

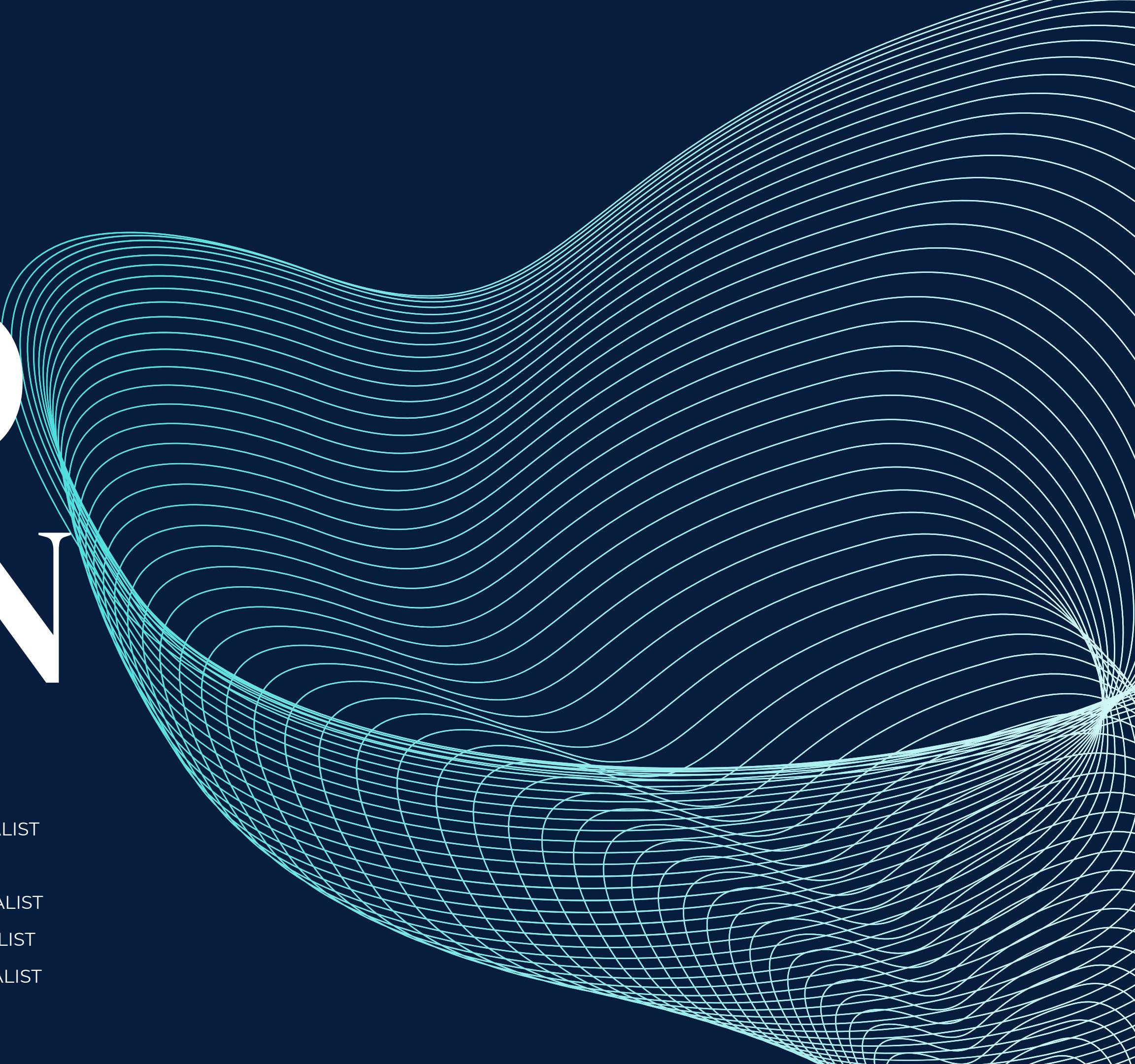
DATA SA

Mayo 2022

TELCO CHURN

EQUIPO:

- LEANDRO MÁRQUEZ – DATA SCIENCE SPECIALIST
- JAVIER GOHLKE – DATA SCIENCE SPECIALIST
- FEDERICO MÁRQUEZ – DATA SCIENCE SPECIALIST
- GUILLERMO FLORES – DATA SCIENCE SPECIALIST
- FRANCISCO RIVEROS – DATA SCIENCE SPECIALIST



Índice de contenidos

03	PRESENTACION
06	OBJETIVOS
07	CONFORMACION DE LOS DATOS
08	ANALISIS EXPLORATORIO DE LOS DATOS
18	SEGMENTACION DE CLIENTES
20	MODELOS DE MACHINE LEARNING
25	CONCLUSIONES
26	TABLA DE VERSIONADO

Presentacion

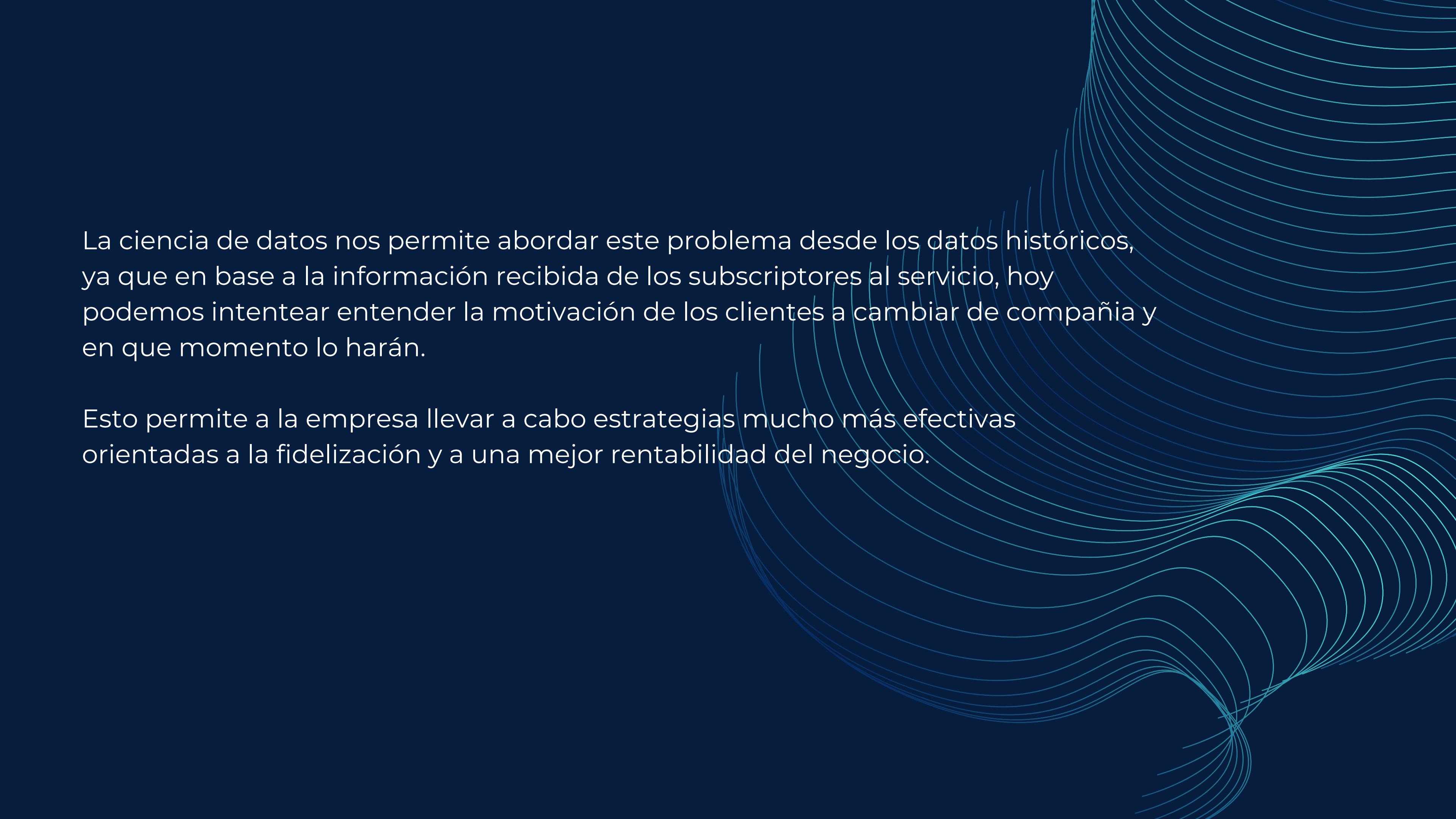
¿POR QUE LOS CLIENTES QUIEREN DEJAR LA EMPRESA?

Hace 20 años los desafíos de las operadoras de telecomunicaciones eran bastante simples, solo se tenían que preocupar que los clientes pudieran hacer llamadas y enviar mensajes de textos, asegurando estos dos servicios los clientes estaban bastante conformes.



Con el tiempo llegaron nuevos servicios, tecnologías como 3G, LTE y el avance del uso de internet para el consumo masivo trajo aparejado nuevos desafíos a las TELCO, no solo a nivel Técnico sino también en como proveer un servicio al cliente que sea satisfactorio.

Actualmente la competencia en el mercado y las exigencias de los clientes es mucho mayor por lo que entender cuáles son las razones por las que un cliente deja la compañía son de gran valor. ya que como sabemos siempre es menos costoso retener un cliente existente que obtener uno nuevo.



La ciencia de datos nos permite abordar este problema desde los datos históricos, ya que en base a la información recibida de los subscriptores al servicio, hoy podemos intentar entender la motivación de los clientes a cambiar de compañía y en qué momento lo harán.

Esto permite a la empresa llevar a cabo estrategias mucho más efectivas orientadas a la fidelización y a una mejor rentabilidad del negocio.

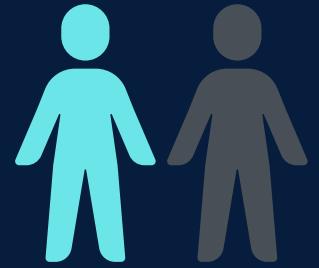
Objetivos

Buscamos responder las preguntas a las siguientes preguntas planteadas:

- ¿Existen patrones que nos permitan entender el comportamiento y las características de los clientes que dejan la compañía?
- ¿Es posible modelar un algoritmo que permita predecir si los clientes abandonarán el servicio?
- ¿Es posible segmentar a los clientes de acuerdo a ciertos parámetros definidos?



Conformacion de los Datos



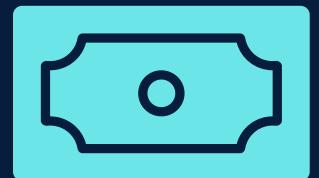
INFORMACION DEMOGRAFICA

* Genero / Edad / Estado civil / Hijos



SERVICIOS CONTRATADOS

* Servicio telefónico / Servicio de Internet / Servicio de Televisión / Soporte / etc.



INFORMACION DE LA CUENTA

* Tipo de contrato / Antigüedad / Metodos de pago / Facturación

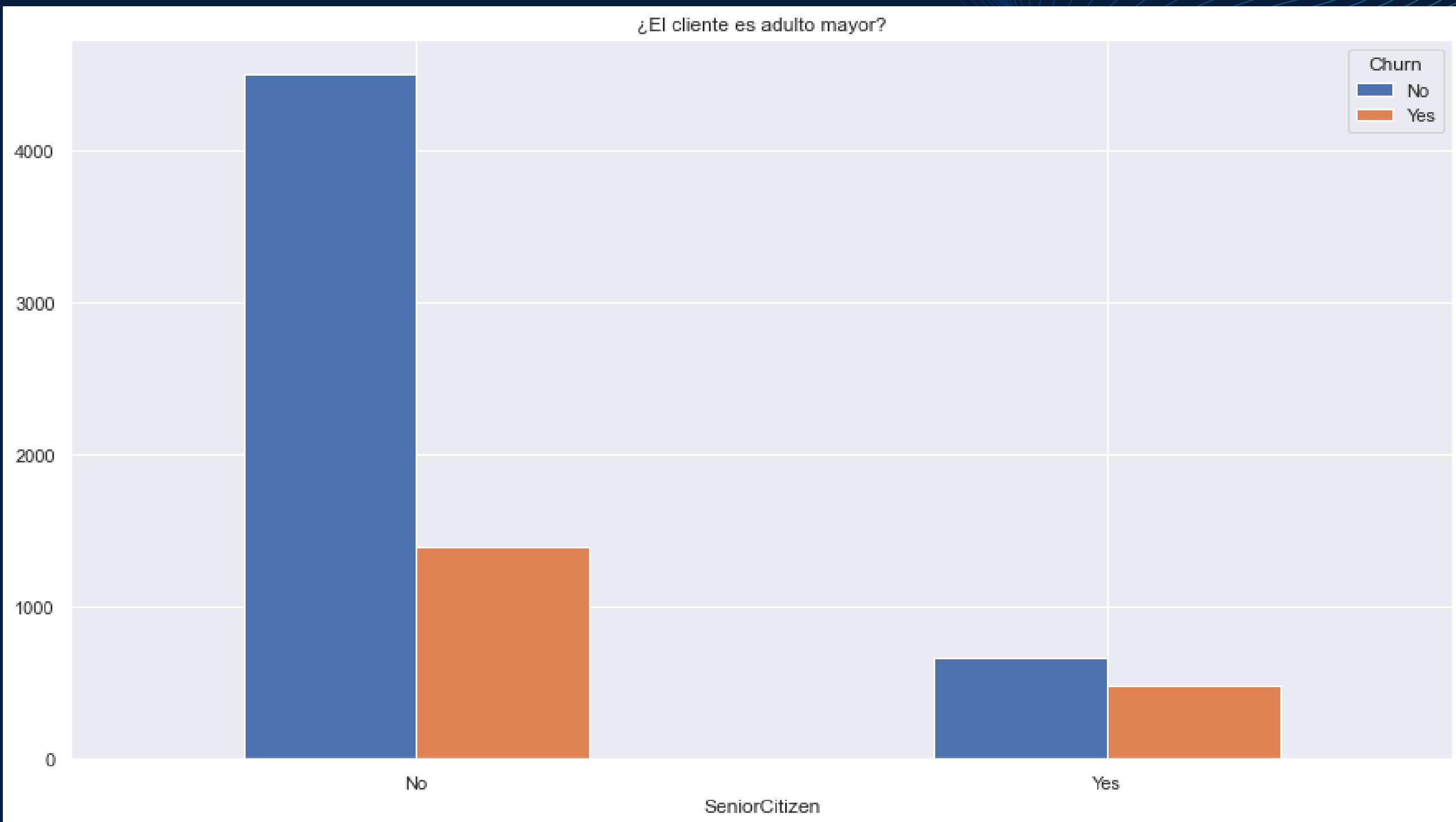


CHURN

* Abandono

Análisis exploratorio

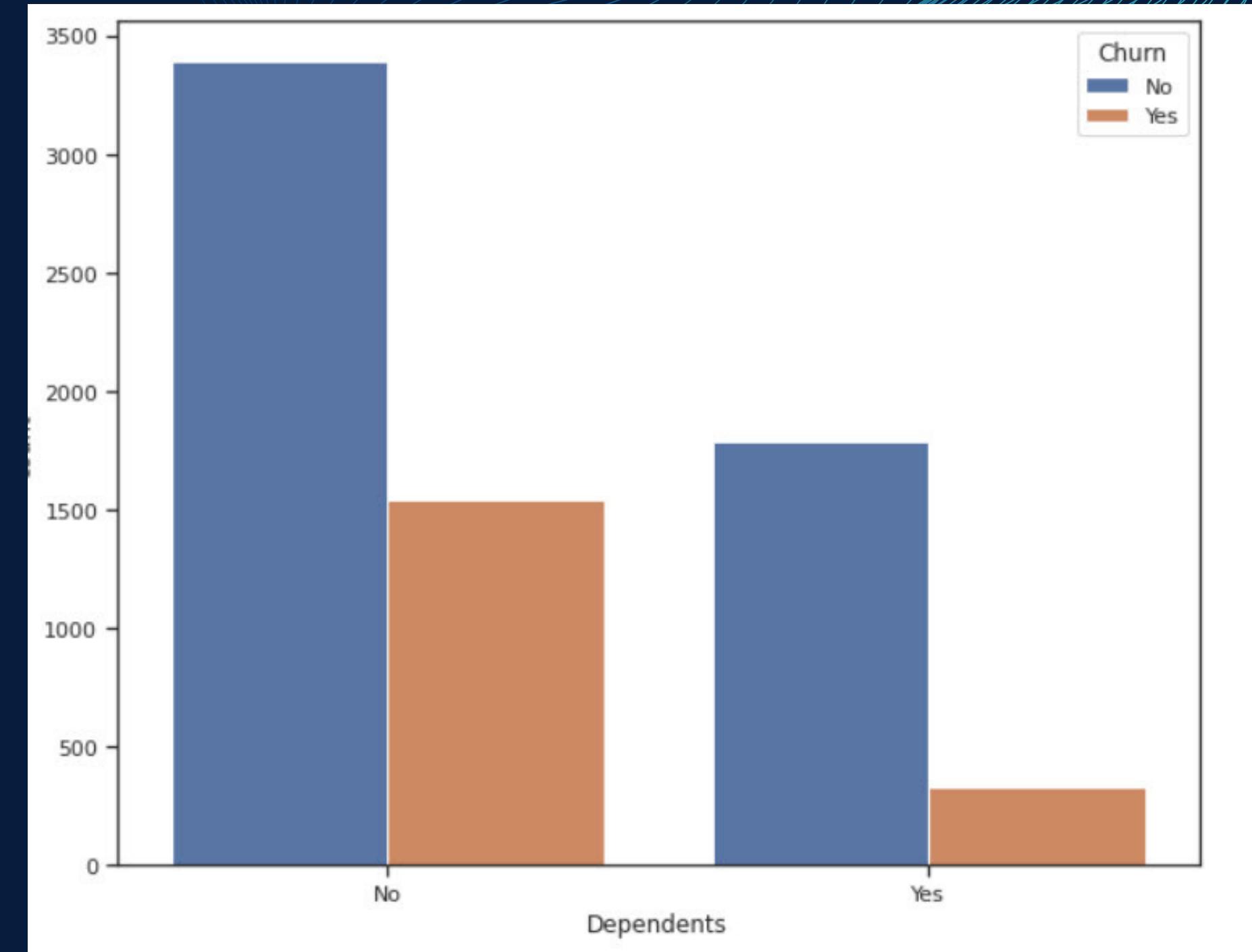
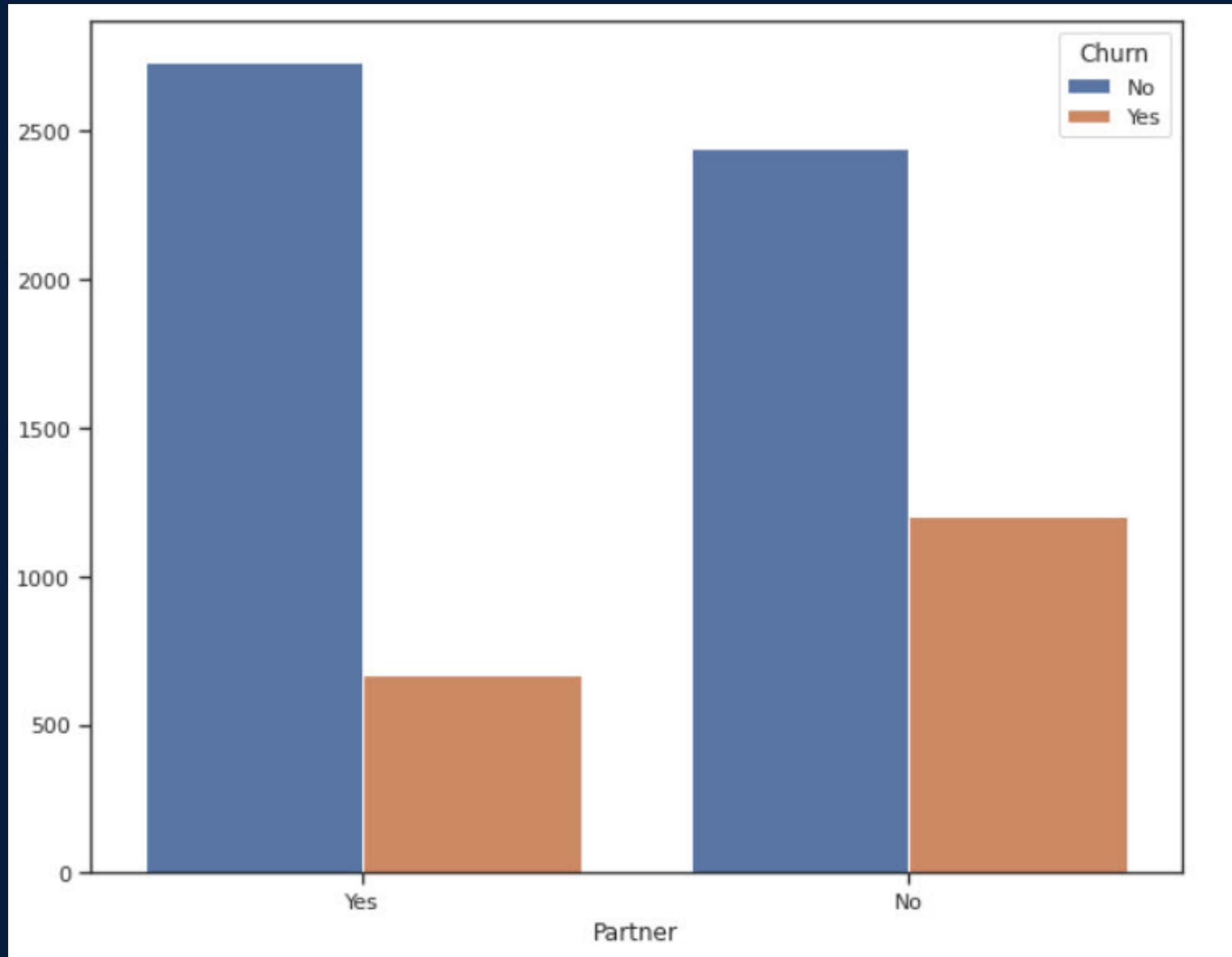
Según demografía



Los jubilados tienden a presentar mayores tasa de abandono

Análisis exploratorio

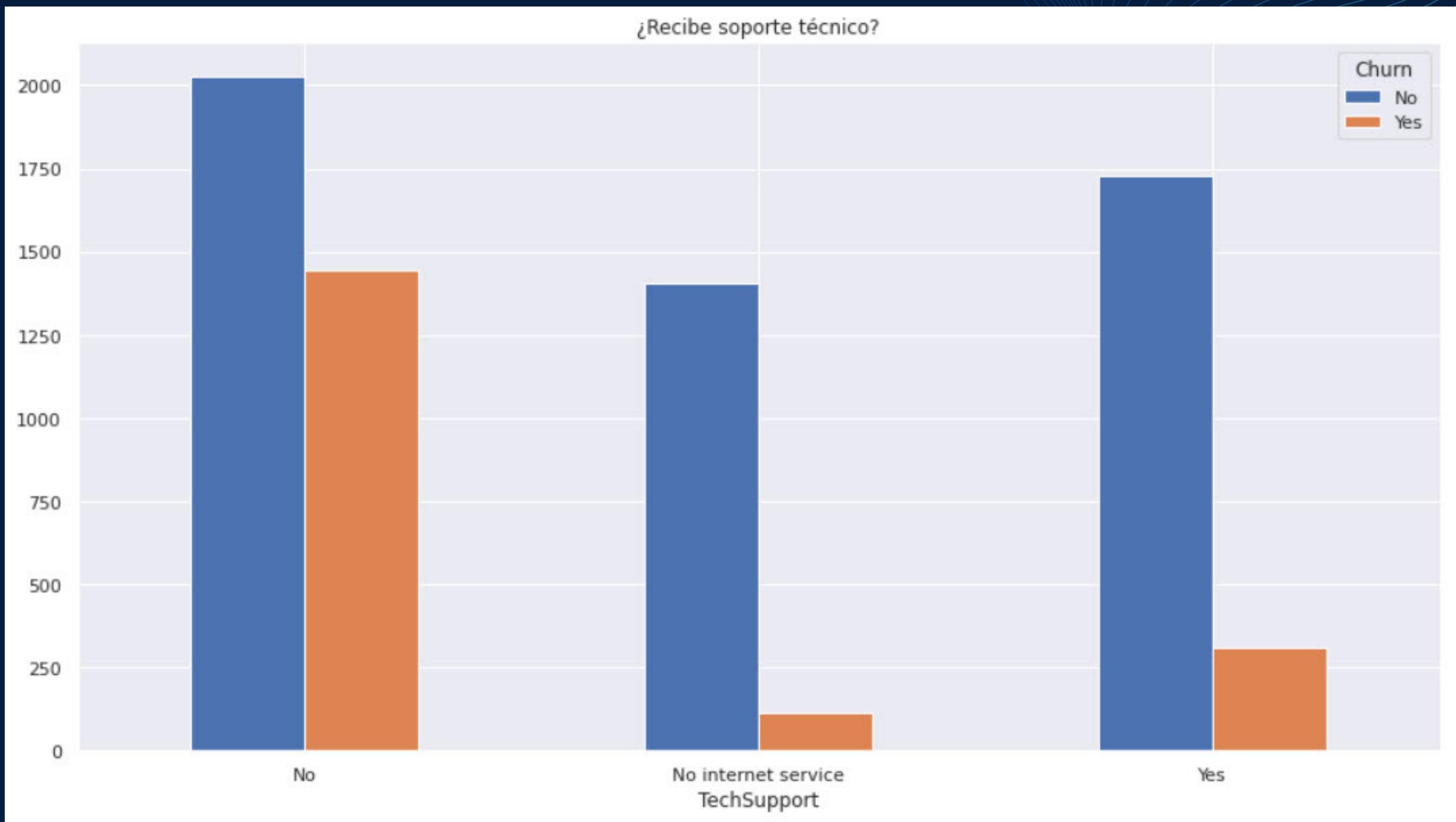
Según demografía



Los clientes Solteros y sin hijos son mas propensos a abandonar el servicio

Análisis exploratorio

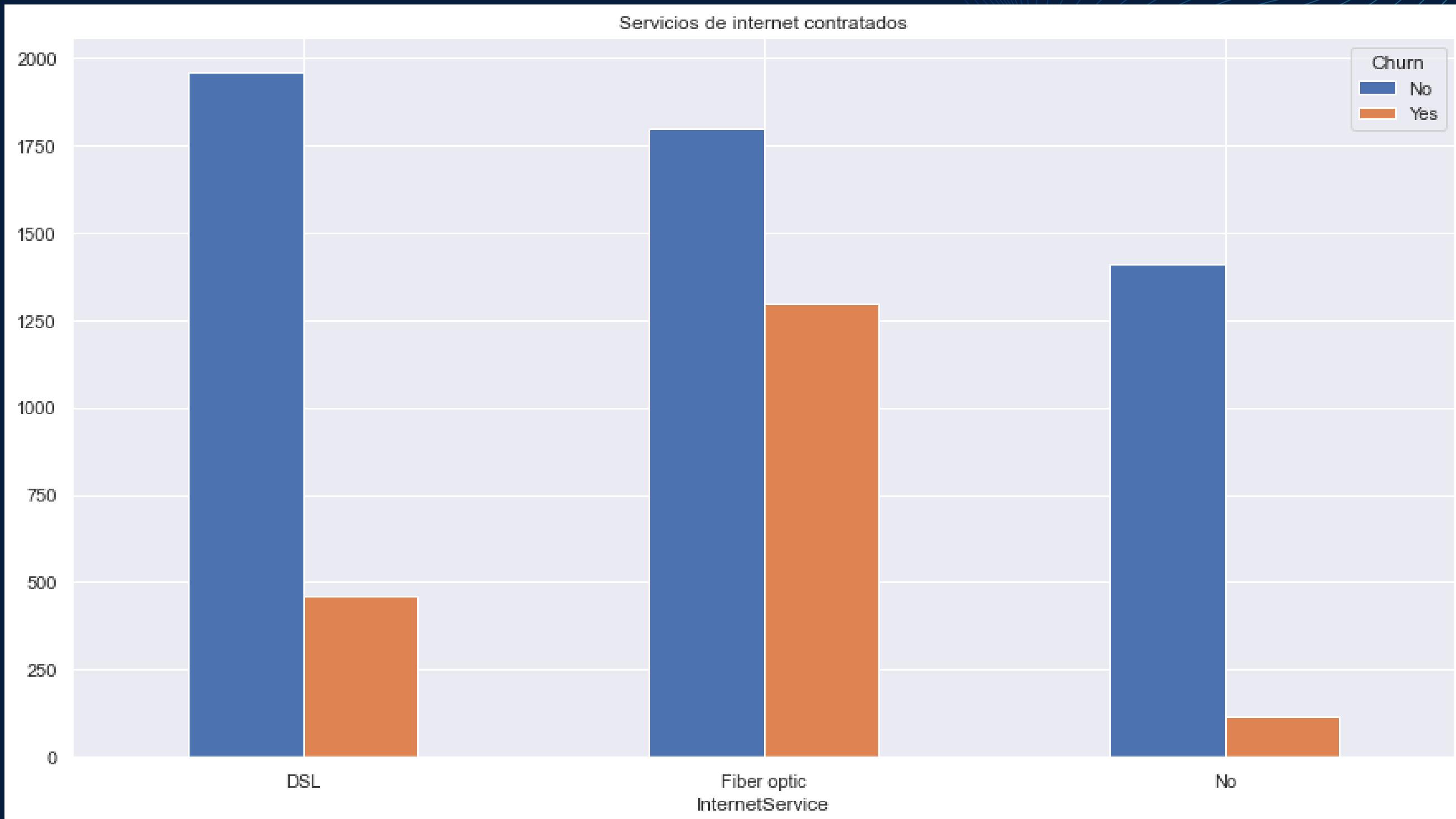
Según servicio



Clientes con servicio de internet sin soporte técnico presentan mayor abandono

Análisis exploratorio

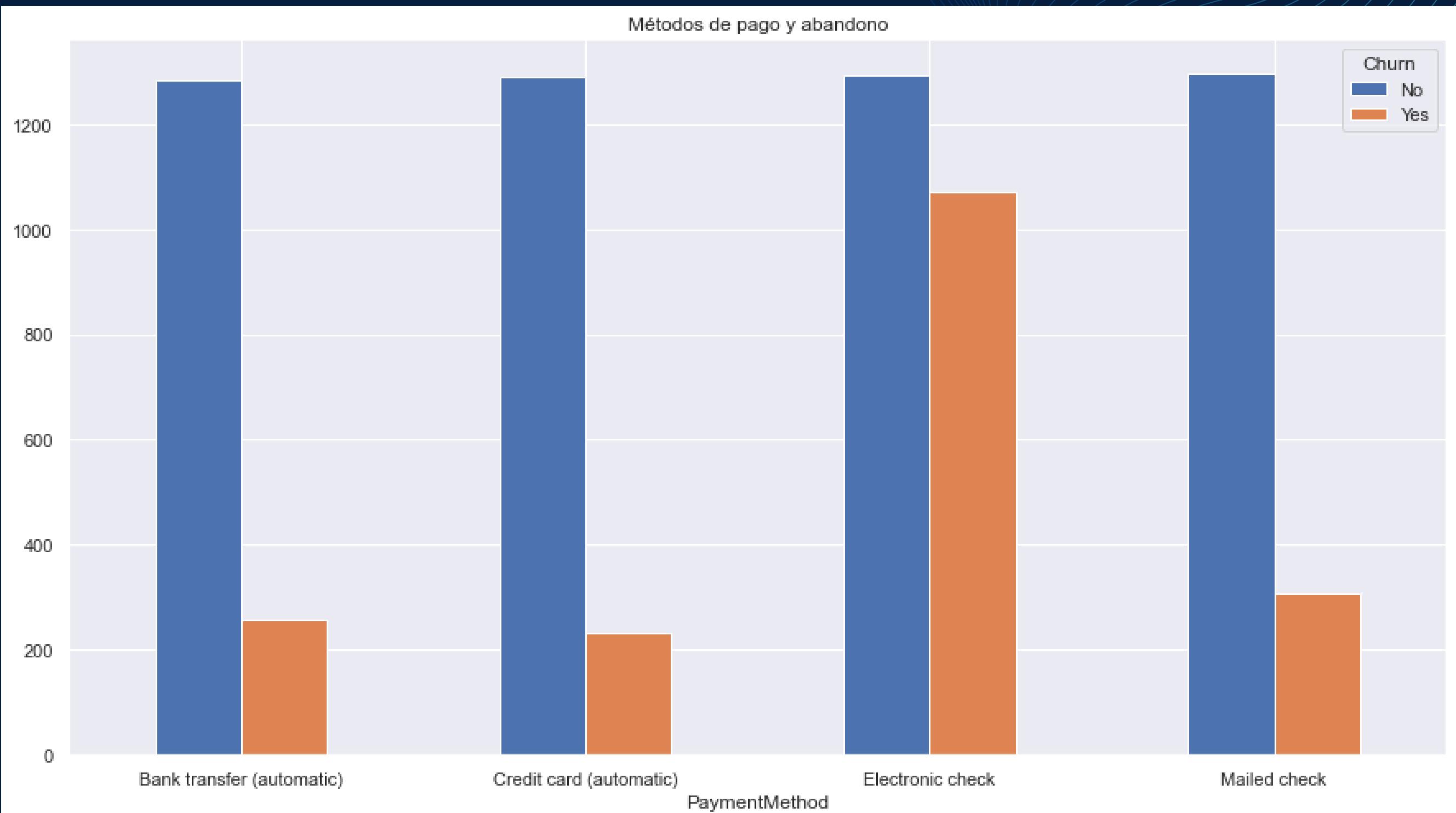
Según servicio



Hay un claro peso de abandonos dado por los clientes que tienen internet por fibra óptica. Probablemente, el servicio esté dando problemas técnicos o sea muy caro.

Análisis exploratorio

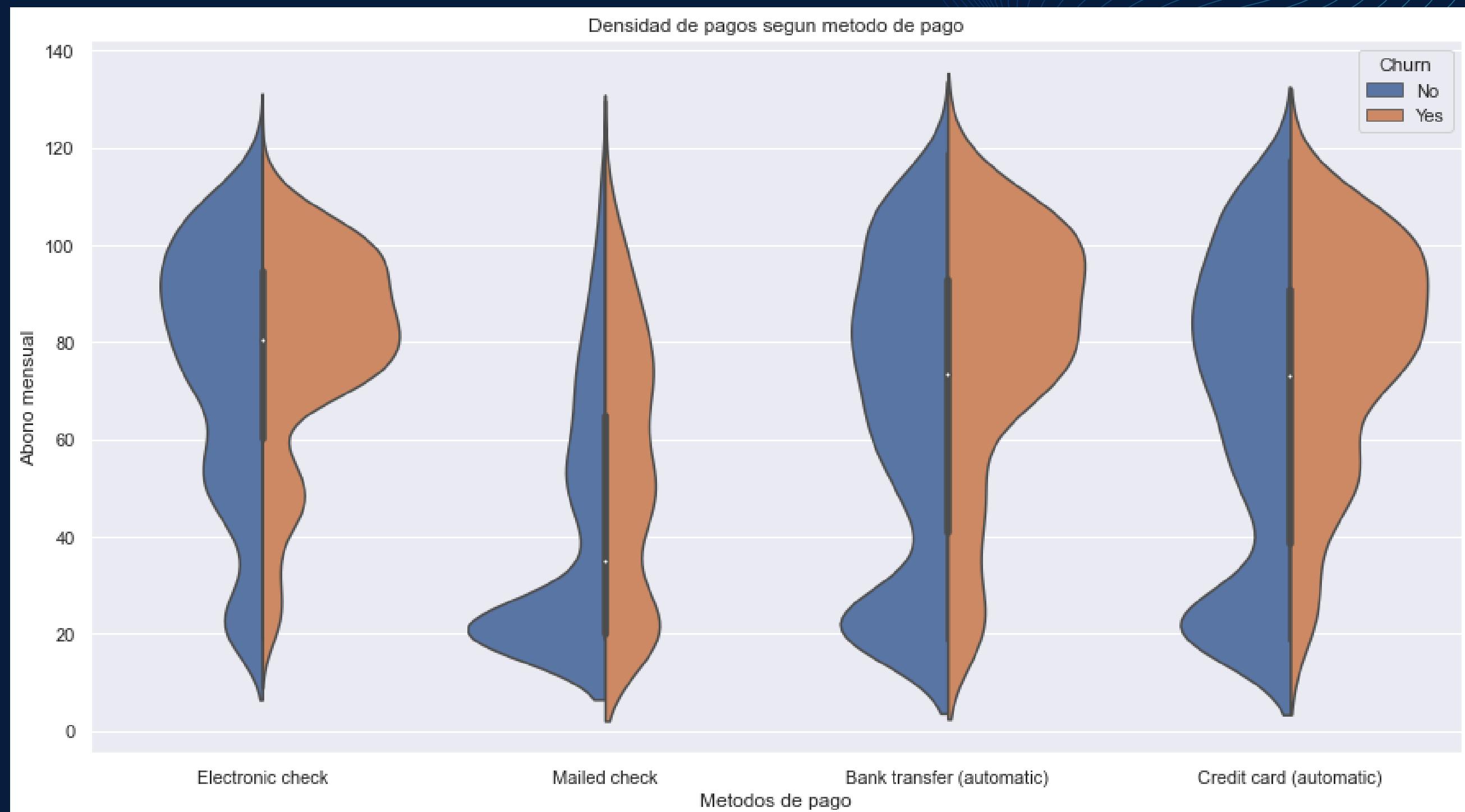
Según condición contractual (Método de pago)



Se observa una fuerte tendencia al abandono con medios de pago mediante cheque electrónico

Análisis exploratorio

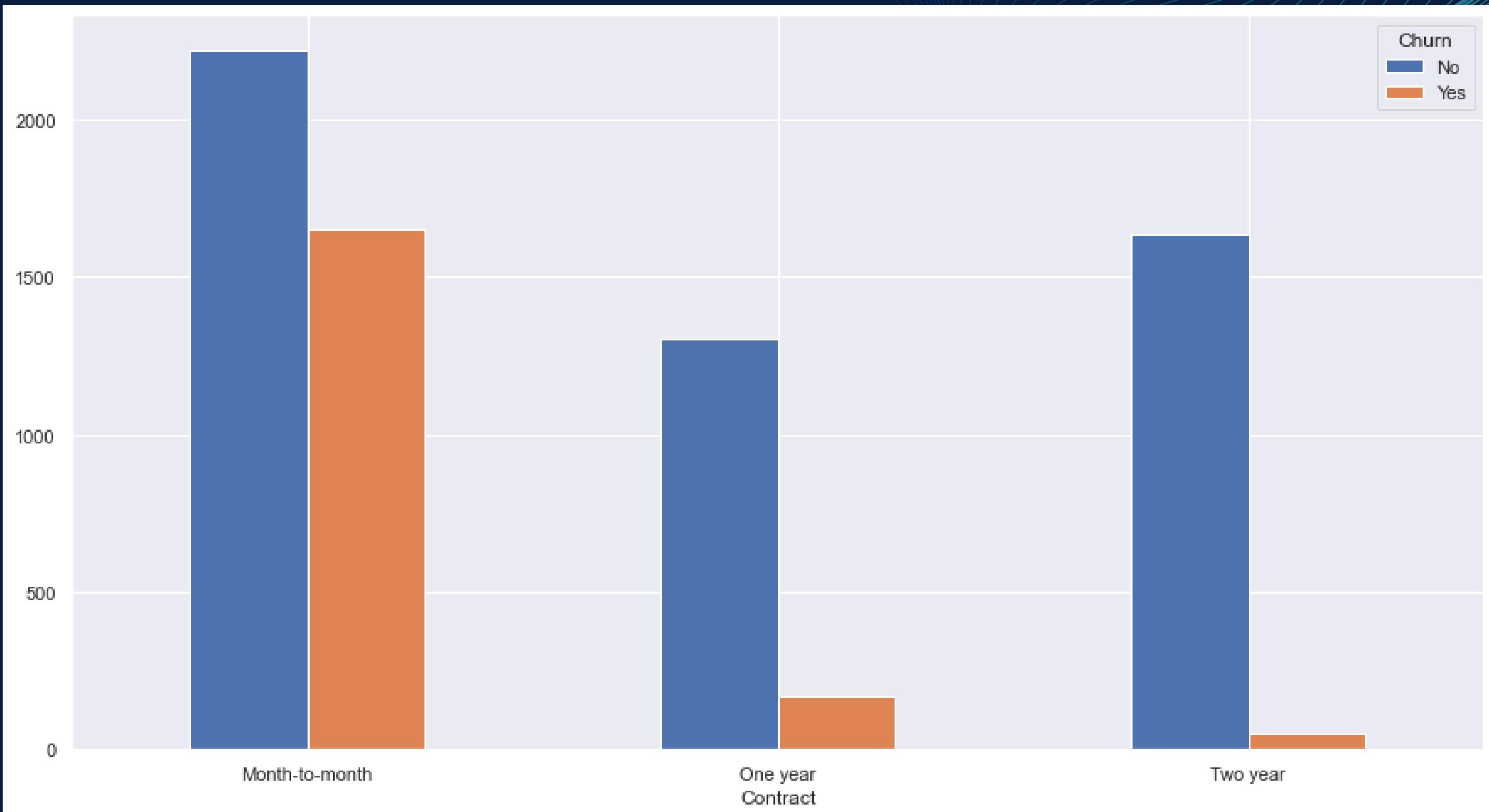
Según condición contractual (Método de pago VS Costo del servicio)



Se concluye que los abandonos se concentran en zonas de cobros altos y aun mas pronunciadamente en medio de pago con cheque electrónico

Análisis exploratorio

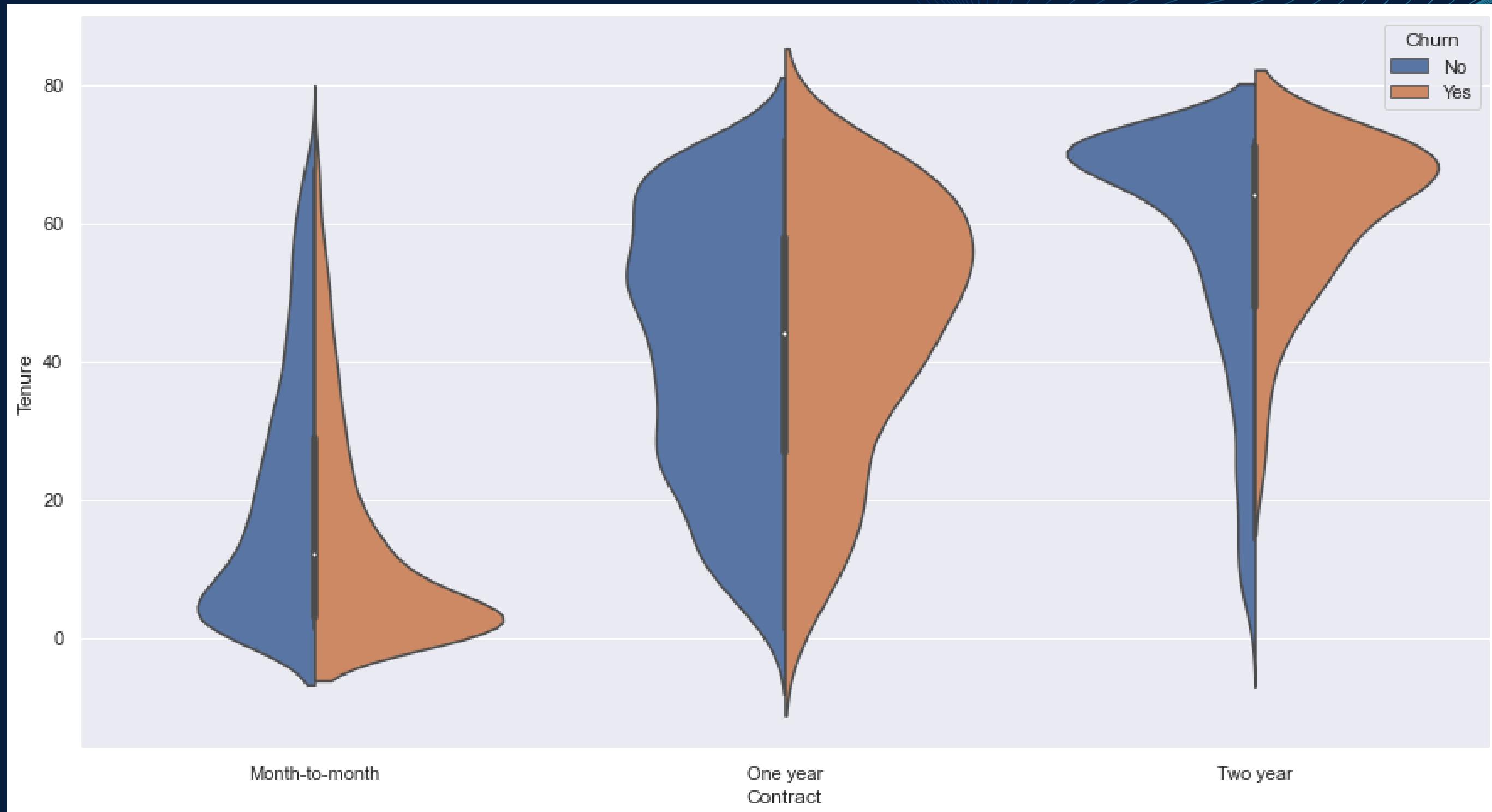
Según condición contractual (Tipo de contrato)



Los contratos de larga duración favorecen a la retención de los clientes

Análisis exploratorio

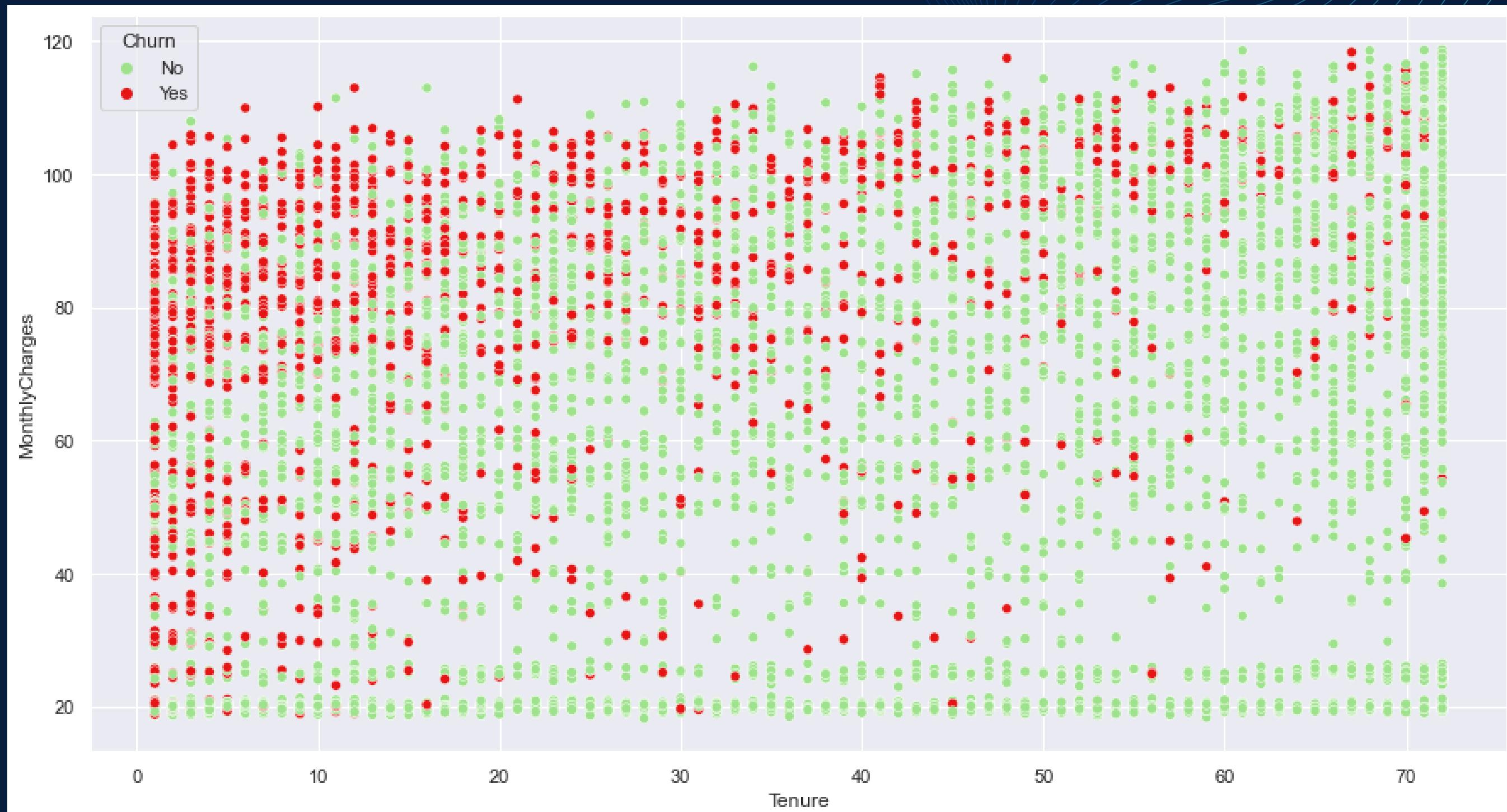
Según condición contractual (Tipo de contrato Vs Permanencia)



Con este gráfico podemos reafirmar lo concluido anteriormente, ya que contratos largos estén alineados con permanencias extendidas

Análisis exploratorio

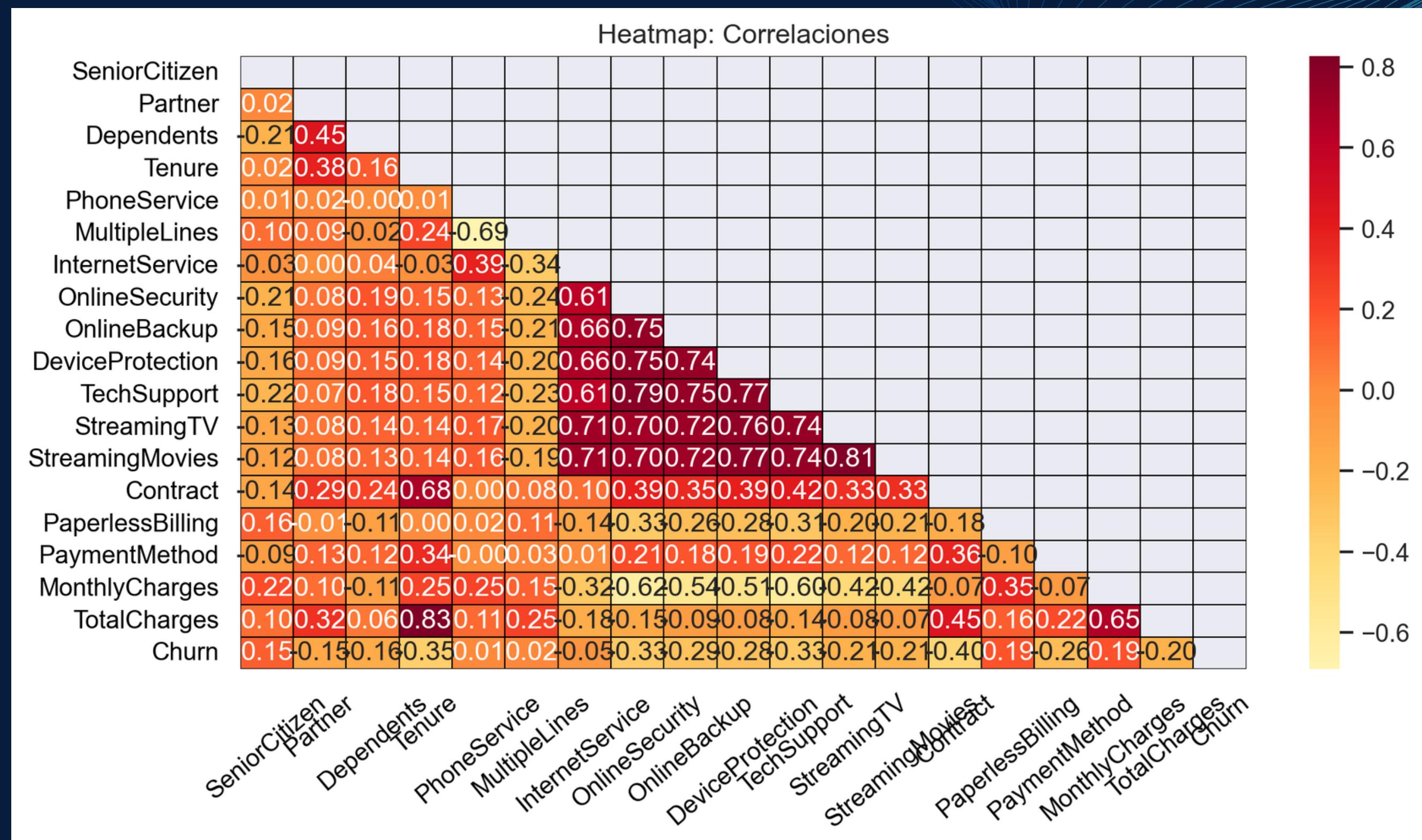
Según condición contractual (Tipo de contrato Vs Permanencia)



En este otro gráfico se puede ver que hay mayor concentración de abandonos en tramos de altos niveles de facturación. Por esta razón, podemos dar por sentado que el precio es realmente relevante en la decisión de los clientes.

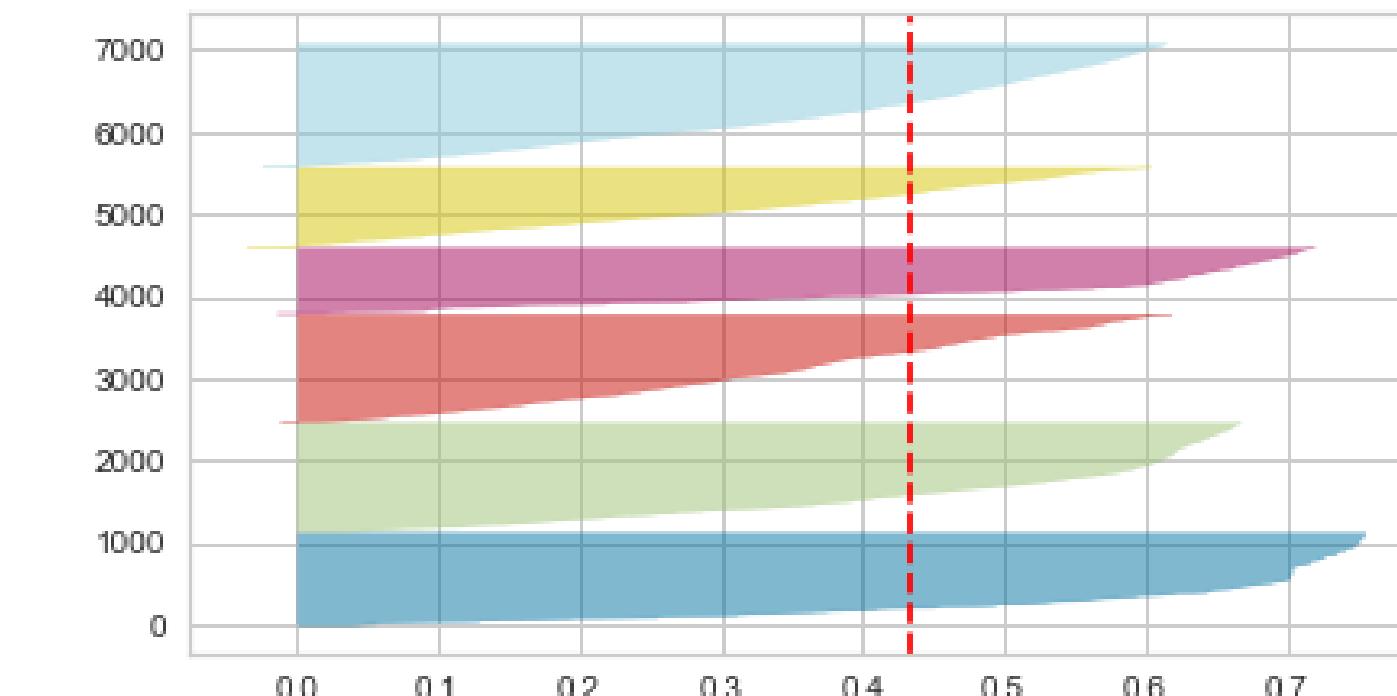
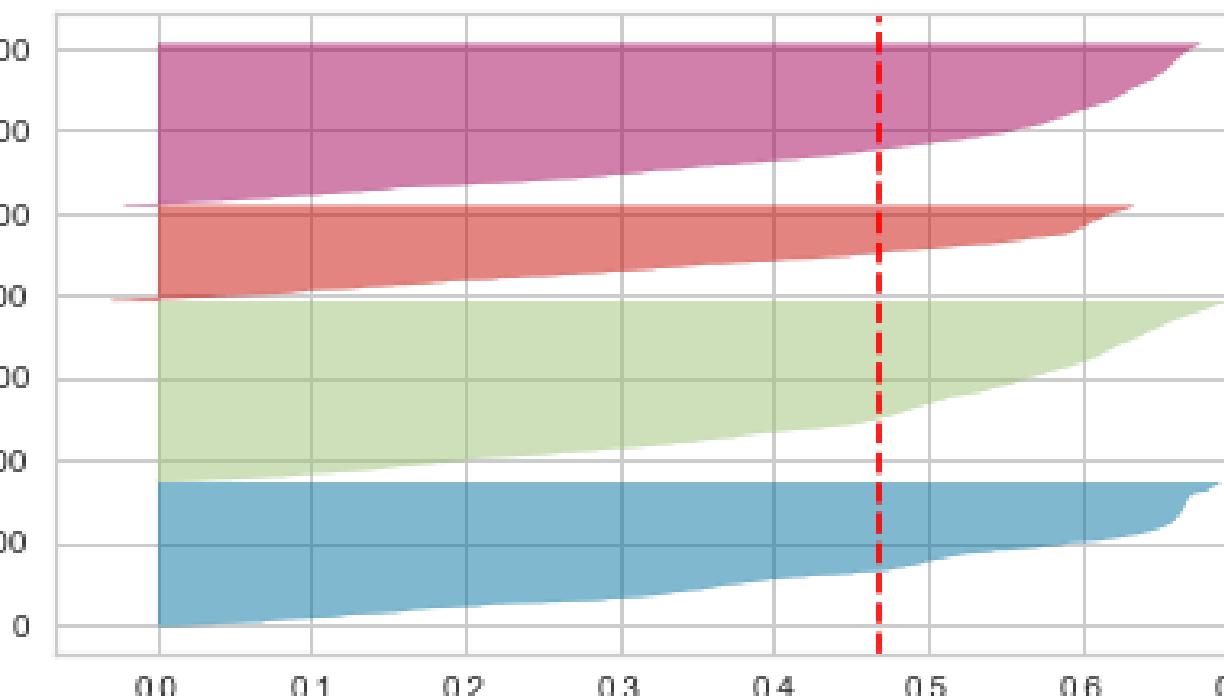
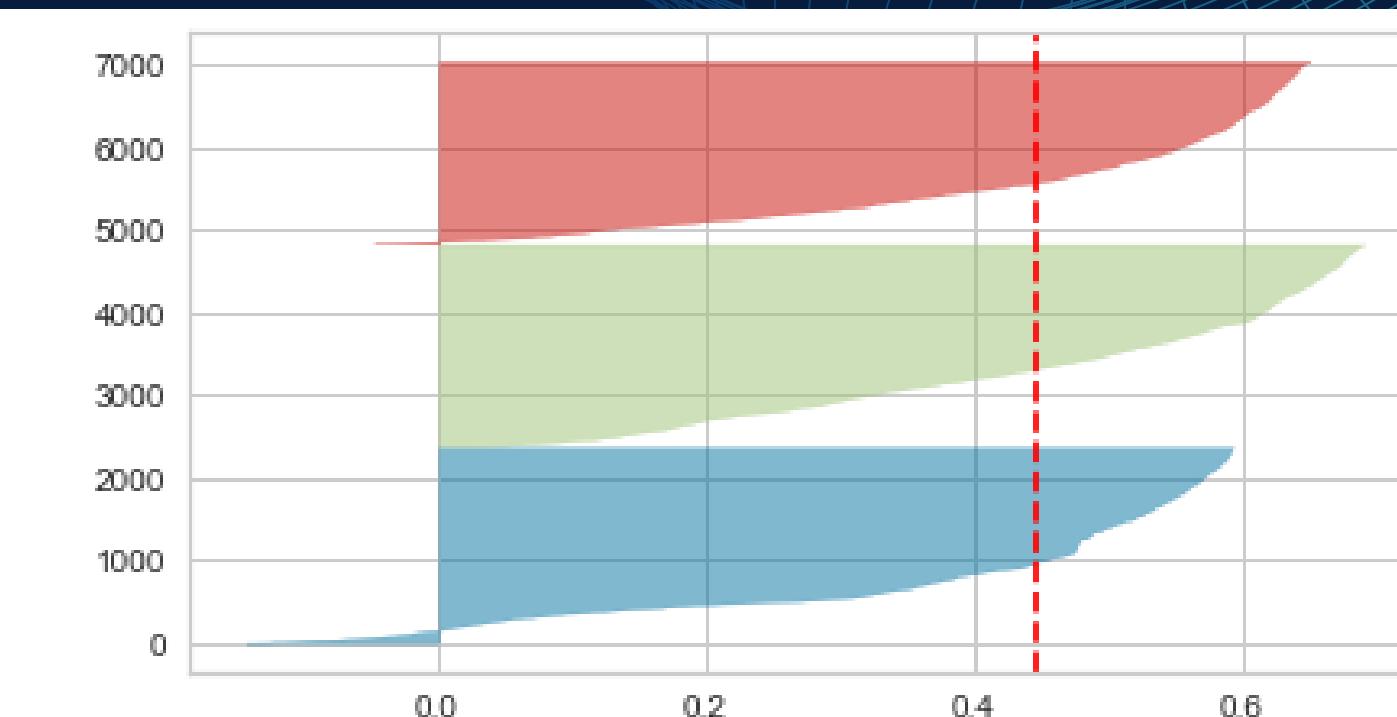
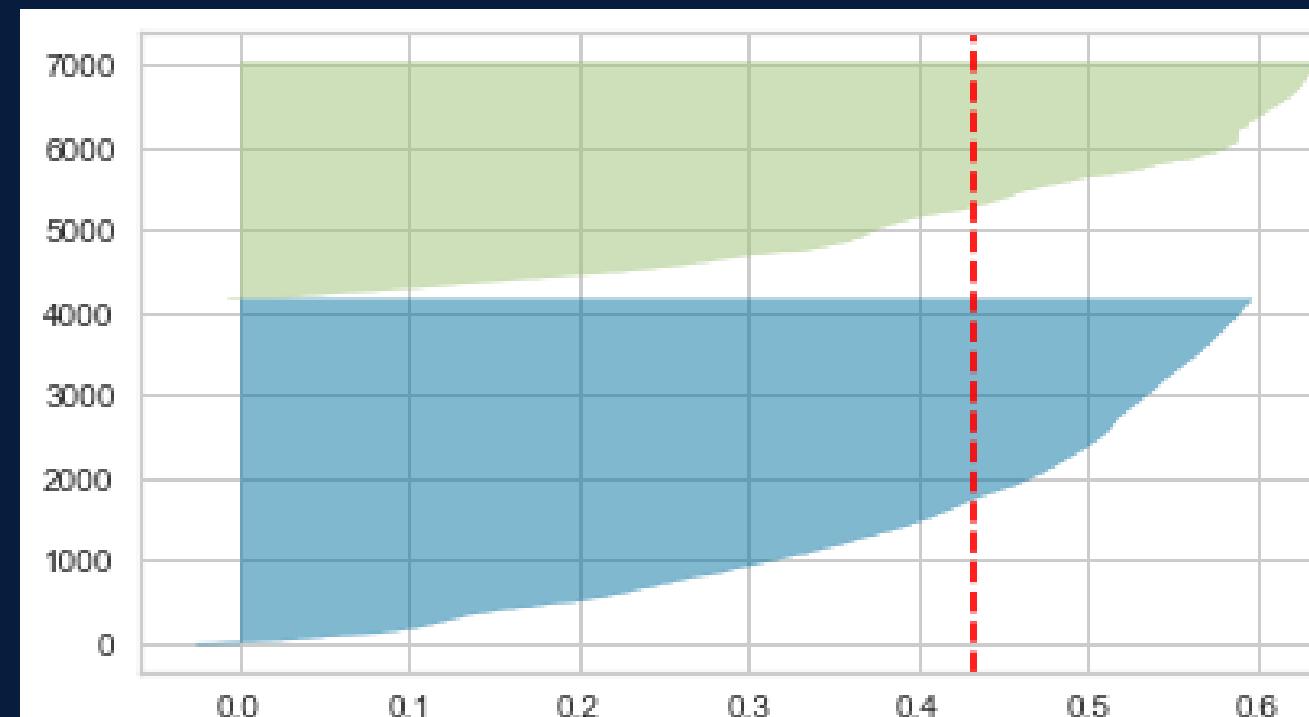
Limpieza de datos

A fines de eliminar aquella información que no es relevante, estudiamos la influencia de cada variables, así como sus correlaciones, logrando al final trabajar con un set de datos menor y por lo tanto mas performante desde el punto de visto del procesamiento y de la calidad de los resultados posteriores.



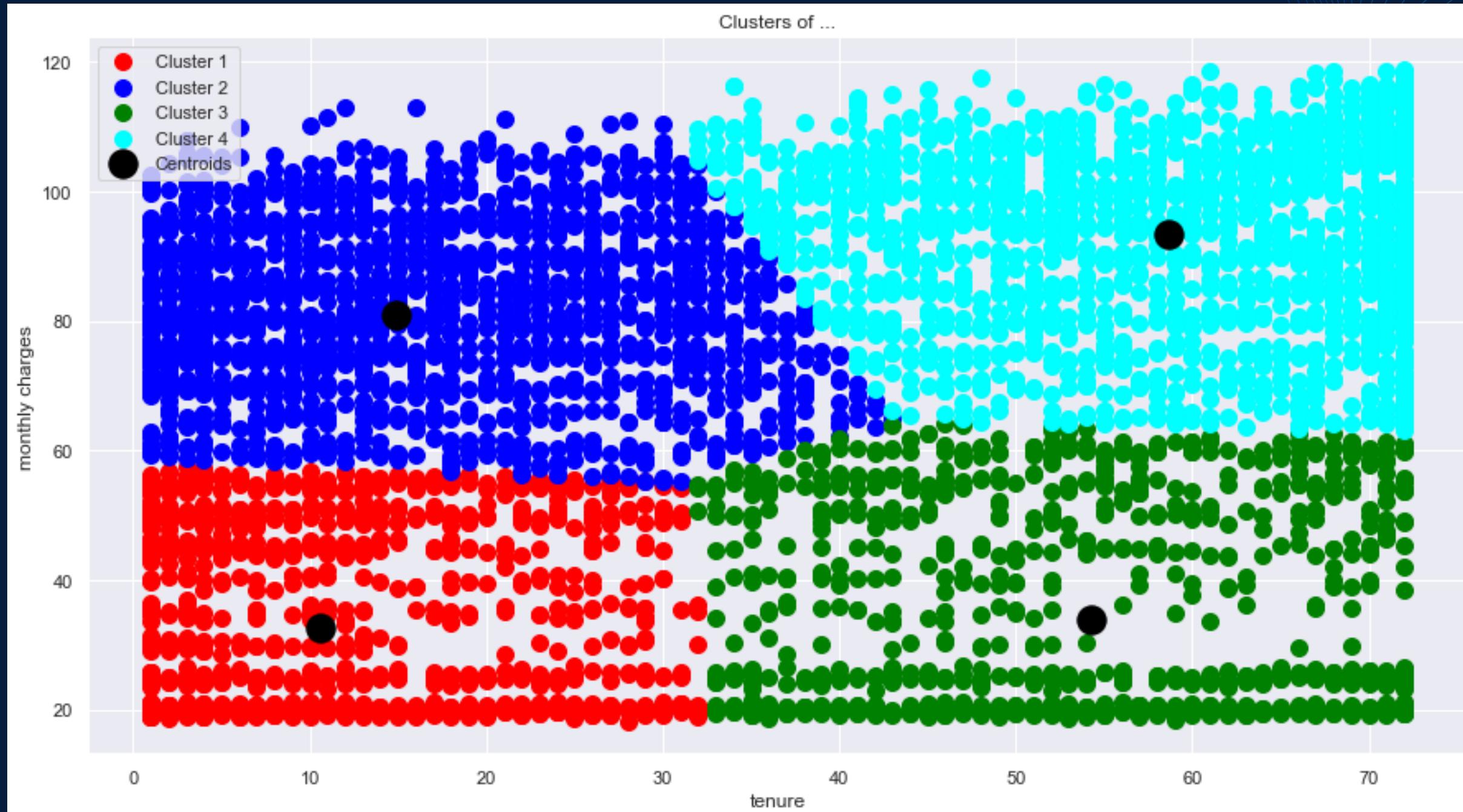
Segmentación de clientes

Para segmentar la clientela, hemos utilizado algoritmos de agrupamiento no supervisados, como el "K means clustering" y "agglomerative clustering". El N° de grupos óptimo según el método de Silohuete es 4, obteniendo así la mejor segmentación posible:



Segmentación de clientes

Resultado del agrupamiento:



Segun se observa, los clientes son agrupados en 4 segmentos:

- 1) Clientes con ALTO tiempo de permanencia y BAJA facturacion mensual,
- 2) Clientes con BAJO tiempo de permanencia y BAJA facturacion mensual,
- 3) Clientes con BAJO tiempo de permanencia y ALTA facturacion mensual,
- 4) Clientes con ALTO tiempo de permanencia y ALTA facturacion mensual.

Modelos de predicción de abandono

Para lograr predecir que clientes abandonarán la compañía, hemos trabajado con diferentes modelos de Machine Learning:



Decision Tree



K Near neighbors



**Support vector
machines**



**Algoritmos de
Bagging
& Boosting**

Metodología: Estudio de indicadores de Accuracy, Specificity ,Recall,Precision y F1 Score



Modelos de predicción de abandono

OPTIMIZACIÓN Y MEJORA DE LOS MODELOS:

para optimizar la calidad de predicción así como para asegurarnos de que los resultados son acertados, hemos utilizado distintos procedimientos y métodos:

- **PCA (Método para reducir la dimensionalidad y mejorar la calidad de predicción)**
- **Métodos de ensamble (Para reducir las predicciones erróneas)**
- **Ajuste de hiperparámetros (Para optimizar los resultados de cada modelo)**
- **Validación cruzada (Para aumentar la confiabilidad de cada modelo)**



Modelos de predicción de abandono

RESULTADOS

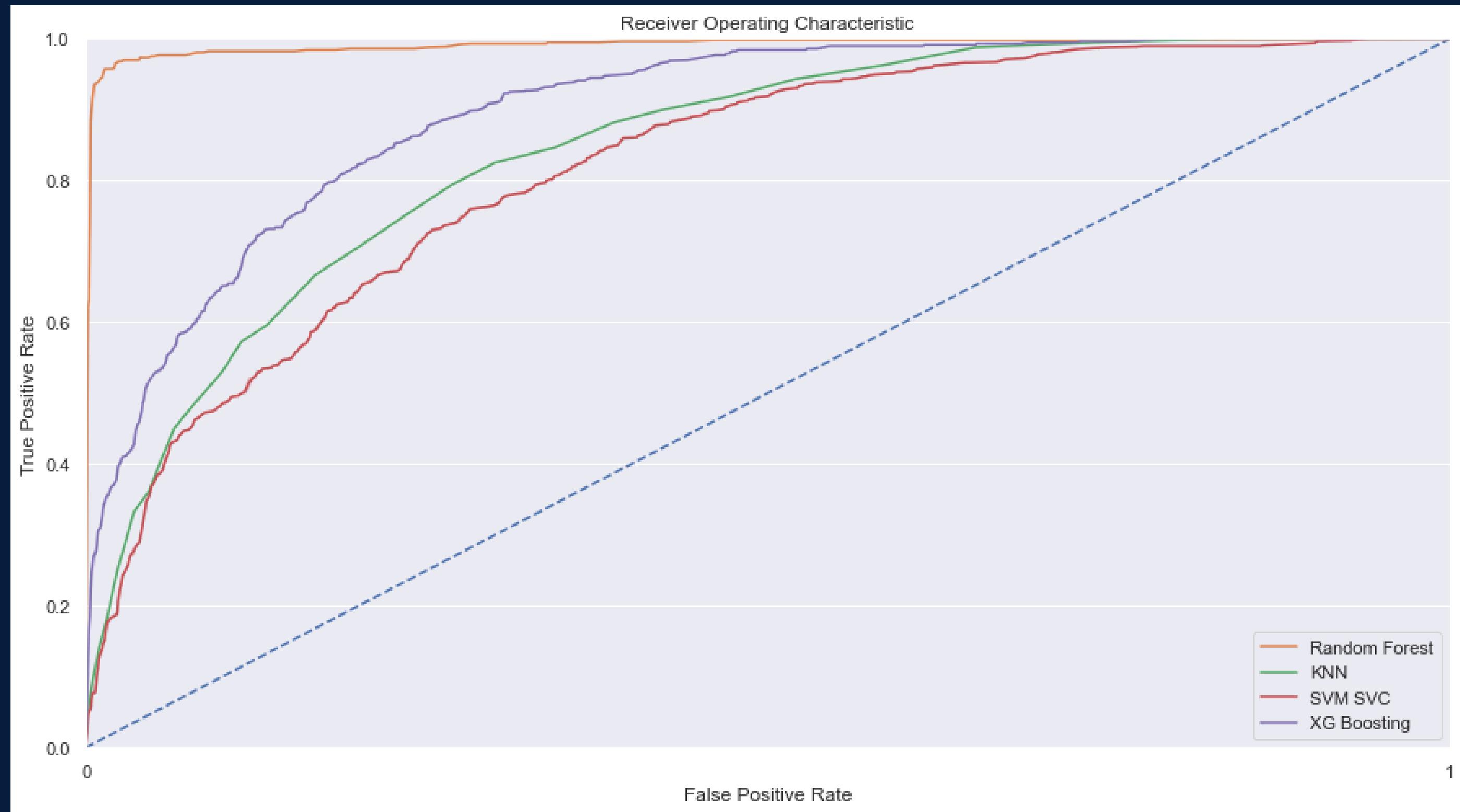
A continuación presentamos una comparativa de las métricas obtenidas entre los 4 modelos predominantes:

Métrica	Random forest	KNN	SVM SVC	XG Boosting
Accuracy (Exactitud)	81,39%	78,98%	78,41%	78,41%
Specificity	92,40%	92,46%	90,72%	90,33%
Recall (Sensibilidad)	50,80%	41,71%	44,39%	50,27%
Precision	70,90%	66,67%	63,36%	65,28%
F1 Score	59,19%	51,32%	52,20%	50,27%

Modelos de predicción de abandono

RESULTADOS

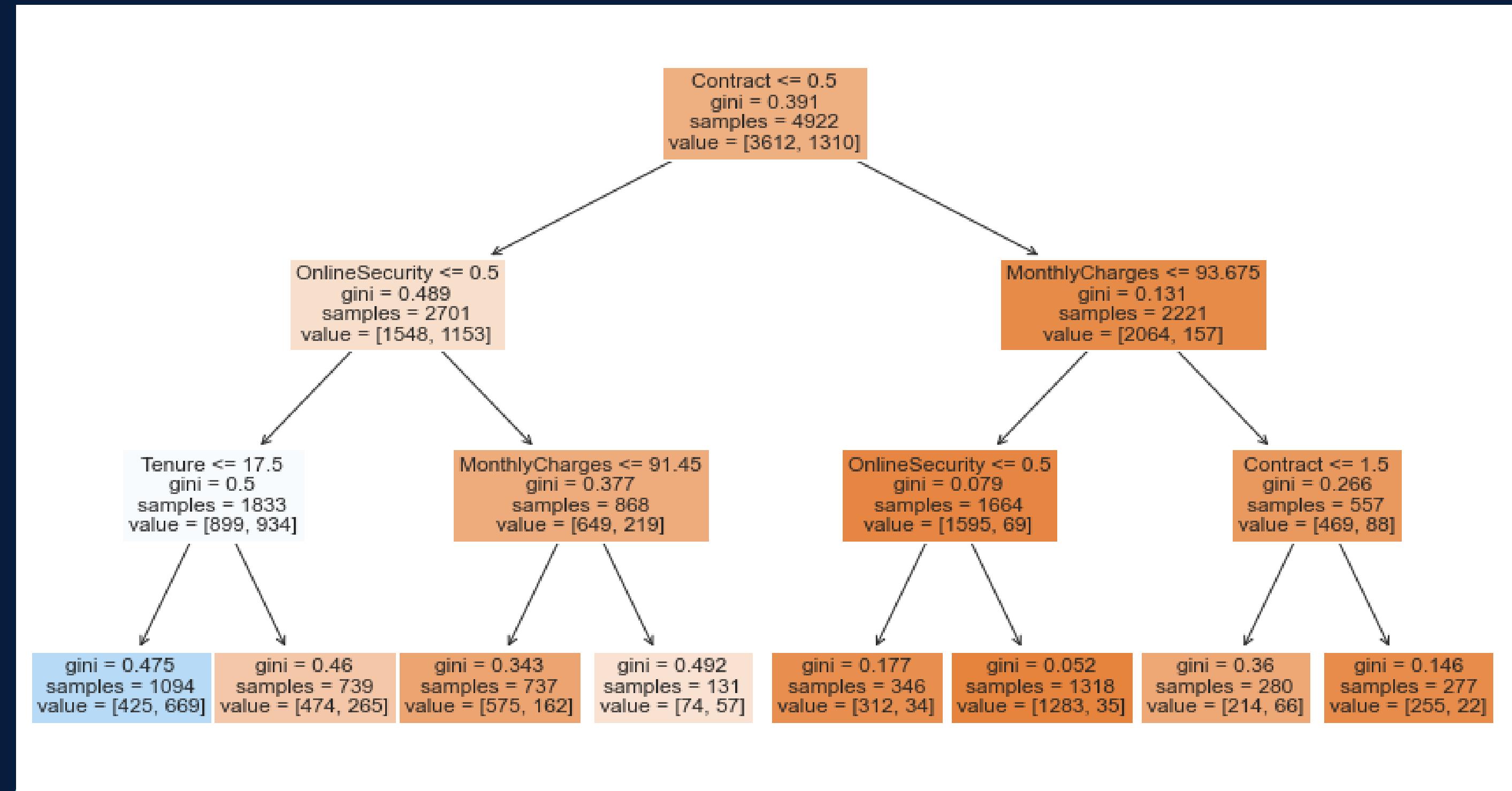
Para resumir los resultados obtenidos anteriormente, utilizaremos el indicador ROC el cual es un gráfico que ilustra la capacidad de diagnóstico de un sistema clasificador binario a medida que varía su umbral de discriminación.



Modelos de predicción de abandono

ALGORITMO SELECCIONADO:

El modelo elegido es "Random Forest", el cual es un método de "baGging" aplicado a arboles de decisión



Conclusiones:

En base a los datos proporcionados y el análisis se logró:

- Generar indicadores basados en características demográficas y de negocio que indican tendencias claras en cuanto a tipos de clientes que dejan la compañía. Esto permitirá a la compañía tomar medidas adecuadas a cada caso.
- Una Segmentación de 4 grupos bien definidos de clientes, que permiten elaborar estrategias de negocio.
- Un modelo de machine learning optimizado para predicción de abandonos, lo cual permite analizar riesgos y beneficios en el lanzamiento de campañas para retención de clientes.

Tabla de versionado:

DESCRIPCION	VERSION	FECHA
Primer entrega Proyecto Final	1	21/03/2022
Segunda entrega Proyecto Final	2	18/04/2022
Tercer Entrega Proyecto Final	3	04/05/2022
Proyecto Final	4	31/05/2022

