

Prevención de Evasión (Churn) en Servicios de Telecomunicaciones

Factores determinantes de abandono y como prevenirlos.

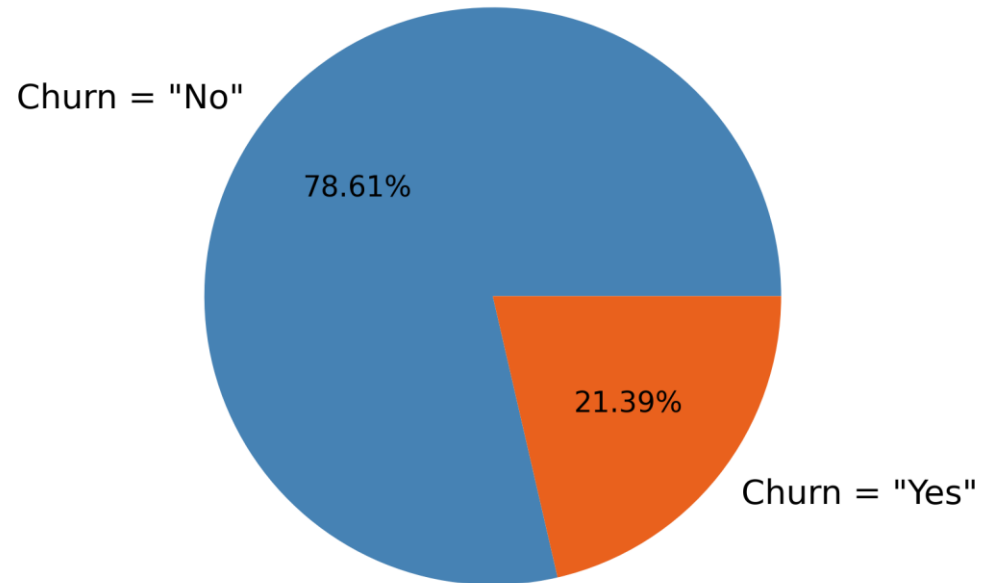
Objetivos:



- Desarrollar modelos de aprendizaje automático para predecir la probabilidad de abandono del cliente.
- Identificar los factores más relevantes asociados a la cancelación del servicio y los que contribuyen a la retención, contrastando los resultados obtenidos con las hipótesis exploratorias planteadas en etapas anteriores del análisis.
- Diseñar estrategias de retención basadas en los hallazgos obtenidos, orientadas a decisiones de negocio

Contexto del problema:

Porporción de Monto Total (ChargesTotal)
Retención y Evasión



Monto total de evasión: \$ 3589987.55

La empresa **TelecomX** enfrenta una **alta tasa de evasión de clientes (Churn)**, especialmente durante las **primeras etapas del ciclo de vida** del cliente.

Además, se ha identificado un grupo de **clientes antiguos** que también han decidido cancelar sus servicios, lo que representa una **pérdida significativa de ingresos y lealtad**.

Actualmente, la **evasión asciende hasta \$ 3,589,987, lo cual representa más del 20% de los ingresos totales**, generando un **impacto directo en la rentabilidad**.

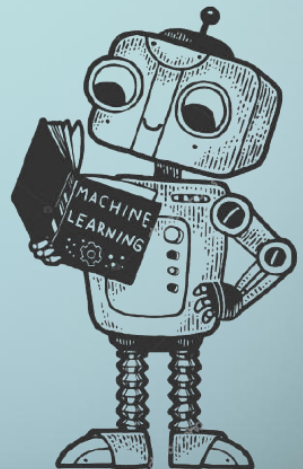
Este proyecto busca **comprender qué factores impulsan la cancelación del servicio** y desarrollar una herramienta predictiva que permita a la empresa **anticipar y prevenir la pérdida de clientes**.

Previsión de Evasión (Churn)



Modelo seleccionado:

- Se seleccionó como modelo final (*Champion Model*) una **Random Forest**, debido a su capacidad de generalización y desempeño sobre los datos de prueba. Cumpliendo con un **85% de evasiones detectadas**.
- El enfoque principal fue **maximizar la Sensibilidad (Recall)**, es decir, minimizar los falsos negativos (evasiones no detectadas), **sin sacrificar exceso de Precisión**, para que las campañas de retención sean efectivas y **eficientes en costos**.



Variación de métricas:

Variación de Recall

Model	Recall Train	Recall Test	Variación
Best Random Forest	0.8769	0.7273	-14.96%
Best Logistic Regression	0.6383	0.6229	-1.54%
Best XGBoost Classifier	0.7555	0.7138	-4.17%
Best Support Vector Machine	0.5996	0.5488	-5.08%

Variación de F1-score

Model	F1-score Train	F1-score Test	Variación
Best Random Forest	0.8673	0.6362	-23.11%
Best Logistic Regression	0.6301	0.5929	-3.72%
Best XGBoost Classifier	0.7591	0.619	-14.01%
Best Support Vector Machine	0.6126	0.5525	-6.01%

El modelo de **Regresión Logística** fue seleccionado por su **alto Recall sostenido en datos nuevos, a pesar de mostrar cierto sobreajuste a los datos de entrenamiento.**

- Mantiene un **Recall del 72.73%** en datos de prueba.
- Presenta el mejor balance general entre **Recall y Precisión, con un valor F1-score de 63.62%.**
- Esto sugiere un modelo **más robusto y confiable para detectar evasión (Churn).**

Umbral de decisión

Modificación de umbrales - Variación Recall

Model	Recall (Umbral=0.5)	Recall (Umbral modificado)	Variación
Best Random Forest	0.7273	0.862	13.47%
Best Logistic Regression	0.6229	0.862	23.91%
Best XGBoost Classifier	0.7138	0.8552	14.14%
Best Support Vector Machine	0.5488	0.862	31.32%

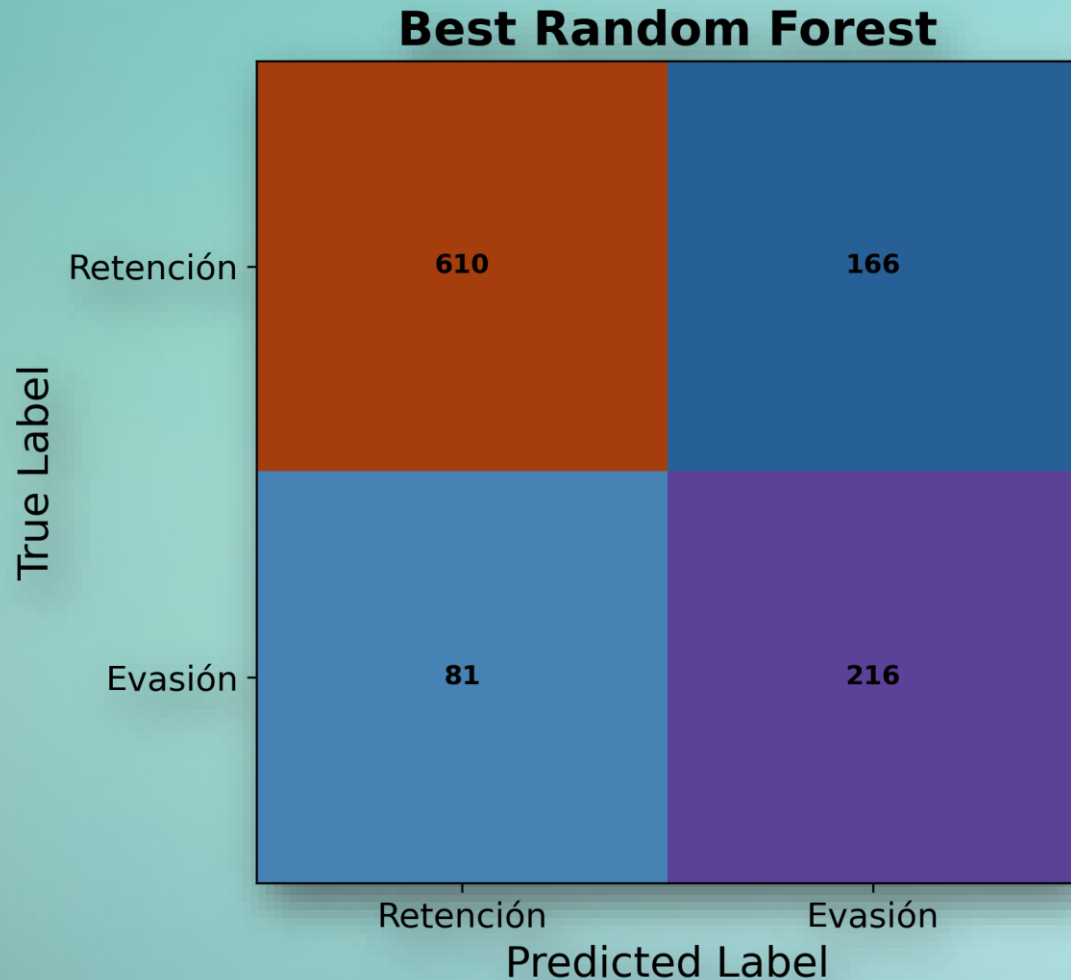
Modificación de umbrales - Variación Precision

Model	Precision (Umbral=0.5)	Precision (Umbral modificado)	Variación
Best Random Forest	0.5654	0.4914	-7.40%
Best Logistic Regression	0.5657	0.4539	-11.18%
Best XGBoost Classifier	0.5464	0.4695	-7.69%
Best Support Vector Machine	0.5563	0.4571	-9.92%

Para obtener una **Sensibilidad (Recall)** de al menos 85%, el modelo **Random Forest** es el que sacrifica la menor proporción de **Precisión**.

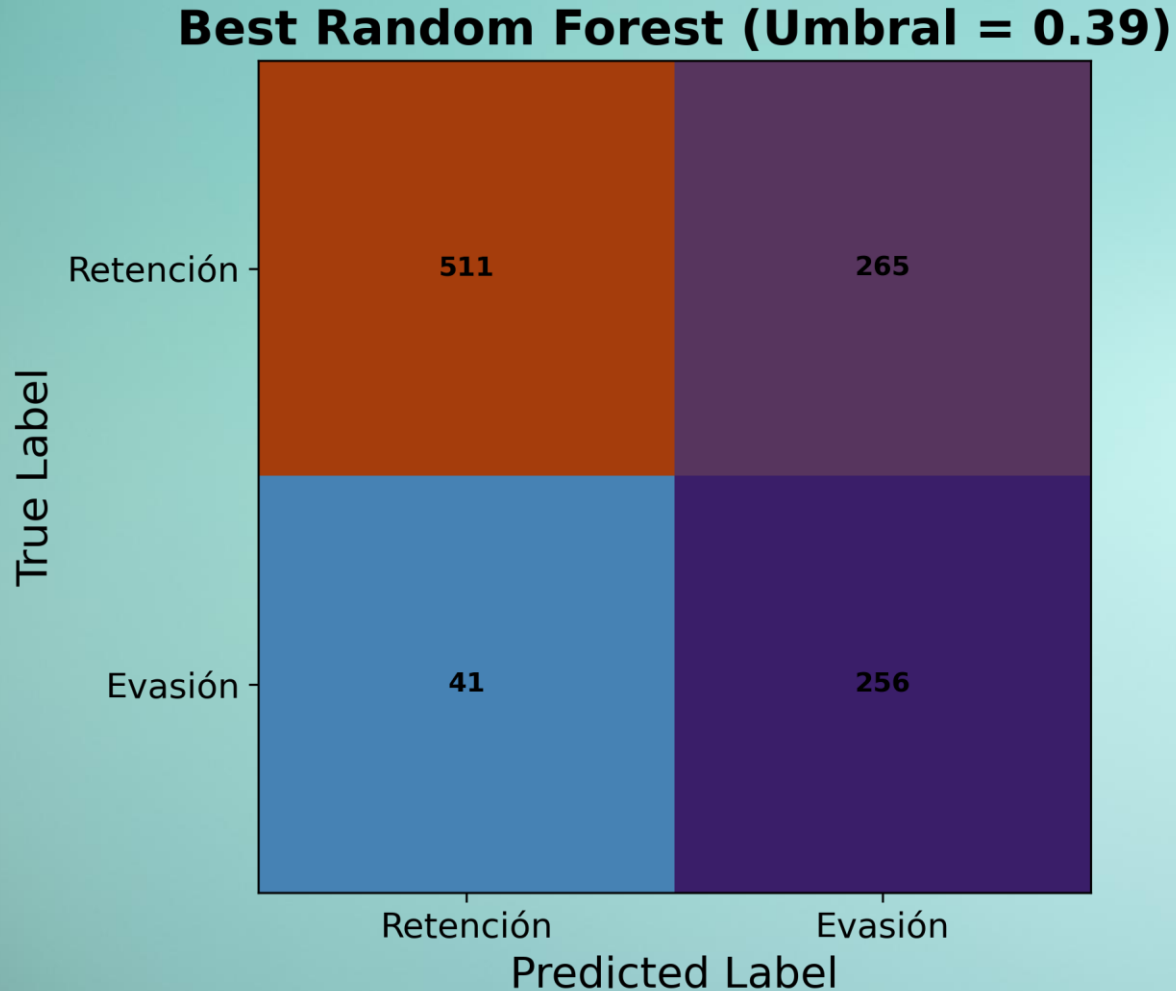
Esto permite identificar correctamente el 85% de casos de abandono, y a su vez clasificar correctamente el 50% de clientes que no planean cancelar sus servicios, ideal para campañas de retención eficientes.

Matriz de Confusión (datos de prueba):



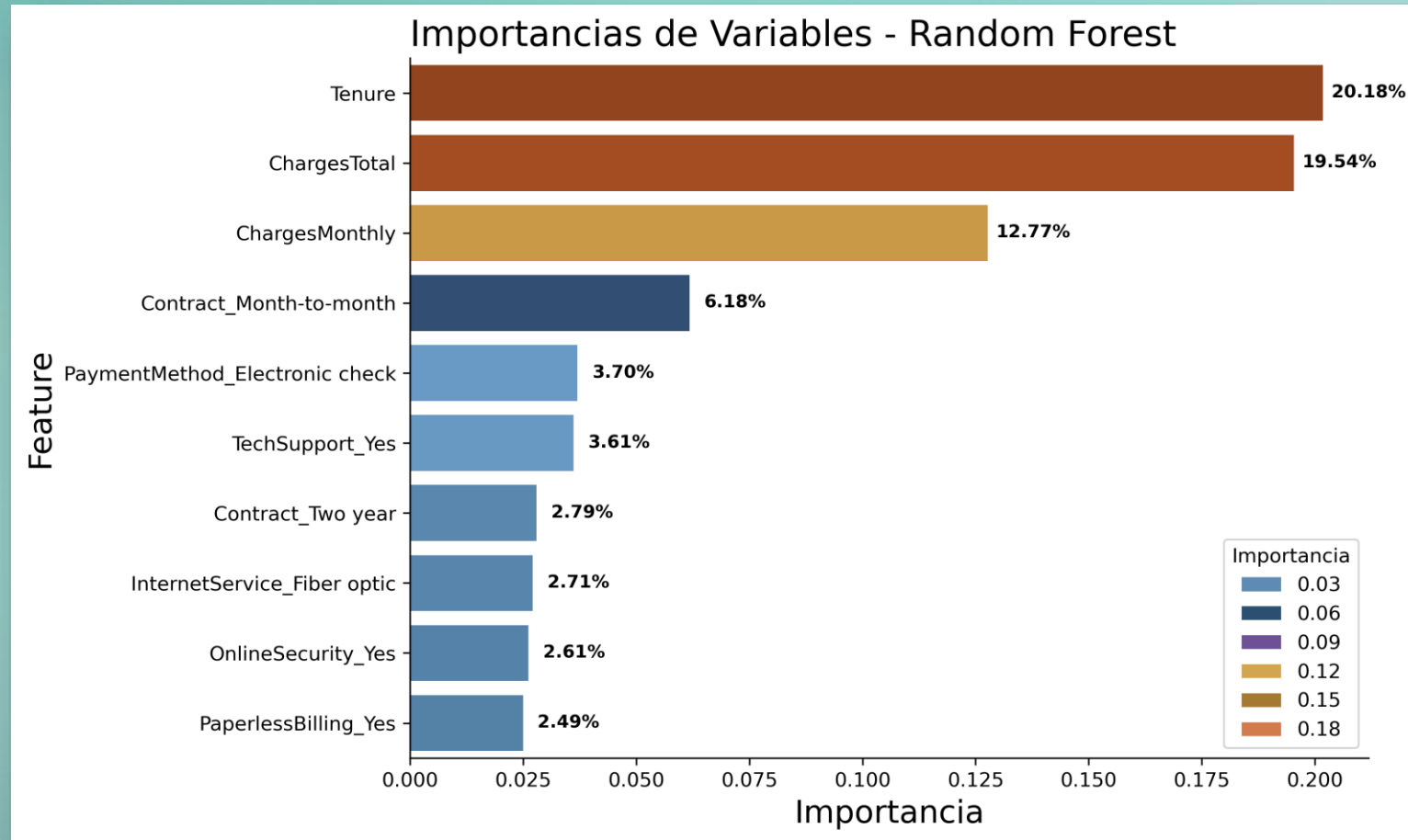
- Sin modificar el umbral de decisión, el modelo logró **identificar correctamente el 72% de los clientes que efectivamente abandonan el servicio.**
- Este resultado permite **anticiparse a 7 de cada 10 abandonos reales** con un margen aceptable de falsos positivos.
- Y a su vez, identificar correctamente el **78% de clientes que han cancelado sus servicios.**

Matriz de Confusión (umbral modificado):



- Al modificar el umbral de decisión del modelo para aumentar el porcentaje de abandonos identificados, para conseguir un **85% de aciertos**, el modelo identifica correctamente el **65% de clientes que no han cancelado sus servicios**.

Factores influyentes en la Evasión:



A partir del modelo de **Random Forest** entrenado, se identificaron las **10 variables más influyentes** en la probabilidad de que un cliente cancele el servicio.

Una de las desventajas de este modelo es que no permite evaluar cuales son las variables de riesgo y cuales son las variables protectoras frente a la Evasión (Churn), solo discernir cuales fueron las características utilizadas para clasificar los clientes.

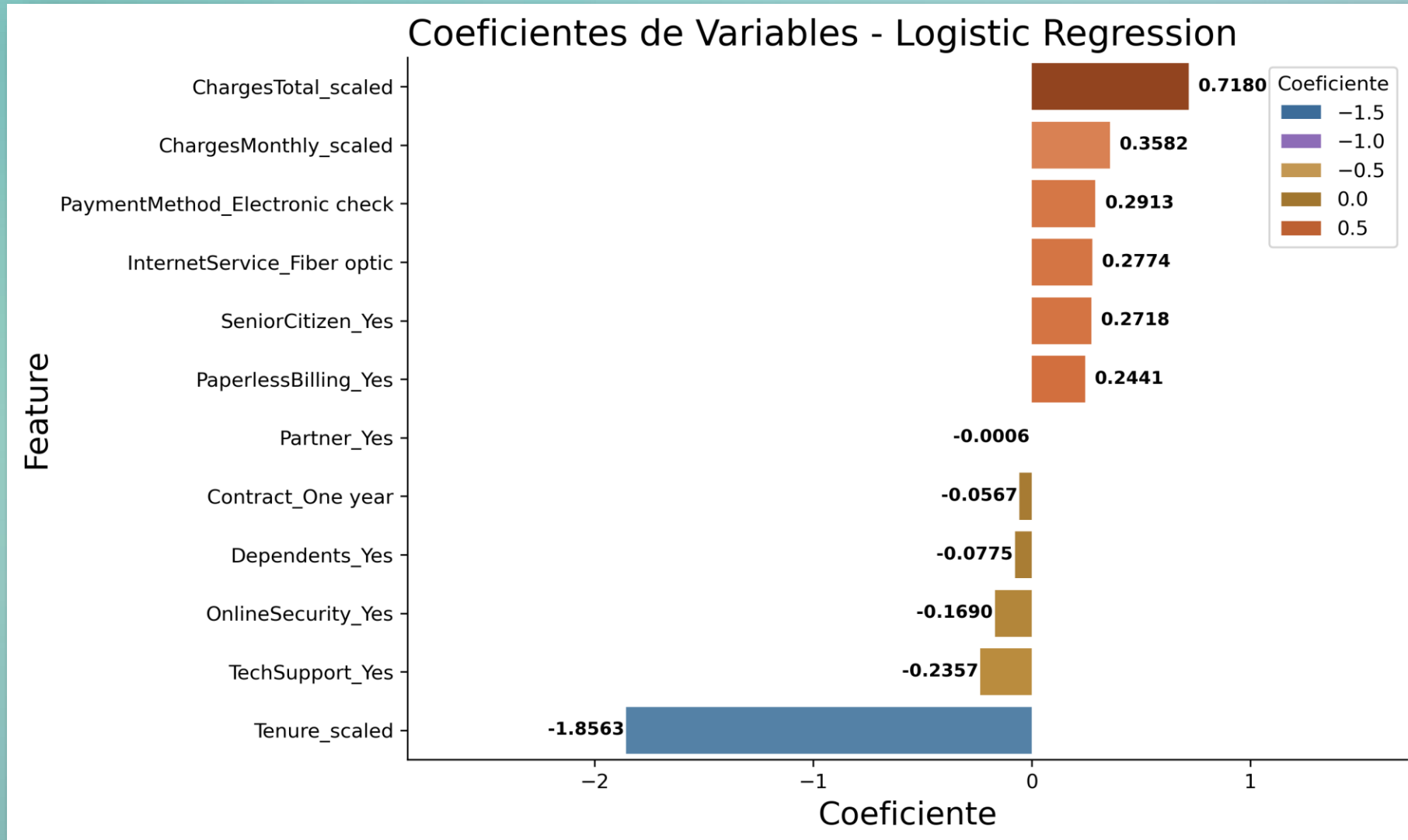
Análisis de apoyo: Regresión Logística

Dado que para diseñar estrategias de retención efectivas, resulta imprescindible entender cómo y cuánto afecta cada variable en las probabilidades de evasión.

Debido a que las importancias del modelo **Random Forest** no permite evaluar la dirección o magnitud del impacto, se complementará el análisis contrastado con los coeficientes determinados por la **Regresión Logística** seleccionada como representante de dicha familia.

Esto permitirá entender los focos de riesgo y ejes protectores contra la Evasión de clientes

Coeficientes Regresión Logística



Coeficientes Regresión Logística

Variables de Impacto en la Evasión (Churn)

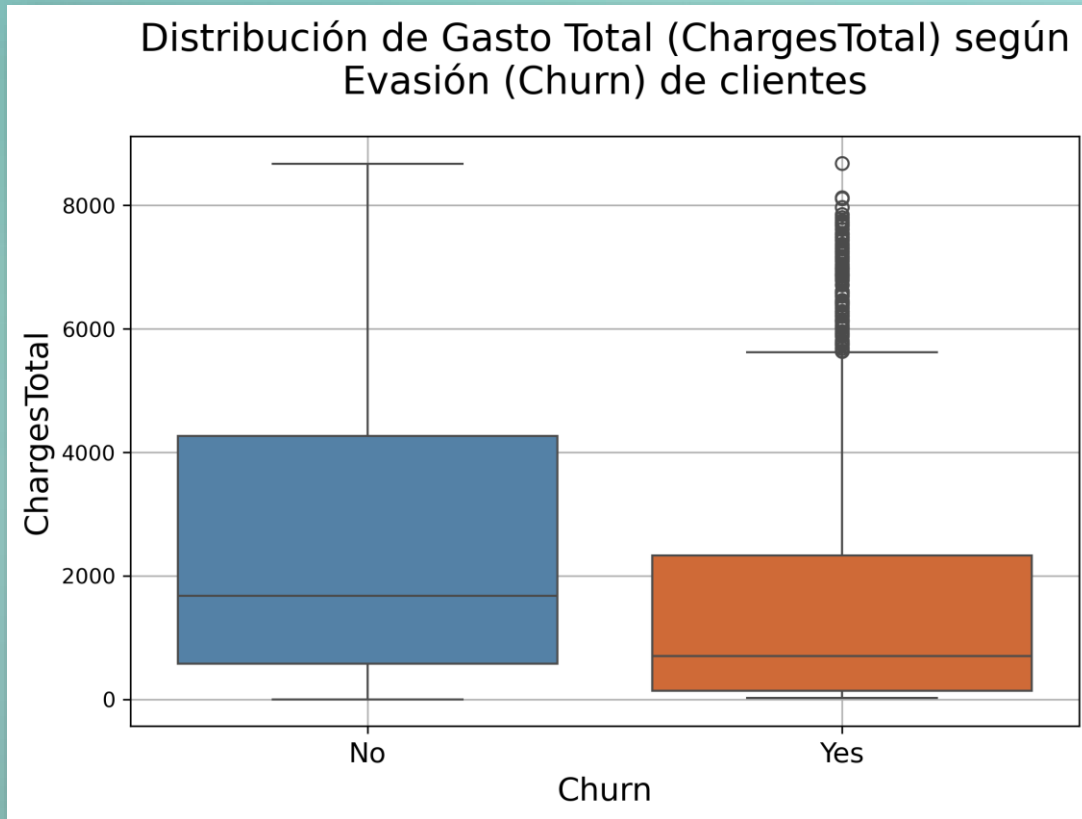
Feature	Coeficiente	Odds Ratio	Interpretación
ChargesTotal_scaled	0.718	2.0503	Aumenta las chances en: 105.03%
ChargesMonthly_scaled	0.3582	1.4307	Aumenta las chances en: 43.07%
PaymentMethod_Electronic check	0.2913	1.3382	Aumenta las chances en: 33.82%
InternetService_Fiber optic	0.2774	1.3197	Aumenta las chances en: 31.97%
SeniorCitizen_Yes	0.2718	1.3124	Aumenta las chances en: 31.24%
PaperlessBilling_Yes	0.2441	1.2765	Aumenta las chances en: 27.65%
Partner_Yes	-0.0006	0.9994	Disminuye las chances en: 0.06%
Contract_One year	-0.0567	0.9449	Disminuye las chances en: 5.51%
Dependents_Yes	-0.0775	0.9254	Disminuye las chances en: 7.46%
OnlineSecurity_Yes	-0.169	0.8445	Disminuye las chances en: 15.55%
TechSupport_Yes	-0.2357	0.79	Disminuye las chances en: 21.0%
Tenure_scaled	-1.8563	0.1563	Disminuye las chances en: 84.37%

Factores de riesgo



Gasto Total (ChargesTotal):

ChargesTotal_scaled	0.718	2.0503	Aumenta las chances en: 105.03%
---------------------	-------	--------	---------------------------------



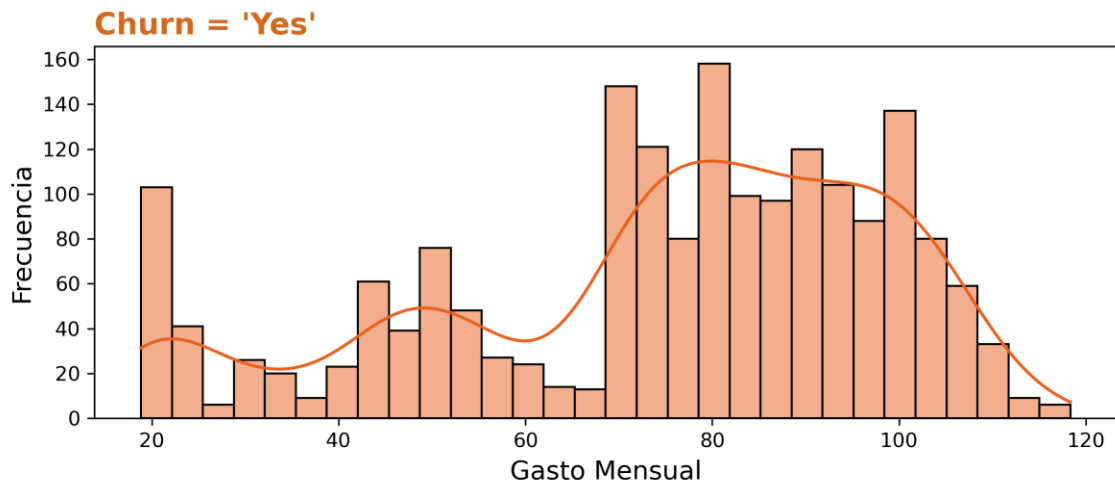
- **Gasto Total** surge como el factor de riesgo más importante. Un aumento de aproximadamente \$3000 en esta variable (equivalente a 1 unidad robusta) se asocia con un incremento de hasta **105.03%** en las probabilidades relativas de Evasión.
- Curiosamente, el Gasto Total suele crecer con la antigüedad del cliente, lo cual revela una anomalía: algunos clientes antiguos y valiosos deciden abandonar la empresa. Este segmento debe ser analizado con mayor profundidad para comprender los motivos detrás de su decisión y prevenir futuras pérdidas.

Datos originales (antes del modelado)

Gasto Mensual (ChargesMonthly):

ChargesMonthly_scaled	0.3582	1.4307	Aumenta las chances en: 43.07%
-----------------------	--------	--------	--------------------------------

Distribución de Gasto Mensual para clientes que abandonan



- El **Gasto Mensual** es identificado como la segunda variable que presenta mayor riesgo de evasión según la **Regresión Logística**, y **Random Forest** la posiciona como la segunda variable con mayor importancia para clasificar a los clientes.
- Ya que esta variable fue escalada, el impacto de esta se mide en unidades robustas (\$54.45 para esta variable), **lo que significa que cada \$54.45 de gasto mensual, las probabilidades relativas de abandono aumentan un 43.07%.**

Datos originales (antes del modelado)

Facturas Digitales:

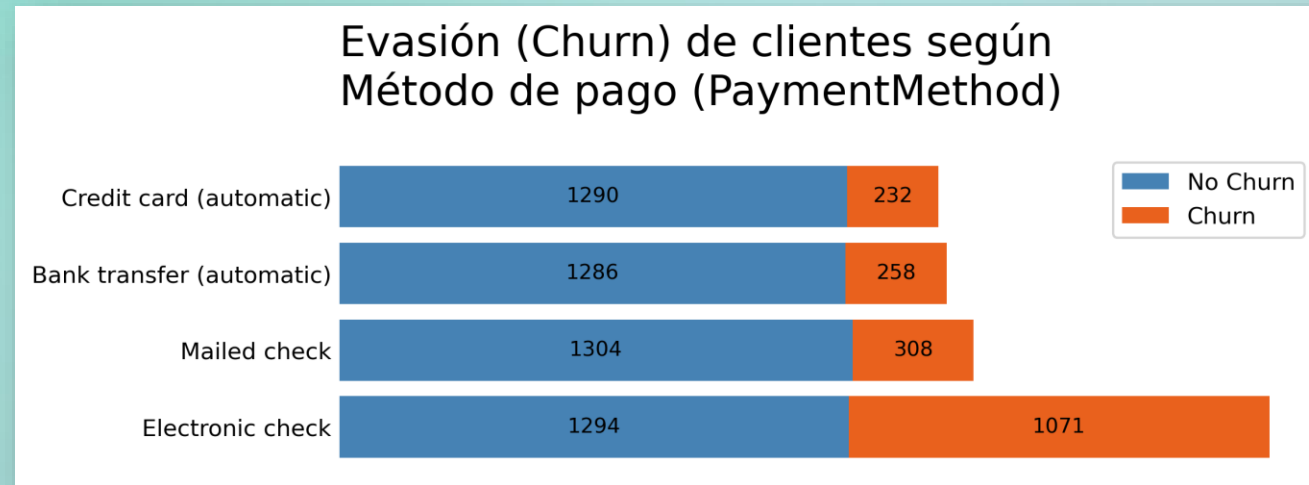
Facturación Digital (Paperless Billing) y

Método de pago: Factura electrónica (Payment Method: Electronic Check)

PaymentMethod_Electronic check	0.2913	1.3382	Aumenta las chances en: 33.82%
PaperlessBilling_Yes	0.2441	1.2765	Aumenta las chances en: 27.65%

- El **impacto negativo** de estas dos variables sobre el abandono, es un fuerte indicador de que **existen problemas relacionados al servicio de facturación digital**.
- El data set no contiene información que permita entender cuáles son las insatisfacciones respecto a este tipo de servicio. Pero debido a su influencia en el abandono, es pertinente que recolectar datos para evaluar y mejorar este área de negocios.

Lo anterior, confirma la hipótesis planteada en la etapa previa del proyecto, donde se detectó que el método de pago **Factura Electrónica (*Electronic Check*)** contaba con la **mayor proporción de Evasión** en relación a los métodos restantes.



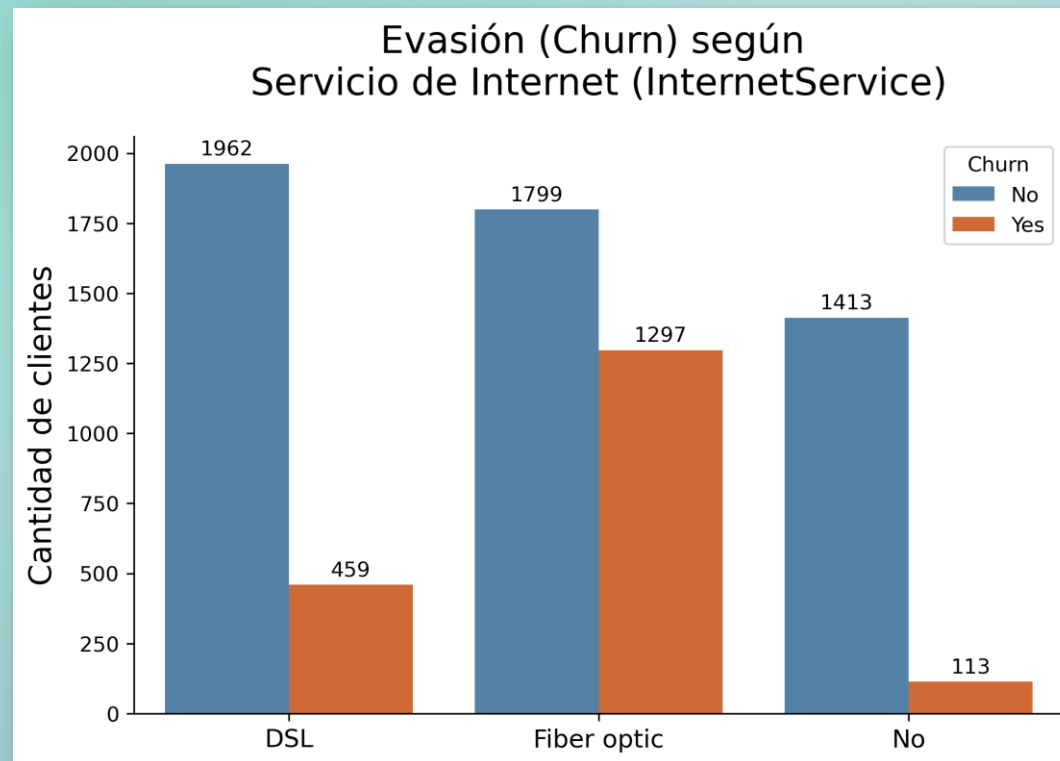
Datos originales (antes del modelado)

Servicio de Fibra Óptica:

InternetService_Fiber optic	0.2774	1.3197	Aumenta las chances en: 31.97%
-----------------------------	--------	--------	--------------------------------

- Clientes que cuentan con servicio de **Fibra Óptica** tienen una **probabilidad relativa de cancelación 31.97% mayor**, que aquellos que cuenta con servicio **DSL** (*variable de referencia*).
- No obstante, los datos disponibles no permiten identificar con certeza las causas de esta insatisfacción. Se recomienda **profundizar el análisis** mediante encuestas o entrevistas dirigidas a clientes que hayan cancelado este servicio, así como incorporar métricas técnicas como **latencia, velocidad, interrupciones del servicio** y otros factores que puedan influir en la **experiencia del cliente**

Esto confirma lo observado en el análisis exploratorio donde se descubrió que la **proporción de cancelación** para clientes que contaban con dicho servicio era **significativamente mayor** que aquellos con servicio DSL o sin servicio de internet.



Datos originales (antes del modelado)

Condición de Adulto Mayor (Senior Citizen):

SeniorCitizen_Yes	0.2718	1.3124	Aumenta las chances en: 31.24%
-------------------	--------	--------	--------------------------------

Medios de Pago (PaymentMethod) utilizados por Adultos Mayores (Senior Citizen) que abandonan

Payment Method	Proporción (%)
Bank transfer (automatic)	11.13
Credit card (automatic)	13.03
Electronic check	66.6
Mailed check	9.24

Datos originales (antes del modelado)

- Si bien esta variable no es utilizada por el modelo **Random Forest**, presenta un coeficiente significativo en la **Regresión Logística**, lo cual se encuentra en consonancia con lo observado durante la etapa de exploración.
- Este resultado puede vincularse con **posibles barreras tecnológicas**, particularmente en lo referido a **medios de pago digitales como el "Electronic Check"**. La adopción forzada de herramientas electrónicas podría generar fricciones y descontento en este grupo etario, afectando su experiencia general con el servicio.

El impacto elevado que esta variable en el abandono, confirman que **este sector demográfico se asocia a mayor probabilidad de evasión**, lo que indica que para reducir la tasa de abandono, será necesario mejorar la experiencia de dicho sector dentro de la empresa.

Evasión (Churn) - Adulto Mayor (SeniorCitizen)

SeniorCitizen	Tasa de Abandono (%)
Yes	41.68
No	23.61

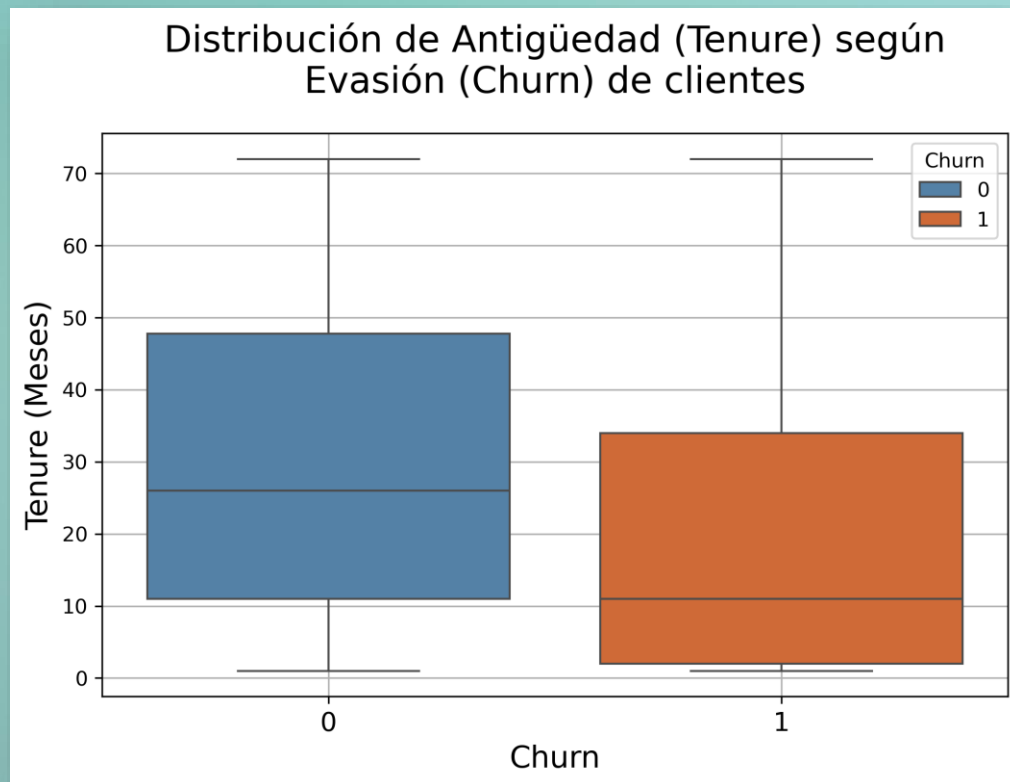
Datos originales (antes del modelado)

Factores de protección



Antigüedad (Tenure):

Tenure_scaled	-1.8563	0.1563	Disminuye las chances en: 84.37%
---------------	---------	--------	----------------------------------



- La **antigüedad** representa el **mayor factor de protección frente a la evasión**. Un cliente que alcanza los **37 meses de permanencia** reduce su probabilidad relativa de abandono en un **84.37%** en comparación con un cliente nuevo (referencia).
- Incluso después de balancear las clases (datos utilizados para el modelado), el **50% de los abandonos** se producen dentro de los **primeros 11 meses** del ciclo de vida.

Tipos de contrato:

Contrato de dos años (Two year) y Contrato de un año (One year)

Contract_One year	-0.0567	0.9449	Disminuye las chances en: 5.51%
-------------------	---------	--------	---------------------------------

- Los resultados del modelo indican que, a medida que **aumenta el compromiso contractual del cliente**, las **probabilidades relativas de abandono disminuyen** significativamente en comparación con aquellos que eligen contratos mensuales (Contract *Month-to-month*, referencia del modelo).
 - **Contrato de un año:** reduce la probabilidad relativa de evasión en un **5.51%**.
- Esto sugiere que los **contratos de mayor duración** actúan como un **importante factor de retención**, lo que resalta la **eficacia de políticas comerciales que incentiven acuerdos de largo plazo** para reducir la tasa de cancelación.

El peso otorgado a ambas variables confirma lo observado en la etapa exploratoria, donde se identificó que el **42.71%** de los clientes con **contratos mensuales** (Contract *Month-to-month*) **abandonaban la empresa**.

Este hallazgo resalta la importancia de diseñar **políticas comerciales que incentiven compromisos contractuales de mediano o largo plazo**, como una medida efectiva para **reducir la tasa de evasión**.

Evasión (Churn) - Tipo de Contrato (Contract)

Contract	Tasa de Abandono (%)
Month-to-month	42.71
One year	11.27
Two year	2.83

Datos originales (antes del modelado)

Servicio adicionales:

Soporte Técnico (Tech Support) y Seguridad Online (Online Security)

OnlineSecurity_Yes	-0.169	0.8445	Disminuye las chances en: 15.55%
TechSupport_Yes	-0.2357	0.79	Disminuye las chances en: 21.0%

- La **contratación de servicios adicionales**, particularmente el soporte técnico y la seguridad online, se asocia a una **reducción significativa en la probabilidad de abandono del cliente**.
 - **Soporte técnico**: disminuye la probabilidad relativa de evasión en **21%**.
 - **Seguridad online**: disminuye la probabilidad relativa de evasión en **15.55%**.

Si bien en la etapa de exploración se planteó que mientras más servicios tuviese un cliente, menos propenso era a abandonar la empresa, a través del presente análisis se determinó que **no depende directamente de la cantidad, sino del tipo de servicio**, ya que, **Seguridad Online y Tech Support actúan como protectores**, pero el **servicio de Streaming de Películas es un factor de riesgo**.

Evasión (Churn) - Cantidad de Servicios Contratados

Cantidad de Servicios	Tasa de Abandono (%)
1	10.92
2	30.97
3	44.92
4	36.48
5	31.34
6	25.55
7	22.49
8	12.41
9	5.29

Datos originales (antes del modelado)

Resumen Ejecutivo



- Se ha desarrollado una herramienta predictiva basada en modelos de Machine Learning capaz de identificar correctamente entre **8 y 9 de cada 10 clientes que abandonan** la empresa, permitiendo **anticipar cancelaciones de servicio y actuar de forma proactiva**.
- Si bien la evasión se concentra en las **etapas tempranas del ciclo de vida del cliente**, se detecta una **anomalía preocupante**: clientes de **larga antigüedad y alto valor** también deciden cancelar el servicio. Este grupo contradice los patrones generales y requiere ser analizado con mayor profundidad, especialmente considerando que la variable **Antigüedad (Tenure)** actúa como el **mayor factor de protección frente al abandono**.
- Los contratos de **mediana duración** (contratos anuales) se destacan como una **estrategia eficaz de fidelización**, disminuyendo significativamente las probabilidades de evasión.
- Por el contrario, se detecta **insatisfacción con los servicios de Fibra Óptica**, lo que podría estar vinculado a problemas técnicos no reflejados en los datos disponibles.
- Finalmente, se identifica un **descontento entre personas mayores de 65 años** asociado al uso del **sistema de facturación electrónica**, lo que sugiere la presencia de **barreras tecnológicas** que podrían mitigarse mediante acompañamiento personalizado o alternativas de facturación más accesibles.

Sugerencias Estratégicas



1. Campañas de retención para clientes antiguos

Identificar y analizar en profundidad al grupo atípico de clientes de **alta antigüedad y alto valor** que decidieron cancelar el servicio. Desarrollar **promociones exclusivas** y acciones de fidelización específicas para este segmento, con el objetivo de reforzar el vínculo con la empresa y prevenir pérdidas inesperadas.

2. Fidelización de Adultos Mayores

Diseñar **planes especiales dirigidos a clientes de 65 años o más**, que incluyan servicios como **Tech Support sin costo adicional**. Esta medida busca **reducir barreras tecnológicas**, mejorar la experiencia de uso y disminuir la tasa de abandono en un segmento demográfico que presenta una **alta vulnerabilidad a la evasión**.

3. Incentivos por Contratación Combinada y Contratos Anuales o Bianuales

Implementar **descuentos o beneficios progresivos** por la contratación de **múltiples servicios** (por ejemplo: Phone Service, Tech Support, Online Security), **bajo contratos de duración mínima de un año**. Esta estrategia tiene como objetivo **superar el umbral crítico de deserción temprana**, y se apoya en la relación positiva entre **diversidad de servicios, antigüedad y fidelización**.

4. Monitoreo y Mejora de los Servicio de Fibra Óptica y Streaming

Establecer un sistema continuo de **monitoreo de calidad y feedback del cliente** para los servicios de **Fiber Optic y Streaming Movies**, ambos asociados a mayor riesgo de abandono. Se recomienda complementar con análisis de métricas internas (como tickets de soporte, interrupciones de servicio, tiempos de respuesta), y **evaluaciones cualitativas** (encuestas, entrevistas, reseñas) que permitan detectar **puntos críticos** y generar planes de mejora concretos.

Este análisis demuestra cómo el Machine Learning, aplicado con criterio estratégico, permite desatar el poder de los datos para transformar el conocimiento en acción. Comprender al cliente desde la evidencia no solo mejora la retención, sino que habilita decisiones inteligentes y sostenibles.

Gracias por su tiempo. Estoy abierto a preguntas, sugerencias y colaboración.

Ignacio Majo – Data Scientist Jr.
Proyecto parte del programa Oracle Next Education



[LinkedIn](#)



[GitHub](#)