

Desgaste de clientes bancarios

¿Cuáles son los indicios que advierten que un cliente abandonará el servicio de tarjeta de crédito?

AUTOR: Barbara Iummato

AGENDA

- 01 | Contexto y Audiencia
- 02 | Hipótesis/Preguntas de Interés
- 03 | Metadata
- 04 | Análisis Exploratorio
- 05 | Insights y Recomendaciones



CONTEXTO Y AUDIENCIA

Contexto

Actualmente en la Argentina más de la mitad de la población no cuenta con una cuenta bancaria. Adicionalmente cerca del 85% de los clientes de bancos prefieren resignar beneficios antes de pasar por la experiencia de cambiar de banco y su tarjeta de crédito asociada.

Sin embargo, la migración de clientes de un banco a otro sucede cada día con más frecuencia. Estos datos me llevaron a analizar las principales causas que generan un desgaste en la relación cliente-banco.

Actualmente existen muchas estrategias de marketing orientadas a la obtención de clientes y muy pocas orientadas a la retención. Sin embargo, cada año el banco al que pertenezco pierde 8 M USD en clientes que abandonan el servicio. ¿Y si pudiéramos anticiparnos a esta pérdida identificando un modelo predictivo asociado al comportamiento y características de nuestros clientes?

Sabemos que nuestros clientes reciben, al menos, 4 propuestas al año para cambiarse de banco por distintos beneficios. Sumado a la alta competencia, contamos con estadísticas que nos muestran que el 20% de nuestros clientes presenta al menos un reclamo durante su tiempo como cliente. Conocemos los principales motivos de queja y le damos seguimiento, sin embargo, no pudimos establecer el patrón entre los motivos de reclamo, la frecuencia y la deserción. Por lo tanto precisamos desarrollar un modelo predictivo que nos permita predecir el porcentaje de deserción.

Audiencia

Este análisis intenta contestar, con evidencia, las preguntas del párrafo anterior por lo cuál puede ser de utilidad los responsables de areas de marketing de bancos comerciales y responsables del área de créditos.



PREGUNTAS DE INTERÉS

Preguntas claves

¿Qué lleva a un cliente a cambiar de banco y tarjeta de crédito?

¿Qué indicios existen para predecir que un cliente está por abandonar el servicio bancario?

Preguntas adicionales

¿En qué porcentaje los clientes abandonan el servicio?

¿Hay relación entre la **edad** de los clientes y el abandono del servicio?

¿Los clientes de **mayor edad** están abandonando más que los más jóvenes?

¿Cómo se distribuye (%) los clientes que abandonan el servicio según el **rango de sus ingresos** informados?

¿Existe correlación entre mi variable target (abandono de clientes) y el resto de las variables del presente trabajo?

¿Hay diferencias significativas en la distribución del **límite crediticio** entre los clientes y los que abandonan?



RESUMEN METADATA

Total clientes
10.127



♂ 4.769

♀ 5.358

Datos de nuestros clientes

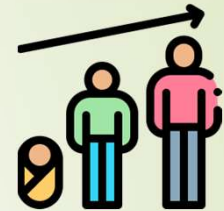


variable target
Cliente desgastado

16% Clientes desgastados

84% Clientes activos

Prom edad
47 años



Prom
antigüedad 44
meses



Los datos fueron extraídos de Kaggle. Para mayor detalle ingresar al siguiente [link](#).



ANÁLISIS **EXPLORATORIO**

¿Quiénes son nuestros clientes?

Los gráficos de la derecha, nos muestran que el 84% de los clientes son clientes activos mientras el 16% restante corresponde a clientes desgastados.

Si analizamos la misma variable por género encontramos que 17% de los clientes de género femenino son clientas desgastadas mientras que el 83% son clientes estables. Para el género masculino la proporción es de 15% para los clientes desgastados y 85% para los clientes estables. La visualización de la derecha nos permite inferir que la distribución por género es bastante homogénea.

Clientes y género



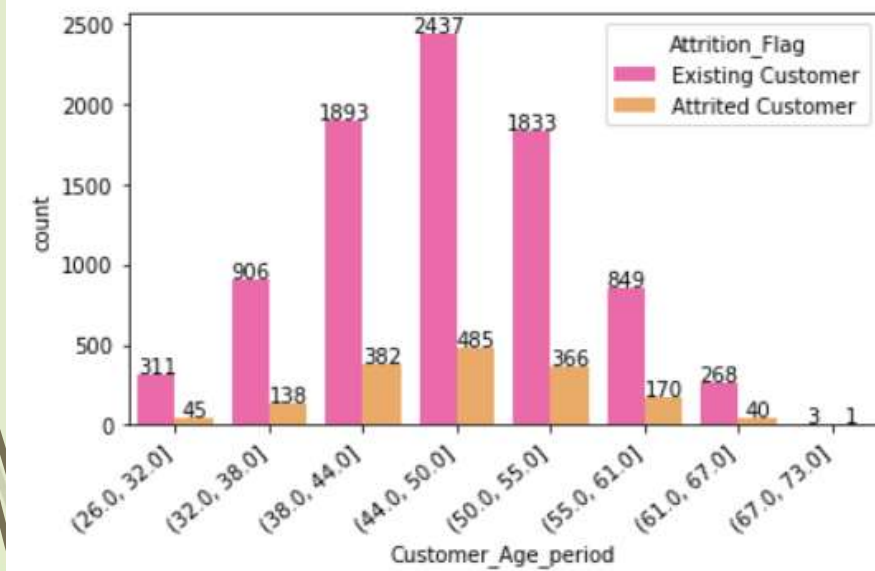
EDAD DE NUESTROS CLIENTES

¿Qué edad tienen nuestros clientes?

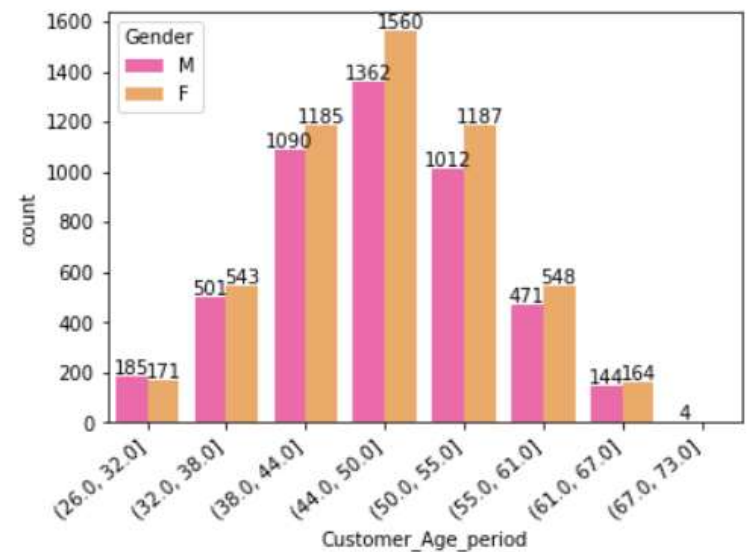
Hay más clientes existentes en las edades entre 32 y 38 años y menos clientes attrited en estas edades. También, hay más clientes attrited en las edades entre 50 y 55 años. Estas tendencias sugieren que los bancos pueden enfocarse en la retención de clientes y la atención adecuada a las necesidades de los clientes en estas edades específicas.

Si visualizamos la distribución de la edad de los clientes abiertos por género vemos que se mantiene la proporción de cada género en cada rango etáreo previamente evidenciado. Resulta relevante conservar este dato para alimentar nuestro modelo predictivo.

Distribución de abandon del servicio por rango etáreo

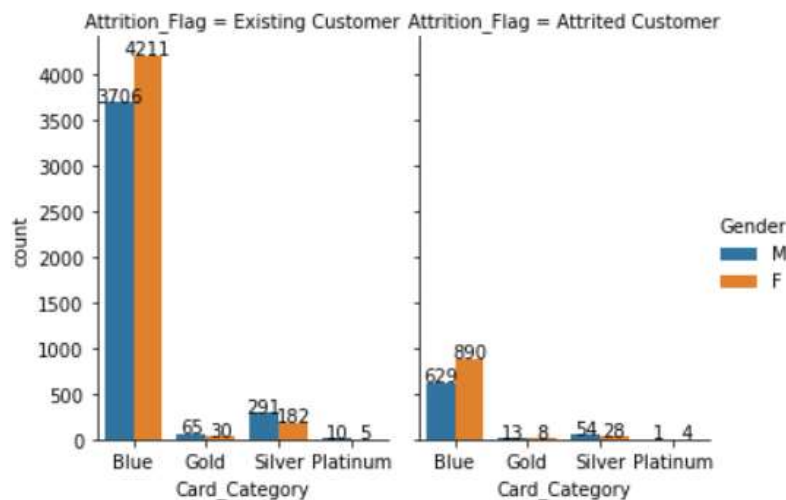


Distribución de edad y género de los clientes



TIPO DE TARJETA DE CREDITO

Distribución de clientes abiertos por tarjeta de crédito, género y variable target



¿Qué tarjeta de crédito utilizan nuestros clientes?

Los hombres (M) tienden a tener tarjetas de crédito de las categorías más altas, como Gold y Platinum.

Las mujeres (F) tienden a tener tarjetas de crédito de las categorías más bajas, como Blue.

Aproximadamente el 40% de los clientes existentes (existing customer) tienen tarjetas de crédito de la categoría Blue, lo que sugiere que la categoría Blue es popular entre los clientes existentes.

El 15% de los clientes existentes tienen tarjetas de crédito de la categoría Gold, lo que sugiere que la categoría Gold también es popular entre los clientes existentes.

Aproximadamente el 42% de los clientes que abandonaron el servicio (attrited customer) tienen tarjetas de crédito de la categoría Blue, lo que sugiere que la categoría Blue es popular entre los clientes que abandonaron el servicio.

Identificando cada % podremos alimentar a nuestro modelo predictivo y predecir, según el género y tarjeta de crédito, la probabilidad de que un cliente abandone el servicio.

55%

52%



MESES DE ANTIGÜEDAD

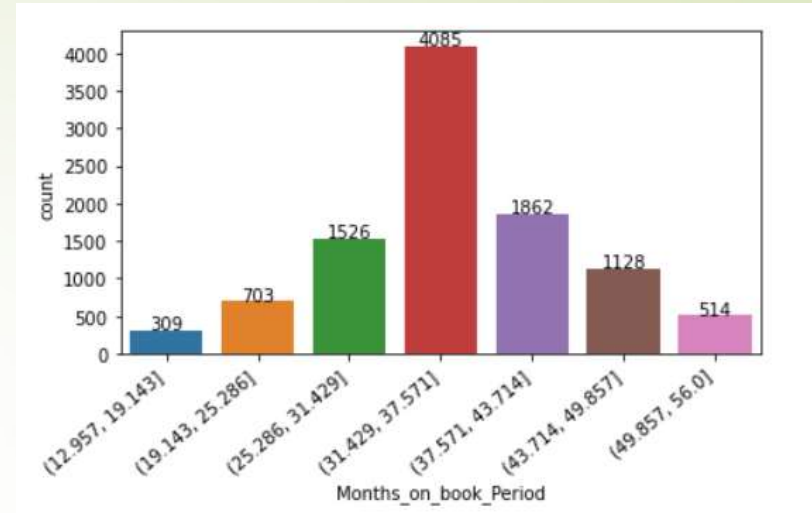
¿Cómo es la distribución de la antigüedad de nuestros clientes y el abandon del servicio?

Existe una alta concentración de clientes en el rango de 31,9 a 37,5 meses. Además, se producen caídas importantes en el número de clientes de mayor duración.

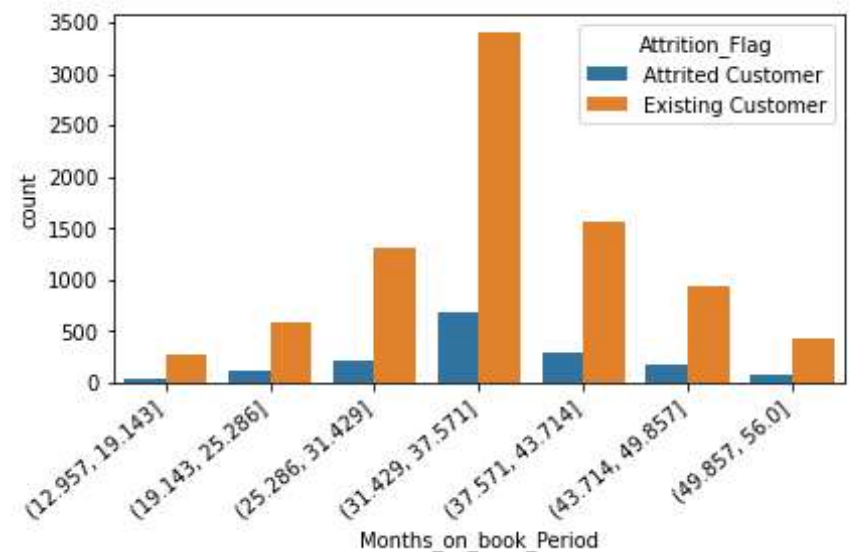
Si analizamos nuestra variable target, vemos que los clientes con un historial más corto tienden a tener niveles de satisfacción más altos y abandonan con menos frecuencia en comparación con aquellos con un historial hasta 37 meses. Esto podría sugerir que dirigirse a estos clientes a corto plazo puede ser más beneficioso para la empresa.

Si bien el análisis inicial sugiere un patrón de mayor número de clientes a corto plazo, complementaremos esta investigación con otros factores para obtener una comprensión más completa del comportamiento y las preferencias de los clientes.

Antigüedad de nuestros Clientes (en meses)



Antigüedad de nuestros Clientes abiertos por nuestra variable target (en meses)



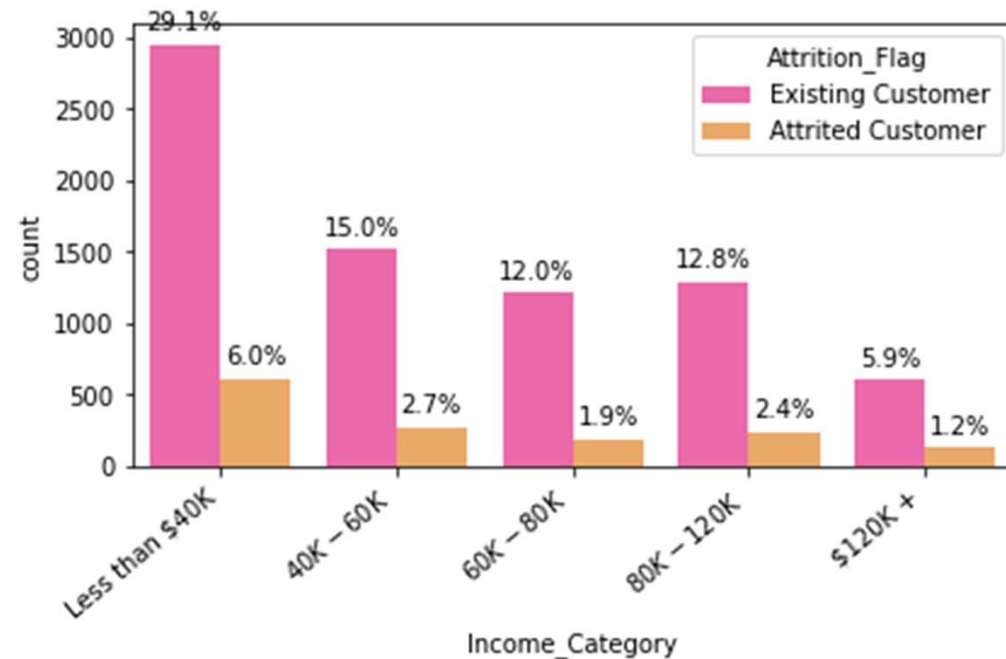
RANGO DE INGRESOS

¿Hay relación entre el rango de ingresos y nuestra variable target?

Existe una fuerte correlación entre los ingresos y la deserción de clientes. El gráfico muestra la distribución de los niveles de ingresos entre los clientes existentes y los clientes que se han abandonado el servicio. Esta fuerte dependencia permite asignar una probabilidad de abandono a cada nuevo cliente.

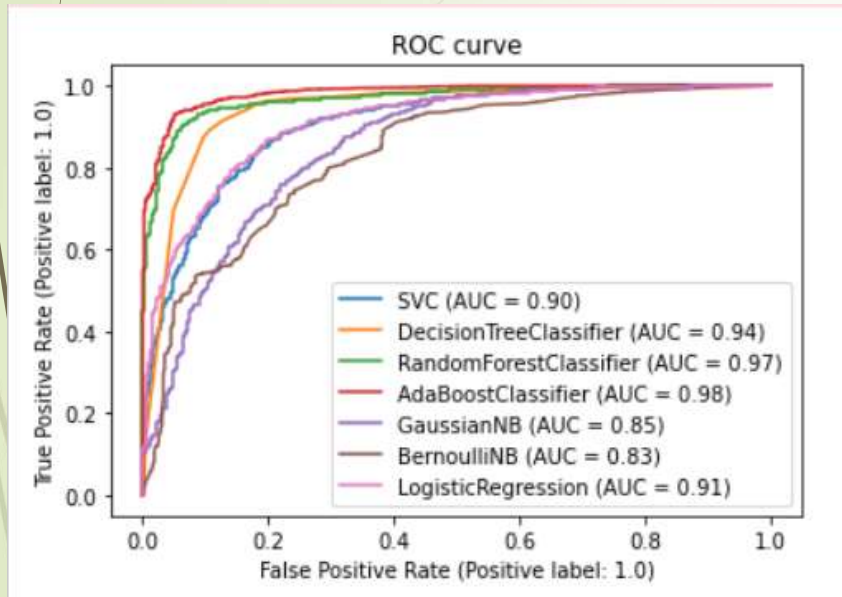
Los clientes existentes con ingresos inferiores a 40K tienen más probabilidades de abandonar el negocio. Esta categoría de ingresos tiene el porcentaje de deserción más alto.

Podemos orientar nuestros esfuerzos de marketing específicos de acuerdo a cada nivel ingresos. La satisfacción del clientes en cada rango está sujeta a distintas percepciones. Con esta información podemos centrarnos en la satisfacción de cliente con mayor tasa de rentabilidad.



MODELOS PREDICIVOS

¿Cuál es el modelo predictivo con mayor precisión?



De acuerdo a nuestro análisis de la matriz de confusión podemos observar que el mejor clasificador es AdaBoost con precisión = 0.88, recall = 0.82. Si realizamos la curva de ROC obtenemos la métrica AUC - 0.98 que resulta consistente con nuestra matriz de confusión.

Las métricas de AdaBoostClassifier son bastante buenas. La matriz de confusión muestra que el modelo hizo 396 predicciones correctas para "Existing Customer" y 2497 predicciones correctas para "Attrited Customer".

AdaBoostClassifier:

	precision	recall	f1-score	support
Existing Customer	0.88	0.81	0.84	488
Attrited Customer	0.96	0.98	0.97	2551
accuracy			0.95	3039
macro avg	0.92	0.90	0.91	3039
weighted avg	0.95	0.95	0.95	3039

```
[[ 396  92]
 [  54 2497]]
```





INSIGHTS & **RECOMENDACIONES**

INSIGHTS & RECOMENDACIONES

Insights

- ❑ La variable Edad tiene una probabilidad de 0,979 de ser dependiente de nuestra variable target. Esto quiere decir que cada vez que un nuevo cliente se suma podremos predecir la probabilidad de que abandone el servicio de acuerdo a su edad.
- ❑ Las siguientes variables tienen una dependencia alta a nuestra variable target:
 - 12 meses de inactividad
 - Total saldo disponible
 - Total importe de transacción
 - Total de transacciones
- Nuestro modelo predictivo podrá evaluar para cada nuevo cliente la probabilidad de que abandone el servicio según las variables mencionadas. Utilizaremos el clasificador es AdaBoost con *precisión = 0.88, recall = 0.82

