

Concesión de créditos otorgados por la Corporación Financiera Nacional: un análisis exploratorio de datos EDA

Joseph Paúl Delgado Quijije
Universidad Laica Eloy Alfaro de Manabí, ULEAM
joseph.delgado@pg.uleam.edu.ec
Manta, Ecuador

Luis Cedeño-Valarezo
Escuela Superior Politécnica Agropecuaria
de Manabí Manuel Félix López
icedeno@espam.edu.ec
Calceta, Ecuador

DOI: <https://doi.org/10.56124/encriptar.v8i16.009>

Resumen

El presente estudio aborda la concesión de créditos otorgados por la Corporación Financiera Nacional (CFN) en el Ecuador, considerando su rol como entidad pública financiera orientada al fomento productivo. El objetivo fue analizar, mediante técnicas de análisis exploratorio de datos (EDA), el comportamiento de los créditos entregados entre 2022 y 2024. Se empleó un enfoque cuantitativo, descriptivo y de corte transversal, utilizando datos secundarios obtenidos de la plataforma de datos abiertos de la CFN. La base incluye variables cuantitativas (monto, número de operaciones) y categóricas (tipo de crédito, operación, estado de operación y provincia). El análisis se realizó con el lenguaje RStudio a través de paquetes del entorno tidyverse, se aplicaron procedimientos de limpieza, análisis univariado, bivariado y visualización.

Los resultados evidencian que la mayor proporción de los montos otorgados son inferiores a los USD 10.000, aunque existen valores atípicos significativos. También, se descubre la existencia de una relación inversa entre el monto y la frecuencia de operaciones, destacando que los créditos más altos se concedieron a pocos beneficiarios. Las provincias de Guayas y Pichincha concentran la mayor parte de esas operaciones, lo que evidencia un sesgo geográfico en su distribución. Por otro lado, el producto "crédito" y el estado "Original" fueron los más frecuentes, con una elevada variabilidad. Se concluye que la CFN gestiona recursos con un patrón en algunos productos y territorios, lo que plantea incógnitas en la equidad del acceso al financiamiento público y su alcance redistributivo.

Palabras clave: Concesión, Créditos CFN, EDA.

CREDIT GRANTING BY THE NATIONAL FINANCIAL CORPORATION: AN EXPLORATORY DATA ANALYSIS (EDA).

ABSTRACT

This study examined the credit allocation granted by the Corporación Financiera Nacional (CFN) in Ecuador, considering its role as a public financial institution aimed at fostering productive sectors. The objective was to analyze the behavior of granted loans between 2022 and 2024 through exploratory data analysis (EDA) techniques. A quantitative, descriptive, and cross-sectional approach was applied, using secondary data retrieved from CFN's open data platform. The dataset included quantitative variables (loan amount, number of operations) and categorical variables (type of credit, operation type, loan status, and province). The analysis was conducted using RStudio and tidyverse packages, including data cleaning, univariate and bivariate analysis, and visualization techniques. The results revealed that most of the loans granted were under USD 10,000, though notable outliers were present. An inverse relationship was identified between loan amounts and the number of operations, indicating that higher-value loans were granted to fewer beneficiaries. The provinces of Guayas and Pichincha concentrated the majority of credit operations, suggesting a geographic bias in credit distribution. Moreover, the "credit" product and "original" loan status were the most prevalent, both with high variability. It was concluded that CFN's financial resources were channeled through a concentrated pattern across specific products and regions, raising concerns about the equity and redistributive reach of public financing.

1. Introducción

El acceso al financiamiento productivo en el Ecuador está anclado en el desarrollo económico y en el tejido empresarial. En este contexto, la Corporación Financiera Nacional CFN, como banca de segundo piso, desempeña un rol importante en el sector bancario público con el objetivo de canalizar recursos hacia la economía local, a través de líneas de crédito a instituciones financieras privadas (Corporación Financiera Nacional CFN, 2025).

Cabe señalar que en los últimos años se ha generado interés desde la academia y por parte del gobierno central evaluar la efectividad de estos instrumentos financieros, si se toma en consideración el efecto postpandemia. No obstante, han sido insuficientes los estudios en los que se considere el análisis de datos con sus técnicas, como el análisis exploratorio de datos EDA por sus siglas en inglés, que sirve como argumento del presente estudio.

Wickham, Çetinkaya-Rundel y Grolemund (2019) consideran que no es un proceso formal, riguroso ni regido por un conjunto de reglas estrictas, es más que un estado mental, además mencionan que el EDA debe ser libre de investigar cualquier idea que se le ocurra, algunas de ellas pueden prosperar, en cambio, otras se encontrarán con un callejón sin salida. Finalmente, se considera como parte importante en cualquier análisis.

Los autores abordan un estudio que examina los patrones de transacción, los límites en los créditos y su categoría de comercio facilitando información sobre el comportamiento del consumidor, en ese sentido, se aplica el análisis exploratorio de datos a las tarjetas de crédito de un determinado banco, además las técnicas aplicadas al presente estudio se recogen en la estadística descriptiva y análisis de correlación. (Agarwal et al., 2024)

Romero J. y Cáceres A. (2024) en su investigación aplican métodos descriptivos con el objetivo de examinar los factores que afectan la decisión de los consumidores de manta de productos de aseo y limpieza, en ese mismo

orden de ideas, se concluye que el análisis EDA a partir de su exploración allana el camino de la modelación estadística.

Los autores Abdou y Pointon (2011) abordan el problema crediticio desde una perspectiva cuantitativa, haciendo el uso de técnicas estadísticas, ellos explican que el credit scoring es una herramienta importante y que, aunque no existe una técnica única e ideal, los métodos clásicos como los avanzados permiten recorrer caminos para mejorar la inclusión financiera, a través de una calificación eficiente.

El trabajo investigativo de Angilella y Mazzù (2013) contribuye con un diseño de modelos riesgos crediticios adaptables a los entornos PYMEs, en el afán de alcanzar la meta la evaluación de los datos se erige como el punto de partida en la construcción del modelo, permitiendo superar las limitaciones inherentes al mismo.

Bakoben et al. (2017) presentan en cambio un enfoque metodológico robusto con el fin de identificar el riesgo crediticio. Además, los autores declaran que la aplicación de técnicas estadísticas en la data real de tarjetas de crédito permite validar la metodología aplicada en contextos operativos auténticos.

Los autores Baesens et al. (2015) se refieren al credit scoring e introduce múltiples métricas de rendimiento como AUC, KS-statistic, entre otras y la aplicación de data dispersa para lograr los resultados. En esa dinámica, se comparan métodos tradicionales como la regresión logística con nuevos enfoques como boosting y bagging, generando una fuerte competición.

Bravo et al. (2020) su investigación representa un avance en el desarrollo de modelos de scoring, la novedad se basa en la combinación metodológica interdisciplinaria, ciencia de datos, teoría económica y teoría de redes de manera conjunta ayudan a resolver problemas sociales, siendo

consecuente, con el presente artículo la exploración de los datos es de suma importancia para la trilogía declarada por los autores.

Hooman et al. (2016) declaran que la utilización sistemática de técnicas estadísticas y de minería de datos permite identificar limitaciones y fortalezas con la comparación de técnicas relacionadas con la precisión, costo computacional, interpretabilidad y sensibilidad. Por lo tanto, contribuye a la evolución de la metodología y la evaluación del riesgo crediticio.

McGrath et al. (2018) exponen en su investigación desarrollar explicaciones interpretables para las decisiones de modelos de créditos, entonces para alcanzar el objetivo, se empieza con la revisión de las variables para identificar la presencia o no de correlación para evitar el problema de la multicolinealidad y esto ayuda a mejorar la estabilidad del modelo y su interpretabilidad. Entonces, el análisis EDA es importante para el acompañamiento en la investigación.

Pardo (2020) en su estudio demuestra a partir de una regresión logística permite identificar factores asociados al incumplimiento crediticio, previamente para la estructura del perfil crediticio se realiza un análisis exploratorio de datos EDA antes de su estimación. Adicional, se aplica una matriz de correlación con el propósito de verificar la presencia o ausencia de multicolinealidad. Finalmente, se evaluaron variables como edad, ingresos, género, tipo de empleo, entre otras.

Sariannidis et al. (2020) entre sus interrogantes plantean evaluar cómo varía la probabilidad de incumplimiento considerando factores demográficos y crediticios, en ese mismo orden y dirección, previo a la modelación se aplica métricas descriptivas y revisión de correlaciones parte del análisis exploratorio. Por otro lado, también ayuda a comprender la estructura de la data previo a la aplicación de modelos de aprendizaje.

Zambrano et al. (2024) en su artículo se hace referencia a la recolección de información para explorar las características y patrones del perfil del cliente,

en ese mismo sentido, se plantean herramientas estadísticas para alcanzar dicho fin. Por otro lado, la importancia de la exploración de la data resuelve en gran medida el objetivo de la investigación.

Delgado et al. (2024) analizan la cartera de crédito de las Cooperativas de Ahorro y Crédito del segmento 1 en Ecuador entre 2018 y 2022. Se analiza la morosidad de información recogida de fuente oficial de la Superintendencia de Economía Popular y Solidaria y mediante técnicas exploratorias descriptivas, los autores concluyen que las cooperativas demostraron ser más resistentes frente a la crisis económica en paralelo con otros sectores.

2. Metodología (Materiales y métodos)

El presente artículo se desarrolla en el contexto del análisis de concesión de créditos otorgados por la Corporación Financiera Nacional CFN en el Ecuador, cuyo rol principal es la canalización de recursos financieros productivos. El estudio aborda un periodo entre 2022 y 2024 con la intención de analizar por medio de técnicas de EDA, como se comportan los créditos y su distribución, en relación con las diversas variables categóricas y cuantitativas.

La base de datos fue obtenida de los datos abiertos de la CFN que incluye variables monto otorgado (valor de las operaciones agrupadas en miles de dólares), tipo de crédito (producto o línea de crédito otorgada), tipo de operación (clasificación por operaciones de crédito y operaciones contingentes), número de operaciones, provincia, sector (agrupación por sectores económicos), entre otras, y está estructurada en csv.

Para el procesamiento y análisis de la base de datos, se utiliza el programa estadístico R junto con el entorno de desarrollo RStudio. Además, se aplican principalmente las paqueterías tidyverse, ggplot2, dplyr, skimr, readr

y corrplot con la intención primero de limpiar la data y, posteriormente, generar gráficos, realizar un análisis estadístico descriptivo y la observación de patrones.

Con respecto al diseño metodológico, este es un enfoque cuantitativo, no experimental, tipo descriptivo y de corte transversal, así pues, se encuadra en el paradigma empírico-analítico, teniendo en cuenta que, se parte desde un análisis sistemático de data real observacional. En ese mismo orden, el EDA fue considerado como la técnica principal en la identificación de patrones, relaciones, distribuciones, datos atípicos, estructuras dentro de la base de datos.

El proceso analítico contempló cuatro etapas: (1) limpieza y verificación de datos (detección de valores faltantes y errores), (2) análisis univariado de variables, (3) análisis bivariado entre variables cuantitativas y categóricas, y (4) visualizaciones multivariadas y agrupamientos iniciales. Este abordaje permitió obtener una comprensión preliminar del comportamiento de las operaciones crediticias de la CFN, en vista de orientar investigaciones posteriores que requieran la formulación de modelos estadísticos inferenciales.

En el proceso de análisis se considera cuatro etapas:

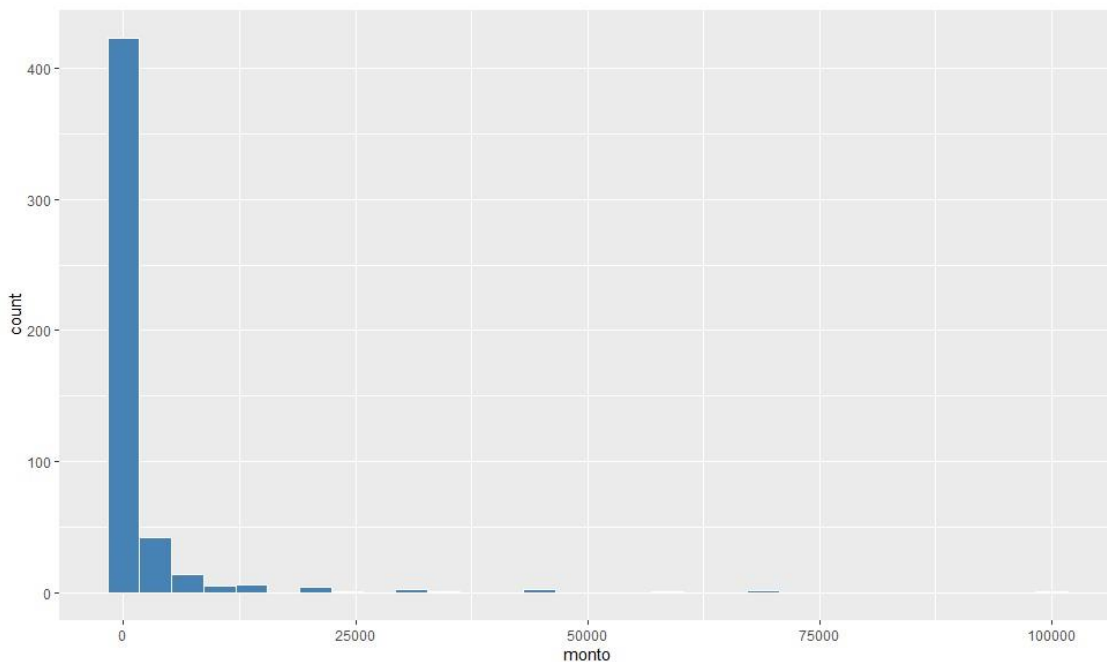
- 1.Limpieza y comprobación de valores faltantes (NA's) y errores
- 2.Análisis univariado de las variables consideradas en la base de datos
- 3.Análisis bivariado entre las variables cuantitativas y cualitativas
- 4.Visualización de los datos analizados

En el mismo orden de la idea anterior, este abordaje permite lograr una comprensión inicial con respecto al comportamiento de las operaciones de créditos otorgadas por la CFN, con el objetivo de orientar futuras investigaciones que puedan requerir a posteriori formulación de modelos estadísticos de carácter inferencial.

3. Resultados (análisis e interpretación de los resultados)

El análisis que corresponde a la variable monto evidencia una leve asimetría positiva, como se observa en la Figura 1, la mayor parte de los créditos otorgados se sitúan por debajo de los USD 10.000, sin embargo, un número minoritario representan valores elevados, por lo tanto, dicho fenómeno genera una distribución con sesgo o asimetría positiva. En ese mismo orden de ideas, dicha concentración indica una estructura de financiamiento donde los créditos de menor valor predominan, caso contrario sucede con los montos altos, estos en cambio distorsionan las medidas de tendencia central.

Figura 1. Distribución del monto de los créditos.

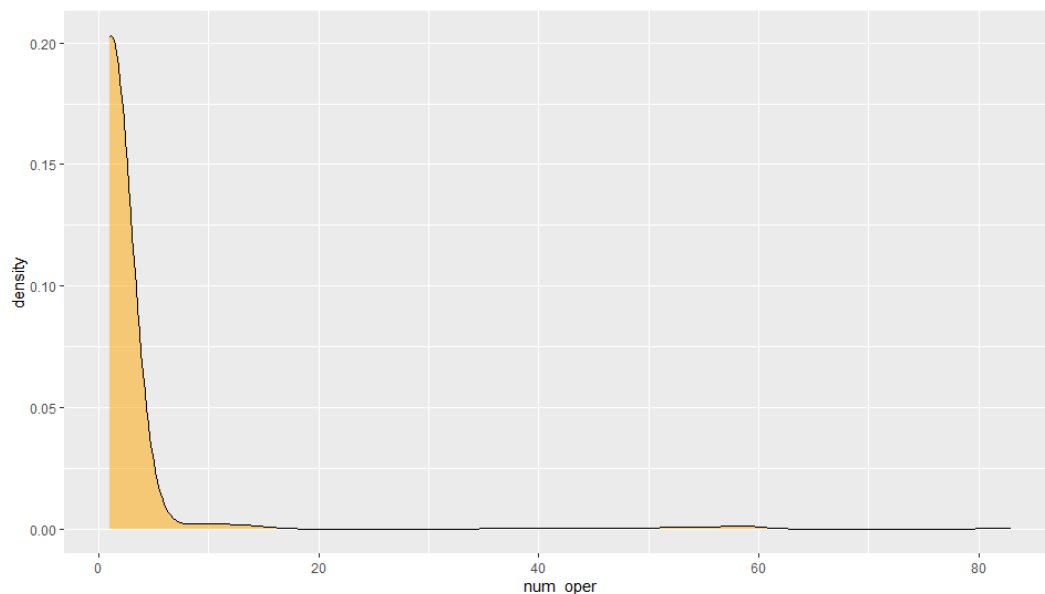


En cambio, como se observa en la Figura 2 la variable número de operaciones, su densidad indica que la mayor parte de los registros corresponde a pocas operaciones, es decir, la minoría con una caída de manera exponencial en correspondencia con el incremento en el número de

operaciones. En cambio, se identifican outliers en los rangos que muestran valores entre 40 y 80 operaciones, lo que sugiere que los beneficiarios accedieron a múltiples desembolsos, esto ocurre también por el sector, el subsector y el tipo de actividad económica.

Desde otra perspectiva, más crítica los patrones observados respaldan la hipótesis de la concentración de los créditos en ciertos sectores de la economía. Aunque pareciera que el acceso a los créditos está al alcance de los demandantes, no obstante, el peso financiero de algunos créditos grandes podría condicionar el efecto redistributivo esperado. En ese mismo orden y dirección, coincide Mazzucato (2019) que propone que el valor económico no ostenta una dependencia única de la cantidad de operaciones, más bien de la asignación de los recursos públicos.

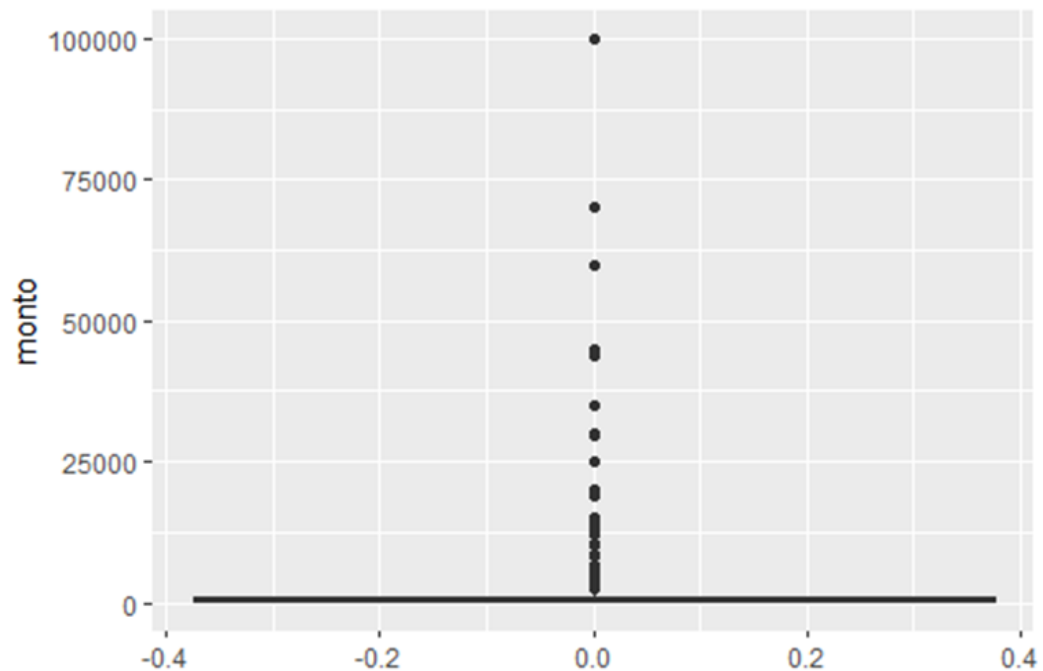
Figura 2. Densidad del número de operaciones.



El gráfico de Box Plot para la variable monto otorgado refuerza la

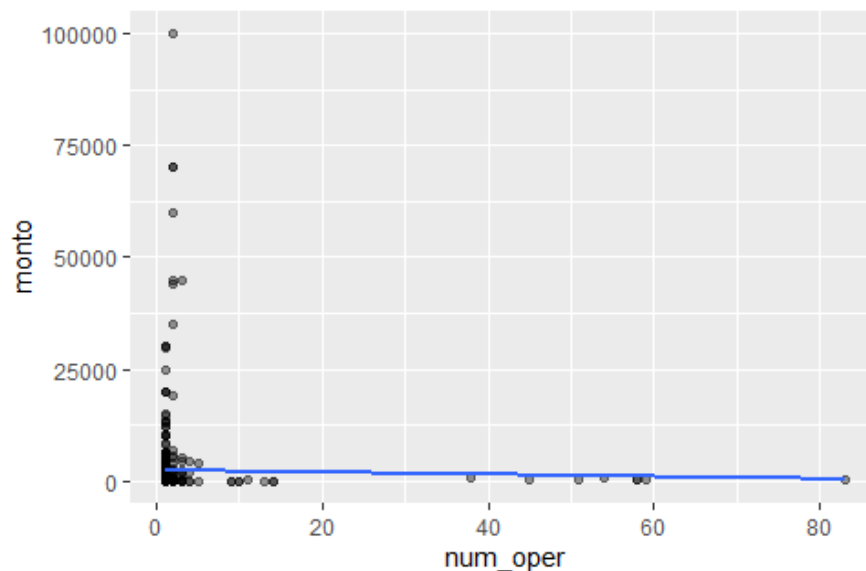
presencia de múltiples outliers o valores atípicos. En ese mismo sentido, la mayor parte de los montos se concentran en la clase inferior, mientras tanto un grupo minoritario de operaciones supera los valores centrales, alcanzando así hasta los USD 100.000. Esto deja al descubierto la necesidad de analizar los créditos desde una perspectiva diferente, es decir, que no sea el limitante de los promedios, caso contrario, que se contemple la variabilidad y la concentración de los recursos.

Figura 3. Boxplot del monto Total del Crédito.



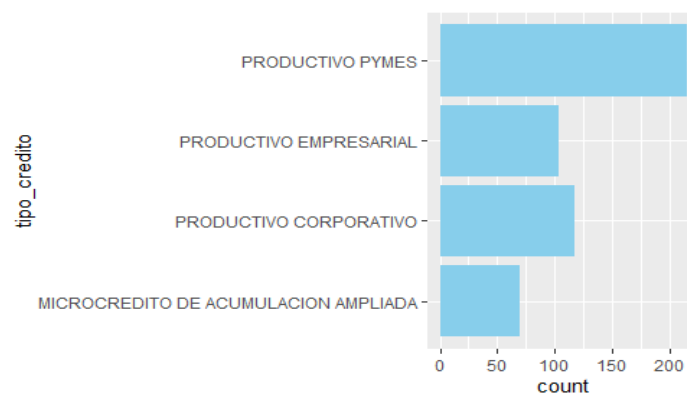
Al contrastar entre el monto y el número de operaciones, se observa que los valores más altos de crédito tienden a concentrarse en aquellos deudores con operaciones limitadas. En cambio, aquellos que realizaron múltiples operaciones acceden a montos más bajos. Entonces se evidencia una relación inversa, esto se puede deber a políticas internas de asignación o a diferencias en el perfil de riesgo entre sujetos institucionales y los empresarios.

Figura 4. Dispersión entre monto y número de operaciones



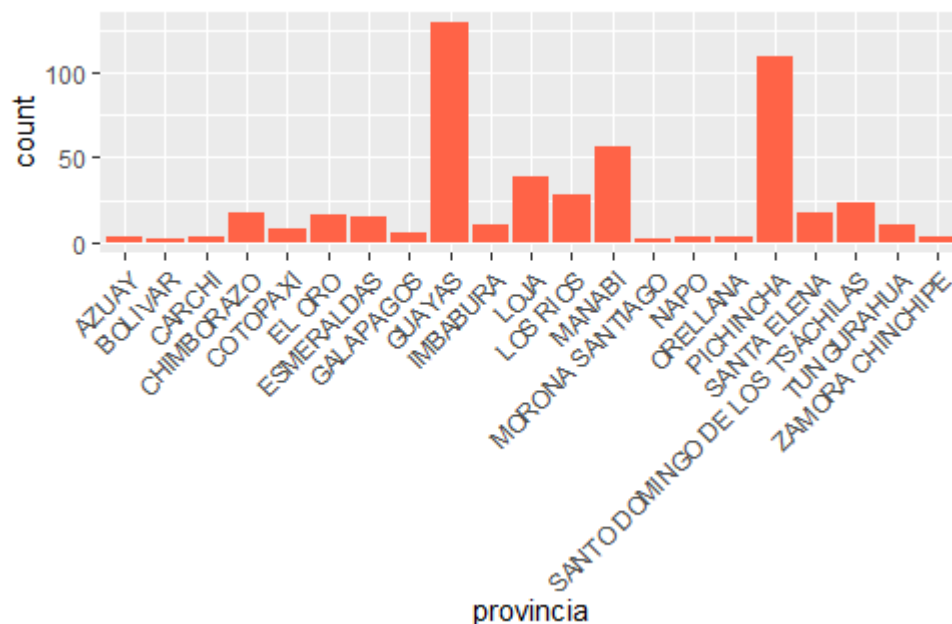
Como se puede observar en la Figura 5 se muestra que el tipo de crédito con mayor frecuencia corresponde a 'productivo pymes', el mismo evidencia una tendencia hacia el financiamiento pequeñas y medianas empresas. Caso contrario, se evidencia con 'microcrédito de acumulación ampliada' que representa una menor cantidad de operaciones en contrastación con el resto de tipos de créditos, en ese orden, se abren posibilidades de generar estrategias con la finalidad de estimular este tipo de crédito.

Figura 5. Gráfico de barras de tipo de crédito



En cuanto a la forma de la distribución geográfica (Figura 6), en las provincias de Guayas y Pichincha se concentran la mayor cantidad de registros de préstamos, en comparación con el resto de las provincias hay una diferencia muy amplia. En ese mismo orden y dirección, esto se debe a que estas provincias ostentan una mayor representación y actividad económica, por ejemplo, con Bolívar, Morona Santiago, entre otras, que representan préstamos más bajos lo que indica una penetración débil en el sistema financiero o una demanda limitada de productos crediticios en dichas zonas.

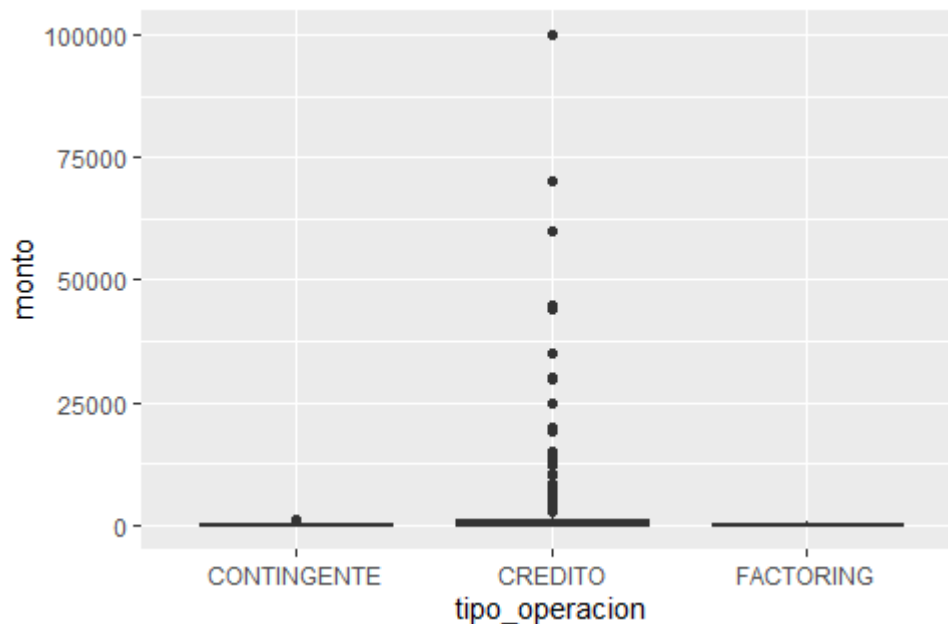
Figura 6. Gráfico de barras de tipo de crédito por provincia



El análisis comparativo (Figura 7) del monto otorgado según el tipo de operación confirma que el producto "crédito" es el principal medio para el financiamiento. En el análisis de los datos se muestra que los montos grandes y dispersos se relacionan con el tipo de operación, en ese mismo orden, es que se sugiere que la institución financiera canaliza sus recursos en una mayor proporción a partir de los créditos directos, mientras tanto las operaciones

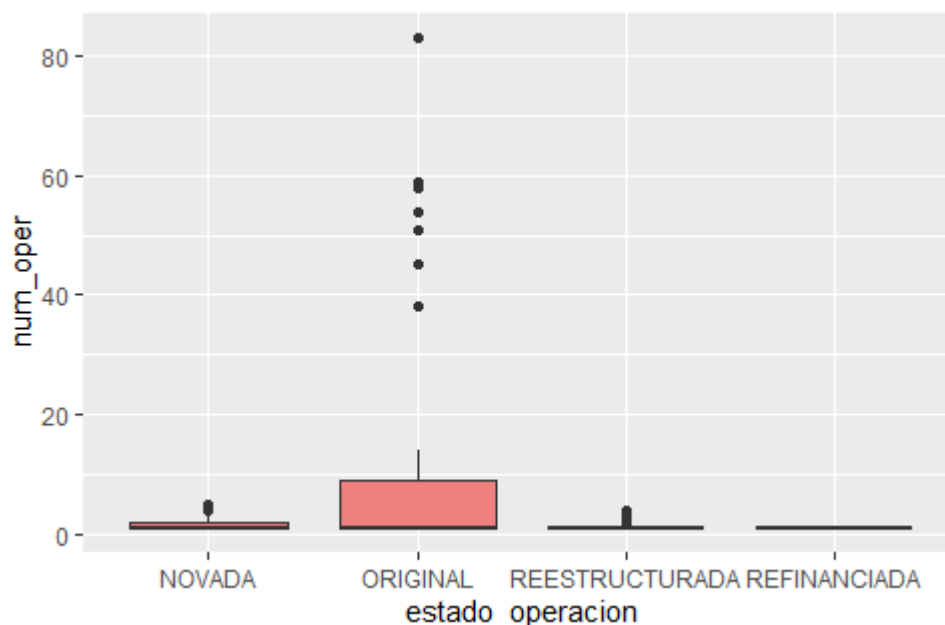
contingentes y factoring operan montos menores y con poca variabilidad.

Figura 7. Boxplot del monto según tipo de operación (comparativo)



Respecto al estado del número de operaciones, la Figura 8 evidencia que la mayoría de las operaciones se encuentran en estado original, a diferencia de las demás categorías.” Esta categoría presenta una gran variabilidad respecto al número de operaciones, con valores atípicos por encima de 80. Caso contrario, sucede con las operaciones novadas, reestructuradas y refinanciadas, estas tienden a presentar un menor registro con valores muy similares, lo cual puede presentar procesos con mayor control. Dicho patrón puede ser de gran utilidad para mitigar el riesgo operativo.

Figura 8. Boxplot del número de operaciones según estado de operación.



4. Conclusiones

El EDA de los créditos de la CFN (2022–2024) confirma una estructura asimétrica: muy alta frecuencia de operaciones de bajo monto junto con una concentración de montos elevados en pocas operaciones, y una relación inversa entre monto y número de operaciones; este patrón, sumado a la concentración territorial, tensiona los objetivos de equidad e inclusión y orienta acciones correctivas y de focalización.

En términos territoriales, Guayas y Pichincha lideran los registros frente a provincias de menor dinamismo, lo que exige metas explícitas de desconcentración y una estrategia de expansión inteligente (brigadas, ventanillas móviles y alianzas con GAD, MIPYMES y cooperativas de ahorro y crédito) priorizando territorios rezagados.

En la mezcla de productos, el liderazgo de Productivo PYMES y la baja participación del Microcrédito de acumulación ampliada sugieren impulsar instrumentos de microfinanzas con metas de cobertura por provincia, cuotas para mujeres, jóvenes y ruralidad, líneas con garantías parciales y documentación simplificada para reducir fricciones de acceso.

Operativamente, el predominio del estado “original” con alta variabilidad y outliers demanda tableros de alertas tempranas, revisión de reglas de originación y un protocolo de tratamiento de valores extremos (winsorización, reglas de negocio y validaciones) para disminuir riesgo operativo y sesgos en el seguimiento.

Para monitorear la equidad de la colocación, se recomienda institucionalizar métricas de concentración: HHI y Gini por provincia y por producto, participaciones Top-k y objetivos anuales de reducción (p. ej., bajar el HHI provincial en 10% y elevar la participación del microcrédito en 3–5 p.p. anuales), complementadas con indicadores per cápita (monto y operaciones por 10.000 habitantes) y por PIB provincial.

La CFN podría pilotear esquemas de originación digital con pre-evaluaciones ligeras y scoring para microcrédito, fondos de garantía para reducir requerimientos de colateral y asistencia técnica a PYMES en gestión financiera, encadenamientos productivos y formalización, evaluando impactos con diseños cuasi-experimentales y pruebas A/B. Estas acciones deben ir acompañadas de un marco de reproducibilidad: publicar URL y fecha de descarga del dataset, diccionario de variables, versión/licencia y un anexo con el código en R mínimo para replicar el flujo (lectura, limpieza, cálculo de métricas y gráficos); actualizar el EDA trimestralmente para realimentar metas y decisiones.

Finalmente, dada la naturaleza exploratoria del estudio, los próximos pasos deben incluir modelos de conteo para el número de operaciones (con

offset poblacional), regresiones cuantílicas para montos, clusterización territorial para tipologías de provincias y paneles de seguimiento operativo que vinculen variabilidad, outliers y desempeño con acciones correctivas; todo ello consolidará una política crediticia pública más inclusiva, territorialmente balanceada y basada en evidencia.

5. Referencias

- Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management*, 18(1–2), 59–88.
- Aldaz Delgado, N. L., & Yaguache Maza, D. M. (2024). *Análisis de la cartera de crédito de las cooperativas de ahorro y crédito del Ecuador del segmento 1 (2018-2022)*. Revista RECIAMUC, 8(1), 156–165. [https://doi.org/10.26820/reciamuc/8.\(1\).ene.2024.156-165](https://doi.org/10.26820/reciamuc/8.(1).ene.2024.156-165)
- Agarwal, A., Prabha, S., & Yadav, R. (2024). Exploratory Data Analysis for Banking and Finance: Unveiling Insights and Patterns. *arXiv.Org*, abs/2407.11976. <https://doi.org/10.48550/arxiv.2407.11976>
- Angilella, S., & Mazzù, S. (2013). The financing of innovative SMEs: A multicriteria credit rating model. *International Journal of Production Economics*, 146(2), 281–290.
- Bakoben, M., Bellotti, T., & Adams, N. M. (2017). Identification of Credit Risk Based on Cluster Analysis of Account Behaviours. *arXiv: Statistical Finance*. <https://doi.org/10.1080/01605682.2019.1582586>
- Bravo, C., Óskarsdóttir, M., & Sarraute, C. (2020). The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. *arXiv*. <https://doi.org/10.48550/arXiv.2002.09931>
- Corporación Financiera Nacional CFN. (2025). *segundo piso*. cfn.fin.ec: <https://www.cfn.fin.ec/servicios/segundo-piso/>
- Lessmann, S., Baesens, B., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124–136. <https://doi.org/10.1016/J.EJOR.2015.05.030>
- Hooman, A., Marthandan, G., Yusoff, W. F. W., Omid, M., & Karamizadeh, S. (2016). Statistical and data mining methods in credit scoring. *Journal of Developing Areas*, 50(5), 371–381.
- McGrath, R., Costabello, L., Le Van, C., Sweeney, P., Kamiab, F., Shen, Z., & Lecue, F. (2018). Interpretable credit application predictions with counterfactual explanations. *arXiv*. <https://doi.org/10.48550/arXiv.1811.05245>
- Pardo, M. A. (2020). Modelización del riesgo crediticio en una cooperativa de ahorro y crédito mediante regresión logística. *Revista Colombiana de Estadística*. <https://doi.org/10.12804/revistas.urosario.edu.co/empresa/a.8266>
- Romero González J. y Cáceres Larreátegui A. Influencia de redes sociales sobre la decisión de compra de productos de aseo y limpieza en Manta-Manabí 2024. RefCalE <https://doi.org/10.56124/refcale.v13i1.009>

- Sariannidis, N., Papadakis, S., Garefalakis, A., Lemonakis, C., & Kyriaki-Argyro, T. (2020). *Default avoidance on credit card portfolios using accounting, demographical and exploratory factors: Decision making based on machine learning (ML) techniques*. *Annals of Operations Research*, 294, 715–739. <https://doi.org/10.1007/s10479-019-03188-0>
- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2019). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly.
- Zambrano Vera, M. F., Bermúdez Alvarado, A. L., & Cáceres Larreátegui, A. L. (2024). Caracterización del perfil del cliente para propuesta de modelo de negocio e-commerce. *Revista Electrónica Formación y Calidad Educativa (REFCaE)*, 12(2), 84-89. <https://doi.org/10.56124/refcale.v12i2.005>