

Trabajo Integrador Final

Introducción

El objetivo del trabajo integrador es aplicar los conocimientos teóricos vistos en la materia, principalmente la exploración y preparación de datos, para la aplicación de reglas de asociación y detección de secuencias.

Entregables

Se debe entregar un script con el código utilizado para resolver el trabajo (correctamente comentado) y un informe en formato PDF con las respuestas a las actividades planteadas.

Fecha límite de entrega: miércoles 26/05 a las 23:59 hs.

Contexto del problema

El comercio electrónico se ha convertido en un nuevo canal para apoyar el desarrollo de las empresas. A través del comercio electrónico, las empresas pueden acceder y establecer una mayor presencia en el mercado proporcionando canales de distribución más baratos y eficientes para sus productos o servicios. El comercio electrónico también ha cambiado la forma en que la gente compra y consume productos y servicios. Muchas personas recurren a sus ordenadores o dispositivos inteligentes para pedir productos, que pueden ser entregados fácilmente en sus hogares.

Un minorista quiere dirigirse a los clientes con sugerencias sobre el conjunto de artículos que es más probable que un cliente compre, lo que permitirá aumentar el compromiso de los clientes, mejorar su experiencia e identificar su comportamiento.

Datos

Se trata de un conjunto de datos de transacciones de ventas del comercio electrónico (venta en línea) con sede en el Reino Unido durante un año. Esta tienda londinense vende regalos y artículos para el hogar para adultos y niños a través del sitio web desde 2007. Sus clientes proceden de todo el mundo y suelen hacer compras directas para ellos. Aunque también hay pequeñas empresas que compran al por mayor y venden a otros clientes a través de canales de venta minoristas.

El conjunto de datos contiene más de 500000 filas y 7 columnas entre el 01/12/2010 y el 09/12/2011. Las variables disponibles son:

- BillNo (categórica): número único de seis dígitos asignado a cada transacción.
- Itemname (categórica): nombre del producto/artículo.
- Quantity (numérica): la cantidad de cada producto por transacción. Los valores negativos indican transacciones anuladas.

- Date (fecha): la fecha y hora en que se generó cada transacción.
- Price (numérica): el precio de cada producto por unidad en libras esterlinas (£).
- CustomerID (categórica): número único de cinco dígitos asignado a cada cliente.
- Country (categórico): nombre del país donde reside el cliente.

Actividades

- a) Explore los datos y presente sus características principales.
- b) Utilizando las gráficas que considere adecuadas, muestre: compras por cliente, compras por país e ítems por compra.
- c) ¿Cómo ha sido la evolución de las ventas a lo largo de los meses?
- d) Encuentre el número de transacciones e ítems.
- e) Encuentre un conjunto de secuencias frecuentes para un soporte mínimo de 2%.
- f) ¿Cuáles son los 10 productos más comprados?
- g) Encuentre las secuencias más frecuentes que tienen 2 o más elementos y un soporte mayor a 3%.
- h) Transforme la base de datos original (secuencial) en una base transaccional en formato "basket", para obtener las reglas de asociación de las transacciones que se originaron solamente en el Reino Unido. Considere que las compras del mismo día de un cliente son una única transacción. (soporte mínimo 1% y confianza mínima 70%). Muestre las 10 reglas de mayor *lift*.