

Fase I. Anteproyecto

Estudiantes: Graciela Rivera Picado | Luis E. Vargas Porras

1. Descripción

[Situación ficticia con fines académicos]

Recientemente Target Global lanzó a inicios del Q4, 2022 su primera tienda online donde desean colocar sus productos más populares e incluir un motor de recomendación en su página oficial, utilizando así la popularidad de un producto para impulsar las ventas del otro. Además, desean incrementar sus ventas por este canal, agrupando los productos asociados y aplicar técnicas de precios o paquetes promocionales entre artículos.

Uno de los motivos que los directores de Target Global decidieron la incorporación de este nuevo canal de ventas, es debido a la compra de un número significativo de “Tea cups” debido a las recientes fiestas en Turquía. Sin embargo, no esperaban que el reciente terremoto en el país fuera a impactar sus inventarios y añadiendo a lo anterior, el margen de beneficio de estos productos es tan bajo que no pueden permitirse un descuento promocional sin afectar sus beneficios.

Nos solicitaron que realizáramos un análisis de los datos de los tickets de compra de la tienda de un periodo de un año para saber qué artículos se compran habitualmente juntos, y poder con ellos solucionar el problema de negocio el cual se enfrentan actualmente.

2. Fase de entendimiento del negocio

2.1 Determinar los objetivos de negocio

Objetivos de negocio: Identificar aquellas transacciones de órdenes de compra de Target Global que sean altamente asociadas entre ellas con el fin de aumentar las promociones de estos productos para el Q1, 2023 y sirva de información para el motor de recomendación de su tienda online.

Criterios de éxito (en términos de negocio):

- 1) Aumentar en un 3% las promociones de los productos que tenga una fuerte asociación para el próximo Q3, 2023.
- 2) Aumentar las ventas del negocio en 5% para el próximo Q3, 2023, y un 7% para el Q4, 2023 respectivamente.

2.2 Evaluación de la situación actual

2.2.1 Inventario de recursos: La lista de recursos disponibles para este proyecto se cita a continuación:

Personal y Recurso Humano: Se tiene a su disposición por tiempo completo a dos profesionales en el área de ciencia de datos y analítica, nivel senior.

Data: En cuanto a las fuentes de datos que se analizan, se tiene a disposición un repositorio de datos en el cual se tienen los permisos de acceso. El data set se describe de forma muy general seguidamente:

El data set “Online Retail” contiene un total de 541909 filas (transacciones), y cuenta con ocho atributos:

Atributo	Tipo de Dato	Descripción
InvoiceNo	int	ID de la factura. Si este código empieza por la letra "C", el pedido ha sido anulado.
StockCode	int	ID del Stock
Description	string	Descripción del producto comprado
Quantity	int	El número de cada producto (artículo) comprado por transacción.
InvoiceDate	Fecha/Hora	La fecha y hora en que se completó cada transacción.
UnitPrice	float	Precio por unidad
CustomerID	int	ID del client
Country	String	País de la compra

Recursos computacionales y de software: En cuanto a recursos computacionales y de software, se requerirá la instalación previa de cuadernos en jupyter para correr los algoritmos en Python, así como sus librerías correspondientes. En cuanto a herramientas de BI y de visualización, se requerirá la instalación y la licencia de Tableau para la visualización de los datos.

2.2.2 Supuestos

En cuanto a la calidad y consistencia de los datos que se poseen, se asume que éstos han sido previamente extraídos, transformados y cargados al datamart, el cual tiene únicamente los datos puntuales que se necesitan para responder todas las preguntas que el negocio requiere; por lo que su número de columnas es limitado, excluyendo aquellos sin importancia.

Adicionalmente a esto, al modelo analítico que se aplicará (Algoritmo A priori) se tiene como supuesto general (*propio del algoritmo*) de que cualquier subconjunto de un conjunto de elementos frecuentes; debe ser también frecuente.

En cuanto a las limitaciones del proyecto, únicamente se analizarán los productos cuyas transacciones han sido efectuadas en la página online, por lo que se deberá de analizar por separado cómo es el comportamiento de los productos en tiendas físicas.

2.2.3 Restricciones

En cuanto a las restricciones del proyecto se tiene la disponibilidad del presupuesto que la empresa dispondrá para el análisis del proyecto e integración de la solución.

2.2.4 Riesgos y contingencias.

Los riesgos y contingencias del proyecto para este análisis son:

Riesgo	Contingencia
Calidad y cantidad de la información en la base de datos	En caso de que haya inconsistencia de los datos o el número de entradas no sea suficientemente representativo, se solicitará a la empresa la corrección de estos temas para su debido tratamiento.
Asociación baja entre productos al concluir el análisis propuesto	Como plan de contingencia se tendrá que aplicar otros algoritmos distintos al A priori y se analizarán otros similares para evaluar su desempeño.
Sesgo de resultados debido a factores como la ubicación o geografía de las compras hechas por los usuarios	Como plan de contingencia en caso de requerirlo se dirigirá el análisis por sector en vez de global.

2.2.4 Beneficios (Del proyecto de MD al negocio)

En cuanto a beneficios se tienen:

1. Incremento en las ventas del negocio lo que se traduce en mayores ingresos o ganancias de la empresa.
2. Rotación de productos con poca venta.
3. Optimización de inventario, lo que impacta directamente el “cash flow” de la empresa y el servicio al cliente ya que se contará con disponibilidad del producto; según la preferencia de este.
4. Menor riesgo en obsolescencia de productos.
5. Distribución estratégica de portafolio de productos, lo que mejora el marketing de productos al identificar espacios premium, y la colocación estratégica en bodega de productos para mejorar los tiempos de reemplazo o suministro de los productos hacia los anaqueles o góndolas.

2.3 Determinar los objetivos de minería de datos

Objetivos de minería de datos: Desarrollar un modelo algorítmico A priori para identificar la fuerza de asociación entre los productos de la tienda Online Target Global.

Criterios de éxito (desde la perspectiva de minería de datos).

Para nuestro valor mínimo de soporte, especificaremos el 3%, lo que significa que solo se devolverán los artículos cuya frecuencia de aparición en los datos sea al menos del 3%.

3. Referencia base de datos:

<https://www.kaggle.com/code/rockystats/apriori-algorithm-or-market-basket-analysis/data?select=data-2.csv>