

Informe de Análisis Exploratorio de Datos

En el presente informe, llevaremos a cabo un análisis exploratorio exhaustivo de los conjuntos de datos proporcionados. A continuación, proporcionaremos una breve descripción de cada conjunto de datos, este conjunto de datos se ha limpiado previamente para eliminar datos inconsistentes o faltantes, lo que nos permite centrarnos directamente en el análisis exploratorio de los datos limpios, ofreciendo así una visión inicial de los aspectos en los que nos enfocaremos durante nuestro trabajo.

- **Churn for Bank Customers:**

Este conjunto de datos aborda la tasa de abandono (churn) de clientes bancarios. Contiene información sobre diversos atributos de los clientes, como edad, género, saldo de la cuenta, historial de transacciones, entre otros.

- **Bank Marketing Dataset:**

Este conjunto de datos se centra en las campañas de marketing bancario posicionado en Portugal. Incluye información sobre los resultados de las campañas de marketing dirigidas a los clientes del banco, como detalles de contacto, características socioeconómicas, resultados de la campaña anterior y si el cliente suscribió o no a un depósito a plazo.

- **Bank Term Deposit Predictions:**

El conjunto de datos "Bank Term Deposit Predictions" se centra en la suscripción a un depósito a plazo en un banco de Portugal. Contiene una variedad de características del cliente, como edad, ocupación, estado civil, nivel educativo y saldo promedio anual, así como detalles sobre campañas de marketing anteriores y resultados. A diferencia del conjunto de datos "Bank Marketing Dataset", que se enfoca en los resultados de campañas de marketing específicas, este conjunto incluye más información detallada sobre los clientes y sobre el estado social y económico del país, lo que permite un análisis más exhaustivo de los factores que influyen en las decisiones de suscripción a depósitos a plazo.

Objetivo del Análisis

El propósito fundamental de este análisis es profundizar en la comprensión de los datos disponibles, destacando patrones, tendencias y posibles anomalías que puedan ser de relevancia para la toma de decisiones en el contexto bancario. Para lograrlo, se lleva a cabo un análisis exhaustivo que prioriza la identificación de datos clave, omitiendo aquellos con poca relevancia para obtener una visión más clara y significativa.

Este estudio aborda una exploración detallada de las características demográficas de los clientes, así como de los factores vinculados a las campañas de marketing del banco. El objetivo es proporcionar resultados sólidos y dinámicos que permitan una observación efectiva de los datos a través de la herramienta Tableau.

Churn for Bank Customers

Este conjunto de datos se distingue por contener información altamente personalizada sobre los clientes del banco. El objetivo de analizar este conjunto de datos es comprender y visualizar los factores que pueden influir en la retención de clientes. Además, proporciona insights sobre el perfil demográfico de los clientes del banco y ayuda a definir estrategias para la retención y protección de los clientes existentes.

Campos:

RowNumber: Corresponde al número de registro (fila) y no tiene efecto en el resultado.

CustomerId: Contiene valores aleatorios y no afecta la salida del cliente del banco.

Surname: El apellido del cliente no tiene impacto en su decisión de dejar el banco.

CreditScore: Puede tener un efecto en la pérdida de clientes, ya que un cliente con un puntaje de crédito más alto es menos probable que deje el banco. (suposición)

Geography: La ubicación del cliente puede afectar su decisión de dejar el banco.

Gender: Es interesante explorar si el género juega un papel en la salida del cliente del banco.

Age: Esto es ciertamente relevante, ya que los clientes mayores son menos propensos a dejar su banco que los más jóvenes. (suposición)

Tenure: Se refiere al número de años que el cliente ha sido cliente del banco. Normalmente, los clientes más antiguos son más leales y menos propensos a dejar el banco.

Balance: También es un muy buen indicador de la pérdida de clientes, ya que las personas con un saldo más alto en sus cuentas son menos propensas a dejar el banco en comparación con aquellas con saldos más bajos. (suposición)

NumOfProducts: Se refiere al número de productos que un cliente ha adquirido a través del banco.

HasCrCard: Indica si un cliente tiene o no una tarjeta de crédito. Esta columna también es relevante, ya que las personas con una tarjeta de crédito son menos propensas a dejar el banco. (suposición)

IsActiveMember: Los clientes activos son menos propensos a dejar el banco. (suposición)

EstimatedSalary: Al igual que con el saldo, las personas con salarios más bajos son más propensas a dejar el banco en comparación con aquellas con salarios más altos.
(suposición)

Exited: Indica si el cliente dejó o no el banco.

Esta descripción fue proporcionada por el usuario que compartió el conjunto de datos. Sin embargo, es importante tener en cuenta que algunas de estas descripciones pueden ser suposiciones y no se ha verificado su veracidad. Por lo tanto, uno de los primeros pasos en el análisis será verificar la precisión de estas descripciones y determinar si corresponden con la realidad de los datos.

Análisis:

El conjunto de datos contiene exactamente 10000 registros, no tienen ningún dato nulo y las columnas

	CreditScore	Age	Tenure	Balance	NumOfProducts	EstimatedSalary
mean	650.53	38.92	5.01	76485.89	1.53	100090.24
std	96.65	10.49	2.89	62397.41	0.58	57510.49
min	350.00	18.00	0.00	0.00	1.00	11.58
25%	584.00	32.00	3.00	0.00	1.00	51002.11
50%	652.00	37.00	5.00	97198.54	1.00	100193.92
75%	718.00	44.00	7.00	127644.24	2.00	149388.25
max	850.00	92.00	10.00	250898.09	4.00	199992.48

Esta información se obtiene de los datos numéricos del conjunto de datos junto con Python, concretamente pandas, usando la función “describe()”, para entender mejor los datos que nos salieron veamos la descripción de cada categoría:

- mean: la media de los valores en la columna.
- std: la desviación estándar de los valores en la columna.
- min: el valor mínimo en la columna.
- 25%: el percentil 25, también conocido como el primer cuartil indica el valor por debajo del cual se encuentra el 25% de los datos.
- 50%: el percentil 50, también conocido como la mediana, es el valor que se encuentra en el centro de la distribución de los datos cuando están ordenados de menor a mayor.
- 75%: el percentil 75, también conocido como el tercer cuartil, indica el valor por debajo del cual se encuentra el 75% de los datos.
- max: el valor máximo en la columna.

Con la información proporcionada, podemos entender mejor los valores numéricos de nuestro conjunto de datos. Observamos tres valores percentiles que representan los

cuartiles. Al analizarlos detenidamente, podemos detectar valores extraños, outliers o tendencias en los datos.

- **CreditScore:** Esta variable varía de 300 a 850. Al observar los percentiles, notamos una distribución normal. La desviación estándar indica que la variación en este campo no es significativa, lo que sugiere que no deberíamos tener problemas.
- **Age:** En esta columna, encontramos más dificultades. Si observamos los cuartiles, donde el 75% es 44, muy por debajo del máximo de 92, podemos ver que la distribución de datos está centrada en personas entre 32 y 44 años. Aunque el banco puede tener usuarios de todas las edades, si la mayoría son jóvenes, esto podría impactar en nuestro análisis.
- **Tenure:** Representa los años que el cliente ha estado en el banco. Con una media de 5, observamos una fidelidad potencialmente alta de los usuarios, considerando que el máximo es 10. Sin embargo, esto dependerá de otros factores, que analizaremos posteriormente.
- **Balance:** Encontramos valores de 0 (posiblemente nuevos usuarios o cuentas inactivas) y un valor atípico máximo que difiere significativamente de nuestros percentiles.
- **NumOfProducts:** Sabiendo que el máximo es de 4 productos y que la media está cerca de 2, vemos que la mayoría de los usuarios tienen solo un producto del banco. Esto se confirma al observar que la media es 1 y el 75% es 2, con esto nos damos cuenta de que difícilmente un cliente puede llegar a tener 3 o 4 productos.
- **EstimatedSalary:** Con una media de 100090 y un valor máximo cercano a 20000, vemos una variedad en los salarios estimados. Sin embargo, la desviación estándar y la diferencia entre el tercer percentil y el valor máximo sugieren la presencia de outliers, aunque la distribución abarque desde salarios bajos hasta relativamente altos, sin llegar a considerar a los clientes millonarios.

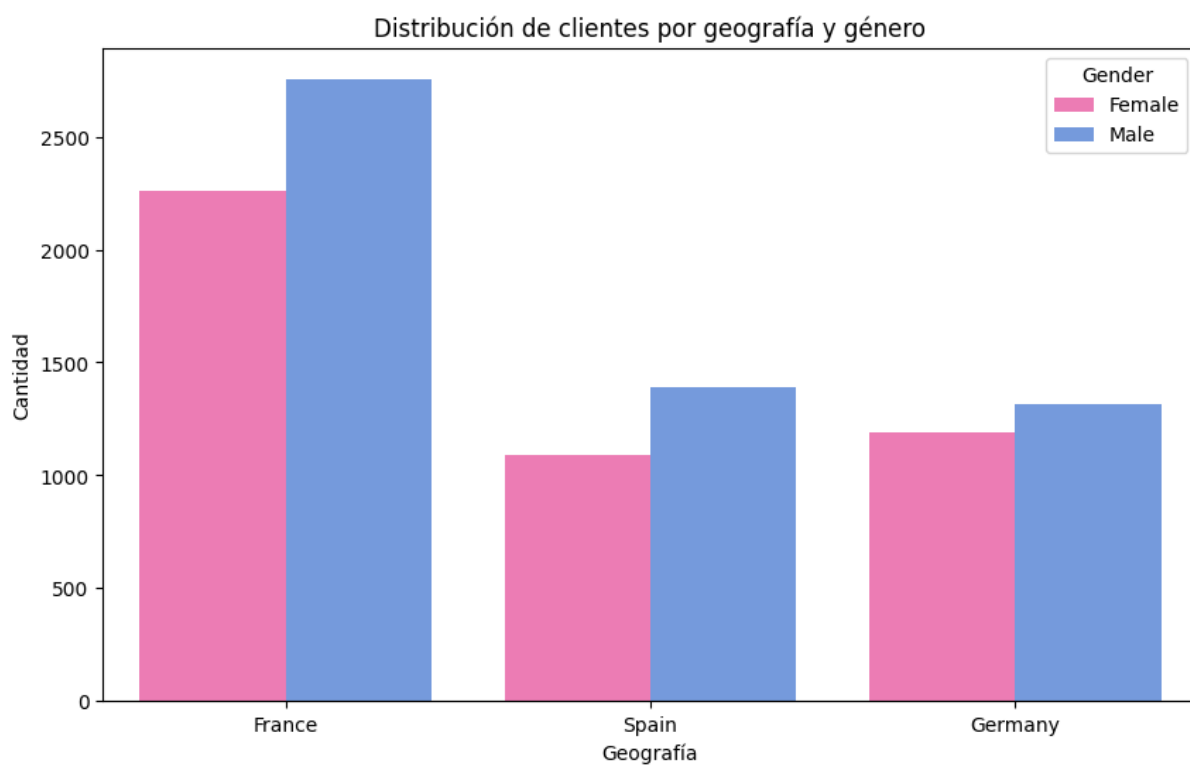
Datos Demográficos:

Para comenzar a analizar de manera general y luego especificar la información, es fundamental comprender qué tipo de clientes tiene el banco. Los campos más importantes para esto son Gender, Age y Geography.

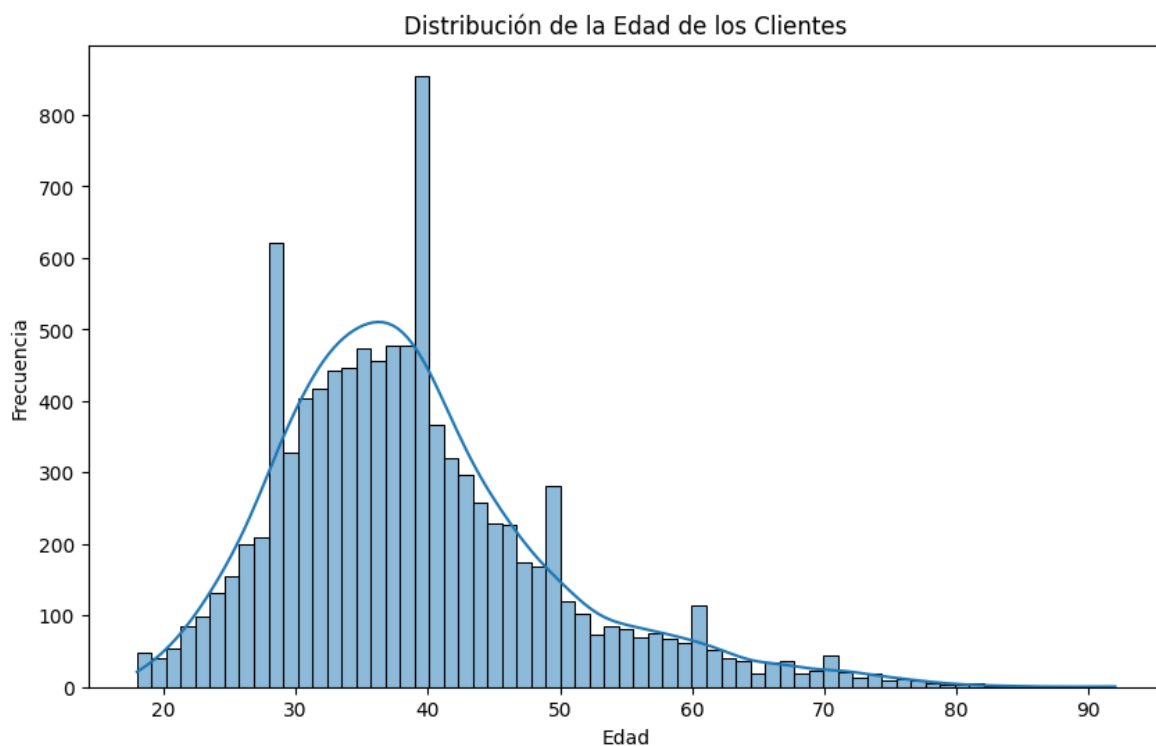
Distribución de clientes por geografía y género:

En la base de datos, encontramos clientes de tres países: Francia con 5014 clientes, España con 2477 clientes y Alemania con 2509 clientes. Además, el género de los clientes se clasifica únicamente como Male (hombres) y Female (mujeres).

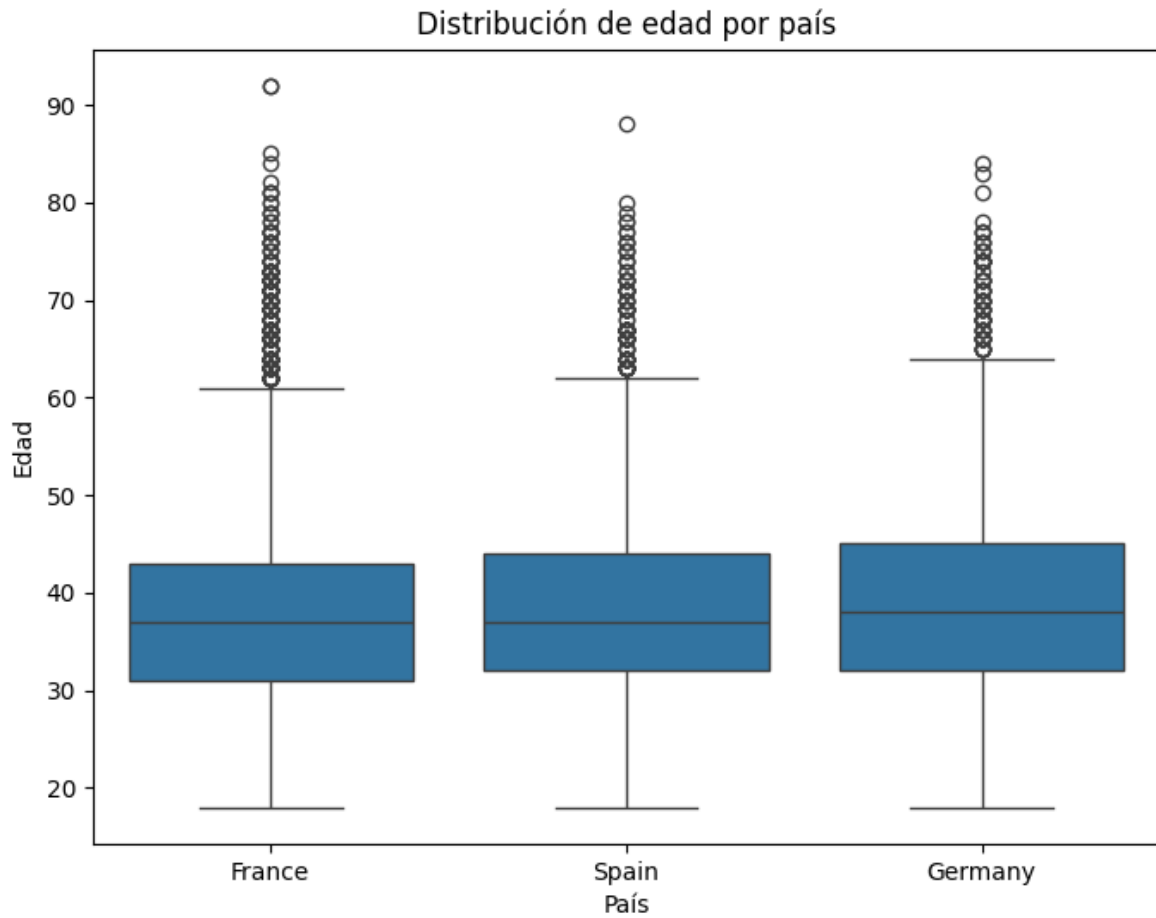
Observamos que no hay una variación significativa en la cantidad de clientes según el género en ninguna de las ciudades, aunque en todas ellas siempre hay más hombres que mujeres.



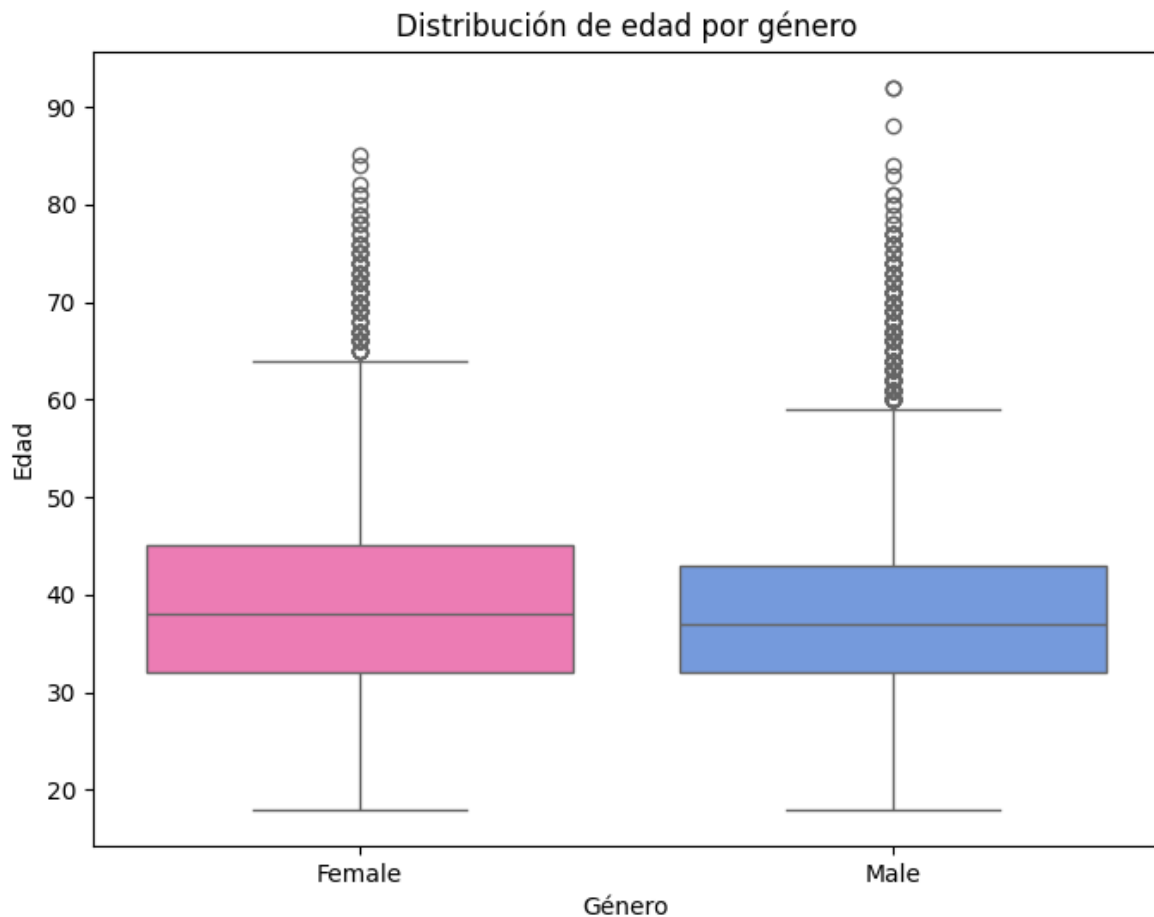
Como se mencionó anteriormente, hay una amplia variación en la edad de los clientes, que abarca desde los 18 hasta los 92 años. Esta diversidad etaria muestra que el banco atiende a una amplia gama de edades, lo que puede influir en las estrategias de retención y fidelización de clientes.



Además, al analizar la distribución de la edad en los tres países, notamos que no hay grandes diferencias. De hecho, la media de edad es muy similar entre los países, lo que indica una distribución de edades bastante homogénea.



Observamos una tendencia similar en cuanto a la distribución de la edad por género. Aunque los valores extremos mayores son más comunes en las mujeres, podemos considerar un rango normal de edad desde los 18 hasta poco más de los 58 años para ambos géneros. Esta distribución sugiere que el banco tiene una clientela diversificada en términos de edad, pero con una mayor presencia de clientes más jóvenes y de mediana edad.

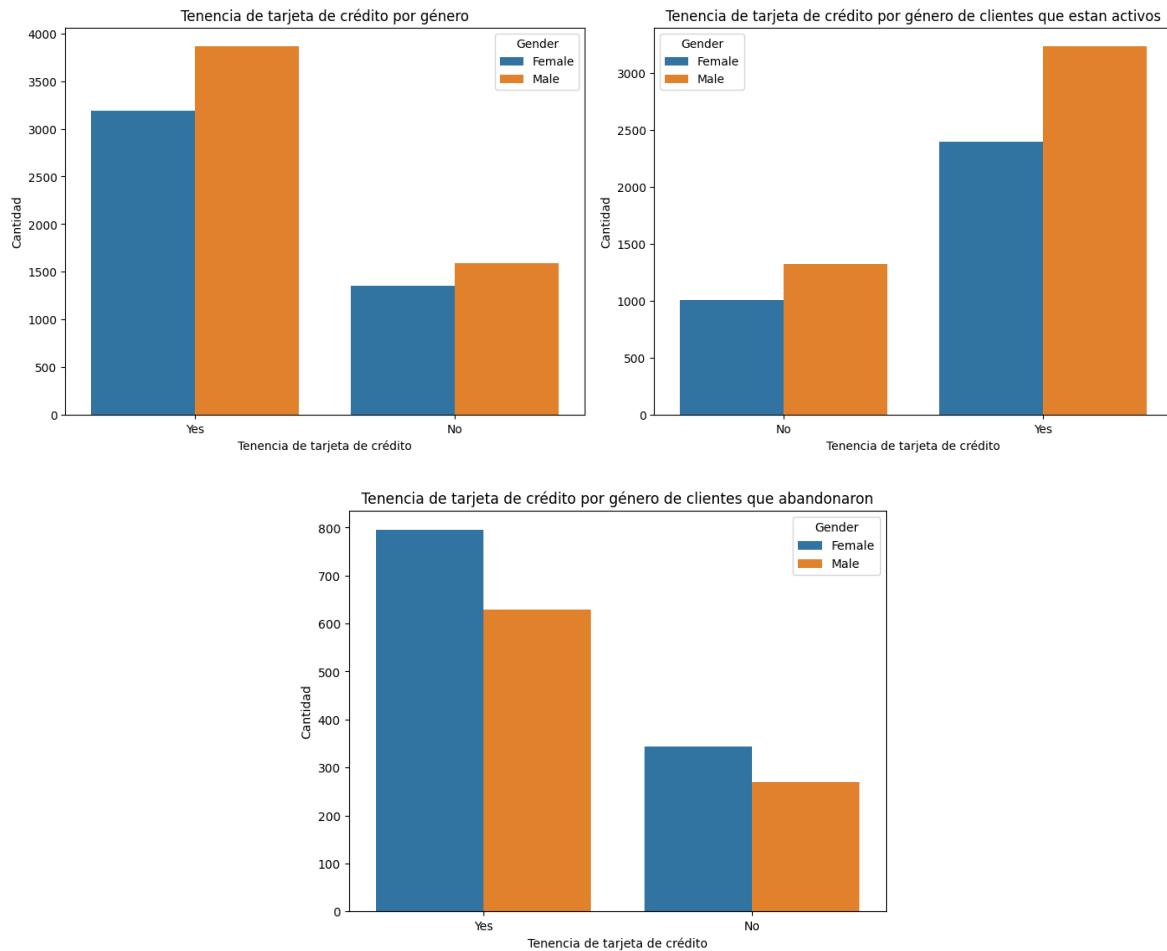


Datos bancarios:

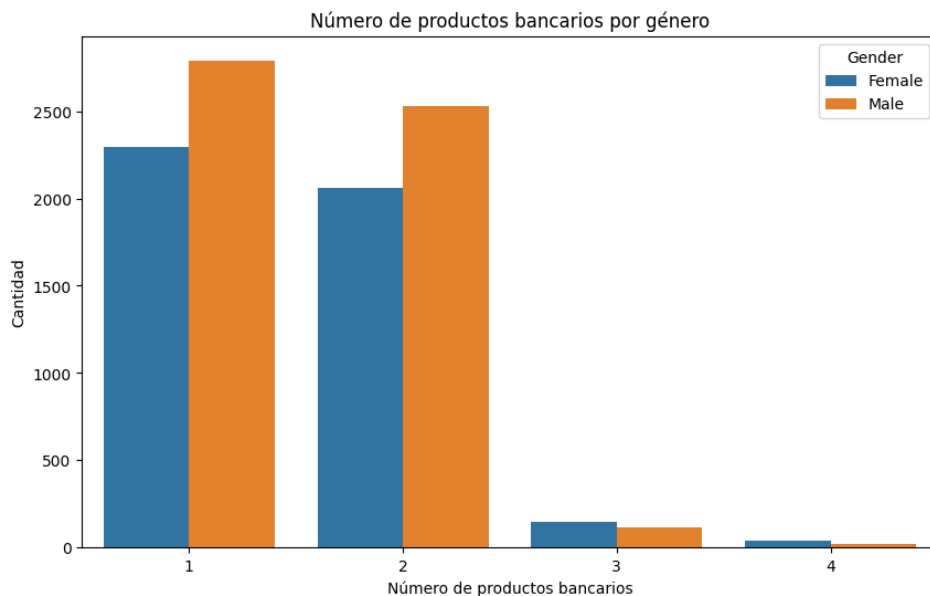
Los datos bancarios proporcionan información valiosa sobre la cantidad de clientes que poseen tarjetas de crédito, su conocimiento sobre los productos del banco, sus saldos y su fidelidad.

Observamos que hay una gran aceptación de las tarjetas de crédito entre los clientes de ambos sexos. Sin embargo, la posesión de una tarjeta de crédito no parece ser un factor determinante para la fidelidad del cliente. Al analizar a los clientes que han dejado el banco, encontramos que casi 1,400 de los 10,000 clientes que poseen una tarjeta de crédito se han desvinculado.

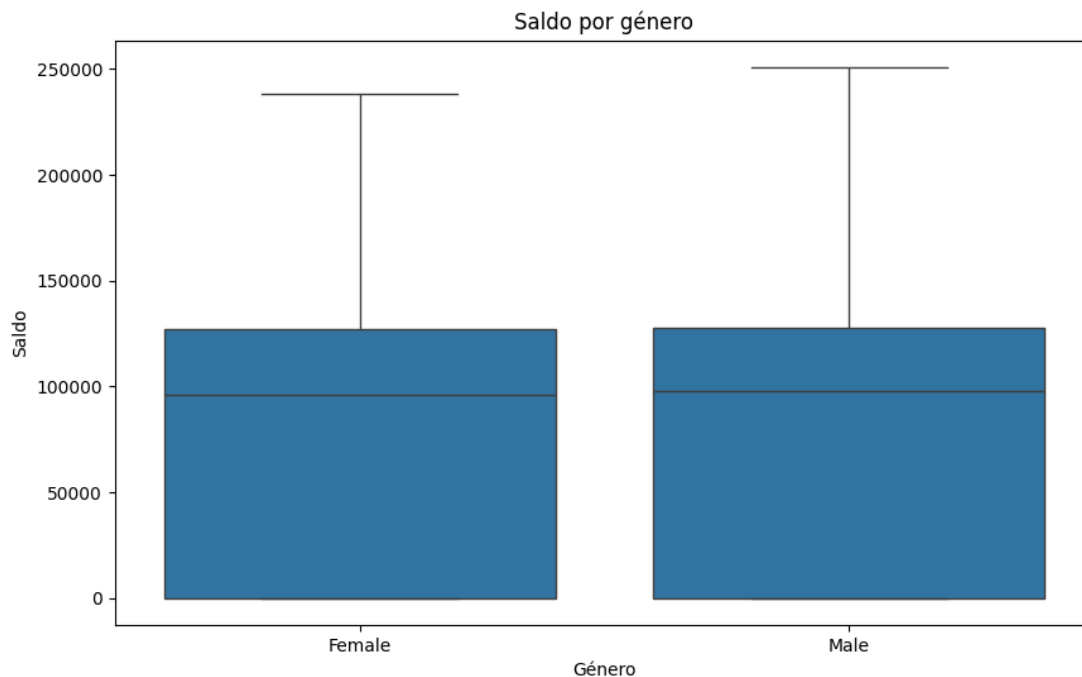
Esto sugiere que, aunque las tarjetas de crédito son populares, no son suficientes para retener a los clientes. Otros factores, posiblemente relacionados con la satisfacción del cliente, los servicios ofrecidos o las condiciones del mercado, podrían estar influyendo en su decisión de abandonar el banco.



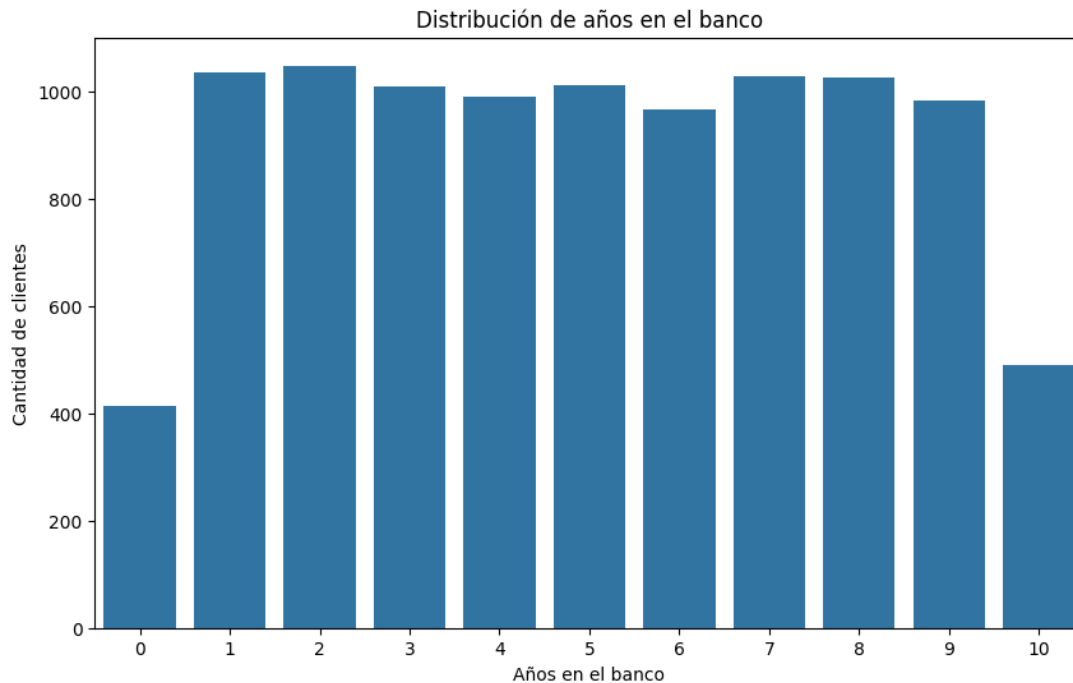
También analizamos la cantidad de productos bancarios que poseen los clientes, diferenciados por sexo, y confirmamos la tendencia observada anteriormente. La mayoría de los clientes tienen entre uno y dos productos bancarios. Sin embargo, la proporción de clientes con tres o cuatro productos es significativamente menor, representando menos del 0.5% de aquellos que tienen al menos un producto bancario.



En cuanto al balance de los clientes, observamos que, aunque existen valores extremos, no se identifican valores atípicos significativos. Esto indica que los saldos de los clientes están dentro de un rango esperado y podemos considerar que están normalizados para el banco.



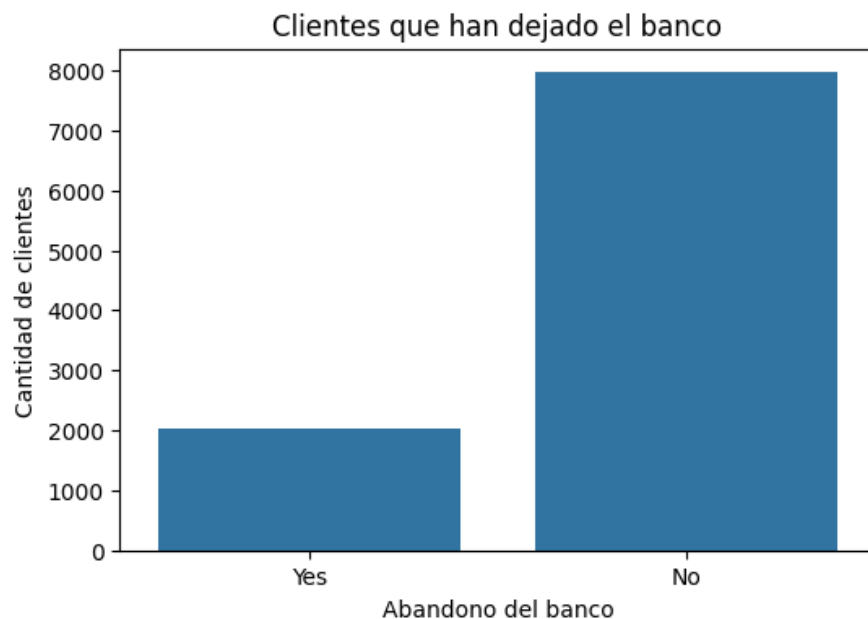
Por último, analizamos la distribución de la fidelidad de los clientes en el banco a lo largo de los años. Observamos que la mayoría de los clientes han permanecido con el banco durante varios años, con una notable excepción en los años 0 y 10, que tienen menos clientes. Esto sugiere que, aunque el banco goza de una buena fidelidad entre sus clientes, no está captando nuevos clientes en la medida que podría ser óptima.



Análisis de abandono:

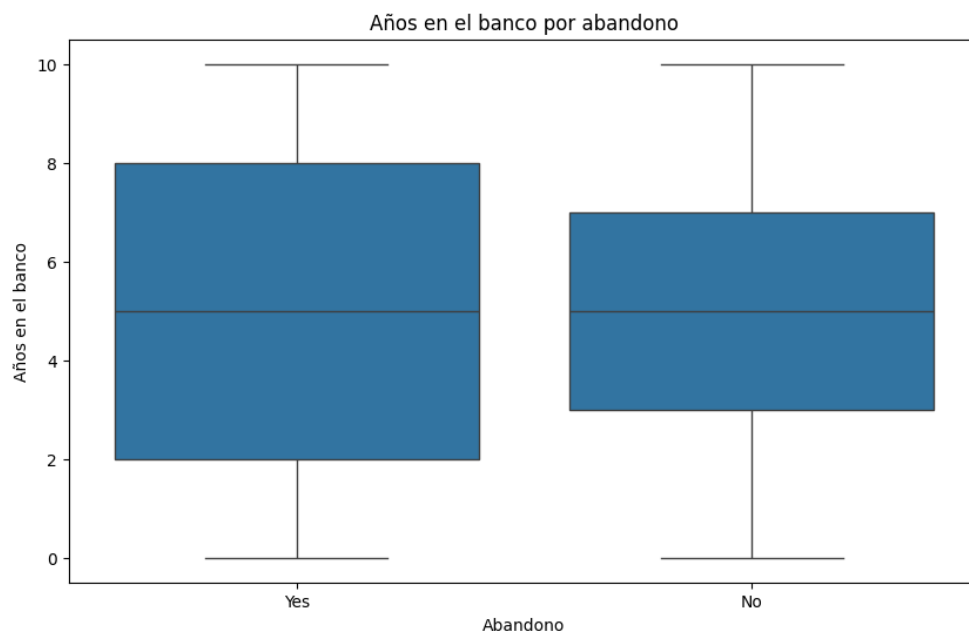
Ahora procederemos a analizar los datos de abandono de clientes en relación con los datos previamente examinados para identificar tendencias o información relevante que pueda influir en la retención de clientes.

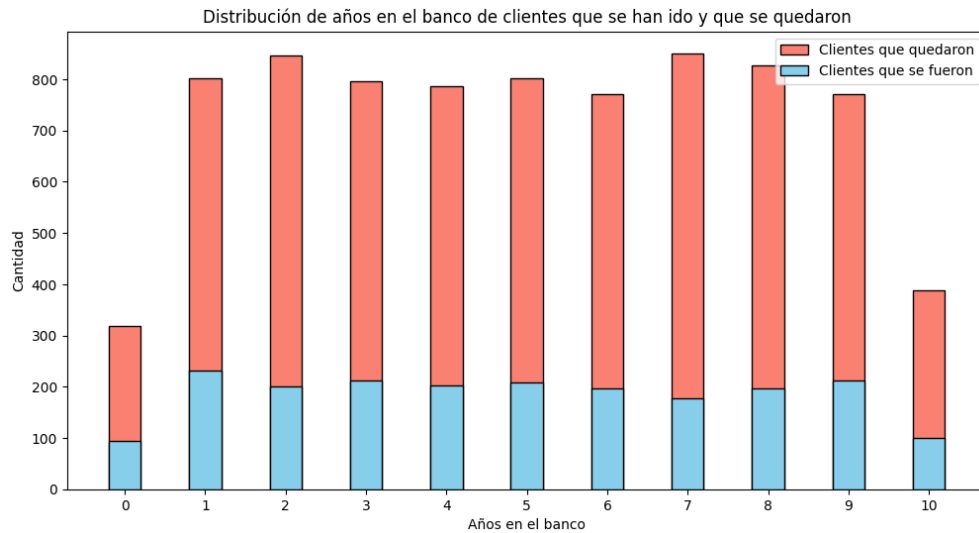
En primer lugar, es importante determinar cuántos clientes permanecen con nosotros y cuántos se han marchado durante el período de tiempo analizado. Observamos que aproximadamente el 25% de los clientes han abandonado el banco. Si bien este porcentaje puede ser preocupante, es crucial considerar el contexto temporal. Sin embargo, al no contar con una especificación temporal precisa, como una fecha completa, resulta difícil evaluar completamente la gravedad de la situación.



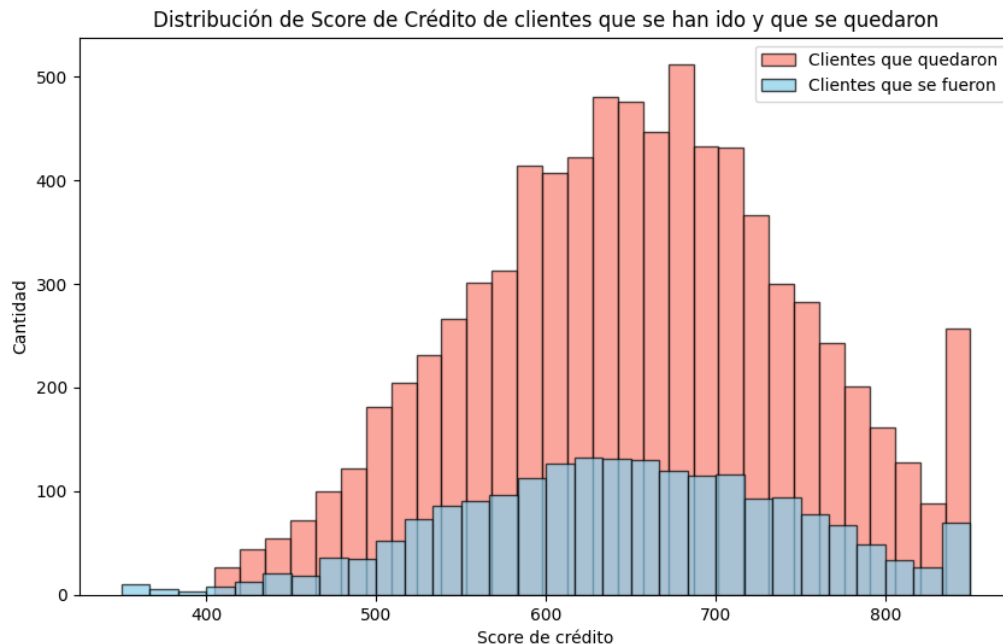
Como podemos observar, la fidelidad de los clientes no parece estar influenciada por la duración de su relación con el banco, ya que la mediana se mantiene constante a lo largo de los años. Esto sugiere que es fundamental reforzar las estrategias de atención al cliente, independientemente de su antigüedad en el banco.

Implementar estrategias destinadas a retener a los clientes por más tiempo, ofreciendo mejores beneficios y servicios personalizados, podría ser una opción efectiva para mejorar la fidelidad y fortalecer la relación con los clientes a largo plazo.

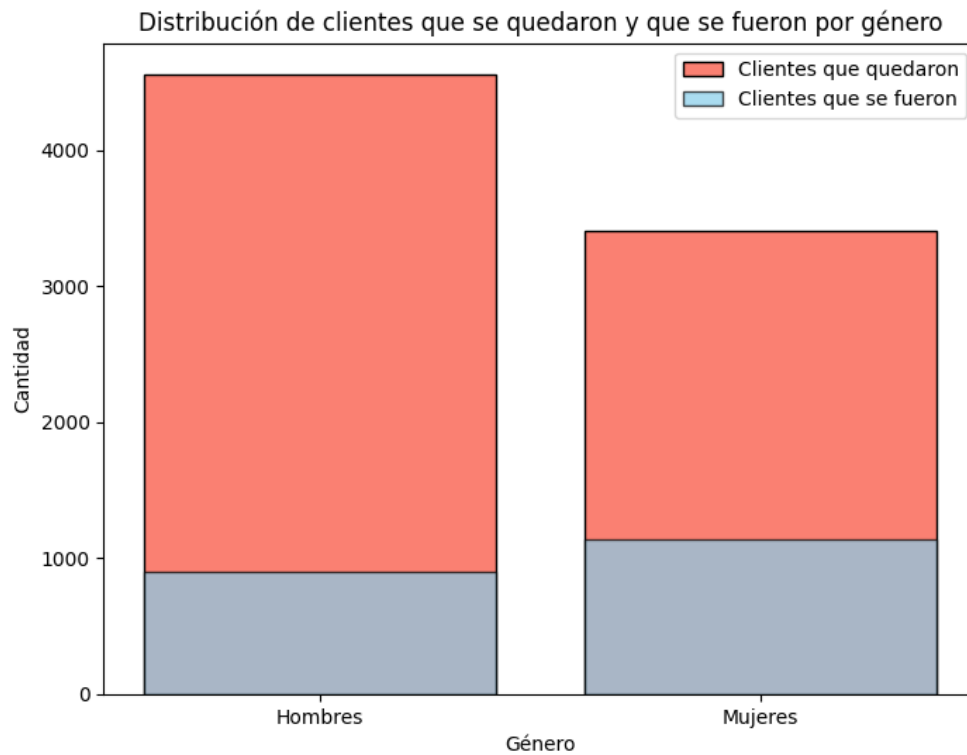




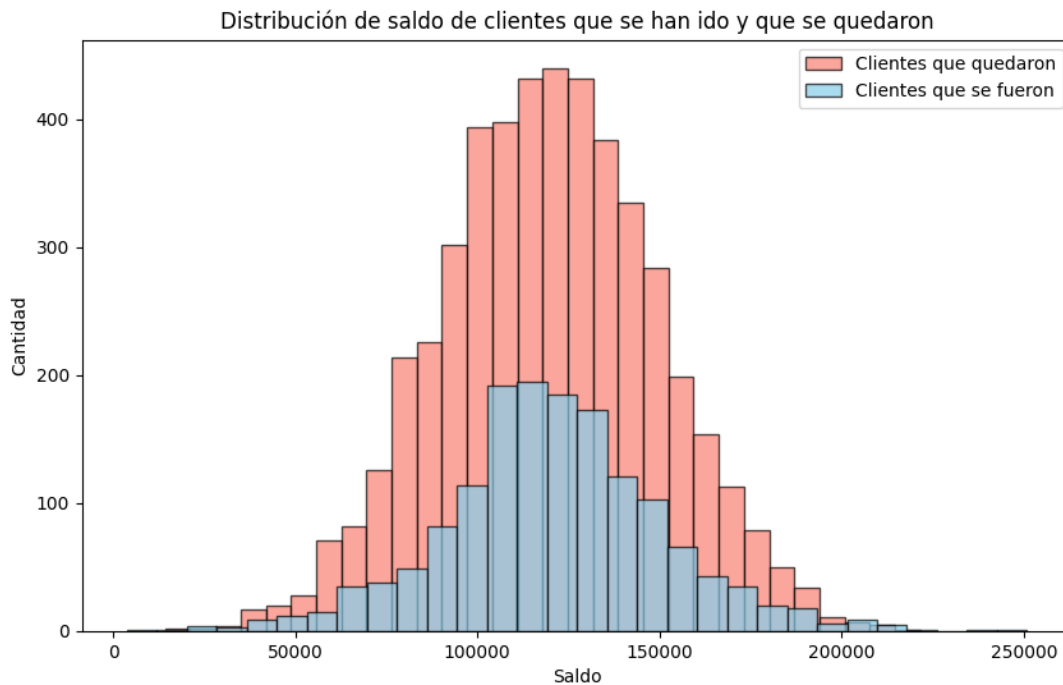
Continuando con las suposiciones, podemos observar que un puntaje crediticio alto no parece tener un impacto significativo en la decisión de los clientes de mantenerse en el banco. Esto se evidencia al comparar las distribuciones de aquellos con puntajes altos y bajos, las cuales son muy similares. Aunque es cierto que algunos clientes con los puntajes más bajos han abandonado el banco, no es concluyente determinar que el puntaje crediticio sea una de las principales razones para permanecer en él.



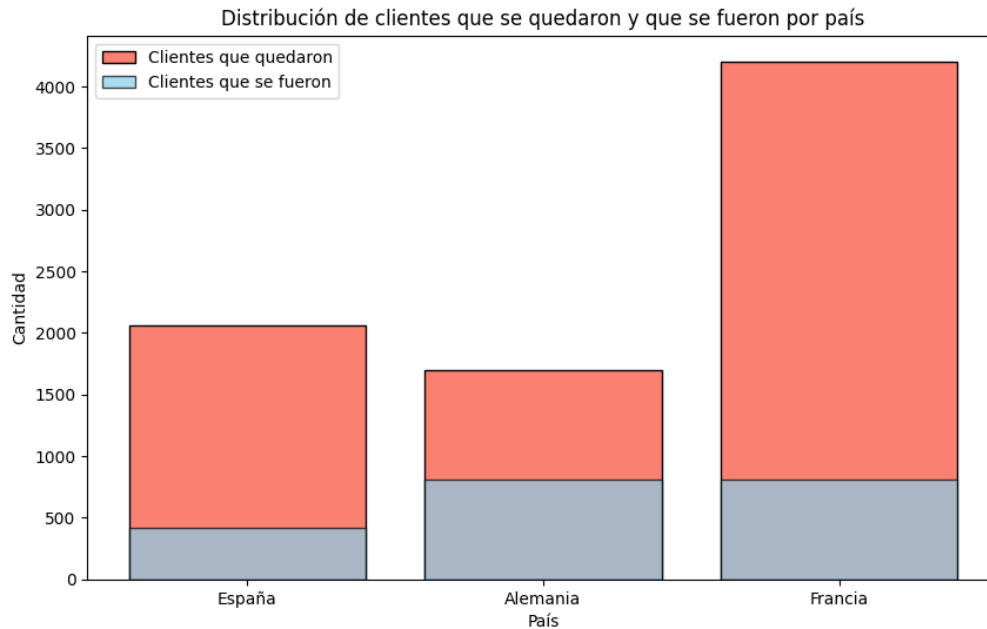
En cuanto al género, podemos observar que, a pesar de que hay menos mujeres en nuestra muestra, estas representan la mayoría de los clientes que han dejado el banco. Esto sugiere que podría existir una relación entre el género y la probabilidad de abandono del banco.



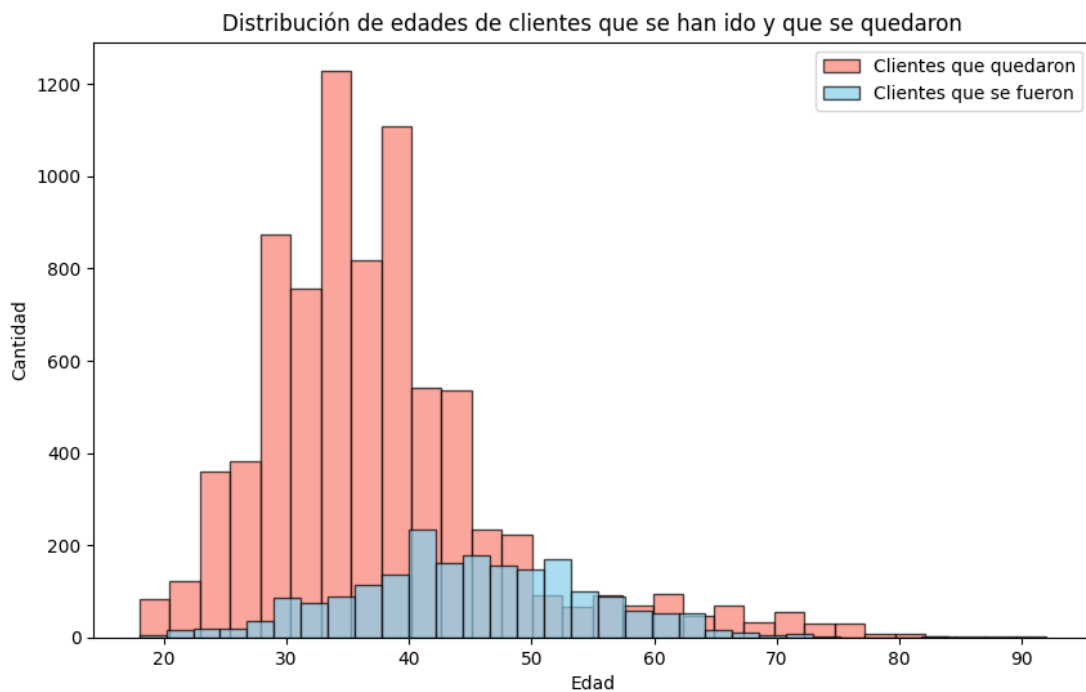
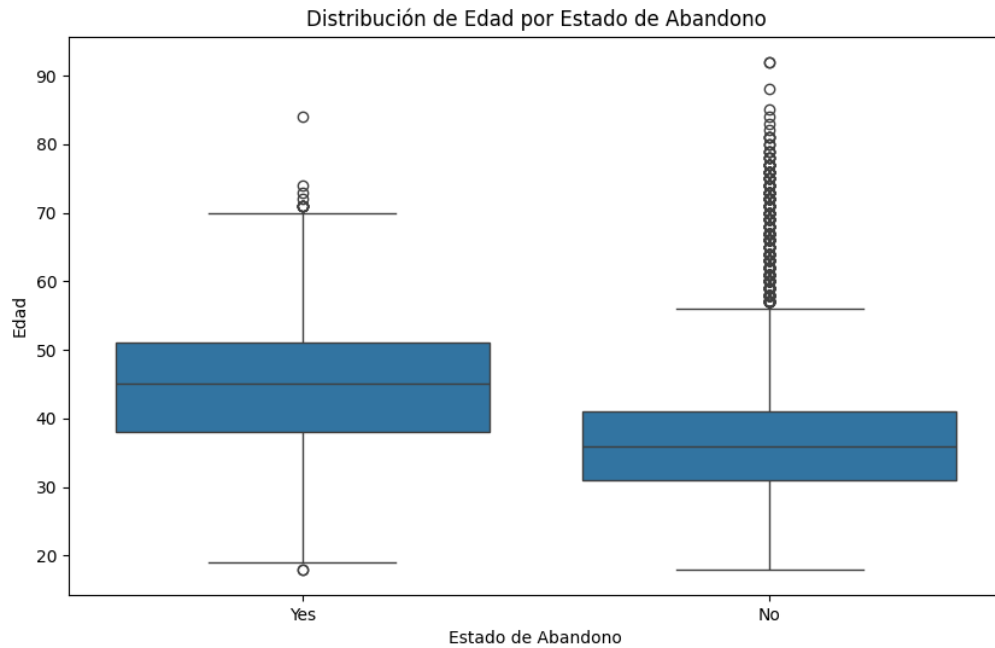
Al analizar la distribución de saldos, no parece haber una relación evidente con la decisión de los clientes de abandonar el banco. Por lo tanto, la suposición de que el saldo tenga un impacto directo en esta decisión no parece ser cierta a simple vista.



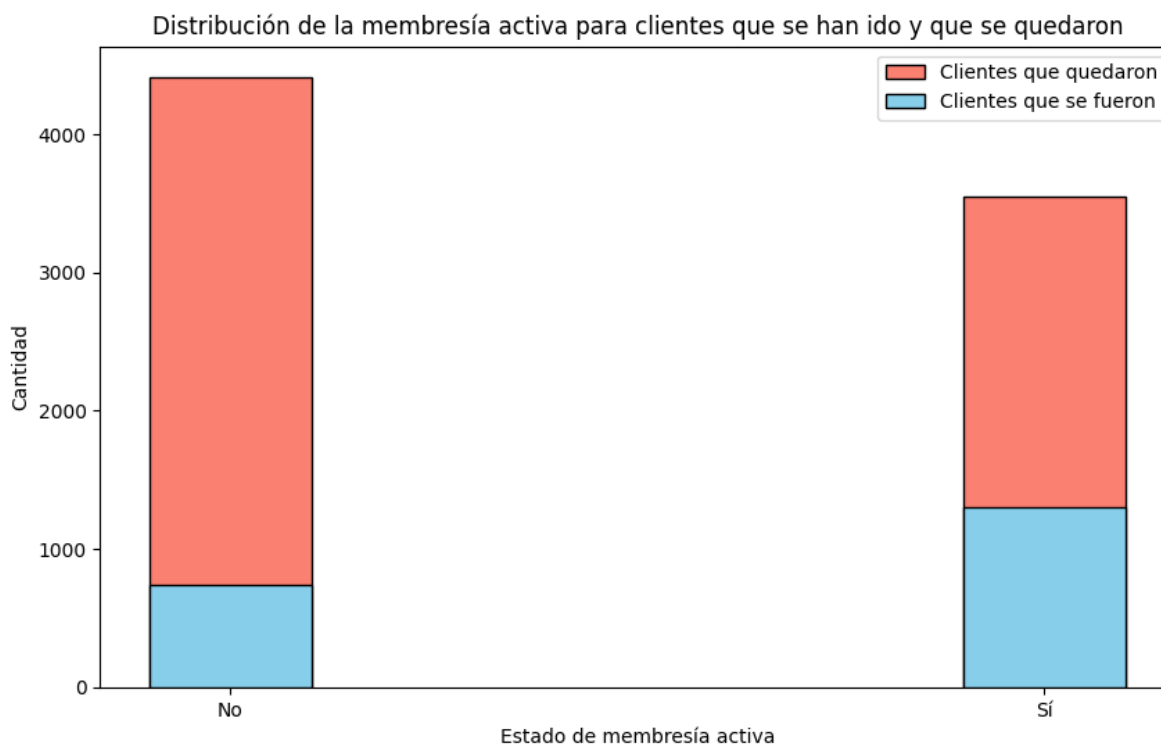
Otro aspecto interesante son las nacionalidades de los clientes. Observamos que, aunque hay menos clientes alemanes en comparación con otras nacionalidades, estos son los que más se han desvinculado del banco. Mientras que Francia también presenta una cantidad considerable de clientes que se han desvinculado, en términos de la proporción de clientes en el banco, no hay comparación, ya que Francia es el país con más clientes.



Otra relación interesante se observa al analizar la edad de los clientes. Podemos notar que las personas que tienen entre 35 y un poco más de 60 años tienden a abandonar el banco con mayor frecuencia. De hecho, en algunos rangos de edad, hay menos clientes vinculados que no vinculados, lo que sugiere una tendencia significativa en la decisión de abandono en esos grupos de edad.



Resulta interesante observar que, contrariamente a la suposición inicial, los clientes más activos también han mostrado una tendencia a abandonar el banco en mayor medida que aquellos que son menos activos. Esta observación sugiere que la actividad de los clientes parece ser un indicador determinante de su lealtad al banco, contradiciendo así otra suposición inicial.



Para concluir, al analizar el mapa de correlación entre todas estas variables numéricas y QUE cliente abandone el banco, confirmamos las conclusiones anteriores. La edad emerge como un factor importante, seguido por el balance de la cuenta, aunque en menor medida. Además, el estado de la membresía y el género del cliente también parecen estar relacionados con la probabilidad de abandono.

Sin embargo, para otras variables no podemos establecer una relación directa con la estabilidad del cliente en el banco. En lugar de centrarse únicamente en estas variables, abordar los problemas que son más evidentes puede ayudar a reducir la cantidad de clientes perdidos y mejorar la retención en el banco.

Bank Marketing Dataset:

El conjunto de datos de Marketing Bancario nos brinda una visión más amplia que su predecesor, permitiéndonos explorar una variedad de variables. Enfoquémonos en las más relevantes para el resultado final: la aceptación de una suscripción a un depósito a plazo.

Después de realizar la limpieza de datos, no se identificaron necesidades específicas de tratamiento. No se encontraron valores nulos, lo que sugiere que los análisis posteriores reflejarán fielmente el contenido y la integridad de este conjunto de datos.

Estos son los campos:

1. **age:** Edad (numérico)

2. **job:** Tipo de trabajo (categórico)
3. **marital:** Estado civil (categórico)
4. **education:** Nivel educativo (categórico)
5. **default:** ¿Tiene crédito en mora? (categórico)
6. **housing:** ¿Tiene préstamo hipotecario? (categórico)
7. **loan:** ¿Tiene préstamo personal? (categórico)

Relacionado con el último contacto de la campaña actual: 8. **contact:** Tipo de comunicación del contacto (categórico)

9. **month:** Último mes de contacto del año (categórico)
10. **day_of_week:** Último día de la semana de contacto (categórico)
11. **duration:** Duración del último contacto, en segundos (numérico)

Otros atributos: 12. **campaign:** Número de contactos realizados durante esta campaña y para este cliente (numérico)

13. **pdays:** Número de días transcurridos desde que el cliente fue contactado por última vez en una campaña anterior (numérico)
14. **previous:** Número de contactos realizados antes de esta campaña y para este cliente (numérico)
15. **poutcome:** Resultado de la campaña de marketing anterior (categórico)

Atributos de contexto social y económico: 16. **emp.var.rate:** Tasa de variación del empleo - indicador trimestral (numérico)

17. **cons.price.idx:** Índice de precios al consumidor - indicador mensual (numérico)
18. **cons.conf.idx:** Índice de confianza del consumidor - indicador mensual (numérico)
19. **euribor3m:** Tasa Euribor a 3 meses - indicador diario (numérico)
20. **nr.employed:** Número de empleados - indicador trimestral (numérico)

Variable de salida (objetivo deseado): 21. **y:** ¿El cliente ha suscrito un depósito a plazo? (binario)

Análisis numérico:

Este conjunto de datos, a diferencia del anterior, también incluye información sobre el país de origen del cliente. Esta variable adicional podría tener un impacto significativo en las decisiones que toman los clientes, por lo que es importante tenerla en cuenta durante nuestro análisis.

	age	duration	campaign	pdays	previous	emp_var_rate	cons_price_idx	euribor3m
mean	40.02	258.32	2.57	962.46	0.17	0.08	93.58	3.62
std	10.42	259.31	2.77	186.94	0.49	1.57	0.58	1.73
min	17.00	0.00	1.00	0.00	0.00	-3.40	92.20	0.63
25%	32.00	102.00	1.00	999.00	0.00	-1.80	93.08	1.34
50%	38.00	180.00	2.00	999.00	0.00	1.10	93.75	4.86
75%	47.00	319.00	3.00	999.00	0.00	1.40	93.99	4.96
max	98.00	4918.00	56.00	999.00	7.00	1.40	94.77	5.04

Rápidamente identificamos que enfrentamos un caso similar con la variable de edad. Sin embargo, otro aspecto destacado es la duración máxima de la llamada, que alcanza los 4918 segundos, equivalente a aproximadamente una hora y media. Esta cifra representa un outlier significativo, ya que el tercer percentil se sitúa en menos del 10% de esa cantidad. A pesar de esto, no se procedió a tratar este dato debido a la amplia distribución de outliers en la duración de las llamadas. Eliminar una cantidad tan grande de datos podría ser perjudicial para el análisis, y crear rangos sería sesgado, ya que podrían abarcar la mayoría de los datos o dejar la mayoría de los outliers fuera del análisis, lo que afectaría la interpretación de la información.

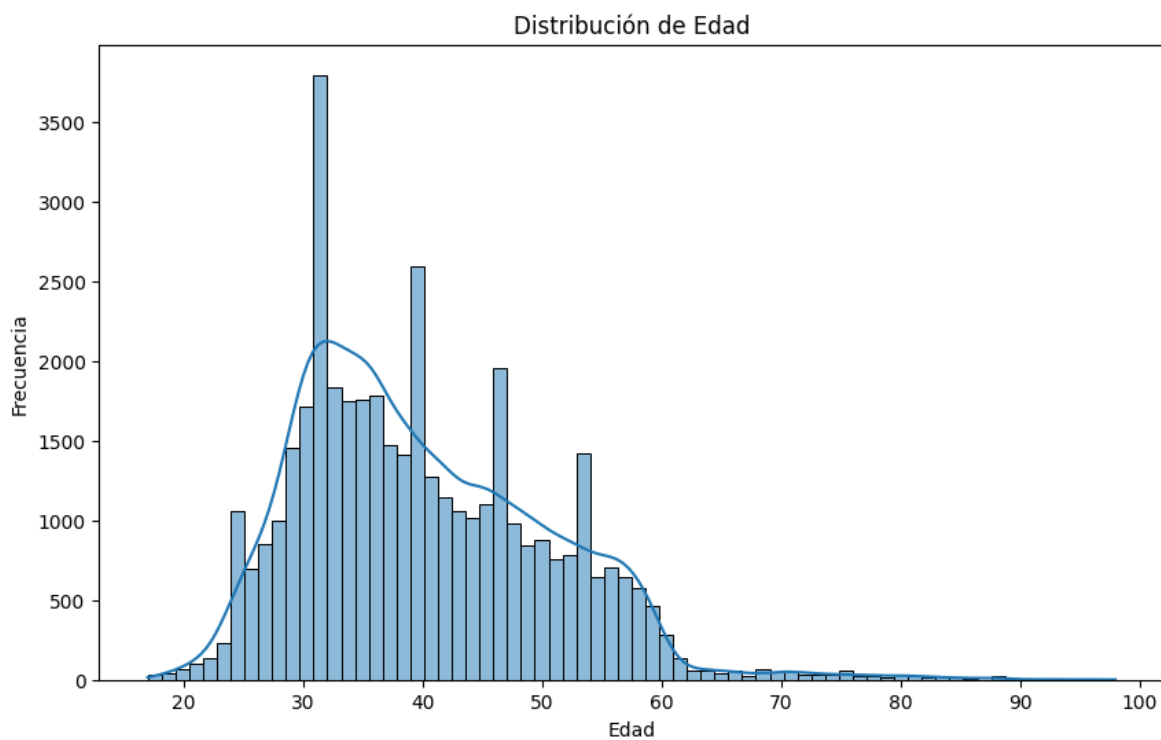
Además, observamos que la variable de campaña también presenta valores extremos. Respecto a "pdays", se destaca que incluso el 25% más bajo de los datos indica que una gran cantidad de clientes no han sido contactados previamente. Esto es similar a la variable "previous", donde desde el 25% de los datos vemos que una parte significativa de los clientes no tiene historial de contactos anteriores.

Por último, al analizar las variables sociales y económicas, observamos un buen equilibrio, dado que están determinadas por entidades gubernamentales. Aunque es posible que experimenten fluctuaciones bruscas, es común que países como Portugal mantengan cierta estabilidad durante varios trimestres.

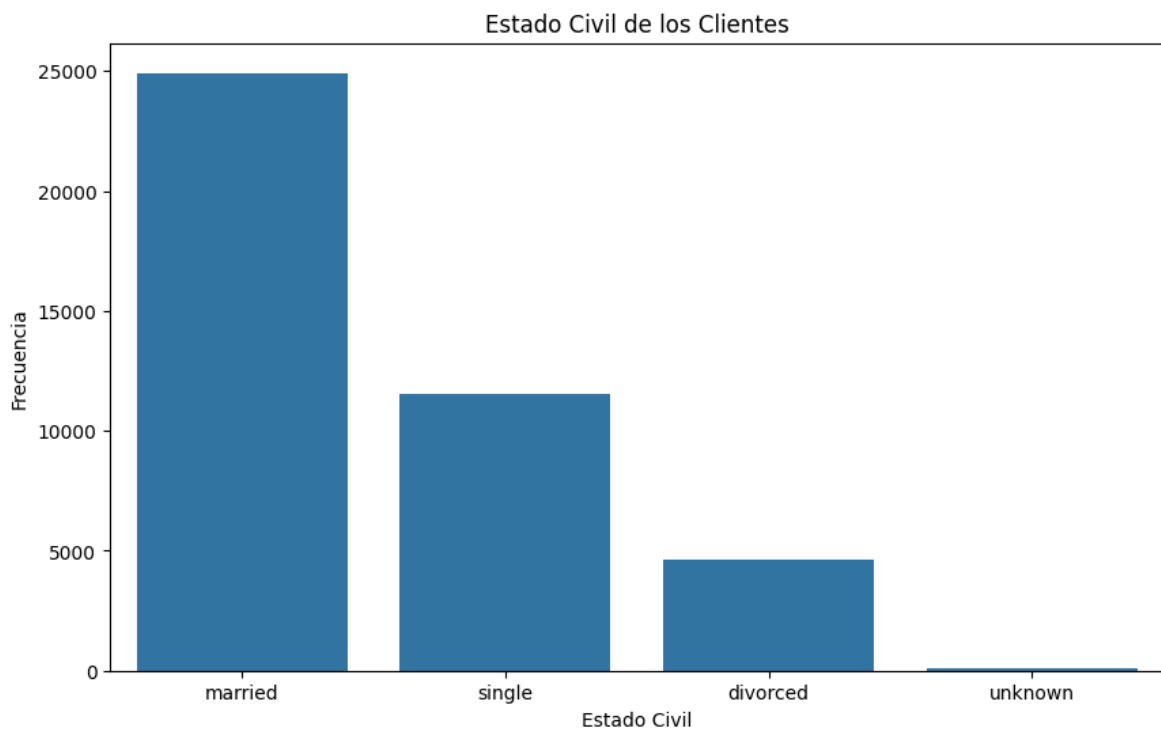
Este análisis destaca varios aspectos importantes del conjunto de datos, incluyendo la identificación de outliers, la distribución de contactos previos y la estabilidad de las variables económicas y sociales.

Análisis Demográfico:

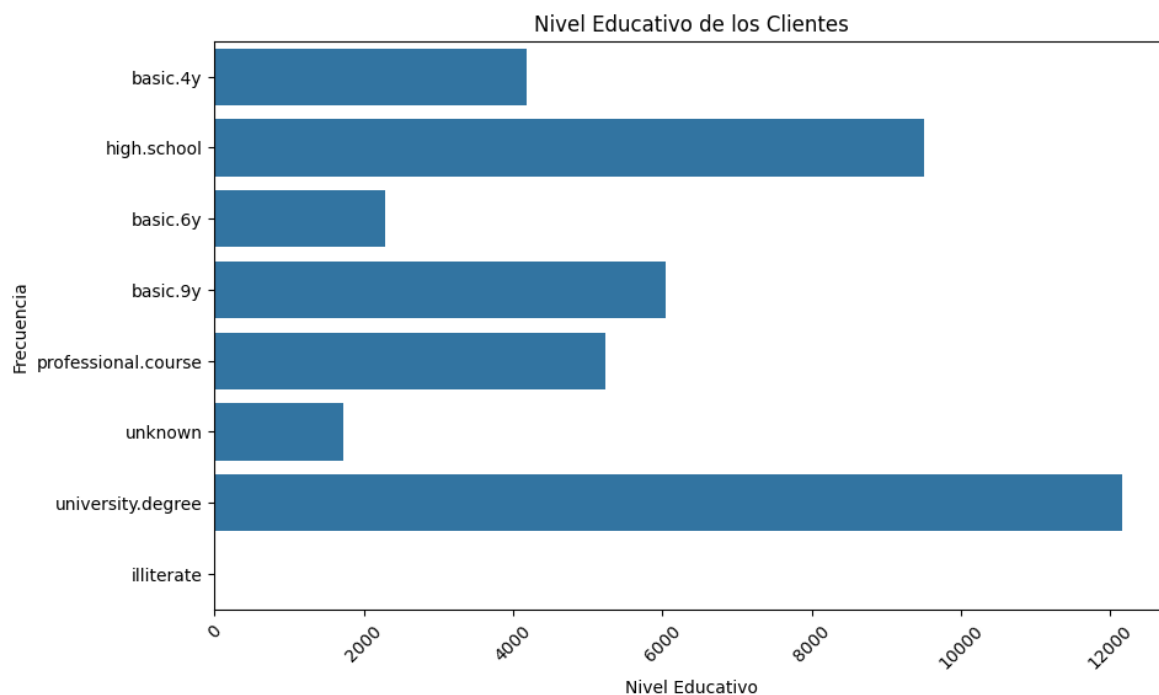
Empezamos por la edad, esta nos permite ver fácilmente que la mayoría de los clientes están centrados entre los 28 y poco más de los 50 años



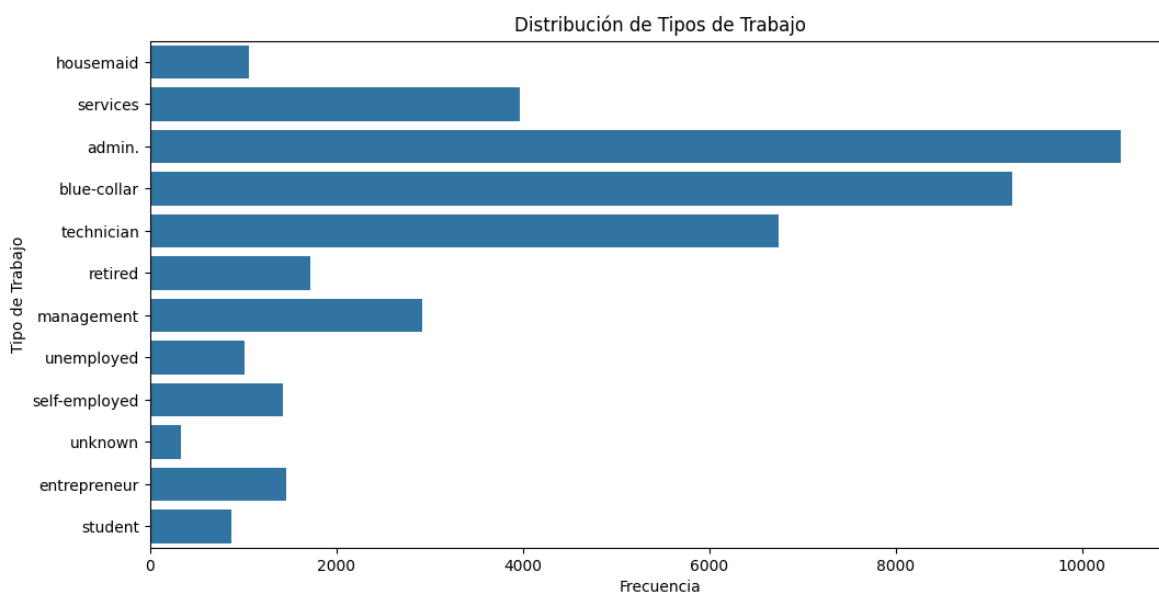
Estado civil de los clientes:



Nivel de educación:

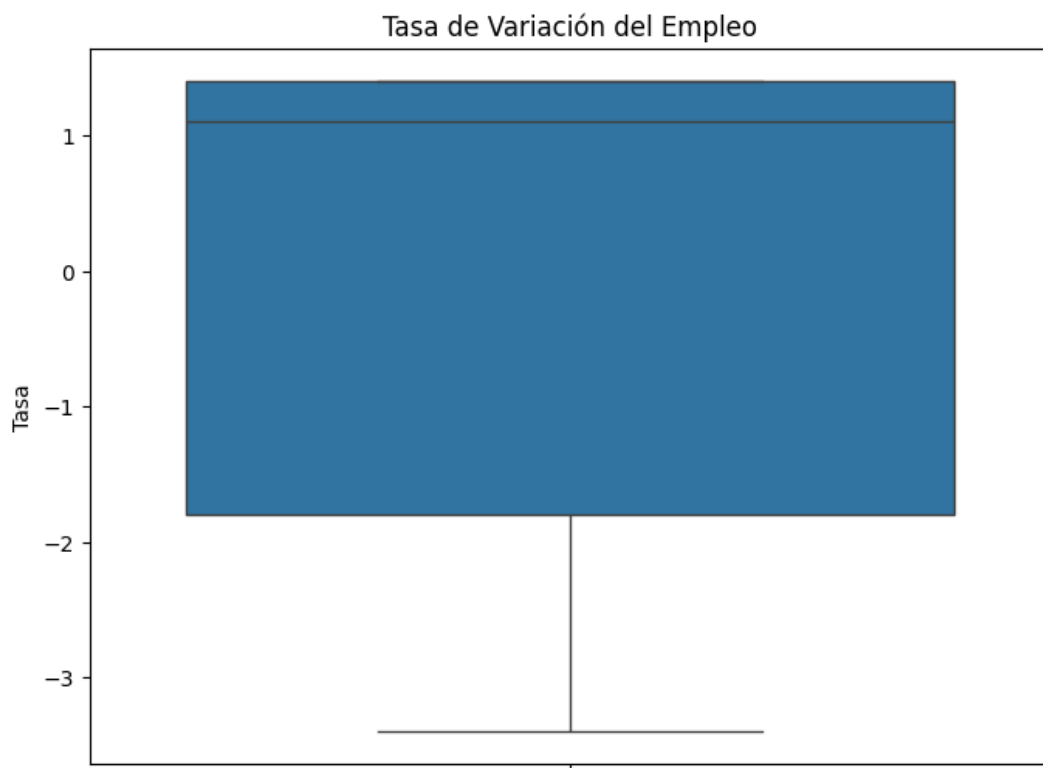


Tipos de trabajo:

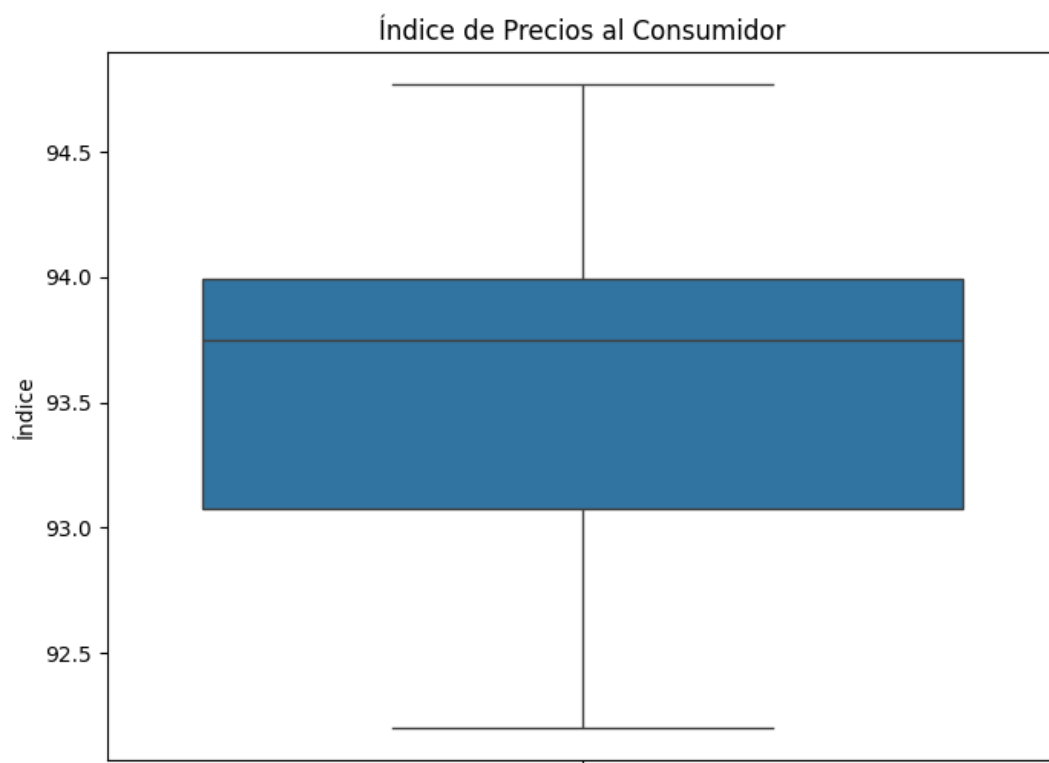


Variables económicas y sociales:

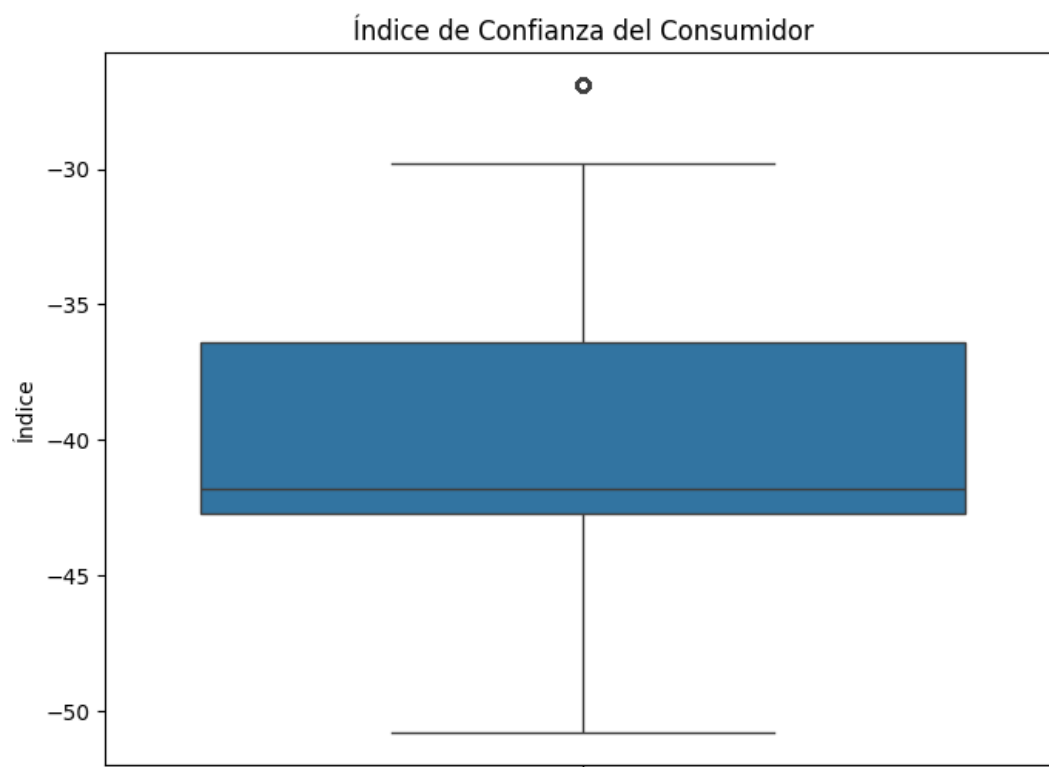
tasa de variación de empleo:



Índice de precios al consumidor:



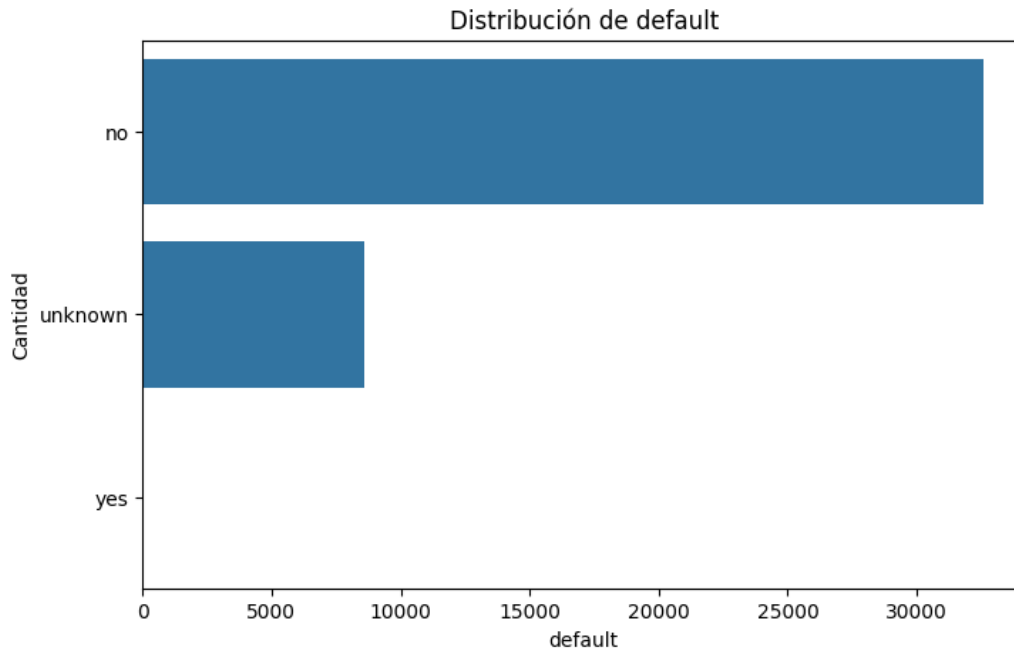
Índice de confianza del consumidor:



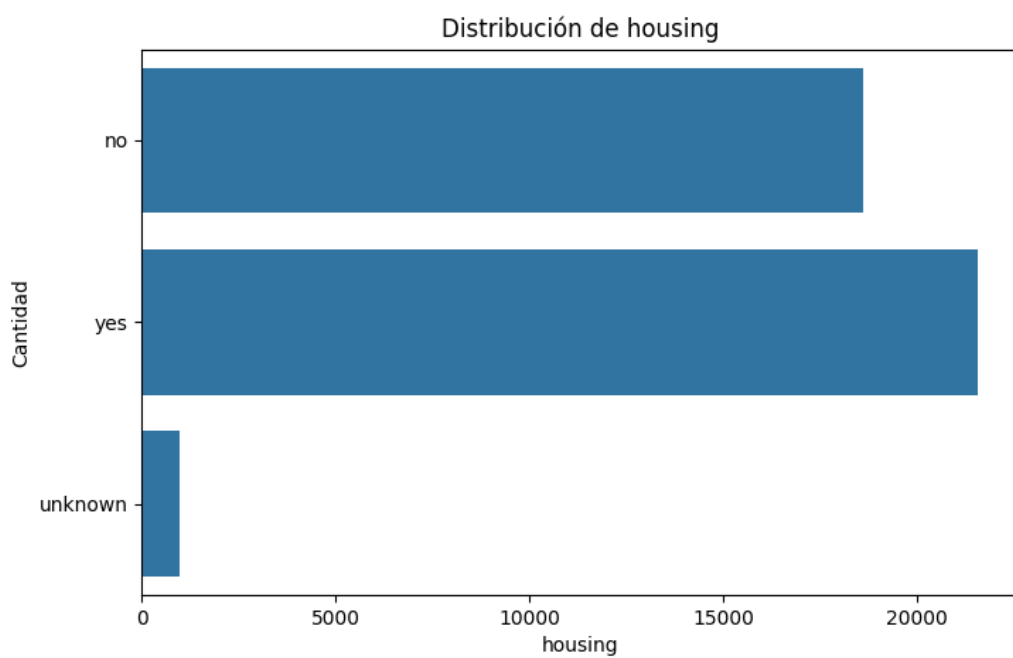
Datos Bancarios:

Aquí podremos visualizar una variedad de aspectos, incluyendo los tipos de contacto utilizados, los meses con mayor actividad de contacto con los clientes, la cantidad de clientes en mora, los diferentes tipos de préstamos, entre otros.

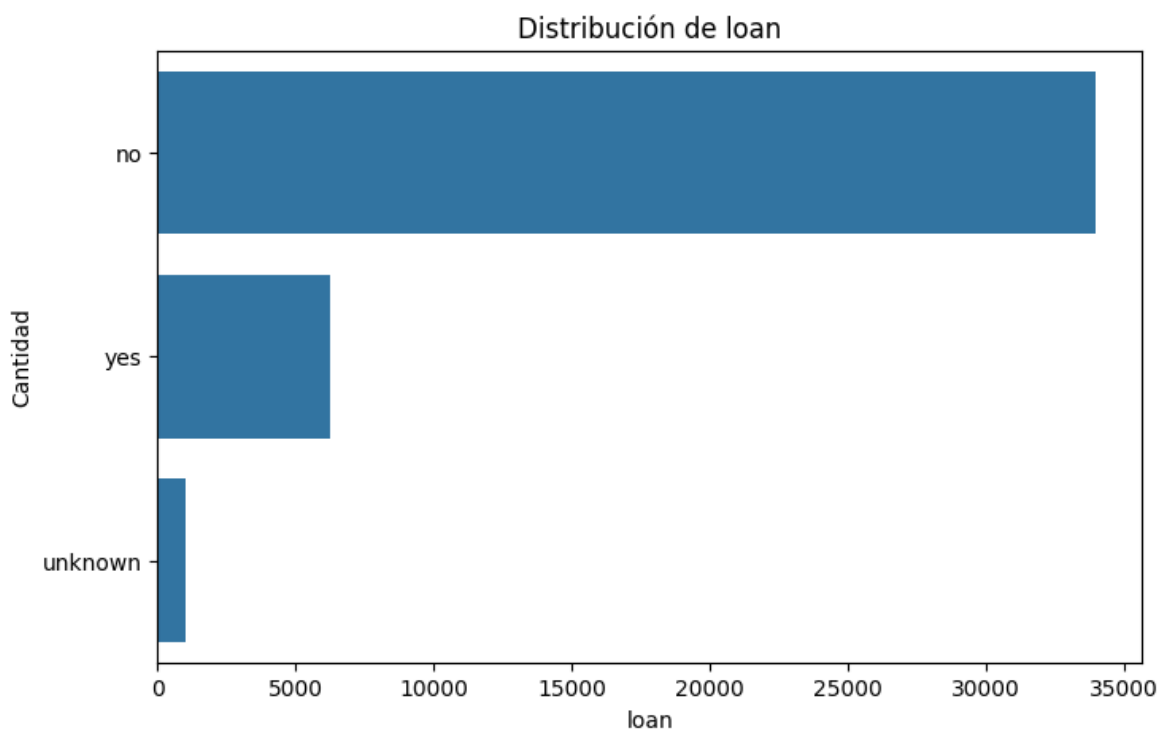
Personas en mora: es interesante ya que las personas que están en mora no son contactadas (realmente tienen solo 3 contactos)



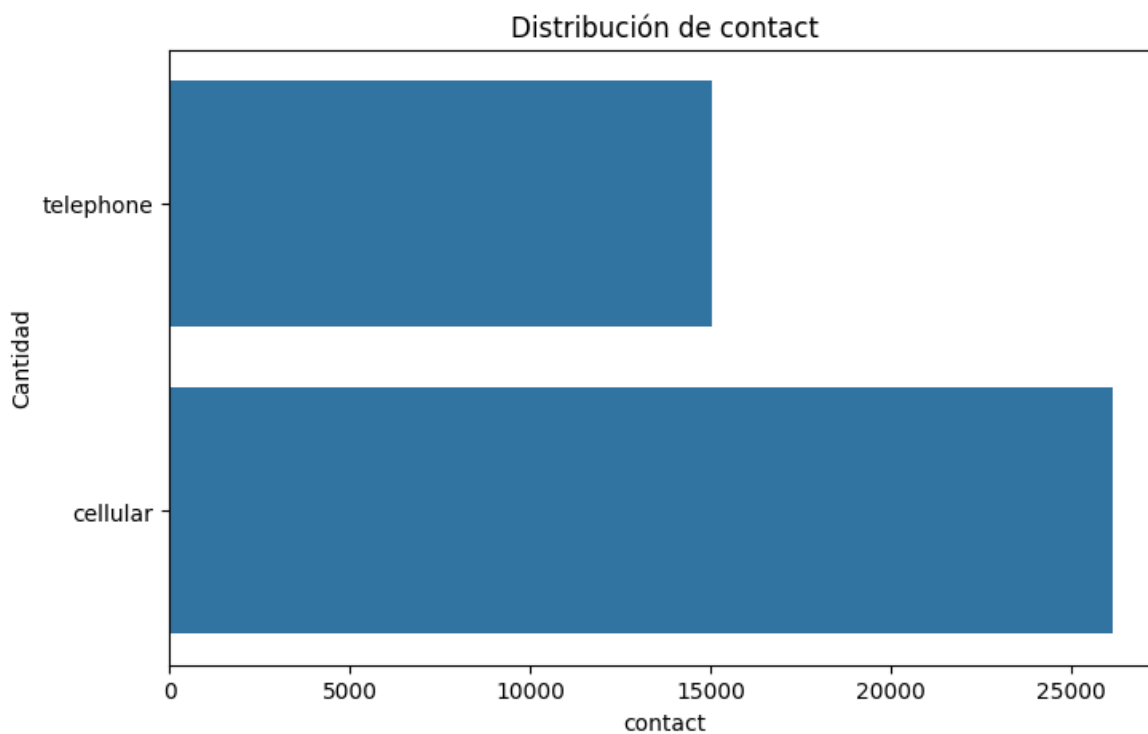
Crédito hipotecario:



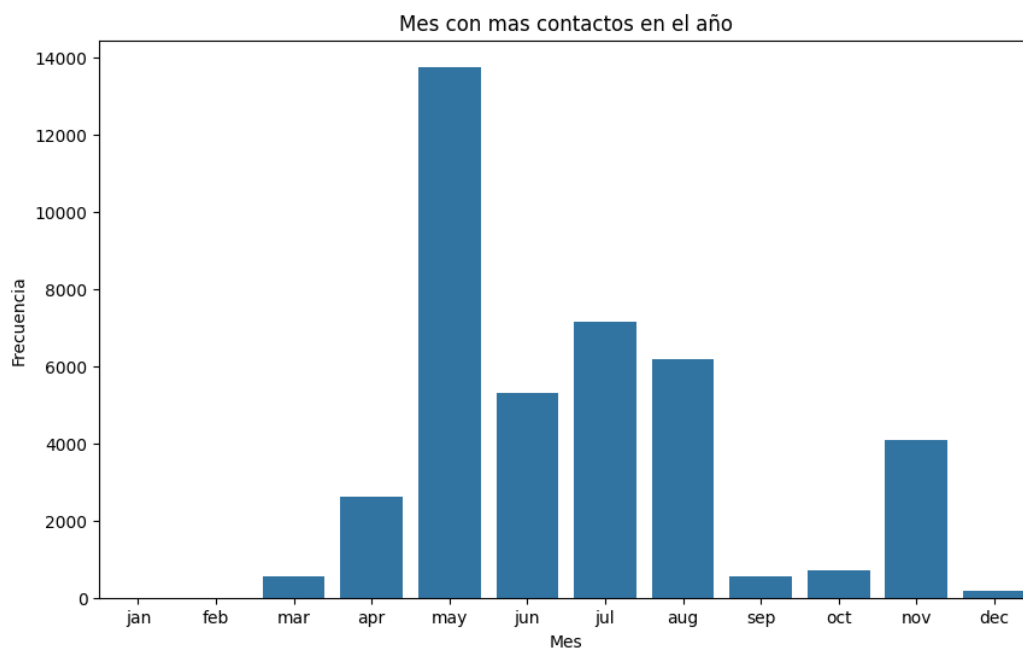
Crédito personal:



Tipo de contacto:



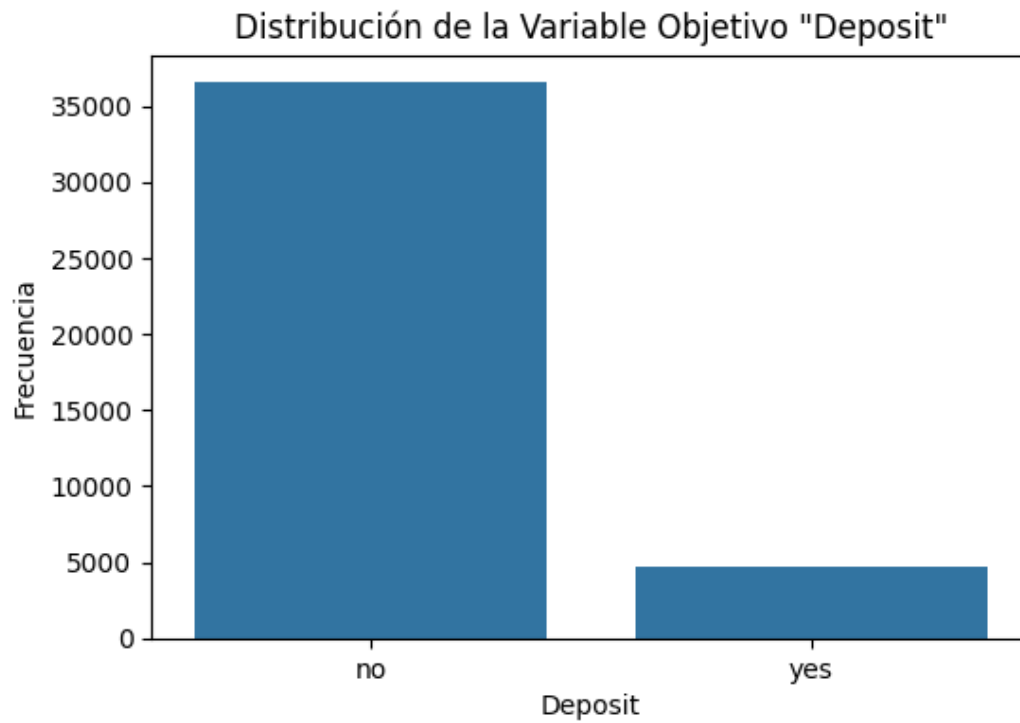
Mes con mas contactos en el año:



Variables importantes en la decisión de aceptación de depósito:

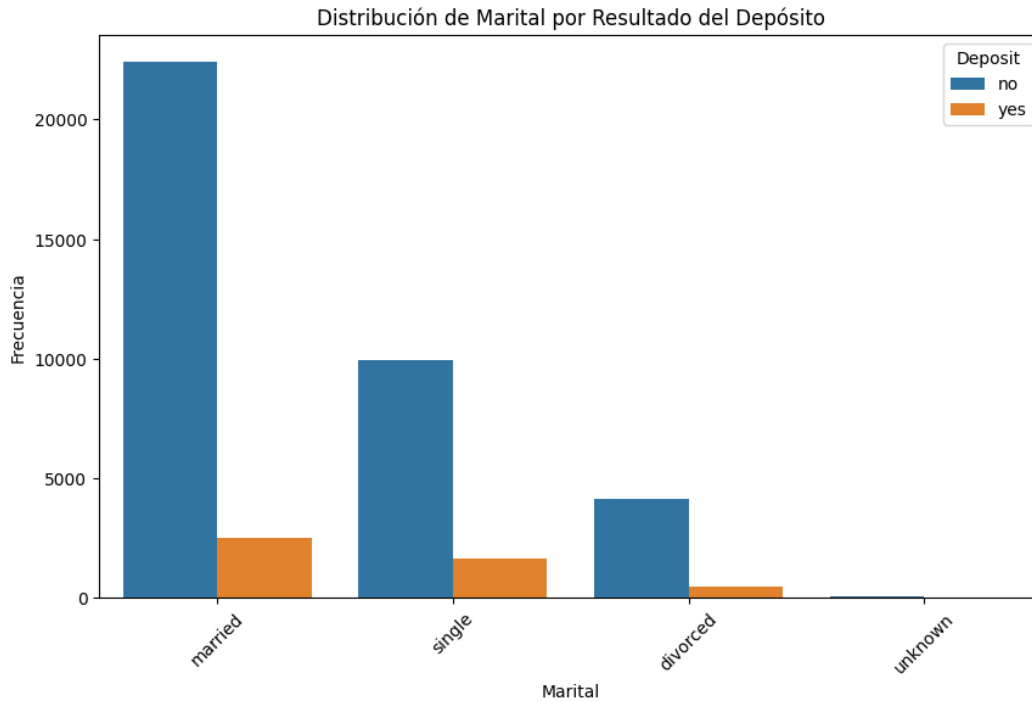
Ahora exploraremos aquellos campos que podrían ser relevantes en la decisión de suscribirse a un depósito a plazo.

Cuántos tienen un depósito:

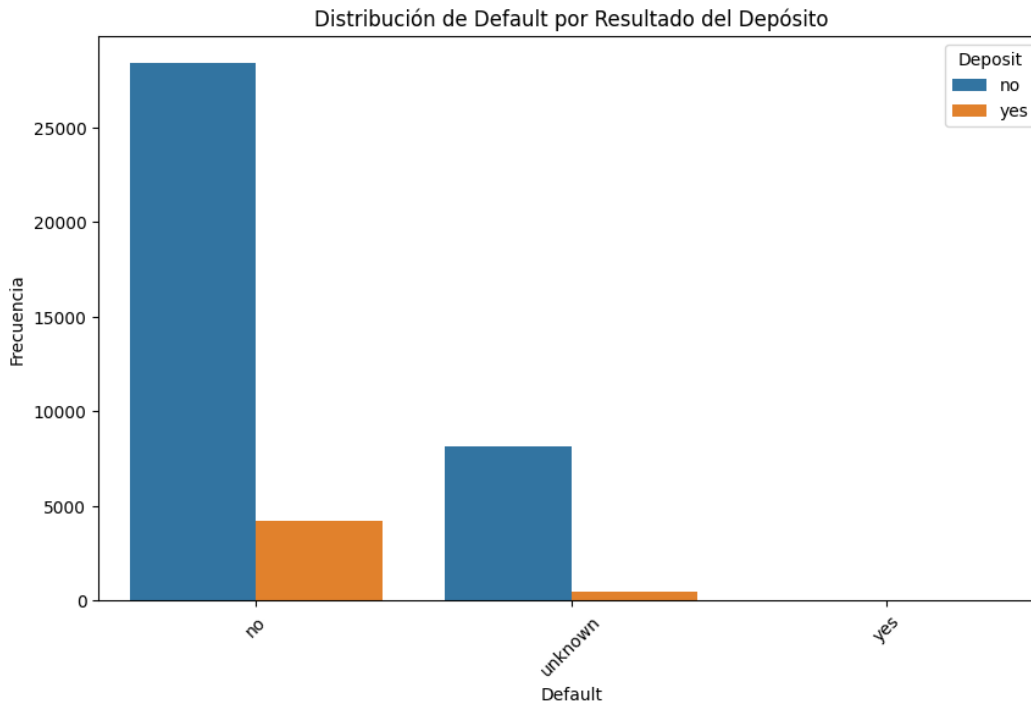


Para los clientes en la base de datos, el estado civil de soltero presenta un porcentaje más

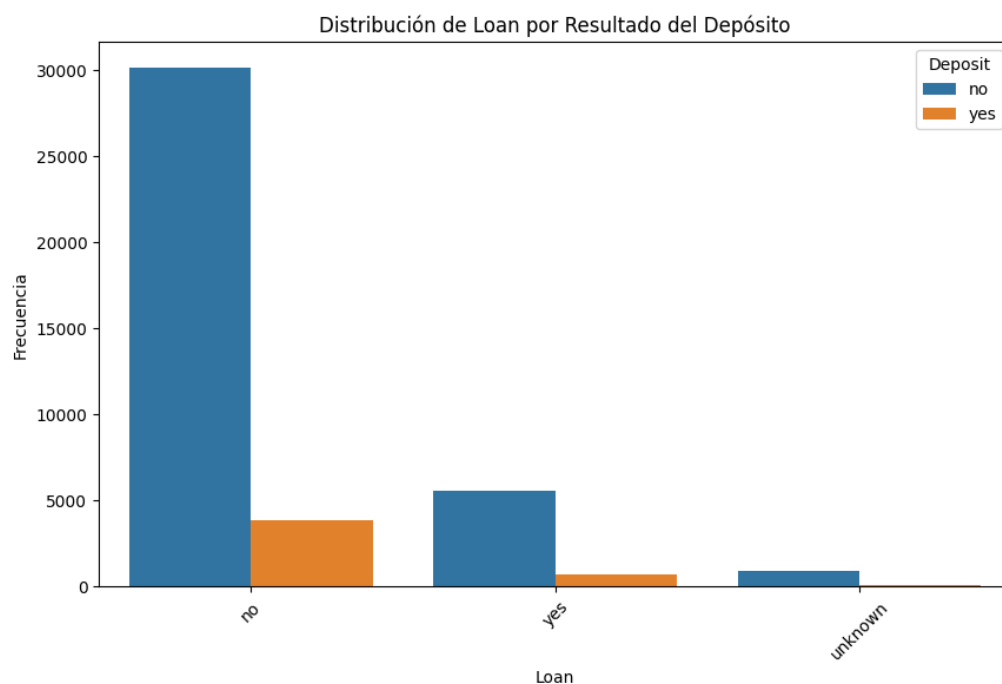
alto de acceso al depósito en comparación con aquellos que están casados o divorciados.



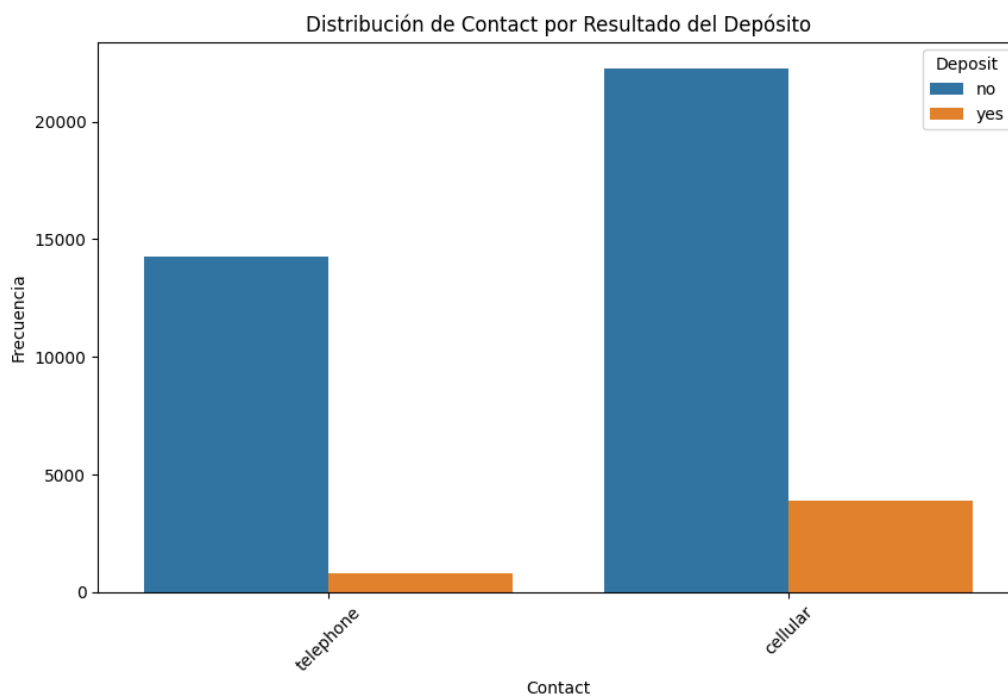
También es importante tener en cuenta que las personas que se encuentran en mora no son contactadas. Esto sugiere que no solo se trata de estar en mora, sino que aparentemente estas personas tampoco pueden acceder a este tipo de beneficios, lo que puede influir en su decisión de suscribirse a un depósito a plazo.



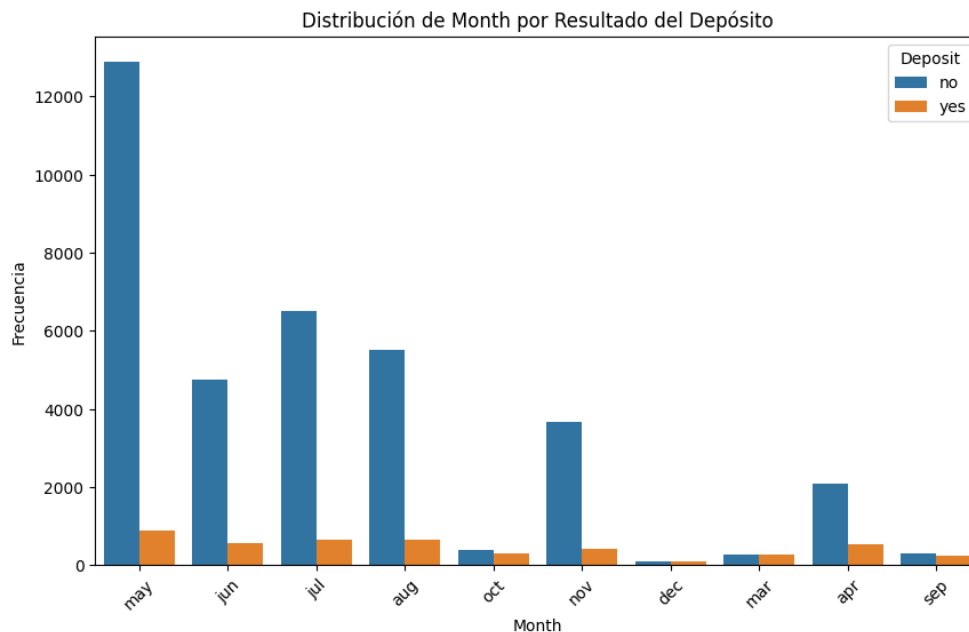
Es posible que la falta de un crédito personal influya en la disposición de una persona para aceptar un depósito a plazo.



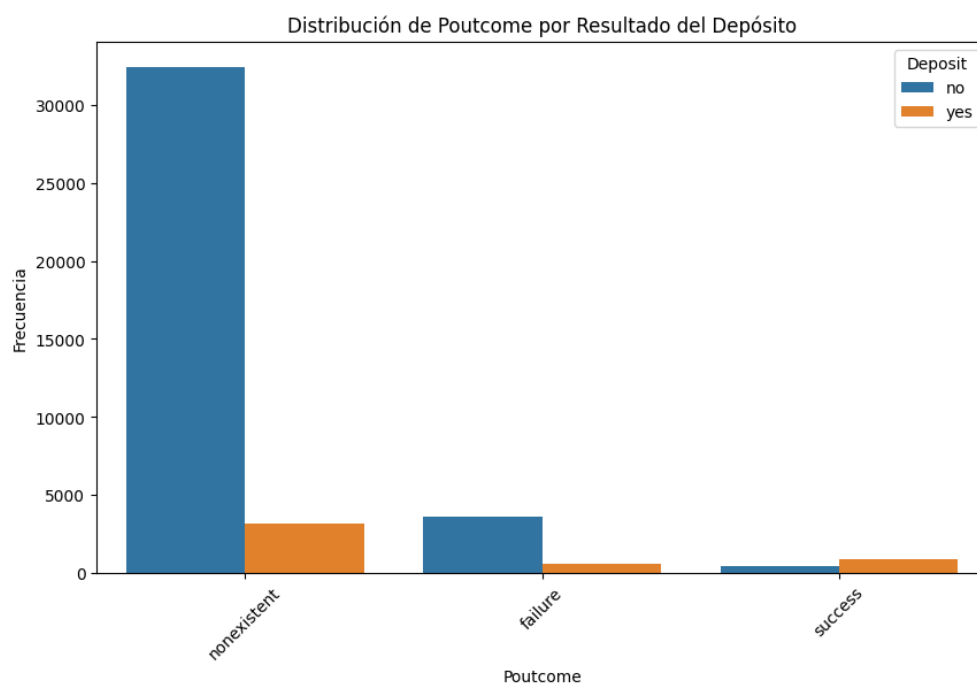
Es probable que el contacto por celular facilite el acceso a la persona y, por ende, aumente la probabilidad de que acepte el depósito.



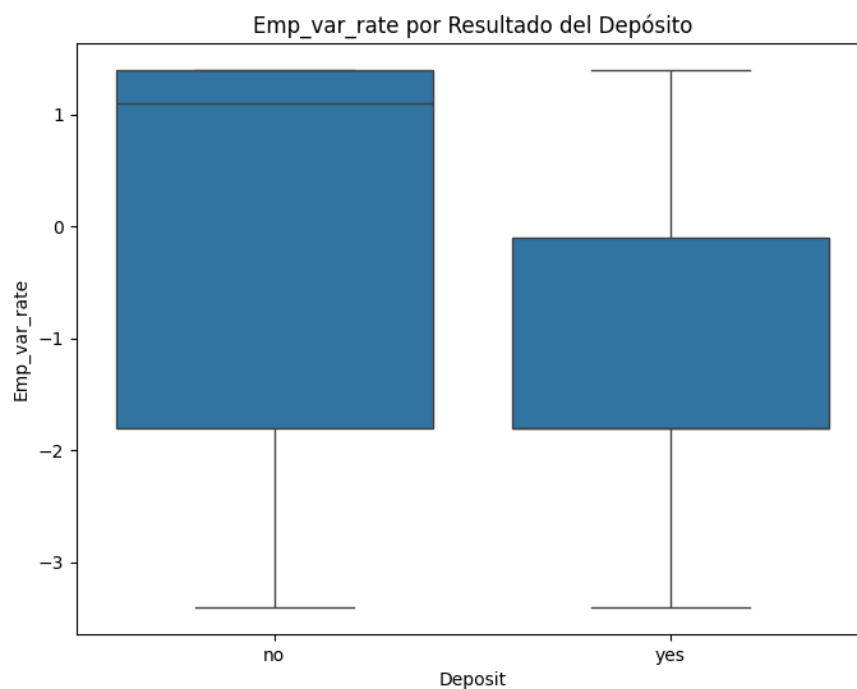
Aunque mayo es el mes con más contactos, abril destaca como uno de los meses más propicios para que los clientes accedan al depósito.



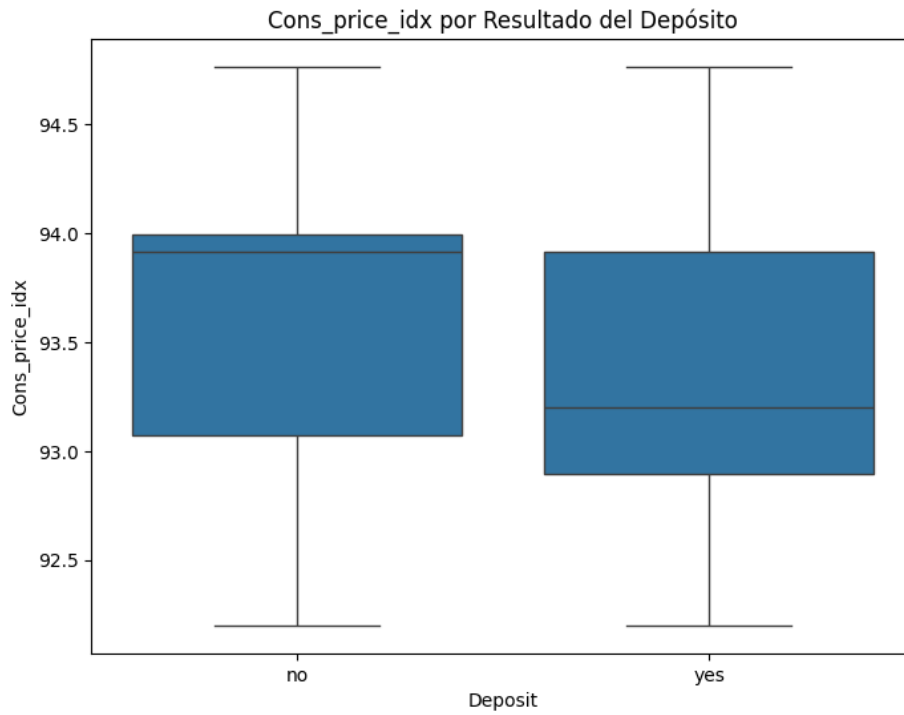
Aquí podemos observar que, si un cliente accedió al producto en una campaña anterior, tiene una probabilidad notablemente alta de acceder al nuevo. Este porcentaje es incluso más alto que cualquier otro dato en términos proporcionales.



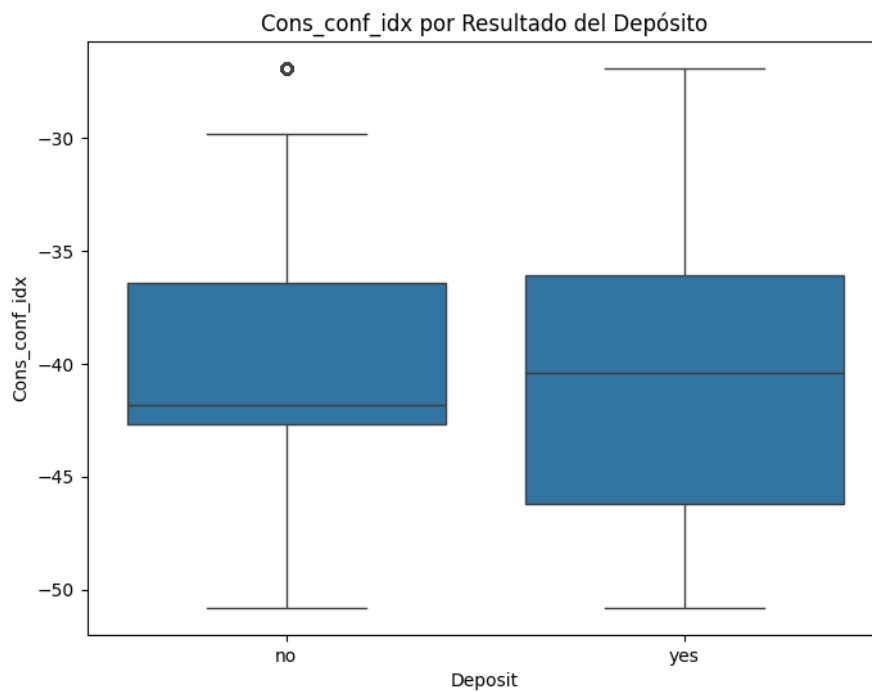
La tasa de variación del empleo también desempeña un papel importante, ya que cuanto más alta sea, más difícil puede resultar que un cliente acepte. Por otro lado, cuando se mantiene en un término medio, ese momento se vuelve óptimo para acceder al cliente.



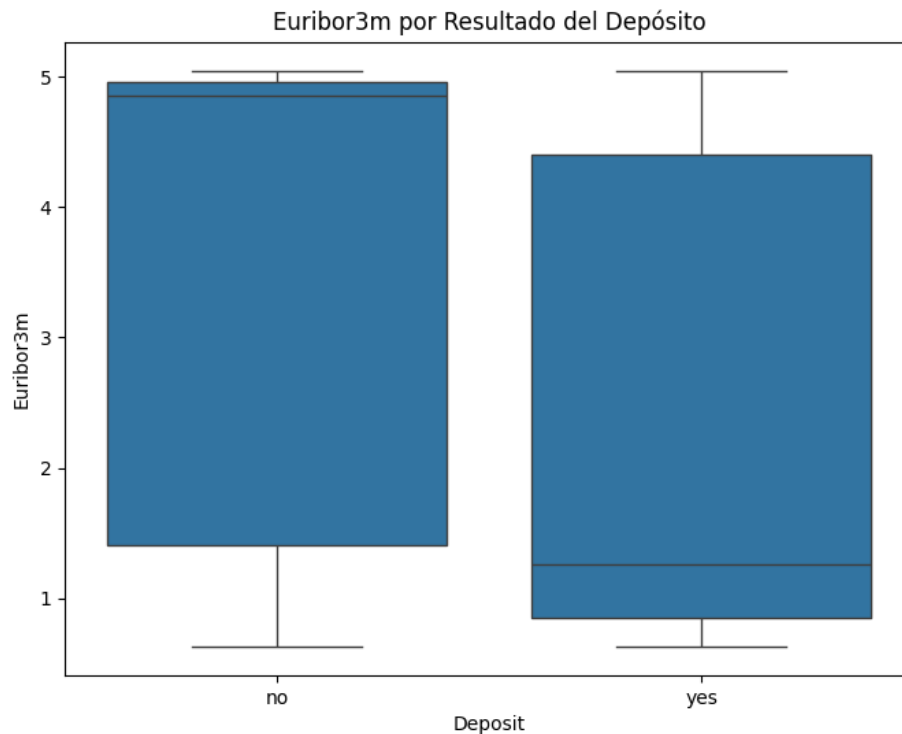
También es importante destacar que si el Índice de Precios al Consumidor (IPC) es más bajo, se presentan más oportunidades para acceder al cliente.



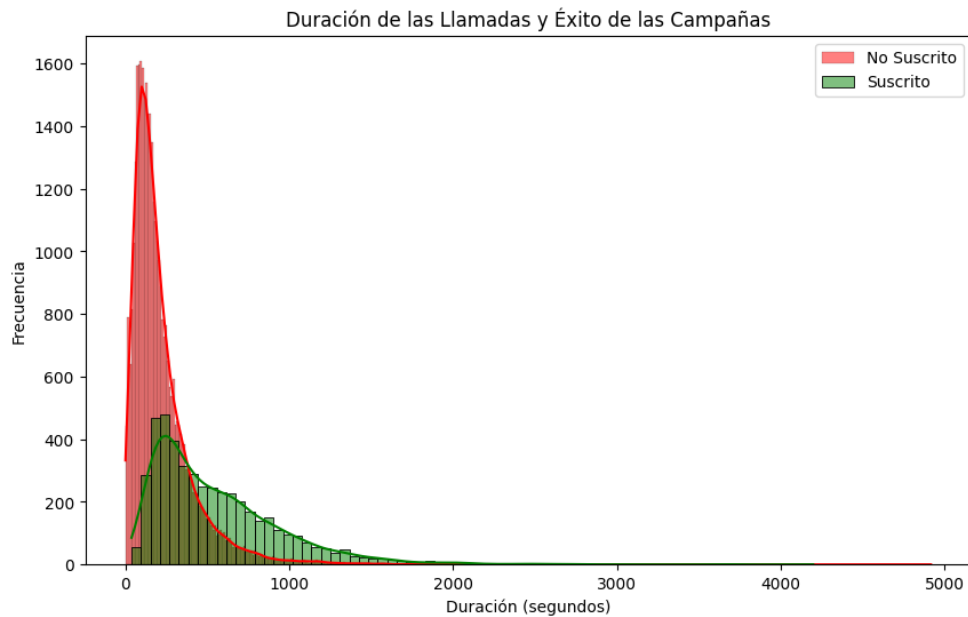
Asimismo, el Índice de Confianza también sugiere que valores intermedios son óptimos para ofrecer el producto al cliente.



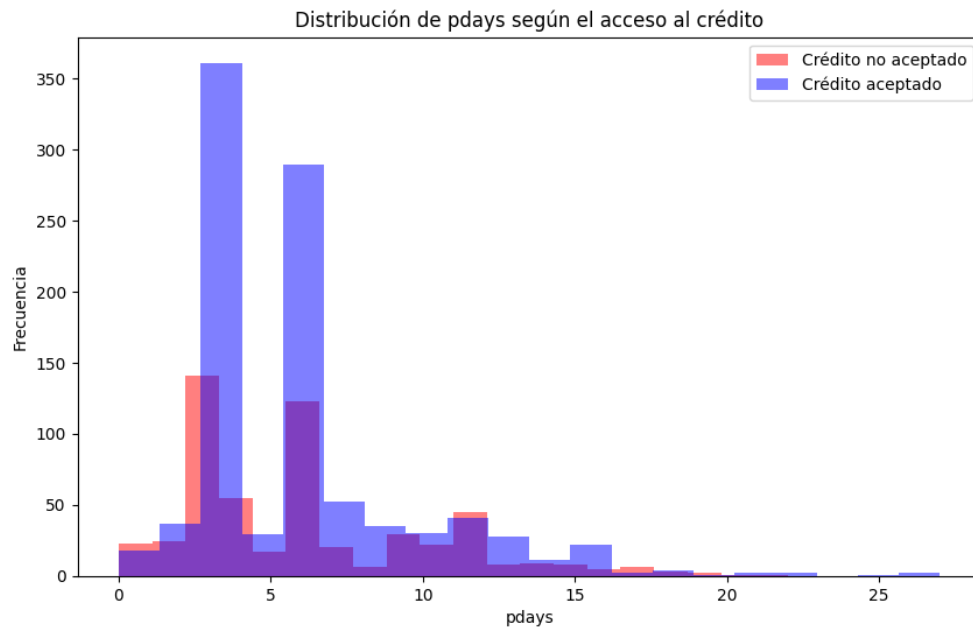
También es evidente que cuanto más baja sea la tasa Euribor, más fácil resulta la oferta al cliente



La duración de las llamadas también desempeña un papel crucial. Aunque pueden extenderse considerablemente en nuestros datos, se evidencia que entre los 500 segundos (8 minutos) y cerca de los 1500 segundos (25 minutos), la suscripción del cliente es mucho más efectiva. Este fenómeno puede deberse a que en ese intervalo se desarrollan conversaciones más interesantes, lo que permite al agente persuadir al cliente con mayor facilidad.

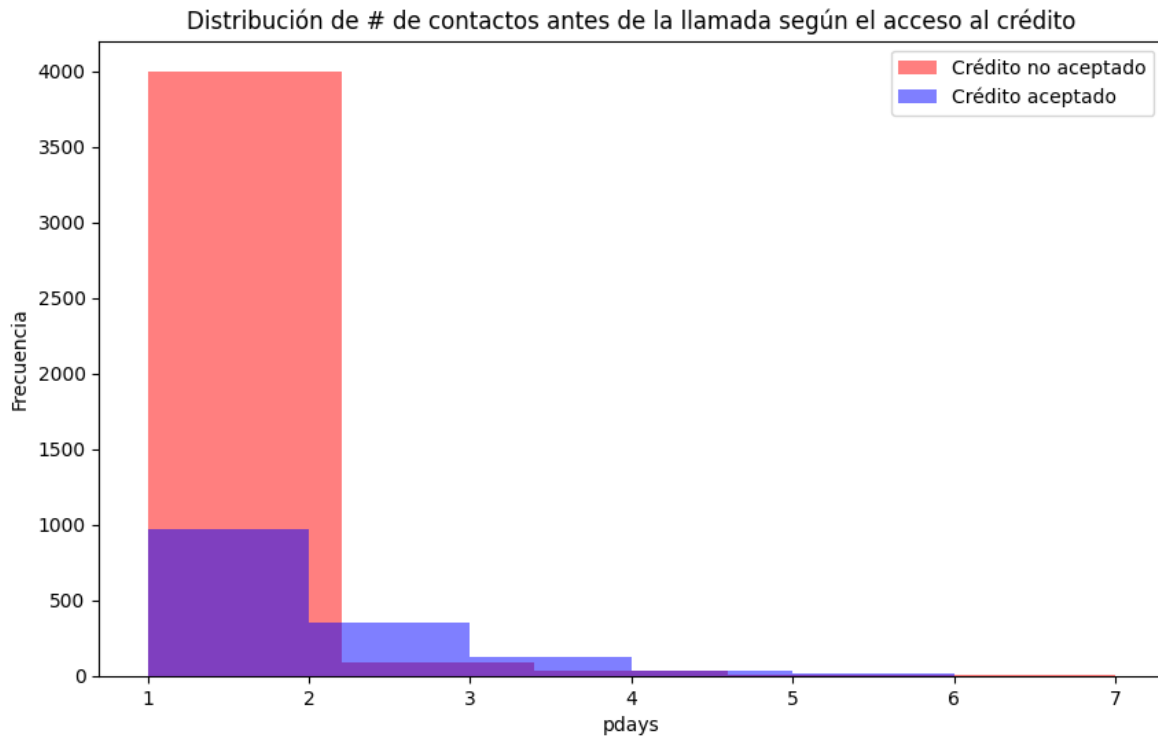


Cuando eliminamos a todos los clientes que no han sido contactados, podemos observar datos aún más interesantes, como en el caso de "pdays". Esto sugiere que entre aproximadamente 3 a 7 días después de haber sido contactado por última vez, la oferta del depósito puede ser más satisfactoria para el cliente.

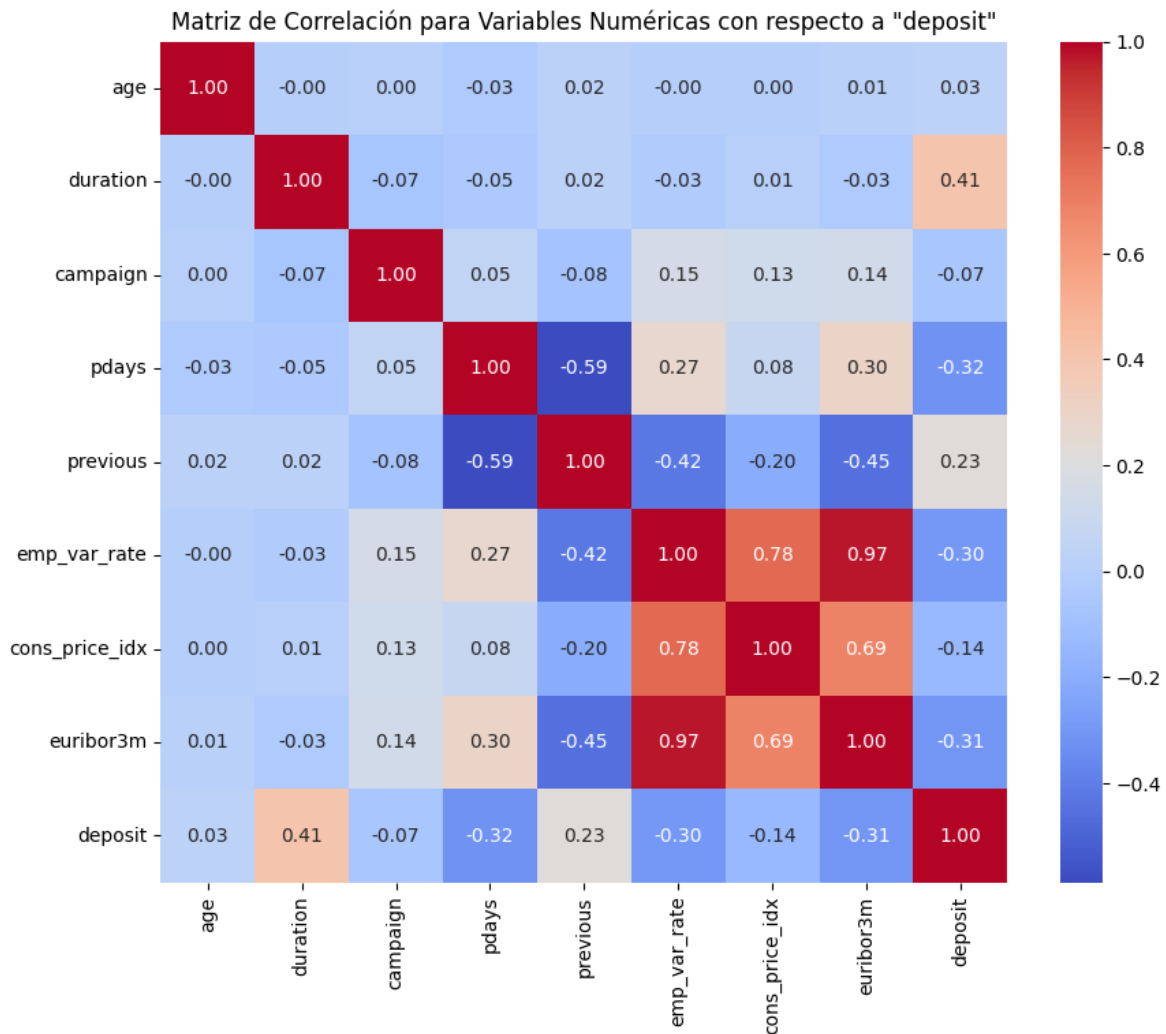


Es importante tener en cuenta que, si bien estos datos son llamativos, al eliminar a los clientes no contactados, la cantidad de datos para el análisis se reduce significativamente, de 41176 a 1515 datos. Esto puede generar sesgos o incompletitudes en el análisis debido a la reducción del tamaño de la muestra.

Asimismo, esto se refleja en el número de contactos realizados antes de la campaña actual. Al centrarnos en los clientes que vuelven a ser contactados, observamos que entre el segundo y cuarto contacto hay más probabilidades de que el cliente acepte el depósito. Sin embargo, es importante recordar que esta conclusión se basa en una muestra reducida de datos y puede no ser representativa del conjunto completo de clientes.



Por último, nuestro mapa de correlación nos revela estas relaciones y la influencia significativa que tienen las variables descritas en la decisión de aceptar el depósito. Es importante destacar que este análisis se basa en el conjunto de datos completo, lo que significa que algunas relaciones pueden no ser tan claras como en análisis anteriores debido a la inclusión de todas las variables numéricas.



Bank Term Deposit Predictions:

Por ultimo el conjunto de datos de depósitos, este no se diferencia tanto al anterior, la diferencia radica en muy pocas columnas y el tipo de dato, por lo que el análisis no va a ser tan distinto y los campos son muy similares.

Campos:

- **Age:** La edad del cliente.
- **Job:** La ocupación del cliente.
- **Marital:** El estado civil del cliente.
- **Education:** El nivel educativo del cliente.
- **Default:** Indica si el cliente tiene o no créditos en mora.
- **Balance:** El saldo de la cuenta del cliente.
- **Housing:** Indica si el cliente tiene o no un préstamo de vivienda.
- **Contact:** El método utilizado para contactar al cliente (por ejemplo, teléfono, celular).

- **Day:** El día del mes en que se realizó el último contacto con los clientes.
- **Duration:** La duración (en segundos) del último contacto con los clientes durante una campaña.
- **Campaign:** Número de contactos realizados durante esta campaña para cada cliente.
- **pdays:** Número de días transcurridos desde el último contacto desde la campaña anterior.
- **previous:** Número de contactos realizados antes de esta campaña y para este cliente.
- **poutcome:** Resultado de la campaña de marketing anterior.
- **y / deposit:** Indica si el cliente ha suscrito un depósito a plazo.

En este caso, identificamos datos repetidos que eran exactamente iguales, lo que sugiere una alta probabilidad de que se trate del mismo cliente. Estos duplicados fueron eliminados, lo que resultó en la eliminación de solo 12 datos. Además, no se encontraron valores atípicos ni fue necesario realizar ningún otro ajuste en los datos.

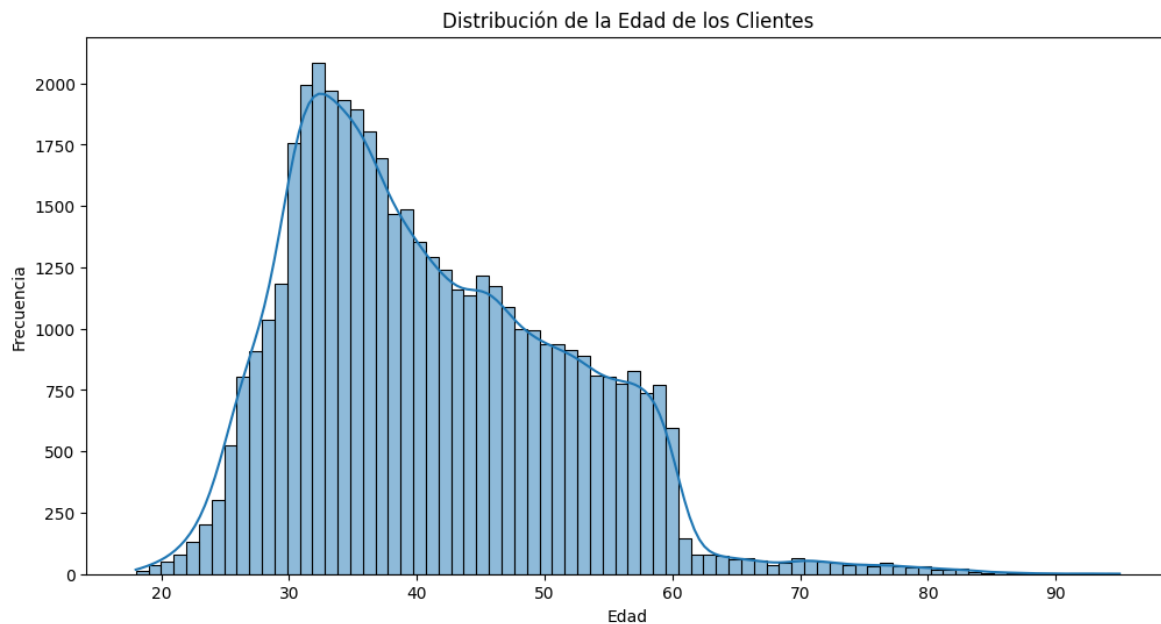
Análisis:

	age	balance	duration	campaign	pdays	previous
mean	40.94	1362.27	258.16	2.76	857.57	0.57
std	10.62	3044.77	257.53	3.10	303.25	1.91
min	18.00	-8019.00	0.00	1.00	1.00	0.00
25%	33.00	72.00	103.00	1.00	999.00	0.00
50%	39.00	448.00	180.00	2.00	999.00	0.00
75%	48.00	1428.00	319.00	3.00	999.00	0.00
max	95.00	102127.00	4918.00	63.00	999.00	58.00

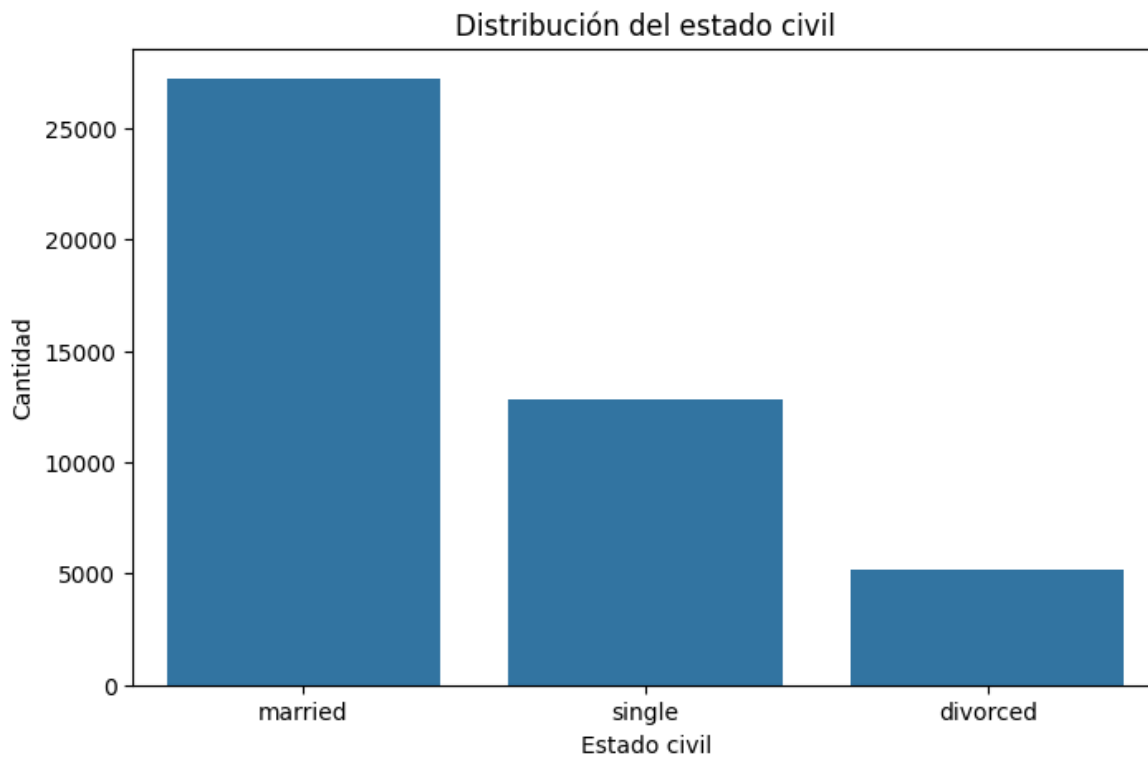
Aquí, la única peculiaridad notable es que el saldo también puede ser negativo. Por lo demás, comparte las mismas "anomalías" que el conjunto de datos anterior.

Análisis demográfico:

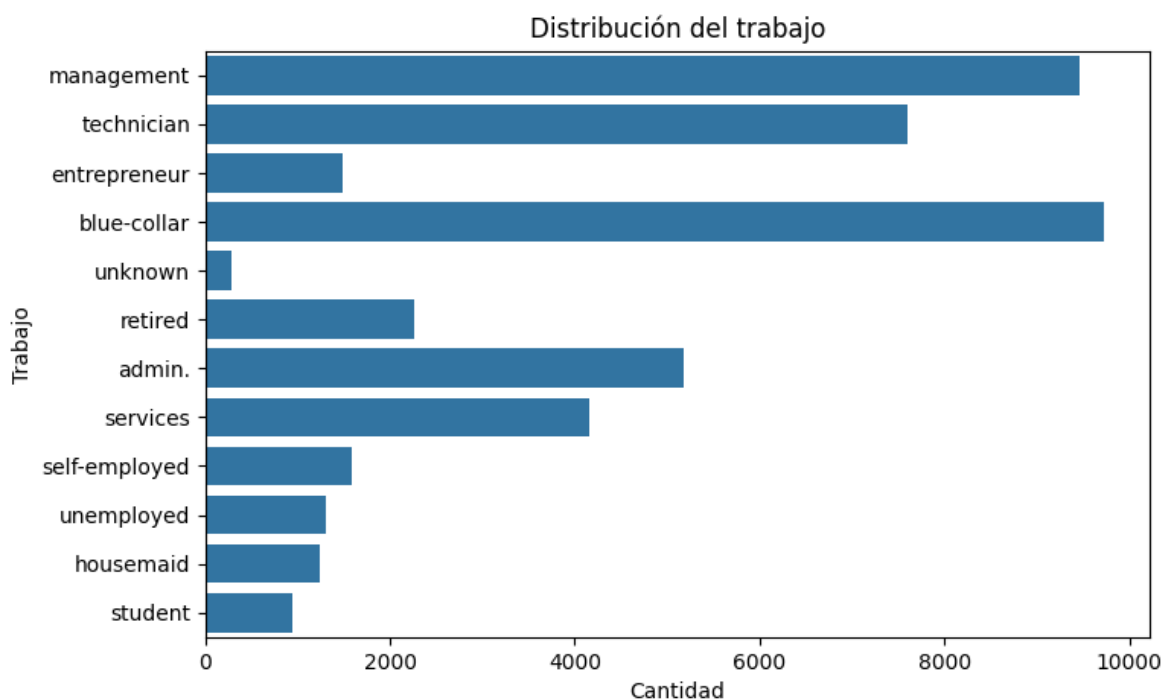
La distribución de la edad se extiende un poco más que en los datos anteriores.



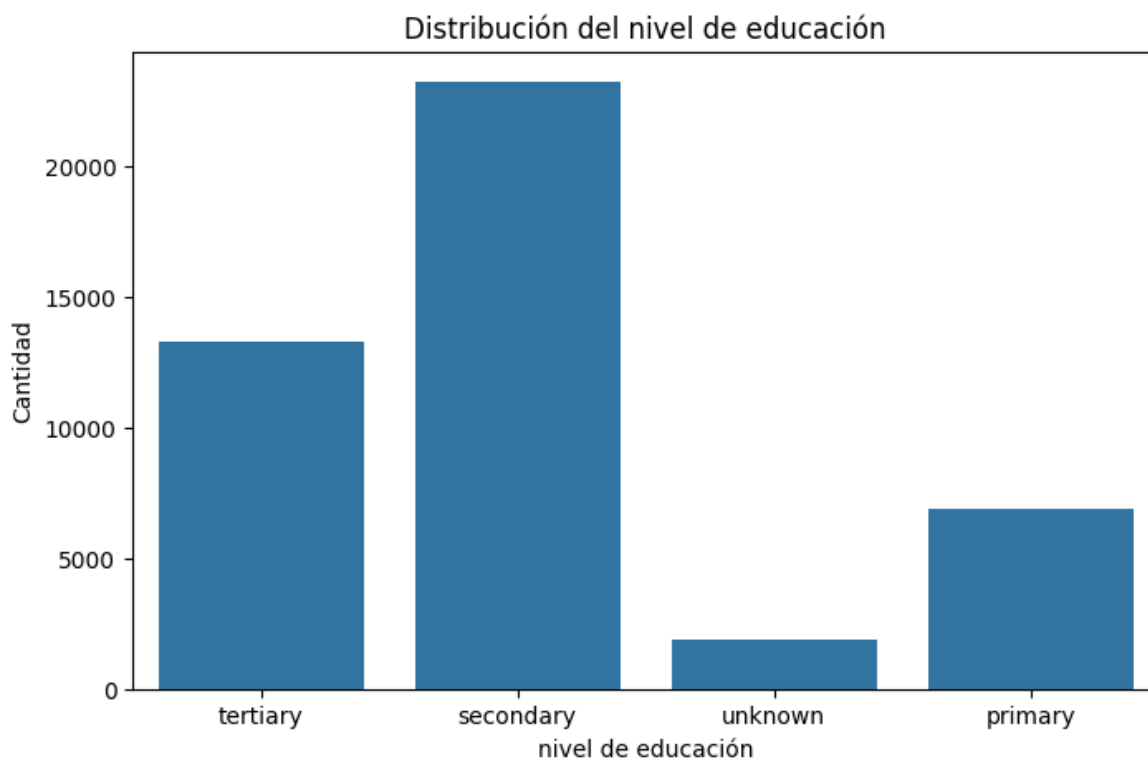
Estado civil:



Distribución de trabajo:

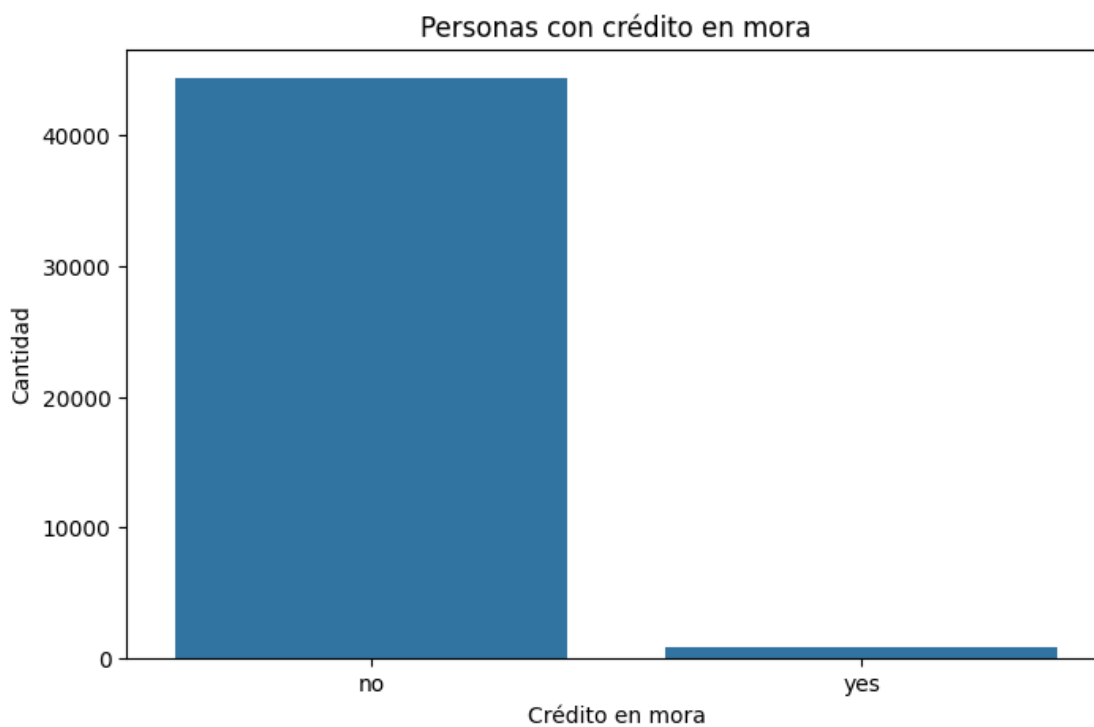


Distribución de nivel educativo:

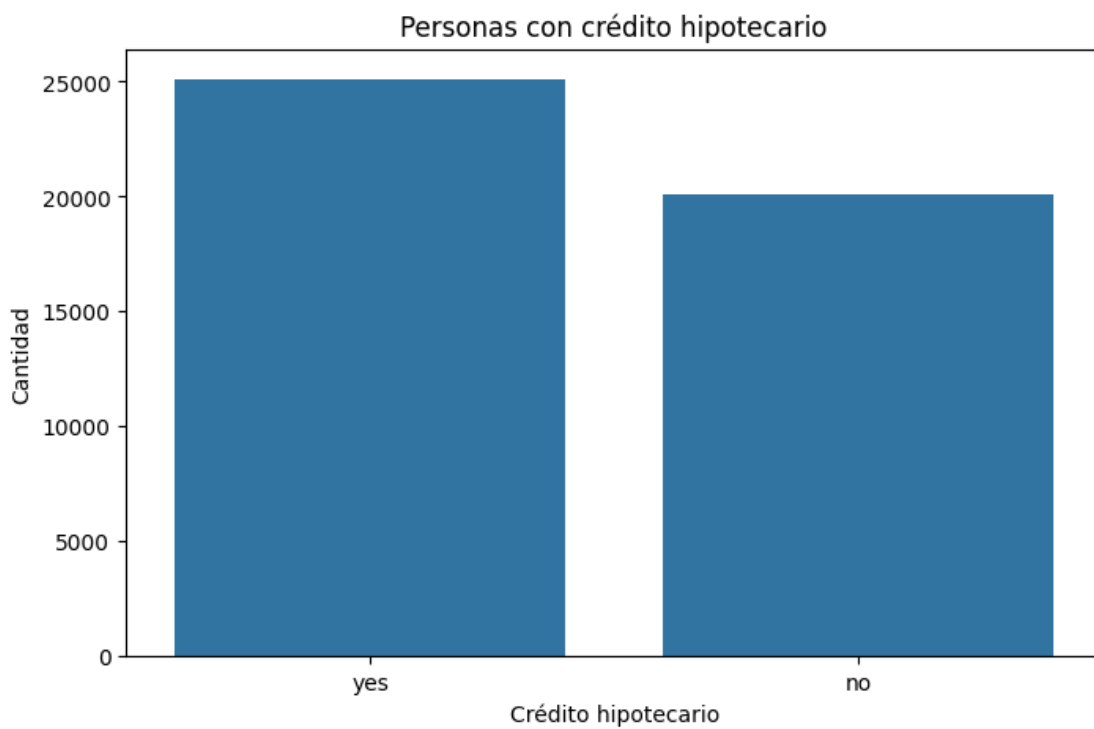


Datos Bancarios

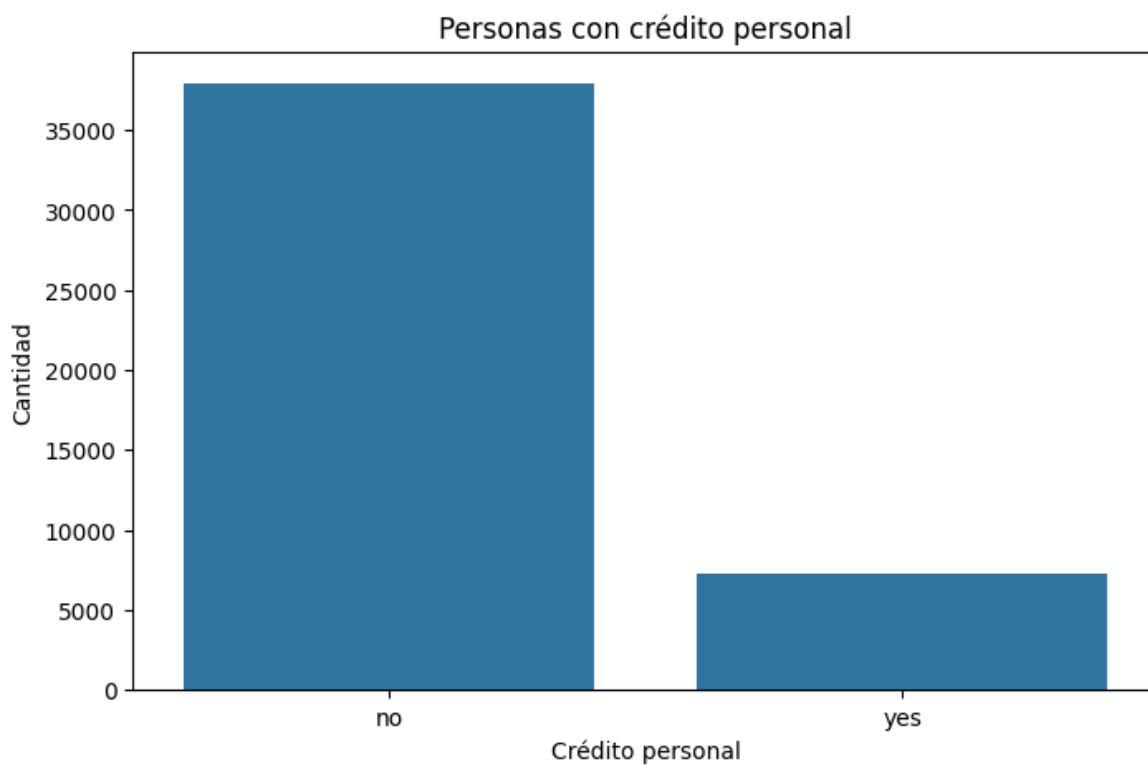
Personas con créditos en mora:



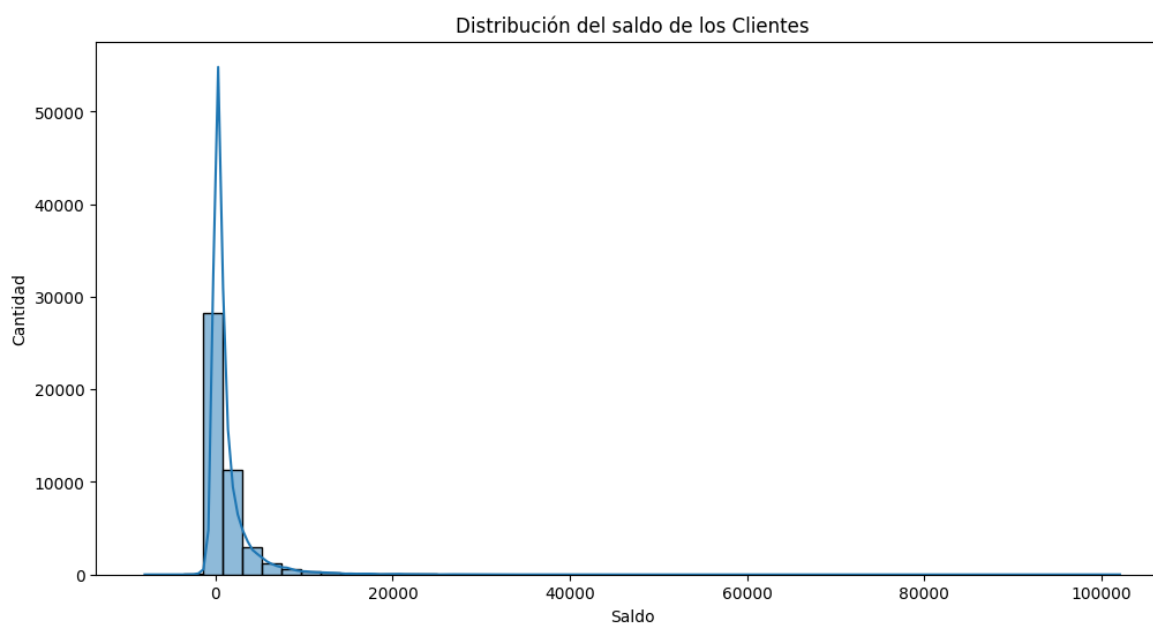
Personas con créditos hipotecarios:



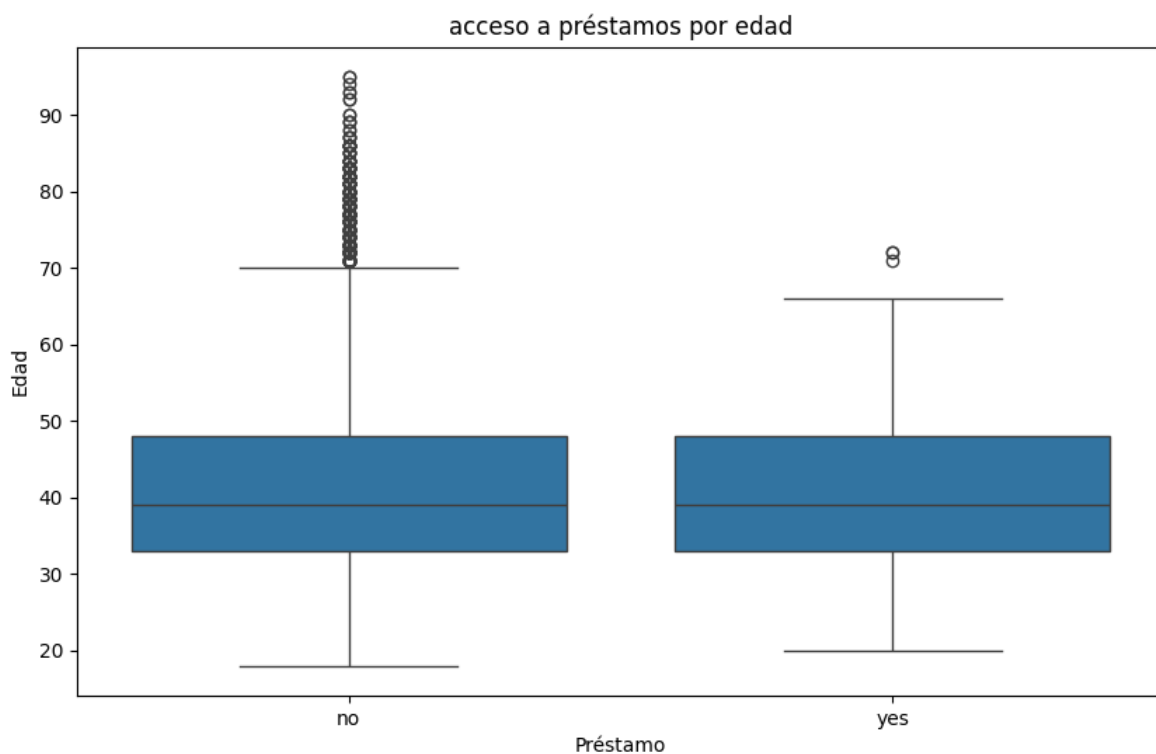
Crédito personal:



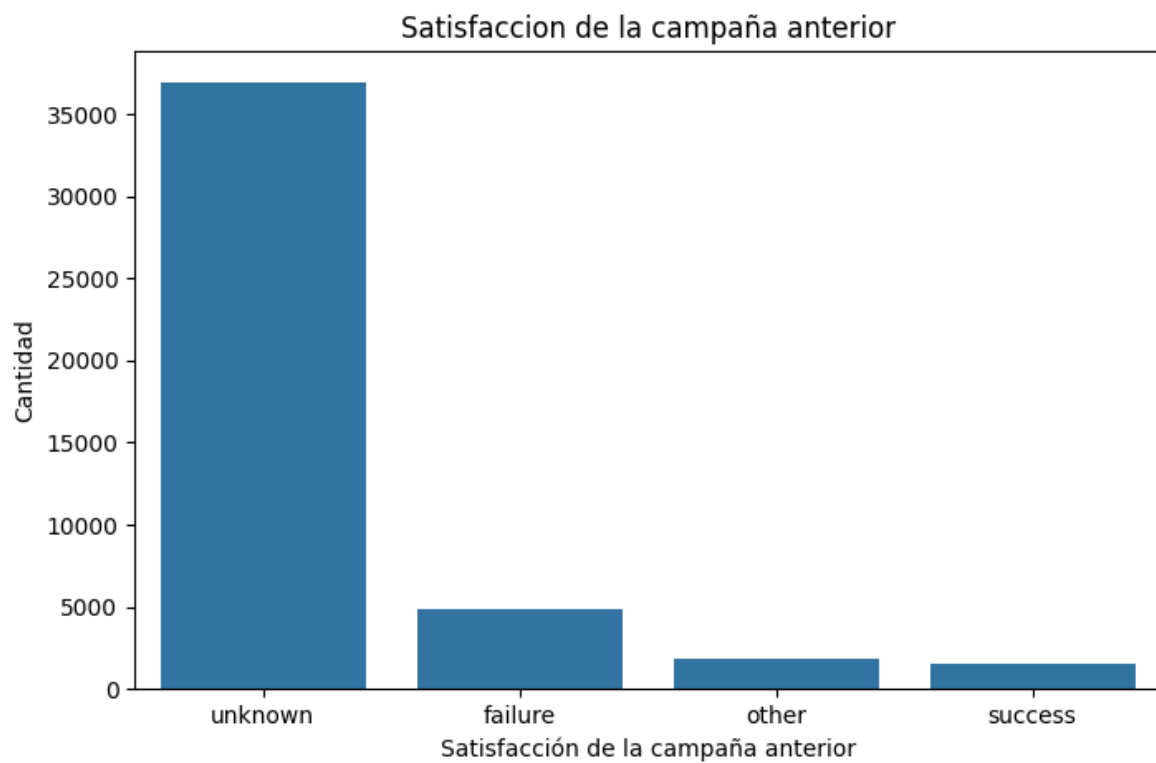
Distribución del saldo de los clientes:



Acceso a prestamos por edad:

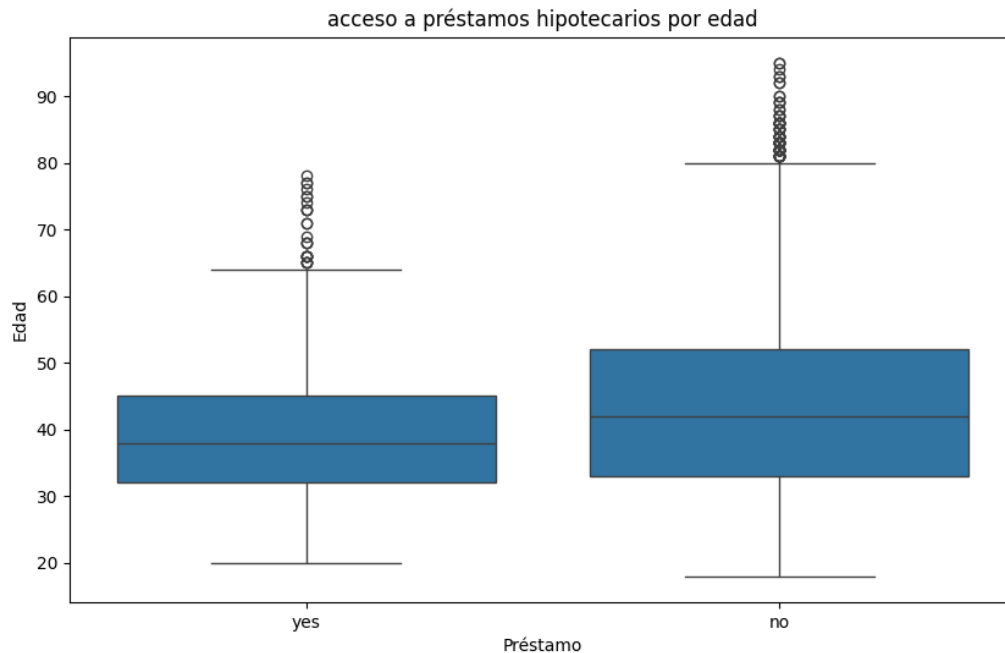


Satisfacción con la campaña anterior:



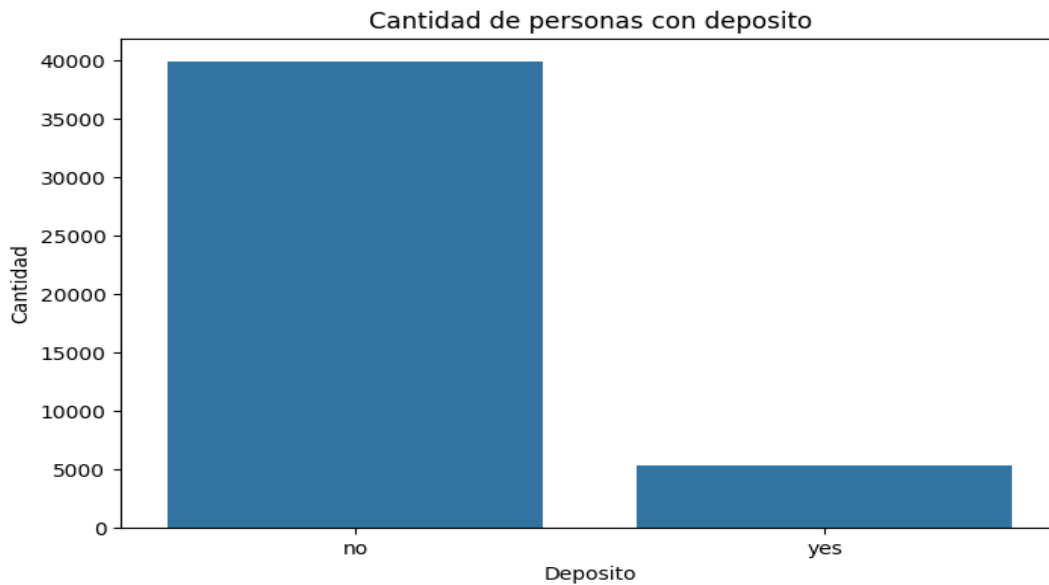
Acceso a préstamos hipotecarios por edad:

Es interesante observar cómo hay un rango de personas jóvenes que acceden a este tipo de préstamos hipotecarios. Esto también refleja la capacidad de la empresa para otorgar préstamos y su disposición para atender a una variedad de clientes en diferentes etapas de la vida.

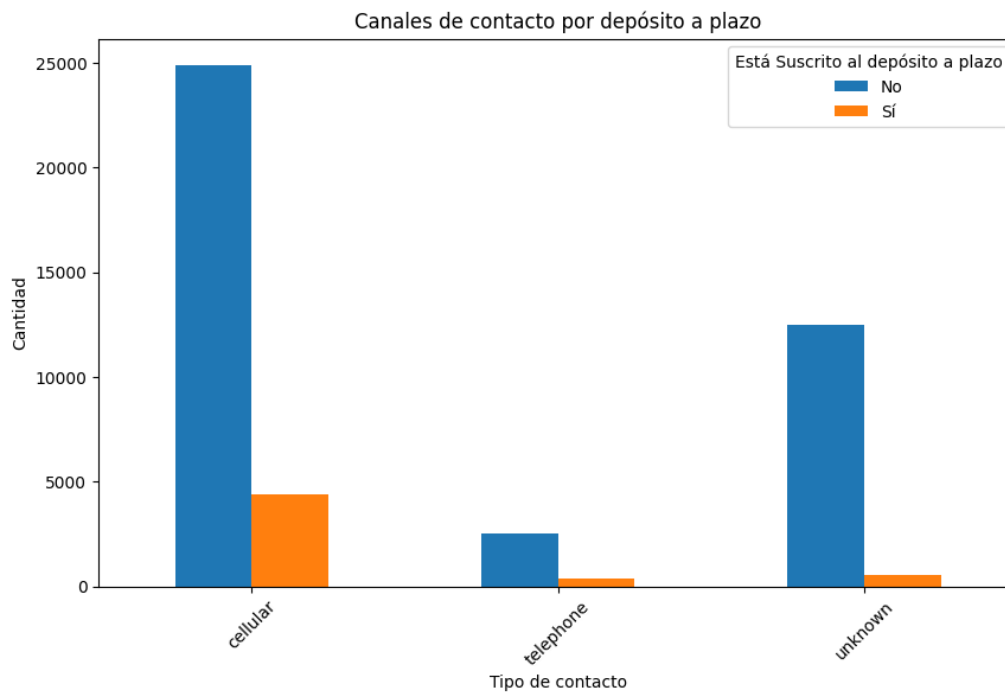


Variables importantes en la decisión de aceptación de depósito:

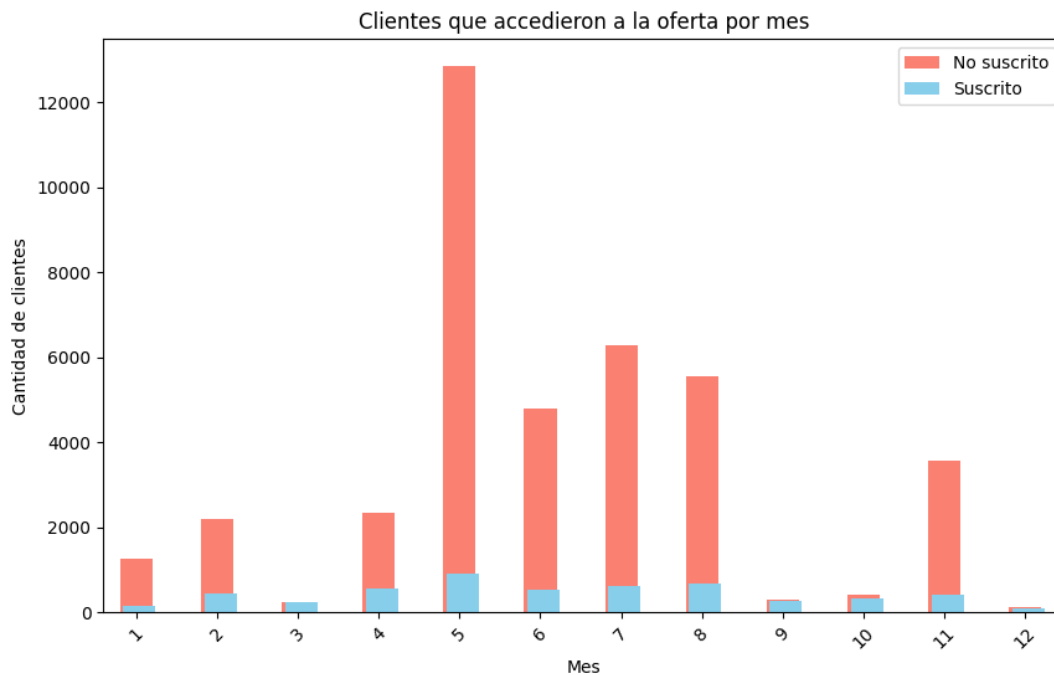
De los 45211 registros en total, 5289 clientes accedieron al depósito, lo que representa aproximadamente el 12% del total de clientes. Esto sugiere que el depósito tiene un nivel de aceptación considerable entre la base de clientes.



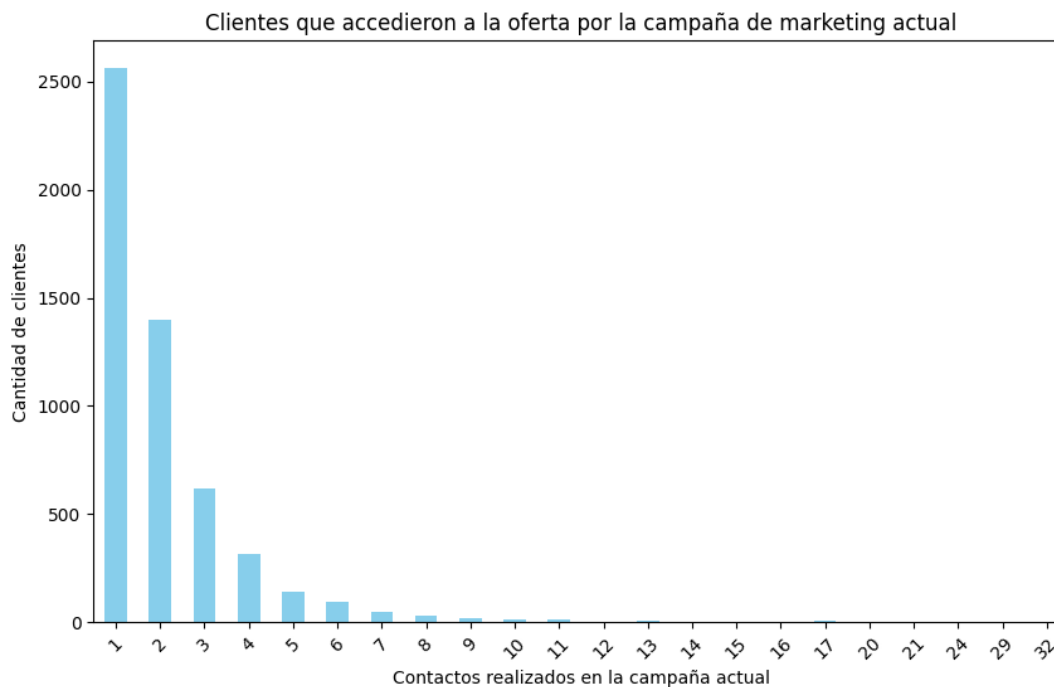
Nuevamente, el canal de atención puede facilitar el contacto y por lo tanto la venta



Como se puede observar, en los meses con menos contactos también se logró una mayor efectividad. Esto sugiere que la cantidad de contactos no siempre refleja la calidad de las ofertas, y que, en ocasiones, una estrategia más selectiva puede conducir a mejores resultados en términos de aceptación de depósitos.

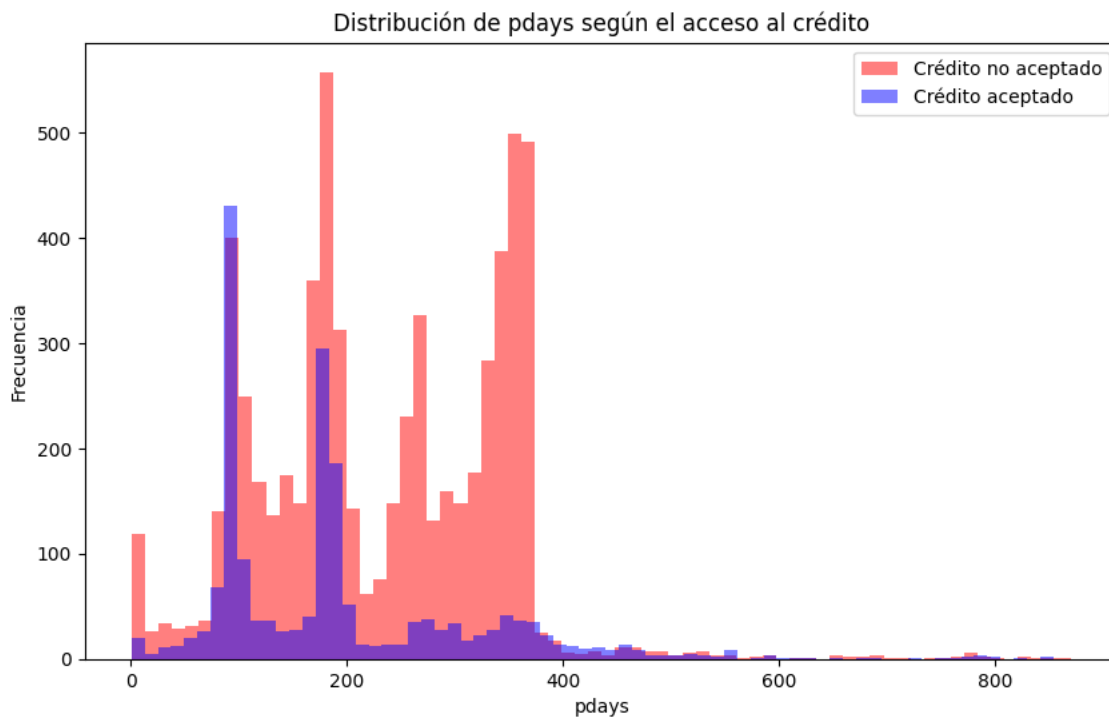


También se evidencia que aquellos clientes que son contactados entre 1 y 4 veces tienen más oportunidades de realizar la venta. Esto sugiere que realizar más llamadas no garantiza necesariamente que se concrete la venta en algún momento.

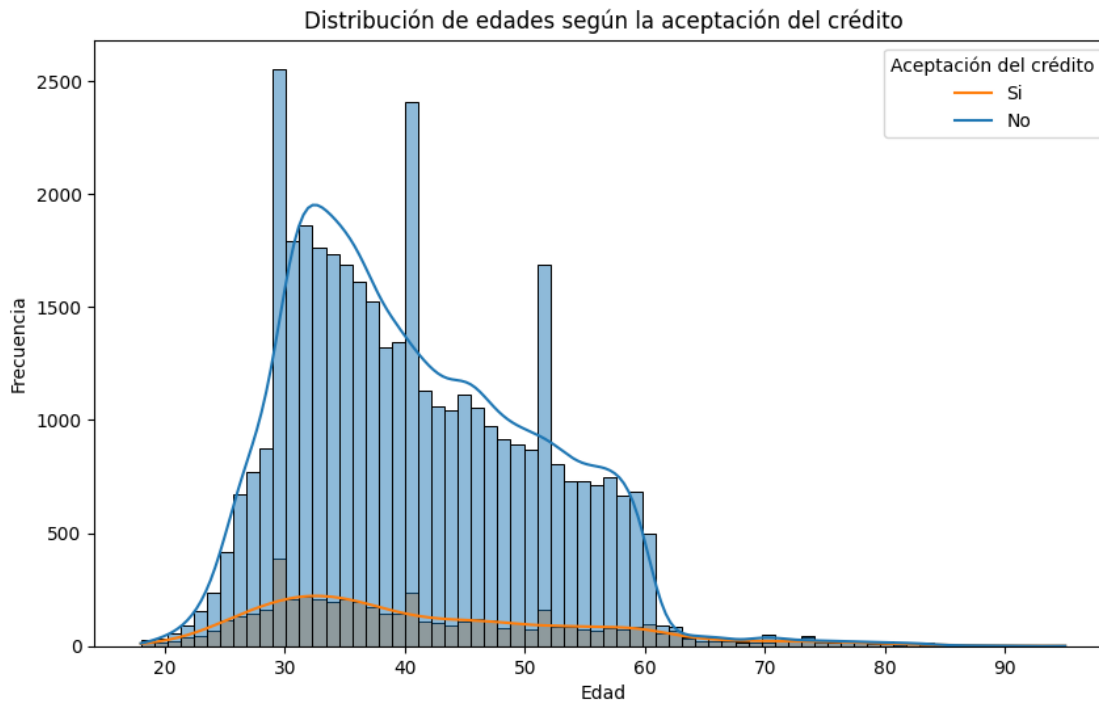


También es evidente que el número de días transcurridos desde el último contacto de la campaña anterior puede tener una influencia significativa. Se observa que cuando han transcurrido menos de 200 días, la estadística de aceptación es más alta que en cualquier

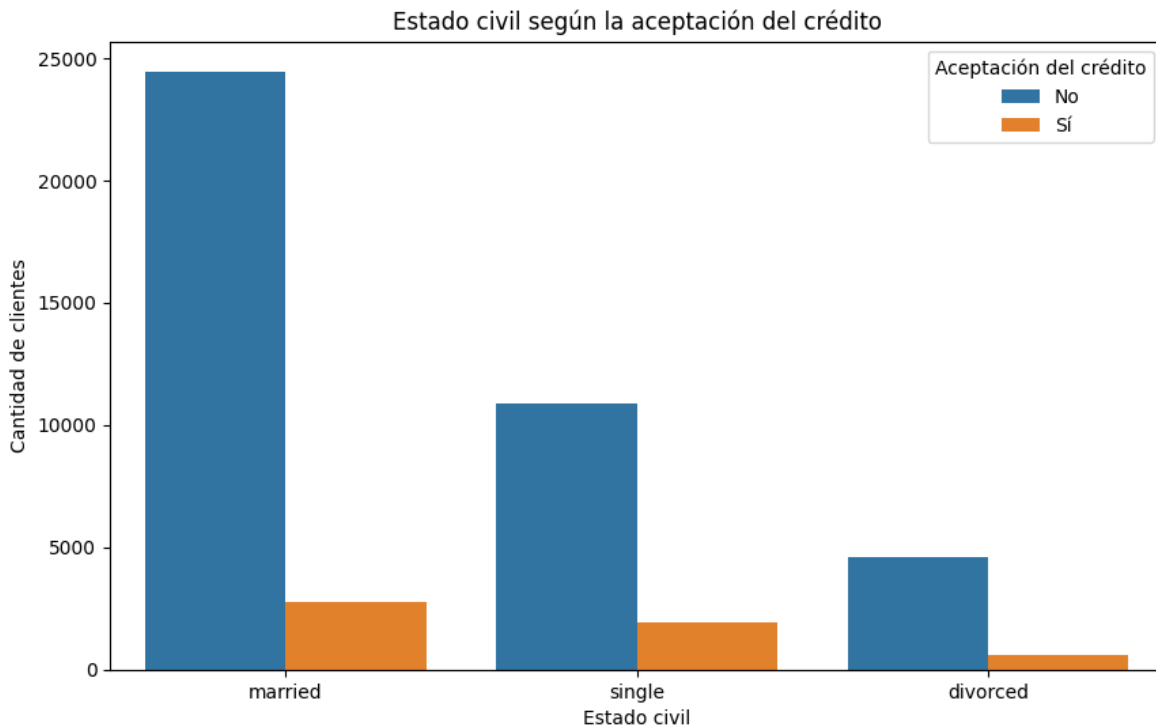
otro momento, lo que sugiere que existe una ventana de tiempo óptima para retomar el contacto con el cliente.



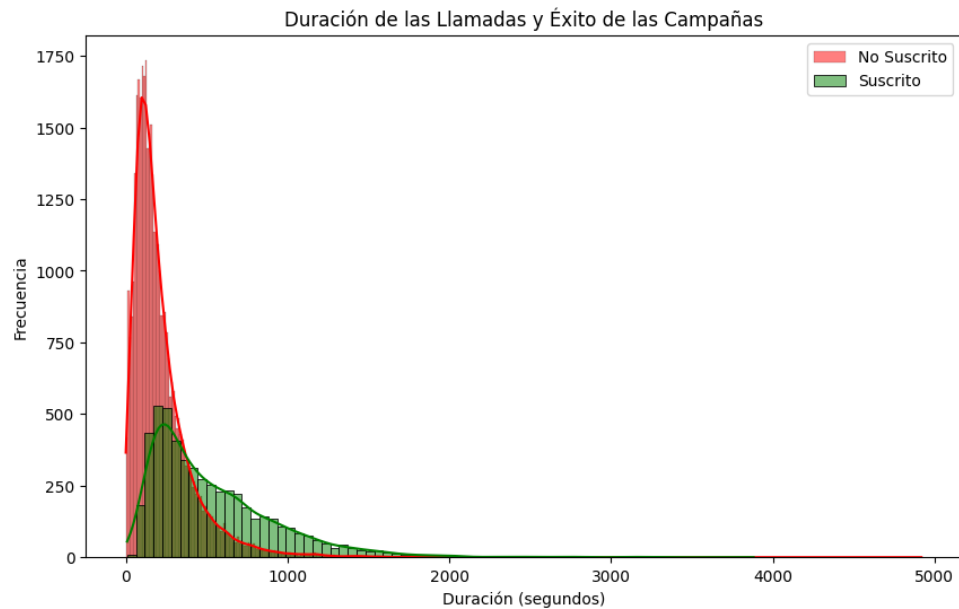
En este caso, se observa que los adultos mayores de 60 años tienen el porcentaje más alto de aceptación.



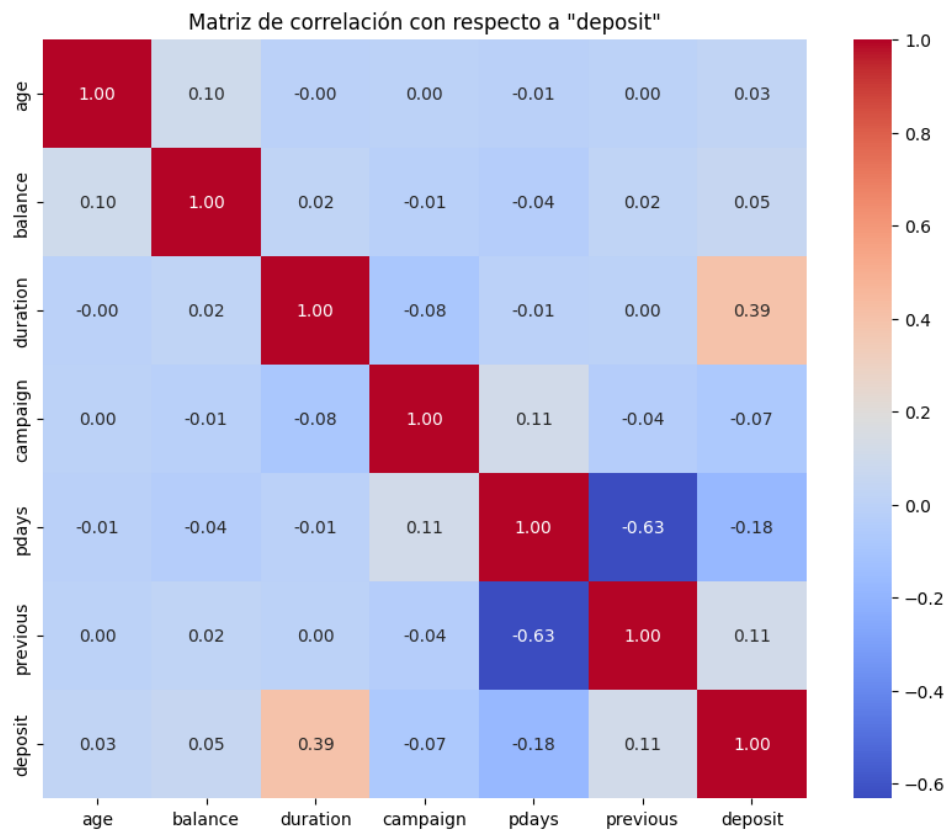
Al igual que en el caso anterior, los clientes en estado marital "soltero" tienen un porcentaje de aceptación más alto.



Y de manera similar al caso anterior, la duración de la llamada también desempeña un papel importante en la aceptación del depósito.



Nuevamente observamos correlaciones muy similares, especialmente en relación con la duración de las llamadas. Estos hallazgos respaldan aún más la idea de que podemos aprovechar estos datos para optimizar nuestras campañas y realizar ventas más efectivas.



Conclusiones

El análisis reveló una serie de factores significativos que influyen en la aceptación de depósitos a plazo por parte de los clientes. Entre ellos, se destacan:

- La edad: Se observó que los adultos mayores de 60 años tienen el porcentaje más alto de aceptación.
- Estado civil: Los clientes solteros mostraron un porcentaje de aceptación más alto que aquellos en otras categorías maritales.
- Duración de la llamada: Se encontró que una duración más prolongada de la llamada está asociada con una mayor probabilidad de aceptación del depósito.
- Contacto previo: Los clientes contactados entre 1 y 4 veces mostraron más probabilidades de aceptar el depósito.
- Días transcurridos desde el último contacto: Se identificó que un período de menos de 200 días desde el último contacto de la campaña anterior se asocia con una mayor tasa de aceptación.
- Ocupación: Se observaron patrones interesantes en la aceptación según la ocupación del cliente, con ciertas categorías mostrando una mayor probabilidad de aceptación que otras.

Estos hallazgos ofrecen información valiosa para mejorar las estrategias de marketing y ventas, permitiendo una segmentación más precisa de los clientes y una mayor efectividad en las campañas de promoción de depósitos a plazo.

Abandono

Se observó que un 25% de los clientes abandonaron el banco, lo que podría ser preocupante dependiendo del período de tiempo considerado. Se identificaron patrones interesantes relacionados con la edad, la actividad de los clientes, la situación crediticia y otros factores que podrían estar asociados con la decisión de los clientes de abandonar el banco.

Se recomienda implementar estrategias dirigidas específicamente hacia las mujeres, así como iniciativas para el cuidado y la retención de clientes a lo largo del tiempo. Ofrecer beneficios adicionales y lanzar campañas para atraer nuevos clientes puede contribuir a mejorar la reputación del banco y atraer una mayor cantidad de clientes que se sientan fieles a la institución financiera.