

PREDICCIÓN AVANZADA DE LTV, CAC Y ROMI PARA SHÓWZ

PRESENTED BY

LUCIANA VALDIVIESO
NAHOMY SALAZAR
MATEO SALGADO
PABLO HERRERA
JOSUE CARDENAS

https://github.com/Pablouski7/proyecto_final_DM

El Reto de Marketing en Showz

- Contexto: Showz invierte en diversos canales de marketing digital para adquirir clientes.
- Problema:
 - ¿Se asigna el presupuesto de manera óptima para maximizar el retorno?
 - ¿Qué canales atraen a los clientes más valiosos a largo plazo?
 - ¿Cuál es el costo real de adquisición por fuente?
- Necesidad: Una estrategia basada en datos para mejorar el ROMI (Retorno de la Inversión en Marketing).



Objetivos del Proyecto

Meta Principal: Desarrollar un sistema predictivo para guiar la asignación del presupuesto de marketing y maximizar la rentabilidad.



Predecir el LTV (Valor del Cliente a 180 días).
Estimar el CAC (Costo de Adquisición por fuente a 30 días).



Identificar los factores clave (drivers) de LTV y CAC.



Recomendar una estrategia de inversión basada en ROMI predicho.

Datos y Preparación

360k

Visitas

Sesiones de usuarios

50k

Ordenes

Compras e ingresos

2.5k

Costos

Gasto diario en marketing
por fuente

Proceso Clave

- Limpieza y estandarización de datos.
- Ingeniería de Características Avanzada: Creación de ~20 variables nuevas (Comportamiento, Temporales, Marketing, Cohortes).
- Cálculo de LTV_180 y CAC_source_30 como variables objetivo.
- Generación de un dataset final a nivel de usuario.



Enfoque de Modelado Predictivo

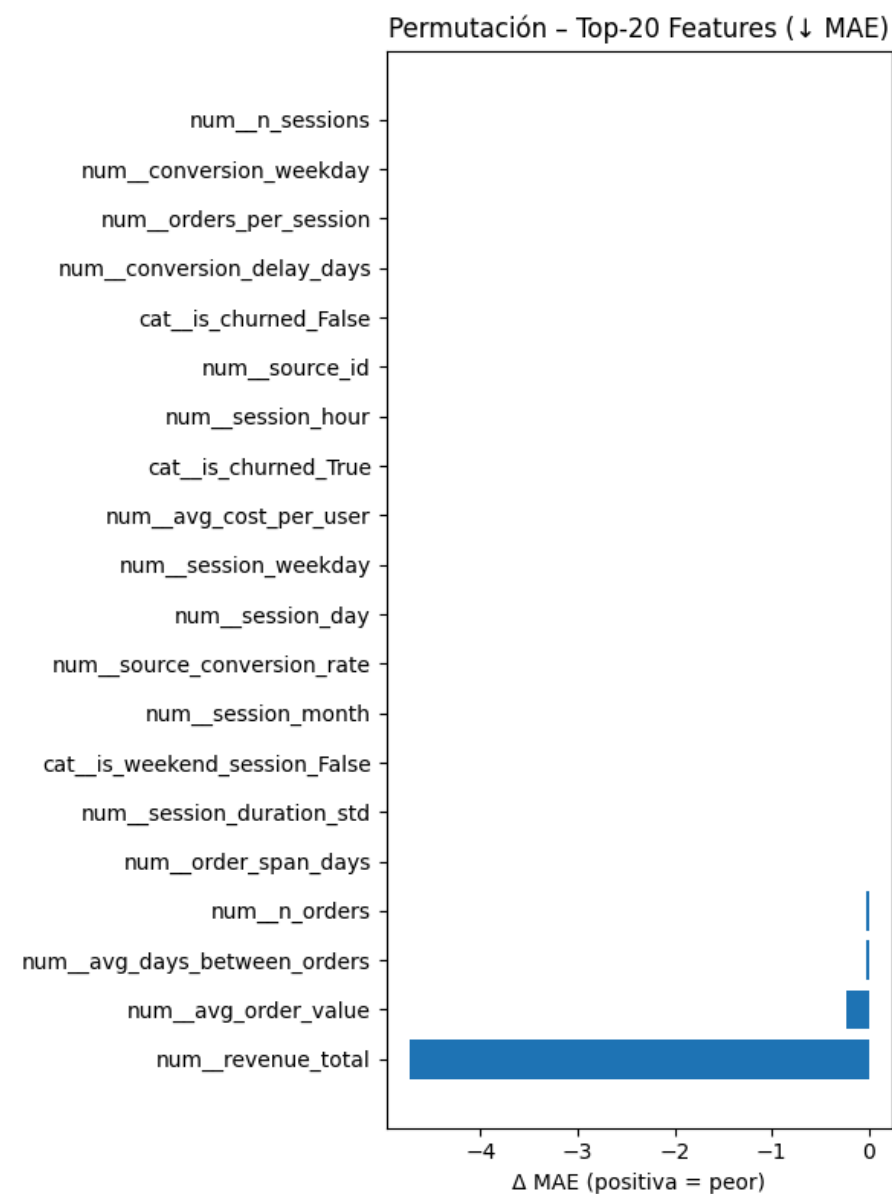
- Tarea: Dos modelos de Regresión (LTV y CAC).
- Modelos Evaluados: Lineales, Árboles (Random Forest, XGBoost), Ensambladores (Stacking), Gradient Boosting.
- Validación Rigurosa:
 - División Temporal: Entrenamiento (2017) vs. Validación (H1 2018) – Simula predicción futura.
 - Optimización: Validación Cruzada Temporal (TimeSeriesSplit) + GridSearchCV.
- Modelo Seleccionado: Gradient Boosting optimizado mostró el mejor rendimiento predictivo en la validación temporal para ambas tareas.



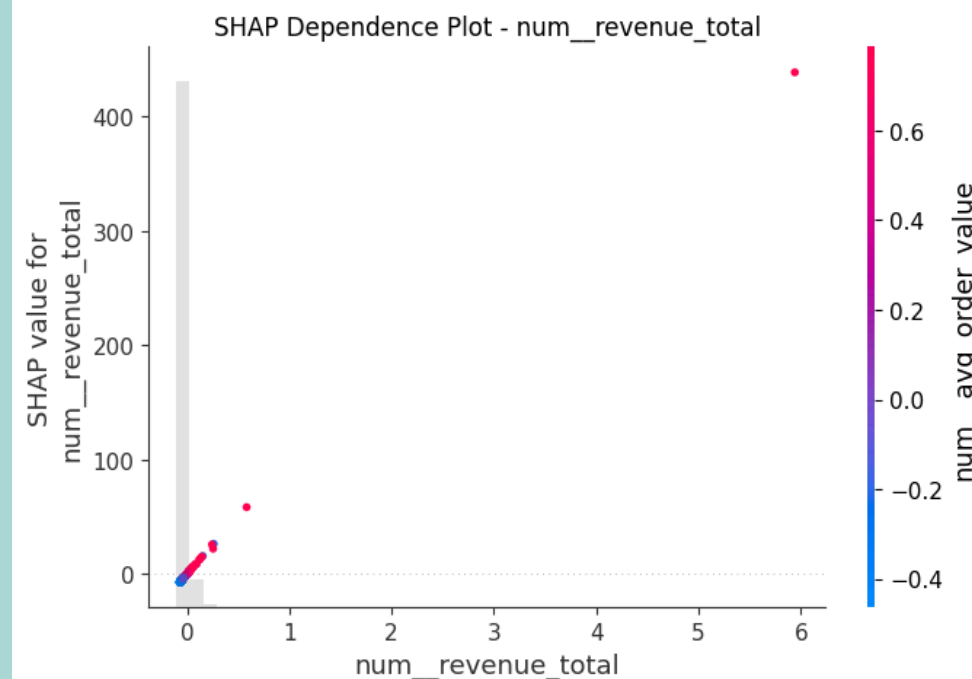
Qué Impulsa el Valor del Cliente (LTV)?



Qué Impulsa el Valor del Cliente (LTV)?



La gráfica muestra que las variables más importantes para predecir el CAC son revenue_total, avg_order_value y avg_days_between_orders, ya que al alterarlas el error del modelo aumenta notablemente. Esto indica que el comportamiento de compra influye directamente en el costo de adquisición.



El gráfico SHAP muestra que a mayor revenue_total, mayor es el impacto positivo en la predicción del CAC, especialmente cuando el avg_order_value también es alto, indicando que usuarios con más ingresos y ticket promedio elevan el costo de adquisición.

Qué Impulsa el Costo de Adquisición (CAC)?

Otros Drivers Relevantes (SHAP/Gain):

Costo Promedio por Usuario (Fuente) (avg_cost_per_user): Relación directa y fuerte con el CAC predicho.

Mes de la Sesión (session_month): Factor temporal que influye en el costo.

Tasa de Conversión (Fuente) (source_conversion_rate): Fuentes con menor conversión tienden a tener CAC predicho más alto.

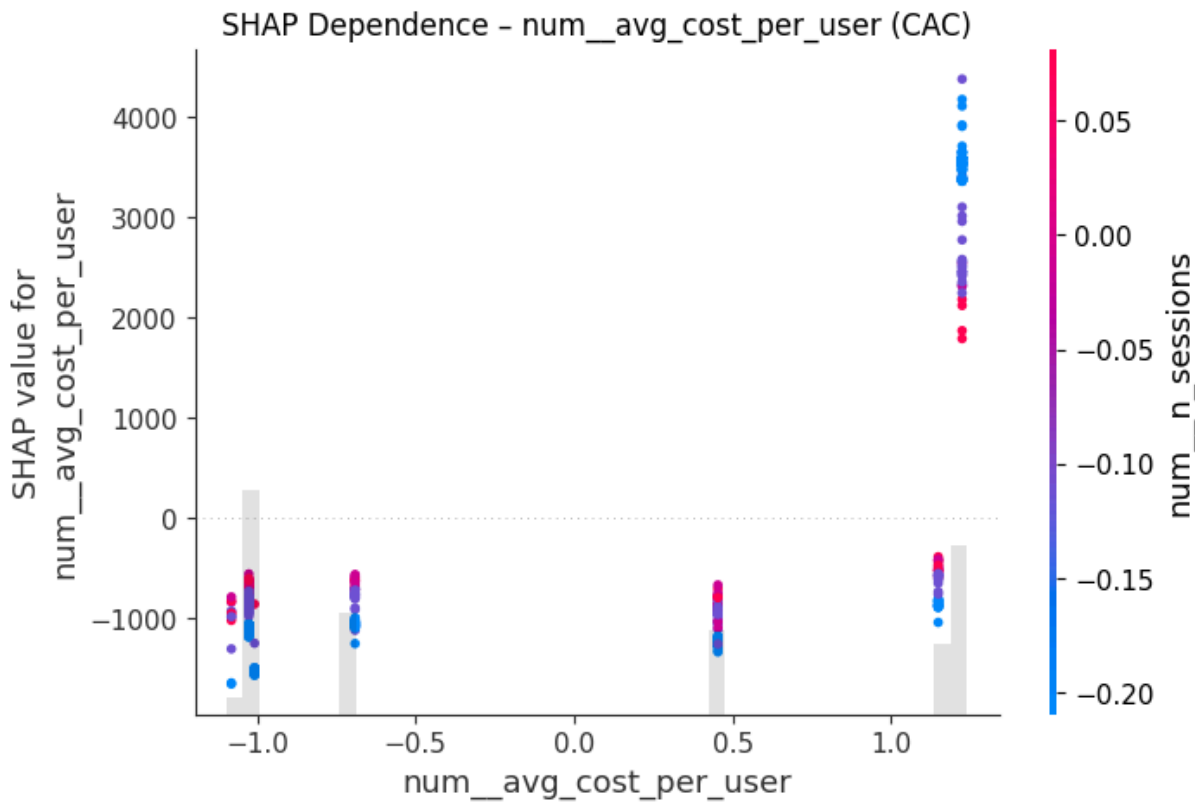
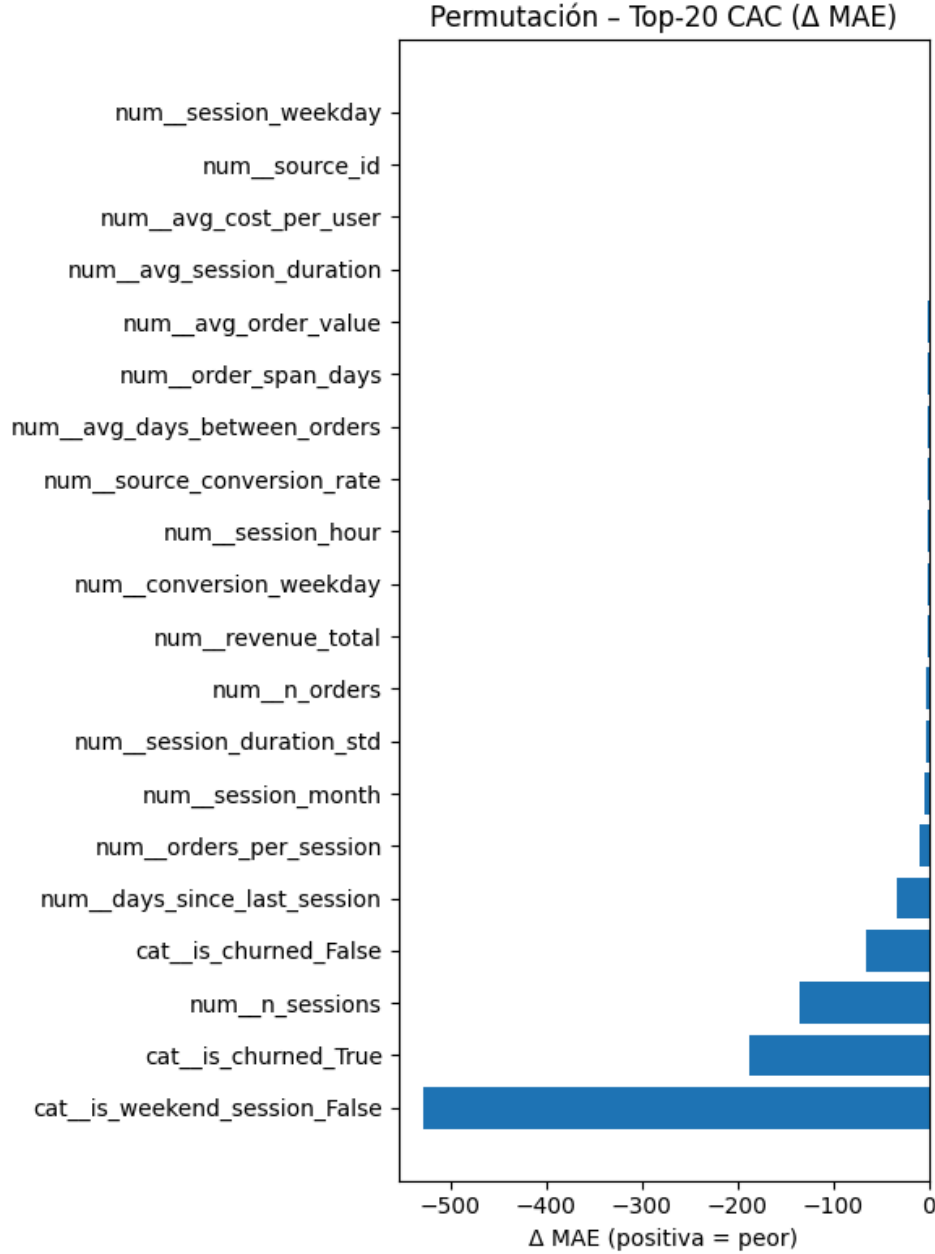
Drivers Clave (Impacto en Error):



¿Sesión entre Semana? (is_weekend_session_False): Factor muy importante según permutación.



¿Usuario Inactivo? (is_churned_True): Reactivar usuarios inactivos es costoso.



Simulación de Estrategias de Presupuesto

Objetivo: Comparar el retorno esperado de diferentes formas de asignar un presupuesto total de \$30,000, usando las predicciones de LTV y CAC para estimar el ROMI por fuente.

ESCENARIOS:

Baseline: Asignación igualitaria (\$3,333 / fuente activa).



+10% a Fuente 10: Aumento táctico a la fuente con mayor ROMI predicho inicial.

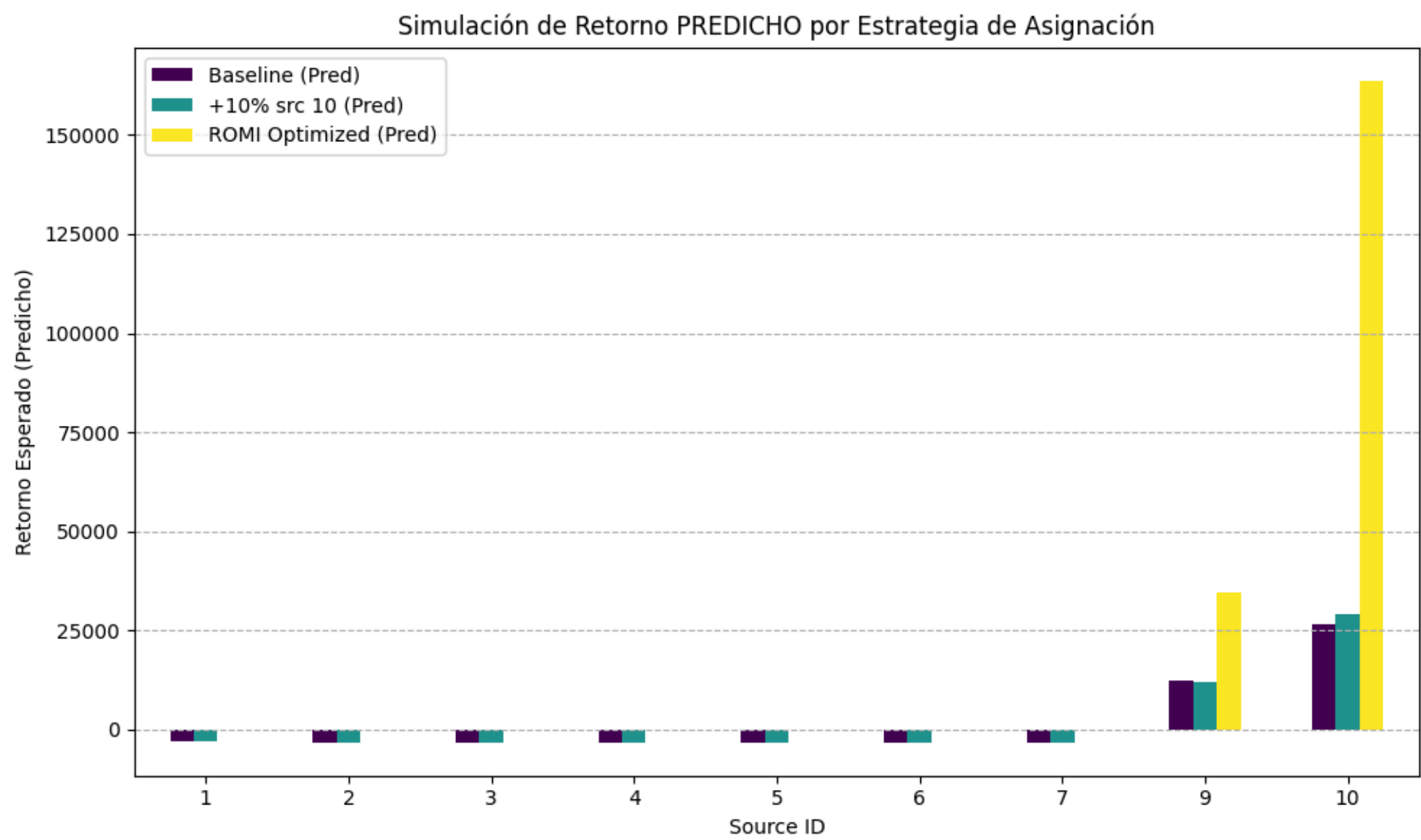


ROMI Optimized: Asignación proporcional solo a fuentes con ROMI predicho positivo Fuente 10 y 9



Resultados de la Simulación: Retorno Estimado

La optimización basada en ROMI predicho genera un retorno significativamente mayor.



Retornos Totales Predichos:

Baseline: ~\$15,623

+10% a Fuente 10: ~\$18,079

ROMI Optimized: ~\$198,392

ESTRATEGIA RECOMENDADA: ROMI OPTIMIZED

A

Acción:

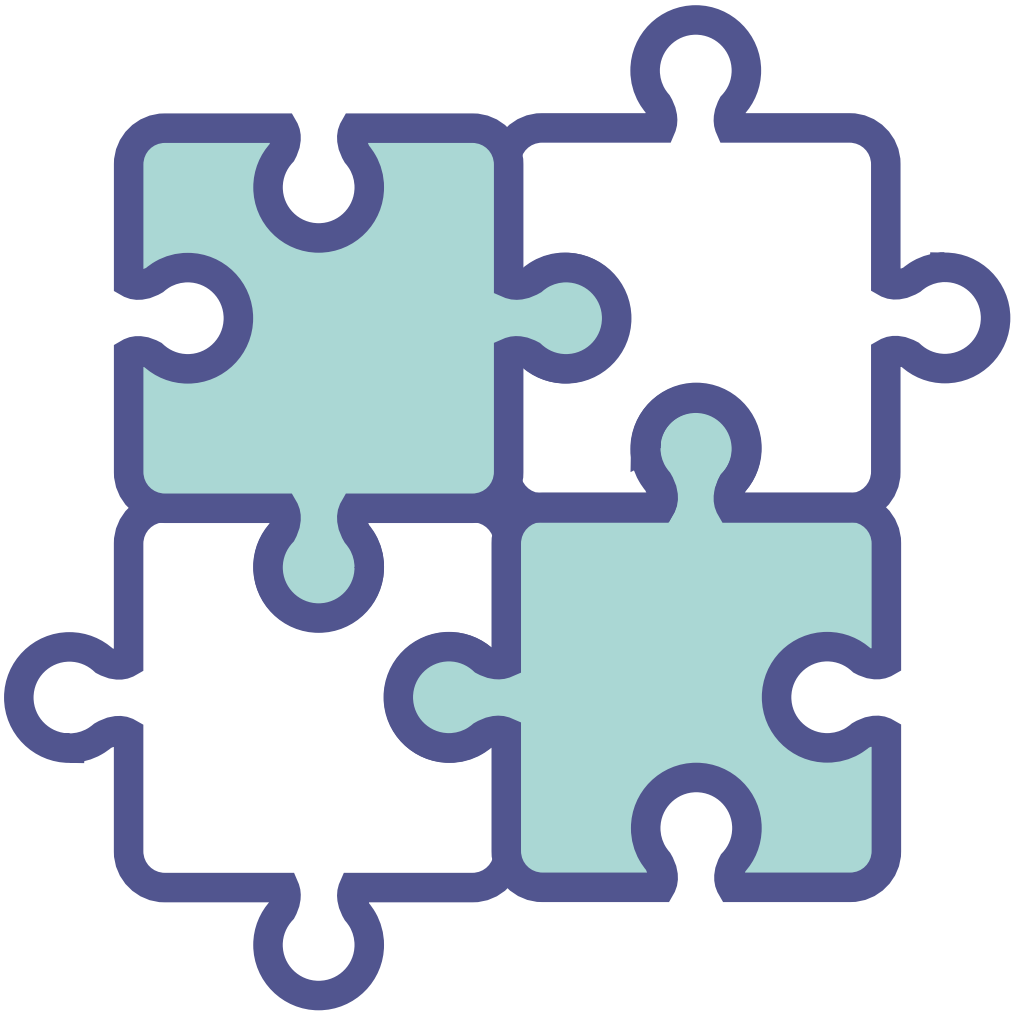
Reasignar el presupuesto de \$30,000 para enfocar la inversión en las fuentes con mayor potencial de rentabilidad demostrado por las predicciones.

F

FUENTE

Fuente 10: 68.5% (\$20,549)
Fuente 9: : 31.5% (\$9,451.28)

Asignación Sugerida:



O

Otras Fuentes:

0% (o presupuesto mínimo para exploración/prueba)

I

Impacto Estimado:

Ganancia potencial de ~\$183K sobre la estrategia baseline (basado en predicciones).

Conclusiones y Próximos Pasos

El modelado predictivo identifica una oportunidad clara para multiplicar el ROMI (~12x vs. baseline predicho).

Las fuentes 9 y 10 son los motores de rentabilidad actuales.

El valor histórico y el costo por usuario son drivers clave, pero el contexto temporal y de comportamiento (ej. actividad entre semana, churn) también son cruciales para CAC.

Próximos Pasos:



Implementar: Ajustar gradualmente la asignación de presupuesto hacia las fuentes 9 y 10.



Monitorear: Validar las predicciones midiendo LTV, CAC y ROMI reales por fuente.



Iterar: Re-entrenar los modelos periódicamente (ej. trimestral) con datos nuevos y refinar las características.

GRACIAS