

UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Negocios



Implementación de una estrategia de *marketing* basada en modelo supervisado para definir clientes propensos a retirar efectivo con tarjeta de crédito

Trabajo de graduación presentado por

Valeria María Recinos Vidal

para optar al grado académico de Licenciada en International Marketing and Business Analytics

Guatemala

2025

UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Negocios



Implementación de una estrategia de *marketing* basada en modelo supervisado para definir clientes propensos a retirar efectivo con tarjeta de crédito

Trabajo de graduación presentado por

Valeria María Recinos Vidal

para optar al grado académico de Licenciada en International Marketing and Business Analytics

Guatemala

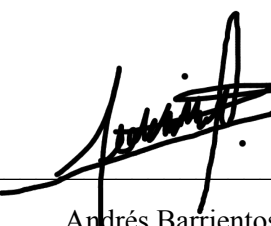
2025

Vo.Bo. :

(f) _____

Vicente Herrera

Tribunal examinador:

(f) _____

Andrés Barrientos

(f) _____

Kevin Gálvez

(f) _____

Fecha de aprobación: Guatemala, junio 2025

ÍNDICE

I. INTRODUCCIÓN	1
II. OBJETIVOS.....	2
III. JUSTIFICACIÓN.....	3
IV. MARCO TEÓRICO.....	4
V. ANTECEDENTES	6
VI. METODOLOGÍA.....	7
VI. RESULTADOS	52
VII. CONCLUSIONES.....	53
VIII. RECOMENDACIONES.....	54
IX. BIBLIOGRAFÍA.....	55

LISTA DE TABLAS

Tabla 1. Variables recopiladas	7
Tabla 2. Transformación de datos: agrupación	10
Tabla 3. Transformación de datos: indicadores	10
Tabla 4. Transformación de datos: representaciones de tiempo	11
Tabla 5. Transformación de datos: ratios	12
Tabla 6. Transformación de datos: codificación de atributos categóricos	12
Tabla 7. Crecimiento anual de retiradores	13
Tabla 8. Cantidad de retiradores en fecha de pago de tarjeta de crédito	27
Tabla 9. Crecimiento anual de transacciones	28
Tabla 10. Matriz de correlación final	39
Tabla 11. Conjunto de datos desequilibrado	40
Tabla 12. Conjunto de datos equilibrado	40
Tabla 13. Conjunto de datos equilibrado	41

LISTA DE ILUSTRACIONES

Ilustración 1. retiradores por año (2021-2024)	13
Ilustración 2. Proporción de clientes retiradores vs. clientes con TC	14
Ilustración 3. Cantidad de retiradores por mes (2021-2024)	15
Ilustración 4. Género de retiradores	16
Ilustración 5. Género de retiradores por año (2021-2024).....	16
Ilustración 6. Edad de retiradores	17
Ilustración 7. Nivel socioeconómico.....	18
Ilustración 8. Departamento de origen de retiradores	19
Ilustración 9. Categoría de tarjeta de crédito de retiradores	20
Ilustración 10. Comparación de la antigüedad de tarjeta de crédito de los clientes retiradores vs. no retiradores.....	21
Ilustración 11. Meses transcurridos hasta el primer retiro	22
Ilustración 12. Comparación de uso de saldo revolvente para clientes retiradores vs. no retiradores	23
Ilustración 13. Comparación de uso límite de la tarjeta de crédito.....	24
Ilustración 14. <i>Behavior Score Interno</i> de clientes retiradores.....	25
Ilustración 15. Días entre retiro de efectivo y fecha de pago de tarjeta de crédito	26
Ilustración 16. Transacciones anuales.....	28
Ilustración 17. Transacciones por mes anual	29
Ilustración 18. Cantidad de retiros por cliente	30
Ilustración 19. Análisis de frecuencia mensual de retiros por cliente.....	31
Ilustración 20. Análisis de frecuencia mensual de retiros por cliente II	31
Ilustración 21. Promedio de retiros por mes	32
Ilustración 22. Promedio de monto por retiro	33
Ilustración 23. Representación monetaria de retiros mensuales	33
Ilustración 24. Género por segmento de retiradores	34
Ilustración 25. Edad por segmento de retiradores.....	35
Ilustración 26. Tarjeta asociada a segmento de retiradores	36
Ilustración 27. Segmento de mercado según segmento de retiradores	37
Ilustración 28. Antigüedad de tarjeta en segmentos retiradores	38
Ilustración 29. Matriz de confusión para modelo: árbol de decisión	42
Ilustración 30. Matriz de confusión para modelo: <i>Naive bayes</i>	43
Ilustración 31. Matriz de confusión para modelo: <i>Random forest</i>	44
Ilustración 32. Matriz de confusión para modelo: <i>XGBoost</i>	45
Ilustración 33. Comparación de <i>accuracy</i> de cada modelo	46
Ilustración 34. Comparación de métricas por clase y modelo	46
Ilustración 35. Curva de error en validación cruzada	47
Ilustración 36. <i>Buyer persona</i>	49
Ilustración 37. Pirámide de retiradores frecuentes.....	50
Ilustración 38. Presupuesto representativo	51
Ilustración 39. Proyección de resultados de bosque aleatorio	52

I. INTRODUCCIÓN

El presente proyecto describe una propuesta de implementación de una estrategia de mercadeo utilizando el poder de los datos para impulsar un servicio financiero. El servicio que se desea impulsar a través de estrategias de segmentación es el retiro de efectivo con tarjeta de crédito. Las instituciones financieras manejan grandes cantidades de información para ofrecer sus servicios, lo cual puede resultar desafiante para una estrategia de segmentación. Es por ello que, se utiliza un modelo de predicción clasificatorio para identificar a los clientes propensos a retirar efectivo por primera vez y a impulsar más retiros de los clientes ya existentes. Se emplea la metodología de *CRISP-DM* de la ciencia de datos para detectar el problema, la solución y el modelo con mejor rendimiento para identificar a los segmentos de clientes. Basada en la información recolectada se diseña la estrategia de marketing. A través de este proyecto se concluye que la minería de datos es de alto valor para un plan de marketing.

II. OBJETIVOS

A. General

Implementar una estrategia de marketing con el propósito de aumentar un 10 % el crecimiento anual de la cartera de clientes que retiran efectivo con tarjeta de crédito (TC) e impulsar un segundo retiro mensual en clientes actuales utilizando un modelo predictivo para identificar clientes con potencial para realizar un primer y segundo retiro.

B. Específicos

- Desarrollar una base de datos de clientes con tarjeta de crédito utilizando la información disponible de la institución.
- Desarrollar una base de datos de clientes que han retirado efectivo con su tarjeta de crédito a lo largo del tiempo con la información disponible de la institución.
- Desarrollar e implementar un modelo de aprendizaje supervisado para predecir la probabilidad de retirar efectivo de un cliente con TC basado en datos históricos, evaluado mediante matriz de confusión: tasas de *precisión*, *accuracy* y *recall*.
- Definir el perfil de los clientes retiradores entendiendo su comportamiento a lo largo del tiempo desde la primera transacción.
- Identificar y segmentar clientes con alta probabilidad de realizar retiros de efectivo para diseñar estrategias de marketing personalizadas.
- Segmentar clientes con patrones de retiro recurrentes para fomentar una segunda transacción mensual.
- Desarrollar estrategias de marketing basadas en los segmentos identificados, midiendo el impacto de estas con el porcentaje de clientes que respondieron positivamente con tasas de conversión.

III. JUSTIFICACIÓN

Las instituciones bancarias manejan una extensa cantidad de datos de sus clientes para ofrecer servicios, lo cual representa una ventaja competitiva si se utiliza esta información para tomar decisiones que generen resultados beneficiosos. En el presente proyecto se manejan y analizan datos de los clientes con tarjeta de crédito para que utilicen el servicio de retiro de efectivo en cajeros con esta. Este servicio es un beneficio de la tarjeta de crédito que es atractivo para los clientes porque ofrece accesibilidad inmediata a efectivo, sin necesidad de trámites adicionales, con la disponibilidad del crédito de la tarjeta.

La cantidad de clientes que retiraron efectivo en 2024 creció un 55 %, lo que representó una reducción del 5 % en comparación al año anterior. Otra tendencia afectada es el aumento en transacciones por retiradores anual, disminuyendo un 19 % en 2024 comparado al año anterior. Al aumentar el ritmo de crecimiento de transacciones, se alcanzan dos retiros mensuales por cliente, actualmente con 1.6 retiros al mes. Al aumentar el crecimiento de retiradores en 10 % y dos retiros mensuales, se supera el monto de retiro de los últimos años y se impulsan las tendencias positivas de retiradores y transacciones anuales.

Por consiguiente, se busca identificar a los clientes con *TC* con propensión a retirar por primera vez y a los clientes ya retiradores con un retiro mensual por medio de segmentación. No obstante, desde que se manejan extensas cantidades de datos de los clientes en la institución, un modelo supervisado de clasificación es de gran apoyo para encontrar a los clientes y establecer probabilidades de retiro. Este modelo tiene como propósito optimizar las comunicaciones que ya se realizan a los clientes para incentivar el retiro identificando a los clientes correctos y obtener resultados más rápido.

IV. MARCO TEÓRICO

Para una mejor comprensión del contexto del presente proyecto se describirán las definiciones de los siguientes conceptos a utilizar:

A. Tarjeta de crédito

Es un documento plástico emitido por un banco o institución financiera que se utiliza como método de pago. Este método permite realizar compras mediante crédito límite. Este crédito es otorgado por la institución financiera según el perfil de la persona [1].

B. Retiro de efectivo

Es un servicio que ofrecen las tarjetas de crédito que permite retirar efectivo de cajeros automáticos permitiendo al tarjetahabiente obtener efectivo de manera inmediata. El banco cobra una comisión proporcional al monto de efectivo retirado [2].

C. Método *CRISP-DM*

Abreviatura de *Cross-Industry Standard Process for Data Mining*. Es un método de trabajo en ciencia de datos donde a través de 6 fases: entendimiento del negocio, entendimiento de los datos, preparación de los datos, modelado, evaluación y presentación, se desarrolla un proyecto de minería de datos [3].

D. Estrategia de *marketing*

Enfoque de la organización para alcanzar los objetivos definidos a través de componentes de *marketing*. El plan de *marketing* es la planificación detallada de cómo alcanzar los objetivos planteados en la estrategia de *marketing* [4].

E. *Machine learning*

En español, aprendizaje automático. Es una rama de la inteligencia artificial que usa datos y algoritmos para modelar predicciones o clasificaciones basadas en los datos de entradas. Los algoritmos permiten que la inteligencia artificial mejore la precisión de las predicciones con grandes cantidades de datos [5] [6].

F. Aprendizaje supervisado

Es subcategoría de machine learning. Este tipo de modelaje utiliza un conjunto de variables para describir la relación con la variable llamada target, del resultado deseado. Estos modelos no agrupan ni clasifican datos sin un humano. Existen varios

tipos de modelos para el aprendizaje supervisado que depende de los datos a utilizar [5] [6].

G. Aprendizaje no supervisado

A diferencia del aprendizaje supervisado, el aprendizaje no supervisado descubre patrones o agrupaciones de datos sin necesidad de intervención humana a través de sus algoritmos. Los modelos de aprendizaje no supervisado son utilizados principalmente para agrupación, asociación y reducción de dimensionalidad [6].

H. Sobreajuste

Término de *machine learning* que refiere a un problema en el modelo de los datos cuando un modelo se ajusta demasiado bien a los datos de entrenamiento que causa que el modelo pierda la capacidad de identificar patrones significativos o adecuando los datos nuevos que no se han visto durante el proceso de entrenamiento [7].

I. Árbol de decisión (*decision tree*)

Es un modelo de aprendizaje supervisado de clasificación que segmenta las variables según las características para hacer una predicción de las clases a través de una estructura en forma de árbol; con ramas, nodos y hojas [8].

J. Matriz de correlación

Es una matriz de variables dependientes que describe el grado de relación entre sí. La correlación es un indicador estadístico definido por el coeficiente de relación - R- por medio de una escala que varía de 1 y -1, estos representan la relación perfecta entre las variables, -1 siendo la relación perfecta inversa. $R = 0$ representa ausencia de correlación entre las variables, lo que se traduce a que las variables son independientes [9].

K. Matriz de confusión

Es una matriz que representa las predicciones y los valores reales de las clasificaciones del modelo. Esta matriz puede dar resultados de accuracy, precisión, recall, falsos positivos y positivos falsos [10].

V. ANTECEDENTES

En el año 2023 se hizo el último análisis de retiro de efectivo en la institución. Este análisis tuvo como objetivo mostrar el comportamiento de los clientes que han retirado efectivo a lo largo del tiempo, comparar el comportamiento de clientes que retiran durante días de promoción días estándar y construir el perfil del cliente retirador de efectivo. Se hizo mediante un análisis histórico de retiros de efectivo de 2019 a 2023. Este análisis reveló que el crecimiento interanual de los clientes que retiran efectivo era de 1 %, que el 73 % de clientes retiran efectivo sin promociones y son estos que hacen dos retiros promedio. Otro dato relevante de este análisis es que el 20% de clientes retiran efectivo con y sin promoción, alcanzando 5 retiros promedio. Este análisis construyó el perfil del retirador basándose en tres variables: metal de la tarjeta de crédito, sexo y generación. Con estas variables se concluyó que los clientes retiradores poseen una tarjeta de crédito de categoría baja, siendo la clásica la más baja de categoría, poca relevancia en el género del cliente y siendo los clientes de la generación Z quienes más retiran efectivo.

Este análisis no tuvo un seguimiento con la información obtenida y no ha sido actualizado desde entonces. Por consiguiente, se realiza el siguiente proyecto, con los objetivos de actualizar información, encontrar oportunidades e implementar un proyecto con el conocimiento adquirido para mejorar procesos del servicio. Se pretende que los datos actualizados sean utilizados estratégicamente para la mejora continua y la innovación.

VI. METODOLOGÍA

Para el desarrollo de este proyecto se utiliza la metodología *CRIPS-DM* (*Cross Industry Standard Process for Data Mining*). Esta metodología se basa en las siguientes 6 etapas:

A. Entendimiento de negocio

El retiro de efectivo es un servicio de la tarjeta de crédito que permite a los tarjetahabientes retirar efectivo desde cajeros automáticos o agencias. El monto límite a retirar es según la disponibilidad del límite de la tarjeta. Se cobra comisión para la institución sobre el monto de efectivo retirado, este varía según la categoría de metal de la tarjeta de crédito. Las promociones actuales de retiro de efectivo consisten en ofrecer una tasa de comisión especial en días específicos del mes. Estas promociones se comunican a través de un mensaje de texto dirigido a todos los clientes con tarjeta de crédito y con publicaciones en redes sociales.

B. Entendimiento de los datos

Los datos de retiro de efectivo provienen de la tabla de cartera de clientes con tarjeta de crédito de la base de datos relacionada de la institución. A través de esta base de datos se hace una recopilación de las variables a utilizar acerca de los clientes con tarjeta de crédito fuera de la tabla mencionada. Estos datos son confiables ya que provienen directamente de la *data warehouse* de la institución.

Las variables recopiladas son las siguientes:

Tabla 1. Variables recopiladas

FEC_CARTERA	Fecha	Fecha de la cartera de clientes con tarjeta de crédito
COD_CLIENTE	Numérica	ID del cliente
NUM_CUENTA	Numérica	Número de cuenta asociada al cliente
FEC_COSECHA	Fecha	Fecha de emisión de la tarjeta de crédito
CATEGORIA	Carácter	Categoría a la que pertenece la tarjeta de crédito del cliente
PRODUCTO_AGRUPADOR	Carácter	Tipo de metales de tarjeta de crédito que pertenecen al cliente

PRODUCTO	Carácter	Metal de tarjeta de crédito que pertenecen al cliente
CICLO	Numérica	Número de ciclo en el que se encuentra la tarjeta de crédito desde la cosecha
LIMITE_CUENTA	Numérica	El crédito límite de la cuenta asociada con la tarjeta de crédito
DISPONIBLE_GTQ	Numérica	Monto disponible de límite del crédito otorgado de la cuenta en quetzales
MON_PAGO_CONTADO	Numérica	Monto de dinero que pagó el cliente a la tarjeta de crédito
MON_PAGO_MINIMO	NUMÉRICA	Monto mínimo de dinero que el cliente debe pagar a la tarjeta de crédito
SCORE_INT	NUMÉRICA	Calificación que se le otorga al cliente con respecto a su comportamiento con el crédito histórico interno a la institución
BHV_SCORE_INT	CARÁCTER	Calificación que se le otorga al cliente con respecto a su comportamiento con el crédito histórico interno a la institución en categoría
SCORE_EXT	NUMÉRICA	Calificación que se le otorga al cliente con respecto a su comportamiento con el crédito histórico externo a la institución
BHV_SCORE_EXT	CARÁCTER	Calificación que se le otorga al cliente con respecto a su comportamiento con el crédito histórico externo a la institución en clasificación
NUM_PAGOS	NUMÉRICA	Cantidad de pagos que hizo el cliente a la tarjeta de crédito en el mes
TASA_INTERES	NUMÉRICA	Qué tasa de interés tiene la cuenta actual

CANT_MORAS	NUMÉRICA	Cantidad de retrasos en el pago de la deuda de la tarjeta / cuenta
NUM_RETIRO_EFECTIVO	NUMÉRICA	Cantidad de retiros de efectivo en ese mes de la cuenta asociada al cliente
MON_RETIRO_EFECTIVO	NUMÉRICA	Suma del monto de efectivo retirado en ese mes de la cuenta asociada al cliente
COSECHA_XF	FECHA	Fecha que tomó extrafinanciamiento
OFERTA_XF	NUMÉRICA	Oferta disponible de extrafinanciamiento en el mes según código de cliente
MONTO_XF	NUMÉRICA	El monto restante luego de haber tomado xf
MON_SALDO_VC	NUMÉRICA	El monto restante luego de haber tomado visacuotas
MON_SALDO_XF	NUMÉRICA	El monto restante luego de haber tomado xf
FEC_APERTURA_CMO	FECHA	Fecha apertura de cuenta monetaria
SALDO_PROMEDIO_Cartera_GTQ	NUMÉRICA	Suma del saldo promedio de la cuenta de monetaria
FEC_APERTURA_CAH	FECHA	Fecha apertura de cuenta de ahorro
SALDO_PROMEDIO_Cartera_CAH_GTQ	NUMÉRICA	Suma del saldo promedio de la cuenta de ahorro
SEXO	CARÁCTER	Cliente masculino o femenino (f/m)
EDAD_REAL	NUMÉRICA	Edad del cliente en el mes y año
ESTADO_CIVIL	CARÁCTER	Estado civil del cliente
TIP_CLIENTE	CARÁCTER	Si el cliente es jurídica/ individual
MON_SALDO_IF	NUMÉRICA	El monto restante luego de haber tomado intrafinanciamiento
OFERTA_IF	NUMÉRICA	Oferta disponible de intrafinanciamiento en el mes según código de cliente
DEPARTAMENTO_DEFAULT	CARÁCTER	Departamento de residencia del cliente
SEGMENTO_MERCADO	CARÁCTER	Nivel socioeconómico del cliente según criterios de la institución

FEC_PAGO	FECHA	Fecha de pago de la tarjeta de crédito
----------	-------	--

C. Preparación de datos

Se crea un *dataset* con las variables de interés seleccionadas esperando la relación de esas variables con el retiro de efectivo. El *dataset* es el histórico de los clientes con tarjeta de crédito y las variables desde enero 2021 a diciembre 2024 en meses. Se hacen filtros y se construyen lógicas para estandarizar la información. Entre ellos:

a. Agrupación

Tabla 2. Transformación de datos: agrupación

AGRUPACION_BHV_SCORE_INT	AGRUPACIÓN RESUMIDA DE CALIFICACION OTORGADA AL CLIENTE POR SU RÉCORD CREDITICIO EN 4 CATEGORÍAS
RANGOS_OFERTA_IF	RANGOS DE MILES DE OFERTA DE INTRAFINANCIAMIENTO DEL CLIENTE
RANGOS_OFERTA_XF	RANGOS DE 5 MIL DE OFERTA DE EXTRAFINANCIAMIENTO DEL CLIENTE
RANGOS_SUM_SALDO_CARTERA_CAH_G TQ	RANGOS DE 100, MIL, 5MIL, 10MIL DEL SALDO EN CUENTA AHORRO
RANGOS_SUM_SALDO_CARTERA_CMO_G TQ	RANGOS DE 100, MIL, 5MIL, 10MIL DEL SALDO EN CUENTA MONETARIA
RANGOS_MONTO_PAGO_CONTADO	RANGOS DE MIL, 2MIL, Y 10MIL DE MONTO PAGO CONTADO A LA TC
RANGOS_MONTO_PAGO_MINIMO	RANGOS DE 200, 250, 500, 750, MIL DE MONTO PAGO MINIMO A LA TC

Estos rangos se definieron con base en los cuartiles de cada variable.

b. Indicadores

Tabla 3. Transformación de datos: indicadores

RETIRADORES	CLIENTE QUE RETIRÓ EN EL MES Y AÑO O NO
-------------	---

MAX_RETIRADORES	CLASIFICACIÓN DEL CLIENTE SI FUE RETIRADOR EN ALGÚN MOMENTO O NUNCA RETIRÓ
PAGADORES_CONTADO	EL CLIENTE PAGÓ O NO AL CONTADO LA TARJETA DE CRÉDITO EN EL MES

Estos indicadores se hicieron con funciones: *ifelse* donde se crea el indicador 1 o 0 si la variable cumple con la expresión.

c. Representaciones de tiempo

Tabla 4. Transformación de datos: representaciones de tiempo

MES_CARTERA	FECHA DE LA CARTERA DE CLIENTES CON TARJETA DE CRÉDITO EN FORMATO MES
AÑO_CARTERA	FECHA DE LA CARTERA DE CLIENTES CON TARJETA DE CRÉDITO EN FORMATO AÑO
ANTIGÜEDAD_COSECHA_MESES	MESES TRANSCURRIDOS DESDE LA COSECHA (EMISIÓN) DE LA TC PARA DICIEMBRE 2024
MES_COSECHA	FECHA DE EMISIÓN DE CADA TARJETA DE CRÉDITO EN FORMATO MES
AÑO_COSECHA	FECHA DE EMISIÓN DE CADA TARJETA DE CRÉDITO EN FORMATO AÑO
AÑO_PRIMERA_FEC_TRX	PRIMER RETIRO DEL MES DEL CLIENTE EN FORMATO AÑO
MES_PRIMERA_FEC_TRX	PRIMER RETIRO DEL MES DEL CLIENTE EN FORMATO MES
MIN_FEC_TRX	PRIMER RETIRO DEL MES DEL CLIENTE
AÑO_ULTIMA_FEC_TRX	ULTIMO RETIRO DEL MES DEL CLIENTE EN FORMATO AÑO
MES_ULTIMA_FEC_TRX	ULTIMO RETIRO DEL MES DEL CLIENTE EN FORMATO MES
MAX_FEC_TRX	ULTIMO RETIRO DEL MES DEL CLIENTE
FRECUENCIA_ULT_TRX	CANTIDAD DE DIAS TRANSCURRIDOS DESDE LA ULTIMA TRANSACCION
DIAS_TRX_FECHAPAGO_TC	RESTA DE DIAS DE LA FECHA DE PAGO DE TARJETA DE CRÉDITO Y FECHA DEL PRIMER RETIRO DEL MES
MESES_TRS_ULT_TRX	MESES TRANSCURRIDOS DESDE LA ÚLTIMA TRANSACCIÓN HASTA DICIEMBRE 2024
MESES_XF_COSECHATC	MESES TRANSCURRIDOS ENTRE EMISIÓN DE TC Y TOMA DE XF
MESES_PRIMERATRX_COSECHATC	MESES TRANSCURRIDOS ENTRE EMISIÓN DE TC Y PRIMER RETIRO

Estas representaciones de tiempo se construyeron con la resta de tiempo (meses, días) entre las variables necesarias.

d. Ratios

Tabla 5. Transformación de datos: ratios

UTILIZACION_LIM	PORCENTAJE DEL LIMITE DE CRÉDITO UTILIZADO CON RESPECTO AL LIMITE DE CUENTA
RATIO_SALDO_REVOLVENTE	PORCENTAJE DEL SALDO REVOLVENTE UTILIZADO CON RESPECTO AL TOTAL DEL LIMITE DE CUENTA
UTILIZACION_XF	PORCENTAJE DE USO DE EXTRAFINANCIAMIENTO CON RESPECTO AL MONTO XF
UTILIZACION_IF	PORCENTAJE DE USO DE INTRAFINANCIAMIENTO CON RESPECTO AL MONTO IF

Estas ratios se crearon a partir de la división entre las variables necesarias para formar un porcentaje.

e. Codificación de atributos categóricos

Tabla 6. Transformación de datos: codificación de atributos categóricos

ESTADO_CIVIL	1: SOLTERO / 0: CASADO
TIP_CLIENTE	1: PERSONA / 0: JUDICIAL

Estas codificaciones representan en número un significado categórico. Se crean a partir de la función *ifelse*, si el registro es de un tipo es 1, sino 0.

1. Limpieza de datos

Iniciando con 714,000 clientes en el *dataset* de 2021 a 2024, se realizó una limpieza de clientes para excluir clientes inconsistentes, inactivos y de prueba. Los clientes inconsistentes son los clientes que pertenecieron a la cartera de clientes de tarjeta de crédito, pero en algún momento cancelaron su tarjeta. Para esta limpieza se construyó una lógica basada en el porcentaje de meses que el cliente permanece en el *dataset*. Los clientes inactivos son los que no cuentan con una cuenta activa en el presente. Además, se aplicaron filtros para excluir a los retiradores de prueba de la institución. El *dataset* tuvo una caída de 387,000 clientes, por consiguiente, trabajamos con un *dataset* de 327,000 clientes en la cartera de tarjeta de crédito.

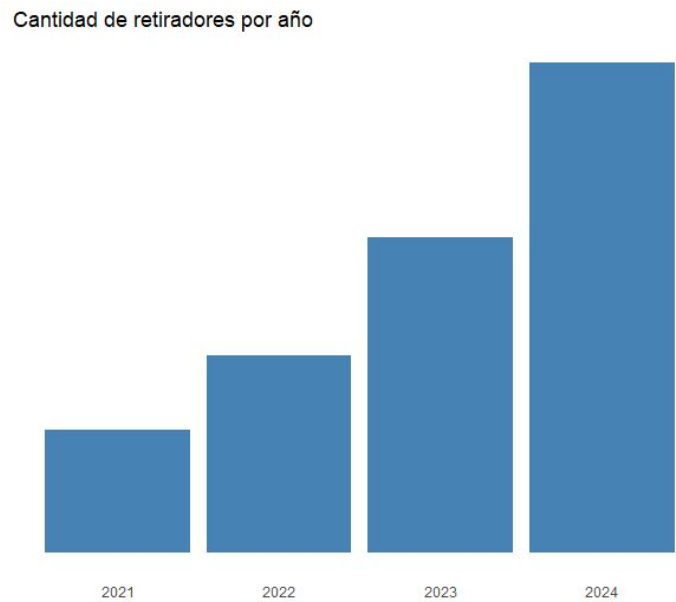
2. Entendimiento de datos (II)

Tras la preparación de los datos, se realiza un análisis exploratorio de los datos. En este se realizan gráficas para entender el comportamiento de las variables, el contexto del *dataset* y conocer al cliente retirador. Para ello se utiliza la herramienta de R, un entorno y lenguaje de programación para analizar información estadísticamente.

Como primera parte del análisis se busca entender el comportamiento de los clientes retiradores a lo largo del tiempo, desde 2021 a 2024. De 2021 a 2024, el 48 % del total de clientes con tarjeta de crédito han retirado efectivo en algún momento en el tiempo.

A más detalle,

Ilustración 1. retiradores por año (2021-2024)



A través de esta visualización observamos que los clientes retiradores han aumentado con el tiempo. Sobre el total de retiradores en este periodo del tiempo, el 10 % fueron retiradores en 2021 y en 2024 el 43 %.

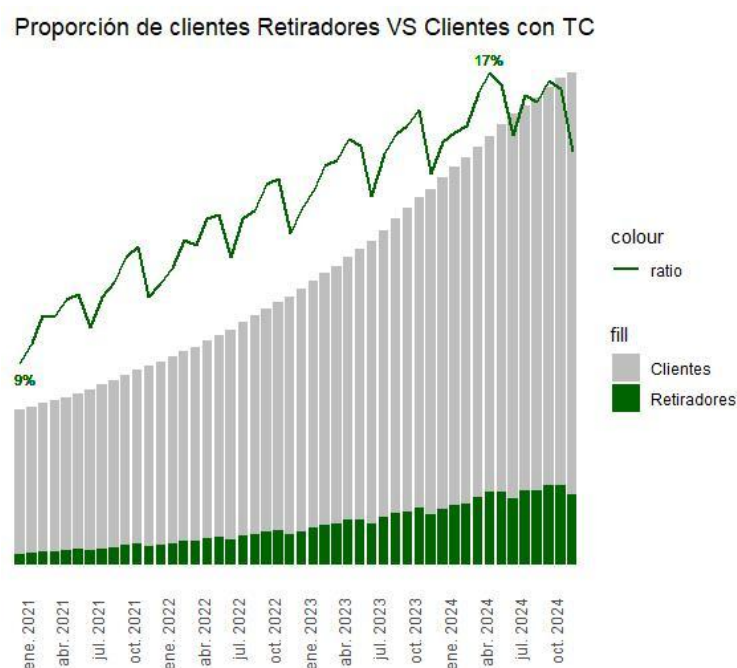
Sin embargo, el ritmo de adopción de este servicio por parte de los clientes ha sido el siguiente:

Tabla 7. Crecimiento anual de retiradores

AÑO	CRECIMIENTO (%) DE RETIRADORES
2021	-
2022	60.2%
2023	60.2%
2024	55.2%

El crecimiento porcentual de clientes retiradores de efectivo ha mantenido el mismo ritmo de 2022 a 2023, sin embargo, ha caído el crecimiento en cinco puntos porcentuales en 2024. Por lo tanto, es importante que en 2025 se prevenga un comportamiento de caída y se recupere el comportamiento previo.

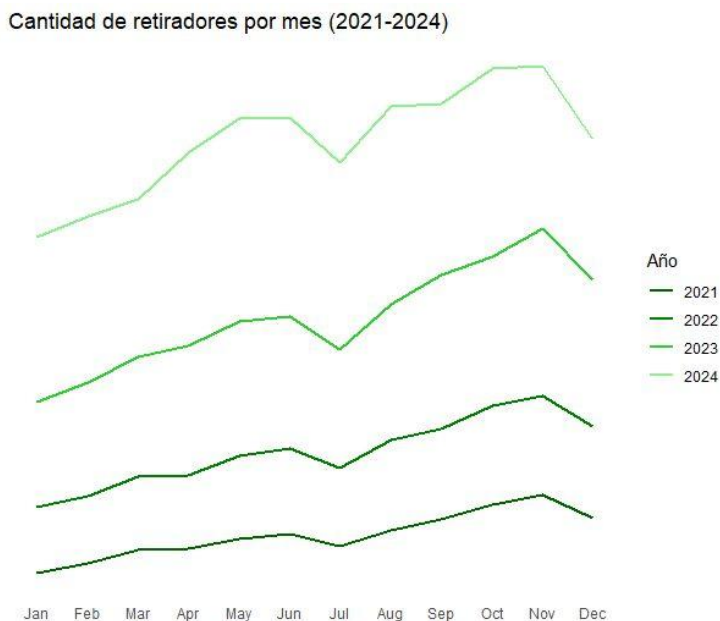
Ilustración 2. Proporción de clientes retiradores vs. clientes con TC



En la Ilustración 2 visualizamos el comportamiento de los clientes retiradores con respecto a los clientes con tarjeta de crédito. La cartera de clientes con tarjeta de crédito ha aumentado en el tiempo eso explica el comportamiento de clientes retiradores. En enero 2021, el ratio de retiradores versus clientes fue del 9 %, el momento con menos conversión en el período. En mayo 2024 fue el mes con más conversión, siendo el 17 % de clientes que retiraron.

En cuanto al comportamiento de retiradores mensuales, se presenta lo siguiente:

Ilustración 3. Cantidad de retiradores por mes (2021-2024)



Los retiradores muestran un patrón similar cada año, donde a inicio de año empiezan a aumentar, caen en julio y retoman de nuevo el crecimiento hasta noviembre. En diciembre la cantidad de retiradores caen nuevamente. Este patrón muestra una estacionalidad cada semestre del año. Este comportamiento se da a explicar por las bonificaciones obligatorias a los guatemaltecos: bono 14 y aguinaldo. Son momentos del año donde los clientes reciben ingresos extraordinarios, por lo que, no les es necesario retirar efectivo.

A continuación, se procede al análisis las variables para perfilar a los clientes retiradores. El análisis se estructura dos partes, la primera orientada a describirlo demográficamente y la segunda centrada en las variables relacionadas con el comportamiento del cliente de la institución.

El análisis demográfico es el siguiente:

Ilustración 4. Género de retiradores

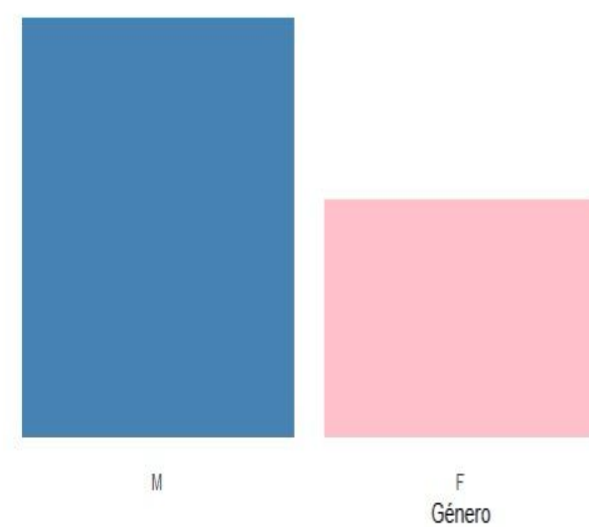
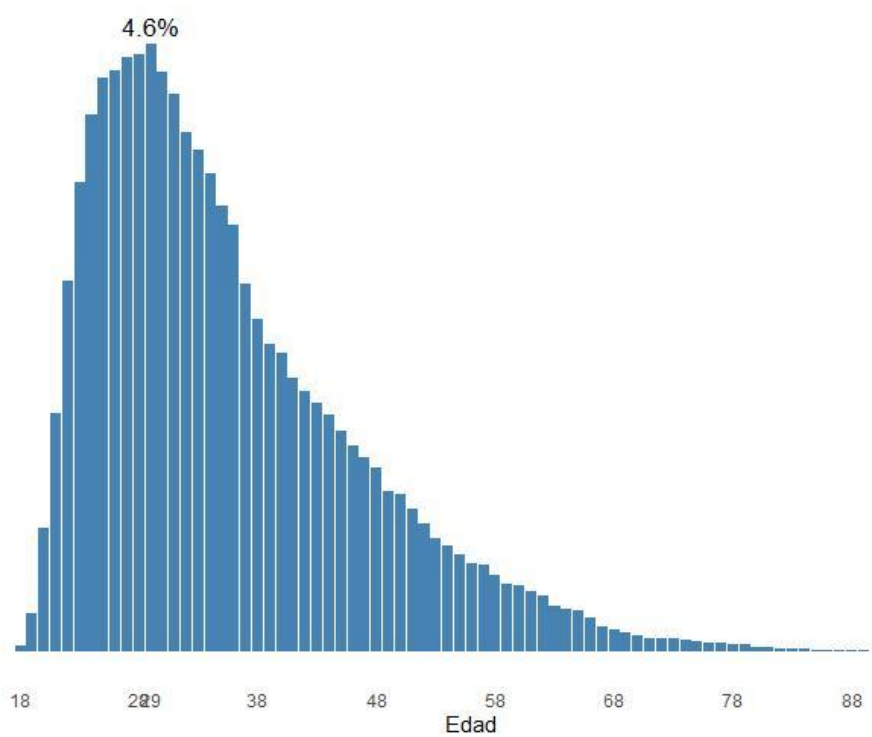


Ilustración 5. Género de retiradores por año (2021-2024)



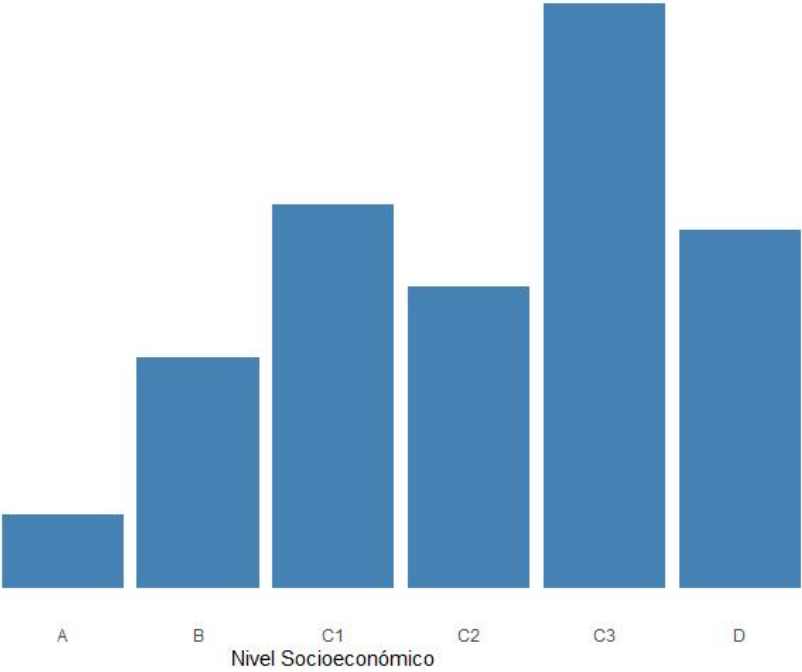
Con relación al género del retirador, donde M para masculino y F para femenino. El 63 % de los retiradores son hombres y el 36 % son mujeres. La diferencia es significativa, por lo que, se quiso observar esta comparación por año. Donde observamos que sí se mantuvo una diferencia durante el tiempo.

Ilustración 6. Edad de retiradores



En cuanto a la última edad de los retiradores (2024), los retiradores se concentran en edades jóvenes. La mayoría de retiradores están en un rango de edad de 24 a 30 años, el 4.6 % de los retiradores tienen 29 años, siendo la edad más popular.

Ilustración 7. Nivel socioeconómico

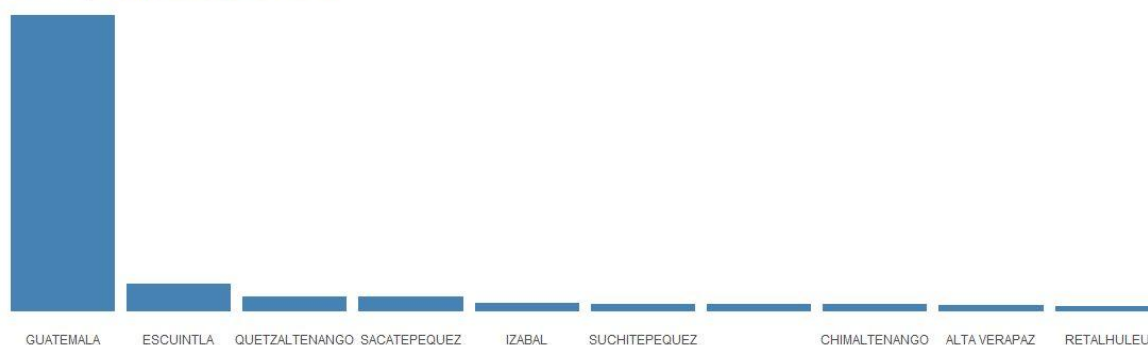


Acerca del nivel socioeconómico de los retiradores, el 30 % son de categoría C3, 19 % C1 y 18 % de nivel D. Con esta visualización concluimos que los clientes son de nivel socioeconómico bajo. Los criterios para definir el nivel socioeconómico de los clientes por el banco son: ingreso económico reportado o estimado, límite aprobado de la tarjeta, producto de tarjeta de crédito asignado e historial crediticio.

Por último, definimos al cliente retirador por departamento de origen.

Ilustración 8. Departamento de origen de retiradores

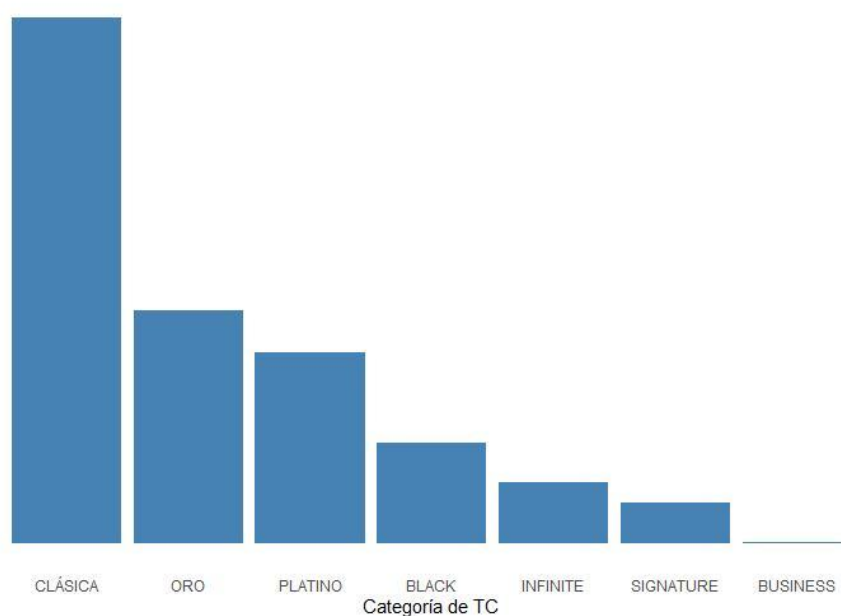
TOP 10 Departamentos con Retiradores



La gran mayoría de retiradores (70 %) se concentran en la ciudad de Guatemala. En la visualización observamos los top 10 departamentos con más retiradores. El 6 % en escuintla y el 4 % en Quetzaltenango. Estas diferencias son bastante significativas. Los departamentos no tuvieron diferencia con los clientes de cartera de tarjeta de crédito. Es decir, son en estos departamentos donde más clientes poseen tarjeta de crédito.

A partir de ahora se analiza al cliente retirador con variables de la institución:

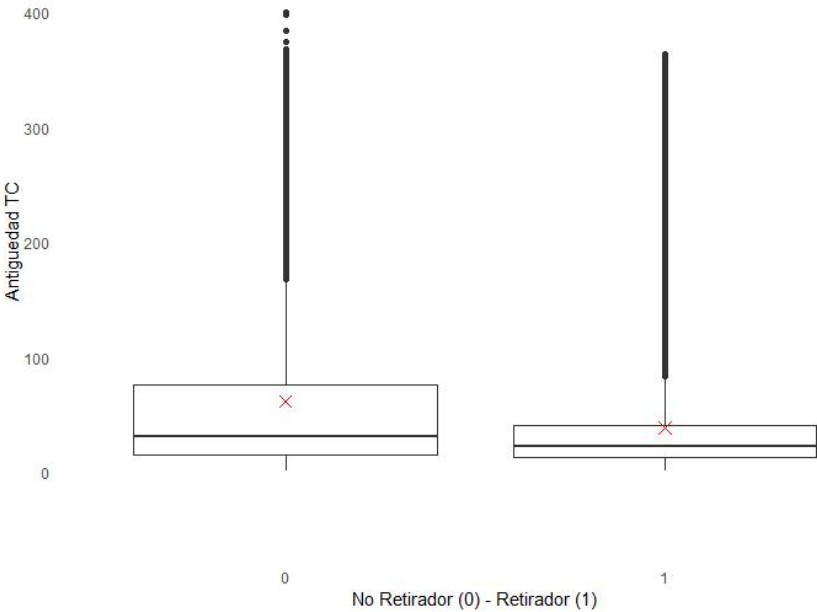
Ilustración 9. Categoría de tarjeta de crédito de retiradores



En la Ilustración 9 se presenta la categoría de tarjeta de crédito asociada a los clientes que realizan retiros de efectivo. Estas categorías agrupan los diferentes productos de tarjeta de crédito. En la gráfica se observa que la mayoría de los clientes retiradores utilizan tarjetas Clásicas (45 %), seguidas por las Oro (20 %) y Platino (16 %).

No obstante, al comparar esta distribución con la cartera general de clientes de tarjeta de crédito, se evidencia la siguiente diferencia: en la cartera, la categoría Platino ocupa el segundo lugar en participación, mientras que la categoría *Infinite* se posiciona en el cuarto lugar. Lo que significa que las categorías de tarjeta más utilizadas por los retiradores no son necesariamente las mismas que predominan en la cartera total de clientes.

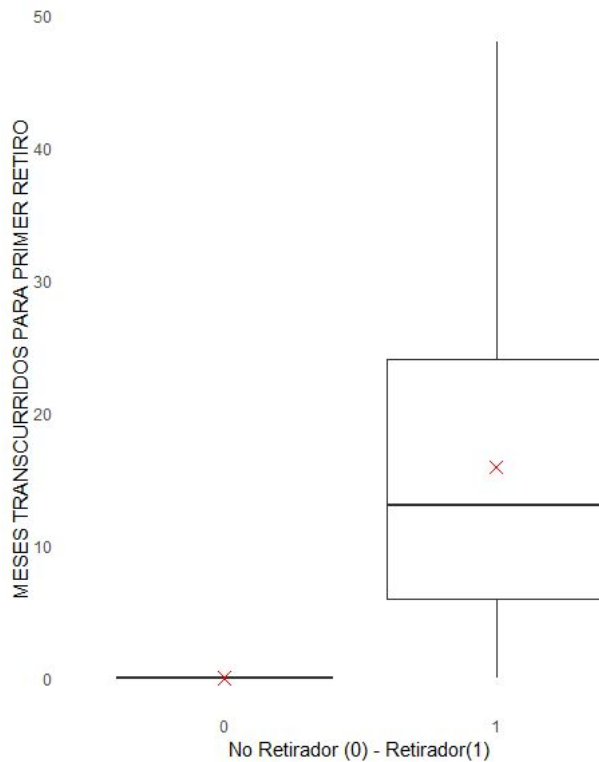
Ilustración 10. Comparación de la antigüedad de tarjeta de crédito de los clientes retiradores vs. no retiradores



En la Ilustración 10 se observa la comparación de la antigüedad de la tarjeta de crédito en meses para los clientes no retiradores y retiradores, son los meses transcurridos desde la emisión de la tarjeta de crédito en diciembre 2024. Se observa una diferencia entre los grupos. La media de antigüedad de la tarjeta de crédito de los clientes no retiradores es de 54 meses y de los retiradores de 33 meses. Lo que sugiere que, en promedio, los clientes no retiradores han tenido su tarjeta durante más tiempo que los retiradores. La mediana tiene una diferencia de 7 meses entre los grupos. La mediana de antigüedad de los clientes no retiradores es de 26 meses y de los retiradores es de 19 meses.

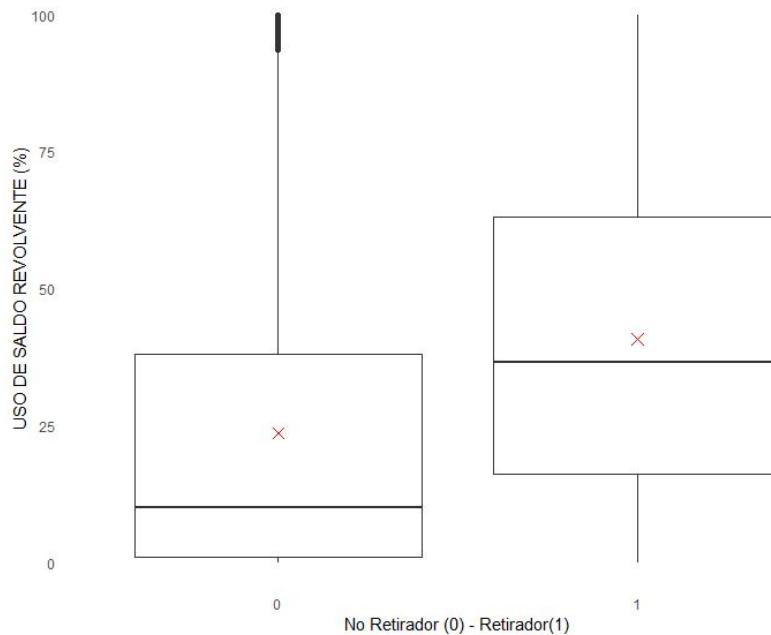
Este análisis sugiere que los clientes retiradores parecen retirar efectivo con menor tiempo de posesión de la tarjeta.

Ilustración 11. Meses transcurridos hasta el primer retiro



En esta gráfica se observan los meses transcurridos de los clientes retiradores desde la fecha de emisión de la tarjeta de crédito hasta su primer retiro. Esta variable tiene el propósito de definir en cuántos meses tarda aproximadamente el cliente en hacer su primer retiro de efectivo. El 25 % de los clientes tarda 7 meses, la mitad de los retiradores se tardan hasta 1 año y en promedio se tardan 18 meses.

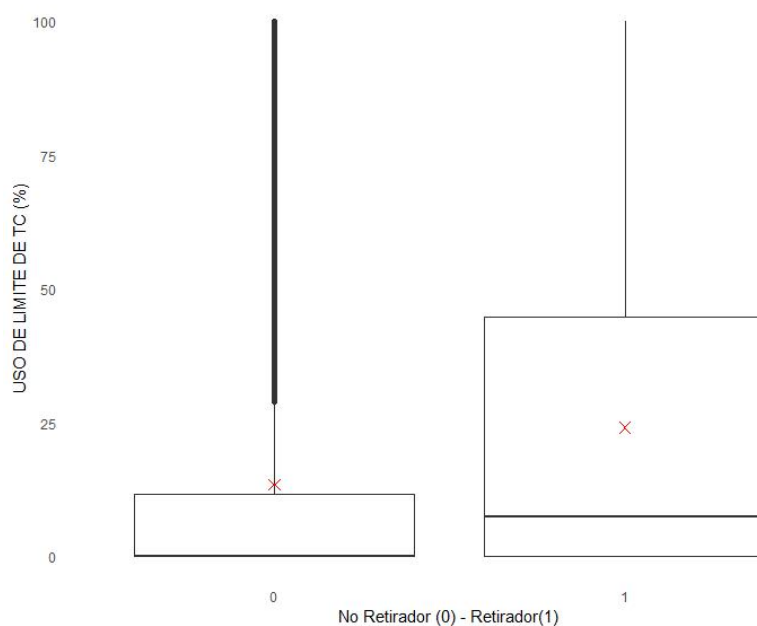
Ilustración 12. Comparación de uso de saldo revolvente para clientes retiradores vs. no retiradores



En esta gráfica se compara el uso de saldo revolvente para cada grupo de clientes, donde se observa una diferencia significativa para cada uno. El saldo revolvente refiere a uso del crédito en extrafinanciamiento, intrafinanciamiento y visacuotas. Los retiradores hacen un mayor uso del crédito, la mitad de ellos hace un uso de 28 % mientras que el 75 % de los no retiradores hace un uso menor, de 24 %.

Esto podría interpretarse como una menor necesidad de acceso a crédito por parte de los no retiradores. A diferencia de los clientes que retiran, quienes presentan una mayor necesidad de liquidez y hacen uso activo de los servicios de la institución.

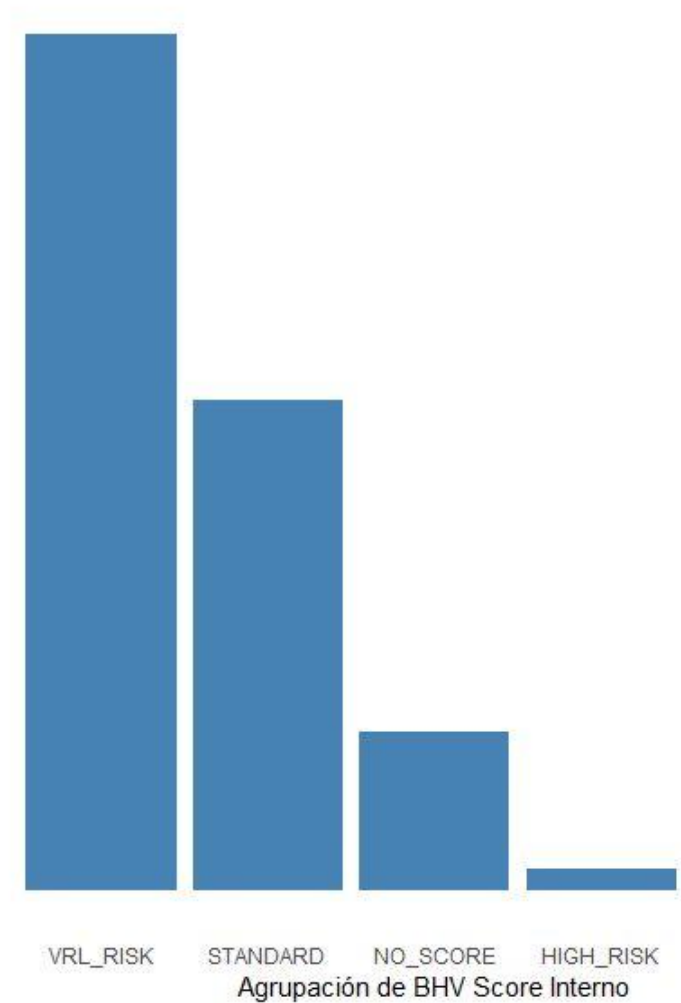
Ilustración 13. Comparación de uso límite de la tarjeta de crédito



En la Ilustración 13 se compara el uso del límite de la tarjeta de crédito por parte de los grupos. En dicha se observa la diferencia, el 75 % de los clientes no retiradores utilizan hasta el 10 % de crédito de la tarjeta, en contraste con los retiradores que el 75 % de clientes hace un 38 % de uso del crédito de la tarjeta. El promedio de los no retiradores es de 13 % y de los retiradores de un 25 %.

Esta visualización hace sentido con lo mencionado anteriormente, los retiradores hacen más uso del crédito otorgado por el banco.

Ilustración 14. *Behavior Score Interno* de clientes retiradores



La Ilustración 14 muestra la agrupación del *score* interno en la que se concentra la mayor cantidad de clientes retiradores. Este *score* interno corresponde a una calificación asignada por el banco basada en el comportamiento del cliente respecto al crédito otorgado, y permite clasificar a los clientes según distintos niveles de riesgo. La agrupación denominada *VRL_RISK* agrupa a los clientes con menor nivel de riesgo, es la que presenta la mayor proporción de retiradores de efectivo. El 56 % de clientes retiradores pertenecen a esta clasificación, 32 % a la clasificación *standard*, 10 % a *NO_SCORE*.

Ilustración 15. Días entre retiro de efectivo y fecha de pago de tarjeta de crédito



En la Ilustración 15 se analiza la variable de días transcurridos desde el retiro de efectivo y fecha de pago de la tarjeta de crédito. Considerando la posibilidad que algunos clientes retiran efectivo para pagar la tarjeta de crédito. En esta ilustración se observa que, en promedio, son dos días diferencia entre el retiro y la fecha de pago de tarjeta. La diferencia de la media y la mediana es de un día, la mediana siendo de tres días. Lo que indica que la mitad de los clientes retiran con al menos tres días de anticipación y el resto, tres días después. Esta diferencia entre la media y la mediana sugiere que hay una concentración de retiros cercanos a la fecha de pago.

A través de esta variable se pudo identificar a los clientes que retiran el mismo día de la fecha de pago de la tarjeta, siendo 69,000 clientes aproximadamente. A continuación, se identifica la cantidad aproximada de clientes con este comportamiento por año y el porcentaje que representan con el total de clientes retiradores:

Tabla 8. Cantidad de retiradores en fecha de pago de tarjeta de crédito

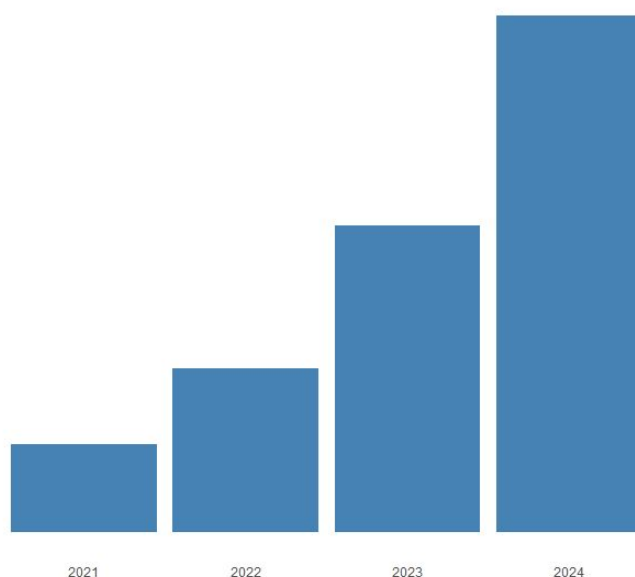
AÑO	CANTIDAD DE RETIRADORES EN FECHA DE PAGO DE TC	% DE CLIENTES RETIRADORES
2021	7,000	4.5%
2022	12,000	4.4%
2023	21,000	4.1%
2024	29,000	4.2%

En conclusión al análisis de clientes que retiran efectivo el mismo día de la fecha de pago de la tarjeta, es únicamente el 4 % de retiradores quienes retiran efectivo para pagar su tarjeta de crédito al contado y prefieren la tasa de interés que cobra la institución por retirar efectivo que el de la tarjeta de crédito.

Tras esta primera parte del análisis se identifican varias características y patrones para conocer a este grupo de clientes. A continuación, se procede a analizar las transacciones para comprender los hábitos de retiro y frecuencia de este servicio.

Ilustración 16. Transacciones anuales

Transacciones por año (2021-2024)



En la visualización se presenta como las transacciones de retiro han aumentado a lo largo de los años. El crecimiento de las transacciones muestra lo siguiente:

Tabla 9. Crecimiento anual de transacciones

AÑO	% DE CRECIMIENTO ANUAL DE RETIROS
2021	-
2022	86%
2023	87%
2024	68%

En 2022 hubo un crecimiento anual en retiros del 86 %, se mantuvo este crecimiento en 2023 con un aumento poco significativo de un punto porcentual. Sin embargo, en 2024 el crecimiento de retiros disminuyó, con un crecimiento de 68 %, siendo 19 % menos al comportamiento del año anterior.

Esto indica que la tendencia en el ritmo de crecimiento de retiros cayó significativamente, provocando una alerta para la institución. Por consiguiente, se busca prevenir que este cambio en la tendencia continúe y recuperar el comportamiento de los años anteriores.

El comportamiento anual en meses es el siguiente:

Ilustración 17. Transacciones por mes anual

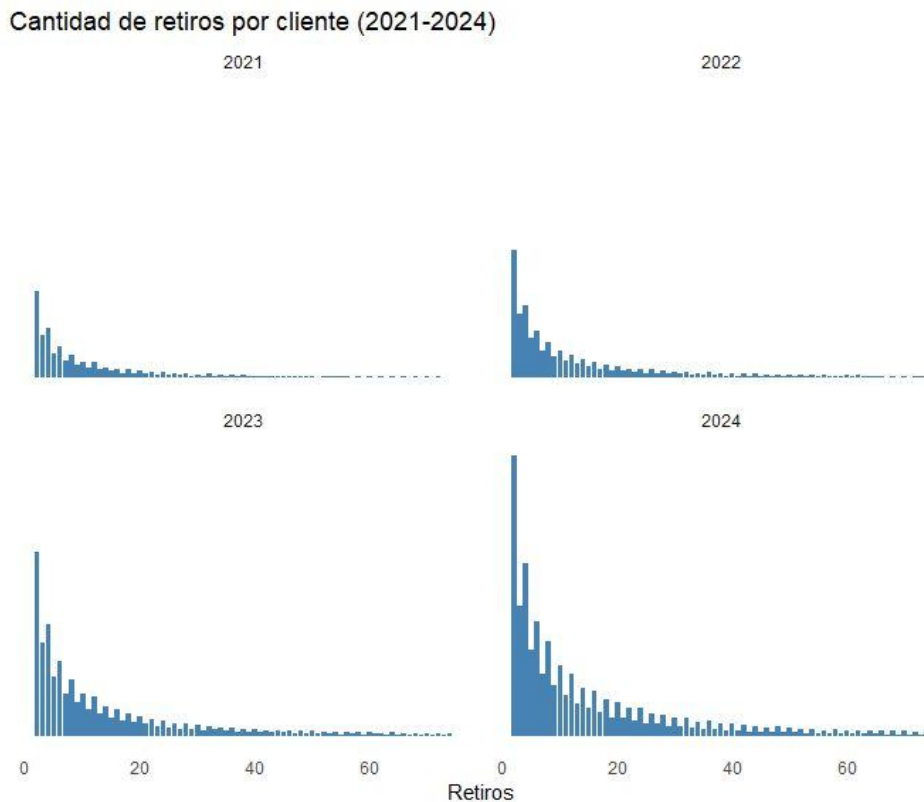
Transacciones por mes (2021-2024)



Las transacciones mensuales comparten el mismo comportamiento de clientes retiradores donde la cantidad de transacciones caen en julio y en diciembre. Estas caídas se deben a las bonificaciones mencionadas anteriormente.

A continuación se muestra la visualización de la cantidad de retiros totales por cada cliente:

Ilustración 18. Cantidad de retiros por cliente



En la Ilustración 18 se observa la cantidad de retiros que hizo el cliente cada año. Cada año muestra que la mayoría de los clientes hacen un retiro. Las gráficas reflejan un sesgo positivo, donde la mayoría de los clientes hacen pocos retiros al año.

En 2024, la mayoría de los clientes (25 %) realizaron entre uno y dos retiros. El 12.5 % hizo un retiro al año y sin mucha diferencia, el 12.3 % hizo dos retiros. Es interesante notar que existen clientes que hacen hasta 100 retiros anualmente.

En este siguiente apartado de análisis, se estimó el promedio de retiros mensuales por cliente al año. Se hizo una suma de los retiros que hizo el cliente por año y se dividió por la cantidad de meses que efectuó estos retiros. En este análisis se puede observar la frecuencia mensual de retiros por cliente.

Ilustración 19. Análisis de frecuencia mensual de retiros por cliente

COD_CLIENTE COD_CLIENTE	AÑO_CARTERA	retiros_totales_x_cliente_anualmente	meses	meses_c_retiro	retiros_x_cliente_mes
100001285	2021	0	12	0	0.000000
100001285	2022	4	12	3	1.333333
100001285	2023	1	12	1	1.000000
100001285	2024	0	12	0	0.000000
100001476	2021	13	12	6	2.166667
100001476	2022	33	12	11	3.000000
100001476	2023	17	12	8	2.125000
100001476	2024	4	12	3	1.333333
100001894	2021	0	8	0	0.000000
100001894	2022	0	12	0	0.000000
100001894	2023	0	12	0	0.000000
100001894	2024	0	12	0	0.000000

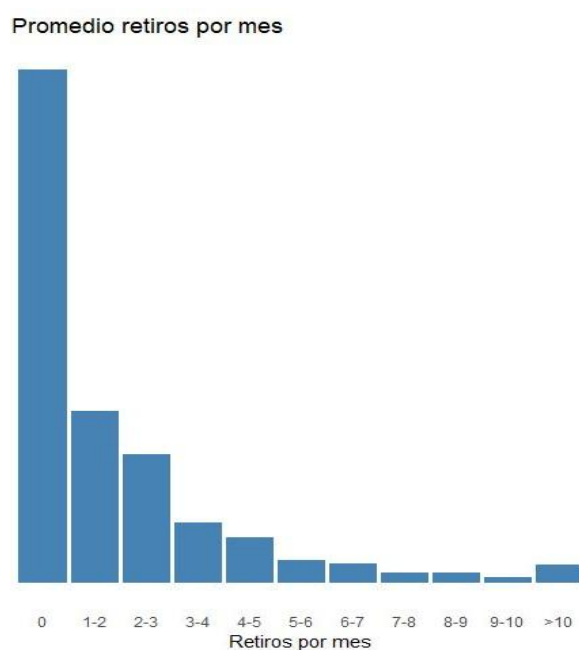
Además, se incluyó el monto promedio retirado por cada transacción.

Ilustración 20. Análisis de frecuencia mensual de retiros por cliente II

COD_CLIENTE COD_CLIENTE	AÑO_CARTERA	retiros_totales_x_cliente_anualmente	meses	meses_c_retiro	retiros_x_cliente_mes	monto_retirado_año	monto_retirado_x_retiro
100001285	2021	0	12	0	0.000000	0.00	NaN
100001285	2022	4	12	3	1.333333	4000.00	1000.00000
100001285	2023	1	12	1	1.000000	500.00	500.00000
100001285	2024	0	12	0	0.000000	0.00	NaN
100001476	2021	13	12	6	2.166667	3300.00	253.84615
100001476	2022	33	12	11	3.000000	11200.00	339.39394
100001476	2023	17	12	8	2.125000	4500.00	264.70588
100001476	2024	4	12	3	1.333333	400.00	100.00000
100001894	2021	0	8	0	0.000000	0.00	NaN
100001894	2022	0	12	0	0.000000	0.00	NaN
100001894	2023	0	12	0	0.000000	0.00	NaN
100001894	2024	0	12	0	0.000000	0.00	NaN

A través de este análisis se identifica a los clientes con mayor frecuencia de retiro al mes. Este análisis permite identificar a los clientes que hacen de 1 a 2 retiros mensuales. Es por esto que, se agruparon a los clientes en rangos de retiros. La visualización de estos retiros mensuales se presenta a continuación:

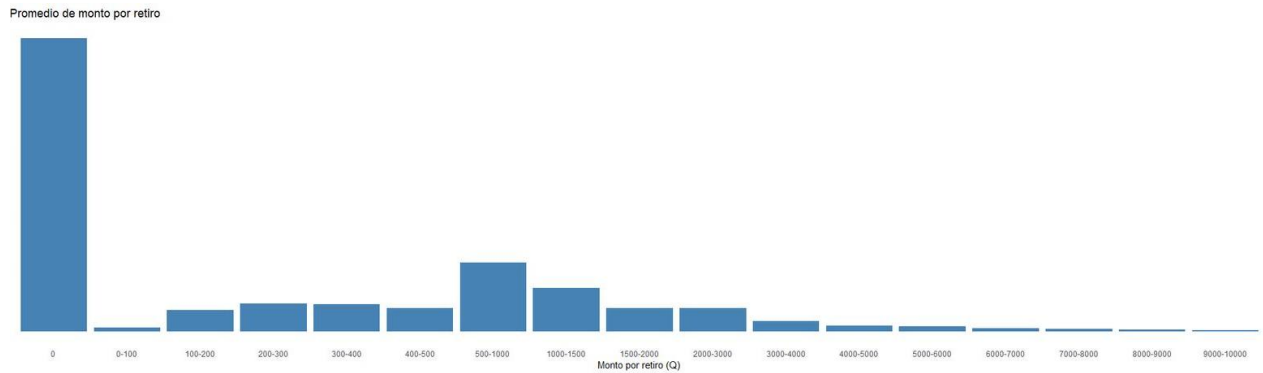
Ilustración 21. Promedio de retiros por mes



Se puede apreciar en la Ilustración 21 que la mayoría de los clientes hacen de 1-2 retiros mensuales, con poca diferencia de los clientes que hacen de 2-3 retiros. El 17 % de clientes realizan de 1-2 retiros, mientras el 13 % realiza de 2-3 retiros. Llama la atención que existen grupos de clientes que efectúan más de 10 retiros al mes.

Luego de analizar la frecuencia, resulta importante observar cuánto dinero retiran en promedio por transacción, para complementar la comprensión del comportamiento de los clientes retiradores.

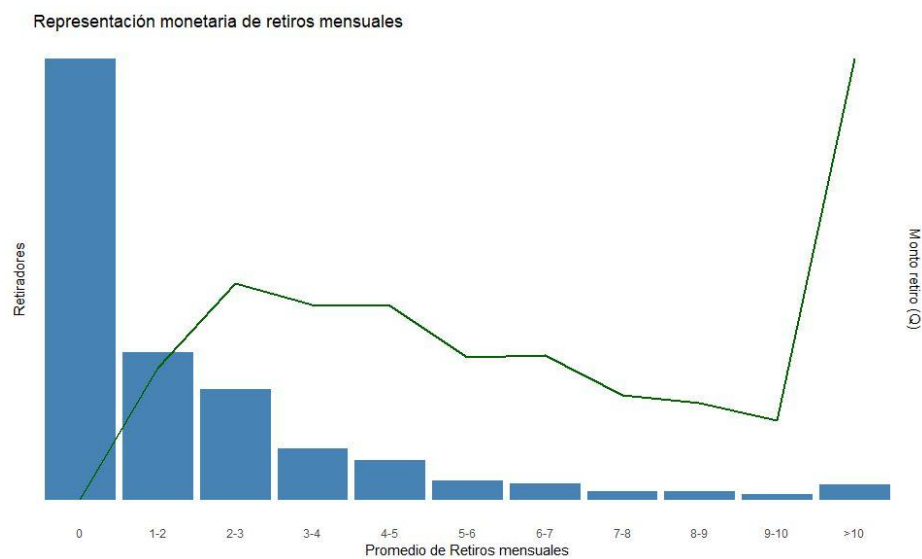
Ilustración 22. Promedio de monto por retiro



Como se muestra en la Ilustración 22 la mayoría de los clientes retiran en promedio de Q500-Q1500 por transacción. Este rango refleja un comportamiento moderado en cuanto al crédito disponible. Además, se observa que montos superiores son menos frecuentes, esto puede relacionarse con el límite de crédito o la intención de los clientes de evitar el pago de intereses altos.

En el siguiente gráfico se revela la relación entre transacciones y su representación monetaria. Con el fin de profundizar en el impacto económico de retiros por cada grupo.

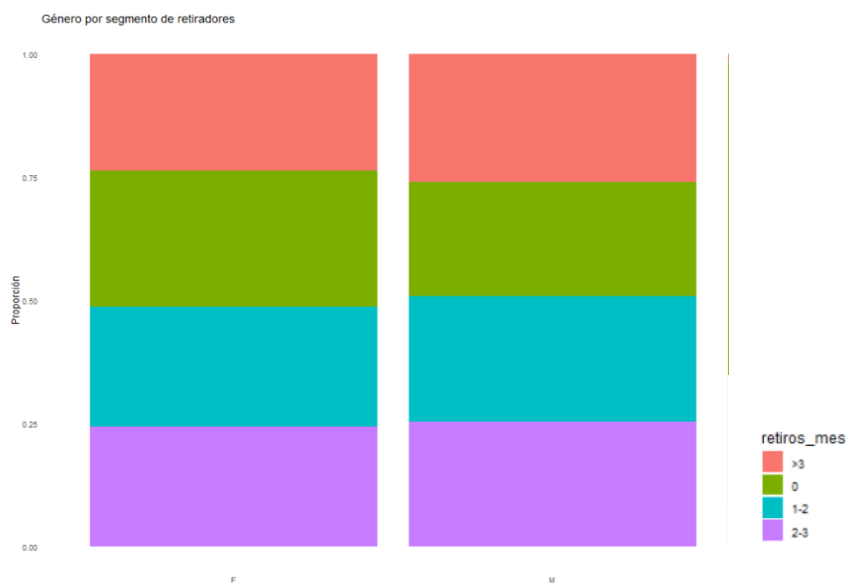
Ilustración 23. Representación monetaria de retiros mensuales



La Ilustración 23 pone en evidencia que los clientes que realizan más de 10 retiros al mes generan mayor impacto monetario, debido a la alta frecuencia de sus transacciones. El siguiente grupo con mayor aporte son los clientes que retiran de 2-3 veces al mes. Este grupo de clientes resulta relevante para la institución, ya que representa un buen equilibrio entre frecuencia de uso y volumen de retiro. Lo convierte en un segmento ideal para la institución y para el modelo de *machine learning*.

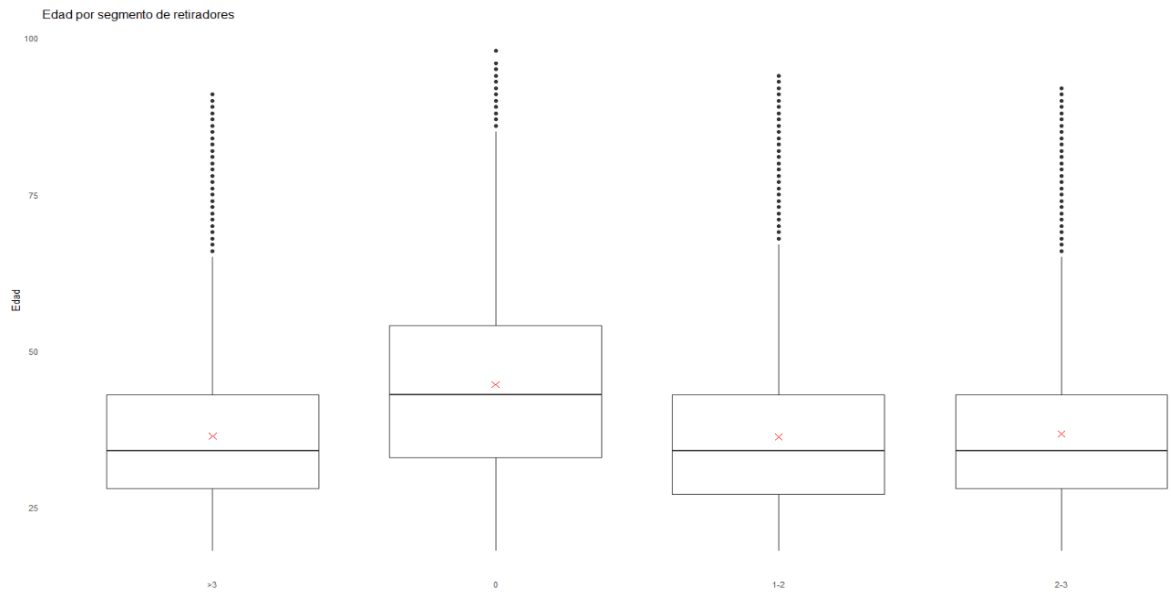
Por consiguiente, se realiza un último análisis para entender el comportamiento que puede presentar este y los grupos de clientes que retiran frecuentemente.

Ilustración 24. Género por segmento de retiradores



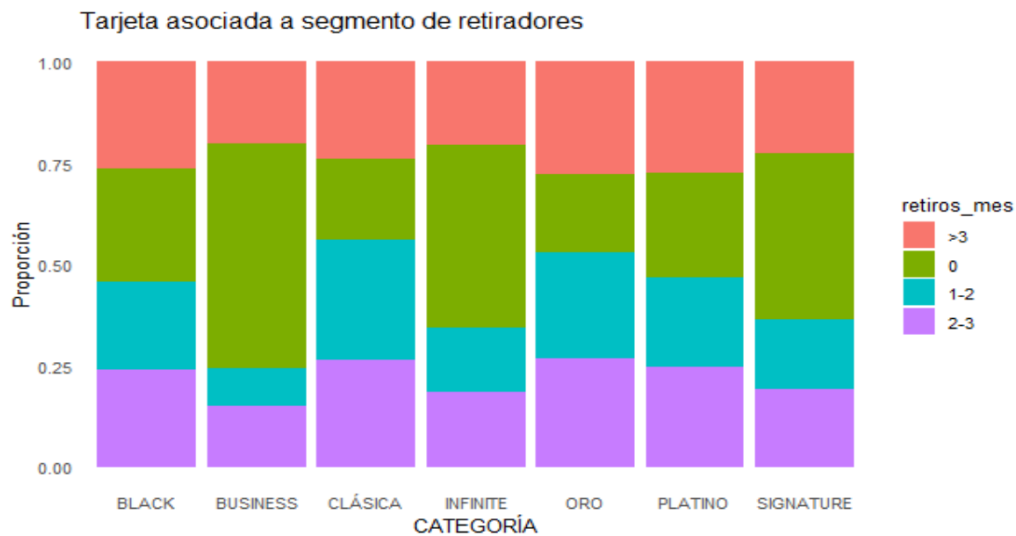
En la Ilustración 24 observamos como no existe variación significativa entre el género de cada grupo.

Ilustración 25. Edad por segmento de retiradores



La edad parece no influir en el segmento de retiradores, todos los grupos mantienen una edad promedio, mediana y cuartiles similares. Sin diferencias significativas. El grupo de clientes que revela una diferencia en la edad es el de los no retiradores, siendo clientes más grandes de edad.

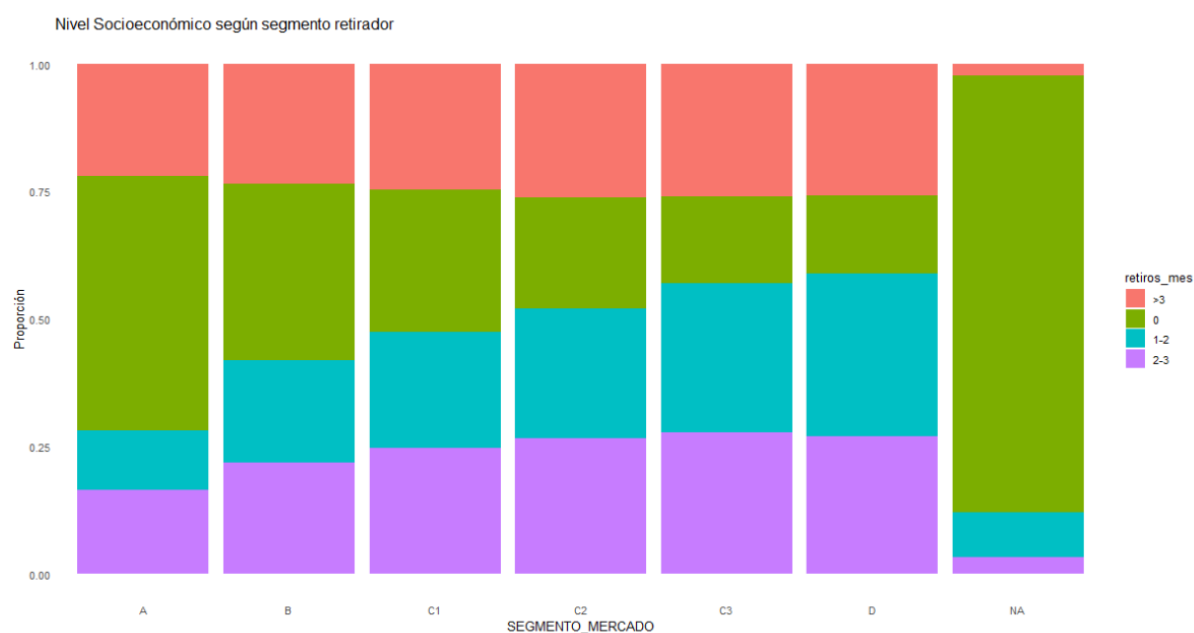
Ilustración 26. Tarjeta asociada a segmento de retiradores



La Ilustración 26 expone la relación entre el segmento de frecuencia de retiro y la categoría de tarjeta de crédito asociada a los clientes. Se observa que el grupo que realiza entre 1 y 2 retiros mensuales posee, en su mayoría, tarjetas de categoría Clásica y Oro. Mientras que, los clientes que efectúan entre 2-3 y mayor a 3 retiros al mes utilizan tarjetas Clásica, Oro, Platinum y Black. Ambos grupos muestran asociación a las mismas categorías de tarjetas.

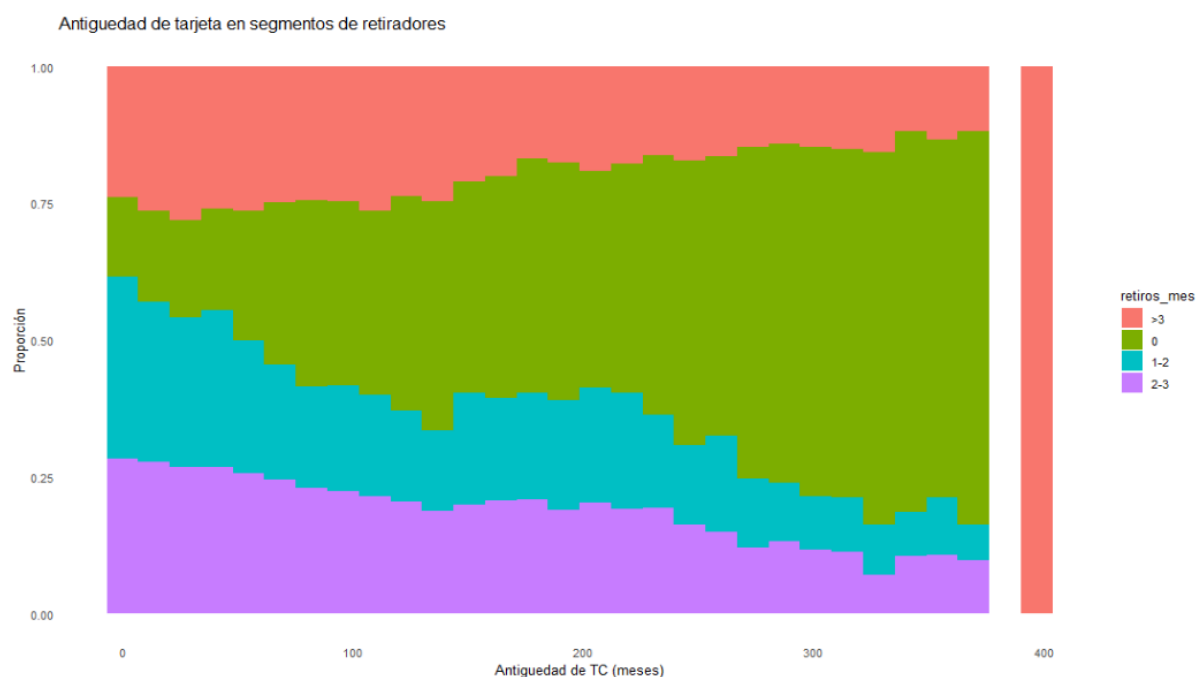
En esta visualización se muestra en contraste, como el grupo de no retiradores posee tarjetas de categorías superiores como Signature, Infinite y Business. Lo cual significa que los clientes no retiradores poseen un perfil más alto y los retiradores frecuentes un perfil bajo según categoría de tarjetas.

Ilustración 27. Segmento de mercado según segmento de retiradores



A partir de la ilustración 27 se puede validar el nivel socioeconómico de cada segmento de retiradores. Mostrando claramente como los clientes no retiradores son de nivel socioeconómico más alto y los retiradores más frecuentes de nivel más bajo. Mostrando una relación inversa. Sin embargo, es interesante como los clientes que efectúan más de tres retiros mensuales, permanecen en todos los niveles.

Ilustración 28. Antigüedad de tarjeta en segmentos retiradores



El propósito de la ilustración 28 es exponer cómo el tiempo del cliente con la tarjeta de crédito influye en la cantidad de retirados. Esta gráfica evidencia que, a menor antigüedad de la tarjeta de crédito, mayor cantidad de retirados. La relación de la antigüedad del cliente con la tarjeta de crédito y la cantidad de retirados está relacionada de forma inversa.

Tras un análisis extenso acerca de comportamientos, patrones y frecuencias de retiro de efectivo, se identifica al perfil de clientes retiradores. Se concluye que los clientes retiradores son hombres en un rango de edad de 24 a 30 años, de nivel socioeconómico bajo, empleados, que viven en la ciudad de Guatemala. Son clientes que utilizan el servicio de retiro de efectivo para pagar su tarjeta de crédito a tiempo con miedo a perder acceso a la tarjeta de crédito.

A partir de este análisis fue posible distinguir al cliente objetivo o *target* con mayor potencial para el banco, identificando un cliente retirador frecuente, que retira de 2 a 3 veces en promedio en el mes, retirando entre Q500-Q1500 por transacción. Este cliente es un nuevo tarjetahabiente de una tarjeta clásica.

Con estas breves descripciones, se construye una plantilla de buyer persona, esta permite personificar y visualizar al cliente ideal, basada en datos reales. Esta herramienta es clave para personalizar campañas de marketing y estrategias de comunicación, buscando una interacción más efectiva con el grupo objetivo.

Modelado y evaluación

En la fase de modelado se tiene como propósito desarrollar el modelo de aprendizaje supervisado de clasificación que pueda predecir qué clientes realizan de 1 a 2 retiros, de 2 a 3 y más de 3 retiros mensuales. Se utiliza la variable que identifica a estos clientes como la variable *target*. Los modelos de aprendizaje supervisados describen las relaciones entre las variables y la variable llamada *target*. Para que la relación hallada por el modelo sea de buen rendimiento, confiable y certera, se analiza la relación de todas las variables con la variable *target* antes de crear los modelos.

Si bien, las variables no muestran un tipo de relación con la variable *target*, se excluyen. Utilizando matrices de correlación, se excluyen hasta 19 variables.

La matriz de correlación que muestra las variables numéricas relacionadas para uso del modelo es la siguiente:

Tabla 10. Matriz de correlación final

VARIABLE	CORRELACIÓN
SEGMENTOS_RETIROS_MES	1
UTILIZACION_LIM	0.143377615
UTILIZACION_IF	0.124947348
Rangos_MON_PAGO_MIN	0.108907292
TASA_INTERES	0.092594158
Rangos_MON_PAGO_CONTADO	0.077843764
UTILIZACION_XF	0.076522094
DIAS_TRX_FECHAPAGO_TC	0.072464567
MESES_XF_COSECHATC	0.033080505
TIP_CLIENTE	0.032506353
Rangos_SUM_SALDO_CARTERA_CAH_GTQ	0.010630284
RATIO_SALDO_REVOLVENTE	0.007748324
Rangos_SUM_SALDO_CARTERA_CMO_GTQ	-0.008270867
PAGADORES_CONTADO	-0.036830694
CANT_MORAS	-0.047389579
ESTADO_CIVIL	-0.089092513
Rangos_OFERTA_IF	-0.099394138
LIMITE_CUENTA	-0.160006403
ANTIGUEDAD_COSECHA_MESES	-0.177789064
EDAD_REAL	-0.214965716

En esta tabla de correlación se puede apreciar el tipo de relación que tiene cada variable con la variable *target*: SEGMENTOS_RETIROS_MES, en orden descendente. Siendo las variables de uso de límite de crédito en la tarjeta, uso de intrafinanciamiento y los montos de pago mínimo a la tarjeta, las variables más relacionadas. Hace sentido ya que entre más retiros de efectivo realiza el cliente, utiliza más el crédito disponible e intrafinanciamiento de la tarjeta. Es interesante cómo el monto pago mínimo a la tarjeta influye en la cantidad de retiros según el nivel de la relación que se muestra. Este monto pago mínimo a la tarjeta puede significar que el cliente retira con intención de cumplir ese pago.

Las variables de tipo carácter que permanecen en el *dataset* son:

- CATEGORÍA
- SEXO
- DEPARTAMENTO
- SEGMENTO_MERCADO

En total, son 24 variables en el *dataset* para construir los modelos.

Como primer paso, se crea un conjunto de datos equilibrado para que las clases no estén desproporcionadas y se evite un sesgo en el modelo. Las clases son los grupos de clientes que no son retiradores, los que retiran de 1 a 2 veces, de 2 a 3 y de 3 en adelante. Se busca un equilibrio perfecto donde el set de datos tenga el mismo porcentaje de clases, en este caso, 20 % para representar cada uno. Este porcentaje nace a partir de crear el conjunto de datos con la misma cantidad de registros para cada clase, siendo la cantidad de registros del grupo con menos registros.

Tabla 11. Conjunto de datos desequilibrado

0	1-2	2-3	>3
6,319,390 registros	116,890 registros	105,847 registros	180,433 registros

Tabla 12. Conjunto de datos equilibrado

0	1-2	2-3	>3
105,847 registros	105,847 registros	105,847 registros	105,847 registros

Una vez esté equilibrado, se divide el set de datos en dos. Donde el set de datos con el 70 % de registros sea para entrenamiento del modelo y el 30 % para prueba del modelo. Este paso es importante para evaluar el rendimiento de los modelos. Los modelos de aprendizaje supervisado usan de *input* el conjunto de datos de entrenamiento y valida el rendimiento sobre el de prueba, esto para que el modelo clasifique sin tener la respuesta de la variable conocida, la variable *target*.

En la Tabla 13 se muestra la cantidad de registros divididos para el entrenamiento y prueba:

Tabla 13. Conjunto de datos equilibrado

Muestra para entrenamiento (70%)	Muestra para prueba (30%)
296,372	127,016

A continuación, se presentan los modelos desarrollados en el marco de este proyecto, junto con la evaluación de su rendimiento. El objetivo es determinar cuál de ellos ofrece un mejor desempeño para predecir con mayor certeza el tipo de cliente retirador. Se utilizan las matrices de confusión para medir el desempeño de cada uno.

Las matrices de confusión reflejan el *accuracy*, que refiere a la proporción de aciertos totales del modelo dividido por la cantidad de predicciones. El *recall* es la tasa de verdaderos positivos sobre el total de valores reales de una clase mientras que *precision* es la cantidad de aciertos sobre las predicciones de la clase. Estas métricas son importantes para decidir sobre el modelo.

1. Árbol de decisión

Se analiza la matriz de confusión del modelo de árbol de decisión. Se busca identificar los indicadores de desempeño que mejor se ajusta a la variable *target* y evaluar las diferentes combinaciones entre el valor de predicción y el real.

Ilustración 29. Matriz de confusión para modelo: árbol de decisión

Confusion Matrix and Statistics				
Prediction	Reference			
	0	1-2	2-3	>3
0	31754	348	389	385
1-2	0	20005	15956	10954
2-3	0	0	1696	0
>3	0	11401	13713	20415
Overall Statistics				
Accuracy : 0.5816				
95% CI : (0.5789, 0.5843)				
No Information Rate : 0.25				
P-Value [Acc > NIR] : < 2.2e-16				
Kappa : 0.4421				
McNemar's Test P-Value : < 2.2e-16				
Statistics by Class:				
	Class: 0	Class: 1-2	Class: 2-3	Class: >3
Sensitivity	1.0000	0.6300	0.05341	0.6429
Specificity	0.9882	0.7175	1.00000	0.7364
Pos Pred Value	0.9659	0.4264	1.00000	0.4484
Neg Pred Value	1.0000	0.8533	0.76015	0.8608
Prevalence	0.2500	0.2500	0.25000	0.2500
Detection Rate	0.2500	0.1575	0.01335	0.1607
Detection Prevalence	0.2588	0.3694	0.01335	0.3585
Balanced Accuracy	0.9941	0.6738	0.52671	0.6896

El *accuracy* de este modelo es de 58.16 %, refiriendo que clasifica correctamente el 58 % de todos los casos. El *recall* es medido a través de la medición: *sensitivity* en la matriz, donde la clase de No Retiradores (0) fue perfectamente clasificada. Sin embargo, nuestra clase con mayor interés (2-3 retiros), tuvo un 5 % de buenos aciertos en los valores reales. La tasa de precisión, reflejada como *Pos Pred Value* indica que la clase de 2-3 retiros fue perfectamente clasificada. No es contradictorio porque el modelo detectó muy pocos casos verdaderos de esta clase, lo que significa que falló en encontrarlos, pero los pocos que detectó, los detectó bien.

Basado en la relevancia de este grupo de clientes, descartamos el modelo por su bajo *recall* y su *accuracy* no significativo.

2. Naive Bayes

Se analiza la matriz de confusión del modelo de *Naive Bayes*. Se busca identificar los indicadores de desempeño que mejor se ajusta a la variable *target* y evaluar las diferentes combinaciones entre el valor de predicción y el real.

Ilustración 30. Matriz de confusión para modelo: *Naive bayes*

Confusion Matrix and Statistics				
Prediction	Reference			
	0	1-2	2-3	>3
0	23821	414	514	792
1-2	6355	23471	20229	17809
2-3	281	3248	5329	5515
>3	1297	4621	5682	7638
Overall Statistics				
Accuracy : 0.4744				
95% CI : (0.4717, 0.4772)				
No Information Rate : 0.25				
P-Value [Acc > NIR] : < 2.2e-16				
Kappa : 0.2992				
McNemar's Test P-Value : < 2.2e-16				
Statistics by Class:				
	Class: 0	Class: 1-2	Class: 2-3	Class: >3
Sensitivity	0.7502	0.7392	0.16782	0.24054
Specificity	0.9819	0.5340	0.90506	0.87823
Pos Pred Value	0.9327	0.3459	0.37076	0.39703
Neg Pred Value	0.9218	0.8600	0.76541	0.77624
Prevalence	0.2500	0.2500	0.25000	0.25000
Detection Rate	0.1875	0.1848	0.04196	0.06013
Detection Prevalence	0.2011	0.5343	0.11316	0.15146
Balanced Accuracy	0.8661	0.6366	0.53644	0.55938

El *accuracy* de este modelo es de 47.44 %, refiriendo que clasifica correctamente menos de la mitad de todos los casos. Es un *accuracy* que representa riesgo. El *recall* es medido a través de la medición: *sensitivity* en la matriz, donde las clases de No Retiradores y de 1 a 2 retiros fueron las mejores. Sin embargo, nuestra clase con mayor interés (2-3 retiros), tuvo un 16% de buenos aciertos en los valores reales. La tasa de precisión, reflejada como *Pos Pred Value* indica que las clases tuvieron un 35 % aproximadamente de aciertos.

Basado en el *accuracy* y tasas de precisión, descartamos este modelo.

3. Bosque aleatorio

Se analiza la matriz de confusión del modelo de bosque aleatorio. Se busca identificar los indicadores de desempeño que mejor se ajusta a la variable *target* y evaluar las diferentes combinaciones entre el

Ilustración 31. Matriz de confusión para modelo: *Random forest*

Confusion Matrix and Statistics				
Prediction	Reference			
	0	1-2	2-3	>3
0	31683	70	114	105
1-2	24	23844	8575	5076
2-3	21	4854	18249	4693
>3	26	2986	4816	21880
Overall Statistics				
Accuracy : 0.7531				
95% CI : (0.7507, 0.7555)				
No Information Rate : 0.25				
P-Value [Acc > NIR] : < 2.2e-16				
Kappa : 0.6708				
McNemar's Test P-Value : < 2.2e-16				
Statistics by Class:				
	Class: 0	Class: 1-2	Class: 2-3	Class: >3
Sensitivity	0.9978	0.7509	0.5747	0.6890
Specificity	0.9970	0.8564	0.8996	0.9178
Pos Pred Value	0.9910	0.6355	0.6560	0.7365
Neg Pred Value	0.9993	0.9116	0.8639	0.8985
Prevalence	0.2500	0.2500	0.2500	0.2500
Detection Rate	0.2494	0.1877	0.1437	0.1723
Detection Prevalence	0.2517	0.2954	0.2190	0.2339
Balanced Accuracy	0.9974	0.8037	0.7371	0.8034

El *accuracy* de este modelo es de 75.31 % refiriendo que clasifica correctamente el 75 % de todos los casos. Es un *accuracy* que no representa mucho riesgo. El *recall* es medido a través de la medición: *sensitivity* en la matriz, donde las clases reflejan una tasa por arriba del 50 % y una tasa de precisión arriba del 60 %.

La clase de no retiradores parece estar bien manejada por el modelo, presentando que detecta todos los casos reales con buena precisión. El segmento de clientes que realizan de 1-2 retiros, es 75 % bien detectado con una precisión de 63 %. El modelo identificó el 57 % de verdaderos positivos para el segmento de 2-3 retiros con una precisión 65 %. La última clase tiene el 68 % de verdaderos positivos con el 73 % de aciertos.

Este modelo parece presentar un buen rendimiento entre el conjunto de variables y la variable *target*. Considerando que tienen porcentajes altos, realistas sin sugerir sobreajuste.

4. XGBoost

Se analiza la matriz de confusión del modelo de XGBoost. Se busca identificar los indicadores de desempeño que mejor se ajusta a la variable *target* y evaluar las diferentes combinaciones entre el valor de predicción y el real.

Ilustración 32. Matriz de confusión para modelo: *XGBoost*

Confusion Matrix and Statistics				
Prediction	Reference			
	0	1-2	2-3	>3
0	31733	54	32	31
1-2	18	31272	639	1123
2-3	3	170	30950	249
>3	0	258	133	30351
Overall Statistics				
Accuracy : 0.9787				
95% CI : (0.9779, 0.9795)				
No Information Rate : 0.25				
P-Value [Acc > NIR] : < 2.2e-16				
Kappa : 0.9716				
McNemar's Test P-Value : < 2.2e-16				
Statistics by Class:				
	Class: 0	Class: 1-2	Class: 2-3	Class: >3
Sensitivity	0.9993	0.9848	0.9747	0.9558
Specificity	0.9988	0.9813	0.9956	0.9959
Pos Pred Value	0.9963	0.9461	0.9865	0.9873
Neg Pred Value	0.9998	0.9949	0.9916	0.9854
Prevalence	0.2500	0.2500	0.2500	0.2500
Detection Rate	0.2498	0.2462	0.2437	0.2390
Detection Prevalence	0.2508	0.2602	0.2470	0.2420
Balanced Accuracy	0.9991	0.9831	0.9851	0.9759

El *accuracy* de este modelo es de 97.87 % refiriendo que clasifica correctamente el 97 % de todos los casos. Es un *accuracy* que no representa riesgo. El *recall* es medido a través de la medición: *sensitivity* en la matriz, donde las clases reflejan una tasa por arriba del 95 % y una tasa de precisión arriba del 94 %.

Este modelo parece presentar un excelente rendimiento entre el conjunto de variables y la variable *target*. Sin embargo, porcentajes muy altos, podrían indicar existencia de sobreajuste en el modelo.

Tras el análisis de rendimiento de cada modelo, se procede a hacer una comparación de métricas de cada uno.

Ilustración 33. Comparación de *accuracy* de cada modelo

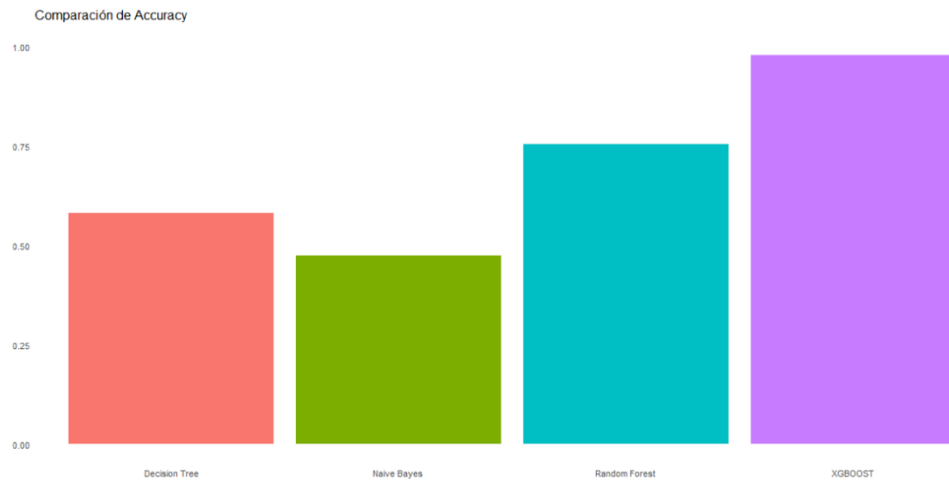
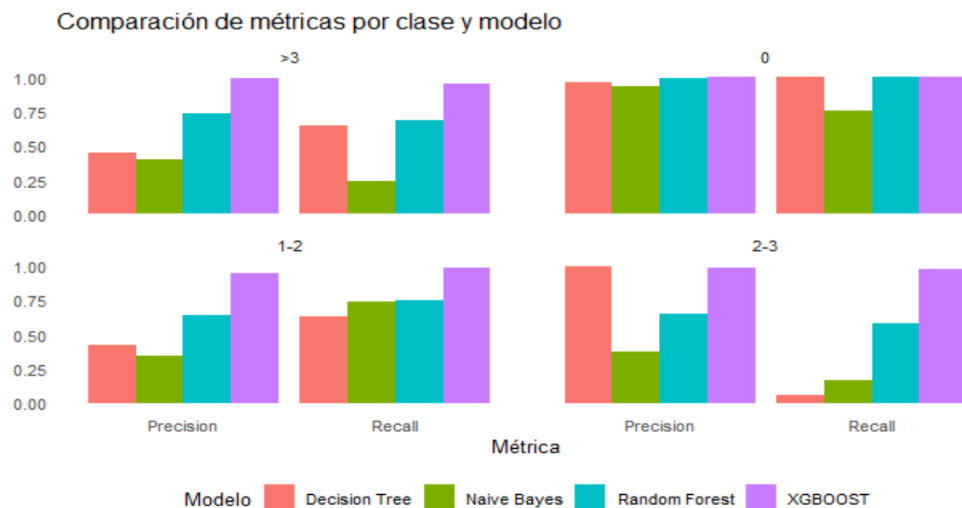


Ilustración 34. Comparación de métricas por clase y modelo



Las ilustraciones 33 y 34 permiten visualizar los resultados mencionados previamente, con el fin de comparar los modelos. Las métricas más altas son las del modelo de XGBOOST.

Para descartar que este modelo tiene sobreajuste y en definitiva es un buen modelo se hacen las siguientes validaciones:

1. Accuracy train > Accuracy test

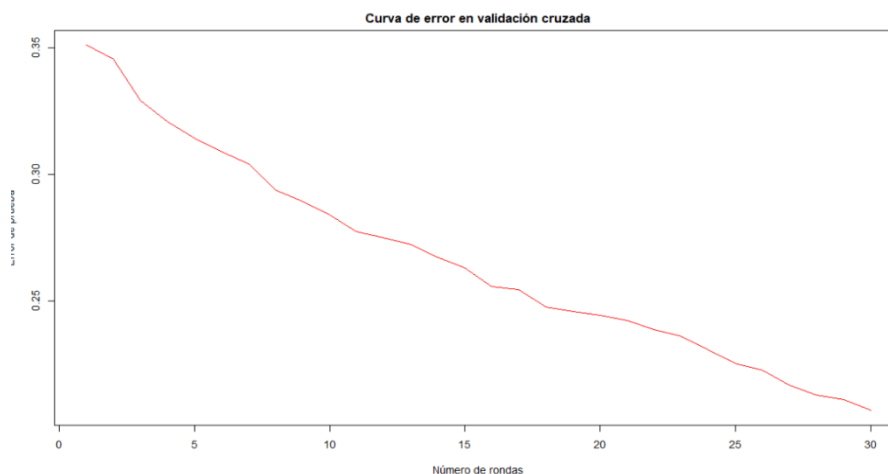
En esta validación se compara el rendimiento del modelo en entrenamiento y prueba. Si el *accuracy* del modelo en entrenamiento es mayor al de prueba, existe sobreajuste. Se

procedió a hacer a verificación y sí cumple con el criterio, sin embargo, es mayor al *accuracy* de prueba por una diferencia del 0.00091 por lo que, no significa un sobreajuste seguro.

2. Validación cruzada

En esta validación se implementa la validación cruzada, conocida con *k-fold cross validation*. Esta validación es un proceso que divide el conjunto de datos en subconjuntos, llamados pliegues, para entrenar el modelo con ellos. Este proceso se repite para evaluar el modelo en los distintos pliegues, siendo las veces necesarias para observar un comportamiento de sobreajuste.

Ilustración 35. Curva de error en validación cruzada



En esta validación no se observa algún comportamiento de sobreajuste en 30 repeticiones. El error tiene tendencia a descender según la cantidad de veces que el modelo es entrenado. Esto significa que el modelo está aprendiendo y mejorando su capacidad de clasificar predicciones. De lo contrario, si la línea tiene tendencia de aumentar, es señal de sobreajuste. Se descarta la posibilidad de existencia de sobreajuste en este modelo.

Para fines académicos, se elige el modelo XGBoost como modelo final. Sugiere que es el modelo perfecto para clasificar a los segmentos de clientes retiradores y no retiradores. Sin embargo, para fines prácticos se elige el modelo de Bosque Aleatorio, el rendimiento de este modelo está más apegado a la realidad. Los modelos implementados en la institución presentan métricas del 70 % por lo que es razonable y confiable.

Tras finalizar la metodología CRISP-DM donde se transformaron, comprendieron, analizaron datos para la construcción de modelos de aprendizajes supervisados, se evaluaron y se decidió el modelo a utilizar, procedemos a diseñar el plan de marketing para este servicio.

Plan de *Marketing*

El plan de *marketing*: “Pirámide de retiradores frecuentes” tiene como objetivo aumentar el 20 % de crecimiento anual de retiros de efectivo en un año. Esto se debe a la caída del 19 % en el ritmo de crecimiento de retiros efectuados en 2024. El método es segmentación de retiradores frecuentes de efectivo identificados a través de un modelo de aprendizaje supervisado para comunicaciones de SMS. Del presente plan se espera una mayor tasa de conversión de las comunicaciones de SMS que se envían a los tarjetahabientes.

El grupo objetivo de este plan de marketing son los retiradores que hacen entre 2 a 3 transacciones en promedio al mes. Este es el grupo objetivo debido a su mayor impacto monetario para la institución. Este grupo objetivo se describe a continuación:

Los retiradores que hacen de 2 a 3 transacciones en promedio al mes son clientes con tarjeta de crédito en un rango de edad de 24 a 30 años, residentes en la ciudad de Guatemala, con un perfil socioeconómico bajo. Son tarjetahabientes recientes de categorías bajas: CLASICAS Y ORO que utilizan el crédito otorgado por el banco en su tarjeta de crédito, con un uso de saldo revolvente del 30 % y una calificación de crédito *VERY LOW RISK*. Su principal método de pago es el efectivo y es propietario de tarjeta de crédito para cubrir emergencias.

Se desarrolló la personificación del grupo objetivo, el *buyer persona*, llamado Kevin Ramírez.

Ilustración 36. *Buyer persona*

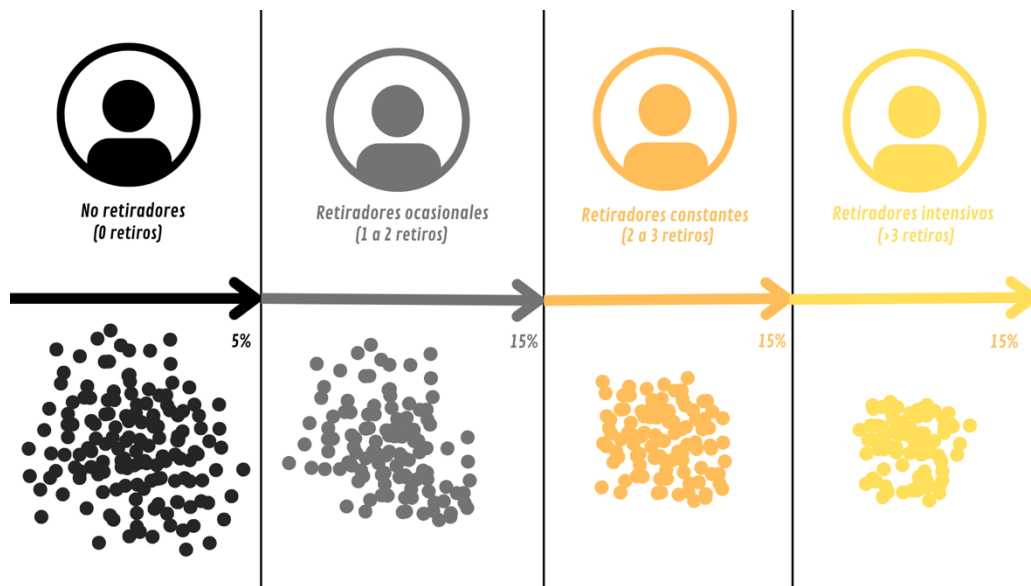
KEVIN RAMIREZ		Demográfica	Psicográfica
		Kevin Ramirez es un joven de 29 años que vive en la ciudad de Guatemala. Es un hombre soltero que aún vive en casa de sus padres. Es empleado de una institución pequeña, con un trabajo operativo.	Kevin es una persona social. Intenta ser responsable pero tiene dificultad para organizarse. Kevin es práctico, no le gusta complicarse. Es resiliente y trabajador, confía en sí mismo para salir adelante. Sus motivaciones son apoyar a su familia, ser independiente económicamente y desea emprender de forma informal para aumentar ingresos.
EDAD 29 años SEXO Hombre CIUDAD Guatemala EDUCACIÓN Graduado de diversificado OCUPACIÓN Operativo ESTADO CIVIL Soltero INGRESO Q3,000-Q4,000	Cliente Es tarjetahabiente de una tarjeta de crédito clásica desde hace dos años. Es un cliente con buen score interno, siendo de clasificación <i>VLR RISK</i>. Hace uso del 25% de crédito de la tarjeta y de 36% del saldo revolving.	Hábitos de compra Hace sus compras buscando las opciones más accesibles. Kevin utiliza su tarjeta de crédito para emergencias o imprevistos. Retira efectivo con su tarjeta de crédito porque utiliza efectivo para sus compras y también para pagar su tarjeta de crédito con miedo a que se la bloqueen.	
Gustos y aficiones - Jugar fútbol - Ver partidos de fútbol - Pasar tiempo con sus amigos - Su entretenimiento son las redes sociales	Objetivos Kevin quiere ahorrar para comprar una motocicleta y cambiar celular.	Retos Los retos de Kevin están relacionados con su inestabilidad económica. Gana lo justo para sobrevivir y no logra ahorrar. Tiene pocas oportunidades para crecer laboralmente.	

Los objetivos de este plan de marketing son:

- Impulsar el 20 % de crecimiento anual de retiros de efectivo en un año.
- Transformar al 5 % del segmento de no retiradores a efectuar su primer retiro de efectivo en 6 meses.
- Incentivar al 15 % de clientes que han realizado un retiro a efectuar el segundo en 3 meses.
- Incentivar al 15 % de clientes que han realizado dos retiros a efectuar el tercero en 3 meses.
- Incentivar al 15 % de clientes que han realizado tres retiros a efectuar más retiros en 3 meses.

La estrategia del plan es basada en la siguiente pirámide:

Ilustración 37. Pirámide de retiradores frecuentes



En esta estrategia buscamos incentivar a cada segmento de clientes, identificados a través del modelo, a realizar una transacción adicional. Estos incentivos consisten en promociones y descuentos para convertir al segmento en un retirador más frecuente. Estas promociones son distintas para cada segmento y se van a comunicar a través de mensaje de texto. El objetivo es convertir a los clientes en retiradores de 2 a 3 transacciones promedio. No obstante, esta estrategia va más allá: también busca reconocer y premiar a aquellos clientes que superan más de 3 transacciones al mes con el fin de fidelizarlos.

Las promociones pueden ser:

- Comisión más baja para el siguiente retiro
- Ofrecer sin costo una cantidad de dinero en el siguiente retiro
- Promoción basada en un mínimo de consumo
- Promoción basada en un mínimo de retiros
- Participación en sorteos al hacer un nuevo retiro
- Conceder una comisión de gracia en un nuevo retiro

A los no retiradores se les ofrecerá participar en esta dinámica de promociones luego de hacer su primer retiro de efectivo como incentivo inicial. Las comunicaciones de SMS

ofreciendo las promociones pueden ser automáticas luego del primer retiro, esperando la acción de retiro del cliente para ofrecer la siguiente.

El presupuesto para esta campaña depende de las promociones aprobadas por la institución, sin embargo, se presenta a continuación un escenario representativo.

Ilustración 38. Presupuesto representativo

Mes	1	2	3	4	5	6	7	8	9	10	11	12
Accuracy	75%	75%	75%	75%	75%	75%	75%	75%	75%	75%	75%	75%
Cantidad de retiradores												
0	39.709,00	35.501,00	40.072,00	39.866,00	34.366,00	30.201,00	27.454,00	41.182,00	26.326,00	41.330,00	27.382,00	42.834,00
1a2	32.414,00	35.873,00	30.957,00	45.585,00	34.692,00	39.771,00	35.184,00	28.986,00	28.618,00	34.089,00	31.941,00	46.996,00
2a3	38.304,00	42.156,00	41.049,00	45.659,00	46.251,00	38.480,00	26.938,00	40.025,00	33.785,00	28.503,00	38.573,00	35.965,00
>3	31.743,00	28.241,00	28.355,00	37.434,00	45.244,00	28.688,00	43.983,00	25.578,00	43.898,00	47.484,00	31.194,00	35.822,00
Ticket promedio												
0 Q	- Q	- Q	- Q	- Q	- Q	- Q	- Q	- Q	- Q	- Q	- Q	-
1a2 Q	500 Q	500 Q	500 Q	500 Q	500 Q	500 Q	500 Q	500 Q	500 Q	500 Q	500 Q	500
2a3 Q	1.000 Q	1.000 Q	1.000 Q	1.000 Q	1.000 Q	1.000 Q	1.000 Q	1.000 Q	1.000 Q	1.000 Q	1.000 Q	1.000
>3 Q	1.500 Q	1.500 Q	1.500 Q	1.500 Q	1.500 Q	1.500 Q	1.500 Q	1.500 Q	1.500 Q	1.500 Q	1.500 Q	1.500
Ingresos												
0 Q	- Q	- Q	- Q	- Q	- Q	- Q	- Q	- Q	- Q	- Q	- Q	-
1a2 Q	16.207.000 Q	17.936.500 Q	15.478.500 Q	22.792.500 Q	17.346.000 Q	19.885.500 Q	17.592.000 Q	14.493.000 Q	14.309.000 Q	17.044.500 Q	15.970.500 Q	23.498.000
2a3 Q	38.304.000 Q	42.156.000 Q	41.049.000 Q	45.659.000 Q	46.251.000 Q	38.480.000 Q	26.938.000 Q	40.025.000 Q	33.785.000 Q	28.503.000 Q	38.573.000 Q	35.965.000
>3 Q	47.614.500 Q	42.361.500 Q	42.532.500 Q	56.151.000 Q	67.866.000 Q	43.032.000 Q	65.974.500 Q	38.367.000 Q	65.847.000 Q	71.226.000 Q	46.791.000 Q	53.733.000
Total Ingresos Q	102.125.500 Q	102.454.000 Q	99.060.000 Q	124.602.500 Q	131.463.000 Q	101.397.500 Q	110.504.500 Q	92.885.000 Q	113.941.000 Q	116.773.500 Q	101.334.500 Q	113.196.000
Costo Unitario												
SMS Q	0,5 Q	0,5 Q	0,5 Q	0,5 Q	0,5 Q	0,5 Q	0,5 Q	0,5 Q	0,5 Q	0,5 Q	0,5 Q	0,5
Descuento 1 (%)	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5
Descuento 2 (%)	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75
Descuento 3 (%)	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
Descuento 4 (%)	1	1	1	1	1	1	1	1	1	1	1	1
Costo clientes efectivos												
0 Q	4.964 Q	4.438 Q	5.009 Q	4.983 Q	4.296 Q	3.775 Q	3.432 Q	5.148 Q	3.291 Q	5.166 Q	3.423 Q	5.354
1a2 Q	12.167.405 Q	13.465.827 Q	11.620.484 Q	17.111.469 Q	13.022.510 Q	14.929.039 Q	13.207.194 Q	10.880.620 Q	10.742.482 Q	12.796.158 Q	11.989.853 Q	17.641.124
2a3 Q	19.166.364 Q	21.093.809 Q	20.539.893 Q	22.846.622 Q	23.142.844 Q	19.254.430 Q	13.479.102 Q	20.027.509 Q	16.905.169 Q	14.262.189 Q	19.300.965 Q	17.995.987
>3 Q	47.626.404 Q	42.372.090 Q	42.543.133 Q	56.165.038 Q	67.882.967 Q	43.042.758 Q	65.990.994 Q	38.376.592 Q	65.863.462 Q	71.243.807 Q	46.802.698 Q	53.746.433
total de clientes efectivos Q	78.965.137 Q	76.936.164 Q	74.708.519 Q	96.128.113 Q	104.052.616 Q	77.230.002 Q	92.680.721 Q	69.289.869 Q	93.514.404 Q	98.307.320 Q	78.096.938 Q	89.388.898
Costo clientes inefectivos												
0	4963,625	4437,625	5009	4983,25	4295,75	3775,125	3431,75	5147,75	3290,75	5166,25	3422,75	5354,25
1a2	4051,75	4484,125	3869,625	5698,125	4336,5	4971,375	4398	3623,25	3577,25	4261,125	3992,625	5874,5
2a3	4788	5269,5	5131,125	5707,375	5781,375	4810	3367,25	5003,125	4223,125	3562,875	4821,625	4495,625
>3	3967,875	3530,125	3544,375	4679,25	5655,5	3586	5497,875	3197,25	5487,25	5935,5	3899,25	4477,75
tal de clientes inefectivos Q	17.771 Q	17.721 Q	17.554 Q	21.068 Q	20.069 Q	17.143 Q	16.695 Q	16.971 Q	16.578 Q	18.926 Q	16.136 Q	20.202
Costo total Q	78.982.908 Q	76.953.885 Q	74.726.074 Q	96.149.181 Q	104.072.685 Q	77.247.145 Q	92.697.416 Q	69.306.840 Q	93.530.982 Q	98.326.246 Q	78.113.075 Q	89.409.100
Utilidades Q	23.142.592 Q	25.500.115 Q	24.333.927 Q	28.453.320 Q	27.390.315 Q	24.150.355 Q	17.807.084 Q	23.578.160 Q	20.410.018 Q	18.447.255 Q	23.221.426 Q	23.796.900

En este presupuesto se estima la cantidad de clientes efectivos por medio del porcentaje de *accuracy* del modelo y los no efectivo (1- *accuracy*).

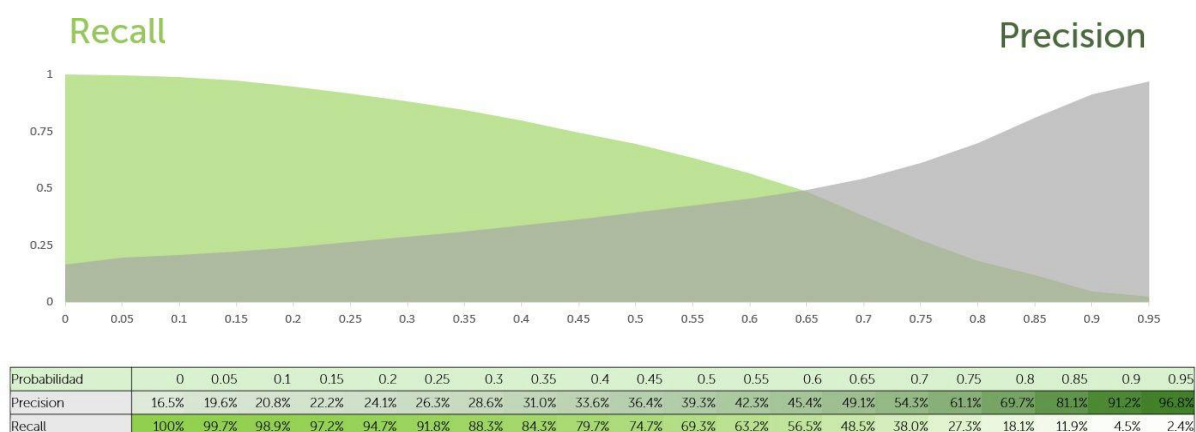
El retorno de inversión (ROI) anual estimado para esta propuesta es del 27.22 %, lo cual significa una expectativa positiva en términos de impacto financiero. Para evaluar los resultados de la campaña, se utilizará esta métrica como indicador clave de rentabilidad. Además, se medirá la tasa de conversión por segmento, con el propósito de identificar el nivel de efectividad alcanzado en cada grupo objetivo.

VI. RESULTADOS

Los resultados esperados tras la implementación del modelo y la campaña están relacionados entre sí. Si bien el modelo tiene resultados positivos, se espera que la campaña de marketing lo refleje. Este proyecto tiene como fin tener impacto positivo en la cantidad de transacciones y retiros de efectivo, con la expectativa de cambiar las tendencias de ritmo de crecimiento para retiro de efectivo desde los primeros tres meses de implementación.

La proyección de resultados del modelo con un *accuracy* del 75% es el siguiente:

Ilustración 39. Proyección de resultados de bosque aleatorio



Esta gráfica permite observar la proyección del comportamiento de *recall* y *precision* del modelo de bosque aleatorio. El comportamiento de *recall* sugiere que el modelo detecta casi todos los verdaderos positivos con un umbral de *precision* que aumenta, refiriendo que comete menos errores, menos falsos positivos.

Este análisis refleja la precisión del modelo y la cantidad de casos positivos que detectará al implementarlo. Es importante para estimar la precisión de las comunicaciones. De las comunicaciones se espera una conversión del 5 % de cada segmento por mes.

Por consiguiente, la expectativa de este proyecto es la optimización de comunicaciones dirigidas hacia el cliente de tarjeta de crédito para ofrecer el servicio de retiro de efectivo. Esta optimización le permitirá a la institución en dirigir sus esfuerzos de comunicación en clientes con potencial a utilizar el servicio.

VII. CONCLUSIONES

Para finalizar, se concluye que a través de un proceso de *data mining* se pueden diseñar estrategias de marketing que permiten la optimización de comunicaciones hacia el cliente por medio de segmentación. El análisis de datos es una herramienta fundamental para definir segmentos y la implementación de un modelo de *machine learning* es clave identificarlos. Además, a través de este proceso se pudo detectar el problema real del servicio y la oportunidad.

El problema del retiro de efectivo no es evidente cuando los números y gráficas de colocación tienen tendencia positiva, este se detectó al observar la tendencia de ritmo de crecimiento, por lo que, este proyecto tiene como principal objetivo detener y cambiar esta tendencia antes de que sea tarde. Este hallazgo no se hubiese reconocido si no fuese por el análisis histórico de datos.

El grupo objetivo fue hallado con un análisis profundo de frecuencia de retiro y de impacto monetario, sin ser los clientes que más retiros hacen. La Ilustración 23 de impacto monetario sobre la frecuencia de retiro dejó de evidencia que son los clientes que realizan de 2 a 3 retiros promedio mensuales quienes otorgan un mayor beneficio al banco.

El perfil del grupo objetivo representa a los clientes retiradores, sin diferencias de características de edad, sexo, tarjeta de crédito, antigüedad como cliente y nivel socioeconómico.

Al querer identificar las características del cliente retirador, se distinguió el perfil del cliente no retirador. El cliente no retirador tiene un perfil evidentemente contrario. Las características más relucientes de distinción son: antigüedad de tarjeta de crédito, es un cliente de muchos años con la TC, de nivel socioeconómico alto en posesión de tarjetas de categorías altas y con poco uso del crédito otorgado por la institución.

En resumen, el poder de la analítica es clave para componentes de marketing. Sin lugar a duda, el análisis profundo permitió la comprensión del negocio, conocer al cliente, identificar al cliente ideal y la implementación de una estrategia de *leads* y fidelización de los clientes ya existentes. Es una herramienta que ofrece mucho valor para crear una estrategia de marketing, entregando una propuesta relevante por su confiabilidad de información.

VIII. RECOMENDACIONES

1. Implementar una campaña para concientizar al cliente sobre el uso del retiro de efectivo en los primeros meses de posesión de tarjeta de crédito.
2. Entrenar el modelo con datos actuales y validar su rendimiento antes de implementarlo.
3. Evaluar distintas combinaciones de promociones para elegir la que mejor impacto financiero represente.
4. Excluir a los clientes no retiradores con alta precisión identificados del modelo de las comunicaciones.
5. Actualizar el modelo periódicamente para darle mantenimiento.
6. Investigar cuáles son las motivaciones y factores decisivos del cliente para retirar.
7. Conocer el grupo objetivo a través de investigaciones cualitativas para mejorar el *buyer persona*.
8. Hacer un análisis de los canales de retiro: cajeros automáticos y agencias.

IX. BIBLIOGRAFÍA

- Bloomenthal, A.** (2023, 15 diciembre). *Credit Card: What It Is, How It Works, and How to Get One*. Investopedia. <https://www.investopedia.com/terms/c/creditcard.asp> [1]
- D Peng, R.** (2016). *Exploratory data analysis with R*. Leanpub.
- Haya, P.** (2022, 5 enero). *La metodología CRISP-DM en ciencia de datos*. Instituto de Ingeniería del Conocimiento. <https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/> [3]
- Provost, F., & Fawcett, T.** (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. [5, 6, 7, 8, 10]
- Sistema de Analisis Estadistico con SPSS.** (n.d.). (n.p.): IICA Biblioteca Venezuela. [9]
- Staff, C.** (2023, 29 noviembre). *Estrategia de marketing: Qué es y cómo crearla*. Coursera. <https://www.coursera.org/mx/articles/marketing-strategy> [4]
- Torres, Y.** (2022). *Si retiras dinero en efectivo de tu tarjeta de crédito, conoce cuánto le pagarás al banco*. *El Economista*. <https://www.eleconomista.com.mx/finanzaspersonales/Si-retiras-dinero-en-efectivo-de-tu-tarjeta-de-credito-conoce-cuanto-le-pagaras-al-banco-20220203-0071.html> [2]