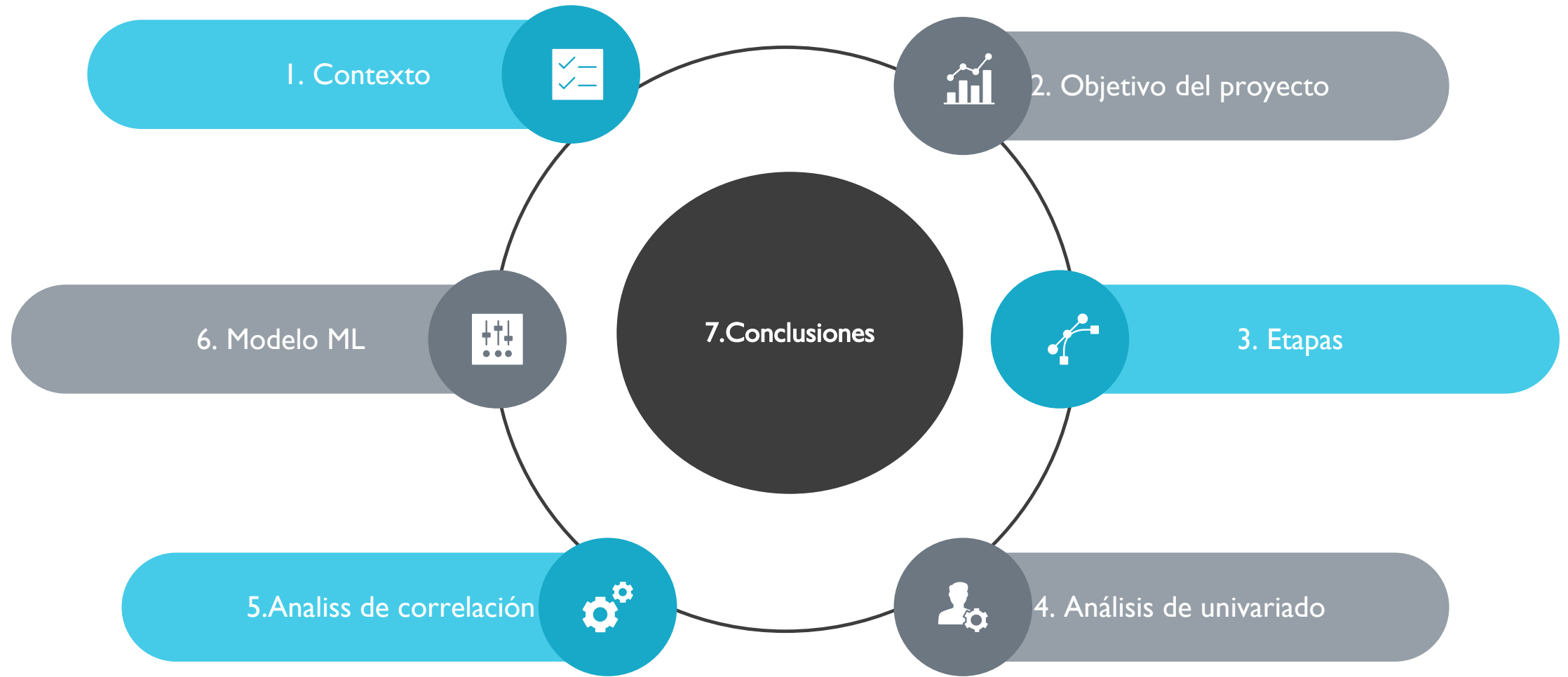


A decorative graphic at the top of the slide consisting of several overlapping geometric shapes. There are three horizontal bars: a dark blue one on the left, a light blue one in the middle, and a light gray one on the right. Overlaid on these are two large, thin-lined triangles pointing downwards, one in cyan and one in light orange.

ANALISIS DE CARTERA DE CLIENTES

Importancia del análisis de clientes y su fidelización.

Índice



CONTEXTO

Una empresa de servicios financieros cuyo rol es brindarle un sistema crediticio a sus clientes, nos encomendó la tarea analizar los datos relacionados con las tarjetas de crédito y, en particular, de estimar los saldos disponibles en estas tarjetas en el futuro. Este análisis es crucial para ayudar a la empresa a tomar decisiones estratégicas sobre sus operaciones y servicios.

A su vez nos solicito que realicemos la probabilidad de que los clientes abandonen la empresa es un proceso complejo que involucra el uso de técnicas avanzadas de análisis de datos y aprendizaje automático.

Objetivos del proyecto



ANÁLISIS DE MERCADO

Identificar los factores que más influyen en la tasa de abandono de clientes y desarrollar un plan de acción para retener a los clientes



ANÁLISIS TÉCNICO

Se utilizan técnicas de análisis de aprendizaje automático para desarrollar un modelo predictivo preciso.



ANÁLISIS FINANCIERO

Analizar los datos relacionados con las tarjetas de crédito, particular los saldos disponibles en estas tarjetas



CUSTOMER EXPERIENCE

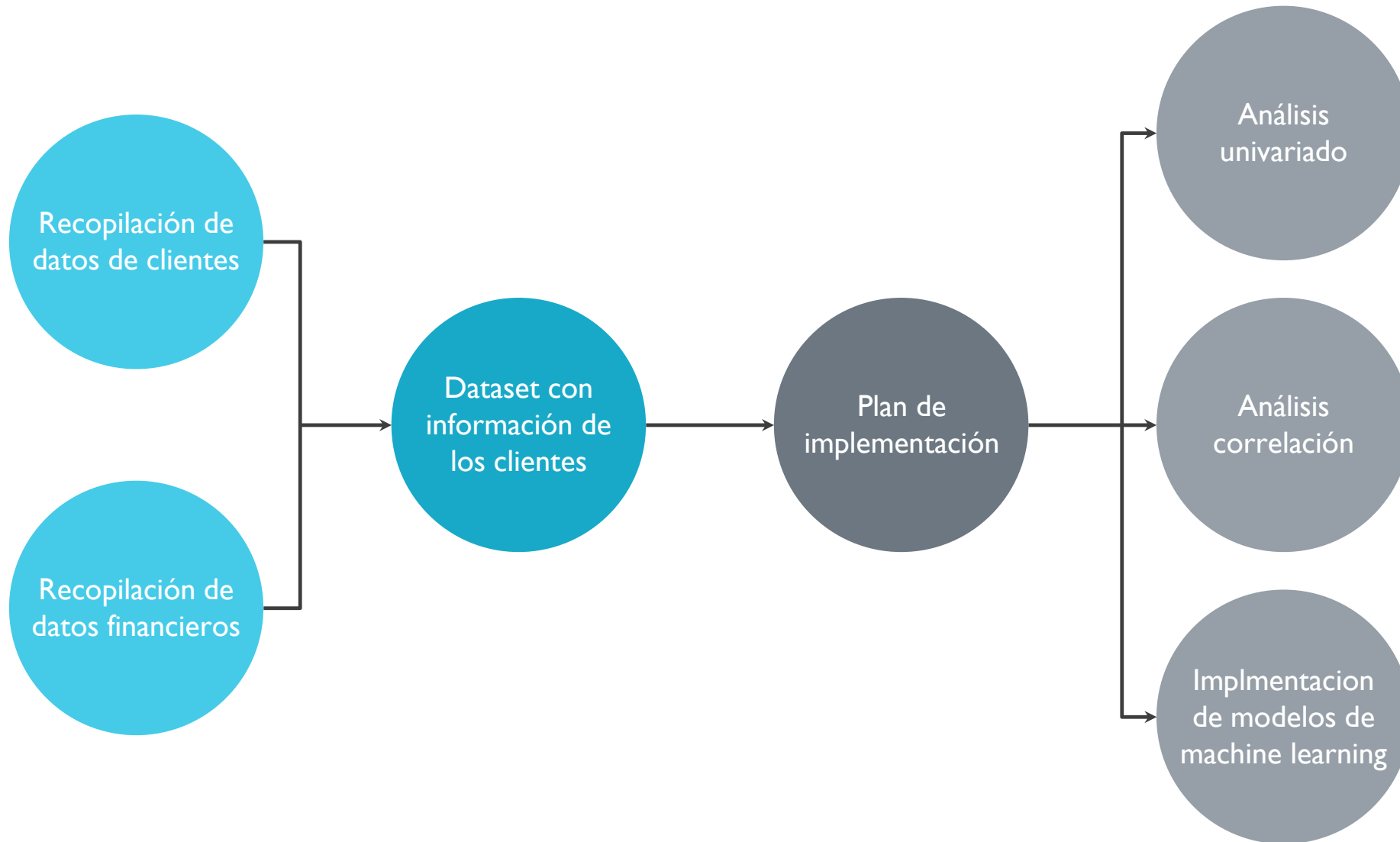
La empresa puede anticiparse a las necesidades de sus clientes y ofrecerles servicios personalizados



ANÁLISIS DE DATOS

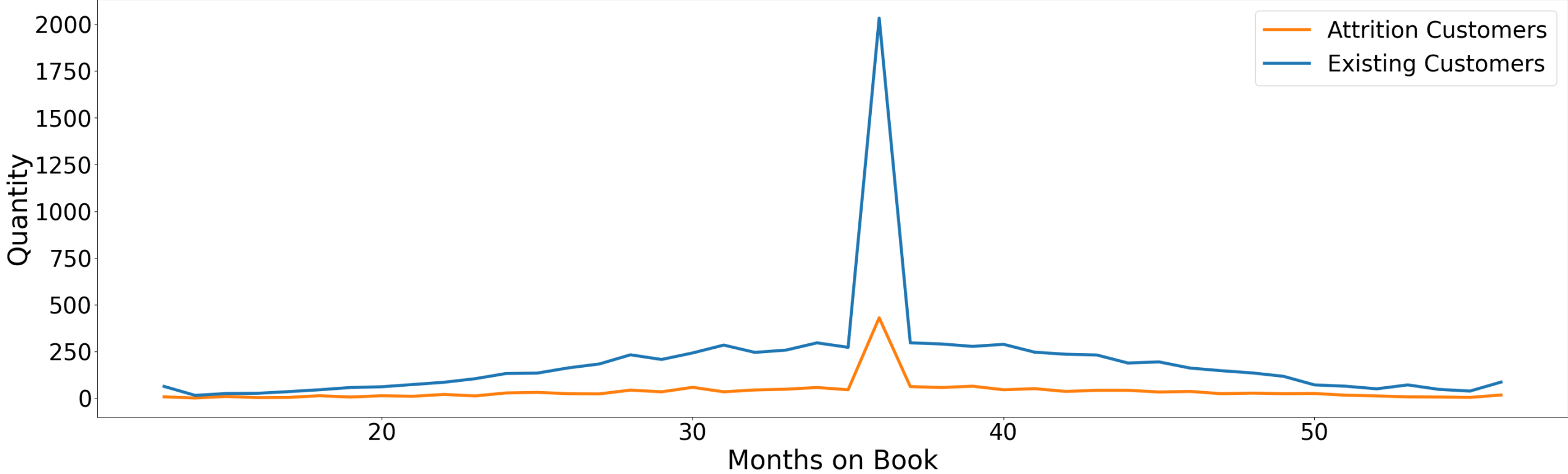
Recopilamos datos históricos de las transacciones de los clientes, como compras realizadas, pagos realizados y límites de crédito

Etapas del proyecto



Estabilidad de clientes

Attrition by Months on Book



Cientes Activos

8500

Se encuentra un pico de clientes que poseen 36 meses de antigüedad en la empresa alcanzando el aproximadamente el 30%

Cientes Dados de baja

1627

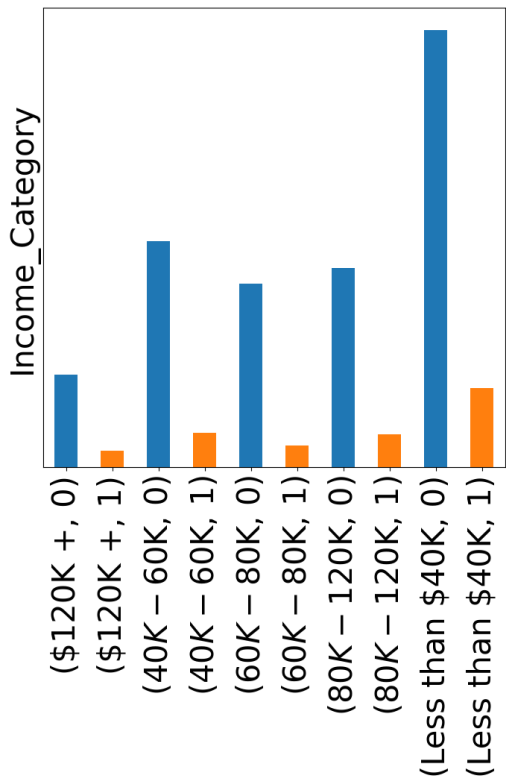
La mayor recisión de clientes se da en los que poseen 36 meses de antigüedad

Resumen

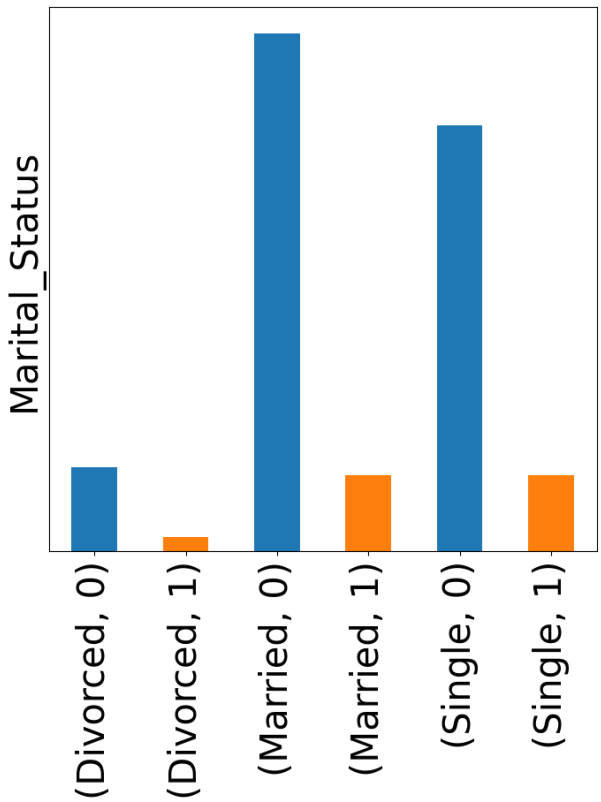
Podemos interpretar que hace un año y medio se hizo una campaña para captar nuevos clientes que fue muy efectiva y se observa la baja capitación de clientes nuevos.

Variables a analizar - Clientes

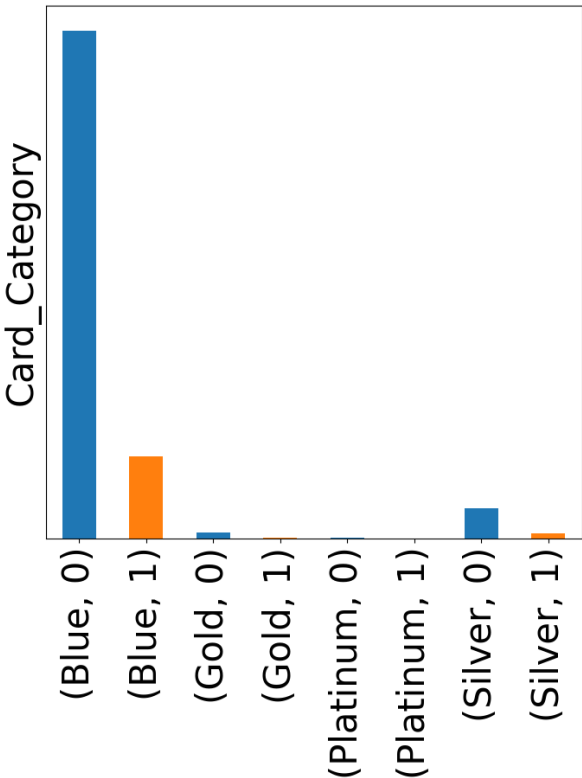
Income category: el 33% de los clientes Activos tienen ingresos menores a \$40,000



Marital status: el 45% de los clientes son casados, cabe destacar que si bien hay pocos clientes divorciados, tiene una alta tasa de cuentas dadas de bajas en relación a las abiertas

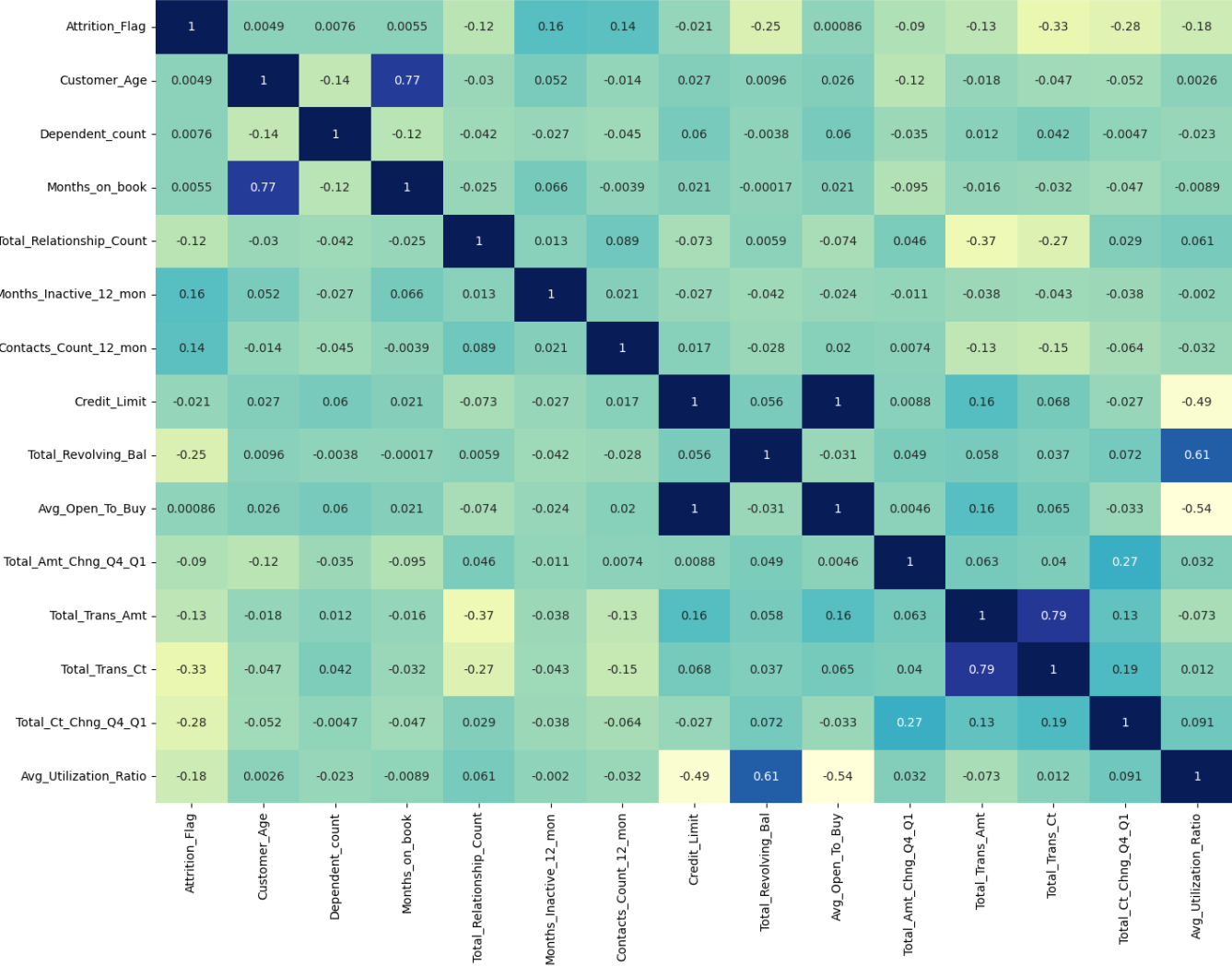


Card_Category: el 80 % de los clientes posee la tarjeta "Blue"



* destacamos que las variables Education_Level

Análisis de correlación



Podemos observar una correlación positiva entre los meses que esta el cliente con la edad del mismo



Se observa una correlación muy fuerte entre el limite de crédito y el promedio de crédito de los últimos 12 meses



Se destaca una correlación negativa entre la posibilidad de que el cliente permanezca o no en la empresa con la cantidad de transacciones netas (count)

Selección de modelo de machine learning

Model Results On Test Data

Model	F1 Score On Test Data
Random Forest	0.87
AdaBoost	0.85
SVM	0.74
XGBoost	0.9

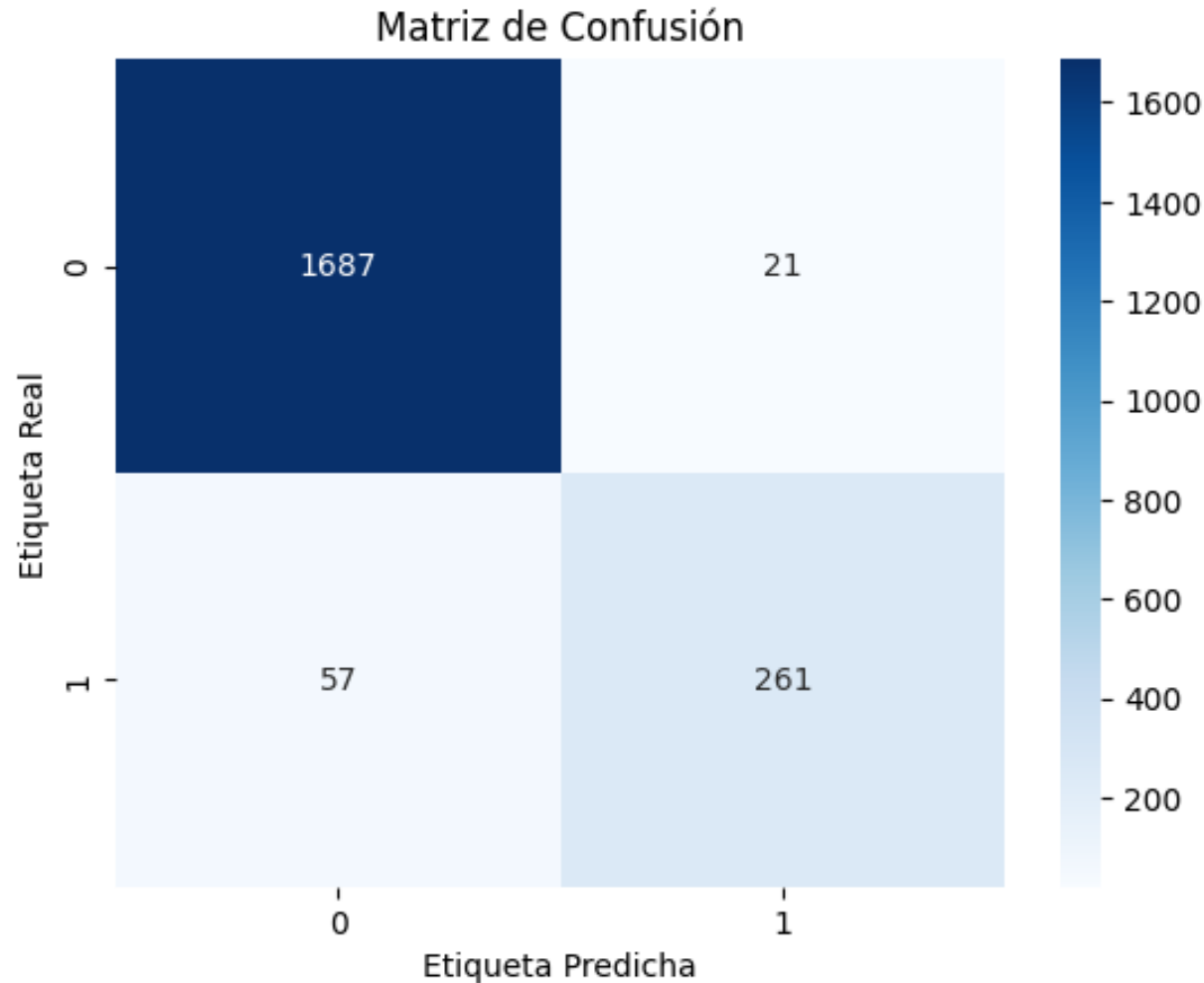
Se emprendió una prueba con el propósito de discernir cuál algoritmo podría sobresalir en términos de eficiencia al trabajar con el conjunto de datos en cuestión.

Los resultados recopilados, tal como se detallan en la tabla superior, arrojan que el algoritmo con un desempeño más sobresaliente es el XGBoost, seguido de cerca por el Random Forest.

Es esencial subrayar que este análisis se ejecutó previo al proceso de depuración y limpieza del conjunto de datos y antes de que se llevara a cabo la fase de ajuste de hiperparámetros. Esta metodología garantiza una evaluación del rendimiento de los algoritmos en un estado de igualdad, permitiendo así identificar sus potencialidades.

En resumen, los datos actuales respaldan la noción de que tanto el XGBoost como el Random Forest presentan habilidades prometedoras en el manejo de este conjunto de datos, y sería recomendable proceder con una optimización de hiperparámetros para alcanzar su máximo potencial.

Random Forest



Accuracy

Muestras clasificadas correctamente sobre el total de muestras: 0.9630

Precision

Muestras positivas clasificadas correctamente: 0.9290

Recall

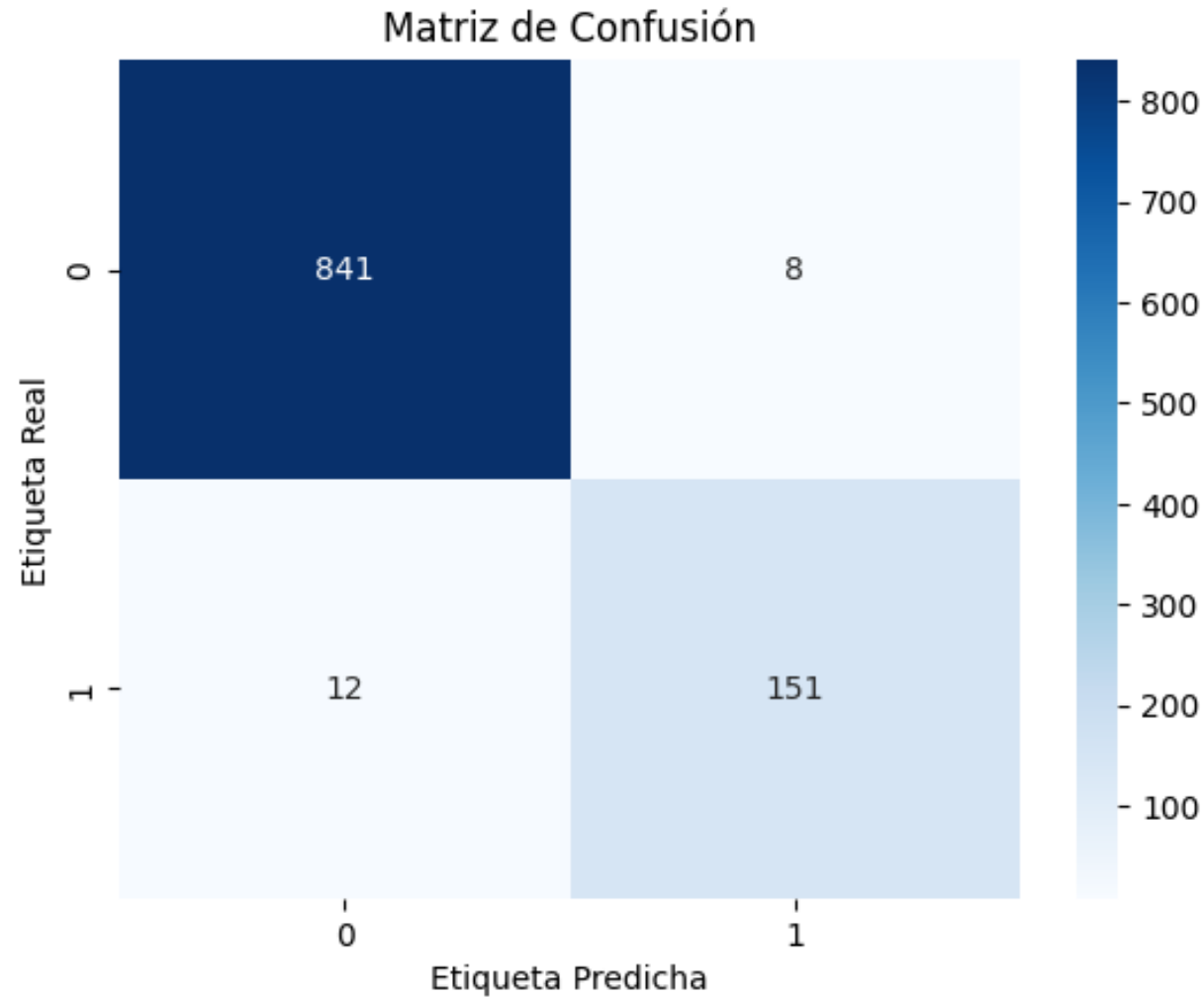
Tasa de verdaderos positivos: 0.8322

F1

Desequilibrio entre las clases: 0.8777

Como conclusión para este modelo podemos decir que es aceptable teniendo en cuenta que posee un f1 de 87.78% aunque lo mas importante es ver la tasa de verdaderos positivos ya que lo que se busca es la exactitud en clientes que puedan llegar a dejar la empresa.

XGBoost



Accuracy

Muestras clasificadas correctamente sobre el total de muestras: 0.9716

Precision

Muestras positivas clasificadas correctamente: 0.9304

Recall

Tasa de verdaderos positivos: 0.8898

F1

Desequilibrio entre las clases: 0.9094

Puntualmente hay un diferencia del 5% en recall (verdaderos positivos) el cual es el punto fuerte de nuestro proyecto. El modelo muestra un buen nivel de confianza y precisión. Se puede observar que el modelo aprendio correctamente y no realiza overfitting ya que al aplicar la validación cruzada los parametros tienden a ser similares.

Conclusiones - FODA

POSITIVO

NEGATIVO

FUERZA

- › Acceso a datos históricos: La empresa tiene acceso a un conjunto de datos históricos de tarjetas de crédito, lo cual proporciona una base sólida para el análisis y la predicción de los saldos disponibles en el futuro.
- › Capacidad de análisis avanzado: El equipo cuenta con habilidades en análisis de datos y aprendizaje automático, lo cual les permite aplicar técnicas avanzadas para obtener información valiosa y realizar predicciones precisas.

PUNTOS DÉBILES

Calidad de los datos: Es posible que los datos históricos de las tarjetas de crédito no estén completamente limpios o sean incompletos.

Falta de experiencia en aprendizaje automático: Si el equipo no cuenta con una sólida experiencia en técnicas de aprendizaje automático, podría requerir capacitación adicional o colaboración con expertos en el campo para aplicar correctamente estas técnicas.

OPORTUNIDAD

Personalización de servicios: La empresa puede identificar patrones y preferencias de los clientes, lo que les permite ofrecer servicios personalizados y adaptados a las necesidades individuales de cada cliente.

Detección temprana de riesgos: Mediante el análisis de datos históricos, la empresa puede identificar patrones de comportamiento de los clientes que indican un mayor riesgo de abandono. Esto les permite tomar medidas preventivas para retener a los clientes en riesgo.

AMENAZA

Competencia en el mercado: Existe una competencia significativa en el mercado de servicios financieros. Otras empresas podrían ofrecer servicios similares o mejores, lo que podría llevar a la pérdida de clientes si no se brinda un valor diferenciado o se satisfacen adecuadamente sus necesidades.

INTERNO

EXTERNO



GRACIAS

Carlos Davoli