



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL
Facultad de Ciencias Naturales y Matemáticas
Ingeniería en Estadística Informática

**“DESARROLLO DE UN MODELO ESTADÍSTICO PARA
MEDIR LA POSIBILIDAD DE MOROSIDAD DE CLIENTES EN
UNA EMPRESA DE RETAIL”**

PROYECTO INTEGRADOR

**Previo a la obtención del Título de:
INGENIERO EN ESTADÍSTICA INFORMÁTICA**

**Presentado por:
Juan Rafael Campoverde López**

Guayaquil – Ecuador

2015

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL FACULTAD DE
CIENCIAS NATURALES Y MATEMÁTICAS

INGENIERÍA EN ESTADÍSTICA E INFORMÁTICA

INFORME DEL PROFESOR DE LA MATERIA INTEGRADORA

Habiendo sido nombrado PROFESOR DE LA MATERIA
INTEGRADORA del señor,

**JUAN RAFAEL
CAMPOVERDE LÓPEZ**

Con el tema "DESARROLLO DE UN MODELO ESTADÍSTICO
PARA MEDIR LA POSIBILIDAD DE MOROSIDAD DE CLIENTES
EN UNA EMPRESA DE RETAIL", previa a la obtención del título de
INGENIERO EN ESTADÍSTICA INFORMÁTICA, me permito
informar que he leído el contenido del proyecto integrador, luego de
lo cual indico que estoy de acuerdo en que el mismo se lo ha
desarrollado conforme a los lineamientos de la Unidad de Titulación
Especial de la ESPOL.



ING. SONNIA PAOLA REYES RAMOS
PROFESOR DE LA MATERIA INTEGRADORA

Guayaquil, 18 de Febrero del 2016

AGRADECIMIENTO

A Dios, por permitirme culminar con esta etapa de mi vida cruzando obstáculos;

A mis padres, abuelos y hermano por estar a mi lado en todo momento, pero en especial a mi madre Msc. Jacqueline López por su apoyo incondicional y palabras de aliento, junto a mi abuelo Segundo López son ejemplo de superación en la vida;

A mis amigos, con los que compartí una etapa de mi vida en la ESPOL, en especial al Ing. Carlos Rodríguez que demostró ser un compañero y amigo;

A mi tutora Ing. Paola Reyes, por la paciencia, palabras de aliento y compromiso que tuvo para poder culminar con el proyecto.

DEDICATORIA

Este trabajo se lo dedico a Dios, por haberme ayudado a cruzar las adversidades, superarme a mi mismo y culminar con este objetivo pendiente en mi vida.

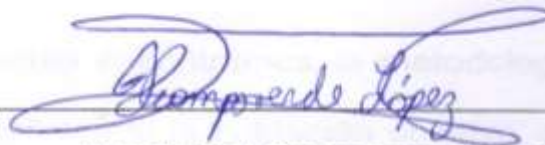
A mis padres, abuelos, hermano y familiares que son pilar importante en mi vida y siempre estarán a mi lado en todo momento; siendo ejemplos que existen retos difíciles pero no imposibles de superar en la vida.

A mis profesores por los conocimientos brindados en las aulas de clases, en especial al M.Sc. Gaudencio Zurita, Ph.D. Francisco Vera, Ing. Sonnia Reyes y Máster Eva María Mera por ser parte fundamental en mi educación con sus conocimientos y experiencias de la vida.

DECLARACIÓN EXPRESA

"La responsabilidad del contenido de este Trabajo Final de Graduación, nos corresponde exclusivamente; y el patrimonio intelectual del mismo a la Escuela Superior Politécnica del Litoral".

(Reglamento de Graduación de la ESPOL)



Juan Rafael Campoverde López

RESUMEN

El presente trabajo nos permite determinar mediante un modelo estadístico la probabilidad de que un cliente sea moroso en la empresa XYZ S.A. La misma es una empresa de retail que se dedica a la venta de productos sobre todo del hogar y nos proporcionó los datos para trabajar. Este estudio tiene como fin determinar el comportamiento de la variable de interés, la misma que será explicada por un conjunto de variables que contienen características personales y socio demográficas de los clientes de la empresa.

En el primer capítulo, tenemos el planteamiento del problema de nuestro estudio, en segundo lugar la justificación del presente trabajo, donde con el modelo estadístico permitirá a la empresa tomar decisiones importantes con los clientes futuros morosos y disminuir la probabilidad de cartera castigada. Y al final planteamos el objetivo general y los objetivos específicos del presente estudio.

En el segundo capítulo encontramos la metodología a utilizar en el presente trabajo. Iniciamos definiendo la población objetivo, que en nuestro caso son los clientes que fueron aperturados en la empresa hasta Junio 2015 en la empresa

XYZ S.A. Luego se procede a describir cada una de las técnicas que utilizaremos en el análisis de datos univariado, inferencial y multivariado.

En el tercer capítulo, tenemos el análisis de datos univariado, inferencial y multivariado. En la parte univariada e inferencial analizaremos las variables a utilizar en el modelo de regresión logística binomial, utilizando tablas, diagramas de barras y pasteles en la parte descriptiva y tablas de contingencia en la parte inferencial para determinar independencia entre las variables. Por tener variables categóricas solo utilizaremos estas técnicas. Y en la parte especializada como técnica multivariada utilizaremos la regresión logística binomial al poseer una variable dicotómica como respuesta.

Para terminar con nuestro estudio, en la última sección encontramos las conclusiones y recomendaciones de mayor importancia que obtuvimos con los resultados del presente trabajo.

Tabla de contenido

AGRADECIMIENTO	ii
DEDICATORIA	iv
RESUMEN	vi
CAPÍTULO I	1
1. Descripción del Problema.....	1
1.1. Antecedentes.....	1
1.2. Definición del problema.....	5
1.3. Justificación del Problema	6
1.4. OBJETIVOS	8
CAPÍTULO II	10
2. Metodología	10
2.1. INTRODUCCIÓN.....	10
2.2. Análisis Estadístico Univariado	11
2.3. Análisis Estadístico Inferencial.....	16
2.4. Regresión Logística Binaria	18
CAPÍTULO III	25
3. Análisis De Datos.....	25
3.1. Análisis Univariado.....	25
3.2. Análisis Inferencial	37
3.3. Regresión Logística.....	53
CONCLUSIONES Y RECOMENDACIONES	64

Índice de Figuras

FIGURA 2.1: DISTRIBUCIÓN LOGÍSTICA	20
---	----

Índice de Cuadros

CUADRO 3.1: TABLA Y DIAGRAMA DE PASTEL DE LOS CLIENTES CON RC	26
CUADRO 3.2: TABLA Y DIAGRAMA DE PASTEL DE LOS CLIENTES VIP	27
CUADRO 3.3: TABLA Y DIAGRAMA DE PASTEL DE LOS CLIENTES SÚPER CRÉDITO	28
CUADRO 3.4: TABLA Y DIAGRAMA DE PASTEL DE LOS CLIENTES POR SEXO	29
CUADRO 3.5: TABLA Y DIAGRAMA DE PASTEL DE LOS CLIENTES MOROSOS	31
CUADRO 3.6: TABLA Y DIAGRAMA DE BARRAS DE LAS PROVINCIAS DE RESIDENCIA DE LOS CLIENTES	32
CUADRO 3.7: TABLA Y DIAGRAMA DE BARRAS DE LA CIUDAD DE RESIDENCIA DE LOS CLIENTES	33
CUADRO 3.8: TABLA Y DIAGRAMA DE BARRAS DE LA REGIÓN DE RESIDENCIA DE LOS CLIENTES	34
CUADRO 3.9: TABLA Y DIAGRAMA DE BARRAS DEL ESTADO CIVIL DE LOS CLIENTES ..	36
CUADRO 3.10: TABLA DE CONTINGENCIA Y DIAGRAMA DE PASTEL: MOROSO VS EMITIÓ RC	38
CUADRO 3.11: TABLA DE CONTINGENCIA Y DIAGRAMA DE PASTEL: MOROSO VS VIP ...40	
CUADRO 3.12: TABLA DE CONTINGENCIA Y DIAGRAMA DE PASTEL: MOROSO VS SÚPER CRÉDITO	42
CUADRO 3.13: TABLA DE CONTINGENCIA Y DIAGRAMA DE PASTEL: MOROSO VS SEXO	44
CUADRO 3.14: TABLA DE CONTINGENCIA Y VALOR P: MOROSO VS PROVINCIA	45
CUADRO 3.15: TABLA DE CONTINGENCIA Y DIAGRAMA DE BARRAS: MOROSO VS CIUDAD	48
CUADRO 3.16: TABLA DE CONTINGENCIA Y VALOR P: MOROSO VS REGIÓN	50
CUADRO 3.17: TABLA DE CONTINGENCIA Y VALOR P: MOROSO VS ESTADO CIVIL	51

Índice de Tablas

TABLA 3.1: CODIFICACIÓN DE LA VARIABLE DEPENDIENTE DEL MODELO DE REGRESIÓN LOGÍSTICA BINOMIAL	53
TABLA 3.2: CODIFICACIÓN DE LAS VARIABLES INDEPENDIENTES DEL MODELO DE REGRESIÓN LOGÍSTICA BINOMIAL	54
TABLA 3.3: RESULTADOS DEL PRIMER MODELO DE REGRESIÓN LOGÍSTICA BINOMIAL	55
TABLA 3.4: RESULTADOS DEL SEGUNDO MODELO DE REGRESIÓN LOGÍSTICA BINOMIAL	56
TABLA 3.5: EXPLICACIÓN DE LOS SIGNOS DE LOS COEFICIENTES ESTIMADOS	59
TABLA 3.6: ODDS DEL MODELO DE REGRESIÓN LOGÍSTICA	60
TABLA 3.7: TABLA DE CLASIFICACIÓN	62

CAPÍTULO I

1. Descripción del Problema

1.1. Antecedentes

El otorgamiento del crédito de parte de muchas empresas comerciales y entidades financieras se ha convertido en una valiosa herramienta para poder incrementar las ventas mediante la captación de clientes pero a su vez puede resultar riesgoso por la falta de recuperación del crédito otorgado generando pérdidas para la empresa.

La empresa XYZ es una empresa comercial muy conocida a nivel nacional, por tal motivo ofrece a sus clientes el crédito directo para que puedan realizar sus compras, pagar de forma cómoda y así la empresa pueda aumentar el volumen de sus ventas pero también aumenta el riesgo crediticio y expone a la empresa. Por tal motivo se debe investigar el riesgo del cliente mediante la implementación

de un modelo matemático que permita analizarlos mediante sus perfiles demográficos y no otorgarle el crédito a quienes sean riesgosos para la empresa XYZ S.A.

Para las empresas comerciales no hay nada mejor que cobrar al contado el servicio o bien proporcionado al cliente, pero en la actualidad la mayor parte de veces esta no es la primera opción por parte del cliente. Por este motivo las empresas comerciales se ven en la necesidad de implementar una herramienta valiosa como es la obtención del crédito directo por parte del cliente, para que de esta forma captar un gran mercado.

Esto hace que la empresa XYZ S.A. con el objetivo de incrementar el volumen de sus ventas y proporcionar un plazo adecuado y cómodo de pago a sus clientes para que estos se sientan atraídos a la empresa, lanza su crédito directo en el año 2003, pero esto conlleva a que la empresa XYZ S.A. no sea la excepción del problema que se adquiere al otorgar crédito a los clientes.

En este capítulo se determinarán y analizarán los objetivos de la investigación, tanto el objetivo específico como los generales. También se planteará el problema que se obtiene al otorgar el crédito directo a los clientes de la empresa XYZ S.A. Se analizarán los perfiles de los clientes que hacen que aumente el riesgo crediticio, para que sean de gran utilidad al área de crédito y comercial al momento de captar un cliente y que su morosidad no sea grande.

1.1.1 DESCRIPCIÓN DE LA EMPRESA

La empresa XYZ S.A. se dedica a la comercialización de productos para el hogar desde hace más de 50 años con 30 almacenes aproximadamente a nivel nacional, distribuidos en la Costa y Sierra.

La empresa XYZ S.A. se ha visto en la necesidad en los últimos años de expandir sus líneas de productos para cubrir las necesidades y expectativas de sus clientes y tener como competir contra empresas similares. A continuación se detallan alguna de las líneas más importantes de la empresa:

- Línea de bebés
- Línea de computación
- Línea de decoración para el hogar
- Línea de artículos deportivos
- Línea de Dormitorio
- Línea de Electrodomésticos
- Línea de Juguetería
- Línea de muebles
- Línea de jardinería

- Línea de productos plásticos
- Línea de Vestir

Para las ventas de sus productos, la empresa XYZ S.A. posee 3 canales los cuales son: **Ventas por almacén** en las que el cliente se acerca al almacén a comprar el producto que desea, **Fonocompras** que son las compras que se realizan por medio del teléfono donde el cliente llama a la empresa a solicitar que le envíen uno ó más productos que ha observado en la revista; y por último el **Canal web** que son las ventas que se realizan por medio del internet donde el cliente ingresa a la página de la empresa y selecciona los artículos que va a comprar.

1.1.2 OBJETIVOS DE LA EMPRESA

Objetivo General

El objetivo general es servir a sus clientes, satisfaciendo sus requerimientos y asegurando una rentabilidad para nuestros accionistas acorde con el valor patrimonial de la empresa.

Objetivos específicos de la empresa

- Fidelizar a sus clientes mediante la venta de productos de buena calidad y a bajo costo en comparación a la competencia.
- Expandirse a nivel nacional mediante la apertura de nuevos almacenes.
- Incrementar sus ventas vía internet, vía telefónica y a través de su tarjeta de crédito propia; a fin de posicionarse como la mejor y primera opción de compra para sus clientes entre las empresas de su tipo.

1.2. Definición del problema

El crédito es una estrategia de los almacenes comerciales que se creó con el objetivo de captar más clientes y poder aumentar sus ventas. Hoy en día la técnica del crédito directo en los locales comerciales se ha masificado provocando que aumenten sus carteras de clientes. Estos créditos son otorgados promocionando una serie de beneficios a sus futuros clientes.

En vista de los nuevos clientes que son captados por el crédito directo para que puedan realizar sus compras, se establece la importancia de estudiarlos

profundamente porque si bien es cierto por un lado se tiene que al captar más clientes aumentan las ventas de la empresa, por otro lado algunos de estos clientes pueden provocar aumento del riesgo crediticio al aumentar su mora no pagando en el tiempo establecido sus cuotas.

Con el fin de no captar clientes que afecten el riesgo crediticio, los índices de morosidad ni la cartera, para que no se vea reflejado en los intereses de la empresa se deben tener en cuenta variables demográficas de los clientes que pueden ser de gran interés y un continuo crecimiento de la empresa tanto en los clientes como en las ventas y la disminución de la mora.

1.3. Justificación del Problema

El problema indicado anteriormente afecta directamente a la empresa XYZ S.A. porque actualmente para evaluar a un futuro cliente se basa en las políticas de crédito definidas por el departamento de crédito y cobranzas y no se consideran otros factores como las variables socio demográficas del cliente lo que permite la captación de un alto número de clientes con falta de pago lo que afecta y

aumenta el riesgo de la empresa XYZ S.A. afectando los índices de mora y riesgo.

Por esto es importante mediante este análisis determinar las características socio demográficas de los clientes morosos, para mediante las características de estos clientes que representan un riesgo para la empresa construir un modelo matemático que ayude al departamento de crédito y cobranza y a la empresa a mejorar los indicadores y de esta forma el crédito directo resulte un buen producto para la empresa XYZ S.A.

Con esta información la empresa podrá tomar decisiones que minimicen los índices de riesgo y morosidad y no se vea afectado con malos clientes.

1.4. OBJETIVOS

1.4.1 Objetivo General

Determinar mediante un modelo matemático las características sociales y demográficas que influyen en los clientes que caen en morosidad basándose en la cartera de clientes de la empresa XYZ S.A.

1.4.2 Objetivos Específicos

- Diagnosticar las características sociales y demográficas en los clientes que caen en morosidad de la empresa XYZ S.A. a partir de Julio del año 2014.
- Valorar los fundamentos teóricos que sustentan las características sociales y demográficas en los clientes que caen en morosidad de la empresa XYZ S.A. a partir de Julio del año 2014.
- Diseñar un modelo matemático que determine las características sociales y demográficas en los clientes que caen en morosidad en la empresa XYZ S.A. a partir de Julio del año 2014.

- Validar el modelo matemático que determine las características sociales y demográficas en los clientes que caen en morosidad en la empresa XYZ S.A. a partir de Julio del año 2014.
- Determinar la cantidad de clientes con crédito directo según la provincia de residencia en el Ecuador.
- Conocer la proporción de clientes con crédito según el género de la persona.
- Determinar la cantidad de clientes con mayor morosidad en los últimos 12 meses.
- Conocer la región del Ecuador donde se encuentran la mayor cantidad de clientes morosos.

CAPÍTULO II

2. Metodología

2.1. INTRODUCCIÓN

Una vez planteado el problema del presente trabajo, la metodología a plantear mediante una función basada en variables tales como: sexo, región, provincia, ciudad para determinar la probabilidad que un cliente sea moroso.

La empresa XYZ S.A. nos proporcionó los datos para el presente trabajo. Donde la población objetivo serán los clientes aperturados hasta el mes de Junio del 2014 y que tengan saldo.

Para tener una mejor visión sobre los clientes morosos en la empresa, utilizaremos las técnicas estadísticas, para analizar cada una de las variables y

realizar una inferencia estadística. Las técnicas a realizar son: Análisis estadístico univariado, inferencial y multivariado(Regresión Logística).

2.2. Análisis Estadístico Univariado

En la presente sección, procederemos a realizar un análisis de forma individual de cada una de las variables mediante cuadros estadísticos, diagramas circulares y de barras.

2.2.1 Diagrama de barras

El diagrama de barras es una herramienta muy útil en el análisis univariado que nos permite representar gráficamente un conjunto de datos para visualizar la distribución de variables. Esta técnica univariada la utilizaremos con las variables discretas: Región, provincia, ciudad de residencia y estado civil del cliente. De esta manera nos permite interpretar mejor el comportamiento de las variables de forma individual.

2.2.2 Diagrama de pastel

El diagrama de pastel es una herramienta muy útil al igual que el diagrama de barras en el análisis univariado. Lo utilizaremos para observar mejor la distribución de variables. En este caso dicotómicas por lo que se diferencian del diagrama de barras, es decir lo utilizaremos con las variables: Cliente moroso, emitió reporte de crédito, club VIP, súper crédito y sexo.

De esta manera entenderemos de forma visual el comportamiento de las variables distribuidas en porcentajes para una mejor interpretación.

2.2.3 Codificación de las variables

A continuación se describirán las 12 variables a utilizar, entre las cuales tenemos la variable dependiente, y qué método descriptivo vamos a utilizar.

Variable 1: Cliente moroso. (Y)

Esta será la variable dependiente, por lo tanto es de tipo dicotómica y hace referencia si se declara en mora o no al cliente dependiendo de sus días de vencido.

Variable 2: Emitió reporte de crédito. (X_1)

Esta variable indica si al cliente se le generó o no el reporte de crédito para que esté enterado de sus valores a pagar.

Variable 3: Club VIP. (X_2)

En la variable club VIP podremos conocer que clientes pertenecen al club VIP, estos clientes tienen beneficios exclusivos a diferencia de los clientes que no son VIP.

Variable 4: Súper crédito. (X_3)

La variable súper crédito muestra si el cliente es o no cliente súper crédito, el súper crédito consiste en un extra cupo que se le otorga a los clientes para comprar electrodomésticos, muebles, celulares y juguetes.

Variable 5: Sexo. (X_4)

Esta variable muestra si el cliente es de sexo masculino o femenino.

Variable 6: Región. (X_5)

La variable región se refiere a la región del país (Costa, Sierra y Otros) donde el cliente se encuentra viviendo. Con esta variable se determinará que región posee la mayoría de los clientes morosos.

Variable 7: Provincia. (X_6)

De igual forma esta variable nos indica la provincia de residencia de los clientes de la empresa XYZ S.A. Las provincias son: Guayas, Pichincha y Otros.

Variable 8: Ciudad. (X_7)

Así mismo la variable ciudad indica la ciudad donde viven los clientes. Las ciudades se agruparon en 3 categorías: Guayaquil (incluye Guayaquil y Durán), Quito y Otros.

Variable 9: Estado civil. (X_8)

Esta variable se refiere al estado civil que tienen los clientes de la empresa. Los estados son: soltero, casado, unión libre, divorciado y viudo.

Variable 10: Edad. (X_9)

La variable Edad indica la edad medida en años del cliente de la empresa XYZ S.A. Esta edad se la considera a partir de la fecha de apertura del crédito del cliente.

2.3. Análisis Estadístico Inferencial

En este capítulo se realizará la inferencia estadística, la técnica bivariada a utilizar entre las variables será, Tablas de Contingencia. Por medio de esta técnica bivariada analizaremos la independencia o no de la variable dependiente con las variables independientes escogidas para el modelo.

2.3.1 Tablas de Contingencia

Se procede a utilizar tablas de contingencia en el presente proyecto porque el mismo consta con variables categóricas; y una de las técnicas bivariadas para realizar inferencia estadística con variables categóricas es, Tablas de

Contingencia, la misma que nos permitirá establecer si existe independencia entre dos variables categóricas.

Las variables categóricas como lo indica su nombre clasifica a cada individuo en una categoría o nivel de las variables.

El contraste de hipótesis para nuestro trabajo sería:

H₀: La variable y es independiente de la variable x

vs

H₁: La variable y no es independiente de la variable x

Este será el contraste de hipótesis considerado en nuestro proyecto en el análisis bivariado para determinar independencia entre la variable dependiente Y con cada una de las variables explicativas o independientes X.

Mediante las tablas de contingencia se determinará si existe evidencia estadística para rechazar o no la hipótesis nula en favor de la hipótesis alterna.

Se podrá obtener el modelo más óptimo, probando varios modelos que expliquen la variable dependiente. Todo esto se determinará contrastando la variable dependiente Y contra cada una de las variables independientes X, para tener conocimiento si existe dependencia entre la variable a ser explicada con las demás variables independientes del modelo.

2.4. Regresión Logística Binaria

Al tener un modelo matemático en el cual una variable de respuesta o dependiente toma dos posibles valores (sí o no, morir o no, enfermar o no) lo que se denomina variable dicotómica, y necesitamos saber si esta variable se ve afectada por un conjunto de variables independientes, lo mejor es utilizar el modelo de regresión logística binaria. En caso de tener la variable dependiente más de 2 eventos de respuesta pasaría a ser una regresión logística multivariada.

La regresión logística puede resultarnos de gran utilidad para:

- Dado los valores de las variables independientes, estimar la probabilidad de que se presente el evento de interés.
- Podemos evaluar la influencia que cada variable independiente tiene sobre la respuesta, en forma de OR. Una OR mayor que 1 indica aumento en la probabilidad del evento y OR menor que 1 indica disminución.

Para la construcción del modelo de regresión logística necesitamos un conjunto de variables predictoras como en la regresión lineal múltiple; y una variable de respuesta dicotómica, diferente a la regresión lineal múltiple donde la variable de respuesta es numérica (Paz, 2009).

La ecuación de la distribución logística es:

$$P(y = 1|x) = \frac{\exp(b_0 + \sum_{i=1}^n b_i x_i)}{1 + \exp(b_0 + \sum_{i=1}^n b_i x_i)}$$

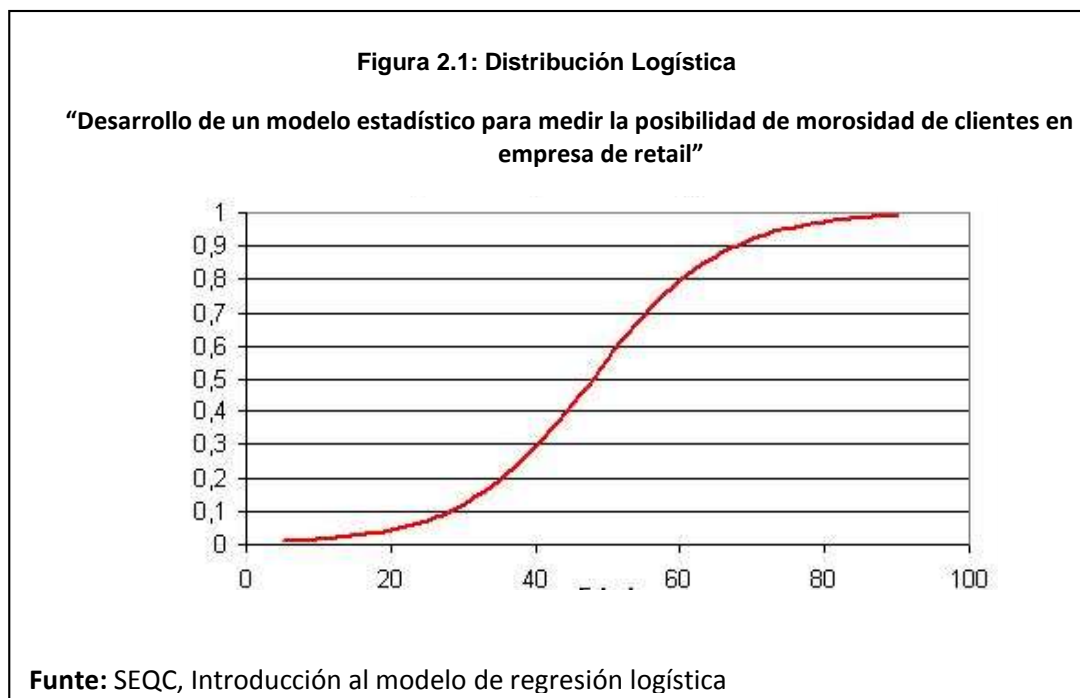
Donde:

$P(y = 1|x)$: es la probabilidad de que y tome el valor 1, en presencia de las co-variables x .

x : es un conjunto de n covariables que forman parte del modelo

b_0 : es la constante del modelo o término independiente

b_i : los coeficientes de las co-variables.



En la figura 2.1 podemos observar la representación gráfica de la distribución logística en la cual notamos que no sigue una forma recta sino una forma sigmoidea.

Si dividimos la expresión anterior para su complemento, obtenemos su ODDS, es decir la probabilidad de que ocurra el evento entre la probabilidad de que no ocurra (Lopez, 2012), obtenemos la siguiente expresión:

$$\frac{P(y = 1|x)}{1 - P(y = 1|x)} = \exp\left(b_0 + \sum_{i=1}^n b_i x_i\right)$$

Si ahora procedemos a sacar el logaritmo natural de la ecuación, se obtiene una ecuación lineal de manejo fácil, idéntico al modelo de regresión lineal.

$$\log\left(\frac{P(y = 1|x)}{1 - P(y = 1|x)}\right) = b_0 + \sum_{i=1}^n b_i x_i$$

Simplificando,

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_i x_i$$

2.4.1 Coeficientes del Modelo

En la regresión logística para la estimación de los coeficientes del modelo usaremos el método de máxima verosimilitud, que consiste en hacer máxima la probabilidad de encontrar los valores de Y obtenidos por los datos de nuestra muestra.

Para las estimaciones de máximo – verosimilias en la regresión logística se emplean métodos iterativos como el de Newton – Raphson, de donde también obtenemos los errores estándar (Valeriano, 2012).

Para comprobar la significancia de los coeficientes estimados en el modelo se emplearán los métodos de: el estadístico de Wald y el estadístico G de razón de verosimilitud.

2.4.2 Estadístico de Wald

Este estadístico contrasta la hipótesis de que un coeficiente aislado es distinto de 0, siguiendo una distribución normal con media 0 y varianza 1. La obtención de significación indica que el coeficiente es distinto a 0 y merece conservarlo en el modelo (Paz, 2009). Y se define como (Abdelrahman, 2010):

$$W = \frac{\hat{\beta}_j^2}{\text{var}(\hat{\beta}_j^2)}$$

2.4.3 Estadístico G de Razón de Verosimilitud (Paz, 2009)

Aquí tratamos de contrastar cada modelo que surge de ir aislando covariables del modelo con el modelo completo. El estadístico sigue una distribución Ji-Cuadrado con 1 grado de libertad(no se asume normalidad). La ausencia de significación

implica que el modelo con ausencia de la covariable eliminada no empeora en comparación con el modelo completo; es decir da los mismo su presencia o ausencia, pero por el principio de parsimonia (obtención del modelo más reducido), esta covariable debe ser eliminada del modelo porque no aporta al modelo. Al no suponer distribución esta prueba es la de mayor aportación.

CAPÍTULO III

3. Análisis De Datos

3.1. Análisis Univariado

En el presente capítulo mostraremos el respectivo análisis descriptivo de cada una de las variables a utilizar en el modelo; es decir de la variable dependiente como de las variables independientes.

3.1.1 Emitió Reporte de Crédito

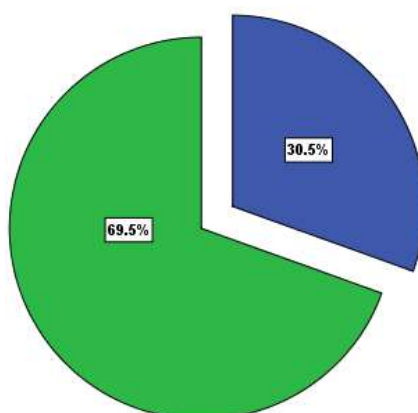
En el Cuadro 3.1 podemos observar que a 69922 clientes se les entregó el reporte de crédito, que equivale al 69,5% mientras que al 30,5% de los clientes no se les entregó el reporte de crédito en el mes de Julio, por lo que la mayoría de los clientes si recibieron su reporte de crédito.

Cuadro 3.1: Tabla y diagrama de pastel de los clientes con RC

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Tabla de frecuencias			
Emitió RC	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada
NO	30710	30,5	30,5
SI	69922	69,5	100
Total	100632	100,0	

■ NO
■ SI



Elaborado por: *J. Campoverde*, Facultad de Ciencias Naturales y Matemáticas

3.1.2 Clientes Club VIP

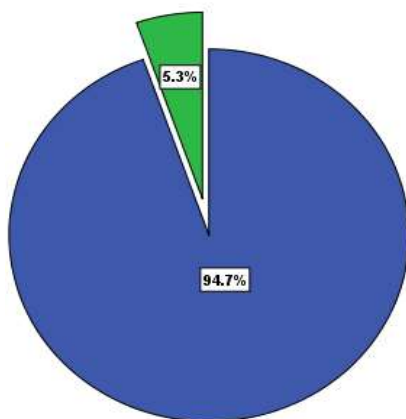
Solo el 5,3% de los clientes se los considera VIP y el 94,7% no lo son como se aprecia en el Cuadro 3.2. Estos clientes tienen beneficios exclusivos y por lo tanto son pocos y muy selectos dependiendo de su comportamiento de compra.

Cuadro 3.2: Tabla y diagrama de pastel de los clientes VIP

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Tabla de frecuencias			
Club VIP	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada
NO	95249	94,7	94,7
SI	5383	5,3	100
Total		100,0	

■ NO
■ SI



Elaborado por: J. Camboverde. Facultad de Ciencias Naturales v Matemáticas

3.1.3 Clientes Súper Crédito

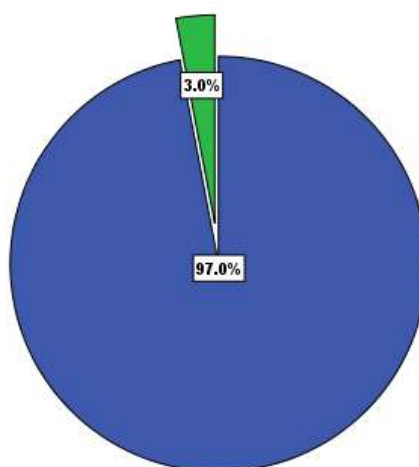
El Cuadro 3.3 nos presenta un producto nuevo que es el Súper Crédito que se trata de un extra cupo para que los clientes puedan adquirir productos de tecnología, muebles y electrodomésticos. Donde observamos que solo el 3% de los clientes aplicaron al Súper Crédito mientras que el 97% no lo han hecho.

Cuadro 3.3: Tabla y diagrama de pastel de los clientes Súper Crédito

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Tabla de frecuencias			
Súper Crédito	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada
NO	97626	97,0	97,0
SI	3006	3,0	100
Total	100632	100	

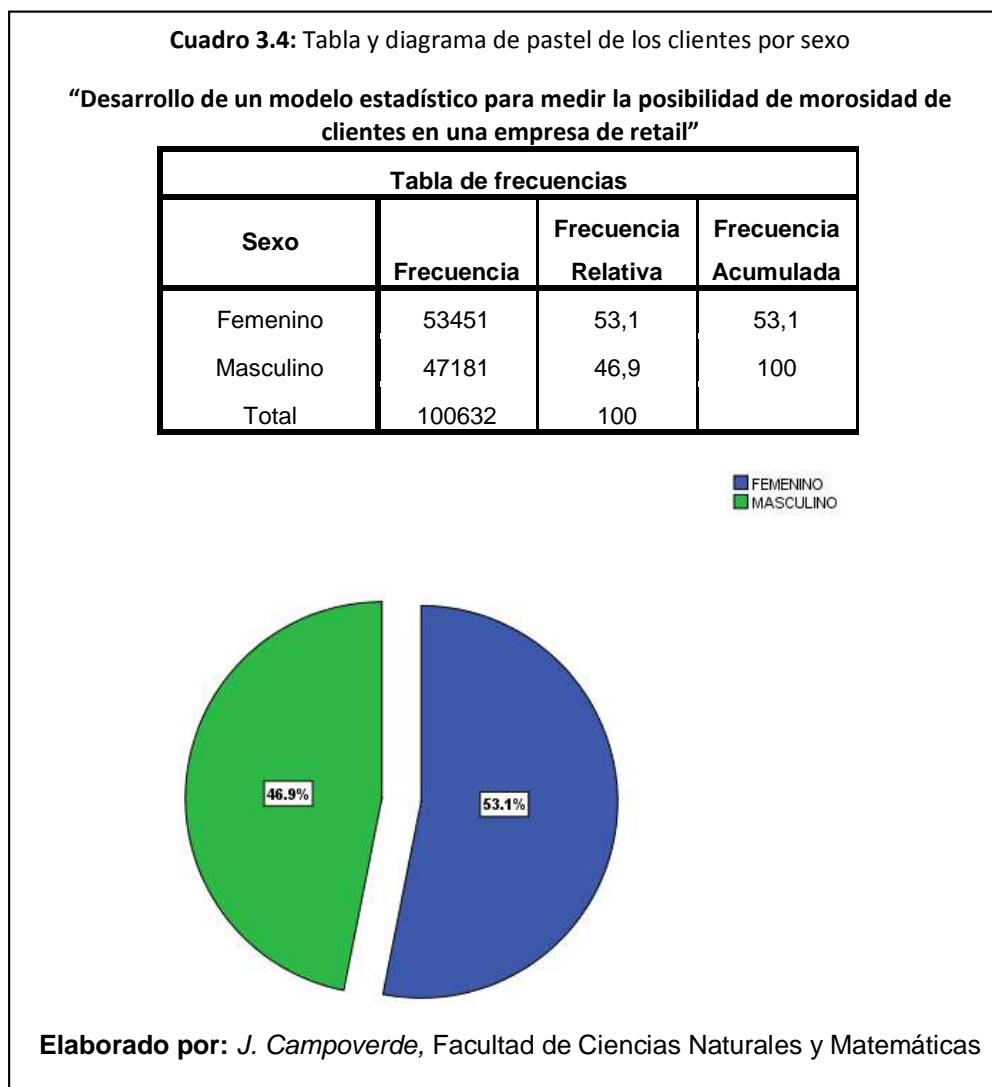
■ NO
■ SI



Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

3.1.4 Sexo del Cliente

El Cuadro 3.4 nos muestra que existen más mujeres que hombres como clientes de la empresa XYZ S.A., esto se comprueba con el 53,1% contra un 46,9% perteneciente al sexo femenino y masculino respectivamente.



3.1.5 Cliente Moroso

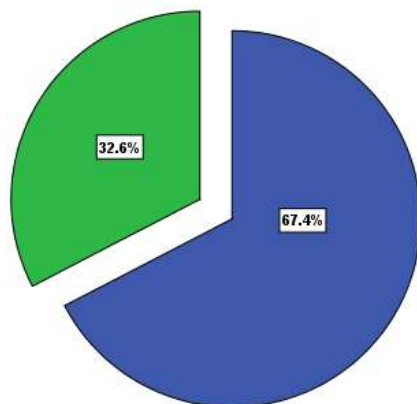
Los clientes morosos son aquellos que de Julio 2014 a Junio2015 en promedio tuvieron una mora mayor a 60 días. En el Cuadro 3.5 se aprecia que el 32,6% de los clientes son morosos y el 67,4% no lo son.

Cuadro 3.5: Tabla y diagrama de pastel de los clientes morosos

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Tabla de frecuencias			
Moroso	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada
No	67797	67,4	67,4
Si	32835	32,6	100
Total	100632	100	

■ NO
■ SI



Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

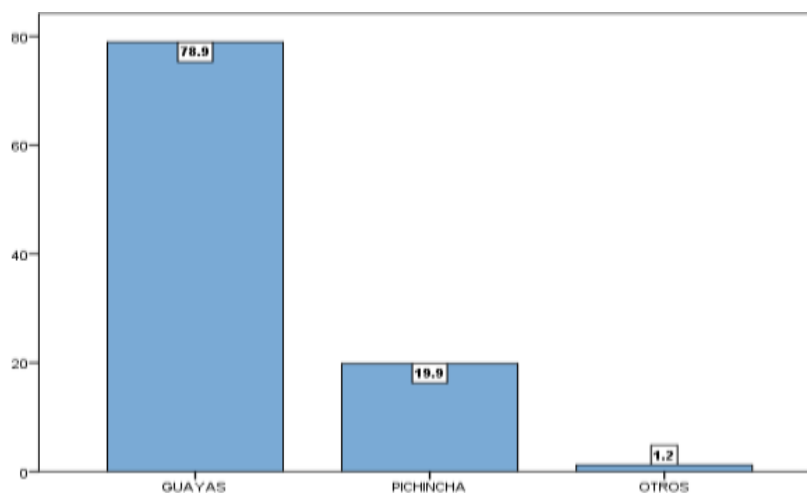
3.1.6 Provincia de Residencia del Cliente

El Cuadro 3.6 nos muestra que en la provincia del Guayas existen más clientes que en las otras provincias. Guayas tiene un 78,9%, Pichincha 19,9% y en otras provincias con sólo 1,2%.

Cuadro 3.6: Tabla y diagrama de barras de las provincias de residencia de los clientes

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Tabla de frecuencias			
Provincia	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada
Guayas	79392	78,9	78,9
Pichincha	19996	19,9	98,8
Otros	1244	1,2	100
Total	100632	100	



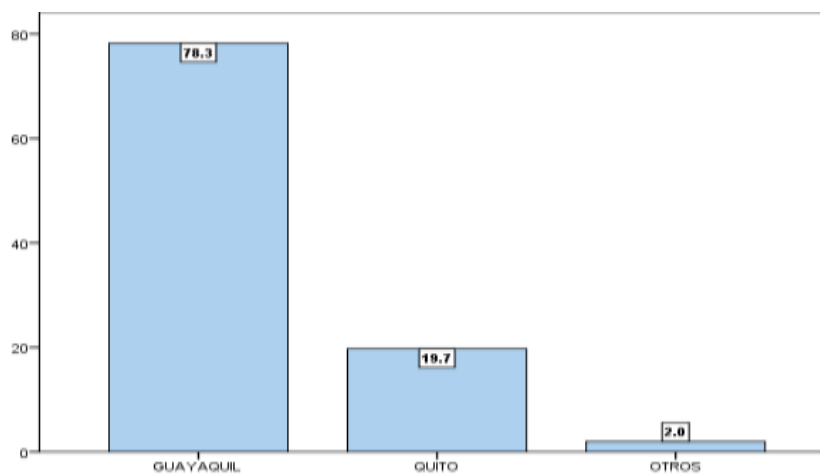
Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

3.1.7 Ciudad de Residencia del Cliente

Con respecto a la ciudad de residencia del cliente, observamos en el Cuadro 3.7 que la mayoría viven en la ciudad de Guayaquil con un 78,3%, seguidos por Quito con el 19,7% y apenas el 2% viven en otras ciudades.

Cuadro 3.7: Tabla y diagrama de barras de la ciudad de residencia de los clientes
“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Tabla de frecuencias			
Ciudad	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada
Guayaquil	78754	78,3	78,3
Quito	19872	19,7	98,0
Otros	2006	2,0	100
Total	100632	100	



Elaborado por: *J. Campoverde*. Facultad de Ciencias Naturales y Matemáticas

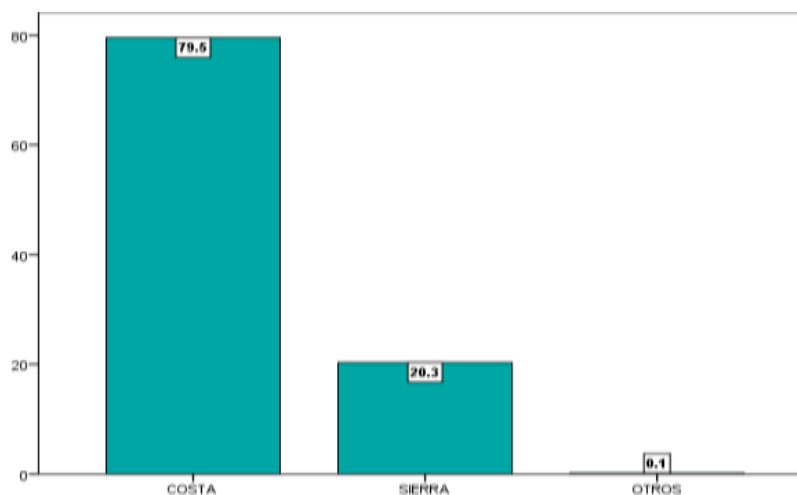
3.1.8 Región de Residencia del Cliente

En el Cuadro 3.8 observamos que la mayoría de los clientes se encuentran en las Costa y Sierra con el 79,5% y 20,3% respectivamente; mientras que en el Oriente y Galápagos solo es 0,1%.

Cuadro 3.8: Tabla y diagrama de barras de la región de residencia de los clientes

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Tabla de frecuencias			
Región	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada
Costa	80039	79,5	79,5
Sierra	20456	20,3	99,9
Otros	137	0,1	100
Total	100632	100	



Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

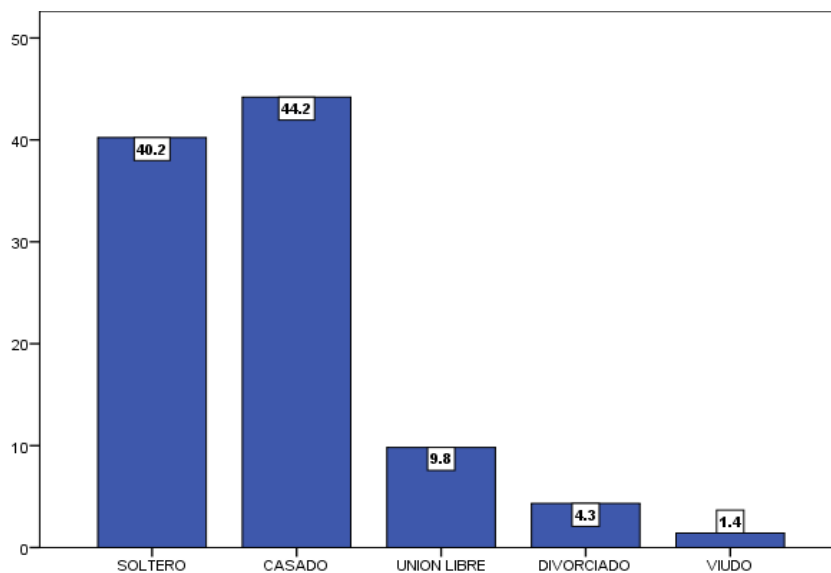
3.1.9 Estado Civil del Cliente

En el Cuadro 3.9 tenemos que el 44,2% de los clientes son casados, 40,2% son solteros, el 9,8% son unión libre, el 4,3% están divorciados y el 1,4 son viudos.

Cuadro 3.9: Tabla y diagrama de barras del estado civil de los clientes

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Tabla de frecuencias			
Estado Civil	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada
Soltero	40484	40,2	40,2
Casado	44478	44,2	84,4
Unión Libre	9885	9,8	94,3
Divorciado	4360	4,3	98,6
Viudo	1425	1,4	100
Total	100632	100	



Elaborado por: *J. Campoverde*, Facultad de Ciencias Naturales y Matemáticas

3.2. Análisis Inferencial

En esta sección del capítulo 3 procederemos a trabajar con las tablas de contingencia para realizar el análisis inferencial. Mediante las tablas de contingencia estableceremos si existe independencia entre la variable a ser explicada y las variables explicativas, observando el valor p de la prueba de hipótesis.

3.2.1 Tabla de Contingencia: Moroso vs Emitió RC

En el Cuadro 3.10 podemos observar que el 90,2% son los clientes declarados no morosos dado que si se les entregó el reporte de crédito, mientras que el 84,6% de los clientes son declarados morosos dado que no se les emitió el reporte de crédito.

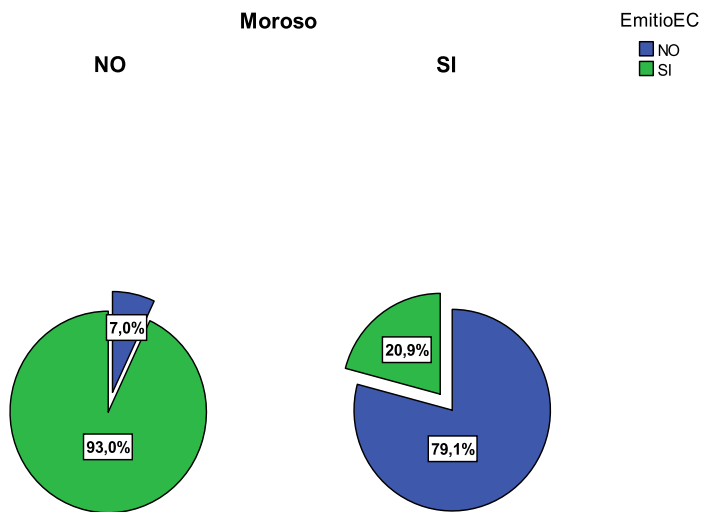
Mediante el valor p de la prueba de hipótesis que es igual a 0,00, cayendo en la zona de rechazo, podemos decir que existe evidencia estadística para negar la

hipótesis nula y concluir que las variables ser morosos y emitió reporte de crédito no son independientes.

Cuadro 3.10: Tabla de contingencia y diagrama de pastel: Moroso vs Emitió RC

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

		Emitió RC	
		NO	SI
Moroso	NO	4727 15.4%	63070 90.2%
	SI	25983 84.6%	6852 9.8%



H_0 : Las variables Moroso y Emitió RC son independientes
 Vs.

H_1 : Las variables Moroso y Emitió RC no son independientes

Chi Cuadrado	Grados de Libertad	Valor p
54322,51	1	0,00

Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

3.2.2 Tabla de Contingencia: Moroso vs Club VIP

El 65,9% y 34,1% de los clientes son declarados no morosos y morosos respectivamente dado que no pertenecen al club VIP de la empresa. Tenemos que el 93% de los clientes son no morosos dado que sí pertenecen al club VIP, mientras que el 7% son morosos dado que sí pertenecen al club VIP como se muestra en el Cuadro 3.11.

Con un valor p de 0,00 podemos decir que existe evidencia estadística para negar la hipótesis nula que decía que las variables moroso y club VIP son independientes. Por lo tanto las variables ser moroso y pertenecer al club VIP no son independientes.

Cuadro 3.11: Tabla de contingencia y diagrama de pastel: Moroso vs VIP

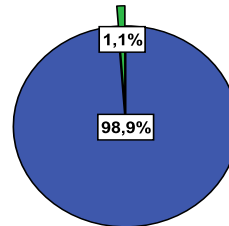
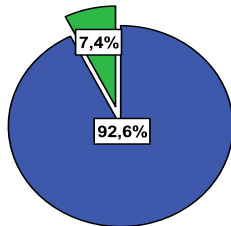
“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

		Club VIP	
		NO	SI
Moroso	NO	62791 65.9%	5006 93.0%
	SI	32458 34.1%	377 7.0%

Moroso

NO
SI

Clubvip
■ NO
■ SI



H_0 : Las variables *Moroso* y *Club VIP* son independientes
Vs.

H_1 : Las variables *Moroso* y *Club VIP* no son independientes

Chi Cuadrado	Grados de Libertad	Valor p
1698,87	1	0,00

Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

3.2.3 Tabla de Contingencia: Moroso vs Súper Crédito

En el Cuadro 3.12 apreciamos que el 91,9% de los clientes son no morosos dado que son clientes súper crédito, mientras que el 8,1% son morosos dado que si son clientes súper crédito. También observamos que el 66,6% y el 33,4% son clientes no morosos y morosos respectivamente dado que no son clientes súper crédito.

Teniendo la hipótesis nula: Las variables ser moroso y ser cliente súper crédito son independientes, mediante el valor p igual a 0,00 podemos concluir que existe evidencia estadística para negar la hipótesis nula. Por lo tanto las variables ser moroso y ser cliente súper crédito no son independientes.

Cuadro 3.12: Tabla de contingencia y diagrama de pastel: Moroso vs Súper Crédito

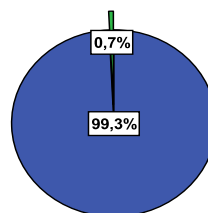
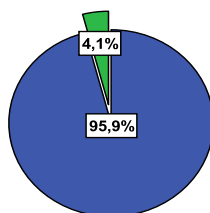
“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

		Súper Crédito	
		NO	SI
Moroso	NO	65034 66,6%	2763 91,9%
	SI	32592 33,4%	243 8,1%

Moroso

NO SI

Super_Credito
■ NO
■ SI



H_0 : Las variables Moroso y Súper Crédito son independientes
Vs.

H_1 : Las variables Moroso y Súper Crédito no son independientes

Chi Cuadrado	Grados de Libertad	Valor p
849,198	1	0,00

Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

3.2.4 Tabla de Contingencia: Moroso vs Sexo

En el Cuadro 3.13 podemos observar que el 65,3% son los clientes declarados no morosos dado que son del sexo masculino, mientras que el 30,8% de los clientes son declarados morosos dado que son del sexo femenino. Mientras que el 69,2% de los clientes no son morosos dado que son del sexo femenino y el 34,7% son clientes morosos dado que son clientes del sexo masculino.

Mediante el valor p de la prueba de hipótesis que es igual a 0,00, cayendo en la zona de rechazo, podemos decir que existe evidencia estadística para negar la hipótesis nula y concluir que las variables ser morosos y sexo no son independientes.

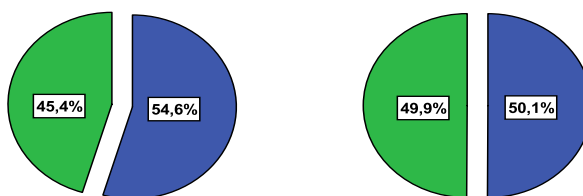
Cuadro 3.13: Tabla de contingencia y diagrama de pastel: Moroso vs Sexo
 “Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

		Sexo	
		Femenino	Masculino
Moroso	NO	37009 69,2%	30788 65,3
	SI	16442 30,8%	16393 34,7%

Moroso

NO SI

Sexo
■ FEMENINO
■ MASCULINO



H_0 : Las variables Moroso y Sexo son independientes
 Vs.

H_1 : Las variables Moroso y Sexo no son independientes

Chi Cuadrado	Grados de Libertad	Valor p
180,95	1	0,00

Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

3.2.5 Tabla de Contingencia: Moroso vs Provincia

Cuadro 3.14: Tabla de contingencia y valor p: Moroso vs Provincia

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

		MOROSO	
		NO	SI
PROVINCIA	GUAYAS	52628 66,3%	26764 33,7%
	PICHINCHA	14175 70,9%	5821 29,1%
	OTROS	994 79,9%	250 20,1%

H_0 : Las variables Moroso y Provincia son independientes
Vs.

H_1 : Las variables Moroso y Provincia no son independientes

Chi Cuadrado	Grados de Libertad	Valor p
243,773	2	0,00

Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

El 66,3% y 33,7% de los clientes son declarados no morosos y morosos respectivamente dado que pertenecen a la provincia del Guayas. Tenemos que el 70,9% de los clientes son no morosos dado que pertenecen a la provincia de Pichincha, mientras que el 29,1% son morosos dado que pertenecen a la provincia de Pichincha. Y el 79,9% son clientes no morosos dado que pertenecen a otras provincias, mientras que el 20,1% son morosos dado que pertenecen a otras provincias como se muestra en el Cuadro 3.14.

Con un valor p de 0,00 podemos decir que existe evidencia estadística para negar la hipótesis nula que decía que las variables morosos y provincia son independientes. Por lo tanto las variables ser moroso y provincia no son independientes.

3.2.6 Tabla de Contingencia: Moroso vs Ciudad

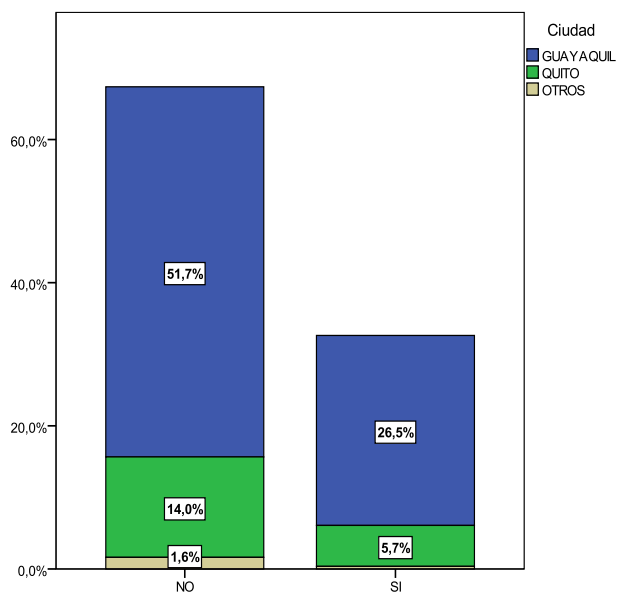
El 66,1% y 33,9% de los clientes son declarados no morosos y morosos respectivamente dado que pertenecen a la ciudad de Guayaquil. Tenemos que el 71% de los clientes son no morosos dado que pertenecen a la ciudad de Quito, mientras que el 29% son morosos dado que pertenecen a la ciudad de Quito. Mientras que el 81,6% y el 18,4% son no morosos y morosos respectivamente dado que pertenecen a otras ciudades como se muestra en el Cuadro 3.15.

Con un valor p de 0,00 podemos decir que existe evidencia estadística para negar la hipótesis nula que decía que la variable moroso y ciudad son independientes. Por lo tanto las variables ser moroso y ciudad no son independientes.

Cuadro 3.15: Tabla de contingencia y diagrama de barras: Moroso vs Ciudad

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

		MOROSO	
		NO	SI
CIUDAD	GUAYAQUIL	52043 66,1%	26711 33,9%
	QUITO	14118 71,0%	5754 29,0%
	OTROS	1636 81,6%	370 18,4%



Moroso

H_0 : Las variables Moroso y Ciudad son independientes
Vs.

H_1 : Las variables Moroso y Ciudad no son independientes

Chi Cuadrado	Grados de Libertad	Valor p
365,037	2	0,00

Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

3.2.7 Tabla de Contingencia: Moroso vs Región

En el Cuadro 3.16 apreciamos que el 66,4% de los clientes son no morosos dado que son clientes de la Costa, mientras que el 33,6% son morosos dado que son clientes de la región Costa. También observamos que el 71,1% y el 28,9% son clientes no morosos y morosos respectivamente dado que son clientes de la región Sierra. Y por último el 47,4% son clientes no morosos y el 52,6% son clientes morosos dado que son de otras regiones del país.

Teniendo la hipótesis nula: Las variables ser moroso y región son independientes, mediante el valor p igual a 0,00 podemos concluir que existe evidencia estadística para negar la hipótesis nula. Por lo tanto las variables ser moroso y región no son independientes.

Cuadro 3.16: Tabla de contingencia y valor p: Moroso vs Región

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

		MOROSO	
		NO	SI
REGIÓN	COSTA	53179 66,4%	26860 33,6%
	SIERRA	14553 71,1%	5903 28,9%
	OTROS	65 47,4%	72 52,6%

H_0 : Las variables Moroso y Región son independientes
Vs.

H_1 : Las variables Moroso y Región no son independientes

Chi Cuadrado	Grados de Libertad	Valor p
188,607	2	0,00

Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

3.2.8 Tabla de Contingencia: Moroso vs Estado Civil

Cuadro 3.17: Tabla de contingencia y valor p: Moroso vs Estado Civil

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

		MOROSO	
		NO	SI
Estado Civil	SOLTERO	25236 62,3%	15248 37,7%
	CASADO	32125 72,2%	12353 27,8%
	UNIÓN LIBRE	6266 63,4%	3619 36,6%
	DIVORCIADO	3065 70,3%	1295 29,7%
	VIUDO	1105 77,5%	320 22,5%

H_0 : Las variables Moroso y Estado Civil son independientes
Vs.

H_1 : Las variables Moroso y Estado Civil no son independientes

Chi Cuadrado	Grados de Libertad	Valor p
1099,378	4	0,00

Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

En el Cuadro 3.17 podemos observar que el 62,3% son los clientes declarados no morosos y el 37,7% de los clientes son declarados morosos dado que son solteros. Mientras que el 72,2% y el 27,8% de los clientes no son y son morosos respectivamente dado que son casados. El 63,4% de los clientes son no morosos dado que son unión libre mientras que el 29,7% de los clientes son morosos dado que son divorciados. El 77,5% y el 22,5% son clientes no morosos y morosos respectivamente dado que son viudos.

Mediante el valor p de la prueba de hipótesis que es igual a 0,00, cayendo en la zona de rechazo, podemos decir que existe evidencia estadística para negar la hipótesis nula y concluir que las variables ser morosos y estado civil no son independientes.

3.3. Regresión Logística

3.3.1 Modelo Estadístico

Dado que tenemos una variable dependiente o de respuesta dicotómica(SI o NO) es decir que solo puede tomar 2 valores, el mejor modelo para explicar esta variable en función de un conjunto de variables independientes es el modelo de Regresión Logística Binomial.

3.3.2 Variables en el modelo

En la Tabla 3.1 podemos observar la variable dependiente o de interés del modelo:

Tabla 3.1: Codificación de la variable dependiente del modelo de Regresión Logística Binomial

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Codificación de la variable dependiente Y		
Variable	Categoría	Codificación
MOROSO	SI	1
	NO	0

Elaborado por: *J. Campoverde*, Facultad de Ciencias Naturales y Matemáticas

A continuación podemos observar las variables independientes del modelo de Regresión Logística Binomial:

Tabla 3.2: Codificación de las variables independientes del modelo de Regresión Logística Binomial

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Codificación de las variables independientes		
Variable	Categoría	Codificación
Emitió EC	SI	1
	NO	0
Club VIP	SI	1
	NO	0
Súper Crédito	SI	1
	NO	0
Provincia	Guayas	1
	Pichincha	2
	Otros	3
Ciudad	Guayaquil	1
	Quito	2
	Otros	3
Región	Costa	1
	Sierra	2
	Otros	3
Sexo	Femenino	0
	Masculino	1
Estado Civil	Soltero	1
	Casado	2
	Unión Libre	3
	Divorciado	4
	Viudo	5
Edad		

Elaborado por: *J. Campoverde*, Facultad de Ciencias Naturales y Matemáticas

3.3.3 Estimación de los Coeficientes

Para determinar los coeficientes del primer modelo de regresión logística binomial de nuestro trabajo utilizaremos como dijimos anteriormente el método de Máxima Verosimilitud. El software estadístico R-Studio nos proporcionará los coeficientes del modelo. En la Tabla 3.3 podemos observar los resultados del primer modelo.

Tabla 3.3: Resultados del primer modelo de Regresión Logística binomial

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Variables	$\hat{\beta}_i$	Std. Error	Estadístico de Wald	Valor p
(Intercept)	25.628.926	0.0466963	54.884	< 2e-16 ***
BASE\$EmitioECSi	-38.591.461	0.0209327	-184.360	< 2e-16 ***
BASE\$Clubvipsi	-0.6149014	0.0603251	-10.193	< 2e-16 ***
BASE\$Super_CreditoSi	-0.6481703	0.0769884	-8.419	< 2e-16 ***
BASE\$ProvinciaOtros	0.0368354	0.1983587	0.186	0.85268
BASE\$ProvinciaPichincha	0.0179865	0.2818253	0.064	0.94911
BASE\$CiudadOtros	-0.7406247	0.1332943	-5.556	2.76e-08 ***
BASE\$CiudadQuito	-0.0725163	0.1716191	-0.423	0.67263
BASE\$RegionOtros	0.8054421	0.3049055	2.642	0.00825 **
BASE\$RegionSierra	-0.1034037	0.2248801	-0.460	0.64565
BASE\$SexoMasculino	0.3529731	0.0205078	17.212	< 2e-16 ***
BASE\$EstadocivilDivorciado	0.2750143	0.0514794	5.342	9.18e-08 ***
BASE\$EstadocivilSoltero	0.1687582	0.0229263	7.361	1.83e-13 ***
BASE\$EstadocivilUnion Libre	0.2169651	0.0350266	6.194	5.85e-10 ***
BASE\$Estadocivilviudo	0.1531799	0.0936861	1.635	0.10204
BASE\$Edad	-0.0298081	0.0009989	-29.841	< 2e-16 ***
AIC: 69298				

Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

En la Tabla 3.4 observamos la estimación del segundo modelo que calculamos.

Tabla 3.4: Resultados del segundo modelo de Regresión Logística binomial
“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

VARIABLES	$\hat{\beta}_i$	Std. Error	Estadístico de Wald	Valor p
(Intercept)	25.628.502	0.0466914	54.889	< 2e-16 ***
BASE\$EmitioECSi	-38.591.282	0.0209324	-184.362	< 2e-16 ***
BASE\$Clubvipsi	-0.6149363	0.0603245	-10.194	< 2e-16 ***
BASE\$Super_CreditoSi	-0.6482063	0.0769887	-8.419	< 2e-16 ***
BASE\$CiudadOtros	-0.7239414	0.0945176	-7.659	1.87e-14 ***
BASE\$CiudadQuito	-0.0731045	0.1317414	-0.555	0.57896
BASE\$RegionOtros	0.8257577	0.2814364	2.934	0.00335 **
BASE\$RegionSierra	-0.0848289	0.1303151	-0.651	0.51508
BASE\$SexoMasculino	0.3529856	0.0205075	17.213	< 2e-16 ***
BASE\$EstadocivilDivorciado	0.2750392	0.0514797	5.343	9.16e-08 ***
BASE\$EstadocivilSoltero	0.1687536	0.0229259	7.361	1.83e-13 ***
BASE\$EstadocivilUnion Libre	0.2169370	0.0350262	6.194	5.88e-10 ***
BASE\$EstadocivilViudo	0.1531620	0.0936855	1.635	0.10208
BASE\$Edad	-0.0298071	0.0009988	-29.842	< 2e-16 ***
AIC: 69294				

Elaborado por: *J. Campoverde*, Facultad de Ciencias Naturales y Matemáticas

Como podemos observar el AIC del segundo modelo es menor que el del primer modelo de regresión logística por lo tanto resulta más efectivo tomar el segundo modelo para predecir nuestra variable dependiente.

Por medio del valor p de las variables podemos darnos cuenta que vivir en Quito, en la Sierra y el estado civil viudo no son factores significativos para el modelo porque el valor de significancia es mayor a 0,05.

Este modelo se lo utiliza para determinar la probabilidad de que un cliente sea moroso en función de sus características, en otras palabras, la probabilidad de que $y=1$.

$$P(y = 1|x) = \frac{e^{\eta}}{1 + e^{\eta}} = \frac{1}{1 + e^{-\eta}}$$

Donde,

$$\eta = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p$$

Por lo tanto nuestro modelo quedaría expresado de la siguiente manera:

$$\begin{aligned} \eta = & 2.5628502 - 3.8591282 \text{ Emitió EC (Si)} - 0.6149363 \text{ Club VIP (Si)} \\ & - 0.6482063 \text{ Super Credito (Si)} - 0.7239414 \text{ Ciudad (Otros)} \\ & - 0.0731045 \text{ Ciudad (Quito)} + 0.8257577 \text{ Región (Otros)} \\ & - 0.0848289 \text{ Región (Sierra)} + 0.3529856 \text{ Sexo (Masculino)} \\ & + 0.2750392 \text{ Estado Civil (Divorciado)} \\ & + 0.1687536 \text{ Estado Civil (Soltero)} \\ & + 0.2169370 \text{ Estado Civil (Unión Libre)} \\ & + 0.1531620 \text{ Estado Civil (Viudo)} - 0.0298071 \text{ Edad} \end{aligned}$$

El modelo nos indica que los clientes que viven en Otras Regiones del país son los de mayor probabilidad de ser morosos; mientras que los clientes de sexo Masculino, Estado Civil Divorciado, Unión Libre, Soltero y Viudo aumentan la probabilidad de ser morosos; mientras que a los que se les emitió Estado de Cuenta, son Club VIP, Súper Crédito, de la Ciudad de Quito, Otros, de la Región Sierra y de mayor Edad disminuyen esta probabilidad.

Tabla 3.5: Explicación de los signos de los coeficientes estimados

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Co-variables	Categoría	Signo	Explicación
Emitió EC	Si	Negativo	El emitir EC disminuye la probabilidad de ser moroso
Club VIP	Si	Negativo	El ser cliente VIP disminuye la probabilidad de ser moroso
Súper Crédito	Si	Negativo	Ser cliente Súper Crédito disminuye la probabilidad de ser moroso
Ciudad	Otros	Negativo	Pertenecer a otras ciudades disminuye la probabilidad de ser moroso
	Quito	Negativo	Pertenecer a Quito disminuye la probabilidad de ser moroso
Región	Otros	Positivo	Ser de Otra Región del país aumenta la probabilidad de ser moroso
	Sierra	Negativo	Vivir en la Sierra disminuye la probabilidad de ser moroso
Sexo	Masculino	Positivo	Ser Masculino aumenta la probabilidad de ser moroso
Estado Civil	Divorciado	Positivo	El estar Casado disminuye la probabilidad de ser moroso
	Soltero	Positivo	
	Unión Libre	Positivo	
	Viudo	Positivo	
Edad		Negativo	El tener más edad disminuye la probabilidad de ser moroso

Elaborado por: *J. Campoverde*, Facultad de Ciencias Naturales y Matemáticas

3.3.4 Resultados de la Razón de Propensión (Odds Ratio)

Como se revisó en el capítulo anterior, la Razón de Propensión (ODDS), se los obtiene si dividimos la probabilidad de que ocurra el evento entre la probabilidad de que no ocurra. Es decir:

$$ODDS = \frac{\frac{p_i}{1-p_i}}{\frac{p_{i+1}}{1-p_{i+1}}} = \exp(\beta_i)$$

Tabla 3.6: ODDS del modelo de Regresión Logística
 “Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Co-variables	Categorías	ODDS	I.C para exp (95%)	
			Inferior	Superior
Constante		12,97273927	11,84045516	14,2186124
Emitió EC	Si	0.02108637	0.02023657	0.0219671
Club VIP	Si	0.54067534	0.47961329	0.6075924
Súper Crédito	Si	0.52298299	0.44855830	0.6066322
Ciudad	Otros	0.48483753	0.40219784	0.5825343
	Quito	0.92950366	0.71800211	1,2030999
Región	Otros	2,28361038	1,32182906	3,9716976
	Sierra	0.91866944	0.71172400	1,1859276
Sexo	Masculino	1,42331063	1,36727268	1,4817263
Estado Civil	Divorciado	1,31658224	1,18987465	1,4559160
	Soltero	1,18382844	1,13181451	1,2382411
	Unión Libre	1,24226590	1,15973447	1,3304167
	Viudo	1,16551381	0.96914836	1,3990937
Edad		0.97063277	0.96873244	0.9725329

Elaborado por: J. Campoverde, Facultad de Ciencias Naturales y Matemáticas

Analizando la Tabla 3.6, observamos lo siguiente:

- La persona que se le emite el Reporte de Crédito tiene 98% menos posibilidad de ser moroso del que no se le emite el Reporte de Crédito.
- La persona de Otra ciudad tiene 51% menos posibilidad de ser moroso que la persona de la Ciudad de Guayaquil.
- La persona de mayor edad tiene el 3% menos posibilidad de ser moroso que el de menor edad
- La persona que vive en la Ciudad de Quito tiene 7% menos posibilidad de ser moroso que la persona de Guayaquil.
- La persona de Otra región del país tiene 2.28 veces más posibilidad de ser moroso que la persona de la región Costa del país.
- La persona de sexo Masculino tiene 42% más posibilidad de ser moroso que la persona de sexo Femenino.
- La persona de estado civil Divorciada tiene 32% más posibilidad de ser moroso que la Casada.

3.3.5 Bondad de Ajuste

3.3.5.1 Tabla de Clasificación

Tabla 3.7: Tabla de Clasificación

“Desarrollo de un modelo estadístico para medir la posibilidad de morosidad de clientes en una empresa de retail”

Tabla de clasificación^a

Observado			Pronosticado		
			Moroso		Porcentaje correcto
			NO	SI	
Paso 1	Moroso	NO	63074	4723	93,0
		SI	6857	25978	79,1
		Porcentaje global			88,5

a. El valor de corte es ,500

Elaborado por: *J. Campoverde*, Facultad de Ciencias Naturales y Matemáticas

En la Tabla 3.7 observamos la clasificación tanto de lo pronosticado como lo observado con respecto a la variable dependiente Moroso.

Notamos que de los 67797 clientes observados que fueron no morosos, tenemos que 63074 se pronosticaron como no morosos en los observados; es decir el 93% se pronosticaron correctamente y 4723 fueron pronosticados incorrectamente.

También observamos que de 32835 observaciones de clientes morosos, el 79,1% fue pronosticado correctamente y 6857 clientes fueron pronosticados incorrectamente.

El modelo nos proporciona un 88,5% de correcta clasificación en forma general.

CONCLUSIONES Y RECOMENDACIONES

Al finalizar nuestro estudio obtuvimos las correspondientes conclusiones y recomendaciones.

Conclusiones

1. En la empresa XYZ S.A. la mayoría de los clientes son de sexo femenino con un 53,1%.
2. La mayoría de los clientes de la empresa se encuentran al día; es decir no son morosos.
3. En la ciudad de Guayaquil se encuentran concentrados la mayor parte de los clientes con 78,3%.
4. La mayor cantidad de clientes se encuentran casados y solteros que representan el 44,2% y 40,2% respectivamente.
5. De los clientes no morosos la mayoría son las mujeres con 54,6% sobre los hombres, por lo que parece que las mujeres son más responsables con el pago.

6. La mayoría de los clientes morosos se encuentran en la ciudad de Guayaquil, puede ser por lo que tiene la mayoría de clientes.
7. Los solteros son los clientes más morosos, puede ser por lo que al ser solteros no son responsables con sus obligaciones.
8. Se puede decir que la variable Provincia no es significativa en el modelo para explicar la variable dependiente.
9. El ser de otra región del país aumenta la probabilidad de ser moroso y el hecho de que se emita el estado de cuenta disminuye la probabilidad de ser moroso.
10. La persona de Otra región del país tiene 2.28 veces más posibilidad de ser moroso que la persona de la región Costa del país.
11. La persona de sexo Masculino tiene 42% más posibilidad de ser moroso que la persona de sexo Femenino.
12. Se puede decir que las mujeres que son casadas y de más edad son las menos morosas.

Recomendaciones

1. Se recomienda una constante depuración de las bases de datos para evitar inconsistencia en la información.
2. Tener presente el sexo, la edad y el estado civil de la persona que llegue a solicitar crédito a la empresa porque estas características demográficas influyen en la morosidad del cliente.
3. Se podrían establecer nuevos puntos de pago en Otras regiones del país para los clientes que no viven ni en la Costa ni en la Sierra dado que en la actualidad el único medio de pago es en el almacén y no poseemos almacenes en Otras regiones que no sean ni Costa ni Sierra.
4. Tener informado al cliente de sus obligaciones de pago como la fecha máxima de pago y los valores a cancelar por medio del Reporte de Crédito porque esta es un factor de suma importancia para disminuir la morosidad del cliente.
5. Se recomienda exigir a los promotores y puntos de captación de clientes llenar las solicitudes de crédito de forma correcta y con mayor información de parte del cliente, información como el sueldo, número de hijos, gastos, información bancaria, vivienda, etc. para obtener mejores resultados.

Bibliografía

1. Abdelrahman, A. (2010). *Applying logistic regression model to the second primary cancer data*. Allahabad: Pushpa Publishing House.
2. Lopez, F. J. (Noviembre de 2012). *Apuntes y videos de Bioestadística*. Obtenido de www.bioestadistica.uma.es/baron/apuntes/ficheros/cap08.pdf
3. Paz, J. M. (23 de 01 de 2009). www.seqc.es. Obtenido de www.seqc.es/dl.asp?175.145.205.255.15.30...
4. Valeriano, K. A. (2012). Ingeniera en Estadística e Informática. *Estadística y distribución espacial de la migración en el Ecuador según censo 2010*. Guayaquil, Guayas, Ecuador.

Anexo: Códigos de Programación en R

Lectura de datos

```
setwd("C:/Users/Johnny Campoverde/Desktop/Juan/Materia_Integradora/Base")
```

```
BASE=read.table("BASE2.txt",header=TRUE,sep=" ")
```

```
View(BASE)
```

Estructura de los datos

```
str(BASE)
```

Modelo de Regresión Logística Binomial

```
m1=glm(BASE$Moroso~BASE$EmitioEC+BASE$Clubvip+BASE$Super_Credito+  
BASE$Provincia+BASE$Ciudad+BASE$Region+BASE$Sexo+BASE$Estadocivil+  
BASE$Edad,family=binomial())
```

Resumen del modelo

```
summary(m1)
```

ODDS

```
coef(m2)
```

```
exp(coef(m2))
```