**Datos para trabajar**: Marketing Campaign. Boost the profit of a marketing campaign

https://www.kaggle.com/rodsaldanha/arketing-campaign

**Contexto**

Este dataset se creó originalmente para predecir quién responderá a una oferta de un producto o servicio. Un modelo de respuesta puede proporcionar un impulso significativo a la eficiencia de una campaña de marketing al aumentar las respuestas o reducir los gastos.

En nuestro caso vamos a aprovechar este conjunto de datos para conocer mejor al cliente que está oculto en él.

**Objetivo del ejercicio**

Objetivo principal:

Definición que tipologías de clientes hay en esta empresa y cómo conseguir vender más a cada tipo.

Objetivos segundarios:

Aplicación de las técnicas vistas en clase:

* Análisis exploratorio de los datos.
* Detección y tratamiento de outliers y missing
* Creación de variables nuevas que consideréis necesarias para el análisis.
* Reducción de dimensiones para visualización y selección de posteriores algoritmos
* Clusterización (segmentación) de clientes.
* Analisis de asociaciones

La aportación clave es la Interpretación de negocio de este ejercicio. Seleccionar las variables y las técnicas siempre teniendo en cuenta esto

**Contenido del fichero**

* AcceptedCmp1 - 1 si el cliente aceptó la oferta en la 1ª campaña, 0 en caso contrario
* AcceptedCmp2 - 1 si el cliente aceptó la oferta en la 2ª campaña, 0 en caso contrario
* AcceptedCmp3 - 1 si el cliente aceptó la oferta en la 3ª campaña, 0 de lo contrario
* AcceptedCmp4 - 1 si el cliente aceptó la oferta en la 4ª campaña, 0 en caso contrario
* AcceptedCmp5 - 1 si el cliente aceptó la oferta en la 5ª campaña, 0 en caso contrario
* Response (target) - 1 si el cliente aceptó la oferta en la última campaña, 0 en caso contrario
* Complain - 1 si el cliente se quejó en los últimos 2 años
* DtCustomer - fecha de inscripción del cliente en la empresa
* Education - nivel de educación del cliente
* Marital - estado civil del cliente
* Kidhome - número de niños pequeños en el hogar del cliente
* Teenhome - número de adolescentes en el hogar del cliente
* Income - ingresos anuales del hogar del cliente
* MntFishProducts - cantidad gastada en productos de pescaderia en los últimos 2 años
* MntMeatProducts - cantidad gastada en productos cárnicos en los últimos 2 años
* MntFruits - cantidad gastada en productos de frutas en los últimos 2 años
* MntSweetProducts - cantidad gastada en productos dulces en los últimos 2 años
* MntWines - cantidad gastada en productos vitivinícolas en los últimos 2 años
* MntGoldProds - cantidad gastada en productos oro en los últimos 2 años
* NumDealsPurchases - número de compras realizadas con descuento
* NumCatalogPurchases - número de compras realizadas utilizando el catálogo
* NumStorePurchases - número de compras realizadas directamente en tiendas
* NumWebPurchases - número de compras realizadas a través del sitio web de la empresa
* NumWebVisitsMonth - número de visitas al sitio web de la empresa en el último mes
* Recency - número de días desde la última compra

Variables que podemos crear para el trabajo (y las que se os ocurran porque salen muchas mas que son interesantes)

* Variable ***Age*** en sustitución de la variable *Year\_birth*
* Variable ***Spending*** como la suma de la cantidad gastada en las 6 categorías de productos
* Variable ***Seniority*** como el número de meses que el cliente está inscrito en la empresa
* Variable ***Marital\_Status*** para agrupar los diferentes estados civiles en sólo 2 categorías integrales: En pareja vs Solo
* Variable ***Education*** ya sea como pregrado o posgrado
* Variable ***Children*** como el número total de niños en el hogar
* Variable ***Has\_child*** como una variable binaria igual a Sí si los clientes tienen 1 hijo o más

Para cada variable relacionada con las compras:

Definimos segmentos de clientes para cada producto en función de sus gastos:

1. No consumidor: Clientes con 0 cantidad de gasto

2. Bajo consumo: Clientes por debajo del 1er tercil

3. Consumidor frecuente: Clientes entre el 1er y 3er tercil

4. Mayor consumidor: Clientes por encima del 3er tercil

**Segmentación:**

1. Normalizar datos

2. Calcular el número óptimo de clústeres: Define el número de clústeres utilizando la puntuación de Silhoutte, Calinski-Harabasz, (en knime se llama Cascade simple k means ) y el método Elbow.

3. La solución podría ser algo así:

• Cluster 0 se compone de clientes antiguos con altos ingresos y alta cantidad de gastos (Stars)

• Cluster 1 se compone de nuevos clientes con ingresos por debajo de la media y una pequeña cantidad de gastos (Need attention)

• Cluster 2 se compone de nuevos clientes con altos ingresos y alta cantidad de gastos (High potential)

• Cluster 3 se compone de clientes antiguos con ingresos por debajo de la media y una pequeña cantidad de gastos (leaky bucket)

**Detalles de la tarea**

Eres un analista de marketing y el director de marketing te ha dicho que las campañas de marketing recientes no han sido tan efectivas como se esperaba que fueran. Debes analizar el conjunto de datos para comprender este problema y proponer soluciones basadas en datos.

**Tarea 1: Análisis exploratorio de datos**

¿Hay valores nulos o valores atípicos? ¿Cómo los manejarás?

¿Hay alguna variable que justifique transformaciones?

¿Hay alguna variable útil que pueda crear con los datos dados?

¿Notas algún patrón o anomalía en los datos? ¿Puedes trazarlos?

**Tarea2: Análisis estadístico**

Ejecuta pruebas estadísticas en forma de regresiones para responder a estas preguntas y proponer recomendaciones de acción basadas en datos a su CMO. Asegúrate de interpretar sus resultados con jerga no estadística para que tu CMO pueda entender tus hallazgos.

¿Qué factores están significativamente relacionados con el número de compras en la tienda?

¿A Estados Unidos le va significativamente mejor que al resto del mundo en términos de compras totales?

Su supervisor insiste en que las personas que compran oro son más conservadoras. Por lo tanto, las personas que gastaron una cantidad superior al promedio en oro en los últimos 2 años tendrían más compras en la tienda. Justificar o refutar esta afirmación utilizando una prueba estadística apropiada.

El pescado tiene ácidos grasos Omega 3 que son buenos para el cerebro. En consecuencia, ¿los "candidatos a doctorado casados" tienen una relación significativa con la cantidad gastada en pescado? ¿Qué otros factores están significativamente relacionados con la cantidad gastada en pescado? (Sugerencia: use su conocimiento de las variables / efectos de interacción)

¿Existe una relación significativa entre la región geográfica y el éxito de una campaña?

**Tarea 3: Visualización de datos**

Por favor, trace y visualice las respuestas a las siguientes preguntas.

¿Qué campaña de marketing tiene más éxito?

¿Cómo es el cliente promedio para esta empresa?

¿Qué productos están funcionando mejor?

¿Qué canales tienen un rendimiento inferior?

**Tarea 4: Recomendaciones y conclusiones**

Reúne todo, desde las Secciones 1 a 3 y proporciona recomendaciones / sugerencias basadas en datos a su CMO.

**Evaluación**

Me gustaría ver un proceso bien definido de análisis exploratorio y estadístico con conclusiones perspicaces.

Exploración de datos - ¿Se discutieron los datos correctamente? ¿Qué tan bien se analizaron los datos? ¿Hay visualizaciones útiles? ¿Aporta el alumno alguna técnica nueva a través de esta presentación? Una buena presentación será informativa y estimulante.

Análisis estadístico - ¿Se utilizaron las pruebas estadísticas correctas? ¿Qué tan bien se interpretó el resultado estadístico? Una buena presentación de los resultados sin el uso de ninguna jerga estadística es importante.

Recomendación empresarial - ¿Las recomendaciones estaban vinculadas a su análisis en las Secciones 1-3? ¿Están basados en datos y enfocados en conceptos de marketing como objetivos, canales o productos?

Documentación - ¿Su código, workflow y docuemento están bien documentados para que un lector pueda entender lo que hizo? ¿Se citan claramente sus fuentes? Un análisis de alta calidad debe ser conciso y claro en cada paso para que la justificación sea fácil de seguir y el proceso sea reproducible.