

# Informe Detallado

## Telecom X: Predicción de Cancelación (Churn)

Objetivo: Este informe busca identificar los principales factores que impulsan la cancelación de clientes en Telecom X, basándose en el análisis de datos y el rendimiento de modelos predictivos, para proponer estrategias de retención efectivas y personalizadas.

### 1. Factores Principales que Influyen en la Cancelación de Clientes.

El análisis de las variables del conjunto de datos, complementado con el entrenamiento de modelos de clasificación y la evaluación de la importancia de las características, ha permitido identificar los siguientes factores como los más influyentes en la probabilidad de que un cliente abandone el servicio:

#### Tipo de Contrato.

- **Contratos Mensuales** (`Tipo de Contrato_Mensual`): Es la característica más dominante en la predicción del abandono según el modelo *XGBoost* (importancia por Defecto de 0.483556, por Gain de 55.954838, y por Cobertura de 304.512360). También es relevante en el análisis de *Random Forest*. Tiene **fuerte correlación positiva (0.405) con el abandono**. Esto indica que los clientes con contratos mensuales tienen una probabilidad significativamente mayor de cancelar su servicio.
- **Contratos Bianuales** (`Tipo de Contrato_Bianual`): Muestra una **fuerte correlación negativa (-0.302)** con el abandono. Su presencia es importante en las métricas de *XGBoost* (0.483556 por Defecto, 12.263952 por Gain, 259.895508 por Cobertura). Esto sugiere que los contratos de mayor duración son un factor protector significativo contra el churn.

#### Servicio de Internet (`Servicio de Internet_Fibra Óptica`).

- La suscripción al servicio de Fibra Óptica presenta una **correlación positiva (0.308)** con el abandono. Es una característica importante en *XGBoost* (17.124043 por Gain, 224.965012 por Cobertura) y en *Random Forest*. Esto es un hallazgo crítico que sugiere **problemas potenciales en la calidad del servicio**, la satisfacción del cliente o la relación costo-beneficio percibida por los usuarios de **fibra óptica**.

#### Costos (`sqr_costo_total`, `sqr_costo_mensual`).

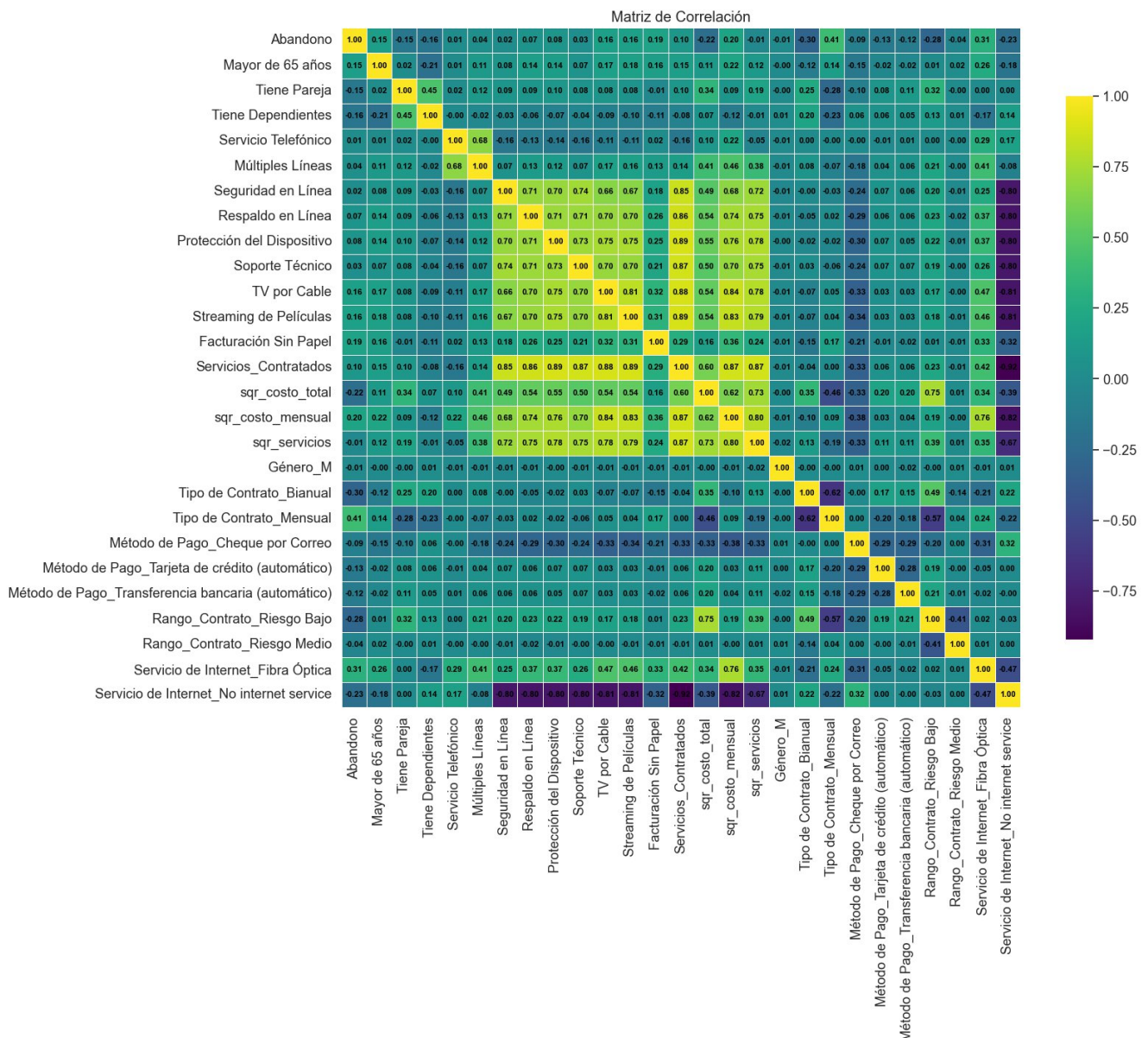
- Las transformaciones de Costo Total (correlación negativa de -0.223) y Costo Mensual (correlación positiva de 0.203) tienen un impacto significativo. El análisis con **Random Forest resalta la mayor prominencia de las variables de costo**, situándolas en las primeras posiciones en importancia nativa y por reducción de impureza. Esto sugiere que tanto un costo total bajo (o una relación costo-duración percibida como baja) como un costo mensual alto pueden influir en la decisión de abandono.

#### Servicios Adicionales (`Seguridad y Soporte Técnico`).

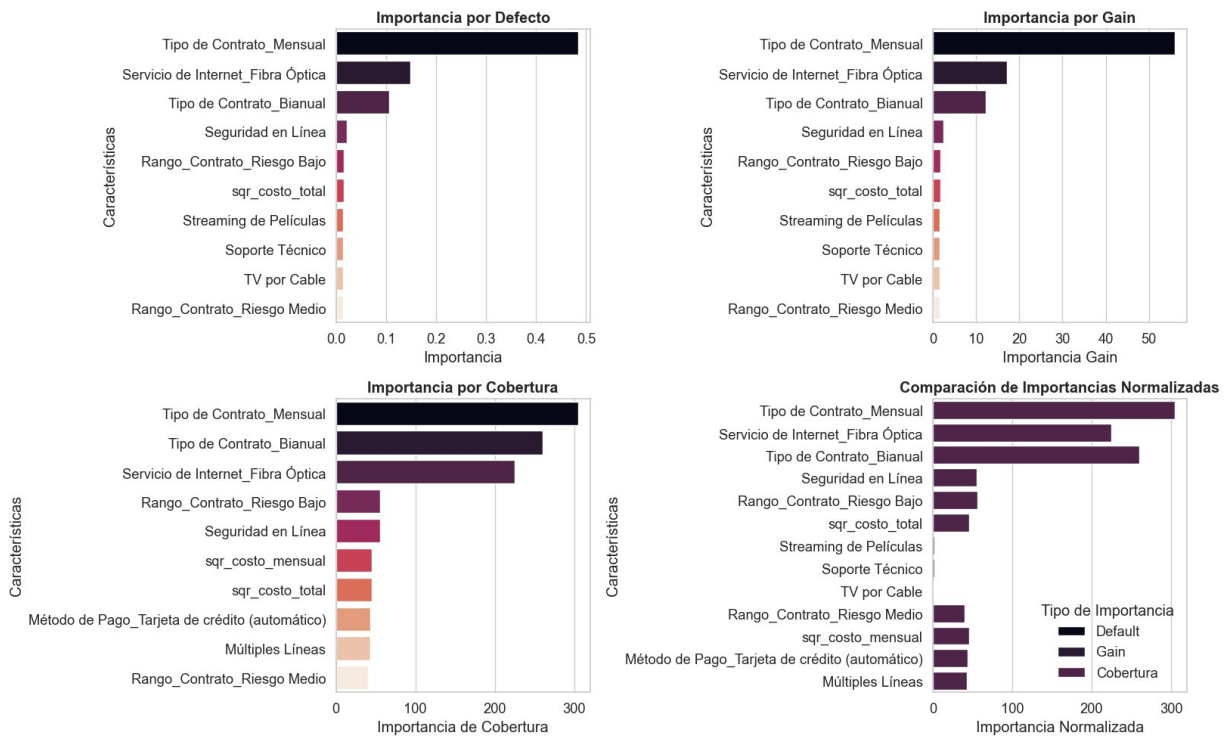
- Servicios como "Seguridad en Línea" y "Soporte Técnico" están fuertemente relacionados con la reducción de la probabilidad de abandono. **Aparecen consistentemente en las métricas de importancia de ambos modelos**. Esto sugiere que estos servicios aumentan la lealtad del cliente.

## Otros Factores Relevantes.

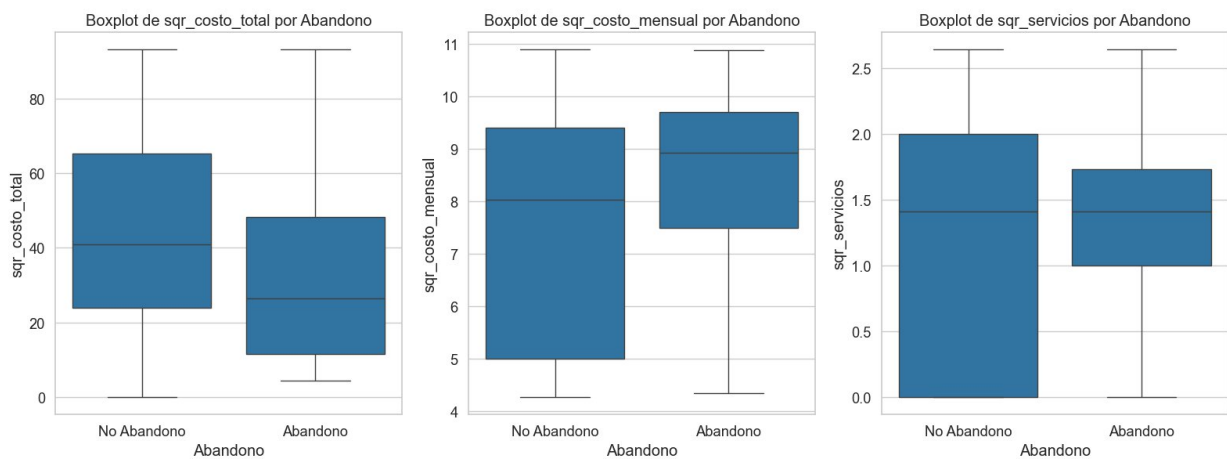
- **Facturación Sin Papel:** Presenta una correlación positiva (0.192) con el abandono.
- **Edad (Mayor de 65 años):** Muestra una correlación positiva (0.151), indicando que este segmento es más propenso al churn.
- **Situación Familiar (Tiene Dependientes, Tiene Pareja):** Correlaciones negativas (-0.164 y -0.150 respectivamente), sugiriendo que tener dependientes o pareja disminuye la probabilidad de abandono.
- **Método de Pago (Método de Pago\_Tarjeta de crédito (automático)):** Muestra una correlación negativa (-0.134). El Método de Pago\_Cheque electrónico es el más frecuente y puede ser un punto a considerar.
- **Rango de Contrato (Rango\_Contrato\_Riesgo Bajo):** Tiene una correlación negativa (-0.284), lo cual es esperable ya que esta categoría agrupa contratos de mayor duración.
- **Matriz de Correlación.**



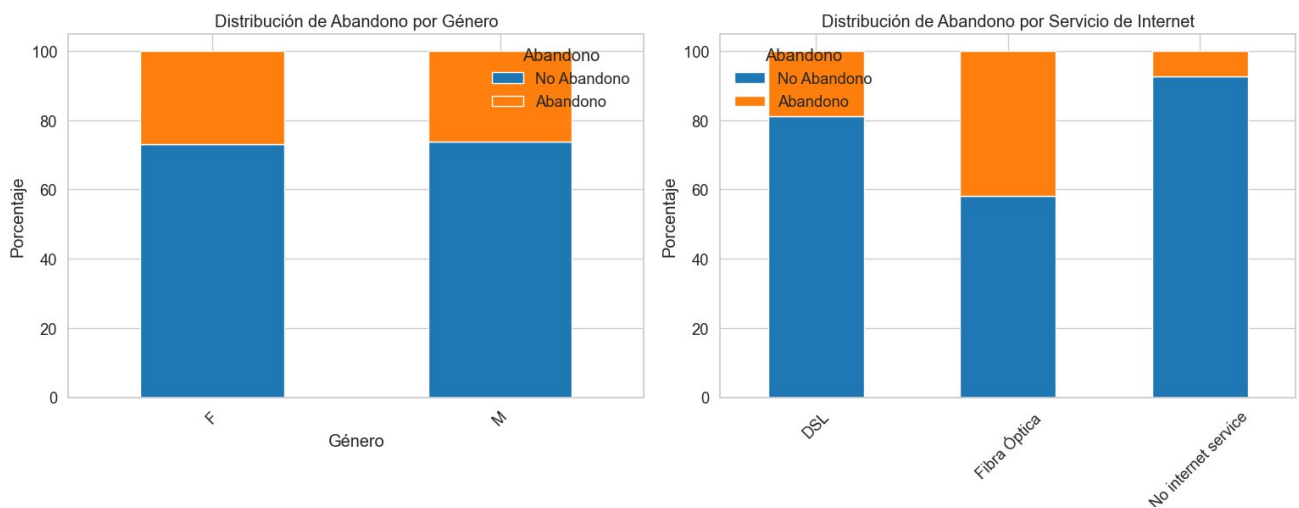
- **Importancia de Características**

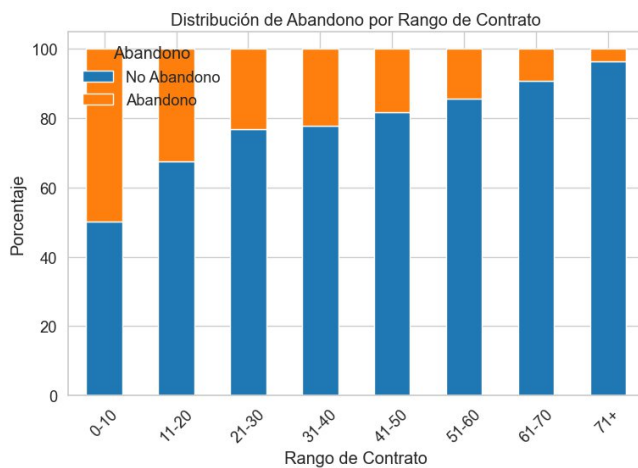
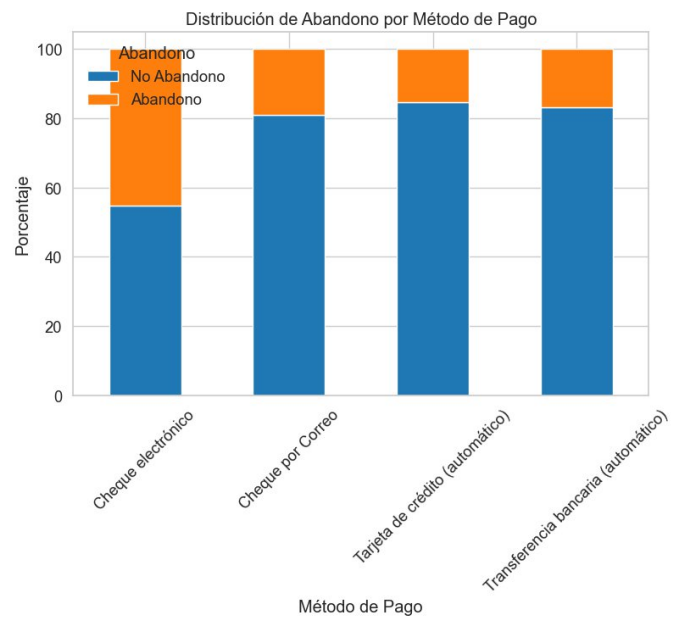
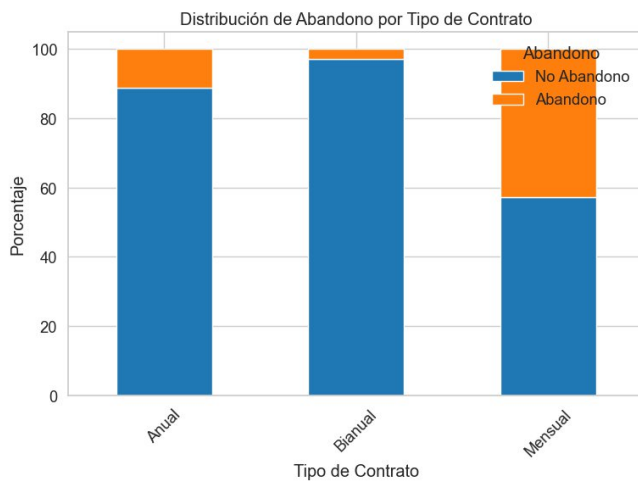


## • Boxplots para Variables Numéricas



## • Abandono en Variables Categóricas





## 2. Rendimiento de los Modelos.

Se entrenaron y evaluaron tres modelos de clasificación: Random Forest, CatBoost y XGBoost.

- Modelos sin ajustar**

	Random Forest	CatBoost	XGBoost.
Accuracy	0.7821	0.7857	0.7771
Precision	0.6175	0.6169	0.5943
Recall	0.4706	0.5080	0.5053
F1-score	0.5341	0.5572	0.5462

- Modelos ajustados**

	Random Forest	CatBoost	XGBoost.
Accuracy	0.7942	0.7935	0.7786
Precision	0.6458	0.6446	0.6020
Recall	0.4973	0.4947	0.4893
F1-score	0.5619	0.5598	0.5398

- **Modelos ajustados con variables reducidas**

	Random Forest	CatBoost	XGBoost.
Accuracy	0.8318	0.6425	0.9383
Precision	0.7370	0.8211	0.9042
Recall	0.5695	0.7132	0.8583
F1-score	0.6425	0.5454	0.8807

- **Rendimiento Inicial y Overfitting:** Inicialmente, XGBoost mostró un alto sobreajuste (una diferencia del 16.64% entre la precisión en entrenamiento del 94.50% y en prueba del 77.86%). Random Forest y CatBoost presentaron un sobreajuste moderado. Ese sobreajuste se trabajó posteriormente en la reducción de variables y
- **Optimización y Reducción de Variables:** Se implementó una estrategia de reducción de variables, eliminando aquellas con menor importancia. Esta optimización resultó en una mejora significativa en el rendimiento de todos los modelos, reduciendo el sobreajuste.
- **Matriz de Confusión:**
  - Verdaderos Positivos (VP): Casos positivos correctamente identificados
  - Verdaderos Negativos (VN): Casos negativos correctamente identificados
  - Falsos Positivos (FP): Casos negativos erróneamente identificados como positivos
  - Falsos Negativos (FN): Casos positivos erróneamente identificados como negativos

Random Forest	CatBoost	XGBoost
Matriz de confusión: [[959 76] [161 213]]	Matriz de confusión: [[953 82] [170 204]]	Matriz de confusión: [[1001 34] [ 53 321]]
<ul style="list-style-type: none"> <li>• Mayor número de falsos positivos (76)</li> <li>• Más falsos negativos (161)</li> </ul> <b>Menor precisión en la clasificación. Tiende a sobrestimar la clase positiva</b>	<ul style="list-style-type: none"> <li>• 82 falsos positivos</li> <li>• 170 falsos negativos</li> </ul> <b>Menor precisión en la clasificación de la clase positiva</b>	<ul style="list-style-type: none"> <li>• Mejor rendimiento general</li> <li>• Muy pocos falsos positivos (34)</li> <li>• Pocos falsos negativos (53)</li> </ul> <b>Alta capacidad para identificar correctamente ambas clases</b>

- Los tres modelos son mejores clasificando la clase negativa (0)
  - XGBoost parece ser el modelo más robusto con variables reducidas.
- **Validación cruzada:** Aunque para XGBoost reducido (0.7894) fue ligeramente inferior a la de CatBoost Reducido (0.8020) y Random Forest Reducido (0.7979), el rendimiento superior en el conjunto de prueba y las mejoras significativas en Precisión y Recall tras el ajuste del umbral (alcanzando 0.96 de precisión y 0.95 de recall con un **umbral de 0.45**) **justifican su selección como el modelo óptimo**. La fuente indica que esta diferencia es estadísticamente aceptable y no compromete la generalización del modelo.

### 3. Selección del Modelo Final

**XGBoost con variables reducidas:** Tras la reducción de variables, XGBoost fue el modelo que experimentó los mejores resultados posterior a la reducción de variable y ajuste del umbral. Se observó un aumento del Accuracy de 0.7786 a 0.9383 (+20.52%) en el conjunto de prueba. Sus métricas clave fueron: Precisión: 0.9042, Recall: 0.8583, F1-score: 0.8807.

### 4. Estrategias de Retención Propuestas

Basado en los factores de influencia identificados y el rendimiento del modelo final, se proponen las siguientes estrategias de retención.



## Enfoque en Clientes con Contrato Mensual

- **Incentivar la Migración:** Ofrecer a clientes con contratos mensuales planes bianuales o **anuales** con descuentos, beneficios exclusivos o servicios adicionales gratuitos.
- **Programas de Fidelización:** Crear programas de recompensas o fidelización para clientes específicos (mensuales) con ofertas personalizadas, descuentos graduales por antigüedad o acceso anticipado a nuevas promociones.
- **Revisión de Precios:** Analizar si el costo-beneficio percibido de los contratos mensuales es adecuado o si hay oportunidades para ajustar precios y paquetes para aumentar su atractivo y retención.

## Mejora de la Experiencia con el Servicio de Fibra Óptica.

- **Diagnóstico de Problemas:** Realizar una investigación profunda sobre las razones de la alta correlación de la fibra óptica con el abandono. Esto puede incluir encuestas de satisfacción, análisis de tickets de soporte técnico relacionados con la fibra, y monitoreo de la estabilidad de la conexión.
- **Refuerzo de Soporte Técnico:** Asegurar un soporte técnico de alta calidad y tiempos de respuesta rápidos para los usuarios de fibra óptica, dado que el soporte técnico es un factor clave de retención.
- **Comunicación de Valor:** Resaltar los beneficios exclusivos y la fiabilidad de la fibra óptica, y cómo se diferencian de la competencia.

## Promoción y Empaquetamiento de Servicios de Valor Añadido.

- **Venta Cruzada y Agrupación:** Promocionar activamente servicios como "Seguridad en Línea" y "Soporte Técnico" a todos los clientes, especialmente a los nuevos y aquellos en riesgo. Se pueden ofrecer en paquetes atractivos para aumentar su adopción.
- **Educación al Cliente:** Informar a los clientes sobre cómo estos servicios adicionales mejoran su experiencia y seguridad, y cómo pueden prevenir problemas futuros.

## Atención Personalizada a Segmentos de Riesgo.

- **Clientes Mayores de 65 años:** Diseñar campañas de comunicación claras y sencillas, ofertas adaptadas a sus necesidades (ej., paquetes de voz y datos simplificados, soporte técnico especializado para personas mayores).
- **Clientes sin Dependientes/Pareja:** Considerar estrategias que fomenten la lealtad individual, como beneficios por suscripción a largo plazo o programas de referencia.

## Optimización de Métodos de Pago y Facturación.

- **Revisar "Facturación Sin Papel":** Investigar si existen dificultades o insatisfacción con el sistema de facturación electrónica. Ofrecer alternativas o mejorar la usabilidad del sistema para aquellos clientes que puedan preferir la facturación tradicional o tener problemas con la digital.
- **Incentivar Métodos de Pago Automatizados:** Promocionar el uso de tarjetas de crédito o transferencias bancarias automáticas, ya que estos métodos se correlacionan negativamente con el abandono.