción estratificación o segmentación de datos y filtrado de variables. La metodología CHAID realiza la segmentación en base al estadístico χ^2 , desarrollándose en su mayoría con el enfoque de Pearson:

$$\chi^2 = \sum \sum \frac{(f_{ij} - f_{ij}^2)^2}{(f_{ij}^2)}$$

Donde f_{ij} es la frecuencia observada f_{ij}^2 y es la frecuencia esperada si hay independencia.

CAPÍTULO III

ASPECTOS METODOLÓGICOS DE LOS DATOS

Para realizar la segmentación de clientes y posterior análisis de las características y comportamiento de los subsegmentos identificados, utilizaremos la base de datos de los 29.520 clientes de la institución financiera la cual se recoge al inicio de la relación cliente – institución.

Dicha base contiene los clientes aceptados tanto en agencias nacionales como en los corresponsales internacionales.

A continuación se detallan las variables que serán utilizadas en los modelos a considerar. El número de variables a utilizar en el análisis corresponde a un total de 6.

3.1 VARIABLES QUE SE USARÁN EN EL ANÁLISIS

➤ **Productos o servicios**, que son mecanismos o instrumentos financieros que de conformidad con la ley, oferta la institución financiera y cuya codificación responde a:

1 = Cuentas Bancarias

2 = Remesas

3 = Créditos

4 = Inversiones

Al analizar los datos de los productos / servicios que ofrece la institución financiera, tenemos que el 60% accede a la apertura de cuenta sea ésta de

ahorro o corriente, el 39,4% es usada para el envío y recepción de giros o remesas de familiares, el 0,3% solicita créditos y el 0,4% establece algún tipo de inversión, el resumen se muestra en la Tabla 1 y la Ilustración 2.

		PRODUCTO			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	CUENTA BANCARIA	17703	60,0	60,0	60,0
	REMESAS	11620	39,4	39,4	99,3
	CREDITO	76	,3	,3	99,6
	INVERSION	121	,4	,4	100,0
	Total	29520	100,0	100,0	

Tabla 1 Frecuencia Producto Servicio

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

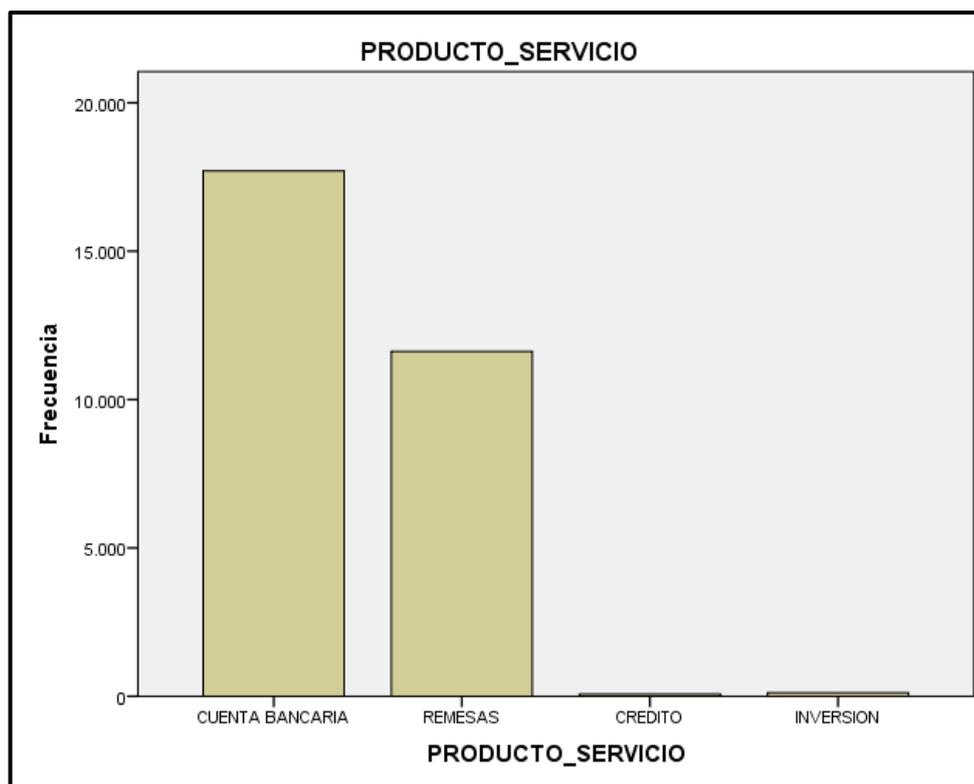


Ilustración 2 Histograma de Productos Servicios

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

No obstante por el mínimo porcentaje que representan los créditos y las inversiones se considerarán para los análisis respectivos únicamente:

1 = Cuentas Bancarias

2 = Remesas

- **Ingresos mensuales** son la fuente principal de ingresos para determinar la capacidad adquisitiva de cada cliente.

De acuerdo a los datos analizados e ilustrados en la Tabla 2 estadísticos de variable, podemos especificar que en la variable “Ingresos”, se obtuvo que la media de los valores que los clientes perciben es de \$ 2.156,21, el ingreso con mayor frecuencia absoluta, lo vemos reflejado en la moda, siendo éste de \$ 600.

Estadísticos		
INGRESOS		
N	Válido	29140
	Perdidos	380
Media		2156,2171
Error estándar de la media		107,92203
Mediana		800,0000
Moda		600,00
Mínimo		10,00
Máximo		1891553,00
Percentiles	25	400,0000
	50	800,0000
	75	1600,0000

Tabla 2 Estadístico de Variable Ingresos

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Para el histograma de la variable “Ingresos” (Ilustración 3), se realizó la agrupación de los valores como se muestra en la Tabla 3 de frecuencia, los resultados de la nueva variable”.

		INGRESOS (Agrupada)			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	< 500	10278	34,8	35,3	35,3
	501 - 1500	11486	38,9	39,4	74,7
	1501 - 2500	3344	11,3	11,5	86,2
	>2501	4032	13,7	13,8	100,0
	Total	29140	98,7	100,0	
Perdidos	Sistema	380	1,3		
Total		29520	100,0		

Tabla 3 Frecuencia Ingresos (Agrupados)

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

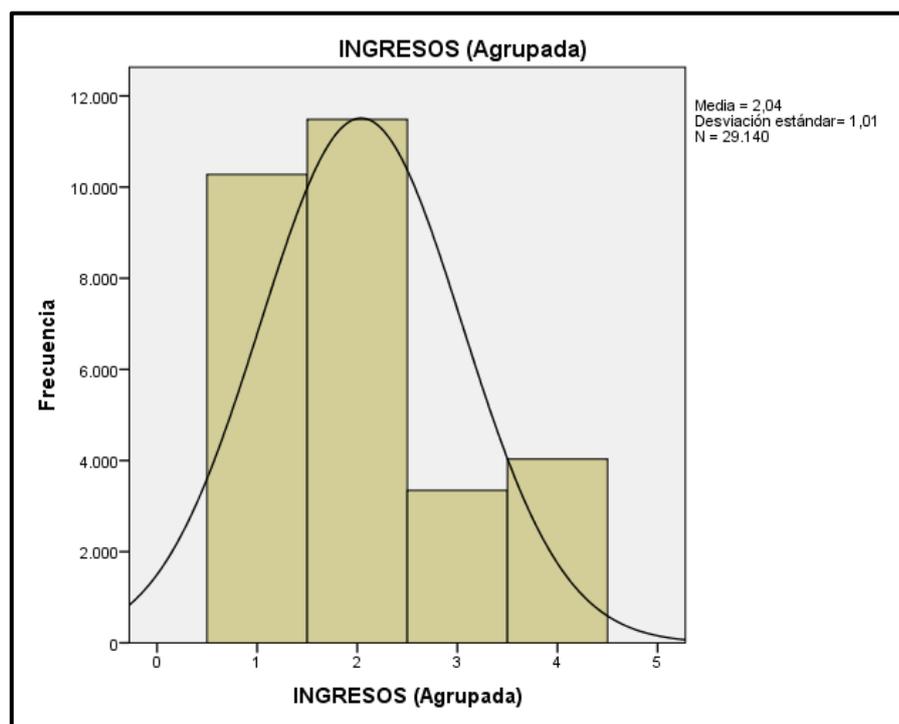


Ilustración 3 Histograma Ingresos (Agrupados)

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

- **Actividad económica.-** labor que habitualmente desempeña el cliente al inicio de la relación comercial y cuya codificación responde a:

1 = Riesgo Bajo

2 = Riesgo Medio

3 = Riesgo Alto

Dicha categorización ha sido dada en base a la siguiente Tabla 4, cuya agrupación a sido segmentada en categorías de riesgo que podrían representar para una institución financiera acorde a las 40 Recomendaciones dadas por el Grupo de Acción Financiera Internacional. (Grupo Acción Financiera Internacional, 2012)

Nivel de Riesgo.- Categoría en el que la institución del sistema financiero ubica a un cliente por el riesgo que representa, entendiéndose como riesgo la probabilidad de ocurrencia e impacto de un evento, en base a los parámetros que permiten evaluar las circunstancias y características particulares de cada actividad económica.

NIVEL	CÓDIGO	DESCRIPCIÓN	NIVEL DE RIESGO 1/2/3
1	A	AGRICULTURA, GANADERÍA, SILVICULTURA Y PESCA.	
3	A011	CULTIVO DE PLANTAS NO PERENNES.	2
3	A012	CULTIVO DE PLANTAS PERENNES.	2
3	A013	PROPAGACIÓN DE PLANTAS.	1
3	A014	GANADERÍA.	1
3	A015	CULTIVO DE PRODUCTOS AGRÍCOLAS EN COMBINACIÓN CON LA CRÍA DE ANIMALES (EXPLOTACIÓN MIXTA).	1
3	A016	ACTIVIDADES DE APOYO A LA AGRICULTURA Y LA GANADERÍA Y ACTIVIDADES POSCOSECHA.	1
3	A017	CAZA ORDINARIA, MEDIANTE TRAMPAS Y ACTIVIDADES DE SERVICIOS CONEXAS.	1
3	A021	SILVICULTURA Y OTRAS ACTIVIDADES FORESTALES.	1
3	A022	EXTRACCIÓN DE MADERA.	1
3	A023	RECOLECCIÓN DE PRODUCTOS FORESTALES DISTINTOS DE LA MADERA.	1
3	A024	SERVICIOS DE APOYO A LA SILVICULTURA.	1
3	A031	PESCA	3
3	A032	ACUICULTURA.	1
1	B	EXPLOTACIÓN DE MINAS Y CANTERAS.	
3	B051	EXTRACCIÓN DE CARBÓN DE PIEDRA.	3
3	B052	EXTRACCIÓN DE LIGNITO.	3

3	B061	EXTRACCIÓN DE PETRÓLEO CRUDO.	3
3	B062	EXTRACCIÓN DE GAS NATURAL.	3
3	B071	EXTRACCIÓN DE MINERALES DE HIERRO.	3
3	B072	EXTRACCIÓN DE MINERALES METALÍFEROS NO FERROSOS.	3
3	B081	EXTRACCIÓN DE PIEDRA, ARENA Y ARCILLA.	3
3	B089	EXPLOTACIÓN DE MINAS Y CANTERAS N.C.P.	3
3	B091	ACTIVIDADES DE APOYO PARA LA EXTRACCIÓN DE PETRÓLEO Y GAS NATURAL.	3
3	B099	ACTIVIDADES DE APOYO PARA LA EXPLOTACIÓN DE OTRAS MINAS Y CANTERAS.	3
		RIAS MANUFACTURERAS.	
3	C101	ELABORACIÓN Y CONSERVACIÓN DE CARNE.	1
3	C102	ELABORACIÓN Y CONSERVACIÓN DE PESCADOS, CRUSTÁCEOS Y MOLUSCOS.	1
3	C103	ELABORACIÓN Y CONSERVACIÓN DE FRUTAS, LEGUMBRES Y HORTALIZAS.	1
3	C104	ELABORACIÓN DE ACEITES Y GRASAS DE ORIGEN VEGETAL Y ANIMAL.	1
3	C105	ELABORACIÓN DE PRODUCTOS LÁCTEOS.	1
3	C106	ELABORACIÓN DE PRODUCTOS DE MOLINERÍA, ALMIDONES Y PRODUCTOS DERIVADOS DEL ALMIDÓN.	1
3	C107	ELABORACIÓN DE OTROS PRODUCTOS ALIMENTICIOS.	1
3	C108	ELABORACIÓN DE ALIMENTOS PREPARADOS PARA ANIMALES.	1
3	C110	ELABORACIÓN DE BEBIDAS.	1
3	C120	ELABORACIÓN DE PRODUCTOS DE TABACO.	2
3	C131	HILATURA, TEJEDURA Y ACABADOS DE PRODUCTOS TEXTILES.	2
3	C139	FABRICACIÓN DE OTROS PRODUCTOS TEXTILES.	2
3	C141	FABRICACIÓN DE PRENDAS DE VESTIR, EXCEPTO PRENDAS DE PIEL.	2
3	C142	FABRICACIÓN DE ARTÍCULOS DE PIEL.	2
3	C143	FABRICACIÓN DE ARTÍCULOS DE PUNTO Y GANCHILLO.	2
3	C151	CURTIDO Y ADOBO DE CUEROS; FABRICACIÓN DE MALETAS, BOLSOS DE MANO Y ARTÍCULOS DE TALABARTERÍA Y GUARNICIONERÍA; ADOBO Y TEÑIDO DE PIELES.	2
3	C152	FABRICACIÓN DE CALZADO.	2
3	C161	ASERRADO Y ACEPILLADURA DE MADERA.	2
3	C162	FABRICACIÓN DE HOJAS DE MADERA PARA ENCHAPADO Y TABLEROS A BASE DE MADERA.	2
3	C170	FABRICACIÓN DE PAPEL Y DE PRODUCTOS DE PAPEL.	2
3	C181	IMPRESIÓN Y ACTIVIDADES DE SERVICIOS RELACIONADOS CON LA IMPRESIÓN.	2
3	C182	REPRODUCCIÓN DE GRABACIONES.	2
3	C191	FABRICACIÓN DE PRODUCTOS DE HORNO DE COQUE.	2
3	C192	FABRICACIÓN DE PRODUCTOS DE LA REFINACIÓN DEL PETRÓLEO.	2
3	C201	FABRICACIÓN DE SUSTANCIAS QUÍMICAS BÁSICAS, DE ABONOS Y COMPUESTOS DE NITRÓGENO Y DE PLÁSTICOS Y CAUCHO	2

		SINTÉTICO EN FORMAS PRIMARIAS.	
3	C202	FABRICACIÓN DE OTROS PRODUCTOS QUÍMICOS.	3
3	C203	FABRICACIÓN DE FIBRAS ARTIFICIALES.	2
3	C210	FABRICACIÓN DE PRODUCTOS FARMACÉUTICOS, SUSTANCIAS QUÍMICAS MEDICINALES Y PRODUCTOS BOTÁNICOS DE USO FARMACÉUTICO.	3
3	C221	FABRICACIÓN DE PRODUCTOS DE CAUCHO.	2
3	C222	FABRICACIÓN DE PRODUCTOS DE PLÁSTICO.	2
3	C231	FABRICACIÓN DE VIDRIO Y PRODUCTOS DE VIDRIO.	2
3	C239	FABRICACIÓN DE PRODUCTOS MINERALES NO METÁLICOS N.C.P.	3
3	C241	INDUSTRIAS BÁSICAS DE HIERRO Y ACERO.	1
3	C242	FABRICACIÓN DE PRODUCTOS PRIMARIOS DE METALES PRECIOSOS Y METALES NO FERROSOS.	3
3	C243	FUNDICIÓN DE METALES.	1
3	C251	FABRICACIÓN DE PRODUCTOS METÁLICOS PARA USO ESTRUCTURAL, TANQUES, DEPÓSITOS, RECIPIENTES DE METAL Y GENERADORES DE VAPOR.	1
3	C252	FABRICACIÓN DE ARMAS Y MUNICIONES.	3
3	C259	FABRICACIÓN DE OTROS PRODUCTOS ELABORADOS DE METAL, ACTIVIDADES DE TRABAJOS DE METALES.	1
3	C261	FABRICACIÓN DE COMPONENTES Y TABLEROS ELECTRÓNICOS.	1
3	C262	FABRICACIÓN DE ORDENADORES Y EQUIPO PERIFÉRICO.	1
3	C263	FABRICACIÓN DE EQUIPO DE COMUNICACIONES.	1
3	C264	FABRICACIÓN DE APARATOS ELECTRÓNICOS DE CONSUMO.	1
3	C265	FABRICACIÓN DE EQUIPOS DE MEDICIÓN, PRUEBA, NAVEGACIÓN, CONTROL Y DE RELOJES.	2
3	C266	FABRICACIÓN DE EQUIPO DE IRRADIACIÓN, Y EQUIPO ELECTRÓNICO DE USO MEDICO Y TERAPÉUTICO.	2
3	C267	FABRICACIÓN DE INSTRUMENTOS ÓPTICOS Y EQUIPO FOTOGRÁFICOS.	2
3	C268	FABRICACIÓN DE SOPORTES MAGNÉTICOS Y ÓPTICOS.	2
3	C271	FABRICACIÓN DE MOTORES, GENERADORES, TRANSFORMADORES ELÉCTRICOS Y APARATOS DE DISTRIBUCIÓN Y CONTROL DE LA ENERGÍA ELÉCTRICA.	2
3	C272	FABRICACIÓN DE PILAS, BATERÍAS Y ACUMULADORES.	2
3	C273	FABRICACIÓN DE CABLES Y DISPOSITIVOS DE CABLEADO.	2
3	C274	FABRICACIÓN DE EQUIPO ELÉCTRICO DE ILUMINACIÓN.	2
3	C275	FABRICACIÓN DE APARATOS DE USO DOMÉSTICO.	2
3	C279	FABRICACIÓN DE OTROS TIPOS DE EQUIPO ELÉCTRICO.	2
3	C281	FABRICACIÓN DE MOTORES Y TURBINAS, EXCEPTO MOTORES PARA AERONAVES, VEHÍCULOS AUTOMOTORES Y MOTOCICLETAS.	2
3	C282	FABRICACIÓN DE MAQUINARIA DE USO ESPECIAL.	2
3	C291	FABRICACIÓN DE VEHÍCULOS AUTOMOTORES.	2
3	C292	FABRICACIÓN DE CARROCERÍAS PARA VEHÍCULOS AUTOMOTORES; FABRICACIÓN DE REMOLQUES Y SEMIRREMOLQUES.	3

3	C293	FABRICACIÓN DE PARTES, PIEZAS Y ACCESORIOS PARA VEHÍCULOS AUTOMOTORES.	3
3	C301	CONSTRUCCIÓN DE BUQUES Y OTRAS EMBARCACIONES.	3
3	C302	FABRICACIÓN DE LOCOMOTORAS Y MATERIAL RODANTE.	3
3	C303	FABRICACIÓN DE AERONAVES Y NAVES ESPACIALES Y MAQUINARIA CONEXA.	3
3	C304	FABRICACIÓN DE VEHÍCULOS MILITARES DE COMBATE.	3
3	C309	FABRICACIÓN DE EQUIPO DE TRANSPORTE N.C.P.	3
3	C310	FABRICACIÓN DE MUEBLES.	2
3	C321	FABRICACIÓN DE JOYAS, BISUTERÍA Y ARTÍCULOS CONEXOS.	3
3	C322	FABRICACIÓN DE INSTRUMENTOS MUSICALES.	2
3	C323	FABRICACIÓN DE ARTÍCULOS DE DEPORTE.	2
3	C324	FABRICACIÓN DE JUEGOS Y JUGUETES.	2
3	C325	FABRICACIÓN DE INSTRUMENTOS Y MATERIALES MÉDICOS Y ODONTOLÓGICOS.	2
3	C329	OTRAS INDUSTRIAS MANUFACTURERAS N.C.P.	2
3	C331	REPARACIÓN DE PRODUCTOS ELABORADOS DE METAL, MAQUINARIA Y EQUIPO.	2
3	C332	INSTALACIÓN DE MAQUINARIA Y EQUIPO INDUSTRIALES.	2
1	D	SUMINISTRO DE ELECTRICIDAD, GAS, VAPOR Y AIRE ACONDICIONADO.	
3	D351	GENERACIÓN, TRANSMISIÓN Y DISTRIBUCIÓN DE ENERGÍA ELÉCTRICA.	2
3	D352	FABRICACIÓN DE GAS; DISTRIBUCIÓN DE COMBUSTIBLES GASEOSOS POR TUBERÍAS.	2
3	D353	SUMINISTRO DE VAPOR Y DE AIRE ACONDICIONADO.	2
1	E	DISTRIBUCIÓN DE AGUA; ALCANTARILLADO, GESTIÓN DE DESECHOS Y ACTIVIDADES DE SANEAMIENTO.	
3	E360	CAPTACIÓN, TRATAMIENTO Y DISTRIBUCIÓN DE AGUA.	1
3	E370	EVACUACIÓN DE AGUAS RESIDUALES.	1
3	E381	RECOLECCIÓN DE DESECHOS.	1
3	E382	TRATAMIENTO Y ELIMINACIÓN DE DESECHOS.	1
3	E383	RECUPERACIÓN DE MATERIALES.	1
3	E390	ACTIVIDADES DE DESCONTAMINACIÓN Y OTROS SERVICIOS DE GESTIÓN DE DESECHOS.	1
1	F	CONSTRUCCIÓN.	
3	F410	CONSTRUCCIÓN DE EDIFICIOS.	3
3	F421	CONSTRUCCIÓN DE CARRETERAS Y LÍNEAS DE FERROCARRIL.	3
3	F422	CONSTRUCCIÓN DE PROYECTOS DE SERVICIOS PÚBLICOS.	3
3	F429	CONSTRUCCIÓN DE OTRAS OBRAS DE INGENIERÍA CIVIL.	3
3	F431	DEMOLICIÓN Y PREPARACIÓN DEL TERRENO.	3
3	F432	INSTALACIONES ELÉCTRICAS Y DE FONTANERÍA Y OTRAS INSTALACIONES PARA OBRAS DE CONSTRUCCIÓN.	3
3	F433	TERMINACIÓN Y ACABADO DE EDIFICIOS.	3
3	F439	OTRAS ACTIVIDADES ESPECIALIZADAS DE CONSTRUCCIÓN.	3
1	G	COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACIÓN DE VEHÍCULOS AUTOMOTORES Y MOTOCICLETAS.	

3	G451	VENTA DE VEHÍCULOS AUTOMOTORES.	3
3	G452	MANTENIMIENTO Y REPARACIÓN DE VEHÍCULOS AUTOMOTORES.	3
3	G453	VENTA DE PARTES, PIEZAS Y ACCESORIOS PARA VEHÍCULOS AUTOMOTORES.	2
3	G454	VENTA, MANTENIMIENTO Y REPARACIÓN DE MOTOCICLETAS Y DE SUS PARTES, PIEZAS Y ACCESORIOS.	2
3	G461	VENTA AL POR MAYOR A CAMBIO DE UNA COMISIÓN O POR CONTRATO.	3
3	G462	VENTA AL POR MAYOR DE MATERIAS PRIMAS AGROPECUARIAS Y ANIMALES VIVOS.	2
3	G463	VENTA AL POR MAYOR DE ALIMENTOS, BEBIDAS Y TABACO.	2
3	G464	VENTA AL POR MAYOR DE ENSERES DOMÉSTICOS.	2
3	G465	VENTA AL POR MAYOR DE MAQUINARIAS EQUIPOS Y MATERIALES.	2
3	G466	OTRAS ACTIVIDADES DE VENTA AL POR MAYOR ESPECIALIZADA.	2
3	G469	VENTA AL POR MAYOR DE OTROS PRODUCTOS NO ESPECIALIZADOS.	2
3	G471	VENTA AL POR MENOR EN COMERCIOS NO ESPECIALIZADOS.	1
3	G472	VENTA AL POR MENOR DE ALIMENTOS, BEBIDAS Y TABACO EN COMERCIOS ESPECIALIZADOS.	1
3	G473	VENTA AL POR MENOR DE COMBUSTIBLES PARA VEHÍCULOS AUTOMOTORES EN COMERCIOS ESPECIALIZADOS.	1
3	G474	VENTA AL POR MENOR DE EQUIPO DE INFORMACIÓN Y DE COMUNICACIONES EN COMERCIOS ESPECIALIZADOS.	1
3	G475	VENTA AL POR MENOR DE OTROS ENSERES DOMÉSTICOS EN COMERCIOS ESPECIALIZADOS.	1
3	G476	VENTA AL POR MENOR DE PRODUCTOS CULTURALES Y RECREATIVOS EN COMERCIOS ESPECIALIZADOS.	1
3	G477	VENTA AL POR MENOR DE OTROS PRODUCTOS EN COMERCIOS ESPECIALIZADOS.	1
3	G478	VENTA AL POR MENOR EN PUESTOS DE VENTA Y MERCADOS.	1
3	G479	VENTA AL POR MENOR NO REALIZADA EN COMERCIOS, PUESTOS DE VENTA O MERCADOS.	1
1	H	TRANSPORTE Y ALMACENAMIENTO.	
3	H491	TRANSPORTE POR FERROCARRIL.	2
3	H492	OTRAS ACTIVIDADES DE TRANSPORTE POR VÍA TERRESTRE.	3
3	H493	TRANSPORTE POR TUBERÍAS.	1
3	H501	TRANSPORTE MARÍTIMO Y DE CABOTAJE.	3
3	H502	TRANSPORTE POR VÍAS DE NAVEGACIÓN INTERIORES.	3
3	H511	TRANSPORTE DE PASAJEROS POR VÍA AÉREA.	3
3	H512	TRANSPORTE DE CARGA POR VÍA AÉREA.	3
3	H521	ALMACENAMIENTO Y DEPÓSITO.	3
3	H522	ACTIVIDADES DE APOYO AL TRANSPORTE.	3

3	H531	ACTIVIDADES POSTALES Y DE MENSAJERÍA.	3
3	H532	ACTIVIDADES DE MENSAJERÍA.	3
1	I	ACTIVIDADES DE ALOJAMIENTO Y DE SERVICIO DE COMIDAS.	
3	I551	ACTIVIDADES DE ALOJAMIENTO PARA ESTANCIAS CORTAS.	3
3	I552	ACTIVIDADES DE CAMPAMENTOS, PARQUES DE VEHÍCULOS DE RECREO Y PARQUES DE CARAVANAS.	3
3	I559	OTRAS ACTIVIDADES DE ALOJAMIENTO.	3
3	I561	ACTIVIDADES DE RESTAURANTES Y DE SERVICIO MÓVIL DE COMIDAS.	3
3	I562	SUMINISTRO DE COMIDAS POR ENCARGO Y OTRAS ACTIVIDADES DE SERVICIO DE COMIDAS.	2
3	I563	ACTIVIDADES DE SERVICIO DE BEBIDAS.	2
1	J	INFORMACIÓN Y COMUNICACIÓN.	
3	J581	PUBLICACIÓN DE LIBROS, PERIÓDICOS Y OTRAS ACTIVIDADES DE PUBLICACIÓN.	1
3	J582	PUBLICACIÓN DE PROGRAMAS INFORMÁTICOS.	1
3	J591	ACTIVIDADES DE PRODUCCIÓN DE PELÍCULAS CINEMATOGRAFICAS, VÍDEOS Y PROGRAMAS DE TELEVISIÓN.	1
3	J592	ACTIVIDADES DE GRABACIÓN DE SONIDO Y EDICIÓN DE MÚSICA.	1
3	J601	TRANSMISIONES DE RADIO.	1
3	J602	PROGRAMACIÓN Y TRANSMISIONES DE TELEVISIÓN.	1
3	J611	ACTIVIDADES DE TELECOMUNICACIONES ALÁMBRICA.	1
3	J612	ACTIVIDADES DE TELECOMUNICACIONES INALÁMBRICAS.	1
3	J613	ACTIVIDADES DE TELECOMUNICACIONES POR SATÉLITE.	1
3	J619	OTRAS ACTIVIDADES DE TELECOMUNICACIONES.	1
3	J620	ACTIVIDADES DE PROGRAMACIÓN INFORMÁTICA Y DE CONSULTORÍA DE INFORMÁTICA Y ACTIVIDADES CONEXAS.	1
3	J631	PROCESAMIENTO DE DATOS, HOSPEDAJE Y ACTIVIDADES CONEXAS; PORTALES WEB.	1
3	J639	OTRAS ACTIVIDADES DE SERVICIOS DE INFORMACIÓN.	1
1	K	ACTIVIDADES FINANCIERAS Y DE SEGUROS.	
3	K641	INTERMEDIACIÓN MONETARIA.	3
3	K642	ACTIVIDADES DE SOCIEDADES DE CARTERA.	3
3	K643	FONDOS Y SOCIEDADES DE INVERSIÓN Y ENTIDADES FINANCIERAS SIMILARES.	3
3	K649	OTRAS ACTIVIDADES DE SERVICIOS FINANCIEROS, EXCEPTO LAS DE SEGUROS Y FONDOS DE PENSIONES.	3
3	K651	SEGUROS.	3
3	K652	REASEGUROS.	3
3	K653	FONDOS DE PENSIÓN.	3
3	K661	ACTIVIDADES AUXILIARES DE LAS ACTIVIDADES DE SERVICIOS FINANCIEROS, EXCEPTO LAS DE SEGUROS Y FONDOS DE PENSIONES.	3
3	K662	ACTIVIDADES AUXILIARES DE SEGUROS Y FONDOS DE PENSIONES.	3

3	K663	ACTIVIDADES DE GESTIÓN DE FONDOS.	3
1	L	ACTIVIDADES INMOBILIARIAS.	
3	L681	ACTIVIDADES INMOBILIARIAS REALIZADAS CON BIENES PROPIOS O ARRENDADOS.	3
3	L682	ACTIVIDADES INMOBILIARIAS REALIZADAS A CAMBIO DE UNA RETRIBUCIÓN O POR CONTRATO.	3
1	M	ACTIVIDADES PROFESIONALES, CIENTÍFICAS Y TÉCNICAS.	
3	M691	ACTIVIDADES JURÍDICAS.	3
3	M692	ACTIVIDADES DE CONTABILIDAD, TENEDURÍA DE LIBROS Y AUDITORIA; CONSULTORÍA FISCAL.	3
3	M701	ACTIVIDADES DE OFICINAS PRINCIPALES.	2
3	M702	ACTIVIDADES DE CONSULTORÍA DE GESTIÓN.	2
3	M711	ACTIVIDADES DE ARQUITECTURA E INGENIERÍA Y ACTIVIDADES CONEXAS DE CONSULTORÍA TÉCNICA.	2
3	M712	ENSAYOS Y ANÁLISIS TÉCNICOS.	2
3	M721	INVESTIGACIONES Y DESARROLLO EXPERIMENTAL EN EL CAMPO DE LAS CIENCIAS NATURALES Y LA INGENIERÍA.	2
3	M722	INVESTIGACIONES Y DESARROLLO EXPERIMENTAL EN EL CAMPO DE LAS CIENCIAS SOCIALES Y LAS HUMANIDADES.	2
3	M731	PUBLICIDAD.	2
3	M732	ESTUDIOS DE MERCADO Y ENCUESTAS DE OPINIÓN PÚBLICA.	2
3	M741	ACTIVIDADES ESPECIALIZADAS DE DISEÑO.	2
3	M742	ACTIVIDADES DE FOTOGRAFÍA.	2
3	M749	OTRAS ACTIVIDADES PROFESIONALES, CIENTÍFICAS Y TÉCNICAS	2
3	M750	ACTIVIDADES VETERINARIAS.	1
1	N	ACTIVIDADES DE SERVICIOS ADMINISTRATIVOS Y DE APOYO.	
3	N771	ALQUILER DE VEHÍCULOS AUTOMOTORES.	3
3	N772	ALQUILER DE EFECTOS PERSONALES Y ARTÍCULOS DE USO DOMÉSTICO.	1
3	N773	ALQUILER DE OTROS TIPOS DE MAQUINARIA, EQUIPO Y BIENES TANGIBLES.	3
3	N774	ARRENDAMIENTO DE PROPIEDAD INTELECTUAL Y PRODUCTOS SIMILARES, EXCEPTO OBRAS PROTEGIDAS POR DERECHOS DE AUTOR.	2
3	N781	ACTIVIDADES DE AGENCIAS DE EMPLEO.	2
3	N782	ACTIVIDADES DE AGENCIAS DE EMPLEO TEMPORAL.	2
3	N783	OTRAS ACTIVIDADES DE DOTACIÓN DE RECURSOS HUMANOS.	2
3	N791	ACTIVIDADES DE AGENCIAS DE VIAJES Y OPERADORES TURÍSTICOS.	3
3	N799	OTROS SERVICIOS DE RESERVAS Y ACTIVIDADES CONEXAS.	2
3	N801	ACTIVIDADES DE SEGURIDAD PRIVADA.	1
3	N802	ACTIVIDADES DE SERVICIOS DE SISTEMAS DE SEGURIDAD.	1
3	N803	ACTIVIDADES DE INVESTIGACIÓN.	1
3	N811	ACTIVIDADES COMBINADAS DE APOYO A INSTALACIONES.	1
3	N812	ACTIVIDADES DE LIMPIEZA.	1
3	N813	ACTIVIDADES DE PAISAJISMO Y SERVICIOS DE MANTENIMIENTO CONEXOS.	1

3	N821	ACTIVIDADES ADMINISTRATIVAS Y DE APOYO DE OFICINA.	2
3	N822	ACTIVIDADES DE CENTROS DE LLAMADAS.	1
3	N823	ORGANIZACIÓN DE CONVENCIONES Y EXPOSICIONES COMERCIALES.	1
3	N829	ACTIVIDADES DE SERVICIOS DE APOYO A LAS EMPRESAS N.C.P.	1
1	O	ADMINISTRACIÓN PÚBLICA Y DEFENSA; PLANES DE SEGURIDAD SOCIAL DE AFILIACIÓN OBLIGATORIA.	
3	O841	ADMINISTRACIÓN DEL ESTADO Y APLICACIÓN DE LA POLÍTICA ECONÓMICA Y SOCIAL DE LA COMUNIDAD.	1
3	O842	PRESTACIÓN DE SERVICIOS A LA COMUNIDAD EN GENERAL.	1
3	O843	ACTIVIDADES DE PLANES DE SEGURIDAD SOCIAL DE AFILIACIÓN OBLIGATORIA.	1
1	P	ENSEÑANZA.	
3	P851	ENSEÑANZA PREPRIMARIA Y PRIMARIA.	1
3	P852	ENSEÑANZA SECUNDARIA.	1
3	P853	ENSEÑANZA SUPERIOR.	1
3	P854	OTROS TIPOS DE ENSEÑANZA.	1
3	P855	ACTIVIDADES DE APOYO A LA ENSEÑANZA.	1
1	Q	ACTIVIDADES DE ATENCIÓN DE LA SALUD HUMANA Y DE ASISTENCIA SOCIAL.	
3	Q861	ACTIVIDADES DE HOSPITALES Y CLÍNICAS.	1
3	Q862	ACTIVIDADES DE MÉDICOS Y ODONTÓLOGOS.	1
3	Q869	OTRAS ACTIVIDADES DE ATENCIÓN DE LA SALUD HUMANA.	1
3	Q871	ACTIVIDADES DE ATENCIÓN DE ENFERMERÍA EN INSTITUCIONES.	1
3	Q872	ACTIVIDADES DE ATENCIÓN EN INSTITUCIONES PARA PERSONAS CON RETRASO MENTAL, ENFERMOS MENTALES Y TOXICÓMANOS.	1
3	Q873	ACTIVIDADES DE ATENCIÓN EN INSTITUCIONES PARA PERSONAS DE EDAD Y PERSONAS CON DISCAPACIDAD.	1
3	Q879	OTRAS ACTIVIDADES DE ATENCIÓN EN INSTITUCIONES.	1
3	Q881	ACTIVIDADES DE ASISTENCIA SOCIAL SIN ALOJAMIENTO PARA PERSONAS DE EDAD Y PERSONAS CON DISCAPACIDAD.	1
3	Q889	OTRAS ACTIVIDADES DE ASISTENCIA SOCIAL SIN ALOJAMIENTO.	1
1	R	ARTES, ENTRETENIMIENTO Y RECREACIÓN.	
3	R900	ACTIVIDADES CREATIVAS, ARTÍSTICAS Y DE ENTRETENIMIENTO.	3
3	R910	ACTIVIDADES DE BIBLIOTECAS, ARCHIVOS, MUSEOS Y OTRAS ACTIVIDADES CULTURALES.	2
3	R920	ACTIVIDADES DE JUEGOS DE AZAR Y APUESTAS.	3
3	R931	ACTIVIDADES DEPORTIVAS.	2
3	R932	OTRAS ACTIVIDADES DE ESPARCIMIENTO Y RECREATIVAS.	3
1	S	OTRAS ACTIVIDADES DE SERVICIOS.	
3	S941	ACTIVIDADES DE ASOCIACIONES EMPRESARIALES, PROFESIONALES Y DE EMPLEADORES.	3
3	S942	ACTIVIDADES DE SINDICATOS.	3
3	S949	ACTIVIDADES DE OTRAS ASOCIACIONES.	3
3	S951	REPARACIÓN DE COMPUTADORAS Y EQUIPO DE	1

		COMUNICACIONES.	
3	S952	REPARACIÓN DE EFECTOS PERSONALES Y ENSERES DOMÉSTICOS.	1
3	S960	OTRAS ACTIVIDADES DE SERVICIOS PERSONALES.	1
1	T	ACTIVIDADES DE LOS HOGARES COMO EMPLEADORES; ACTIVIDADES NO DIFERENCIADAS DE LOS HOGARES COMO PRODUCTORES DE BIENES Y SERVICIOS PARA USO PROPIO.	
3	T970	ACTIVIDADES DE LOS HOGARES COMO EMPLEADORES DE PERSONAL DOMÉSTICO.	1
3	T981	ACTIVIDADES NO DIFERENCIADAS DE LOS HOGARES COMO PRODUCTORES DE BIENES Y SERVICIOS PARA USO PROPIO.	1
3	T982	ACTIVIDADES NO DIFERENCIADAS DE LOS HOGARES COMO PRODUCTORES DE SERVICIOS PARA USO PROPIO.	1
1	U	ACTIVIDADES DE ORGANIZACIONES Y ÓRGANOS EXTRATERRITORIALES.	
3	U990	ACTIVIDADES DE ORGANIZACIONES Y ÓRGANOS EXTRATERRITORIALES.	3

Tabla 4 Actividades económicas por Riesgo

Fuente: Recomendaciones Grupo de Acción Financiera Internacional - GAFI

En referencia a la actividad económica, del total de los usuarios se obtuvo que el 64,4% están categorizados como “Riesgo Bajo”, y el 35,2% está vinculado a la categoría de “Riesgo Alto”, tal y como se muestra en la Tabla 4 y la Ilustración 4.

ACTIVIDAD_ECONOMICA					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	RIESGO BAJO	19010	64,4	64,4	64,4
	RIEGO MEDIO	131	,4	,4	64,8
	RIESGO ALTO	10377	35,2	35,2	100,0
	Total	29518	100,0	100,0	
Perdidos	Sistema	2	,0		
Total		29520	100,0		

Tabla 5 Frecuencia de Actividad Económica

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

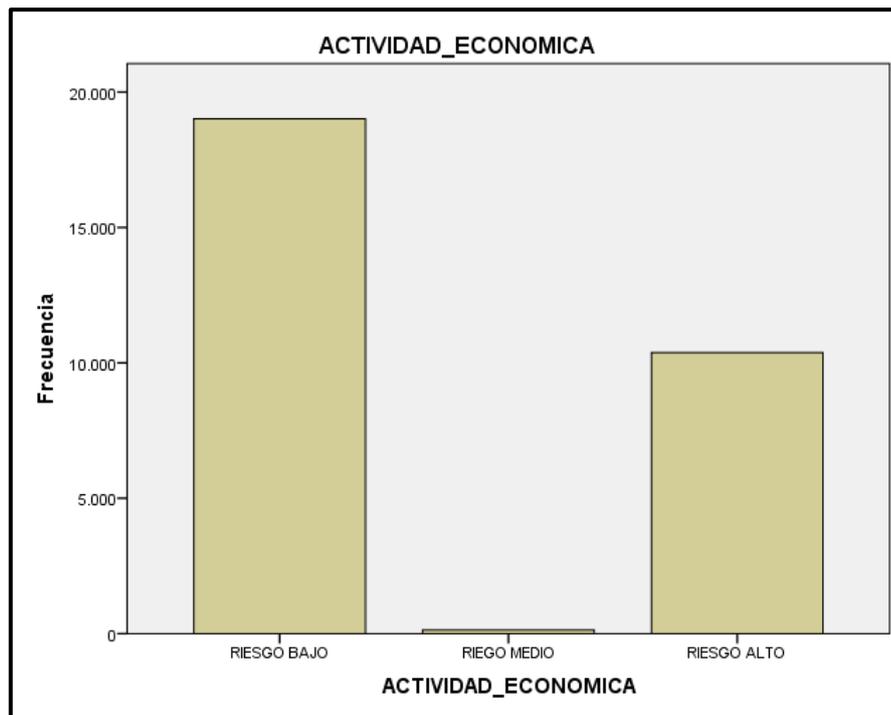


Ilustración 4 Histograma de Actividad Económica

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

- **Nacionalidad.-** condición que reconoce a una persona la pertenencia a un estado o nación, y para la aplicación de los análisis la variable está codificada como:

1 = Ecuatoriana

2 = Otros

Sin embargo a continuación se muestran los resultados de las nacionalidades de los clientes, el 97,3% de la nacionalidad de los clientes es ecuatoriana como se observa en la tabla 6 e ilustración 6, y en porcentajes más pequeños tenemos la estadounidense, colombiana y venezolana.

NACIONALIDAD2				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	381	1,3	1,3	1,3
ALEMANA	1	,0	,0	1,3
ARGENTINA	4	,0	,0	1,3
BELGICA	1	,0	,0	1,3
BOLIVIANA	2	,0	,0	1,3
BRASILEÑA	2	,0	,0	1,3
CAMERUN	1	,0	,0	1,3
CANADIENSE	1	,0	,0	1,3
CHECA	1	,0	,0	1,3
CHILENA	10	,0	,0	1,4
CHINA	4	,0	,0	1,4
COLOMBIANA	177	,6	,6	2,0
COREA DEL NORTE	4	,0	,0	2,0
COREA DEL SUR	1	,0	,0	2,0
CUBANA	36	,1	,1	2,1
DOMINICANA	4	,0	,0	2,1
ECUATORIANA	28728	97,3	97,3	99,5
ESPAÑOLA	15	,1	,1	99,5
ESTADOUNIDENSE	59	,2	,2	99,7
FRANCESA	2	,0	,0	99,7
HOLANDESA	1	,0	,0	99,7
HONDUREÑA	1	,0	,0	99,7
INDIA	2	,0	,0	99,7
INGLESA	2	,0	,0	99,7
ITALIANA	1	,0	,0	99,7
LIBANES	3	,0	,0	99,7
LIBIA	1	,0	,0	99,7
MEXICANA	1	,0	,0	99,7
PANAMEÑO	1	,0	,0	99,8
PERUANA	22	,1	,1	99,8
POLACA	2	,0	,0	99,8
PUERTORRIQUEÑO	3	,0	,0	99,8
RUSA	2	,0	,0	99,9
SALVADOREÑA	2	,0	,0	99,9
URUGUAYA	3	,0	,0	99,9
VENEZOLANA	39	,1	,1	100,0
Total	29520	100,0	100,0	

Tabla 6 Frecuencia de Nacionalidades

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

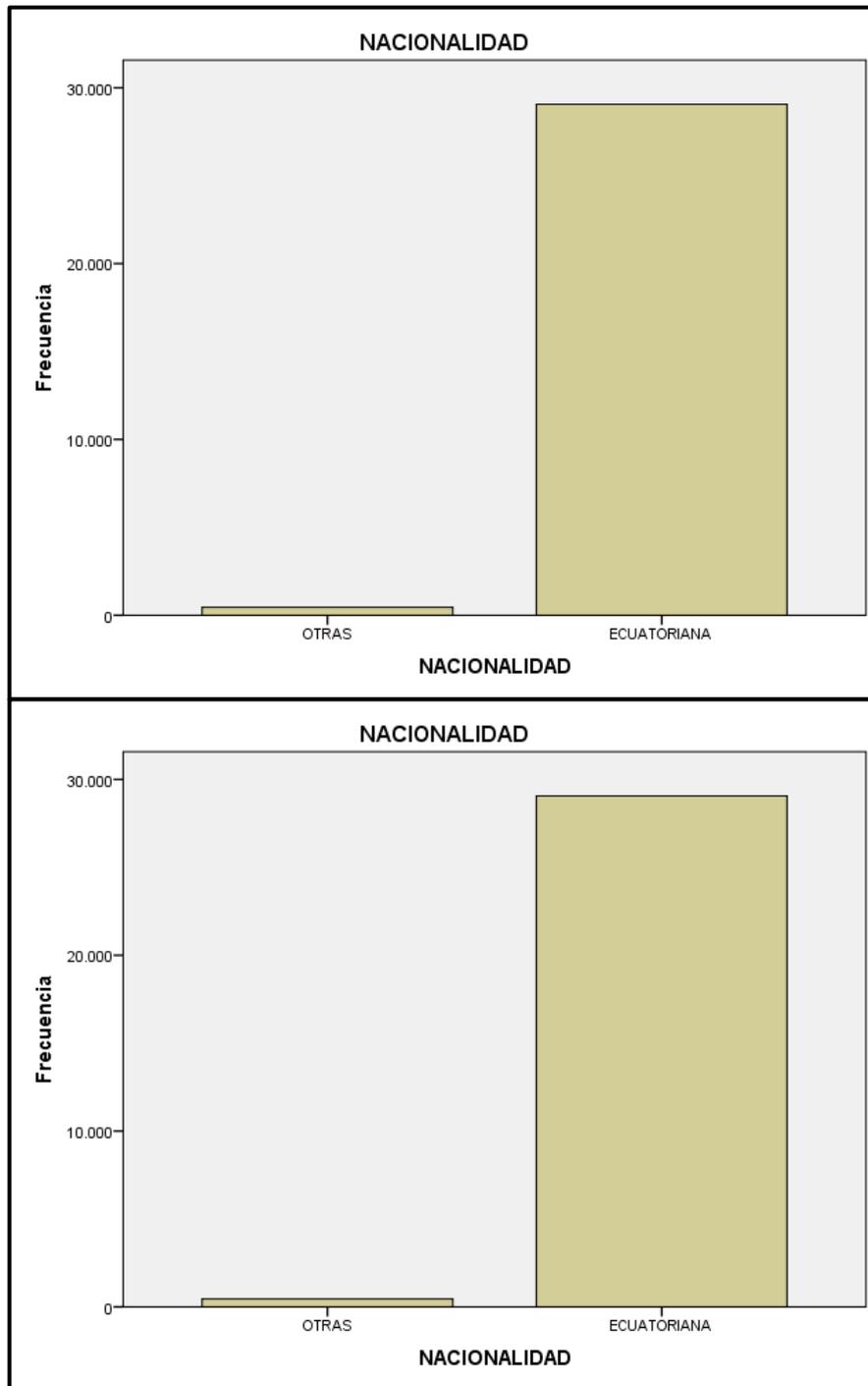


Ilustración 5 Histograma de Nacionalidades

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

- **Sexo.-** conjunto de características físicas, biológicas, anatómicas y fisiológicas de los seres humanos, que los definen como hombre o mujer, siendo:

1 = Masculino

2 = Femenino

Como se muestra en la Tabla 7 y la Ilustración 7, en cuanto al sexo de los clientes, existe una relación muy pareja, obteniéndose así que el 57,4% pertenecen al sexo femenino y el 42,6% al sexo masculino.

SEXO					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	M	12565	42,6	42,6	42,6
	F	16955	57,4	57,4	100,0
	Total	29520	100,0	100,0	

Tabla 7 Frecuencia de Sexo

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

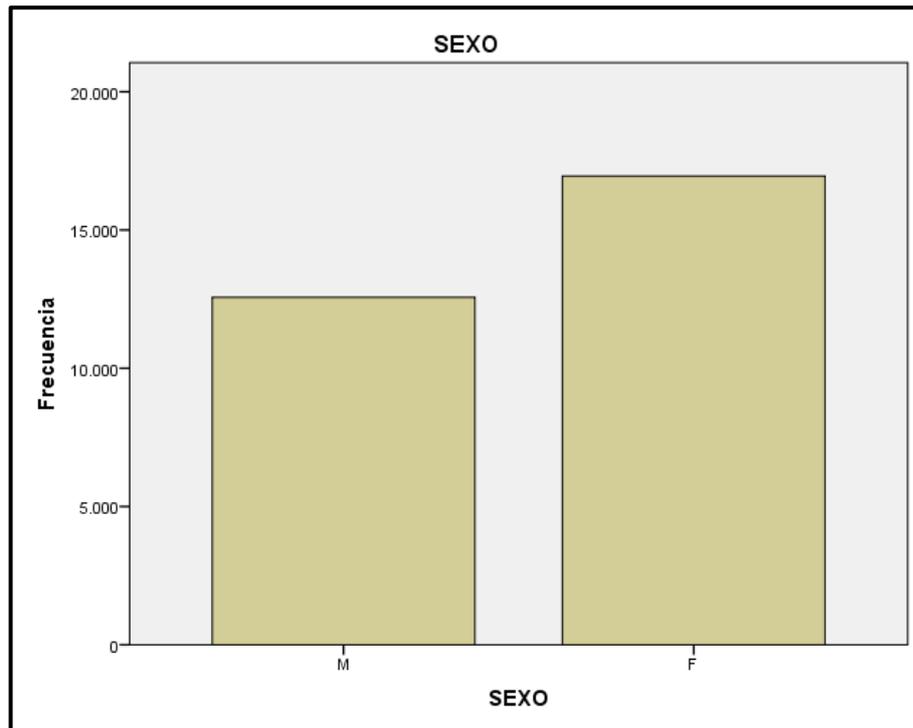


Ilustración 6 Histograma de Sexo

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

- **Patrimonio.-** conjunto de bienes propios de una persona o una institución, susceptibles de estimación económica.

Estadísticos		
PATRIMONIO		
N	Válido	29140
	Perdidos	380
Media		28611,4038
Error estándar de la media		730,07211
Mediana		2022,5000
Moda		,00
Mínimo		,00
Máximo		11554894,00
Percentiles	25	300,0000
	50	2022,5000
	75	25300,0000

Tabla 8 Estadístico de la Variable Patrimonio

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

De acuerdo a los datos analizados e ilustrados en la Tabla 8 estadísticos de las variables, podemos especificar que en la variable “Patrimonio”, se obtuvo que la media de los valores que los clientes perciben es de \$ 28.611,65, y la mediana de \$2.022,50.

Para el histograma de la variable “Patrimonio” se realizó la agrupación de los valores como se muestra en la Tabla 9 e Ilustración 8, mostrando los resultados de la nueva variable “Patrimonio agrupado”.

PATRIMONIO (Agrupada)					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	< 10000	19333	65,5	66,3	66,3
	10001 - 20000	1665	5,6	5,7	72,1
	> 20000	8142	27,6	27,9	100,0
	Total	29140	98,7	100,0	
Perdidos	Sistema	380	1,3		
Total		29520	100,0		

Tabla 9 Frecuencia de Patrimonio (Agrupado)

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

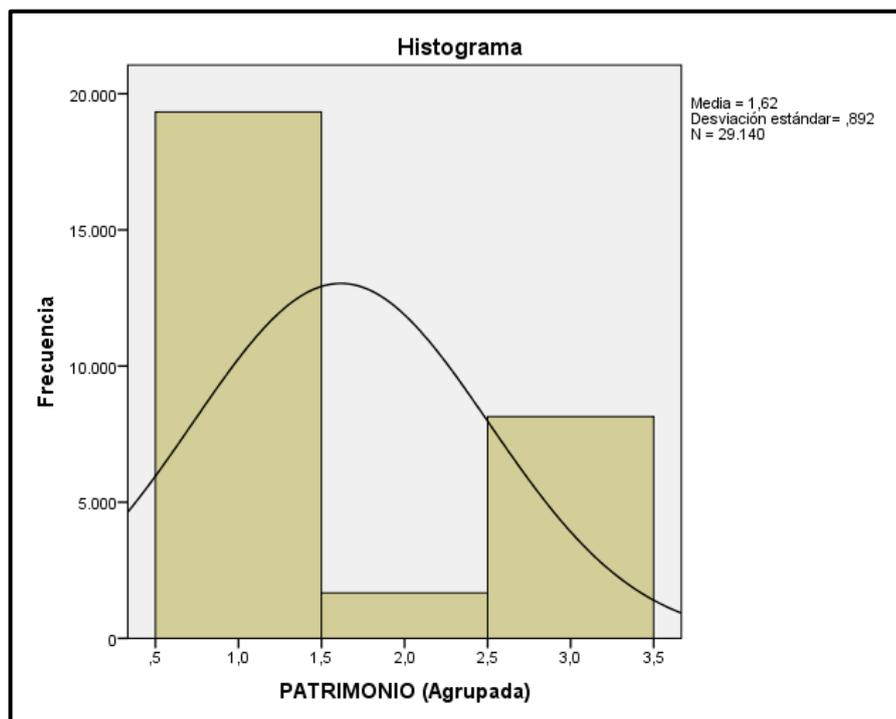


Ilustración 7 Histograma de Patrimonio (Agrupado)

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

CAPÍTULO IV

RESULTADOS

4.1 ANÁLISIS DE CONTINGENCIA

Dado que las variables utilizadas en el análisis son de carácter cualitativo, realizamos un análisis de contingencia como punto inicial para posteriormente aplicar otras técnicas.

Para realizar el análisis de contingencia se construye tablas donde se presenta la información de dos o más variables, generalmente cualitativas; para realizar comparaciones de relación e independencia se definen categorías de una variable a través de la frecuencia o porcentaje de una segunda variable, obteniendo un análisis descriptivo que resuma las categorías de las variables involucradas.

A partir de esto se puede generar un análisis de inferencia, para establecer relaciones sobre las variables objeto de estudio utilizando la prueba de independencia Chi cuadrado.

Hipótesis 1:

- H_0 : Las variables Ingresos Agrupados y Producto/Servicio son independientes.
- H_1 : Existe dependencia entre las variables.

Tabla cruzada INGRESOS (Agrupada)*PRODUCTO_SERVICIO				
% del total		PRODUCTO_SERVICIO		
		CUENTA BANCARIA	REMESAS	Total
INGRESOS (Agrupada)	< 500	20,4%	14,8%	35,3%
	501 - 1500	22,1%	17,3%	39,4%
	1501 - 2500	7,2%	4,3%	11,5%
	>2501	10,7%	3,1%	13,8%
Total		60,4%	39,6%	100,0%

Tabla 10 Tabla de contingencia Ingresos - Productos/Servicios

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Los resultados de la tabla cruzada de la variable “Ingresos Agrupados” con la variable “Producto/Servicio” (véase Tabla 10), nos indica que el 22,1% de los clientes mantienen ingresos en el rango de \$501 y\$1500 y eligen como producto/servicio la apertura de cuenta bancaria, siendo éste el porcentaje más alto de la comparación realizada.

Pruebas de chi-cuadrado			
	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	611,176 ^a	3	,000
Razón de verosimilitud	648,890	3	,000
Asociación lineal por lineal	404,520	1	,000
N de casos válidos	29140		

a. 0 casillas (0,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 1323,37.

Tabla 11 Prueba chi-cuadrado Ingreso - Producto/Servicio

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Como se muestra en la Tabla 11 de la prueba chi-cuadrado y cuya significación asintótica es $< 0,05$ podemos concluir las variables Ingresos y ProductoServicio son dependientes (Sig. 0,000), por lo tanto se rechaza la Hipótesis H_0 .

Hipótesis 2:

- H_0 : Las variables Ingresos Agrupados y Actividad Económica son independientes.
- H_1 : Existe dependencia entre las variables.

Tabla cruzada INGRESOS (Agrupada)*ACTIVIDAD_ECONOMICA					
% del total		ACTIVIDAD_ECONOMICA			Total
		RIESGO BAJO	RIEGO MEDIO	RIESGO ALTO	
INGRESOS (Agrupada)	< 500	25,7%	0,1%	9,5%	35,3%
	501 - 1500	26,1%	0,2%	13,2%	39,4%
	1501 - 2500	6,4%	0,0%	5,0%	11,5%
	>2501	6,3%	0,1%	7,5%	13,8%
Total		64,4%	0,4%	35,2%	100,0%

Tabla 12 Tabla de contingencia Ingreso - Actividad Económica

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Los resultados obtenidos de la variable “Ingresos Agrupados” con la variable “Actividad económica” (véase Tabla 12), podemos destacar que, el 26,1% de los clientes desempeñan actividades económica de “Riesgo Bajo”, de los cuales su mayor porcentaje de ingresos se mantiene en la escala \$501- \$1500.

Pruebas de chi-cuadrado			
	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	1090,852 ^a	6	,000
Razón de verosimilitud	1067,997	6	,000
Asociación lineal por lineal	1069,990	1	,000
N de casos válidos	29140		

a. 0 casillas (0,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 14,80.

Tabla 13 Prueba chi-cuadrado Ingreso - Actividad Económica

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Podemos concluir que existe una asociación entre los Ingresos de los clientes y las Actividades Económicas a las que están involucrados como se muestra en la Tabla 13 del resultado de la prueba chi cuadrado y cuya significación asintótica es $< 0,05$. (Sig 0,000), por lo tanto se rechaza la hipótesis H_0

Hipótesis 3:

- H_0 : Las variables Ingresos Agrupados y Nacionalidad son independientes.
- H_1 : Existe dependencia entre las variables.

% del total		NACIONALIDAD		
		OTRAS	ECUATORIAN A	Total
INGRESOS (Agrupada)	< 500	0,4%	34,9%	35,3%
	501 - 1500	0,6%	38,9%	39,4%
	1501 - 2500	0,2%	11,3%	11,5%
	>2501	0,3%	13,5%	13,8%
Total		1,4%	98,6%	100,0%

Tabla 14 Tabla de contingencia Ingresos - Nacionalidad

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Los resultados de la variable “Ingresos Agrupados” con la variable “Nacionalidad” (Tabla 14), nos indica que los clientes con “Nacionalidad Ecuatoriana” tienen ingresos menores a \$1500 en mayor proporción con el 73,8% en total.

	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	19,434 ^a	3	,000
Razón de verosimilitud	18,283	3	,000
Asociación lineal por lineal	17,458	1	,000
N de casos válidos	29140		

a. 0 casillas (0,0%) han esperado un recuento menor que 5.
El recuento mínimo esperado es 47,28.

Tabla 15 Prueba chi-cuadrado Ingreso - Nacionalidad

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

La prueba chi cuadrado y cuya significación asintótica es 0,00 nos indica que existe una relación asociativa entre las variables “Ingresos” y “Nacionalidad ya que el valor Sig es < 0,05, se rechaza la hipótesis nula H_0 (Tabla 15).

Hipótesis 4:

- H_0 : Las variables Ingresos Agrupados y Patrimonio son independientes.
- H_1 : Existe dependencia entre las variables.

Tabla cruzada INGRESOS (Agrupada)*PATRIMONIO (Agrupada)					
% del total		PATRIMONIO (Agrupada)			Total
		< 10000	10001 - 20000	> 20000	
INGRESOS (Agrupada)	< 500	29,1%	1,4%	4,8%	35,3%
	501 - 1500	26,8%	2,2%	10,3%	39,4%
	1501 - 2500	6,1%	0,9%	4,5%	11,5%
	>2501	4,4%	1,2%	8,3%	13,8%
Total		66,3%	5,7%	27,9%	100,0%

Tabla 16 Tabla de contingencia Ingreso - Patrimonio

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

De la variable “Ingresos Agrupados” con relación a la variable “Patrimonio Agrupado” como se muestra en la Tabla 16, podemos destacar que, el 29,1% de los clientes cuyo mayor porcentaje de patrimonio es menor a \$10000, percibe ingresos menores a \$500 y de los clientes que perciben ingresos en el rango de \$501 - \$1500 representan el 10,3% con un patrimonio menor a \$20000.

Pruebas de chi-cuadrado			
	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	19,434 ^a	3	,000
Razón de verosimilitud	18,283	3	,000
Asociación lineal por lineal	17,458	1	,000
N de casos válidos	29140		

a. 0 casillas (0,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 47,28.

Tabla 17 Prueba chi-cuadrado Ingreso - Patrimonio

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Como se muestra en la Tabla 17 de la prueba chi cuadrado y cuya significación asintótica es $< 0,05$ (Sig 0,00) podemos concluir que existe una asociación entre los ingresos de los clientes y el patrimonio de cada uno, por tal razón se rechaza la hipótesis H_0 .

Hipótesis 6:

- H_0 : Las variables Ingresos Agrupados y Sexo son independientes.
- H_1 : Existe dependencia entre las variables.

Tabla cruzada INGRESOS (Agrupada)*SEXO				
% del total		SEXO		Total
		M	F	
INGRESOS (Agrupada)	< 500	13,8%	21,5%	35,3%
	501 - 1500	15,8%	23,6%	39,4%
	1501 - 2500	5,3%	6,1%	11,5%
	>2501	7,7%	6,2%	13,8%
Total		42,5%	57,5%	100,0%

Tabla 18 Tabla de contingencia Ingreso - Sexo

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

El cruce de la variable “Ingresos Agrupados” con la variable “Sexo”, nos indica que el 23,6% de los ingresos de los clientes de sexo “Femenino”, representados por el 57,5% del total de la población, se encuentran en una escala de ingresos de \$501 y \$1500 (Tabla 18).

Pruebas de chi-cuadrado			
	Valor	df	Significación asintótica (bilateral)
Chi-cuadrado de Pearson	378,873 ^a	3	,000
Razón de verosimilitud	375,295	3	,000
Asociación lineal por lineal	325,012	1	,000
N de casos válidos	29140		

a. 0 casillas (0,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 1422,63.

Tabla 19 Prueba chi-cuadrado Ingreso - Sexo

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Mediante la prueba chi cuadrado visualizamos en la Tabla 19 que la significación asintótica es $< 0,05$ por lo tanto existe relación entre las variables “Ingresos” y “Sexo” (Sig 0,00), por ende se rechaza la hipótesis nula H_0 .

Cabe indicar que se obviaron las demás transacciones porque los resultados no fueron representativos.

Adicionalmente, se realizaron evaluaciones de regresión lineal y análisis discriminante como técnicas estadísticas multivariantes, sin embargo no se obtuvieron resultados significativos para el planteamiento de un modelo, por lo tanto se mostrará los resultados del análisis de regresión logística y árbol de decisiones.

4.2 REGRESIÓN LOGÍSTICA

Para nuestro caso es muy importante considerar que la regresión logística no asume los supuestos de la regresión lineal, particularmente el de normalidad, linealidad y homocedasticidad. Puede manejar cualquier tipo de relación no necesariamente lineal, ya que aplica una transformación logarítmica no lineal.

Como las variables explicativas pueden ser continuas o discretas (categóricas u ordinales), no necesitan ser independientes pero de serlo, la regresión da una solución más estable.

Se deben incluir en el modelo todas las variables significativas para obtener un modelo que ajuste bien los datos.

Modelo

- Variable dependiente: Producto Servicio

- Variable independiente: Ingresos

 Patrimonio

 Nacionalidad

 Actividad Económica

 Sexo

Se utilizó regresión logística para conocer la probabilidad de que un cliente realice una remesa (**Producto/Servicio**) dados los **Ingresos, Patrimonio, Nacionalidad, Actividad Económica, Sexo** al momento de solicitar la admisión a la institución financiera. Para la variable dependiente sólo se consideró dos categorías puesto que el porcentaje de clientes que piden un préstamo o hacen una inversión es apenas el 0.6%.

Como se mencionó en capítulos previos, la variable a explicar a través de las variables predictoras será la variable Producto/Servicio, la cual indica si un cliente escoge o no la apertura de cuenta bancaria cuando solicita la admisión. La codificación se muestra en las siguientes tablas.

Codificación de variable dependiente	
Valor original	Valor interno
CUENTA BANCARIA	0
REMESAS	1

Tabla 20 Codificación de variable Producto/Servicio

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Codificaciones de variables categóricas				
		Frecuencia	Codificación de parámetro	
			(1)	(2)
ACTIVIDAD_ECONOMICA	RIESGO BAJO	18761	1,000	,000
	RIEGO MEDIO	129	,000	1,000
	RIESGO ALTO	10250	,000	,000
NACIONALIDAD	OTRAS	412	1,000	
	ECUATORIANA	28728	,000	
SEXO	M	12397	1,000	
	F	16743	,000	

Tabla 21 Codificación Variables Independientes

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Cuando el modelo no incluye variables explicativas (modelo sólo con constante), el porcentaje de clasificación global correcta del modelo es 60,40%, clasificando correctamente (100%) a todos los casos en que los clientes abren una cuenta bancaria, pero incorrectamente a los que hacen un giro o remesa (Tabla 22)

Observado	PRODUCTO_SERVICIO	Pronosticado		Porcentaje correcto
		CUENTA BANCARIA	REMESAS	
Paso 0	CUENTA BANCARIA	17608	0	100,0
	REMESAS	11532	0	,0
Porcentaje global				60,4

a. La constante se incluye en el modelo.
b. El valor de corte es ,500

Tabla 22 Tabla de clasificación

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

	B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 0 Constante	-,423	,012	1248,168	1	,000	,655

Tabla 23 Variables en la ecuación - Modelo 1

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

La Tabla 23, el $\text{Exp}(\beta_0) = 0,655$, es menor a 1 lo que nos indica que la probabilidad de que el cliente escoja como producto/servicio el de giros o remesas no es mayor que la probabilidad de que el producto/servicio seleccionado por el cliente sea el abrir una cuenta bancaria.

		Variables en la ecuación					
		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	INGRESOS	,000	,000	49,928	1	,000	1,000
	PATRIMONIO	,000	,000	59,347	1	,000	1,000
	ACTIVIDAD_ECONOMICA			1721,449	2	,000	
	ACTIVIDAD_ECONOMICA (1)	1,189	,029	1703,281	1	,000	3,283
	ACTIVIDAD_ECONOMICA (2)	-,260	,271	,923	1	,337	,771
	SEXO(1)	-,971	,027	1331,736	1	,000	,379
	NACIONALIDAD(1)	-,151	,114	1,773	1	,183	,859
	Constante	-,744	,027	745,035	1	,000	,475

a. Variables especificadas en el paso 1: NACIONALIDAD.

Tabla 24 Variables incluidas en el modelo

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

La prueba de Wald contrasta la hipótesis de que un coeficiente aislado es distinto de 0, y sigue una distribución normal de media 0 y varianza 1.

Si el valor p es menor que 0.05, se rechaza H0 que afirma que el coeficiente asociado es igual a cero, por lo que se concluye que la variable es significativa en el modelo; caso contrario, si el valor p es mayor 0.05 se concluye que la variable no influye en variable de respuesta.

En la Tabla 24 se muestran los resultados de los coeficientes estimados, su error estándar, el estadístico de Wald y el valor p.

En nuestro modelo las variables que resultaron significativas son: Actividad Económica (1), Sexo (1), Ingresos, Patrimonio puesto que el valor p es cercano a 0 (Sig = 0), lo cual indica que es influyente en la explicación de la variable.

La oportunidad de que un cliente requiera realizar una remesa aumenta 3.283 veces cuando el cliente realiza una actividad con Riesgo Alto en relación a una actividad de Riesgo Bajo.

La oportunidad de que un hombre realice una remesa es 0,379 veces que si lo hiciera una mujer, es decir la probabilidad de que un hombre realice una remesa es menor a que lo haga una mujer.

Con los betas estimados en la Tabla 24 el modelo se lo plantearía de la siguiente manera:

$$\frac{\text{Producto/Servicio}}{1-\text{Producto/Servicio}} = 1/(1 - \exp (-0,744 + 1,189 \text{ Actividad Económica Riesgo Bajo} - 0,971 \text{ Sexo Femenino}))$$

A través de la ecuación es posible predecir si un cliente escogerá la apertura de cuenta bancaria o no. La variable dependiente es dicotómica y el punto de corte es 0,5, por esta razón los valores mayores iguales a 0,5 nos llevan a pronosticar que realizará giros/remesas, y las menores a 0,5 que realizarán la apertura una cuenta.

Prueba de Hosmer y Lemeshow			
Paso	Chi-cuadrado	gl	Sig.
1	156,473	8	,000

Tabla 25 Prueba de Hosmer y Lemeshow

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

La prueba de Hosmer y Lemeshow permite determinar la bondad de ajuste en los modelos de regresión logística, el procedimiento consiste en calcular para cada observación del conjunto de datos las probabilidades de la variable dependiente que predice el modelo, agruparlas y calcular, a partir de ellas, las frecuencias esperadas y compararlas con las observadas mediante la prueba chi-cuadrado.

La prueba da como resultado un valor $p = 0,000$, indicando que no existe evidencia suficiente para determinar que es o no un buen ajuste.

		PRODUCTO_SERVICIO = CUENTA BANCARIA		PRODUCTO_SERVICIO = REMESAS		Total
		Observado	Esperado	Observado	Esperado	
Paso 1	1	2481	2567,430	433	346,570	2914
	2	2394	2410,549	520	503,451	2914
	3	2002	2035,275	904	870,725	2906
	4	2166	1984,040	748	929,960	2914
	5	1906	1885,228	1009	1029,772	2915
	6	2028	1843,704	885	1069,296	2913
	7	1358	1397,741	1556	1516,259	2914
	8	1072	1180,084	1842	1733,916	2914
	9	1086	1137,512	1782	1730,488	2868
	10	1115	1166,438	1853	1801,562	2968

Tabla 26 Tabla de observados vs esperados

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

La Tabla 26 nos muestra los valores observados vs esperados, en la cual podemos visualizar que en todos los intervalos los valores predichos se aproximan a los observados.

Observado		Pronosticado		Porcentaje correcto
		PRODUCTO_SERVICIO CUENTA BANCARIA	REMESAS	
Paso 1	PRODUCTO_SERVICIO CUENTA BANCARIA	13355	4253	75,8
	REMESAS	4872	6660	57,8
Porcentaje global				68,7

a. El valor de corte es ,500

Tabla 27 Tabla de clasificación

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Esta tabla clasifica a los individuos de la población según la concordancia de los valores observados con los predichos o estimados del modelo. El porcentaje de falsos positivos es igual al 14,59%, mientras que el porcentaje de

falsos negativos es del 16,72%. El porcentaje global de valores correctamente predichos es del 68,7%.

4.3 ARBOLES DE DECISIÓN

Para el árbol de decisión dado el **Producto/Servicio** que seleccione el cliente se modelará con las variables **Ingresos, Patrimonio, Nacionalidad, Actividad Económica, Sexo** al momento de solicitar la admisión a la institución financiera, como se muestra en el resumen del modelo en la Tabla 28.

Resumen del modelo		
Especificaciones	Método de crecimiento	CHAID
	Variable dependiente	PRODUCTO_SERVICIO
	Variables independientes	ACTIVIDAD_ECONOMICA, SEXO, NACIONALIDAD, INGRESOS (Agrupada), PATRIMONIO (Agrupada)
	Validación	Ninguna
	Máxima profundidad del árbol	3
	Casos mínimos en nodo padre	100
	Casos mínimos en nodo hijo	50
	Resultados	Variables independientes incluidas
	Número de nodos	24
	Número de nodos terminales	15
	Profundidad	3

Tabla 28 Resumen del Modelo - Árbol de decisión

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Se obtiene el 68,9% de probabilidad de acierto, con un 31,1% de error de clasificación (Tabla 29 y 30).

Observado	Pronosticado		Porcentaje correcto
	CUENTA BANCARIA	REMESAS	
CUENTA BANCARIA	13949	3951	77,9%
REMESAS	5236	6384	54,9%
Porcentaje global	65,0%	35,0%	68,9%

Método de crecimiento: CHAID
Variable dependiente: PRODUCTO_SERVICIO

Tabla 29 Tabla de clasificación CHAID – Modelo 4

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Riesgo	
Estimación	Desv. Error
,311	,003

Método de crecimiento:
CHAID
Variable dependiente:
PRODUCTO_SERVICIO

Tabla 30 Tabla de Riesgo - Modelo 4

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

A continuación se visualizará el árbol de decisión dividido en cada Nodo, en la Ilustración 9, podemos indicar que la primera ramificación corresponde a la actividad económica, y del 64,4% que comprende la categoría “Riesgo Bajo”, el 35,2% a la categoría “Riesgo Alto” y el 0,5% a actividades de “Riesgo Medio”.

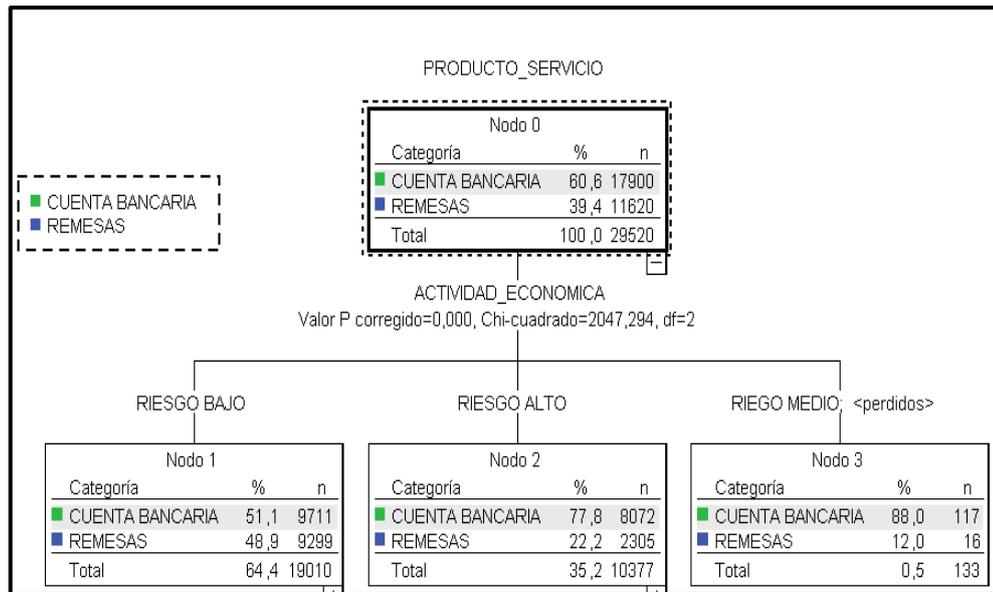


Ilustración 8 Árbol de decisión - Nodo 0

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

En el Nodo 1, como se muestra en la Ilustración 10, tenemos que del 64,4% de clasificados como actividades económicas de “Riesgo Bajo”, el 51,1% corresponde a cuenta bancaria y el 48,9% al producto giros/remesas, dando apertura a la siguiente ramificación que concierne a la variable “Sexo”, del que podemos indicar que 38,4% correspondiente a sexo femenino el 60,1% se inclina por el producto de giros/remesas, mientras que del 26% de los clientes de sexo masculino, el 67,6% se inclina por el producto cuenta bancaria.

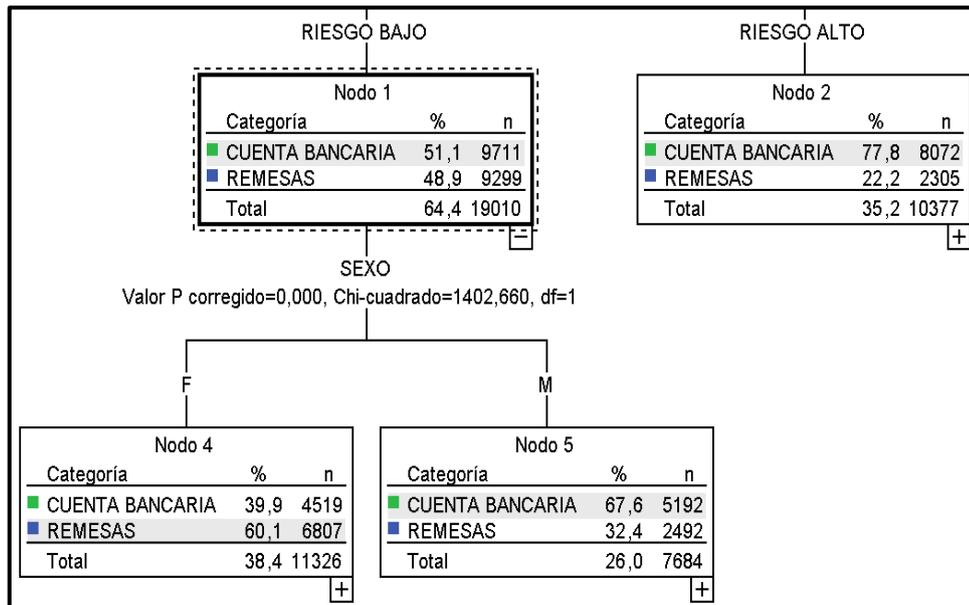


Ilustración 9 Árbol de decisión – Nodo 1

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Continuando de forma horizontal, nos encontramos con el Nodo 4, que concierne al sexo femenino y que de acuerdo a la Ilustración 11, la siguiente subdivisión son los Ingresos, donde del 31,6% de los clientes con ingresos mayores iguales a \$501 y menores iguales a \$1500, el 62,1% transacciona giros/remesas. Ilustración 11.

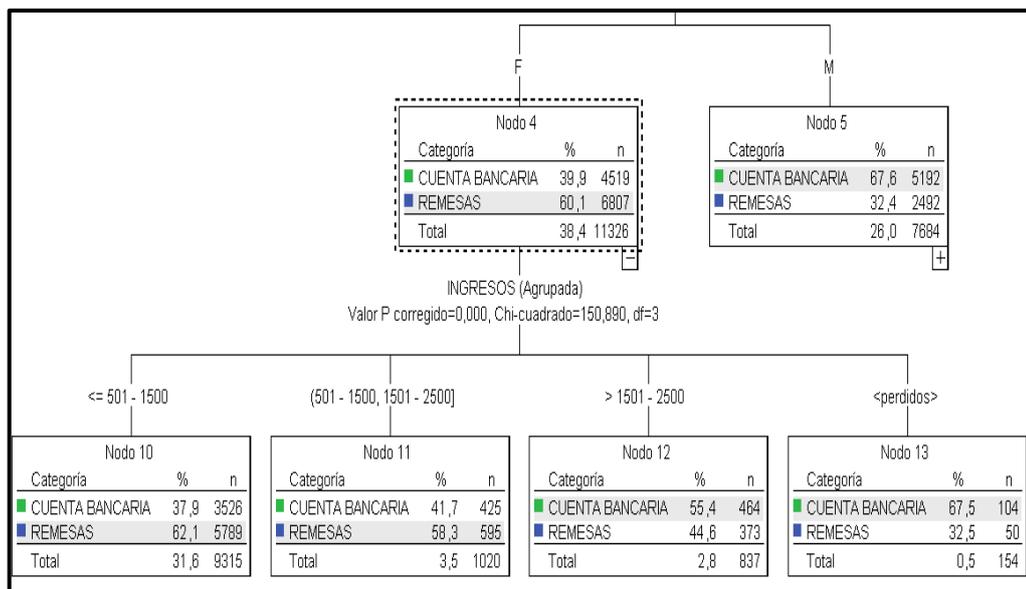


Ilustración 10 Árbol de decisión - Nodo 4

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

Mientras que del Nodo 5, correspondiente al sexo masculino, tenemos que el 22,4% tienen ingresos en el rango de \$1501 - \$2500, siendo el 65,6% de éstos quienes eligen apertura de cuenta bancaria como producto/servicio, como se visualiza en la Ilustración 12.

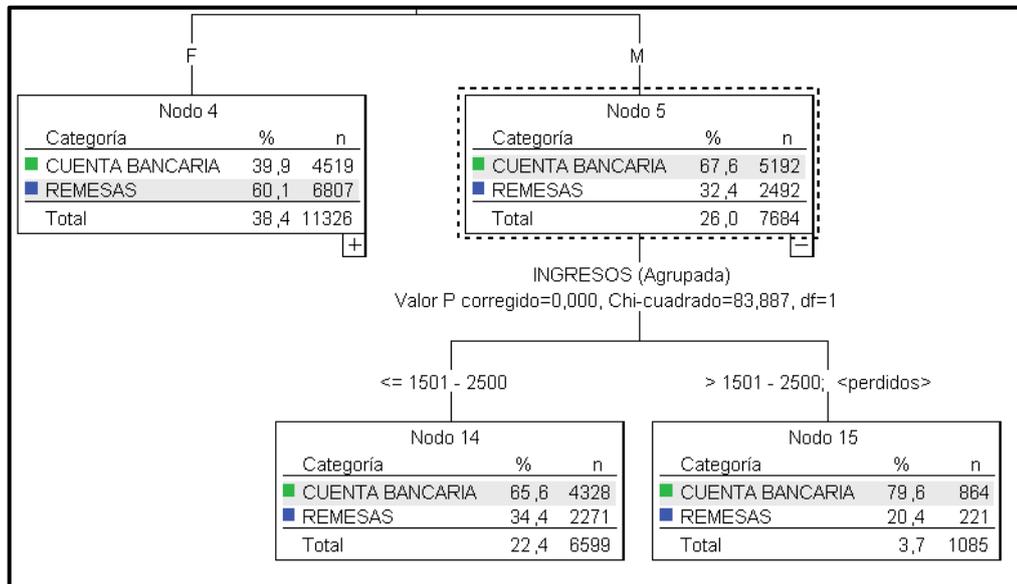


Ilustración 11 Árbol de decisión - Nodo 5

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

En el Nodo 2, perteneciente a la categoría “Riesgo Alto”, no existe una subdivisión en cuanto al sexo a éste nivel a diferencia del Nodo 1, pero si existe un mayor detalle en los rangos de ingresos que se encuentran los clientes.

La Ilustración 13 nos da la división, siendo que el 13 % de los ingresos se encuentran en el rango de \$501 - \$1500, donde el 71,6% clasifica a cuentas bancarias como producto.

Del Nodo 7, Ilustración 15, tenemos que en el rango de ingresos de \$501 - \$1500, el mayor porcentaje de cuenta bancaria es solicitado por el sexo masculino con un 78,8% perteneciente al 5,5% de la población total.

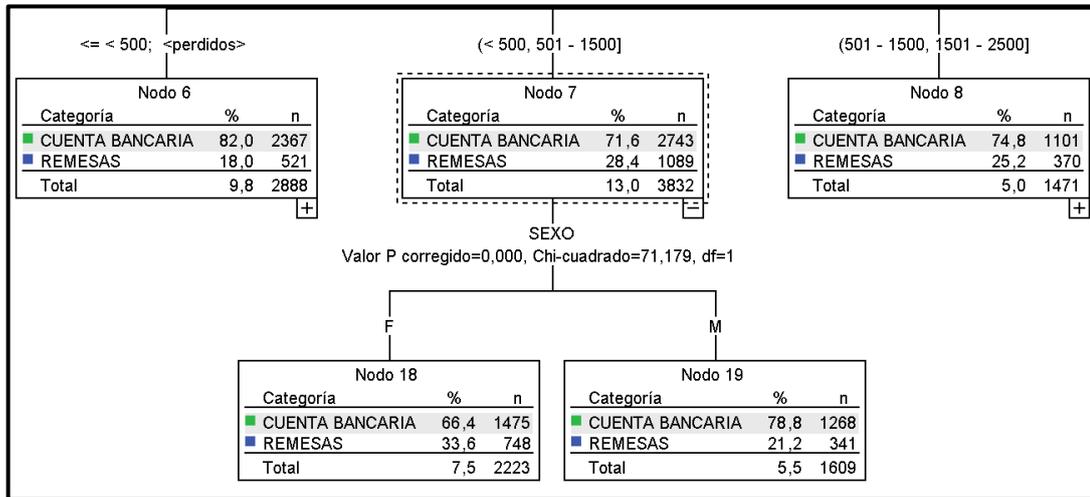


Ilustración 14 Árbol de decisión - Nodo 7

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

El Nodo 8, representa el 5% de los ingresos en las escalas de \$501-\$1500 y \$1500 - \$2500, donde el 2,6% pertenece al sexo femenino.

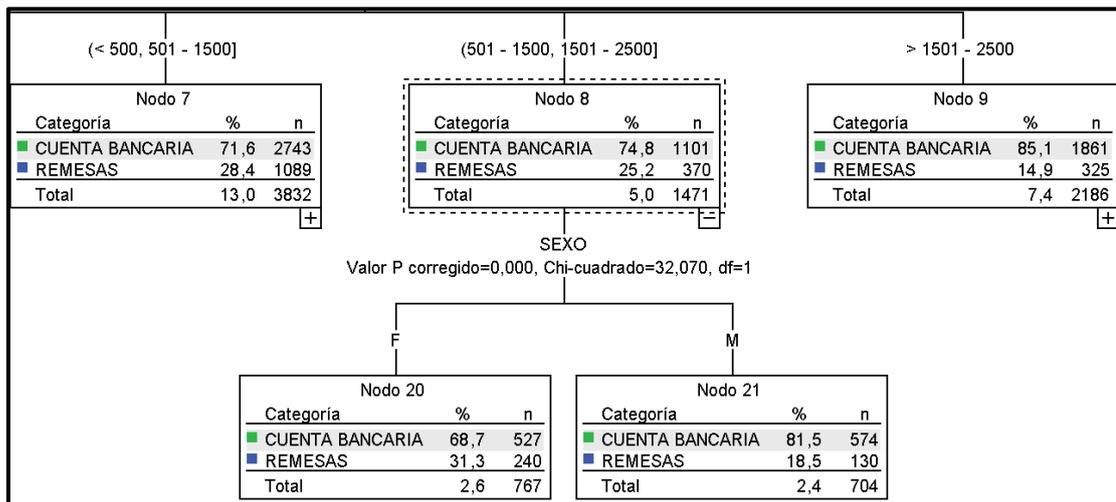


Ilustración 15 Árbol de decisión - Nodo 8

Fuente: Elaboración propia a partir de la base de datos de los clientes de la institución financiera.

CONCLUSIONES

- Dado el Producto/Servicio y considerando para un mejor perfilamiento de clientes las variables Ingresos, Patrimonio, Actividad Económica, Sexo y Nacionalidad declarado por el cliente, se tiene un acierto del 68,7% utilizando Regresión Logística y el 68,9% de probabilidad de acierto con el Modelo de Árbol de decisiones.
- Con el modelo de Regresión Logística se tiene que la oportunidad de que un cliente requiera realizar una remesa aumenta 3.283 veces cuando el cliente realiza una actividad económica categorizada como “Riesgo Bajo” en relación a una actividad de “Riesgo Alto”, dentro de las actividades más frecuentes en la categoría “Riesgo Bajo” tenemos: Actividades No Económicas, Empleados bajo relación de dependencia y Servicios comunales y sociales.
- La oportunidad de que un hombre realice una remesa es 0,379 veces que si lo hiciera una mujer, es decir la probabilidad de que un hombre realice una remesa es menor a que lo haga una mujer de acuerdo a los resultados del modelo de Regresión Logística.
- En el modelo de árbol de decisión, tenemos que el porcentaje de mayor clasificación corresponde a actividades económicas de “Riesgo Bajo” (64,4%), que de dicha categoría existe una leve preferencia en la selección del producto “Cuenta Bancaria” (51,1%) en referencia al producto “Giros/Remesas” (48,9%).
- En cuanto a la variable “Sexo” podemos indicar que las mujeres eligen en mayor proporción (38,4%) el producto “Giros/Remesas” (60,1%), a diferencia de los hombres (26%) que prefieren “Cuenta Bancaria” (67,6%)

- En el Nodo 2 en la categoría “Riesgo Alto”, no existe una subdivisión en cuanto a la variable “Sexo” a éste nivel a diferencia del Nodo 1, pero si existe un mayor detalle en los rangos de ingresos en el que se encuentran los clientes, lo que quiere decir que para la categoría “Riesgo Bajo” existen clientes con mayor influencia en la variable “Sexo”, que para la categoría “Riesgo Alto”, donde influye más la categoría ingresos.
- La variable “Patrimonio” únicamente es influyente para Actividades Económicas en categoría de “Riesgo Alto”, cuyos ingresos son mayores a \$1500 y menores a \$2500 y no influye la variable “Sexo”.
- De acuerdo a los problemas planteados en el presente proyecto, podemos concluir que las variables socioeconómicas son las que nos permite conocer el comportamiento habitual que debería cumplir como cliente posterior a su aceptación en la institución, mientras que las variables demográficas no son relevantes, ya que no mejoraron ninguno de los métodos en los que fueron probados.
- Tanto el método de regresión logística, como los árboles de decisión clasifican aceptablemente a todos los clientes (personas naturales), siendo ambas metodologías útiles.

RECOMENDACIONES

Utilizando como base el modelo obtenido en Árbol de decisión para el perfil de comportamiento de clientes recomendamos:

- Automatizar los parámetros por segmentos de clientes de tal manera que mes a mes se pueda monitorear la transaccionalidad de los clientes y cuyos movimientos no deberán superar el límite de ingresos declarado.

- En caso que el sistema genere una alerta basados en el perfil levantado y la transaccionalidad incurrida, se deberá solicitar actualización de datos del cliente para conocer las nuevas fuentes de ingreso o en su defecto remitirlo al área pertinente para el seguimiento respectivo y se tomen decisiones oportunas en caso de que la investigación de como resultado actividades ilícitas.

BIBLIOGRAFÍA

- Barnett, N. (1969). Beyond Market Segmentation. *Harvard Business Review*.
- Bowen. (2014). *Social responsibilities of the businessman*. New York: Harper.
- Calantone, R., & Sawyer, A. (1978). *The Stability of Benefit Segments*. Journal of Marketing Research.
- Cuadras, C. (2014). *Nuevos métodos de Análisis Multivariantes*. Barcelona: CMC Edition.
- Frank, R., Massy, W., & Wind, Y. (1972). *Market Segmentation*. Prentice Hall.
- Grover, R., & Srinivasan, V. (1987). *A Simultaneous Approach to Market Segmentation and Market Structuring*. American Marketing Association.
- Grupo Acción Financiera Internacional. (2012, Febrero 16). *fatf-gafi.org*. Retrieved from fatf-gafi.org: [http://www.fatf-gafi.org/publications/?hf=10&b=0&s=desc\(fatf_releasedate\)](http://www.fatf-gafi.org/publications/?hf=10&b=0&s=desc(fatf_releasedate))
- Hair, J., Anderson, R., Tatham, R., & Black, W. (1999). *Análisis Multivariante*. Madrid: Prentice Hall.
- Kotler, P. (1988). *Dirección de mercadotecnia: Análisis, Planeación y Control*. Mexico: Editorial Diana.
- Mahajan, V., & Jain, A. (1978). *An approach to normative segmentation*. Journal of Marketing.
- Pride, W., & Ferrell, O. (1982). *Marketing: Decisiones y Conceptos Básicos*. Mexico: Interamericana.
- Salgado Carrión, J. (1981). *La práctica del muestreo*. Madrid: Ortega Martínez.
- Sampieri. (2012). *Metodología de la Investigación Científica*. España: UCES.
- Superintendencia de Bancos y Seguros. (2008). Normas generales para las instituciones del sistema financiero. In J. Bancaria, *Del Control Interno* (pp. 1-27). Quito: Junta Bancaria.

Superintendencia de Bancos y Seguros. (2013). Normas Generales para las Instituciones del Sistema Financiero. In J. Bancaria, *Del Control Interno* (pp. 818-852). Quito: Junta Bancaria.

Uriel, E. (1995). *Análisis de datos. Series temporales y Análisis multivariante*. Madrid: AC.