

Secuencias frecuentes 2

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) Encuentre el número de transacciones e ítems.
- d) Encuentre un conjunto de secuencias frecuentes para un soporte mínimo de 2%.
- e) Encuentre las secuencias más frecuentes que tienen 2 o más elementos y un soporte mayor a 3%.