

Respuestas:

1. **Regresión (Análisis de series temporales):** para analizar cuánto tiempo transcurre antes de que se haga evidente el impacto producido por las campañas de mercadeo lanzadas a través de cada canal en los últimos 5 años, se puede realizar un análisis de series temporales con base en las cifras de ventas para cada canal de mercadeo, a fin de determinar los cambios en las actividades de ventas una vez que las campañas de mercadeo fueron lanzadas. Primero, los datos de ventas de la tienda y los datos de ventas online deben ser agregados para obtener una sola cifra de ventas cada hora. Se puede analizar el lapso transcurrido entre el inicio de una campaña y cualquier cambio en las cifras de ventas por hora para determinado canal de mercadeo a fin de establecer la efectividad del canal de mercadeo.
2. **Correlación:** se realiza una correlación entre las ventas generadas por campaña y el presupuesto designado a dicha campaña para cada canal de mercadeo, a fin de determinar si gasta la misma cantidad en cada canal de mercadeo tiene como resultado el mismo nivel de aumento de las ventas en todos los canales. Se deben agregar las cifras de ventas para cada campaña de mercadeo por canal durante los últimos cinco años. Se puede calcular un coeficiente de correlación de Spearman o Pearson, dependiendo del tipo de relación observada entre el presupuesto y las ventas para un canal de mercadeo.
3. **Regresión (Regresión lineal múltiple):** se puede aplicar la regresión lineal múltiple para desarrollar un modelo que permita encontrar la mejor asignación de presupuesto para cada canal de mercadeo y predecir las ventas resultantes. Las cifras de ventas pueden ser asignadas como la variable de respuesta, mientras que cada canal de mercadeo —como prensa, correo físico, televisión, radio, correos electrónicos y anuncios online— puede ser asignado como la variable explicativa. Se puede utilizar la magnitud de cada parámetro estimado por el modelo de regresión lineal múltiple para cada canal de mercadeo como una medida para la asignación de presupuesto.
4. **Clasificación (Regresión logística):** debido a que la variable de respuesta que se debe predecir es de naturaleza binaria (un cliente responderá o no a una oferta), es posible utilizar la regresión logística para construir un modelo de clasificación que pueda estimar la posibilidad de que un cliente responda a una campaña de mercadeo. Para esto, se debe combinar la información sobre dichos clientes que han respondido en el pasado con el dataset de clientes que contiene información sobre el salario de un cliente utilizando un identificador común. Posteriormente, se puede construir el modelo de clasificación utilizando el dataset combinado. Después de construir el modelo, aquellos clientes existentes que tienen valores desconocidos en relación sobre si respondieron o no a una campaña, pueden ser procesados para determinar qué clientes tienen una mayor probabilidad de responder a una campaña.
5. **Clustering (k-medias):** para obtener nuevos clientes, se deben agrupar los datos censales de cada código postal utilizando k-medias, a fin de hallar aquellos códigos postales que tienen una naturaleza similar. Los valores de característica deben ser normalizados para que las características de mayor valor no dominen a las de menor valor. Se deben guardar los valores de característica faltantes, asignándoles valores simulados apropiados, para que el número total de instancias se mantenga razonablemente alto y se generen clusters significativos. De acuerdo con los requisitos empresariales,

el valor de k se puede definir como 3 para que se ajuste al agrupamiento lógico de las tres clases de clientes de las campañas de mercadeo.

6. **Clustering (k-medias):** Se puede utilizar nuevamente el agrupamiento (Clustering) mediante k -medias para segmentar los clientes existentes y así encontrar grupos de clientes similares. Además, los datos agregados de ventas mensuales por categoría de producto por cliente se pueden combinar con el registro personal del cliente. Combinar una variedad de datasets generará un conjunto de características más amplio, lo que conducirá a clusters más significativos de clientes. El conjunto de características del cliente estará compuesto por 30 categorías de productos (cantidad de artículos comprados por categoría como la característica de valor) y cualquier característica relevante del registro personal del cliente, como edad y salario. Al no contar con ningún valor lógico de k que pueda derivarse de los requisitos del negocio, se debe probar el algoritmo de k -medias con diferentes valores de k hasta que se encuentren grupos de clientes homogéneos y significativos. Después, se puede utilizar la característica de valor del centroide de cada cluster para comprender las agrupaciones de clientes, lo que también se puede utilizar para enviar promociones específicas a los clientes de cada cluster.
7. **Regresión (Análisis de series temporales):** para determinar si las ventas de algunos artículos siguen ciertos patrones, se puede realizar un análisis de series temporales de los datos del programa de fidelidad de los clientes. Debido a la probabilidad de que haya una gran cantidad de datos, la cantidad de artículos vendidos se puede agregar a nivel semanal para determinar la estacionalidad. Esto permitirá que los gerentes de las tiendas almacenen los artículos sin tener un inventario excesivo durante los periodos de ventas bajas.
8. **Reglas de asociación:** con miras a abordar el problema de escasez y exceso de artículos, se pueden analizar los datos a nivel de tienda del programa de fidelidad de los clientes extrayendo las reglas de asociación, las cuales ayudarán a determinar los patrones de compra en una tienda en particular. Luego de conocer qué los artículos generalmente se compran juntos a nivel de tienda (además de ayudar con un mejor inventario), permite determinar mejor cómo se deben ubicar los productos tanto en las tiendas tradicionales como en la tienda online, se pueden combinar los registros de compras de los clientes contenidos en los datos del programa de fidelidad de clientes y de ventas online a clientes, y así obtener una visualización coherente del comportamiento de compras de los clientes, seguido de la extracción de reglas de asociación. Al encontrar los artículos que generalmente se compran juntos, dichos artículos se pueden agrupar tanto en las tiendas tradicionales como en las tiendas online, a fin de aumentar las ventas. Conocer los artículos que generalmente se compran juntos también permite crear un motor de recomendación básico para la tienda online que les sugiera artículos a los clientes a medida que añaden diferentes artículos a sus canastas virtuales.
9. **Clasificación:** en vista de que se cuenta con los datos de ejemplo relacionados con los clientes que registran incumplimientos en sus pagos, se puede utilizar un modelo de clasificación para predecir si un cliente incumplirá en el pago de su saldo. Se puede utilizar un algoritmo de reglas de clasificación para desarrollar un modelo basado en reglas que el equipo financiero comprenda fácilmente. Se deben usar las reglas de clasificación ya que se trata de un problema de clasificación en una sola clase,

además de la detección de fenómenos excepcionales (personas financieramente estables que incumplen en el pago de sus saldos en algunas ocasiones). Para desarrollar el modelo de clasificación, primero se deben combinar los datos del ERP (clientes morosos y puntuales con sus pagos) con los datos de CRM específicos de todos los clientes (ingresos, edad, propiedad de viviendas, y empleo), incluyendo aquellos para los cuales se dispone de información de incumplimiento y aquellos que aún no tienen una tarjeta de crédito. El siguiente paso es entrenar al clasificador utilizando dichos registros para los cuales se dispone de información de incumplimiento de pagos, y luego clasificar los demás registros con base en las reglas aprendidas. Se pueden utilizar técnicas de evaluación de modelos, tales como la matriz de confusión, para lograr la máxima exactitud en la clasificación.

10. **Regresión (Regresión lineal múltiple):** para desarrollar un modelo que pueda establecer un límite de crédito específico para cada cliente, se pueden utilizar múltiples regresiones lineales. Al igual que el modelo para clasificación predeterminada, se deben combinar los datos de ERP (total de reembolsos hechos) y los datos de CRM específicos de los clientes (ingresos, edad, propiedad de viviendas y empleo). Para crear el modelo, se puede utilizar el total de reembolsos realizados (límite crediticio) como la variable de respuesta, mientras que los ingresos, la edad, la propiedad de viviendas y el empleo se pueden utilizar como las variables explicativas.