

Proyecto Ciencia de Datos

Jesús Abraham BAUTISTA LÓPEZ¹, Alejandra Irene HINOSTROZA CALDAS², Juan Saidt RÍOS GARCÍA³

Temas Selectos de Física Computacional 2. Semestre 2023-2

Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad Universitaria, Av. Universidad # 3000, Ciudad de México, 04510, México

Resumen

En un panorama donde la fidelidad del cliente es un objetivo crucial para cualquier institución financiera, comprender los motivos detrás de la deserción de un servicio y ser capaces de anticiparse a ella se ha convertido en una tarea imperativa para los bancos. Por lo que usando una base de datos bancaria se aplicarán los aprendizajes obtenidos a lo largo del curso para dar una hipótesis, analizar y posteriormente, argumentar una solución a diversos planteamientos sobre el uso de tarjetas de crédito.

Objetivo general

En este proyecto nos proponemos analizar las principales características de los clientes que abandonan el servicio de tarjeta de crédito de un banco. Así se podría reconocer anticipadamente aquellos clientes que en un futuro cercano podrían renunciar a su línea de crédito en el banco y así tomar medidas proactivas y dirigirse a estos clientes de manera oportuna, brindándoles opciones y servicios mejorados que los convenzan de mantener su relación con el banco.

Hipótesis

Hoy en día, existe una tendencia creciente hacia el uso de créditos bancarios debido a factores como el acceso más amplio a servicios financieros, la digitalización de la banca y la mayor conciencia sobre los beneficios y ventajas de utilizar tarjetas de crédito. Sin embargo, es probable que persistan diferencias significativas en el uso de créditos entre diferentes segmentos de la población, como aquellos con ingresos medios-bajos y altos. Se espera que los clientes de ingresos medios-bajos muestren una mayor disposición a utilizar créditos si se les ofrece un límite de crédito adecuado y se les brindan beneficios atractivos, mientras que los clientes de ingresos medios-altos puedan tener una preferencia hacia tarjetas de mayor nivel.

Introducción

Muchos de los bancos alrededor del mundo proveen el servicio de tarjetas de crédito. Una tarjeta de crédito funciona similarmente a un préstamo por un corto

periodo de tiempo, donde debes saldar tu deuda antes de una fecha especificada en tu periodo de servicio. Este servicio es sumamente popular, ya que las tarjetas de crédito se presentan como una manera conveniente de pagar los gastos de alimentación, renta, boletas, compras, o cualquier cosa que se requiera. Aparte de esto, también permite disfrutar de diferentes beneficios como descuentos exclusivos, ofertas y facilidades.

Se analizará el dataset **Credit Card Customers Prediction** que puede encontrarse en **Kaggle.com**, el cual consta de 10,000 clientes que mencionan su edad, salario, estado civil, límite de tarjeta de crédito, categoría de tarjeta de crédito, etc. Con el fin de reconocer anticipadamente aquellos clientes que en un futuro cercano renunciarían a su línea de crédito en el banco para tomar medidas proactivas y dirigirse a estos clientes de manera oportuna.

En un entorno empresarial altamente competitivo, la retención de clientes es fundamental para el éxito y el crecimiento de una empresa. Para comprender mejor los factores que influyen en la retención de clientes, es crucial analizar las variables demográficas de los usuarios y su impacto en la retención de clientes.

Resultados

Análisis Univariado

La primera variable, `attrition_flag`, revela que el 84 % de los clientes se mantienen activos, mientras que el 16 % han rescindido su contrato. Esta diferencia sugiere que la mayoría de los clientes muestran una lealtad sólida hacia la empresa, lo cual es un indicador positivo de retención.

En cuanto a la variable `gender`, se observa que el número de clientes masculinos y femeninos es casi equitativo,

³j._.abraham@ciencias.unam.mx

³alejandra.hinostroza@ciencias.unam.mx

³juansaidtriosgarcia@ciencias.unam.mx

lo que sugiere que el género no desempeña un papel significativo en la participación o cancelación de servicios. Esta igualdad de género puede ser un aspecto positivo, ya que indica una amplia aceptación y adopción de los servicios ofrecidos.

La educación de los clientes, medida por la variable `education_level`, también puede influir en la retención. El hecho de que el 30 % de los clientes tengan un título universitario sugiere que este grupo puede ser más consciente de los beneficios a largo plazo de mantener una relación comercial estable. Sin embargo, es importante tener en cuenta que una proporción significativa de clientes no tiene educación formal o su nivel educativo es desconocido, lo que implica la necesidad de adaptar las estrategias de retención para abordar diferentes perfiles educativos.

La variable `marital_status` revela que el 7 % de los clientes son divorciados, mientras que casi la mitad están casados. Esto sugiere que los clientes casados pueden tener una mayor necesidad o preferencia por los servicios ofrecidos, lo que se puede aprovechar para mejorar las estrategias de retención dirigidas específicamente a este grupo.

La distribución de la variable `income_category` también tiene implicaciones en la retención de clientes. Más de un tercio de los usuarios tienen el nivel de ingresos más bajo, lo que indica la necesidad de adaptar los precios y servicios ofrecidos para atraer y retener a este segmento. Además, solo el 7 % de los clientes tienen ingresos superiores a 120 mil dólares americanos, lo que puede indicar la existencia de oportunidades de venta adicionales para segmentos de ingresos más altos.

Por último, la variable `card_category` muestra que una pequeña cantidad de usuarios ha escalado su nivel de crédito, a pesar de tener la capacidad para hacerlo. Esto revela una brecha en la comunicación o percepción de valor por parte de los clientes, lo que destaca la importancia de educar y promover las ventajas y beneficios de las categorías superiores de tarjetas de crédito.

Para el siguiente análisis se comienza con la variable "dependent_count". Según los datos presentes, se puede inferir que el promedio de dependientes para una persona se sitúa entre 2 y 3 personas. Los extremos, es decir, aquellos sin dependientes o con el número máximo de dependientes (5), representan el menor porcentaje en comparación con el promedio.

Luego, se analiza la variable "total_relationship_count", que muestra la cantidad de productos contratados por

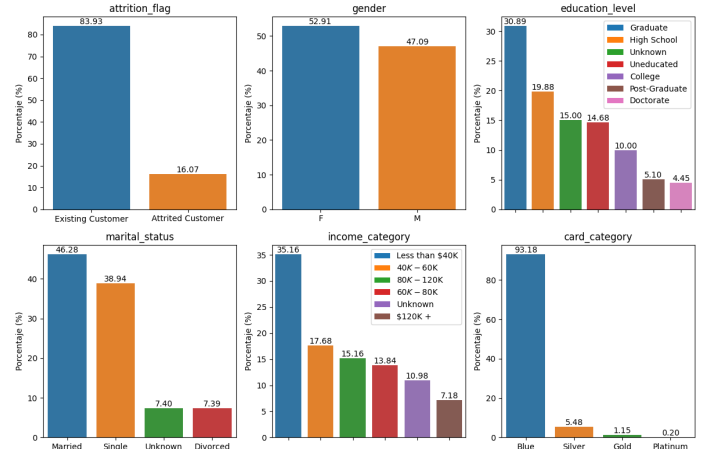


Figura 1: Análisis univariado 1.

los clientes. El promedio general se encuentra entre 3 y 6 productos. No hay una diferencia significativa entre los clientes que tienen 4 o 6 productos, mientras que aquellos con 3 productos representan un 4 % adicional en comparación con las categorías mencionadas anteriormente. Por otro lado, los clientes con 1 o 2 productos contratados tienen menos de la mitad de la proporción en comparación con las categorías anteriores.

Posteriormente, se mencionan las variables `months_inactive_12_mon` y `contacts_count_12_mon`. Se destaca una relación estrecha entre los porcentajes de clientes que presentan diferentes períodos de inactividad y el número de contactos realizados durante esos meses. Aunque puede haber algunas variaciones, la tendencia general muestra un crecimiento similar en ambos casos.

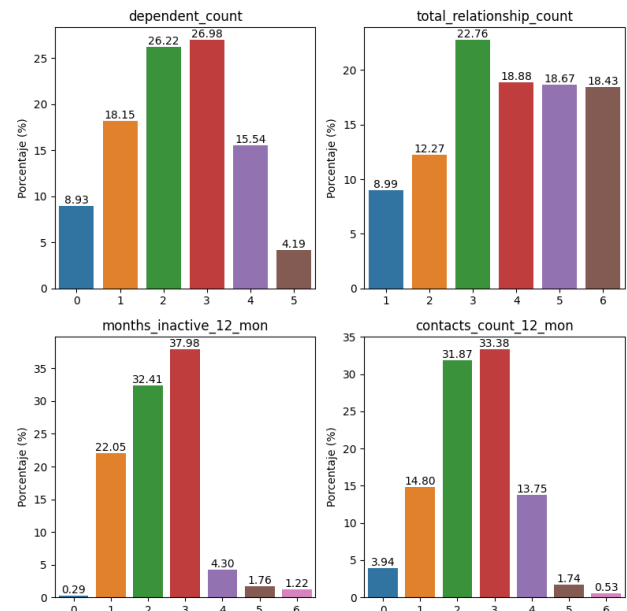


Figura 2: Análisis univariado 2.

Por otro lado, usando gráficos violinPlot observaremos el comportamiento de diferentes categorías, representadas en las Figuras 3 y 4.

El análisis de los datos revela algunas características importantes en cuanto a las distribuciones y los valores atípicos presentes en las variables. A excepción de `customer_age` "months_on_book", se observa que las distribuciones son multimodales, lo que significa que hay múltiples modas o picos en los datos. Además, se identifican muchos valores atípicos, es decir, puntos que se encuentran muy alejados de la mayoría de los datos.

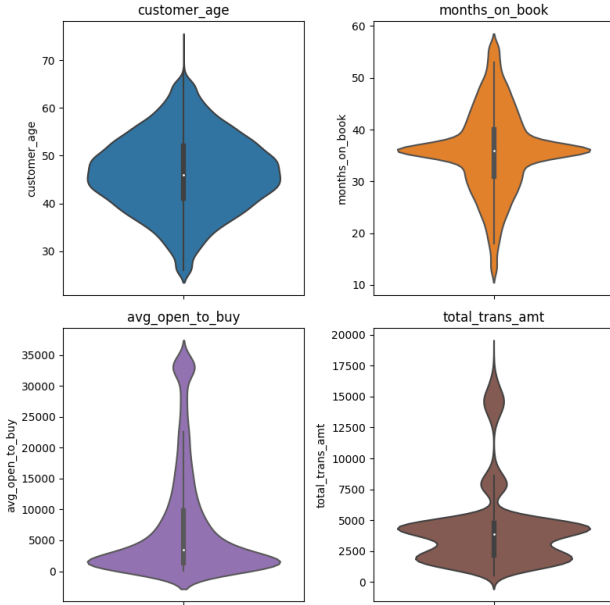


Figura 3: Análisis univariado 3.

Dado que existen valores atípicos y distribuciones multimodales, es recomendable utilizar la mediana en lugar de la media para hacer interpretaciones y comparaciones entre las categorías de datos, como `card_category` o `income_category`. La mediana es una medida de tendencia central robusta ante valores atípicos, ya que no se ve afectada por ellos de la misma manera que la media.

Análisis vibariado

En el gráfico izquierdo [ver Fig. 5] se muestra el porcentaje de clientes que eligieron diferentes tipos de tarjetas (Blue, Silver, Gold, Platinum). La tarjeta Blue es la más común, seguida de Silver, Gold y Platinum. Sin embargo, en el gráfico derecho [ver Fig. 5] se revela que la elección del tipo de tarjeta no está vinculada a los ingresos del cliente. Aunque la mayoría de los que eligieron la tarjeta Blue tienen ingresos inferiores a 40K, también hay clientes con ingresos más altos. Además, se observa que personas de bajos ingresos eligieron la tarjeta Platinum. En resumen, el nivel de ingresos no

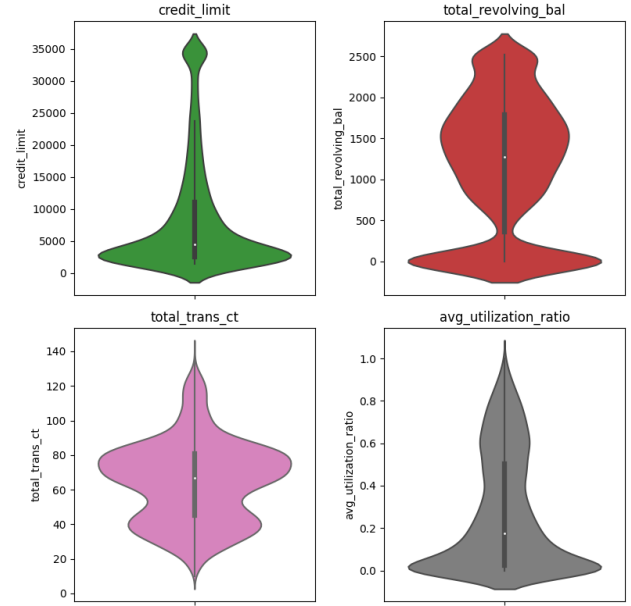


Figura 4: Análisis univariado 4.

determina la elección de tarjeta, lo que implica que no se pueden hacer suposiciones intuitivas al reemplazar datos desconocidos.

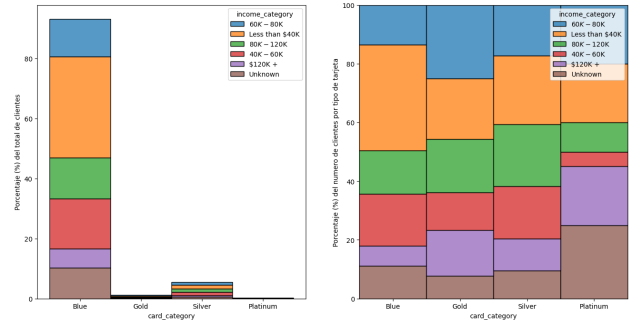


Figura 5: Análisis vibariado 1.

A partir del gráfico anterior [Fig. 5], creamos una tabla [Fig. 6] donde se hace evidente que el valor medio del número de meses de inactividad y la antigüedad del cliente en el banco es el mismo para las 4 tarjetas. Sin embargo, el número medio de productos adquiridos varía según el tipo de tarjeta, excepto para los clientes Platinum. Además, el número de veces que los clientes se han contactado con el banco es similar entre los tipos de tarjeta.

Se observa que el banco muestra un mayor nivel de contacto con los clientes Platinum, lo cual es comprensible, ya que es su servicio más exclusivo y no pueden permitirse perder a estos clientes. Sin embargo, el banco también debería prestar atención a los clientes Blue, quienes adquieren más productos en comparación con los Silver.

	total_relationship_count	months_on_book	months_inactive_12_mon	contacts_count_12_mon	months_on_book
Blue	4.0	36.0	2.0	2.0	36.0
Gold	2.5	36.0	2.0	2.0	36.0
Platinum	2.0	36.0	2.0	3.0	36.0
Silver	3.0	36.0	2.0	2.0	36.0

Figura 6: Análisis vibariado 2.

En el gráfico KDE [Fig. 7] se muestra la distribución del saldo rotativo total en diferentes tipos de tarjetas. La mayoría de los clientes de la tarjeta Blue no tienen saldo rotativo, mientras que en los otros tipos de tarjetas el máximo se encuentra alrededor de 1500, lo cual contradice nuestra hipótesis.

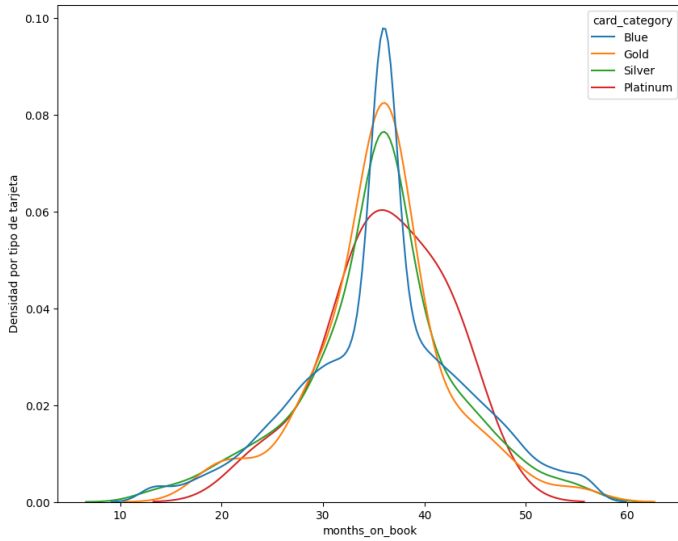


Figura 7: Análisis vibariado 3.

Esto sugiere que los clientes de tarjetas de mayor categoría gastan más de lo que su crédito les permite y dejan saldos impagados. Para retener a estos clientes, el banco podría considerar implementar facilidades de pago. Además, esto podría motivar a aquellos con tarjetas menos exclusivas y saldos rotativos altos a optar por servicios más costosos.

En el gráfico [ver Fig. 8], se observa que la mayoría de los clientes retirados tenían un saldo rotativo cercano a 0, mientras que los clientes activos presentan una distribución más amplia y multimodal, con valores más altos de saldo rotativo. Esto refuta la hipótesis de que los clientes dejaron de usar la tarjeta debido a los intereses generados por saldos impagados. Es probable que existan otros motivos por los cuales los clientes dejaron de utilizar la tarjeta.

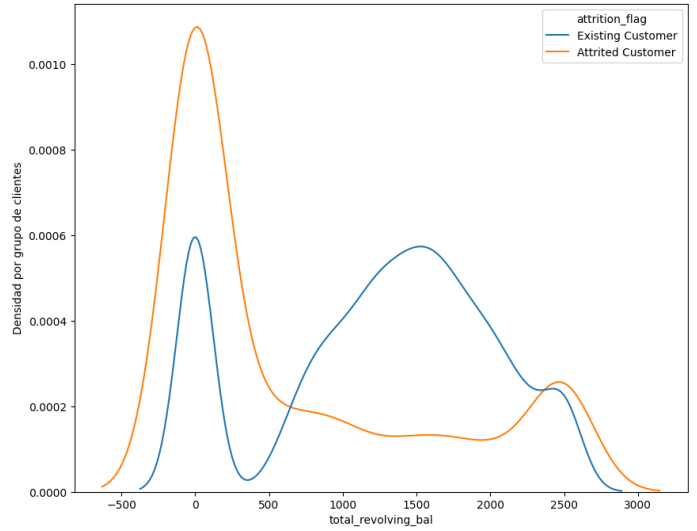


Figura 8: Análisis vibariado 4.

Análisis Multivariado

El gráfico de barras [Fig. 9] muestra el uso promedio de la tarjeta según el tipo y nivel de ingresos de los clientes. Contrario a nuestra hipótesis, los usuarios de la tarjeta Blue son los que más la utilizan, independientemente de su nivel de ingresos. Además, se observa que los clientes de menor ingreso tienden a utilizar más la tarjeta, sin importar su tipo. Esto sugiere la posibilidad de implementar estrategias dirigidas a clientes de bajos ingresos para mantener su alto nivel de uso, como promociones adecuadas a su estilo de vida y límite de crédito. Asimismo, es importante fomentar el mayor uso de las tarjetas por parte de los clientes de mayores ingresos, aprovechando su límite de crédito más alto.

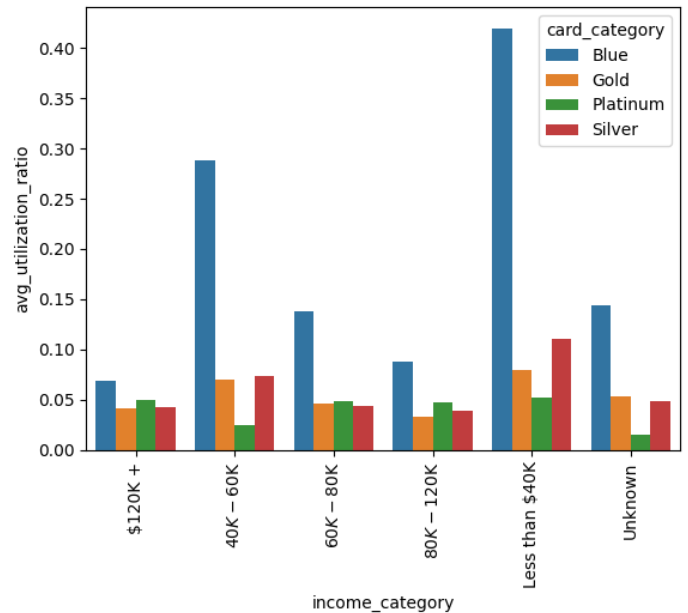


Figura 9: Análisis multivariado 1.

Haciendo un *joinplot* [ver Fig. 10] comparando el número total de transacciones versus el monto total de las transacciones para los clientes existentes y retirados del servicio, el gráfico muestra el número total de transacciones y el monto total de los clientes que cancelaron su tarjeta en comparación con los clientes existentes. Contrario a nuestra hipótesis, no se observa una diferencia significativa en el número de transacciones entre los clientes cancelados y los existentes. Además, ninguno de los clientes cancelados muestra un patrón de un uso intenso seguido de una cancelación repentina.

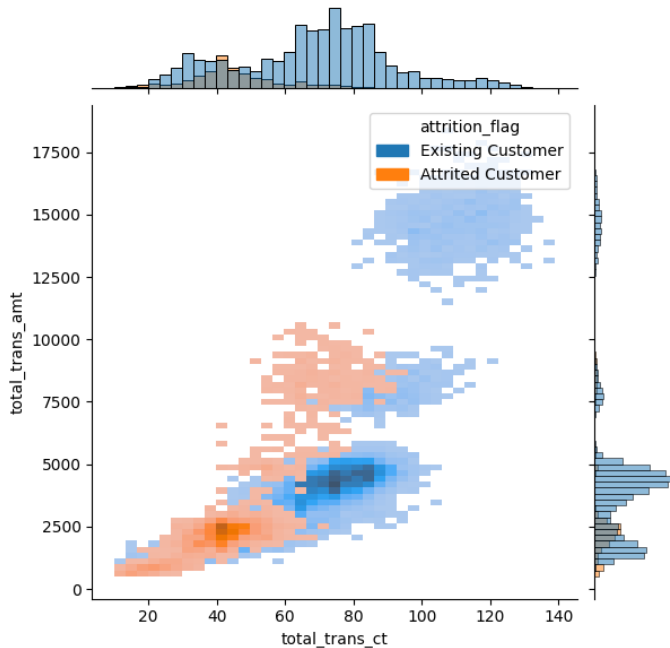


Figura 10: Análisis multivariado 2.

Análisis y Discusión de Resultados

El análisis de las variables demográficas proporciona información valiosa para comprender mejor la retención de clientes. La mayoría de los clientes se mantienen activos, lo que es alentador para la empresa. Aunque el género parece no tener un impacto significativo, los niveles educativos, el estado civil y la categoría de ingresos pueden influir en las estrategias de retención. Es importante adaptar las estrategias de retención para satisfacer las necesidades y preferencias específicas de cada grupo demográfico, brindando un valor diferenciado y fomentando una relación sólida y duradera con los clientes.

El análisis de estas variables [ver Figura 2] proporciona una visión general sobre los clientes y su comportamiento. Los resultados indican que la mayoría de los clientes tienen un número moderado de dependientes

y contratan entre 3 y 6 productos. Además, se observa una relación estrecha entre la inactividad y el número de contactos realizados, sugiriendo que los clientes menos inactivos tienden a tener más interacciones con la empresa. Estos hallazgos son valiosos para comprender mejor el perfil de los clientes y pueden ayudar a mejorar las estrategias de atención al cliente y personalización de servicios en función de sus necesidades y comportamientos.

Por ejemplo, se detectó que los clientes que rescindieron de su servicio de tarjeta de crédito presentaron una inactividad de 3 meses, lo que podría utilizarse como una alerta para el banco (así contacta al cliente y ofrece facilidades al alcanzarse este tiempo de inactividad). Asimismo, la mayoría de desertores de la tarjeta Blue resultaron ser las personas a cargo de una familia (según nuestra interpretación y análisis) por lo que sería ideal planear estrategias para adecuar esta tarjeta a este público objetivo cuando hay peligro de deserción. Algo similar sucede con los clientes Platinum, el que este tipo de tarjeta presente el mayor porcentaje de deserción, es un fuerte indicador que deben analizar los beneficios y servicios de esta tarjeta y así adecuarlo a estos clientes.

Conclusiones

A partir del análisis de datos realizado anteriormente, podemos dar las siguientes soluciones a la problemática principal:

Aumentar el límite de crédito para clientes de ingresos medios-bajos puede fomentar su uso de tarjetas de crédito.

Se deben ofrecer más beneficios en tarjetas de crédito de niveles superiores para atraer a clientes con ingresos medios-altos.

Las tarjetas de crédito azules, aunque comunes, tienen una menor cantidad de transacciones promedio, por lo que se recomienda ofrecer recompensas o promociones específicas para incentivar su uso.

Los clientes con ingresos por debajo de \$40,000 constituyen un grupo grande y propietarios de más productos, por lo que se sugiere mantener un contacto frecuente con ellos mediante promociones y beneficios relacionados con las tarjetas de crédito.

Estas estrategias pueden mejorar el uso de tarjetas de crédito y fortalecer la relación con los clientes.

Referencias

- [1] Kaggle. Recuperado el 29 de Mayo de 2023, de <https://www.kaggle.com/datasets>
- [2] Credit Card Customer Churn Predictive Analytics. Recuperado el 3 de junio de 2023, de <https://blog.devgenius.io/credit-card-customer-churn-predictive-analytics-b012ff8c385d>
- [3] What Is Credit Card Churning?. Recuperado el 3 de junio de 2023, de <https://www.forbes.com/advisor/credit-cards/what-is-credit-card-churning/>