

Business Intelligence

- Reto 1 - Clustering -

1 Objetivos y Entregables

1.1 Objetivo General

Usar los datos para crear campañas para 2-5 grupos de clientes. Cada uno debe contener al menos el 10% de los datos totales. Se espera una presentación que expone los diferentes grupos de clientes y cuales campañas podrían ser apropiados para ellos.

1.2 Objetivos Subordenados

1. Encuentren una agrupación de clientes que es razonable en el sentido que cada uno de los grupos tiene una característica especial que justifica una campaña distinta a los demás.

Esa selección debe ser justificada basado en tecnologías de clustering aplicados a los datos originales (tal vez con un pre-procesamiento para limpieza pero sin eliminar atributos que podrían ser relevantes).

La selección de la cantidad de grupos de clientes también tiene que ser justificada desde métricas sobre clustering.
2. Creen una visualización 2D o 3D que mejor muestre los diferentes grupos (la separabilidad debe ser clara).
3. Creen visualizaciones que justifiquen la interpretación de los grupos. Es decir, las gráficas deberían mostrar claramente las características de las personas en los diferentes clusters y así explicar que hace diferente cada cluster de los demás.

1.3 Entregables

1. una presentación de 10 - 12 minutos (al menos 3 minutos por integrante). Puede ser sostenido en español o inglés.
2. un solo Python notebook en la que se realiza todo el estudio y donde se generan todas las gráficas que muestran durante de la presentación.

La presentación es en la semana del 8 - 12 de abril en el horario normal de la clase.

2 Métodos

Generalmente no hay restricciones frente a las herramientas que pueden usar. Está permitido usar técnicas que no se han visto en clase, *pero serán sujeto a preguntas en la sustentación!*. Así que todo lo que entregan también tienen que comprender técnicamente. Algunas recomendaciones incluyen:

1. revisar si se puede eliminar atributos o derivar nuevos atributos a partir de los que ya existen. Sin embargo, atributos pueden ser eliminados únicamente si hay una justificación porque *imposiblemente* contribuyen algo razonable al análisis.
2. considerar pre-procesamiento de los datos (transformaciones).
3. considerar la eliminación de atípicos.
4. considerar muchos tipos de visualizaciones (pueden explorar seaborn para más ideas).
5. considerar conceptos adicionales no vistos en clase que pueden ser útiles en el análisis (pueden consultar herramientas IA).

Pero de nuevo: No incluyan componentes que no entienden!

Sin embargo, en realidad ni siquiera tienen que usar técnicas de clustering si piensan que pueden lograr los objetivos mejor de otra forma.

3 Rúbrica de Evaluación

Concepto	100%	80%	50%	20%
Determinación de la cantidad de grupos	se escogió la mejor cantidad según métricas justificadas para el caso	la selección de grupos se justifica usando varias métricas apropiadas	se han analizado diferentes números y escogió la mejor	se presenta una cantidad aceptable de grupos
Identificación de los grupos	se optimizó la representación de datos para encontrar el mejor agrupamiento	han tenido en cuenta aspectos aleatorios de algoritmos de agrupación	han analizado diferentes agrupaciones y escogido la mejor según una métrica	se usaron algoritmos de clustering para identificar una agrupación aceptable
Visualización de grupos	hay visualizaciones para diferentes cantidades de grupos, justificando la selección final, y visualizaciones interactivas en 3D para estudiar los grupos en más detalle	hay visualizaciones que muestran lo que hace especial a cada uno de los clusters	la visualización muestra claramente los grupos y se pueden apreciar sus ubicaciones y bordes entre ellos	existe una visualización de los grupos y la distribución entre ellos
Presentación de la solución	hacen una interpretación elocuente de los resultados e pueden argumentar limitaciones de su análisis y como se podría mejorarlo	todas las decisiones y afirmaciones están bien fundamentadas (el <i>por qué</i>)	en la presentación y las preguntas se evidencia que los integrantes comprenden su propia solución (el <i>cómo</i>)	la presentación respeta el tiempo y cubre todos los objetivos subordinados
Propuestas de campañas (Interpretación de los grupos)	la forma planeada de ejecución de las campañas es innovadora y apropiado para los grupos (el <i>como</i>)	las campañas son significativamente diferentes y justificadas (el <i>por qué</i>)	las campañas son razonables según las características de cada grupo (el <i>qué</i>)	se propone una campaña coherente formulado para cada grupo que tiene sentido frente al negocio de que se trata

4 Información de Campos

4.1 Datos de Clientes

- ID: Identificador único del cliente.
- Year_Birth: Año de nacimiento del cliente.
- Education: Nivel de educación del cliente.
- Marital_Status: Estado civil del cliente.
- Income: Ingresos anuales del hogar del cliente.
- Kidhome: Número de niños en el hogar del cliente.
- Teenhome: Número de adolescentes en el hogar del cliente.
- Dt_Customer: Fecha de inscripción del cliente en la compañía.
- Recency: Número de días desde la última compra del cliente.
- Complain: 1 si el cliente presentó una queja en los últimos 2 años, 0 en caso contrario.

4.2 Productos

- MntWines: Cantidad gastada en vino en los últimos 2 años.
- MntFruits: Cantidad gastada en frutas en los últimos 2 años.
- MntMeatProducts: Cantidad gastada en carne en los últimos 2 años.
- MntFishProducts: Cantidad gastada en pescado en los últimos 2 años.
- MntSweetProducts: Cantidad gastada en dulces en los últimos 2 años.
- MntGoldProds: Cantidad gastada en productos de oro en los últimos 2 años.

4.3 Promociones

- NumDealsPurchases: Número de compras realizadas con descuento.
- 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 en caso 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: 1 si el cliente aceptó la oferta en la última campaña, 0 en caso contrario.

4.4 Lugares

- NumWebPurchases: Número de compras realizadas a través del sitio web de la compañía.
- NumCatalogPurchases: Número de compras realizadas utilizando un catálogo.
- NumStorePurchases: Número de compras realizadas directamente en tiendas.
- NumWebVisitsMonth: Número de visitas al sitio web de la compañía en el último mes.