

Desafío 2 - Segmentación de Clientes

En este desafío tendrás la oportunidad de poner a prueba los conceptos aprendidos durante la sesión. Los ejercicios están diseñados para reforzar practicar lo explicado en clases y poder implementar un caso real.

Lee todo el documento antes de comenzar el desarrollo **individual**, para asegurarte de tener el máximo de puntaje y enfocar bien los esfuerzos. Asegúrate de seguir las instrucciones específicas en cada ejercicio y de completar los requerimientos adicionales, si los hubiera. ¡A disfrutar aprendiendo!

Tiempo asociado: 4 horas cronológicas

Descripción

En el apasionante mundo del comercio minorista, el supermercado "FreshMart" se ha convertido en el lugar preferido para realizar las compras del día a día. Con el objetivo de brindar una experiencia de compra más personalizada, "FreshMart" ha decidido emplear técnicas de aprendizaje no supervisado para segmentar a sus clientes y diseñar estrategias de marketing adaptadas a sus necesidades y preferencias.

Con un amplio conjunto de datos recopilados durante el último año, que incluyen información de compras y perfiles de clientes, "FreshMart" cuenta con una valiosa fuente de información para el análisis. Ahora, el desafío radica en aplicar técnicas de clustering innovadoras para descubrir segmentos de clientes que comparten patrones de compra similares y comportamientos de consumo.

El objetivo principal de la segmentación de clientes es identificar grupos específicos dentro de la base de clientes de "FreshMart" que puedan beneficiarse de promociones y campañas de marketing personalizadas. La idea es poder caracterizar estos segmentos y ponerle un nombre a cada uno.

Una vez que se logre una segmentación efectiva, se abrirán las puertas a oportunidades emocionantes. El análisis detallado de cada segmento permitirá comprender los patrones de compra, las preferencias de productos y la disposición a participar en programas de fidelidad. Esto allanará el camino para el diseño de estrategias de marketing personalizadas, como descuentos exclusivos, promociones especiales y recomendaciones de productos relevantes.

El objetivo final de "FreshMart" es mejorar la satisfacción del cliente y optimizar la rentabilidad del supermercado. Al emplear técnicas de clustering avanzadas y aprovechar los datos históricos disponibles, "FreshMart" estará en condiciones de ofrecer una experiencia de compra más personalizada y adaptada a las necesidades de cada segmento de clientes. Esto consolida la fidelidad de los clientes existentes y atrae a nuevos clientes que buscan un supermercado que se preocupe por sus preferencias individuales.

La data entregada esta a nivel de boleta y tiene los siguientes campos:

- **InvoiceNo:** Nro identificador de cada boleta
- **InvoiceDate:** Fecha de la boleta
- **CustomerId:** Nro identificador de cada usuario
- **Quantity:** Cantidad de unidades totales compradas
- **price_total:** Precio total de la boleta
- **StockCode:** Cantidad de unidades diferentes compradas

A partir de estos datos se te solicita:

1. Leer el archivo **Retail_Invoices.xlsx** y realizar un análisis de calidad de datos, para definir la calidad de la data entregada considerando valores anómalos, valores faltantes, etc. Limpia lo que sea necesario y realiza un análisis exploratorio de los datos considerando gráficos de distribuciones y correlaciones entre estos.
2. Generar variables para la segmentación. Notar que la granularidad de la data es por boleta y se necesita segmentar clientes, por lo que es necesario generar variables a nivel de clientes. Se propone generar al menos RFM (Recencia, frecuencia y monto).
3. Realizar un análisis exploratorio y las transformaciones necesarias a las nuevas variables.
4. Aplicar al menos 2 algoritmos de clustering. Buscar el número indicado de clusters que pueden tener y visualizar los cluster resultantes de los algoritmos.
5. Evaluar la calidad de los clustering de forma numérica y realizar una explicación de cada segmento asignándoles un nombre descriptivo.

Requerimientos

1. Analiza y limpia los datos entregados considerando los requerimientos y genera las variables para la segmentación correspondientes. **(3 puntos)**
2. Aplica algoritmos de clustering, calculando la cantidad óptima y visualizándolos. **(4 puntos)**
3. Evalúa la calidad de los clusterings y los explica, aplicándolos a la segmentación. **(3 puntos)**



¡Mucho éxito!

Consideraciones y recomendaciones

- Debes entregar la solución en un archivo de Jupyter Notebook, con el código y las explicaciones necesarias
- Toma en cuenta que la granularidad de la data es a nivel de boleta y la segmentación se pide a nivel de clientes.
- Las variables a considerar en la segmentación van a definir los cluster a realizar, por lo que es importante la elección para la interpretación posterior.
- Cualquier supuesto que se tome de las variables a utilizar y transformaciones a realizar tienen que estar en el documento.
- Toma en cuenta que no son personas necesariamente expertas en retail, por lo que el criterio experto no aplica, pero si es necesaria la imaginación y el criterio en base a los datos para la descripción final de los clusters.