

Tipo de documento: Tesis de maestría

Master in Management + Analytics

Predicción de Churn Voluntario en plataforma de Comercio online

Autoría: Sánchez, Sofía Danila

Fecha de defensa de la tesis: 2023

¿Cómo citar este trabajo?

Sánchez, S. (2023) "Predicción de Churn Voluntario en plataforma de Comercio online". [*Tesis de maestría. Universidad Torcuato Di Tella*]. Repositorio Digital Universidad Torcuato Di Tella

<https://repositorio.utdt.edu/handle/20.500.13098/12041>

El presente documento se encuentra alojado en el Repositorio Digital de la Universidad Torcuato Di Tella bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 2.5 Argentina (CC BY-NC-SA 2.5 AR)
Dirección: <https://repositorio.utdt.edu>



**UNIVERSIDAD
TORCUATO DI TELLA**

MASTER IN MANAGEMENT + ANALYTICS

**PREDICCIÓN DE CHURN VOLUNTARIO EN
PLATAFORMA DE COMERCIO ONLINE**

TESIS

Sánchez Sofía Danila

Mayo 2023

Tutor: Brenda Calabró

Resumen

Este trabajo aborda la problemática de retención de clientes en el contexto de una empresa que ofrece servicios digitales a otras empresas comerciales.

Dado que el servicio funciona en un esquema de comisiones y no de suscripción, se plantea un enfoque de prevención de *churn* transaccional, en el cual el foco es que el cliente continúe operando con la empresa.

Para ello, se identifican y analizan las principales variables relacionadas a los clientes que abandonan la plataforma. A partir de este análisis, se realiza el entrenamiento de tres tipos de modelos de aprendizaje automático (Regresión Logística, *Gradient Boosting* y *Random Forest*) y se compara su performance a través de la curva ROC.

Finalmente, se optimizan los hiperparámetros del modelo seleccionado a través de Grid Search, y se utilizan las predicciones para generar campañas de retención enfocadas en los grupos de clientes que poseen mayor riesgo de abandono.

Abstract

This research tries to address the problem of customer retention in the context of a company that offers digital services to other commercial companies.

Since the platform works on a commission scheme and not a subscription, a transactional churn prevention approach is proposed, where the focus is on the client continuing to operate with the company.

The main variables that are related to customers who leave the platform are identified. From this analysis, three types of machine learning models are trained (Logistic Regression, Gradient Boosting, and Random Forest) and their performance is compared through the ROC curve.

Finally, the hyperparameters of the selected model are optimized through Grid Search, and the predictions are used to generate retention campaigns focused on the customer groups that have the highest risk of churn.

Índice

1. Introducción	4
1.1. Contexto	4
1.1.1. Solución de pagos	5
1.1.2. Churn transaccional	6
1.2. Problema	7
1.3. Objetivo	8
2. Métodos y procedimientos	11
2.1. Data	11
2.1.1. Fuente de datos	11
2.1.2. Análisis univariado	14
2.1.3. Análisis multivariado	18
2.2. Técnicas de clasificación	24
2.2.1. Algoritmos de aprendizaje supervisado	24
2.2.2. Criterio de evaluación de modelos	30
2.2.3. Métricas de evaluación de modelos	32
3. Resultados	34
3.1. Performance de los modelos	34
3.2. Modelo final	40
3.3. Estabilidad del modelo	41
4. Aplicaciones del modelo predictivo	45
5. Conclusiones	49
6. Bibliografía	51
7. Anexos	53
7.1. Sistema de pagos en Brasil	53
7.2. Ingeniería de atributos	54
7.3. Análisis exploratorio - gráficos adicionales	56

1. Introducción

1.1. Contexto

Este trabajo analiza como caso de negocio la retención de clientes de una empresa que ofrece software como un servicio (SaaS) a otras empresas. Es una plataforma de comercio electrónico que ofrece el servicio de creación de una tienda online autogestionable para pequeñas y grandes empresas en Brasil. En otras palabras, provee una solución tecnológica que permite crear un sitio de e-commerce, personalizar el diseño a medida y ofrecer en ese entorno productos y/o servicios a la venta.

El servicio que ofrece está diseñado para que no requiera conocimientos técnicos: tanto el proceso de creación como la gestión del día a día son intuitivos para que emprendedores con poca experiencia en el comercio online puedan configurar y gestionar su tienda. En este sentido, logra una diversa cartera de comerciantes, distribuidos en un amplio rango de facturación.

Para cubrir todas las necesidades que rodean a un comercio digital, la empresa se encuentra integrada con un ecosistema de soluciones tecnológicas. Este ecosistema consiste de una red de proveedores, que brindan diversos productos y servicios que complementan la tienda online. Entre los servicios complementarios podemos encontrar soluciones de medios de pago, medios de envío¹, plataformas de marketing² y, aplicaciones de gestión, entre muchas otras. El presente trabajo se focaliza en los medios de pago.

Durante los últimos dos años, la empresa desarrolló un servicio de gestión de medios de pago de forma interna para potenciar el crecimiento de los comercios con una solución hecha a medida de sus necesidades. Éstos se integran a la plataforma de comercio electrónico y gestionan las transacciones financieras online para que los dueños de las tiendas puedan realizar cobros en sus negocios mediante tarjetas, transferencias, etc.

En la subsección a continuación se describe en mayor profundidad la solución de pagos generada de forma interna por la empresa. Para mayor información sobre el contexto del sistema de pagos electrónicos general en el cual se enmarca, referirse al [Anexo 7.1](#). Asimismo, en segundo lugar, se aborda la definición de "*churn* transaccional" en el contexto de la empresa para distinguirlo de otras instancias de *churn* que no son el foco de este trabajo. En la siguiente sección se cubren los aportes de la literatura realizados al problema del *churn* voluntario para finalmente, definir el objetivo del presente trabajo.

¹ Empresas de logística que se encargan del proceso de envío para que sea realizado de la forma más automatizada posible para que los productos lleguen a los compradores.

² Ofrecen servicios como la creación de anuncios, disparo de SMS, envíos de e-mail marketing y creación de cupones.

1.1.1. Solución de pagos

Dentro del ecosistema de pagos brasileño, la solución de pagos interna ha sido desarrollada especialmente para su uso en las tiendas creadas por esta plataforma de comercio electrónico. Cumple el rol de una pasarela de pagos, concentrando el intenso flujo de intercambio de información entre varias partes necesarias para el procesamiento de un pago. La integración con la tienda es fácil y sencilla, ya que todas las acciones relacionadas con los pagos de ventas se encuentran disponibles directamente en el panel de administración de la tienda virtual. Esto significa que los comerciantes pueden gestionar su comercio electrónico y sus medios de pago en un solo lugar, lo que agiliza y simplifica el proceso.

Asimismo, al ser una solución interna diseñada para una cartera diversa de comerciantes, se caracteriza por su flexibilidad, permitiendo adaptarse a las necesidades y preferencias de los comerciantes en términos de disponibilidad de fondos. La plataforma ofrece diferentes formas de pago en la tienda, como el boleto, el pix y la tarjeta de crédito³. Los cobros se acreditan de manera inmediata, excepto en el caso de las tarjetas de crédito, cuya acreditación puede variar entre 2, 14 y 30 días, dependiendo de la elección del comerciante en función al costo que esté dispuesto a pagar.

Antes del desarrollo de esta solución, los ingresos de la empresa estaban vinculados únicamente a la suscripción que los comerciantes abonaban para tener su tienda virtual y a una comisión variable basada en el monto de ventas mensuales (facturación). No se percibía ningún tipo de ingreso o comisión por permitir la integración de servicios de terceros en la plataforma. Sin embargo, al ofrecer una plataforma interna de pagos, la empresa ahora puede generar ingresos adicionales por el procesamiento de pagos, lo que contribuye a expandir su rentabilidad y diversificar sus fuentes de ingresos.

En este sentido, la retención de comerciantes y la captación de nuevos clientes son aspectos clave en la estrategia de la empresa para mantener su crecimiento y rentabilidad a largo plazo. Si bien esta solución es elegida por gran parte de los comerciantes, la empresa sigue dedicando esfuerzos a la adquisición de nuevas tiendas con un perfil de mayor facturación. Es importante tener en cuenta que cuando una de estas tiendas deja de utilizar ese servicio en la plataforma, tiene un impacto significativo en las métricas de negocio, como el volumen de transacciones, la facturación y la rentabilidad.

³ El boleto es una opción para pagos en efectivo, pix es un sistema de pagos instantáneos y la tarjeta de crédito brinda la posibilidad de pagar a crédito. En particular, el que suele ser más relevante tanto por su rentabilidad así como la mayor complejidad en el cobro y análisis de la transacción es la tarjeta de crédito. Para más detalles, referirse al Anexo 7.1.

En resumen, esta solución de pagos interna proporciona a los comerciantes una forma conveniente y centralizada de gestionar los pagos en sus tiendas virtuales. A través de diversas opciones de pago, un proceso integrado y la generación de ingresos adicionales para la empresa, esta solución contribuye a la expansión y rentabilidad de la plataforma de comercio electrónico. Es por ello que, para garantizar el crecimiento sostenido del negocio, es importante abordar los desafíos asociados con la retención de clientes y la adquisición de nuevos comerciantes.

1.1.2. *Churn* transaccional

El *churn* se puede diferenciar en dos grupos principales: voluntario e involuntario. Mientras que el primero ocurre cuando el comerciante decide dejar de utilizar un servicio por sus propios medios, el segundo depende únicamente de una decisión de la empresa de concluir la relación (pueden existir distintos motivos como falta de pago, fraude, etc.). En este trabajo, se concentrará en prevenir el *churn* del primer grupo, el voluntario.

Debido a la naturaleza del servicio, donde las transacciones son la única instancia de uso y pago, no se refleja directamente la decisión de dejar de utilizarlo a través de una cancelación de suscripción o una baja del servicio. Es posible que un comerciante tenga la integración activa e instalada, pero esté utilizando otro subadquirente para procesar los pagos. Por lo tanto, en lugar de definir el *churn* como un fenómeno permanente, se ha establecido en base a un período de inactividad, donde no se han realizado procesamientos de pagos, pero se han registrado transacciones en la tienda.

Un comerciante se considera activo en la plataforma de pagos si ha procesado al menos un pago a través de esa integración en el último mes. Por otro lado, se considera que un comerciante previamente activo realiza *churn* cuando, a pesar de tener ventas a través de otros subadquirentes, deja de tener actividad en la solución interna de pagos durante los últimos 30 días.

La métrica de *churn* transaccional se define como la cantidad de comerciantes que han realizado *churn* en el último mes dividido por el número de comerciantes activos al comienzo del período. Dado que la variable está construida de esta manera, se requiere esperar un máximo de 30 días después de observar una transacción a través de la plataforma de pagos interna para determinar si ese comerciante continuará utilizando los servicios o no.

1.2. Problema

Actualmente, el equipo de negocio dedica grandes esfuerzos a la adquisición de clientes con un nivel de facturación alto o medio, pero no realiza acciones de retención proactivas para evitar el abandono. En este aspecto, la mayoría de las acciones están orientadas a recuperar a un *churn*, y poseen una baja tasa de efectividad. Si bien la tasa de *churn* estuvo disminuyendo en los últimos meses, por mejoras propias de la plataforma, no deja de ser un indicador relevante monitoreado por el equipo.

La literatura proporciona amplia evidencia sobre el impacto positivo de las acciones de retención en comparación con los costos asociados a las campañas de adquisición (Rust & Zahorik, 1993; Athanassopoulos, 2000). Por ello, las empresas destinan múltiples esfuerzos en detectar clientes con mayor riesgo de *churn* para que las campañas se enfoquen donde puedan generar un mayor diferencial.

Sin embargo, la mayoría de los estudios existentes (Burez & Van den Poel, 2007; Coussement & Van den Poel, 2008) se han centrado en contextos business-to-consumer (B2C), donde las muestras y tasas de *churn* son más amplias y el costo de perder un cliente es más limitado en comparación con el ámbito business-to-business (B2B). A su vez, con la profundización del desarrollo de la industria de comercio electrónico, se observa que las barreras para cambiar de servicios/productos se han debilitado, por lo que los clientes poseen más facilidad de cambiar de proveedores versus otras industrias y hace que la problemática sea aún más relevante.

Existen trabajos previos que han focalizado en la predicción de *churn* en un contexto B2B, pero con foco únicamente en industrias de comercio mayorista y retail. Es el caso de Gatterman et al. (2022), que desarrollan un modelo de predicción de *churn* no-contractual y, comprueban a través de un experimento, que el contacto con clientes con alta probabilidad de *churn*, reduce el *churn* en la población de forma significativa. Asimismo, contribuyen a la literatura a través de la inclusión en el modelo de variables como la relación con la empresa (qué tan reciente fue su último contacto) además de las variables evaluadas comúnmente dentro del esquema *Recency, Frequency* y *Monetary value* (RFM).

Otras líneas de trabajo, avanzan un paso más y no sólo intentan predecir el *churn* sino que también entender cómo esto afecta en la rentabilidad del negocio. En Tamaddoni et al. (2014) se observa que además de modelar la probabilidad de *churn*, proponen una campaña destinada a maximizar la rentabilidad de la misma con la utilización de las predicciones del modelo. En otro trabajo, De Caigny et al. (2021) hacen una revisión de literatura sobre *churn* no contractual en B2B y testean modelos para predecir no solo la probabilidad de *churn* de cada cliente, sino la probabilidad de *churn* neta que resulta de

aplicar una campaña de retención. Si bien no se podría aplicar en este trabajo es una rama de investigación que se podría continuar para futuras iteraciones.

En cuanto al tipo de modelos de aprendizaje automático utilizados para este tipo de problemáticas, existe una amplia variedad de posibilidades. Como en Tamaddoni et al. (2015) que realizan un análisis comparativo de distintas técnicas de modelado de probabilidad de *churn*, encontrando que en contextos donde la tasa de *churn* es más baja, se observa una mejor performance de la regresión logística.

Si bien estos trabajos aportan conocimientos valiosos, ninguno hasta el momento ha abordado la problemática de retención de clientes en la industria tecnológica en un entorno B2B como es nuestro caso actual referente a la retención de clientes en una plataforma de pagos en Brasil. A través de esta investigación, se busca generar conocimientos y recomendaciones prácticas para mejorar la retención de clientes y la rentabilidad en este contexto específico de creciente relevancia.

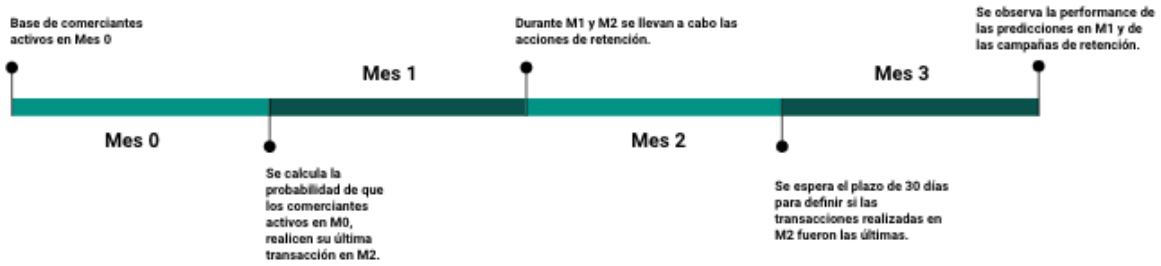
1.3. Objetivo

Con la intención de brindar una solución proactiva al *churn* voluntario, este trabajo se propone desarrollar un modelo predictivo de *churn* para ser aplicado a campañas de retención a los clientes con mayor riesgo de *churn*. A través de sus predicciones, la empresa podrá destinar esfuerzos de retención de forma focalizada, generando campañas de retención sustentables sin perjudicar la rentabilidad del producto en general.

El objetivo principal de esta investigación será predecir con una antelación de 2 meses si un comerciante dejará de transaccionar a través de la plataforma. Luego, en función a estas predicciones se evaluará el impacto de las campañas de retención realizadas.

En el diagrama a continuación se puede observar la línea temporal de las predicciones. A partir de la base de comerciantes activos identificados en el mes inicial (M0), a comienzos de M1 se generan las probabilidades de que cada uno de ellos realice su última transacción en los próximos 2 meses (en M2). No se intenta predecir si la última transacción es en M1 ya que dejaría poco tiempo para llevar a cabo las estrategias de retención del equipo. Recién a finales de M3 se observa si efectivamente se cumplieron o no las probabilidades predichas en M1, dejando transcurrir el plazo de 30 días de inactividad.

Figura 1.1: Ventana temporal de las predicciones



De esta forma, se optimiza el modelo para orientar las acciones a la población con más probabilidad de que se llegue a retener a tiempo. Sin embargo, esta decisión afecta la agilidad con la que se puede identificar una pérdida de performance del modelo. Para acotar este riesgo, se implementará el Índice de Estabilidad de la Población⁴ para anticiparnos a posibles cambios en la distribución de la probabilidad de *churn*.

Uno de los principales desafíos de este abordaje, es que la tasa de *churn* medida en función a nuestra variable objetivo (última transacción a 2 meses) es baja, manteniéndose en un rango de entre 2% y 6% en el período bajo análisis y, por ende, reflejando un alto desbalance en el dataset a modelar.

Por otro lado, la cantidad de observaciones es acotada, ya que la solución de pagos tiene menos de un año de historia y en ese período ha crecido constantemente su adopción y ha tenido importantes cambios en su desarrollo. Por lo que es sumamente importante monitorear que la performance del modelo se mantenga en los meses más recientes para no afectar la estabilidad de las predicciones.

Con estas problemáticas presentes, se decidió realizar el análisis comparativo de los modelos a través del área bajo la curva (AUC)⁵ y el porcentaje de churners identificados correctamente. Complementando este análisis con métricas sobre la estabilidad del modelo en el período posterior del análisis.

El presente trabajo de investigación se estructura de la siguiente manera:

- En la primera sección hemos cubierto el contexto de la empresa, el producto bajo análisis, la literatura previa relevante así como también, planteado el objetivo principal de esta investigación.

⁴ El Índice de Estabilidad de la Población, PSI por sus siglas en inglés, es una métrica para determinar si ha cambiado la distribución de una variable en dos momentos distintos.

⁵ Métrica para evaluar performance de modelos de aprendizaje supervisado. Equivale a la probabilidad de que un clasificador clasifique una instancia positiva elegida al azar más alta que una instancia negativa elegida al azar

- La segunda sección se compone de dos partes principales, la primera realiza una descripción de la construcción de las variables para el modelo y un análisis exploratorio de las mismas. En la segunda, se hace un análisis en profundidad de los distintos algoritmos de aprendizaje relevantes para este caso de negocio y los distintos criterios de evaluación a ser utilizados.
- En la tercera sección del trabajo, se encuentran los resultados comparativos de los modelos aplicados, se explica cuál fue el que tuvo mejor performance y se analiza la estabilidad del modelo en el período fuera de la muestra.
- Finalmente, en la cuarta sección, a modo de conclusión, se hace un breve resumen de las iniciativas de retención que se llevaron a cabo en función a las predicciones del modelo, la técnica a través de la cual se estimó el impacto y los resultados finales en términos de la métrica de *churn*.

2. Métodos y procedimientos

2.1. Data

2.1.1. Fuente de datos

Los datos proporcionados por la empresa para llevar a cabo el desarrollo de este trabajo se concentran principalmente en cuatro fuentes: ventas totales de cada tienda, transacciones procesadas por la solución interna, características asociadas a la tienda y el vínculo establecido con la empresa (contactos, tipo de relación). Debajo se presenta un breve explicación del contenido de cada tabla y se mencionan las columnas relevantes para el proyecto.

El primer conjunto de datos concentra todas las órdenes de compra de la tienda (sin importar la integración de pagos utilizada), con su monto asociado en reales y datos sobre el método de pago y el proveedor a través de la cual se realizó. Son aproximadamente 9 millones de registros de los cuales no se observa ninguno con datos nulos. Se encuentra a nivel de la orden y los principales campos son:

1. *order_id*: clave primaria que identifica cada orden dentro de la empresa.
2. *store_id*: clave foránea que identifica la tienda en que se produjo la orden.
3. *completed_at: timestamp* que contiene fecha y hora en que se completó el pago de la orden.
4. *total*: monto en moneda local de la orden (en este caso, reales).
5. *payment_method*: método de pago (boleto, pix, tarjeta de crédito, otros).
6. *payment_gateway*: proveedor que procesó el pago.

El segundo conjunto de datos es un complemento del primero, ya que contiene para cada orden que fue procesada por la solución de pagos interna, cada intento de pago de la orden y su estado correspondiente. No existen registros nulos en esta tabla, la cual contiene más de 5 millones de registros. Se encuentra a nivel de transacción (una orden puede tener más de una transacción si tuvo intentos de pago fallidos) y los principales campos son:

1. *transaction_id*: clave primaria que identifica cada transacción de la solución de pagos.
2. *order_id*: clave foránea que identifica la orden asociada a la transacción.
3. *estado*: estado de la transacción.
4. *created_at: timestamp* que contiene fecha y hora en que se creó la transacción.
5. *updated_at: timestamp* que contiene fecha y hora en que se actualizó la transacción.

El estado del intento de pago inicialmente puede ser pagado o rechazado. Este último, ocurre cuando no puede concluirse la compra por falta de fondos o riesgo de fraude. Más allá de estos dos casos, el estado de las transacciones que prosperaron pueden sufrir

modificaciones en el tiempo cambiando a *chargeback* o *refunded* (en esos casos el campo de *updated_at* registra la fecha de la modificación). Mientras que *chargeback* son transacciones que luego manifestaron ser fraudulentas, por lo que implica una pérdida y una mala experiencia tanto para el comerciante como para la plataforma de pagos; el estado *refunded* implica una cancelación del pedido y devolución de monto por parte del comerciante. De todas formas, el mayor porcentaje de estos cambios se concentra en el primer mes luego de la compra, por lo que no afectaría los análisis realizados en este trabajo.

El tercer conjunto de datos, concentra todas las características propias de la tienda en sí, prácticamente invariables en el tiempo, como la fecha de creación, fecha de pago, tipo de plan y el segmento de los productos que vende. Posee aproximadamente 2 millones de tiendas/filas.

1. *store_id*: clave primaria que identifica a la tienda.
2. *created_at*: *timestamp* con la fecha y hora de la creación de la tienda.
3. *current_plan*: plan de comercio digital contratado por el comerciante.
4. *current_segment*: columna categórica con una clasificación en función a la cantidad de órdenes de la tienda en los últimos 90 días⁶.
5. *vertical*: categoría de los productos que vende la tienda. Se encuentra construida con información proporcionada por el comerciante y en caso de que esté incompleta se asigna el segmento a través de un modelo predictivo⁷.

El cuarto y último conjunto de datos hace referencia a la relación que tiene el comerciante con la empresa, en particular, todos los contactos que hayan existido con el equipo de soporte (600.000 registros aproximadamente), contiene:

1. *ticket_id*: clave primaria que identifica a cada ticket/contacto.
2. *created_at*: *timestamp* con la fecha y hora de la creación del ticket.
3. *store_id*: clave foránea que identifica la tienda que inició el ticket. Este campo, posee un 25% de registros nulos, los cuales la mayoría corresponden a casos de tickets que no fueron iniciados por comerciantes sino por socios comerciales o consumidores en las tiendas por lo que se excluyeron del análisis.
4. *topic*: tópico al que hace referencia el ticket. A través de este campo, se logran identificar los tickets que corresponden a consultas sobre la plataforma de pagos únicamente.

⁶ Esta variable es nula en un 10% de los casos, siempre que la tienda tenga menos de 30 días en la plataforma y, por ende, no hay información suficiente de ventas para clasificarla.

⁷ El porcentaje de nulos, aún considerando los casos taggeados por el modelo es de un 25%, que en esos casos se les asigna como categoría “*unknown*”.

5. *satisfaction_rating*: json con el rating numérico de la encuesta de satisfacción, motivo principal de la respuesta, y comentarios adicionales. La tasa de respuesta de las encuestas de satisfacción luego de un contacto es aproximadamente del 15%, por lo que en un 85% de los casos estos registros son nulos, y no son incluidos en el análisis.

En función a las fuentes mencionadas, se realizó un importante trabajo de ingeniería de atributos para construir un dataset que consolide información a nivel de la tienda y de forma mensual. En el [Anexo 7.2](#) se profundiza en el detalle de la construcción de las variables. Debajo se observa un resumen de las variables construidas:

Grupo de variable	VARIABLES	Nombre abreviado	Definición	Tipo de variable
Volumen de negocio	Gross Merchandise Value (GMV)	<i>gmv</i>	Volumen de ventas en términos monetarios del comerciante (Facturación)	Numérica
	Gross Merchandise Value (GMV) en Tarjeta de crédito	<i>gmv_cc</i>	Volumen de ventas en términos monetarios del comerciante (Facturación) por tarjeta de crédito	Numérica
	Tier	<i>tier</i>	Clasificación de los comerciantes por rango de GMV. Mayor GMV, menor Tier.	Categórica
	Transacciones	<i>trx</i>	Volumen de ventas en cantidades del comerciante	Numérica
	Transacciones en la plataforma de pagos	<i>trx_p</i>	Volumen de ventas en cantidades del comerciante en la plataforma de pagos	Numérica
	Share of Wallet (SoW)	<i>sow</i>	Volumen de ventas procesadas por plataforma (R\$)/Volumen de ventas totales (R\$)	Numérica
	Share of Wallet (SoW) en Tarjeta de crédito	<i>sow_cc</i>	Volumen de ventas procesadas por plataforma por tarjeta (R\$)/Volumen de ventas totales por tarjeta (R\$)	Numérica
	Ticket promedio	<i>avg_ticket</i>	Volumen de ventas (R\$)/ Cantidad de ventas totales	Numérica

Periodicidad	Periodicidad de ventas en días	<i>days_btw_trx</i>	Cantidad de días promedio que transcurren entre transacciones	Numérica
	Periodicidad de ventas en días en plataforma de pagos	<i>days_btw_trx_p</i>	Cantidad de días promedio que transcurren entre transacciones con la plataforma	Numérica
	Ratio de periodicidad de ventas	<i>days_btw_trx_ratio</i>	Cantidad de días promedio entre transacciones en la plataforma / Cantidad de días promedio entre transacciones en la tienda en general	Numérica
Características del cliente	Segmento	<i>vertical</i>	Tipo de producto que comercializa	Categórica
	Antigüedad en plataforma de pagos	<i>tenure_mont_h_p</i>	Cantidad de meses que transcurrieron desde su primera transacción a través de la plataforma de pagos	Numérica
	Antigüedad de su tienda virtual	<i>tenure_mont_h</i>	Cantidad de meses que transcurrieron desde la creación de su tienda virtual	Numérica
Relación con empresa	Contacto con atención al cliente	<i>contacts</i>	Contactos con atención al cliente	Categórica
	Contacto con atención al cliente con calificación negativa	<i>bad_csat</i>	Contactos con atención al cliente con calificación de satisfacción negativa	Categórica
Riesgo y Fraude	% de la facturación por tarjeta rechazado	<i>rejection_ratio</i>	Volumen de ventas que fue rechazada al procesar el pago (R\$)/Volumen de ventas totales (R\$)	Numérica
	% de la facturación con chargeback	<i>chargebacks_ratio</i>	Volumen de devoluciones (R\$)/Volumen de ventas totales (R\$)	Numérica

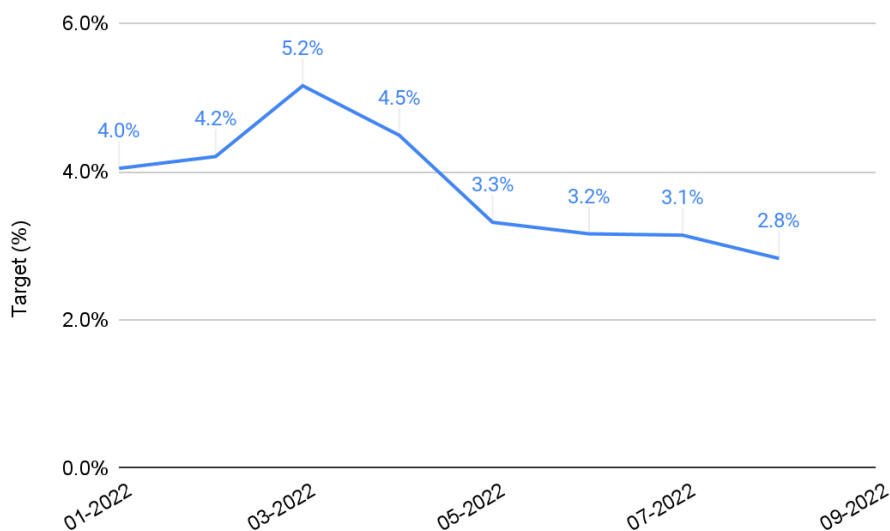
2.1.2. Análisis univariado

Se cuenta con un dataset de 21.000 observaciones (Enero 2022 a Julio 2022 inclusive), donde cada observación refiere a las características de un comerciante activo en la plataforma de pago en un mes en particular. La variable *target* dicotómica toma el valor 1 cuando el comerciante realizó su última transacción en los próximos 2 meses, 0 en caso contrario.

Debajo se puede observar la evolución de la variable objetivo por mes (cantidad de *churners* en los próximos 2 meses / cantidad de observaciones). El promedio durante el período bajo análisis fue del 3.2%, con una tendencia decreciente. Este indicador refleja una evolución acorde al desarrollo y madurez que fue ganando la solución de pagos desde que

surgió (comienzos de 2021), que tuvo como correlato un crecimiento de la base de comerciantes con mayor antigüedad y una mejor tasa de retención.

Gráfico 2.1: Churners (%) por mes de actividad



En principio, se realizó un análisis sobre la distribución de cada variable explicativa. Dentro del grupo de indicadores asociados al volumen de negocio, encontramos que aquellas referidas al monto y cantidad de transacciones tienen un sesgo hacia la derecha. Aproximadamente un 13% de las observaciones se considerarían *outliers*, tomando como su definición aquellos registros superiores al límite del cuartil 75% + rango intercuartil * 1.5.

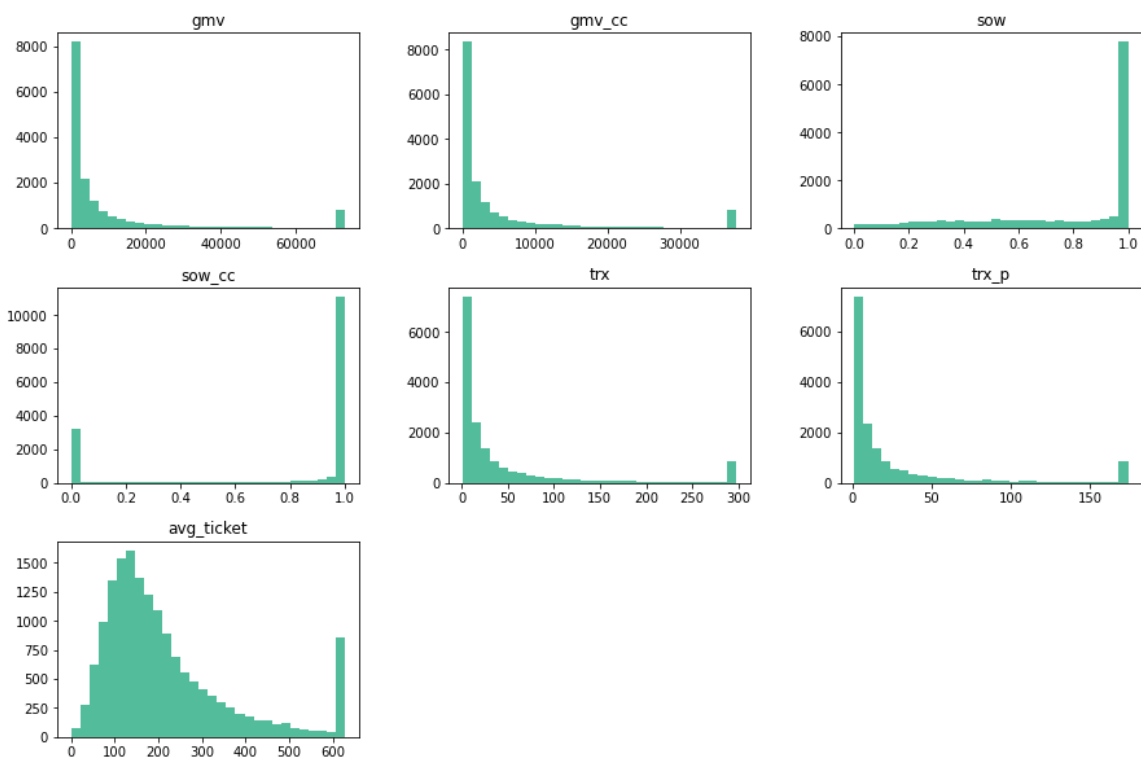
Tabla 2.2: Distribución de las variables del volumen de negocio

Variable	Mínimo valor	Q25	Q50	Q75	Máximo valor	% de outliers
GMV	0.2	615.8	2,344.0	8,821.2	2,093,512.1	13.3%
GMV en Tarjeta de Crédito	0.0	265.1	1,145.6	4,507.9	1,474,057.0	13.6%
SoW	0.0	0.5	1.0	1.0	1.0	0.0%
SoW en Tarjeta de crédito	0.0	0.7	1.0	1.0	1.0	0.0%
Transacciones	1.0	4.0	13.0	42.0	11,044.0	12.7%
Transacciones en la plataforma de pagos	1.0	3.0	8.0	26.0	7,484.0	12.8%
Ticket promedio	0.1	114.7	170.5	264.6	15,389.1	7.6%

Estos valores corresponden a comerciantes que tienen un nivel superior de transacciones que el resto de las tiendas y que si bien es importante que estén incluidos en el análisis, pueden llegar a afectar las conclusiones de este análisis al tener valores tan extremos. Como resulta importante en este trabajo mantener la forma de la distribución de las variables, y ante la limitación para aplicar logaritmos ya que pueden existir observaciones en 0, se ha decidido que para disminuir los valores extremos, se realice un *capping*⁸ de estas variables. Que algunas variables tengan valores cercanos a 0 tienen una lógica del negocio. Por ejemplo, en el caso del GMV en tarjeta de crédito puede tener valor 0, si el comerciante no tiene configurado ese método de pago en sus ventas (vende por boleto o pix). El *Share of Wallet (SoW)*, si bien no puede tomar el valor exacto de 0 (ya que implicaría que el comerciante no tuvo ninguna transacción en la plataforma de pagos interna) si puede tender a 0 en casos donde el comerciante utilice muy poco esa integración.

A los registros que pertenezcan al 5% de los valores más extremos, se les asigna el valor del percentil 95. Debajo se puede observar cómo queda la distribución de cada variable luego de aplicar el *capping*.

Gráfico 2.2: Volumen de negocio - Histogramas



⁸ Es uno de los métodos disponibles para el manejo de outliers sin excluir las observaciones. Se asigna un valor predeterminado a todas aquellas observaciones mayores a cierto umbral.

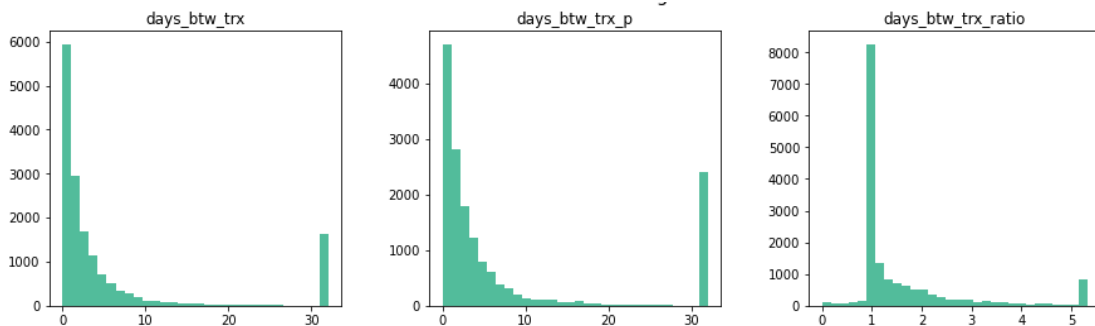
Sobre las variables de periodicidad de ventas, es importante resaltar que por su confección (cantidad de días que transcurren entre las ventas dentro de un mismo mes) cuando el comerciante tuvo solo una venta en el mes, por default se le asigna el valor 32. Por ende, en análisis de cuartiles si bien los considera como *outliers*, son casos donde no tuvieron más de una venta en el mes y por ende, válidos de estar incluidos en el dataset.

Tabla 2.3: Distribución de las variables de periodicidad

Variable	Mínimo valor	Q25	Q50	Q75	Máximo valor	% de outliers
Periodicidad de ventas	0.0	0.6	1.8	4.8	32.0	13.4%
Periodicidad de ventas en plataforma de pagos	0.0	0.9	2.4	7.0	32.0	16.7%
Ratio de periodicidad	0.0	1.0	1.0	1.7	1,323.7	11.7%

Diferente es el ratio entre estas variables, ya que ante valores muy pequeños la periodicidad de ventas de la tienda en general, el ratio tiende a infinito. Por ello, también para esta variable se aplicó un *capping* del 5%, manteniéndose entre valores de 0 y 5.3, como se puede observar en el histograma a continuación.

Gráfico 2.3: Periodicidad de ventas - Histogramas



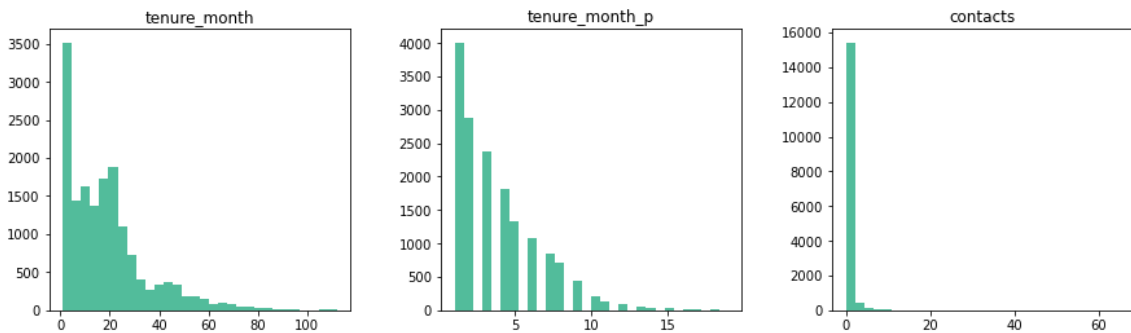
En cuanto a los indicadores de antigüedad, encontramos que como mínimo hay comerciantes con un mes de vida de su tienda virtual hasta 112 meses como máximo, mientras que la mediana es de 16 meses. La antigüedad en la plataforma de pagos es significativamente menor, con una mediana de 3 meses y como máximo de 19 meses. No consideramos que estos datos sean *outliers*.

La cantidad de contactos con el equipo de soporte, presenta una alta concentración en 0 (sin contactos sobre la plataforma de pago). En promedio, se contactó un 20% de la base en el mes.

Tabla 2.4: Distribución de las características y relación con la empresa

Variable	Mínimo valor	Q25	Q50	Q75	Máximo valor	% de outliers
Antigüedad de su tienda virtual	1	6	16	25	112	4.8%
Antigüedad en plataforma de pagos	1	2	3	5	19	4.0%
Contacto con atención al cliente	0.0	0.0	0.0	0.0	65.0	20.3%

Gráfico 2.4: Características y contactos - Histogramas



Por último, se observa que también existe una gran concentración de valores en torno al 0 para las variables de *chargeback* y rechazos⁹. Por lo que el análisis de cuartiles califica como *outliers* casos que simplemente tuvieron algún tipo de *chargeback* o fraude.

2.1.3. Análisis multivariado

Luego del análisis univariado y de realizar las transformaciones necesarias a las variables, se realizó un análisis sobre las interacciones entre las variables a través de una matriz de correlación. La misma refleja los valores de correlación¹⁰, que miden el grado de relación lineal entre cada par de variables. En función de ella podemos destacar aquellas correlaciones mayores a 0.3 (en términos absolutos):

- El *tier* posee una correlación negativa con el GMV y las transacciones, que se debe a cómo está construida la variable. Este indicador se genera en función a la facturación de cada comerciante en los últimos 3 meses. Tiers más pequeños concentran comerciantes con mayor GMV. Más allá de reflejar la facturación, es

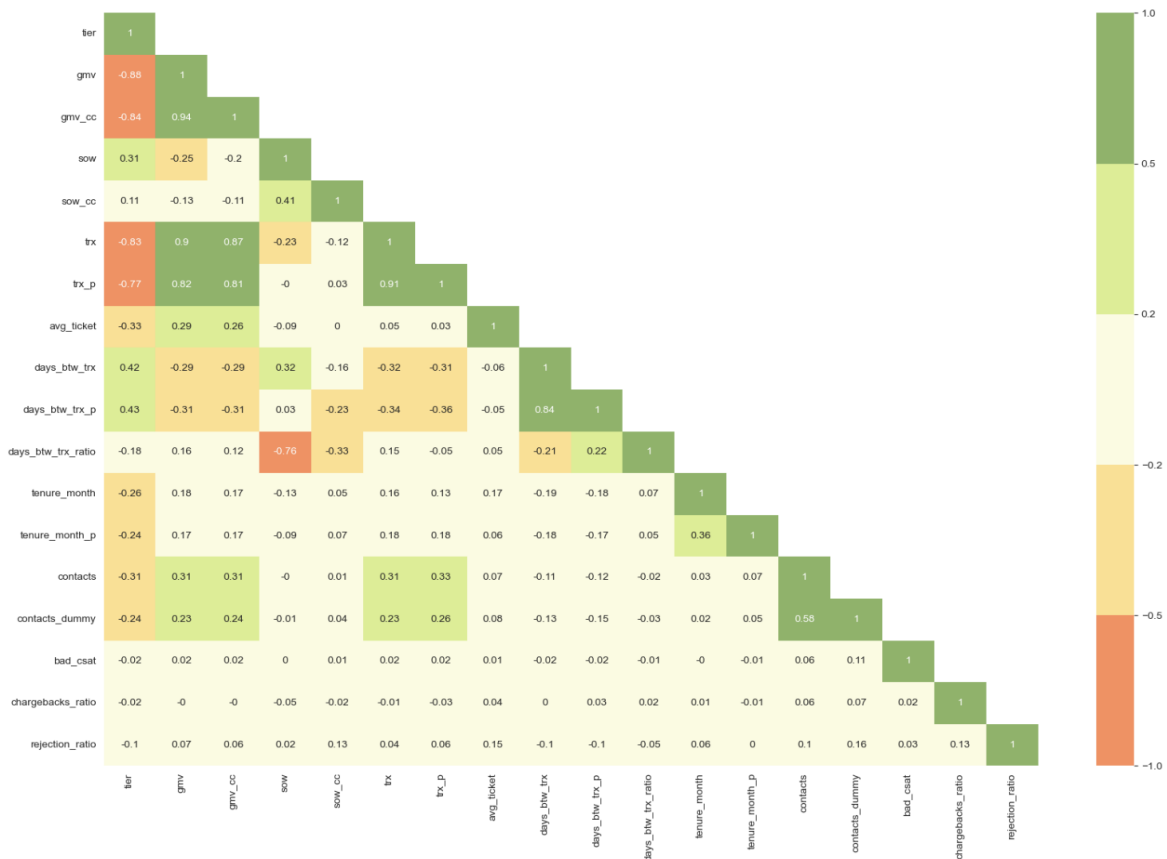
⁹ Ver Tabla 7.1 y Gráfico 7.1 en [Anexo 7.3](#)

¹⁰ Los valores de correlación se pueden ubicar entre -1 (correlación negativa perfecta) y 1 (correlación positiva perfecta), tomando 0 en caso que no exista una correlación lineal entre las variables.

importante resaltar que es el indicador utilizado por los equipos de negocio para diferenciar los servicios ofrecidos.

- El *Share of Wallet* (*SoW*) (general y el de tarjeta de crédito), tiene una correlación negativa con el GMV, comerciantes más grandes poseen un *SoW* menor. Esto se debe a que los comerciantes de mayor facturación son los que tienen más posibilidades de encontrar otros proveedores y tasas más bajas de procesamiento. A su vez, el *SoW* correlaciona de forma negativa con el ratio de periodicidad de transacciones.
- La periodicidad de las transacciones (*days_bt看_trx*), correlaciona de forma positiva con el *tier* de los comerciantes, asociado a que *tiers* más altos tienen una menor cantidad de transacciones en el mes. A su vez, es interesante resaltar que si bien la cantidad de transacciones es el componente principal del cálculo de la periodicidad (para más detalle ver [Anexo 7.2](#)), solo posee una correlación de 0.3 con esa variable.
- En cuanto a la antigüedad de los comerciantes, es interesante resaltar que posee una correlación negativa con el *tier*, y se encuentra relacionado que ante la masificación de la solución de pagos comenzaron a ingresar de forma más tardía comerciantes con menor facturación.
- Los contactos con el equipo de soporte tienen una correlación positiva con el volumen del negocio, ya que ante más operaciones y más uso de la plataforma es más probable que precise ayuda del equipo de soporte.
- De este análisis visual, no se consigue concluir algo en particular en relación a las variables de riesgo y fraude.
- Se observa que el resto de las relaciones entre variables arrojan correlaciones cercanas a 0.

Gráfico 2.5: Matriz de correlación



Por otro lado, se realizó un análisis sobre las variables en relación a la variable dependiente a predecir (*target*). En la tabla a continuación, se encuentra la media y desviación estándar del SoW para cada uno de los grupos en conjunto con el estadístico del test de medias¹¹ y su p-valor asociado. De esta forma, se encontró que la media de SoW es significativamente menor para el grupo próximo a ser *churn*, lo que da como indicio que este tipo de comerciantes utiliza menos la plataforma de pagos en relación a las ventas que realiza en general en su tienda.

En el gráfico de la derecha, se realizó un boxplot¹² sobre esta variable en cada uno de los grupos, donde además de observarse la clara diferencia entre las medianas y los cuartiles de cada grupo, se destaca que la distribución de la variable para el grupo que se mantendrá activo tiene un claro sesgo positivo con respecto al *target*.

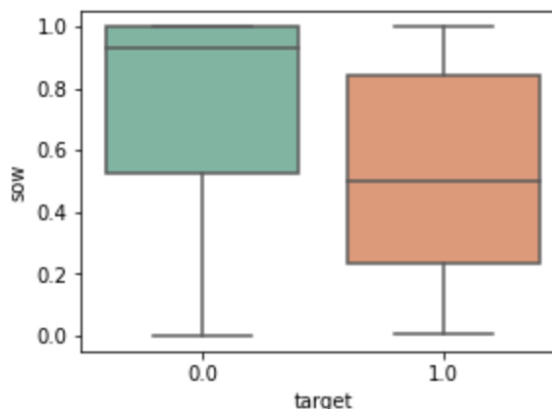
¹¹ Test que considera como hipótesis nula que la diferencia de la media entre los grupos es 0. El estadístico sigue la distribución T de Student.

¹² Método estadístico para visualizar la distribución de una variable, el cual refleja los mínimos, máximos, mediana, primer y tercer cuartil de la variable bajo análisis.

Tabla 2.5: Indicadores estadísticos de SoW

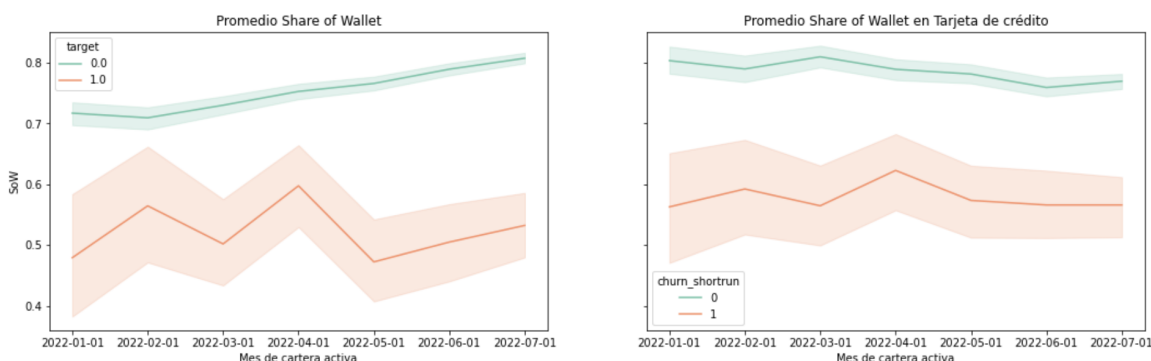
	SoW	
	Target	No churners
Mean	0.53	0.75
SD	0.33	0.30
T-test Statistic	-13.47	
P Value	0.000 ***	

Gráfico 2.6: Distribución de SoW



Estos resultados se sostienen si observamos la diferencias de medias mes a mes, tanto para SoW en general como para el SoW específico de tarjeta de crédito. De forma consistente, el promedio de SoW para el grupo retenido es un 40% mayor que el grupo que realiza *churn*.

Gráfico 2.7: Promedio de SoW



Dada la relevancia de SoW, se generó la variación mes a mes de este indicador para cada comerciante y si bien también refleja diferencias significativas entre los grupos, el signo parece ser contrario a lo que se esperaba¹³. La media de la tasa de crecimiento de SoW es del 58% para el grupo target, mientras que es del 34% para los comerciantes que no realizan *churn*. Si bien esta diferencia es significativa entre los grupos, el resultado parece ser poco robusto en el tiempo ya que para el grupo que realiza *churn* se observa una dispersión mucho más alta y la diferencia deja de ser significativa¹⁴.

¹³ Ver Tabla 7.2 y Gráfico 7.2 en [Anexo 7.3](#)

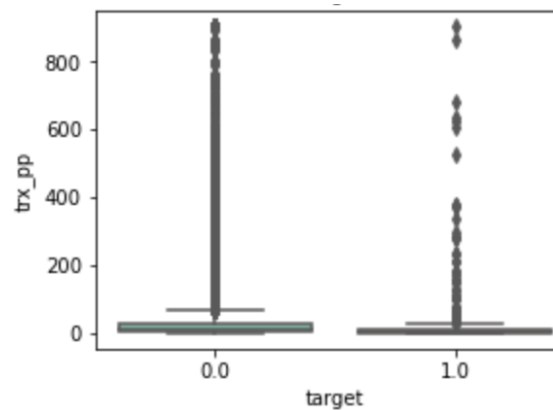
¹⁴ Ver Gráfico 7.3 en [Anexo 7.3](#)

Dentro del conjunto de variables de volumen de negocio, otra variable que refleja un comportamiento interesante es la cantidad de transacciones, donde la media de quienes realizan *churn* es un 25% menor que el grupo retenido (al 10% de significatividad). Por ende, los comerciantes que realizan menor cantidad de ventas poseen mayor riesgo de ser *churn*.

Tabla 2.6: Indicadores estadísticos de Transacciones

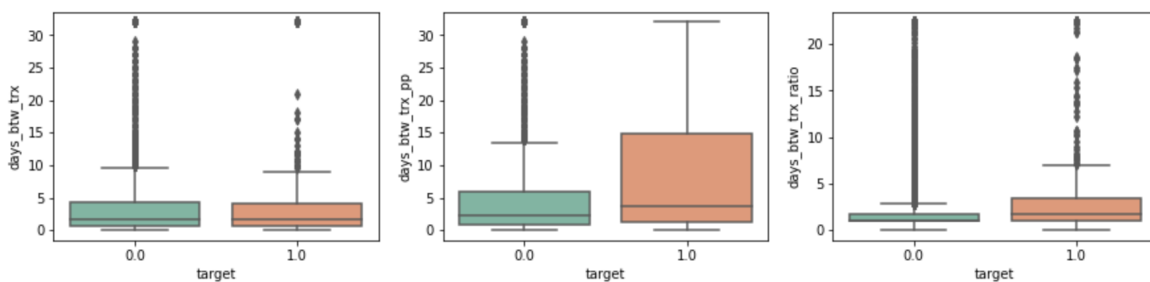
	Cantidad de ventas en la plataforma	
	Target	No churners
Mean	33	44
SD	105	122
T-test Statistic	-2.15	
P Value	0.032 **	

Gráfico 2.8: Distribución de Transacciones



En cuanto a la frecuencia de las transacciones, la cantidad promedio de días que transcurren entre transacciones de la plataforma de pagos es mayor que en la tienda en general para quienes se encuentran próximos a hacer *churn*. Este resultado es consistente también en el tiempo¹⁵.

Gráfico 2.9: Distribución de Periodicidad de transacciones



¹⁵ Ver Gráfico 7.4 en [Anexo 7.3](#)

Tabla 2.7: Indicadores estadísticos de Periodicidad

	Cantidad de días promedio entre transacciones con la plataforma		Cantidad de días promedio entre transacciones ratio (Plataforma de pagos/Total)	
	Target	No churners	Target	No churners
Mean	10	7	4.0	1.9
SD	12	10	5.7	2.7
T-test Statistic	5.28		7.30	
P Value	0.000 ***		0.000 ***	

Las variables que consideran la antigüedad del comerciante no arrojaron conclusiones consistentes en este análisis.

En cuanto a las variables de riesgo, se encontró que el ratio de *chargeback* tiene un impacto en *churn* de más corto plazo (mes actual y próximo) y no tan así sobre la ventana de tiempo que se encuentra en análisis en este estudio (2 meses). Por ende, las acciones preventivas de *churn* que estén focalizadas en este tipo de factores de riesgos deben ser realizadas a través de mejoras del sistema de antifraude aplicado por la empresa, ya que una vez que se observa un alto ratio de *chargeback* para un comerciante queda poco tiempo para retenerlo.

En cuanto al ratio de rechazos, no se pueden llegar a resultados consistentes sobre la media ya que si bien el indicador es mayor para los futuros *churns*, la diferencia con quienes no realizan *churn* no es significativa.

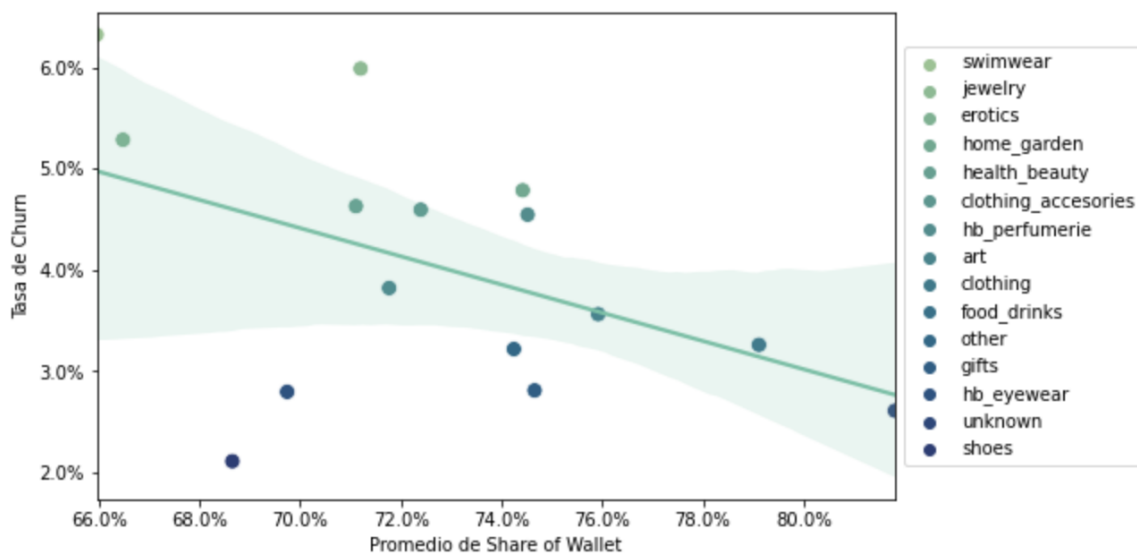
Por último, sobre la relación con la empresa, el grupo que será *churn* tiene un menor ratio de contacto que el resto¹⁶. Posiblemente, el hecho de que el comerciante se contacte para resolver una duda/problema que posee con el producto, actúa como una proxy del interés que tiene sobre el mismo. En cuanto a la calificación de CSAT, no se observan diferencias.

Además de las variables numéricas con las que se estuvo realizando este análisis, es interesante profundizar en la interacción de las mismas con la variable de segmentos. Existen 67 segmentos posibles, pero los primeros 15 concentran el 86% del total de las observaciones. En Gráfico 7.6 en [Anexo 7.3](#) se puede observar la tasa de *churn* promedio para estos rubros, donde se destacan los segmentos de *swimwear*, *jewelry* y *erotics* con un mayor porcentaje de abandono.

¹⁶ Ver Tabla 7.3 y Gráfico 7.5 en [Anexo 7.3](#)

Como identificamos anteriormente, el *SoW* es un potencial predictor de la tasa de *churn*, por lo que puede ser que la mayor tasa de *churn* de alguno de estos rubros esté relacionado a que en ellos la tasa de *SoW* es menor. En el gráfico a continuación se puede observar esta relación negativa, e identificar que de todas formas los segmentos mencionados tienen una tasa de *churn* por encima de la esperada al *SoW* promedio. También se destaca la categoría de *shoes* que dado su nivel de *SoW* tiene una tasa de *churn* sustancialmente menor.

Gráfico 2.10: Tasa de churn y SoW por segmento



En líneas generales, a través del análisis multivariado se destaca la relación que existe entre el *SoW* y la tasa de *churn*, por lo que es la que tiene un mayor potencial predictivo sobre este comportamiento. A su vez, se encontró que variables como la periodicidad de las transacciones, su volumen y los contactos con la empresa también se encontrarían relacionadas a una menor probabilidad de *churn*. A través de este tipo de análisis, no se observó una diferencia significativa por categoría de los productos que venden los comerciantes.

2.2. Técnicas de clasificación

2.2.1. Algoritmos de aprendizaje supervisado

Existen diferentes técnicas de clasificación comúnmente utilizadas en el modelado de la predicción de *churn*. En Verbeke et al (2012), realiza una lista de las diferentes técnicas observadas en la literatura.

En general, se pueden diferenciar dos grandes grupos de algoritmos de clasificación en función al número de clasificadores: modelos de algoritmo único y modelos de ensamble

(Tamaddoni, 2015). En este trabajo, se realizará una comparación entre un algoritmo perteneciente al primer grupo (regresión logística) y dos conjuntos de algoritmos en el contexto del segundo grupo (*Random Forest* y *Gradient Boosting*).

En casos como la predicción de churn, donde la variable dependiente es binaria, la regresión lineal ordinaria no es aplicable ya que permite que la variable dependiente caiga fuera del rango de 0-1. Para evitar este problema, es necesario modelar la variable dependiente $p(X)$ utilizando una función que arroje como *output* valores entre 0 y 1. Muchas funciones se adaptan a esta limitación, en este trabajo se utilizará la función logística:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (1)$$

La función logística siempre generará una curva con forma de S, donde para las probabilidades altas tenderá a 1, pero nunca por encima de ese valor, y para las probabilidades bajas se acercará a 0 pero nunca tomará un valor negativo.

Luego de manipular la fórmula anterior podemos obtener el log odds o logit, una fórmula que es lineal en X. A diferencia de un esquema de regresión lineal donde β_1 representa el cambio promedio en Y asociado por un aumento de una unidad en X, en el modelo de regresión logística, el aumento en X de una unidad modifica el log odds por β_1 .

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X \quad (2)$$

Se pueden destacar 3 grandes ventajas sobre las regresiones logísticas. En primer lugar, posee una interpretación clara y directa. Los coeficientes del modelo representan el cambio en el logaritmo de odds de la clase positiva para un cambio unitario en el predictor, lo que puede ser útil para entender la influencia de cada variable. En segundo lugar, dado que es un modelo lineal, es menos probable que sobreajuste los datos en comparación con los modelos más complejos, especialmente cuando el número de características es grande en relación con el número de observaciones. Por último, siendo modelos más sencillos, tienden a ser más rápidos para entrenar y requerir menos recursos computacionales. Esto puede ser especialmente útil en contextos donde se dispone de una gran cantidad de datos o se necesita realizar una validación cruzada exhaustiva o reentrenamientos frecuentes del modelo.

La facilidad de uso y la solidez de los resultados han hecho de la regresión logística un clasificador binario popular entre los académicos de marketing. Por lo tanto, se utilizará la regresión logística como técnica de referencia para comparar el rendimiento con respecto a modelos más sofisticados. De esta forma, se podrá ver si la complejidad añadida de

técnicas más sofisticadas vale la pena para predecir el *churn* cuando en su lugar se pueden utilizar métodos más simples como la regresión logística.

Dentro del grupo de modelos de ensamble, los que se destacan por su interpretabilidad y performance son aquellos que utilizan como base árboles de decisión. Los árboles de decisión constituyen una familia de algoritmos de aprendizaje supervisado no paramétrico, es decir, no asumen ninguna forma paramétrica para las distribuciones de clases ni una estructura fija para el árbol.

Un árbol de decisión se encuentra compuesto de nodos internos de decisión, ramas y nodos terminales. En cada nodo, se evalúan cada división de observaciones posible según las distintas variables en los datos, y se elige aquella que genera nodos más puros (con la mayor similitud posible entre las observaciones). Este proceso de división se repite de forma recursiva de arriba hacia abajo hasta que todos o la mayoría de los registros se hayan clasificado bajo etiquetas de clase específicas en los nodos terminales. Que todos los puntos de datos se clasifiquen o no como conjuntos homogéneos depende en gran medida de la complejidad del árbol de decisión.

En las tareas de clasificación, las divisiones en ramas se definen en función de su bondad de ajuste para el objetivo de aprendizaje que se cuantifica mediante una medida de impureza. Es decir, los datos se dividen en ramas para que los nodos resultantes sean lo más puros posible. Una medida común para la impureza es el índice de gini o la entropía, que reflejan valores bajos cercanos a 0 cuando el nodo es perfectamente puro. El proceso de aprendizaje de los árboles de decisión consiste en iterar todos los nodos que aún no son puros, evaluar cada posible posición de división y elegir la que tenga la mínima gini/entropía.

Entre las principales ventajas de trabajar con árboles de decisión, se destaca la interpretabilidad y la sencillez. Como contracara, por sí solos árboles de decisión no poseen gran capacidad predictiva y se los considera poco robustos, ante mínimos cambios en los datos la performance puede variar ampliamente. Sin embargo, al combinar múltiples árboles de decisión la performance puede ser mejorada sustancialmente.

Los métodos de ensamble se conciben como algoritmos de aprendizaje que construyen un conjunto de clasificadores y luego clasifican nuevos puntos de datos tomando un voto ponderado de sus predicciones. Existen 3 razones principales por la que esto conduce a una mejora en la performance por sobre los métodos de aprendizaje estándar (Dietterich, 2000):

En primer lugar, un algoritmo de aprendizaje puede verse como una búsqueda en un espacio H de hipótesis para identificar la mejor hipótesis en el espacio. El problema estadístico surge cuando la cantidad de datos de entrenamiento disponibles es demasiado pequeña en comparación con el tamaño del espacio de hipótesis. Sin datos suficientes, el algoritmo de aprendizaje puede encontrar muchas hipótesis diferentes en H que dan la misma precisión en los datos de entrenamiento. Al construir un conjunto a partir de todos estos clasificadores precisos, el algoritmo puede promediar sus votos y reducir el riesgo de elegir el clasificador incorrecto.

En segundo lugar, se debe a una cuestión computacional. Muchos algoritmos de aprendizaje funcionan mediante la realización de algún tipo de búsqueda que puede atascarse en los óptimos locales. Los algoritmos de árbol de decisión emplean una regla de división codiciosa para hacer crecer el árbol de decisión: evalúan las divisiones una por vez y deciden qué división elegir independientemente de las posibles divisiones subsiguientes. Por lo tanto, un conjunto construido mediante la ejecución de la búsqueda local desde muchos puntos de partida diferentes puede proporcionar una mejor aproximación a la verdadera función desconocida que cualquiera de los clasificadores individuales.

Por último, en muchas aplicaciones de aprendizaje automático, la verdadera función f no puede representarse mediante ninguna de las hipótesis en H . Incluso con algoritmos muy flexibles como árboles de decisión, como la muestra de entrenamiento es finita, explorarán un conjunto finito de hipótesis y dejarán de buscar cuando encuentran uno que se ajuste a los datos de entrenamiento. Formar sumas ponderadas de hipótesis a partir de H puede expandir el espacio de funciones representables.

En cuanto a cómo se construye el conjunto, se han desarrollado diferentes técnicas. Algunos de ellos funcionan particularmente bien con clasificadores de árboles de decisión, más específicamente: bagging y boosting.

Bagging

El *Bagging* es una técnica de ensamble que busca disminuir la varianza en las predicciones de los modelos de aprendizaje estadístico. La premisa subyacente se basa en un principio estadístico fundamental: dado un set de observaciones independientes Z_1, \dots, Z_n siendo la varianza de cada una σ^2 , la varianza de la media de las observaciones (\bar{Z}) será $\frac{\sigma^2}{n}$. De manera intuitiva, esto significa que el promedio de un conjunto de observaciones reduce la varianza, lo cual podría aumentar la asertividad de la predicción.

Una forma de implementar este concepto es tomando múltiples conjuntos de entrenamiento, construyendo un modelo separado para cada uno, y luego promediando las predicciones. En la práctica, sin embargo, rara vez se dispone de múltiples conjuntos de entrenamiento. En su lugar, se puede recurrir a la toma de muestras bootstrap, donde se generan múltiples muestras con reemplazo a partir de un solo conjunto de entrenamiento. Cada una de estas muestras se usa para entrenar un modelo separado, y las predicciones se promedian.

En el contexto de los árboles de decisión, *bagging* permite a cada árbol crecer a su máxima capacidad, resultando en modelos individuales con bajo sesgo y alta varianza. Al promediar las predicciones, se busca reducir la varianza global sin incrementar sustancialmente el sesgo.

Una mejora sobre la técnica de ensamblado simple es la que aplica el algoritmo *Random Forest*. En lugar de permitir que cada árbol considere todos los predictores en cada división, *Random Forest* limita a cada árbol a un subconjunto aleatorio de predictores. Esto disminuye la correlación entre los árboles, lo que puede resultar en un rendimiento mejorado. Al permitir que los árboles se entrenen de forma independiente, *Random Forest* facilita la computación paralela, lo que puede acelerar sustancialmente el proceso de entrenamiento.

Hay dos hiperparámetros clave que deben tenerse en cuenta al aplicar técnicas de *bagging*:

- El número de clasificadores (o árboles de decisión) a generar: A diferencia de otros métodos, en *bagging* y *Random Forest*, aumentar el número de clasificadores no suele llevar al sobreajuste. Por tanto, se puede seleccionar este parámetro basándose en el punto en el que la tasa de error se estabiliza.
- La cantidad de variables a considerar al construir cada árbol: Este parámetro controla el grado de descorrelación entre los árboles. Cuanto menor sea el número de características utilizadas, menos correlacionados estarán los clasificadores, lo que puede mejorar la precisión del ensamble. Sin embargo, también puede aumentar el costo computacional, por lo que debe elegirse con cuidado.

En resumen, las técnicas de *bagging* y *Random Forest* representan formas poderosas y flexibles de mejorar la precisión de los modelos de aprendizaje estadístico, especialmente aquellos basados en árboles de decisión.

Boosting

El *boosting* es una poderosa técnica de ensamble que, al igual que el *bagging*, puede aplicarse a varios métodos de aprendizaje estadístico. No obstante, para el propósito de esta explicación, limitaremos su aplicación al contexto de árboles de decisión. Mientras que

el *bagging* permite el entrenamiento independiente de los clasificadores, el *boosting* adopta un enfoque más sofisticado y secuencial.

El proceso de *boosting* implica el entrenamiento de árboles de decisión de manera secuencial, en donde cada árbol se construye utilizando la información obtenida de los árboles que se han entrenado previamente. Este enfoque ofrece una ventaja significativa al ayudar a prevenir el sobreajuste, que es un problema común cuando se intenta crear un árbol de decisión grande. En lugar de crear un árbol grande, el *boosting* busca aprender lentamente mediante la construcción de árboles más pequeños y manejables.

En términos más técnicos, cada árbol en el proceso de *boosting* se construye con el objetivo de predecir los residuos, o errores, del clasificador anterior, en lugar de intentar predecir directamente la variable dependiente. Este enfoque significa que cada nuevo árbol se enfoca en "corregir" las predicciones erróneas del árbol anterior, mejorando así la función de decisión en las áreas donde se está desempeñando deficientemente.

En el proceso de *boosting*, hay tres hiperparámetros principales que juegan un papel importante en su rendimiento:

- El número de clasificadores entrenados: Este parámetro tiene un impacto significativo en la posibilidad de sobreajuste. A diferencia del algoritmo *Random Forest*, en el que se pueden entrenar numerosos clasificadores sin mucha preocupación por el sobreajuste, en *boosting*, un número excesivamente grande de clasificadores puede llevar a un ajuste excesivo del modelo a los datos de entrenamiento.
- La tasa de aprendizaje: Este parámetro, a menudo denotado como λ , controla la velocidad a la que el algoritmo de *boosting* aprende de los errores del clasificador anterior. Valores muy pequeños de λ harán que el aprendizaje sea más lento, y requerirán un mayor número de clasificadores para obtener un buen rendimiento. Sin embargo, una tasa de aprendizaje demasiado alta puede hacer que el modelo se ajuste demasiado a los datos de entrenamiento, lo que reduce su capacidad de generalización a nuevos datos.
- La máxima profundidad de los árboles: Este parámetro de regularización controla el número máximo de divisiones que puede tener un solo árbol. Una profundidad más grande puede permitir que el modelo capture interacciones más complejas, pero también puede hacer que el modelo sea más propenso al sobreajuste. Por lo tanto, este parámetro necesita ser ajustado cuidadosamente para equilibrar la capacidad de capturar interacciones y la posibilidad de sobreajuste.

En resumen, el *boosting* es una técnica avanzada que mejora la precisión de los modelos de aprendizaje automático al aprender de manera secuencial a partir de los errores de los clasificadores anteriores.

2.2.2. Criterio de evaluación de modelos

Antes de definir las métricas con las que se evaluarán los modelos, es importante detenerse en la definición de las muestras de datos donde se entrenará el modelo y se evaluará su performance. Uno de los mayores peligros al aplicar métodos de aprendizaje automático es que el modelo puede llegar a ser muy bueno clasificando las observaciones con las que entrena, pero no puede generalizar para observaciones que no ha visto previamente.

Una de las formas para reducir este riesgo es la validación cruzada. Este proceso mide la capacidad de generalización de diferentes modelos probándolos en nuevos datos, no vistos durante el entrenamiento, y luego eligiendo el más preciso en este conjunto de datos.

Existen diferentes técnicas disponibles para realizar la validación cruzada. Una de las estrategias usadas consiste en dividir el conjunto disponible de observaciones en dos partes: un conjunto de entrenamiento y un conjunto de validación. El modelo se ajusta usando solo el conjunto de entrenamiento, y el modelo entrenado se usa para predecir respuestas en el conjunto de validación. El principal inconveniente de usar este enfoque tiene que ver con el hecho de que solo se usa un subconjunto de las observaciones para entrenar el modelo (menos datos significan que el aprendizaje del modelo será más pobre) y el desempeño se mide solo en un subconjunto de observaciones (por lo tanto, el error medido puede ser más variable, ya que se calcula con menos datos).

Como en este trabajo, uno de los objetivos es que el modelo no solo tenga una buena performance en datos fuera de la muestra, sino que también su validez se sostenga en el tiempo, se dividió el conjunto de datos en 3 partes: el set de entrenamiento, de validación y de testeo. El set de entrenamiento concentra el 75% de las observaciones del período enero a julio seleccionadas de forma aleatoria. El 25% restante se asignó al set de validación, utilizado para optimizar la performance de los modelos entrenados.

El set de testeo concentra todas las observaciones de agosto y septiembre, lo que permite analizar cómo sería la performance del modelo si fuera aplicado a sets de datos completamente nuevos. Sólo será utilizado en última instancia para comprobar que el mejor modelo entrenado con datos de enero a julio mantiene su performance y estabilidad temporal en los meses más recientes.

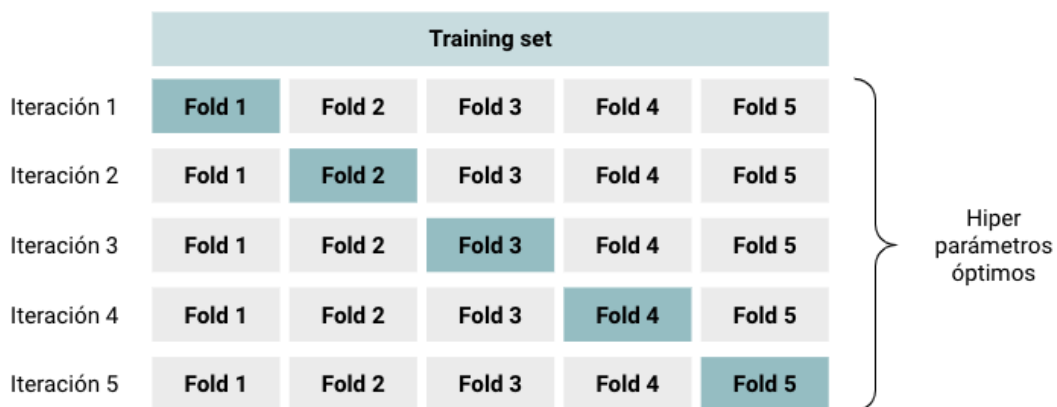
Figura 2.1: División del dataset



Además, se utilizará un enfoque de *k-folds*, el cual implica dividir aleatoriamente el conjunto de observaciones en k grupos del mismo tamaño. El modelo se ajusta k veces, cada una de ellas usando $k - 1$ *folds* para entrenamiento y el restante omitido para validación, mientras se rota el *fold* que queda afuera. Al final del proceso, hay k estimaciones diferentes del rendimiento del modelo que se promedian para estimar el rendimiento de la validación cruzada k veces.

De esta forma, se podrá comparar el rendimiento entre diferentes combinaciones de hiper parámetros para el mismo algoritmo de aprendizaje automático. Hay una razón principal por la que esto tiene sentido: al optimizar los hiperparámetros, se debe ajustar una cantidad significativa de modelos. Si todos ellos están entrenados y validados con la misma división estática de validación, se podría terminar sobre-ajustando los datos de validación.

Figura 2.2: Validación cruzada



Al combinar este conjunto de estrategias, por un lado, se obtiene un conjunto de validación lo suficientemente grande que permitiría obtener una medida suficientemente sólida de la capacidad de generalización de los modelos, mientras que a través de *k-fold* se evita un sobreajuste en los datos de validación al optimizar los hiper parámetros.

2.2.3. Métricas de evaluación de modelos

Para evaluar el rendimiento predictivo de los modelos construidos, se han empleado dos criterios: el área bajo la curva ROC (Receiver Operating Characteristic) y la *cumulative lift curve*.

La forma más simple de datos producidos por un modelo de clasificación es una matriz de confusión de conteos de casos clasificados correctamente y mal clasificados de cada clase. En un caso de predicción de abandono, un clasificador binario predice el estado de un cliente como *churner* o no *churner*. Dicho clasificador podría clasificar a un verdadero *churner* como "*churner*" (verdadero positivo), a un *churner* real como "no *churner*" (falso negativo), a un no *churner* real como "no *churner*" (verdadero negativo), y a un no *churner* real como "*churner*" (falso positivo).

Usando la matriz de confusión, se puede extraer la tasa de *false positives* (FP/N) y la tasa de *true positives* (TP/P). Al graficar los pares FP y TP para diferentes umbrales de decisión, se puede lograr la curva característica operativa del receptor (ROC) (Bradley, 1997). En otras palabras, una curva ROC traza las tasas de verdaderos positivos frente a las tasas de falsos positivos para un sistema clasificador binario a medida que varía su umbral de decisión.

Sin embargo, dado que la curva ROC brinda el rendimiento de los modelos con dos medidas, puede haber casos en los que aumentar la tasa de verdaderos positivos provoque un aumento en la tasa de falsos positivos. En tales casos, la comparación de dos modelos de forma gráfica se convertiría en una tarea desafiante. Para comparar clasificadores es necesario reducir la medida del desempeño a un solo valor escalar y ahí es donde AUC-ROC entra en escena (Fawcett, 2006). Dado que es una porción de un área de un cuadrado unitario, se encuentra entre 0 y 1: cuanto mayor sea su valor, mejor será el clasificador. Este valor es equivalente a la probabilidad de que un clasificador clasifique una instancia positiva elegida al azar más alta que una instancia negativa elegida al azar.

Para abordar este problema, en este estudio se compara el desempeño de los modelos en función del área bajo su curva ROC (AUC). Mientras que un clasificador aleatorio posee un AUC de 0.5, un clasificador ideal tiene un AUC igual a 1. Sobre esta base, todos los clasificadores en la práctica tienen un AUC mayor que 0.5 y menor que 1. El área bajo la curva ROC se calcula de la siguiente manera:

$$AUC = \int_0^1 \frac{TP}{P} d\frac{FP}{N} = \frac{1}{P \cdot N} \int_0^N TP dFP. \quad (3)$$

Uno de los principales beneficios de utilizar el área bajo la curva como una de las métricas de performance es que no se encuentra afectada por el importante desbalance entre la clase positiva y la negativa presente en el dataset.

Como segundo criterio de evaluación se emplea la *cumulative lift curve* (Jamal & Bucklin, 2006). El enfoque de esta medida está en los clientes con la mayor probabilidad de ser positivos (es decir, *churner* en el caso del modelo de *churn*) y se define como la proporción de positivos en un segmento, dividida por la proporción de positivos en todo el conjunto de prueba (Berry & Linoff, 2004). Por ejemplo, si no se utiliza ningún modelo, se espera que al dirigirse al N% de la base de clientes con incentivos, se apunte al N% de los abandonos reales. Sin embargo, mediante el uso de un modelo de clasificación con una elevación del percentil n-ésimo superior de c , se espera que $c \times N\%$ de todos los abandonos reales reciban el incentivo, cuando se apunta al N% superior de clientes. En otras palabras, para que un modelo se desempeñe mejor que el clasificador aleatorio, su elevación del decil superior c debe ser mayor que 1. Por lo tanto, una elevación alta del decil superior es un indicador del desempeño prometedor del clasificador.

3. Resultados

3.1. Performance de los modelos

Como se mencionó previamente, el objetivo es encontrar aquel modelo que posea una mejor performance con el menor esfuerzo posible. Una mayor cantidad de variables no sólo afecta el tiempo de cómputo, el cual una vez ejecutado deja de tener tanto impacto para el uso que se le quiere dar, sino más bien complejiza el mantenimiento del mismo. Con este enfoque, en una primera instancia se experimentó con los 3 tipos de algoritmos previamente mencionados, con sus parámetros en default, y 3 sets distintos de *features*, para identificar el mejor acercamiento posible al problema y acotar el espacio de variables a ser utilizadas.

El primer set concentra sólo dos variables, el *Share of Wallet*, que como se observó en el análisis exploratorio es la que mayor relación tiene con la tasa de *churn*, y el *tier* del comerciante a modo control, ya que el *SoW* se encuentra correlacionado negativamente a la facturación.

En el segundo set de variables, se incluyen, además de *SoW* y *tier*, todas las variables numéricas restantes¹⁷.

Por último, en el tercer set, se incluye además de las variables numéricas, la única variable categórica con la que se cuenta, el segmento de la tienda. Para incluirla en el modelo, en una etapa del procesamiento se aplica *OneHotEncoder*¹⁸, para convertir en variables binarias cada uno de los segmentos disponibles. Como consecuencia, se pasa de tener 15 a 82 variables, por lo que escala significativamente el tiempo de cómputo y afecta la capacidad de predicción de los modelos ya que la cantidad de observaciones no es tan alta.

En la tabla a continuación se pueden observar para cada combinación de modelo y el set de variables las principales métricas de performance:

- Área bajo la curva en los conjuntos de entrenamiento y validación.
- Área bajo la curva en el conjunto de validación, considerando sólo los comerciantes de mayor facturación (tiers menor e igual a 6).
- Ratio de verdaderos positivos para los primeros 3 percentiles de la distribución.

¹⁷ Además se incluyen las siguientes variables: 'gmv', 'gmv_cc', 'sow_diff', 'trx', 'trx_p', 'days_btw_trx_p', 'days_btw_trx_ratio', 'avg_ticket', 'contacts_dummy', 'tenure_month', 'tenure_month_p', 'chargebacks_ratio', 'rejection_ratio'.

¹⁸ *OneHotEncoder* es una técnica de preprocesamiento de datos que convierte variables categóricas en un formato numérico. Para cada categoría única en la variable, crea una nueva variable binaria que toma el valor 1 si la observación pertenece a esa categoría, y 0 en caso contrario.

Modelo	Nombre	Set de datos	Train AUC	Validation AUC	Validation AUC (Tiers<=6)	Overfitting (%)
Logistic Regression	Tier + SoW	1	0.721	0.722	0.620	-0.1%
Random Forest	Tier + SoW	1	0.742	0.705	0.597	5.2%
Gradient Boosting	Tier + SoW	1	0.782	0.714	0.603	9.6%
Logistic Regression	Variables numéricas	2	0.732	0.721	0.636	1.5%
Random Forest	Variables numéricas	2	0.768	0.728	0.635	5.5%
Gradient Boosting	Variables numéricas	2	0.832	0.736	0.644	13.0%
Logistic Regression	Variables numéricas + Segmento	3	0.746	0.718	0.648	3.9%
Random Forest	Variables numéricas + Segmento	3	0.774	0.735	0.639	5.2%
Gradient Boosting	Variables numéricas + Segmento	3	0.833	0.737	0.638	12.9%

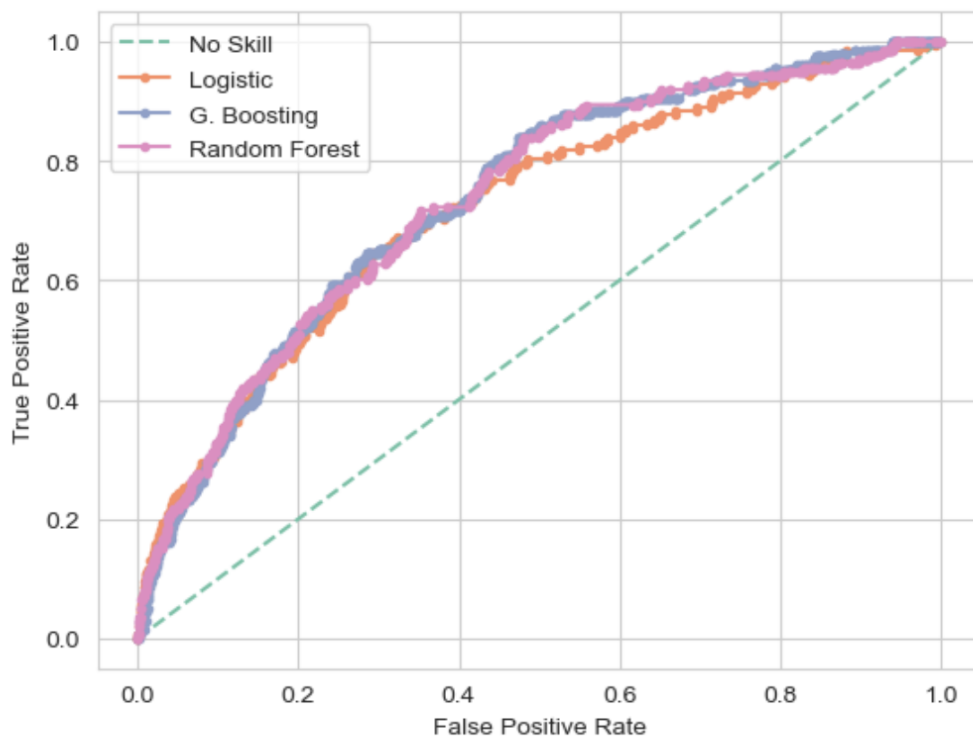
La inclusión de nuevas variables en el set 2 y 3, no refleja una mejora significativa en los indicadores de performance de los modelos. En general el conjunto de variables donde se obtiene la mejor performance es el set 2, que concentra todas las variables numéricas mencionadas previamente. La inclusión de la clasificación de las tiendas por segmento no parece tener un impacto significativo en la performance de los modelos ni en entrenamiento ni validación. En parte se debe a que la mayor capacidad explicativa de las tasas de *churn* por segmento se debe también al *SoW* de esas aperturas, y que, en los casos donde esta relación difiere, son de segmentos con pocas observaciones que podrían llevar a resultados poco robustos.

Es relevante destacar que si bien el modelo que posee mejores indicadores es *Gradient Boosting* (AUC score de 0.736 en validación), tanto el algoritmo de *Random Forest* (AUC score de 0.728 en validación) como la regresión logística tienen una performance cercana

(AUC score de 0.721 en validación), la cual es similar al baseline de este trabajo (regresión logística con solo dos variables -SoW + Tier-).

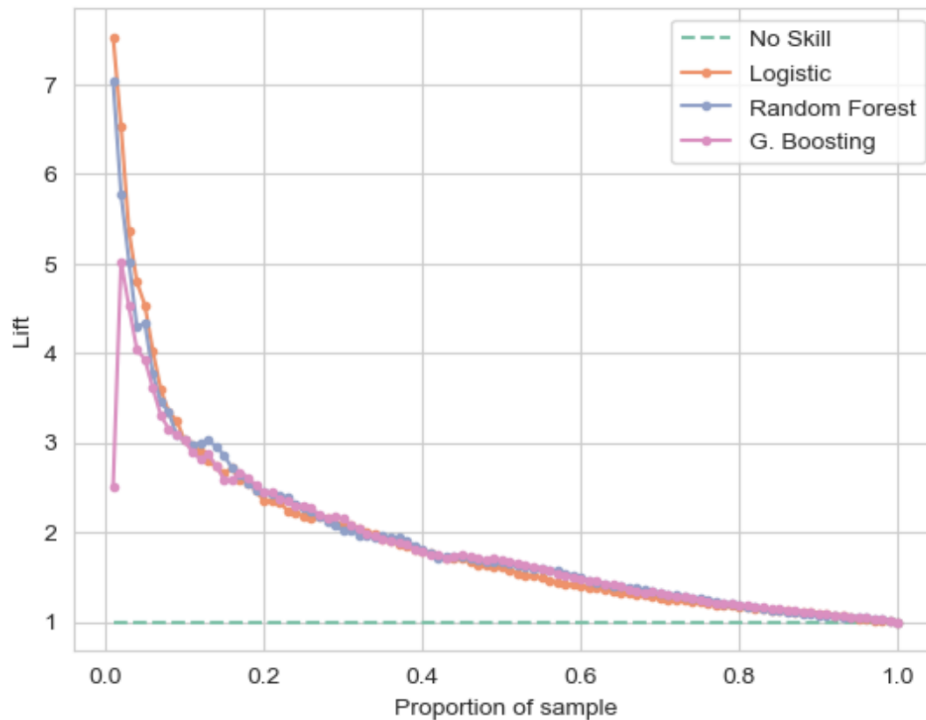
Debajo se puede observar la curva ROC para los modelos procesados con el segundo set de variables. En todos los casos, la performance es mayor a los de un escenario de aleatorización completa (*No Skill*) y se observa como la curva de *Gradient Boosting* es la que más logra acercarse a la esquina superior izquierda.

Gráfico 3.1: Curva ROC - Set 2 de variables



A pesar de ser el modelo con mayor *overfitting*, *Gradient Boosting* posee el mayor *gain* en el ordenamiento de la probabilidad de *churn*, concentrando la mayor cantidad de true positives en los 3 percentiles superiores (64.6% vs. un promedio de 61.9%). Debajo se puede observar la *Lift curve* para los 3 modelos versus el caso de aleatorización completa. Si bien la regresión logística predomina en los registros dentro del 5% superior, luego la performance de *Gradient Boosting* llega a igualarse y hasta mejorar levemente con respecto a ese baseline.

Gráfico 3.2: Curva Lift - Set 2 de variables



Los hiperparámetros que fueron considerados en esta comparación fueron:

- *Logistic*: {C = 1.0, max_iter = 400, class_weight='balanced'}
- *Random Forest*: {n_estimators= 100, max_features= sqrt(n_features), max_depth=3, class_weight='balanced'}
- *Gradient Boosting*: {n_estimators=100, learning_rate = 0.1, max_depth=3}

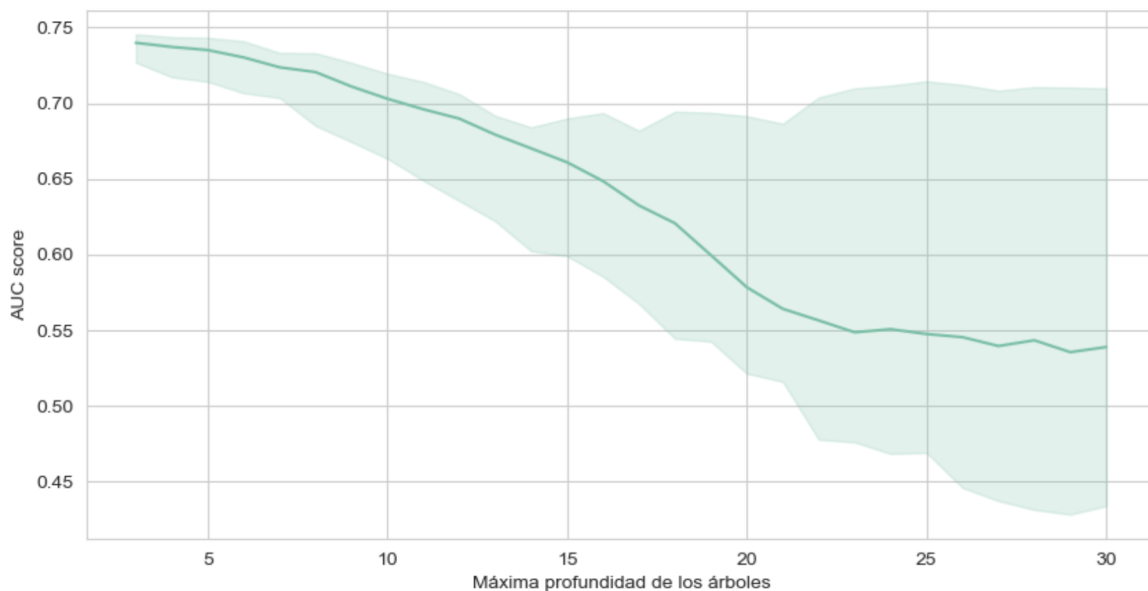
Para realizar la optimización de los hiperparámetros se utilizó *Grid Search*. Esta metodología consiste en definir un espacio de búsqueda estructurado, entrenando y evaluando el modelo para cada combinación de hiperparámetros. Al término del proceso, se selecciona la configuración que ofrece el mejor rendimiento según la métrica de evaluación seleccionada.

A pesar de su eficacia, la búsqueda en cuadrícula puede ser computacionalmente exigente, especialmente con un gran número de hiperparámetros y/o un amplio espacio de búsqueda. Por ello, para casos más complejos se pueden considerar métodos más eficientes como la búsqueda aleatoria o la optimización bayesiana. Sin embargo, la búsqueda en cuadrícula sigue siendo una opción sólida y de fácil implementación para la optimización de hiperparámetros.

Dado que en este caso, tanto el número de *features* como la cantidad de observaciones son acotadas a través de *Grid Search* se realizó una búsqueda en un espacio amplio de combinaciones de hiperparámetros, para encontrar aquella que maximice el área debajo de la curva con *Gradient Boosting*. Como resultado, se procesaron 2.688 clasificadores, en los que se varió la profundidad de los árboles, la cantidad de estimadores y la tasa de aprendizaje.

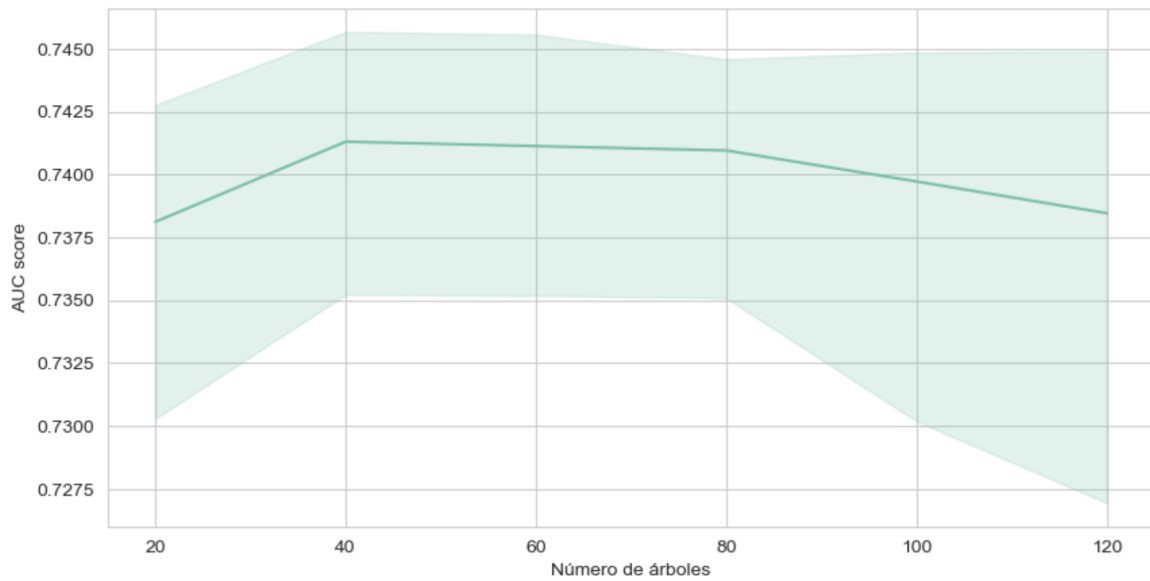
Debajo se puede observar cómo en la medida que se permite que los árboles ganen mayor profundidad, el modelo tiende a hacer *overfitting*, reduciendo su performance en el set de testeo de validación cruzada. Por ende, la mejor combinación de parámetros permite una profundidad máxima de 3.

Gráfico 3.3: Profundidad de los árboles



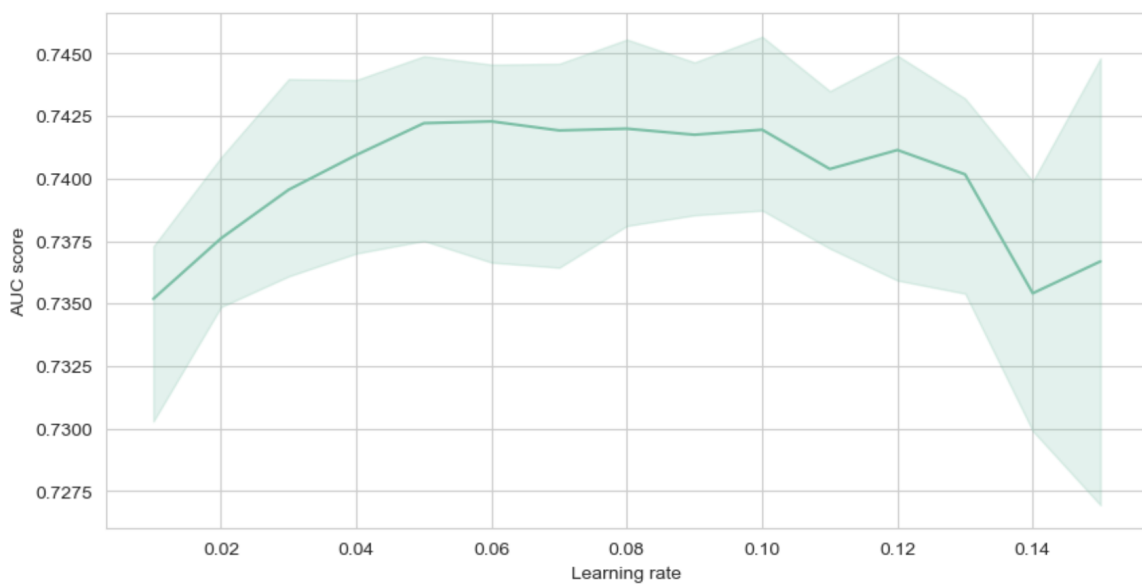
Considerando sólo las combinaciones donde la profundidad máxima de los árboles es de 3, encontramos que la cantidad de árboles que logran una mayor performance se encuentra entre los 40 - 60 estimadores.

Gráfico 3.4: Número de árboles



Finalmente, si observamos el *score* AUC con respecto a la tasa de aprendizaje, se destaca que en el rango de 0.08 a 0.1 se logra llegar a valores similares de performance.

Gráfico 3.5: Tasa de aprendizaje



La combinación final con la se obtiene la mejor performance (AUC *score* promedio de 0.745) es:

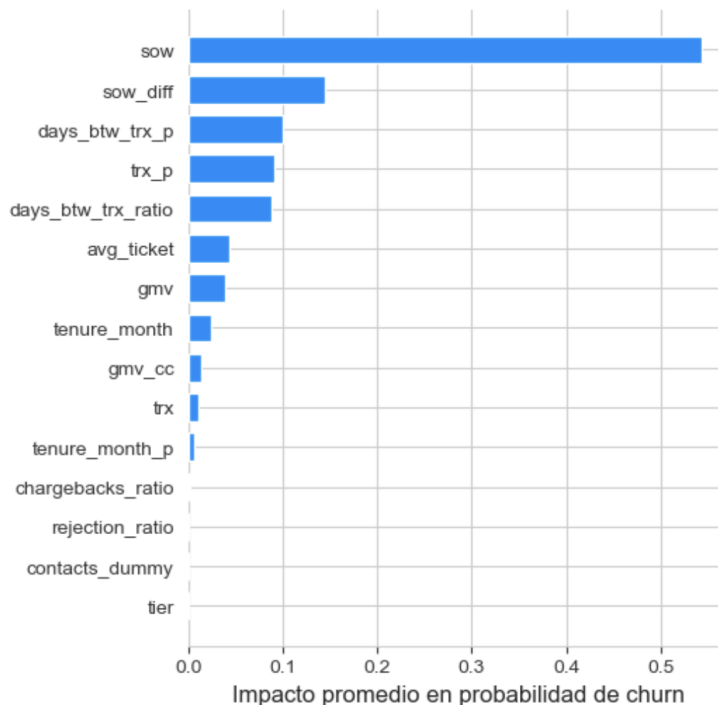
- *Gradient Boosting*: {*n_estimators* = 40, *learning_rate* = 0.1, *max_depth*=3}

La principal variable que actúa como predictor en el modelo seleccionado es *Share of Wallet* (en niveles y en diferencias). Valores más altos tienden a reducir la probabilidad de *churn* de forma significativa.

A su vez, las variables que reflejan el volumen transaccionado de la tienda también tienen una mayor importancia relativa en el modelo (transacciones, días entre transacciones, GMV, ticket promedio).

Finalmente, aquellas variables con menor peso en el modelo final se encuentran asociadas a el ratio de chargebacks, rechazos y los contactos con la empresa.

Gráfico 3.6: Importancia de variables explicativas



3.2. Modelo final

Con el modelo seleccionado a través de la optimización de los hiperparámetros en *Gradient Boosting*, se procesó nuevamente el modelo eliminando aquellas variables explicativas que tenían impacto nulo en el resultado final. De esta forma podremos comparar los resultados finales en entrenamiento, validación y testeo con el modelo que definimos como baseline previamente: la regresión logística con sólo dos variables explicativas (*SoW* y *Tier*).

Se destaca que si bien *Gradient Boosting* posee una mayor tasa de *overfitting* (la performance cae un 7.8% en el set de validación), la regresión logística pierde poder

predictivo en el conjunto de datos de testeo. Mientras que el modelo optimizado posee una mejora de la performance en ese set de datos, los cuales pertenecen a los meses de agosto y septiembre y por ende son más relevantes y recientes para el caso de uso actual.

Gráfico 3.7: Comparación de performance

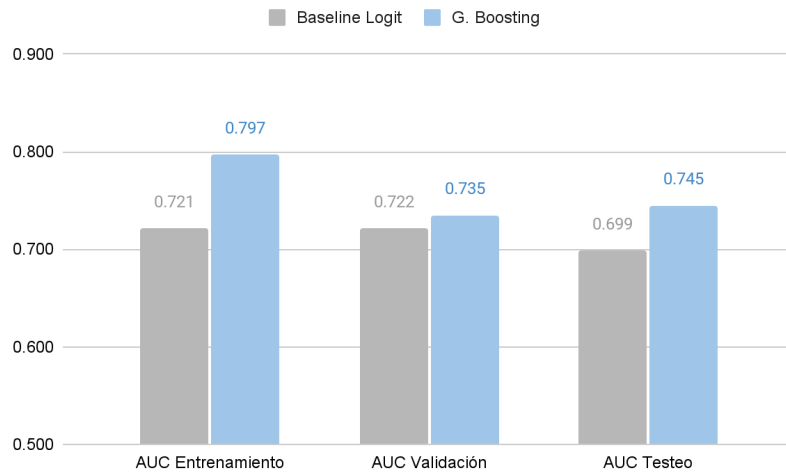


Gráfico 3.8: Curva ROC

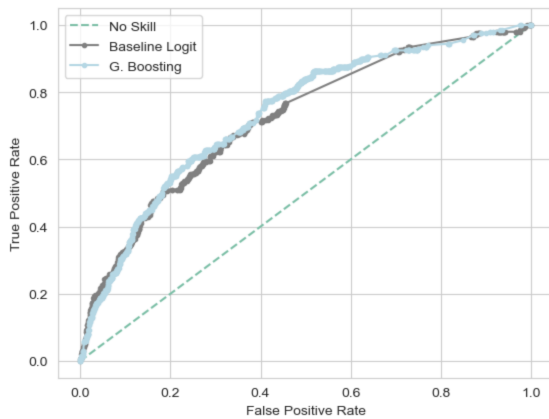
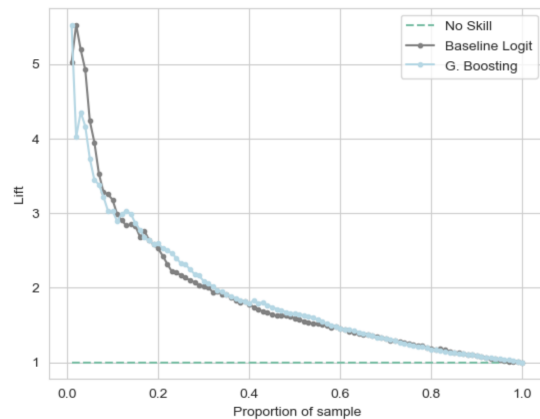


Gráfico 3.9: Curva Lift



3.3. Estabilidad del modelo

Existen diversas fuentes de cambio en la distribución de las poblaciones que pueden hacer que un modelo vaya perdiendo efectividad con el tiempo. Estas fuentes pueden deberse a cambios socioeconómicos de los comerciantes, de las políticas comerciales de la empresa, cuestiones no incluidas en el modelo como estacionalidades, entre otros.

Para monitorear que estos factores no influyen en la robustez del modelo en el corto plazo, se pueden generar indicadores sobre la estabilidad del *output* del modelo así como de las *features*.

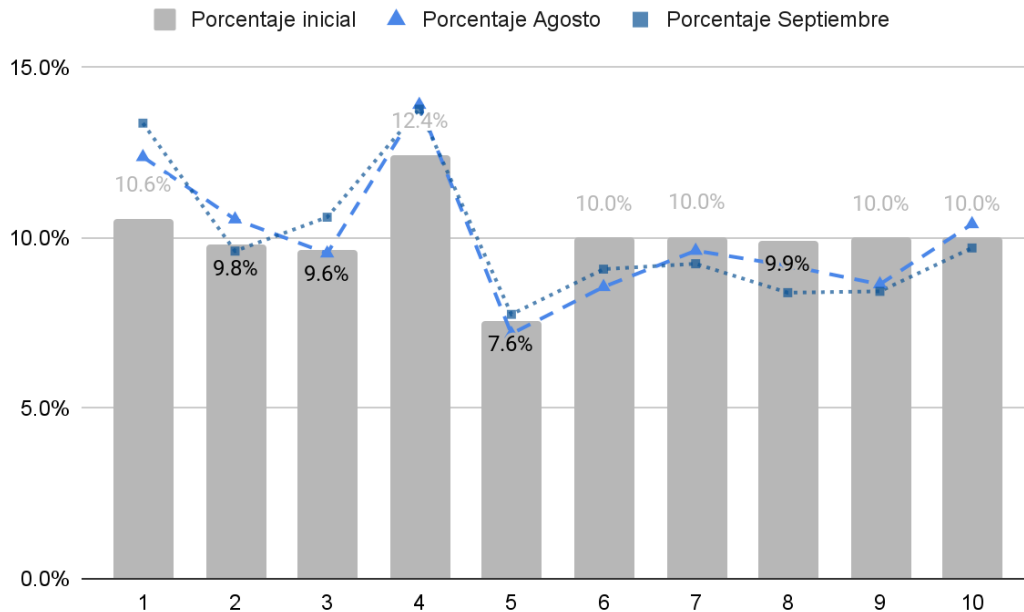
En este trabajo, se aplicará el Índice de Estabilidad de la Población (PSI por sus siglas en inglés), una medida para determinar si ha cambiado significativamente la población en dos momentos diferentes. El PSI puede ayudar a identificar *drifts* en los datos, es decir, cambios en la distribución de los datos con el tiempo que pueden afectar la efectividad del modelo.

Una importante ventaja de utilizar este índice a diferencia de otras métricas de rendimiento del modelo, es que se puede calcular sin necesidad de tener los datos reales de nuestra variable dependiente. En este contexto, donde la variable *target* tiene un rezago de 3 meses para evaluar su performance, es de suma importancia lograr cierta anticipación a una pérdida del poder predictivo que pueda tener el algoritmo.

A continuación, se analizará cómo cambió la distribución de la población en términos de la variable predicha (la probabilidad de *churn*) del modelo. Para ello, primero, creamos 10 contenedores de nuestros datos de entrenamiento del mes de julio dividiendo el rango de la variable pronosticada en secciones del mismo tamaño. Luego comparamos el porcentaje de la población dentro de cada contenedor con los nuevos datos generados a partir del set de testeo (agosto y septiembre).

En el gráfico a continuación se puede observar la distribución de las observaciones por el mes bajo análisis. Se destaca que el porcentaje inicial no es uniforme, ya que existen muchos registros en el mes de julio que tienen un *score* que cae cerca del punto de corte del contenedor 4 con respecto al 5.

Gráfico 3.10: Distribución de la probabilidad de churn



De todas formas, la distribución de los meses subsiguientes permanece similar a la del mes inicial. Con estos datos se procedió a calcular el índice de estabilidad poblacional de la siguiente forma:

$$PSI_{Agosto} = \sum (Porcentaje julio (\%) - Porcentaje agosto (\%)) * \ln\left(\frac{Porcentaje julio (\%)}{Porcentaje agosto (\%)}\right) \quad (4)$$

No existe un umbral universalmente aceptado para determinar cuando el valor del PSI es alto o no (Yurdakul, 2018). El mismo puede depender del contexto específico y de la tolerancia al cambio en la distribución de los datos. De todas formas, convencionalmente, se toman en cuenta los siguientes parámetros:

- $PSI < 0.1$ = La población no ha cambiado y podemos mantener el modelo
- $0.1 \leq PSI < 0.25$ = La población ha cambiado levemente, y es recomendable evaluar los impactos de estos cambios
- $PSI \geq 0.25$ = Los cambios en la población son significativos y el modelo debe volver a entrenarse o incluso rediseñarse.

Para los meses de agosto y septiembre, se obtuvo un PSI de 0.0105 y 0.0159 respectivamente. Por lo que el modelo sostiene su validez por fuera del período de entrenamiento.

Así como resulta una herramienta útil para detectar cambios en la performance del modelo de forma anticipada, el PSI tiene algunas limitaciones que es importante tener en cuenta al querer replicar el análisis. El indicador puede ser sensible a la selección del número de bins

y al tamaño de la muestra, lo que puede afectar su estabilidad y fiabilidad. A su vez, si bien indica que existe un cambio en la distribución de los datos, no proporciona información directa sobre la magnitud de este cambio o cómo puede afectar el rendimiento del modelo.

4. Aplicaciones del modelo predictivo

Considerando el análisis exploratorio sobre la tasa de *churn* y el modelo estimado para lograr predecir la probabilidad de inactividad por parte de los comerciantes, se trabajó con el equipo de negocio para orientar y generar acciones con la intención de incrementar la tasa de retención del producto.

Por un lado, a través de la identificación de cuáles eran las variables más relevantes en el modelo, se realizaron algunas recomendaciones al negocio tales como:

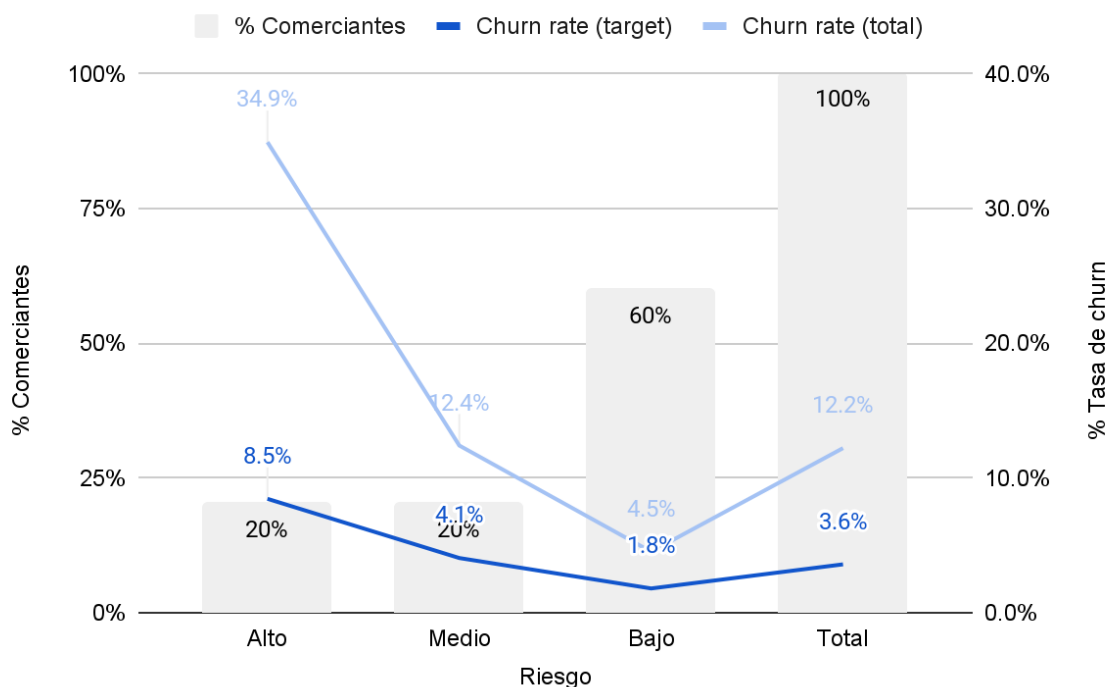
- Incentivar un mayor uso del producto, a través del cual lograr incrementar el *Share of Wallet*, principal predictor de *churn*. Complementar con información cualitativa (a través de encuestas y entrevistas) en comerciantes con un *SoW* bajo/Riesgo de *churn* alto para entender los principales problemas que poseen.
- Priorizar las mejoras del sistema antifraude del producto, ya que una aceleración en el crecimiento de los indicadores de riesgo tiene un impacto rápido en la tasa de *churn* de corto plazo.

Por otro lado, se comenzó a utilizar las predicciones del modelo para identificar aquellos comerciantes que deberían incluirse en campañas de retención. Para ello, se dividió de forma mensual la base de comerciantes en función a la probabilidad de *churn* en 3 grupos:

- Riesgo alto: comerciantes que se ubicaron en los primeros dos percentiles del *score*.
- Riesgo medio: comerciantes que se ubicaron en el 3° y 4° percentiles del *score*.
- Riesgo bajo: concentra al resto de los comerciantes (60% de la población inicial).

En el gráfico a continuación se puede observar la tasa de *churn* para cada segmento de riesgo para la base que se encontraba activa en el mes de octubre. En el grupo de alto riesgo, el porcentaje de comerciantes que realiza *churn* a los 3 meses es del 8.5%, una tasa 2 veces mayor que la total. A su vez, si incluimos en la métrica a los comerciantes que realizaron *churn* en menos de 3 meses, la tasa es del 34.9%, casi 3 veces mayor que la general.

Gráfico 4.1: Segmentación de riesgo



Se realizaron tres tipos de campañas distintas en función al rango de facturación de los comerciantes:

1. Los comerciantes pertenecientes a *tiers* 1 a 4 ya poseen un agente de cuenta como su punto de contacto y reciben atención personalizada con el objetivo de prevenir el *churn*. Para la selección de a quién contactar, previamente consideraban otras variables, pero comenzaron a incluir en el análisis de cada perfil el riesgo de *churn* del modelo (de forma mensual).
2. Para los *tiers* de 5 y 6, se les envió una comunicación que ofrecía un método de contacto personalizado para recibir su opinión sobre el producto y, luego del contacto, se realizaba una aplicación de un descuento en las comisiones cobradas por 3 meses. En este caso, la acción estuvo orientada al grupo de alto riesgo.
3. Finalmente, para *tiers* 7 a 10 se les envió una comunicación dentro de la plataforma y por mail destacando beneficios sobre el producto. La campaña se enfocó en el grupo de alto y medio riesgo.

En el primer caso, se decidió no medir la performance de la campaña en sí, dado el tamaño de la muestra acotado y el riesgo que implicaba tener una muestra control muy grande. Pero para las otras dos iniciativas, se analizó el impacto a través de un experimento aleatorio controlado (A/B test).

En líneas generales, este tipo de experimentación permite identificar si efectivamente las acciones o cambios realizados tuvieron algún tipo de efecto en la métrica de interés. En un principio, se plantea una hipótesis a ser testada que, en este caso, sería que las acciones de retención logren reducir la tasa de *churn*.

En una siguiente instancia, se diseña el experimento dividiendo a la población objetivo en dos grupos principales: tratamiento y control. El primer grupo se encuentra sujeto a las acciones de retención mientras que al segundo se lo excluye de la iniciativa. El objetivo permite inferir el impacto de las acciones, a través de un test de hipótesis para entender si la diferencia en la tasa de *churn* entre los dos grupos es significativa o no. Estos resultados dependen de un supuesto esencial, y es que la probabilidad de los sujetos de ser expuestos a cada versión del experimento tiene que ser la misma. En el caso de los experimentos controlados, esto se logra asignando los grupos de forma aleatoria.

A continuación, se realiza un breve resumen de las iniciativas medidas a través de esta metodología.

Resultados iniciativas del segmento de facturación media

El tamaño de la muestra fue de 447 casos, comerciantes con alto riesgo de churn pertenecientes a los tiers 5 y 6. De forma aleatoria, se dividieron en dos grupos de igual tamaño.

Dado que el tamaño de la muestra es acotado, no es posible aplicar una prueba de hipótesis utilizando tests paramétricos para evaluar la significatividad de la diferencia entre los grupos en nuestra variable de interés. Los tests paramétricos, como la prueba t de Student, asumen que los datos siguen una distribución específica. Con un tamaño de muestra grande, el teorema del límite central a menudo justifica esta suposición incluso cuando la población no está normalmente distribuida. Sin embargo, en un tamaño de muestra pequeño esta suposición puede no ser válida.

En este caso se optó por evaluar los resultados a través de tests no paramétricos. Estos no hacen suposiciones sobre la distribución de los datos por lo que arrojan resultados más robustos ante muestras pequeñas (Corder et.al., 2011). Aplicando la prueba de chi-cuadrado, el mínimo efecto en retención con el que se podría rechazar la hipótesis nula¹⁹ es de 3 puntos porcentuales con un nivel de significatividad del 15%.

¹⁹ La prueba de Chi-cuadrado plantea como hipótesis nula que no existe una relación estadística entre dos categorías.

Sin embargo, no se observaron diferencias significativas entre los grupos en el ratio de *churn*. El grupo tratamiento tuvo una tasa de *churn* a 3 meses del 1.4%, mientras que el grupo control fue de 1.8% (0.04 puntos porcentuales de diferencia).

En parte, esto se encuentra explicado por el bajo alcance que tuvo la iniciativa. Las acciones dependían no solo de que el comerciante abriera la comunicación sino que también se comunicará con agentes del equipo de atención al cliente para recibir acompañamiento y un descuento en caso de descontento. Tan sólo 5 comerciantes llegaron a esa instancia (el 2% del grupo tratamiento).

Tabla 4.2: Resultados - Segmento de facturación media

	# Comerciantes	Churn rate (target)
Control	221	1.8%
Tratamiento	226	1.4%

Resultados iniciativas del segmento de facturación baja

El grupo que recibió la comunicación fue de 1.281 comerciantes, mientras que el grupo control fue de 1.338. En este caso, dado que la muestra era mayor, se utilizaron test paramétricos (t-test) para definir el mínimo efecto detectable. Considerando una tasa de *churn overall* del 3.2%, el mínimo efecto detectable es de 1.3 puntos porcentuales con un nivel de significatividad del 5%.

Sin embargo, no se observan diferencias significativas en la tasa de retención entre los grupos. El grupo que recibió la comunicación tuvo una tasa de *churn* del 3.1%, mientras que en el grupo control fue de 3.4%. De todas formas, es interesante destacar que aquellos comerciantes que abrieron la comunicación la tasa fue del 2.4%. Sin embargo, no es posible asociar este efecto a la iniciativa de retención en sí, ya que probablemente es un subgrupo con mayor propensión a abrir comunicaciones del producto y, por ende, más leal al mismo.

Tabla 4.3: Resultados - Segmento de facturación baja

	# Comerciantes	Churn rate (target)
Control	1,338	3.4%
Tratamiento	1,281	3.1%
Tratamiento que abrió la comunicación	494	2.4%

5. Conclusiones

El presente trabajo planteó la problemática de retención de clientes para una plataforma de comercio electrónico en un contexto B2B. En este marco, se decidió realizar un análisis sobre las principales variables que influyen en la tasa de *churn* y dejar en funcionamiento un modelo que permita de forma mensual predecir la probabilidad de *churn* de cada cliente.

En cuanto al análisis exploratorio realizado, se lograron identificar las principales características que diferencian a los comerciantes próximos a realizar *churn*. En primer lugar, se destaca que aquellos comerciantes con menor *Share of Wallet*, presentan un mayor riesgo de abandonar el servicio. Por último, las variables asociadas a fraude son potenciales factores que influyen el *churn* de corto plazo.

En función a estos hallazgos, se recomendó al equipo de negocio realizar acciones preventivas para incentivar que los comerciantes utilicen en mayor proporción la plataforma de pagos, impulsando una mejora en la métrica de *Share of Wallet*. A su vez, se transmitió la importancia de las mejoras en el sistema antifraude de la plataforma ya que es determinante en el *churn* y no puede ser compensado por acciones comerciales.

Con el objetivo de encontrar el mejor modelo predictivo se generó un esquema para la evaluación comparativa de distintos modelos considerando las necesidades de este caso de uso. Una muestra desbalanceada por la cual se empleó una métrica como AUC para el comparativo y teniendo en cuenta la mayor importancia de tiendas con mayor facturación donde se monitorea la performance en ese grupo específico.

Con este esquema, se encontró que el *Gradient Boosting* entrenado con las variables numéricas es el algoritmo que posee la mejor performance en el dataset de validación. De todas formas, la misma sólo es levemente superior al modelo utilizado como baseline (Regresión Logística con solo dos variables). Este hallazgo está en línea con los *findings* de Tamaddoni et al. (2015), donde para casos de muestras pequeñas la regresión logística posee resultados mejores que otros métodos de aprendizaje automático.

Si bien la complejidad del primero es mayor, tanto en tiempo de entrenamiento como en la cantidad de variables que utiliza, este esfuerzo adicional se ve trasladado a una performance significativamente mayor en el set de testeo que considera los últimos dos meses de información. Desde el punto de vista de la estabilidad del modelo elegido, también observamos que en los dos meses siguientes la proyección de la probabilidad de *churn* se mantiene estable.

Finalmente, a través de las predicciones del modelo se evaluaron dos conjuntos de acciones de retención a través de un experimento aleatorio controlado, lo que permitió inferir el impacto de las mismas y guiar al equipo de negocio en los próximos pasos. Como las primeras iniciativas no arrojaron resultados, se continuará trabajando en redefinir el foco de las campañas de retención hasta conseguir observar un impacto significativo en nuestra variable de interés, la tasa de *churn*.

Existen ciertas problemáticas que no han llegado a ser abordadas en este trabajo y que podrían guiar futuras líneas de investigación que sirvan de base para retroalimentar el modelo y mejorar su performance. En principio, en lo referido a la definición del problema y la variable objetivo, sería relevante focalizar en aquellos comerciantes que sean más propensos a ser retenidos, factor que se podría tener en cuenta una vez que se comiencen a aplicar campañas de retención que tengan impacto en la tasa de *churn*. A su vez, si bien controlamos por cambios en la distribución de la variable predicha, no se generó un sistema de monitoreo de cambios de las variables explicativas incluidas en el modelo. Esto podría afectar su performance en un futuro ante cambios del producto que a veces se reflejan en los indicadores considerados. De todas formas, en la medida que la plataforma gane madurez y se consolide más en el mercado, estos factores de cambio que afectan la relación entre las variables deberían disminuir en el tiempo.

Este trabajo contribuye en la literatura de *churn* voluntario en el contexto B2B y procura orientar a las empresas, especialmente tecnológicas, que se enfrenten a este tipo de problemas. A través de un análisis detallado y un modelo predictivo de alta precisión, se ofrece una hoja de ruta para la identificación de factores de riesgo de *churn*, la implementación de estrategias de retención efectivas y la evaluación continua del rendimiento de tales iniciativas. Más allá de la identificación de las variables que impactan en el *churn*, este estudio proporciona evidencia empírica de la utilidad de los métodos de aprendizaje automático en el diseño y monitoreo de estrategias de retención.

6. Bibliografia

- Athanassopoulos, A. D. (2000). Customer satisfaction cues to support market segmentation and explain switching behavior. *Journal of business research*, 47(3), 191-207.
- Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7), 1145-1159.
- Burez, J., & Van den Poel, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 32(2), 277-288.
- Corder, G. W., & Foreman, D. I. (2011). Nonparametric statistics for non-statisticians.
- Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert systems with applications*, 34(1), 313-327.
- De Caigny, A., Coussement, K., Verbeke, W., Idbenjra, K., & Phan, M. (2021). Uplift modeling and its implications for B2B customer churn prediction: A segmentation-based modeling approach. *Industrial Marketing Management*, 99, 28-39.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings 1* (pp. 1-15). Springer Berlin Heidelberg.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861-874.
- Jamal, Z., & Bucklin, R. E. (2006). Improving the diagnosis and prediction of customer churn: A heterogeneous hazard modeling approach. *Journal of Interactive Marketing*, 20(3-4), 16-29.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.
- Lam, S. Y., Shankar, V., Erramilli, M. K., & Murthy, B. (2004). Customer value, satisfaction, loyalty, and switching costs: an illustration from a business-to-business service context. *Journal of the academy of marketing science*, 32(3), 293-311.
- Gattermann-Itschert, T., & Thonemann, U. W. (2022). Proactive customer retention management in a non-contractual B2B setting based on churn prediction with random forests. *Industrial Marketing Management*, 107, 134-147.

- Gordini, N., & Veglio, V. (2017). Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. *Industrial Marketing Management*, 62, 100-107.
- Rust, R. T., & Zahorik, A. J. (1993). Customer satisfaction, customer retention, and market share. *Journal of retailing*, 69(2), 193-215.
- Tamaddoni, A., Stakhovych, S., & Ewing, M. (2014). Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*, 43(7), 1258-1268.
- Tamaddoni, A., Stakhovych, S., & Ewing, M. (2015). Comparing churn prediction techniques and assessing their performance: A contingent perspective. *Journal of service research*, 19(2), 123-141.
- Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European journal of operational research*, 218(1), 211-229.
- Yurdakul, B. (2018). *Statistical properties of population stability index*. Western Michigan University.

7. Anexos

7.1. Sistema de pagos en Brasil

En el ecosistema de pagos en Brasil, existen diferentes medios de pago electrónico que un consumidor puede seleccionar al realizar una compra en una tienda virtual. Estos incluyen el boleto, el pix y la tarjeta de crédito, cada uno con su propio sistema de aprobación en el sistema bancario.

- El boleto es una opción popular que permite a los consumidores generar un código de barras para pagar en efectivo en bancos, casas de lotería o tiendas autorizadas. Es ampliamente utilizado, especialmente por aquellos que no tienen acceso a una cuenta bancaria o no desean utilizar tarjetas de crédito. Sin embargo, su proceso de pago puede ser más lento, ya que implica la impresión o generación del boleto físico y su posterior pago en una institución autorizada.
- El pix, por otro lado, es un sistema de pagos instantáneos desarrollado por el Banco Central de Brasil. Permite realizar transferencias y pagos electrónicos de forma rápida y segura las 24 horas del día, los 7 días de la semana. Utiliza identificadores únicos, como el CPF (Registro de Identificación Fiscal) o el número de teléfono, para facilitar las transacciones entre personas y empresas. El pix se ha vuelto muy popular debido a su conveniencia y velocidad, y ha ganado terreno como una opción de pago en línea.
- La tarjeta de crédito, por su parte, es ampliamente utilizada en Brasil y permite a los consumidores realizar compras en línea y en establecimientos físicos con una línea de crédito otorgada por el emisor de la tarjeta. Es un medio de pago ampliamente aceptado y ofrece comodidad y flexibilidad, ya que los pagos se pueden realizar en cuotas y se beneficia de programas de recompensas y protección al consumidor. Sin embargo, requiere una relación previa con una entidad financiera y está sujeta a límites de crédito y tasas de interés.

En este ecosistema, la transacción representa el movimiento de dinero cuando se intercambia por un producto o servicio. El proceso de pago implica un intenso flujo de intercambio de información entre varias partes, como pasarelas de pago, subadquirentes y/o adquirentes, marcas de tarjetas y bancos emisores.

En términos generales, el proceso de pago involucra a un subadquirente que inicialmente procesa el pago y transmite los datos al adquirente. El adquirente es responsable de procesar el pago y a través de su red de socios acreditados permite que una tienda ofrezca diferentes condiciones de pago a sus clientes. El adquirente recibe la información del pago, la procesa y la transmite a la marca de la tarjeta (en el caso de pagos con tarjeta de crédito)

y al banco emisor. Si todos los actores aprueban la transacción, el adquirente se encarga de transferir los fondos cobrados al cliente por el banco emisor a la cuenta de la tienda.

Las marcas de tarjetas (o banderas) son empresas que establecen las reglas de negocio para las compras realizadas con tarjeta de crédito. Cada marca de tarjeta define los patrones que los adquirentes deben seguir al procesar las transacciones realizadas con esa tarjeta. Mantienen una estrecha relación con los adquirentes para garantizar la aceptación de sus tarjetas en los establecimientos, promoviendo su uso y ampliando la base de usuarios. En el flujo de pago con tarjeta de crédito, la marca de la tarjeta recibe la información de compra de la tienda a través del adquirente, valida las condiciones (como el número de cuotas) y envía los datos al banco emisor para verificar si el comprador tiene límite disponible.

El banco emisor es la institución financiera responsable de los medios de pago, ya sean boletos bancarios, tarjetas de débito o crédito. Al recibir los datos de una compra, el banco emisor autoriza o rechaza la transacción según el límite de crédito o el saldo disponible en la cuenta del usuario. Si es aprobada, el banco efectúa el cobro del valor correspondiente. Es importante tener en cuenta este concepto, ya que permite configurar condiciones especiales para pagos realizados a través de bancos específicos. Además, en el caso de los boletos registrados, cada banco tiene su propia forma de emisión.

7.2. Ingeniería de atributos

Variable *target*

Para generar la variable a predecir, en un principio se construyó a nivel tienda y fecha de la orden, la cantidad de días que transcurrieron hasta la siguiente orden (en caso de que no exista siguiente orden se asignó un número superior a 30). Luego, se filtraron todos los casos con más de 30 días de diferencia, quedando todas las tiendas con su fecha de última transacción asociada. Finalmente, se transformaron los datos para obtener para cada tienda y mes si había realizado su última transacción.

El mismo ejercicio se realizó considerando sólo las órdenes procesadas por la plataforma de pagos interna. Los meses en que coincidía la última transacción del comerciante en las dos formas de cálculo, se consideró que no era un *churn* específico de la integración de pagos, sino que en verdad había dejado de tener ventas en su tienda online. Los meses en que se observa una última transacción de una tienda con la plataforma de pagos interna

pero no hay una última transacción de la tienda en general, se asigna un 1 a la variable target, considerando que ese comerciante realizó churn.

Variables sobre el volumen de negocio

Las variables dentro de este grupo se construyeron con los datos de la primera fuente de información, la cual contenía todas las órdenes pagas que se habían generado en cada tienda. Se agruparon estos datos por tienda y mes para obtener la facturación, la cantidad de transacciones y el ticket promedio.

En función al promedio de la facturación de los últimos 3 meses, se calcula el *tier*, variable numérica entera que toma valores del 1 al 10. La asignación del valor se hace en función a rangos fijos de facturación definidos por el negocio. Valores más bajos de *tier*, concentran a una menor cantidad de comerciantes pero con mayor nivel de facturación. Esta variable es de suma importancia en la empresa, ya que es cómo se clasifica a todos los comerciantes y se alinean tanto condiciones del producto como del servicio en función a ello.

A través del *payment_method*, se logra diferenciar las variables sólo a través de tarjeta de crédito, el cual resulta el método de pago más rentable para la empresa pero, a su vez, el más costoso para el comerciante.

El *Share of Wallet (SoW)*, es una métrica que captura qué porcentaje de la facturación del comerciante se procesa a través de la plataforma de pagos interna. Por ende, se calcula como el ratio entre Facturación donde el *payment_gateway* es la empresa/Facturación total del comerciante.

Variables sobre periodicidad

En este caso, las variables también se calcularon en función al primer dataset mencionado, agrupando la información por tienda y mes. La forma de cálculo de la periodicidad de ventas es:

$$\text{Periodicidad de transacciones} = \frac{\text{Diferencia de días entre la última fecha de venta y la primera (en el mismo mes)}}{\text{Cantidades de transacciones en el mes} - 1} \quad (5)$$

Si el comerciante solo tuvo una venta en todo el mes, se imputa por default el valor 32.

Se generó esta variable para todas las órdenes en general y también en específico para aquellas que utilizaron la plataforma de pagos interna. Luego se procedió a realizar el ratio entre ellas, para entender si la periodicidad en la plataforma de pagos era mayor o menor a la de la tienda en general.

VARIABLES SOBRE CARACTERÍSTICAS DEL CLIENTE Y RELACIÓN CON LA EMPRESA

En este caso, se utilizó el tercer dataset para obtener una variable categórica que identifica la categoría de los productos que vende la tienda, y en función a la fecha de creación de la tienda, se obtuvo la antigüedad de la tienda virtual a comienzos de cada mes. Para determinar la antigüedad del comerciante utilizando la plataforma de pagos interna, se tomó la fecha de la primera transacción que fue procesada por este medio, obtenida del dataset 2.

En cuanto a las variables sobre la contactabilidad del comerciante con el equipo de atención al cliente, del dataset 3 se filtraron todos los tickets donde la variable tópico contenía palabras claves asociadas con la plataforma de pagos interna. A su vez, se extrajo del campo json, la cantidad de contactos que tenían una calificación de la calidad de atención negativa, asignando un 1 en ese caso y 0 caso contrario. Luego se procedió a agrupar los datos a nivel de tienda y mes para realizar la unión con el resto de las variables.

VARIABLES SOBRE RIESGO Y FRAUDE

Como se mencionó anteriormente, en el dataset que contiene las transacciones que se procesan en la plataforma de pagos, existe el campo de estado que refleja si la transacción fue exitosa, y en caso contrario le asigna el motivo. En función a esta información, se construyeron dos variables principales que el negocio también suele monitorear en sus indicadores de riesgo:

- % con rechazo: refleja el monto de las transacciones con estado *failed*, sobre el total de la facturación del comerciante.
- % con *chargeback*: refleja el monto de las transacciones con estado *chargeback*, sobre el total de la facturación del comerciante. En este caso, la fecha considerada no es la de la creación de la transacción sino la fecha de la actualización de la misma (fecha en que la transacción pasa de estado *paid* a *chargeback*).

7.3. Análisis exploratorio - gráficos adicionales

Tabla 7.1: Distribución de las variables de riesgo

Variable	Mínimo valor	Q25	Q50	Q75	Máximo valor	% de outliers
% de la facturación con chargeback	0.0	0.0	0.0	0.0	1.0	6.3%

% de la facturación por tarjeta rechazado

0.0 0.0 0.0 0.3 1.0 6.0%

Gráfico 7.1: Riesgo y Fraude - Histogramas

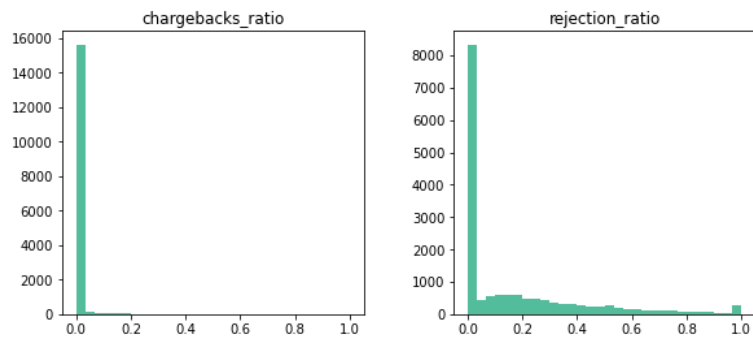


Tabla 7.2: Indicadores estadísticos de la variación mensual de SoW

	SoW (MoM)	
	Target	No churners
Mean	0.58	0.34
SD	1.80	1.36
T-test Statistic	2.60	
P Value	0.010 ***	

Gráfico 7.2: Distribución de SoW MoM

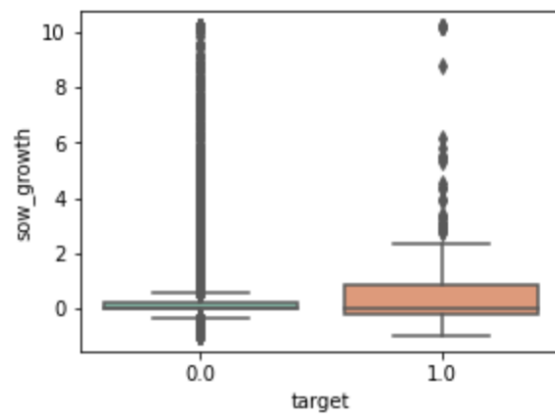


Gráfico 7.3: Promedio de la variación mensual del SoW

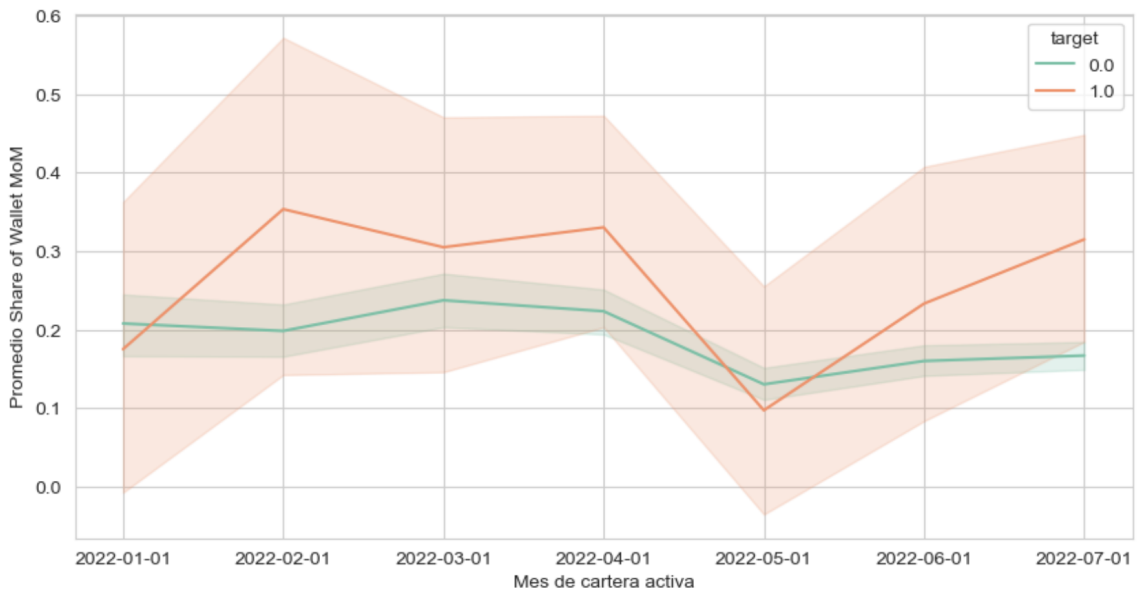


Gráfico 7.4: Promedio del ratio de periodicidad de transacciones

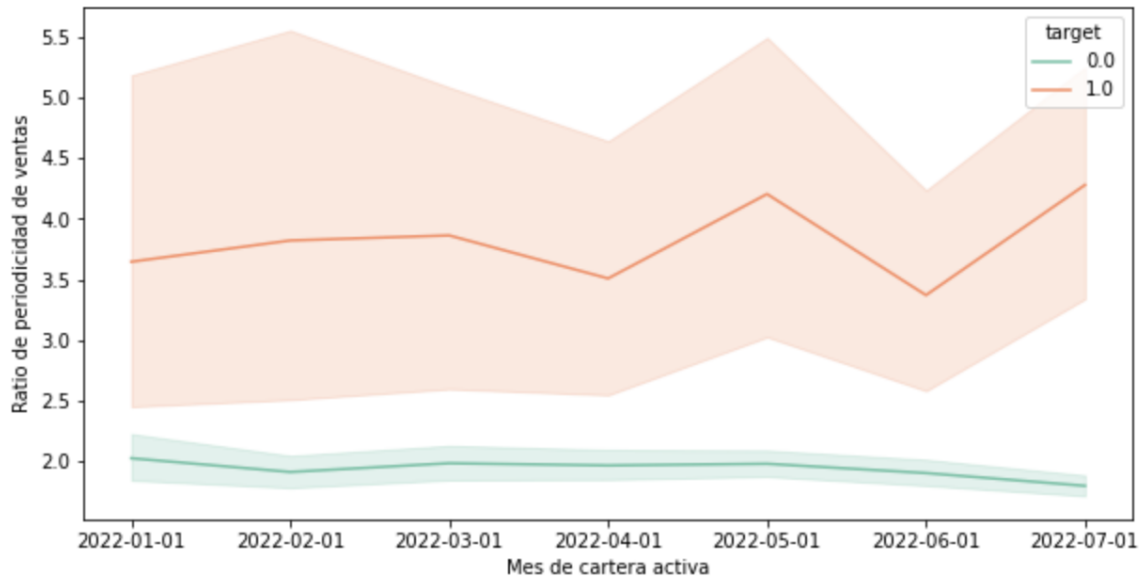


Tabla 7.3: Contactos con Support

	Contacts with Support		Contacts with Support (dummy)	
	Target	No churners	Target	No churners
Mean	0.3	0.5	0.15	0.21
SD	1.2	1.5	0.36	0.41
T-test Statistic		-1.75		-3.69
P Value		0.080 **		0.000 ***

Gráfico 7.5: Promedio del ratio de contacto

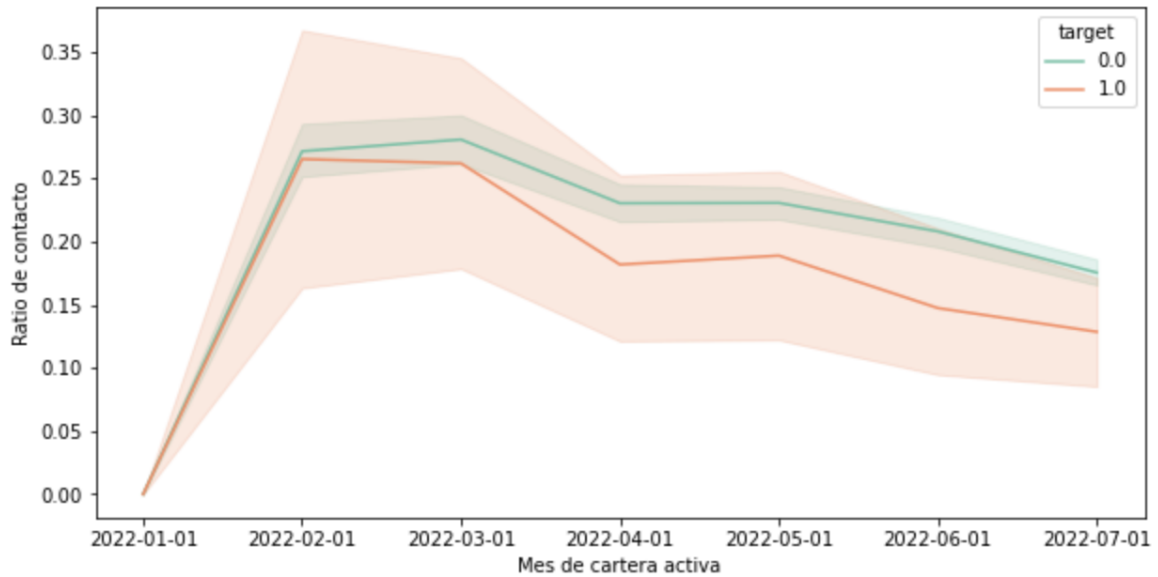


Gráfico 7.6: Tasa de churn por segmento

