Universidad Peruana de Ciencias Aplicadas

“Informe de Trabajo Final”

Carrera: Ingeniería de Software

Sección: 14085

Docente:

Carlos Fernando Montoya Cubas

Curso:

Data Mining & Data Analysis

Nombre del Dataset:

Online Retail Dataset: Análisis del Sector Retail: UK Online Retail Dataset

GRUPO 3

Apellidos y Nombres	Código
Xiao Lian Li Zegarra	U202118784
Eduardo Andre Chero Emé	U20201F282
Tatiana Medalith Paucar De La Cruz	U20211f955

Lima, 4 de diciembre del 2025

1. Introducción

En este informe presentaremos la interpretación de un Dataset de una tienda retail en UK el cual tiene como objetivo principal identificar los patrones emergentes en las compras generadas en los periodos de tiempo entre el 1ro de diciembre del 2010 al 09 de diciembre del 2011.
Para este análisis usaremos minería de datos transaccionales, reglas de asociación, itemsets closed y maximales así como calcular el Growth rate para identificar Emerging Patterns y Jumpling Emerging Patterns

2. Descripción del Dataset

El dataset incluye una carga de datos de 541919 instancias divididas en los siguientes atributos:

- InvoiceNo: Número de factura. Nominal, un número entero de 6 dígitos asignado de forma única a cada transacción. Si este código comienza con la letra "c", indica una cancelación.
- StockCode: Código de producto (artículo). Nominal, un número entero de 5 dígitos asignado de forma única a cada producto.
- Description: Nombre del producto (artículo). Nominal.
- Quantity: Cantidad de cada producto (artículo) por transacción. Numérico.
- InvoiceDate: Fecha y hora de la factura. Numérico, el día y la hora en que se generó cada transacción.
- UnitPrice: Precio unitario. Numérico, precio del producto por unidad en libras esterlinas.
- CustomerID: Número de clientes. Nominal, un número entero de 5 dígitos asignado de forma única a cada cliente.
- Country: Nombre del país. Nominal, el nombre del país donde reside cada cliente.

Características principales:

Transacciones únicas	18532 facturas
Items unicos	3877 productos diferentes
Clientes únicos	3921 clientes registrados
Países unicos	38 países

Para la limpieza del dataset se generaron los siguientes criterios:

- Eliminación de transacciones que contengan valores nulos en ‘CustomerID’ y ‘Description’

- Eliminación de transacciones que contengan una cantidad de productos o ‘Quantity’ negativos
- Estandarización de ‘CustomerID’ a formato entero

3. Particionamiento del dataset

Se crearon dos particiones principales para este dataset, enfocados en el análisis en base a dos dimensiones diferentes de un comportamiento de compra.

Partición semestral: La división de semestres nos permite identificar patrones por estaciones a lo largo del año, resaltando las épocas primaverales y navideñas

```
======================================================================
PARTICIÓN 1: DIVISIÓN TEMPORAL (SEMESTRES)
======================================================================

RESULTADOS:
Semestre 1: 146,488 transacciones | Ventas: £3,421,091.76 | Ticket: £462.12
Semestre 2: 251,436 transacciones | Ventas: £5,490,316.14 | Ticket: £493.16
```

Como esperábamos, el segundo semestre muestra 72% más transacciones y 60% más en ventas totales. El ticket promedio también aumenta £69 (16%), confirmando que los clientes gastan más por compra en la temporada navideña.

Partición geográfica (UK vs Internacional): Otra forma de analizar este dataset es viendo el comportamiento cultural entre el mercado del país de origen y el mercado internacional.

```
======================================================================
PARTICIÓN 2: DIVISIÓN GEOGRÁFICA (UK vs INTERNACIONAL)
======================================================================

RESULTADOS:
UK: 354,345 transacciones          | Ventas: £7,308,391.55 | Ticket: £438.97
Internacional: 43,579 transacciones | Ventas: £1,603,016.35 | Ticket: £849.51
Países internacionales: 36
```

Los resultados confirman nuestras expectativas: UK representa el 89% de las transacciones, pero los clientes internacionales gastan casi el doble por transacción (£849 vs £439), sugiriendo que compran al por mayor o productos de mayor valor.

Itemsets frecuentes: Si un conjunto de productos aparece juntos en al menos un 2% de las transacciones se les denomina frecuentes, esto nos ayuda a entender si algunos productos guardan una relación con otros y poder ser usados en promociones para atraer mayor interés.

Itemsets closed: Un itemset es closed si ningún superconjunto tiene el mismo soporte, estos representan en síntesis patrones maximales en frecuencia, por lo que agregar algún producto extra disminuye su soporte, estos itemsets nos ayudan a reducir la redundancia para un mejor manejo de la información.

Itemsets maximales: si un itemset es maximal significa que es frecuente, pero ninguno de sus superconjuntos lo son, tienden a representar mayores combinaciones de productos que tengan un soporte mínimo, es útil para identificar cuando un usuario compra una cesta de productos más llena.

Partición	Itemsets frecuentes	Reglas	Itemsets closed	Itemsets maximales
S1	267	76	41	233
S2	291	100	54	237
UK	235	67	36	200
Internacional	598	554	319	384

- El semestre 2 tiene más patrones que el Semestre 1, lo que indica que los clientes compraron una mayor variedad de combinaciones de productos.
- El mercado internacional es mucho más diverso, con casi 2.5 veces más patrones que UK.
- Las reglas en Internacional son 8 veces más numerosas, señal de que las relaciones entre productos son más complejas.

4. Evaluación de patrones

Se implementó la evaluación de patrones emergentes comparando el comportamiento de itemsets entre particiones mediante el **Growth Rate (GR):**

GR = soporte en partición 2 / soporte en partición 1

4.1. Tipos de Patrones Identificados

Se encontraron tres tipos de patrones:

- Patrones de alto crecimiento: productos que aumentaron su frecuencia de compra en más del 50%
- Patrones de alto decrecimiento: productos que disminuyeron su frecuencia en más del 50%
- Patrones nuevos: productos que aparecieron por primera vez en la segunda partición

4.2. Análisis Temporal: Primer vs Segundo Semestre

Al comparar ambas particiones temporales se encontró:

Tipo de Patron	Cantidad	Significado
Total analizados	407	Patrones con mínimo soporte
Crecimiento	7	Mayor demanda en el segundo semestre
Decrecimiento	117	Productos con menor demanda, por estación o descontinuados
Nuevos	140	Nuevos productos introducidos

4.3. Análisis Geográfico: UK vs Internacional

Al realizar el análisis geográfico:

Tipo de Patron	Cantidad	Significado
Total Analizados	709	Mayor diversidad internacional
Crecimiento	37	Preferencias internacionales
Decrecimiento	114	Productos específicos de UK
Nuevos	474	Catálogo para exportación

4.4. Conclusiones del análisis

Comportamiento Temporal:

- 140 productos nuevos en S2 indican lanzamientos estratégicos para el segundo semestre
- 117 productos desaparecen o decrecen fuertemente: son artículos de verano
- El negocio es claramente estacional y predecible
- Los productos con mayor crecimiento (>100%) son oportunidades para aumentar inventario en próximas temporadas

Diferencias Geográficas:

- 474 productos exclusivos para mercado internacional: catálogo diferenciado por región
- Productos infantiles y decorativos crecen internacionalmente
- Productos con referencias culturales británicas funcionan solo en UK

Implicaciones para el Negocio:

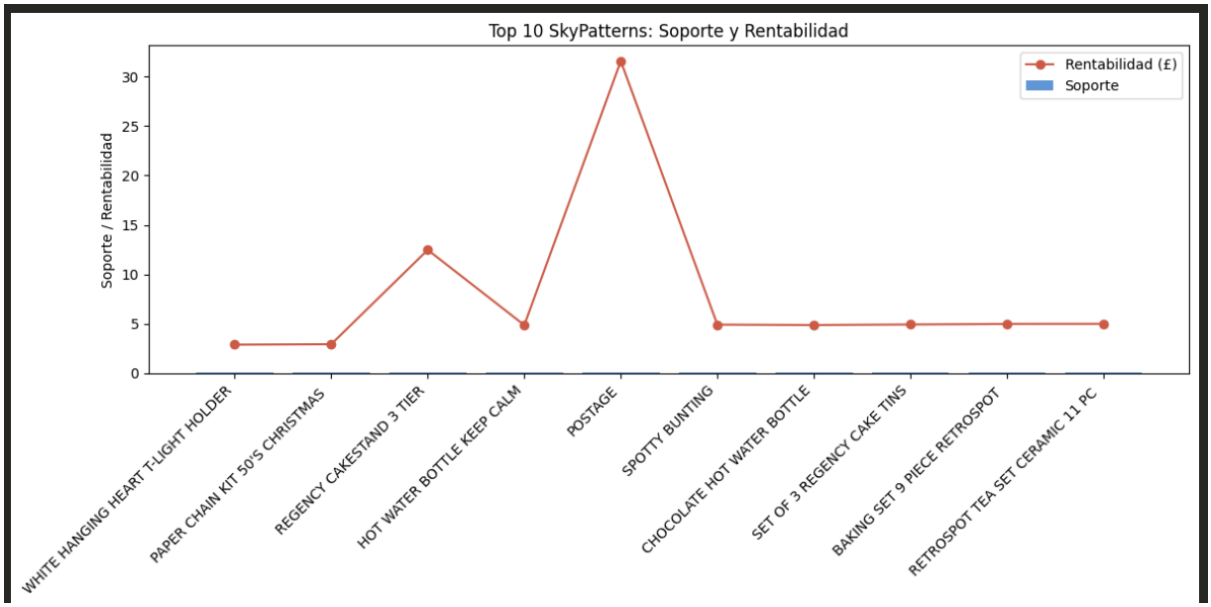
- Inventario: Ajustar stock 2-3 meses antes según patrones estacionales identificados
- Marketing: Estrategias diferenciadas UK vs Internacional según productos con alto GR
- Catálogo: Descontinuar productos con decrecimiento sostenido
- Predicción: Este análisis permite anticipar demanda futura con mayor precisión

5. Sky Pattern

En esta sección se aplicó la técnica **SkyPatterns**, que permite identificar los productos más valiosos considerando varias métricas al mismo tiempo:

- **Support:** qué tan frecuentemente se compra un producto.
- **Growth Rate:** si sus ventas aumentan o disminuyen.
- **Size:** si el patrón incluye uno o varios ítems.
- **Rentabilidad:** precio promedio del producto.

De un total de 407 patrones, se encontraron 23 patrones dominantes. Estos son los productos que destacan simultáneamente en las cuatro métricas y pueden considerarse los más importantes para el negocio.



1. Los patrones dominantes muestran que los mejores productos no solo se venden mucho, sino que también tienen precios dentro del rango óptimo (£3 – £5), lo cual maximiza los ingresos.
2. El envío tiene un precio promedio de £31.57 y un soporte del 6%. Esto evidencia que muchos clientes pagan envíos premium, generando un aporte económico importante.
3. Gran parte de los productos más importantes son sets, ya que ofrecen mayor margen de ganancia y mantienen niveles de venta similares a los productos individuales.

6. Conclusiones

- Algunos productos se venden mucho mejor cuando se compran juntos. Estos combos pueden ser entre 15 y 140 veces más fuertes que cuando se venden por separado. Son los que generan más ganancias.

- Los clientes del Reino Unido y los del resto del mundo compran de manera muy diferente, así que la tienda debería usar catálogos y estrategias distintas para cada grupo.

- Los clientes internacionales gastan el doble por compra comparado con los del Reino Unido. Pero, como representan solo 11% de las compras totales, existe una gran oportunidad de crecimiento si la tienda se enfoca más en ellos.

- En la segunda mitad del año (S2), el 34.4% de los productos son nuevos, lo que significa que el catálogo cambia bastante. Por eso, la tienda necesita manejar el inventario de forma muy flexible y rápida, porque los productos entran y salen constantemente.