

UNIVERSIDAD TÉCNICA DE AMBATO



FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL

MAESTRÍA EN MATEMÁTICA APLICADA

Tema: “MODELO MATEMÁTICO PARA MEDIR LA PROBABILIDAD DE LA INCURRENCIA EN MORA DE CRÉDITOS UTILIZANDO REGRESIÓN LINEAL EN UNA INSTITUCIÓN FINANCIERA DE LA CIUDAD DE AMBATO”

Trabajo de titulación previo a la obtención del Grado Académico de Magister en Matemática Aplicada

Modalidad de titulación Proyecto de Desarrollo

Autor: Ing. Luis Alberto Marcalla Pilamunga

Director: Profesor Edgar Johni Bustamante Romero, Ph.D.

Ambato – Ecuador

2022

APROBACIÓN DEL TRABAJO DE TITULACIÓN

A la unidad académica de titulación de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial.

El Tribunal receptor de la defensa del Trabajo de Titulación presidido por la Ingeniera Elsa Pilar Urrutia Urrutia, Magíster integrado por los señores, Ingeniera Clara Augusta Sánchez Benítez, Magíster, e Ingeniero José Vicente Morales Lozada, PhD., designados por la Unidad Académica de Titulación de la Universidad Técnica de Ambato, para receptor el Trabajo de Titulación con el tema: “Modelo matemático para medir la probabilidad de la incurrencia en mora de créditos utilizando regresión lineal en una institución financiera de la ciudad de Ambato”, elaborado y presentado por el señor Ingeniero Luis Alberto Marcalla Pilamunga, para optar por el Grado Académico de Magíster en Matemática Aplicada; una vez escuchada la defensa oral del Trabajo de Titulación el Tribunal aprueba y remite el trabajo para uso y custodia en las bibliotecas de la Universidad Técnica de Ambato.

Ing. Elsa Pilar Urrutia Urrutia Mg.
Presidente y Miembro del Tribunal de Defensa

Ing. Clara Augusta Sánchez Benítez, Mg.
Miembro del Tribunal de Defensa

Ing. José Vicente Morales Lozada, PhD.
Miembro del Tribunal de Defensa

AUTORÍA DEL TRABAJO DE TITULACIÓN

La responsabilidad de las opiniones, comentarios y críticas emitidas en el trabajo de titulación presentado con el tema: “Modelo matemático para medir la probabilidad de la incurrencia en mora de créditos utilizando regresión lineal en una institución financiera de la ciudad de Ambato”, le corresponde exclusivamente a: Ing. Luis Alberto Marcalla Pilamunga, autor bajo la dirección del Profesor Edgar Johni Bustamante Romero, Ph.D., director del Trabajo de Investigación; y el patrimonio intelectual a la Universidad Técnica de Ambato.

Ing. Luis Alberto Marcalla Pilamunga

AUTOR

Profesor Edgar Johni Bustamante Romero, Ph. D.

DIRECTOR

DERECHOS DE AUTOR

Autorizo a la Universidad Técnica de Ambato, para que el Trabajo de Titulación, sirva como un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la Institución.

Cedo los Derechos de mi Trabajo de Titulación, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad Técnica de Ambato.

Ing. Luis Alberto Marcalla Pilamunga

C.I.: 1725100257

ÍNDICE GENERAL

CONTENIDO

PORTADA.....	ii
APROBACIÓN DEL TRABAJO DE TITULACIÓN.....	ii
AUTORÍA DEL TRABAJO DE TITULACIÓN	iii
DERECHOS DE AUTOR	iv
ÍNDICE DE TABLAS	vii
ÍNDICE DE FIGURAS.....	viii
AGRADECIMIENTO	ix
DEDICATORIA	x
RESUMEN EJECUTIVO	xi
EXECUTIVE SUMMARY.....	xiii
CAPÍTULO I.....	1
EL PROBLEMA DE INVESTIGACIÓN.....	1
1.1 Introducción.....	1
1.2 Justificación.....	4
1.3 Objetivos	5
1.3.1 General.....	5
1.3.2 Específicos	5
CAPITULO II	6
ANTECEDENTES INVESTIGATIVOS.....	6
2.1 Estado del Arte	6
CAPITULO III.....	11
MARCO METOLOGICO.....	11
3.1 Ubicación.....	11
3.2 Equipos y materiales	11
3.3 Tipo de investigación	11
3.4 Prueba de Hipótesis	12
3.5 Población	12
3.6 Recolección de información.....	12
3.6.1 Análisis y revisión de estructuras SEPS	12
3.6.2 Extracción de la información.....	13
3.7 Lenguajes de programación y software estadísticos	13
3.7.1 Python	14
3.7.2 Anaconda	14

3.7.3 Python y análisis estadísticos.....	14
3.7.4 R.....	15
3.7.2 RStudio	15
3.7.2 R vs Python.....	15
3.8 Procesamiento de la información y análisis estadístico	17
3.9 Variables.....	19
CAPITULO IV	23
RESULTADOS Y DISCUSIÓN	23
4.1 Análisis descriptivo de las variables.....	24
4.2 Modelación matemática	43
CAPÍTULO V	48
CONCLUSIONES Y RECOMENDACIONES.....	48
5.1 Conclusiones	48
5.2 Recomendaciones	49
5.3 Bibliografía.....	50
5.4 Anexos.....	53

ÍNDICE DE TABLAS

Tabla 1. Cuadro comparativo entre lenguaje Python y R	18
Tabla 2. Ejemplo de uso de variables artificiales.....	21
Tabla 3. Descripción de variables utilizadas en el estudio.....	21
Tabla 4. Cuadro de morosidad por periodos	25
Tabla 5. Cuadro de tabulación de variable periodicidad de pago	26
Tabla 6. Cuadro de tabulación de variable oficina concesión.....	27
Tabla 7. Cuadro de tabulación de variable garantía	29
Tabla 8. Cuadro de tabulación de variable tipo de crédito.....	30
Tabla 9. Cuadro de tabulación de variable destino financiero	31
Tabla 10. Cuadro de tabulación de variable act. econ. receptora de la operación	35
Tabla 11. Cuadro de tabulación de variable provincia destino.	37
Tabla 12. Cuadro de tabulación de variable actividad económica de titular	39
Tabla 13. Cuadro de tabulación de variable nivel de estudio	41
Tabla 14. Cuadro de tabulación de variable tipo de vivienda.....	41
Tabla 15. Cuadro de tabulación de variable género.....	42

ÍNDICE DE FIGURAS

Figura 1. Esquema de envío de estructuras a SEPS	15
Figura 2. Mapa del proceso de ciencia de datos.....	17
Figura 3. Gráfica de barras de préstamos morosos y no morosos por año.....	25
Figura 4. Gráfica de barras de morosidad por periodicidad de pago	27
Figura 5. Gráfica de barras de morosidad por oficina concesión.....	28
Figura 6. Gráfica de barras de morosidad por tipos de garantías.....	30
Figura 7. Gráfica de barras de morosidad por tipos de crédito	30
Figura 8. Gráfica de barras de morosidad por provincia destino.	38
Figura 9. Gráfica de barras de morosidad por actividad económica de titular.	40
Figura 10. Gráfica de barras de morosidad por tipo de vivienda.....	42
Figura 11. Gráfica de barras de morosidad por género.....	42
Figura 12. Resultados de primera corrida de modelo en R.....	44
Figura 13. Resultados de segunda corrida de modelo en R	45

AGRADECIMIENTO

A Dios quien me ha dado la vida y me ha permitido alcanzar un logro más.

A mis padres y hermanas quienes han sido un apoyo incondicional durante toda mi vida.

A mis compañeros de aula y colegas de trabajo que directa o indirectamente contribuyeron a mi investigación.

A mi director y evaluadores de tesis por todo el apoyo y la sabiduría brindada para la finalización del presente trabajo.

Luis Marcalla

DEDICATORIA

Dedico el presente trabajo de investigación a la Universidad Técnica de Ambato, que con sus aulas y docentes han influido en mi carácter y me han dado las herramientas necesarias para mi desarrollo personal y profesional.

Luis Marcalla

UNIVERSIDAD TÉCNICA DE AMBATO
FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E
INDUSTRIAL
MAESTRÍA EN MATEMÁTICA APLICADA

TEMA:

“Modelo matemático para medir la probabilidad de la incurrencia en mora de créditos utilizando regresión lineal en una institución financiera de la ciudad de Ambato”

AUTOR: Ing. Luis Alberto Marcalla Pilamunga

DIRECTOR: Profesor Edgar Johni Bustamante Romero, Ph. D.

LÍNEA DE INVESTIGACIÓN:

- Tecnología de la información y sistemas de control

FECHA: Septiembre 20, 2022

RESUMEN EJECUTIVO

El presente trabajo aborda el análisis de las concesiones de créditos en instituciones financieras de la ciudad de Ambato, con la finalidad de poder estimar la probabilidad de retorno de la inversión, disminuyendo el riesgo y estableciendo convenios entre el prestatario y el prestador. Las variables predictoras que para el presente estudio son todas categóricas, pasaron a formar variables ficticias o dummy al incorporarse al modelo, y la variable dependiente, que puede tomar solo dos posibles valores, 0 o 1, la misma que estima la probabilidad de mora; la presente investigación considera ámbitos socioeconómicas, sociales, territoriales e incluso de género para predecir si un titular de un préstamo cumplirá con diligencia los pagos pactados.

La institución que prestó su colaboración para el diseño del modelo fue la “Cooperativa de ahorro y crédito Chibuleo Ltda.”, ubicada en el centro de la ciudad de Ambato, y vigente en el mercado financiero los últimos 16 años. La cooperativa perteneciente al sector de economía popular y solidaria ha venido fortaleciendo la económica local a través de los productos y servicios que ofertan a la comunidad.

La metodología aplicada en la cooperativa para revisar si un socio es un buen pagador

es realizado con ayuda de diferentes burós y con la exhaustiva diligencia de un analista de crédito, que decide si conceder o no un préstamo. Esta metodología es útil, sin embargo, no considera el histórico de préstamos concedidos con las mismas características. El modelo realizado llena ese vacío y nos una mejor panorámica de los posibles escenarios futuros.

Los datos fueron extraídos desde la base de datos de la cooperativa, una base de datos relacional en SQL Server 2014, mismos datos que fueron cargados a un software estadístico para su procesamiento y analítica. Se utilizó R, un lenguaje de programación enfocado al análisis estadístico de grandes volúmenes de información, además de un ambiente gráfico dispuesto por RStudio versión 2020.02.1. Al ser de tipo categóricas todas las variables estudiadas, y no continuas, se procedió a utilizar caso particular de los modelos lineales generalizados, modelos logit, el mismo sirvió para calcular la probabilidad de recaída en mora: 1 para indicar la incurrencia de mora y 0 para predecir préstamos sin novedades en los pagos. Para el desarrollo del modelo se utiliza el método de máxima verosimilitud con ayuda de R y las variables fueron sometidas a pruebas para determinar su significancia con un nivel de confianza de 95%.

Descriptores: Score crediticio, probabilidad mora, regresión lineal, logit, variables categóricas, variables ficticias, regresión logística

UNIVERSIDAD TÉCNICA DE AMBATO
FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL
MAESTRÍA EN MATEMÁTICA APLICADA

TEMA:

“Mathematical model to measure the probability of credit delinquency using linear regression in a financial institution in the city of Ambato”

AUTHOR: Ing. Luis Alberto Marcalla Pilamunga

DIRECTED BY: Profesor Edgar Johni Bustamante Romero, Ph. D.

LINE OF RESEARCH:

- Information technology and control systems

FECHA: September 20th, 2022

EXECUTIVE SUMMARY

This paper deals with the analysis of credit concessions in financial institutions in the city of Ambato, in order to estimate the probability of return on investment, reducing risk and granting agreements between the borrower and the lender. The intrinsic characteristics of the research, of the predictor variables and the independent variable that were used resulted in the application of a generalized linear model with the binomial logit option. The predictor variables that for the present study are all categorical, became fictitious or dummy variables when incorporated into the model, and the dependent variable, which can take only two possible values, 0 or 1, the same one that estimates the probability of default; grant to the present investigation a different approach that considers the socioeconomic, social, territorial and even gender environment to predict if a loan holder will diligently comply with the payments in the agreed installments.

The institution that lent its collaboration for the design of the model was the "Cooperative of savings and credit Chibuleo Ltda.", located in the center of the city of Ambato, and in force in the financial market for the last 16 years. The cooperative

belonging to the popular and solidarity economy sector has been promoting the local economy through the products and services it offers.

The methodology applied in the cooperative to check if a member is a good payer is carried out with the help of different bureaus and with the thoroughness of a credit analyst, who decides whether or not to grant a loan; This methodology is useful, however, the history of loans granted with the same characteristics is not considered. The model made fills that gap and gives us a better overview of the possible scenarios in the future.

The data was extracted from the cooperative's database, a relational database in SQL Server 2014, which was loaded into statistical software for processing and analysis. R is used, a programming language focused on a statistical analysis of large volumes of data, in addition to a graphic environment RStudio version 2020.02.1. As the variables studied were categorical, and not continuous, a version of linear regression, logit regression, was used to calculate the probability of relapse into default: 1 to indicate the occurrence of default on a loan bonus and 0 to predict loans without news in payments. For the development of the model, the maximum likelihood method is used with the help of R and the variables were tested to determine their significance with a confidence level of 95%.

Keywords: Credit score, default probability, linear regression, logit, categorical variable, dummy variable, logistic regression

CAPÍTULO I

EL PROBLEMA DE INVESTIGACIÓN

1.1 Introducción

La economía de nuestro país es un tema ampliamente abordado a lo largo de los años, con el uso de indicadores que nos dan señales de desarrollo o recesión económica del que hemos sido partícipe. Nuestra economía es dolarizada, abierta y dependiente de los ingresos del petróleo y las exportaciones que aportan el 50% del PIB [1], lo que la hace vulnerable a factores globales; sin duda la reciente emergencia sanitaria generó una profunda recesión que provocó un repunte de la pobreza, esta crisis incrementó los desbalances macroeconómicos que el país intentaba subsanar desde el fin del boom de los precios del petróleo que se situó entre los años 2010 y 2014 [2]. La banca también ha presentado rezagos en su desarrollo y planes de crecimiento por las consideraciones antes mencionadas, y con ello presionándose en busca de llevar sus productos y servicios a sus clientes y poder alcanzar la anhelada reactivación económica.

El sistema financiero ecuatoriano conformado por instituciones privadas, públicas, compañías de seguros y compañías auxiliares del sistema financiero no han sido ajenas a las coyunturas globales y nacionales a los que nuestro país se ha visto afectado. Teniendo que idear y aplicar nuevas estrategias dentro sus procesos normales con el fin de amortiguar en algo la caída de sus indicadores de gestión [3]. La cartera de crédito, piedra angular para el crecimiento de las instituciones financieras ha desarrollado nuevos métodos especializados a la hora de tomar decisiones, teniendo que analizar las cualidades del solicitante de préstamo a la vez que estudiar las coyunturas actuales para medir la probabilidad de retorno del valor a prestar.

Los productos crediticios de acuerdo con la necesidad que se pretende cubrir está clasificada por la Superintendencia de Bancos y Seguros de la siguiente manera: “créditos comerciales”, que consiste en el uso de los pasivos de las empresas, a corto plazo, de la manera menos costosa, “créditos de consumo”, se otorga para la adquisición de bienes o servicios, con plazos entre 1 y dos años (corto plazo), “créditos para vivienda”, se otorgan a persona naturales para la adquisición, construcción, reparación, remodelación y mejoramiento de viviendas propias, y por último, “créditos para microempresas”, la manera común cómo funciona el microcrédito es a través de la

concesión de un préstamo pequeño, cuyo monto está determinado por la capacidad de pago del prestatario, durante un período corto [3].

La solidez y confianza que puede presumir una institución financiera se apoya sobre su calificación de riesgo, la calificación de riesgos se puede definir como “la solvencia y la capacidad de las entidades del sistema financiero nacional para administrar los riesgos con terceros y cumplir sus obligaciones con el público será calificada sobre la base de parámetros mínimos que incluyan una escala uniforme de calificación de riesgo por sectores financieros, de acuerdo con las normas que al respecto emita la Junta de Política y Regulación Monetaria y Financiera”[4]. Las empresas calificadoras de riesgo aceptadas por la Junta Bancaria y por la Superintendencias de bancos y seguros (SBS), realizan una evaluación de cada empresa implementado un sistema que incluye una escala uniforme de calificación de riesgo, cuyos resultados promueven exhortos sobre la capacidad y solvencia de las instituciones del sistema financiero. Sin duda los indicadores de mora en cartera colocada es un factor que se toma en cuenta para una buena calificación, además de otros factores como son: el riesgo económico, riesgo de la industria, posición del mercado, manejo y gestión, suficiencia de capital y eficiencia. Por eso la importancia de aminorar las probabilidades de que un sujeto de crédito no cumpla con sus obligaciones pactadas en el momento de la adjudicación.

En el caso de las cooperativas de ahorro y crédito y demás instituciones de la economía popular y solidaria, su ente regulador dada por la Junta de Política y Regulación Monetaria es la Superintendencia de Economía Popular y Solidaria (SEPS) [5], la SEPS es encargada de la vigilancia, auditoría, intervención, control y supervisión de las actividades financieras de las instituciones que tiene a su cargo. En función a lo anterior las cooperativas están obligadas a enviar grandes cantidades de datos a la SEPS, también denominadas estructuras, las estructuras son remitidas por las entidades controladas a través del sistema de acopio “Gestión de envío de información”, disponible en la página web de la Superintendencia de Economía Popular y Solidaria [6]. Existen varias estructuras, cada una enfocada en un producto o servicio en particular, la estructura operaciones concedidas C01 tiene periodicidad mensual, contiene operaciones de crédito y contingentes que se hayan concedido, novado, refinanciado o reestructurado en el mes que corresponde a la fecha de corte, se estructura de un total de 33 columnas como son cédula del titular, interés normal, número de operación, actividad económica destino de recursos, entre otras. La estructura saldos de operaciones C02, también con periodicidad mensual, en este se

detallan los saldos de operaciones de crédito y contingentes que aún se encuentran activas, así como las cancelaciones de aquellas operaciones que fueron reportadas previamente [7]. También existen estructuras enfocadas a inversiones, portafolio de servicios, balances financieros, contables entre otras.

La institución financiera donde se realiza la investigación, la cooperativa de ahorro y crédito Chibuleo Ltda., está sujeta a acatar las normas y regulaciones de sus entes reguladores, sin descuidar siempre su objetivo de dar un servicio de calidad y calidez a sus cerca de 20000 socios. En el caso de la otorgación de crédito, se ha venido perfeccionando una metodología de evaluación de riesgo de cada solicitud de crédito entrante, siendo un proceso completo e integral, con sus respectivas etapas, y cada una de estas con un fin en particular, en las etapas iniciales se revisa la documentación presentada, garantes disponibles, capacidad de pago, antecedentes penales y coactivas. En las etapas subsiguientes se verifica la información presentada en las etapas iniciales con ayuda de sistemas auxiliares, página de consejo de la judicatura para revisar algún posible antecedente penal del solicitante, consultas a enormes bases de datos, DataBook para consulta de información de patrimonio de socios y garantes, además de la consulta a central de riesgo o burós. La cooperativa ha venido trabajando en los últimos años con el proveedor de score crediticio 'Equifax 360', que provee un histórico acerca del desarrollo de las obligaciones que el solicitante mantuvo en otras entidades financieras, cooperativas o casas comerciales. De acuerdo con las políticas internas, la institución se abstiene de otorgar una operación crediticia a personas con score menor a 700. Para el análisis de cada solicitud de crédito dentro de la cooperativa los analistas recurren a toda la información que esté disponible, esta tarea se vuelve complicada debido que los datos reposan sin más, sin ser procesados y/o analizados. Con el fin de obtener una mejor comprensión de los datos generados en los procesos relacionados a negocios, han aparecido disciplinas que involucran: estadística, minería de datos, aprendizaje automático, inteligencia artificial y analítica preventiva; transformar los datos en conocimiento útil se ha vuelto necesario para los directivos de las empresas en la toma de decisiones, sean estas de cualquier índole. En la presente investigación se pretende aplicar estas nuevas técnicas y así generar estrategias e identificar los factores que influyen en el incremento de mora en la cartera colocada, mediante la preparación de datos, realización de análisis, limpieza, agregación y la manipulación en un análisis avanzado.

1.2 Justificación

En los últimos años el sector económico ha venido creciendo en el sentido de que se han incrementado las personas que se han integrado al sistema financiero nacional, con una simple apertura de una cuenta, también la otorgación de créditos es más accesible para los sectores vulnerables y además el uso de las tarjetas de débito/crédito es más sustancioso debido a las facilidades y seguridades que este provee, todo lo anteriormente dicho hacen notar la premisa [8]. La dinamización de la economía nacional es uno de los principales retos para el gobierno nacional, que luego de la recesión debido a factores internos y externos se vio severamente afectada. El impulso del sector privado, la asequibilidad de créditos a los sectores menos privilegiados, y el fortalecimiento del sistema financiero nacional con ayuda de políticas que contribuyan a su fortalecimiento, son algunas de las acciones que el gobierno ha empezado a tomar.

Dentro del sector financiero se encuentran: el sector financiero público, el sector financiero privado y el sector financiero popular y solidario, donde cada una de ellas persigue objetivos concretos en pro del crecimiento económico y social del país. Tal y como existen entidades en el sector financiero también podemos encontrar varios segmentos de crédito, a las que las personas naturales y jurídicas pueden acceder, como son: consumo, microcrédito, inmobiliario, productivo entre otras [3] . Habiendo un mercado tan grande se hace necesario la calificación cuantitativa o también denomina score crediticio de los solicitantes de crédito para estimar la probabilidad de la devolución del dinero prestado con determinadas condiciones iniciales; siempre con el componente tecnológico de por medio como son tareas de consulta de score crediticio (Equifax), datos sobre patrimonio (Data Book) y gestión de grandes volúmenes de datos (Big Data).

El sistema financiero no es ajeno a la transformación tecnológica que en la actualidad hemos sido espectadores, utilizándolas siempre en el ofrecimiento de nuevos productos y servicios personalizados. Cada vez se ve más el uso de chatbots que simulan la interacción con una persona en tiempo real para consultas y realización de transacciones bancarias, la implementación de inteligencias artificiales para la detección temprana de fraudes y las nuevas herramientas para proteger la privacidad de los usuarios denotan la nueva era que estamos experimentando. Porque no dar un

paso más para la banca en el ámbito de la otorgación de créditos, la posibilidad de estimar el grado de seguridad de que una inversión retornará en el tiempo acordado, con ayuda de las técnicas de inteligencia artificial como es la regresión lineal, contribuirá sustanciosamente al acreedor y deudor.

En el sector financiero es fundamental innovarse constantemente en lo que tiene que ver con los productos y servicios que prestan a la sociedad, por supuesto, con la ayuda de las nuevas tecnologías disponibles. Conociendo las necesidades actuales de la banca y la metodología que utilizan para la otorgación de créditos en un segmento determinado, y como un aporte más a dicha metodología se presente realizar el presente proyecto denominado: “Modelo matemático para medir la probabilidad de la incurrencia en mora de créditos utilizando regresión lineal en una institución financiera de la ciudad de Ambato”. El mismo que pretende realizar un análisis exhaustivo de los factores que propician el incumplimiento de los pagos pactados como una herramienta adicional para la concesión de créditos, además de aportar una perspectiva más acerca de la economía nacional.

1.3 Objetivos

1.3.1 General

Modelizar matemáticamente la probabilidad de la incurrencia en mora de créditos utilizando regresión lineal en una institución financiera de la ciudad de Ambato.

1.3.2 Específicos

- Realizar un análisis exploratorio de las variables dependientes e independientes de las personas que alguna vez incurrieron en mora.
- Evaluar los lenguajes de programación e IDE's licenciados y no licenciados para el análisis e interpretación de datos.
- Implementar el modelo para medir la probabilidad de incurrencia en mora utilizando las variables y metodología analizada.

CAPITULO II

ANTECEDENTES INVESTIGATIVOS

2.1 Estado del Arte

En este apartado se tomaron en cuenta ciertas investigaciones, las cuales sirvieron como base bibliográfica para el desarrollo del presente trabajo de investigación:

Brenes Humberto [9] en su obra titulada “Aplicación del análisis de regresión lineal simple para la estimación de los precios de las acciones de Facebook, Inc.” utiliza un algoritmo de regresión lineal bajo el método de Mínimos Cuadrados Ordinarios (MCO) con el fin de pronosticar el promedio mensual de las acciones de la empresa Facebook. Para la realización de la ecuación utilizó el valor de las acciones de Facebook en el periodo de mayo del 2012 hasta octubre del 2017 como variable independiente(x) y el tiempo en meses como variable dependiente(y). En la definición de la significancia de su regresión de una sola variable utilizó la prueba estadística F con n-2 grados de libertad y nivel de confianza de 90% con el fin de asegurar que las variables en estudio estén linealmente relacionadas, además de estadísticos t en la constante y pendiente.

Llaugel Felipe & Fernández Ana [10] en su trabajo titulado “Evaluación del uso de modelos de regresión logística para el diagnóstico de instituciones financieras” realizan un estudio de la viabilidad financiera de los múltiples bancos del sector financiero dominicano, con la aplicación regresión logística y basándose en los datos contables que estas instituciones reportan diariamente a la Superintendencia de Bancos. La elección de este tipo de regresión se debió a que la variable respuesta, que era dicotómica o binaria, es decir podía tener dos posibles resultados, entendiendo como 1 como bancos colapsados y 0 bancos no colapsados; para las variables independientes se tomaron indicadores financieros de cada institución, $I7$ (Cartera de Créditos Vigentes + Inversiones) / Activos, $I19$ (Disponibilidades+ Inversiones / Total Activos) e $I25$ (Activos Productivos / Total Activos). La regresión fue implementada con ayuda del software estadístico SPSS versión 11.5, llegando a un 98.9% de precisión según las pruebas estadísticas incluidas en el mismo paquete de software, posibilitando una nueva manera para la evaluación de entidades supervisadas de manera automática, sin deslegitimar un estudio más profundo por parte de los analistas.

Jiménez García, Alvarado Jesús & Blanco Jiménez [11] realizaron una investigación denominada “Estimation of vineyard leaf area by linear regression” en donde estudiaron 6 viñedos, ubicados en el centro de la ciudad de España, la investigación se realizó con el fin de evaluar el potencial productivo e identificar los rangos lumínicos y térmicos predominantes en la producción de uvas, a la vez de validar el método que habían venido utilizando en sus producciones anuales. Para tal efecto diseñaron un modelo matemático para estimar los factores ideales que conllevan a la crianza de la planta y crecimiento de la uva, las variables analizadas fueron el área floral, la medida de los brotes primarios y laterales de la planta. Se tomaron al azar 25 plantas de cada viñedo y se utilizaron métodos no destructivos, aunque este significó mayores costos y utilización de protocolos de calibración y muestreos específicos. La duración del estudio fue de tres años en donde se registraron los valores de las medidas de las plantas, además de factores climáticos, dando como resultado el modelo matemático que ayudó a tener una mejor comprensión y optimización de los diferentes métodos que los campesinos utilizaban para la siembra de este tipo de plantas.

Toscano Orlando & Benalcázar Freddy [12] realizaron una investigación denominada “Modelo matemático para determinar la calidad de servicio en el transporte público urbano en la ciudad de Ambato”, donde se trata de percibir la satisfacción de un usuario del transporte público urbano, a través de la construcción de un modelo de regresión binomial. La variable dependiente fue calidad de servicio y las variables predictoras fueron tiempo de espera, trato al usuario, estado de las unidades de transporte, forma de conducción del transportista, todas estas variables resultaron ser de tipo dicotómica, es decir solo tenían dos posibles valores (0 o 1 para el modelo). Para un adecuado tratamiento del modelo, se sometió a las variables en estudio a pruebas para determinar si existían relación entre ellas y la variable predictora. De acuerdo a la curva ROC se logró determinar la frontera de decisión de 61.42 %, el cual es el umbral para determinar si el servicio de transporte urbano es seguro o inseguro, y con la ayuda de la matriz de confusión y pruebas estadísticas realizadas se logró concluir que el modelo clasifica correctamente 308 observaciones, de las cuales 205 observaciones son clasificadas como seguras representando el 51,25% de los datos totales, mientras 103 observaciones fueron clasificadas como inseguras, que equivalen al 25,75%.

Sarría Almeida Shirley Patricia [13] en obra titula “Probabilidad de incurrir en mora de los usuarios de una tarjeta de crédito marca propia en Cali”, realiza un análisis estadístico y técnico acerca de los factores socioeconómicos que pueden incidir en el no pago de tarjetas de crédito marca propia de una compañía de financiamiento comercial en la ciudad de Cali. En el trabajo menciona cuestiones subyacentes a este tipo de ámbito financiero (concesión de tarjetas de créditos), donde menciona que la intermediación financiera es clave para el desarrollo y cada vez presenta financiaciones y ampliación de su margen de acción, un ejemplo de ello, son las tarjetas marca propia que se dirigen a muchos sectores poblacionales en especial los estratos medios, medios bajos y bajos; además de estimar que la probabilidad que el prestatario cancele su obligación está relacionada con el monto aprobado, ya que el reembolso prometido $(1 + r)B$ disminuye a medida que aumenta la tasa de préstamo. Concluyendo en la comparación de modelos de regresión logit y probit con sus respectivas métricas de evaluación y la utilización en su totalidad de variables categóricas para estimar los factores de riesgo que propician la incurrencia en mora por parte de los tarjetahabientes.

Fei Zhou & Yu Ting [14] en su estudio clínico ‘Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study’, realiza un análisis retrospectivo de los factores clínicos que incidieron en la mortalidad de pacientes adultos hospitalizados con COVID-19 en Wuhan, a los primeros días de diciembre de 2019, en la ciudad de Wuhan (capital de la provincia de Hubei en China); donde brotó una neumonía desconocida hasta entonces y que fue designada como COVID-2019 en febrero de 2020 por la OMS (Organización Mundial de la Salud). Para el estudio se incluyeron 191 pacientes, 135 del hospital de Jinyintan y 20 del hospital Wuhan Pulmonary, de los cuales 137 fueron dados de alta y 54 fallecieron en el hospital. Con la utilización de la regresión multivariable se demostró las probabilidades crecientes de muerte hospitalaria asociadas con la edad avanzada (odds ratio 1.10, IC del 95 % 1.03–1.17, aumento por año; $p=0.0043$), fallos multisistémica de órganos vitales (5.65, 2.61–12.23; $p<0.0001$), además la duración de la excreción viral (0-20 días), pero el SARS-CoV-2 fue detectable hasta la muerte en los no sobrevivientes.

Stéphane Laurent [15] en su trabajo titulado ‘Aortic stiffness is an independent

predictor of all-cause and cardiovascular mortality in hypertensive patients’, establece una relación directa entre la rigidez arterial y la mortalidad por hipertensión arterial. Utiliza un modelo univariado de análisis de regresión logística a 1980 pacientes que asistieron a la consulta externa de hipertensión del Hospital de Broussais entre 1980 y 1996, con quienes se realizó una medición de la rigidez arterial. Los resultados arrojaron que la rigidez arterial está asociada significativamente con la mortalidad cardiovascular y por todas las causas en un modelo univariado de análisis de regresión logística (odds ratio para la VOP de 5 m/s fue 2,14 [intervalo de confianza del 95 %, 1,71 a 2,67, $P < 0,0001$] y 2,35 [intervalo de confianza del 95 %] intervalo, 1,76 a 3,14, $P < 0,0001$], respectivamente). En cambio, con la aplicación de un modelo multivariado de análisis de regresión logística, la rigidez arterial se asoció significativamente con la mortalidad por todas las causas, independientemente de la enfermedad cardiovascular previa, la edad, y la diabetes.

Andrew Levey, Lesley Stevens & Schmid Christopher [16] en su obra titulada “A New Equation to Estimate Glomerular Filtration Rate”, desarrollaron una nueva ecuación de estimación de tasa de filtración glomerular (TFG), utilizando regresión lineal de mínimos cuadrados para relacionar la TFG medida con la creatinina sérica y las características clínicas disponibles en todas las bases de datos. Las variables predictoras incluyeron creatinina sérica, edad, raza (negra frente a blanca y otra) y sexo en todos los modelos, y variables adicionales en algunos modelos (diabetes [sí/no], trasplante previo de órganos [sí /no], y peso, según lo asignado por los estudios individuales). El estudio fue financiado por un acuerdo de cooperación con el Instituto Nacional de Diabetes y Enfermedades Digestivas y Renales, lo que les permite una participación sustancial en el diseño del estudio y en la recopilación, análisis e interpretación de los datos. R en su versión 2 y SAS versión 9.1 arrojaron una diferencia mediana mejorada (sesgo), IQR, P_{30} , y la raíz del error cuadrático medio (PAGS-0.001 para todos), consistentes entre los estudios y subgrupos definidos por edad, sexo, raza, estado de trasplante e índice de masa corporal.

Ochoa Juan, Galeano Galeano & Agudelo Luis Gabriel [17] en su obra titulada ‘Construcción de un modelo de scoring para el otorgamiento de crédito en un a entidad financiera’, donde indican que por las regulaciones de Colombia acerca de la implementación de estrategias que garanticen la solidez de las instituciones

financieras, se disponen a construir un modelo de scoring para el otorgamiento de crédito mediante una metodología de análisis discriminante. La base de datos que utilizaron (créditos de 24.786 personas, con fecha de corte en noviembre de 2009), fue facilitada por una cooperativa financiera de mediano tamaño, y para la ordenación de las variables y cálculos necesarios se utilizó el paquete estadístico SPSS en su versión número 18. La variable explicada o variable dependiente, resultó ser el número de días en el que la institución financiera considera que se entra en mora. Entre las variables independientes para el modelo propuesto están: Oficina, Categoría, Monto, Garantía, Reestructurado, Edad, Ocupación, etc.; como se verifica existe una variedad de variables, entre continuas y categóricas. El producto de la investigación desembocó en el diseño de una ecuación que disminuye en gran medida los riesgos que pudiera sufrir la cartera, desde el otorgamiento hasta el aseguramiento de liquidez en las instituciones financieras, asegurando el crecimiento de este sector.

CAPITULO III

MARCO METOLOGICO

3.1 Ubicación

El presente trabajo de investigación se realizó en la Cooperativa de Ahorro y Crédito Chibuleo Ltda. y su oficina matriz ubicada entre las calles Espejo 12-78 y Av. 12 de noviembre, ciudad de Ambato, provincia de Tungurahua.

3.2 Equipos y materiales

Una laptop MSI Core 7 con conexión a internet para consulta de documentación teórica y cursos de actualización. Motor de base de base SQL Server, donde se encontraban los datos a analizar, lenguaje Transact SQL para la extracción de datos sobre las bases de datos. Herramientas ofimáticas para la manipulación y limpieza de datos, previo ingreso a modelo diseñado. Para el desarrollo del modelo se utilizó el lenguaje de programación R y su IDE (Integrated Development Environment) RStudio versión 2020.02.1.

3.3 Tipo de investigación

Investigación aplicada

La investigación fue de tipo aplicada debido a que los resultados arrojados en la presente investigación serán de utilidad para la toma de decisión en la concesión de créditos y además las ecuaciones resultantes serán el punto base para aplicar modelos matemáticos más complejos dentro del ciclo de vida de cartera de créditos.

Investigación exploratoria

La investigación fue de tipo exploratoria debido a que el investigador no disponía de una investigación similar y aplicada a la economía local, por lo que se revisaron las normas vigentes que la institución financiera acataba, además del estudio de los procesos donde se generan los datos prescindibles para el estudio.

Investigación Cuantitativa

La investigación que de tipo cualitativa debido que al iniciar la investigación se

recolectó datos históricos de la cartera de créditos de la institución financiera entre los periodos 2017 - 2021 mediante los cuales se realizaron análisis matemáticos y estadísticos.

Investigación Correlacional

La investigación fue de tipo correlacional debido a que se estudió la relación existente entre las variables descriptoras o variables independientes y la variable dependiente, también si la ausencia o presencia de ciertas variables independientes podría ocasionar cambios significativos en el modelo propuesto.

3.4 Prueba de Hipótesis

Un modelo de regresión lineal permite identificar y preponderar los factores que inciden en el incumplimiento de pagos de un préstamo por parte de un socio de la cooperativa.

3.5 Población

La población está constituida por los préstamos otorgados por parte de la cooperativa y de todas sus oficinas, comprendidas entre el periodo enero 2017 y diciembre 2021. En detalle a la información de los préstamos constan condiciones socioeconómicas de los prestatarios, así como el plazo en días, monto solicitado, días de atrasos, tasa, etc.

3.6 Recolección de información

3.6.1 Análisis y revisión de estructuras SEPS

Las estructuras enviadas por parte de cooperativa a la Superintendencia de Economía Popular y Solidaria (SEPS), será la fuente y suministro necesario para la realización del modelo de regresión lineal. Las estructuras se reportan periódicamente a este organismo de control y contiene información relacionada a las operaciones de cartera de crédito y contingentes, conforme a las resoluciones emitidas por parte Junta de Política y Regulación Monetaria y Financiera. Una representación gráfica de lo anteriormente indicado se representa en la figura 1.

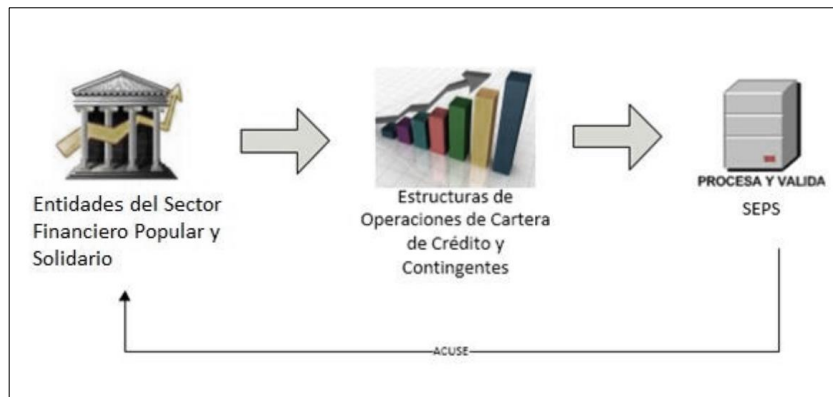


Figura 1. Esquema de envío de estructuras a SEPS
Fuente: Manual técnico SEPS

Cada estructura tiene un código y una razón de uso por parte de la SEPS, en la estructura C01 se reportan las operaciones de crédito concedidas por parte de la institución hasta la fecha corte, la estructura C02 con envío periódico contiene los saldos de operaciones de crédito y cancelaciones de aquellas operaciones que fueron reportadas anteriormente. La institución envía además otras estructuras relacionadas con los estados, servicios financieros, depósitos, socios, riesgo y liquidez. Para la presente investigación se tomará las estructuras que están relacionadas directamente con el ciclo de vida de las operaciones de créditos (C01 y C02).

3.6.2 Extracción de la información

Determinadas las variables a utilizarse en el estudio, se necesitó el diseño y conformación de la consulta en base de datos a ejecutarse en el correspondiente sistema de gestión de base de datos (SGBD), el cual es un SQL Server 2014. La consulta consistió en realizar uniones algebraicas entre la estructura C01 donde constan los datos con los que nace una operación de crédito, y la estructura C02 donde yacen el comportamiento de la mora, pagos e interés de estas, con utilización del número de la operación como factor común. Realizada la consulta de base de datos se procedió a ejecutarla en el motor de base de datos de la cooperativa, obteniéndose un total de 53887 registros y exportándola a un formato xlsx (Excel) para su posterior procesamiento y análisis. En primera instancia la presencia de variables categóricas (actividad económica, oficina concesión, tipo crédito, etc.) se presentan en mayor medida que variables continuas (tasa, monto solicitado, plazo en días, etc.).

3.7 Lenguajes de programación y software estadísticos

3.7.1 Python

Python es un lenguaje de scripting orientado a objetos y multiplataforma, preparado para realizar cualquier tipo de programa, desde servicios de red hasta aplicaciones Windows o incluso, páginas web. No necesita ser compilado para poder ejecutarlo, lo que ofrece ventajas como la rapidez de desarrollo. Posee una gran comunidad de programadores, debido a su sencillez y a sus amplias posibilidades, sobre todo las agregadas en los últimos años, ya que facilita trabajar con big data, inteligencia artificial, data science y machine learning, entre muchos otros campos en auge[18].

3.7.2 Anaconda

Es una de los IDE's que tiene Python para su utilización, además de paquetes para la aplicación de ciencia de datos, análisis predictivo, procesamiento de grandes volúmenes de información y machine learning. Es un administrador de entorno como un administrador de paquetes, lo que ayuda a los usuarios a preparar todo lo que necesitan en sus máquinas para sus aplicaciones y análisis de datos. Algunas de las herramientas que podemos encontrar en Anaconda son: JupyterLab, Spyder, PyCharm, entre otras[19].

3.7.3 Python y análisis estadísticos

El uso de Python en análisis e interpretación de datos es muy común entre los estadistas y programadores debido a los grandes ventajas que esta provee, como son: fácil de aprender, lo que significa que también es uno de los mejores para construir equipos más grandes de desarrolladores experimentados; flexible, se ejecuta en casi todas las plataformas, incluidas Windows y MacOS; eficiente, utiliza menos líneas que otros lenguajes para realizar la misma acción [20]. Las librerías utilizadas en analítica con Python son:

- 1) Pandas: El objetivo principal de esta biblioteca es manipular y analizar los datos con la ayuda de las funciones incluidas en ella. Un gran volumen de datos estructurados también se puede manejar con esta biblioteca.
- 2) SciPy: Se usa específicamente para llevar a cabo tareas de ciencia de datos, también es útil en el campo de la computación científica. Provee funcionalidades para resolver tareas de programación informática y problemas matemáticos científicos.
- 3) Matplotlib: La visualización de datos es importante para cualquier empresa. Esta

biblioteca no se limita a dibujar gráficos de barras, histogramas, circulares, sino que también es capaz de hacer figuras de alto nivel. [21]. En la figura 2 se muestra el ciclo de procesamiento y minería de datos.

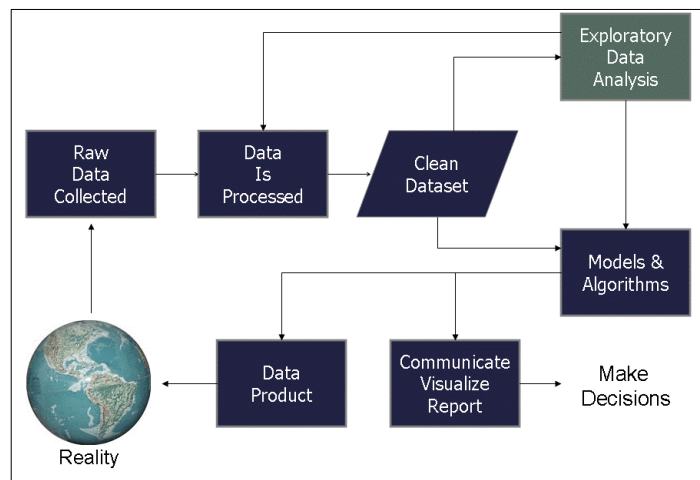


Figura 2. Mapa del proceso de ciencia de datos

Fuente: [21]

3.7.4 R

Es un entorno de software libre (licencia GNU GLP) y lenguaje de programación interpretado, es decir, ejecuta las instrucciones directamente, sin una previa compilación del programa a instrucciones en lenguaje máquina. El término entorno, en R, se refiere a un sistema totalmente planificado y coherente, en lugar de una acumulación de herramientas específicas e inflexibles, como suele ser el caso en otros softwares de análisis de datos [22].

3.7.2 RStudio

Es un entorno de desarrollo integrado que permite acceder a toda la potencia del lenguaje de programación R, dedicado al diseño de gráficos y a la computación estadística. Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo. R Studio está disponible para Linux, Mac y Windows. [23].

3.7.2 R vs Python

Para la presente investigación se analizó dos lenguajes de programación muy reconocidos en términos de análisis estadísticos, como son R y Python, ambos lenguajes presentan grandes ventajas en uno u otro ámbito. Para la elección de que software utilizar se realizó un cuadro comparativo en donde se calificó el desenvolvimiento de ambos lenguajes en distintos ámbitos. Las calificaciones

otorgadas oscilan en el rango de 1-5, siendo 5 la mejor calificación y 1 la calificación más baja. Un punto en donde R se colocó en primer lugar es donde se indica que R es un lenguaje estadístico utilizado para el análisis y representación visual de datos, este punto resulta útil en el estudio debido a que el enfoque es netamente estadístico y no se lo pretende integrar a algún otro sistema de la cooperativa, en caso contrario Python sería una mejor alternativa[24]. En cuanto librerías disponibles para análisis de datos, quien gana es R, que al ser un lenguaje creado específicamente para la estadística abarca muchos más conceptos y aristas[25]. Por las implicaciones del uso de variables categóricas, R es el candidato ideal, debido que se encarga de la creación de las variables sin dificultades, muy contrario a Python donde se necesita tratar este tipo de variables de manera especial, lo que conlleva tiempo y recursos.

Tabla 1. Cuadro comparativo entre lenguaje Python y R

Aplicaciones	Python es un lenguaje de propósito general que se utiliza para la implementación y el desarrollo de varios proyectos.	2
	R es un lenguaje estadístico utilizado para el análisis y representación visual de datos.	5
Machine learning	Python es el aprendizaje profundo y las aplicaciones web a gran escala.	4
	R es adecuado para el aprendizaje estadístico al tener bibliotecas poderosas para experimentar y explorar datos.	4
Bibliotecas disponibles	Python tiene muchas bibliotecas. Sin embargo, puede ser complejo entenderlos todos.	3
	R tiene menos bibliotecas en comparación con Python y es fácil de saber.	5
Aprendizaje	Python tiene una sintaxis simple y es fácil de aprender.	4
	R tiene una sintaxis relativamente compleja y la curva de aprendizaje no es sencilla.	3
Rendimiento	Los paquetes estadísticos de Python son menos potentes.	3
	Los paquetes estadísticos de R son muy potentes.	5
Integración	Python se utiliza principalmente cuando el análisis de datos debe integrarse con aplicaciones web.	4
	R generalmente se usa cuando la tarea de análisis de datos requiere computación (análisis) y procesamientos independientes.	4
Campo laboral	Desarrolladores e ingenieros de software.	4

	Matemáticos, estadistas.	4
Variables categóricas	En Python se debe especificar $k-n$ variables ficticias por cada nivel de una variable categórica.	2
	En R las variables categóricas no necesitan ser especificadas en niveles lo que disminuye la utilización de recursos.	5

Fuente: Elaborado por autor en base a investigación realizada.

De acuerdo con la tabla de puntuación de los leguajes de analítica, R alcanzó un puntaje de 35 frente a los 26 de Python, posicionándose en el lenguaje seleccionado para el diseño y desarrollo del modelo a implementarse. De acuerdo con el estudio anterior el procesamiento, análisis, y visualización de los datos disponibles se realizaría en RStudio versión 2022.02.1

3.8 Procesamiento de la información y análisis estadístico

El procesamiento de la información se define como el conjunto de actividades secuenciales y concretas mediante los cuales se pretende ordenar, almacenar y preparar los archivos con la información captada, asegurando su congruencia con el fin de proceder a su explotación para la presentación de resultados estadísticos [26].

La información fue extraída directamente de la base de datos (SQL Server) y posteriormente procesada en el paquete Microsoft Excel, en donde se realizó un primer análisis mediante gráficas y tabulaciones de cada variable categórica. Luego de lo cual se conformaría el data set en formato .csv para la carga en RStudio.

En el proceso de determinación de variables para la presente investigación se observó que por el giro de negocio (financiero), existen connotaciones bien marcadas en cuanto a características de un préstamo, como son destino financiero, actividad económica, lugar de concesión, etc., por lo que se decidió tomar las mismas como variables categóricas.

Modelo regresión lineal

La regresión lineal es un método estadístico que trata de modelar la relación entre una variable continua y una o más variables independientes mediante el ajuste de una ecuación lineal. Dependiendo del contexto, a la variable modelada se le conoce como variable dependiente o variable respuesta, y a las variables independientes como

regresores, predictores o features.

El modelo de regresión lineal (Legendre, Gauss, Galton y Pearson) considera que, dado un conjunto de observaciones $\{Y_i, X_{i1}, \dots, X_{ip}\}_{i=1}^n$, la media μ de la variable respuesta y se relaciona de forma lineal con la o las variables regresoras $X_1 \dots X_p$ acorde a la ecuación[27]:

$$\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Modelo Logit (Regresión logística binaria)

El modelo Logit es un tipo de regresión lineal con la diferencia de que la variable dependiente es dicotómica, es decir solo puede tener dos posibles valores (generalmente toma valores 1,0). El modelo es frecuentemente utilizado en situaciones prácticas en donde la respuesta puede tomar únicamente dos valores: 1, presencia (con probabilidad p); y 0, ausencia (con probabilidad $1-p$)[28].

El modelo logístico establece la siguiente relación entre la probabilidad de que ocurra el suceso, dado que el individuo presenta los valores $(X = x_1, X = x_2, \dots, X = x_k)$ [29]:

$$P[Y = 1/x_1, x_2, \dots, x_k] = \frac{1}{1 + e^{(-\beta_0 - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_k x_k)}}$$

Tomando logaritmos neperianos en la expresión anterior, se obtiene una expresión lineal para el modelo:

$$\text{Logit}[P(Y = 1)] = \text{Ln} \left[\frac{P[Y = 1/x_1, x_2, \dots, x_k]}{1 - P[Y = 1/x_1, x_2, \dots, x_k]} \right] = \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Variables artificiales o dummy

Las variables explicativas de tipo nominal con más de dos categorías deben ser incluidas en un modelo de regresión definiendo variables artificiales para cada nivel. Si una variable nominal (raza, religión, grupo sanguíneo, etc.) consta de k categorías deben crearse entonces $(k - 1)$ variables dicotómicas que son las llamadas variables artificiales asociadas a la variable nominal. Las $(k - 1)$ variables dicotómicas se denotan por $(Z_1, Z_2, \dots, Z_{k-1})$. A cada categoría o clase de la variable nominal le corresponde un conjunto de valores de los Z_i con el cual se identifica dicha clase.

La manera más usual de definir estas $(k - 1)$ variables es la siguiente: si el sujeto

pertenece a la primera categoría, entonces las $(k - 1)$ variables artificiales tiene el valor de 0 [29].

$$\text{Si } Z_1 = 1 \Rightarrow Z_2 = 0, Z_3 = 0, \dots, Z_{k-1} = 0$$

$$\text{Si } Z_2 = 1 \Rightarrow Z_1 = 0, Z_3 = 0, \dots, Z_{k-1} = 0$$

$$\text{Si } Z_3 = 1 \Rightarrow Z_1 = 0, Z_2 = 0, \dots, Z_{k-1} = 0$$

$$\text{Si } Z_{k-1} = 1 \Rightarrow Z_1 = 0, Z_2 = 0, Z_3 = 0, \dots, Z_{k-2} = 0$$

En esta línea, si la variable nominal de interés es el grupo sanguíneo (tipo 0, tipo A, tipo B, tipo AB), entonces se tendrían los siguientes valores de las variables artificiales para cada grupo sanguíneo:

Tabla 2. Ejemplo de uso de variables artificiales

Grupo sanguíneo	Z ₁	Z ₂	Z ₃
0	0	0	0
A	1	0	0
B	0	1	0
AB	0	0	1

Fuente: [29]

Si se ajusta un modelo que incluya una variable nominal con k clases, ésta debe ser sustituida por las $(k - 1)$ variables artificiales, y a cada una de ellas corresponderá su respectivo coeficiente.

3.9 Variables

Tabla 3. Descripción de variables utilizadas en el estudio

Nombre variable	Ítems	Hipótesis
Periodo de pago	1. Al vencimiento 2. Anual (360 DIAS) 3. Bimensual (60 DIAS) 4. Mensual (30 DIAS) 5. Quincenal (15 DIAS) 6. Semanal (7 DIAS) 7. Semestral (180 DIAS) 8. Trimestral (90 DIAS)	Se espera que la elección de elegir la frecuencia de pago refleja claras tendencias de no pago de cuotas.
Oficina concesión	1. Matriz 2. Machachi 3. Latacunga 4. Quito sur 5. Salcedo 6. Riobamba 7. Sangolquí 8. Pujilí 9. Pelileo 10. Otavalo 11. Ibarra 12. Tulcán 13. Cayambe 14. Mayorista 15. Izamba 16. Calderón 17. Quito centro 18. Duran 19. Milagro 20. Santo domingo 21. Guayaquil 22. Guaranda 23. Loja 24. La libertad 25. Lago agrio 26. Tena 27. Machala 28. Tarqui 29. Coca 30. Portoviejo 31. Manta 32. Babahoyo 33. Quevedo 34. La mana 35. Puyo	Se espera que algunas provincias del país reflejen mayor riesgo en el pago de cuotas.
Garantías	1. Garantías reales y/o financieras 2. Garantía personal de un tercero 3. Garantía personal de un tercero y garantía reales y/o financieras 4. Garantía quirografaria 5. Codeudores 6. Garantía personal de un tercero y codeudor	Se espera que personas que una garantía solida presenten menor probabilidad de no pago.

Tipo de crédito	1. Productivo corporativo 2. Productivo empresarial 3. Productivo pymes 4. Consumo 5. Educativo 6. Educativo Social 7. Vivienda interés público 8. Vivienda interés social 9. Inmobiliario 10. Microcrédito minorista 11. Microcrédito de acumulación simple 12. Microcrédito de acumulación ampliada	Se espera que ciertos tipos de crédito sean más propensos a recaer mora.
Destino financiero	1. Capital de Trabajo 2. Activos Fijos Tangibles: Terrenos, edificios y construcción de infraestructura, para fines productivos y comerciales 3. Activos Fijos Tangibles: Equipos, maquinaria y otros bienes de capital, a excepción de vehículos, para fines productivos y comerciales 4. Activos Fijos Tangibles: Vehículos pesados para fines productivos y comerciales 5. Activos Fijos Tangibles: Vehículos livianos para fines productivos y comerciales 6. Activos Fijos Tangibles: Vehículos livianos con combustibles fósiles 7. Activos Fijos Tangibles: Otros activos fijos tangibles que no constituyen bienes de capital, para fines productivos y comerciales 8. Activos Fijos Intangibles: Derechos de propiedad industrial. 9. Activos Fijos Intangibles: Adquisición de franquicias, marcas, pago de regalías, licencias y otros activos fijos intangibles. 10. Adquisición de servicios 11. Crédito directo otorgado a personas jurídicas no residentes para la adquisición de exportaciones producidos por residentes de la economía ecuatoriana 12. Reestructuración de Pasivos y Pago de Obligaciones 13. Microcrédito otorgado para consumo de microempresarios 14. Vivienda para microempresarios (Vivienda de interés público e inmobiliario destinado para la adquisición de vivienda de microempresarios) 15. No productivas (Para operaciones inmobiliarias, vivienda de interés público, consumo prioritario y crédito educativo, excepto vivienda de interés público e inmobiliario destinado a la adquisición de vivienda para microempresarios)	Se espera que ciertos destinos financieros no sean fiables para conceder préstamos.
Actividad económica receptora de la operación	1. No tiene actividad económica 2. Agricultura, ganadería, silvicultura y pesca. 3. Explotación de minas y canteras. 4. Industrias manufactureras. 5. Suministro de electricidad, gas, vapor y aire acondicionado. 6. Distribución de agua; alcantarillado, gestión de desechos y actividades de saneamiento. 7. Construcción. 8. Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas. 9. Transporte y almacenamiento. 10. Actividades de alojamiento y de servicio de comidas. 11. Información y comunicación. 12. Actividades financieras y de seguros. 13. Actividades inmobiliarias. 14. Actividades profesionales, científicas y técnicas. 15. Actividades de servicios administrativos y de apoyo. 16. Administración pública y defensa; planes de seguridad social de afiliación obligatoria. 17. Enseñanza. 18. Actividades de atención de la salud humana y de asistencia social. 19. Artes, entretenimiento y recreación. 20. Otras actividades de servicios. 21. Actividades de los hogares como empleadores; actividades no diferenciadas de los hogares como productores de bienes y servicios para uso propio. 22. Actividades de organizaciones y órganos extraterritoriales.	Se espera que ciertas actividades económicas no sean fiables para conceder préstamos.
Destino provincia	1. Matriz 2. Machachi 3. Latacunga 4. Quito Sur 5. Salcedo 6. Riobamba 7. Sangolquí 8. Pujilí 9. Pelileo 10. Otavalo 11. Ibarra 12. Tulcán 13. Cayambe 14. Mayorista 15. Izamba 16. Calderón 17. Quito Centro 18. Durán 19. Milagro 20. Santo Domingo 21. Guayaquil 22. Guaranda 23. Loja 24. La libertad 25. Lago Agrio 26. Tena 27. Machala 28. Tarqui 29. Coca 30. Portoviejo 31. Manta 32. Babahoyo 33. Quevedo 34. La Maná 35. Puyo	Se espera que algunas provincias del Ecuador presenten mayor riesgo en la concesión de préstamos que otras.
Nivel de estudios	1. Sin estudios 2. Primaria 3. Secundaria 4. Formación intermedia (técnica - Tecnología) 5. Universitaria 6. Postgrado	Se espera que personas con mayor formación académica presenten menor riesgo en la concesión de préstamos.

Tipo de vivienda	1. Propia no hipotecada 2. Propia hipotecada 3. Vive con familiares 4. Prestada Arrenda	Se espera que personas con casas propias presenten menor incidencia en mora en préstamos.
Actividad económica de titular	1. No tiene actividad económica 2. Agricultura, ganadería, silvicultura y pesca. 3. Explotación de minas y canteras. 4. Industrias manufactureras. 5. Suministro de electricidad, gas, vapor y aire acondicionado. 6. Distribución de agua; alcantarillado, gestión de desechos y actividades de saneamiento. 7. Construcción. 8. Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas. 9. Transporte y almacenamiento. 10. Actividades de alojamiento y de servicio de comidas. 11. Información y comunicación. 12. Actividades financieras y de seguros. 13. Actividades inmobiliarias. 14. Actividades profesionales, científicas y técnicas. 15. Actividades de servicios administrativos y de apoyo. 16. Administración pública y defensa; planes de seguridad social de afiliación obligatoria. 17. Enseñanza. 18. Actividades de atención de la salud humana y de asistencia social. 19. Artes, entretenimiento y recreación. 20. Otras actividades de servicios. 21. Actividades de los hogares como empleadores; actividades no diferenciadas de los hogares como productores de bienes y servicios para uso propio. 22. Actividades de organizaciones y órganos extraterritoriales.	Se espera que ciertas actividades económicas no sean fiables para conceder préstamos.
Género	1. Masculino 2. Femenino	Se espera encontrar tendencias en cuanto al género del titular en cuanto a la incidencia en mora de préstamos.
Incurrir en mora	1. Si (0) 2. No (1)	Variable dependiente, puede tomar solo dos valores: 1 (recae en mora), 0 (no recae en mora.)

Fuente: Elaborado por autor.

- 1) **Prioridad de pago.** - Indica la periodicidad con la que el sujeto deberá realizar los pagos de los dividendos del crédito
- 2) **Oficina concesión.** - Oficina de la entidad financiera en que se instrumentó la operación.
- 3) **Garantías.** - Determina el tipo de garantía para la operación de crédito concedida.
- 4) **Tipo de crédito.** - Identifica al tipo de crédito otorgado por la entidad, según normativa vigente.
- 5) **Destino financiero.** - Indica el uso financiero al cual está destinada la operación concedida.
- 6) **Actividad económica receptora de la operación.** - Identifica la actividad económica a la cual se destinarán realmente los fondos de la operación concedida

- 7) **Destino provincia.** - Provincia en el Ecuador hacia donde se destinan los fondos de la operación concedida.
- 8) **Actividad económica de titular.** - Actividad económica a la que se dedica el sujeto de riesgo.
- 9) **Nivel de estudios.** - Representa el nivel de educación del sujeto que se espera con el otorgamiento del crédito.
- 10) **Tipo de vivienda.** - Indica el tipo de vivienda en la que habita el sujeto.
- 11) **Género.** - Indica el género de titular de la operación.
- 12) **Incurre en mora.** – Representa si un préstamo incurrió en mora, a partir de los primeros 30 días de la primera cuota vencida. Variable respuesta o dependiente.

CAPITULO IV

RESULTADOS Y DISCUSIÓN

Para la extracción de datos se determinó que la consulta en base de datos estaría comprendida entre los años 2017 hasta 2021, obteniendo como resultado un data set de 53887 registros que contenían datos de préstamos morosos y no morosos.

En un primer análisis se identificó la cantidad de préstamos que habían caído en mora por año de operaciones que mantenía la cooperativa; a primera vista los préstamos morosos no fueron equiparables en ningún año a los préstamos no morosos, relativamente los analistas de crédito habían realizado un buen diagnóstico al momento de calificar si un sujeto podía ser acreedor a un producto crediticio.

Tabla 4. Cuadro de morosidad por periodos

Año	Morosos	No morosos	Préstamos	Proporción	% moroso
2017	1317	11876	13193	9,02	9,98
2018	1503	10887	12390	7,24	12,13
2019	1381	9589	10970	6,94	12,59
2020	841	7604	8445	9,04	9,96
2021	491	8398	8889	17,10	5,52
TOTAL	5533	48354	53887		

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

Como se indica en la tabla 4 el porcentaje de préstamos morosos con respecto a la totalidad de préstamos otorgados no supera el 12,59 % que fue el año 2019. El mejor año resulta el 2021 con un 5,52 % de morosidad. El año 2017 resultó un buen año con respecto a la colocación de cartera, unos 13193 préstamos otorgados y un porcentaje de morosidad no incluido entre los mayores porcentajes. La columna proporción indica el número de préstamos morosos por cada préstamo no moroso, una notable novedad es que el año 2021 está con la morosidad más baja, aproximadamente 17.10 préstamos morosos por cada préstamo no moroso, indicativo que quizá ese año no fue uno de los mejores y requiriendo de un análisis de mayor profundidad.

La figura 3 remarca lo anteriormente dicho acerca de la colocación de cartera, y haciendo notar que los préstamos morosos son mucho menores a los préstamos no morosos, lo que podría dificultar el diseño del modelo matemático, que al no tener un balance adecuado entre casos éxitos (préstamos morosos) y casos no éxitos (préstamos no morosos), dificulta identificar los casos exitosos y poder diseñar un modelo óptimo.

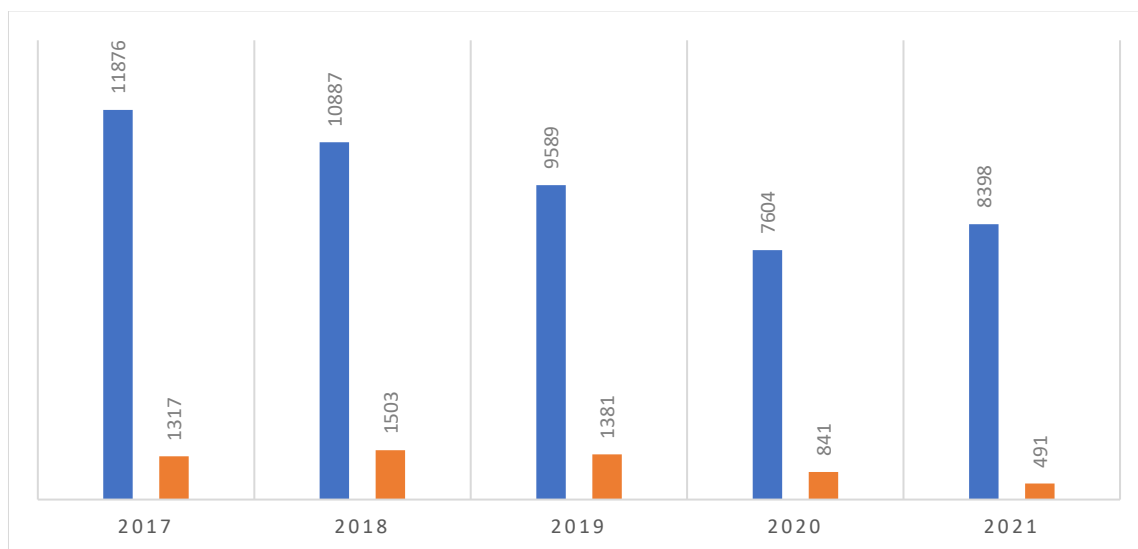


Figura 3. Gráfica de barras de préstamos morosos y no morosos por año.

Fuente: Elaborado por el investigador con los datos extraídos de la base de datos.

4.1 Análisis descriptivo de las variables

Periodicidad de pago. – De acuerdo con la tabla 5, el periodo comúnmente elegido por parte de los clientes de la cooperativa es ‘Mensual’ (30 días) con 85.38 % de los préstamos otorgados, en segundo lugar, se ubica la periodicidad ‘Al vencimiento’ con un 10.06 % de los préstamos. En cuanto morosidad por periodo de pago, ‘Anual’ tiene un número considerable con un 44.44 % pero con la mínima cantidad de préstamos otorgados, tan solo 9 préstamos, este dato se podría obviar por su baja ocurrencia.

Tabla 5. Cuadro de tabulación de variable periodicidad de pago

PERIODICIDAD PAGO	NO MOROSOS	MOROSO	PRÉSTAMOS	PORC. PRÉSTAMOS	PORC. MOR. POR PER. PAGO	POR. MOR. TOTAL
Al vencimiento	5171	2	5173	10,06%	0,04%	0,00%
Anual (360 DIAS)	5	4	9	0,02%	44,44%	0,01%
Bimensual (60 DIAS)	212	0	212	0,41%	0,00%	0,00%
Mensual (30 DIAS)	38928	4998	43926	85,38%	11,38%	9,72%
Quincenal (15 DIAS)	16	0	16	0,03%	0,00%	0,00%
Semanal (7 DIAS)	16	1	17	0,03%	5,88%	0,00%
Semestral (180 DIAS)	751	43	794	1,54%	5,42%	0,08%
Trimestral (90 DIAS)	1207	92	1299	2,52%	7,08%	0,18%
Total	46306	5140	51446	100,00%		

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

De acuerdo con la figura 4, se puede notar de mejor manera que el tipo de pago más frecuente es de tipo ‘Mensual’, que sobrepasa por mucho a los demás tipos de periodicidad. Por el motivo anterior el tipo de pago mensual también tiene la mayor tasa de morosidad.

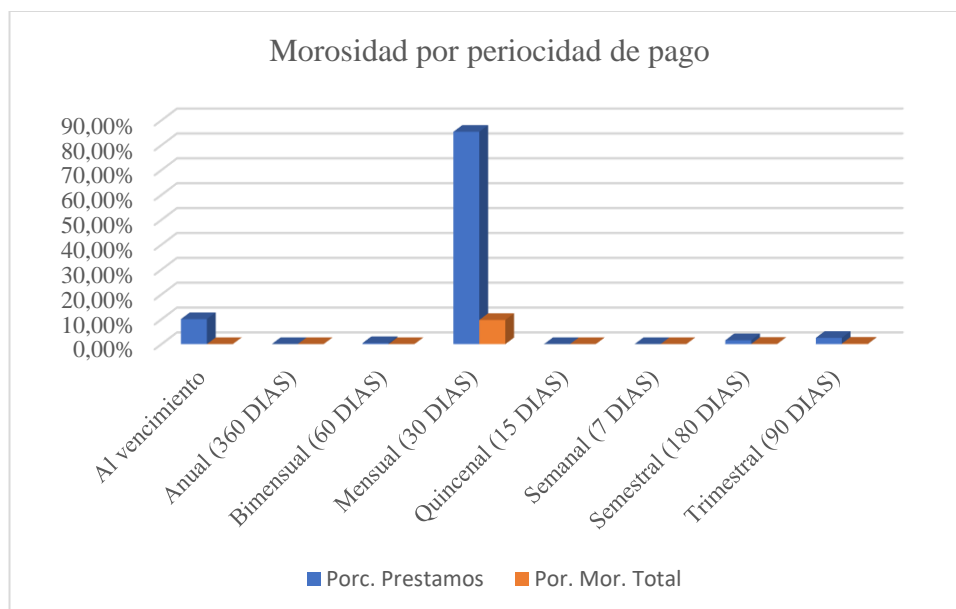


Figura 4. Gráfica de barras de morosidad por periodicidad de pago.

Fuente: Elaborado por el investigador con los datos extraídos de la base de datos.

Oficina concesión. – Segunda variable a analizar es la oficina en donde se generó la operación de crédito, para lo cual se realizó una estratificación de préstamos morosos y no morosos por oficina. En la tabla 6 se muestra los préstamos morosos y no morosos por oficina, además de su porcentaje en relación con los préstamos no morosos en esa oficina y total de préstamos morosos. ‘Mayorista’ es la oficina que presenta mayor mayoría por préstamos otorgados por oficina con un 21.40 % como consta en la columna Porc. Mor. Oficina (Porcentaje de morosidad por oficina), oficina ubicada en la ciudad de Ambato, sector de mercado mayorista. ‘Matriz’ presenta mayor morosidad por total de préstamos; no debería ser novedad debido a que, por la historia de fundación de la cooperativa, matriz es quien más tiempo ha estado en operación y mayores créditos a otorgado. ‘Machachi’ tiene un porcentaje considerable en préstamos morosos por oficina y total, con un 10.40% y 8.29% respectivamente.

Tabla 6. Cuadro de tabulación de variable oficina concesión

Oficina	No morosos	Moroso	Préstamos	Porc. Mor. Oficina	Porc. Mor. Total
Matriz	16208	1645	17853	9,21%	32,00%
Mayorista	1112	304	1416	21,47%	5,91%
Izamba	182	16	198	8,08%	0,31%

Machachi	3671	426	4097	10,40%	8,29%
Latacunga	4518	462	4980	9,28%	8,99%
Salcedo	2848	356	3204	11,11%	6,93%
Quito sur	3603	496	4099	12,10%	9,65%
Riobamba	4999	408	5407	7,55%	7,94%
Sangolquí	2210	263	2473	10,63%	5,12%
Pujilí	706	74	780	9,49%	1,44%
Pelileo	917	66	983	6,71%	1,28%
Otavaló	1393	229	1622	14,12%	4,46%
Ibarra	996	182	1178	15,45%	3,54%
Tulcán	943	112	1055	10,62%	2,18%
Cayambe	902	79	981	8,05%	1,54%
Calderón	282	2	284	0,70%	0,04%
Quito centro	69	1	70	1,43%	0,02%
Duran	149	4	153	2,61%	0,08%
Milagro	182	4	186	2,15%	0,08%
Santo Domingo	273	8	281	2,85%	0,16%
Guayaquil	143	3	146	2,05%	0,06%
		5140	51446		100,00%

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

En la figura 5 constan los préstamos morosos y no morosos por oficina. A parte de notar la baja morosidad que cada oficina, no se puede observar picos salvo levemente ‘Matriz’ en lo relacionado a morosidad lo que podría indicar a priori que la variable ‘Matriz’ en lo relacionado a morosidad lo que podría indicar a priori que la variable oficina concesión no podría ser un factor determinante en el proceso de realización del modelo.

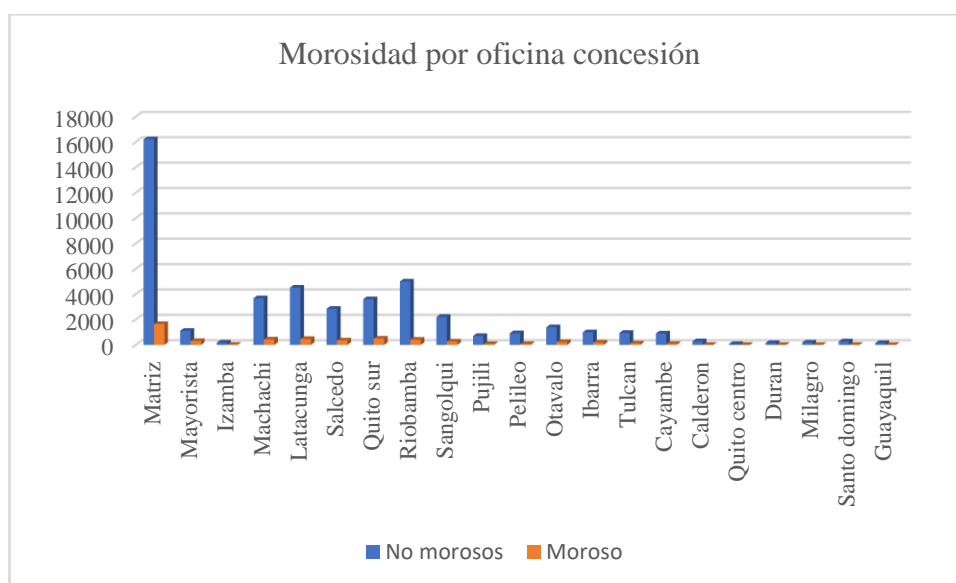


Figura 5. Gráfica de barras de morosidad por oficina concesión.

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

Garantías. – La variable garantía presenta un comportamiento diferente a la variable oficina concesión, tiene marcadas tendencias cuando la garantía se trata de ‘Garantías reales y/o financieras’, con un 5.18% de los préstamos totales. Pero eso puede deberse a que ese tipo de garantía está presente en un gran parte de los préstamos otorgados, con un 56.23% de los préstamos totales y por eso ese pico notable.

Tabla 7. Cuadro de tabulación de variable garantía

Oficina	No morosos	Morosos	Préstamos	Porc. Préstamos	Porc. Mor. Garantía	Por. Mor. Total
Garantías reales y/o financieras	16208	1645	17853	56,23%	9,21%	5,18%
Garantía personal de un tercero	1112	304	1416	4,46%	21,47%	0,96%
Garantía personal de un tercero y garantía reales y/o financieras	182	16	198	0,62%	8,08%	0,05%
Garantía quirografaria	3671	426	4097	12,90%	10,40%	1,34%
Codeudores	4518	462	4980	15,69%	9,28%	1,46%
Garantía personal de un tercero y codeudor	2848	356	3204	10,09%	11,11%	1,12%
Total		3209	31748	100,00%		

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

De acuerdo con la figura 6 se puede observar que la garantía ‘Garantía personal de un tercero’ se presenta en más ocasiones en Porc. Mor. Garantía (Porcentaje de morosidad por garantía) con un 21.47 %, sin embargo, tiene un porcentaje bajo de morosidad con respecto al número total de préstamos otorgados con un 0.05 %, un porcentaje menor con respecto a las demás garantías. Con el análisis realizado se denota que el tipo de garantía si es un factor determinante para la consecuencia de morosidad en préstamos. Es una consideración a tener en cuenta el diseño y desarrollo del modelado.

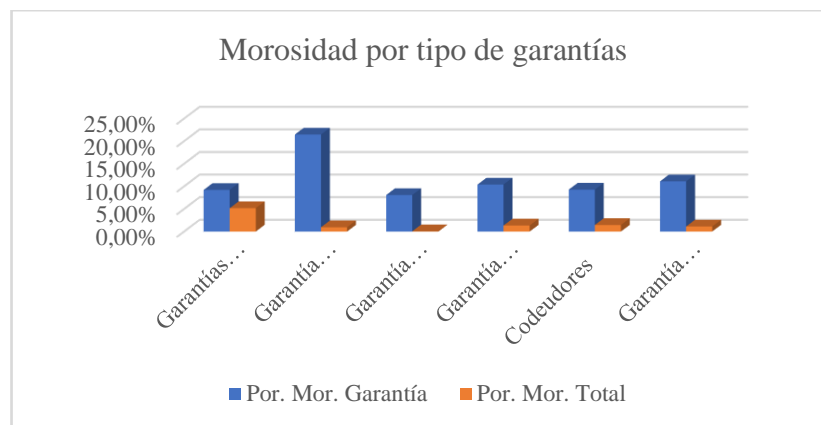


Figura 6. Gráfica de barras de morosidad por tipos de garantías.

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

Tabla 8. Cuadro de tabulación de variable tipo de crédito

Tipo de crédito	No morosos	Moroso	Préstamos	Porc. Préstamos	Por. Mor. Tipo de crédito	Por. Mor. Total
Consumo ordinario	2724	107	2831	5,50%	3,78%	0,21%
Consumo prioritario	17246	810	18056	35,10%	4,49%	1,57%
Microcrédito minorista	5505	381	5886	11,44%	6,47%	0,74%
Microcrédito de acumulación simple	18448	2866	21314	41,43%	13,45%	5,57%
Microcrédito de acumulación ampliada	2383	976	3359	6,53%	29,06%	1,90%
Total			51446	100,00%		

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

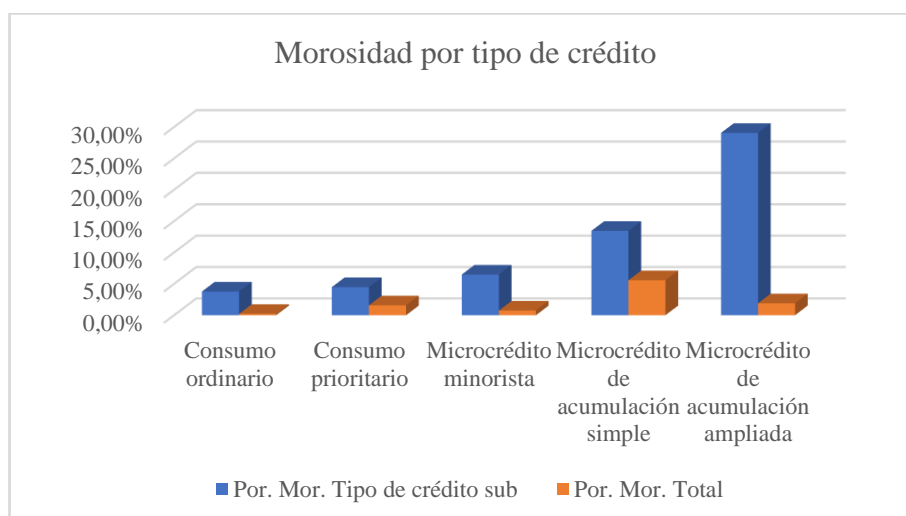


Figura 7. Gráfica de barras de morosidad por tipos de crédito.

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

Destino financiero. - De acuerdo con la tabla 9, ‘No productivas’ y ‘Capital de trabajo’ tienen la mayor cantidad de préstamos otorgados, con un 28.76 % y 28.19 % respectivamente, adicional, ‘Capital de trabajo’ presenta mayor porcentaje de morosidad por destino financiero, con 10.78 %. ‘Capital de trabajo’ se ubica en el primer puesto de acuerdo con morosidad total con un 3.04%, lo que indica que capital de trabajo no le hace honor a su nombre y no es una garantía de generador de riqueza.

Tabla 9. Cuadro de tabulación de variable destino financiero

Destino financiero	No morosos	Moroso	Préstamos	Porc. Préstamos	Porc. Mor. Por Des. Finan.	Por. Mor. Total
No productivas	14191	604	14795	28,76%	4,08%	1,17%
Capital de trabajo	12940	1564	14504	28,19%	10,78%	3,04%
Activos Fijos Tangibles	7971	1313	9284	18,05%	14,14%	2,55%
Activos Fijos Intangibles	301	75	376	0,73%	19,95%	0,15%
Adquisición de servicios	3221	403	3624	7,04%	11,12%	0,78%
Reestructuración de Pasivos y Pago de Obligaciones	5184	1041	6225	12,10%	16,72%	2,02%
Microcrédito otorgado para consumo	2498	140	2638	5,13%	5,31%	0,27%
Total			51446	100,00%		

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

Actividad económica receptora de la operación. - De acuerdo con la tabla 10, la actividad económica receptora de la operación con más presencia es ‘No tiene actividad económica’, con un sustancioso 40.60%. Esta actividad económica está ligado a los tipos de crédito ‘Consumo’, que tiene como destino la adquisición de bienes y servicios de consumo propio. En cuanto a morosidad por actividad económica receptora, se ignora la actividad ‘Actividades de organizaciones y órganos extraterritoriales’ por contar con 0.01% de préstamos otorgados, y se considera la actividad económica ‘Transporte y almacenamiento’ con un porcentaje de préstamos otorgados de 8.08% y 22.10% de morosidad por actividad económica. De acuerdo con esto ‘Transporte y almacenamiento’ sobresale en la actividad económica que presenta mayor riesgo de no cumplimiento de pagos.

Tabla 10. Cuadro de tabulación de variable actividad económica receptora de la operación

Actividad Económica Receptora	No morosos	Moroso	Préstamos	Porc. Préstamos	Porc. Mor. Por Act. Eco. Recep.	Porc. Mor. Total
No tiene actividad económica	19970	917	20887	40,60%	4,39%	1,78%
Agricultura, ganadería, silvicultura y pesca.	5474	718	6192	12,04%	11,60%	1,40%
Explotación de minas y canteras.	52	6	58	0,11%	10,34%	0,01%
Industrias manufactureras.	4733	704	5437	10,57%	12,95%	1,37%
Suministro de electricidad, gas, vapor y aire acondicionado.	69	14	83	0,16%	16,87%	0,03%
Distribución de agua; alcantarillado, gestión de desechos y actividades de saneamiento.	16	1	17	0,03%	5,88%	0,00%
Construcción.	2428	290	2718	5,28%	10,67%	0,56%
Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas.	5172	768	5940	11,55%	12,93%	1,49%
Transporte y almacenamiento.	3239	919	4158	8,08%	22,10%	1,79%
Actividades de alojamiento y de servicio de comidas.	991	163	1154	2,24%	14,12%	0,32%
Información y comunicación.	141	16	157	0,31%	10,19%	0,03%
Actividades financieras y de seguros.	99	11	110	0,21%	10,00%	0,02%
Actividades inmobiliarias.	142	31	173	0,34%	17,92%	0,06%

Actividades profesionales, científicas y técnicas.	688	110	798	1,55%	13,78%	0,21%
Actividades de servicios administrativos y de apoyo.	757	135	892	1,73%	15,13%	0,26%
Administración pública y defensa; planes de seguridad social de afiliación obligatoria.	134	13	147	0,29%	8,84%	0,03%
Enseñanza.	205	33	238	0,46%	13,87%	0,06%
Actividades de atención de la salud humana y de asistencia social.	220	26	246	0,48%	10,57%	0,05%
Artes, entretenimiento y recreación.	94	19	113	0,22%	16,81%	0,04%
Otras actividades de servicios.	826	131	957	1,86%	13,69%	0,25%
Actividades de los hogares como empleadores; actividades no diferenciadas de los hogares como productores de bienes y servicios para uso propio.	855	113	968	1,88%	11,67%	0,22%
Actividades de organizaciones y órganos extraterritoriales.	1	2	3	0,01%	66,67%	0,00%
Total	46306	5140	51446	100,00%		

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

Provincia destino. – Como se muestra en la tabla 11, la provincia donde más se han destinado los fondos por préstamos, resultó la provincia de ‘Tungurahua’ con un 39.37 %; segundo lugar le sigue ‘Pichincha’ con un 23.40 %. Las provincias con datos sobresalientes son ‘Cotopaxi’ y ‘Chimborazo’. ‘Cotopaxi’ con 17.17 % de préstamos otorgados, 9.87 % de morosidad por Provincia destino y 16.96 % de la morosidad total. ‘Chimborazo’ con 10.55 % de préstamos otorgados, 7.55 % de morosidad por Provincia destino y 7.98 % de la morosidad total. Los hallazgos anteriores hacen notar que la variable Provincia destino es un factor a tener en cuenta en la concesión de un préstamo e importante variable para el diseño de modelo propuesto.

Tabla 11. Cuadro de tabulación de variable provincia destino.

Provincia destino	No morosos	Moroso	Préstamos	Porc. Préstamos	Porc. Mor. Por Prov. Destino	Porc. Mor. Total
Tungurahua	18291	1963	20254	39,37%	9,69%	38,19%
Azuay	7	1	8	0,02%	12,50%	0,02%
Bolívar	86	32	118	0,23%	27,12%	0,62%
Carchi	955	110	1065	2,07%	10,33%	2,14%
Cotopaxi	7959	872	8831	17,17%	9,87%	16,96%
Chimborazo	5020	410	5430	10,55%	7,55%	7,98%
El Oro	0	1	1	0,00%	100,00%	0,02%
Esmeraldas	4	1	5	0,01%	20,00%	0,02%
Guayas	494	16	510	0,99%	3,14%	0,31%
Imbabura	2363	405	2768	5,38%	14,63%	7,88%
Loja	2	0	2	0,00%	0,00%	0,00%
Los Ríos	43	8	51	0,10%	15,69%	0,16%
Manabí	4	4	8	0,02%	50,00%	0,08%
Morona S.	2	4	6	0,01%	66,67%	0,08%
Napo	3	2	5	0,01%	40,00%	0,04%
Pastaza	13	2	15	0,03%	13,33%	0,04%
Pichincha	10759	1277	12036	23,40%	10,61%	24,84%
Galápagos	2	0	2	0,00%	0,00%	0,00%
Sucumbíos	5	9	14	0,03%	64,29%	0,18%
Orellana	1	0	1	0,00%	0,00%	0,00%
Santo Dom.	288	23	311	0,60%	7,40%	0,45%
Santa Elena	5	0	5	0,01%	0,00%	0,00%
Total	46306	5140	51446	100,00%		

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

De acuerdo con la figura 8 se puede notar que la morosidad de las provincias en relación con la morosidad total (Por. Mor. Total), está condicionada por la cantidad de préstamos por oficina en la relación a total de préstamos otorgados. Esto supone una

relación proporcional entre el número de préstamos otorgados y préstamos morosos; denota intrínsecamente que la variable provincia no es un factor para considerar en el diseño del modelo propuesto.

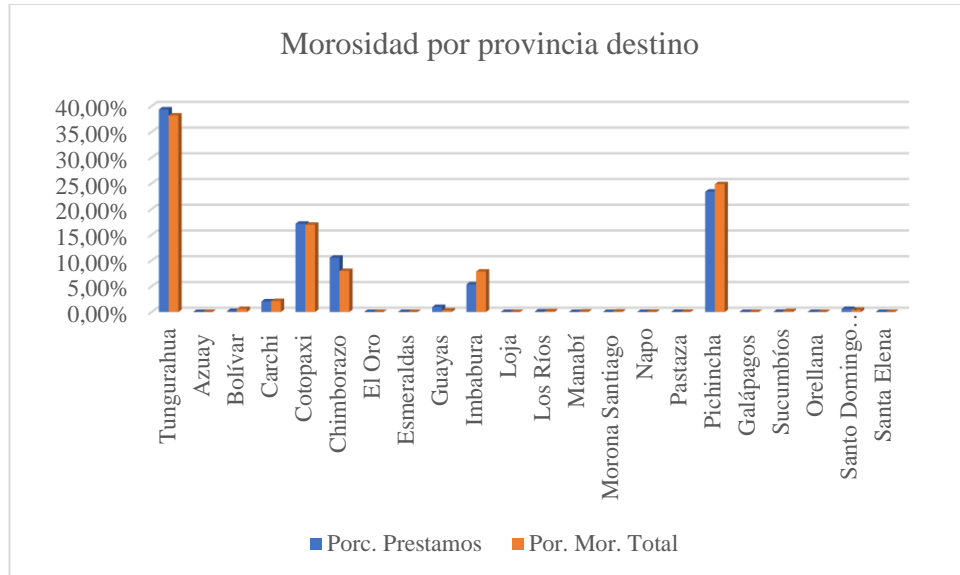


Figura 8. Gráfica de barras de morosidad por provincia destino.
Fuente: Elaborado por autor con los datos extraídos de la base de datos.

Actividad económica de Titular. – De acuerdo con la tabla 12, la actividad económica de titular con más presencia en los préstamos otorgados es ‘Actividad no económica’ con un 28.83 %, en segundo lugar, le sigue ‘Comercio al por mayor y al por menor’ con 14.78%. Lo anterior indica que la mayor parte recursos de los préstamos está dirigido a compra de bienes y servicios sin aparente fin productivo como es el caso de las microempresas. En cuanto a morosidad total en los primeros lugares su ubica ‘Actividad no económica’ con 19,73% y ‘Transporte y almacenamiento’ con 19.74%; la presencia nuevamente de ‘Actividad no económica’ es debido a que existen más préstamos con esta característica y por eso también más morosidad.

Tabla 12. Cuadro de tabulación de variable actividad económica de titular.

Actividad económica de Titular	No morosos	Moroso	Préstamos	Porc. Préstamos	Porc. Mor. Por Act. Titular	Porc. Mor. Total
Actividad no económica	13817	1014	14831	28,83%	6,84%	19,73%
Agricultura, ganadería, silvicultura y pesca.	6240	688	6928	13,47%	9,93%	13,39%
Explotación de minas y canteras.	65	5	70	0,14%	7,14%	0,10%
Industrias manufactureras.	6068	718	6786	13,19%	10,58%	13,97%
Suministro de electricidad, gas, vapor y aire acondicionado.	61	12	73	0,14%	16,44%	0,23%
Distribución de agua	20	1	21	0,04%	4,76%	0,02%
Construcción.	2757	291	3048	5,92%	9,55%	5,66%
Comercio al por mayor y al por menor	6832	774	7606	14,78%	10,18%	15,06%
Transporte y almacenamiento.	3582	881	4463	8,68%	19,74%	17,14%
Actividades de alojamiento y de servicio de comidas.	1161	145	1306	2,54%	11,10%	2,82%
Información y comunicación.	159	11	170	0,33%	6,47%	0,21%
Actividades financieras y de seguros.	161	13	174	0,34%	7,47%	0,25%
Actividades inmobiliarias.	186	30	216	0,42%	13,89%	0,58%
Actividades profesionales, científicas y técnicas.	1222	125	1347	2,62%	9,28%	2,43%
Actividades de servicios administrativos y de apoyo.	987	155	1142	2,22%	13,57%	3,02%
Administración pública y defensa	253	20	273	0,53%	7,33%	0,39%
Enseñanza.	526	38	564	1,10%	6,74%	0,74%
Actividades de atención de la salud humana y de asistencia.	324	25	349	0,68%	7,16%	0,49%
Artes, entretenimiento y recreación.	110	19	129	0,25%	14,73%	0,37%
Otras actividades de servicios.	728	88	816	1,59%	10,78%	1,71%
Actividades de los hogares como empleadores	1039	87	1126	2,19%	7,73%	1,69%
Actividades de organizaciones y órganos extraterritoriales.	8	0	8	0,02%	0,00%	0,00%
Total	46306	5140	51446	100,00%		

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

De acuerdo con figura 9 y valores de tabla 12, las actividades económicas que presentan novedad son: ‘Comercio al por mayor y al por menor’ que tiene 14.78% de los préstamos otorgados y 15.06% de morosidad total; ‘Transporte y almacenamiento’ tiene 8.68% de los préstamos otorgados y 17.14% de morosidad total. El estudio anterior indica que la actividad económica tiene relevancia y condiciona en el no pago de las obligaciones de crédito.

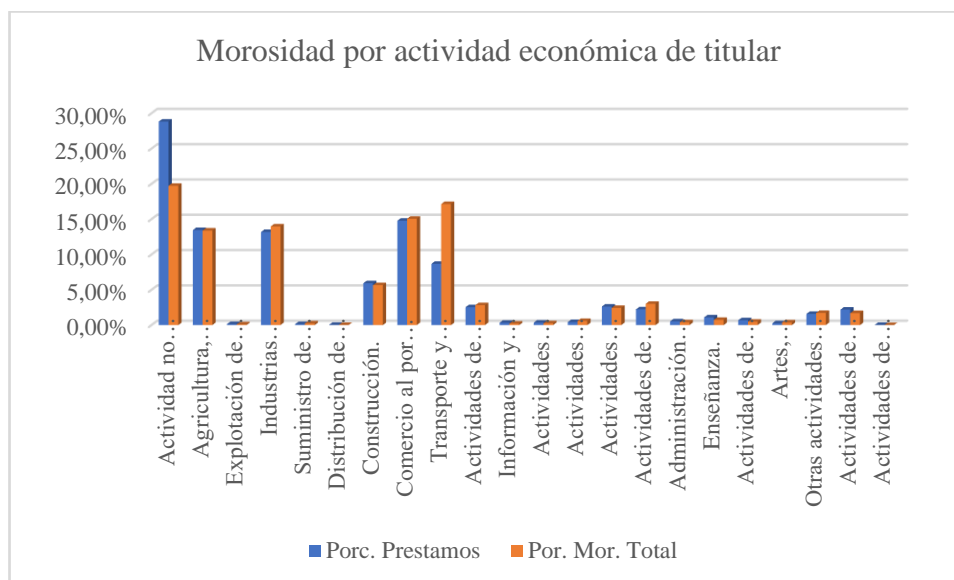


Figura 9. Gráfica de barras de morosidad por actividad económica de titular.
Fuente: Elaborado por autor con los datos extraídos de la base de datos.

Nivel de estudios. - De acuerdo con la tabla 13, la mayor cantidad de préstamos conocidos ha sido a personas con nivel de educación ‘Primaria’, con un 46.27 %, en segundo lugar, está el nivel de educación ‘Secundaria’, con un 36.65 %. El mayor porcentaje de morosidad por nivel de educación es ‘Sin estudios’, con un porcentaje de 10.77 % y le siguen ‘Primaria’ y ‘Secundaria’ con 10.36% y 10.44% respectivamente. Los niveles de educación con mayor morosidad global son ‘Primaria’ y ‘Secundaria’, ‘Primaria’ en primer lugar con 47.96% y ‘Secundaria’ con 38.29%. Los niveles de educación con relativa baja en los porcentajes de morosidad global, son ‘universitaria’ con 10.43 % y ‘postgrado’ con 0.02%. Lo anterior hace notar que las personas con mayor formación profesional (universitaria, postgrado) tienden a cumplir con sus obligaciones a cabalidad, contrario lo que sucede del otro extremo; sin duda la preparación académica da mejores oportunidades de prosperidad.

Tabla 13. Cuadro de tabulación de variable nivel de estudio.

Nivel de estudios	No morosos	Moroso	Préstamos	Porc. Préstamos	Por. Mor. Por Nivel Estudio	Por. Mor. Total
Sin estudios	580	70	650	1,26%	10,77%	1,36%
Primaria	21337	2465	23802	46,27%	10,36%	47,96%
Secundaria	16887	1968	18855	36,65%	10,44%	38,29%
Formación intermedia (técnica - Tecnología)	1087	100	1187	2,31%	8,42%	1,95%
Universitaria	6398	536	6934	13,48%	7,73%	10,43%
Postgrado	17	1	18	0,03%	5,56%	0,02%
Total	46306	5140	51446			

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

Tipo vivienda. –De acuerdo con la tabla 14, la mayor cantidad de préstamos otorgados fue concedido a personas con casa ‘Propia no hipotecada’ con un 52.27 %, en segundo lugar, esta ‘Vive con familiares’ con 36.73%. Por el contrario, la menor cantidad de préstamos otorgados fue realizado a personas con casas prestadas, como lo representa el 4.39%. En cuanto a la morosidad total se puede observar que ‘Propia no hipotecada’ y ‘Vive con familiares’ con un 50.47 % y 37.92 % respectivamente, estos lugares pudieran ser debido a la proporcionalidad con la cantidad de préstamos otorgados.

Tabla 14. Cuadro de tabulación de variable tipo de vivienda.

Tipo de vivienda	No morosos	Moroso	Préstamos	Porc. Préstamos	Por. Mor. Por Tipo Viv.	Por. Mor. Total
Propia hipotecada	866	108	974	1,89%	11,09%	2,10%
Propia no hipotecada	24297	2594	26891	52,27%	9,65%	50,47%
Arrendada	4001	480	4481	8,71%	10,71%	9,34%
Prestada	196	9	205	0,40%	4,39%	0,18%
Vive con familiares	16946	1949	18895	36,73%	10,31%	37,92%
Total	46306	5140	51446	100,00%		

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

De acuerdo con la figura 10, se puede apreciar de mejor manera que ‘Propia no hipotecada’ y ‘Vive con familiares’ son más frecuentes tanto en préstamos morosos como en préstamos no morosos. Se puede notar que la cantidad de préstamos morosos en cuanto a ‘Tipo de vivienda’ es proporcional al número de préstamos otorgados.

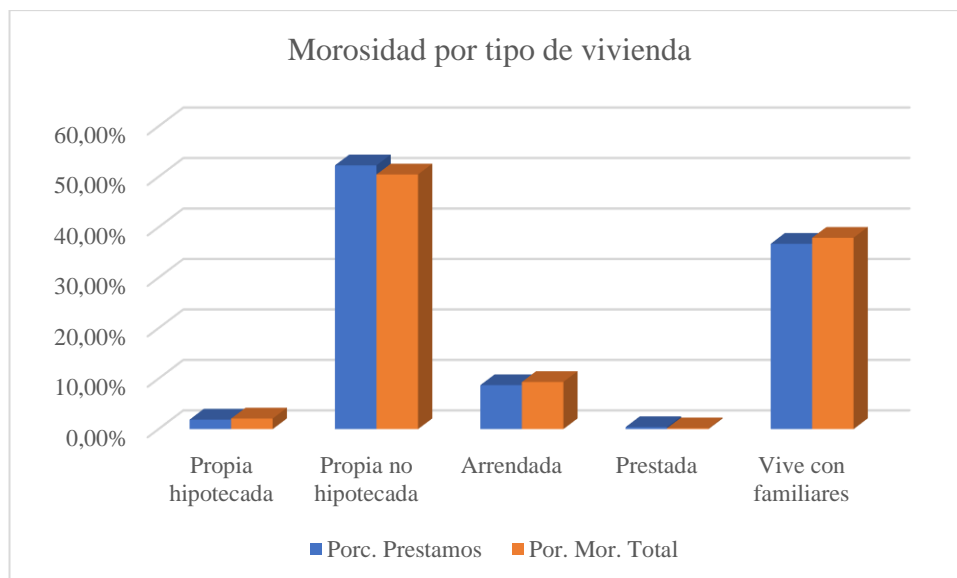


Figura 10. Gráfica de barras de morosidad por tipo de vivienda.

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

Género. - De acuerdo con la tabla 15 y gráfico 11, mayoritariamente los préstamos fueron otorgados a hombres con 53.05 % de los préstamos. Le antecede el género femenino con un 46.95% de préstamos. Con respecto a morosidad total, se puede observar que el género masculino se presenta en primer lugar, con 58.87% sobre 41.13% del género femenino. Lo anterior se puede deberse al volumen de préstamos de hombres.

Tabla 15. Cuadro de tabulación de variable género.

Género	No morosos	Moroso	Préstamos	Porc. Préstamos	Por. Mor. Género	Por. Mor. Total
Masculino	24264	3026	27290	53,05%	11,09%	58,87%
Femenino	22042	2114	24156	46,95%	8,75%	41,13%
Total	46306	5140	51446	100,00%		

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

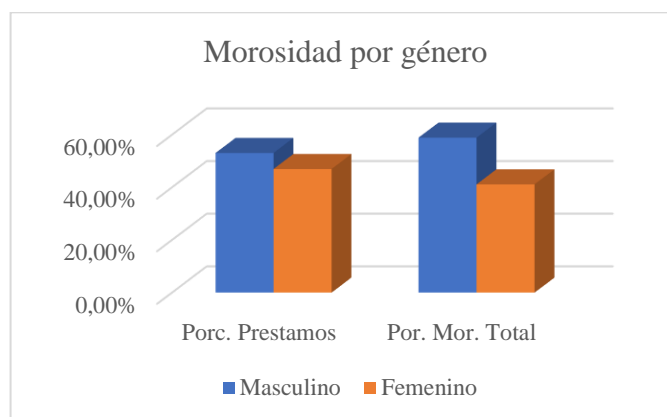


Figura 11. Gráfica de barras de morosidad por género.

Fuente: Elaborado por autor con los datos extraídos de la base de datos.

4.2 Modelación matemática

Luego de la generación del set de datos con las variables antes analizadas se procedió a la codificación y posterior generación del modelo utilizando RStudio, arrojando los siguientes resultados.

```

Call:
glm(formula = prest_moros ~ perio_pago + ofic_concecion + garantia +
    tipo_credito + destino_financiero + act_econo_recep_oper +
    provinc_destino + act_econo_titular + nivel_estud + tipo_viviend +
    genero, family = "binomial", data = datoscategoria)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8328  -0.4917  -0.3672  -0.2328   4.0256

Coefficients: (2 not defined because of singularities)
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -10.930136   1.504819  -7.263 3.77e-13 ***
perio_pagoAnual (360 DIAS)\n      -6.485731   386.914968  -0.017 0.986626
perio_pagoBimensual (60 DIAS)    3.545338    0.874469   4.054 5.03e-05 ***
perio_pagoMensual (30 DIAS)\n    4.888857    0.710053   6.885 5.77e-12 ***
perio_pagoQuincenal (15 DIAS)   -7.313578   209.854194  -0.035 0.972199
perio_pagoSemana (7 DIAS)       4.750091    1.257395   3.778 0.000158 ***
perio_pagoSemestral (180 DIAS)  3.843968    0.728453   5.277 1.31e-07 ***
perio_pagoTrimestral (90 DIAS)\n 4.434544    0.718501   6.172 6.75e-10 ***
ofic_concecionCAYAMBE            2.312601    0.722482   3.201 0.001370 **
ofic_concecionDURAN              1.243902    1.054052   1.180 0.237954
ofic_concecionGUAYAYUIL          0.917277    1.091449   0.840 0.400672
ofic_concecionIBARRA             3.554164    0.801446   4.435 9.22e-06 ***
ofic_concecionIZAMBA             2.431379    0.788107   3.085 0.002035 **
ofic_concecionLATACUNGA          2.850982    0.741209   3.846 0.000120 ***
ofic_concecionMACHACHI           2.657789    0.714143   3.722 0.000198 ***
ofic_concecionMATRIZ             2.801841    0.737926   3.797 0.000147 ***
ofic_concecionMAYORISTA          3.592087    0.738926   4.861 1.17e-06 ***
ofic_concecionMILAGRO            0.940327    1.054832   0.891 0.372689
ofic_concecionNOTAVALO           3.472111    0.799090   4.345 1.39e-05 ***
ofic_concecionPELILEO            2.351411    0.749486   3.137 0.001705 **
ofic_concecionPUJILI             2.805475    0.749934   3.741 0.000183 ***
ofic_concecionQUITO CENTRO       0.777944    1.236838   0.629 0.529363
ofic_concecionQUITO SUR          2.851231    0.713896   3.994 6.50e-05 ***
ofic_concecionRIOBAMBA           2.752122    0.845475   3.255 0.001133 **
ofic_concecionSALCEDO            3.024175    0.740737   4.083 4.45e-05 ***
ofic_concecionSANGOLQUI          2.938398    0.715463   4.107 4.01e-05 ***
ofic_concecionSANTO DOMINGO      0.576368    0.877534   0.657 0.511307
ofic_concecionTULCAN             3.951464    0.948769   4.165 3.12e-05 ***
garantiagC                       0.515162    0.052521   9.809 < 2e-16 ***
garantiagP                       0.714174    0.051870  13.769 < 2e-16 ***
garantiagQ                       0.354124    0.052481   6.748 1.50e-11 ***
garantiagR                       -0.065979    0.098518  -0.670 0.503043
garantiagT                       0.638363    0.087259   7.316 2.56e-13 ***
tipo_creditoMA                    2.106801    0.247232   8.522 < 2e-16 ***
tipo_creditoMM                    0.921364    0.249640   3.691 0.000224 ***
tipo_creditoMS                    1.614142    0.244438   6.603 4.02e-11 ***
tipo_creditoON                    1.173461    0.259569   4.521 6.16e-06 ***
tipo_creditoPN                    0.593692    0.213467   2.781 0.005416 **
provinc_destinoBolivar            -0.362887    1.109546  -0.327 0.743623
provinc_destinoChimborazo        -2.234704    1.264906  -1.767 0.077279
provinc_destinoCotopaxi          -1.277824    1.166308  -1.096 0.273248
provinc_destinoCotopaxi          -1.078666    1.094933  -0.985 0.324554
provinc_destinoEl Oro            14.748027   882.744065  0.017 0.986670
provinc_destinoEsmeraldas        0.949451    1.619707   0.586 0.557750
provinc_destinoGallapagos        -12.894225   623.682982  -0.021 0.983505
provinc_destinoGuayas            -0.624079    1.221336  -0.511 0.609365
provinc_destinoImbabura          -1.275715    1.156373  -1.103 0.269939
provinc_destinoLoja              -12.733960   616.989236  -0.021 0.983534
provinc_destinoLos Rios          -1.055984    1.172044  -0.901 0.367601
provinc_destinoManabí            0.474459    1.423369   0.333 0.738881
provinc_destinoMorona Santiago   0.870773    1.407484   0.619 0.536132
provinc_destinoNapó              -0.557561    1.456269  -0.383 0.701817
provinc_destinoOrellana          -13.309685   882.744052  -0.015 0.987970
provinc_destinoPastaza           -0.232103    1.357053  -0.171 0.864197
provinc_destinoPichincha         -0.809251    1.101617  -0.735 0.462581
provinc_destinoSanta Elena       -12.755667   382.821707  -0.033 0.973419
provinc_destinoSanto Domingo De Los Tsáchilas -0.017722    1.160044  -0.015 0.987811 **
provinc_destinoSucumbios         0.359864    1.232716   0.292 0.770342
provinc_destinoTungurahua        -0.945771    1.086517  -0.870 0.384049 **
destino_financieroRP              0.865963    0.070975  12.201 < 2e-16 ***
destino_financieroTE              0.224798    0.158284   1.420 0.155544
destino_financieroTF              NA           NA           NA           NA
destino_financieroTL              0.225846    0.075509   2.991 0.002781 **
destino_financieroTO              0.161374    0.108180   1.492 0.135772
destino_financieroTP              0.694726    0.117734   5.901 3.62e-09 ***
destino_financieroTT              -0.093470    0.085590  -1.092 0.274802
act_econo_recep_operB             -0.710301    0.943825  -0.753 0.451704
act_econo_recep_operC             0.055290    0.129666   0.426 0.669812
act_econo_recep_operD             0.079788    0.475537   0.168 0.866752
act_econo_recep_operE             -0.221632    1.296785  -0.171 0.864296
act_econo_recep_operF             -0.118985    0.191862  -0.620 0.535154
act_econo_recep_operG             0.021840    0.132246   0.165 0.868825
act_econo_recep_operH             0.078665    0.142740   0.551 0.581560
act_econo_recep_operI             0.300032    0.193442   1.551 0.120897
act_econo_recep_operJ             -0.091218    0.373946  -0.244 0.807281
act_econo_recep_operK             -0.289857    0.381242  -0.760 0.447077
act_econo_recep_operL             0.210495    0.401552   0.524 0.600137
act_econo_recep_operM             0.072231    0.185632   0.389 0.699558
act_econo_recep_operO             -0.415039    0.188048  -2.207 0.027307 *
act_econo_recep_operO             -0.603206    0.346379  -1.741 0.081602
act_econo_recep_operP             0.352461    0.244634   1.441 0.149650
act_econo_recep_operQ             -0.138828    0.287486  -0.483 0.629163
act_econo_recep_operR             -0.016935    0.457413  -0.037 0.970466
act_econo_recep_operS             0.046490    0.163179   0.285 0.775717
act_econo_recep_operT             0.296818    0.182676   1.625 0.104198
act_econo_recep_operU             2.727743    1.269385   2.149 0.031645 *

```

act_econo_titularL68	0.376528	0.405820	0.928	0.353501
act_econo_titularM69	-0.262466	0.330571	-0.794	0.427208
act_econo_titularM70	0.331337	0.307011	1.079	0.280483
act_econo_titularM71	0.299033	0.515359	0.580	0.561752
act_econo_titularM72	-0.007652	0.199802	-0.038	0.969450
act_econo_titularM73	0.891330	0.474859	1.877	0.060512
act_econo_titularM74	-0.192019	0.259303	-0.741	0.458984
act_econo_titularM75	0.177720	0.815864	0.218	0.827561
act_econo_titularN77	0.960002	0.232879	4.122	3.75e-05 ***
act_econo_titularN78	0.505795	0.772962	0.654	0.512880
act_econo_titularN79	0.922124	0.451347	2.043	0.041048 *
act_econo_titularN80	-1.136840	0.729420	-1.559	0.119102
act_econo_titularN81	0.146057	0.650570	0.225	0.822364
act_econo_titularN82	0.381739	0.188960	2.020	0.043361 *
act_econo_titularO84	-0.087659	0.279413	-0.314	0.753730
act_econo_titularP85	-0.185023	0.225421	-0.821	0.411767
act_econo_titularQ86	-0.307048	0.316966	-0.969	0.332690
act_econo_titularQ87	0.114286	0.476131	0.240	0.810306
act_econo_titularQ88	0.511669	1.170614	0.437	0.662043
act_econo_titularR90	0.011979	0.480010	0.025	0.980091
act_econo_titularR91	15.098757	882.743500	0.017	0.986353
act_econo_titularR92	0.258908	1.198585	0.216	0.828979
act_econo_titularR93	0.928383	0.629646	1.474	0.140360
act_econo_titularS01	0.045930	0.177623	0.259	0.795960
act_econo_titularS02	-0.602549	0.359151	-1.678	0.093405
act_econo_titularS03	-0.059887	0.186948	-0.320	0.748709
act_econo_titularS04	-0.232689	0.128056	-1.817	0.069204
act_econo_titularS05	-0.095096	0.104387	-0.911	0.362301
act_econo_titularS94	0.181563	0.391607	0.464	0.642909
act_econo_titularS95	0.611724	0.272745	2.243	0.024907 *
act_econo_titularS96	-0.053978	0.208760	-0.259	0.795970
act_econo_titularT97	-0.384195	0.203620	-1.887	0.059184
act_econo_titularT98	-0.591313	0.391031	-1.512	0.130485
act_econo_titularU99	-11.001060	305.227016	-0.036	0.971249
nivel_estudPostgrado	-0.744673	1.074759	-0.693	0.488388
nivel_estudPrimaria	0.184243	0.112912	1.632	0.102735
nivel_estudSecundaria	0.198671	0.112834	1.761	0.078283
nivel_estudSin estudios\n	0.352502	0.173241	2.035	0.041876 *
nivel_estudUniversitaria	0.119188	0.120695	0.988	0.323390
tipo_viviendPrestada	-0.886932	0.350626	-2.530	0.011420 *
tipo_viviendProptia hipotecada	0.013209	0.121700	0.109	0.913567
tipo_viviendProptia no hipotecada	-0.160647	0.057987	-2.770	0.005599 **
tipo_viviendvive con familiares	-0.124405	0.057259	-2.173	0.029805 *
generoMASCULINO	0.082696	0.036256	2.281	0.022555 *

Figura 12. Resultados de primera corrida de modelo en R.

Fuente: Elaborado por autor con resultados de RStudio.

Una vez observados los coeficientes de regresión resultantes y de acuerdo con [30] que menciona que una correlación es significativa para un modelo si su p-valor es inferior a 0.05 con un nivel de confianza de 95%. El presente modelo propone que la mayor parte de las variables analizadas resulta ser significativa para el modelo, a excepción de ‘Actividad económica receptora de la operación de crédito’, ‘Provincia destino’, ‘Nivel de estudio’ y ‘Actividad económica del titular’. Como se menciona en [28], las variables ficticias necesariamente necesitan ser afectadas en su totalidad y no una parte de ella (para cada variable categórica), por lo que eliminar parcialmente las variables artificiales para una variable categórica no es una opción. A priori, este descarte de variables: ‘Actividad económica receptora de la operación de crédito’, ‘Provincia destino’ y ‘Actividad económica del titular’ nos indican que las mismas no son factores determinantes para la propensión de mora en créditos. No así, las variables ‘Prioridad de pago’, ‘Oficina concesión’, ‘Tipo de garantía’, ‘Tipo de crédito’, ‘Destino financiero’, ‘Tipo de vivienda’, ‘Genero’ que si presentan significancia para el presente modelo. Luego del descarte de variables se procede a correr nuevamente el script de generación del modelo, arrojando los nuevos resultados expuestos a continuación.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-11.6048447	1.0306137	-11.260	< 2e-16 ***
perio_pago[T. Anual (360 DIAS)\n]	-4.4195514	143.0637908	-0.031	0.975356
perio_pago[T. Bimensual (60 DIAS)]	3.5730703	0.8718871	4.098	4.17e-05 ***
perio_pago[T. Mensual (30 DIAS)\n]	4.8811482	0.7099723	6.875	6.19e-12 ***
perio_pago[T. Quincenal (15 DIAS)]	-5.3398439	76.6244139	-0.070	0.944442
perio_pago[T. Semanal (7 DIAS)]	4.779716	1.2585226	3.798	0.000146 ***
perio_pago[T. Semestral (180 DIAS)]	3.8283289	0.7279571	5.259	1.45e-07 ***
perio_pago[T. Trimestral (90 DIAS)\n]	4.4078478	0.7183730	6.136	8.47e-10 ***
ofic_concecion[T. CAYAMBE]	2.2409533	0.7214443	3.106	0.001895 **
ofic_concecion[T. DURAN]	1.4206944	0.8752456	1.623	0.104548
ofic_concecion[T. GUAYAQUIL]	1.0778040	0.9215054	1.170	0.242157
ofic_concecion[T. IBARRA]	3.0673441	0.7161828	4.283	1.84e-05 ***
ofic_concecion[T. IZAMBA]	2.2730708	0.7610864	2.987	0.002821 **
ofic_concecion[T. LATACUNGA]	2.5066316	0.7131804	3.515	0.000440 ***
ofic_concecion[T. MACHACHI]	2.6027951	0.7133555	3.649	0.000264 ***
ofic_concecion[T. MATRIZ]	2.6293850	0.7118781	3.694	0.000221 ***
ofic_concecion[T. MAYORISTA]	3.4255353	0.7147281	4.793	1.64e-06 ***
ofic_concecion[T. MILAGRO]	1.0871417	0.8743650	1.243	0.213739
ofic_concecion[T. OTAVALO]	3.0437036	0.7152290	4.256	2.09e-05 ***
ofic_concecion[T. PELILEO]	2.1791798	0.7232206	3.013	0.002585 **
ofic_concecion[T. PUJILI]	2.4557585	0.7225680	3.399	0.000677 ***
ofic_concecion[T. QUITO CENTRO]	0.6200861	1.2363404	0.502	0.615984
ofic_concecion[T. QUITO SUR]	2.8289689	0.7132006	3.967	7.29e-05 ***
ofic_concecion[T. RIOBAMBA]	2.2271800	0.7133824	3.122	0.001796 ***
ofic_concecion[T. SALCEDO]	2.6836664	0.7139046	3.759	0.000170 ***
ofic_concecion[T. SANGOLQUI]	2.8822704	0.7146655	4.033	5.51e-05 ***
ofic_concecion[T. SANTO DOMINGO]	1.3889922	0.7972896	1.679	0.093068
ofic_concecion[T. TULCAN]	2.5213122	0.7186242	3.509	0.000451 ***
garantia[T. GC]	0.5128765	0.0521882	9.827	< 2e-16 ***
garantia[T. GP]	0.7158416	0.0512438	13.969	< 2e-16 ***
garantia[T. GQ]	0.3591621	0.0520443	6.901	5.16e-12 ***
garantia[T. GR]	-0.0090095	0.0958146	-0.094	0.925085
garantia[T. GT]	0.6913230	0.0855474	8.081	6.41e-16 ***
tipo_credito[T. MA]	2.2873995	0.2274572	10.056	< 2e-16 ***
tipo_credito[T. MM]	1.0498460	0.2302124	4.560	5.11e-06 ***
tipo_credito[T. MS]	1.7540427	0.2249409	7.798	6.30e-15 ***
tipo_credito[T. ON]	1.1584295	0.2584609	4.482	7.39e-06 ***
tipo_credito[T. PN]	0.5874673	0.2127026	2.762	0.005746 **
destino_financiero[T. CT]	0.1250328	0.0640371	1.953	0.050878
destino_financiero[T. IF]	0.6141214	0.1451278	4.232	2.32e-05 ***
destino_financiero[T. MC]	0.3059926	0.1142116	2.679	0.007380 **
destino_financiero[T. OT]	0.3765519	0.1080923	3.484	0.000495 ***
destino_financiero[T. RP]	0.8859648	0.0704313	12.579	< 2e-16 ***
destino_financiero[T. TE]	0.2461447	0.1568429	1.569	0.116562
destino_financiero[T. TL]	0.2364375	0.0748610	3.158	0.001587 **
destino_financiero[T. TO]	0.1449709	0.1075065	1.348	0.177502
destino_financiero[T. TP]	0.7923754	0.1145947	6.915	4.69e-12 ***
destino_financiero[T. TT]	-0.1223955	0.0850577	-1.439	0.150159
tipo_viviend[T. Prestada]	-0.9331389	0.3490701	-2.673	0.007513 **
tipo_viviend[T. Propia hipotecada]	-0.0001839	0.1206557	-0.002	0.998784
tipo_viviend[T. Propia no hipotecada]	-0.1531863	0.0571088	-2.682	0.007311 **
tipo_viviend[T. Vive con familiares]	-0.1400243	0.0566590	-2.471	0.013460 **
genero[T. MASCULINO]	0.0906977	0.0321081	2.825	0.004732 **

Figura 13. Resultados de segunda corrida de modelo en R.

Fuente: Elaborado por autor con resultados de RStudio

Interpretación de resultados

El coeficiente estimado de la intersección, que a la vez es el valor del logaritmo de odds de que un préstamo recaiga en mora teniendo en cuenta que todas las variables artificiales mostradas en la figura 13 toman el valor de 0, equivalente a un odds de -11.6048447, aplicando la inversa del logaritmo natural $e^{-11.6048447} = 9.122e^{-6}$, lo que se corresponde a una probabilidad de $p = \frac{e^{-11.6048447}}{1+e^{-11.6048447}} = 9.122e^{-6}$. En síntesis, la probabilidad de que un préstamo recaiga en mora cuando el valor de una variable categórica toma el valor por default (establecida por R), es de $9.122e^{-6}$.

Acorde al modelo, el logaritmo de los odds de que un préstamo tenga pagos quincenales esta negativamente relacionado con el no pago del préstamo (coeficiente de regresión = -5.3398439). Esto denota que cuanto esta variable artificial posea el valor de 1 (sea elegida entre los demás tipos de periodicidad de pagos), el logaritmo de odds de la variable morosidad (variable dependiente) se decremente en promedio

0.00479 unidades. Aplicando la inversa del logaritmo natural $e^{-5.3398439} = 0,00479$ unidades.

El coeficiente de regresión de ‘Quito centro’ 0.620086, siendo la menor cantidad registrada en cuanto a variable ‘Oficina concesión’, además está relacionado positivamente con el cumplimiento de pagos de un préstamo. El logaritmo de odds se incrementa en solo 0.620086 unidades cuando esta variable artificial toma el valor de 1. Por otro lado, aplicando la inversa del logaritmo natural $e^{0.620086} = 1,85908$, lo que indica que cuando esta variable esta activa (tiene valor de 1), los odds se incrementan en promedio 1,85908 unidades.

La ecuación resultante se formula de la siguiente manera:

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots + \beta_k X_k$$

$$Z = \ln(\text{odds})$$

$$\begin{aligned} \ln(\text{odds}) = & \beta_0 + \beta_1(\text{Perio_Anual}) + \beta_2(\text{Perio_Bimensual}) + \beta_3(\text{Perio_Mensual}) \\ & + \beta_4(\text{Perio_Quincenal}) + \beta_5(\text{Perio_Semanal}) + \beta_6(\text{Perio_Semestral}) \\ & + \beta_7(\text{Perio_Trimesral}) + \beta_8(\text{Oficina_Concesion_Cayambe}) \\ & + \beta_9(\text{Oficina_Concesion_Duran}) + \beta_{10}(\text{Oficina_Concesion_Guayaquil}) \\ & + \beta_{11}(\text{Oficina_Concesion_Ibarra}) + \beta_{12}(\text{Oficina_Concesion_Izamba}) \\ & + \beta_{13}(\text{Oficina_Concesion_Latacunga}) + \beta_{14}(\text{Oficina_Concesion_Machachi}) \\ & + \beta_{15}(\text{Oficina_Concesion_Matriz}) + \beta_{16}(\text{Oficina_Concesion_Mayorista}) \\ & + \beta_{17}(\text{Oficina_Concesion_Milagro}) + \beta_{18}(\text{Oficina_Concesion_Otavalo}) \\ & + \beta_{19}(\text{Oficina_Concesion_Pelileo}) + \beta_{20}(\text{Oficina_Concesion_Pujili}) \\ & + \beta_{21}(\text{Oficina_Concesion_Quito_Centro}) + \beta_{22}(\text{Oficina_Quito_Sur}) \\ & + \beta_{23}(\text{Oficina_Concesion_Riobamba}) + \beta_{24}(\text{Oficina_Concesion_Salcedo}) \\ & + \beta_{25}(\text{Oficina_Concesion_Sanngolqui}) + \beta_{26}(\text{Oficina_Concesion_Santo_Domingo}) \\ & + \beta_{27}(\text{Oficina_Concesion_Tulcan}) + \beta_{28}(\text{Garantia_GC}) + \beta_{29}(\text{Garantia_GP}) \\ & + \beta_{30}(\text{Garantia_GQ}) + \beta_{31}(\text{Garantia_GR}) + \beta_{32}(\text{Garantia_GT}) + \beta_{33}(\text{Credito_MA}) \\ & + \beta_{34}(\text{Credito_MM}) + \beta_{35}(\text{Credito_MS}) + \beta_{36}(\text{Credito_ON}) + \beta_{37}(\text{Credito_PN}) \\ & + \beta_{38}(\text{Destino_Financiero_CT}) + \beta_{39}(\text{Destino_Financiero_IF}) \\ & + \beta_{40}(\text{Destino_Financiero_MC}) + \beta_{41}(\text{Destino_Financiero_OT}) \\ & + \beta_{42}(\text{Destino_Financiero_RP}) + \beta_{43}(\text{Destino_Financiero_TE}) \\ & + \beta_{44}(\text{Destino_Financiero_TL}) + \beta_{45}(\text{Destino_Financiero_TO}) \\ & + \beta_{46}(\text{Destino_Financiero_TP}) + \beta_{47}(\text{Destino_Financiero_TT}) \\ & + \beta_{48}(\text{Tipo_Vivienda_Prestada}) + \beta_{49}(\text{Tipo_Vivienda_Propia_Hipotecada}) \\ & + \beta_{50}(\text{Tipo_Vivienda_Propia_No_Hipotecada}) \\ & + \beta_{51}(\text{Tipo_Vivienda_Vive_Con_Familiares}) + \beta_{52}(\text{Genero}) \end{aligned}$$

$$\begin{aligned} \ln(\text{odds}) = & -11,6048447 - 4,4195514(\text{Perio_Anual}) + 3,5730703(\text{Perio_Bimensual}) \\ & + 4,8811482(\text{Perio_Mensual}) - 5,3398439(\text{Perio_Quincenal}) \\ & + 4,7797716(\text{Perio_Semanal}) + 3,8283289(\text{Perio_Semestral}) \\ & + 4,4078478(\text{Perio_Trimesral}) + 2,2409533(\text{Oficina_Concesion_Cayambe}) \\ & + 1,4206944(\text{Oficina_Concesion_Duran}) + 1,077804(\text{Oficina_Concesion_Guayaquil}) \\ & + 3,0673441(\text{Oficina_Concesion_Ibarra}) + 2,2730708(\text{Oficina_Concesion_Izamba}) \\ & + 2,5066316(\text{Oficina_Concesion_Latacunga}) \\ & + 2,6027951(\text{Oficina_Concesion_Machachi}) + 2,629385(\text{Oficina_Concesion_Matriz}) \\ & + 3,4255353(\text{Oficina_Concesion_Mayorista}) \\ & + 1,0871417(\text{Oficina_Concesion_Milagro}) \\ & + 3,0437036(\text{Oficina_Concesion_Otavalo}) + 2,1791798(\text{Oficina_Concesion_Pelileo}) \\ & + 2,4557585(\text{Oficina_Concesion_Pujili}) \end{aligned}$$

$$\begin{aligned}
&+0,6200861(\text{Oficina_Concesion_Quito_Centro}) + 2,8289689(\text{Oficina_Quito_Sur}) \\
&+ 2,22718(\text{Oficina_Concesion_Riobamba}) + 2,6836664(\text{Oficina_Concesion_Salcedo}) \\
&+ 2,8822704(\text{Oficina_Concesion_Sangolqui}) \\
&+ 1,3389922(\text{Oficina_Concesion_Santo_Domingo}) \\
&+ 2,5213122(\text{Oficina_Concesion_Tulcan}) + 0,5128765(\text{Garantia_GC}) \\
&+ 0,7158416(\text{Garantia_GP}) + 0,3591621(\text{Garantia_GQ}) - 0,0090095(\text{Garantia_GR}) \\
&+ 0,691323(\text{Garantia_GT}) + 2,2873995(\text{Credito_MA}) + 1,049846(\text{Credito_MM}) \\
&+ 1,7540427(\text{Credito_MS}) + 1,1584295(\text{Credito_ON}) + 0,5874673(\text{Credito_PN}) \\
&+ 0,1250328(\text{Destino_Financiero_CT}) + 0,6141214(\text{Destino_Financiero_IF}) \\
&+ 0,3059926(\text{Destino_Financiero_MC}) + 0,3765519(\text{Destino_Financiero_OT}) \\
&+ 0,8859648(\text{Destino_Financiero_RP}) + 0,2461447(\text{Destino_Financiero_TE}) \\
&+ 0,2364375(\text{Destino_Financiero_TL}) + 0,1449709(\text{Destino_Financiero_TO}) \\
&+ 0,7923754(\text{Destino_Financiero_TP}) + -0,1223955(\text{Destino_Financiero_TT}) \\
&- 0,9331389(\text{Tipo_Vivienda_Prestada}) \\
&- 0,0001839(\text{Tipo_Vivienda_Propia_Hipotecada}) \\
&- 0,1531863(\text{Tipo_Vivienda_Propia_No_Hipotecada}) \\
&- 0,1400243(\text{Tipo_Vivienda_Vive_Con_Familiares}) + 0,0906977(\text{Genero})
\end{aligned}$$

El cálculo de la probabilidad de recaída en mora de un préstamo con pagos mensuales, en oficina mayorista, con garantía personal, para un tipo de crédito microcrédito de acumulación simple, con destino financiero microcrédito otorgado para consumo de microempresas, con tipo de vivienda propia hipotecada y para una persona de género masculino se definiría de la siguiente manera:

$$\begin{aligned}
\ln(odds) = &-11,6048447 - 4,4195514(0) + 3,5730703(0) + 4,8811482(1) - 5,3398439(0) \\
&+ 4,7797716(0) + 3,8283289(0) + 4,4078478(0) + 2,2409533(0) + 1,4206944(0) \\
&+ 1,077804(0) + 3,0673441(0) + 2,2730708(0) + 2,5066316(0) + 2,6027951(0) \\
&+ 2,629385(0) + 3,4255353(1) + 1,0871417(0) + 3,0437036(0) + 2,1791798(0) \\
&+ 2,4557585(0) + 0,6200861(0) + 2,8289689(0) + 2,22718(0) + 2,6836664(0) \\
&+ 2,8822704(0) + 1,3389922(0) + 2,5213122(0) + 0,5128765(0) + 0,7158416(0) \\
&+ 0,3591621(0) - 0,0090095(0) + 0,691323(1) + 2,2873995(0) + 1,049846(0) \\
&+ 1,7540427(1) + 1,1584295(0) + 0,5874673(0) + 0,1250328(0) + 0,6141214(0) \\
&+ 0,3059926(1) + 0,3765519(0) + 0,8859648(0) + 0,2461447(0) + 0,2364375(0) \\
&+ 0,1449709(0) + 0,7923754(0) + -0,1223955(0) - 0,9331389(0) - 0,0001839(1) \\
&- 0,1531863(0) - 0,1400243(0) + 0,0906977(1)
\end{aligned}$$

$$\begin{aligned}
\ln(odds) = &-11,6048447 + 4,8811482(1) + 3,4255353(1) + 0,691323(1) + 1,7540427(1) \\
&+ 0,3059926(1) - 0,0001839(1) + 0,0906977(1)
\end{aligned}$$

$$\ln(odds) = -0,4562891$$

$$Odd = \text{antilog}(-0,4562891) = e^{-0,4562891} = 0,633630628$$

$$p = \frac{odds}{1 + odds}$$

$$p = \frac{0,633630628}{1,633630628} = 0,38$$

La probabilidad de recaída de mora en un préstamo con las características antes mencionadas es de un $p = 0.38$

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

En el proceso de estudio de variables se pudo notar notables sesgos, debido a que existían mayor número de individuos con esas características dentro de la población. Para estos casos es necesario aplicar mecanismos para equipar las características de los individuos para construir un modelo confiable, además de la aplicación de pruebas de testeo especializadas.

El análisis de las variables arrojó conclusiones relevantes acerca de las características intrínsecas del nicho de mercado donde la institución se desenvuelve. El destino financiero más frecuente por la que solicita un préstamo es 'Capital de trabajo' y 'Activos Fijos Tangibles'. La mayoría de los socios de la cooperativa solo ha culminado la educación primaria y es el segmento que tiene la mayor cantidad de préstamos otorgados.

Los factores dominantes para la recaída de mora en créditos según la valoración inicial en el modelado son 'Prioridad de pago', 'Oficina concesión', 'Tipo de garantía', 'Tipo de crédito', 'Destino financiero', 'Tipo de vivienda' y 'Genero'. Comprobando que las variables que inciden en la mora de crédito no necesariamente pueden ser continuas (tasa nominal, valor de la operación, plazo en días, tasa efectiva anual, score crediticio, etc.), lo que posibilita un nuevo campo de estudio en el modelado matemático en operaciones crediticias.

El procesamiento de la información histórica de préstamos resultó en un modelo logístico que predice la probabilidad de que un préstamo sea moroso, con un 95% de nivel de confianza y p-value de las variables predictoras ≤ 0.05 . Además de las definiciones de los logaritmos de odds (coeficientes de modelo), dispuestas para cada variable artificial, las cuales indican si existe relación inversa logarítmica natural entre las variables predictoras y la morosidad de préstamos otorgados.

5.2 Recomendaciones

Se recomienda realizar algún otro tipo de modelo matemático o probabilístico, con ello se planea encontrar similitudes entre los resultados de ambos modelos, además de detectar pesos coincidentes de cada variable en el nuevo y antiguo modelo. La realización de un segundo modelo puede ayudar a realizar una calibración superior en el modelo propuesto.

Se recomienda profundizar los estudios en los modelos donde las variables categóricas se prioricen, que según la investigación realizada existen diferentes herramientas con esta notación, como es el caso del modelo log-lineales para tablas de contingencia, que estudia e interpreta los efectos sobre las frecuencias observadas entre los niveles de varias variables categóricas, así como sus interrelaciones. El estudio de las variables en este tipo de modelos puede dar resultados o enfoques que no se obtuvieran utilizando la común regresión lineal matemática.

Se recomienda aplicar el presente estudio a otras áreas de la cooperativa con el fin de ser soporte para el entendimiento de las condiciones que propician la asequibilidad de los productos y servicios por parte de los clientes y socios, además de ser una herramienta adicional a la generación de estrategias de negocios en lo relacionado a optimización de procesos internos y creación de productos y servicios personalizados.

5.3 Bibliografía

- [1] D. Carillo, “Política económica en Ecuador y su capacidad de impacto en el desarrollo,” 2018, pp. 30–60. Accessed: Aug. 06, 2022. [Online]. Available: <https://revistas.uasb.edu.ec/index.php/eg/article/view/1010>
- [2] D. Baquero Méndez and J. D. Mieles López, “Los booms petroleros: ¿Qué cambió en los últimos 40 años?,” *Desarrollo, historia económica, petróleo*, 2015.
- [3] A. Soria and O. Salome, “La calificación crediticia por el sistema financiero y el derecho a la honra y al buen vivir,” Universidad Regional Autónoma de los Andes “Uniandes,” Ambato, 2017.
- [4] Superintendencia de bancos, “Calificación de Riesgo Instituciones Financieras 2021,” *Superintendencia de Banco, Entidades Controladas*, 2021. Accessed: Aug. 06, 2022. [Online]. Available: <https://www.superbancos.gob.ec/bancos/calificacion-de-riesgo-instituciones-financieras-2021/>
- [5] M. Herrera, “La economía popular y solidaria: el caso ecuatoriano enfocado en el sector financiero cooperativista,” Universidad Andina Simón Bolívar , Quito, 2018. Accessed: Aug. 06, 2022. [Online]. Available: <https://repositorio.uasb.edu.ec/bitstream/10644/6053/1/T2546-MRI-Herrera-La%20economia.pdf>
- [6] Superintendencia de Economía Popular y Solidaria, “Manual técnico ‘Tablas de Información,’” *www.seps.gob.ec*, pp. 1–146, 2022. Accessed: Aug. 06, 2022. [Online]. Available: https://www.seps.gob.ec/wp-content/uploads/1.4.-Manual-de-Tablas-de-Información_V19.0.pdf
- [7] I. y C. Intendencia de Información Técnica, “Manual Técnico de Estructuras de Datos,” 2022. Accessed: Aug. 06, 2022. [Online]. Available: <https://www.seps.gob.ec/wp-content/uploads/9-Manual-Tecnico-Segmentos-4-y-5-version-3.0.pdf>
- [8] S. Álvarez and P. Guzmán, “Situación macroeconómica en 2021 y perspectivas 2022 ,” *colegiodeeconomistas.org*, Quito, pp. 1–15, Jan. 2021. Accessed: Aug. 06, 2022. [Online]. Available: <https://colegiodeeconomistas.org.ec/wp-content/uploads/2022/01/perspectivas-macroeconomicas-ecuador-2022-rev-2.pdf>
- [9] H. A. Brenes González, “Aplicación del análisis de regresión lineal simple para la estimación de los precios de las acciones de Facebook, Inc,” *REICE: Revista Electrónica de Investigación en Ciencias Económicas*, vol. 5, no. 10, pp. 133–155, Jan. 2018, doi: 10.5377/reice.v5i10.5535.
- [10] A. I. F. and F. A. L., “Evaluación del uso de modelos de regresión logística para el

- diagnóstico de instituciones financieras,” *Cienc Soc*, vol. XXXVI, pp. 590–627, 2011, [Online]. Available: <https://www.redalyc.org/articulo.oa?id=87022786002>
- [11] P. Sanchez-de-Miguel *et al.*, “Estimation of vineyard leaf area by linear regression,” *Spanish Journal of Agricultural Research*, vol. 9, no. 1, p. 202, Mar. 2011, doi: 10.5424/sjar/20110901-354-10.
- [12] O. R. Toscano Ramos and F. G. Benalcázar Palacios, “Modelo matemático para determinar la calidad de servicio en el transporte público urbano en la ciudad de Ambato,” Universidad Técnica de Ambato, Ambato, 2021.
- [13] Shirley Patricia Sarria Almeida, *Probabilidad de incurrir en mora de los usuarios de una tarjeta de crédito marca propia en Cali*, vol. 1. Bogota, 2013.
- [14] F. Zhou and T. Yu, “Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study,” *The Lancet*, vol. 395, no. 10229, pp. 1054–1062, Mar. 2020, doi: 10.1016/S0140-6736(20)30566-3.
- [15] S. Laurent, P. Boutouyrie, and I. Gautier, “Aortic Stiffness Is an Independent Predictor of All-Cause and Cardiovascular Mortality in Hypertensive Patients,” *Hypertension*, vol. 37, no. 5, pp. 1236–1241, May 2001, doi: 10.1161/01.HYP.37.5.1236.
- [16] A. S. Levey, L. A. Stevens, C. H. Schmid, and Y. (Lucy) Zhang, “A New Equation to Estimate Glomerular Filtration Rate,” *Ann Intern Med*, vol. 150, no. 9, p. 604, May 2009, doi: 10.7326/0003-4819-150-9-200905050-00006.
- [17] C. Ochoa, W. Galeano, and L. Agudelo, “Construcción de un modelo de scoring para el otorgamiento de crédito en un entidad financiera,” *Perfil de Coyuntura Económica No. 16*, vol. 16, pp. 1–30, Dec. 2010.
- [18] A. Downey, J. Elkner, and C. Meyers, *How to Think Like a Computer Scientist - Learning with Python*. Quality Books, 2008.
- [19] C. Monroy, “El lenguaje Python y su potencial en el desarrollo de software de inteligencia artificial,” 2018.
- [20] Aptude, “Por qué amamos Python para la ciencia de datos,” 2020. Accessed: Aug. 06, 2022. [Online]. Available: <https://aptude.com/es/Ciencia-de-los-datos/entrada/por-que%20amamos-Python-para-la-ciencia-de-datos/>
- [21] Google, “Python Master Guide.” 2018. Accessed: Aug. 06, 2022. [Online]. Available: <https://sites.google.com/view/learn-python-data-science/home>
- [22] Y. Rosseel, “Package for Structural Equation Modeling,” *J Stat Softw*, vol. 48, no. 2, 2012, doi: 10.18637/jss.v048.i02.
- [23] L. E. Vargas and E. M. Fuquen, *Introducción al análisis de datos con RStudio*. Colombia: Cenipalma, 2021. Accessed: Aug. 06, 2022. [Online]. Available: <https://www.interviewbit.com/blog/python-vs-r/>

- [24] Interviewbit, “Python Vs R: Know The Difference,” 2022.
- [25] D. Johnson, “R Vs Python: What’s the Difference?,” *Guru99*, pp. 1–20, 2022. Accessed: Aug. 06, 2022. [Online]. Available: <https://www.guru99.com/r-vs-python.html>
- [26] Instituto Nacional de Estadística y Geografía, *Procesamiento de la información*. México, 2012. Accessed: Sep. 02, 2022. [Online]. Available: https://www.snieg.mx/documentacionportal/normatividad/vigente/doctos_genbasica/procesamiento_informacion.pdf
- [27] R. Montero Granado, *Modelos de regresión lineal múltiple*. 2016.
- [28] C. M. Salcedo Poma, “Modelo de Regresión Logística,” 2012, pp. 20–250.
- [29] S. Fernández, “REGRESION LOGISTICA,” 2010, Accessed: Sep. 02, 2022. [Online]. Available: <https://www.estadistica.net/ECONOMETRIA/CUALITATIVAS/LOGISTICA/regresion-logistica.pdf>
- [30] A. Molina, “¿Qué significa realmente el valor de p?,” *Scielo*, pp. 5–10, Dec. 2017, Accessed: Sep. 13, 2022. [Online]. Available: https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1139-76322017000500014

5.4 Anexos

Anexo 1: Cooperativa Chibuleo en centro de Ambato



Anexo 2: Data set en Excel para diseño de modelo

Archivo Inicio Insertar Disposición de página Fórmulas Datos Revisar Vista Programador Ayuda Acrobat Power Pivot ¿Qué desea hacer?												
110 : S05												
	A	B	C	D	E	F	G	H	I	J	K	L
1	prest_moros	perio_pago	ofic_concecion	garantia	tipo_credito	destino_fina	act_econo_n	provinc_destinc	act_econo_titular	nivel_estud	tipo_viviend	genero
2	False	Mensual (30 DIAS)	MACHACHI	CO	PN	OT	Z	Pichincha	S05	Primaria	Vive con familiar	MASCULINO
3	False	Mensual (30 DIAS)	SANGOLQUI	GR	PN	OT	Z	Pichincha	C32	Formacion intern	Propia no hipote	MASCULINO
4	False	Mensual (30 DIAS)	QUITO SUR	GC	PN	OT	Z	Pichincha	S05	Secundaria	Propia no hipote	MASCULINO
5	False	Al vencimiento	MATRIZ	GR	PN	OT	Z	Tungurahua	G47	Primaria	Vive con familiar	FEMENINO
6	False	Mensual (30 DIAS)	MATRIZ	GC	MS	AS	G	Tungurahua	G47	Secundaria	Arrendada	MASCULINO
7	False	Mensual (30 DIAS)	SANGOLQUI	CO	PN	OT	Z	Pichincha	C32	Primaria	Propia no hipote	FEMENINO
8	True	Mensual (30 DIAS)	QUITO SUR	GC	MA	TO	C	Pichincha	C32	Secundaria	Propia hipotecad	FEMENINO
9	False	Mensual (30 DIAS)	QUITO SUR	CO	PN	OT	Z	Pichincha	S05	Secundaria	Propia no hipote	MASCULINO
10	False	Mensual (30 DIAS)	QUITO SUR	CO	PN	OT	Z	Pichincha	S05	Secundaria	Propia no hipote	MASCULINO
11	False	Mensual (30 DIAS)	MATRIZ	CO	PN	OT	Z	Tungurahua	P85	Universitaria	Propia no hipote	MASCULINO
12	False	Mensual (30 DIAS)	SANGOLQUI	GR	PN	OT	Z	Pichincha	M72	Secundaria	Propia no hipote	FEMENINO
13	False	Mensual (30 DIAS)	QUITO SUR	GT	PN	OT	Z	Pichincha	S05	Universitaria	Arrendada	MASCULINO
14	False	Mensual (30 DIAS)	MATRIZ	GC	PN	OT	Z	Tungurahua	S05	Secundaria	Vive con familiar	FEMENINO
15	False	Mensual (30 DIAS)	MATRIZ	GR	PN	OT	Z	Tungurahua	G45	Primaria	Arrendada	FEMENINO
16	False	Mensual (30 DIAS)	MATRIZ	GQ	PN	OT	Z	Tungurahua	S04	Universitaria	Propia no hipote	FEMENINO
17	False	Mensual (30 DIAS)	SANGOLQUI	GC	PN	OT	Z	Pichincha	M74	Formacion intern	Vive con familiar	MASCULINO
18	False	Mensual (30 DIAS)	MATRIZ	GC	ON	TF	Z	Tungurahua	S04	Universitaria	Propia no hipote	MASCULINO
19	False	Mensual (30 DIAS)	MATRIZ	GR	PN	OT	Z	Tungurahua	G47	Universitaria	Vive con familiar	MASCULINO
20	False	Mensual (30 DIAS)	SANGOLQUI	GR	PN	OT	Z	Pichincha	S96	Formacion intern	Prestada	MASCULINO
21	True	Mensual (30 DIAS)	LATACUNGA	CO	PN	OT	Z	Cotopaxi	S05	Primaria	Propia no hipote	MASCULINO
22	False	Mensual (30 DIAS)	QUITO SUR	GQ	PN	OT	Z	Pichincha	S05	Primaria	Propia no hipote	FEMENINO
23	False	Al vencimiento	QUITO SUR	GR	PN	OT	Z	Pichincha	C31	Primaria	Propia no hipote	FEMENINO
24	False	Mensual (30 DIAS)	QUITO SUR	GC	PN	OT	Z	Pichincha	G47	Secundaria	Vive con familiar	FEMENINO
25	False	Mensual (30 DIAS)	MATRIZ	CO	PN	OT	Z	Tungurahua	S05	Universitaria	Propia no hipote	MASCULINO
26	False	Mensual (30 DIAS)	MATRIZ	GR	PN	OT	Z	Tungurahua	S96	Secundaria	Prestada	FEMENINO
27	False	Mensual (30 DIAS)	SANGOLQUI	GP	PN	OT	Z	Pichincha	O84	Universitaria	Vive con familiar	MASCULINO
28	False	Mensual (30 DIAS)	MATRIZ	CO	PN	OT	Z	Tungurahua	S01	Primaria	Propia no hipote	MASCULINO
29	False	Mensual (30 DIAS)	SANGOLQUI	GP	MS	AS	C	Pichincha	C32	Formacion intern	Vive con familiar	FEMENINO
30	False	Mensual (30 DIAS)	MATRIZ	GC	PN	OT	Z	Cotopaxi	S05	Secundaria	Arrendada	FEMENINO
31	False	Mensual (30 DIAS)	MATRIZ	GR	PN	OT	Z	Tungurahua	C15	Secundaria	Vive con familiar	FEMENINO
32	True	Mensual (30 DIAS)	MACHACHI	GP	MS	CT	C	Pichincha	C32	Secundaria	Propia no hipote	FEMENINO
33	False	Mensual (30 DIAS)	SANGOLQUI	GP	PN	OT	Z	Pichincha	S04	Universitaria	Vive con familiar	FEMENINO
34	False	Mensual (30 DIAS)	MACHACHI	GC	PN	OT	Z	Pichincha	S04	Secundaria	Vive con familiar	MASCULINO
35	False	Mensual (30 DIAS)	SANGOLQUI	GC	PN	OT	Z	Pichincha	O86	Formacion intern	Propia no hipote	MASCULINO
36	False	Mensual (30 DIAS)	QUITO SUR	GC	ON	TF	Z	Pichincha	S05	Formacion intern	Vive con familiar	MASCULINO
37	True	Mensual (30 DIAS)	MAYORISTA	GC	ON	TF	Z	Tungurahua	T97	Secundaria	Propia no hipote	FEMENINO
38	True	Mensual (30 DIAS)	MACHACHI	GC	PN	OT	Z	Pichincha	S05	Secundaria	Propia no hipote	MASCULINO
39	False	Mensual (30 DIAS)	SANGOLQUI	CO	PN	OT	Z	Pichincha	C32	Secundaria	Vive con familiar	MASCULINO
40	False	Mensual (30 DIAS)	QUITO SUR	CO	PN	OT	Z	Pichincha	N82	Secundaria	Propia no hipote	FEMENINO
41	False	Mensual (30 DIAS)	QUITO SUR	GP	PN	OT	Z	Pichincha	S05	Formacion intern	Vive con familiar	MASCULINO

Anexo 3: Script en Transact-SQL para la extracción de datos

```
SELECT
IIF((SELECT TOP 1 C02_D.DIASMOROSIDAD FROM
FBS_ESTRUCTURAS.C02_DETALLE C02_D
WHERE C02_D.SECUENCIALPRÉSTAMO = C01_D.SECUENCIALPRÉSTAMO
ORDER BY C02_D.DIASMOROSIDAD DESC)>=30,1,0) MOROSIDAD_VALOR,
(SELECT TOP 1 C02_D.DIASMOROSIDAD FROM FBS_ESTRUCTURAS.C02_DETALLE
C02_D
WHERE C02_D.SECUENCIALPRÉSTAMO = C01_D.SECUENCIALPRÉSTAMO
ORDER BY C02_D.DIASMOROSIDAD DESC) MOROSIDAD_DIAS,
PM.SECUENCIAL,
PM.NUMEROPRÉSTAMO,
PM.FECHAADJUDICACION,
--VARIABLE PERIODICIDAD DE PAGO
P_P.VALOR PERIODICIDAD_PAGO_VALOR, P_P.DESCRIPCION
PERIODICIDAD_PAGO_DESCRIPCION,
--VARIABLE OFICINA CONCECION
O.VALOR OFICINA_CONCECION_VALOR, O.CIUDAD OFICINA_CONCECION_CIUDAD,
D.NOMBRE OFICINA_CONCECION_AGENCIA,
--VARIABLE TIPO DE GARANTÍA
G_C.VALOR GARANTIA_VALOR,G_C.DESCRIPCION GARANTIA_DESCRIPCION,
--VARIABLE TIPO CREDITO
T_C.VALOR_ITEM TIPO_CREDITO_VALOR_ITEM,
T_C.DESCRIPCION_ITEM TIPO_CREDITO_DESCRIPCION_ITEM,
T_C.VALOR_LISTA TIPO_CREDITO_VALOR,T_C.DESCRIPCION_LISTA
TIPO_CREDITO_DESCRIPCION_LISTA,
-- VARIABLE DESTINO FINANCIERO
D_F.VALOR_ITEM_2 DESTINO_FINANCIERO_VALOR_2,
D_F.DESCRIPCION_ITEM DESTINO_FINANCIERO_DESCRIPCION,
D_F.DESCRIPCION_LISTA DESTINO_FINANCIERO_DESCRIPCION_LISTA,
D_F.VALOR_DESCRIPCION_LISTA
DESTINO_FINANCIERO_VALOR_DESCRIPCION_LISTA,
--VARIABLE ACTIVIDAD ECONOMICA RECEPTORA DE OPERACIÓN
IIF((C01_D.TIPOCREDITO = 'ON' OR C01_D.TIPOCREDITO = 'PN'
OR C01_D.TIPOCREDITO = 'CO' OR C01_D.TIPOCREDITO = 'VP' OR
C01_D.TIPOCREDITO = 'IN'
OR C01_D.TIPOCREDITO = 'EV'), 'NNNNNNN', C01_D.ACTIVIDADOPERACION)
ACTIVIDAD_ECONOMICA_OPERACION_CODIGO_SEPS,
IIF((C01_D.TIPOCREDITO = 'ON' OR C01_D.TIPOCREDITO = 'PN' OR
C01_D.TIPOCREDITO = 'CO'
OR C01_D.TIPOCREDITO = 'VP' OR C01_D.TIPOCREDITO = 'IN' OR
C01_D.TIPOCREDITO = 'EV'),
'No tiene actividad económica', D_ACTIVIDADOPERACION.NOMBRE)
ACTIVIDAD_ECONOMICA_OPERACION_DETALLES ,
IIF((C01_D.TIPOCREDITO = 'ON' OR C01_D.TIPOCREDITO = 'PN' OR
C01_D.TIPOCREDITO = 'CO'
OR C01_D.TIPOCREDITO = 'VP' OR C01_D.TIPOCREDITO = 'IN' OR
C01_D.TIPOCREDITO = 'EV'),
1, A_E.VALOR )
ACTIVIDAD_ECONOMICA_OPERACION_VALOR,
IIF((C01_D.TIPOCREDITO = 'ON' OR C01_D.TIPOCREDITO = 'PN' OR
C01_D.TIPOCREDITO = 'CO'
OR C01_D.TIPOCREDITO = 'VP' OR C01_D.TIPOCREDITO = 'IN' OR
C01_D.TIPOCREDITO = 'EV'),
'No tiene actividad económica', A_E.DESCRIPCION)
ACTIVIDAD_ECONOMICA_OPERACION_DESCRIPCION,
```

```

--VARIABLE PROVINCIA DESTINO
PR.VALOR PROVINCIA_DESTINO_VALOR, PR.DESCRIPCION
PROVINCIA_DESTINO_DESCRIPCION,
--VARIABLE INGRESO DE TITULAR
(SELECT R_SBU.VALOR FROM [FBS_ESTRUCTURAS].[RANGOS_SBU] R_SBU
WHERE C01_D.TOTALINGRESOSUJETO BETWEEN R_SBU.MONTOINICIAL AND
R_SBU.MONTOIFINAL)
INGRESO_RANGO_SBU_VALOR,
C01_D.TOTALINGRESOSUJETO INGRESO_RANGO_SBU_DESCRIPCION,
--VARIABLE ACTIVIDAD ECONOMICA DE TITULAR DE OPERACION
C01_D.ACTIVIDADSUJETO ACTIVIDAD_ECONOMICA_TITULAR_CODIGO,
AC_E.DESCRIPCION_ITEMLISTA ACTIVIDAD_ECONOMICA_TITULAR_DESCRIPCION,
AC_E.VALOR_ITEMLISTA ACTIVIDAD_ECONOMICA_TITULAR_VALOR,
AC_E.DESCRIPCION_ITEM ACTIVIDAD_ECONOMICA_TITULAR_DETALLES,
--VARIABLE DE NIVEL DE ESTUDIO
N_E.VALOR NIVEL_ESTUDIO_VALOR, N_E.DESCRIPCION
NIVEL_ESTUDIO_DESCRIPCION,
--VARIABLE TIPO DE VIVIENDA
T_V.VALOR TIPO_VIVIENDA_VALOR, T_V.DESCRIPCION
TIPO_VIVIENDA_DESCRIPCION,
C01_D.PREDICIDADPAGO,
P_P.VALOR,
C01_D.VALOROPERACION,
C01_D.TASAINTERESNOMINAL,
C01_D.TASAEFFECTIVAANUAL,
C01_D.TOTALINGRESOSUJETO,
C01_D.TOTALEGRESOSUJETO,
C01_D.PATRIMONIOSUJETO,
N_E.VALOR,
IIF((T_V.VALOR = 3),1,0)
FROM FBS_ESTRUCTURAS.C01_DETALLE C01_D
INNER JOIN [FBS_ESTRUCTURAS].[PERIODICIDAD_PAGO] P_P ON
P_P.CODIGO=C01_D.PREDICIDADPAGO
INNER JOIN FBS_CARTERA.PRÉSTAMOMAESTRO PM ON PM.SECUENCIAL =
C01_D.SECUENCIALPRÉSTAMO
INNER JOIN FBS_ORGANIZACIONES.OFICINA O ON O.SECUENCIALDIVISION =
PM.SECUENCIALOFICINA
INNER JOIN FBS_GENERALES.DIVISION D ON D.SECUENCIAL =
O.SECUENCIALDIVISION
INNER JOIN [FBS_ESTRUCTURAS].[GARANTES_CODEUDORES] G_C ON G_C.CODIGO
= C01_D.GARANTIAS
INNER JOIN [FBS_ESTRUCTURAS].[TIPO_CREDITO] T_C ON T_C.CODIGO =
C01_D.TIPOCREDITO
INNER JOIN [FBS_ESTRUCTURAS].[DESTINO_FINANCIERO] D_F ON
D_F.CODIGO=C01_D.DESTINOFINANCIERO
INNER JOIN [FBS_ESTRUCTURAS].[ACTIVIDAD_ECON_RECEPOPER] A_E ON
A_E.CODIGO = SUBSTRING(C01_D.ACTIVIDADOPERACION,1,1)
INNER JOIN [FBS_ESTRUCTURAS].[PROVINCIA] PR ON PR.CODIGO =
C01_D.DESTINOPROVINCIA
INNER JOIN [FBS_ESTRUCTURAS].[ACTIVIDAD_ECONOMICA_SUJETO] AC_E ON
AC_E.CODIGO = C01_D.ACTIVIDADSUJETO
INNER JOIN [FBS_ESTRUCTURAS].[NIVEL_ESTUDIOS] N_E ON N_E.CODIGO =
C01_D.NIVELESTUDIOS
INNER JOIN [FBS_ESTRUCTURAS].[TIPO_VIVIENDA] T_V ON T_V.CODIGO =
C01_D.TIPOVIVIENDA
--ACTIVIDAD ECONOMICA DE SUJETO/TITULAR

```



```

INNER JOIN FBS_ESTRUCTURAS.ACTIVIDAD_ECONOMICA_SUJETO A_E_S ON
A_E_S.CODIGO = C01_D.ACTIVIDADSUJETO
--ACTIVIDAD ECONOMICA RECEPTORA DE OPRECIÓN
INNER JOIN FBS_GENERALES.DIVISION D_ACTIVIDADOPERACION ON
D_ACTIVIDADOPERACION.CODIGO = C01_D.ACTIVIDADOPERACION
INNER JOIN FBS_GENERALES.NIVELDIVISION ND ON ND.SECUENCIAL =
D_ACTIVIDADOPERACION.SECUENCIALNIVEL
INNER JOIN FBS_GENERALES.TIPODIVISION TD ON TD.SECUENCIAL =
ND.SECUENCIALTIPODIVISION
WHERE C01_D.FECHACONCESION>='01/01/2017'
AND PM.CODIGOESTADOPRÉSTAMO='Z'
AND TD.CODIGO='AEC'
AND C01_D.ACTIVIDADOPERACION <> 'N/A' --CREDITOS MIGRADOS
ORDER BY PM.SECUENCIAL ASC

```

Anexo 4: Segmento de codificación del modelado en R

```

install.packages("dplyr")
#carga de datos

datosCategoria <- read.csv(file =
'C:/Users/Usuario/Documents/EstructurasCsv/variableCategorica_3_ori_r.csv')

#Establecimiento de factores de nivel - Variables categórica
datosCategoria$prest_moros <- as.factor(datosCategoria$prest_moros)
datosCategoria$perio_pago <- as.factor(datosCategoria$perio_pago)
datosCategoria$ofic_concecion <- as.factor(datosCategoria$ofic_concecion)
datosCategoria$garantia <- as.factor(datosCategoria$garantia)
datosCategoria$tipo_credito <- as.factor(datosCategoria$tipo_credito)
datosCategoria$destino_financiero <- as.factor(datosCategoria$destino_financiero)
datosCategoria$sact_econo_recep_oper <- as.factor(datosCategoria$sact_econo_recep_oper)
datosCategoria$provinc_destino <- as.factor(datosCategoria$provinc_destino)
datosCategoria$sact_econo_titular <- as.factor(datosCategoria$sact_econo_titular)
datosCategoria$nivel_estud <- as.factor(datosCategoria$nivel_estud)
datosCategoria$tipo_viviend <- as.factor(datosCategoria$tipo_viviend)
datosCategoria$genero <- as.factor(datosCategoria$genero)

#Nuevo dataframe con variables
datosCategoriaDescarteVaria <- data.frame(datosCategoria$prest_moros,
      datosCategoria$perio_pago,
      datosCategoria$ofic_concecion,
      datosCategoria$ofic_concecion,
      datosCategoria$garantia,
      datosCategoria$tipo_credito,
      datosCategoria$destino_financiero,
      datosCategoria$nivel_estud,
      datosCategoria$tipo_viviend
)

#Conversión a factor en nuevo modelo
datosCategoriaDescarteVaria$prest_moros <-
as.factor(datosCategoriaDescarteVaria$prest_moros)
datosCategoriaDescarteVaria$perio_pago <-
as.factor(datosCategoriaDescarteVaria$perio_pago)

```

```

datosCategoriaDescarteVaria$ofic_concecion <-
as.factor(datosCategoriaDescarteVaria$ofic_concecion)
datosCategoriaDescarteVaria$garantia <- as.factor(datosCategoriaDescarteVaria$garantia)
datosCategoriaDescarteVaria$tipo_credito <-
as.factor(datosCategoriaDescarteVaria$tipo_credito)
datosCategoriaDescarteVaria$destino_financiero <-
as.factor(datosCategoriaDescarteVaria$destino_financiero)
datosCategoriaDescarteVaria$nivel_estud <-
as.factor(datosCategoriaDescarteVaria$nivel_estud)
datosCategoriaDescarteVaria$tipo_viviend <-
as.factor(datosCategoriaDescarteVaria$tipo_viviend)
datosCategoriaDescarteVaria$genero <- as.factor(datosCategoriaDescarteVaria$genero)

tabla <- table(datosCategoria$prest_moros, datosCategoria$perio_pago,
              dnn = c("prest_moros", "perio_pago"))
addmargins(tabla)

tabla_frecuencias <- prop.table(tabla)*100
addmargins(tabla_frecuencias)

#generación de modelo con todas las variables iniciales

modelo_glm <- glm(prest_moros ~ perio_pago + ofic_concecion + garantia + tipo_credito
                 + destino_financiero + act_econo_recep_oper + provinc_destino +
act_econo_titular
                 + nivel_estud + tipo_viviend + genero, data = datosCategoria ,
                 family = "binomial")
summary(modelo_glm)

modelo_glm <- glm(prest_moros ~ perio_pago + ofic_concecion + garantia + tipo_credito
                 + destino_financiero + tipo_viviend + genero,
                 data = datosCategoria ,
                 family = "binomial")
summary(modelo_glm)

sig.var<- summary(modelo_glm)$coeff[-1,4] <0.01
names(sig.var)[sig.var == TRUE]

#modelo con variables no descartadas.

modeloVarDesc_glm <- glm(prest_moros ~ perio_pago + ofic_concecion + garantia +
tipo_credito
                        + destino_financiero + nivel_estud + tipo_viviend + genero, data =
datosCategoriaDescarteVaria ,
                        family = "binomial")

#resumen de modelado

summary(modelo_glm)

```