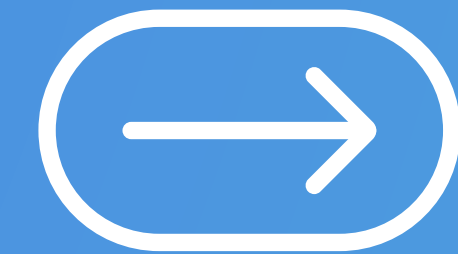


Churn en un Call Center



MLOps con CRISP-DM

CONTEXTO DEL NEGOCIO

01



Cobranza del Banco Estrellitas

Pierden clientes en cobranza por abandono de llamada



02



Aumento de exposición

Dinero no utilizable y por ende pérdidas del banco.



03



Estrategias de retención

No se mantienen a los clientes dentro de las llamadas.

Objetivos

Disminución

- Abandonamiento de clientes
- Exposición de la mora

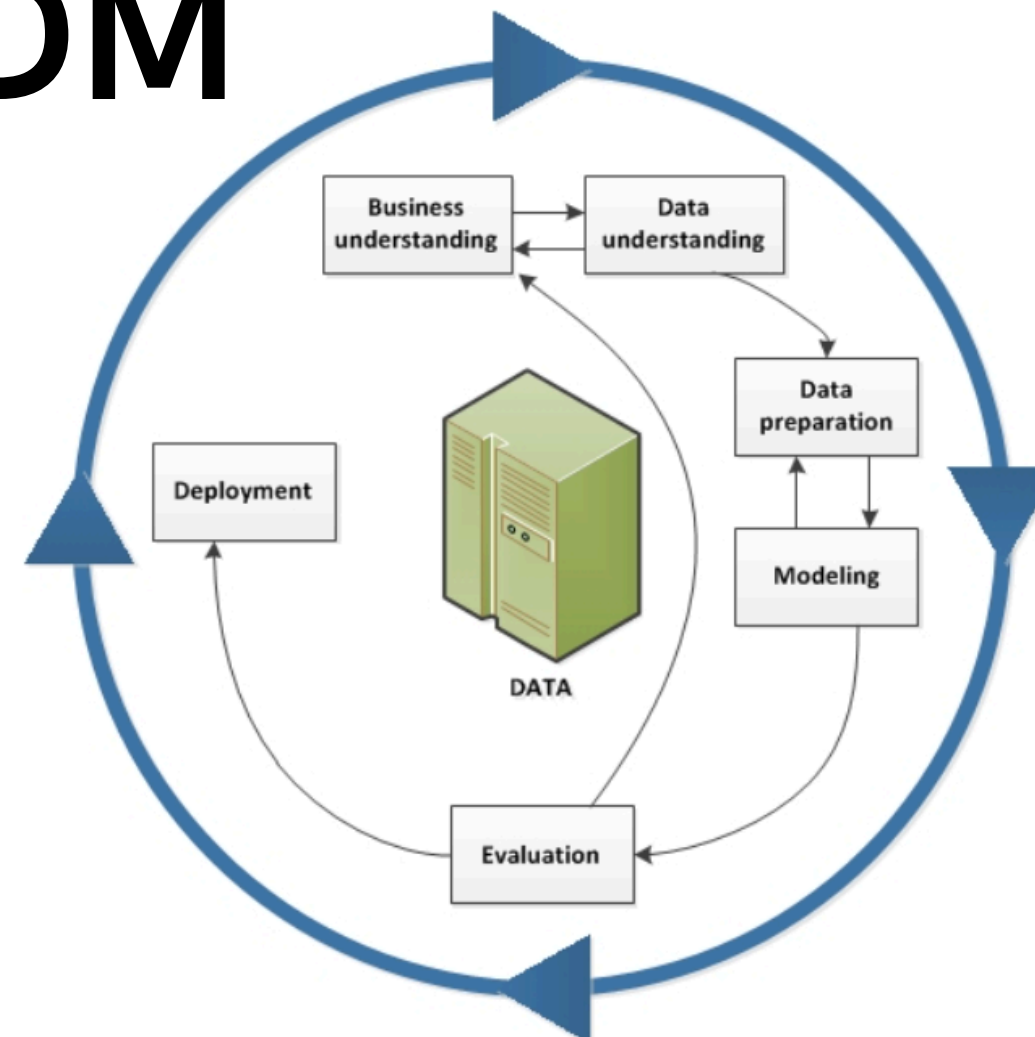
Aumento

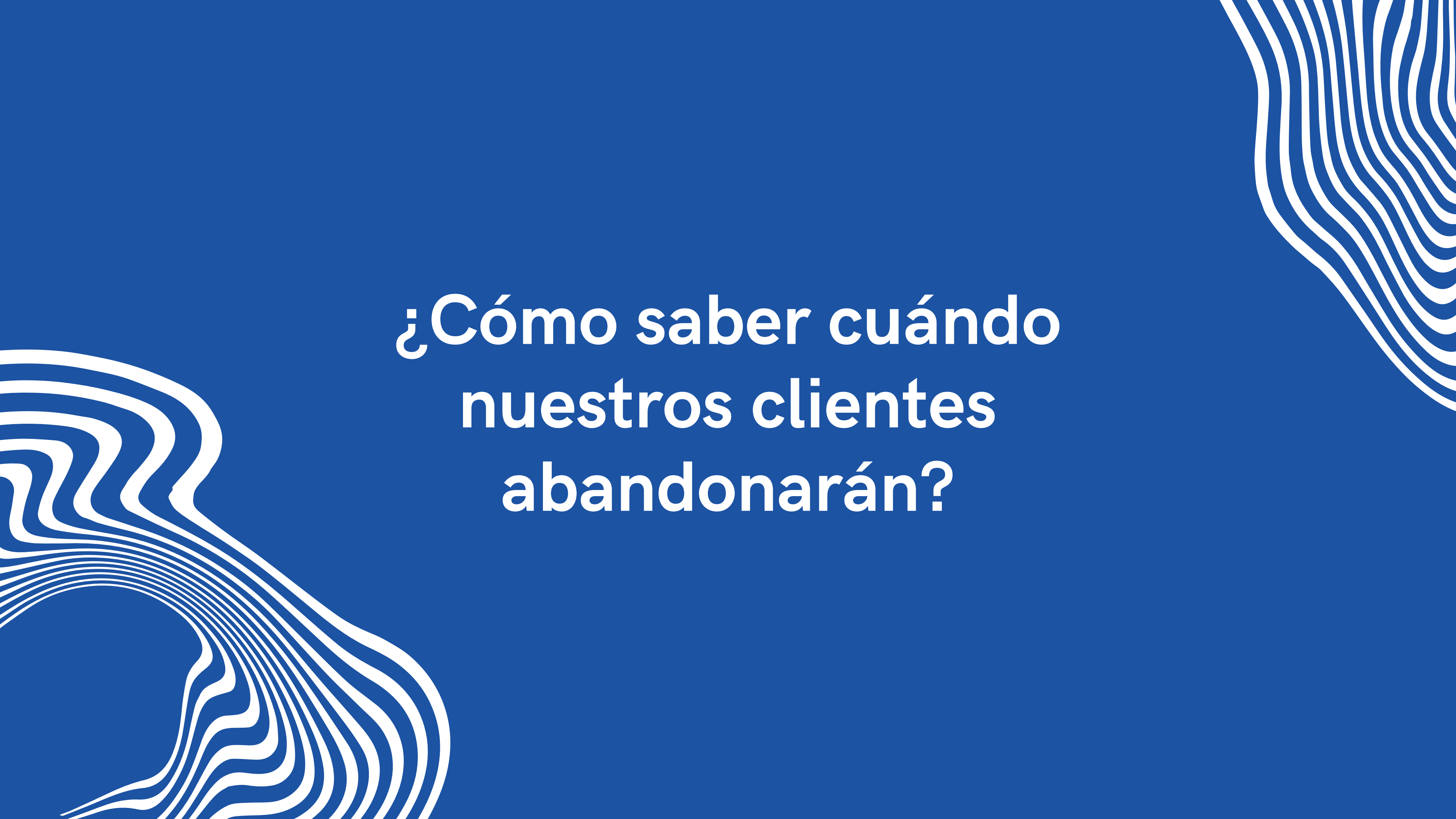
- Cartera de clientes
- Desempeño al momento de retener clientes

Recuperación

- Del dinero en mora
- Clientes que se daban por perdidos

CRISP-DM



The background is a solid blue color. It features decorative white wavy lines that flow from the left and right edges towards the center, framing the text. The lines are of varying thickness and create a sense of movement.

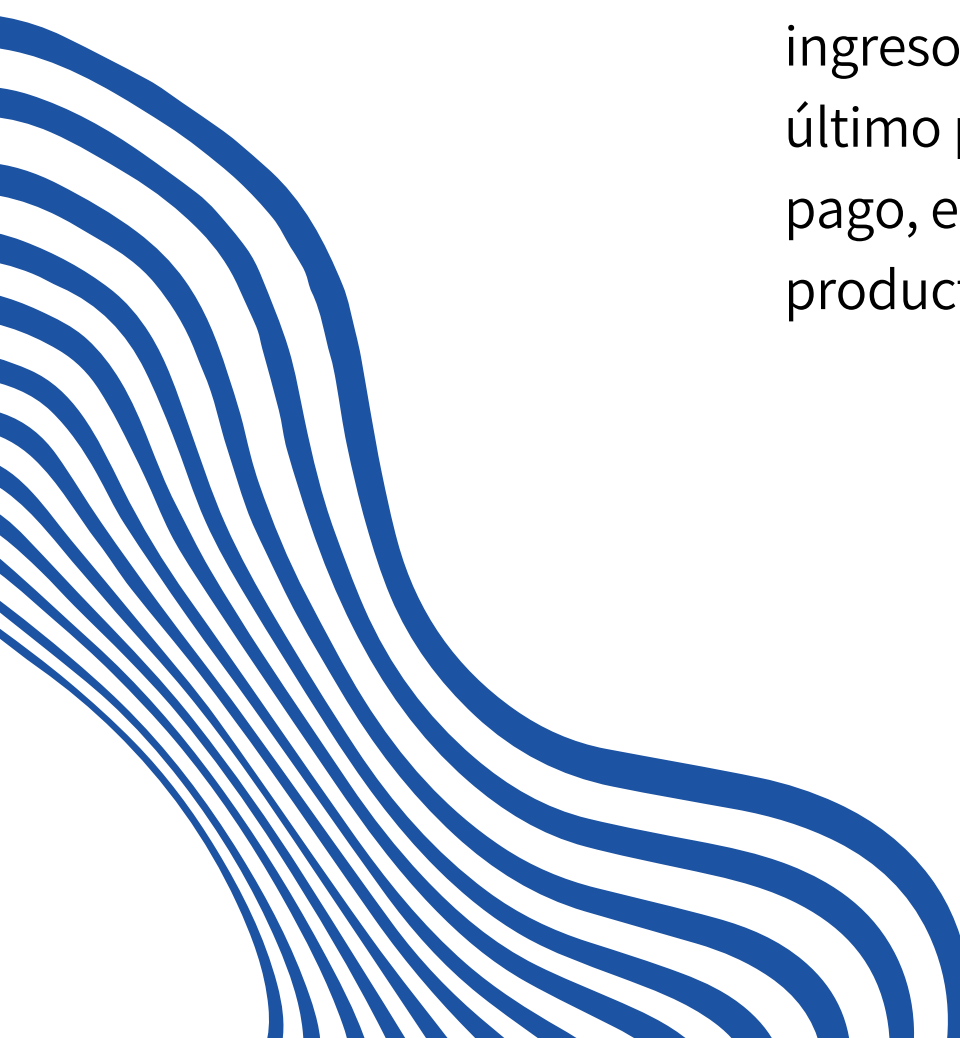
**¿Cómo saber cuándo
nuestros clientes
abandonarán?**



Dataset sintético

Cliente

Datos del cliente como id, edad, ingreso mensual, región, días del último pago, cantidad del último pago, educación, tipo de producto en riesgo.



Llamada

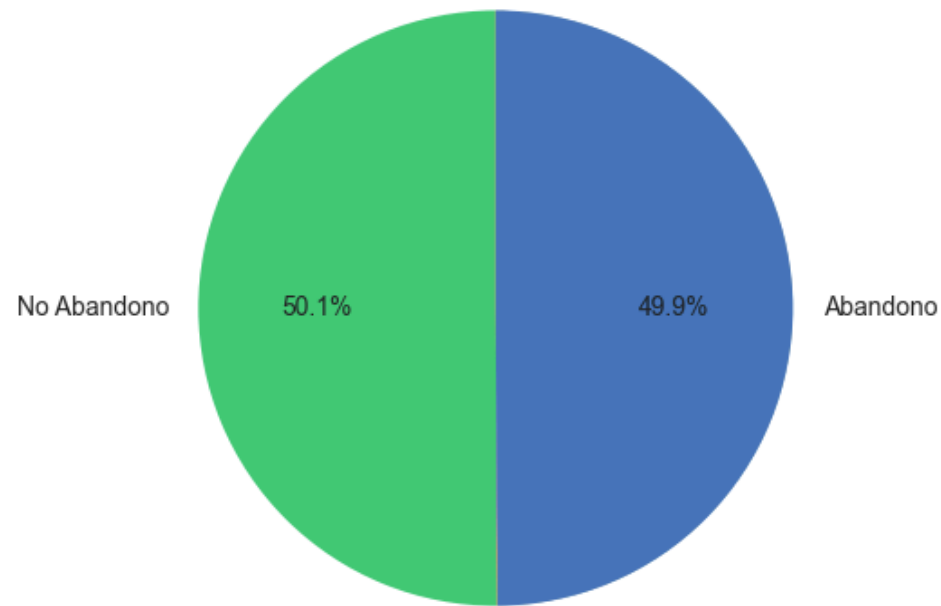
Datos como: # intentos de contacto, tiempo de espera, volumen de saturación, canal de contacto, hora de la llamada.

Riesgo

Datos como: saldo en mora, días en mora, segmento de riesgo, score crediticio, si abandono o no.

Insights principales

Distribución de Abandono (Porcentaje)



Correlación de Abandono

- + Score crediticio → - Abandono
- + Días mora → + Abandono
- + Recarga del call center → + Abandono
- + Saldo en mora → + Abandono

Tendencias de Abandono

- Crédito personal → 69.11%
- Llamada → 64.52%
- Riesgo alto → 69.59%

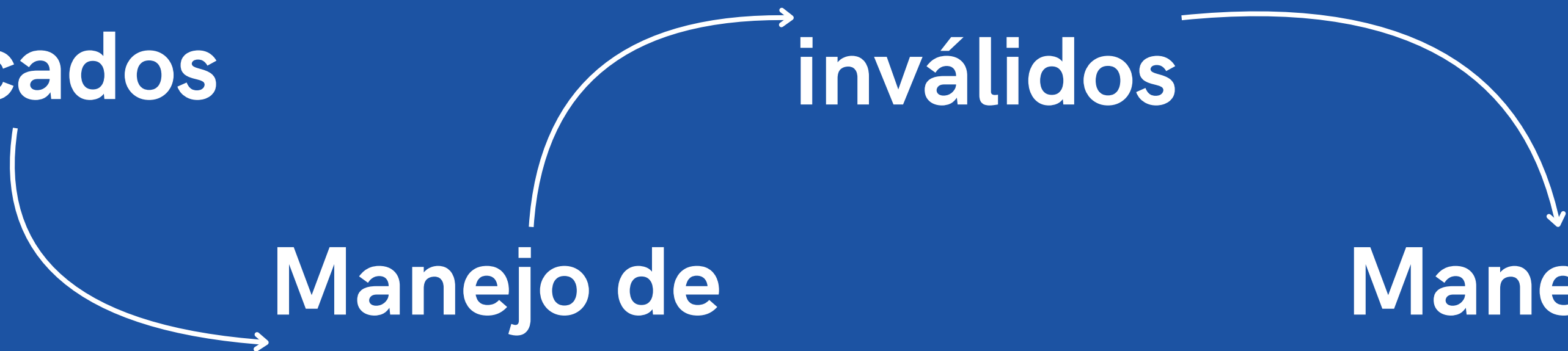
Transformación de datos

Datos
Duplicados

Valores
inválidos


Manejo de
fechas

Manejo
de
atípicos





Ingeniería de características

- Inferir la fecha de nacimiento a partir de la edad.
 - Inferir los años de antigüedad del cliente.
 - Llamadas en días de oficina
 - Día de la semana en la que ocurre la llamada
 - Mes cuándo ocurre la llamada
 - Año cuándo ocurre la llamada
- 

Validación de datos

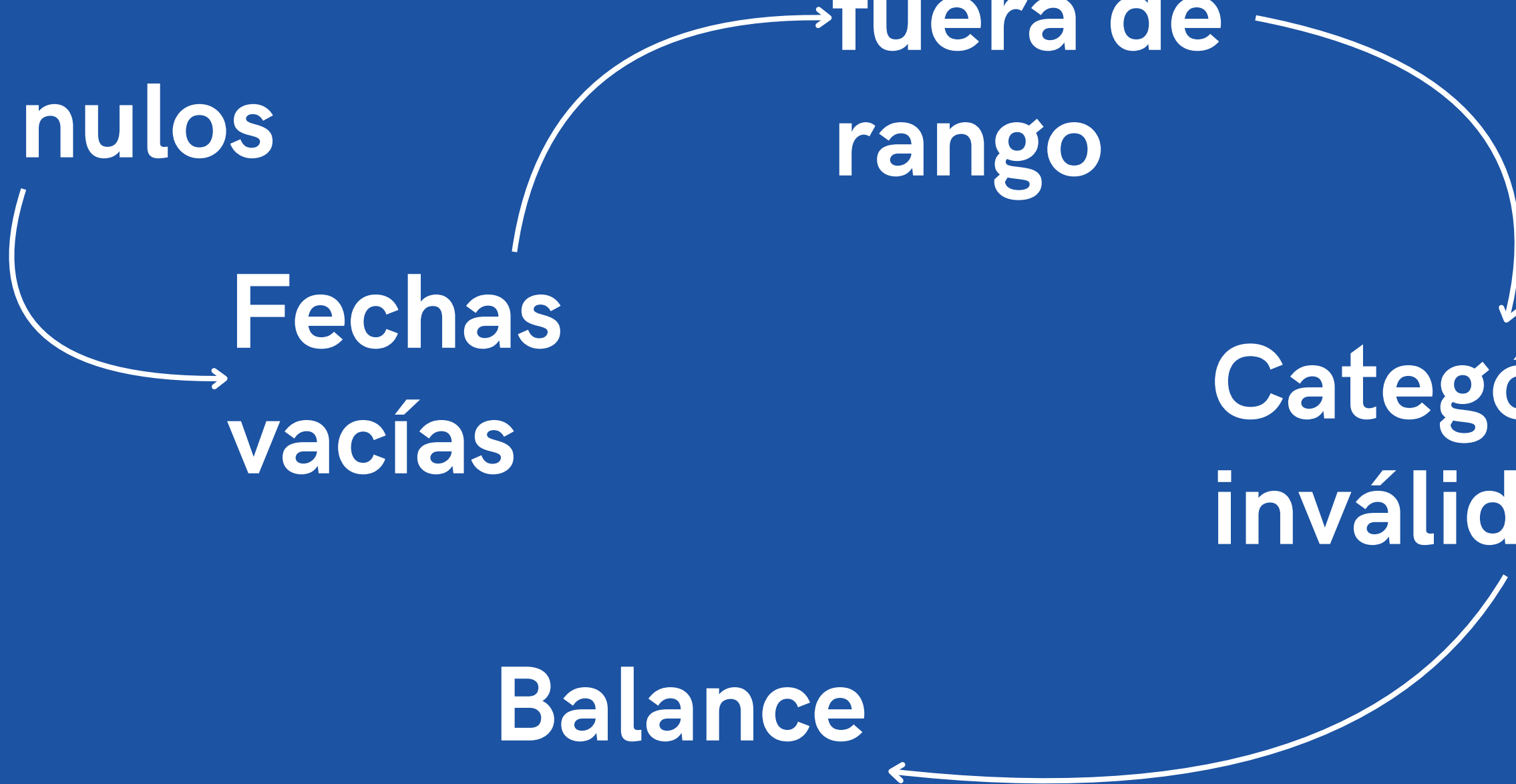
Datos nulos

Fechas
vacías

Valores
fuera de
rango




















Categorógicos
inválidos

Balance
objetivo



MLFlow

Experimentación

<input type="checkbox"/>	Run Name	Created 	Duration	accuracy	f1_score	log_loss	precision	recall	roc_auc	learning_rate	max_depth	n_estimators	n_features	n_samples_train	n_samples_val
<input type="checkbox"/>	 LightGBM	 1 minute ago	7.6s	0.85844748...	0.85699222...	0.43871333...	0.85868991...	0.85530123...	0.88683463...	0.2	3	50	2239	14863	7665
<input type="checkbox"/>	 XGBoost	 1 minute ago	8.8s	0.86027397...	0.85898617...	0.43672230...	0.85977859...	0.85819521...	0.88814646...	0.1	3	100	2239	14863	7665
<input type="checkbox"/>	 RandomForest	 3 minutes ago	8.5s	0.84918460...	0.84821428...	0.41824514...	0.84665792...	0.84977637...	0.88254446...	-	20	200	2239	14863	7665
<input type="checkbox"/>	 LogisticRegression	 19 minutes ago	17.4s	0.82191780...	0.82173174...	0.46031930...	0.81587136...	0.82767692...	0.87573057...	-	-	-	2239	14863	7665
<input type="checkbox"/>	 LogisticRegression	 34 minutes ago	2.2s	-	-	-	-	-	-	-	-	-	-	-	-
<input type="checkbox"/>	 LightGBM	 1 hour ago	5.3s	0.85844748...	0.85699222...	0.43871333...	0.85868991...	0.85530123...	0.88683463...	0.2	3	50	2239	14863	7665
<input type="checkbox"/>	 XGBoost	 1 hour ago	12.1s	0.86027397...	0.85898617...	0.43672230...	0.85977859...	0.85819521...	0.88814646...	0.1	3	100	2239	14863	7665
<input type="checkbox"/>	 RandomForest	 1 hour ago	5.7s	0.84918460...	0.84821428...	0.41824514...	0.84665792...	0.84977637...	0.88254459...	-	20	200	2239	14863	7665
<input type="checkbox"/>	 LogisticRegression	 1 hour ago	13.8s	0.82191780...	0.82173174...	0.46031930...	0.81587136...	0.82767692...	0.87573057...	-	-	-	2239	14863	7665

Hiperparámetros

XGBoost:

- Estimadores: 50, 100, 200, 300
- Max Depth: 3, 5, 7, 9
- Learning Rate: 0.01, 0.1, 0.2, 0.3

Random Forest:

- Estimadores: 50, 100, 200, 300
- Max Depth: 5, 10, 15, 20
- Min Muestras a Separar: 2, 5, 10

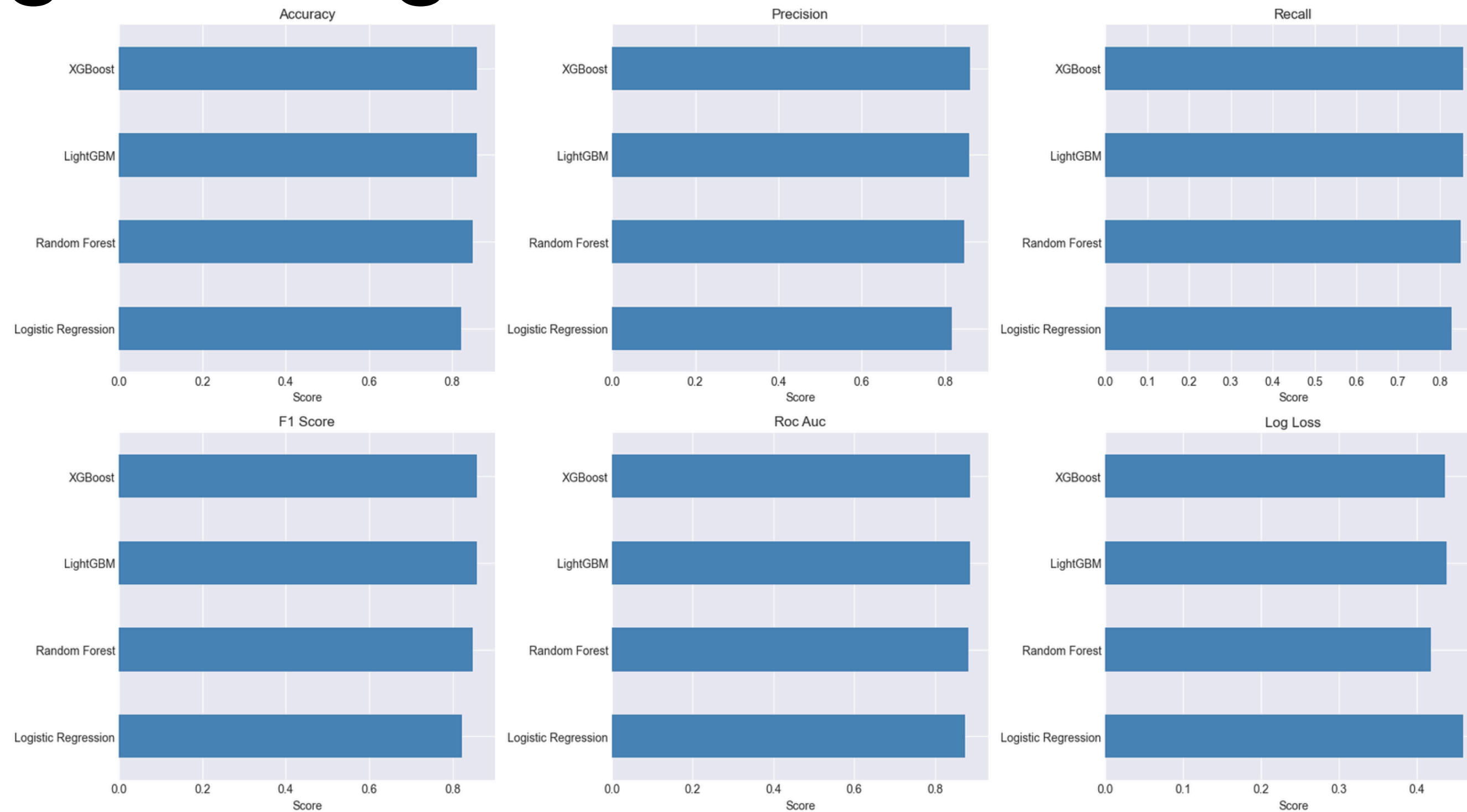
LightGBM:

- Estimadores: 50, 100, 200, 300
- Max Depth: 3, 5, 7, 9, -1
- Learning Rate: 0.01, 0.1, 0.2, 0.3

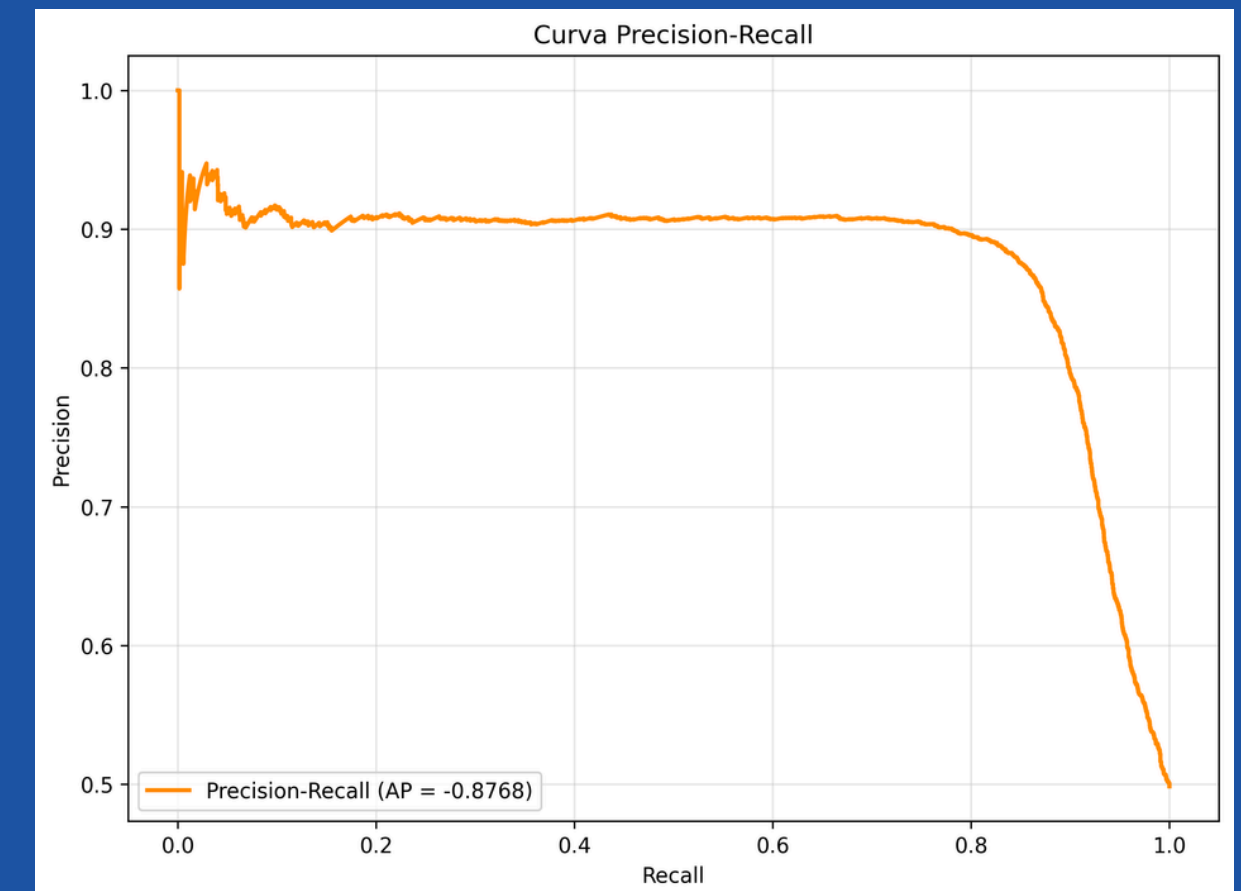
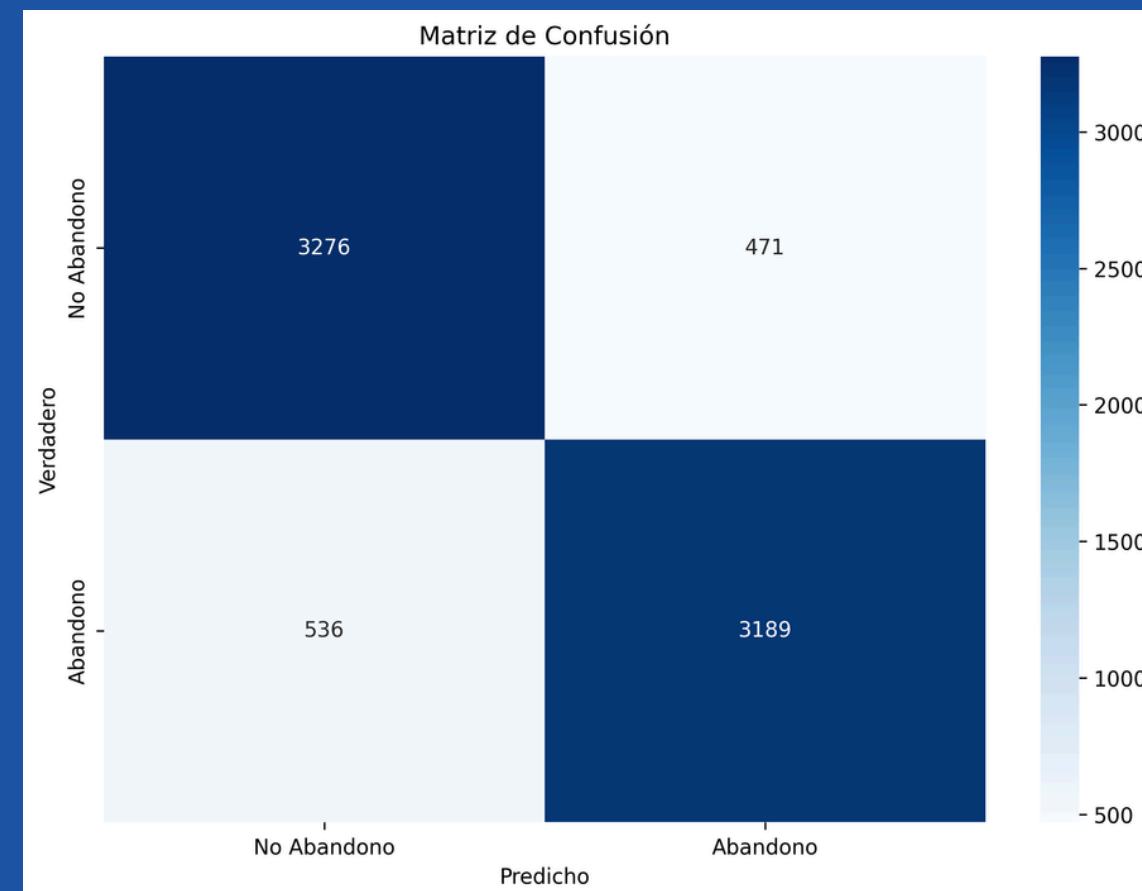
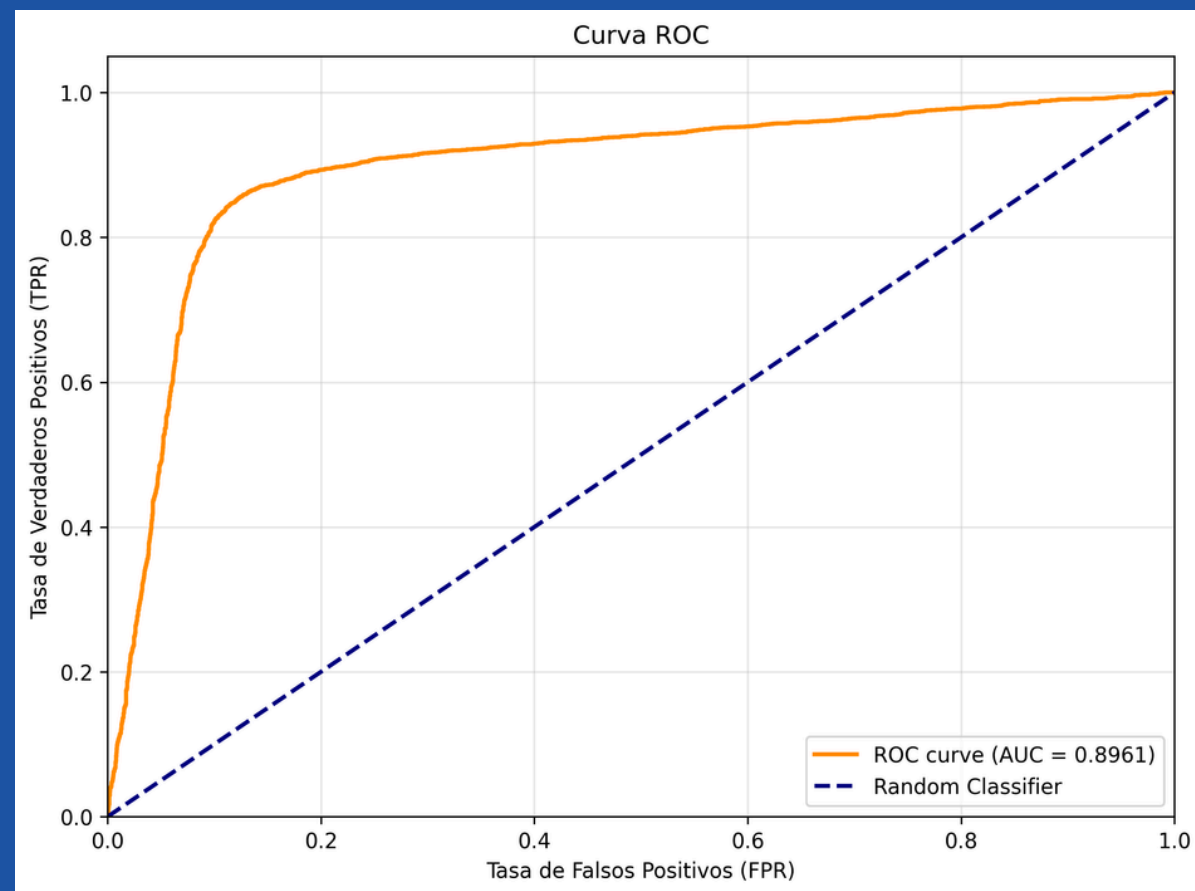
Regresión Logística:

- C:0.001, 0.01, 0.1, 1, 10, 1000
- Penalty: 11, 12
- Max Iter: 1000, 2000

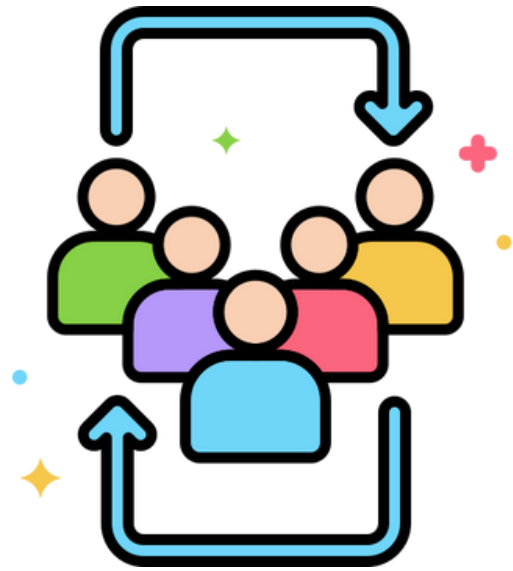
XGBoost v LightGBM v Random Forest v Logistic Regression



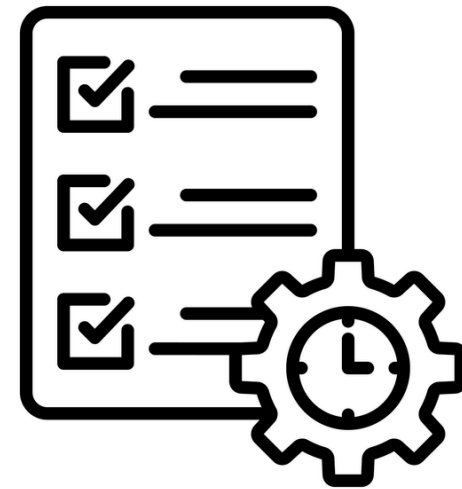
Modelo seleccionado XGBoost



Mejoras



Decisiones dinámicas



Priorización



Menos Abandono



Recuperación

Conclusiones

- La estandarización permite facilidad en despliegue
- El proyecto deja una base sólida para incorporar más modelos con mayor rapidez.
- El desarrollo de modelos en MLFlow facilita su monitoreo