

## Descripción del proyecto

La cadena de gimnasios Model Fitness está desarrollando una estrategia de interacción con clientes basada en datos analíticos.

Uno de los problemas más comunes que enfrentan los gimnasios y otros servicios es la pérdida de clientes. ¿Cómo descubres si un/a cliente ya no está contigo? Puedes calcular la pérdida en función de las personas que se deshacen de sus cuentas o no renuevan sus contratos. Sin embargo, a veces no es obvio que un/a cliente se haya ido: puede que se vaya de puntillas.

Los indicadores de pérdida varían de un campo a otro. Si un usuario o una usuaria compra en una tienda en línea con poca frecuencia, pero con regularidad, no se puede decir que ha huido. Pero si durante dos semanas no ha abierto un canal que se actualiza a diario, es motivo de preocupación: es posible que tu seguidor o seguidor/a se haya aburrido y te haya abandonado.

En el caso de un gimnasio, tiene sentido decir que un/a cliente se ha ido si no viene durante un mes. Por supuesto, es posible que estén en Cancún y retomen sus visitas cuando regresen, pero ese no es un caso típico. Por lo general, si un/a cliente se une, viene varias veces y luego desaparece, es poco probable que regrese.

Con el fin de combatir la cancelación, Model Fitness ha digitalizado varios de sus perfiles de clientes. Tu tarea consiste en analizarlos y elaborar una estrategia de retención de clientes.

Tienes que:

- Aprender a predecir la probabilidad de pérdida (para el próximo mes) para cada cliente.
- Elaborar retratos de usuarios típicos: selecciona los grupos más destacados y describe sus características principales.
- Analizar los factores que más impactan la pérdida.
- Sacar conclusiones básicas y elaborar recomendaciones para mejorar la atención al cliente:
  - identificar a los grupos objetivo;
  - sugerir medidas para reducir la rotación;
  - describir cualquier otro patrón que observes con respecto a la interacción con los clientes.

## Instrucciones para completar el proyecto

### Paso 1. Descargar los datos

Model Fitness te aportó archivos CSV que contienen los datos sobre la cancelación de un mes en concreto e información del mes que lo precedía. El *dataset* incluye los siguientes campos:

- 'Churn' — la cancelación para el mes en cuestión
- Campos de *dataset* actuales:
  - Datos del usuario del mes anterior
    - 'gender'.
    - 'Near\_Location' — si el/la usuario/a vive o trabaja en el vecindario donde se encuentra el gimnasio.
    - 'Partner' — si el/la usuario/a trabaja en una compañía asociada (el gimnasio tiene empresas asociadas cuyos empleados obtienen descuentos; en esos casos el gimnasio almacena información sobre los empleadores de los clientes).
    - Promo\_friends — si el/la usuario/a originalmente se inscribió mediante una oferta “trae a un/a amigo/a” (se utilizó el código promocional de un/a amigo/a cuando pagaron el primer abono).
    - 'Phone' — si el/la usuario/a aportó el número de teléfono.
    - 'Age'.
    - 'Lifetime' — el tiempo (en meses) desde que el/la usuario/a llegó por primera vez al gimnasio.
- Datos del registro de visitas y compras y datos sobre el estado actual de la membresía:
  - 'Contract\_period' — 1 mes, 3 meses, 6 meses o 1 año.
  - 'Month\_to\_end\_contract' — los meses que faltan hasta que expire el contrato.
  - 'Group\_visits' — si el/la usuario/a participa en sesiones grupales.
  - 'Avg\_class\_frequency\_total' — frecuencia media de visitas por semana a lo largo de la vida del cliente.
  - 'Avg\_class\_frequency\_current\_month' — frecuencia media de visitas por semana durante el mes en curso.
  - 'Avg\_additional\_charges\_total' — cantidad total de dinero gastado en otros servicios del gimnasio: cafetería, productos deportivos, cosméticos, masajes, etc.

Ruta de archivo: /datasets/gym\_churn\_us.csv. [Descargar el dataset](#).

### Llevar a cabo el análisis exploratorio de datos (EDA)

- Observa el *dataset*: ¿contiene alguna característica ausente? Estudia los valores promedio y la desviación estándar (utiliza el método describe()).

- Observa los valores medios de las características en dos grupos: para las personas que se fueron (cancelación) y para las que se quedaron (utiliza el método `groupby()`).
- Traza histogramas de barras y distribuciones de características para aquellas personas que se fueron (cancelación) y para las que se quedaron.
- Crea una matriz de correlación y muéstrala.

### **Paso 3. Construir un modelo para predecir la cancelación de usuarios**

Crea un modelo de clasificación binaria para clientes donde la característica objetivo es la marcha del usuario o la usuaria el mes siguiente.

- Divide los datos en conjuntos de entrenamiento y validación utilizando la función `train_test_split()`.
- Entrena el modelo en el set de entrenamiento con dos métodos:
  - regresión logística;
  - bosque aleatorio.
- Evalúa la exactitud, precisión y *recall* para ambos modelos utilizando los datos de validación. Utilízalos para comparar los modelos. ¿Qué modelo dio mejores resultados?

Recuerda indicar el parámetro `random_state` cuando dividas los datos y defines el algoritmo.

### **Paso 4. Crear clústeres de usuarios/as**

Deja de lado la columna con datos sobre la cancelación e identifica los clústeres de objetos (usuarios/as):

- Estandariza los datos.
- Utiliza la función `linkage()` para crear una matriz de distancias basada en la matriz de características estandarizada y trazar un dendrograma. Nota: ¡renderizar el dendrograma puede llevar tiempo! Utiliza el gráfico resultante para estimar el número de clústeres que puedes destacar.
- Entrena el modelo de *clustering* con el algoritmo K-means y predice los clústeres de clientes. (Deja que el número de clústeres sea  $n=5$  para que sea más sencillo comparar los resultados con los del resto del estudiantado. Sin embargo, en la vida real, nadie te dará tales pistas, así que tendrás que decidir basándote en el gráfico del paso anterior.)
- Mira los valores medios de característica para los clústeres. ¿Hay algo que te llame la atención?
- Traza distribuciones de características para los clústeres. ¿Notas algo?

- Calcula la tasa de cancelación para cada clúster (utiliza el método `groupby()`).  
¿Difieren en términos de tasa de cancelación? ¿Qué grupos son propensos a irse y cuáles son leales?

### **Paso 5. Saca conclusiones y haz recomendaciones básicas sobre el trabajo con clientes**

Llega a conclusiones y formula recomendaciones con respecto a la estrategia para la interacción y retención de clientes.

No necesitas entrar en detalles. Bastarán tres o cuatro principios esenciales y ejemplos de su implementación en forma de pasos de *marketing* específicos.