

Modelo metodológico para la originación eficiente de créditos en el sector de economía popular
urbano en Colombia

Erik Andres Naranjo Gonzalez

Plan de trabajo de grado para el título de Especialista en Estadística

Director

Heivar Yesid Rodríguez Pinzón

Magíster en Ciencias Económicas

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Matemáticas

Especialización en Estadística

Bucaramanga

2026

Tabla de Contenido

	Pág.
Introducción	7
1. Descripción del problema	10
2. Antecedentes	11
3. Justificación	14
4. Objetivos	15
4.1 Objetivo general	15
4.2 Objetivos específicos	16
5. Marco teórico	16
6. Metodología	20
7. Resultados	26
8. Conclusiones	66
Referencias Bibliográficas	68
Apéndices	70

Lista de Tablas

	Pág.
Tabla 1. Descripción de variables.....	21
Tabla 2. Top 10 de ciudades del portafolio popular productivo urbano.....	48
Tabla 3. Resultado del muestreo aleatorio estratificado.....	58
Tabla 4. Variables con mayor colinealidad.....	59
Tabla 5. Matriz de confusión.....	64
Tabla 6. Indicadores Estadísticos.....	64
Tabla 7. Score.....	66

Lista de Figuras

	Pág.
Figura 1. Resumen estadístico de las variables numéricas	27
Figura 2. Gráfico Box Plot – Y= Default X= Plazo.....	27
Figura 3. Gráfico Box Plot – Y= Default X= Tasa	28
Figura 4. Gráfico Box Plot – Y= Default X= Monto Desembolsado.....	29
Figura 5. Gráfico Box Plot – Y= Default X= Cuota	30
Figura 6. Gráfico Box Plot – Y= Default X= Ventas.....	31
Figura 7. Gráfico Box Plot – Y= Default X= Disponible	32
Figura 8. Gráfico Box Plot – Y= Default X= Activos Fijos	33
Figura 9. Gráfico Box Plot – Y= Default X= Gastos de Negocio.....	34
Figura 10. Gráfico Box Plot – Y= Default X= Gastos Familiares	35
Figura 11. Gráfico Box Plot – Y= Default X= Experiencia del negocio	36
Figura 12. Gráfico Box Plot – Y= Default X= Tiempo de posesión	37
Figura 13. Gráfico Box Plot – Y= Default X= Numero de hijos	38
Figura 14. Gráfico Box Plot – Y= Default X= Liquidez	39
Figura 15. Gráfico Box Plot – Y= Default X= Activo sobre patrimonio.....	40
Figura 16. Gráfico Box Plot – Y= Default X= Pasivo sobre activo.....	41
Figura 17. Gráfico Box Plot – Y= Default X= Capacidad de pago sobre patrimonio	42
Figura 18. Gráfico Box Plot – Y= Default X= Días de mora t-1	43
Figura 19. Gráfico Box Plot – Y= Default X= Días de mora t-3	44
Figura 20. Gráfico Box Plot – Y= Default X= Días de mora t-6.....	44

Figura 21. Gráfico Box Plot – Y= Default X= Días de mora t-9	45
Figura 22. Gráfico Box Plot – Y= Default X= Días de mora t-12	46
Figura 23. Gráfico Barras Y= frecuencia X= Categoría genero	47
Figura 24. Gráfico Barras Y= frecuencia X= Categoría producto	47
Figura 25. Gráfico Barras Y= frecuencia X= Categoría departamento	49
Figura 26. Gráfico Barras Y= frecuencia X= Categoría oficina top 15	50
Figura 27. Gráfico Barras Y= frecuencia X= Destino	51
Figura 28. Gráfico Barras Y= frecuencia X= Macro Sector Top 15.....	52
Figura 29. Gráfico Barras Y= frecuencia X= Segmento.....	53
Figura 30. Gráfico Barras Y= frecuencia X= Tipo de crédito	54
Figura 31. Gráfico Barras Y= frecuencia X= Tipo de seguro.....	55
Figura 32. Gráfico Barras Y= frecuencia X= Tipo de vivienda.....	56
Figura 33. Gráfico Barras Y= frecuencia X= Estrato	57
Figura 34. Gráfico Barras Y= frecuencia X= Tipo de ingreso.....	57
Figura 35. Matriz de correlación (V de Cramer)	60
Figura 36. Modelo logístico optimizado	61
Figura 37. Curva ROC	65

Lista de Apéndices

	Pág.
Apéndice A. Nota explicativa sobre los títulos del cuadro de estadísticos descriptivos:.....	70

RESUMEN

Título: modelo metodológico para la originación eficiente de créditos en el sector de economía popular urbano en Colombia.*

Autor: Erik Andres Naranjo Gonzalez**

Palabras clave: Crédito, Puntaje, Originación, Microcrédito, Regresión Logística, Probabilidad, Apetito de Riesgo, Gestión de Riesgo de crédito

Descripción:

Este estudio desarrolla un modelo metodológico para optimizar la originación de créditos en el sector de economía popular productiva urbana en Colombia, respondiendo a la necesidad de una evaluación de riesgo eficiente ante las nuevas modalidades de crédito impulsadas por el Gobierno Nacional mediante el decreto 455 de marzo de 2023.

Bajo un enfoque cuantitativo y predictivo, se estimó un modelo de regresión logística utilizando una base de datos real de una entidad financiera con registros entre abril de 2023 y mayo de 2025. La investigación identificó las variables determinantes del riesgo crediticio, destacando el comportamiento histórico de mora reciente (días de mora en t-1) como el predictor más fuerte del incumplimiento, seguido por variables como el tipo de cliente (nuevo vs. renovado) y el género.

El modelo resultante demostró un desempeño predictivo robusto, alcanzando un Área Bajo la Curva (AUC) del 73.5%, un coeficiente GINI del 40.92% y un estadístico KS del 30.36%, validando su capacidad para discriminar efectivamente entre clientes cumplidos e incumplidos. Finalmente, se generó un score crediticio que permite segmentar el riesgo en niveles bajo, medio y alto, proporcionando una herramienta práctica para maximizar la relación riesgo-rentabilidad, asegurar la sostenibilidad del portafolio de microcrédito en este segmento y definir el apetito de riesgo de la entidad.

*Trabajo de grado

**Facultad de ciencias. Escuela de matemáticas. Director Heivar Yesid Rodríguez Pinzón

ABSTRACT

Title: Methodological Model for Efficient Credit Origination in Colombia's Urban Popular Economy Sector

Author: Erik Andres Naranjo Gonzalez

Keywords: Credit, Scoring, Origination, Microcredit, Logistic Regression, Probability, Risk Appetite, Credit Risk Management.

Descripción:

This study develops a methodological model to optimize credit origination in the urban productive popular economy sector in Colombia, addressing the need for efficient risk assessment regarding the new credit modalities promoted by the National Government through Decree 455 of March 2023.

Using a quantitative and predictive approach, a logistic regression model was estimated using a real database from a financial entity with records between April 2023 and May 2025. The research identified the determinant variables of credit risk, highlighting recent historical delinquency behavior (days past due in t-1) as the strongest predictor of default, followed by variables such as client type (new vs. renewed) and gender.

The resulting model demonstrated robust predictive performance, reaching an Area Under the Curve (AUC) of 73.5%, a Gini coefficient of 40.92%, and a KS statistic of 30.36%, validating its ability to effectively discriminate between compliant and defaulting clients. Finally, a credit score was generated to segment risk into low, medium, and high levels, providing a practical tool to maximize the risk-return relationship, ensure the sustainability of the microcredit portfolio in this segment, and define the entity's risk appetite.

*Bachelor Thesis

**Faculty of Science. School of Mathematics. Advisor Heivar Yesid Rodríguez Pinzón

Introducción

Los intermediarios financieros vigilados por la Superintendencia Financiera de Colombia han visto limitada la originación de créditos hacia el sector microcrediticio, debido a los niveles de tasa de usura en la economía colombiana. Este contexto restringe el apetito de riesgo de las entidades y reduce el acceso al crédito formal por parte de los microempresarios.

En respuesta a esta situación, el Gobierno Nacional expidió el Decreto 455 del 29 de marzo de 2023, mediante el cual se establecen nuevas modalidades de crédito orientadas a promover la inclusión financiera, mediante la generación de eficiencias en los costos financieros que faciliten el acceso al crédito formal acorde con las necesidades y perfiles de los sectores productivos, con el objetivo de impulsar la actividad económica de los sectores más vulnerables de la población, con especial énfasis en las zonas rurales. En este marco, se crean las siguientes modalidades de crédito: crédito popular productivo rural, crédito popular productivo urbano, crédito productivo rural, crédito productivo urbano y crédito productivo de mayor monto.

En este sentido, se hace necesario desarrollar modelos estadísticos prospectivos, que optimicen la discriminación de clientes con buen hábito de pago de los que no por cada una de las nuevas modalidades, buscando potencializar el volumen de desembolsos, maximizar la relación riesgo rentabilidad y aumentar el acceso al crédito del sector microcrediticio popular.

Lo anterior, mediante la aplicación metodológica de un modelo de regresión logística generalizado que permita estimar la probabilidad de incumplimiento en un horizonte de tiempo de 12 meses, para la modalidad de crédito popular urbano, buscando la mayor discriminación

entre clientes con buen hábito de pago de los que no, mediante variables demográficas, financieras y de comportamiento.

1. Descripción del problema

Los establecimientos de crédito están expuestos al riesgo de crédito como parte inherente de su actividad económica. Este riesgo se define como la posibilidad de que la entidad incurra en pérdidas y disminuya el valor de sus activos debido al incumplimiento de las obligaciones por parte de un deudor o contraparte. Su gestión requiere la implementación de etapas de identificación, medición, control y monitoreo a lo largo del ciclo de crédito de originación, seguimiento y recuperación de la cartera de créditos, definiendo un marco de apetito de riesgo, el cual debe estar alineado a los objetivos de rentabilidad del negocio de intermediación financiera.

En este contexto, las entidades financieras buscan expandir su cartera de crédito manteniendo niveles de riesgo aceptables que permitan alcanzar las rentabilidades esperadas. En el caso del portafolio de microcrédito, segmentado en cinco submodalidades en las cuales el perfil de riesgo de los deudores presenta diferencias significativas derivadas de factores como las fuentes de ingreso, liquidez, solvencia, nivel de endeudamiento, actividad económica, hábitos de pago, garantías, ubicación geográfica, riesgos emergentes y tasas de originación. Por ello, resulta fundamental contar con políticas, procedimientos y metodologías que faciliten la adecuada identificación y discriminación entre clientes sujetos de crédito y aquellos que no lo son.

En consecuencia, los establecimientos de crédito enfrentan la necesidad de optimizar el proceso de originación, con el propósito de incrementar el saldo del portafolio, controlar el riesgo de crédito y maximizar la rentabilidad. Para ello, es indispensable el desarrollo e implementación de modelos estadísticos que permitan una evaluación más efectiva y eficiente de los potenciales deudores.

2. Antecedentes

En la literatura latinoamericana existe un volumen amplio de estudios sobre modelos predictivos para microcrédito, haciendo uso de la regresión logística como técnica central para estimar probabilidades de incumplimiento, debido a su interpretabilidad, su alineación con marcos regulatorios y su utilidad práctica en procesos de originación y seguimiento de cartera. Este enfoque se ha consolidado especialmente en instituciones microfinancieras y entidades de crédito de la región, donde la disponibilidad de información es limitada y los modelos deben balancear precisión y simplicidad.

- Regresión logística como herramienta principal para la predicción del riesgo

Rayo Cantón, Lara Rubio y Camino Blasco (2010) desarrollan un modelo de *credit scoring* para una institución de microfinanzas peruana, demostrando que la regresión logística binaria es apropiada para clasificar clientes entre pagadores e incumplidos. Los autores muestran que la técnica permite integrar variables de cliente, operación y entorno, logrando un modelo con

equilibrio entre capacidad predictiva y claridad interpretativa para ser aplicado por analistas de riesgo.

Del mismo modo, Arenas Díaz, Boccardi Rodríguez y Piñeyrúa Ibáñez (2012) confirman la utilidad de los modelos logísticos en microcrédito, comparándolos con árboles de decisión y mostrando que el logit ofrece estimaciones más estables y comprensibles para el oficial de crédito. Además, destacan el reto de construir modelos con bases pequeñas o sesgadas, situación frecuente en entidades microfinancieras.

El estudio de Tamara-Ayus, Trujillo-Hernández y Mosquera-Peña (2022), aplicado específicamente a una entidad colombiana, refuerza esta tendencia. Los autores implementan un modelo score basado en regresión logística para mejorar el proceso de aprobación de microcréditos, concluyendo que la inclusión de variables socioeconómicas, demográficas y de comportamiento incrementa sustancialmente la capacidad discriminante del modelo. El artículo resalta que el uso de regresión logística en entidades financieras colombianas contribuye a estandarizar el proceso de evaluación crediticia y reduce la subjetividad del análisis tradicional.

- Validación de modelos y métricas utilizadas

Los estudios revisados coinciden en el uso de indicadores clásicos para evaluar modelos logísticos: la curva ROC, el AUC, las matrices de confusión, así como indicadores de ajuste como el estadístico χ^2 y el R^2 de Nagelkerke.

En particular, Rayo Cantón et al. (2010) incorporan variables macroeconómicas que mejoran la estabilidad del modelo ante cambios del ciclo económico, mientras que Tamara-Ayus et al. (2022) destacan la importancia de realizar validaciones *out-of-sample* para garantizar que el modelo conserve su poder predictivo en nuevos periodos de originación.

Por otro lado, Rodríguez Avellaneda (2018) compara modelos logísticos, probit, árboles y redes neuronales en el contexto colombiano y evidencia que, aunque algunas técnicas presentan mejor ajuste, la regresión logística sigue siendo esencial por su claridad interpretativa y su facilidad para identificar variables clave como garantía, historial de mora o caracterización del deudor.

- Aplicación al sector financiero colombiano

El trabajo de Tamara-Ayus et al. (2022) representa una evidencia sólida del uso de modelos logísticos en microcrédito colombiano, demostrando cómo estas herramientas transforman los procesos de aprobación mediante metodologías cuantitativas alineadas con los estándares regulatorios de la Superintendencia Financiera de Colombia.

Asimismo, Rodríguez Avellaneda (2018) muestra que los modelos estadísticos aplicados a entidades colombianas permiten segmentar clientes y diseñar estrategias de mitigación de riesgo, especialmente en la fase de recuperación. Sus hallazgos son relevantes para instituciones locales que deben conciliar rigurosidad estadística, facilidad operativa y cumplimiento regulatorio.

- Consideraciones regulatorias y desafíos comunes

Asimismo, Los estudios revisados resaltan que la regresión logística se adapta adecuadamente a las exigencias de marcos regulatorios como Basilea II y Basilea III, que requieren estimaciones de PD confiables para procesos de gestión de capital y provisiones.

Sin embargo, también enfatizan limitaciones como: bases de datos pequeñas o con baja proporción de incumplimiento, sesgos derivados de trabajar únicamente con créditos aprobados, problemas de calidad de datos en el segmento microcredicio y necesidad de realizar validaciones temporales (out-of-time).

Estas consideraciones son especialmente relevantes para el sistema financiero colombiano, que ha venido fortaleciendo la analítica del riesgo de crédito en entidades demicrofinanzas, bancos especializados y cooperativas.

3. Justificación

El desarrollo de un modelo de regresión logística para la modalidad de microcrédito popular productivo urbano se justifica en la necesidad de capturar las particularidades y heterogeneidades que caracterizan a este segmento del portafolio. Cada submodalidad de microcrédito presenta diferencias sustanciales en variables de comportamiento, financieras y demográficas.

Estas diferencias generan estructuras de riesgo no homogéneas, por lo que la estimación de un único modelo global podría conducir a resultados sesgados o imprecisos al no reflejar adecuadamente las dinámicas específicas de cada segmento. En contraste, la estimación de modelos individuales permite capturar con mayor precisión las variables determinantes del incumplimiento dentro de cada grupo, mejorando la capacidad predictiva y la discriminación entre clientes de buen y mal hábito de pago.

La regresión logística, por su naturaleza probabilística y su idoneidad para variables dependientes binarias (como el cumplimiento o incumplimiento del pago), se constituye en una herramienta adecuada para estimar la probabilidad de incumplimiento en cada segmento.

Adicionalmente, su interpretación facilita la identificación de los factores que inciden significativamente en el riesgo crediticio, proporcionando insumos valiosos para la gestión del riesgo, la definición de políticas de originación y la toma de decisiones estratégicas.

En síntesis, la construcción de un modelo de regresión logística específico para la modalidad de microcrédito popular productivo urbano contribuye a fortalecer la gestión del riesgo de crédito, generando una evaluación más precisa, diferenciada y eficiente de los deudores, lo que se traduce en una asignación de recursos más efectiva y en un mejor equilibrio en la relación riesgo-rentabilidad, sin limitar el crecimiento del portafolio.

4. Objetivos

4.1 Objetivo general

Desarrollar un modelo metodológico que permita optimizar la originación de créditos en el sector de economía popular productiva urbano en Colombia, a través de la identificación, estimación y validación de variables determinantes del riesgo crediticio, con el fin de mejorar la eficiencia en la evaluación de deudores, fortalecer la gestión del riesgo y contribuir a la sostenibilidad del portafolio de microcrédito.

4.2 Objetivos específicos

Diseñar y aplicar un modelo estadístico de regresión logística que estime la probabilidad de incumplimiento para la modalidad popular productiva urbano de microcrédito, considerando las particularidades de este segmento.

Evaluar el desempeño y capacidad predictiva del modelo estimado, mediante métricas de validación como la curva ROC, el estadístico KS y GINI con el propósito de medir su efectividad en la discriminación del riesgo.

Generar un score crediticio con fundamento en la probabilidad de incumplimiento, que permita la originación eficiente de créditos, manteniendo un equilibrio entre el crecimiento del portafolio y el control del riesgo de crédito en el segmento popular productivo urbano.

5. Marco teórico

El riesgo de crédito se define como la posibilidad de que una entidad financiera incurra en pérdidas derivadas del incumplimiento de las obligaciones contractuales por parte de un deudor o contraparte. En el contexto de la intermediación financiera, este riesgo constituye uno de los principales factores que afectan la estabilidad, rentabilidad y sostenibilidad de las instituciones de crédito.

La gestión del riesgo de crédito implica el desarrollo de modelos prospectivos que permitan anticipar comportamientos de incumplimiento y adoptar decisiones informadas sobre la colocación

de recursos. En Colombia, este proceso se desarrolla en el marco del Sistema de Administración del Riesgo de Crédito (SARC), regulado por la Superintendencia Financiera de Colombia, el cual incorpora la gestión de la modalidad microcrédito, la cual agrupa actividades económicas desarrolladas en unidades productivas de bajo nivel de formalización y limitado acceso a servicios financieros tradicionales. En esta línea, el microcrédito se consolida como un instrumento clave para la inclusión financiera, al facilitar el acceso al crédito, incentivando la dinámica de la economía y fortaleciendo la estabilidad económica de este sector.

Sin embargo, la naturaleza de este segmento implica mayores niveles de riesgo crediticio, asociados a la inestabilidad de las fuentes de ingreso, niveles de endeudamiento relativamente altos y una capacidad de pago restringida. A ello se suma la ausencia de información contable o financiera estructurada, la falta de historial crediticio formal, la insuficiencia de garantías reales y la exposición a riesgos sociales, ambientales y macroeconómicos, agravados en algunos casos por la localización geográfica dispersa o de difícil acceso de los deudores.

En consecuencia, la modelación estadística del riesgo de crédito se orienta a estimar la probabilidad de incumplimiento, la cual representa la probabilidad de que un cliente no cumpla con sus obligaciones contractuales en un horizonte de tiempo determinado. Esta estimación se realiza mediante modelos de regresión logística, aunque también se emplean técnicas como árboles de decisión, cadenas de markov o redes neuronales, entre otras.

El modelo de regresión logística ha sido ampliamente utilizado en el ámbito financiero por su interpretabilidad, robustez y capacidad para manejar variables cualitativas y cuantitativas. Su aplicación permite estimar la probabilidad de que ocurra un evento binario, en este caso el incumplimiento y el no incumplimiento, a partir de un conjunto de variables explicativas. Matemáticamente, el modelo se expresa como:

$$P(Y_i = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}}$$

donde $P(Y=1|X)$ representa la probabilidad de incumplimiento, X_i son las variables explicativas (demográficas, financieras o de comportamiento) y β_i los coeficientes asociados que determinan la influencia de cada variable sobre la probabilidad de default. El signo de cada β_i indica la dirección del efecto (positivo o negativo), mientras que su magnitud refleja la intensidad de dicho impacto. Para facilitar la interpretación, se utiliza el odds ratio (OR), definido como:

$$OR_i = e^{\beta_i}$$

El OR_i indica cuánto se multiplican las probabilidades de incumplimiento por cada unidad de incremento en la variable X_i , manteniendo las demás constantes. Si $OR_i > 1$, la variable aumenta la probabilidad de incumplimiento; si $OR_i < 1$, la reduce.

Una vez estimado el modelo, su validación y capacidad discriminadora se evalúan mediante diversas métricas estadísticas:

1. El estadístico KS mide la capacidad del modelo para discriminar entre buenos y malos pagadores. Se calcula como la diferencia máxima entre las funciones de distribución acumulada de las probabilidades predichas para ambos grupos. Un valor de KS superior al 40% indica un modelo con buena capacidad de discriminación, mientras que valores inferiores al 20% sugieren bajo poder predictivo.

2. El índice GINI se deriva de la curva ROC y representa el grado de desigualdad o separación entre las distribuciones de buenos e incumplidos. Se calcula mediante la relación:

$$GINI = 2 \times AUC - 1$$

donde el AUC (Area Under the Curve) es el área bajo la curva ROC. Un GINI cercano a 1 indica excelente discriminación, mientras que valores cercanos a 0 representan un modelo sin poder predictivo. En modelos crediticios, un GINI entre 0.4 y 0.6 se considera adecuado.

3. La curva ROC (Receiver Operating Characteristic) ilustra gráficamente la capacidad del modelo para clasificar correctamente a los individuos. Representa la tasa de verdaderos positivos (sensibilidad) frente a la tasa de falsos positivos (1- especificidad) para diferentes umbrales de probabilidad. El área bajo la curva (AUC) resume el desempeño global:

- $AUC > 0.9$: Excelente discriminación
- $0.8 \leq AUC < 0.9$: Muy buena
- $0.7 \leq AUC < 0.8$: Aceptable
- $AUC < 0.7$: Débil

En el contexto del riesgo de crédito, estas métricas permiten no solo evaluar la capacidad del modelo para diferenciar entre clientes con buen hábito de pago y de los que no, sino también validar su estabilidad y consistencia en el tiempo mediante procesos de backtesting y monitoreo continuo, de acuerdo con los lineamientos del Sistema de Administración del Riesgo de Crédito (SARC) establecido por la Superintendencia Financiera de Colombia.

Dado que la modalidad popular productiva urbano presentan diferencias significativas en el perfil de riesgo de los clientes y las condiciones del crédito, se aplicará un enfoque segmentado, lo que permitirá mejorar la capacidad explicativa y predictiva del modelo, buscando optimizar la asignación de recursos, reducir la tasa de incumplimiento y mejorar la rentabilidad ajustada por riesgo.

6. Metodología

El tipo de investigación empleado es de tipo aplicada, descriptiva y predictiva, dado que busca desarrollar y poner en práctica un modelo metodológico que mejore la eficiencia en los procesos de originación de crédito dentro del sector de economía popular, contribuyendo directamente a la gestión del riesgo crediticio y a la rentabilización del portafolio. Para construir el modelo, se requiere realizar un análisis descriptivo de las variables y el comportamiento de los deudores de microcrédito popular productivo urbano dentro del horizonte de tiempo analizado, identificando las características sociodemográficas, financieras y crediticias de los deudores, con el fin de comprender los factores que inciden en el incumplimiento.

Lo anterior, con la finalidad de construir un modelo estadístico prospectivo, mediante una regresión logística, con el objetivo de estimar la probabilidad de incumplimiento, siendo esta la posibilidad de presentar mora mayor a 30 días, aplicación de castigo y/o acogerse a la ley de insolvencia, durante los 12 meses siguientes a la fecha de corte. Esto permite asignar un perfil de riesgo a cada deudor conforme a la modalidad popular, desde el análisis de la originación del crédito.

Respecto a los datos empleados en el estudio corresponden a un portafolio de crédito originado desde abril de 2023 a mayo de 2025 de una entidad financiera que atiende el sector microcrediticio popular, vigilada por la Superintendencia Financiera de Colombia. Cada observación representa un deudor con información correspondiente a variables demográficas, financieras y de comportamiento crediticio, las cuales se describen a continuación:

Tabla 1.*Descripción de variables*

No.	Variable	Descripción	Unidad	Categorización		
1	Genero	Identifica el sexo del solicitante	Categoría nominal	Jurídica		0
				Hombre		1
				Mujer		2
2	Producto	Clasificación del tipo de productocrediticio solicitado	Categoría nominal	CR62		1
				CR67		2
				CR69		3
				CR61		4
				CR68		5
				CR66		6
3	Plazo	El período total del crédito agrupado en rangos de tiempo encuartiles.	Meses	LI	LS	Categoría
				0	51,89	1
				52	63,0	2
				63	67,0	3
				67		4
5	Desembolso	Clasifica el valor desembolsado al cliente en cuartiles	Pesos COP	LI	LS	Categoría
				-	1.500.000	1
				1.500.000	2.500.000	2
				2.500.000	4.000.000	3
		4.000.000		4		
6	Cuota	Categoriza el valor de la cuota de pago del crédito en cuartiles.	Pesos COP	LI	LS	Categoría
				-	162.896	1
				162.896	228.292	2
				228.292	322.548	3
				322.548		4
7	Ventas	Agrupa los niveles de ventas mensuales del negocio del solicitante en cuartiles.	Pesos COP	LI	LS	Categoría
				-	3.000.000	1
				3.000.000	4.550.000	2
				4.550.000	7.732.475	3
				7.732.475		4
8	Ciudad	Ciudad en la cual se originó el crédito.	Categoría nominal	Se categoriza cada una de las 257 ciudades en orden alfabético toman el valor de 1 hasta 257		
9	Departamento	Departamento del país donde se registra la operación crediticia.	Categoría nominal	Se categoriza cada una de los 14 departamentos en orden alfabético toman el valor de 1 hasta 14		

No.	Variable	Descripción	Unidad	Categorización		
10	Oficina	Identifica la oficina comercial responsable del crédito	Catagórica nomina	Se categoriza cada una de las 94 ciudades en orden alfabético toman el valor de 1 hasta 94		
11	Zonal	Región comercial asignada por la entidad financiera.	Catagórica nomina	Se categoriza cada una de las 14 zonas en orden alfabético toman el valor de 1 hasta 14		
12	Destino	Clasifica la Finalidad del crédito (capital de trabajo, inversión, Agro, Negocio, etc.).	Catagórica nomina	Se categoriza cada uno de los 15 destinos en orden alfabético toman el valor de 1 hasta 15		
13	Macro Sector	Clasificación del sector económico del cliente.	Catagórica nomina	Se categoriza cada uno de los 32 macro sectores en orden alfabético toman el valor de 1 hasta 32		
14	Garantía	Tipo de garantía asociada a la operación	Catagórica nomina	USAID	1	
				Garantía No Idónea	2	
				Sin Garantía	3	
15	Ventas	Clasificación interna del tipo de negocio	Catagórica nomina	Comercio	1	
				Manufactura	2	
				Servicios	3	
				Sin Info	4	
16	Segmento	Segmento comercial del cliente según la entidad	Catagórica nomina	Pequeño Empresario	1	
				Autoempleado	2	
				Pequeño AgroEmpresario	3	
				Empresario Emergente	4	
				AgroEmpresario Emergente	5	
17	Tipo de Zona	Tipo de área geográfica donde opera el cliente	Catagórica nomina	Rural Agro	1	
				Urbano	2	
				Rural Urbano	3	
18	Tipo de Crédito	Si es cliente nuevo o renovado	Catagórica nomina	R	1	
				N	2	
19	Disponible	Rango del disponible financiero del cliente en cuartiles.	Pesos COP	LI	LS	Categoría
				-	250.000	1
				250.000	450.000	2
				450.000	780.000	3
				780.000		4
20	Activos Fijos	Valor de activos fijos del negocio en cuartiles	Pesos COP	LI	LS	Categoría
				-	5.200.000	1
				5.200.000	21.500.000	2
				21.500.000	73.330.000	3

No.	Variable	Descripción	Unidad	Categorización		
				73.330.000		4
				LI	LS	Categoría
21	Gastos de Negocio	Rangos de los gastos operativos mensuales del negocio en cuartiles.	Pesos COP	-	100.000	1
				100.000	250.000	2
				250.000	750.000	3
				750.000		4
				LI	LS	Categoría
22	Gastos Familiares	Rango del gasto mensual del hogar del solicitante en cuartiles	Pesos COP	-	1.000.000	1
				1.000.000	1.280.000	2
				1.280.000	1.896.000	3
				1.896.000		4
				LI	LS	Categoría
23	Experiencia	Meses de experiencia en la actividad económica clasificado en cuartiles	Meses	-	48	1
				48	84	2
				84	150	3
				150		4
				LI	LS	Categoría
24	Tiempo de Posesión	Tiempo que el solicitante lleva con el negocio o activo productivo clasificado en cuartiles	Meses	-	36	1
				36	72	2
				72	120	3
				120		4
25	Tipo de Seguro Medico	Tipo de régimen de salud	Categórica nominal	Contributivo		1
				Subsidiado		2
				Ninguno		3
26	Tipo de Vivienda	Tipo de vivienda del cliente.	Categórica nominal	Propia		1
				Familiar		2
				Arrendada		3
				Viviente		4
27	Estrato	Nivel socioeconómico colombiano del 1 al 6.	Categórica Ordinal	Se categoriza cada uno de los 6 estratos en orden de menor a mayor asignando el valor de 1 hasta 6		
				LI	LS	Categoría
28	Número de Hijos	Número de hijos o dependientes económicos por cuartiles	Numérica	-	1	1
				1	2	2
				2	3	3
				3		4
29	Nivel Educativo	Máximo nivel educativo alcanzado.	Categórica nominal	0		0
				Ninguno o Preescolar		1
				Básica Primaria		2

No.	Variable	Descripción	Unidad	Categorización	
				Básica Sec 6 a 9	3
				Media 10 A 11	4
				Técnico	5
				Superior Universita 1 a 4 anos	6
				Superior Universita 5 años o	7
				M	
				Especialista	8
				Magister o Doctorado	9
				0	0
				Soltero(A)	1
				Casado(A)	2
				Casado Extranjero doc	2
				invalido	
30	Estado Civil	Estado civil del solicitante	Categórica nominal	Casado/convivencia menor edad	2
				Casado(a) dinero al instante	2
				Union libre < 2 años	3
				Union libre > 2 años	3
				Divorciado-separado(a)	4
				Viudo(a)	5
				04. Autoempleado	1
				04, Autoempleado	1
				02. Propietario Negocio	2
				Pequeñ	
				02, Propietario Negocio	2
				Pequeñ	
				02. Pequeño Empresario	3
				Servicios	
				Pequeño Empresario Comercio	4
31	Tipo de Ingreso	Tipo de ingreso principal del solicitante	Categórica nominal	Pequeño Empresario	5
				Producción	
				01, Empresario Emergente	6
				01. Empresario Emergente	6
				03. Pequeño productor	7
				agropecuario	
				03, Pequeño productor	7
				agropecuario	
				17. Productor Agropecuario de	8
				bajos ingresos	

No.	Variable	Descripción	Unidad	Categorización		
				17, Productor Agropecuario de bajos ingresos	8	
				07. Pensionad	9	
32	Sector	Sector económico donde opera el cliente.	Categoría nominal	Comercio	1	
				Producción	2	
				Servicios	3	
				Agropecuario	4	
33	Liquidez Calculada	Estimación de la capacidad de pago segmentado en cuartiles	Pesos COP	LI	LS	Categoría
				-	576.104	1
				576.104	922.000	2
				922.000	1.717.854	3
				1.717.854		4
34	Activo / Patrimonio	Relación entre activos y patrimonio segmentado en cuartiles	Porcentaje	LI	LS	Categoría
				-	1,00	1
				1,0	1,04	2
				1,0	1,19	3
				1,2		4
35	Pasivo / Activo	Porcentaje de endeudamiento del separado por cuartiles	Porcentaje	LI	LS	Categoría
				-	0,02	1
				0,0	0,05	2
				0,0	0,12	3
				0,2		4
36	Capacidad de pago / Patrimonio	Relación entre capacidad de pago y patrimonio, segmentado en cuartiles	Porcentaje	LI	LS	Categoría
				-	0,00	1
				0,0	0,04	2
				0,1	0,16	3
				0,1		4
37	Días de Mora t-1	Días de mora en el último mes.	Días	LI	LS	Categoría
				-	30,00	1
				31,0	60,00	2
				61,0	90,00	3
				91,0		4
38	Días de Mora t-3	Días de mora hace 3 meses de la fecha de corte.	Días	LI	LS	Categoría
				-	30,00	1
				31,0	60,00	2
				61,0	90,00	3
				91,0		4
39	Días de Mora t-6	Días de mora hace 6 meses de la fecha de corte.	Días	LI	LS	Categoría
				-	30,00	1

No.	Variable	Descripción	Unidad	Categorización		
				31,0	60,00	2
				61,0	90,00	3
				91,0		4
				LI	LS	Categoría
				-	30,00	1
40	Días de Mora t-9	Días de mora hace 9 meses de la fecha de corte.	Días	31,0	60,00	2
				61,0	90,00	3
				91,0		4
				LI	LS	Categoría
				-	30,00	1
41	Días de Mora t-12	Días de mora hace 12 meses de la fecha de corte.	Días	31,0	60,00	2
				61,0	90,00	3
				91,0		4

7. Resultados

Esta sección tiene como finalidad proporcionar un análisis descriptivo sobre las variables independientes y su relación con la variable dependiente, así como la estimación del modelo de regresión logística para el segmento popular urbano de la modalidad microcrédito.

A continuación, se realiza el análisis de las variables numéricas:

Figura 1.

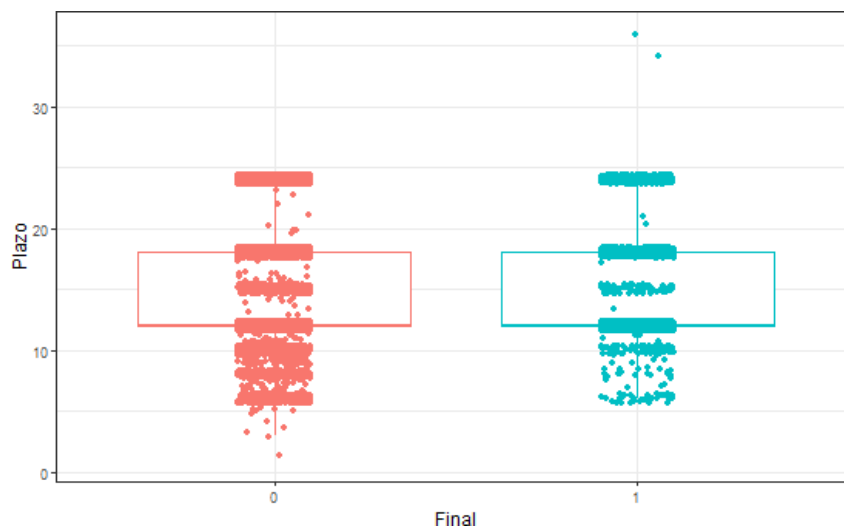
Resumen estadístico de las variables numéricas

	vars	n	mean	sd	median	trimmed	mad	min	max	range
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Plazo	1	21578	14.51	4.15	12.00	14.02	0.00	1.00	36.00	35.00
Tasa	2	21578	60.80	7.95	63.00	60.61	13.34	44.00	72.00	28.00
MontoDesembolsado	3	21578	2494539.66	1525998.74	2000000.00	2281900.82	1482600.00	580000.00	8541000.00	7961000.00
Cuota	4	21578	221361.68	102438.40	213804.52	209935.61	104127.71	60178.32	3124665.66	3064487.34
Ventas	5	21578	4330493.59	2319647.65	3778750.00	4020598.42	1946283.15	561000.00	13200000.00	12639000.00
DISPONIBLE	6	21578	417180.68	262470.55	350000.00	386221.63	222390.00	0.00	1370000.00	1370000.00
ACTIVOSFIJOS	7	21578	34837943.07	39140102.29	16075000.00	28083119.38	20748987.00	0.00	163050000.00	163050000.00
GASTOSNEGOCIO	8	21578	306320.83	315394.88	200000.00	249190.77	192738.00	0.00	1475000.00	1475000.00
GASTOSFAM	9	21578	1227942.23	429123.01	1125000.00	1177270.17	333585.00	0.00	2922000.00	2922000.00
EXPNEGOCIO	10	21578	89.71	64.94	60.00	80.58	53.37	2.00	252.00	250.00
TIEMPOSESION	11	21578	81.85	62.27	60.00	72.57	53.37	6.00	246.00	240.00
NUMHIJOS	12	21578	2.07	1.46	2.00	1.98	1.48	0.00	6.00	6.00
LiquidezCalculada	13	21578	828471.41	510401.57	688736.50	748332.42	371741.93	65637.00	3010941.00	2945304.00
Activo/Patrimonio	14	21578	1.06	0.09	1.02	1.04	0.03	1.00	1.43	0.43
Pasivo/Activo	15	21578	0.05	0.07	0.02	0.04	0.03	0.00	0.30	0.30
Cap_Pat	16	21578	0.07	0.06	0.04	0.06	0.05	0.00	0.27	0.27
DiasMora(-1)	17	21578	0.40	2.96	0.00	0.00	0.00	0.00	59.00	59.00
DiasMora(-3)	18	21578	0.20	1.44	0.00	0.00	0.00	0.00	32.00	32.00
DiasMora(-6)	19	21578	0.14	1.00	0.00	0.00	0.00	0.00	19.50	19.50
DiasMora(-9)	20	21578	0.14	1.00	0.00	0.00	0.00	0.00	19.50	19.50
DiasMora(-12)	21	21578	0.14	1.00	0.00	0.00	0.00	0.00	19.50	19.50

El plazo: de las obligaciones presenta una mediana de 12 meses, lo que indica que la mayoría de los créditos otorgados corresponden a períodos cortos, característica típica del microcrédito. La dispersión es baja y la distribución es ligeramente asimétrica hacia la izquierda, evidenciando que los plazos más largos (hasta 36 meses) son menos frecuentes.

Figura 2.

Gráfico Box Plot – Y= Default X= Plazo

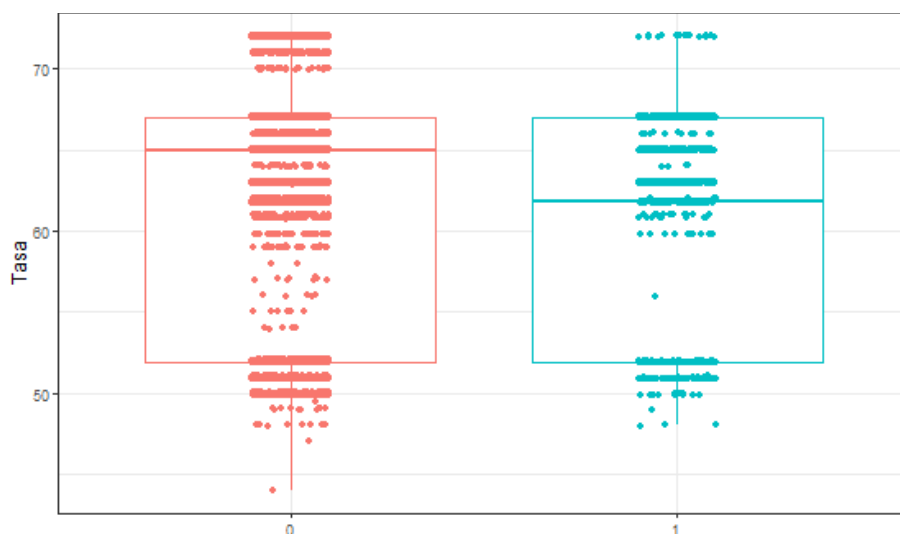


En la gráfica anterior se observa que los créditos en incumplimiento presentan una mediana de plazo ligeramente mayor, lo que sugiere que las obligaciones con períodos más largos tienden a estar más expuestas al riesgo de deterioro. Además, el grupo de incumplimiento presenta mayor dispersión, reflejando una mayor heterogeneidad en la duración de los créditos. Ambos grupos exhiben valores atípicos, especialmente en plazos elevados. Aunque existen diferencias entre los dos grupos, se observa una superposición considerable de las distribuciones, por lo que el Plazo aporta información relevante, pero no es una variable completamente discriminante por sí sola. No obstante, su relación con el incumplimiento justifica su inclusión en el modelo logístico.

Las tasas de interés: se mantienen dentro de un rango estrecho (44 % – 72 %), con una mediana cercana al 63 %. La baja dispersión confirma que este parámetro es regulado institucionalmente, por lo que su variación entre clientes es mínima y no presenta valores extremos relevantes.

Figura 3.

Gráfico Box Plot – Y= Default X= Tasa

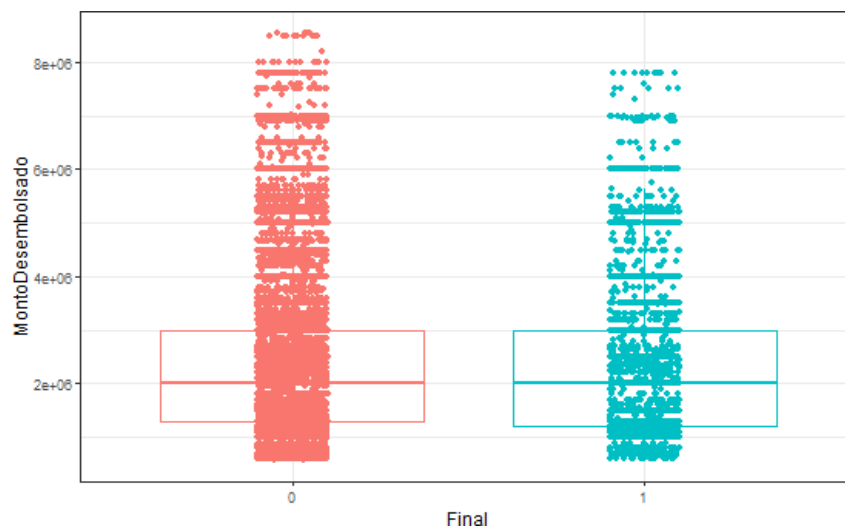


En la gráfica anterior se evidencia que la mediana de la Tasa es ligeramente menor en el grupo que incumple, lo que indica que, en estos datos, las tasas más altas no están asociadas a mayor riesgo de incumplimiento. Ambos grupos presentan amplia dispersión y valores atípicos, y sus distribuciones se superponen de forma importante. Por ello, la Tasa no es una variable claramente discriminante por sí sola, aunque puede aportar información complementaria en el modelo logístico.

El monto desembolsado: muestra una mediana de dos millones de pesos, con una media algo superior, lo que sugiere la presencia de operaciones de mayor tamaño dentro del portafolio urbano. Aun así, los valores permanecen dentro de los límites normales de la oferta de microcrédito, reflejando la coexistencia de microempresas de distintos tamaños.

Figura 4.

Gráfico Box Plot – Y= Default X= Monto Desembolsado



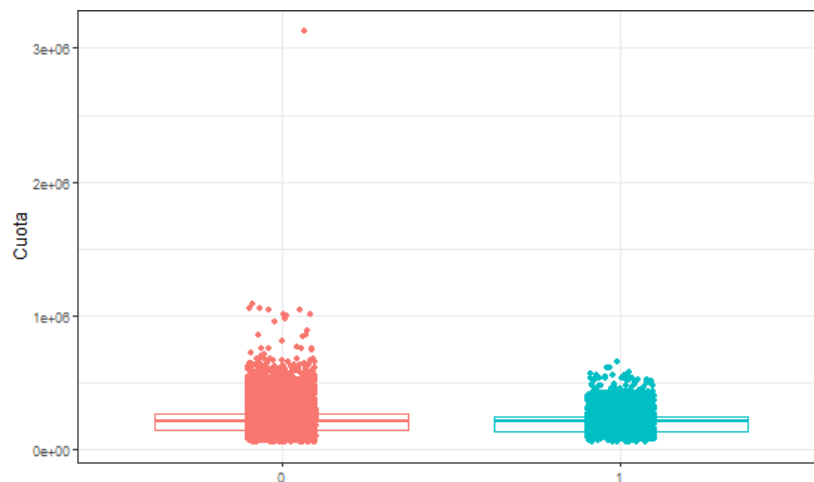
En la gráfica anterior se evidencia que la mediana del Monto Desembolsado es muy similar entre los créditos que incumplen y los que no incumplen, lo que indica que el monto otorgado no

presenta diferencias claras entre los dos grupos. Aunque existen valores muy altos en ambos casos, estos corresponden a outliers que elevan la dispersión, pero no cambian la posición central de la distribución.

Las cuotas: periódicas presentan un comportamiento similar al del monto desembolsado, con mediana alrededor de 213 mil pesos. Existen algunos casos con cuotas significativamente más altas, que elevan la dispersión, posiblemente asociados a créditos de mayor monto o menor plazo.

Figura 5.

Gráfico Box Plot – Y= Default X= Cuota

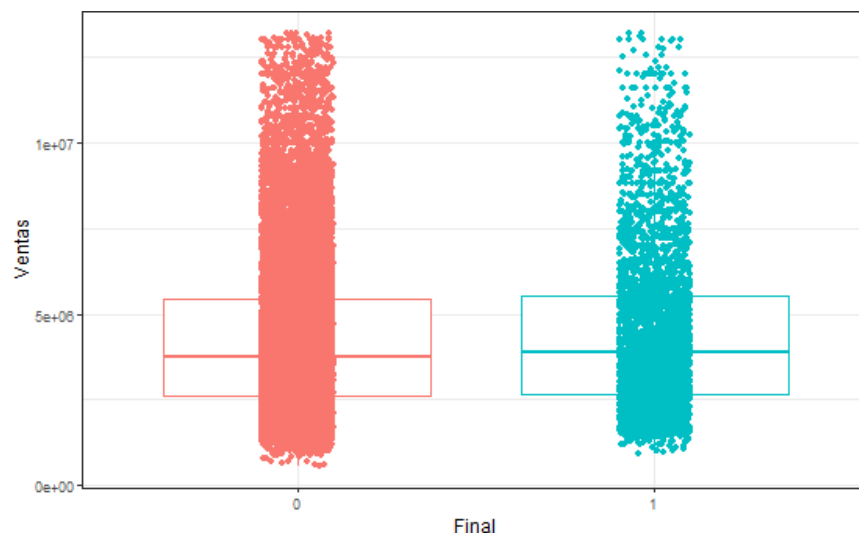


En la gráfica anterior se evidencia que la variable Cuota presenta una distribución altamente asimétrica, con numerosos valores atípicos de gran magnitud en ambos grupos. El rango intercuartílico es similar entre las dos categorías, lo que indica que, para la mayoría de las observaciones, los niveles de cuota no difieren significativamente entre quienes incumplen y quienes no. Esto sugiere que la variable Cuota por sí sola no posee un poder discriminatorio fuerte en la separación de ambos grupos.

Las ventas: registradas por los clientes presentan una mediana cercana a 3.8 millones de pesos, aunque la media asciende a 4.3 millones debido a la presencia de negocios con ingresos considerablemente superiores al promedio. Esta variabilidad es típica del sector, donde conviven actividades informales y microempresas consolidadas.

Figura 6.

Gráfico Box Plot – Y= Default X= Ventas

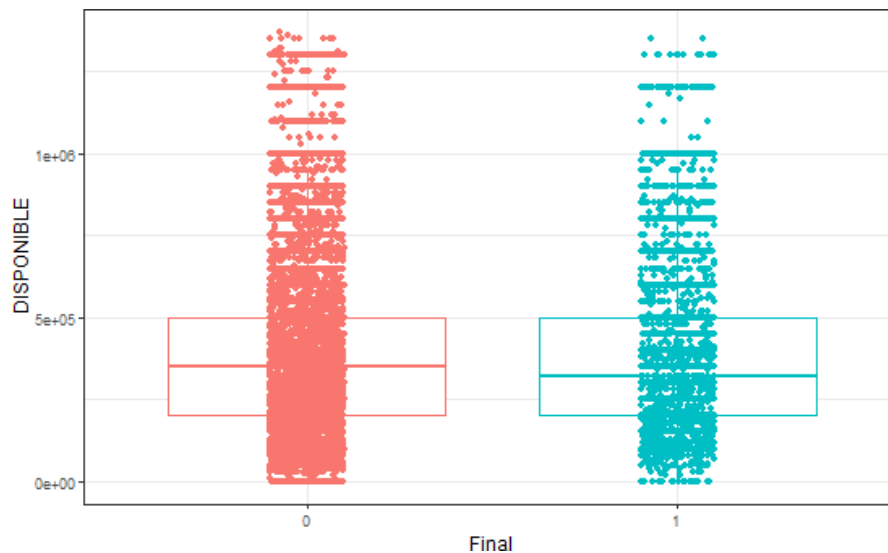


En la gráfica anterior se evidencia que la variable Ventas presenta una alta dispersión y una marcada asimetría positiva en ambos grupos, evidenciada por la gran cantidad de valores atípicos. El rango intercuartílico es similar entre ambos casos, lo que indica que, en términos centrales, los niveles de ventas no exhiben diferencias relevantes entre quienes incumplen y quienes no.

El disponible: entendido como liquidez inmediata, presenta valores moderados, con mediana de 350 mil pesos. La distribución está influenciada por algunos negocios con disponibilidades mucho mayores, reflejando diferencias marcadas en la capacidad de liquidez entre clientes.

Figura 7.

Gráfico Box Plot – Y= Default X= Disponible

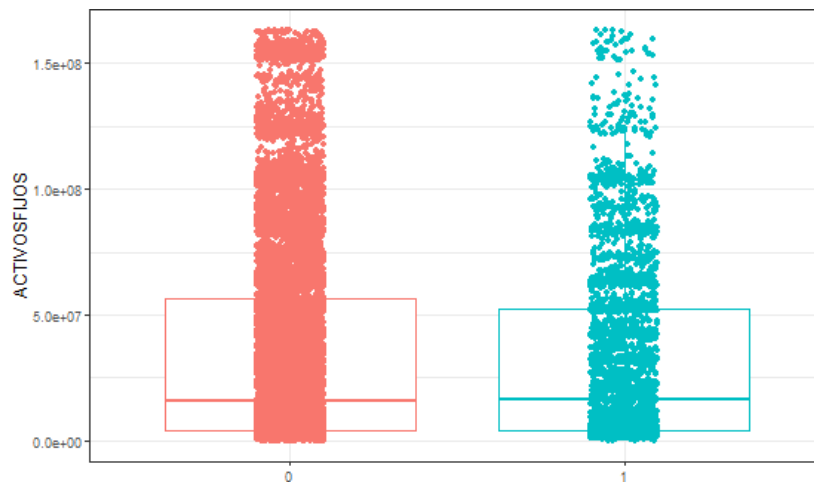


En la gráfica anterior se evidencia que la variable Disponible muestra una distribución altamente dispersa y con fuerte asimetría positiva en ambos grupos de la variable Final (0 = no incumple, 1 = incumple). Se observa un número considerable de valores atípicos, lo que indica que algunos clientes presentan disponibilidades significativamente mayores que el resto. El rango intercuartílico es muy similar entre los dos grupos, lo que sugiere que, en su comportamiento central, el nivel de recursos disponibles no presenta diferencias sustanciales entre quienes incumplen y quienes no.

Los activos fijos evidencian una alta heterogeneidad: mientras muchos negocios operan con infraestructura mínima, algunos registran montos muy superiores, lo que incrementa la variabilidad. La mediana cercana a 1.6 millones de pesos confirma que la mayoría de los microempresarios cuenta con equipamiento básico.

Figura 8.

Gráfico Box Plot – Y= Default X= Activos Fijos

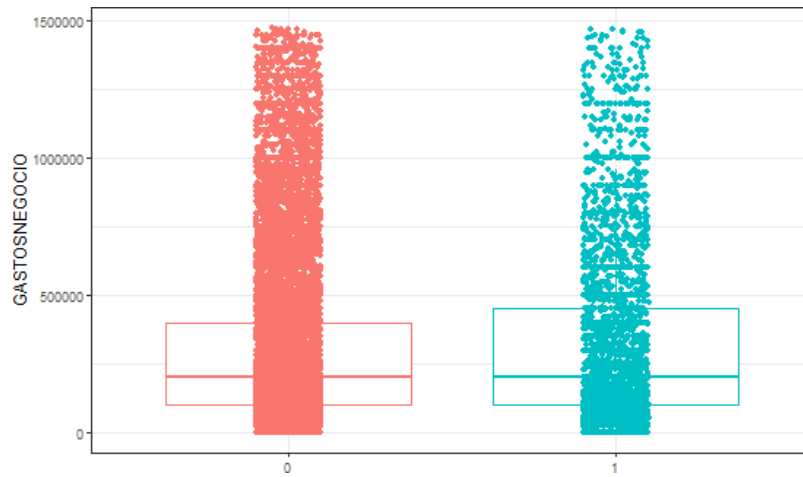


En la gráfica anterior se evidencia que la variable Activos Fijos, segmentado entre clientes incumplidos (1) y no incumplidos (0), evidencia que ambas distribuciones son muy similares. Tanto las medianas como los rangos intercuartílicos muestran valores cercanos entre los dos grupos, mientras que se observa una alta dispersión y presencia de numerosos valores extremos en ambos casos. Estos resultados indican que Activos Fijos no presenta una capacidad discriminante significativa entre clientes que incumplen y aquellos que no.

Los gastos de negocio y gastos familiares: muestran valores acordes a los niveles de actividad económica del segmento urbano. Los gastos del negocio tienen una mediana de 200 mil pesos, mientras que los gastos familiares rondan 1.1 millones. La dispersión está asociada a diferencias en tamaño, actividad y composición familiar de los hogares.

Figura 9.

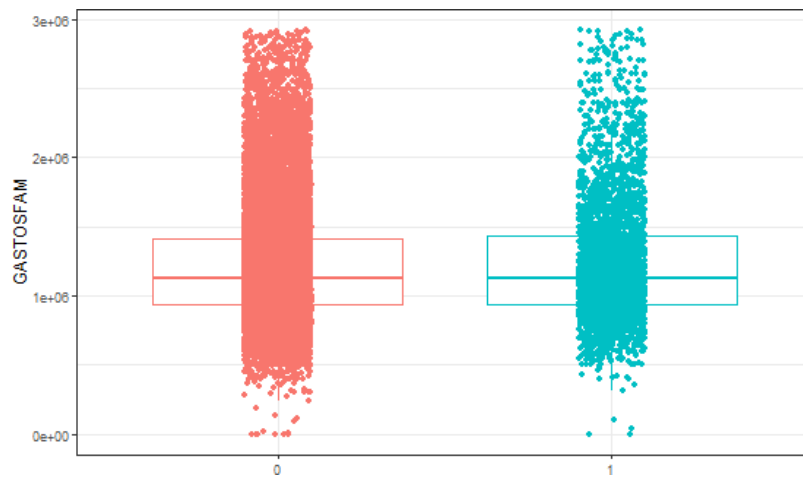
Gráfico Box Plot – Y= Default X= Gastos de Negocio



En la gráfica anterior se evidencia que la variable Gastos de Negocio, comparando a los clientes no incumplidos (0) e incumplidos (1), muestra que ambas distribuciones son muy similares. En ambos grupos se observa una alta dispersión y una gran cantidad de valores atípicos, lo que evidencia una marcada heterogeneidad en los niveles de gasto. Asimismo, las medianas y los rangos intercuartílicos presentan diferencias mínimas entre los dos segmentos, lo que indica que esta variable no permite distinguir de manera clara el comportamiento de pago de los clientes.

Figura 10.

Gráfico Box Plot – Y= Default X= Gastos Familiares

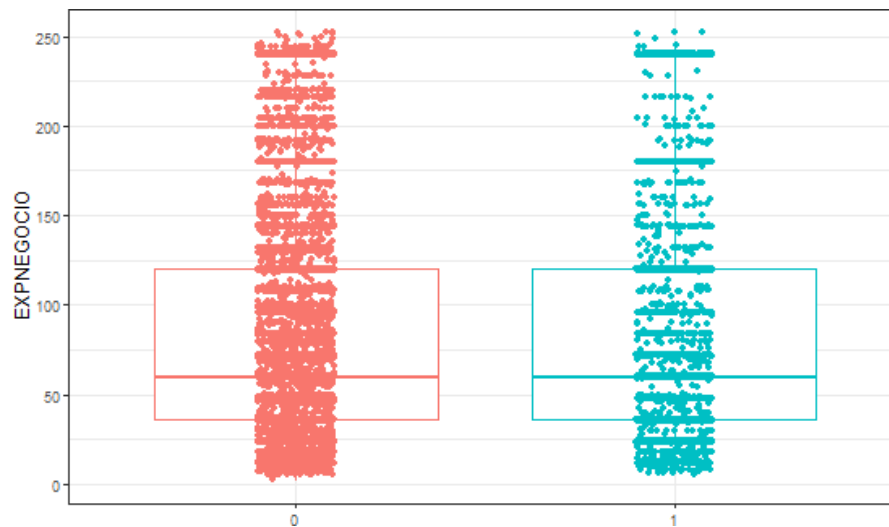


En la gráfica anterior se evidencia que la variable Gastos Familiares muestra distribuciones altamente similares en ambos grupos. Se observa una amplia dispersión y una gran presencia de valores extremos, lo que refleja diferencias significativas en los niveles de gasto familiar dentro de cada categoría. Las medianas y los rangos intercuartílicos son casi equivalentes, por lo que no se evidencian diferencias sustanciales entre los clientes que cumplen y los que incumplen con sus obligaciones crediticias.

La experiencia del negocio y el tiempo de posesión de la actividad: presentan medianas de 60 y 72 meses respectivamente, indicando que la mayoría de los microempresarios cuentan con entre 5 y 7 años de trayectoria. Sin embargo, existen algunos negocios con décadas de operación, lo que explica la asimetría positiva.

Figura 11.

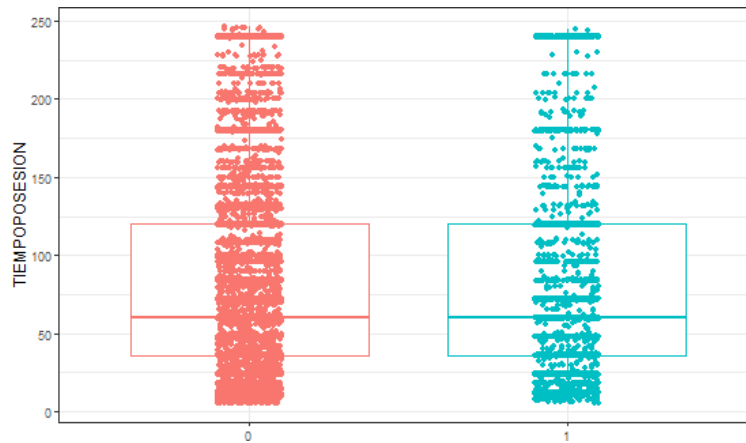
Gráfico Box Plot – Y= Default X= Experiencia del negocio



En la gráfica anterior se observa que ambos grupos presentan medianas y rangos intercuartílicos muy similares, así como una alta dispersión y numerosos valores atípicos en ambos casos. Esto indica que la variable experiencia del negocio no muestra diferencias significativas entre los dos tipos de clientes.

Figura 12.

Gráfico Box Plot – Y= Default X= Tiempo de posesión

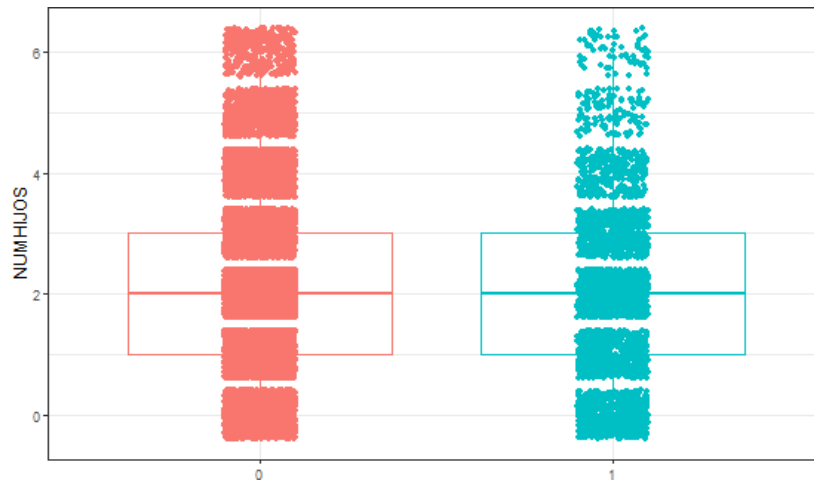


En la gráfica anterior se evidencia que la variable Tiempo de Posesión del Negocio muestra que tanto los clientes no incumplidos (0) como los incumplidos (1) presentan distribuciones muy similares. Las medianas, los rangos intercuartílicos y la dispersión general prácticamente no difieren entre los grupos, y ambos muestran una presencia considerable de valores atípicos.

Número de hijos: El tamaño promedio del hogar es estable, con una mediana de dos hijos. No se observan valores extremos significativos tras la depuración, lo que sugiere información consistente con patrones demográficos urbanos.

Figura 13.

Gráfico Box Plot – Y= Default X= Numero de hijos

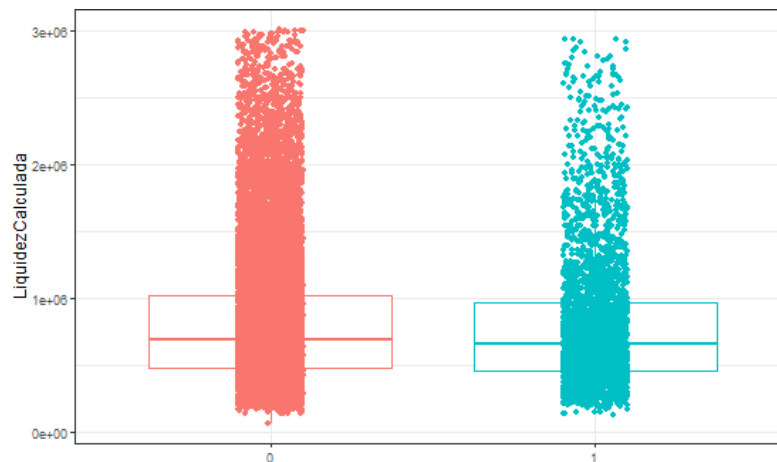


En la gráfica anterior se evidencia que la variable Número de Hijos muestra distribuciones muy similares entre los clientes no incumplidos (0) e incumplidos (1). Las medianas son prácticamente iguales y el rango intercuartílico no presenta diferencias relevantes entre ambos grupos. Aunque existen algunos valores más altos, estos aparecen en ambos tipos de clientes, por lo que no representan una distinción significativa. En conjunto, esta variable no evidencia capacidad discriminante para diferenciar el riesgo de incumplimiento, por lo que su aporte al modelo de regresión logística resulta limitado.

La liquidez: calculada presenta una distribución amplia, pero con valores representativos alrededor de 688 mil pesos. La presencia de algunos negocios con liquidez elevada incrementa la media y la variabilidad, reflejando diferencias en la capacidad operativa.

Figura 14.

Gráfico Box Plot – Y= Default X= Liquidez

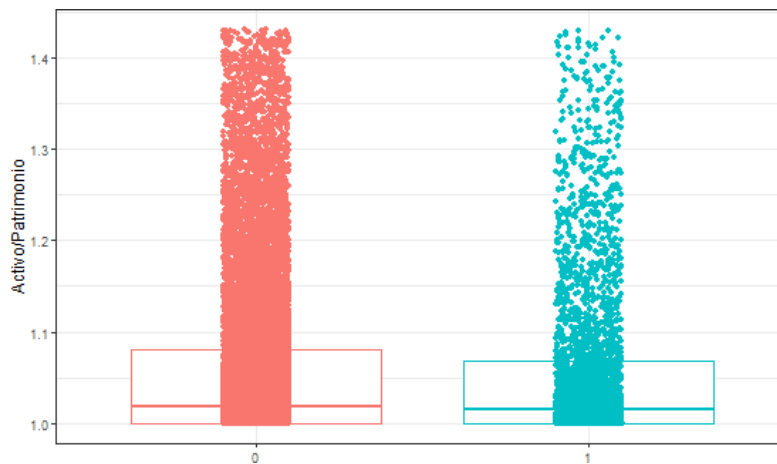


En la gráfica anterior se evidencia que no existen diferencias sustanciales entre ambos grupos, ya que tanto las medianas como los rangos intercuartílicos presentan valores muy similares. Esto indica que, en términos de liquidez, los clientes que incumplen y los que no incumplen exhiben comportamientos similares. Adicionalmente, se observa una alta dispersión de los datos, junto con una gran cantidad de valores atípicos, lo cual es común en variables financieras debido a su naturaleza asimétrica y a la presencia de montos extremos.

Los indicadores financieros (Activo/Patrimonio, Pasivo/Activo y Cap_Pat): en general, muestran niveles controlados de endeudamiento y una relación activa-patrimonio estable, con medianas cercanas a 1. Estos resultados son coherentes con microempresas que operan con recursos propios, con niveles moderados de apalancamiento.

Figura 15.

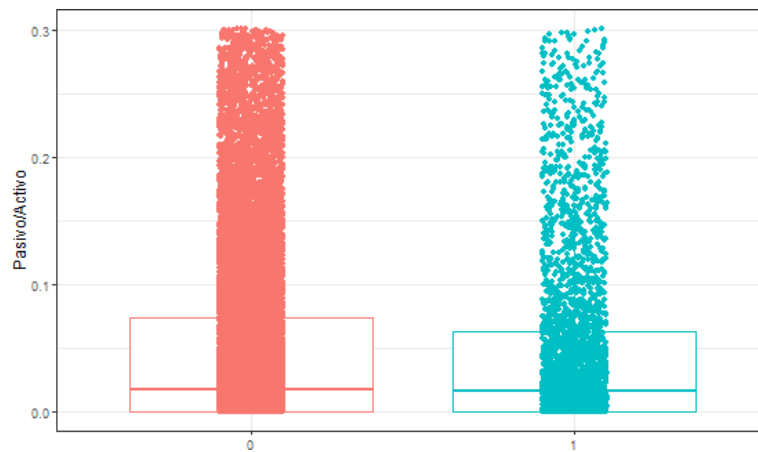
Gráfico Box Plot – Y= Default X= Activo sobre patrimonio



En la gráfica anterior se evidencia que la variable Activo/Patrimonio representa un indicador financiero de apalancamiento que permite evaluar el nivel de solvencia y riesgo estructural del cliente. El boxplot muestra que ambas distribuciones presentan medianas y rangos intercuartílicos muy similares, lo que indica que este indicador no evidencia diferencias relevantes entre los dos grupos. También se observa una alta concentración de valores atípicos, característica común en variables financieras de apalancamiento.

Figura 16.

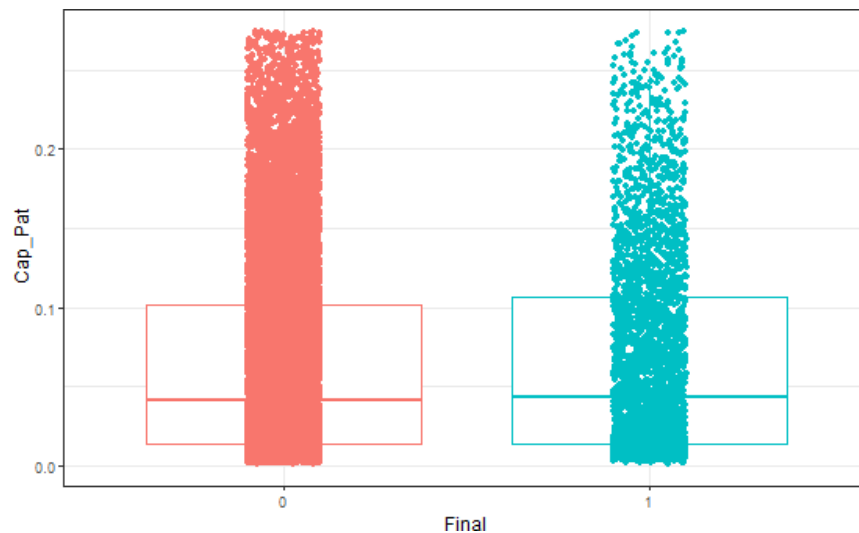
Gráfico Box Plot – Y= Default X= Pasivo sobre activo



En la gráfica anterior se evidencia que la razón Pasivo/Activo, que representa el nivel de endeudamiento de los clientes, muestra que las distribuciones de los grupos no incumplidos (0) e incumplidos (1) son muy similares, con medianas y rangos intercuartílicos prácticamente equivalentes. Esto indica que el nivel de endeudamiento no presenta diferencias marcadas entre ambos tipos de clientes. Asimismo, se observa una alta dispersión y presencia de valores atípicos, lo cual es habitual en indicadores financieros donde existen clientes con niveles de endeudamiento significativamente superiores al promedio.

Figura 17.

Gráfico Box Plot – Y= Default X= Capacidad de pago sobre patrimonio

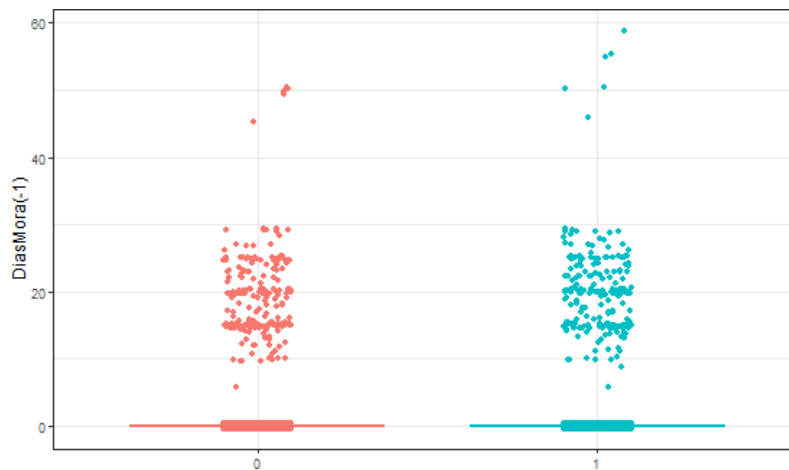


En la gráfica anterior se evidencia que la variable Capacidad de Pago sobre Patrimonio (Cap_Pat) muestra que no existen diferencias relevantes entre los clientes no incumplidos (0) y los incumplidos (1). Las medianas y los rangos intercuartílicos son similares en ambos grupos. Asimismo, la presencia de valores atípicos es comparable para las dos categorías.

Las variables de comportamiento: registran días de mora en diferentes ventanas de tiempo presentan mediana igual a cero en todos los casos, lo cual confirma que la mayoría de los clientes se mantiene al día en sus obligaciones. Sin embargo, existen casos aislados con días de mora mayores, que aportan información relevante para la estimación del riesgo de incumplimiento.

Figura 18.

Gráfico Box Plot – Y= Default X= Días de mora t-1



En la gráfica anterior se evidencia que la variable Días de mora en t-1 evidencia diferencias más claras entre los clientes no incumplidos (0) y los incumplidos (1). Aunque ambos grupos presentan una mediana cercana a cero, el grupo de incumplidos muestra una mayor dispersión y más valores atípicos, alcanzando niveles de mora considerablemente superiores. Esta mayor concentración de días de atraso previos en el grupo 1 sugiere que la mora histórica es un indicador relevante del riesgo de incumplimiento, aportando capacidad discriminante al modelo de regresión logística.

En la figura 19., se presenta la variable Días de mora en t-3 la cual muestra un patrón similar al observado en t-1: aunque la mediana es cercana a cero para ambos grupos, el grupo de incumplidos (1) exhibe mayor dispersión y más valores atípicos en comparación con los no incumplidos (0). En particular, los clientes incumplidos presentan atrasos más elevados y frecuentes en este periodo histórico.

Figura 19.

Gráfico Box Plot – Y= Default X= Días de mora t-3

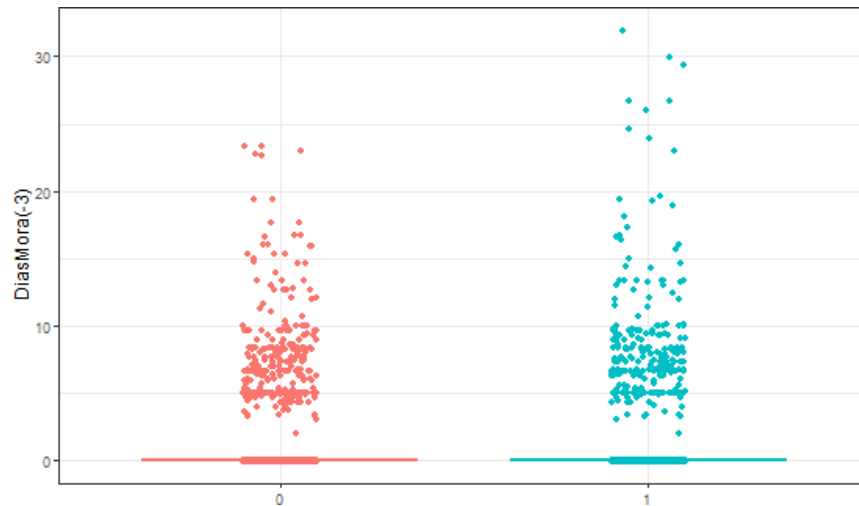
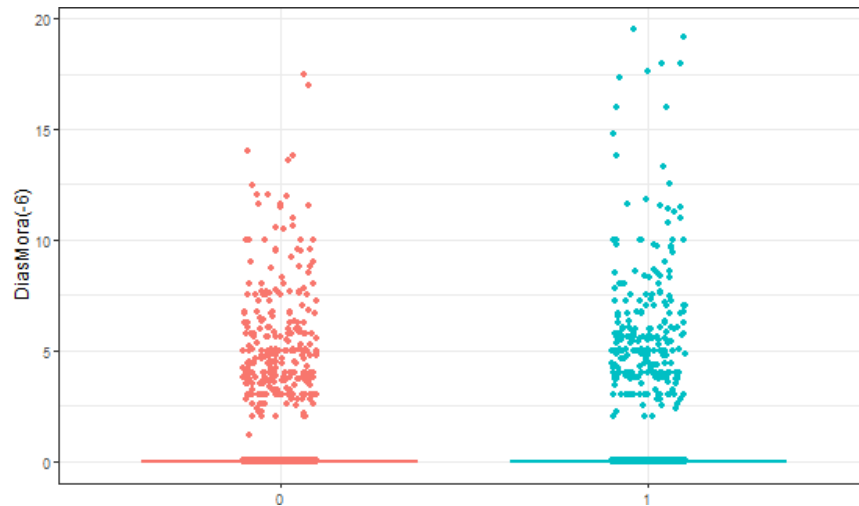


Figura 20.

Gráfico Box Plot – Y= Default X= Días de mora t-6

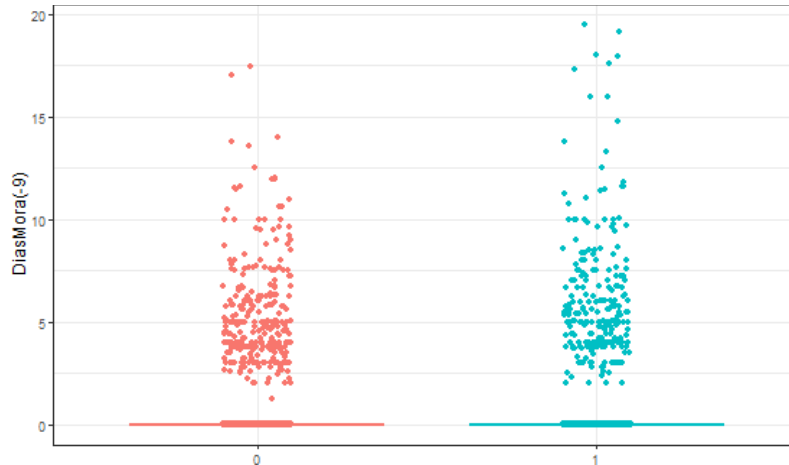


En la gráfica anterior se presenta la variable “Días de mora en t-6” muestra una diferenciación clara entre los estados de incumplimiento. Los clientes incumplidos presentan valores significativamente superiores y una mayor dispersión respecto a los no incumplidos,

concentrando incluso observaciones por encima de 15 días de mora. En contraste, los clientes no incumplidos exhiben niveles bajos y menos variables. Figura 21

Figura 21.

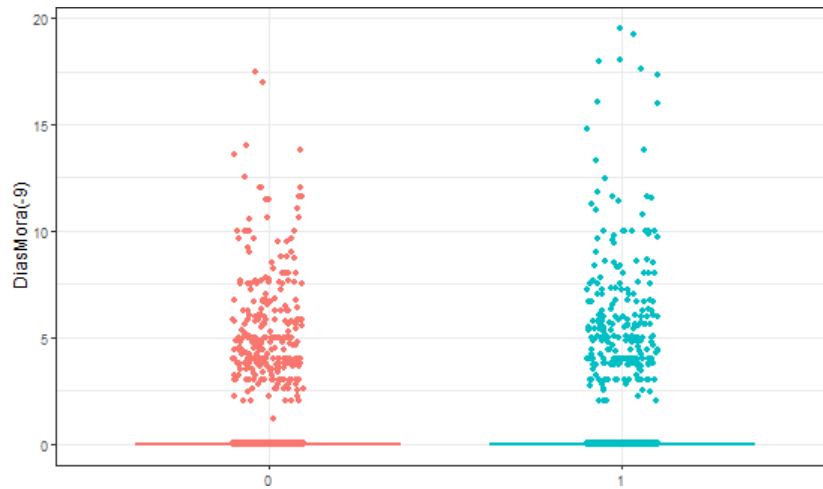
Gráfico Box Plot – Y= Default X= Días de mora t-9



En la gráfica anterior se presenta la variable “Días de mora en t-9”, la cual mantiene el mismo patrón discriminante observado en los periodos anteriores. Los clientes incumplidos registran valores de mora más altos y con mayor dispersión respecto a los no incumplidos, concentrando observaciones que alcanzan niveles superiores a 15 días. En contraste, los clientes no incumplidos presentan una distribución más acotada y valores sustancialmente menores.

Figura 22.

Gráfico Box Plot – Y= Default X= Días de mora t-12

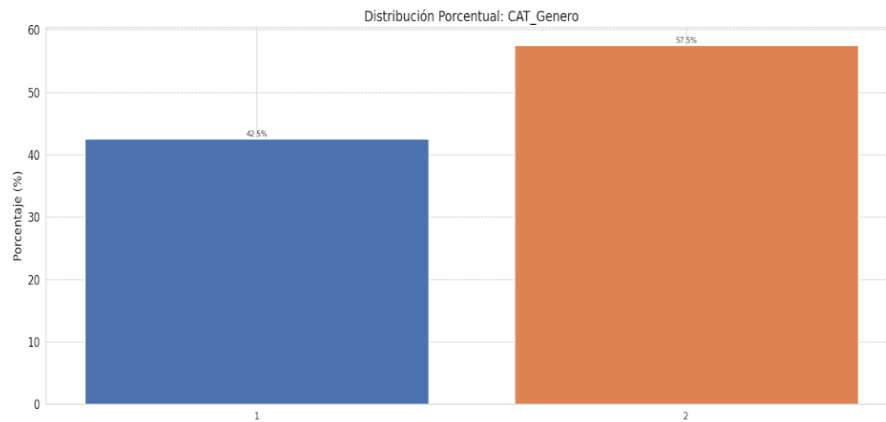


En la gráfica anterior se presenta la variable “Días de mora en t-12”, la cual continúa mostrando un comportamiento diferenciado entre los dos estados de incumplimiento. Los clientes incumplidos presentan valores de mora más elevados y con mayor dispersión, alcanzando registros superiores a 15 días, mientras que los clientes no incumplidos mantienen niveles bajos y una variabilidad más reducida.

A continuación, se realiza el análisis de las variables categóricas: Genero:

Figura 23.

Gráfico Barras Y= frecuencia X= Categoría genero

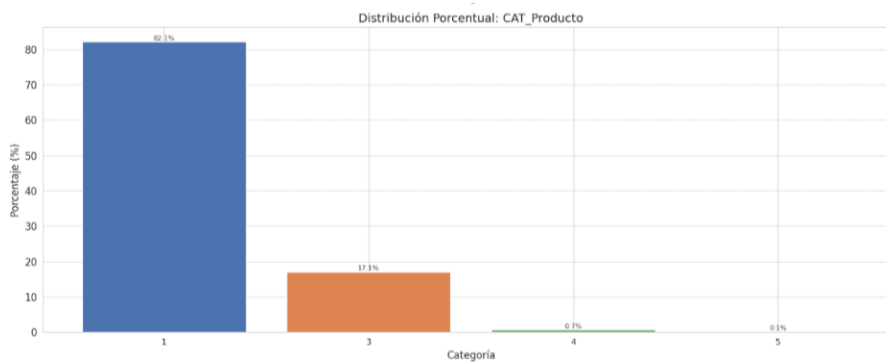


La gráfica muestra una clara concentración del portafolio en la Categoría 2 (Femenino), con una participación significativamente mayor del 57,5%.

Producto:

Figura 24.

Gráfico Barras Y= frecuencia X= Categoría producto



La gráfica revela una fuerte y alta concentración en la Categoría 1 del producto, participando en el 82,1%.

Ciudad:

La cartera se distribuye a un número muy elevado de categorías (ciudades), lo que significa que el total del 100% está distribuido en una base geográfica amplia. Sin embargo, el top 10 de las ciudades concentran el 33,12% del portafolio popular productivo urbano

Tabla 2.

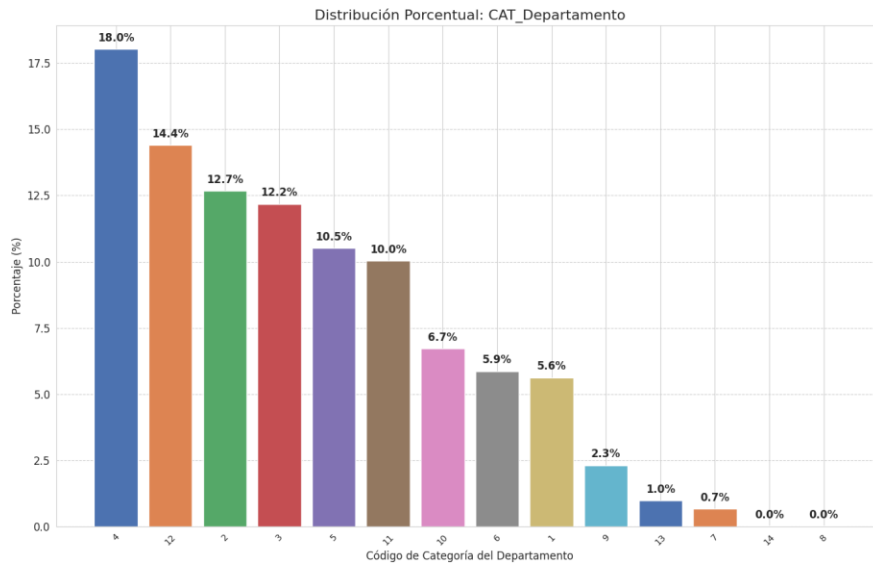
Top 10 de ciudades del portafolio popular productivo urbano

Numero	Variable	Porcentaje Individual	Concentración Acumulada
1	62	4.94%	4.94%
2	32	3.99%	8.94%
3	201	3.98%	12.92%
4	245	3.79%	16.71%
5	3	3.20%	19.91%
6	210	2.99%	22.91%
7	14	2.82%	25.73%
8	128	2.52%	28.25%
9	253	2.49%	30.74%
10	68	2.38%	33.12%

Departamento:

Figura 25.

Gráfico Barras Y= frecuencia X= Categoría departamento

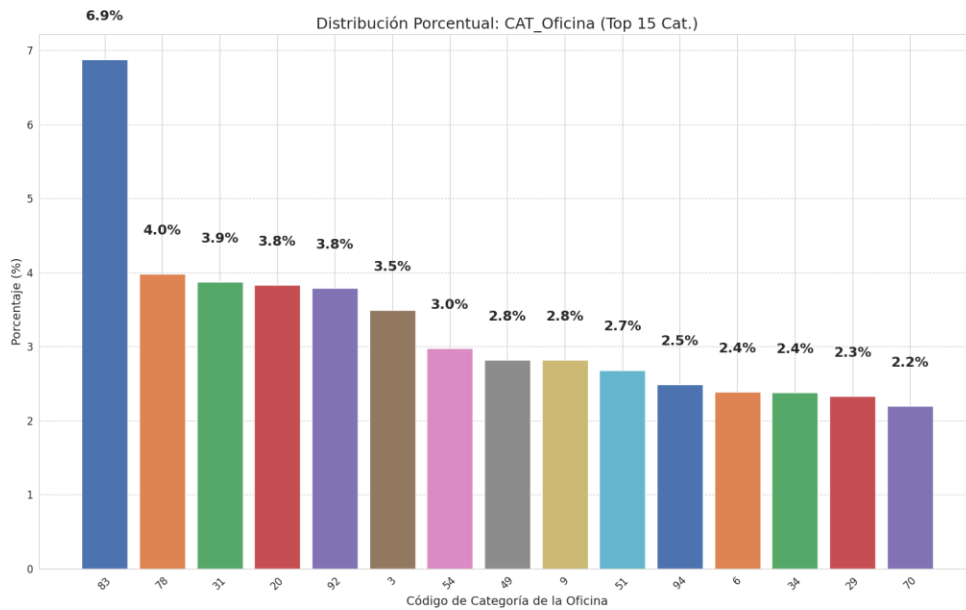


La gráfica ilustra que la cartera está significativamente concentrada en 6 departamentos los cuales corresponden a Bolívar, Sucre, Cesar, Magdalena, Santander y norte de Santander, lo cual concluye que hay una concentración significativa del 77.82% en la región norte del país.

Oficina:

Figura 26.

Gráfico Barras Y= frecuencia X= Categoría oficina top 15

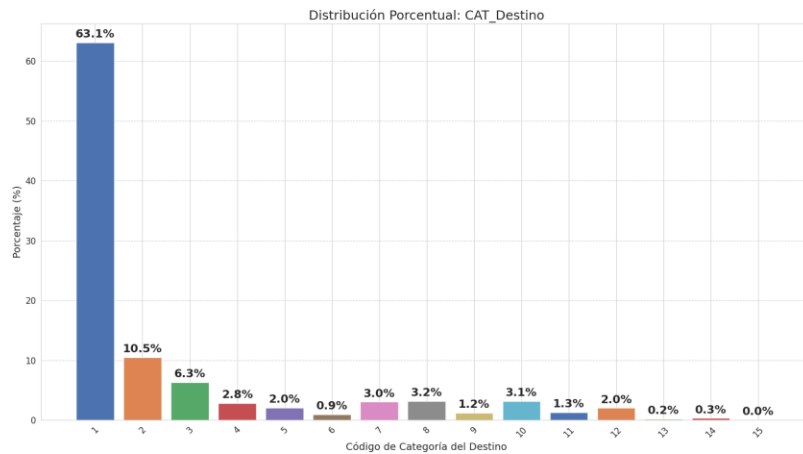


La gráfica confirma que un número muy reducido de oficinas (las primeras categorías) concentra la mayor parte del volumen de la cartera.

Destino:

Figura 27.

Gráfico Barras Y= frecuencia X= Destino

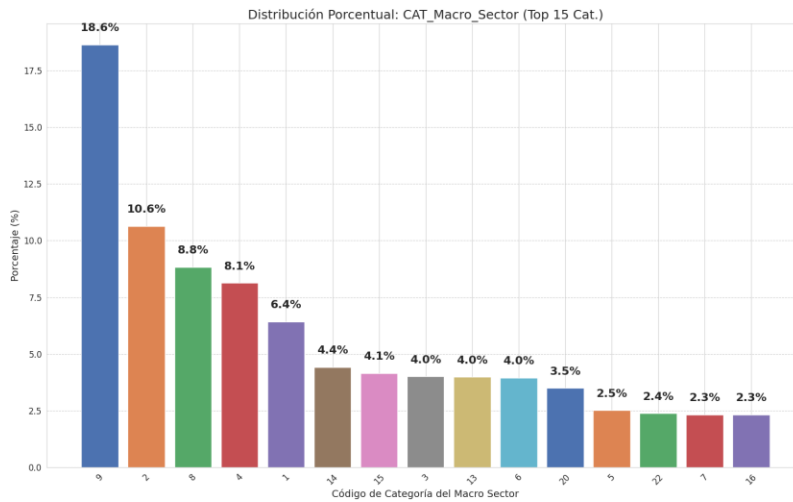


La gráfica muestra que la Categoría 1 (Capital de Trabajo para Negocio) es el destino de crédito dominante, concentrando el 63.1% de la muestra. Lo cual refleja que la mayoría de la financiación está ligada a la operación diaria y a la liquidez inmediata del microempresario.

Macro Sector:

Figura 28.

Gráfico Barras Y= frecuencia X= Macro Sector Top 15

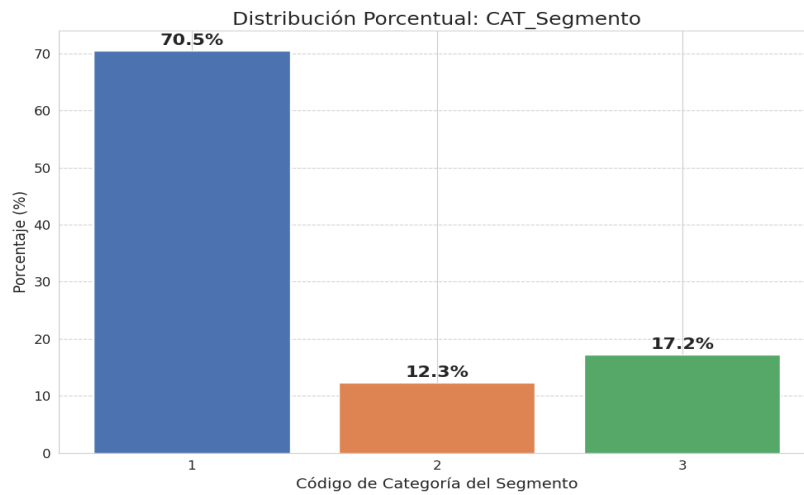


La gráfica revela que la cartera está concentrada en un 18.6% en el Macro Sector 9 (Agricultura, ganadería, silvicultura y pesca) y otros sectores adyacentes 2 (Restaurantes y cafeterías), 8 (Servicios sociales y salud) y 4 (Comercio al por menor de alimentos y víveres), lo cual evidencia una diversificación del portafolio en diferentes macro sectores, donde el 47.4% corresponde a el sector Comercios.

Segmento:

Figura 29.

Gráfico Barras Y= frecuencia X= Segmento

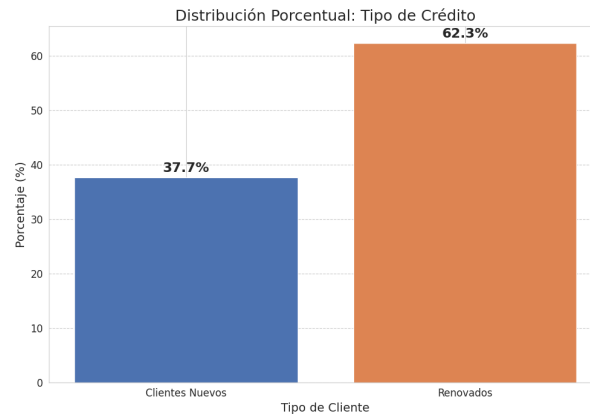


La cartera se concentra en la categoría 1 (pequeño empresario) con el 70.5%, lo que confirma que la entidad está enfocada en su público objetivo popular. La gran mayoría de la cartera responde al riesgo típico de la base microempresarial volatilidad de ingresos y sensibilidad a la demanda.

Tipo de Crédito:

Figura 30.

Gráfico Barras Y= frecuencia X= Tipo de crédito

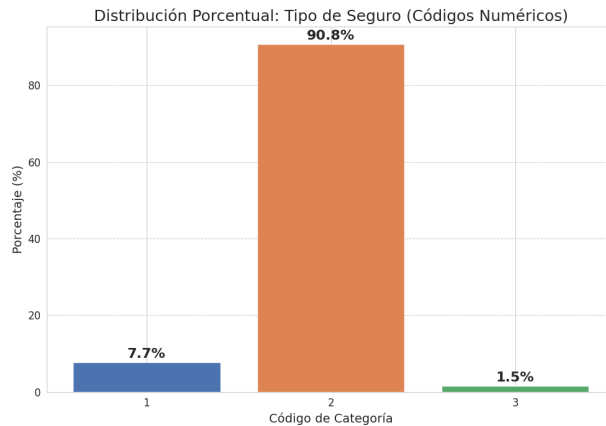


La gráfica muestra que la mayor parte de la cartera urbana, el 62.3% está compuesta por clientes Renovados. Esta composición mitiga el riesgo de crédito, ya que los clientes renovados generalmente tienen un historial de pago conocido y menor probabilidad de mora que los clientes nuevos.

Tipo de seguro:

Figura 31.

Gráfico Barras Y= frecuencia X= Tipo de seguro

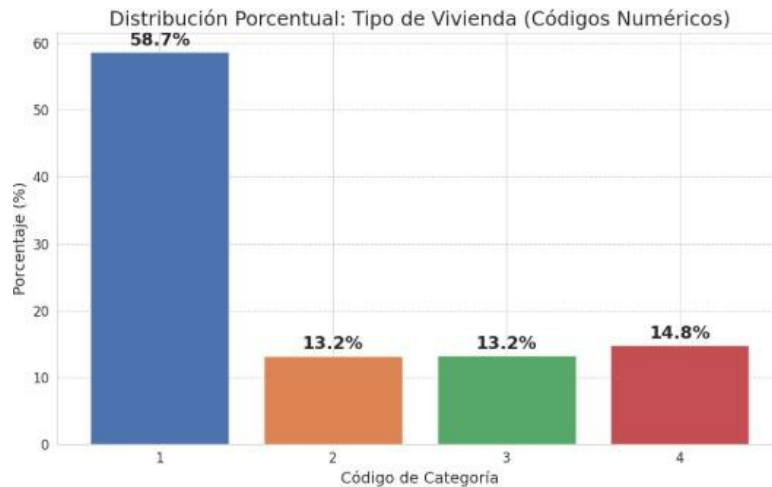


La gráfica muestra que la gran mayoría de los clientes de la cartera urbana el 91 se encuentran bajo el régimen de seguridad social subsidiado. Esta concentración es un indicador clave de la vulnerabilidad socioeconómica de la base de clientes. El régimen subsidiado está dirigido a la población sin capacidad de pago para cotizar, confirmando que el portafolio atiende al segmento de menores ingresos y mayor riesgo ante choques económicos.

Tipo de vivienda:

Figura 32.

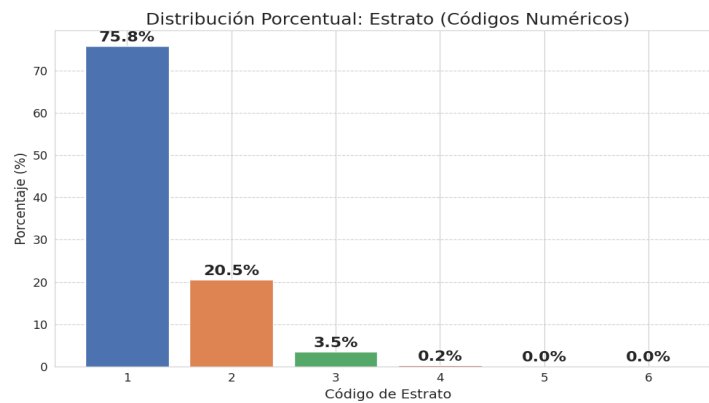
Gráfico Barras Y= frecuencia X= Tipo de vivienda



La Categoría 1 (Propia) domina la cartera con el 58.7%. Si se combina con la Categoría 2 (Familiar) 13.2%, aproximadamente el 71.9% de la cartera reside en una vivienda que no representa un gasto fijo de arrendamiento. La alta proporción de vivienda propia y familiar es un factor mitigante de riesgo, ya que el deudor tiene menos presión sobre el flujo de caja personal, lo que libera recursos para el pago de la cuota.

Figura 33.

Gráfico Barras Y= frecuencia X= Estrato

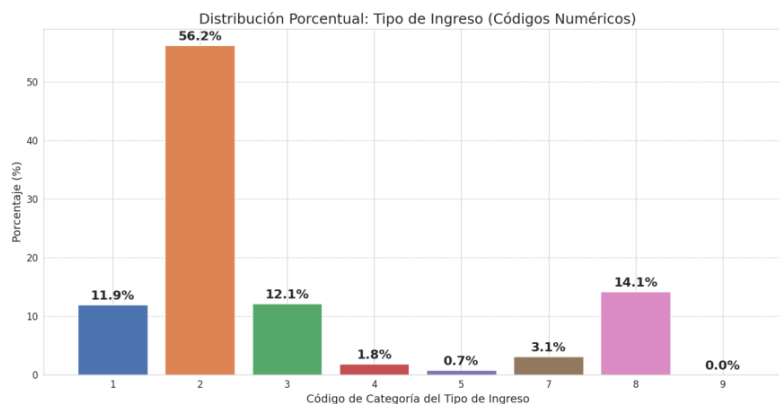


La gráfica muestra una concentración significativa de la cartera en los Estratos 1 (75.8%) y 2 (20.5%), representando el 96.3% del total de clientes. Esta composición confirma que la entidad atiende al segmento de mayor vulnerabilidad socioeconómica. Esto implica un riesgo sistémico alto, ya que la capacidad de pago de estos clientes se ve afectada de manera inmediata a choques macroeconómicos como la inflación o el aumento de tasas.

Tipo de Ingreso:

Figura 34.

Gráfico Barras Y= frecuencia X= Tipo de ingreso



La Categoría 2 (Propietario Negocio Pequeño) concentra el 56.2% de la cartera. Si se suma a la Categoría 1 (Autoempleado) (11.9%), más del 65% de la cartera depende de ingresos no fijos, lo que aumenta el riesgo de volatilidad ante choques económicos.

Estimación del modelo logístico para el segmento popular productivo urbano con el 80% con un default del 30%:

- Selección de la base de entrenamiento y base de prueba:

Dado el default de la base total es del 15.6%, mediante la aplicación de muestreoaleatorio con reemplazo, se logra la proporción 70% Buenos / 30% Malos en la base de entrenamiento mediante Oversampling (repetición de los casos de mora). De este modo, se ajusta el balance en la base de entrenamiento con el objetivo de que el modelo aprenda adecuadamente y que la evaluación del desempeño sea representativa.

Tabla 3.

Resultado del muestreo aleatorio estratificado

Conjunto de Datos	Proporción	Registros	Tasa Default
Original	100%	21.578	15.60%
Entrenamiento	80%	17.262	30%
Prueba	20%	4.315	15.59%

- Análisis de correlación entre variables independientes:

Se realiza el análisis de asociación (colinealidad) entre las variables categóricas mediante el cálculo del coeficiente V de Cramér, que está basado en la prueba de Chi-Cuadrado. El V de Cramér mide la fuerza de la asociación entre dos variables, donde valores cercanos a 1.0 indican una colinealidad fuerte. Las variables de mora fueron excluidas por su varianza casi nula.

La siguiente tabla lista los pares de variables con la mayor colinealidad, las cuales son candidatas para excluir para simplificar el modelo de regresión logística y evitar problemas de interpretación (multicolinealidad).

Tabla 4.

Variables con mayor colinealidad

Variable 1	Variable 2	V de Cramér	Decisión
CAT_Oficina	CAT_Zonal	0.9979	Se excluye del modelo la variable CAT_Oficina y CAT_Zonal
CAT_Ciudad	CAT_Departamento	0.9938	Se excluye del modelo la variable CAT_Ciudad y CAT_Departamento
CAT_Activo__PAT	CAT_Pasivo_Activo	0.9938	Se excluye del modelo la variable CAT_Pasivo_Activo
CAT_Segmento	CAT_Tipo_de_Ingreso	0.9073	Se excluye del modelo la variable CAT_Tipo_de_Ingreso
CAT_Macro_Sector	CAT_Sector	0.8134	Se excluye del modelo la variable CAT_Sector
CAT_Experiencia	CAT_Tiempo_de_Posecion	0.7297	Se excluye del modelo la variable CAT_Tiempo_de_Posecion
CAT_Plazo	CAT_Desembolso	0.7908	Se excluye del modelo la variable CAT_Plazo
CAT_Macro_Sector	CAT_Segmento	0.7581	Se excluye del modelo la variable CAT_Macro_Sector
CAT_Destino	CAT_Segmento	0.7163	Se excluye del modelo la variable CAT_Destino
CAT_Producto	CAT_Segmento	0.7082	Se excluye del modelo la variable CAT_Producto
CAT_Desembolso	CAT_Cuota	0.7011	Se excluye del modelo la variable CAT_Cuota

- **Estimación del modelo logístico generalizado:**

Se realiza la estimación del modelo de regresión logística generalizado sobre la base de entrenamiento estratificada, utilizando el conjunto final de variables limpias de colinealidad, 23 y se adiciona la variable de mora en t -1, para un total de 24 variables. A su vez se utiliza el método de optimización del modelo logístico, mediante el criterio de AIC (Akaike Information Criterion), el cual mide qué tan bien se ajusta el modelo penalizando el número de variables. Un AIC mas bajo indica mejor ajuste del modelo logístico.

Figura 36.

Modelo logístico optimizado

	Coefficiente <td><td>	Odds_Ratio <td><td>	IC_2.5 <td><td>	IC_97.5 <td><td>	p_value <td><td>
(Intercept)	-2.65531607	0.07027662	0.05360536	0.09176254	1.475005e-83
CAT_Genero2	-0.24671110	0.78136639	0.71404589	0.85509068	8.076791e-08
CAT_ClasificacionVentas2	0.06854693	1.07095088	0.91329000	1.25194230	3.940735e-01
CAT_ClasificacionVentas3	-0.142274215	0.86697759	0.76145924	0.98580801	3.021273e-02
CAT_ClasificacionVentas4	0.48519001	1.62448365	1.46964220	1.79529630	2.031940e-21
CAT_Segmento2	-0.27018179	0.76324074	0.66037740	0.87935351	2.159653e-04
CAT_Segmento3	-0.27007656	0.76332105	0.65812670	0.88401627	3.325601e-04
CAT_TipodeZona2	0.15720647	1.17082247	1.03792026	1.32359312	1.127881e-02
CAT_TipodeZona3	0.14997041	1.16179987	1.00752695	1.33950835	3.896975e-02
CAT_TipoCredito2	0.75808364	2.13418244	1.93316443	2.35861762	1.782313e-30
CAT_Disponible2	-0.13318658	0.87530176	0.78856918	0.97141097	1.228164e-02
CAT_Disponible3	-0.14741125	0.86293902	0.76810259	0.96883630	1.280162e-02
CAT_Disponible4	-0.07923668	0.97118658	0.84198843	1.11560836	6.811822e-01
CAT_Experiencia2	-0.11893149	0.88786863	0.79297923	0.99331260	3.844742e-02
CAT_Experiencia3	-0.10162453	0.90336868	0.80846437	1.00885648	7.198320e-02
CAT_Experiencia4	-0.24483189	0.78283613	0.68044367	0.89892834	5.660138e-04
CAT_TipodeSeguro2	0.33557328	1.39888191	1.17577270	1.67504291	1.989697e-04
CAT_TipodeSeguro3	0.07509442	1.07798593	0.71691841	1.58443454	7.097700e-01
CAT_Hijos2	0.12598467	1.13426478	1.01614360	1.26602977	2.468108e-02
CAT_Hijos3	0.14137646	1.15185820	1.01589518	1.30523542	2.699045e-02
CAT_Hijos4	0.12681251	1.13520416	0.98816385	1.30270671	7.199872e-02
CAT_EstadoCivil2	-0.22866366	0.79559608	0.67333920	0.93685140	6.627212e-03
CAT_EstadoCivil3	0.06767257	1.07001490	0.96049062	1.19212227	2.194508e-01
CAT_EstadoCivil4	0.23663833	1.26698281	1.12268903	1.42934149	1.221557e-04
CAT_EstadoCivil5	-0.02964861	0.97078660	0.74526637	1.25019854	8.220332e-01
CAT_ActivoPAT2	0.22659631	1.25432340	1.12727652	1.39570231	3.195435e-05
CAT_ActivoPAT3	0.22062833	1.24685993	1.11404385	1.39528956	1.218019e-04
CAT_ActivoPAT4	0.19362656	1.21364298	1.0434569	1.42017719	1.661620e-02
CAT_DiasMora12	2.15504772	8.62830195	2.43062230	34.17149798	9.616964e-04

El modelo logístico desarrollado permite identificar los factores que influyen de manera significativa en la probabilidad de incumplimiento de los clientes pertenecientes al portafolio de microcrédito urbano. La interpretación se realiza a partir de los coeficientes estimados y sus

respectivos Odds Ratios (OR), lo cual proporciona una medida clara de cómo cambian las probabilidades de incumplimiento frente a una categoría de referencia.

La variable Género mostró un coeficiente negativo (-0.2467), lo cual corresponde a un Odds Ratio de 0.78. Esto indica que los clientes pertenecientes a esta categoría presentan 22% menor probabilidad de incumplir frente a la categoría base. Este hallazgo sugiere un comportamiento de pago más estable dentro de este grupo, contribuyendo a la discriminación del riesgo crediticio.

La categoría ClasificacionVentas3 (Servicios) presenta un coeficiente negativo (-0.1427) y un OR de 0.867, lo que implica una reducción del 13% en la probabilidad de incumplimiento respecto a la categoría base (Comercio). Por su parte, la categoría ClasificacionVentas4 (Sin Info) presenta un OR de 1.62, lo que incrementa la probabilidad de incumplimiento en 62 %.

Las categorías segmento2 (Autoempleado) y segmento 3 (Pequeño AgroEmpresario), presentan odds ratios de 0.76 y 0.76, respectivamente, con efectos negativos sobre la probabilidad de incumplimiento. Esto indica que los clientes de estos segmentos tienen aproximadamente 24% menor probabilidad de incumplir, lo cual puede sugerir modelos de negocio más estables o mayor formalización relativa.

Las categorías relacionadas con la zona (2 y 3) presentan odds ratios entre 1.17 y 1.16, lo que representa un incremento cercano al 16–17% en la probabilidad de incumplimiento.

La variable TipoCredito2 (Nuevos) con un odds ratio de 2.13, constituye uno de los predictores fuertes del modelo. Los clientes en esta categoría tienen más del doble de probabilidad de incumplir respecto a los clientes renovados.

Las categorías 2 y 3 de la variable Disponible presentan odds ratios de 0.87 y 0.86, respectivamente. Esto implica que niveles intermedios de esta variable generan una reducción

moderada (13–14%) en la probabilidad de incumplimiento, posiblemente reflejando una mayor disponibilidad de recursos o liquidez para atender la obligación crediticia.

La variable Experiencia 2 y 4, presentan un odd ratio del 0.88 y 0.78, respectivamente, reduciendo el riesgo en un 11% y 22%. La experiencia del microempresario muestra un efecto protector frente al incumplimiento; mayores niveles de trayectoria se asocian con una administración más sólida del negocio.

La variable Tipo de Seguro2 (Subsidiado) con un odd ratio de 1.40, esta categoría incrementa la probabilidad de incumplimiento en 40%.

Respecto a la variable Número de Hijos (CAT_Hijos2 y CAT_Hijos3) ambas categorías presentan odds ratios de 1.13 y 1.15, respectivamente. Esto se traduce en incrementos del 13– 15% en la probabilidad de incumplimiento, sugiriendo que mayores cargas familiares podrían afectar la disponibilidad de ingresos para el pago del crédito.

En cuanto a la variable Estado Civil 2 (Casado) presentan odd ratio de 0.79, reduciendo la probabilidad en 21%. Mientras que Estado Civil 4 (Divorciado) genera un odd ratio de 1.27, aumentando la probabilidad de incumplimiento en 27%.

Para la variable activo sobre patrimonio, los tres niveles significativos presentan odd ratio entre 1.25 y 1.21, lo que implica un incremento entre 21% y 25% del riesgo de incumplimiento a medida que aumenta el indicador.

Este comportamiento indica que, aunque los clientes poseen activos, su estructura financiera puede no traducirse en mayor capacidad de pago, probablemente por bajos niveles de liquidez.

Finalmente, la variable días mora en t-1, es la más fuerte del modelo, con un odds ratio de 8.63, los clientes en esta categoría (mora mayor a 30 días) tienen 8.6 veces mayor probabilidad

de incumplir en comparación con la categoría base. Esto demuestra que el comportamiento histórico de mora reciente es el predictor más determinante del riesgo en microcrédito urbano.

- Evaluación del modelo sobre la base de prueba:

El modelo se evalúa sobre 4.315 registros de prueba, manteniendo la tasa de mora real del 15.6% y utilizando un umbral de clasificación de 0.5.

Tabla 5.

Matriz de confusión

Real	Predicción 0	Predicción 1
0 (Clientes Buenos)	3.200	442
1 (Clientes Default)	400	273

Accuracy (Exactitud): El modelo clasifica correctamente al 80.4% de los clientes en general.

Specificity (Detección Buenos): Clasifica 88% clientes buenos correctamente.

Sensitivity (Detección Morosos): El modelo detecta al 40.6% de los morosos reales.

- Indicadores de validación estadística

Tabla 6.

Indicadores Estadísticos

KS	GINI	AUC
30.36%	40.92%	73.5%

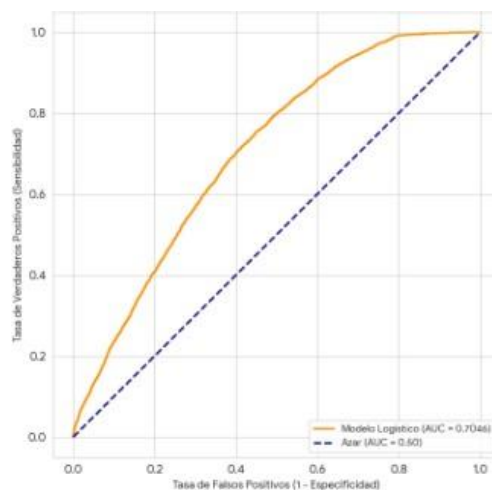
El valor KS 30.36% indica que, en el punto de corte óptimo 0.1628, la diferencia máxima acumulada entre la proporción de clientes en default y la proporción de clientes buenos pagadores es del 30.36%, siendo una discriminación aceptable.

Un valor de GINI de 40.92% indica que el modelo tiene un poder de separación bueno entre los clientes en default y los clientes buenos. Un GINI superior al 40% es un resultado sólido para un modelo de riesgo crediticio basado en variables de originación, confirmando la capacidad predictiva de las variables seleccionadas.

Respecto al área bajo la curva AUC el modelo tiene un buen poder discriminante. Existe una probabilidad del 73.5% de que el modelo clasifique a un cliente default con una probabilidad más alta que a un cliente buen pagador, seleccionados aleatoriamente.

Figura 37.

Curva ROC



- Score de originación

Tabla 7.*Score*

Rangos	Score	Riesgo	Probabilidad de Incumplimiento
609	1000	Riesgo Bajo	4.65%
577	608	Riesgo Medio	10.63%
0	576	Riesgo Alto	16.19%

La segmentación por nivel de score refleja una adecuada discriminación del riesgo, permitiendo a generar políticas y asignación de tasas conforme al riesgo al momento de la originación de créditos, maximizando la relación riesgo rentabilidad del portafolio.

8. Conclusiones

El desarrollo del modelo metodológico permitió identificar y validar las variables determinantes del riesgo crediticio en el segmento de economía popular productiva urbano en Colombia, facilitando un proceso de originación de créditos más eficiente y sistemático. Esto contribuye a una evaluación más precisa de los deudores y fortalece la gestión del riesgo en la compañía.

La implementación del modelo estadístico de regresión logística demostró ser adecuado para estimar la probabilidad de incumplimiento en este segmento, considerando sus

particularidades. El modelo mostró capacidad predictiva robusta y estable, lo que permite tomar decisiones fundamentadas sobre la aprobación de créditos.

Mediante métricas de validación como la curva ROC, el estadístico KS y el índice GINI, se confirmó que el modelo tiene un desempeño satisfactorio en la discriminación del riesgo crediticio, asegurando que las decisiones de otorgamiento se basen en información confiable y cuantificable.

El score crediticio derivado del modelo proporciona un instrumento práctico para la originación de créditos, permitiendo asignar tasas y condiciones de colocación de acuerdo con la exposición al riesgo, equilibrando el crecimiento del portafolio con el control del riesgo crediticio.

Además, el modelo establece las bases para una estrategia proactiva de gestión de riesgo, enfocando los recursos en el monitoreo de la mora temprana y en la revisión de los filtros de renovación y clasificación de ventas, asegurando así decisiones de crédito más informadas y efectivas.

La aplicación de este modelo al segmento de economía popular urbana, compuesto por microempresarios, facilita la inclusión financiera, ofreciendo acceso a crédito de manera responsable y ajustada al riesgo. De esta manera, la compañía no solo mejora su gestión de riesgo, sino que también apoya el desarrollo y fortalecimiento de los microempresarios, fomentando el crecimiento económico local y promoviendo oportunidades sostenibles en este sector.

Referencias Bibliográficas

- Arenas Díaz, M. A., Boccardi Rodríguez, P., & Piñeyrúa Ibáñez, A. (2012). *Credit scoring: evaluación del riesgo crediticio de la cartera de microcréditos de una institución financiera en Uruguay*.
- Comité de Supervisión Bancaria de Basilea. (2006). *Convergencia internacional de medidas y normas de capital: marco revisado (versión integral)*. Banco de Pagos Internacionales.
- Doménech Massons, J. M., & Navarro Pastor, J. B. (2025). *Regresión logística binaria, multinomial, de Poisson y binomial negativa*. Barcelona, España: Graunt21.
- Fernández, A., & Llorca, M. (2018). *Modelos de scoring crediticio y gestión del riesgo de crédito*. Madrid: Ediciones Pirámide.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). Wiley.
- Peña, D. (2010). *Regresión y diseño de experimentos*. Madrid: Alianza Editorial.
- Rayo Cantón, S., Lara Rubio, J., & Camino Blasco, D. (2010). *Un modelo de credit scoring para instituciones de microfinanzas en el marco de Basilea II*. *Journal of Economics, Finance and Administrative Science*.
- República de Colombia. (2010, 15 de julio). *Decreto 2555 de 2010: Por el cual se recogen y reexpiden las normas en materia del sector financiero, asegurador y del mercado de valores, y se dictan otras disposiciones*.

República de Colombia. (2023, 29 de marzo). *Decreto 455 de 2023: Por el cual se modifica el Decreto 2555 de 2010 en lo relacionado con el Sistema de Administración de Riesgo de Crédito (SARC) y se dictan otras disposiciones.*

República Rodríguez Avellaneda, A. L. (2018). *Modelo scoring para recuperar cartera de microcrédito*. Fundación Universitaria Los Libertadores.

Superintendencia Financiera de Colombia. (2020). *Circular básica contable y financiera (Capítulo XXXI: Sistema Integral de Administración de Riesgos – SIAR)*. Bogotá, Colombia

Tamara-Ayus, A. L., Trujillo-Hernández, K. P., & Mosquera-Peña, V. D. (2022). Implementación de un modelo score para el otorgamiento en una cartera de microcrédito. *Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA*, 23(1), 53–75.

Apéndices

Apéndice A. Nota explicativa sobre los títulos del cuadro de estadísticos descriptivos:

Vars: corresponde al índice asignado a cada variable dentro del conjunto de datos analizado.

n: indica el número total de observaciones válidas (sin valores perdidos) que fueron utilizadas para el cálculo de las estadísticas.

Mean (media): representa el promedio aritmético de la variable, es decir, el valor central alrededor del cual se distribuyen los datos.

Sd (standard deviation): desviación estándar, medida que refleja la variabilidad o dispersión de los datos respecto a la media.

Median (mediana): valor que divide la distribución en dos partes iguales. Es una medida robusta de tendencia central, especialmente útil en presencia de valores extremos.

Trimmed (media recortada): media calculada luego de excluir un porcentaje de los valores más extremos (generalmente el 10 %), reduciendo así la influencia de outliers.

Mad (median absolute deviation): desviación absoluta mediana, medida robusta de dispersión que, a diferencia de la desviación estándar, no se ve afectada de manera significativa por valores atípicos.

Min y max: valores mínimo y máximo observados para cada variable, respectivamente.

Range: rango total de la variable, calculado como la diferencia entre el valor máximo y el mínimo

(range = max – min).