



DEPARTAMENTO DE POSGRADOS

**Regresión cuantílica para la detección de transacciones
anómalas en canales electrónicos: estudio de caso en una
cooperativa de ahorro y crédito ecuatoriana.**

Magíster en Estadística Aplicada.

Autor: Omar Antonio Pinos Guillén.

Director: Iván Mendoza Vázquez.

Cuenca, Ecuador 2025

DEDICATORIA

A mi querida esposa, por su amor, paciencia y apoyo incondicional a lo largo de este proceso académico. Su compañía y comprensión fueron esenciales para alcanzar esta meta.

A mis hijos **Josué, David, Sarita y Ana María** quienes desde el cielo interceden cada día ante Dios por mí; y a mis hijos **Antonio, Gabriel y José Miguel** que tengo junto a mi lado, cuyo cariño, alegría e inspiración me impulsaron a continuar y dar siempre lo mejor de mí.

Este trabajo representa no solo un logro personal, sino también un homenaje a su amor, a su presencia – terrenal y celestial – y al apoyo que me sostuvo en cada etapa de este camino.

AGRADECIMIENTO

Expreso mi más sincero agradecimiento a la **Universidad del Azuay**, institución que me brindó la oportunidad de formarme profesionalmente y desarrollar las capacidades necesarias para la realización de este trabajo de titulación.

Mi especial gratitud al **Ing. Iván Andrés Mendoza Vázquez**, director de esta tesis, por su valiosa orientación, asesoramiento técnico e intelectual, así como por su compromiso y acompañamiento constante durante el desarrollo de este trabajo integrador.

Extiendo también mi reconocimiento a todos los **docentes de la Universidad del Azuay**, quienes, a través de cada módulo impartido, compartieron con generosidad sus conocimientos y experiencias, contribuyendo de manera significativa a mi crecimiento académico y profesional.

A todos ellos, mi más profundo agradecimiento por ser parte esencial de este logro.

RESUMEN

Este estudio aborda la detección de transacciones anómalas en canales electrónicos de una cooperativa de ahorro y crédito mediante regresión cuantílica. El modelo se estimó en el cuantil extremo ($\tau = 0,97$) y alcanzó pseudo- R^2 de Koenker & Machado (1999) igual a 0,434, lo que evidencia capacidad explicativa moderada-alta y la identificación de umbrales efectivos para separar operaciones anómalas y normales. La validez se contrastó con algoritmos supervisados, tras balancear clases y dividir la base 70/30 (entrenamiento/prueba). Entre ellos, Random Forest obtuvo el mejor desempeño promedio por canal: F1-score 71%, AUC-ROC 0,70 y accuracy 77%. Los resultados indican que la estrategia es replicable en el sector financiero y útil para el monitoreo operativo y la gestión de riesgos en canales electrónicos.

PALABRAS CLAVES

Regresión cuantílica, Transacciones anómalas, Canales electrónicos, Monto transaccional, Random Forest, W-Part, K - means, K-NN.

ABSTRACT

This study addresses the detection of anomalous transactions in the electronic channels of a savings and credit cooperative using quantile regression. The model was estimated at the extreme quantile ($\tau = 0,97$) and achieved a Koenker & Machado (1999) pseudo- R^2 of $= 0.434$, which demonstrates a moderate-to-high explanatory capacity and effective thresholds for separating anomalous and normal operations.

Model validity was contrasted with supervised algorithms after balancing classes and splitting the dataset 70/30 (training/testing). Among these, Random Forest obtained the best average performance per channel: F1-score of 71%, AUC-ROC of 0.70, and an accuracy of 77%.

The results indicate that this strategy is replicable in the financial sector and useful for operational monitoring and risk management in electronic channels.

KEYWORDS

Quantile Regression, Anomalous Transactions, Electronic Channels, Transactional Amount, Random Forest, W-Part, K-means, K-NN.

ÍNDICE DE CONTENIDO

DEDICATORIA.....	i
AGRADECIMIENTO	ii
Resumen	iii
ABSTRACT	iv
INDICE DE FIGURAS, TABLAS Y ANEXOS.....	vi
1. Introducción.....	1
1.2. Descripción del problema.....	3
1.3. Alcance.	3
1.4. Estructura general del documento.	5
1.4.1. Laboratorio: Enfoque descriptivo inferencial.....	5
1.4.2. Laboratorio: Enfoque Machine Learning.....	5
1.4.3. Laboratorio: Enfoque Toma de Decisiones	6
1.4.4. Laboratorio: Enfoque Robustez y Validación.....	6
2. Objetivo General	6
3. Laboratorio: Enfoque Descriptivo e Inferencial	6
3.1. Objetivos específicos	6
3.2. Método	7
3.2.1. Preparación y análisis exploratorio de datos	7
3.2.2. Modelado del monto transaccional mediante (modelo lineal de regresión múltiple) .	7
3.2.3. Modelado del monto transaccional mediante (regresión cuantílica)	8
3.2.4. Detección de anomalías por umbral estadístico.....	9
3.3. Discusión y resultados	10
3.3.1. Análisis estadístico descriptivo del perfil transaccional de los clientes	10
3.3.2. Análisis estadístico descriptivo del perfil sociodemográfico y económico ...	14
3.3.3. Análisis estadístico inferencial y prueba de hipótesis.....	18
3.3.3.1. Pruebas de normalidad	18
3.3.3.2. Prueba de Wilcoxon, y Kruskal-Wallis para determinar diferencia de medianas entre grupos	20
3.3.3.3. Análisis post-hoc mediante test de Tukey con corrección de Bonferroni	21
3.3.3.4. Matriz de correlación de Spearman	22
3.3.3.5. Regresión lineal múltiple	23

3.3.3.6.	Regresión cuantílica ($\tau = 50$) para estimar el monto transaccional	25
3.3.3.7.	Regresión cuantílica para $\tau = 0,97$: estimación del umbral superior del monto transaccional	28
	Análisis de residuos del modelo de regresión cuantil $\tau = 0,97$	31
3.3.3.8.	Estadísticas generales de las transacciones anómalas que superan el umbral superior del modelo de regresión cuantil ($\tau = 0,97$)	32
3.4.	Discusión.....	33
3.5.	Conclusiones del laboratorio.....	33
4.	Laboratorio: Enfoque Machine Learning.....	35
4.1	Objetivos específicos	35
4.2	Método	35
4.3	Discusión y resultados	35
4.4	Conclusiones del laboratorio.....	41
5.	Laboratorio: Enfoque Toma de Decisiones	42
5.1.	Objetivos específicos:	42
5.2.	Métodos:	43
5.3.	Discusión y resultados	43
5.4.	Conclusiones del laboratorio.....	48
6.	Laboratorio: Enfoque Robustez y validación	49
6.1.	Objetivos específicos:	49
6.2.	Método	49
6.3.	Discusión y resultados	50
6.4.	Conclusiones del laboratorio.....	53
7.	Conclusiones.....	54
8.	Bibliografía.	56
	Anexos	57

INDICE DE FIGURAS, TABLAS Y ANEXOS

Tabla 1.	Lista de variables consideradas para el modelo.....	4
Tabla 2.	Estadísticas descriptivas del monto transaccional por tipo de canal.....	11
Tabla 3.	Número de transacciones máximas realizadas por día y por canal	13
Tabla 4.	Tiempo en horas promedio entre transacciones por canal	14
Tabla 5.	Estadísticas de tendencia central del monto transaccional por tipo de sexo	15

Tabla 6. Estadísticas del monto transaccional por nivel de instrucción	16
Tabla 7. Monto transaccional por edad edades generacionales	18
Tabla 8. Análisis <i>post-hoc</i> Tukey HSD entre niveles de estudio	22
Tabla 9. Resultados modelo de regresión lineal múltiple	24
Tabla 10. Regresión cuantílica ($\tau=0,50$): determinantes del percentil 50 del monto	26
Tabla 11. Resultados del modelo de regresión cuantílica ($\tau = 0,97$)	29
Tabla 12. Número de transacciones anómalas por tipo de canal.....	32
Tabla 13. Comparación de valores de pseudo- R^2 reportados en estudios del sector bancario mediante regresión cuantílica	33
Tabla 14. Resumen de desempeño del modelo Random Forest por canal virtual.....	36
Tabla 15. Resumen de desempeño del modelo W PART por canal virtual	37
Tabla 16. Reglas de clasificación para canal Web Transaccional por tipo de clase	38
Tabla 17. Reglas de clasificación para canal APP Móvil por tipo de clase	39
Tabla 18. Perfilamiento transaccional en canal Web Transaccional	44
Tabla 19. Perfilamiento transaccional en canal APP Móvil	44
Tabla 20. Perfilamiento transaccional en canal ATMS	45
Tabla 21. Perfilamiento transaccional en canal Corresponsales No Bancarios	45
Tabla 22. Perfilamiento transaccional en canal APP MÓVIL PYMES	46
Tabla 23. Perfilamiento sociodemográfico clúster anómalo por canal electrónico	46
Tabla 24. Validación de la técnica K Medias	47
Tabla 25. Métricas de desempeño del modelo KNN, clase positiva anómalos	48
Tabla 26. Métricas de validación de los dos modelos de regresión cuantílica	50
Tabla 27. Métricas de validación de los modelos de regresión cuantílica	51
Tabla 28. Métricas de validación de los modelos supervisados y no supervisados	53
Tabla 29. Métricas de validación modelos K medias y KNN	53
Tabla 30. Límites reglamentarios definidos por canal electrónico	57
Tabla 31. Conceptos transaccionales incluidos en el análisis, según canal electrónico	57
Figura 1. Evolución de las transacciones físicas y electrónicas	1
Figura 2. Umbral máximo calculado para detección de transacciones anómalas	10
Figura 3. Histograma con curva de normalidad de la densidad y monto transaccional	11
Figura 4. Histograma de las transacciones atípicas por canal transaccional	12
Figura 5. Horario habitual transaccional	13
Figura 6. Monto promedio transaccional por tipo de sexo y canal	15
Figura 7. Monto promedio transaccional por rango de ingreso del cliente	16
Figura 8. Monto promedio transaccional por actividad económica del cliente	17
Figura 9. Gráfico QQ variable del monto transaccional	19
Figura 10. Matriz de correlación de Spearman	23
Figura 11. Gráfico de dispersión de los residuos modelo de regresión cuantil $\tau = 0,97$	32
Figura 12. Gráficos de diagnóstico de validación de los modelos de regresión cuantílica ...	52

Figura 13. Árbol de decisión Rpart canal Botón De Pagos	58
Figura 14. Árbol de decisión Rpart canal Corresponsales no bancarios	58
Figura 15. Árbol de decisión Rpart canal ATMS.....	59
Figura 16. Árbol de decisión Rpart canal APP Móvil Pymes	59
Figura 17. Proyección PCA de los clústeres generados por el algoritmo K-medias por canal	60

1. Introducción

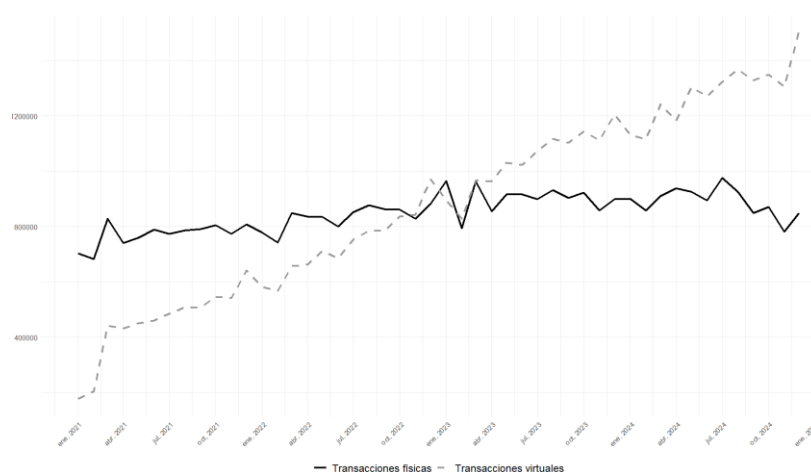
La innovación y el desarrollo de la banca digital se han consolidado como factores estratégicos esenciales para las instituciones financieras, al facilitar la inclusión, optimizar la eficiencia operativa y fomentar la fidelización de los clientes.

Asimismo, potencian los servicios de intermediación financiera —tanto en la captación de depósitos como en la colocación de créditos— generando impactos positivos en aspectos intangibles como la cercanía con el cliente, y proyectan una buena imagen reputacional de la institución ante la sociedad.

De otro lado, la rápida digitalización de los servicios financieros ha transformado la forma en que las personas operan su dinero. En Ecuador, el peso de los canales electrónicos sobre el total de operaciones ha crecido de manera sostenida desde la pandemia, con implicaciones directas para la gestión del riesgo operativo y de fraude en las entidades del sector financiero formal.

A junio de 2024, los canales electrónicos concentran el 62% del total de transacciones de la Cooperativa de Ahorro y Crédito analizada, con $\approx 1,27$ millones operaciones mensuales, mientras que las transacciones presenciales realizadas en oficina se mantienen relativamente estables. La Figura 1 muestra la evolución mensual desde marzo de 2021 donde se aprecia una clara transición hacia el uso de los canales electrónicos.

Figura 1. Evolución de las transacciones físicas y electrónicas



Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

La institución mediante la innovación ha desarrollado a la fecha un total de seis canales electrónicos (Web transaccional, App móvil, App móvil PYMES, Botón de pagos, Cajeros automáticos y Corresponsales no bancarios) en el cual mantiene controles como bloqueos automáticos por IP/país de alto riesgo y límites de monto transaccional máximo diario por canal

Para el abordaje de este trabajo de titulación se depuraron las operaciones de egreso (retiros y transferencias salientes) —por su impacto directo en el saldo— y se consideraron únicamente clientes con al menos dos transacciones realizadas en el mes de estudio, a fin de asegurar densidad mínima de evidencia. El conjunto final integra 128.074 clientes únicos que realizaron transacciones en estos canales, quienes describen el perfil sociodemográfico y económico de los clientes (edad, género, ingresos, actividad económica, nivel de educación) y del perfil transaccional (canal, horario, monto acumulado, monto última transacción (Y_{t-1}), media, mediana, moda, mínimo, máximo del monto transaccional histórico, saldo en ahorros vista, intensidad), permitiendo una caracterización diferenciada por canal.

En un estudio anterior realizado por la institución se aplicó un modelo de regresión lineal múltiple para estimar el monto transaccional promedio en el cual se obtuvo un coeficiente de determinación R^2 cercano al 70%; sin embargo, los diagnósticos del modelo evidenciaron violaciones sustantivas a supuestos clásicos (no normalidad de residuos, heterocedasticidad, no linealidad), lo que invalida la inferencia y conduce a umbrales de alerta imprecisos cuando se usan reglas de tipo “media $\pm k \cdot \sigma$ ” en distribuciones con colas pesadas y valores atípicos, con lo cual no fue replicable el modelo. Estos hallazgos fueron corroborados también en un primer abordaje de análisis estadístico descriptivo e inferencial desarrollado en el laboratorio 1 del presente trabajo de titulación como etapa inicial de diagnóstico obteniendo el mismo resultado.

En ese sentido, y acudiendo a la literatura reciente se evidencia que el aprendizaje automático cumple un papel protagónico en la detección de anomalías y fraude transaccional. No obstante, su implementación enfrenta retos operativos bien documentados, como (i) desbalance de clases (la cantidad de anómalos son muy escasos frente a transacciones normales), (ii) el concept drift (los patrones de uso y de fraude cambian en el tiempo) y (iii) los costos de falsos positivos. Las revisiones sistemáticas y marcos de referencia proponen combinar enfoques supervisados y no supervisados, así como integrar gobernanza con métricas de negocio, según Hernández et al. (2024).

En paralelo, han ganado tracción los métodos contextuales que comparan una observación con su entorno de referencia (usuarios “similares”). En esta línea, la regresión cuantílica —y en particular los Quantile Regression Forests— son modelos robustos usados para estimar cuantiles condicionales y construir umbrales personalizados de normalidad, con ventajas de

interpretabilidad y robustez frente a colas pesadas y asimetrías típicas de montos y frecuencias en el contexto del sector bancario. Estudios recientes como el desarrollado por Zhong & Matthijs (2023) también exploran técnicas cuantílicas para manejo de outliers y detección de anomalías en series y datos financieros. Además. Desde la base teórica, Koenker & Bassett (1978) demostraron que la regresión cuantílica no requiere supuestos de normalidad ni varianza constante, lo que la hace especialmente robusta frente a la presencia de valores atípicos, heterocedasticidad y distribuciones asimétricas, el cual es replicable para estudios en el sector bancario. En síntesis, la regresión cuantílica y sus extensiones de bosque ofrecen una vía interpretable, robusta y personalizable para definir bandas de normalidad y umbrales de alerta en la detección de comportamientos transaccionales anómalos en canales electrónicos que, en combinación con técnicas de aprendizaje supervisado y no supervisado, enriquecen la evidencia empírica sobre el comportamiento atípico.

1.2. Descripción del problema.

¿Puede un esquema de umbrales cuantílicos personalizados reducir los falsos positivos y mejorar la detección temprana de transacciones anómalas en los canales electrónicos de una COAC del segmento 1, manteniendo a la vez explicabilidad y estabilidad operativa?

Para responder a esta pregunta, se estiman cuantiles condicionales (anómalos y normales) de variables de uso por canal —monto, frecuencia, intensidad y horario— en función del perfil del cliente y del propio canal. Una transacción se clasifica como anómala cuando excede el cuantil 97, es decir el percentil 97 de la banda de normalidad esperada para dicho perfil, generando un puntaje de anomalía cuantílico que habilita la priorización operacional de alertas, en contraste si es menor a este cuantil se considera como transacción normal.

El desempeño del enfoque se contrasta con líneas base de aprendizaje automático:

- Supervisadas: Random Forest (entrenado con la etiqueta binaria *anómala/normal* generada por el modelo cuantílico), R-part, W-PART y k-NN (clasificación).
- No supervisadas: k-means (segmentación).

La comparación se realiza con métricas orientadas al negocio (reducción de falsos positivos y detección oportuna) y con criterios de mantenibilidad (estabilidad temporal e interpretabilidad).

1.3. Alcance.

El estudio se circunscribe a personas naturales, a operaciones de egreso y a los seis canales electrónicos mencionados. Se excluyen transacciones sin impacto neto en el saldo de depósitos, recaudaciones y ajustes, así como a los clientes con actividad transaccional

esporádica¹ con el fin de asegurar pertinencia operativa y estabilidad estadística del modelo, la base de datos es tomada de la propia institución financiera con corte al mes de junio del 2024

La base de datos está compuesta por un total de 13 variables numéricas y 5 variables categóricas que permiten describir tanto el perfil sociodemográfico y económico como el comportamiento transaccional de los clientes la cual además fue anonimizada. A continuación, la siguiente Tabla 1 detalla dichas variables.

Tabla 1. Lista de variables consideradas para el modelo

Variables	Tipo	Observación
Servicio	Nominal	Canal utilizado (Web, APP Móvil, APP Pagos, APP Pagos pymes, ATMs; y, Corresponsales No bancarios)
Nro. transacciones	Numérica entera	Total, de transacciones por canal y cliente en el mes de estudio.
Monto acumulado	Numérica continua	Suma total de montos transaccionados por canal.
Media	Numérica continua	Promedio histórico mensual de montos transaccionados por canal.
Mediana	Numérica continua	Valor central de los montos transaccionados por canal.
Moda	Numérica continua	Valor más frecuente de transacción por canal.
Hora promedio	Numérica continua	Hora promedio transaccional por cliente y por canal transaccional
Monto penúltima transacción Yt-1	Numérica continua	Valor del penúltimo monto registrado por canal.
Máx. transacciones por día	Numérica entera	Mayor número de transacciones diarias realizadas en el mes.
Tiempo medio entre transacciones	Numérica continua	Velocidad transaccional promedio por canal (en horas).
Género	Nominal	Sexo del cliente.
Edad cumplida	Numérica entera	Edad
Nivel estudio	ordinal	Nivel de estudio de los clientes, aplica solo para personas naturales
Ingresos mensuales	Numérica continua	Ingreso mensual monetario del cliente.
Actividad del cliente	Nominal	indica el tipo de actividad económica a la que se dedica el cliente
Saldo promedio vista	Numérica continua	Promedio mensual del saldo en cuenta de ahorros.
Día con mayor transacción	Numérica entera	Día del mes con mayor número de transacciones por canal.

Nota. Elaboración (Omar Pinos, 2025).

¹ Hace referencia a clientes que en el mes de estudio hayan realizado al menos dos transacciones en cualquiera de los canales electrónicos que utilizaron

1.4. Estructura general del documento.

1.4.1. Laboratorio: Enfoque descriptivo inferencial

Este apartado da cuenta del objetivo general y específicos del laboratorio con énfasis al análisis estadístico descriptivo e inferencial, donde previo al abordaje del mismo se realizó:

- Preparación y análisis exploratorio de los datos (EDA)
- Análisis estadístico descriptivo: síntesis por canal y perfil (tendencia central: media/mediana; dispersión: DE/IQR; percentiles, histogramas de frecuencias), y análisis de temporalidad (horarios, estacionalidad).
- Análisis estadístico Inferencial:
 - Contraste de normalidad tripartita (Shapiro–Wilk/K–S/Anderson–Darling),
 - Heterocedasticidad (Breusch–Pagan/White/Levene),
 - Independencia de errores (Durbin–Watson) y colinealidad (VIF).
Comparaciones entre grupos mediante
 - Análisis post-hoc utilizando la prueba TukeyHSD (con corrección de Bonferroni) ANOVA/ Wilcoxon rank/
 - Matriz de correlación de Spearman contrastaciones según aplique.
- Modelos exploratorios: regresión lineal múltiple como línea base con diagnóstico de supuestos (normalidad, homocedasticidad, linealidad, colinealidad) y regresión cuantílica preliminar en $\tau = 0.50$ (mediana) y $\tau = 0.97$ (cola superior) para motivar el uso de bandas condicionales y el umbral de alerta. Adicional se aborda resultados, discusión y las conclusiones del laboratorio.

1.4.2. Laboratorio: Enfoque Machine Learning

En este laboratorio se aborda modelos de aprendizaje supervisados desarrollados para la detección de transacciones anómalas, y los indicadores de desempeño alcanzados en cada uno de ellos, el contenido se desarrolló de la siguiente manera:

- Objetivo general y específicos,
- Técnicas de modelos supervisados (Random Forest, W – Part, R part)
- Balanceado de la base 60% transacciones normales, 40% transacciones anómalas para cada canal electrónico.
- Entrenamiento de la base 70% y 30% prueba
- Indicadores de desempeño (Accuracy, Sensibilidad, especificidad, precisión, F1 score, coeficiente Kappa, ROC (AUC), 95% IC)
- Resultados
- Discusión
- Conclusiones.

1.4.3. Laboratorio: Enfoque Toma de Decisiones

En este laboratorio se aborda nuevamente modelos de aprendizaje no supervisados y supervisados desarrollados para segmentación de socios con perfil transaccional anómalo, y del vecino más cercano. Los resultados de indicadores de desempeño alcanzados en cada uno de ellos, el contenido se desarrolló de la siguiente manera:

- Objetivo general y específicos,
- Técnicas de agrupamiento (K medias)
 - Entrenamiento de la base 70% y 30% prueba
 - Indicadores de desempeño (coeficiente silueta, índice de pureza)
 - Segmentación de clientes con perfil anómalo por canal
- Técnica supervisada (KNN)
 - Entrenamiento de la base 70% y 30% prueba
 - Indicadores de desempeño (Accuracy, Sensibilidad, especificidad, precisión, F1 score, coeficiente Kappa, ROC (AUC), 95% IC)
- Resultados
- Discusión
- Conclusiones.

1.4.4. Laboratorio: Enfoque Robustez y Validación

Este laboratorio aborda la validación y robustez de los modelos realizados dentro del trabajo integrador, esto es Modelo de regresión cuantílica, Random Forest, W – Part, K medias, y KNN, el cual además cuenta con el apartado de resultados, discusión y conclusiones

2. Objetivo General

Mediante la ciencia de datos desarrollar modelos predictivos que permitan identificar el comportamiento transaccional anómalo de los clientes que usan canales electrónicos de una Cooperativa de Ahorro y Crédito del Ecuador

3. Laboratorio: Enfoque Descriptivo e Inferencial

3.1. Objetivos específicos

- Identificar las variables sociodemográficas, económicas y transaccionales que presentan mayor influencia sobre el monto transaccional no habitual, mediante técnicas de análisis estadístico e inferencial.
- Establecer umbrales y patrones de referencia (monto, horario habitual, frecuencia, velocidad transaccional) para detectar desviaciones respecto al comportamiento

transaccional habitual, considerando características sociodemográficas y económicas de los clientes.

- Mediante el modelo de regresión cuantílica determinar la mediana del monto transaccional habitual por canal con $\tau = 0.50$ y el umbral superior con $\tau = 0.97$ para determinar montos transaccionales anómalos y normales

3.2. Método

La presente investigación adopta un enfoque cuantitativo, empleando técnicas de estadística aplicada con el objetivo de desarrollar un modelo predictivo robusto que permita identificar transacciones anómalas en los canales electrónicos de una institución financiera del Ecuador. Se utiliza un diseño metodológico de investigación descriptivo, correlacional y explicativo, basado en el análisis de registros históricos de transacciones realizadas por los clientes, estructurado en cinco etapas:

3.2.1. Preparación y análisis exploratorio de datos

- Evaluación de la calidad de los datos (valores faltantes, duplicados, outliers)
- Limpieza de datos (tratamiento de valores extremos, manejo de registros con información incompleta)
- Estadística descriptiva profundizando el análisis del **i)** monto transaccional, **ii)** velocidad transaccional (promedio en horas entre transacciones realizadas por canal) **iii)** horario habitual; y, **iv)** frecuencia mensual de transacciones.
- Generación de tablas y gráficos de comportamiento típico por tipo de canal
- Identificación de perfiles transaccionales habituales diferenciados por canal electrónico.

3.2.2. Modelado del monto transaccional mediante (modelo lineal de regresión múltiple)

A través de un primer diagnóstico para estimar el monto transaccional por canal electrónico se desarrolla un modelo de regresión lineal múltiple considerando las variables (edad, tipo de género, actividad económica, monto acumulado transaccional, ingresos, saldo promedio en ahorro vista, nivel de estudios, y el tipo de canal).

3.2.3. Modelado del monto transaccional mediante (regresión cuantílica)

En esta etapa, se estima el monto transaccional utilizando un modelo de regresión cuantílica, técnica que modela cuantiles condicionales de $Y | X$; en particular, la mediana ($\tau = 0,50$) y un umbral superior (p. ej., $\tau = 0,97$) en función de variables sociodemográficas, económicas y de comportamiento transaccional bancario del cliente.

Definición de cuantil

Hyndman & Fan (1996) definen el cuantil de orden τ como el valor x_τ tal que la probabilidad acumulada satisface

$$F_x = (x_\tau) = P[X \leq x_\tau] = \tau.$$

Este valor delimita la proporción τ de observaciones inferiores o iguales dentro de la población.

Definición de Cuantil Condicional

El concepto de cuantil puede extenderse al contexto condicional. Para una variable aleatoria Y condicionada a un conjunto de covariables $X = x$, el cuantil condicional de orden τ se define como el valor $Q_Y(\tau | x)$ que satisface:

$$P = Y \leq Q_Y(\tau | x) | X = x = \tau.$$

De esta forma, la regresión cuantil estima la relación entre los cuantiles de Y y las covariables X , permitiendo analizar efectos heterogéneos a lo largo de la distribución de Y Koenker & Bassett (1978).

Función objetivo

El estimador $\hat{\beta}(\tau)$ se obtiene minimizando la pérdida asimétrica de chequeo o check:

$$\widehat{\beta\tau} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho\tau(y_i - x_i'\beta), \quad \rho\tau(u) = u(\tau - I\{u < 0\}).$$

Este enfoque no impone normalidad ni homocedasticidad de los residuos y es robusto frente a valores atípicos y distribuciones asimétricas Koenker & Bassett (1978).

Bondad de ajuste

Para evaluar la bondad de ajuste del modelo de regresión cuantílica se emplea el valor pseudo R de Koenker & Machado (1999) indicador análogo al coeficiente de determinación de la regresión lineal, definido como la proporción de la desviación absoluta total explicada por el modelo en relación con un modelo nulo (solo intercepto).

En consecuencia, esta metodología permite obtener predicciones más representativas del comportamiento transaccional real de los clientes, especialmente en contextos financieros donde los datos presentan colas largas, sesgos y comportamientos extremos que no son capturados adecuadamente por modelos basados en la media.

Ecuación 1. Estimación de la mediana del monto transaccional ($\tau = 50$):

$$\hat{Y}_{\tau=50} = \beta_0 + \beta_0 \chi_1 + D_1 \chi_2 + D_1 \chi_2 + D_2 \chi_3 + D_3 \chi_4 + \beta_2 \chi_5 + \beta_3 \chi_6 + \beta_4 \chi_7 + \beta_5 \chi_8 + \mu$$

Ecuación 2. Estimación del umbral superior ($\tau = 97$):

$$\hat{Y}_{\tau=97} = \beta_0 + \beta_0 \chi_1 + D_1 \chi_2 + D_1 \chi_2 + D_2 \chi_3 + D_3 \chi_4 + \beta_2 \chi_5 + \beta_3 \chi_6 + \beta_4 \chi_7 + \beta_5 \chi_8 + \mu$$

Donde:

$\hat{Y}_{\tau}$ = Valor estimado del monto transaccional para cuantil τ .

χ_1 = Edadc umplida (var. numérica)

χ_2 = Nivel de estudio (var. dummies)

χ_3 = Actividad del cliente (var. dummies)

χ_4 = Tipo de canal electrónico (var. dummies)

χ_5 = Total ingresos monetarios (var. numérica)

χ_6 = Saldo ahorros vista (var. numérica)

χ_7 = Monto última transacción (var. numérica)

χ_8 = Monto acumulado (var. numérica)

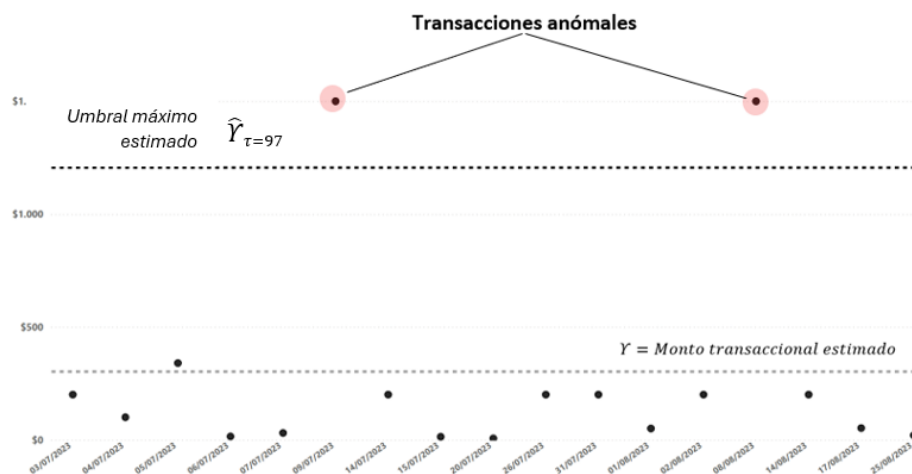
μ = Término de error para el cuantil estimado

3.2.4. Detección de anomalías por umbral estadístico

En esta etapa, se identifican transacciones como normales o anómalas en función de la desviación respecto al umbral superior predicho. Se adopta el siguiente criterio técnico:

- Si el monto observado de una transacción supera el valor estimado del cuantil 97 ($\hat{Y}_{\tau=97}$) esta es clasificada como transacción anómala.
- Si se encuentra debajo del umbral, se clasifica como transacción normal.

La Figura 2 muestra la distribución de las transacciones identificadas como anómalas bajo este enfoque.

Figura 2. Umbral máximo calculado para detección de transacciones anómalas

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

3.3. Discusión y resultados

3.3.1. Análisis estadístico descriptivo del perfil transaccional de los clientes

Los resultados estadísticos descriptivos dan cuenta que los canales electrónicos APP Móvil, y Web transaccional presentan una media del monto transaccional más elevadas en comparación con el resto de los canales, lo cual era previsible considerando que no poseen restricciones en el monto máximo permitido por transacción.

Asimismo, ambas plataformas muestran una alta dispersión en sus datos, reflejada en sus desviaciones estándar (APP Móvil: USD 857,44; Web Transaccional: USD 1.371,74), así como en sus amplios rangos intercuartílicos (IQR), lo cual sugiere la presencia de comportamientos extremos tanto en la cola inferior como superior de las distribuciones.

La Tabla 2 resume los principales estadísticos de tendencia central y dispersión por canal, incluyendo medidas como la media, mediana, moda, desviación estándar, percentiles 5 y 95, y los límites intercuartílicos de normalidad.

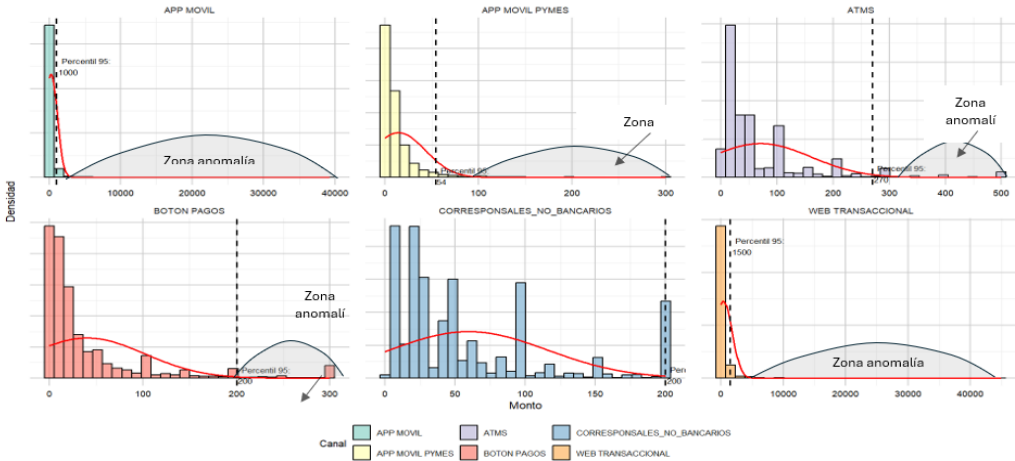
Tabla 2. Estadísticas descriptivas del monto transaccional por tipo de canal

Canal	Media	Mediana	Moda	Desviación_Estándar	Percentil5	Q1	Q3	Percentil95	Mínimo	Máximo	IQR	Límite_Inferior	Límite_Superior
APP MOVIL	247.57	55.0	20.00	857.44	5.000	20.00	190	1000	0.01	39300.00	170.00	20.00	445.00
APP_MOVIL_PYMES	14.45	5.5	0.35	28.72	0.350	2.25	15	54	0.01	300.00	12.75	2.25	34.12
ATMS	70.82	40.0	20.00	90.10	5.000	20.00	100	270	5.00	500.00	80.00	20.00	220.00
BOTON PAGOS	40.62	17.5	20.00	61.71	1.500	5.00	45	200	0.01	300.00	40.00	5.00	105.00
CORRESPONSALES_NO_BANCARIOS	59.61	40.0	20.00	56.33	6.000	20.00	98	200	0.50	200.25	78.00	20.00	215.00
WEB TRANSACCIONAL	407.90	100.0	100.00	1371.74	9.915	35.00	300	1500	0.00	45000.00	265.00	35.00	697.50

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

La Figura 3 muestra los histogramas de densidad de cada canal con su respectiva curva de normalidad teórica, en donde se ha definido una zona anómala (sombreada) como aquella correspondiente a transacciones con montos superiores al percentil 95. Este enfoque proporciona una aproximación no paramétrica para la identificación de transacciones atípicas, especialmente útil en distribuciones asimétricas y leptocúrticas como las observadas.

Figura 3. Histograma con curva de normalidad de la densidad y monto transaccional



Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

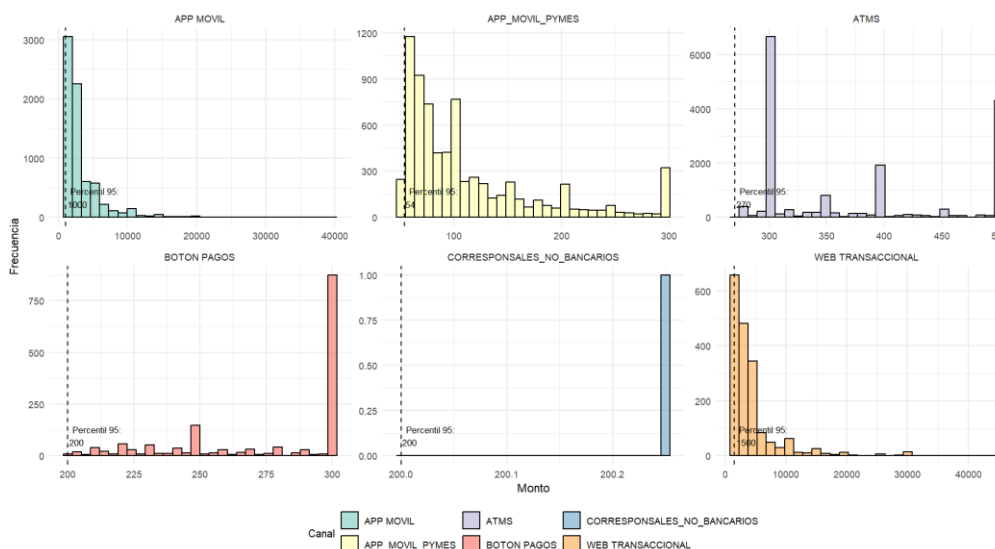
La Figura 4 presenta exclusivamente el histograma de frecuencia de las transacciones que superan el percentil 95, es decir, aquellas consideradas atípicas o anómalas en función del

comportamiento transaccional habitual de cada canal sin considerar aún el perfil sociodemográfico y económico de los clientes. Se evidencia además que:

- En APP Móvil y Web Transaccional, las transacciones atípicas se concentran principalmente entre USD 1.000 y USD 5.000.
- En Botón de Pagos, los montos atípicos se agrupan en torno a USD 300.
- En ATMs, los valores anómalos oscilan entre USD 300 y USD 500.
- En APP Móvil PYMES y Corresponsales No Bancarios, el comportamiento atípico se encuentra mayoritariamente por debajo de los USD 300.

Estos hallazgos permiten identificar umbrales operativos relevantes y posibles focos de revisión desde una perspectiva de riesgo.

Figura 4. Histograma de las transacciones atípicas por canal transaccional



Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

La siguiente Tabla 3 presenta un resumen del número máximo de transacciones diarias realizadas por cliente, desagregado por canal transaccional. En general, se observa que el promedio bordea a un total de 2 transacciones máximas por día y por canal, con valores medianos consistentes con dicha tendencia.

No obstante, al considerar el percentil 97, se identifican valores significativamente más altos para ciertos canales. Por ejemplo, en el canal BOTÓN DE PAGOS, el 97% de los clientes no supera las 9 transacciones máximas diarias, mientras que en APP MÓVIL y WEB TRANSACCIONAL este umbral se sitúa en 6 y 7 transacciones, respectivamente.

Este análisis permite establecer un criterio de referencia para la detección de patrones atípicos, donde se propone que cualquier observación que supere el percentil 97 del número máximo de transacciones diarias para su canal correspondiente podría considerarse una posible anomalía operativa o comportamiento irregular.

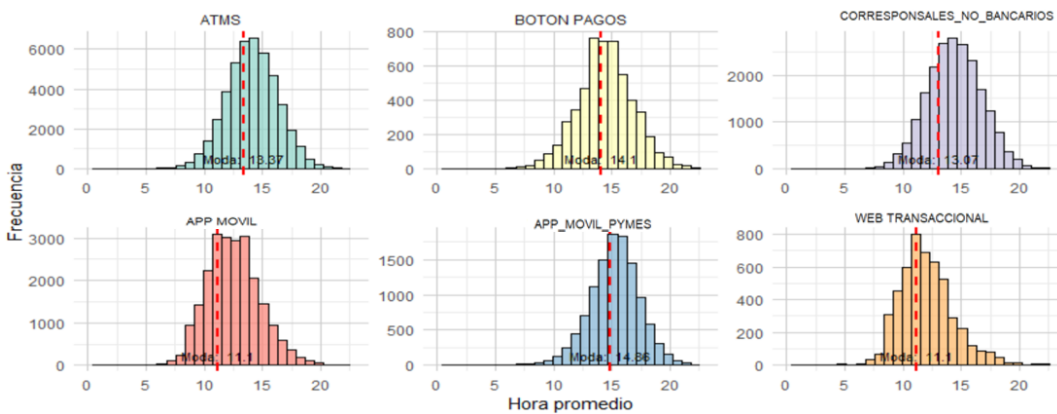
Tabla 3. Número de transacciones máximas realizadas por día y por canal

Servicio	media	Mediana	min	max	per.97
APP MOVIL	2.21	2	1	46	6
APP MOVIL_PYMES	2.33	2	1	40	6
ATMS	1.56	1	1	16	3
BOTON PAGOS	3.28	3	1	28	9
CORRESPONSAL_NO_BANCARIO	1.20	1	1	5	2
TRANSACCIONES_WEB	2.42	2	1	37	7

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

En cuanto al horario habitual transaccional la siguiente Figura 5 evidencia que la mayoría de clientes realizan transacciones entre las 13:00 a 14:00 pm excepto en los canales de APP Móvil y Web transaccional donde el horario habitual es a las 11:00 am.

Figura 5. Horario habitual transaccional



Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

La Tabla 4 presenta la velocidad transaccional medida como el tiempo promedio (en horas) entre dos transacciones consecutivas realizadas por un cliente, desagregado por canal. En promedio, los clientes realizan una nueva transacción entre 1,5 a 2 horas, aunque este valor varía según el tipo de canal utilizado.

Se observa que los canales APP Móvil PYMES y Web Transaccional presentan los menores tiempos medianos entre transacciones (60,00 y 73,23 horas, respectivamente), lo que indica una mayor frecuencia de uso y, por tanto, una mayor velocidad transaccional. En contraste, canales como Corresponsales No Bancarios y ATMs presentan tiempos promedio más altos (superiores a 95 horas), lo que sugiere un uso menos recurrente.

Este indicador permite establecer un umbral de normalidad en los patrones de comportamiento del cliente. Cuando se detectan transacciones realizadas en un intervalo de tiempo significativamente inferior al habitual por canal, se podría estar frente a una actividad atípica o potencialmente anómala, dado que la velocidad transaccional incrementada puede corresponder a comportamientos fuera de lo común del cliente o a posibles usos indebidos del canal.

Tabla 4. Tiempo en horas promedio entre transacciones por canal

Servicio	Tiempo promedio	Tiempo Máx.	Mediana
Atms	97,03	354,33	84.32
Botón de pagos	108,82	353,00	72.00
Corresponsales no bancarios	111,59	365,62	100.79
App móvil	90,56	351,50	80.00
App móvil pymes	72,45	353,83	60.00
Web Transaccional	84,33	340,50	73.23

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

3.3.2. Análisis estadístico descriptivo del perfil sociodemográfico y económico

Al analizar los indicadores estadísticos del monto transaccional por género (véase Tabla 5), se observa que los hombres presentan un monto promedio transaccional mayor (USD 136,72) en comparación con las mujeres (USD 107,20). Esta diferencia también se refleja en los valores de la mediana (USD 40 vs. USD 35) y el percentil 95 (USD 859,59 para hombres frente a USD 652,00 para mujeres), lo que indica que las transacciones más altas tienden a concentrarse en el grupo masculino.

No obstante, también se evidencia una mayor dispersión en los montos transaccionales del grupo masculino, con una desviación estándar (DE) de USD 471,92 frente a USD 365,14 para el grupo femenino. Esto sugiere una mayor variabilidad en los montos que transaccionan los hombres.

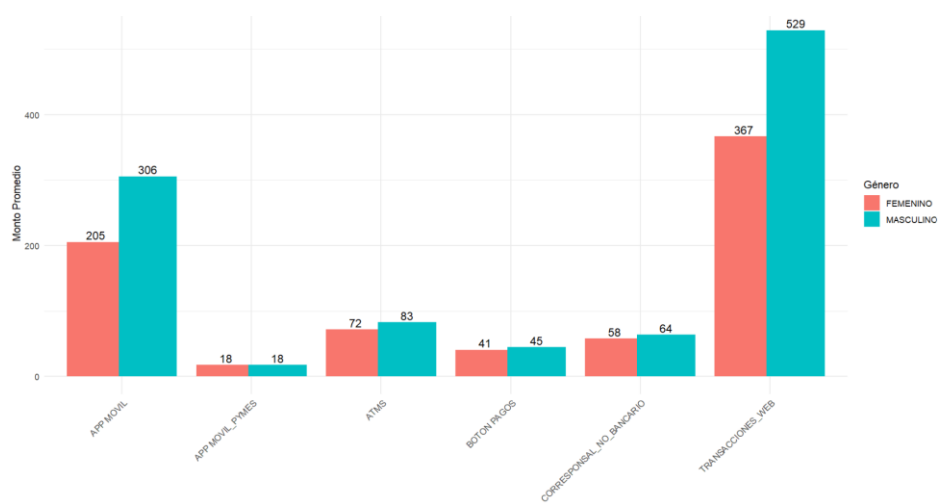
Tabla 5. Estadísticas de tendencia central del monto transaccional por tipo de sexo

genero	Media	Mediana	Moda	DE	Percentil5	Q1	Q3	Percentil95	Mínimo	Máximo	IQR	Límite_Inferior	Límite_Superior
FEMENINO	107.20	35	20	365.14	1.53	11.25	125.00	652.00	0.01	39300	113.75	11.25	295.625
MASCULINO	136.72	40	20	471.92	1.50	11.27	151.75	859.59	0.00	45000	140.48	11.27	362.470

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Como se puede observar en la Figura 6, el canal con mayores diferencias de monto promedio por género es el de APP Móvil, donde los hombres transaccionan significativamente en promedio montos más altos respecto a las mujeres. Esta tendencia se acentúa aún más en el canal Web Transaccional, con un promedio masculino de USD 529 frente a USD 367 de las mujeres. En el resto de los canales (APP Móvil Pymes, ATMs, Botón Pagos y Corresponsal no Bancario), las diferencias son mínimas, con montos similares entre ambos géneros.

Figura 6. Monto promedio transaccional por tipo de sexo y canal



Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

La siguiente Tabla 6, da cuenta del monto promedio transaccional en los canales electrónicos de acuerdo al nivel de instrucción de los clientes, en cual se aprecia que al incrementar el nivel de estudios de estos el monto promedio transaccional también aumenta.

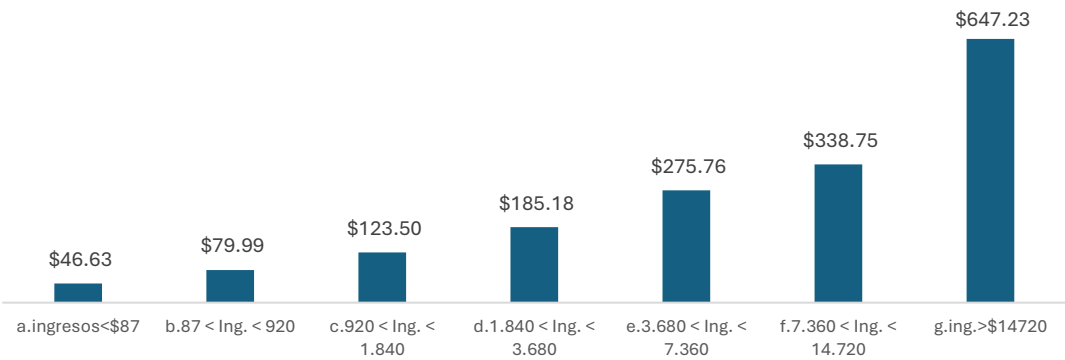
Tabla 6. Estadísticas del monto transaccional por nivel de instrucción

descripcion_nivel_estudio	Media	Mediana	Moda	Percentil5	Percentil95
SIN ESTUDIOS	83.41	40.00	20	2.50	415.00
PRIMARIA	97.52	40.00	20	2.58	500.00
SECUNDARIA	107.52	35.16	20	1.38	715.02
FORMACION INTERMEDIA (TECNICA - TECNOLOGIA)	111.69	33.00	20	0.99	938.80
NO DEFINIDO	124.80	40.00	20	1.35	1020.72
UNIVERSITARIA	145.29	41.00	20	1.07	1166.44
POSTGRADO	152.43	37.50	20	0.83	1441.99

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

La Figura 7 muestra de manera clara una relación directamente proporcional entre el nivel de ingresos mensuales de los clientes y el monto promedio transaccional realizado a través de canales electrónicos. A medida que el ingreso aumenta, también lo hace el monto promedio de las transacciones

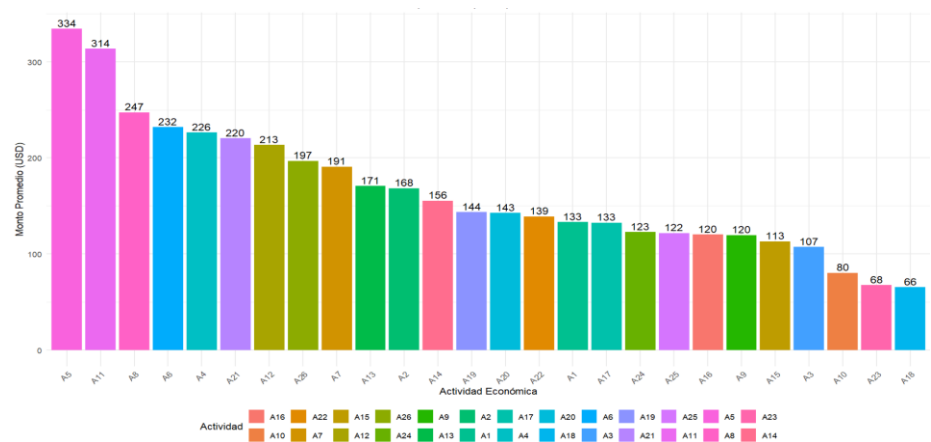
Figura 7. Monto promedio transaccional por rango de ingreso del cliente



Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Asimismo, se evidencia que el monto transaccional promedio varía significativamente en función de la actividad económica a la que se dedica el cliente. Por ejemplo, los clientes cuya actividad principal corresponde a Actividades Financieras y de Seguros (A5)² presentan un monto transaccional promedio de USD 334, el más alto entre los grupos analizados. En contraste, los clientes identificados como Estudiantes (A18) registran un monto promedio de apenas USD 66, ubicándose en el nivel más bajo como se ve en la siguiente Figura 8

Figura 8. Monto promedio transaccional por actividad económica del cliente



Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

² "ACTIVIDADES DE ALOJAMIENTO Y DE SERVICIO DE COMIDAS." ~ "A1",
"ACTIVIDADES DE ATENCION DE LA SALUD HUMANA Y DE ASISTENCIA SOCIAL." ~ "A2",
"ACTIVIDADES DE LOS HOGARES COMO EMPLEADORES; ACTIVIDADES NO DIFERENCIADAS DE LOS HOGARES COMO PRODUCTORES DE BIENES Y SERVICIOS PARA USO PROPIO." ~ "A3",
"ACTIVIDADES DE SERVICIOS ADMINISTRATIVOS Y DE APOYO." ~ "A4",
"ACTIVIDADES FINANCIERAS Y DE SEGUROS." ~ "A5",
"ACTIVIDADES INMOBILIARIAS." ~ "A6",
"ACTIVIDADES PROFESIONALES, CIENTIFICAS Y TECNICAS." ~ "A7",
"ADMINISTRACION PUBLICA Y DEFENSA; PLANES DE SEGURIDAD SOCIAL DE AFILIACION OBLIGATORIA." ~ "A8",
"AGRICULTURA, GANADERIA, SILVICULTURA Y PESCA." ~ "A9",
"AMA DE CASA" ~ "A10",
"ARTES, ENTRETENIMIENTO Y RECREACION." ~ "A11",
"COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACION DE VEHICULOS AUTOMOTORES Y MOTOCICLETAS." ~ "A12",
"CONSTRUCCION." ~ "A13",
"DISTRIBUCION DE AGUA; ALCANTARILLADO, GESTION DE DESECHOS Y ACTIVIDADES DE SANEAMIENTO." ~ "A14",
"EMPLEADO PRIVADO" ~ "A15",
"EMPLEADO PUBLICO" ~ "A16",
"ENSEÑANZA." ~ "A17",
"ESTUDIANTE" ~ "A18",
"EXPLOTACION DE MINAS Y CANTERAS." ~ "A19",
"INDUSTRIAS MANUFACTURERAS." ~ "A20",
"INFORMACION Y COMUNICACION." ~ "A21",
"JUBILADO" ~ "A22",
"NO DEFINIDO" ~ "A23",
"OTRAS ACTIVIDADES DE SERVICIOS." ~ "A24",
"SUMINISTRO DE ELECTRICIDAD, GAS, VAPOR Y AIRE ACONDICIONADO." ~ "A25",
"TRANSPORTE Y ALMACENAMIENTO." ~ "A26"

En relación a la edad generacional, se puede apreciar que existe diferencia en el monto promedio transaccional, de tal manera que al pasar de una generación menor a una mayor el monto promedio transaccional crece también, siendo así una variable de comportamiento esperado como se ve en la siguiente Tabla 7.

Tabla 7. Monto transaccional por edad edades generacionales

edad_generacional	Media	Mediana	Moda	Percentil5	Percentil95	Mínimo	Máximo	IQR
a.Generacion Z	62.20	20.0	10	1.00	387.60	0.01	20300	82.50
b.Generacion Millenials	113.71	37.5	20	1.19	840.00	0.01	39300	132.50
c.Generacion X	130.13	50.0	20	2.68	910.00	0.01	30000	146.12
d.Baby Boomers	144.26	50.0	100	5.00	856.93	0.15	45000	155.00

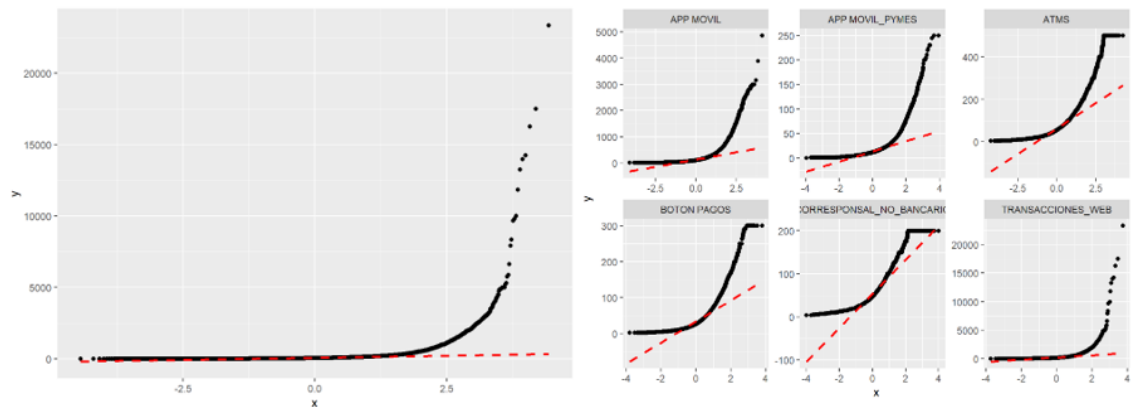
Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

3.3.3. Análisis estadístico inferencial y prueba de hipótesis

A partir del análisis estadístico inferencial realizado, se identificó que las variables numéricas —como el monto transaccional, los ingresos, y el saldo promedio— no presentan una distribución normal, según las pruebas de normalidad aplicadas. En consecuencia, se optó por utilizar pruebas de hipótesis no paramétricas, con el objetivo de evaluar si existen diferencias estadísticamente significativas entre grupos para dichas variables. Este procedimiento resulta clave para determinar cuáles variables presentan un comportamiento diferenciado entre segmentos y, por tanto, si deben ser consideradas en etapas posteriores de modelado predictivo.

3.3.3.1. Pruebas de normalidad

La siguiente Figura 9 presenta el gráfico Q-Q en el cual se puede apreciar claramente que la variable monto no sigue una distribución normal a nivel general y por tipo de canal electrónico.

Figura 9. Gráfico QQ variable del monto transaccional

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

En ese sentido, para validar la información visual, se aplicó la prueba tripartita para verificar si la variable Monto transaccional sigue una distribución normal, con las siguientes pruebas:

- **Kolmogorov-Smirnov**
- **Anderson-Darling**
- **Lilliefors (con versión corregida de Kolmogorov-Smirnov)**

En todos los casos, se obtuvo un valor de $p < 2.2e-16$, lo cual es significativamente menor que el umbral de significancia convencional ($\alpha = 0.05$).

Hipótesis planteadas:

- H_0 : La variable Monto transaccional sigue una distribución normal.
- H_1 : La variable Monto transaccional no sigue una distribución normal.

Conclusión estadística:

Dado que en todas las pruebas el p -value es inferior a 0.05, se rechaza la hipótesis nula (H_0). Por tanto, se concluye con evidencia estadística suficiente que la variable Monto transaccional no sigue una distribución normal.

Con el objetivo de determinar si la variable ingresos mensuales presenta una distribución normal, se aplicó la prueba tripartita obteniendo los siguiente resultados:

- **Kolmogorov-Smirnov:** $D = 0.49438$, $p < 2.2e-16$
- **Anderson-Darling:** $A = 42769$, $p < 2.2e-16$
- **Lilliefors:** $D = 0.49438$, $p < 2.2e-16$

Hipótesis planteada:

- H_0 : La variable Ingresos sigue una distribución normal.
- H_1 : La variable Ingresos no sigue una distribución normal.

Interpretación:

En todos los casos, el valor p es inferior al umbral de significancia de 0.05, lo que indica que se debe rechazar la hipótesis nula (H_0). Por tanto, se concluye que la variable Ingresos no sigue una distribución normal.

Para verificar si la variable saldo promedio en cuentas de ahorro a la vista sigue una distribución normal, se aplicaron igualmente las pruebas de Kolmogorov-Smirnov y Anderson-Darling. A continuación se presentan los resultados obtenidos:

- Kolmogorov-Smirnov: $D = 0.41639$, $p < 2.2e-16$
- Anderson-Darling: $A = 27082$, $p < 2.2e-16$

Hipótesis planteadas:

- H_0 : La variable Saldo Promedio Ahorros Vista sigue una distribución normal.
- H_1 : La variable Saldo Promedio Ahorros Vista no sigue una distribución normal.

Interpretación:

En ambas pruebas estadísticas, el valor p resultó ser menor al nivel de significancia de 0.05. Por tanto, se rechaza la hipótesis nula (H_0), concluyendo que la variable Saldo Promedio en Ahorros Vista no sigue una distribución normal.

3.3.3.2. Prueba de Wilcoxon, y Kruskal-Wallis para determinar diferencia de medianas entre grupos

Hipótesis:

- H_0 : Las medianas del monto transaccional entre hombre y mujer son iguales
- H_1 : Las medianas del monto transaccional entre hombre y mujer NO son iguales

Prueba aplicada: Wilcoxon rank sum test (equivalente a la prueba de Mann-Whitney para dos muestras independientes)

Resultado del test:

- Estadístico $W = 1454460496$
- Valor- $p < 2.2e-16$

Interpretación:

Dado que el valor-p es significativamente menor a 0.05, se rechaza la hipótesis nula (H_0). Por lo tanto, se concluye que existe una diferencia estadísticamente significativa entre las medianas del monto transaccional entre hombres y mujeres.

Hipótesis:

H_0 : Las medianas del monto transaccional entre los tipo de actividades que se dedican los clientes son iguales

H_1 : Las medianas del monto transaccional entre los tipo de actividades que se dedican los clientes NO son iguales

Para contrastar si las medianas del monto promedio transaccional difieren entre los distintos grupos de actividades económicas, se aplicó la prueba no paramétrica de Kruskal-Wallis.

El resultado obtenido ($\chi^2 = 3878.3$, $df = 25$, $p\text{-value} < 2.2e-16$) lo que permite rechazar la hipótesis nula (H_0) con un nivel de significancia del 5%, concluyéndose que las medianas del monto transaccional varían significativamente entre los distintos tipos de actividades económicas. Esto sugiere que la actividad económica desempeña un rol importante en el comportamiento transaccional de los clientes en canales electrónicos.

3.3.3.3. Análisis post-hoc mediante test de Tukey con corrección de Bonferroni

De otro lado, mediante el análisis post-hoc utilizando el test TukeyHSD (con corrección de Bonferroni) sobre un modelo ANOVA como aproximación comparativa, revelaron que las diferencias en las medianas del monto transaccional no son homogéneas entre los 26 grupos de actividades; por el contrario, se identificaron pares de actividades con diferencias significativas con valores p ajustados menores a 0.05, lo que refuerza la heterogeneidad observada en el comportamiento financiero de los usuarios por tipo de actividad.

- **H_0 :** Las medianas del monto transaccional entre el nivel de estudios de los clientes son iguales
- **H_1 :** Las medianas del monto transaccional entre el nivel de estudios de los clientes NO son iguales

Mediante la prueba no paramétrica de Kruskal-Wallis, se obtuvo el resultado de ($\chi^2 = 667.19$, $df = 6$, $p\text{-value} < 2.2e-16$) lo que permite rechazar la hipótesis nula (H_0) con un nivel de significancia del 5%, concluyendo que las medianas del monto transaccional entre los grupos de nivel de estudios difieren significativamente.

Mediante la aplicación de la prueba post-hoc también da cuenta de la diferencia de medianas del monto promedio transaccional entre niveles de estudios de los clientes, en el cual se destaca principalmente la diferencia entre un postgradista que tienen en promedio un

comportamiento de monto transaccional más alto respecto al resto de niveles de estudio, lo que además cumple con el supuesto esperado. La siguiente Tabla 8 da cuenta de de la diferencia de medias del monto transaccional intra grupos más significativos.

Tabla 8. Análisis *post-hoc* Tukey HSD entre niveles de estudio

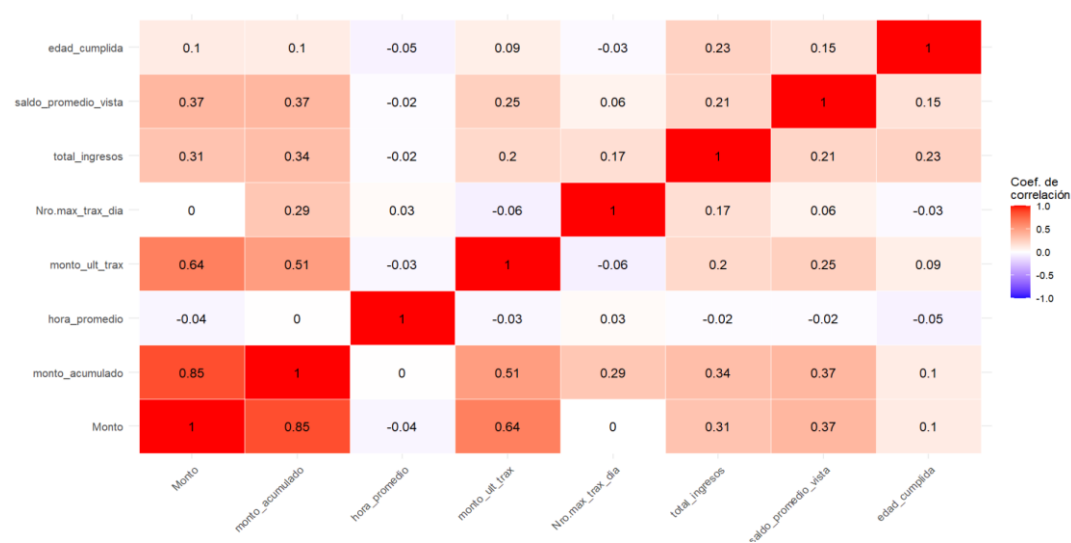
Comparación	diferencia	p_adj	Interpretación
Primaria – No definido	-22,6862	0,000	Primaria < No definido
Sin estudios – No definido	-28,90377	0,005	Sin estudios < No definido
Universitaria – No definido	17,43713	0,015	Universitaria > No definido
Primaria – Postgrado	-45,97729	0,000	Primaria < Postgrado
Secundaria – Postgrado	-36,25467	0,000	Secundaria < Postgrado
Sin estudios – Postgrado	-52,19486	0,000	Sin estudios < Postgrado
Tecnólogo – Postgrado	-28,34829	0,008	Tecnólogo < Postgrado
Tecnólogo – Primaria	17,62902	0,007	Tecnólogo > Primaria
Universitaria – Primaria	40,12337	0,000	Universitaria > Primaria
Universitaria – Secundaria	30,46042	0,000	Universitaria > Secundaria
Universitaria – Sin estudios	46,3409	0,000	Universitaria > Sin estudios
Universitaria – Tecnólogo	22,49433	0,000	Universitaria > Tecnólogo

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

3.3.3.4. Matriz de correlación de Spearman

Con el objetivo de identificar relaciones significativas entre las variables transaccionales del cliente, se construyó una matriz de correlación basada en el coeficiente de Spearman, luego de aplicar una transformación logarítmica natural a los datos para homogeneizar la escala de las variables. Como se observa en la siguiente Figura 10, se evidencia una alta correlación entre el monto transaccional y el monto acumulado, así como una asociación moderada con el monto de la última transacción, también se evidencia una correlación entre ligera y moderada entre variables como el ingreso y saldo en ahorros vista respecto a la variable monto y monto acumulado. En contraste, variables como la hora promedio de transacción o el número máximo de transacciones por día presentan correlaciones débiles, lo que sugiere independencia parcial respecto a los monto transaccional realizado.

Figura 10. Matriz de correlación de Spearman



Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

3.3.3.5. Regresión lineal múltiple

Entre los principales hallazgos se tiene que el modelo cuenta con un buen coeficiente de determinación R^2 de 0,63 y un R^2 ajustado de 0,63 dando cuenta de que las variables analizadas explican en un 63% el comportamiento del monto transaccional de los clientes en función de su perfil socio demográfico, económico y del comportamiento transaccional bancario, de otro lado, se obtiene que las variables más importantes que influyen en el comportamiento son: el monto acumulado, edad, sexo masculino, tipo de canal, monto última transacción, saldo en ahorros vista, rango de ingresos altos, tiempo promedio entre transacciones, número máximo de transacciones por día, actividades económicas como: información y comunicación, actividades financieras y de seguros (véase la siguiente Tabla 9).

Tabla 9. Resultados modelo de regresión lineal múltiple

Variable	Coefficiente	Error estándar	Valor t	Pr(> t)	Nivel de Significancia
(Intercepto)	84,9605	17,9573	4,7312	0	***
Millennials	-0,8357	1,9105	-0,4374	0,662	
Generación X	-1,451	2,4365	-0,5955	0,552	
Babby Boomers	3,04	4,0285	0,7546	0,451	
Silent Generación	19,8522	10,6029	1,8723	0,061	
Rango ingresos \$87<Ing.<\$470	14,608	16,1767	0,903	0,367	
Rango ingresos \$471<Ing.<\$940	16,912	16,2457	1,041	0,298	
Rango ingresos \$941<Ing.<\$1.410	24,7025	16,2881	1,5166	0,129	
Rango ingresos \$1.411<Ing.<\$1.880	32,2542	16,3778	1,9694	0,049	*
Rango ingresos \$1.881<Ing.<\$2.350	30,2871	16,5582	1,8291	0,067	
Rango ingresos >\$2.350	34,3637	16,375	2,0985	0,036	*
Sexo Masculino	6,9166	1,503	4,6018	0	***
Nivel estudio No definido	7,1998	8,9647	0,8031	0,422	
Nivel estudio Posgrado	17,3937	10,7408	1,6194	0,105	
Nivel estudio Primaria	-0,7688	5,878	-0,1308	0,896	
Nivel estudio Secundaria	0,9304	5,7734	0,1611	0,872	
Nivel estudio Sin estudios	5,1381	10,13	0,5072	0,612	
Nivel estudio Universitaria	4,4455	5,8717	0,7571	0,449	
Servicio APP MOVIL_PYMES	-58,7345	2,603	-22,5642	0	***
Servicio ATMS	-59,7883	1,9313	-30,9582	0	***
Servicio BOTON PAGOS	-20,938	3,2687	-6,4056	0	***
Servicio CORRESPONSAL NO BANCARIO	-66,9757	2,4967	-26,8261	0	***
Servicio TRANSACCIONES_WEB	73,4384	3,5483	20,6969	0	***
Ama de casa	3,0718	5,7606	0,5332	0,594	
Artes, entretenimiento y recreación	160,6264	28,4965	5,6367	0	***
Comercio al por mayor y al por menor	-2,4109	5,6746	-0,4249	0,671	
Construcción	1,7659	7,3642	0,2398	0,811	
Distribución de agua	1,0575	81,106	0,013	0,99	
Empleado privado	4,233	5,2609	0,8046	0,421	
Empleado público	-3,5793	5,3782	-0,6655	0,506	
Enseñanza	7,2224	11,5837	0,6235	0,533	
Estudiante	-1,9101	5,9879	-0,319	0,75	
Explotación de minas y canteras	8,5287	13,8752	0,6147	0,539	
Atención de la salud humana	-2,8038	8,6293	-0,3249	0,745	
Industrias manufactureras	6,597	6,5286	1,0105	0,312	
Información y comunicación	-42,5005	14,6326	-2,9045	0,004	**
Jubilado	-9,1132	8,2473	-1,105	0,269	
No definido	-11,7332	81,106	-0,1447	0,885	
Otras actividades de servicios	1,3744	9,1004	0,151	0,88	
Suministros de electricidad	-11,7746	20,3623	-0,5783	0,563	
Transporte y almacenamiento	2,5118	6,9346	0,3622	0,717	
Actividades de los hogares como empleadores	-0,6804	15,5946	-0,0436	0,965	

Servicios de administración y de apoyo	-1,0334	10,3804	-0,0995	0,921	
Actividades financieras y de seguros	-95,0178	20,2429	-4,6939	0	***
Actividades inmobiliarias	14,5741	15,6065	0,9338	0,35	
Actividades profesionales, científicas y técnicas	3,7727	6,8304	0,5523	0,581	
Administración pública y defensa	-9,5156	29,4802	-0,3228	0,747	
Agricultura, ganadería, silvicultura y pesca	-0,8439	5,908	-0,1428	0,886	
Saldo promedio Ahorro Vista	0,0038	0,0002	25,0272	0	***
Monto última Transacción	0,4142	0,0016	263,048	0	***
Tiempo promedio entre transacciones	0,0334	0,0071	4,7226	0	***
Monto Acumulado	0,0551	0,0003	187,4582	0	***
Número máximo de transacciones por día	-24,9984	0,5812	-43,0083	0	***

Nota. Se reportan los coeficientes no estandarizados (β), error estándar (EE), valor t y valor p.

Las categorías base son: Edad generacional Z; Rango de ingresos < \$ 87; Sexo Femenino; Actividades de alojamiento y servicios de comida; Nivel de educación Formación intermedia (Técnica y Tecnología).

$p < .05$ *, $p < .01$ **, $p < .001$ ***.

Error estándar residual = 97,54; $R^2 = 0,6301$; R^2 ajustado = 0,63;

$F(52, 123\ 205) = 4037$; $p < .001$.

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

En relación con el cumplimiento de supuestos, se evidencia que no cumple con el supuesto de normalidad al desarrollar la prueba tripartita todas las pruebas obtuvieron valor p -value menor al 0,05.

Asimismo, se evidencia mediante la prueba de Breuch-Pagan que no se cumple con el supuesto de varianzas iguales de los residuos lo que afecta los intervalos de confianza, siendo el valor obtenido menor al valor p -value 0.05.

Se comprobó también que el modelo a pesar de no cumplir con el supuesto de normalidad y homocedasticidad, sin embargo, no presenta multicolinealidad en las variables explicativas, siendo el valor VIF (Factor de inflación de varianza) menor a 5 en todas las ellas.

3.3.3.6. Regresión cuantílica ($\tau = 50$) para estimar el monto transaccional

A continuación, se presenta los resultados del modelo de regresión cuantílica estimado para el cuantil $\tau = 50$, correspondiente a la mediana del monto transaccional. El análisis se realizó mediante el método de mínimos absolutos ponderados, utilizando errores estándar obtenidos por *bootstrapping* con 100 remuestreos, lo que permite una estimación robusta de la significancia estadística de los coeficientes.

El modelo presenta un pseudo R^2 de Koenker & Machado (1999) igual a 0,298, lo cual indica una capacidad explicativa baja sobre el comportamiento de la mediana del monto

transaccional. La siguiente Tabla 10 reporta coeficientes, error estándar bootstrap, estadístico t, p-valor e IC95%, ordenados por significancia ($|t|$) de mayor a menor de las variables que aportan al modelo y su respectivo nivel de significancia.

Tabla 10. Regresión cuantílica ($\tau=0,50$): determinantes del percentil 50 del monto

Variables	Coeficiente	EE (boot)	Valor t	Valor p	IC95%	Nivel significancia
Constante	24.479	1,544	15,855	0,000	[21.453, 27.505]	***
Millennials	1.663	0,146	11,376	0,000	[1.377, 1.950]	***
Generación X	4.516	0,296	15,231	0,000	[3.935, 5.097]	***
Babby Boomers	8.908	0,593	15,019	0,000	[7.746, 10.071]	***
Silent Generación	17.378	3,253	5,342	0,000	[11.002, 23.754]	***
Nivel estudio POSTGRADO	-2.718	0,804	-3,38	0,001	[-4.294, -1.142]	***
Nivel estudio PRIMARIA	1.466	0,523	2,804	0,005	[0.441, 2.490]	**
Nivel estudio SECUNDARIA	1.078	0,506	2,132	0,033	[0.087, 2.069]	*
Sexo Masculino	0,793	0,139	5,704	0,000	[0.520, 1.065]	***
Rango ingresos ing.>\$2.350	6.309	1,377	4,582	0,000	[3.611, 9.008]	***
Rango ingresos \$87<Ing.<\$470	-4.793	1,29	-3,717	0,000	[-7.320, -2.265]	***
Rango ingresos \$471<Ing.<\$940	-3.377	1,286	-2,627	0,009	[-5.897, -0.858]	**
Rango ingresos \$941<Ing.<\$1.410	-2.719	1,286	-2,115	0,034	[-5.239, -0.199]	*
Rango ingresos \$1.411<Ing.<\$1.880	-2.491	1,297	-1,921	0,055	[-5.032, 0.051]	•
Servicio APP MOVIL_PYMES	-21.110	0,398	-53,023	0,000	[-21.890, -20.330]	***
Servicio BOTON PAGOS	-16.302	0,408	-39,935	0,000	[-17.102, -15.502]	***
Servicio CORRESPONSAL NO BANCARIO	-9.476	0,351	-26,997	0,000	[-10.164, -8.788]	***
Servicio ATMS	-9.376	0,358	-26,211	0,000	[-10.077, -8.675]	***
Servicio TRANSACCIONES WEB	16.127	0,94	17,158	0,000	[14.284, 17.969]	***
Comercio al por mayor y al por menor	2.855	0,632	4,515	0,000	[1.616, 4.094]	***
Transporte y almacenamiento	5.279	0,923	5,718	0,000	[3.469, 7.089]	***
Agricultura, ganadería, silvicultura y pesca	1.989	0,58	3,431	0,001	[0.853, 3.126]	***
Estudiante	-1.466	0,534	-2,745	0,006	[-2.513, -0.419]	**
Empleado Privado	-1.241	0,48	-2,585	0,010	[-2.181, -0.300]	**
Administración pública y defensa	7.359	3,633	2,026	0,043	[0.238, 14.481]	*
Servicios administrativos y de apoyo	2.501	1,328	1,883	0,060	[-0.102, 5.104]	•
Explotación de minas y canteras	3.666	2,001	1,832	0,067	[-0.257, 7.588]	•
Saldo promedio vista	0,001	0,000	15,047	0,000	[0.001, 0.001]	***
Monto última transacción	0,504	0,005	110,838	0,000	[0.509, 0.499]	***
Tiempo promedio entre transacciones	0,022	0,001	23,137	0,000	[0.023, 0.021]	***
Número máximo de transacciones por día	-1,1547	0,06	-18,95	0,000	[-1,094, -1,2147]	***
Atención de la salud humana	-0,03	0,784	-0,038	0,970	[-1.567, 1.508]	
Nivel estudio UNIVERSITARIA	-0,065	0,529	-0,123	0,902	[-1.102, 0.972]	
Distribución de agua	-9,563	44,785	-0,214	0,831	[-97.342, 78.215]	
Rango ingresos \$1.881<Ing.<\$2.350	-0,285	1,296	-0,22	0,826	[-2.825, 2.256]	
No definido	-4,646	16,632	-0,279	0,780	[-37.244, 27.952]	
Actividades de los hogares como empleadores	0,469	1,376	0,341	0,733	[-2.228, 3.167]	
Construcción	0,333	0,804	0,414	0,679	[-1.244, 1.909]	

Artes, entretenimiento y recreación	1,496	3,033	0,493	0,622	[-4.448, 7.441]
Actividades financieras y de seguros	1,861	3,146	0,591	0,554	[-4.305, 8.027]
Suministros de electricidad	-0,814	1,31	-0,622	0,534	[-3.381, 1.753]
Actividades profesionales, científicas y técnicas	-0,434	0,645	-0,673	0,501	[-1.698, 0.830]
Enseñanza	0,902	1,308	0,69	0,490	[-1.662, 3.466]
Actividades inmobiliarias	4,691	6,102	0,769	0,442	[-7.269, 16.651]
Ama de casa	-0,526	0,558	-0,943	0,346	[-1.619, 0.567]
Información y Comunicación	-1,489	1,578	-0,944	0,345	[-4.581, 1.603]
Otras actividades de servicios	-0,695	0,693	-1,004	0,316	[-2.053, 0.662]
Industrias Manufactureras	0,69	0,601	1,148	0,251	[-0.488, 1.868]
Empleado Público	-0,609	0,485	-1,256	0,209	[-1.560, 0.342]
Jubilado	1,303	1,028	1,267	0,205	[-0.712, 3.317]
Nivel estudio SIN ESTUDIOS	1,232	0,841	1,464	0,143	[-0.417, 2.880]
Nivel estudio NO DEFINIDO	0,992	0,665	1,492	0,136	[-0.312, 2.296]

Nota. Se reportan los coeficientes, error boot (EE), valor t, valor p, IC95%, y Nivel significancia.

Las categorías base son: Edad generacional Z; Rango de ingresos < \$ 87; Sexo Femenino; Actividades de alojamiento y servicios de comida; Nivel de educación Formación intermedia (Técnica y Tecnología).

$p < .05$ *, $p < .01$ **, $p < .001$ ***.

P pseudo R^2 de Koenker & Machado = 0,298

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Entre los resultados más relevantes se destaca lo siguiente:

- **Edad generacional:** Los clientes que pertenecen a los grupos generacionales con edad más alta como Baby Boomers, y Generación X, muestran montos transaccionales medianos significativamente mayores respecto al grupo de referencia (Generación Z) lo cual tiene el comportamiento esperado de acuerdo con el análisis exploratorio de datos realizado previamente
- **Rango de ingresos:** Se observa una relación creciente entre el nivel de ingresos y el monto transaccional, en ese sentido, los clientes con ingresos superiores a USD 2.350 registran un incremento significativo en la mediana del monto transaccional de USD 6,30 veces más alto respecto al promedio de la categoría base (ingresos < USD 87) ($p < 0.001$) manteniendo constantes las demás variables.
- **Género:** Ser hombre se asocia con un monto transaccional mediano mayor USD 0.79 más respecto al tipo de género Femenino ($p < 0.001$), manteniendo constantes las demás variables.
- **Nivel de estudios:** El tener nivel de estudios secundarios y primarios tienden en promedio a realizar transacciones superiores respecto a los que tienen formación

intermedia (técnica – tecnología) , mientras que los clientes que tienen **postgrado** presenta un efecto negativo en comparación al grupo de referencia (Formación intermedia (técnica - tecnología), es decir que en promedio los clientes que tienen posgrado realizan en promedio montos transaccionales de menor valor respecto al grupo de referencia.

- **Tipo de canal utilizado:** Los canales APP Móvil PYMES, Botón de pagos, Corresponsal no bancario y ATMs se asocian con una disminución significativa del monto transaccional, en comparación con el canal de referencia (APP Móvil)
- **Actividad del cliente:** Algunas categorías de actividad económica, como Comercio al por mayor y menor, Transporte y Almacenamiento; y, Agricultura, ganadería y silvicultura, tienen un efecto positivo y significativo sobre la mediana del monto transaccional respecto al grupo de referencia (Actividades de alojamiento y de servicio de comidas) indicando diferencias en el comportamiento financiero por tipo de cliente, en contraste con actividades económicas como estudiantes, y empleado público.
- **Variables financieras continuas:** Las variables saldo promedio en cuenta vista, monto de la última transacción, tiempo promedio entre transacciones, y número de transacciones realizadas por día son predictores positivos y altamente significativos, destacando que, por cada incremento unitario en estas variables, se incrementa también la mediana del monto transaccional.

Estos resultados reflejan un comportamiento transaccional diferenciado según el perfil sociodemográfico y financiero del cliente, lo cual resulta clave para el diseño de mecanismos predictivos de alertas por anomalías.

3.3.3.7. Regresión cuantílica para $\tau = 0,97$: estimación del umbral superior del monto transaccional

Con el objetivo de identificar los factores que explican los montos transaccionales más elevados y establecer un umbral de referencia para la detección de posibles transacciones anómalas, se estimó un modelo de regresión cuantílica con un valor de $\tau = 0,97$, correspondiente al percentil 97 de la distribución del monto transaccional. Este enfoque permitió modelar los extremos superiores sin que la estimación se viera distorsionada por valores atípicos o la no normalidad del residuo.

El ajuste presenta un pseudo- R^2 de 0,434, lo cual indica una capacidad explicativa moderada en este cuantílico extremo. La siguiente Tabla 11 presenta los coeficientes, error estándar bootstrap, estadístico t, p -value, y IC95%, ordenadas por significancia ($|t|$) de mayor a menor.

Tabla 11. Resultados del modelo de regresión cuantílica ($\tau = 0,97$)

Variables	Coeficiente	EE (boot)	Valor t	Valor p	IC95%	Nivel significancia
Constante	232.377	30,293	7,671	0,000	[173.004, 291.750]	***
Sexo Masculino	5.615	1,304	4,305	0,000	[3.059, 8.172]	***
Servicio APP MOVIL_PYMES	-227.447	8,351	-27,235	0,000	[-243.815, -211.078]	***
Servicio ATMS	-182.969	7,906	-23,143	0,000	[-198.465, -167.473]	***
Servicio BOTON PAGOS	-187.754	8,478	-22,146	0,000	[-204.371, -171.138]	***
Servicio CORRESPONSAL_NO_BANCARIO	-201.637	8,083	-24,946	0,000	[-217.480, -185.794]	***
Servicio TRANSACCIONES_WEB	229.772	43,729	5,254	0,000	[144.063, 315.480]	***
Babby Boomers	11.369	5,208	2,183	0,029	[1.162, 21.575]	*
Generación X	3.942	2,101	1,876	0,061	[-0.176, 8.060]	•
Nivel estudio POSTGRADO	-16.058	8,733	-1,839	0,066	[-33.175, 1.059]	•
Rango ingresos.>\$2.350	69.440	29,076	2,388	0,017	[12.452, 126.427]	*
Saldo promedio vista	0,011	0,002	6,176	0,000	[0.007, 0.014]	***
Monto última transacción	1.134	0,04	28,125	0,000	[1.055, 1.213]	***
Tiempo promedio entre transacciones	0,115	0,008	13,613	0,000	[0.098, 0.131]	***
Comercio al por mayor y al por menor	20.415	4,387	4,653	0,000	[11.816, 29.014]	***
Nivel estudio NO DEFINIDO	-6,315	8,499	-0,743	0,457	[-22.974, 10.343]	
Nivel estudio SIN ESTUDIOS	3,257	10,327	0,315	0,752	[-16.984, 23.499]	
Nivel estudio UNIVERSITARIA	-5,747	7,886	-0,729	0,466	[-21.204, 9.710]	
Nivel estudio PRIMARIA	-0,164	7,839	-0,021	0,983	[-15.529, 15.200]	
Nivel estudio SECUNDARIA	-0,114	7,802	-0,015	0,988	[-15.406, 15.179]	
Silent Generación	6,654	10,591	0,628	0,53	[-14.104, 27.412]	
Millennials	0,672	1,075	0,626	0,532	[-1.434, 2.779]	
Rango ingresos \$1.881<Ing.<\$2.350	15,023	27,932	0,538	0,591	[-39.724, 69.769]	
Rango ingresos \$1.411<Ing.<\$1.880	1,573	27,658	0,057	0,955	[-52.637, 55.783]	
Rango ingresos \$941<Ing.<\$1.410	-2,341	27,356	-0,086	0,932	[-55.958, 51.276]	
Rango ingresos \$471<Ing.<\$940	-3,814	27,306	-0,14	0,889	[-57.333, 49.705]	
Rango ingresos \$87<Ing.<\$470	-9,508	27,279	-0,349	0,727	[-62.975, 43.958]	
Artes, entretenimiento y recreación	51,281	3122,449	0,016	0,987	[-6068.667, 6171.229]	
No definido	-4,488	29,15	-0,154	0,878	[-61.622, 52.646]	
Suministro de electricidad	9,147	59,112	0,155	0,877	[-106.711, 125.005]	
Estudiante	0,664	2,992	0,222	0,824	[-5.200, 6.528]	

Administración Pública y defensa	43,549	189,626	0,23	0,818	[-328.115, 415.214]
Jubilado	-2,229	8,307	-0,268	0,788	[-18.511, 14.053]
Actividades de los hogares como empleadores	3,318	11,321	0,293	0,769	[-18.871, 25.506]
Enseñanza	5,047	15,114	0,334	0,738	[-24.577, 34.671]
Explotación de minas y canteras	-3,047	8,948	-0,34	0,733	[-20.585, 14.491]
Empleado público	-1,243	3,068	-0,405	0,685	[-7.256, 4.771]
Agricultura, ganadería, silvicultura y pesca	2,079	3,738	0,556	0,578	[-5.246, 9.405]
Actividades profesionales, científicas y técnicas	2,946	5,264	0,56	0,576	[-7.370, 13.263]
Otras actividades de servicios	4,095	6,203	0,66	0,509	[-8.062, 16.253]
Atención de la salud humana	10,389	12,753	0,815	0,415	[-14.606, 35.385]
Ama de casa	2,521	3,043	0,829	0,407	[-3.442, 8.485]
Actividades financieras y de seguros	253,501	303,355	0,836	0,403	[-341.070, 848.072]
Industrias manufactureras	4,324	4,835	0,894	0,371	[-5.151, 13.800]
Información y comunicación	18,745	18,361	1,021	0,307	[-17.243, 54.732]
Transporte y almacenamiento	8,328	7,859	1,06	0,289	[-7.075, 23.731]
Distribución del agua	193,515	149,089	1,298	0,194	[-98.698, 485.727]
Actividades inmobiliarias	57,905	44,254	1,308	0,191	[-28.833, 144.643]
Empleado privado	3,798	2,898	1,311	0,19	[-1.881, 9.478]
Servicios administrativos y de apoyo	33,423	22,201	1,506	0,132	[-10.089, 76.936]
Construcción	15,772	9,944	1,586	0,113	[-3.719, 35.262]

Nota. Se reportan los coeficientes, error boot (EE), valor t, valor p, IC95%, y Nivel significancia.

Las categorías base son: Edad generacional Z; Rango de ingresos < \$ 87; Sexo Femenino; Actividades de alojamiento y servicios de comida; Nivel de educación Formación intermedia (Técnica y Tecnología).

$p < .05$ *, $p < .01$ **, $p < .001$ ***.

P pseudo R^2 de Koenker & Machado = 0,434

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

A continuación se presentan los principales hallazgos del modelo.

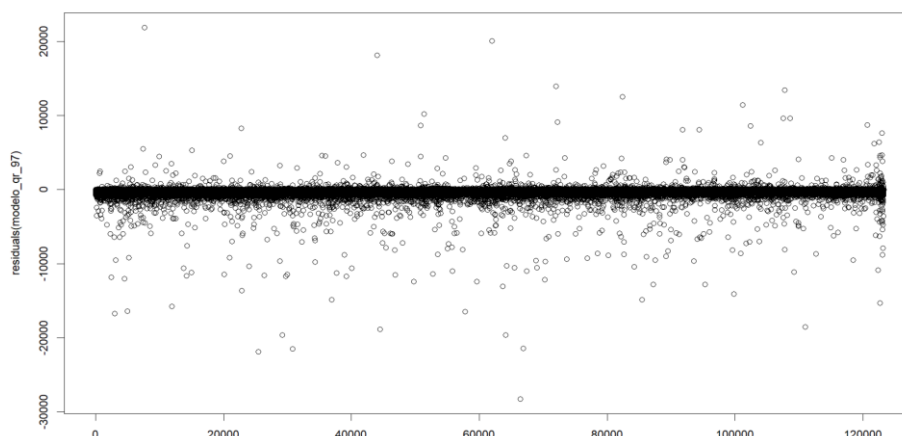
- **Ingresos mensuales:** Los individuos con ingresos superiores a USD 2.350 experimentan un aumento estadísticamente significativo de **\$69,44** en el monto transaccional en el percentil 97 en comparación con el grupo de referencia (ingresos < USD 87). Esto resalta una alta diferencia transaccional en la cola superior de la distribución.
- **Género:** “Ser de género Masculino se asocia con un aumento altamente significativo de **USD 5,61** en el monto transaccional en el percentil 97 en comparación con el género femenino. Esta diferencia absoluta en la cola superior de las transacciones es estadísticamente fuerte.

- **Nivel educativo:** Manteniendo las demás variables constantes, pertenecer al nivel de estudio Posgrado se asocia con una disminución de U\$D 16,058 unidades monetarias en el monto transaccional en el percentil 97 de la distribución (respecto a la categoría de nivel de formación intermedia).
- **Canales de servicio electrónico:** Los canales como APP Móvil PYMES, Corresponsales no bancarios, ATMS, y BOTÓN DE PAGOS se relacionan con disminuciones altamente significativas en el monto transaccional en este percentil con reducciones entre USD 227,44; USD 201,63; USD182,96; y USD187,75 respectivamente en relación al grupo de referencia (App Móvil) ($p < 0,001$), lo que sugiere patrones operativos diferenciados respecto a montos altos.
- **Actividad económica:** La actividad de comercio al por mayor y menor tiene una relación positiva en el incremento de la mediana del monto transaccional en USD 20,41 con respecto al grupo de referencia (Actividades de alojamiento y de servicio de comidas) ($p < 0,001$).
- **Comportamiento transaccional:** El saldo promedio en ahorros vista, el monto de la última transacción y el tiempo promedio entre transacciones por hora se relacionan positivamente con el aumento del monto transaccional en el percentil 97 ($p < 0,001$ en todos los casos).
- **Edades generacionales:** las generaciones Baby Boomers, y X son significativas y aportan al modelo, en ese sentido el hecho de que los clientes se encuentren dentro de estas categorías la mediana del monto transaccional será superior respecto a la categoría base (generaciónZ)

Estos hallazgos permiten caracterizar el perfil sociodemográfico y comportamental de los clientes que operan con montos transaccionales elevados en los canales electrónicos, y proporcionan insumos fundamentales para el establecimiento de umbrales de referencia orientados a la detección de transacciones anómalas.

Análisis de residuos del modelo de regresión cuantil $\tau = 0,97$

Mediante el gráfico de dispersión de los residuos se evidencia que la mayoría se concentran alrededor del eje horizontal (valor 0), sin evidencia clara de patrones no lineales, heterocedasticidad, lo que da cuenta que el modelo logra explicar de manera razonable los valores esperados del cuantil 97 del monto transaccional mediano (Mediana) en función de los predictores considerados.

Figura 11. Gráfico de dispersión de los residuos modelo de regresión cuantil $\tau = 0,97$ 

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

3.3.3.8. Estadísticas generales de las transacciones anómalas que superan el umbral superior del modelo de regresión cuantil ($\tau = 0,97$)

A partir de los resultados obtenidos mediante la aplicación del modelo de regresión cuantílica, se utilizó el cuantil superior ($\tau = 0.97$) como umbral de corte para identificar transacciones anómalas. Este enfoque permitió detectar 24.928 transacciones anómalas de un total de 146.483 observaciones, lo que representa una proporción global de del 17%. No obstante, se observa una marcada heterogeneidad en la distribución de anomalías según el tipo de canal, como se detalla en la Tabla 12.

Tabla 12. Número de transacciones anómalas por tipo de canal

Servicio	NORMAL	ANOMALO	total	% Anómalo
App móvil	4.555	3.118	7.673	40,64
App móvil pymes	28.121	6.396	34.517	18,53
Atms	60.897	10.808	71.705	15,07
Botón pagos	3.937	1.138	5.075	22,42
Corresponsal no bancario	22.487	2.583	25.070	10,3
Transacciones Web	1.558	885	2.443	36,23

TOTAL	121.555	24.928	146.483	17,01
-------	---------	--------	---------	-------

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

3.4. **Discusión**

Estudios realizados en el sector bancario mediante modelos de regresión cuantílica como los desarrollados por Damane (2024); Alvarado & Pinos (2017; y, Koenker & Hallock (2001), han reportado estimaciones en los cuantiles 90 a 97 con valores de pseudo R^2 entre 0,35 y 0,57, lo que indica un nivel de bondad de ajuste entre moderado y alto.

Estos resultados son coherentes con el valor obtenido en el presente estudio pseudo R^2 igual a 0,434 para el cuantil 97 el cual evidencia una capacidad explicativa moderada del modelo aplicado al umbral superior del monto transaccional, para la detección de transacciones anómalas en canales electrónicos. (véase la siguiente Tabla 13).

Tabla 13. Comparación de valores de pseudo- R^2 reportados en estudios del sector bancario mediante regresión cuantílica

Referencia	Contexto	Cuantil	Pseudo R^2	Interpretación
Koenker & Hallock (2001)	Wage inequality	0.90	~0.35	Moderado
Alvarado & Pinos (2017)	Estimación de ingresos de la población ecuatoriana una propuesta desde la regresión cuantílica	0.5	0,46	Moderado
Fosu (2024)	Inclusión financiera (MPRA)	0.75–0.95	0.36–0.57	Moderado–alto
Trabajo de titulación	Regresión cuantílica para la detección de transacciones anómalas en canales electrónicos: estudio de caso en una cooperativa de ahorro y crédito ecuatoriana.	0,5–0,97	0,29–0,43	Moderado

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

3.5. **Conclusiones del laboratorio**

Luego de realizado el análisis estadístico descriptivo, se puede concluir que existen canales electrónicos con un marcado nivel de monto promedio transaccional superior en los canales de APP Móvil y Web Transaccional respecto al resto, esto a razón que el resto de los canales

cuentan límites operativos institucionales de monto transaccional como es el caso de los corresponsales no bancarios, ATMS, APP Móvil Pymes; y, Botón de pagos.

Se comprobó que variables como el monto transaccional, ingresos mensuales y saldo en ahorros vista no siguen una distribución normal. En consecuencia, se aplicaron pruebas no paramétricas para evaluar diferencias en la mediana de variables categóricas como sexo, nivel de instrucción, actividad económica y rango de ingresos. En todos los casos, se rechazó la hipótesis nula, lo que evidencia diferencias significativas entre grupos de estas categorías de análisis.

Para reforzar estos hallazgos, se implementó un análisis post hoc mediante la prueba de Tukey, el cual confirmó cuáles subgrupos presentaban diferencias significativas entre sí. Estos hallazgos sustentan la inclusión de dichas variables como predictores claves en los modelos de regresión cuantílica empleados para estimar tanto la mediana del monto transaccional ($\tau = 0.5$) como su umbral superior ($\tau = 0.97$), fundamentales para la detección de comportamientos atípicos en canales electrónicos.

La estimación de modelos de regresión cuantílica para $\tau = 0.5$ y $\tau = 0.97$ permitió caracterizar el comportamiento típico y extremo del monto transaccional en canales electrónicos. El modelo para la mediana ($\tau = 0.5$) mostró una estructura explicativa sólida (pseudo- $R^2 = 0.298$), mientras que el cuantil superior ($\tau = 0.97$) ofreció un ajuste superior (pseudo- $R^2 = 0.434$), capturando mejor el comportamiento transaccional en el cuantil 97 lo cual es importante ya que este es el modelo que termina el umbral superior para detección de transacciones anómalas.

Ambos modelos confirmaron que factores sociodemográficos como la edad generacional, el nivel de ingresos, el género y el nivel educativo influyen significativamente en el monto transaccional. De igual forma, variables conductuales del perfil transaccional bancario como el saldo promedio en cuenta vista, el monto de la última transacción y el tiempo entre transacciones mostraron relaciones positivas y estadísticamente significativas, destacando su relevancia como predictores del comportamiento financiero.

Particularmente, el modelo para el cuantil 0.97 evidenció un mejor ajuste (pseudo- $R^2 = 0.434$) y reveló que ciertos canales electrónicos como App Móvil PYMES, Corresponsales No Bancarios y Botón de Pagos se asocian con montos transaccionales inferiores en los percentiles más altos, lo cual sugiere diferencias estructurales en los patrones operativos de dichos canales.

Este hallazgo evidencia la presencia de un número significativo de transacciones que se desvían del comportamiento esperado, según el perfil transaccional habitual. Sin embargo,

se observa un desbalance en la distribución de clases entre transacciones normales y anómalas, lo que podría limitar el desempeño de modelos supervisados de clasificación.

Por tanto, se recomienda aplicar técnicas de balanceo de clases (como sobremuestreo, submuestreo o generación sintética de datos) previo a la construcción de modelos de segmentación o predicción, con el objetivo de mejorar la capacidad de detección de comportamientos no habituales en canales electrónicos.

4. Laboratorio: Enfoque Machine Learning

4.1 Objetivos específicos

- Identificar las variables sociodemográficas, económicas y transaccionales que presentan mayor influencia sobre el monto transaccional no habitual, mediante técnicas de análisis estadístico y aprendizaje automático.
- Mediante el establecimiento del umbral superior definido en el laboratorio 1 a través de la regresión cuantílica, definir reglas de clasificación mediante modelos supervisados de aprendizaje automatizado
- Evaluar la precisión, sensibilidad y desempeño de los modelos desarrollados, mediante métricas adecuadas para la detección de transacciones anómalas, como matriz de confusión, AUC-ROC, F1-score y tasa de falsos positivos.

4.2 Método

Se entrenaron varios modelos supervisados entre ellos Random Forest, W-Part, y R-Part por canal electrónico, tomando la etiqueta anómalo y normal predicho del modelo de regresión cuantílica con tau 97 (percentil 97), además se realizó una partición del 70% para entrenamiento y 30% para prueba sobre una base de datos balanceada con una participación del 60% de transacciones normales frente a un 40% de transacciones anómalas con la finalidad de tener clases balanceadas en cada canal electrónico.

4.3 Discusión y resultados

Los experimentos muestran desempeños elevados del clasificador cuando se entrena y evalúa por canal, con especial rendimiento en App Móvil. En el presente trabajo se reportan matrices de confusión, métricas clásicas (exactitud, sensibilidad, especificidad) y AUC-ROC, junto con intervalos de confianza, lo que aporta una base estadística adecuada para comparar modelos entre canales.

Entre los principales hallazgos se tiene que el mejor desempeño se observó en el canal App Móvil, con un *Accuracy* de 84,93%, *F1 Score* de 82,00% y un *AUC* de 91,32%, destacando su alta capacidad predictiva para transacciones anómalas.

El canal de Corresponsales No Bancarios presentó también resultados sólidos (*F1 Score* = 74,10%, *AUC* = 85,84%), seguido por Transacciones Web y Botón de Pagos, con *F1 Scores* de 73,40% y 66,72%, respectivamente.

Los canales ATMs (*F1 Score* = 64,70%) y App Móvil – PYMES (*F1 Score* = 54,40%) reflejaron un desempeño inferior, especialmente en sensibilidad, lo que indica menor capacidad del modelo para detectar correctamente las transacciones anómalas en estos entornos. No obstante, estas métricas pueden considerarse aceptables para fines de monitoreo transaccional continuo, particularmente como filtros preliminares. En este contexto, el modelo Random Forest podría complementarse para estos canales con otros enfoques basados en modelos de aprendizaje automatizado. La siguiente Tabla 14 presenta el resumen de las métricas de desempeño del modelo Random Forest por tipo de canal, donde la clase positiva fue “anómalo”.

Tabla 14. Resumen de desempeño del modelo Random Forest por canal virtual

Métricas ³	App móvil	Transacciones Web	Atms	Corresponsales no bancarios	Botón de pagos	App móvil - pymes
Accuracy (%)	83,93%	78,31	72,60%	79,10%	74,59%	65,79%
	83,40% -	74,90% -	71,6% -	77,22% -	71,53% -	64,43% -
95% IC	86,30%	81,40%	73,49%	80,89%	77,48%	67,13%
Sensibilidad o recall (%)	83,40%	72,10%	64,06%	75,06%	64,12%	50,96%
Especificidad (%)	86,00%	82,73%	78,00%	81,86%	81,52%	75,75%
Precisión (positivo) (%)	80,60%	74,81%	65,40%	69,81%	69,70%	58,52%
F1score (%)	82,00%	73,40%	64,70%	74,10%	66,72%	54,40%
Coeficiente Kappa (%)	69,02%	55,12%	42,00%	56,77%	46,26%	27,29%
Área bajo la curva ROC (AUC) (%)	91,32%	77,40%	79,02%	85,84%	79,81%	71,41%

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

³ Las métricas se obtuvieron a partir de modelos Random Forest entrenados sobre muestras balanceadas (60% normal, 40% anómalo) por canal, con un entrenamiento 70% y 30% Para cada modelo, se aplicó una partición del 70% para entrenamiento y 30% para prueba. La clase positiva fue definida como transacción anómala.

Asimismo, mediante el análisis de importancia de variables del modelo Random Forest reveló qué, en la mayoría de los canales electrónicos, las principales características predictoras para discriminar entre transacciones anómalas y normales fueron las siguientes, ordenadas según contribución al índice de Gini:

- Monto acumulado
- Monto última transacción
- Frecuencia transaccional
- Saldo en ahorros vista promedio
- Ingresos/actividad del cliente

Para mayor detalle sobre las métricas asociadas a la importancia de variables predictoras en cada modelo Random Forest, así como las gráficas correspondientes a las curvas ROC y los valores de AUC por canal electrónico, revisar el apartado de Anexos.

El modelo W PART, aplicado sobre muestras balanceadas por canal, mostró un desempeño adecuado en la detección de transacciones anómalas, aunque inferior al observado con Random Forest. El mejor rendimiento se obtuvo en el canal App Móvil, con un F1 Score de 80,73%, sensibilidad del 85,41% y accuracy del 83,49%, evidenciando buena capacidad predictiva.

Canales como Transacciones Web, Botón de Pagos y Corresponsales No Bancarios mantuvieron métricas consistentes, con F1 Scores entre 68% y 70%, adecuados para monitoreo transaccional. En contraste, App Móvil – PYMES y ATMs presentaron los desempeños más bajos (F1 Score de 62,31% y 64,24%, respectivamente), con menor capacidad de detección y discriminación. A pesar de su menor precisión global, W PART destaca por su interpretabilidad, siendo útil en entornos donde se requiere trazabilidad y explicación de decisiones mediante la implementación de reglas de clasificación, como se aprecia en la siguiente Tabla 15.

Tabla 15. Resumen de desempeño del modelo W PART por canal virtual

Métricas	App móvil	Transacciones Web	Atms	Corresponsales no bancarios	Botón de pagos	App móvil pymes
Accuracy	83,49%+/- 1,10	75,46% +/- 3,61	69,86% +/- 0,82	75,25% +/- 1,39	71,74% +/- 3,22	66,97% +/- 1,11
Sensibilidad (%)	85,41%	72,31%	67,64%	72,44%	77,59%	68,39%
Especificidad (%)	82,38%	77,56%	71,34%	77,13%	67,84%	66,03%
Precisión (%)	76,63%	68,23%	61,14%	67,86%	61,66%	57,30%
F1 score	80,73%	70,22%	64,24%	70,07%	68,77%	62,31
Coeficiente Kappa	66,40%	47,00%	38,00%	43,00%	41,00%	33,00%

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Mediante validación cruzada, se obtuvo un total de 116 reglas de decisión para el canal Web Transaccional. A continuación, se presentan las cinco reglas más relevantes por clase según cobertura y especificidad (véase la siguiente Tabla 16)

Tabla 16. Reglas de clasificación para canal Web Transaccional por tipo de clase

Tipo de clase	Detalle	Comentario técnico
ANÓMALO	· Clientes de sexo femenino con monto transaccional acumulado del mes alto > USD 6.315; Saldo en ahorros vista alto > USD 1.961; pero última transacción realizada baja.	Clientes mujeres con monto transaccional alto acumulado en el mes, pero micro transacciones recientes.
ANÓMALO	· Nivel de instrucción primaria Saldo en ahorros vista alto > USD 3.264; monto transaccional acumulado del mes alto > 2.447; y monto de última transacción <= USD 501	Clientes con instrucción primaria, pero con saldo en ahorros vista alto y monto transaccional del mes alto
ANÓMALO	· Clientes que se dedican a actividades profesionales, científicas y técnicas (A7), de educación universitaria, monto última transacción <= USD 328, y saldo en ahorros vista > USD 3.368	Precisión perfecta; perfil instruido con micro transacciones atípicas
ANÓMALO	Clientes que se dedican al Comercio al por mayor y menor (A12), con ingresos altos, educación universitaria, monto acumulado al mes > USD 5.857; y, monto última transacción <= 380	Usuarios independientes, con estudio superior e ingresos altos
ANÓMALO	Clientes que se dedican a actividades profesionales, científicas y técnicas (A7), edad generacional Millennials, monto acumulado mensual > USD 2.894	Segmento joven con acumulación transaccional elevada.
NORMAL	Saldo ahorros vista < = USD 1.182	Bajo saldo refleja patrón transaccional habitual.
NORMAL	Monto acumulado mensual <= USD 1031,57	Usuarios con poca actividad acumulada presentan bajo riesgo
NORMAL	Clientes que se dedican a la actividad de transporte y almacenamiento	Este tipo de actividad se comporta consistentemente normal
NORMAL	Frecuencia transaccional diaria > 3, ingresos medios, saldo ahorros vista < USD 7.657; monto última transacción > 69	Actividad transaccional sostenida y moderada, sin señales de alarma.
NORMAL	Clientes que se dedican al Comercio al por mayor y menor (A12), de edad generacional Millennials	Jóvenes con patrones web estables.

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Para el canal APP MÓVIL se obtuvo un total de 316 reglas de decisión. A continuación, se presentan las cinco reglas más relevantes por clase y nivel de precisión más alta (véase la siguiente Tabla 17).

Tabla 17. Reglas de clasificación para canal APP Móvil por tipo de clase

Tipo de clase	Reglas:	Cobertura (n / errores)	Comentario técnico
ANOMALO	Monto acumulado > 1904 AND saldo vista > 2082.86 AND monto_ult_trax <= 896.75 AND frecuencia trans > 22.44 AND Actividad cliente = Construcción (A13)	54 / 5	Regla de alta cobertura y precisión. Acumulación media-alta, frecuencia moderada, cliente específico.
ANOMALO	Monto acumulado > 2033 AND saldo vista > 1981.71 AND monto_ult_trax <= 896.75 AND frecuencia trans > 22.44 AND Actividad cliente = Agricultura, ganadería, silvicultura, y pesca (A9)	65 / 7	Perfil típico de anomalía: acumulación alta, frecuencia estable y tipo de actividad definida.
ANOMALO	Monto acumulado > 2033 AND saldo vista > 1981.71 AND monto_ult_trax <= 1400 AND Actividad cliente = Actividades atención en la salud humana (A2) AND frecuencia trans > 22.44	45 / 4	Buen poder predictivo en usuarios de frecuencia media con montos controlados.
ANOMALO	Monto acumulado > 2998 AND Nro.max_trax_dia <= 10 AND nivel estudio = UNIVERSITARIA AND Actividad cliente = Construcción (A13)	28/8	Identifica perfil universitario de clase media-alta con actividad no típica.
ANOMALO	Frecuencia trans > 110.8 AND monto acumulado > 1180 AND monto_ult_trax <= USD 216.35 AND Actividad cliente = A16	22 / 0	Regla perfecta (sin errores) . Alta frecuencia con montos bajos en usuarios con historial elevado.
NORMAL	Saldo vista <= USD 305.145	1175 / 1	Regla dominante y casi perfecta . Bajo saldo es patrón fuertemente normal.
NORMAL	Monto acumulado <= 503.25 AND rango ingresos = c. USD 471<Ing.< USD 940	279 / 0	Regla totalmente precisa . Bajo nivel de ingresos y acumulado identifica clientes regulares.
NORMAL	Monto acumulado <= USD 308	274 / 0	Otra regla perfecta: micro transacciones de bajo volumen como comportamiento normal.
NORMAL	Monto acumulado <= 997.18 AND Monto_ult_trax > 88	107 / 1	Actividad básica con montos individuales moderados. Alta confiabilidad.
NORMAL	Saldo vista <= 1023.43 AND Actividad cliente = Ama de casa A10 AND frecuencia trans <= 88.33	32 / 1	Perfil de saldo bajo y actividad limitada en un tipo de cliente común.

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Para los canales Corresponsales no bancarios, ATMS, APP Móvil Pymes; y, Botón de pagos se realizó en análisis de reglas de clasificación mediante el algoritmo R- Part, esto a razón de que era más interpretable sus resultados para revisión del árbol ir a la parte de los anexos. A continuación, los principales resultados.

CORRESPONSALES NO BANCARIOS las reglas principales son:

- Cuando se presentan combinaciones de monto acumulado > USD 132, monto última transacción < USD 113, saldo vista > USD 101, monto última transacción < USD 58; y, saldo vista > USD 5.225 (Clientes con alto monto acumulado, transacciones recientes pequeñas y saldos intermedios tienden a presentar comportamiento anómalo).
- Cuando se presentan combinaciones de monto acumulado > USD 132, monto última transacción < USD 113, saldo vista < USD 101.
- Cuando se presentan combinaciones de monto acumulado > USD 132, monto última transacción < USD 113, saldo vista > USD 101, monto última transacción < USD 58; y, monto acumulado > USD 256 (A pesar de un monto de transacción algo mayor (≥ 58), si la acumulación es muy alta, se activa alerta).
- Cuando se presentan combinaciones de monto acumulado > USD 132, monto última transacción < USD 113, saldo vista > USD 101, monto última transacción < USD 58; y, saldo vista > USD 5.252 (Casos muy específicos donde el saldo en ahorros vista es demasiado alto y las transacciones pequeñas).

Para ATMS las principales reglas de clasificación son:

- Cuando se presentan combinaciones de monto acumulado > USD 208, monto última transacción < USD 103, saldo vista > USD 153, monto última transacción < USD 78. (Cuando el cliente acumula montos altos, realiza una transacción pequeña (<78) y tiene un saldo moderadamente alto, es más probable que sea una transacción anómala).
- Cuando se presentan combinaciones de monto acumulado > USD 208, monto última transacción < USD 103, saldo vista > USD 153, monto última transacción < USD 78; y, monto acumulado > USD 393 (Incluso con una transacción reciente de tamaño medio, si el monto acumulado es muy alto, hay probabilidad de anómalo).
- Cuando se presentan combinaciones de monto acumulado > USD 208, monto última transacción < USD 103, saldo vista > USD 153, monto última transacción < USD 78; y, monto acumulado > USD 393 (patrón claro de anomalía aparece cuando los clientes tienen altos saldos acumulados y realizan transacciones pequeñas).

Finalmente, tanto para los canales BOTÓN DE PAGOS, y APP MÓVIL PYMES se obtuvieron las mismas reglas de clasificación mediante el algoritmo R-Part aplicado, las cuales fueron:

- Cuando se presentan combinaciones de monto acumulado > USD 49, ingresos < USD 1.982, monto última transacción < USD 27, saldo vista < USD 4.292; y, monto acumulado > USD 121
- Cuando se presenta monto acumulado < USD 49 es una transacción normal

En conjunto, la estrategia de etiquetado cuantílico permitió construir bandas de normalidad personalizadas por canal, superando los supuestos de normalidad y homocedasticidad que limitan a los modelos lineales, y habilitando un clasificador supervisado con alto rendimiento y buena interpretabilidad (importancias de Random Forest, reglas PART). No obstante, para una evaluación más fiel es recomendable reportar AUC-PR, estimar intervalos de confianza del AUC (DeLong/*bootstrap*), y evitar fuga de objetivo generando las etiquetas cuantílicas exclusivamente en los pliegues de entrenamiento. Finalmente, de acuerdo con Liu et al. (2008) es importante incorporar detectores no supervisados como Isolation Forest/LOF lo cual permitirá ampliar la cobertura frente a patrones emergentes y facilitará la vigilancia de concept drift en operación.

4.4 Conclusiones del laboratorio

Las técnicas desarrolladas mediante modelos de aprendizaje supervisado como Random forest, W Part; y, Rpart dan cuenta de que el nivel de precisión en la detección de transacciones anómalas es robusto y alto para el canal APP Móvil cuya métrica de presión F1 score fue del 82% el porcentaje de predicción correcta bajo la curva ROC fue del 91,31

De otro lado para canales electrónicos como Corresponsales no bancarios, Web transaccional; y, Botón de pagos obtuvieron valores de métricas de precisión sólidos con un valor promedio del F1 score de un promedio del 70% para el modelo Random Forest; y, un ligeramente más bajo en las métricas obtenidas mediante el algoritmo W Part.

También se ha visto que existe una baja capacidad predictiva de los casos de transacciones anómalas (clase positiva) para los canales ATMS y APP Móvil Pymes con una métrica del F1 score promedio del 60% para ambos canales, sin embargo, se recomienda que es un modelo suficiente para monitoreo que se puede adoptar con reglas de decisión obtenidas de técnicas como W Part o Rpart que complementarían el seguimiento oportuno de transacciones correctamente identificadas como anómalas.

Las reglas principales obtenidas para los canales APP Móvil y Web transaccional mediante al algoritmo W Part dan cuenta que las transacciones anómalas van en función principalmente de la combinación de varias variables como es el monto transaccional acumulado del mes el

cual oscila en un promedio superior que va entre los USD 2.000 a USD 6.315, cuando provienen de sexo femenino, nivel de instrucción primaria, frecuencia transaccional rápida es decir que no han pasado en promedio más de 22 horas entre transacción, que se dediquen a actividades como profesionales científicos y técnicos, salud humana, construcción; cuando disponen de saldo promedio de ahorros vista elevados es decir muy por encima del promedio entre USD 800 a USD 3.400; también es fuerte el comportamiento de la última transacción realizada el cual supera los USD 300.

Los modelos RPART aplicados por canal permitieron identificar reglas claras asociadas a comportamientos anómalos. En Corresponsales No Bancarios y ATMs, las anomalías se caracterizan por montos acumulados altos, combinados con transacciones recientes pequeñas y saldos vista elevados, lo que indica posibles operaciones fragmentadas o inusuales frente a su capacidad financiera.

En App Móvil PYMES y Botón de Pagos, se detectaron como anómalas las combinaciones de monto acumulado elevado con ingresos bajos y últimas transacciones pequeñas, mientras que acumulaciones bajas (< 49 USD) se asocian consistentemente a comportamiento normal. Estos resultados confirman que la anomalía transaccional responde a interacciones no lineales entre variables financieras y sociodemográficas, y que los árboles de decisión permiten capturar dichas reglas de forma interpretable y eficaz.

Estos hallazgos evidencian que los comportamientos anómalos no se explican únicamente por montos absolutos, sino por combinaciones específicas entre acumulación previa, monto reciente, saldo disponible e ingresos del cliente, lo que justifica el uso de técnicas no lineales y orientadas a reglas como RPART para complementar modelos predictivos tradicionales en la vigilancia transaccional.

5. Laboratorio: Enfoque Toma de Decisiones

5.1. Objetivos específicos:

- Identificar a través del modelo de agrupamiento K medias el número óptimo de clusters, que permita segmentar de manera eficiente los perfiles sociodemográficos y de comportamiento transaccional de los clientes que presentan comportamientos no habituales (anómalos) en el monto transaccional por canal electrónico, con el propósito de generar información estratégica para el diseño de controles orientadas a la mitigación de riesgos de fraude.
- Evaluar el desempeño del modelo de clasificación supervisado KNN para la detección de transacciones anómalas y normales por tipo de canal electrónico, identificando el número óptimo de vecinos (K) que maximice su precisión.

5.2. Métodos:

Para el cumplimiento del objetivo se adoptó un enfoque híbrido, que combina técnicas de aprendizaje no supervisado y supervisado, con el propósito de segmentar y clasificar clientes que presentan comportamientos atípicos en sus montos transaccionales por canal electrónico.

En la primera fase, se aplicó el algoritmo de agrupamiento K-means para identificar clústeres homogéneos basados en variables sociodemográficas y transaccionales. El número óptimo de grupos se determinó mediante la evaluación de la suma de cuadrados intra-cluster (WSS), el método del codo y la métrica de silueta.

Esta segmentación permitió establecer perfiles de comportamiento diferenciados que sirvieron como insumo para la fase supervisada. Las variables numéricas fueron previamente escaladas y se incluyeron las siguientes:

- X1: Monto
- X2: Edad
- X3: Ingresos
- X4: Saldo promedio ahorros vista
- X5: Monto última transacción Y_{t-1}
- X6: Tiempo promedio en horas entre transacciones
- X7: Monto acumulado
- X8: Número máximo de transacciones diarias

Posteriormente, se utilizó el algoritmo K-Nearest Neighbors (KNN) binario sobre el conjunto de transacciones (Normales y Anómalas), para clasificar nuevas observaciones y validar la coherencia de los clústeres definidos. Para este modelo, la base de datos fue la misma que se utilizó en el laboratorio 2 es decir con la base de datos balanceada. Se fijó “Anómalo” como clase positiva. El número de vecinos (k) se optimizó mediante validación cruzada, priorizando métricas de precisión, sensibilidad y F1 Score para maximizar la detección de transacciones anómalas.

5.3. Discusión y resultados

La segmentación obtenida mediante K-Medias mostró una alta coherencia interna en los canales APP Móvil y Web, en concordancia con estudios previos que destacan la eficacia de esta técnica para identificar patrones atípicos en transacciones electrónicas. Sin embargo, la baja cohesión observada en otros canales refleja la heterogeneidad del comportamiento financiero, como también lo evidencian análisis comparativos de métodos no supervisados,

por su parte, el modelo KNN alcanzó buen equilibrio entre precisión y sensibilidad, coincidiendo con investigaciones que confirman su efectividad en detección de fraudes financieros, aunque con menor robustez frente a Random Forest.

Entre los principales hallazgos tenemos que el modelo de agrupamiento K-means determinó que la segmentación óptima corresponde a dos clústeres para la mayoría de los canales electrónicos y tres clústeres para ATMs. El análisis intra-clúster permitió identificar segmentos con mayor presencia de comportamientos anómalos y perfilar a dichos clientes según canal.

En Web Transaccional y APP Móvil, los usuarios de estos segmentos presentan montos promedio por transacción elevados (USD 8.876,58 y USD 3.421,23, respectivamente), saldos en ahorro vista superiores a USD 23.000, y montos acumulados mensuales mayores a USD 53.000, con una alta frecuencia operativa (intervalos de 34–42 horas y hasta 8 y 3 transacciones diarias). Predominan Millennials, con ingresos > USD 2.350, principalmente hombres con educación secundaria o universitaria, vinculados a comercio, construcción, servicios financieros y manufactura (véanse las siguientes Tablas 18 y 19 respectivamente

Tabla 18. Perfilamiento transaccional en canal Web Transaccional

Clúster	Monto promedio	Saldo promedio Vista	Monto promedio última transacción	Media del monto acumulado mes	Nro. máximo de transacciones diarias	Ingresos promedio del cliente	Edad promedio del cliente	Tiempo en horas entre transacciones	% de transacciones anómalas
1 normal	787,01	4.972,12	274,45	5.785,66	4	4.062,97	42	66,27	39,16
2 anómalo	12.258,48	50.413,37	4.590,49	55.224,84	3	278.717,74	44	49,34	89,09

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Tabla 19. Perfilamiento transaccional en canal APP Móvil

Clúster	Monto promedio transaccional	Saldo promedio en Ahorros Vista	Monto promedio última transacción	Media del monto acumulado mes	Nro. máximo de transacciones diarias	Ingresos promedio del cliente	Edad promedio del cliente	Tiempo en horas entre transacciones	% de transacciones anómalas
1 anómalo	3.421,23	23.828,74	1.178,52	53.203,52	8	25.325,97	42,55	33,94	73,57
2 normal	411,24	3.023,97	150,43	3.644,48	3	2.081,66	35,93	63,01	38,59

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

En otros canales, los perfiles presentan menor intensidad financiera. En ATMs, destacan saldos en ahorro de USD 4.346,68, montos acumulados mensuales de USD 1.690,45 y frecuencias de 70 horas, con Millennials (49,9%) y Generación X (31,3%), hombres (56,5%), ingresos de USD 471–940 y educación secundaria (véase la siguiente Tabla 20).

Tabla 20. Perfilamiento transaccional en canal ATMS

Clúster	Monto promedio	Saldo promedio Vista	Monto promedio última transacción	Media del monto acumulado o mes	Nro. máximo de transacciones diarias	Ingresos promedio del cliente	Edad promedio del cliente	Tiempo en horas entre transacciones	Media umbral superior	% de transacciones anómalas
1 anómalo	283,48	4.346,68	158,13	1.690,45	2	1.930,08	42	70,38	285,22	61,2
2 normal	99,19	1.220,04	57,32	382,24	1	866,34	39	131,84	147,62	35,74
3 normal	77,25	950,32	42,94	653,96	2	812,13	36	51,62	118,26	36,06

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Los Corresponsales No Bancarios agrupan clientes de baja actividad (1 transacción diaria), montos acumulados bajos (USD 667,65) y alta dispersión temporal (112 horas), predominando Millennials (51,3%), hombres (60%) y educación primaria. Botón de Pagos muestra un perfil más balanceado, con mayoría femenina (57%), 3 transacciones diarias, Millennials (53,4%), ingresos de USD 471–940 y educación secundaria (véase la siguiente Tabla 21)

Tabla 21. Perfilamiento transaccional en canal Corresponsales No Bancarios

Clúster	Monto promedio	Saldo promedio Vista	Monto promedio última transacción	Media del monto acumulado o mes	Nro. máximo de transacciones diarias	Ingresos promedio del cliente	Edad promedio del cliente	Tiempo en horas entre transacciones	Media umbral superior	% de transacciones anómalas
1 normal	50,58	741,90	32,80	256,76	1,4	700,04	35	120,38	90,88	29,95
2 anómalo	150,51	1.891,55	88,53	667,65	1,3	1.186,93	41	112,42	147,47	58,98

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

En el canal APP Móvil PYMES, aunque el monto promedio por transacción es bajo (USD 43,64), destaca por la mayor frecuencia operativa (24 horas entre transacciones y hasta 6 diarias), con Millennials (67,8%), mayoría femenina (56,7%), ingresos de USD 471–940, y actividad en empleo privado, comercio y servicios de alojamiento (véase la Tabla 22)

Tabla 22. Perfilamiento transaccional en canal APP MÓVIL PYMES

Clúster	Monto promedio	Saldo promedio Vista	Monto promedio o última transacción	Media del monto acumulado mes	Nro. máximo de transacciones diarias	Ingresos promedio del cliente	Edad promedio del cliente	Tiempo en horas entre transacciones	Media umbral superior	% de transacciones anómalas
1 normal	17,44	813,23	9,19	160,48	3	946,46	31	56,85	34,23	37,53
2 anómalo	43,64	2.349,13	19,48	569,49	6	1.720,41	35	24,58	64,54	45,58

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

La siguiente Tabla 23 presenta la segmentación de los clientes que se encuentran dentro de los clusters con mayor presencia de transacciones inusuales o anómalas por canal electrónico.

Tabla 23. Perfilamiento sociodemográfico clúster anómalo por canal electrónico

App móvil	Web transaccional	Atms	Corresponsales no bancarios	Botón de pagos	App móvil pymes
Sexo: Masculino 55,2%	Sexo: Masculino 60%	Sexo: Masculino 56,5%	Sexo: Masculino 60%	Sexo: Femenino 57%	Sexo: Femenino 56,7%
Edad generacional: Millenials: 57,9%	Edad generacional: Millenials: 52,7% y Generación X 30,9%	Edad generacional: : Millenials: 49,9% y Generación X 31,3%	Edad generacional: Millenials: 51,3% y Generación X 26,2%	Edad generacional: Millenials: 53,4% y Generación X 24%	Edad generacional: Millenials: 67,80%
Rango de ingresos del cliente: > = USD 2,350	Rango de ingresos del cliente: > = USD 2,350	Rango de ingresos del cliente: > = USD 471 a \$940	Rango de ingresos del cliente: > = USD 471 a \$940	Rango de ingresos del cliente: > = USD 471 a \$940	Rango de ingresos del cliente: > = USD 471 a \$940

Actividad cliente: Comercio al por mayor y menor, Empleado privado, Construcción, enseñanza, Actividades financieras, Actividades inmobiliarias	Actividad cliente: Comercio al por mayor y menor, Empleado privado, Agricultura y ganadería; y, Industrias Manufactureras	Actividad cliente: Empleado privado, Empleado Público, Agricultura y ganadería; y, Comercio al por mayor y menor	Actividad cliente: Empleado privado, Empleado Público, Agricultura y ganadería; y, Comercio al por mayor y menor	Actividad cliente: Empleado privado, Empleado Público, Comercio al por mayor y menor; y, Amas de casa	Actividad cliente: Empleado privado, Empleado Público, Comercio al por mayor y menor; y, Actividades de alojamiento y servicios de comida
Nivel de estudio: Secundaria	Nivel de estudio: Universitaria	Nivel de estudio: Secundaria	Nivel de estudio: Primaria	Nivel de estudio: Secundaria	Nivel de estudio: Secundaria

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

A continuación, se presenta los principales resultados de validación del modelo K medias, los cuales evidencian una buena segmentación interna en APP MÓVIL y WEB (> 0,70 silueta). Los demás canales muestran baja cohesión (< 0,40), aunque la pureza es moderada, destacando BOTÓN DE PAGOS (0,71). La siguiente Tabla 24 presenta los resultados de validación.

Tabla 24. Validación de la técnica K Medias

Métricas:	App móvil	Transacciones Web	Atms	Corresponsales no bancarios	Botón de pagos	App móvil - pymes
Coefficiente Silueta	0,71	0,72	0,16	0,22	0,30	0,27
Índice de pureza	0,62	0,61	0,64	0,66	0,71	0,60

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Modelo supervisado KNN, se optimizó el número de vecino k , por canal mediante validación cruzada, modelando la variable etiqueta “Anómalo” como clase positiva. En el conjunto de prueba se obtuvo un área bajo la curva (AUC) entre 0,77–0,82 y precisiones (accuracy) de 68,4 – 81,80% según canal electrónico, priorizando sensibilidad para reducir falsos negativos. Los mejores desempeños se observaron en Web Transaccional ($k=7$) y Corresponsales No Bancarios ($k=15$) con AUC=0,82 y sensibilidades de 58,7% y 64,8%, respectivamente. En

ATMs (k=20) y Botón de Pagos (k=14), el modelo obtuvo AUC de 0,79 y 0,77, y precisiones (accuracy) del 71–72% y sensibilidades en torno al 57–59%. Para APP Móvil (k=12) y APP Móvil PYMES (k=11), el desempeño fue AUC de 0,77–0,78, accuracy del 71% y sensibilidades de 55–62%. Estos resultados confirman que KNN, ajustado por canal, logra un equilibrio entre precisión y detección de anomalías, aunque su rendimiento es inferior al observado en modelos más robustos como Random Forest. La siguiente Tabla 25 presenta las métricas de desempeño por canal electrónico. Para más información ir a la parte de anexos.

Tabla 25. Métricas de desempeño del modelo KNN, clase positiva anómalos

Métricas:	App móvil	Transacciones Web	Atms	Corresponsales no bancarios	Botón de pagos	App móvil - pymes
Accuracy (%)	67,13%	76,24%	71,93%	74,23%	70,33%	71,17%
95% IC	66,03% - 68,22%	73,46% - 78,87%	71,18% - 72,67%	72,61% - 75,81%	67,98% - 72,61%	70,23% - 71,10%
Sensibilidad recall (%)	61,78%	59,95%	59,08%	65,53%	56,01%	62,38%
Especificidad (%)	70,70%	87,02%	80,52%	80,06%	79,91%	77,19%
Precisión positivos (%)	59,01%	75,32%	66,99%	68,75%	65,09%	65,21%
F1score (%)	60,30%	66,76%	62,78%	67,11%	60,20%	63,76%
Coeficiente Kappa (%)	32,31%	48,65%	40,39%	45,94%	36,77%	39,85%
Área bajo la curva ROC (AUC) (%)	68,40%	81,80%	78,60%	80,80%	76,70%	77,90%

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

5.4. Conclusiones del laboratorio

El análisis de segmentación mediante K-means reveló patrones diferenciados en clientes con comportamientos transaccionales atípicos, destacando que la mayoría de los canales se segmentan en dos clústeres y ATMs en tres. Los canales Web Transaccional y APP Móvil concentran perfiles de alto volumen financiero, con montos por transacción y acumulados mensuales significativamente superiores, y mayor frecuencia operativa, predominando Millennials con ingresos altos y educación secundaria o universitaria. En contraste, ATMs, Corresponsales y Botón de Pagos presentan perfiles de actividad más moderada o baja, con ingresos medios y predominancia de Millennials.

La validación del modelo K-Medias evidenció una segmentación adecuada en los canales digitales, con buena cohesión interna y coherencia moderada respecto a las clases reales, confirmando su utilidad para la clasificación inicial.

El modelo KNN, ajustado por canal, alcanzó un área bajo la curva (AUC) de 0,77–0,82 y precisiones (accuracy) de 70–76%, priorizando la detección de anómalos

Estos resultados sugieren que Random Forest es el modelo más robusto para clasificación, el cual llegó a obtener un AUC de hasta 91%, mientras que KNN puede servir como herramienta complementaria por su simplicidad y adaptabilidad a nuevos datos.

6. Laboratorio: Enfoque Robustez y validación

6.1. Objetivos específicos:

- Validar los supuestos y la robustez y validación de los modelos desarrollados en los diferentes laboratorios, precisando que supuestos corresponden a cada enfoque y la pertinencia de su uso en el contexto para la detección de transacciones anómalas en canales virtuales.

6.2. Método

Se estimaron dos modelos de regresión cuantílica: $\tau = 50$ (mediana del monto transaccional), y $\tau = 97$ (umbral superior para detección de transacciones anómalas).

La validación de los modelos se realizó a partir del análisis de los residuos, evaluando los supuestos de ortogonalidad, linealidad, multicolinealidad, e independencia.

Cabe precisar que, a diferencia de la regresión por MCO, en la regresión cuantílica no se reporta el MSE, sino la pérdida de cuantil. No obstante, de forma complementaria se aplicaron pruebas de normalidad, homocedasticidad, y determinación (pseudo- R^2) como diagnósticos complementarios, aun cuando no son requerimientos estrictos en regresión cuantílica.

La significancia conjunta de los predictores se evaluó mediante ANOVA cuantílico (test de Wald), comparando modelo completo vs el modelo nulo.

En el segundo laboratorio, se implementaron modelos supervisados (árbol de decisión, Random Forest, y W-Part) y no supervisados (K Medias), las validaciones se basan en métricas de desempeño predictivo para modelos supervisados, y índices de calidad de segmentación para K Medias. Las métricas de validación de cada modelo son:

En el segundo laboratorio, se implementaron modelos supervisados (árbol de decisión, Random Forest, W-Part; y, R - Part.), las validaciones se basan en métricas de desempeño predictivo para modelos supervisados, y índices de calidad de segmentación para K Medias. Las métricas de desempeño y validación de cada modelo son:

- Random Forest: área bajo la curva ROC (AUC), F1-score, Coeficiente Kappa, Accuracy, Sensibilidad, Especificidad, Precisión
- W-Part: Accuracy, F1-score, Coeficiente Kappa, Sensibilidad, Especificidad, Precisión

Mediante machine learning se aplicó el modelo no supervisado K Medias para segmentación y el modelo supervisado KNN para medir la aproximación entre vecinos más cercanos de acuerdo con la etiqueta anómalo (clase positiva) y transacción normal. Las métricas de desempeño y validación de cada modelo son:

- K medias: Coeficiente Silueta, y el índice de pureza
- KNN: Accuracy, Sensibilidad, Especificidad, Precisión, F1-score, Coeficiente Kappa, área bajo la curva ROC(AUC)

Finalmente, se procede a la validación de los modelos y la robustez de estos modelos desarrollados en los tres laboratorios.

6.3. Discusión y resultados

Los resultados confirman la validez técnica de los modelos aplicados, evidenciando que la regresión cuantílica identifica adecuadamente los cuantiles críticos del comportamiento transaccional, incluso bajo supuestos no clásicos.

La técnica K-Medias mostró buena segmentación interna en APP Móvil y Web, resultados consistentes con investigaciones recientes que destacan su utilidad para la detección no supervisada de fraudes financieros. Asimismo, el modelo KNN logró un equilibrio entre precisión y sensibilidad, confirmando lo reportado según Gu et la. (2019) respecto a su efectividad en clasificación supervisada, aunque con menor robustez frente a Random Forest.

Los resultados obtenidos del modelo de regresión cuantílica evidencian que en ambos cuantiles los predictores son altamente significativos como se aprecia en la Tabla 26.

Tabla 26. Métricas de validación de los dos modelos de regresión cuantílica

Cuantil (τ)	F value	p-value	Conclusión
0.50 (Mediana)	4032.9	< 0.001	Predictores altamente significativos para la mediana.
0.97 (Umbral superior)	540.38	< 0.001	Predictores significativos para la detección de anomalías.

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

En cuanto a los supuestos, los diagnósticos sobre los residuos en la regresión cuantílica $\tau = 50$, y $\tau = 97$ mostraron no normalidad y heterocedasticidad, y posibles no linealidades en los

parámetros, mientras que la independencia y multicolinealidad fueron aceptables. La siguiente Tabla 27 presenta los resultados de cumplimiento de los supuestos analizados.

Tabla 27. Métricas de validación de los modelos de regresión cuantílica

Supuestos	Prueba:	$\tau=0.50$	$\tau=0.97$	$\tau=0.50$	$\tau=0.97$	$\tau=0.50$	$\tau=0.97$
		Estadístico		P-value		Cumple supuesto	
Normalidad	Jarque_Bera	X-squared = 1.6542e+10	A = 20793	2.2e-16	2.2e-16	NO	NO
	Anderson Darling	A = 30342	A = 20793	2.2e-16	2.2e-16	NO	NO
	Kolmogorov Smirnov	D = 0.34216	D = 0.31763	2.2e-16	2.2e-16	NO	NO
Homocedasticidad	Breusch–Pagan	BP = 8141.6	BP = 7915.1	2.2e-16	2.2e-16	NO	NO
	Levene	Df = 5	Df = 5	2.2e-16	2.2e-16	NO	NO
Ortogonalidad	Durbin y Watson	DW = 1.986	DW = 1.9866	0,0069	0,009	SI	SI
Linealidad	Reset test	RESET = 714.98	RESET = 155.65	2.2e-16	2.2e-16	NO	NO
Determinación	pseudo-R ²	0,44	0,3				

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

El contraste tipo Wald confirmo que el modelo del cuantil 97 completo presenta un ajuste estadísticamente significativo frente al modelo nulo ($F = 540.38$; $p < 0.001$), lo que demuestra que, de forma conjunta, las variables explicativas contribuyen significativamente a la variación del monto transaccional en el cuantil analizado.

Modelo nulo: solo intercepto (sin predictores)

$$Mediana = \beta_0$$

Modelo completo; incluye todos los predictores

$$Mediana = \beta_0 + \beta_0 \chi_1 + D_1 \chi_2 + D_1 \chi_2 + D_2 \chi_3 + D_3 \chi_4 + \beta_2 \chi_5 + \beta_3 \chi_6 + \beta_4 \chi_7 + \beta_5 \chi_8$$

Donde:

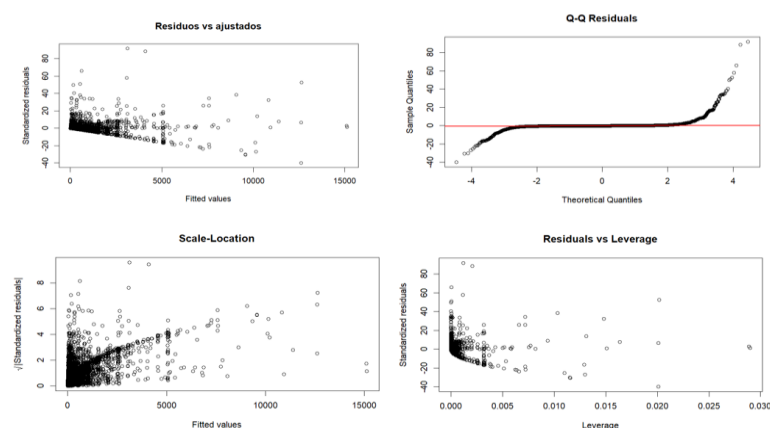
$$H_0 = \beta_1 = \beta_2 = \dots \beta_k = 0$$

$$H_1 = \text{Al menos una } \beta_j \neq 0$$

Al ser el valor p-value $< 2.2e-16$ se rechaza la hipótesis nula.

De acuerdo a la Figura 12 de diagnóstico se podría concluir que los residuos no siguen una forma matemática clara en ambos modelos de regresión cuantílica, más bien muestran heterocedasticidad y colas pesadas, lo que confirma que los supuestos clásicos de normalidad y homocedasticidad no se cumplen confirmado también por el gráfico QQ y lo desarrollado en el laboratorio 1.

Figura 12. Gráficos de diagnóstico de validación de los modelos de regresión cuantílica



Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

De otro lado, a nivel general se aprecia que el modelo de aprendizaje supervisado Random Forest cuenta con las mejores métricas de validación en cuanto a los modelos supervisados, en relación al modelo de agrupamiento K Medias se puede apreciar que cuenta con un coeficiente de silueta muy buena alrededor de 0.7 en los canales APP Móvil, y Web Transaccional, indicando una buena separación y segmentación intra grupos, en el resto de canales este indicador es moderado, y de pureza aceptable indicando que los grupos son homogéneos en un promedio del 0.6. La siguiente Tabla 28 se evidencia las métricas de validación analizadas. Lo que en general indica que los modelos son válidos para ser aplicados en el monitoreo transaccional para detección de transacciones anómalas en los canales virtuales.

Tabla 28. Métricas de validación de los modelos supervisados y no supervisados

Modelo	Métrica principal	Resultado	Interpretación
Árbol de decisión	AUC, F1-score	0,85; 0,78	Buen desempeño, alta interpretabilidad
Random Forest	AUC, F1-score, Kappa	0,91; 0,82; 0,69	Mejor modelo predictivo, robusto
W-PART	AUC, Reglas	0,83; 5 reglas clave	Más interpretabilidad, menor desempeño
K-means	Silueta, Pureza	0,70 APP Móvil, y Web Transaccional; 0.31 resto de canales; Pureza promedio 64% por canal electrónico	Segmentación moderada, perfiles diferenciados

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Asimismo, mediante la validación de las métricas se evidencia un nivel de segmentación aceptable con dos clústeres definidos, en cuanto al modelo KNN da cuenta que la curva ROC indica un buen porcentaje de clasificación correcta de los vecinos cercanos como se aprecia en la Tabla 29.

Tabla 29. Métricas de validación modelos K medias y KNN

Modelo	Métricas principales	Conclusión técnica
K-means	Índice de silueta promedio = 0,31; Pureza = 64%	Segmentación aceptable; identifica 2 clústeres (3 en ATMs), diferenciando perfiles con alta concentración de anomalías en clientes de mayores ingresos y montos elevados.
KNN	AUC = 0,77–0,82; Accuracy = 70–76%	Clasificación adecuada; prioriza sensibilidad para reducir falsos negativos; modelo simple y adaptable

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

6.4. Conclusiones del laboratorio

Los modelos aplicados resultaron adecuados para la detección de transacciones anómalas, mostrando buen desempeño y coherencia general. Como limitación, se reconoce que algunos supuestos clásicos no se cumplen estrictamente en la regresión cuantil, aunque la evidencia científica respalda su validez aún bajo estas condiciones. Por otra parte, el modelo Random

Forest se destaca por su replicabilidad y sólidos indicadores de desempeño, mientras que la técnica de agrupamiento K-Medias evidencia una segmentación homogénea, especialmente efectiva para diferenciar transacciones normales y anómalas en los canales APP Móvil y Web Transaccional, donde no existen umbrales transaccionales definidos por la institución financiera.

7. Conclusiones

Mediante el análisis exploratorio estadístico (EDA) se pudo determinar que en canales electrónicos que no cuentan con límite establecido por la institución como es el caso de la Web transaccional y App Móvil, el histograma de frecuencias y de densidad dio cuenta que clientes que realicen transacciones con montos superiores a USD 1.000 en adelante representan el 5% de los montos más altos y podrían considerarse como transacciones anómalas, en el resto de los canales al contar con límite restringido el comportamiento atípico percentil 95 son cercanos al propio límite restringido por la institución.

El análisis estadístico descriptivo dio cuenta también que existe diferencia del monto transaccional con relación al perfil socio demográfico y económico de los clientes, es ese sentido, se tiene que a mayor rango de ingreso incrementa el monto transaccional promedio, igualmente sucede con la edad generacional, y el nivel de estudio. El tipo de género también influye en el monto transaccional, en ese sentido el sexo masculino sobresale en relación al femenino, la actividad económica a la que se dedica el cliente también es una variable de peso a considerarse.

La estadística inferencial pudo determinar que las variables monto transaccional, ingresos, saldo promedio en ahorros vista no siguen una distribución normal por lo cual, se aplicó pruebas no paramétricas para validar la diferencia de medianas entre grupos y el monto transaccional, dando como resultado que las variables esperadas como tipo canal, género, actividad económica, edad generacional, rango de ingresos, nivel de estudio, difieren significativamente sus medianas respecto al monto transaccional, no rechazando la hipótesis del investigador.

Se concluye que el modelo de regresión cuantílica empleada para obtener la mediana del monto transaccional y del umbral superior en función del perfil sociodemográfico, económico y del perfil transaccional habitual del cliente es robusto, donde se obtuvo para el cuantil 97 un valor pseudo R^2 de 0,43 el cual es considerada de acuerdo a la literatura científica como un valor de moderado a alto, el modelo permitió determinar el las variables que más aportan a determinar los clientes que pueden pasar del umbral son: monto de la última transacción realizada, tiempo promedio entre transacciones, saldo promedio en ahorros vista, canal web transaccional, clientes que se dedican a la actividad del comercio al por mayor y menor,

género masculino, clientes con ingreso mayores a USD 2.350, clientes con edades generacionales Baby Boomers, y generación X.

Con los valores predichos obtenidos del modelo cuantílico 97 se pudo determinar que existen de las 146.483 transacciones aplicadas en el modelo 24.928 corresponden a transacciones etiquetadas como anómalas lo que corresponde a una participación del 17%.

Mediante la técnica de balanceo de clases (anómalas/normal) y una partición del 70% para entrenamiento y 30% las mejores métricas de desempeño se obtuvieron en el modelo supervisado Random Forest el cual, las etiquetas fueron tomadas del modelo de regresión cuantílica, lo cual ratifica y valida la probidad del modelo desarrollado. Las métricas de desempeño fueron en promedio por canal: Accuracy 75,66%, Curva bajo la roca (AUC) 80,82%; F1-score 69,3%

Los modelos W-part, R – Part, y KNN también presentaron buenas métricas de desempeño, pero no llegaron a sobre salir como el modelo Random Forest, sin embargo, se obtuvo información muy importante de reglas de clasificación y de los árboles de decisión los cuales ayudarán sin duda a la institución a contar con insumos con reglas para monitoreo transaccional.

Con la aplicación de la técnica de agrupamiento K medias se llega a concluir que el óptimo de número de clusters por canal es $k = 2$, a excepción del canal ATMs que es $k = 3$, asimismo, la calidad de la partición es moderada, la silueta media es 31% por canal lo que indica una separación limitada intra cluster a excepción de los canales App Móvil y Web transaccional que tuvieron una silueta promedio del 70% cada una, la pureza media fue del 64% respecto de la etiqueta de referencia sugiere correspondencia intermedia y cierto solapamiento entre clústeres.

Se concluye que, mediante la validación y robustez de los modelos desarrollados, que el modelo de regresión cuantílica satisface los indicadores requeridos como también el modelo Random Forest pudiendo tomarse ambos como referentes para la institución para replicarlo en sistemas informáticos para la detección y monitoreo de transacciones anómalas.

8. Bibliografía.

- Alvarado, J., & Pinos, O. (2017). *Estimación de ingresos de la población ecuatoriana una propuesta desde la regresión cuantílica*. Obtenido de Cuestiones Económicas Volumen 27: <https://estudioseconomicos.bce.fin.ec/index.php/RevistaCE/article/view/396>
- Damane, M. (2024). *Quantile regression analysis of the economic impact of business and household credit in Lesotho*. Obtenido de <https://mpra.ub.uni-muenchen.de/121954/>
- Gu, X., Akoglu, L., & Rinaldo, A. (2019). Statistical analysis of nearest neighbor methods for anomaly detection. *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*. Curran Associates, Inc. Obtenido de <https://proceedings.neurips.cc/paper/2019/hash/805163a0f0f128e473726ccd a5f91bac-Abstract.html>
- Hernández, L., Bustamante, L., Gutierrez, F., Moreno, J., & Rodriguez, M. (2024). Financial fraud detection through the application of machine learning techniques: A literature review. *Humanities and Social Sciences Communications*, 11. doi:10.1057/s41599-024-03606-0
- Hyndman, R., & Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, 50(4), 361 - 365. doi:10.1080/00031305.1996.10473566
- Koenker, R., & Hallock, K. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 143–156. doi:10.1257/jep.15.4.143
- Koenker, R., & Machado. (1999). *Journal of the American Statistical Association*. Obtenido de Goodness of fit and related inference processes for quantile regression : <https://doi.org/10.1080/01621459.1999.10473882>
- Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50. doi:10.2307/1913643
- Liu, F., Ting, K., & Zhou, Z.-H. (2008). Isolation forest. *2008 Eighth IEEE International Conference on Data Mining (ICDM)*. IEEE. doi:10.1109/ICDM.2008.17
- Zhong, L., & Matthijs, v. (2023). Explainable contextual anomaly detection using quantile regression forests. *Data Mining and Knowledge Discovery*, 37(6), 2517 - 2563. doi:10.1007/s10618-023-00967-z

Anexos

Tabla 30. Límites reglamentarios definidos por canal electrónico

Canal transaccional	Monto máximo transaccional	Observaciones
APP Móvil	Sin límite	El cliente establece el cupo máximo diario de forma manual.
Web transaccional	Sin límite	El cliente establece el cupo máximo diario de forma manual.
APP Móvil PYMES	USD 300,00	Monto máximo de retiro diario.
Cajeros automáticos (ATM)	USD 500,00	Monto máximo de retiro por día.
Botón de pagos	USD 300,00	Monto máximo de retiro por día.
Corresponsal no bancario	USD 200,00	Monto máximo de retiro por día.

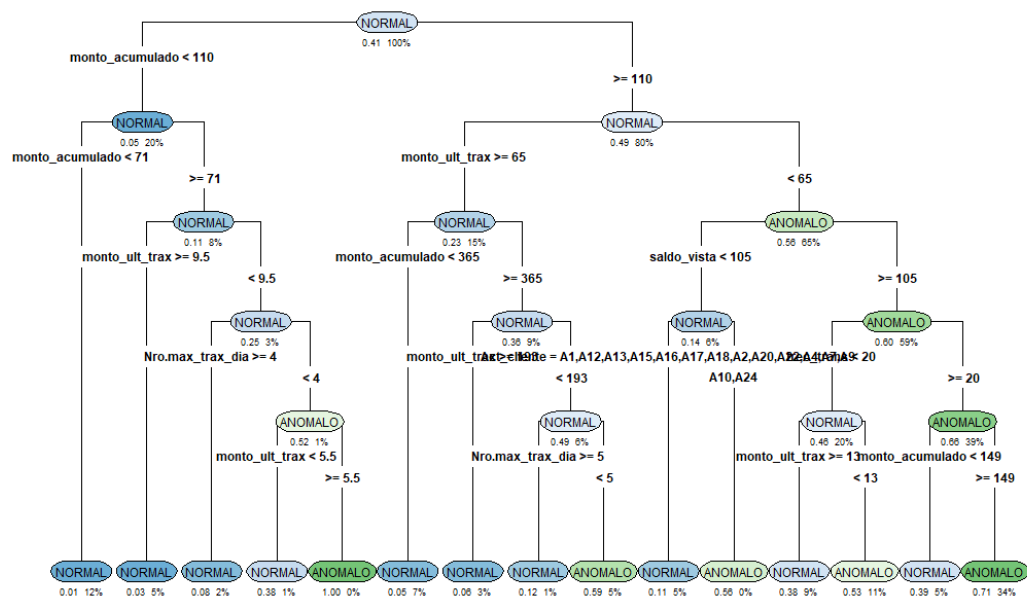
Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Tabla 31. Conceptos transaccionales incluidos en el análisis, según canal electrónico

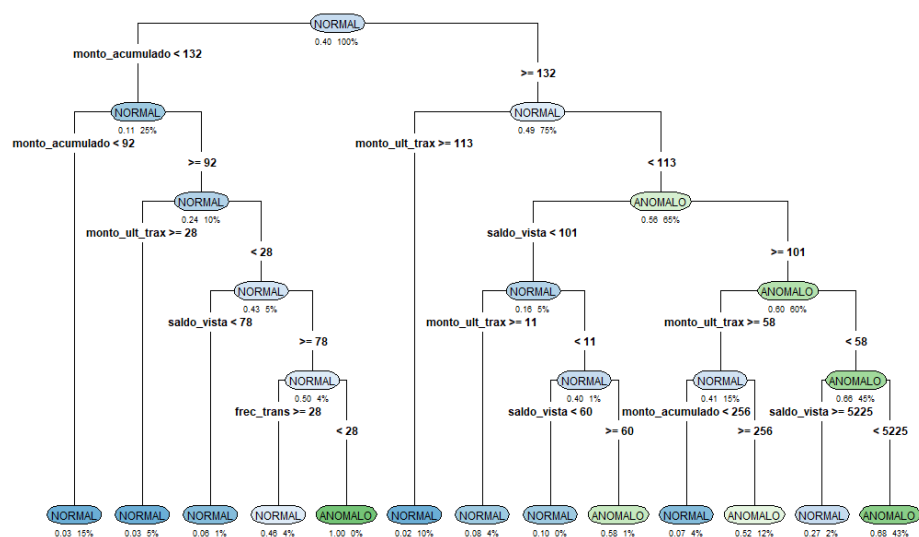
Descripción:	Código transacción
Cajeros Automáticos ATMS	Retiro
Web Transaccional	Transferencias en cuentas de ahorro de la IFI ⁴ ; y, Transferencias a otras IFIS
APP Móvil	Transferencias en cuentas de ahorro de la IFI; y, Transferencias a otras IFIS
APP Móvil PYMES	Transferencias en cuentas de ahorro de la IFI; y, Transferencias a otras IFIS
Botón de pagos	Transferencias en cuentas de ahorro de la IFI; y, Transferencias a otras IFIS
Corresponsales No Bancarios	NDRETP (nota débito por retiro)

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

⁴ IFI siglas que corresponde al significado de institución financiera

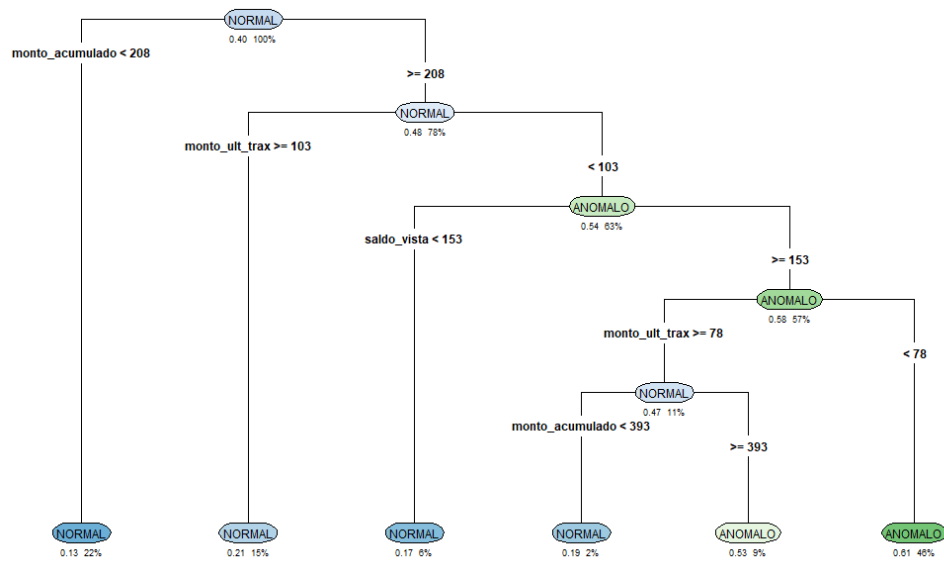
Figura 13. Árbol de decisión Rpart canal Botón De Pagos

Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Figura 14. Árbol de decisión Rpart canal Correspondientes no bancarios

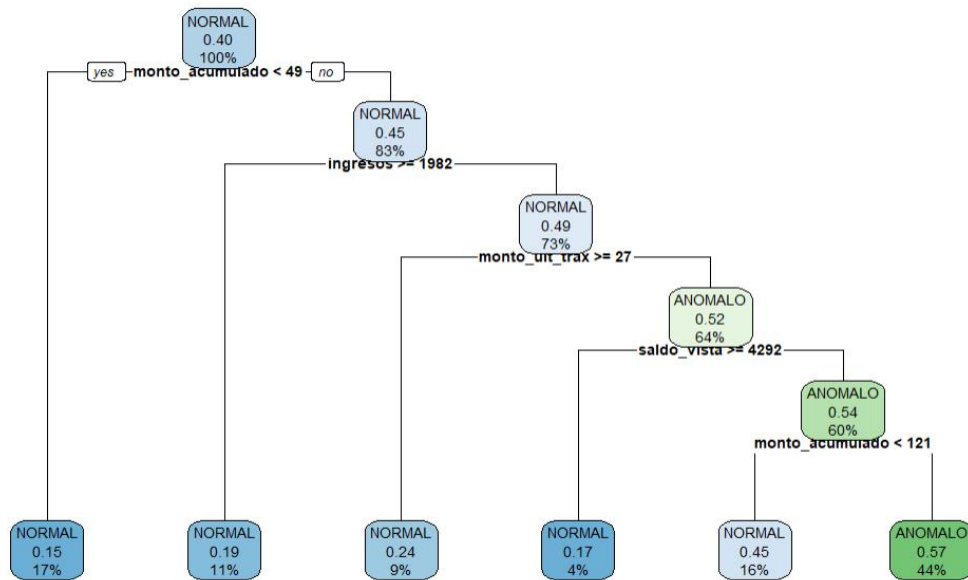
Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Figura 15. Árbol de decisión Rpart canal ATMS

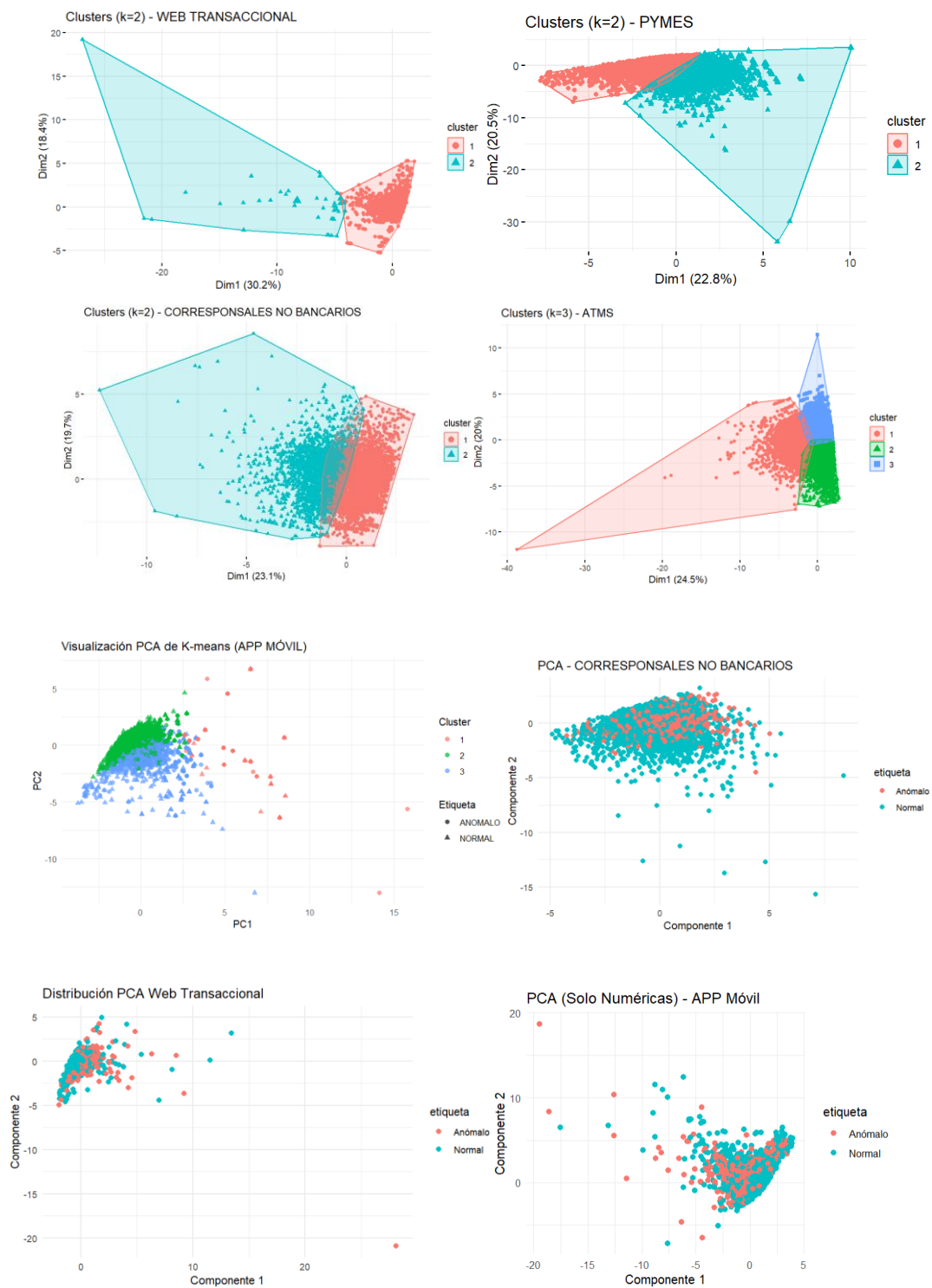


Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Figura 16. Árbol de decisión Rpart canal APP Móvil Pymes



Fuente. Elaboración: Omar Pinos (2025), con base en la entidad bancaria investigada, corte: junio de 2024.

Figura 17. Proyección PCA de los clústeres generados por el algoritmo K-medias por canal

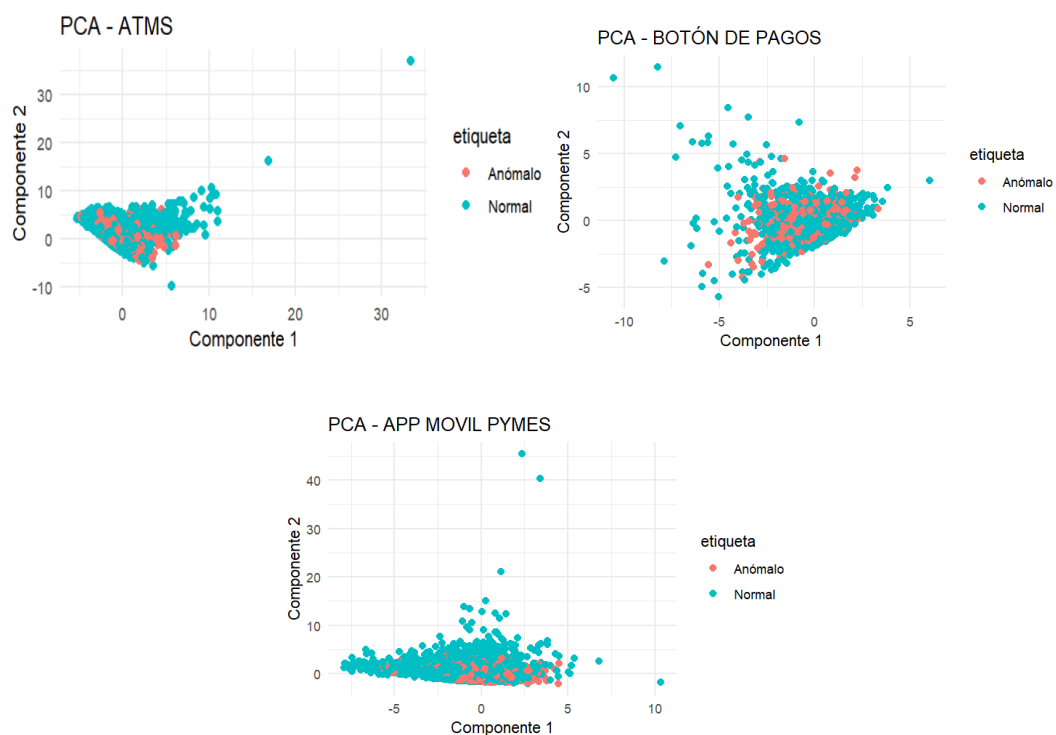


Figura 18. Curvas ROC–AUC del clasificador KNN por canal

