

Informe

EDA - Análisis Exploratorio de Datos | Empresa ficticia de telefonía

Isaac Mohamed Laaouaj Madrouni

1. Familiarización de los datos y objetivo

El objetivo es realizar un análisis exploratorio que permita definir la búsqueda de “churners” (clientes que cambian de compañía).

Tenemos tres conjuntos de datos de datos. Cada uno contiene información relevante de la empresa:

- Clientes: Contiene información sociodemográfica de los clientes a fecha 2023-10-31.
- Billing: contiene la facturación emitida a los clientes en todo el año 2023.
- Tenure: contiene la información relativa a los compromisos de permanencia durante el año 2023.

2. Limpieza de los datos

La fase de limpieza de datos se ha centrado en preservar la integridad de los datos, manteniendo sus patrones y tendencias originales. De esta forma nos aseguramos la coherencia y consistencia de los datos, lo cual es fundamental para llevar a cabo un análisis exploratorio preciso y efectivo.

Durante el proceso de limpieza, se ha trabajado para eliminar cualquier tipo de inconsistencia, error o valor atípico que pudiera distorsionar los resultados del análisis.

3. Perspectivas de negocios

3.1 Segmento principal de la empresa

Analizamos la columna `monthlycharges`, que representa los ingresos mensuales generados por los clientes.

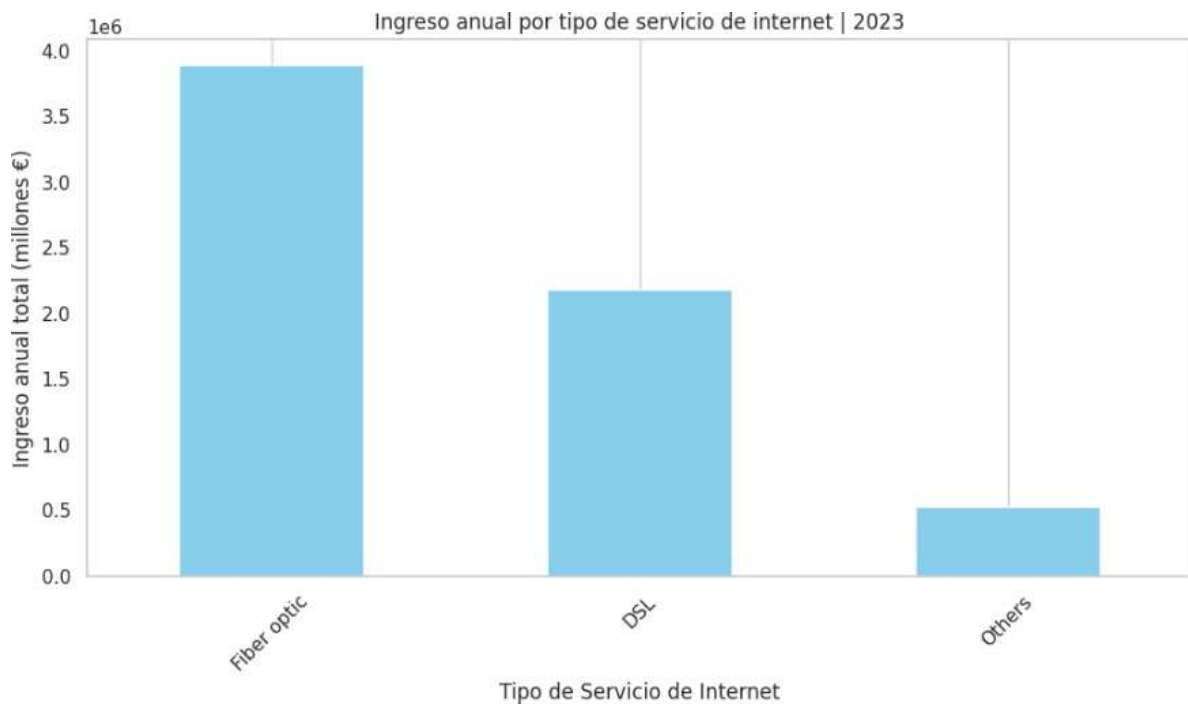
Calculamos la suma total de los ingresos mensuales agrupados por el tipo de servicio contratado por los clientes, para saber nuestra facturación anual (2023) de cada servicio. Esto nos brinda una visión detallada de cómo se distribuyen los ingresos en función de las diferentes opciones de servicios ofrecidos a los clientes. El resultado es el siguiente:

Fibra óptica	ADSL	Otros
3,891,497.12 €	2,183,183.80 €	534,687.35 €

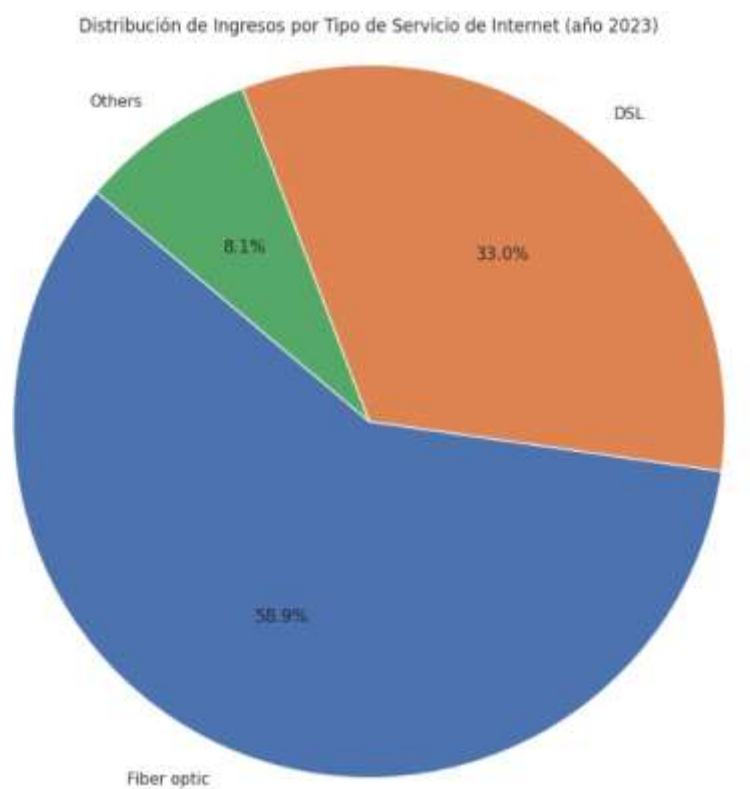
El sector con más peso en nuestra facturación del año 2023 es el de Fibra óptica representando una parte significativa de nuestros ingresos anuales.

Su contribución al total de ingresos ha sido notablemente superior en comparación con otros servicios de conectividad que ofrecemos.

Realizamos la gráfica de barras de los ingresos del año 2023 por tipo de servicio:



Lo que nos indica que nuestra distribución de los ingresos en función del tipo de servicio en porcentaje respecto al total de la facturación es el siguiente:



3.2 Clientes más propensos a irse de la compañía

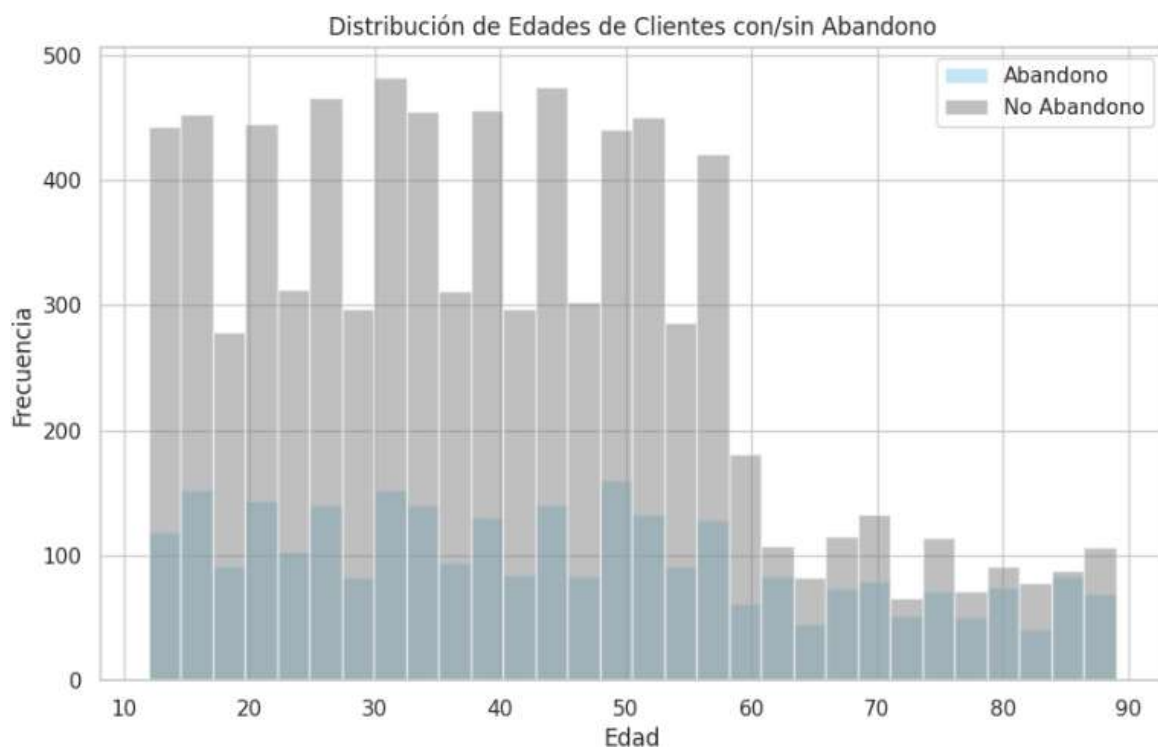
Para comprender y crear un perfil de los clientes más propensos a abandonar el servicio, utilizaremos una combinación de datos geográficos, demográficos y de negocio. Estos datos nos permitirán identificar y caracterizar a los clientes que tienden a abandonar el servicio de la compañía.

3.2.1 Demográficos:

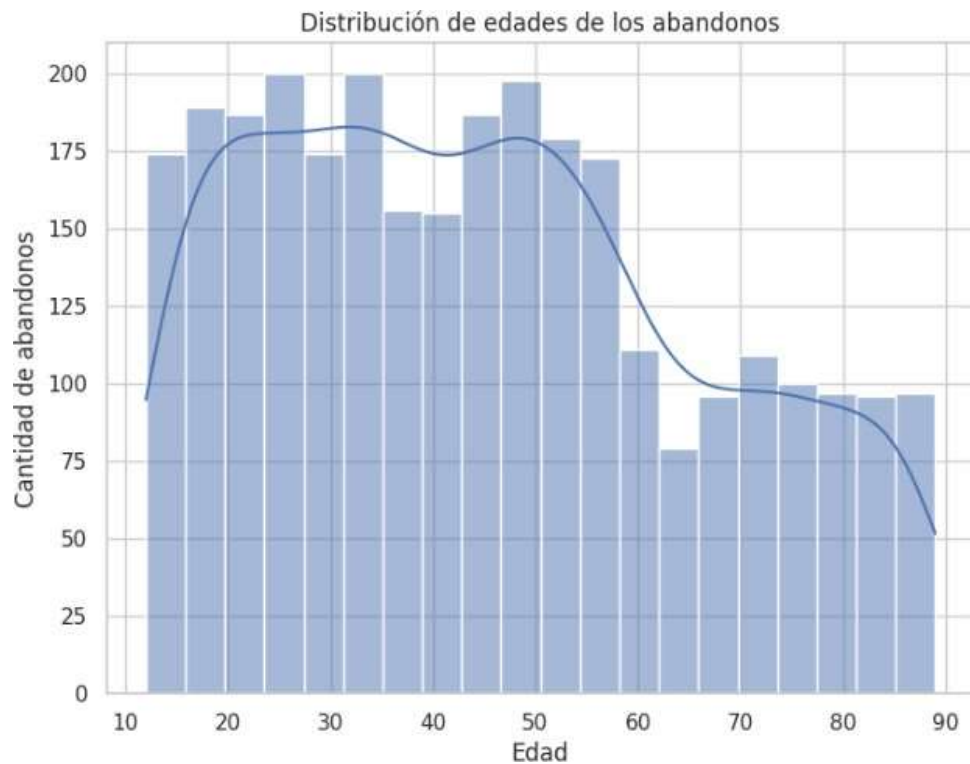
Los aspectos demográficos, como la edad, el sexo, el estado civil y el número de hijos, desempeñan un papel crucial en la comprensión de los clientes propensos a abandonar el servicio. Podemos examinar si ciertos grupos demográficos, como clientes más jóvenes o mayores, hombres o mujeres, solteros o casados, o aquellos con cierta cantidad de hijos, tienen una mayor tendencia al abandono.

Edad

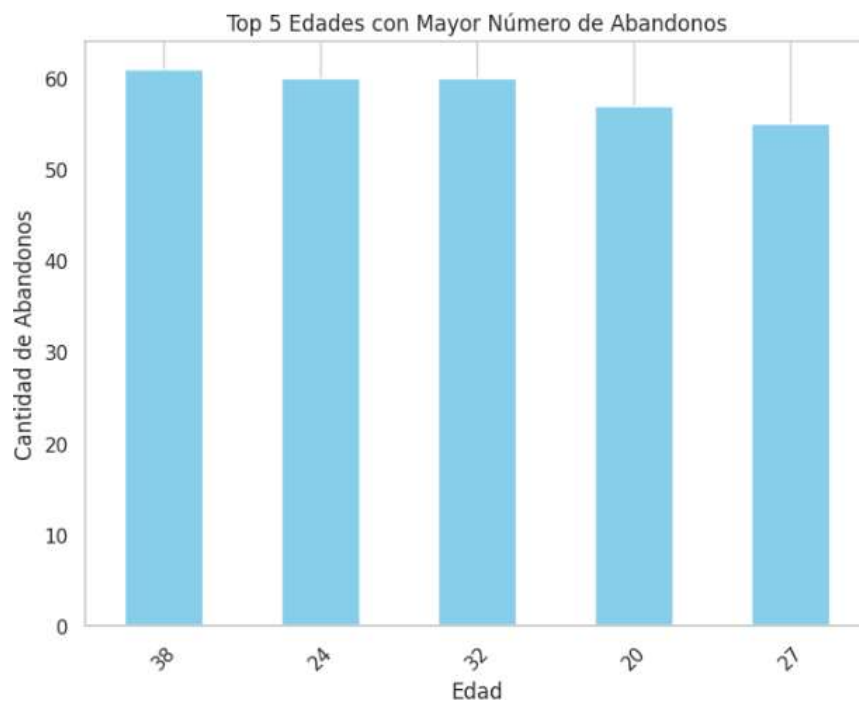
Podemos observar la distribución de la edad, diferenciando entre los clientes que han abandonado nuestro servicio y aquellos que siguen siendo clientes activos. Se aprecia una relativa similitud en la distribución por edades entre ambos grupos:



A edades más avanzadas menos abandono pronunciamos, concluyendo que los abandonos ocurren más en grupos de edades adultas:



Las 5 edades con más abandono en nuestra empresa constituyen el rango de entre 20 hasta 38 años.



Género

Analizamos las tres categorías de los datos referente al género (Hombre, Mujer y otros) mediante una tabla de contingencias tanto numérico como percentual:

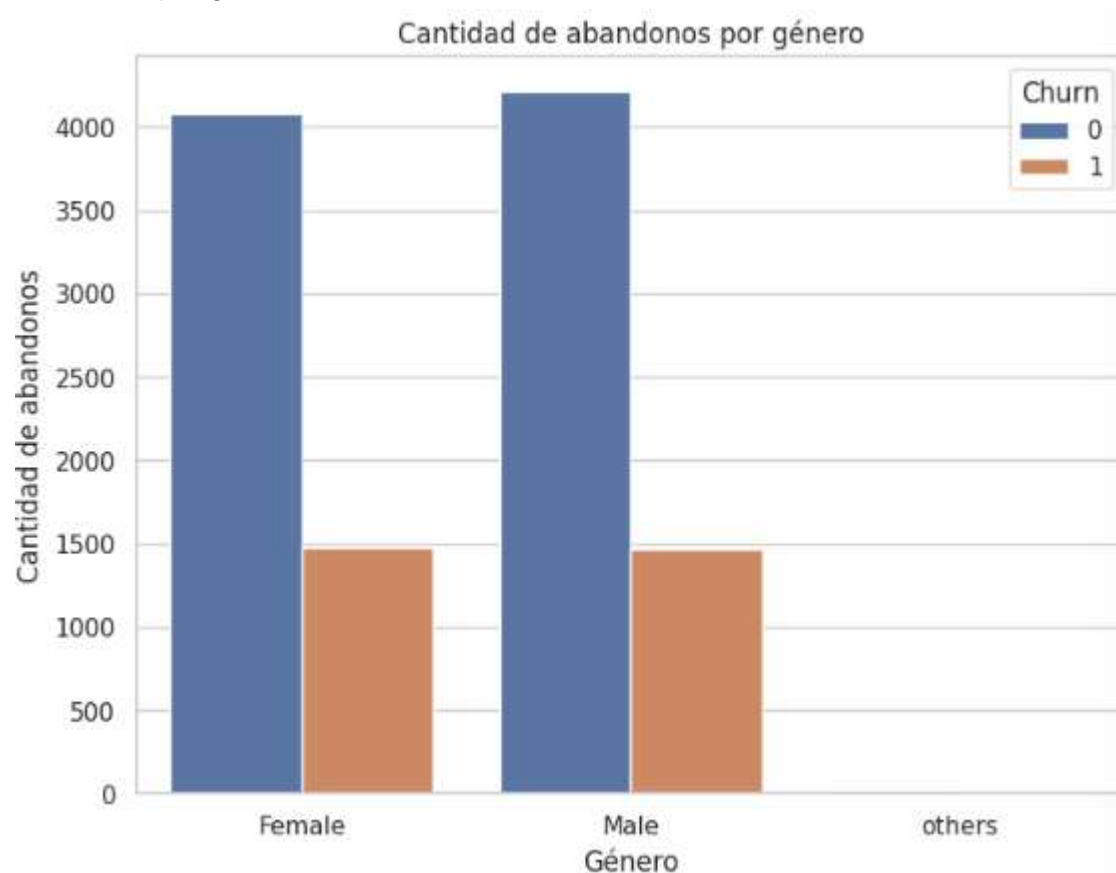
	Clientes que han abandonado	Clientes que no han abandonado	Todos
Femenino	4076	1476	5552
Masculino	4211	1470	2681
Otros	20	11	31
Suma total	8307	2957	11264

	Clientes que han abandonado	Clientes que no han abandonado
Femenino	73.41%	26.58%
Masculino	74.12%	25.87%
Otros	64.51%	35.48%

Observamos que porcentualmente tanto el género masculino como femenino mantienen una igualdad en el abandono y por tanto en el no abandono.

Resulta notable que los clientes que no han especificado su género presentan una tasa de abandono del 64.51%. Este porcentaje es significativamente alto considerando su pequeña representación dentro del total de clientes. Esto sugiere que la empresa debería prestar atención al tratamiento o consideración hacia aquellos colectivos que no se identifican como masculinos o femeninos.

El siguiente diagrama de barras muestra una representación general de la cantidad de abandonos por género:

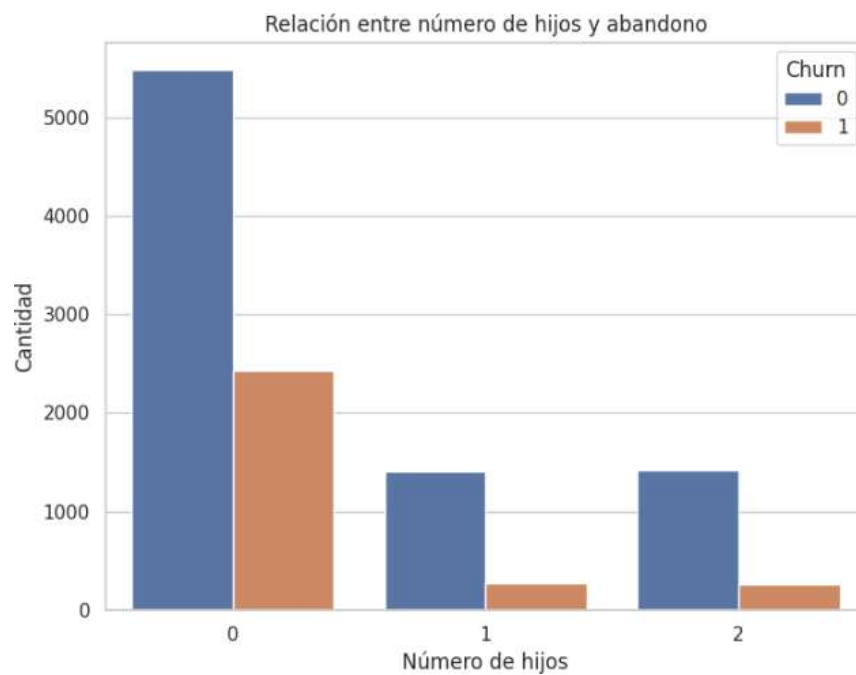


Número de hijos

Los clientes que han abandonado son más propensos a no tener ningún hijo a cargo:

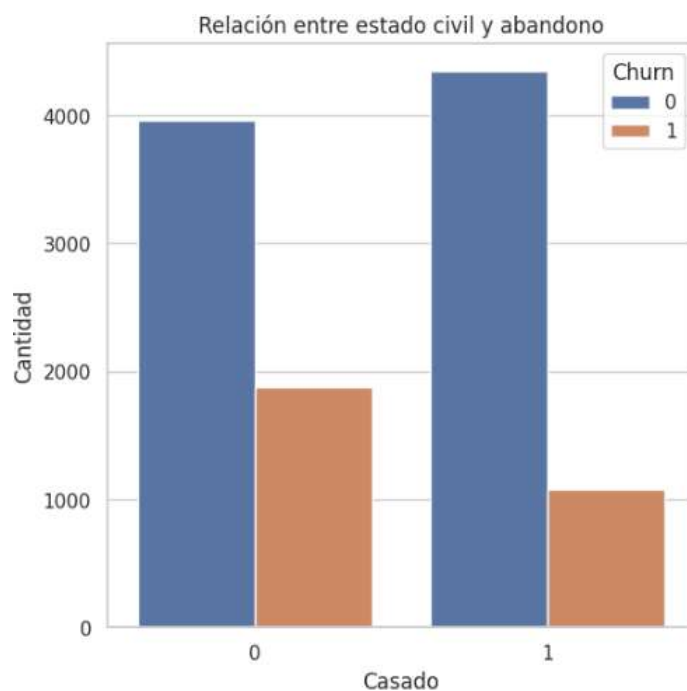
Cantidad de hijos	Tasa de abandono en fundación de la cantidad de hijos
0	30.60%
1	16.3%
2	15.53%

Podemos visualizarlo mejor mediante el siguiente diagrama de barras, aunque es totalmente coherente que el quien tiene mayor tasa de abandono son los que no tienen hijos ya que son los clientes que más hay:



Estado civil

Visualizamos si el estado civil influye en los abandonos de parte de los clientes:



Observando el diagrama si podemos ver una cierta relación en que los solteros o divorciados han abandonado más en comparación con los clientes que se encuentran casados:

Estado civil	% Abandono
No casado	32.18%
Casado	19.85%

Por lo tanto, los clientes que no se encuentran casados tienden más a abandonar nuestros servicios.

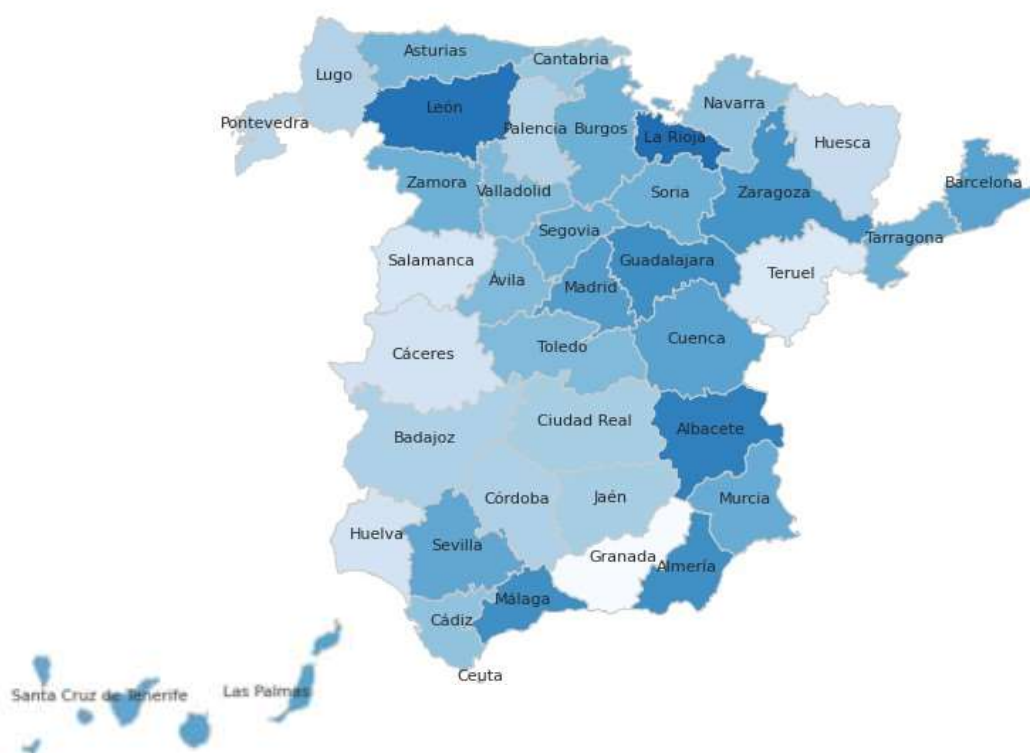
Conclusión perfil demográfico del cliente que abandona

Se trata de un cliente de entre 20 y 38 años, indiferentemente del género (aunque en especial énfasis a los otros géneros que nos sean femenino ni masculino), sin hijos y con un estado civil diferente a casado (soltero, viudo o viuda, divorciado, etc.)

3.2.2 Geográficos:

El análisis geográfico, específicamente la provincia donde residen los clientes, puede ser un factor relevante para identificar patrones de abandono en áreas geográficas específicas. Podríamos observar si existen provincias donde el abandono es más frecuente en comparación con otras.

Cientes con mayor abandono por provincia en España



Como más pronunciado sea el azul, más clientes han abandonado el año 2023, y como menos azul es menor ha sido el número de abandonos.

Las provincias con más abandono de clientes han sido:

Provincia	Abandonos
Ceuta	85
La Rioja	73
León	72
Albacete	70
Guadalajara	67

Conclusión cliente que abandona según su geografía:

Cliente posicionado geográficamente en Ceuta, La Rioja, León, Albacete o Guadalajara

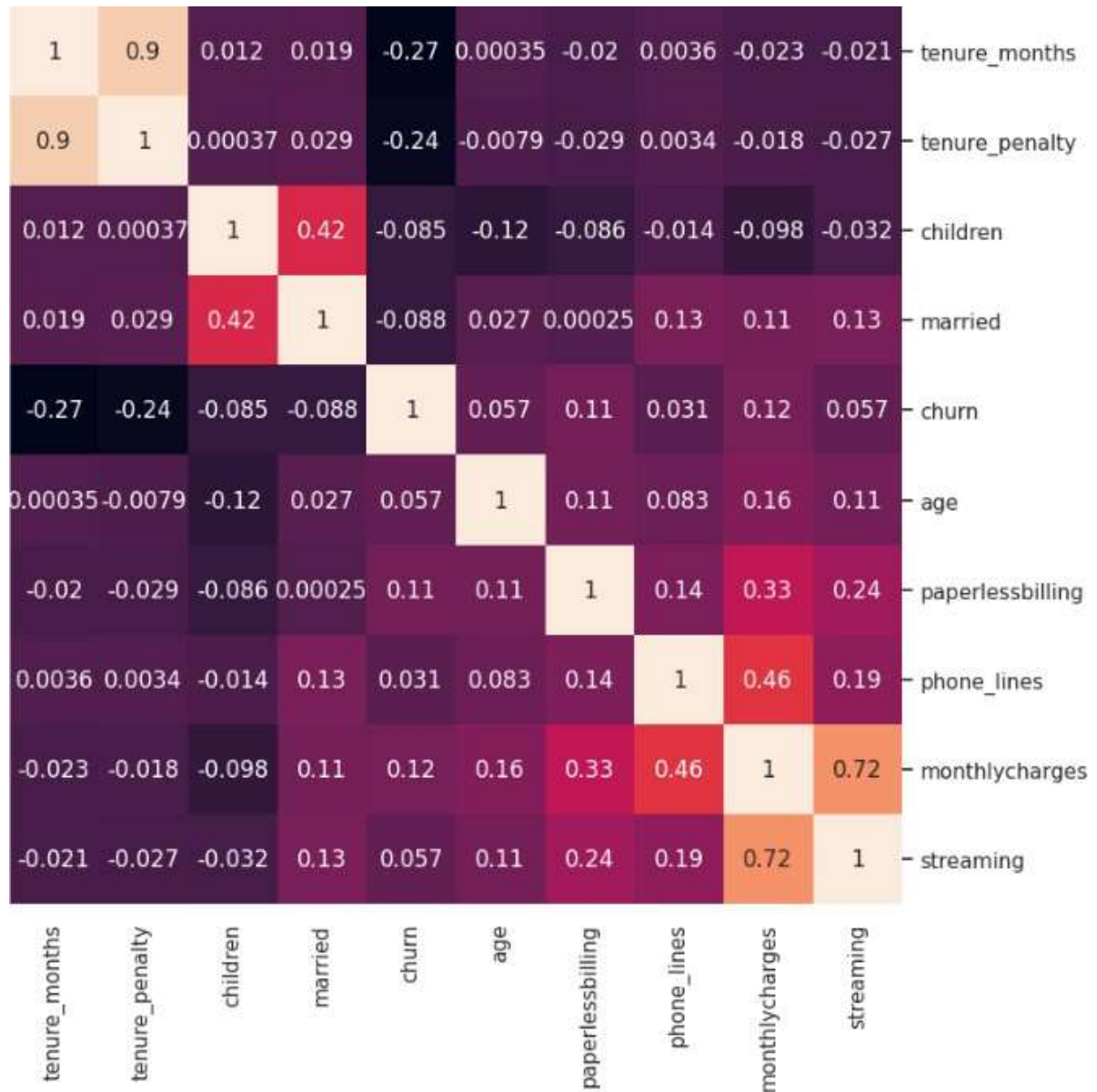
3.2.3 Datos de negocio:

El análisis de los datos relacionados con el tipo de servicio más propenso al abandono ofrece información valiosa. En este caso, sabemos que el servicio de ****Fiber optic**** es más susceptible a abandonos. Por lo tanto, considerar este servicio como un factor clave en el análisis nos permite enfocar estrategias de retención específicas para este segmento.

Servicio	Abandonos
Fibra óptica	758
ADLS	257
Otros	69

3.3 ¿Por qué se van los clientes? ¿Cómo podríamos frenar su marcha?

Visualizamos las posibles correlaciones que hay entre los datos que disponemos:



Observamos que en la columna "churn":

- Existe una correlación negativa (muy leve) con las variables "tenure_months" y "tenure_penalty".
- Se evidencia una correlación positiva (muy leve) con "paperlessbilling" y "monthlycharges".

Estos hallazgos sugieren que la duración de la permanencia en el servicio y las penalizaciones asociadas están relacionadas de forma negativa con los abandonos.

Por otro lado, la elección de facturación electrónica y los cargos mensuales muestran una relación positiva con los abandonos. A fin de comprender esta relación en mayor profundidad, examinaremos estos aspectos con mayor detenimiento.

Análisis de la duración de permanencia:

Se observa que los clientes que abandonan el servicio muestran una menor cantidad de meses de permanencia en comparación con aquellos que continúan siendo clientes activos:

	count	mean	std	min	25%	50%	75%	max
churn								
0	3104.0	8.512564	3.449909	1.0	6.0	8.0	12.0	14.0
1	1084.0	2.655904	1.535860	1.0	2.0	3.0	4.0	14.0

Análisis de penalización por permanencia:

Observamos que los clientes que decidieron abandonar nuestros servicios presentan una penalización monetaria menor en comparación con aquellos que optaron por continuar con nosotros.

	count	mean	std	min	25%	50%	75%	max
churn								
0	3104.0	153.062870	74.660085	14.67	91.9475	144.23	202.7650	362.30
1	1084.0	47.936891	31.709688	10.13	23.9200	42.26	63.9575	299.21

Evaluación de cargos mensuales:

Observamos que, en promedio, los clientes que abandonaron en 2023 tuvieron un costo mensual más alto en sus facturas en comparación con los clientes que actualmente siguen siendo vigentes.

Esta diferencia en el importe medio de los cargos mensuales podría ser un factor relevante para comprender el abandono de clientes durante el año 2023.

	count	mean	std	min	25%	50%	75%	max
churn								
0	3104.0	61.024662	31.021594	18.25	25.1000	64.125	88.6125	118.65
1	1084.0	74.765913	24.931341	18.85	56.4125	79.900	94.4500	118.35

¿Cómo podríamos frenar la marcha de futuros clientes?

En base a la información analizada podemos sugerir las siguientes acciones:

- **Proporcionar incentivos para la permanencia** para retener a los futuros clientes: Esto implica ofrecer beneficios a largo plazo a aquellos que decidan quedarse con el servicio durante más tiempo. Estos beneficios pueden tomar la forma de descuentos, promociones, ofertas especiales u otros incentivos atractivos.
- **Otorgar un descuento considerable en la factura mensual a cambio de un compromiso prolongado con el servicio.** Un aumento en la penalización por abandono, especialmente dirigido a los clientes que han experimentado un incremento progresivo en sus facturas mensuales. Este tipo de estrategia no solo fomenta la lealtad del cliente, sino que también ofrece un estímulo adicional para mantener una relación a largo plazo con la empresa.

3.4 Campaña de retención

Como ya hemos visto anteriormente, el [segmento con más clientes que abandonan los servicios es la Fibra Óptica](#). Necesitamos saber el mes con más clientes que abandonan nuestros servicios:

Meses	Clientes que abandonan en total
Marzo	375
Abril	423
Mayo	419
Junio	425
Julio	437
Agosto	428
Septiembre	450

Como el segmento de clientes más propenso al abandono es el que utiliza Fibra óptica, sugerimos dirigir la estrategia de retención específicamente a este grupo.

La campaña consistirá en la aplicación de un atractivo descuento del 10% exclusivamente para aquellos clientes que tengan contratado el servicio de Fibra óptica. Se planea lanzar esta campaña un mes antes del período con mayor número de abandonos.

La duración prevista para esta campaña será de un mes, iniciando el 01/08 y concluyendo el 01/09. En cuanto al impacto en la facturación, se prevé lo siguiente:

El impacto en la facturación debido al descuento del 10% sobre la mensualidad de Fibra Óptica durante la campaña (agosto).	84,146.78 €
---	--------------------

Si la campaña es efectiva (sin tener en cuenta ningún margen de error) la empresa evitará una pérdida de **35,311.43 €**, que no se asemeja a las pérdidas ocasionadas por la promoción pero garantiza una fidelidad del cliente para los próximos meses.

4. Conclusiones

Facturación Anual por Servicio Se ha identificado la distribución de ingresos anuales (en euros) por cada tipo de servicio ofrecido a los clientes. Los ingresos más altos provienen del servicio de Fibra Óptica (3,891,497.12 €), seguido por ADSL (2,183,183.80 €) y Otros (534,687.35 €). Esto proporciona una visión detallada de cómo se distribuyen los ingresos entre las diferentes opciones de servicios.

Perfil del Cliente Propenso al Abandono: Se ha delineado un perfil de cliente más propenso a abandonar el servicio. Se trata de individuos de entre 20 y 38 años, sin hijos, con un estado civil distinto a casado, y ubicados geográficamente en Ceuta, La Rioja, León, Albacete o Guadalajara. Además, se hace énfasis en aquellos clientes cuyo género no es ni masculino ni femenino.

Análisis de Correlación: Se han identificado correlaciones leves pero significativas. Existe una correlación negativa leve con las variables "tenure_months" y "tenure_penalty", y una correlación positiva también leve con "paperlessbilling" y "monthlycharges".

Acciones de Retención de Clientes: Basándose en el análisis, se sugieren estrategias específicas para frenar la pérdida de futuros clientes. Esto incluye proporcionar incentivos para la permanencia, ofreciendo descuentos o beneficios a largo plazo a aquellos que decidan quedarse más tiempo con el servicio.

Impacto de la Campaña de Retención: Se ha calculado el impacto en la facturación debido a la campaña de descuento del 10% sobre la mensualidad de Fibra Óptica durante el mes de agosto, resultando en 84,146.78 €. Se plantea que, si la campaña es efectiva, la empresa evitará una pérdida de 35,311.43 €, lo que garantizaría la fidelidad del cliente para los próximos meses, a pesar de no igualar las pérdidas ocasionadas por la promoción inicial.