UNIVERSIDAD NACIONAL DE ENTRE RÍOS

FACULTAD DE INGENIERÍA

Minería de Datos

Informe del Trabajo Integrador Final

Ing.Juan Aued,Lic.Mariana Blanco

Gareis Daiana, Letelier Sherly, Schmidt Leandro

2025

Introducción

El comercio electrónico se ha convertido en una herramienta clave para el crecimiento de muchas empresas. A través de este canal, las organizaciones pueden llegar a más clientes y reducir los costos de distribución de sus productos. Además, permite conocer mejor los hábitos de consumo y mejorar la experiencia de compra.

Este trabajo se centra en una tienda en línea polaca que vende ropa para embarazadas. El objetivo es analizar los datos de navegación de los usuarios para entender su comportamiento y así generar recomendaciones de productos. Esto puede ayudar a aumentar el compromiso del cliente, mejorar el servicio y optimizar las estrategias de venta.

La información utilizada corresponde al registro de clics realizados por los usuarios durante cinco meses del año 2008. Incluye detalles como país de origen, productos vistos, colores, precios y categorías, entre otros. A partir de este conjunto de datos, se aplican técnicas de minería de datos para extraer patrones de comportamiento y reglas de asociación.

Objetivo General

Analizar el comportamiento de navegación de los usuarios en una tienda en línea de ropa para embarazadas, utilizando herramientas de minería de datos en R, con el fin de identificar patrones de compra, generar recomendaciones de productos y extraer reglas de asociación que mejoren la experiencia del cliente y el rendimiento comercial de la empresa.

Exploración de la base de datos

Trabajamos con el archivo e-shop clothing 2008.csv, el cual contiene información sobre las sesiones de navegación realizadas por usuarios en una tienda en línea de ropa para embarazadas durante los meses de abril a agosto de 2008.

Al cargar la base de datos, observamos que cuenta con 165.474 registros distribuidos en 14 columnas. Cada fila representa un clic realizado durante una sesión de navegación. Las columnas iniciales del dataset, como se observa al cargar los datos, presentaban nombres genéricos en inglés (year, month, day, order, etc.). Esto dificulta una interpretación directa y un manejo intuitivo ya que el análisis y los gráficos se iban a realizar en español.

Para abordar esta cuestión y facilitar la comprensión del conjunto de datos, se procedió a renombrar las variables. Este renombramiento se realizó utilizando la información proporcionada en el archivo de texto Data description "e-shop clothing 2008".txt (originalmente en inglés), que detalla el significado de cada columna.

Tabla 1. Descripción de variables del registro de compras

Variable original	**Descripción**	**Tipo**
year	Representa el año en el que se realizó la compra en la página web	Cuantitativa Discreta
moth	Representa el mes en el que se realizó la compra en la página web	Cuantitativa Discreta
day	Representa el dia en el que se realizó la compra en la página web	Cuantitativa Discreta
order	Representa la secuencia de clics durante una sesión	Cuantitativa Discreta
country	Representa el país desde el cual se realizó la compra en la página web	Cualitativa Nominal
session.ID	Identificador de la sesión de navegación. Agrupa todos los clics realizados por un mismo usuario durante una sesión específica	Cualitativa Nominal
page.1.main. category	Categoría principal del productor visualizado. los valores posibles son: - 1(pantalones) - 2(faldas) - 3(blusas) - 4(ofertas)	Cualitativa Nominal
page.2..clothing.mod	Código del producto visualizado. Esta variable identifica a cuál de los 217 modelos de ropa corresponde el clic	Cualitativa Nominal
colour	Color del producto visualizado. Está codificado con valores del 1 al 14, que representan colores como beige, negro, azul, rojo, blanco, entre otros.	Cualitativa Nominal
location	Posición de la fotografía del producto en la pantalla. La interfaz está dividida en seis secciones,	Cualitativa Nominal
model.photography	Tipo de toma de la fotografía del modelo. Puede ser "frontal" (en face) o "de perfil".	Cualitativa Nominal
price	Precio del producto en dólares estadounidenses (USD).	Cuantitativa Discreta

price.2	Indica si el precio del producto es mayor al precio promedio dentro de su categoría. Codificada con: 1 (sí) y 2 (no), posteriormente convertida a "1" (sí) y "0" (no).	Cualitativa Ordinal
page	Número de página del sitio web donde se realizó el clic. Los valores van del 1 al 5.	Cualitativa Ordinal

El archivo .txt fue igualmente fundamental para realizar la conversión de códigos numéricos a factores categóricos. Muchas de las columnas, como country, page 1 (main category), colour, location, model photography y price 2, contenían valores numéricos (ej. 1, 2, 3) que por sí solos no revelaban su significado. Gracias a la descripción detallada en el .txt, fue posible mapear estos números a etiquetas de texto significativas

A continuación, describimos los principales aspectos observados durante la exploración:

- **Estructura de los datos:** El conjunto de datos final, después del renombramiento y la transformación de variables, se presenta como un data.table con 165.474 observaciones y 14 variables. Las variables numéricas originales se mantuvieron para mes, día, clicks_sesion, sesión_ID, precio, y num_pagina_sitio.

  Decidimos eliminar la variable **año** debido a su redundancia, dado que el archivo de datos abarcaba únicamente el período comprendido entre abril y agosto del mismo año.

- **Datos faltantes:** Se realizó una verificación de valores nulos (NA) en cada columna. Afortunadamente, no se detectó ningún valor faltante en el dataset, lo que simplifica el proceso de limpieza y preparación de datos.

- **Filas duplicadas:** También se verificó la presencia de filas completamente duplicadas. El análisis reveló que no existen filas duplicadas en el conjunto de datos, asegurando que cada registro sea único.

- **Rango de fechas:** Los datos corresponden exclusivamente al año 2008. Los meses registrados van desde abril (4) hasta agosto (8). Los días del mes varían entre 1 y 31, cubriendo el período completo de estos meses.

- **Sesiones:** El dataset contiene un total de 34.397 sesiones únicas (sesión_ID), lo que indica la cantidad de interacciones distintas de usuarios con el sitio web.

- **Precios:** La variable precio (en dólares estadounidenses) muestra una amplia distribución. Según el summary(df), los precios oscilan desde un mínimo de 1 USD hasta un máximo de 500 USD, con una media de 44.8 USD y una mediana de 35 USD, lo que sugiere que hay algunos productos con precios significativamente más altos que la mayoría. La columna clicks_sesion (order original) va de 1 a 217, indicando la profundidad de clics dentro de una sesión.

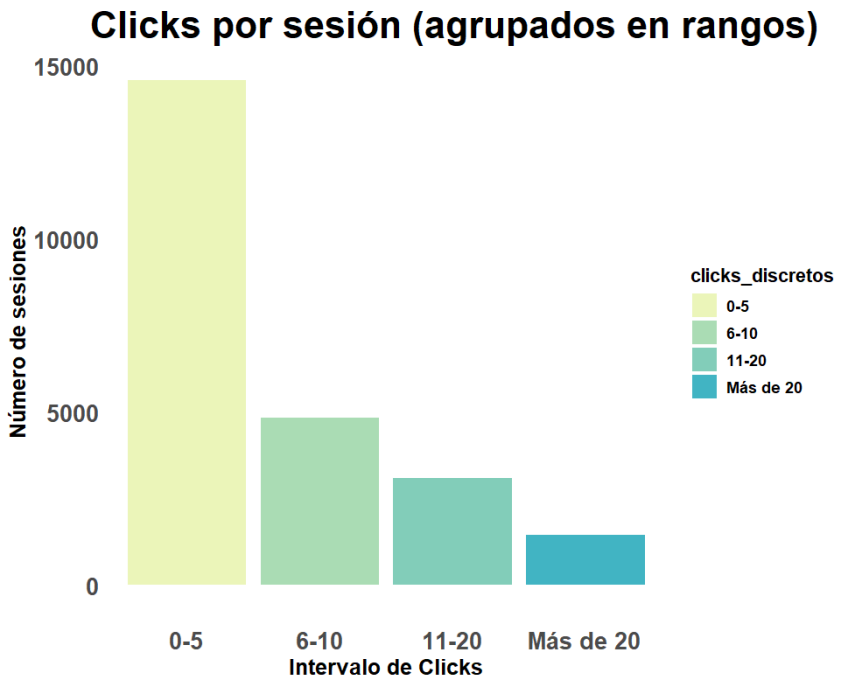
Mapeo y conversión de variables numéricas a factores:

Para la preparación de los datos fue importante convertir las representaciones numéricas de varias variables a sus respectivas etiquetas descriptivas, facilitando la interpretación:

- ★ *País* originalmente country*:* En el dataset original la columna se llamaba country. Luego, la columna país se transformó de códigos numéricos a nombres de países, incluyendo 47 categorías distintas como "Australia", "Alemania", "España", "Reino Unido", "USA", entre otros, y categorías genéricas como "no identificado" o dominios (.com, .org).
- ★ *Categorías de producto (categoria) originalmente (Originalmente page 1 (main category)):* La columna categoria se mapeó de códigos numéricos a categorías de ropa: "pantalones", "faldas", "blusas" y "ofertas".
- ★ *Colores (color) originalmente colour:* La variable color se convirtió de valores numéricos a una lista de 14 colores descriptivos, como "beige", "negro", "azul", "rojo", "blanco", etc.
- ★ *Ubicación de la foto (ubicacion_foto) originalmente location:* Esta variable se transformó para reflejar la posición de la foto del producto en la pantalla, con 6 categorías como "arriba izquierda", "arriba en el medio", "abajo derecha", etc.
- ★ *Fotografía (fotografia) Originalmente model photography:* La variable fotografía (originalmente model photography con valores 1 y 2) se mapeó a "frontal" para el valor 1 y "de perfil" para el valor 2. Esto permite entender la perspectiva de la foto del modelo.
- ★ *Precio mayor al promedio de la categoría (Originalmente price 2):* Esta columna, que indica si el precio de un producto era superior al promedio de su categoría, se transformó en una variable categórica (factor) con etiquetas "0" (no) y "1" (sí), a partir de los valores originales 1 y 2.

Exploración por medio de gráficos:

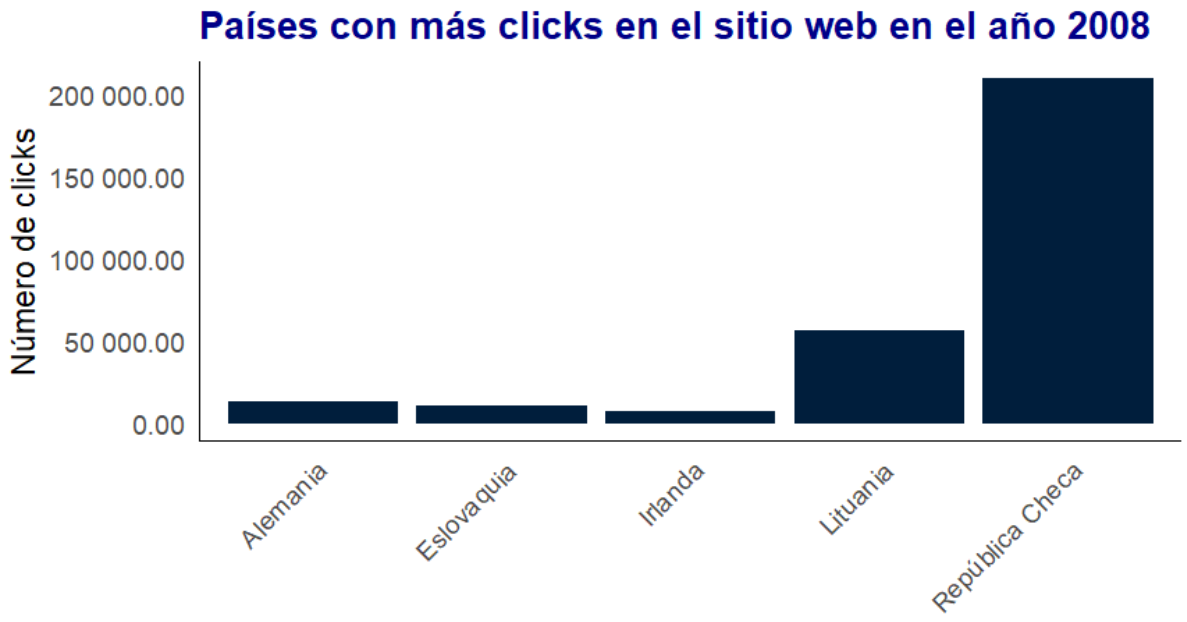
Figura 1. Cantidad de clicks por sesión



Fuente: Elaboración propia.

En la **Figura 1**, se puede observar que la mayoría de las sesiones registran entre 0 y 5 clicks, con una diferencia considerable respecto a los demás rangos. También podemos observar que a medida que aumenta la cantidad de clicks por sesión, la frecuencia disminuye progresivamente.

Figura 2. Sesiones por país.

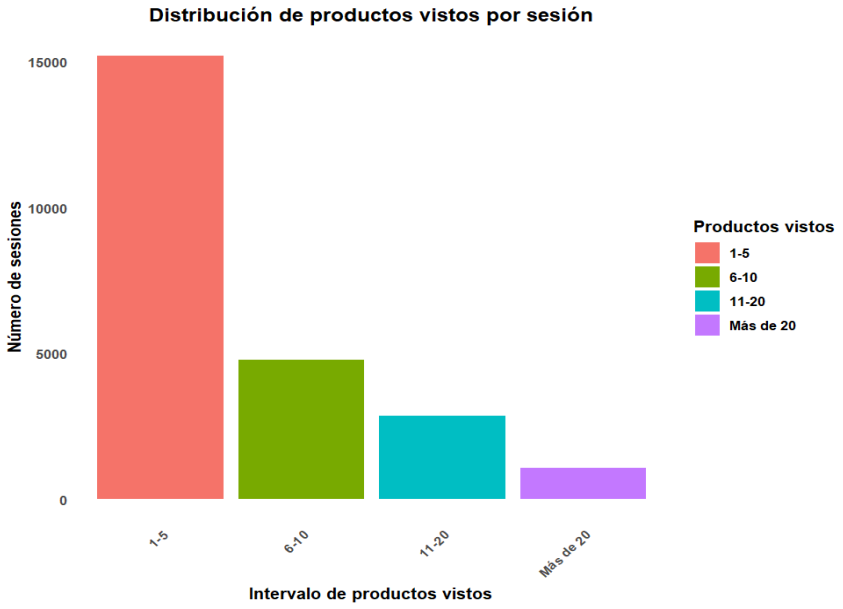


Fuente: elaboración propia

En la **Figura 2,** se ha excluido a Polonia, ya que, al ser el país de origen de la tienda en línea, concentra naturalmente la mayor parte de los clics. Su inclusión habría opacado la visualización del comportamiento del resto de los países, dificultando la identificación de otras regiones con participación relevante. Al remover a Polonia, se logra una mejor comparación entre los demás países y dominios.

Se puede observar que República Checa es el país con mayor número de clicks (mayor a 200 mil), seguida por Lituania (mayor de 50 mil), Alemania, Eslovaquia e Irlanda.

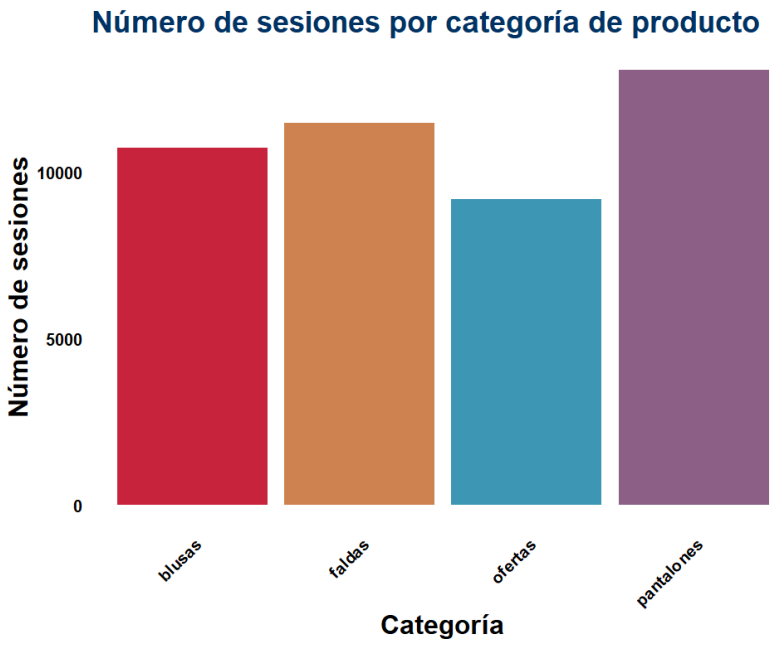
3. Productos vistos por sesión.



Fuente: Elaboración propia.

La **Figura 3**, muestra una mayor concentración de sesiones en el rango de 1 a 5 productos vistos, con una disminución progresiva en la cantidad de sesiones a medida que aumenta el número de productos visualizados.

Figura 4.Categoría de producto por número de sesiones.



Fuente: Elaboración propia.

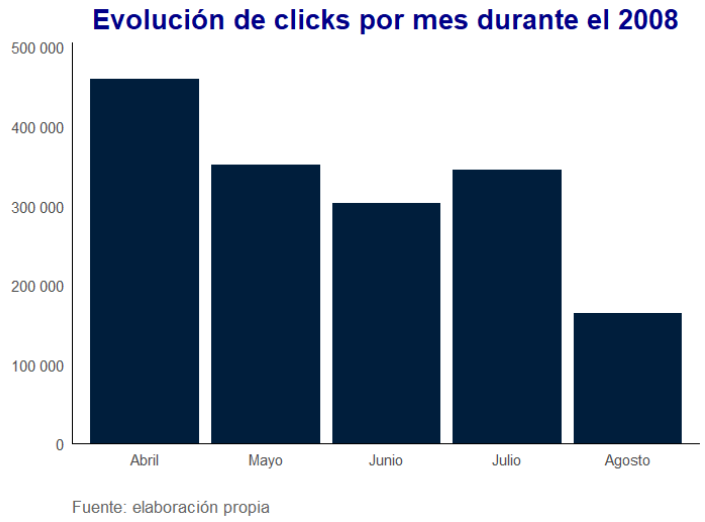
La **Figura 4** presenta la distribución del número de sesiones según la categoría de producto visualizado. Se observa que las categorías "pantalones", "faldas" y "blusas" concentran un volumen similar de sesiones, con una leve predominancia de la primera, mientras que la categoría "ofertas" registra el menor número.

La exploración gráfica de los datos nos permitió obtener una visión inicial clara y entender un poco más el comportamiento de los usuarios dentro del sitio web. Esto nos facilitó la **identificación de patrones preliminares** y **relaciones entre variables** que guiaron el desarrollo de las siguientes actividades analíticas.

Actividades:

¿Cómo ha sido la evolución de los clicks de navegación a lo largo de los meses estudiados?

Figura 5.Evolución de clicks por mes, durante el 2008



Fuente: Elaboración propia.

En la Figura 5 se evidencia que los meses con mayor interacción en la página son abril y mayo, siendo abril el que registra el pico más alto, con más de 450.000 clics. Este comportamiento puede atribuirse tanto al inicio del verano europeo como a la significativa cantidad de ofertas promocionales típicas de ese mes. A partir de abril se observa una disminución en la cantidad de clics alcanzando su punto más bajo en agosto, con menos de 200.000. No obstante, entre mayo y julio la actividad se mantiene relativamente elevada, con una leve recuperación en julio respecto a junio.

Encuentre el número de transacciones e ítems (pensando cada sesión como una transacción).

Para obtener una comprensión detallada del comportamiento de navegación de los usuarios, se realizó un análisis cuantitativo centrado en dos métricas clave: el número de transacciones individuales, entendidas como sesiones únicas de usuario. Dentro de estas transacciones, los **ítems** se refieren a los productos específicos (modelo_ropa) que los usuarios visualizaron.

- **Número de Transacciones (Sesiones Únicas):** Se identificaron 24026 sesiones únicas en el conjunto de datos. Este valor representa la cantidad total de interacciones independientes de usuarios con la tienda online.
- **Número de Ítems Únicos (Modelos de Ropa Distintos):** Se registraron 217 modelos de ropa únicos en todo el conjunto de datos. Esta cifra indica la diversidad de productos que, al menos una vez, fueron visualizados por los usuarios.

Encuentre un conjunto de itemsets frecuentes para un soporte mínimo de 2% y con una longitud mínima de 2 ítems.

Para identificar conjuntos de ítems que aparecen frecuentemente juntos, utilizamos los algoritmos Apriori y Eclat. Para aplicar estos algoritmos de la librería arules, es indispensable transformar el data.frame en un objeto de tipo transactions, ya que esta es la estructura de datos que esperan estos algoritmos. En este contexto, cada sesión_ID se considera una transacción, y los ítems dentro de cada transacción son los modelo_ropa visitados durante esa sesión. Esto nos permite identificar combinaciones populares de modelos de productos.

Preparación de Datos: Transformación a Objeto transactions:

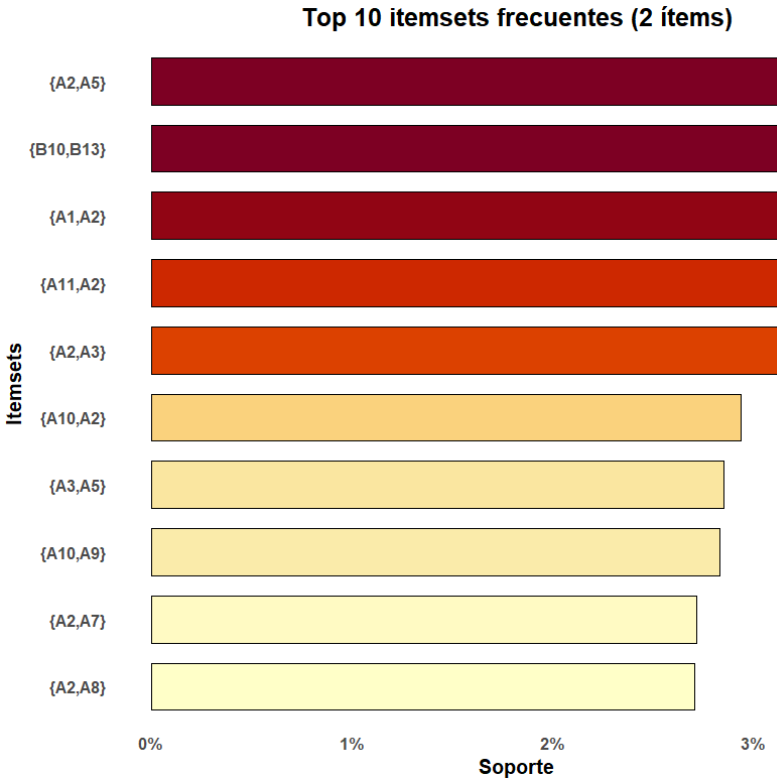
Creamos un data.frame auxiliar que solo contiene las columnas sesión_ID y modelo_ropa, asegurando que cada combinación de sesión y modelo se registre una única vez (distinct()). Posteriormente, agrupamos los modelo_ropa por cada sesión_ID en una lista, donde cada elemento de la lista representa una transacción.

Análisis de Itemsets Frecuentes con Eclat:

Definimos un soporte mínimo del 2% (0.02) y una longitud mínima de 2 ítems para los conjuntos frecuentes. Esto implica que sólo se considerarán aquellos grupos de al menos dos ítems que coexistan en al menos el 2% de todas las sesiones.

Al aplicar el algoritmo Eclat sobre el conjunto de transacciones de sesiones de usuario, identificamos varios itemsets frecuentes con el soporte mínimo establecido del 2%.

Figura 9. Itemsets frecuentes para un soporte mínimo de 2%



	items	support	count
[1]	{A2, A5}	0.03637726	874
[2]	{B10, B13}	0.03629401	872
[3]	{A1, A2}	0.03583618	861
[4]	{A11, A2}	0.03433780	825
[5]	{A2, A3}	0.03379672	812
[6]	{A10, A2}	0.02934321	705
[7]	{A3, A5}	0.02846916	684
[8]	{A10, A9}	0.02826105	679
[9]	{A2, A7}	0.02713727	652
[10]	{A2, A8}	0.02701240	649

Fuente: Elaboración propia.

El gráfico de barras horizontal y la tabla adjunta presentan exactamente los mismos resultados. Ambos muestran los Top 10 itemsets frecuentes que cumplen con los criterios de soporte y longitud establecidos.

Nos revelan que, al aplicar el algoritmo Eclat, se identificaron varios grupos de dos ítems que aparecen juntos. Los itemsets en la parte superior del gráfico y en los primeros puestos de la tabla (como `(A2, A6)` y `(B10, B13)`) son los más recurrentes, lo que sugiere una fuerte asociación entre estos modelos de ropa en las sesiones de usuario.

Encuentre las reglas de asociación para los datos de navegación correspondientes a Polonia, en la categoría “blusas”. Para un soporte mínimo de 2% y una confianza de 20%. Muestre las 10 reglas de mayor soporte.

Se utiliza algoritmo **Apriori** (permite obtener reglas claras y fáciles de interpretar entre productos, lo cual resulta útil para entender cómo navegan los usuarios. A diferencia de otros métodos, brinda resultados concretos que se pueden aplicar directamente en recomendaciones), se identificaron las 10 reglas de asociación con mayor soporte, respetando los umbrales definidos. Estas reglas fueron representadas mediante un grafo interactivo (**Figura 8**).

En la **Figura 8** se destaca **C5 y C17:**

Como producto de destino (RHS):

- C5:
    - Aparece como resultado frecuente cuando los clientes compran **C17, C7 o C14**.
    - Especialmente fuerte con C7 (confidence 29.4% y lift 1.95), indicando una **relación por afinidad**.
- C17:
    - Aparece como consecuencia de haber comprado **C12, C7 y C11**.
    - Especialmente notable con C12 (confidence 37.7%, lift 2.91), lo que sugiere que **cuando alguien compra C12, hay una fuerte probabilidad de que también adquiera C17**.

Como producto base (LHS):

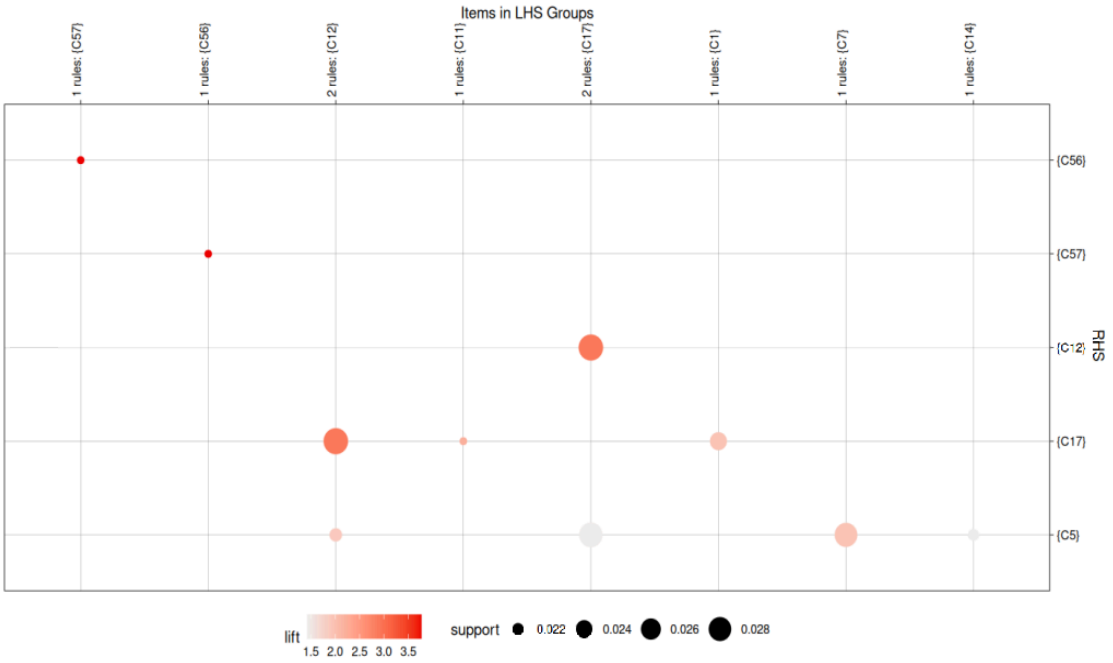
- C5:
    - Alguien que compra C5 también tiende a comprar C12, aunque la fuerza de esta relación es más débil (confidence 21.4%).
- C17:
    - También genera reglas salientes, hacia C12 y C5.

- La simetría con C12 (mismo lift de 2.91) confirma que **C17 y C12 están altamente correlacionados**, pero no de forma intercambiable (la confianza varía).

La presencia de reglas como {C12} => {C17} y {C17} => {C12} que se repiten, significa que cuando se compra **C12**, hay una alta probabilidad de que también se compre **C17 y así con otras reglas**, aunque ambos tienen el **mismo soporte** (porque la co-ocurrencia es la misma), tienen **diferente confidence (confianza).**

Figura 8. Reglas de mayor soporte y Grafo para Polonia en categoría “blusas”.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{C12}	=> {C17}	0.02972049	0.3772455	0.07878288	2.910523	252
[2]	{C17}	=> {C12}	0.02972049	0.2292994	0.12961434	2.910523	252
[3]	{C17}	=> {C5}	0.02842316	0.2192903	0.12961434	1.454900	241
[4]	{C7}	=> {C5}	0.02806935	0.2945545	0.09529426	1.954247	238
[5]	{C1}	=> {C17}	0.02405944	0.2559598	0.09399693	1.974780	204
[6]	{C12}	=> {C5}	0.02229036	0.2829341	0.07878288	1.877151	189
[7]	{C14}	=> {C5}	0.02193655	0.2147806	0.10213469	1.424980	186
[8]	{C57}	=> {C56}	0.02158273	0.3426966	0.06297912	3.768774	183
[9]	{C56}	=> {C57}	0.02158273	0.2373541	0.09093053	3.768774	183
[10]	{C11}	=> {C17}	0.02158273	0.2872841	0.07512678	2.216453	183



Fuente: Elaboración propia.

g. Encuentre las reglas para la República Checa, en la misma categoría del ítem anterior, pero para un soporte mínimo de 4% y una confianza de 25%. Muestre las 10 reglas de mayor soporte.

Se empleó el algoritmo Apriori, dado el umbral mínimo definido (4% de soporte y 25% de confianza), el modelo generó únicamente **7 reglas válidas**, lo que indica una menor densidad de asociaciones fuertes en este segmento específico del mercado.

En la **Figura 9,** se identifican**:**

Como producto de destino (RHS):

- **C56:**
  Aparece como resultado frecuente cuando los clientes compran C57 y C50.
  Especialmente fuerte con C57 (confidence 42.8% y lift 2.99), lo que indica una fuerte complementariedad entre ambos productos.
- **C57:**
  Surge como consecuencia cuando los clientes compran C56, mostrando una relación de reciprocidad. Sin embargo, la confianza disminuye (34.1%), aunque el lift se mantiene igual (2.99), lo que implica que el vínculo es fuerte pero no simétrico en comportamiento.
- **C5:**
  Es destino frecuente cuando se compran productos como C50, C49 y C17.
  Las asociaciones son moderadas: C50 → C5 tiene confidence de 49.1% (lift 1.45), mientras que C49 → C5 y C17 → C5 tienen menores confianzas.
- **C40:**
  Aparece como producto resultante tras comprar C29 y C41.
  Ambas relaciones presentan lifts altos (≈4.08) y confianzas superiores al 50%, indicando una co-ocurrencia muy fuerte.

Como producto base (LHS):

- **C57:**
  Cuando se compra C57, hay una fuerte tendencia a que también se compre C56.
  La confianza (42.8%) y el lift (2.99) son indicadores de una relación robusta y relevante para acciones de recomendación directa.
- **C56:**
  Actúa también como disparador hacia C57, aunque con menor confianza (34.1%)..
- **C50 y C49:**
  Ambos productos generan reglas salientes hacia C5, con confianzas del 49.1% y 41.7% respectivamente.

- **C29 y C41:**

  Funcionan como disparadores para C40.

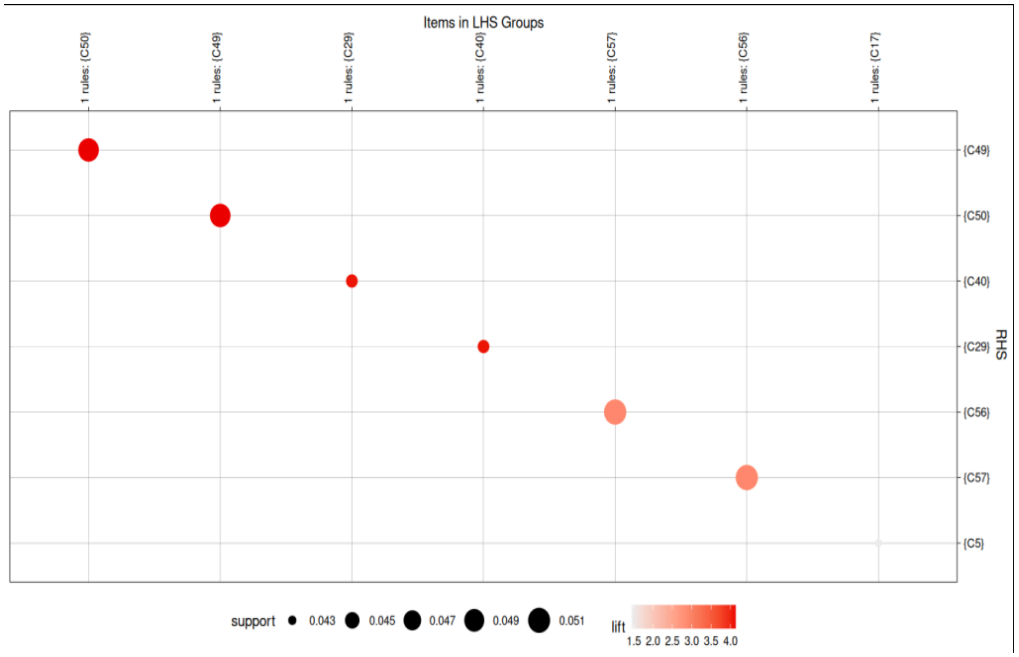
  Sus reglas presentan confianzas del 50.6% y 50.9%, con lifts superiores a 4.

- **C17:**

  También figura como producto base hacia C5, aunque la confianza (27.6%) es menor que la de otros pares y el lift (1.45) indica una relación útil pero no prioritaria comparada con las otras reglas.

Figura 9. Reglas de mayor soporte y Grafo para República Checa en categoría “blusas”.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{C57}	=> {C56}	0.05119736	0.4275862	0.11973576	2.909028	62
[2]	{C56}	=> {C57}	0.05119736	0.3483146	0.14698596	2.909028	62
[3]	{C50}	=> {C49}	0.04954583	0.4918033	0.10074319	4.135929	60
[4]	{C49}	=> {C50}	0.04954583	0.4166667	0.11890999	4.135929	60
[5]	{C29}	=> {C40}	0.04376548	0.5096154	0.08587944	4.087048	53
[6]	{C40}	=> {C29}	0.04376548	0.3509934	0.12469034	4.087048	53
[7]	{C17}	=> {C5}	0.04293972	0.2765957	0.15524360	1.456337	52



Fuente: Elaboración propia.

h. Compare los resultados de los dos países. ¿Qué conclusión sobre los consumidores puede obtener de los dos resultados?

La comparación entre ambos países evidencia diferencias significativas en los patrones de navegación y comportamiento de compra. En Polonia, se observa una estructura de navegación más conectada, con asociaciones frecuentes entre productos que

permiten identificar trayectorias claras dentro del catálogo. El modelo C5 se destaca como nodo central, participando tanto en reglas entrantes como salientes, lo que indica una alta interacción entre productos relacionados. La presencia de relaciones bidireccionales, como la que existe entre C12 y C17, sugiere un comportamiento de compra más sistemático y orientado a la complementariedad.

Por otro lado, los consumidores en la República Checa presentan un patrón más fragmentado, con un menor número de reglas de mayor soporte y sin un producto que domine claramente la red de asociaciones. Si bien existen vínculos fuertes, como los observados entre C56 y C57 o entre C29 y C40, estos son más puntuales y menos interconectados, con menor exploración del catálogo en comparación con el caso polaco.

i. Encuentre las secuencias más frecuentes que tienen más de un elemento (ítem) y un soporte mayor a 2%.

Para resolver esta consigna utilizaremos el algoritmo cSPADE (Sequential Pattern Discovery using Equivalence classes), ya que necesitamos encontrar secuencias a lo largo del tiempo, no simplemente asociaciones

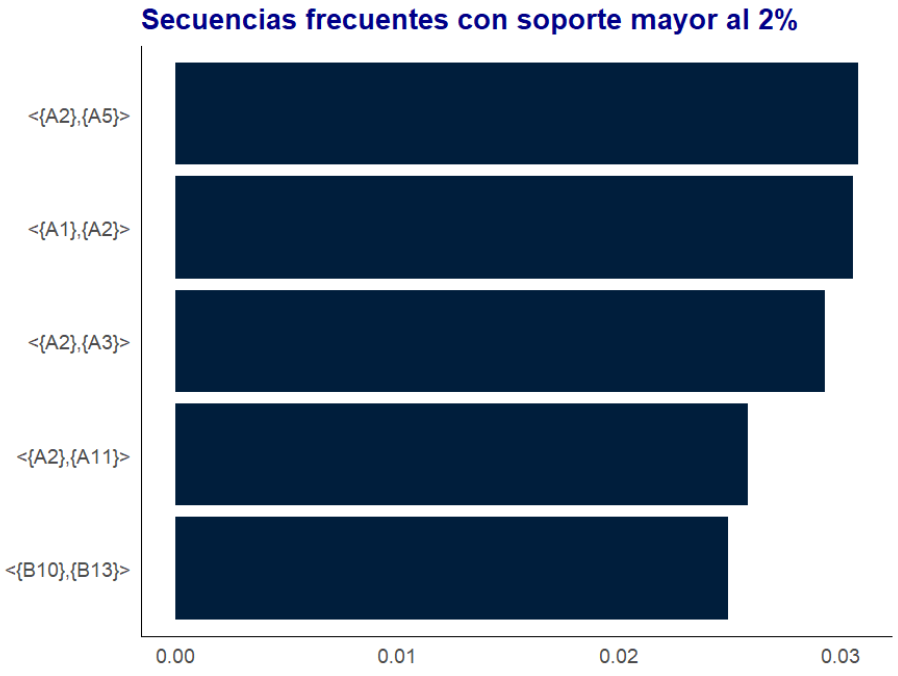
Para identificar patrones de comportamiento de los usuarios en la navegación de la tienda, transformamos los datos del archivo original en secuencias de ítems por sesión. Cada sesión fue considerada como una secuencia temporal, y cada click (orden de navegación) dentro de la sesión como un evento.

Los productos se identificaron por su modelo de ropa, y se agruparon por sesión y orden para construir una secuencia con los productos vistos en cada click. Posteriormente, estas secuencias se procesaron con cSPADE del paquete arruleSequence, que permite detectar patrones secuenciales frecuentes.

Se pudo encontrar un total de 165474 secuencias, y 217 items. El número de secuencias corresponde al mismo valor que el total de filas del dataset, esto se debe a que cada secuencia representa un evento único (click) dentro de una sesión, por otro lado la cantidad de ítems frecuentes corresponden a la cantidad de modelos de ropa que aparecen en patrones secuenciales recurrentes a lo largo de todas las sesiones analizadas. Esto nos indica que, aunque hay muchas secuencias-eventos individuales, solo un subconjunto de ítems se repite con suficiente frecuencia para superar el umbral de soporte establecido.

Se encontraron un total de 138 secuencias de las cuales solo 18 poseen un soporte mayor al 2% y tienen más de un elemento.

Figura 10. Secuencias frecuentes que tienen más de 1 ítem y con soporte mayor a 2%.



Fuente: elaboración propia

Modelos más frecuentes en secuencia:

El modelo A2 aparece en la mayoría de las secuencias frecuentes, siendo uno de los productos más visitados en el sitio y probablemente actúe como "puerta de entrada" hacia otros productos.

Patrones secuenciales comunes:

Existen combinaciones frecuentes como A2 → A5, A1 → A2 y A2 → A3, que indican que muchos usuarios navegan por estos productos en ese orden. Estos patrones podrían representar intereses relacionados en términos de estilo, categoría o tipo de prenda.

Conclusión:

El análisis exhaustivo de los datos de navegación de la tienda online de ropa para embarazadas nos proporcionó información relevante sobre el comportamiento de los usuarios y el comportamiento del comercio electrónico. Al aplicar diversas técnicas de minería de datos, desde la exploración inicial hasta la identificación de patrones secuenciales, hemos logrado una comprensión más detallada de cómo los clientes interactúan con los productos y las categorías del sitio.

La etapa de exploración de la base de datos fue fundamental para asegurar la calidad y consistencia de la información, permitiendo el renombramiento y mapeo de variables para una interpretación clara. Los gráficos exploratorios revelaron tendencias importantes, como la alta concentración de clics y productos visualizados en las primeras etapas de las sesiones, así como la distribución geográfica del tráfico, destacando a Polonia como el mercado principal y a República Checa con una presencia significativa. La evolución mensual de los clics durante 2008 mostró picos de actividad marcados en abril, seguidos por incrementos menores pero destacados en mayo y julio.

La identificación de itemsets frecuentes a través del algoritmo Eclat nos permitió identificar combinaciones de productos que los usuarios visualizan conjuntamente, revelando relaciones implícitas entre diferentes modelos de ropa.

Por otro lado, el análisis de reglas de asociación específicas para los mercados de Polonia y República Checa, dentro de la categoría "blusas", nos permitió comparar el comportamiento del consumidor. Se observaron diferencias notables en la densidad y estructura de las asociaciones, sugiriendo patrones de navegación más interconectados en Polonia y más fragmentados en la República Checa.

Finalmente, el análisis de secuencias frecuentes utilizando el algoritmo cSPADE identificamos los caminos más comunes que siguen los usuarios al navegar por el sitio. La recurrencia de ciertos modelos de ropa, como el A2, como puntos de entrada o nodos clave en estas secuencias, proporciona información para optimizar el flujo de navegación y las rutas de descubrimiento de productos.

Bibliografía:

Introducción a la minería de datos.pdf. (s/f). Google Docs. Recuperado el 13 de junio de 2025, de https://drive.google.com/file/d/1CgMlun-aYQXgaoSkc5hr-uSUMCP0TUdJ/view

(GUÍA PARA LA PRESENTACIÓN DE GRÁFICOS ESTADÍSTICOS). Gob.pe. https://www.inei.gob.pe/media/MenuRecursivo/metodologias/libro.pdf

https://www.ibm.com/docs/es/spss-modeler/18.5.0?topic=details-specifying-filters-rules